



Saturating mutagenesis screen to dissect enhancer RNA function

by

Petra Celadova

A thesis submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

School of the Biosciences

The University of Sheffield

Sheffield, United Kingdom

March 2022

Acknowledgements

First of all, I am most deeply grateful to Dr. Dan Bose, for letting me undertake my doctoral research in his lab, for his gentle guidance, excellent advice, kind listening ear and cups of tea. His patience, thoughtfulness and steady support have been of invaluable help on my PhD journey, and I feel incredibly fortunate to have had Dan for my supervisor.

I was extremely lucky in being able to share the lab and office spaces with some of the loveliest and kindest people I ever had the fortune to meet. My amazing, brilliant, caring, daring, extraordinary fellow PhDs were the light on many a grey day in the lab, and my thankfulness to Katie Gelder, Nicola Carruthers and Laura Harrison goes beyond the realm of words. I am also grateful to our amazing Dr. Archana Shah, for leading by brave example and always being ready to offer help. Our wonderful Master Students over the years, including Christian Todaro, Naimah Hameed and Chris Callaghan, Luc Costello Heaven, Monika Olejnik and Molly John, gave me the opportunity to learn and teach, laugh and commiserate. I would also like to thank everyone in Dr. Emma Thomson's group, whom we've shared our lab space with over the last three years – most especially, Dr. Emma Thomson, for her caring support and ready help and advice, Holly Sutherland, for her kindness and cake, and Dr. Moni Feigenbutz, for her company and contagious laughter. I'm also grateful to Peter Daniels and Josh Zilate, whose presence always further lovelified the friendly atmosphere in our E21 office.

I would also like to thank Prof. Stuart Wilson, whose tissue culture space I was allowed to use in the first year of my PhD, and his group, who were very accommodating and welcoming to me. Amongst them, I'm deeply indebted to the lovely, brilliant Vicky Porteous-Medley, who kept me great company during long hours of tissue culturing, taught me lots of useful shortcuts and gave me piles of invaluable tips. I am also most thankful to Dr. Nicolas Viphacone, Dr. Helen Patrick and Dr. Llywelyn Griffith, for their useful tips and help around the Wilson lab, and their kindness, caring, good humour and general loveliness.

I am also grateful to my advisors, Dr. Emma Thomson and Dr. Ian Sudbery, whose doors were always open to me when I wished to ask for scientific or pastoral help.

Furthermore, I would like to thank the Medical School Flow Cytometry Core Service at The University of Sheffield – Sue Clark, Kay Hopkinson and Julie Swales were a great help during my FACS-sorting and flow cytometry experiments.

Additionally, I would also like to thank all the people that supported me non-scientifically as a PhD student and a member of the former Department of Molecular Biology and Biotechnology and current School of the Biosciences. Most notably, I'd like to express my thanks to Emily Cooper, who was a fast, knowledgeable and reliable help on all things admin, as well as an inspiration as a person, always

cheerful, supportive and kind. Before her, Linda Harris was an equally great and reliable help with PGR admin. Furthermore, the Think Ahead team prepared many of the personal and professional development opportunities that I have taken over the years, and I'm very grateful for that. I'm also thankful to Martin Price, who was always kind and helpful, great at fixing things and whose cheerful presence illuminated the corridors. Mark Johnson was also always very helpful whenever technical emergencies and non-emergencies arose.

There are so many other people who have directly or indirectly supported me on my PhD journey, including the Firth Court porters, the cleaning ladies, the Estates and Facilities Management, the IT Services, the Department of Finance, the Student Union and Student Service Information Desk, the Postgraduate Research Administration Team, the MBB PGR Committee, the safety officers – Ruth Roberts and Sarah Noble-Longster, the former MBB Departmental Manager – Jenny Miller – and the former Head of Department – Prof. Mike Williamson, and my gratitude goes to every single one of them.

I'm also grateful to Prof. Andy Sharrocks and Dr. Ian Sudbery, for accepting the task of examining my thesis and carrying out my viva.

Last but not least, I would like to thank my parents, for their financial support during my BSc and MSc studies.

Abstract

DNA elements known as enhancers are key players in physiological and aberrant transcriptional gene regulation, and understanding the mechanisms underlying their function is therefore of paramount importance. Enhancers can upregulate transcription of their target genes several hundred times, often in cell- and tissue-specific manner. Enhancers are widely transcribed in a bidirectional fashion, and the resulting RNA species is termed enhancer RNAs (eRNAs). Accumulating evidence suggests functional roles for eRNAs, including scaffolding of protein complexes, modulating enzymatic properties, acting as decoys for repressive factors and participating in chromosomal looping. However, the mechanistic basis for eRNA function remains unclear, and this work aimed to contribute to the elucidation of this question.

To address the enigma of eRNA functionality, we decided to work with an enhancer regulating the expression of the *TAL1* oncogene in Jurkat cells, a model system that is well-defined and features convenient reporter activity through a decrease in cell proliferation. A portion of the *TAL1* enhancer is transcribed into eRNAs (*TAL1*-eRNAs or *eTAL1s*), which were shown to play a role in the upregulation of *TAL1* in preliminary experiments in our lab. To map the functionally important regions of *TAL1*-eRNAs, we set off to dissect the transcribed portion of the *TAL1* enhancer via a saturating CRISPR/Cas9-mediated mutagenesis. Here, I describe the establishment and validation of a genetically modified Jurkat cell line for this CRISPR/Cas9 screen, and set forth the design of the screen itself. These results represent a set of solid building blocks for the execution of the saturating mutagenesis of the *TAL1* enhancer, the results of which will provide further insights into the workings of the *TAL1* enhancer and the *TAL1*-eRNAs, and by extension, will contribute to our understanding of the mechanisms behind eRNA functionality.

Table of contents

Acknowledgements.....	2
Abstract	4
Table of contents	5
List of figures.....	11
List of tables.....	16
List of appendices	17
List of abbreviations.....	18
1 Introduction	29
1.1 Non-coding genome	29
1.2 Enhancers	30
1.2.1 An introduction to enhancers – The importance and history of enhancers	30
1.2.2 Chromatin features and functional states of enhancers	31
1.2.3 Transcriptional regulation by enhancers	33
1.2.3.1 Establishment of enhancer-promoter contacts	33
1.2.3.2 Modes of enhancer regulation	37
1.2.4 Super-enhancers.....	39
1.3 Enhancer RNAs.....	40
1.3.1 The biogenesis of eRNAs.....	40
1.3.2 Structural features and other characteristics of eRNAs	42
1.3.3 Function of enhancer transcription in gene regulation	43
1.3.3.1 Class I eRNAs	43
1.3.3.2 Class II eRNAs.....	44
1.3.3.3 Class III eRNAs.....	45
1.3.3.3.1 eRNA roles in chromatin remodelling.....	45
1.3.3.3.2 eRNA roles in chromatin looping.....	47
1.3.3.3.3 eRNA roles in regulation of transcription machinery	48
1.3.3.3.4 eRNAs with <i>trans</i> roles.....	50

1.4	CRISPR/Cas9-based screening of non-coding genome.....	51
1.4.1	A case for CRISPR.....	51
1.4.2	Non-coding genome editing with CRISPR/Cas9-based screens	53
1.4.2.1	CRISPR/Cas9 screens introducing small insertions and deletions	53
1.4.2.2	CRISPR/Cas9 screens introducing long deletions	58
1.4.2.3	CRISPR/dCas9-based transcriptional regulation.....	61
1.4.2.3.1	CRISPR interference.....	62
1.4.2.3.2	CRISPR activation.....	64
1.4.2.4	Lessons from CRISPR/Cas9-based screens of non-coding genome	66
1.5	<i>TAL1</i> enhancer in Jurkat cells.....	68
1.6	CRISPR/Cas9-mediated dissection of the <i>TAL1</i> super-enhancer in Jurkat cells.....	73
2	Materials and methods	76
2.1	Cloning and plasmid preparation.....	76
2.1.1	Buffers and media.....	76
2.1.2	Molecular biology kits.....	76
2.1.3	Plasmids	77
2.1.4	Polymerase Chain Reaction (PCR)	78
2.1.5	Phenol:chloroform extraction and ethanol precipitation.....	79
2.1.6	Agarose electrophoresis	79
2.1.7	Transformation of bacterial cells.....	79
2.1.8	Plasmid preparation and sequencing	80
2.1.8.1	Preparation of pRetroX-TRE3G-FLAG-TAL1	80
2.1.8.2	Preparation of the plasmids for CRISPR validation of the genetically modified cell line.....	81
2.1.8.3	Preparation of the plasmids for the CRISPRi validation of the genetically modified cell line.....	83
2.2	Cell culture	84
2.2.1	Cell lines	84
2.2.2	Growth media and other tissue culture solutions.....	85
2.2.3	Selection antibiotics.....	86

2.2.4	Routine culturing procedures.....	86
2.2.4.1	Sub-culturing of Jurkat cells and their derivative cell lines	86
2.2.4.2	Sub-culturing of HEK293T cells and HEK293-derivative cell lines.....	87
2.2.5	Culturing procedures during antibiotic selection	87
2.2.6	Culturing procedures following FACS	89
2.3	Viral transduction.....	90
2.3.1	Transfection of packaging cells	90
2.3.2	Transduction with unconcentrated virus.....	91
2.3.3	Transduction with concentrated virus.....	91
2.4	Neon electroporation	93
2.4.1	siRNAs & ASO	94
2.5	Flow cytometry and fluorescence-activated cell sorting (FACS).....	95
2.5.1	Glossary and technical parameters	95
2.5.2	Preparations	95
2.5.3	Fluorescence-activated cell sorting (FACS)	95
2.5.4	Flow cytometry.....	96
2.6	Cell assays.....	97
2.6.1	Cell proliferation with cytotoxic drugs.....	97
2.6.2	Luciferase assay.....	97
2.6.3	Cell viability measurements using alamarBlue HS.....	97
2.6.4	Proliferation with doxycycline.....	98
2.7	RT-qPCR	100
2.7.1	RNA isolation	100
2.7.2	DNase I treatment and phenol:chloroform extraction	100
2.7.3	Reverse transcription.....	100
2.7.4	Quantitative PCR	101
2.8	Western Blot	102
2.8.1	Buffers.....	102
2.8.2	Antibodies	102
2.8.3	Cell harvest and bicinchoninic acid (BCA) protein assay.....	103

2.8.4	SDS electrophoresis	103
2.8.5	Protein transfer	103
2.8.6	Western blotting.....	104
2.9	CRISPR efficiency assay.....	105
2.10	Design tools and genome viewers.....	106
2.11	Statistical testing	107
3	Establishment of a cell line suitable for CRISPR/Cas9-mediated mutagenesis of the <i>TAL1</i> enhancer	108
3.1	Introduction	108
3.1.1	Notes on methods for gene introduction into Jurkat cells	109
3.2	Establishment of a Jurkat cell line stably expressing Tet-On 3G transactivator protein....	110
3.3	Establishment of a Jurkat cell line expressing doxycycline-inducible FLAG-TAL1.....	114
3.4	Establishment of a Jurkat cell line expressing doxycycline-inducible FLAG-TAL1 and Cas9.....	119
3.5	Discussion	121
4	<i>TAL1</i> downregulation reduces the proliferation rate of the newly established cell line	126
4.1	Introduction	126
4.2	Guide RNA sequences for CRISPR validation	127
4.3	sgRNAs are cleaving at the expected loci	129
4.4	First experiments show a massive cell death problem and inconsistencies	132
4.5	Selection using FACS.....	134
4.6	Optimized viral preparation improves transduction efficiency	136
4.7	Optimized transduction fails to solve the massive cell death problem	137
4.8	siRNA <i>TAL1</i> knockdown slows proliferation in Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9	139
4.9	Neon electroporation fails to solve the massive cell death problem and inconsistent behaviour in Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9	142
4.10	Exploring the role of hygromycin in the massive cell death problem	146
4.11	Understanding the influence of cell density on the survival and proliferation rates in Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9, and reconsidering the experimental time frame.....	151
4.12	The role of blasticidin, the last piece of the puzzle	155

4.13	A comparison of the effect of <i>TAL1</i> knockdown between all Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 clones.....	159
4.14	Discussion	162
5	Growth rate of the newly established cell line upon <i>TAL1</i> downregulation can be partially rescued by the doxycycline-inducible FLAG-TAL1.....	170
5.1	Introduction	170
5.2	Doxycycline induces FLAG-TAL1 expression in Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 and leads to increased proliferation rates over the first 2 days of culture	172
5.3	Effect of doxycycline on the proliferation rate of the newly established cell line.....	173
5.4	Optimization of the use of doxycycline for rescue in Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9.....	177
5.5	Using FACS as selection in Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 validation experiment generates a clearer trend, but kills all cells in combination with doxycycline	185
5.6	Doxycycline partially rescues proliferation rates in Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells of the final line upon <i>TAL1</i> downregulation	189
5.7	Transduction optimization and evaluation of the new packaging cell line, 293FT	191
5.8	Validation of doxycycline rescue using CRISPR interference (CRISPRi)	196
5.9	Discussion	200
6	Initial considerations and experimental outline of the CRISPR/Cas9-mediated mutagenesis of the <i>TAL1</i> enhancer	207
6.1	The adjusted experimental design	207
6.2	sgRNA design and library preparation.....	210
6.3	Using iBARs for internal replicates	212
6.4	Vectors and cloning strategies.....	213
6.5	Viral packaging	217
6.6	Multiplicity of infection (MOI)	217
6.7	Representation and cell numbers	219
6.8	Genomic DNA isolation.....	222
6.9	Library preparation.....	222
6.10	Next-generation sequencing.....	224
6.11	Hit validation.....	224

6.12	Discussion	225
7	Discussion.....	229
7.1	The particularities and problematics of working with eRNAs	229
7.2	Our chosen approach to studying the functionality of eRNAs.....	231
7.3	Further potential adjustments to our current system	234
7.4	Other CRISPR/Cas-based strategies for the dissection of the <i>TAL1</i> enhancer in Jurkat cells.....	235
7.5	Conclusion.....	243
	Bibliography.....	246
	Appendices	272

List of figures

Figure 1.1: Functional enhancer states.....	31
Figure 1.2: Proposed mechanisms of enhancer-promoter contact establishment.....	35
Figure 1.3: Modes of enhancer regulation.....	38
Figure 1.4: The effects of the act of eRNA transcription.	45
Figure 1.5: eRNA binding can activate CBP, resulting in increased histone acetylation.....	46
Figure 1.6: Roles of eRNAs in promoter-enhancer looping.	48
Figure 1.7: Role of eRNAs in RNAPII pausing.....	49
Figure 1.8: CRISPR/Cas9 editing and the two pathways to repair DSBs.	52
Figure 1.9: CRISPR screen strategies.....	56
Figure 1.10: CRISPR/Cas9-mediated pairwise genomic deletions.....	58
Figure 1.11: Transcriptional regulation by CRISPR/dCas9-fusions.	62
Figure 1.12: De novo MYB-binding site at the <i>TAL1</i> enhancer locus. (<i>redacted</i>).....	69
Figure 1.13: MYB and members of TAL1 co-activator complex bind at the mutation site at the <i>TAL1</i> enhancer. (<i>redacted</i>)	70
Figure 1.14: The <i>TAL1</i> enhancer and TAL1 co-activator complex.	71
Figure 1.15: eRNA transcription at the <i>TAL1</i> enhancer in Jurkat cells.....	72
Figure 1.16: CRISPR/Cas9-mediated saturating mutagenesis of the transcribed portion of the <i>TAL1</i> enhancer in Jurkat cells.....	74
Figure 3.1: The three individual steps in the establishment of the final cell line for the CRISPR/Cas9 mediated mutagenesis of the <i>TAL1</i> super-enhancer.....	109
Figure 3.2: Tet-transactivator protein in action.	111
Figure 3.3: Proliferation of Jurkat cells with various concentrations of G418.	111
Figure 3.4: Single cell sorting and clonal expansion.	112
Figure 3.5: Luminescence of monoclonal Tet-transactivator Jurkat cell lines transduced with luciferase.	113

Figure 3.6: Jurkat cell proliferation with puromycin over the course of 5 days.....	116
Figure 3.7: Expression of FLAG-TAL1 in doxycycline-induced monoclonal Jurkat-Tet3G-TRE3G-FLAG-TAL1 clones.	117
Figure 3.8: Background FLAG-TAL1 expression in selected uninduced monoclonal Jurkat-Tet3G-TRE3G-FLAG-TAL1 cell lines (85-XX).....	118
Figure 3.9: FLAG-TAL1 and endogenous TAL1 protein expression in monoclonal Jurkat-85-TRE3G-FLAG-TAL1, clones 15 and 16.	119
Figure 3.10: Jurkat-Tet3G-TRE3G-FLAG-TAL1 proliferation with a range of blasticidin concentrations.	120
Figure 3.11: Cas9, TAL1 and FLAG-TAL1 expression in selected Jurkat-85-15-Cas9 clones.....	121
Figure 4.1: Expected behaviour of the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line in CRISPR/siRNA validation experiments.	126
Figure 4.2: An overview of sgRNAs for the CRISPR validation of Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 addition to TAL1.....	128
Figure 4.3: CRISPR analysis of a sample of Jurkat cells edited with an sgRNA targeting the 5'-untranslated region of TAL1 (TAL4 sgRNA).....	130
Figure 4.4: TAL1 expression in wild-type Jurkat cells 3 days upon TAL1 knockout.	131
Figure 4.5: Proliferation of Jurkat cells with hygromycin.	132
Figure 4.6: Proliferation rates of Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells transduced with unconcentrated viral supernatants encoding various sgRNAs targeting <i>TAL1</i> gene and <i>TAL1</i> enhancer.	133
Figure 4.7: Proliferation and TAL1 expression of differentially CRISPR-edited, FACS-sorted Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells.	135
Figure 4.8: Transduction optimization.....	136
Figure 4.9: Proliferation rates of Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells transduced with concentrated viral particles containing various sgRNAs targeting <i>TAL1</i> gene and <i>TAL1</i> enhancer. ...	138
Figure 4.10: Nucleofection optimization for the delivery of small oligonucleotides.	140
Figure 4.11: siRNA-mediated <i>TAL1</i> knockdown in Jurkat and Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells.	141

Figure 4.12: Nucleofection optimization.	143
Figure 4.13: Proliferation rates of Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells transfected with various sgRNAs targeting <i>TAL1</i> gene and <i>TAL1</i> enhancer.....	144
Figure 4.14: Wild-type Jurkat cells are met with a growth challenge upon <i>TAL1</i> knockout.....	145
Figure 4.15: Proliferation of all the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 monoclonal cell lines compared to the proliferation of the maternal Jurkat cells.....	147
Figure 4.16: Proliferation of the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 clones with a range of hygromycin concentrations.....	148
Figure 4.17: Influence of hygromycin concentration on proliferation of <i>TAL1</i> -edited Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells and <i>TAL1</i> -edited maternal Jurkat cells.	150
Figure 4.18: Proliferation patterns of unchallenged Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells (clone 85-15-4).....	152
Figure 4.19: The influence of maintaining an optimal cell density on proliferation of <i>TAL1</i> -edited Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells and <i>TAL1</i> -edited maternal Jurkat cells, along with blasticidin supplementation.....	154
Figure 4.20: The influence of blasticidin addition into the selection media in the course of the <i>TAL1</i> -editing of Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells.....	156
Figure 4.21: Proliferation of blasticidin-conditioned Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells after <i>TAL1</i> -CRISPR challenge.....	158
Figure 4.22: Proliferation of <i>TAL1</i> -challenged monoclonal Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell lines compared to maternal Jurkat cells.	160
Figure 4.23: <i>TAL1</i> expression in maternal Jurkat cells and in clones of Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 upon CRISPR/Cas9-mediated knockout.....	161
Figure 5.1: CRISPR/Cas9-mediated saturating mutagenesis of the transcribed portion of the <i>TAL1</i> enhancer in Jurkat cells (copy from Figure 1.16).	171
Figure 5.2: Titration of doxycycline for the purposes of FLAG-TAL1 induction in the final line, Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9.....	172
Figure 5.3: Viability of Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 upon a 24-hour and 48-hour incubation with increasing concentrations of doxycycline.	173

Figure 5.4: The effects of cultivation of TAL1-challenged Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells in doxycycline.....	174
Figure 5.5: The effect of long-term doxycycline exposure on variously edited Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells.....	176
Figure 5.6: Titration of doxycycline for the purposes of FLAG-TAL1 induction in the final line, Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9, II.....	178
Figure 5.7: Proliferation of edited Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells with a range of doxycycline concentrations in selective media.....	180
Figure 5.8: Proliferation of edited Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells with a range of doxycycline concentrations in selective media.....	182
Figure 5.9: TAL1 expression in edited Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells cultivated with a range of doxycycline concentrations in selective media.....	182
Figure 5.10: Proliferation rates of variously edited Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells cultivated in a range of doxycycline concentrations.....	184
Figure 5.11: Proliferation of Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 after TAL1 challenge and FACS-sorting.....	186
Figure 5.12: Proliferation of control or <i>TAL1</i> enhancer-edited, FACS-sorted Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells with doxycycline.....	188
Figure 5.13: The effects of a range of doxycycline concentrations on the proliferation of variously edited Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells.....	190
Figure 5.14: Transduction optimization for further FACS-based doxycycline-rescue experiments...	192
Figure 5.15: Representative results from the transduction optimization experiment as described in Figure 5.14.....	194
Figure 5.16: 293FT packaging does not result in increased autofluorescence.....	195
Figure 5.17: The sensitivity of Jurkat-Tet3G-TRE3G-FLAG-TAL1 to blasticidin.....	198
Figure 5.18: <i>TAL1</i> -CRISPR interference in Jurkat-Tet3G-TRE3G-FLAG-TAL1 cell line using single sgRNAs.....	199
Figure 5.19: <i>TAL1</i> -CRISPR interference in Jurkat-Tet3G-TRE3G-FLAG-TAL1 cell line using combinations of sgRNAs.....	200

Figure 6.1: Comparison of the original and the adjusted design for the CRISPR/Cas9-mediated mutagenesis of the <i>TAL1</i> super-enhancer in the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line.	208
Figure 6.2: Workflow for the CRISPR/Cas9-mediated saturating mutagenesis of the <i>TAL1</i> enhancer, reimagined.	209
Figure 6.3: Experimental sgRNAs targeting the <i>TAL1</i> super-enhancer in Jurkat cells.	212
Figure 6.4: The sgRNA ^{iBAR} -Cas9 complex.	213
Figure 6.5: Improvements in the sgRNA scaffolding for CRISPR/Cas9-mediated mutagenesis of the <i>TAL1</i> enhancer.	214
Figure 6.6: The adjustment of the original lentiGuide-Hygro-dTomato plasmid.	214
Figure 6.7: Comparison of the original and adjusted cloning strategies.	216
Figure 6.8: Illumina sequencing library preparation strategy.	223
Figure 7.1: Workflow for the CRISPR/Cas9-mediated saturating mutagenesis of the <i>TAL1</i> enhancer, reimagined (copy from Figure 6.2).	232
Figure 7.2: Targeted integration of sgRNAs into the genome based on HDR recombination that could be used as an alternative to our system.	236
Figure 7.3: Transcriptional regulation of Cas9 expression that could be used as an alternative to our system.	238
Figure 7.4: Translational regulation of Cas9 expression that could be used as an alternative to our system.	241
Figure 7.5: Transcriptional regulation of sgRNA expression that could be used as an alternative to our system.	242

List of tables

Table 2.1: Plasmids used in this doctoral work.	77
Table 2.2: The composition of a 50 µl PCR reaction.	78
Table 2.3: The parameters for a PCR reaction run.	78
Table 2.4: Primers for pRetroX-TRE3G-FLAG-TAL1 cloning.....	80
Table 2.5: Primers for the cloning of lentiGuide-Hygro-dTomato plasmids for CRISPR/Cas9 validation of the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line.	82
Table 2.6: Primers for the cloning of Lenti-(BB)-EF1a-KRAB-dCas9-P2A-BlastR plasmids for CRISPRi validation of the Jurkat-Tet3G-TRE3G-FLAG-TAL1 cell line.....	83
Table 2.7: Mammalian cell lines used in this doctoral work.	84
Table 2.8: Preparation of transfection mixtures.	90
Table 2.9: siRNAs and ASO used for transfection by electroporation with the Neon device.	94
Table 2.10: qPCR primers used in this doctoral work.	101
Table 2.11: Antibodies used in this doctoral work.	102
Table 2.12: Primers for CRISPR/Cas9 efficiency assays used in this doctoral work.	105
Table 2.13: Signs used to designate p-values in statistical testing.....	107
Table 4.1: sgRNAs used for the CRISPR/Cas9 validation of the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line's dependence on TAL1.	129
Table 6.1: Starting cell numbers required for the CRISPR/Cas9-mediated saturating mutagenesis of the <i>TAL1</i> super-enhancer in Jurkat cells, for common representation and MOI ranges.....	220

List of appendices

Appendix 1: sgRNA sequences for the CRISPR/Cas9-mediated mutagenesis of the <i>TAL1</i> enhancer.	272
Appendix 2: The sequence of the polynucleotide fragment synthesized for the adjustment of the lentiGuide-Hygro-dTomato plasmid.	296

List of abbreviations

18S	ribosomal RNA molecule that sediments at 18 svedberg
3C	chromosome conformation capture
3D	three-dimensional
4-OHT	4-hydroxytamoxifen
AAVS1	adeno-associated virus integration site 1
AID	activation induced deaminase
AR	androgen receptor
arCas9	allosterically regulated Cas9
ARID5B	AT-rich interactive domain-containing protein 5B
<i>ARIEL</i>	ARID5B-inducing enhancer associated long non-coding RNA
ASO	antisense oligonucleotide
ATAC	assay for transposase-accessible chromatin
AU	arbitrary unit
BB	base-balancer
BCA	bicinchoninic acid
BCL11A	B-cell lymphoma/leukaemia 11A (transcription factor)
bHLH	basic helix-loop-helix
blast	blasticidin
BRAF	v-RAF (rapidly accelerated fibrosarcoma) murine sarcoma viral oncogene homolog B1
BRD4	Bromodomain-Containing Protein 4
Cas9	CRISPR-associated protein 9
CBP	cyclic adenosine monophosphate response element binding protein (CREB)-binding protein
<i>CCAT1</i>	colon cancer associated transcript 1 (lncRNA)

<i>CCAT1-5L</i>	<i>CCAT1</i> (colon cancer associated transcript 1) lncRNA with a 5'-extension of 4,700 bp (5'-long) compared to <i>CCAT1</i>
CCR5	chemokine (C-C motif) receptor 5
CD8/69	cluster of differentiation 8/69
CDK9	cyclin-dependent kinase 9
cDNA	complementary DNA
Cga	chorionic gonadotropin alpha
Ch	change of media (in transduction optimization experiments)
CHD1	chromodomain-helicase-DNA-binding protein 1
ChIA-PET	chromatin interaction analysis by paired-end tag sequencing
ChIP	chromatin immunoprecipitation
ChromRNA-seq	chromatin-enriched RNA fractions sequencing
CIP	chemically induced proximity
CMV	cytomegalovirus
cPPT/CTS	central polypurine tract/central termination sequence of HIV-1
cRACE	circular rapid amplification of cDNA ends
Cre	'causes recombination' (enzyme)
CRISPR	clustered regularly interspaced short palindromic repeats
CRISPRa	CRISPR activation
CRISPRi	CRISPR interference
crRNA	CRISPR RNA
C _T	threshold cycle
CTCF	CCCTC-binding factor
CTD	C-terminal domain (of the RNA Polymerase)
CUL3	cullin-3
Cy3	cyanine 3

dCas9	(catalytically) dead Cas9
DMEM	Dulbecco's Modified Eagles Medium
DMSO	dimethylsulfoxide
DN1S	dominant negative version of p53-binding protein 1 - short
DNA	deoxyribonucleic acid
DNMT3A	DNA methyltransferase 3 alpha
dNTP	deoxynucleoside triphosphate
DOX	doxycycline
DPBS	Dulbecco's Phosphate Buffered Saline
DPE	downstream promoter element (motif)
DRB	dichloro-1- β -D-ribofuranosyl-benzimidazole
DRE	dehydration-responsive element (motif)
^{DRR} eRNA	eRNA from the DNA damage response enhancer element (lncRNA)
DSB	double-strand break
DSIF	dichloro-1- β -D-ribofuranosyl-benzimidazole (DRB) sensitivity-inducing factor
DTT	dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
E12	splice variant of E2A (transcription factor)
E2-2	E2-box-binding (transcription factor) 2 (also known as TCF4 – transcription factor 4)
E2A	Ephrussi-box 2 (E2-box)-binding (transcription factor) alpha
E47	splice variant of E2A (transcription factor)
EDTA	ethylenediaminetetraacetic acid
EGFP	enhanced green fluorescent protein
ELISA	enzyme-linked immunosorbent assay
EMX1	Empty Spiracles Homeobox 1

ENCODE	Encyclopaedia of DNA Elements
<i>Env</i>	envelope (viral gene coding for viral envelope)
eQTL	expression quantitative trait locus
ER	oestrogen receptor
ER-LBD	ligand-binding domain of the oestrogen receptor
ER- α	oestrogen receptor α
eRNA	enhancer RNA
<i>eTAL1</i>	<i>TAL1</i> enhancer RNA (also known as <i>TAL1-eRNA</i>)
<i>Evf2</i>	<i>embryonic ventral forebrain-1</i> (lncRNA)
FACS	fluorescence-activated cell sorting
FBS	foetal bovine serum
FISH	fluorescence <i>in situ</i> hybridization
FITC	fluorescein isothiocyanate
FKBP	FK506 binding protein 12
FokI	(endonuclease from) <i>Flavobacterium okeanokoites</i>
Fos	FBJ (Finkel, Biskis, and Jinkins) murine osteosarcoma viral oncogene homolog
FRB	FKBP rapamycin binding domain
FSC-A	Forward scatter – area (under the voltage curve of the detected pulse)
FSC-H	Forward scatter – height (of the voltage curve of the detected pulse)
FSC-W	Forward scatter – width (of the voltage curve of the detected pulse)
FW	forward (primer)
<i>Gag</i>	group antigens (denotes a group of viral genes coding for matrix protein, capsid protein and nucleocapsid protein)
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
GATA1/3	GATA-binding factor 1/3 (transcription factors)
GFP	green fluorescent protein

GP2-293	<i>gag</i> and <i>pol</i> -expressing HEK293-based cell line
gRNA	guide RNA
H3K27ac	acetylation of histone 3 at lysine 27
H3K27me3	trimethylation at lysine 27 in histone 3
H3K4me1	monomethylation at lysine 4 in histone 3
H3K4me2	dimethylation at lysine 4 in histone 3
H3K4me3	trimethylation at lysine 4 in histone 3
H3K9me3	trimethylation at lysine 9 in histone 3
HAP1	haploid 1 (cell line)
HAT	histone acetyltransferase
HDAC3/6	histone deacetylase 3/6
HDR	homology-directed repair
HEB	HeLa E-box binding protein (also known as TCF12, transcription factor 12)
HEK293	human embryonic kidney (cell line)
HER2	human epidermal growth factor receptor 2
HIV-1	human immunodeficiency virus 1
hnRNPU/K	heterogeneous nuclear ribonucleoprotein U/K
HPRT1	hypoxanthine phosphoribosyltransferase 1
HPSE	heparase
HRP	horseradish peroxidase
HS2	DNaseI hypersensitivity site 2
Hsp90	heat shock protein 90
Huh7.5 _{oc}	human hepatoma-derived, stably expressing <i>Oct1</i> (cell line)
hygro	hygromycin
iBAR	internal barcode
ICE	inference of CRISPR edits

IgG	immunoglobulin G
IGV	Integrative Genomics Viewer
IL2RA	interleukin 2 receptor subunit alpha
Inr	initiator element (motif)
IRA	internal replicate analysis
kb	kilobase
kDa	kilodalton
Kdm6a/b	lysine demethylase 6a
KLF1	Krüppel-like factor 1 (transcription factor)
KLK3	Kallikrein-related peptidase 3
KRAB	Krüppel-associated box domain
LB	Luria Broth
LBD	ligand-binding domain
LCR	(β -globin) locus control region
LDA	lineage dropout analysis
LDB1	LIM (Lin-11/Isl-1/Mec-3) domain-binding protein 1 (transcription factor)
LDS	lithium dodecyl sulphate
Lea	leaving media unchanged until the day of FACS (in transduction optimization experiments)
LINE	long interspersed nuclear element
LNA	locked nucleic acid
lncRNA	long non-coding RNA
loxP	locus of X-over P1
LSD1	lysine-specific demethylase 1
<i>M. tuberculosis</i>	<i>Mycobacterium tuberculosis</i>
Mb	mega base pair

MED1/12	Mediator complex subunit 1/12
MERA	multiplexed editing regulatory assay
mESC	mouse embryonic stem cells (cell line)
miRNA	micro RNA
MLL1/2/3/4	mixed-lineage leukaemia 1/2/3/4 (enzyme)
MOI	multiplicity of infection
MoMuLV	Moloney murine leukaemia virus
MOPS	3-(N-morpholino)propane sulfonic acid
MPRA	massively parallel reporter assay
MR1	Multiplexing Read 1 Illumina sequencing primer
mRNA	messenger RNA
MTE	motif ten element (motif)
MuTE	mutation of the <i>TAL1</i> enhancer
MYB	myeloblastosis (transcription factor)
MYC	myelocytomatosis (transcription factor)
MyoD	myogenic differentiation factor
MyoG	myogenin
<i>Nanog</i>	Tìr nan Òg ('Land of the Young' in Irish Gaelic) (transcription factor involved in embryonic stem cell self-renewal)
ncRNA	non-coding RNA
ncRNA-a	activating non-coding RNA
<i>NEAT1</i>	Nuclear Enriched Abundant Transcript 1 (also known as nuclear paraspeckle assembly transcript 1; lncRNA)
NELF	negative elongation factor
NES	nuclear export sequence
NF1/2	neurofibromin 1/2

NHEJ	non-homologous end joining
NLS	nuclear localization sequence
Npas4	neuronal PAS (Per-Arnt-Sim) domain protein 4
Oct4	octamer-binding transcription factor 4 (also known as POU5F1)
P-TEFb	positive transcription elongation factor b
P/S	penicillin/streptomycin
p300	300-kD protein (enzyme)
p53	protein 53 (transcription factor)
PAM	proto-spacer adjacent motif
PBS	phosphate-buffered saline
PCAF	p300/CBP-associated factor
PCR	polymerase chain reaction
PD-1	programmed cell death protein 1
PE150	paired-end sequencing of 150 nt from each end of a product
PGK	3-phosphoglycerate kinase
pgRNA	paired guide RNA
PNK	polynucleotide kinase
<i>Pol</i>	polymerase (denotes a group of viral genes coding for reverse transcriptase, integrase and protease)
POU5F1	POU (Pit-1/Oct1/2/Unc-86) domain, class 5, transcription factor 1 (also known as Oct4)
PRC2	Polycomb repressive complex 2
PRO-seq	precision run-on sequencing
PROMPT	promoter upstream transcript
Ptet-1	Tetracycline-responsive promoter 1
PVDF	polyvinylidene (di)fluoride

<i>PVT1</i>	<i>Plasmacytoma</i> variant translocation 1 (lncRNA)
qPCR	quantitative PCR
RACE	rapid amplification of cDNA ends
RAD21	radiation-sensitive mutant 21
RBR	RNA-binding region
<i>Rev</i>	regulator of expression of virion proteins (viral gene)
REV	reverse (primer)
RNA	ribonucleic acid
RNAPII	RNA polymerase II
RNAPIII	RNA Polymerase III
ROS1	repressor of silencing 1 (transcription factor in <i>Arabidopsis thaliana</i>)
ROSA	reverse orientation splice acceptor
RPM	rounds per minute
RPMI	Roswell Park Memorial Institute (cell medium)
<i>Rpp25</i>	ribonuclease P protein subunit p25
RRE	Rev response element
rRNA	ribosomal RNA
RT	reverse transcription
RT-qPCR	reverse transcription quantitative real-time polymerase chain reaction
RUNX1	Runt-related transcription factor 1
<i>S. pyogenes</i>	<i>Streptococcus pyogenes</i>
SAP	shrimp alkaline phosphatase
SDS	sodium dodecyl sulphate
SERPINB2	SERPIN (serine proteinase inhibitor) family B member 2
sgRNA	single guide RNA
SHAPE	selective 2'-hydroxyl acylation analyzed by primer extension

shRNA	short/small hairpin RNA
SINE	short interspersed nuclear element
siRNA	small/short interfering RNA
SMC3	structural maintenance of chromosomes protein 3
SNP	single nucleotide polymorphism
snRNA	small nuclear RNA
snRNP	small nuclear ribonucleoprotein
<i>sp.</i>	species
SSC-A	Side scatter – area (under the voltage curve of the detected pulse)
SSC-H	Side scatter – height (of the voltage curve of the detected pulse)
SSC-W	Side scatter – width (of the voltage curve of the detected pulse)
SV40	simian virus 40
T-ALL	T-cell acute lymphoblastic leukaemia
TAD	topologically associated domain
TAE	Tris-acetic acid-EDTA (buffer)
TAF1	TBP-associated factor 1
TAL1	T-cell acute lymphocytic leukaemia protein 1 (transcription factor)
TALE	transcription activator-like effector
TALEN	transcription activator-like effector nuclease
TATA	TA-rich DNA motif
TBP	TATA-binding protein
TBS	Tris-buffered saline
TCT	a pyrimidine-rich motif (CTCTTTCC)
TdGF1	teratocarcinoma-derived growth factor 1
Tet	tetracycline
Tet-O	tetracycline operator

Tet-On	tetracycline-inducible gene expression system
Tet-On 3G	tetracycline-inducible gene expression system, 3 rd generation
Tet1	Ten-Eleven Translocation 1
TF	transcription factor
TFIIB	transcription factor II B
tracrRNA	<i>trans</i> -activating CRISPR RNA
TRE	tetracycline-responsive element
tRNA	transfer RNA
U	unit
UPBS	unique primer binding site
URE	unmarked regulatory region
UTR	untranslated region
VP16	virion protein 16
VSV-G	vesicular stomatitis virus G
WDR82	WD (tryptophan-aspartic acid) repeat-containing protein 82
YY1	Yin Yang 1 (transcription factor)
ZFN	zinc finger nuclease
<i>Zfp42</i>	zinc finger protein 42 homolog

1 Introduction

1.1 Non-coding genome

While protein-coding genes enjoy the main spotlight within the genome, a vast majority of the genome – over 98% (ENCODE Project Consortium, 2012) – doesn't translate into proteins. Once believed to be 'junk', the latest computational estimates suggest that at least 4.51% of the human non-coding genome is under purifying selection, strongly indicating functionality (Huber et al., 2019). Moreover, when probed by a host of molecular biology methods, about four fifths of the human genome displayed participation in at least one biochemical event in at least one of the 147 surveyed cell types (ENCODE Project Consortium, 2012). Importantly, 93% of disease- and trait-associated genetic variants identified in genome-wide association studies were found to fall within non-coding regions (Maurano et al., 2012). Taken together, these findings imply an extremely important role for the non-coding genome.

Transcribed and regulatory elements in the genome are associated with various combinations of specific chromatin signatures, and these can be used to identify and classify such elements. This underpins the idea of epigenetic code (Nightingale et al., 2006; Turner, 2007). The differential signatures include various biochemical marks on histones spanning or flanking the regulatory features (Barski et al., 2007; Heintzman et al., 2007; Kouzarides, 2007; Mikkelsen et al., 2007; Wang et al., 2008), DNA methylation (Ball et al., 2009; Meissner et al., 2008), DNA accessibility (Buenrostro et al., 2015; Simon et al., 2012; Song et al., 2011; Thurman et al., 2012), and transcription factor binding (Cheng et al., 2012; Fu et al., 2008; Wang et al., 2012).

Because of the vastness of non-coding genome, its thorough mapping was dependent on the development of high-throughput methods for chromatin examination. At the forefront of these efforts was (and remains to be) the international ENCODE consortium. ENCODE have developed a number of high-throughput approaches for the functional annotation of whole genomes, which they then applied to a wide range of human cell types (ENCODE Project Consortium, 2011, 2012; Roadmap Epigenomics Consortium et al., 2015). From this Herculean endeavour arose an expansive, publicly accessible database containing dozens of terabytes of information on functional genomic elements (<https://www.encodeproject.org/>).

With the application of machine learning, this abundant data could begin to be classified into biologically relevant categories (Ernst and Kellis, 2012; Hoffman et al., 2012, 2013). Apart from the bodies of transcribed genes, transcription start sites and repressed regions, the algorithms identified

and labelled various functional non-coding features within the collections, such as candidate transcriptional terminators, insulators, promoters and enhancers (Hoffman et al., 2013).

1.2 Enhancers

1.2.1 An introduction to enhancers – The importance and history of enhancers

Enhancers are sequences in the genome that can positively regulate the expression of nearby genes, often over long distances and independently of their orientation (Serfling et al., 1985).

Enhancer activity was first discovered in 1981, in a 72-bp long fragment from simian virus 40 (SV40) that increased transcription of linked genes in an orientation-independent manner up to 200 times (Banerji et al., 1981; Moreau et al., 1981). The discovery of the first cellular enhancer, which tissue-specifically regulates expression of a mouse immunoglobulin heavy-chain gene, came soon afterwards (Banerji et al., 1983; Gillies and Morrison, 1983). Since then, many more enhancers have been described. In fact, the ENCODE Project endeavour has identified almost 400,000 putative enhancers in human genome, which represents roughly twenty times the number of protein-coding genes (ENCODE Project Consortium, 2012).

Enhancer DNA interacts with an extensive network of transcription factors in a cell-type-specific manner, orchestrating embryonal development and maintaining cell homeostasis (Coppola et al., 2016; Reiter et al., 2017; Spitz and Furlong, 2012). Their immense functional importance is further underlined by the fact that mutations in enhancers often contribute to or even cause disease, including type 2 diabetes and Alzheimer's disease, as well as some autoimmune diseases and cancer types (Carullo and Day, 2019; Maurano et al., 2012; Miguel-Escalada et al., 2015; Sur and Taipale, 2016).

The general prerequisite for the transcriptional activation of gene expression by enhancers is the physical contact of an active-state enhancer and its associated transcriptional machinery with the promoter of the target gene and its associated protein complexes, but there are still many gaps in our understanding of how exactly this accomplishes the target gene activation (reviewed in (Panigrahi and O'Malley, 2021)).

1.2.2 Chromatin features and functional states of enhancers

For the complex, dynamic and tissue-specific coordination of the cellular transcriptional response to function in a correct and timely fashion, not all enhancers can be in an active state all of the time. Rather, they often alternate between different states, which are characterized by different chromatin properties and include latent, poised, primed and active enhancer states (Figure 1.1).

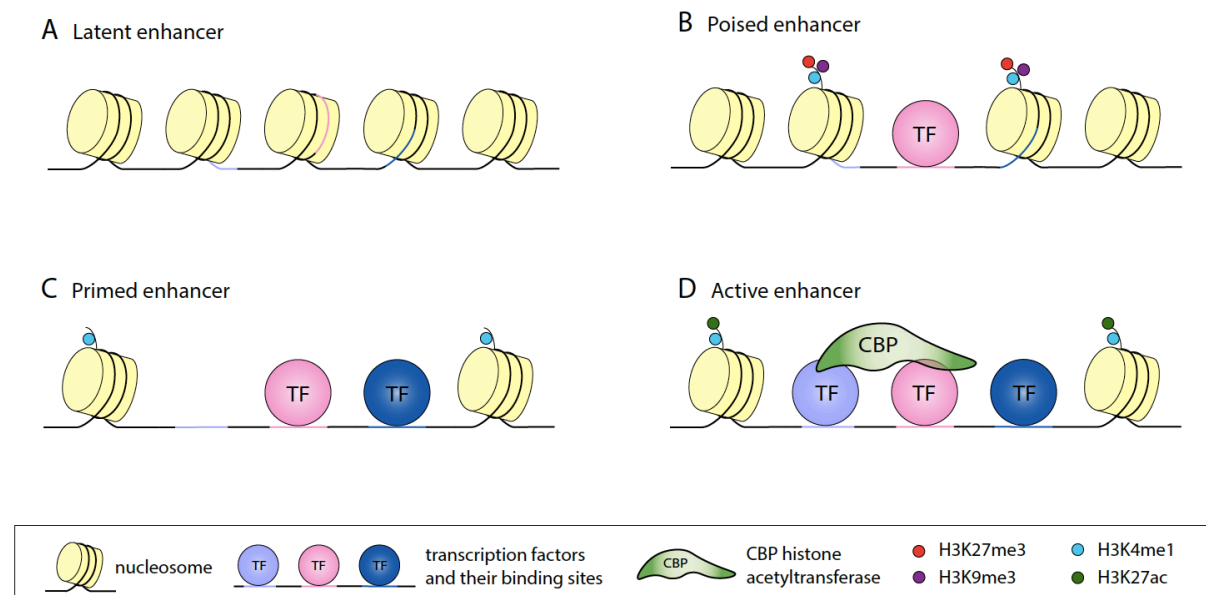


Figure 1.1: Functional enhancer states.

(A) Latent enhancers bear no distinguishing histone marks and bind no transcription factors (TF), but possess TF binding sites and await potential activation. **(B)** Poised enhancers are found in partially open chromatin, and may already be binding some TFs. Adjacent histones are decorated with the classic enhancer mark H3K4me1 and repressive marks H3K27me3 and H3K9me3. **(C)** Primed enhancers bind some TFs in open chromatin and sport the enhancer-defining H3K4me1 modification, waiting for additional signals to become fully activated. **(D)** Also located in open chromatin, active enhancers are enriched in TF and co-factor binding and their signature includes the H3K4me1 and H3K27ac histone marks.

A unifying feature of all enhancers is that their underlying DNA sequence is enriched in transcription factor binding sites. However, most transcription factors can only bind to DNA where the chromatin is open and accessible. This is the case for active enhancers (Figure 1.1D), which are depleted for the canonical, more stable nucleosomes and partly populated by highly labile nucleosomes containing the specialized histone variants H2A.Z and H3.3 (Jin and Felsenfeld, 2007; Jin et al., 2009). The conventional, less mobile nucleosomes flanking the clusters of transcription factor binding sites in

active enhancers bear specific post-translational modifications, including acetylation of histone 3 at lysine 27, H3K27ac (Creyghton et al., 2010) and a certain level of methylation at lysine 4 in histone 3. Initially, monomethylation at lysine 4 in histone 3, H3K4me1, was considered the defining hallmark of active enhancers (Heintzman et al., 2007, 2009), but later studies strongly suggested that the degree of H3K4 methylation was a function of the local transcription intensity, with more active enhancers often sporting H3K4me2/3 rather than, or in addition to, H3K4me1 (Core et al., 2014; Henriques et al., 2018; Pekowska et al., 2011). Further to active enhancer signatures, the enriched transcription factor occupancy attracts the binding of transcriptional co-factors, most notably the histone acetyltransferases CBP (cyclic adenosine monophosphate response element binding protein (CREB)-binding protein)/p300 (300-kD protein) (Visel et al., 2009). The DNA at active enhancers is hypomethylated, consistent with the enrichment in Ten-Eleven Translocation 1 (Tet1) methylcytosine dioxygenase, a DNA hydroxylase which catalyses the first step of the DNA demethylation process (Pulakanti et al., 2013; Stadler et al., 2011; Tahiliani et al., 2009). In sum, DNase I hypersensitivity, p300/CBP occupancy and an enrichment in H3K27ac and H3K4me1/2/3 histone marks constitute the main characteristic signature of active enhancers.

On the other side of the spectrum, some enhancers may be found in closed chromatin, with no transcription or co-transcription factor occupancy, unmarked by any histone modifications. Such latent enhancers (Figure 1.1A) become activated upon appropriate stimulation (Ostuni et al., 2013).

Other enhancers found in closed chromatin, closed enhancers, are marked with the active H3K4me1 enhancer mark, but at the same time actively repressed in their endogenous context by the deposition of the repressive H3K27me3 histone mark, which precludes the key active modification, H3K27ac (Arnold et al., 2013).

The important transition from closed to open chromatin at enhancers was repeatedly shown to be initiated by so-called 'pioneer' factors, a special class of transcription factors capable of binding to their respective recognition motifs within the context of closed chromatin (Ghisletti et al., 2010; Gualdi et al., 1996; Sérandour et al., 2011; Zaret, 2020). These factors often bind in cooperation with other lineage-determining, signal-dependent or collaborating transcription factors (Adams and Workman, 1995; Boyes and Felsenfeld, 1996; Heinz et al., 2010; Miller and Widom, 2003). Such cooperation between transcription factors can increase their chances for the successful eviction of the nucleosome from the closed chromatin region, which in turn opens the possibility of further transcription factor and co-factor binding. In this way, poised (Figure 1.1B) or primed (Figure 1.1C) enhancer states are established (reviewed in (Heinz et al., 2015)).

Poised enhancers are found in regions of low nucleosomal density and are enriched in p300 occupancy and H3K4me1, but, similarly to closed enhancers, are blocked from acquiring the active H3K27 acetylation mark by the trimethylation of the same residue (Bernstein et al., 2006; Rada-Iglesias et al., 2011). A large group of enhancers was also found to be in an intermediate state between poised and active – enriched in H3K4me1 and devoid of any H3K27 modifications, the expression levels of genes controlled by these intermediate-state enhancers were higher than the expression levels of genes associated with poised enhancers, but lower than the expression levels of genes linked with active enhancers (Zentner et al., 2011). A further distinguishing factor between these two groups was the presence of the H3K9me3 mark at poised enhancers (Zentner et al., 2011), a modification associated with repressed or silenced chromatin (Peters et al., 2002).

On the whole, enhancer states present an intricate dynamic continuum rather than a simple on/off switch, in consistency with the complexity of the transcriptional networks they regulate.

1.2.3 Transcriptional regulation by enhancers

1.2.3.1 Establishment of enhancer-promoter contacts

Apart from the various states of enhancer activation, the transcription of a gene is further regulated by the physical distances between its promoter and any of its potential enhancers. In the nucleus, chromatin regions are segregated into two main compartments based on transcriptional activity (Lieberman-Aiden et al., 2009). On the megabase level, chromatin is further organized into ‘topologically associated domains’ (TADs), bounded by the insulator protein CCCTC-binding factor (CTCF) binding sites, highly transcribed genes or short interspersed nuclear element (SINE) retrotransposons, which all serve as barriers from the spread of heterochromatin (Dixon et al., 2012; Nora et al., 2012). An even more detailed genome-wide study shows that chromatin might be divided into domains smaller than TADs, with a median length of 185 kb (Rao et al., 2014). These so-called ‘contact domains’ are also often bounded by CTCF binding sites, show consistent histone modification patterns and often form loops (Rao et al., 2014).

Many studies indicate that it is through looping of the chromatin in three-dimensional space that enhancers are brought in the proximity of the appropriate gene promoters, and sequestered from other gene promoters where the interaction is undesirable (Downen et al., 2014; Hnisz et al., 2016a). The earliest evidence for chromosomal looping was presented almost 40 years ago (Dunn et al., 1984),

and reinforced since by many independent lines of evidence (Amano et al., 2009; Chambeyron and Bickmore, 2004; Tolhuis et al., 2002). Cohesin and CTCF are strongly implicated in the formation of loops larger than 100 kb (Figure 1.2) (Phillips-Cremins et al., 2013; Wendt et al., 2008). The latest models indicate that cohesin is loaded onto DNA in a CTCF-independent manner (Wendt et al., 2008), initially forming a small DNA loop. Cohesin then travels along the DNA strand in opposite directions, extruding the intervening DNA into a dynamically expanding chromatin loop, until its further movement is prevented by a pair of CTCF insulators (Fudenberg et al., 2016, 2018; Sanborn et al., 2015). Importantly, the two CTCF molecules have to be bound in a convergent orientation to be able to stop the progress of the cohesin ring and anchor the extruded loop (Rao et al., 2014; Sanborn et al., 2015). Loop formation by extrusion is a highly dynamic process (Fudenberg et al., 2016), but contact frequency patterns do emerge from an overlay of data from a large number of cells (Fudenberg et al., 2018).

While the formation of these larger, insulating loop domains seems to be largely dependent on CTCF and cohesin, the genome-wide impact on transcriptional regulation upon the depletion of either of these architectural proteins is surprisingly quite minimal (Hyle et al., 2019; Nora et al., 2017; Rao et al., 2017; Schwarzer et al., 2017). Loss of cohesin led to a somewhat enhanced compartmentalization pattern between active and inactive chromatin (Rao et al., 2017; Schwarzer et al., 2017), while CTCF depletion resulted in a minor reduction in compartmentalization (Nora et al., 2017). The preservation of compartmentalization following CTCF or cohesin loss indicates that there are other organizational forces at play here, governed by principles independent from the CTCF/cohesin-mediated loop formation. Interestingly, one of the recent cohesin-depletion studies pointed out that histone signatures remained largely unaffected upon cohesin loss (Rao et al., 2017). In keeping with this finding, a new study suggested a model where compartmentalization of the genome is driven by the attraction of similar histone modifications, although the forces behind this process were not discussed (Nichols and Corces, 2021).

Enhancer-promoter contacts may be also established, reinforced or maintained by smaller loops than the ones that arise from the collaboration of cohesin and CTCF (Figure 1.2A). The CTCF binding sites are usually distal from the regulatory elements within the domain that they insulate, and the loops that result from CTCF/cohesin interactions generally span 100 kb – 1 Mb of DNA (Phillips-Cremins et al., 2013). Smaller loops (<100 kb) can be formed directly between enhancers and promoters by the cooperation of cohesin with Mediator, the scaffold protein for transcriptional machinery (Kagey et al., 2010; Phillips-Cremins et al., 2013). Furthermore, enhancer-promoter looping was also shown to be facilitated by the interaction of the ubiquitously expressed Yin Yang 1 (YY1) transcription factor (Weintraub et al., 2017). YY1 binds to its recognition motif on promoters and enhancers, its binding

stabilized by an interaction with RNA (Sigova et al., 2015); YY1 then mediates the loop formation by dimerization (Weintraub et al., 2017). A similar instance of CTCF/cohesin-independent enhancer-promoter looping was described in erythrocytes, where the β -globin promoter-locus control region (LCR) enhancer loop was mediated by the dimerization of the LDB1 transcription factor and aided by the KLF1 factor (Krivega and Dean, 2017). This evidence collectively suggests that the interaction between transcription factors and co-factors may be extremely important for bringing enhancers and promoters into a closer contact both within and outside the context of CTCF/cohesin-insulated loop domains.

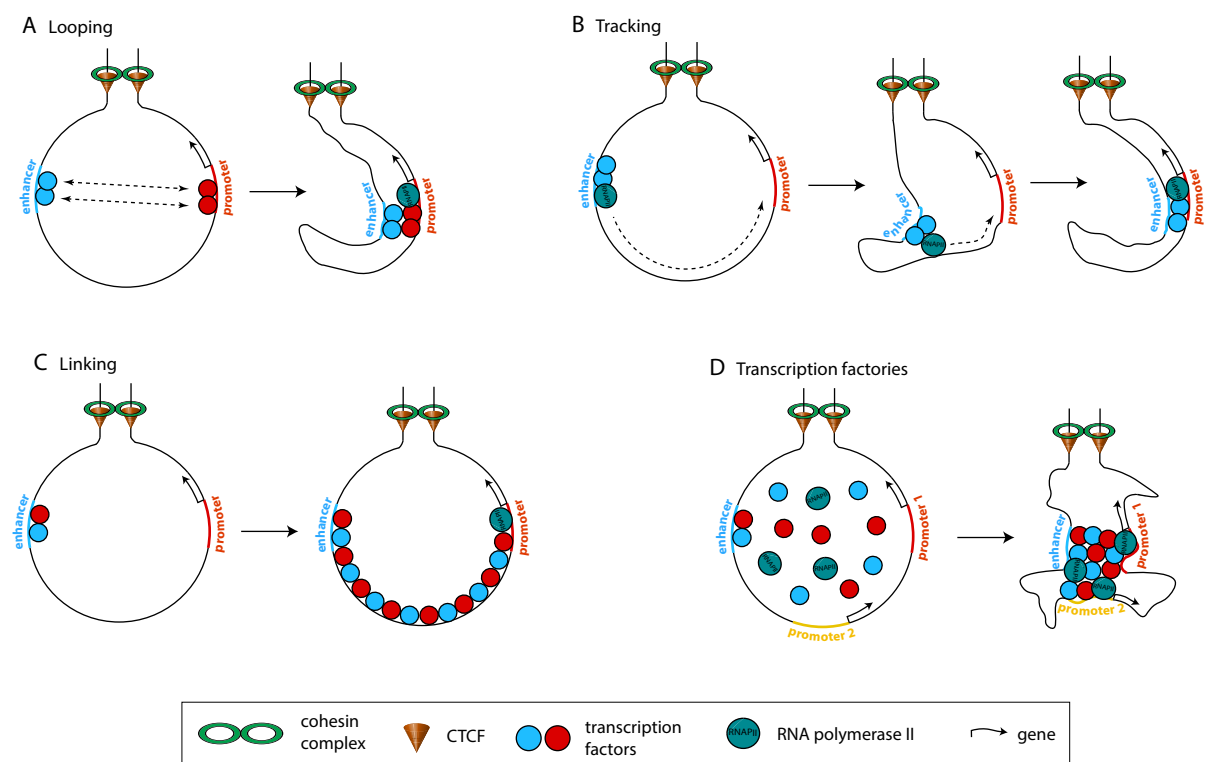


Figure 1.2: Proposed mechanisms of enhancer-promoter contact establishment.

(A) Looping can be mediated by the interactions between TFs bound at enhancers and TFs bound in promoter regions. **(B)** Tracking of the RNAPII along a DNA strand can help deliver TFs from enhancer to promoter regions. **(C)** Short-distance linking of TFs from enhancers to promoters can help loading TFs on promoters. **(D)** Transcription of genes may proceed from dynamic transcription factories – assemblies of RNAPII and other transcriptional machinery. **(A-D)** Cohesin complex in collaboration with CCCTC-binding factor (CTCF) forms bigger loops and sequesters contact domains or TADs.

Another mechanism that may be aiding in the establishment of enhancer-promoter contacts is tracking (Figure 1.2B). The idea was first proposed a quarter of a century ago (Blackwood and

Kadonaga, 1998; Kong et al., 1997), and some limited evidence was since presented in its support (Hatzis and Talianidis, 2002; Wang et al., 2005). According to this model, enhancers and their associated transcriptional complexes are pulled into the proximity of their target promoters by virtue of RNA polymerase II (RNAPII) transcription from the enhancer towards the target promoter (Hatzis and Talianidis, 2002; Wang et al., 2005). Histone acetylation of the intervening DNA portion between the enhancer and the promoter, and the generation of sense enhancer RNAs were described as the by-products and evidence of the progress of the RNAPII along the DNA strand (Kong et al., 1997; Wang et al., 2005). Another proposed mechanism for enhancer-promoter interaction is linking (Figure 1.2C), where transcription factors can create a bridge between the enhancer and the target promoter by oligomerization (Bulger and Groudine, 1999; Morcillo et al., 1997), but direct evidence for this is lacking. Conceivably, either of these mechanisms would likely only be able to mediate short-range interactions. Therefore, tracking and linking could play supporting roles in the establishment of contacts between proximally situated regulatory elements (Furlong and Levine, 2018).

An additional mechanism posited for the facilitation of enhancer-promoter contact are so-called transcription factories (Figure 1.2D). This model was built on the observation of discrete RNAPII foci, where genes were proposed to travel to be transcribed (reviewed in (Sutherland and Bickmore, 2009)). These transcription factories were at first believed to be spatiotemporally stable, but more recent evidence suggests that they are, in fact, highly dynamic structures, persisting on average only 5 – 8 seconds (Cho et al., 2016; Cisse et al., 2013). Phase separation has been implicated in the assembly of these structures (reviewed in (Hnisz et al., 2017)).

A cooperative model of enhancer action, where multiple requirements have to be fulfilled for transcriptional activation, is consistent with a number of observations (Furlong and Levine, 2018). A vast majority of enhancer-promoter loops seem to be already set up before the activation of the transcription from the target promoter; the final activation step may for example take shape of the recruitment of a specific transcription factor (Ghavi-Helm et al., 2014; Jin et al., 2013; Rubin et al., 2017). Other enhancer-promoter contacts, however, are not pre-established, and are formed *de novo*, for example in cell differentiation (Rubin et al., 2017).

Overall, many mechanisms of enhancer-promoter interaction have been described, including tracking, transcription factories and various forms of looping. These are not mutually exclusive; on the contrary, different mechanisms may likely contribute to the same enhancer-promoter interaction.

1.2.3.2 Modes of enhancer regulation

The factors governing the choice of the preferred interaction between enhancers and promoters seem to be as numerous as the mechanisms that can contribute to the interactions (Figure 1.3). One highly important factor seems to be the mutual proximity of the promoter and its candidate enhancer (Figure 1.3A). This was well-illustrated in a genome-wide study in a human leukaemia cell line, where two thirds of non-intronic enhancers were found to regulate their nearest gene (Gasperini et al., 2019). However, this left 33% of enhancer-promoter pairs that skipped the nearest gene. In some cases, this could be because enhancers are separated from their closest gene by a TAD boundary, even though cross-TAD enhancer-promoter contacts are also not uncommon (29% of all assessed enhancer-promoter pairs in (Gasperini et al., 2019)).

In other cases, inherent preferences of individual enhancers toward a specific type of promoter may be at play (Figure 1.3B; (Juven-Gershon et al., 2008; Sharpe et al., 1998; Zabidi et al., 2015)). For example, in *Drosophila*, a widespread enhancer-promoter specificity was described, dividing enhancers into a group that preferentially associated with promoters of developmental genes, sporting motifs such as TATA, DPE, Inr and MTE, and a group that preferred to interact with promoters of housekeeping genes, bearing the DRE or TCT motif (Zabidi et al., 2015).

On the other hand, many enhancers are shared between or competed for by multiple different gene promoters, especially within the context of the given regulatory domain (Figure 1.3C; (Cho et al., 2018; Fulco et al., 2016, 2019; Klann et al., 2017; Ohtsuki et al., 1998; Symmons et al., 2014)). There seems to be competition, for example, for the β -globin LCR enhancers (Klann et al., 2017) or the enhancers regulating GATA1 and HDAC6 expression (Fulco et al., 2016). Interestingly, Cho and colleagues describe a case of competition for intragenic enhancers, where the transcription of a long non-coding RNA, *PVT1*, prevents the MYC oncogene promoter from engaging with the enhancers located within the body of the *PVT1* gene (Cho et al., 2018).

Some genes may also be regulated by more than one enhancer (Figure 1.3D; (Bender et al., 2012; Kieffer-Kwon et al., 2013; Rosenbauer et al., 2004)). In these cases, enhancers often work in an additive fashion, each of them mediating a partial increase in the transcription rate of the target gene (Bender et al., 2012; Kieffer-Kwon et al., 2013; Rosenbauer et al., 2004). Other enhancers function in a redundant manner, buffering for potential mutations in or losses of individual enhancers to prevent phenotypic consequences (Hong et al., 2008; Osterwalder et al., 2018). Finally, some enhancers, such as the activation induced deaminase (AID) enhancers E1 and E2, seem to act in synergy with each other – the ablation of either one of the pair leads to a near complete abrogation of AID expression (Kieffer-Kwon et al., 2013).

There is usually a number of potential enhancer-promoter interactions within a regulatory unit, and various mechanisms contribute to the selection of the preferred contacts, including enhancer-promoter specificity, competition for enhancers and enhancer sharing. Multiple enhancers may regulate a single gene in an additive, synergistic or redundant manner.

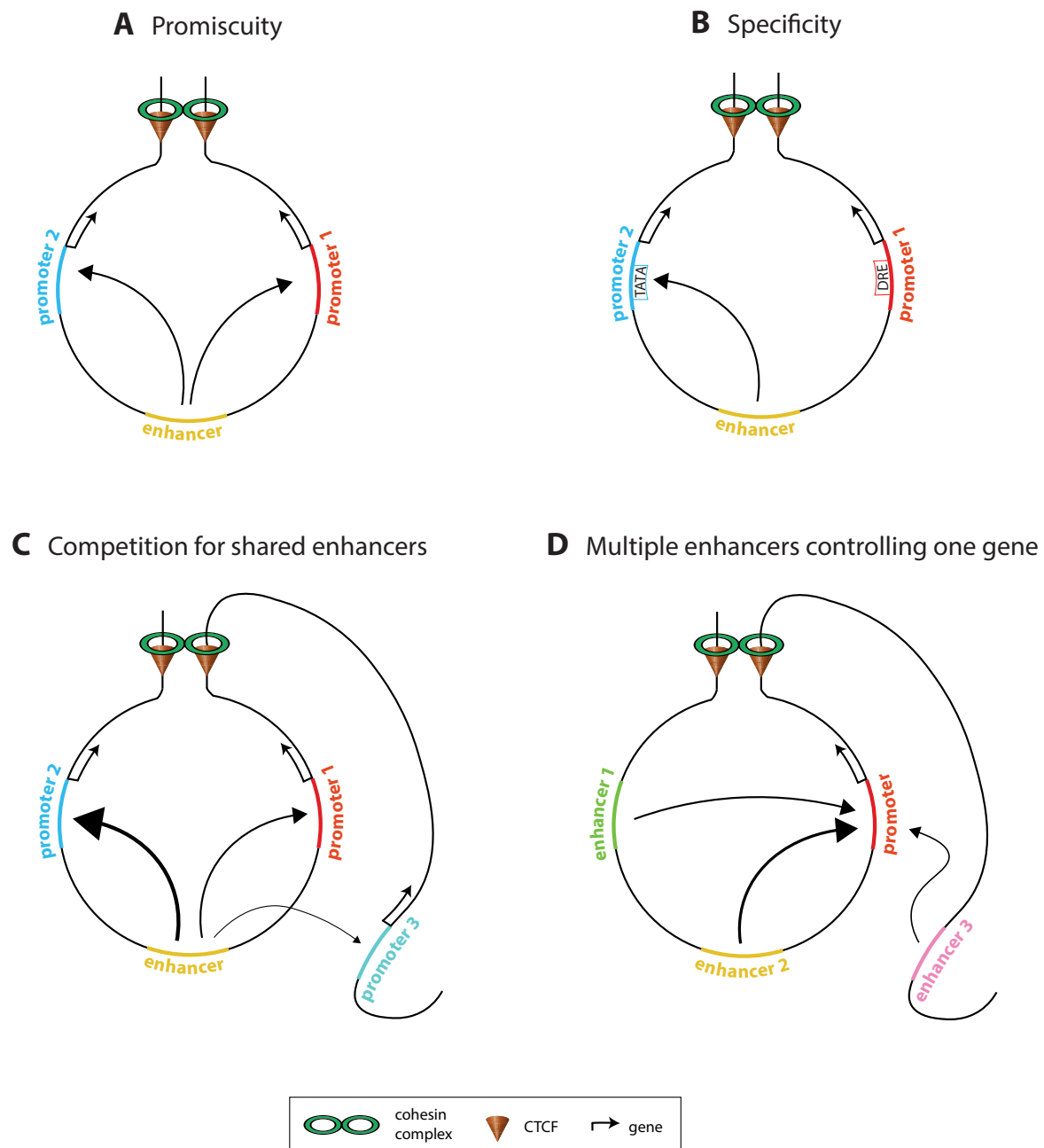


Figure 1.3: Modes of enhancer regulation.

(A) Some enhancers can work with any type of promoter, while (B) other enhancers prefer to communicate with specific types of promoters. (C) A single enhancer can contribute to the regulation of multiple genes, while sometimes, (D) multiple enhancers can control transcription of a single gene.

1.2.4 Super-enhancers

Regions of clustered enhancer elements are known as super-enhancers (Hnisz et al., 2013; Whyte et al., 2013). Super-enhancers usually span several kilobases (kb) of DNA in length (Parker et al., 2013), and compared to classical enhancers display an unusual enrichment in the marks typically associated with enhancers, including high levels of cell-type-specific master transcription factors, RNAPII, Mediator, cohesin and p300/CBP binding, as well as high levels of H3K27 acetylation (Hnisz et al., 2013; Whyte et al., 2013). Concomitantly, super-enhancers generally drive higher levels of transcription than typical enhancers (Whyte et al., 2013). Importantly, it was shown across many dozens of different cell types that a vast majority of super-enhancers is responsible for the regulation of cell-type-determining genes (Hnisz et al., 2013; Parker et al., 2013; Whyte et al., 2013). Super-enhancers typically bind high levels of cell-type-specific master transcription factors, and are very sensitive to the reduction in their amounts; the Young lab speculate that this vulnerability might have evolved to facilitate transitions between developmental states or cell-specific gene expression programmes (Hnisz et al., 2013; Whyte et al., 2013). Perhaps unsurprisingly, super-enhancers are enriched in disease-associated single nucleotide polymorphisms (SNPs) compared to typical enhancers, and dysregulated or de novo super-enhancers are implicated in a remarkably wide spectrum of cancers (Hnisz et al., 2013; Lovén et al., 2013; Parker et al., 2013).

According to expectation, super-enhancer activity is typically confined into CTCF/cohesin-defined loops (Downen et al., 2014). Finally, a recent CRISPR interference (CRISPRi)-based multiplexed analysis study offers new insights into the workings of super-enhancers (Xie et al., 2017). Intriguingly, most of the studied super-enhancers had only one or two main contributing constituents, whose effect on the expression of the associated genes was anywhere between 18 and 88.9%. The rest of the constituents didn't elicit any measurable gene downregulation when perturbed individually, although in some cases, a loss of a combination of these weak constituents led to a measurable effect on gene expression. This data suggests a hierarchical structure between super-enhancer constituents, with weaker constituents supporting the one or two main ones in a redundant fashion (Xie et al., 2017).

Importantly, the model system we work with in this thesis is a cancerous de novo super-enhancer regulating the expression of TAL1 in Jurkat cells (Chapter 1.5, (Mansour et al., 2014)).

1.3 Enhancer RNAs

It had long been believed that transcription by RNA polymerase II (RNAPII) only originated from promoter regions. Since the late 1980s, accumulating evidence pointed towards the fact that this notion is untrue. In 2005, Cheng and colleagues showed in their systematic study of 10 human chromosomes that an extensive portion of non-coding regions in the genome is being transcribed into polyadenylated and non-polyadenylated RNA transcripts, finally confirming that non-coding RNAs (ncRNAs) are an abundant species rather than an exceptional oddity (Cheng et al., 2005). In a later, more comprehensive study of the whole human genome in 15 cell lines, the ENCODE annotation group concluded that RNA transcripts may arise from up to three quarters of the human genome (Djebali et al., 2012).

The advent of transcriptome-wide approaches marked the beginning of an era in which a myriad of ncRNA species with distinct functions has been discovered, dispelling, time and time again, well-established misconceptions about the limits of what RNAs can do (reviewed by (Cech and Steitz, 2014)). In 2010, the sea of emerging ncRNA classes grew to encompass yet another novel species, which was unearthed in great numbers in the mouse genome. When Kim and colleagues realized that many enhancers bound RNAPII, they inquired whether these regions were also being transcribed. Having employed high-throughput RNA sequencing in mouse cortical neurons, the group found that a subset of enhancers indeed gave rise to RNAs (Kim et al., 2010). De Santa and collaborators made the same discovery in activated mouse macrophages (De Santa et al., 2010).

Both groups noticed that the transcription from the enhancers dynamically changed in response to a stimulus, and Kim and co-workers also noted that the resulting enhancer RNAs (eRNAs) seemed to arise from those enhancers that actively promoted transcription of their target genes (De Santa et al., 2010; Kim et al., 2010). This correlation was later shown to be so strong and pervasive that several studies proposed eRNA transcription could be used as a robust annotation method for active enhancers (Andersson et al., 2014a; Tyssowski et al., 2018). In keeping with their key role in cell-fate determination, enhancer transcription generally represents the first wave of transcriptional response in mammalian cells, preceding mRNA transcription (Arner et al., 2015).

1.3.1 The biogenesis of eRNAs

Transcription can only occur at enhancers that are in an active state (Chapter 1.2.2). Nucleosomes have been evicted from such enhancers by binding of pioneer factors (Adams and Workman, 1995;

Boyes and Felsenfeld, 1996; Ghisletti et al., 2010; Gualdi et al., 1996; Heinz et al., 2010; Miller and Widom, 2003; Sérandour et al., 2011; Zaret, 2020), and transcriptional machinery has been recruited to them. This machinery includes RNAPII (De Santa et al., 2010; Kim et al., 2010; Koch et al., 2011), general transcription factors, such as TATA-binding protein (TBP) and TBP-associated factor 1 (TAF1) (Heintzman et al., 2007; Koch et al., 2011), and transcription co-factors, including histone acetyltransferases (HAT; CBP/p300) responsible for the deposition of the H3K27ac activating enhancer mark (Tie et al., 2009; Visel et al., 2009), histone methyltransferases (MLL1, MLL2/4, MLL3), which deposit the H3K4me1/2 activating signature (Kaikkonen et al., 2013), and H3K27me2/3 demethylases (Kdm6a/b), which are responsible for removing the repressive methylation imparted to chromatin by the Polycomb repressive complex 2 (PRC2) (Agger et al., 2007; Boyer et al., 2006). Kdm6a additionally associates with MLL2 (Issaeva et al., 2007) and helps to recruit MLL4 (Li et al., 2017), and Kdm6b associates with CBP (Kyzar et al., 2019), illustrating a part of the orchestrated endeavour that leads to enhancer activation.

Transcription proceeds along the enhancer with the help of Bromodomain-Containing Protein 4 (BRD4), which facilitates the movement of RNAPII along the chromatin by interacting with acetylated histones (Kanno et al., 2014). RNAPII at enhancers is typically phosphorylated on tyrosine-1 of the C-terminal domain (CTD) (Descostes et al., 2014), with predominant phosphorylation on serine-5 (initiating RNAPII) in comparison to serine-2 (elongating RNAPII) (Koch et al., 2011). This representation is consistent with the short length of typical eRNA transcripts, tied to the generally early termination of transcription at enhancers, due in turn to the common occurrence of polyadenylation sites at enhancers (Andersson et al., 2014a; Ntini et al., 2013). This is in contrast with long RNA transcripts such as messenger RNA (mRNA), whose bodies are marked by an underrepresentation of polyadenylation sites, allowing, in cooperation with an enrichment in U1 small nuclear ribonucleoprotein (snRNP) binding sites, productive elongation along the full length of the transcript (Almada et al., 2013; Andersson et al., 2014a; Kaida et al., 2010; Ntini et al., 2013). The proximity of the RNAPII initiation site and the polyadenylation signal, such as encountered at enhancers, also contributes to the instability of the resulting transcripts, possibly owing to a facilitated interaction between the nuclear cap-binding complex and the exosome, which is responsible for the degradation of eRNAs (Andersen et al., 2013; Andersson et al., 2014a).

Transcription termination on enhancers is also connected with the Integrator complex and the WD repeat-containing protein 82 (WDR82) (Austena et al., 2015; Lai et al., 2015). The perturbation of the Integrator complex led to an increased association of eRNAs with RNAPII, a concomitant drop in mature eRNA levels and a significant rise in polyadenylation of eRNAs, implicating the Integrator in the 3'-end cleavage of newly synthesized eRNA transcripts to release them from RNAPII (Lai et al.,

2015). The depletion of WDR82 resulted in the synthesis of elevated levels of unnaturally long eRNA transcripts, suggesting a role of WDR82 in the recognition of polyadenylation signals by the elongating RNAPII (Austena et al., 2015).

Overall, enhancer transcription seems to be a rather coordinated process, with a number of similarities to the RNAPII transcription from promoters (reviewed in (Li et al., 2016)). RNAPII is enabled to initiate transcription in the environment of active enhancers, elongating with the help of BRD4 before the usually short eRNA transcripts are terminated with the aid of Integrator and WDR82.

1.3.2 Structural features and other characteristics of eRNAs

Following the establishment of eRNAs as a novel RNA species, a host of studies subjected them to detailed scrutiny. They were found to be predominantly enriched in the nuclear fraction (Andersson et al., 2014a; Djebali et al., 2012), short (median 346 nt) and generally unstable (Andersson et al., 2014a) and rather scarce (with approximately 5 – 15 copies per cell for the majority of examined eRNAs and 70 – 95 molecules per cell for the most abundant ones) (Li et al., 2013). For the most part, eRNAs were shown to be capped, but rarely spliced (only 5% of all eRNAs), owing to a lack of U1 splice sites within the body of the transcripts (Andersson et al., 2014a). The majority of eRNAs were repeatedly found to be non-polyadenylated and bidirectionally transcribed (Andersson et al., 2014a; Djebali et al., 2012; Kim et al., 2010), although some exceptions and contradictions have emerged (Andersson et al., 2014a; Djebali et al., 2012; Koch et al., 2011; Kouno et al., 2019).

Koch and co-workers noticed that longer eRNA transcripts often arose from more active enhancers, were more likely to be transcribed unidirectionally and to be polyadenylated (Koch et al., 2011). Several studies reported bidirectional transcription from specific enhancers with a dominant transcription from one strand (Hsieh et al., 2014; Pulakanti et al., 2013; Schaukowitch et al., 2014). Such a dominance of one transcript over its counterpart may have functional implications (Hsieh et al., 2014). Interestingly, for seemingly unidirectionally transcribed elements and bidirectional transcription with a dominant transcript, it may in fact be that the transcription does proceed bidirectionally in a more or less equal measure, but one of the transcripts is exosome-sensitive, while the other is not (Andersson et al., 2014b). In an intriguing twist, the work of Kouno and colleagues showed that while transcription from many enhancers may seem to proceed bidirectionally when assessed globally, an overwhelming majority of the same loci are, in fact, transcribed unidirectionally on a single-cell level (Kouno et al., 2019).

Apart from the variation in length, processing and abundance, the world of eRNA also seems to boast a kaleidoscopic range of secondary structures. According to computational analyses, many of the structures within eRNAs mirrored those known from other RNA species, including transfer RNAs (tRNAs), long non-coding RNAs (lncRNAs), micro RNAs (miRNAs) and small nuclear RNAs (snRNAs), while some were previously unobserved (Cheng et al., 2015; Ren et al., 2017). Given the close relationship between RNA structure and function, this adumbrates the potentially multifarious roles that eRNAs might play in transcriptional regulation.

1.3.3 Function of enhancer transcription in gene regulation

The seemingly well-regulated process of eRNA transcription (Chapter 1.3.1, further reviewed in (Arnold et al., 2019)), the similarities between the structures of eRNAs and other RNA transcripts with known functions (Cheng et al., 2015; Ren et al., 2017), and, perhaps most importantly, the fact that there is a positive correlation between the level of transcription of genes and the transcriptional activity of the enhancers in their vicinity (Kim et al., 2010), begets the question of what role, if any, eRNAs play in the mechanism of transcriptional enhancement.

Based on the body of published research, Li and colleagues proposed a classification of eRNAs according to their function (Li et al., 2016). The categorization derives from our current understanding that enhancer function may be linked to transcription-unrelated mechanisms, even though the vast majority of active enhancers are transcribed (class I eRNAs), or enhancer function may lie within the act of transcription, rather than the transcript (class II eRNAs), or depend on the transcripts themselves (class III eRNAs). Some of these functions may be non-exclusive.

1.3.3.1 Class I eRNAs

Out of the imperfection of RNAPII fidelity, the probability of transcription initiation is only about 200 times lower at a random locus than at an average correct site, or about 10^4 times lower than at a maximally active promoter site, as determined in yeast (Struhl, 2007). And indeed, compared to transcripts arising from protein-coding genes, non-coding RNAs, including eRNAs, are generally of a much lower abundance (Djebali et al., 2012). Moreover, the high accessibility of chromatin at enhancers can cause an increased incidence of random transcription initiation. Furthermore,

enhancer regions display a low level of conservation across species, as, for example, by comparison of liver enhancers in 20 mammalian species (Villar et al., 2015). Taken together, these findings may point towards a lack of biological function of eRNAs. In such a case, the purpose of an enhancer would be limited to providing promoter regions with transcriptional machinery.

1.3.3.2 Class II eRNAs

In an early endeavour, Cho and co-workers isolated RNAPII complexes containing several known chromatin remodellers, including the histone acetyltransferases p300/CBP and p300/CBP-associated factor (PCAF) (Cho et al., 1998). A later, more systematic experiment in yeast showed that RNAPII can, in fact, bind over 100 different proteins and ribosomal proteins, by means of its differentially phosphorylated CTD (Phatnani et al., 2004). A recent work convincingly linked H3K4 mono- and dimethylation at enhancers to eRNA transcription (Kaikkonen et al., 2013). Importantly, methylation levels did not decrease upon depletion of eRNAs by locked nucleic acid (LNA) antisense oligonucleotides (ASO), which strongly indicated that the methylation was not dependent on eRNAs, only on the act of transcription itself (Kaikkonen et al., 2013). Taken together, this evidence suggests that as RNAPII moves along an enhancer region while elongating an eRNA transcript, the chromatin-remodelling machinery riding on its CTD tail gets to work, dramatically changing the landscape of local chromatin (Figure 1.4A). In other instances, the act of transcription at enhancers may interfere with transcription of other genes (Figure 1.4B), as exemplified by Onodera's study of two antisense eRNAs whose increased level of transcription was correlated with preferential expression of shorter isoforms of nearby protein-coding genes (Onodera et al., 2012), or Cinghu's work on intragenic enhancer transcription which attenuated the transcription of the host gene (Cinghu et al., 2017).

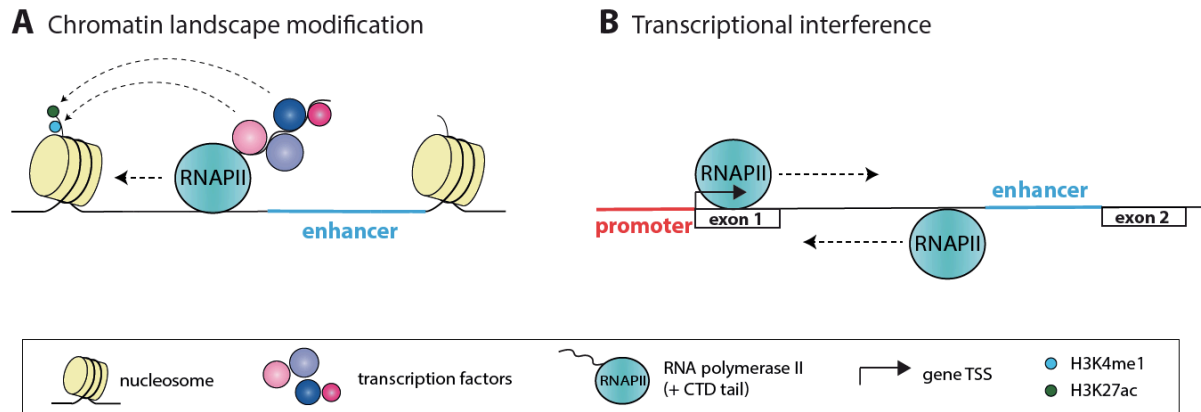


Figure 1.4: The effects of the act of eRNA transcription.

(A) The CTD tail of RNAPII binds a number of chromatin modifiers, which can deposit positive marks on proximal histones as RNAPII tracks along the DNA strand. **(B)** Transcription from enhancers may interfere with a productive elongation of a gene.

1.3.3.3 Class III eRNAs

Over the recent years, a number of functions for eRNA has been proposed, including roles in chromatin remodelling, promoter-enhancer looping and RNAPII pause release.

1.3.3.3.1 eRNA roles in chromatin remodelling

Mounting evidence suggests that eRNAs contribute to the alterations of chromatin landscape. For example, in an early study, the knockdown of eRNAs associated with the myogenic regulatory factors MyoG and MyoD led to a reduction in chromatin accessibility at the gene loci (Mousavi et al., 2013). Another study showed that cells stably transfected with short hairpin RNA (shRNA) targeting eRNA from the enhancer of α -subunit chorionic gonadotropin alpha (Cga) displayed diminished binding of a known chromatin remodeller (chromodomain-helicase-DNA-binding protein 1, CHD1), but also a dramatic drop in active H3K4me3 and H3K27ac marks and an increase in repressive H3K27me3 mark compared to wild-type cells (Pnueli et al., 2015). The majority of these effects were observed both at the relevant promoter and enhancer region (Pnueli et al., 2015). Furthermore, an shRNA-mediated knockdown of Epstein-Barr virus super-enhancer RNAs led to a decrease in H3K27ac marks on the super-enhancer (Liang et al., 2016), and an antisense oligonucleotide (ASO)-mediated knockdown of

two eRNAs associated with the prominent inflammatory monocyte gene *SERPINB2* reduced H3K27ac and H3K4me3 at the *SERPINB2* promoter (Shi et al., 2017).

A couple of later studies provided direct links between eRNAs and their influence on H3K27ac levels (Bose et al., 2017; Jiao et al., 2018). Jiao and co-workers demonstrated that binding of the heparase (*HPSE*) eRNA to heterogeneous nuclear ribonucleoprotein U (hnRNPU) promoted the interaction of hnRNPU with the histone acetyltransferase p300 and their recruitment to the heparase super-enhancer (Jiao et al., 2018). In a genome-wide study, Bose and colleagues demonstrated that the histone acetyltransferase CREB-binding protein (CBP) bound a wide range of eRNAs (amongst other RNA species), likely in a locus-specific manner. Interestingly, different species of eRNA elicited distinct and concentration-dependent response patterns of CBP activation (Bose et al., 2017). The study put forward a mechanistic explanation as well. RNA binding to CBP was shown to be predominantly realized through a highly basic, disordered and evolutionary conserved RNA-binding region (RBR) within the CBP-HAT domain (Bose et al., 2017). The CBP-RBR and adjacent residues form a loop which blocks the active site of CBP (Thompson et al., 2004). This loop can be displaced by acetylation (Thompson et al., 2004) and by RNA binding (Bose et al., 2017). The loop displacement frees the active site, allowing for substrate binding and rendering the CBP-HAT domain active (Figure 1.5; (Bose et al., 2017; Thompson et al., 2004)).

Overall, eRNAs can influence chromatin accessibility and histone modifications at their enhancer of origin and their target promoters by direct or indirect recruitment of chromatin remodellers, such as p300/CBP.

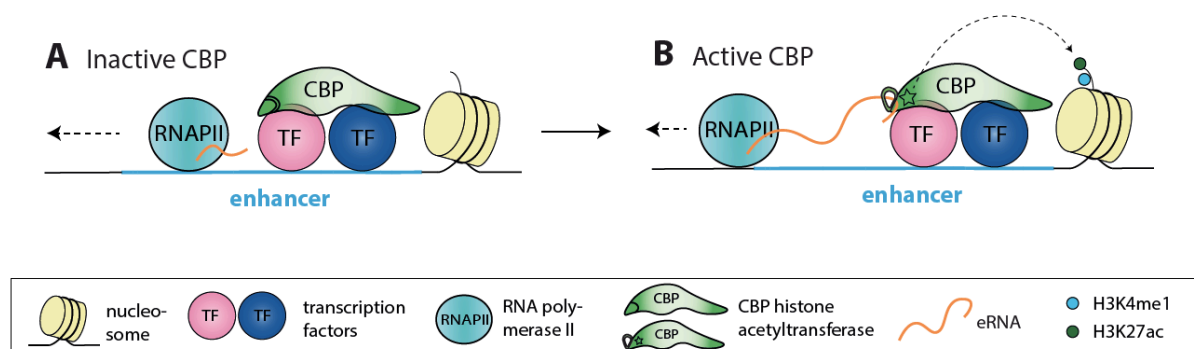


Figure 1.5: eRNA binding can activate CBP, resulting in increased histone acetylation.

(A) A loop blocks an active site in CBP, preventing its enzymatic activity. **(B)** eRNA binding can displace the loop, open access to the active site (denoted by a star), and render the enzyme active. (Bose et al., 2017)

1.3.3.3.2 eRNA roles in chromatin looping

Enhancer RNAs also seem to be implicated in chromatin looping. An early paper pointed out that actively transcribing enhancers were more likely to participate in looping than non-transcribing ones (Sanyal et al., 2012). Since then, several works have connected the loss of eRNAs with impaired DNA looping.

Firstly, using both small interfering RNAs (siRNAs) and LNA ASOs, Li and colleagues showed that depletion of several eRNAs from oestrogen receptor α (ER- α)-regulated enhancers led to a loss of promoter-enhancer looping in human breast cancer cells; additionally, the interrogated eRNAs were found to interact with SMC3 (structural maintenance of chromosomes protein 3) and RAD21 (radiation-sensitive mutant 21), two components of the cohesin complex (Li et al., 2013). Further studies documented interactions between the SMC1 and SMC3 (structural maintenance of chromosomes protein 1/3) subunits of cohesin and two enhancer lncRNAs, *Evf2*, contributing to the regulation of interneuron diversity (Cajigas et al., 2018) and ^{DRR}eRNA, playing a role in the regulation of muscle cell differentiation (Tsai et al., 2018).

Moreover, several studies reported interactions between eRNAs and the Mediator complex. Lai and collaborators investigated a couple of enhancer-like activating ncRNAs (ncRNA-a) in human HEK293 cells and concluded that they facilitated looping by interaction with the MED1 and MED12 subunits of Mediator complex (Lai et al., 2013). Association with Mediator complex (MED1 subunit) and a corresponding loss of looping and downregulation of target genes upon siRNA-mediated knockdown was also demonstrated for the sense eRNA originating from Kallikrein-related peptidase 3 (KLK3) enhancer, one of the most potent androgen-receptor (AR)-bound regulatory elements in human prostate cancer cells (Hsieh et al., 2014). Another pro-oncogenic eRNA, *ARIEL*, co-precipitated with MED12 in a significant measure, and *ARIEL* knockdown led to an almost entire loss of looping between the target *ARID5B* enhancer and promoter (Tan et al., 2019).

In addition, heparase (HPSE) eRNA was found to increase enhancer-promoter looping through the above-mentioned facilitation of the hnRNPU-p300 interaction (Jiao et al., 2018); super-enhancer lncRNA *CCAT1-5L* promoted local looping by binding other RNA species and hnRNPK, which was shown to dimerize and recruit RNAPII in turn (Cai et al., 2020).

Furthermore, eRNA interaction with the common transcription factor Yin-Yang 1 (YY1), mentioned earlier in connection with looping (Chapter 1.2.3.1), led to a modest, yet significant increase in YY1 binding to its DNA recognition motif in murine embryonic stem cells (Figure 1.6A; (Sigova et al., 2015)). Notably, similarly to the eRNA-CBP interactions, different RNAs bound to YY1 with different affinities (Bose et al., 2017; Sigova et al., 2015).

Finally, other studies noted the role of eRNAs in looping without putting forward a mechanistic explanation (Liang et al., 2016; Pnueli et al., 2015).

In sum, the direct contributions of eRNAs to chromatin looping described to date seem to concern their interactions with the Mediator complex, the cohesin complex, YY1 or the hnRNP family members. In a lot of these interactions, eRNAs are presumed to act as scaffolding, helping to tie the complexes of transcriptional machinery together (Figure 1.6B).

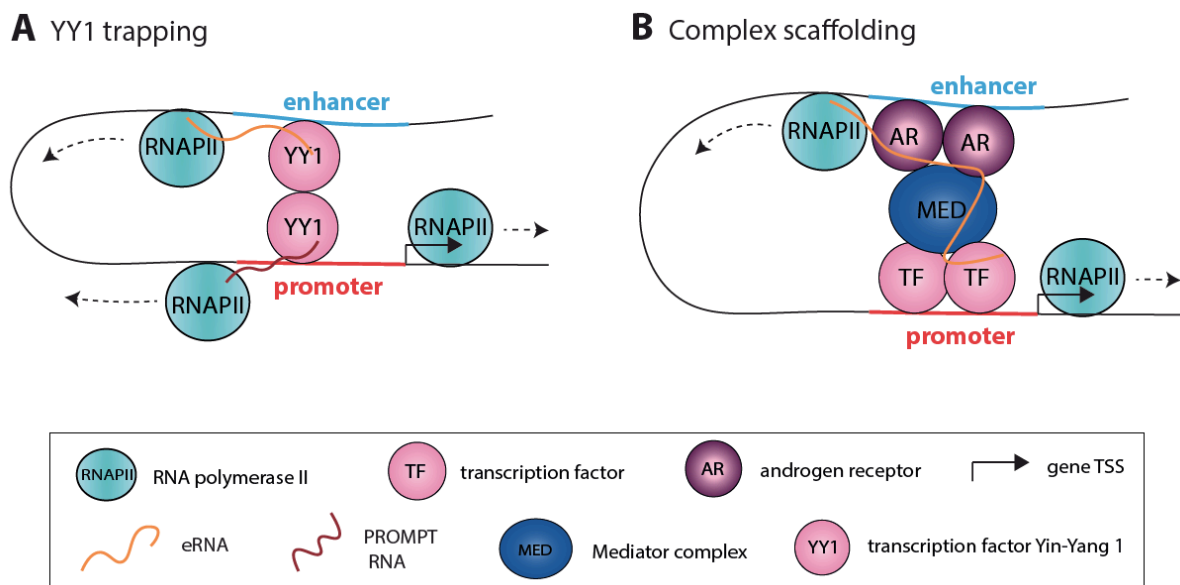


Figure 1.6: Roles of eRNAs in promoter-enhancer looping.

(A) Nascent transcripts, such as eRNAs or promoter upstream transcripts (PROMPTs), interact with Yin-Yang 1 (YY1), stabilizing its binding to enhancers and promoters. YY1 facilitates looping by dimerization (Sigova et al., 2015). **(B)** eRNA may act as scaffolding, for example in androgen-receptor (AR)-associated complexes with Mediator (Hsieh et al., 2014).

1.3.3.3.3 eRNA roles in regulation of transcription machinery

Many lines of evidence suggest that eRNAs can also regulate transcriptional machinery.

A host of studies found RNAPII levels decreased at enhancers and their target promoters after the knockdown of the corresponding eRNAs (Cai et al., 2020; Lai et al., 2013; Maruyama et al., 2014; Mousavi et al., 2013; Rahnamoun et al., 2018; Tan et al., 2019; Yang et al., 2016). Sometimes the recruitment of RNAPII was indirect through another factor, such as BRD4 (Rahnamoun et al., 2018) or

hnRNPK (Cai et al., 2020), although more often no mechanism was put forward, and a direct role of eRNAs in the recruitment of RNAPII remains an intriguing possibility.

In a fashion similar to the ‘trapping’ mechanism described for YY1 (Sigova et al., 2015), eRNAs were found to increase the stability of BRD4 binding to acetylated histones by direct interaction with the two BRD4 bromodomains (Rahnamoun et al., 2018). Importantly, BRD4 not only plays a major role in transcriptional elongation (Chapter 1.3.1), but also contributes to RNAPII recruitment, as well as RNAPII pause release (Jang et al., 2005; Kanno et al., 2014; Rahnamoun et al., 2018; Winter et al., 2017).

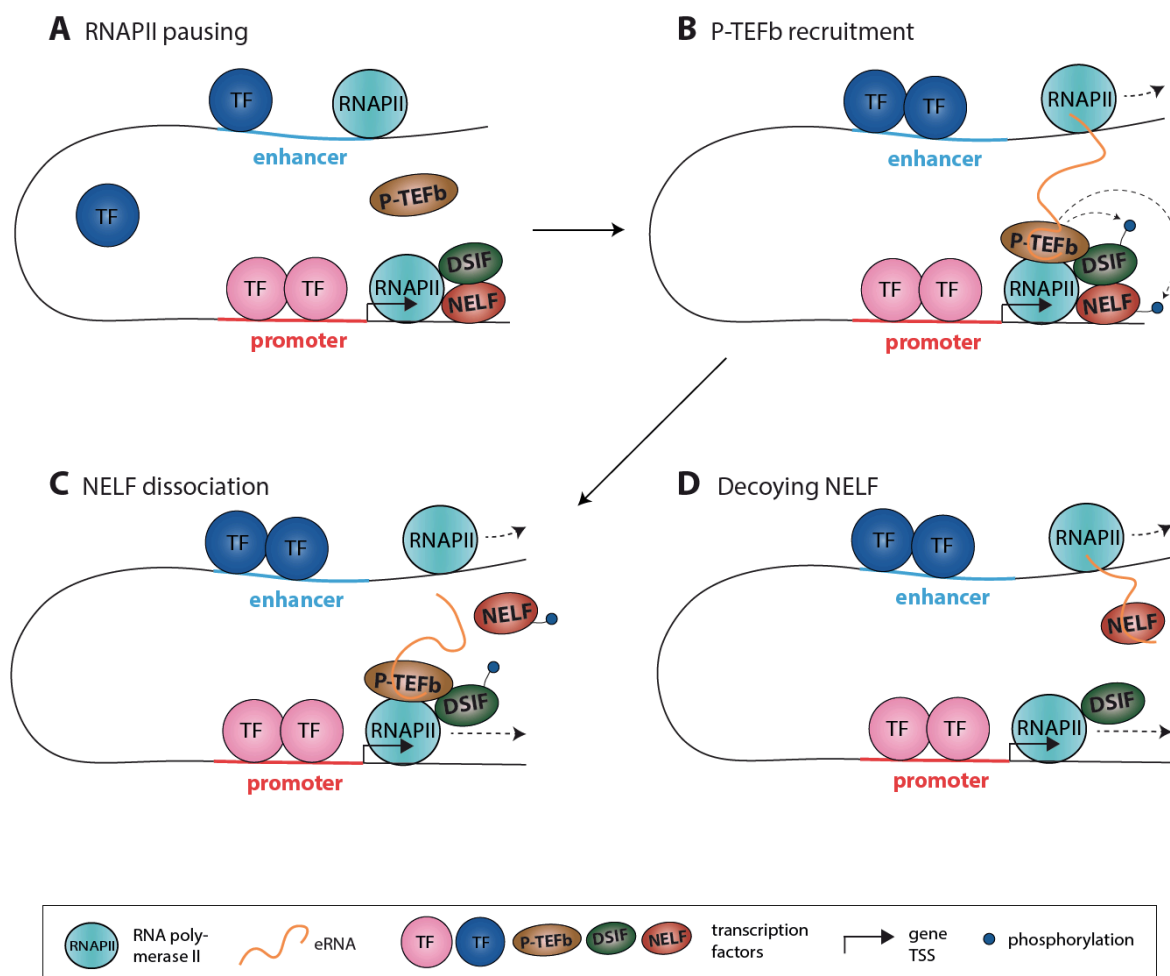


Figure 1.7: Role of eRNAs in RNAPII pausing.

(A) RNAPII pausing is initiated by the binding of DSIF, which recruits the negative elongation factor NELF. **(B)** eRNAs can contribute to the recruitment of the positive elongation factor P-TEFb, which phosphorylates both DSIF and NELF and thus precipitates the dissociation of NELF. (Shi et al., 2017) **(C)** The dissociation of NELF allows RNAPII to enter productive elongation phase. **(D)** eRNAs can also act as decoy for NELF, thereby lessening the frequency of RNAPII pausing (Schaukowitch et al., 2014).

Finally, eRNAs were further implicated in RNAPII pause release by direct interaction with the positive transcription elongation factor b (P-TEFb) and negative elongation factor (NELF) complexes (Figure 1.7; (Schaukowitch et al., 2014; Shi et al., 2017)). RNAPII pausing is a widespread transcriptional control mechanism, which takes place on both promoters and enhancers (Henriques et al., 2018). In pausing, RNAPII is stabilized by the dichloro-1- β -D-ribofuranosyl-benzimidazole (DRB) sensitivity-inducing factor (DSIF) and NELF (Figure 1.7A) until such a time that P-TEFb is recruited. P-TEFb phosphorylates NELF and DSIF, promoting the dissociation of NELF from the paused RNAPII and the subsequent release of RNAPII into transcriptional elongation (Figure 1.7C; reviewed in (Core and Adelman, 2019)). Enhancer RNAs were shown to come into this process at two different steps. Firstly, by binding to the CDK9 (cyclin-dependent kinase 9) subunit of P-TEFb, eRNAs support the recruitment of P-TEFb to the paused RNAPII complex (Figure 1.7B, C; (Shi et al., 2017)). Secondly, eRNAs seem to act as a decoy for NELF, competing with the nascent transcripts for binding to the RNA recognition motif of the NELF-E subunit, hence facilitating the dissociation of NELF from the paused RNAPII complex (Figure 1.7D; (Schaukowitch et al., 2014)).

Overall, eRNAs influence the recruitment of various transcriptional machinery, including RNAPII, BRD4, P-TEFb and NELF.

1.3.3.3.4 eRNAs with *trans* roles

Most eRNAs seem to exert their function within the local environment they have arisen from – acting in *cis*. Investigation suggests that some eRNAs, albeit a minority, may act on targets on different chromosomes (*trans* function) (Alvarez-Dominguez et al., 2017; Hsieh et al., 2014; Mousavi et al., 2013; Ørom et al., 2010; Tsai et al., 2018). Enhancer RNAs with proposed *trans* functions tend to be lnc-eRNAs, polyadenylated and/or otherwise processed. Post-transcriptional processing prolongs the half-life of such *trans*-acting eRNAs, which may in turn allow them to find their distant targets (Tsai et al., 2018).

1.4 CRISPR/Cas9-based screening of non-coding genome

1.4.1 A case for CRISPR

Many methods have been employed to pinpoint and study the function of enhancers, and while they all have contributed to the bulk of our current knowledge, all of them also come with their own set of limitations (reviewed in detail in (Gasperini et al., 2020)). While the examination of primary DNA structure for TF binding motifs and scanning of chromatin for enhancer-associated biochemical signatures can point us in the right direction as to the whereabouts of regulatory elements, these approaches tell us little of the interactions that the identified elements partake in, as enhancers often don't regulate the linearly closest gene and quite commonly also regulate multiple ones (Chapter 1.2.3.2). Methods capturing the three-dimensional conformations of chromatin, such as chromosome conformation capture (3C) and its derivatives for the analysis of selected local architectures (Dekker et al., 2002; Dostie et al., 2006; Zhao et al., 2006), or so-called Hi-C with the potential to map whole genomes (Lieberman-Aiden et al., 2009), do come a step closer to unveiling the enhancer-promoter interactomes, but the pitfall remains that spatial proximity of regulatory elements doesn't automatically warrant an active interaction (Chapter 1.2.3.2). On the other side, some highly practical information about active enhancer-promoter interactions can be gleaned from studying the correlations between genetic sequence and variation in gene expression on the level of human populations (expression quantitative trait locus, eQTL) (GTEx Consortium, 2017), but clearly there are various ethical limitations to these studies, including the restriction to naturally occurring variation. These particular limitations can be easily circumvented in massively parallel reporter assays (MPRA) (Patwardhan et al., 2009), whereby the functionality of many different candidate enhancer sequences, cloned as a library into a reporter vector, can be tested in a single experiment. However, MPRA takes candidate enhancers completely out of their biological context, rendering their results potentially irrelevant on a systemic level.

This is why the programmable perturbation of DNA sequence within a biological system observed through the lens of transcriptional output of the perturbed cells has gained a lot of traction, offering both a high customizability and biological relevance. First arrivals in this department were zinc finger nucleases (ZFNs) ((Kim et al., 1996); reviewed by (Jo et al., 2015)), much later followed by transcription activator-like effector nucleases (TALENs) (Christian et al., 2010; Pu et al., 2015). Both ZFNs and TALENs present a fusion between site-specific DNA-binding proteins and the catalytic domain from the FokI endonuclease, which exerts nucleolytic activity upon dimerization. To allow for the dimerization, the fusion proteins are targeted to both sides of the cut site. The target regions on the

DNA are recognized through a succession of specific amino acids in the DNA-binding domains of the fusion proteins, which can thus be engineered to recognize any desired target sequence (Christian et al., 2010; Segal et al., 2003). While both TALENs and ZFNs are highly programmable, the need to change the primary structure of these nucleases to reach each individual target sequence presents a challenge for a transition into a high-throughput format.

This challenge was addressed most effectively by the discovery of a new tool for programmable genome editing: CRISPR/Cas9 (clustered regularly interspaced short palindromic repeats (CRISPR) and the CRISPR-associated protein 9 (Cas9); (Cong et al., 2013; Jinek et al., 2013; Mali et al., 2013)). This system does not require any changes on the protein level, because the specificity of the Cas9 nuclease is defined by a short associated RNA molecule. This short, specific CRISPR RNA (crRNA) forms a structure with the scaffolding *trans*-activating CRISPR RNA (tracrRNA) that interacts with Cas9 and guides it to its target DNA sequence (Deltcheva et al., 2011; Jinek et al., 2012). Without a loss of activity, crRNA and tracrRNA can be genetically combined into one chimeric RNA molecule, known as a single guide RNA (sgRNA) (Jinek et al., 2012).

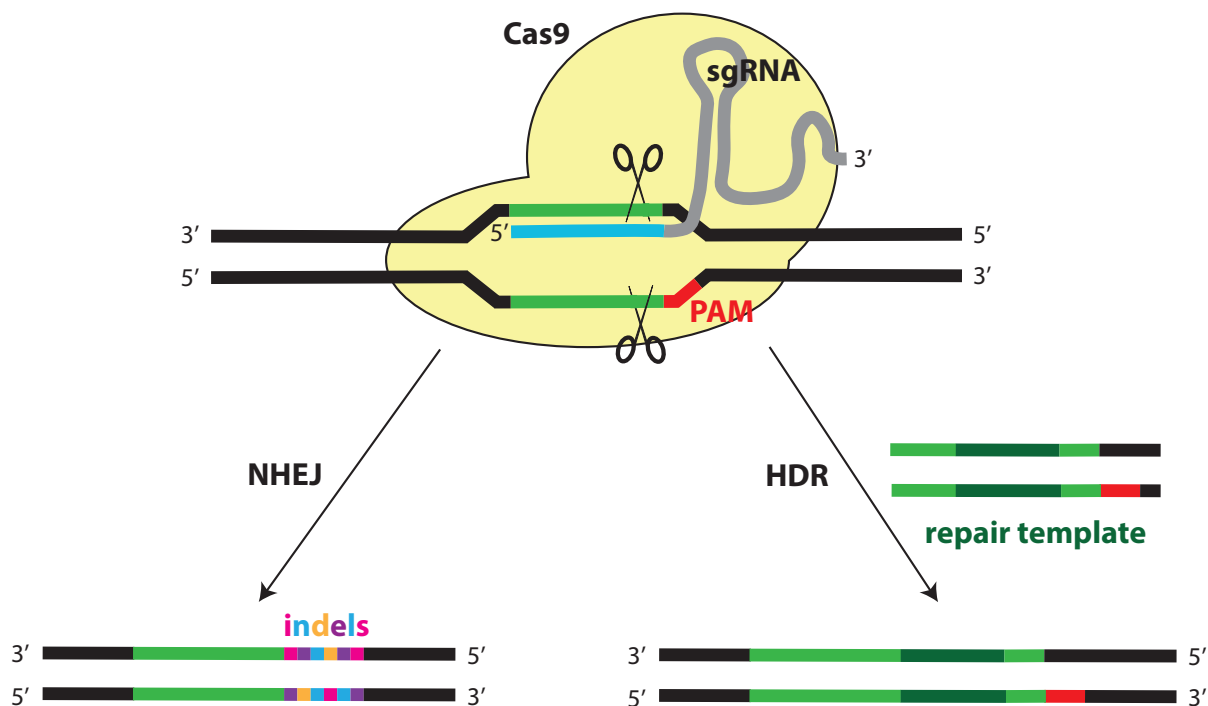


Figure 1.8: CRISPR/Cas9 editing and the two pathways to repair DSBs.

Cas9 (yellow) homes in on the target sequence (green) thanks to the specific portion of the sgRNA (blue) and the PAM recognition motif (red). Cas9 cuts on both strands, and the resulting double-strand break (DSB) is repaired either by non-homologous end joining (NHEJ), which often causes small local mutations (indels), or homology-directed repair (HDR), which uses a template to make a precise fix.

An sgRNA directs the Cas9 nuclease to cut at the selected DNA site, which can be any sequence that precedes the proto-spacer adjacent motif (PAM; for example, NGG in case of the most commonly used Cas9 from *S. pyogenes*) (Figure 1.8; (Jinek et al., 2012; Sapranasauskas et al., 2011)). The Cas9 enzyme inflicts a double-strand break (DSB) to the DNA three nucleotides upstream of the PAM (Jinek et al., 2012). In eukaryotic cells, DSBs can be repaired by either non-homologous end joining (NHEJ), or homology-directed repair (HDR) (Figure 1.8; (Scully et al., 2019)). While HDR fixes DSBs with high precision based on a DNA template, NHEJ – quick, efficient and generally much more common than HDR – is a less precise ligation mechanism that often introduces small insertions or deletions (indels) into the repaired DNA strand. This can result in the disruption of the normal function of the NHEJ-repaired regions, a feature exploited in functional CRISPR/Cas9-based experiments.

1.4.2 Non-coding genome editing with CRISPR/Cas9-based screens

The system of our choice, CRISPR/Cas9, has previously been called the King of Genome Editing Tools (Bannikov and Lavrov, 2017), although it might be argued that it is more like Robin Hood, a renowned marksman with a highly accurate aim. CRISPR and the associated proteins were first discovered in bacteria and described as their adaptive immune systems, protecting them from phages and plasmids (reviewed e.g. in (Wright et al., 2016)). A number of groups soon started tapping into the potential of CRISPR/Cas9 for programmable mammalian genome editing (Cong et al., 2013; Jinek et al., 2013; Mali et al., 2013), and it was only a short while before the first CRISPR/Cas9-based screen appeared (Shalem et al., 2014). The first screens based on CRISPR/Cas9-mediated genome editing targeted protein-coding regions (Parnas et al., 2015; Shalem et al., 2014; Zhou et al., 2014), but the application of this screening approach on non-coding regions closely followed suit (Chapter 1.4.2.1). A whole region deletion strategy based on CRISPR/Cas9 soon appeared as an alternative method for the dissection of regulatory landscapes (Chapter 1.4.2.2). Simultaneously, new methods utilizing a catalytically dead Cas9 (dCas9) fused to various effector domains were developed, perfect for the dissection of regulatory elements (Chapter 1.4.2.3).

1.4.2.1 CRISPR/Cas9 screens introducing small insertions and deletions

The focus of CRISPR/Cas9 screens on non-coding genome has been broad, although with potential downstream therapeutic intervention often in mind: BCL11A (B-cell lymphoma/leukaemia 11A) in the

centre of Canver's study (Canver et al., 2015) is a repressor of foetal haemoglobin and a therapeutic target for β -hemoglobinopathies; p53, which is in the spotlight of Korkmaz's work, is a well-known tumour-suppressor; ER- α , a focus of the second part of this study, has a mitogenic effect in breast cancer (Korkmaz et al., 2016). Another study focused on describing the regulatory landscapes of three genes (*NF1*, *NF2* and *CUL3*) causing susceptibility to BRAF protein-kinase inhibitor vemurafenib in melanoma (Sanjana et al., 2016), while yet another explores the regulatory genomic context of *PD-1*, a gene whose expression marks the state of exhaustion in CD8+ T-cells (Sen et al., 2016).

Most studies to date that endeavoured CRISPR/Cas9 screening of non-coding genome were limited to investigating the regulatory elements belonging to either just one gene (Canver et al., 2015, 2020; Diao et al., 2016; Sen et al., 2016) or a few (Rajagopal et al., 2016; Sanjana et al., 2016), allowing for detailed mapping of the regions. While lacking in this sort of detail, genome-wide studies of regulatory landscapes provide invaluable information about the more global roles of functional elements in physiological and pathological states and processes. For example, one study used CRISPR/Cas9 screening to uncover p53-bound enhancers that play a role in oncogene-induced senescence, an important tumour-suppressive mechanism (Korkmaz et al., 2016). In the same study, the authors also surveyed a subset of ER- α -bound enhancers, identifying enhancers that drive proliferation of breast cancer cells addicted to ER- α .

The methods employed to select the portions of non-coding genome for dissection varied between groups: many made use of the fact that functional regions may be predicted by means of active or open chromatin marks, chromosome conformation, transcription factor binding sites, or, to a limited degree, homology with the genome of a related organism, while others chose to work with whole regions proximal to the gene of interest. Canver et al. used DNase I hypersensitivity to pinpoint three candidate regulatory elements of human *BCL11A*. By homology, they identified orthologues of these candidate enhancers in mouse (Canver et al., 2015). Diao and co-workers focused on 174 putative regulatory elements contained in the same topological associated domain as their gene of interest, *POU5F1* (also known as *Oct4*). These candidate regions were selected because they bore marks of enhancers, contained CTCF binding sites and/or displayed DNase I hypersensitivity (Diao et al., 2016). The 2020 Canver's study of the regulatory landscapes of the same gene expanded the surveyed elements to *trans*-regulatory elements, choosing target regions genome-wide based on an advanced method for mapping accessible chromatin – assay for transposase-accessible chromatin coupled with high-throughput sequencing (ATAC-seq) (Canver et al., 2020). ATAC-seq was also used by Sen et al. to identify putative regulatory regions of *PD-1*, a major marker of exhausted CD8+ T-cells, in their study of the non-coding landscapes in exhausted versus functional effector CD8+ T-cells (Sen et al., 2016). Korkmaz and co-workers chose sgRNA target sites for their CRISPR/Cas9 screens based on either p53

or ER- α binding, as determined by ChIP-seq, coupled with the presence of enhancer hallmarks – various histone marks for the p53-bound enhancers, eRNA expression for ER- α -bound elements (Korkmaz et al., 2016). Rajagopal and colleagues screened the regulatory landscape of four different mESC-specific genes: for three of them (*Nanog*, *Rpp25* and *Zfp42*), the authors chose to dissect regions with enhancer-like signatures in the proximity of the target genes, as well as distal loci that displayed physical interaction in ChIA-PET. In case of the last gene, *Tdgf1*, they decided to simply target the 40 kb of genome most proximal to the gene in an unbiased mutagenesis (Rajagopal et al., 2016). Similarly, Sanjana et al. tiled their sgRNAs without bias across the regions 100 kb upstream and 100 kb downstream from each of their genes of interest (Sanjana et al., 2016). While this latter, unbiased screening strategy may uncover regulatory regions which do not bear the classic hallmarks of regulatory function (Rajagopal et al., 2016; Sanjana et al., 2016), the approach focused on pre-selected regions allows the inspection of a larger portion of the non-coding landscape.

While the techniques to perform the CRISPR/Cas9 screens themselves are multitudinous, the unifying feature is the search for differentially represented sgRNAs. Broadly, two main strategies are employed to narrow down the pool of candidate sgRNAs (Figure 1.9). Firstly, survival (Figure 1.9A) or dropout (Figure 1.9B) screens rely, respectively, on an advantage or disadvantage that the CRISPR/Cas9-inflicted mutations will lend the cells in the experimental group over the cells in the control group, or the same cells at an early time point. An example of a survival screen is Sanjana's study of regions that convey sensitivity to the BRAF inhibitor vemurafenib in melanoma cells: sgRNAs that disrupted regions responsible for enhancing the expression of the genes causing sensitivity to vemurafenib (as identified in (Shalem et al., 2014)) were enriched, as they provided an advantage to cells in vemurafenib culture (Sanjana et al., 2016). An example of a dropout screen is Korkmaz's study of ER- α -bound enhancers. In ER- α -addicted breast cancer cell lines, sgRNAs targeting ER- α -bound enhancers that are depleted after a period of culture pinpoint enhancers that drive proliferation in these cells (Korkmaz et al., 2016). The second strategy is based on the enrichment of populations with differential expression of the gene of interest or its endogenous fluorescent tag (Figure 1.9C). For example, Sen and co-workers used an antibody against their gene of interest, *PD-1*, to sort EL4 cells transduced with an sgRNA library targeting putative regulatory regions of *PD-1* into PD-1^{low} and PD-1^{high} groups, and by comparison, found eight positive regulatory elements for *PD-1* (Sen et al., 2016). Canver and colleagues, in their study of *BCL11A* regulatory landscapes, FACS-sorted their cells into populations with high and low expression of foetal haemoglobin as an inverse correlative measure of *BCL11A* levels (Canver et al., 2015). Meanwhile, some authors chose to endogenously tag their genes of interest with EGFP and then used FACS to sort the cells into groups according to GFP intensity (Canver et al., 2020; Diao et al., 2016; Rajagopal et al., 2016).

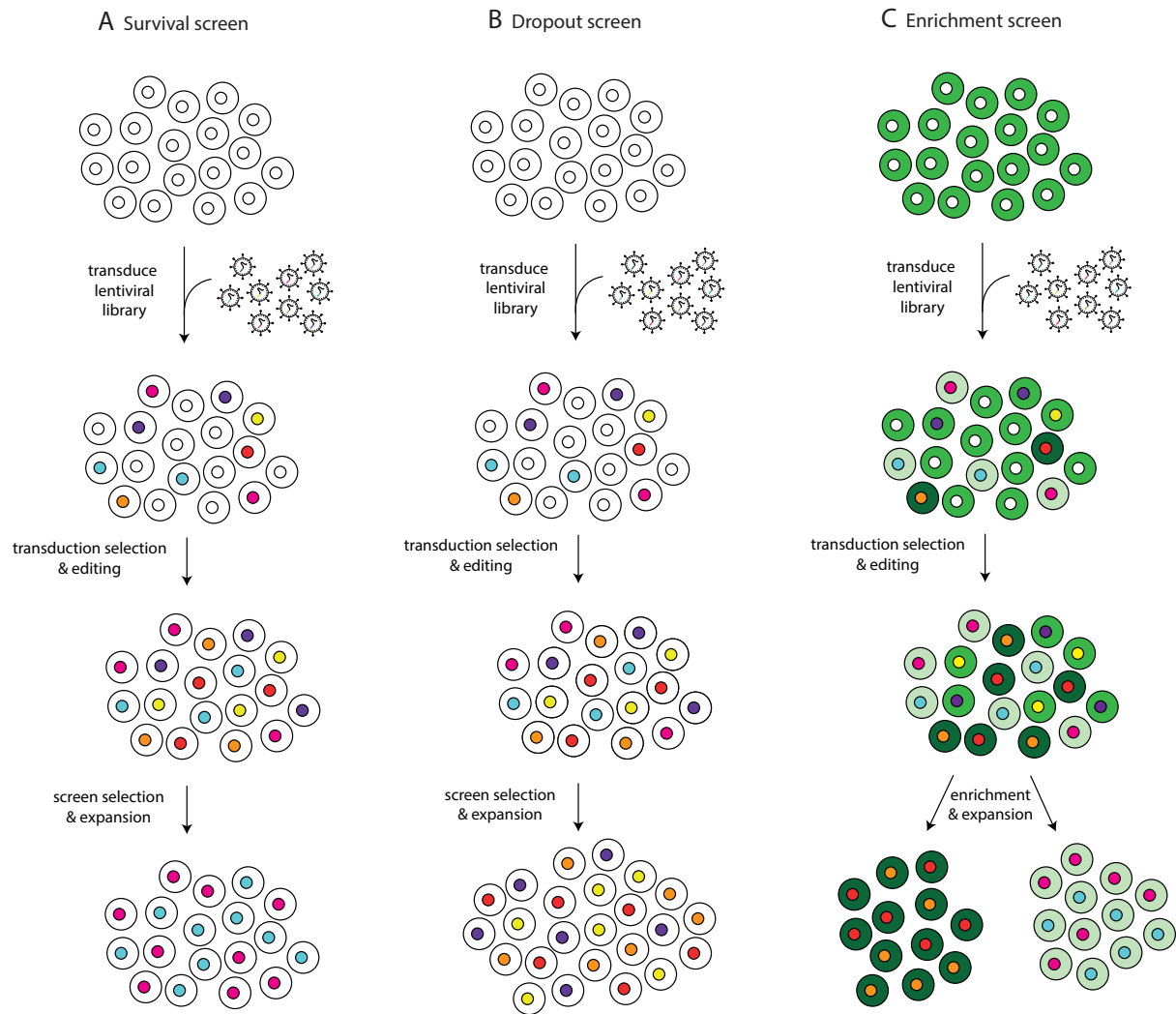


Figure 1.9: CRISPR screen strategies.

(A) A survival screen is designed to discover what mutations lead to an increased survival or resistance in response to a specific condition (e.g. a therapeutic drug). This type of screen usually results in a pool of cells with a handful of different modifications. **(B)** A dropout screen identifies mutations that are either lethal or increase sensitivity under the chosen screen conditions. Most of the mutations are expected to be represented in the final sample. **(C)** In an enrichment screen, the expression of a gene of interest is linked to a fluorescent signal, and cells with different levels of fluorescence can be enriched using FACS.

When it comes to Cas9 nuclease, some studies prefer to use cell lines stably expressing Cas9 (Canver et al., 2015, 2020; Sanjana et al., 2016; Sen et al., 2016), while others simply introduce Cas9 at the same time as the sgRNA library (Korkmaz et al., 2016; Rajagopal et al., 2016).

Most of the studies relied on transduction for stable integration of sgRNAs into the genome of the examined cells. This strategy requires working at low multiplicity of infection (MOI) to achieve the introduction of only a single sgRNA into the majority of the cells. However, excitingly, one study introduced a different approach that also resulted in stably integrated sgRNAs, but allowed the use of electroporation for sgRNA library delivery. Rajagopal and colleagues created a mouse embryonic stem cell line with one copy of a stably integrated cassette containing a dummy guide RNA (gRNA) hairpin under control of U6 promoter. The guide RNA portion of the dummy hairpin could then be replaced by gRNAs from a focused library via homology-directed repair. The method requires adding homology arms to all the gRNAs in the library by PCR. This library is then electroporated into the cell line along with Cas9 and a plasmid coding for an sgRNA targeting the dummy portion of the integrated hairpin (Rajagopal et al., 2016). This approach guarantees the integration of a single guide RNA in each cell without the necessity of targeting only a small number of cells to begin with (working at low MOI), and with the need for cloning the library into a plasmid removed, the CRISPR/Cas9 screen itself can be put into motion within hours from when the oligomers arrive. On the other hand, there is a dependency on the level to which the cell line uses the homology-directed repair pathway, and it necessitates the establishment of a stable cell line before this experiment can commence.

When it came to validating the hits from the screen, most groups chose to clone the individual sgRNAs into a vector, transduce their original cells with it and assess the expression of the target gene (Korkmaz et al., 2016; Rajagopal et al., 2016; Sanjana et al., 2016; Sen et al., 2016), and in one case, the change in histone modifications at the target sites (ChIP for H3K27ac and H3K4me2 at enhancers; (Sanjana et al., 2016)). Both Canver's studies validated their screen results by deleting the implicated regions via pairwise sgRNA deletions (Canver et al., 2015, 2020) and one of them also by transducing the top-scoring sgRNA individually into a different cell line (Canver et al., 2015). Diao and co-workers validated the candidate regulatory elements that emerged from their screen in a classic reporter assay, cloning them individually into a luciferase plasmid (Diao et al., 2016). In the same study, the authors also confirmed the direct effect of the sgRNA-inflicted mutations by preparing monoallelic clonal deletions of some of the candidate regions, either on the same allele as their EGFP-tagged gene, or the wild-type allele, and comparing the EGFP output from the two clonal cultures (Diao et al., 2016). Overall, the percentage of false-positive hits in the CRISPR/Cas9 screens was very low. Many of the hits mapped onto predicted transcription factor binding sites (Canver et al., 2015; Rajagopal et al., 2016; Sanjana et al., 2016; Sen et al., 2016).

In sum, a number of CRISPR-Cas9-based editing screens dissected regulatory landscapes of potential therapeutic targets in either a saturating mutagenesis fashion, or by pre-selecting regions based on telltale signs of regulatory elements. A library of sgRNAs was most often virally transduced into cells

that either stably expressed Cas9, or were co-transduced with it. Important regions were discovered either by monitoring the survival or the dropout of the differentially edited cells, or by measuring – directly or indirectly – the levels of expression of the target gene. All studies confirmed the screen hits by at least one, and more often multiple means of validation.

1.4.2.2 CRISPR/Cas9 screens introducing long deletions

While CRISPR/Cas9 screens employing single guide RNAs have proved to be a source of valuable information, some scientists argue that such screens can fail to identify a number of regulatory elements. The reasoning behind this is that the mutations caused by a single gRNA often don't cause enough perturbation in the sequence to significantly impair the functionality of a regulatory element, not least because frameshifts are irrelevant in non-coding genome. A number of groups proposed that CRISPR/Cas9-based screens introducing long deletions with a pair of gRNAs could remedy such a lack of sensitivity (Figure 1.10; (Diao et al., 2017; Gasperini et al., 2017; Zhu et al., 2016)). Indeed, this claim was proved legitimate in Bing Ren's group, who used a long deletion-based CRISPR/Cas9 screen on a region they previously dissected using CRISPR/Cas9-mediated mutagenesis with sgRNAs, and found five new regulatory elements that scored as negative in the previous study (Diao et al., 2016, 2017).

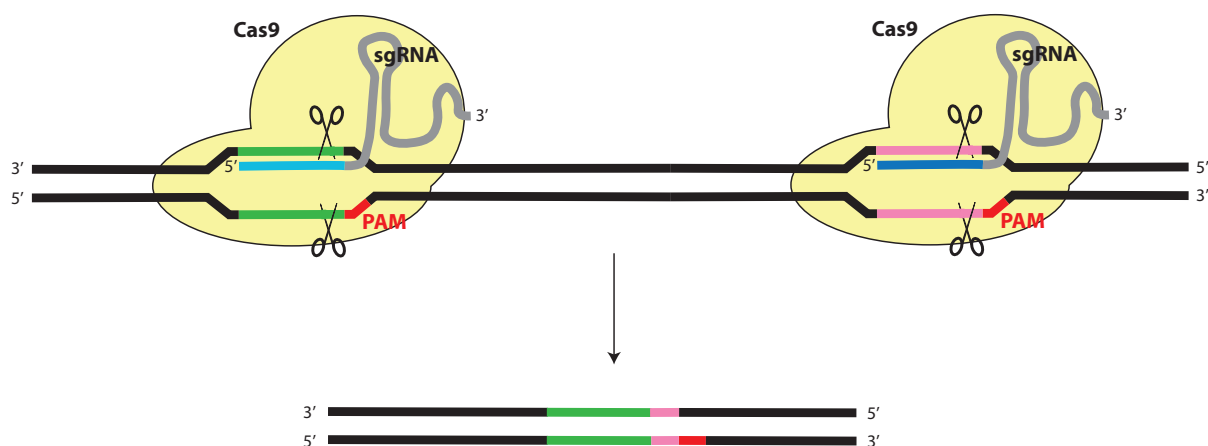


Figure 1.10: CRISPR/Cas9-mediated pairwise genomic deletions.

These deletions are achieved by the CRISPR/Cas9 system when the Cas9 nuclease is directed to two target regions, typically several kb apart, within one cell.

The sensitivity of a paired guide RNA (pgRNA) screen can be further improved by introducing overlapping genomic deletions, thus increasing the coverage of the screen (Diao et al., 2017; Gasperini et al., 2017). For example, Diao and co-workers, in their screen of *POU5F1* regulatory landscape, introduced 2-kb deletions throughout the 2-Mb locus with a 1.9-kb overlap between two adjacent deletions, achieving a 20-fold coverage (Diao et al., 2017). Similarly, Gasperini and colleagues programmed overlapping 1-kb and 2-kb deletions tiling the non-coding genome around *HPRT1*, achieving a median 27-times redundancy per base (Gasperini et al., 2017).

A further advantage of the paired guide RNA CRISPR/Cas9-mediated deletion screen is a lower dependence on an even distribution of PAM sites (Diao et al., 2017; Gasperini et al., 2017). In a similar vein, it is not hard to imagine that this approach can also help, to an extent, with screening regions containing repetitive elements.

Finally, this twist on the classic CRISPR/Cas9-mediated mutagenesis also allows to screen larger portions of the genome due to a significantly reduced requirement for the number of guide RNA vectors (Diao et al., 2017).

Along with these multitudinous advantages, however, this approach also brings many challenges and limitations that are not associated with single guide RNA-mediated CRISPR/Cas9 screens. Perhaps the most troublesome among these hurdles is the fact that the introduction of paired guide RNAs often results in a large percentage of various unintended by-products, including inversions and indels at the gRNA target sites with or without the intended deletion (Canver et al., 2014). The percentage of deletions versus other modifications that take place seems to be largely dependent on the cell line, size of the deletion, and even the individual pair of sgRNAs (Canver et al., 2014; Zhu et al., 2016). This can be partially mitigated by increasing the coverage of the screen, as discussed above, although sometimes even that can fail to preclude artefacts, as exemplified by the false positive signals in Gasperini's study of *HPRT1*'s regulatory landscapes (Gasperini et al., 2017). These signals were generated by deletions that were on target, but with incorrect boundaries, and thus extending into transcribed portions of the gene. Such editing outcomes were rare initially, but became strongly enriched during the selection process, and were only uncovered by meticulous hit validation by sequencing (Gasperini et al., 2017). Another strategy to circumvent the problem of mixed editing outcomes is to establish deletion clones, as done in Yamazaki's study of *NEAT1* lncRNA in the context of paraspeckle assembly (Yamazaki et al., 2018), although this reduces the screen to an arrayed format.

An early challenge of long deletion CRISPR/Cas9 screens has been finding means to increase the efficiency of the deletion editing. In the first proof-of principle experiment, Cong and colleagues

observed a 1.6% efficacy when endeavouring to delete a 119-bp fragment from human *EMX1* locus (Cong et al., 2013). This experiment was done in the easy-to-transfect HEK293FT cell line with a plenitude of plasmid material (800 ng per well on a 24-well plate) and samples were collected 72 hours post transfection. This is a rather low efficiency for a pooled screen, where cells need to be transduced at a low MOI. Fortunately, subsequent experiments showed that increasing the incubation time after the introduction of an sgRNA pair can substantially improve the editing efficiency (Gasperini et al., 2017; Zhu et al., 2016). Gasperini and colleagues achieved 10 – 20% efficiency depending on the system used and when assessed a week after transduction (Gasperini et al., 2017). Zhu and co-workers observed a similar percentage of deletions a week post transduction, but also noted the editing continued to take place after that, reaching a plateau about 15 days post transduction. Depending on the pair of sgRNAs used, they observed a final efficiency of genomic deletions between 50 and 95% (Zhu et al., 2016). Time really appears to be a healer in this particular scenario, and if the setup of the experiment permits, cells should be allowed to proliferate for at least 2 weeks after the introduction of the paired sgRNA library to achieve the best genomic deletion efficiencies possible (Zhu et al., 2016).

Another early obstacle to pooled long genomic deletion CRISPR/Cas9 screens was the lack of a suitable cloning technique for pools of paired sgRNA vectors. Two groups have published independent methodology articles tackling this issue (Aparicio-Prat et al., 2015; Vidigal and Ventura, 2015), proposing similar two-step cloning protocols. Both approaches use a pool of synthesized DNA oligonucleotides, in which each oligonucleotide includes a unique pair of gRNAs. These oligonucleotides are first inserted into an intermediate structure containing one Polymerase III promoter and one gRNA scaffold, and a second Polymerase III promoter and gRNA scaffold are then added in the next step. Either of the cloning techniques provides a straightforward way for cloning pools of paired gRNAs for long deletion CRISPR/Cas9 screens (Aparicio-Prat et al., 2015; Vidigal and Ventura, 2015).

Yet another hurdle can be the levels of viral recombination within the paired gRNA pool (Aparicio-Prat et al., 2015; Gasperini et al., 2017; Vidigal and Ventura, 2015). Fortunately, this can be easily prevented to a large degree by using two different Polymerase III promoters, such as human U6 and human H1 promoters (Aparicio-Prat et al., 2015; Gasperini et al., 2017), or human and chimeric mouse-human U6 promoter (Vidigal and Ventura, 2015).

While not exclusively an issue of CRISPR/Cas9 screens introducing long genomic deletions, the ambiguity that can be caused by differential editing on the (usually) two alleles can present a complex problem in a long deletion CRISPR/Cas9 screen, considering the breadth of potential unintended products (Canver et al., 2014). An elegant solution to this problem is the use of haploid cells, such as

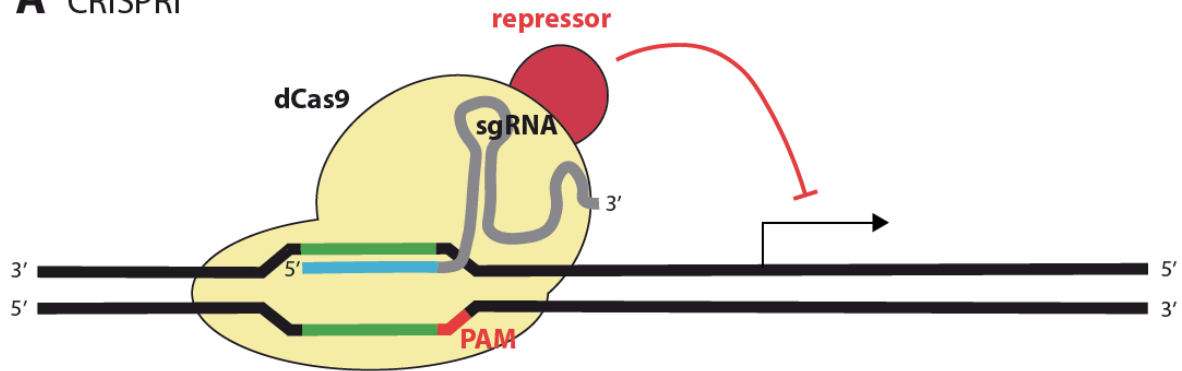
the human HAP1 cell line (Gasperini et al., 2017; Yamazaki et al., 2018). Another graceful solution is using a cell line with an endogenous EGFP tag of the gene of interest on one allele to pinpoint *cis*-regulatory elements by examining a population of cells that has low EGFP fluorescence, but relatively unchanged global levels of the gene of interest (Diao et al., 2017). A similar strategy based on the same cell line was used by the same group for their sgRNA CRISPR/Cas9 screen (Diao et al., 2016), as discussed in the previous chapter. As always, and especially if no such solution is applied to the screen at hand, rigorous hit validation by multiple means needs to take place. For example, Zhu and colleagues, in their search for lncRNAs that play either a positive or a negative role in liver cancer cell (Huh7.5_{OC}) proliferation, validated their hits by editing with additional pgRNAs, as well as with CRISPRi or CRISPRa (Zhu et al., 2016).

In sum, pairwise genomic deletions can have the advantage of increasing the sensitivity of CRISPR/Cas9-based editing screens, but this comes at a cost of a number of technical challenges.

1.4.2.3 CRISPR/dCas9-based transcriptional regulation

While screens using CRISPR/Cas9 for non-coding genome editing have proved to be a source of invaluable information, and extensive optimization efforts have been deployed, some of the associated limitations still remain. This includes variable efficiency of editing and size of the effect, unpredictable editing by-products and differential editing of the targeted locus on each allele in non-haploid cell lines (Gasperini et al., 2020). Transcriptional regulation based on catalytically dead Cas9 mutant (dCas9), an inactive version of Cas9 nuclease (Qi et al., 2013), can alleviate some of these problems, as well as potentially provide an orthologous method to confirm results from editing CRISPR-Cas9 screens. Transcriptional silencing based on CRISPR/dCas9 fusions with transcriptional repressors is collectively known as CRISPR interference (CRISPRi; Figure 1.11A; (Qi et al., 2013)); transcriptional activation achieved with CRISPR/dCas9 fusions with transcriptional activators is termed CRISPR activation (CRISPRa; Figure 1.11B; (Bikard et al., 2013; Cheng et al., 2013; Gilbert et al., 2013; Perez-Pinera et al., 2013)).

A CRISPRi



B CRISPRa

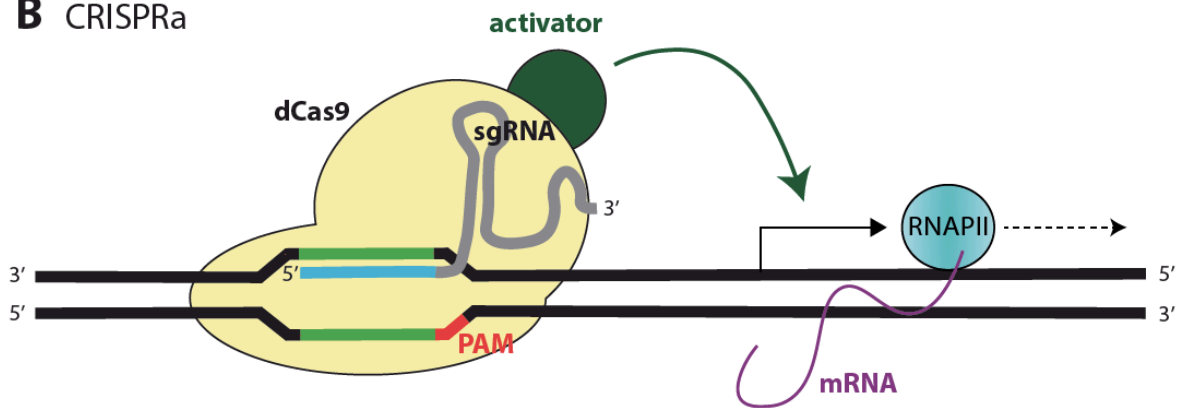


Figure 1.11: Transcriptional regulation by CRISPR/dCas9-fusions.

(A) Fusing a repressor, such as the KRAB domain, to a dCas9 and targeting the complex via an sgRNA to transcribed regions results into transcriptional interference (CRISPRi). **(B)** Transcriptional activators, such as VP16 as a repeating unit, targeted via the CRISPR/dCas9 system to transcribed regions, activate or increase transcription of the target genes (CRISPRa).

1.4.2.3.1 CRISPR interference

The idea of using a dCas9 as a transcriptional block was first successfully realized in *E. coli*, with up to 1000-fold repression levels (Qi et al., 2013). While the group also tested the system in human HEK293 cells, they only achieved a two-fold decrease in gene expression with the most effective of their sgRNAs. In their following paper, the same group created dCas9 fusions with three different domains with known transcriptional repression function, identifying a dCas9 fusion with Krüppel-associated box domain (KRAB) as the most effective one (Gilbert et al., 2013). KRAB domain, a 75-amino acid, highly conserved region present in many zinc finger proteins, was described as a transcriptional repressor a quarter of a century ago (Margolin et al., 1994), and the mechanism of the gene silencing

it mediates has been the subject of a number of subsequent studies (Groner et al., 2010; Schultz et al., 2002; Sripathy et al., 2006). When using the dCas9-KRAB fusion for epigenetic editing in a GFP-expressing HEK293 cell line, Gilbert and colleagues observed at least a three-fold decrease in GFP expression for 6 out of 8 sgRNAs, and a 15-fold decrease for the best sgRNA (Gilbert et al., 2013). Furthermore, they could knock down endogenous genes in HeLa cells by means of the dCas9-KRAB fusion, although with a lower efficiency, between two- and five-fold; and, in important news for non-coding genome, also observed gene silencing upon targeting its promoter with the same fusion protein. In a following study, yet the same group systematically assessed tens of thousands of different sgRNAs, performing first CRISPRi screens and creating a set of rules for efficient, on-target gene repression with typical knockdown efficiencies between 80 and 99% (Gilbert et al., 2014).

The first attempt at targeting an enhancer region with CRISPRi was carried out in Gao's study, which compared the efficiencies of TALE and dCas9 platforms for gene silencing via enhancer targeting, and pronounced dCas9-KRAB a superior repressor to TALE-KRAB (Gao et al., 2014). By performing CRISPRi of a specific DNaseI hypersensitivity site (HS2) within the globin locus control region, a complex regulatory element for haemoglobin gene expression, Thakore and co-workers showed that dCas9-KRAB binding is highly specific in genome-wide context and results in highly specific gene repression, accompanied by near-perfect specificity of H3K9me3 deposition and DNase I hypersensitivity changes (Thakore et al., 2015).

A number of groups went on to use dCas9-KRAB in a screen format (Fulco et al., 2016, 2019; Gasperini et al., 2019; Klann et al., 2017; Xie et al., 2017). Two of the studies concentrated on surveying the non-coding regulatory landscape of one gene at a time with high density of sgRNAs (Fulco et al., 2016; Klann et al., 2017). For example, for their dissection of a 74-kb region surrounding GATA1, Fulco and colleagues designed a library of sgRNAs with a 16-bp average spacing and used a sliding window approach in their evaluation, averaging the effects of 20 consecutive sgRNAs. This helped them to deal with the common problem of a varying sgRNA efficiency while providing them with high-confidence results (Fulco et al., 2016). Other groups instead chose to focus on a higher number of candidate enhancers at the cost of limiting the number of sgRNAs used to 2 – 4 per each element (Fulco et al., 2019; Gasperini et al., 2019; Xie et al., 2017). As one example, a method introduced by Fulco and co-workers, CRISPRi-FlowFISH, uses FACS to bin-sort cells labelled for an RNA of interest by fluorescence *in situ* hybridization (FISH), upon an infection with a CRISPRi sgRNA library at low MOI (Fulco et al., 2019). While infections at low MOI have been the golden standard for CRISPR screens in general, two recent studies described how performing a CRISPRi screen at higher MOI, when coupled with single cell RNA sequencing, can increase power while reducing the number of cells needed in an experiment (Gasperini et al., 2019; Xie et al., 2017). The study from Hon lab probed 71 enhancer constituents of

15 super-enhancers in a set of experiments using between 1.1 and 3.2 sgRNAs per cell, pinpointing elements with an effect between 18 to 88.9% of target gene expression (Xie et al., 2017). Gasperini and co-workers went even further, assessing the activity of 5779 candidate enhancers with a median of 28 sgRNAs per cell, allowing them to discern effects ranging from 1.4% to 97.5% of target gene expression (Gasperini et al., 2019). While these multiplexed experiments can be done on a large scale with a much-improved power compared to classic low MOI screens, and have the ability to disentangle penetrance (the percentage of cells in which the given enhancer is active) and the actual contribution of the given enhancer to the target gene expression, this emerging type of screen still suffers from several limitations common to CRISPR-based screens, and additionally bears the high costs associated with single cell sequencing (Gasperini et al., 2019; Xie et al., 2017).

A number of other groups proceeded to explore the functionality of dCas9 fusions with other domains known to impart repressive epigenetic marks, with varying success (Huang et al., 2017; Kearns et al., 2015; Kwon et al., 2017; Liu et al., 2016; Vojta et al., 2016). The most interesting study from the perspective of enhancer dissection was the work of Kearns and colleagues (Kearns et al., 2015), showing that a dCas9 fusion with lysine-specific demethylase 1 (LSD1), a H3K4/K9-demethylase essential for enhancer silencing (Whyte et al., 2012), efficiently repressed gene transcription through epigenetic changes when targeted to enhancers, but not promoters (Kearns et al., 2015). This selectivity of dCas9-LSD1 was in opposition to the broader specificity of dCas9-KRAB, which indiscriminately repressed transcription when targeted to both enhancers and promoters (Kearns et al., 2015). The other fusions were only used to target promoters: employing a DNA methyltransferase fusion, dCas9-DNMT3A (DNA methyltransferase 3 alpha), consistently achieved specific methylation at CpG islands accompanied by strong target gene repression (Huang et al., 2017; Liu et al., 2016; Vojta et al., 2016), while the use of a histone deacetylase fusion, dCas9-HDAC3 (histone deacetylase 3), led to opposing results under different experimental conditions (Kwon et al., 2017).

Overall, dCas9-KRAB and dCas9-LSD1 proved to be useful fusions for CRISPR interference at enhancers.

1.4.2.3.2 CRISPR activation

CRISPR activation, or CRISPRa, emerged at the same time as its inhibitory twin, CRISPRi, and was also first used for targeting promoters (Bikard et al., 2013; Cheng et al., 2013; Gilbert et al., 2013; Perez-Pinera et al., 2013). The first dCas9-based activators were mostly created using the minimal

transcriptional activation domain VP16 (Greaves and O'Hare, 1989; Triezenberg et al., 1988) as a repeating unit. VP16 acts as a scaffolding for a number of transcription factors (Hirai et al., 2010). Initially, three VP16 domains were fused into VP48, without any success at endogenous gene activation (Cheng et al., 2013), but more fruitful fusions followed suit – four VP16 domains fused into VP64 (Gao et al., 2014; Gilbert et al., 2013; Perez-Pinera et al., 2013) and ten VP16 domains fused into VP160 (Cheng et al., 2013; Gao et al., 2014). Gao's study showed that the efficiency of VP64 and VP160 was comparable (Gao et al., 2014). Additional transactivators used in early studies included the ω subunit of the bacterial RNAP (Bikard et al., 2013) and the p65 activation domain (Gilbert et al., 2013). Overall, the dCas9-VP16-based fusions exhibited a high genome-wide specificity (Cheng et al., 2013; Hilton et al., 2015; Perez-Pinera et al., 2013; Polstein et al., 2015) and a synergistic effect of multiple sgRNAs against the same regulatory element (Cheng et al., 2013; Perez-Pinera et al., 2013).

Several groups used dCas9-VP64 and/or VP160 to target enhancers (Gao et al., 2014; Ginley-Hidinger et al., 2019; Hilton et al., 2015; Sano et al., 2020; Simeonov et al., 2017). In an unbiased saturating screen of the 135-kb CD69 locus, Simeonov and colleagues showed that dCas9-VP64, in conjunction with a targeting sgRNA library, was able to pinpoint known regulatory elements (Simeonov et al., 2017). The group then proceeded to screen 178 kb of genome surrounding *IL2RA*, discovering previously undefined regulatory regions within the *IL2RA* super-enhancer. Using dCas9-VP64 together with individual sgRNAs or sgRNA tandems, Sano and colleagues targeted multiple potential enhancer and promoter regions of genes involved in cardio-specific differentiation, in a hope to trigger the development of cardiomyocytes upon their loss in a heart attack, with favourable results in a rat model (Sano et al., 2020). In general, most studies that assessed the target gene enrichment in response to dCas9-VP64 or dCas9-VP160-mediated activation through enhancer targeting found the mRNA levels of the genes of interest significantly elevated several fold compared to controls (Gao et al., 2014; Ginley-Hidinger et al., 2019; Sano et al., 2020). However, Hilton and colleagues found a very limited, if any, activation potential of dCas9-VP64 at all five tested enhancers, even when multiple sgRNAs were administered for synergy (Hilton et al., 2015).

Apart from setting forth a discouraging set of results for dCas9-VP64-mediated activation of gene expression from enhancer regions, Hilton's work also introduced a new promising transactivating fusion protein, dCas9-p300 (Hilton et al., 2015). p300 is a histone acetyltransferase (Ogryzko et al., 1996) responsible for the deposition of the active H3K27ac marks on chromatin (Tie et al., 2009). In the experimental context of Hilton's study, the effects of the dCas9-p300 fusion were extremely powerful when targeted to either enhancers or promoters, even by individual sgRNAs (Hilton et al., 2015). For example, selected individual sgRNAs against enhancers increased transcription of the target genes over 20 times. Somewhat at odds with this report, even when focused at different loci in a

different cell line, a later study presented dCas9-p300 as only a modest transcriptional activator, largely dependent on the RNAPII amounts already present at the targeted site (Ginley-Hidinger et al., 2019). Nevertheless, Chen and co-workers successfully modulated transcriptional activity in neurons by targeting dCas9-p300 to *Fos* and *Npas4* enhancers (Chen et al., 2019), and Klann and colleagues employed the dCas9-p300 fusion for screens of the beta-globin locus and the *HER2* regulatory landscapes, finding a generally good correlation with a CRISPRi screen they carried out alongside (Klann et al., 2017). Interestingly, however, Kuscu and colleagues reported that sequences displaying no features of regulatory activity could be turned into enhancer-like elements by virtue of dCas9-p300 targeting (Kuscu et al., 2019). As a result of novel H3K27 acetylation deposited by the dCas9-p300 fusion, these induced enhancers formed new 3D connections and increased expression of nearby genes. This finding, along with the fact that some active regulatory elements failed to be identified in Klann's dCas9-p300 screen, means that caution should be strongly exercised when making conclusions about existing regulatory elements within a region from gain-of-function experiments, with a loss-of-function experiment ideally performed as a complementary method (Klann et al., 2017; Kuscu et al., 2019).

Finally, several other transactivating fusions with varying degrees of potency have been described, including dCas9-Tet1 human demethylase (Liu et al., 2016), dCas9-CBP acetyltransferase from *Drosophila* (Sajwan and Mannervik, 2019) and dCas9-ROS1 demethylase from *Arabidopsis* (Devesa-Guerra et al., 2020).

In sum, a usually efficient CRISPR activation can be achieved by using dCas9 fusions with VP64, VP160 or p300, although the method may often lead to false positive as well as false negative results. Therefore, rigorous hit validation, perhaps along a complementary CRISPRi screen, are needed to lend credibility to CRISPRa results.

1.4.2.4 Lessons from CRISPR/Cas9-based screens of non-coding genome

Screens based on the CRISPR/Cas9 technology have been successfully deployed to dissect various portions of non-coding genome to reveal functional relationships between specific regulatory elements and their target genes. Moreover, though, some of the discussed studies also came to locate new regulatory elements and to discover (or confirm) more general truths about transcriptional regulation.

The potential for discovery of new regulatory elements belongs to those studies that venture to dissect the non-coding landscape in an unbiased fashion. In several such endeavours, a number of regulatory regions was observed that didn't coincide with features habitually associated with enhancers or regulatory regions in general (Diao et al., 2017; Rajagopal et al., 2016; Sanjana et al., 2016). Rajagopal and his colleagues described these genomic elements, which are often over 1 kb long and sensitive to base substitution in parts, as unmarked regulatory regions (UREs), noting that they influence their target gene expression at a similar level to some distal enhancers (Rajagopal et al., 2016).

Another interesting group of regulatory elements found by virtue of CRISPR-Cas9-based screens are temporary enhancers. These regions showed high DNase I hypersensitivity and gave weak, but statistically significant positive results in a luciferase reporter assay. In cell culture, they temporarily, yet detectably downregulated expression of their target gene, *POU5F1* (Diao et al., 2016, 2017). The authors theorized that the reason why these elements only function in a temporary fashion may be that they are only needed transiently or that they are replaceable by other similar elements upon their disruption. Speculating over the potential roles of such elements, Diao and co-workers suggest that they may be needed at transcription initiation, or function as a scaffold of the local chromatin architecture, or facilitate the association of their target gene with transcription factories (Diao et al., 2016).

Examining the nature of interactions between regulatory elements, some groups noticed that promoters can act as enhancers for neighbouring genes (Diao et al., 2017; Gasperini et al., 2019; Rajagopal et al., 2016). For example, Rajagopal and colleagues noted that while the non-coding landscapes of each of the four genes that they studied were different, a common denominator was the regulation of each of the genes by promoters of other genes, some of which were up to several Mb away (Rajagopal et al., 2016). Diao and collaborators found that there was a total of 18 promoters which participated in the regulation of their gene of interest, *POU5F1*, in human embryonic stem cells (Diao et al., 2017).

Some of the studies also confirmed established beliefs about enhancers. Several groups observed a competition for shared enhancers and/or regulation of multiple genes by a single enhancer (Fulco et al., 2016, 2019; Klann et al., 2017), and a single gene was often reported to be regulated by multiple enhancers (Diao et al., 2017; Fulco et al., 2016; Gasperini et al., 2019; Ginley-Hidinger et al., 2019). Furthermore, the range of effects of each particular regulatory element on its target gene expression was shown to be extremely wide, from meagre 1-2% contributions to almost a full responsibility for the gene transcription (Fulco et al., 2019; Gasperini et al., 2019). Finally, two of the works concluded

that the level of evolutionary conservation of the dissected regulatory regions was low between humans and non-primate species (Canver et al., 2015; Sanjana et al., 2016). Canver's study identified mouse orthologues of the human enhancers they were studying, but found their activity to be different from their human counterparts (Canver et al., 2015). In Sanjana's work, elements identified as regulatory, while conserved in primates, were 1.7 times less likely to occupy regions of genome near conservation peaks among mammals (Sanjana et al., 2016).

Overall, CRISPR-Cas9-based screens proved to be useful not only in unveiling the functional importance of regulatory elements in connection to their target genes, but also in pinpointing new functional regulatory elements and confirming previously observed phenomena.

1.5 *TAL1* enhancer in Jurkat cells

The *TAL1* super-enhancer in Jurkat cells is a monoallelic regulatory element that supports an aberrantly high expression of TAL1 (T-cell acute lymphocytic leukaemia protein 1) and drives an increased proliferation in the cell line (Mansour et al., 2014).

TAL1 is a class B basic helix-loop-helix (bHLH) protein (Hsu et al., 1994a) that functions as a transcriptional regulatory factor (Hsu et al., 1994b). TAL1 utilizes its bHLH domain to dimerize with E proteins, which are class A bHLH proteins such as E12 and E47 (E2A gene products), E2-2 and HEB (Hsu et al., 1994b). TAL1 heterodimers recognize the consensual E-box sequence, AACAGATGGT (Hsu et al., 1994b).

TAL1 plays a critical role in hematopoietic regulation (reviewed by (Porcher et al., 2017)). In normal development, TAL1 is silenced at common lymphoid progenitor stage (Zhang et al., 2005). In T-cell acute lymphoblastic leukaemias (T-ALL), which are haematological malignancies characterized by the accumulation of transformed thymocytes under maturation arrest, an aberrant expression of TAL1 is associated with up to 45% of paediatric T-ALL cases, as well as 10 – 15% of adult cases (Bardelli et al., 2021; Sanda et al., 2012). This TAL1 dysregulation may owe to one of multiple genetic abnormalities, including the t(1;14) (p32;q11) chromosome translocation (Begley et al., 1989; Bernard et al., 1990; Chen et al., 1990; Finger et al., 1989), the 90kb tal^d deletion (Brown et al., 1990) and a de novo *TAL1* enhancer creation 7.5 kb upstream from *TAL1* (Figure 1.12, Figure 1.13 and Figure 1.14; (Liu et al., 2017; Mansour et al., 2014)).

Figure 1.12 has been redacted for third-party copyright reasons.

Please find the original Figure in Mansour et al., 2014, where it is designated Figure 2A.

Figure 1.12: De novo MYB-binding site at the *TAL1* enhancer locus.

Mutations (in red) in T-ALL cell lines (Jurkat and MOLT-3) and in 8 different T-ALL patients all create new MYB binding motifs in the same region of the *TAL1* enhancer locus (from (Mansour et al., 2014)).

The -7.5 kb *TAL1* enhancer is a monoallelic feature brought into existence by mutations that introduce a novel MYB transcription factor binding site (“mutation of the *TAL1* enhancer”, MuTE; (Mansour et al., 2014)). Such newly introduced MYB binding motifs (GACGTA) were found in 5.5% or 5.7% of examined samples from paediatric and young adult T-ALL patients (Figure 1.12; (Liu et al., 2017; Mansour et al., 2014)). Moreover, not one, but two consecutive novel MYB binding sites were discovered upon examination of the *TAL1* enhancer locus in the Jurkat cells (Figure 1.12; (Mansour et al., 2014)).

A ChIP-seq experiment in Jurkat cells confirmed the presence of MYB at the MuTE site and revealed the binding of multiple other transcription factors, including GATA3, RUNX1 and HEB, as well as TAL1 itself and the histone acetyltransferase CBP (Figure 1.13; (Mansour et al., 2014)). An enrichment for RNAPII and Mediator protein spanned downstream from the MuTE site for over 20 kb (Figure 1.13), as did the aberrant H3K27 acetylation, which also spread for many kb in the upstream direction from the MuTE site (Mansour et al., 2014). 3C mapping of the regulatory landscape of *TAL1* in Jurkat cells confirmed high-frequency contacts between the *TAL1* promoter and the region around 500 bp upstream from the MuTE site (Zhou et al., 2013), and the CRISPR-Cas9-mediated excision or disruption of the aberrant MYB-binding sites led to a profound drop in *TAL1* levels in Jurkat cells (Mansour et al., 2014). Collectively, these findings strongly indicate that this region in Jurkat cells is a de novo super-enhancer largely contributing to the regulation of *TAL1* expression (Figure 1.14).

Figure 1.13 has been redacted for third-party copyright reasons.

Please find the original Figure in Mansour et al., 2014, where it is designated Figure 3A.

Figure 1.13: MYB and members of TAL1 co-activator complex bind at the mutation site at the *TAL1* enhancer.

ChIP-seq tracks (in reads per million reads) show the presence of MYB, GATA3, HEB, RUNX1, TAL1, CBP, RNAPII, and Mediator 1 (MED1) at the MuTE site in Jurkat cells (from (Mansour et al., 2014)).

The fact that MYB knockdown in Jurkat cells resulted in significant *TAL1* downregulation, coupled with the fact that the CRISPR/Cas9-mediated removal of one or both of the MYB-binding sites in the *TAL1* enhancer in Jurkat cells led to a 55 – 85% decrease in *TAL1* expression, suggested that MYB is the first factor to bind to the MuTE locus and nucleate the transcriptional complex present at the *TAL1* enhancer in Jurkat cells (Mansour et al., 2014). In all likelihood, upon binding, MYB recruits CBP, which in turn acetylates chromatin locally, opening it up for other transcription factors, including GATA3, RUNX1 and HEB (Mansour et al., 2014). These transcription factors, including TAL1, form a so-called TAL1 (co-activator) complex (Figure 1.14). Interestingly, a number of TAL1 co-activator complex members, namely MYB, TAL1, GATA3 and RUNX1, were found to occupy each other's and their own

regulatory elements and to be positively interconnected in an autoregulatory loop (Mansour et al., 2014; Sanda et al., 2012).

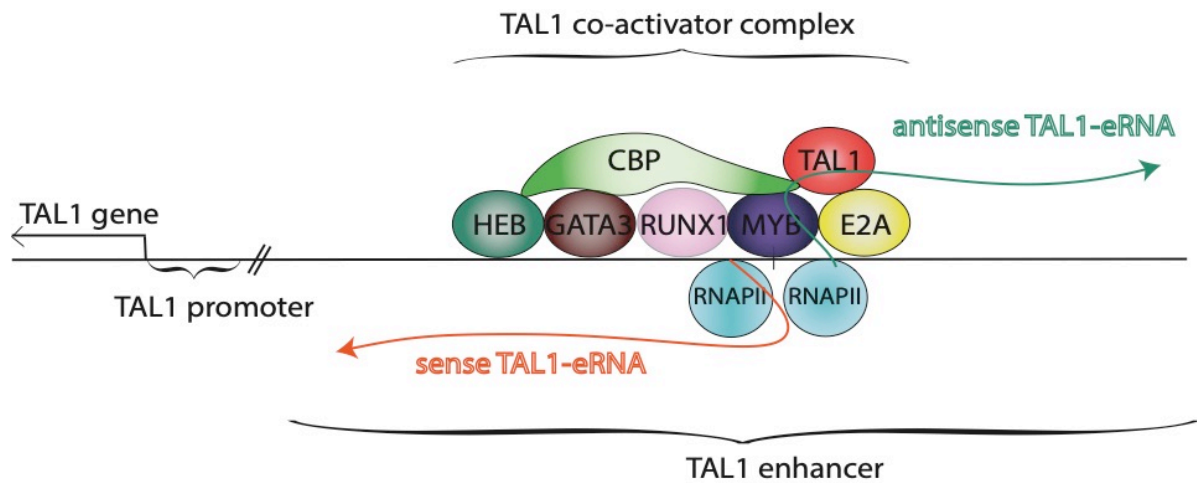


Figure 1.14: The *TAL1* enhancer and TAL1 co-activator complex.

The *TAL1* enhancer is a de novo enhancer found in Jurkat cells and some other T-ALL cells, centred around a novel MYB-binding site around 7.5 kb upstream from *TAL1* transcription start site. MYB binding initiates the formation of the TAL1 complex, which includes CBP, RUNX1, GATA3, HEB, E2A and TAL1 itself. RNAPII is recruited to the enhancer and transcribes a portion of it in a bidirectional manner.

Finally, a bidirectional eRNA transcription from the *TAL1* super-enhancer in Jurkat cells has been documented by precision run-on sequencing (PRO-seq) (Figure 1.15A, B; GEO accession: GSM1613182; (Danko et al., 2015)), confirming the existence of a sense and antisense *TAL1*-eRNA, or *eTAL1* (structure predictions in Figure 1.15C, D). It is these two eRNAs that are in the spotlight of this doctoral work.

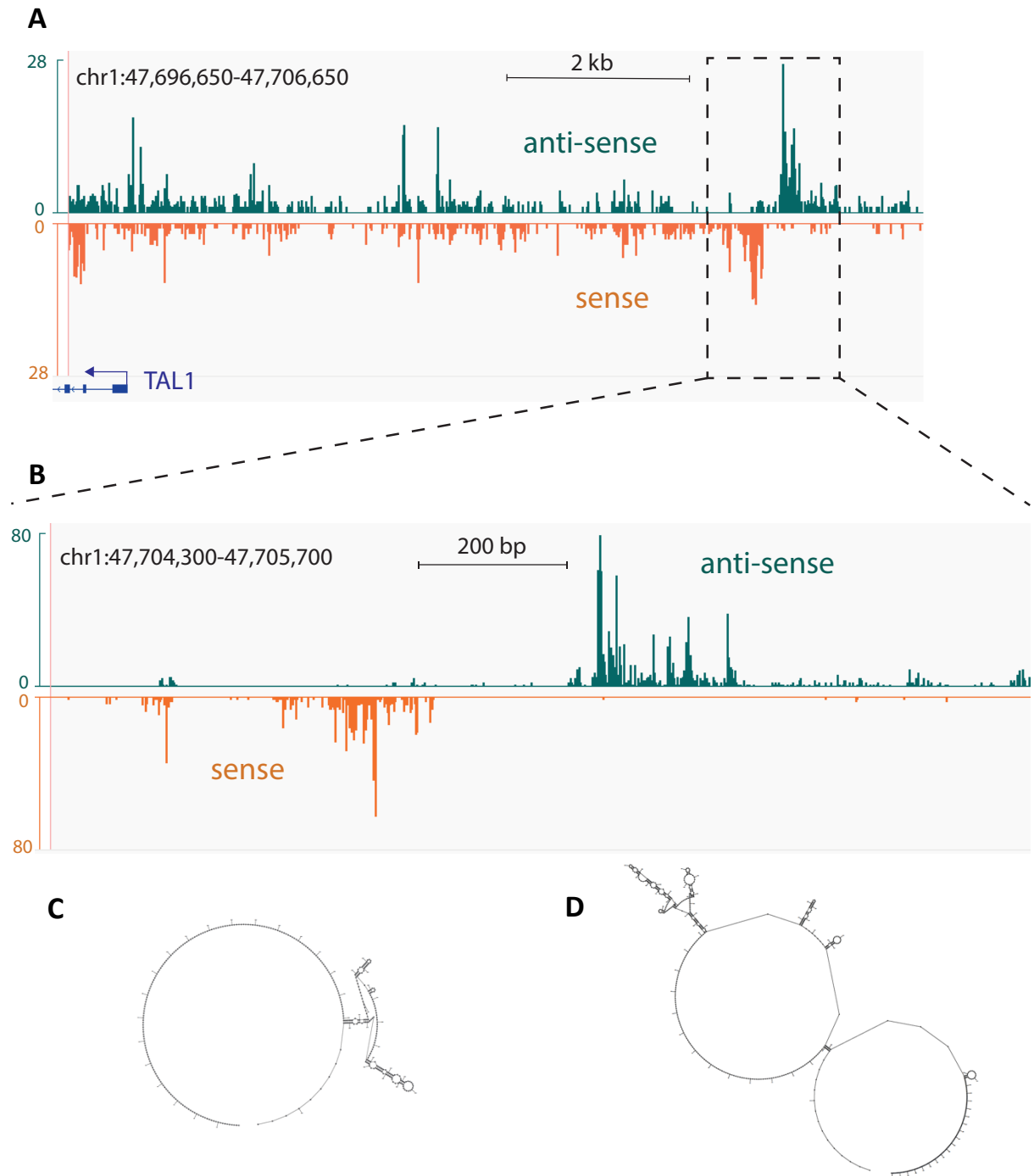


Figure 1.15: eRNA transcription at the *TAL1* enhancer in Jurkat cells.

(A) PRO-seq of the *TAL1* enhancer region documents bidirectional transcription; **(B)** detailed PRO-seq tracks in the region of the *TAL1* enhancer surrounding the MYB binding site spotlight the sense (orange) and antisense (green) eRNAs (Danko et al., 2015). The figure was made in IGV, the Integrative Genomics Viewer (Robinson et al., 2011), GRCh37/hg19 assembly. **(C, D)** Predictions of the secondary structure of **(C)** sense *TAL1*-eRNA and **(D)** antisense *TAL1*-eRNA, constructed using the Fold algorithm on the RNAstructure Web Server (Reuter and Mathews, 2010), which predicts the lowest free energy structure.

1.6 CRISPR/Cas9-mediated dissection of the *TAL1* super-enhancer in Jurkat cells

While a small wealth of information about eRNAs has been collected and a number of functions proposed for the species (Chapter 1.3), the mechanistic basis for eRNA function remains unclear. To address the enigma of eRNA functionality, we have decided to work with the de novo *TAL1* enhancer (or *TAL1* super-enhancer – these two terms are used interchangeably throughout this work) in Jurkat cells. This choice is supported by a number of reasons.

Firstly, *TAL1* enhancer in Jurkat cells is very well-defined (Chapter 1.5). Secondly, the aberrantly high *TAL1* levels are crucial for the proliferation of the cell line (Palii et al., 2011; Palomero et al., 2006; Sanda et al., 2012), which provides a convenient means of assessing changes in *TAL1* expression. Thirdly, the levels of *TAL1* in Jurkat cells are largely dependent on the *TAL1* enhancer, as evidenced by an almost complete abrogation of *TAL1* expression following the core *TAL1* enhancer removal (to the point that continued viability of cells edited in this manner had to be maintained by the expression of an exogenous *TAL1*), and a *TAL1* level drop to 15 – 45% of wild-type Jurkat *TAL1* levels in cells disrupted at the *TAL1* enhancer MYB-binding site by small, CRISPR/Cas9-mediated indels (Mansour et al., 2014). Finally, preliminary experiments from our lab (Dr. Dan Bose) have implied a functional role for *TAL1*-eRNAs – an ASO-/LNA ASO-mediated knockdown of the sense *TAL1*-eRNA led not only to a substantial drop in the levels of the *TAL1*-eRNA transcript itself, but also resulted in concurrently decreased levels of *TAL1* mRNA.

To study whether an eRNA has a function in itself or not, it is often sufficient to knock the transcript down with ASO or siRNA. However, to study what structural determinants lend functionality to an eRNA transcript, an alternative approach is called for, one that allows to inspect the functionality of discrete regions within the eRNA transcript. For this purpose, we have decided to use a CRISPR/Cas9-mediated saturating mutagenesis screen (Canver et al., 2015).

Because any significant reduction in *TAL1* levels will lead to a decrease in Jurkat cell proliferation (Palii et al., 2011; Palomero et al., 2006; Sanda et al., 2012), and because *TAL1*-eRNAs seem to play a role in the aberrant *TAL1* upregulation, any mutations in the *TAL1* enhancer that lead to a damage in functionally important regions of *TAL1*-eRNAs should be disfavoured. The CRISPR/Cas9-mediated mutagenesis screen is based on this reasoning – sgRNAs targeting critical regions of *TAL1* enhancer should be underrepresented in the experimental samples.

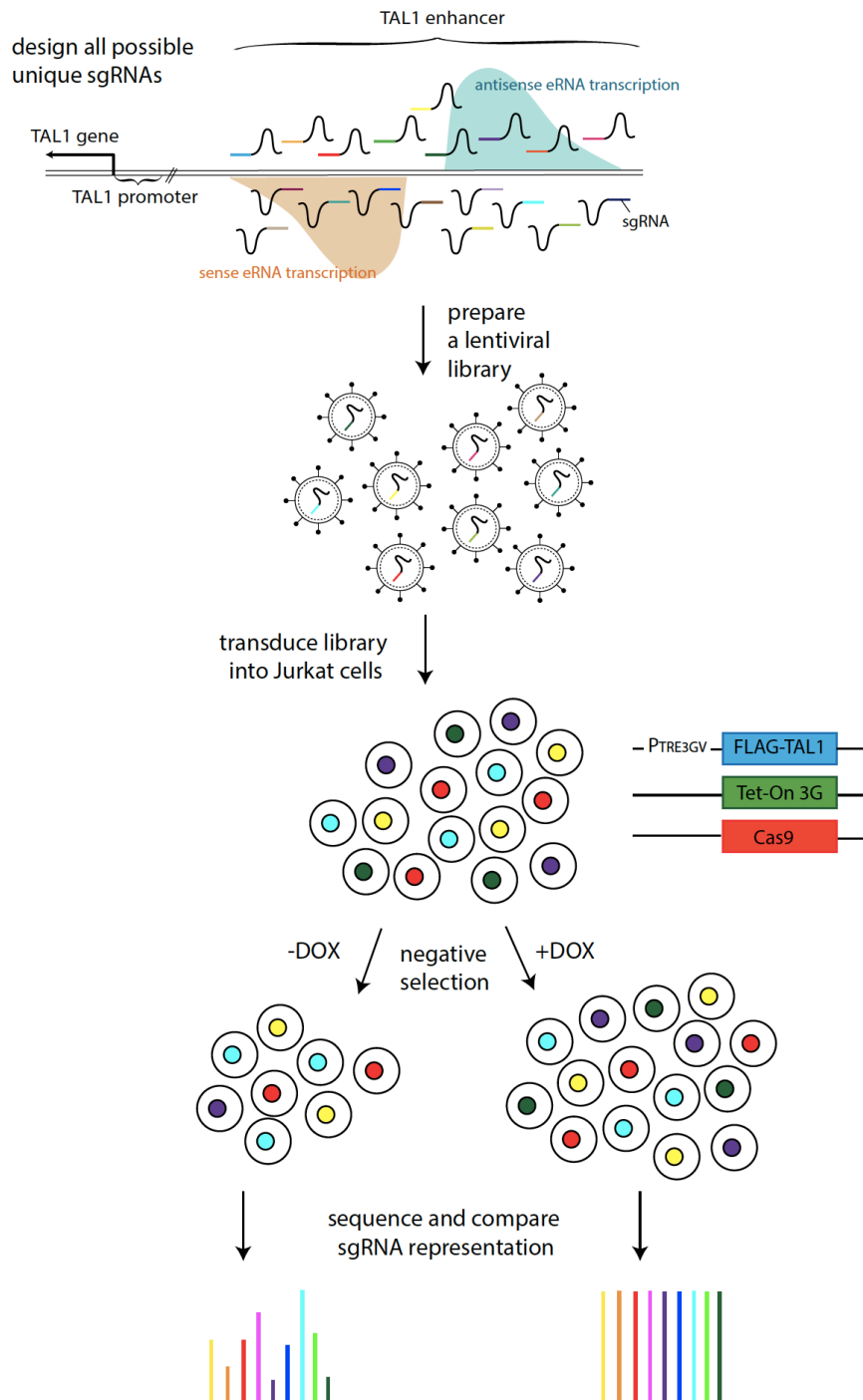


Figure 1.16: CRISPR/Cas9-mediated saturating mutagenesis of the transcribed portion of the *TAL1* enhancer in Jurkat cells.

All unique sgRNAs targeting the transcribed portion of the *TAL1* enhancer will be cloned into a plasmid and packaged into a lentiviral library for delivery into genetically modified Jurkat cells, which can inducibly express an exogenous FLAG-TAL1 and stably express Cas9 as a component of the CRISPR/Cas9 system. Upon the transduction of the library, the control group will be supplemented with doxycycline to enable the FLAG-TAL1 rescue, while the experimental group will be left to its own devices. After the editing has taken place and the editing effects have had a chance to manifest, genomic DNA from the cells will be isolated for sequencing, and the sgRNA representation will be compared between the experimental and the control group.

To provide for a control group with equal representation of sgRNAs, we decided to use a system of rescue (Figure 1.16). To this end, we have chosen to establish a cell line with doxycycline-inducible expression of exogenous FLAG-tagged TAL1. By comparison of the sgRNA representation in the experimental and the control group, we aim to build a map of important functional regions within the *TAL1*-eRNA molecules.

Chapter 3 of this work will deal with the establishment of the Jurkat cell line with a doxycycline-inducible expression of FLAG-TAL1 and a stable expression of Cas9. Chapters 4 and 5 will then tell of the validation of this cell line, and Chapter 6 will describe the necessary considerations for the CRISPR/Cas9-based dissection of the *TAL1* enhancer in Jurkat cells.

By learning about how the functionality of *TAL1*-eRNAs is conveyed, we hope to contribute to a more general understanding of how eRNAs exert their functions.

2 Materials and methods

2.1 Cloning and plasmid preparation

2.1.1 Buffers and media

50x TAE: 2 M Tris base, 1 M glacial acetic acid, 50 mM EDTA

Luria Broth (LB) media: each 25 g of LB Broth granules (Melford) was dissolved in 1 L Milli-Q® water and sterilized by autoclaving for 20 min at 121°C

LB agar: each 32 g of LB Agar granules, Low Salt Formula (Melford) was dissolved in 1 L Milli-Q® water and sterilized by autoclaving for 20 min at 121°C

2.1.2 Molecular biology kits

PCR product purification: Monarch® PCR DNA Cleanup Kit (NEB)

Gel extraction: QIAquick® Gel Extraction Kit (QIAGEN)

Plasmid DNA extraction and purification (small scale): Monarch® Plasmid Miniprep Kit (NEB)

Plasmid DNA extraction and purification (medium scale): QIAGEN® Plasmid Plus Midi Kit (QIAGEN)

2.1.3 Plasmids

Plasmid	Bacterial resistance	Mammalian selection marker	Source
pRetroX-Tet3G	Ampicillin	Neomycin	Clontech
pRetroX-TRE3G	Ampicillin	Puromycin	Clontech
pRetroX-TRE3G-Luc Control	Ampicillin	Puromycin	Clontech
pVSV-G	Ampicillin	-	Clontech
lentiCas9-Blast	Ampicillin	Blasticidin	Addgene (#52962)
pMD2.G	Ampicillin	-	Addgene (#12259)
pRSV-Rev	Ampicillin	-	Addgene (#12253)
pMDLg/pRRE	Ampicillin	-	Addgene (#12251)
pSIN4-EF1a-TAL1-IRES-Puro	Ampicillin	Puromycin	Addgene (#61065)
lentiGuide-Hygro-dTomato	Ampicillin	Hygromycin	Addgene (#99376)
lentiCas9-Blast	Ampicillin	Blasticidin	Addgene (#52962)
pSpCas9(BB)-2A-GFP (PX458)	Ampicillin	-	Addgene (#48138)
pAcGFP1-C1	Ampicillin	Neomycin	Clontech
Lenti-(BB)-EF1a-KRAB-dCas9-P2A-BlastR	Ampicillin	Blasticidin	Addgene (#118154)

Table 2.1: Plasmids used in this doctoral work.

2.1.4 Polymerase Chain Reaction (PCR)

Each PCR reaction was done in a 50 µl volume:

Component	50 µl reaction
H ₂ O	to 50 µl
10x Pfu Ultra II reaction buffer	5 µl
DMSO (Sigma)	2.5 µl
100 µM forward primer (IDT)	0.25 µl
100 µM reverse primer (IDT)	0.25 µl
10 mM dNTP mix (Roche)	5 µl
PfuUltra II Fusion High-Fidelity Polymerase (Agilent)	0.5 µl
DNA template	1 µl

Table 2.2: The composition of a 50 µl PCR reaction.

A ProFlex PCR thermocycler (Applied Biosystems) was employed to run the reaction:

CYCLE STEP	TEMPERATURE	TIME	CYCLES
Initial denaturation	95°C	5 min	1
Denaturation	95°C	45 s	30
Annealing	55 – 59°C (depending on the primer)	45 s	
Extension	72°C	25 s/kb	
Final extension	72°C	15 min	1

Table 2.3: The parameters for a PCR reaction run.

2.1.5 Phenol:chloroform extraction and ethanol precipitation

DNA preparations in aqueous solutions containing unwanted ions or proteins were isolated using the phenol:chloroform extraction and purified using ethanol precipitation. Briefly, nuclease-free H₂O (Sigma-Aldrich) was added to the sample to bring the volume up to 350 µl. Then, an equal volume of phenol:chloroform:isoamyl alcohol, 25:24:1, pH 8.0 (Sigma) was added and the sample was mixed by vigorous shaking for 15 seconds before a centrifugation at 12,000 x g for 10 minutes. The aqueous fraction was removed into a fresh tube, and 0.1 volumes of 3 M sodium acetate pH 5.0, 1 µl GlycoBlue and 2.5 volumes of cold 100% ethanol were added. The sample was mixed by inverting, then incubated at -80°C for >60 minutes to precipitate the DNA. Afterwards, the sample was spun at 20,000 x g and 4°C for 15 minutes, the supernatant was removed, and the pellet washed with 70% and 80% cold ethanol in two subsequent steps, each wash followed by a centrifugation for 5 minutes at 20,000 x g and 4°C. After the final spin and removal of the 80% ethanol, the DNA pellet was allowed to air-dry briefly before resuspension in H₂O or a desired buffer.

2.1.6 Agarose electrophoresis

UltraPure™ Agarose (Thermo Fisher) and 1x TAE buffer were used to prepare a 1% agarose gel, which was supplemented with SYBR™ Safe DNA Gel Stain (Thermo Fisher) to a 1:10,000 dilution. Once the gel solidified, it was placed in a gel tank, which, in turn, was filled with 1x TAE. Samples were loaded into the wells along with an appropriate standard (Quick-Load® 1 kb DNA Ladder or Quick-Load® Purple 100 bp DNA Ladder, both NEB) and the gel was run at 80 – 120 V for 0.5 – 1.5 hours. Gels were visualized in blue light and photographed using a G:BOX Chemi-XRQ (Syngene).

2.1.7 Transformation of bacterial cells

Competent *E. coli* cells of the Stbl3 strain (genotype: $F^-mcrB\ mrrhsdS20(r_B^-, m_B^-)\ recA13\ supE44\ ara-14\ galK2\ lacY1\ proA2\ rpsL20(Str^R)\ xyl-5\ \lambda^-\ leumtl-1$; Thermo Fisher) were thawed on ice for 10 minutes. Then, 1 pg – 100 ng (1 – 5 µl) of plasmid DNA was added to 50 µl of the Stbl3 cells in an Eppendorf tube. After a gentle mix, the suspension was placed on ice for 30 minutes, then heat-shocked at 42°C for 40 seconds before being returned to ice for 5 minutes. 450 µl of room-temperature SOC media (Melford) was added to the cell suspension, and cells were incubated for 1 hour at 37°C with vigorous

shaking (500 RPM in a table-top shaking incubator) before being spun down at 3000 RPM for 5 minutes. 400 µl of the media was discarded and the cells were resuspended in the leftover medium and plated out on LB agar plates containing 100 µg/ml ampicillin. These plates were incubated overnight (12 – 16 hours) at 37°C and checked for colonies in the morning.

2.1.8 Plasmid preparation and sequencing

Selected colonies from the overnight plates were each transferred into a 50 ml Falcon tube with 5 ml LB media containing 100 µg/ml ampicillin. These suspensions were incubated for 12 – 16 hours at 37°C with vigorous shaking (220 RPM on a shaking platform). The plasmid was isolated and purified from the cells using the Monarch® Plasmid Miniprep Kit (NEB), and sequenced by Eurofins Genomics.

To generate larger amounts of the desired plasmids, Stbl3 cells were transformed and plated as described above, and a 5 ml starter culture was prepared from a selected colony. 100 – 500 µl of the starter culture was used to inoculate 100 ml of LB media containing 100 µg/ml ampicillin in a 250 ml conical flask, which was then incubated for 16 hours at 37°C and 220 RPM on a shaking platform. The plasmid was isolated and purified using the QIAGEN® Plasmid Plus Midi Kit (QIAGEN).

2.1.8.1 Preparation of pRetroX-TRE3G-FLAG-TAL1

This plasmid was prepared by Dr. Dan Bose. FLAG-TAL1 was first amplified from an existing plasmid (Addgene plasmid #61065; (Elcheva et al., 2014)) using PCR with the following primers:

Primer name	Sequence
TRE-FLAG-TAL1_RC_FW	catcatggatccccATGGACTACAAGGACGACGATGAC
TRE-FLAG-TAL1_RC_REV	catcatgaattcgatccTCACCGAGGGC

Table 2.4: Primers for pRetroX-TRE3G-FLAG-TAL1 cloning.

The PCR product was purified using the Monarch® PCR DNA Cleanup Kit (NEB). The purified PCR product and the pRetroX-TRE3G plasmid (Clontech) were digested using EcoRI-HF and BamHI-HF (10

– 17 U/1 µg DNA; both NEB) and run on an agarose gel. Appropriate bands were excised and purified from the gel using the blue light Dark Reader transilluminator (DR-45M, Clare Chemical Research) and the QIAquick® Gel Extraction Kit (QIAGEN), and the two components were then mixed and ligated using T4 DNA ligase (400 U/ca. 100 ng DNA; NEB). 5 µl of the resulting ligation mixture was used to transform 50 µl of competent Stbl3 *E. coli* cells, as described above.

2.1.8.2 Preparation of the plasmids for CRISPR validation of the genetically modified cell line

The plasmids used for different CRISPR/Cas9 editing experiments were all based on lentiGuide-Hygro-dTomato (Addgene plasmid #99376, (Ho et al., 2017)) and cloned using a slightly modified protocol from the Zhang lab (Sanjana et al., 2014; Shalem et al., 2014). Briefly, two oligos were designed and ordered (IDT) for each sgRNA to be cloned, in the following form:

Oligo 1 (up): 5' – CACCGNNNNNNNNNNNNNNNNNNNN – 3'

Oligo 2 (down): 3' – CNNNNNNNNNNNNNNNNNNNNCAAA – 5',

where N₂₀ is the unique sequence of the sgRNA. The overhangs are compatible with BsmBI digestion of the lentiGuide-Hygro-dTomato plasmid.

The following oligos (standard desalted, lyophilized) were utilized:

sgRNA name/sequence	Target	Oligo name	Oligo sequence	Source/Design tool
<u>MYB1</u> CACAGAAAGAC GGTTAGGAAA	TAL1 enhancer – MYB binding site	sg_MYBbs-up	CACCGCACAGAAAG ACGGTTAGGAAA	(Mansour et al., 2014)
		sg_MYBbs-down	AAACTTTCCTAACC GTCTTCTGTGC	
<u>TAL1</u> GAGTGGAGATC CTATTCAGA	TAL1 5'-UTR – endogenous TAL1	sg_TAL1_ex1.1-up	CACCGGAGTGGAG ATCCTATTCAGA	CRISPick (Doench et al., 2016; Sanson et al., 2018)
		sg_TAL1_ex1.1-down	AAACTCTGAATAGG ATCTCCACTCC	

sgRNA name/sequence (cont.)	Target (cont.)	Oligo name (cont.)	Oligo sequence (cont.)	Source/Design tool (cont.)
<u>TAL4</u> GAATAGGATCTC CACTCCGC	TAL1 5'-UTR – endogenous TAL1	sg_TAL1_ex1.2-up	CACCGGAATAGGAT CTCCACTCCGC	CRISPick (Doench et al., 2016; Sanson et al., 2018)
		sg_TAL1_ex1.2-down	AAACGCGGAGTGG AGATCCTATTCC	
<u>TAL7</u> GCGGCCCTTTAA GTCTCTCG	TAL1 exon (2/4) – endogenous TAL1 + FLAG-TAL1	sg_TAL1_ex2.1-up	CACCGCGGCCCTT TAAGTCTCTCG	CRISPick (Doench et al., 2016; Sanson et al., 2018)
		sg_TAL1_ex2.1-down	AAACCGAGAGACTT AAAGGGCCGCC	
<u>TAL10</u> TGAGGCTGTAG AGCAGCGCG	TAL1 exon (2/4) – endogenous TAL1 + FLAG-TAL1	sg_TAL1_ex2.2-up	CACCGTGAGGCTGT AGAGCAGCGCG	CRISPick (Doench et al., 2016; Sanson et al., 2018)
		sg_TAL1_ex2.2-down	AAACCGCGCTGCTC TACAGCCTCAC	
<u>SCR1</u> GCTGATCTATCG CGGTCGTC	non-targeting – scrambled negative control	sg_Scramble1-up	CACCGGCTGATCTA TCGCGGTCGTC	(Lawhorn et al., 2014)
		sg_Scramble1-down	AAACGACGACCGC GATAGATCAGCC	

Table 2.5: Primers for the cloning of lentiGuide-Hygro-dTomato plasmids for CRISPR/Cas9 validation of the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line.

First, the lentiGuide-Hygro-dTomato plasmid (5 µg) was digested with BsmBI (25 U; NEB), followed by heat inactivation, phenol:chloroform extraction and ethanol precipitation. The purified linearized plasmid was then dephosphorylated by an incubation with Shrimp Alkaline Phosphatase (rSAP; 2 U; NEB) and gel purified. Meanwhile, each pair of oligos was phosphorylated and annealed, at a 10 µM concentration, in a 10 µl reaction with T4 Polynucleotide Kinase (PNK; 5U; NEB) in 1x T4 Ligation Buffer (NEB), in a thermocycler with the temperature first set to 37°C for 30 minutes, then increased up to 95°C for 5 minutes and afterwards ramped down to 25°C at 5°C/min. The annealed oligos were then ligated with the BsmBI-digested plasmid in an overnight 16°C reaction with the T4 DNA Ligase (400 U/50 ng DNA; NEB), and transformed into Stbl3 *E. coli* cells the next day.

2.1.8.3 Preparation of the plasmids for the CRISPRi validation of the genetically modified cell line

The CRISPR/dCas9-KRAB plasmids were based on the Lenti-(BB)-EF1a-KRAB-dCas9-P2A-BlastR backbone (Addgene plasmid #118154, unpublished), which uses the same cloning mechanism as lentiGuide-Hygro-dTomato. These plasmids were therefore prepared in the same way as the CRISPR/Cas9 plasmids, using the following oligos:

sgRNA name/sequence	Target	Oligo name	Oligo sequence	Source/Design tool
<u>MYB1</u> CACAGAAAGAC GGTTAGGAAA	TAL1 enhancer – MYB binding site	sg_MYBbs-up	CACCGCACAGAAA GACGGTTAGGAAA	(Mansour et al., 2014)
		sg_MYBbs-down	AAACTTTCCTAACC GTCTTTCTGTGC	
<u>PRO1</u> TGAGTGGGATT ACAGCGCGT	TAL1 promoter	sg_PRO1_up	CACCGTGAGTGGG ATTACAGCGCGT	CRISPick (Doench et al., 2016; Sanson et al., 2018)
		sg_PRO1_down	AAACACGCGCTGTA ATCCCACTCAC	
<u>PRO2</u> GATTACAGCGC GTCGGTGGA	TAL1 promoter	sg_PRO2_up	CACCGGATTACAGC GCGTCGGTGGA	CRISPick (Doench et al., 2016; Sanson et al., 2018)
		sg_PRO2_down	AAACTCCACCGACG CGCTGTAATCC	
<u>PRO3</u> AGGCCTCTCAGC GAAAAAGG	TAL1 promoter	sg_PRO3_up	CACCGAGGCCTCTC AGCGAAAAAGG	CRISPick (Doench et al., 2016; Sanson et al., 2018)
		sg_PRO3_down	AAACCCTTTTTCGCT GAGAGGCCTC	
<u>PRO4</u> CCACACCGCAGC GTAAGTGC	TAL1 promoter	sg_PRO4_up	CACCGCCACACCGC AGCGTAACTGC	CRISPick (Doench et al., 2016; Sanson et al., 2018)
		sg_PRO4_down	AAACGCAGTTACGC TGCGGTGTGGC	
<u>SCR1</u> GCTGATCTATCG CGGTCGTC	non-targeting – scrambled negative control	sg_Scramble1-up	CACCGGCTGATCTA TCGCGGTCGTC	(Lawhorn et al., 2014)
		sg_Scramble1-down	AAACGACGACCGC GATAGATCAGCC	

Table 2.6: Primers for the cloning of Lenti-(BB)-EF1a-KRAB-dCas9-P2A-BlastR plasmids for CRISPRi validation of the Jurkat-Tet3G-TRE3G-FLAG-TAL1 cell line.

2.2 Cell culture

2.2.1 Cell lines

Cell line	Description	Source
Jurkat	Human T-cell acute lymphoblastic leukaemia	Dr. Helen Bryant, The University of Sheffield
Jurkat-Tet3G	Jurkat cells stably expressing the tetracycline-dependent transcription activator (Tet-transactivator) protein	In house
Jurkat-Tet3G-TRE3G-FLAG-TAL1	Jurkat cells containing stably expressed Tet-transactivator protein and doxycycline-inducible FLAG-TAL1	In house
Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9	Jurkat cells containing stably expressed Tet-transactivator protein, doxycycline-inducible FLAG-TAL1, and stably expressed Cas9	In house
HEK293T	Human embryonic kidney cell line stably expressing the SV40 large T antigen	Prof. Stuart Wilson, The University of Sheffield
GP2-293	Viral packaging cell line derived from HEK293	Clontech
293FT	HEK293-based cell line modified to produce high viral titres (fast-growing and stably expressing the SV40 large T antigen)	Thermo Fisher

Table 2.7: Mammalian cell lines used in this doctoral work.

2.2.2 Growth media and other tissue culture solutions

Jurkat and its derivatives (standard cell culture, with the exceptions below): Roswell Park Memorial Institute (RPMI) 1640 Medium (Thermo Fisher) supplemented with 10% tetracycline-free (Tet-free) Fetal Bovine Serum (FBS), certified, heat inactivated, United States (Thermo Fisher) and 1% Penicillin Streptomycin (P/S) Solution (Thermo Fisher)

Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9, last experiment in Chapter 4.12 and all experiments in Chapter 4.13: RPMI 1640 supplemented with 10% Tet-free FBS, 1% P/S, and 5 µg/ml blasticidin

Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9, all experiments in Chapter 5: RPMI 1640 supplemented with 10% Tet-free FBS, 1% P/S, 1 mg/ml G418, 0.25 µg/ml puromycin, and 5 µg/ml blasticidin

HEK293T and GP-293: Cytiva HyClone™ Dulbecco's High Glucose Modified Eagles Medium (DMEM, Thermo Fisher) supplemented with 10% Gibco™ FBS, qualified, heat inactivated, E.U.-approved, South America Origin (Thermo Fisher) and 1% P/S for regular passaging and DMEM with 10% Tet-Free FBS when using for virus production for experiments in Jurkat cells

293FT: DMEM supplemented with 10% FBS, 1% P/S and 500 µg/ml G418 for regular passaging and DMEM with 10% Tet-Free FBS when using for virus production for experiments in Jurkat cells

Dulbecco's Phosphate Buffered Saline (DPBS), Ca²⁺- and Mg²⁺-free (Sigma-Aldrich)

Trypsin 0.25% dissociation reagent (Thermo Fisher)

2.2.3 Selection antibiotics

Neomycin: G418 Sulphate, BioVision

Puromycin: Puromycin Dihydrochloride, Sigma-Aldrich

Blasticidin: Blasticidin S Hydrochloride, Melford

Hygromycin: Hygromycin B, Thermo Fisher

2.2.4 Routine culturing procedures

All cells were grown at 37°C and in 5% CO₂.

All cells were counted using an Improved Neubauer Haemocytometer (Hawksley).

Cryopreservation medium for all cell lines consisted of their respective full medium supplemented with 10% DMSO (V/V) and extra FBS to a final concentration of 20% (V/V). 1 ml aliquots containing 2 – 4 x 10⁶ cells/ml in cryopreservation medium were frozen at -80°C using Mr. Frosty™ Freezing Container (Thermo Scientific) and transferred into -196°C the next day.

1 ml aliquots of all cells were stored in liquid nitrogen at -196°C until needed, then thawed rapidly in a water bath at 37°C.

2.2.4.1 Sub-culturing of Jurkat cells and their derivative cell lines

Upon defrosting, Jurkat cells and their derivative cell lines were transferred into a Falcon tube with 9 ml appropriate full media, then centrifuged at 200 x g for 7 minutes. The supernatant was removed and the cell pellet resuspended in 10 – 13 ml full media, and the cell suspension was transferred into a new T-75 flask. Cells were passaged every 3 – 4 days outside of experiments, with cell density never below 1 x 10⁵ cells/ml or over 3 x 10⁶ cells/ml. For routine passaging, cells were split in a 1:8 to 1:13 ratio by mixing the desired part of the grown cell suspension with appropriate fresh media. For passaging before experiments, cells were spun down at 200 x g for 7 minutes, old medium was aspirated, pellet loosened by tapping and resuspended in the required media volume. Cells were then split according to the specific experimental needs.

2.2.4.2 Sub-culturing of HEK293T cells and HEK293-derivative cell lines

The thawed suspensions of HEK293T cells and HEK293-derivative cell lines were transferred directly onto a fresh 10 cm dish with 9 ml of appropriate media. The dish was cultivated for 16 – 24 hours to allow the cells to adhere, then the media was changed to remove the DMSO traces from the cryopreservation medium. HEK293T cells and related cell lines were sub-cultured at ratios from 1:6 to 1:12, passaged generally every 3 – 4 days. The media from the dish was first aspirated, then the dish was washed with DPBS, 0.25% Trypsin was added and the dish was tilted from side to side to make sure the solution reached the whole surface. Then, the Trypsin solution was aspirated, and the dish was incubated at 37°C for 3 – 5 minutes. After the incubation, the dish was tapped on the sides to lift the cells off the surface completely, full medium was added, cells were resuspended in the medium and split according to the desired ratio.

2.2.5 Culturing procedures during antibiotic selection

An antibiotic pressure was often employed to select for cells with a stable integration of the gene of interest, which in suitable plasmids is linked to antibiotic resistance. (Although in some plasmids, the gene of interest and antibiotic resistance marker are under control of a single promoter and the expression of the one is therefore indelibly linked to the expression of the other, this is not the case in any of our plasmids. In all our plasmids containing antibiotic resistance cassettes, the link between the gene of interest and the antibiotic resistance cassette is by proximity only. This is the most common practice and the connection is usually sufficiently strong between the expression of the two genes.)

Antibiotics were first added 24 hours after transduction, except for the first cell line establishment in Chapter 3.2 and some of the tested conditions in the hygromycin influence examination in Chapter 4.10, where a 48-hour interval was employed.

Cells grown in antibiotic selection media were first passaged as and when needed, before the examination of proliferation and survival patterns in Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 (Chapter 4.11) showed the need for more frequent passaging.

- In the antibiotic selection during the cell line establishment (Chapter 3), cells were passaged every 2 – 3 days with a complete change of media.

- In all the validation experiments (Chapter 4 and Chapter 5), a control group sample was always chosen (usually, non-targeting sgRNA, plasmid without sgRNA, non-targeting siRNA) with the highest expected proliferation rate, and the same volume adjustments that were made to the control sample were made to all the other samples in the experimental set. In practice, this meant that while retaining the same volume, control samples and experimental samples were most often adjusted to different cell densities.
 - In the first validation experiment with unconcentrated virus (Chapter 4.4), first experiments with concentrated virus (Chapter 4.7), first Neon electroporation experiments in Jurkat and Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 (Chapter 4.9) and the experiment examining the influence of different hygromycin treatments (Chapter 4.10), cells were left growing without passaging up to 8 days at a time. Passaging was dependent on detecting slight changes in media colour, based on the phenol red indicator in RPMI media, or on detecting expansion by examination of cell density under the microscope, or, in some cases, on counting.
 - In the experiment concerned with the influence of maintaining an optimal cell density in Chapter 4.11, media was changed at least every 4 days, and the cell density adjusted every 2 days to between 3.5×10^5 and 5.5×10^5 cells/ml in the control group.
 - In the blasticidin influence experiment (Chapter 4.12), cells in the control group were maintained at a density between 1×10^5 and 1×10^6 cells/ml, with a change of media at least every 4 days.
 - In the CRISPR/Cas9 validation experiment comparing the performance of the different Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 clones and the Jurkat cell line (Chapter 4.13), cell density in the control group was adjusted to between 1×10^5 and 2×10^5 cells/ml every 2 days, and medium was changed at least every 4 days.
 - In the doxycycline rescue validation experiments (Chapter 5), cells were passaged every 2 – 3 days, medium was changed at least every 4 days and the cell density in the control group was adjusted to between 1×10^5 and 2×10^5 cells/ml every 2 – 3 days.

2.2.6 Culturing procedures following FACS

In a number of experiments, FACS was used to select cells of desirable qualities (Chapter 2.5).

In Chapter 3, FACS was used to single-cell sort cells of an expanded polyclonal cell line, selected for by antibiotic pressure, into 96-well plates. These plates were regularly checked for expanding clones under microscope, and when a slight change in colour of the media indicated that the clone was steadily expanding, the contents of the well were transferred into a bigger volume and a bigger receptacle. In this manner, viable clones travelled from 96-well plates to 24-well plates to 12-well plates, and potentially 6-well plates, T-25 or T-75 flasks, depending on how many cells were required for testing. Selected clones were then expanded according to the regular culturing procedures described above.

In some experiments in Chapter 4 and 5, FACS was used for the enrichment of transduced cells. In the CRISPR/Cas9 validation experiment with unconcentrated virus (Chapter 4.5), cells were monitored for signs of expansion under the microscope, as well as counted, and transferred into larger volumes and bigger receptacles as they expanded. In the FACS-sorted proof-of-principle and doxycycline rescue experiments in Chapter 5.5, cells were spun down every 3 days at 200 x g for 10 minutes, supernatant removed, and fresh media added to reach a cell density of 1×10^5 cells/ml in the control (SCR1 – 0 ng/ml DOX) group, regardless of expansion. The same volume adjustments that were made to the control sample were made to all the other samples in the experimental set.

2.3 Viral transduction

2.3.1 Transfection of packaging cells

Viral particles for lentiviral or retroviral transduction were prepared by transfection of the packaging GP2-293 cell line (pRetroX-Tet3G, pRetroX-TRE3G-FLAG-TAL1 and pRetroX-TRE3G-Luc Control) or HEK293T cell line (lentiCas9-Blast, lentiGuide-Hygro-dTomato, lentiCas9-Blast and Lenti-(BB)-EF1a-KRAB-dCas9-P2A-BlastR). One day before transfection, the packaging cells were plated at $8 - 9.6 \times 10^4$ cells/cm² of surface area in antibiotic-free, Tet-free media in 6-well plates (2 ml), T-25 flasks (3 – 4 ml) or T-75 flasks (9 – 12 ml).

The transfection mix was then prepared using plasmid DNA and the Lipofectamine 3000 reagent kit (Thermo Fisher) according to manufacturer's instructions. Briefly, Lipofectamine 3000 was mixed with Opti-MEM medium (Thermo Fisher) in one tube; required plasmids and the P3000 Reagent were mixed with Opti-MEM medium in another tube; for example:

		MoMuLV-based retroviral system	HIV-1-based 3 rd generation lentiviral system		
		GP2-293 cells (1 well on a 6-well plate)	HEK293T cells (1 well on a 6-well plate)	HEK293T cells (T-25)	HEK293T cells (T-75)
TUBE 1	Opti-MEM	125 µL	125 µL	250 µL	750 µL
	Lipofectamine 3000	7.5 µl	7.5 µl	15 µl	45 µl
TUBE 2	Opti-MEM	125 µL	125 µL	250 µL	750 µL
	Transfer plasmid	2 µg	2 µg	4 µg	12 µg
	Envelope plasmid	2 µg pVSV-G	0.5 µg pMD2.G	1 µg pMD2.G	3 µg pMD2.G
	Packaging plasmid pRSV-Rev	-	0.75 µg	1.5 µg	4.5 µg
	Packaging plasmid pMDLg/pRRE	-	0.75 µg	1.5 µg	4.5 µg
	P3000 Reagent	8 µL	8 µL	16 µL	48 µL

Table 2.8: Preparation of transfection mixtures.

Afterwards, the contents of the two tubes were combined, gently mixed and incubated at room temperature for 15 – 20 minutes. Resulting complexes were added into respective wells. Medium was changed 15 – 17 hours post transfection, replaced by antibiotic-free, Tet-free media.

2.3.2 Transduction with unconcentrated virus

24 hours after the medium change, the viral supernatant was collected and filtered through a 0.45 μm filter to remove the cell debris, but not the viral particles. For cell line establishment (Chapter 3), fresh antibiotic-free, Tet-free medium was added to the packaging cells after the removal of the viral supernatant. In all the transduction experiments with unconcentrated virus, polybrene (Merck) was added to the filtrate so that the final concentration in the resulting transduction suspension would be 8 $\mu\text{g}/\text{ml}$. Jurkat cells or the appropriate derivative cell line were pelleted at 200 x g for 7 minutes and resuspended in fresh medium. 8×10^5 cells in a total of 2 ml RPMI medium were mixed with the virus-containing media and transferred into a well on a 6-well plate. For cell line establishment (Chapter 3), the transduced Jurkat cells were pelleted and transduced for a second time 24 hours later, using a newly prepared, 0.45 μm filtered viral medium from the packaging cells supplemented with 2 ml fresh RPMI medium and polybrene at the same final concentration as before. In all the transduction experiments with unconcentrated virus, 24 hours after the final transduction, viral medium was replaced by selection medium (depending on selection marker, containing either 2000 $\mu\text{g}/\text{ml}$ G418, or 0.5 $\mu\text{g}/\text{ml}$ puromycin, or 10 $\mu\text{g}/\text{ml}$ blasticidin, or 500 $\mu\text{g}/\text{ml}$ hygromycin).

2.3.3 Transduction with concentrated virus

24 hours after the medium change, the media containing virus was collected and spun down at 500 x g for 10 minutes. The viral supernatant was carefully removed into a fresh tube, mixed with Lenti-X Concentrator (Takara) in a 3:1 ratio by gentle inversion, and incubated at 4°C for >30 minutes (usually overnight). After that, these samples were centrifuged at 1,500 x g for >45 minutes at 4°C. Supernatants were carefully removed and the off-white pellet containing viral particles was resuspended in DPBS to a concentration of 0.25 units/ μl (1 unit, 1 U, was equal to the amount of virus that 2.33 cm^2 packaging cells, grown as detailed above, produced in 24 hours). If not used immediately, these viral stocks were aliquoted and stored at -80°C.

To transduce the Jurkat-Tet3G-TRE3G-FLAG-TAL1 or Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line, 1.6×10^5 cells per well were seeded into 24-well plates in 0.8 ml Tet-free, antibiotic-free media with the final concentration of 8 $\mu\text{g/ml}$ polybrene. Concentrated virus was added to each well as required, and the suspension was mixed well with a P1000 pipette tip. 1 – 17 U virus was used for optimization, 9 U for the first experiments with concentrated virus (Chapter 4.7) and 10 U for all the other experiments since then (Chapter 4.12 and 4.13; Chapter 5). Where antibiotic selection was employed, 24 hours after transduction, viral medium was replaced by selection medium (depending on selection marker, this was 5 – 10 $\mu\text{g/ml}$ blasticidin and/or 250 – 500 $\mu\text{g/ml}$ hygromycin, as detailed for each experiment in the Results Chapters). Where FACS was to be performed, the viral medium was changed at 24 hours post transduction for full, Tet-free medium, apart from some of the samples in the second transduction optimization experiment, where the viral medium was left on until the day of FACS, i.e. until day 2 or day 3, as required per experimental setup (Chapter 5.7), and the FACS-sorted experiment with 293FT-generated virus in Chapter 5.5, where the viral medium was left unchanged until day 3.

2.4 Neon electroporation

Cells were passaged a day before transfection so that they would be in log-phase on the following day. Transfections using the Neon Transfection System (Thermo Fisher) were carried out according to manufacturer's instructions.

On the day of transfection, for each condition, 5 ml of antibiotic-free, Tet-free media was prewarmed in a T-25 flask. Required plasmids, ASOs or siRNAs for each condition were placed in sterile 1.5 ml Eppendorf tubes. In the proof-of-principle Neon experiments, 5 or 15 µg of pAcGFP1-C1 or pSpCas9(BB)-2A-GFP (PX458) were used, or 50 or 100 nM Cy3-ASO (control; Table 2.9 in Chapter 2.4.1). In the first experiment with Jurkat-Tet3G-TRE3G-FLAG-TAL1 (Chapter 4.9), 15 µg of lentiGuide-Hygro-dTomato plasmids containing sgRNAs was used. In all the following Jurkat-Tet3G-TRE3G-FLAG-TAL1 experiments, 7.5 µg of the same plasmids was used. In all the experiments with Jurkat cells, 7.5 µg sgRNA-containing lentiGuide-Hygro-dTomato plasmids and 7.5 µg lentiCas9-Blast was used for each sample. In siRNA knockdown experiments, 50 nM siRNA was used (Table 2.9 in Chapter 2.4.1). 3 ml Electrolytic Buffer E2 was transferred into a Neon Tube, and placed in the Neon Pipette Station.

Cells were counted, spun down at 200 x g for 7 minutes, washed with Ca²⁺- and Mg²⁺-free DPBS and centrifuged again. The cell pellet was resuspended in Resuspension Buffer R to a cell density of 2 x 10⁷ cells/ml. 110 µl of thus prepared cells was added to each of the tubes containing plasmids, and the contents of the tube were mixed gently. 100 µl of each of the cell suspensions was then carefully aspirated into a Neon Tip with a Neon Pipette, avoiding air bubbles, and placed into the Neon Pipette Station so that the Neon Tip was submerged in the Electrolytic Buffer. Electroporation was then carried out at 1,350 V, 3 pulses of 10 ms. Cells were immediately transferred into the prewarmed media in T-25 flasks, and flasks were transferred into the cell incubator. Antibiotic selection was added 24 hours post transfection, except in some of the tested conditions in the hygromycin influence examination in Chapter 4.10, where a 48 hour interval was employed.

2.4.1 siRNAs & ASO

siRNA name	Forward	Reverse
TAL1 A	CCUAUGAGAUGGAGAUUACUGAUGG	CCAUCAGUAAUCUCCAUCUCAUAGG
TAL1 B	ACACCAAAGUUGUGCGGCGUAUCUU	AAGAUACGCCGCACAACUUUGGUGU
TAL1 C	GCCUGGCCAUGAAGUAUAUCAACUU	AAGUUGAUUAACUUCAUGGCCAGGC
control 0	#462000, high GC Duplex	from the Stealth RNAi™ siRNA Negative Control Kit (catalogue number 12935100, Thermo Fisher)
control 1	#462001, medium GC Duplex	
control 2	#462002, low GC Duplex	
Cy3-ASO (control)	TCACCTTCACCCTCTCC	GGAGAGGGTGAAGGTGA

Table 2.9: siRNAs and ASO used for transfection by electroporation with the Neon device.

2.5 Flow cytometry and fluorescence-activated cell sorting (FACS)

2.5.1 Glossary and technical parameters

FSC – forward scatter

SSC – side scatter

(FSC/SSC)-A – area (under the voltage curve of the detected pulse)

(FSC/SSC)-H – height (of the voltage curve of the detected pulse)

(FSC/SSC)-W – width (of the voltage curve of the detected pulse)

dTomato = Cy3 (BD FACSMelody) – 561 nm yellow-green laser, 605LP mirror, 613/18 filter

dTomato (LSRII) – 488 nm blue laser, 550LP mirror, 575/26 filter

GFP = FITC (BD FACSMelody) – 488 nm blue laser, 507LP mirror, 527/32 filter

2.5.2 Preparations

Prior to flow cytometry measurements or FACS-based cell sorting, cells were always washed at least once in the appropriate fresh media, twice where the medium wasn't changed after viral transduction. Before the final resuspension in Tet-free media, the pellet was loosened by tapping, to break up as many clumps as possible. All flow cytometry and FACS experiments were carried out at The Medical School Flow Cytometry Core Facility (The University of Sheffield), largely with technical support. Cells were transferred to and from the Core Facility on ice, in tubes or plates. Where plates were used, sides were sealed with parafilm for the journey to minimize the potential for contamination and contain potential spillage.

2.5.3 Fluorescence-activated cell sorting (FACS)

FACS was used for single-cell sorting in monoclonal cell line establishment (Chapter 3) and for the sorting of transduced populations (Chapters 4 and 5). For the cell line establishment in Chapter 3, cells from polyclonal cell lines selected by antibiotic pressure were single-cell sorted using the BD FACSAria

flu cell sorter. Cells were gated for viability (FSC-A vs SSC-A) and for singlets (FSC-A vs FSC-H), and the resulting population of live singlets was sorted into 96-well plates containing 100 μ l of fresh or pre-conditioned Tet-free media (media in which a healthy culture of Jurkat cells has been allowed to proliferate for 24 – 48 hours, spun down for 5 min at 500 x g and filtered through a 0.45 μ m filter to remove any residual cells) without selection antibiotics.

FACS for the sorting of sgRNA-positive populations according to dTomato fluorescence from the transduced lentiGuide-Hygro-dTomato plasmid (Chapters 4 and 5) was carried out using the BD FACSMelody cell sorter. Cells were gated for a live population (FSC-A vs SSC-A), then live singlets (FSC-H vs FSC-W and/or SSC-H vs SSC-W), and then dTomato-positive live singlets (dTomato-A vs SSC-A). Autofluorescence in the MYB1 vs SCR1 experiment in Chapter 5.7 was monitored in the FITC channel (GFP-A vs SSC-A). In the CRISPR/Cas9 validation experiment with unconcentrated virus in Chapter 4.5, 1×10^4 cells were collected into Tet-free Jurkat media on a 96-well plate. In the FACS-based sorting in Chapter 5.5, 1×10^5 cells were collected into 0.5 ml fresh Tet-free Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 maintenance media on a 24-well plate.

2.5.4 Flow cytometry

Flow cytometry was used to assess the percentage of transduced or transfected cells in proof-of-principle and optimization experiments. 10,000 live singlet events were collected for each measured sample. BD FACSMelody was used to measure the virus amount transduction optimizations in Chapter 4.6 and the Neon transfection proof-of-principle experiments in Chapters 4.8 and 4.9. Live cells were gated from the FSC-A vs SSC-A plot, singlets from the FSC-H vs FSC-W and/or SSC-H vs SSC-W plots. In the proof of principle experiment, GFP-positive cells were identified in a GFP-A vs SSC-A plot and Cy3-positive cells in a Cy3-A vs SSC-A plot. In the optimization experiment, dTomato-positive cells were visualized in a dTomato-A vs SSC-A plot. LSRII was used for the measurements in the transduction optimization in Chapter 5.7. Live cells were identified on a FSC-A vs SSC-A plot, singlets on a FSC-A vs FSC-H plot, and dTomato-positive cells on a dTomato-A vs SSC-A plot (different settings to BD FACSMelody, see technical parameters above).

2.6 Cell assays

2.6.1 Cell proliferation with cytotoxic drugs

To establish the minimum lethal dose of antibiotics for Jurkat and its derivative cell lines, equal numbers of cells per well were incubated either in a 6-well or a 12-well plate with increasing concentrations of each drug. Tested concentrations ranged from 500 to 4000 µg/ml for G418, from 0.125 to 2 µg/ml for puromycin, from 5 to 30 µg/ml for blasticidin, and from 75 to 4000 µg/ml for hygromycin. Viability of cell culture was assessed at different time points depending on the drug tested, using haemocytometer and Trypan Blue stain.

2.6.2 Luciferase assay

GP2-293 cells, plated the day before at 6×10^5 per well in a 6-well plate, were transfected as described in Chapter 2.3.1, using pRetroX-TRE3G-Luc vector (Takara). Transduction medium was prepared by mixing filtered viral medium from GP2-293 cells 48 hours post transfection with fresh RPMI at a 3:7 ratio, and polybrene was added at a final concentration of 8 µg/ml. 200 µl of this mixture was transferred into each well of a 96-well plate with 2×10^4 cells of each tested clonal cell line. 1000 ng/ml doxycycline was added 24 and 72 hours post transduction. Luciferase assay was performed 96 hours post transduction using Pierce Firefly Luciferase Flash Assay Kit (Thermo Fisher) according to the manual, using white 96-well plates (SARSTEDT) and plate reader VICTOR 3 (Perkin Elmer) programmed for luciferase assay measurements.

2.6.3 Cell viability measurements using alamarBlue HS

The alamarBlue HS reagent was used according to manufacturer's instructions. Briefly, cell samples were mixed well with a pipette tip of an appropriate size. Then, 90 µl of the cell suspension was taken from the sample in duplicate to sextuplicate, depending on the available material, and transferred into a 96-well plate. 90 µl of cell-free medium served as blank. In case of doxycycline rescue experiments, 90 µl media with the adequate doxycycline concentration served as blank for cells that were grown in doxycycline. When all the samples were ready in the 96-well plate, 10 µl of alamarBlue

HS was added to each well, and the contents were mixed by vigorous shaking. The plates were then incubated in a cell incubator at 37°C for 1 – 24 hours, depending on the number of cells present. Afterwards, fluorescence emission at 590 nm was measured following an excitation at 560 nm using a plate reader (Hidex Sense Beta, Type 425-311, Lablogic). In some cases, plates were kept in the dark at 4°C for 1 – 2 days before measurement. In the experiments before Chapter 4.10 (hygromycin influence), linear response was assumed (as per manufacturer's information) and relative viability counted. From Chapter 4.10 onwards, standard curves were employed to obtain absolute cell numbers.

2.6.4 Proliferation with doxycycline

Doxycycline (DOX, Thermo Fisher) was used in our experiments to induce the expression of FLAG-TAL1 in the Jurkat-Tet3G-TRE3G-FLAG-TAL1 and the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell lines, and luciferase in pRetroX-TRE3G-Luc-transduced Jurkat-Tet3G cells.

For functionality testing in Chapter 3.2, Jurkat-Tet3G cells transduced with pRetroX-TRE3G-Luc plasmid were induced with 1000 ng/ml DOX 24 and 72 hours post transduction, before the luciferase assay was carried out 96 hours post transduction. In Chapter 3.3, Jurkat-Tet3G-TRE3G-FLAG-TAL1 cells were incubated with a single dose of 500, 750 or 1000 ng/ml DOX for 48 hours; in Chapter 3.4, Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells were incubated with a single dose of 1000 ng/ml DOX for 48 hours.

To find the appropriate range of concentrations to use in the doxycycline rescue experiments (Chapters 5.2 and 5.4), Western blots for TAL1 were first conducted 24 and 48 hours post-induction with cells incubated with a single dose of 50 – 1000 ng/ml DOX, and then, in a subsequent experiment, 24, 48 and 72 hours after induction with cells incubated with a single dose of 50 – 500 ng/ml DOX.

In all the DOX rescue experiments, cells were incubated with DOX from the day of transduction. The subsequent supplementation of DOX varied depending on the experiment, as follows:

- In the first doxycycline rescue experiment with a panel of sgRNAs (Chapter 5.3), cells were supplemented with 400 ng/ml DOX every 2 days until day 15, then every 3 days until day 21. This was the only experiment where DOX was supplemented at several time points without a change of media and based on the 24 hour half-life of DOX in media.

- In all the following experiments, cells were spun down at 200 x g for 7 minutes, old media was removed and replaced by fresh DOX media, to attain a more precise control of the DOX concentration in the media.
- In the first rescue experiment with edited cells (Chapter 5.4), cells were supplemented with 6.25 – 400 ng/ml DOX every 3 days. Prior to this experiment, edited cells were kept in 400 ng/ml DOX, lowered to 200 ng/ml DOX for one passage immediately preceding the start of this experiment.
- In the second rescue experiment with edited cells (Chapter 5.4), the cells were supplemented with 0.1 – 50 ng/ml DOX every 3 days. Prior to this experiment, edited cells were kept in 400 ng/ml DOX for the first 26 days, then 200 ng/ml DOX for 4 days and without DOX for the remaining 16 days.
- In the second experiment where rescue was assessed from the time of transduction (Chapter 5.4), 6.25, 25 or 200 ng/ml DOX was supplemented every 3 days.
- In the rescue experiment that employed FACS (Chapter 5.5), 6.25, 25, 50 or 200 ng/ml DOX was supplemented every 3 days.
- Finally, in the last rescue experiment in this work (Chapter 5.6), cells were supplemented with 200 or 400 ng/ml DOX every 3 days.

2.7 RT-qPCR

2.7.1 RNA isolation

Around 10^6 cells were pelleted at $200 \times g$ for 5 minutes, washed with PBS, transferred into an Eppendorf tube and recentrifuged at $200 \times g$ for 5 minutes. Cell pellet was resuspended in 1 ml TRIzol (Thermo Fisher) and 200 μ l of chloroform was added to the suspension. The tube was then shaken vigorously for 15 seconds, incubated at room temperature for 2 minutes and shaken again for 15 seconds. The sample was spun down at $16,000 \times g$ for 15 minutes and supernatant was transferred into a clean tube. 1 μ l of GlycoBlue Coprecipitant (Thermo Fisher) and 500 μ l of isopropanol were added to the supernatant and sample was mixed by inversion of the tube, followed by a 10 minute incubation at room temperature. Sample was centrifuged at $16,000 \times g$ for 15 minutes and supernatant was discarded. Pellet was washed in turn with 1 ml 70% ethanol and 1 ml 80% ethanol, each time followed by centrifugation at $16,000 \times g$ for 5 minutes and discarding of the ethanol. After a final spin at $16,000 \times g$ for 5 minutes, residual ethanol was aspirated from the tube and pellet was briefly air-dried before being reconstituted in 16 μ l H_2O .

2.7.2 DNase I treatment and phenol:chloroform extraction

To digest contaminating DNA, samples were mixed with 2 μ l 10x Turbo DNase Buffer (Thermo Fisher) and 2 μ l TURBO DNase (4 U; Thermo Fisher) and incubated at $37^\circ C$ for 30 minutes. RNA from the tested cells was isolated using phenol:chloroform extraction and ethanol precipitation, as described in Chapter 2.1.5. The resulting pellet was dissolved in 12 μ l H_2O .

2.7.3 Reverse transcription

Reverse transcription (RT) was carried out using the High-Capacity cDNA Reverse Transcription Kit (Thermo Fisher) according to manufacturer's instructions. Briefly, 2 μ g of RNA sample were added to 10 μ l of 2x concentrated RT master mix containing 2 μ l of 10x RT Buffer, 0.8 μ l of 100 mM dNTPs, 2 μ l of 10x RT Random Primers, 1 μ l (50 U) of MultiScribe Reverse Transcriptase, and H_2O up to 20 μ l. Samples were incubated at $25^\circ C$ for 10 minutes, followed by 2 hours at $37^\circ C$ and a brief inactivation at $85^\circ C$ for 5 minutes.

2.7.4 Quantitative PCR

SensiMix SYBR Hi-ROX Kit (Bioline) was employed for qPCR experiments. Each reaction was performed at least in duplicate. Each tube contained 5 μ l 2x SensiMix SYBR, 0.25 μ l 10 μ M forward primer, 0.25 μ l 10 μ M reverse primer and 4.5 μ l template cDNA. The primers, synthesized by IDT and purified by standard desalting procedure, were as follows:

	Forward	Reverse
FLAG-TAL1	GACGACGATGACAAGACCGA	TGGCGACGCCGTTTCAG
18S	GTAACCCGTTGAACCCATT	CCATCCAATCGGTAGTAGCG

Table 2.10: qPCR primers used in this doctoral work.

qPCR reactions were carried out in Rotor Gene 6000 Real-Time PCR Machine (Corbett Research). Initial 10 minute denaturation at 95°C was followed by 45 – 55 cycles of 15 seconds of denaturation at 95°C, 15 seconds of annealing at 58 – 59°C and 25 seconds of extension at 72°C and a plate read, and a subsequent melt curve measurement according to the instrument recommendations. Evaluation was performed in Rotor-Gene 6000 Series Software 1.7 using the comparative quantitation method. Melting curves were analyzed to confirm the specificity of the RT-qPCR reactions.

2.8 Western Blot

2.8.1 Buffers

RIPA Cell Lysis Buffer: 50 mM Tris HCl (pH 8.0 at 4°C), 100 mM NaCl, 2 mM MgCl₂, 1% Triton X-100, 0.1% sodium deoxycholate, 0.1% SDS; freshly added before each use: 1 mM DTT, 1/100 cOmplete protease inhibitor cocktail tablet (Roche)/100 µl, 500 U benzonase/100 µl

MOPS Buffer: 50 mM MOPS, 50 mM Tris, 1 mM EDTA, 0.1% SDS

Transfer Buffer: 25 mM Tris, 192 mM glycine, 20% methanol

TBS Buffer: 20 mM Tris, 150 mM NaCl, pH 7.6

TBS-T Buffer: 0.1% Tween 20 in TBS Buffer

Blocking Solution: 5% skimmed milk in TBS-T Buffer

PBS: PBS 100 ml Tablets (Melford)

2.8.2 Antibodies

Antibody	Source	species	type	Dilution used
α-GAPDH	Proteintech (60004-1-Ig)	mouse	monoclonal	1:10,000
α-TAL1	Proteintech (55317-1-AP)	rabbit	polyclonal	1:1,000
α-TAL1	Abcam (ab155195)	rabbit	polyclonal	1:1,000
α-TAL1	Santa Cruz (sc-393287)	mouse	monoclonal	1:200
α-FLAG	Proteintech (66008-3-Ig)	mouse	monoclonal	1:1,000
α-Cas9	Abcam (ab191468)	mouse	monoclonal	1:500
α-mouse IgG	Proteintech (SA00001-1)	goat	polyclonal	1:10,000
α-rabbit IgG	Proteintech (SA00001-2)	goat	polyclonal	1:10,000

Table 2.11: Antibodies used in this doctoral work.

2.8.3 Cell harvest and bicinchoninic acid (BCA) protein assay

Cells were harvested by centrifugation at 400 x g for 5 minutes, washed with PBS and transferred into an Eppendorf tube. Cells were pelleted at 400 x g for 5 minutes, resuspended in 30 – 100 µl RIPA Cell Lysis Buffer and incubated for 10 – 20 minutes, until the suspension lost its viscosity. Cell debris was then pelleted at 20,000 x g for 10 minutes. Supernatant was transferred into a clean low adhesion (LoBind) Eppendorf tube on ice and protein content in the lysate was established by Pierce BCA Protein Assay Kit (Thermo Fisher) according to manufacturer's instructions.

2.8.4 SDS electrophoresis

Invitrogen Novex NuPAGE LDS Sample Buffer (Fisher Scientific) was added to each sample, followed by a brief protein denaturation at 70°C for 10 minutes. The contents of the tube were collected by brief centrifugation. XCell SureLock Mini-Cell Electrophoresis System (Thermo Fisher) gel tank was filled with MOPS buffer and 500 µl of Invitrogen Novex NuPAGE Antioxidant (Thermo Fisher) was added in the upper buffer chamber. Samples of cell lysate, usually containing between 30 and 50 µg of protein per lane (same amount across all groups), were loaded onto a NuPAGE 4 – 12% Bis-Tris gradient gel (Thermo Fisher) along with the SeeBlue Plus2 Pre-Stained Standard (Thermo Fisher). Electrophoresis was carried out at 120 V for 1 hour and 40 minutes.

2.8.5 Protein transfer

SDS electrophoresis was followed by a transfer of proteins from gel to polyvinylidene (di)fluoride (PVDF) membrane or nitrocellulose membrane. In Chapter 3.3, wet transfer was carried out to a PVDF membrane (activated by a 1 minute incubation in methanol) in Transfer Buffer at 4°C overnight (16 hours) at the constant current of 60 mA. From Chapter 3.4 onwards, the transfers were made to a nitrocellulose membrane using the semi-dry Trans-Blot Turbo Transfer System (Bio-Rad) and the ready-to-use Trans-Blot Turbo Mini 0.2 µm Nitrocellulose Transfer Packs (Bio-Rad), at 1.3 Amp for 20 minutes.

2.8.6 Western blotting

After the transfer, the membrane bearing transferred proteins was rinsed twice with TBS and blocked in Blocking Solution for at least 40 minutes before the desired primary antibody was added. The membrane was then incubated with the primary antibody overnight at 4°C on a rocking platform. On the following day, the membrane was washed three times in TBS-T for >5 minutes, then incubated with the appropriate HRP-conjugated secondary antibody in Blocking Solution for >30 minutes before being washed three times in TBS-T for >5 minutes once again. The membrane was then exposed to a freshly combined solution of HRP substrate (SuperSignal™ West Pico Chemiluminescent Substrate, Thermo Fisher) for 5 minutes and chemiluminescence was measured in a G:BOX Chemi-XRQ imager (Syngene).

2.9 CRISPR efficiency assay

10⁴ Jurkat cells co-electroporated with sgRNA-containing lentiGuide-Hygro-dTomato and lentiCas9-Blast were collected 3 days after transfection and vortexed with 0.5 ml of QuickExtract™ DNA Extraction Solution in an Eppendorf tube for 15 seconds. The mixture was then incubated for 6 minutes at 65°C, vortexed again for 15 seconds, and incubated for 2 min at 98°C. 5 µl of this extract was then used to set up a PCR, as set out in Chapter 2.1.4, with the following primers:

sgRNA	Forward	Reverse
MYB1	TCTCACCACCTTGCTCTCCTG	GTGTGTGTCTCCTGAACGGT
TAL1, TAL4	CTGTACCACCCGGATACAGC	CACTGAACCAGACCGATCCC
TAL7*	TCCAGCTAGCGCTGAGTTTC	TTGTCTGTCTGTCTGCCCAC
TAL7*	AGGCACACTCTTTCCTGGTA	TTTGTCTGCCTCTTCTCACG
TAL7*	ACATGCAGGCACACTCTTTC	GTCTGCCTCTTCTCACGTCT
TAL10*	GACACCAACCGGAGTACAGG	GAGCCCCCAGTCATCGAAC
TAL10*	GCCGGTTTCTATTCCAACCC	GCCCCCAGTCATCGAACT
TAL10*	AGCGCTGAGTTTCCGAATGA	GAGCCCCCAGTCATCGAAC
TAL7, TAL10*	GCCACTGGGTTTAAAACGACCTCC	GTGTTTCTGTCGCTGGCTGTATTCC
TAL7, TAL10*	CTGGGTTTAAAACGACCTCCTCTC	CTCCTGATTTCCCCTTCTGTGTTTC
TAL7, TAL10*	CACTGGGTTTAAAACGACCTCCTC	CTTCCTCCATCTCTGTCCTTTTGG
TAL7, TAL10*	CACGCACACTCTCTCTCACAGAAGG	CTCTCTGCCCTTCTCCCCTTTACC
TAL7, TAL10*	CCACAAACATCTACCTCTGCTCAC	CTAACTTTGTCTGCCTCTTCTCACG

Table 2.12: Primers for CRISPR/Cas9 efficiency assays used in this doctoral work.

*These primers were all tested in turn in attempts to amplify the region of *TAL1* where the TAL7 and TAL10 sgRNAs cut, but none of the reactions returned specific products.

For each case of editing to be confirmed, two reactions were set up, one from the DNA of an edited sample and one from the DNA of a control sample (SCR1). The PCR products were purified using the Monarch® PCR DNA Cleanup Kit (NEB) and sent for sequencing. The resulting sequences were compared using The Synthego ICE Analysis tool, 2019, v2.

2.10 Design tools and genome viewers

Predictions of secondary eRNA structures in Chapter 1.5 were generated using the Fold algorithm on the RNAstructure Web Server (<https://rna.urmc.rochester.edu/RNAstructureWeb/>, (Reuter and Mathews, 2010)).

Figures featuring genomic tracks were made using IGV, the Integrative Genomics Viewer (GRCh37/hg19 assembly, (Robinson et al., 2011)), with some help of the UCSC Genome Browser (<http://genome.ucsc.edu>, GRCh37/hg19 assembly, (Kent et al., 2002)).

All sgRNAs, except where stated that these were adopted from published works, were designed using the Broad Institute's CRISPick (<https://portals.broadinstitute.org/gppx/crispick/public>, (Doench et al., 2016; Sanson et al., 2018)).

The iBARs for the CRISPR/Cas9-mediated saturating mutagenesis of the *TAL1* enhancer were generated by Dr. Dan Bose using the Bioconductor platform (<https://bioconductor.org>, (Buschmann and Bystriykh, 2013)).

2.11 Statistical testing

Statistical testing was done in SPSS (version 28.0.1.1). All statistical tests were two-tailed unpaired T-tests.

sign	p-value
NS	not significant
*	$p \leq 0.05$
**	$p \leq 0.01$
***	$p \leq 0.001$

Table 2.13: Signs used to designate p-values in statistical testing.

3 Establishment of a cell line suitable for CRISPR/Cas9-mediated mutagenesis of the *TAL1* enhancer

3.1 Introduction

In order to carry out a CRISPR/Cas9-mediated mutagenesis of the transcribed portion of the *TAL1* enhancer in Jurkat cells, there is a need for a specific cell line on the background of the maternal one. This is to provide for a control group for the experiment. Because the mutagenesis of some of the regions within the *TAL1* enhancer is expected to cause TAL1 downregulation, and because Jurkat cells are naturally sensitive to TAL1 deficiency (Palii et al., 2011; Palomero et al., 2006; Sanda et al., 2012), we predict that cells with disfavoured mutations will be underrepresented in the experiment. As a minimum, this should be the case for mutations that affect transcription factor binding at the enhancer, as demonstrated before for MYB (Mansour et al., 2014). Acting as rescue for cells in the control group, inducible expression of TAL1 should ensure that there is no selective pressure in the control group and the representation of all the mutations inflicted by CRISPR/Cas9 stays the same throughout the experiment.

Aside from a doxycycline-inducible expression of TAL1, which provides for a control in this particular case of CRISPR/Cas9-mediated mutagenesis, a useful feature in a cell line for any CRISPR/Cas9 screen is an expression of Cas9. Cas9, an RNA-guided DNA-nuclease, is an integral part of the CRISPR-Cas9 complex. By stably integrating a gene for Cas9 in the genome of a target cell line, one can rid the system of a variable in the form of oscillating levels of the Cas9 nuclease, which are a likely result in the case when a *Cas9* plasmid is co-transduced at the time of the experiment itself. In this system, Cas9 is under control of a stable promoter. This allows for potential negative effects of the enzyme on the cells to manifest themselves equally in both the experimental and control samples.

The genetically modified cell line for the CRISPR/Cas9-mediated mutagenesis of the *TAL1* enhancer was built on the background of the Jurkat cell line and established in three consecutive steps, with each of the cell lines being more complex than the previous one (Figure 3.1). First, a monoclonal Tet-transactivator protein-expressing cell line was established, on the background of which a monoclonal doxycycline-inducible FLAG-TAL1 cell line was built. The second cell line was then used as background for a monoclonal doxycycline-inducible FLAG-TAL1 cell line stably expressing Cas9. All the cell lines created in this process were monoclonal, to reduce the variety a polyclonal cell population would bring into the experiment, and to allow for a streamlined selection of clones that displayed the most

favourable traits. This chapter will describe the step-by-step process of the establishment of this cell line.

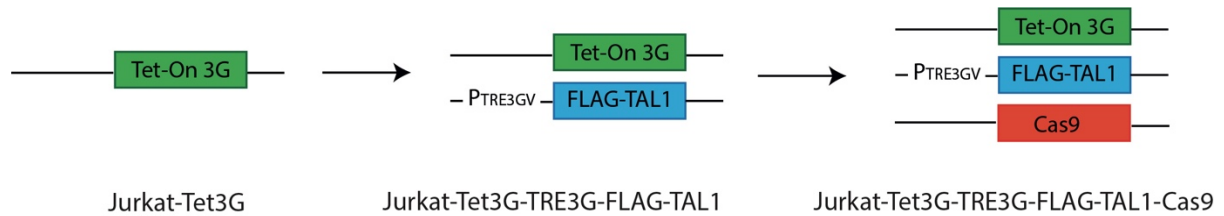


Figure 3.1: The three individual steps in the establishment of the final cell line for the CRISPR/Cas9 mediated mutagenesis of the *TAL1* super-enhancer.

Tet-On 3G – gene for a 3rd generation Tet-transactivator protein; FLAG-TAL1 – *TAL1* gene with an N-terminal FLAG-tag; Cas9 – gene for the Cas9 nuclease; P_{TRE3GV} – a doxycycline-inducible promoter, 3rd generation.

3.1.1 Notes on methods for gene introduction into Jurkat cells

While the establishment of stable cell lines is common, a distinct challenge in establishing a stable Jurkat cell line is its notorious lack of amenability to transfection, which is a rather common characteristic of suspension cells. I tried optimizing chemical transfection of Jurkat cells for the purposes of cell line validation, using Lipofectamine 3000, a lipid nanoparticle formula for difficult-to-transfect cell lines from Thermo Fisher, as well as GeneJuice, a non-lipid reagent from Sigma-Aldrich. Using a plasmid containing EGFP, there was no observable change in fluorescence in any of the transfected cell samples when measured on a plate reader (data not shown). When measured on a flow cytometer, transfection of Jurkat cells using GeneJuice resulted in a 0.2% EGFP-positive population. This figure, however small, was recorded only about 36 hours post transfection, and as such pertains to transiently transfected cells, out of which statistically a very small percentage comes to integrate the gene of interest into their genome. With the odds of chemical transfection against us, we decided to use viral transduction as a means of introduction of the desired genes into the experimental system.

Transduction makes use of retroviral particles for delivery of genetic material into cells. Not only is this method more efficient for hard-to-transfect cells compared to chemical transfection, it is also designed for the integration of the genes of interest into the genome of the host cells (reviewed in (Cockrell and Kafri, 2007)). Briefly, the virion particle, consisting of viral genomic RNA, structural proteins and key enzymatic machinery (reverse transcriptase, integrase) fuses with the host cell

membrane, penetrating into the cytoplasm (reviewed in (Suzuki and Craigie, 2007)). Here, reverse transcription of the viral RNA genome takes place, with the use of the viral reverse transcriptase and an aid of some host cellular proteins. Thus generated viral DNA, complexed with the viral and cellular proteins, is then pulled into the proximity of the nucleus by hijacking the cytoskeleton of the host cell. Some retroviruses, such as HIV-1, are capable to access the nucleus of their host cells by means of active transport through nuclear pores, while others, such as MoMuLV (Moloney murine leukaemia virus), rely on the breakage of the nuclear envelope during mitosis (reviewed in (Suzuki and Craigie, 2007)). Upon entry into the nucleus, the integration into the host genome ensues, an action completed by the joint efforts of the viral integrase and cellular double-strand break DNA repair machinery (Smith and Daniel, 2006). The method of retroviral transduction harnesses this excellent infection potential for the purposes of scientific research and medicine, and we chose to use a classic transduction protocol to make the desired genetic modifications in the Jurkat cell line.

3.2 Establishment of a Jurkat cell line stably expressing Tet-On 3G transactivator protein

As a first step towards the final cell line, I established a cell line that expresses the Tet-On 3G transactivator protein. A transcriptional activator (transactivator) protein, such as the Tet-On 3G transactivator protein, is an indispensable part of any tetracycline-controlled gene expression system, such as the one used in the final cell line to control expression of exogenous FLAG-TAL1. For transcriptional activation to take place, doxycycline, a tetracycline derivative, needs to bind to the transactivator protein, which responds by a change in its conformation (Figure 3.2). In this altered state, transactivator protein becomes able to bind to a tetracycline-responsive promoter that is inactive under normal circumstances. Upon binding, the transactivator protein recruits transcriptional machinery and chromatin modifiers to the promoter, enabling the transcription of the gene under control of this element (Gossen et al., 1995). The transcriptional activator used in this cell line, Tet-On 3G transactivator protein, is a highly active and highly doxycycline-sensitive variant of a transcriptional activator, developed by spontaneous virus evolution (Zhou et al., 2006) from earlier variants (Baron et al., 1997; Gossen et al., 1995; Urlinger et al., 2000).

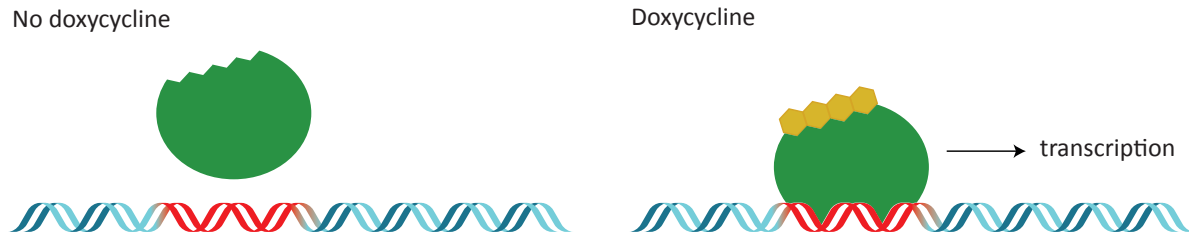


Figure 3.2: Tet-transactivator protein in action.

Doxycycline binding (yellow hexagons) triggers a conformational change in Tet-On 3G transactivator protein (green), allowing it to bind to a tetracycline-responsive promoter (red) and activate transcription of a downstream gene of interest. Figure was adapted from Retro-X™ Tet-On® 3G Inducible Expression System User Manual (Clontech, 2013).

To deliver the Tet-On 3G transactivator protein into Jurkat cells, I used the pRetroX-Tet3G plasmid (Clontech). This is a plasmid for retroviral delivery, containing Tet-On 3G transactivator protein under control of CMV promoter. As a selection marker, the plasmid includes a gene for aminoglycoside phosphotransferase, which confers resistance against the antibiotic G418.

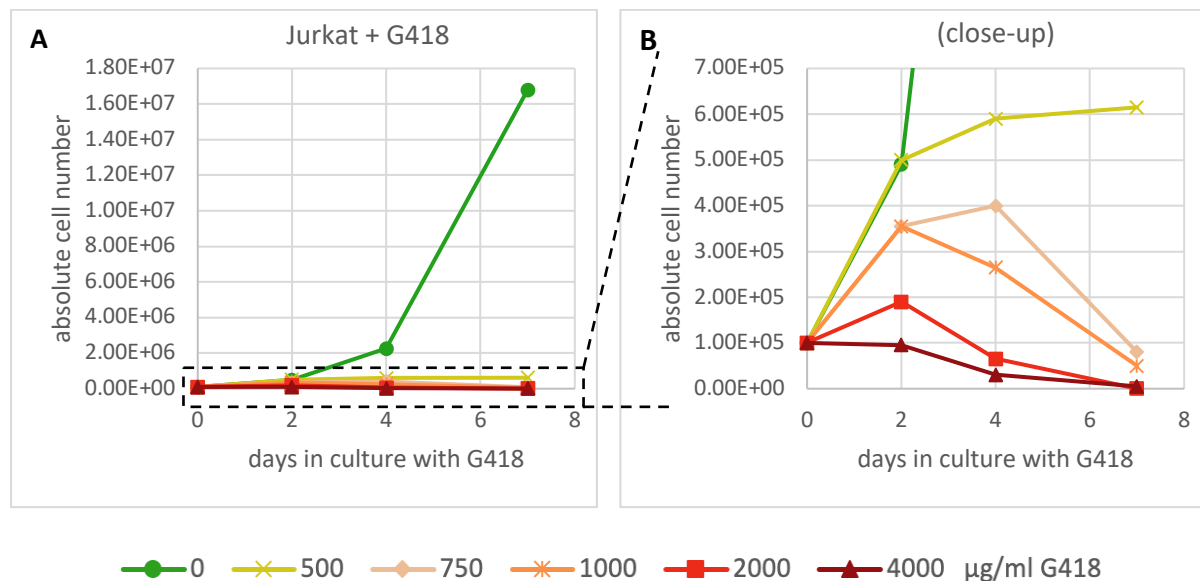


Figure 3.3: Proliferation of Jurkat cells with various concentrations of G418.

Cells were counted on a haemocytometer using Trypan Blue stain. Graphs show absolute cell numbers in the course of the experiment. **(B)** is a close-up of the dashed box in **(A)**. In **(A)**, proliferation curves for 750 – 2000 µg/ml G418 are partially obscured by the proliferation curve for 4000 µg/ml G418. Results show data from a single replicate.

To determine the appropriate concentration of G418 for the selection of transduced cells, untransduced Jurkat cells were first tested for their intrinsic sensitivity to this antibiotic. The lowest concentration of the drug at which there were no live cells left after 7 days of culture, 2000 $\mu\text{g/ml}$, was defined as the minimum lethal dose (Figure 3.3).

Next, as a means of introduction of Tet-On 3G transactivator protein into Jurkat cells, a transduction was carried out using viral supernatant from GP2-293 cells following co-transfection with pRetroX-Tet3G and the viral envelope vector pVSV-G. Transduced cells were selected for by an incubation with 2000 $\mu\text{g/ml}$ G418, giving rise to a polyclonal Tet-On 3G transactivator protein cell line. To obtain individual clones, two rounds of single-cell sorting into 96-well plates were performed 13 and 19 days after the selection antibiotic was added. Cells for single-cell sorting were selected using basic flow cytometry data (Figure 3.4A, B). Healthy live cells were first gated based on their forward and side scatter parameters (Figure 3.4A) – forward scatter roughly corresponds with the particle size, while side scatter is indicative of the internal complexity of the cell. Single cells were then selected based on the correlation between the area under the voltage curve of the detected pulse for forward scatter and the height of this curve (Figure 3.4B). First signs of expanding clones appeared within a week after the sorting (Figure 3.4C). Overall, 88 clones were designated and their expansion monitored.

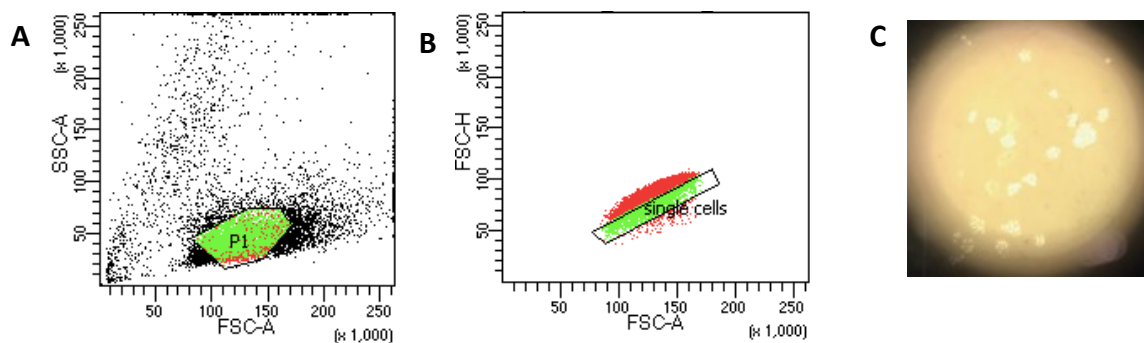


Figure 3.4: Single cell sorting and clonal expansion.

(A, B) Flow cytometry figures showing the selection of the population of live, single cells for single-cell sorting. **(A)** Scatter plot shows gating for a population of healthy live cells (P1). **(B)** Scatter plot shows gating for single cells. SSC-A, Side scatter – area (under the voltage curve of the detected pulse); FSC-A, Forward scatter – area (under the voltage curve of the detected pulse); FSC-H, Forward scatter – height (of the voltage curve of the detected pulse). **(C)** Expanding Jurkat-Tet-On 3G transactivator protein clone, 11 days post sorting.

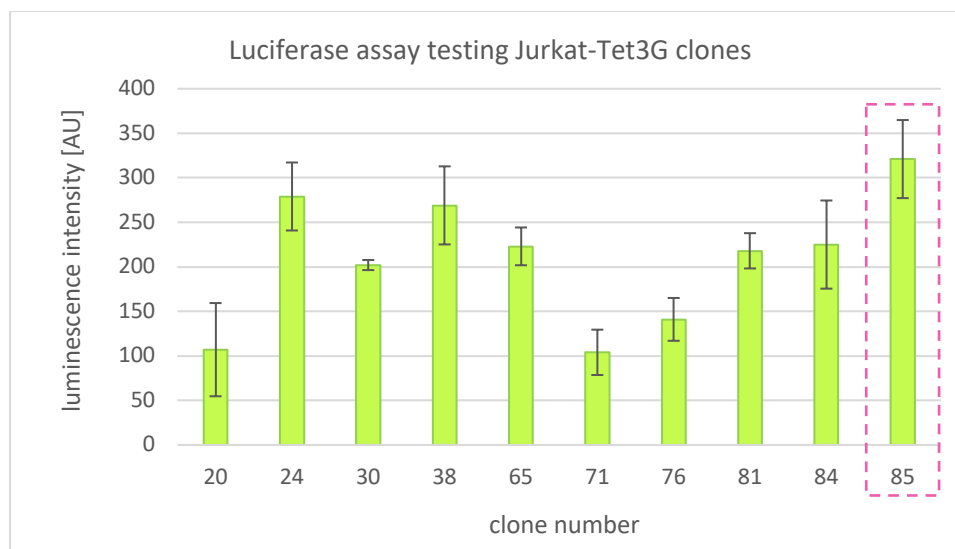


Figure 3.5: Luminescence of monoclonal Tet-transactivator Jurkat cell lines transduced with luciferase.

Luminescence intensity of the individual clones in this graph is expressed in arbitrary units (AU). To account for background luminescence, the luminescence intensity value of untransduced Jurkat cells was subtracted from the luminescence intensity produced by each of the luciferase-transduced Jurkat-Tet3G clones. Error bars represent standard deviation from 2 technical replicates. Pink dashed box outlines the selected clone.

To select a clone for further work, the fastest expanding monoclonal cell lines were tested for Tet-On 3G transactivator protein functionality and activity using luciferase assay. Selected monoclonal cell lines were transduced with a packaged pRetroX-TRE3G-Luc plasmid (Clontech), which contains a gene for luciferase under control of P_{TRE3GV} , a tetracycline-responsive promoter that will be used in the next step of building the final cell line (Figure 3.5). The cell cultures were supplemented with 1000 ng/ml doxycycline 24 and 72 hours after transduction to activate the transactivator protein and induce luciferase expression, before being harvested 96 hours post transduction. Lysates from the tested cell lines were subjected to luciferase assay using Firefly Luciferase Flash Assay Kit (Thermo Scientific). Detecting any luminescence above background (luminescence of cells untransduced with luciferase) was to be taken as an indicator of Tet-On 3G transactivator protein functionality in the given cell line, and luminescence intensity was going to serve as a measure of Tet-transactivator protein activity. Luminescence measurements uncovered more than three-fold differences in luminescence intensity between the monoclonal cell lines (Figure 3.5). Based on these results, clone 85, which had the highest luminescence intensity of all the tested clones and was one of the fastest growing clones, was chosen for further work. Frozen stocks were prepared from this clone (designated Jurkat-85, or 85) and two others (24 and 84) that also displayed a high luminescence and a good growth rate.

3.3 Establishment of a Jurkat cell line expressing doxycycline-inducible FLAG-TAL1

Continuing with the establishment of the final Jurkat cell line for CRISPR/Cas9-saturating mutagenesis of the *TAL1* enhancer, I proceeded to establish a cell line that expressed exogenous, doxycycline-inducible FLAG-tagged TAL1.

As explained earlier, a positive doxycycline-mediated transcriptional control is based on the binding of this tetracycline derivative to Tet-transactivator protein, a highly effective kind of which is expressed in abundance in the newly established Jurkat-Tet3G cell line, and the subsequent binding of this complex to a tetracycline-responsive promoter. The promoter to be used in this cell line is designated P_{TRE3GV} , and is a version of the P_{TRE3G} promoter (Loew et al., 2010) optimized for improved performance in lentiviruses and retroviruses by Clontech (Retro-X™ Tet-On® 3G Inducible Expression System User Manual, Clontech, 2013).

The P_{TRE3G} promoter (Loew et al., 2010) is in itself an optimized version of the *E. coli*-inspired tetracycline-responsive promoter Ptet-1, which consists of a minimal CMV promoter sequence downstream of seven repeats of 19-bp tetracycline operator sequences, in a head-to-tail orientation (Gossen and Bujard, 1992). These repeats are bound by homodimers of variously engineered Tet-transactivator proteins, triggering either transcriptional repression (Gossen and Bujard, 1992) or activation (Gossen et al., 1995). When used for transcriptional activation, however, the original Ptet-1 promoter displayed a propensity to generate background expression of the downstream gene of interest at levels that were non-negligible in some settings. To minimize the background levels of expression from the promoter while maintaining its high inducibility, Loew and colleagues carefully tinkered with the original, randomizing the spacing between the tetracycline operators and altering the minimal CMV promoter by introducing consensus TATA-box and TFIIB-binding site sequences, truncating the 5'-UTR, eliminating binding sites for transcription factors, and removing hairpin structures that could trigger cap-independent translation initiation (Loew et al., 2010). The resulting tetracycline-responsive promoter displayed more than a hundred times lower background expression compared to Ptet-1, with the inducibility in a transient transfection being 50,000-fold, and 10,000-fold in a stably integrated system (Loew et al., 2010). The P_{TRE3GV} is a version of this potent promoter, further optimized for retroviral delivery (Retro-X™ Tet-On® 3G Inducible Expression System User Manual, Clontech, 2013).

In the second intermediate cell line, the newly introduced gene for TAL1 is tagged with a FLAG-epitope tag at its N-terminus. The FLAG epitope, a short, eight amino acid-long tag (DYKDDDDK), serves as an identifier of the exogenous protein as well as a tag for purification if necessary.

For the establishment of the cell line with doxycycline-inducible expression of FLAG-TAL1, there is a need for a plasmid coding for the sequence in question, under control of a tetracycline-responsive promoter. For this purpose, I set off to clone a collated sequence of the three exons coding for TAL1, without any intronic or untranslated regions, and with the FLAG-coding sequence tagged onto its N-terminus (as per plasmid pSIN4-EF1a-TAL1-IRES-Puro, Addgene plasmid #61065, (Elcheva et al., 2014)), under the control of the P_{TRE3GV} promoter in the pRetroX-TRE3G plasmid (Clontech) adjusted to express the hygromycin resistance gene instead of the puromycin resistance gene. This adjustment (Dr. Dan Bose) was done with the future CRISPR/Cas9 transductions in mind, as the plasmid we planned to use for the downstream CRISPR/Cas9 experiments encoded the puromycin resistance cassette. If the puromycin resistance gene was already stably integrated in the genome, we couldn't use it as a selection marker in the future. However, while the cloning of this pRetroX-TRE3G-Hygro-FLAG-TAL1 plasmid seemed to have gone successfully, upon cell transduction and application of hygromycin selection pressure we repeatedly failed to observe any surviving cells in our cell line establishment attempts. Upon having sequenced the hygromycin resistance gene, we found that the resistance cassette was compromised. In the interest of time, the FLAG-TAL1 sequence was cloned into the available pRetroX-TRE3G plasmid with the original puromycin resistance gene (Dr. Dan Bose) and downstream plans were changed with regards to the plasmid to be used for CRISPR editing.

The cell line with doxycycline-inducible expression of FLAG-TAL1 was prepared by the transduction of a selected monoclonal Jurkat cell line (Jurkat-85) stably expressing Tet-On 3G transactivator protein with retrovirus carrying the FLAG-TAL1 gene under control of the P_{TRE3GV} promoter in the pRetroX-TRE3G plasmid. Cells were cultured in selection medium (0.5 $\mu\text{g}/\text{ml}$ puromycin, as determined by a dose response experiment in Jurkat cells, Figure 3.6) for a week before being single-cell sorted into 96-well plates.

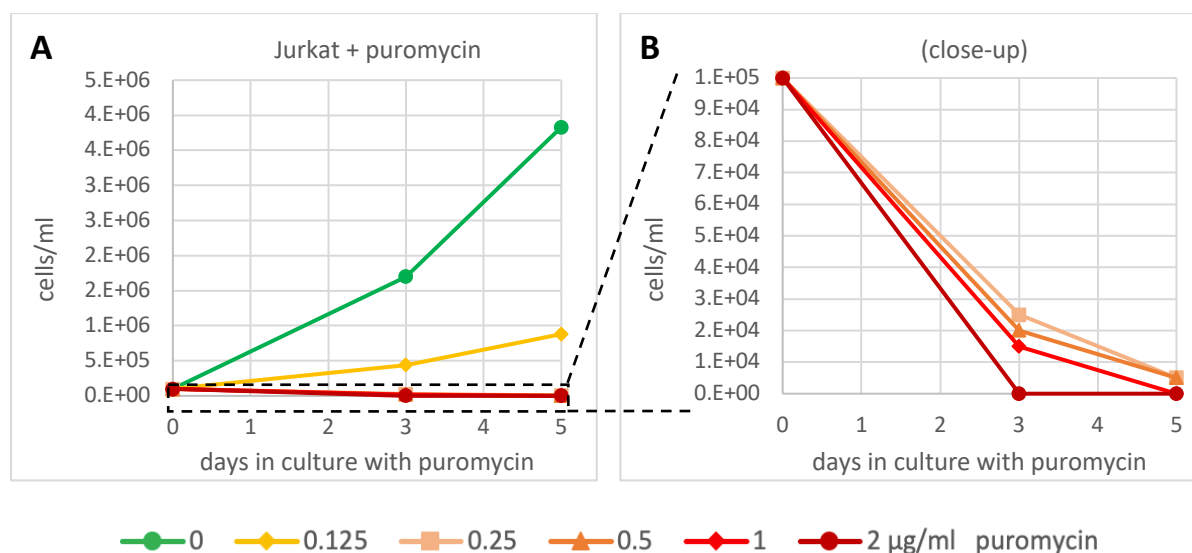


Figure 3.6: Jurkat cell proliferation with puromycin over the course of 5 days.

Jurkat cells were cultured with a range of puromycin concentrations and counted on a haemocytometer with the use of Trypan Blue stain. **(B)** is a close-up of the dashed box in **(A)**. In **(A)**, proliferation curves for 0.25 – 1 µg/ml puromycin are obscured by the proliferation curve for 2 µg/ml puromycin. Results are from one replicate.

Several different conditions were tested for clone expansion, including using pre-conditioned media (filtered medium in which a healthy culture of Jurkat cells was allowed to proliferate for 24 – 48 hours) and different types of 96-well plates. The clones grown in the pre-conditioned media expanded faster than the clones growing in unconditioned media, and were more numerous, with about 50% extra colonies compared to clones expanding in fresh media. There was no difference between using round-bottom and flat-bottom plates for clonal expansion of transduced Jurkat cells in terms of clone viability, expansion rates or number of clones obtained from each type of plate. In the most prolific plates, over a quarter of wells gave a viable clone.

Upon expansion, selected clones were subjected to a three-step elimination screening process to find amongst them the monoclonal cell line with the most suitable behaviour: high inducibility with the lowest possible background expression of FLAG-TAL1, and with the inducibility preferably easily fine-tuned by increasing the concentration of doxycycline.

As a first step towards choosing the best-suited cell line, all clones that were satisfactorily expanded within 4 weeks of single-cell sorting were tested for their *FLAG-TAL1* expression upon doxycycline induction using reverse transcription quantitative PCR (RT-qPCR; Figure 3.7) with one of the primers directed to the FLAG tag.

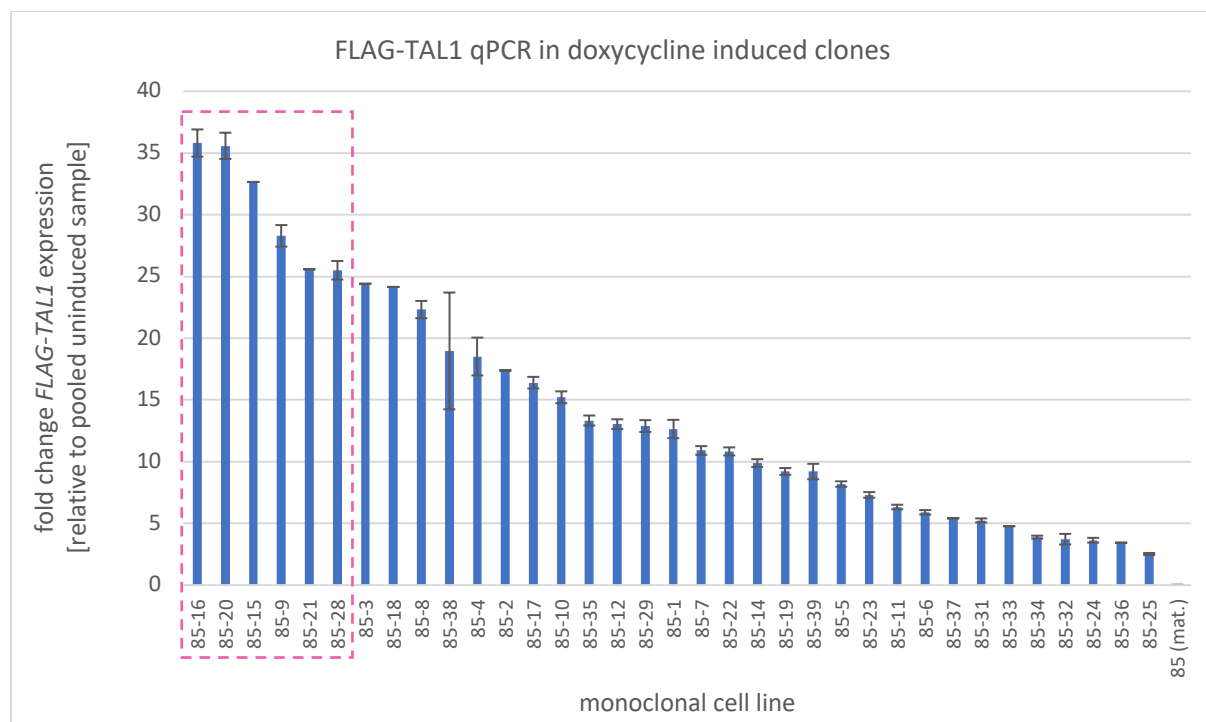


Figure 3.7: Expression of FLAG-TAL1 in doxycycline-induced monoclonal Jurkat-Tet3G-TRE3G-FLAG-TAL1 clones.

Various Jurkat-Tet3G-TRE3G-FLAG-TAL1 clones (85-XX) were induced with 1000 ng/ml doxycycline and their FLAG-TAL1 expression was assessed by RT-qPCR 48 hours later. Bars represent a fold change in FLAG-TAL1 expression relative to a pooled sample of uninduced 85-XX clones. 85, Jurkat-Tet3G (clone 85, maternal line to 85-XX; no *FLAG-TAL1* in the genome). Data normalized to *18S* rRNA expression. Results show data from a single biological replicate. Error bars represent standard deviation from 3 technical replicates. Pink dashed box outlines the clones that were taken into the next stage of screening.

As a second step, six clones with the highest induced *FLAG-TAL1* mRNA levels were screened for the levels of background (uninduced) *FLAG-TAL1* expression by RT-qPCR in two ways (Figure 3.8). Firstly, uninduced *FLAG-TAL1* mRNA expression in Jurkat-Tet3G-TRE3G-FLAG-TAL1 cell lines was compared to the signal obtained from their maternal cell line (Jurkat-Tet3G, clone 85; Figure 3.8A). Secondly, *FLAG-TAL1* levels in the individual doxycycline-induced Jurkat-Tet3G-TRE3G-FLAG-TAL1 clones were compared to the *FLAG-TAL1* levels in each of the uninduced samples (Figure 3.8B).

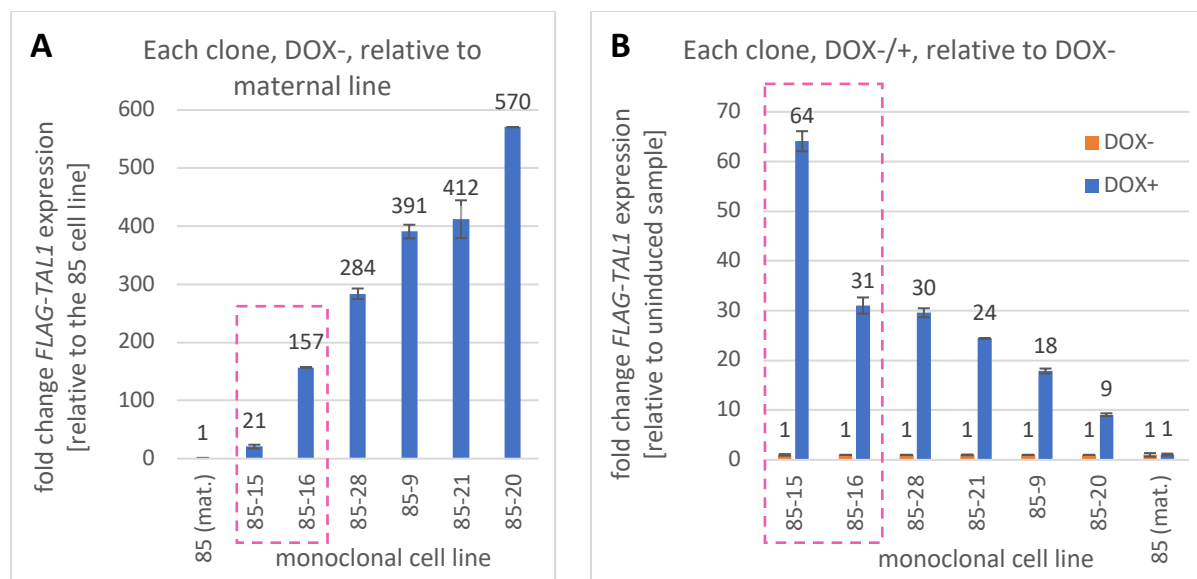


Figure 3.8: Background FLAG-TAL1 expression in selected uninduced monoclonal Jurkat-Tet3G-TRE3G-FLAG-TAL1 cell lines (85-XX).

FLAG-TAL1 levels were determined by RT-qPCR, data were normalized to 18S rRNA expression. **(A)** Background expression of FLAG-TAL1 relative to maternal cells. Bars represent fold change in FLAG-TAL1 mRNA levels of uninduced 85-XX clones normalized to Jurkat-85 (85, no FLAG-TAL1 in the genome). **(B)** Induced vs uninduced FLAG-TAL1 levels. Bars represent the ratio between FLAG-TAL1 mRNA levels in each of the induced (1000 ng/ml doxycycline (DOX+), 48-hour incubation) and the uninduced (no doxycycline, DOX-) 85-XX sample. Results show data from a single biological replicate. Error bars represent standard deviation from 3 technical replicates. Pink dashed box outlines the clones that were taken into the final stage of clone selection.

Finally, the two clones with the lowest background expression of FLAG-TAL1 among the tested clones (clone 15 and 16) were subjected to increasing levels of doxycycline and their FLAG-TAL1 inducibility assessed by Western blotting (Figure 3.9A). Clone 16 showed a better correlation of FLAG-TAL1 protein levels with the concentration of doxycycline within the tested range (red arrow), but also displayed additional unexpected bands on the Western blot membrane probed with anti-FLAG antibody (purple arrows), which was not the case for clone 15. Within the range of doxycycline used, the exogenous TAL1 expression levels were at a similar level as that of the major endogenous TAL1 isoforms combined (compare red and blue arrows, the two major isoforms run at approximately 40 kDa and 44 kDa; Figure 3.9). As clone 15 had a much lower background expression of FLAG-TAL1 (Figure 3.8) and showed no unidentifiable FLAG products (Figure 3.9), clone 15 was chosen for further work.

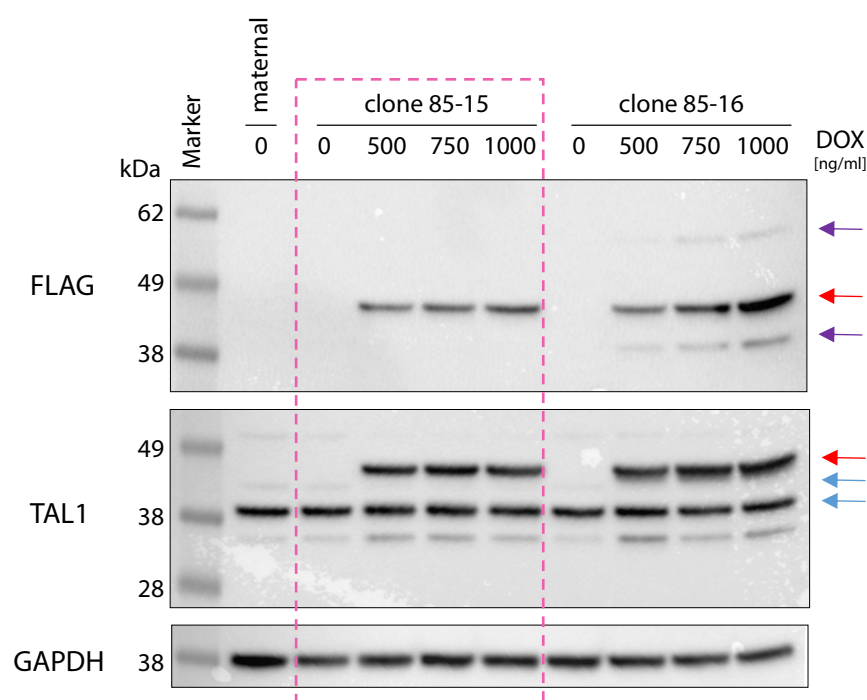


Figure 3.9: FLAG-TAL1 and endogenous TAL1 protein expression in monoclonal Jurkat-85-TRE3G-FLAG-TAL1, clones 15 and 16.

TAL1 expression was assayed by Western blotting. Doxycycline (DOX) concentrations are indicated in ng/ml, induction period was 48 hours. Purple arrows indicate unidentifiable FLAG products. Red arrows indicate FLAG-TAL1, blue arrows endogenous TAL1. Antibodies used were anti-FLAG (66008-3-Ig, Proteintech); anti-TAL1 (55317-1-AP, Proteintech) and anti-GAPDH (60004-1-Ig, Proteintech). The clone selected for further work is highlighted in a pink dashed box.

3.4 Establishment of a Jurkat cell line expressing doxycycline-inducible FLAG-TAL1 and Cas9

As the third and final step towards the establishment of the cell line to be used in the CRISPR/Cas9-mediated mutagenesis of the *TAL1* enhancer, the existing Jurkat cell line with doxycycline-inducible expression of FLAG-TAL1 was to be equipped with a gene for the Cas9 enzyme.

Cas9 is an RNA-guided nuclease which cuts both the DNA strands wherever it is directed by a short guide RNA sequence (Chapter 1.4.1; reviewed e.g. in (Bannikov and Lavrov, 2017)). It was first described as a part of a bacterial defence system against viruses (Barrangou et al., 2007), and later repurposed as a scientific tool for genome editing due to its fantastic programmability (Cong et al., 2013; Jinek et al., 2013; Mali et al., 2013). The cuts in DNA resulting from Cas9 nuclease activity are

usually repaired by non-homologous end joining, which is imperfect and often leads to insertions and deletions (indels) in the repaired sequence. A more precise way of fixing the Cas9-inflicted cuts is homology-directed repair, which is generally less common, although the odds depend on the particular cell line, locus and cell-cycle phase (reviewed in (Bannikov and Lavrov, 2017)). With the planned CRISPR-Cas9 screen, we rely on the non-homologous end joining repair pathway to create indels that will disrupt the natural sequence of the *TAL1* enhancer and thereby potentially change its potency.

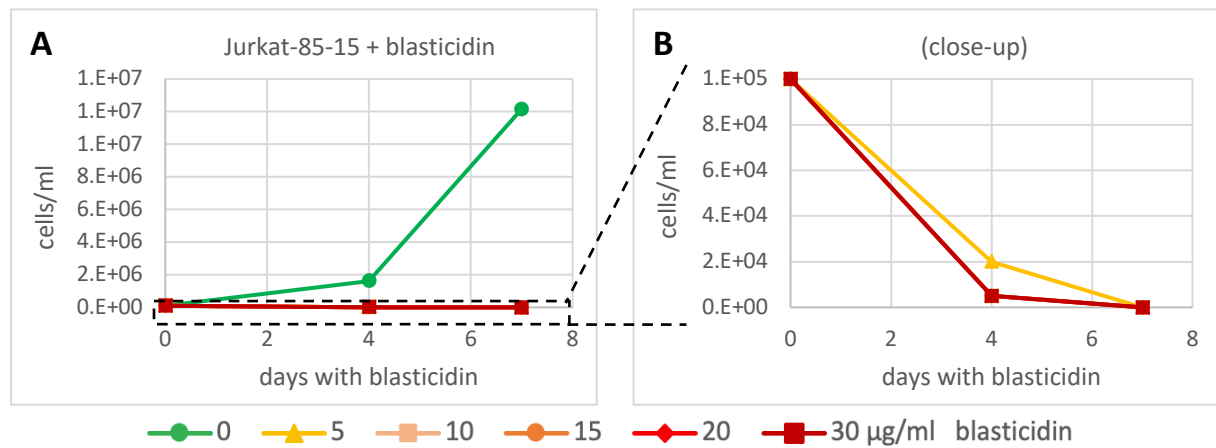


Figure 3.10: Jurkat-Tet3G-TRE3G-FLAG-TAL1 proliferation with a range of blasticidin concentrations. Jurkat-Tet3G-TRE3G-FLAG-TAL1 (clone 85-15) cells were cultured with blasticidin (0 – 30 µg/ml) for 7 days and counted on day 4 and day 7 on a haemocytometer using the Trypan Blue stain. (B) is a close-up of the dashed box in (A). Proliferation curves for 10 – 20 µg/ml blasticidin are obscured by the proliferation curve for 30 µg/ml blasticidin. Results are from a single replicate.

To build the final cell line with a doxycycline-inducible expression of FLAG-TAL1 and stable expression of the Cas9 nuclease, a monoclonal Jurkat-Tet3G-TRE3G-FLAG-TAL1 cell line (Jurkat-85-15) was transduced with lentivirus containing the gene for Cas9 and a blasticidin resistance cassette. After 10 days of incubation in selection medium (10 µg/ml blasticidin, the experimentally determined minimal lethal dose of the antibiotic, Figure 3.10), surviving cells were single-cell sorted into 96-well plates.

Very few clones expanded into cell lines, so there was no need for a streamlined screening process this time. The four clones that expanded sufficiently within 5 weeks after single-cell sorting were tested for stable Cas9 expression using Western blot, and three of them were found to be positive for Cas9 (Figure 3.11A). These clones were further tested to confirm that the FLAG-TAL1 inducible

expression had been preserved (Figure 3.11B). Clone 85-15-4 was selected for future work, as the cell line appeared to be morphologically the healthiest out of all the positive clones.

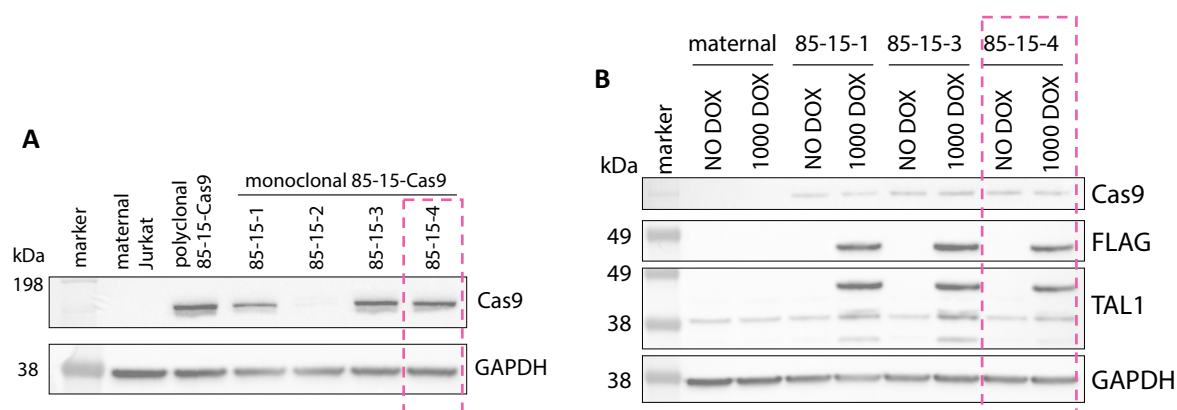


Figure 3.11: Cas9, TAL1 and FLAG-TAL1 expression in selected Jurkat-85-15-Cas9 clones.

(A) Expression of Cas9 in selected Jurkat-85-15-Cas9 clones. **(B)** Expression of Cas9, FLAG-TAL1 and endogenous TAL1 in selected Jurkat-85-15-Cas9 clones without doxycycline induction (NO DOX) or upon 48-hour induction with 1000 ng/ml doxycycline (1000 DOX). Maternal, maternal Jurkat cells. Antibodies: Cas9 antibody (ab191468, Abcam), FLAG antibody (66008-3-Ig, Proteintech), TAL1 antibody (55317-1-AP, Proteintech) and GAPDH antibody (60004-1-Ig, Proteintech). Pink dashed box highlights the selected clone of the final cell line.

3.5 Discussion

The first step in conducting the CRISPR/Cas9-mediated mutagenesis of the coding portion of the *TAL1* enhancer was the establishment of an experimental system, a monoclonal Jurkat cell line with a doxycycline-inducible expression of FLAG-TAL1 and a stable expression of the Cas9 nuclease, Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9. The final monoclonal cell line was established in a stepwise manner, through two intermediate monoclonal cell lines, to ensure that all the features newly introduced into the experimental system were of the highest possible functionality and efficacy. The large differences between the luminescence intensity of the tested Jurkat-Tet3G clones (Figure 3.5 in Chapter 3.2), the striking variation in FLAG-TAL1 inducibility and background expression in the Jurkat-Tet3G-TRE3G-FLAG-TAL1 clones (Figure 3.7 and Figure 3.8 in Chapter 3.3), and the variability in the condition and proliferation rates in the clones of the final line, Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 (Chapter 3.4), fully justified this approach.

The step-by-step approach not only allowed for an effective, streamlined selection of the best quality features, it also significantly decreased the number of clones that had to be screened. Besides, judging from the available evidence – especially the low number of clones obtained for Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 and the condition of these final lines – it is also quite likely that the step-wise approach was the only way to establish this complicated genetically modified cell line at all: a synchronous introduction of all the features might have been too stressful for any of the cells to survive the process.

The process of establishment of stable cell lines is a common practice, but not without potential stumbling blocks. In the case of Jurkat cells, the challenges include extremely low transfectability and limited options for obtaining monoclonal populations.

We circumvented the problem of low transfectability by using viral transduction, a method often employed for hard-to-transfect cells (Chapter 3.1.1). Early results from Jurkat transfection experiments completely justified the use of the transduction method, which, while highly efficient, also has its drawbacks. The main disadvantage of the approach these days is an increased demand for time and labour, which explains why it often isn't the first method of choice for genetic manipulation. Another drawback of using virus for delivering a genetic payload is the inherent danger it may present, although there has been a fantastic progress in the area of developing relatively safe viral vector systems for transduction (reviewed in (Gouvarchin Ghaleh et al., 2020)).

In the establishment of our stable cell lines, I used two different viral systems: Tet-transactivator protein and FLAG-TAL1 were introduced by gamma-retroviral infection (Chapters 3.2 and 3.3), while Cas9 was transduced using a 3rd generation lentiviral packaging (Chapter 3.4). The main difference between these two systems is that lentiviruses, as opposed to gamma-retroviruses, are able to cross an intact nuclear membrane and as a result can infect not only dividing, but also non-dividing cells (Chapter 3.1.1). However, as Jurkat cells divide about once in every 24 hours, this distinction wasn't a worrisome one. Nevertheless, it is important to be mindful of what type of system one is using, as each type of transfer plasmid (plasmid coding the gene of interest) can be efficiently packaged only by a set of compatible components, including the appropriate viral enzymatic machinery and structural proteins. The viral envelope is interchangeable between systems; the vesicular stomatitis virus glycoprotein (VSV-G) envelope is most commonly used due to its wide tropism (reviewed in (Hastie et al., 2013)), which is now understood to be owing to the interaction of the VSV-G with the ubiquitously expressed low-density lipoprotein receptor (Finkelshtein et al., 2013; Nikolic et al., 2018).

The plasmid containing Tet-transactivator protein (pRetroX-Tet3G) and the plasmid that *FLAG-TAL1* was cloned into (pRetroX-TRE3G) were both part of a MoMuLV retrovirus-based doxycycline-inducible

expression system (Retro-X™ Tet-On 3G Inducible Expression System, Clontech; Chapters 3.2 and 3.3). This system also features GP2-293 cells, a HEK293-derived cell line with stably integrated MoMuLV *Gag* and *Pol* genes (structural proteins and indispensable viral enzymes, respectively). To prepare retroviral particles using this cell line, a viral envelope plasmid is co-transfected with the transfer plasmid. The use of GP2-293 eliminates the need for co-delivering an additional (*Gag-Pol*) plasmid into the cells, which is conducive to higher viral titres.

Cas9 was introduced into our intermediate Jurkat-Tet3G-TRE3G-FLAG-TAL1 cell line from an HIV-1-based transfer plasmid, lentiCas9-Blast (Addgene plasmid #52962, (Sanjana et al., 2014)), using a 3rd generation lentiviral packaging ((Dull et al., 1998); Chapter 3.4). Regular HEK293T cells were used for the packaging, which necessitated the co-transfection of four different plasmids: the transfer plasmid coding for Cas9; a plasmid coding for the VSV-G viral envelope, pMD2.G (Addgene plasmid #12259, (Dull et al., 1998)); a plasmid encoding the main structural proteins of the virion (*Gag*) and the enzymatic machinery (*Pol*), pMDLg/pRRE (Addgene plasmid #12251, (Dull et al., 1998)); and a plasmid coding for the essential HIV-1-regulatory protein Rev, pRSV-Rev (Addgene plasmid #12253, (Dull et al., 1998)). Rev facilitates nuclear export of viral RNA transcripts by binding to the Rev-responsive element, an RNA motif which is included on the *Gag-Pol*-coding plasmid, pMDLg/pRRE.

Both the gamma-retroviral and the lentiviral transduction were efficient in introducing genes of interest into Jurkat cells, in a stark contrast with the inefficiency of common chemical transfection methods (Chapter 3.1.1). Polyclonal lines expressing the proteins of interest were obtained 1 – 2 weeks after transduction, depending on the selection antibiotic (Chapters 3.2 – 3.4).

Once polyclonal lines were established and tested for the expression of our proteins of interest, the next step was to create monoclonal cell lines, for the reasons outlined above. Growing clonal lines from single cells can be a challenge in mammalian cell culture for several different reasons, largely depending on the specific cell line. Perhaps the most notable challenge in producing monoclonal Jurkat cell lines is that they grow in suspension, which makes it impossible to use the easier, more cell-friendly methods that are habitually employed for generating monoclonal adherent cell lines, such as trypsin discs and cloning rings. Limiting dilution, where cells are aliquoted into plates at a concentration that should theoretically ensure monoclonality, remains an option for Jurkat cells, but the presence of a single cell in each well at the start of the experiment is a statistical function and cannot be easily confirmed. Therefore, limiting dilution is often repeated several times to increase the probability of producing a true monoclonal cell line, which is a time-consuming endeavour. The method that we chose to use instead is single-cell sorting.

Single-cell sorting is in principle the same as limiting dilution, but more sophisticated in execution, and largely without the statistical uncertainty. The single-cell deposition efficiency of this method generally exceeds 99% (99.7% for BD FACSMelody, BD FACSMelody™ Cell Sorter Brochure, BD, 2018), and it is done using fluorescence-activated cell sorters. These instruments sort cells based on flow cytometry data, which can include fluorescence intensity, but not as a necessity: it is possible to sort cells according to their size and granularity, which are reflected in the values of their forward and side scatter (Chapter 3.2). This information allows to select with a relatively high accuracy for healthy, living, single cells. When single-cell sorting a polyclonal cell line, which, such as in our case, has undergone an antibiotic selection, no other information is necessarily required. These minimal requirements, along with the high confidence in single-cell deposition, make single-cell sorting using fluorescence-activated cell sorters an attractive option. The drawback of this approach is the physical strain it puts on the sorted cells, and the fact that after the cells have gone through this strain, they are separated from other cells, which can introduce further stress. While the strain from FACS can be at least partially alleviated by using lower flow rates when sorting, the stress stemming from isolation can be helped by using pre-conditioned media. Overall, when optimal conditions were used, single cell sorting using a FACS machine provided us with a sufficient number of clonal Jurkat cell lines every time (Chapters 3.2 – 3.4).

The subsequent streamlined testing of the obtained monoclonal cell lines using a variety of methods, including luciferase assay (Chapter 3.2), RT-qPCR (Chapter 3.3) and Western blotting (Chapters 3.3 and 3.4), allowed us to choose the best amongst the available clones at each step. In the first instance, based on its strong proliferation and a strong performance in the luciferase assay, clone 85 of the Jurkat-Tet3G cell line, expressing the Tet-On 3G transactivator protein, was selected to serve as a background for the doxycycline-inducible FLAG-TAL1 cell line (Chapter 3.2). It can, however, also serve as a background for other potential doxycycline-inducible Jurkat cell lines that we might want to establish in the future. Next, clone 85-15 of the Jurkat-Tet3G-TRE3G-FLAG-TAL1 cell line was chosen for its high induced levels and low background expression of FLAG-TAL1 (Chapter 3.3). Finally, clone 85-15-4 of the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line was selected for its apparent morphological health compared to its two other competitors, 85-15-1 and 85-15-3 (Chapter 3.4).

The establishment of the final cell line, Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9, took overall longer than the preparation of the two previous cell lines, and produced only four expanding clones in the first month (Chapter 3.4). This was from the same number of 96-well plates as with the two intermediate cell lines, which yielded, in a shorter timeframe, ten to twenty times more viable clones. Furthermore, all of the newly established final monoclonal cell lines grew at a slower rate compared to the maternal cell line, and their cultures contained a higher percentage of morphologically unsound cells. Taken

together, these observations suggest that the stable expression of the Cas9 enzyme takes its toll on the cells.

However, there are also distinct advantages in using a cell line with a stably integrated Cas9. The effectiveness of such a system surpasses that of a co-delivery of sgRNA and Cas9 about 100 times (Sanjana et al., 2014), likely due to the sizeable nature of the enzyme (Montalbano et al., 2017). The stable integration together with clonal selection also leads to higher accuracy and better reproducibility in pooled screens due to a steady level of Cas9 expression in all cells. Finally, stable expression of Cas9 in cell lines designed for CRISPR/Cas9 screens is a common feature (for example, (Canver et al., 2015; Shi et al., 2015)) and has been previously used in Jurkat cells as well (Chi et al., 2016; Shang et al., 2018). These reasons led us to adopt the stable Cas9 strategy.

Concerning doxycycline concentrations for the rescue, from the first data that we collected, the inducible FLAG-TAL1 seems to be amply responsive to doxycycline. While higher concentrations of doxycycline can be toxic to cells, Jurkat cells seemed to be mostly unperturbed (if a little slowed down in growth) even at the highest doxycycline concentration used, 1000 ng/ml, which, according to current evidence, is likely much higher than the concentration that will be needed for rescue. It is also important to note that while Tet-On 3G is highly sensitive to doxycycline (a tetracycline-group compound), it is poorly induced by tetracycline itself, so care has to be taken to use specifically doxycycline for the induction of this system. Finally, it is also paramount that certified, good quality tetracycline-free FBS be used for all the doxycycline-sensitive experiments in the cell line. As the members of the tetracycline class are used as antibiotics in cattle, they may be contained in regular serum, at levels that can activate doxycycline-inducible transcription, hence interfering with doxycycline-sensitive experiments.

Overall, the stepwise transduction, FACS-based single-cell sorting and streamlined testing led to the establishment of a viable monoclonal cell line expressing doxycycline-inducible FLAG-TAL1 and constitutive Cas9: Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9.

4 TAL1 downregulation reduces the proliferation rate of the newly established cell line

4.1 Introduction

In the last chapter, we have followed the steps leading to the establishment of the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line. This new cell line had next to be validated. Two main points had to be confirmed: firstly, that the newly established cell line responds to TAL1 downregulation in the same way the maternal cell line does, i.e. by a drop in proliferation rate; secondly, that this drop in growth rate can be rescued by the addition of doxycycline, which induces the expression of exogenous TAL1. This chapter deals with the validation of the former and the following chapter provides details about the experiments that were done in attempts to validate the latter.

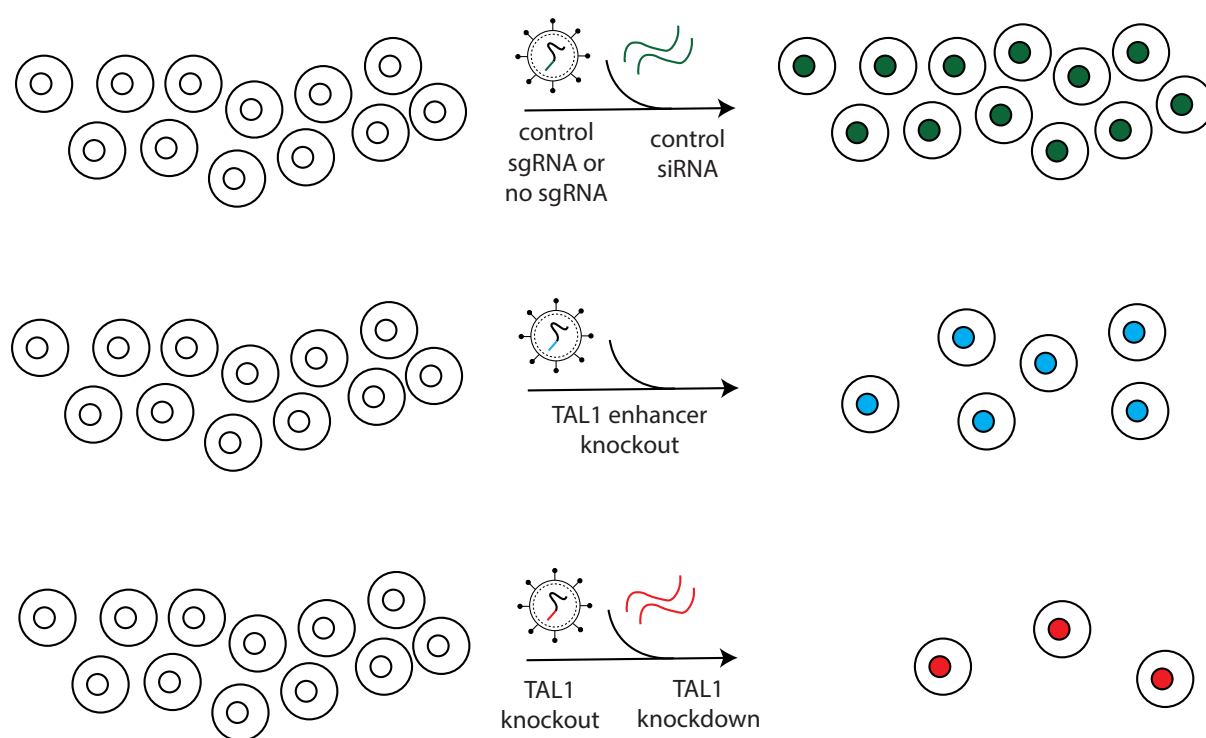


Figure 4.1: Expected behaviour of the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line in CRISPR/siRNA validation experiments.

Control cells (green) proliferate unhampered. *TAL1* enhancer knockout (blue) causes some growth challenge, *TAL1* knockout or knockdown (red) leads to a serious proliferation challenge.

In broad terms, in an ideal experiment that validates the sensitivity of the newly established cell line to TAL1 abrogation, cells would show a clear pattern of decreased proliferation after *TAL1* knockout or *TAL1* knockdown, and possibly, in the case of the knockout, an eventual complete stop of proliferation and cell death (Figure 4.1). This would be in contrast to control cells, transduced or transfected with non-targeting sgRNAs or siRNAs, whose TAL1 expression has not been compromised, and which should continue proliferating and expanding as a cell line. We would expect the viability of cells whose TAL1 expression is downregulated to a degree (but not completely abrogated) to fall between that of the control group and the group where TAL1 is completely abolished. This should be the case when targeting the *TAL1* enhancer, for example. As a general rule, we would expect that the more downregulated the expression of TAL1, the slower the cells would proliferate. Meanwhile, we would fully anticipate survival in control cell samples that were not TAL1-challenged (Figure 4.1).

4.2 Guide RNA sequences for CRISPR validation

For the screen to be viable, we expected the proliferation of the newly established cell line to be reduced by transfection of sgRNAs targeting *TAL1*. It was because the planned screen of the *TAL1* enhancer is based in CRISPR/Cas9-mediated mutagenesis that we first set off to validate the dependence of the new cell line on the endogenous TAL1 levels using CRISPR. As the new cell line, Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9, already expresses the Cas9 enzyme (Figure 3.11 in Chapter 3.4), the only remaining requirement for CRISPR editing to take place in these cells is the introduction of the RNA component – the single guide RNA (sgRNA). The sgRNA directs the Cas9 enzyme to the regions in the genome where editing is desired (Figure 1.8 in Chapter 1.4.1).

To confirm that Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 had preserved the addiction to the TAL1 transcription factor, I designed four guide RNA sequences targeting the *TAL1* gene, out of which two target the 5'-untranslated region (5'-UTR) of TAL1 (designated as TAL1, TAL4) and two target the first exon of TAL1 (designated as TAL7, TAL10; Figure 4.2 and Table 4.1). I also adopted, from a published paper (Mansour et al., 2014), a guide RNA sequence targeting the MYB-binding site in the *TAL1* enhancer, whose mutation should be sufficient to seriously impair the function of this regulatory element and therefore also cause a marked decrease in TAL1 expression. This guide RNA sequence targeting the MYB-binding site was designated as MYB1 (Figure 4.2 and Table 4.1).

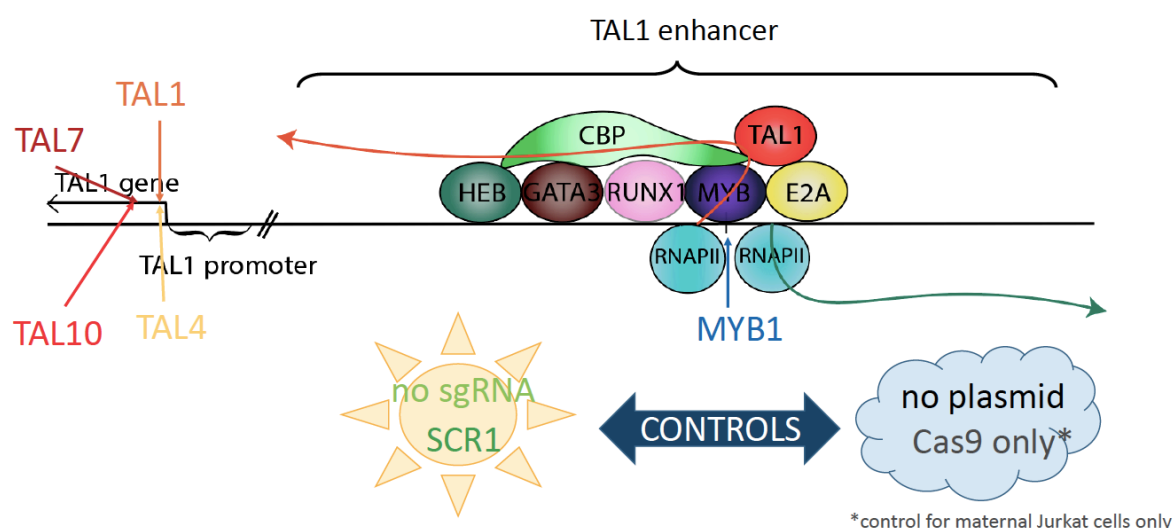


Figure 4.2: An overview of sgRNAs for the CRISPR validation of Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 addition to TAL1.

TAL1 and TAL4 target the 5'-UTR of *TAL1*; TAL7 and TAL10 target the first exon of *TAL1*; MYB1 targets the MYB-binding site at the *TAL1* enhancer. All of these guide sequences are cloned under control of the U6 promoter and in front of a full sgRNA scaffold sequence in the lentiGuide-Hygro-dTomato plasmid. Negative controls (without predicted negative effects on growth rate) are the lentiGuide-Hygro-dTomato plasmid containing a scrambled sgRNA (SCR1) and the lentiGuide-Hygro-dTomato plasmid containing no guide RNA sequence (no sgRNA). Cells without the plasmid will experience a lack of selection marker and can therefore be used as positive control for selection pressure. In the maternal line, where experiments were later performed for comparison (Chapters 4.3; 4.9; 4.10; 4.11; 4.13), the lack of *Cas9* in the genome calls for the introduction of a plasmid coding for *Cas9* alongside the sgRNA-containing plasmid. In these cells, using only the *Cas9*-encoding plasmid and not the sgRNA-containing plasmid was used as an additional positive control for selection pressure (*Cas9* only).

The five experimental guide RNA sequences, along with a non-targeting control – a scrambled guide RNA sequence (Lawhorn et al., 2014) – were cloned into the lentiGuide-Hygro-dTomato plasmid (Addgene plasmid #99376, (Ho et al., 2017)), which contains a single guide RNA scaffold, a hygromycin resistance gene and a gene for the orange fluorescent protein dTomato. In a number of the following experiments, the lentiGuide-Hygro-dTomato plasmid without a guide RNA sequence was used as an additional or alternative negative control, designated as “no sgRNA”. Positive controls for selection pressure in the CRISPR validation experiments were generally cells that didn’t receive the lentiGuide-Hygro-dTomato plasmid, designated either as “virus only” in transduction experiments, or “PBS” in transfection experiments.

Overall, five plasmids of the lentiGuide-Hygro-dTomato backbone encoding sgRNAs designed to decrease TAL1 expression in Jurkat cells were prepared – TAL1, TAL4, TAL7, TAL10 and MYB1 (Figure 4.2 and Table 4.1).

sgRNA name	Target sequence	Target region
TAL1	GAGTGGAGATCCTATTCAGA	<i>TAL1</i> 5'-UTR – endogenous <i>TAL1</i>
TAL4	GAATAGGATCTCCACTCCGC	
TAL7	GCGGCCCTTTAAGTCTCTCG	<i>TAL1</i> exon (2/4) – endogenous <i>TAL1</i> + <i>FLAG-TAL1</i>
TAL10	TGAGGCTGTAGAGCAGCGCG	
MYB1	CACAGAAAGACGGTTAGGAAA	<i>TAL1</i> enhancer – MYB binding site
SCR1	GCTGATCTATCGCGGTCGTC	non-targeting control
no sgRNA	-	control using plasmid without an sgRNA
no plasmid/PBS	-	control for antibiotic selection (hygromycin)
Cas9 only	-	control for second antibiotic selection in Jurkat cell experiments (blasticidin)

Table 4.1: sgRNAs used for the CRISPR/Cas9 validation of the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line's dependence on *TAL1*.

4.3 sgRNAs are cleaving at the expected loci

To test whether the newly designed sgRNAs targeting *TAL1* and the *TAL1* enhancer cleave at the expected sites, we introduced the sgRNAs into wild-type (wt) Jurkat cells by nucleofection and checked for CRISPR editing 3 days later by sequencing analysis. Figure 4.3 shows such a CRISPR-editing analysis using The Synthego ICE Analysis tool (v2) for a sample edited with the TAL4 sgRNA. Figure 4.3A shows the base calls from Sanger sequencing for both the experimental and the control samples around the guide sequence (this information comes from .ab1 files, which are part of the package of sequencing results). The sequencing signal is clean in both samples before the cut site (black dashed line); past the cut site, the base calls in the experimental sample become mixed, while the control signal continues to be clean. The software uses these differences to determine the actual cut site (Figure 4.3A), the percentage of edited cells (Figure 4.3B) and even the distribution of the indels (Figure 4.3C, D).

Using the sequencing analysis and The Synthego ICE Analysis tool (v2), we confirmed editing at the expected loci in the case of TAL1 and TAL4 sgRNAs, which occurred with a 10% efficiency for the TAL1-edited sample (data not shown) and 18% efficiency for the TAL4-edited sample (Figure 4.3).

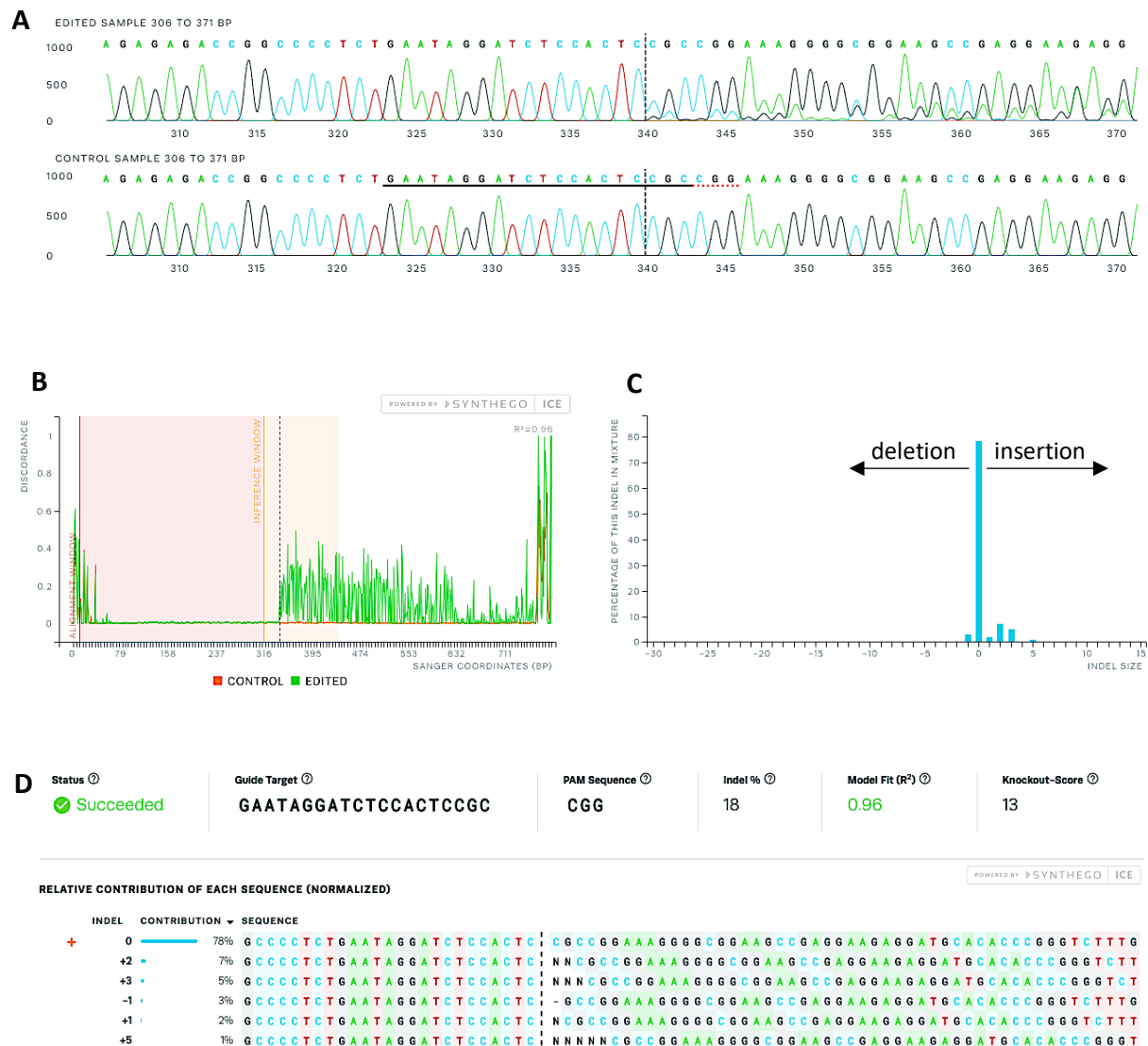


Figure 4.3: CRISPR analysis of a sample of Jurkat cells edited with an sgRNA targeting the 5'-untranslated region of TAL1 (TAL4 sgRNA).

Jurkat cells were co-electroporated with a plasmid coding for Cas9 and the lentiGuide-Hygro-dTomato plasmid containing TAL4-sgRNA. Jurkat cells co-electroporated with a plasmid coding for Cas9 and the non-targeting, scrambled sgRNA, SCR1, were used as the control. **(A)** A comparison of the sequence around the cut site in the control (SCR1) and the edited sample (TAL4). The TAL4 sgRNA sequence is underlined, the red dotted line indicates the PAM sequence. The vertical dashed line represents the cut site, as determined by the analysis software. **(B)** A discordance plot shows the portion of the sequencing signal in the experimental sample (TAL4, green) which disagrees with the control sample (SCR1, orange). The vertical black dashed line indicates the position of the cut site. **(C)** A distribution of indels within the entire experimental sample (TAL4) genome population, inferred by the software. **(D)** A more detailed view of the mix of sequences present in the experimental sample (TAL4) and their relative proportions. The vertical dashed line represents the cut site. The orange plus sign (+) denotes unedited sequence. The analysis was performed using The Synthego ICE Analysis tool (v2).

For samples edited with TAL7 and TAL10 sgRNAs, I wasn't able to amplify the correct sequence, and therefore couldn't confirm the editing outcomes by sequencing analysis. However, a Western blot from cells collected on day 3 post-transfection showed a clear TAL1 downregulation for the TAL7- and TAL10-edited samples (Figure 4.4).

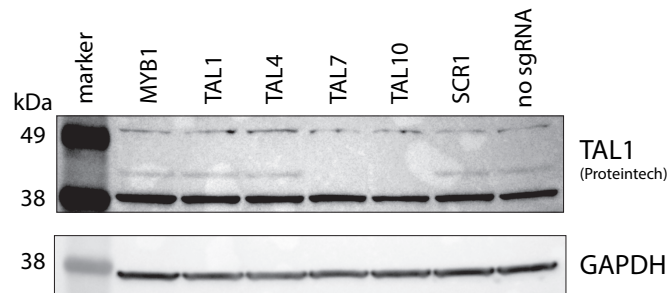


Figure 4.4: TAL1 expression in wild-type Jurkat cells 3 days upon TAL1 knockout.

Jurkat cells were co-electroporated with a plasmid coding for Cas9 (blastocidin resistance) and a plasmid containing an sgRNA (hygromycin resistance) against one of the following regions: *TAL1* enhancer (MYB1), *TAL1* 5'-UTR (TAL1, TAL4) or *TAL1* exonic region (TAL7, TAL10). Negative controls for the experiment were a non-targeting sgRNA (SCR1) and a plasmid with no sgRNA (no sgRNA). Positive controls for selection pressure were cells mock-transfected with PBS (no plasmid), or transfected only with the Cas9 plasmid (Cas9 only). Cells were cultured in 10 µg/ml blastocidin and 500 µg/ml hygromycin from day 1 post nucleofection to select for cells that contained both the plasmids. In the Western blot, GAPDH served as loading control. Antibodies used were anti-TAL1 antibody (55317-1-AP, Proteintech) and anti-GAPDH antibody (60004-1-Ig, Proteintech).

The *TAL1* enhancer-edited sample (MYB1 sgRNA) was impossible to check by sequencing analysis, as *TAL1* enhancer is a monoallelic feature caused by a mutation at exactly the site we are targeting with MYB1 sgRNA – this means that even in a wild-type Jurkat cell sample, this region will have mixed base calls in this region, regardless of any editing (data not shown).

4.4 First experiments show a massive cell death problem and inconsistencies

The first attempt at CRISPR validation of Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 dependence on TAL1 was carried out in a very similar way to how the stable cell lines were established. The aforementioned plasmids containing anti-TAL1 sgRNAs were packaged using 3rd generation lentiviral packaging in HEK293T cells; the resulting particles were used for the transduction of Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells (clone 85-15-4), which stably express Cas9. We chose to select for transduced cells using hygromycin (rather than the other available marker, the fluorescent protein dTomato) due to the simplicity and ease of use of antibiotic selection pressure. We determined the minimal lethal dose previously in maternal Jurkat cells as 500 µg/ml (Figure 4.5). Virus only control cells demonstrated no viability 7 days post transduction, indicating that the selection with hygromycin was successful.

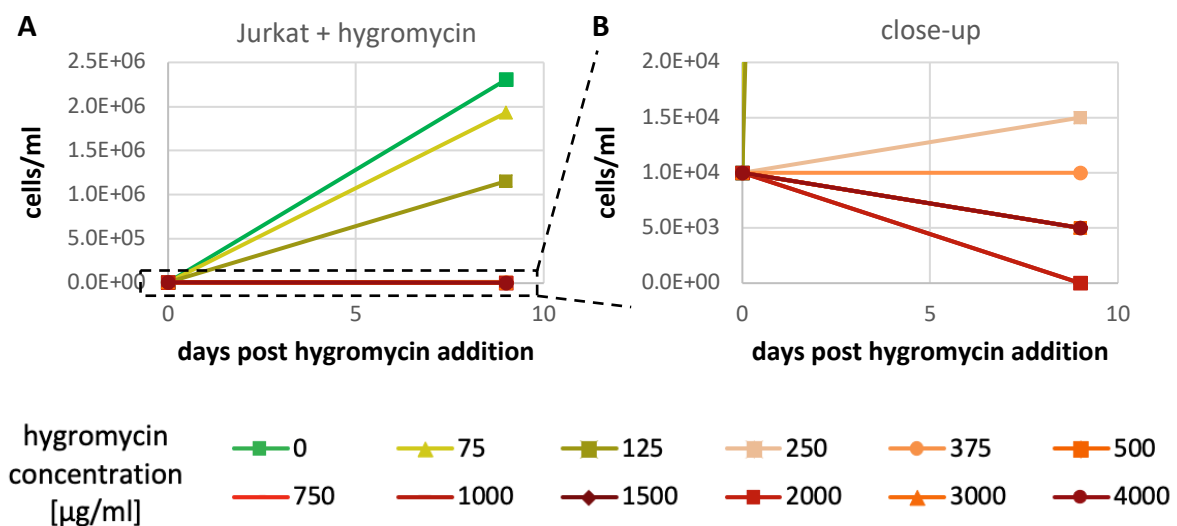


Figure 4.5: Proliferation of Jurkat cells with hygromycin.

Jurkat cells were cultivated with a range of hygromycin concentrations (0 – 4000 µg/ml) and counted on a haemocytometer using Trypan Blue; graphs show data from a single replicate. **(B)** shows a close up view of **(A)**, zooming in on the proliferation pattern of cells challenged with 250 – 4000 µg/ml hygromycin. In **(B)**, proliferation curves for 500, 750, 1500 and 3000 µg/ml hygromycin are obscured by the proliferation curve for 4000 µg/ml hygromycin, and the proliferation curve for 1000 µg/ml hygromycin is obscured by the proliferation curve for 2000 µg/ml hygromycin.

Viability measurements using alamarBlue HS reagent showed some unexpected behaviour, with most experimental groups faring better 7 days after transduction than the non-targeting control cells

(Figure 4.6). Furthermore, an examination of the cells under the microscope revealed virtually no surviving cells a week after transduction. This led us to hypothesize that the unexpected behaviour might only be a product of working with very low cell numbers.

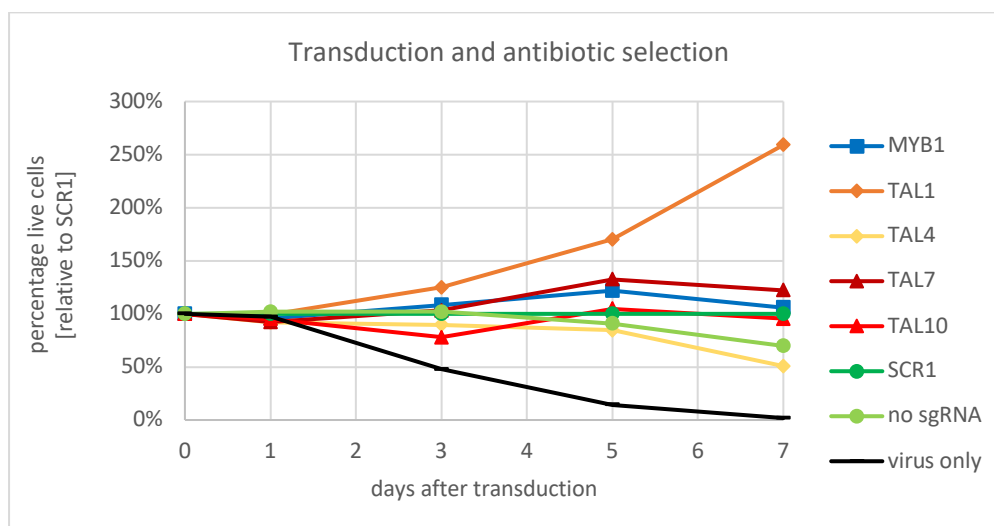


Figure 4.6: Proliferation rates of Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells transduced with unconcentrated viral supernatants encoding various sgRNAs targeting *TAL1* gene and *TAL1* enhancer.

Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 (clone 85-15-4) cells were transduced with fresh, unconcentrated viral particles containing the lentiGuide-Hygro-dTomato plasmid encoding a hygromycin resistance cassette and either an sgRNA targeting the *TAL1* gene (TAL1, TAL4, TAL7 or TAL10), or an sgRNA targeting the MYB-binding site at the *TAL1* enhancer (MYB1). Controls were transduced with viral particles containing the lentiGuide-Hygro-dTomato plasmid encoding a hygromycin resistance cassette and either a scrambled sgRNA (SCR1), or the plasmid without an sgRNA (no sgRNA). Positive controls for antibiotic selection were cells transduced with cargo-less viral particles (virus only). 500 µg/ml hygromycin was added to cells 24 hours post transduction. Viability, expressed as percentage of living cells relative to the SCR1 sample, was measured at indicated time points using alamarBlue HS reagent. These results are from a single replicate.

Such a massive cell death, however, was surprising, as it was unobserved in any of the previous transduction experiments, where cell lines were always clearly expanding by the end of a week of selection. One of the potential culprits was identified as the antibiotic of choice – for example, an unfortunate mutation in the resistance gene in our plasmid, an improper expression of the resistance gene from our plasmid or a different kinetics of expression of this resistance gene compared to the resistance cassettes I used previously could all be to blame.

4.5 Selection using FACS

The quickest way to see whether the massive cell death was connected to the hygromycin selection in any way was to make use of the other selectable marker in the sgRNA-bearing plasmid, the fluorescent protein dTomato. This strategy circumvented the need for hygromycin selection and instead, employed FACS to select cells with presumed sgRNA expression based on fluorescence intensity.

Importantly, the dTomato-based approach gave rise to the expected behaviour in terms of relationships between the proliferation rates of the individual samples over time (Figure 4.7A). The cells transduced with *TAL1* exon-targeting sgRNAs (TAL7 and TAL10) expanded much slower than the control samples (SCR1, no sgRNA), while cells transduced with *TAL1* 5'-UTR-targeting sgRNAs (TAL1, TAL4) and the *TAL1* enhancer-targeting sgRNA (MYB1) expanded slower than the control cells, but faster than the cells edited at *TAL1* exon.

Crucially, FACS data immediately highlighted the very low efficiency of transduction seen using our original transduction protocol, with only ~1.62% of cells transduced successfully (dTomato positive cells in Figure 4.7B). This potentially explains the high cell mortality seen using hygromycin resistance in our first experiment.

Western blotting showed a decreased *TAL1* expression with the MYB1-sgRNA on day 27 after transduction, while the TAL7- and TAL10-edited samples showed an almost complete abrogation of *TAL1* expression (Figure 4.7C). However, *TAL1* levels in the *TAL1* 5'-UTR-edited samples remained largely unperturbed (Figure 4.7C).

In the first repeat of this experiment, while the cells expanded into cultures eventually, the cell numbers at the beginning of the experiment were extremely low (as little as 5×10^2 cells in a sample on day 3 after sorting). This was likely a combined effect of the stress caused by FACS and the low numbers of collected cells (1×10^4 per sample). The latter was due to a very low transduction efficiency, between 0.94 and 2.29% (Figure 4.7B and data not shown). While cells in the first repeat of this experiment grew slowly, but expanded eventually, in the second repeat, the cells didn't recuperate after the sorting at all and died off by day 10 after FACS (data not shown).

Instrumentally, the experiments in this chapter pointed out the low transduction efficiency we were achieving with our concurrent transduction protocol, highlighting this as the potential primary issue causing the massive cell death we observed in the experiment described in the previous chapter.

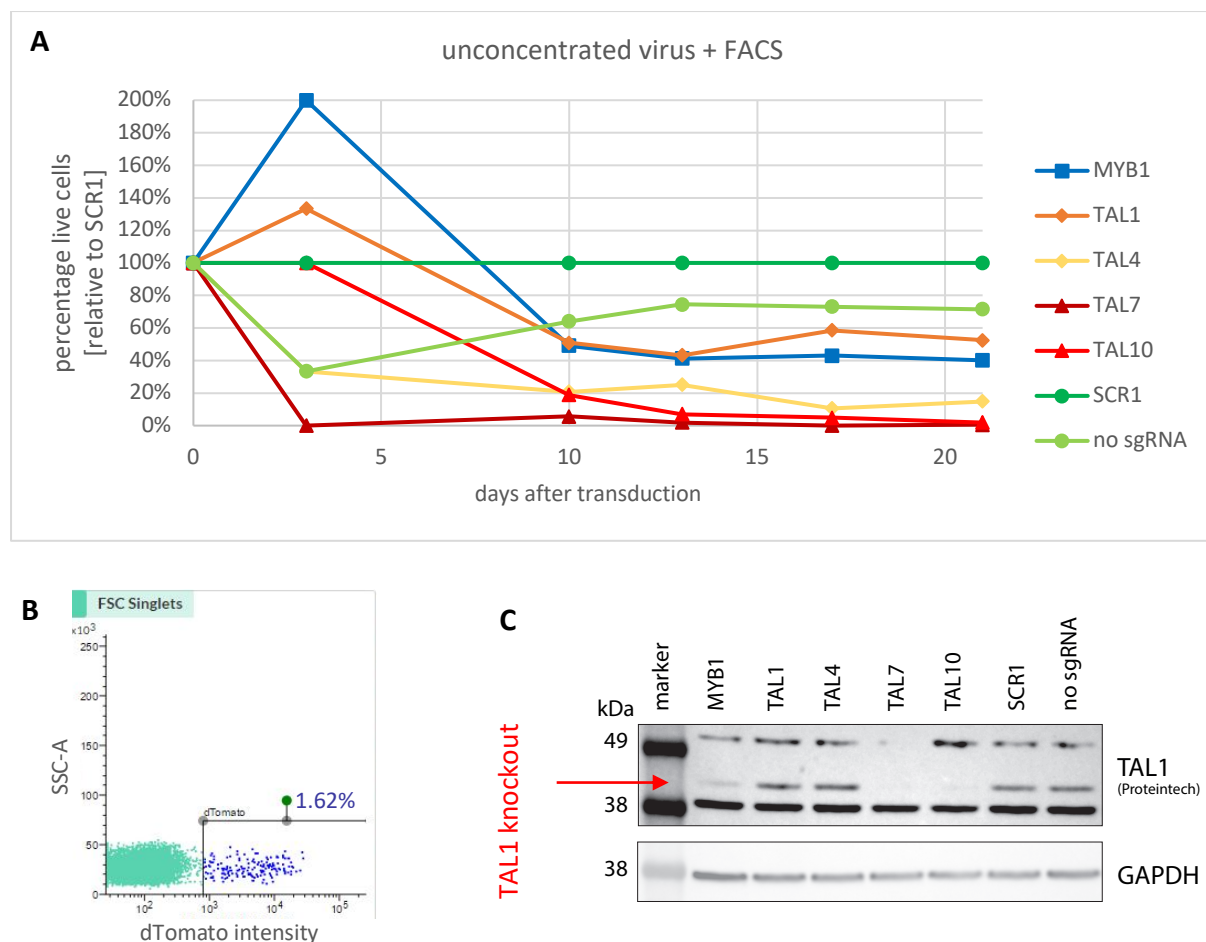


Figure 4.7: Proliferation and TAL1 expression of differentially CRISPR-edited, FACS-sorted Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells.

Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 (clone 85-15-4) cells were transduced using fresh, unconcentrated viral particles containing the lentiGuide-Hygro-dTomato plasmid encoding a gene for the dTomato orange fluorescent protein and either an sgRNA targeting the *TAL1* gene (TAL1, TAL4, TAL7 or TAL10), or an sgRNA targeting the MYB-binding site at the *TAL1* enhancer (MYB1). Controls were cells transduced with viral particles containing a plasmid encoding the dTomato orange fluorescent protein and either a scrambled sgRNA (SCR1), or the plasmid without an sgRNA (no sgRNA). 1×10^4 transduced cells were collected 24 hours post transduction using FACS (FACSMelody, BD) based on their dTomato fluorescence. **(A)** Percentage of living cells in the course of the experiment, relative to SCR1. Day 0 on the graph is the day of FACS. Live cells were counted on a haemocytometer using the Trypan Blue stain at the indicated time points. These results are from a single replicate. **(B)** A flow cytometry image of the SCR1 sample at the time of FACS. **(C)** A Western blot shows TAL1 expression in each sample on day 27 after transduction. GAPDH serves as loading control. Antibodies used were anti-TAL1 (55317-1-AP, Proteintech) and anti-GAPDH (60004-1-Ig, Proteintech).

4.6 Optimized viral preparation improves transduction efficiency

In the belief that low transduction efficiency was the root cause of the massive cell death and the behavioural inconsistencies we had observed in our first attempts at validating the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line, we decided to explore the potential of increasing the transduction efficiency by way of optimization of the transduction protocol.

To that end, we employed a viral concentrator formula, Lenti-X Concentrator (Takara), which allows to increase the viral titre up to 100 times. I transfected HEK293T packaging cells with a control plasmid (lentiGuide-Hygro-dTomato-SCR1) and the appropriate viral components and prepared concentrated lentivirus from the supernatant collected after the first 24 hours and the second 24 hours after the transfection medium was changed. I transduced Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 with different amounts of the concentrated lentivirus and measured dTomato fluorescence using a flow cytometer in all groups 2 days later.

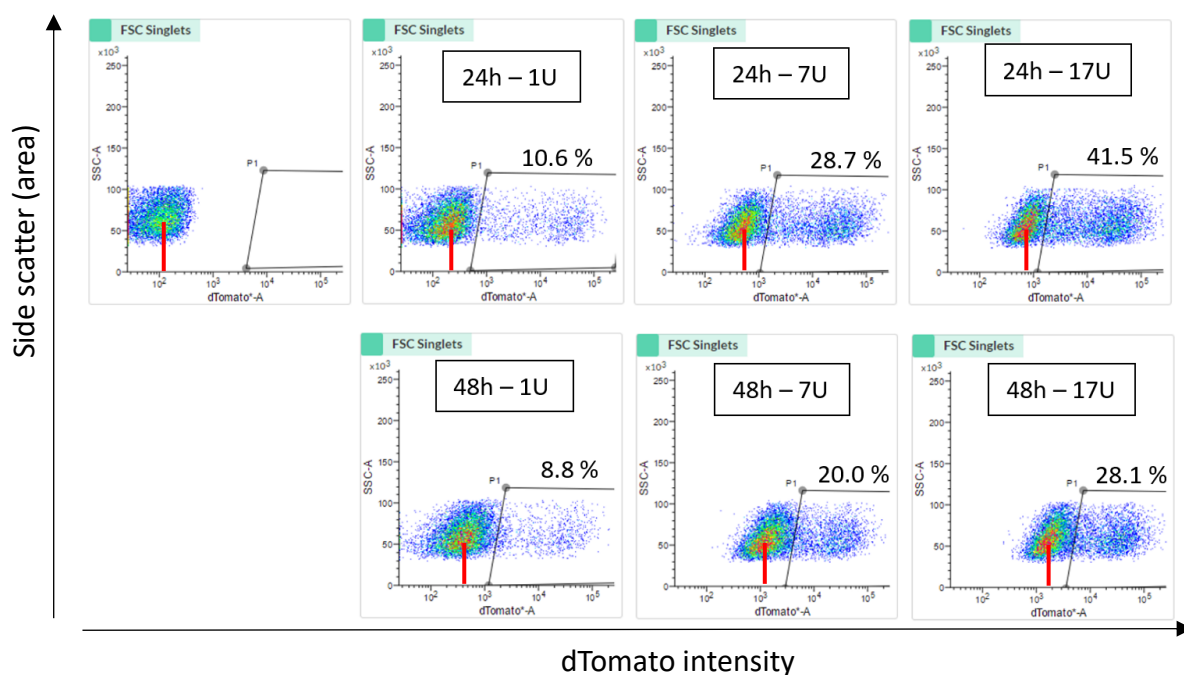


Figure 4.8: Transduction optimization.

Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells were transduced with increasing amounts (in units, U) of concentrated viral supernatant containing packaged template for dTomato orange fluorescent protein. Fluorescence was measured 48 hours after transduction using flow cytometry (FACSMelody, BD). 1U represents the number of viral particles produced by 2.33 cm² packaging cells in 24 hours. Viral supernatant produced in the first 24 hours after the change of medium was collected and concentrated (24 h) and replaced by fresh medium, which was again collected and concentrated 24 hours later (48 h). Numeric value indicates the percentage of dTomato-positive cells in each sample. Red line approximately indicates median fluorescence of the main population in the dTomato channel.

Interestingly, even the same amount of virus in concentrated form as was used for the early supernatant transduction attempts in the previous chapter (defined as 1 unit, 1 U, which was equal to the amount of virus that 2.33 cm² packaging cells can produce in 24 hours) gave better results in terms of percentage of transduced cells (10.6%, Figure 4.8) compared to the 1 – 2% efficiency observed previously in the transductions with unconcentrated virus (Figure 4.7 in Chapter 4.5 and data not shown). This increase could also be partly due to the extended incubation time allowed after the transduction of the sgRNA-bearing plasmids (24 hours in the previous experiment, 48 hours in this experiment). As expected, transducing cells at a higher multiplicity of infection (MOI) led to an increase in transduced cells; the correlation was roughly logarithmic. However, this increase in transduction efficiency seemed to have come at a price of increased stress to the cells, as inferred from the shift of the main population towards higher fluorescence intensity (red line, Figure 4.8), suggesting autofluorescence of untransduced cells. Nevertheless, the viability of the cells was comparable in all groups (data not shown).

The viral supernatant from the first 24 hours after transfection medium was changed on the packaging cells produced better results than its counterpart collected at 24 hours after the first batch of viral supernatant was harvested (48 hours after the initial media change; Figure 4.8). This would have been either due to a higher content of viral particles, or a higher potency thereof.

Based on the results in this chapter, we decided to use concentrated viral stocks for future experiments, made from supernatants collected 24 hours after transfection media change on the packaging cell line.

4.7 Optimized transduction fails to solve the massive cell death problem

With the new, efficient twist on the transduction method in hand, we set off to assess whether the massive dying observed in the early experiments in all groups, including negative controls, could be avoided by increased transduction efficiency. In the next set of experiments, which were again a CRISPR/Cas9-based validation screen of the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line, I used 9 U of concentrated viral particles containing the various anti-*TAL1* sgRNAs (Figure 4.2 and Table 4.1 in Chapter 4.2) to transduce the newly established cell line, instead of 1 U unconcentrated viral particles as before. We chose to select for the transduced cells with antibiotic pressure (500 µg/ml hygromycin) rather than with FACS, because it presents a simpler strategy and is arguably less stressful for the cells.

In the first repeat of this experiment (Figure 4.9A), anti-*TAL1* sgRNA-edited cells (apart from TAL10-edited cells) showed an initial (day 4) drop in proliferation rate according to expectation, but by day 8, the differences between the groups became minimal. In the second repeat of this experiment (Figure 4.9B), the TAL7- and TAL10-edited samples were showing the expected drop in proliferation compared to controls on day 8, while the *TAL1* 5'-UTR-targeting and *TAL1* enhancer-targeting sgRNAs had an unexpected opposite effect of lending the edited cells a proliferative advantage over the control group. Moreover, the two negative controls (SCR1 and no sgRNA) acted quite differently to each other in the second repeat of the experiment. The general tendencies seemed grossly inconsistent between the two repeats.

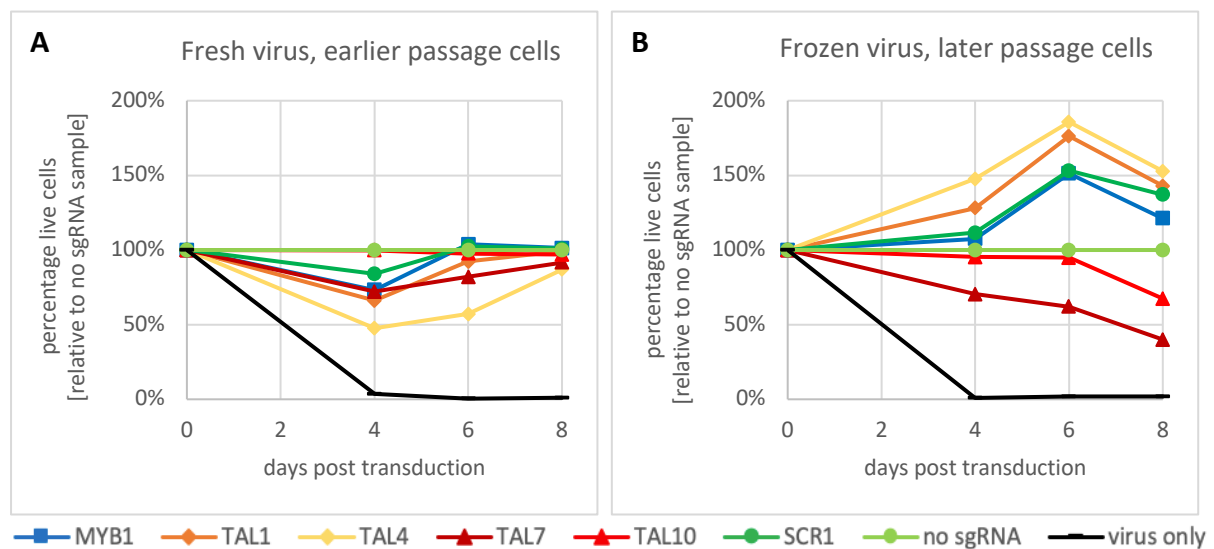


Figure 4.9: Proliferation rates of Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells transduced with concentrated viral particles containing various sgRNAs targeting *TAL1* gene and *TAL1* enhancer.

(A, B) Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 (clone 85-15-4) cells were transduced with 9 U of concentrated viral particles containing the lentiGuide-Hygro-dTomato plasmid encoding a hygromycin resistance cassette and either an sgRNA targeting the *TAL1* gene (TAL1, TAL4, TAL7 or TAL10), or an sgRNA targeting the MYB-binding site at the *TAL1* enhancer (MYB1), or a non-targeting, scrambled sgRNA (SCR1), or the lentiGuide-Hygro-dTomato plasmid with no sgRNA cloned in (no sgRNA). 500 µg/ml hygromycin was added to cells 24 hours post transduction. Viability, expressed as percentage of living cells relative to the no sgRNA sample, was measured at indicated time points using alamarBlue HS. The experiment in (A) was done using a fresh viral aliquot, whereas the experiment in (B) was done using a frozen viral aliquot; also the passaging was slightly different in each case, and the cells used for the experiment in (B) were about a week older than the cells used for the experiment in (A).

Furthermore, by day 8 of the experiment, again, none of the groups, including control cells, showed any signs of expansion into cultures, which was always the case by the end of week 1 in my previous

transduction experiments with Jurkat cells (Chapter 3), and were instead dying off. Curiously, however, cells that were transduced with empty viral particles (virus only), and served therefore as a positive control for antibiotic selection, were almost completely dead by day 4, in contrast to the cells in all the other groups, which received the hygromycin resistance gene during transduction (Figure 4.9A, B). This was consistent across the two repeats, which otherwise differed in almost every other respect, and indicated that the hygromycin resistance cassette from lentiGuide-Hygro-dTomato was, at least to a degree, successfully expressed and functional.

Taken together, the experiments in this chapter showed that the improved transduction method didn't solve the problem of extensive cell death in the CRISPR/Cas9-based validation screen in Jurkat-TRE3G-FLAG-TAL1-Cas9. Furthermore, the results indicated that the hygromycin resistance cassette was functional. These findings therefore strongly suggested that neither the low transduction efficiency, nor a faulty hygromycin resistance cassette were at the heart of the issue that was preventing us from a successful, viable CRISPR/Cas9 validation of the newly established cell line, and prompted us to look for a potential solution elsewhere.

4.8 siRNA *TAL1* knockdown slows proliferation in Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9

Because of such disheartening results from the CRISPR/Cas9-based validation of *TAL1* dependence in Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9, we decided to employ a different method for the validation of the cell line to see whether our CRISPR/Cas9 method needed further optimization, or whether the cell line was just not behaving in the expected manner. To this end, we chose to carry out an siRNA-mediated knockdown of *TAL1* and measure the effect on cell proliferation.

Because chemical transfection methods are generally grossly inefficient in Jurkat cells, we instead chose to employ our latest technological addition, a Neon Transfection System (Thermo Fisher), that transfects cells based on electroporation. I firstly trialled this system for the delivery of small oligonucleotides (such as siRNAs or ASOs) into Jurkat cells with striking success (Figure 4.10). A control Cy3-labelled antisense oligonucleotide (ASO) was introduced into 100% of cells in the sample when the concentration of the ASO in the cell culture was as little as 50 nM (Figure 4.10). Therefore, this method seemed promising for the purposes of siRNA introduction into a Jurkat-based cell line.

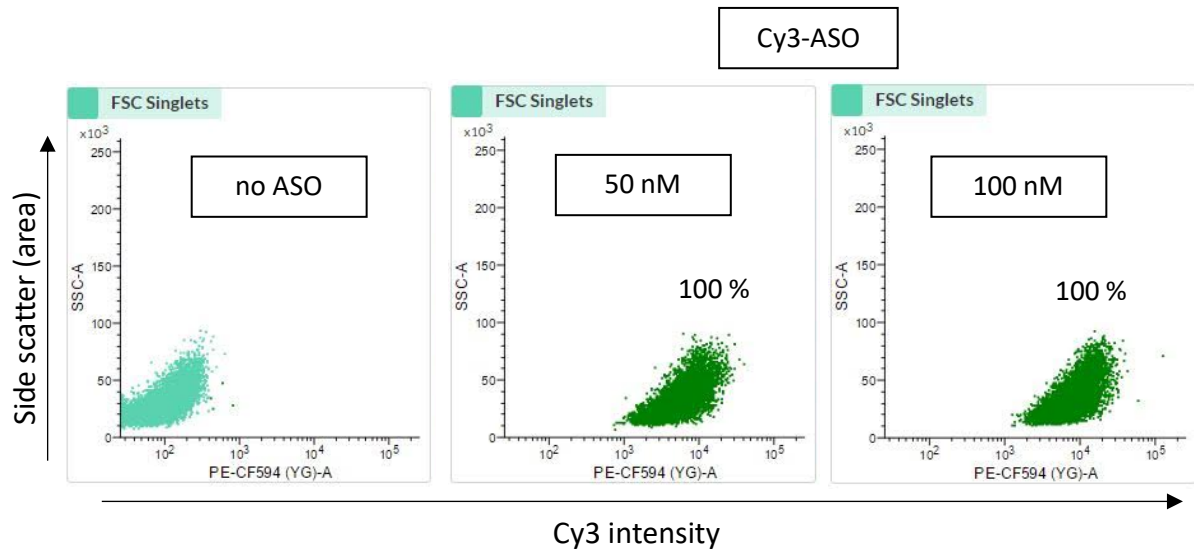


Figure 4.10: Nucleofection optimization for the delivery of small oligonucleotides.

Jurkat cells were electroporated using Neon Transfection System (Thermo Fisher) with control Cy3-labelled antisense oligonucleotide (Cy3-ASO). The final Cy3-ASO concentration in the culture medium was either 50 nM or 100 nM. Figure shows flow cytometry images from 10,000 cells, with Cy3 fluorescence intensity on the x-axis.

To confirm that *TAL1* knockdown decreases Jurkat cell proliferation in the engineered cell lines, anti-*TAL1* siRNAs of three kinds were delivered into Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells (clone 85-15-4) and viability was monitored in regular intervals using the alamarBlue HS Cell Viability Reagent (Figure 4.11B). As a control measure, the same experiment was also carried out in the maternal Jurkat cell line (Figure 4.11A). Downregulation in TAL1 expression was confirmed by a Western blot (Figure 4.11C). The proliferation of both the cell lines was negatively affected by the downregulation of TAL1 expression upon the siRNA-mediated *TAL1*-knockdown.

This set of experiments confirmed that both wild-type Jurkat cells and the newly established cell line, Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9, maintain their dependence on TAL1 for continued proliferation, a key characteristic for the planned CRISPR/Cas9 screen.

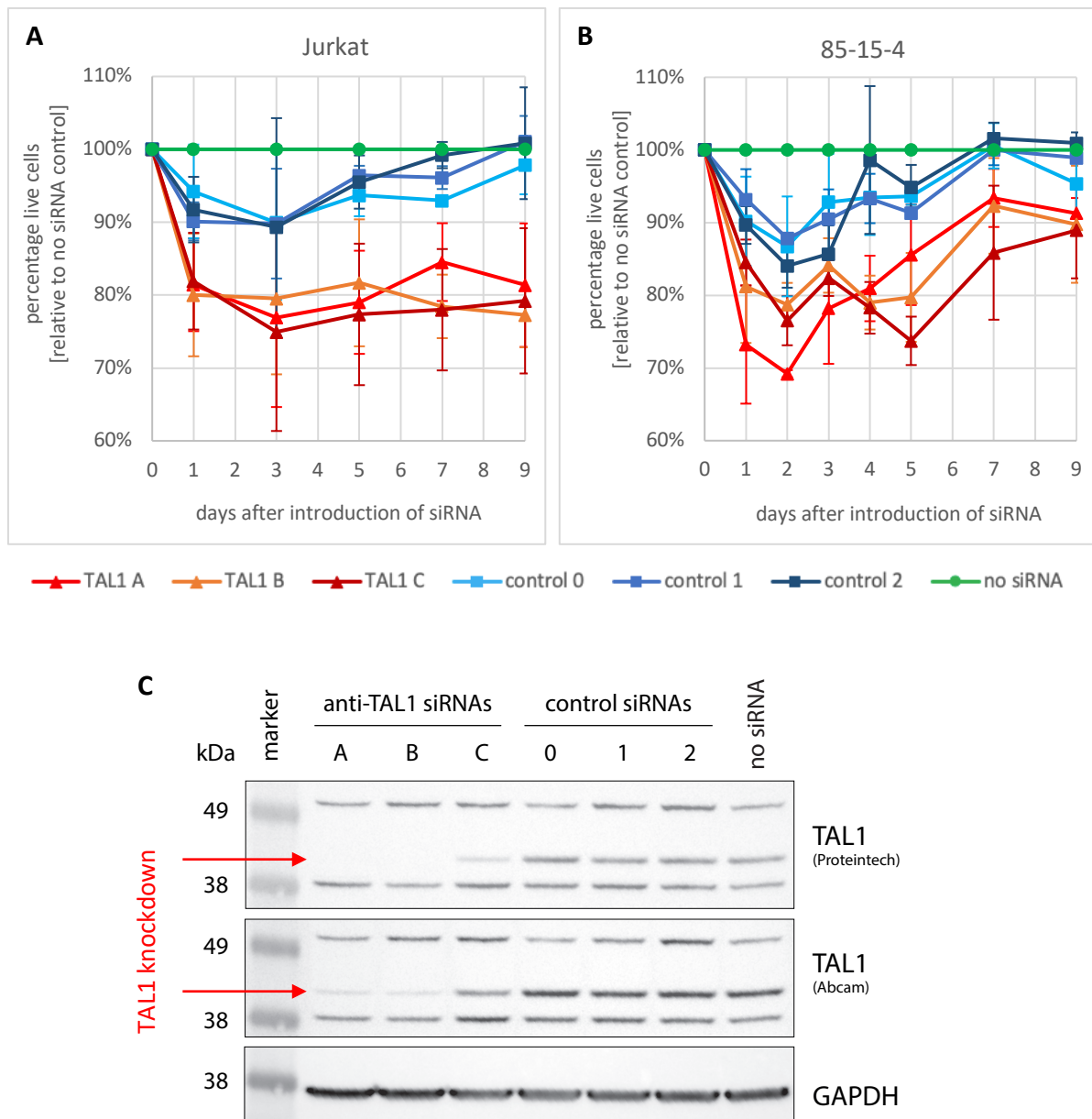


Figure 4.11: siRNA-mediated *TAL1* knockdown in Jurkat and Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells.

siRNAs (3 anti-*TAL1* siRNAs, 3 non-targeting controls) were delivered into cells on day 0 via nucleofection (Neon Transfection System, Thermo Fisher) to a final concentration of 50 nM. 'No siRNA' is a PBS-treated sample. **(A, B)** Percentage of living Jurkat **(A)** or Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 **(B)** cells after siRNA-mediated *TAL1* knockdown relative to the PBS-treated sample (no siRNA), as measured with alamarBlue HS Cell Viability Reagent. Error bars represent standard deviation from 3 biological replicates. **(C)** A Western blot shows TAL1 expression in each experimental sample of Jurkat cells on day 3 after transfection. GAPDH serves as loading control. Anti-TAL1 antibodies used were from Proteintech (55317-1-AP) and Abcam (ab155195), anti-GAPDH antibody was from Proteintech (60004-1-Ig).

4.9 Neon electroporation fails to solve the massive cell death problem and inconsistent behaviour in Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9

As the results from the *TAL1*-mediated knockdown experiments showed that the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line maintained its addiction to TAL1, the next step was to try to optimize the ill-fated CRISPR/Cas9 validation screen further. This was still important, because the planned mutagenesis of the *TAL1* enhancer is based on CRISPR technology, not on siRNA-mediated knockdown.

While not entirely sufficient for the validation of the cell line, the experiments with siRNAs lent us the idea of trying out Neon electroporation instead of transduction for the introduction of the sgRNA-bearing plasmid. The main advantages of this approach in this particular situation were considered the high transfection efficiency (for example, 94% transfection efficiency 24 hours after electroporation with 10 µg of an EGFP-encoding plasmid, Neon transfection system protocol for Jurkat cells on www.invitrogen.com, accessed 02/03/2022) and the easy scalability of the method. More substantial samples with a large percentage of transfected cells were hoped to provide consistency across individual repeats and a better starting point for long-term survival. Furthermore, in the transduction method, a higher transduction efficiency requires a higher number of viral particles, which can be somewhat stressful to the cells (Figure 4.8 in Chapter 4.6), and in combination with other stressors (antibiotic selection, FACS) can conceivably lead to increased cytotoxicity. In Neon electroporation, however, there is no reason why transfection efficiency and toxicity should be inherently correlated. According to the Neon Transfection System protocol for Jurkat cells, the viability of Jurkat cells electroporated with 10 µg of an EGFP-encoding plasmid was close to 98% at 24 hours after the Neon electroporation (www.invitrogen.com, accessed 02/03/2022). Finally, as an additional benefit, Neon transfection is much less onerous than transduction, saves time and cuts on a number of experimental variables.

In a proof-of-principle experiment, two differently sized GFP-coding plasmids, 4.7 kb and 9.3 kb, were introduced into Jurkat cells using the Neon Transfection System (Thermo Fisher; Figure 4.12A, B). The transfection efficiency was positively correlated with the amount of DNA used and negatively correlated with the size of the plasmid, with a maximum efficiency of 72.7% achieved for the 4.7 kb plasmid and 38.3% achieved for the 9.3 kb plasmid. In both cases, this far exceeded the 1.6% efficiency achieved in our first transduction experiments with unconcentrated viral particles (Figure 4.7 in Chapter 4.5). The transfection efficiency achieved with 15 µg of the 9.3 kb plasmid, which is close to

the size of our experimental lentiGuide-Hygro-dTomato plasmids (9.5 kb), was comparable to the efficiency achieved with the highest tested MOI in our transduction optimization experiment using concentrated virus (Figure 4.8 in Chapter 4.6) – but while the cells transduced with 17 U of viral particles showed signs of intense stress in the form of a shift in their autofluorescence values, the cells transfected with the Neon Transfection System seemed completely unperturbed (compare red lines in Figure 4.8 in Chapter 4.6 and Figure 4.12). I therefore decided to use Neon electroporation to transfect constructs for my validation experiments.

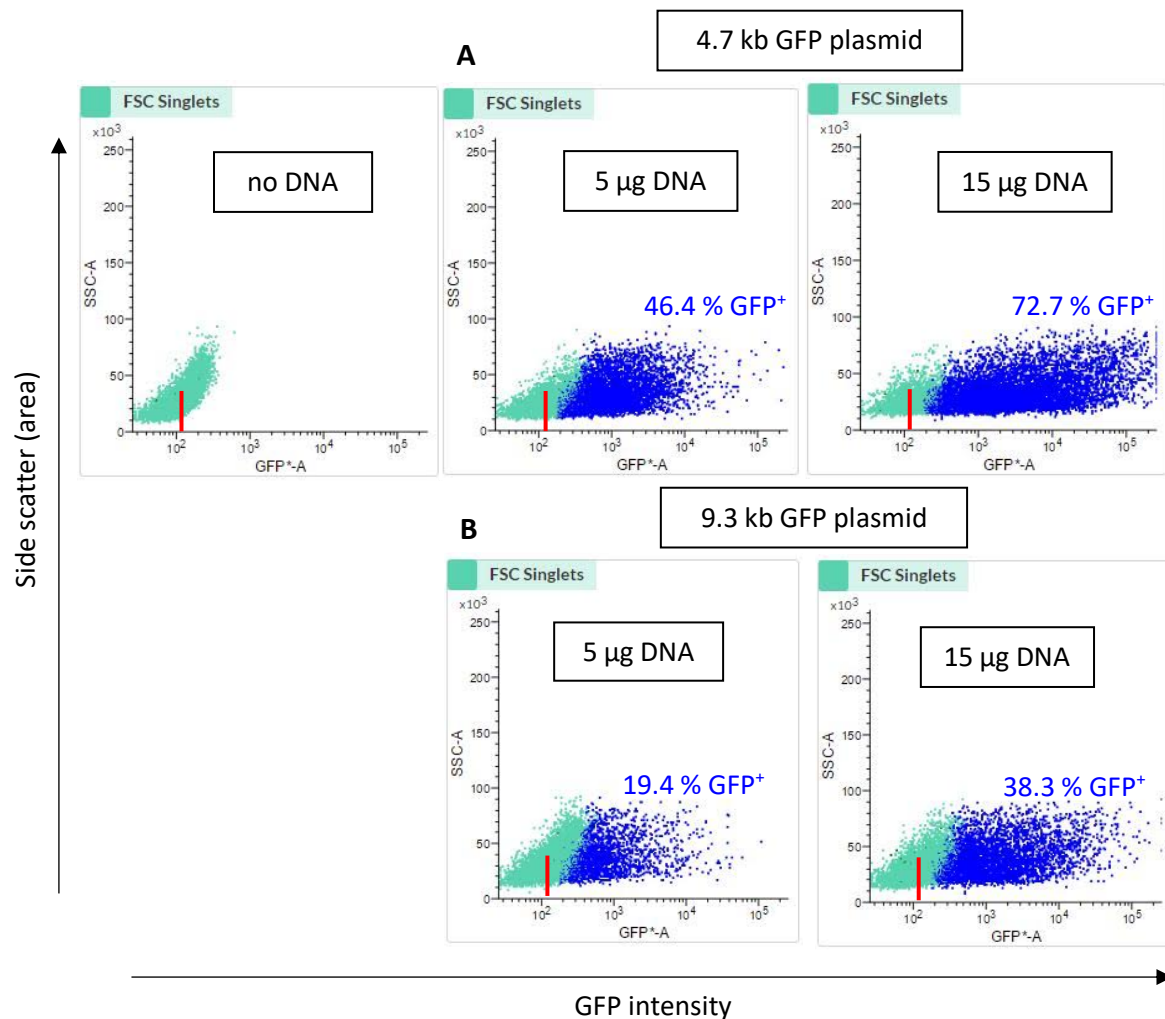


Figure 4.12: Nucleofection optimization.

Jurkat cells were electroporated using Neon Transfection System (Thermo Fisher) with **(A)** a 4.7 kb plasmid containing GFP (pAcGFP1-C1, Clontech); **(B)** a 9.3 kb plasmid containing GFP (pSpCas9(BB)-2A-GFP (PX458), Addgene plasmid #48138, (Ran et al., 2013)). Percentage of transfected cells was assessed 48 hours after transfection using flow cytometry. Results are from a single replicate. Red line approximately indicates median fluorescence of the untransfected population in the GFP channel.

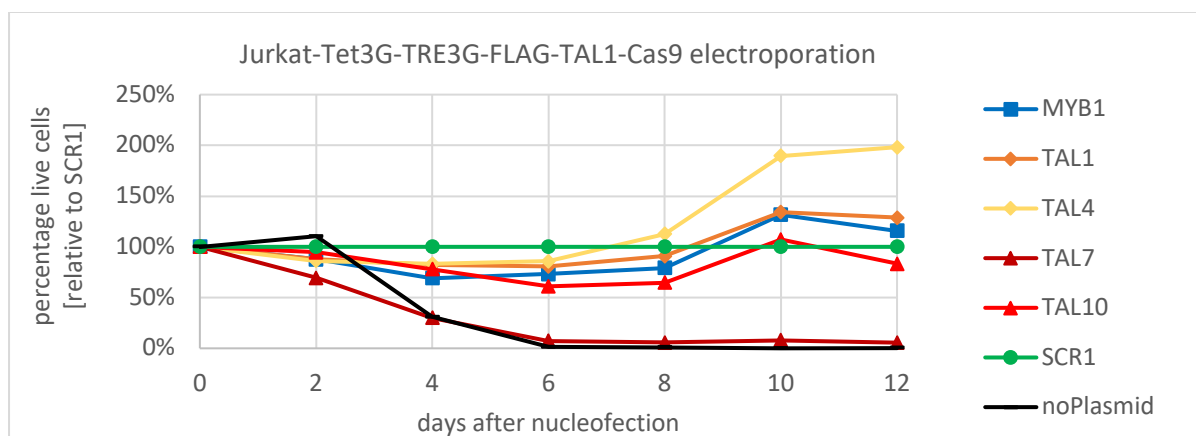


Figure 4.13: Proliferation rates of Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells transfected with various sgRNAs targeting *TAL1* gene and *TAL1* enhancer.

Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 (clone 85-15-4) cells were transfected by electroporation with the lentiGuide-Hygro-dTomato plasmid encoding a hygromycin resistance cassette and either an sgRNA targeting the *TAL1* gene (TAL1, TAL4, TAL7 or TAL10), or an sgRNA targeting the MYB-binding site at *TAL1* enhancer (MYB1), or a scrambled sgRNA (SCR1) as a negative control. 500 µg/ml hygromycin was added to cells 24 hours post transfection. Viability, expressed as percentage of living cells relative to the SCR1 sample, was measured every 2 days using alamarBlue HS. These results are from a single replicate.

I proceeded to carry out the CRISPR validation experiment in Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9, using the Neon electroporation method and the anti-*TAL1* sgRNA-containing plasmids described earlier. While these plasmids are primarily designed for transduction, we were unable to conceive of any reason why they shouldn't work for transfection as well as transduction. Cells were electroporated and selection with 500 µg/ml hygromycin was applied after 24 h. The results from the first part of this experiment looked promising, with a number of surviving cells and the TAL1-challenging sgRNAs having the expected effect of decrease in proliferation rate (Figure 4.13). However, about a week into the experiment, the proliferation rate of the cells bearing the experimental sgRNAs began to surpass the proliferation rate of the non-targeting control sgRNA (SCR1), in all cases except for the TAL7-edited sample, which had virtually died off after the first week. This occurrence was similar to the results seen in one of the previous transduction experiments (Figure 4.9B in Chapter 4.7). However, in contrast to the transduction experiment, in which this phenomenon started only 4 days after hygromycin had been added, here the change in proliferation patterns started occurring 8 days following the addition of hygromycin.

In addition to the strange rebound of the experimental proliferation rates from day 8 onwards, there were again no visible signs that the cells were expanding into cell lines. The viability was only assessed in relative terms (with alamarBlue HS), but an expanding cell culture can be easily identified under the microscope, and by a colour change in the growth media. While the bigger initial size of the sample for Neon electroporation (2×10^6 cells, compared to 1.6×10^5 cells used for my previous transduction experiments) helped to maintain a pool of live cells for a slightly longer time compared to the transduction experiments, none of the Neon-transfected cell cultures were expanding by day 12 post electroporation.

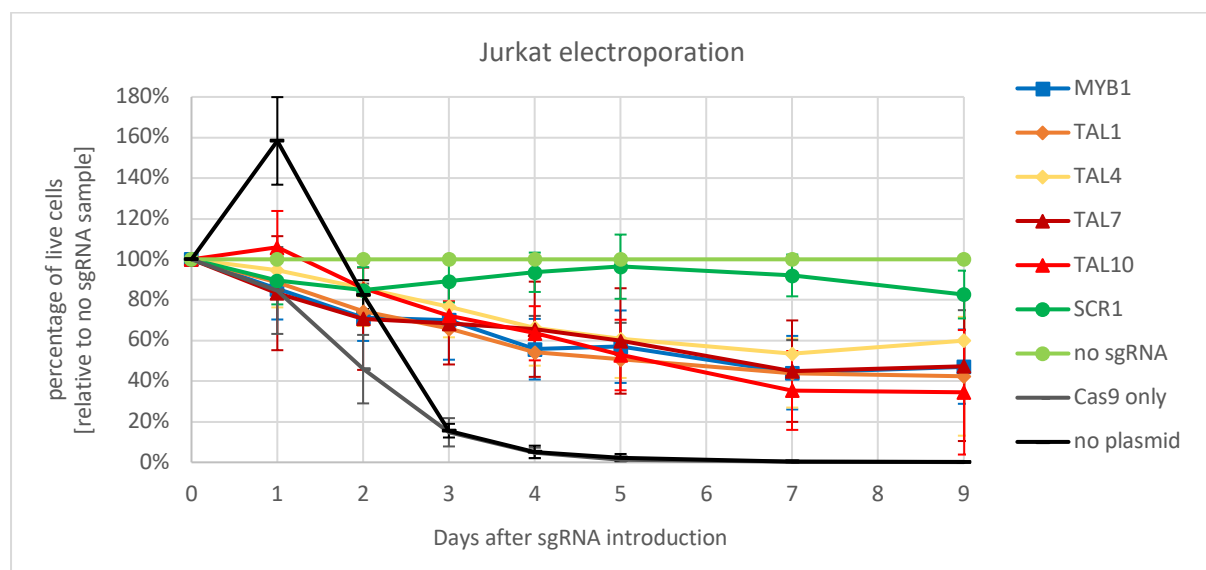


Figure 4.14: Wild-type Jurkat cells are met with a growth challenge upon *TAL1* knockout.

Jurkat cells were co-electroporated with lentiCas9-Blast (Addgene plasmid #52962, (Sanjana et al., 2014)) coding for Cas9 (blasticidin resistance) and the lentiGuide-Hygro-dTomato plasmid (hygromycin resistance) containing an sgRNA against one of the following regions: *TAL1* enhancer (MYB1), *TAL1* 5'-untranslated region (TAL1, TAL4) or *TAL1* exonic region (TAL7, TAL10). Negative controls for the experiment were a plasmid with a non-targeting sgRNA (SCR1) and a plasmid with no sgRNA (no sgRNA). Positive controls for selection pressure were cells mock-transfected with PBS (no plasmid), or transfected only with the Cas9 plasmid (Cas9 only). Cells were cultured in 10 µg/ml blasticidin and 500 µg/ml hygromycin from day 1 post electroporation to select for cells that contained both the plasmids. Graph shows cell viability in each sample, as measured with alamarBlue HS Cell Viability Reagent, and normalized to no sgRNA sample viability. Error bars represent standard deviation from 3 biological replicates.

To investigate whether this behaviour was specific to Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9, the same experiment was carried out in the maternal Jurkat cell line (Figure 4.14). Here, the behaviour of the

experimental samples was much more in keeping with our expectations. All targeting sgRNAs consistently caused a reduction in growth compared to control sgRNAs; the numbers of cells in the experimental samples were falling steadily from the beginning of the experiment and reached down to 35 – 60% of no sgRNA control by day 9. The cell numbers were visibly high throughout the course of each repeat of this experiment. In short, the CRISPR/Cas9 validation experiment, as designed for the validation of the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line, seemed to be perfectly functional in the maternal Jurkat line.

In the maternal Jurkat cell experiment, the TAL10 sgRNA elicited the most profound decrease in proliferation rate, and the no sgRNA control performed better than the SCR1 control. To reduce the complexity of subsequent validation experiments, the experimental plasmid bearing TAL10-sgRNA and the control plasmid bearing no sgRNA were selected for further experiments on the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line.

In sum, while using Neon electroporation instead of transduction enabled us to easily increase the number of starting cells in the CRISPR/Cas9-based validation experiment, the approach did not save any of the transfected Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell groups, control or experimental, from ultimately succumbing to death. Moreover, strange and unexpected rebound patterns were observed in the proliferation of the anti-TAL1 sgRNA-edited cell samples. On the other hand, the CRISPR/Cas9 validation experiment appeared to be fully functional in wild-type Jurkat cells, both in terms of expected outcomes and cell survival. Taken together, these findings indicated that the protocol for the CRISPR/Cas9-based validation experiment required further tweaking for the use with the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line.

4.10 Exploring the role of hygromycin in the massive cell death problem

As the CRISPR/Cas9 validation screen proved to be functional in the maternal Jurkat cell line, and the TAL1 dependence of the newly established Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line was confirmed by siRNA-mediated knockdown, it seemed to logically follow that the CRISPR/Cas9 validation screen should also work in the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line. However, with Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9, there was the problem of the massive cell death following application of hygromycin, and the problem of unexpected and often inconsistent results, with a potential connection between the two issues. Taken together, the observations we had made up to this point seemed to indicate that further adjustments were needed in the CRISPR/Cas9 validation

screen protocol for the use with the newly established cell line. And to find the right adjustments to be made, we looked for the differences between the maternal cell line and the genetically modified one.

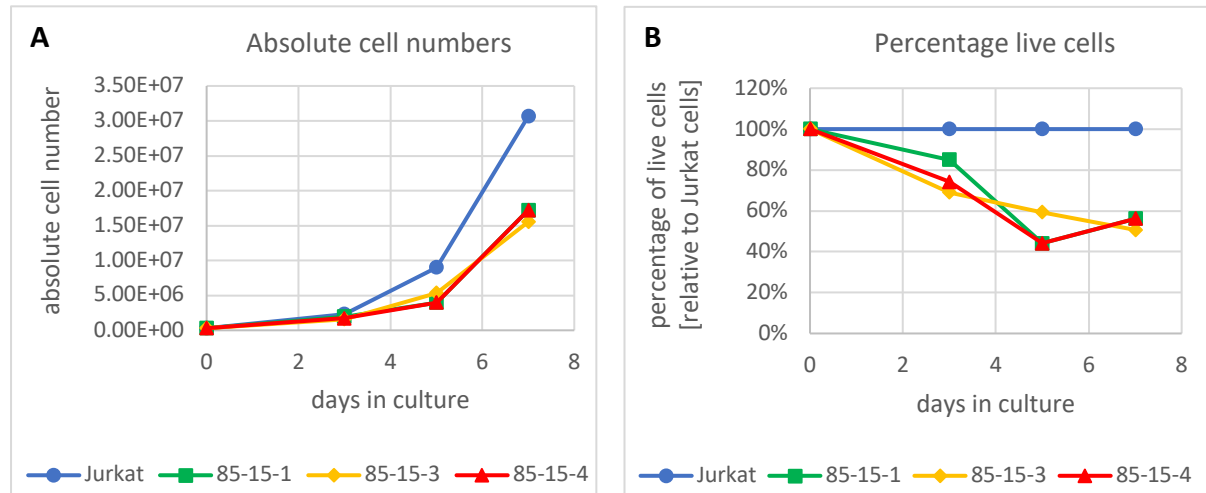


Figure 4.15: Proliferation of all the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 monoclonal cell lines compared to the proliferation of the maternal Jurkat cells.

The same set of data is expressed as **(A)** absolute cell numbers, **(B)** percentage of living cells relative to maternal Jurkat. Cells were counted on a haemocytometer using Trypan Blue; graphs show data from a single replicate.

Perhaps the most perceptible difference between the maternal Jurkat cell line and the newly established Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 clonal lines is one of overall fitness. Unchallenged Jurkat cells grow faster than any of the final clonal Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 lines (Figure 4.15); they are also morphologically healthier and more viable. It is a likely speculation that as a result, they would also be more resilient and robust. This led us to hypothesize that gentler conditions were perhaps needed throughout the course of the CRISPR/Cas9 validation experiment in order for the genetically modified cell lines to survive the challenges the screening conditions posed.

As the first port of call, hygromycin concentrations were taken into question. We hypothesized that if the newly established lines are less resilient than the maternal cells, perhaps their sensitivity to antibiotics has increased. In order to test whether the 85-15-4 line that we had used until now was uniquely susceptible to hygromycin, we compared all the newly established monoclonal cell lines in a minimal lethal dose experiment, where Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 along with the maternal Jurkat cells were cultured with a range of hygromycin concentrations for 7 days (Figure 4.16). Notably, all the cell lines seemed to perform with high comparability in the hygromycin challenge. Even so, the

experiment served as an invitation to consider if a lower concentration of hygromycin could possibly be used.

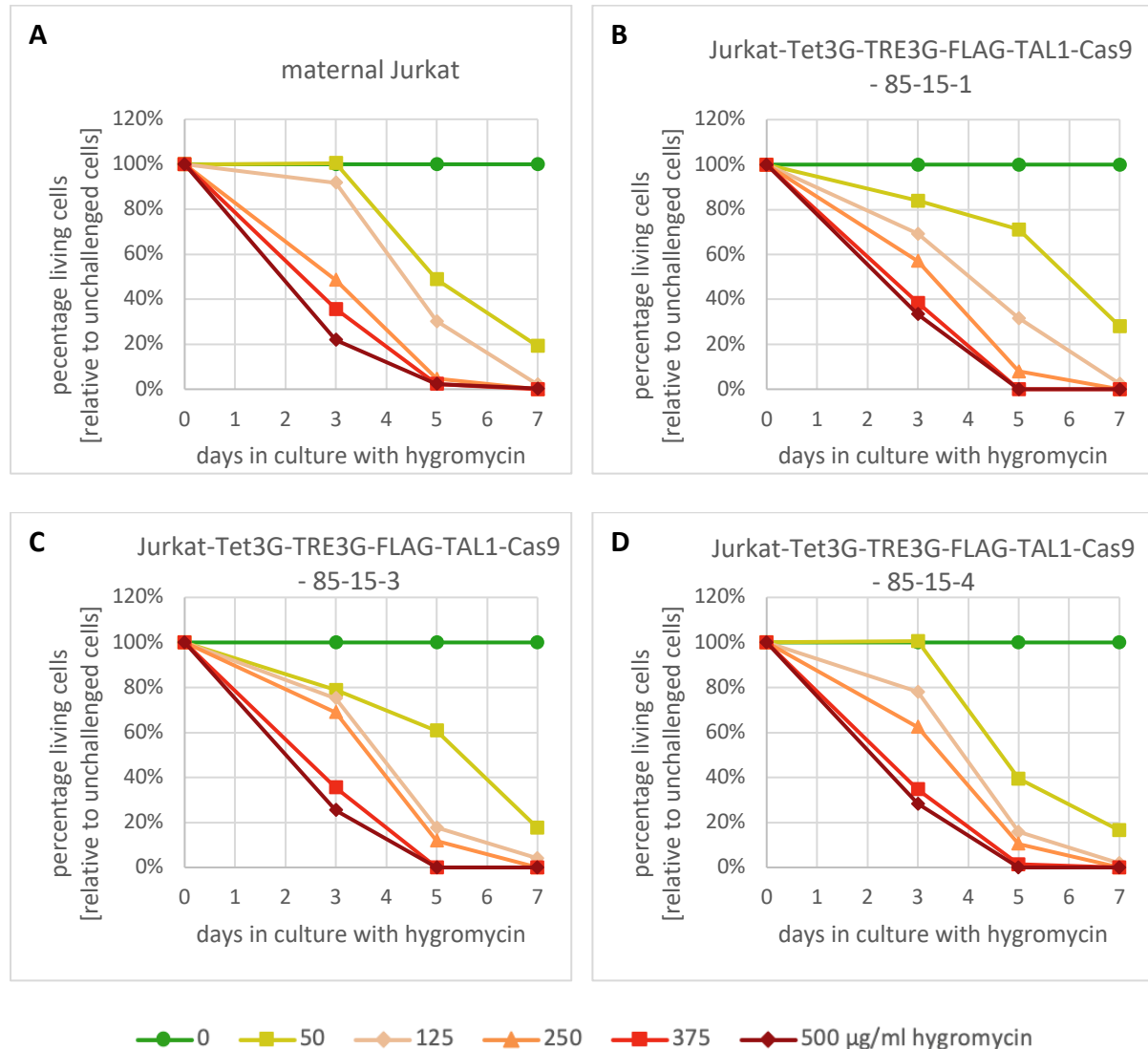


Figure 4.16: Proliferation of the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 clones with a range of hygromycin concentrations.

Graphs show proliferation rates of the maternal Jurkat cell line (**A**) and the three Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 clones (**B – D**) when cultured with 0 – 500 µg/ml hygromycin. Cells were counted on a haemocytometer using Trypan Blue; graphs show data from a single replicate.

Cells in this experiment were passaged on day 3, 5 and 7, in contrast to the Jurkat cells in the earlier minimal lethal dose experiment, which were left to proliferate without any passaging (at the time, only a very approximate result was required). The results from the new experiment were just a little

different from the results of the earlier experiment (Figure 4.5 in Chapter 4.4). 500 µg/ml and 375 µg/ml hygromycin killed virtually all cells by day 5. After 7 days of culture, there were no living cells of any cell line in the 250 µg/ml hygromycin group. The 125 µg/ml hygromycin group allowed the survival of a very few cells by the end of the week. Therefore, lower concentrations of hygromycin than 500 µg/ml can be used for selection if we are mindful of the varying timeframes.

To test the influence of the hygromycin concentration on the outcome of the CRISPR/Cas9 validation experiment, Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 (clone 85-15-4) and Jurkat cells were electroporated with either the lentiGuide-Hygro-dTomato-TAL10-sgRNA plasmid targeting the *TAL1* gene, or the plasmid without an sgRNA (no sgRNA), with PBS used as a positive control for hygromycin resistance. Different hygromycin concentrations were added at either 24 or 48 hours post transfection (Figure 4.17). The two different time points for adding the selection antibiotic were another variable to investigate: if it takes a longer time for the expression of the hygromycin resistance cassette to start, adding the antibiotic at a later time point might potentially affect the survival rates beneficially. The relative viability of the cells in each group was assessed using alamarBlue HS, but absolute numbers were also monitored to determine whether the cells were expanding into viable cultures.

In wild-type Jurkat cells, according to expectation, the *TAL1*-knockout cells (TAL10) showed a pronounced decrease in proliferation relative to the negative control (no sgRNA) transfected cells. This trend was independent of the concentration of hygromycin or the time point that hygromycin was added (Figure 4.17A – E). This was not the case in Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9, however. Strange patterns emerged with this cell line. Under every tested hygromycin condition, *TAL1*-knockout cells were outgrowing the control by day 11 (Figure 4.17F – J). The trend was more pronounced at higher hygromycin concentrations, regardless of the time of hygromycin addition.

Importantly, absolute numbers (calculated from the alamarBlue HS viability data using a standard curve, data not shown) of both Jurkat and Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells in all groups, control and experimental, steadily declined after a short initial expansion (day 0 to day 3) and a plateau (day 3 to day 5) until the experiment was ended on day 11. While this would be an acceptable observation for the *TAL1*-knockout cells, in keeping with the Jurkat cell addiction to TAL1, the expansion of the control cells in this experiment is absolutely crucial – in our planned screen, we need to be able to collect a substantial sample of edited cells 2 to 3 weeks after the introduction of the sgRNA library.

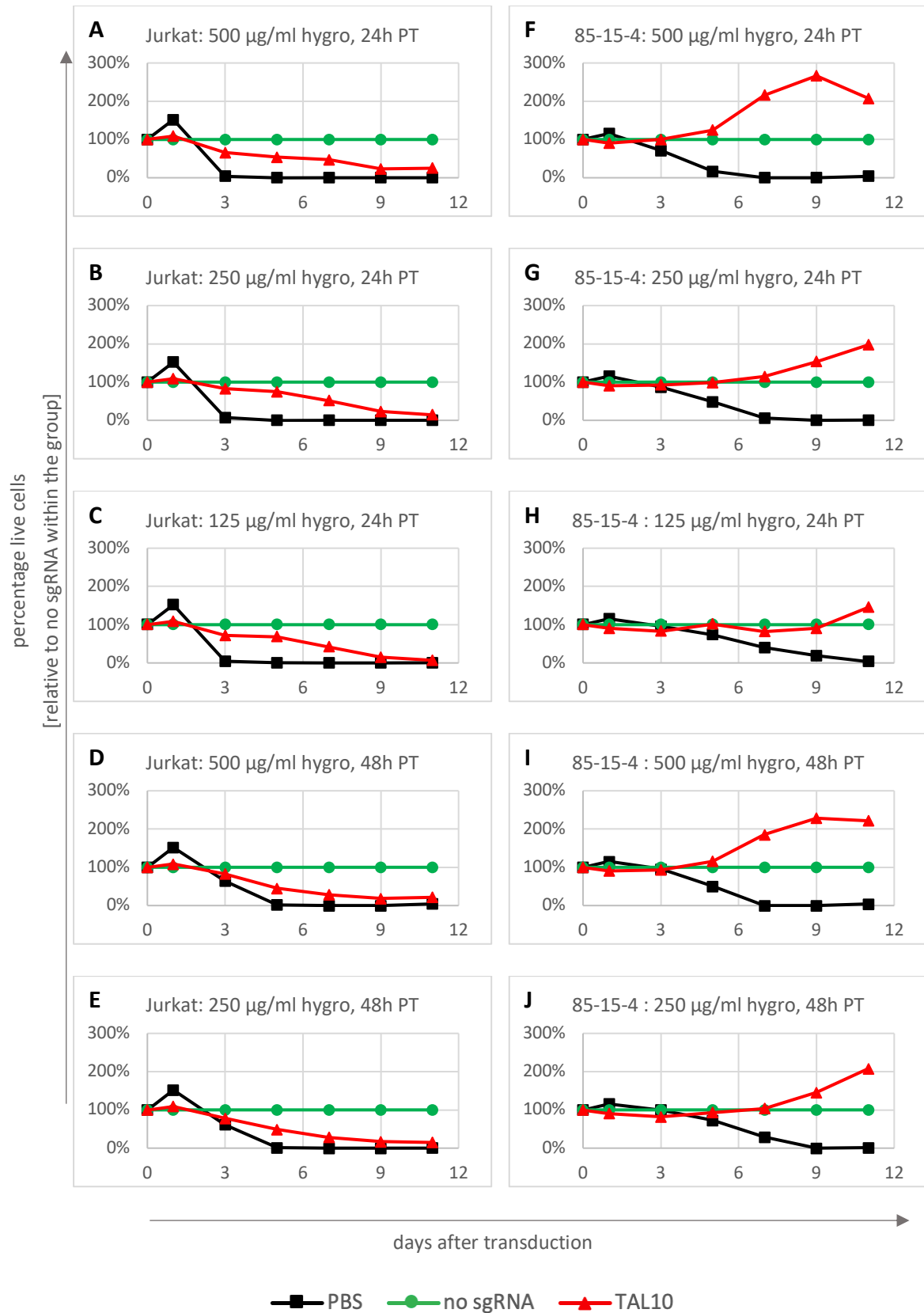


Figure 4.17: Influence of hygromycin concentration on proliferation of *TAL1*-edited Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells and *TAL1*-edited maternal Jurkat cells.

Both maternal Jurkat cells (**A – E**) and Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 (clone 85-15-4; **F – J**) were transfected by electroporation with the lentiGuide-Hygro-dTomato plasmid encoding a hygromycin

resistance cassette and either an sgRNA targeting the *TAL1* gene (TAL10), or no sgRNA (no sgRNA) as negative control. The maternal line was co-transfected with the lentiCas9-Blast plasmid coding for Cas9 and a blasticidin resistance marker. Hygromycin (hygro) was added to cells either 24 hours post transfection (24h PT; **A – C** and **F – H**) or 48 hours post transfection (48h PT; **D, E, I, J**) to different concentrations: 500 µg/ml (**A, D, F, I**); 250 µg/ml (**B, E, G, J**) or 125 µg/ml (**C, H**). The maternal line (**A – E**) was also grown in 10 µg/ml blasticidin for the selection of Cas9-positive cells. Viability was measured every 2 days using alamarBlue HS reagent and is expressed as percentage of viable cells in the no sgRNA sample within each group. These results are from a single replicate of a single experiment.

The steady decline in cell numbers in the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 control (no sgRNA) group confirmed previous observations, but in the control Jurkat cells was unexpected, at odds with the previous observation. However, this previous observation was made without absolute cell counting; Jurkat cells proliferate faster than the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells (Figure 4.15), and could therefore have reached higher cell numbers throughout the early-stage expansion (day 0 to 3) in the CRISPR/Cas9 validation experiment, which could then potentially have obscured the trend of the steady decline in cell numbers.

In summary, while the tolerance to hygromycin was extremely alike in Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 clones and the maternal Jurkat cells, changes in hygromycin concentration in the CRISPR/Cas9 validation experiment caused a pronounced difference in the proliferation patterns of the edited Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells, but not the edited maternal Jurkat cells. Furthermore, both the cell lines failed to show any signs of expansion over the course of 11 days under any of the tested conditions. We had to look for other factors to adjust.

4.11 Understanding the influence of cell density on the survival and proliferation rates in Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9, and reconsidering the experimental time frame

The Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line grows relatively well when unchallenged, if somewhat slower than its maternal counterpart. Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells are commonly passaged every 3 – 4 days, often only by discarding a portion of the cell suspension and adding fresh medium in a desired split ratio. However, in the CRISPR/Cas9 validation experiments until

this point, cells were often left in the same medium for more than 4 days at a time, similarly to when clones are being expanded into cell lines.

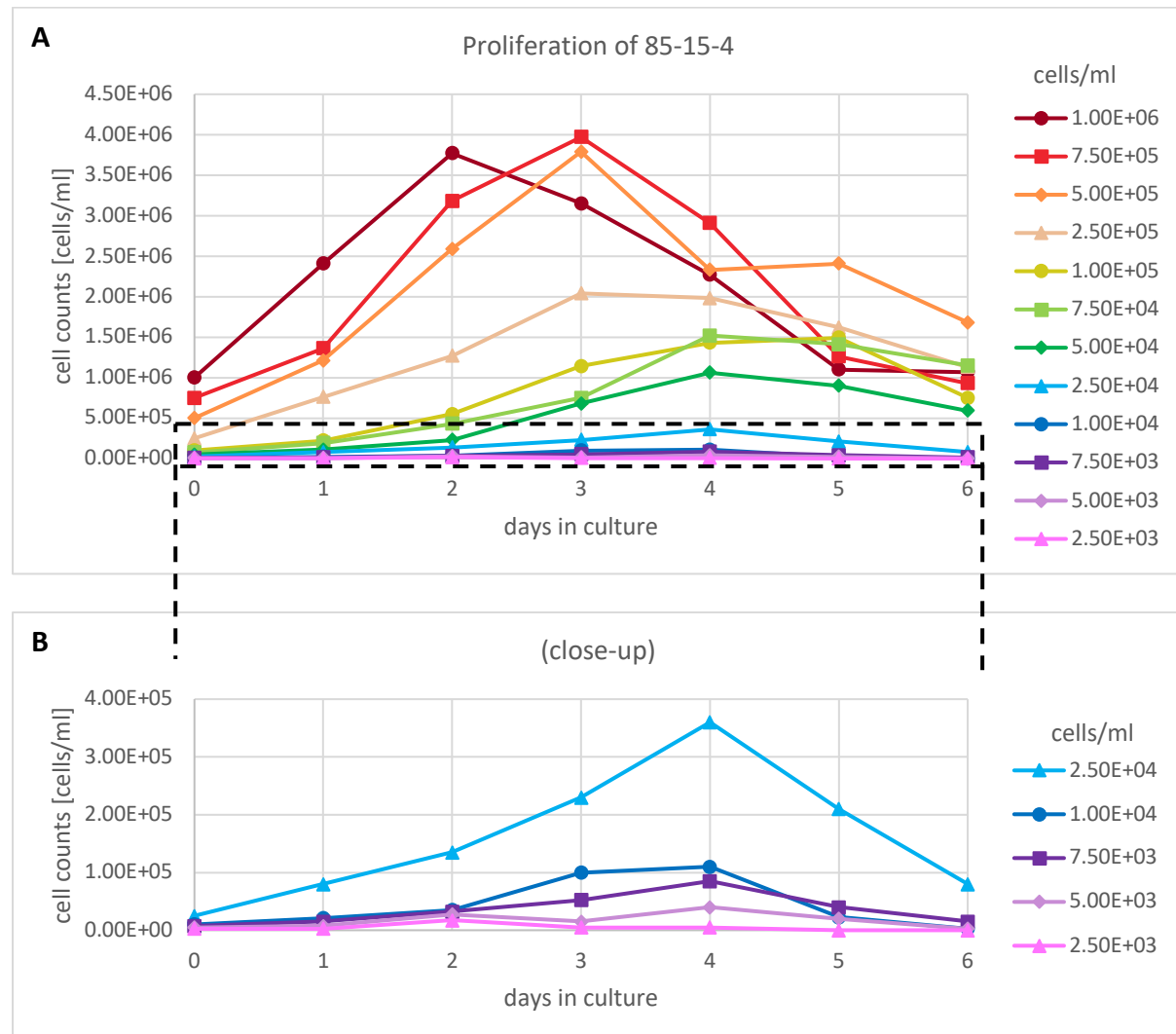


Figure 4.18: Proliferation patterns of unchallenged Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells (clone 85-15-4).

Cell culture was started on day 0 in a range of cell densities. Cells were not passaged, but mixed and counted every 24 hours on a haemocytometer using Trypan Blue stain. **(A)** All concentrations. **(B)** is a close-up on the lower starting concentration curves, area as designated by the dashed box in **(A)**. Results are from a single replicate.

To investigate whether passaging conditions might be a contributing factor in the massive cell death issue, I analyzed the proliferation patterns of Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 (clone 85-15-4) by growing these cells from a range of starting cell densities, without passaging, and by counting the cells

every day for 6 days (Figure 4.18A and B). In this experiment, cultures as sparsely populated as 2.5×10^3 cells/ml and as dense as 1.0×10^6 cells/ml made a healthy start. After 2 days, cell density in the two most populated samples reached a plateau (this was around 4×10^6 cells/ml) and started plummeting. Cultures that started at cell densities between 1.0×10^5 and 5.0×10^5 cells/ml proliferated roughly exponentially for 3 days and cultures that started at 2.5×10^4 to 7.5×10^4 cells/ml proliferated roughly exponentially for 4 days before the cell numbers started to drop. There was a mixed success in the expansions of the low density samples ($\leq 1.0 \times 10^4$ cells/ml; Figure 4.18B). Importantly, we learned that while the range of cell densities for a healthy Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 culture is broad, the cells seem to require a change of media at least every 4 days, even when the media is not obviously spent (change of colour due to the pH indicator).

To put the new findings into practice, a CRISPR/Cas9 validation experiment was conceived where the medium was to be changed at least once in 4 days and the cell density was to be adjusted every 2 days to between 3.5 and 5.5×10^5 cells/ml. The choice of cell densities in the higher range of the exponential growth window was designed to further assist the survival of the stressed cells. To further relieve the cells, 250 $\mu\text{g/ml}$ hygromycin was used instead of the previously employed 500 $\mu\text{g/ml}$, in accordance with the results in the previous chapter (Figure 4.17 in Chapter 4.10).

As another adjustment, after reviewing the results from the comparative experiments between Jurkat cells and Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9, blasticidin was to be added not only to Jurkat cells, but also to Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9, which were previously not grown in blasticidin because the gene for Cas9 was already incorporated into their genome. This incorporation should technically be stable, but it is conceivable that the newly established cell line, less robust than its maternal counterpart, could be losing the *Cas9* gene under the pressures of the validation experiment. To prevent this potential loss of Cas9, blasticidin, which was the selection marker for Cas9 during the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line establishment, was supplemented into the hygromycin selection media in the CRISPR/Cas9 validation experiment at a maintenance concentration of 5 $\mu\text{g/ml}$.

In this new tweak on the CRISPR/Cas9 validation screen experiment, neither of the cell lines was again expanding by day 11 of the experiment (Figure 4.19A, B). However, when cells were checked on day 17 after transduction, after having been left to grow without passaging or cell density adjustments (as a final hope), both the no sgRNA control and the TAL10-edited Jurkat cells showed signs of expansion, and grew into healthy cultures in the following days (Figure 4.19A). None of the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell groups started expanding by day 21, when the experiment was ended (Figure 4.19B).

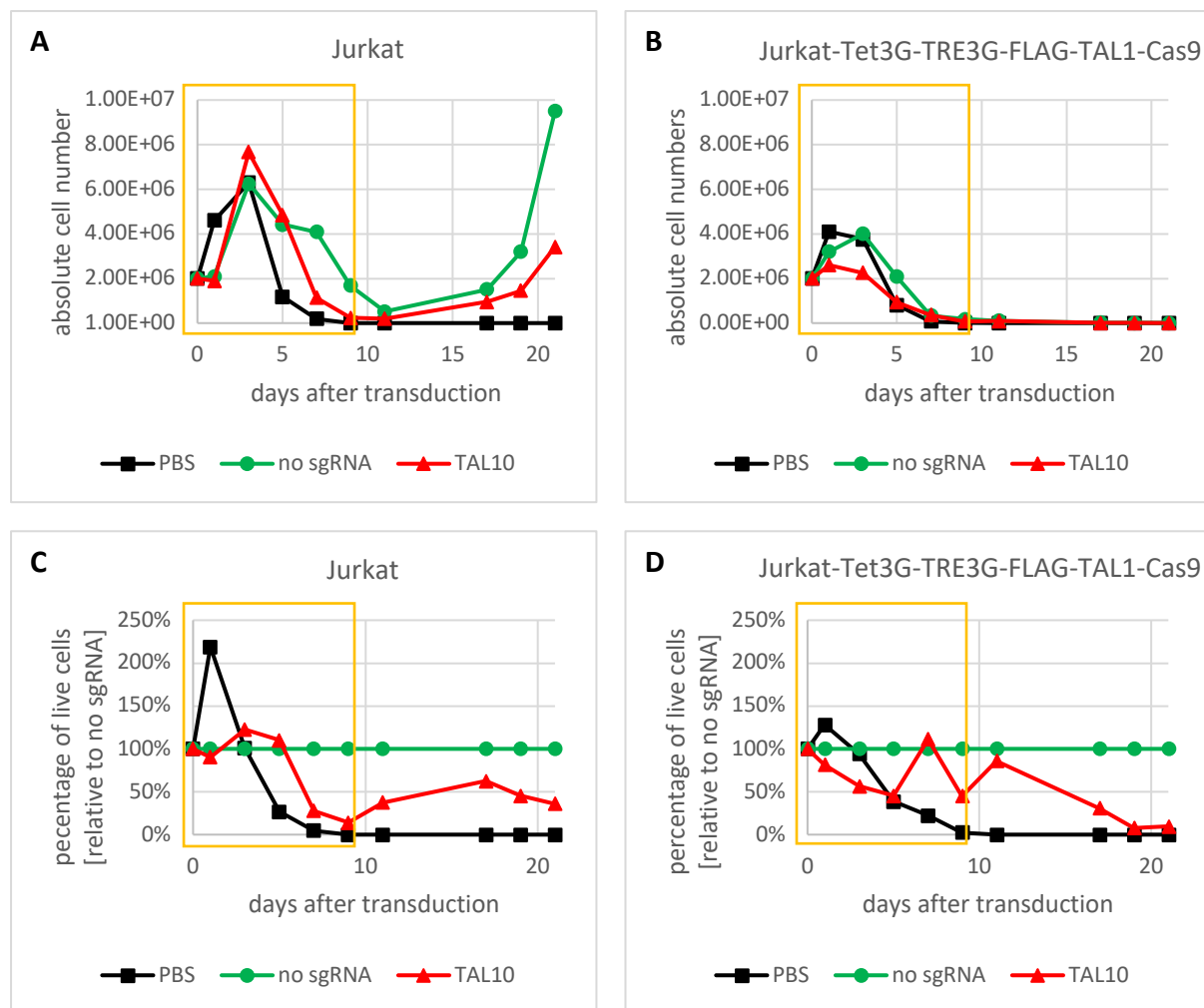


Figure 4.19: The influence of maintaining an optimal cell density on proliferation of *TAL1*-edited Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells and *TAL1*-edited maternal Jurkat cells, along with blasticidin supplementation.

Maternal Jurkat cells (**A**, **C**) and Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 (clone 85-15-4; **B**, **D**) were transfected by electroporation with the lentiGuide-Hygro-dTomato plasmid encoding a hygromycin resistance cassette and either an sgRNA targeting the *TAL1* gene (TAL10), or no sgRNA (no sgRNA) as negative control. Positive control for antibiotic selection was a PBS mock-transfected sample. The maternal line was co-transfected with the lentiCas9-Blast plasmid coding for Cas9 and a blasticidin resistance marker. Cells were kept in 250 µg/ml hygromycin and 5 µg/ml blasticidin from 24 hours post transduction. Cell density was adjusted every 2 days throughout the initial – selection – phase (until day 9, **yellow rectangle in graphs**) to between 3.5 and 5.5×10^5 cells/ml in the control group. The media volumes in the other groups were adjusted based on the control group, such that the cell density in these other groups could have been outside the $3.5 - 5.5 \times 10^5$ cells/ml density window set for the control group. In the outgrowth phase following day 9, cells were not passaged, but counted at the indicated time points to check for potential expansion later into the course of the experiment. (**A**, **B**) Absolute cell numbers throughout the experiment. (**C**, **D**) Percentage of live cells relative to no sgRNA sample in each group. Cells were counted on a haemocytometer using Trypan Blue stain. Results are from a single replicate.

Encouragingly, however, the relationship between the proliferation rate of the TAL10-edited cells and the no sgRNA control was in the expected pattern for both Jurkat (Figure 4.19C) and Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 (Figure 4.19D), with the no sgRNA controls overgrowing the *TAL1*-knockouts. This highlighted the importance of the blasticidin addition into the selection media for maintenance of Cas9 expression.

In sum, the adjustments implemented so far (lower hygromycin concentration, regular passaging of cells, cell density adjustments) did not prevent the massive cell death of Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 in the CRISPR/Cas9 validation experiment. Furthermore, Jurkat cells showed us that we might need to rethink the timeframes for this experiment – expansion into proliferating cell lines might be a much more lengthy process in this particular case than we expected based on previous data. Finally, the behaviour of Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 in the last experiment suggested that blasticidin in the selection media might potentially remedy the issue of unexpected proliferation patterns in the CRISPR/Cas9 validation and should therefore be subject for further testing.

4.12 The role of blasticidin, the last piece of the puzzle

Even though a number of factors, including hygromycin concentration and cell passaging patterns, were adjusted for the CRISPR/Cas9 validation screen protocol, we were still encountering the problem of the massive cell death and the problem of unexpected proliferation patterns in Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9. The latest experimental evidence indicated that the addition of blasticidin into the selection media after the introduction of the sgRNA plasmid into the cells could remedy the latter issue, possibly by improving maintenance of the *Cas9* construct. This was a possibility that we were keen to explore further.

To this end, the CRISPR/Cas9 validation experiment was carried out again in Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 (Figure 4.20). Cells were transfected with either the lentiGuide-Hygro-dTomato plasmid with the TAL10-sgRNA (TAL10), or lentiGuide-Hygro-dTomato without an sgRNA as a control (no sgRNA) using Neon electroporation with all the wellbeing adjustments made to date. To determine whether blasticidin made a difference to the cell proliferation under these conditions, cells were grown without blasticidin or with the addition of 5 µg/ml blasticidin to the selection media.

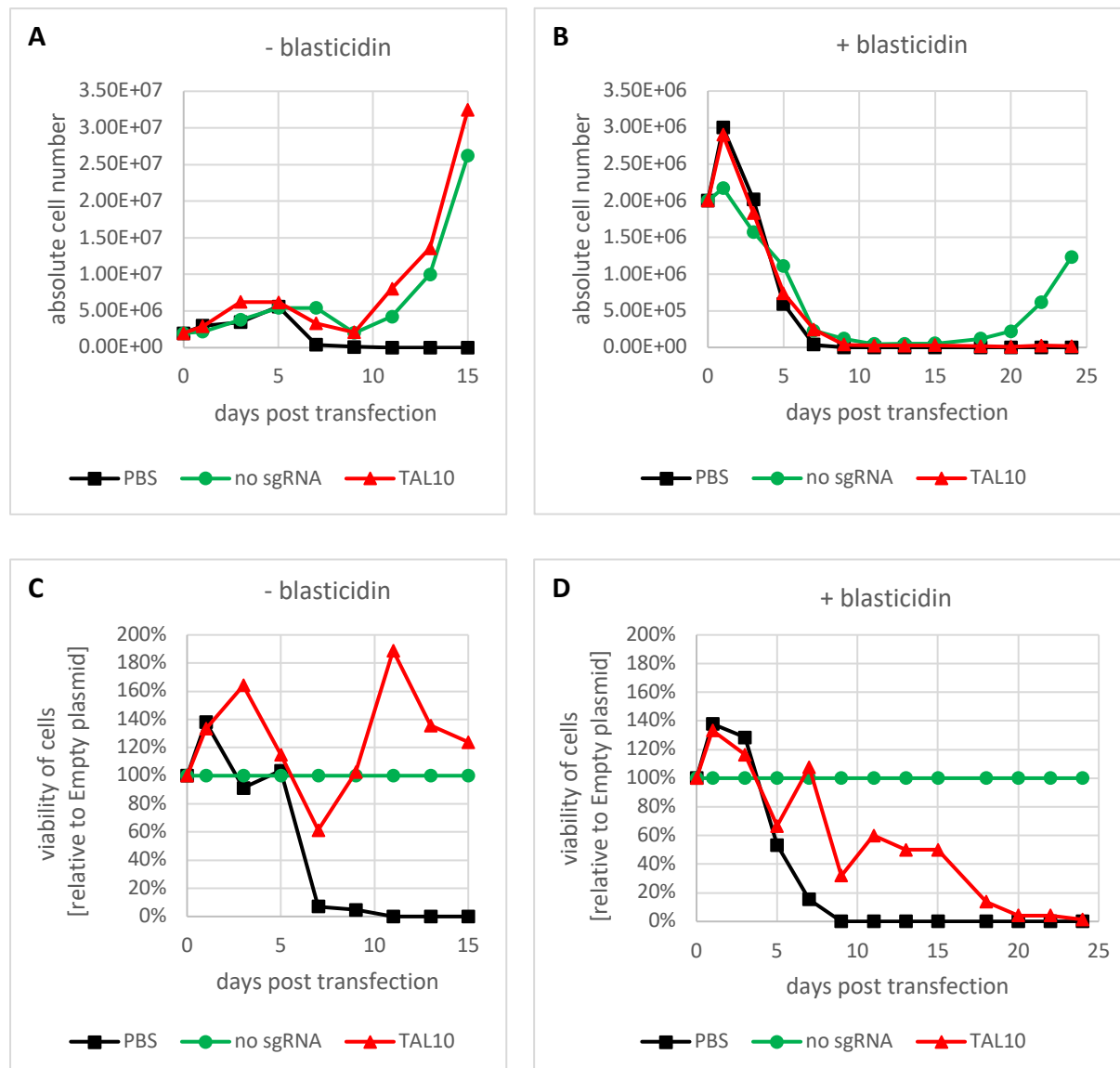


Figure 4.20: The influence of blasticidin addition into the selection media in the course of the *TAL1*-editing of Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells.

Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 (clone 85-15-4) were transfected by electroporation with lentiGuide-Hygro-dTomato encoding a hygromycin resistance cassette and either an sgRNA targeting the *TAL1* gene (TAL10), or no sgRNA (no sgRNA) as negative control. Positive control for antibiotic selection were cells mock-transfected with PBS. Cells were kept in 250 µg/ml hygromycin and either without blasticidin (**A, C**) or with 5 µg/ml blasticidin (**B, D**) from 24 hours after transduction. Cell density in the control group was maintained above 1×10^5 cells/ml and below 1×10^6 cells/ml throughout the experiment. The media volumes in the other groups were adjusted based on the control group, such that the cell density in these other groups could have been outside the $1 \times 10^5 - 1 \times 10^6$ cells/ml density window set for the control group. (**A, B**) Absolute cell numbers throughout the experiment. (**C, D**) Percentage of live cells relative to no sgRNA sample in each group. Cells were counted on a haemocytometer using Trypan Blue stain. Results are from a single replicate.

In the group whose selection media contained blasticidin, the relationship between the proliferation patterns of the *TAL1*-edited and control cells was as expected (and as previously repeatedly observed for Jurkat cells; Figure 4.20D; Figure 4.14 in Chapter 4.9; Figure 4.17A – E in Chapter 4.10; Figure 4.19C in Chapter 4.11), with *TAL1*-edited cells showing reduced proliferation relative to the negative no sgRNA control. In the group that was cultured without blasticidin, the proliferation patterns were decisively different from the blasticidin group, with the *TAL1*-edited cells displaying a proliferative advantage over the no sgRNA control group (Figure 4.20C). These observations, in addition to the experiment with added blasticidin in the previous chapter (Figure 4.19 in Chapter 4.11), strongly pointed towards the importance of blasticidin addition into the selection media for the CRISPR/Cas9 validation experiment in Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9.

In terms of viability, however, the no-blasticidin group had the advantage: after an initial expansion, cell numbers dropped for a few days before the cells started expanding from day 9 onwards (Figure 4.20A). In comparison with earlier results, this suggested that the regular cell density adjustments, which we adopted in the previous chapter, were having a positive effect on the viability of the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line. Meanwhile, the control (no sgRNA) cells in the blasticidin group fell to extremely low numbers before starting to make a very slow expansion from about 2 weeks into the experiment; the *TAL1*-edited cells in this group didn't manage to expand at all (Figure 4.20B).

The addition of blasticidin into the selection media in the CRISPR/Cas9 validation experiment in Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 seemed to be the antidote to the strange proliferation patterns we had been observing; at the same time, however, its negative impact on the viability of the cell line was astounding. This suggested that the loss of Cas9 (and the connected blasticidin resistance) was quite pervasive; the question offered itself whether it was only the stress of the CRISPR/Cas9 experiment that was to blame, or whether the cell line had been losing the Cas9 expression even before being confronted with the pressures of the CRISPR/Cas9 experiment. To address this potential issue, I started supplementing the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 growth medium with blasticidin (5 µg/ml). This step was designed to eliminate from the culture such cells that have lost their Cas9 expression, maintaining Cas9 expression in the growing culture.

With Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 thus preconditioned, I was able to strongly mitigate the viability drop in the CRISPR/Cas9 validation experiment. Figure 4.21 shows absolute cell numbers of Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 (clone 85-15-4), control (no sgRNA) and *TAL1*-edited, in the course of a transduction experiment using concentrated viral supernatants and 5 µg/ml blasticidin and 250 µg/ml hygromycin added 24 hours post transduction. In this experiment, cells were counted

every 2 days, media changed at least every 4 days and cell density adjusted every 2 days to between 1 and 2×10^5 cells/ml in the control group. Expansion of the cells into a cell line started between day 5 and day 7 after transduction in this set-up (Figure 4.21). Importantly, with this final optimized experiment we have returned to transducing cells rather than electroporating them, as this is the method we are preferentially planning to use for the CRISPR/Cas9-mediated mutagenesis of the *TAL1* enhancer.

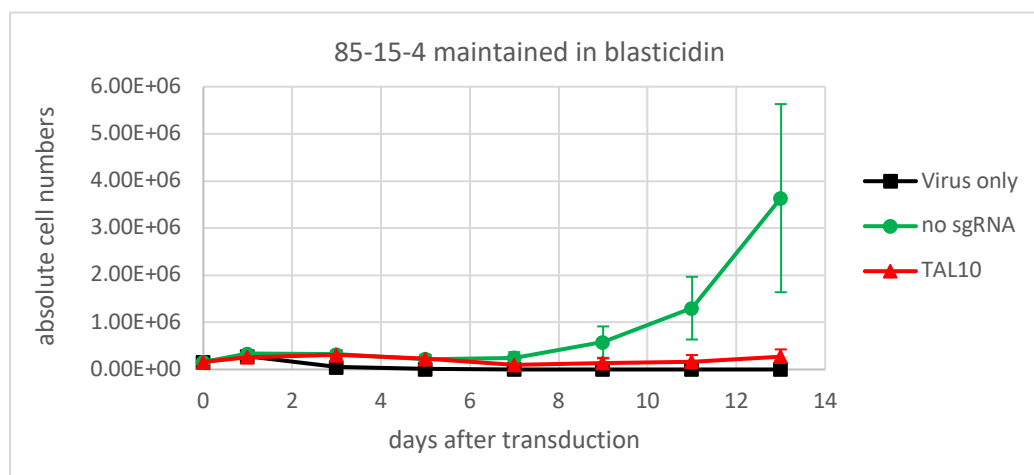


Figure 4.21: Proliferation of blasticidin-conditioned Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells after *TAL1*-CRISPR challenge.

Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line (clone 85-15-4) was transduced with viral particles encoding a hygromycin resistance cassette from the lentiGuide-Hygro-dTomato plasmid and either an sgRNA targeting the *TAL1* gene (TAL10), or no sgRNA (no sgRNA). Empty viral vector (virus only) was included as a positive control for antibiotic selection. Cells were cultured in 5 µg/ml blasticidin from at least a week prior to the experiment, and 250 µg/ml hygromycin was added 24 hours after transduction. Every 2 days, cell density was adjusted to between 1 and 2×10^5 cells/ml according to the cell counts in the negative control group, medium was changed at least every 4 days. Graph shows the percentage of live cells relative to the no sgRNA sample, as counted on a haemocytometer using Trypan Blue stain. Error bars indicate standard deviation of the mean of three biological replicates.

In brief, the presence of blasticidin in both the maintenance and the selection media in the CRISPR/Cas9 validation protocol plays a key role in the functionality of the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line. Blasticidin supplementation, in combination with a number of the other adjustments and improvements, helped us arrive at a functional, reproducible and optimized CRISPR/Cas9 validation protocol for the use with the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line.

4.13 A comparison of the effect of *TAL1* knockdown between all Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 clones

The understanding of the importance of blasticidin supplementation of Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 culture media allowed us to formulate a functional protocol for the *TAL1*-dependence CRISPR/Cas9 validation experiment in Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9. We next used this optimized protocol to compare and validate the behaviour of the three monoclonal Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell lines, 85-15-1, 85-15-3 and 85-15-4.

The optimized protocol for the CRISPR/Cas9-based validation experiment included the following changes compared to the original transduction protocol used for the establishment of the stable cell lines: preconditioning of the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells in 5 µg/ml blasticidin for at least a week prior to the experiment; transducing the cells with 9 U of concentrated viral particles; selection media containing a lowered selection dose of 250 µg/ml hygromycin and a maintenance dose of 5 µg/ml blasticidin; adjustment of cell densities to between $1 - 2 \times 10^5$ cells/ml every 2 days during the course of the CRISPR/Cas9 experiment, with a change of the selection growth media at least every 4 days.

To carry out the optimized experiment, the maternal Jurkat cells and the cells of the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line were again transduced with the lentiGuide-Hygro-dTomato plasmid bearing the *TAL1* exon-targeting sgRNA (TAL10), or the lentiGuide-Hygro-dTomato plasmid without an sgRNA (no sgRNA) as a control; Jurkat cells were co-transduced with lentiCas9-Blast at the same time, to deliver the editing enzyme into their genome. Cells transduced with an empty vector (virus only) served as a positive control for hygromycin selection pressure. While in the first part of the experiment (day 0 – 5), the TAL10-edited cells outcompeted the no sgRNA control sample at some of the time points, from day 7 into the experiment, the viability in the TAL10 group was diminished in all cell lines compared to the no sgRNA control. Furthermore, the viability in the TAL10 groups relative to no sgRNA controls decreased steadily from day 7 until the experiment was ended on day 13. This was in keeping with the expected behaviour.

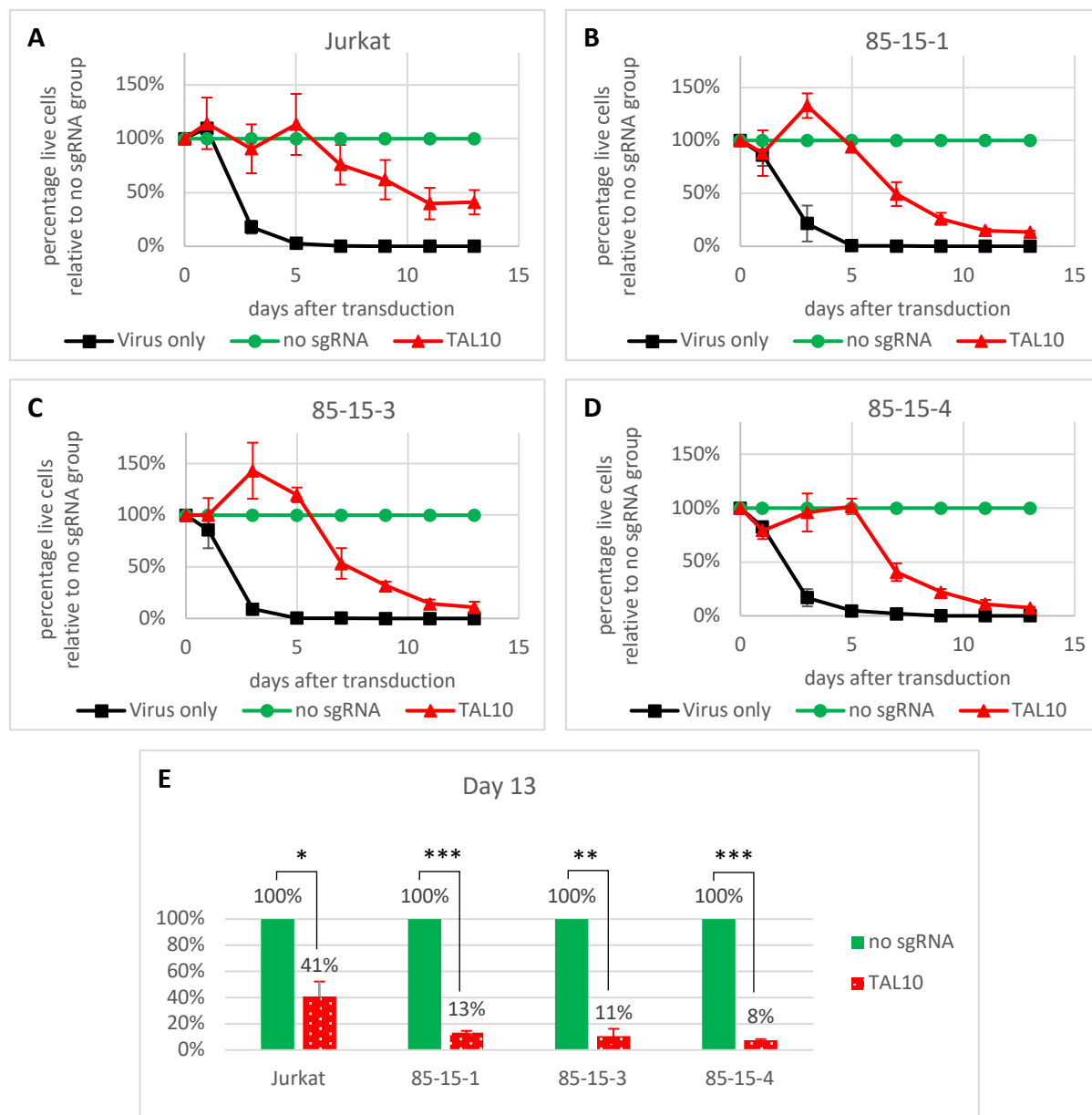


Figure 4.22: Proliferation of TAL1-challenged monoclonal Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell lines compared to maternal Jurkat cells.

Maternal Jurkat cells (**A**) and three clones of the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line (85-15-1, -3 and -4; **B – D**) were transduced with viral particles encoding a hygromycin resistance cassette from the lentiGuide-Hygro-dTomato plasmid and either an sgRNA targeting the *TAL1* gene (TAL10), or no sgRNA (no sgRNA) as a non-targeting control. Positive control for antibiotic selection was empty vector (virus only). Cells were kept in 250 µg/ml hygromycin and 5 µg/ml blasticidin from 24 hours after transduction. Every 2 days, cell density was adjusted to between 1 and 2 x 10⁵ cells/ml according to the cell counts in the no sgRNA control group, medium was changed at least every 4 days. (**A – D**) Graphs show the percentage of live cells relative to no sgRNA sample in each group, i.e. Jurkat (**A**), 85-15-1 (**B**), 85-15-3 (**C**), and 85-15-4 (**D**). (**E**) Bar graph shows percentage of live cells in control (no sgRNA) vs *TAL1*-edited (TAL10) samples on day 13 post transduction for maternal Jurkat cells and all Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 clones, relative to the no sgRNA control within the group. Cells were counted on a haemocytometer using Trypan Blue stain. Results are from three biological replicates; error bars indicate standard deviation of the mean. Statistical analysis in (**E**) is a two-tailed unpaired T-test. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

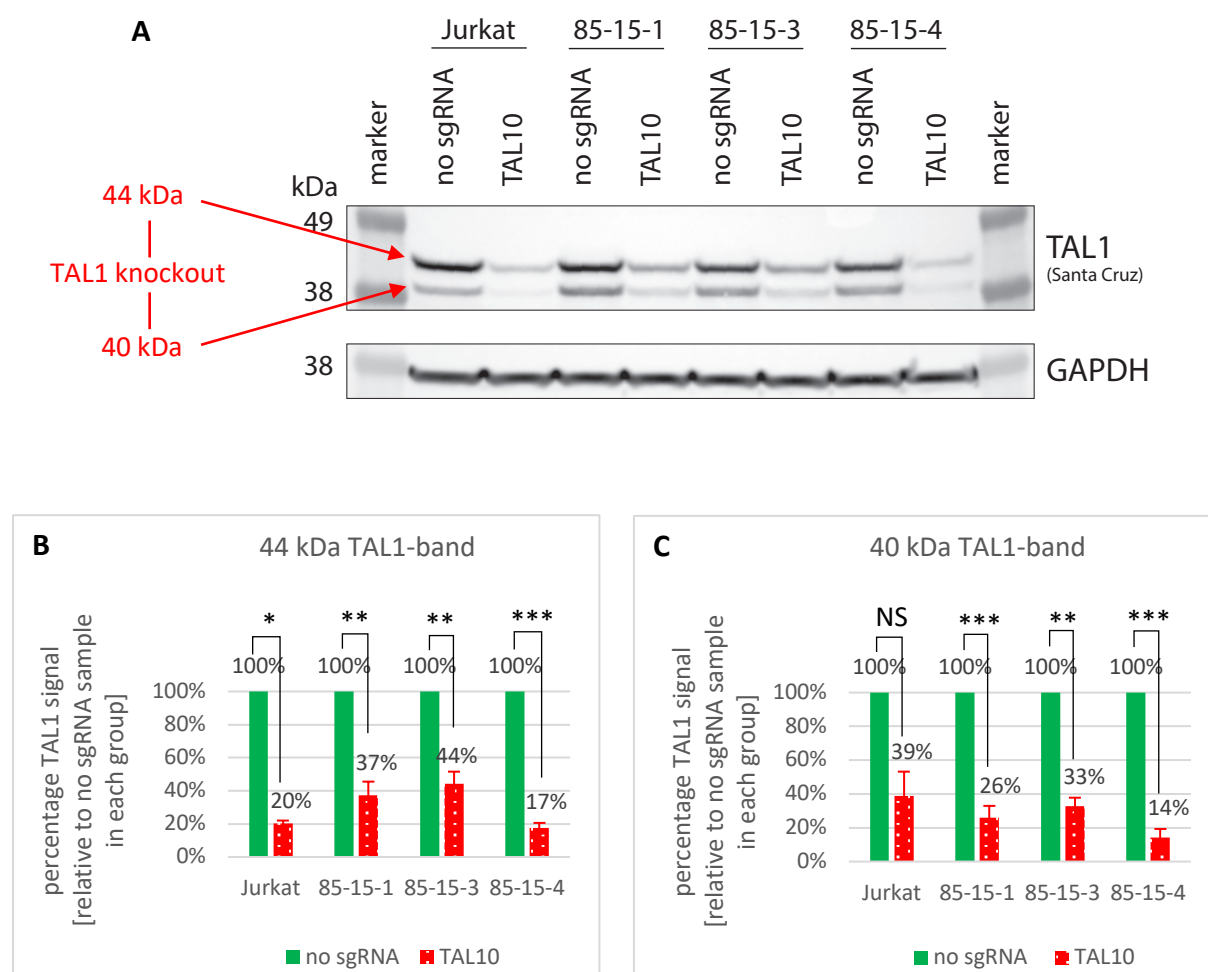


Figure 4.23: TAL1 expression in maternal Jurkat cells and in clones of Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 upon CRISPR/Cas9-mediated knockout.

Cells were treated as described in Figure 4.22 legend. Cells for Western blot analysis were harvested on day 21 after transduction. **(A)** A representative Western blot. Membranes were probed with anti-TAL1 antibody (sc-393287, Santa Cruz) and anti-GAPDH antibody (60004-1-Ig, Proteintech). **(B, C)** Relative downregulation of the 44 kDa TAL1 band **(B)** or the 40 kDa TAL1 band **(C)** normalized to GAPDH and compared to the no sgRNA sample in each group. Band intensities were read from the Western blot images using the Fiji software (Schindelin et al., 2012). Results are from two (Jurkat) or three (all Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 clones) biological replicates, error bars indicate standard deviation of the mean. Statistical analysis in **(B)** and **(C)** is a two-tailed unpaired T-test. NS – not significant, * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

This validation experiment confirmed a preserved TAL1-dependence in all the three clones of Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 (Figure 4.22). Importantly, in comparison to the maternal Jurkat cell line (Figure 4.22A), all the three clones (Figure 4.22B – D) showed a more marked downregulation in the proliferation rate of the TAL10-edited sample compared to the no sgRNA control (Figure 4.22E). On the last day of the experiment, day 13 after transduction, the number of cells in the TAL10-edited

sample of 85-15-4 dropped on average to only 8% of the number of cells in the control sample (Figure 4.22D, E). This was the most pronounced decrease in proliferation rate recorded among the tested cell lines, although the other two clones were not far behind with the average drop in the number of cells in the TAL10-edited sample to 13% in 85-15-1 and 11% in 85-15-3 (Figure 4.22E). In Jurkat cells, the decrease was to 41% (Figure 4.22E). This highlighted a further benefit of the stable Cas9 expression in Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line compared to the expression from a co-transduced Cas9 plasmid.

To confirm that the drop in the proliferation rates was indeed caused by diminished TAL1 expression, and to compare the magnitude of TAL1 downregulation in all the cell lines, a Western blot was performed, which showed a marked decrease in the two main TAL1 isoforms, observed at 40 kDa and 44 kDa, in all the *TAL1*-edited samples (Figure 4.23A). This decrease was again most pronounced in 85-15-4 (Figure 4.23B and C), although overall, the measure of TAL1 downregulation by Western blot did not seem to tally with the decrease in proliferation rate (data not shown).

Overall, a CRISPR/Cas9-mediated TAL1 challenge was detrimental to the proliferation rates of all the three clones of Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9; the timeframe for the differences to play out was about 2 weeks from the time of transduction. Clone 85-15-4 performed the best of the three Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 monoclonal cell lines, with the most marked decrease in viability compared to a no sgRNA control (8%) and the most pronounced reduction in TAL1 expression (14 and 17% of control for the 40 kDa and the 44 kDa band, respectively).

4.14 Discussion

In this chapter, we described the first half of the two-part validation process of the newly established Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line – the confirmation of its dependence on the transcription factor TAL1. This addiction is well-described in the Jurkat cell line (Palii et al., 2011; Palomero et al., 2006; Sanda et al., 2012) and the planned CRISPR/Cas9-mediated screen of the *TAL1* enhancer hinges on this property, so it is paramount that the newly established cell line had preserved this TAL1-dependence. We perceived a potential danger to this property in the multiple genetic modifications we had introduced into the cell line and the long months of culturing, often in stressful conditions, which were nevertheless inevitable in the process of the establishment of the new cell line.

In practical terms, the viability of the downstream CRISPR screen experiment is dependent on our cells showing a drop in *TAL1* expression upon *TAL1* gene or *TAL1* enhancer knockout or *TAL1* knockdown, coincident with a decrease in proliferation. Furthermore, the control cells, targeted with non-targeting sgRNAs or siRNAs, need to demonstrate the ability to survive the screening conditions and expand into healthy cultures.

Because a well-functioning CRISPR/Cas9-based experiment for the *TAL1*-dependence validation in the Jurkat-Tet3G-TRE3G-FLAG-*TAL1*-Cas9 cell line would give a good basis for the protocol for the planned CRISPR/Cas9 screen of the *TAL1* enhancer, we first attempted to use and optimize this method. We were going to test five different guide RNA sequences for this purpose, yielding, upon cloning into a plasmid with an sgRNA scaffold, two sgRNAs targeting the *TAL1* exonic region (TAL7, TAL10), two sgRNAs targeting the 5'-untranslated region of *TAL1* (TAL1, TAL4), and one sgRNA targeting the MYB-binding site at the *TAL1* enhancer (MYB1; Figure 4.2 and Table 4.1 in Chapter 4.2). The exonic sgRNAs targeted the first of the *TAL1* exons – the consensus is that the closer to the 5'-end of the gene Cas9 cuts, the more effective the resulting knockout in terms of (dis)functionality of the gene (Doench et al., 2014). While the exonic sgRNAs would knockout both the endogenous *TAL1* and the newly introduced *FLAG-TAL1* in Jurkat-Tet3G-TRE3G-FLAG-*TAL1*-Cas9, the *TAL1* 5'-UTR-targeting sgRNAs were designed to only target the endogenous *TAL1*, leaving the *FLAG-TAL1* intact, which means that the proliferation rate of the cells edited with these sgRNAs should be rescuable by the intact, doxycycline-inducible *FLAG-TAL1* – a useful feature for the second half of the validation process. The sgRNA designated here as MYB1, which targets the MYB binding site at the monoallelic de novo *TAL1* enhancer in Jurkat cells, had been used previously in one of the papers our project is based on (Mansour et al., 2014). In this study, *TAL1* expression was decreased considerably in MYB1-edited Jurkat cells, by 55 – 85% compared to the maternal cell line.

To introduce these sgRNAs into Jurkat-Tet3G-TRE3G-FLAG-*TAL1*-Cas9, we initially chose viral transduction, as this is the traditional way of performing CRISPR/Cas9-based screens (Shang et al., 2017). Transduction is the go-to method because of its high efficiency, even in difficult-to-transfect cells, and because of its unparalleled rate of stable integration (Dong and Kantor, 2021). For example, based on published data, the integration frequency of the gene of interest following transduction with HIV-1-based viral particles was about 1 in 3 – 4 cells (Bayer et al., 2008; Kantor et al., 2011), while the integration of the DNA of interest following electroporation was reported to be anywhere between 1 in 10^3 to 10^6 cells depending on cell type (Potter, 2003). The integration of the exogenous DNA into the genome is paramount, as the typical output of a CRISPR/Cas9 screening experiment is the sequencing of the introduced sgRNAs.

From the first of our experimental endeavours, we observed two serious problems arising: firstly, a lack of long-term survival not only in the TAL1-CRISPR-edited, but also the non-targeting control Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 in the context of the CRISPR/Cas9 validation experiment (“the massive cell death issue”); and secondly, unexpected and inconsistent proliferation patterns upon the sgRNA introduction, where TAL1-challenged cells often outperformed the non-targeting controls. Our initial suspicions of malfunctioning or delayed hygromycin resistance gene or protein expression were replaced by flow cytometry evidence of extremely low transduction efficiency (Figure 4.7 in Chapter 4.5).

To fix the low transduction efficiency, we have introduced a step to concentrate the viral stock. The method of our choice used a chemical reagent (Lenti-X Concentrator, Takara) to increase the viral titre up to 100-fold, without the common need for ultracentrifugation. The procedure also allowed us to obtain a purer preparation – in the concentration process, PBS replaces the medium from the packaging cells, which is likely to be depleted, contain stress factors and other undesirable substances; additionally, the medium used for the HEK293T packaging cell line is of a different kind (DMEM) than the medium used for Jurkat cells (RPMI). From optimization results, the method seemed highly efficient. However, when applied in the context of the CRISPR/Cas9 validation experiment, the promising new step failed to help with either the massive cell death or the inconsistency of the results (Figure 4.9 in Chapter 4.7).

Suspecting that perhaps the newly established cell line was too dysfunctional to be used for TAL1-challenge-based experiments, we turned to a completely different method of validation, siRNA-mediated knockdown of *TAL1*. Encouragingly, this method confirmed that TAL1 depletion poses a challenge for the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line (Figure 4.11 in Chapter 4.8), and the cell line is, in this sense, functional.

This meant that further optimization was needed for the CRISPR/Cas9 experiment. The first adjustment we made was a transition from transduction to electroporation with the Neon Transfection System (Thermo Fisher). Neon electroporation is highly efficient (Figure 4.12 in Chapter 4.9) and easy to scale up, although much of the transfection is only transient (Potter, 2003). The integration of the sgRNA sequence into the genome is important from the viewpoint of the ultimate goal of this work, the CRISPR/Cas9 screen. The low rate of plasmid integration into the genome encountered in all transfection methods is in contrast with transduction, which is much better suited for integrating new sequences into the genome and is consequently also the commonly employed method for CRISPR/Cas9 screens in the literature. Nevertheless, a small percentage of transfected cells do integrate the new genetic information into their genome; these cells can be selected for by

standard methods, and so Neon electroporation would potentially be a viable alternative to transduction.

Swapping transduction for electroporation in the CRISPR/Cas9 validation experiment did not, however, solve either the massive cell death issue or the unexpected and inconsistent results problem in Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9. Curiously, however, using the same experimental setup (+Cas9 co-transfection), we saw a clear and reproducible pattern of TAL1-dependency in the maternal Jurkat cell line (Figure 4.14 in Chapter 4.9).

The next avenue we explored in order to alleviate either or both of the problems we were encountering with Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 in the context of the CRISPR/Cas9 experiment were the possibilities for a gentler treatment of the cell line throughout the experiment. The cell line is clearly more sensitive and less robust than its maternal counterpart, and a gentler approach might be needed in order to keep the cells alive in the course of the CRISPR/Cas9 experiment, which brings a combination of stressors to the cells. Firstly, we reconsidered the hygromycin concentration used for selection of the sgRNA-transfected cells and halved it. Secondly, we ventured to gain a deeper understanding of the proliferation patterns of the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line in order to find optimal conditions for the passaging of these cells.

The exploration of the latter yielded some interesting information (Figure 4.18 in Chapter 4.11). Cells of a wide range of starting cell densities, from 2.5×10^4 to 1.0×10^6 cells/ml, seemed to proliferate well and roughly exponentially without any passaging unless they either reached the plateau (4×10^6 cells/ml), or became overgrown, as indicated by a change of colour due to the pH indicator in the media. This successful proliferation pattern, however, only continued until day 4 of the experiment. Cells in all the cultures started dying off after that point, even cells in cultures which were not visibly overgrown and the medium colour didn't indicate the need for a change.

In trying to explain this phenomenon, we arrived at two possible, yet quite implausible hypotheses: the first one considered that even when the colour of the media didn't indicate a need for an exchange, some of the nutrients may already have been depleted, and the cells were therefore not being nourished properly; the second hypothesis was concerned with the fact that L-glutamine, a common additive in growth media, is unstable and over time degrades to ammonia (Tritsch and Moore, 1962), whose build-up may have a toxic influence on the cells (Schneider et al., 1996), although this is not usual at the concentrations of ammonia that can arise from L-glutamine degradation in media (Jagušić et al., 2016). Whatever the reason of the demise of the cells was, it brought to our attention the importance of the change of the growth media at least every 4 days for this cell line.

Another important message of the proliferation experiment was that Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells are not always comfortably expanding at very low cell densities (under 2.5×10^4 cells/ml). This may be tied to catalase concentrations in the medium. Catalase is an enzyme that breaks down toxic reactive oxygen species; it was previously recognized as a survival factor in high-density activated T-cell cultures, and its protective powers were demonstrated in low-density cultures (Ma et al., 2010). Higher cell-density cultures secrete enough catalase for their own protection from oxidative apoptosis; it follows that lower cell density cultures may not (Ma et al., 2010). While this behaviour was not reported in Jurkat cells, it is not a stretch of imagination that these cells, being T-lymphocytes as well, could behave similarly. Conceivably, this could also become more pronounced under stress; therefore, while unchallenged Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells may be happy at cell densities above 2.5×10^4 cells/ml, it is conceivable that the cells undergoing the CRISPR/Cas9 validation experiment will require higher cell densities for their survival. It is also important to recognize that while the cell density in an unchallenged culture generally increases over time, this may not be the case when antibiotic selection is applied to transfected cells. Cell density may actually be decreasing in such cultures for a time. A practical implication of these inferences, in conjunction with the available results, is to keep the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells in inherently stressful experiments at higher cell densities, at least above 1×10^5 cells/ml starting cell density and potentially higher, up to 7.5×10^5 cells/ml where cell density adjustments and media exchange are done every 2 days, up to 2.5×10^5 cells/ml where these are done every 3 days.

Another important discovery came with the exploration of the role of blasticidin, the resistance to which is linked to the Cas9 expression in our cell line. When selection media was supplemented with blasticidin in our CRISPR/Cas9 TAL1-dependence validation experiment in Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9, the resulting proliferation patterns of *TAL1*-edited cells relative to control were as observed with the maternal Jurkat cells and as expected, whereas when blasticidin was omitted from the selection media, we observed different, varying and unexpected patterns. This clearly pointed towards the paramount importance of exerting selection pressure linked to the *Cas9* gene during the CRISPR/Cas9 experiments. As previously discussed, Cas9 is a bulky enzyme and likely has negative effects on the proliferation rate of the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line, and as such is a prime candidate for silencing or complete loss of expression levels, especially in response to further stress. Using blasticidin during the CRISPR/Cas9 validation screen is designed to keep the population of cells Cas9-positive.

Going a step further, we hypothesised that the loss of Cas9 expression might be an ongoing process in Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 even under normal conditions. To counter this potentiality, we started growing the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells in blasticidin. This had an

immensely beneficial effect on the performance of the cell line in the CRISPR/Cas9 validation experiment carried out with blasticidin in the selection media (compare Figure 4.20B in Chapter 4.12 and Figure 4.21 in Chapter 4.12), and along with the aforementioned adjustments led to the establishment of a protocol where the cells were able to survive and expand into polyclonal cell lines.

In answer to this development, we also started supplementing the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 growth media with G418 and puromycin, to prevent the potential loss of the other exogenes. There is a number of reasons why many labs avoid the maintenance use of antibiotics. In general, the addition of each new component into the growth medium presents an additional variable. Cells cultured in low antibiotic concentrations can develop resistance phenotypes; antibiotics can mask contamination and infections. Furthermore, if the gene of interest and the selectable marker are not expressed from a single promoter, the expression of the two is not indelibly linked, and the gene of interest may be silenced regardless of the fate of the resistance gene. Despite all that, it is common practice to culture stable cell lines in maintenance antibiotic concentrations, and in the case of our newly established cell line, this seems to be both a functional and a necessary measure.

Considering the two major issues, the unexpected behaviour and the inconsistencies in proliferation patterns were chiefly due to the presumed loss of Cas9 expression, and were largely remedied by supplementing the growth media with blasticidin; the massive cell death issue was possibly owing to a combination of factors, and was counteracted by a combination of measures – increasing the sgRNA-introduction efficiency, lowering the hygromycin concentration for selection, adopting optimal passaging patterns and maintaining cells in blasticidin-supplemented growth media prior to the experiment.

An important consideration was the timeframe of the experiment, from at least two perspectives. Firstly, a key factor to bear in mind is the amount of time it takes for the CRISPR editing to take place. Various timeframes were reported in connection with various experimental systems; for example, Shalem and colleagues found over 90% of their EGFP-expressing cells edited by day 11 post transduction with Cas9 and their best targeting sgRNA (Shalem et al., 2014); Zhu and colleagues explored the kinetics of paired gRNAs action in a cell line with stable Cas9 expression, which reached a plateau around 2 weeks after transduction (Zhu et al., 2016). From our Western blot experiments, it was clear that the *TAL1*-exon-targeting sgRNAs (TAL7, TAL10) acted very fast, as 3 days after transfection, there already was a considerable decrease in TAL1 levels (Figure 4.4 in Chapter 4.3). The *TAL1* enhancer-targeting sgRNA, MYB1, needed more time to act, as 3 days post transfection, the decrease in TAL1 levels wasn't yet noticeable (Figure 4.4 in Chapter 4.3), while 4 weeks post transduction the downregulation of TAL1 was unmistakeable (Figure 4.7C in Chapter 4.5). This delay

in TAL1 downregulation after MYB1 vs TAL7/TAL10 targeting could have been due to the fact that TAL7 and TAL10 target the *TAL1* gene directly, whereas MYB1 targets a regulatory element responsible for *TAL1* expression, and the effects may therefore have taken longer to play out; or it could have been due to a different speed of action of each of the sgRNAs. Taken together, the kinetics of CRISPR editing seem to depend not only on the particular system and experimental setup, but also on the particular sgRNA being used.

Secondly, regarding the timeframe of our CRISPR/Cas9 validation experiment in Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9, a curious occurrence was observed in the final experiment (Figure 4.22 in Chapter 4.13) and likely, in various forms, in the preceding experiments as well. Before the viability rate of the *TAL1*-edited sample sank beneath the viability rate of the control sample, it rose for several days, often after an initial dip in the first 24 hours. We theorize that this is due to the addition of hygromycin selection 24 hours after transduction. The hygromycin resistance might likely take a longer time to come into functional existence in comparison with the sgRNAs, which are shorter and function as RNA transcripts, while the hygromycin resistance cassette requires not only transcription, but also a translation and post-translational processing to become a functioning entity. Hygromycin, just as most other antibiotics, acts faster on more actively dividing cells. Therefore, in the early phase of the experiment, the decrease in proliferation rate due to *TAL1*-editing may actually provide the edited cells with a level of protection against the antibiotic, and as a result, for some time, the *TAL1*-edited cells may be outgrowing the controls. Therefore, enough time has to be allowed for this effect to be neutralized and for the effect of the TAL1 depletion to fully manifest itself.

When the three clones of the final Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line were compared to each other in terms of their behaviour in the CRISPR/Cas9 validation experiment, there were only small differences between them (Figure 4.22 in Chapter 4.13). 85-15-1 and 85-15-3 had a more pronounced reaction to the addition of hygromycin than 85-15-4, which could have been due to the higher proliferation rates of 85-15-1 and 85-15-3 controls compared to the 85-15-4 control (data not shown). It could also be attributed to a more effective TAL1 downregulation in 85-15-4 *TAL1*-edited samples compared to the other two cell lines (Figure 4.22 in Chapter 4.13). In any case, all the clones of Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 were dependent on TAL1 expression, and 85-15-4 seemed to be the best-performing clone.

Interestingly, all three clones of Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 showed a more pronounced phenotype after *TAL1* knockout than the maternal cell line (Figure 4.22 in Chapter 4.13), which was likely due to the fact that the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line enters the experiment

with a Cas9 already stably integrated in the genome, whereas Jurkat cells are co-transduced with Cas9 at the time of the sgRNA introduction.

In summary, through a number of various experiments, we succeeded in the optimization of the CRISPR/Cas9 protocol for the use with the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line and showed that the newly established cell line has preserved its addiction to TAL1.

5 Growth rate of the newly established cell line upon TAL1 downregulation can be partially rescued by the doxycycline-inducible FLAG-TAL1

5.1 Introduction

For the purposes of our planned CRISPR/Cas9 screen, we established the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line with a doxycycline-inducible expression of FLAG-TAL1 and stable expression of Cas9 (Chapter 3). Before the screen itself could be carried out, the behaviour of this cell line had to be examined and two particular points validated. Firstly, it was key to ascertain that the newly established cell line has preserved its dependency on the transcription factor TAL1. The validation on this point was eventually successful, although it required extensive optimization of experimental conditions (Chapter 4). Nevertheless, we ultimately arrived at a robust and reproducible protocol for a CRISPR/Cas9-based validation of TAL1-dependency, which formed a useful basis both for the validation experiments in this chapter and for the planned CRISPR/Cas9 screen.

This chapter describes the testing of the second key characteristic of the cell line indispensable for the successful undertaking of the planned CRISPR/Cas9 screen (Figure 5.1) – the doxycycline-inducible FLAG-TAL1-mediated rescue of proliferation rate in response to a TAL1 downregulation. In general, the idea of the rescue rests in the intention to functionally replace the depleted endogenous TAL1 by the exogenous doxycycline-inducible FLAG-TAL1. In practice, we expect doxycycline-treated samples of TAL1-challenged Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells (except for the cell samples edited with sgRNAs that also target the exogenous FLAG-TAL1) to proliferate at the same speed as doxycycline-treated Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells transduced with non-targeting sgRNAs.

This behaviour is key for our planned CRISPR/Cas9 screen. If the rescue is functional, we would expect an sgRNA library containing anti-*TAL1* sgRNAs to be equally represented in doxycycline-treated Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells, while differentially represented in untreated Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells. The comparison between the sgRNA representation in the doxycycline-treated (control) sample and untreated (experimental) sample should then highlight sgRNAs whose editing has a potential to downregulate TAL1.

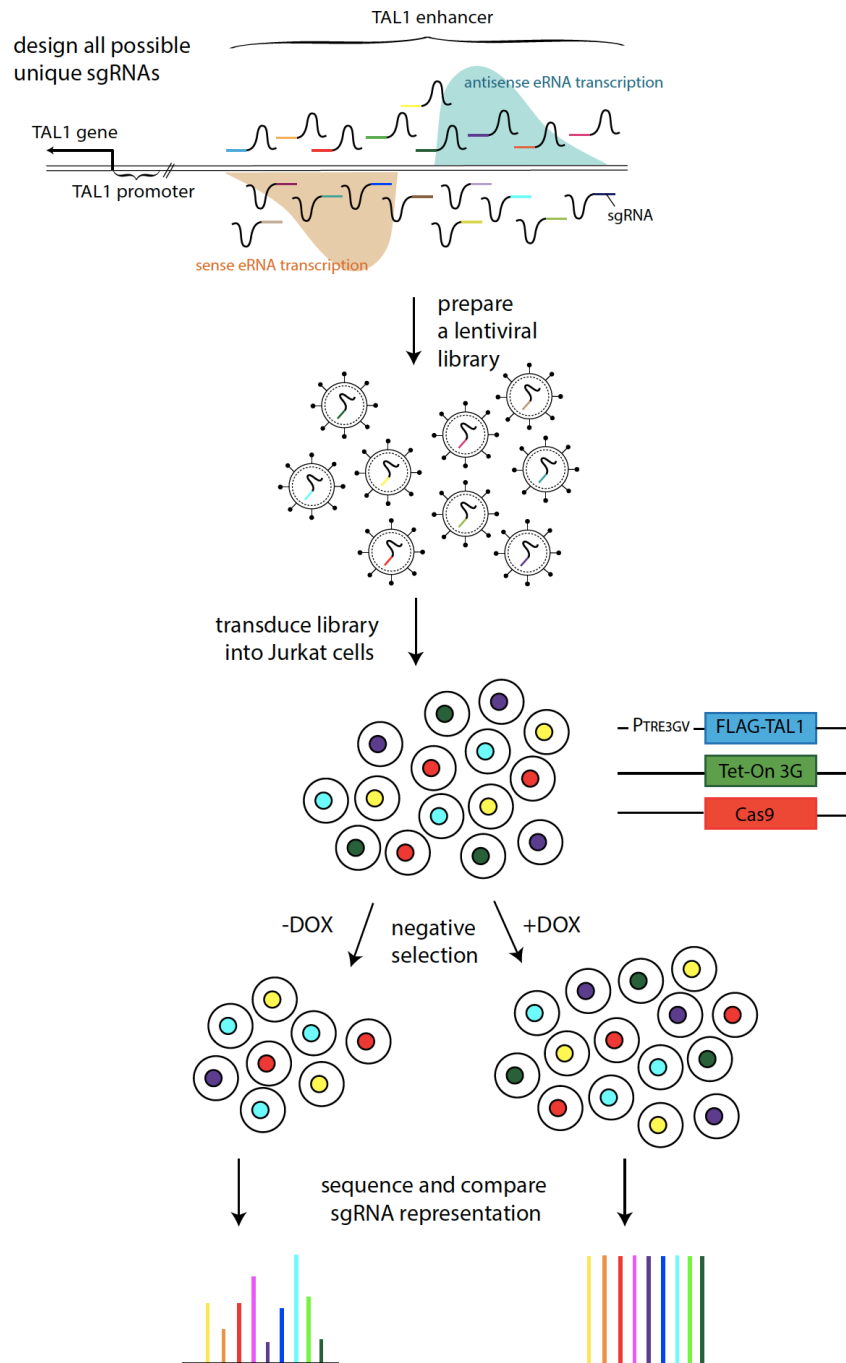


Figure 5.1: CRISPR/Cas9-mediated saturating mutagenesis of the transcribed portion of the *TAL1* enhancer in Jurkat cells (copy from Figure 1.16).

All unique sgRNAs targeting the transcribed portion of the *TAL1* enhancer will be cloned into a plasmid and packaged into a lentiviral library for delivery into genetically modified Jurkat cells, which can inducibly express an exogenous FLAG-TAL1 and stably express Cas9 as a component of the CRISPR/Cas9 system. Upon the transduction of the library, the control group will be supplemented with doxycycline to enable the FLAG-TAL1 rescue, while the experimental group will be left to its own devices. After the editing has taken place and the editing effects have had a chance to manifest, genomic DNA from the cells will be isolated for sequencing, and the sgRNA representation will be compared between the experimental and the control group.

5.2 Doxycycline induces FLAG-TAL1 expression in Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 and leads to increased proliferation rates over the first 2 days of culture

In order to form an idea about what concentration of doxycycline is needed for the FLAG-TAL1-mediated rescue, Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 (clone 85-15-4) cells were incubated with increasing doxycycline concentrations for 24 hours (Figure 5.2A) and 48 hours (Figure 5.2B). No doxycycline addition was made after the initial supplementation. Taking into account both the endogenous TAL1 isoforms, represented by a 40 kDa and a 44 kDa band on Western blot, the addition of 750 ng/ml doxycycline induced an amount of FLAG-TAL1 closest to the endogenous TAL1 amounts after 24 hours in culture (Figure 5.2A), while at 48 hours, 400 ng/ml doxycycline induced FLAG-TAL1 expression at levels comparable with the endogenous TAL1 levels in uninduced cells (Figure 5.2B).

The proliferation rate of the doxycycline-induced Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells during the first 2 days after induction was elevated compared to uninduced controls and positively correlated with the concentration of doxycycline up to the concentration of 750 ng/ml and up to an increase in cell numbers of around 25% relative to an uninduced sample (Figure 5.3). This could be explained by the cell line's addiction to TAL1 and requirement for TAL1 expression for proliferation.

Taken together, Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells expressed increasing levels of FLAG-TAL1 when induced by increasing concentrations of doxycycline, and proliferated faster with doxycycline in a concentration-dependent manner.

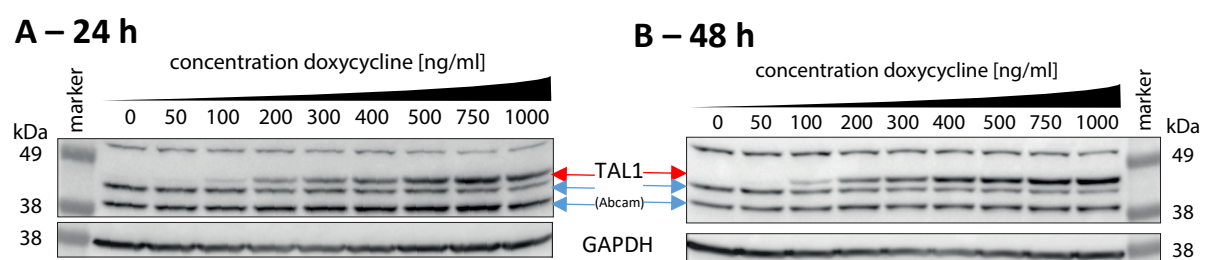


Figure 5.2: Titration of doxycycline for the purposes of FLAG-TAL1 induction in the final line, Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9.

Western blots show FLAG-TAL1 expression 24 hours (**A**) and 48 hours (**B**) after doxycycline induction. **Red arrows** represent FLAG-TAL1, **blue arrows** point out the two endogenous TAL1 isoforms. Membranes were probed with anti-TAL1 antibody (ab155195, Abcam) and anti-GAPDH antibody (60004-1-Ig, Proteintech).

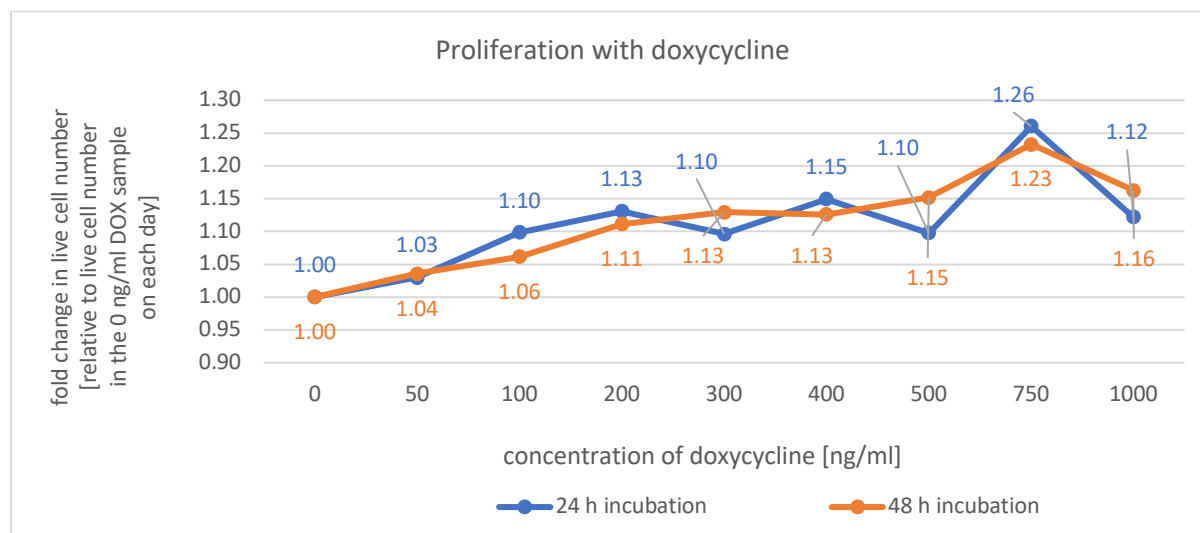


Figure 5.3: Viability of Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 upon a 24-hour and 48-hour incubation with increasing concentrations of doxycycline.

Measured by alamarBlue HS fluorescence. Results show data from a single replicate.

5.3 Effect of doxycycline on the proliferation rate of the newly established cell line

In a first attempt to put the FLAG-TAL1 rescue into action using CRISPR/Cas9 editing of the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line, cells were transduced with either the lentiGuide-Hygro-dTomato plasmid carrying hygromycin resistance and one of the five previously described anti-*TAL1* sgRNAs, which are expected to generate a growth defect, or the lentiGuide-Hygro-dTomato plasmid carrying hygromycin resistance, but no sgRNA. The optimized protocol for Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 transduction was used. To induce the expression of FLAG-TAL1, cells were supplemented with 400 ng/ml doxycycline every 2 days until day 15, then every 3 days until the end of the experiment. Viability measurements on day 13 (Figure 5.4A) and day 21 (Figure 5.4B) showed a level of rescue in the cells transduced with the *TAL1* exon-targeting sgRNAs (TAL7 and TAL10) and the *TAL1* enhancer-targeting sgRNA (MYB1), while there was no such trend discernible in the cells transduced with the *TAL1* 5'-UTR-targeting sgRNAs (TAL1 and TAL4). These two sgRNAs, in fact, didn't seem to cause a growth challenge in the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line at all, which was in contrast with the effect exhibited on the wild-type Jurkat cells (Figure 4.14 in Chapter 4.9).

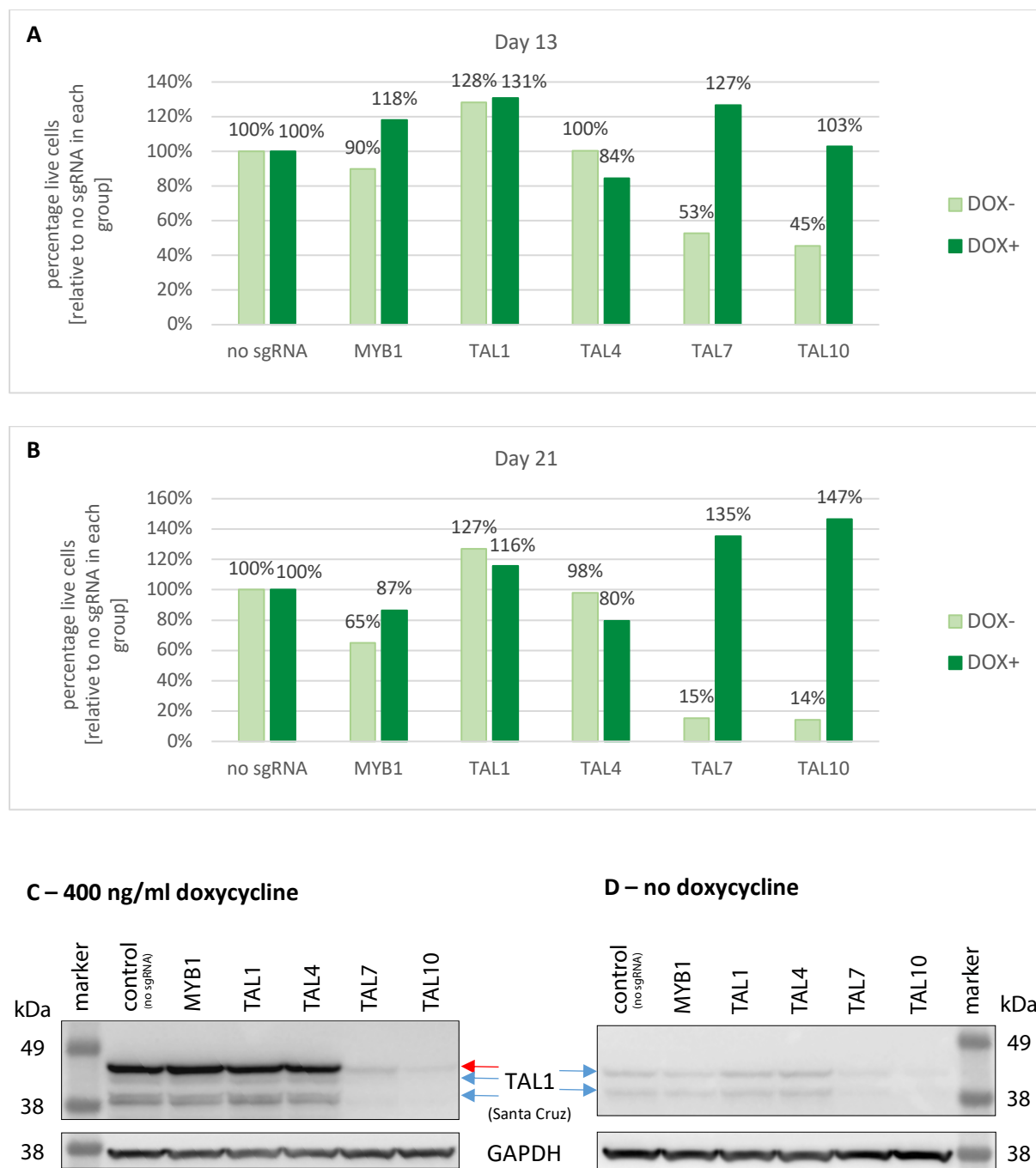


Figure 5.4: The effects of cultivation of TAL1-challenged Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells in doxycycline.

Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells were transduced with a panel of sgRNAs – an sgRNA targeting MYB-binding site in the *TAL1* enhancer (MYB1), sgRNAs targeting the 5'-UTR of *TAL1* (TAL1, TAL4) and sgRNAs targeting *TAL1* exons (TAL7, TAL10) – and cultured from the point of transduction (day 0) either without the addition of doxycycline or with the addition of 400 ng/ml doxycycline. Viability in all groups was measured using alamarBlue HS on day 13 post transduction (**A**) and day 21 post transduction (**B**) and is normalized to viability levels in the control groups, transduced with plasmid without sgRNA (no sgRNA), for each doxycycline concentration separately. DOX- (light green bars) – cells cultured without doxycycline; DOX+ (dark green bars) – cells cultured in 400 ng/ml doxycycline. Results show data from a single replicate. (**C**) and (**D**) show Western blots of lysates prepared from samples treated as outlined above and harvested on day 24 after transduction. (**C**)

Samples maintained in selection media with 400 ng/ml doxycycline; **(D)** Samples grown in selection media without doxycycline. **Red arrow** represents FLAG-TAL1, **blue arrows** point out the two endogenous TAL1 isoforms. Membranes were probed with anti-TAL1 antibody (sc-393287, Santa Cruz) and anti-GAPDH antibody (60004-1-Ig, Proteintech).

Western blots on day 24 post transduction confirmed the expected endogenous TAL1 downregulation in the MYB1, TAL7 and TAL10-edited samples (Figure 5.4C and D) and showed an abundant FLAG-TAL1 expression in the control, MYB1-, TAL1- and TAL4-edited samples, as well as low-level FLAG-TAL1 expression in the TAL7- and TAL10-edited samples (Figure 5.4C). This latter was somewhat unexpected, as the TAL7- and TAL10-sgRNAs target both the endogenous TAL1 and the exogenous FLAG-TAL1. However, the truth remains that as with each knockout, there will always be a proportion of non-deleterious mutations in an edited population, even in such cases where the editing efficiency reaches close to 100%. As a result, a partial rescue of TAL7- or TAL10-edited cells is a real possibility.

Apart from these findings, the first CRISPR/Cas9 rescue experiment further strongly indicated that long-term exposure of Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 to doxycycline might have a negative effect on the proliferation rate of the cell line (Figure 5.5). Comparing the expansion of the control cells in the rescue experiment, we noted that while both groups eventually expanded into cell lines (Figure 5.5A), the relative viability of the doxycycline-cultured control group surpassed that of the no-doxycycline control group in the early stages of the experiment, consistent with the increase in proliferation we observed in the titration experiment in the last chapter, then plummeted in the second half of the experiment (Figure 5.5B). Similarly, the viability of the MYB1-edited sample in later stages of the experiment was higher in the no-doxycycline group compared to the group supplemented with doxycycline (Figure 5.5C). However, importantly, from day 15 after transduction, the viability rates of the control and MYB1-edited sample were much more similar to each other in the doxycycline group than in the no-doxycycline group (Figure 5.5C). Figure 5.5D shows again the viability measurements from day 21 of the experiment, but here all of them are shown relative to the no doxycycline control group, further highlighting the relationship between the viability in the two groups.

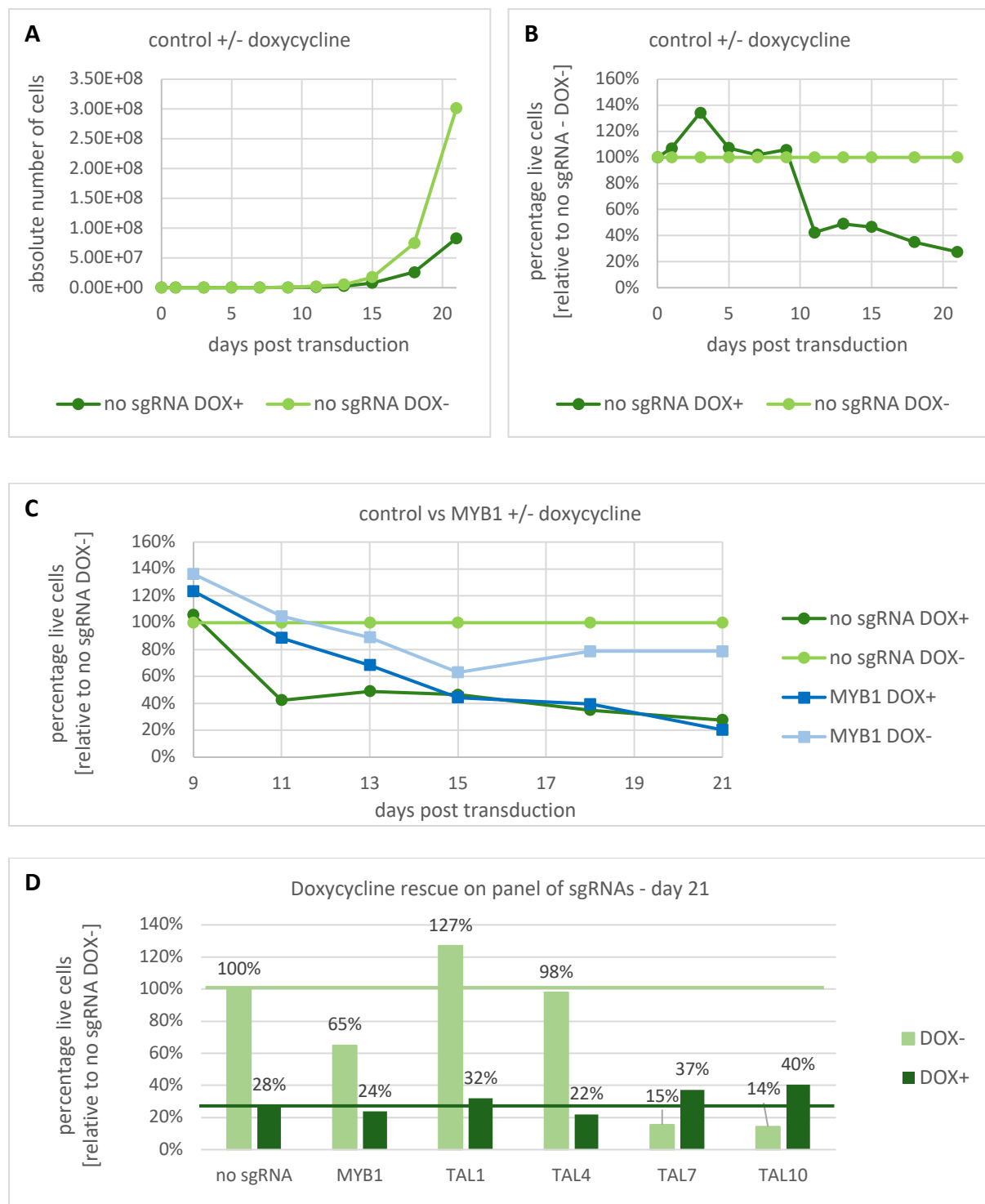


Figure 5.5: The effect of long-term doxycycline exposure on variously edited Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells.

From the day of transduction (day 0), cells were cultured in either selection media without doxycycline (lighter colours), or selection media containing 400 ng/ml doxycycline (darker colours). **(A)** Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells were transduced with control plasmid containing no sgRNA (no sgRNA); live cells were counted every 2 to 3 days with the use of a haemocytometer and the Trypan Blue stain. **(B)** Percentual representation of **(A)**. **(C)** A comparison between the growth of cells transduced with either the control plasmid (no sgRNA, green) or the plasmid containing sgRNA targeting *TAL1* enhancer (MYB1, blue), with or without the addition of 400 ng/ml doxycycline. Cells

were counted with the aid of a haemocytometer and the Trypan Blue stain. NB this graph starts on day 9. **(D)** Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells were transduced with a panel of sgRNAs as detailed in Figure 5.4. Bars represent cell viability on day 21 as assessed by alamarBlue HS, relative to control (no sgRNA) group without addition of doxycycline. DOX+ (dark green bars) – cells cultured in 400 ng/ml doxycycline. Horizontal bars represent 100% of control in DOX+/DOX-. NB this is the same experiment as in Figure 5.4B, with different normalization. **(A)** to **(D)** show results from a single replicate.

In sum, while the addition of doxycycline seemed to have a rescue-like effect on some of the edited samples, linked to it was also a marked decrease in the growth rate of the control cells. The strong FLAG-TAL1 bands we observed on our day 24 Western blot (Figure 5.4C) along with the cytotoxicity that arose in the doxycycline-supplemented group (Figure 5.5) made us hypothesize that the doxycycline concentration used and/or the frequency of doxycycline supplementation might be excessively high.

5.4 Optimization of the use of doxycycline for rescue in Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9

Because the first rescue experiment suggested that the doxycycline concentration used might have been too high, we repeated the doxycycline titration experiment first. We started using a new anti-TAL1 antibody (sc-393287, Santa Cruz) in the lab which, from comparison, seemed to be more sensitive to the FLAG-TAL1 signal than the anti-TAL1 antibody from Abcam (ab155195; data not shown). Employing the new anti-TAL1 antibody showed that much lower concentrations than 400 ng/ml of doxycycline might actually be needed for the rescue, and that the expression of FLAG-TAL1 was still high 72 hours post induction (Figure 5.6), suggesting that in the rescue validation experiment, a supplementation of Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 with doxycycline every 3 days (rather than every 2 days) should be sufficient to keep FLAG-TAL1 at desirable levels.

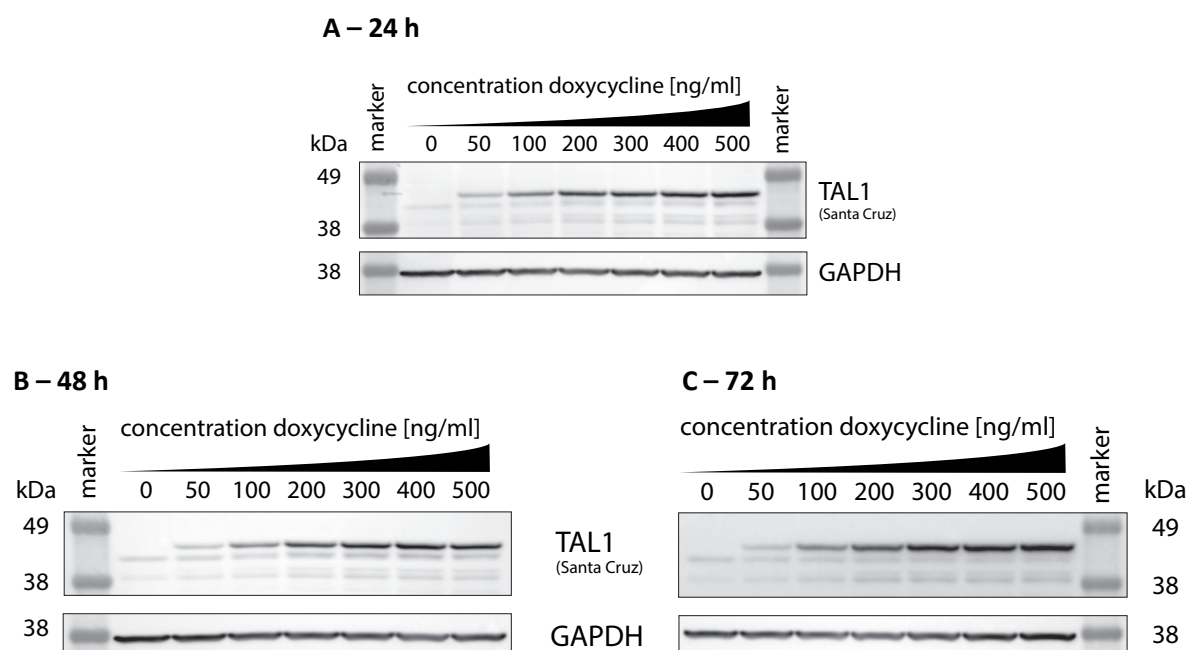


Figure 5.6: Titration of doxycycline for the purposes of FLAG-TAL1 induction in the final line, Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9, II.

Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells were incubated in media containing increasing concentrations of doxycycline. Samples were collected 24 hours **(A)**, 48 hours **(B)** and 72 hours **(C)** post induction. Membranes were probed with anti-TAL1 antibody (sc-393287, Santa Cruz) and anti-GAPDH antibody (60004-1-Ig, Proteintech).

While the Western blots gave us some idea of the range of doxycycline concentrations to consider, the concentration of doxycycline that would serve best in the rescue validation experiment was something that needed to be closer determined in the experimental system itself. For that purpose, we carried out a 15-day long doxycycline-rescue experiment in the polyclonal cell lines that had been edited either by the MYB1-sgRNA, or the TAL10-sgRNA (as obtained from the doxycycline-rescue experiment in Chapter 5.3). These cells, along with a control group, were cultured with a range of doxycycline concentrations.

Proliferation of TAL10-edited cells was positively influenced by concentrations over 50 ng/ml doxycycline, with the best effect produced by a concentration of 200 ng/ml (Figure 5.7A). In contrast, however, doxycycline concentrations of 200 and 400 ng/ml influenced the proliferation of control and MYB1-edited cells most negatively of all tested concentrations. Concentrations of doxycycline below 50 ng/ml seemed to initially encourage the growth of the control, although not MYB1-edited cells (Figure 5.7A).

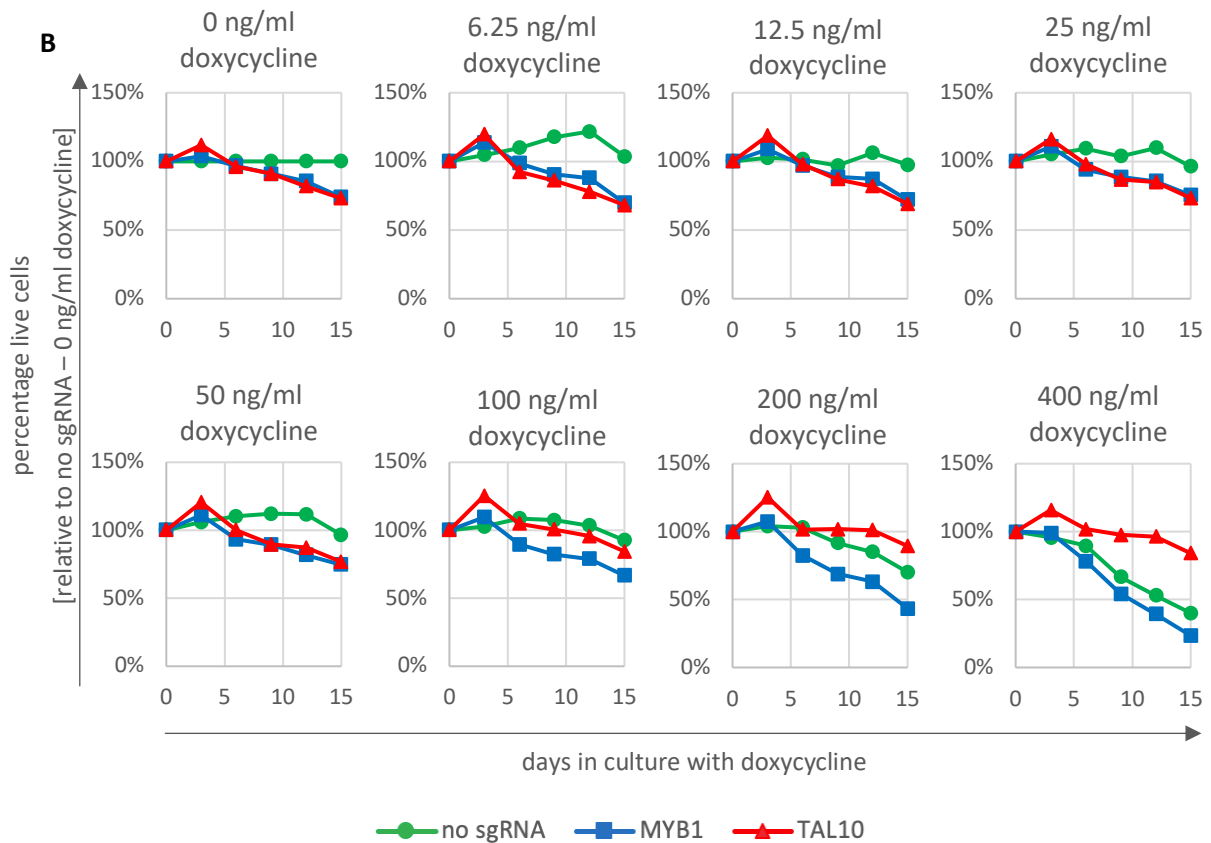
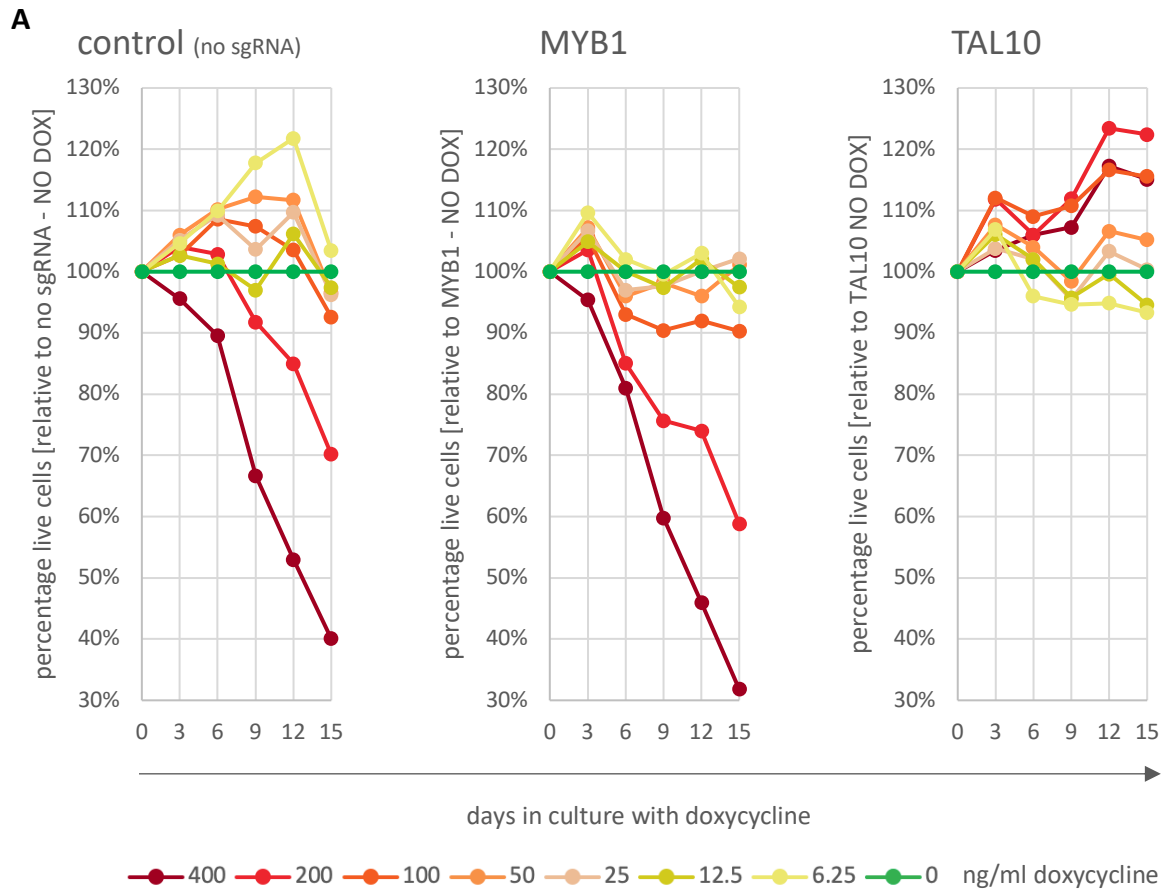


Figure 5.7: Proliferation of edited Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells with a range of doxycycline concentrations in selective media.

Experiment was done with cells edited with an sgRNA targeting MYB-binding site at the *TAL1* enhancer (MYB1), cells edited with an sgRNA targeting the *TAL1* gene (TAL10) and a plasmid without an sgRNA (no sgRNA) as control. This experiment started 30 days after transduction with the appropriate sgRNAs. Prior to this experiment, edited cells were kept in 400 ng/ml doxycycline, lowered to 200 ng/ml doxycycline for one passage immediately preceding the start of this experiment. Cell viability was measured every 3 days with alamarBlue HS and is expressed as **(A)** percentage of live cells relative to the 0 ng/ml doxycycline sample within each group, grouped by the type of cell line editing, and **(B)** percentage of live cells relative to the 0 ng/ml doxycycline sample in the control (no sgRNA) group, shown as single concentration graphs. These results are from a single replicate of a single experiment.

Single concentration graphs (Figure 5.7B) highlight the relationships between all the samples, as they are all normalized to the viability of the control sample cultured without doxycycline. Looking at the final point (day 15) for simplicity, the increasing doxycycline concentrations have elicited a spectrum of responses. In the relationship between the control and the TAL10-edited sample, there seemed to be doxycycline concentrations ('low', up to 50 ng/ml) that made little or no difference to the relationship, a doxycycline concentration that brought the viability rate of the two samples close together (100 ng/ml) and concentrations of doxycycline ('high', over 200 ng/ml) that caused the TAL10-edited sample to outgrow the control sample. However, there was no such clear relationship between the MYB1-edited sample and the control (Figure 5.7B).

A plausible reason for this lack of a spectrum of effects in the relationship between the MYB1-edited sample and the control was that we were looking in the wrong range of doxycycline concentrations. The *TAL1* enhancer knockouts tend to have more *TAL1* expression preserved compared to the *TAL1* exon knockouts, as we had repeatedly observed from our Western blotting (e.g. Figure 4.7C in Chapter 4.5). We therefore argued that we might need to look into a lower range of doxycycline concentrations to find the concentration of doxycycline optimal for the rescue of cells with *TAL1* enhancer mutations. In the next experiment, we set off to explore the effects of various doxycycline concentrations in the range between 0 and 50 ng/ml on the MYB1-edited Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells.

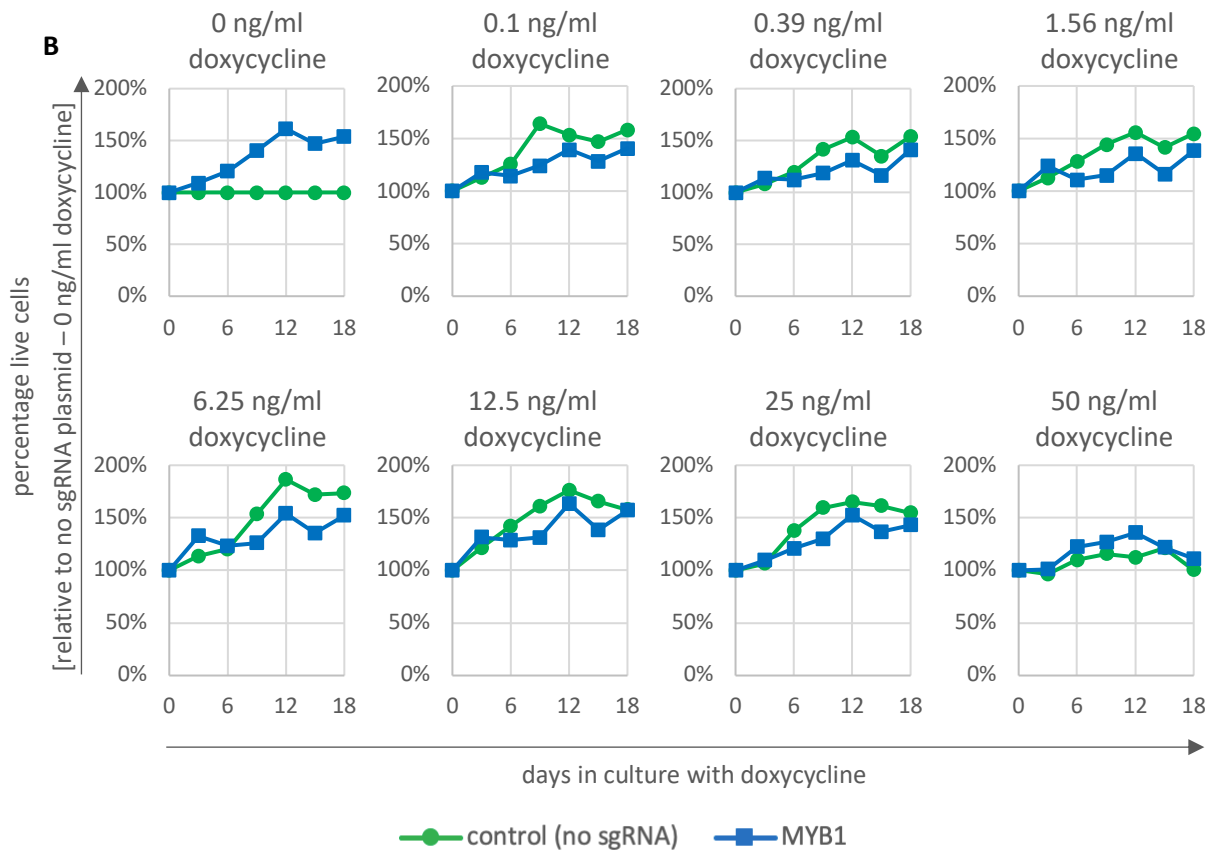
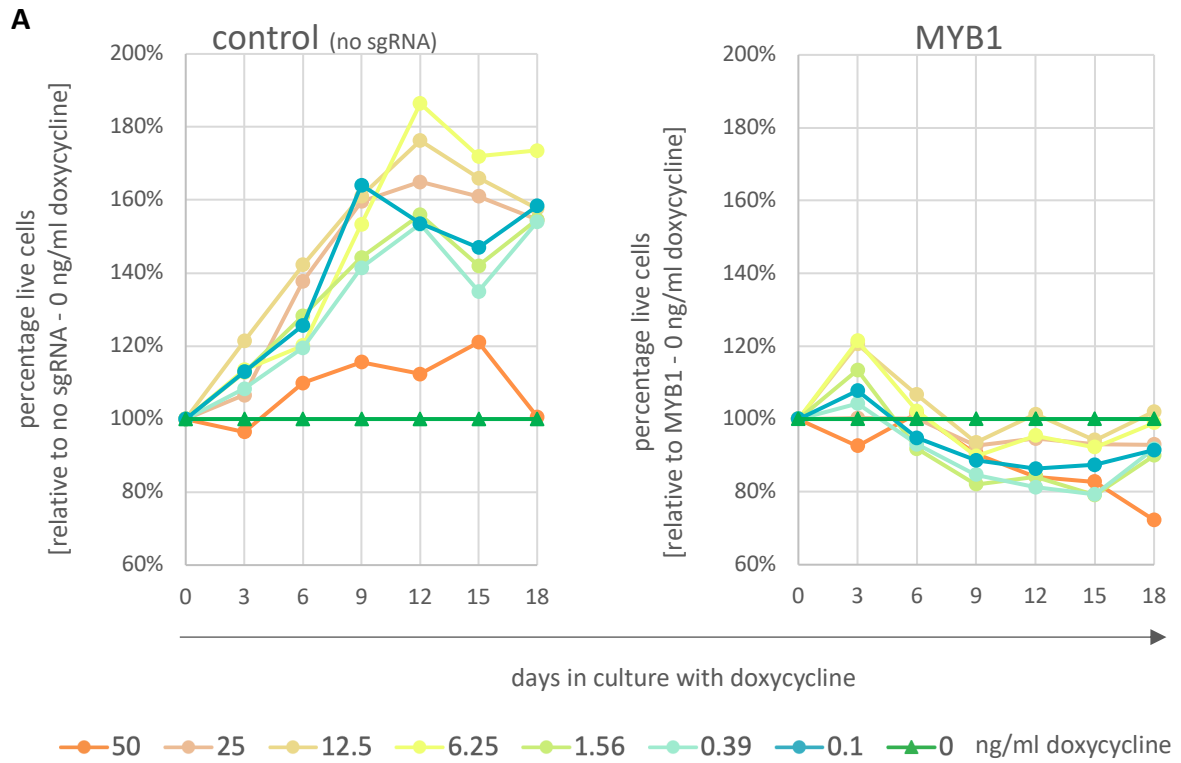


Figure 5.8: Proliferation of edited Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells with a range of doxycycline concentrations in selective media.

Experiment was done with cells edited with an sgRNA targeting MYB-binding site at the *TAL1* enhancer (MYB1) and a plasmid without sgRNA (no sgRNA) as control. This experiment started 46 days after transduction with the appropriate sgRNAs. Prior to this experiment, edited cells were kept in 400 ng/ml doxycycline for the first 26 days, then 200 ng/ml doxycycline for 4 days and without doxycycline for the remaining 16 days. Cell viability was measured every 3 days with alamarBlue HS and is expressed as **(A)** percentage of live cells relative to the 0 ng/ml doxycycline sample within each group, grouped by the type of cell line editing, and **(B)** percentage of live cells relative to the 0 ng/ml doxycycline sample in the control (no sgRNA) group, shown as single concentration graphs. These results are from a single replicate of a single experiment.

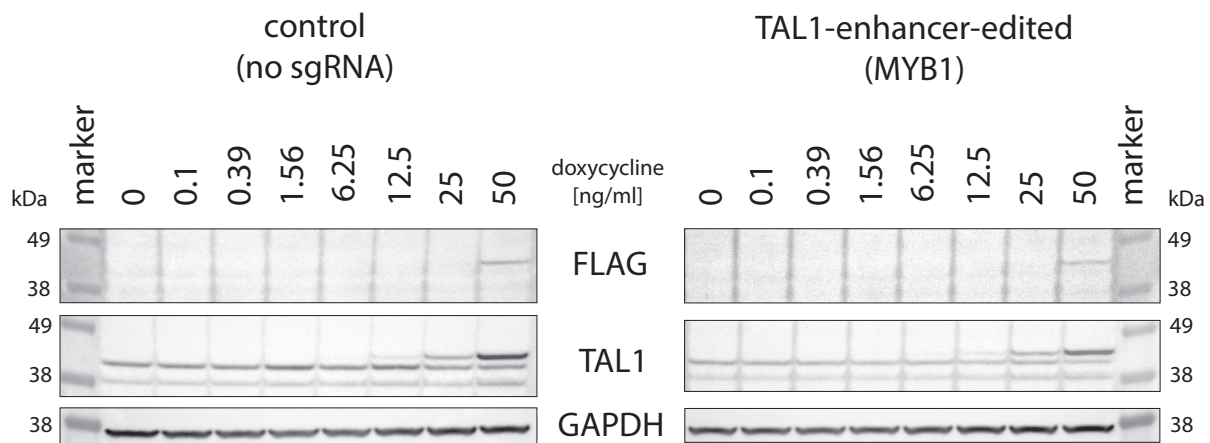


Figure 5.9: TAL1 expression in edited Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells cultivated with a range of doxycycline concentrations in selective media.

Cells were treated as described in Figure 5.8 legend. Western blots were performed on day 18 from the start of the experiment, and obtained using an anti-FLAG antibody (66008-3-Ig, Proteintech), anti-TAL1 antibody (sc-393287, Santa Cruz) and anti-GAPDH antibody (60004-1-Ig, Proteintech).

As in the previous experiment, we worked with cells that were already edited – a polyclonal population generated in the most recent CRISPR/Cas9 experiment (Chapter 5.3). Yet again, none of the selected doxycycline concentrations seemed to be able to boost the viability rate of the MYB1-edited sample, which was in stark contrast with the behaviour of the control sample (Figure 5.8A). Doxycycline concentrations between 12.5 and 50 ng/ml seemed to produce a similar response in the MYB1-edited and the control cells (Figure 5.8B). The expression of FLAG-TAL1 was at the expected levels in both the polyclonal cell lines (Figure 5.9). However, most unexpectedly, in this experiment, the MYB1-edited

cells cultured without any addition of doxycycline were growing faster than the control cells (Figure 5.8B).

While the source of the edited cells used for both of the two most recently discussed experiments was the same (i.e. the CRISPR/Cas9 rescue experiment in Chapter 5.3), the two experiments differed, apart from the range of doxycycline concentrations used, in the treatment of the cells prior to the first day of the experiment. Prior to the first set of experiments (Figure 5.7), cells were cultured in rather high doxycycline concentrations – 400 ng/ml doxycycline for 26 days followed by 200 ng/ml for 4 days; in case of the second set of experiments (Figure 5.8), cells were kept without doxycycline for 16 days before the start of the cultivation. Furthermore, while both of these experiments used already edited polyclonal cell lines, the first experiment used cells that were cultured for about 4 weeks since the introduction of the sgRNAs, the second experiment used cells that were cultured for about 6 weeks since the transduction. These were practical implications of the experimental timelines that we work with.

Over the course of the experiments with the edited cells, it became clear that the inducible FLAG-TAL1 rescue system in the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line is very sensitive to the particulars of the doxycycline treatment. Moreover, another possibility that was brought into consciousness was that over time, the edited polyclonal cell lines may undergo a selection process where cells with the least disruptive mutations gradually take over the initially more heterogenous population of cells.

To address the potential problems stemming from these observations, we decided to test out selected doxycycline concentrations within the experimental system of the CRISPR/Cas9 rescue validation (Figure 5.10), rather than to keep testing the already edited cells. The concentration of 200 ng/ml doxycycline was selected as the best concentration to rescue the TAL10-edited sample, 25 ng/ml doxycycline was in the middle of the range of concentrations that worked best for the MYB1-edited cells, and 6.25 ng/ml doxycycline was chosen as a low boundary, a concentration which boosted the viability of the control sample the most, while affecting the other two groups very little. Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells were transduced with either the TAL10-sgRNA, the MYB1-sgRNA or a control plasmid, and the cultures were supplemented with doxycycline on the day of transduction. The proliferation of these cells was then monitored over 3 weeks, with alamarBlue HS assays conducted on day 13 (Figure 5.10A and C) and on day 22 (Figure 5.10B and D). Graphs show the cell viabilities on the given days relative to SCR1. This control was reintroduced for a new comparison in a transduction experiment in the newly established cell line – the data that led us to select the plasmid without an sgRNA came from a nucleofection experiment in Jurkat cells – and SCR1 proved a better control in the context of the transduction CRISPR/Cas9 experiment in the new cell line.

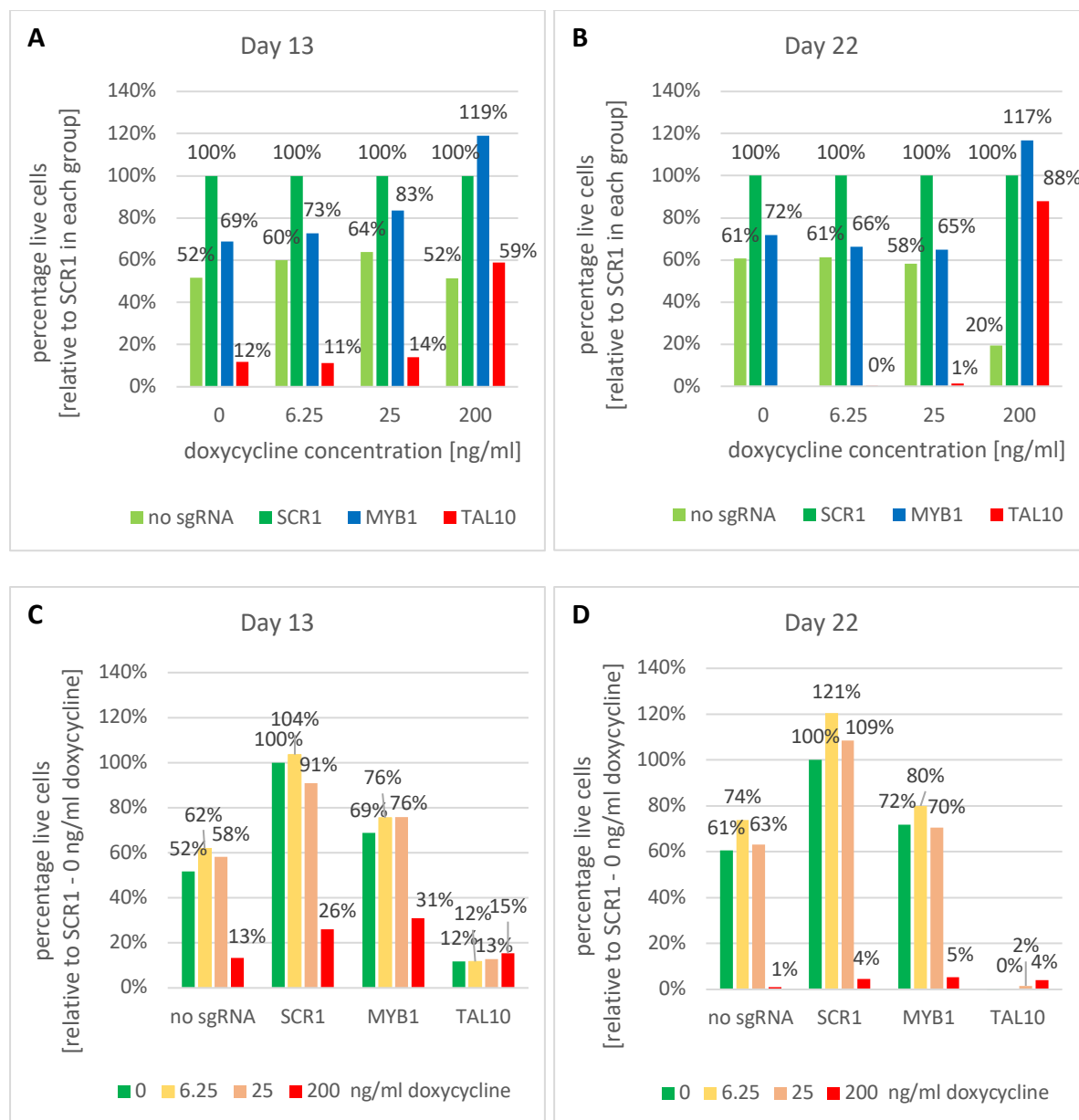


Figure 5.10: Proliferation rates of variously edited Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells cultivated in a range of doxycycline concentrations.

Cells were cultured in selection media with the relevant doxycycline concentration (0 – 200 ng/ml) from the day of transduction (day 0) with plasmids containing sgRNAs targeting either *TAL1* gene (TAL10) or *TAL1* enhancer (MYB1), or with control plasmids containing either a scrambled sgRNA (SCR1), or no sgRNA (no sgRNA). Cells were counted on day 13 (A, C) and day 22 (B, D) after transduction using alamarBlue HS. (A, B) Graphs show the percentage of live cells in each sample normalized to SCR1 in each group defined by a specific doxycycline concentration in selection media. (C, D) Graphs show percentage of live cells in each sample relative to the viability of the SCR1 – 0 ng/ml doxycycline sample. (A) to (D) show results from a single replicate. (A) and (C), as well as (B) and (D), are using the same data presented in a different form.

In Figure 5.10A and B, viability data was normalized to the SCR1 sample viability in each group defined by doxycycline concentration. These graphs confirm that 200 ng/ml was the best concentration of doxycycline for the rescue of the TAL10-edited sample. In this format of the experiment, the results also suggested this concentration might be the best for the rescue of the MYB1-edited sample, especially over a longer period of 22 days (Figure 5.10B), where the supplementation of 200 ng/ml doxycycline was the only tested condition that produced a clear rescue effect (117% of SCR1).

In Figure 5.10C and D, data from the same experiment was normalized to the SCR1 sample cultured without doxycycline, and the data was grouped according to the type of editing. This view better shows the interplay of the rescue effect and the long-term negative impact of doxycycline on the proliferation rate of the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line. When presented this way, the data highlights that the doxycycline-mediated rescue effect takes place at the heavy cost of a strikingly negative effect of doxycycline on the proliferation of control and MYB1-edited cells (for example, the viability of the SCR1 sample cultured in 200 ng/ml doxycycline for 22 days sinks to 4% of the SCR1 sample untreated with doxycycline, Figure 5.10D).

Taken together, the experiments in this chapter showed that the rescue of proliferation with the doxycycline-inducible FLAG-TAL1 in the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line is very sensitive to the doxycycline concentration used, as well as the particular stage of the experiment evaluated. An accumulation of benign mutations during the course of the experiments might also be at play. This means that all further rescue attempts should be carried out within the context of the transduction experiment, rather than from polyclonal cell lines representing already edited cells, and with a highly consistent doxycycline treatment schedule.

5.5 Using FACS as selection in Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 validation experiment generates a clearer trend, but kills all cells in combination with doxycycline

As the results from the previous chapter showed, the rescue of the proliferation rate of the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line was dependent not only on the concentration of doxycycline, but also on the point in the course of the experiment that we assessed it. These sensitivities, we realized in the light of our observations in the previous chapter, might have been in part due to the combined effect of hygromycin and doxycycline. Both of these agents have both a positive *and* a

negative effect on the proliferation of Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells in the rescue validation experiment. The early positive effect of hygromycin on the proliferation rate of *TAL1* knockouts (Figure 4.22 in Chapter 4.13) results in the requirement of a 2- to 3-week long cultivation for the system to reach an equilibrium state where the relative viability of the cells correlates with the toxicity of the editing. At the same time, such long cultivation of Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 in doxycycline is toxic to the cells. Therefore, the hope arose that removing the hygromycin selection could provide clearer rescue results in a shorter timeframe.

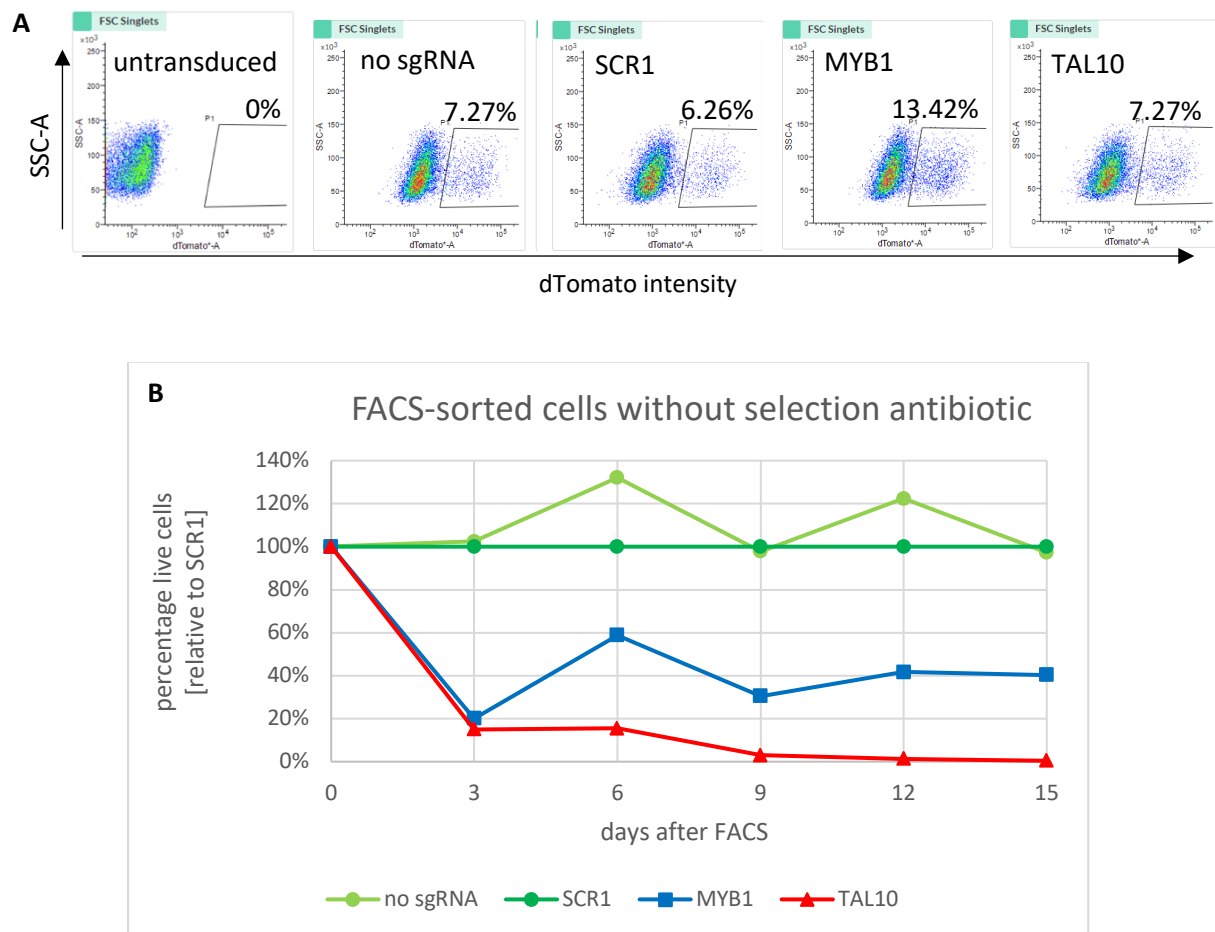


Figure 5.11: Proliferation of Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 after TAL1 challenge and FACS-sorting.

Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 were transduced with sgRNAs targeting the *TAL1* gene (TAL10) or *TAL1* enhancer (MYB1), or one of the negative controls, scrambled sgRNA (SCR1) or a plasmid containing no sgRNA (no sgRNA). 5×10^4 transduced cells were collected 2 days after transduction using FACS, and passaged regularly to ensure optimal cell density. **(A)** Flow cytometry images of the transduced cells 2 days after transduction, at the time of FACS. Y-axis shows the distribution of dTomato fluorescence intensity (dTomato is a fluorescent marker present in the sgRNA plasmids). **(B)** Graph shows percentage of live cells in the variously edited samples relative to SCR1, starting on the day of FACS (day 0). Results show data from a single replicate.

To avoid using the hygromycin selection, we used the other marker present in the sgRNA plasmids, the dTomato orange fluorescent protein, to sort transduced cells using FACS. We had previously abandoned this method but had since optimized the transduction protocol for a higher transduction efficiency, which addressed the primary issue we had with the method (low cell numbers, Chapters 4.5 and 4.6).

To test the feasibility of this approach, Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells were transduced with the TAL10, MYB1, SCR1 and no sgRNA lentiGuide-Hygro-dTomato plasmids and FACS-sorted 48 hours later (Figure 5.11A). No addition of hygromycin was made; cells were counted and passaged regularly. Proliferation reached expected levels 3 days after FACS (5 days after transduction) with both the TAL10- and MYB1-edited cells gravely affected in growth compared to the controls (Figure 5.11B). Taking the hygromycin selection process out of the equation seemed to simplify the relationships between the differentially edited cells enormously, and shorten the timeframe in such a manner that the longer-term negative effects of doxycycline might potentially not even have to come into play.

However, the transduction efficiencies recorded in the latest experiment (6.26% for SCR1, done with 10 U of viral stock, Figure 5.11A) were considerably lower than in the previous optimization experiments (for example, 28.7% SCR1-positive cells after a transduction with 7 U of viral stock, Figure 4.8 in Chapter 4.6). To further improve the chances of the promising potential of using FACS within the doxycycline-rescue experiment, we decided to extensively optimize transduction in the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line again, examining the influence of previously untested factors. The results of this optimization are provided in Chapter 5.7. Based on this process, we decided to use 293FT cells (a HEK293-derived cell line developed for the generation of high viral titres) for lentiviral packaging instead of the HEK293T cells used previously, transduce with freshly prepared virus, leave the virus in the cell suspension until FACS, and sort on day 3 rather than day 2 after transduction.

We proceeded to carry out a doxycycline rescue experiment in Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells with a FACS-sorting step. Cells were transduced with fresh 293FT-generated virus and sorted on day 3 post transduction (Figure 5.12), as the optimization experiment suggested was best for efficiency (Figure 5.14 in Chapter 5.7). Cells were kept in a range of doxycycline concentrations from the day of transduction. While FACS-sorted cells unsupplemented with doxycycline started expanding within several days after the sorting, none of the cells supplemented with doxycycline made a recovery within the scope of the experiment (Figure 5.12A and B). This strongly indicated that the addition of doxycycline to Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells distressed by the FACS-sorting had a profound negative effect on the proliferation of these cells, and effectively prevented us from

using FACS-sorting as a strategy to select transduced Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells in the context of the doxycycline rescue validation experiment.

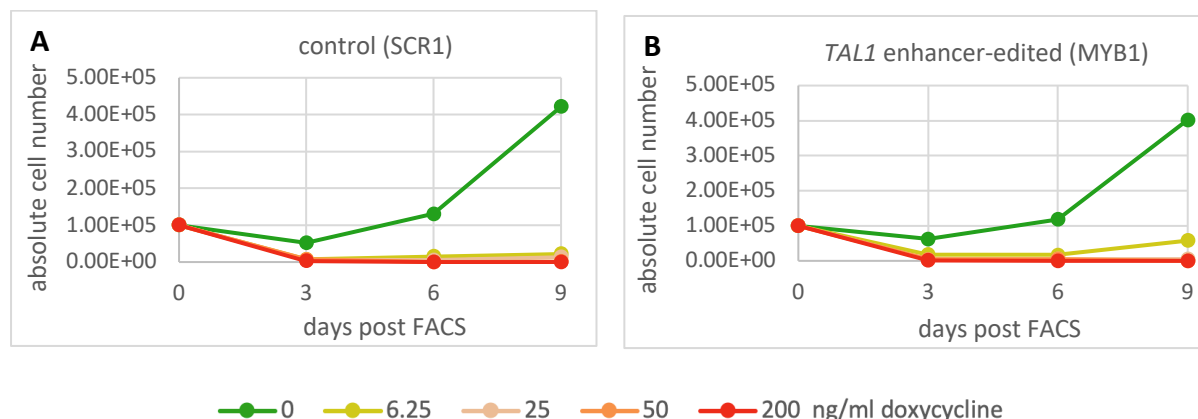


Figure 5.12: Proliferation of control or *TAL1* enhancer-edited, FACS-sorted Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells with doxycycline.

Cells were transduced with fresh 293FT-generated virus containing the RNA copy of the gene for dTomato fluorescent protein and either a control scrambled sgRNA (SCR1), or an sgRNA targeting the MYB-binding site in the *TAL1* enhancer (MYB1). Cells were cultivated in various concentrations of doxycycline (0 – 200 ng/ml) from the day of transduction. 3 days upon transduction (day 0 on the graphs), FACS was performed to collect 1×10^5 dTomato positive cells (as detailed in Figure 5.16 in Chapter 5.7). Cells were then counted with the use of Trypan Blue and a haemocytometer and passaged every 3 days. **(A, B)** Absolute numbers of living cells in each sample, control cells transduced with a scrambled sgRNA in **(A)** and *TAL1* enhancer (MYB1)-edited cells in **(B)**. **(A, B)** Results are from a single replicate.

It also is worth noting that the differences between the cell numbers in the SCR1 control group and the MYB1-edited group culture without doxycycline were very small (MYB1 viability on day 9 was 95% of SCR1 viability), which is in stark contrast with the previous experiment done with HEK293T-generated virus (MYB1 viability on day 9 was 30% of SCR1 viability, Figure 5.11).

In sum, sorting transduced Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells on FACS for the doxycycline rescue experiment, despite its initial promise, did not seem to be a viable strategy for running the rescue experiment (and the future screen), as the combined pressures of the FACS process and the doxycycline presence interfered strongly with Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell survival.

5.6 Doxycycline partially rescues proliferation rates in Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells of the final line upon TAL1 downregulation

Because our data from the FACS-based doxycycline rescue experiment (Figure 5.12 in Chapter 5.5) indicated a lack of feasibility for this approach, we decided to return to using hygromycin for the selection process. We wanted to show a reproducible doxycycline rescue in Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells that would normally be retarded in growth by anti-*TAL1* CRISPR-editing. We hoped to create a robust rescue protocol that could be used as a basis for the CRISPR/Cas9 saturating mutagenesis of the *TAL1* enhancer, where the doxycycline-rescued cells would form the control group with equal representation of control and TAL1-challenging sgRNAs.

Using the optimized transduction protocol (described in Chapter 4.13), Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells were transduced with the MYB1, TAL10 or SCR1 lentiGuide-Hygro-dTomato plasmids and subjected to 250 µg/ml hygromycin selection. The cells were treated with 0, 200 or 400 ng/ml doxycycline from the day of transduction, regularly passaged and assessed for viability using alamarBlue HS 22 days after transduction (Figure 5.13). In the context of this experiment, we again observed the complicated interplay of the rescue effect (Figure 5.13A) and the longer-term negative impact of doxycycline on the proliferation of Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 (Figure 5.13B).

Firstly, Figure 5.13A shows the viability of the cells in each tested sample relative to the viability of the non-targeting SCR1 control in each doxycycline group (0, 200 or 400 ng/ml). Grouping our data in this way is important because it mimics the fashion in which the planned CRISPR/Cas9 screen experiment will be carried out, with all cells either untreated or treated with a single doxycycline concentration. This grouping shows us that in the 0 ng/ml doxycycline group, the non-targeting SCR1 control group was more viable on day 22 than both the MYB1-edited cells (87% of SCR1) and TAL1-edited cells (1% of control), which is in keeping with our expectation. When 200 or 400 ng/ml doxycycline was added to these cells, both MYB1-edited and TAL10-edited cells became more viable than the SCR1 control, the TAL10-edited cells more so than the MYB1-edited cells. This data suggests that the rescue was happening, which is in keeping with the general trends observed in previous experiments throughout this chapter (e.g., Figure 5.4 in Chapter 5.3 and Figure 5.10 in Chapter 5.4), although the proliferation of the individual samples was largely varied, in contrast to our initial expectation of an equal proliferation in all samples within a single doxycycline concentration group (Chapter 5.1).

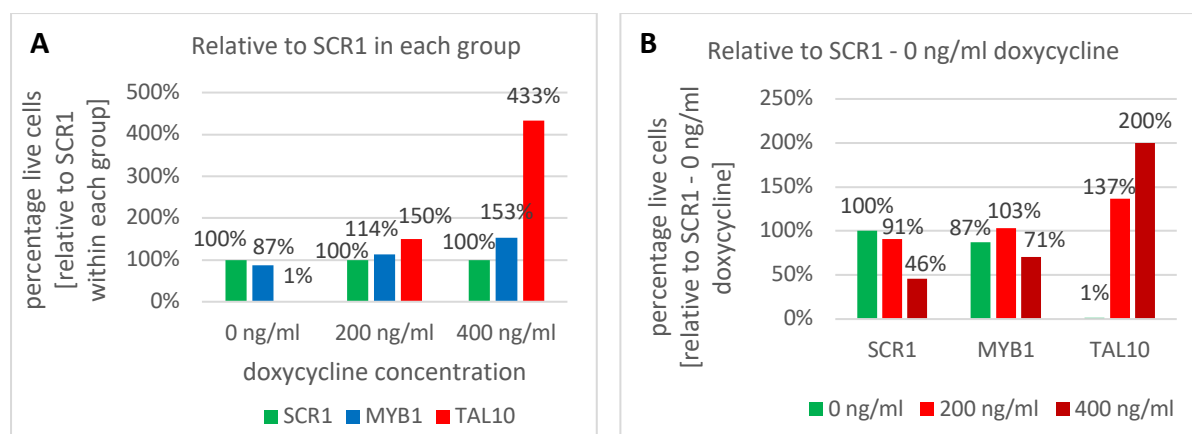


Figure 5.13: The effects of a range of doxycycline concentrations on the proliferation of variously edited Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells.

Cells were cultured with the relevant doxycycline concentration (0, 200 or 400 ng/ml) from the day of transduction (day 0) with sgRNAs targeting either the *TAL1* gene (TAL10) or *TAL1* enhancer (MYB1), or a non-targeting scrambled sgRNA (SCR1). Viral particles were generated in 293FT cells and used fresh after concentration, medium was changed a day after transduction to introduce hygromycin selection, then every 3 days to maintain doxycycline levels. Viability of cells was assessed using alamarBlue HS reagent on day 22 and is expressed relative to SCR1 within each group (**A**) or relative to SCR1 in the 0 ng/ml doxycycline group (**B**). Results are from a single replicate.

Secondly, the negative effect of doxycycline on unedited cell proliferation came into plain view when cell viabilities in the tested samples were plotted out as relative to the viability of the sample transduced with non-targeting sgRNA (SCR1) and untreated with doxycycline (Figure 5.13B). The viability of SCR1-transduced cells proliferating in 400 ng/ml doxycycline for 22 days was only 46% of the viability of the sample untreated with doxycycline.

Meanwhile, the absolute number of cells in the TAL10-edited group treated with 400 ng/ml doxycycline was twice as high as in the SCR1 0 ng/ml doxycycline group, which pointed towards the potential existence of at least a third type of doxycycline effect (or a mixture of additional influences) on Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells, apart from the FLAG-TAL1 doxycycline-mediated rescue effect and the cytotoxicity. However, it is worth mentioning that we did not observe this effect previously (Figure 5.10 in Chapter 5.4), where the experimental setup was identical apart from the difference in the packaging cell line used for the production of the virus. On day 22, absolute numbers of TAL10-edited cells treated with 200 ng/ml doxycycline in the previous experiment were 4% those of doxycycline-untreated SCR1 sample (Figure 5.10D in Chapter 5.4), whereas in the most recent experiment (Figure 5.13B), absolute numbers of TAL10-edited cells treated with 200 ng/ml doxycycline were 137% of the absolute numbers in the doxycycline-untreated SCR1 sample. This further testifies to the overall lack of robustness of the whole experimental setup.

An additional issue with this experiment was that the number of surviving cells was unexpectedly low throughout the selection process (as low as 3×10^4 cells in a sample on days 7 – 10). In comparison with earlier experiments with HEK293T-packaged virus, the absolute cell numbers were similar or even lower on any given day. Furthermore, in the two following repeats of this experiment, the transduced cells failed to expand and didn't survive the selection process at all (data not included). This was very surprising, given that we had established that the transduction efficiency of the 293FT-packaged viral preparation was at least 4-fold higher than the transduction efficiency of the HEK293T-packaged virus (Figure 5.14 in Chapter 5.7), and therefore we would expect a 4-fold higher survival rate. Moreover, as one of the requirements for a viable screen is the continued expansion of the control doxycycline-untreated cell lines (Chapter 4.1), as well as control and edited doxycycline-treated cell lines (Chapter 5.1), the experimental setup with the 293FT-generated virus used in this chapter is infeasible, as insufficient cell numbers survive the process.

Overall, the effects of doxycycline on the proliferation of the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line in the context of the rescue validation experiment are, at the very least, two-fold – the negative effect due to doxycycline cytotoxicity on a longer-term scale (>1 week), and the positive effect stemming from the induction of the exogenous FLAG-TAL1, to which the cell line is addicted. The comparison of our attempts at doxycycline rescue led us to conclude that the results are extremely sensitive to a number of factors, including the transduction conditions and the choice of sgRNA, achieved transduction efficiency, doxycycline supplementation regime and concentration, and possibly other unknown factors. Unfortunately, this renders the system impracticable for the use with the planned CRISPR/Cas9 screen, as the effects of the mutations in the *TAL1*-eRNAs on the proliferation of the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line are expected to be very subtle, and would highly likely get covered up by the effects of doxycycline.

5.7 Transduction optimization and evaluation of the new packaging cell line, 293FT

Due to the relatively low transduction efficiencies recorded in the experiment in Chapter 5.5 (e.g., 6.26% for SCR1), we decided to further optimize our transduction protocol to improve the transduction rate for the use with the FACS-based approach to the doxycycline rescue experiment (Chapter 5.5).

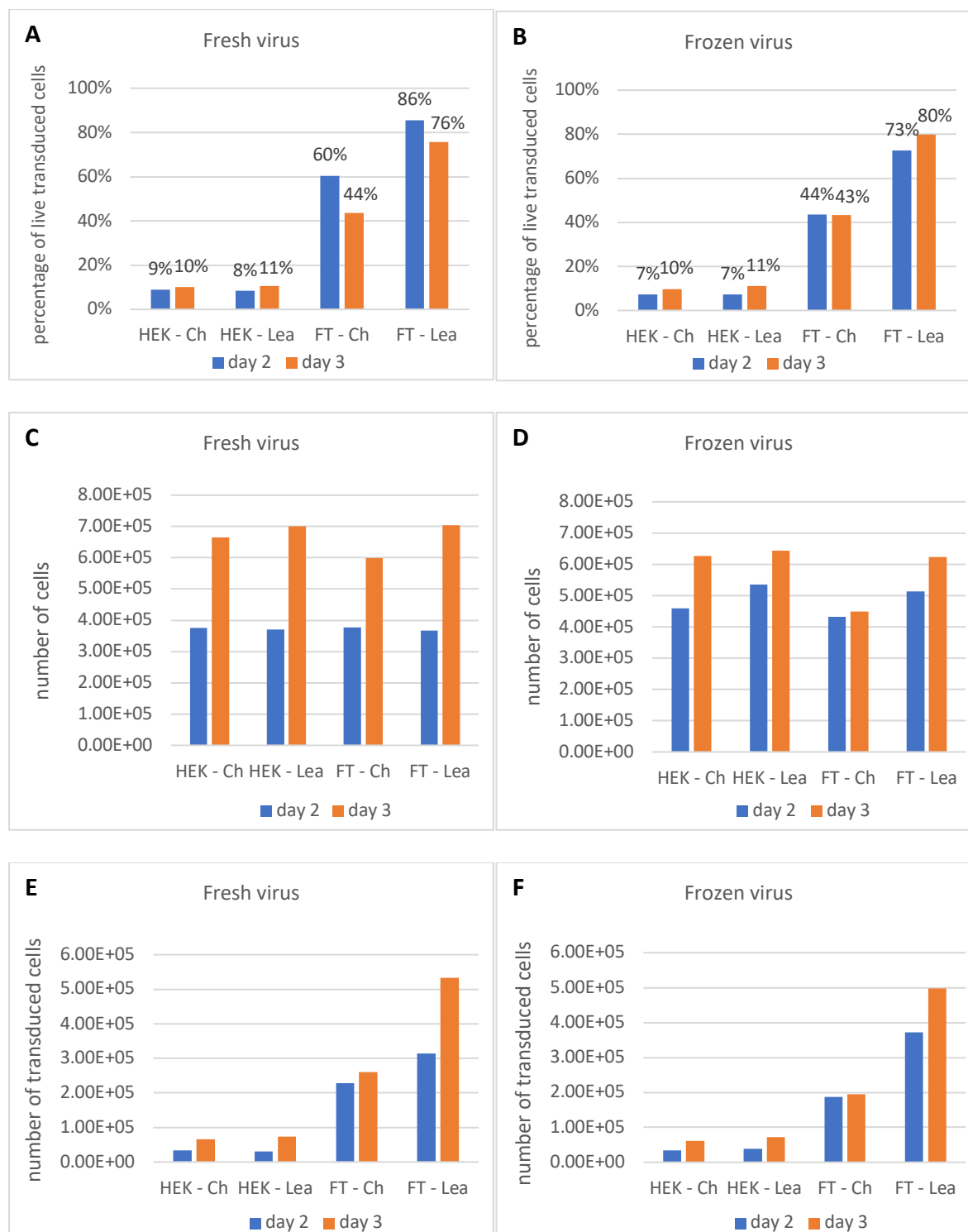


Figure 5.14: Transduction optimization for further FACS-based doxycycline-rescue experiments.

Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells were transduced with a plasmid containing dTomato as a fluorescent marker and a scrambled sgRNA (SCR1). The SCR1 virus was generated either in HEK293T cells (HEK), or in 293FT cells (FT) and was used either immediately (fresh), or frozen for 24 hours at -80°C and used after thawing (frozen). Medium containing virus was either changed on day 1 after transduction (Ch), or left on until the time of flow cytometry (Lea). The same number of arbitrary units, based on virus produced by cm^2 of cells, was used for all transductions. Data was collected on day 2 and 3 post transduction. **(A, B)** Percentage of transduced (dTomato-positive) cells in each sample with

fresh **(A)** or frozen **(B)** virus, as determined by flow cytometry. **(C, D)** Absolute cell numbers in each of the samples with fresh **(C)** or frozen **(D)** virus. Viability was measured using alamarBlue HS. **(E, F)** Absolute numbers of transduced cells in each sample 2 and 3 days after transduction with fresh **(E)** or frozen **(F)** virus. This is a combination of flow cytometry data and alamarBlue HS cell counts. **(A – F)** are data from a single replicate.

This time, we didn't work with an increasing viral titre, but instead tested the influence of different transduction conditions on the transduction efficiency: the difference between a fresh and a frozen viral stock, the difference between changing and not changing media after transduction, as well as the difference between a 2-day and a 3-day incubation period after the transduction. Furthermore, a newly acquired packaging cell line, 293FT (Thermo Fisher), was tested against the routinely used HEK293T cells. 293FT is a HEK293-based cell line genetically modified to yield high viral titres.

To carry out the optimization experiment, 10 U of lentiGuide-Hygro-dTomato-SCR1 virus produced in either HEK293T, or 293FT was transduced into Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells either immediately after concentration with Lenti-X Concentrator, or after 24 hours of storage at -80°C. The viral media was then either changed a day after transduction, according to the protocol used hitherto, or left on until the day of the flow cytometry assay on day 2 or 3 after transduction, exposing the cells to the viral particles for 24 or 48 hours longer than in the former case.

The biggest increase in transduction efficiency was observed in relation to the employment of the new packaging cell line, 293FT (Figure 5.14). While Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells transduced with the HEK293T-generated virus achieved 7 – 11% efficiency, the 293FT-generated virus from the same initial number of packaging cells seemed to achieve transduction efficiencies of 43 – 86% (Figure 5.14A, B). The number of surviving Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells was very similar between the groups (Figure 5.14C and D). A function of the absolute cell numbers in each sample and the transduction efficiency of each of the conditions, the number of transduced cells was highest in the group transduced by fresh 293FT-generated virus on day 3 after transduction, where media wasn't changed 24 hours post transduction (Figure 5.14E). Overall, there was little difference between the use of fresh and frozen viral stocks in terms of the resulting efficiency, and day 3 after transduction provided a larger number of transduced cells than day 2 (Figure 5.14E and F).

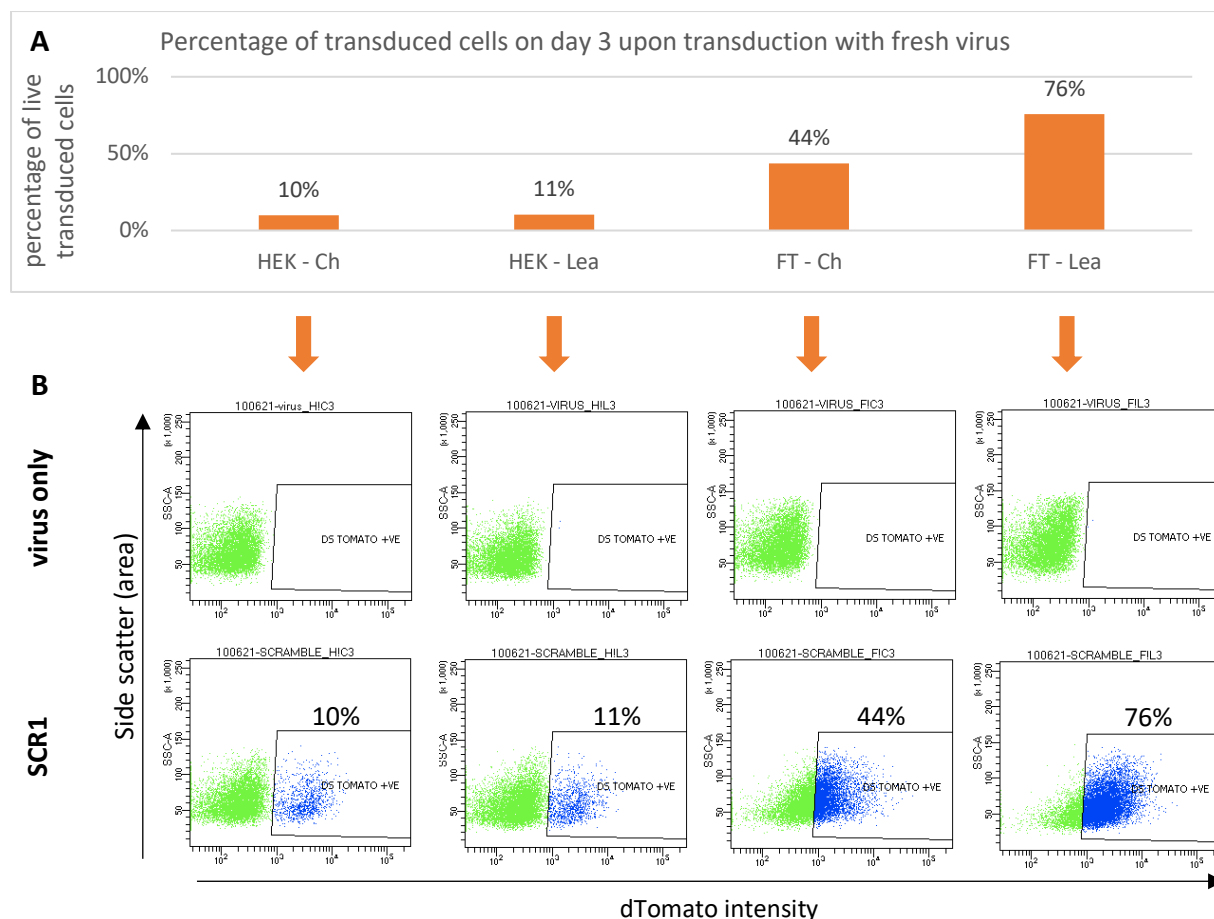


Figure 5.15: Representative results from the transduction optimization experiment as described in Figure 5.14.

(A) Percentage of dTomato-positive Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells in samples transduced, 3 days prior, with fresh viral particles encoding the dTomato fluorescent marker and a scrambled sgRNA (SCR1). Virus was generated in either the HEK293T (HEK) or 293FT (FT) cell line, medium was changed on day 1 post transduction (Ch) or left until day 3 (Lea). This Figure is using the same data as Figure 5.14A. **(B)** Flow cytometry images of cells transduced as described in **(A)** with either SCR1 or with empty viral particles (virus only) as negative control. Fluorescence intensity of dTomato is plotted on the x-axis. Gates for dTomato-positive (ds Tomato +ve) cells were based on negative control (virus only) samples, each of which was treated in the same way as the respective experimental sample.

A closer look at the flow cytometry images from the transduction optimization experiment showed a difference in the appearance of the populations of Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells in the samples which were transduced using the HEK293T-generated virus, and those that were transduced using the 293FT-generated virus (Figure 5.15B). In cells transduced using the HEK293T-generated virus, we observed two distinct populations – a main population of untransduced cells, of low dTomato intensity, and a smaller population of transduced cells with a decisively higher dTomato

intensity (gated and highlighted in blue). In the cells transduced using the 293FT-generated virus, there were no such two distinct populations. Rather, we observed a shift of the whole population towards higher dTomato intensities. When compared to the virus only control, the shift seems to testify to a large percentage of dTomato-positive cells (Figure 5.15B). However, we had previously seen similar shifts in the populations of cells transduced with an increasing viral titre of HEK293T-generated virus (Figure 4.8 in Chapter 4.6), which we then attributed to autofluorescence due to increased stress. However, in the previous experiment, untransduced cells were used as control, whereas here, we compared the samples transduced with the dTomato plasmid to samples transduced with empty viral particles (virus only). This led us to believe that we were looking at dTomato fluorescence, not autofluorescence of stressed cells, as we reasoned that the empty viral vector should cause as much stress as functional viral particles.

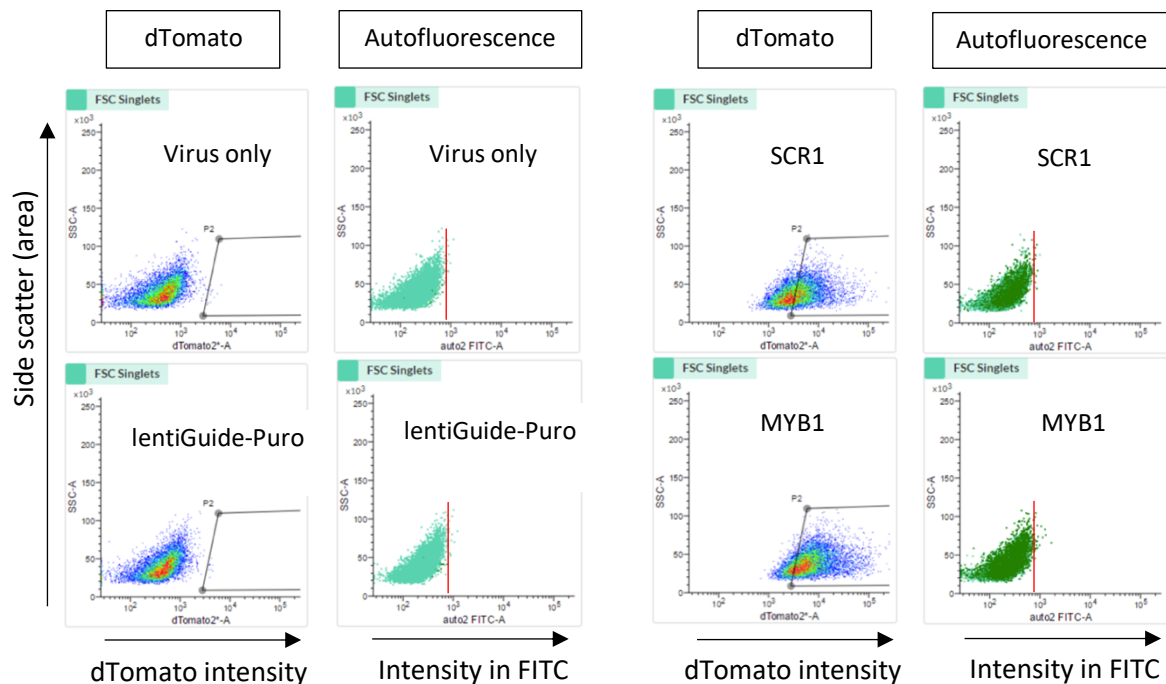


Figure 5.16: 293FT packaging does not result in increased autofluorescence.

Flow cytometry on cells transduced with: empty viral particles (virus only); viral particles bearing a transfer plasmid without a gene for a fluorescent protein (lentiGuide-Puro); non-editing control plasmid coding for a scrambled sgRNA and dTomato (SCR1); sgRNA targeting the MYB-binding site in *TAL1* enhancer and dTomato (MYB1). Flow cytometry data shows the intensity of fluorescence in the dTomato channel and the autofluorescence in the FITC channel. The MYB1- and SCR1-transduced cells were used for the experiment in Figure 5.12 (Chapter 5.5), and the gating shows the area from which cells were collected in the experiment. Gating for dTomato positive cells was first done on negative controls (virus only, lentiGuide-Puro) and then adjusted to include only the top 50% of SCR1 population for collection from FACS. For easy comparison, a red bar is drawn at 8×10^2 AU of fluorescence intensity in all plots showing fluorescence in the FITC channel.

To gain more confidence in this matter, we transduced the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells with either dTomato-encoding plasmids (SCR1 or MYB1) or with non-fluorescent controls (either empty viral particles, or viral particles carrying a plasmid without a fluorescent marker). We then recorded fluorescence of these cells 3 days post transduction in the dTomato channel and in the FITC (fluorescein isothiocyanate) channel. FITC is a common channel where autofluorescence may be observed, while dTomato fluorescence would be undetectable in this channel. Therefore, any shifts in this channel could easily be attributed to autofluorescence (resulting from cellular stress), and warn us that the shift we observed in the dTomato-transduced samples may not be solely due to dTomato fluorescence. We saw the expected shifts in the dTomato channel, but no shifts in the FITC channel. This deepened our confidence that we were looking at a bona fide effect of dTomato fluorescence, rather than an autofluorescence resulting from the experimental stress.

In sum, our optimization led us to start using the 293FT packaging cell line instead of the previously used HEK293T, and to FACS-sort on day 3 rather than on day 2 after transduction. We decided to leave the viral particles in the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell suspension until the day of FACS, and to keep using fresh virus in our future transduction experiments, as per the usual practice, even though the difference between the efficiency achieved by the transduction of fresh vs frozen virus was minimal.

5.8 Validation of doxycycline rescue using CRISPR interference (CRISPRi)

Repeated failures to develop a viable protocol for the doxycycline rescue of CRISPR/Cas9-edited cells drove us to look for alternative avenues to demonstrate the potential of the doxycycline rescue in a more robust fashion. While it was becoming clear that we would not be able to use the doxycycline rescue in the planned pooled CRISPR/Cas9-mediated saturating mutagenesis of the *TAL1* enhancer in the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line, we nourished a hope that we could still use the doxycycline rescue feature for later validation experiments in an arrayed format (more on this in Chapter 6). The general trends in all the *TAL1*-CRISPR/Cas9 doxycycline rescue experiments we have done strongly suggested that the doxycycline rescue was functional, but we wanted to evidence this in a more robust and reproducible fashion.

While siRNAs served us very well in showing the *TAL1* dependency in the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line in Chapter 4.8, demonstrating the potential of FLAG-TAL1 doxycycline rescue with

these would be problematic, as our siRNAs target not only the endogenous *TAL1*, but also the exogenous doxycycline-inducible FLAG-*TAL1*. Instead, we decided to employ *TAL1*-CRISPR interference (CRISPRi). CRISPRi has proved to be a robust approach to transcriptional regulation, and with the application of a set of rules to sgRNA design/selection, the method typically results in knockdown efficiencies of 80 – 99% (reviewed in Chapter 1.4.2.3.1).

Unfortunately, *TAL1*-CRISPRi could not be carried out in the final cell line, Jurkat-Tet3G-TRE3G-FLAG-*TAL1*-Cas9, because it expresses Cas9, which would compete for the experimental sgRNAs with the dCas9-KRAB in CRISPRi. On that account, we decided to attempt the *TAL1*-CRISPRi in the second intermediate cell line, Jurkat-Tet3G-TRE3G-FLAG-*TAL1*, which only differs from the final line by the lack of Cas9 expression. This difference should not have any impact on the functionality of the inducible FLAG-*TAL1*, and the results of a CRISPRi validation in the intermediate cell line could be therefore reasonably extended to the final line.

For the *TAL1*-CRISPRi in the intermediate cell line, the Lenti-(BB)-EF1a-KRAB-dCas9-P2A-BlastR (Addgene plasmid #118154, unpublished) plasmid was used, containing a gene for blasticidin resistance, a gene for dCas9-KRAB and an sgRNA scaffolding. We cloned the following guide sequences into the plasmid: (1) PRO1 and PRO2, targeting the core promoter of *TAL1* (as defined in (Andersson and Sandelin, 2020)); (2) PRO3 and PRO4, targeting the 5'-UTR of *TAL1* (top anti-*TAL1* CRISPRi sgRNAs identified by CRISPick (Doench et al., 2016; Sanson et al., 2018)); (3) MYB1, targeting the MYB-binding site at the *TAL1* enhancer in Jurkat cells (Mansour et al., 2014); and (4) SCR1, a scrambled non-targeting sequence (Lawhorn et al., 2014). We also planned to use a plasmid without sgRNA (no sgRNA) as an alternative control.

Due to former experience with hygromycin in the final cell line, the intermediate cell line, Jurkat-Tet3G-TRE3G-FLAG-*TAL1* (clone 85-15), was tested for its sensitivity to blasticidin. Wild-type Jurkat and 85-15 cells were grown in media supplemented with 1 – 10 µg/ml blasticidin for 7 days, and viability was measured using the alamarBlue HS Reagent on day 3, 5 and 7. The Jurkat-Tet3G-TRE3G-FLAG-*TAL1* cell line exhibited a very similar sensitivity to the antibiotic as the maternal cell line (Figure 5.17A and B).

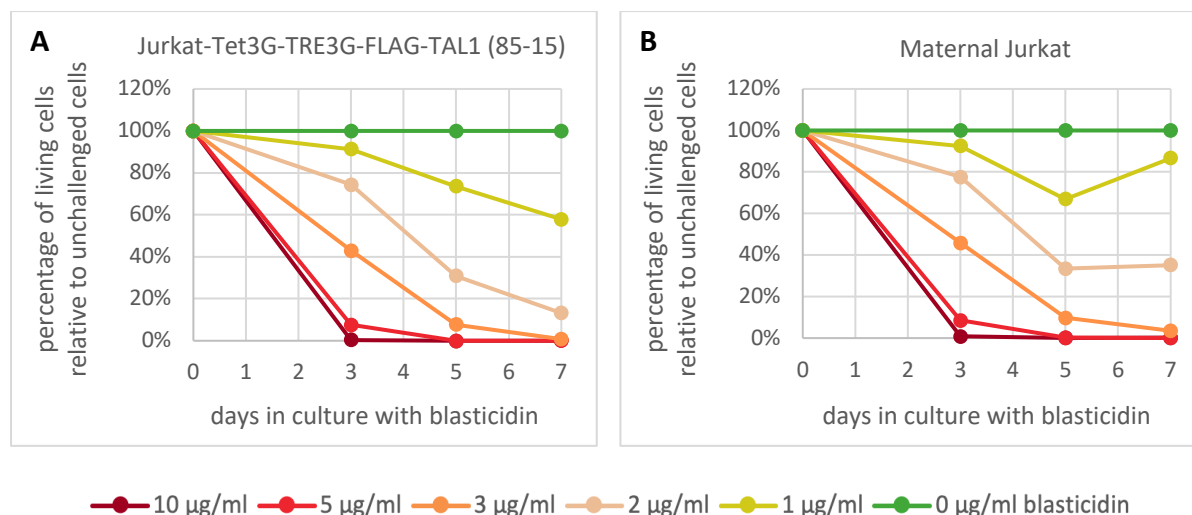


Figure 5.17: The sensitivity of Jurkat-Tet3G-TRE3G-FLAG-TAL1 to blasticidin.

Proliferation of **(A)** Jurkat-Tet3G-TRE3G-FLAG-TAL1, the intermediate cell line, and **(B)** Jurkat, the maternal cell line, with a range of blasticidin concentrations. Cell viability was assessed using alamarBlue HS reagent. Graphs show data from a single replicate.

Having demonstrated the expected behaviour under blasticidin selection, Jurkat-Tet3G-TRE3G-FLAG-TAL1 cells were then transduced with the CRISPRi plasmids and their viability was monitored with alamarBlue HS between day 1 and day 26. Surprisingly, apart from the MYB1-CRISPRi group, the *TAL1*-CRISPRi didn't elicit the expected drop in proliferation (Figure 5.18A). We wondered whether this might have been down to blasticidin creating a similar early advantage for transduced cells as hygromycin was in the case of the *TAL1*-CRISPR in the final line, and to test this hypothesis, we used 1×10^5 cells from each of the samples on day 13 post transduction to set up an independent culture (Figure 5.18B). The trends, however, remained similar, and only the MYB1-CRISPRi led to a decided drop in proliferation rate in both the experimental setups (Figure 5.18A and B).

To investigate whether CRISPRi reduced TAL1 expression according to the primary expectation, we carried out a Western blot for all groups on day 26 of the experiment. The Western blot confirmed TAL1 downregulation in all the experimental groups apart from PRO2 (Figure 5.18C), which made the lack of effect on proliferation seem mystifying.

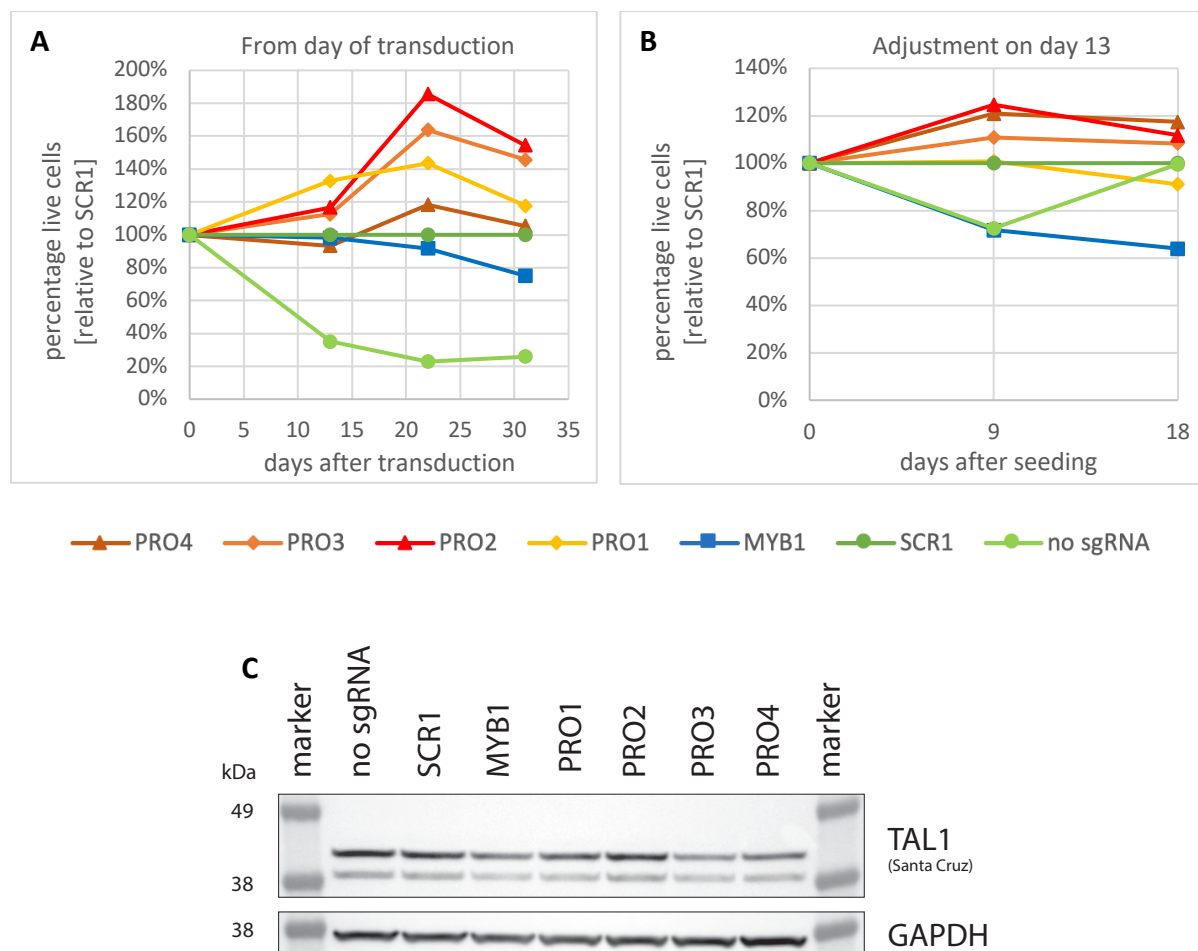


Figure 5.18: *TAL1*-CRISPR interference in Jurkat-Tet3G-TRE3G-FLAG-TAL1 cell line using single sgRNAs.

Cells were transduced with viral particles encoding a blasticidin resistance gene, dCas9-KRAB and various sgRNAs for *TAL1*-CRISPRi, targeting either different sites in the *TAL1* promoter (PRO1 – PRO4) or the MYB-binding site at the *TAL1* enhancer (MYB1); negative controls were cells transduced with virus encoding the blasticidin resistance marker, dCas9-KRAB and either scrambled sgRNA (SCR1) or no sgRNA (no sgRNA). Cells were supplemented with 5 µg/ml blasticidin on day 1 after transduction. **(A)** Cells in all groups were counted using alamarBlue HS reagent 13, 22 and 31 days post transduction. **(B)** 1×10^5 cells from each of the samples on day 13 upon transduction (i.e., cells selected with blasticidin for 12 days, therefore with presumed CRISPRi) were used to start an independent culture, and viability was measured 9 and 18 days into the experiment using alamarBlue HS. **(C)** TAL1 expression in all CRISPRi samples, as determined by Western blot on day 26 after transduction. Antibodies used were anti-TAL1 antibody (sc-393287, Santa Cruz) and anti-GAPDH antibody (60004-1-Ig, Proteintech). Results are from a single replicate.

To achieve a more pronounced CRISPRi effect, the literature suggests combining several targeting sgRNAs (Qi et al., 2013). Therefore, we next tried to use various combinations of selected sgRNAs (MYB1, PRO1, PRO3) in our CRISPRi experiment in the intermediate cell line. The experiment was

otherwise set up in the same way as the previous one. On day 13, cell viability was assessed using alamarBlue HS (Figure 5.19), but similarly to the previous experiment, we again observed the unexpected proliferation-boosting effect in all *TAL1*-CRISPRi groups relative to the control groups. This experiment was therefore abandoned on day 13.

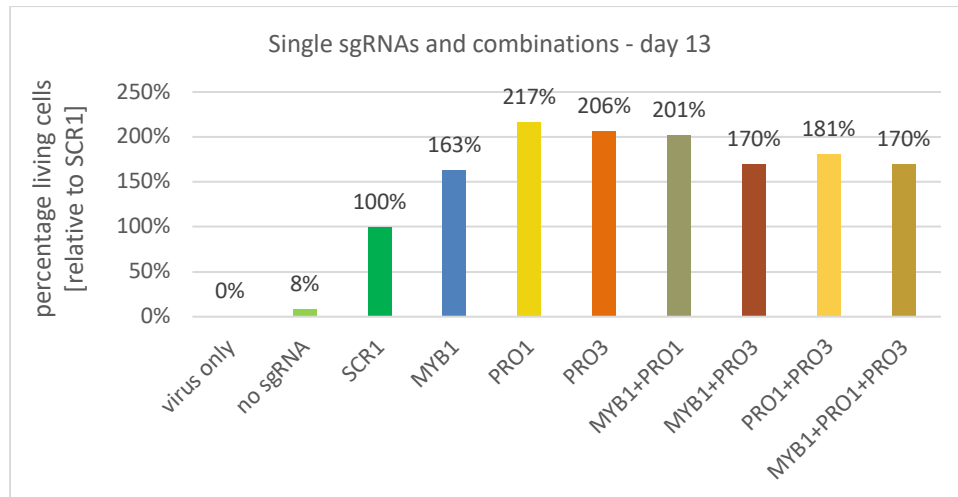


Figure 5.19: *TAL1*-CRISPR interference in Jurkat-Tet3G-TRE3G-FLAG-TAL1 cell line using combinations of sgRNAs.

Cells were transduced with either selected single sgRNAs as described in Figure 5.18, or various combinations thereof. Cell viability was measured using alamarBlue HS reagent 13 days post transduction. Graph shows data from a single replicate.

Overall, while *TAL1*-CRISPRi decreased the expression of TAL1 (Figure 5.18C), there was no accompanying decrease in growth rate (Figure 5.18A, B and Figure 5.19) – rather the opposite was true. This indicated that *TAL1*-CRISPRi in the Jurkat-Tet3G-TRE3G-FLAG-TAL1 cell line, in the tested setup, could not be used for the validation of the functionality of the doxycycline-inducible FLAG-TAL1 rescue without significant additional work.

5.9 Discussion

In this chapter, we tried to validate the rescue potential of the doxycycline-inducible FLAG-TAL1 in the newly established cell line, which is a non-negotiable feature in the context of the CRISPR/Cas9-mediated screen of the transcribed portion of the *TAL1* enhancer as we had originally planned it. The

growth of Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells is hampered upon TAL1 challenge, as we have shown in Chapter 4; we need to be able to counteract this trend by inducing the exogenous FLAG-TAL1 in our genetically modified cell line.

To demonstrate this rescue ability of the doxycycline-inducible FLAG-TAL1 in our Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line, we first planned to build on the optimized CRISPR/Cas9 transduction protocol from the previous chapter and compare the behaviour of the TAL1-challenged Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells cultured with or without the addition of doxycycline.

In an early set of experiments, we noted that increasing concentrations of doxycycline induced an increasing FLAG-TAL1 expression in Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 (Figure 5.2 in Chapter 5.2 and Figure 5.6 in Chapter 5.4), and that 400 ng/ml doxycycline supplementation of unedited cells initially caused an increase in proliferation (Figure 5.3 in Chapter 5.2 and Figure 5.5B in Chapter 5.3), but over time led to a drop in proliferation compared to uninduced, unedited cells (Figure 5.5B in Chapter 5.3). This pointed towards a certain toxicity of either the doxycycline itself, or the doxycycline-induced FLAG-TAL1 in a longer-term culture of Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells. Despite the unexpected negative effect of doxycycline, we also observed a rescue-like effect on the TAL1-challenged cells (Figure 5.4A and B in Chapter 5.3), which led us to go on to explore the options of reducing the toxic effect of doxycycline.

Firstly, while the half-life of doxycycline in media is 24 hours, we discovered that there is still an abundance of FLAG-TAL1 expressed in the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells 72 hours post induction even when as little as 50 ng/ml doxycycline is used (Figure 5.6 in Chapter 5.4). This reflected the high sensitivity of the Tet-On system in our cell line to doxycycline, and indicated that a pool of FLAG-TAL1 mRNA/protein might take days to deplete. Taken together, these results showed that doxycycline supplementation doesn't need to be renewed more often than every third day.

Next, we explored the effects of doxycycline concentrations lower than 400 ng/ml on polyclonal TAL1-challenged Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell lines (TAL10- and MYB1-edited cells established in the first attempt at doxycycline rescue in Chapter 5.3; Figure 5.7, Figure 5.8 and Figure 5.9 in Chapter 5.4). While these experiments on already edited cells provided us with a slightly better idea of the range of doxycycline concentrations to use, these experiments couldn't be used as validation experiments per se, especially not in the case of MYB1-edited cells, which seemed unrescuable in this context. This was somewhat surprising; we suspected that polyclonal cell lines that resulted from the initial editing were possibly undergoing an ongoing selection process where the edited cells with the most viable mutations were becoming more and more prominent as time progressed. This could, in theory, continue to the point where the surviving cells in the MYB1-edited group were no longer TAL1-

challenged. The comparison of the band intensities on the Western blots from the rescue experiment on the edited cells (Figure 5.9 in Chapter 5.4; intensity measurements not shown), along with the fact that in the latter experiment from edited cells, MYB1-edited cells without doxycycline grew faster than the control cells (Figure 5.8 in Chapter 5.4), somewhat confirmed these suspicions and indicated the need to carry out validation experiments that start from the day of the sgRNA introduction. Nevertheless, why the MYB1-edited cells were almost completely resistant to any viability booster effect, such as we saw with the control cells at lower doxycycline concentrations (Figure 5.7 and Figure 5.8 in Chapter 5.4), remains a mystery.

A rescue validation experiment carried out from the day of transduction and using three different doxycycline concentrations again indicated a level of rescue, especially in the cultures incubated with 200 ng/ml doxycycline, but the rescue effect was once more mixed with a strong negative effect on proliferation rate.

This negative effect of doxycycline could be avoided if the experimental timeframe was shortened. Conducive to that, CRISPR/Cas9-editing seems to take place rather fast in the Jurkat cell line (for example, TAL7- and TAL10-edited samples showed strong TAL1 downregulation on day 3 after the introduction of the sgRNAs, Figure 4.4 in Chapter 4.3), as well as in our Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line (Western blot data from day 3 after sgRNA introduction were not included in this work). On the other hand, we previously observed that selecting cells with hygromycin leads to an early proliferation advantage in TAL1-challenged cells (e.g. Figure 4.22 in Chapter 4.13), possibly due to the fact that antibiotics in general act faster on more rapidly dividing cells. These two countereffects need time to play themselves out. We previously tried to take the hygromycin selection out of the equation by means of FACS-sorting of the transduced cells; this approach failed at the time due to low transduction efficiency. Since then, however, we had successfully raised the transduction efficiency many fold by means of optimization, and therefore a return to this method held a distinct promise. First results reinforced this hope – the viability of the TAL1-challenged cells sank to the expected relative levels within days after the sorting (Figure 5.11 in Chapter 5.5). However, when FACS-sorted cells were cultivated with doxycycline for the FLAG-TAL1-mediated rescue, a massive cell death occurred as a result, in both control (SCR1) and experimental (MYB1) groups, even with the lowest of doxycycline concentrations used (Figure 5.12 in Chapter 5.5). Conceivably, the immense stress of sorting with FACS rendered the cells more vulnerable to the negative effects of doxycycline and at an earlier time point in the experiment. These results revealed the unfeasibility of using FACS in conjunction with doxycycline rescue.

The drop in viability of the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line in response to doxycycline could be either due to the toxicity of the exogenous FLAG-TAL1 and its overexpression, or due to the toxicity of doxycycline to the cells. Literature suggest that some types of cancer cells, as well as primary cells, are sensitive to doxycycline at concentrations of 500 ng/ml and higher, with a reduction in proliferation rate (Ahler et al., 2013; Dijk et al., 2020; Liu et al., 1999; Xie et al., 2008) and disruptions to glycolysis (Ahler et al., 2013; Dijk et al., 2020) and normal mitochondrial function (Ahler et al., 2013; Dijk et al., 2020; Moullan et al., 2015). Some changes in gene expression were already measurable at 100 ng/ml doxycycline in selected cell lines (Ahler et al., 2013). In Jurkat cells in particular, high doxycycline concentrations were shown to lead to apoptosis (Liu et al., 1999). The doxycycline concentrations used in this study (5000 ng/ml and above) were much higher than the concentrations used in our experiments, but the observed effects were radical even at 24 hours after doxycycline supplementation. It is therefore plausible that over a longer time period, lower doxycycline concentrations can also become detrimental to the Jurkat growth rate.

In our doxycycline rescue validation experiments in the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line, we likely observed a combination of the rescue effect of the doxycycline-induced FLAG-TAL1 and the effect of doxycycline cytotoxicity. Potentially, there might have been a third element to the effect of doxycycline on the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell proliferation, similar to the hygromycin effect discussed earlier: doxycycline killing more rapidly dividing cells faster than cells that are slower to proliferate. Nevertheless, encouraged by the information from Western blotting, which repeatedly testified to an ample FLAG-TAL1 expression in doxycycline-treated samples (e.g., Figure 5.2 in Chapter 5.2; Figure 5.6 in Chapter 5.4 and Figure 5.9 in Chapter 5.4), and the early proliferative advantage observed in unedited or non-targeting sgRNA-transduced doxycycline-treated Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells (e.g., Figure 5.3 in Chapter 5.2; Figure 5.5B in Chapter 5.3; Figure 5.7 and Figure 5.8 in Chapter 5.4), we believe we were able to discern a distinct rescue effect of doxycycline on the TAL7- and TAL10-edited cells, and a less pronounced, but rather consistent rescue effect on MYB1-edited cells (Figure 5.4A, B in Chapter 5.3; Figure 5.10 in Chapter 5.4; Figure 5.13 in Chapter 5.6).

The rescue effect on TAL7- and TAL10-edited cells (*TAL1* exon-edited) was unanticipated; both these sgRNAs target both the endogenous *TAL1* and the exogenous *FLAG-TAL1*, and a lack of rescue effect was therefore expected in these cells. However, as with all CRISPR editing, some of the cells within these samples retained some level of functionality of the exogenous FLAG-TAL1, and as a result, their fate could be altered by doxycycline addition. The partial retainment of the doxycycline-inducible FLAG-TAL1 in TAL7- and TAL10-edited cell populations is also evidenced by Western blotting (Figure 5.4C in Chapter 5.3).

On the other hand, cells edited with TAL1 and TAL4 sgRNAs (targeting the *TAL1* 5'-UTR) in our initial experiment in this chapter (Figure 5.4 in Chapter 5.3) did not display any signs of being rescued by doxycycline, or indeed any signs of needing a rescue – the proliferation rate of these cells without doxycycline was at least comparable to the controls. This was at odds with an earlier experiment where TAL1 and TAL4 sgRNAs lowered the growth rate of Jurkat cells (Figure 4.14 in Chapter 4.9), although this earlier experiment was done in the maternal cell line, by electroporation rather than transduction, and by co-transfection of Cas9 and sgRNAs rather than just the introduction of the sgRNAs. TAL1 and TAL4 sgRNAs were therefore abandoned in the context of the doxycycline rescue validation experiment.

The rescue effect of doxycycline on the MYB1-edited (*TAL1* enhancer-edited) Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells showed steadily across the experiments where doxycycline rescue was administered and monitored from the day of sgRNA transduction (Figure 5.4A, B in Chapter 5.3; Figure 5.10 in Chapter 5.4; Figure 5.13 in Chapter 5.6), although not in the experiments where already selected populations of MYB1-edited cells were used (Figure 5.7 and Figure 5.8 in Chapter 5.4).

The doxycycline-rescue effects were strongly dependent on doxycycline concentration and treatment schedule, the sgRNA used, but also differed between the individual experimental setups, and the time points in the experiment when we did our measurements (Figure 5.4A, B in Chapter 5.3; Figure 5.7, Figure 5.8 and Figure 5.10 in Chapter 5.4; Figure 5.13 in Chapter 5.6). Other factors also seemed to play a role, such as the growth rate of the cell line immediately prior to the experiment, the passaging and the transduction efficiency achieved in each individual case.

Upon finding that FACS-sorting couldn't be used because the cell line was too sensitive to withstand the joint stress of the FACS and the doxycycline, we returned to the hygromycin selection setup, however flawed, to generate replicates for the doxycycline rescue validation experiment. The experiments were done using the 293FT-generated viral particles, as our new transduction optimization experiment indicated this packaging cell line produced much higher viral titres than the regular HEK293T cell line. However, in the first experiment, the absolute numbers of surviving cells at each point (data not shown) were at least as low or even lower than in case of the experiments done with the HEK293T-produced virus. In the second and third repeat of this experiment, the cells failed to expand into polyclonal cell lines, and died off by 2 weeks into the experiment. These findings suggested that something about the exchange of the HEK293T-generated virus for the 293FT-generated virus caused increased stress to the cells in the context of the transduction experiment. A higher toxicity of the 293FT-generated viral particles, due to their quality or quantity, could conceivably be at fault, but in samples where cells transduced with 293FT-generated virus were left

without selection pressure, the proliferation was hardly inhibited (data not shown), making this explanation somewhat weak. The other explanation that offered itself was the fact that the flow cytometry images (Figure 5.15 in Chapter 5.7) could have been read wrongly, and the whole population shift was not due to the intake of the sgRNA plasmids by the whole population, but rather due to autofluorescence. This would be in keeping with the patterns we believed to have seen in the first optimization experiment (Figure 4.8 in Chapter 4.6), although the possibility of the opposite explanation also arises (We evaluated the first set of optimization experiments wrongly, and did in fact see a whole-population shift due to successful transduction.). The autofluorescence theory would also be in line with the unexpectedly small difference between the viability of MYB1-edited cells and the SCR1 control cells in the doxycycline rescue experiment with 293FT-generated virus that employed FACS (Figure 5.12 in Chapter 5.5) – the explanation would be that most of the sorted cells were in fact not successfully transduced, but only autofluorescing. To address this issue further, we checked for autofluorescence in the FITC channel (a channel where autofluorescence is common, and where we would expect to see some if there was any in the dTomato channel; personal communication, Dr. Jan Svoboda) – but we found none. Taken together, it is quite possible that the transduction efficiency obtained with the 293FT-produced virus was perhaps nowhere near as high as we initially thought, and indeed was perhaps even lower than with the HEK293T-generated viral particles. On the other hand, it is also plausible that if the viral titre was high, as our original evaluation would suggest (Figure 5.15 in Chapter 5.7), then the combined stress of a high viral titre and hygromycin selection could have caused the massive cell death in the last set of doxycycline rescue validation experiments. To explore this matter further, we could titrate the virus either by physical titration (ELISA for Gag (p24), (Logan et al., 2004)), and/or by functional titration using the dTomato fluorescence. If the viral titre of the 293FT-generated viral stock turns out to be a lot higher than that of HEK293T-generated viral stock prepared under the same conditions, further experiments might be in order to find out whether employing a lower MOI might help with the viability of cells in validation experiments. If these experiments are successful, we could use the 293FT cell line to generate the viral stocks of the sgRNA library for our CRISPR/Cas9 screen.

In any case, to try to avoid all the pitfalls of the doxycycline rescue validation experiment in Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells encountered so far, we decided to completely change tack and attempt to show a doxycycline rescue in the intermediate cell line, Jurkat-Tet3G-TRE3G-FLAG-TAL1, upon *TAL1*-CRISPR interference (*TAL1*-CRISPRi). The hope for this approach was that it might potentially provide us with a more robust system than the CRISPR-knockout. The experiment couldn't be performed in the final line, as Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 stably expresses Cas9, which would compete for the provided sgRNAs with the dCas9-KRAB for CRISPRi, but we felt the results of a

doxycycline rescue experiment in Jurkat-Tet3G-TRE3G-FLAG-TAL1 could be extrapolated to the final cell line. Unfortunately, the CRISPRi experiment in the intermediate cell line did not show the expected drops in proliferation upon the attempted interference with *TAL1* expression (Figure 5.18A, B and Figure 5.19 in Chapter 5.8), despite the fact that TAL1 downregulation did take place in these cells (Figure 5.18C in Chapter 5.8). As we repeatedly observed that *TAL1* knockout definitely had a negative impact on the growth rate of the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line, as well as Jurkat cells (e.g., Figure 4.14 in Chapter 4.9; Figure 4.22 in Chapter 4.13), it seemed as though something in the experimental setup was interfering with the natural proliferation rates of the cells. Possibly, this could have been the antibiotic selection process again, only in this case, because the CRISPRi effects were not as strong as the CRISPR effects (compare e.g. Figure 4.23 in Chapter 4.13 or Figure 5.4C in Chapter 5.3 with Figure 5.18C in Chapter 5.8), it was taking much longer to clear up. However, the effect persisted even when already edited and selected cells were started in independent cultures (Figure 5.18B in Chapter 5.8). Furthermore, the stark differences between the two controls, no sgRNA and SCR1, were also quite startling. In sum, the *TAL1*-CRISPRi provided us with highly unexpected results; to start to understand the factors contributing to the mystery, we would probably want to begin by cloning the sgRNAs and the dCas9-KRAB into a plasmid sporting a fluorescent marker, which would allow us to measure the transduction efficiency, use FACS to collect positive cells for a proliferation experiment to see if the trend changes, and carry out potential optimizations.

In terms of other potential ways of validating the doxycycline rescue in the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line, we briefly considered attempting CRISPRi in the final line with dCas12-KRAB, but logistical issues prevented us from embarking on this course. The rescue could not be carried out using our siRNAs for *TAL1* knockdown, as these siRNAs target both the endogenous and the exogenous *TAL1*. Overall, the intricate interplay of the effects of doxycycline, hygromycin selection and sensitivity of the final cell line within and without the constraints of the rescue experiment render the system impracticable for the use in a pooled screen format with the originally planned doxycycline rescue. However, since the Jurkat-Tet3G-TRE3G-FLAG-TAL1 cell line displays the TAL1-dependency phenotype, it can still be used for a pooled screen where the control sample is collected at the beginning of the experiment (Chapter 6). Furthermore, it can also be used for the validation of individual sgRNAs with doxycycline rescue in an arrayed format, as we can trace the proliferation of the cell line in answer to each of the individual sgRNAs throughout the course of the experiment, rather than just check at a defined final end-point.

6 Initial considerations and experimental outline of the CRISPR/Cas9-mediated mutagenesis of the *TAL1* enhancer

For the reasons outlined in the previous chapter, most prominently the double impact of doxycycline on the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line, the newly established cell line seems to be unfit for use with the original screen design, where the control group would be treated with doxycycline (Figure 6.1A in Chapter 6.1). However, as the cells do demonstrate a growth defect when the *TAL1* enhancer is targeted by CRISPR/Cas9 (Chapter 4), they can still be used for the CRISPR/Cas9-mediated mutagenesis, with the employment of an adjusted protocol (Figure 6.1B in Chapter 6.1).

6.1 The adjusted experimental design

In the adjusted protocol for the CRISPR/Cas9-mediated mutagenesis of the *TAL1* enhancer (Figure 6.1B, Figure 6.2), we will take an advantage of the stably transduced Cas9 in the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line, and the fact that we have successfully validated this cell line's dependence on TAL1, having shown that both a CRISPR/Cas9-mediated *TAL1* knockout (e.g., Figure 4.22 in Chapter 4.13) and *TAL1* enhancer knockout (e.g., Figure 5.4 in Chapter 5.3) lead to a decrease in proliferation in Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9.

In the adjusted experimental design, the difference in sgRNA abundance reflecting the altered growth phenotypes will be provided by the comparison of a control group that will be collected at an early time point ($T = 0$) and an experimental group collected some time into the experiment ($T = \text{exp.}$; Figure 6.1B; Figure 6.2).

The control group of cells will be harvested at an early time point ($T = 0$). This will be likely on day 2 or 3 after transduction – at a time when the genomic incorporation of the sgRNAs will hopefully have taken place, but either the editing still hasn't occurred, or at the very least, the effects of the editing haven't had the chance to play themselves out yet. Alternatively, we might deep-sequence the sgRNA plasmid library as the initial time point (Canver et al., 2018), or we might transduce the sgRNA library into a cell line that is either independent of TAL1 expression, and/or does not express Cas9 (e.g., HEK293 cells or the Jurkat-Tet3G-TRE3G-FLAG-TAL1 cell line).

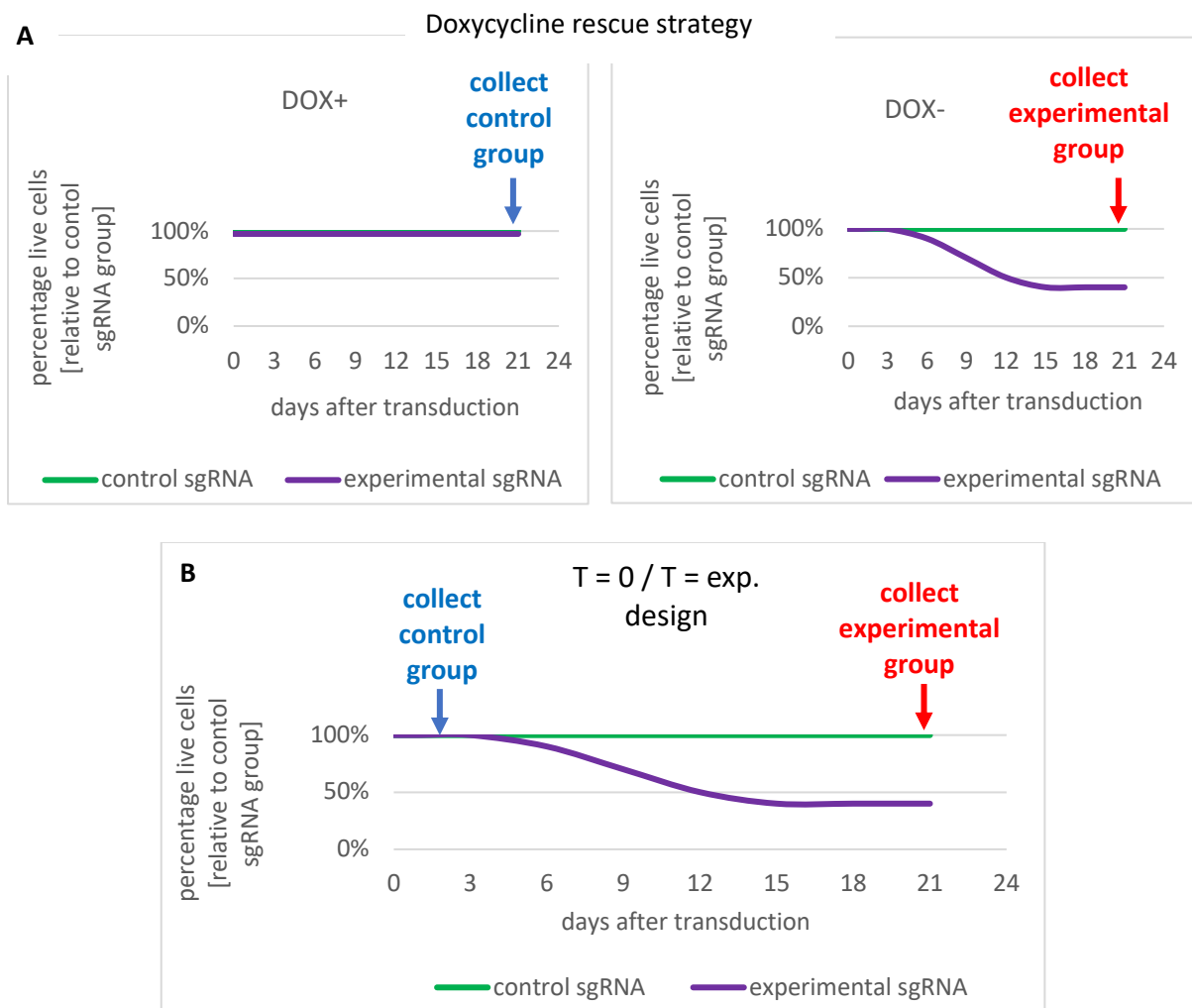


Figure 6.1: Comparison of the original and the adjusted design for the CRISPR/Cas9-mediated mutagenesis of the *TAL1* super-enhancer in the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line.

(A) In the original doxycycline (DOX)-based rescue design, both the control (DOX+) and the experimental (DOX-) sample were to be collected on the last day of the experiment (approximately 3 weeks post sgRNA library introduction). Control sample was to be treated with doxycycline (DOX) to induce the exogenous FLAG-TAL1 expression, which was meant to rescue the proliferation rate of any TAL1-challenged cells, bringing the growth rates of TAL1-challenged and control cells to the same level. **(B)** The new approach operates without the doxycycline rescue, collecting a control ($T = 0$) sample early on in the course of the CRISPR experiment, and the experimental ($T = \text{exp.}$) sample at the end of the 3-week proliferation period.

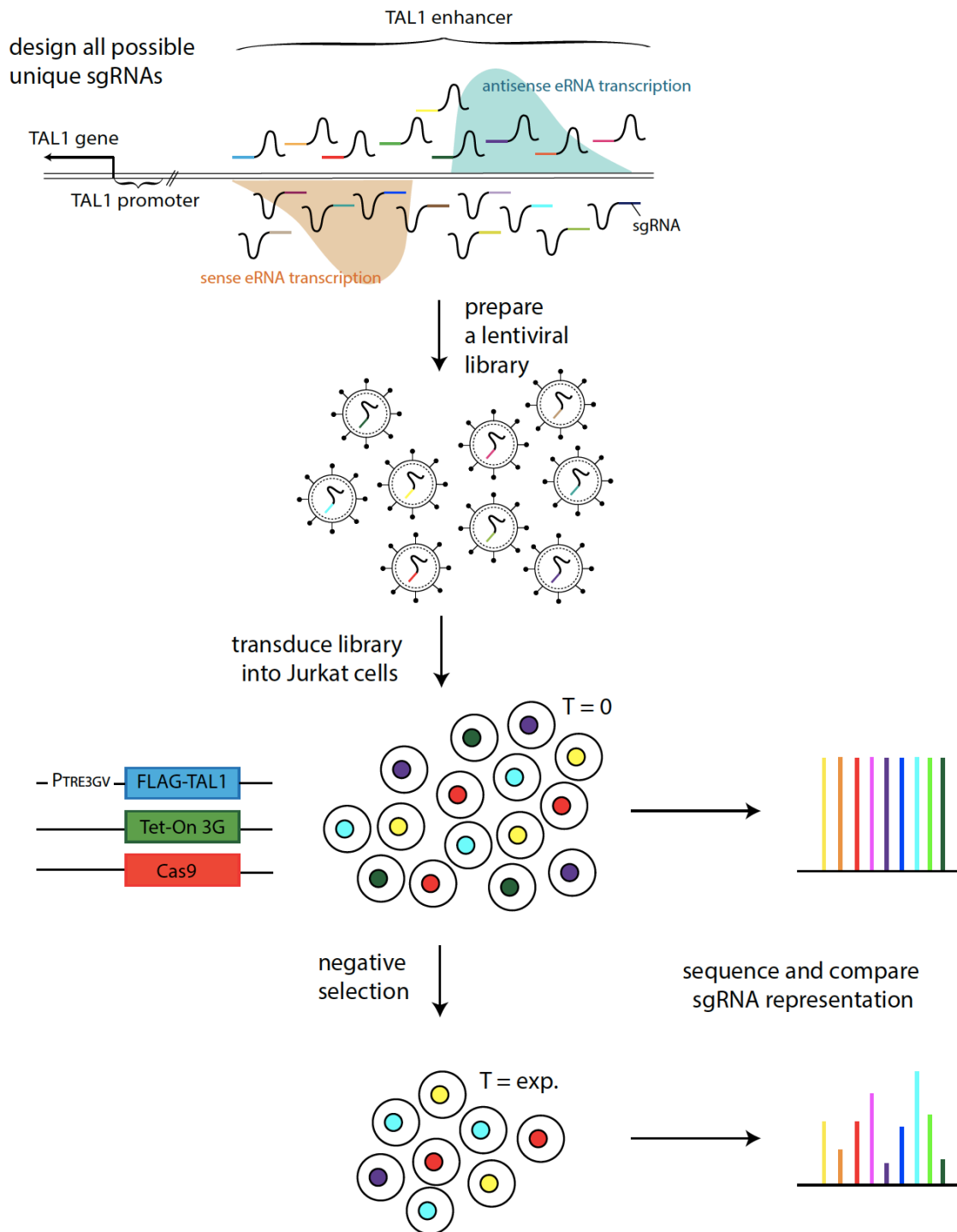


Figure 6.2: Workflow for the CRISPR/Cas9-mediated saturating mutagenesis of the *TAL1* enhancer, reimaged.

All unique *sgRNAs* targeting the transcribed portion of the *TAL1* enhancer are designed, synthesized and cloned into a plasmid, which is then encased in lentivirus packaging. Such lentiviral library with an equal representation of all designed *sgRNAs* is then transduced into the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line. At an early time point ($T = 0$), cells are collected to provide for a control group. After negative selection has taken place, the remainder of the cells is collected as the experimental group. Both samples are sequenced, abundance of *sgRNAs* is compared. *sgRNAs* responsible for functionally important mutations are expected to be depleted in the experimental group.

Before or at $T = 0$, the selection for transduced cells will commence, to keep cell numbers at a manageable level throughout the experiment. Edited cells can technically be selected for either by FACS, or by hygromycin selection. The advantage of employing FACS is the avoidance of the need for hygromycin selection, which can obfuscate the results possibly by acting on edited cells in more than one way (Chapter 4). On the other hand, FACS and the associated time on ice, as well as prolonged periods at room temperature, are very stressful for the cells, especially since they have already been stressed by viral infection, and result in significant cell losses and/or growth retardation in the sorted cells. Both methods might be employed in due course as we optimize the experimental protocol for the CRISPR/Cas9-screen.

The experimental sample ($T = \text{exp.}$) will be collected when the effects of editing have had the time to come into effect, at around 3 weeks from the introduction of the sgRNA library, representative of approximately 17 cell doublings (Figure 6.1B).

Genomic DNA from both the control and the experimental group will be isolated and a sequencing library prepared. By comparing the abundance of sgRNAs in the two samples, as obtained from the sequencing data, we will pinpoint important regions within the *TAL1* enhancer. We will then focus on these regions to understand whether their functional importance is connected to underlying transcription factor binding motifs on the DNA, or if they might be pointing towards important structural features within the *TAL1*-eRNA transcripts.

6.2 sgRNA design and library preparation

To design the library of sgRNAs for the CRISPR/Cas9-mediated mutagenesis of the *TAL1* enhancer, there is a need for experimental sgRNAs (targeting the *TAL1* enhancer) and control sgRNAs (with expected results on the cell line). We used the following information, tools and approaches to design the library.

The sgRNAs for the screen were designed using the Broad Institute's CRISPick (Doench et al., 2016; Sanson et al., 2018). Based on the presence of PAM sites, the software generates all possible sgRNAs targeting a selected region, but flags sgRNAs targeting more than one region in the genome and sgRNAs with problematic structural features, such as poly-T sequences. This allows users to avoid subpar or non-unique sgRNAs.

For our screen, we wanted to dissect the portion of the *TAL1* enhancer that is transcribed into eRNAs. I set off to pinpoint this region by performing a rapid amplification of cDNA ends (RACE, (Frohman et al., 1988)). RACE uses PCR to amplify the full length of unknown transcripts based on the knowledge of a short part of the sequence within the body of the transcript. In the first instance, I attempted circular RACE (cRACE), which is a variation on RACE that enables a simultaneous identification of both the ends of the transcript by circularization with ligase followed by nested inverse PCR (McGrath, 2011). Unfortunately, from the number of rounds of cRACE I performed, none of the obtained PCR products corresponded to either of the *eTAL1*s, which I was trying to amplify. I then tried the more classic take on RACE, which uses universal anchor sequences attached to either 3'- or 5'-end of the transcript to amplify either the sequence 3' or 5' from the short known sequence. I used the SMARTer® RACE 5'/3' Kit (Takara) for the job, but the sequences of all the products that I managed to amplify again did not correspond to the *eTAL1* sequences. It is highly likely that we did not manage to amplify the *eTAL1* sequences because of their low abundance.

Instead of the information we were hoping to acquire by RACE, we decided to let ourselves be guided by a published set of precision run-on sequencing (PRO-seq) data obtained for Jurkat cells (Figure 6.3, GEO accession: GSM1613182, (Danko et al., 2015)). PRO-seq is a highly sensitive method that captures the 3'-ends of nascent transcripts to a precision of several nucleotides, creating a quantitative map of transcriptional activity (Mahat et al., 2016), and can thus be used with a high confidence.

For our screen, we chose to tile the *TAL1* enhancer-targeting sgRNAs across a 2400 bp stretch of the *TAL1* enhancer centred on the de novo MYB binding site. This region covers not only the signal from the Jurkat PRO-seq, but also the signal from an unpublished ChromRNA-seq (Dr. Ramin Shiekhatter, personal communication). All possible unique sgRNAs tiling this region were designed; some regions, however, were left untiled, because repeating elements didn't allow for unique sgRNA design. (Figure 6.3).

Positive controls for the experiment, i.e. sgRNAs causing mutations which are expected to slow down the proliferation of the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line, will target the members of the TAL1 complex (*TAL1*, *MYB*, *GATA3*, *HEB*, *RUNX1*, *CBP*) and the *TAL1* promoter region. Negative controls, i.e. sgRNAs that shouldn't affect the proliferation of the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line, will include sgRNAs targeting a safe harbour locus in Jurkat cells (Aznauryan et al., 2021), sgRNAs targeting a locus within the *TAL1* TAD without known function and no PRO-seq or chromatin modification signal (Danko et al., 2015; Hnisz et al., 2016b), and non-targeting sgRNAs (Canver et al., 2017). In total, there will be around 640 unique sgRNAs (see Appendix 1 for sequences).

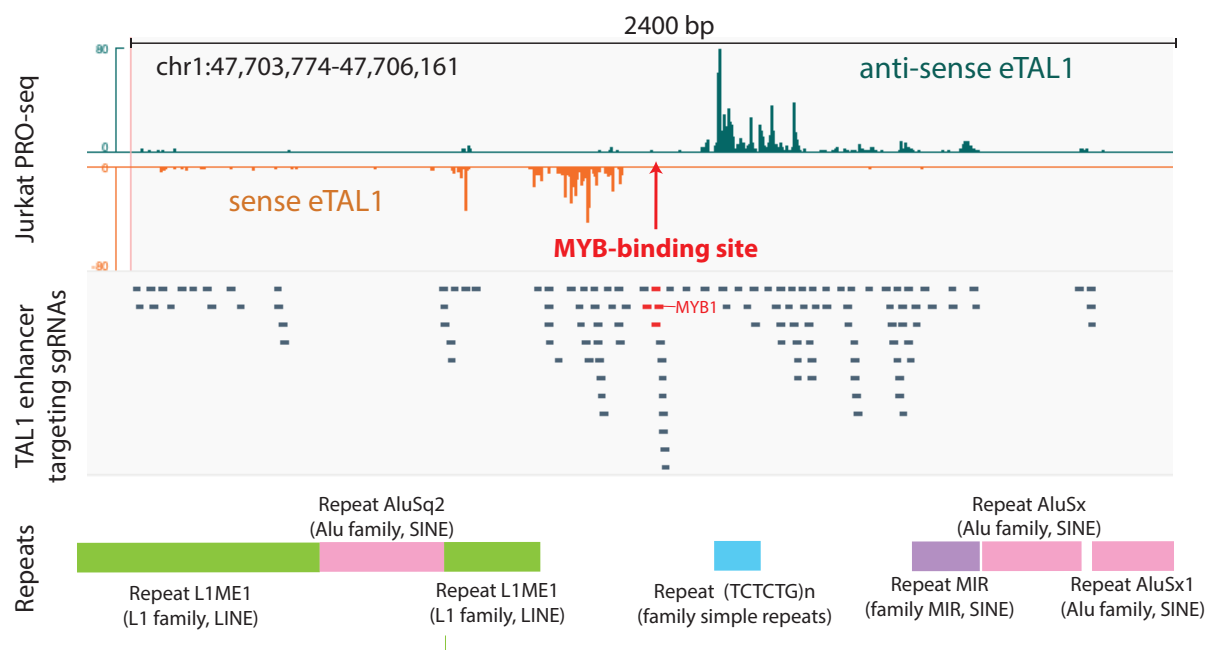


Figure 6.3: Experimental sgRNAs targeting the *TAL1* super-enhancer in Jurkat cells.

Experimental sgRNAs were designed against the *TAL1* enhancer sequence in Jurkat cells, +/- 1200 bp from the MuTE site (mutation of the *TAL1* enhancer; MYB-binding site is labelled in red). The sgRNAs span the active transcription region as per the Jurkat PRO-seq data (GEO accession: GSM1613182) (Danko et al., 2015) plus generous overhangs on each side. MYB-binding site targeting sgRNAs are in red, MYB1 sgRNA is labelled. Repeat elements in the region of the *TAL1* enhancer are shown in lime green (long interspersed nuclear elements, LINEs), pink/violet (short interspersed nuclear elements, SINEs) and blue (simple repeats). Figure was made using IGV, the Integrative Genomics Viewer (Robinson et al., 2011), GRCh37/hg19 assembly, and repeats were added according to information from the UCSC Genome Browser (<http://genome.ucsc.edu>, GRCh37/hg19 assembly, (Kent et al., 2002)).

6.3 Using iBARs for internal replicates

To increase the accuracy and the precision of our screen, we plan to introduce a barcoding strategy that will supply us with internal replicates for each of our sgRNAs within each biological replicate of our CRISPR/Cas9 screening experiment. Several strategies have recently been developed for sgRNA barcoding (Michlits et al., 2017; Schmierer et al., 2017; Zhu et al., 2019), and these are further discussed in Chapter 6.12. From these approaches, we chose to use the internal barcode (iBAR) strategy, in the manner described by Zhu and co-workers (Zhu et al., 2019). In this work, four 6-nt long randomly generated barcodes were assigned to each sgRNA, providing internal repeats that led to decreased background noise, increased accuracy and enhanced statistical significance of results. The barcode was inserted into the tetra loop of the sgRNA scaffolding, which, together with an extension of the stem loop in question, placed it physically outside of the sgRNA/Cas9 complex (Figure 6.4). This

ensured that the adjustment bore minimal impact on the functionality of the ribonucleoprotein (RNP) complex (Zhu et al., 2019).

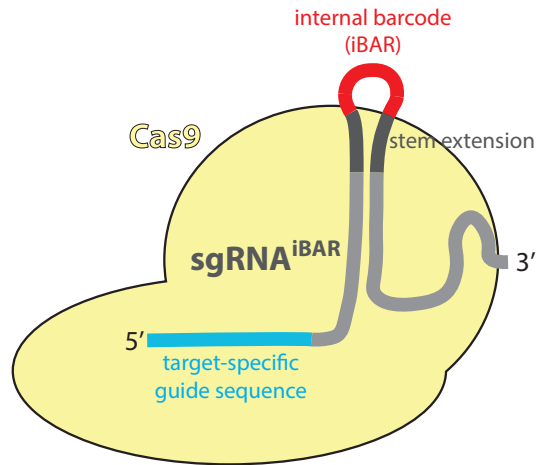


Figure 6.4: The sgRNA^{iBAR}-Cas9 complex.

If extended, the upper stem loop in the sgRNA scaffold juts out of the Cas9/sgRNA complex, providing a convenient spot for alterations such as internal barcodes (iBARs). (Zhu et al., 2019)

In our own work, we chose to use a unique iBAR for each sgRNA^{iBAR}, and the iBARs were therefore made 10 nt in length so that we could reach a sufficient number of them. The iBARs were generated by Dr. Dan Bose using the Bioconductor platform (<https://bioconductor.org>, (Buschmann and Bystrykh, 2013)).

6.4 Vectors and cloning strategies

To house the sgRNAs for the CRISPR/Cas9-mediated mutagenesis of the transcribed portion of the *TAL1* enhancer, we plan to use the lentiGuide-Hygro-dTomato plasmid backbone (Addgene plasmid #99376; (Ho et al., 2017)), which will be packaged using pMDLg/pRRE, pRSV-Rev and pMD2.G (Addgene plasmids #12251, #12253 and #12259; (Dull et al., 1998)), as per our previous experiments (Chapter 4 and 5). This is advantageous because having worked with these plasmids previously, we already have optimized protocols for their use. Furthermore, the lentiGuide-Hygro-dTomato plasmid will allow us to use either hygromycin selection or FACS to select for transduced cells, lending us some flexibility in the experimental design.

We will adjust the lentiGuide-Hygro-dTomato plasmid in several ways, to meet the cloning needs associated with the iBAR strategy, and for increased efficiency and result clarity. Firstly, we will introduce changes into the sgRNA scaffolding that should lead to a higher stability and efficiency of our sgRNAs – an ‘A-U flip’ in the lower stem and an extension of the upper stem (Figure 6.5B; (Chen et al., 2013; Zhu et al., 2019)). The elongated upper stem will be capped with the iBAR (Figure 6.4; Figure 6.5B; (Zhu et al., 2019)). Secondly, to make use of the iBAR sgRNA cloning strategy, we will have to alter the location of the BsmBI restriction sites in our plasmid (Figure 6.6; (Zhu et al., 2019)). Finally, to ensure high purity and lessen the potential for cross-contamination of our samples with other lentiGuide constructs used in the lab, we will introduce a unique primer binding site (UPBS) in front of the U6 promoter (Figure 6.6; (Michlits et al., 2020)).

We chose to introduce all the adjustments in a single step by means of a custom-made polynucleotide fragment (Figure 6.6; Appendix 2), both for simplicity and to avoid the need for PCR in the cloning procedure. The latter is important as the lentiGuide-Hygro-dTomato plasmid has a number of repeating elements, which can be problematic features to amplify by PCR. Using available restriction enzyme sites (PpuMI/XmaI), the fragment will replace the region starting with the U6 promoter and ending with the central polypurine tract/central termination sequence of HIV-1 (cPPT/CTS) in the lentiGuide-Hygro-dTomato plasmid, introducing the unique primer binding site (UPBS), adjusted locations for the BsmBI restriction sites, and a part of the optimized sgRNA scaffolding (sgRNA scaffold right (R), Figure 6.6).

The other part of the sgRNA scaffold (sgRNA scaffold left, (L)) is to be cloned into the adjusted lentiGuide-Hygro-dTomato plasmid as a part of the oligonucleotide containing the sgRNA targeting sequence and the iBAR in the sgRNA^{iBAR} library cloning step (Figure 6.7B). This is because the iBAR sequence is localized between the left and the right part of the scaffold, so that it forms a loop at the end of the first stem of the sgRNA scaffolding (Figure 6.5B). The cloning into the adjusted lentiGuide-Hygro-dTomato plasmid requires the synthesis of a 90-bp-long oligo library. Each oligo will contain (left to right): left-side overhang and BsmBI restriction site, target-specific guide sequence, left part of the sgRNA scaffolding, iBAR, right-side overhang with a BsmBI restriction site. The sgRNA^{iBAR} library will then be cloned into the adjusted lentiGuide-Hygro-dTomato plasmid using BsmBI-mediated Golden Gate Assembly (Figure 6.7B). This is in contrast with the original cloning strategy, where BsmBI restriction was used to remove a long stuffer sequence between the U6 promoter and the sgRNA scaffold in the original lentiGuide-Hygro-dTomato plasmid, and pairs of annealed oligonucleotides representing the target-specific guide sequence with appropriate overhangs were ligated into the

plasmid in place of the stuffer (Figure 6.7A). For the sgRNA^{iBAR} oligo library synthesis and cloning of the library, we will employ the services of a bioscience company, Twist Bioscience.

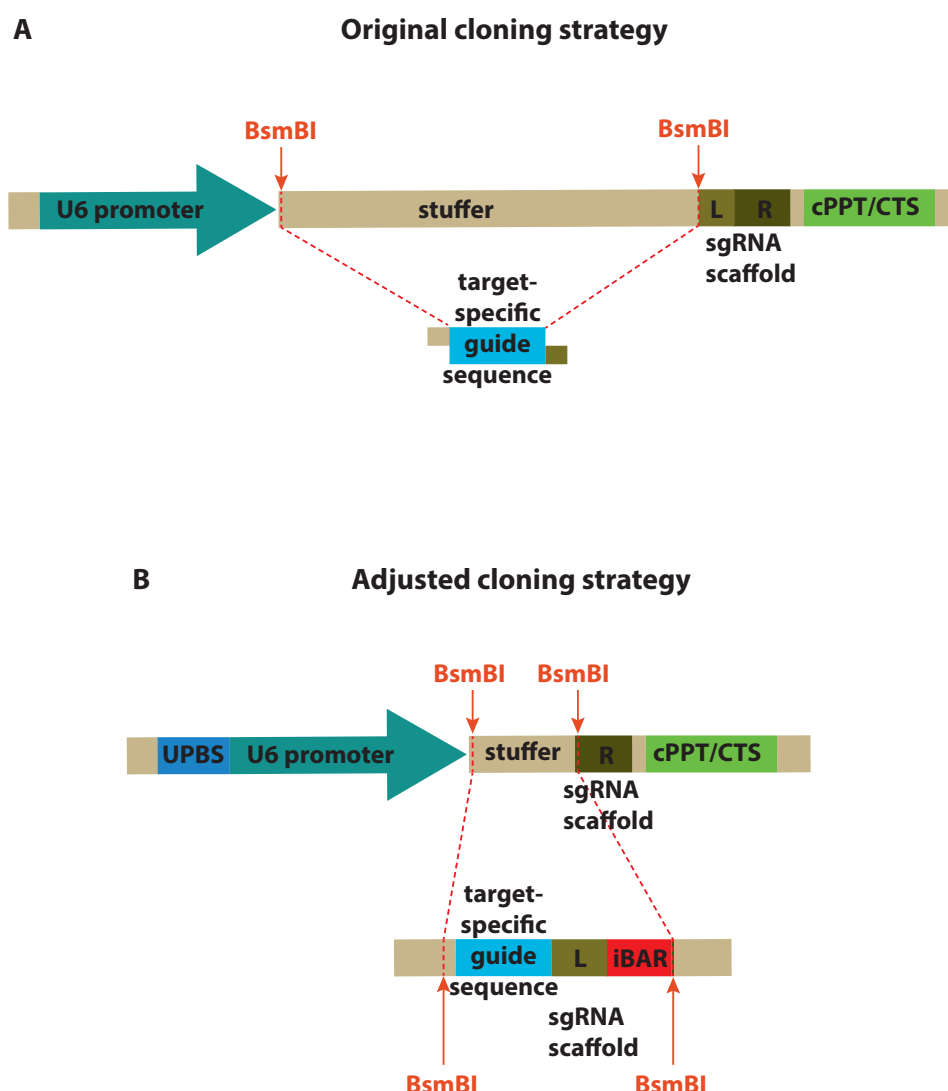


Figure 6.7: Comparison of the original and adjusted cloning strategies.

(A) Cloning of a pair of annealed oligonucleotides (25 nt in length) with appropriate, single-stranded overhangs into the original, BsmBI-digested lentiGuide-Hygro-dTomato plasmid. **(B)** A double-stranded oligonucleotide (90 nt in length) containing appropriate overhangs, BsmBI restriction sites, target-specific guide sequence, left part of the sgRNA scaffold (L, see Figure 6.5) and an iBAR is digested using BsmBI and cloned into a BsmBI-digested adjusted lentiGuide-Hygro-dTomato plasmid. UPBS – unique primer binding site; R – right side of the sgRNA scaffolding; cPPT/CTS – central polypurine tract/central termination sequence of HIV-1.

6.5 Viral packaging

We will likely use the HEK293T cell line to generate the lentivirus. While virus generated in 293FT cells showed some initial promise (Chapter 5.7), it was also associated with high cell mortality in Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 transduction experiments (Chapter 5.6; discussed in Chapter 5.9). Unless we find a solution to the latter, we will turn to our originally employed cell line, HEK293T.

6.6 Multiplicity of infection (MOI)

Cells will be transduced with concentrated viral particles at a low multiplicity of infection (MOI), to ensure that most cells either receive just one sgRNA, or no sgRNA at all. Using low MOI has been the golden standard for CRISPR/Cas9-mediated screens since their conception, because the combined effects of multiple sgRNAs in a single cell may confound the screen results. However, the advent of sgRNA barcoding diminished the need for low MOI, as the barcoding strategies are efficient at disentangling the effects of individual sgRNAs in experiments where multiple sgRNAs are present per cell. Nevertheless, it has been shown that screens carried out both with barcodes and at a low MOI can unearth elements that play relatively weak roles within the examined context, which would go otherwise unnoticed (Zhu et al., 2019). As we are expecting the mutations in *TAL1*-eRNAs to only have a subtle (if any) effect on the overall *TAL1* expression and cell viability, using a low MOI for our screen, along with the iBARs, should increase the probability of identifying the regions important for the *TAL1*-eRNA function. Another reason for keeping the MOI low in our experiment is the tiled nature of our screen – in cells receiving more than one of the experimental sgRNAs, pairwise deletions might result instead of local indels.

Low MOI screens habitually use a range of MOIs between 0.1 and 0.5 (Diao et al., 2017; Doench et al., 2016; Gasperini et al., 2017; Klann et al., 2017; Korkmaz et al., 2016; Parnas et al., 2015; Sen et al., 2019; Shi et al., 2015; Zhu et al., 2016), which corresponds to a 9.5 – 39.3% transduction efficiency. CRISPR screens with Jurkat cells were previously carried out at low MOIs, 0.3 – 0.5 (Shang et al., 2018; Wang et al., 2018). MOI is inherently linked to transduction efficiency through the Poisson distribution equation:

$$P(x) = \frac{\lambda^x * e^{-\lambda}}{x!} ,$$

where x is an input variable that represents the number of sgRNAs a single cell receives, $P(x)$ is the probability of a cell receiving x sgRNAs, and λ is MOI (the average rate of event occurrence). As a practical example, if MOI = 0.3, then out of the definition of MOI, each 100 cells will bear, on average, 30 sgRNAs between them. The Poisson distribution teaches us how these 30 sgRNAs will be statistically distributed amongst any given 100 cells:

$$P(0) = \frac{0.3^0 * e^{-0.3}}{0!} = 0.741$$

$$P(1) = \frac{0.3^1 * e^{-0.3}}{1!} = 0.222$$

$$P(2) = \frac{0.3^2 * e^{-0.3}}{2!} = 0.033$$

$$P(3) = \frac{0.3^3 * e^{-0.3}}{3!} = 0.0033$$

$$P(4) = \frac{0.3^4 * e^{-0.3}}{4!} = 0.000$$

This means that in a culture transduced at MOI = 0.3, there will be 74.1% cells containing no sgRNA whatsoever, 22.2% cells containing one sgRNA, 3.3% cells containing two sgRNAs, and 0.3% cells containing three sgRNAs. Taken together, 25.9% cells will be transduced (by one or more sgRNAs).

This indelible link between the MOI and the transduction efficiency can be exploited for the purposes of the functional titration of viral stocks. Conveniently, our system, which includes the fluorescent protein dTomato in the transfer vector, allows for the measurement of transduction efficiency by flow cytometry. Fluorescent proteins have been used for functional viral titration before (Reiser, 2000; Zhang et al., 2002, 2004). Briefly, increasing viral stock volumes are used to transduce the same number of cells in parallel, and after a period of incubation, fluorescence in cells is measured using flow cytometry. The amount of viral stock necessary for a transduction at any chosen MOI can then be worked out from the titration curve. In a similar vein, if the transfer vector contains an antibiotic resistance cassette, functional titration can also be achieved by measuring viability in antibiotically challenged versus unchallenged cell populations transduced with different volumes of viral stock (Diao et al., 2017; Doench et al., 2016; Joung et al., 2017).

A range of methods is also available for physical titration of the virus – for example, ELISA assays for determining the levels of viral proteins, such as Gag (p24) (Logan et al., 2004), or RT-qPCR for various

viral components (Kutner et al., 2009; Scherr et al., 2001). However, as a very small percentage (0.1 – 1%) of the total viral particles may be infectious, physical titration assays should be run from lysates or genomic DNA of transduced cells, rather than the viral stocks or supernatants, to provide acceptable levels of accuracy (Scherr et al., 2001).

Considering the available choices for the titration of our viral stocks, a functional titration through flow cytometry for dTomato is our least time-consuming and most precise option to assess the amount of virus necessary for our chosen MOI. To integrate the functional titration into the context of our CRISPR/Cas9 screen, virus will be generated according to the optimized protocol (Chapter 2.3), concentrated using Lenti-X Concentrator (Chapter 2.3.3), and frozen at -80°C in aliquots for at least 24 hours before the functional titration is performed. Because the functional titration of the virus needs to precede the CRISPR/Cas9 screen transduction, the CRISPR screen will always have to be carried out with viral stocks that had gone through a freeze-thaw cycle. Performing the functional titration with viral stocks that had been frozen and thawed before use will account for the small potential differences in the potency of a fresh and a frozen viral stock (Figure 5.14 in Chapter 5.7).

6.7 Representation and cell numbers

Several factors influence the required initial and ongoing cell number for the CRISPR/Cas9 screen experiment. These include the number of sgRNAs, the number of iBARs used for each sgRNA, the representation we want to maintain for each unique sgRNA/iBAR combination (the number of cells that are transduced with the sgRNA/iBAR combination at any given time point, also known as coverage), and the MOI at which we decide to work.

The number of sgRNAs in our case is 640. The recommended coverage for CRISPR screen experiments ranges between 300- and 1000-fold per unique sgRNA (Miles et al., 2016). The original iBAR studies used 4 iBARs per sgRNA, at a representation of 100- or 500-fold for each sgRNA/iBAR combination (cumulatively, 400- or 2000-fold coverage per each sgRNA, (Zhu et al., 2019)), although we decided to work with 9 iBARs per sgRNA on the practical account that this will further increase the accuracy of our results compared to the 4 iBARs per sgRNA strategy, and it is at no extra cost from Twist Bioscience, who will be preparing our sgRNA^{iBAR} library. The lowest MOI used in the original iBAR study was 0.3 (Zhu et al., 2019).

To work out how many cells would be needed to start a CRISPR/Cas9 experiment under a specific set of circumstances, the equation would be:

starting cell number

$$= \frac{\text{number of sgRNAs} * \text{number of iBARs per sgRNA} * \text{representation per sgRNA/iBAR}}{\text{MOI}}$$

Table 6.1 shows the required starting cell numbers for a range of low MOIs and representations in our CRISPR/Cas9-mediated mutagenesis, which uses 640 unique sgRNAs, each with 9 different iBARs. For example, if we choose to proceed with MOI = 0.3 and a representation of 250-fold per sgRNA/iBAR (2250-fold per sgRNA cumulatively), we will require 4.8×10^6 cells to begin the screen. Throughout the duration of the screen, the cells will be maintained at such numbers so that the coverage of the sgRNA^{iBAR} library is always at least as high as the initially chosen representation.

representation per sgRNA/iBAR (-fold)	cumulative representation per sgRNA (-fold)	Cells required to start the screen experiment at MOI =				
		0.1	0.2	0.3	0.4	0.5
500	4500	2.88E+07	1.44E+07	9.60E+06	7.20E+06	5.76E+06
250	2250	1.44E+07	7.20E+06	4.80E+06	3.60E+06	2.88E+06
100	900	5.76E+06	2.88E+06	1.92E+06	1.44E+06	1.15E+06
50	450	2.88E+06	1.44E+06	9.60E+05	7.20E+05	5.76E+05

Table 6.1: Starting cell numbers required for the CRISPR/Cas9-mediated saturating mutagenesis of the *TAL1* super-enhancer in Jurkat cells, for common representation and MOI ranges.

The example expounded in the main text is highlighted in pink.

The selected conditions directly influence practical parameters for the CRISPR/Cas9 screening experiment, such as the volume of media and Lenti-X Concentrator required, and the time needed for FACS-sorting of these cells (should we use FACS for the selection of transduced cells). For example, using the transduction protocols established for our validation experiments, 4.8×10^6 cells would be transduced in 24 ml media, at a cell density of 2×10^5 cells/ml, which could take place in a T-75 flask. Alternatively, as recommended in a recent review paper for increased transduction efficiency (Canver et al., 2018), cells could be transduced at a higher cell density of 5×10^5 cells/ml in small aliquots on a 24-well plate. Using the latest figures from optimization of virus generation, the making of the

concentrated lentiviral stock needed to transduce 4.8×10^6 cells at $\text{MOI} = 0.3$ would require around 300 ml Lenti-X Concentrator with the HEK293T cell line, according to the most recent optimization experiments (Figure 5.14 in Chapter 5.7). If the interpretation of our 293FT-generated virus transduction efficiency results is correct (Chapter 5.7; discussed in Chapter 5.9), and provided we find a solution to the extensive cell death of Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells treated with the 293FT-generated virus (Chapter 5.6; discussed in Chapter 5.9), we would be able to use the 293FT cells to generate the virus, which would bring down the requirement for Lenti-X Concentrator down to 40 – 70 ml (according to the optimization experiment, Figure 5.14 in Chapter 5.7). Assuming the highest possible sorting flow rate, 3000 cells/s, the net time required to FACS-sort 4.8×10^6 cells should be about 28 minutes. This set of conditions would constitute a feasible screening experiment.

It might also need to be taken into consideration, however, that in reality a lot of cells do not survive the FACS, and that to maintain the representation as high as initially, we would need to collect more cells to account for the deaths. The available absolute cell number counts from my previous FACS experiments (Chapter 5.5; absolute cell number data not shown), compared to the absolute cell number counts of unsorted cells (for example, Figure 4.15A in Chapter 4.10), suggest that the death toll of FACS might be as high as 9 out of 10 cells for our rather sensitive Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line, although an initial post-FACS decrease in proliferation, brought about by the stress of the procedure, likely contributes to this apparent (and highly approximate) rate of mortality. If we assume that the approximate survival rate of FACS-sorted Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells is 1 in 10, we would have to sort 4.8×10^7 cells (instead of 4.8×10^6 cells) to maintain the 250-fold sgRNA/iBAR representation, which would take around 5 hours. With this in mind, we might need to employ a lower representation of 100-fold per sgRNA/iBAR, which would achieve a very respectable 900-fold cumulative representation per sgRNA. This would require the sorting of 1.92×10^7 cells (Table 6.1), which would take ca. 1 hour and 45 minutes. These numbers could be further decreased by the employment of a slightly higher MOI – for example, if we transduce at $\text{MOI} = 0.5$, the number of cells we would need to sort would drop to 1.15×10^7 cells, which at the top recommended sorting speed of 3000 cells/s would take just over an hour to FACS-sort. Taken together, a range of feasible and high quality CRISPR/Cas9 screen execution options are available to us with our design.

6.8 Genomic DNA isolation

After samples for $T = 0$ and $T = \text{exp.}$ (Figure 6.1B and Figure 6.2) have both been collected, genomic DNA will be isolated from the cells. Both samples will be obtained from enough cells to maintain the chosen representation per sgRNA/iBAR. For instance, if we're working with 5760 sgRNA/iBAR combinations, the representation of 100-fold per sgRNA/iBAR, and the MOI of 0.3, we will need to collect at least 1.92×10^6 cells at $T = 0$ (before sorting – 74.1% of these cells are untransduced), and at least 4.8×10^5 cells at $T = \text{exp.}$ (sorted cells only include transduced cells). As an average human diploid cell contains around 6.6 pg DNA (Forslund et al., 2003), 1.92×10^6 cells ($T = 0$) should result into 12.7 μg of DNA material and 4.8×10^5 cells ($T = \text{exp.}$) into 3.17 μg of DNA material. The genomic DNA isolation will be carried out using a commercially available kit (for example, The Blood & Cell Culture DNA Mini/Midi Kit from QIAGEN) and according to manufacturer's instructions.

6.9 Library preparation

A two-step PCR will be employed to amplify the genome-integrated sgRNA cassettes from the isolated genomic DNA (Figure 6.8).

In the first step, we will make use of the unique primer binding site (UPBS) in front of the U6 promoter, which is comprised of 11 nt completely unique sequence and 11 nt of the P7 Illumina adapter sequence. The reverse primer for the first PCR step will be composed of the sequence corresponding to ca. 20 nt of the sgRNA scaffold sequence immediately 3'-from the iBAR; a sample barcode; a short stretch of balancer DNA; and a part of the Multiplexing Read 1 Illumina sequencing primer (Figure 6.8). The balancer DNA is inserted to base-balance the sequence immediately following the Multiplexing Read sequencing primer, which is necessary for an optimal Illumina performance (Optimizing Cluster Density on Illumina Sequencing Systems technical note, Illumina, 2016), and is used whenever the sequence immediately following the Multiplexing Read sequencing primer would otherwise be non-random, such as a sample barcode (e.g., (van der Lelij et al., 2020)).

In the second PCR step of the sequencing library preparation, the rest of P7 Illumina adapter sequence will be attached to the 5'-end of the product from the first PCR step, and the rest of the Multiplexing Read 1 Illumina sequencing primer, along with the P5 Illumina adapter, will be attached to the 3'-end of the product (Figure 6.8).

Illumina sequencing library preparation strategy

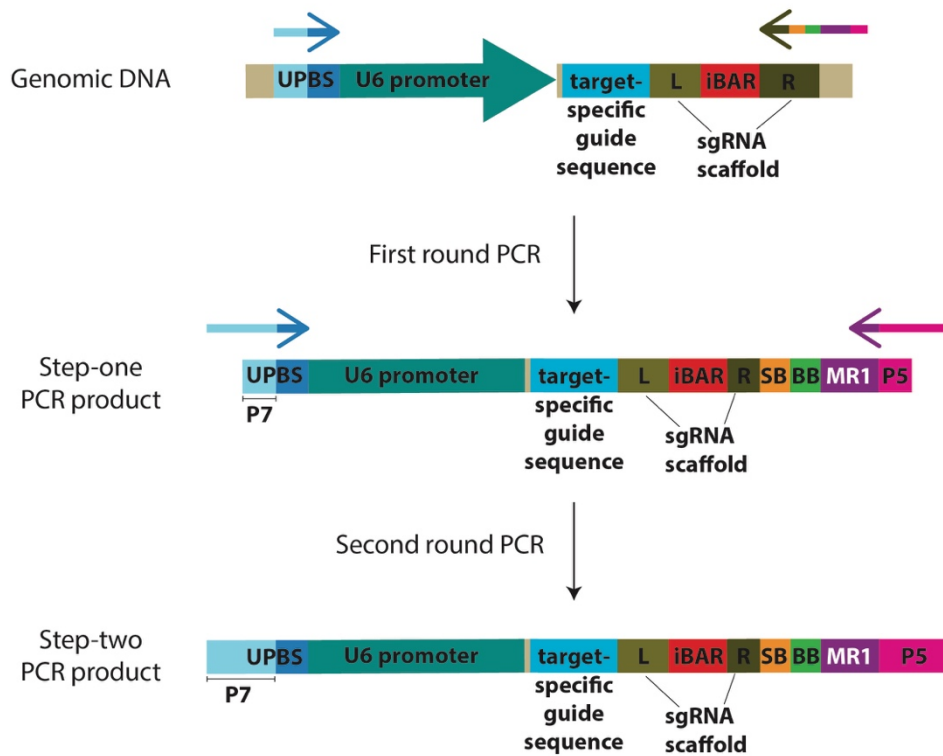


Figure 6.8: Illumina sequencing library preparation strategy.

Two rounds of PCR are employed to gain the desired product for Illumina sequencing. The first round amplifies the sgRNA cassette and the iBAR from the genomic DNA. The second round introduces the rest of the appropriate Illumina adapter sequences. UPBS – unique primer binding site; L/R – the left(5’)/right(3’) part of the sgRNA scaffolding; P7 – Illumina P7 adapter; SB – sample barcode; BB – base balancer; MR1 – Multiplexing Read 1 Illumina sequencing primer; P5 – Illumina P5 adapter. The colourful arrows indicate primers.

We will use a high fidelity polymerase (e.g., PfuUltra II Fusion HS DNA Polymerase) for both the amplifications, to minimize PCR errors. As template, we shall use 1 µg of the isolated genomic DNA per PCR (van der Lelij et al., 2020) or less, and set up multiple PCR reactions as needed to maintain the selected representation of the sgRNA^{iBAR} library. To reduce the potential for PCR bias, the amplifications will take place at low cycle numbers. The products will be pooled, subjected to agarose electrophoresis, isolated and cleaned using a commercially available kit (Canver et al., 2018).

6.10 Next-generation sequencing

To sequence our sgRNA^{iBAR} library, we will use Illumina paired-end sequencing of 150 nt from each end of our product (PE150) on the NovaSeq 6000 S4 platform. All the information that we need (sample barcode, iBAR and sgRNA sequence) will be acquired from one direction, but paired-end sequencing is the most cost-effective option with Novogene, which is the company that we will commission to carry out the sequencing of our library.

A recent review recommended, for a library of 1×10^5 elements, a sequencing depth of $1 - 2 \times 10^7$ reads at $T = 0$ and 1×10^8 reads at $T = \text{exp.}$ when the screen is based on negative selection and expected changes are subtle (Miles et al., 2016). For our modest library of 5760 sgRNA/iBARs, this would translate to $5.76 \times 10^5 - 1.15 \times 10^6$ reads for $T = 0$ and 5.76×10^6 reads for $T = \text{exp.}$ Illumina NovaSeq 6000 S4 is capable of delivering 2.66×10^9 reads per run if required.

For help with the evaluation of our data, we have enlisted the help of the Sheffield Bioinformatics Core. In general terms, the data will be trimmed to only contain the sgRNA sequences, the iBARs and the barcodes. The sgRNA sequences will be mapped onto the reference sgRNA library, and we will assess the fold change between the reads at $T = \text{exp.}$ and the reads at $T = 0$. The iBARs will be used as internal replicates. A piece of software based on the MAGeCK algorithm (Li et al., 2014) – MAGeCK^{iBAR} (Zhu et al., 2019) – has been developed for the analysis of sgRNA^{iBAR} library sequencing results.

6.11 Hit validation

However much we will be able to increase the accuracy of our screen results with the barcoding approach, there still remains the need for validating the potential hits from our CRISPR/Cas9 screen. As a minimum, we expect to register the sgRNAs that target the TAL1 complex TF binding sites, such as the MYB1 sgRNA in our validation experiments (Chapter 5).

To validate the results of the screening experiment, we will clone each of the sgRNAs that registered as a hit individually into a plasmid and transduce the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line with these plasmids in an arrayed screen format, monitoring the cell viability for each sample. We will also attempt doxycycline rescue in this format, where each differentially transduced sample can be tracked individually. We will test for *TAL1* decrease with RT-qPCR and *TAL1* downregulation with Western blotting. To better understand the nature of the changes leading to the loss of *TAL1* expression, we will also employ ChIP-seq for histone modifications associated with enhancers, such

as H3K27ac, and for the TAL1 complex TFs (e.g., MYB, RUNX1, GATA3). The latter will be instrumental in helping us discern between the hits caused by the loss of TF binding at the *TAL1* enhancer, and the hits that might have resulted from unfavourable changes in the structure of either of the *TAL1*-eRNAs. Finally, individual sgRNA transductions will enable us to sequence the *TAL1* enhancer locus, allowing us to learn about the mutations linked to each individual hit. The range of mutations associated with each hit can be further studied by the generation of mutated eRNA sequences *in vitro*, and the assessment of their performance in *in vitro* assays (e.g., the CBP-HAT activity assay).

6.12 Discussion

In the light of the validation results for doxycycline rescue in the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line, which showed that doxycycline had mixed effects on the cells, we decided that the best way to proceed with our CRISPR/Cas9 screen would be to procure a control sample from an early stage in the experiment, instead of a doxycycline-rescued control sample collected at the same time as the experimental sample (Figure 6.1 and Figure 6.2 in Chapter 6.1).

There would have been several advantages to the doxycycline rescue method. The first advantage would have been the removal of potentially toxic sgRNAs from the control sample, which can otherwise emerge as false positive hits. However, with a combination of hit validation methods, we should be able to distinguish between the true and the false positive hits from our adjusted $T = 0/T = \text{exp. screen}$. Secondly, avoiding the need for a $T = 0$ control sample would have also been advantageous because technically, there is no perfect $T = 0$ – one must find a reasonable compromise to allow enough time for the incorporation of the sgRNAs into the genome of the host cell line, but to avoid as much of the editing effects as possible. This might prove problematic to achieve in our case, as some of our validation experiments suggested that the editing effects in the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells might be manifesting as soon as 1 day post transduction (Figure 4.22 in Chapter 4.13). Furthermore, the sgRNA integration into the genome does not happen at the same time in all the cells and the editing may take place at different rates with different sgRNAs. These problems could be solved by using the information from the plasmid library sequencing (Canver et al., 2018), although this approach, in turn, does not account for potential differences in sgRNA integration efficiency. However, the collection of a $T = 0$ sample from cells, as well as using the plasmid library sequencing data as the control group, are common approaches, which have been successfully applied in the past (Canver et al., 2018). Yet another potential solution would be to transduce the plasmid library into a

cell line that does not express the Cas9 enzyme, and/or is not dependent on TAL1 (personal communication, Dr. Julian Jude). Taken together, even if they might all have their small potential disadvantages, there is a range of options to choose from in terms of a $T = 0$ control sample for our CRISPR/Cas9-mediated mutagenesis of the *TAL1* enhancer in the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line.

In addition to the two time points, $T = 0$ and $T = \text{exp.}$, we might also choose to collect samples at times between the start and the end of the screen (previously done in e.g. (Dixit et al., 2016; Thompson et al., 2021; Wang et al., 2019)). While collecting a single experimental end-point sample is common practice in CRISPR/Cas9 screens, sequencing samples collected at extra time points can sometimes provide additional information (Bock et al., 2022). In our case, this might be useful especially if we end up using hygromycin for the selection of transduced cells in our screen, because the antibiotic selection can partly cloud the effects of CRISPR editing (Chapter 4), and learning more about the representation of individual sgRNAs throughout the course of the screening experiment could help us flag sgRNAs that could go unnoticed in the comparison of the starting point and the end-point sample.

For our screen, we will use experimental sgRNAs that cover the transcribed portion of the *TAL1* enhancer (20% of our sgRNA library), and a large number of positive (14%) and negative (66%) control sgRNAs. The changes we expect from the experimental sgRNAs are subtle – MYB1 sgRNA, which is designed to bring the entire *TAL1* enhancer down by damaging the MYB-binding site, and is therefore the experimental (*TAL1* enhancer-targeting) sgRNA with the highest possible effect on cell growth, generated a modest 13% decrease in proliferation compared to control on day 22 in our most recent experiment (Figure 5.13 in Chapter 5.6), although with the use of FACS, there was a 60% drop in cell proliferation compared to control on day 15 after FACS-sorting (Figure 5.11 in Chapter 5.5). The growth challenge resulting from *eTAL1* alterations is expected to be much lower. This is why we chose to employ an especially high number of negative control sgRNAs: the hope is that a more robust baseline might help us to better distinguish the small, but real changes in sgRNA representation from the background.

Another approach we will integrate in order to increase the signal to noise ratio is sgRNA barcoding. While CRISPR/Cas9 editing is a powerful tool, it may often result in harmless or even favourable mutations rather than deleterious ones. Innocuous mutations may then confound the overall screen results. To tackle this issue, several recent papers (Michlits et al., 2017; Schmierer et al., 2017; Zhu et al., 2019) described a few different ways for tracking the separate fates of different cells or subgroups of cells transduced with the same sgRNA in CRISPR/Cas9 experiments, providing an insight into the distribution of the range of effects caused by an sgRNA, rather than just the average of these effects

in the whole of the cell population. Two groups used a large number of randomly generated barcodes, which they randomly combined with the sgRNAs in a pooled cloning step, so that each sgRNA was associated with many different barcodes, and each barcode was associated with many different sgRNAs (Michlits et al., 2017; Schmierer et al., 2017). The advantage of this approach is that the large number of the sgRNA-barcode combinations enables a so-called lineage dropout analysis (LDA), in which each single editing outcome (originating from one cell), or a very small group of editing outcomes (originating from a very small group of cells), can be tracked separately. The drawback of this approach is that it can be quite technically challenging to achieve an equal sgRNA representation in the library, because the sgRNA-barcode library has to be cloned in two steps; as a result, bioscience companies shy away from this option (personal communication, Dr. Julian Jude). The other study that introduced barcodes in recent years was the previously mentioned iBAR paper (Zhu et al., 2019). This study uses barcodes that are randomly generated, but non-randomly assigned. Each sgRNA is combined with a small number of barcodes (this can be the same set for each sgRNA); the sgRNA-barcode combination is synthesized in one step, as a single oligonucleotide and cloned into an sgRNA expression vector. This approach doesn't allow for LDA, but increases the precision and accuracy through enabling internal replicate analysis (IRA), where each unique sgRNA-barcode combination acts as an internal replicate within the pooled screen. This approach to sgRNA library preparation is offered and recommended by Twist Bioscience, which is the company whose services we are planning to use.

To improve the stability and the efficiency of our sgRNAs, we will change a couple of things in the sgRNA scaffolding in our plasmid. Firstly, an 'A-U flip' at the base of the first stem loop (Figure 6.5B in Chapter 6.4) will remove four consecutive U's, a putative termination sequence for RNA Polymerase III (RNAPIII) (Chen et al., 2013; Nielsen et al., 2013) – the enzyme that transcribes sgRNAs from the commonly used U6 promoter in mammalian cells. Secondly, an extension of the stem leading up to the tetra loop (Figure 6.5B in Chapter 6.4) will contribute to an enhanced sgRNA-Cas9 assembly (Chen et al., 2013). Both of these alterations were tested separately and jointly in a dCas9-EGFP imaging system, and their combination was shown to have the most favourable impact on increasing the precise localization of the incapacitated enzyme and decreasing the background noise (Chen et al., 2013). The altered sgRNA scaffolding (with both the 'A-U flip' and the extended stem) was also tested in a small essential gene CRISPR/Cas9 screen against the natural design, and found supreme in terms of performance and speed of editing, without any apparent off-target effects (this was inferred from the negative control sgRNA enrichment, rather than detailed testing by global genome sequencing) (Cross et al., 2016). The doubly-improved sgRNA structure was also successfully used in a CRISPRi screen (Qi et al., 2013), as well as in the iBAR-CRISPR/Cas9 screens in Zhu's study (Zhu et al., 2019).

While our strategy for the CRISPR/Cas9-mediated mutagenesis of the *TAL1* super-enhancer originally relied on doxycycline rescue for a control sample, an alternative strategy is available to us in the form of an early time point control sample. The associated challenges can be hopefully overcome by a combination of measures, including a large number of negative sgRNA controls present in the screen, barcoding, and rigorous hit validation.

7 Discussion

The aim of my doctoral project was to functionally dissect the transcribed portion of the *TAL1* super-enhancer in Jurkat cells, to learn more about the workings of the two enhancer RNAs (eRNAs) that arise from this enhancer, the sense and the antisense *eTAL1*. In a very broad sense, this endeavour was to further contribute to the gargantuan mass of information that has been collected to date about the non-coding genome. The enormous extent of existing knowledge about non-coding genome reflects its paramount importance in health and disease. At the same time, however, there are still many gaps in our knowledge to fill.

Within the vastness of the non-coding genome (98% of human genome, or about 3 billion base pairs, (ENCODE Project Consortium, 2012), Chapter 1.1), our lab is interested in enhancers, regulatory elements that are capable of increasing transcription rates of their target genes up to several hundred times (Chapter 1.2). This makes enhancers one of the key players in transcriptional regulation and highlights their immense importance of their correct functioning. The repercussions of enhancer malfunction can be multifarious and severe: embryonic lethality, developmental defects, autoimmune disorders, Alzheimer's disease, cancer. A substantial body of research has been conducted on the subject of enhancers since their discovery in 1981, but much still remains obscured or unexamined.

7.1 The particularities and problematics of working with eRNAs

One of the largely obfuscated areas in our understanding of enhancers is the importance, biology, structure and function of eRNAs, transcripts that arise from enhancers in a genome-wide (and not yet fully understood) manner (Chapter 1.3; (De Santa et al., 2010; Kim et al., 2010)). Since their discovery as a ubiquitous RNA species in 2010, relatively sporadic reports of the functional implications of eRNA existence have been trickling into the vast pool of enhancer knowledge. There are multiple reasons as to why eRNAs as a species have not enjoyed more spotlight.

Firstly, eRNAs generally exist in relatively low numbers (one study reported eRNA numbers per cell to be less than 100 in the most abundant of cases (Li et al., 2013)), which can make it difficult to obtain enough material for an experiment. The low abundance is likely due to the instability of a typical eRNA molecule. This is also the reason why a lot of information that we have about eRNA function comes from probing eRNAs that are lncRNAs at the same time (i.e., lncRNAs arising from enhancers). These e-lncRNAs are usually polyadenylated and much more stable than classic eRNA transcripts.

Secondly, further to why studying especially the function of eRNAs is complicated, many methods available to study the functionality of eRNAs do not make it possible to disentangle the function of the act of transcription of the eRNA and the function of the transcript in its own right. Enhancer RNA synthesis may be blocked at various steps with a palette of inhibitory substances, such as flavopiridol, α -amanitin, actinomycin D or triptolide, but while some of these are specific to RNAPII (Martin et al., 2020), none of these are specific to eRNA transcription, let alone eRNA transcription at a specific locus. Enhancer RNA transcription can be specifically silenced by CRISPR interference (CRISPRi) (Chapter 1.4.2.3.1; (Blank-Giwojna et al., 2019; Che et al., 2020; Huang et al., 2021)) or specifically activated by CRISPR activation (CRISPRa) (Chapter 1.4.2.3.2; (Blank-Giwojna et al., 2019; Carullo et al., 2020)), but both of these methods, while highly targeted and clinically relevant, do not allow to distinguish between the effect of enhancer transcription and the role of the eRNA transcript per se. Similarly, long CRISPR-mediated deletions (Chapter 1.4.2.2), which can be used to excise a part (or the whole) of a transcribed portion of an enhancer (Lidschreiber et al., 2021; Sakaguchi et al., 2018), again fail to unambiguously ascribe a role to the eRNA molecule in itself.

A strategy that can distinguish between the role of the act of eRNA transcription and the role of the eRNA transcript in its own right is the specific knockdown of the selected eRNA transcript using siRNAs, shRNAs or (LNA) ASOs. This is how the majority of the discoveries about eRNA transcript function were made (Chapter 1.3.3.3). While bringing us a step closer to understanding what the functions of eRNAs are, eRNA knockdown cannot answer how the eRNAs carry these functions out.

The question of how eRNAs work is a tricky one to tackle, and to our knowledge, it has not been attempted previously. By comparison with other RNA species, it is a highly likely speculation that the eRNA functionality will be closely connected to its structure (Chapter 1.3.2). We decided to base our investigation of eRNA function on this hypothesis.

To exploit the relationship between structure and function, we chose to employ the CRISPR/Cas9 technology (Chapter 1.4.1) in a screen format (Chapter 1.4.2.1) to generate point mutations in the underlying DNA sequence of a pair of selected functionally important eRNAs and then assess the functional consequences of the genetic manipulations. This assessment should allow us to start mapping the functionally important regions within these eRNAs.

7.2 Our chosen approach to studying the functionality of eRNAs

For our experimental system, we decided to use the transcribed region of the monoallelic *TAL1* super-enhancer in Jurkat cells, which gives rise to a sense and an antisense eRNA (*eTAL1*, Chapter 1.5, (Mansour et al., 2014)). Previous experiments from our lab confirmed that as a minimum, the sense *eTAL1* transcript plays an important role in the upregulation of *TAL1* (Chapter 1.6). *TAL1* is a key transcription factor in the Jurkat cell line; survival and growth of the cells is dependent on high *TAL1* levels (Palii et al., 2011; Palomero et al., 2006; Sanda et al., 2012). Taken together, this means that in such cells where the CRISPR-inflicted mutations disrupt important regions or structures and render the *eTAL1s* less functional, *TAL1* levels will drop, and so will, concomitantly, the proliferation of these cells.

Concerning methodology, we chose to dissect the transcribed portion of the *TAL1* enhancer by a CRISPR-Cas9-mediated saturating mutagenesis in a dropout screen format (Figure 1.9 in Chapter 1.4.2.1, Figure 1.16 in Chapter 1.6; Figure 6.1 in Chapter 6.1; Figure 7.1). This means using a pooled library containing all the possible unique sgRNAs against the selected sequence, introducing this library into Jurkat cells, and assessing which sgRNAs caused a decrease in proliferation as a consequence of a drop in *TAL1* levels.

CRISPR/Cas9-mediated saturating mutagenesis has been done for enhancers previously (Canver et al., 2015, 2020), but without a focus on eRNAs. In one of the papers, in which *BCL11A* enhancer was dissected, the top hits were mapped onto transcription factor binding sites (Canver et al., 2015). In the other study, the approach was used to identify regulatory elements of *Oct4*, rather than to dissect them (Canver et al., 2020). Instead, we set off to work with an enhancer that has already been identified and very well described (Mansour et al., 2014). Similarly to the *BCL11A* enhancer saturating mutagenesis (Canver et al., 2015), we would expect our top hits to map onto the binding motifs for the transcription factors important for the *TAL1* enhancer transcription. These would be the transcription factors of the *TAL1* complex at the heart of the *TAL1* enhancer that bind directly to DNA – GATA3, RUNX1, MYB, HEB and likely also E2A (Figure 1.13 and Figure 1.14 in Chapter 1.5; (Mansour et al., 2014)). Apart from the presumably strong hits that damage the transcription factor binding sites, however, we would also look for potentially weaker hits that disrupt the body of the *eTAL1* transcripts in ways that prevent them to fully perform their function. These are the hits that are of primary interest to us, the hits that would begin to tell the story of the important functional regions within eRNAs.

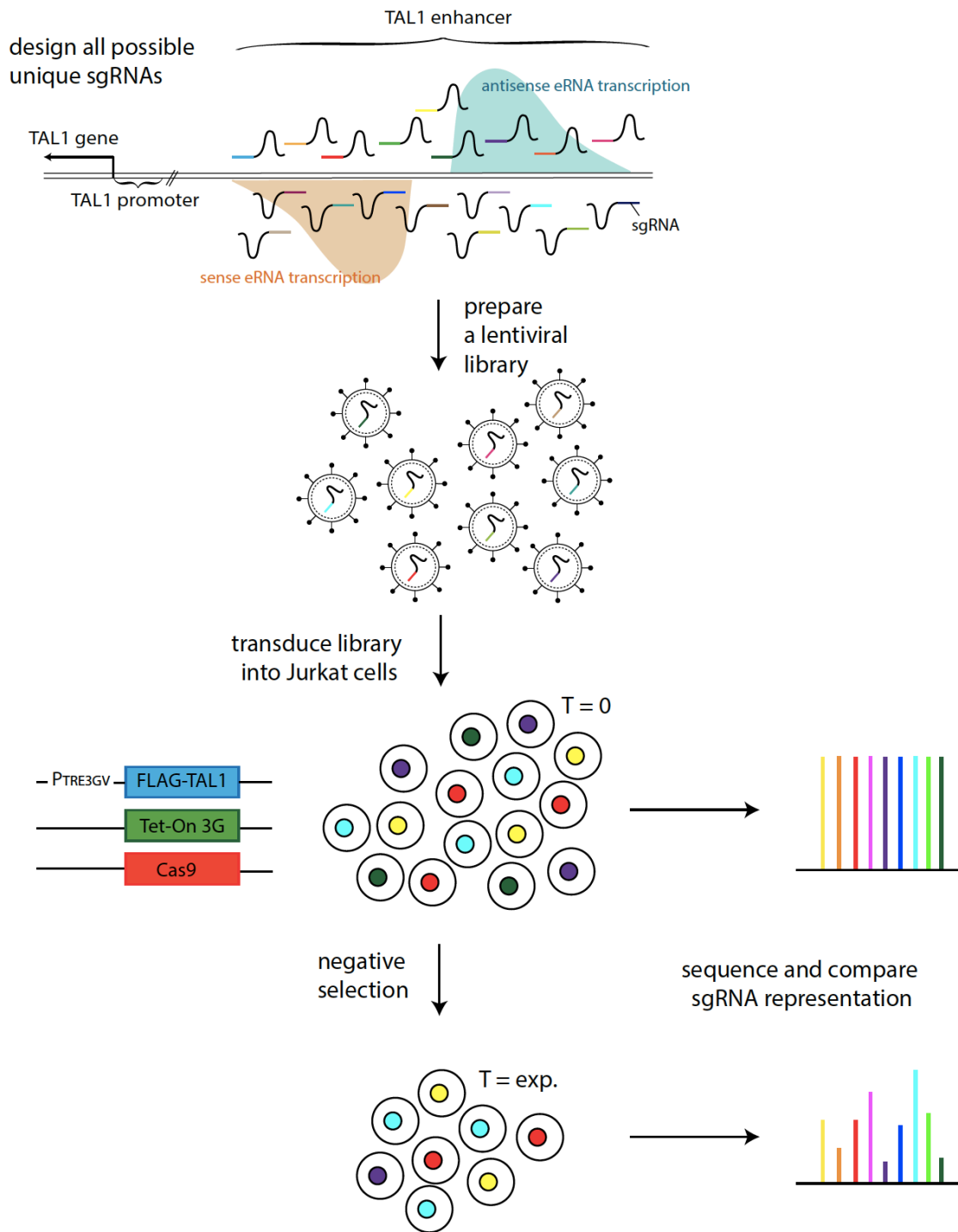


Figure 7.1: Workflow for the CRISPR/Cas9-mediated saturating mutagenesis of the *TAL1* enhancer, reimaged (copy from Figure 6.2).

All unique sgRNAs targeting the transcribed portion of the *TAL1* enhancer are designed, synthesized and cloned into a plasmid, which is then encased in lentivirus packaging. Such lentiviral library with an equal representation of all designed sgRNAs is then transduced into the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line. At an early time point ($T = 0$), cells are collected to provide for a control group. After negative selection has taken place, the remainder of the cells is collected as the experimental group. Both samples are sequenced, abundance of sgRNAs is compared. sgRNAs responsible for functionally important mutations are expected to be depleted in the experimental group.

In our CRISPR/Cas9-mediated mutagenesis of the *TAL1* enhancer, we are interested in the sgRNA representation in an experimental group of Jurkat cells that were affected by *TAL1* enhancer mutations and the potentially linked decrease in TAL1, and left to proliferate for 2 – 3 weeks to allow for the effects of the inflicted mutations to play out. As a reference base point, a control group with an equal representation of all the sgRNAs in the pooled *TAL1* enhancer-targeting library is needed. To this end, we originally planned to use a group of cells expressing an exogenous, doxycycline-inducible FLAG-TAL1 as rescue. This rescue strategy was to ensure that any lack of TAL1 caused by the CRISPR/Cas9 targeting of the *TAL1* enhancer during the 2 – 3 weeks of proliferation after the sgRNA library introduction would be replenished by the exogenous variant. In other words, both the control and the experimental cells would be collected after a period of proliferation at the same time point, except the control group would have been treated with doxycycline for the whole time to keep a steady store of TAL1 for potential rescue of growth rate in affected cells, whereas the experimental group would have been left to its own devices (Figure 1.16 in Chapter 1.6).

For this purpose, on the background of the Jurkat cell line, I established the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line, which expressed a doxycycline-inducible FLAG-TAL1 (Chapter 3). The cell line also boasted a stable expression of Cas9, which removed the need for Cas9 co-transduction at the time of the introduction of the sgRNA library. This removed cell-to-cell variability between Cas9 expression levels (Chapter 3.4) and likely also decreased the reaction time of the system (compare graphs in Figure 4.22 in Chapter 4.13).

Because the cells were genetically modified multiple times and passaged for many months before one of them became the final monoclonal Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line, we first tested whether the cell line had preserved its dependability on TAL1, a feature key for a viable CRISPR/Cas9 screen as we planned it. If the cell line had lost its TAL1 addiction during the months of passaging and genetic alterations, the rescue strategy would have been unviable. Fortunately, after a set of optimizations, we succeeded in proving that the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line was dependent on TAL1 in the same way as the Jurkat cell line is (Chapter 4).

The second aspect of the newly established cell line crucial for a viable CRISPR/Cas9 screen as we planned it was a well-functioning doxycycline-mediated FLAG-TAL1 rescue (Chapter 5). In the process of confirming the functionality of this feature, we realized that the exposure of Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cells to doxycycline caused cytotoxicity, and that while we observed some rescue-like effects on the cell line upon doxycycline treatment of TAL1-challenged cells, the measure of the rescue was variable and heavily dependent on a number of factors, which included the type of sgRNA (Chapter 5). Realizing the lack of robustness of the doxycycline rescue approach was irremediable, we

proceeded to look for another way the screen could be carried out, without the use of the FLAG-TAL1 rescue.

While the doxycycline rescue strategy that we chose to employ to begin with has its theoretical benefits and an indisputable elegance to it, it is a common approach to collect control cells early on in a CRISPR/Cas9 screening experiment, ideally before the effects of Cas9 editing start taking place, but after the sgRNAs have been introduced into the genome (discussed in Chapter 6). There are also other strategies that allow for the procurement of a control group with an equal sgRNA representation (sequencing of the plasmid stock; transduction of the library into a different cell line that is not sensitive to TAL1 depletion). Therefore, we were able to continue with the planning of the CRISPR/Cas9 screen in the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line (Chapter 6). The benefit of employing this cell line for the pooled CRISPR/Cas9 screening experiment is that we can use the doxycycline rescue on it in an arrayed format once we proceed into the hit validation stage, as the viability of the cells can be monitored throughout the rescue process in this format, allowing us to follow the viability patterns, rather than just to check an end point without reference to prior developments (discussed in Chapter 6). Moreover, the cell line already expresses Cas9, so practically, we are also avoiding the need to establish and validate a new cell line.

7.3 Further potential adjustments to our current system

A couple of further adjustments could be applied to the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line to potentially improve the system for the CRISPR/Cas9-mediated mutagenesis of the *TAL1* enhancer, to achieve an increased precision, accuracy and sensitivity.

Firstly, a useful addition to the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line would be an EGFP tag on the endogenous TAL1 on the allele that carries the *TAL1* enhancer in Jurkat cells (Chapter 1.5, (Mansour et al., 2014)). This would furnish us with a convenient way of monitoring TAL1 expression levels and allow us to carry out an enrichment screen instead of, or alongside the currently planned dropout screen. Enrichment screens can be very sensitive and powerful (Miles et al., 2016), which would be a valued asset in our experimental system since we are expecting only small differences in TAL1 expression upon *eTAL1* targeting. On the other hand, there is always the unwelcome possibility that the EGFP-tagging would influence the functionality of the endogenous protein, as well as the difficulty of obtaining a monoallelic EGFP tag. Importantly, EGFP-based enrichment screens of non-coding genome have been successfully carried out before (Canver et al., 2020; Diao et al., 2016, 2017; Rajagopal et al., 2016).

Secondly, a further adjustment to our system that could theoretically enable us to acquire even more power and sensitivity would be the deletion of the *TAL1* gene on the allele that does not feature the monoallelic *TAL1* enhancer. That way, the only source of TAL1 in the Jurkat cell line would be the allele where TAL1 expression is under control of the *TAL1* super-enhancer, simplifying the experimental system as it currently is. This would be technically similar to using the haploid cell line HAP1, in which each chromosome, and by extension, each allele, is represented only once (Gasperini et al., 2017; Yamazaki et al., 2018). However, data from Mansour's study strongly indicated that the TAL1 overexpression in Jurkat cells is almost exclusively due to the *TAL1* super-enhancer, with the other allele contributing possibly a few percent of the overall *TAL1* expression at best (Mansour et al., 2014). Therefore, this alteration might just be more work than it is worth.

7.4 Other CRISPR/Cas-based strategies for the dissection of the *TAL1* enhancer in Jurkat cells

Aside from the alterations described above, there would be many other potential approaches to performing the *TAL1* enhancer CRISPR/Cas9-mediated saturating mutagenesis in Jurkat cells starting from the maternal cell line. An elegant choice would be Rajagopal's multiplexed editing regulatory assay (MERA; described in Chapter 1.4.2.1; Figure 7.2; (Rajagopal et al., 2016)), in which a single copy of a cassette containing the U6 promoter and a dummy sgRNA is first integrated into the cells; then sgRNAs, with homology arms attached by a PCR reaction, are introduced into the cells; a double-strand break (DSB) is induced at the dummy sgRNA site; and the dummy site is replaced by an sgRNA from the introduced library via homology-directed repair (HDR). This strategy would help us avoid the need for plasmid library cloning (cutting the work load, the time requirement and the costs) and ensure a single sgRNA integration per cell, as well as permit us to use the much simpler Neon electroporation approach instead of the rather onerous transduction procedure. For the targeted integration of the sgRNA expression cassette, Rajagopal and colleagues used the mouse *ROSA* locus as they worked with mouse embryonic stem cells (Rajagopal et al., 2016); in Jurkat cells, we could use, for example, the adeno-associated virus integration site 1 (AAVS1) or the chemokine (C-C motif) receptor 5 (CCR5) safe harbour loci (Lombardo et al., 2011), or potentially even acquire the Flp-In-Jurkat cell line (Thermo Fisher). MERA relies on HDR (Rajagopal et al., 2016); the HDR rates in Jurkat cells vary immensely depending on the integration site (Schubert et al., 2021), for example with around 18% efficiency at the CCR5 locus following the electroporation of the Cas9/GFP donor template complex (Jayavaradhan et al., 2019). The HDR rate can be further increased (for instance, from 18 to 25% in the previous

example) by the employment of a NHEJ-suppressive Cas9-DN1S fusion, DN1S being a dominant negative version of the pro-NHEJ factor p53-binding protein 1 (Jayavaradhan et al., 2019). A small issue with MERA is that there is no selection mechanism for transfected cells, although with the modest size of our sgRNA library, this would be a perfectly manageable matter. Taken together, MERA would be a viable and advantageous alternative option for our CRISPR/Cas9-mediated saturating mutagenesis of the *TAL1* enhancer in Jurkat cells.

HDR-based sgRNA cloning

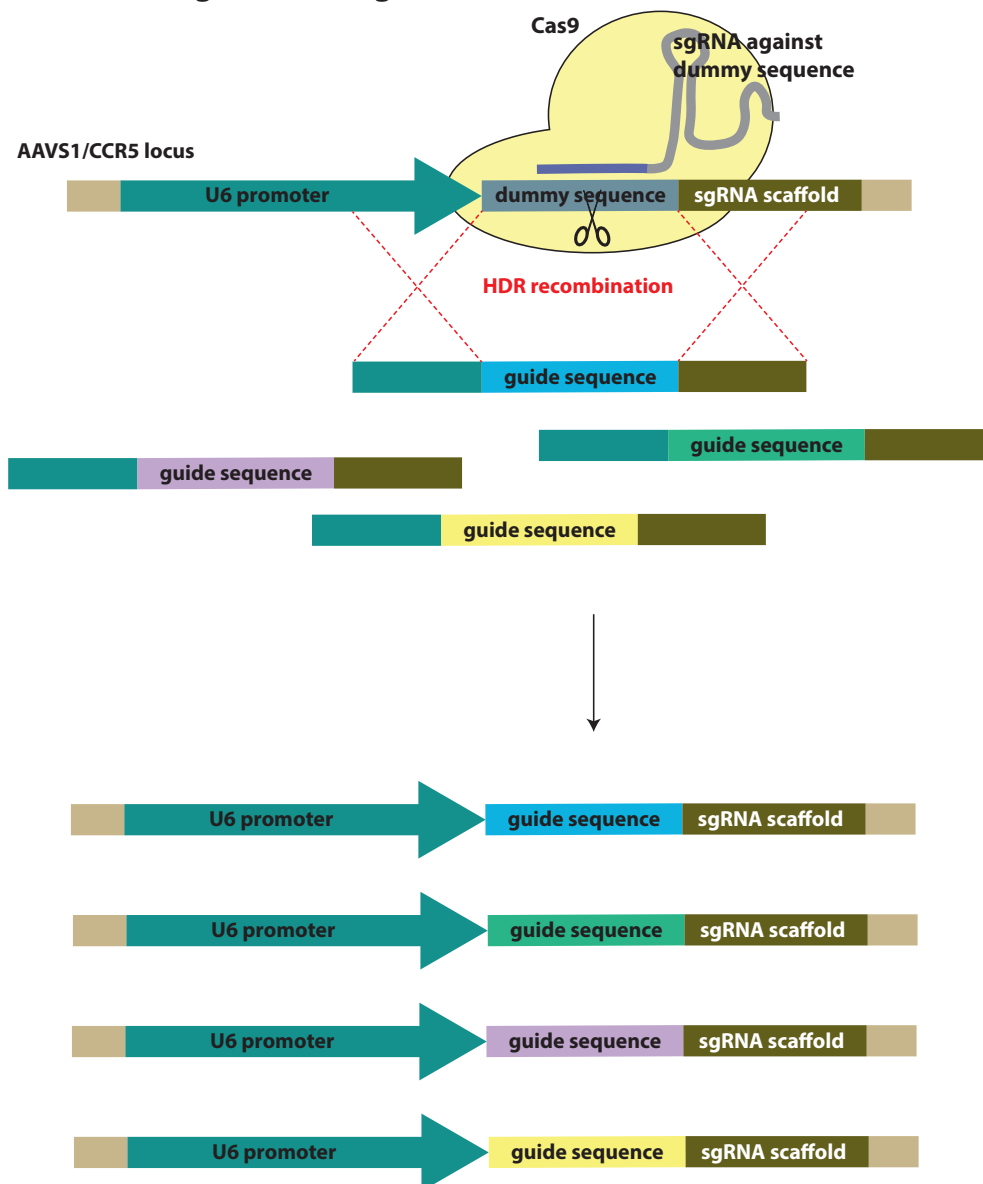


Figure 7.2: Targeted integration of sgRNAs into the genome based on HDR recombination that could be used as an alternative to our system.

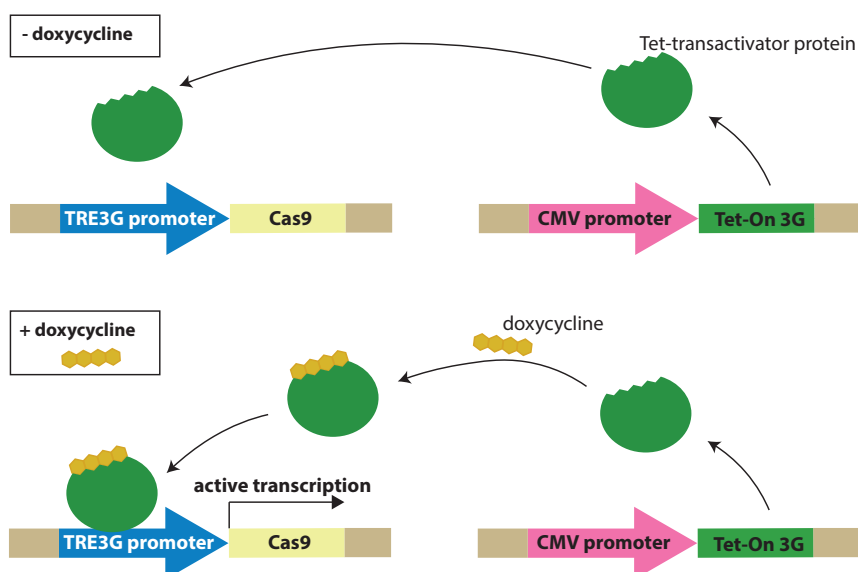
Method published in (Rajagopal et al., 2016).

Another option would be the establishment of a cell line with an inducible Cas9 expression on the background of our first intermediate cell line, Jurkat-Tet3G (Figure 7.3A). Inducible Cas9 has been previously used in several studies (Cao et al., 2016; González et al., 2014; Lundin et al., 2020; Wang et al., 2014) and a plasmid for doxycycline-inducible Cas9 expression is also commercially available from Takara (the Lenti-X Tet-On 3G CRISPR-Cas9 System). In this setup, doxycycline would only be needed in the culture for a short time, to allow for the CRISPR/Cas9 editing to happen, which would help us avoid the problem of doxycycline cytotoxicity observed with longer term supplementation. Furthermore, an inducible expression of Cas9 could also help us avoid the lack of robustness that we observe in the Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell lines, which is likely linked, at least in part, to the stable expression of the enzyme, rather bulky (160 kDa) and exogenous in nature. On the other hand, as we have seen with our doxycycline-inducible FLAG-TAL1, the Tet-on system that we are using can be fairly leaky (Chapter 3.3), and the levels of Cas9 needed for on-target editing are very low (Shen et al., 2019). Therefore, we would definitely have to establish monoclonal cell lines, and be extremely rigorous when choosing the one to work with (possibly, RT-qPCR to narrow down the options, but followed by stringent functional testing – for example with an anti-TAL1 sgRNA – to make sure the background expression does not lead to editing).

Aside from the doxycycline-inducible Cas9 system, which could be established in a single step on the background of our Tet-transactivator protein-expressing Jurkat cell line (Chapter 3.2), there is a host of other tried and described strategies to regulate the expression of Cas9 in a cell line, with many of them more leak-proof than the Tet-on system (reviewed in (Zhang et al., 2019)). One approach is to place a loxP-poly-A-loxP cassette between the Cas9 promoter and the Cas9 gene (Figure 7.3B; (Hans et al., 2021; Platt et al., 2014)). Cas9 expression in this system is activated at the desired timepoint by the Cre recombinase-mediated removal of the loxP-flanked poly-A signal. This can be as straightforward as the introduction of a plasmid encoding the gene for Cre recombinase (Platt et al., 2014). A potential disadvantage of this system is that the Cas9 cannot be turned off once it's been switched on.

Transcriptional regulation of Cas9 expression

A Doxycycline-inducible Cas9



B Cre/lox inducible Cas9



Figure 7.3: Transcriptional regulation of Cas9 expression that could be used as an alternative to our system.

(A) Doxycycline-inducible Cas9 cell line could be built on the background of our Jurkat-Tet3G cell line. (B) A Cre-inducible Cas9 system could be established in one step by the introduction of a Cas9 controlled by CMV promoter with an interposed floxed (poly-A)_n cassette ('stop').

Another approach to generating inducible Cas9 is to use the chemically induced proximity (CIP) method, in which two binding partners, each fused to a portion of Cas9, are brought into physical proximity by a small molecule, precipitating the unification of the associated Cas9 enzyme. Such regulation has previously been achieved with the use of rapamycin, which leads to the binding between the 12 kDa FK506 binding protein 12 (FKBP) and the 12 kDa FKBP rapamycin binding domain (FRB) (Figure 7.4A; (Nguyen et al., 2016; Zetsche et al., 2015)). To avoid high background editing rates due to the self-assembly of the split Cas9, Zetsche and colleagues sequestered the N-terminal part of the Cas9 fused with FRB in the cytoplasm using a nuclear export sequence (NES) (Zetsche et al., 2015), while Nguyen and co-workers sequestered both of the Cas9 moieties in the cytoplasm by adding a further fusion with the ligand-binding domain (LBD) of the oestrogen receptor (ER) (Nguyen et al.,

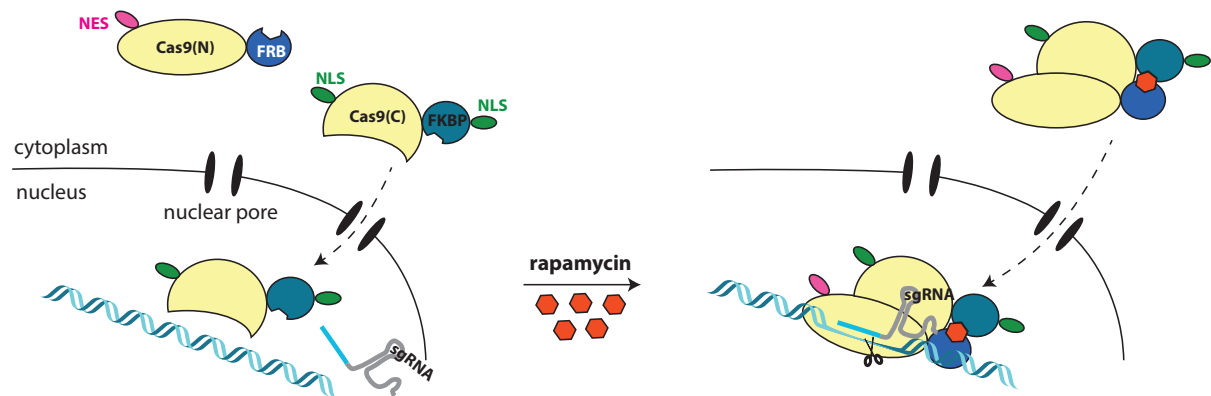
2016). Under normal circumstances, ER is sequestered in the cytoplasm by heat shock protein 90 (Hsp90), but upon ligand binding (oestrogen or ER modulators, such as 4-hydroxytamoxifen (4-OHT)), the ER-LBD undergoes a conformational change, Hsp90 binding is lost and the ER with the bound ligand translocates into the nucleus (reviewed e.g. in (Okat, 2018)). Both groups were able to achieve efficient editing with the reconstituted Cas9 upon rapamycin/4-OHT induction and very low levels of background editing in the absence of the inductors (Nguyen et al., 2016; Zetsche et al., 2015).

A set of further interesting Cas9 induction strategies has been based on the 4-OHT-mediated regulation of the ER-LBD or its oestrogen-insensitive mutants (Figure 7.4B – D; (Buskirk et al., 2004; Davis et al., 2015; Oakes et al., 2016; Zhao et al., 2018)). Firstly, Cas9 regulation was accomplished with a 4-OHT-responsive intein (self-splicing protein element) (Figure 7.4B; (Davis et al., 2015)). A directed evolution in *M. tuberculosis* rendered a version of ER-LBD flanked by a pair of intein that underwent self-cleavage only upon 4-OHT stimulation (Buskirk et al., 2004). This intein was inserted into the Cas9 structure, blocking the enzyme activity until the self-splicing of the intein was induced by 4-OHT (Davis et al., 2015). While this system showed an increased on-target specificity compared to wild-type Cas9, and an only slightly diminished activity in the presence of 4-OHT compared to the wild-type enzyme, the differences between 4-OHT-treated and 4-OHT-untreated samples ranged, depending on the target, between 3.4 – 9.6-fold (Davis et al., 2015). While this seems like a relatively high background compared to the Tet-On approach in our experience (Chapter 3.2), it is essential to mention that these experiments were done in polyclonal cell lines, and it is therefore highly likely that if monoclonal cell lines were established and tested, much lower background editing levels could be achieved.

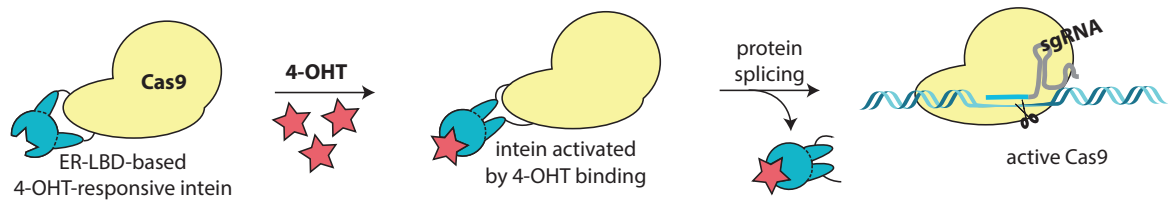
Secondly, regarding the use of 4-OHT ER-LBD-inducible systems, a recent study made use of the fact that ER translocates from the cytoplasm into the nucleus upon ligand binding. Zhao and co-workers added two NES sequences and two modified, selectively 4-OHT-responsive ER-LBDs to the C-terminus of their Cas9, creating a 4-OHT-inducible Cas9 system based on nuclear translocation (Figure 7.4C; (Zhao et al., 2018)). The efficiency of this translocating Cas9 upon induction was about twice lower than the efficiency of a nucleus-localized Cas9, and the difference in on-target editing efficiencies between a 4-OHT-treated and 4-OHT-untreated sample was between 5- and 15.7-fold depending on the individual experiment (Zhao et al., 2018). These experiments were again done in polyclonal cell lines, and so it is virtually without a doubt that a much less leaky performance could be achieved through monoclonality. Zhao and colleagues compared their translocation-based Cas9 system to a standard doxycycline-inducible one, and in their setup with polyclonal cell lines, their system achieved a 2.8-times lower background level of editing than the doxycycline-inducible one.

Translational regulation of Cas9 expression

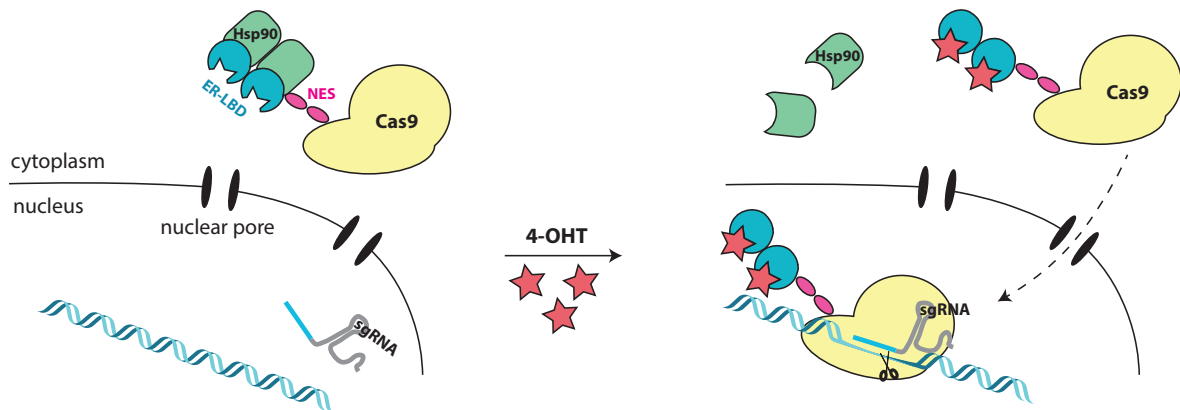
A Split Cas9 regulated by chemically induced proximity (CIP)



B 4-OHT-responsive intein



C 4-OHT/ER-LBD-based nuclear translocation system



D Allosterically regulated Cas9



Figure 7.4: Translational regulation of Cas9 expression that could be used as an alternative to our system.

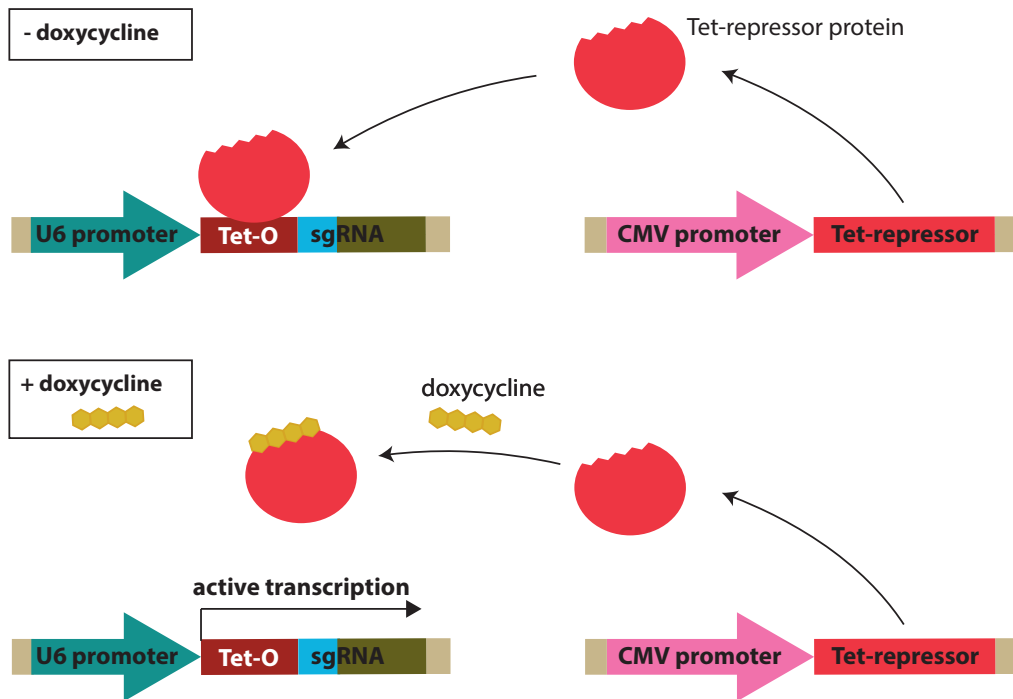
(A) Cas9 split into two portions, N- and C-terminal (Cas9(N), Cas9(C)), is reconstituted after the addition of rapamycin, which induces the binding of FKBP and FRB, two light-weight domains, each fused to one of the Cas9 moieties. NES – nuclear export sequence; NLS – nuclear localization sequence; FKBP – FK506 binding protein 12; FRB – FKBP rapamycin binding domain. Figure adapted from (Zetsche et al., 2015). **(B)** An oestrogen receptor ligand binding domain (ER-LBD)-based intein blocks Cas9 activity in the absence of 4-hydroxytamoxifen (4-OHT); 4-OHT supplementation triggers the self-splicing of the intein and activates Cas9. Figure adapted from (Davis et al., 2015). **(C)** Cas9 targeted into cytoplasm by a NES and fused to ER-LBD is translocated into the nucleus upon 4-OHT addition. **(D)** Cas9 engineered to be allosterically incapacitated by a fused ER-LBD undergoes an allosteric change upon 4-OHT binding to the ER-LBD, becoming activated.

Finally, in the realm of 4-OHT ER-LBD-inducible systems, Oakes and colleagues created an allosterically regulated Cas9 (arCas9) by first preparing a library of Cas9 with ER-LBD insertions at various positions within the structure of the enzyme and then scanning this library for positive response in the presence of 4-OHT and lack of response in the absence of 4-OHT (Figure 7.4D; (Oakes et al., 2016)). The authors first targeted the arCas9 into the nucleus with a nuclear localization sequence (NLS), and achieved a 6-fold difference in editing between the 4-OHT-treated and 4-OHT untreated samples. As this represented a rather high background of 10.9% editing efficiency in the absence of 4-OHT, Oakes and co-workers went on to employ nuclear translocation as a second level of arCas9 regulation. By simply removing the NLS, the group succeeded in further decreasing the background of their inactivated arCas9 editing to just over 1% editing efficiency, although the editing efficiency of the exported arCas9 in the presence of 4-OHT decreased from 66% to 30% due to the added regulatory step (Oakes et al., 2016). Once again, these results were obtained from polyclonal cell lines, and could therefore be undoubtedly improved by establishing clonal ones.

The 4-OHT-inducible Cas9 strategies proved to be effective systems that offer tight, reversible and dose-dependent regulation of Cas9 activity (Buskirk et al., 2004; Davis et al., 2015; Oakes et al., 2016; Zhao et al., 2018). As an additional benefit, apart from the easy-to-supplement 4-OHT, the systems do not rely on any other components to function (in the way that the Tet-On system relies on the Tet-transactivator protein, for example), which would make the potential establishment and testing of monoclonal cell lines easy and relatively fast to accomplish.

Transcriptional regulation of sgRNA expression

A Doxycycline-inducible sgRNA



B Cre/lox inducible sgRNA



Figure 7.5: Transcriptional regulation of sgRNA expression that could be used as an alternative to our system.

(A) A Tet-operator (Tet-O) interposed between the U6 promoter and the sgRNA sequence binds the Tet-repressor protein in the absence of doxycycline, blocking the progress of transcription from the U6 promoter. Upon addition, doxycycline binds to the Tet-repressor protein, inhibiting its ability to bind the Tet-O. Without the Tet-repressor barring the way, transcription of the sgRNA from the U6 promoter can proceed. **(B)** A floxed poly-T cassette ('loxP-stop-loxP') is placed past the targeting sequence (blue) and left part of the scaffolding (sgRNA-L), interrupting the transcription of the full sgRNA, rendering a truncated, non-functional sgRNA lacking the right part of the scaffolding (sgRNA-R). Upon delivery/activation, Cre recombinase removes the loxP-stop-loxP site, which enables the transcription of a full uninterrupted functional sgRNA sequence.

While the choice of options for effective Cas9 regulation is wide, a further alternative to render our CRISPR/Cas9 system inducible would be to regulate the expression of the sgRNAs (Figure 7.5), rather than the Cas9 enzyme. Previously, this was achieved by either a doxycycline-inducible system (Aubrey et al., 2015) or a Cre/loxP system (Chylinski et al., 2019). The system based on doxycycline induction (Figure 7.5A) employed a Tet-operator site placed between the sgRNA and its promoter. The binding of the Tet-repressor to the Tet-operator in the absence of doxycycline blocked the transcription of the sgRNA. Added doxycycline would then bind to the Tet-repressor, precipitating its dissociation from the Tet-operator, and allowing the transcription of the sgRNA (Aubrey et al., 2015). There were several systems described based on the Cre/loxP system (Chylinski et al., 2019). Amongst these, the system that would be most applicable to our CRISPR/Cas9-mediated mutagenesis of the *TAL1* enhancer would be the Cre/loxP activation system featuring a floxed poly-A signal inserted into the sgRNA scaffolding (Figure 7.5B). The poly-A signal could be removed by the action of the Cre recombinase, rendering the sgRNA scaffolding fully transcribable and the sgRNA active (Chylinski et al., 2019). Both the doxycycline-inducible sgRNA system and the Cre/loxP-based sgRNA regulatory system were shown to be highly efficient and tightly regulated (Aubrey et al., 2015; Chylinski et al., 2019).

Finally, to increase the resolution of our CRISPR-based saturating mutagenesis of the *TAL1* enhancer, we could employ multiple Cas variants in our screen (Canver et al., 2017), not just the *S. pyogenes* Cas9, which is the most commonly used Cas9 species and also the one stably integrated in our Jurkat-Tet3G-TRE3G-FLAG-TAL1-Cas9 cell line. Because different Cas nucleases have different PAM recognition sites, the gaps between the editing sites in a saturating mutagenesis shorten with the concomitant use of multiple Cas nucleases, providing a higher resolution (Canver et al., 2017). For example, our *S. pyogenes* Cas9 (PAM = NGG) could be complemented by *S. pyogenes* VQR variant Cas9 (PAM = NGA) (Canver et al., 2017), Cas12a from *Acidaminococcus* sp. *BV3L6* (PAM = TTTV, where V = A, C or G) (Gier et al., 2020) or one of the enhanced versions of Cas12a with a significantly broadened range of PAMs (Kleinstiver et al., 2019).

7.5 Conclusion

In this doctoral work, I have set forth the set of experiments and designs that will ultimately lead to a CRISPR/Cas9-mediated saturating mutagenesis of the transcribed portion of the *TAL1* super-enhancer in Jurkat cells, with the view of identifying the molecular determinants of *TAL1*-eRNA functionality (Chapter 1.6).

The understanding of the importance of eRNA transcripts in transcriptional regulation and its implications for health and disease has been growing steadily over the last decade (Chapter 1.3), but to our knowledge, no studies to date have ventured to probe the ways in which eRNAs exert their functions. Here, we hope to pinpoint the functionally important regions in *TAL1*-eRNA structure by inflicting small mutations to the underlying DNA sequence of the transcribed portion of the *TAL1* enhancer, then assessing the changes in the functionality of the enhancer. The *TAL1* enhancer is a powerful regulatory feature in Jurkat cells, responsible for the expression of >95% of all Jurkat-generated *TAL1* (Mansour et al., 2014). *TAL1* drives the fast proliferation of the Jurkat cell line, and a drop in *TAL1* expression ties in with a decrease in the growth rate of the cells (Palii et al., 2011; Palomero et al., 2006; Sanda et al., 2012). *TAL1*-eRNA transcripts have been shown to play a role in the expression of *TAL1* (Chapter 1.6), and therefore we argued that any mutations leading to a functional defect in *eTAL1s* would lead to a drop in proliferation in the (genetically modified) Jurkat cell line.

To perform our screen, I have established and validated a cell line that stably expresses the *S. pyogenes* Cas9, as well as a doxycycline-inducible FLAG-*TAL1* (Jurkat-Tet3G-TRE3G-FLAG-*TAL1*-Cas9; Chapters 3 – 5). The latter feature had been intended for our original design where a control group (aiming to provide an equal sgRNA representation) would be rescued with the addition of doxycycline, but this design had been abandoned because of the unexpected negative effects of doxycycline on the genetically modified Jurkat cell line. Instead, we have chosen to proceed with an alternative strategy where a control group, representing an equal sgRNA distribution, would be collected before or at the very beginning of the screen (Figure 7.1). The newly established and validated cell line, Jurkat-Tet3G-TRE3G-FLAG-*TAL1*-Cas9, is fit for the purpose.

We integrated several cutting-edge approaches into our *TAL1* enhancer CRISPR/Cas9 screen design that will increase the precision, accuracy and power of our screen, as well as decrease potential background effects (Chapter 6). Amongst these are an improved sgRNA scaffolding design and the use of internal barcodes (iBARs). The combined effects of these adjustments will provide us with a better chance at uncovering the relatively small effects that we are expecting to be caused by the *TAL1*-eRNA mutations.

The CRISPR/Cas9-mediated mutagenesis of the *TAL1* enhancer was conceived to contribute the first set of results towards the goal of uncovering the underlying mechanisms of eRNA functionality. The results of this experimental work, which will hopefully highlight the functionally important regions within *TAL1*-eRNA transcripts, will form a basis for further study of the *eTAL1s*, e.g. in vitro functional assays, such as the histone acetyltransferase (HAT) assay, or structural assays, such as the selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE). Furthermore, the methodology, while not

original in conception, is original in application, and if this screen proves to be successful, the strategy could be adopted for the study of other eRNAs in future endeavours to further illuminate the molecular determinants and underlying mechanics of eRNA action.

Overall, this study hopes to have laid a solid basis for a contribution to the pool of eRNA knowledge, by combining a number of recently developed strategies, and leading a previously unattempted foray into the field of underlying mechanisms of eRNA action.

Bibliography

- Adams, C.C., and Workman, J.L. (1995). Binding of disparate transcriptional activators to nucleosomal DNA is inherently cooperative. *Mol. Cell. Biol.* *15*, 1405–1421. .
- Agger, K., Cloos, P.A.C., Christensen, J., Pasini, D., Rose, S., Rappsilber, J., Issaeva, I., Canaani, E., Salcini, A.E., and Helin, K. (2007). UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. *Nature* *449*, 731–734. .
- Ahler, E., Sullivan, W.J., Cass, A., Braas, D., York, A.G., Bensinger, S.J., Graeber, T.G., and Christofk, H.R. (2013). Doxycycline alters metabolism and proliferation of human cell lines. *PLoS One* *8*, e64561. .
- Almada, A.E., Wu, X., Kriz, A.J., Burge, C.B., and Sharp, P.A. (2013). Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature* *499*, 360–363. .
- Alvarez-Dominguez, J.R., Knoll, M., Gromatzky, A.A., and Lodish, H.F. (2017). The Super-Enhancer-Derived alncRNA-EC7/Bloodlinc Potentiates Red Blood Cell Development in trans. *Cell Rep.* *19*, 2503–2514. .
- Amano, T., Sagai, T., Tanabe, H., Mizushima, Y., Nakazawa, H., and Shiroishi, T. (2009). Chromosomal dynamics at the *Shh* locus: limb bud-specific differential regulation of competence and active transcription. *Dev. Cell* *16*, 47–57. .
- Andersen, P.R., Domanski, M., Kristiansen, M.S., Storvall, H., Ntini, E., Verheggen, C., Schein, A., Bunkenborg, J., Poser, I., Hallais, M., et al. (2013). The human cap-binding complex is functionally connected to the nuclear RNA exosome. *Nat. Struct. Mol. Biol.* *20*, 1367–1376. .
- Andersson, R., and Sandelin, A. (2020). Determinants of enhancer and promoter activities of regulatory elements. *Nat. Rev. Genet.* *21*, 71–87. .
- Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., et al. (2014a). An atlas of active enhancers across human cell types and tissues. *Nature* *507*, 455–461. .
- Andersson, R., Andersen, P.R., Valen, E., Core, L.J., Bornholdt, J., Boyd, M., Heick Jensen, T., and Sandelin, A. (2014b). Nuclear stability and transcriptional directionality separate functionally distinct RNA species. *Nat. Commun.* *5*, 5336. .
- Aparicio-Prat, E., Arnan, C., Sala, I., Bosch, N., Guigó, R., and Johnson, R. (2015). DECKO: Single-oligo, dual-CRISPR deletion of genomic elements including long non-coding RNAs. *BMC Genomics* *16*, 846. .
- Arner, E., Daub, C.O., Vitting-Seerup, K., Andersson, R., Lilje, B., Drabløs, F., Lennartsson, A., Rönnerblad, M., Hrydziusko, O., Vitezic, M., et al. (2015). Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. *Science* *347*, 1010–1014. .
- Arnold, C.D., Gerlach, D., Stelzer, C., Boryń, Ł.M., Rath, M., and Stark, A. (2013). Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* *339*, 1074–1077. .
- Arnold, P.R., Wells, A.D., and Li, X.C. (2019). Diversity and Emerging Roles of Enhancer RNA in Regulation of Gene Expression and Cell Fate. *Front Cell Dev Biol* *7*, 377. .

- Aubrey, B.J., Kelly, G.L., Kueh, A.J., Brennan, M.S., O'Connor, L., Milla, L., Wilcox, S., Tai, L., Strasser, A., and Herold, M.J. (2015). An inducible lentiviral guide RNA platform enables the identification of tumor-essential genes and tumor-promoting mutations in vivo. *Cell Rep.* **10**, 1422–1432. .
- Austena, L.M.I., Barozzi, I., Simonatto, M., Masella, S., Della Chiara, G., Ghisletti, S., Curina, A., de Wit, E., Bouwman, B.A.M., de Pretis, S., et al. (2015). Transcription of Mammalian cis-Regulatory Elements Is Restrained by Actively Enforced Early Termination. *Mol. Cell* **60**, 460–474. .
- Aznauryan, E., Yermanos, A., Kinzina, E., Kapetanovic, E., Milanova, D., Church, G.M., and Reddy, S.T. (2021). Discovery and validation of novel human genomic safe harbor sites for gene and cell therapies.
- Ball, M.P., Li, J.B., Gao, Y., Lee, J.-H., LeProust, E.M., Park, I.-H., Xie, B., Daley, G.Q., and Church, G.M. (2009). Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat. Biotechnol.* **27**, 361–368. .
- Banerji, J., Rusconi, S., and Schaffner, W. (1981). Expression of a β -Globin Gene Is Enhanced by Remote SV40 DNA Sequences. *Cell* **27**, 299–308. .
- Banerji, J., Olson, L., and Schaffner, W. (1983). A Lymphocyte-Specific Cellular Enhancer Is Located Downstream of the Joinings Region in Immunoglobulin Heavy Chain Genes. *Cell* **33**, 729–740. .
- Bannikov, A.V., and Lavrov, A.V. (2017). CRISPR/CAS9, the king of genome editing tools. *Mol. Biol.* **51**, 514–525. .
- Bardelli, V., Arniani, S., Pierini, V., Di Giacomo, D., Pierini, T., Gorello, P., Mecucci, C., and La Starza, R. (2021). T-Cell Acute Lymphoblastic Leukemia: Biomarkers and Their Clinical Usefulness. *Genes* **12**. <https://doi.org/10.3390/genes12081118>.
- Baron, U., Gossen, M., and Bujard, H. (1997). Tetracycline-controlled transcription in eukaryotes: novel transactivators with graded transactivation potential. *Nucleic Acids Res.* **25**, 2723–2729. .
- Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712. .
- Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837. .
- Bayer, M., Kantor, B., Cockrell, A., Ma, H., Zeithaml, B., Li, X., McCown, T., and Kafri, T. (2008). A large U3 deletion causes increased in vivo expression from a nonintegrating lentiviral vector. *Mol. Ther.* **16**, 1968–1976. .
- Begley, C.G., Aplan, P.D., Davey, M.P., Nakahara, K., Tchorz, K., Kurtzberg, J., Herschfield, M.S., Haynes, B.F., Cohen, D.I., Waldmann, T.A., et al. (1989). Chromosomal Translocation in a Human Leukemic Stem-Cell Line Disrupts the T-Cell Antigen Receptor δ -Chain Diversity Region and Results in a Previously Unreported Fusion Transcript. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 2031–2035. .
- Bender, M.A., Ragoczy, T., Lee, J., Byron, R., Telling, A., Dean, A., and Groudine, M. (2012). The hypersensitive sites of the murine β -globin locus control region act independently to affect nuclear localization and transcriptional elongation. *Blood* **119**, 3820–3827. .
- Bernard, O., Guglielmi, P., Jonveaux, P., Cherif, D., Gisselbrecht, S., Mauchauffe, M., Berger, R., Larsen, C.-J., and Mathieu-Mahul, D. (1990). Two distinct mechanisms for the SCL gene activation in the t(1;14) translocation of T-cell leukemias. *Genes Chromosomes Cancer* **1**, 194–208. .

- Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315–326. .
- Bikard, D., Jiang, W., Samai, P., Hochschild, A., Zhang, F., and Marraffini, L.A. (2013). Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res.* **41**, 7429–7437. .
- Blackwood, E.M., and Kadonaga, J.T. (1998). Going the distance: a current view of enhancer action. *Science* **281**, 60–63. .
- Blank-Giwojna, A., Postepska-Igielska, A., and Grummt, I. (2019). lncRNA KHPS1 Activates a Poised Enhancer by Triplex-Dependent Recruitment of Epigenomic Regulators. *Cell Rep.* **26**, 2904-2915.e4. .
- Bock, C., Datlinger, P., Chardon, F., Coelho, M.A., Dong, M.B., Lawson, K.A., Lu, T., Maroc, L., Norman, T.M., Song, B., et al. (2022). High-content CRISPR screening. *Nat Rev Methods Primers* **2**. <https://doi.org/10.1038/s43586-021-00093-4>.
- Bose, D.A., Donahue, G., Reinberg, D., Shiekhata, R., Bonasio, R., and Berger, S.L. (2017). RNA Binding to CBP Stimulates Histone Acetylation and Transcription. *Cell* **168**, 135-149.e22. .
- Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K., et al. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**, 349–353. .
- Boyes, J., and Felsenfeld, G. (1996). Tissue-specific factors additively increase the probability of the all-or-none formation of a hypersensitive site. *EMBO J.* **15**, 2496–2507. .
- Briner, A.E., Donohoue, P.D., Goma, A.A., Selle, K., Slorach, E.M., Nye, C.H., Haurwitz, R.E., Beisel, C.L., May, A.P., and Barrangou, R. (2014). Guide RNA functional modules direct Cas9 activity and orthogonality. *Mol. Cell* **56**, 333–339. .
- Brown, L., Cheng, J.T., Chen, Q., Siciliano, M.J., Crist, W., Buchanan, G., and Baer, R. (1990). Site-specific recombination of the tal-1 gene is a common occurrence in human T cell leukemia. *EMBO J.* **9**, 3343–3351. .
- Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr. Protoc. Mol. Biol.* **109**, 21.29.1-21.29.9. .
- Bulger, M., and Groudine, M. (1999). Looping versus linking: toward a model for long-distance gene activation. *Genes Dev.* **13**, 2465–2477. .
- Buschmann, T., and Bystrykh, L.V. (2013). Levenshtein error-correcting barcodes for multiplexed DNA sequencing. *BMC Bioinformatics* **14**, 272. .
- Buskirk, A.R., Ong, Y.-C., Gartner, Z.J., and Liu, D.R. (2004). Directed evolution of ligand dependence: Small-molecule-activated protein splicing. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 10505–10510. .
- Cai, Z., Cao, C., Ji, L., Ye, R., Wang, D., Xia, C., Wang, S., Du, Z., Hu, N., Yu, X., et al. (2020). RIC-seq for global in situ profiling of RNA-RNA spatial interactions. *Nature* **582**, 432–437. .
- Cajigas, I., Chakraborty, A., Swyter, K.R., Luo, H., Bastidas, M., Nigro, M., Morris, E.R., Chen, S., VanGompel, M.J.W., Leib, D., et al. (2018). The Evf2 Ultraconserved Enhancer lncRNA Functionally and Spatially Organizes Megabase Distant Genes in the Developing Forebrain. *Mol. Cell* **71**, 956-972.e9. .

Canver, M.C., Bauer, D.E., Dass, A., Yien, Y.Y., Chung, J., Masuda, T., Maeda, T., Paw, B.H., and Orkin, S.H. (2014). Characterization of Genomic Deletion Efficiency Mediated by Clustered Regularly Interspaced Palindromic Repeats (CRISPR)/Cas9 Nuclease System in Mammalian Cells. *J. Biol. Chem.* **289**, 21312–21324. .

Canver, M.C., Smith, E.C., Sher, F., Pinello, L., Sanjana, N.E., Shalem, O., Chen, D.D., Schupp, P.G., Vinjamur, D.S., Garcia, S.P., et al. (2015). BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature* **527**, 192–197. .

Canver, M.C., Lessard, S., Pinello, L., Wu, Y., Ilboudo, Y., Stern, E.N., Needleman, A.J., Galactéros, F., Brugnara, C., Kutlar, A., et al. (2017). Variant-aware saturating mutagenesis using multiple Cas9 nucleases identifies regulatory elements at trait-associated loci. *Nat. Genet.* **49**, 625–634. .

Canver, M.C., Haeussler, M., Bauer, D.E., Orkin, S.H., Sanjana, N.E., Shalem, O., Yuan, G.-C., Zhang, F., Concordet, J.-P., and Pinello, L. (2018). Integrated design, execution, and analysis of arrayed and pooled CRISPR genome-editing experiments. *Nat. Protoc.* **13**, 946–986. .

Canver, M.C., Tripathi, P., Bullen, M.J., Olshansky, M., Kumar, Y., Wong, L.H., Turner, S.J., Lessard, S., Pinello, L., Orkin, S.H., et al. (2020). A saturating mutagenesis CRISPR-Cas9 mediated functional genomic screen identifies cis- and trans- regulatory elements of Oct4 in murine ESCs.

Cao, J., Wu, L., Zhang, S.-M., Lu, M., Cheung, W.K.C., Cai, W., Gale, M., Xu, Q., and Yan, Q. (2016). An easy and efficient inducible CRISPR/Cas9 platform with improved specificity for multiple gene targeting. *Nucleic Acids Res.* **44**, e149. .

Carullo, N.V.N., and Day, J.J. (2019). Genomic Enhancers in Brain Health and Disease. *Genes* **10**. <https://doi.org/10.3390/genes10010043>.

Carullo, N.V.N., Phillips, R.A., Iii, Simon, R.C., Soto, S.A.R., Hinds, J.E., Salisbury, A.J., Revanna, J.S., Bunner, K.D., Ianov, L., Sultan, F.A., et al. (2020). Enhancer RNAs predict enhancer-gene regulatory links and are critical for enhancer function in neuronal systems. *Nucleic Acids Res.* **48**, 9550–9570. .

Cech, T.R., and Steitz, J.A. (2014). The noncoding RNA revolution-trashing old rules to forge new ones. *Cell* **157**, 77–94. .

Chambeyron, S., and Bickmore, W.A. (2004). Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. *Genes Dev.* **18**, 1119–1130. .

Che, W., Ye, S., Cai, A., Cui, X., and Sun, Y. (2020). CRISPR-Cas13a Targeting the Enhancer RNA-SMAD7e Inhibits Bladder Cancer Development Both in vitro and in vivo. *Front Mol Biosci* **7**, 607740. .

Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.-W., Park, J., Blackburn, E.H., Weissman, J.S., Qi, L.S., et al. (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155**, 1479–1491. .

Chen, L.-F., Lin, Y.T., Gallegos, D.A., Hazlett, M.F., Gómez-Schiavon, M., Yang, M.G., Kalmeta, B., Zhou, A.S., Holtzman, L., Gersbach, C.A., et al. (2019). Enhancer Histone Acetylation Modulates Transcriptional Bursting Dynamics of Neuronal Activity-Inducible Genes. *Cell Rep.* **26**, 1174-1188.e5. .

Chen, Q., Cheng, J.T., Tasi, L.H., Schneider, N., Buchanan, G., Carroll, A., Crist, W., Ozanne, B., Siciliano, M.J., and Baer, R. (1990). The tal gene undergoes chromosome translocation in T cell leukemia and potentially encodes a helix-loop-helix protein. *EMBO J.* **9**, 415–424. .

- Cheng, A.W., Wang, H., Yang, H., Shi, L., Katz, Y., Theunissen, T.W., Rangarajan, S., Shivalila, C.S., Dadon, D.B., and Jaenisch, R. (2013). Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell Res.* 23, 1163–1171. .
- Cheng, C., Alexander, R., Min, R., Leng, J., Yip, K.Y., Rozowsky, J., Yan, K.-K., Dong, X., Djebali, S., Ruan, Y., et al. (2012). Understanding transcriptional regulation by integrative analysis of transcription factor binding data. *Genome Res.* 22, 1658–1667. .
- Cheng, J., Kapranov, P., Drenkow, J., Dike, S., Brubaker, S., Patel, S., Long, J., Stern, D., Tammanna, H., Helt, G., et al. (2005). Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science* 308, 1149–1154. .
- Cheng, J.-H., Pan, D.Z.-C., Tsai, Z.T.-Y., and Tsai, H.-K. (2015). Genome-wide analysis of enhancer RNA in gene regulation across 12 mouse tissues. *Sci. Rep.* 5, 12648. .
- Chi, S., Weiss, A., and Wang, H. (2016). A CRISPR-Based Toolbox for Studying T Cell Signal Transduction. *Biomed Res. Int.* 2016, 5052369. .
- Cho, H., Orphanides, G., Sun, X., Yang, X.J., Ogryzko, V., Lees, E., Nakatani, Y., and Reinberg, D. (1998). A human RNA polymerase II complex containing factors that modify chromatin structure. *Mol. Cell. Biol.* 18, 5355–5363. .
- Cho, S.W., Xu, J., Sun, R., Mumbach, M.R., Carter, A.C., Chen, Y.G., Yost, K.E., Kim, J., He, J., Nevins, S.A., et al. (2018). Promoter of lncRNA Gene PVT1 Is a Tumor-Suppressor DNA Boundary Element. *Cell* 173, 1398-1412.e22. .
- Cho, W.-K., Jayanth, N., English, B.P., Inoue, T., Andrews, J.O., Conway, W., Grimm, J.B., Spille, J.-H., Lavis, L.D., Lionnet, T., et al. (2016). RNA Polymerase II cluster dynamics predict mRNA output in living cells. *Elife* 5. <https://doi.org/10.7554/eLife.13617>.
- Christian, M., Cermak, T., Doyle, E.L., Schmidt, C., Zhang, F., Hummel, A., Bogdanove, A.J., and Voytas, D.F. (2010). Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* 186, 757–761. .
- Chylinski, K., Hubmann, M., Hanna, R.E., Yanchus, C., Michlits, G., Uijttewaal, E.C.H., Doench, J., Schramek, D., and Elling, U. (2019). CRISPR-Switch regulates sgRNA activity by Cre recombination for sequential editing of two loci. *Nat. Commun.* 10, 5454. .
- Cinghu, S., Yang, P., Kosak, J.P., Conway, A.E., Kumar, D., Oldfield, A.J., Adelman, K., and Jothi, R. (2017). Intragenic Enhancers Attenuate Host Gene Expression. *Mol. Cell* 68, 104-117.e6. .
- Cisse, I.I., Izeddin, I., Causse, S.Z., Boudarene, L., Senecal, A., Muresan, L., Dugast-Darzacq, C., Hajj, B., Dahan, M., and Darzacq, X. (2013). Real-time dynamics of RNA polymerase II clustering in live human cells. *Science* 341, 664–667. .
- Cockrell, A.S., and Kafri, T. (2007). Gene delivery by lentivirus vectors. *Mol. Biotechnol.* 36, 184–204. .
- Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819–823. .
- Coppola, C.J., C Ramaker, R., and Mendenhall, E.M. (2016). Identification and function of enhancers in the human genome. *Hum. Mol. Genet.* 25, R190–R197. .

Core, L., and Adelman, K. (2019). Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. *Genes Dev.* **33**, 960–982. .

Core, L.J., Martins, A.L., Danko, C.G., Waters, C.T., Siepel, A., and Lis, J.T. (2014). Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat. Genet.* **46**, 1311–1320. .

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 21931–21936. .

Cross, B.C.S., Lawo, S., Archer, C.R., Hunt, J.R., Yarker, J.L., Riccombeni, A., Little, A.S., McCarthy, N.J., and Moore, J.D. (2016). Increasing the performance of pooled CRISPR-Cas9 drop-out screening. *Sci. Rep.* **6**, 31782. .

Danko, C.G., Hyland, S.L., Core, L.J., Martins, A.L., Waters, C.T., Lee, H.W., Cheung, V.G., Kraus, W.L., Lis, J.T., and Siepel, A. (2015). Identification of active transcriptional regulatory elements from GRO-seq data. *Nat. Methods* **12**, 433–438. .

Davis, K.M., Pattanayak, V., Thompson, D.B., Zuris, J.A., and Liu, D.R. (2015). Small molecule-triggered Cas9 protein with improved genome-editing specificity. *Nat. Chem. Biol.* **11**, 316–318. .

De Santa, F., Barozzi, I., Mietton, F., Ghisletti, S., Polletti, S., Tusi, B.K., Muller, H., Ragoussis, J., Wei, C.-L., and Natoli, G. (2010). A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol.* **8**, e1000384. .

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. *Science* **295**, 1306–1311. .

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., and Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602–607. .

Descostes, N., Heidemann, M., Spinelli, L., Schüller, R., Maqbool, M.A., Fenouil, R., Koch, F., Innocenti, C., Gut, M., Gut, I., et al. (2014). Tyrosine phosphorylation of RNA polymerase II CTD is associated with antisense promoter transcription and active enhancers in mammalian cells. *Elife* **3**, e02105. .

Devesa-Guerra, I., Morales-Ruiz, T., Pérez-Roldán, J., Parrilla-Doblas, J.T., Dorado-León, M., García-Ortiz, M.V., Ariza, R.R., and Roldán-Arjona, T. (2020). DNA Methylation Editing by CRISPR-guided Excision of 5-Methylcytosine. *J. Mol. Biol.* **432**, 2204–2216. .

Diao, Y., Li, B., Meng, Z., Jung, I., Lee, A.Y., Dixon, J., Maliskova, L., Guan, K.-L., Shen, Y., and Ren, B. (2016). A new class of temporarily phenotypic enhancers identified by CRISPR/Cas9-mediated genetic screening. *Genome Res.* **26**, 397–405. .

Diao, Y., Fang, R., Li, B., Meng, Z., Yu, J., Qiu, Y., Lin, K.C., Huang, H., Liu, T., Marina, R.J., et al. (2017). A tiling-deletion-based genetic screen for cis-regulatory element identification in mammalian cells. *Nat. Methods* **14**, 629–635. .

Dijk, S.N., Protasoni, M., Elpidorou, M., Kroon, A.M., and Taanman, J.-W. (2020). Mitochondria as target to inhibit proliferation and induce apoptosis of cancer cells: the effects of doxycycline and gemcitabine. *Sci. Rep.* **10**, 4363. .

- Dixit, A., Parnas, O., Li, B., Chen, J., Fulco, C.P., Jerby-Arnon, L., Marjanovic, N.D., Dionne, D., Burks, T., Raychowdhury, R., et al. (2016). Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* 167, 1853–1866.e17. .
- Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376–380. .
- Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., et al. (2012). Landscape of transcription in human cells. *Nature* 489, 101–108. .
- Doench, J.G., Hartenian, E., Graham, D.B., Tothova, Z., Hegde, M., Smith, I., Sullender, M., Ebert, B.L., Xavier, R.J., and Root, D.E. (2014). Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat. Biotechnol.* 32, 1262–1267. .
- Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* 34, 184–191. .
- Dong, W., and Kantor, B. (2021). Lentiviral Vectors for Delivery of Gene-Editing Systems Based on CRISPR/Cas: Current State and Perspectives. *Viruses* 13. <https://doi.org/10.3390/v13071288>.
- Dostie, J., Richmond, T.A., Arnaout, R.A., Selzer, R.R., Lee, W.L., Honan, T.A., Rubio, E.D., Krumm, A., Lamb, J., Nusbaum, C., et al. (2006). Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res.* 16, 1299–1309. .
- Downen, J.M., Fan, Z.P., Hnisz, D., Ren, G., Abraham, B.J., Zhang, L.N., Weintraub, A.S., Schuijers, J., Lee, T.I., Zhao, K., et al. (2014). Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* 159, 374–387. .
- Dull, T., Zufferey, R., Kelly, M., Mandel, R.J., Nguyen, M., Trono, D., and Naldini, L. (1998). A third-generation lentivirus vector with a conditional packaging system. *J. Virol.* 72, 8463–8471. .
- Dunn, T.M., Hahn, S., Ogden, S., and Schleif, R.F. (1984). An operator at -280 base pairs that is required for repression of araBAD operon promoter: addition of DNA helical turns between the operator and promoter cyclically hinders repression. *Proc. Natl. Acad. Sci. U. S. A.* 81, 5017–5020. .
- Elcheva, I., Brok-Volchanskaya, V., Kumar, A., Liu, P., Lee, J.-H., Tong, L., Vodyanik, M., Swanson, S., Stewart, R., Kyba, M., et al. (2014). Direct induction of haematoendothelial programs in human pluripotent stem cells by transcriptional regulators. *Nat. Commun.* 5, 4372. .
- ENCODE Project Consortium (2011). A user’s guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol.* 9, e1001046. .
- ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74. .
- Ernst, J., and Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* 9, 215–216. .
- Finger, L.R., Kagan, J., Christopher, G., Kurtzberg, J., Hershfield, M.S., Nowell, P.C., and Croce, C.M. (1989). Involvement of the TCL5 gene on human chromosome 1 in T-cell leukemia and melanoma. *Proc. Natl. Acad. Sci. U. S. A.* 86, 5039–5043. .

- Finkelshtein, D., Werman, A., Novick, D., Barak, S., and Rubinstein, M. (2013). LDL receptor and its family members serve as the cellular receptors for vesicular stomatitis virus. *Proc. Natl. Acad. Sci. U. S. A.* *110*, 7306–7311. .
- Forslund, O., Antonsson, A., Higgins, G., Ly, H., Delius, H., Hunziker, A., and de Villiers, E.M. (2003). Nucleotide sequence and phylogenetic classification of candidate human papilloma virus type 92. *Virology* *312*, 255–260. .
- Frohman, M.A., Dush, M.K., and Martin, G.R. (1988). Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci. U. S. A.* *85*, 8998–9002. .
- Fu, Y., Sinha, M., Peterson, C.L., and Weng, Z. (2008). The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome. *PLoS Genet.* *4*, e1000138. .
- Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., and Mirny, L.A. (2016). Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep.* *15*, 2038–2049. .
- Fudenberg, G., Abdennur, N., Imakaev, M., Goloborodko, A., and Mirny, L.A. (2018). Emerging Evidence of Chromosome Folding by Loop Extrusion.
- Fulco, C.P., Munschauer, M., Anyoha, R., Munson, G., Grossman, S.R., Perez, E.M., Kane, M., Cleary, B., Lander, E.S., and Engreitz, J.M. (2016). Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science* *354*, 769–773. .
- Fulco, C.P., Nasser, J., Jones, T.R., Munson, G., Bergman, D.T., Subramanian, V., Grossman, S.R., Anyoha, R., Doughty, B.R., Patwardhan, T.A., et al. (2019). Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nat. Genet.* *51*, 1664–1669. .
- Furlong, E.E.M., and Levine, M. (2018). Developmental enhancers and chromosome topology. *Science* *361*, 1341–1345. .
- Gao, X., Tsang, J.C.H., Gaba, F., Wu, D., Lu, L., and Liu, P. (2014). Comparison of TALE designer transcription factors and the CRISPR/dCas9 in regulation of gene expression by targeting enhancers. *Nucleic Acids Res.* *42*, e155. .
- Gasperini, M., Findlay, G.M., McKenna, A., Milbank, J.H., Lee, C., Zhang, M.D., Cusanovich, D.A., and Shendure, J. (2017). CRISPR/Cas9-Mediated Scanning for Regulatory Elements Required for HPRT1 Expression via Thousands of Large, Programmed Genomic Deletions. *Am. J. Hum. Genet.* *101*, 192–205. .
- Gasperini, M., Hill, A.J., McFaline-Figueroa, J.L., Martin, B., Kim, S., Zhang, M.D., Jackson, D., Leith, A., Schreiber, J., Noble, W.S., et al. (2019). A Genome-wide Framework for Mapping Gene Regulation via Cellular Genetic Screens. *Cell* *176*, 377–390.e19. .
- Gasperini, M., Tome, J.M., and Shendure, J. (2020). Towards a comprehensive catalogue of validated and target-linked human enhancers. *Nat. Rev. Genet.* *21*, 292–310. .
- Ghavi-Helm, Y., Klein, F.A., Pakozdi, T., Ciglar, L., Noordermeer, D., Huber, W., and Furlong, E.E.M. (2014). Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* *512*, 96–100. .

Ghisletti, S., Barozzi, I., Mietton, F., Polletti, S., De Santa, F., Venturini, E., Gregory, L., Lonie, L., Chew, A., Wei, C.-L., et al. (2010). Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages. *Immunity* 32, 317–328. .

Gier, R.A., Budinich, K.A., Evitt, N.H., Cao, Z., Freilich, E.S., Chen, Q., Qi, J., Lan, Y., Kohli, R.M., and Shi, J. (2020). High-performance CRISPR-Cas12a genome editing for combinatorial genetic screening. *Nat. Commun.* 11, 3455. .

Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A., et al. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* 154, 442–451. .

Gilbert, L.A., Horlbeck, M.A., Adamson, B., Villalta, J.E., Chen, Y., Whitehead, E.H., Guimaraes, C., Panning, B., Ploegh, H.L., Bassik, M.C., et al. (2014). Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell* 159, 647–661. .

Gillies, S.D., and Morrison, S.L. (1983). A Tissue-specific Transcription Enhancer Element Is Located in the Major Intron of a Rearranged Immunoglobulin Heavy Chain Gene. *Cell* 33, 717–728. .

Ginley-Hidinger, M., Carleton, J.B., Rodriguez, A.C., Berrett, K.C., and Gertz, J. (2019). Sufficiency analysis of estrogen responsive enhancers using synthetic activators. *Life Sci Alliance* 2. <https://doi.org/10.26508/lsa.201900497>.

González, F., Zhu, Z., Shi, Z.-D., Lelli, K., Verma, N., Li, Q.V., and Huangfu, D. (2014). An iCRISPR platform for rapid, multiplexable, and inducible genome editing in human pluripotent stem cells. *Cell Stem Cell* 15, 215–226. .

Gossen, M., and Bujard, H. (1992). Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc. Natl. Acad. Sci. U. S. A.* 89, 5547–5551. .

Gossen, M., Freundlieb, S., Bender, G., Müller, G., Hillen, W., and Bujard, H. (1995). Transcriptional activation by tetracyclines in mammalian cells. *Science* 268, 1766–1769. .

Gouvarchin Ghaleh, H.E., Bolandian, M., Dorostkar, R., Jafari, A., and Pour, M.F. (2020). Concise review on optimized methods in production and transduction of lentiviral vectors in order to facilitate immunotherapy and gene therapy. *Biomed. Pharmacother.* 128, 110276. .

Greaves, R., and O’Hare, P. (1989). Separation of requirements for protein-DNA complex assembly from those for functional activity in the herpes simplex virus regulatory protein Vmw65. *Journal of Virology* 63, 1641–1650. .

Groner, A.C., Meylan, S., Ciuffi, A., Zangger, N., Ambrosini, G., Déneraud, N., Bucher, P., and Trono, D. (2010). KRAB-zinc finger proteins and KAP1 can mediate long-range transcriptional repression through heterochromatin spreading. *PLoS Genet.* 6, e1000869. .

GTEx Consortium (2017). Genetic effects on gene expression across human tissues. *Nature* 550, 204–213. .

Gualdi, R., Bossard, P., Zheng, M., Hamada, Y., Coleman, J.R., and Zaret, K.S. (1996). Hepatic specification of the gut endoderm in vitro: cell signaling and transcriptional control. *Genes Dev.* 10, 1670–1682. .

Hans, S., Zöller, D., Hammer, J., Stucke, J., Spieß, S., Kesavan, G., Kroehne, V., Eguiguren, J.S., Ezhkova, D., Petzold, A., et al. (2021). Cre-Controlled CRISPR mutagenesis provides fast and easy conditional gene inactivation in zebrafish. *Nat. Commun.* **12**, 1125. .

Hastie, E., Cataldi, M., Marriott, I., and Grdzelishvili, V.Z. (2013). Understanding and altering cell tropism of vesicular stomatitis virus. *Virus Res.* **176**, 16–32. .

Hatzis, P., and Talianidis, I. (2002). Dynamics of enhancer-promoter communication during differentiation-induced gene activation. *Mol. Cell* **10**, 1467–1477. .

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* **39**, 311–318. .

Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., et al. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108–112. .

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589. .

Heinz, S., Romanoski, C.E., Benner, C., and Glass, C.K. (2015). The selection and function of cell type-specific enhancers. *Nat. Rev. Mol. Cell Biol.* **16**, 144–154. .

Henriques, T., Scruggs, B.S., Inouye, M.O., Muse, G.W., Williams, L.H., Burkholder, A.B., Lavender, C.A., Fargo, D.C., and Adelman, K. (2018). Widespread transcriptional pausing and elongation control at enhancers. *Genes Dev.* **32**, 26–41. .

Hilton, I.B., D'Ippolito, A.M., Vockley, C.M., Thakore, P.I., Crawford, G.E., Reddy, T.E., and Gersbach, C.A. (2015). Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat. Biotechnol.* **33**, 510–517. .

Hirai, H., Tani, T., and Kikyo, N. (2010). Structure and functions of powerful transactivators: VP16, MyoD and FoxA. *Int. J. Dev. Biol.* **54**, 1589–1596. .

Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-André, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947. .

Hnisz, D., Day, D.S., and Young, R.A. (2016a). Insulated Neighborhoods: Structural and Functional Units of Mammalian Gene Control. *Cell* **167**, 1188–1200. .

Hnisz, D., Weintraub, A.S., Day, D.S., Valton, A.-L., Bak, R.O., Li, C.H., Goldmann, J., Lajoie, B.R., Fan, Z.P., Sigova, A.A., et al. (2016b). Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* **351**, 1454–1458. .

Hnisz, D., Shrinivas, K., Young, R.A., Chakraborty, A.K., and Sharp, P.A. (2017). A Phase Separation Model for Transcriptional Control. *Cell* **169**, 13–23. .

Ho, S.-M., Hartley, B.J., Flaherty, E., Rajarajan, P., Abdelaal, R., Obiorah, I., Barretto, N., Muhammad, H., Phatnani, H.P., Akbarian, S., et al. (2017). Evaluating Synthetic Activation and Repression of Neuropsychiatric-Related Genes in hiPSC-Derived NPCs, Neurons, and Astrocytes. *Stem Cell Reports* **9**, 615–628. .

- Hoffman, M.M., Buske, O.J., Wang, J., Weng, Z., Bilmes, J.A., and Noble, W.S. (2012). Unsupervised pattern discovery in human chromatin structure through genomic segmentation. *Nat. Methods* 9, 473–476. .
- Hoffman, M.M., Ernst, J., Wilder, S.P., Kundaje, A., Harris, R.S., Libbrecht, M., Giardine, B., Ellenbogen, P.M., Bilmes, J.A., Birney, E., et al. (2013). Integrative annotation of chromatin elements from ENCODE data. *Nucleic Acids Res.* 41, 827–841. .
- Hong, J.-W., Hendrix, D.A., and Levine, M.S. (2008). Shadow enhancers as a source of evolutionary novelty. *Science* 321, 1314. .
- Hsieh, C.-L., Fei, T., Chen, Y., Li, T., Gao, Y., Wang, X., Sun, T., Sweeney, C.J., Lee, G.-S.M., Chen, S., et al. (2014). Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation. *Proc. Natl. Acad. Sci. U. S. A.* 111, 7319–7324. .
- Hsu, H.L., Wadman, I., Tsan, J.T., and Baer, R. (1994a). Positive and negative transcriptional control by the TAL1 helix-loop-helix protein. *Proc. Natl. Acad. Sci. U. S. A.* 91, 5947–5951. .
- Hsu, H.L., Huang, L., Tsan, J.T., Funk, W., Wright, W.E., Hu, J.S., Kingston, R.E., and Baer, R. (1994b). Preferred sequences for DNA recognition by the TAL1 helix-loop-helix proteins. *Mol. Cell. Biol.* 14, 1256–1265. .
- Huang, Y.-H., Su, J., Lei, Y., Brunetti, L., Gundry, M.C., Zhang, X., Jeong, M., Li, W., and Goodell, M.A. (2017). DNA epigenome editing using CRISPR-Cas SunTag-directed DNMT3A. *Genome Biol.* 18, 176. .
- Huang, Z., Liang, N., Goñi, S., Damdimopoulos, A., Wang, C., Ballaire, R., Jager, J., Niskanen, H., Han, H., Jakobsson, T., et al. (2021). The corepressors GPS2 and SMRT control enhancer and silencer remodeling via eRNA transcription during inflammatory activation of macrophages. *Mol. Cell* 81, 953-968.e9. .
- Huber, C.D., Kim, B.Y., and Lohmueller, K.E. (2019). Population genetic models of GERP scores suggest pervasive turnover of constrained sites across mammalian evolution (bioRxiv).
- Hyle, J., Zhang, Y., Wright, S., Xu, B., Shao, Y., Easton, J., Tian, L., Feng, R., Xu, P., and Li, C. (2019). Acute depletion of CTCF directly affects MYC regulation through loss of enhancer-promoter looping. *Nucleic Acids Res.* 47, 6699–6713. .
- Issaeva, I., Zonis, Y., Rozovskaia, T., Orlovsky, K., Croce, C.M., Nakamura, T., Mazo, A., Eisenbach, L., and Canaani, E. (2007). Knockdown of ALR (MLL2) reveals ALR target genes and leads to alterations in cell adhesion and growth. *Mol. Cell. Biol.* 27, 1889–1903. .
- Jagušić, M., Forčić, D., Brgles, M., Kutle, L., Šantak, M., Jergović, M., Kotarski, L., Bendelja, K., and Halassy, B. (2016). Stability of Minimum Essential Medium functionality despite L-glutamine decomposition. *Cytotechnology* 68, 1171–1183. .
- Jang, M.K., Mochizuki, K., Zhou, M., Jeong, H.-S., Brady, J.N., and Ozato, K. (2005). The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. *Mol. Cell* 19, 523–534. .
- Jayavaradhan, R., Pillis, D.M., Goodman, M., Zhang, F., Zhang, Y., Andreassen, P.R., and Malik, P. (2019). CRISPR-Cas9 fusion to dominant-negative 53BP1 enhances HDR and inhibits NHEJ specifically at Cas9 target sites. *Nat. Commun.* 10, 2866. .

- Jiao, W., Chen, Y., Song, H., Li, D., Mei, H., Yang, F., Fang, E., Wang, X., Huang, K., Zheng, L., et al. (2018). HPSE enhancer RNA promotes cancer progression through driving chromatin looping and regulating hnRNPU/p300/EGR1/HPSE axis. *Oncogene* 37, 2728–2745. .
- Jin, C., and Felsenfeld, G. (2007). Nucleosome stability mediated by histone variants H3.3 and H2A.Z. *Genes Dev.* 21, 1519–1529. .
- Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K., and Felsenfeld, G. (2009). H3.3/H2A.Z double variant-containing nucleosomes mark ‘nucleosome-free regions’ of active promoters and other regulatory regions. *Nat. Genet.* 41, 941–945. .
- Jin, F., Li, Y., Dixon, J.R., Selvaraj, S., Ye, Z., Lee, A.Y., Yen, C.-A., Schmitt, A.D., Espinoza, C.A., and Ren, B. (2013). A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* 503, 290–294. .
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821. .
- Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., and Doudna, J. (2013). RNA-programmed genome editing in human cells. *Elife* 2, e00471. .
- Jo, Y.-I., Kim, H., and Ramakrishna, S. (2015). Recent developments and clinical studies utilizing engineered zinc finger nuclease technology. *Cell. Mol. Life Sci.* 72, 3819–3830. .
- Joung, J., Konermann, S., Gootenberg, J.S., Abudayyeh, O.O., Platt, R.J., Brigham, M.D., Sanjana, N.E., and Zhang, F. (2017). Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nat. Protoc.* 12, 828–863. .
- Juven-Gershon, T., Hsu, J.-Y., and Kadonaga, J.T. (2008). Caudal, a key developmental regulator, is a DPE-specific transcriptional factor. *Genes Dev.* 22, 2823–2830. .
- Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., et al. (2010). Mediator and cohesin connect gene expression and chromatin architecture. *Nature* 467, 430–435. .
- Kaida, D., Berg, M.G., Younis, I., Kasim, M., Singh, L.N., Wan, L., and Dreyfuss, G. (2010). U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* 468, 664–668. .
- Kaikkonen, M.U., Spann, N.J., Heinz, S., Romanoski, C.E., Allison, K.A., Stender, J.D., Chun, H.B., Tough, D.F., Prinjha, R.K., Benner, C., et al. (2013). Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. *Mol. Cell* 51, 310–325. .
- Kanno, T., Kanno, Y., LeRoy, G., Campos, E., Sun, H.-W., Brooks, S.R., Vahedi, G., Heightman, T.D., Garcia, B.A., Reinberg, D., et al. (2014). BRD4 assists elongation of both coding and enhancer RNAs by interacting with acetylated histones. *Nat. Struct. Mol. Biol.* 21, 1047–1057. .
- Kantor, B., Bayer, M., Ma, H., Samulski, J., Li, C., McCown, T., and Kafri, T. (2011). Notable reduction in illegitimate integration mediated by a PPT-deleted, nonintegrating lentiviral vector. *Mol. Ther.* 19, 547–556. .
- Kearns, N.A., Pham, H., Tabak, B., Genga, R.M., Silverstein, N.J., Garber, M., and Maehr, R. (2015). Functional annotation of native enhancers with a Cas9-histone demethylase fusion. *Nat. Methods* 12, 401–403. .

- Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. *Genome Res.* **12**, 996–1006. .
- Kieffer-Kwon, K.-R., Tang, Z., Mathe, E., Qian, J., Sung, M.-H., Li, G., Resch, W., Baek, S., Pruett, N., Grøntved, L., et al. (2013). Interactome maps of mouse gene regulatory domains reveal basic principles of transcriptional regulation. *Cell* **155**, 1507–1520. .
- Kim, T.-K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., et al. (2010). Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182–187. .
- Kim, Y.G., Cha, J., and Chandrasegaran, S. (1996). Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 1156–1160. .
- Klann, T.S., Black, J.B., Chellappan, M., Safi, A., Song, L., Hilton, I.B., Crawford, G.E., Reddy, T.E., and Gersbach, C.A. (2017). CRISPR-Cas9 epigenome editing enables high-throughput screening for functional regulatory elements in the human genome. *Nat. Biotechnol.* **35**, 561–568. .
- Kleinstiver, B.P., Sousa, A.A., Walton, R.T., Tak, Y.E., Hsu, J.Y., Clement, K., Welch, M.M., Horng, J.E., Malagon-Lopez, J., Scarfò, I., et al. (2019). Engineered CRISPR-Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. *Nat. Biotechnol.* **37**, 276–282. .
- Koch, F., Fenouil, R., Gut, M., Cauchy, P., Albert, T.K., Zacarias-Cabeza, J., Spicuglia, S., de la Chapelle, A.L., Heidemann, M., Hintermair, C., et al. (2011). Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nat. Struct. Mol. Biol.* **18**, 956–963. .
- Kong, S., Bohl, D., Li, C., and Tuan, D. (1997). Transcription of the HS2 enhancer toward a cis-linked gene is independent of the orientation, position, and distance of the enhancer relative to the gene. *Mol. Cell. Biol.* **17**, 3955–3965. .
- Korkmaz, G., Lopes, R., Ugalde, A.P., Nevedomskaya, E., Han, R., Myacheva, K., Zwart, W., Elkon, R., and Agami, R. (2016). Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. *Nat. Biotechnol.* **34**, 192–198. .
- Kouno, T., Moody, J., Kwon, A.T.-J., Shibayama, Y., Kato, S., Huang, Y., Böttcher, M., Motakis, E., Mendez, M., Severin, J., et al. (2019). C1 CAGE detects transcription start sites and enhancer activity at single-cell resolution. *Nat. Commun.* **10**, 360. .
- Kouzarides, T. (2007). Chromatin modifications and their function. *Cell* **128**, 693–705. .
- Krivega, I., and Dean, A. (2017). LDB1-mediated enhancer looping can be established independent of mediator and cohesin. *Nucleic Acids Res.* **45**, 8255–8268. .
- Kuscu, C., Mammadov, R., Czikora, A., Unlu, H., Tufan, T., Fischer, N.L., Arslan, S., Bekiranov, S., Kanemaki, M., and Adli, M. (2019). Temporal and Spatial Epigenome Editing Allows Precise Gene Regulation in Mammalian Cells. *J. Mol. Biol.* **431**, 111–121. .
- Kutner, R.H., Zhang, X.-Y., and Reiser, J. (2009). Production, concentration and titration of pseudotyped HIV-1-based lentiviral vectors. *Nat. Protoc.* **4**, 495–505. .
- Kwon, D.Y., Zhao, Y.-T., Lamonica, J.M., and Zhou, Z. (2017). Locus-specific histone deacetylation using a synthetic CRISPR-Cas9-based HDAC. *Nat. Commun.* **8**, 15315. .

- Kyzar, E.J., Zhang, H., and Pandey, S.C. (2019). Adolescent Alcohol Exposure Epigenetically Suppresses Amygdala Arc Enhancer RNA Expression to Confer Adult Anxiety Susceptibility. *Biol. Psychiatry* 85, 904–914. .
- Lai, F., Orom, U.A., Cesaroni, M., Beringer, M., Taatjes, D.J., Blobel, G.A., and Shiekhataar, R. (2013). Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* 494, 497–501. .
- Lai, F., Gardini, A., Zhang, A., and Shiekhataar, R. (2015). Integrator mediates the biogenesis of enhancer RNAs. *Nature* 525, 399–403. .
- Lawhorn, I.E.B., Ferreira, J.P., and Wang, C.L. (2014). Evaluation of sgRNA target sites for CRISPR-mediated repression of TP53. *PLoS One* 9, e113232. .
- van der Lelij, P., Newman, J.A., Lieb, S., Jude, J., Katis, V., Hoffmann, T., Hinterndorfer, M., Bader, G., Kraut, N., Pearson, M.A., et al. (2020). STAG1 vulnerabilities for exploiting cohesin synthetic lethality in STAG2-deficient cancers. *Life Sci Alliance* 3. <https://doi.org/10.26508/lsa.202000725>.
- Li, W., Notani, D., Ma, Q., Tanasa, B., Nunez, E., Chen, A.Y., Merkurjev, D., Zhang, J., Ohgi, K., Song, X., et al. (2013). Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature* 498, 516–520. .
- Li, W., Xu, H., Xiao, T., Cong, L., Love, M.I., Zhang, F., Irizarry, R.A., Liu, J.S., Brown, M., and Liu, X.S. (2014). MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol.* 15, 554. .
- Li, W., Notani, D., and Rosenfeld, M.G. (2016). Enhancers as non-coding RNA transcription units: recent insights and future perspectives. *Nat. Rev. Genet.* 17, 207–223. .
- Li, X., Zhang, Q., Shi, Q., Liu, Y., Zhao, K., Shen, Q., Shi, Y., Liu, X., Wang, C., Li, N., et al. (2017). Demethylase Kdm6a epigenetically promotes IL-6 and IFN- β production in macrophages. *J. Autoimmun.* 80, 85–94. .
- Liang, J., Zhou, H., Gerdt, C., Tan, M., Colson, T., Kaye, K.M., Kieff, E., and Zhao, B. (2016). Epstein-Barr virus super-enhancer eRNAs are essential for MYC oncogene expression and lymphoblast proliferation. *Proc. Natl. Acad. Sci. U. S. A.* 113, 14121–14126. .
- Lidschreiber, K., Jung, L.A., von der Emde, H., Dave, K., Taipale, J., Cramer, P., and Lidschreiber, M. (2021). Transcriptionally active enhancers in human cancer cells. *Mol. Syst. Biol.* 17, e9873. .
- Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289–293. .
- Liu, J., Kuszynski, C.A., and Baxter, B.T. (1999). Doxycycline induces Fas/Fas ligand-mediated apoptosis in Jurkat T lymphocytes. *Biochem. Biophys. Res. Commun.* 260, 562–567. .
- Liu, X.S., Wu, H., Ji, X., Stelzer, Y., Wu, X., Czauderna, S., Shu, J., Dadon, D., Young, R.A., and Jaenisch, R. (2016). Editing DNA Methylation in the Mammalian Genome. *Cell* 167, 233–247.e17. .
- Liu, Y., Easton, J., Shao, Y., Maciaszek, J., Wang, Z., Wilkinson, M.R., McCastlain, K., Edmonson, M., Pounds, S.B., Shi, L., et al. (2017). The genomic landscape of pediatric and young adult T-lineage acute lymphoblastic leukemia. *Nat. Genet.* 49, 1211–1218. .

- Loew, R., Heinz, N., Hampf, M., Bujard, H., and Gossen, M. (2010). Improved Tet-responsive promoters with minimized background expression. *BMC Biotechnol.* *10*, 81. .
- Logan, A.C., Nightingale, S.J., Haas, D.L., Cho, G.J., Pepper, K.A., and Kohn, D.B. (2004). Factors influencing the titer and infectivity of lentiviral vectors. *Hum. Gene Ther.* *15*, 976–988. .
- Lombardo, A., Cesana, D., Genovese, P., Di Stefano, B., Provasi, E., Colombo, D.F., Neri, M., Magnani, Z., Cantore, A., Lo Riso, P., et al. (2011). Site-specific integration and tailoring of cassette design for sustainable gene transfer. *Nat. Methods* *8*, 861–869. .
- Lovén, J., Hoke, H.A., Lin, C.Y., Lau, A., Orlando, D.A., Vakoc, C.R., Bradner, J.E., Lee, T.I., and Young, R.A. (2013). Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* *153*, 320–334. .
- Lundin, A., Porritt, M.J., Jaiswal, H., Seeliger, F., Johansson, C., Bidar, A.W., Badertscher, L., Wimberger, S., Davies, E.J., Hardaker, E., et al. (2020). Development of an ObLiGaRe Doxycycline Inducible Cas9 system for pre-clinical cancer drug discovery. *Nat. Commun.* *11*, 4903. .
- Ma, Q., Wang, Y., Lo, A.S.-Y., Gomes, E.M., and Junghans, R.P. (2010). Cell density plays a critical role in ex vivo expansion of T cells for adoptive immunotherapy. *J. Biomed. Biotechnol.* *2010*, 386545. .
- Mahat, D.B., Kwak, H., Booth, G.T., Jonkers, I.H., Danko, C.G., Patel, R.K., Waters, C.T., Munson, K., Core, L.J., and Lis, J.T. (2016). Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). *Nat. Protoc.* *11*, 1455–1476. .
- Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. *Science* *339*, 823–826. .
- Mansour, M.R., Abraham, B.J., Anders, L., Berezovskaya, A., Gutierrez, A., Durbin, A.D., Etchin, J., Lawton, L., Sallan, S.E., Silverman, L.B., et al. (2014). Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. *Science* *346*, 1373–1377. .
- Margolin, J.F., Friedman, J.R., Meyer, W.K., Vissing, H., Thiesen, H.J., and Rauscher, F.J., 3rd (1994). Krüppel-associated boxes are potent transcriptional repression domains. *Proc. Natl. Acad. Sci. U. S. A.* *91*, 4509–4513. .
- Martin, R.D., Hébert, T.E., and Tanny, J.C. (2020). Therapeutic targeting of the general RNA polymerase II transcription machinery. *Int. J. Mol. Sci.* *21*, 3354. .
- Maruyama, A., Mimura, J., and Itoh, K. (2014). Non-coding RNA derived from the region adjacent to the human HO-1 E2 enhancer selectively regulates HO-1 gene induction by modulating Pol II binding. *Nucleic Acids Res.* *42*, 13599–13614. .
- Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., et al. (2012). Systematic localization of common disease-associated variation in regulatory DNA. *Science* *337*, 1190–1195. .
- McGrath, P.T. (2011). Characterizing cDNA ends by circular RACE. *Methods Mol. Biol.* *772*, 257–265. .
- Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B., et al. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* *454*, 766–770. .

Michlits, G., Hubmann, M., Wu, S.-H., Vainorius, G., Budusan, E., Zhuk, S., Burkard, T.R., Novatchkova, M., Aichinger, M., Lu, Y., et al. (2017). CRISPR-UMI: single-cell lineage tracing of pooled CRISPR-Cas9 screens. *Nat. Methods* **14**, 1191–1197. .

Michlits, G., Jude, J., Hinterndorfer, M., de Almeida, M., Vainorius, G., Hubmann, M., Neumann, T., Schleiffer, A., Burkard, T.R., Fellner, M., et al. (2020). Multilayered VBC score predicts sgRNAs that efficiently generate loss-of-function alleles. *Nat. Methods* **17**, 708–716. .

Miguel-Escalada, I., Pasquali, L., and Ferrer, J. (2015). Transcriptional enhancers: functional insights and role in human disease. *Curr. Opin. Genet. Dev.* **33**, 71–76. .

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.-K., Koche, R.P., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553–560. .

Miles, L.A., Garippa, R.J., and Poirier, J.T. (2016). Design, execution, and analysis of pooled in vitro CRISPR/Cas9 screens. *FEBS J.* **283**, 3170–3180. .

Miller, J.A., and Widom, J. (2003). Collaborative competition mechanism for gene activation in vivo. *Mol. Cell. Biol.* **23**, 1623–1632. .

Montalbano, A., Canver, M.C., and Sanjana, N.E. (2017). High-Throughput Approaches to Pinpoint Function within the Noncoding Genome. *Mol. Cell* **68**, 44–59. .

Morcillo, P., Rosen, C., Baylies, M.K., and Dorsett, D. (1997). Chip, a widely expressed chromosomal protein required for segmentation and activity of a remote wing margin enhancer in *Drosophila*. *Genes Dev.* **11**, 2729–2740. .

Moreau, P., Hen, R., Wasylyk, B., Everett, R., Gaub, M.P., and Chambon, P. (1981). The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. *Nucleic Acids Res.* **9**, 6047–6068. .

Moullan, N., Mouchiroud, L., Wang, X., Ryu, D., Williams, E.G., Mottis, A., Jovaisaite, V., Frochaux, M.V., Quiros, P.M., Deplancke, B., et al. (2015). Tetracyclines Disturb Mitochondrial Function across Eukaryotic Models: A Call for Caution in Biomedical Research. *Cell Rep.* **10**, 1681–1691. .

Mousavi, K., Zare, H., Dell’orso, S., Grontved, L., Gutierrez-Cruz, G., Derfoul, A., Hager, G.L., and Sartorelli, V. (2013). eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci. *Mol. Cell* **51**, 606–617. .

Nguyen, D.P., Miyaoka, Y., Gilbert, L.A., Mayerl, S.J., Lee, B.H., Weissman, J.S., Conklin, B.R., and Wells, J.A. (2016). Ligand-binding domains of nuclear receptors facilitate tight control of split CRISPR activity. *Nat. Commun.* **7**, 1–10. .

Nichols, M.H., and Corces, V.G. (2021). Principles of 3D compartmentalization of the human genome. *Cell Rep.* **35**, 109330. .

Nielsen, S., Yuzenkova, Y., and Zenkin, N. (2013). Mechanism of eukaryotic RNA polymerase III transcription termination. *Science* **340**, 1577–1580. .

Nightingale, K.P., O’Neill, L.P., and Turner, B.M. (2006). Histone modifications: signalling receptors and potential elements of a heritable epigenetic code. *Curr. Opin. Genet. Dev.* **16**, 125–136. .

- Nikolic, J., Belot, L., Raux, H., Legrand, P., Gaudin, Y., and Albertini, A. (2018). Structural basis for the recognition of LDL-receptor family members by VSV glycoprotein. *Nat. Commun.* 9, 1029. .
- Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 485, 381–385. .
- Nora, E.P., Goloborodko, A., Valton, A.-L., Gibcus, J.H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L.A., and Bruneau, B.G. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* 169, 930-944.e22. .
- Ntini, E., Järvelin, A.I., Bornholdt, J., Chen, Y., Boyd, M., Jørgensen, M., Andersson, R., Hoof, I., Schein, A., Andersen, P.R., et al. (2013). Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality. *Nat. Struct. Mol. Biol.* 20, 923–928. .
- Oakes, B.L., Nadler, D.C., Flamholz, A., Fellmann, C., Staahl, B.T., Doudna, J.A., and Savage, D.F. (2016). Profiling of engineering hotspots identifies an allosteric CRISPR-Cas9 switch. *Nat. Biotechnol.* 34, 646–651. .
- Ogryzko, V.V., Schiltz, R.L., Russanova, V., Howard, B.H., and Nakatani, Y. (1996). The transcriptional coactivators p300 and CBP are histone acetyltransferases. *Cell* 87, 953–959. .
- Ohtsuki, S., Levine, M., and Cai, H.N. (1998). Different core promoters possess distinct regulatory activities in the *Drosophila* embryo. *Genes Dev.* 12, 547–556. .
- Okat, Z. (2018). Molecular dynamics of estrogen receptors. *Eurasian J. Med. Oncol.* <https://doi.org/10.14744/ejmo.2018.76476>.
- Onodera, C.S., Underwood, J.G., Katzman, S., Jacobs, F., Greenberg, D., Salama, S.R., and Haussler, D. (2012). Gene isoform specificity through enhancer-associated antisense transcription. *PLoS One* 7, e43511. .
- Ørom, U.A., Derrien, T., Beringer, M., Gumireddy, K., Gardini, A., Bussotti, G., Lai, F., Zytynski, M., Notredame, C., Huang, Q., et al. (2010). Long noncoding RNAs with enhancer-like function in human cells. *Cell* 143, 46–58. .
- Osterwalder, M., Barozzi, I., Tissières, V., Fukuda-Yuzawa, Y., Mannion, B.J., Afzal, S.Y., Lee, E.A., Zhu, Y., Plajzer-Frick, I., Pickle, C.S., et al. (2018). Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature* 554, 239–243. .
- Ostuni, R., Piccolo, V., Barozzi, I., Polletti, S., Termanini, A., Bonifacio, S., Curina, A., Prosperini, E., Ghisletti, S., and Natoli, G. (2013). Latent enhancers activated by stimulation in differentiated cells. *Cell* 152, 157–171. .
- Palii, C.G., Perez-Iratxeta, C., Yao, Z., Cao, Y., Dai, F., Davison, J., Atkins, H., Allan, D., Dilworth, F.J., Gentleman, R., et al. (2011). Differential genomic targeting of the transcription factor TAL1 in alternate haematopoietic lineages. *EMBO J.* 30, 494–509. .
- Palomero, T., Odom, D.T., O’Neil, J., Ferrando, A.A., Margolin, A., Neuberg, D.S., Winter, S.S., Larson, R.S., Li, W., Liu, X.S., et al. (2006). Transcriptional regulatory networks downstream of TAL1/SCL in T-cell acute lymphoblastic leukemia. *Blood* 108, 986–992. .
- Panigrahi, A., and O’Malley, B.W. (2021). Mechanisms of enhancer action: the known and the unknown. *Genome Biol.* 22, 108. .

- Parker, S.C.J., Stitzel, M.L., Taylor, D.L., Orozco, J.M., Erdos, M.R., Akiyama, J.A., van Bueren, K.L., Chines, P.S., Narisu, N., NISC Comparative Sequencing Program, et al. (2013). Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc. Natl. Acad. Sci. U. S. A.* *110*, 17921–17926. .
- Parnas, O., Jovanovic, M., Eisenhaure, T.M., Herbst, R.H., Dixit, A., Ye, C.J., Przybylski, D., Platt, R.J., Tirosh, I., Sanjana, N.E., et al. (2015). A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks. *Cell* *162*, 675–686. .
- Patwardhan, R.P., Lee, C., Litvin, O., Young, D.L., Pe'er, D., and Shendure, J. (2009). High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. *Nat. Biotechnol.* *27*, 1173–1175. .
- Pekowska, A., Benoukraf, T., Zacarias-Cabeza, J., Belhocine, M., Koch, F., Holota, H., Imbert, J., Andrau, J.-C., Ferrier, P., and Spicuglia, S. (2011). H3K4 tri-methylation provides an epigenetic signature of active enhancers. *EMBO J.* *30*, 4198–4210. .
- Perez-Pinera, P., Kocak, D.D., Vockley, C.M., Adler, A.F., Kabadi, A.M., Polstein, L.R., Thakore, P.I., Glass, K.A., Ousterout, D.G., Leong, K.W., et al. (2013). RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat. Methods* *10*, 973–976. .
- Peters, A.H.F.M., Mermoud, J.E., O'Carroll, D., Pagani, M., Schweizer, D., Brockdorff, N., and Jenuwein, T. (2002). Histone H3 lysine 9 methylation is an epigenetic imprint of facultative heterochromatin. *Nat. Genet.* *30*, 77–80. .
- Phatnani, H.P., Jones, J.C., and Greenleaf, A.L. (2004). Expanding the functional repertoire of CTD kinase I and RNA polymerase II: novel phosphoCTD-associating proteins in the yeast proteome. *Biochemistry* *43*, 15702–15719. .
- Phillips-Cremins, J.E., Sauria, M.E.G., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S.K., Ong, C.-T., Hookway, T.A., Guo, C., Sun, Y., et al. (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* *153*, 1281–1295. .
- Platt, R.J., Chen, S., Zhou, Y., Yim, M.J., Swiech, L., Kempton, H.R., Dahlman, J.E., Parnas, O., Eisenhaure, T.M., Jovanovic, M., et al. (2014). CRISPR-Cas9 knockin mice for genome editing and cancer modeling. *Cell* *159*, 440–455. .
- Pnueli, L., Rudnizky, S., Yosefzon, Y., and Melamed, P. (2015). RNA transcribed from a distal enhancer is required for activating the chromatin at the promoter of the gonadotropin α -subunit gene. *Proc. Natl. Acad. Sci. U. S. A.* *112*, 4369–4374. .
- Polstein, L.R., Perez-Pinera, P., Kocak, D.D., Vockley, C.M., Bledsoe, P., Song, L., Safi, A., Crawford, G.E., Reddy, T.E., and Gersbach, C.A. (2015). Genome-wide specificity of DNA binding, gene regulation, and chromatin remodeling by TALE- and CRISPR/Cas9-based transcriptional activators. *Genome Res.* *25*, 1158–1169. .
- Porcher, C., Chagraoui, H., and Kristiansen, M.S. (2017). SCL/TAL1: a multifaceted regulator from blood development to disease. *Blood* *129*, 2051–2060. .
- Potter, H. (2003). Transfection by electroporation. *Curr. Protoc. Mol. Biol.* *Chapter 9*, Unit 9.3. .
- Pu, J., Frescas, D., Zhang, B., and Feng, J. (2015). Utilization of TALEN and CRISPR/Cas9 technologies for gene targeting and modification. *Exp. Biol. Med.* *240*, 1065–1070. .

- Pulakanti, K., Pinello, L., Stelloh, C., Blinka, S., Allred, J., Milanovich, S., Kiblawi, S., Peterson, J., Wang, A., Yuan, G.-C., et al. (2013). Enhancer transcribed RNAs arise from hypomethylated, Tet-occupied genomic regions. *Epigenetics* 8, 1303–1320. .
- Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 152, 1173–1183. .
- Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* 470, 279–283. .
- Rahnamoun, H., Lee, J., Sun, Z., Lu, H., Ramsey, K.M., Komives, E.A., and Lauberth, S.M. (2018). RNAs interact with BRD4 to promote enhanced chromatin engagement and transcription activation. *Nat. Struct. Mol. Biol.* 25, 687–697. .
- Rajagopal, N., Srinivasan, S., Kooshesh, K., Guo, Y., Edwards, M.D., Banerjee, B., Syed, T., Emons, B.J.M., Gifford, D.K., and Sherwood, R.I. (2016). High-throughput mapping of regulatory DNA. *Nat. Biotechnol.* 34, 167–174. .
- Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* 8, 2281–2308. .
- Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1665–1680. .
- Rao, S.S.P., Huang, S.-C., Glenn St Hilaire, B., Engreitz, J.M., Perez, E.M., Kieffer-Kwon, K.-R., Sanborn, A.L., Johnstone, S.E., Bascom, G.D., Bochkov, I.D., et al. (2017). Cohesin Loss Eliminates All Loop Domains. *Cell* 171, 305–320.e24. .
- Reiser, J. (2000). Production and concentration of pseudotyped HIV-1-based gene transfer vectors. *Gene Ther.* 7, 910–913. .
- Reiter, F., Wienerroither, S., and Stark, A. (2017). Combinatorial function of transcription factors and cofactors. *Curr. Opin. Genet. Dev.* 43, 73–81. .
- Ren, C., Liu, F., Ouyang, Z., An, G., Zhao, C., Shuai, J., Cai, S., Bo, X., and Shu, W. (2017). Functional annotation of structural ncRNAs within enhancer RNAs in the human genome: implications for human disease. *Sci. Rep.* 7, 15518. .
- Reuter, J.S., and Mathews, D.H. (2010). RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* 11, 129. .
- Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., et al. (2015). Integrative analysis of 111 reference human epigenomes. *Nature* 518, 317–330. .
- Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–26. .
- Rosenbauer, F., Wagner, K., Kutok, J.L., Iwasaki, H., Le Beau, M.M., Okuno, Y., Akashi, K., Fiering, S., and Tenen, D.G. (2004). Acute myeloid leukemia induced by graded reduction of a lineage-specific transcription factor, PU.1. *Nat. Genet.* 36, 624–630. .

Rubin, A.J., Barajas, B.C., Furlan-Magaril, M., Lopez-Pajares, V., Mumbach, M.R., Howard, I., Kim, D.S., Boxer, L.D., Cairns, J., Spivakov, M., et al. (2017). Lineage-specific dynamic and pre-established enhancer-promoter contacts cooperate in terminal differentiation. *Nat. Genet.* **49**, 1522–1528. .

Sajwan, S., and Mannervik, M. (2019). Gene activation by dCas9-CBP and the SAM system differ in target preference. *Sci. Rep.* **9**, 18104. .

Sakaguchi, Y., Nishikawa, K., Seno, S., Matsuda, H., Takayanagi, H., and Ishii, M. (2018). Roles of Enhancer RNAs in RANKL-induced Osteoclast Differentiation Identified by Genome-wide Cap-analysis of Gene Expression using CRISPR/Cas9. *Sci. Rep.* **8**, 7504. .

Sanborn, A.L., Rao, S.S.P., Huang, S.-C., Durand, N.C., Huntley, M.H., Jewett, A.I., Bochkov, I.D., Chinnappan, D., Cutkosky, A., Li, J., et al. (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E6456–E6465. .

Sanda, T., Lawton, L.N., Barrasa, M.I., Fan, Z.P., Kohlhammer, H., Gutierrez, A., Ma, W., Tatarek, J., Ahn, Y., Kelliher, M.A., et al. (2012). Core transcriptional regulatory circuit controlled by the TAL1 complex in human T cell acute lymphoblastic leukemia. *Cancer Cell* **22**, 209–221. .

Sanjana, N.E., Shalem, O., and Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784. .

Sanjana, N.E., Wright, J., Zheng, K., Shalem, O., Fontanillas, P., Joung, J., Cheng, C., Regev, A., and Zhang, F. (2016). High-resolution interrogation of functional elements in the noncoding genome. *Science* **353**, 1545–1549. .

Sano, T., Ito, T., Ishigami, S., Bandaru, S., and Sano, S. (2020). Intrinsic activation of cardiosphere-derived cells enhances myocardial repair. *J. Thorac. Cardiovasc. Surg.* <https://doi.org/10.1016/j.jtcvs.2020.05.040>.

Sanson, K.R., Hanna, R.E., Hegde, M., Donovan, K.F., Strand, C., Sullender, M.E., Vaimberg, E.W., Goodale, A., Root, D.E., Piccioni, F., et al. (2018). Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. *Nat. Commun.* **9**, 1–15. .

Sanyal, A., Lajoie, B.R., Jain, G., and Dekker, J. (2012). The long-range interaction landscape of gene promoters. *Nature* **489**, 109–113. .

Sapranaukas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., and Siksnys, V. (2011). The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res.* **39**, 9275–9282. .

Schaukowitch, K., Joo, J.-Y., Liu, X., Watts, J.K., Martinez, C., and Kim, T.-K. (2014). Enhancer RNA facilitates NELF release from immediate early genes. *Mol. Cell* **56**, 29–42. .

Scherr, M., Battmer, K., Blömer, U., Ganser, A., and Grez, M. (2001). Quantitative determination of lentiviral vector particle numbers by real-time PCR. *Biotechniques* **31**, 520, 522, 524, passim. .

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682. .

Schmierer, B., Botla, S.K., Zhang, J., Turunen, M., Kivioja, T., and Taipale, J. (2017). CRISPR/Cas9 screening using unique molecular identifiers. *Mol. Syst. Biol.* **13**, 945. .

- Schneider, M., Marison, I.W., and von Stockar, U. (1996). The importance of ammonia in mammalian cell culture. *J. Biotechnol.* **46**, 161–185. .
- Schubert, M.S., Thommandru, B., Woodley, J., Turk, R., Yan, S., Kurgan, G., McNeill, M.S., and Rettig, G.R. (2021). Optimized design parameters for CRISPR Cas9 and Cas12a homology-directed repair. *Sci. Rep.* **11**, 19482. .
- Schultz, D.C., Ayyanathan, K., Negorev, D., Maul, G.G., and Rauscher, F.J., 3rd (2002). SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes Dev.* **16**, 919–932. .
- Schwarzer, W., Abdennur, N., Goloborodko, A., Pekowska, A., Fudenberg, G., Loe-Mie, Y., Fonseca, N.A., Huber, W., Haering, C.H., Mirny, L., et al. (2017). Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**, 51–56. .
- Scully, R., Panday, A., Elango, R., and Willis, N.A. (2019). DNA double-strand break repair-pathway choice in somatic mammalian cells. *Nat. Rev. Mol. Cell Biol.* **20**, 698–714. .
- Segal, D.J., Beerli, R.R., Blancafort, P., Dreier, B., Effertz, K., Huber, A., Kokschi, B., Lund, C.V., Magnenat, L., Valente, D., et al. (2003). Evaluation of a modular strategy for the construction of novel polydactyl zinc finger DNA-binding proteins. *Biochemistry* **42**, 2137–2148. .
- Sen, D.R., Kaminski, J., Barnitz, R.A., Kurachi, M., Gerdemann, U., Yates, K.B., Tsao, H.-W., Godec, J., LaFleur, M.W., Brown, F.D., et al. (2016). The epigenetic landscape of T cell exhaustion. *Science* **354**, 1165–1169. .
- Sen, P., Lan, Y., Li, C.Y., Sidoli, S., Donahue, G., Dou, Z., Frederick, B., Chen, Q., Luense, L.J., Garcia, B.A., et al. (2019). Histone Acetyltransferase p300 Induces De Novo Super-Enhancers to Drive Cellular Senescence. *Mol. Cell* **73**, 684-698.e8. .
- Sérandour, A.A., Avner, S., Percevault, F., Demay, F., Bizot, M., Lucchetti-Miganeh, C., Barloy-Hubler, F., Brown, M., Lupien, M., Métivier, R., et al. (2011). Epigenetic switch involved in activation of pioneer factor FOXA1-dependent enhancers. *Genome Res.* **21**, 555–565. .
- Serfling, E., Jasin, M., and Schaffner, W. (1985). Enhancers and eukaryotic gene transcription. *Trends Genet.* **1**, 224–230. .
- Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelsen, T., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G., et al. (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84–87. .
- Shang, W., Wang, F., Fan, G., and Wang, H. (2017). Key elements for designing and performing a CRISPR/Cas9-based genetic screen. *J. Genet. Genomics* **44**, 439–449. .
- Shang, W., Jiang, Y., Boettcher, M., Ding, K., Mollenauer, M., Liu, Z., Wen, X., Liu, C., Hao, P., Zhao, S., et al. (2018). Genome-wide CRISPR screen identifies FAM49B as a key regulator of actin dynamics and T cell activation. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E4051–E4060. .
- Sharpe, J., Nonchev, S., Gould, A., Whiting, J., and Krumlauf, R. (1998). Selectivity, sharing and competitive interactions in the regulation of Hoxb genes. *EMBO J.* **17**, 1788–1798. .
- Shen, C.-C., Hsu, M.-N., Chang, C.-W., Lin, M.-W., Hwu, J.-R., Tu, Y., and Hu, Y.-C. (2019). Synthetic switch to minimize CRISPR off-target effects by self-restricting Cas9 transcription and translation. *Nucleic Acids Res.* **47**, e13. .

- Shi, J., Wang, E., Milazzo, J.P., Wang, Z., Kinney, J.B., and Vakoc, C.R. (2015). Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. *Nat. Biotechnol.* **33**, 661–667. .
- Shi, L., Song, L., Maurer, K., Zhang, Z., and Sullivan, K.E. (2017). SERPINB2 is regulated by dynamic interactions with pause-release proteins and enhancer RNAs. *Mol. Immunol.* **88**, 20–31. .
- Sigova, A.A., Abraham, B.J., Ji, X., Molinie, B., Hannett, N.M., Guo, Y.E., Jangi, M., Giallourakis, C.C., Sharp, P.A., and Young, R.A. (2015). Transcription factor trapping by RNA in gene regulatory elements. *Science* **350**, 978–981. .
- Simeonov, D.R., Gowen, B.G., Boontanart, M., Roth, T.L., Gagnon, J.D., Mumbach, M.R., Satpathy, A.T., Lee, Y., Bray, N.L., Chan, A.Y., et al. (2017). Discovery of stimulation-responsive immune enhancers with CRISPR activation. *Nature* **549**, 111–115. .
- Simon, J.M., Giresi, P.G., Davis, I.J., and Lieb, J.D. (2012). Using formaldehyde-assisted isolation of regulatory elements (FAIRE) to isolate active regulatory DNA. *Nat. Protoc.* **7**, 256–267. .
- Smith, J.A., and Daniel, R. (2006). Following the path of the virus: the exploitation of host DNA repair mechanisms by retroviruses. *ACS Chem. Biol.* **1**, 217–226. .
- Song, L., Zhang, Z., Grasfeder, L.L., Boyle, A.P., Giresi, P.G., Lee, B.-K., Sheffield, N.C., Gräf, S., Huss, M., Keefe, D., et al. (2011). Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. *Genome Res.* **21**, 1757–1767. .
- Spitz, F., and Furlong, E.E.M. (2012). Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.* **13**, 613–626. .
- Sripathy, S.P., Stevens, J., and Schultz, D.C. (2006). The KAP1 corepressor functions to coordinate the assembly of de novo HP1-demarcated microenvironments of heterochromatin required for KRAB zinc finger protein-mediated transcriptional repression. *Mol. Cell. Biol.* **26**, 8623–8638. .
- Stadler, M.B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Schöler, A., van Nimwegen, E., Wirbelauer, C., Oakeley, E.J., Gaidatzis, D., et al. (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480**, 490–495. .
- Struhl, K. (2007). Transcriptional noise and the fidelity of initiation by RNA polymerase II. *Nat. Struct. Mol. Biol.* **14**, 103–105. .
- Sur, I., and Taipale, J. (2016). The role of enhancers in cancer. *Nat. Rev. Cancer* **16**, 483–493. .
- Sutherland, H., and Bickmore, W.A. (2009). Transcription factories: gene expression in unions? *Nat. Rev. Genet.* **10**, 457–466. .
- Suzuki, Y., and Craigie, R. (2007). The road to chromatin - nuclear entry of retroviruses. *Nat. Rev. Microbiol.* **5**, 187–196. .
- Symmons, O., Uslu, V.V., Tsujimura, T., Ruf, S., Nassari, S., Schwarzer, W., Ettwiller, L., and Spitz, F. (2014). Functional and topological characteristics of mammalian regulatory domains. *Genome Res.* **24**, 390–400. .
- Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., et al. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–935. .

- Tan, S.H., Leong, W.Z., Ngoc, P.C.T., Tan, T.K., Bertulfo, F.C., Lim, M.C., An, O., Li, Z., Yeoh, A.E.J., Fullwood, M.J., et al. (2019). The enhancer RNA ARIEL activates the oncogenic transcriptional program in T-cell acute lymphoblastic leukemia. *Blood* 134, 239–251. .
- Thakore, P.I., D'Ippolito, A.M., Song, L., Safi, A., Shivakumar, N.K., Kabadi, A.M., Reddy, T.E., Crawford, G.E., and Gersbach, C.A. (2015). Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nat. Methods* 12, 1143–1149. .
- Thompson, N.A., Ranzani, M., van der Weyden, L., Iyer, V., Offord, V., Droop, A., Behan, F., Gonçalves, E., Speak, A., Iorio, F., et al. (2021). Combinatorial CRISPR screen identifies fitness effects of gene paralogues. *Nat. Commun.* 12, 1302. .
- Thompson, P.R., Wang, D., Wang, L., Fulco, M., Pediconi, N., Zhang, D., An, W., Ge, Q., Roeder, R.G., Wong, J., et al. (2004). Regulation of the p300 HAT domain via a novel activation loop. *Nat. Struct. Mol. Biol.* 11, 308–315. .
- Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B., et al. (2012). The accessible chromatin landscape of the human genome. *Nature* 489, 75–82. .
- Tie, F., Banerjee, R., Stratton, C.A., Prasad-Sinha, J., Stepanik, V., Zlobin, A., Diaz, M.O., Scacheri, P.C., and Harte, P.J. (2009). CBP-mediated acetylation of histone H3 lysine 27 antagonizes *Drosophila* Polycomb silencing. *Development* 136, 3131–3141. .
- Tolhuis, B., Palstra, R.J., Splinter, E., Grosveld, F., and de Laat, W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol. Cell* 10, 1453–1465. .
- Triebenberg, S.J., Kingsbury, R.C., and McKnight, S.L. (1988). Functional dissection of VP16, the trans-activator of herpes simplex virus immediate early gene expression. *Genes Dev.* 2, 718–729. .
- Tritsch, G.L., and Moore, G.E. (1962). Spontaneous decomposition of glutamine in cell culture media. *Exp. Cell Res.* 28, 360–364. .
- Tsai, P.-F., Dell'Orso, S., Rodriguez, J., Vivanco, K.O., Ko, K.-D., Jiang, K., Juan, A.H., Sarshad, A.A., Vian, L., Tran, M., et al. (2018). A Muscle-Specific Enhancer RNA Mediates Cohesin Recruitment and Regulates Transcription In trans. *Mol. Cell* 71, 129-141.e8. .
- Turner, B.M. (2007). Defining an epigenetic code. *Nat. Cell Biol.* 9, 2–6. .
- Tyssowski, K.M., DeStefino, N.R., Cho, J.-H., Dunn, C.J., Poston, R.G., Carty, C.E., Jones, R.D., Chang, S.M., Romeo, P., Wurzelmann, M.K., et al. (2018). Different Neuronal Activity Patterns Induce Different Gene Expression Programs. *Neuron* 98, 530-546.e11. .
- Urlinger, S., Baron, U., Thellmann, M., Hasan, M.T., Bujard, H., and Hillen, W. (2000). Exploring the sequence space for tetracycline-dependent transcriptional activators: Novel mutations yield expanded range and sensitivity. *Proc. Natl. Acad. Sci. U. S. A.* 97, 7963–7968. .
- Vidigal, J.A., and Ventura, A. (2015). Rapid and efficient one-step generation of paired gRNA CRISPR-Cas9 libraries. *Nat. Commun.* 6, 8083. .
- Villar, D., Berthelot, C., Aldridge, S., Rayner, T.F., Lukk, M., Pignatelli, M., Park, T.J., Deaville, R., Erichsen, J.T., Jasinska, A.J., et al. (2015). Enhancer evolution across 20 mammalian species. *Cell* 160, 554–566. .

- Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., et al. (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**, 854–858. .
- Vojta, A., Dobrinić, P., Tadić, V., Bočkor, L., Korać, P., Julg, B., Klasić, M., and Zoldoš, V. (2016). Repurposing the CRISPR-Cas9 system for targeted DNA methylation. *Nucleic Acids Res.* **44**, 5615–5628. .
- Wang, B., Wang, M., Zhang, W., Xiao, T., Chen, C.-H., Wu, A., Wu, F., Traugh, N., Wang, X., Li, Z., et al. (2019). Integrative analysis of pooled CRISPR genetic screens using MAGeCKFlute. *Nat. Protoc.* **14**, 756–780. .
- Wang, J., Zhuang, J., Iyer, S., Lin, X., Whitfield, T.W., Greven, M.C., Pierce, B.G., Dong, X., Kundaje, A., Cheng, Y., et al. (2012). Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Res.* **22**, 1798–1812. .
- Wang, M., Chen, Q., Liu, H., Chen, H., Wang, B., Zhang, W., Mading, A., Li, W., Li, S., Liu, H., et al. (2018). CRISPR/Cas9 screens Reveal Dasatinib Targets of Inhibiting T cell Activation and Proliferation.
- Wang, Q., Carroll, J.S., and Brown, M. (2005). Spatial and temporal recruitment of androgen receptor and its coactivators involves chromosomal looping and polymerase tracking. *Mol. Cell* **19**, 631–642. .
- Wang, T., Wei, J.J., Sabatini, D.M., and Lander, E.S. (2014). Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**, 80–84. .
- Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Peng, W., Zhang, M.Q., et al. (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat. Genet.* **40**, 897–903. .
- Weintraub, A.S., Li, C.H., Zamudio, A.V., Sigova, A.A., Hannett, N.M., Day, D.S., Abraham, B.J., Cohen, M.A., Nabet, B., Buckley, D.L., et al. (2017). YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell* **171**, 1573-1588.e28. .
- Wendt, K.S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T., et al. (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* **451**, 796–801. .
- Whyte, W.A., Bilodeau, S., Orlando, D.A., Hoke, H.A., Frampton, G.M., Foster, C.T., Cowley, S.M., and Young, R.A. (2012). Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. *Nature* **482**, 221–225. .
- Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319. .
- Winter, G.E., Mayer, A., Buckley, D.L., Erb, M.A., Roderick, J.E., Vittori, S., Reyes, J.M., di Iulio, J., Souza, A., Ott, C.J., et al. (2017). BET Bromodomain Proteins Function as Master Transcription Elongation Factors Independent of CDK9 Recruitment. *Mol. Cell* **67**, 5-18.e19. .
- Wright, A.V., Nuñez, J.K., and Doudna, J.A. (2016). Biology and Applications of CRISPR Systems: Harnessing Nature’s Toolbox for Genome Engineering. *Cell* **164**, 29–44. .
- Xie, J., Nair, A., and Hermiston, T.W. (2008). A comparative study examining the cytotoxicity of inducible gene expression system ligands in different cell types. *Toxicol. In Vitro* **22**, 261–266. .

- Xie, S., Duan, J., Li, B., Zhou, P., and Hon, G.C. (2017). Multiplexed Engineering and Analysis of Combinatorial Enhancer Activity in Single Cells. *Mol. Cell* 66, 285-299.e5. .
- Yamazaki, T., Souquere, S., Chujo, T., Kobelke, S., Chong, Y.S., Fox, A.H., Bond, C.S., Nakagawa, S., Pierron, G., and Hirose, T. (2018). Functional Domains of NEAT1 Architectural lncRNA Induce Paraspeckle Assembly through Phase Separation. *Mol. Cell* 70, 1038-1053.e7. .
- Yang, Y., Su, Z., Song, X., Liang, B., Zeng, F., Chang, X., and Huang, D. (2016). Enhancer RNA-driven looping enhances the transcription of the long noncoding RNA DHRS4-AS1, a controller of the DHRS4 gene cluster. *Sci. Rep.* 6, 20961. .
- Zabidi, M.A., Arnold, C.D., Schernhuber, K., Pagani, M., Rath, M., Frank, O., and Stark, A. (2015). Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. *Nature* 518, 556–559. .
- Zaret, K.S. (2020). Pioneer Transcription Factors Initiating Gene Network Changes. *Annu. Rev. Genet.* 54, 367–385. .
- Zentner, G.E., Tesar, P.J., and Scacheri, P.C. (2011). Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome Res.* 21, 1273–1283. .
- Zetsche, B., Volz, S.E., and Zhang, F. (2015). A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nat. Biotechnol.* 33, 139–142. .
- Zhang, J., Chen, L., Zhang, J., and Wang, Y. (2019). Drug Inducible CRISPR/Cas Systems. *Comput. Struct. Biotechnol. J.* 17, 1171–1177. .
- Zhang, X.-Y., La Russa, V.F., Bao, L., Kolls, J., Schwarzenberger, P., and Reiser, J. (2002). Lentiviral vectors for sustained transgene expression in human bone marrow-derived stromal cells. *Mol. Ther.* 5, 555–565. .
- Zhang, X.-Y., La Russa, V.F., and Reiser, J. (2004). Transduction of bone-marrow-derived mesenchymal stem cells by using lentivirus vectors pseudotyped with modified RD114 envelope glycoproteins. *J. Virol.* 78, 1219–1229. .
- Zhang, Y., Payne, K.J., Zhu, Y., Price, M.A., Parrish, Y.K., Zielinska, E., Barsky, L.W., and Crooks, G.M. (2005). SCL expression at critical points in human hematopoietic lineage commitment. *Stem Cells* 23, 852–860. .
- Zhao, C., Zhao, Y., Zhang, J., Lu, J., Chen, L., Zhang, Y., Ying, Y., Xu, J., Wei, S., and Wang, Y. (2018). HIT-Cas9: A CRISPR/Cas9 Genome-Editing Device under Tight and Effective Drug Control. *Mol. Ther. Nucleic Acids* 13, 208–219. .
- Zhao, Z., Tavoosidana, G., Sjölander, M., Göndör, A., Mariano, P., Wang, S., Kanduri, C., Lezcano, M., Sandhu, K.S., Singh, U., et al. (2006). Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat. Genet.* 38, 1341–1347. .
- Zhou, X., Vink, M., Klaver, B., Berkhout, B., and Das, A.T. (2006). Optimization of the Tet-On system for regulated gene expression through viral evolution. *Gene Ther.* 13, 1382–1390. .
- Zhou, Y., Kurukuti, S., Saffrey, P., Vukovic, M., Michie, A.M., Strogantsev, R., West, A.G., and Vetrie, D. (2013). Chromatin looping defines expression of TAL1, its flanking genes, and regulation in T-ALL. *Blood* 122, 4199–4209. .

Zhou, Y., Zhu, S., Cai, C., Yuan, P., Li, C., Huang, Y., and Wei, W. (2014). High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature* 509, 487–491. .

Zhu, S., Li, W., Liu, J., Chen, C.-H., Liao, Q., Xu, P., Xu, H., Xiao, T., Cao, Z., Peng, J., et al. (2016). Genome-scale deletion screening of human long non-coding RNAs using a paired-guide RNA CRISPR-Cas9 library. *Nat. Biotechnol.* 34, 1279–1286. .

Zhu, S., Cao, Z., Liu, Z., He, Y., Wang, Y., Yuan, P., Li, W., Tian, F., Bao, Y., and Wei, W. (2019). Guide RNAs with embedded barcodes boost CRISPR-pooled screens. *Genome Biol.* 20, 20. .

<https://www.addgene.org/99376/>

BD FACSMelody™ Cell Sorter Brochure, BD, 2018

<https://www.encodeproject.org/>

IBM Corp. Released 2020. IBM SPSS Statistics for Macintosh, Version 28.0. Armonk, NY: IBM Corp

www.invitrogen.com, accessed 02/03/2022

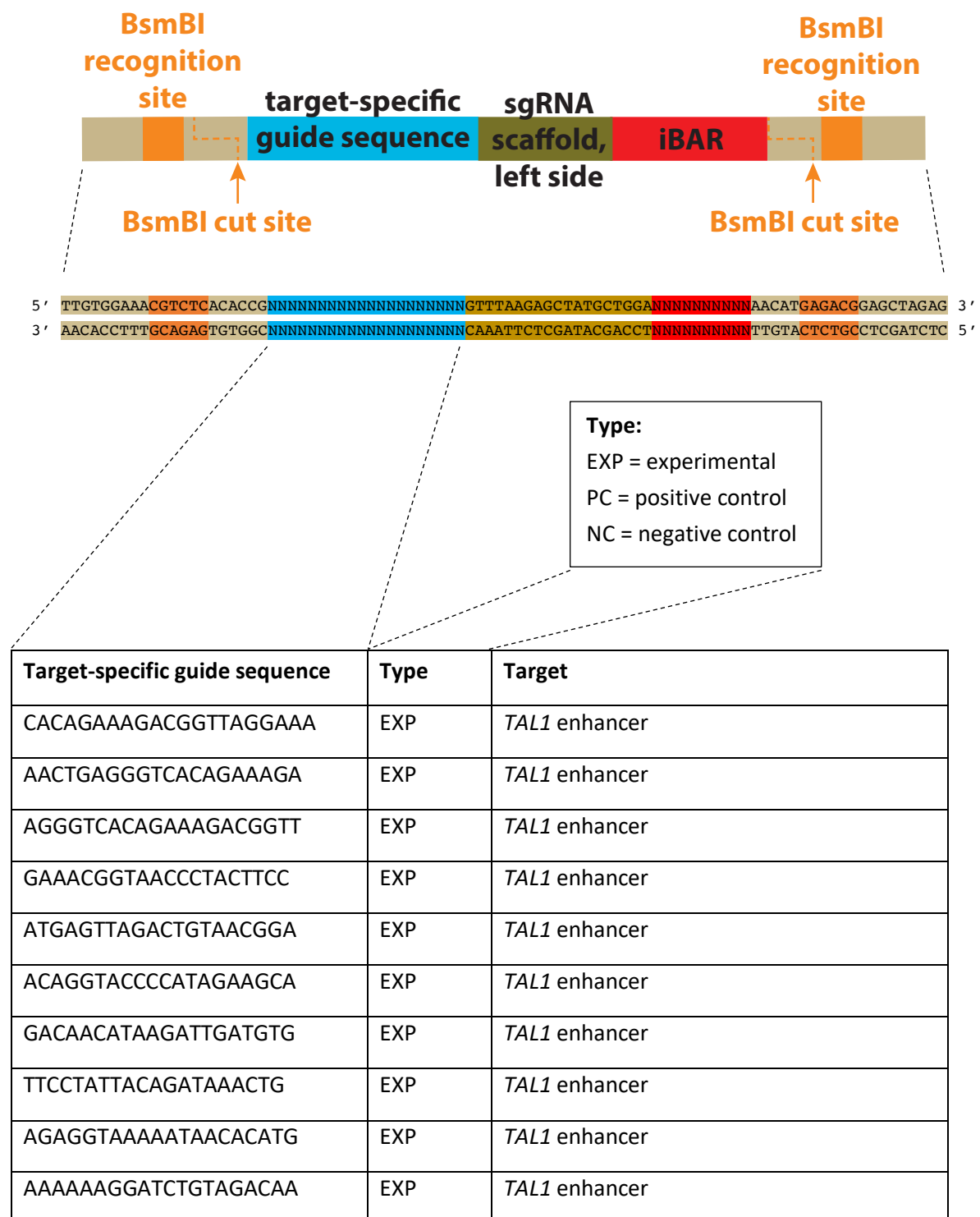
Optimizing Cluster Density on Illumina Sequencing Systems technical note, Illumina, 2016

Retro-X™ Tet-On® 3G Inducible Expression System User Manual, Clontech, 2013

Synthego Performance Analysis, ICE Analysis. 2019. v2.0. Synthego; [accessed 23/3/2020]

Appendices

Appendix 1: sgRNA sequences for the CRISPR/Cas9-mediated mutagenesis of the *TAL1* enhancer.



Target-specific guide sequence	Type	Target
AATGGCAAAGCTCAGATCAG	EXP	<i>TAL1</i> enhancer
CATGAGTTAGACTGTAACGG	EXP	<i>TAL1</i> enhancer
CTATCAGCTATATGGGACTG	EXP	<i>TAL1</i> enhancer
ACATGCGCTTAAAATAGAGA	EXP	<i>TAL1</i> enhancer
AGAGGCATTCTAGTCCAGGG	EXP	<i>TAL1</i> enhancer
G TTCAGGAGACACACACGCC	EXP	<i>TAL1</i> enhancer
AAAGGATCTGTAGACAAGGG	EXP	<i>TAL1</i> enhancer
TCCTATTACAGATAAACTGA	EXP	<i>TAL1</i> enhancer
CAAAAAAGGATCTGTAGACA	EXP	<i>TAL1</i> enhancer
TATAGCTGATTAAGAGCACA	EXP	<i>TAL1</i> enhancer
ATGTGCTCTTAATCAGCTAT	EXP	<i>TAL1</i> enhancer
AGGCGTGTGTGTCTCCTGAA	EXP	<i>TAL1</i> enhancer
GGACAACATAAGATTGATGT	EXP	<i>TAL1</i> enhancer
GTTTATCTGTAATAGGAATG	EXP	<i>TAL1</i> enhancer
ATGCTAATCAGGAGAGCAAG	EXP	<i>TAL1</i> enhancer
TGGACAACATAAGATTGATG	EXP	<i>TAL1</i> enhancer
GAGGACATATGACTGCAAAG	EXP	<i>TAL1</i> enhancer
TATCTGTAATAGGAATGGGG	EXP	<i>TAL1</i> enhancer
TGCGAAAGAGTTCTCCAAAG	EXP	<i>TAL1</i> enhancer
CAGTCCAGAACCCATGCCAC	EXP	<i>TAL1</i> enhancer
TTAATGTACAGGTGTTCTTG	EXP	<i>TAL1</i> enhancer
AGTCTAACTCATGCTGCTCA	EXP	<i>TAL1</i> enhancer
GGTCCAGCCTGTCGGAATG	EXP	<i>TAL1</i> enhancer
AGGAGATATAAATGCAGCTG	EXP	<i>TAL1</i> enhancer
AGGAAGGAGGCAATGATGAG	EXP	<i>TAL1</i> enhancer
CTCTCCTGATTAGCATACCC	EXP	<i>TAL1</i> enhancer
GCCTGGTACACTGGATGCAG	EXP	<i>TAL1</i> enhancer
GACTGTAACGGAGGGCTCCA	EXP	<i>TAL1</i> enhancer

Target-specific guide sequence	Type	Target
GAAGTGATGTAACCTGCTTA	EXP	<i>TAL1</i> enhancer
CTGTGGCCCTCTCCTCCACA	EXP	<i>TAL1</i> enhancer
CAGCTATAGGCTGTGTCCTG	EXP	<i>TAL1</i> enhancer
CCTGTTAGAAAGGCATAAGC	EXP	<i>TAL1</i> enhancer
ACATCAATCTTATGTTGTCC	EXP	<i>TAL1</i> enhancer
TAAAAATAACACATGAGGCC	EXP	<i>TAL1</i> enhancer
CGGCATGCCTTGCTTCTATG	EXP	<i>TAL1</i> enhancer
CAGCATGAGTTAGACTGTAA	EXP	<i>TAL1</i> enhancer
TCTGCTCCTCTTTCCTACCC	EXP	<i>TAL1</i> enhancer
GGAGAGCAAGTGGTGAGAGG	EXP	<i>TAL1</i> enhancer
GAGACCAAAAAAGAGGAAGG	EXP	<i>TAL1</i> enhancer
GGGTTCTGGA CTGGGCACAC	EXP	<i>TAL1</i> enhancer
CTGAACAGAGACCAAAAAAG	EXP	<i>TAL1</i> enhancer
GGTGAATGGTAAACAAATTG	EXP	<i>TAL1</i> enhancer
TCAGGAGAGCAAGTGGTGAG	EXP	<i>TAL1</i> enhancer
AACAACAGGCTCAGTGAAAG	EXP	<i>TAL1</i> enhancer
CGCCTGGTACACTGGATGCA	EXP	<i>TAL1</i> enhancer
GATTTGGAAAGTCACCGTTC	EXP	<i>TAL1</i> enhancer
TTGCCACATTCCCGACAGGC	EXP	<i>TAL1</i> enhancer
AATATAGACCTCTCAACAAC	EXP	<i>TAL1</i> enhancer
GAATGGGGTGGGGCAACCAC	EXP	<i>TAL1</i> enhancer
AAAGCTCAGATCAGGGGTGA	EXP	<i>TAL1</i> enhancer
AGGCAGATTAAAGACAGAGA	EXP	<i>TAL1</i> enhancer
TTCTAACAGGTCCAGCCTGT	EXP	<i>TAL1</i> enhancer
TCACTGAGCCTGTTGTTGAG	EXP	<i>TAL1</i> enhancer
ACAGGGCCGGGGTAGGAAAG	EXP	<i>TAL1</i> enhancer
TGCTCAGGGCCAGGCACACA	EXP	<i>TAL1</i> enhancer
ACAGGGAACCTTCTGCATAA	EXP	<i>TAL1</i> enhancer

Target-specific guide sequence	Type	Target
CAATTCAACTATCCATCAAC	EXP	<i>TAL1</i> enhancer
TGGGTCTGGACTGGGCACA	EXP	<i>TAL1</i> enhancer
CCGGCAAAGCTAATTTCTAA	EXP	<i>TAL1</i> enhancer
TTCACATCCTCAGCAATACT	EXP	<i>TAL1</i> enhancer
ACGCCTGGTACACTGGATGC	EXP	<i>TAL1</i> enhancer
TCTATCAGCTATATGGGACT	EXP	<i>TAL1</i> enhancer
GGCACACGGGCTGGGCCGTG	EXP	<i>TAL1</i> enhancer
GAGACATCTGCCAGGAAGTA	EXP	<i>TAL1</i> enhancer
ATCTGTAATAGGAATGGGGT	EXP	<i>TAL1</i> enhancer
ACAGGCTGGACCTGTTAGAA	EXP	<i>TAL1</i> enhancer
TTGACTATAAACGTTCTTTG	EXP	<i>TAL1</i> enhancer
CTAATGGCAAAGCTCAGATC	EXP	<i>TAL1</i> enhancer
TAAGAGCTCCAGCTGTGCAC	EXP	<i>TAL1</i> enhancer
GGTTGACTTTCACGCCACCC	EXP	<i>TAL1</i> enhancer
TAGGCTGTGTCCTGTGGCAT	EXP	<i>TAL1</i> enhancer
TCCCTTGAACCCACCTCCC	EXP	<i>TAL1</i> enhancer
ACCCTCAGTTTATCTGTAAT	EXP	<i>TAL1</i> enhancer
TTCTATCACTGTTAATGTAC	EXP	<i>TAL1</i> enhancer
GACACACACGCCTGGTACAC	EXP	<i>TAL1</i> enhancer
TCCCCTGCATCCAGTGTACC	EXP	<i>TAL1</i> enhancer
TTTATAAACTGCACATTCAT	EXP	<i>TAL1</i> enhancer
ACCCGGGAGGTGGGGTTCAA	EXP	<i>TAL1</i> enhancer
TGAGGATGTGAAGGAGCAAC	EXP	<i>TAL1</i> enhancer
GCTGGGCCGTGTGGAGGAGA	EXP	<i>TAL1</i> enhancer
CAGTTTATCTGTAATAGGAA	EXP	<i>TAL1</i> enhancer
GGCTCCAGGGTATGCTAATC	EXP	<i>TAL1</i> enhancer
TTCTCCAGCAACATTGTTGA	EXP	<i>TAL1</i> enhancer
CAAAAATTACTCAGCCACTT	EXP	<i>TAL1</i> enhancer

Target-specific guide sequence	Type	Target
TCGGCATGCCTTGCTTCTAT	EXP	<i>TAL1</i> enhancer
AGGCACACAGGGCACAAAAA	EXP	<i>TAL1</i> enhancer
TTGCAGTGAGAAGGAGAGGC	EXP	<i>TAL1</i> enhancer
AGACTGTAACGGAGGGCTCC	EXP	<i>TAL1</i> enhancer
TTGTTGAGAGGTCTATATTC	EXP	<i>TAL1</i> enhancer
ACACGGGCTGGGCCGTGTGG	EXP	<i>TAL1</i> enhancer
CTGGACTGGGCACACGGGCT	EXP	<i>TAL1</i> enhancer
CGGGAATGTGGCAAGAGAAT	EXP	<i>TAL1</i> enhancer
AGAAGTGATGTAACCTTGCTT	EXP	<i>TAL1</i> enhancer
ATTGATCAGAAAAGGTTTCC	EXP	<i>TAL1</i> enhancer
ATCCCTTGAACCCACCTCC	EXP	<i>TAL1</i> enhancer
CCAGCTTATGCCTTTCTAAC	EXP	<i>TAL1</i> enhancer
GTCTATCAGCTATATGGGAC	EXP	<i>TAL1</i> enhancer
AGAAATTAGCTTTGCCGGCC	EXP	<i>TAL1</i> enhancer
AAGAGATGATAAGAGATAAA	EXP	<i>TAL1</i> enhancer
AAGCTGGTCTATCAGCTATA	EXP	<i>TAL1</i> enhancer
CTGTGGCATGGGTTCTGGAC	EXP	<i>TAL1</i> enhancer
CAGTCTAACTCATGCTGCTC	EXP	<i>TAL1</i> enhancer
CAAGGGAGGAACTGAATTAA	EXP	<i>TAL1</i> enhancer
TCTAACAGGTCCAGCCTGTC	EXP	<i>TAL1</i> enhancer
AGTTTATCTGTAATAGGAAT	EXP	<i>TAL1</i> enhancer
ATTCCTTCAACAATGTTGC	EXP	<i>TAL1</i> enhancer
AGCTGGTCTATCAGCTATAT	EXP	<i>TAL1</i> enhancer
AATGAAAAGAAGGGTAGAAA	EXP	<i>TAL1</i> enhancer
CTATATGGGACTGGGGAGAA	EXP	<i>TAL1</i> enhancer
AACTCATGCTGCTCAGGGCC	EXP	<i>TAL1</i> enhancer
TACACTGGATGCAGGGGATT	EXP	<i>TAL1</i> enhancer
TTTCAGCTGTCTCAGTAAAT	EXP	<i>TAL1</i> enhancer

Target-specific guide sequence	Type	Target
ATAGGCTGTGTCTGTGGCA	EXP	<i>TAL1</i> enhancer
TGTGGCATGGGTTCTGGACT	EXP	<i>TAL1</i> enhancer
AGCTTCTGCCAGAGTCCAGC	EXP	<i>TAL1</i> enhancer
AAATGAAAAGAAGGGTAGAA	EXP	<i>TAL1</i> enhancer
ATCGGCATGCCTTGCTTCTA	EXP	<i>TAL1</i> enhancer
CCATTAGAAATTAGCTTTGC	EXP	<i>TAL1</i> enhancer
TCTGGACTGGGCACACGGGC	EXP	<i>TAL1</i> enhancer
CCTGCAGTTACGCTGCGGTG	PC	<i>TAL1</i> proximal regulatory elements
AGGCCTCTCAGCGAAAAAGG	PC	<i>TAL1</i> proximal regulatory elements
ATCTCCACTCCGCCGAAAG	PC	<i>TAL1</i> proximal regulatory elements
GAGGAAGAGGATGCACACCC	PC	<i>TAL1</i> proximal regulatory elements
CCAGACCGATCCCAGTTGGA	PC	<i>TAL1</i> proximal regulatory elements
ACTATCCCTTCGCGGTGTAG	PC	<i>TAL1</i> proximal regulatory elements
ACTATTGCCTTTCCCAACA	PC	<i>TAL1</i> proximal regulatory elements
TAAGCAGGGAGGTGTCTACG	PC	<i>TAL1</i> proximal regulatory elements
AATAGTCTTCAGACTCTGGT	PC	<i>TAL1</i> proximal regulatory elements
GAGTGCGTTCATGAGAACTG	PC	<i>TAL1</i> proximal regulatory elements
CCGAGTGTGGTGTGCCTGCG	PC	<i>TAL1</i> proximal regulatory elements
GCGTTGGCTGCTTCTAAGTG	PC	<i>TAL1</i> proximal regulatory elements
AAACGCAGAAGGGCCTCGAA	PC	<i>TAL1</i> proximal regulatory elements
TGCGTACGATTGTGCTCCGT	PC	<i>TAL1</i> proximal regulatory elements
CATTTGGCCATAATGGCCG	PC	<i>TAL1</i> proximal regulatory elements
TGAGTGGGATTACAGCGCGT	PC	<i>TAL1</i> proximal regulatory elements
CAGAGATAAGGCACTGCCGC	PC	<i>TAL1</i> proximal regulatory elements
GTTCCAGGCCTCGTTAGCAT	PC	<i>TAL1</i> proximal regulatory elements
CCACACCGCAGCGTAACTGC	PC	<i>TAL1</i> proximal regulatory elements
GTGGGATTACAGCGCGTCGG	PC	<i>TAL1</i> proximal regulatory elements
GCGGCCCTTTAAGTCTCTCG	PC	<i>TAL1</i>

Target-specific guide sequence	Type	Target
TATGAGATGGAGATTACTGA	PC	TAL1
GCGCCCAGTTCGATGACTGG	PC	TAL1
GTTGGTGGTGAACATAGGGA	PC	TAL1
GATGTGTGGGGATCAGCTTG	PC	TAL1
CGCGCCCGCCTCGGTTACAG	PC	TAL1
GGTCTGCACAGCTCGGTGGT	PC	TAL1
GTGACCCCCAGCTAGAGGGA	PC	TAL1
TGAGGCTGTAGAGCAGCGCG	PC	TAL1
AGACTTAAAGGGCCGCGACG	PC	TAL1
AGTCTGGAAGCGTCACTTG	PC	MYB
ACCAGGCACACAAGAGACTG	PC	MYB
TATTTACATGTAACGCTACA	PC	MYB
CGTCGGAAGGTGGAACAGGA	PC	MYB
ACCCGGGGTAGCTGCATGTG	PC	MYB
AGAAATACGGTCCGAAACGT	PC	MYB
CTCCACTCCATCTCTGCCAG	PC	MYB
AGTTATTGCCAATTATCTCC	PC	MYB
GACGCATTGTAGAATTCCAG	PC	MYB
ACTGGAATTCTACAATGCGT	PC	MYB
TCCAAGACGTCCATCCACCA	PC	GATA3
CAGGGAGTGTGTGAACTGTG	PC	GATA3
GGAGCTGTACTCGGGCACGT	PC	GATA3
GGTAGGGATCCATGAAGCAG	PC	GATA3
CTGGAGTCGTCCCACTCCCG	PC	GATA3
AGCCTGGGGTGGACAGCGAT	PC	GATA3
AGGCCCCGTCCAGCACAGGT	PC	GATA3
GCCATCTCGCCGCCACAGTG	PC	GATA3
GGGCAACCTCGACCCCACTG	PC	GATA3

Target-specific guide sequence	Type	Target
CGACGAGGAGGCTCCACCCA	PC	GATA3
CCAATGTCCAGCTTTCATCG	PC	HEB
GTGCTAGCGAACATACTGGT	PC	HEB
AGTCGATTAGGAGCCCATGA	PC	HEB
CTGGGAAGCAGTCAATTCAG	PC	HEB
TGTGCTGCGGAACCATGCTG	PC	HEB
AGAGGTGAAGGTGATCCAAC	PC	HEB
ACTACTGGATTGAGACATGT	PC	HEB
GTCTCCTGAGACAAGATCTG	PC	HEB
CTCCATGACTCTGCAGCGCT	PC	HEB
CCGCGATGAAAGCTGGACAT	PC	HEB
CTGATCGTAGGACCACGGTG	PC	RUNX1
TGCTCCCCACAATAGGACAT	PC	RUNX1
GGATGTTCCAGATGGCACTC	PC	RUNX1
GGTCATTAAATCTTGCAACC	PC	RUNX1
GAGTGGTTCAGGGAGGCACG	PC	RUNX1
GAGCCCAGGCAAGATGAGCG	PC	RUNX1
GAGGTGCTGGCCGACCACCC	PC	RUNX1
CCAGCAACGCCCATTTCACC	PC	RUNX1
GACTGATCGTAGGACCACGG	PC	RUNX1
GTATTGGTAGGACTGATCGT	PC	RUNX1
TAATTAATCAGGCTTCACAA	PC	CBP
TGGCAACTGGACGTTCCCCA	PC	CBP
ATTGCCCCCTCCAAACACG	PC	CBP
CTTAGCCCACTGATGAACGA	PC	CBP
TCCAGCACACGACACCACCT	PC	CBP
CAGGACGGTACTTACGTCTG	PC	CBP
TTTGTCGTGAAGATGCACAA	PC	CBP

Target-specific guide sequence	Type	Target
ATGATGGAGGAGGATTTGCA	PC	CBP
TGCAGCCAGTGAAACCACTG	PC	CBP
CTGGACAGAGTGGTTCATTG	PC	CBP
TTCACGTGCCAGTACAGGGT	PC	Tet-transactivator protein
AGAGAAACAGTACGAAACCC	PC	Tet-transactivator protein
TAGCCCCGTCGCGATGTGAG	PC	Tet-transactivator protein
GCCAGTGTAAGTGGCCCA	PC	Tet-transactivator protein
TGCCTGTCCAGCATCTCGAT	PC	Tet-transactivator protein
GCCATGACTCGCCTTCCAGG	PC	Tet-transactivator protein
GTATCGAAGGCCTGACGACA	PC	Tet-transactivator protein
CAGTGTAAGTGGCCACGG	PC	Tet-transactivator protein
CTGTACGCTCTGTCCGCGT	PC	Tet-transactivator protein
GACGGGGCTAAAGTGCATCT	PC	Tet-transactivator protein
TATTTACTATGATAGGGTGA	NC	safe harbour GSH2
TATTTAAGCCACTAGGTGCT	NC	safe harbour GSH2
GTACTTAAGCCTCTAGAAAC	NC	safe harbour GSH2
CTGGCATTATTTACTATGAT	NC	safe harbour GSH2
TGGCATTATTTACTATGATA	NC	safe harbour GSH2
TCTTATACCTTCTCTCATC	NC	safe harbour GSH2
CTGCATTGTCCAGTTTCTAG	NC	safe harbour GSH2
GTATTTAAGCCACTAGGTGC	NC	safe harbour GSH2
AGAACTCTGAGTGTTACGGA	NC	safe harbour GSH2
TGGACATGAAATGAATACCC	NC	safe harbour GSH2
GCCCTGAAAGAACCAACTTG	NC	safe harbour GSH2
ATATCGTGCTAATGTCAGTG	NC	safe harbour GSH2
CATATCGTGCTAATGTCAGT	NC	safe harbour GSH2
ATCCACGTCCTGGTTATCAA	NC	safe harbour GSH2
CAGAAGTCGCAACCCCAAGT	NC	safe harbour GSH2

Target-specific guide sequence	Type	Target
CAATTCTAGAATCCACGTCC	NC	safe harbour GSH2
TCCACGTCTGGTTATCAAA	NC	safe harbour GSH2
TAAAGGGGATGAGCAAGCCA	NC	safe harbour GSH2
GCGACTTCTGAGTCATCAGA	NC	safe harbour GSH2
CATCAGAAGGATAGCCAGAG	NC	safe harbour GSH2
TGTCTCCAGAATGATGATGA	NC	safe harbour GSH2
GCCCTTTGATAACCAGGACG	NC	safe harbour GSH2
CTCCAGAATGATGATGAGGG	NC	safe harbour GSH2
ACAAAGATGCAGACATGTAC	NC	safe harbour GSH2
AGATGAACACATTGTGTGAA	NC	safe harbour GSH2
TTGTCTCCAGAATGATGATG	NC	safe harbour GSH2
GGGCCCTGAAAGAACCAACT	NC	safe harbour GSH2
GGCCCTGAAAGAACCAACTT	NC	safe harbour GSH2
GGCACTGCCCTTTGATAACC	NC	safe harbour GSH2
CAGGACGTGGATTCTAGAAT	NC	safe harbour GSH2
TATAAGAACTCTGAGTGTTA	NC	safe harbour GSH2
CTCACTCTAAGGTAAGTAAC	NC	safe harbour GSH2
GAATAAAGGAGAATCATAAA	NC	safe harbour GSH2
CAATGTGTTTCATCTCCTCTC	NC	safe harbour GSH2
ACCCAGGTACAGAAGAATAA	NC	safe harbour GSH2
TCCTTTATTCTTCTGTACCT	NC	safe harbour GSH2
ACCCCAAGTTGGTTCTTTCA	NC	safe harbour GSH2
GTCCACCCTCATCATCATTC	NC	safe harbour GSH2
AGTTGGTTCTTTCAGGGCCC	NC	safe harbour GSH2
CTCCTTTATTCTTCTGTACC	NC	safe harbour GSH2
GAAACTGCACTCCCAATGAC	NC	safe harbour GSH2
GGCTCTGGCATCTCACTCTA	NC	safe harbour GSH2
GCACTCCCAATGACAGGCTC	NC	safe harbour GSH2

Target-specific guide sequence	Type	Target
AGATGCCAGAGCCTGTCAAT	NC	safe harbour GSH2
CTAATGTCAGTGGGGTGAGA	NC	safe harbour GSH2
ACAGATACAGTGTATTAAAA	NC	safe harbour GSH2
TTATGTTAAGTATCCTACAG	NC	safe harbour GSH2
TGAATAGGAGAACCCACTGT	NC	safe harbour GSH2
ATTTCTATCACTGATATCCA	NC	safe harbour GSH2
GATTTGTCATTGAAATTCCC	NC	safe harbour GSH2
AATTTCTATCACTGATATCC	NC	safe harbour GSH2
TTTGATACAAGTGAAAGTCA	NC	safe harbour GSH2
GACTATCCAAAGTCAGACAT	NC	safe harbour GSH2
GAGGAACTTTGTCTTAAGTG	NC	safe harbour GSH2
CATCAGACTTGATAGCACTG	NC	safe harbour GSH2
TTGATACAAGTGAAAGTCAT	NC	safe harbour GSH2
AAAGCAGAGTCATCCCTCCT	NC	safe harbour GSH2
TTCAAGGATGAGCTTGGATG	NC	safe harbour GSH2
GAGGGCTTCAAGGATGAGCT	NC	safe harbour GSH2
AGGAACTTTGTCTTAAGTGA	NC	safe harbour GSH2
GAGCAGGCAGAAGGGATGGG	NC	safe harbour GSH2
CTTCAAGGATGAGCTTGGAT	NC	safe harbour GSH2
GGTTCAGACCATGCCAGAGC	NC	safe harbour GSH2
ACCATGCCAGAGCAGGCAGA	NC	safe harbour GSH2
AAGGATGAGCTTGGATGGGG	NC	safe harbour GSH2
TGTCTTAAGTGAGGGCTTCA	NC	safe harbour GSH2
CCAGAGCAGGCAGAAGGGAT	NC	safe harbour GSH2
GCTTCAAGGATGAGCTTGGGA	NC	safe harbour GSH2
CCCTTCTGCCTGCTCTGGCA	NC	safe harbour GSH2
CCCATCCCTTCTGCCTGCTC	NC	safe harbour GSH2
GAGAACCTAAAACCCTGCTT	NC	safe harbour GSH2

Target-specific guide sequence	Type	Target
ACTGCTAGGAATGCCGAAGC	NC	safe harbour GSH2
AGAACAAATGAAACCAGGCA	NC	safe harbour GSH2
CAAAAAGAACAAATGAAACC	NC	safe harbour GSH2
AACTGAACTGTGACACTGCT	NC	safe harbour GSH2
CAGGGCAATTAATAGGCAAA	NC	safe harbour GSH2
AACCAGGCAGGGCAATTAAT	NC	safe harbour GSH2
TGCCTATTAATTGCCCTGCC	NC	safe harbour GSH2
TCATATCGTGCTAATGTCAG	NC	safe harbour GSH2
TCCAAGACATTCCTTTGAC	NC	safe harbour GSH2
TCTGAGATATTCATCAGTTT	NC	safe harbour GSH2
CAGCCTAACGGTGTTTAAGG	NC	safe harbour GSH2
CTCAGGTAAAATACCACCAT	NC	safe harbour GSH2
CATAAAGGAGATACAACCTCT	NC	safe harbour GSH2
CAGTGATACATATCACACAA	NC	safe harbour GSH2
ACATAAAGGAGATACAACCTC	NC	safe harbour GSH2
ACTCAGGTAAAATACCACCA	NC	safe harbour GSH2
AGTTTGATTTCAGTGGTGA	NC	safe harbour GSH2
TTTCCTCCTTAAACACCGTT	NC	safe harbour GSH2
TTAAACACCGTTAGGCTGCT	NC	safe harbour GSH2
GAATCAAGTTTGATTTCAG	NC	safe harbour GSH2
AAACTTGATTCAAGTTACTC	NC	safe harbour GSH2
AAAAAGGATGCCACCCATGG	NC	safe harbour GSH2
AAGCAGCCTAACGGTGTTTA	NC	safe harbour GSH2
AAGCTGACTGGGAAAGAAGT	NC	safe harbour GSH2
TGGGAAAGAAGTGGGCATGA	NC	safe harbour GSH2
TAAGCTGACTGGGAAAGAAG	NC	safe harbour GSH2
ACCTGTCAAAGGAATGTCTT	NC	safe harbour GSH2
AATGTTGTAGTCTTCCAGAA	NC	safe harbour GSH2

Target-specific guide sequence	Type	Target
TCAGCACTACAGAGACAATC	NC	safe harbour GSH2
CTCTTTAGAAATGCTATTTG	NC	safe harbour GSH2
CACCTGTCAAAGGAATGTCT	NC	safe harbour GSH2
CAAGTGGTGATGGTTTCGTT	NC	safe harbour GSH2
CAATTATAAATAAGCTGACT	NC	safe harbour GSH2
GTCATGTGAAATTTGTGAAT	NC	safe harbour GSH2
TTTGGCACAACCTGAAATAT	NC	safe harbour GSH2
AAGTTGTGCCAAAGAGATTT	NC	safe harbour GSH2
GGTTTGATCCAAAATCTCTT	NC	safe harbour GSH2
TCATTGAGCTCTAAATATTT	NC	safe harbour GSH2
CTTTGGCACAACCTGAAATA	NC	safe harbour GSH2
CTGCTAGGAATGCCGAAGCA	NC	safe harbour GSH2
ACTATCCAAAGTCAGACATG	NC	safe harbour GSH2
AAAAAAAAGGATGCCACCCA	NC	safe harbour GSH2
TATGTTAAGTATCCTACAGT	NC	safe harbour GSH2
TCAAAATAAACACCTGTCAA	NC	safe harbour GSH2
TCAATATAACATTGTGTCAA	NC	safe harbour GSH2
AGACAATCTGGAGAAAACAG	NC	safe harbour GSH2
GAGCTCAATGAGTCTAACT	NC	safe harbour GSH2
GGGGGTATCTATCTGAACTG	NC	silent site within TAL1 TAD
CTGGAAGAATGGTCCGAACA	NC	silent site within TAL1 TAD
GATGACGATGAGTCCCAGTG	NC	silent site within TAL1 TAD
ACACCACACAATGAGCTGAT	NC	silent site within TAL1 TAD
TCTAGTTTAAAAAATAGCAC	NC	silent site within TAL1 TAD
TATTCAAAGAGTACAATCA	NC	silent site within TAL1 TAD
GTCCTAGAGGAGTGCCTCCA	NC	silent site within TAL1 TAD
ATGACACGGTCCAGGATAGG	NC	silent site within TAL1 TAD
ATGCCCTGAAATGTGAGGTG	NC	silent site within TAL1 TAD

Target-specific guide sequence	Type	Target
TAGGGTAAGACTACAAGCAT	NC	silent site within TAL1 TAD
TGCAAAACAGCCTGTCCACAC	NC	silent site within TAL1 TAD
TTTAGGCAAAGATGCGATGT	NC	silent site within TAL1 TAD
CCTGTGAGCTCCCTTTGGCA	NC	silent site within TAL1 TAD
AGGAATGCCAGACCAAAACA	NC	silent site within TAL1 TAD
CACAGGAAGACACCTACGAA	NC	silent site within TAL1 TAD
AACTGGGACTCATCGTCAT	NC	silent site within TAL1 TAD
CCGTGCCAAAGGGAGCTCAC	NC	silent site within TAL1 TAD
TCATTTAACACACCTTTCGT	NC	silent site within TAL1 TAD
CAGTGCACAGAGAGAATTGG	NC	silent site within TAL1 TAD
CCCTCCCCATCCATTTAAGC	NC	silent site within TAL1 TAD
CCTGCTTAAATGGATGGGGA	NC	silent site within TAL1 TAD
AATGGTCCGAACAGGGAAGT	NC	silent site within TAL1 TAD
TGGTACTTTAGAACAGTGCT	NC	silent site within TAL1 TAD
TTTCTCTCTCTAGGTCCTAG	NC	silent site within TAL1 TAD
CAGCTTATGGATGGTGACTG	NC	silent site within TAL1 TAD
TCCACATGCCCTGAAATGTG	NC	silent site within TAL1 TAD
GAGAAAATCTGCAACAACTG	NC	silent site within TAL1 TAD
CGCATCTCTTAGGTCCTGCA	NC	silent site within TAL1 TAD
GATGACACGGTCCAGGATAG	NC	silent site within TAL1 TAD
AACACCTGCTTAAATGGATG	NC	silent site within TAL1 TAD
AGCTCCACAGCCGTGCCAAA	NC	silent site within TAL1 TAD
GGAATTCAACAGCTTATGGA	NC	silent site within TAL1 TAD
CAGCACCCGGAGGACATCTC	NC	silent site within TAL1 TAD
GGGATTGAGGAGAGGGCATG	NC	silent site within TAL1 TAD
AACTAGAAGAGCAGCAAGCT	NC	silent site within TAL1 TAD
ATCTGCAACAACTGGGGTTG	NC	silent site within TAL1 TAD
ACTGTCTGAATCAGAAGACG	NC	silent site within TAL1 TAD

Target-specific guide sequence	Type	Target
GCTTGTAGTCTTACCCTAAG	NC	silent site within TAL1 TAD
TTTAAAAATAGCACCGGCC	NC	silent site within TAL1 TAD
AGAGATAATACAGATAATAC	NC	silent site within TAL1 TAD
ACTAACCTGTGAGCTCCCTT	NC	silent site within TAL1 TAD
ATTTAAGCAGGTGTTGCTTA	NC	silent site within TAL1 TAD
CTGCTTAAATGGATGGGGAG	NC	silent site within TAL1 TAD
CGATGAGTCCCAGTGTGGAC	NC	silent site within TAL1 TAD
AAGGGAGCTCACAGGTTAGT	NC	silent site within TAL1 TAD
TAAGCAGGTGTTGCTTAGGG	NC	silent site within TAL1 TAD
AAGCAGGTGTTGCTTAGGGC	NC	silent site within TAL1 TAD
TGATGATGACCGCATCTCTT	NC	silent site within TAL1 TAD
CACACCACACAATGAGCTGA	NC	silent site within TAL1 TAD
CCTCCAGGTTGGGAGGAGCA	NC	silent site within TAL1 TAD
TTTGGTCTGGCATTCTGGG	NC	silent site within TAL1 TAD
CATTGTCTTCTAGACCACTT	NC	silent site within TAL1 TAD
GCTGGAAGAATGGTCCGAAC	NC	silent site within TAL1 TAD
CTGGACCGTGTCCATCCCTGC	NC	silent site within TAL1 TAD
AGCCGTGGAGGCACTCCTCT	NC	silent site within TAL1 TAD
CGGCAGGGATGACACGGTCC	NC	silent site within TAL1 TAD
AGAGCAGCAAGCTGGGATTG	NC	silent site within TAL1 TAD
AGAGAAAATCTGCAACAACT	NC	silent site within TAL1 TAD
GCAGTGAGCAGGCTGTGGAA	NC	silent site within TAL1 TAD
GCAACACCTGCTTAAATGGA	NC	silent site within TAL1 TAD
GACCGTGTCCATCCCTGCCGG	NC	silent site within TAL1 TAD
GGCCTCCGGCAGGGATGACA	NC	silent site within TAL1 TAD
TTTCAGGGCATGTGGAGTCA	NC	silent site within TAL1 TAD
ATTTAAGGGGTACAAGTGGC	NC	silent site within TAL1 TAD
ACCTGCTTAAATGGATGGGG	NC	silent site within TAL1 TAD

Target-specific guide sequence	Type	Target
CCCTGCTCCTCCCAACCTGG	NC	silent site within TAL1 TAD
TCCCAGCACAGATCCCTCCC	NC	silent site within TAL1 TAD
ATTGTCTTCTAGACCACTTA	NC	silent site within TAL1 TAD
TGTGGAAGGGATCCAGGAGA	NC	silent site within TAL1 TAD
TAGCTCCACAGCCGTGCCAA	NC	silent site within TAL1 TAD
ACAATGAGCTGATGGGAGAT	NC	silent site within TAL1 TAD
GCAAACAGCCTGTCCACACT	NC	silent site within TAL1 TAD
GGTGA CTGAGGCCTCCGGCA	NC	silent site within TAL1 TAD
AGAAGAGAGCACCTCCAGGT	NC	silent site within TAL1 TAD
AGCAAGCTGGGATTGAGGAG	NC	silent site within TAL1 TAD
TCCTGGGAGGGATCTGTGCT	NC	silent site within TAL1 TAD
TGCGTCCTGAGATGTCCTCC	NC	silent site within TAL1 TAD
GAGCTGATGGGAGATGGGCC	NC	silent site within TAL1 TAD
GGGATCTGTGCTGGGAATAG	NC	silent site within TAL1 TAD
TTCAGACAGTGCTTGAACT	NC	silent site within TAL1 TAD
CACAATGAGCTGATGGGAGA	NC	silent site within TAL1 TAD
AGGATGGAAGAGCAATTTAA	NC	silent site within TAL1 TAD
CATTTAAGCAGGTGTTGCTT	NC	silent site within TAL1 TAD
AATCTGCAACA ACTGGGGTT	NC	silent site within TAL1 TAD
TGTGGGAAAGTCTTGCA TTT	NC	silent site within TAL1 TAD
GCAAGCTGGGATTGAGGAGA	NC	silent site within TAL1 TAD
GACAGGCTGTTTG CAGCACC	NC	silent site within TAL1 TAD
AGAGGTCTCGAGAGCCTTGC	NC	silent site within TAL1 TAD
AGCTTCTCATCCCCCTATCC	NC	silent site within TAL1 TAD
GGGATGACACGGTCCAGGAT	NC	silent site within TAL1 TAD
CAGGACGCAGCTCCATGTTT	NC	silent site within TAL1 TAD
CAACACCTGCTTAAATGGAT	NC	silent site within TAL1 TAD
GAATGGTCCGAACAGGGAAG	NC	silent site within TAL1 TAD

Target-specific guide sequence	Type	Target
GAGAGCACCTCCAGGTTGGG	NC	silent site within TAL1 TAD
TATTATCTGTATTATCTCTA	NC	silent site within TAL1 TAD
GGATGACACGGTCCAGGATA	NC	silent site within TAL1 TAD
TGGCCACACCTCACATTTCA	NC	silent site within TAL1 TAD
TCACTGCCCACTTCCTGTT	NC	silent site within TAL1 TAD
AGCTCCCTTTGGCACGGCTG	NC	silent site within TAL1 TAD
CTGCGTCCTGAGATGTCCTC	NC	silent site within TAL1 TAD
AAATATCTTGCTTGTCCTCC	NC	silent site within TAL1 TAD
AAACAGTGACAGAGAGAAT	NC	silent site within TAL1 TAD
GGGACATGTGGGTGCAGCCG	NC	silent site within TAL1 TAD
AAATCTGCAACAACCTGGGGT	NC	silent site within TAL1 TAD
GCAGGCTGTGGAAGGGATCC	NC	silent site within TAL1 TAD
ATAGCACCGGCCAGGTGCGG	NC	silent site within TAL1 TAD
TTGGCACGGCTGTGGAGCTA	NC	silent site within TAL1 TAD
TTGGTCTGGCATTCTGGGA	NC	silent site within TAL1 TAD
TTCCTGGGAGGGATCTGTGC	NC	silent site within TAL1 TAD
TGGTGACTGAGGCCTCCGGC	NC	silent site within TAL1 TAD
AAGCTCTCTCCAACCTGCAGC	NC	silent site within TAL1 TAD
AGCCATCACCATCCTTCTCC	NC	silent site within TAL1 TAD
GGCAGTGAGCAGGCTGTGGA	NC	silent site within TAL1 TAD
TCCAACCTGCAGCTGGAAGAA	NC	silent site within TAL1 TAD
GAGAGAAAATCTGCAACAAC	NC	silent site within TAL1 TAD
ACCTCCAGGTTGGGAGGAGC	NC	silent site within TAL1 TAD
TGATGAAAAGAGTCACTGGG	NC	silent site within TAL1 TAD
ACAGAGATGACCCTGAAAGG	NC	silent site within TAL1 TAD
AAAGGCAAAATTAGTAACAA	NC	silent site within TAL1 TAD
GACGTACATTTGGGACATGT	NC	silent site within TAL1 TAD
CCAAGTATTTAACAGACTAA	NC	silent site within TAL1 TAD

Target-specific guide sequence	Type	Target
GGATGATGAAAAGAGTCACT	NC	silent site within TAL1 TAD
AGAAACCTTGAGCTTCCAGA	NC	silent site within TAL1 TAD
GTTTCTATTACATACCTCT	NC	silent site within TAL1 TAD
ATGACAGAGATGACCCTGAA	NC	silent site within TAL1 TAD
CTGAAGAGTAGCAGTCTCAG	NC	silent site within TAL1 TAD
TCTGTCTCCAGATATCTGCA	NC	silent site within TAL1 TAD
AGAAAGGAATGTGAGTCTGG	NC	silent site within TAL1 TAD
TAAAGGCAAAATTAGTAACA	NC	silent site within TAL1 TAD
GAGACTGCTACTCTTCAGGC	NC	silent site within TAL1 TAD
ATCCTCAGCTCCATGCTCAG	NC	silent site within TAL1 TAD
TTGAGCTTCCAGAAGGAGGT	NC	silent site within TAL1 TAD
AACCTTGAGCTTCCAGAAGG	NC	silent site within TAL1 TAD
AGGGGCAGGATTCACGAGTT	NC	silent site within TAL1 TAD
TGGAGAAAGGAATGTGAGTC	NC	silent site within TAL1 TAD
CAGGATTCACGAGTTTGGTT	NC	silent site within TAL1 TAD
GGAGGTCAGAACTGAGGTCC	NC	silent site within TAL1 TAD
TCCTCAGCTCCATGCTCAGA	NC	silent site within TAL1 TAD
GCCCTCTGAGCATGGAGCTG	NC	silent site within TAL1 TAD
TTAGTCTGTAAATACTTGG	NC	silent site within TAL1 TAD
ATTCACATACCTCTGGGAAT	NC	silent site within TAL1 TAD
GGTTTCTATTACATACCTC	NC	silent site within TAL1 TAD
GGGAAGCTCACCACTTTCA	NC	silent site within TAL1 TAD
TATTCACATACCTCTGGGAA	NC	silent site within TAL1 TAD
TGCTCTGTGCCCTCTGAGCA	NC	silent site within TAL1 TAD
CACAGTTTCTCTCTGATC	NC	silent site within TAL1 TAD
AGTCACTGGGAGGCTGGATT	NC	silent site within TAL1 TAD
CTCTGTCTCCAGATATCTGC	NC	silent site within TAL1 TAD
CTCTGAGACTGCTACTCTTC	NC	silent site within TAL1 TAD

Target-specific guide sequence	Type	Target
AGCAAGGGGTGTCCAATGTG	NC	silent site within TAL1 TAD
TAGCCCAAAAGTTGAACCAG	NC	silent site within TAL1 TAD
GTGGCTTCTTGACTGAGCAA	NC	silent site within TAL1 TAD
AGAAGCCACATCCAAACCCT	NC	silent site within TAL1 TAD
CCAAGTGTGAAAGGGTCCGA	NC	silent site within TAL1 TAD
TCCAAGTGTGAAAGGGTCCG	NC	silent site within TAL1 TAD
GGACAGATTCTCTGGCAAAG	NC	silent site within TAL1 TAD
AAGATTATTTAGTTGAGGGG	NC	silent site within TAL1 TAD
TAATAAAGATTATTTAGTTG	NC	silent site within TAL1 TAD
TGGCTTCTTGACTGAGCAAG	NC	silent site within TAL1 TAD
ATAAAGATTATTTAGTTGAG	NC	silent site within TAL1 TAD
AGATTATTTAGTTGAGGGGA	NC	silent site within TAL1 TAD
ACAAAAGCTCCAAGTGTGAA	NC	silent site within TAL1 TAD
TGTGGCTTCTTGACTGAGCA	NC	silent site within TAL1 TAD
AGGGTCCGAGGGTTTGATG	NC	silent site within TAL1 TAD
AATACATAAATGGGTGATCT	NC	silent site within TAL1 TAD
AATAAAGATTATTTAGTTGA	NC	silent site within TAL1 TAD
AATCTTTATTATAGAACTC	NC	silent site within TAL1 TAD
TGTGAAAGGGTCCGAGGGTT	NC	silent site within TAL1 TAD
GACAGATTCTCTGGCAAAGT	NC	silent site within TAL1 TAD
CCCTCGGACCTTTCACACT	NC	silent site within TAL1 TAD
AGGGGTGTCCAATGTGTGGC	NC	silent site within TAL1 TAD
TAATACATAAATGGGTGATC	NC	silent site within TAL1 TAD
TGCAATTATAATTTATAGA	NC	silent site within TAL1 TAD
TGAAGATGACTTTAAAAAGC	NC	silent site within TAL1 TAD
TGGGCAGTGAGAGGTGTTGC	NC	silent site within TAL1 TAD
TCGAGCGTTGATCGACAACG	NC	non-targeting (Canver et al. 2017)
AGTCGCTTTATACGCCGCTT	NC	non-targeting (Canver et al. 2017)

Target-specific guide sequence	Type	Target
CAGGGCATTACGCGAGTCGG	NC	non-targeting (Canver et al. 2017)
CGCCGCTCGCTATCGTTATT	NC	non-targeting (Canver et al. 2017)
TCTGCGTGAGCGTAATACGC	NC	non-targeting (Canver et al. 2017)
CCGATATCGCGATCGTCGGT	NC	non-targeting (Canver et al. 2017)
ATTCGACCCCGTAGCGCGGG	NC	non-targeting (Canver et al. 2017)
TAATGTCGAACCGACGGGTC	NC	non-targeting (Canver et al. 2017)
CGTCGATAGGGTCCGATTCTG	NC	non-targeting (Canver et al. 2017)
CTCGTAGGCCTTTATCGCGC	NC	non-targeting (Canver et al. 2017)
CGCGGATGCTCGTACTCCGA	NC	non-targeting (Canver et al. 2017)
CACCGGCCGATCGGGTCGAT	NC	non-targeting (Canver et al. 2017)
ACGACTACCAGCGGCGGTAA	NC	non-targeting (Canver et al. 2017)
CGCCGATATATGGCGGTCAA	NC	non-targeting (Canver et al. 2017)
TCATAGGCCTTTCGCCGGCG	NC	non-targeting (Canver et al. 2017)
ACGGGCTCGTGTCGCAATAG	NC	non-targeting (Canver et al. 2017)
GTGACATTCCGTTGCGCGAG	NC	non-targeting (Canver et al. 2017)
TGAGCTTCATCGGGCGCGAC	NC	non-targeting (Canver et al. 2017)
TGCGGAATGCATTGACCGC	NC	non-targeting (Canver et al. 2017)
CGCGCGCCGGGTAAGTTAGC	NC	non-targeting (Canver et al. 2017)
GTTGCGTTAACGCTTAACGC	NC	non-targeting (Canver et al. 2017)
TTTCGCCGACGCAACACGTA	NC	non-targeting (Canver et al. 2017)
GGTTTGCGGTCCGTTACCCG	NC	non-targeting (Canver et al. 2017)
ATTCGCGGCGGCGGATAACC	NC	non-targeting (Canver et al. 2017)
TTGCAGCAGTTACGGCGCGC	NC	non-targeting (Canver et al. 2017)
GTAGACGTCGCCTGCGACGT	NC	non-targeting (Canver et al. 2017)
CGTGCCGTCCGTTACTGATAC	NC	non-targeting (Canver et al. 2017)
TCCATGGCGCGTTCCGATAC	NC	non-targeting (Canver et al. 2017)
CAGCAATCTACGCTCGCGAC	NC	non-targeting (Canver et al. 2017)
GTATCAGGCCTGTACGACGG	NC	non-targeting (Canver et al. 2017)

Target-specific guide sequence	Type	Target
ACCCATCGTCGGCGACCGGA	NC	non-targeting (Canver et al. 2017)
GTTCACCCCCGGCGAACGTT	NC	non-targeting (Canver et al. 2017)
TTCAGCGTAGCCCGAACCGT	NC	non-targeting (Canver et al. 2017)
CGCCCCGGTAAACTCGATCA	NC	non-targeting (Canver et al. 2017)
CCCGCGAATTTAGTGCCGAA	NC	non-targeting (Canver et al. 2017)
CAACTCCGTTACGCCGAGTC	NC	non-targeting (Canver et al. 2017)
AGCGCGCGCGAAACACTACT	NC	non-targeting (Canver et al. 2017)
AGCGACGTACCGGACGCTAA	NC	non-targeting (Canver et al. 2017)
CCACCGTTGTTCCCGCGTAT	NC	non-targeting (Canver et al. 2017)
TCTTGTAAGATCCCGCGATT	NC	non-targeting (Canver et al. 2017)
CCAACCGTTGGGTCGACGTT	NC	non-targeting (Canver et al. 2017)
TTGGCCTATCGTATCGCGGC	NC	non-targeting (Canver et al. 2017)
ATCGATATTTGACGCTAGTC	NC	non-targeting (Canver et al. 2017)
CGCCTAACTACTACGGCGGG	NC	non-targeting (Canver et al. 2017)
GGGCATCGCAGTAATACGCG	NC	non-targeting (Canver et al. 2017)
TCGTCGAGCTGGTTATCGAC	NC	non-targeting (Canver et al. 2017)
CCTTCGACTAGTGACGCCGG	NC	non-targeting (Canver et al. 2017)
CCGCTAGATGTTCAACGCGC	NC	non-targeting (Canver et al. 2017)
CGTTATGCGCGTATACCGCC	NC	non-targeting (Canver et al. 2017)
TCGATAGTGCGCCGATTAC	NC	non-targeting (Canver et al. 2017)
CGCCGATGAGAATAGCGTCG	NC	non-targeting (Canver et al. 2017)
AAAAGCGCTCGGTACGGATC	NC	non-targeting (Canver et al. 2017)
TTTCGTCCAGTCGTGCCGTA	NC	non-targeting (Canver et al. 2017)
TTATAGGTCGGCGAGTCGAC	NC	non-targeting (Canver et al. 2017)
TAACTACGAGGACGAATCGT	NC	non-targeting (Canver et al. 2017)
CCTCCGGCGCACCGCGAAAC	NC	non-targeting (Canver et al. 2017)
ATCGCGCGTCCTTCGGCAAT	NC	non-targeting (Canver et al. 2017)
GTAGTGTGCGCCCGACGGCA	NC	non-targeting (Canver et al. 2017)

Target-specific guide sequence	Type	Target
AGGTACCCGCCTGACGCGAT	NC	non-targeting (Canver et al. 2017)
CATTGCGTACATGCGTAGCG	NC	non-targeting (Canver et al. 2017)
ACGTCGCGATTGACGGTTCC	NC	non-targeting (Canver et al. 2017)
TGTCGGCAAATGCGACCGTT	NC	non-targeting (Canver et al. 2017)
ACCCGCGGCATATTCGCCGT	NC	non-targeting (Canver et al. 2017)
ACCGACCGCGACTCACGTCG	NC	non-targeting (Canver et al. 2017)
CGATCTCGGACGATTACGGC	NC	non-targeting (Canver et al. 2017)
CAACGCAACACGTCCGCGGC	NC	non-targeting (Canver et al. 2017)
AACCTCGCAGAGTCGACTCG	NC	non-targeting (Canver et al. 2017)
GCCGTAGGCATACGCGATAG	NC	non-targeting (Canver et al. 2017)
CTACTCGACTCCGGTCGGGT	NC	non-targeting (Canver et al. 2017)
CCGAGTTACTTGCGCGAGTT	NC	non-targeting (Canver et al. 2017)
AAACGCGGCACGTACGTCAG	NC	non-targeting (Canver et al. 2017)
TACATAGGTACGCGAACCCG	NC	non-targeting (Canver et al. 2017)
AGTGGATCGAGTCGCGTATT	NC	non-targeting (Canver et al. 2017)
AATTGCCCCCGTTGCGCGT	NC	non-targeting (Canver et al. 2017)
GCGCGATAACCCGTTCCGTC	NC	non-targeting (Canver et al. 2017)
CCGACCTTGATCGCGAAAAT	NC	non-targeting (Canver et al. 2017)
GCGGACCGTCCAACGGAAGT	NC	non-targeting (Canver et al. 2017)
TACGTGCCGCTAATCGTCTG	NC	non-targeting (Canver et al. 2017)
CTCGGATGCAGCGCGTTATA	NC	non-targeting (Canver et al. 2017)
ACGATGACGCGTTCGCATCT	NC	non-targeting (Canver et al. 2017)
GGCCGGCGGCGTATCAACGG	NC	non-targeting (Canver et al. 2017)
CCAAGGCGCCTTCGACGTCG	NC	non-targeting (Canver et al. 2017)
AGATCCCCGGTCCGTACGGA	NC	non-targeting (Canver et al. 2017)
GTTATAATTATGCGGTCCGC	NC	non-targeting (Canver et al. 2017)
CCGTTGCATCGCGTACCGAT	NC	non-targeting (Canver et al. 2017)
GGCCGTGCGGACATGATTCTG	NC	non-targeting (Canver et al. 2017)

Target-specific guide sequence	Type	Target
ATATAGGATCATCTACGCGC	NC	non-targeting (Canver et al. 2017)
TTCGACTACATCGTTCGTCG	NC	non-targeting (Canver et al. 2017)
GTATCCGGAGTACGTTCGGC	NC	non-targeting (Canver et al. 2017)
GCTATGAATGACGTCCGCGA	NC	non-targeting (Canver et al. 2017)
TCACCTATGCGCGCGATTCT	NC	non-targeting (Canver et al. 2017)
GTTACGCTCGAATACGAGTC	NC	non-targeting (Canver et al. 2017)
CCGATTCGTTCTAGACGGGG	NC	non-targeting (Canver et al. 2017)
TGGTACGGAAGGTACGTCCG	NC	non-targeting (Canver et al. 2017)
CGCTTGACGTATGAACGATC	NC	non-targeting (Canver et al. 2017)
ATAATCGGGGCGACGACGAC	NC	non-targeting (Canver et al. 2017)
GTCAGTGCATACGACTCG	NC	non-targeting (Canver et al. 2017)
ATACTACACGAGGACGTCTGT	NC	non-targeting (Canver et al. 2017)
TCGAGCTAATTACGCGACCT	NC	non-targeting (Canver et al. 2017)
AGTCGAAAGACGCGCTATCG	NC	non-targeting (Canver et al. 2017)
GCGCGCGCGCTATCGTGGTC	NC	non-targeting (Canver et al. 2017)
TAATCACGGAGACGCGCTAG	NC	non-targeting (Canver et al. 2017)
CAATGTCGCCCACCCGATCG	NC	non-targeting (Canver et al. 2017)
CCGCGAGTCGGTTTGTCGCA	NC	non-targeting (Canver et al. 2017)
CGCTCTACCGGTCCCGATAA	NC	non-targeting (Canver et al. 2017)
GACGCGCACGTGCGTTGCA	NC	non-targeting (Canver et al. 2017)
TTAGGATCGGAACGACTCGA	NC	non-targeting (Canver et al. 2017)
GAAACTCGATTACGTGTCGT	NC	non-targeting (Canver et al. 2017)
GTGAAAGTACGCCCCGACGA	NC	non-targeting (Canver et al. 2017)
GCCGAACCCCCGGCGTATAA	NC	non-targeting (Canver et al. 2017)
TTATCGATACGCGCTACCGT	NC	non-targeting (Canver et al. 2017)
GTCGAGACCAACGTCGATTG	NC	non-targeting (Canver et al. 2017)
TATTTGTGTTAACGCGGTCG	NC	non-targeting (Canver et al. 2017)
AGGTTACGATCGGACTCGAA	NC	non-targeting (Canver et al. 2017)

Target-specific guide sequence	Type	Target
GTGACGACCACGAACTCGTA	NC	non-targeting (Canver et al. 2017)
TCGAATGACGACGCCGTTCA	NC	non-targeting (Canver et al. 2017)
ACCTGGCGTAACGATTACGA	NC	non-targeting (Canver et al. 2017)
CGGATCATCTACGTACCGCC	NC	non-targeting (Canver et al. 2017)
GATGTGCACCCGTCGCACGC	NC	non-targeting (Canver et al. 2017)
CCATACGATCAGACTTCGCG	NC	non-targeting (Canver et al. 2017)
CGCCTTATTTACCGGGCGCA	NC	non-targeting (Canver et al. 2017)
TCGATTGCCGACTGTACCGC	NC	non-targeting (Canver et al. 2017)
ATTGAGGCGACACGTACGCG	NC	non-targeting (Canver et al. 2017)
CACGCTTACGGCGTACTTCG	NC	non-targeting (Canver et al. 2017)
AATTACTCGATGCAACGCGT	NC	non-targeting (Canver et al. 2017)
GCTGATCTATCGCGGTCGTC	NC	non-targeting (SCR1) (Lawhorn, Ferreira, and Wang 2014)

Appendix 2: The sequence of the polynucleotide fragment synthesized for the adjustment of the lentiGuide-Hygro-dTomato plasmid.

ttcagaccacactcccaaccccgaggggagccagaGCATACGAGATAGCTAGCCACCgagggcctatt
 tcccatgattccttcataatttgcatatacgatacaaggctgttagagagataattagaattaatttga
 ctgtaaacacaaagatatttagtacaaaatacgtgacgtagaaagtaataatttcttggttagtttgca
 gttttaaaattatgttttaaaatggactatcatatgcttacgtaacttgaaagtatttcgatttctt
 ggctttatatatcttgtggaaaggacgaaacaccggagacggttcattgcaggatcaaattcttacaaca
 ttttaattgtttcatagggataacagggtaatcataatacatttaacaacatgtgatccagctgcgcgtc
 tcgaacagcatagcaagttttaaataaggctagtcggttatcaacttgaaaaagtggcaccgagtcggt
 gcgtttttttaagcttggcgtaactagatcttgagacaaatggcagtattcatccacaaattttaaaaga
 aaaggggggattgggggtacagtgcaggggaaagaatagtagacataatagcaacagacatacaaac
 taaagaattacaaaaacaaattacaaaaattcaaaattttcgggtttattacaggggacagcagagatc
 cactttggcgccggctcgagggggcccgggtgcaaagatggataaagttttaaac

RpuMI restriction site

part of P7 Illumina primer

Unique primer binding site

U6 promoter

BsmBI restriction sites

sgRNA scaffolding

A-U flip

cPPT/CTS

XmaI restriction site

A 735 bp fragment to replace the original lentiGuide-Hygro-dTomato sequence between bp 8904 and bp 11337 as per the published map of lentiGuide-Hygro-dTomato (<https://www.addgene.org/99376/>, (Ho et al., 2017)). The diagram detailing the replacement is in Figure 6.6 in Chapter 6.4.