

**“Cut-and-Run”: A New Way By Which Aberrant
V(D)J Recombination Causes Genome Instability**

Xiaoling Wang

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

The University of Leeds
Faculty of Biological Sciences
School of Molecular and Cellular Biology

September 2018

The candidate confirms that the work submitted is her own and that appropriate credit has been given where reference has been made to the work of others.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

The right of Xiaoling Wang to be identified as Author of this work has been asserted by her in accordance with the Copyright, Designs and Patents Act 1988.

© 2018 The University of Leeds and Xiaoling Wang

Acknowledgements

First and foremost, I would like to thank my supervisors Dr. Joan Boyes, Prof. Peter Stockley and Dr. Roman Tuma for all of their guidance and support during this project. I would especially like to thank Dr. Joan Boyes for her meticulous care and attention during my project, always giving me a lot of motivation and hopes, and also for her quick and extremely helpful comments on thesis drafts. I thank Dr. Roman Tuma for his help in RAG protein purification and Prof. Peter Stockley for kindly allowing me to use some of his equipment.

I would like to thank Dr Chris Kirkham, for proposing the “cut-and-run” hypothesis and setting up the initial conditions of *in vitro* RAG cutting assays, as well as DNase I footprint assays. I would also like to thank the other members of the Boyes lab, especially, Daniel Thwaites, for always giving me a lot of suggestions and kindly allowing me to use his full-length RAG proteins and Krab virus, Dr James Scott for his work on the LAM-HTGTS assays, giving me a better understanding of “cut-and-run” and for his help in FACS, Alastair Smith and Zeqian Gao for their assistance in making single cell clone. I am very grateful to Dr. Nikesh Patel, Dr. Simon White and Dr. Alex Borodavka from the Stockley and Tuma labs for their help with fluorescent labelling.

I would like to thank all of my friends in Leeds, for taking care of me for a long period of time even though I am always a bit annoying. All the craziness, laughter and tears we shared will definitely become one of my most amazing memories.

Finally, I would like to give my huge thanks to my mom and dad, my brother and sister in law, as well as my lovely little niece, for all of their support, advice and love throughout my studies and life. The moments being together with you are without doubt my happiest time. You all mean the world to me.

Abstract

V(D)J recombination is essential to create a highly diverse repertoire of antigen receptors. However, this process carries an inherent risk because it involves breaking genomic DNA. Indeed, errors in this process are a major cause of leukaemias and lymphomas. Previous studies showed that the recombination by-product, the excised signal circle (ESC), can be a source of genome instability. The Boyes lab showed recently that the ESC triggers cutting at a recombination signal sequence (RSS) but the ESC itself is largely uncut i.e. asymmetric cleavage occurs in an ESC/RSS synaptic complex. Since the ESC remains bound by recombinase (RAG) proteins, this suggests that the ESC can trigger cutting at one RSS and then “run” to trigger cutting at further RSS or cryptic RSS (cRSS) in a reaction that was named “cut-and-run”. This further implies that the ESC could trigger numerous double strand breaks (DSBs) in the genome, leading to genome instability.

I further investigated the mechanism of cut-and-run. Firstly, I examined the reaction *in vitro* and showed that the asymmetric cleavage is due to RAGs simultaneously binding to both RSSs in the ESC and blocking cutting at heptamer/heptamer boundary of ESC. Next, I showed this asymmetric cutting also occurs *in vivo*. Moreover, if I mutate one of the RSSs in the ESC, both ESC and RSS are cut.

Next, I investigated the dangers of cut-and-run. Firstly, I showed that it generates one blunt signal end and one hairpinned coding end that are not readily repaired. Secondly, ESC was found to trigger cutting at cRSS *in vitro*. Thirdly, I showed that this reaction can cause DSBs throughout the genome. Lastly, the lambda and kappa ESCs were found to persist to mature B cells.

Together, these investigations suggest that cut-and-run is a dangerous reaction by providing substrates that potentially lead to chromosome instabilities.

Contents

Acknowledgements	iii
Abstract	iv
Contents	v
List of Figures	xi
List of Tables	xiv
Abbreviations	xv
Chapter 1: Introduction	1
A) The adaptive immune system	1
1.1 Innate and adaptive immunity	1
1.2 B and T lymphocytes	2
1.2.1 B lymphocytes	2
1.2.2 T lymphocytes	3
1.3 Antigen receptors	4
1.3.1 Immunoglobulins	5
1.3.2 T cell receptors.....	10
B) V(D)J recombination	12
1.4 Generation of antigen receptor diversity	12
1.5 Overview of V(D)J recombination.....	14
1.6 The RAG proteins	20
1.7 High mobility group box (HMGB) proteins in V(D)J recombination.....	25
1.8 Assembly of RAG-RSS complexes	27
1.9 RAG cleavage of RSSs.....	30
1.10 Basis of the 12/23 rule	32
1.11 “Beyond the 12/23” restriction	35
C) Regulation of V(D)J recombination	37
1.12 Ordered recombination	37
1.13 B and T cell development.....	37
1.13.1 B cell development.....	37
1.13.2 T cell development.....	40
1.14 Regulation of RAG expression	41
1.15 RSS accessibility and V(D)J recombination.....	44
1.16 Histone modification and V(D)J recombination.....	44

1.16.1 Histone acetylation and V(D)J recombination	45
1.16.2 Histone methylation and V(D)J recombination	45
1.17 Non-coding transcription and V(D)J recombination	46
1.18 Allelic exclusion.....	47
1.18.1 The stochastic model.....	48
1.18.2 The probabilistic model	49
1.18.3 The instructive model.....	49
1.18.4 Feedback inhibition models	50
D) Generation of additional antigen receptor diversity	51
1.19 Activation-induced cytidine deaminase (AID)	51
1.20 Somatic hypermutation	52
1.21 Class switch recombination	53
E) Errors in V(D)J recombination.....	55
1.22 Recognition errors in V(D)J recombination	56
1.23 End donation	61
1.24 RAG mediated transposition	63
1.25 ESC reintegration	65
1.26 A newly uncovered aberrant recombination reaction: cut-and-run.....	67
F) Aims	70
Chapter 2: Materials and Methods	71
A) Common buffers and Media.....	71
2.1 Common buffers.....	71
2.2 Tissue culture media and solutions	74
B) Manipulation of DNA	74
2.3 Restriction enzyme digestion of DNA.....	74
2.4 Traditional PCR.....	75
2.5 Agarose gel electrophoresis.....	75
2.6 DNA ligation	75
2.7 Preparation of chemically competent <i>E. coli</i>	76
2.8 Transformation of <i>E. coli</i>	76
2.9 Phenol/chloroform extraction	77
2.10 Ethanol precipitation.....	77
2.11 Gel purification of DNA	78
2.12 Small-scale preparation of plasmid DNA (Mini-prep).....	78

2.13 Large-scale preparation of plasmid DNA using Qiagen Kits (Maxi-prep)	79
2.14 Large-scale preparation of plasmid DNA by the cesium chloride (CsCl) method (Maxi-prep)	80
2.15 Preparation of glycerol stocks for long-term storage of plasmid DNA	81
C) Expression and purification of RAG proteins	81
2.16 Transfection of HEK-293T cells with RAG expression vectors	81
2.17 Purification of core RAG proteins from HEK-293T cells	82
2.18 Purification of core RAG1 plus full-length RAG2 proteins from HEK-293T cells	82
2.19 SDS-Polyacrylamide gel electrophoresis of proteins	83
2.20 Coomassie brilliant blue staining of polyacrylamide gel	83
2.21 Estimation of protein concentration by SDS-PAGE and densitometry	84
2.22 Identification of tagged RAG proteins by western blotting	84
D) <i>In vitro</i> RAG cutting assay	85
2.23 Gel purification of DNA oligonucleotides	85
2.24 Radiolabelling of oligonucleotides	86
2.25 Annealing oligonucleotides	86
2.26 <i>In vitro</i> RAG cutting assays	87
2.27 Electrophoretic mobility shift assay (EMSA)	87
2.28 Supershift assay	88
2.29 Signal end release assay	88
2.30 Preparation of fluorescently- labelled RSSs and ESC	88
2.31 DNase I footprint analysis	89
2.32 Preparation of G+A ladders	90
E) <i>In vivo</i> RAG cutting assay	91
2.33 Generation and concentration of lentivirus	91
2.34 Transfection of NIH3T3 cells	91
2.35 Spinfection of REH cells	92
2.36 Analysis of level of transfection or transduction by flow cytometry	92
2.37 Recovery of DNA from mammalian cell by Hirt extraction	92
2.38 Detection of RAG cutting in REH cells	93
2.39 Analysis of RAG cleavage by Quantitative Polymerase Chain Reaction (qPCR)	93

2.40 Calculation of virus copies per cell	94
2.41 Ligation-mediated PCR (LM-PCR) assay	95
2.42 Immunofluorescence to detect γ H2AX foci.....	96
2.43 Apoptosis assay	97
F) B Cells isolation.....	98
2.44 Preparation of B cells for FACs with NH ₄ Cl treatment.....	98
2.45 Isolation of B cells by FACS.....	98
2.46 Genomic DNA extraction	102
2.47 Total RNA extraction and reverse transcription	102
2.48 Nested Polymerase Chain Reaction (Nested PCR).....	103
G) Generation of Artemis-deficient REH cell line.....	105
2.49 Knock-down of Artemis in REH cells by shRNA.....	105
2.50 Generation of Artemis-deficient REH cells using CRISPR-Cas 9 targeting	107
2.51 Generation of Artemis-deficient REH cell lines using the KRAB repressor	109
2.52 Detection of loss of Artemis expression.....	109
2.52.1 Determination of Artemis expression by qPCR.....	109
2.52.2 Detection of Artemis mutations by PCR and sequencing.....	110
2.52.3 Determination of Artemis deficiency by western blotting	110
H) Oligonucleotides.....	111
I) Plasmid maps	120
Chapter 3: Investigation of the molecular basis of asymmetric ESC/RSS cleavage in vitro	142
A) Introduction	142
B) Results.....	143
3.1 Expression and purification of core and full-length RAG proteins.	143
3.2 Comparison of asymmetric cutting by core RAG1 plus core RAG2 proteins and core RAG1 plus full-length RAG2 proteins.....	146
3.3 Asymmetric cleavage is due to RAG proteins binding to both RSSs within the ESC.....	149
3.4 Both RAG1 and RAG2 are present in the ESC/RAG complex.....	151
3.5 Mutating one side of the ESC restores symmetric cleavage within an ESC/RSS synaptic complex	153

3.6 Increasing the spacer between two RSSs within the ESC causes ESC cleavage.....	158
3.7 Analysis of RAG/mutant ESC complexes	159
C) Discussion	162
3.8 Core and full-length RAG protein purification.....	162
3.9 Asymmetric cleavage is due to RAG binding to both RSSs within the ESC.....	164
3.10 The RAG2 C-terminus and cut-and-run reaction.....	166
Chapter 4: Cut-and-run occurs in vivo via the same mechanism as in vitro.....	167
A) Introduction	167
B) Results.....	169
4.1 Asymmetric RSS/ESC cleavage occurs <i>in vivo</i>	169
4.2 Mutating one side of the ESC restores symmetric cleavage <i>in vivo</i>	172
4.3 Asymmetric cutting is observed with both full-length and core RAGs...	173
4.4 SCR7 facilitates the detection of consensus 12-RSS signal ends <i>in vivo</i>	175
4.5 Cut-and-run occurs at physiological concentrations of RAGs	178
C) Discussion	181
4.6 Asymmetric cutting <i>in vivo</i>	181
4.7 SCR7 and cut-and-run reaction.....	182
Chapter 5: Dangers of the cut-and-run reaction	185
A) Introduction	185
B) Results.....	189
5.1 The ESC triggers cutting at cRSS <i>in vitro</i>	189
5.2 Analysis of post-cleavage complex stability	190
5.3 The nature of the ends generated by cut-and-run	193
5.4 The ESC triggers double-strand breaks <i>in vivo</i>	196
5.5 The ESC triggers apoptosis	200
5.6 The ESC persists to mature B cell stages	202
C) Discussion	209
5.7 The ESC can trigger cutting at cRSSs	209
5.8 The ends generated by cut-and-run reaction	211
5.9 The persistence of ESC	213

Chapter 6: Discussion	218
6.1 The occurrence and mechanism of the cut-and-run reaction.....	218
6.2 Dangers posed by cut-and-run	221
6.3 The likelihood of cut-and-run occurring B and T cells.....	223
6.4 Cut-and-run likely provides substrates for chromosomal translocations	226
6.5 Cut-and-run and reintegration.....	228
6.6 Future plans	229
6.7 Conclusions	231
References	232

List of Figures

Figure 1.1 Antigen receptor variable exons are assembled by V(D)J recombination.....	5
Figure 1.2 The structure of an antibody.	6
Figure 1.3 The immunoglobulin heavy chain locus.....	8
Figure 1.4 The immunoglobulin kappa chain locus.....	9
Figure 1.5 The immunoglobulin lambda chain locus.....	10
Figure 1.6 The structure of a T cell receptor.	11
Figure 1.7 The conserved structure of a recombination signal sequence (RSS)13	
Figure 1.8 Overview of the cleavage and joining phases of V(D)J recombination.....	17
Figure 1.9 Non-homologous end joining pathway.....	18
Figure 1.10 Generation of junctional diversity by adding P and N nucleotides19	
Figure 1.11 The core and non-core domains of RAG proteins	22
Figure 1.12 Schematic of RAG-RSS complexes.....	28
Figure 1.13 Schematic representation of RAG-mediated catalysis steps	31
Figure 1.14 The 12/23 rule.....	33
Figure 1.15 Beyond the 12/23 rule.....	33
Figure 1.16 B cell development.....	40
Figure 1.17 Model for somatic hypermutation	53
Figure 1.18 Class switch recombination (CSR)	54
Figure 1.19 Recognition errors in V(D)J recombination.....	59
Figure 1.20 End donation.....	63
Figure 1.21 RAG mediated transposition reactions	65
Figure 1.22 ESC reintegration.....	67
Figure 1.23 Cut-and-run.....	69
Figure 2.1 Sort template for primary pre-B cells	99
Figure 2.2 Sort template for primary spleen IgM+ cells.....	100
Figure 2.3 Sort template for primary spleen IgG+ high cells.....	101
Figure 2.4 Lambda BAC DNA standard curve for measuring the dsDNA concentration.....	105
Figure 2.5 Schematic of Tet-pLKO-puro and its mechanism of action.....	107
Figure 2.6 Schematic of the CRISPR-Cas9 system.....	108

Figure 2.7 Map of pEF-McR1 (pEFXC-MBP-cRAG1).....	120
Figure 2.8 Map of pEF-McR2 (pEFXC-MBP-cRAG2).....	121
Figure 2.9 Map of pJH548 (flRAG1).....	122
Figure 2.10 Map of pEFflRAG2.....	1243
Figure 2.11 Map of pEFXC-cRAG2.....	124
Figure 2.12 Map of pEFXC-Myc-MBP-cRAG1.....	125
Figure 2.13 Map of pEFXC-HA-MBP-cRAG2.....	126
Figure 2.14 Map of pJ β 2.7+.....	127
Figure 2.15 Map of pcDNA3.1-12-RSS (p12RSS).....	128
Figure 2.16 Map of pcDNA3.1-23-RSS (p23RSS).....	129
Figure 2.17 Map of pJ δ 1+.....	130
Figure 2.18 Map of pcDNA3.1-ESC (pESC).....	131
Figure 2.19 Map of pWPI-12RSS.....	132
Figure 2.20 Map of pWPI-23RSS.....	133
Figure 2.21 Map of pWPI-ESC.....	134
Figure 2.22 Map of pWPI-Ruby2_12RSS.....	135
Figure 2.23 Map of pWPI_Ruby2_23RSS.....	136
Figure 2.24 Map of pWPI_Ruby2_ESC.....	137
Figure 2.25 Map of pWPI-ESC12d7.....	138
Figure 2.26 Map of pWPI-ESC12d9.....	139
Figure 2.27 Map of pWPI-ESC23d7.....	140
Figure 2.28 Map of pWPI-ESC23d9.....	141
Figure 3.1 RAG proteins purified from HEK293T cells.....	145
Figure 3.2 Asymmetric cleavage by RAG proteins.....	147
Figure 3.3 Purification of full-length RAG proteins.....	148
Figure 3.4 <i>In vitro</i> cutting with full-length RAG1 plus full-length RAG2.....	149
Figure 3.5 Asymmetric cleavage is due to RAG proteins binding to both RSSs within the ESC.....	151
Figure 3.6 Both RAG1 and RAG2 are present in the ESC/RAG complex....	153
Figure 3.7 Mutation of the 12-RSS nonamer in the ESC restores symmetric RSS/ESC cutting.....	155
Figure 3.8 Mutation of 23-RSS CAC in the ESC restores symmetry to 23-RSS/ESC cutting.....	156

Figure 3.9 Mutation of 12-RSS CAC in the ESC does not restores symmetric RSS/ESC cutting.....	157
Figure 3.10 ESC is cleaved by increasing the spacer between the two RSSs within the ESC.....	159
Figure 3.11 Mutation of the 12-RSS nonamer in the ESC eliminates RAG binding.....	160
Figure 3.12 Mutation of 12-RSS CAC in the ESC does not eliminate RAG binding but mutation of 23-RSS CAC does decrease RAG binding.....	161
Figure 4.1 Cleavage of transfected 12-RSS/ESC substrates is asymmetric <i>in vivo</i>	171
Figure 4.2 Mutation of the RSS in the ESC restores symmetric cutting within an RSS/ESC complex <i>in vivo</i>	173
Figure 4.3 Comparison of asymmetric RSS/ESC cleavage by core RAG proteins and full-length RAG proteins	175
Figure 4.4 Symmetric 12-RSS/23-RSS cleavage <i>in vivo</i>	176
Figure 4.5 Asymmetric RSS/ESC cleavage <i>in vivo</i>	177
Figure 4.6 Cut-and-run occurs at physiological concentrations of RAGs.....	180
Figure 5.1 The ESC triggers cutting at a cryptic RSS	190
Figure 5.2 Signal ends are released faster following ESC/RSS cutting.....	193
Figure 5.3 The nature of ends generated by cut-and-run reaction.....	195
Figure 5.4 Flow chart for γ H2Ax foci detection.....	198
Figure 5.5 Immunofluorescence to detect γ H2Ax X foci.....	199
Figure 5.6 The ESC triggers cell death	201
Figure 5.7 The ESC triggers cell apoptosis.....	202
Figure 5.8 The ESC is present in both pre-B and splenic B cells	204
Figure 5.9 RAG1 expression at different B cell developmental stages	205
Figure 5.10 Fold changes of ESC/Recombination in different B cell development stages	207
Figure 5.11 Absolute quantification of ESC/Recombination during B cell development.....	208

List of Tables

Table 2.1 Oligonucleotides used for the RAG cutting assays <i>in vitro</i>	111
Table 2.2 Primers used to detect cutting and normalization	113
Table 2.3 Primers used for <i>in vivo</i> RAG cutting assays	113
Table 2.4 Primers used in LM-PCR and nested PCR assays	114
Table 2.5 Primers used to detect ESC and recombination	115
Table 2.6 Oligonucleotides to generate Artemis deficient cell lines	116
Table 2.7 Oligonucleotides used for cloning	117
Table 2.8 Plasmids used during this project	117
Table 5.1 ESC formation and RAG expression during B cell developmental stages.....	214

Abbreviations

aa	Amino acid
AID	Activation-induced cytidine deaminase
ALL	Acute lymphocytic leukaemia
APCs	Antigen presenting cells
APE1/2	Apurinic/apyrimidinic endonuclease 1/2
β -Me	β -mercaptoethanol
BAC	Bacterial artificial chromosome
BCR	B cell receptor
BER	Base excision repair
bp	Base pair
BSA	Bovine serum albumin
cDNA	Coding DNA
CDR	Complementarity determining region
CE	Coding end
CEC	Coding end complex
ChIP	Chromatin immunoprecipitation
ChIP-seq	Chromatin immunoprecipitation sequencing
CMV	Cytomegalovirus
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeat
cRSS	Cryptic recombination sequence
Cryo-EM	Cryo-electron microscopy
CsCl	Cesium chloride
CSR	Class switch recombination
CV	Column volume
dC	Deoxycytidines
dH ₂ O	Deionised water
DMEM	Dulbecco's modified Eagle's medium
DMSO	Dimethyl sulphoxide
DN	Double negative
DNA	Deoxyribonucleic acid
DNA-PKcs	DNA-dependent protein kinase catalytic subunit
DP	Double positive
DSB	Double-strand break

DTT	Dithiothreitol
dU	Deoxyiracils
<i>E.coli</i>	<i>Escherichia coli</i>
EDTA	Ethylenediaminetetraacetic acid
EGFP	Enhanced green fluorescent protein
EMSA	Electrophoretic mobility shift assay
ESC	Excised signal circle
EtBr	Ethidium bromide
FACS	Fluorescence activated cell sorting
FCS	Foetal calf serum
FISH	Fluorescence in-situ hybridisation
FITC	Fluorescein isothiocyanate
FRs	Framework regions
GCs	Germinal centers
GFP	Green fluorescent protein
GST	Glutathione S-transferase
H3K4me3	tri-methylated histone H3 lysine 4
HA	Human influenza hemagglutinin
HRP	Horse radish peroxidase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HMGB	High mobility group protein B
HSC	Haematopoietic stem cell
HSC1/2	HMGB1 single-RSS containing complex 1 or 2
IDT	Integrated DNA Technologies
IFN- γ	Interferon- γ
Ig	Immunoglobulin
IL	Interleukin
kDa	Kilodalton
KOAc	Potassium acetate
KPNA1	Karyopherin alpha 1
KRAB	Kruppel-associated box
LAM-HTGTS	Linear amplification-mediated high-throughput genome-wide translocation sequencing
LB	Lysogeny broth

LiCl	Lithium chloride
LMPCR	Ligation-mediated PCR
LM-TECA	Ligation-mediated transferred-end capture assay
LncRNA	Long non-coding RNA
Mb	Mega bases
MBN	Mung bean nuclease
Mbr	Major breakpoint region
MBP	Maltose binding protein
MHC	Major histocompatibility complex
MMR	Mismatch repair
MOPS	3-(N-morpholino) propanesulfonic acid
Mtc	Major translocation cluster
NaOAc	Sodium acetate
NaPi	Sodium phosphate buffer
NBD	Nomamer binding domain
NEB	New England Biolabs
NH ₄ Cl	Ammonium chloride
NHEJ	Non-homologous end joining
NK	Natural killer
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PC	Paired complex
PCC	Post-cleavage complex
PCR	Polymerase chain reaction
PE	Phycoerythrin
PEI	Polyethyleneimine
PFA	Paraformaldehyde
PHD	Plant homeodomain
PK	Proteinase K
PMSF	Phenylmethanesulfonylfluoride
PNK	Polynucleotide kinase
Pol λ/μ	Polymerases λ/μ
PVDF	Polyvinylidene fluoride

qPCR	Quantitative polymerase chain reaction
RAG	Recombination activating gene
RING	Really interesting new gene
RNA	Ribonucleic acid
ROS	Reactive oxygen species
RPMI	Roswell Park Memorial Institute
RSS	Recombination signal sequence
RT	Room temperature
SC	Synaptic complex
SC1/2	Single RSS containing complex 1/2
SCID	Severe combined immunodeficiency
SDS	Sodium dodecyl sulphate
SE	Signal end
SEC	Signal end complex
sgRNA	Single guide RNA
SHM	Somatic hypermutation
shRNA	Small hairpin RNA
SOB	Super optimal broth
ssDNA	Single stranded DNA
T _a	Annealing temperature
TAE	Tris-acetate-EDTA
TBE	Tris-borate-EDTA
TBS	Tris-buffered saline
TCR	T cell receptor
TdT	Terminal deoxynucleotidyl transferase
TE	Tris EDTA
TEA promoter	T early α promoter
TNF- α	Tumour necrosis factor α
TSSs	Transcription start sites
UNG	Uracil DNA glycosylate
V(D)J	Variable (Diversity) Joining
VprBP	Vpr binding protein
XRCC4	X-ray repair cross-complementing protein 4
ZDD	Zinc-binding dimerization domain

Chapter 1:

Introduction

A) The adaptive immune system

1.1 Innate and adaptive immunity

All multicellular organisms suffer attack from pathogenic microbes, such as bacteria, viruses, parasites and fungi. In order to avoid infections, vertebrates have developed a highly sophisticated immune system, recognizing and destroying invading pathogenic microbes. This defence against microbes is generally divided into innate immunity and adaptive immunity. The former is the first line of defence and the latter is more specific and complicated. These two immune systems mutually cooperate to protect organisms from attack. Innate and adaptive immunity differ with respect to how quickly and for how long they respond to pathogens, their specificity for different classes of microbes and their effector cell types (Alberts et al., 2002).

The innate immune system is comprised of proteins and cells which are ready to mobilize and fight microbes when infections occur. There are five main components of the innate immune system: physical epithelial barriers, phagocytic leukocytes, dendritic cells, circulating plasma proteins and natural killer (NK) cells. The components of this system non-specifically recognize common features of pathogens and destroy them effectively. For example, cells such as polymorphonuclear leukocytes recognise foreign particles and infected cells, ingest them in a phagocytic vacuole or endosome and break them down into small innocuous products. Nevertheless, the innate immune system alone is not capable of destroying all infectious organisms, nor can it cope with micro-organisms that have already adapted to avoid these non-specific mechanisms. Thus, the more specific and complicated adaptive immune system has developed in vertebrates. In contrast to the innate immune system, the components of the adaptive immune system are normally silent. Once pathogens evade or overcome innate immune defences, these

components are activated to protect the organism by creating potent mechanisms for neutralizing or even eliminating the invading microbes. The adaptive immune responses are mainly divided into humoral immunity, which is mediated by antibodies produced by B lymphocytes, and cell-mediated immunity, which is mediated by T lymphocytes (Murphy, 2012).

1.2 B and T lymphocytes

Both B and T lymphocytes arise from haematopoietic stem cells (HSCs). HSCs give rise to all other blood lineages via the process of haematopoiesis which occurs in the red bone marrow (Birbrair & Frenette, 2016). HSCs give rise to both the common myeloid progenitor (CMP) and the common lymphocyte progenitor (CLP). CMP differentiation results in the production of: macrophages, monocytes, basophils, neutrophils, erythrocytes, eosinophils, and megakaryocytes, which differentiate to platelets. The CLP is responsible for the production of B lymphocytes, T lymphocytes and natural killer cells (Till & MuCulloch, 1961).

B and T lymphocytes are essential to the adaptive immune response, as they function by expressing highly specific and varied cell surface receptors, immunoglobulin (Ig) molecules and T cell receptor (TCR), respectively, which can specifically recognise and bind an antigen. Each antigen receptor has high specificity towards a single antigen, even a single amino acid residue difference can prevent an antigen receptor from binding to its cognate antigen. Although both B and T lymphocytes express similar cell surface receptors, they elicit significantly different responses when they encounter a pathogen.

1.2.1 B lymphocytes

B lymphocytes, or B cells, comprise 10% to 20% of circulating peripheral lymphocytes and mainly recognise circulating antigens that can have many chemical structures. They function in the humoral immunity component of the adaptive immune system by secreting antibodies (Murphy, 2012). B cells are the only cells that express antibodies on their surface, allowing the B cell to bind to a specific antigen. When stimulated by a microbial antigen, the antigen

is processed into peptides and presented on the cell surface with the major histocompatibility complex (MHC) molecules. These peptides are then recognized by T helper cells, which then activate the B cell (Parker, 1993), causing cell division and differentiation into plasma cells. Plasma cells are long-lived and non-proliferating cells that are able to secrete large amounts of antibodies, which travel in the lymphatic system and blood plasma (LeBien & Tedder, 2008; Nutt et al., 1999). Subsequently, these antibodies bind to specific antigens, neutralizing or destroying the microbe in collaboration with other cells of the immune system such as macrophages and natural killer cells (Murphy, 2012). After a response, memory B cells persist, circulating through the body, “remembering” the particular antigen and initiating a stronger and faster antibody response upon re-exposure to the antigen (Kurosaki et al., 2015).

1.2.2 T lymphocytes

T lymphocytes, also known as T cells, comprise 60% to 70% of circulating peripheral lymphocytes. Unlike B cells, the majority of T cells (>95%), are only able to recognize peptide fragments when they are presented by MHC proteins on the surfaces of antigen presenting cells. T cells play a central role in cell-mediated immunity of the adaptive immune system (Murphy, 2012). They can be distinguished from other lymphocytes by the presence of a T cell receptor on the cell surface (Alberts et al., 2002). When stimulated, two major subsets of T cells are activated.

The first, CD4+ “helper” T cells, which secrete interleukins (IL) that help B cells to produce antibodies and activate macrophages to eliminate endocytosed microbes (Janssen et al., 2003; Murphy et al. 2007; Williams & Bevan, 2007). Helper T cells can be further classified into four different groups: type 1 CD4+ T effector cells (T_H1 cells), type 2 CD4+ T effector cells (T_H2 cells), T_H17 cells and regulatory T cells (T_{reg}) (Kurts, 2008; Parker, 1993; Yamane & Paul, 2013). T_H1 cells secrete tumour necrosis factor α (TNF- α) and interferon- γ (IFN- γ) to recruit macrophages to combat intracellular bacterial and protozoal infections (Kurts, 2008; Mosmann, 1992). In contrast, T_H2 cells secrete IL-4, IL-

5 and IL-13 to provide a response against extracellular parasites by recruiting a different class of immune effectors, namely B cells (Kurts, 2008; Stevens et al., 1988). T_H17 cells are not only targeted against extracellular bacterial and fungicidal infections, but are also involved in combating inflammatory and autoimmune diseases (Kurts, 2008; Murphy et al., 2007; Noack & Miossec, 2014). T_{reg} cells mainly function to suppress response to self-antigens, decreasing the chance of autoimmune diseases (Beissert et al., 2006; Murphy et al., 2007; Noack & Miossec, 2014; Parker, 1993; Yamane & Paul, 2013). The second, CD8+ “cytotoxic” T cells, which also can secrete soluble mediators but play a more essential role in directly killing tumour or virus-infected cells (Janssen et al., 2003; Shedlock & Shen, 2003; Sun et al., 2004). After an infection has resolved, memory T cells persist long-term, to provide the immune system with a “memory” of past infections and to be ready to quickly expand to large numbers of effector T cells upon re-exposure to their cognate antigen.

1.3 Antigen receptors

Antigen receptors, known as immunoglobulins (Igs) and T cell receptors (TCRs), produced by B and T cells, respectively, play a vital role in the adaptive immune response, identifying and neutralizing pathogens such as bacteria and viruses and mediating a variety of effector functions. A diverse repertoire of antigen receptors is essential for defence against the many infections faced daily. With the number of unique immunoglobulins and the TCRs totalling around one billion per individual, such magnitude of diversity would be impossible to produce from individual genes, even if all the 20-25,000 known protein coding genes were used in the human genome. The majority of this observed level of diversity is generated by a process known as V(D)J recombination (Figure 1.1), which catalyses the recombination of gene segments within the antigen receptor loci of B and T cells (Gellert, 2002).

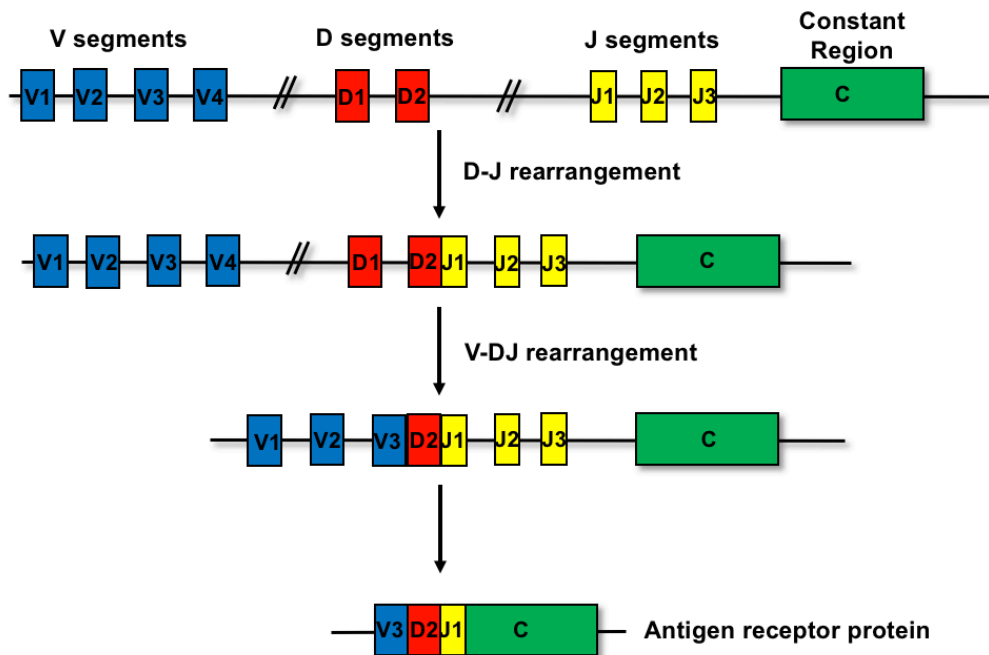


Figure 1.1 Antigen receptor variable exons are assembled by V(D)J recombination

Assembly of a complete variable exon takes place in two steps (in the case of an Ig heavy chain gene or a T-cell receptor β or δ gene). First, a D and a J segment are chosen from among several possibilities and brought together to form a D-J fusion sequence. A V region is then selected and joined with the D-J sequence to form a complete VDJ exon. Ig light chain genes and T-cell receptor α or γ genes rearrange in a single step, involving V-J recombination, as D segments are absent from these loci. C denotes the first constant region exon.

1.3.1 Immunoglobulins

Immunoglobulins, or antibodies, are large, Y-shaped proteins, encoded by three distinct loci - *IgH*, *Ig κ* and *Ig λ* . Immunoglobulins consist of two identical heavy chains and two identical light chains, linked by disulphide bridges between two conserved cysteine residues (Figure 1.2). Both of heavy chain and light chain can be divided functionally into two distinct domains: a variable (V) region, via which immunoglobulins recognize and bind antigens (Charles, 2001; Litman et al., 1993) and a constant (C) region, that is similar in all antibodies, providing scaffolds for the V regions and signalling the immune responses upon antigen binding.

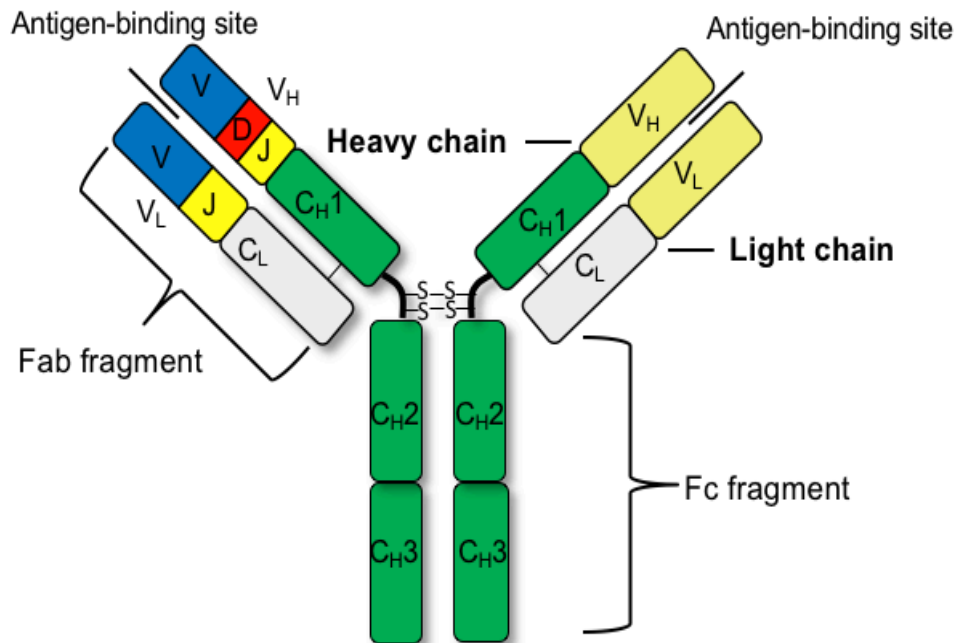


Figure 1.2 The structure of an antibody.

An antibody is composed of two identical heavy chains and two identical light chains, connected by disulphide bonds (shown as -S-S-). Each light chain contains a variable region (V_L) and a constant region (C_L) and each heavy chain contains a variable region (V_H) and three constant regions (C_{H1} , C_{H2} and C_{H3}). The antigen binding sites are located in variable regions of the light and heavy chains. V(D)J recombination combines V, D and J gene segments to form the V_H domain, and V and J gene segments to form the V_L domain.

The variable domains are generated by means of V(D)J rearrangement events, and can then be subjected to somatic hypermutation (Section 1.22) after exposure to antigen to allow affinity maturation and increased diversity. The V region includes three sequence variability and complementarity determining regions, known as CDR. The three CDRs of the heavy chain are paired with the three CDRs of the light chain, forming the antigen binding site. The amino acids contained in different CDRs vary greatly in developing B cells, thus being able to generate various antigen binding sites and recognise a wide range of pathogens. Four invariant framework regions (FRs) surround between the three CDRs, providing a necessary scaffold for the CDRs and facilitating the antigen binding.

In mammals, there are five types of Ig heavy chains: α , δ , ϵ , γ and μ (Charles, 2001), found in IgA, IgD, IgE, IgG and IgM antibodies, respectively (Rhoades &

Pflanzer, 2002). Furthermore, IgG is further subdivided into five subtypes (IgG1, IgG2a, IgG2b, IgG3 and IgG4). The antibodies can either exist as monomers (IgG, IgE, IgD) or dimers (IgA) or pentamers (IgM). The different heavy chains vary in size and composition. For example, α and γ have approximately 450 amino acids, whilst μ and ϵ contain approximately 550 amino acids. Heavy chains α , ϵ and γ have a C region composed of three tandem immunoglobulin constant domains plus a hinge region between the first (C_{H1}) and second (C_{H2}) domains for added flexibility, whilst heavy chains ϵ and μ have a C region that consists of four Ig domains. Prior to binding specific antigens, all of the antibodies are expressed as membrane-bound IgM or IgD subtype on native B cells. Upon antigen binding, isotype switching is activated, resulting other subtype immunoglobulins. These different types of antibodies differ in their functional locations, biological properties and ability to deal with diverse antigens. For example, IgE immunoglobulins can bind to allergens, trigger histamine release from mast cells and protect against parasitic worms. Moreover, IgM is either expressed on the surface of B cells or secreted. It forms pentamers to increase avidity in the secreted form and functions by eliminating pathogens in the early stages of B cell mediated immunity before there is sufficient IgG.

The murine immunoglobulin heavy chain locus (*IgH*) spans 3 Mbp on chromosome 12 and is ordered into distinct segments encoding the variable (V_H), diversity (D_H), joining (J_H), and constant (C_H) regions (Figure 1.3). Eight constant regions encode the various isotypes (C_{H1} , C_{H2} , C_{H1} , C_{H2a} , C_{H2b} , C_{H3} , C_{H4} and C_{H5}). Twelve D_H and four J_H segments are located immediately upstream of the C_{H1} region. There are around 120 V_H gene segments, belonging to 15 partly intermixed families (Brodeur & Riblet, 1984). Each V_H , D_H and J_H coding gene fragments is associated with one or two recombination signal sequences (Section 1.4) that function as recognition elements for RAG proteins (Section 1.6) (Schatz & Spanopoulou, 2005; Jung et al., 2006). Specifically, each V_H and J_H gene segment is flanked by a two turn recombination signal sequence (23-RSS) and a pair of one turn recombination signal sequences (12-RSSs) associates each D_H gene segment.

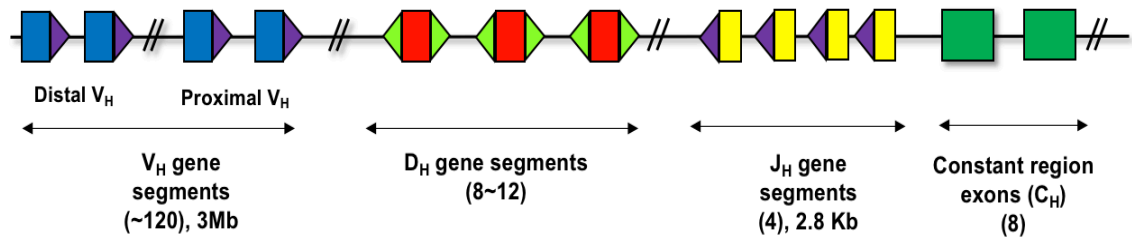


Figure 1.3 The immunoglobulin heavy chain locus

A simplified schematic of the murine immunoglobulin heavy chain (*IgH*) locus. *IgH* contains about 120 V_H gene segments (shown as blue rectangles) associated with 23-RSSs (shown as purple triangles), 12 D_H gene segments (shown as red rectangles) flanked by 5' 12-RSSs and 3' 12-RSSs (shown as light green triangles), 4 J_H gene segments (shown as yellow rectangles) associated with 23-RSSs (shown as purple triangles) and 8 C_H gene segments (shown as dark green rectangles).

There are two types of mammalian Ig light chains: lambda (*Igλ*) and kappa (*Igκ*) (Charles, 2001). The two light chains of each antibody are always identical in isotype and have identical function. Every light chain consists of one variable region and one constant region. The approximate length of a light chain is 211-217 amino acids. It was documented that the λ light chain gene locus only recombines if κ rearrangement fails (Hieter et al., 1981; Sun et al., 2012). Intriguingly, the ratio of antibodies that include a kappa light chain compared to a lambda light chain varies considerably among species: 66% κ / 34% λ in humans, 95% κ / 5% λ in mice, 52% κ / 48% λ in pigs, 8% κ / 92% λ in horses, 9% κ / 91% λ in cows, and the λ light chain is the only one expressed in ducks and chickens (Arun et al., 1996; Gibson 1974; Gray et al., 1967; Kessler et al., 1981; Reynaud et al., 1985; Tallmadge et al., 2014). Although the reason for this is not known yet, it has been suggested that the predominant light chain used may be attributed to the number of germline V genes present in that species (Almagro et al., 1998). This noticeable lack of conservation in Ig light chain usage and gene recombination order is an interesting phenomenon.

In mice, the *Igκ* locus, is spread over 3.2 Mb on chromosome 6, and is composed of 174 V_{κ} gene segments, 5 J_{κ} gene segments and 1 constant region exon. Of the V_{κ} elements, 94-96 are functional as are 4 of J_{κ} elements (Fig 1.4; Giudicelli et al., 2005). one third of the V_{κ} gene segments contain frameshift mutations or stop codons that preclude them from generating functional protein (Schroeder & Cavacini, 2010). It is of interest that, at *Igκ* locus, V_{κ} gene segments are located in two orientations, either in the forward or reverse orientation, leading to recombination occurring either by deletion or inversion (Zachau, 1993). If the recombination occurs by deletion, the intervening DNA is excised from the genome, forming an excised signal circle (ESC). If the recombination occurs by inversion, the signal and coding joints are both maintained in the genome, as described further in Section 1.5.

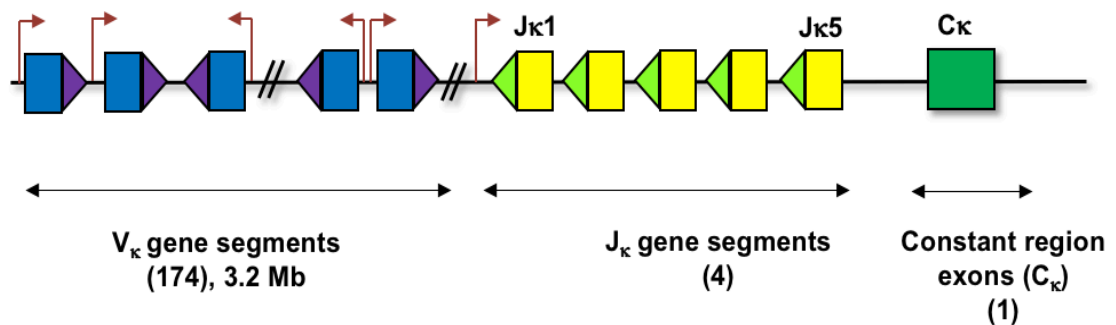


Figure 1.4 The immunoglobulin kappa chain locus

A simplified schematic of the murine immunoglobulin kappa chain (*Igκ*) locus. *Igκ* consists of 174 V_{κ} gene segments (shown as blue rectangles) associated with either 5' 23-RSSs (shown as purple triangles) or 3' 23-RSSs, 5 J_{κ} gene segments (shown as yellow rectangles) flanked by 12-RSSs (shown as light green triangles) and 1 constant region exon (shown as dark green rectangle).

The murine immunoglobulin lambda (*Igλ*) locus is the smallest antigen receptor locus, spread over 240 kb on chromosome 16. *Igλ* consists of three functional V_{λ} gene segments and three J_{λ} gene segments (Figure 1.5). Furthermore, a separate constant region segment lies downstream of each J gene segment (Eisen & Reilly, 1985). The *Igλ* locus includes two likely duplicated gene clusters (Eisen & Reilly, 1985), with each cluster consisting of 1-2 V gene segments followed by 2 J/C regions, in the order $V_{\lambda 2}-V_{\lambda X}-JC_{\lambda 2}-JC_{\lambda 4}-V_{\lambda 1}-JC_{\lambda 3}-JC_{\lambda 1}$ (Dildrop et al., 1987; Storb et al., 1989). Because $JC_{\lambda 4}$ is a

pseudogene, only four main rearrangements are possible: $V_{\lambda 1}$ - $J_{\lambda 1}$, $V_{\lambda 2}$ - $J_{\lambda 2}$, $V_{\lambda X}$ - $J_{\lambda 2}$ and $V_{\lambda 1}$ - $J_{\lambda 3}$, with the ratio 3:2:2:1 (Nadel et al., 1990; Sanchez et al., 1991). Nevertheless, recombination between $V_{\lambda 2}$ and $J_{\lambda 1}$ or $J_{\lambda 3}$ can also occur, although the frequency is rather low (Reilly et al., 1982).

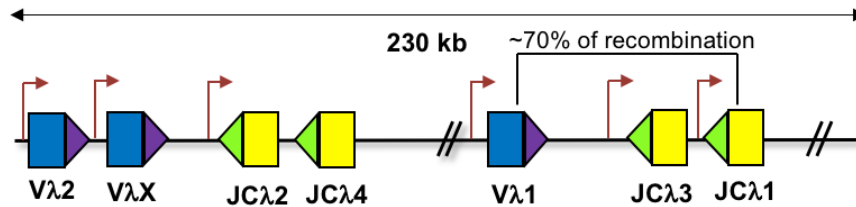


Figure 1.5 The immunoglobulin lambda chain locus

A simplified schematic of the murine immunoglobulin lambda chain (*Igλ*) locus. *Igλ* includes three functional V_{λ} gene segments (shown as blue rectangles) associated with 23-RSSs (shown as purple triangles) and three functional J_{λ} gene segments (shown as yellow rectangles) flanked by 12-RSSs (shown as light green triangles). In addition, each J_{λ} gene segment is associated with its own constant region exon (not shown). Furthermore, about 70% of recombination events occur between $V_{\lambda 1}$ and $J_{\lambda 1}$.

1.3.2 T cell receptors

T cell receptors (TCRs) are disulphide-linked membrane-anchored heterodimeric proteins (Figure 1.6), comprising two chains, either an alpha (α) chain and a beta (β) chain, or a gamma (γ) chain and a delta (δ) chain. The TCR γ chain has greater sequence homology to TCR α , whereas the TCR δ chain is more closely related to TCR β (Chien et al., 1987; Saito et al., 1984). In humans, the majority of T cell receptors (approximately 95%), are composed of α chains and β chains, whilst the remainder (about 5%) consist of γ chains and δ chains (Holtmeier & Kabelitz, 2005). T cell receptors are extremely similar to immunoglobulin proteins, composed of two extracellular domains: a variable region, that binds to the antigen/major histocompatibility complex (MHC) complex, and a constant region that is proximal to the cell membrane. The loci encoding TCR genes also contain multiple V, D and J gene segments for the TCR β and δ loci whilst the TCR α and γ loci have multiple V and J gene segments. During lymphocyte development, these gene segments are rearranged to generate a unique antigen receptor for each cell. Consequently, the T cell receptor is the topological equivalent of an antigen-

binding fragment of the antibody, both being part of the immunoglobulin superfamily. However, distinct from Igs that are both secreted and expressed on the cell surface and can recognize any type of organic antigen, TCRs are only expressed on the cell surface and can only recognize small peptide antigens which are presented on the cell surface via MHC proteins. In addition, the specificity and affinity of TCR-antigen interactions is generally much lower than that of Ig-antigen interactions.

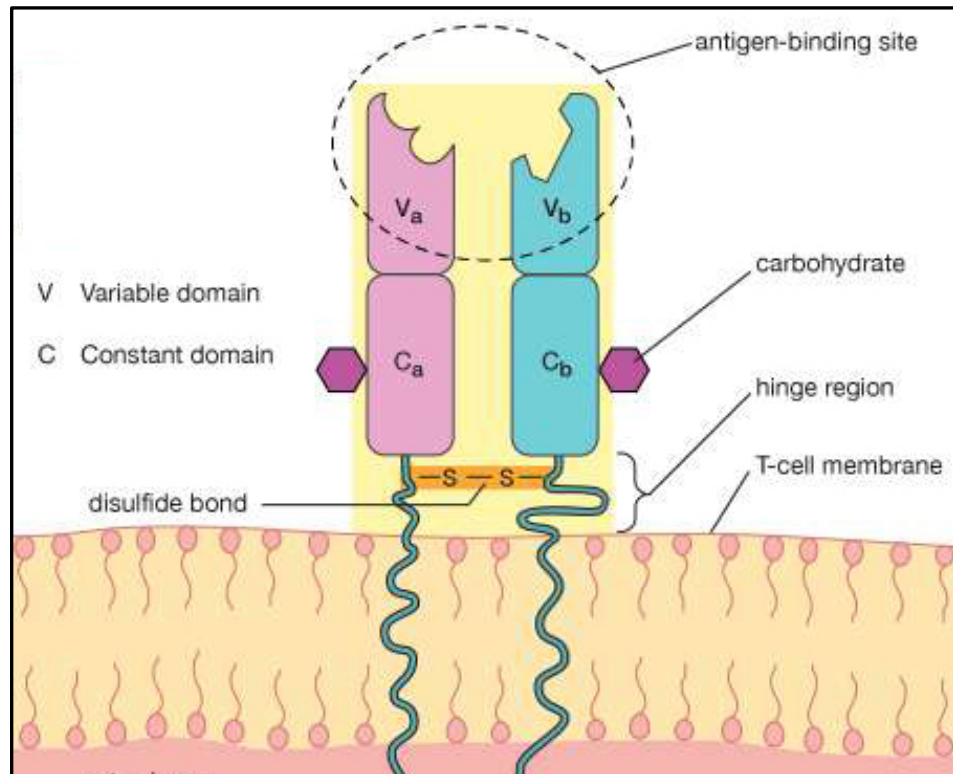


Figure 1.6 The structure of a T cell receptor.

A TCR contains two polypeptide chains: α and β or δ and γ , linked by a disulphide bond. α and γ chains are the equivalent of antibody light chains, whilst β and δ chains equate to antibody heavy chains. TCRs are always membrane-bound. Each chain is composed of one variable domain that binds to the antigen/MHC complex and one constant region that is proximal to the cell membrane, followed by a transmembrane domain and a short intracellular tail.

B) V(D)J recombination

1.4 Generation of antigen receptor diversity

The prerequisite for the adaptive immune system being able to efficiently bind and attack potential pathogens is the generation of a varied repertoire of antigen receptors. The number of antibodies and T cell receptors circulating within the human body at any specific time is up to one billion, which is drastically much more than the number of protein coding genes (approximately 20-25,000) in the entire genome (Janeway et al., 2001). A fundamental question therefore is: How does it manage to generate a huge diversity repertoire of antigen receptors? Studies in the early 1980s showed that, antibodies and T cell receptors are assembled by mixing and matching different gene segments in developing B and T cell by a process termed V(D)J recombination.

Antigen receptors (Igs and TCRs) contain numerous copies of variable (V), diversity (D), and joining (J) gene segments. During V(D)J recombination, D to J gene segments rearrangement also takes place first, followed by V to DJ rearrangement. During gene expression, the variable domain exon combined with a constant region coding exon by RNA splicing. This stochastic assembly of coding exons, together with the imprecise nature by which these coding exons joined, plays an essential role in the generation of the huge diversity of antigen receptors, which are able to recognize a vast range of pathogens (Gellert, 2002).

The V, D and J gene segments, which are used to produce the diverse range of immunoglobulins and T cell receptors, are immediately flanked by recombination signal sequences (RSSs) that are recognized by the V(D)J recombinase (Bassing et al., 2002). An RSS consists of a conserved heptamer ('CACAGTG'), which lies adjacent to the gene segment and serves as the cleavage target, and a conserved nonamer (consensus 'ACAAAACC'), separated by a 12 \pm 1 or 23 \pm 1 bp less-conserved spacer (termed 12-RSS and 23-RSS, respectively) (Ramsden et al. 1994; Schatz & Swanson, 2011; Figure 1.7). Whilst the nucleotide sequences of natural RSSs show significant

variability, the first three nucleotides of the heptamer 'CAC' are essential and show the highest sequence conservation, whereas the remaining heptamer nucleotides are less important (Hesse et al., 1989). The nonamer sequence is less well conserved, with only a few highly conserved positions (as shown in Figure 1.4; Roth, 2014). The spacer is poorly conserved, but the length is highly conserved (Ramsden et al., 1994; Cowell et al., 2004). Notably, the length of the spacer corresponds to approximately one (12 bp) or two turns (23 bp) of the DNA helix.



Figure 1.7 The conserved structure of a recombination signal sequence (RSS)

A consensus RSS consists a conserved heptamer (shown in red) and a conserved nonamer (shown in green), separated by a highly variable 12(\pm 1) bp or 23(\pm 1) bp spacer. The heptamer is always found contiguous to the coding V, D or J gene segment and serves a cleavage site whilst nonamer serves as a binding site for RAG1. The most highly conserved positions of the heptamer and nonamer are shaded in dark blue. The spacer has no conserved sequence but has a fixed length of either 12 or 23 base pairs.

Recombination is subject to the 12/23 rule, which means that recombination only occurs between gene segments that are adjacent to RSSs of different spacer lengths (i.e., one has a "12-RSS" and the other one has a "23-RSS") (van Gent et al., 1996; Bassing et al., 2000; Tonegawa, 1983). In loci comprised of only V and J gene segments, all of the V gene segments are flanked by RSSs with the same size spacer length, and all of the J gene segments are associated with RSSs of the opposite type, hence recombination only happens between V and J gene segments. For example, at the *Ig κ* locus, all V gene segments are flanked by 23-RSSs and all J gene segments are flanked by 12-RSSs. In loci consisting of V, D and J gene segments, the arrangement of RSSs can be in two different ways. For example, at the *TCR β* and δ loci, V gene segments are flanked by 23-RSSs, D gene segments are associated with two RSSs with dissimilar spacer length (a 5' 12-RSS and a 3'

23-RSS), and J gene segments are flanked by 12-RSS. However, at the mouse *IgH* loci, both V and J gene segments are associated with 23-RSSs, and D gene segments are associated with two RSSs with same size spacer (a 3' 12-RSS and a 5' 12-RSS). This is an important feature in the regulation of V(D)J recombination (Tonegawa, 1983), since it ensures only productive assembly of gene segments are permitted, such as D-to-J, thus avoiding non-productive rearrangements such as V-to-V. Nevertheless, these non-productive rearrangements do occur, albeit at low frequencies (Parkinson et al., 2015). Furthermore, not all 12- and 23-RSSs are recombined with equal efficiency, for example, little to no direct recombination occurs between V β 23-RSSs and J β 12-RSSs in the *TCR β* locus, even though this combination being permissible (Drejer-Teel et al., 2007).

1.5 Overview of V(D)J recombination

V(D)J recombination is an irreversible genetic recombination reaction that is responsible for generating antigen receptor diversity during the early stages of B and T cell maturation (Hozumi & Tonegawa, 1976; Roth, 2014). V(D)J recombination can also catalyse aberrant DNA rearrangements in developing lymphocytes, which can be life-threatening if they trigger the genesis of common lymphoid neoplasms (Mullighan et al., 2008; Papaemmanuil et al., 2014; Tsujimoto et al., 1985; Zhang et al., 2012).

The process of V(D)J recombination can be physiologically separated into two distinct phases (Figure 1.8): a cleavage phase which is mediated by lymphoid-specific RAG proteins, RAG1 and RAG2 (Oettinger et al., 1990; McBlane et al., 1995) and a joining phase which is mediated by the non-homologous end joining (NHEJ) pathway of DNA double-strand break (DSB) repair (Schatz & Swanson, 2011).

RAG proteins are expressed by recombination activating genes, *Rag1* and *Rag2*, respectively, mainly in developing lymphocytes and therefore, the cleavage phase is lymphoid-specific (Fugmann et al., 2000a; Gellert, 2002; Oettinger, 1999). During this phase, RAG proteins bring the 12-RSS and the 23-RSS into close vicinity, forming a stable synaptic complex and introducing a

single-strand nick precisely in the DNA between the coding flank and the first base of the RSS (before the heptamer). This process is essentially energetically neutral and gives rise to a 5' phosphate group and a free 3' hydroxyl group on the same strand. The resulting reactive 3'OH group then attacks the phosphodiester bond of opposite strand in a direct transesterification reaction (McBlane et al., 1995), a reaction catalyzed by the recombinase, resulting in the formation of DNA double-strand breaks (DSBs) at each RSS (Fugmann et al., 2000b). This transesterification reaction generates two DNA ends: a blunt 5' phosphorylated end on the signal segment which, in some cases, is important for maintenance of the linear integrity of the chromosome (Bassing et al., 2002; Beth et al., 2012) and a covalently sealed hairpin structure on the coding segment which is essential for antigen receptor gene assembly (Beth et al., 2012; Gellert, 2002; Schatz & Swanson, 2011).

Although nicking can occur individually on either RSS, in the presence of the physiological divalent metal ion Mg^{2+} , transesterification requires assembly of a synaptic complex that includes both a 12- and a 23-RSS (Hiom & Gellert, 1997; Roth, 2014), providing a molecular basis for the 12/23 rule. DNA nicking and hairpin formation occurs on both strands simultaneously in a complex which is known as a synaptic cleavage complex (Agrawal et al., 1997; Curry et al., 2005; Fugmann et al., 2000a; Schatz & Ji, 2011). After cleavage, RAG proteins transiently maintain both the hairpinned coding ends and blunt signal ends together as a stable post-cleavage complex (PCC; Agrawal & Schatz, 1997; Hiom & Gellert, 1998; Fugmann et al., 2000a). These single broken DNA ends are then repaired by NHEJ pathway (Figure 1.9).

For the blunt signal ends, the repair reaction starts with the a Ku70-Ku80 dimer binding to the 5' phosphorylated DNA ends (Gu et al., 1997). Ku70-Ku80 then recruits DNA repair protein, DNA ligase IV:XRCC4 complex, to ligate the blunt ends together into a circular extrachromosomal piece of DNA, which contains all of the intervening sequences from between the coding segments, known as an excised signal circle (ESC) and normally defined as the by-product of V(D)J recombination (Grawunder et al., 1997; Lee et al., 2004; Lieber, 2010).

The coding ends are processed further prior to their ligation by several events that ultimately lead to junctional diversity (Lewis et al., 1994). Initially, Ku70-Ku80 binds to the DNA ends (Gu et al., 1997) and recruits DNA-dependent protein kinase catalytic subunit (DNA-PKcs), forming a complex that leads to DNA-PKcs auto-phosphorylation (Gottlieb & Jackson, 1993), which in turn recruits and contributes to activation of XRCC4 and Artemis (Blier et al., 1993; Helmink et al., 2012; Swanson, 2004). The DNA-PKcs:Artemis complex then opens the hairpin structures of the coding end at a random site (Ma et al., 2005). If they are opened at the apex of the hairpin, a blunt DNA end will be formed, but in many cases, the opening is "off-center". In this case, an overhanging flaps will be formed. The extra bases remaining on one strand are known as palindromic (P) nucleotides due to the palindromic nature of the sequence produced when DNA repair enzymes resolve the overhang (Figure 1.10; Lu et al., 2007). Next, XRCC4, Cernunnos, and DNA-PKcs align the DNA ends and recruit terminal deoxynucleotidyl transferase (TdT), a template-independent DNA polymerase that adds non-templated (N) nucleotides to the coding ends, providing further variation (Figure 1.8; Komori et al., 1993; Mahajan et al., 1999). The addition is mostly random, but TdT does exhibit a preference for G/C nucleotides (Gauss & Lieber, 1996). Then, following pairing of the two coding ends, exonucleases remove bases from the coding ends and this may include any P or N nucleotides that may have formed. DNA polymerases λ and μ (Pol λ and Pol μ , respectively) then insert additional nucleotides as needed to make the two ends compatible for joining. Finally, the processed coding ends are ligated together by the DNA ligase IV:XRCC4 complex (van Gent & van der Burg, 2007; Lieber, 2010). Notably, two thirds of recombination events are joined out of frame because of the imprecise joining of these processed coding ends, which is the price that is paid to ensure the generation of a highly diverse antigen receptor repertoire.

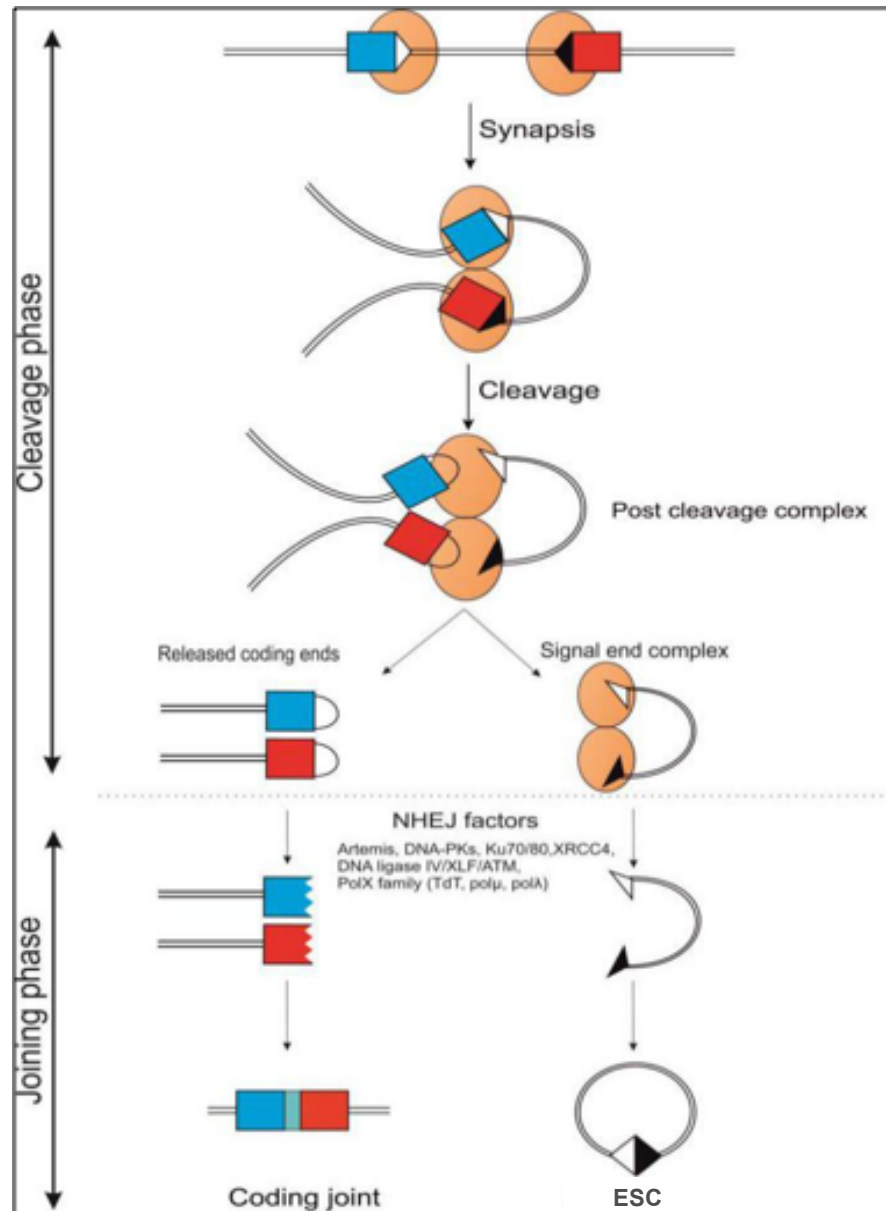


Figure 1.8 Overview of the cleavage and joining phases of V(D)J recombination

The cleavage phase is initiated by binding of RAG proteins (orange ovals) to an RSS (triangle), which then capture an appropriate partner RSS to form a synaptic complex. Upon synapsis, RAG proteins complete nicking and hairpin formation, generating blunt signal ends and hairpinned coding ends. After cleavage, the signal ends are still bound to the RAG proteins in a signal end complex until repaired by the non-homologous end joining (NHEJ) machinery. Coding ends are released from RAG proteins and imprecisely repaired by NHEJ pathway (indicated by light blue). Ligation of the signal joint results in an episomal circle – the excised signal circle (ESC) (Adapted from Schatz & Swanson, 2011).

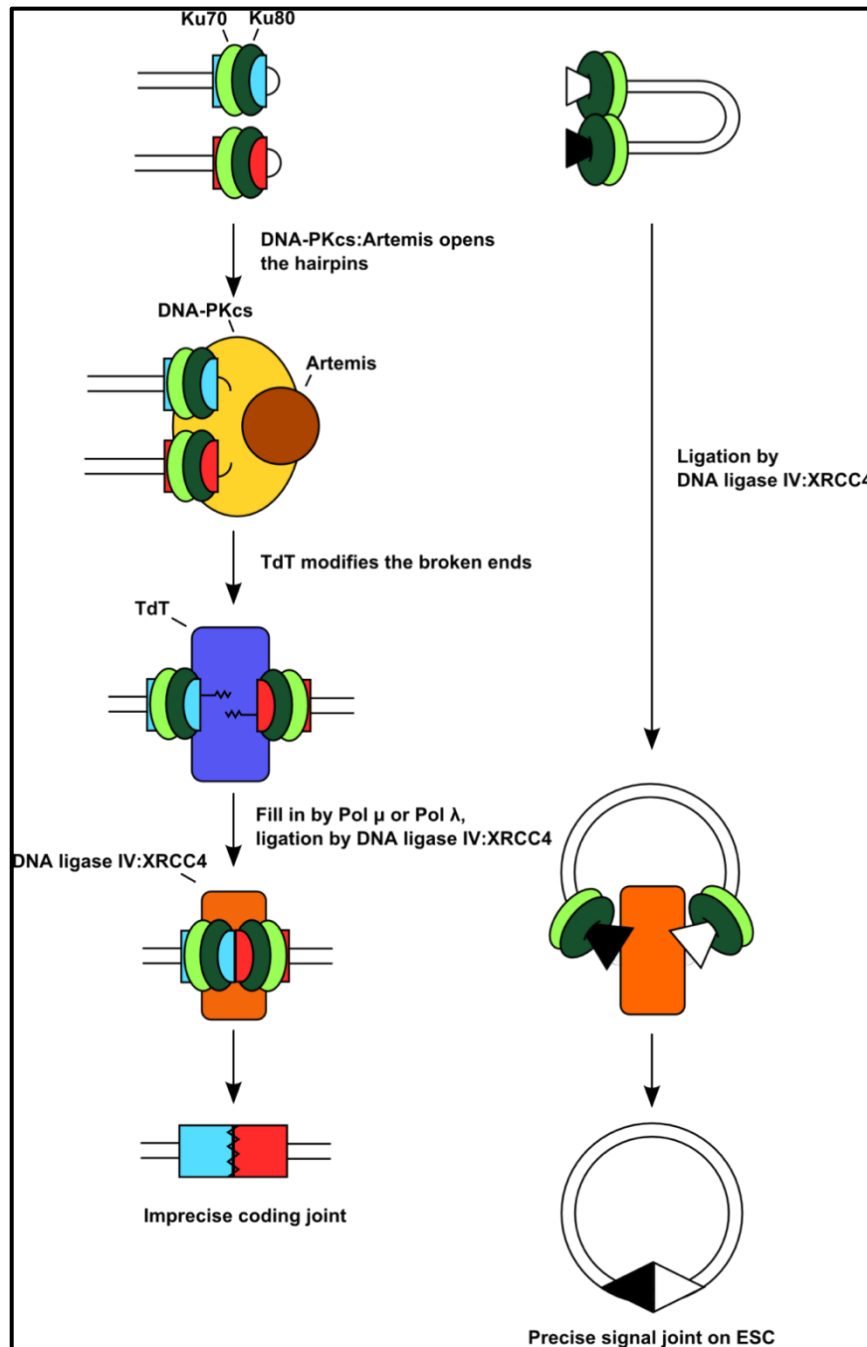


Figure 1.9 Non-homologous end joining pathway

A Ku70-Ku80 dimer binds each of the broken ends. The signal ends are ligated directly by DNA ligase IV:XRCC4, resulting in a signal joint on an excised signal circle (ESC). The joining of coding ends is more complicated. Firstly, DNA-PKcs:Artemis complex is recruited by Ku70-Ku80 dimer, opening the hairpinned DNA ends. Then, a lymphoid-specific polymerase (TdT) is recruited to add non-templated nucleotides, creating further junctional variation. Next, Pol λ or Pol μ fill in any remaining gaps, making the two ends compatible for joining. Finally, the DNA ligase IV/ XRCC4 complex ligates the ends together to create a coding joint (Adapted from Murphy et al., 2007).

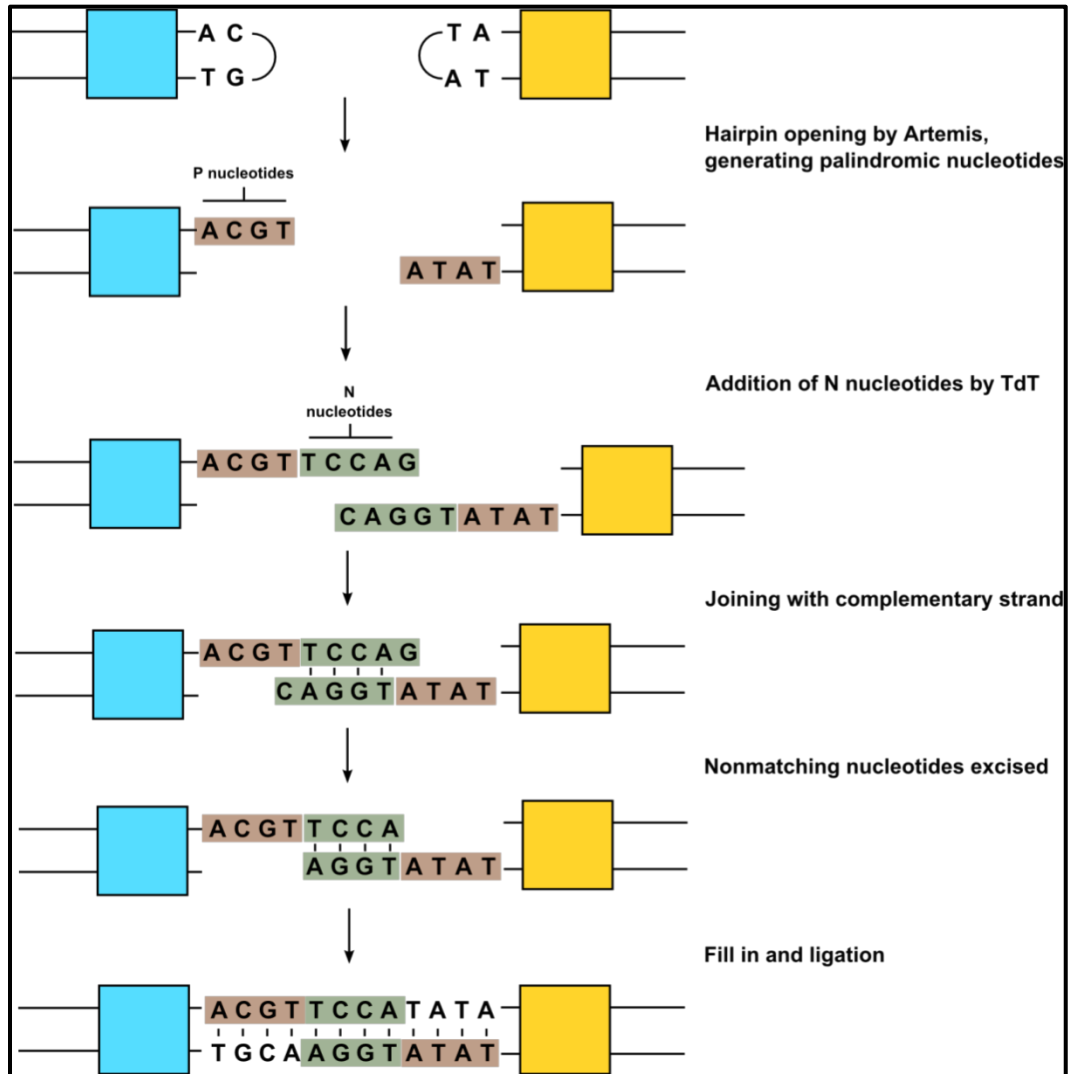


Figure 1.10 Generation of junctional diversity by adding P and N nucleotides

When DNA-PKcs:Artemis opens a hairpinned coding end in a “off-center” way, an overhang known as palindromic (P) nucleotides is formed (indicated by orange boxes). In addition, the lymphoid-specific polymerase TdT is recruited to add non-templated (N) nucleotides randomly (indicated by green boxes). Following this, two strands are annealed by complementary base pairing, and any additional nucleotides that are not matched are removed by exonucleases. This addition of P and N nucleotides further increase antigen receptor diversity.

1.6 The RAG proteins

The proteins, encoded by lymphoid-specific recombination activating genes 1 and 2 (*Rag1* and *Rag2*), play a key role in V(D)J recombination (Figure 1.11) (Schatz et al. 1989; Oettinger et al. 1990), cooperating with the non-lymphoid-specific DNA binding factor, HMGB1 or HMGB2 (Section 1.7), to initiate V(D)J recombination by carrying out DNA cleavage (van Gent et al., 1997). Both RAG1 and RAG2 are expressed predominantly in early B and T cells and are essential for the generation of mature B and T lymphocytes (Lieber, 2010). RAG1 was first identified by the finding that V(D)J recombination could be initiated by transfecting *Rag1* cDNA into fibroblasts, although the recombination efficiency was rather low (Schatz et al., 1989). This was explained by the subsequent observation that the co-transfection of *Rag1* and *Rag2* increases the recombination efficiency *in vitro* by over 1000-fold compared to with *Rag1* or *Rag2* alone (Oettinger et al., 1990). Because *Rag1*-deficient and *Rag2*-deficient mice and humans exhibit identical phenotypes characterized by a complete absence of V(D)J recombination and lymphocyte development (Mombaerts et al. 1992; Shinkai et al. 1992), the two RAG proteins are presumed both to be essential for V(D)J recombination. In addition, it has been documented that mutation of *Rag1* or *Rag2* results in severe combined immunodeficiency (SCID) and less severe Omenn syndromes (Notarangelo et al. 1999; Jones & Simkus, 2009). Furthermore, apart from being important for the rearrangement of antigen receptor genes, RAG-mediated DSBs could impact cellular processes and outcomes by affecting genetic and epigenetic programs (Lescale & Deriano, 2017).

RAG1 and RAG2 are greatly conserved in all jawed vertebrates. It was suggested that they were introduced into the vertebrate genome by horizontal gene transfer from an unknown donor organism (Chatterji et al., 2004; Fugmann, 2010; Thompson, 1995). It is proposed that RAG1 originates from an ancient transposable element of the *Transib* transposon (Fugmann, 2010; Kapitonov & Jurka, 2005; Thompson, 1995), although definitive evidence for this remained elusive (Litman et al., 2010). A more recent study provided a further evidence for the origins of *Rag1* by testing the ability of two RAG1-

related proteins to initiate V(D)J recombination (Carmona et al., 2016). In this study, the investigators employed both the H_ztransib transposase from *H. zea* and the purple sea-urchin RAG1-like protein and showed that they cannot initiate V(D)J recombination alone, whilst, recombination of an extrachromosomal substrate can efficiently occur by co-transfection of mouse *Rag2*, with a level only three-fold lower than with mouse *Rag1* (Carmona et al., 2016). Consequently, the investigators suggest that RAG1 arose from the original acquisition of a *Transib* transposase gene, and that RAG2 was then obtained from a *Transib* transposable element to result in the *Rag1-Rag2* transposon. As a prerequisite for efficient V(D)J recombination, RAG2 was suggested to emerge in jawed vertebrates (Kapitonov & Jurka, 2005), the origin of *Rag2*, nevertheless, remains unknown (Carmona et al., 2016).

Being located on the same chromosome, approximately 8 kb apart in opposite transcriptional directions, the *Rag1* and *Rag2* genes are each encoded by a single exon (Oettinger et al., 1990; Schatz et al., 1992). The conserved domains of both RAG1 and RAG2 are shown in Figure 1.9. Mouse full-length RAG1 is composed of 1040 amino acids whilst full-length RAG2 consists of 527 amino acids, with molecular weights of 119 kDa and 58 kDa, respectively. The “core” regions of RAG proteins are sufficient for catalyzing V(D)J recombination and were frequently used *in vitro* as they are much easier to purify and are the minimum regions of both proteins that are required for RSS cleavage *in vitro* and *in vivo* (Dudley et al., 2003; Sadofsky et al., 1993; Sadofsky et al., 1994). The “core” region of RAG1 contains amino acids 384 to 1008 (Silver et al., 1993; Sadofsky et al., 1994), whereas the “core” region of RAG2 has been defined as amino acids 1 to 387 (Cuomo et al., 1994; Sadofsky et al., 1994). Moreover, non-core region of RAGs were also found important for effective regulation of V(D)J recombination *in vivo* (Akamatsu et al., 2003; Curry & Schlissel, 2008; Dudley et al., 2003; Liang et al., 2002), as discussed later in this chapter (Section 1.6).

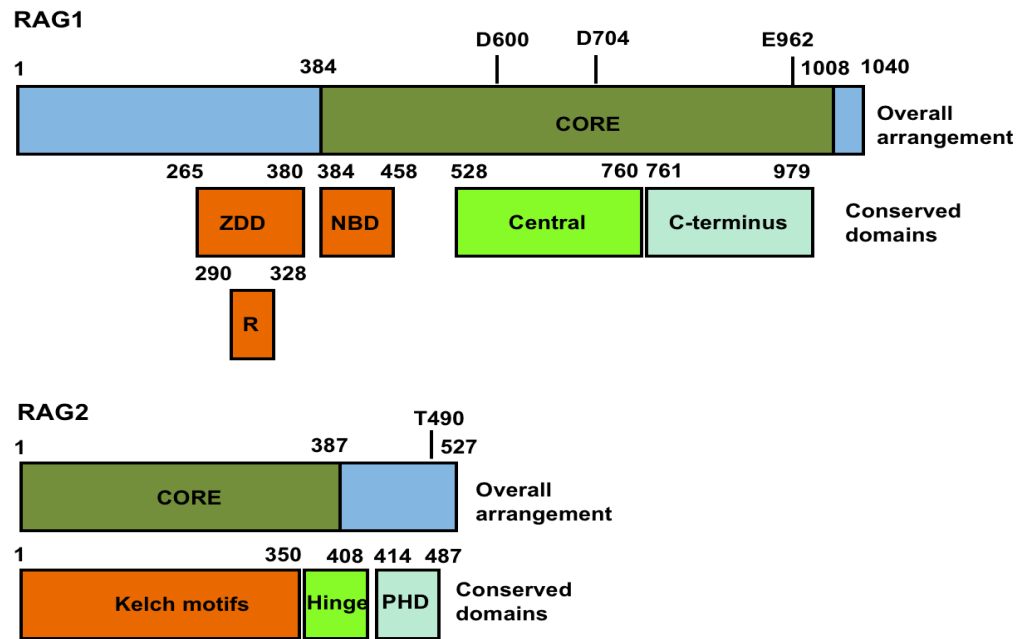


Figure 1.11 The core and non-core domains of RAG proteins

The RAG proteins can be functionally divided into core and non-core domains. Regions of RAG1 and RAG2 required (core) or dispensable (non-core) for supporting basic recombination activity are shaded dark green or blue, respectively. The conserved domains within these proteins and their location are shown underneath the maps of full-length RAG1 and RAG2. RAG1 directly binds to DNA via its nonamer binding domain (NBD), central domain and C-terminus domain. RAG2 mainly functions as a cofactor by enhancing sequence-specific recognition of the RSS by RAG1 and increasing RAG1 cleavage activity via its kelch motifs. The C-terminus of RAG2 contains plant homeodomain (PHD) finger that recognises and binds to tri-methylated histone H3 lysine 4 (H3K4me3). The non-core and core regions of RAG2 are connected by a flexible acidic hinge. R = RING (Really Interesting New Gene) domain, ZDD = Zinc-binding Dimerisation Domain (Adapted from Swanson & Schatz, 2011).

RAG1 and RAG2 play vastly different roles in the DNA cleavage reaction. RAG1 is the workhorse for DNA cleavage (Teng et al., 2015). The core region of RAG1 contains a highly conserved catalytic triad of acidic residues termed a DDE motif (residues D₆₀₀, D₇₀₈ and E₉₆₂), which form the major active site and is essential for nicking the DNA strand by coordinating a divalent metal cofactor and forming the DNA hairpin (Fugmann et al., 2000a; Kim et al., 1999; Landree et al., 1999). *In vitro* RAG cutting assays showed that both Mg²⁺ and Mn²⁺ allow RAGs to cleave, but, the metal ion used *in vivo* is most likely to be Mg²⁺, since RAG cleavage obeys 12/23 rule in the presence of Mg²⁺ but not Mn²⁺ (van Gent et al., 1996). The core region of RAG1 is also structurally separated into three different domains: the N-terminal nonamer binding

domain (NBD), the central domain and the C-terminal domain. The NBD directly binds to the AT-rich nonamer of 12- and 23-RSSs (Difilippantonio et al., 1996; Spanopoulou et al., 1996; Yin et al., 2009). In addition, crystallization studies indicated that two RAG1 NBDs extensively contact with each other and with the two nonamers, which suggests that the NBD also participates in facilitating the formation and stability of synapsis complex between two RSSs (Yin et al., 2009). Moreover, the cryo-electron microscopy (cryo-EM) structure of RAG proteins resolved by Ru et al showed that the nonamer binding domain of RAG1 is also involved in the enforcement of the 12/23 rule (Ru et al., 2015). The central domain specifically binds to the conserved heptamer of the RSSs (Peak et al., 2003), and displays nicking activity on single-stranded DNA. The C-terminal domain non-specifically binds to double-stranded DNA (Arbuckle et al., 2001).

The non-core region of RAG1 also plays a key role in regulation of V(D)J recombination *in vivo*, as well as in ensuring the efficiency and fidelity of the reaction (Jones & Simkus, 2009). The non-core N-terminus of RAG1 consists of a basic motif and a zinc-binding dimerization domain (ZDD, residues 265-380). The latter contains a RING (Really Interesting New Gene) finger domain (residues 290-328) and a classical C₂H₂ zinc finger domain (355-376). The basic motif participates in the interaction with karyopherin alpha 1 (KPNA1), facilitating nuclear localisation of RAG1 (Spanopoulou et al., 1995; Simkus et al., 2009). The RING domain, the prototypical member of E3 ubiquitin (Ub) ligases, is involved in the autoubiquitination of K233, potentially as well as the ubiquitination of KPNA1 (Jones & Gellert, 2003; Simkus et al., 2007; Simkus et al., 2009). It was suggested that the RING domain could recruit a multisubunit E3 Ub ligase complex via Vpr binding protein (VprBP), disrupting developmental arrest of B cells at the pro-B-to-pre-B transition (Kassmeier et al., 2012). Recombination products analysed in these cells displayed a greatly increased number of non-templated nucleotides and a higher mutation frequency (Kassmeier et al., 2012). Consequently, RAG1-mediated ubiquitination is likely to have an effect on the regulation of the joining of RAG cleavage products. The zinc finger domain has been proposed to mediate dimerization of RAG1 monomers (Bellon et al., 1997; Rodgers et al., 1996).

RAG2 functions as an essential cofactor by enhancing sequence-specific recognition of the RSS by RAG1 and increasing RAG1 cleavage activity (Swanson, 2004; Teng et al., 2015; Zhao et al., 2009). Core RAG2 is consist of a six-bladed β -propeller conformation (Kelch motifs), which are thought to mediate contacts with RAG1 (Aidinis et al., 2000; Calleebaut & Mornon, 1998). Consistent with this, a recent crystal structure study found that loops between these propeller domains interact with many regions of RAG1 (Kim et al., 2015). Remarkably, one of these interaction regions is located near the active site of RAG1, suggesting that RAG2 could promote the formation of the RAG1 active site and DNA cleavage (Kim et al., 2015). Cryo-EM studies further imply that the interaction between RAG1 and RAG2 could be further triggered by RSS binding (Ru et al., 2015). Moreover, this structure also suggests that RAG2 is involved in DNA binding to the coding gene region (Ru et al., 2015).

Whist the RAG2 C-terminus is not necessary for V(D)J recombination *in vitro*, it is very important for optimal V(D)J recombination *in vivo* (Akamatsu et al., 2003). The RAG2 C-terminus contains a PHD finger which recognises and binds to tri-methylated histone H3 lysine 4 (H3K4me3). This directs the RAG proteins to active chromatin (Elkin et al., 2005; Matthews et al., 2007; Ramon-Maiques et al., 2007) since H3K4me3 is a ubiquitous chromatin marker of active transcription start sites (TSSs) (Liu et al., 2007; Matthews et al., 2007; Teng et al., 2015). The RAG2 non-core region at the protein's C-terminus also plays a role in regulating genome stability (Curry & Schlissel, 2008; Deriano et al., 2011; Shimazaki et al., 2012). The RAG2 C-terminus is connected to the core domain via a flexible acidic hinge which appears to maintain the stability of the RAG post-cleavage complex (Coussens et al., 2013; West et al., 2005). Indeed, the absence of the RAG2 C-terminus leads to aberrant recombination events, such as transpositions (Curry et al., 2007; Curry & Schlissel, 2008; Talukder et al., 2004). In addition, loss of the RAG2 C-terminus results in increased levels of double-strand breaks and chromosome translocations by impairing proper end joining (Steen et al., 1999; Coussens et al., 2013). Specifically, it increases the chances of the broken DNA ends being joined via the error-prone alternative non-homologous end joining rather than the classical NHEJ pathway (Corneo et al., 2007; Talukder et al., 2004). The RAG2 C-terminus therefore is important for maintaining genome stability

(Coussens et al., 2013; Derianol et al., 2011; Gigi et al., 2014; Zhang et al., 2011).

The RAG proteins bind to RSSs to initiate V(D)J recombination as a complex of a RAG1 dimer and a RAG2 dimer (Swanson, 2002a). Initially, it was thought two RAG1/RAG2 heterodimers assemble on two different RSSs separately before forming a synaptic complex. However, a later study suggests that this heterodimer firstly assembles on one RSS, introduces nicking of one strand before capturing the partner RSS (Eastman & Schatz, 1997; Mundy et al., 2002; Swanson, 2002a). Both RAG1 subunits are necessary for the binding of each RSS: one subunit binds the start positions of the heptamer, and the other binds the more terminal positions of the RSS (Ru et al., 2015). In addition, for the coding end DNA, the RAG1-RAG2 monomer binding is exclusive to one RSS: the start positions are bound by RAG1, whilst the more terminal regions are bound by RAG2 (Ru et al., 2015)

1.7 High mobility group box (HMGB) proteins in V(D)J recombination

The high mobility group box (HMGB) proteins are some of the most evolutionarily conserved proteins (Tsong et al., 2014). HMGB1 is a member of the HMGB protein family with a molecular weight of 25 to 30 kDa in mammals (Bianchi & Manfredi, 2007). In mice, HMGB1 consists of 215 amino acids and three domains including two DNA binding domains, HMGB A box (9-79 aa) and HMG B box (95-163 aa), as well as the C-terminal acidic tail (186-215 aa; Kang et al., 2014). It is an essential and ubiquitous non-specific DNA architectural factor that can bind in the minor groove of DNA, facilitating DNA bending towards the major groove (Malarkey & Churchill, 2012; Weir et al., 1993).

Whilst RAG proteins by themselves are able to carry out DNA binding and cleavage at RSSs *in vitro* (McBlane et al., 1995), further investigations found that HMGB proteins can stimulate the cutting of RSSs *in vitro* by 7- to 100-fold in a dose-dependent manner (Sawchuk et al., 1997; van Gent et al., 1997). In addition, the presence of HMGB proteins, including HMGB1 or HMGB2, can

decrease the K_d of the RAG complex for both 12-RSS and 23-RSS, facilitating the formation of a 12/23 complex (Lovely et al., 2015).

Several studies found that cleavage by the RAG proteins alone is more efficient at a 12-RSS than a 23-RSS. However, cleavage is increased, particularly at a 23-RSS, by adding HMGB1 or HMGB2 (McBlane et al., 1995; Sawchuk et al., 1997). In addition, it has been shown that, during V(D)J recombination, HMGB1 is recruited by interacting with a 23-RSS, inducing a bend in the conserved spacer. This suggests that DNA bending contributes by bringing the conserved heptamer and nonamer closer together, facilitating RAG binding (van Gent et al., 1997; Little et al., 2013). Consistent with this, single molecule fluorescence experiments showed that HMGB1 and RAG proteins cooperate to bring the 23-RSS heptamer and nonamer closer together (Ciubotaru et al., 2013). HMGB1 has also shown to stabilize the bend of the 12-RSS heptamer and localize the RAG complex to its target cleavage site at 5' of the heptamer (Zagelbaum et al., 2016). In addition, footprint assays found that, in the presence of HMGB1, the RAG/23-RSS complex resembles the RAG/12-RSS complex more closely (Yoshida et al., 2000; Swanson, 2002a). Furthermore, the cryo-electron microscopy studies of the mouse and zebrafish RAG/RSS complex showed that a single HMG box (likely Box A) binds to the 12-RSS spacer, close to the nonamer-spacer boundary by bending $\sim 60^\circ$, whereas two HMG boxes bind to the 23-RSS spacer by bending $\sim 120^\circ$, such that the additional binding falls in the middle of the spacer to induce a large DNA bend (Kim et al., 2015; Ru et al., 2015).

HMGB1/2 has been shown to be essential to facilitate V(D)J recombination *in vitro*. However, there is little to no evidence showing that HMGB1/2 is necessary *in vivo*. When the HMGB1 was added to a RAG cleavage assay where a 12-RSS was packaged in a nucleosome, the cleavage efficiency is significantly increased, suggesting that HMGB1/2 can increase RAG accessibility to RSSs *in vivo* (Kwon et al., 1998). Furthermore, consistent with this, an *in vivo* recombination assay found that the recombination efficiency of an extrachromosomal substrate increased by more than three-fold upon transfection of extra HMGB1/2, further implying that HMGB proteins are able to

facilitate V(D)J recombination *in vivo* (Aidinis et al., 1999). In contrast, however, one study found that the levels of antigen receptors is normal in HMGB1 knockout mice, indicating that HMGB proteins are not necessary for V(D)J recombination *in vivo* (Calogero et al., 1999). It is possible, however, that HMGB2 which is expressed at normal levels in these HMGB1 knockout mice, could probably make up for the deficiency of HMGB1 (Calogero et al., 1999). In agreement with this, an *in vitro* RAG cleavage showed that the cleavage efficiency was stimulated to similar levels by adding either HMGB1 or HMGB2 (van Gent et al., 1997).

1.8 Assembly of RAG-RSS complexes

RAG1 binds an isolated RSS with a K_d of 29-41 nM, measured by fluorescence anisotropy (Ciubotaru et al., 2003), or 92-114 nM, when measured by electrophoretic mobility shift assay (EMSA) (Rodgers et al., 1999; Zhao et al., 2009) in the absence of RAG2 by forming a stable dimer (Swanson & Desiderio, 1999). RAG1 can specifically and non-specifically contact the DNA phosphate backbone in the centre of the nonamer region of the RSS, forming a RAG1-RSS complex. The binding also can extend into the spacer region by several base pairs (Swanson & Desiderio, 1998). However, in the presence of RAG2, RAG1 binds a single RSS with a K_d of 12-25 nM, measured by fluorescence anisotropy and EMSA (Wang et al., 2012; Zhao et al., 2009). RAG protein binding to a single RSS forms two main complexes *in vitro*, termed SC1 (single RSS containing complex 1) and SC2 (single RSS containing complex 2), respectively (Swanson, 2002b) (Figure 1.12). SC1 contains two molecules of RAG1 and one molecule of RAG2 whilst SC2 contains two molecules of RAG1 and two molecules of RAG2 (Swanson, 2002b). Notably, in these complexes, RAG2 molecules can freely-re-assort, exchanging without completely dissociating from the complex (Swanson, 2002b). Furthermore, in the presence of both RAG1 and RAG2 plus HMGB1, RAG proteins bind a single 12- or 23-RSS with K_d of 4.1 ± 1.6 nM or 8.8 ± 4.0 nM, respectively (Lovely et al., 2015). In the presence of HMGB1, RAG protein binding to a single RSS forms HMGB1-SC1 (HSC1) and HMGB1-SC2 (HSC2), respectively (Figure 1.12). Nicking of an RSS can take place in both HSC1 and HSC2, independent of synapsis with a partner RSS (Eastman & Schatz, 1997;

Swanson, 2002a; Swanson, 2002b). A second RSS can combine with HSC2 without recruiting additional RAG proteins, forming a paired complex (PC; Swanson, 2002b; Figure 1.12).

In addition, RAG1 molecules bind to the nonamer of one RSS, but cut the heptamer of the partner RSS (Yin et al., 2009; Swanson, 2001), since the RAG1 active site functions *in trans*. The interactions between RAG and DNA are further extended to the region containing the heptamer/coding boundary and the whole spacer in the presence of RAG2 (Swanson & Desiderio, 1998). However, in contrast to studies that showed that a synaptic complex includes two molecules of RAG1, it was suggested by several other studies that a synaptic complex contains three or more RAG1 molecules (Mundy et al., 2002; Mo et al., 1998). Complexes with this stoichiometry, however, are specific to RAG proteins expressed with GST tags which could contribute to oligomerisation of its protein conjugates (Niedziela-Majka et al., 1998). Consequently, the accepted dogma is that the RAG-RSS PC contains a 12-/23-RSS pair, with two molecules each of RAG1 and RAG2 (Swanson, 2004).

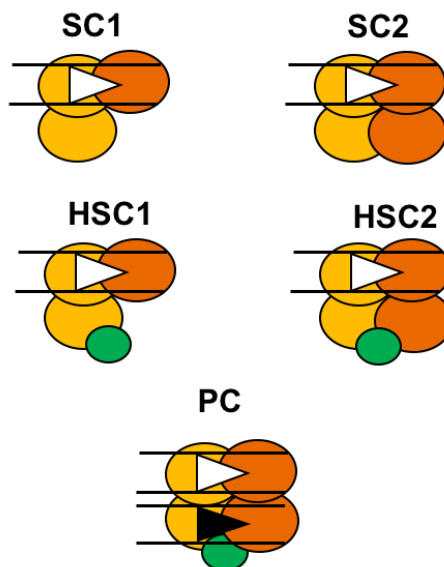


Figure 1.12 Schematic of RAG-RSS complexes

In the absence of HMGB1, RAG proteins binding to a single RSS forms single complex 1 (SC1) and single complex 2 (SC2). SC1 contains two molecules of RAG1 and one molecule of RAG2 whilst SC2 contains two molecules of RAG1 and two molecules of RAG2. Addition of HMGB1 results in the formation of HMGB1-SC1 (HSC1) and HMGB1-SC2 (HSC2). RAG heterotetramer can form on a single RSS, and incorporation of a second RSS into HSC2 results in the formation of a paired complex (PC). Open and closed triangles = 12- or 23-RSSs, yellow oval = RAG1, orange oval = RAG2, green oval = HMGB1.

There are two possible models for the formation of a PC containing two RSSs: RAG association and RSS capture (Swanson, 2004). In the association model, RAG proteins bind to two RSSs separately, forming two distinct SCs, and then these two SCs associate via protein-protein interactions. In the RSS capture model, RAG proteins, with the assistance of HMGB1, bind a single RSS first, forming SC2, and then capture a second RSS, without the addition of any extra RAG molecules (Swanson, 2004). RSS capture is the generally accepted model, for two reasons. Firstly, both PC and SC2 include two molecules of RAG1 and two molecules of RAG2 (Swanson, 2002b; Mundy et al., 2002), and thus SC2 assembled on one RSS can be potentially altered to form a PC by adding a partner RSS (Mundy et al., 2002). Secondly, addition of 12- and 23-RSS step by step could greatly increase the cleavage efficiency compared to mixing the separately assembled 12- and 23-RSS complexes (Jones & Gellert, 2002).

It was documented that when the RAGs bind to a 12-RSS first, the efficiency of PC formation is much higher than when the 23-RSS is bound first (Jones & Gellert, 2002). In addition, it was found that nicking occurs at endogenous 12-RSSs at the *IgH*, *Igκ* and *TCRα* loci in lymphoid cells, whilst the 23-RSSs are not nicked (Curry et al., 2005). According to these findings, it was initially thought that the RAG proteins show a preferential binding to a 12-RSS and subsequently capture a 23-RSS partner. Nevertheless, this preferential binding is considered to be unlikely for a number reasons. In an early study, a recombination assay was carried out using a truncated HMGB1 where the acidic C terminus is absent (Jones & Gellert, 2002). Crucially, the acidic tail functions in increasing the DNA-bending activity of HMGB1 (Stors, 1998). Because HMGB1 is not necessary for 12-RSS binding, the lack of acidic tail is unlikely to affect initial 12-RSS binding but could have an influence on subsequent 23-RSS binding.

The “12 first” model of RSS capture has also been questioned further by two later studies. The first study looked at endogenous RSS nicking at the *TCRβ* locus *in vivo* and *in vitro* and found that the nicking frequency of 23-RSSs associated with Dβ1 gene segments was much higher than that of 12-RSSs

associated with J β gene segments (Franchii et al., 2009), suggesting that the RAG proteins are likely to bind to the 23-RSSs first at the *TCR β* locus and then subsequently capture a 12-RSS partner. Furthermore, because the sequence of the RSS, as well as the coding gene segments can also influence the nicking rate of RSS, measurement of RSS nicking seems not to be a reliable way to determine RAG binding (Cowell et al., 2004; Feeney et al., 2004; Gerstein & Lieber, 1993; Yu & Lieber, 1999).

Notably, in a more convincing chromatin immunoprecipitation sequencing (ChIP-seq) study, RAG1 and RAG2 binding at the antigen receptor loci was determined directly in developing primary lymphocytes and the results showed that RAG proteins don't show preference for either 12- or 23-RSSs (Ji et al., 2010). Instead, it is the location of the RSS that plays a role in determining RAG binding. If the RSS is located in active chromatin, where recombination centres form, RAG binding is much more likely to happen, since in these regions, RNA polymerase II and H3K4me3 accumulate to high levels, both of which are essential for the regulation of V(D)J recombination. Thus, as previously proposed, the favoured model of RAG-RSS binding *in vivo* is RSS capture, and not RAG association. However, there are several factors that have an effect on the order of RSS binding, such as the sequence of the flanking coding gene segment, the closeness of the heptamer and nonamer sequence to the consensus sequence and the chromatin environment of the RSS (SSadofsky et al., 1995; chatz & Ji, 2011; Ji et al., 2010).

1.9 RAG cleavage of RSSs

RAGs cleave RSSs by generating double-strand breaks through two consecutive steps of catalysis: nicking and hairpin formation (Figure 1.13; Fugmann et al., 2000b; Gellet, 2002; Oettinger, 1999). Nicking can occur at either a single RSS or a paired RSS. After forming a paired complex (PC), RAGs interact with the heptamer and adjacent coding sequence much more closely and assume a more open conformation, thus avoiding a steric clash with intact RSS substrates (Ru et al., 2018). Moreover, when the PC is formed, DNA is distorted, exposing the cleavage site (Akamatsu & Oettinger, 1998;

Nagawa et al., 1998). Because of the preferential recognition of RAG1 for a single-stranded heptamer sequence, it is thought that after binding to an RSS, the RAG proteins generate a nick at the boundary between the heptamer and coding end and then induce localised base-pair melting at that region (Schatz & Swanson, 2011). Consistent with this, it was found that abasic residues or base-pair mismatches in the coding flank could stimulate RAG cleavage (Cuomo et al., 1996; Grundy et al., 2007; Ramsden et al., 1996). However, in contrast to this, a recent cryo-EM study showed that, prior to formation of the nick, the intact RSS substrates are melted at the CAC/GTG base pair of the heptamer, which is facilitated by the piston-like movement of the isolation domain (ID; aa 738-976) of RAG1 and the RAG dimer opening (Ru et al., 2018). Subsequently, RSS melting, coupled with corkscrew DNA rotation, result in placement of the scissile phosphate into the active site of RAG1, thus allowing the formation of a DNA nick (Ru et al., 2018). Following nicking of both RSSs, the two RAG monomers rotate relative to each other, forming a fully closed conformation (Kim et al., 2015; Ru et al., 2015; Ru et al., 2018). The RAG dimer closure, highly distorted nicked DNAs, and extensively recognized CAC/GTG of the heptamer then facilitate hairpin formation at the coding flank (Ru et al., 2018).

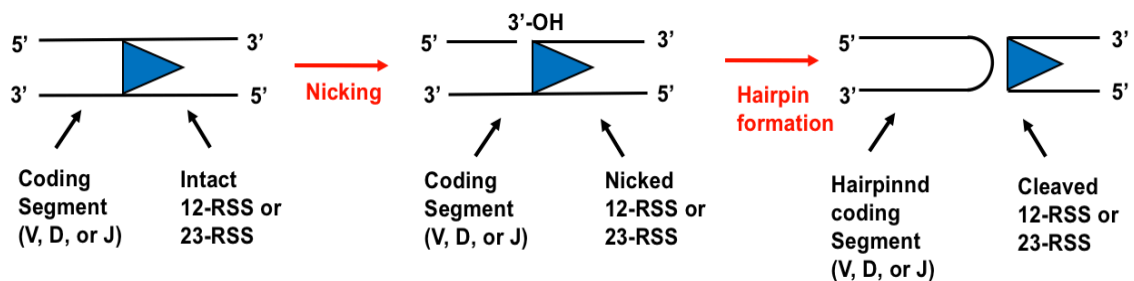


Figure 1.13 Schematic representation of RAG-mediated catalysis steps

The RAG proteins cleave RSSs in two consecutive steps: nicking and hairpin formation. A nick is introduced at the 5' end of the RSS (shown as blue triangle) heptamer, leaving a 3'-OH on the flanking coding gene segment. The 3'-OH then attacks the opposite strand in a direct transesterification reaction to generate a blunt signal end and a hairpin coding end.

A vital step for hairpin formation by the related Tn5 transposases is the stabilisation of a “flipped” thymidine residue extruding from DNA helix, via assembling interactions with a conserved tryptophan residue (Davies et al., 2000; Ru et al., 2015). This contributes to the formation of the closed hairpin loop by nucleophilic attack of the nicked DNA on the complementary DNA

strand. Several studies found a similarly flipped thymidine residue in the RAG PC complex (Bischerour et al., 2009; Nishihara et al., 2008; Ru et al., 2015). Furthermore, four candidate aromatic residues (W893, W956, Y935 and F971), which are able to stabilize a flipped thymidine, have been confirmed in RAG1 by mutational studies (Bischerour et al., 2009; Lu et al., 2006; Grundy et al., 2007). It is proposed that all of these four residues are very important in base stacking, since there is no evidence to show that a single residue could completely satisfy all of the requirements for base flipping (Bischerour et al., 2009).

1.10 Basis of the 12/23 rule

Efficient V(D)J recombination takes place between a pair of recombination signal sequences with dissimilar spacer lengths, one with a 12-bp spacer (12-RSS) and the other with a 23-bp spacer (23-RSS). This restriction is the so-called 12/23 rule (Figure 1.14; Akamatsu et al., 1994; Early et al., 1980; Hesse et al., 1989; Lewis & Hesse, 1991; Sakano et al., 1980; Tonegawa, 1983), which ensures the correct joining of the recombining gene segments (West & Lieber, 1998). The recombination efficiency between two RSSs with dissimilar spacer lengths is much higher than that between two RSSs with the same spacer length, by approximately 30 times *in vivo* (Tonegawa, 1983). It has been documented that normal mechanism of V(D)J recombination involves assembly of a 12-/23-RSS pair into a synaptic complex prior to cleavage, suggesting that the 12/23 rule is enforced at or before the cleavage step of V(D)J recombination *in vivo* (Steen et al., 1996). In addition, consistent with this, several studies have indicated that, together with the help of HMGB1, the RAG proteins establish the 12/23 rule at a very early stage of recombination, possibly during the formation of synaptic complex (Fugmann et al., 2000b; West & Lieber, 1998). Moreover, it has been demonstrated that the presence of RSS partner with dissimilar spacer length dramatically stimulates hairpin formation (Hiom & Gellert, 1998; van Gent et al., 1996; Steen et al., 1996; West & Lieber, 1998). The fact that hairpin formation can only occur in the complex containing a pair of RSSs could partly explained by the fact that the RAG1 active site is contributed *in trans* (Yin et al., 2009; Swanson, 2001), as

the RSS is cleaved by the RAG1 that binds to partner RSS. Nevertheless, this does not explain why the two RSSs must be of different spacer lengths.

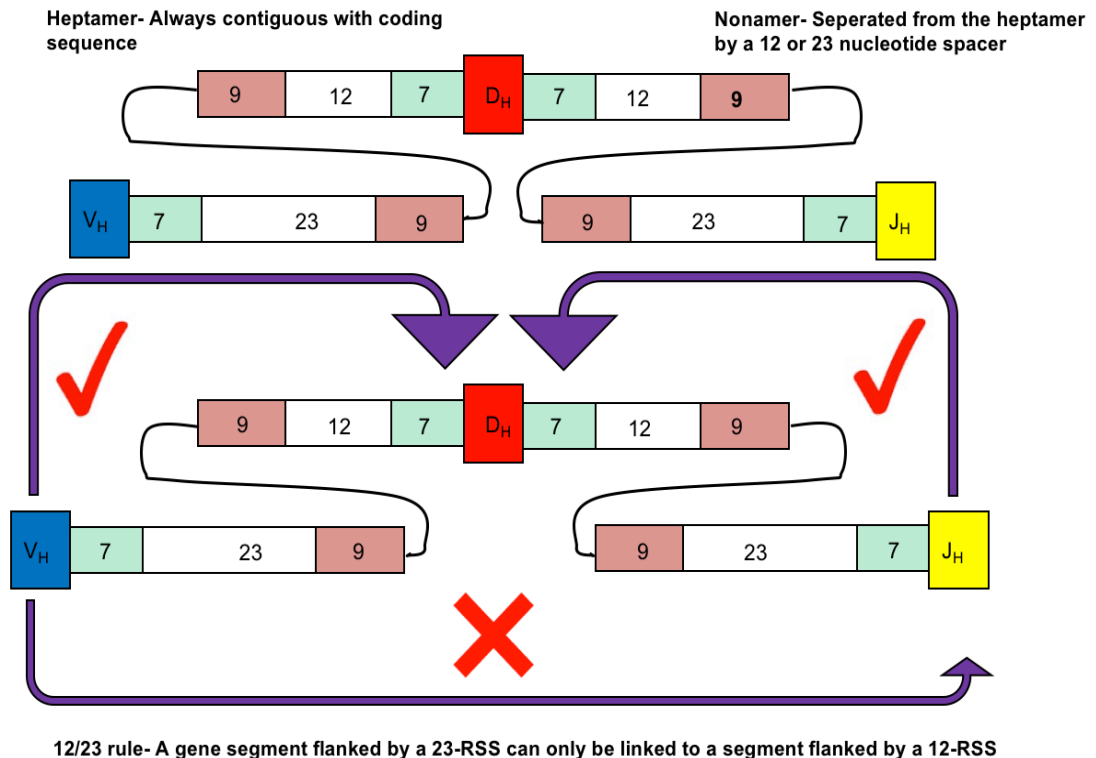


Figure 1.14 The 12/23 rule

During V(D)J recombination, efficient events only occur between a pair of RSSs with different spacer lengths, one with a 12-bp spacer (12-RSS) and the other with a 23-bp spacer (23-RSS). For instance, in the mice *IgH* locus, the recombination between D gene segment (shown as red rectangles) associated with a 5' 12-RSS and V gene segment (shown as blue rectangles) associated with a 3' 23-RSS is allowed, as well as the recombination between D gene segment (shown as red rectangles) associated with a 3' 12-RSS and J gene segment (shown as yellow rectangles) associated with a 5' 23-RSS. However, the recombination between V gene segment (shown as blue rectangles) flanked by a 3' 23-RSS and J gene segment (shown as yellow rectangles) flanked by a 5' 23-RSS is not permitted.

The Gellert group proposed that after binding to the first RSS, the RAG1-RAG2 dimer bound to the RSS forms either a 12-RSS-like or a 23-RSS-like isomeric conformation, according to which RSS is bound first (Jones & Gellert, 2002). When a RAG1-RAG2 dimer binds to a 12-RSS and generates a nick in the DNA within the SC2 complex, the nonamer binding domain of the free RAG1 has to search for an accessible 23-RSS partner in the presence of competition from other non-specific DNA and 12-RSSs. Possibly the free RAG1 in the 12-RSS SC2 complex seeks a nonamer sequence of an incoming

RSS, as that is the major RAG-RSS interactions domain. After binding to a second nonamer, the RAG1-RAG2 dimer then searches for a heptamer in the adjacent sequence. Because this RAG1/RAG2 heterodimer is held in a 23-RSS-binding conformation by the bound 12-RSS, the incoming heptamer sequence must be 23 bp away from the nonamer to tightly associate the RAG1 active site. Therefore, it is only likely for a PC to form two nicked substrates when it includes a 12/23 pair (Swanson et al., 2009). The nicked substrates are then converted to hairpin structures, which is also an important checkpoint in the 12/23 rule, as a flipped-out base is only detectable at a pre-nicked 12-RSS when there is a 23-RSS partner (Nishihara et al., 2008).

Once binding to the second nicked RSS occurs, the conformation of the complex changes to activate the active site of the RAG1 subunit that has already bound to the heptamer of the first RSS. This could possibly involve either the accurate alignment of one or more aromatic residues in the DDE motif and/or at the position which are suspected to stabilize a flipped base. Several studies found that two gain-of-function RAG1 mutants could catalyze efficient hairpin formation in Mg^{2+} without a corresponding partner (Kriatchko et al., 2006; Nishihara et al., 2008). In addition, the cleavage efficiency is not increased in the presence of partner. These results indicate that these two gain-of-function RAG1 mutants are defective in recognising synapsis of a 12/23 pair. Interestingly, both mutations are very close to the DDE motif, suggesting these residues could possibly participate in the conformational change process to ensure signal synaptic complex formation.

In 2015, Ru et al resolved the cryo-EM structure of the complex containing RAGs and two RSSs and provided the first convincing explanation of how the 12/23 rule is enforced (Ru et al., 2015). Flexibility of the NBD dimer and plasticity of RSS conformation are the two basic requirements for the 12/23 rule. The HMGB proteins (Section 1.7) assist in the plasticity of RSS conformation. The RAG proteins symmetrically contact the heptamer of both the 12- and 23-RSS at a fixed position. However, RAG binding to the nonamer of the 12- and 23-RSS is asymmetric. The RAGs bound one type of RSS change the conformation of the NBD dimer which then only enables binding of

the other type of RSS (Ru et al., 2015). When the RAGs are bound to a 12-RSS first, the NBD dimer tilts towards the 12-RSS and only enables the subsequent binding of a second more extended nonamer of the 23-RSS. In contrast, if the RAGs are bound to a 23-RSS first, the NBD tilts away from the nonamer and only enables a 12-RSS to subsequently bind (Ru et al., 2015).

1.11 “Beyond the 12/23” restriction

RSSs can impose significant constraints on variable region gene assembly beyond enforcing the 12/23 rule (Drejer-Teel et al., 2007). This restriction, known as the “beyond the 12/23 rule” (B12/23; Figure 1.15), has been identified at the *TCR β* locus (Bassing et al., 2000; Hughes et al., 2003; Sleckman et al., 2000), where $V\beta$ RSSs and 3' $D\beta$ RSSs contain 23 bp spacers and 5' $D\beta$ RSSs and $J\beta$ RSSs contain 12 bp spacers. Nevertheless, recombination between the $V\beta$ gene segments and the $J\beta$ gene segments almost never happens even though it is permissible by the 12/23 rule (Section 1.10). The occurrence of B12/23 restriction in the *TCR β* locus is not due to the competition between the 12-RSSs flanking the $D\beta$ and $J\beta$ gene segments or the arrangement of $V\beta$ and $J\beta$ RSSs (Sleckman et al., 2000; Bassing et al., 2000), as it is reproducible on extrachromosomal V(D)J recombination substrates in non-lymphoid cells (Jung et al., 2003; Olaru et al., 2003; Tillman et al., 2003). This also suggests that this restriction is not affected either by the nuclear environment of the *TCR β* locus or lymphoid-specific factors. In addition, the B12/23 restriction is not due to chromatin activation of the different regions of the *TCR β* locus at different stages, since it has been documented that both $V\beta$ and $DJ\beta$ regions are active at the same developmental stage (Jackson & Krangel, 2006; Mathieu et al., 2000).

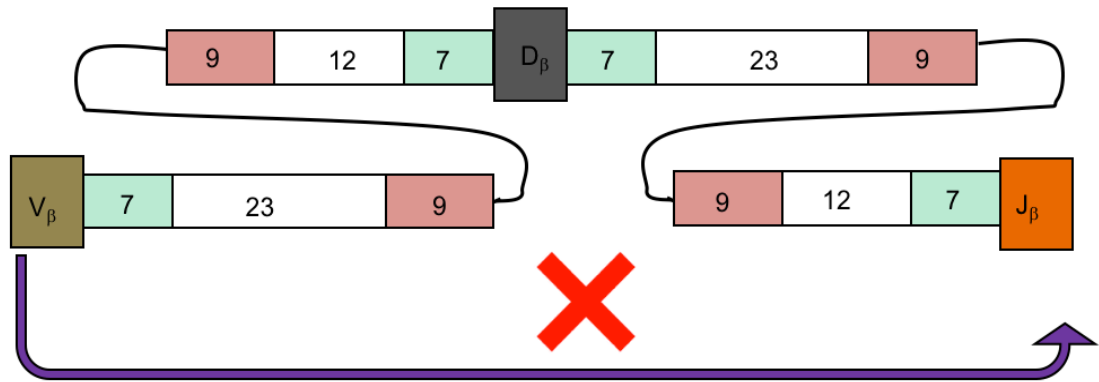


Figure 1.15 Beyond the 12/23 rule

In the mice *TCRβ* locus, V_{β} gene segment (shown as dark tan rectangle) is associated with a 3' 23-RSS, D_{β} gene segment (shown as light black rectangle) is flanked by a 5' 12-RSS and a 3' 23-RSS, and J_{β} gene segment (shown as orange rectangle) is associated with a 3' 12-RSS. However, recombination between the V_{β} gene segments and the J_{β} gene segments almost never happens even though it is permissible by the 12/23 rule. This is so called "beyond the 12/23 rule".

The contributions of the coding flank, heptamer, nonamer, and spacer sequences to the B12/23 rule have been studied *in vivo* and *in vitro* (Bassing et al., 2000; Hughes et al., 2003; Jung et al., 2003; Olaru et al., 2004; Sleckman et al., 2000; Wu et al., 2003; Wu et al., 2007). Several studies indicated that the B12/23 restriction is critically dependent on recombination signal nonamer and spacer sequences (Hughes et al., 2003; Jung et al., 2003), which could potentially explain why RAG binding to D_{β} 12-RSSs is much more efficient than to J_{β} 12-RSSs. However, EMSA studies showed binding to J_{β} 12-RSS is comparable to binding to D_{β} 12-RSS, albeit substantially weaker than to a consensus 12-RSS (Drejer-Teel et al., 2007). In addition, the ChIP experiments by Ji et al did not imply a RAG binding preference (Ji et al., 2010).

The B12/23 rule is largely or entirely enforced at the nicking and pairing steps in the DNA cleavage phase of V(D)J recombination by the RAG proteins, the *TCRβ* RSSs, and their flanking coding sequences (Tillman et al., 2004). Based on the observation that the D_{β} 23-RSSs contain a binding site for the AP1 transcription factor (Wang et al., 2008), a potential explanation for this phenomenon was proposed. c-Fos, which is a component of AP1, could promote the D_{β} -to- J_{β} rearrangement by binding the D_{β} 23-RSSs and greatly

increasing RAG association with these RSSs. In addition, it was shown that if the AP1 site is mutated, this effect can be eliminated (Wang et al., 2008). Furthermore, an *in vitro* study performed in 2007 found that synapsis and nicking of J β 12-RSSs is much more efficient with D β 23-RSSs compared to that with V β 23-RSSs in the absence of c-Fos (Drejer-Teel et al., 2007). Consequently, whilst the molecular basis of the B12/23 restriction is still not understood completely, the beyond 12/23 rule is most likely imposed by a combination of RAG interactions with c-Fos/AP1, and by as-yet unidentified intrinsic features of the TCR β RSSs.

C) Regulation of V(D)J recombination

1.12 Ordered recombination

V(D)J recombination must be strictly regulated, as it is a dangerous process, involving the breaking and joining of DNA in developing B and T cells. V(D)J recombination is tightly controlled by various mechanisms that occur at a number of different levels. Firstly, V(D)J recombination is regulated depending on cell lineage, for example, the expression of both RAG proteins only occurs in cells in the lymphoid lineage. Moreover, immunoglobulin loci and T cell receptor loci normally recombine in B cells and T cells, respectively. Secondly, antigen receptor locus rearrangement is regulated at the level of cell development, such that recombination of the immunoglobulin heavy chain stage (*IgH*) locus only occurs in pro-B cells, and the light chain loci (kappa or lambda) recombine mainly in pre-B cells once recombination at *IgH* locus has been successfully achieved. In addition, complete rearrangement usually only takes place on one allele, known as allelic exclusion (Section 1.18).

1.13 B and T cell development

1.13.1 B cell development

B cells pass through a number of development stages both before and after exposure to antigens. Commitment of a haematopoietic stem cell (HSC) to the B cell lineage primarily relies on the expression of Pax5, a transcription factor,

which has been referred to as “the guardian of B cell identity” (Nutt et al., 1999). B-cell development can be divided into two distinct phases: the first phase is the antigen-independent precursor-B cell differentiation, which happens in bone marrow, and the second phase is antigen-dependent B-cell maturation, which occurs in peripheral lymphoid organs, such as spleen, where most B cells continue their development into mature follicular B cells. The purpose of the first phase, i.e. precursor-B cell differentiation, is to create immunoglobulin (Ig) receptors, which occurs by classical ordered V(D)J recombination of the genes encoding the Ig heavy (*IgH*) and Ig light (*Igκ* or *Igλ*) chains (Ghia et al., 1996; van Zelm et al., 2005). Recombination firstly takes place at the *IgH* locus by rearrangement of D_H to J_H gene segments, followed by V_H to DJ_H joining. Following productive VDJ_H recombination, the resulting product, known as μ chain, is expressed, prior to formation of a complex with the surrogate light chains, VpreB and lambda 5 (Bassing et al., 2002; Melchers et al., 1993). After expression of this complex and formation of the pre-B-cell receptor (pre-BCR) on the cell surface, the pre-B cell undertakes several cycles of proliferation, progressing from the large pre-B cell stage to the small pre-B cell stage (Geier & Schlissel, 2006), prior to inducing Ig light chain (*Igκ* and *Igλ*) rearrangements (Gorman & Alt, 1998; van Zelm et al., 2005). After successfully generating a functional Ig receptor, precursor-B cells migrate to the peripheral lymphoid organs and become part of the naïve B-cell pool. Each precursor-B cell generates a single Ig receptor and therefore, the peripheral B-cell pool contains a diverse range of antigen-specific receptors. Naïve B lymphocytes are short-lived cells and there is a high turnover due to the continuous production of newly-generated cells. However, once stimulated by a microbial antigen, a naïve mature B lymphocyte that recognizes the antigen via its specific Ig receptor, undergoes clonal proliferation and differentiation, thereby generating a large number of long-lived and non-proliferating memory B cells and plasma cells that produce and secrete large amount of antigen-specific Ig molecules (LeBien & Tedder, 2008; Nutt et al., 2015). Subsequently, these antibodies bind to antigens, neutralizing or destroying the microbe in collaboration with other cells of the immune system such as macrophages and natural killer cells (Murphy, 2012).

After a response, memory B cells persist, to “remember” the particular antigen and initiate a stronger and faster antibody response upon re-exposure to the antigen (Kurosaki et al., 2015). Early in this response, highly specialized germinal centers (GCs) are created. In these GCs, B lymphocytes proliferate in network of follicular dendritic cells that present complete unprocessed antigen coupled to complement with their CD21 complement receptors. Furthermore, B lymphocytes can take up the antigen, process it and present it to helper T lymphocytes, which subsequently provide additional support, mainly by cytokine production and the expression of CD40L that interacts with CD40 on B lymphocytes. The CD40-CD40L interaction initiates additional molecular processes in the B lymphocyte: somatic hypermutation (SHM) of Ig variable regions to adapt and optimize the antigen-binding affinity (affinity maturation), and Ig class switch recombination (CSR) to change the effector function of the Ig molecule. Thus, whereas precursor B-cell differentiation in bone marrow results in a large repertoire of B-cell antigen receptors (BCR), in the periphery antigen-specific B lymphocytes are selected, modified, and specifically expanded to generate a highly specific and powerful immune response.

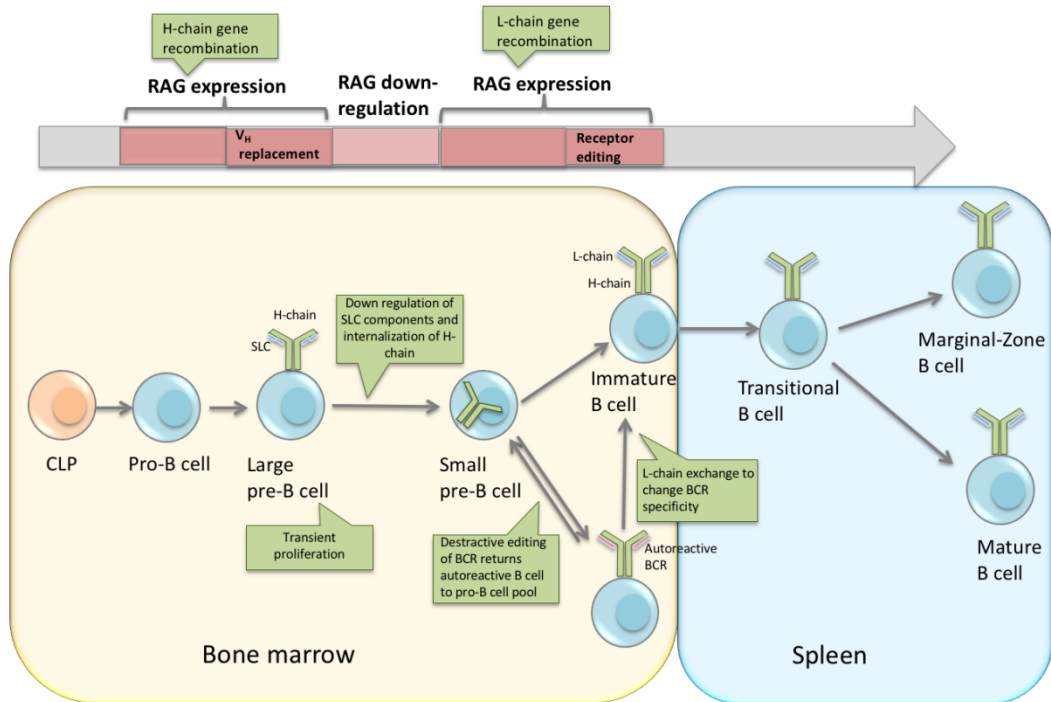


Figure 1.16 B cell development.

Phenotypically, the main stages of early B cell development, the pro-B, pre-B and immature B cells, are characterized by expression of specific cell surface markers and by successive rearrangements of the immunoglobulin (Ig) heavy (H) and light (L) gene segments. Successful rearrangement of the H-chain gene segments in pro-B cells allows the expression of the pre-B cell receptor (pre-BCR). Pre-BCR expression initiates progression into the pre-B cell stages and induces L-chain rearrangement. After this, the B lymphocyte precursors have reached the final stage of antigen-independent development. These immature B cells express surface IgM and are able to migrate from the bone marrow (BM) to the spleen. There, they complete maturation by developing into either follicular or marginal zone (MZ) B cells.

1.13.2 T cell development

T cell development is also divided into a series of maturation steps that can be identified based on the expression of different cell surface markers. The majority of cells in the thymus generate $\alpha\beta$ T cells, whilst approximately 5% give rise to the $\gamma\delta$ TCR. The thymocytes at the earliest developing stage are named double negative (DN) cells due to the lack of expression of the co-receptors CD4 and CD8. Subsequently, the DN cells express CD44, which is an adhesion molecule, and CD25, which is the interleukin receptor α chain. Cells that express CD25 but lack expression of CD44 undergo a process

named beta-selection. Here, cells which have already successfully rearranged their *TCR β* chain locus are selected. Then, a pre-TCR is produced by pairing the β chain with the surrogate chain, pre-T α (Janssen et al., 2003; Williams & Bevan, 2007). The pre-TCR then produces a complex with CD3 molecules, leading to the further β chain loci rearrangement, as well as the further differentiation by up-regulation and expression of CD4 and CD8 (Krangel, 2009). These cells are called double positive (DP) cells. Cells that do not undertake beta-selection die by apoptosis.

Double positive cells recombine their *TCR α* chain loci, to generate an $\alpha\beta$ -TCR. Subsequently, these cells undergo positive selection. Double positive cells interact with self-antigens in the context of major histocompatibility complex (MHC) class I or II molecules (Janssen et al., 2003; Shedlock & Shen, 2003; Sun et al., 2004). If the cells interact with antigen/MHC with an appropriate affinity, they will survive. However, if they engage antigen/MHC with a weaker affinity, they will die by apoptosis. Following this, the surviving cells migrate into the inner medulla region of the thymus to undergo negative selection. They are presented with self-antigens on antigen presenting cells (APCs), for instance, macrophages and dendritic cells. Those T cells that interact with antigen too strongly undergo apoptosis. During negative selection, most developing thymocytes die. After selection, down-regulation of either co-receptor results in either naïve CD4 or CD8 single positive cells. These cells exit in the thymus and circulate through the periphery.

1.14 Regulation of RAG expression

Expression of the RAG proteins is the first level at which V(D)J recombination is regulated. RAG proteins are only expressed at specific stages of B and T cell development, thereby restricting the V(D)J recombination of antigen receptor loci in a lineage- and time-specific manner (Grawunder et al., 1995; Wilson et al., 1994). There is evidence for two distinct waves of RAG expression corresponding to V(D)J recombination in the developing B and T cells (Grawunder et al., 1995; Wilson et al., 1994). RAG proteins are firstly expressed in pro-B and pro-T cell (CD4⁻/CD8⁻ double negative T cell) stages,

supporting *IgH* and *TCR β* locus recombination, respectively. After successful formation of a pre-BCR or pre-TCR, RAG protein expression levels undergo a dramatic decrease, followed by rapid proliferation to the large pre-B cell or pre-T cell (CD4⁺/CD8⁺ double positive T cell) stages (Grawunder et al., 1995; Wilson et al., 1994). Afterwards, RAG proteins are up-regulated, allowing recombination of light chain loci to proceed. Once recombination has successfully occurred, RAG expression is down-regulated by the expression of BCR and positive selection of the expressed TCR, in immature B cell stage and T cells, respectively, preventing further recombination (Brandle et al., 1992).

RAG proteins can also be re-expressed at later stages of B cell development in secondary lymphoid organs, such as spleen and lymph nodes. If autoreactive antigen receptors are expressed by B cells, they can be inactivated by either experiencing anergy or clonal deletion. RAG expression is then reactivated and terminated by further rearrangement of the light chain loci, changing the specificity of previously autoreactive B cell receptor, in a process named receptor editing, and generating a functional antigen receptor with a somewhat altered specificity (Gay et al., 1993; McGargill et al., 2000; Nemazee, 2006; Tiegs et al., 1993). RAG proteins were also found to be expressed in peripheral B cell germinal centres (Hikida et al., 1996; Han et al., 1996), as well as in some mature B cells (Qin et al., 1999), whilst their function at these B cell developmental stages is still unknown.

In addition to control by lymphoid-specific transcription factors, *Rag* expression is also controlled by ubiquitous transcription factors. The *Rag1* promoter has been demonstrated to interact with E2A, Ikaros and NF-Y (Kurioka et al., 1996; Fuller & Storb, 1997; Brown et al., 1997), whilst the *Rag2* promoter includes potential binding sites for PAX5, LEF-1, Myb and GATA3 (Lauring & Schelissel, 1999; Kishi et al., 2000; Fong et al., 2000; Wang et al., 2000; Kishi et al., 2002; Jin et al., 2002; Zarrin et al., 1997). Since the *Rag2* promoter is activated specially in lymphoid cells (Wang et al., 2000; Lauring & Schlissel, 1999), the lymphoid specific expression of the recombinase seems to be conferred by *Rag2* expression, whilst the regulation of this promoter is slightly

different in both B and T cells. Conversely, *Rag1* promoter can be activated in other cell types, such as neuronal cells (Chun et al., 1991).

Some distal elements were also identified that contribute to the regulation of *Rag1* and *Rag2* expression. One clear example is Erag, which is a highly conserved enhancer located 22 kb upstream of the *Rag2* promoter (Hsu et al., 2003). In a reporter construct containing this region, the expression of GFP driven by either the *Rag1* or *Rag2* promoter was increased, whilst targeted deletion of Erag could partially influence B cell development, leading to a deficiency in V(D)J recombination (Hsu et al., 2003). In addition, consistent with previous studies of the promoters, the presence of potential binding sites for Ikaros, E2A, Pax5, LEF-1 and GATA was confirmed by sequence analysis of Erag. Notably, in a non-lymphoid cell line, E2A was able to activate the *Rag1* or *Rag2* promoter in the presence of Erag (Hsu et al., 2003). It is of interest that T cell development is not affected by Erag, suggesting that expression of RAG1 and RAG2 in T cells is only controlled by their respective promoters (Hsu et al., 2003).

As well as the transcriptional regulation described above, RAG protein activity is also controlled during the cell cycle via post-translational modifications (Lin & Desiderio, 1994). The cyclin-dependent kinase, CDK2/Cyclin restricts RAG2 activity to the G₀/G₁ phases of the cell cycle by phosphorylating threonine-490 in the RAG2 C-terminus (Lin & Desiderio, 1994; Li et al., 1996; Lee & Desiderio, 1999), resulting in the degradation of RAG2 by the ubiquitin-proteasome pathway. The cell cycle regulation of RAG2 ensures that DSBs are not formed during rapid proliferation of lymphocytes. Consistent with this, aberrant V(D)J recombination occurs if the RAG2 is enforcedly expressed at other stages of the cell cycle (Lee & Desiderio, 1999). Because RAG2 expression is restricted to G₀/G₁ stages of the cell cycle and efficient V(D)J recombination requires both RAG1 and RAG2, V(D)J recombination is effectively restricted to the G₀/G₁ stages. In addition, this mechanism influences the DSB repair pathway used, since the NHEJ pathway is the dominant way to repair DNA during the G₁ phase. Therefore, this mechanism also inhibits DNA repair by alternative pathways, such as homologous recombination alternative NHEJ, thus helping to ensure genomic stability.

1.15 RSS accessibility and V(D)J recombination

V(D)J recombination of the antigen receptor loci occurs at a specific stage of B and T cell development and rearrangement of distinct gene segments always occurs in a certain order. Three immunoglobulin loci (*IgH*, *Igλ* and *Igκ*) are rearranged completely in B cells, whilst four TCR loci (α , β , γ and δ) are assembled only in T cells. Since recombination of these loci depends on the same proteins (RAG1 and RAG2) and conserved RSSs, this suggests that there are mechanisms to control the chromatin accessibility to restrict V(D)J recombination to the correct lineage specifically (Yancopoulos & Alt., 1985; Stanhope-Baker et al., 1996; Roth & Roth, 2000). Indeed, two early studies in the 1980s proposed a “accessibility hypothesis”, which could explain the developmental regulation of the antigen receptor loci. This hypothesis pointed out that when the antigen receptor gene segments are not rearranging, they are not accessible to the recombinase. During cell development, the chromatin structure changes specifically increase the accessibility of the correct antigen receptor loci, thus allowing the RAG proteins binding to RSS and initiate recombination only at the specific stage of cell development (Alt et al, 1984; Yancopoulos & Alt, 1985). This accessibility coincides with transcription from the unrearranged V genes (germline transcriptions) (Subrahmanyam & Sen, 2010; Schatz & Ji, 2011; Kwon et al., 1998). In accordance with this, it was found that when the RSSs are packaged into reconstituted nucleosomes, their accessibility to RAG binding was blocked, therefore inhibiting V(D)J recombination (Kwon et al., 1998; Golding et al., 1999; Kwon et al., 2000; McBlane & Boyes, 2000). In addition, nucleosomes are favourably positioned over RSSs (Baumann et al., 2003), indicating that prior to V(D)J recombination being initiated, the nucleosomes must be remodelled. Furthermore, numerous studies have shown that particular histone modifications and other alterations in chromatin are involved in increased V(D)J recombination.

1.16 Histone modification and V(D)J recombination

In addition to nucleosome remodelling, V(D)J recombination is also carefully controlled by ordered histone modifications, such as hypermethylation of lysine

4 of histone H3 (H3K4me_{2/3}), and histone H3/H4 hyperacetylation (Choi et al., 2013; Chowdhury & Sen, 2001; Xu & Feeney, 2009; Morshead et al., 2003; Goldmit et al., 2005; Chakraborty et al., 2009). Only the loci undergoing recombination have hypermethylation of lysine 4 of histone H3 and acetylation of histone H3/H4 (Hsieh & Lieber, 1992; Xu et al., 2008; Xu et al., 2009; Subrahmanyam & Sen, 2010).

1.16.1 Histone acetylation and V(D)J recombination

Numerous studies showed that increased H3/H4 acetylation is strongly correlated with the regulation of ordered V(D)J recombination (Espinoza & Feeney, 2005; Huang & Muegge, 2001; Maes et al., 2001; McBlane & Boyes, 2000; Roth & Roth, 2000; Ye et al., 2001). In 2000, a study by McMurry and Krangel first proposed that H3 hyperacetylation is closely associated with accessibility to RAG proteins for V(D)J recombination at the *TCR α* loci (McMurry & Krangel, 2000). Further evidence for the role of histone acetylation in the developmental regulation of V(D)J recombination at the *IgH* locus suggested that histone acetylation was regulated both locally and universally (Chowdhury & Sen, 2001). In addition, histone acetylation was proposed to contribute to increased RSS accessibility to RAG proteins by increasing the accessibility of nucleosome remodelling complexes to their nucleosome targets (Nightingale et al., 2007). Whilst a lot of studies showed that histone acetylation is strongly associated with the initiation of developmental V(D)J recombination, histone acetylation alone is insufficient for full locus activation (Sikes et al., 2002; Tripathi et al., 2002; Hesslein et al., 2003), which suggests that additional mechanisms are involved.

1.16.2 Histone methylation and V(D)J recombination

H3K4me₃ plays a key role in efficient V(D)J recombination. In 2003, Morshead et al first observed increased H3K4me₃ at the *IgH* and *TCR β* loci (Morshead et al., 2003). Subsequently, several studies examined the recombination of the *Ig κ* locus at different B cell development stages and found that increased levels of H3K4me₃ is associated with the initiation of V(D)J recombination

(Fitzsimmons et al., 2007; Goldmit et al., 2005; Perkins et al., 2004; Xu & Feeney, 2009). Whilst the RAG complex initiates V(D)J recombination by selectively recognizing the conserved RSS sequence in open chromatin, it is targeted to regions of increased H3K4me3 through the direct interaction via the PHD finger of RAG2 (Liu et al., 2007; Matthews et al., 2007), which could possibly explain the molecular basis for the relationship between increased H3K4me3 and V(D)J recombination (Matthews et al., 2007; Ramon-Maiques et al., 2007). *In vivo* RAG2 binding also strongly associates with H3K4me3 in recombination centers (Ji et al., 2010). Consistent with this, mutations in the RAG2 PHD finger (Y415, M443 and W453) showed a marked loss of RAG2 binding to H3K4me3 peptides, resulting a noticeable reduction in the antigen receptor genes rearrangement in mice (Liu et al., 2007; Matthews et al., 2007). In addition, several PHD finger mutations are documented in human severe combined immune deficiency (SCID) or Omenn syndrome (Sobacchi et al., 2006). Furthermore, the interaction between H3K4me3 and the RAG complex was found to enhance the catalytic activity of the RAG complex by stimulating the nicking and hairpin formation steps of the cleavage stage (Matthews et al., 2007; Ramon-Maiques et al., 2007; Grundy et al., 2010). Consequently, the H3K4me3 appears to positively regulate the kinetic reaction of the RAG proteins, thus affecting V(D)J recombination efficiency. Moreover, it was also proposed that H3K4me3 can also lead to misrecognition of cryptic RSS by the RAG complex, leading to chromosome translocations (Shimazaki & Lieber, 2014).

1.17 Non-coding transcription and V(D)J recombination

In addition to histone modifications, non-coding transcription, also known as sterile transcription, is also proposed to be tightly associated with the initiation of V(D)J recombination. The transcripts of the unrearranged antigen receptor loci (non-coding transcripts) was first identified in the mid 1980s (Blackwell et al., 1986; Lennon & Perry, 1985; Schlissel & Baltimore, 1989; Yancopoulos & Alt, 1985). Non-coding transcription generates transcripts that do not encode functional proteins. However, non-coding transcription was suggested to be an unmistakable sign that preparation is underway for V(D)J recombination of AgR loci (Stubington & Corcoran, 2013). Some investigators found that when

V(D)J recombination is activated, these transcripts were simultaneously up-regulated (Duber et al., 2003; Engel et al., 1999), implying the association between the non-coding transcription and the regulation of V(D)J recombination. Consistent with this, the Krangel laboratory showed that non-coding transcription is critical for the regulation of recombination of $TCR\alpha$ genes. They observed that when a transcriptional terminator element was inserted downstream of the T early α (TEA) promoter, transcription at the $TCR\alpha$ locus was suppressed (Abarrategui & Krangel, 2006; Abarrategui & Krangel, 2007). This suppression not only inhibited the rearrangement between $V\alpha$ and $J\alpha$ gene segments, but also greatly reduced the levels of H3 acetylation, H3K4me2, H3K4me3 and H3K36me3 at the downstream $J\alpha$ gene segments. This indicates that TEA promoter-initiated non-coding transcription increases the activation of downstream RSSs and regulates the addition of histone modifications, thereby controlling their chromatin structure.

Moreover, several studies have shown that H2A/H2B dimers must be transiently evicted from nucleosomes to enable the RNA polymerase II to transcribe efficiently (Belotserkovskaya et al., 2003; Orphanides et al., 1999). Notably, a more recent study showed that the initiation of V(D)J recombination does not occur, even in the presence of all of the histone modifications previously associated with recombination. Instead, it was found that RSS accessibility is strongly correlated with the level of non-coding transcription. Moreover, this accessibility was found to depend on the transient eviction of H2A-H2B dimers at RSSs (Bevington & Boyes, 2013). Consequently, this gave an explanation for how nucleosomes are remodelled for the activation of V(D)J recombination. In addition, the authors argue that this mechanism may be essential to restrict the number of RAG-mediated DSBs by making the RSSs available only transiently, thus helping to maintain genomic stability.

1.18 Allelic exclusion

Effective rearrangement only takes place on one allele of each antigen receptor locus, ensuring that only one receptor is expressed on the surface of each lymphocyte (Sonoda et al., 1997). The homologous allele either remains

in the germline state, has an incomplete D-J rearrangement (in the case of the *IgH* locus or *TCR β* locus), or carries a non-productive rearrangement, for example when two gene segments are joined out of frame. The phenomenon of mono-allelic expression is known as allelic exclusion, but the mechanisms by which this occurs are still poorly understood. The current genetic models to explain the establishment of allelic exclusion fall into four categories. These include the stochastic model, asynchronous recombination models and feedback inhibition models (Vettermann & Schlissel, 2010).

1.18.1 The stochastic model

In the stochastic model, allelic exclusion is suggested to occur due to the low probability of generation of a productive VDJ or VJ exon on two alleles simultaneously, since the probability of V(D)J recombination products being rearranged in the correct reading frame is very low, resulting in the high probability of generation of a non-productive (out-of frame) allele (Cohn & Langman, 1990; Coleclough et al., 1981; Krangel, 2009; Schlimgen et al., 2008; Schlissel, 2002). A 3D immune-fluorescence in-situ hybridisation (FISH) study in Krangel's lab provided evidence in favour of this model by showing the allelic exclusion on *TCR β* alleles (Schlimgen et al., 2008). In this study, the authors found that *TCR β* alleles in double negative (CD4⁻/CD8⁻) thymocytes associated with a high frequency either with pericentromeric heterochromatin or the nuclear lamina (Schlimgen et al., 2008). Notably, associated *TCR β* alleles were distributed stochastically in double negative thymocyte nuclei and were less likely to undergo V β -to-D β J β recombination in CD4⁻/CD8⁻ thymocytes. Therefore, it was suggested that stochastically distributed and frequent associations with inhibitory subnuclear compartments lowers the possibility that two *TCR β* alleles undergo V β -to-D β J β recombination at the same time (Schlimgen et al., 2008). Nevertheless, this model would still allow productive rearrangement of the *IgH* locus on both alleles in approximately 20% of newly developing lymphocytes (Wabl & Steinberg, 1992), which is much more than the observed about 1% of cells allelically included (Barreto & Cumano, 2000). This implies that additional mechanisms might exist.

1.18.2 The probabilistic model

Asynchronous recombination models are divided into two sub-types: the probabilistic model and the instructive model. Asynchronous recombination models describe how the recombination process avoids the simultaneous rearrangement of the two alleles in one cell. These models suggest allelic exclusion is controlled by the accessibility of antigen receptor alleles within chromatin structure, thus facilitating the RAG recombinase to recognize only one allele of any given time and precluding biallelic rearrangements.

In the probabilistic model, it was proposed that a low efficiency of recombination because of limited chromatin accessibility leads to asynchronous recombination of each allele (Liang et al., 2004; Schlimgen et al., 2008; Perry et al., 1980). A study by Liang et al at the *Ig κ* locus provided evidence in support of this model (Liang et al., 2004). Here, it was proposed that the probability for the locus to be activated is very low because of the competition between the alleles for limited levels of transcription factors. It was documented that only 5% kappa alleles are highly activated in developing pre-B cells. Therefore, activation of two alleles in the same cell is highly unlikely. In this study, a GFP reporter was inserted into the *J κ* region of the locus, to act as a indicator of the exact time of locus activation for germline transcription. Nonetheless, in contrast to this, Taylor et al showed that the level of GFP expression does not actually represent the level of germline transcription (Taylor et al., 2009), implying that the locus could be activated to a higher level than that previously reported (Liang et al., 2004), thus causing doubt of the validity of this model of allelic exclusion.

1.18.3 The instructive model

In the instructive model, it was suggested that the asynchronous replication of each allele established during early embryogenesis epigenetically marks one allele, making it to be the first allele available for recombination in developing B cells (Mostoslavsky et al., 2001; Vettermann & Schlissel, 2010). This model is supported by FISH studies by Mostoslavsky et al, where they observed that

both Ig alleles were asynchronously replicated during early embryogenesis and their asynchronous replication timing was maintained during development (Mostoslavsky et al., 2001). In addition, they also found that the early replicating allele was preferentially recombined and the late replicating allele only undergoes rearrangement later. In accordance with this, a study at the *Igκ* locus showed that, in pro-B cells, even though both *Igκ* loci are located in the centre of the nucleus, the late replicating allele is preferentially recruited to pericentric heterochromatin upon differentiation to pre-B cells (Fitzsimmons et al., 2007; Goldmit et al., 2005; Kosak et al., 2002; Roldan et al., 2005), thereby decreasing the accessibility of the late replicating allele to the RAG proteins.

1.18.4 Feedback inhibition models

Apart from the above models of allelic exclusion, feedback inhibition models have also been proposed for allelic exclusion. Feedback inhibition models suggest that the gene products or intermediates of antigen receptor gene rearrangement inhibit the recombination process, thus maintaining allelic exclusion (Vettermann & Schlissel, 2010).

In the feedback inhibition models, it was suggested that functional antigen receptor gene products are firstly assembled into either pre-BCR or pre-TCR, initiating signals to inhibit recombination of the other allele (Alt et al., 1980; Alt et al., 1982; Alt et al., 1984). Typically, this signalling pathway suppresses further rearrangements of antigen receptor loci either by eliminating the accessibility to the RAG proteins or by downregulation of RAG expression (Grawunder et al., 1995). For example, in B cells, pre-BCR signalling reduces the active levels of the transcription factor STAT5 by reducing the sensitivity to the cytokine IL-7, thereby subsequently decreasing histone acetylation at the *IgH* locus and consequently reducing the accessibility of gene segments for recombination (Bertolino et al., 2005; Chowdhury & Sen, 2003; Stanton & Brodeur, 2005). In T cells, pre-TCR signalling decreases histone acetylation at the TCR β locus by eliminating binding of the E2A transcription factor, eventually suppressing recombination between V_{β} gene segments and DJ_{β} gene segments (Agata et al., 2007). In addition, it was also found that pre-BCR and

pre-TCR signalling is relevant to the chromatin compaction at the antigen receptor heavy chain loci, thus further diminishing the recombination efficiency by arranging gene segments further away from each other (Roldan et al., 2005; Jackson & Krangel, 2005). Once recombination at the antigen receptor light chain locus is achieved successfully, the newly expressed BCR or TCR initiates the second stage of feedback inhibition signals, therefore inhibiting any further recombination events by down-regulating RAG expression (Turka et al., 1991; Maes et al., 2000).

D) Generation of additional antigen receptor diversity

In addition to generation of antibody diversity by mixing and matching different V, D and J gene segments, antibody diversity in mature B cells can be increased even further by somatic hypermutation (SHM; Section 1.20) and class switch recombination (CSR; Section 1.21), both of which are initiated by activation-induced cytidine deaminase (AID) (Dudley et al., 2005; Neuberger 2008).

1.19 Activation-induced cytidine deaminase (AID)

Activation-induced cytidine deaminase (AID) is selectively expressed in activated germinal centre B cells within secondary lymphoid organ and converts deoxycytidine (dC) to deoxyuracil (dU) on single-stranded DNA (ssDNA; Torres et al., 2008). Transcription of the switch region upstream of the C_H1 domain or of an immunoglobulin V domain opens the DNA helix to generate ssDNA, which can then be deaminated by AID, forming mismatched dU/dG DNA base pairs. Uracil DNA glycosylase (UNG), which is an abasic excision repair (BER) protein, removes the mismatched dU base to create a basic site. Differential repair of the lesion results in either SHM or CSR. In addition, the mismatch repair (MMR) proteins, such as MSH2 and MSH6, can also recognise and process the dU:dG mismatch.

1.20 Somatic hypermutation

In mammals, somatic hypermutation (SHM) of immunoglobulin genes is essential for the generation of high-affinity antibodies and an effective immune response (Marianes & Zimmerman, 2010). SHM is a process which is induced during B cell proliferation within the microenvironment of the GC. During this process, point mutations accumulate in the antibody variable regions of both the rearranged immunoglobulin heavy and light chains, with a rate that is approximately 10^6 -fold higher than that of the background mutation in other genes (Harris et al., 1999; Luo et al., 2004), increasing the specificity of the antigen-binding site. SHM does not happen in T cells, and therefore, V(D)J recombination itself is responsible for the entire TCR variation which can also explain the lower affinity of TCR-antigen interactions compared to antibody-antigen interactions.

SHM can be functionally divided into a deamination stage and a repair stage. Firstly, once a B cell is stimulated by an antigen, the expression of the AID is upregulated by the transcription factor Pax5 and E2A. Subsequently, AID initiates SHM by deaminating a dC residue in ssDNA to dU, leading to a mismatched dU:dG (Figure 1.17). Since AID can only act on ssDNA, it is essential that transcription occurs through the immunoglobulin loci to generate single-stranded transcription bubbles prior to SHM (Peled et al., 2008). It has been suggested that the transcription-mediated addition of H3K4me3 is tightly associated with the recruitment of AID to SHM and CSR hotspots in chromatin (Begum et al., 2012).

The mismatched dU:dG could be repaired in one of the following three ways: DNA replication, BER and MMR. DNA replication is the simplest way to repair this mismatched dU:dG. During DNA replication, when a uracil residue is encountered by the DNA polymerase, it is simply misread as a deoxythymidine, leading to a C to T mutation, or a G to A mutation on the opposite strand. During BER, the uracil residue is excised by UNG, generating an abasic site, which is then converted to a ssDNA break by apurinic/apyrimidinic endonuclease 1 (APE1). Following this, the gap is filled in with random base by an error prone DNA polymerase, such as DNA pol μ or

REV1 (Peled et al., 2008). During MMR, the mismatched dU:dG is recognised either by a MSH2-MSH3 or a Msh2-Msh6 heterodimer, which subsequently recruits a set of proteins, MLH1 and PMS2, and generates a single-strand nick at a site near the mismatched base (Peled et al., 2008). Following this, exonuclease I excises the mismatch, as well as several nucleotides surrounding the nick. Next, error prone polymerases (DNA Pol ϵ or δ) are recruited to repair the gap, and eventually DNA ligase I is recruited to ligate the ends of the mismatch (Peled et al., 2008).

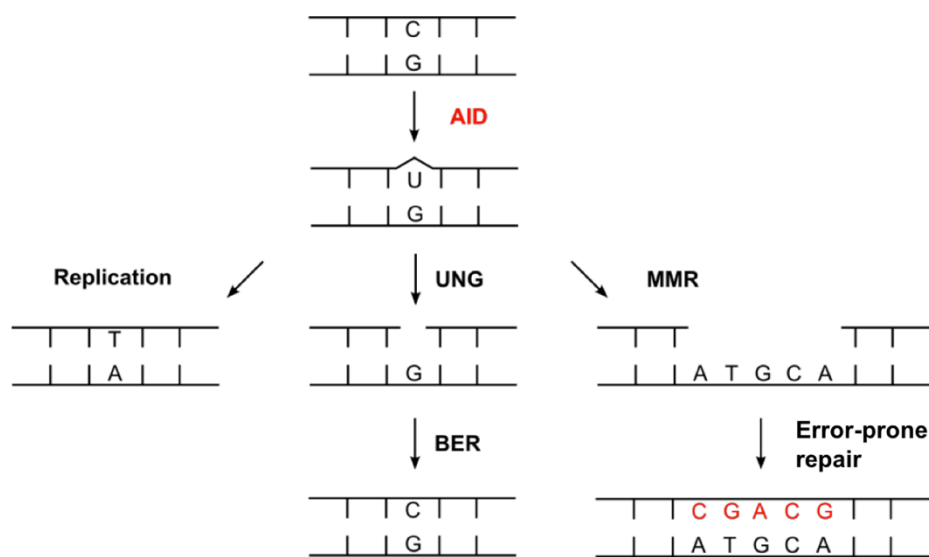


Figure 1.17 Model for somatic hypermutation

Somatic hypermutation is initiated by AID deaminating dCs to dUs, which is then processed by one of three ways: DNA replication, base excision repair (BER) or mismatch repair (MMR). During DNA replication, a U is misrecognized as a T, leading to a C to T point mutation. During BER, the uracil residue is removed by uracil DNA glycosylase (UNG), leaving an abasic site, which is then repaired by an error prone polymerase. During MMR, bases surrounding the lesion are also mutated whilst being repaired.

1.21 Class switch recombination

Class switch recombination (CSR) of the immunoglobulin heavy chain, also known as isotype switching, is mediated by an intrachromosomal deletional recombination event between μ , γ , α , or ϵ constant region genes (Figure 1.18). This recombination event is upregulated during immune responses by a regulatory region that lies 5' of the constant region genes. Similar to SHM, the initiation of CSR is also mediated by AID. However, unlike SHM which leads to

the generation of point mutations, CSR requires the generation of DSBs, which are then processed by the NHEJ pathway. Even though naïve B cells are capable of switching to any immunoglobulin isotype, the switch is directed by cytokines secreted by T cells to specific isotypes mainly by regulating the transcription through switch regions, as only transcriptionally active switch regions are subject to AID mutagenesis (Stavnezer et al., 2008).

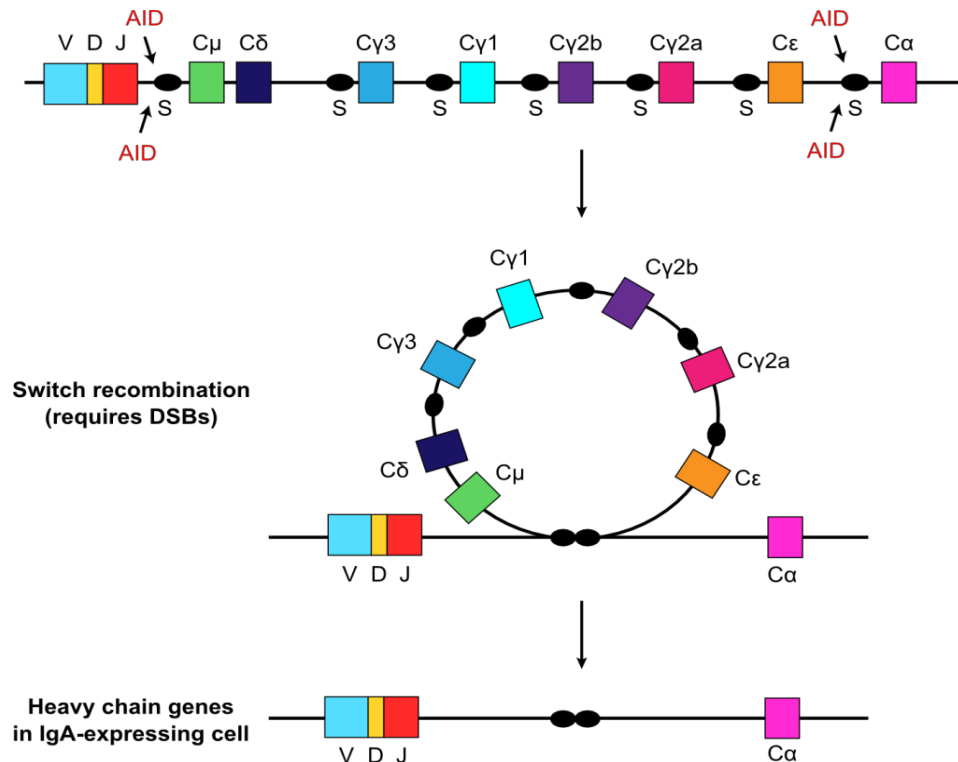


Figure 1.18 Class switch recombination (CSR)

CSR is initiated by transcription through switch (S) regions (ovals) and requires AID and components of several DNA repair pathways. During CSR, the intervening DNA sequence between participating S regions is excised as a switch circle and a new constant region gene is juxtaposed downstream of the variable region exons (Adapted from Stavnezer & Schrader, 2006).

CSR is initiated by the action of AID on both strands of a switch (S) region. Firstly, non-coding transcription through the C region of the *IgH* locus generates localised regions of ssDNA at switch regions, which are located upstream of each *IgH* constant region exon except C δ (Stavnezer et al., 2008). S regions are 2 kb long and consist of multiple repeats of the short G-

rich tandem DNA sequences (generally 20-80 bp for each), which forms the R loops to stabilize ssDNA (Muramatsu et al., 2000; Stavnezer et al., 2008). AID then deaminates cytosine residues and introduces numerous dU residues in the S region on each strand. Similar to BER, these mismatched dU:dG are recognized and removed by UNG, forming an apyrimidinic ribose residue, which is then edited by APE1 or APE2, thus leading to the formation of single-stranded nicks. The generation of multiple nicks in the S regions on both strands thereby makes it highly likely that two single-stranded nicks close enough to each other on opposite strands to subsequently form a DSB. Notably, prior to joining two DSBs at donor and acceptor switch regions, 3' and 5' overhangs generated must first be processed by ERCC1-XPF excision and filling in by DNA polymerases, respectively. DSBs at donor and acceptor switch regions are then joined by the NHEJ pathway via the same mechanism as during V(D)J recombination (Stavnezer et al., 2008).

E) Errors in V(D)J recombination

Despite being well-regulated, V(D)J recombination is prone to errors, generating aberrant DNA rearrangements during lymphocyte development. Indeed, after the discovery of V(D)J recombination, chromosomal rearrangements involving antigen-receptor loci were subsequently reported both in B- and T- cell neoplasms (Haluska et al., 1986; Kagan et al., 1987; Tsujimoto et al., 1985). More recently, next-generation sequencing technologies have allowed the genomic landscapes of these malignancies to be studied with greatly increased resolution (Roth, 2014). Aberrant rearrangement events found in lymphoid neoplasms contain chromosome translocations, relatively small deletions and inversions (Haydu et al., 2012; Mullighan et al., 2008; Zhang et al., 2012), as well as reinsertion of V(D)J recombination by-product, the ESC (Vanura et al., 2007). It has been reported that about 90% of lymphomas and 50% of leukaemias are associated with chromosomal translocations (Marculescu et al., 2006; Nambiar & Raghavan, 2011). In addition, several studies have emphasized the importance of deletions that influence numerous genes suspected or implicated in

tumorigenesis (Gigi et al., 2014; Mendes et al., 2014; Mijuskovic et al., 2012; Mijuskovic et al., 2015; Papaemmanuil et al., 2014).

Evidence that these chromosome changes are RAG-mediated includes: firstly, they take place between pairs of cryptic RSSs (cRSSs) that closely resembling authentic RSSs (Papaemmanuil et al., 2014); secondly, they often obey the 12/23 rule; and thirdly, reciprocal signal joints have been detected (Mijuskovic et al., 2015). Consequently, different types of aberrant V(D)J recombination events play key roles in initiating oncogenic transformation. Aberrant V(D)J recombination events have been grouped into four types: RAG-mediated recognition errors, end donation, RGA-mediated transposition, as well as ESC reintegration.

1.22 Recognition errors in V(D)J recombination

RAG proteins are known to mistakenly recognize suboptimal RSS-like sites at off-target genomic locations, causing DNA DSBs outside of the V(D)J loci, and thereby leading to chromosomal translocations and lymphoid malignancies (Tsai et al., 2008). The first type of errors is target recognition errors (Figure 1.19), whereby RAGs recognise an authentic RSS and a cryptic RSS, initiating cleavage and recombination (Roth, 2014; Showe & Croce, 1987; Zhang & Swanson, 2008). Because RSSs are quite small and consensus heptamer/nonamer sequences are not absolutely required, it is not surprising that cRSSs are recognized and used as an authentic RSS for V(D)J recombination (Lewis et al., 1997; Zhang & Swanson, 2008). In addition, Lewis et al estimated that there are over 10 million cryptic RSSs throughout the human genome with one cRSS roughly every 1-2 kb of genomic sequence, thereby greatly increasing the possibility of mis-recognition by RAG proteins and generating a huge scope for chromosomal translocations (Lewis et al., 1997).

Misrecognition of cRSSs by RAGs and using them as substrates for V(D)J recombination has long been suggested as a possible machinery for chromosomal translocations. By analysing the source of translocations in lymphoid cancers, it has been found that cRSSs strongly resemble authentic

recombination signal sequences at chromosomal breakpoints (Roth, 2003). In a chromosome translocation event involving an RSS/cRSS pair, an authentic RSS within an antigen-receptor locus is recombined with a cRSSs adjacent to proto-oncogenes (Figure 1.19a). This leads to a chromosome translocation where the oncogene is placed under the strong transcriptional control of the antigen receptor locus. An example of this cryptic recombination that leads to human lymphoid neoplasms is the t(8;14) chromosome translocations where the *c-myc* oncogene on chromosome 8 recombined with a J_H region on chromosome 14, leading to the B-cell malignancies Burkitt's lymphoma and acute lymphoblastic leukaemia (ALL) (Haluska et al., 1986; Manolov & Manolova, 1972; Robbiani & Nussenzweig, 2013). These events can result in inappropriate expression of the *c-myc* gene because of the presence of transcriptional regulatory elements from the heavy chain antigen receptor locus. Furthermore, recombination events including a cRSS/RSS pair can also deregulate oncogene through amplification, possibly through a breakage-fusion-bridge mechanism (Nam & Rabbitts, 2006; Raghaven et al., 2001; Zhu et al., 2002). For example, *LMO2* gene, usually located on chromosome 11, can be translocated to chromosome 14 and fuse with the TCR D or J gene segments (t(11;14)(p13;q11)) (Raghaven et al., 2001). In healthy cells, *LMO2* is essential for the early stages of haematopoiesis since it is a transcription factor. However, in these abnormal cells, the function of *LMO2* is deregulated by being placed under the transcriptional control of the active TCR promoter, thereby contributing to T cell acute lymphoblastic lymphoma (T-ALL) (Raghaven et al., 2001). Moreover, *Hox11*, *Ttg-1*, *TAL1*, and *TAL2* were also found to translocate to chromosome 14, leading to T-ALL cancers (Raghaven et al., 2001; Onozawa & Aplan, 2012).

Recent whole genome sequencing showed aberrant recombination also occurs between pairs of cRSSs either *in trans* (Figure 1.19b), causing a chromosome translocation as in T-ALL cases involving translocations between the *SCL* locus and T-cell receptor gene segments (Onozawa & Aplan, 2012), or *in cis* (Figure 1.19c), resulting in an excised signal joint and a deletional coding joint. It is of interest that, even though inversion events between cRSSs also are expected, they are seldom observed (Papaemmanuil et al., 2014). A

clear example for illegitimate V(D)J recombination involving two cRSSs on two non-antigen receptor loci is the *STIL-TAL1* rearrangement (Aplan et al., 1990; Brown et al., 1990), which present in about 16% of adult T-ALL patients and 26% of childhood T-ALL patients (Delabesse et al., 1997). The *STIL-TAL1* rearrangement involves a recurrent deletion of approximately 90 kb of DNA from chromosome 1. This site-specific interstitial deletion disrupts the *TAL1* locus and fuses intron 1 of the *TAL1* gene with the 5' regulatory region and exon 1 of the *STIL* gene (Aplan et al., 1990; Brown et al., 1990; Collazo-Garcia et al., 1995; Izraeli et al., 1997). This mutant allele results in a fusion mRNA transcript where exon 1 of *STIL* is spliced to exon 3 of *TAL1*. Because exon 1 of *STIL* and the first two exons of *TAL1* don't have the coding ability, the fusion mRNA does not encode a fusion protein, but instead encodes a full-length *TAL1* protein, this interstitial deletion ultimately brings *TAL1* under the control of *STIL* regulatory elements, resulting in mis-expression of *TAL1*. These deletional recombination events between cRSS pairs also have been identified in *Izgf1*, *Notch1*, *PTEN* and other important genes in lymphoid neoplasms in mice and humans (Ashworth et al., 2010; Haydu et al., 2012; Gigi et al., 2014; Mendes et al., 2014; Mijuskovic et al., 2012; Mullighan et al., 2008; Papaemmanuil et al., 2014). These are now considered to be major drivers of oncogenic transformation in lymphocytes (Roth, 2014).

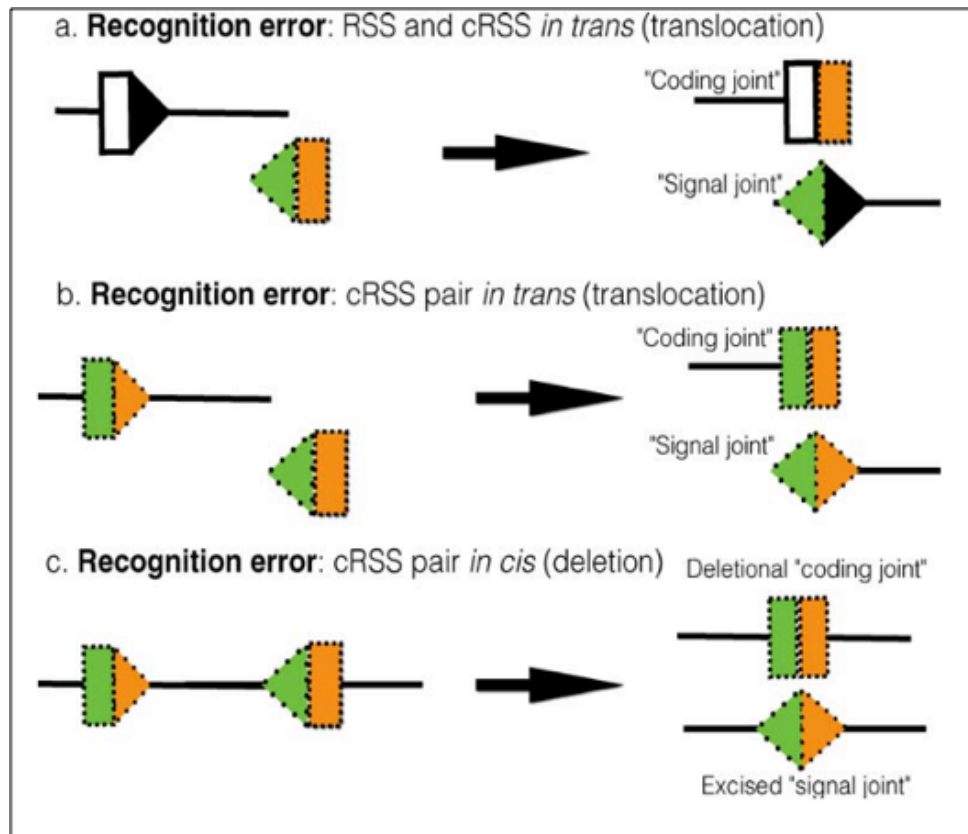


Figure 1.19 Recognition errors in V(D)J recombination

Three types of recognition error are shown. (a) Recombination occurs between an authentic RSS (blank triangle) with its associated coding flank (white box) and a cryptic RSS (cRSS; green triangle) with its associated coding flank (orange box), located on a separate DNA molecule. Recombination produces a *trans*-rearrangement, with a pseudo-coding joint and a pseudo-signal joint. (b) The recombinase recognizes a pair of cRSSs located on separated DNA molecules. These recombine, generating a reciprocal chromosome translocation. The two products bear a pseudo-coding joint and a pseudo-signal joint. (c) The recombinase recognizes a pair of cRSSs located on the same DNA molecule and generates a deletion, forming a pseudo-coding joint and an excised circle containing a pseudo-signal joint (Adapted from Roth, 2014).

Another type of recognition error involves RAG-mediated cutting at non-B-form DNA structures (Roth, 2014). A typical example of this type of error is t(14;18) chromosomal translocations which occurs in nearly all follicular lymphomas (Raghavan et al., 2001). In this case, the *BCL2* gene located at band q21 of chromosome 18 is translocated to the heavy chain locus on chromosome 14 and recombined with a J gene segment (Raghavan et al., 2001; Tsujimoto et al., 1984; Tsujimoto et al., 1985). In normal physiological conditions, *BCL2*

functions by reducing the release of mitochondrial cytochrome c, thereby preventing caspase 3 and 9 mediated apoptosis (Kridel et al., 2012). However, in the abnormal cells, *BCL2* expression is up-regulated by the J_H enhancer, thereby contributing to survival of these cells by decreasing their capacity to undergo apoptosis. Numerous studies have investigated the mechanism by which the prevalent t(14;18) translocation occurs (Raghavan et al., 2001; Raghavan et al., 2004; Raghavan & Lieber, 2006; Tsujimoto et al., 1984; Tsujimoto et al., 1985). It was found that translocations in the *BCL2* gene generally happen in the exon 3 within a 150 bp sequence termed the major breakpoint region (Mbr). This region can be further divided into three distinct sections named Mbr 1, 2 and 3, respectively (Raghavan & Lieber, 2006). Notably, Mbr 1 and 3 form a triplex DNA conformation due to the enrichment of purine/pyrimidine (Dayn, 1992). The DNA structure at Mbr 1 and 3, therefore, was proposed to be in a non-B form. It has also been demonstrated by numerous *ex vivo* and *in vitro* cleavage assays, that non-B structures are able to be cleaved by the RAG proteins (Raghavan et al., 2005). Furthermore, RAG proteins have been shown to cut other sections of DNA which adopt a non-B DNA form, for instance, heterologous loops (Nambiar & Raghaven, 2011). Together, these data imply that RAGs cut the non-B form DNA at Mbr, leading to translocation of the *BCL2* gene.

Nevertheless, in contrast, increasing evidence implies that chromosome translocations in developing lymphocytes are not entirely due to RAG recognition errors (Brandt & Roth, 2009). Firstly, RAG proteins do not always efficiently recognise and cleave all of the cRSSs, which has been demonstrated by *in vivo* recombination experiments using extra-chromosomal substrates (Marculescu et al., 2002; Raghavan et al., 2001). These assays showed that the *LMO2* cRSS undergoes recombination with an efficiency 27-fold lower than that of authentic RSS, whilst, no recombination events were detected with a *Hox11* cRSS (Marculescu et al., 2002; Raghavan et al., 2001). Secondly, a lot of translocation breakpoints do not take place at heptamer boundary of the candidate cryptic RSS, which is not in agreement with the translocation occurring via RAG cleavage activity. Lastly, some of the breakpoints include short direct repeats, suggesting that cleavage generates a

single-stranded overhang, which is also not in agreement with RAG cleavage activity (Bakhshi et al., 1987; Koppers & Dalla-Favera, 2001).

1.23 End donation

The second type error is named end donation (Figure 1.20; Lewis, 1994; Marculescu et al. 2002). End donation is one of the most common errors in V(D)J recombination (Roth, 2003; Koppers, 2005) although it is only partly linked with the V(D)J recombination machinery (Raghaven et al., 2005). In this reaction, the V(D)J recombinase creates a DSB at an authentic RSS, which is then joined with a random DSB formed by an unrelated process, such as ionizing radiation, reactive oxygen species (ROS), or physical or mechanical stress on the DNA duplex (Lieber, 2010), leading to chromosome translocations or insertions of signal-ended fragments into another chromosomal location (Marculescu et al., 2002; Vanura et al., 2007). It was documented that 30-40% of chromosome translocations found in follicular and mantle cell lymphomas are associated with end donation (Jager et al., 2000; Welzel et al., 2001).

The cleaved RSS, since it is a target for RAG proteins, appears to be generated via errors in V(D)J recombination, although the mechanism of cleavage is unknown. It seems improbable that the break is caused during normal V(D)J recombination as both the broken coding and signal ends are held together in a tight synapse by RAG proteins until the coding ends are re-joined by the NHEJ machinery (Kirkham, 2014; Lieber, 2010). Some evidence suggests that the coding ends are bound less tightly than signal ends *in vitro* (Hiom & Gellert, 1997; Lee et al., 2004; Wang et al., 2012), and this weaker association might provide the coding ends with an opportunity to join with other single DSB via the NHEJ pathway (Kirkham, 2014). As mentioned above, the *BCL2* translocation is an example showing that end donation is a possible cause for aberrant V(D)J recombination (Raghaven & Lieber, 2006). In this translocation, RAG proteins create a DSB at an RSS flanking a J gene segment at *IgH* locus on chromosome 14, which is then joined to a donated DSB cleaved by RAG proteins at the non-B form DNA structure found in the *BCL2* gene located on chromosome 18 (Raghaven & Lieber, 2006). In

addition, the *BCL1/IgH* fusion t(11;14)(q32;q21), where *the BCL1* gene on chromosome 11 was placed next to the *IgH* on chromosome 14 resulting in overexpression of the cyclin D1 gene located about 120 kb upstream the *BCL1* locus, provides further evidence for end donation being a cause of chromosome translocation (Marcelescu et al., 2002; Nyvold et al., 2007). In this reaction, only the breaks at the *IgH* locus are mediated by V(D)J recombination, whilst the breaks at the *BCL1* major translocation cluster (Mtc) are initiated by other yet-known mechanisms.

Although end donation clearly occurs, the NHEJ machinery acts to protect against this being a very frequent event. Specifically, experiments performed in NHEJ-deficient mice showed that these mice show a high association with lymphoid cancers (Jankovic et al., 2008). In particular, the absence of NHEJ proteins involved in the detection of DSBs, for instance, DNA-PKcs and the serine threonine kinase ATM, severely influences the development of lymphomas (Zhu et al., 2002). DNA-PKcs and ATM are able to transduce the DNA damage signal, activating other DSB repair proteins (Abraham, 2003). Loss of DNA-PKcs can inhibit recruitment of further NHEJ proteins and result in joining defects (Burma & Chen, 2004). Loss of ATM leads to an autosomal recessive disease termed ataxia telangiectasia, contributing to severe immunodeficiency and chromosome translocations (Barlow et al., 1996).

Even though the immune system has developed some ways to inhibit end donation, the fact that end donation is associated with as many as 40% of lymphoid cancers found in follicular and mantle cell lymphomas, nevertheless, suggests that end donation significantly contributes to the development of lymphoid cancers.

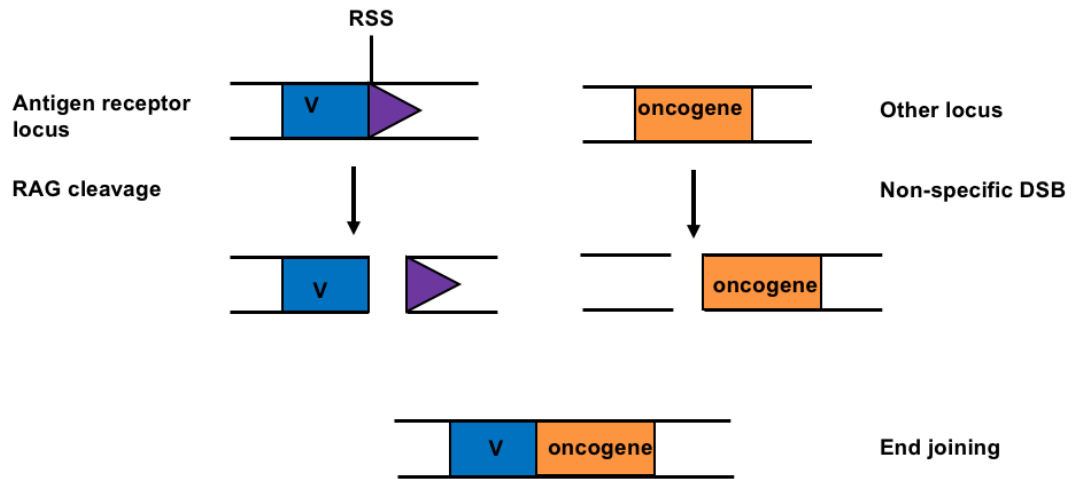


Figure 1.20 End donation

In this model, a DSB formed during the generation of antigen-receptor by RAG cleavage escapes the normal V(D)J recombination pathway and is joined to a non-specific DSB created by some other means on another locus, forming a chromosome aberration.

1.24 RAG mediated transposition

The third class of the aberrant recombination reaction involves the recombination by-product, the ESC. This can be re-inserted into the genome via a transposition-like reaction. The ESC is a relatively long-lived structure, despite the fact that it is eventually lost from cells. For example, the ESC persists in chicken (*Gallus gallus domesticus*) T cells for as long as 2 weeks (Kong & Chen-lo, 1999) and in primates, the persistence is thought to be substantially longer (Sodora et al., 2000). In addition, there is no degradation observed in precursor B cells, which implies that the ESC is able to persist into mature B cells in some cases (Kong & Chen-lo, 1999; Sodora et al., 2000).

Because the ESC doesn't have the coding capacity, it would be reasonable to ask why is it formed at all and how does it contribute to genome instability. ESC formation was previously thought to neutralize potentially reactive signal ends which could in theory undergo additional V(D)J recombination reactions. However, studies in the late 1990s showed that the ESC can also undergo a transposition reaction in a sequence independent manner, thus posing a threat to genome stability (Agrawal et al., 1998; Hiom et al., 1998). Hiom et al. suggested that RAG proteins can initiate transposition in two different ways. In

the first model, it was suggested that during RAG cleavage at a single RSS, the free 3'-OH group formed could undergo a transesterification attack on DNA located on another chromosome, thereby generating a branched molecule. Resolution of this structure would finally result in the joining of the coding gene segment with the DNA from the other chromosome through a hairpin intermediate (Figure 1.21A). The second model of transposition involves the ESC. In this reaction, ESC is cleaved by RAG proteins between heptamer/heptamer boundary, thus forming linearized signal ends, which then insert into the genomic DNA by taking part in a dual transposition on antiparallel strands of a DNA fragment in another chromosome (Hiom et al., 1998; Figure 1.21B).

Numerous studies have pointed out the similarity between RAG-mediated V(D)J recombination and transposition, consistent with the hypothesis that the RAG proteins evolved from a transposase which integrated into the genome of jawed vertebrates (Hansen & McBlane, 2000). Similar to RAGs, transposases specifically recognise short DNA sequences flanking the mobile gene segment and subsequently generate DSBs between the recognition sequences and the gene segment by a direct, one-step *trans*-esterification mechanism (van Gent et al., 1996). Therefore, cut-and-paste transposition reaction is very similar to V(D)J recombination.

Indeed, the RAG-mediated transposition reaction was found to be very frequent *in vitro* (Agrawal et al., 1998; Hiom et al., 1998). However, only two cases of transposition have ever been recorded *in vivo* (Messier et al., 2003; Messier et al., 2006). A study performed in human T cell line described an excised signal-end fragment which is originally from the T cell receptor locus integrating into an intron of the *HPRT* gene (Messier et al., 2003). However, notably, to date, there is no description of a case of lymphoma or leukaemia which is caused by RAG-mediated transposition. The low level of events *in vivo* suggested that transposition is unlikely to play a significant role in haematological malignancies (Ramsden et al., 2010). Therefore, the question arose of why do RAG-mediated transposition reactions take place so frequently *in vitro* but not *in vivo*? The initial *in vitro* studies were performed

using core RAG proteins, where the regions dispensable for the basic recombination reaction are removed for ease of purification, indicating that the non-core region of RAG proteins might play a role in inhibiting RAG-mediated transposition *in vivo*. Indeed, it was found that full-length RAG2 does not support efficient transposition, in contrast to core RAG2 (Swanson et al., 2004). The finding that the RAG2 C-terminus provides protection against transposition activity *in vivo* (Elkin et al., 2003; Messier et al., 2003; Messier et al., 2006), thus makes transposition an unlikely source of translocations in B cells.

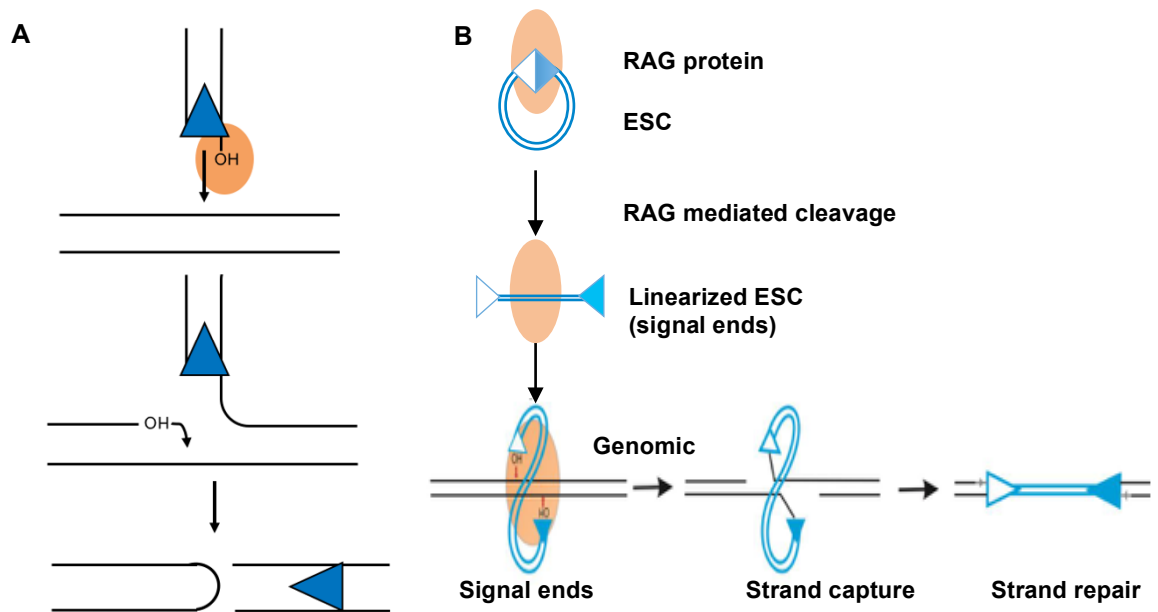


Figure 1.21 RAG mediated transposition reactions

(A) In transposition, a region of DNA is cleaved and incorporated into an unrelated site in the genome. The RAG proteins nick in between the coding gene and the RSS, generating a free 3' hydroxyl group. The reactive group can perform a transesterification reaction involving a nucleophilic attack on a phosphodiester bond in target DNA. **(B)** In this reaction, ESC is cleaved by RAG proteins to generate signal ends with free hydroxyl groups in the ends, which then attack and insert genomic DNA at the assistance of RAG proteins. The resulting strands are then repaired by DNA polymerase.

1.25 ESC reintegration

The last class of aberrant V(D)J recombination also involves the ESC. Similar to transposition, reintegration involves reinsertion of DNA fragment into the genome. But unlike transposition that shows no target specificity, the ESC

reintegrated into the genome at a genuine RSS or a cRSS i.e. ESCs can be rebound by the RAG proteins by virtue of the two RSSs, and reintegrates into the genome, forming a chromosomal signal joint and a pseudo-hybrid joint at RSS or cRSS (Arnal & Roth, 2007) (Figure 1.22).

The process of reintegration of the ESC has been described both *ex-vivo* and *in vivo*. (Vanura et al., 2007; Curry et al., 2007). Firstly, an *ex-vivo* experiment performed by Vanura et al showed that ESC can undergo *trans*-V(D)J recombination. Crucially, they pointed that the frequency of this *trans*-V(D)J recombination of ESC was the same as that of two canonical RSSs (Vanura et al., 2007). Simultaneously, this phenomenon was investigated *in vivo* using thymocyte DNA from mice. The presence of the same pseudo-hybrid joints suggests that reintegration of ESC also takes place *in vivo*.

A ligation-mediated transferred-end capture assay (LM-TECA) performed in murine thymocytes and splenocytes by Curry et al provided further evidence for ESC reintegration (Curry et al., 2007). This assay could recover the integrated RSS ends and subsequently detect if these RSS ends were indicative of the hallmark pseudo-hybrid joint. It was shown that half of the end insertions out of 48 observations showed loss of nucleotides to different extents, thereby indicating the formation of the of pseudo-hybrid joints which have processed signal ends. In addition, the fact that no target-site duplications were observed suggests that these events are not due to transpositions or translocations (Curry et al., 2007). It is of importance that, the frequency of reintegration events in core RAG2 mutant thymocytes was sevenfold greater, suggesting that RAG2 C-terminus serves to inhibit ESC reintegration, like ESC transposition (Curry et al., 2007).

In most cases, this reaction obeys the 12/23 rule (Vanura et al., 2007) and appears to proceed via a reversal of V(D)J recombination. Given that ESC reintegration relies on intermolecular *trans*-V(D)J recombination, the efficiency of this reaction appears to be significantly greater than transposition (Reddy et al., 2006). A study by Vanura *et al* estimated a rate of one reintegration event out of every 100,000 V(D)J recombinations. Considering the average number

of new lymphocytes generated every day, these data suggest that by as many as 5,000 ESC reintegration events could occur per genome per day (Vanura et al., 2007). Consequently, reintegration could be a significant pathway in the development of haematological malignancies. This is a particularly dangerous reaction if reintegration occurs adjacent to an oncogene (Figure 1.22).

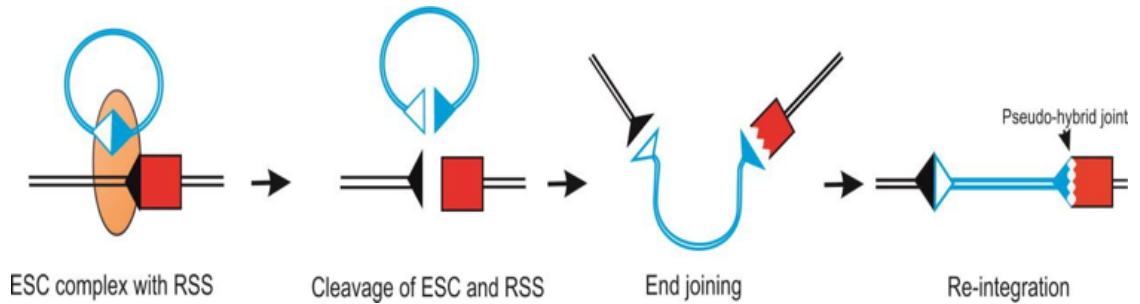


Figure 1.22 ESC reintegration

In this reaction, the ESC is bound by RAG proteins to an authentic or cryptic RSS to form a synaptic complex. Two cleavage sites are then formed. One is at ESC heptamer/heptamer boundary and the other is at RSS/coding end boundary. Cleaved ESC ends and the cleaved RSS are joined together by NHEJ pathway, generating a chromosomal signal joint. In addition, a pseudo-hybrid joint is formed, in which one end of ESC is joined with the coding end by imprecise end processing.

1.26 A newly uncovered aberrant recombination reaction: cut-and-run

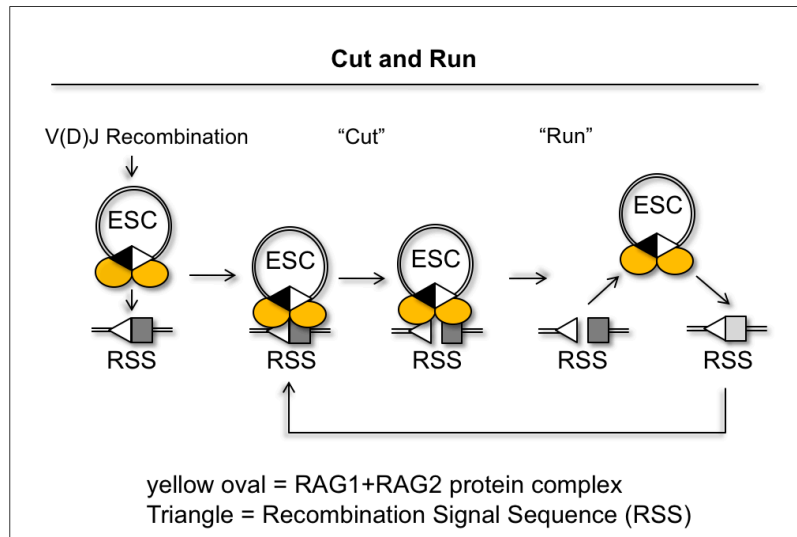
Recently, the Boyes lab discovered a new mechanism by which ESC could trigger genomic instability and they have named this reaction “cut-and-run” (Figure 1.23A; Kirkham, 2014). This is based on the observation in an *in vitro* RAG cleavage assay where when RAGs bind to the ESC and form a synaptic complex with an RSS, cleavage is asymmetric. Specially, in this *in vitro* RAG cleavage assay, Chris Kirkham found that, when a radioactively-coding end-labelled 50 bp or 61 bp oligonucleotide carrying the sequence of a 12- or 23-RSS was incubated with unlabelled 61 bp or 50 bp oligonucleotide carrying the sequence of a complementary partner 23- or 12-RSS or 79 bp oligonucleotide carrying the sequence of an ESC (two head-to-head RSSs), the ESC could greatly enhance cutting at an RSS, generating a 16 bp RSS cleavage product. However, in contrast, when a radioactively-labelled 79 bp oligonucleotide carrying the sequence of an ESC was paired with a unlabelled

oligonucleotide carrying the sequence of an RSS, a 34 bp ESC cleavage product was barely observed (Figure 1.23B). This was not due to limiting protein concentrations as increasing protein concentration by 8-fold did not increase ESC cutting. This implies that the ESC stimulates cutting at RSSs far more frequently than the cutting at the ESC. This further suggests that *in vivo*, the ESC could trigger cutting at many RSSs, leading to the increased risk of broken chromosomes/ translocations/ errors in repair. This new reaction is named “cut-and-run” since the ESC potentially triggers cutting at one RSS and then “runs” to trigger cutting at a second RSS. This process would enable multiple cycles of ESC mediating “cutting” at genomic RSSs or cRSSs followed by “running” to another site in the genome until the ESC becomes cleaved, or until the RAG proteins are downregulated, potentially generating numerous DSBs throughout the genome at RSSs and cRSSs. The DSBs generated by cut-and-run could be a source for the major chromosome translocations associated with aberrant V(D)J recombination reactions, and in particular, end donation. Consistent with this idea, Kirkham et al also showed that following cutting, the coding end is released at a 12- or 23-RSS but RAGs remain bound to the ESC.

The normal biology of ESC is thought to be non-functional by-product of V(D)J recombination and will be eventually lost with cell division. However, in the cut-and-run reaction, the ESC is a very important component which could continually stimulate cleavage at RSSs or cRSSs in the genome, generating numerous DSBs and potentially leading to genome instability.

Cut-and-run reaction was discovered to occur *in vitro* and the feature of this reaction is asymmetric cleavage between RSS and ESC. However, the molecular basis of this asymmetric cleavage was not completely understood from Chris Kirkham’s work. Therefore, in the first part of this thesis, I further investigate the molecular basis of cut-and-run.

A



B

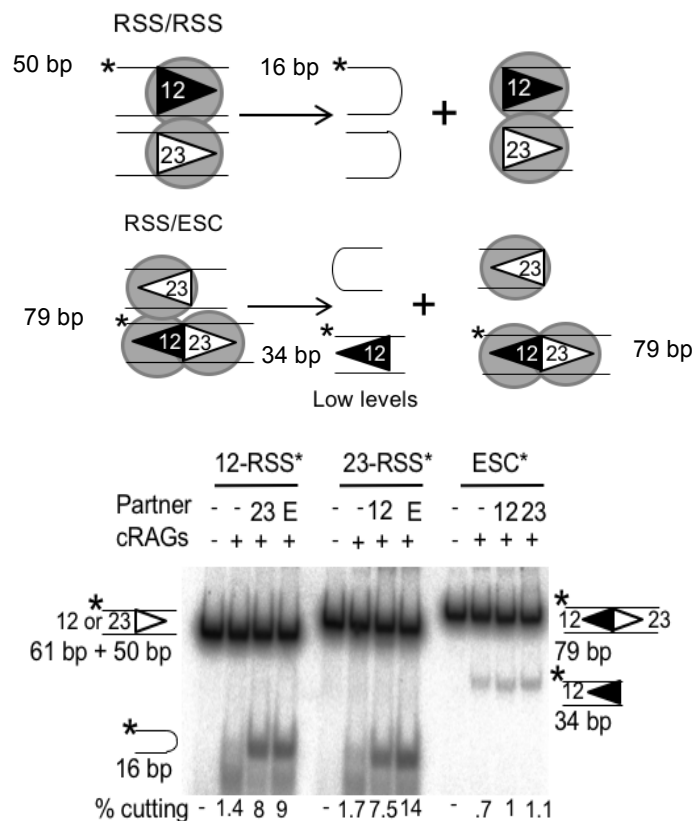


Figure 1.23 Cut-and-run

(A) The model of cut-and-run. After binding to a ESC and forming a synaptic complex with an RSS, the RAG proteins cleave the RSS but not the ESC. Following cleavage, the RSS is released from the post-cleavage complex, whilst the RAG/ESC complex remains intact, allowing the ESC to potentially run to trigger cutting at another RSS or cRSS in the genome.

(B) *In vitro* RAG cleavage at RSSs and ESC. When the RAG proteins, together with HMGB1, are incubated with a radioactively-labelled RSS and a unlabelled RSS or partner, the labelled RSS is cut and the ESC partner enhances this cutting compared to RSS partner. In contrast, when the ESC is radioactively-labelled, the cutting is significantly lower than that of RSS. This figure was prepared by Joan Boyes.

F) Aims

The overall aim of this project is to determine if cut-and-run is indeed a significant cause of genome instability that could underpin a major class of chromosome translocations, leading to lymphomas and leukaemias. Notably, in leukaemias triggered by the *ETV6/RUNX1* translocation, B cells are blocked at the pre-B stage of development and continue to undergo many rounds of V(D)J recombination. This gradually increases genome instability and causes progression from a pre-leukeamic state to full blown leukaemia. Whole genome sequencing of ALL patients showed that many chromosome aberrations map to cRSS, implying that aberrant recombination is a major cause of the increased genome instability. Although quite a lot of data has been generated in favour of the cut-and-run model, there are still some outstanding questions to determine if cut-and-run is truly a cause of genome instability. Therefore, the key questions I want to ask are:

- (1) What is the molecular basis of cut-and-run reaction *in vitro*?
- (2) Does cut-and-run occur *in vivo* and is this via the same mechanism as *in vitro*?
- (3) Does the ESC trigger cutting at cRSS *in vitro*?
- (4) How dangerous is cut-and-run reaction?

These aims are addressed in the following results chapters. In the first results chapter, the molecular basis of asymmetric cleavage was investigated by using mutated ESCs in *in vitro* RAG cutting assays, via bandshift assays, DNase I footprint assays, as well as supershift assays. In the second results chapter, the molecular basis of cut-and-run reaction was further investigated by *in vivo* RAG cutting assays with core RAG proteins or full-length RAG proteins in NIH3T3 cells or REH cells where the RAGs are present at physiological condition. Finally, in the third results chapter, the dangers of cut-and-run reaction was investigated by examining the presence and potential risks posed by the ESC.

Chapter 2:

Materials and Methods

A) Common buffers and Media

2.1 Common buffers

Alkaline lysis buffer I

50 mM	Glucose
25 mM	Tris-HCl (pH 8.0)
10 mM	EDTA (pH 8.0)

Alkaline lysis buffer II

0.2 M	NaOH
1% (w/v)	SDS

Alkaline lysis buffer III

3 M	KOAc
5 M	Acetic acid

DNA loading buffer (6 x)

15%	Ficoll
10 mM	Tris
1 mM	EDTA
0.01% (w/v)	Bromophenol blue
0.01% (w/v)	Xylene cyanol

Formamide loading buffer

0.2 ml	0.5 M EDTA, pH 8.0
10 mg	bromphenol blue [0.1% (w/v)]
10 mg	xylene cyanol [0.1% (w/v)]
10 ml	formamide

Lysogeny Broth (LB)

1% w/v	Bacto-tryptone
0.5% w/v	Yeast Extract
0.5% w/v	NaCl

LB-Agar (used for plating transformed *E. coli*)

1.5% w/v	Agar
1% w/v	Bacto-tryptone
0.5% w/v	Yeast Extract
0.5% w/v	NaCl

Ampicillin or Kanamycin was added to a final concentration of 50 µg/ml to make this media selective.

Phosphate-buffered saline (PBS)

137 mM	NaCl
2.7 mM	KCl
4.3 mM	Na ₂ HPO ₄
1.47 mM	KH ₂ PO ₄

Adjusted to pH 7.4

Protein loading buffer (4X)

20%	Glycerol
200 mM	Tris-HCl (pH 6.8)
568 mM	β-Mercaptoethanol
2%	SDS
0.2% w/v	Bromophenol Blue

SDS PAGE running buffer

192 mM	Glycine
25 mM	Tris-HCl (pH 8.5)
0.1% w/v	SDS

Sodium phosphate buffer (NaPi) pH 7.231.6 ml 1 M NaH₂PO₄68.4 ml 1 M Na₂HPO₄Make final adjustments to pH with NaH₂PO₄ (acid) and Na₂HPO₄ (base)**Super Optimal Broth (SOB)**

2% w/v Bacto-tryptone

0.5% w/v Yeast Extract

0.05% w/v NaCl

0.019% w/v KCl

TAE (Tris-acetate-EDTA) buffer

40 mM Tris-acetate

1 mM EDTA

TBE (Tris-borate-EDTA) buffer

90 mM Tris-base

90 mM Boric acid

2 mM EDTA (pH 8.0)

TBS (Tris-buffered saline)

50 mM Tris-HCl (pH 7.4)

150 mM NaCl

2 mM KCl

TE (Tris-EDTA) buffer

10 mM Tris (pH 8.0)

1 mM EDTA

TGS

25 mM Tris (pH 8.0)

192 mM Glycine

0.1% SDS

2. 2 Tissue culture media and solutions

Dulbecco's Modified Eagle Medium (DMEM)-complete

500 ml	DMEM medium (Sigma-Aldrich, D5671)
10% (v/v)	Fetal Calf Serum (Perbio Bioscience; heat treated at 56 °C for 40 minutes)
120 mM	L-Glutamine
2.5 ml	Penicillin/streptomycin (pen/strep; 100 U/ml final concentration for streptomycin and 100 µg/ml concentration for penicillin)

RPMI-1640 Medium

500 ml	RPMI-1640 medium (Sigma-Aldrich, R0883)
10% (v/v)	Fetal Calf Serum (Perbio Bioscience; heat treated at 56 °C for 40 minutes)
120 mM	L-Glutamine
2.5 ml	Penicillin/streptomycin (pen/strep; 100 U/ml final concentration for streptomycin and 100 µg/ml concentration for penicillin)
50 µM	β-Mercaptoethanol

Polyethyleneimine (PEI, 1 mg/ml)

Linear PEI (25 kDa, Alfa Aesar #043896.01) was dissolved in water heated to ~80°C then cooled to room temperature. The pH was adjusted to 7.0 with HCl, and the solution was filter-sterilised and stored in aliquots at -80°C. Working aliquots were kept at 4°C for up to 1 month.

B) Manipulation of DNA

2.3 Restriction enzyme digestion of DNA

All of the restriction enzymes used were obtained from New England Biolabs (NEB). Digestion reactions were set up in the recommended buffer and were incubated for 3-16 hours at the recommended temperature. Typically, the digestion reaction volumes were 10 µl, containing 2 units of restriction enzymes per µg of DNA and 1 µl of 10 x buffer, for diagnostic digestions. For

preparative digestions, the reactions had up to 10 µg DNA in a final volume of 200 µl.

2.4 Traditional PCR

Typically, PCR reactions were set up on ice and contained 1 x ThermoPol buffer (20 mM Tris-HCl [pH 8.8], 10 mM (NH₄)₂SO₄, 10 mM KCl, 2 mM MgSO₄, 0.1% Triton® X-100), 0.2 µM forward and reverse primers, 200 µM each dNTP, 1 ng plasmid DNA template and 2.5 units of Taq DNA polymerase (NEB #M0237) in a 50 µl reaction. The thermocycling conditions were as follow: pre-denaturation at 95°C for 2 minutes, followed by 30 cycles of denaturation at 95°C for 5 seconds, annealing at T_m for 30 seconds and elongation at 72 °C for 1 min/kb, with a final elongation at 72 °C for 10 minutes. The annealing temperature (T_a) of oligonucleotides was calculated using the OligoCalc oligonucleotide properties calculator (Kibbe, 2007).

2.5 Agarose gel electrophoresis

Different sizes of DNA fragments were separated by agarose gel electrophoresis. The concentration of agarose varied according to the sizes of fragments being separated. Typically, for small fragments (<500 bp) and large fragments (>1000 bp), 2% agarose gels and 1% agarose gels were used for electrophoresis, respectively. 1% agarose gel was made by dissolving 0.4 g of agarose in 40 ml 1x TAE buffer. To this, 5 µl of ethidium bromide (10 mg/ml) was added to allow the DNA to be visualised when exposed to UV light. Before loading on the gel, 1/6 volume of 6 x DNA loading buffer was added to the DNA samples. The agarose gel was electrophoresed in 1x TAE buffer in a Bio-Rad Sub-Cell GT tank at 80 V for 45-75 minutes to allow sufficient separation of the DNA fragments.

2.6 DNA ligation

DNA ligation was carried out using T4 DNA ligase (NEB). Typically, the reaction volume for ligation was 10 µl, containing 1 µl of 10 x T4 DNA ligase buffer (50 mM Tris-HCl [pH 7.5], 10 mM MgCl₂, 1 mM ATP, 10 mM DTT), 1

unit of T4 DNA Ligase (NEB #M0202S), 50 ng of digested and dephosphorylated vector DNA and 3-fold molar excess of digested insert DNA. Both overhanging “sticky” end ligations and blunt end ligations were set up on ice and incubated at 16°C for 16 h before transformation into competent *E. coli*. Vector-only control ligations were carried out alongside each vector plus insert ligation reaction.

2.7 Preparation of chemically competent *E. coli*

CaCl₂ treated competent DH5α *E. coli* cells were prepared using a variation of the method described by Cohen in 1972 (Sambrook et al., 1989). Firstly, DH5α *E. coli* was grown on a LB Agar plate without antibiotics at 37°C overnight. Following this, a single colony was inoculated into 50 ml LB medium in a 250-ml flask. Following incubation at 37°C overnight, 4 ml was transferred to a 2-litre flask containing 400 ml LB. This was then grown at 37°C to an OD⁶⁰⁰ of 0.375 (typically 2 to 3 hours post inoculation). Following this, eight 50 ml aliquots of cells were made, cooled on ice for 5 to 10 minutes, and were harvested by centrifugation at 1600 x *g* for 7 minutes at 4°C. Each cell pellet was gently resuspended in 10 ml of ice-cold CaCl₂ solution (60 mM CaCl₂, 15% glycerol, 10 mM PIPES [pH 7]) before centrifugation at 1100 x *g* for 5 minutes at 4°C. Following resuspension in 10 ml ice-cold CaCl₂ solution, cells were incubated for 30 minutes on ice, before centrifugation at 1100 x *g* for 5 minutes. Each cell pellet was finally resuspended in 2 ml ice-cold CaCl₂ and 200 µl aliquots of cells were made and frozen immediately at -80°C, until required.

2.8 Transformation of *E. coli*

Usually, the transformation was carried out with chemically competent DH5α cells, prepared as above. Firstly, DH5α cells were fully thawed on ice and a 50 µl aliquot was placed in a 1.5 ml centrifuge tube. Secondly, 2-10 ng of DNA or 5 µl of ligation product was added to the DH5α cells and the sample was mixed gently and incubated on ice for 30 min. The cells were then transformed by heat shock in a 42°C water bath for 90 seconds, followed by a 3 minutes recovery on ice. Subsequently, 300 µl of pre-warmed SOB was added and the

cells were incubated at 37°C for 45-60 min (generally 60 min for ligations, 45 min for stock plasmids), after which an appropriate amount of cells were plated on LB-agar plates containing the selective antibiotic for the plasmid transformed. Typically, 10 µl and 150 µl were plated for transformation of purified plasmid stock and ligation reactions, respectively. The plates were incubated at 37°C overnight.

2.9 Phenol/chloroform extraction

Phenol/chloroform extraction was used to remove contaminating proteins and SDS from DNA samples. An equal volume of 1:1 (v/v) phenol/chloroform mixture was added to the DNA sample. The mixture was then vortexed for about 1 min, thus producing an emulsion of the two phases. The sample was then centrifuged for 2 minutes at 15,700 x *g* at room temperature, which separates the two phases. The upper phase (containing DNA) was transferred to a fresh Eppendorf tube, and the tube containing the lower phases (proteins and phenol/chloroform) was discarded. This was repeated to fully remove contaminants. The DNA phase was then ethanol precipitated (Section 2.10).

2.10 Ethanol precipitation

DNA was precipitated by adding one tenth volume of 3 M NaOAc [pH 5.2] and two volumes of 100% ethanol. 1 µl of 20 mg/ml glycogen was added to make the DNA pellet visible when the amount of DNA is < 1 µg. The samples were mixed well by inversion and put on dry ice for five minutes, followed by centrifugation at 20,817 x *g* for 10 min at 4°C. Immediately, the supernatant was carefully removed. After addition of 1 ml of 70% ethanol to wash the pellet, the sample was centrifuged at 20,817 x *g* for 5 min at 4°C. Once precipitated, the supernatant was discarded and the DNA was allowed to completely air dry and then was resuspended in appropriate volume of TE (pH 8.0).

2.11 Gel purification of DNA

DNA was separated by agarose gel electrophoresis (Section 2.5) and visualized by ultraviolet light for the minimum amount of time to minimize nicking of DNA. Desired fragments were excised from the gel using a sterile scalpel blade and transferred to pre-washed dialysis tubing (with a molecular weight cut-off of 80 kDa) containing a minimal amount of 1x TAE. Fragments were electroeluted at 100 V for 1 h. Polarity of the current was reversed for ~20 s to recover DNA that had stuck to the interior surface of the dialysis tubing. DNA-containing TAE buffer was transferred to 1.5 ml tubes, and the DNA was purified by 4 cycles of phenol/chloroform extraction (Section 2.9) and 1 cycle of ethanol precipitation (Section 2.10) and resuspended in appropriate amount of TE.

2.12 Small-scale preparation of plasmid DNA (Mini-prep)

Mini-prep of plasmid DNA were carried out as follows: A single bacterial colony was picked from an LB-agar plate using a sterile pipette tip and transferred into 2 ml of LB medium containing the appropriate selective antibiotic. Cultures were then incubated overnight at 37°C with shaking at 200 rpm/min. The next day, 1.5 ml of the culture was transferred to a microfuge tube and centrifuged at 12,000 x g for 30 s. The supernatant was then discarded and the pellets were thoroughly resuspended in 100 µl of ice-cold solution I (50 mM glucose, 25 mM Tris [pH 8.0], 10 mM EDTA [pH 8.0]) by vigorous vortexing. Then, 200 µl of freshly prepared solution II (0.2 M NaOH, 1% SDS) was added, mixed gently by inverting 5-6 times, and stored on ice for 5 minutes. 150 µl of solution III (3 M potassium acetate, 5 M glacial acetic acid) was then added and mixed by vortexing gently in an inverted position for 10 s. The tubes were then placed on ice for 5 minutes, followed by centrifugation at 12,000 x g for 5 min at 4°C. The supernatant was then transferred to a fresh 1.5 ml centrifuge tube and an equal volume of phenol/chloroform was added and mixed by vortexing for 1 min. Before transferring the supernatant into a fresh tube, the mixture was centrifuged at 12,000 g for 2 min at 4°C. Two volumes of ethanol at room temperature was then added and mixed well to precipitate the double-strand DNA. The mixture was allowed to stand for 5 min at room temperature,

followed by centrifuging at 12,000 x *g* for 10 min at 4°C. The supernatant was then removed carefully and thoroughly and the pellet was rinsed by adding 1 ml of 70% ethanol at 4°C. The mixture was then centrifuged at 12,000 x *g* for 5 min. The supernatant was then removed carefully and thoroughly and the pellet was air-dried for 10 min. The DNA was resuspended in 50 µl of TE containing 2 µl of DNase-free pancreatic RNase A (10 mg/ml).

2.13 Large-scale preparation of plasmid DNA using Qiagen Kits (Maxi-prep)

The DNA was extracted according to the manufacturer's recommended instructions. Specifically, 500 ml of an overnight bacterial culture was harvested by centrifuging at 6,000 x *g* for 15 min at 4°C. The supernatant was discarded and the bacteria pellet was resuspended and mixed well by pipetting up and down with 10 ml of buffer P1, followed by addition of 10 ml of buffer P2. The sample was mixed thoroughly by inverting several times and was then incubated at room temperature for 5 min. 10 ml of pre-chilled buffer P3 was added to the sample, which was then vigorously inverted 5-6 times to mix thoroughly. Following 15 minutes' incubation on ice, the lysate was centrifuged at 15,000 x *g* for 30 min at 4°C and the supernatant was re-centrifuged at 15,000 x *g* for 15 min at 4°C to clear the last pieces of debris. After equilibration with 10 ml of buffer QBT by gravity flow, the supernatant was applied to the maxi-prep column. The column was then washed with 2 x 30 ml of buffer QC and the DNA was eluted with 15 ml of buffer QF into a clean 50 ml tube, followed by addition of 0.7 volumes of room-temperature isopropanol and centrifugation at 15,000 x *g* for 30 min at 4°C to precipitate the DNA. The supernatant was carefully decanted. The DNA pellet was washed with 5 ml of room-temperature 70% ethanol and centrifuged at 15,000 x *g* for 10 min at 4°C. The supernatant was carefully discarded. Before dissolving in suitable volume of TE, the DNA pellet was thoroughly air-dried for at least 10 min. The concentration of DNA was then measured using a NanoDrop spectrophotometer (A^{260}).

2.14 Large-scale preparation of plasmid DNA by the cesium chloride (CsCl) method (Maxi-prep)

Pure DNA was extracted from *E. coli* using the alkaline lysis method. 500 ml of bacteria culture containing selective antibiotic was incubated at 37°C overnight with shaking. After harvesting the cells by centrifuging at 5000 x *g* for 20 min, the supernatant was discarded and the pellet was resuspended in 40 ml of solution I (50 mM glucose, 25 mM Tris-HCl [pH8.0], 10 mM EDTA). 80 ml of freshly prepared solution II (0.2 M NaOH, 1% SDS) was then added and mixed by swirling gently before incubation on ice for 5 min. This was followed by addition of 40 ml of cold solution III (5 M potassium acetate, acetic acid) and mixing thoroughly by vigorously shaking. Before centrifuging at 5000 x *g* for 15 min at 4°C, the sample was kept on ice for 15 min. The supernatant was then filtered into a 250 ml of measuring cylinder and transferred into a 250 ml of centrifuge bottle. 0.6 volumes of isopropanol were added and mixed, followed by centrifuging at 8000 x *g* for 15 min. The supernatant was discarded and the pellet was washed in 70% ethanol. After centrifuging at 8000 x *g* for 5 min, the pellet was resuspended in 10 ml of TE [pH 8.0]. The suspension was then transferred to a 50 ml Sorvall centrifuge tube and 10 ml of 5 M lithium chloride (LiCl) was added to precipitate RNA. The sample was put on ice for 10 min, followed by centrifuging at 10,000 x *g* for 10 min. The supernatant was transferred into fresh tubes and two volumes of ethanol were added to precipitate the DNA. After centrifuging at 10,000 x *g* for 10 min, the pellet was washed in 70% ethanol and resuspended in 3.21 ml of TE. 3.55 g of solid CsCl was added and dissolved completely, followed by addition of 0.34 ml of 10 mg/ml ethidium bromide (EtBr). After transferring into 5 ml tubes and adding CsCl/TE (1mg/ml) to fill the tubes, the sample was then centrifuged in a Vti65.2 fixed angle vertical rotor at 55,000 x *g* overnight at 16°C. The next day, the highly stained lower band containing the supercoiled plasmid DNA was removed into a fresh 50 ml of centrifuge tube by inserting a needle through the side of the centrifuge tube. The EtBr was extracted from the DNA at least three times with CsCl-saturated isopropanol. Three volumes of 70% ethanol were added, followed by centrifuging at 12,000 x *g* for 10 min. The pellet was washed with 2 x 10 ml of 70% ethanol. After centrifuging, the supernatant was

discarded and the pellet was air-dried thoroughly and resuspended in 500 μ l of TE. 16.6 μ l of 3 M NaOAc and 3 μ l of 10 mg/ml RNAase A were added. The mixture was incubated at 37°C for 30 min to remove residual RNA. 10 μ l of 20% SDS and 10 μ l of 10 mg/ml proteinase K (PK) were then added and incubated at 37°C for 30 min to remove protein, followed by four cycles of Phenol/chloroform extractions (Section 2.9) and 1 cycle of ethanol precipitation (Section 2.10). The DNA pellet was resuspended in 500 μ l of TE and the concentration was measured using a NanoDrop spectrophotometer (A^{260}).

2.15 Preparation of glycerol stocks for long-term storage of plasmid DNA

For long-term storage of plasmids, glycerol stocks were made by transferring 625 μ l of cells cultured for 16 hours for large scale plasmid purification, to a sterile 1.5 ml screw-cap tube, adding 375 μ l sterile 40% glycerol and stored at 4°C for 8 hours before placing at -80°C for long term storage.

C) Expression and purification of RAG proteins

2.16 Transfection of HEK-293T cells with RAG expression vectors

The day before transfection, 10 cm² dishes (typically 5 per transfection) were seeded with 3.5×10^6 HEK-293T cells/dish in complete DMEM and incubated at 37°C/5% CO₂ overnight. The following day, the medium was changed to fresh complete medium 3 hours before transfection. HEK-293T cells were then co-transfected with either pEF-MBP-core RAG1 (pEF-McR1) and pEF-MBP-core RAG2 (pEF-McR2) or pEF-MBP-full-length RAG2 (pEF-MflR2) or pEF-Myc-MBP-core RAG1 (pEF-MMcR1) and pEF-HA-MBP-core RAG2 (pEF-HMcR2) using polyethylenimine (PEI). The transfection solution was prepared in antibiotic- and serum-free opti-MEM. 5 μ g of each plasmid and 30 μ g of PEI were diluted in 500 μ l of opti-MEM, respectively. Opti-MEM diluted PEI was then added to opti-MEM diluted DNA and the DNA/PEI mixture was then vortexed thoroughly for 10 s, followed by incubation at room temperature for 15 min. 1 ml of DNA/PEI mixture was then added drop-wise to each dish and distributed evenly by gentle rocking back and forward. After incubation for 48 hours, the medium was aspirated, and the cells were harvested in 2 ml of

sterile phosphate-buffered saline-EDTA (137 mM NaCl, 27 mM KCl, 43 mM KH_2PO_4 , 14 mM Na_2HPO_4 , 2 mM EDTA) per plate and collected in a 50-ml conical tube on ice. The tubes were then centrifuged at 1000 x *g* for 3 min at 4°C and the supernatant was removed and the cell pellets washed twice in ice cold PBS. Cell pellets were frozen in a dry ice and stored at -80°C until use.

2.17 Purification of core RAG proteins from HEK-293T cells

All steps in purification of core RAG or tagged core RAG proteins were performed at 4°C. For purification, cell pellets (typically from 10 dishes per purification) were thawed on ice and resuspended in a total of 10 ml of lysis buffer (0.5 M NaCl, 10 mM NaP_i [pH 7.2], 0.25% TWEEN-20, 1 mM DTT, 1 mM benzamidine, 1 x complete protease inhibitors with EDTA (Roche) and 0.1 mM PMSF). Cells were then lysed using 20 strokes of a Dounce homogenizer with tight fitting pestle. The lysates were clarified by centrifugation at 11,000 x *g* (Beckman SW55Ti rotor, 30,000 rpm) for 40 min at 4°C. The supernatant was bound to amylose resin (NEB, # E8021, Ipswich, MA, USA; 1ml) equilibrated in buffer A (0.5 M NaCl, 10 mM NaP_i [pH 7.2], 0.25% TWEEN-20 and 1 mM DTT). The column was washed with 10 ml buffer A, followed by washing with 10 ml buffer B (0.5 M NaCl, 10 mM NaP_i [pH 7.2] and 1 mM DTT). The MBP-RAG proteins were eluted with buffer C (0.5 M NaCl, 10 mM NaP_i [pH 7.2], 1 mM DTT and 10 mM maltose). Fractions were analyzed by SDS-PAGE, and RAG-containing fractions were pooled and dialysed twice for 3 h against dialysis buffer (25 mM Tris-HCl [pH8.0], 150 mM KCl, 2 mM DTT, 10% glycerol). Dialyzed protein was aliquoted and snap-frozen on dry ice and stored at -80°C.

2.18 Purification of core RAG1 plus full-length RAG2 proteins from HEK-293T cells

For purification core RAG1 plus full-length RAG2 proteins, each thawed cell pellet was resuspended in a total of 1.5 ml of cold RSB buffer (10 mM Tris pH 7.4, 10 mM NaCl, 5 mM MgCl_2 , 0.5 % (v/v) NP-40, 1 mM PMSF) and incubated for 5 minutes on ice. Following this, 2.25 ml of cold LSB buffer (20 mM Tris [pH

7.4], 1 M NaCl, 0.2 mM MgCl₂, 0.2 % (v/v) NP-40, 1 mM PMSF) was added to each pellet and the mixture was mixed with gentle agitation at 4 °C for 1 hour. The lysates were centrifugated at 11,000 x g (Beckman SW55Ti rotor, 30,000 rpm) for 40 min at 4 °C and the supernatants were bound to amylose resin (NEB, # E8021, Ipswich, MA, USA; 1ml) equilibrated in 1:1.5 RSB/LSB by rocking on the roller for 30 minutes at 4 °C. Following this, the mixture was applied to an econo-column and the flowthrough was saved for further analysis. Then, the column was washed with 5 ml of RSB/LSB buffer (RSB:LSB = 1:1.5), followed by washes with 4 ml of WB buffer (20 mM Tris [pH 7.4], 0.5 M NaCl and 5 mM MgCl₂). The proteins were eluted with EB buffer (20 mM Tris [pH 7.4], 0.5 M NaCl and 5 mM MgCl₂, 10 mM maltose). Fractions were analyzed by SDS-PAGE, and RAG-containing fractions were pooled and dialysed twice for 3 h against dialysis buffer (25 mM Tris-HCl [pH8.0], 150 mM KCl, 2 mM DTT, 10% glycerol). Dialyzed protein was aliquoted and snap-frozen in a dry ice and stored at -80°C.

2.19 SDS-Polyacrylamide gel electrophoresis of proteins

Before loading onto protein gels, protein samples were mixed well with one-third volume of 4 x protein loading buffer (0.2 M Tris [pH 6.8], 40% glycerol, 4% SDS, 600 mM β-Mercaptoethanol, 0.05% bromophenol blue) by vortexing and heating in boiling water for 3 min, to ensure that the proteins were denatured. Typically, protein samples were analyzed on a 10% (37.5:1 acrylamide:bis-acrylamide) polyacrylamide separating gel, with a 3% (37.5:1 acrylamide:bis-acrylamide) stacking gel to concentrate the samples. Stacking and separating gel solutions were prepared according to the protocol in Sambrook et al. (1989) and cast using the BioRad Protean III system. Denatured samples were loaded onto the gel and separated for 1 h at 180 V. Following electrophoresis, proteins were visualized by Coomassie brilliant blue staining (Section 2.21) or western blotting (Section 2.23).

2.20 Coomassie brilliant blue staining of polyacrylamide gel

After SDS-PAGE (Section 2.19), the glass plates were carefully separated and the polyacrylamide gel was immersed in fixing solution (10% (v/v) acetic acid,

25% (v/v) isopropanol) and incubated on an orbital shaker at 25°C for 10 min. Fixing solution was then replaced with Coomassie brilliant blue staining solution (0.006% (w/v) Coomassie Brilliant Blue G-250, 10% (v/v) acetic acid) for 30 min, followed by destaining the gel with 10% (v/v) acetic acid until the background was clear.

2.21 Estimation of protein concentration by SDS-PAGE and densitometry

The concentration of purified proteins was estimated by running a range of known BSA concentrations on an SDS-PAGE (Section 2.19) gel alongside an aliquot of purified protein. Gels were stained with Coomassie brilliant blue (Section 2.20) and photographed using a SynGene white light transilluminator. The band densities were measured using AIDA Image Analyser software (Raytest, Germany, v.4.14). A BSA standard curve was plotted and the equation of the line of best fit was used to estimate the amount of purified protein in each band.

2.22 Identification of tagged RAG proteins by western blotting

Electrophoresed SDS-PAGE (Section 2.19) gel were transferred to polyvinylidene fluoride (PVDF) membrane (Hybond P, GE Healthcare #10600023) for western blot analysis. PVDF membrane was first soaked in 100% methanol and then washed with ddH₂O for three times. The soaked PVDF membrane was then immersed in transfer buffer (150 mM glycine, 20 mM Tris-HCl [pH 8.0], 0.038 % (w/v) SDS, 20% (v/v) methanol) for 10 minutes, together with the gel and six pieces of Whatman® 3MM filter paper. Three pieces of pre-treated Whatman® 3MM filter paper were placed on bottom, followed by PVDF membrane, SDS-PAGE gel and three more Whatman 3MM filter paper. The proteins on the gel were then transferred onto PVDF membrane by running a semi-dry blotter at 0.68 mA/cm² for 1 hour. Non-specific binding sites were blocked in 5 % (w/v) non-fat dried milk (Marvel) in TBS-T (150 mM NaCl, 50 mM Tris-HCl [pH 7.6], 0.05% Tween®20) for 1 hour at room temperature on a roller. The blocked membrane was then hybridized in 5 ml of 1:5000 diluted anti-Myc (abcam #ab9106) (for Myc-tagged RAG1) or

anti-HA (abcam #ab137838) (for HA-tagged RAG2) antibodies in 5% (w/v) non-fat dried milk (Marvel) in TBS-T for overnight at 4 °C, followed by washing the membrane for 5 times with TBS-T to remove unbound primary antibody, replacing the TBS-T every 5 minutes. The secondary hybridization was with 1:20,000 diluted horse radish peroxidase (HRP)-conjugated polyclonal anti-goat (Alpha Diagnostics #30222-200) (for anti-Myc) or anti-rabbit (GE Healthcare #NXA931) (for anti-HA) antibody in 5% (w/v) non-fat dried milk (Marvel) in TBS-T for 1 hour at room temperature. After this, the unbound secondary antibody was removed as above by washing the membrane five times with TBS-T. Blots were processed by adding 1 ml of 1:1 mix of luminol enhancer and peroxide solution (ThermoPierce Chemiluminescence Kit #34077) onto the membrane for 3 minutes, prior to being exposed to Syngene G:Box Chemi XRQ imaging system. Blots were stripped (0.36 M Tris [pH 6.8], 12% SDS, 100 mM β -mercaptoethanol) for 15 minutes at 50 °C, washed and blocked prior to reprobing.

D) *In vitro* RAG cutting assay

2.23 Gel purification of DNA oligonucleotides

The oligonucleotides (Table 2.1) were synthesised by Sigma-Aldrich (Poole, UK) or Integrated DNA Technologies (IDT) and purified by desalting. After synthesis, DNA oligonucleotides were gel-purified using the “crush-and-soak” method (Ellington & Pollard, 2001). The lyophilized oligonucleotides were resuspended in 200 μ l of formamide loading buffer, mixed thoroughly by vortexing and boiled for 3 min. The samples were then loaded onto a pre-run 12% (19:1) denaturing polyacrylamide gel containing 7 M urea and 1x TBE (12% denaturing gel) for 1 h at 500 V. Following electrophoresis, one lane was excised and stained with EtBr for 10 min, whilst the rest of gel was wrapped in cling film. After staining, the EtBr-stained gel slice was aligned with the remainder of the gel to locate the unstained oligonucleotide. The oligonucleotides were then excised using a sterile scalpel blade. The gel slice was finely diced on a clean glass plate with elution buffer (TE, 100 mM NaOAc, 0.5% SDS) and then transferred to a 1.5 ml tube, filled with elution

buffer. Tubes were boiled for 3 min and then freeze-thawed 3 times in dry ice and a 42°C water bath before being rotated on a wheel at room temperature (RT) overnight for the oligonucleotide to diffuse from the gel. The next day, the samples were centrifuged at full speed for 2 min to remove the polyacrylamide and the supernatant was then transferred to a fresh 1.5 ml centrifuge tube. Residual small pieces of polyacrylamide were removed by filtering through a 0.2 µm syringe filter unit. The filter was finally flushed with 200 µl of TE to recover all the eluted oligonucleotides. The volume of oligonucleotide solution was reduced to 400 µl by repeated butanol extraction, followed by phenol-chloroform extraction and ethanol precipitation. DNA pellets were resuspended in 100 µl of dH₂O and the concentration of purified oligonucleotide was measured using a NanoDrop spectrophotometer (A²⁶⁰).

2.24 Radiolabelling of oligonucleotides

Purified oligonucleotides were 5' end labelled with γ -³²P ATP using T4 polynucleotide kinase (T4 PNK, NEB, #M0201S). Labelling reactions contained 10 pmol of the oligonucleotide strand to be labelled (for *in vitro* RAG cutting assay and signal end release assay, the top strand and bottom strand were labelled, respectively), 10 units T4 PNK, 1.5 µl 10x T4 PNK buffer (70 mM Tris [pH 7.6], 10 mM MgCl₂, 5 mM DTT) and 2 µl γ -³²P ATP (3000 Ci/mmol, Perkin Elmer) in a total volume of 15 µl. These were incubated at 37°C for 30 min. The volume was then increased to 50 µl with TE, prior to removing unincorporated γ -³²P ATP by centrifugation for 2 minutes at 700 x g in an illustra MicroSpin G-25 column (GE Healthcare). The labelled oligonucleotides were then annealed with 10 pmol of its unlabelled complementary strand.

2.25 Annealing oligonucleotides

Oligonucleotide annealing reactions containing 100 pmol of each of the complementary oligonucleotide strands, 0.1 M NaCl and 44 µl TE in a total volume of 100 µl were heated in screw-cap tubes for 2 min in a boiling water bath followed by cooling gradually to room temperature overnight. The final

concentration of annealed oligonucleotide was 100 nM. Labelled and unlabeled oligonucleotides were annealed in the same way except that 10 pmol of labelled oligo were annealed.

2.26 *In vitro* RAG cutting assays

Typically, RAG cutting reactions contained 1 µl of 10 x buffer (25 mM MOPS [pH 7.0], 100 µg/ml BSA, 50 mM potassium acetate, 1 mM MgCl₂, and 1 mM DTT), 30 ng (120 nM) HMGB1, 2 nM radiolabeled oligonucleotide, 20 nM unlabeled partner oligonucleotide and about 80 nM of purified RAG proteins in a total volume of 10 µl. These were incubated at 37°C for 15 min for authentic RSSs. The cleavage reactions were stopped by adding 3.3 µl of 4 x stop buffer (200 mM Tris-HCl [pH 8.0], 0.7 mg/ml proteinase K (PK), 20 mM EDTA, 0.4% SDS, 6 x DNA loading dye), prior to incubation at 37°C for 30 min. PK and SDS were added in the stop buffer to degrade the RAG proteins at the end of the reaction. This step is necessary to ensure accurate quantification of the RAG cleavage level on preventing some of the radiolabeled oligonucleotide from being bound by HMGB1 and RAG proteins, which becomes stuck in the well during electrophoresis. After loading onto a 12% (19:1) polyacrylamide gel in 1x TAE, the samples were separated for 90 min at 170 V. The glass plates were then carefully separated. Due to the high percentage of polyacrylamide, the gel is prone to crack when drying. In order to prevent this, 70% methanol was used to soak the gel on an orbital shaker for 20 min at room temperature. Gels were then transferred on to Whatman® 3 MM chromatography paper and dried for 1 h using a Bio-Rad gel drier with an 80°C heated lid. Following this, the dried gel was exposed to a phosphorimager screen for at least 2 hours prior to visualization using an FLA-1500 phosphorimager (Fujifilm). The cleavage percentage of cleaved and uncleaved DNA was quantified using AIDA Image Analyser software (Raytest, Germany, v.4.14).

2.27 Electrophoretic mobility shift assay (EMSA)

RAG-RSS binding assays were carried out in the same way as cleavage reactions, except that MgCl₂ was replaced by CaCl₂. After incubation, the samples were immersed in ice for 5 min, prior to loading onto a 4% (19:1)

polyacrylamide gel in 0.5x TBE pre-chilled to 4°C in the cold room. The gels were run and dried as the same as for cleavage assays except electrophoresis was for 2 hours at 4°C. Moreover, the gels were not soaked in 70% methanol. The dried gel was exposed to a phosphorimager screen for at least 2 hours prior to visualization using an FLA-1500 phosphorimager (Fujifilm).

2.28 Supershift assay

Supershift assays were carried out in the same way as RAG-RSS binding assays, except that the oligos used were fluorescently-labelled and the RAG proteins were Myc-tagged core RAG1 and HA-tagged core RAG2. After incubation for 15 minutes at 37 °C, 0.5 µl anti-Myc antibody (Abcam #ab9106) or anti-HA antibody (Abcam #ab137838) was added and the reactions were immersed in ice for 10 min, prior to loading onto a 4% (19:1) polyacrylamide gel in 0.5 x TBE pre-chilled to 4°C in the cold room. The gels were run for 3 hours at 4°C and visualized using an FLA-1500 phosphorimager (Fujifilm).

2.29 Signal end release assay

The reactions were set up as for *in vitro* RAGs cutting assays as described above except the oligonucleotides were labelled on the other strand to detect signal ends. Following cutting for 15 minutes at 37°C, the reactions were challenged at different temperatures ranging from 37 °C to 60 °C for 30 minutes. Half of the reaction was resolved on an 8% polyacrylamide native gel whilst the remainder was resolved on a 10% cutting gel. End release was normalised to the level of cutting with core RAG1 plus core RAG2 or full length RAG2.

2.30 Preparation of fluorescently- labelled RSSs and ESC

To make fluorescently-labelled 12-RSS and 23-RSS, annealed 12-RSS and 23-RSS oligonucleotides were used as templates for labelling by PCR, respectively. To make fluorescently-labelled ESC, the pBS-ESC-Label was made. Specifically, the annealed ESC oligonucleotides were ligated to the pBS-Label vector which was modified from pBS-12RSS-Label (a gift from

Daniel Thwaites) by removing the 12RSS sequence. The resulting pBS-ESC-Label was digested with Hind III and Xba I at 37 °C for 4 hours, followed by gel-purification by electroelution to make template for ESC labelling. Fluorescently-labelled RSSs or ESC were prepared by PCR (50 µl final volume) containing 0.2 µM fluorescently-labelled universal_488_fwd primer and 0.2 µM unlabelled RSS_ or ESC_labelling_rvs primer (Table 2.7), 1 ng DNA template, 0.2 µM dNTP mix, 1 x ThermoPol buffer and 2.5 units of Taq DNA polymerase (NEB #M0237). The thermocycling conditions are as follows: initial denaturation at 94 °C for 2 minutes, followed by 30 cycles of 94 °C for 5 seconds, 60°C for 20 seconds, 72°C for 20 seconds, with a final extension at 72°C for 2 minutes. The PCR reaction was loaded onto a 10 % acrylamide gel and run for 3 hours at 4°C at 170 V. The target bands were excised and purified by electroelution at 60 V for 2 hours, followed by 4 x phenol/chloroform extraction (Section 2.9) and 1 x ethanol precipitation (Section 2.10).

2.31 DNase I footprint analysis

The ESC footprint probe was prepared by digestion of 100 µg of ESC containing plasmid (pcDNA3.1-ESC) with Nhe I and Apa I. The digestion products were gel-purified via electroelution, followed by 2 x phenol/chloroform extraction and 1 x ethanol precipitation. The DNA pellet was resuspended in 100 µl of dH₂O. The concentration was then measured using a NanoDrop spectrophotometer and the number of moles was calculated. The labelling reaction contained 5 pmol of digested and purified ESC fragment, 50 µCi of [³²P] dCTP, 5 µl of 10 x NEB buffer 2, 10 units of Klenow DNA polymerase in a 50 ul reaction. This was incubated for 25 min at room temperature, followed by addition of 2 µl of 5 mM dNTP mix and incubation for 5 min at room temperature to completely fill-in the sticky Nhe I end. After removing unincorporated nucleotides with a G-25 spin column (GE Healthcare) and ethanol precipitation, the DNA pellets were washed twice with 70% ethanol and resuspended in 50 µl of TE buffer [pH 8.0], generating a labelled ESC with a concentration of 100 nM.

Binding reactions contained 1x binding buffer (25 mM MOPS [pH 7.0], 42 mM KOAc, 0.1 mg/ml BSA, 1 mM CaCl₂, 2 mM DTT), 20% DMSO, ESC probe

(10,000-15,000 counts per minutes per lane), 0.2 µg polydI.dC, and 100 nM cRAG proteins in a total volume of 100 µl. After incubation for 10 minutes at 37 °C, 0.5 µl 1M MgCl₂ and 2 units of DNase I (diluted in 40% glycerol) mix was added. Subsequently, the samples were incubated at room temperature for 1 min, followed by addition of 3 µl 0.5 M EDTA (15 mM final concentration) to stop the DNase I reactions. The volume of sample was increased to 400 µl with TE (10 mM Tris, 15 mM EDTA). The samples were then purified by phenol/chloroform extraction and ethanol precipitation, using glycogen to visualize the small DNA pellet. This was then resuspended in 3 µl of formamide loading buffer, boiled for 3 min prior to running on a pre-run 6% (19:1) denaturing polyacrylamide gel alongside the appropriate G+A ladder for 45 min at 38 mA. Following this, gels were transferred to Whatman® 3MM paper and dried using a BioRad vacuum drier for 45 min with lid heated to 80°C. The dried gel was exposed to a phosphorImager screen for at least 2 hours and visualised using an FLA-1500 PhosphorImager (FujiFilm).

2.32 Preparation of G+A ladders

To determine the identity of each band in the DNase I footprint assay, a G+A ladder was prepared using the same radiolabelled ESC probe as used in footprint. The G+A ladder was prepared as follows: one fifth (10 µl) of the labelled probe stock was ethanol precipitated to remove residual salt and resuspended in 10 µl dH₂O, followed by addition of 25 µl of 100% formic acid and thorough mixing. After 2 minute's incubation at room temperature, the sample was added to 200 µl stop buffer (300 mM NaOAc [pH 5.2], 0.1 mg/ml yeast tRNA, 0.1 mM EDTA [pH 8.0]) and 150 µl of 100% ethanol chilled to -20°C to quench the reaction. Following this, DNA was precipitated by centrifuging at 14,000 x *g* for 10 min and washed twice with 70% ethanol. Subsequently, the air-dried pellet was resuspended in 200 µl of dH₂O, followed by 1 x ethanol precipitation and the DNA pellet was dried using a SpeedVac prior to being resuspended in 70 µl of 10% (v/v) piperidine. The resuspended sample was then incubated at 90°C for 30 min and the sample transferred to a new 1.5 ml centrifuge tube and piperidine was evaporated using a SpeedVac. The DNA pellet was resuspended in 30 µl of dH₂O and transferred to a fresh tube to minimise piperidine contamination, followed by being dried again using

the SpeedVac. The DNA pellet was resuspended in 20 μ l dH₂O and transferred to a new tube and dried again using the SpeedVac. The final DNA pellet was then resuspended in 10 μ l of formamide loading buffer and boiled for 3 min before being stored at 4°C.

E) *In vivo* RAG cutting assay

2.33 Generation and concentration of lentivirus

12-RSS, 23-RSS and ESC lentiviruses were generated by transfection relevant plasmids into HEK293T cells. Specifically, HEK293T cells were seeded at a density of 2.5×10^6 cells/plate five 10 cm² dishes. The following day, the medium was changed to antibiotic-free medium and three hours later, cells were transfected with 30 μ g PEI and 4.9 μ g of lentiviral vector (pWPI-12RSS, pWPI-23RSS or pWPI-ESC), 2.6 μ g of the LV packaging vector, pR8.74 D64V integrase deficient mutant or pR8.74 (Addgene Plasmid #22036) and 2.5 μ g of the vector that expresses the coat protein, pMD2.G (Addgene Plasmid #12259). After 48 hours, the growth medium containing the virus was harvested by centrifugation at 1600 rpm for 3 minutes to remove cell debris. The supernatant was then added quarter volume of sucrose buffer (50 mM Tris-HCl [pH 7.4], 100 mM NaCl, 0.5 mM EDTA, 10% sucrose), followed by centrifugation at 11,000 $\times g$ for 4 hrs. The supernatant was discarded and the pellet was resuspended in appropriate amount of PBS overnight at 4°C. Aliquots of the concentrated virus were stored at -80 °C until use.

2.34 Transfection of NIH3T3 cells

Transfection of NIH3T3 cells was performed using polyethyleneimine (PEI). The day before transfection, 1×10^5 of NIH3T3 cells were plated per well in a 6-well plate. The culture medium was changed 3 h before transfection to antibiotic-free medium. DNA and PEI were diluted in same volume of opti-MEM and then mixed together by vortexing thoroughly. After 15 minutes' incubation at room temperature, the DNA/PEI mixture was then added drop-wise to each well on the 6-well plate and mixed gently by rocking back and forth. For *in vivo* RAG cutting assays, cells were transfected with 0.5 μ g of

RSS containing plasmid and/or 0.5 µg of ESC plasmid, and 1 µg of each of full-length RAG1 (pJH548) and core RAG2 (pEFXC-cRAG2). The cells were harvested 48 h post-transfection. Plasmid DNA was extracted according to the HIRT method (Section 2.37), and resuspended in 30 µl of water. 1 ng of DNA was analysed by qPCR (Section 2.39).

2.35 Spinfection of REH cells

Typically, REH cells were seeded into a 12-well dish at a density of 1×10^6 cells/well and mixed with 100 µl of concentrated or 500 µl of unconcentrated lentivirus with 8 µg/ml of polybrene, followed by centrifugation at $800 \times g$ for 90 min at 32 °C. 500 µl of medium was then added to cells to dilute polybrene which is harmful to the cells. 4 hours post-transduction, cells were centrifuged at $800 \times g$ for 10 min at 32 °C. The supernatant was carefully removed and 2 ml of fresh medium was added, prior to incubation at 37°C.

2.36 Analysis of level of transfection or transduction by flow cytometry

For both adherent cells and suspending cells, 20% of transfected or transduced cells were used to check efficiency. Cells were harvested by washing twice with PBS. For adherent cells, 1 ml trypsin was added and for suspending cells, cells were collected directly, prior to centrifugation at $600 \times g$ for 3 minutes. Cells were then resuspended in 400 µl of FACS buffer (PBS, 2% FCS, 25 mM HEPES-KOH [pH 7.9], 1 mM EDTA) and filtered through a 40 µm filter. The transfections of green fluorescence protein expressing cells and untransfected cells were determined by flow cytometry using a Becton Dickinson BD-LSRFortessa.

2.37 Recovery of DNA from mammalian cell by Hirt extraction

DNA was isolated from the transfected or transduced cells by Hirt extraction. Specifically, culture medium was gently aspirated and the cells were washed twice with 3 ml of room temperature PBS. After completely removing the PBS, 0.8 ml of Hirt solution I (0.6% SDS, 10 mM Tris [pH 8.0], 1 mM EDTA [pH 8.0]) was added and mixed by swirling the plate gently to lyse the cells. The

samples were incubated at room temperature for 10 min, followed by tilting back and forth and swirling gently for 30 s. 0.2 ml of room temperature Hirt solution II (5 M NaCl, 10 mM Tris [pH 8.0], 1 mM EDTA [pH 8.0]) was then added and mixed by tilting and swirling gently and briefly. After 2 minute's incubation at room temperature, the samples were tilted and swirled again to lyse the cells. The mixture was then transferred into 1.5 ml standard polypropylene centrifuge tube. The samples were stored at 4°C overnight. The following day, the samples were centrifuged at 15,000 x g for 40 min at 4°C. The supernatant was transferred to fresh tube, taking care to avoid white clumps, followed by 2 x phenol/chloroform extraction. 0.5 ml of butanol was then added and mixed by vortexing, followed by centrifuging at 12,000 x g for 1 min. The top butanol layer was then removed to waste and the aqueous layer was equally distributed into two fresh tubes, followed by 1x ethanol precipitation (this ethanol precipitation was done with 200 µl 7.5 M NH₄OAc rather than NaOAc) and washing with 70% ethanol. The pellet was then resuspended in appropriate volume of TE. The resulting DNA was then stored at -20°C until use.

2.38 Detection of RAG cutting in REH cells

Cells were transduced by spinfection (Section 2.35). After 48 hrs' incubation at 37 °C, the cells were mixed well and 1/5 volume of cells were used to carry out FACs analysis to check infection efficiency (Section 2.36), ensuring it at 35%~55%. The remaining cells were used to recover DNA by Hirt extraction (Section 2.37). The resulting DNA was used to analyse RAG cleavage by qPCR (Section 2.39) using the primers given in Table 2.3.

2.39 Analysis of RAG cleavage by Quantitative Polymerase Chain Reaction (qPCR)

The primers used to determine the amount of cut plasmid and the total amount of plasmid are given in Tables 2.2 and 2.3. Typically, a qPCR reaction contained 5 µl of SYBR green master mix (2x) (NoROX SensiMix, Bioline, #PB201405), 1 µl of diluted template DNA, 8 pmol of each primer, and ddH₂O to a total volume of 10 µl. qPCR reactions were carried using a Corbett Rotor-

Gene 6000 machine and a three step plus melt programme. The temperature conditions are as follows: 95°C for 3 min, followed by 40 cycles of 95°C for 5 s, annealing at T_m for 15 s and extension at 72°C for 10 s. A melt curve was generated by detecting fluorescence as the temperature is increased from 72°C to 95°C to test if a single, specific product was amplified. The results were analyzed using the Corbett Rotor-Gene 6000 Series Software (v.1.7, build 87). In each case, a standard curve of samples containing the amplified region was analyzed to calculate the relative amount of product in each of the unknown samples. A no-template water control was included in each run to test for contamination.

2.40 Calculation of virus copies per cell

Genomic DNA was isolated from 1×10^6 transduced or non-transduced REH cells (Section 2.46). Following this, a 680 bp fragment was amplified from the genome using the primer pair genomicDNA_PCR(680bp)_fwd and genomicDNA_PCR(680bp)_rvs (Table 2.4), with an annealing temperature of 59 °C. This PCR product was purified by gel electrophoresis and electroelution (Section 2.11), followed 4 x phenol/chloroform extractions (Section 2.9) and 1 x ethanol precipitation (Section 2.10). The DNA pellet was then resuspended in 50 μ l of TE and the concentration determined using a nanodrop spectrophotometer. The copies of the genomic fragment per ml were calculated using the following formula:

$$\text{number of copies (molecules)} = \frac{X \text{ ng} * 6.0221 \times 10^{23} \text{ molecules/mole}}{(N * 660 \text{ g/mole}) * 1 \times 10^9 \text{ ng/g}}$$

where X = amount of amplicon (ng)

N = length of dsDNA amplicon

660 g/mole = average mass of 1 bp dsDNA

The concentrations of pWPI-12RSS, pWPI-23RSS and pWPI-ESC were similarly measured. The genomic fragment and each of the pWPI plasmids were diluted to 1×10^{10} copies/ml and a 10-fold serial dilution ranging from 1×10^8 to 1×10^2 copies/ml was generated and used as a standard curve for qPCR. The copies of genomic DNA in the test samples was determined using

the primer pair genomicDNA_qPCR(200 bp)_fwd and genomicDNA_qPCR(200 bp)_rvs (Table 2.3). The copies of 12-RSS, 23-RSS and ESC viruses were determined using the primer pairs, 12RSS_virus_norm_fwd plus virus_norm_rvs, 23RSS_virus_norm_fwd plus 23RSS_virus_norm_rvs, and ESC_virus_norm_fwd plus virus_norm_rvs, respectively. The copies of 12-RSS, 23-RSS or ESC-containing viruses per cell were calculated by dividing the copies of the 12-RSS, 23-RSS or ESC by the copies of the genomic sample and then multiplying by 2 (chromosome 14, from which the 680 bp fragment was amplified, is diploid in REH cells; <https://www.dsmz.de/catalogues/details/culture/ACC-22.html>).

2.41 Ligation-mediated PCR (LM-PCR) assay

In vivo RAG cutting reactions were performed by transfection of 0.5 µg of 23-RSS and ESC containing plasmids into 293T cells together with 1 µg of RAG1 and RAG2 expression plasmids. 48 hours post-transfection, cells were harvested and plasmid DNA was recovered via Hirt extraction (Section 2.42). To open potential hairpin ends, 5 µg Hirt extracted DNA was treated with 33 U of mung bean nuclease (MBN; NEB #M0250S) for 30 min at 30 °C in MBN buffer (30 mM NaCl, 50 mM sodium acetate, 1 mM ZnSO₄, pH 5.0), followed by 4 x phenol/chloroform extractions (Section 2.9) and 1 x ethanol precipitation (Section 2.10). To ensure that the ends treated by MBN were blunt, T4 DNA Polymerase (NEB #M0203S) was used to fill in the ends. Each reaction contained MBN-treated DNA, 1x NEbuffer™ 2.1 (50 mM NaCl, 10 mM Tris-HCl, 10 mM MgCl₂, 100 µg/ml BSA pH 7.9), 100 µM dNTPs and 1 unit of T4 DNA Polymerase and was carried out at 12 °C for 15 minutes. The reaction was stopped by adding EDTA to a final concentration of 10 mM and heating to 75°C for 20 minutes, followed by 4x phenol/chloroform extractions (Section 2.9) and 1x ethanol precipitation (Section 2.10). Ligation reactions contain 5 pmol of annealed FM25/11 (Table 2.4), 5 ng of 23-RSS plasmid digested with BamH I and blunted as positive control or 50 ng of untreated or MBN treated and blunted DNA sample, 3 µl of 10 x T4 DNA ligase buffer (50 mM Tris-HCl, 10 mM MgCl₂, 1 mM ATP, 10 mM DTT, pH 7.5), 200 units of T4 DNA ligase (NEB #M0202S) in 30 µl final volume and were carried out at 16 °C overnight.

After ligation, DNA was purified by 4 x phenol/chloroform extractions (Section 2.9) and 1 x ethanol precipitation (Section 2.10) and resuspension of the sample in 10 μ l of dH₂O. LM-PCR products were detected by nested PCR (Section 2.54). In the first round, PCR reactions contained 1x ThermoPol reaction buffer (20 mM Tris-HCl, 10 mM (NH₄)₂SO₄, 10 mM KCl, 2 mM MgSO₄, 0.1% Triton ® X-100, pH 8.8), 300 nM of each dNTP, 0.6 μ M each of RSS_fwd_1st (signal end) or 23-RSS_rvs_1st (coding end) and FM25 primers (Table 2.4), 1 μ l of DNA template and 2 units of Vent DNA polymerase (NEB #M0254S) in a 30 μ l reaction volume. Thermocycling conditions were as follows: pre-denature at 94 °C for 5 min, followed by 18 cycles of 94 °C for 30 s, 56 °C (signal end) or 63 °C (coding end) for 20 s and 72 °C for 20 s. A 7 min of final extension was performed. In the second round of PCR, 1 μ l of the first round PCR product was used in a similar reaction to the first round, except that the RSS_fwd_2nd (signal end) or 23-RSS_rvs_2nd (coding end) (Table 2.4) primers were used. Thermocycling conditions were the same as the first round except the annealing temperature was 60 °C (signal end) or 65 °C (coding end) and 28 cycles were used. Reaction products were resolved on a 2% agarose gel and visualized by UV light.

2.42 Immunofluorescence to detect γ H2AX foci

For imaging cells using confocal microscope, cells need to be seeded on coverslips. Before transfection (for Cos7 cells) or transduction (for REH cells), coverslips were rinsed in 100% ethanol and placed in a 6-well dish. After air-drying, a meniscus of poly-L-lysine was added and left for 3-5 minutes. After washing the coverslips in 2 ml of PBS and allowing to air dry, 1x10⁵ cultured Cos7 cells or 1x10⁶ cultured REH were plated in the dishes. The Cos7 cells were cultured at 37°C for overnight before transfection whereas the REH cells were transduced directly. For transfection, 0.25 μ g of full-length RAG1 and 0.25 μ g of core RAG2 expression plasmids, with or without 0.5 μ g of ESC-carrying plasmid were co-transfected into the Cos7 cells using PEI as described in Section 2.34. For transduction, 100 μ l of concentrated control or ESC viruses were added into wells, followed by spinfection (Section 2.35) at 800 x g for 90 minutes. After incubation for 24 hrs, cells were carefully washed in 2 ml of PBS, followed by fixing in 2 ml 4% (v/v) paraformaldehyde in PBS for

10 min at room temperature. Cells were then carefully washed three times with 2 ml of PBS, prior to permeabilising with 2 ml 1% (v/v) Triton® X-100 in PBS for 10 min at RT. After washing the cells three times with 2 ml of PBS, a solution of 5% (w/v) BSA in PBS was added to block non-specific antibody-binding for 1 hour at 37°C. After aspirating the blocking solution, 1 ml of primary anti-phospho-histone γ H2Ax (Ser 139) antibody (Cell Signalling #2577) (1:500 dilution in PBS/1% (w/v) BSA) was added. After incubation for 1 hour at 37°C, the primary antibody was aspirated and the coverslips were carefully washed 5 x with PBS. Subsequently, 100 μ l of secondary FITC-anti-rabbit (1:500 dilution in PBS/1% (w/v) BSA; for Cos7 cells) or 1 ml of AlexaFluor568 (red)-labelled secondary antibody (Invitrogen #A-11011; 1:400 dilution in PBS/1% (w/v); for REH cells) was added. After incubation for 1 hour at 37°C, the secondary antibody was aspirated and the coverslips were carefully washed 5 x with PBS. Finally, 2.5 μ g DAPI was added to each well and left for 5 minutes, followed by carefully washing 5 x with PBS. Before analysis using a confocal microscope (Bioimaging Zeiss LSM880 Inverted + Airyscan), the coverslips were mounted onto slide using ProLong™ Gold Antifade Mountant (Invitrogen P36934).

2.43 Apoptosis assay

Before transduction, 1×10^6 cultured cells were seeded in a 12-well dish. 500 μ l of unconcentrated pWPI or ESC lentiviruses were added to the cells. After 48 hrs' incubation at 37 °C, cells were used for an apoptosis assay using eBioscience™ Annexin V Apoptosis Detection Kit APC (Invitrogen 88-8007). Specifically, 5×10^5 cells were collected by centrifugation and resuspended in 500 μ l of 1x binding buffer. 5 μ l of APC-conjugated Annexin V was added to 100 μ l of the cell suspension, followed by incubation for 10-15 minutes at room temperature. 5 μ l of Propidium Iodide Staining Solution was then added and the cells were analysed by flow cytometry using a LSRFortessa (Becton Dickinson).

F) B Cells isolation

2.44 Preparation of B cells for FACs with NH₄Cl treatment

Mouse femurs or spleens were collected from 5-7 week old mice by Joan Boyes. To collect bone marrow, femurs were flushed with 10 ml PBS, followed by centrifugation at 600 x g for 3 minutes. To collect spleen, cells finely diced pieces of spleen were placed onto a strainer attached to a 50 ml conical tube, followed by being pressed through the strainer and washing with PBS before centrifugation at 600 x g for 3 minutes. The cell pellet was then resuspended in 10 ml of 0.168 M ammonium chloride (NH₄Cl) and incubated at room temperature for 10 minutes to lyse red blood cells. Cells were then washed with 40 ml of ice cold PBS and resuspended in 1 ml of FACS staining solution (PBS, 2% FCS, 25 mM Hepes-KOH [pH 7.9], 1 mM EDTA) in a 50 ml falcon tube.

2.45 Isolation of B cells by FACS

Staining of cells prior to sorting by flow cytometry was as follows. Before being sorted, cells were stained with different antibodies. Each tube (from one mouse) of pre-treated bone marrow cells were stained with 6 µl of Fluorescein Isothiocyanate (FITC) labelled anti-CD19 antibody (BD Pharmingen™ # 553785) and 6 µl of R-phycoerythrin (PE) labelled anti-CD43 antibody (BD Pharmingen™ #553271) for pre-B cells or 10 µl of Fluorescein Isothiocyanate (FITC) labelled anti-IgM (BD Pharmingen™ #553408) for bone marrow IgM cells. Each tube (from one mouse) of pre-treated spleen cells were stained with 10 µl of Fluorescein Isothiocyanate (FITC) labelled anti-IgM (BD Pharmingen™ #553408) or 15 µl R-phycoerythrin (PE) labelled anti-IgG antibody (eBioscience #12-4010-82) at room temperature for 10 minutes, in the dark. Cells were washed with PBS by centrifugation at 600 x g for 3 minutes and resuspended in 0.6 ml of FACS staining solution. Samples were filtered through a 40 µm filter before analysis of the expression of cell surface makers using a Becton Dickinson FACS Aria machine to sort the various B cell types. Sort templates for pre-B cells, spleen IgM cells and spleen IgG cells are shown in Figures 2.1, 2.2 and 2.3, respectively. Bone marrow IgM cells were

sorted using the pre-B cell template. Following sorting, cells were recovered by centrifuging at $500 \times g$ for 10 minutes at 4°C and were either processed immediately to make genomic DNA and total RNA, or stored at -80°C until use.

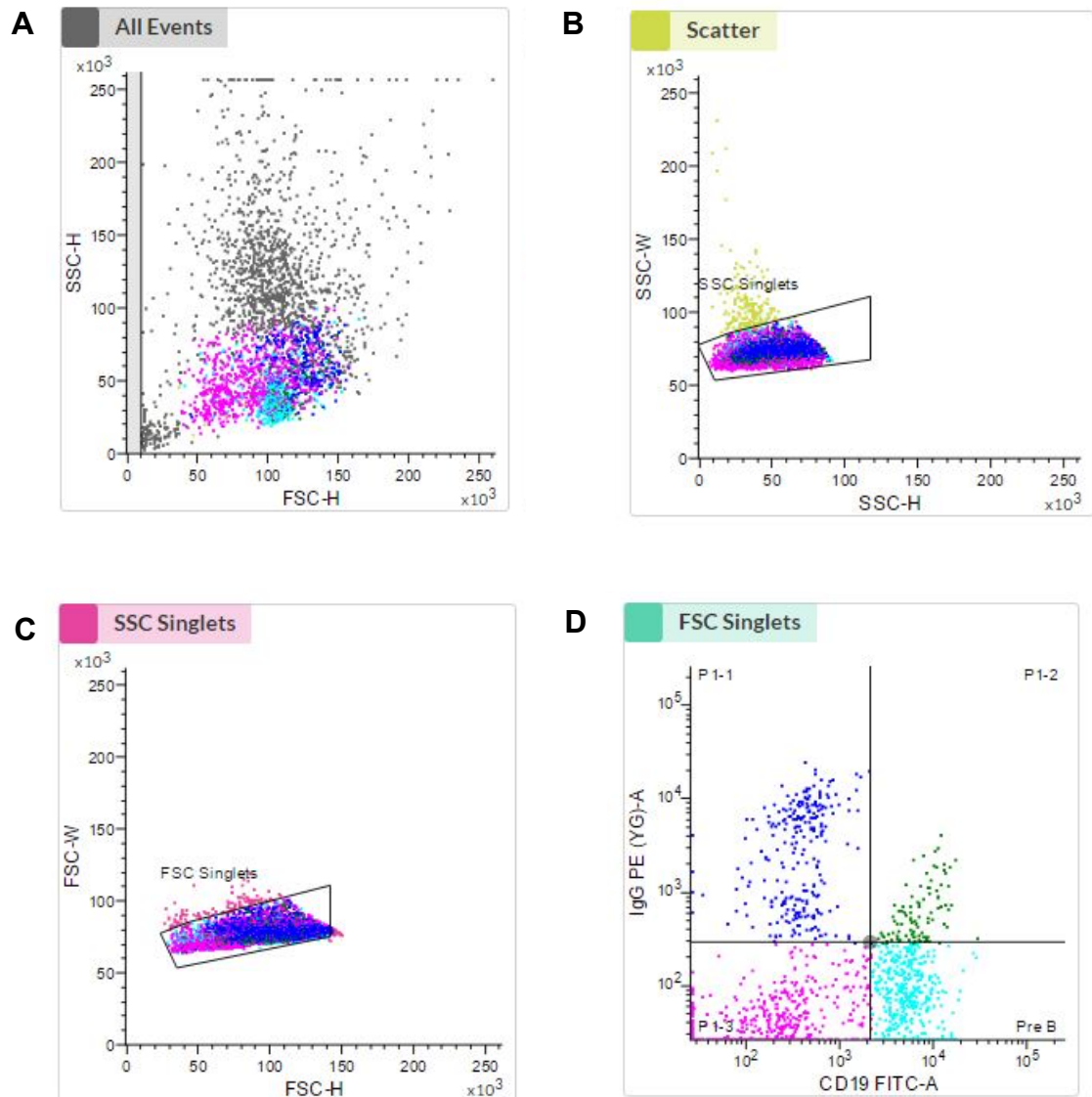


Figure 2.1 Sort template for primary pre-B cells

Bone marrow cells were flushed from the femurs of 5 to 7 week old mice and lymphocytes were isolated. (A) Lymphocytes are first gated on forward scatter (FSC-A) and side scatter (SSC-A).

(B) Single lymphocytes are gated on side scatter (SSC-H and SSC-W).

(C) Single lymphocytes are gated on forward scatter (FSC-H and FSC-W).

(D) Pre-B cells ($\text{CD}19^+/\text{CD}43^-$) are gated based on their staining with anti-CD19FITC (CD19 FITC-A) and anti-CD43 (CD43 PE(YG)-A).

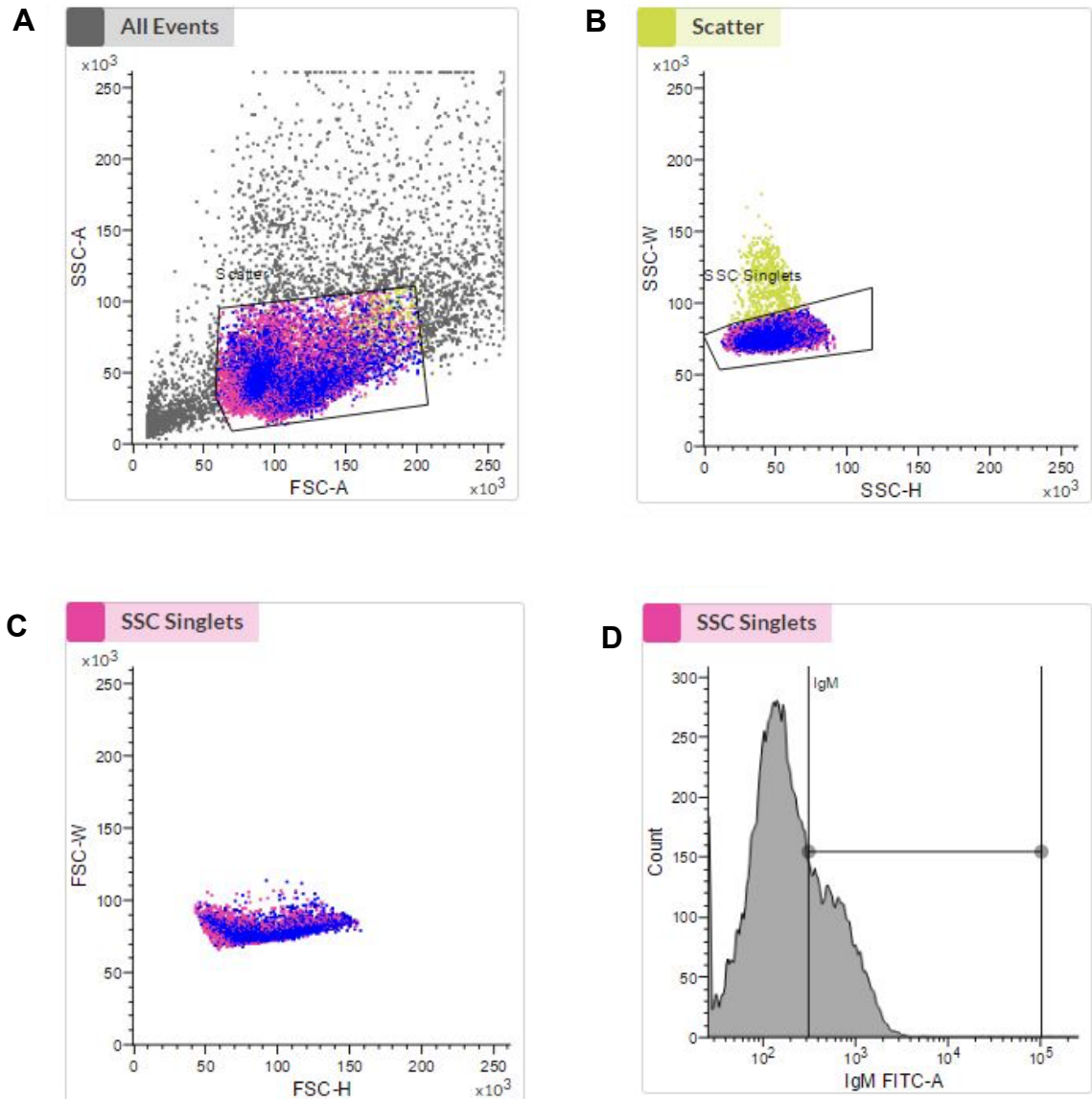


Figure 2.2 Sort template for primary spleen IgM+ cells

Cells were collected from the spleen of 5 to 7 week old mice and lymphocytes were isolated.

(A) Lymphocytes are first gated on forward scatter (FSC-A) and side scatter (SSC-A).

(B) Single lymphocytes are gated on side scatter (SSC-H and SSC-W).

(C) Single lymphocytes are gated on forward scatter (FSC-H and FSC-W).

(D) Spleen IgM+ cells are gated based on their staining with anti-IgM (IgM FITC-A).

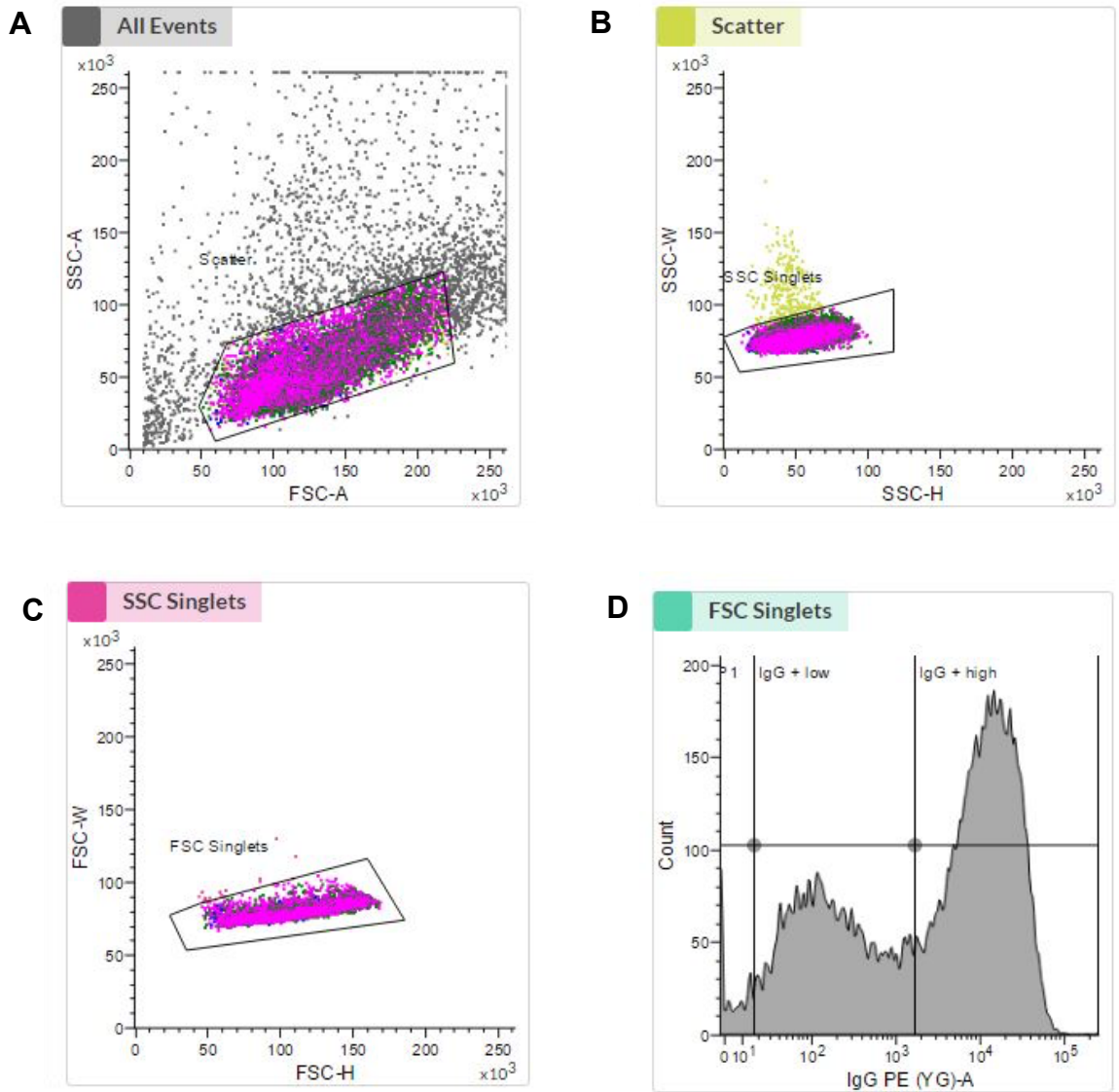


Figure 2.3 Sort template for primary spleen IgG+ high cells

Cells were collected from the spleen of 5 to 7 week old mice and lymphocytes were isolated.

(A) Lymphocytes are first gated on forward scatter (FSC-A) and side scatter (SSC-A).

(B) Single lymphocytes are gated on side scatter (SSC-H and SSC-W).

(C) Single lymphocytes are gated on forward scatter (FSC-H and FSC-W).

(D) Spleen IgG+ cells are gated based on their staining with anti-IgG (IgG PE (YG)-A).

2.46 Genomic DNA extraction

1 x 10⁶ cells were recovered by centrifugation at 600 x g for 3 minutes and the cell pellet washed twice with ice cold PBS, followed by resuspension in 500 µl of genomic DNA digestion buffer (200 mM NaCl, 10 mM Tris-HCl [pH 7.5], 2 mM EDTA, 0.2% SDS) and mixed thoroughly. To sample, 0.4 mg/ml proteinase K was added and the reaction incubated at 56 °C with rotation overnight. The next day, an equal volume of isopropanol and 1 µl of glycogen were added and mixed by inversion, followed by centrifugation at 14,000 x g for 5 minutes at room temperature. The DNA pellet was then washed with 1 ml of 70% ethanol and resuspended in 50 µl of TE. The concentration was determined using NanoDrop spectrophotometer (A²⁶⁰).

2.47 Total RNA extraction and reverse transcription

2.5 x 10⁶ cells were recovered by centrifugation at 600 x g for 3 minutes. The supernatant was removed and 500 µl of TRIzol (Sigma #T3934) was added into the pellets and mixed thoroughly by pitting up and down several times and vortexing, followed by incubation at RT for 5 minutes. 100 µl of chloroform was added and mixed by shaking the tubes for 15 seconds, followed by incubation at RT for 2-3 minutes. After centrifuging at 14,000 x g for 15 minutes, the aqueous phase was transferred to a fresh tube and 250 µl of isopropanol was added to precipitate the RNA, followed by incubation at RT for 10 minutes and centrifugation at 14,000 x g for 10 minutes at 4 °C. The total RNA pellet was washed with 1 ml of 80% ethanol and redissolved in 20 µl of sterile water. Following this, 2 µg total RNA was treated with 2 units of DNase I (RNase-free) (NEB #M0303S) in a 100 µl reaction containing 1 x DNase I reaction buffer (2.5 mM MgCl₂, 10 mM Tris-HCl, 0.5 mM CaCl₂ pH 7.6) for 1-2 hours at 37 °C. The reaction was then stopped by phenol/chloroform extraction and ethanol precipitation and the RNA pellet was resuspended in 10 µl of dH₂O. Subsequently, 5 µl of the total RNA was used in a reverse transcription reaction containing 1 µl 10 mM dNTP mix and 1 µl of Oligo (dT)₁₅. After heating at 65 °C for 5 minutes, the reaction was incubated on ice for at least 1 minute. The reaction was then collected by brief centrifugation and 1 x first-

strand buffer (50 mM Tris-HCl, pH 8.3, 75 mM KCl, 3 mM MgCl₂), 10 mM DTT and 20 units of RNase Inhibitor (Applied Biosystems™ #N8080119) were added, followed by incubation at 37 °C for 2 minutes. Subsequently, 200 units of M-MLV reverse transcriptase (Invitrogen™ #28025013) was added and the reaction was then incubated at 37 °C for 50 minutes, followed by heat inactivation at 70 °C for 15 minutes, phenol/chloroform extraction and ethanol precipitation. The resulting cDNA pellet was resuspended in 40 µl of dH₂O and diluted 1:10. 2 µl of the diluted cDNA was used in qPCR analyses.

2.48 Nested Polymerase Chain Reaction (Nested PCR)

Nested PCR, which is a modification of traditional PCR, was designed to improve specificity and sensitivity. Nested PCR contains two primer sets and two successive PCR assays. The first set of primers anneal to sequences upstream from the second set of primers and are used in first round PCR. Amplicons generated from the first round PCR are used as a template for the second amplification step with a second set of primers. To check how long the ESC persists, nested PCR assays were used to check and quantify the copies of ESC and recombination for the following genes: V λ 1J λ 1, V κ 3J κ 5, V κ 11J κ 5 and V κ 16J κ 5. To quantify the V λ 1J λ 1 ESC, a single round of qPCR was set up as described in Section 2.39. This contained 20 ng genomic DNA and 4 pmol of each primer (Vend_2nd and Jend_3). To detect V λ 1J λ 1 recombination, the V κ 3J κ 5, V κ 11J κ 5 and V κ 16J κ 5 ESCs and their respective levels of recombination, first round PCR reactions contained 1x ThermoPol reaction buffer (20 mM Tris-HCl, 10 mM (NH₄)₂SO₄, 10 mM KCl, 2 mM MgSO₄, 0.1% Triton ® X-100, pH 8.8), 200 nM of each dNTP, 0.2 µM each primer (Table 2.4 and Table 2.5), 20 ng of genomic DNA template and 1 unit of Taq DNA polymerase (NEB #M0237) in a 25 µl reaction volume. Thermocycling conditions were as follows: pre-denature at 94 °C for 3 min, followed by 18 cycles of 94 °C for 20 s, T_m for 20 s and 72 °C for 20 s. A 7 min of final extension was performed. 1 µl of 1 in 10 diluted first round PCR product was used as the template for a second round of qPCR (Section 2.39) using the primers shown in Table 2.4 and 2.5. The absolute amount of ESC and recombination was then quantified using a standard curve.

The standard curve was generated with a serial of concentration-known DNA samples. Specifically, the first round PCR product (generated by 35 PCR cycles) was purified as described as Section 2.11. The concentration of the PCR product was determined using the QuantiFluor™ dsDNA system (Promega #E2670) according to manufacturer's protocol (Figure 2.4). Briefly, QuantiFluor™ dsDNA Dye was diluted 1:400 in 1 x TE and mixed well by vortexing to make the QuantiFluor™ dsDNA Dye working solution. The standard DNA (λ DNA standard, 100 ng/ μ l) was serially diluted to 1 ng/ μ l and 0.1 ng/ μ l, respectively. Following this, 200 μ l of QuantiFluor™ dsDNA Dye working solution was pipetted into each well of a 96-well plate. Diluted standard DNA ranging from 0.125 ng to 10 ng (0.125 ng, 0.25 ng, 0.5 ng, 1 ng, 2 ng, 4 ng, 6 ng, 8 ng and 10 ng) was added into each duplicate wells to make a standard curve to measure the concentration of the unknown dsDNA. 1-10 μ l of purified first round ESC or recombination PCR product was added into 200 μ l of QuantiFluor™ dsDNA Dye working solution. The prepared samples were mixed thoroughly by pipetting up and down and incubated at room temperature for 5 minutes, protected from light. The fluorescence (504 nm_{Ex}/531 nm_{Em}) of each well was then measured using a plate reader. The dsDNA concentration was calculated as follows: subtract the fluorescence of the blank sample (1 x TE buffer) from all of the standard and unknown samples. Use the corrected data from the DNA standards to generate a standard curve of fluorescence versus DNA concentration. The DNA concentration of sample was determined from the standard curve and the DNA number of copies was calculated as described as Section 2.40.

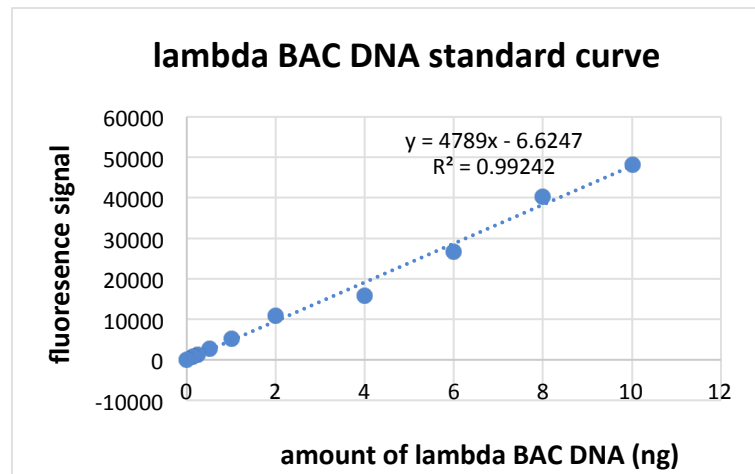


Figure 2.4 Lambda BAC DNA standard curve for measuring the dsDNA concentration

The fluorescence signal of different amounts of a lambda BAC DNA standard ranging from 0-10 ng as determined with QuantiFluor™ dsDNA Dye. The results were then plotted to generate a lambda BAC DNA standard curve. The Y value indicates the fluorescence signal and the X value indicates the dsDNA concentration. After measuring the fluorescence signal of concentration-unknown samples, the y value was put into the formula generated from the lambda BAC DNA standard curve to calculate the concentration of dsDNA.

G) Generation of Artemis-deficient REH cell line

2.49 Knock-down of Artemis in REH cells by shRNA

A common method for the generation of stable knockdown of gene expression in cells is small hairpin RNA (shRNA) by to inhibit mRNA transcription (Benson et al., 2006; Kanungo, 2017; Silva et al., 2004; Wiederschain et al., 2009). ShRNA targeted against Artemis were designed with the online software (<http://sirna.wi.mit.edu>) and synthesized by Integrated DNA Technologies (IDT; Table 2.6). The products were then annealed in a reaction containing 0.45 μ M of each oligo, 2.5 μ l of 10 x annealing buffer (1 M NaCl, 100 mM Tris-HCl, pH 7.4) in a 25 μ l reaction volume, followed by boiling for 5 min and cooling to room temperature. 0.1125 μ M annealed oligos were then ligated to 20 ng of gel-purified digested Tet-pLKO-puro (Addgene Plasmid #21915) or pLKO-EGFP (derived from Tet-pLKO-puro) in a 10 μ l reaction containing 1 μ l of 10 x ligase buffer and 1 μ l of T4 DNA ligase. The resulting pLKO-Artemis-puro or pLKO-EGFP-Artemis clones were identified by enzyme digestion and Sanger

sequencing. 4.9 µg of the correct pLKO-Artemis-puro or pLKO-EGFP-Artemis was co-transfected into HEK293T cells plated in 10-cm dishes (two 10-cm dishes were needed per viral construct) together with 2.6 µg of pR8.74 (Addgene Plasmid #22036) and 2.5 µg of pMD2.G (Addgene Plasmid #12259) to generate lentiviruses for delivering shRNA and knocking down of Artemis expression. 48 hours post-transduction, viral media was collected in 15 ml falcons and centrifuged for 10 min at 600 x g to pellet cell debris. Next, the viral media was filtered by syringe through a 0.45 µM filter. 1×10^6 REH cells were infected with 500 µl of unconcentrated Artemis-puro or Artemis-GFP lentiviruses together with 8 µg/ml of polybrene to improve transduction efficiency by spinfection (Section 2.35) at 800 x g for 90 min at 32 °C. Two days post-transduction, 1 µg/ml of puromycin was added to the REH cells and maintained selection for 10 days. The Tet-pLKO-Puro system is an inducible lentiviral shRNA vector that relies on tetracycline (Tet) to regulate shRNA expression (Figure 2.5; Wiederschain et al., 2009). Therefore, after selection, 100 ng/ml doxycycline (a derivative of tetracycline) was added to the cells to induce the expression of shRNA to knock down Artemis. 1×10^6 induced cells were then collected at 48, 72 and 96 hours post-induction, respectively. Total RNA was extracted and reverse-transcribed. The resulting cDNA was used to check the transcription level of Artemis by qPCR. cDNA prepared from normal REH cells was used as a control.

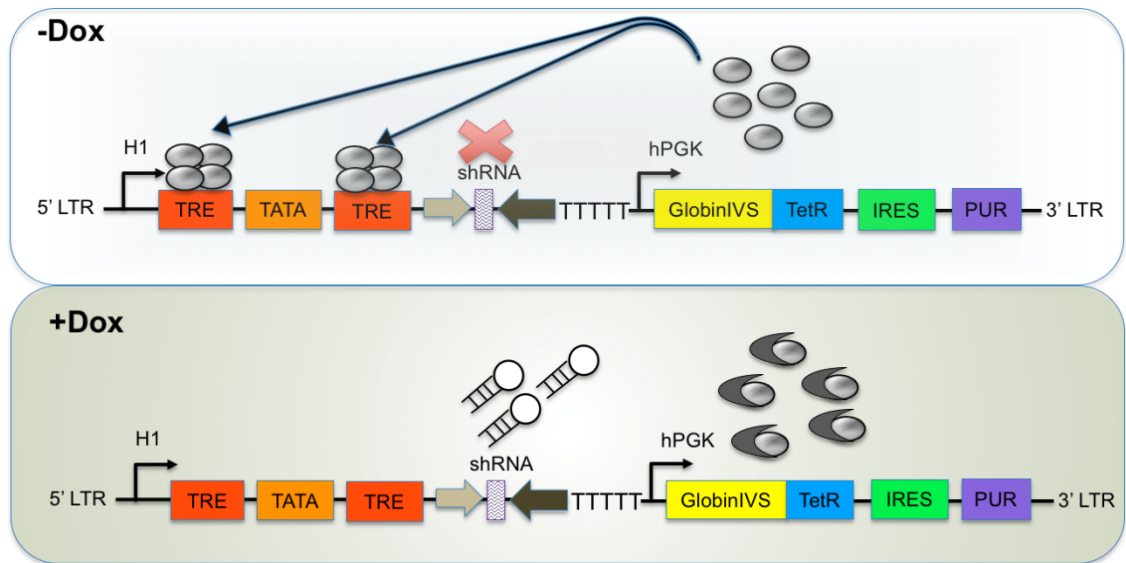


Figure 2.5 Schematic of Tet-pLKO-puro and its mechanism of action

The Tet-pLKO-puro vector includes the necessary elements for packaging, reverse transcription and integration, which are required for subsequent production of lentiviral particles. More importantly, this vector also encompasses the necessary components for the inducible expression of shRNA in the target cells. In the absence of tetracycline/doxycycline, the expression of shRNA is repressed by TetR proteins. Upon addition of appropriate amount of tetracycline/doxycycline, the expression of shRNA is triggered, allowing for the knockdown of gene of interest.

2.50 Generation of Artemis-deficient REH cells using CRISPR-Cas 9 targeting

Another universally-used method for generation of specific loss-of-function cell lines is the Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) and CRISPR-associated protein (Cas) system (Rojo et al., 2018). Therefore, CRISPR-Cas 9 system (Figure 2.6) was also used to generate an Artemis-deficient cell line. A single guide RNA (sgRNA) against the transcription start site of Artemis (Table 2.6) was designed with the online software (<http://crispr.mit.edu>) and the relevant forward and reverse DNA oligonucleotides were synthesized by Integrated DNA Technologies (IDT). The products were then phosphorylated and annealed in a reaction containing 100 pmol of each oligo, 1 μ l of 10 x T4 DNA ligase buffer and 0.5 μ l of T4 PNK in a 10 μ l reaction volume at 37 $^{\circ}$ C for 30 min, followed by boiling for 5 min and cooling to room temperature. 2 μ l of the 1:250 diluted annealed oligos were

then digested with BsmBI and ligated to 100 ng of lentiCRISPR v2 (Addgene Plasmid #52961) in a 20 μ l reaction containing 2 μ l of 10 x Buffer 3.1 (1 M NaCl, 500 mM Tris-HCl [pH 7.9], 100 mM MgCl₂, 1 mg/ml BSA), 1 μ l of 10 mM ATP, 1 μ l of BsmBI and 0.5 μ l of T7 DNA ligase. The reactions were performed via thermocycling using the conditions: 6 cycles of 37 °C for 5 min, 23 °C for 5 min and 1 cycle of 50 °C for 30 min. 10 μ l of the product was transformed into *Endura* competent cells to generate single colony for mini-prep (Section 2.12) and verified by Sanger sequencing. 4.9 μ g of resulting lentiCRISPR-Artemis was co-transfected into 293T cells together with 2.6 μ g of pR8.74 (Addgene Plasmid #22036) and 2.5 μ g of pMD2.G (Addgene Plasmid #12259) to generate Artemis KO lentivirus. 1x10⁶ REH cells were infected with 500 μ l of unconcentrated Artemis KO lentivirus together with 8 μ g/ml of polybrene by spinfection (Section 2.35) at 800 x *g* for 90 min at 32 °C. Two days after spinfection, 5 μ g/ml of puromycin was added to the cells and maintained selection for 10 days. After selection, 5 x 10⁴ cells were harvested and used for extraction of total proteins.

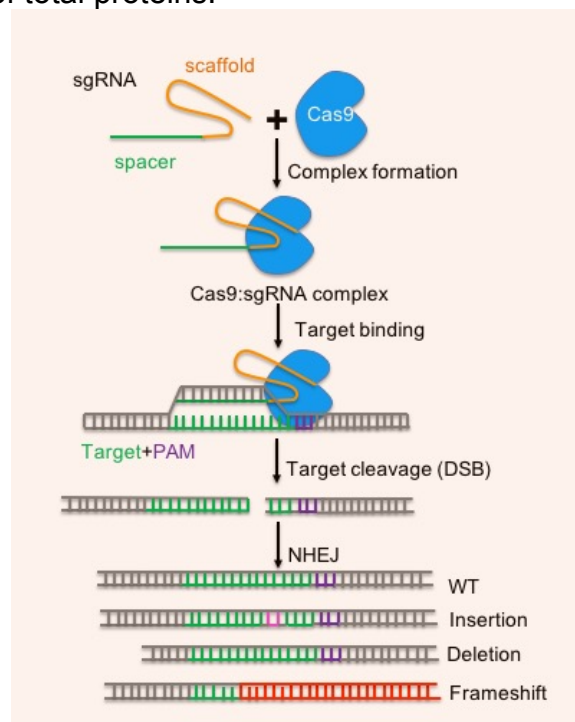


Figure 2.6 Schematic of the CRISPR-Cas9 system

Cas9 protein interacts with sgRNA, forming a Cas9:sgRNA complex. The complex then binds to target genomic DNA and causes DNA cleavage at ~3-4 nucleotides upstream of the Protospacer Adjacent Motif (PAM) sequence, leading to a blunt double-strand break (DSB). This DSB is then mostly repaired by error-prone NHEJ pathway, leading to different mutations, for instance, insertions, deletions or frameshift mutations and thereby generating a loss-of-function mutation in the target gene of interest.

2.51 Generation of Artemis-deficient REH cell lines using the KRAB repressor

Due to the instability of CRISPR-Cas9 generated Artemis-deficient REH cell lines, Kruppel-associated box (KRAB) effector was employed to fuse with Cas9 protein, thereby repressing Artemis promoter expression by targeting to relevant sgRNA (Beerli et al., 2000; Cong et al., 2012; Gilbert et al., 2013). Two sgRNAs against the Artemis promoter (Apro) (Table 2.6) were designed and processed as above (Section 2.50), which target to Cas9-KARB fusion protein. These were then ligated into the vector lenti sgRNA(MS2)_zeo (Addgene Plasmid #61427). The Apro-lentivirus was generated as described in Section 2.33. 1×10^6 REH cells were infected with 1 ml of unconcentrated Apro lentivirus and/or 1 ml of lentivirus expressing the KRAB repressor (generated by Daniel Thwaites) together with 8 $\mu\text{g}/\text{ml}$ of polybrene by spinfection (Section 2.36) at $800 \times g$ for 90 min at $32 \text{ }^\circ\text{C}$. Two days after spinfection, 5 $\mu\text{g}/\text{ml}$ of blasticidin and 100 $\mu\text{g}/\text{ml}$ of zeocin was added to select cells. After selection for 10 days, 5×10^4 cells were harvested and used for extraction of total proteins.

2.52 Detection of loss of Artemis expression

2.52.1 Determination of Artemis expression by qPCR

2.5×10^6 Artemis-deficient REH cells generated by the shRNA system and normal REH cells (for control) were harvested by centrifuging at $600 \times g$ for 3 minutes at $4 \text{ }^\circ\text{C}$. The total RNA was extracted as described (Section 2.47). Artemis cDNA was amplified by qPCR (Section 2.39) using the primers Artemis_262_FWD and Artemis_396_RVS (Table 2.6), with an annealing of $60 \text{ }^\circ\text{C}$. The relative Artemis transcription level was then determined by normalizing to a human housekeeping gene, HPRT, with the primers HPRT_FWD_human and HPRT_RVS_human (Table 2.6).

2.52.2 Detection of Artemis mutations by PCR and sequencing

1 x10⁶ Artemis-deficient REH generated by the CRISPR-Cas9 system and normal REH cells (for control) were harvested by centrifuging at 600 x g for 3 minutes at 4 °C. The genomic DNA was extracted as described (Section 2.46). The relevant fragment of the Artemis gene was amplified by PCR using primers Artemis_F and Artemis_R (Table 2.6), with an annealing temperature of 63 °C. The PCR product was purified by electroelution (Section 2.11) and sequenced using 2.5 μM of the primer Artemis_F. The sequencing result was analyzed by MacVector. The Artemis deficient cells were then analysed for protein expression by western blotting (Section 2.52.3).

2.52.3 Determination of Artemis deficiency by western blotting

5 x 10⁴ Artemis-deficient REH generated by the CRISPR-Cas9 system and normal REH cells (for control) were harvested by centrifugation at 600 x g for 3 minutes at 4 °C, and the cell pellet was resuspended in 50 μl of 1:3 mix lysis buffer (30% glycerol, 0.15 M Tris [pH 6.7], 5% SDS) and RIPA (50 mM NaCl, 0.5% NaDOC, 25 mM Tris-HCl [pH 8.2], 0.5% NP-40, 0.1% SDS) supplemented with 1 μl of 50 x protease inhibitor cocktail. The samples were then boiled for 5 minutes, followed by centrifugation at 14,000 x g for 10 minutes at 4 °C to remove cell debris. 15 μl of supernatant was added into 5 μl of 4 x protein loading dye (100 mM Tris-HCl [pH 6.8], 25 mM EDTA, 20% glycerol, 4% SDS, 2% β-mercaptoethanol, 0.04% Bromophenol blue). The samples were mixed well and boiled again for 2 minutes before SDS-PAGE (Section 2.20). Artemis proteins were identified using a primary anti-Artemis antibody (St John's Laboratory #STJ91710) diluted 1:2500 and β-tubulin control was visualized using anti-β-tubulin (ORIGENE #TA347064) diluted 1:5000 in 5% (w/v) non-fat dried milk (Marvel) in TBS-T. The secondary antibodies were diluted 1:20,000 in 5% (w/v) non-fat dried milk (Marvel) in TBS-T and were horse radish peroxidase (HRP)-conjugated polyclonal anti-rabbit antibody (for Artemis) and HRP-anti-mouse antibody (for β-tubulin). The relative expression level of Artemis was determined by normalizing the intensity of the Artemis signal to the β-tubulin signal.

H) Oligonucleotides

The oligonucleotides and plasmids used during this project are shown in Tables 2.1-2.8. The primers were synthesised either by Sigma-Aldrich (Poole, UK) or Integrated DNA Technologies (IDT) and purified by desalting. The concentration of the primers were adjusted to 100 pmol/μl in water before use.

Table 2.1 Oligonucleotides used for the RAG cutting assays *in vitro*

Oligonucleotide	Sequence (5'-3')	Reference
12-RSS_TOP	GATCTGGCCTGTCTTACACAGTGCTACAGACTGGAACAAA AACCCCTGCAG	McBlane et al. (1995)
12-RSS_BOTTOM	CTGCAGGGTTTTTGTTCAGTCTGTAGCACTGTGTAAGAC AGGCCAGATC	McBlane et al. (1995)
23-RSS_TOP	GATCTGGCCTGTCTTACACAGTGGTAGTACTCCACTGTCT GGCTGTACAAAAACCCTGCAG	McBlane et al. (1995)
23-RSS_BOTTOM	CTGCAGGGTTTTTGTACAGCCAGACAGTGGAGTACTACC ACTGTGTAAGACAGGCCAGATC	McBlane et al. (1995)
ESC_TOP	CTGCAGGGTTTTTGTTCAGTCTGTAGCACTGTGCACAGT GGTAGTACTCCACTGTCTGGCTGTACAAAAACCCTGCAG	Neiditch et al. (2002)
ESC_BOTTOM	CTGCAGGGTTTTTGTACAGCCAGACAGTGGAGTACTACAC TGTGCACAGTGCTACAGACTGGAACAAAAACCCTGCAG	Neiditch et al. (2002)
ESC12d9_TOP	CTGCAGTCCAGTCTGTAGCACTGTGCACAGT GGTAGTAC TCCACTGTCTGGCTGTACAAAAACCCTGCAG	-
ESC12d9_BOTTOM	CTGCAGGGTTTTTGTACAGCCAGACAGTGGAGTACTACC ACTGTGCACAGTGCTACAGACTGGACTGCAG	-
ESC12dCAC_TOP	CTGCAGGGTTTTTGTTCAGTCTGTAGCACTCACAGTGGT AGTACTCCACTGTCTGGCTGTACAAAAACCCTGCAG	-
ESC12dCAC_BOTTOM	CTGCAGGGTTTTTGTACAGCCAGACAGTGGAGTACTACC ACTGTGAGTGCTACAGACTGGAACAAAAACCCTGCAG	-
ESC23dCAC_TOP	CTGCAGGGTTTTTGTTCAGTCTGTAGCACTGTGAGTGGT AGTACTCCACTGTCTGGCTGTACAAAAACCCTGCAG	-
ESC23dCAC_BOTTOM	CTGCAGGGTTTTTGTACAGCCAGACAGTGGAGTACTACC ACTCACAGTGCTACAGACTGGAACAAAAACCCTGCAG	-
LMO2_RSS_TOP	GATCTGGCCTGTCTTACACAGTAGTAGTAGGCTGTGCAA TAATTCTGCAG	-
LMO2_RSS_BOTTOM	CTAGACCGGACAGAATGTGTCATCATCATCCGACACGTT ATTAAGACGTC	-
Ttg1_RSS_TOP	GATCTGGCCTGTCTTACACAGTGGTAGTACTCCACTGTCT GGCTGTACTCTGGCACTGCAG	-
Ttg1_RSS_BOTTOM	CTAGACCGGACAGAATGTGTCACCATCATGAGGTGACAG ACCGACATGAGACCGTGACGTC	-

SIL_RSS_TOP	GATCTGGCCTGTCTTACACTCTGGTAGTACTCCACTGTCT GGCTGTTGTTGAAGCTGCAG	-
SIL_RSS_BOTTOM	CTAGACCGGACAGAATGTGAGACCATCATGAGGTGACA GACCGACAACAACCTCGACGTC	-
ESC_6bp spacer_TOP	CTGCAGGGTTTTTGTCCAGTCTGTAGCACTGTGAGGTAG CACAGTGGTAGTACTCCACTGTCTGGCTGTACAAAAACCC TGCAG	
ESC_6bp spacer_BOTTOM	CTGCAGGGTTTTTGTACAGCCAGACAGTGGAGTACTACCA CTGTGCTACCTCACAGTGCTACAGACTGGAACAAAAACCC TGCAG	
ESC_9bp spacer_TOP	CTGCAGGGTTTTTGTCCAGTCTGTAGCACTGTGAGGTAG CTACACAGTGGTAGTACTCCACTGTCTGGCTGTACAAAA CCCTGCAG	
ESC_9bp spacer_BOTTOM	CTGCAGGGTTTTTGTACAGCCAGACAGTGGAGTACTACCA CTGTGTAGCTACCTCACAGTGCTACAGACTGGAACAAAA CCCTGCAG	
ESC_12bp spacer_TOP	CTGCAGGGTTTTTGTCCAGTCTGTAGCACTGTGTACAGG TAGCTACACAGTGGTAGTACTCCACTGTCTGGCTGTACAA AAACCCTGCAG	
ESC_15bp spacer_BOTTOM	CTGCAGGGTTTTTGTACAGCCAGACAGTGGAGTACTACCA CTGTGTAGCTACCTGTACACAGTGCTACAGACTGGAACAA AAACCCTGCAG	
ESC_12bp spacer_TOP	CTGCAGGGTTTTTGTCCAGTCTGTAGCACTGTGTACGTC CAGGTAGCTACACAGTGGTAGTACTCCACTGTCTGGCTGT ACAAAAACCCTGCAG	
ESC_15bp spacer_BOTTOM	CTGCAGGGTTTTTGTACAGCCAGACAGTGGAGTACTACCA CTGTGTAGCTACCTGACGTACACAGTGCTACAGACTGGAA CAAAAACCCTGCAG	
J1_RSS_TOP	GATCTGGCCTGTCTTACACTGTGCTACAGACTGGAGCAA AAAACCTGGAG	-
J1_RSS_BOTTOM	CTGCAGGTTTTTGTCTCCAGTCTGTAGCACAGTGTAAGA CAGGCCAGATC	-
J2_RSS_TOP	GATCTGGCCTGTCTTACACAATGCTACAGACTGGACCC AAAACCCTGGAG	-
J2_RSS_BOTTOM	CTGCAGGGTTTTGGGTCCAGTCTGTAGCATTGTGTAAG ACAGGCCAGATC	-
J3_RSS_TOP	GATCTGGCCTGTCTTACACAGTGCTACAGACTGGACCC TAAACCCTGGAG	-
J3_RSS_BOTTOM	CTGCAGGGTTTAGGGTCCAGTCTGTAGCACTGTGTAAG ACAGGCCAGATC	-

Table 2.2 Primers used to detect cutting and normalization

Plasmid	Primers used to detect cutting	Primers used for normalization
J β 2.7+	re-int 1 and re-int 2	Neo_3_fwd and Neo_3_rev
J δ 1+	ESC_cut_fwd and ESC_cut_rev	Jd_norm_fwd and Jd_norm_rev
12-RSS	RSS_cut_fwd and RSS_cut_rev	Neo_3 fwd and Neo_3 rev
23-RSS	23-RSS_fwd and 23-RSS_rev	Neo_3 fwd and Neo_3 rev
ESC	ESC_cut_fwd and ESC_cut_rev	Jd_norm_fwd and Jd_norm_rev
ESC12d7	ESC_cut_fwd and ESC_cut_rev	Jd_norm_fwd and Jd_norm_rev
ESC12d9	ESC_cut_fwd and ESC_cut_rev	Jd_norm_fwd and Jd_norm_rev
ESC23d7	ESC_cut_fwd and ESC_cut_rev	Jd_norm_fwd and Jd_norm_rev
ESC23d9	ESC_cut_fwd and ESC_cut_rev	Jd_norm_fwd and Jd_norm_rev
pWPI-12RSS	12RSS_cutting_fwd_new and 12RSS_cutting_new_rvs	12RSS_virus_norm_fwd and virus_norm_rvs
pWPI-23RSS	23RSS_cutting_fwd and 23RSS_cutting_rvs	23RSS_virus_norm_fwd and 23RSS_virus_norm_rvs
pWPI-ESC	ESC_cut_fwd and ESC_cut_rev	ESC_virus_norm_fwd and virus_norm_rvs
pWPI-ESC12d7	ESC_cut_fwd and ESC_cut_rev	ESC_virus_norm_fwd and 23RSS_virus_norm_fwd
pWPI-ESC12d9	ESC_cut_fwd and ESC_cut_rev	ESC_virus_norm_fwd and virus_norm_rvs
pWPI-ESC23d7	ESC_cut_fwd and ESC_cut_rev	ESC_virus_norm_fwd and virus_norm_rvs
pWPI-ESC23d9	ESC_cut_fwd and ESC_cut_rev	ESC_virus_norm_fwd and 23RSS_virus_norm_fwd

Table 2.3 Primers used for *in vivo* RAG cutting assays

Oligonucleotide	Sequence (5'-3')	Tm(°C)
re-int1	GAGCTCGGGGAGCCTTAG	60
re-int 2	AATCTCACCTGTGACCGTGAG	60
RSS_cut_fwd	CTGGCTAGCGTTTAACTTAAGC	60
RSS_cut_rev	TCTGCAGAATTCCACCACAC	60
23-RSS_cut_fwd	AAACATGTCCGGTCTGTCACC	60
23-RSS_cut_rev	TCTGCAGAATTCCACCACAC	60
ESC_cut_fwd	CTGGCTAGCGTTTAACTTAAGC	60
ESC_cut_rev	TGTTTTTCCAAAGCGGTCGAG	60

Neo_3_rev	TTTCGCTTGGTGGTCAATG	60
Neo_3_fwd	TGCTCCTGCCGAGAAAAGTATC	60
Jd_norm_fwd	GTCGCCTTCTTGACATTACTC	60
Jd_norm_rev	AACAACAGATGGCTGGCAAC	60
12RSS_cutting_fwd_new	CCTGTGCTGGGAGACCTG	60
12RSS_cutting_new_rvs	AGACGCCCGAATCTCACC	60
23RSS_cutting_fwd	CTGGCTAGCGTTTAAACTTAAGC	60
23RSS_cutting_rvs	CTATGATCGATTCACTGACTGTAGA	60
12RSS_virus_norm_fwd	GCAGATATCCAGCACAGTGG	60
Virus_norm_rvs	CCCGTAGTTTTAGAAAGGCACAG	60
23RSS_virus_norm_fwd	GGTACAGTGCAGGGGAAAGAATAG	60
23RSS_virus_norm_rvs	GTGAATCGATCATAGTCTAGAGGG	60
ESC_virus_norm_fwd	CGTCGCCTTCTTGACATTAC	60
genomicDNA_qPCR (200 bp)_fwd	ACCATGTCCTTCCCGATCTG	58
genomicDNA_qPCR (200 bp)_rvs	TGGGTCCACTCTGAAGGAAC	58
genomicDNA_PCR (680 bp)_fwd	TCATTCCCAGTGTTGATGCC	55
genomicDNA_PCR (680 bp)_rvs	TGGGTCCACTCTGAAGGAAC	55

Table 2.4 Primers used in LM-PCR and nested PCR assays

Primers	Sequence (5'-3')	Tm(°C)
FM25	GCGGTGACTCGGGAGATCTGAAGTG	
FM11	CACTTCAGATC	
RSS_fwd_1 st	CCCACTGCTTACTGGCTTATCG	60
RSS_fwd_2 nd	CGACTCACTATAGGGAGACCCAAG	56
23RSS_rvs_1 st	GGCAAACAACAGATGGCTGG	65
23RSS_rvs_2 nd	TAGAAGGCACAGTCGAGGCTG	63
Jk5_end_1st	CCTGTGCATCAATAGAAGATCC	53
Jk5_end_2nd	CTTAGGCTTCTGAGACCAGT	60
Jk5_cdjt_1st	GAAGTACTTTAACTCCTAACATG	55
Jk5_cdnd_1st	GTAATTACGTTTCAGCTCCAGC	60
Vk3_end_1st	GTAATTTTACGAATAAACCTATAGC	53

Vk3_end_2nd	GAGCCTCAGTGAGCTGCAAC	60
Vk3_cdnd_1st	GAATATTATGGCACAAGTTAATGCAG	55
Vk3_cdnd_2nd	CCTCATCTATGCTGCATCCAACG	60
Vk16-104_end_1st	CATTCTCCATGGGTTTCATGCT	53
Vk16-104_end_2nd	GCCCAACATCTGGTTTCTGC	60
Vk16-104_cdjt_1st	GCCTGGTATCAAGAGAAACC	55
Vk16-104_cdnd_1st	GGATCCACTTTGCAATCTGG	60
Vk11-125_end_1st	CTTTCATCCAATACTGTGGCAC	53
Vk11-125_end_2nd	GAAGGAGCTGATGGACCAGC	60
Vk11-125_cdjt_1st	CAGGACTCAGAATGAACACG	55
Vk11-125_cdjt_2nd	CTATGGTGCAAGCAACTTGAAG	60
V1Sacl_F	GCCAACTGGGTCCAAGAAAAAC	55
J1Sty_R	GCACCTCAAGTCTTGGAGAG	55
V1RSS_F	GGCACAGACTGAGGATGAG	55
J1RSS_R	CAGTCAGTTTGGTTCCTCCA	55
Vend_2nd	CCTATGAGGACATATGGATCCTGGGAAGAAGG	67
Jend_3	CCTATGAGGACATATGGATCCTGGGAAGAAGG	67
RAG1_F	ACAGCCGGAGATACCCAGTCCACG	65
RAG1_R	GGTGGATGGAGTCAACATCTGCCT	65
GAPDH_F	ACTTCTTGTGCAGTGCCAGC	56
GAPDH_R2	GCACACTTCGCACCAGCATC	56
IntgenIII_F	CAAGGAAAGGCCAACCAATA	53
IntgenIII_R	TAACCCTTCCCCAGCTCTT	53

Table 2.5 Primers used to detect ESC and recombination

product	First round PCR		Second round PCR	
	primers	Tm (°C)	primers	Tm (°C)
Vλ1Jλ1 ESC	Vend_2nd Jend_3	67		
Vλ1Jλ1 recombination	V1Sacl_F J1Sty_R	55	V1RSS_F J1RSS_R	60
Vκ3Jκ5 ESC	Vk3_end_1st Jk5_end_1st	53	Vk3_end_2nd Jk5_end_2nd	60
Vκ3Jκ5 recombination	Vk3_cdnd_1st Jk5_cdjt_1st	55	Vk3_cdnd_2nd Jk5_cdnd_1st	60
Vκ11Jκ5 ESC	Vk11-125_end_1st Jk5_end_1st	53	Vk11-125_end_2nd Jk5_end_2nd	60
Vκ11Jκ5 recombination	Vk11-125_cdjt_1st Jk5_cdjt_1st	55	Vk11-125_cdjt_2nd Jk5_cdnd_1st	60

Vk16Jk5 ESC	Vk16-104_end_1st Jk5_end_1st	53	Vk16-104_end_2nd Jk5_end_2nd	60
Vk16Jk5 recombination	Vk16-104_cdjt_1st Jk5_cdjt_1st	55	Vk16-104_cdjt_2nd Jk5_cdnd_1st	60

Table 2.6 Oligonucleotides to generate Artemis deficient cell lines

Oligonucleotide	Sequence (5'-3')
Artemis_PCR_F(703bp)	TAGGAGAATTGCTTGAAC
Artemis_PCR_R(703bp)	CAGGAGAATTGCTTGAAC
Artemis_F	GGCACAAACCCATGATCCCAG
Artemis_R	GGCAGATAGCTTGAGGTCAGG
Artemis_knockout_oligo 1	CACCGTCTTAGAGCTGGTCCGAAGC
Artemis_knockout_oligo 2	AAACGCTTCGGACCAGCTCTAAGAC
Artemis_shRNA_oligo 1	CCGGAAAGTTCCATACCCGGAAACTCTCGAGAGTTTCCGGGTATGG AACTTTTTTTTTG
Artemis_shRNA_oligo 2	AATTCAAAAAAAAGTTCCATACCCGGAAACTCTCGAGAGTTTCCGGG TATGGAACTTT
Artemis_262_FWD	GCGGCTTATGGCTATGAATATCTG
Artemis_396_RVS	CTGAGTGTTGCGGTCTGTTG
Artemis_ShRNA_F_new	CCGGAAGTGAAGAGAGCTAGAACAGCTCGAGCTGTTCTAGCTCTCT TCAGTTTTTTTT
Artemis_ShRNA_R_new	AATTAATAAAAACTGAAGAGAGCTAGAACAGCTCGAGCTGTTCTAGCT CTCTTCAGTT
HPRT_FWD_human	GACCAGTCAACAGGGGACAT
HPRT_RVS_human	AACAATTTCGTGGGGTCTTTTT
Scramble_ShRNA_F	CCGGCCTAAGGTTAAGTCGCCCTCGCTCGAGCGAGGGCGACTTAAC CTTAGGTTTTT
Scramble_ShRNA_R	AATTAATAAACCTAAGGTTAAGTCGCCCTCGCTCGAGCGAGGGCGAC TTAACCTTAGG
SgRNA_Apro_184_F	CACCGGGCTGCGTTCGGCCGCCCAA
SgRNA_Apro_184_R	AAACTTGGGCGGCCGAACGCAGCCC
SgRNA_Apro_374_F	CACCGGGCGGGAATGAGCCTTGAC
SgRNA_Apro_374_R	AAACGTGCAAGGCTCATTCCCGCCC

Table 2.7 Oligonucleotides used for cloning

Oligonucleotide	Sequence (5'-3')
ESC23d7_fwd	GATCCCTGCAGGGTTTTTGTTCAGTCTGTAGCACTGTGGTAGTACTCCACT GTCTGGCTGTACAAAAACCCCTGCAGC
ESC23d7_rev	TCGAGCTGCAGGGTTTTTGTACAGCCAGACAGTGGAGTACTACCACTGTGCT ACAGACTGGAACAAAAACCCCTGCAGG
T7_fwd	TAATACGACTCACTATAGGG
bGH_rvs	TAGAAGGCACAGTCGAGG
pWPI_EGFP remove_MluI_fwd	TAACCCGGGAGCTAGTCGAGCTCAAGC
pWPI_EGFP remove_XmaI_fwd	AGCACGCGTGGTATTATCATCGTGTTCCTCAAAGG
Ruby_MluI_F	AGCACGCGTGATGGTGTCTAAGGGCGAA
Ruby_XmaI_R	TAACCCGGGCTTACTTGTACAGCTCGTC
pBS_Label_Open_F	ATCGAATTCCTGCAGCCC
pBS_Label_Open_R	TAAGACAGGCCAGATCATCAAGC
Universal_488_fwd	TGATCTGGCCTGTCTTA
RSS_labelling_rvs	GATCTGCAGGGTTTTTGT
ESC_labelling_rvs	GCTGCAGGAATTCGAT
Myc_MBP_cRAG1_F	ATGGAGCAGAACTCATCTCAGAAGAGGATCTGATGGAAATCGAAGAAGGTA AACTG
HA_MBP_cRAG2_F	ATGTACCCATACGATGTTCCAGATTACGCTATGGAAATCGAAGAAGGTAAACT G
Tags_MBP_cRAGs_R	GGTGGTCTAGAGAATTCCTCACG

2.8 Plasmids used during this project

Plasmid	Description
pEF-McR1	The cDNA for core RAG1 (amino acids 384-1008) was cloned in frame with a maltose binding protein tag into the Xba I and Bam HI sites of pEF-XC (Mizushima and Nagata, 1990).
pEF-McR2	The cDNA for core RAG2 (amino acids 1-387) was cloned in frame with a maltose binding protein tag into the Xba I and Bam HI sites of pEF-XC (Mizushima & Nagata, 1990).
pEF-MfIR2	The cDNA for full-length RAG2 was cloned in frame with a maltose binding protein tag into the Xba I and Bam HI sites of pEF-XC (Mizushima & Nagata, 1990).
pEF-Myc-MBP-cRAG1	The Myc tag was cloned into pEF-McR1 after the start code of the maltose binding protein tag by Q5 site-directed mutagenesis PCR using a forward primer containing Myc tag sequence and a reverse primer (Table 2.7).

pEF-HA-MBP-cRAG2	The HA tag was cloned into pEF-McR2 after the start code of the maltose binding protein tag by Q5 site-directed mutagenesis PCR using a forward primer containing HA tag sequence and a reverse primer (Table 2.7).
pJβ2.7+	J β 2.7 carries a non-consensus 12-RSS from the human TCR β J region 5'-GGTTT G CATGCGGGGGTGCACCT C CGTG-3' cloned into the polylinker of pcDNA3.1
pJδ1+	The non-consensus natural ESC sequence formed by linking Ki/Jk3 was cloned into the Pst I and Sal I sites of the polylinker of pcDNA3.1.
pcDNA3.1_ESC	Oligonucleotides containing a consensus ESC oligonucleotide with terminal Bam HI and Xho I sites were annealed and cloned into the polylinker of pcDNA3.1. An additional 120 bp Sal I/Bam HI fragment was cloned into the Xho I site of the polylinker to enable the ESC plasmid to be distinguished from the 12- and 23-RSS plasmids
p23RSS	Oligonucleotides containing a consensus 23-RSS oligonucleotide with terminal Bam HI and Hind III sites were annealed and ligated into pcDNA3.1 cut with Bam HI and Hind III. To remove a cRSS in the polylinker of pcDNA3.1, the vector was digested with Eco RI and Xba I and the oligonucleotide sequence AATTCTCTACAGTCAGTGAATCGATCATAGT was ligated in to give a unique sequence that could be used for as a primer for PCR.
p12RSS	Oligonucleotides containing a consensus 12-RSS oligonucleotide with terminal Bam HI and Hind III sites ligated into the polylinker of pcDNA3.1. To remove a cRSS in the polylinker of pcDNA3.1, the vector was digested with Eco RI and Xba I and the oligonucleotide sequence AATTCGTCTCACATCGATGAACTATTGTATGT was ligated in to give a unique sequence that could be used for as a primer for PCR.
pESC12d7+ and pESC23d7+	The ESC sequence in pESC was replaced with a mutant ESC sequence in which the heptamer in the 12-RSS or 23-RSS within the ESC was deleted (ESC12d7 and ESC23d7, respectively).
pESC12d9+ and pESC23d9+	The ESC sequence in pESC was replaced with a mutant ESC sequence in which the nonamer in the 12-RSS or 23-RSS within the ESC was mutated (ESC12d9 and ESC23d9, respectively).
pJH548 (RAG1 expression vector)	The full length mouse RAG1 cDNA was cloned into pCDM8 so that the cDNA was expressed via the CMV immediate early promoter (Sadofsky et al 1993).
pCS2MT-cRAG1 (core RAG1 expression vector)	Core mouse RAG1 cDNA (amino acids 384-1008) was cloned into the pCS2MT vector (Turner & Weintraub, 1994)
pEFfIRAG2 (full length RAG2 expression vector)	The full length mouse RAG2 cDNA was cloned into the pEF-XC vector (Mizushima & Nagata, 1990) so that the cDNA was expressed via the strong EF1 α promoter.
pEFXC-cRAG2 (core RAG2 expression vector)	Core mouse RAG2 cDNA (amino acids 1-387) was cloned into the pEF-XC vector (Mizushima & Nagata, 1990) so that the cDNA was expressed via the strong EF1 α promoter.
pWPI	Second generation bicistronic lentiviral vector to allow for simultaneous expression of a transgene and EGFP marker. A gift from Didier Trono (Addgene plasmid #12254)

pWPI-12RSS	The 12RSS sequence was amplified from p12RSS using the T7 forward and bGH reverse primers (Table 2.7). The PCR product was cloned into the PmeI site of pWPI
pWPI-23RSS	As above but the 23-RSS was amplified from p23RSS
pWPI-ESC and pWPI-mutant ESC vectors	As above but the ESC signal joint sequence was amplified from pESC+ whilst the mutant ESC sequences were amplified from the respective mutant pESC+ plasmids
pWPI-Ruby-12-RSS	The GFP in pWPI-12RSS was removed by amplifying vector with primers pWPI-EGFPremove_MluI_fwd and pWPI_EGFPremove_XmaI_rvs (Table 2.7). A Ruby fragment was amplified from pcDNA_Ruby with primers Ruby_XmaI_fwd and Ruby_MluI_rvs (Table 2.7). After digestion of the vector and Ruby PCR products with MluI and XmaI. Ruby fragment was ligated with GFP-removed pWPI-12RSS vector, resulting in pWPI-Ruby-12-RSS
pWPI-Ruby-23-RSS	As above but the GFP was removed from pWPI-23RSS vector
pWPI-Ruby-ESC	As above but the GFP was removed from pWPI-ESC vector

I) Plasmid maps

Maps for all plasmids used during this project were shown below and were prepared using SnapGene Viewer software (Version 2.5).

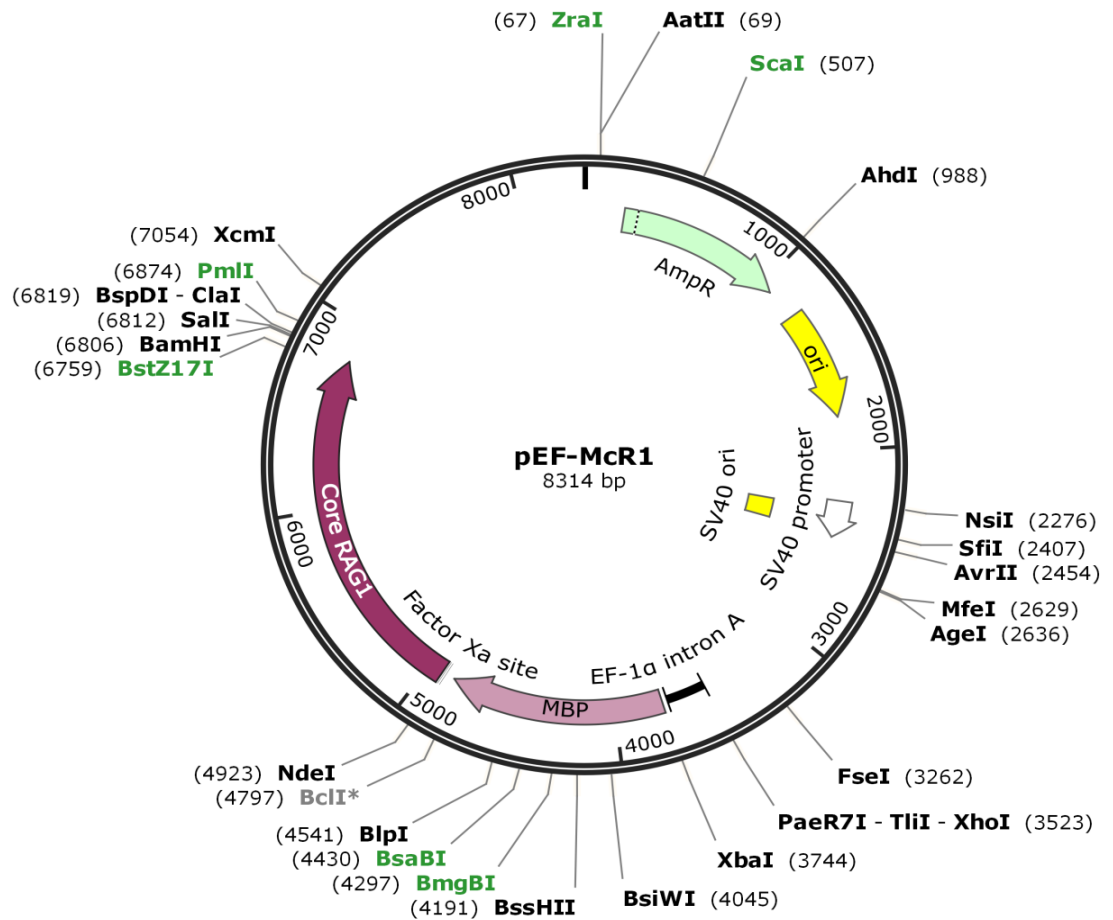


Figure 2.7 Map of pEF-McR1 (pEFXC-MBP-cRAG1)

All the unique restriction enzyme sites are shown. AmpR = Ampicillin resistance gene, ori = origin of replication, SV40 = Simian virus 40, EF-1 α intron A = truncated intron upstream of the start codon of elongation factor-1 α , MBP = maltose-binding protein.

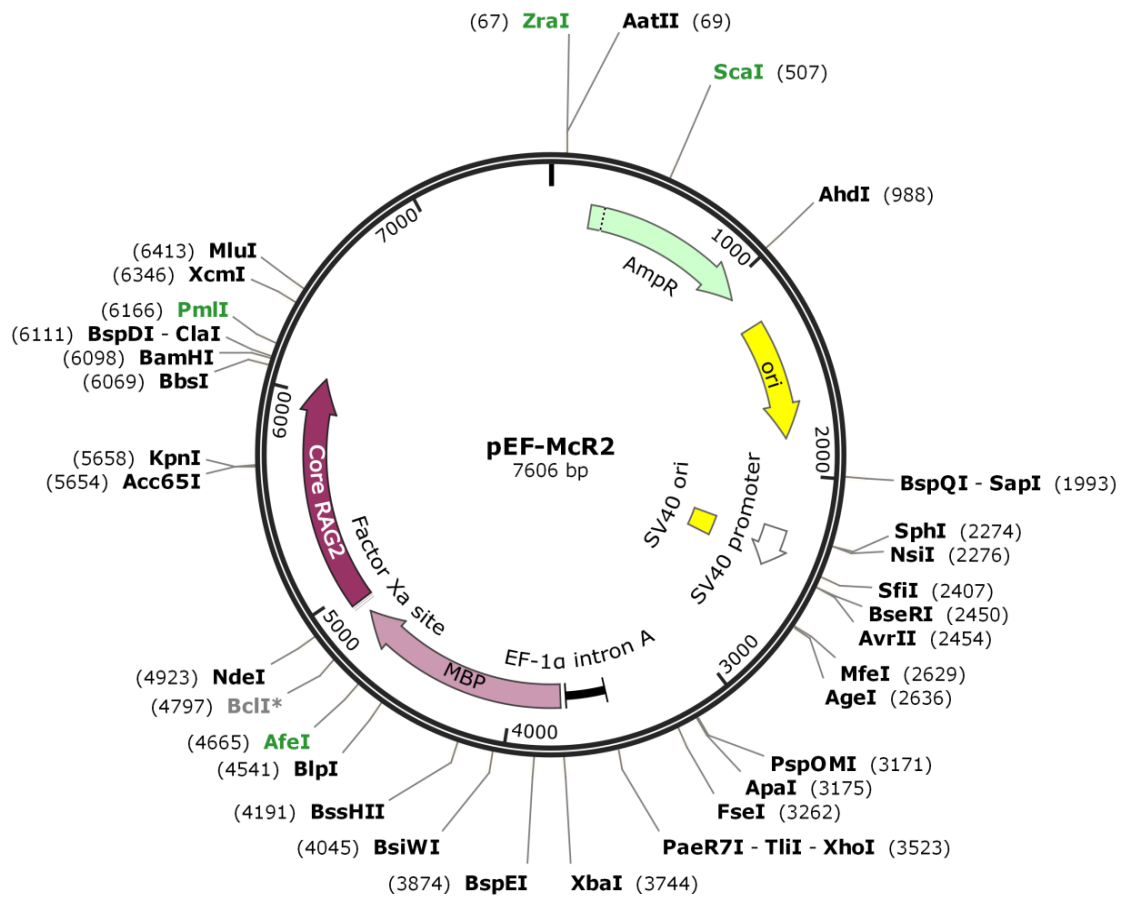


Figure 2.8 Map of pEF-McR2 (pEFXC-MBP-cRAG2)

All the unique restriction enzyme sites are shown. AmpR = Ampicillin resistance gene, ori = origin of replication, SV40 = Simian virus 40, EF-1 α intron A = truncated intron upstream of the start codon of elongation factor-1 α , MBP= maltose-binding protein.

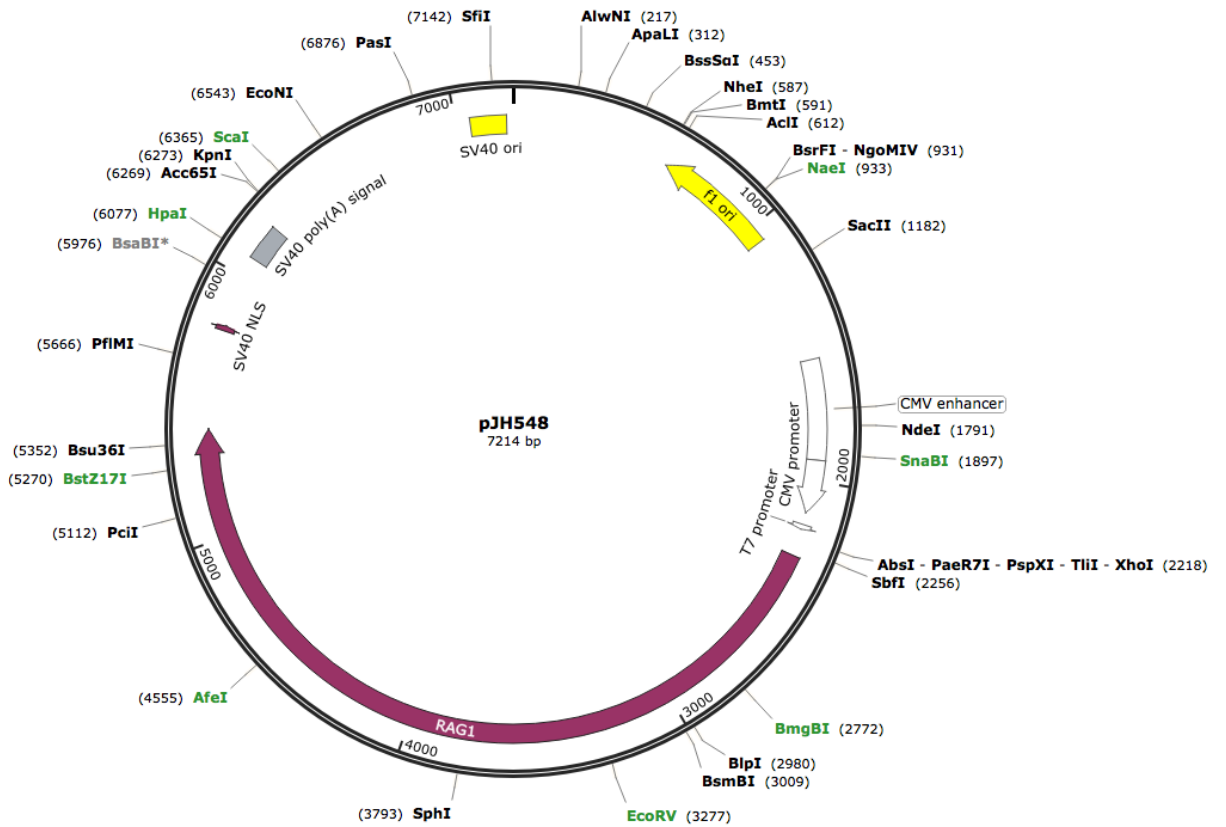


Figure 2.9 Map of pJH548 (fIRAG1)

All the unique restriction enzyme sites are shown. Ori = origin of replication, SV40 = Simian virus 40, CMV = Human cytomegalovirus promoter

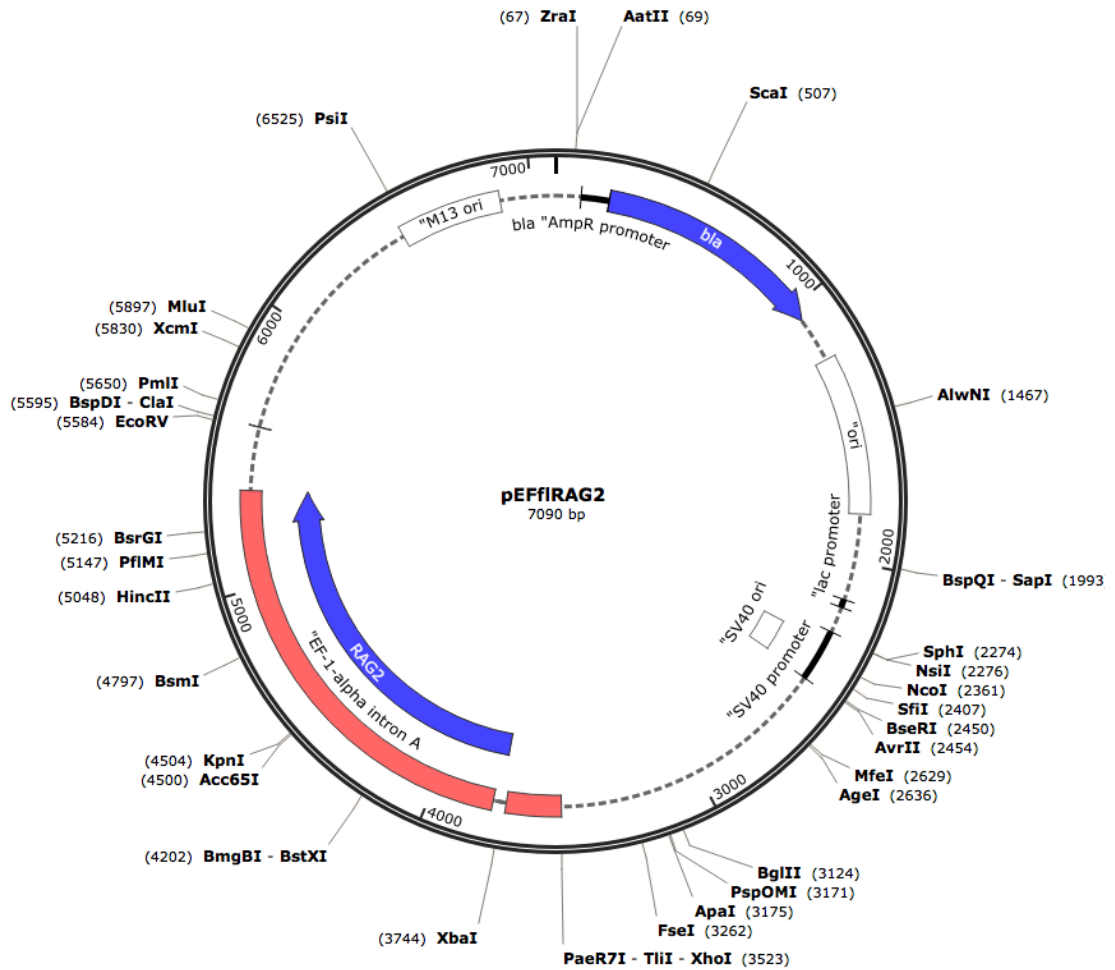


Figure 2.10 Map of pEFfIRAG2

All the unique restriction enzyme sites are shown. Ori = origin of replication, SV40 = Simian virus 40, EF-1 α intron A = truncated intron upstream of the start codon of elongation factor-1 α , bla = beta-lactamase.

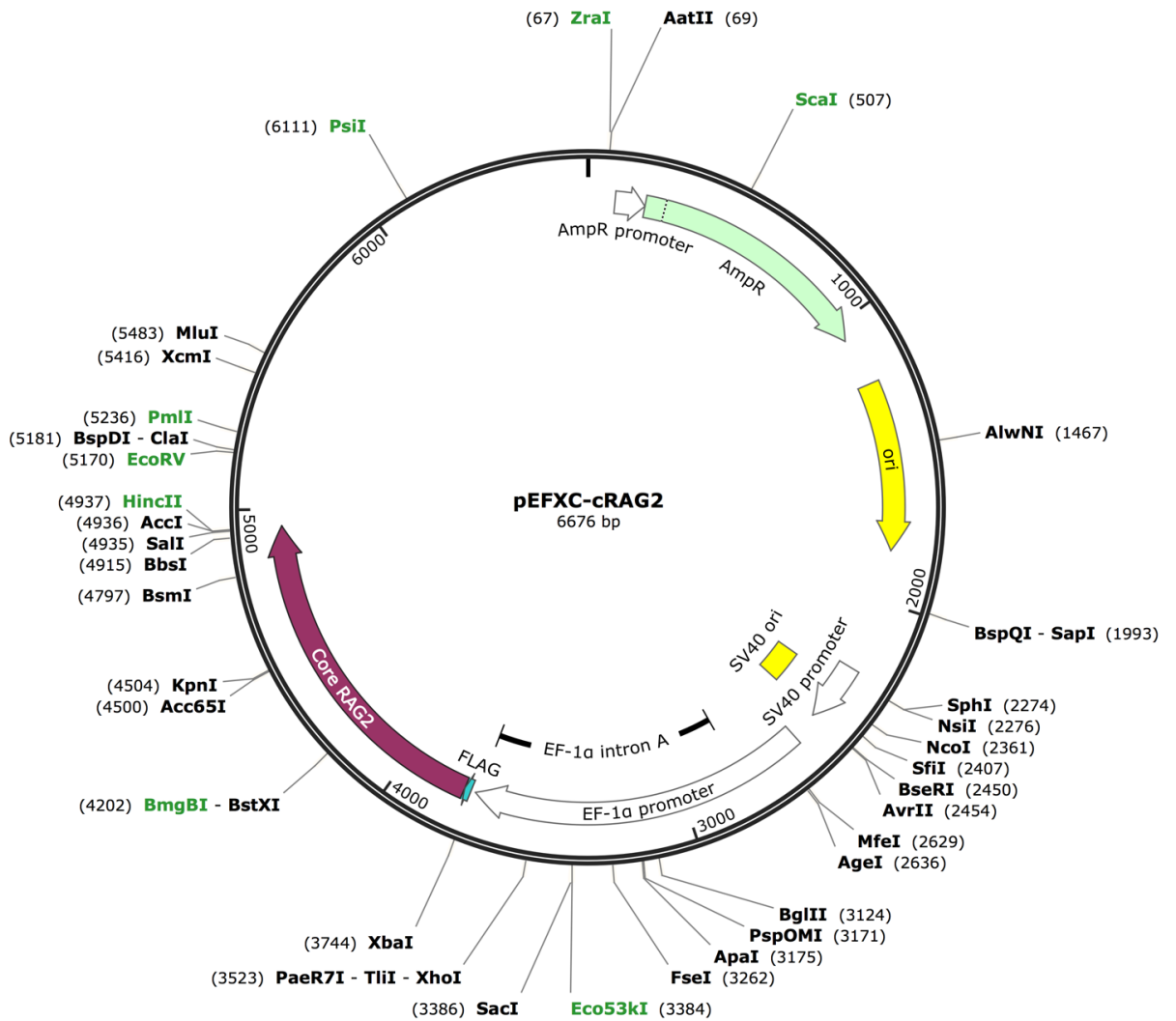


Figure 2.11 Map of pEFXC-cRAG2

All the unique restriction enzyme sites are shown. AmpR = Ampicillin resistance gene, ori = origin of replication, SV40 = Simian virus 40, EF-1 α intron A = truncated intron upstream of the start codon of elongation factor-1 α , FLAG = DYKDDDDK= FLAG tag

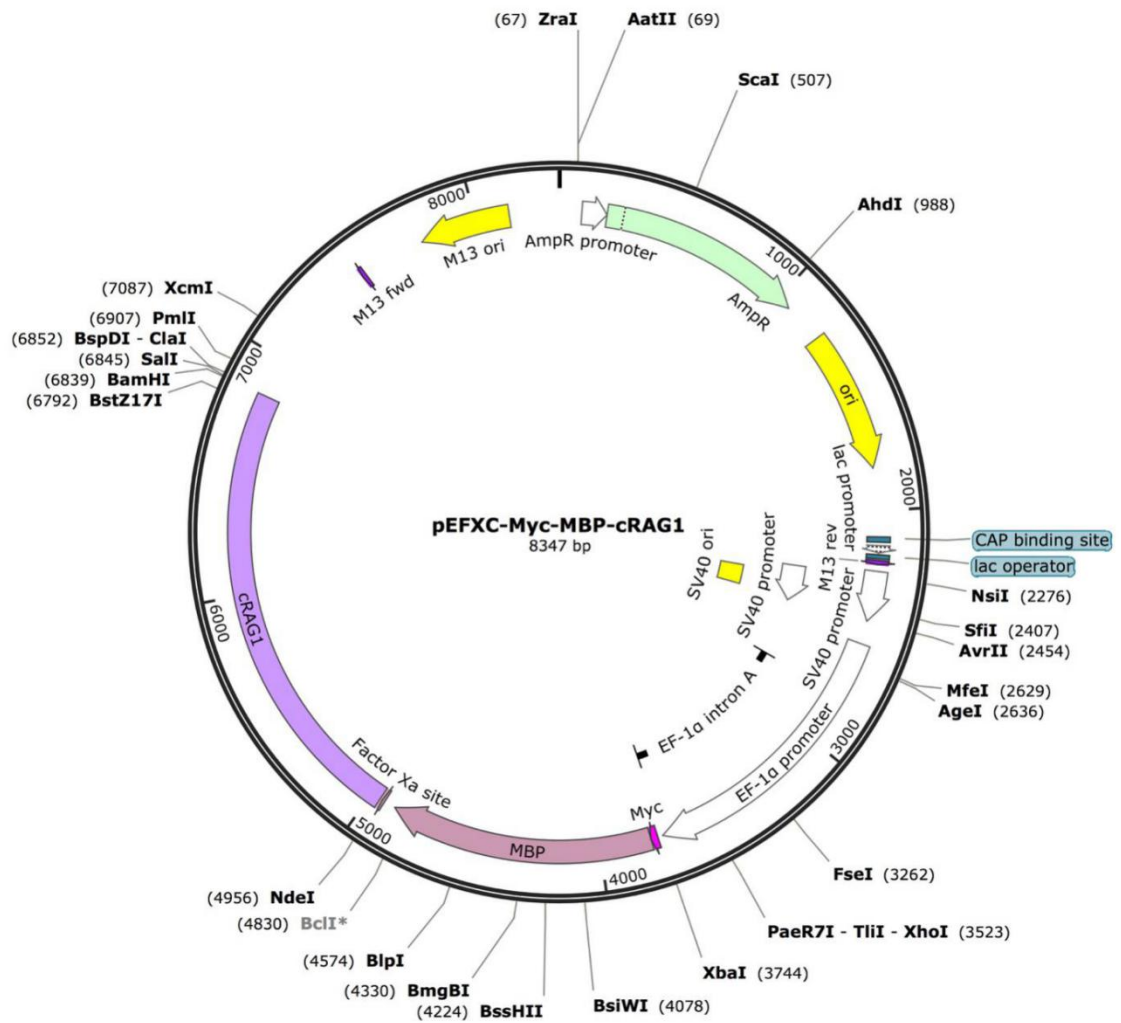


Figure 2.12 Map of pEFXC-Myc-MBP-cRAG1

A Myc tag was inserted into pEFXC-MBP-cRAG1 in frame with MBP. All the unique restriction enzyme sites are shown. AmpR = Ampicillin resistance gene, ori = origin of replication, SV40 = Simian virus 40, EF-1 α intron A = truncated intron upstream of the start codon of elongation factor-1 α , MBP = maltose-binding protein.

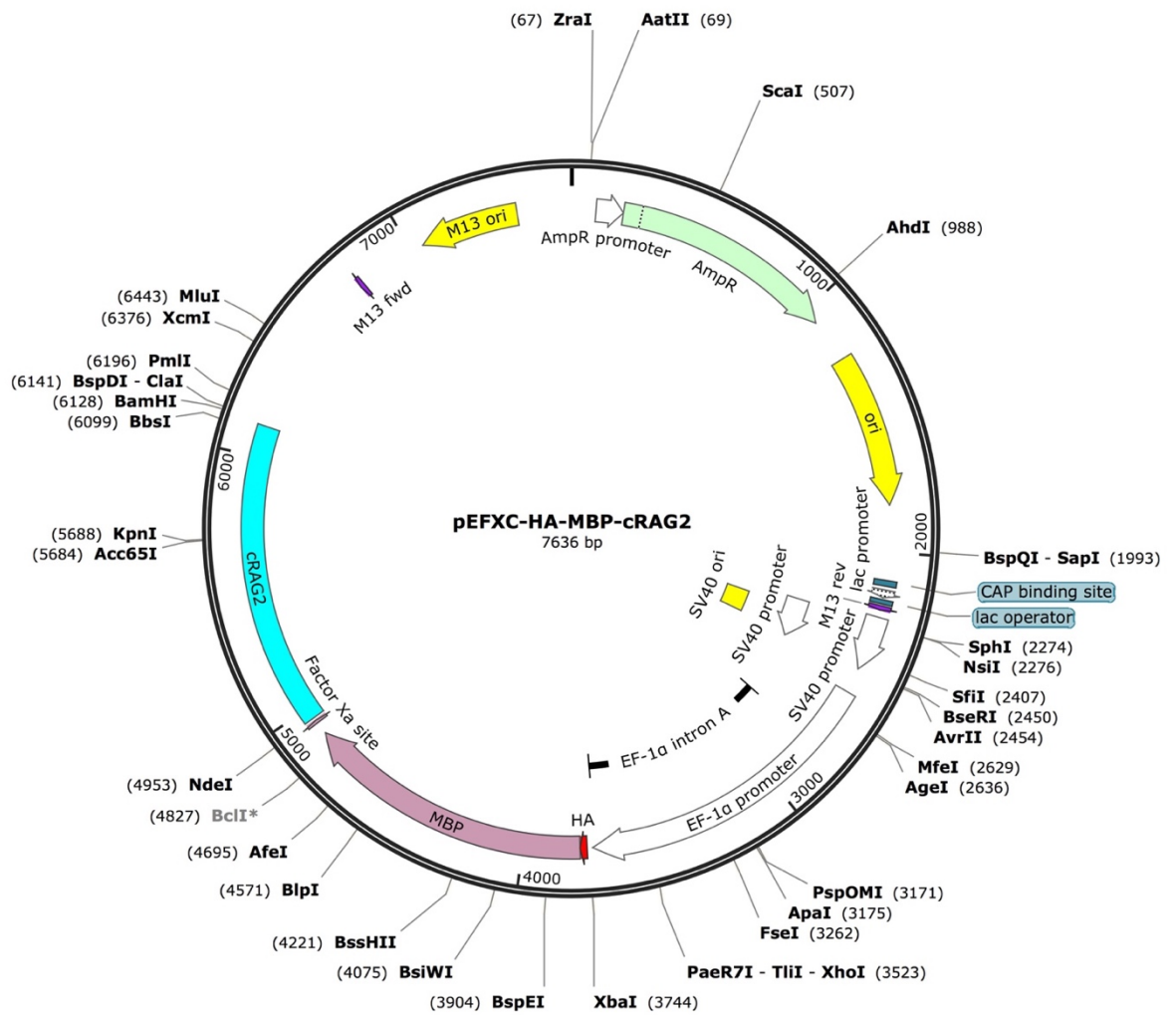


Figure 2.13 Map of pEFXC-HA-MBP-cRAG2

A HA tag was inserted into pEFXC-MBP-cRAG2 in frame with MBP. All the unique restriction enzyme sites are shown. AmpR = Ampicillin resistance gene, ori = origin of replication, SV40 = Simian virus 40, EF-1 α intron A = truncated intron upstream of the start codon of elongation factor-1 α , MBP = maltose-binding protein.

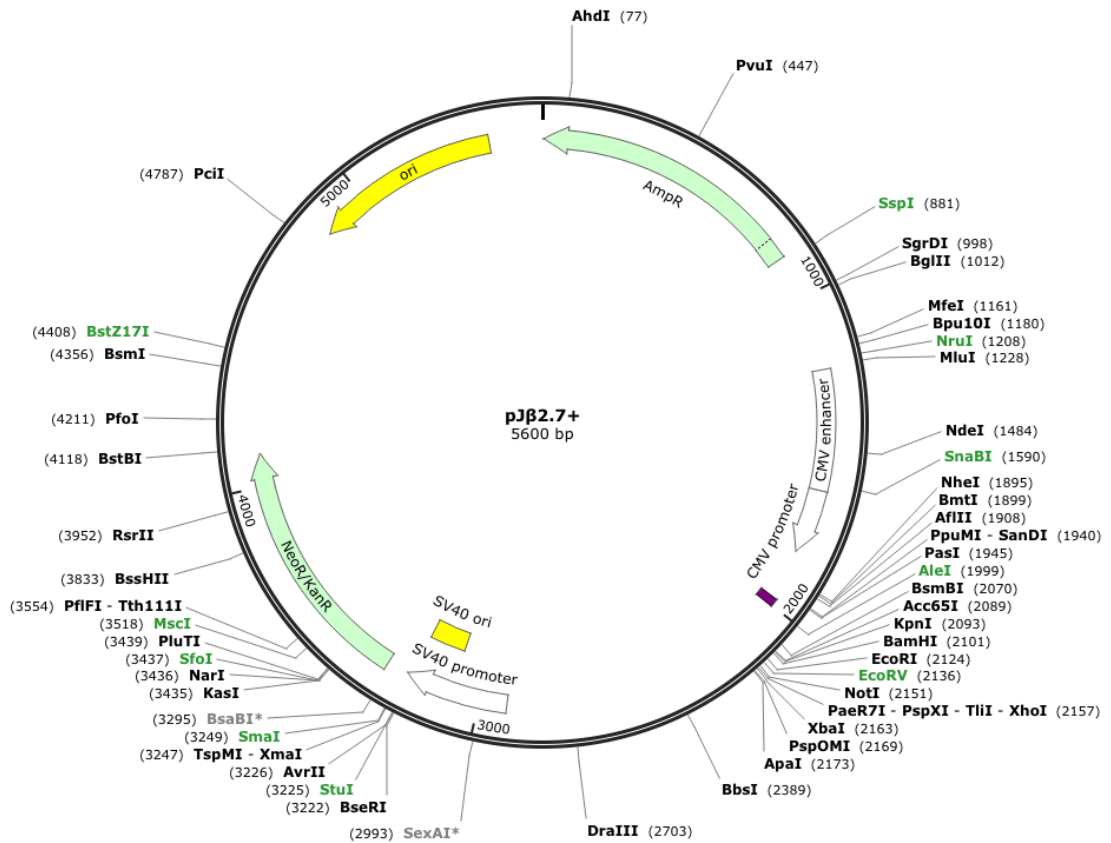


Figure 2.14 Map of pJβ2.7+

pJβ2.7+ contains the CMV promoter upstream of a non-consensus 12-RSS to enable transcription through the 12-RSS. All the unique restriction enzyme sites are shown. Ori = origin of replication, AmpR = Ampicillin resistance gene, SV40 = Simian virus 40, CMV = human cytomegalovirus, NeoR/KanR = neomycin/kanamycin resistance gene.

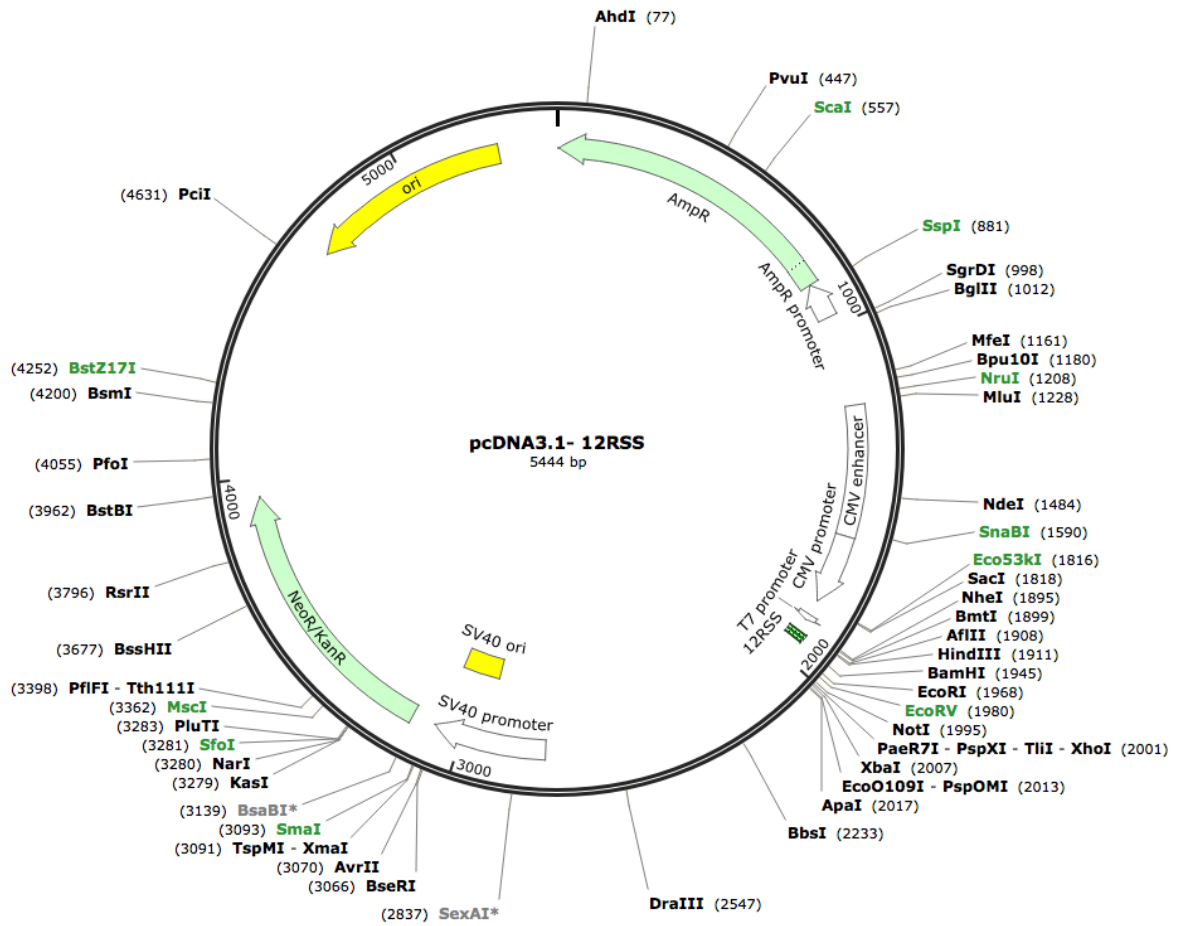


Figure 2.15 Map of pcDNA3.1-12-RSS (p12RSS)

pcDNA3.1-12-RSS contains the CMV promoter upstream of a consensus 12-RSS to enable transcription through the 12-RSS. All the unique restriction enzyme sites are shown. Ori = origin of replication, AmpR = Ampicillin resistance gene, SV40 = Simian virus 40, CMV = human cytomegalovirus promoter, NeoR/KanR = neomycin/kanamycin resistance gene.

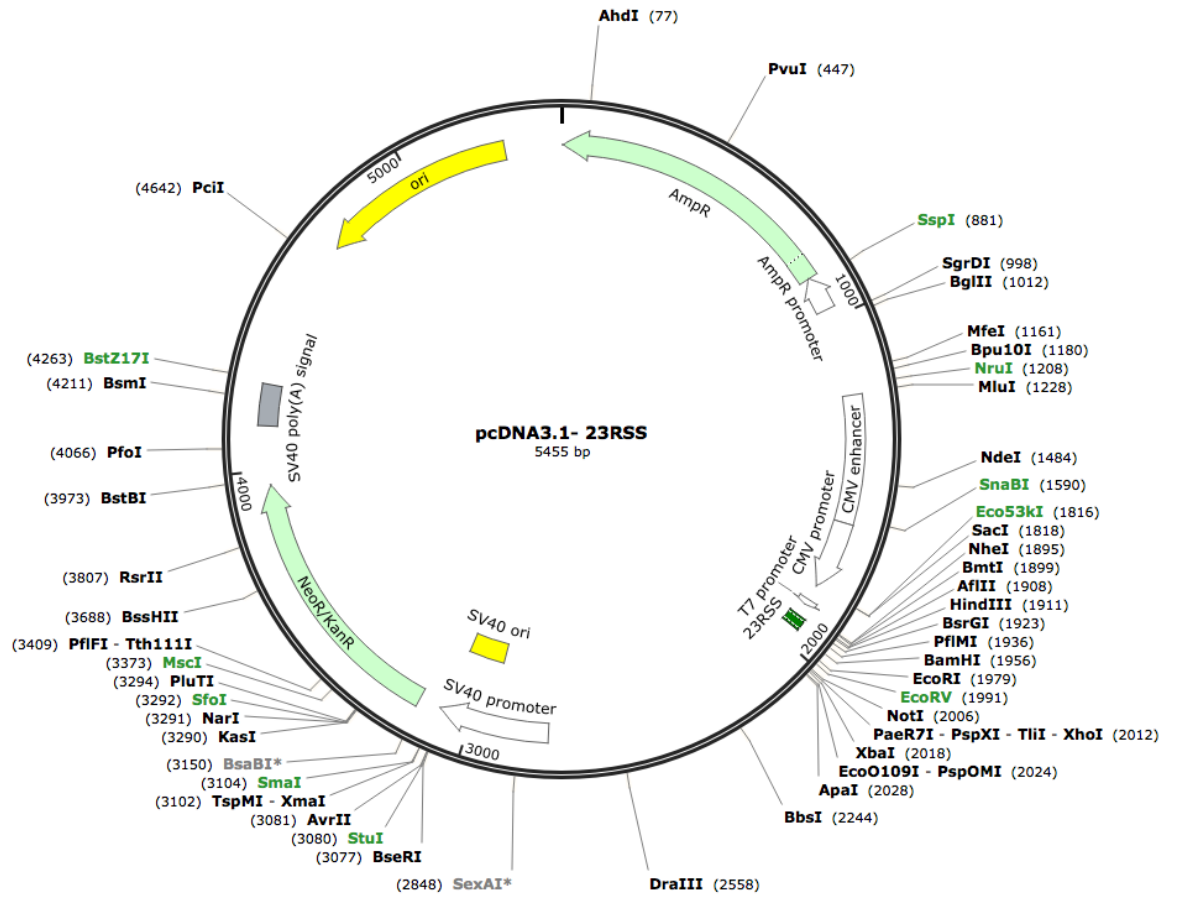


Figure 2.16 Map of pcDNA3.1-23-RSS (p23RSS)

pcDNA3.1-23-RSS contains the CMV promoter upstream of a consensus 23-RSS to enable transcription through the 23-RSS. All the unique restriction enzyme sites are shown. Ori = origin of replication, AmpR = Ampicillin resistance gene, SV40 = Simian virus 40, CMV = human cytomegalovirus promoter, NeoR/KanR = neomycin/kanamycin resistance gene.

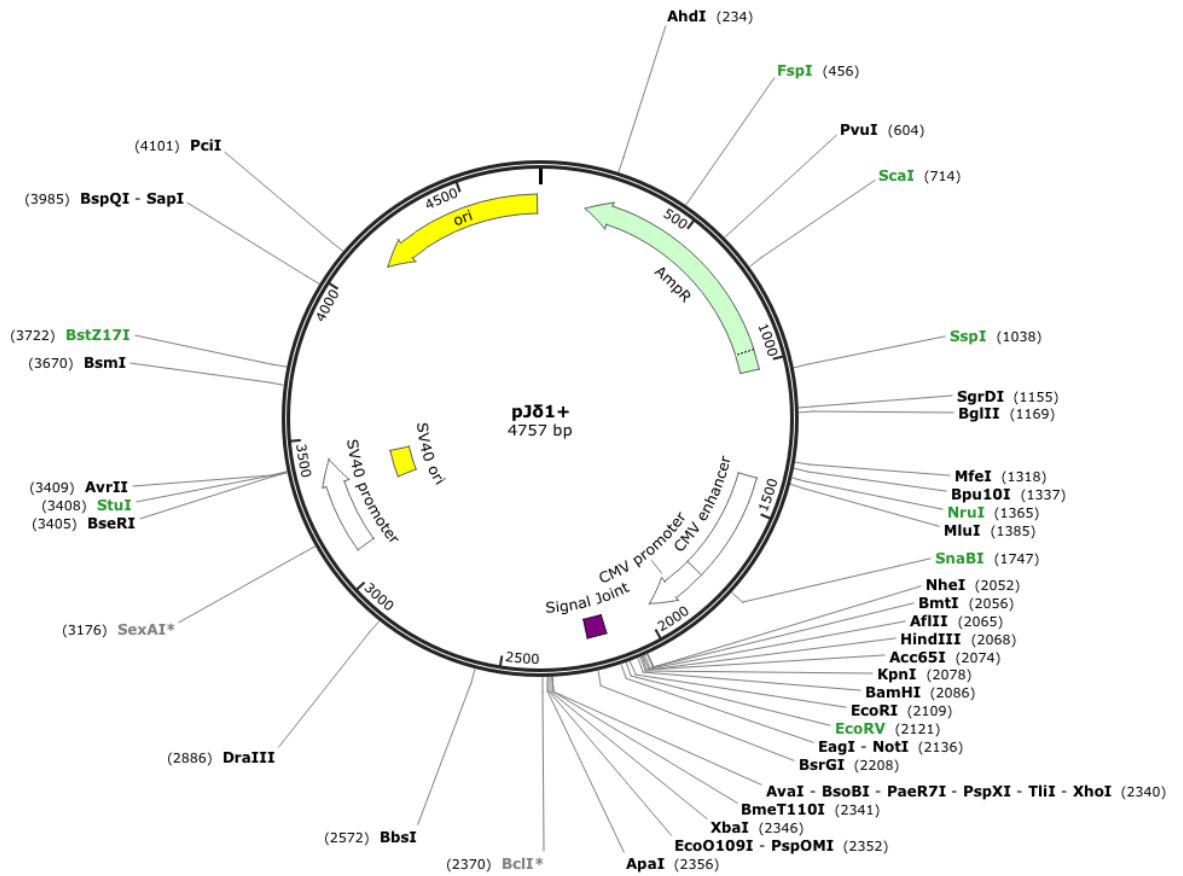


Figure 2.17 Map of pJδ1+

pJδ1+ contains two head to head RSSs, where one is a non-consensus 12-RSS and the other is a non-consensus 23-RSS. All the unique restriction enzyme sites are shown. Ori = origin of replication, AmpR = Ampicillin resistance gene, CMV = human cytomegalovirus promoter, Signal Joint = head-to-head 12-RSS-23-RSS, SV40 = Simian virus 40.

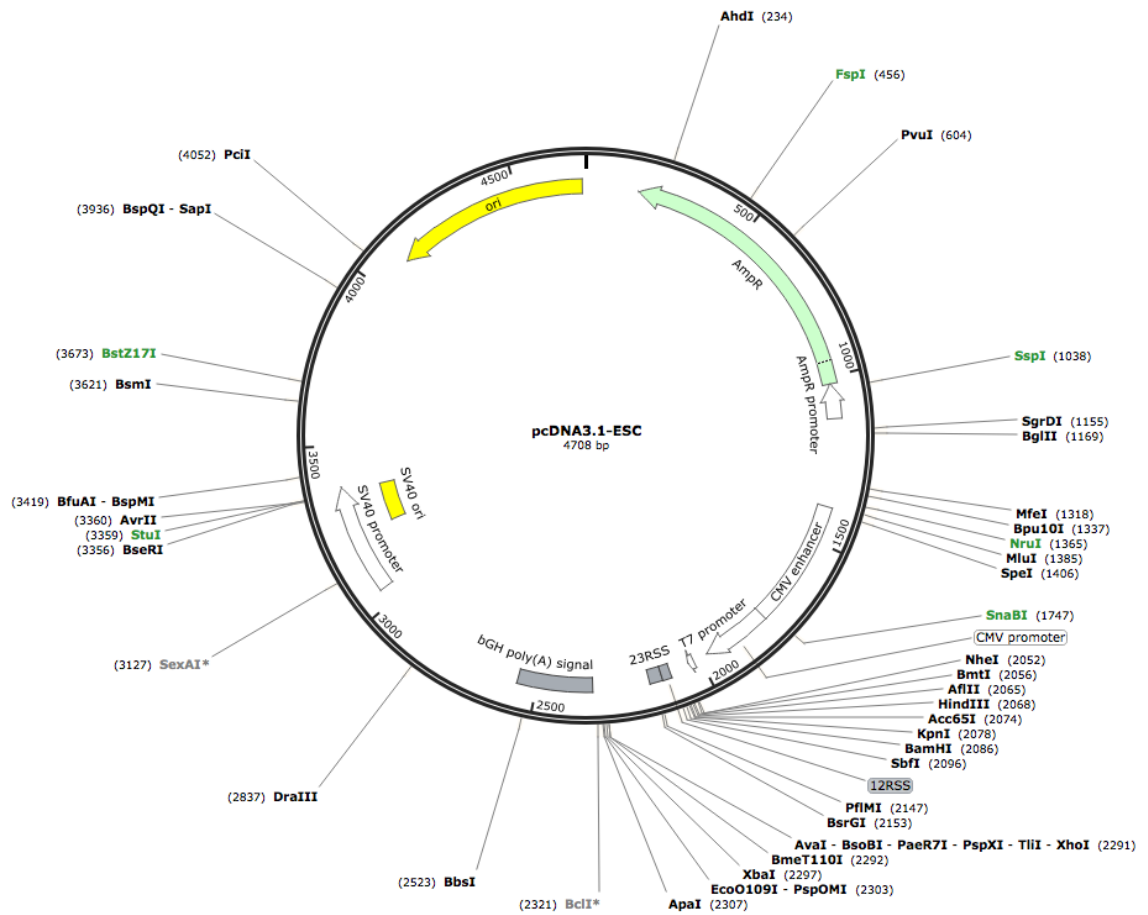


Figure 2.18 Map of pcDNA3.1-ESC (pESC)

pcDNA3.1-ESC contains a consensus 12-RSS and a consensus 23-RSS. All the unique restriction enzyme sites are shown. Ori = origin of replication, Amp^r = Ampicillin resistance gene, CMV = human cytomegalovirus promoter, bGH poly(A) signal = bovine growth hormone polyadenylation signal, SV40 = Simian virus 40.

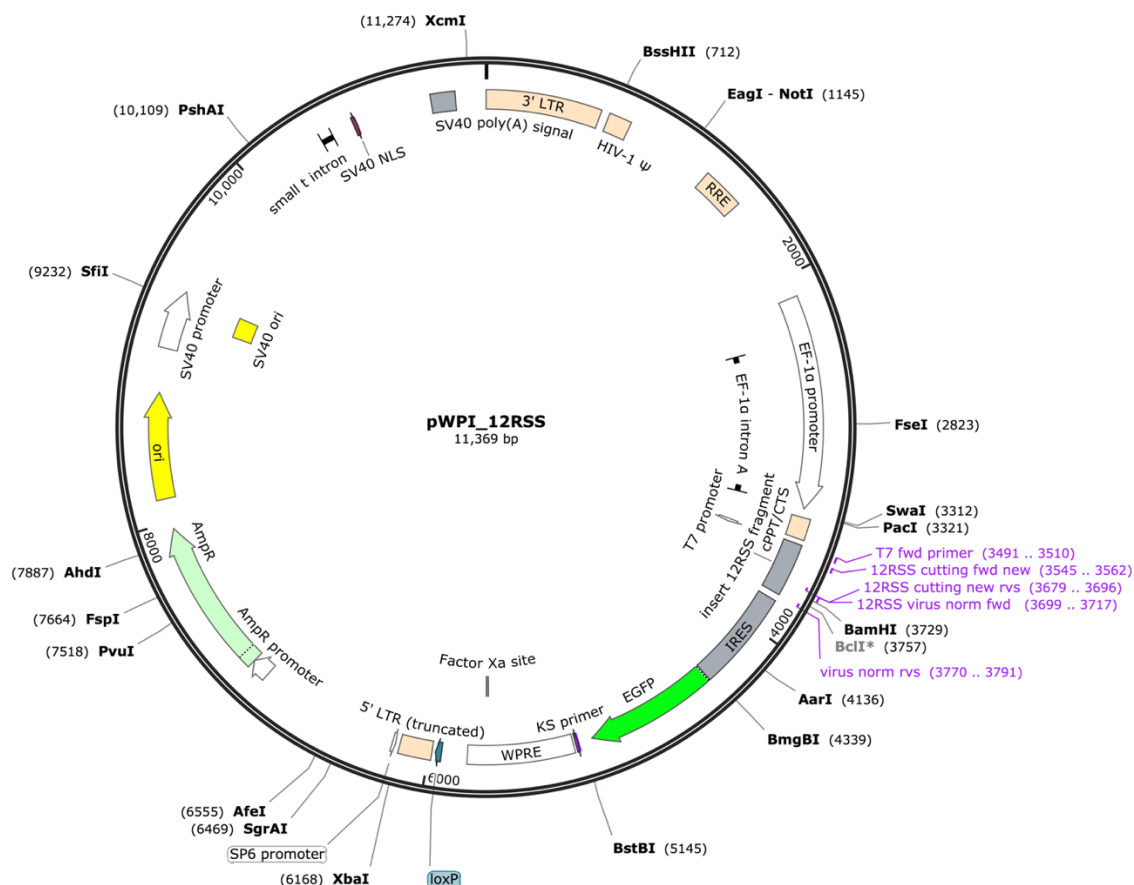


Figure 2.19 Map of pWPI-12RSS

A 12RSS fragment was amplified from p12RSS and cloned into the pWPI vector by blunt end ligation. All the unique restriction enzyme sites are shown. Ori = origin of replication, AmpR = Ampicillin resistance gene, CMV = human cytomegalovirus promoter, bGH poly(A) signal = bovine growth hormone polyadenylation signal, SV40 = Simian virus 40, EF-1 α intron A = truncated intron upstream of the start codon of elongation factor-1 α , IRES = internal ribosome entry site, EGFP = enhanced GFP, WPRE = woodchuck hepatitis virus post-transcriptional regulatory element. Main primers used are shown in purple.

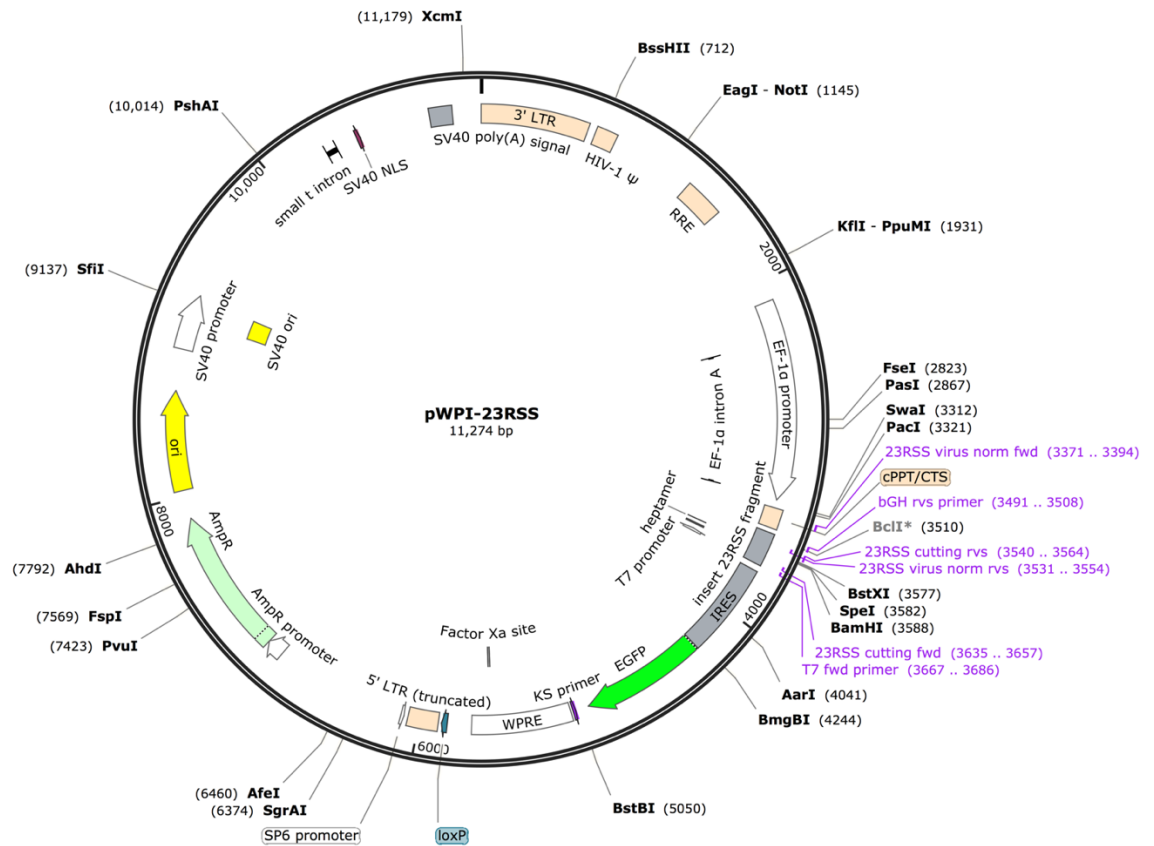


Figure 2.20 Map of pWPI-23RSS

A 23RSS fragment was amplified from p23RSS and cloned into the pWPI vector by blunt end ligation. All the unique restriction enzyme sites are shown. Ori = origin of replication, AmpR = Ampicillin resistance gene, CMV = human cytomegalovirus promoter, bGH poly(A) signal = bovine growth hormone polyadenylation signal, SV40 = Simian virus 40, EF-1 α intron A = truncated intron upstream of the start codon of elongation factor-1 α , IRES = internal ribosome entry site, EGFP = enhanced GFP, WPRE = woodchuck hepatitis virus post-transcriptional regulatory element. Main primers used are shown in purple.

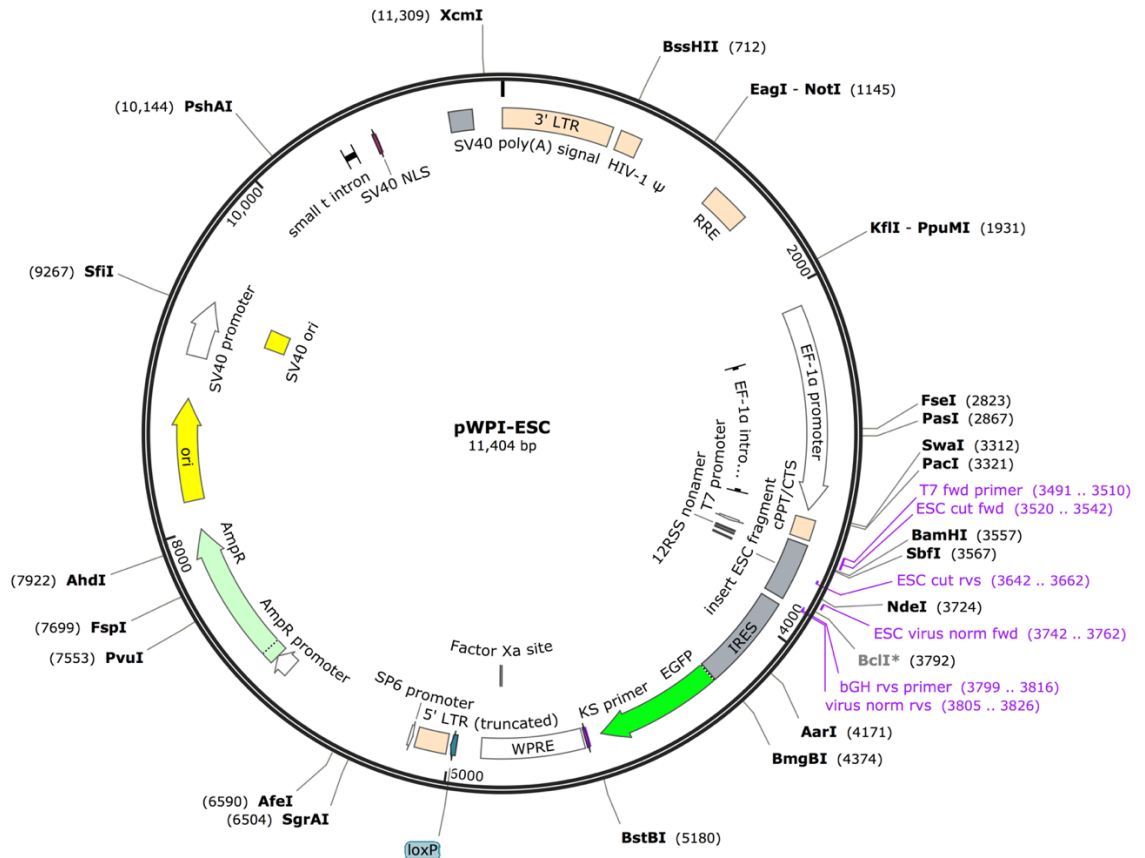


Figure 2.21 Map of pWPI-ESC

An ESC fragment was amplified from pESC and cloned into the pWPI vector by blunt end ligation. All the unique restriction enzyme sites are shown. Ori = origin of replication, AmpR = Ampicillin resistance gene, CMV = human cytomegalovirus promoter, bGH poly(A) signal = bovine growth hormone polyadenylation signal, SV40 = Simian virus 40, EF-1 α intron A = truncated intron upstream of the start codon of elongation factor-1 α , IRES = internal ribosome entry site, EGFP = enhanced GFP, WPRE = woodchuck hepatitis virus post-transcriptional regulatory element. Main primers used are shown in purple.

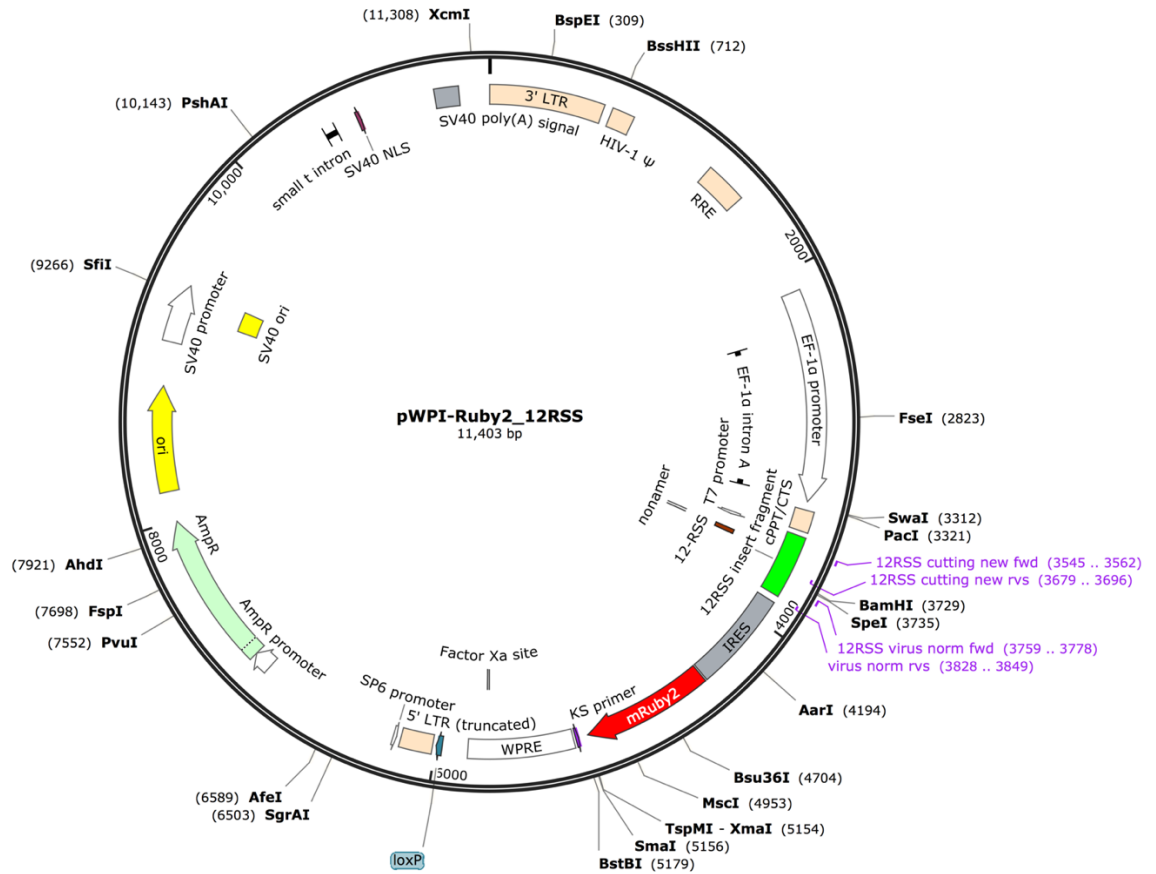


Figure 2.22 Map of pWPI-Ruby2_12RSS

The EGFP fragment in pWPI-12RSS was replaced by mRuby2 by cloning. All the unique restriction enzyme sites are shown. Ori = origin of replication, AmpR = Ampicillin resistance gene, CMV = human cytomegalovirus promoter, bGH poly(A) signal = bovine growth hormone polyadenylation signal, SV40 = Simian virus 40, EF-1 α intron A = truncated intron upstream of the start codon of elongation factor-1 α , IRES = internal ribosome entry site, mRuby2 = monomeric red fluorescent protein derived from a variant of the *Entacmaea quadricolor* proteins eqFP611, mRuby (Kredel et al., 2009; Lam et al., 2012), WPRE = woodchuck hepatitis virus post-transcriptional regulatory element. Main primers used are shown in purple.

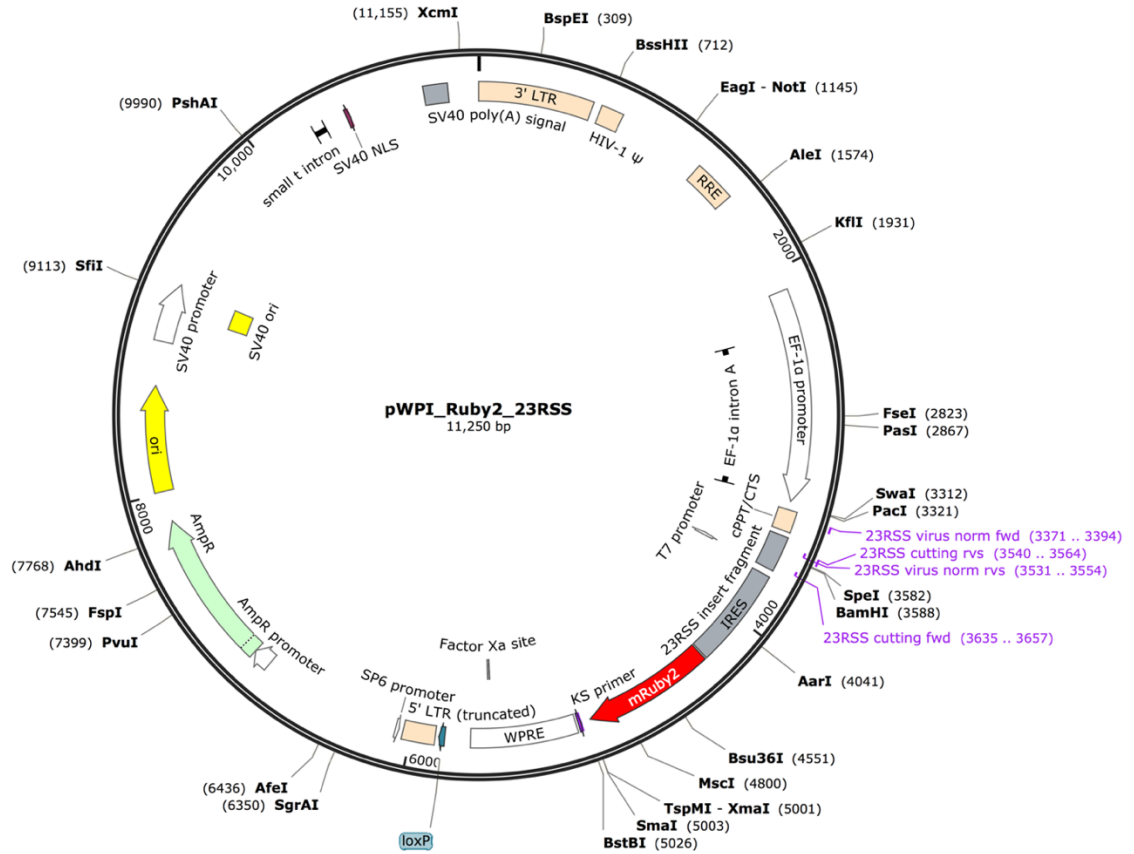


Figure 2.23 Map of pWPI_Ruby2_23RSS

The EGFP fragment in pWPI-23RSS was replaced by mRuby2 by cloning. All the unique restriction enzyme sites are shown. Ori = origin of replication, AmpR = Ampicillin resistance gene, CMV = human cytomegalovirus promoter, bGH poly(A) signal = bovine growth hormone polyadenylation signal, SV40 = Simian virus 40, EF-1 α intron A = truncated intron upstream of the start codon of elongation factor-1 α , IRES = internal ribosome entry site, mRuby2 = monomeric red fluorescent protein derived from a variant of the *Entacmaea quadricolor* proteins eqFP611, mRuby (Kredel et al., 2009; Lam et al., 2012), WPRE = woodchuck hepatitis virus post-transcriptional regulatory element. Main primers used are shown in purple.

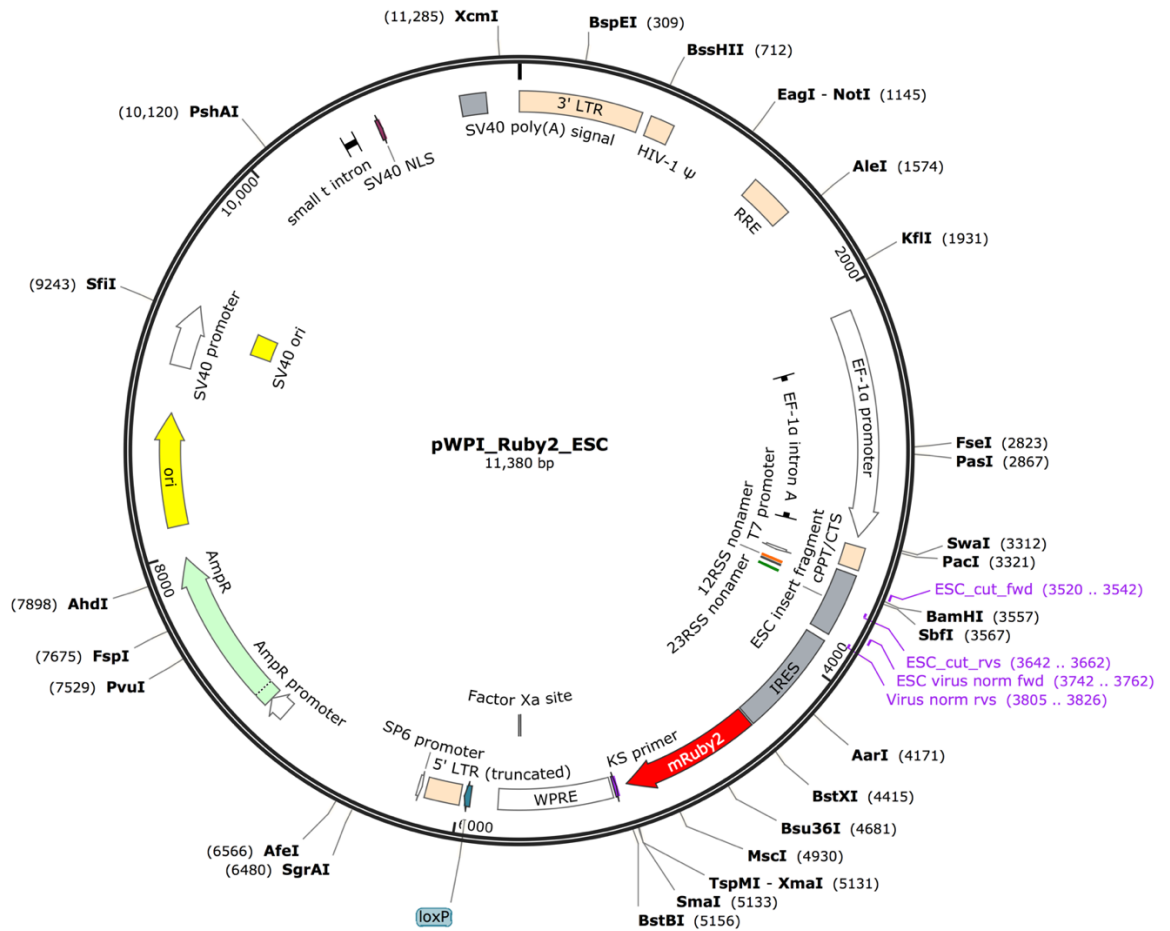


Figure 2.24 Map of pWPI_Ruby2_ESC

The EGFP fragment in pWPI-ESC was replaced by mRuby2 by cloning. All the unique restriction enzyme sites are shown. Ori = origin of replication, AmpR = Ampicillin resistance gene, CMV = human cytomegalovirus promoter, bGH poly(A) signal = bovine growth hormone polyadenylation signal, SV40 = Simian virus 40, EF-1 α intron A = truncated intron upstream of the start codon of elongation factor-1 α , IRES = internal ribosome entry site, mRuby2 = monomeric red fluorescent protein derived from a variant of the *Entacmaea quadricolor* proteins eqFP611, mRuby (Kredel et al., 2009; Lam et al., 2012), WPRE = woodchuck hepatitis virus post-transcriptional regulatory element. Main primers used are shown in purple.

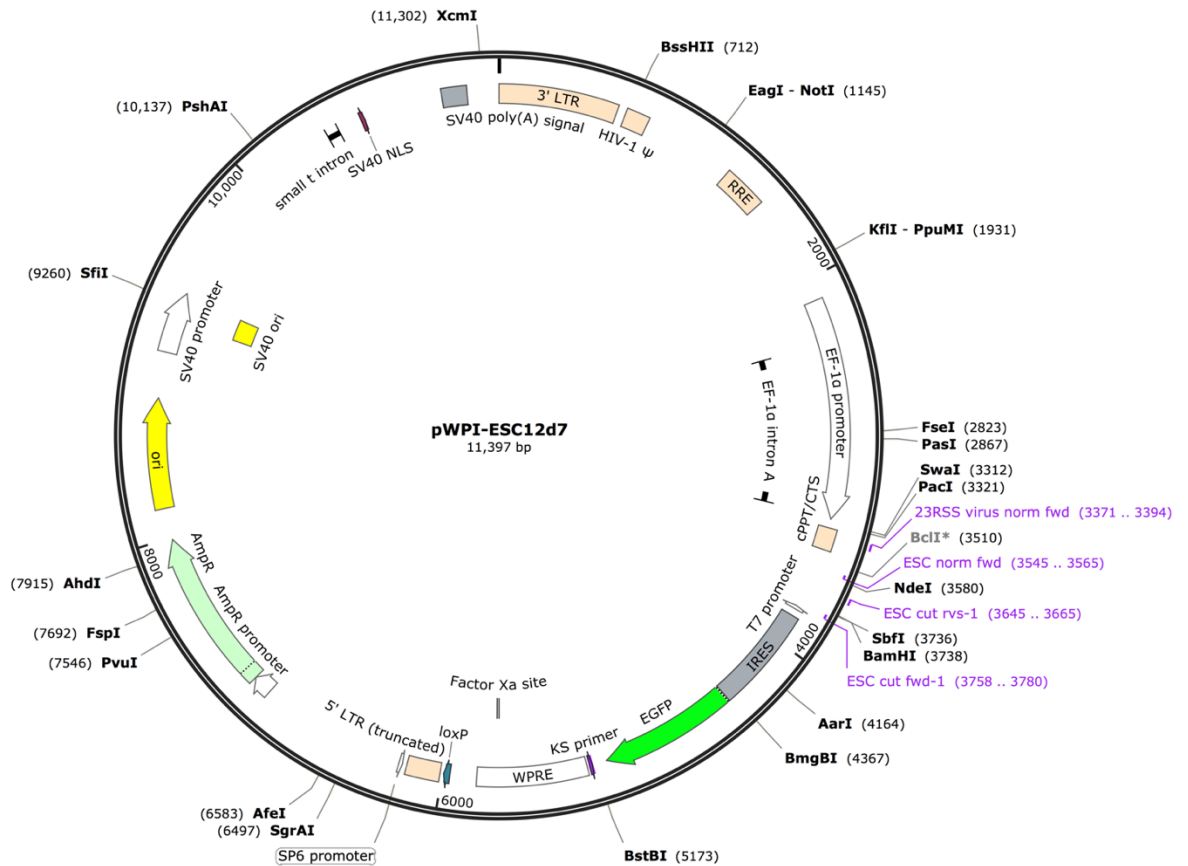


Figure 2.25 Map of pWPI-ESC12d7

The ESC12d7 fragment was amplified from pESC12d7 (not shown, generated by Chris Kirkham) and cloned into pWPI vector by blunt end ligation. All the unique restriction enzyme sites are shown. Ori = origin of replication, AmpR = Ampicillin resistance gene, CMV = human cytomegalovirus promoter, bGH poly(A) signal = bovine growth hormone polyadenylation signal, SV40 = Simian virus 40, EF-1 α intron A = truncated intron upstream of the start codon of elongation factor-1 α , IRES = internal ribosome entry site, EGFP = enhanced GFP, WPRE = woodchuck hepatitis virus post-transcriptional regulatory element. Main primers used are shown in purple.

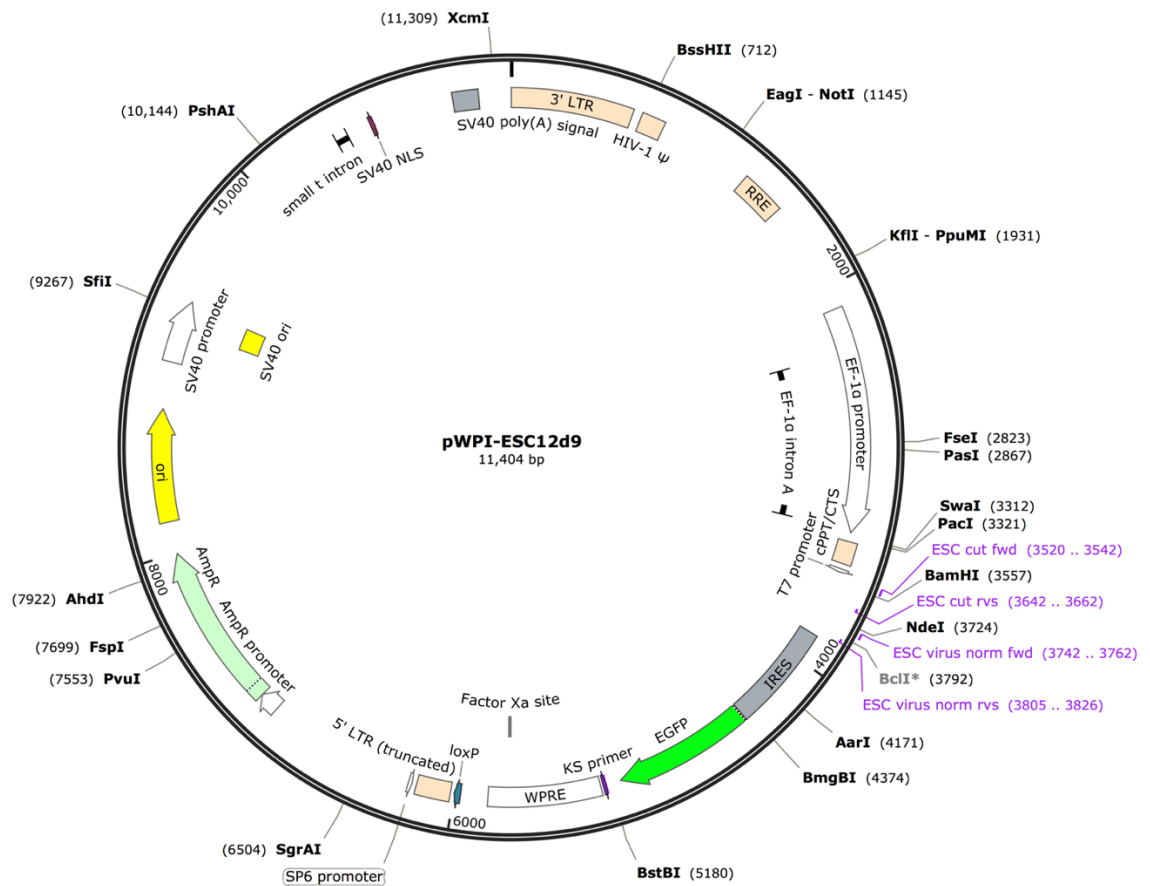


Figure 2.26 Map of pWPI-ESC12d9

The ESC12d9 fragment was amplified from pESC12d9 (not shown, generated by Chris Kirkham) and cloned into pWPI vector by blunt end ligation. All the unique restriction enzyme sites are shown. Ori = origin of replication, AmpR = Ampicillin resistance gene, CMV = human cytomegalovirus promoter, bGH poly(A) signal = bovine growth hormone polyadenylation signal, SV40 = Simian virus 40, EF-1 α intron A = truncated intron upstream of the start codon of elongation factor-1 α , IRES = internal ribosome entry site, EGFP = enhanced GFP, WPRE = woodchuck hepatitis virus post-transcriptional regulatory element. Main primers used are shown in purple.

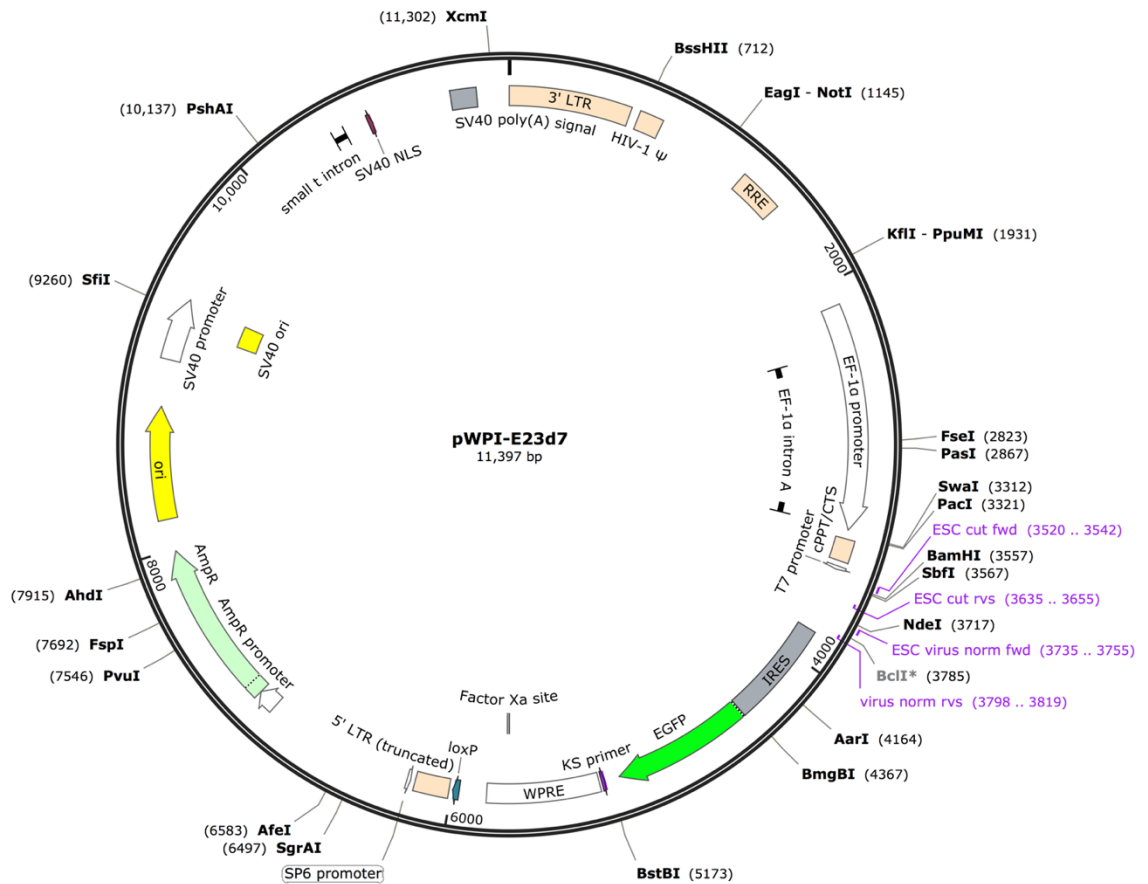


Figure 2.27 Map of pWPI-ESC23d7

The ESC23d7 fragment was amplified from pESC23d7 and cloned into pWPI vector by blunt end ligation. All the unique restriction enzyme sites are shown. Ori = origin of replication, AmpR = Ampicillin resistance gene, CMV = human cytomegalovirus promoter, bGH poly(A) signal = bovine growth hormone polyadenylation signal, SV40 = Simian virus 40, EF-1 α intron A = truncated intron upstream of the start codon of elongation factor-1 α , IRES = internal ribosome entry site, EGFP = enhanced GFP, WPRE = woodchuck hepatitis virus post-transcriptional regulatory element. Main primers used are shown in purple.

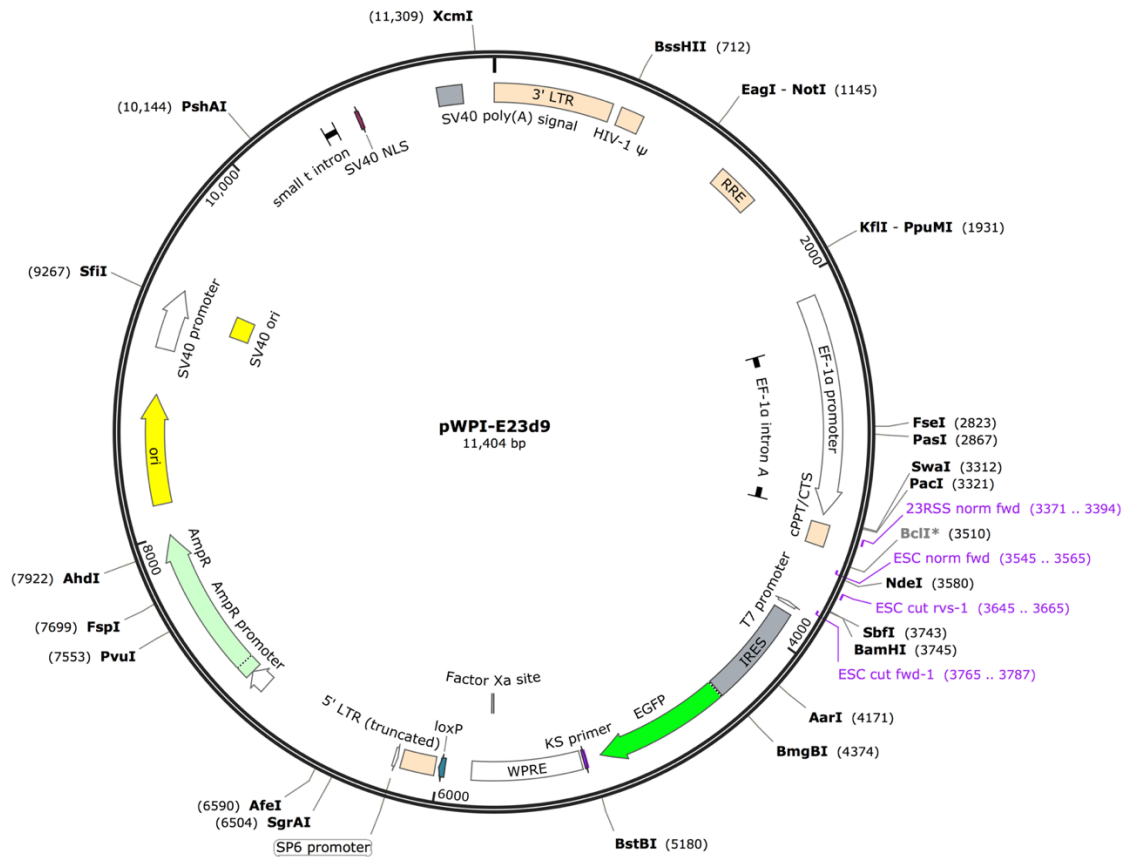


Figure 2.28 Map of pWPI-ESC23d9

The ESC23d9 fragment was amplified from pESC23d9 (not shown, generated by Chris Kirkham) and cloned into pWPI vector by blunt end ligation. All the unique restriction enzyme sites are shown. ori = origin of replication, AmpR = Ampicillin resistance gene, CMV = human cytomegalovirus promoter, bGH poly(A) signal = bovine growth hormone polyadenylation signal, SV40 = Simian virus 40, EF-1 α intron A = truncated intron upstream of the start codon of elongation factor-1 α , IRES = internal ribosome entry site, EGFP = enhanced GFP, WPRE = woodchuck hepatitis virus post-transcriptional regulatory element. Main primers used are shown in purple.

Chapter 3:

Investigation of the molecular basis of asymmetric ESC/RSS cleavage *in vitro*

A) Introduction

V(D)J recombination is prone to making mistakes during the generation of antigen receptor diversity (Hozumi & Tonegawa, 1976; Roth, 2014), by catalysing aberrant DNA rearrangements in developing lymphocytes, which can potentially trigger leukaemias and lymphomas via chromosome translocations, deletions and mutations (Kuppers, 2005; Mullighan et al., 2008; Papaemmanuil et al., 2014; Tsujimoto et al., 1985; Zhang et al., 2012).

During the V(D)J recombination, the signal ends are ligated to form the V(D)J recombination by-product, the ESC, which was previously thought to be inert. However, a variety of studies later demonstrated that the ESC has the potential to trigger genomic instability by different mechanisms (Arnal & Roth, 2007; Ramsden et al., 2010). Firstly, the ESC was found to contribute to genomic instability by being reinserted back into the genome randomly in a non-sequence specific reaction called transposition (Hiom et al., 1998). However, to date, there is only two pieces of evidence to show that this reaction occurs *in vivo* (Elkin et al., 2003; Messier et al., 2003; Messier et al., 2006), since it is almost totally inhibited by the RAG2 C-terminus. Later, several studies suggested that the ESC can reinsert into the genome in a sequence-specific manner at genuine or cryptic RSS in a reaction termed reintegration (Curry et al., 2007; Vanura et al., 2007). Since there are over 10 million cryptic RSSs (Lewis et al., 1997), some of which are located closely to oncogenes (Boehm et al., 1988; Kuppers & Dalla-Favera, 2001), reintegration of ESCs has the potential to be highly dangerous if it occurs next to an oncogene. This is because if an ESC carrying promoters of non-coding transcript is inserted next to an oncogene, these promoters can cause increased oncogene expression.

Notably, Chris Kirkham from the Boyes lab uncovered a new aberrant V(D)J recombination named cut-and-run. In this reaction, an RSS was readily cleaved by RAG proteins when incubated with appropriate RSS partner or ESC partner, whilst there was little cutting of the ESC when incubated with either a 12-RSS or 23-RSS partner i.e. asymmetric cleavage occurred in synaptic complexes between an ESC and RSS substrates. Since the ESC is formed from two RSSs in a head-to-head configuration, it was proposed that the asymmetric cleavage is possibly due to simultaneous RAG binding to both RSSs within the ESC, thus preventing efficient cleavage of the ESC at the heptamer/heptamer boundary (Kirkham, 2014). Moreover, it was also found that after cleavage, the cut RSS is released from the post-cleavage complex, whereas the ESC remains bound by RAG proteins, which suggests that the ESC is potential to trigger cleavage at further RSSs or cryptic RSSs in the genome in a reaction they named “cut-and-run”.

In this chapter, I further investigated the molecular basis of asymmetric cleavage between RSS and ESC substrates. Since all of the *in vitro* RAG cleavage assays by Chris Kirkham were carried out using core RAG proteins and since the RAG2 C-terminus was suggested to inhibit transposition, to investigate if the RAG2 C-terminus also inhibits the cut-and-run reaction, the first part of this chapter describes the expression and purification of core and full-length RAG proteins. I then use these to compare the asymmetric cutting by core RAG1 plus core RAG2 and cutting by core RAG1 plus full-length RAG2. Then, the hypothesis that RAGs bind to both RSSs in the ESC was further investigated by DNase I footprint assays on the ESC. In addition, I performed *in vitro* RAG binding and cutting assays on ESC mutants to further test the hypothesis that RAGs bind to both RSSs of the ESC.

B) Results

3.1 Expression and purification of core and full-length RAG proteins.

In vitro analysis of RAG activity (e.g. in binding and cutting assays) requires purified RAG1 and RAG2 proteins. Numerous previous studies investigating the biochemistry of V(D)J recombination used truncated RAG1 and RAG2 due

to the difficulty in purifying full-length RAG proteins. The minimal regions of RAG proteins for catalysing efficient V(D)J recombination were identified using an *in vivo* recombination assay on extrachromosomal substrates (Cuomo & Oettinger, 1994; Kirch et al., 1996; Sadofsky et al., 1993; Sadofsky et al., 1994) and were frequently used *in vitro* (Dudley et al., 2003). For the RAG1, it was found that, the core region contains amino acids 384-1008 of 1040, including a nonamer binding domain which directly binds to the nonamer of 12- and 23-RSSs (Difilippantonio et al., 1996; Spanopoulou et al., 1996; Yin et al., 2009) and a highly conserved DDE motif which is the major active site and essential for DNA nicking and hairpin formation (Fugmann et al., 2000; Kim et al., 1999; Landree et al., 1999). For RAG2, the core region includes amino acids 1-387 of 527. This excludes the RAG2 C-terminus which has a conserved plant homeodomain (PHD) finger which recognises H3K4me3 and directs RAG binding to active chromatin (Elkin et al., 2005; Matthews et al., 2007; Ramon-Maiques et al., 2007), and an acidic “hinge” region (Matthews et al., 2007; Ramon-Maiques et al., 2007) which is potentially involved in maintaining the stability of the RAG post-cleavage complex (Coussens et al., 2013; West et al., 2005). In addition, Core RAG2 does include six kelch domains that are vital for interaction with RAG1 (Aidinis et al., 2000; Calleebaut & Mornon, 1998).

Both core and full-length RAG proteins were expressed and purified from HEK293T cells. The HEK293T cell line, which expresses the SV40 large T-antigen, is a widely used cell line for eukaryotic expression of proteins since they enable the amplification of transfected plasmids containing the SV40 origin of replication (DuBridge et al., 1987). Core RAG1 and core RAG2 or full-length RAG2 proteins were co-expressed in HEK293T cells using expression plasmids (pEF-MBP-cRAG1 and pEF-MBP-cRAG2 or pEF-MBP-flRAG2) generated by the Boyes lab, since co-expression increases both the yield and activity of the RAG proteins compared to when they are expressed separately (Swanson, 2004). In these vectors, the open reading frames (ORFs) of RAG1 and RAG2 are fused to an N-terminal maltose binding protein (MBP) tag (Figure 3.1A), to facilitate the purification of RAG proteins by amylose affinity chromatography. An example of proteins purified via amylose column

chromatography is given in Figure 3.1B. Similar to previous reports, the yield of RAG2 was consistently lower than that of RAG1, for both core and full-length proteins (McBlane et al., 1995; Kirkham, 2014). In addition, to test if RAG1 and RAG2 simultaneously bind to the ESC, c-Myc tagged cRAG1 and HA tagged cRAG2 plasmids were constructed and co-transfected in HEK293T cells to express and purify tagged RAG proteins (Figure 3.1A and B), for use in supershift assays (see Figure 3.6).

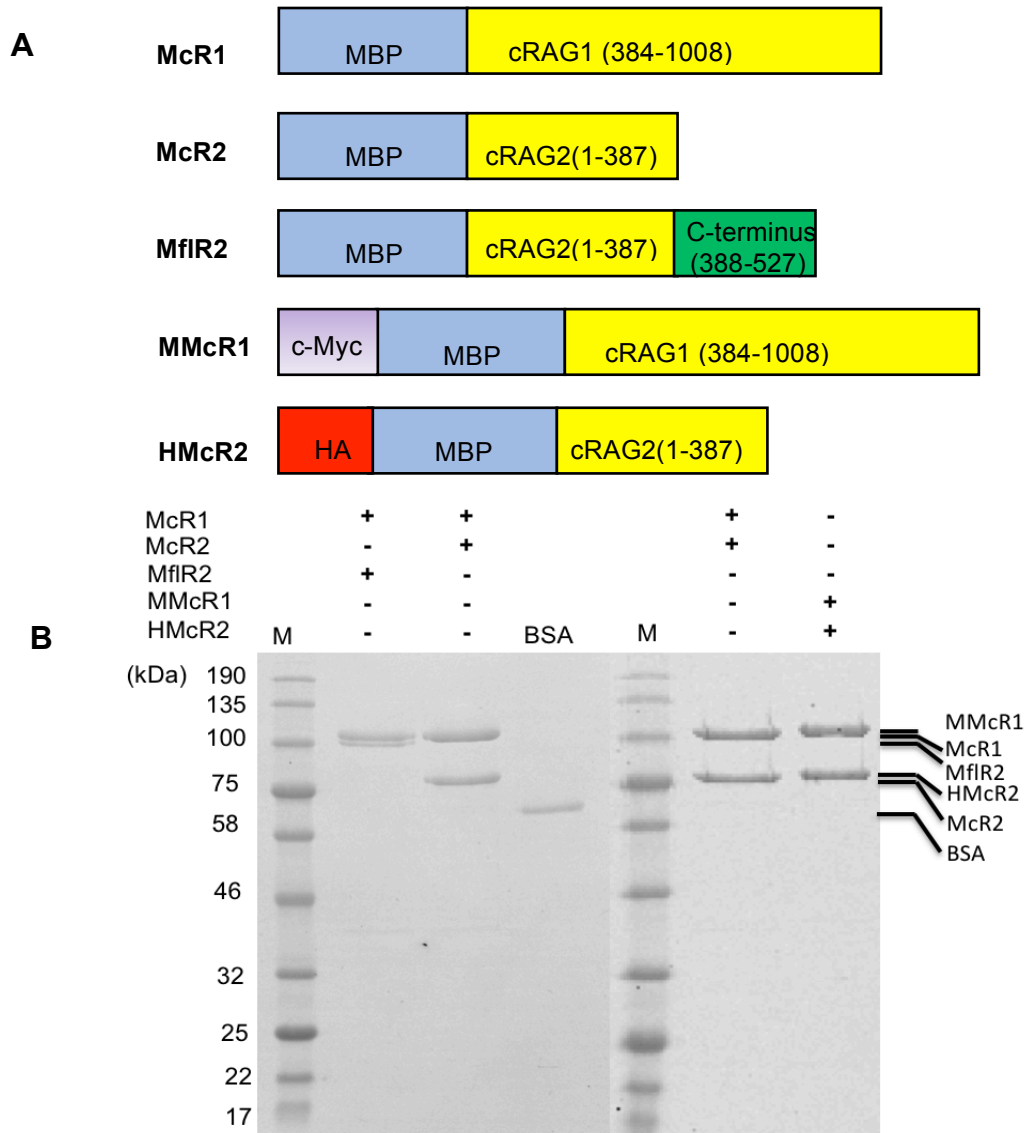


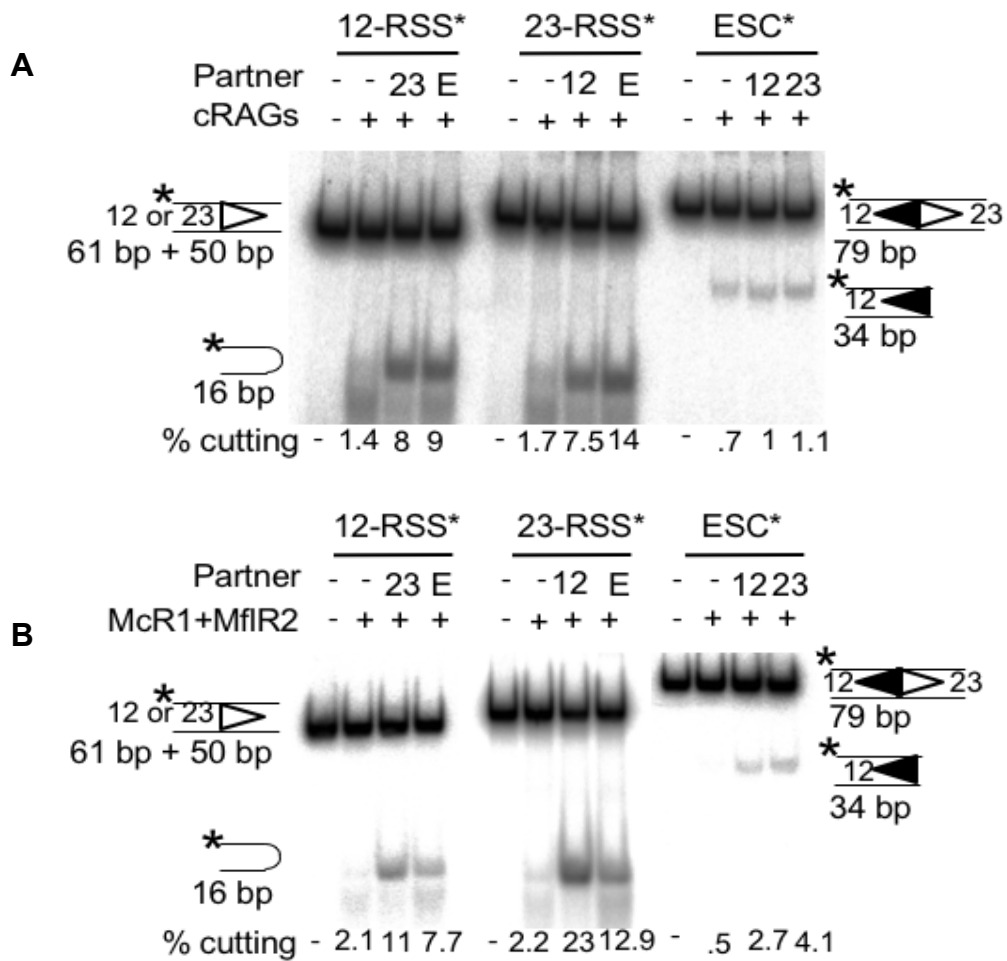
Figure 3.1 RAG proteins purified from HEK293T cells

(A) Schematic of tagged core or full-length RAG proteins. McR1 = MBP-tagged core RAG1; McR2 = MBP-tagged core RAG2; MflR2 = MBP-tagged full-length RAG2; MMcR1 = cMyc-MBP-tagged core RAG1; HMcR2 = HA-MBP-tagged core RAG2. **(B)** Purified tagged proteins. 400 ng BSA was used to normalize the concentration of purified RAG proteins. M: Protein Marker (NEB P7706); McR1 = 111 kDa; McR2 = 85 kDa; MflR2 = 102 kDa; MMcR1 = 112.2 kDa; HMcR2 = 86.1 kDa.

3.2 Comparison of asymmetric cutting by core RAG1 plus core RAG2 proteins and core RAG1 plus full-length RAG2 proteins

V(D)J recombination obeys the 12/23 rule whereby coupled cleavage occurs only between a 12-RSS and a 23-RSS (Tonegawa et al., 1983). The ESC consists of a 12-RSS and a 23-RSS in a head-to-head configuration. Previous experiments in the Boyes lab have shown that there is an asymmetric cutting when an ESC oligonucleotide is paired with an oligonucleotide carrying either a 12- or 23-RSS i.e. the RSS is cut well whilst the ESC is barely cut. By contrast, when an RSS oligonucleotide is incubated with corresponding RSS partner, the cutting at two RSSs is equivalent, with 8 - 9% cutting for 12-RSS, 7.5 - 11% cutting for 23-RSS, but only 1 - 1.1% cutting for ESC (Figure 3.2A). However, in this assay, core RAG1 and core RAG2 were used to initiate cutting at the DNA substrates. The RAG2 C-terminus has been shown to inhibit transposition and reintegration reactions *in vivo* (Curry et al., 2007; Elkin et al., 2003). Therefore, to test if RAG2 C terminus also influences asymmetric cleavage in the RSS/ESC synaptic complex, an *in vitro* cleavage assay with core RAG1 (McR1) plus full-length RAG2 (MflR2) proteins was carried out. As shown in Figure 3.2B, when a radioactively labelled 12-RSS was incubated with an unlabelled 23-RSS or ESC partner, 7.7 - 11% 12-RSS was cleaved by core RAG1 and full-length RAG2 after incubation for 15 minutes at 37 °C (Figure 3.2B, lanes 3 and 4). Likewise, when a radioactively labelled 23-RSS was paired with an unlabeled 12-RSS or ESC, the 23-RSS was also cleaved efficiently, with 13 - 23% of the substrates being digested (Figure 3.2B, lanes 7 and 8). In contrast, when a radioactively labelled ESC was incubated with a 12- or 23-RSS partner, the cleavage efficiency was significantly lower at 2.7 - 4.1%, which is just a little higher than cleavage of a 12- and 23-RSS in the absence of partner RSS (Figure 3.2B, lanes 2, 5, 11 and 12). These data suggest that asymmetric cleavage also occurs with full-length RAG2 and thus this property of RAG cutting is not totally inhibited by the RAG2 C terminus. Interestingly, it was found that, whilst the cleavage efficiency for a 12-RSS is similar when incubated with core RAG1 plus core RAG2 compared to core RAG1 plus full-length RAG2 (Figure 3.2A lanes 3 and 4, Figure 3.2B lanes 3 and 4), the percent cutting of a 23-RSS with core RAG1 plus full-length RAG2

is almost twice as much as that observed with core RAG1 plus core RAG2 (Figure 3.2A lanes 7 and 8, Figure 3.12B lanes 7 and 8), possibly implying that the RAG2 C-terminus has an effect on 23-RSS binding or nicking.



To examine the role of non-core regions, I also tried to co-express and purify full-length RAG1 plus full-length RAG2 from HEK293T cells (Figure 3.3). However, unfortunately, the activity was very low. These were used in an *in vitro* RAG cleavage assay but asymmetric cleavage was not observed, which although unexpected, is consistent with previous data (Kirkham, 2014). It is also notable that the cleavage efficiency of the labelled substrate is similar to or even higher than that when no partner was added (Figure 3.4). The reason for this low activity possibly is due to the purification method used, which is further discussed in the discussion of this chapter. In addition, full-length RAG proteins are easily degraded during purification (Figure 3.3B). However, such low activity means that unfortunately, these data are not fully conclusive.

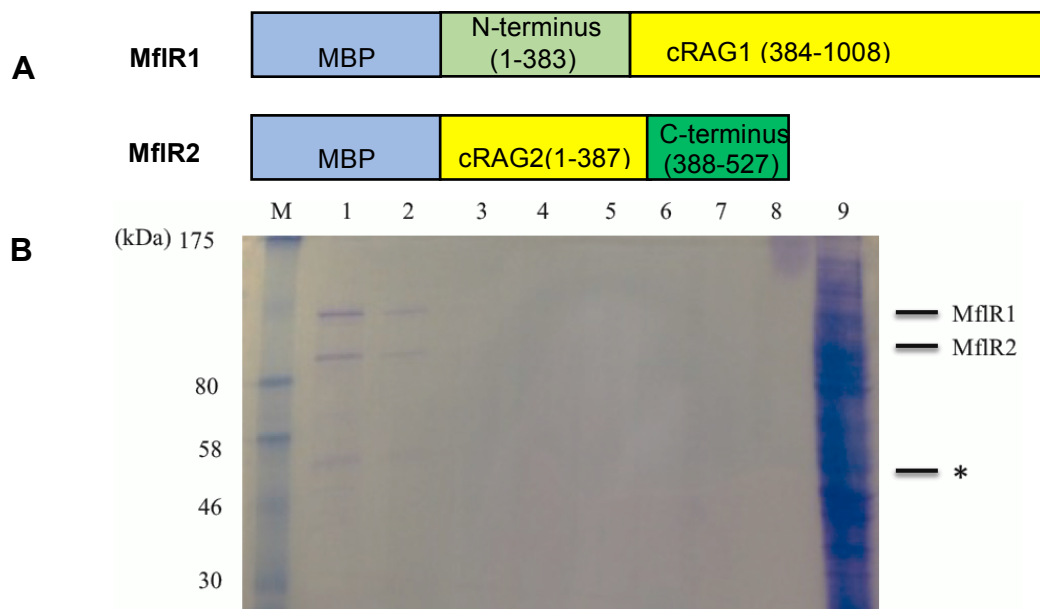


Figure 3.3 Purification of full-length RAG proteins

(A) Schematic of MBP tagged full-length RAG proteins. MfIR1 = MBP-tagged full-length RAG1; MfIR2 = MBP-tagged full-length RAG2. **(B)** MBP-tagged flRAG proteins purified from transfected HEK-293T cells. M: Protein Marker (NEB P7708s), Lane 1-8 are different fraction eluted from an amylose column, lane 9 is the flowthrough. MfIR1 = 162 kDa, MfIR2 = 102 kDa, * = degradation product.

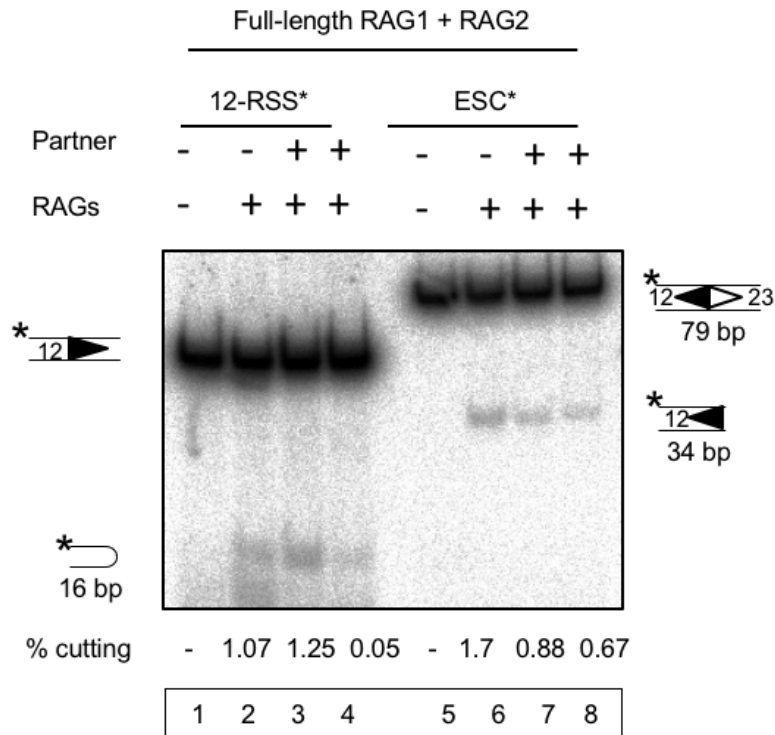


Figure 3.4 *In vitro* cutting with full-length RAG1 plus full-length RAG2

The radioactively labelled oligonucleotides are indicated by an asterisk above each set of lanes. Radiolabelled 12-RSS or ESC were incubated with an unlabelled partner as indicated, HMGB1 and full-length RAG1 plus full-length RAG2 proteins. Following cutting, the products were separated on a 12% native polyacrylamide gel and the level of cleavage was determined by quantification of the intensity of cleaved substrates against the intensity of total substrate.

3.3 Asymmetric cleavage is due to RAG proteins binding to both RSSs within the ESC

Asymmetric RSS/ESC cleavage is proposed to be due to RAG proteins simultaneously binding to both RSSs within the ESC, preventing the cutting at heptamer/heptamer boundary. To test this idea, a DNase I footprint assay was carried out using purified core RAG proteins and a 258 bp radioactively labelled DNA fragment containing the wild-type or mutated ESC sequence. If the radioactively labelled DNA is bound by proteins, it will be protected from DNase I digestion and regions of enhanced cutting and protected regions will be observed compared to naked DNA. As expected, after binding RAG proteins and digestion with DNase I, regions of enhanced cutting, indicated by stars, and protected regions, indicated by bars, were observed (Figure 3.5). There is also a low level of enhanced cutting observed at heptamer/heptamer

boundary (Figure 3.5). This appears to be due to cutting by the RAG proteins themselves, not DNase I. This was shown by a control experiment in which the radioactively labelled ESC was incubated with RAG proteins but where DNase I was not added. The increased intensity at the heptamer/heptamer boundary was also observed (data not shown).

It was demonstrated that RAG cleavage at RSSs can be influenced by the flanking coding gene sequences (Sadofsky et al., 1995). Within ESC, both RSSs work as a flanking sequences for each other. When RAG proteins bind to both RSSs within the ESC, the RAGs might bend the DNA to expose the heptamer/heptamer boundary to some extent, allowing a single strand DNA nicking at this site. It is notable that in the DNase I footprint, the protected blank regions are equivalent on both RSSs in the ESC (Figure 3.5). In addition, if one RSS in the ESC was mutated, the RSS cannot be protected by RAGs completely, and if both RSSs in the ESC were mutated, this ESC mutant cannot be protected by RAGs at all (Figure 3.5). These results suggest that RAG proteins can bind to both RSSs simultaneously, blocking most of the cleavage at heptamer/heptamer boundary, which is consistent with previous data (Kirkham, 2014) and the cut-and-run hypothesis.

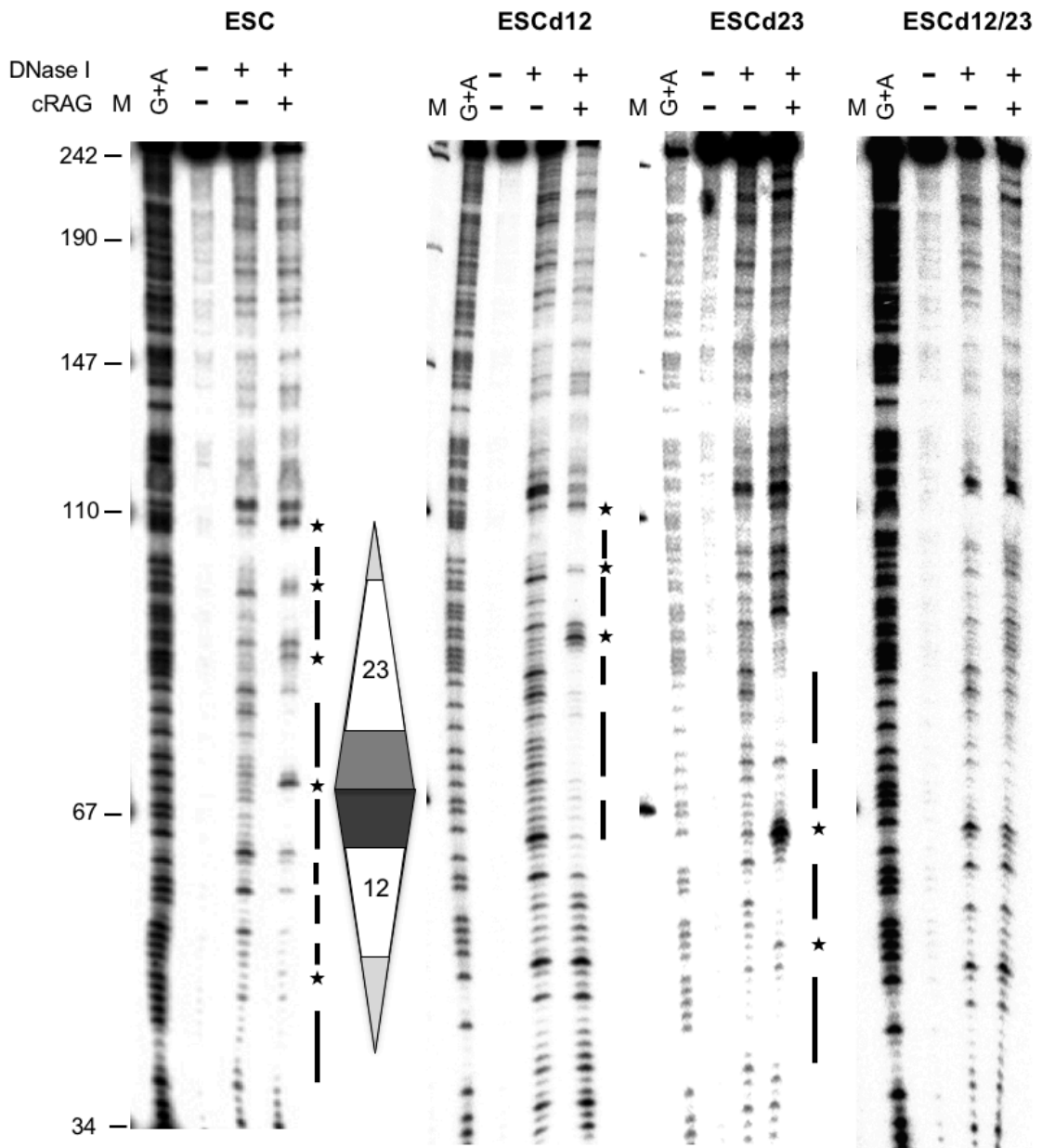


Figure 3.5 Asymmetric cleavage is due to RAG proteins binding to both RSSs within the ESC

DNase I footprint on ESC and ESC mutants. Following binding of RAG proteins to DNA in the presence of Ca^{2+} , DNase I was added for 1 minute at room temperature. Reactions were purified and separated on a 6% denaturing polyacrylamide gel. M shows the DNA markers (pUC18 digested with HpaII), the size of which are given at the side of the gel. G+A shows the marker generated by Maxam and Gilbert sequencing. The position of the RSSs, as determined from the G+A marker ladder is shown to the right. Stars indicate enhanced DNase I cutting in the presence of RAGs. Bars indicate protected regions. ESCd12 means the heptamer and the nonamer of 12-RSS in the ESC are mutated. ESCd23 means the heptamer and the nonamer of 23-RSS in the ESC are mutated. ESCd12/23 means the heptamer and the nonamer of both RSSs in the ESC are mutated.

3.4 Both RAG1 and RAG2 are present in the ESC/RAG complex

Previous studies have shown upon RSS binding by RAG proteins, two complexes are formed, termed SC1 and SC2. Upon addition of HMGB1, these are shifted to HSC1 and HSC2, respectively (Swanson, 2002). HSC1 contains two molecules of RAG1 and one molecule of RAG2 whereas HSC2 has two molecules of each RAG protein (Swanson, 2002). In addition, a study by Chris Kirkham found that RAG proteins form an additional higher molecular weight complex with a lower mobility than HSC1 and HSC2 when binding to ESC substrates (Kirkham, 2014). From the results of the DNase I footprint assay, it would be expected that the slower migrating complex results from RAG binding to both RSSs within the ESC.

To test if both RAG1 and RAG2 are indeed present in this higher molecular weight complex, a supershift assay was carried out using Myc-tagged core RAG1, HA-tagged core RAG2 and a labelled ESC. Consistent with previous studies, in this assay, upon RAG binding to a 12-RSS or an ESC, two and three complexes were observed for 12-RSS/RAG binding and ESC/RAG binding, respectively, in the presence of 23-RSS partner (Figure 3.6, lanes 1 and 4; indicated by red arrows). Upon addition of anti-Myc or anti-HA antibodies, these complexes are shifted to much higher molecular complexes (Figure 3.6, lanes 2, 3, 5 and 6; indicated by green arrows). Specifically, consistent with previous studies, both HSC1 and HSC2 are supershifted after addition of anti-Myc antibody, indicating that RAG1 is present in both complexes (Figure 3.6, lanes 2 and 5). Likewise, HSC2 is supershifted to a similar position upon the addition of anti-HA antibody to that when the anti-Myc antibody is added, whilst HSC1 supershifted to a lower position upon addition of anti-HA compared to the addition of anti-Myc, with some unshifted HSC1 remaining (Figure 3.6, lanes 3 and 6).

Importantly, the higher molecular weight complex formed with the ESC is supershifted to a much higher position upon addition either anti-Myc or anti-HA antibody, suggesting that this complex contains both RAG1 and RAG2. However, it is still not known exactly how many molecules of each RAG

subunit are present in this higher molecular weight ESC/RAG complex. Theoretically, since RAG proteins bind to RSS in a RAG1-RAG2 heterodimer (Swanson, 2002) and ESC contains two RSSs, four molecules of RAG1 and four molecules of RAG2 could be present in this complex.

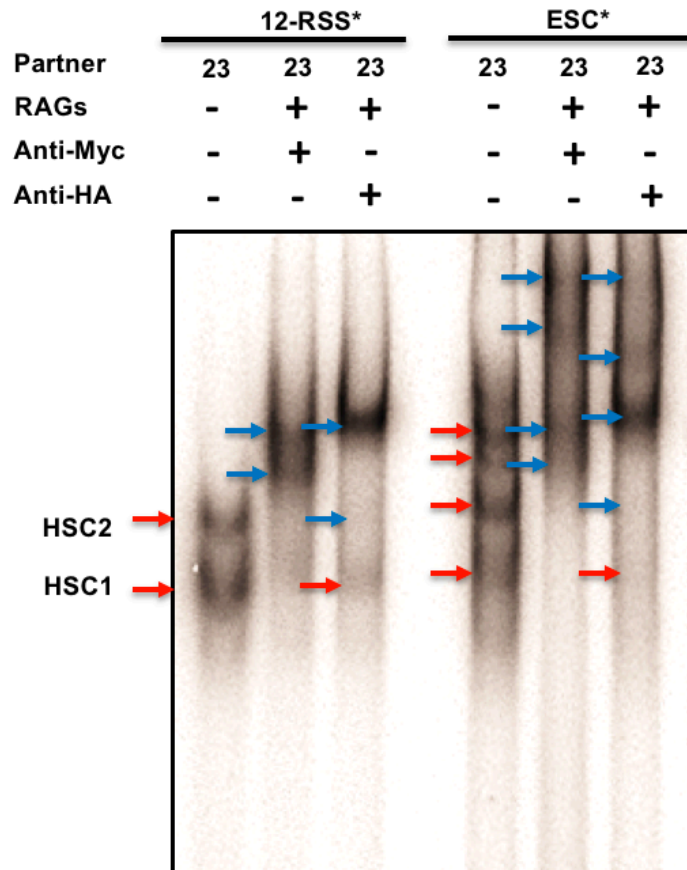


Figure 3.6 Both RAG1 and RAG2 are present in the ESC/RAG complex

The radioactively labelled 12-RSS and ESC oligonucleotides are indicated by an asterisk above each set of lanes. A radioactively labelled 12-RSS or ESC was incubated with an unlabelled 23-RSS partner, HMGB1, Myc-tagged core RAG1 and HA-tagged core RAG2 proteins at 37°C for 15 minutes. Following binding, anti-Myc or anti-HA antibodies were added into the sample to incubated on ice for 10 minutes. The samples were then loaded onto a 4% polyacrylamide gel. Non-supershifted bands are indicated by the red arrows and shifted bands are indicated by blue arrows.

3.5 Mutating one side of the ESC restores symmetric cleavage within an ESC/RSS synaptic complex

If RAG cutting of the ESC is indeed blocked by RAG binding to both RSSs, then we would predict that mutating one RSS in the ESC should allow the ESC

to be cleaved. Therefore, different ESC mutants were used in *in vitro* cutting assays with core RAG1 plus core RAG2 proteins to determine if mutation of one side of RSS within the ESC can restore symmetric cleavage within the ESC/RSS synaptic complex.

Firstly, ESC12d9, in which the 12-RSS nonamer within ESC is mutated, was used. Since the nonamer is the primary DNA binding domain for RAG proteins, the mutation of nonamer would be expected to show decreased even no RAG binding to 12-RSS within the ESC, thereby potentially allowing ESC cleavage. Moreover, ESC12d9 contains a mutated 12-RSS and a wild-type 23-RSS, and therefore will pair well with a functional 12-RSS partner but not with a 23-RSS partner. Therefore, ESC12d9 is expected to be efficiently cut when incubated with a 12-RSS partner and barely cut when incubated with a 23-RSS partner.

As expected, the results showed that when a radioactively labelled 12-RSS was incubated with a 23-RSS partner or an ESC12d9 partner, the level of cutting was almost the same, at 9.4% and 9.1%, respectively (Figure 3.7, lanes 3 and 4). Likewise, when the labelled 23-RSS was incubated with a 12-RSS, cutting occurred efficiently (10%; Figure 3.7, lane 8). However, when a labelled 23-RSS was incubated with ESC12d9, the percentage of cutting is low (2.6%; Figure 3.7, lane 9). Conversely, the consensus 23-RSS within ESC12d9 can pair well with a 12-RSS partner, giving a 5 % cutting, but the mutated 12-RSS within ESC12d9 does not pair with a 23-RSS partner, leading to a lower percentage of cutting, of 1.3% (Figure 3.7, lanes 13 and 14). Since ESC12d9 is cut efficiently when incubated with an unlabelled 12-RSS and at almost the same level as when a 12-RSS is incubated with an ESC, these data suggest that the mutation of 12-RSS nonamer in the ESC almost fully restores symmetry to 12-RSS/ESC cutting, which is consistent with the idea that mutating one RSS in ESC restores symmetric cleavage within an ESC/RSS synaptic complex.

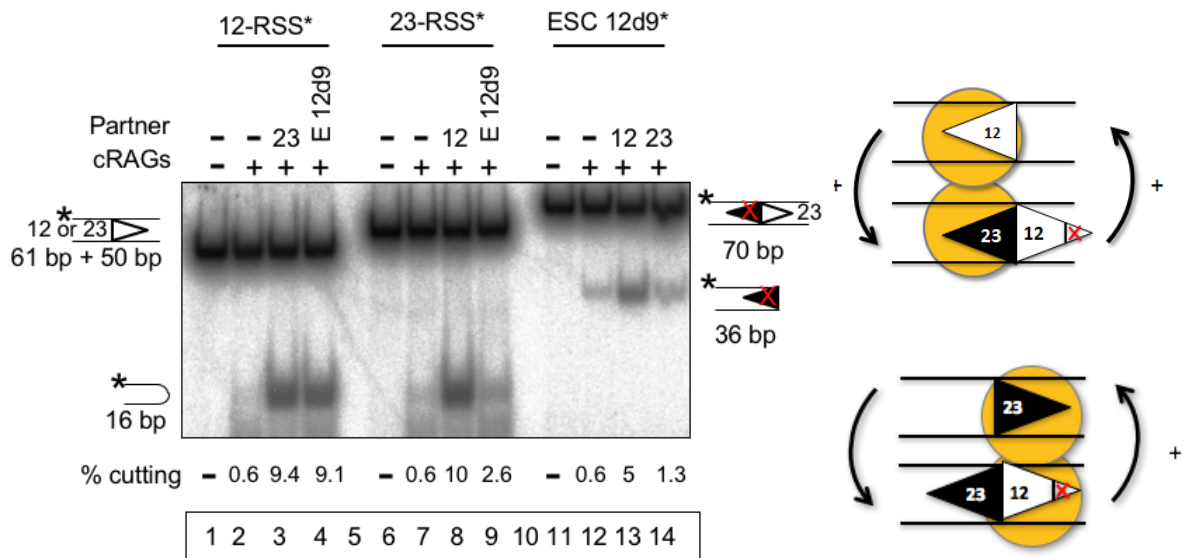


Figure 3.7 Mutation of the 12-RSS nonamer in the ESC restores symmetric RSS/ESC cutting

The radioactively labelled oligonucleotides are indicated by an asterisk above each set of lanes. A radiolabelled 12-RSS, 23-RSS or ESC12d9 was incubated with an unlabelled partner as indicated, as well as HMGB1 and core RAG proteins. Following cutting, the products were separated on a 12% native polyacrylamide gel and the level of cleavage was determined by quantification of the intensity of cleaved substrates against the intensity of total substrates.

The first three nucleotides of the heptamer 'CAC' are critical for RAG cutting and show the highest level of sequence conservation (Hesse et al., 1989; Ramsden et al., 1996). This therefore raises the question: If just part of the heptamer in the ESC is mutated, does this also restore symmetric cutting? To test this, the effect of mutation of the CAC in the 23-RSS heptamer within the ESC (ESC23dCAC) was next investigated. In this case, the ESC retains a consensus 12-RSS, which is expected to pair with a consensus 23-RSS. Consistent with this, addition of cold ESC23dCAC to labelled 23-RSS gave a high level of cutting (11.3%; Figure 3.8, lane 9). Likewise, in the converse pairing i.e. labelled ESC23dCAC plus cold 23-RSS, strong cutting of the ESC was also observed (Figure 3.8, lane 14), indicating symmetric cleavage has been restored. Also, as might be expected with a mutated 23-RSS in the ESC, we see poor cutting when ESC23dCAC was paired with 12-RSS. When a labelled 12-RSS was incubated with a cold ESC23dCAC, or the converse pairing, the cutting percentage was significantly reduced, to 1.2% and 0.7%, respectively (Figure 3.8, lane 4 and 13). These data suggest that mutation of

the 23-RSS CAC within ESC prevents efficient cutting when paired with the 12-RSS whilst pairing between a 23-RSS and the 12-RSS within ESC23dCAC proceeds normally, implying that RAG binding to the heptamer/heptamer boundary can be disrupted by just mutation of the critical bases of the heptamer in the ESC, thus restoring symmetric RSS/ESC cutting to some extent.

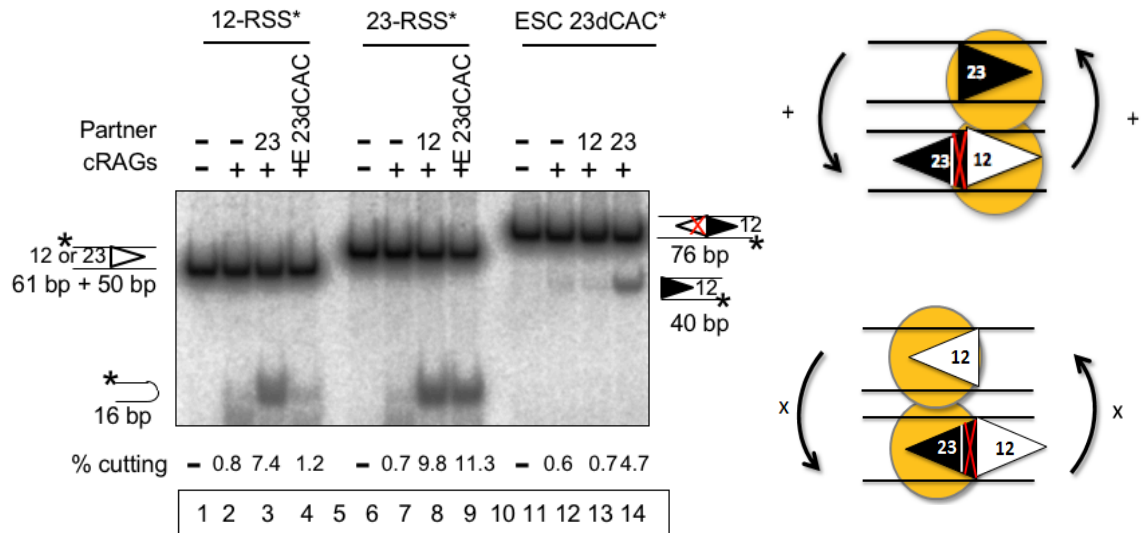


Figure 3.8 Mutation of 23-RSS CAC in the ESC restores symmetry to 23-RSS/ESC cutting

The radioactively labelled oligonucleotides are indicated by an asterisk above each set of lanes. Radiolabelled 12-RSS, 23-RSS or ESC23dCAC was incubated with an unlabelled partner as indicated, together with HMGB1 and core RAG proteins. Following cutting, the products were separated on a 12% native polyacrylamide gel and the level of cleavage was determined by quantification of the intensity of cleaved substrates against the intensity of total substrates.

Next, the effect of mutation of CAC in the 12-RSS heptamer within ESC (ESC12dCAC) was investigated. In this case, a consensus 23-RSS is retained within ESC, which is expected to pair with a consensus 12-RSS. Therefore, efficient cutting is expected when it is incubated with a 12-RSS but not a 23-RSS. However, surprisingly, the results showed that addition of either an unlabeled 12-RSS or an unlabelled 23-RSS to the labelled ESC12dCAC did not result in any cutting whatsoever (Figure 3.9, lanes 13 and 14). Conversely, the reverse pairing, cutting occurred as expected i.e. cold ESC12dCAC plus labelled 12-RSS or labelled 23-RSS (Figure 3.9, lanes 4 and 9), with cutting of

the 12-RSS being more efficient than the 23-RSS, as expected. These data suggest that mutation of 12-RSS CAC within ESC does not restore symmetric cutting at RSS/ESC. Because the primary RSS/RAG interactions take place between the RAG1 nonamer binding domain and the RSS nonamer (Ru et al., 2015; Yin et al., 2009), deleting the CAC of the 12-RSS in the ESC might not completely prevent RAG binding. However, this could not explain why mutation of 23-RSS CAC restores RSS/ESC symmetric cleavage. A potential explanation of this is given, however, by examining RAG/ESC complexes (Figure 3.6).

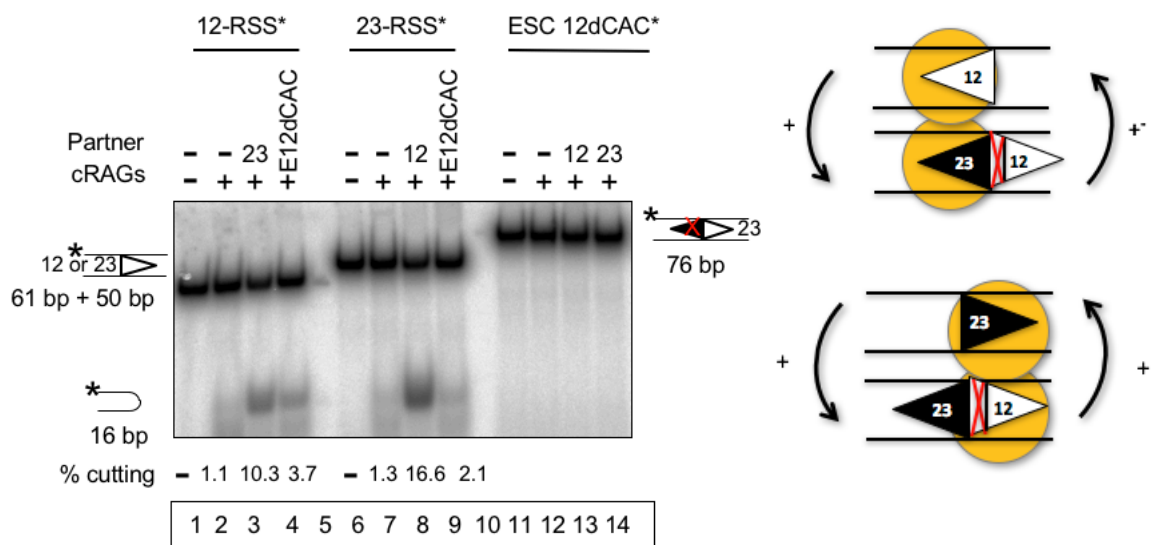


Figure 3.9 Mutation of 12-RSS CAC in the ESC does not restore symmetric RSS/ESC cutting

The radioactively labelled oligonucleotides are indicated by an asterisk above each set of lanes. A radiolabelled 12-RSS, 23-RSS or ESC12dCAC was incubated with an unlabelled partner as indicated, together with HMGB1 and core RAG proteins. Following cutting, the products were separated on a 12% native polyacrylamide gel and the level of cleavage was determined by quantification of the intensity of cleaved substrates against the intensity of total substrates.

Overall, these data support the hypothesis that RAG proteins bind simultaneously to both sides of the ESC, leading to asymmetric cleavage.

3.6 Increasing the spacer between two RSSs within the ESC causes ESC cleavage

It was documented that RAG contacts extend 12 bp into the flanking coding gene segment in a single RSS/RAG complex and 16 bp in a paired RSSs/RAG complex (Nagawa et al., 2002; Nagawa et al., 2004). Within an ESC, the second RSS forms the flanking coding gene segment of the first RSS, therefore the extended RAG binding into the adjacent heptamer can explain why ESC cutting is blocked upon RAG binding to both RSSs. If this is the case, we would predict that cleavage of ESC will be possible if the spacer between the two RSSs within the ESC is increased sufficiently. To test this idea, the oligonucleotide, ESC15-spacer, in which 15 nucleotides were inserted between the two RSSs within the ESC, was radio-labelled and used in an *in vitro* RAG cutting assay. If both RSSs within the ESC15 spacer are cleaved, we would expect two bands rather than just one band as is observed with the normal ESC. Since the coding end of 12-RSS within ESC15-spacer was radio-labelled, when the 23-RSS was added as a partner, cutting occurs at the 12-RSS of ESC15-spacer and a 34 bp fragment is expected. When the 12-RSS is added as a partner, cutting occurs at the 23-RSS within ESC15-spacer and a 49 bp fragment is expected. Moreover, since each RSS within the ESC15 spacer can act as a partner for the other, it is also possible that we will get cutting due to self-pairing.

What we actually observed was two bands after RAGs cutting at ESC15-spacer upon incubation of ESC15-spacer with RAGs even in the absence of partner (Figure 3.10, lane 6). This indicates that both RSSs within the ESC15 spacer can function as a partner for each other without the addition of a partner RSS. Furthermore, increased cleavage of 12- or 23-RSSs in the ESC15-spacer was observed when a 23- or 12-RSSs partner was added compared to cutting of ESC15-spacer alone (Figure 3.10, lanes 6-8). Notably, the cleavage at ESC15-spacer is greater than the normal ESC, which is consistent with the idea that cleavage at heptamer/heptamer boundary in the ESC is normally blocked due to RAG binding to both RSSs in the ESC.

Increasing the spacer between the two RSSs within the ESC can release the cleavage site to some extent, restoring the cutting at the ESC.

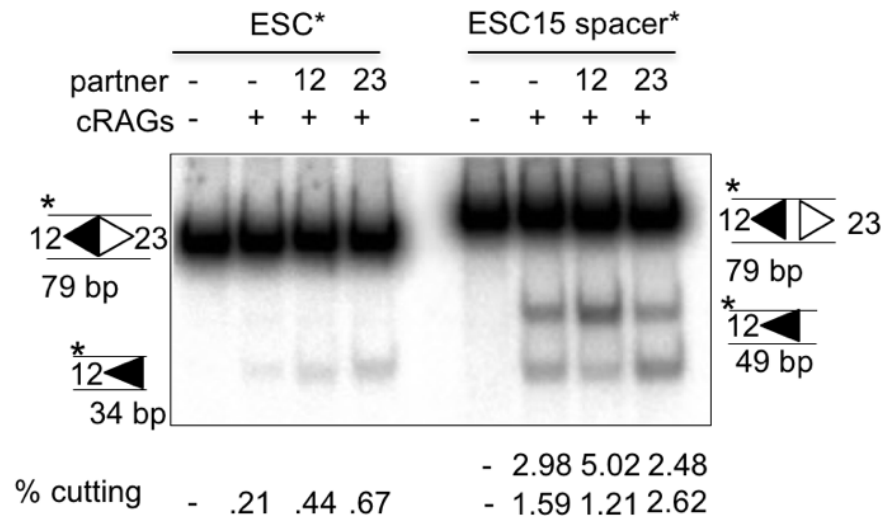


Figure 3.10 Increasing the spacer between the two RSSs within the ESC allows cleavage of the ESC

ESC and ESC15-spacer *in vitro* RAG cutting assay. The radioactively labelled oligonucleotides are indicated by an asterisk above each set of lanes. Radio-labelled ESC or ESC15-spacer were incubated with an unlabeled partner as indicated. HMGB1 proteins were expressed by myself and purified by Joan Boyes and core proteins were purified by Chris Kirkham. Following cutting, the products were separated on a 12% native polyacrylamide gel and the level of cleavage was determined by quantification of the intensity of cleaved substrates against the intensity of total substrates.

3.7 Analysis of RAG/mutant ESC complexes

The mutagenesis experiments suggest that RAG binding to both RSSs in a wild-type ESC blocks its cleavage. To further investigate the molecular basis of this, band-shift experiments were performed. Initially, bandshifts were performed with a labelled 12-RSS or ESC. Consistent with previous data (Kirkham, 2014) that suggest the RAG proteins form an additional complex with the ESC, two bands are observed with 12-RSS whereas three bands, including a higher molecular complex are visible with ESC. If RAG binding to both sides of RSSs within the ESC simultaneously is represented by the higher molecular weight complex, then it would be expected that mutating one of the RSSs should reduce the abundance of this complex or entirely prevent its formation. Consistent with this idea, deletion of the nonamer of the 12-RSS

within the ESC reduces the abundance of high molecular weight complex relative to the other RAG complexes (HSC1 and HSC2) (Figure 3.11, lane 11). Interestingly, mutation of 12-RSS CAC within ESC does not reduce the abundance of the higher molecular weight complex (Figure 3.12, lane 3). This suggests that mutation of the 12-RSS CAC does not eliminate RAG binding to both RSSs in the ESC, which is consistent with the result that mutation of 12-RSS CAC in the ESC does not restore symmetry to RSS/ESC cutting. However, mutation of the CAC within the 23-RSS does reduce the abundance of high molecular weight complex relative to the other RAG complexes (Figure 3.12, lane 7) i.e. ESC23dCAC can eliminate RAG binding, which is consistent with cutting data where symmetric cleavage within an RSS/ESC complex is restored (Figure 3.12, lanes 3 and 7).

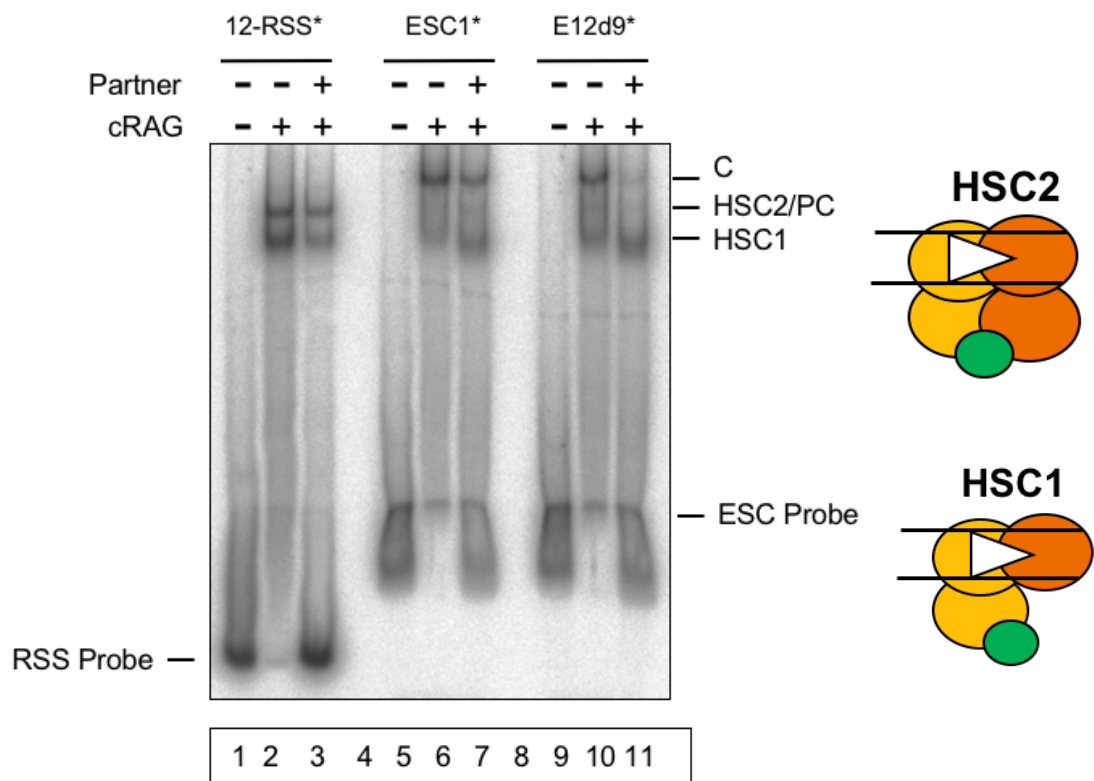


Figure 3.11 Mutation of the 12-RSS nonamer in the ESC eliminates RAG binding

To compare the RAG complexes that form with different substrates, radiolabelled 12-RSS, ESC1 or ESC12d9 oligonucleotides were incubated in the presence of cRAG proteins, HMGB1 and unlabelled partner substrate for 10 min at 37°C, and run on a 4% polyacrylamide gel. C = higher molecular weight RAG complex formed with ESC, HSC1/2 = HMGB1 single complex 1/2, PC = paired complex, yellow ovals indicate RAG1, orange ovals indicate RAG2, blue ovals indicate HMGB1 and triangles indicate RSSs.

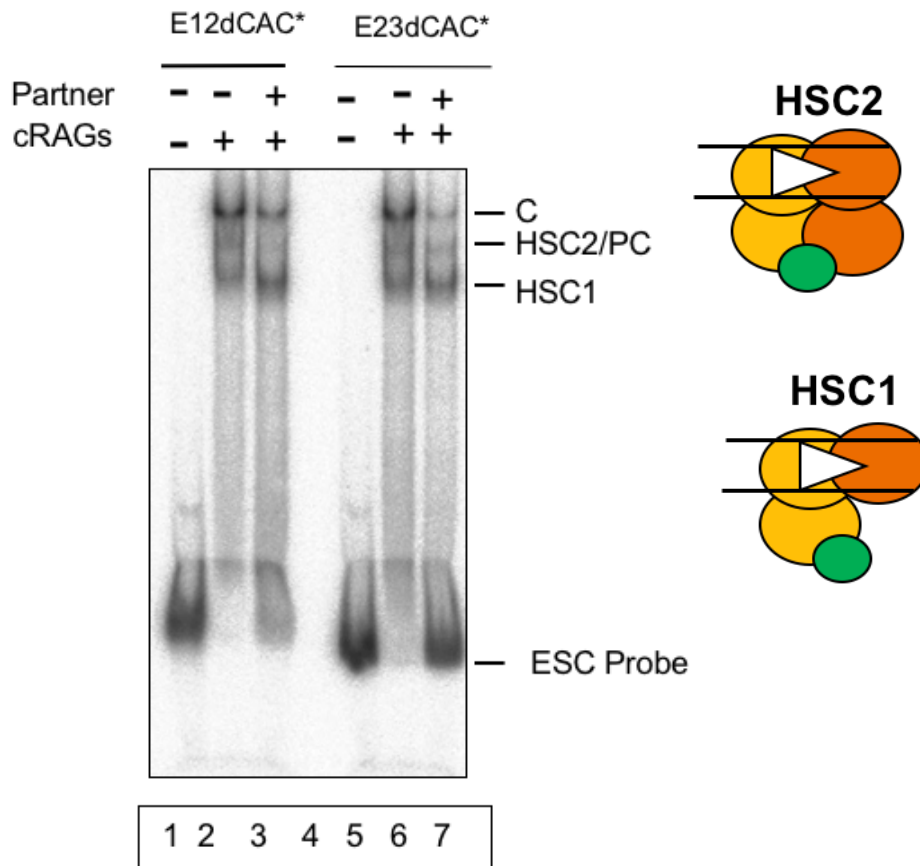


Figure 3.12 Mutation of 12-RSS CAC in the ESC does not eliminate RAG binding but mutation of 23-RSS CAC does decrease RAG binding

To compare the RAG complexes that form with different substrates, radiolabelled 12-RSS, ESC1, ESC12dCAC or ESC23dCAC oligonucleotides were incubated in the presence of cRAG proteins, HMGB1 and unlabelled partner substrate for 10 min at 37 °C, and run on a 4% polyacrylamide gel. C = higher molecular weight RAG complex formed with ESC, HSC1/2 = HMGB1 single complex 1/2, PC = paired complex, yellow ovals indicate RAG1, orange ovals indicate RAG2, blue ovals indicate HMGB1 and triangles indicate RSSs.

C) Discussion

Cut-and-run is a new aberrant V(D)J recombination reaction which was proposed to result in genomic instability (Kirkham, 2014). In this reaction, asymmetric cleavage between an RSS and an ESC occurs i.e. the RSS is cut whilst the ESC is barely cut. This asymmetric RSS/ESC cleavage was proposed to be due to RAG proteins simultaneously binding to both RSSs in the ESC and then blocking its cutting. In this chapter, further evidence was provided to support this idea using *in vitro* cleavage assays with core or full-length RAG proteins as well as DNase I footprint assays.

3.8 Core and full-length RAG protein purification

RAG proteins are vital for the cleavage step of V(D)J recombination (Oettinger et al., 1990; Schatz et al., 1989) and *in vitro* cleavage assays require purified RAG proteins. Various previous studies identified the minimal core regions of RAG1 and RAG2 that are necessary for recombination of extrachromosomal substrates *in vivo* by transfection of numerous truncated RAG1 and RAG2 constructs (Cuomo & Oettinger, 1994; Kirch et al., 1996; Sadofsky et al., 1993; Sadofsky et al., 1994; Silver et al., 1993). Since purified full-length proteins were found to be insoluble and/or inactive, numerous biochemical studies of V(D)J recombination used the purified core RAG proteins (Elkin et al., 2003; Gellert, 2002). Core RAG1 and core RAG2 proteins have been purified from different systems: insect cells via a baculovirus expression system or HeLa cells via a vaccinia virus expression system (McBlane et al., 1995; Kirkham, 2014), or HEK293T cells with an expression vector transfection system (Bergeron et al., 2006; Grundy et al., 2009; Kirkham, 2014; Sawchuk et al., 1997; Swanson & Desiderio, 1999). In addition, it was noted that core RAG proteins purified from HEK293T cells are more active than those purified from insect cells (Kirkham, 2014). Furthermore, a study by Chris Kirkham point out that the full-length RAG proteins purified using a vaccinia/HeLa-based system is insoluble whilst proteins expressed in HEK293T cells via an expression vector/transfection system can yield soluble full-length RAG proteins (Kirkham, 2014). Therefore, in these studies, I used HEK293T cells to co-express and

purify core RAG1 plus core RAG2 or full-length RAG1 plus full-length RAG2. Unfortunately, the concentration of purified full-length RAG1 plus full-length RAG2 is much lower than that of core RAG1 plus core RAG2. In addition, when they were used for *in vitro* cleavage assays, they showed substantially low activity and displayed no asymmetric RSS/ESC cleavage which was consistently observed with core RAG proteins. The lack of asymmetric cleavage is, however, consistent with a previous study (Kirkham, 2014). This extremely low activity of full-length RAG proteins is possibly due to the low concentration of full-length RAG proteins. When I tried to concentrate the purified full-length RAG proteins using PEG 8000, the proteins became severely degraded. This low concentration and poor activity unfortunately means that the results with full-length RAG1 plus full-length RAG2 are currently unreliable.

Even though purification of full-length RAG proteins at high concentrations and activity is difficult, it might not be unachievable. Firstly, I purified full-length RAG proteins from 10 dishes. It may be possible to get higher concentration proteins by increasing the transfected cells. Secondly, I successfully co-expressed and purified core RAG1 plus full-length RAG2 from HEK293T cells by using different method from core RAG1 plus core RAG2 purification (Section 2.18 and 2.19). The lysis buffer used in these two purification methods is different as the buffer for purification of core RAG1 plus full-length RAG2 is more gentle (Section 2.18 and 2.19). In addition, when purification of core RAG1 plus full-length RAG2, the Dounce step was removed to decrease the possibility of protein degradation. This method might be also suitable for purification of full-length RAG1 plus full-length RAG2.

To purify proteins efficiently, the RAG proteins used in the *in vitro* assays are fused with MBP, which is quite large (approximately 40 kDa). MBP was found to be able to enhance solubility of RAG proteins (McBlane et al., 1995). It is currently not known if MBP significantly affects the activity of the RAG proteins. However, the fact that similar results were observed *in vivo* (Chapter 4) to those *in vitro*, implies that the effect of MBP is minimal. In addition, numerous previous studies have been performed using MBP-RAG fusion proteins to

investigate the *in vitro* V(D)J recombination assays (Grundy et al., 2009; McBlane et al., 1995; Rahman et al., 2006; Swanson & Desiderio, 1999).

3.9 Asymmetric cleavage is due to RAG binding to both RSSs within the ESC

The binding of RAG proteins was proposed to extend 12 bp and 16 bp beyond the RSS into the flanking coding gene segment in a single and paired RSS/RAG complexes, respectively (Nagawa et al., 2002; Nagawa et al., 2004). The ESC contains a pair of RSSs in a head-to-head configuration. Therefore, the second RSS forms the flanking coding gene segment of the first RSS. Consequently, one hypothesis suggests that asymmetric ESC/RSS cleavage occurs due to RAG proteins binding to both sides of the ESC simultaneously, thus extending to the adjacent RSS and preventing efficient RAG cleavage of ESC at the heptamer/heptamer boundary. A prediction of this hypothesis is that if one side of ESC is mutated, then this should restore symmetric ESC/RSS cleavage since this should release the cutting site at heptamer/heptamer boundary. Consistent with this hypothesis, DNase I footprint data showed that RAG proteins can protect from DNase I digestion on both RSSs of the ESC and almost 100% protection of some regions of both heptamers was observed. These data suggest that both RSSs within ESC are simultaneously occupied, which is also consistent with higher molecular weight complex observed in bandshift experiments (Figures 3.8 and 3.9). A problem with the footprint experiment is that the background is high, thus obscuring some signals and contributing to the relatively weak footprint in some places. One way to overcome this would be to gel purify the probe following radio-labelling. It should also be noted that the footprint assays were carried out with core RAG proteins due to relatively low activity of full-length RAG proteins. If active full-length RAG proteins could be purified, to better mimic the physiological situation, the footprint assay should be repeated with full-length RAG proteins.

Consistent with the cut-and-run hypothesis and footprint assays, mutating one RSS within ESC restores symmetric cutting in RAG cleavage assays *in vitro*.

RAG1 binds to the nonamer of the RSS and also contacts the heptamer, together with RAG2. When the heptamer, nonamer or CAC in the heptamer of one RSS within ESC was deleted or mutated, RAG binding to that RSS is decreased and the ESC cutting site is released. This is also consistent with the bandshift results where the intensity of high molecular weight complex with mutant ESCs (ESC12d9 and ESC23dCAC) is decreased relative to HSC1 and HSC2/PC. However, interestingly, this decrease for ESC12dCAC was not observed and notably this corresponds with the result that asymmetric cleavage is not restored when ESC12dCAC was incubated with either 12-RSS or 23-RSS. An interesting question is why there is a difference between ESC23dCAC and ESC12dCAC. This might be accounted for several possibilities. Firstly, RAGs contact the 12-/23-RSS spacers differently and binding to the 12-RSS is more stable. Secondly, RAGs could find a pseudo heptamer sequence in the ESC12dCAC oligonucleotides and thus still block ESC cutting. Thirdly, after binding by RAG proteins, proper bending of RSSs needs to occur to allow RAG cleavage. It is possible that the 23-RSS is more bent, thus exposing the cleavage site.

It is notable that the cutting efficiency at the ESC 15-spacer is still not as high as at the 12- or 23-RSS, even though it is higher than for the wild-type ESC (Figure 3.10). The reason for this possibly is that the increased spacer between two RSSs within the ESC is not long enough. I also performed *in vitro* RAG cleavage assay with ESC 6-spacer, ESC 9-spacer and ESC 12-spacer and none of them was cleaved (data not shown) until the spacer length was increased to 15 bp. The low cutting efficiency at ESC 15-spacer suggests that the RAG binding to both RSSs within the ESC 15-spacer is still partially blocked the cleavage site which is consistent with the idea that the binding of RAG proteins extends 16 bp beyond the RSS into the flanking coding gene in a paired RSS/RAG complexes (Nagawa et al., 2004). Therefore, the cleavage site in the ESC would possibly be released by increasing the spacer length further, to 18 bp or even longer.

3.10 The RAG2 C-terminus and cut-and-run reaction

Previously, due to the problematic of purification of full-length RAG1 and full-length RAG2, all of the *in vitro* RAG assays were performed with core RAG1 plus core RAG2. This does not perfectly mimic the *in vivo* conditions. In addition, core RAG proteins can carry out transposition *in vitro*, catalysing the attack of the signal ends into an unrelated target sequence (Agrawal et al., 1998; Hiom et al., 1998; Melek & Gellert, 2000), whilst further evidence showed RAG2 C-terminus can protect against transposition *in vivo* (Elkin et al., 2003; Deriano et al., 2011). Furthermore, RAG2 C-terminus was also shown to prevent reintegration (Curry et al., 2007). Therefore, to test if the asymmetric RSS/ESC cleavage also occurs with full-length RAG proteins, and if cut-and-run can also be inhibited by RAG2 C-terminus *in vitro*, core RAG1 plus full-length RAG2 was used to perform *in vitro* RAG cleavage assays using the same conditions in the *in vitro* cutting assay as with core RAG1 plus core RAG2. By comparing with the cutting that occurred with core RAG1 plus core RAG2, it was found that whilst there is a slightly increased cleavage efficiency with core RAG1 plus full-length RAG2, but nevertheless, asymmetric RSS/ESC cleavage is still maintained (Figure 3.2 A and B). Firstly, this result suggests that RAG2 C-terminus is unlikely to inhibit cut-and-run reaction at least *in vitro*. Secondly, given that the only difference between these two experiments is the RAG proteins used, it is reasonable to speculate that this increased cleavage efficiency is possibly due to the presence of RAG2 C-terminus increasing the accessibility of RAG binding to RSSs, leading to increased cleavage, which is in agreement with previous studies showing that RAG2 C-terminus is critical for optimal V(D)J recombination (Akamatsu et al., 2003).

Chapter 4:

Cut-and-run occurs *in vivo* via the same mechanism as *in vitro*

A) Introduction

In Chapter 3, I further investigated the molecular basis of asymmetric RSS/ESC cleavage *in vitro* by using oligonucleotides carrying RSS or ESC (two head-to-head RSS) sequences and showed that this asymmetric RSS/ESC cleavage is due to RAG simultaneously binding to two RSSs within the ESC, thus blocking the cleavage site at the heptamer/heptamer boundary of the ESC. In addition, to some extent, this asymmetric cleavage can be restored by increasing the spacer between two RSSs within the ESC or mutating one side of RSSs within the ESC. However, these experiments were carried out using either core RAG1 plus core RAG2 or core RAG1 plus full-length RAG2, and under physiological conditions, it is full-length RAG1 and full-length RAG2 that function as the recombinase for cleavage of the RSSs. Therefore, the aim of this chapter is to determine if cut-and-run occurs *in vivo* via the same mechanism as *in vitro*. In this thesis, *in vivo* means in cell lines.

Recombination experiments *in vitro* are usually investigated with short oligonucleotides (McBlane et al., 1995; van Gent et al., 1996), which is what I used in chapter 3. To investigate the cut-and-run reaction *in vivo*, extrachromosomal substrates are often transfected into lymphoid and non-lymphoid cells. In 1987, Hesse et al first used an extrachromosomal V(D)J recombination substrate, pJH201, which contains two recombination signal sequences, to investigate V(D)J recombination by transfection of this substrate into a murine pre-B cell line, PD31 (Hesse et al, 1987). Later, Liber et al modified this extrachromosomal recombination substrate to pJH200 by removing the kappa transcription enhancer and used it to investigate the mechanism and regulation of lymphoid-specific V(D)J recombination (Lieber et al., 1988). Subsequently, a human assay system for V(D)J recombination was first generated by transfection of extrachromosomal recombination substrates into the lymphoid cell lines, REH and Nalm-6 (Gauss & Lieber, 1993). Both of

REH and Nalm-6 cells are human pre-B cell lines that were originally derived from patients with acute lymphoblastic leukaemia (ALL; Hurwitz et al., 1979; Rosenfeld et al., 1977). They were used to perform *in vivo* V(D)J recombination assays since these two human pre-B cell lines have been reported to possess TdT activity (Minowada et al., 1978; Rosenfeld et al., 1977), have recombinations at their *IgH*, *Igκ*, and *Igλ* loci (Korsmeyer et al., 1983), and have early rearrangement at their T-cell receptor α/δ loci (Hansen-Hagge et al., 1992). Later, many other investigators demonstrated V(D)J recombination activity in lymphoid cells by transfection of extrachromosomal substrates (Fisher et al., 2017; Pinsonneault et al., 2007; Raghavan et al., 2001; Um et al., 2013). In addition to the lymphoid system, extrachromosomal recombination substrates were also transfected into non-lymphoid cells, together with RAG expression plasmids, to investigate V(D)J recombination. This was first performed by Steen et al in 1996 (Steen et al., 1996) and afterward, numerous other studies also investigated recombination activity by transfection of extrachromosomal recombination substrate into non-lymphoid cells (Han et al., 1997; Han et al., 1999; Vaura et al., 2007).

As under appropriate conditions *in vitro*, RAG cleavage occurs most efficiently in a coupled fashion at a 12/23 RSS pair *in vivo* (Eastman et al., 1996; Kim & Oettinger, 1998; Steen et al., 1997; Tonegawa et al., 1983; van Gent et al., 1996), as predicted by the 12/23 rule. Several studies have demonstrated that when the two RSSs are located on distinct DNA molecules (*in trans*), the V(D)J recombination efficiency is much lower than that when the two RSSs are positioned on the same DNA molecule (*in cis*) (Hesse et al., 1987; Tevelev & Schatz, 2000). Since *trans* V(D)J recombination events might be associated with oncogenic chromosomal translocations, it is not surprising that *trans* V(D)J rearrangements are relatively rare (Bailey & Rosenberg, 1997; Han et al., 1999; Tycko et al., 1989). Nevertheless, several studies found that *trans* cleavages of plasmid substrates *in vivo* proceeds efficiently both in fibroblasts and in lymphoid cells (Han et al., 1999; Tevelev & Schatz, 2000). Moreover, further studies showed when the two RSSs reside on different DNA molecules, signal ends are joined more efficiently, but coding ends are barely joined, suggesting that intermolecular recombination (i.e. recombination between two

RSSs located on different DNA molecules) is repressed at the level of coding joint formation (Han et al., 1999). Indeed, RAG binding and cleavage of RSSs located on different DNA molecules occurs with the same efficiency as when RSSs are located on the same DNA molecule (Han et al., 1999; Steen et al., 1996). In addition, stimulation of cleavage of RSSs obeys the 12/23 rule no matter whether they are located on different or the same DNA molecule. Synapsis of an ESC and a legitimate or cryptic genomic RSS is an inherently intermolecular interaction. In other words, if a RAG cleavage reaction occurs between an ESC and an RSS in the genome, it must be a *trans* cleavage reaction. In the proposed cut-and-run reaction, after forming a synaptic complex between an ESC and an RSS, the RSS is cleaved whilst the ESC remains intact (Kirkham, 2014). This asymmetric cleavage reaction doesn't involve in a joining step. Consequently, two separate plasmids carrying RSS and ESC sequence could be used to examine whether cleavage between RSS and ESC is asymmetric or not *in vivo*.

Therefore, in this chapter, to test if asymmetric RSS/ESC cleavage occurs *in vivo*, NIH3T3 mouse fibroblast cells, a non-lymphoid cell line, were transfected with plasmids carrying either an RSS or/and ESC sequence, together with RAG expression plasmids. Following this, quantitative PCR (qPCR) was used to determine the level of cleavage. If the cut-and-run reaction occurs *in vivo*, the cleavage of these transfected substrates should be asymmetric. In addition, to test if asymmetric RSS/ESC cleavage also occurs at physiological concentrations of RAGs, a human pre-B cell line, REH, was transduced with lentiviruses including RSS or ESC sequence, followed by FACS to check the infection efficiency and qPCR to calculate the cleavage efficiency.

B) Results

4.1 Asymmetric RSS/ESC cleavage occurs *in vivo*

To test if asymmetric ESC/RSS cleavage also occurs *in vivo*, NIH3T3 fibroblast cells were co-transfected with a plasmid carrying a non-consensus 12-RSS (pJ β 2.7+, Figure 2.14; Figure 4.1A) and a plasmid carrying either a

consensus (Figure 2.18) or a non-consensus ESC (pJ δ 1+, Figure 2.17; Figure 4.1A), together with full-length RAG1 expression vector (pJH548; Sadofsky et al., 1992) and a core RAG2 expression vector (pEFXC-cRAG2). Even though the signal sequences reside on plasmid substrates, and not in the context of a chromosome, it has been reported that transient transfection of plasmids allows the substrates to chromatinise to some extent (Cereghini & Yaniv, 1984; Reeves et al., 1985). Therefore, to cause transcription through the signal sequence, a strong human cytomegalovirus (CMV) promoter sequence was cloned upstream of the recombination signal sequences in all plasmid substrates. This is beneficial because transcription increases recombination efficiency by improving accessibility, for instance, via the eviction of H2A/H2B dimers from nucleosomes to allow RAG cleavage (Bevington & Boyes, 2013). Transfected cells were harvested 48 hours post-transfection and plasmid DNA was extracted from the harvested cells via Hirt extraction (Section 2.37). The cleavage efficiency of the RSS or ESC substrate was determined by a qPCR assay using a pair of primers that amplify across the RAG cleavage site. Therefore, only intact substrate is amplified. By comparing the level of intact plasmid with the total amount of plasmid, cleavage at the RSS could be calculated. The data were then normalized to a unique region of each plasmid substrate by qPCR amplification to control for recovery efficiencies.

To optimize the *in vivo* RAG cleavage conditions, 0.1 μ g and 0.5 μ g of non-consensus 12-RSS and non-consensus ESC substrates were firstly transfected, together with 1 μ g of each RAG expression plasmids. The results showed that the intact 12-RSS substrates decreased by 11% and 9% when 0.5 μ g and 0.1 μ g of J β substrates were transfected into NIH3T3 cells, respectively, whilst J δ , which contains non-consensus ESC sequence, remained uncut no matter how much substrate was transfected (Figure 4.1B). These data therefore suggest that, the cleavage of transfected RSS/ESC substrate is asymmetric *in vivo* which is consistent with our *in vitro* data and the cut-and-run hypothesis. Since slightly better cleavage was achieved when 0.5 μ g of 12-RSS substrates was transfected, this condition was used for all of the subsequent *in vivo* RAG cleavage assays.

To further test if this asymmetric cleavage also occurs with a consensus ESC, 0.5 μg of $J\beta$ and a consensus ESC-containing plasmid were transfected, along with 1 μg of pJH548 and 1 μg of pEFXC-cRAG2. Again, asymmetric cleavage between the 12-RSS and the ESC was observed, with 12% cutting at the non-consensus 12-RSS substrates but no cutting of the consensus ESC substrates (Figure 4.1C). These data indicate that, asymmetric cleavage also occurs between a non-consensus RSS and a consensus ESC.

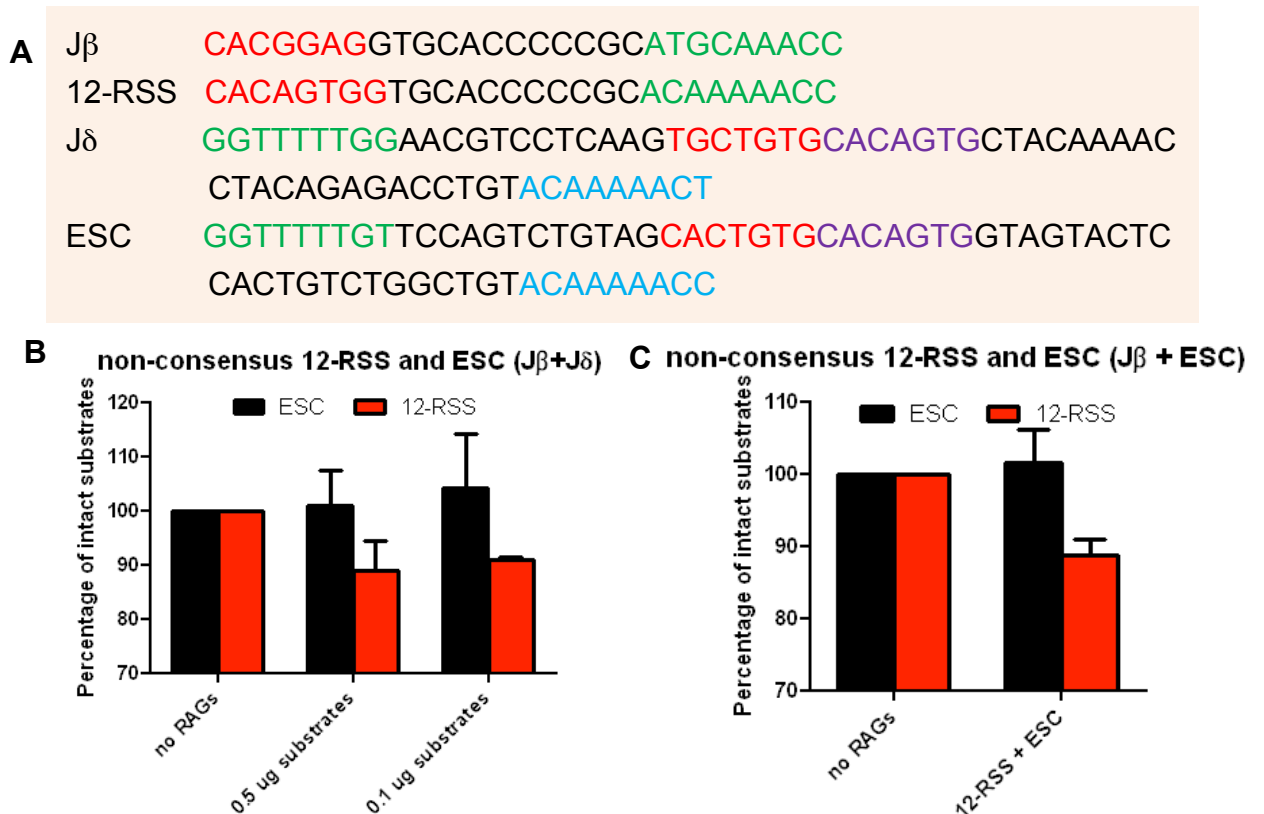


Figure 4.1 Cleavage of transfected 12-RSS/ESC substrates is asymmetric *in vivo*

(A) The sequences of non-consensus and consensus 12-RSS and ESC. The 12-RSS heptamer and nonamer are in red and green, respectively. The 23-RSS heptamer and nonamer within the ESC are in purple and blue, respectively. **(B)** Titration of RSS and ESC substrates for the *in vivo* RAG cleavage assay. NIH3T3 cells were transfected with a full-length RAG1 expression vector, a core RAG2 expression vector, and ESC and RSS substrate plasmids. Plasmids were extracted after 48 hrs and the level of RSS or ESC cleavage was measured by qPCR. Data were normalised to unique regions within each plasmid, and the values given are relative to no RAG controls. Error bars show the SEM of three independent transfections. **(C)** Asymmetric cleavage occurs between non-consensus 12-RSS and consensus ESC.

4.2 Mutating one side of the ESC restores symmetric cleavage *in vivo*

To further test if cut-and-run occurs via the same mechanism *in vivo* as *in vitro* i.e. asymmetric cleavage is due to RAG binding to both RSSs within ESC, blocking cutting site at heptamer/heptamer boundary, the ESC was mutated so that RAGs can only bind to one RSS in the ESC. Mutated ESC plasmids, pESC12d7 and pESC12d9, where the heptamer or nonamer of 12-RSS in the ESC was mutated, respectively, were co-transfected with either the non-consensus 12-RSS or consensus 23-RSS containing plasmids, with pJH548 and pEFXC-cRAG2, followed by qPCR to determine the cutting efficiency. pESC12d7, pESC12d9 and pESC23d9 had already been generated by the Boyes lab. pESC23d7 was cloned by inserting an annealed ESC oligonucleotide in which the heptamer of 12-RSS is deleted into the pcDNA3.1 vector.

As shown in figure 4.2A, when the heptamer of 12-RSS within the ESC was deleted, making the 12-RSS within the ESC essentially unavailable, 20% cutting of ESC12d7 was observed and 10% cutting was seen at the 12-RSS (J β) (Figure 4.2A). Likewise, when the nonamer of 12-RSS within ESC was mutated, 11% and 10% cutting were observed for ESC12d9 and 12-RSS, respectively (Figure 4.2A). These data suggest that the symmetric cleavage is restored when 12-RSS within ESC is mutated.

Similarly, when cells were transfected with a 23-RSS and a wild-type ESC substrates, the intact 23-RSS substrate decreased by 14%, but the intact ESC substrates did not decrease at all. Whilst when transfected with 23-RSS and ESC mutants (ESC23d7 and ESC23d9), 14% and 16% cutting was observed at ESC23d7 and ESC23d9, respectively, whereas 12-18% of 23-RSS substrates were cleaved. These data suggest that the cleavage between 23-RSS and ESC is asymmetric *in vivo*, and that symmetric 23-RSS/ESC cleavage can be restored by deletion either the heptamer or the nonamer in the 23-RSS within the ESC.

Together, these results are consistent with the *in vitro* RAG cleavage assays, implying that cut-and-run occurs by the same mechanism *in vivo* as *in vitro*.

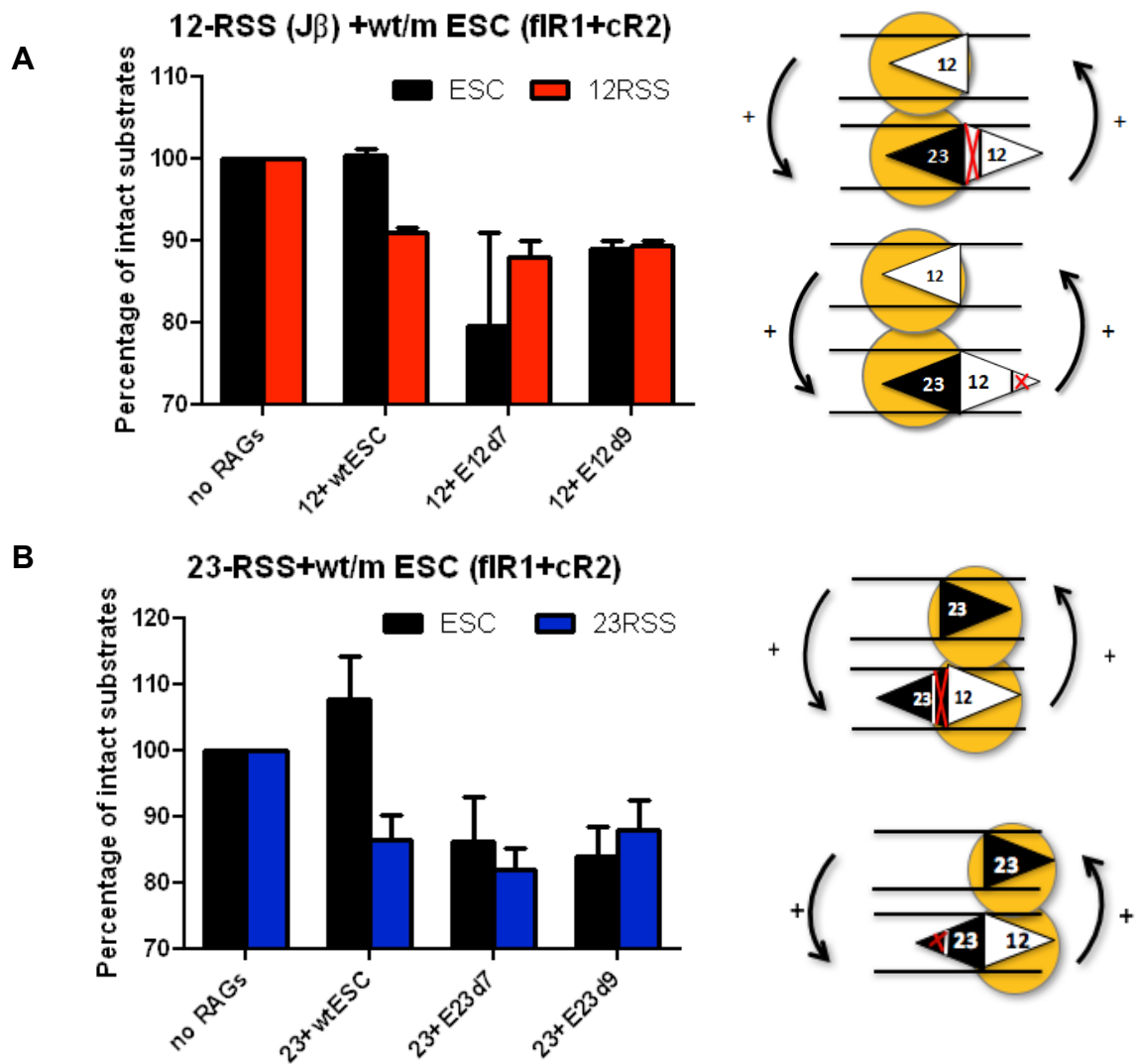


Figure 4.2 Mutation of the RSS in the ESC restores symmetric cutting within an RSS/ESC complex *in vivo*

To determine if mutation of one RSS in the ESC restores symmetric cleavage to an ESC/RSS pairing *in vivo*, NIH3T3 cells were transfected with a full-length RAG1 expression vector, a core RAG2 expression vector, 12-RSS or 23-RSS plasmid and mutant ESC plasmids. Plasmids DNA extraction and data analysis are shown as Figure 4.1. Error bars show the SEM of three independent transfections.

4.3 Asymmetric cutting is observed with both full-length and core RAGs

The previous data show that asymmetric RSS/ESC cleavage by full-length RAG1 and core RAG2 occurs *in vivo*. However, it is still not known if this asymmetric cleavage also occurs with full-length RAG1 plus full-length RAG2. This is important because the RAG2 C-terminus was shown to inhibit

transposition and reintegration and both these reactions involve integration of the ESC with target DNA sequences (Curry et al., 2007; Deriano et al., 2011; Elkin et al., 2003). Whilst we found cut-and-run is not inhibited by the RAG2 C-terminus *in vitro* (Figure 3.2), there is no data to show this is the case *in vivo*. Therefore, to test this, NIH3T3 cells were transfected with 12- or 23-RSS plasmids and a wild-type ESC plasmid, along with either full-length RAG1 plus core or full-length RAG2 expression plasmids or core RAG1 plus core RAG2 expression plasmids. The results show that both 12- and 23-RSS were cleaved, albeit at slightly different levels, whilst the ESC remains intact no matter whether full-length or core RAG expression plasmids were transfected (Figure 4.3). This suggests that asymmetric RSS/ESC cleavage occurs *in vivo* no matter whether core or full-length RAG proteins are present, which is consistent with our *in vitro* RAG cleavage assays (Figure 3.4). This also indicates that, as *in vitro*, the RAG2 C-terminus does not inhibit the cut-and-run reaction *in vivo*. In addition, within the 12-RSS and ESC pairing, it was found that cutting of the 12-RSS is somewhat more efficient by full-length RAG1 plus full-length RAG2 than by full-length RAG1 plus core RAG2. However, this was not observed in the 23-RSS and ESC pairing. In contrast, it was found that, the intact 23-RSS substrates decreased more when co-transfected with core RAG1 and core RAG2 expression plasmids, compared to when full-length RAG1 plus core or full-length RAG2 expression plasmids were co-transfected, although the difference is not statistically significant. The reasons for these difference are unclear.

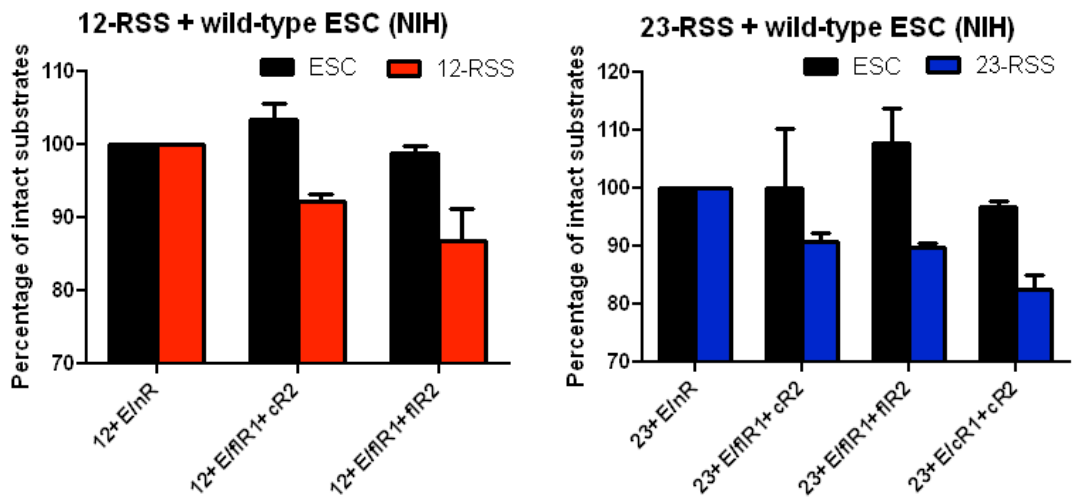


Figure 4.3 Comparison of asymmetric RSS/ESC cleavage by core RAG proteins and full-length RAG proteins

To compare the asymmetric RSS/ESC cleavage by different RAG proteins and to determine cut-and-run is inhibited by the RAG2 C-terminus *in vivo*, NIH3T3 cells were transfected with 12- or 23-RSS plasmids and the wild-type ESC plasmid, along with either full-length RAG1 plus core or full-length RAG2 expression plasmids or core RAG1 plus core RAG2 expression plasmids. Plasmids DNA extraction and data analysis are shown as Figure 4.1. Error bars show the SEM of three independent transfections.

4.4 SCR7 facilitates the detection of consensus 12-RSS signal ends *in vivo*

Experiments using a consensus 12-RSS and consensus ESC pairing were also performed under the same conditions as above. However, no cleavage of the consensus 12-RSS or ESC were could be detected (data not shown). One possible reason for this could be due to the broken ends being repaired by the NHEJ machinery *in vivo*. Therefore, to detect the percentage of cleavage more accurately and easily, the NHEJ machinery was blocked. To this end, 100 μ M of SCR7, which is an inhibitor of DNA ligase IV (Srivastava et al., 2012), was added to the transfected cells 24 hrs post-transfection, thus blocking end-joining by interfering with DNA ligase IV binding to DNA. In addition, to best mimic physiological conditions, full-length RAG1 and full-length RAG2 expressing plasmids (pJH548 and pEFXC-fIRAG2, respectively) were used. Cleavage between a consensus 12-RSS and consensus 23-RSS was firstly examined. As shown in Figure 4.3, cleavage was detectable at both a 12-RSS

and 23-RSS, of 16% and 13%, respectively (Figure 4.4). This suggests that symmetric 12-RSS/23-RSS cleavage occurs *in vivo*, in agreement with the results *in vitro*.

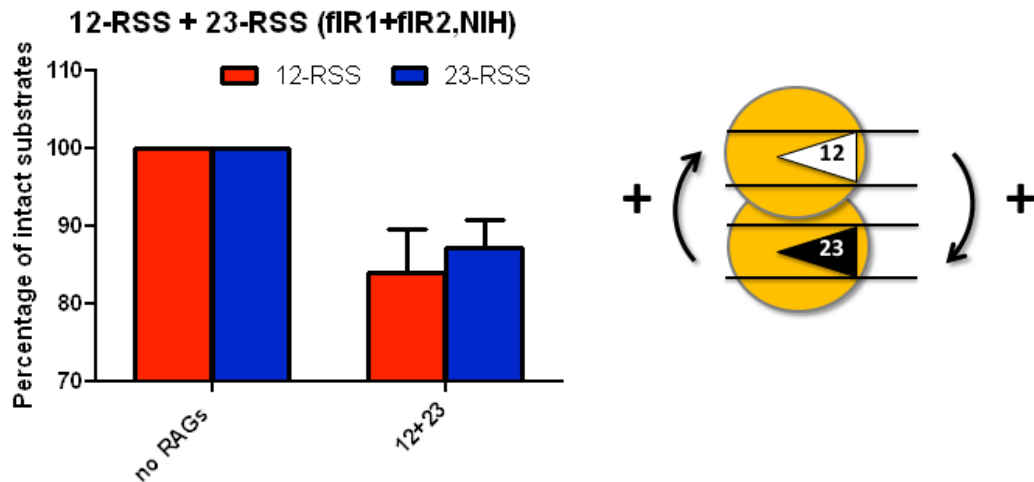


Figure 4.4 Symmetric 12-RSS/23-RSS cleavage *in vivo*

To accurately determine the cleavage *in vivo*, NIH3T3 cells were transfected with a full-length RAG1 expression vector, a full-length RAG2 expression vector, consensus 12-RSS and consensus 23-RSS plasmid. 100 μ M SCR7 was added to the transfected cells 24 hrs post-transfection. Plasmids DNA extraction and data analysis are shown as Figure 4.1. Error bars show the SEM of three independent transfections.

Next, cleavage of a consensus RSS with wild-type or mutated ESC was examined under these conditions i.e. with full-length RAG1 and full-length RAG2, as well as 100 μ M SCR7. In this case, as with full-length RAG1 plus core RAG2, the intact 12- and 23-RSS substrates were decreased by 15% and 10%, respectively, whilst the intact ESC substrates remained at the same level as the no RAG controls (Figure 4.5A and B). In addition, when the consensus RSS was paired with an appropriate ESC mutant, both RSS and the ESC were cleaved (Figure 4.5A and B). Furthermore, to compare with the *in vitro* RAG cleavage assay, the effect of core RAG1 plus core RAG2 was also examined. In agreement with the *in vitro* data, asymmetric 23-RSS/ESC cleavage by core RAG1 plus core RAG2 also takes place *in vivo* and this asymmetric cleavage could be restored by mutation of one of the RSSs in the ESC (Figure 4.5C). It is notable that, by comparing the 23-RSS and wild-type or mutated ESC pairing with or without SCR7 (Figure 4.2B vs Figure 4.5B), the decreased level of intact substrates is quite similar, suggesting that blocking the NHEJ DNA

repair pathway doesn't increase the detectable cleavage level with a 23-RSS whilst it does have an effect with a 12-RSS.

Overall, these data suggest, consistent with *in vitro* RAG cleavage assays, that asymmetric RSS/ESC cleavage occurs *in vivo* with no matter whether full-length or core RAG proteins are used. Moreover, mutating one RSS within the ESC can restore this symmetric cleavage.

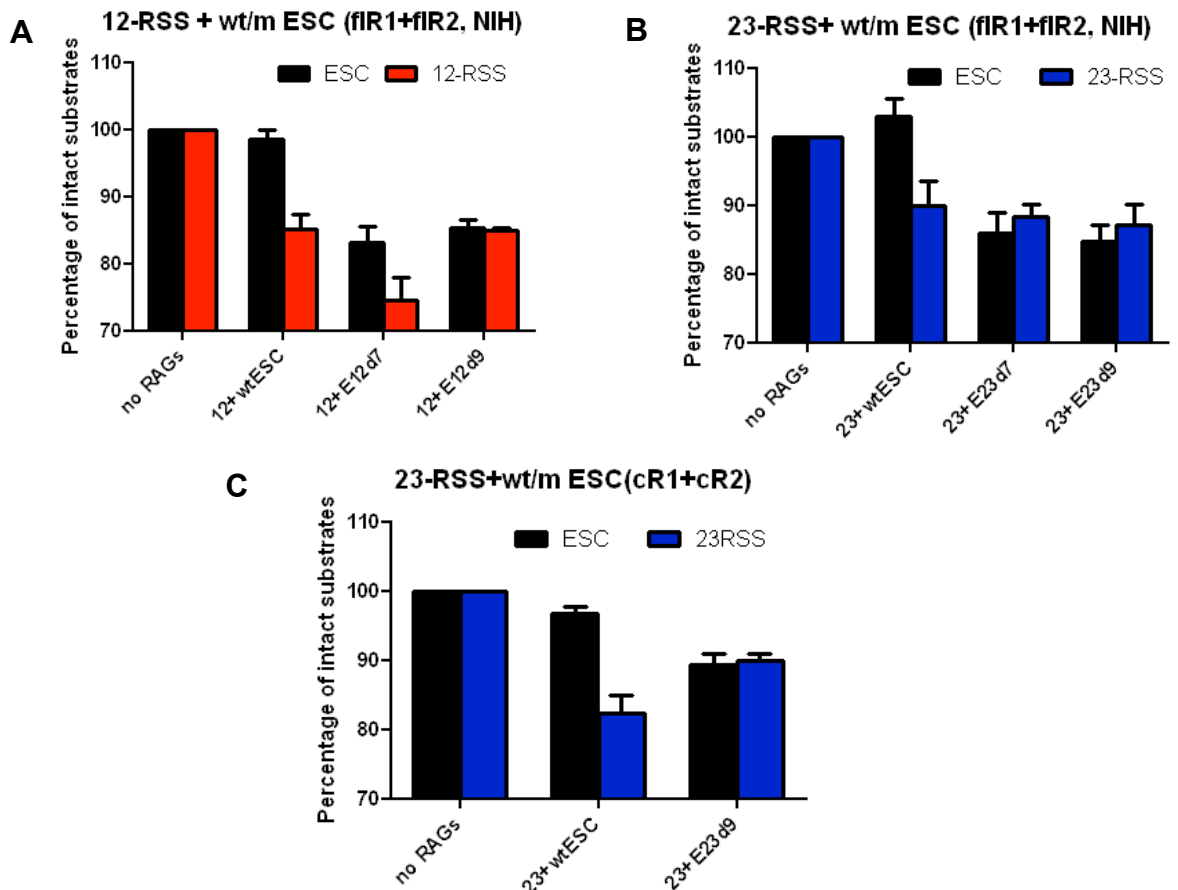


Figure 4.5 Asymmetric RSS/ESC cleavage *in vivo*

To determine the mechanism of asymmetric cleavage *in vivo*, NIH3T3 cells were transfected with a 12- or 23-RSS plasmid and a wild-type or mutated ESC plasmid, along with either full-length RAG1 plus full-length RAG2 expression plasmids or core RAG1 plus core RAG2 expression plasmids. SCR7 addition, plasmids DNA extraction and data analysis are shown as Figure 4.4. Error bars show the SEM of three independent transfections. **(A)** Symmetric 12-RSS/ESC cleavage was restored by mutation of 12-RSS within the ESC with full-length RAG1 plus full-length RAG2 in non-lymphoid cells. **(B)** Symmetric 23-RSS/ESC cleavage was restored by mutation of 23-RSS within the ESC with full-length RAG1 plus full-length RAG2 in non-lymphoid cells. **(C)** Symmetric 23-RSS/ESC cleavage was restored by mutation of 23-RSS within the ESC with core RAG1 plus core RAG2 in non-lymphoid cells.

4.5 Cut-and-run occurs at physiological concentrations of RAGs

The *in vivo* RAG cleavage assays showed above were all carried out in non-lymphoid cells by transfection of all the required substrates. However, under physiological conditions, V(D)J recombination occurs in lymphoid cells. Therefore, to test if cut-and-run also occurs at physiological concentrations of RAGs, lentiviruses carrying a 12-RSS (Figure 2.19), a 23-RSS (Figure 2.20) or an ESC signal joint (Figure 2.21) that also express GFP were generated and transduced into REH cells, a human pre-B cell line derived from an ALL patient (Rosenfeld et al., 1977) where RAGs are expressed at physiological concentrations (Bories et al., 1991). Forty-eight hrs post-transduction, 1/5 transfected REH cells were used for analysis by flow cytometry to measure the transduction efficiency by measuring the GFP expression level; only cells with a 35 - 55% infection efficiency were used in further experiments (Figure 4.6 A-C). The remainder of the cells were used to extract episomal DNA via Hirt extraction (Section 2.37) and the levels of RAG cleavage were measured via qPCR.

When REH cells were co-transduced with lentiviruses carrying a 12- and a 23-RSS, both 12- and 23-RSS were cleaved, with levels of 27% for 12-RSS and 20% for 23-RSS, respectively, relative to samples where no partner was present (Figure 4.6 D), in agreement with previously published data and the *in vivo* RAG cutting assay in NIH3T3 cells carried out by Chris Kirkham (Steen et al., 1996; Kirkham, 2014). When cells were co-transduced with lentiviruses carrying an ESC and either a 12- or 23-RSS, the RSS was cleaved in both cases, with levels of 24% for 12-RSS and 20% for 23-RSS, whilst negligible cleavage of the ESC is detected (Figure 4.6 E and F). These data therefore indicate that under physiological concentrations of RAGs, cleavage within a 12-RSS/23-RSS complex is symmetric, whilst within in an ESC/RSS complex is asymmetric, corresponding to our *in vitro* data and our *in vivo* non-lymphoid cell RAG cleavage assays.

Our *in vitro* and *in vivo* (NIH3T3 cells) data suggest that asymmetric ESC/RSS cleavage is due to RAG binding to both RSSs in the ESC at the same time. To test if the same is true under physiological RAG concentrations, lentiviruses

carrying mutated ESC sequences (Figures 2.25-2.28) were generated and used for the transduction assays in REH cells, together with lentiviruses carrying wild-type 12- or 23-RSSs. If RAG binding to the second RSS indeed blocks ESC cutting, then we predict that the mutated ESC should be cut in the presence of the appropriate partner. As can be seen in Figures 4.6 E and F, this prediction is borne out, i.e. when an ESC mutated at the 12-RSS was paired with a 12-RSS, both the RSS and the ESC were cleaved (Figure 4.6 E). Likewise, for an ESC mutated at the 23-RSS and paired with a 23-RSS (Figure 4.6 F). Together, these data suggest that asymmetric cleavage, thus potentially the cut-and-run reactions, also occurs under physiological concentration of RAGs.

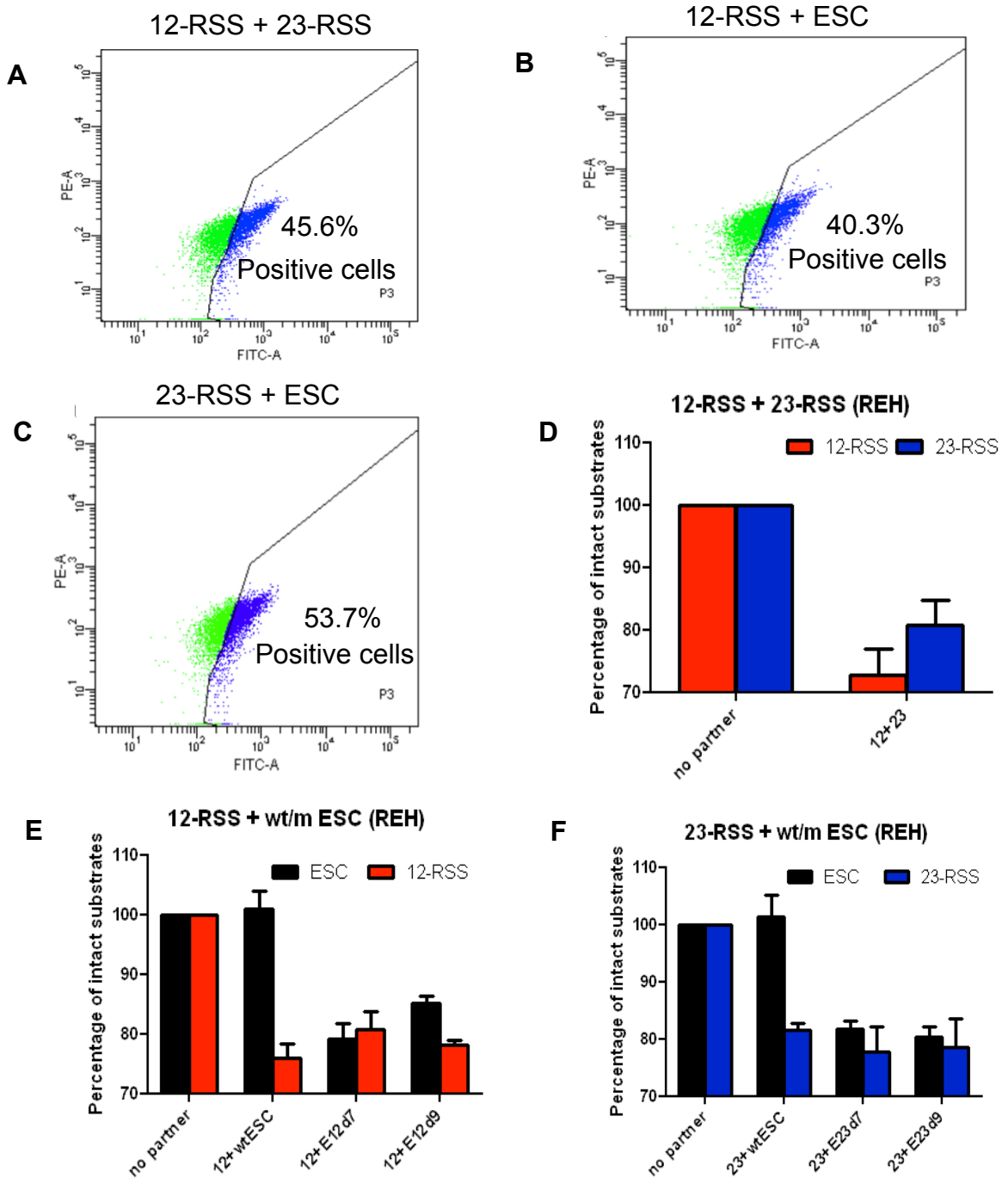


Figure 4.6 Cut-and-run occurs at physiological concentrations of RAGs

To determine if cut-and-run occurs at physiological concentrations of RAGs, human pre-B (REH) cells were transduced with lentiviruses carrying a wild-type or mutant ESC or/and an RSS, followed by FACS to check infection efficiency and qPCR to check cutting at the RSSs or ESC. Data were normalised to unique regions within each virus, and the values given are relative to no partner controls. Error bars show the SEM of three independent transductions. FACS to check the co-infection efficiency of **(A)** 12-RSS and 23-RSS viruses, **(B)** 12-RSS and ESC viruses and **(C)** 23-RSS and ESC viruses. **(D)** Symmetric 12-RSS/23-RSS cleavage under physiological RAG concentrations. Symmetric **(E)** 12-RSS/ESC and **(F)** 23-RSS/ESC cleavage can be restored by mutation of **(E)** 12-RSS and **(F)** 23-RSS within the ESC under physiological RAG concentrations.

C) Discussion

In this chapter, I investigated the cut-and-run reaction both in non-lymphoid and lymphoid cells. As *in vitro*, cut-and-run reaction occurs *in vivo* via asymmetric cleavage between the RSS and the ESC. In addition, *in vivo*, the symmetric cleavage could be restored by mutation of one RSS within the ESC.

4.6 Asymmetric cutting *in vivo*

Asymmetric ESC/RSS cutting is an important feature of cut-and-run. This has been shown *in vitro* by RAG cutting assays and the mechanism for asymmetric ESC/RSS cleavage has been found to be due to RAG binding to both RSSs within ESC and blocking the cutting site at heptamer/heptamer boundary. However, to further determine the significance of asymmetric ESC/RSS cleavage and the occurrence of cut-and-run *in vivo*, RAG cutting assays need to be carried out under physiological conditions. Consistent with *in vitro* RAG cutting assays, *in vivo* RAG cutting assays in NIH3T3 cells also showed that a plasmid carrying the RSS was cut whilst ESC remains intact. However, the cutting percentage is relatively low, at only 10-20%. There are various explanations for this low level of cutting. Firstly, it is difficult to achieve co-transfection of four plasmids (RSS containing plasmid, ESC carrying plasmid, RAG1 and RAG2 expression plasmids) into the same cell. Those cells with only some of these required plasmids will not show cutting. Secondly, the PEI transfection reagent could cause cell death after transfection, thus influencing the result detected i.e. if the transfected cells mostly die, then this could mean that cells transfected with only some of the plasmids are primarily the ones that are analysed. Thirdly, the ESC could reintegrate into the RSS plasmid or the genome (Vanura et al., 2007), making it difficult to determine the frequency of cut-and-run.

The *in vivo* experiments showed that, as *in vitro*, when one side of the RSSs within ESC is mutated, it prevents RAG proteins from binding to both RSSs within ESC, restoring the symmetric ESC/RSS cleavage. This is consistent with *in vitro* RAG cutting assays and DNase I footprint assays. Whilst

asymmetric cutting was found to occur *in vivo* in NIH3T3 cells, this was not at physiological concentrations of RAGs or the ESC. Therefore, human pre-B cells (REH cells), where the RAG proteins are expressed at physiological levels (Bories et al., 1991), were used to check the cut-and-run reaction *in vivo*. In this case, two types of lentivirus were co-transduced and the infection efficiency was determined via FACS and controlled by altering the amount of the viruses used. However, the detected cleavage efficiency is still at most 30%, suggesting that, similar to NIH3T3 cells, it might be difficult to co-transduce two types of viruses into one cell. In addition, in *trans*-recombination, two molecules on separate plasmids or viruses are less likely to meet than in *cis*-recombination where RSSs on same molecule. Nevertheless, the cleavage efficiency in REH cells is consistently higher than that in NIH3T3 cells. One possibility is that V(D)J recombination of an extrachromosomal V(D)J substrate can be enhanced by increased levels of NHEJ proteins such as Ku70, Ku80, as well as Artemis (Pinsonneault et al., 2007). Importantly, Artemis is constitutively hyperphosphorylated in the REH cell line, thus contributing to the enhanced V(D)J recombination (Pinsonneault et al., 2007).

4.7 SCR7 and cut-and-run reaction

The detection of cut-and-run reaction *in vivo* is not a positive experiment i.e. the cutting of the substrates is not quantified directly, but is calculated instead by quantifying the difference of the total intact substrate before and after RAG cleavage. However, it is still not known if the cleaved RSS ends could be repaired by NHEJ pathway. Therefore, if the cut ends are repaired and re-ligated, the substrates would return to being intact, thus affecting the quantification of total intact substrates after RAG cleavage. To detect the cleaved substrates more accurately, the NHEJ pathway can be blocked. In the cut-and-run reaction, only the RSS is cut, but not the ESC, thus generating one signal end and one coding end. Furthermore, I show in Chapter 5, that the nature of the ends generated by cut-and-run is blunt signal ends and hairpinned coding ends (Chapter 5). The repair of both blunt and hairpinned ends will need the DNA ligase IV to seal double-strand breaks during NHEJ (Helmink & Sleckman, 2012). Therefore, an inhibitor of DNA ligase IV, SCR7,

was used to disrupt the NHEJ pathway to amass DSBs via SCR7 binding to the DNA binding domain of ligase IV and interfering with its binding to DSBs (Srivastava et al., 2012). Before addition of SCR7, no detectable cutting of consensus 12-RSS was observed. However, after addition of 100 μ M SCR7, 16% of intact consensus 12-RSS substrates and 13% of consensus 23-RSS substrates were cleaved, respectively (Figure 4.4). These results suggest that SCR7 aids the detection of consensus 12-RSS ends but has no effect on consensus 23-RSS ends which were cut to 14% even in the absence of SCR7.

However, it is still not known why there is a difference between consensus 12- and 23-RSS. One possibility is the concentration of SCR7. Different concentrations of SCR7 ranging from 1 μ M to 100 μ M were used to inhibit end-joining (data not shown) and it was found that the cutting level of consensus 12-RSS is highest when 100 μ M of SCR7 was added. However, the cutting levels of consensus 23-RSS remain consistent no matter how much SCR7 was added. It is possible that, the concentration of SCR7 is still not high enough to inhibit 23-RSS end joining. It has been reported that SCR7 inhibits end joining in a concentration-dependent manner from 50 μ M and incubation with increasing concentrations of SCR7 inhibited the formation of end joining at 200 μ M and above (Srivastava et al., 2012). Another possibility is the timing of SCR7 addition. If the SCR7 is added after one of the ends has already been joined, it will have no effect. It is not known how long the ends generated by cut-and-run reaction can persist before being repaired. For the normal V(D)J recombination reaction, after RAG cleavage, the signal ends are joined slowly and accumulate whilst the coding ends are processed with the deletion and addition of several nucleotides and rapidly joined (Hiom & Gellert, 1997; Schatz & Swanson, 2011). Therefore, the signal ends persist much longer than coding ends. The RAG proteins can stably associate with signal ends for a long time until the RAG proteins are down-regulated at immature B cell stage. However, in transfected NIH3T3 cells, it is unclear how long the ends persist before being repaired and how long the RAG proteins are expressed. Here, I added SCR7 24 hours post-transfection and harvested cells 48 hours post-transfection. It is possible that the addition of SCR7 is too late and, some ends have already been joined. Nevertheless, the fact that symmetric cleavage

occurs between 12-RSS and 23-RSS whilst asymmetric cleavage occurs between RSS and ESC *in vivo* is in agreement with cut-and-run hypothesis and *in vitro* RAG cleavage assays.

Chapter 5:

Dangers of the cut-and-run reaction

A) Introduction

The previous two results chapters investigated the mechanism of the cut-and-run reaction *in vitro* and *in vivo*. In this chapter, I describe further investigation into the dangers of the cut-and-run reaction.

The dangers of the cut-and-run reaction depend on the number of factors. Firstly, where the breaks occur. Many of chromosome translocations and deletions found in lymphoid cancers have the characteristics of aberrant V(D)J recombination (Kuppers & Dalla-Favera, 2001; Marculescu et al., 2006; Schlissel et al., 2006). Cut-and-run reaction is a new aberrant V(D)J recombination. If it causes a DSB at a cryptic RSS in close proximity to a proto-oncogene, this increases the risk of a chromosomal translocation, that could activate the proto-oncogene that potentially contributes to malignant transformation.

Cryptic RSSs, which resemble an authentic recombination signal sequence, was first demonstrated to undergo aberrant rearrangement in 1983, where it was shown that the L10 gene, a sequence unrelated to any V gene and which in the germ-line sequence is flanked by a sequence resembling a V region recombination signal sequence joined to a J κ 1 gene segment (Hochtl & Zachau, 1983). Lewis et al reported that there are approximately 10 million cRSSs in the genome by looking at cryptic joining-signals sequences within extrachromosomal V(D)J recombination substrates (Lewis et al., 1996). However, the majority of them make no positive contribution to antigen receptor gene assembly since they are not located adjacent to any V, D or J coding sequence (Lewis et al., 1996). Subsequent studies showed that these cRSSs can be cut by RAGs and used in translocations, for example LMO2 and Ttg-1, whilst the ability of cRSSs that flank certain proto-oncogenes to undergo V(D)J recombination is 10-100-fold lower than that of a normal RSS (Marculescu et al., 2006; Raghavan et al., 2001). Later, another study showed

that even though RAG proteins can bind the cRSSs throughout the genome, not all of them can be cleaved by RAG proteins since some of them have already been mutated (Teng et al., 2015). The RAG proteins can recognize cRSSs and use some of them coupled with either other cRSSs or *bona fide* RSSs to catalyse translocations or deletions, contributing to genomic instability and malignancy (Curry et al., 2007). More recently, whole genome sequencing studies of acute lymphocytic leukaemia (ALL; Pui et al., 2004) patients carrying the *ETV6/RUNX1* translocation have further suggested the involvement of cRSSs in the generation of chromosome alterations (Papaemmanuil et al., 2014). The *ETV6/RUNX1* translocation can partially stall B cell development at the pre-B cell stage when RAG proteins are still expressed (Tsuzuki et al., 2004), thus providing the opportunity for secondary rearrangement reactions. Most of these aberrant recombination reactions take place at cryptic RSSs which are near the transcriptionally active B cell developmental-specific genes, thereby further stalling B cell progression. This process can eventually lead to full malignant transformation (Papaemmanuil et al., 2014). In patients where full malignant transformation has taken place, almost 50% of the chromosome aberrations contain an RSS or cRSS on only one side of translocation (Papaemmanuil et al., 2014). Notably, a study performed in Boyes lab using linear amplification-mediated high-throughput genome-wide translocation sequencing (LAM-HTGTS) showed that some translocation break-points formed in REH cells in the presence of the ESC map to the same cRSS break sites as found in ALL patients (unpublished data in Boyes lab; Papaemmanuil et al., 2014), suggesting that cut-and-run makes a significant contribute to chromosome alterations in ALL.

Since ESC itself is not cut and remains bound by RAG proteins after RSS cleavage, cut-and-run reaction has been proposed to trigger numerous double-strand breaks (DSBs) at RSSs or cRSSs throughout the genome. In contrast to single-strand DNA breaks, where the genetic information is still available to template repair since it is retained on the complementary strand, faithful restoration of DSBs can be problematic. In addition to loss of genetic information, DSBs can lead to fragmentation, loss or rearrangement of chromosomes, contributing to the genomic instability that drives cancer development (Aparicio et al., 2014).

A second potential danger of the cut-and-run reaction revolves around whether the cut ends are released or not and what type of cut ends are released. Previously, several studies have shown that, in the normal V(D)J recombination, the RAG proteins remain tightly bound to a pair of cleaved signal ends for some time (Agrawal & Schatz, 1997; Jones & Gellert, 2001). Previous studies of the cut-and-run reaction suggest that after formation of post-cleavage complex, the cleaved ends are released whilst the ESC remains bound by the recombinase (Kirkham, 2014). Since the ESC is not cleaved by RAG proteins, it has potential to accelerate the release of orphan ends, thereby increasing the chance of genomic instability associated with other aberrant V(D)J recombination reactions. However, it is still not known what type of ends are released. It is universally known that in the normal V(D)J recombination, two blunt signal ends and two hairpin coding ends are generated and two of the same type of ends are then joined by the NHEJ pathway (Zhu et al, 1996). In the proposed cut-and-run reaction, since only the RSS is cut and the ESC is not, the nature of the ends generated on the RSS containing DNA substrate are not known. However, the nature of the ends generated by cut-and-run reaction is very important. If this reaction generates a hairpinned coding end and a blunt signal end which could be released by RAG proteins, it is potentially very dangerous since RAG proteins are not able to shepherd a single broken end to the NHEJ pathway for repair unless it is still bound by RAG proteins (Lee et al., 2004). Therefore, these single broken ends could pose a huge threat to genomic stability (Agrawal et al., 1998; Curry et al., 2007; Hiom et al., 1998; Marculescu et al., 2002; Tsai et al., 2008; Vanura et al., 2007).

A third factor that will affect how dangerous the cut-and-run reaction is is how long the ESC persists. For the cut-and-run reaction to occur, the ESC and RAG protein are need to co-exist in a cell. Whilst it is still not known whether ESCs can replicate or not, they appear to be a relatively long-lived and remarkably stable structures *in vivo*. For instance, the ESC in chicken (*Gallus gallus domesticus*) T cells was reported to persist for as long as 2 weeks (Kong & Chen-lo, 1999) and this persistence is thought to be even longer in primate thymocytes (Sodora et al., 2000). Furthermore, the ESC was not

degraded in precursor B cells, suggesting that the ESC has potential to persist into mature B cells in some cases (Kong & Chen-Io, 1999; Sodora et al., 2000). This raises the possibility that cut-and-run reaction could also trigger breaks associated with mature B cell cancers.

If RAG proteins also co-exist with the ESC, cut-and-run reaction is extremely likely to happen. It has been reported that there are two distinct waves of RAG expression corresponding to V(D)J recombination in the developing B cells (Grawunder et al., 1995; Wilson et al., 1994). The first wave of RAG expression occurs at pro-B cell stages, allowing the recombination of *IgH* locus. After forming a pre-BCR, the RAG protein expression is down-regulated. Afterwards, the second wave of RAG expression occurs at the large pre-B cell stage, allowing the recombination of light chain locus. Once recombination occurs successfully, the RAG expression is down-regulated by the expression of BCR in immature B cell stage, preventing further recombination (Brandle et al., 1992). However, if B cells express autoreactive antigen receptors, they can be inactivated by re-expression of the RAG proteins which change the specificity of previous autoreactive B cell receptor, in a process named receptor editing (Tiegs et al., 1993; Gay et al., 1993; McGargill et al., 2000; Nemazee, 2006).

In this chapter, I investigate the dangers of cut-and-run reaction. Firstly, since the cut-and-run hypothesis predicts that the ESC has the potential to trigger cutting at cryptic RSSs, I performed experiments to investigate cryptic RSS cutting in the presence of an ESC partner. Secondly, the stability of RAGs/RSSs and RAGs/RSS/ESC post-cleavage complex was investigated by checking the RSS signal end release following cutting. In complementary experiment, the nature of ends generated by the cut-and-run reaction was determined by LM-PCR. Thirdly, the question of whether the ESC triggers double-strand breaks was addressed by checking for γ H2Ax foci in the presence of ESC substrates and RAG proteins. Finally, the persistence of ESC in the developing B cells was investigated by nested PCR.

B) Results

5.1 The ESC triggers cutting at cRSS *in vitro*

To investigate the dangers of the cut-and-run reaction, I firstly asked if the ESC can trigger cutting at cRSSs. Cryptic RSSs tend to be cut less well than conventional RSSs. It is possible that the complex formed with the ESC is incapable of cutting these poorer substrates. The cRSS at the LMO2 locus was shown to be involved in a chromosome translocation to the TCR α/δ locus (Boehm et al., 1988; Boehm & Rabbitts 1989; Rabbitts et al., 1997). To test if the ESC triggers cutting at a cRSS too, a 12-spacer cRSS from the LMO2 locus (Figure 5.1A) was used as the substrate in an *in vitro* RAG cutting assay. I found that the ESC can indeed trigger cutting at the LMO2 cRSS (Figure 5.1B, lane 7 and 8). Although the level of cutting is much lower than that at an authentic RSS (Figure 5.1B, lane 3 and 4), these data nevertheless suggest that ESC could potentially cause dangerous breaks at cRSSs next to protooncogenes. It is also worth noting that the data show that the cutting efficiency of LMO2 is 2.76-fold lower than that of an authentic RSS, which is consistent with data published by Zhang & Swanson where they showed the relative recombination efficiency of an authentic 12-RSS is 2.63-fold higher than that of a LMO2 cryptic RSS (Zhang & Swanson, 2008). However, previous data by Raghavan et al showed cutting of LMO2 is greatly decreased compared to that of an authentic RSS, with an efficiency of 27-fold lower (Raghavan et al., 2001). I also checked cutting at other cRSSs, such as adjacent to the *Ttg-1*, *SCL1* and *Notch1* loci. Unfortunately, the cutting efficiency was too low even when the incubation time was increased to 1 hour. This is possibly due to poorly matched heptamer and nonamer sequences compared to an authentic RSS.

A 12-RSS GATCTGGCCTGTCTTA^{cacagtg}GTAGTAGGCTGT^{acaaaaacc}CTGCAG
 LMO2 GATCTGGCCTGTCTTA^{cacagtA}GTAGTAGGCTGT^{GcaaTaa}TTCTGCAG

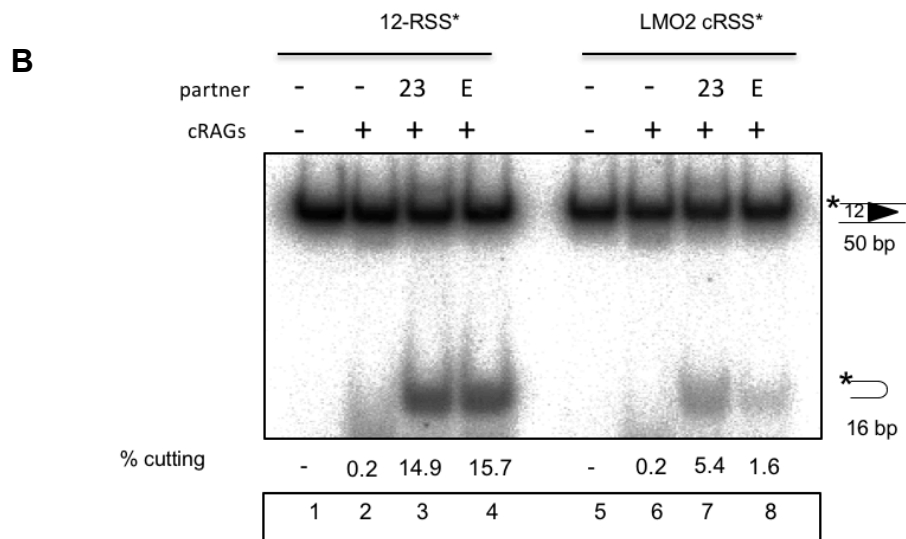


Figure 5.1 The ESC triggers cutting at a cryptic RSS

(A) Sequences of the consensus 12-RSS and LMO2 cryptic RSS. Heptamer sequences are shown in lower-case, coloured red and nonamer sequences are shown in lower-case, coloured blue. The bases at the heptamer and nonamer of the LMO2 cryptic RSS that are different from the consensus 12-RSS are shown in capitals. **(B)** The ESC triggers cutting at the LMO2 cryptic RSS. To test if the ESC triggers cutting at a cRSS, the LMO2 12-spacer cRSS was used as a cutting substrate. The radioactively labelled oligonucleotides are indicated by an asterisk above each set of lanes. Radiolabelled 12-RSS and LMO2 were incubated with an unlabelled partner as indicated, HMGB1 and core RAG proteins. Following cutting, the products were separated on a 12% native polyacrylamide gel and the level of cleavage was determined by quantification of the intensity of cleaved substrates against the intensity of total substrate.

5.2 Analysis of post-cleavage complex stability

It is important to understand the fate of the RSS and ESC following RAG cleavage and specifically, whether the cut RSS remains bound by RAG proteins or whether it is released as broken DNA. This is because RAG proteins are not able to drive the broken ends towards the classical NHEJ pathway for repair unless the broken ends remains bound by RAG proteins in the post-cleavage complexes (Lee et al., 2004). Therefore, if the end is released from the post-cleavage complexes, it could contribute to an enhanced risk of chromosome translocations. Although Chris Kirkham has examined the

fate of the coding ends (Kirkham, 2014), here I examined the fate of the signal ends using a more accepted assay with higher sensitivity. The coding end and signal end release assays are complementary. The coding end release assay demonstrated that RAGs remain bound to the ESC but not RSS. By contrast, signal end release assay indicated that cleaved RSS partner released in ESC/RSS pairing. To this end, purified core or full-length RAG proteins were employed in an *in vitro* RAG cutting assay, followed by challenging the post-cleavage complex at different temperatures, ranging from 37 °C to 60 °C for 30 minutes. Subsequently, half of the reaction was resolved on an 8% native polyacrylamide gel to analyse RAG-substrate complexes, whilst the remainder was electrophoresed on a 10% cutting gel to analyse RAG cleavage products. The stability of the RAG:RSS post-cleavage complexes was determined by quantifying the release of signal ends (SEs) and uncut substrates following temperature challenge (Coussens et al., 2013).

As can be seen in Figure 5.2, the 12-RSS and 23-RSS signal ends are released noticeably earlier from the post-cleavage complexes containing an ESC partner than from complexes involving either a corresponding RSS or mutant ESC partner (Figure 5.2 B-E). Since the RAG2 C-terminus has been shown to play a role in retaining the ends post-cleavage, I also tested if signal ends are released more quickly from ESC/RSS complexes involving full-length RAG2. To this end, a 12-RSS signal end release assay was performed with core RAG1 plus full-length RAG2. Exactly as for core RAG2, 12-RSS signal ends are released faster from the ESC/RSS/RAG complex with full-length RAG2 compared to the RSS/RSS/RAG complex (Figure 5.2 F, G). In addition, by comparing Figure 5.2 D E with F G, it is clear that signal ends are released considerably faster from complex with core RAG2 compared to complex containing the full-length RAG2, which is consistent with previous results from Coussens et al., (Figure 5.2 D, E, F, G; Coussens et al., 2013). Consequently, these data imply that the ESC-stimulated cutting at RSSs contributes to the release of the signal ends. In addition, previous data by Chris Kirkham showed that RAG proteins stay more stably associated with the ESC (Kirkham, 2014). This indicates that the released and cut RSS could potentially undergo joining with another broken DNA end, leading to chromosome alterations, whilst the

ESC-RAGs complex could potentially trigger further RSS or cRSS cleavage in the genome, leading to numerous double-strand breaks.

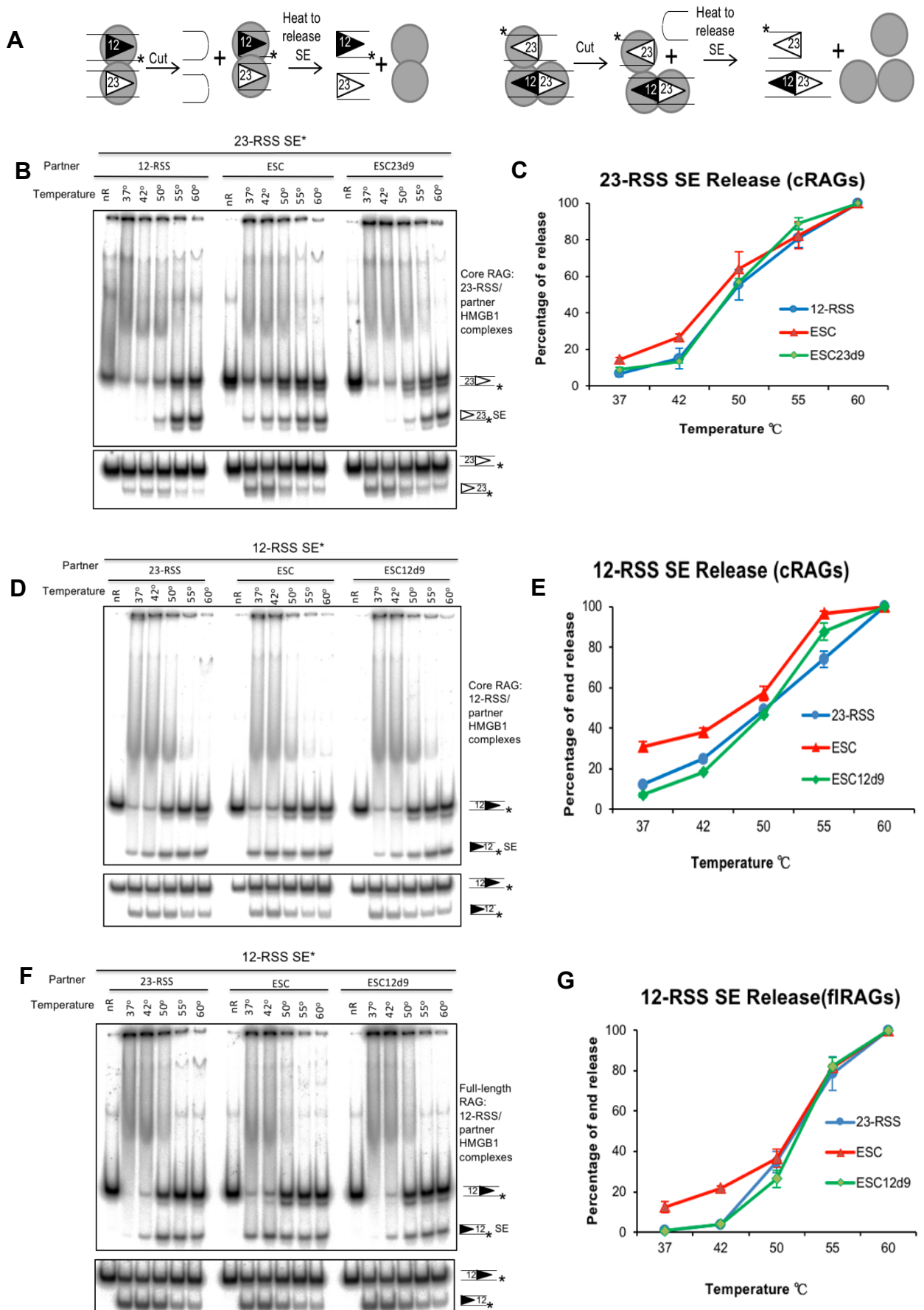


Figure 5.2 Signal ends are released faster following ESC/RSS cutting

The signal end release assay was carried out as a standard *in vitro* RAG cutting assay, using oligonucleotides labelled on the lower strand to detect signal ends. RAG cleavage was performed at 37°C for 15 minutes, followed by incubation at the temperatures indicated for 30 minutes (A). Half of the reaction was loaded on an 8% native gel (B, D and F, upper) whilst the remainder was loaded on a 10% standard cutting gel (B, D and F, lower). The release of SE was then normalised to the level of cutting with core RAG1 plus core RAG2 (B, D) or full-length RAG2 (F). The graphs show the average of three independent experiments; error bars indicate standard deviation (SD).

5.3 The nature of the ends generated by cut-and-run

V(D)J recombination generates two sets of DNA termini: hairpinned coding ends and blunt signal ends (Figure 5.3A). Joining occurs by linking the two hairpin ends and the two coding ends (Zhu et al., 1996). In the cut-and-run reaction, however, it isn't clear what type of ends are generated at the RSS. To investigate the types of ends formed, the ligation-mediated polymerase chain reaction (LMPCR) was used (Figure 5.3B). The LMPCR assay employs a linker primer that is ligated onto blunt DNA ends. After ligation, PCR is performed using a primer specific for target gene and a primer that hybridizes to the linker primer. Because the linker primer is blunt, the broken ends must also be blunt to achieve successful ligation. Hairpin ends cannot be detected by LMPCR as they are unable to ligate to the linker primers. Nevertheless, if the hairpins are opened by Mung bean nuclease (MBN; Kabotyanski et al., 1995), leaving blunt termini, ligation can proceed, allowing the ends to be detected by LMPCR (Figure 5.3B).

An *in vivo* RAG cutting assay was carried out by co-transfection of 23-RSS and ESC plasmid substrates, together with or without full-length RAG1 and core RAG2 expression vectors into NIH/3T3 cells. Plasmid DNA was extracted 48 hours post-transfection. After treatment with or without MBN, samples were analysed via LMPCR and visualized on a 2% agarose gel. As can be seen in Figure 5.3C, standard LMPCR does not detect coding ends in untreated DNA (Figure 5.3C lane 4 upper), although signal ends are readily detectable (Figure 4.7C lane 4 bottom), in agreement with previous studies (Schlissel et al., 1993; Zhu et al., 1996). However, after treatment with MBN and blunting the ends

using T4 DNA polymerase, a 189 bp PCR product of coding ends and a 113 bp PCR product of signal ends were observed, respectively (Figure 5.3C lane 5 upper and bottom). Control experiments, in which the 23-RSS plasmid was digested with *Bam*H I, blunted using T4 DNA polymerase and ligated to linker primers, verified the primers and PCR conditions (Figure 5.3C lane 2, upper and bottom). These results indicate that cut-and-run generates one blunt signal end and one hairpinned coding end. Since NHEJ does not recognize hairpins and the cut ends are released, this suggests that cut-and-run could be a major way of causing orphan DNA breaks during V(D)J recombination, that could lead to chromosome translocations.

In addition, it was notable that after MBN and T4 DNA polymerase treatment, the PCR bands shown on the agarose gel are smeared (Figure 5.3C lane 5 and 6). This is because MBN and T4 DNA polymerase chew back the broken DNA ends to different extents (Chaudhry & Weinfeld, 1995), giving rise to PCR products with different lengths.

During normal V(D)J recombination, after broken ends generation, the NHEJ proteins are recruited to repair them. By contrast, the cut-and-run reaction generates a hairpin coding end and a blunt signal end and it is still not known how these ends are repaired. To test if the ends generated by cut-and-run recruit and bind NHEJ in the same way as the ends generated during normal V(D)J recombination, I attempted to make an Artemis deficient cell line to use in an *in vivo* RAG cutting assay. Here, the level of cutting would be compared between Artemis-deficient cells and wild-type cells to determine if the ends generated by cut-and-run are being properly repaired. If repair proteins are recruited normally to the cut-and-run ends, we predict that the cutting level in Artemis deficient cells will be higher than that in normal REH cells. If not, the cutting level in both kinds of cells should be the same. However, despite several attempts to generate knock-out and knock-down cells (Section 2.51 and 2.52), none gave a stable reduction in Artemis, possibly because this protein is essential.

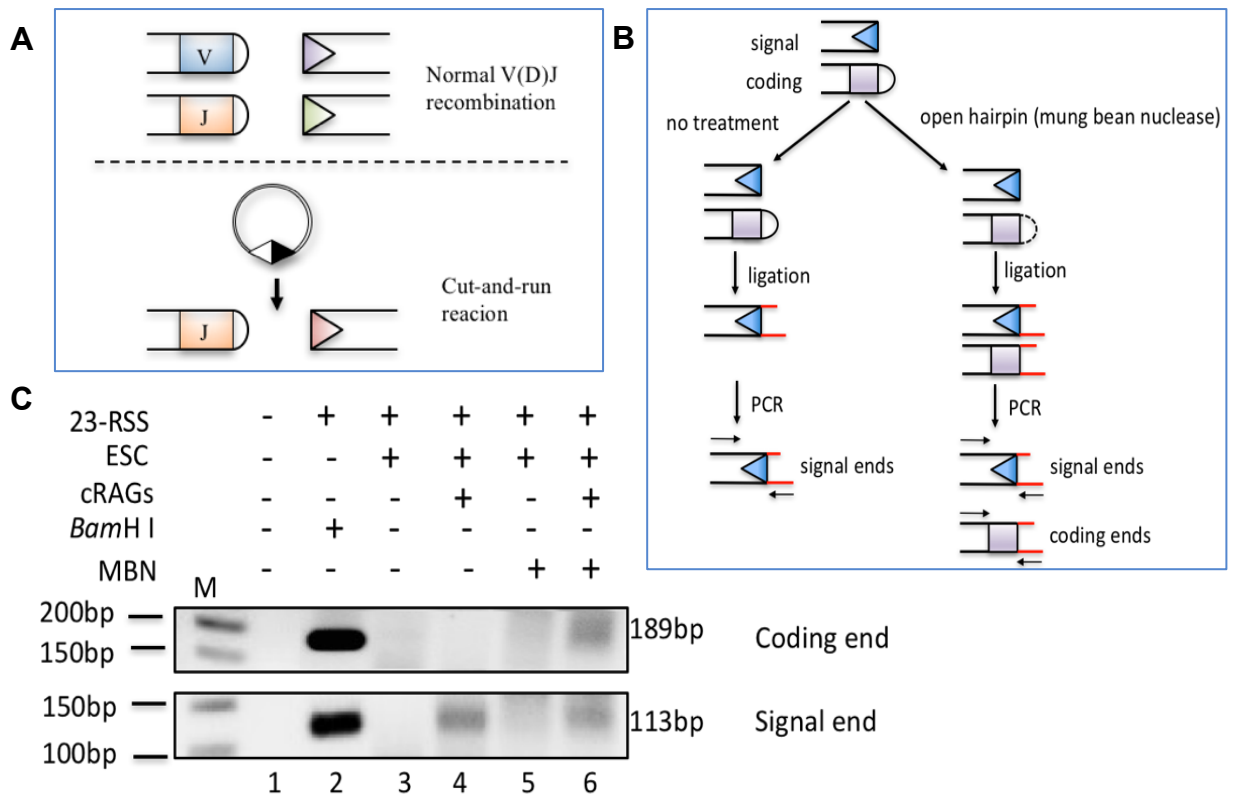


Figure 5.3 The nature of ends generated by cut-and-run reaction

(A) Schematic overview of the products of normal V(D)J recombination and the cut-and-run reaction. In normal V(D)J recombination, hairpins are formed at the coding ends and blunt end are formed at the signal ends and a pair of each type of ends is rejoined by NHEJ pathway. In the cut-and-run reaction, hairpin/blunt ends are likely formed. Symbols are as follows: triangles, signal ends; rectangles, coding ends. **(B)** Schematic diagram of LMPCR-based assay for blunt signal ends and hairpin coding ends. Hairpin coding ends cannot participate in the ligation reaction, and are therefore not amplified. However, opening hairpins by treatment with mung bean nuclease, allows coding ends to be ligated and amplified, giving rise to a product. Symbols are as follows: triangles, signal ends; rectangles, coding ends; arrows, PCR primers; red lines, ligation primers. **(C)** The nature of the ends generated by cut-and-run. To determine the nature of the ends generated by cut-and-run, NIH/3T3 cells were transfected with 23-RSS and ESC substrates plasmids, together with core RAG1 and core RAG2 expression vectors. Plasmids DNA were extracted after 48 hours post-transfection. After treatment with or without MBN, LM-PCR was carried out and the products were visualized on a 2% agarose gel.

5.4 The ESC triggers double-strand breaks *in vivo*

Human cells normally accumulate at least 10,000 DNA double-strand breaks every day (Ohnishi et al, 2009). These are generated by exogenous agents such as radiomimetic drugs and ionizing radiation, as well as natural processes such as V(D)J recombination, stalled replication forks, class switch recombination and reactions that form reactive-oxygen species (Povirk, 2006; Helleday et al., 2007). A prediction of the cut-and-run model is that the ESC should generate DSBs at RSSs and cRSSs in the genome. Both exogenous and endogenous formation of DSBs can be followed by the phosphorylation of the histone, H2Ax. This is a variant of the H2A protein family (Kuo & Yang, 2008). Within a minute of DSBs being formed, γ H2Ax becomes phosphorylated at serine 139, which spreads several Mega bases (Mb) from the break (Rogakou et al., 1998). γ H2Ax foci are readily detectable by immunofluorescence, which I used here to determine if cut-and-run triggers DSBs *in vivo* (Figure 5.4).

To maximise the chances of detecting DSBs, Cos 7 cells were transfected with plasmids carrying an RSS or wild-type or mutated ESC, together with or without RAG1 and RAG2 expression plasmids. The RSSs or ESCs-carrying plasmid includes an SV40 replication origin to cause replication of the plasmid in Cos7 cells and maximise RSS or ESC levels. The presence of DSB, measured by γ H2Ax foci, was compared with control cells where either the RSSs or ESCs or the RAG expression plasmids were omitted. The results show that in the presence of ESC and RAGs, more γ H2Ax foci per cell were detected compared to in the samples containing RSS plus RAGs or the ESC-free or RAGs-free samples (Figure 5.5A, left), suggesting that the ESC can indeed trigger DSBs in the presence of RAGs. This is consistent with the cut-and-run hypothesis. To quantify this, the foci were counted in 450 cells. This showed that only 0-5 foci per cell were detected in 90-100% cells with RSSs or mutated ESC and RAGs or RAGs alone, whilst more than 5 foci per cell were observed in 37% cells transfected with ESC plus RAGs. Indeed, up to 16 foci per cell were detected in 13% ESC plus RAGs transfected cells, resulting an average number of 7.3 foci per ESC plus RAGs transfected cell compared to

nearly zero per RSS or mutated ESC plus RAGs or RAGs alone transfected cell (Figure 5.5A right).

Furthermore, to test if the ESC also triggers DSBs at physiological concentrations of RAGs, REH cells were transduced with an EGFP-expressing lentivirus carrying the ESC sequence or an empty virus as a control, followed by detection of the γ H2Ax foci, indicative of DSBs. In agreement with the results in Cos 7 cells, in the ESC transduced cells, more γ H2Ax foci were detected compared to cells transduced with RSS or mutated ESC or empty virus, with one or more than one foci detected in 37% of ESC transduced cells compared to zero foci observed in 83-86% of RSS or mutated ESC or empty virus transduced cells (Figure 5.5B left and right), suggesting that the ESC can indeed trigger DSBs at physiological concentrations of RAGs. It is notable that fewer γ H2Ax foci were observed in REH cells compared to Cos 7 cells. An explanation of this is only about 3-5 ESC sequences were present per REH cell (Section 2.46) compared to numerous ESCs per Cos7 cell. Overall, these results in Cos 7 and REH cells indicate that the ESC indeed increase the genomic DSBs in the presence of RAG proteins.

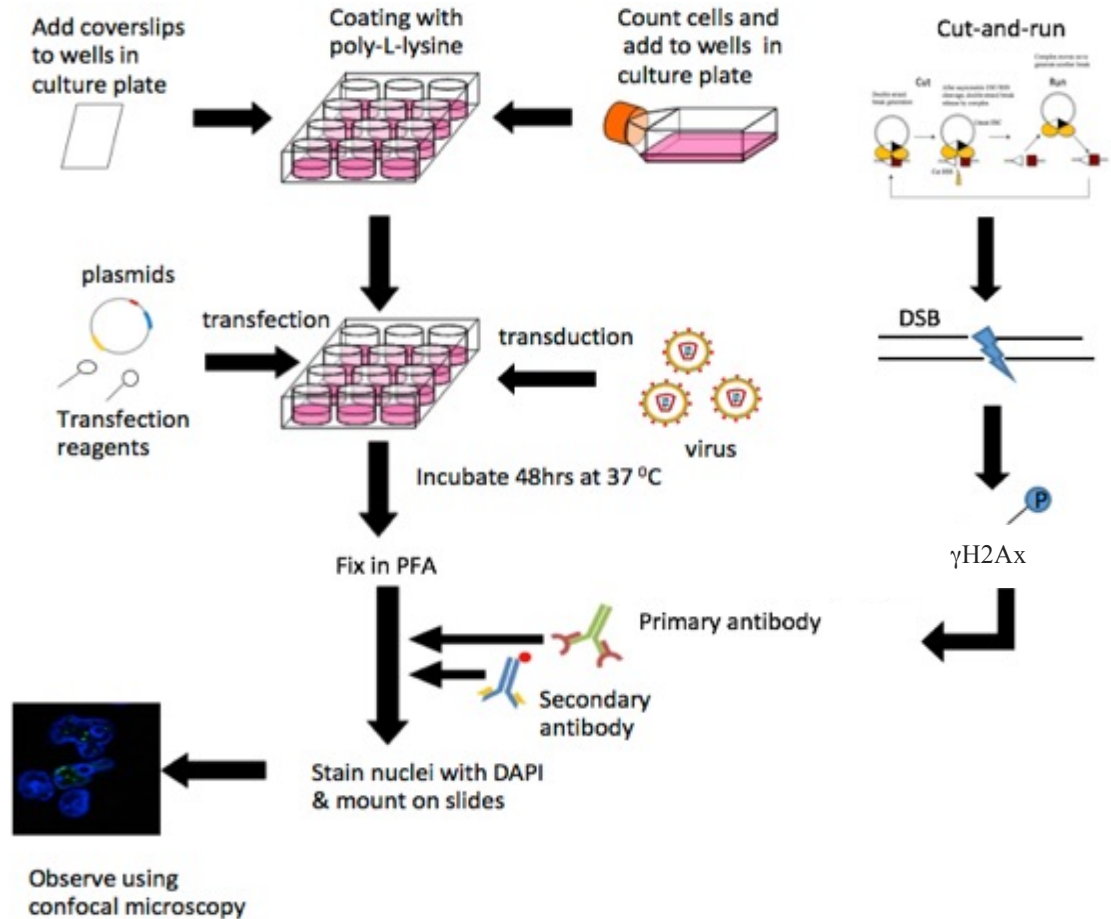


Figure 5.4 Flow chart for γ H2Ax foci detection

To test if ESC triggers double-strand breaks in vivo, Cos 7 or REH cells were transfected or transduced with an ESC-containing plasmid or lentivirus and cultured on coverslips in a 6-well plate. After incubation for 48 hours, cells were fixed in paraformaldehyde (PFA) and hybridized with primary and secondary antibodies. Following this, the nuclei were stained with DAPI and the stained cells were mounted on slides, prior to observation of the images via confocal microscopy.

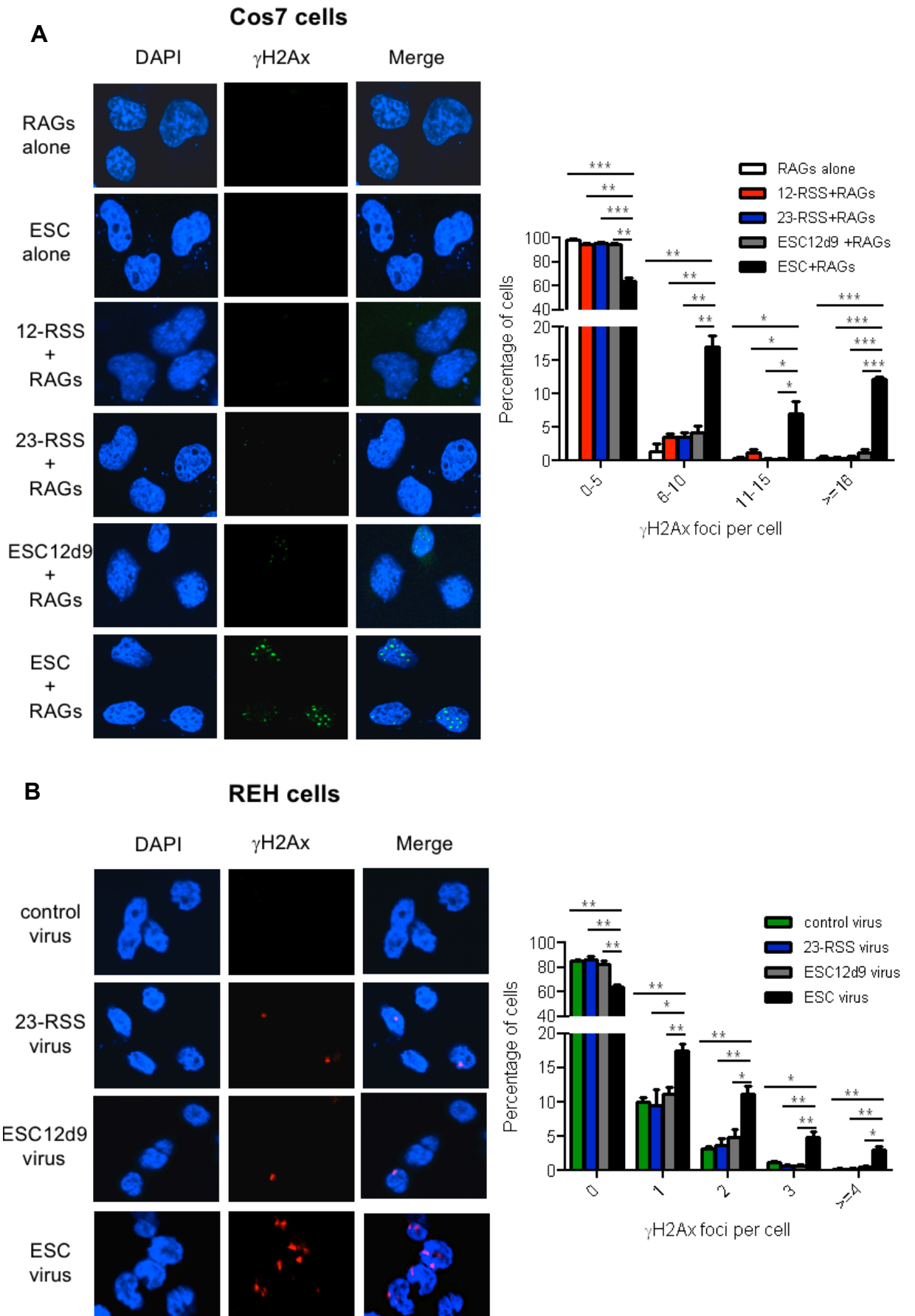


Figure 5.5 Immunofluorescence to detect γ H2Ax foci

(A) γ H2Ax foci were detected in Cos 7 cells transfected with RSS- or wild-type or mutated ESC-carrying plasmid or/and RAG expression vectors. Left: γ H2Ax foci were detected using an AlexaFluor488-labelled (green) secondary antibody Right: Quantification of γ H2Ax foci from 450 cells per transfection condition and three independent experiments. **(B)** γ H2Ax foci were detected in REH cells transduced with EGFP-expressing lentiviruses carrying the wild-type or mutated ESC or RSS or with empty virus. Left: γ H2Ax foci were detected using an AlexaFluor568-labelled (red) secondary antibody. Right: Foci in 450 cells per condition from three independent experiments. Error bars show SD.

5.5 The ESC triggers apoptosis

During the detection of γ H2Ax foci, more cell death was apparent upon transduction with ESC-carrying lentiviruses compared to when the empty virus was transduced. Based on this observation, it was speculated that the ESC could potentially somehow trigger cell death. To verify this phenomenon, REH cells were transduced with lentivirus carrying wild-type or mutated ESC sequence, as well as the empty lentivirus (pWPI) as a control. Seventy-two hours post-transduction, cells were stained with TOPRO, which is only taken up by dead cells, followed by FACs to determine the level of dead cells. As shown in Figure 5.6, the empty lentivirus and mutated ESC-carrying lentivirus caused cell death at a statistically insignificant level, whereas the lentivirus carrying the ESC sequence led to significantly increased cell death (Figure 5.5; Student's t-test, $p=0.0037$ for pWPI vs ESC and $p=0.0005$ for mESC vs ESC). This is consistent with the cell loss observed during the detection of γ H2Ax foci, indicating that the ESC indeed triggers cell death.

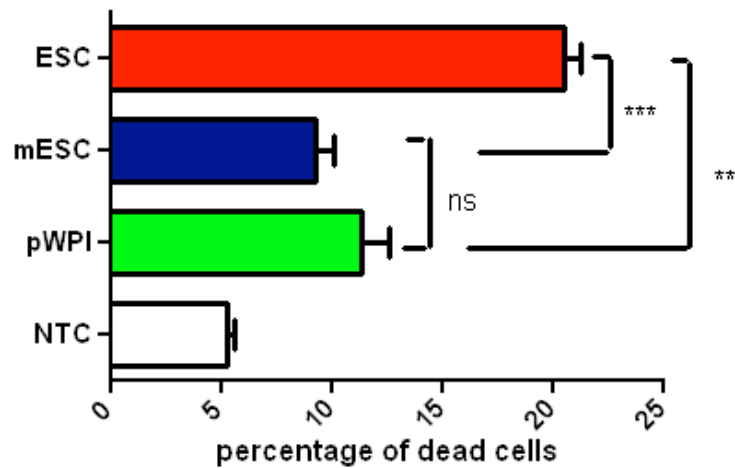


Figure 5.6 The ESC triggers cell death

To determine if ESC can cause cell death, pWPI, wild-type or mutant ESC (mESC) viruses were transduced into human pre-B (REH) cells. 72 hrs post-transduction, cells were collected and washed 3 x with PBS, followed by addition of 0.5 μ l of TOPRO. The percentage of dead cells was detected by FACs. Error bars show the SEM of three independent transductions.

The ESC increases the level of DSBs (Figure 5.5) which potentially can be repaired by two major pathways, termed homologous recombination and non-homologous end joining, respectively (Helleday et al., 2007; Ohnishi et al., 2009). The former is an error-free repair pathway that uses DNA homology to direct DNA repair, whilst the latter is an error-prone repair pathway that religates the broken DNA ends without any requirement for a template but does result in micro additions or deletions (Helleday et al., 2007; Ohnishi et al., 2009). Failure to repair DSBs can result in genomic instability and mutations (Lieber et al., 2006; Povirk, 2006). In addition, unrepaired DNA breaks could also trigger apoptosis via the p53-mediated pathway (Gladdy et al., 2003; Guidos et al., 1996; Helmink & Sleckman, 2012; Perkins et al., 2002; Toledo & Wahl, 2006; Vousden & Lu, 2002). Therefore, it is predicated that ESC triggers cell death by increasing the DSBs, thereby leading to increased apoptosis. To test if this is the case, REH cells were transduced with an EGFP expressing empty or 23-RSS- or wild-type or mutated ESC-carrying lentivirus, followed by FACs to check the infection efficiency and subsequent staining with APC-labelled antibodies against annexin V, an early apoptotic marker, and with propidium iodide (PI) for late apoptotic and necrotic cells. As can be seen in Figure 5.7, even though cells were infected with empty virus (pWPI) and RSS or wild-type or mutated ESC-carrying virus at similar levels, a significantly

increased level of annexin V stained cells were observed in the cells transduced with wild-type ESC-carrying lentivirus compared to those transduced with RSS- or mutated ESC-carrying virus or empty virus (Figure 5.7). This suggests that ESC causes cells to become apoptotic at early stage. These results are in agreement with the observations that more DSBs are generated and there is a greater level of cell death in the presence of the ESC. This is also consistent with the idea that the cut-and-run reaction generates numerous DSBs in the genome.

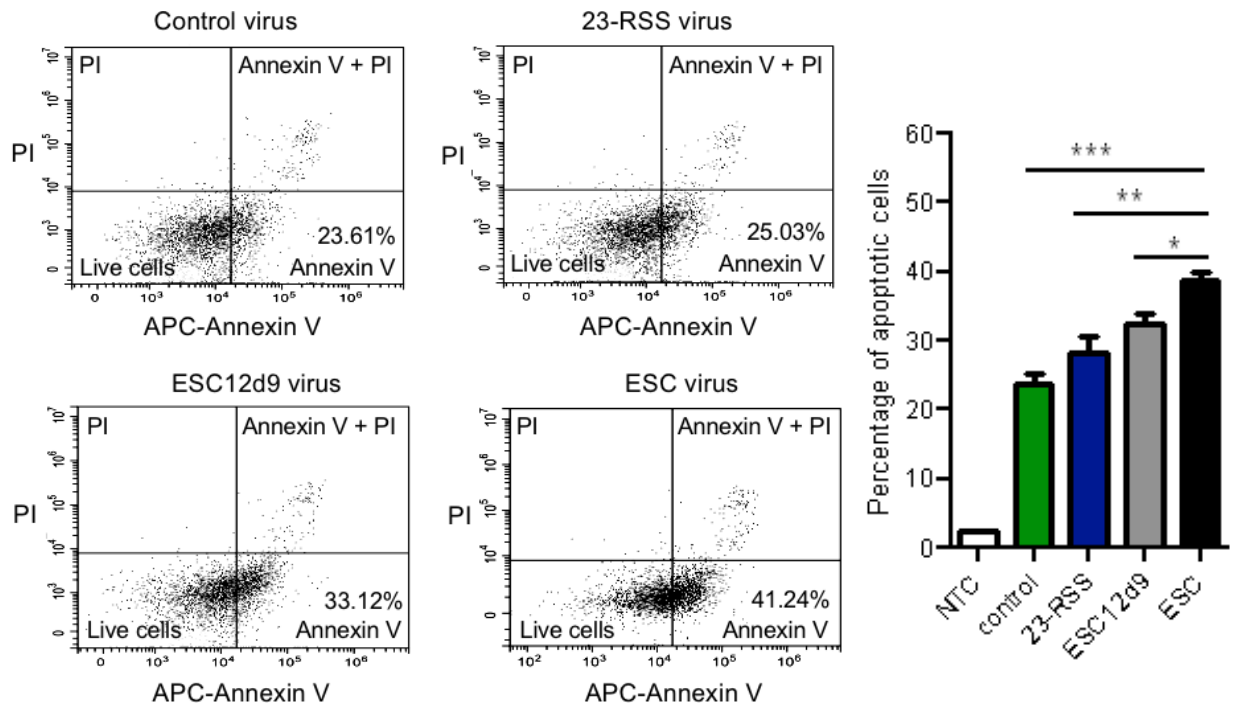


Figure 5.7 The ESC triggers cell apoptosis

Apoptosis in transduced REH cells. Figures show one of three independent experiments. Early stage apoptosis is represented by Annexin V stained cells and late stage apoptosis is indicated by Annexin V and propidium iodide (PI) stained cells. Bar chart shows apoptosis levels at early stage. The data shown are the average number of three independent experiments. Error bars indicate SD. Stars indicate the level of statistical difference by the Student's t-test. * indicates $P < 0.05$, ** indicates $P < 0.01$, *** indicates $P < 0.001$.

5.6 The ESC persists to mature B cell stages

For cut-and-run reaction to occur, the ESC needs to be present at the same time RAG proteins are expressed. The ESC is formed during recombination and generally, one recombination event produces one ESC. Three types of ESC can be generated during the different stages of B cell development: the

IgH ESC, the Ig κ ESC and/or the Ig λ ESC. IgH ESC is generated at pro-B cell stage where the RAG proteins are initially expressed and Ig κ and Ig λ ESCs are formed at pre-B cell stage where the RAG expression becomes up-regulated (Grawunder et al., 1995; Wilson et al., 1994). It was noted that the ESC can persist for at least two weeks in chickens (Kong & Chen-lo, 1999). To test how long the ESC persists in developing B cells, pre-B cells and splenic B cells were used to assess the presence of light chain ESCs. Given that individual ESCs are quite rare, a nested PCR (two rounds of traditional PCR) was employed to detect the ESC and DNA from non-lymphoid cells, NIH3T3, was used as a negative control.

I first detected Ig κ ESC. Notably, at *Ig κ* locus, *V κ* gene segments can be either in forward or reverse orientation, resulting in recombination occurring either by deletion or inversion (Zachau, 1993). If recombination occurs by deletion, the intervening DNA is excised from the genome, forming an ESC. If the recombination occurs by inversion, the signal and coding joints are both maintained in the genome, without giving rise to an ESC (Martinze-Jean et al., 2001). Therefore, a forward orientation *V κ* gene, *V κ 3-1*, and *J κ 5* gene were firstly chosen to check recombination and the presence of the ESC. To detect the ESC generated upon recombination between *V κ 3-1* and *J κ 5* genes and to decrease non-specific PCR, the first round PCR was performed using a pair of primers downstream of RSS flanking the *V κ 3-1* coding gene segment and upstream of the RSS associated with the *J κ 5* coding gene segment. This was followed by a second round of PCR with nested primers to improve specificity. *V κ 3J κ 5* recombination was also determined in a similar way. The results showed that the *V κ 3J κ 5* ESC and *V κ 3J κ 5* recombination could be detected in both pre-B and splenic B cells (Figure 5.8A). Furthermore, the *V κ 3J κ 5* ESC PCR products were verified by Sanger sequencing (Figure 5.8B). Together, these results suggest that the ESC can persist at least to the mature B cell stage.

To further test this, the ESCs generated upon recombination between *V κ 11*, *V κ 16* and *J κ 5* were detected. As for *V κ 3J κ 5*, both the *V κ 11J κ 5* and *V κ 16J κ 5* ESCs were present in pre-B and spleen B cells (Figure 5.8C).

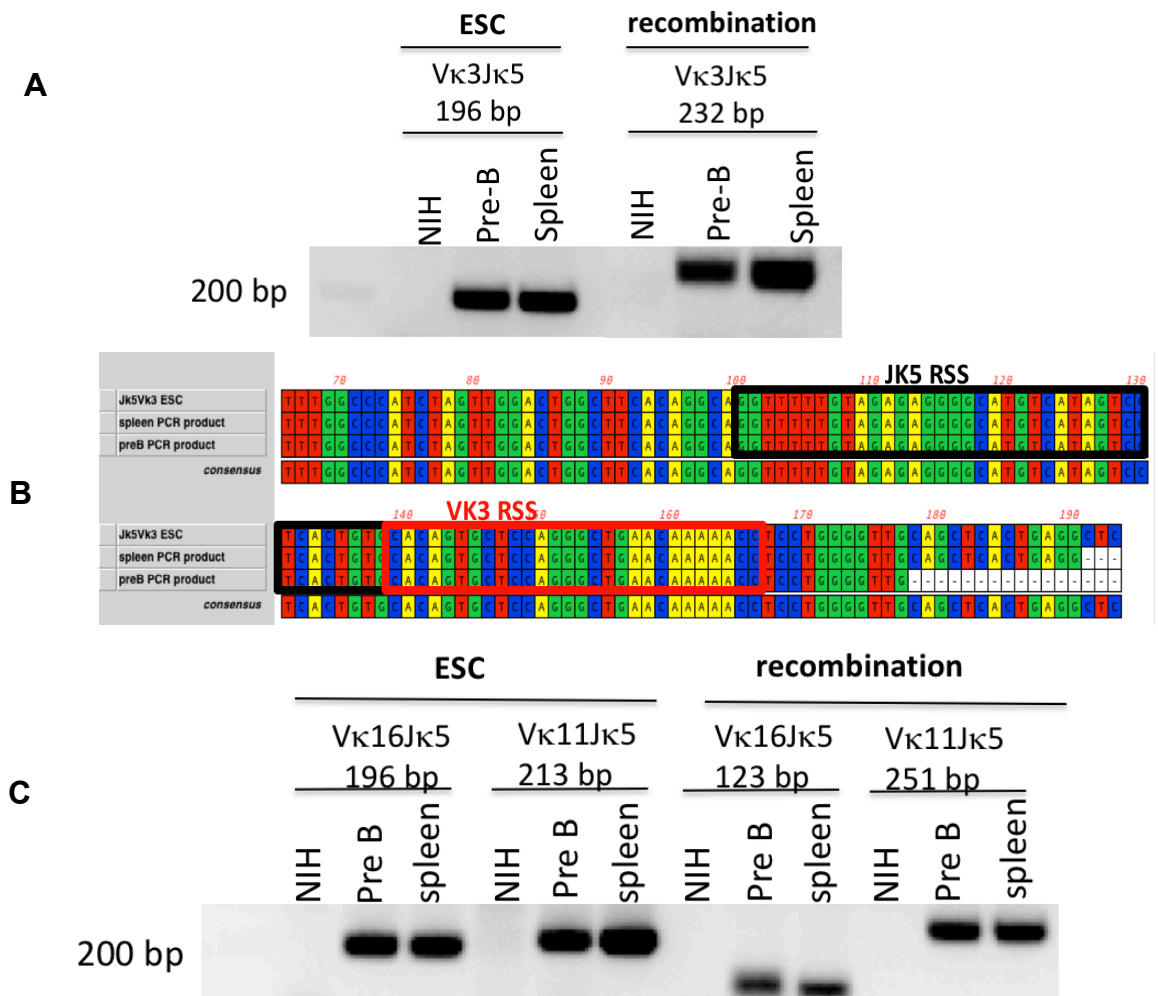


Figure 5.8 The ESC is present in both pre-B and splenic B cells

(A) V κ 3J κ 5 ESC and recombination detection by nested PCR. (B) Sequencing of the V κ 3J κ 5 ESC junctions detected by nested PCR. J κ 5 RSS is shown as a black box and V κ 3 RSS is shown as a red box. (C) V κ 11J κ 5 and V κ 16J κ 5 ESC and recombination detection by nested PCR.

It is notable that more ESCs appears to be present in the splenic B cells compared to pre-B cells. This result suggests that the ESC could potentially replicate in splenic B cells. To test where the replication might occur, B cells at different developmental stages (pre-B, bone marrow IgM⁺, spleen IgM⁺ and spleen IgG⁺) were purified via flow cytometry and genomic DNA and total RNA were extracted. Genomic DNA was used to check the ratio of the ESC and recombination of the respective gene segments, whereas the cDNA was generated via reverse-transcription from the total RNA to check the expression level of RAG proteins.

At the four B cell developmental stages isolated, RAG1 is always expressed whilst with different expression levels. RAG1 was expressed highly at the pre-B stage where light chain locus recombination occurs (Figure 5.9). Subsequently, it dramatically drops to almost 5% and 3% in bone marrow IgM⁺ B cells and spleen IgM⁺ B cells where receptor editing might happen, respectively, relative to the expression level in pre-B cells. Moreover, the RAG1 expression level in spleen IgG⁺ B cells is less than 1% compared to that in pre-B cells (Figure 5.9), which is consistent with previous data (Peng & Eckhardt, 2013). Notably, it was documented that RAG proteins can be re-expressed at late stage of B cell developmental stages for receptor editing or receptor revision (Gay et al., 1993; McGargill et al., 2000; Nemazee, 2006; Tiegs et al., 1993). Therefore, high level RAG expression in IgM⁺ cells might be expected. However, from my result, even though RAG1 is expressed in IgM⁺ cells, the expression level is very low, suggesting that a sub-population of IgM⁺ cells might need to be isolated to detect RAG up-regulation for receptor editing.

RAG1 expression at different B cell developmental stages

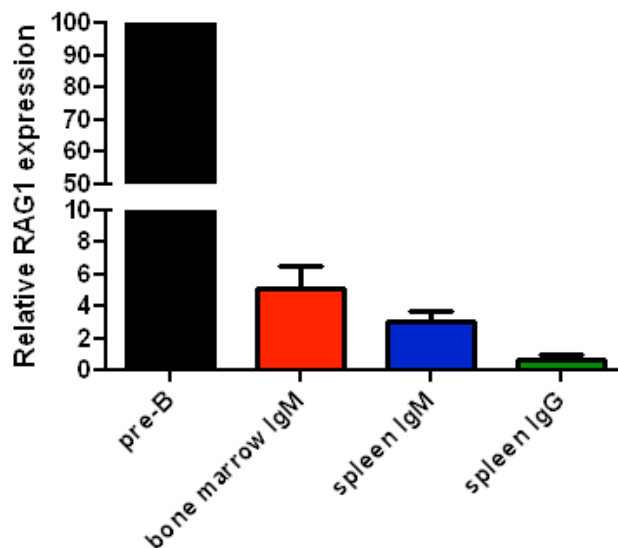


Figure 5.9 RAG1 expression at different B cell developmental stages

B cells at different developmental stages were sorted by FACs. Total RNA was extracted from purified cells and reverse transcribed to cDNA, which was then used for detection of RAG1 expression by qPCR. GAPDH was used to normalize the amounts of cDNA. Error bars show SEM in three independent experiments.

The initial PCR reactions suggested that, more ESC seems to be present in the spleen than in pre-B cells, raising the intriguing possibility that the ESC replicates. Theoretically, if the ESC cannot replicate, it will be diluted with the cell division and the ratio of ESC to recombination should gradually decrease with the cell development. To test the relative levels of the ESC and recombination, the levels of the $V\lambda 1J\lambda 1$ ESC and $V\lambda 1J\lambda 1$ recombination products were determined. This ESC was chosen initially because the $Ig\lambda$ locus is the simplest antigen receptor locus and 70 % recombination occurs between $V\lambda 1$ and $J\lambda 1$, meaning that most $Ig\lambda$ positive cells will have the $V\lambda 1J\lambda 1$ ESC (Boudinot et al., 1995). Using relative quantification, I found that, the ratio of $V\lambda 1J\lambda 1$ ESC compared to $V\lambda 1J\lambda 1$ recombination slowly decreases from the pre-B stage to spleen IgG^+ B cell stage (Figure 5.10 left). This is consistent with lack of replication of the $V\lambda 1J\lambda 1$ ESC .

To test if the $Ig\kappa$ ESCs are also slowly lost during B cell development, the ESC and recombination levels involving $V\kappa 3$, $J\kappa 5$ and $V\kappa 16$, $J\kappa 5$ gene segments were quantified. Contrary to the results with the $Ig\lambda$ locus, the ratios of both $V\kappa 3J\kappa 5$ and $V\kappa 16J\kappa 5$ ESC and recombination steadily increase during B cell development (Figure 5.10 middle and right). The increased ESC in splenic B cells could be explained in several ways: firstly, most of the ESCs might not be fully ligated in pre-B cells and they are gradually ligated during B cell development, since it was reported that ESC remains unligated until RAG expression downregulated (Livak & Schatz, 1996; Ramsden & Gellert, 1995); secondly, some of the ESC could replicate during the B cell development; finally, the recombination products might be lost during B cell development due to secondary rearrangement between the upstream $V\kappa$ gene and downstream RS element in $Ig\kappa$ locus.

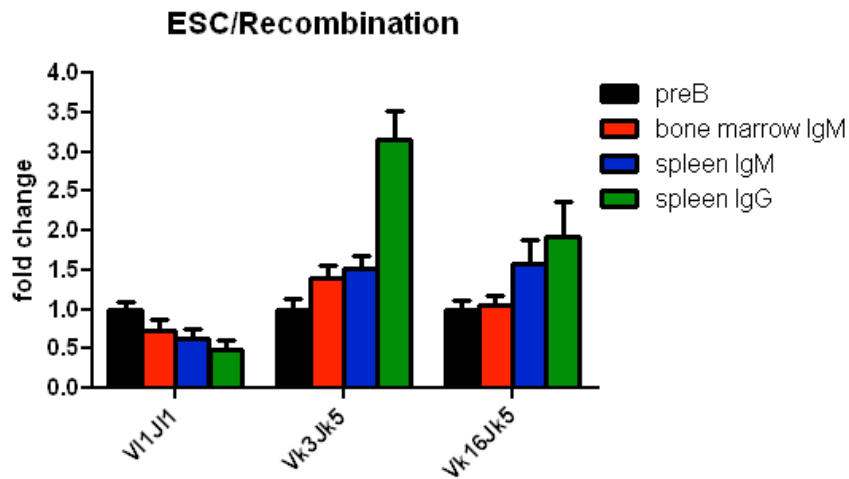


Figure 5.10 Fold changes of ESC/Recombination in different B cell development stages

To test the fold changes of ESC/Recombination, relative quantifications of ESC and recombination were determined using nested PCR. Fold changes show the ratio of ESC/Recombination. Error bars indicate SEM of three independent experiments.

To test if some ESCs remain unligated, the first round ESC and recombination PCR products were gel-purified and the concentrations of PCR products were measured by QuantiFluor™ dsDNA system using standard lambda BAC DNA to generate standard curves of known concentrations of dsDNA. To this end, PCR products of known concentrations were diluted to 1×10^{10} copies/ml and a 10-fold serial dilution ranging from 1×10^4 to 1×10^1 copies/ml was used as standard curves in qPCR to accurately quantify the amount of ESC and recombination at different B cells stages. Consistent with the results of relative quantification, the ratio of Vλ1Jλ1 ESC/recombination decreased from 35% in pre-B cells to 17.5% in spleen IgG⁺ B cells. By contrast, the ratio of Vκ3Jκ5 ESC/recombination increased from 46% to 145%, i.e. more than three-fold, and the ratio of Vκ16Jκ5 ESC/recombination increased from 51% to 99% (Figure 5.11). These results suggest that about 50 - 65% ESCs are unligated in pre-B cells, and it is possible that circular ESCs continue to be generated during B cell development. However, the increase of ESC/recombination in *Igκ* locus raises the intriguing possibility that the ESC can replicate, especially the increase of Vκ3Jκ5 ESC to 145%. In addition, the spleen IgM⁺ cells have undergone somatic hypermutation and many rounds of replications.

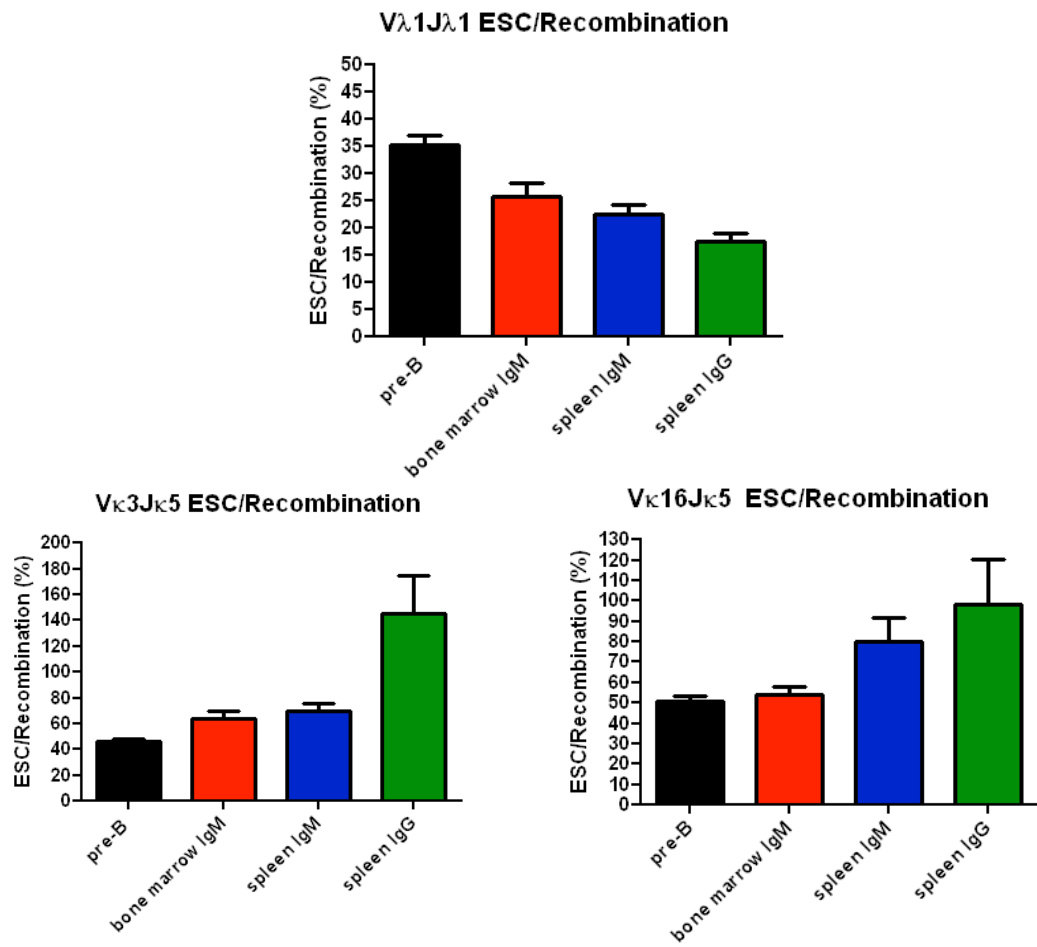


Figure 5.11 Absolute quantification of ESC/Recombination during B cell development

To test if the apparent replication of the ESC in the spleen is in fact due to ESCs not being ligated in pre-B cells, the levels of the ESC and recombination were determined using standard curves of known copy numbers. Error bars show SEM of three independent experiments.

C) Discussion

Cut-and-run has the potential to be a very dangerous reaction, depending on a number of factors. Here, firstly, I showed that the ESC triggers cutting at a cRSS that is associated with the LMO2 protooncogene, which implies that cut-and-run could underpin known leukaemogenic chromosomal translocations. Secondly, I found that cut-and-run generates a blunt signal end and a hairpinned coding end. Since these are not readily recognized by NHEJ machinery, this implies that the broken ends could contribute to genome instability. Thirdly, as predicted by the cut-and-run hypothesis, I showed that this reaction can cause double strand breaks throughout the genome, as indicated by the presence of significantly increased H2A γ X foci in the presence of the ESC plus RAGs compared to the absence of the ESC or RAGs. Finally, three different ESCs were found to persist to the mature B cell stage, implying that cut-and-run reaction could also trigger DNA breaks associated with mature B cells cancers.

5.7 The ESC can trigger cutting at cRSSs

The first factor that affect the dangers of the cut-and-run reaction is the position of double strand breaks formed. Cut-and-run reaction was assumed to generate numerous DSBs *in vivo*. To test this, I performed immunofluorescence assay to detect the γ H2Ax foci which could be phosphorylated and detectable once DSBs are formed (Rogakou et al., 1998). It was found that numerous γ H2Ax foci can be detected in the presence of the ESC and RAGs whilst just a few γ H2Ax foci were detected in the absence of the ESC or RAGs, suggesting that the cut-and-run reaction indeed causes the formation of DSBs. If those DSBs cannot be repaired properly or if they are generated at a dangerous position in the genome, such as proto-oncogene-associated cRSSs, cut-and-run reaction could be extremely dangerous by providing substrates for chromosome translocations via end donation. It has long been known that cRSS resembling an authentic RSS can be potentially targeted by mistake, leading to illegitimate recombination events such as chromosomal deletions, inversions, or translocations (Marculescu et al., 2002).

When such RSSs are located in or near cellular proto-oncogenes, translocation and subsequent upregulation of oncogenes can contribute to malignant transformation (Tycko & Sklar, 1990; Lieber, 1998). In addition, previous studies using extrachromosomal substrates suggest that cRSSs identified near the translocation with breakpoint in the samples such as LMO2, TAL1, Ttg-1 and Hox11 undergo V(D)J recombination with efficiencies ranging from about 30- to 20,000-fold less than *bona fide* V(D)J recombination (Cayuela et al., 1997; Kitagawa et al., 2002). To test if the ESC can trigger cutting at cRSS and contribute to these chromosome changes, I performed an *in vitro* RAG cutting assay with several cRSSs in the presence of the ESC. Significant cutting at LMO2 was observed when it was incubated with either a 23-RSS or an ESC, which suggests that the ESC can indeed trigger cutting at cryptic RSSs. However, there was no obvious cutting observed when Ttg-1 was incubated with the ESC, whilst it could be cut well when the partner was a 12-RSS (data not shown). Notably, even after increasing the incubation time to 1 h, detectable cleavage at two further cRSSs, Notch and Bcl11, was not observed (data not shown). Given that cleavage depends on site-specific RAG recognition of recombination signal sequences, it is likely that this is due to poor RAG binding to some cRSS substrates. Different cRSSs compare more or less well to consensus RSS, even authentic RSSs exhibit some degree of sequence variability in the heptamer and nonamer, and this correlates with how well different cRSSs are cut (Marculescu et al., 2002; Zhang & Swanson, 2008). Furthermore, some investigators found that cutting of LMO2 and Ttg-1 by RAG proteins to be detectable in cell culture by using ligation-mediated PCR, whilst cleavage at the Hox11 and SCL cRSS was not detectable (Zhang & Swanson, 2008). Nevertheless, the fact that cutting at cRSSs catalysed by ESC was observed at the LMO2 cRSS means that ESC could trigger breaks next to some oncogenes in genome, generating the substrates for chromosome translocations, which could pose a great risk to genome stability.

Consistent with this, a study performed in Boyes lab tested if the DNA breaks formed via this cut-and-run reaction indeed underpin chromosome alterations in cancer using ETV6/RUNX1-positive ALL as a model. To this end, James Scott performed the LAM-HTGTS assay where a human pre-B cell line, REH,

was transduced with lentiviruses designed to introduce a CRISPR/Cas9-mediated 'bait' DSB at genomic loci and asked if there are increased translocations of 'prey' DSBs to these sites in the presence of the ESC. The results showed that ESC/RAGs-mediated breaks in REH cells co-localise with 28 chromosome cRSS break sites that are found in ALL patients (unpublished data, Boyes lab; Papaemmanuil et al., 2014), suggesting that cut-and-run could trigger cutting at oncogene-associated cRSSs, significantly contributing to chromosome translocation in ALL.

The amount of damage caused by cut-and-run will likely depend on how well cRSSs throughout the genome correspond to consensus RSSs. The fact that ESC triggers cutting at cRSSs and that the DSBs generated by cut-and-run could be used for chromosome translocation nevertheless suggest that, the cut-and-run reaction is very dangerous by offering DSBs for chromosome alterations. This further suggests that decreasing the presence of the ESC might be contribute to reducing the genome stability and correspondingly decrease the risk of cancer.

5.8 The ends generated by cut-and-run reaction

The second factor that affect the dangers of the cut-and-run reaction is the fate and the nature of the DNA following cleavage. It is well-known that during normal V(D)J recombination, two sets of ends are generated: blunt signal ends and hairpin coding ends, and following cleavage, these ends remain bound by RAG proteins prior to being repaired by the NHEJ DNA repair pathway. The blunt 5' phosphorylated signal ends were proposed to be essential for maintaining the linear integrity of the chromosome (Bassing et al., 2002; Beth, 2012) and the covalently sealed hairpin coding ends are important for antigen receptor gene assembly (Beth, 2012; Gellert, 2002; Schatz & Ji, 2011).

In the newly uncovered cut-and-run reaction, just one coding end and one signal end are generated since the ESC is not cut. If these generated DNA-ends are retained in a RAG post-cleavage complex, they are likely to be repaired by NHEJ pathway even though the RAG proteins are able to mediate transposition of DNA ends containing RSSs (Agrawal et al., 1998; Hiom et al.,

1998). However, if following cutting, the broken RSS is released, RAG proteins would not be able to drive the naked broken end towards the NHEJ pathway for repair (Lee et al., 2004), thereby increasing the risk of chromosome translocations. Chris Kirkham detected the coding end release generated by cut-and-run reaction by a time-course assay and found that following cleavage, the coding ends are released whilst the ESC remains bound by RAG proteins (Kirkham, 2014). To test if this also happens with signal ends, I performed signal end release assay by challenging the post-cleavage complex at different temperatures following RAG cleavage and found that the signal end is released at lower temperatures in the presence of an ESC partner compared to complexes involving either an RSS or mutant ESC partner. These results suggest that the signal ends generated by cut-and-run are released much more readily by RAG proteins, thus making them likely to undergo joining with other DNA ends that might be associated with cancers.

Since both the coding end and signal end generated by cut-and-run reaction are released by RAG proteins, it is important to know the types of the released ends. If both of them are open ends, they could be recognised by Ku protein, followed by recruitment of other proteins in NHEJ pathway to re-ligate the broken end with other open ends in the genome. However, if the signal end is open and the coding end is sealed, the blunt signal end still could be able to be repaired whilst the sealed coding end is left unrepaired, since NHEJ cannot recognise sealed hairpin DNA. To detect the nature of the ends generated at the RSS by cut-and-run, LMPCR was performed. In addition, MBN was used to open any potential sealed ends. The results showed that, without MBN treatment, a PCR product can be observed when signal broken end was ligated to a linker primer but not coding end. The coding end PCR product is only obtained if the sample was treated with MBN to open the hairpin. These results indicate that, cut-and-run reaction generates a blunt signal end and a hairpinned coding end, which can be dangerous since the sealed coding end cannot be repaired and the blunt signal end can be a target for chromosome translocation.

Furthermore, it was found that MBN opens the hairpin structure asymmetrically, since after MBN treatment, a PCR product from coding end still cannot be observed unless T4 DNA polymerase is used to fill in single stranded ends before ligation to a linker primer, which is consistent with previous data by Chaudhry & Weinfeld showing that incubation of 5'-radiolabeled hairpin oligonucleotides with MBN generates a limited number of products resulting from cleavage near the hairpin tip (Chaudhry & Weinfeld, 1995).







5.9 The persistence of ESC

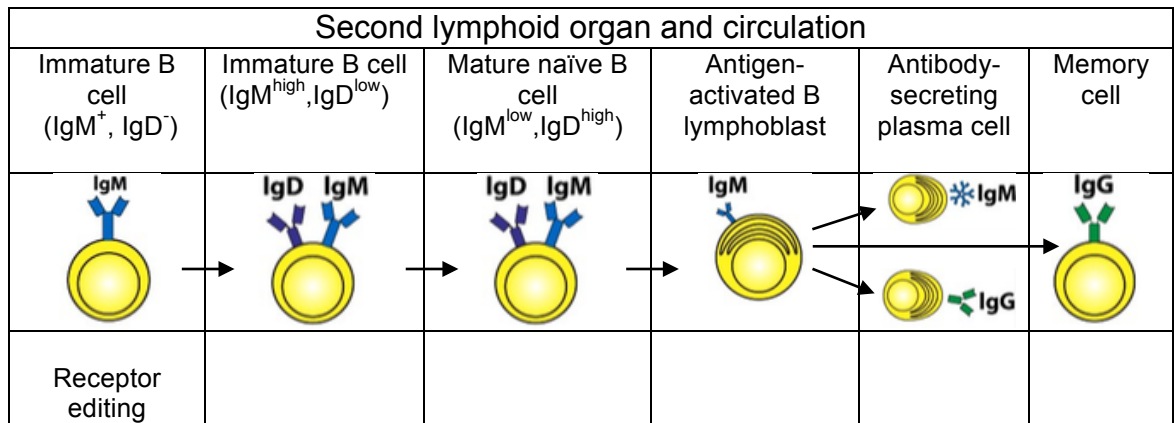
The third factor that affect the dangers of the cut-and-run reaction is the co-existence of ESC and RAGs. ESC is clearly a dangerous structure, associating with cleavage of cRSSs (section 5.1), DSBs generation (section 5.4) and cell apoptosis (section 5.5), etc. Therefore, it is worthwhile to determine how long the ESC persists. To test this, a simple PCR-based method was used. During B cell development, the ESC was previously thought to be gradually decreased by cell division until totally lost. Therefore, B cells at different developmental stages were purified by flow cytometry to determine the presence of the ESC. Notably, I found that the V λ 1J λ 1, V κ 3J κ 5 and V κ 16J κ 5 ESCs can persist at least to mature B cell stage, suggesting that the ESC is a long-lived structure. Since the occurrence of cut-and-run reaction needs ESC and RAG proteins in the same cell, I then performed experiment to detect the RAG expression at different B cell developmental stages. RAG proteins are initially expressed at pro- to pre-B cell stage (Grawunder et al., 1995; Wilson et al., 1994). After formation of a productive product at *IgH* locus, expression of RAG proteins is down-regulated, followed by a second wave of up-regulation at the pre-B cell stage, permitting the rearrangement of the light chain locki, thus forming the lambda or kappa ESCs (Table 5.1).

During the B cell developmental stages from the pre-B stage to spleen IgG⁺ stage, RAG transcripts are exist persistently although the expression levels in bone marrow IgM⁺ and spleen IgM⁺ stages are 20-30-fold less than in pre-B cells, with an even lower expression (about 100-fold lower) in spleen IgG⁺ cells. Nevertheless, these results suggest that the ESC and RAG proteins co-

exist at least to the mature B cell stage, meaning that the dangerous cut-and-run reaction could potentially occur. However, if the concentration of RAGs is too low to bind an RSS or an ESC, the cut-and-run reaction would not be possible. It has been reported that the total RAG concentration in the mammalian pro-B cell nucleus is about 40 nM (Askary et al., 2014) and in the presence of both RAG1 and RAG2 plus HMGB1, RAG1 binds an isolated 12-RSS with a K_d of 4.1 ± 1.6 nM or 8.8 ± 4.0 nM for 23-RSS (Lovely et al., 2015). In addition, I detected the RAG expression at pro- and pre-B cell stages and found that the RAG expression level at the pre-B cell stage is about 1.3-fold higher than that at the pro-B cell stage (data not shown), suggesting that the total RAG concentration in pre-B cell nucleus is about 52 nM. According to these numbers and the RAG expression levels at different B cell developmental stages, it can be calculated that the RAG concentrations in bone marrow IgM^+ and spleen IgM^+ stages is about 2 - 3 nM and less than 1 nM at the spleen IgG^+ stage. This means that it is possible for RAG proteins to bind an RSS at a low level at least in spleen IgM^+ cells, then potentially leading to cut-and-run. Consequently, the co-existence of RAGs and ESC poses a risk for genome stability at least to IgM^+ cells.

Table 5.1 ESC formation and RAG expression during B cell developmental stages

Bone marrow					
Stem cell	Early pro-B cell	Late pro-B cell	Large pre-B cell	Small pre-B cell	Immature B cell
					
μ germline	μ germline	μ DJ	μ VDJ	μ VDJ	μ VDJ
κ/λ germline	κ/λ germline	κ/λ germline	κ/λ germline	κ/λ germline	κ/λ VJ
		IgH ESC	IgH ESC	IgH ESC	Ig λ or Ig κ ESC
		RAG expression	RAG down-regulation		RAG re-expression Receptor editing



During the detection of the ESCs, I found that more ESCs seemed to be present in splenic B cells and thus the amount of ESCs were quantified. Since one recombination event generates one ESC, the amount of recombination can be used to determine the amount of ESC expected per cell. Interestingly, it was found that the lambda ESC gradually decreases as B cell development progresses whilst the kappa ESC appears to steadily increase. The gradual decrease in the lambda ESC suggests that possibly the ESC is not able to replicate. However, if this is the case, why are more and more kappa ESCs detected at later stages of B cell development?

Normally, the λ light chain gene locus is only recombined if there is no productive recombination at the $Ig\kappa$ locus (Hieter et al., 1981; Sun et al., 2012). Correspondingly, in mice, 95% of antibodies include a kappa light chain whilst only 5% contain a lambda light chain (Tallmadge et al., 2014). In addition, $Ig\lambda$ cells will normally have $Ig\kappa$ ESCs since $Ig\kappa$ normally rearranges first. Therefore, it is possible that the lower level of lambda ESCs is hard to accurately detect. Furthermore, one possibility for the presence of gradually increased kappa ESCs is that the kappa ESC is not totally ligated in pre-B cells and continues to be ligated as B cell develop. It was proposed that in normal lymphoid precursors, signal ends stably persist associating with RAG proteins before being joined (Livak & Schatz, 1996; Ramsden & Gellert, 1995; Roth et al., 1992; Schlissel et al., 1993). A necessary requirement for formation of a circular ESC is the removal of the RAG proteins from the signal ends complex (Ramsden & Gellert, 1995). In pre-B cells, RAG protein expression is up-regulated (Brandle et al., 1992; Grawunder et al., 1995; Wilson et al., 1994). Therefore, it is highly possible that the ESC remains

unligated in pre-B cells, which cannot be detected by our PCR-based assays. Following productive IgL rearrangement, RAG expression is down-regulated, thus releasing the broken signal ends and making it possible to ligate and form a circular ESC.

A further reason why the kappa ESCs could persist is the presence of a long non-coding RNA (lncRNA) in all Ig κ ESCs but no similar element in lambda ESCs. Notably, the V λ 1J λ 1 ESC and V κ 3J κ 5 ESC are about the similar size, 21,754 bp compared to 19,708 bp. Nevertheless, a lncRNA, *Gm46961* predicted gene (GeneID: 108169148), is present on all of the kappa ESCs. LncRNAs, which range from 200 bp to several kilobases in size and possess little ability to code proteins, were initially thought to be non-functional by-products of RNA polymerase II transcripts. Indeed, a study performed by Struhl suggests that they are spurious transcriptional noise (Struhl, 2007). Nevertheless, increasing evidence suggests that lncRNAs have a vast diversity of critical functions in cellular development and metabolism, including regulation of the cell cycle, genetic imprinting, chromatin modification, genomic rearrangement, transcription, splicing, translation, and mRNA decay (Huarte et al., 2010; Hung et al., 2011; Kretz et al., 2013; Tsai et al., 2010). A lot of lncRNAs have been found to be associated with gene regulation in mammals at three main levels: epigenetics, transcriptional regulation and post-transcriptional regulation (Guttman et al., 2009; Mercer et al., 2008). It was reported that the act of lncRNA transcription itself can either positively contribute to the expression of adjacent genes by providing a more “open” chromatin domain (Uhler et al., 2007) or silence nearby genes by occupying the transcription factor binding sites (Martens et al., 2004). Interestingly, some investigators found that the promoters of lncRNA are almost as conservative as those of protein-coding genes (Carninici et al., 2005; Derrien et al., 2012; Guttman et al., 2009). Indeed, we found that this lncRNA on *Ig κ* could be expressed as B cells develop but at different levels (data not shown), suggesting that, it is possible that the lncRNA on the kappa ESC might perform critical function that gives cell a survival advantage.

A further explanation for the apparent increase in the Ig κ ESC relative to recombination is the total amount of recombination is gradually decreased.

The kappa recombination product can be lost by secondary rearrangement between V_{κ} gene upstream of that used in primary rearrangement and a downstream RS element in mouse (Durdik et al., 1984; Moore et al., 1985) or κ de in man (Klobeck & Zachau, 1986; Siminovitch et al., 1985) on the same κ allele. Therefore, it is highly likely that the increased ratio of kappa ESC/recombination is due to a combined action of several factors such as continually joined or replicated ESC during B cell developmental stages and gradually decreased amount of recombination products.

Chapter 6:

Discussion

6.1 The occurrence and mechanism of the cut-and-run reaction

Whilst the V(D)J recombination by-product, known as the ESC, was long considered to be inert (Arnal & Roth, 2007), an increasing number of studies have suggested that the ESC might pose a threat to genomic stability, and might contribute to the development of lymphoid cancers. Nevertheless, the mechanism by which it does this is poorly understood. Two important studies in 2007 showed that the ESC can undergo efficient *trans* recombination, becoming re-inserted into the genome (Curry et al., 2007; Vanura et al., 2007). Recently, a new, yet potentially much more frequent reaction, named “cut-and-run”, was uncovered in the Boyes lab (Kirkham, 2014). In this reaction, it was proposed that, the ESC/RAG complex triggers breaks at RSSs whereas the ESC itself is barely cut i.e. an asymmetric RAG cleavage occurs between an ESC and an RSS. Crucially, following RAG cleavage, the double strand DNA breaks generated at RSSs are released, and have the potential to be substrates for chromosome translocation whilst the ESC remains bound by RAG proteins, allowing it to trigger further cleavage at other authentic or cryptic RSSs throughout the genome, thereby leading to genome instability.

The cut-and-run reaction was initially proposed by Chris Kirkham and was investigated *in vitro* using purified consensus RSS and ESC oligonucleotides and core RAG proteins, as well as *in vivo* using extrachromosomal recombination substrates containing non-consensus RSS and ESC (Kirkham, 2014). Since the RAG2 C-terminus was found to inhibit ESC reintegration by 7-fold (Curry et al., 2007), it begs the question of whether the RAG2 C-terminus also inhibits the cut-and-run reaction. To test this, core RAG1 and full-length RAG2 were co-expressed and purified. By comparing of the cleavage activity at an RSS and an ESC by core RAG1 plus core RAG2 with core RAG1 plus full-length RAG2, I found that asymmetric RSS/ESC cleavage occurs in both cases (Figure 3.2). In addition, extrachromosomal substrates

containing consensus RSS and ESC were employed to investigate cut-and-run reaction in non-lymphoid cells, together with core or full-length RAG expression plasmids. No matter whether core or full-length RAG proteins were used, the consensus 12- or 23-RSSs were efficiently cut whilst the consensus ESC remains uncut (Figure 4.3). Furthermore, this asymmetric cleavage was also observed in a lymphoid cell line, REH cells (Figure 4.6), where the RAG proteins are expressed at physiological levels (Bories et al., 1991; Rosenfeld et al., 1997). Together, these results suggest that cut-and-run reaction occurs both *in vitro* and *in vivo*, even at physiological concentrations of RAGs. In addition, given that asymmetric cleavage occurs with core and full-length RAG proteins, cut-and-run reaction is not thought to be inhibited by RAG2 C-terminus either *in vitro* or *in vivo*. Since the cut-and-run is a cleavage reaction and does not involve a joining step, whilst the reintegration involves the end-joining by the NHEJ pathway, the reason why the RAG2 C-terminus has an effect on reintegration but not on cut-and-run reaction could be because RAG2 C-terminus influences the end-joining but not the cleavage step. This is consistent with the studies showed that the RAG post-cleavage complex remains more stable with core RAG2 compared to full-length RAG2 (Figure 5.2; Coussens et al., 2013; West et al., 2005).

Asymmetric cleavage is a fundamental characteristic of the cut-and-run reaction. In a single or paired RSSs/RAG complex, the contact of the RAG proteins to an RSS extends 12 or 16 bp into the coding flank, respectively (Nagawa et al., 2002; Nagawa et al., 2004). The ESC is formed by two RSSs in a head-to-head configuration and the second RSS, therefore, works as the coding flank of the first RSS. Consequently, the asymmetric RSS/ESC cleavage was proposed to be due to RAG proteins simultaneously binding to both RSSs in the ESC, thereby blocking the cutting site at RSS/RSS boundary. The *in vitro* DNase I footprint assay on ESC confirmed this hypothesis by showing that both RSSs in the ESC are protected by RAG proteins from DNase I digestion (Figure 3.5). Furthermore, bandshift assays showed that upon formation of the ESC/RAG complex, an extra higher molecular weight complex was observed compared to the 12- or 23-RSS/RAG complex. Notably, the mutated ESCs formed complexes in which the intensity of this

upper complex was decreased (Figure 3.11, 3.12). In addition, a supershift assay using tagged RAG proteins showed that both RAG1 and RAG2 are contained within this higher molecular weight complex (Figure 3.6). These results are consistent with the idea that both RSSs in the ESC are simultaneously bound by RAG proteins and suggest that mutation of one RSS in the ESC can eliminate RAG binding to ESC. In agreement with this, symmetric RAG cleavage at RSS and ESC can be restored by mutation of one RSS, in either the heptamer or the nonamer, in the ESC both *in vitro* (Chapter 3) and *in vivo* (Chapter 4). Furthermore, increasing the spacer between the two RSSs in the ESC was also found to release the cutting site for both RSSs in the ESC, helping to partially restore symmetric RAG cleavage within an RSS/ESC pairing (Figure 3.10). Together, these results favour the hypothesis that asymmetric RSS/ESC cleavage is due to RAG binding to both RSSs in the ESC at the same time, thus preventing the cutting of the ESC.

Two previous studies have investigated the mechanism of ESC cleavage. Vanura et al showed that an ESC is cleaved via the normal nick-hairpin mechanism (Vanura et al., 2007), which is consistent with the studies performed by the Boyes lab (Kirkham, 2014), whilst Neiditch et al proposed that an ESC was cleaved via an alternative “nick-nick” mechanism (Neiditch et al., 2002). Neiditch et al showed that the RAG proteins firstly bind to one RSS within the ESC and generate a single-strand nick between the two heptamers of the ESC. Then, the recombinase complex bound to the other RSS within the ESC nicks the opposite strand, thereby forming two DSBs with blunt ends (Neiditch et al., 2002). These two alternate models might be due to the different methods used. Vanura et al used PCR to detect ESC reintegration products, whilst Neiditch et al used LMPCR to detect broken ESC ends. Because LMPCR is a very sensitive method, ESC cleavage via the “nick-nick” mechanism was possibly overestimated.

In recent cryo-EM structure studies, Ru et al showed that after binding to intact DNA, the first RAG1-RAG2 monomer cradles the coding flank and heptamer, whilst the second monomer interacts with the heptamer and the spacer. The two NBDs together interact with the nonamer, changing the RAG dimer to a

more open conformation. Subsequently, the intact RSS substrates are melted at the heptamer-coding flank junction by the piston-like inward and outward movement of the ID (especially the $\alpha 15$ - $\alpha 16$ loop) and corkscrew DNA rotation, contributing to positioning of the scissile phosphate in the active site for nicking (Rue et al., 2018). Since two RSSs within the ESC could be bound by RAG proteins, the change of the RAG dimer conformation could be blocked, thereby influencing the melting and nicking of CAC/GTG base pairs in both heptamers in the ESC and consequently inhibiting the cleavage of the ESC.

6.2 Dangers posed by cut-and-run

Cut-and-run was proposed to be a very dangerous reaction by causing RAG cleavage at authentic or cryptic RSSs throughout the genome, generating numerous DSBs and thus presenting a risk to genome stability. However, the threat to genomic stability posed by this reaction is almost totally reliant on four factors: firstly, the stability of the post-cleavage complex in the cut-and-run reaction; secondly, the frequency of the cut-and-run reactions occurring at potential genomic cryptic RSSs adjacent to protooncogene; thirdly, the fidelity of the end-joining machinery; and lastly, how often RAG proteins and ESC co-exist.

In this thesis, signal end release assays showed that the post-cleavage complex in the cut-and-run reaction is much more unstable than that formed during normal V(D)J recombination (Figure 5.2), contributing to an increased risk of chromosome translocations by releasing broken DNA ends. In addition, an *in vitro* RAG cleavage assay showed that ESC can trigger cutting at the LMO2 cRSS (Figure 5.1), which was previously shown to be associated with a chromosome translocation to the *TCR α/δ* locus (Boehm et al., 1988; Rabbitts et al., 1997). RAG cleavage at other cRSSs, such as Ttg-1, was at an undetectable level in the presence of the ESC. This is possibly because of the poor recognition and binding by RAGs to some cRSSs, since different cRSSs match the consensus sequence to different levels, which then impacts on how well different cRSSs are cleaved (Marculescu et al., 2002; Zhang & Swanson, 2008). These results nevertheless suggest that, cut-and-run can occur within a

cRSS/ESC pairing, resulting in broken cRSS ends. Importantly, when these cryptic RSSs are present in locations such as near cellular proto-oncogenes, their cleavage could potentially contribute to chromosome translocations and malignant transformation (Lieber, 1998; Tycko & Sklar, 1990).

Consistent with the cutting at cRSSs, the increased level of γ H2Ax foci in the presence of the ESC and RAG proteins both in non-lymphoid (Cos7) and lymphoid (REH) cells (Figure 5.5) strongly suggests that the ESC/RAG complex could synapse with genome RSSs or cRSSs, giving rise to a number of DSBs. Importantly, the LAM-HTGTS assay was recently performed in the lab by James Scott and this showed that these ESC/RAGs-mediated breaks co-localise with twenty-eight chromosome cRSS breakpoints found in ALL patients (unpublished data, Boyes lab; Papaemmanuil et al., 2014). This provides strong evidence for the dangers of the cut-and-run reaction which potentially triggers cutting at oncogene-associated cRSSs to contribute to genome instability.

Furthermore, the LMPCR assay demonstrated that the cut-and-run reaction generates a blunt signal end and a hairpinned coding end at the RSS (Figure 5.3). These were shown to be released faster by RAG proteins after cleavage compared to the ends generated by the normal V(D)J recombination reaction (Figure 5.2; Kirkham, 2014). The ends generated by cut-and-run are unlikely to be repaired by NHEJ pathway since NHEJ proteins cannot recognize sealed hairpin DNA ends and RAG proteins cannot drive single broken ends to the NHEJ pathway (Lee et al., 2004). Therefore, these broken DNA ends could be repaired by an error-prone process such as alternative NHEJ machinery, disproportionately contributing to the oncogenic genome rearrangement (Gigi et al., 2014), or could be joined to other broken ends via NHEJ, leading to a translocation. Interestingly, it was found that the use of the alternative NHEJ pathway is highly context-dependent: signal ends are more likely to be repaired in this way compared to coding ends (Gigi et al., 2014). This is possibly because the coding ends need to undergo several further processing steps prior to joining. In addition, failure to repair DSBs was reported to lead to genomic instability and mutations (Lieber et al., 2006; Povirk, 2006). However,

several studies showed that unrepaired DSBs could trigger apoptosis via the p53-mediated pathway, which would thus eliminate cells with these dangerous lesions (Gladdy et al., 2003; Guidos et al., 1996; Helmink & Sleckman, 2012; Perkins et al., 2002; Toledo & Wahl, 2006; Vousden & Lu, 2002). Indeed, I found that the accumulated RAGs/ESC-mediated DSBs triggers cell death (Figure 5.6) and contribute to increased apoptosis (Figure 5.7). Together, these findings suggest that cut-and-run is a very dangerous reaction that could lead to genome instability.

6.3 The likelihood of cut-and-run occurring B and T cells

The RAG cleavage assays performed in non-lymphoid and lymphoid cells suggest that cut-and-run occurs *in vivo*. However, these assays were carried out either by transfection of exogenous substrates into non-lymphoid cell line or by transduction of lentivirus carrying RSS and ESC sequences into a lymphoid cell line. It is not known if cut-and-run actually occurs when all of the components needed for this reaction are already within the cells. *In vivo*, the occurrence of cut-and-run is fairly difficult to detect, since it leaves no trace except the broken ends which would also possibly be generated by normal V(D)J recombination rather than the ESC. Nevertheless, several facts indicate that it might occur in B and T cells.

A prerequisite for the occurrence of cut-and-run reaction is the co-existence of all the components needed for this reaction. There are four essential components: an intact ESC, an authentic or cryptic RSS, as well as RAG1 and RAG2. There are over 3000 RSSs in the antigen receptor loci and about 10 million cRSSs throughout the genome (Lewis et al., 1997). However, not all of the cRSSs can be recognised and cleaved RAG proteins, as some of them are either not located in open chromatin or have already been mutated (Teng et al., 2015). The large number of them, nevertheless, provides a good chance for RAGs to bind and cut at least a fraction of them. In addition, the increase of H2A γ X in both non-lymphoid and lymphoid cells in the presence of the ESC suggests that ESCs are capable of synapsis with genomic RSSs or cRSSs to

trigger double-strand breaks. Therefore, if intact ESCs are present at the same time as RAG proteins, cut-and-run reaction is likely to occur.

It is worth considering when RAGs and the ESC might coexist during B cell development. To tackle this, three factors need to be addressed. Firstly, the persistence of endogenous ESC needs to be checked. By checking the persistence of ESCs at different B cell developmental stages (pre-B, bone marrow IgM⁺, spleen IgM⁺ and spleen IgG⁺), I found that the ESC persists to at least mature B cells (spleen IgG⁺), at reasonable levels (Figure 5.10). Numerous previous investigations suggested that following RAG cleavage, a pair of signal ends are tightly bound by RAG proteins and unlikely to be joined to generate a circular ESC until the RAG proteins are downregulated (Agrawal & Schatz, 1997; Jiang et al., 2004; Livak & Schatz, 1996; Zhu et al., 1996). However, analysis of endogenous lambda and kappa ESCs in mouse pre-B cells suggests that about 35 - 50% of the ESCs are intact i.e. circular shortly after they are formed (Fig. 5.11).

Secondly, the expression of RAG proteins during B cell development needs to be investigated. In this study, RAG1 was shown to be expressed at different B cell developmental stages, albeit at a relatively low level at late B cell stages (Figure 5.9). As discussed in Chapter 5, by estimating the concentration of RAG proteins at different B cell developmental stages, it appears it should be possible for RAGs to bind an RSS at least in spleen IgM⁺ cells. Crucially, however, these estimates were made from RNA levels. It is vital to look via western blot at actual protein levels. Such experiments were not possible, however, due to lack of reliable RAG1 antibodies. It is also vital to examine RAG2 levels as both proteins need to be expressed for cutting.

Thirdly, the co-persistence of intact ESC and RAG proteins during B cell development needs to be mapped. There are two cases where the intact ESC and RAG proteins might co-exist. In the first case, if the first rearrangement on one allele of an immunoglobulin or *TCR* locus generates an unproductive product, the second rearrangement will occur on the other allele (Rajewsky, 1996). Therefore, the RAG proteins will co-exist with an ESC from the first

unproductive rearrangement in most lymphocytes. In the second case, following rearrangement of the *IgH* or *TCR β* loci in pro-B or pro-T cells, *IgL* or *TCR α* loci rearrangement occurs at a subsequent stage of development in pre-B or pre-T cells (Bassing et al., 2002). Therefore, the RAG proteins that are up-regulated for rearrangement of the *IgL* or *TCR α* loci will co-exist with the ESC formed during *IgH* or *TCR β* locus rearrangement. Consequently, in these two cases, cut-and-run is highly likely to occur. In addition, RAG proteins were reported to be re-activated for receptor editing to change the antigen receptor specificity in immature B cells if autoreactive antigen receptors are expressed in B cells (Gay et al., 1993; McGargill et al., 2000; Nemazee, 2006; Tiegs et al., 1993). In this case, the re-activated RAG proteins can co-exist with the ESC that generated in the previous rearrangement, resulting in the likelihood for cut-and-run occurring. Furthermore, the increased level of ESC and presence of RAG proteins in late B cells suggests the RAG proteins might be able to bind an RSS and an ESC in mature B cells, thus giving the opportunity for the cut-and-run reaction to occur, potentially resulting in mature B cell cancers.

Further experiments need to be performed to verify the co-existence of circular ESC and RAG proteins. Firstly, the presence of a circular ESC must be verified by the isolation of extrachromosomal circular DNA (eccDNA) from spleen IgG^+ cells via caesium chloride-ethidium bromide (CsCl-EB) density gradient centrifugation, followed by a PCR-based assay to determine the presence of specific circular $Ig\kappa$ or $Ig\lambda$ ESCs. Secondly, the expression of both RAG1 and RAG2 proteins must be determined by western blot at later B cell stages. Thirdly, RAG binding to the ESC also needs to be determined. RAG binding to an RSS is a concentration-dependent reaction (Ciubotaru et al., 2003; Lovely et al., 2015; Zhao et al., 2009). If the concentration of RAG proteins at late B cell stages is lower than that needed for binding, the cut-and-run reaction is unlikely to occur even though the ESC and RAGs co-exist. According to the studies in this thesis, the RAG concentration in late B cells, such as spleen IgG^+ , is about 1 nM. This implies a small amount of binding is possible, since the K_d for RAG protein to a 12- or 23-RSS is about 4.1 ± 1.6 nM or 8.8 ± 4.0 nM, respectively (Lovely et al., 2015). However, since ESC is

composed by two RSSs, a higher RAG concentration might be required. This binding could be tested by a ChIP assay using an anti-RAG1 antibody. Following ChIP, PCR-based assay could be performed to identify the persistence of specific ESC by analysing the recovered DNA with primers which are specific to the ESC. If the ESC and RAG proteins persist at the same time and the ESC is bound by RAG proteins, the intact ESC will be pulled down with RAGs by anti-RAG1 antibody and a specific PCR product will be obtained. Notably, specific primers for endogenous 12- or 23-RSS should be used to check RSS/RAG binding for a positive control to make sure the ChIP assay has been successful. Unfortunately, to date, even though a highly effective anti-RAG1 antibody was generated by the Schatz Group (Little et al., 2013; Subrahmanyam et al., 2012), good quality commercial anti-RAG1 antibodies are not available. Therefore, these experiments could currently only be performed by tagging RAG1 in a cell line.

6.4 Cut-and-run likely provides substrates for chromosomal translocations

During antigen receptor rearrangement, a number of mistakes could occur. Chromosome translocations are one of the errors caused by aberrant RAG activity, which are also the hallmark of many leukaemias and lymphomas (Kuppers, 2005). One of the most common forms of translocation caused by erroneous V(D)J recombination is end donation (Kuppers, 2005; Roth, 2003). This involves a broken DNA end generated by RAG cleavage being joined to another random DSB caused by an unrelated mechanism, such as ionizing radiation (Lieber et al., 2010). However, it is not known how a single broken RSS end is generated, since during V(D)J recombination, paired broken ends are repaired by the NHEJ pathway and no single broken ends remain. The newly-uncovered “cut-and-run” reaction might provide an explanation for how single broken RSSs are generated.

The cut-and-run reaction generates a blunt signal end and a hairpinned coding end, which are unlikely to be repaired normal NHEJ pathway. Therefore, it can potentially cause accumulation of numerous double-strand breaks by triggering

cutting at authentic or cryptic RSSs throughout the genome (Chapter 5). There are millions cryptic RSSs in the human genome, with a distribution frequency in every two kilobases (Lewis et al., 1997), and some non-canonical cryptic RSSs have been found to be involved in chromosomal translocations via end donation (Kuppers, 2005; Papaemmanuil et al., 2014; Roth, 2003). Consistent with this, a recent study by Rommel et al showed that, broken RSS and cRSS ends are employed as substrates for translocation to an engineered DNA break, with a high frequency (Rommel et al., 2017). Therefore, it seems highly likely that cut-and-run produces the broken RSS and cRSS ends, which can serve as substrates for end donation, potentially leading to chromosomal translocation.

A specific example where the cut-and-run reaction could be involved in chromosomal translocations is ALL where patients carry the *ETV6/RUNX1* (*TEL/AML1*) translocation. This translocation normally arises *in utero*, but it cannot lead to malignancy without further mutations. Instead, the *ETV6/RUNX1* translocation stalls B cell development at the pre-B cell stage (Tsuzuki et al., 2004). Importantly, at this stage, RAG proteins are highly expressed, thus providing the opportunity for secondary recombination reactions. Papaemmanuil et al identified 140 chromosome breakpoints that lie in close proximity to RSS-like sequences via whole genome sequencing studies of *ETV6/RUNX1*-positive ALL patients and found that most of these aberrant recombination reactions appear to take place at cRSSs which are adjacent to transcriptionally active B cell development genes. These are potentially cleaved by RAG proteins in the presence of the ESC, thus eventually leading to full malignant transformation (Papaemmanuil et al., 2014). It has been suggested that about half of the RAG-mediated chromosome translocations have an RSS or cRSS on only one side of the translocation (Papaemmanuil et al., 2014). Crucially, exactly in agreement with this, a more recent LAM-HTGTS study performed using the human pre-B cell line, REH, in the Boyes lab showed that cutting occurs at cRSSs and that broken cRSSs are translocated to a DSB with a significant higher frequency in the presence of the ESC. In addition, twenty of translocation breakpoints that are generated in the presence of the ESC, map to exactly the same cRSS

breakpoints as found in twenty-eight ALL patients (Unpublished data in Boyes lab; Papaemmanuil et al., 2014). These data therefore strongly imply that cut-and-run reaction could give rise to single RSS or cRSS breaks which can subsequently be the targets for chromosomal translocation, suggesting that cut-and-run makes a significant contribution to mutation of cancer genes and to chromosome alterations in ALL.

6.5 Cut-and-run and reintegration

As described in the introduction, the ESC is not only the main substrate for cut-and-run, but it is also involved in other reactions such as reintegration. Both cut-and-run and reintegration reactions are very dangerous. Therefore, one may ask which of these reactions is likely to be more frequent and dangerous.

Theoretically, reintegration events do not require ESC to be ligated whilst cut-and-run reactions require an intact ESC. Every normal V(D)J recombination is coupled with the generation of an unligated ESC. However, the formation of a ligated ESC is thought to involve the downregulation of RAG expression (Jones & Gellert, 2001). In this thesis, I found that in pre-B cells, about 35 - 50% of the ESCs are intact, which means 50 - 65% of ESC could undergo reintegration and the remainder might undergo cut-and-run. In addition, the cut-and-run reaction could be prevented by ESC reintegration since reintegration of an ESC is unlikely to be a reversible process. This is because one of the RSS within the ESC is treated as a coding end where non-templated base pairs are added at the pseudo-hybrid joint, thus rendering that RSS useless. Therefore, the frequency of reintegration events could be obviously higher than cut-and-run events. However, the reintegration events have been found to be inhibited by the RAG2 C-terminus (Curry et al., 2007) whilst the cut-and-run events are not influenced by the RAG2 C-terminus (Figure 3.4 and 4.3). From this aspect, ESC reintegration is less likely to occur than the cut-and-run reaction. In addition, once an ESC is ligated and bound by RAGs, cut-and-run reactions are more likely to happen than reintegration since the ESC is not cleaved and can continually trigger DSBs at RSSs or cRSSs throughout the genome. Therefore, asymmetric cleavage, coupled with

the inhibition of ESC reintegration by the RAG2 C-terminus, implies that the frequency of the cut-and-run reaction possibly occurs at much higher levels than ESC reintegration reactions.

Determining which of the cut-and-run reactions and ESC reintegration reactions is more dangerous is not easy. Reintegration of an ESC could possibly upregulate oncogene expression if it takes place at a cryptic RSS which is adjacent to this oncogene. ESC reintegration is essentially a reversal of V(D)J recombination and therefore the DNA damage sensing machinery would not be sensitised to a reintegration event (Arnal & Roth, 2007). Thus, the oncogenic mutations caused by ESC reintegration would not be removed by apoptosis, even though the disruption of an important gene might result in cell death. In addition, once a specific ESC reintegrates into the genome, it cannot be used for further reintegration or cut-and-run events any more. In contrast, as investigated in this thesis, a specific intact ESC can be continually used for the cut-and-run reaction until it is eventually lost, generating numerous DSBs that could be the substrates for chromosomal translocations via end donation, thereby posing a risk to genome stability, even potentially leading to cancers. Even though reintegration reactions have the potential to lead to considerably great damage upon insertion, the low probability of ESC reintegration into a protooncogene (only seven cryptic RSSs was mapped to protooncogenes (Merelli et al., 2010)) and the low frequency of this reaction relative to cut-and-run reaction could dramatically reduce the dangers of ESC reintegration. Together, it appears that the cut-and-run reaction is likely to be more frequent and dangerous than ESC reintegration.

6.6 Future plans

The experiments in this thesis investigated the mechanism of cut-and-run reaction *in vitro* and *in vivo*. In addition, preliminary investigations into the dangers of the cut-and-run reaction were performed. To further determine how dangerous the cut-and-run reaction is, several experiments should be carried out.

Firstly, during the analysis of the persistence of the ESC at different B cell developmental stages, it was found that, even in spleen IgG⁺ cells, the ESCs are present at reasonable levels. However, since this detection depends on the PCR-based assay with a small product across the RSS/RSS junction within the ESC, it is not known whether the ESC persists as extrachromosomal DNA. Indeed, if the ESC contains another RSS, then there is a possibility that this RSS is cleaved, resulting in reinsertion of the ESC into the genome, thereby decreasing the frequency of cut-and-run reaction. Therefore, to test if the ESC truly persists as an extrachromosomal element, genomic DNA should be isolated from spleen IgG⁺ cells and eccDNA should be enriched using CsCl-EB density gradient centrifugation. Following this, LMPCR should be performed to check the presence of specific circular ESC.

Secondly, during cell development, the lambda ESC gradually decreases whilst the kappa ESC dramatically increases. It was found that a specific lncRNA is present on all of the kappa ESCs, but not the lambda ESCs. To check if this lncRNA is important for maintenance of kappa ESCs, the expression of this lncRNA needs to be determined at different B cell developmental stages. Following this, the lncRNA should be knocked out using CRISPR-Cas9 in 103/BCL-2 cells. In these cells, the expression of RAG proteins can be controlled by temperature shift from 33°C to 39°C. If the lncRNA is essential for maintenance of the kappa ESCs and cell survival, after temperature shift, the cells in which the lncRNA is knocked out will be expected to die. In addition, since this lncRNA is not found in *IgH* locus, the presence of IgH ESCs needs to be determined at different B cell developmental stages. If the IgH ESC also decreases, it will further imply that this lncRNA is a reason for the persistence of kappa ESCs. Furthermore, the presence of replication origins on the ESCs needs to be determined. If these exist, it can explain how ESCs persist so readily.

Finally, the occurrence of the cut-and-run reaction relies on the co-existence of ESC and RAG proteins. It has been reported that the RAG proteins can be re-expressed for receptor editing (Gay et al., 1993; McGargill et al., 2000; Nemazee, 2006; Tiegs et al., 1993). However, when I checked the RAG

expression at different B cell developmental stages, high level RAG expression wasn't found in IgM⁺ B cells where receptor editing is thought to occur. To check if RAGs and the ESC co-exist during receptor editing, the correct sub-population of B cells where receptor editing takes place should be sorted, followed by checking the expression of RAG proteins and persistence of ESCs. Receptor editing was observed in IgM⁺ IgD⁻ immature B cells (Hertz & Nemazee, 1997; Nemazee, 2006; Nemazee, 2017), therefore, IgM⁺ IgD⁻ immature B cells should be purified by flow cytometry.

Cut-and-run is a dangerous reaction which seems to underpin genomic instability in *ETV6/RUNX1* positive ALL. A further implication of this work is that cut-and-run has potential roles of in other cancers. A number of leukaemia subtypes, such as hyperdiploid ALL (51-65 chromosomes) and *BCR-ABL* ALL, also occur in immature lymphoid cells (Boer & den Boer, 2017; Paulsson, 2016; Paulsson & Johansson, 2009) and could be triggered by cut-and-run. In addition, cut-and-run potentially occurs in immature B cells when RAGs upregulated for receptor editing. This could generate DSBs that form mutations that contribute to the development of immature and mature B cell cancers.

6.7 Conclusions

Cut-and-run is a newly uncovered aberrant V(D)J recombination reaction whereby an asymmetric RAG cleavage occurs in a synaptic complex containing an RSS and an ESC. This asymmetric cleavage is due to RAG simultaneously binding to both RSSs in the ESC, thereby blocking the cleavage site at RSS/RSS boundary both *in vitro* and *in vivo*. The ESC is able to cause cutting at cRSSs and generate numerous DSBs, suggesting that the cut-and-run reaction could provide substrates for chromosome translocation by generating orphan DNA breaks. This has significant implications for the aetiology of lymphoid cancers. In particular, the ESC is a very long-lived structure that is able to persist to the mature B cell stage, indicating that cut-and-run reaction has the potential to trigger DNA breaks associated with mature B cells cancers.

References

- Abarrategui, I. & Krangel, M.S. (2006). Regulation of T cell receptor-alpha gene recombination by transcription. *Nat Immunol* 7:1109-1115.
- Abarrategui, I. & Krangel, M.S. (2007). Noncoding transcription controls downstream promoters to regulate T-cell receptor alpha recombination. *EMBO J* 26: 4380-4390.
- Abraham, R.T. (2003). Checkpoint signaling: epigenetic events sound the DNA strand-breaks alarm to the ATM protein kinase. *Bioessays* 25(7):627-630.
- Agata, Y., Tamaki, N., Sakamoto, S., Ikawa, T., Masuda, K., Kawamoto, H. & Murre, C. (2007). Regulation of T cell receptor beta gene rearrangements and allelic exclusion by the helix-loop-helix protein, E47. *Immunity* 27(6):871-884.
- Agrawal, A., Eastman, Q.M. & Schatz, D.G. (1998). Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature* 394:744-751.
- Agrawal, A. & Schatz, D.G. (1997). RAG1 and RAG2 form a stable postcleavage synaptic complex with DNA containing signal ends in V(D)J recombination. *Cell* 89: 43-53.
- Aidinis, V., Bonaldi, T., Beltrame, M., Santagata, S., Bianchi, M.E. & Spanopoulou, E. (1999). The RAG1 homeodomain recruits HMG1 and HMG2 to facilitate recombination signal sequence binding and to enhance the intrinsic DNA-bending activity of RAG1-RAG2. *Mol Cell Biol* 19(10):6532-6542.
- Akamatsu, Y., Monroe, R., Dudley, D.D., Elkin, S.K., Gartner, F., Talukder, S.R., Takahama, Y., Alt, F.W., Bassing, C.H. & Oettinger, M.A. (2003). Deletion of the RAG2 C terminus leads to impaired lymphoid development in mice. *Proc Natl Acad Sci U S A* 100:1209-1214.
- Akamatsu, Y. & Oettinger, M.A. (1998). Distinct roles of RAG1 and RAG2 in binding the V(D)J recombination signal sequences. *Mol Cell Biol* 18:4670-4678.
- Akamastu, Y., Tsurushita, N., Nagawa, F., Matsuoka, M., Okazaki, K., Imai, M. & Sakano, H. (1994). Essential residues in V(D)J recombination signals. *J Immunol* 153(10):4520-4529.
- Alberts, B., Johnson, A., Lewis, J., Raff, M., Robert, K. & Walter, P. (2002). Molecular Biology of the Cell, Available at <http://www.ncbi.nlm.nih.gov/books/NBK21070/>.
- Almagro, J.C., Hernandez, I., Ramirez, M.C. & Vargas-Madrado, E. (1998). Structural differences between the repertoires of mouse and human germline genes and their evolutionary implications. *Immunogenetics* 47(5):355-363.
- Alt, F.W., Enea, V., Bothwell, A.L. & Baltimore, D. (1980). Activity of multiple light chain genes in murine myeloma cells producing a single, functional light

chain. *Cell* 21(1):1-12.

Alt, F.W., Rosenberg, N., Casanova, R.J., Tomas, E. & Baltimore, D. (1982). Immunoglobulin heavy-chain expression and class switching in a murine leukaemia cell line. *Nature* 296(5855):325-331.

Alt, F.W., Yancopoulos, G.D., Blackwell, T.K., Wood, C., Thomas, E., Boss, M., Coffman, R., Rosenberg, N., Tonegawa, S. & Baltimore, D. (1984). Ordered rearrangement of immunoglobulin heavy chain variable region segments. *EMBO J* 3:1209-1219.

Anderson, C.W. & Carter T.H. (1996). The DNA-activated protein kinase -- DNA-PK. *Curr Top Microbiol Immunol* 217:91-111.

Aparicio, T., Baer, R. & Gautier, J. (2014). DNA double-strand break repair pathway choice and cancer. *DNA Repair (Amst)* 19:169-175.

Aplan, P.D., Lombardi, D.P., Ginsberg, A.M., Cossman, J., Bertness, V.L. & Kirsch, I.R. (1990). Disruption of the human SCL locus by "illegitimate" V-(D)-J recombinase activity. *Science* 250(4986):1426-1429.

Arbuckle, J.L., Fauss, L.A., Simpson, R., Ptaszek, L.M. & Rodgers, K.K. (2001). Identification of two topologically independent domains in RAG1 and their role in macromolecular interactions relevant to V(D)J recombination. *J Biol Chem* 276: 37093-37101.

Arnal, S.M. & Roth, D.B. (2007). Excised V(D)J recombination byproducts threaten genomic integrity. *Trends Immunol* 28:289-292.

Arun, S.S., Breuer, W. & Hermanns, W. (1996). Immunohistochemical examination of light-chain expression (lambda/kappa ratio) in canine, feline, equine, bovine and porcine plasma cells. *Zentralbl Veterinarmed A* 43(9):573-576.

Ashworth, T.D., Pear, W.S., Chiang, M.Y., Blacklow, S.C., Mastio, J., Xu, L., Kelliher, P., Chan, S. & Aster, J.C. (2010). Deletion-based mechanisms of Notch1 activation in T-ALL: key roles for RAG recombinase and a conserved internal translational start site in Notch1. *Blood* 116(25):5455-5464.

Askary, A., Shimazaki, N., Bayat, N. & Lieber, M.R. (2014). Modeling of the RAG reaction mechanism. *Cell Rep* 7(2):307-315.

Bailey, S.N. & Rosenberg, N. (1997). Assessing the pathogenic potential of the V(D)J recombinase by interlocus immunoglobulin light-chain gene rearrangement. *Mol Cell Biol* 17(2):887-894.

Bakhshi, A., Wright, J.J., Graninger, W., Seto, M., Owens, J., Cossman, J., Jensen, J.P., Goldman, P. & Korsmeyer, S.J. (1987). Mechanism of the t(14;18) chromosomal translocation: structural analysis of both derivative 14 and 18 reciprocal partners. *Proc Natl Acad Sci U S A* 84:2396-2400.

Barlow, C., Hirotsune, S., Paylor, R., Liyanage, M., Eckhaus, M., Collins, F., Shiloh, Y., Crawley, J.N., Ried, T., Tagle, D. & Wynshwa-Boris, A. (1996). Atm-deficient mice: a paradigm of ataxia telangiectasia. *Cell* 86(1):157-171.

Barreto, V. & Cumano, A. (2000). Frequency and characterization of phenotypic Ig heavy chain allelically included IgM-expressing B cells in mice. *J Immunol* 164(2):893-899.

Bassing, C.H., Alt, F.W., Hughes, M.M., D'Auteuil, M., Wehrly, T.D., Woodman, B.B., Gartner, F., White, J.M., Dabidson, L. & Sleckman, B.P. (2000). Recombination signal sequences restrict chromosomal V(D)J recombination beyond the 12/23 rule. *Nature* 405:583-586.

Bassing, C.H., Swat, W. & Alt, F.W. (2002). The mechanism and regulation of chromosomal V(D)J recombination. *Cell* 109 Suppl:S45-55.

Baumann, M., Mamais, A., McBlane, F., Xiao, H. & Boyes, J. (2003). Regulation of V(D)J recombination by nucleosome positioning at recombination signal sequences. *EMBO J* 22: 5197-5207.

Berli, R.R., Dreier, B. & Barbas, C.F. 3rd (2000). Positive and negative regulation of endogenous genes by designed transcription factors. *Proc Natl Acad Sci U S A* 97(4):1495-1500.

Begum, N.A., Stanlie, A., Nakata, M., Akiyama, H. & Honjo, T. (2012). The histone chaperone Spt6 is required for activation-induced cytidine deaminase target determination through H3K4me3 regulation. *J Biol Chem* 287(39):32415-32429.

Beissert, S., Schwarz, A. & Schwarz, T. (2006). Regulatory T cells. *J Invest Dermatol* 126 (1):15-24.

Bellon, S.F., Rodgers, K.K., Schatz, D.G., Coleman, J.E. & Steitz, T.A. (1997). Crystal structure of the RAG1 dimerization domain reveals multiple zinc-binding motifs including a novel zinc binuclear cluster. *Nat Struct Biol* 4:586-591.

Belotserkovskaya, R., Oh, S., Bondarenko, V.A., Orphanides, G., Studitsky, V.M. & Reinberg, D. (2003). FACT facilitates transcription-dependent nucleosome alteration. *Science* 301:1090-1093.

Benedict, C.L., Glifillan, S., Thai, T.H. & Kearney, J.F. (2000). Terminal deoxynucleotidyl transferase and repertoire development. *Immunol Rev* 175:150-157.

Benson, M.D., Kluge-Beckerman, B., Zeldenrust, S.R., Siesky, A.M., Bodenmiller, D.M., Showalter, A.D. & Sloop, K.W. (2006). Targeted suppression of an amyloidogenic transthyretin with antisense oligonucleotides. *Muscle Nerve* 33:609-618.

Bergeron, S., Anderson, D.K. & Swanson, P.C. (2006). RAG and HMGB1 proteins: purification and biochemical analysis of recombination signal

complexes. *Methods Enzymol* 408:511-528.

Bertolino, E., Reddy, K., Medina, K.L., Parganas, E., Ihle, J. & Singh, H. (2005). Regulation of interleukin 7-dependent immunoglobulin heavy-chain variable gene rearrangements by transcription factor STAT5. *Nat Immunol* 6(8):836-843.

Beth, A.H. (2012). Rotational diffusion of membrane proteins: characterization of protein-protein interactions in membranes. *Biophys J* 103(6):1109.

Bevington, S. & Boyes, J. (2013). Transcription-coupled eviction of histones H2A/H2B governs V(D)J recombination. *EMBO J* 32(10):1381-1392.

Bianchi, M.E. & Manfredi, A.A. (2007). High-mobility group box 1 (HMGB1) protein at the crossroads between innate and adaptive immunity. *Immunol Rev* 220:35-46.

Birbrair, A. & Frenette, P.S. (2016). Niche heterogeneity in the bone marrow. *Ann N Y Acad Sci* 1370(1):82-96.

Bischerour, J., Lu, C., Roth, D.B. & Chalmers, R. (2009). Base flipping in V(D)J recombination: insights into the mechanism of hairpin formation, the 12/23 rule, and the coordination of double-strand breaks. *Mol Cell Biol* 29:5889-5899.

Blackwell, T.K., Moore, M.W., Yancopoulos, G.D., Suh, H., Lutzker, S., Selsing, E. & Alt, F.W. (1986). Recombination between immunoglobulin variable region gene segments is enhanced by transcription. *Nature* 324:585-589.

Blier, P.R., Griffith, A.J., Craft, J. & Hardin, J.A. (1993). Binding of Ku protein to DNA. Measurement of affinity for ends and demonstration of binding to nicks. *J Biol Chem* 268(10):7594-601.

Boehm, T., Baer, R., Lavenir, I., Forster, A., Waters, J.J., Nacheva, E. & Rabbitts, T.H. (1988). The mechanism of chromosomal translocation t(11;14) involving the T-cell receptor C delta locus on human chromosome 14q11 and a transcribed region of chromosome 11p15. *EMBO J* 7:385-394.

Boehm, T. & Rabbitts, T.H. (1989). The human T cell receptor genes are targets for chromosomal abnormalities in T cell tumors. *FASEB J* 3(12):2344-2359.

Boer, J.M. & den Boer, M.L. (2017). BCR-ABL1-like acute lymphoblastic leukaemia: From bench to bedside. *Eur J Cancer* 82:203-218.

Bories, J.C., Cayuela, J.M., Loiseau, P. & Sigaux, F. (1991). Expression of human recombination activating genes (RAG1 and RAG2) in neoplastic lymphoid cells: correlation with cell differentiation and antigen receptor expression. *Blood* 78: 2053-2061.

Boudinot, P., Rueff-Juy, D., Drapier, A.M., Cazenave, P.A. & Sanchez, P. (1995). Various V-J rearrangement efficiencies shape the mouse lambda B cell

repertoire. *Eur J Immunol* 25: 2499-2505.

Brandle, D., Muller, C., Rulicke, T., Hengartner, H. & Pircher, H. (1992). Engagement of the T- cell receptor during positive selection in the thymus down-regulates RAG-1 expression. *Proc Natl Acad Sci U S A* 89:9529-9533.

Brandt, V.L. & Roth, D.B. (2009). Recent insights into the formation of RAG-induced chromosomal translocations. *Adv Exp Med Biol* 650:32-45.

Bredemeyer, A.L., Sharma, G.G., Huang, C.Y., Helmink, B.A., Walker, L.M., Khor, K.C., Nuskey, B., Sullivan, K.E., Pandita, T.K., Bassing, C.H. & Sleckman, B.P. (2006). ATM stabilizes DNA double-strand-break complexes during V(D)J recombination. *Nature* 442:466-470.

Brodeur, P. H. & Riblet, R. (1984). The immunoglobulin heavy chain variable region (Igh-V) locus in the mouse. I. One hundred Igh-V genes comprise seven families of homologous genes. *Eur J Immunol* 14(10):922-930.

Brown, L., Cheng, J.T., Chen, Q., Siciliano, M.J., Crist, W. Buchanan, G. & Baer, R. (1990). Site-specific recombination of the tal-1 gene is a common occurrence in human T cell leukemia. *EMBO J* 9(10):3343-3351.

Brown, S.T., Miranda, G.A., Galic, Z., Hartman, I.Z., Lyon, C.J. & Aguilera, R.J. (1997). Regulation of the RAG-1 promoter by the NF-Y transcription factor. *J Immunol* 158(11):5071-5074.

Burma, S. & Chen, D.J. (2004). Role of DNA-PK in the cellular response to DNA double-strand breaks. *DNA Repair (Amst)* 3(8-9):909-918.

Callebaut, I. & Moron, J.P. (1998). The V(D)J recombination activating protein RAG2 consists of a six-bladed propeller and a PHD fingerlike domain, as revealed by sequence analysis. *Cell Mol Life Sci* 54(8): 880-891.

Calogero, S., Grassi, F., Aguzzi, A., Voigtlander, T., Ferrier, P., Ferrari, S. & Bianchi, M.E. (1999). The lack of chromosomal protein Hmg1 does not disrupt cell growth but causes lethal hypoglycaemia in newborn mice. *Nat Genet* 22:276-280.

Carmona, L.M., Fugmann, S.D. & Schatz, D.G. (2016). Collaboration of RAG2 with RAG1-like proteins during the evolution of V(D)J recombination. *Genes Dev.* 30(8): 909-917.

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B. & Wells, C. et al. (2005). The transcriptional landscape of the mammalian genome. *Science* 309(5740):1559-1563.

Cayuela, J.M., Gardie, B. & Sigaux, F. (1997). Disruption of the multiple tumor suppressor gene MTS1/p16(INK4a)/CDKN2 by illegitimate V(D)J recombinase activity in T-cell acute lymphoblastic leukemias. *Blood* 90(9):3720-3726.

Cereghini, S. & Yaniv, M. (1984). Assembly of transfected DNA into chromatin:

structural changes in the origin-promoter-enhancer region upon replication. *EMBO J* 3:1243-1253.

Chakraborty, T., Perlot, T., Subrahmanyam, R., Jani, A., Goff, P.H., Zhang, Y., Ivanova, I., Alt, F.W. & Sen, R. (2009). A 220-nucleotide deletion of the intronic enhancer reveals an epigenetic hierarchy in immunoglobulin heavy chain locus activation. *J Exp Med* 206(5):1019-1027.

Charles J. (2001). Immunobiology. 5th Edition. Garland Publishing.

Chatterji, M. Tsai, C.L. & Schatz, D.G. (2004). New concepts in the regulation of an ancient reaction: transposition by RAG1/RAG2. *Immunol Rev* 200:261-271.

Chaudhry, M.A. & Weinfeld, M. (1995). Induction of double-strand breaks by S1 nuclease, mung bean nuclease and nuclease P1 in DNA containing abasic sites and nicks. *Nucleic Acids Res* 23(9):3805-3809.

Chien, Y. H., Iwashima, M., Kaplan, K. B., Elliott, J. F. & Davis, M. M. (1987). A new T-cell receptor gene located within the alpha locus and expressed early in T-cell differentiation. *Nature* 327:677-82.

Choi, N.M., Loguercio, S., Verma-Gaur, J., Degner, S.C., Torkamani, A., Su, A.I., Oltz, E.M., Artyomov, M. & Feeney, A.J. (2013). Deep sequencing of the murine IgH repertoire reveals complex regulation of nonrandom V gene rearrangement frequencies. *J Immunol* 191(5):2393-2402.

Chowdhury, D. & Sen, R. (2001). Stepwise activation of the immunoglobulin mu heavy chain gene locus. *EMBO J* 20(22):6394-6403.

Chowdhury, D. & Sen, R. (2003). Transient IL-7/IL-7R signaling provides a mechanism for feedback inhibition of immunoglobulin heavy chain gene rearrangements. *Immunity* 18(2):229-241.

Chun, J.J., Schatz, D.G., Oettinger, M.A., Jaenisch, R. & Baltimore, D. (1991). The recombination activating gene-1 (RAG-1) transcript is present in the murine central nervous system. *Cell* 64(1):189-200.

Ciubotaru, M., Ptaszek, L.M., Baker, G.A., Baker, S.N., Bright, F.V. & Schatz, D.G. (2003). RAG1-DNA binding in V(D)J recombination. Specificity and DNA-induced conformational changes revealed by fluorescence and CD spectroscopy. *J Biol Chem* 278:5584-5596.

Ciubotaru, M., Trexler, A.J., Spiridon, L.N., Surleac, M.D., Rhoades, E., Petrescu, A.J. & Schatz, D.G. (2013). RAG and HMGB1 create a large bend in the 23RSS in the V(D)J recombination synaptic complexes. *Nucleic Acids Res* 41:2437-2454.

Cohn, M. & Langman, R.E. (1990). The protecton: the unit of humoral immunity selected by evolution. *Immunol Rev* 115:11-147.

Coleclough, C., Perry, R.P., Karjalainen, K. & Weigert, M. (1981). Aberrant rearrangements contribute significantly to the allelic exclusion of immunoglobulin gene expression. *Nature* 290 (5805):372-378.

Collazo-Garcia, N., Scherer, P. & Aplan, P.D. (1995). Cloning and characterization of a murine SIL gene. *Genomics* 30(3):506-513.

Cong, L., Zhou, R., Kuo, Y.C., Cunniff, M. & Zhang, F. (2012). Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. *Nat Commun* 3:968.

Corneo, B., Wendland, R.L., Deriano, L., Cui, X., Klein, I.A., Wong, S.Y., Arnal, S., Holub, A.J., Weller, G.R., Pancake, B.A. Shah, S., Brandt, V.L., Meek, K. & Roth, D.B. (2007). Rag mutations reveal robust alternative end joining. *Nature* 449:483-486.

Coussens, M.A., Wendland, R.L., Deriano, L., Lindsay, C.R., Arnal, S.M. & Roth, D.B. (2013). RAG2's acidic hinge restricts repair-pathway choice and promotes genomic stability. *Cell Rep* 4:870-878.

Cowell, L.G., Davila, M., Ramsden, D. & Kelsoe, G. (2004). Computational tools for understanding sequence variability in recombination signals. *Immunol Rev* 200:57-69.

Cuomo, C.A., Mundy, C.L. & Oettinger, M.A. (1996). DNA sequence and structure requirements for cleavage of V(D)J recombination signal sequences. *Mol Cell Biol* 16:5683-5690.

Cuomo, C.A. & Oettinger, M.A. (1994). Analysis of regions of RAG-2 important for V(D)J recombination. *Nucleic Acids Res* 22:1810-1814.

Curry, J.D., Geier, J.K. & Schlissel, M.S. (2005). Single-strand recombination signal sequence nicks in vivo: evidence for a capture model of synapsis. *Nat Immunol* 6: 1272-1279.

Curry, J.D., Schulz, D., Guidos, C.J., Danska, J.S., Nutter, L., Nussenzweig, A. & Schlissel, M.S. (2007). Chromosomal reinsertion of broken RSS ends during T cell development. *J Exp Med* 204:2293-2303.

Curry, J.D. & Schlissel, M.S. (2008). RAG2's non-core domain contributes to the ordered regulation of V(D)J recombination. *Nucleic Acids Res* 36:5750-5762.

Davies, D.R., Goryshin, I.Y., Reznikoff, W.S. & Rayment, I. (2000). Three-dimensional structure of the Tn5 synaptic complex transposition intermediate. *Science* 289:77-85.

Dayn, A. Samadashwily, G.M. & Mirkin, S.M. (1992). Intramolecular DNA triplexes: unusual sequence requirements and influence on DNA polymerization. *Proc Natl Acad Sci U S A* 89(23):11406-11410.

Deriano, L., Chaumeil, J., Coussens, M., Multani, A., Chou, Y., Alekseyenko, A.V., Chang, S., Skok, J.A. & Roth, D.B. (2011). The RAG2 C terminus suppresses genomic instability and lymphomagenesis. *Nature* 471:119-123.

Derrien, T., Guigo, R. & Johnson, R. (2012). The Long Non-Coding RNAs: A New (P)layer in the "Dark Matter". *Front Genet* 2:107.

Difilippantonio, M.J., McMahan, C.J., Eastman, Q.M., Spanopoulou, E. & Schatz, D.G. (1996). RAG1 mediates signal sequence recognition and recruitment of RAG2 in V(D)J recombination. *Cell* 87(2):253-262.

Difilippantonio, M.J., Petersen, S., Chen, H.T., Johnson, R., Jasin, M., Kanaar, R., Ried, T. & Nussenzweig, A. (2002). Evidence for replicative repair of DNA double-strand breaks leading to oncogenic translocation and gene amplification. *J Exp Med* 196(4):469-480.

Dildrop, R., Gause, A., Muller, W. & Rajewsky, K. (1987). A new V gene expressed in lambda-2 light chains of the mouse. *Eur J Immunol* 17(5):731-734.

Drejer-Teel, A.H., Fugmann, S.D. & Schatz, D.G. (2007). The beyond 12/23 restriction is imposed at the nicking and pairing steps of DNA cleavage during V(D)J recombination. *Mol Cell Biol* 27:6288-6299.

Duber, S., Engel, H., Rolink, A., Kretschmer, K. & Weiss, S. (2003). Germline transcripts of immunoglobulin light chain variable regions are structurally diverse and differentially expressed. *Mol Immunol* 40:509-516.

DuBrudge, R.B., Tang, P., Hsia, H.C., Leong, P.M., Miller, J.H. & Calos, M.P. (1987). Analysis of mutation in human cells by using an Epstein-Barr Virus shuttle system. *Mol Cell Biol* 7(1):379-387.

Dudley, D.D., Chaudhuri, J., Bassing, C.H. & Alt, F.W. (2005). Mechanism and control of V(D)J recombination versus class switch recombination: similarities and differences. *Adv Immunol* 86:43-112.

Dudley, D.D., Sekiguchi, J., Zhu, C., Sadofsky, M.J., Whitlow, S., DeVido, J., Monroe, R.J., Bassing, C.H. & Alt, F.W. (2003). Impaired V(D)J recombination and lymphocyte development in core Rag-1 expressing mice. *J Exp Med* 198: 143943450.

Durdik, J., Moore, M.W. & Selsing, E. (1984). Novel kappa light-chain gene rearrangements in mouse lambda light chain-producing B lymphocytes. *Nature* 307(5953):749-752.

Early, P., Huang, H., Davis, M., Calame, K. & Hood, L. (1980). An immunoglobulin heavy chain variable region gene is generated from three segments of DNA: VH, D and JH. *Cell* 19(4):981-992.

Eastman, Q.M., Leu, T.M. & Schatz, D.G. (1996). Initiation of V(D)J recombination in vitro obeying the 12/23 rule. *Nature* 380(6569):85-88.

Eastman, Q.M. & Schatz, D.G. (1997). Nicking is asynchronous and stimulated by synapsis in 12/23 rule-regulated V(D)J cleavage. *Nucleic Acids Res* 25:4370-4378.

Eisen, H.N. & Reilly, E.B. (1985). Lambda chains and genes in inbred mice. *Annu Rev Immunol* 3:337-365.

Elkin, S.K., Ivanov, D., Ewalt, M., Ferguson, C.G., Hyberts, S.G., Sun, Z.Y., Prestwich, G.D., Yuan, J., Wagner, G., Oettinger, M.A. & Gozani, O.P. (2005). A PHD finger motif in the C terminus of RAG2 modulates recombination activity. *J Biol Chem* 280: 28701-28710.

Elkin, S.K., Matthews, A.G. & Oettinger, M.A. (2003). The C-terminal portion of RAG2 protects against transposition in vitro. *EMBO J* 22:1931-1938.

Engel, H., Rolink, A. & Weiss, S. (1999). B cells are programmed to activate kappa and lambda for rearrangement at consecutive developmental stages. *Eur J Immunol* 29:2167-2176.

Espinoza, C.R. & Feeney, A.J. (2005). The extent of histone acetylation correlates with the differential rearrangement frequency of individual VH genes in pro-B cells. *J Immunol* 175:6668-6675.

Falzon, M., Fewell, J.W. & Kuff, E.L. (1993). EBP-80, a transcription factor closely resembling the human autoantigen Ku, recognizes single- to double-strand transitions in DNA. *J Biol Chem* 268(14):10546-10552.

Feeney, A.J., Goebel, P. & Esponzoza, C.R. (2004). Many levels of control of V gene rearrangement frequency. *Immunol Rev* 200:44-56.

Fisher, M.R., Rivera-Reyes, A., Bloch, N.B., Schatz, D.G. & Bassing, C.H. (2017). Immature lymphocytes inhibit Rag1 and Rag2 transcription and V(D)J recombination in response to DNA double-strand breaks. *J Immunol* 198(7):2943-2956.

Fitzsimmons, S.P., Bernstein, R.M., Max, E.E., Skok, J.A. & Shapiro, M.A. (2007). Dynamic changes in accessibility, nuclear positioning, recombination, and transcription at the Ig kappa locus. *J Immunol* 179:5264-5273.

Franchini, D.M., Benoukraf, T., Jaeger, S., Ferrier, P. & Payet-Bornet, D. (2009). Initiation of V(D)J recombination by Dbeta-associated recombination signal sequences: a critical control point in TCRbeta gene assembly. *PLoS One* 4:e4575.

Fugmann, S.D. (2010). The origins of the Rag genes--from transposition to V(D)J recombination. *Semin Immunol* 22(1):10-16.

Fugmann, S.D., Lee, A.I., Shockett, P.E., Villey, I.J. & Schatz, D.G. (2000a). The RAG proteins and V(D)J recombination: complexes, ends, and

transposition. *Annu Rev Immunol* 18:495-527.

Fugmann, S.D., Villey, I.J., Ptaszek, L.M. & Schatz, D.G. (2000b). Identification of two catalytic residues in RAG1 that define a single active site within the RAG1/RAG2 protein complex. *Mol cell* 5(1):97-107.

Fuller, K. & Storb, U. (1997). Identification and characterization of the murine Rag1 promoter. *Mol Immunol* 34(12-13):939-954.

Gauss, G. & Lieber, M. (1993). Unequal signal and coding joint formation in human V(D)J recombination. *Mol Cel Biol* 13(7):3900-3906.

Gauss, G. & Lieber, M. (1996). Mechanistic constraints on diversity in human V(D)J recombination. *Mol Cel Biol* 16(1):258-269.

Gay, D., Saunders, T., Camper, S. & Weigert, M. (1993). Receptor editing: an approach by autoreactive B cells to escape tolerance. *J Exp Med* 177:999-1008.

Geier, J.K. & Schlissel, M.S. (2006). Pre-BCR signals and the control of Ig gene rearrangements. *Semin Immunol* 18(1):31-39.

Gellert, M. (2002). V(D)J recombination: RAG proteins, repair factors, and regulation. *Annu Rev Biochem* 71:101-132.

Gerstein, R.M. & Lieber, M.R. (1993). Coding end sequence can markedly affect the initiation of V(D)J recombination. *Genes Dev* 7(7B):1459-1469.

Ghia, P., ten Boekel, E., Sanz, E., de la Hera, A., Rolink, A. & Melchers, F. (1996). Ordering of human bone marrow B lymphocyte precursors by single-cell polymerase chainreaction analyses of the rearrangement status of the immunoglobulin H and L chain gene loci. *J Exp Med* 184(6):2217-2229.

Gibson, D. (1974). Structural studies on normal horse immunoglobulin light chains. Detection of k-type N-terminal sequences. *Biochemistry* 13(13):2776-2785.

Gigi, V., Lewis, S., Shestova, O., Mijuskovic, M., Deriano, L., Meng, W., Luning Prak, E.T. & Roth, D.B. (2014). RAG2 mutants alter DSB repair pathway choice in vivo and illuminate the nature of 'alternative NHEJ'. *Nucleic Acids Res* 42: 6352-6364.

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., Lim, W.A., Weissman, J.S. & Qi, L.S. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* 154(2):442-451.

Giudicelli, V. Chaume, D. Jabado-Michaloud, J. & Lefranc, M.P. (2005). Immunogenetics sequence annotation: the strategy of IMGT based on IMGT-ONTOLOGY. *Stud Health Technol Inform* 116:3-8.

- Gladdy, R.A., Taylor, M.D., Williams, C.J., Grandal, I., Karaskova, J., Squire, J.A., Rutka, J.T., Guidos, C.J. & Danska, J.S. (2003). The RAG-1/2 endonuclease causes genomic instability and controls CNS complications of lymphoblastic leukemia in p53/Prkdc-deficient mice. *Cancer Cell* 3(1):37-50.
- Golding, A., Chandler, S., Ballestar, E., Wolffe, A.P. & Schliessel, M.S. (1999). Nucleosome structure completely inhibits in vitro cleavage by the V(D)J recombinase. *EMBO J* 18:3712-3723.
- Goldmit, M., Ji, Y., Skok, J., Roldan, E., Jung, S., Cedar, H., & Bergman, Y. (2005). Epigenetic ontogeny of the Igk locus during B cell development. *Nat Immunol* 6:198-203.
- Gorman, J.R. & Alt, F.W. (1998). Regulation of immunoglobulin light chain isotype expression. *Adv Immunol* 69:113-181.
- Gottlieb, T.M. & Jackson, S.P. (1993). The DNA-dependent protein kinase: requirement for DNA ends and association with Ku antigen. *Cell* 72(1):131-142.
- Gray, W.R., Dreyer, W.J & Hood, L. (1967). Mechanism of antibody synthesis: size differences between mouse kappa chains. *Science* 155(3761):465-467.
- Grawunder, U., Leu, T.M., Schatz, D.G., Werner, A., Rolink, A.G., Melchers, F. & Winkler, T.H. (1995). Down-regulation of RAG1 and RAG2 gene expression in pre B cells after functional immunoglobulin heavy chain rearrangement. *Immunity* 3:601-608.
- Grawunder, U., Wilm, M., Wu, X., Kulesza, P., Wilson, T.E., Mann, M. & Lieber, M.R. (1997). Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells. *Nature* 388 (6641):492-495.
- Grundy, G.J., Hesse, J.E. & Gellert, M. (2007). Requirements for DNA hairpin formation by RAG1/2. *Proc Natl Acad Sci U S A* 104:3078-3083.
- Grundy, G.J., Ramon-Maiques, S., Dimitriadis, E.K., Kotova, S., Biertumpfel, C., Heymann, J.B., Steven, A.C., Gellert, M. & Yang, W. (2009). Initial Stages of V(D)J Recombination: The Organization of RAG1/2 and RSS DNA in the Postcleavage Complex. *Mol Cell* 35:217-227.
- Grundy, G.J., Yang, W. & Gellert, M. (2010). Autoinhibition of DNA cleavage mediated by RAG1 and RAG2 is overcome by an epigenetic signal in V(D)J recombination. *Proc Natl Acad Sci U S A* 107:22487-22492.
- Gu, Y., Seidl, K.J., Rathbun, G.A., Zhu, C., Manis, J.P., van der Stoep, N., Davidson, L., Cheng, H.L., Sekiguchi, J.M., Frank, K. Stanhope-Baker, P., Schliessel, M.S. & Roth, D.B. (1997). Growth retardation and leaky SCID phenotype of Ku70-deficient mice. *Immunity* 7(5):653-665.
- Guidos, C.J., Williams, C.J., Grandal, I., Knowles, G., Huang, M.T. & Danska, J.S. (1996). V(D)J recombination activates a p53-dependent DNA damage checkpoint in scid lymphocyte precursors. *Genes Dev* 10(16):2038-2054.

Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P., Cabili, M.N., Jaenisch, R., Mikkelsen, T.S., Jacks, T., Hacohen, N., Bernstein, B.E., Kellis, M., Regev, A., Rinn, J.L. & Lander, E.S. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 458(7235):223-227.

Han, J.O., Steen, S.B. & Roth, D.B. (1999). Intermolecular V(D)J recombination is prohibited specifically at the joining step. *Mol Cell* 3:331-338.

Han, S. Dillon, S.R., Zheng, B., Shimoda, M., Schlissel, M.S. & Kelsoe, G. (1997). V(D)J recombinase activity in a subset of germinal center B lymphocytes. *Science* 278(5336):301-305.

Han, S. Zheng, B. Schatz, D.G., Spanopoulou, E. & Kelsoe, G. (1996). Neoteny in lymphocytes: Rag1 and Rag2 expression in germinal center B cells. *Science* 274(5295):2094-2097.

Hansen, J.D. & McBlane, J.F. (2000). Recombination-activating genes, transposition, and the lymphoid-specific combinatorial immune system: a common evolutionary connection. *Curr Top Microbiol Immunol* 248:111-135.

Hansen-Hagge, T.E., Yokota, S., Reuter, H.J., Schwarz, K. & Bartram, C.R. (1992). Human common acute lymphoblastic leukemia-derived cell lines are competent to recombine their T-cell receptor delta/alpha regions along a hierarchically ordered pathway. *Blood* 80(9):2353-2362.

Haluska, F.G., Finver, S., Tsujimoto, Y., Croce, C.M. (1986). The t(8;14) chromosomal translocation occurring in B-cell malignancies results from mistakes in V-D-J joining. *Nature* 324:1585461.

Harris, J.M., McIntosh, E.M. & Muscat, G.E. (1999). Structure/function analysis of a dUTPase: catalytic mechanism of a potential chemotherapeutic target. *J Mol Biol* 288(2):275-287.

Haydu, J.E., De Keersmaecker, K., Duff, M.K., Paietta, E., Racevskis, J., Wiernik, P.H., Rowe, J.M. & Ferrando, A. (2012). An activating intragenic deletion in NOTCH1 in human T-ALL. *Blood* 119:5211-5214.

Hazenbergh, M.D., Verschuren, M.C., Hamann, D., Miedema, F. & van Dongen, J.J. (2001). T cell receptor excision circles as markers for recent thymic emigrants: basic aspects, technical approach, and guidelines for interpretation. *J Mol Med (Berl)* 79: 631-640.

Helleday, T., Lo, J., van Gent, D.G. & Enelward, B.P. (2007). DNA double-strand break repair: from mechanistic understanding to cancer treatment. *DNA Repair (Amst)* 6(7):923-935.

Helmink, B. & Sleckman, B. (2012). The response to and repair of RAG-mediated DNA double-strand breaks. *Annu Rev Immunol* 30:175-202.

- Hertz, M. & Nemazee, D. (1997). BCR ligation induces receptor editing in IgM⁺IgD⁻ bone marrow B cells in vitro. *Immunity* 6:429-436.
- Hesse, J.E., Lieber, M.R., Gellert, M. & Mizuuchi, K. (1987). Extrachromosomal DNA substrates in pre-B cells undergo inversion or deletion at immunoglobulin V-(D)-J joining signals. *Cell* 49:775-783.
- Hesse, J.E., Lieber, M.R., Mizuuchi, K. & Gellert, M. (1989). V(D)J recombination: a functional definition of the joining signals. *Genes Dev* 3:1053-1061.
- Hesslein, D.G., Pflugh, D.L., Chowdhury, D., Bothwell, A.L., Sen, R. & Schatz, D.G. (2003). Pax5 is required for recombination of transcribed, acetylated, 5' IgH V gene segments. *Genes Dev* 17:37-42.
- Hieter, P. A., Korsmeyer, S. J., Waldmann, T. A. & Leder, P. (1981). Human immunoglobulin kappa light-chain genes are deleted or rearranged in lambda-producing B cells. *Nature* 290(5805):368-372.
- Hikida, M., Mori, M., Takai, T., Tomochika, K., Hamatani, K. & Ohmori, H. (1996). Reexpression of RAG-1 and RAG-2 genes in activated mature mouse B cells. *Science* 274(5295):2092-2094.
- Hiom, K. & Gellert, M. (1997). A stable RAG1-RAG2-DNA complex that is active in V(D)J cleavage. *Cell* 88:658-662.
- Hiom, K. & Gellert, M. (1998). Assembly of a 12/23 paired signal complex: a critical control point in V(D)J recombination. *Mol cell* 1(7):1011-1019.
- Hiom, K., Melek, M. & Gellert, M. (1998). DNA transposition by the RAG1 and RAG2 proteins: a possible source of oncogenic translocations. *Cell* 94:463-470.
- Hochtl, J. & Zachau, H.G. (1983). A novel type of aberrant recombination in immunoglobulin genes and its implications for V-J joining mechanism. *Nature* 302(5905):260-263.
- Holtmeier, W. & Kabelitz, D. (2005). Gammadelta T cells link innate and adaptive immune responses. *Chem Immunol Allergy* 86:151-183.
- Hozumi, N. & Tonegawa, S. (1976). Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. *Proc Natl Acad Sci U S A* 73:3628-3632.
- Hsieh, C.L. & Lieber, M.R. (1992). CpG methylated minichromosomes become inaccessible for V(D)J recombination after undergoing replication. *EMBO J* 11(1):315-325.
- Hsu, L.Y., Lauring, J., Liang, H.-E., Greenbaum, S., Cado, D., Zhuang, Y. & Schlissel, M.S. (2003). A conserved transcriptional enhancer regulates RAG gene expression in developing B cells. *Immunity* 19(1):105-117.

Huang, J. & Muegge, K. (2001). Control of chromatin accessibility for V(D)J recombination by interleukin-7. *J Leukoc Biol* 69:907-911.

Huarte, M., Guttman, M., Feldser, D., Garber, M., Koziol, M.J., Kenzelmann-Broz, D., Khalil, A.M., Zuk, O., Amit, I., Rabani, M., Attardi, L.D., Regev, A., Lander, E.S., Jacks, T. & Rinn, J.L. (2010). A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* 142(3):409-419.

Hughes, M.M., Yassai, M., Sedy, J.R., Wehrly, T.D., Huang, C.Y., Kanagawa, O., Gorski, J. & Sleckman, B.P. (2003). T cell receptor CDR3 loop length repertoire is determined primarily by features of the V(D)J recombination reaction. *Eur J Immunol* 33(6):1568-1575.

Hung, T., Wang, Y., Lin, M.F., Koegel, A.K., Kotake, Y., Grant, G.D., Horlings, H.M., Shah, N., Umbricht, C., Wang, P., Wang, Y., Kong, B., Langerod, A., Borresen-Dale, A.L., Kim, S.K., van de Vijver, M., Sukumar, S., Whitfield, M.L., Kellis, M., Xiong, Y., Wong, D.J. & Chang, H.Y. (2011). Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nat Genet* 43(7):621-619.

Hurwitz, R., Hozier, J., LeBien, T., Minowada, J., Gail-Peczalska, K., Kubonishi, I. & Kersey, J. (1979). Characterization of a leukemic cell line of the pre-B phenotype. *Int J Cancer* 23(2):174-180.

Izraeli, S., Colaizzo-Anas, T., Bertness, V.L., Mani, K., Aplan, P.D. & Kirsch, I.R. (1997). Expression of the SIL gene is correlated with growth induction and cellular proliferation. *Cell Growth Differ* 8(11):1171-1179.

Jackson, A.M. & Krangel, M.S. (2005). Allele-Specific Regulation of TCR Variable Gene Segment Chromatin Structure. *J Immunol* 175(8):5186-5191.

Jackson, A.M. & Krangel, M.S. (2006). Turning T-cell receptor beta recombination on and off: more questions than answers. *Immunol Rev* 209:129-141.

Jager, U., Bocskor, S., Le, T., Mitterbauer, G., Bolz, I., Chott, A., Kneba, M., Mannhalter, C. & Nadel, B. (2000). Follicular lymphomas' BCL-2/IgH junctions contain templated nucleotide insertions: novel insights into the mechanism of t(14;18) translocation. *Blood* 95:3520-3529.

Janeway, C.A., Travers, P., Walport, M. & Shlomchik, M.J. (2001). Immunobiology: The Immune System in Health and Disease. 5th edition. New York: Garland Science.

Jankovic, D., Gorello, P., Liu, T., Ehret, S., La Starza, R., Desjobert, C., Baty, F., Brutsche, M., Jayaraman, P.S., Santoro, A., Mecucci, C. & Schwaller, J. (2008). Leukemogenic mechanisms and targets of a NUP98/HHEX fusion in acute myeloid leukemia. *Blood* 111(12):5672-5682.

Janssen, E.M., Lemmens, E.E., Wolfe, T., Christen, U., von Herrath, M.G. &

- Schoenberger, S.P. (2003). CD4+ T cells are required for secondary expansion and memory in CD8+ T lymphocytes. *Nature* 421(6924):852-856.
- Ji, Y., Resch, W., Corbett, E., Yamane, A., Casellas, R. & Schatz, D.G. (2010). The in vivo pattern of binding of RAG1 and RAG2 to antigen receptor loci. *Cell* 141:419-431.
- Jiang, H., Ross, A.E. & Desiderio, S. (2004). Cell cycle-dependent accumulation in vivo of transposition-competent complexes between recombination signal ends and full-length RAG proteins. *J Biol Chem* 279:8478-8486.
- Jones, J.M. & Gellert, M. (2001). Intermediates in V(D)J recombination: a stable RAG1/2 complex sequesters cleaved RSS ends. *Proc Natl Acad Sci U S A* 98(23): 12926-12931.
- Jones, J.M. & Gellert, M. (2002). Ordered assembly of the V(D)J synaptic complex ensures accurate recombination. *EMBO J* 21:4162-4171.
- Jones, J.M. & Gellert, M. (2003). Autoubiquitylation of the V(D)J recombinase protein RAG1. *Proc Natl Acad Sci U S A* 100:15446-15451.
- Jones, J.M. & Simkus, C. (2009). The roles of the RAG1 and RAG2 "non-core" regions in V(D)J recombination and lymphocyte development. *Arch Immunol Ther Exp (Warsz)* 57(2):105-116.
- Jung, D., Bassing, C. H., Fugmann, S. D., Cheng, H. L., Schatz, D. G. & Alt, F. W. (2003). Extrachromosomal recombination substrates recapitulate beyond 12/23 restricted V(D)J recombination in nonlymphoid cells. *Immunity* 18:65-74.
- Jung D., Giallourakis, C., Mostoslavsky, R. & Alt, F.W. (2006). Mechanism and control of V(D)J recombination at the immunoglobulin heavy chain locus. *Annu Rev Immunol* 24:541-570.
- Kabotyanski, E.B., Zhu, C., Kallick, D.A. & Roth, D.B. (1995). Hairpin opening by single-strand-specific nucleases. *Nucleic Acids Res* 23(19):3872-3881.
- Kagan, J., Finan, J., Letofsky, J., Besa, E.C., Nowell, P.C. & Croce, C.M. (1987). Alpha-chain locus of the T-cell antigen receptor is involved in the t(10;14) chromosome translocation of T-cell acute lymphocytic leukemia. *Proc Natl Acad Sci U S A* 84(13):4543-4546.
- Kang, R., Chen, R., Zhang, Q., Hou, W., Wu, S., Cao, L., Huang, J., Yu, Y., Fan, X.G., Yan, Z., Sun, X., Wang, Q., Tsung, A., Billiar, T.R., Zeh, H.J. 3rd, Lotze, M.T. & Tang, D. (2014). HMGB2 in health and disease. *Mol Aspects Med* 40:1-116.
- Kanungo, J. (2017). Puromycin-resistant lentiviral control shRNA vector, pLKO.1 induces unexpected cellular differentiation of P19 embryonic stem cells. *Biochem Biophys Res Commun* 486(2):481-485.

Kapitonov, V.V. & Jurka, J. (2005). RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. *PLoS Biol* 3(6):e181.

Kassmeier, M.D., Mondal, K., Palmer, V.L., Raval, P., Kumar, S., Perry, G.A., Anderson, D.K., Ciborowski, P., Jackson, S., Xiong, Y. & Swanson, P.C. (2012). VprBP binds full-length RAG1 and is required for B-cell development and V(D)J recombination fidelity. *EMBO J* 31:945-958.

Kessler, S., Kim, K.J. & Scher, I. (1981). Surface membrane kappa and lambda light chain expression on spleen cells of neonatal and maturing normal and immune-defective CBA/NB mice: the kappa:lambda ratio is constant. *J Immunol* 127(4):1674-1678.

Kim, D.R., Dai, Y., Mundy, C.L., Yang, W. & Oettinger, M.A. (1999). Mutations of acidic residues in RAG1 define the active site of the V(D)J recombinase. *Genes Dev* 13(23):3070-3080.

Kim, M.S., Lapkouski, M., Yang, W. & Gellert, M. (2015). Crystal structure of the V(D)J recombinase RAG1-RAG2. *Nature* 518(7540):507-511.

Kim, D.R. & Oettinger, M.A. (1998). Functional analysis of coordinated cleavage in V(D)J recombination. *Mol Cell Biol* 18(8):4679-4688.

Kirch, S.A., Sudarsanam, P. & Oettinger, M.A. (1996). Regions of RAG1 protein critical for V(D)J recombination. *Eur J Immunol* 26:886-891.

Kirkham, C.M. (2014). Cut-and-Run: A New Mechanism by which V(D)J Recombination Could Trigger Leukaemia and Lymphoma. University of Leeds.

Kitagawa, Y., Inoue, K., Sasaki, S., Hayashi, Y., Mastuo, Y., Lieber, M.R., Mizoguchi, H., Yokota, J. & Kohno, T. (2002). Prevalent involvement of illegitimate V(D)J recombination in chromosome 9p21 deletions in lymphoid leukemia. *J Biol Chem* 277(48):46289-46297.

Klobeck, H.G. & Zachau, H.G. (1986). The human CK gene segment and the kappa deleting element are closely linked. *Nucleic Acids Res* 14(11):4591-4603.

Kong, F. & Chen-lo, H. (1999). T cell receptor gene deletion circles identify recent thymic emigrants in the peripheral T cell pool. *Proc Natl Acad Sci U S A* 96(4):1536-1540.

Korsmeyer, S.J., Arnold, A., Bakhshi, A., Ravetch, J.V., Siebenlist, U., Hieter, P.A., Sharrow, S.O., LeBien, T.W., Kersey, J.H., Poplack, D.G., Leder, P. & Waldmann, T.A. (1983). Immunoglobulin gene rearrangement and cell surface antigen expression in acute lymphocytic leukemias of T cell and B cell precursor origins. *J Clin Invest* 71(2):301-313.

Kosak, S.T., Skok, J.A., Medina, K.L., Riblet, R., Le Beau, M.M., Fisher, A.G. & Singh, H. (2002). Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. *Science* 296(5565):158-162.

- Krangel, M.S. (2009). Mechanics of T cell receptor gene rearrangement. *Curr Opin Immunol* 21(2):133-139.
- Kredel, S., Oswald, F., Nienhaus, K., Deuschle, K., Rocker, C. Wolff, M., Heilker, R. & Nienhaus, G.U. (2009). mRuby, a bright monomeric red fluorescent protein for labeling of subcellular structures. *PLoS One* 4(2):e4391.
- Kretz, M., Siprashvili, Z., Chu, C., Webster, D.E., Zehnder, A., Qu, K., Lee, C.S., Flockhart, R.J., Groff, A.F., Chow, J., Johnston, D., Kim, G.E., Spitale, R.C., Flynn, R.A., Zheng, G.X., Aiyer, S., Raj, A., Rinn, J.L., Chang, H.Y. & Khavari, P.A. (2013). Control of somatic tissue differentiation by the long non-coding RNA TINCR. *Nature* 493(7431):231-235.
- Kriatchko, A.N., Anderson, D.K. & Swanson, P.C. (2006). Identification and characterization of a gain-of-function RAG-1 mutant. *Mol Cell Biol* 26:4712-4728.
- Kridel, R., Sehn, L.H. & Gascoyne, R.D. (2012). Pathogenesis of follicular lymphoma. *J Clin Invest* 122:3424-3431.
- Kuo, L.J. & Yang, L.X. (2008). Gamma-H2AX - a novel biomarker for DNA double-strand breaks. *In Vivo* 22(3):305-309.
- Kuppers, R. (2005). Mechanisms of B-cell lymphoma pathogenesis. *Nat Rev Cancer* 5(4):251-262.
- Kuppers, R. & Dalla-Favera, R. (2001). Mechanisms of chromosomal translocations in B cell lymphomas. *Oncogene* 20:5580-5594.
- Kurioka, H., Kishi, H., Isshiki, H., Tagoh, H., Mori, K., Kitagawa, T., Nagata, T., Dohi, K. & Muraguchi, A. (1996). Isolation and characterization of a TATA-less promoter for the human RAG-1 gene. *Mol Immunol* 33(13):1059-1066.
- Kurosaki, T., Kometano, K. & Ise, W. (2015). Memory B cells. *Nat Rev Immunol* 15(3):149.
- Kurts, C. (2008). Th17 cells: a third subset of CD4+ T effector cells involved in organ-specific autoimmunity. *Nephrol Dial Transplant* 23(3):816-819.
- Kwon, J., Imbalzano, A.N., Matthews, A. & Oettinger, M.A. (1998). Accessibility of nucleosomal DNA to V(D)J cleavage is modulated by RSS positioning and HMG1. *Mol Cell* 2:829-839.
- Lam, A.J., St-Pierre, F., Gong, Y., Marshall, J.D., Cranfill, P.J., Barid, M.A., McKeown, M.R., Wiedenmann, J., Davidson, M.W., Schnitzer, M.J., Tsien, R.Y. & Li, M.Z. (2012). Improving FRET dynamic range with bright green and red fluorescent proteins. *Nat Methods* 9(10):1005-1012.
- Landree, M.A., Wibbenmeyer, J.A. & Roth, D.B. (1999). Mutational analysis of RAG1 and RAG2 identifies three catalytic amino acids in RAG1 critical for both

cleavage steps of V(D)J recombination. *Genes Dev* 13(23):3059-3069.

Lauring, J. & Schlissel, M.S. (1999). Distinct factors regulate the murine RAG-2 promoter in B- and T-cell lines. *Mol Cell Biol* 19(4):2601-2612.

LeBien, T.W. & Tedder, T.F. (2008). B lymphocytes: how they develop and function. *Blood* 112(5):1570-1580.

Lee, G.S., Neiditch, M.B., Salus, S.S. & Roth, D.B. (2004). RAG proteins shepherd double-strand breaks to a specific pathway, suppressing error-prone repair, but RAG nicking initiates homologous recombination. *Cell* 117:171-184.

Lee, J. & Desiderio, S. (1999). Cyclin A/CDK2 regulates V(D)J recombination by coordinating RAG-2 accumulation and DNA repair. *Immunity* 11(6):771-781.

Lennon, G.G. & Perry, R.P. (1985). C mu-containing transcripts initiate heterogeneously within the IgH enhancer region and contain a novel 5'-nontranslatable exon. *Nature* 318:475-478.

Lescale, C. & Deriano, L. (2017). The RAG recombinase: Beyond breaking. *Mech Ageing Dev* 165(Pt A):3-9.

Lewis, S.M. (1994). The mechanism of V(D)J Joining: lessons from molecular, immunological, and comparative analyses. *Adv Immunol* 56:27-150.

Lewis, S.M., Agard, E., Suh, S. & Czyzyk, L. (1997). Cryptic signals and the fidelity of V(D)J joining. *Mol Cell Biol* 17:3125-3136.

Lewis, S.M. & Hesse, J.E. (1991). Cutting and closing without recombination in V(D)J joining. *EMBO J* 10(12):3631-3639.

Li, A., Rue, M., Zhou, J., Wang, H., Goldwasser, M.A., Neuberg, D., Dalton, V., Zuckerman, D., Lyons, C., Silverman, L.B., Sallan, S.E. & Gribben, J.G. (2004). Utilization of Ig heavy chain variable, diversity, and joining gene segments in children with B-lineage acute lymphoblastic leukemia: implications for the mechanisms of VDJ recombination and for pathogenesis. *Blood* 103(12):4602-4609.

Li, Z. Dordai, D.I., Lee, J. & Desiderio, S. (1996). A conserved degradation signal regulates RAG-2 accumulation during cell division and links V(D)J recombination to the cell cycle. *Immunity* 5(6):575-589.

Liang, H.E., Hsu, L.Y., Cado, D. & Schlissel, M.S. (2004). Variegated transcriptional activation of the immunoglobulin kappa locus in pre-b cells contributes to the allelic exclusion of light-chain expression. *Cell* 118(1):19-29.

Liang, H.E., Hsu, L.Y., Cado, D., Cowell, L.G., Kelsoe, G. & Schlissel, M.S. (2002). The "dispensable" portion of RAG2 is necessary for efficient V-to-DJ rearrangement during B and T cell development. *Immunity* 17(5):639-651.

Lieber, M.R. (1998). Warner-Lambert/Parke-Davis Award Lecture. Pathological

and physiological double-strand breaks: roles in cancer, aging, and the immune system. *Am J Pathol* 153(5):1323-1332.

Lieber, M.R. (2010). The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu Rev Biochem* 79:181-211.

Lieber, M.R., Hesse, J.E., Mizuuchi, K. & Gellert, M. (1988). Studies of V(D)J recombination with extrachromosomal substrates. *Curr Top Microbiol Immunol* 137: 94-99.

Lin, W.C. & Desiderio, S. (1994). Cell cycle regulation of V(D)J recombination-activating protein RAG-2. *Proc Natl Acad Sci U S A* 91(7):2733-2737.

Litman, G.W., Rast, J.P., Shablott, M.J., Haire, R.N., Hulst, M., Roess, W., Litman, R.T., Hinds-Frey, K.R., Zilch, A. & Amemiya, C.T. (1993). Phylogenetic diversification of immunoglobulin genes and the antibody repertoire. *Mol Biol Evol* 10(1):60-72.

Litman, G.W. Rast, J.P. & Fugmann, S.D. (2010). The origins of vertebrate adaptive immunity. *Nat Rev Immunol* 10(8):543-553.

Little, A.J., Corbett, E., Ortega, F. & Schatz, D.G. (2013). Cooperative recruitment of HMGB1 during V(D)J recombination through interactions with RAG1 and DNA. *Nucleic Acids Res* 41:3289-3301.

Liu, Y., Subrahmanyam, R., Chakraborty, T., Sen, R. & Desiderio, S. (2007). A plant homeodomain in RAG-2 that binds Hypermethylated lysine 4 of histone H3 is necessary for efficient antigen-receptor-gene rearrangement. *Immunity* 27:561-571.

Livak, F. & Schatz, D.G. (1996). T-cell receptor alpha locus V(D)J recombination by-products are abundant in thymocytes and mature T cells. *Mol Cell Biol* 16(2):609-618.

Lovely, G.A., Brewster, R.C., Schatz, D.G., Baltimore, D. & Phillips, R. (2015). Single-molecule analysis of RAG-mediated V(D)J DNA cleavage. *Proc Natl Acad Sci U S A* 112(14):E1715-1723.

Lu, H., Schwarz, K. & Lieber, M. (2007). Extent to Which Hairpin Opening by the Artemis:DNA-PKcs complex can contribute to junctional diversity in V(D)J recombination. *Nucleic Acids Res* 35(20):6917-6923.

Luo, Z., Ronai, D. & Scharff, M.D. (2004). Structure/function analysis of a dUTPase: catalytic mechanism of a potential chemotherapeutic target. *J Allergy Clin Immunol* 114(4):726-735.

Ma Y., Schwarz, K. & Lieber, M. (2005). The Artemis:DNA-PKcs endonuclease cleaves DNA loops, flaps, and gaps. *DNA Repair (Amst)* 4(7):845-851.

Ma Y., Lu, H., Schwarz, K. & Lieber, M.R. (2005). Repair of double-strand DNA breaks by the human nonhomologous DNA end joining pathway: the iterative

processing model. *Cell Cycle* 4(9):1193-1200.

Maes, J., Caspi, Y., Rougeon, F., Haimovich, J. & Goodhardt, M. (2000). Secondary V(D)J rearrangements and B cell receptor-mediated down-regulation of recombination activating gene-2 expression in a murine B cell line. *J Immunol* 165 (2):703-709.

Maes, J., O'Neill, L.P., Cavellier, P., Turner, B.M., Rougeon, F. & Goodhardt, M. (2001). Chromatin remodeling at the Ig loci prior to V(D)J recombination. *J Immunol* 167:866-874.

Malarkey, C.S. & Churchill, M.E.A. (2012). The high mobility group box: the ultimate utility player of a cell. *Trends Biochem Sci* 37(12):553-562.

Malu, S., Malshetty, V., Francis, D. & Cortes, P. (2012). Role of non-homologous end joining in V(D)J recombination. *Immunol Res* 54(1-3):233-246.

Manolov, G. & Manolova, Y. (1972). Marker band in one chromosome 14 from Burkitt lymphomas. *Nature* 237(5349):33-34.

Marculescu, R., Le, T., Simon, P., Jaeger, U. & Nadel, B. (2002). V(D)J-mediated translocations in lymphoid neoplasms: a functional assessment of genomic instability by cryptic sites. *J Exp Med* 195:85-98.

Marculescu, R., Vanura, K., Montpellier, B., Roulland, S., Le, T., Navarro, J.M., Jager, U., McBlane, F. & Nadel, B. (2006). Recombinase, chromosomal translocations and lymphoid neoplasia: targeting mistakes and repair failures. *DNA Repair (Amst)* 5(9-10):1246-1258.

Marianes, A.E. & Zimmerman, A.M. (2011). Targets of somatic hypermutation within immunoglobulin light chain genes in zebrafish. *Immunology* 132(2):240-255.

Martens, J.A., Laprade, L. & Winston, F. (2004). Intergenic transcription is required to repress the *Saccharomyces cerevisiae* SER3 gene. *Nature* 429(6991):571-574.

Martinez-Jean, C., Folch, G. & Lefranc, M.P. (2001). Nomenclature and overview of the mouse (*Mus musculus* and *Mus sp.*) immunoglobulin kappa (IGK) genes. *Exp Clin Immunogenet* 18(4):255-279.

Mathieu, N., Hempel, W.M., Spicuglia, S., Verthuy, C. & Ferrier, P. (2000). Chromatin remodeling by the T cell receptor (TCR)-beta gene enhancer during early T cell development: Implications for the control of TCR-beta locus recombination. *J Exp Med* 192:625-636.

Matsuda, F., Ishii, K., Bourvagnet, P., Kuma, K., Hayashida, H., Miyata, T. & Honjo, V. (1998). The complete nucleotide sequence of the human immunoglobulin heavy chain variable region locus. *J Exp Med* 188(11):2151-62.

- Matthews, A.G., Kuo, A.J., Ramon-Maiques, S., Han, S., Champagne, K.S., Ivanov, D., Gallardo, M., Carney, D., Cheung, P., Ciccone, D.N. Walter, K.L., Utz, P.J., Shi, Y., Kutateladze, T.G., Yang, W., Gozani, O. & Oettinger, M.A. (2007). RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination. *Nature* 450:1106-1110.
- McBlane, F. & Boyes, J. (2000). Stimulation of V(D)J recombination by histone acetylation. *Curr Biol* 10:483-486.
- McBlane, J.F., van Gent, D.C., Ramsden, D.A., Romeo, C., Cuomo, C.A., Gellert, M. & Oettinger, M.A. (1995). Cleavage at a V(D)J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps. *Cell* 83 (3)387-395.
- McGargill, M.A., Derbinski, J.M. & Hogquist, K.A. (2000). Receptor editing in developing T cells. *Nat Immunol* 1(4):336-341.
- McMurry, M.T. & Krangel, M.S. (2000). A role for histone acetylation in the developmental regulation of VDJ recombination [see comments]. *Science* 287:495-498.
- Melchers, F., Karasuyama, H., Haasner, D., Bauer, S., Kudo, A., Sakaguchi, N., Jameson, B. & Rolink, A. (1993). The surrogate light chain in B-cell development. *Immunol Today* 14(2):60-68.
- Melek, M. & Gellert, M. (2000). RAG1/2-mediated resolution of transposition intermediates: two pathways and possible consequences. *Cell* 101(6):625-633.
- Melek, M., Gellert, M. & van Gent, D.C. (1998). Rejoining of DNA by the RAG1 and RAG2 proteins. *Science* 280(5361):301-303.
- Melek, M., Jones, J.M., O'Dea, M.H., Pais, G., Burke, T.R., Jr., Pommier, Y., Neamati, N. & Gellert, M. (2002). Effect of HIV integrase inhibitors on the RAG1/2 recombinase. *Proc Natl Acad Sci U S A* 99:134-137.
- Mendes, R.D., Sarmiento, L.M., Cante-Barrett, K., Zuurbier, L., Buijs-Gladdines, J.G., Povoas, V., Smits, W.K., Abecasis, M., Yunes, J.A., Sonneveld, E., Horstmann, M.A., Pieters, R., Barata, J.T. & Meijerink, J.P. (2014). PTEN micro-deletions in T-cell acute lymphoblastic leukemia are caused by illegitimate RAG-mediated recombination events. *Blood* 124:567.
- Mercer, T.R., Dinger, M.E., Mariani, J., Kosik, K.S., Mehler, M.F. & Mattick, J.S. (2008). Noncoding RNAs in long-term memory formation. *Neuroscientist* 14(5):434-445.
- Merelli, I., Guffanti, A., Fabbri, M., Cocito, A, Furia, L., Grazini, U., Bonnal, R.J., Milanesi, L. & McBlane, F. (2010). RSSsite: a reference database and prediction tool for the identification of cryptic Recombination Signal Sequences in human and murine genomes. *Nucleic Acids Res* 38:W262-267.
- Messier, T.L., O'Neill, J.P., Hou, S.M., Nicklas, J.A. & Finette, B.A. (2003). In

vivo transposition mediated by V(D)J recombinase in human T lymphocytes. *EMBO J* 22:1381-1388.

Messier, T.L., O'Neill, J.P. & Finette, B.A. (2006). V(D)J recombinase mediated inter-chromosomal HPRT alterations at cryptic recombination signal sequences in peripheral human t cells. *Hum mutat* 27:829.

Mijuskovic, M., Brown, S.M., Tang, Z., Lindsay, C.R., Efstathiadis, E., Deriano, L. & Roth, D.B. (2012). A streamlined method for detecting structural variants in cancer genomes by short read paired-end sequencing. *PLoS One* 7:e48314.

Mijuskovic, M., Chou, Y.F., Gigi, V., Lindsay, C.R., Shestova, O. Lewis, S.H. & Roth, D.B. (2015). Off-Target V(D)J recombination drives lymphomagenesis and is escalated by loss of the Rag2 C terminus. *Cell Rep* 12(11):1842-1852.

Mimori, T. & Hardin, J.A. (1986). Mechanism of interaction between Ku protein and DNA. *J Biol Chem* 261(22):10375-10379.

Minowada, J., Janossy, G., Greaves, M.F., Tsubota, T., Srivastava, B.I. Morikawa, S. & Tatsumi, E. (1978). Expression of an antigen associated with acute lymphoblastic leukemia in human leukemia-lymphoma cell lines. *J Natl Cancer Inst* 60(6):1269-1277.

Mizushima, S. & Nagata, S. (1990). pEF-BOS, a powerful mammalian expression vector. *Nucleic Acids Res* 18:5322.

Mo, X., Bailin, T. & Sadofsky, M.J. (1999). RAG1 and RAG2 cooperate in specific binding to the recombination signal sequence in vitro. *J Biol Chem* 274: 7025-7031.

Mombaerts, P., Iacomini, J., Johnson, R.S., Herrup, K., Tonegawa, S. & Papaioannou, V.E. (1992). RAG-1-deficient mice have no mature B and T lymphocytes. *Cell* 68(5):869-877.

Moore, M.W., Durdik, J., Persiani, D.M. & Selsing, E. (1985). Deletions of kappa chain constant region genes in mouse lambda chain-producing B cells involve intrachromosomal DNA recombinations similar to V-J joining. *Proc Natl Acad Sci U S A* 82(18):6211-6215.

Morshead, K.B., Ciccone, D.N., Taverna, S.D., Allis, C.D. & Oettinger, M.A. (2003). Antigen receptor loci poised for V(D)J rearrangement are broadly associated with BRG1 and flanked by peaks of histone H3 dimethylated at lysine 4. *Proc Natl Acad Sci U S A* 100:11577-11582.

Mosmann T, R. (1992). T lymphocyte subsets, cytokines, and effector functions. *Ann N Y Acad Sci* 664:89-92.

Mostoslavsky, R., Singh, N., Tenzen, T., Goldmit, M., Gabay, C., Elizur, S., Qi, P., Reubinoff, B.E., Chess, A., Cedar, H. & Bergman, Y. (2001). Asynchronous replication and allelic exclusion in the immune system. *Nature* 414(6860):221-225.

Mullighan, C.G., Miller, C.B., Radtke, I., Phillips, L.A., Dalton, J., Ma, J., White, D., Hughes, T.P., Le Beau, M.M., Pui, C.H., Relling, M.V., Shurtleff, S.A. & Downing, J.R. (2008). BCR-ABL1 lymphoblastic leukaemia is characterized by the deletion of Ikaros. *Nature* 453:1101314.

Mundy, C.L., Patenge, N., Matthews, A.G. & Oettinger, M.A. (2002). Assembly of the RAG1/RAG2 synaptic complex. *Mol Cell Biol* 22:69-77.

Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y. & Honjo, T. (2000). Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* 102(5):553-563.

Murphy, K.M. (2012). Janeway's Immunobiology 8th Edition. New York, NY: Garland Science.

Murphy, K.M., Travers, P. & Walport, M. (2007). *Janeway's Immunobiology*. 7th Ed. Garland Science.

Nadel, B., Cazenave, P.A. & Sanchez, P. (1990). Murine lambda gene rearrangements: the stochastic model prevails over the ordered model. *EMBO J* 9(2):435-440.

Nagawa, F., Hirose, S., Nishizumi, H., Nishihara, T. & Sakano, H. (2004). Joining mutants of RAG1 and RAG2 that demonstrate impaired interactions with the coding-end DNA. *J Biol Chem* 279:38360-38368.

Nagawa, F., Ishiguro, K., Tsuboi, A., Yoshida, T., Ishikawa, A., Takemori, T., Otsuka, A.J. & Sakano, H. (1998). Footprint analysis of the RAG protein recombination signal sequence complex for V(D)J type recombination. *Mol Cell Biol* 18:655-663.

Nagawa, F., Kodama, M., Nishihara, T., Ishiguro, K. & Sakano, H. (2002). Footprint analysis of recombination signal sequences in the 12/23 synaptic complex of V(D)J recombination. *Mol Cell Biol* 22:7217-7225.

Nam, C.H. & Rabbits, T.H. (2006). The role of LMO2 in development and in T cell leukemia after chromosomal translocation or retroviral insertion. *Mol Ther* 13(1):15-25.

Nambiar, M. & Raghavan, S.C. (2011). How does DNA break during chromosomal translocations? *Nucleic Acids Res* 39(14):5813-5825.

Neiditch, M.B., Lee, G.S., Huye, L.E., Brandt, V.L. & Roth, D.B. (2002). The V(D)J recombinase efficiently cleaves and transposes signal joints. *Mol Cell* 9:871-878.

Nemazee, D. (2006). Receptor editing in lymphocyte development and central tolerance. *Nat Rev Immunol* 6(10):728-740.

- Nemazee, D. (2017). Mechanisms of central tolerance for B cells. *Nat Rev Immunol* 17:281-294.
- Neuberger, M.S. (2008). Antibody diversification by somatic mutation: from Burnet onwards. *Immunol Cell Biol* 86(2):124-132.
- Niedziela-Majka, A., Rymarczyk, G., Kochman, M. & Ozyhar, A. (1998). GST-Induced dimerization of DNA-binding domains alters characteristics of their interaction with DNA. *Protein Expr Purif* 14:208-220.
- Nightingale, K.P., Baumann, M., Eberharder, A., Mamais, A., Becker, P.B. & Boyes, J. (2007). Acetylation increases access of remodelling complexes to their nucleosome targets to enhance initiation of V(D)J recombination. *Nucleic Acids Res* 35:6311-6321.
- Nishihara, T., Nagawa, F., Imai, T. & Sakano, H. (2008). RAG-heptamer interaction in the synaptic complex is a crucial biochemical checkpoint for the 12/23 recombination rule. *J Biol Chem* 283:4877-4885.
- Noack, M. & Miossec, P. (2014). Th17 and regulatory T cell balance in autoimmune and inflammatory diseases. *Autoimmun Rev* 13(6):668-677.
- Notarangelo, L.D., Villa, A. & Schwarz, K. (1999). RAG and RAG defects. *Curr Opin Immunol* 11(4):435-442.
- Nutt, S.L., Heavey, B., Rolink, A.G. & Busslinger, M. (1999). Commitment to the B-lymphoid lineage depends on the transcription factor Pax5. *Nature* 401:556-562.
- Nyvold, C.G., Bendix, K., Brandsborg, M., Pulczynski, S., Silkjaer, T. & Hokland, P. (2007). Multiplex PCR for the detection of BCL-1/IGH and BCL-2/IGH gene rearrangements--clinical validation in a prospective study of blood and bone marrow in 258 patients with or suspected of non-Hodgkin's lymphoma. *Acta Oncol* 46(1):21-30.
- Oettinger, M.A. (1990). RAG-1 and RAG-2, adjacent genes that synergistically active V(D)J recombination. *Science* 248:1517-1523.
- Oettinger, M.A. (1999). V(D)J recombination: on the cutting edge. *Curr Opin Cell Biol* 11(3):325-329.
- Ohnishi, T., Mori, E. & Takahashi, A. (2009). DNA double-strand breaks: their production, recognition, and repair in eukaryotes. *Mutat Res* 669(1-2):8-12.
- Olaru, A., Patterson, D. N., Villey, I. & Livak, F. (2003). DNA-Rag protein interactions in the control of selective D gene utilization in the TCR beta locus. *J Immunol* 171:3605-3611.
- Onozawa, M. & Aplan, P.D. (2012). Illegitimate V(D)J recombination involving nonantigen receptor loci in lymphoid malignancy. *Genes Chromosomes Cancer* 51(6):525-535.

- Orphanides, G., Wu, W.H., Lane, W.S., Hampsey, M. & Reinberg, D. (1999). The chromatin- specific transcription elongation factor FACT comprises human SPT16 and SSRP1 proteins. *Nature* 400:284-288.
- Papaemmanuil, E., Rapado, I., Li, Y., Potter, N.E., Wedge, D.C., Tubio, J., Alexandrov, L.B., Van Loo, P., Cooke, S.L. & Marshall, J. et al. (2014). RAG-mediated recombination is the predominant driver of oncogenic rearrangement in ETV6-RUNX1 acute lymphoblastic leukemia. *Nat Genet* 46:116-125.
- Parker, D. (1993). T cell-dependent B cell activation. *Annu Rev Immunol* 11:331-360.
- Parkinson, N.J., Roddis, M., Ferneyhough, B., Zhang, G., Marsden, A.J., Maslau, S., Sanchez-Pearson, Y., Barthlott, T., Humphreys, I.R. Ladell, K., Price, D.A., Ponting, C.P., Hollander, G. & Fischer, M.D. (2015). Violation of the 12/23 rule of genomic V(D)J recombination is common in lymphocytes. *Genome Res* 25(2):226-234.
- Paulsson, K. (2016). High hyperdiploid childhood acute lymphoblastic leukemia: Chromosomal gains as the main driver event. *Mol Cell Oncol* 3(1):e1064555.
- Paulsson, K. & Johansson, B. (2009). High hyperdiploid childhood acute lymphoblastic leukemia. *Genes Chromosomes Cancer* 48(8):637-660.
- Peak, M.M., Arbuckle, J.L. & Rodgers, K.K. (2003). The central domain of core RAG1 preferentially recognizes single-stranded recombination signal sequence heptamer. *J Biol Chem* 278:18235-18240.
- Peled, J.U., Kuang, F.L., Iglesias-Ussel, M.D., Roa, S., Kalis, S.L., Goodman, M.F. & Scharff, M.D. (2008). The biochemistry of somatic hypermutation. *Annu Rev Immunol* 26:481-511.
- Peng, C. & Eckhardt, L.A. (2013). Role of the Igh intronic enhancer E μ in clonal selection at the pre-B to immature B cell transition. *J Immunol* 191(8):4399-4411.
- Perez Rojo, F., Nyman, R.K.M., Johnson, A.A.T., Navarro, M.P., Ryan, M.H., Erskine, W. & Kaur, P. (2018). CRISPR-Cas systems: ushering in the new genome editing era. *Bioengineered* 9(1):214-221.
- Perkins, E.J., Kee, B.L. & Ramsden, D.A. (2004). Histone 3 lysine 4 methylation during the pre- B to immature B-cell transition. *Nucleic Acids Res* 32:1942-1947.
- Perkins, E.J., Nair, A., Cowley, D.O., Van Dyke, T., Chang, Y. & Ramsden, D.A. (2002). Sensing of intermediates in V(D)J recombination by ATM. *Genes Dev* 16(2):159-164.
- Perry, R.P., Kelley, D.E., Coleclough, C., Seidman, J.G., Leder, P., Tonegawa, S., Matthysens, G. & Weigert, M. (1980). Transcription of mouse kappa chain

- genes: implications for allelic exclusion. *Proc Natl Acad U S A* 77(4):1937-1941.
- Pinsonneault, R.L., Vacek, P.M., O'Neill, J.P. & Finette, B.A. (2007). Induction of V(D)J-mediated recombination of an extrachromosomal substrate following exposure to DNA-damaging agents. *Environ Mol Multagen* 48(6):440-450.
- Povirk, L.F. (2006). Biochemical mechanisms of chromosomal translocations resulting from DNA double-strand breaks. *DNA Repair (Amst)* 5(9-10):1199-1212.
- Pui, C.H., Relling, M.V. & Downing, J.R. (2004). Acute lymphoblastic leukemia. *N Engl J Med* 350:1535-1548.
- Qin, X.F., Schwers, S., Yu, W., Papavasiliou, F., Suh, H., Nussenzweig, A., Rajewsky, K. & Nussenzweig, M.C. (1999). *Nature* 397(6717):355-359.
- Qiu, J.X., Kale, S.B., Yarnell Schultz, H. & Roth, D.B. (2001). Separation-of-function mutants reveal critical roles for RAG2 in both the cleavage and joining steps of V(D)J recombination. *Mol Cell* 7:77-87.
- Rabbitts, T.H., Axelson, H., Forster, A., Grutz, G., Lavenir, I., Larson, R., Osada, H., Valge-Archer, V., Wadman, I. & Warren, A. (1997). Chromosomal translocations and leukaemia: a role for LMO2 in T cell acute leukaemia, in transcription and in erythropoiesis. *Leukemia* 11 Suppl 3:271-272.
- Raghavan, S.C. & Lieber, M.R. (2006). DNA structures at chromosomal translocation sites. *Bioessays* 28(5):480-494.
- Raghavan, S.C., Kirsch, I.R. & Lieber, M.R. (2001). Analysis of the V(D)J recombination efficiency at lymphoid chromosomal translocation breakpoints. *J Biol Chem* 276(31):29126-29133.
- Raghavan, S.C., Swanson, P.C., Ma, Y. & Lieber, M.R. (2005). Double-strand break formation by the RAG complex at the bcl-2 major breakpoint region and at other non-B DNA structures in vitro. *Mol Cell Biol* 25:5904-5919.
- Raghavan, S.C., Swanson, P.C., Wu, X., Hsieh, C.L. & Lieber, M.R. (2004). A non-B-DNA structure at the Bcl-2 major breakpoint region is cleaved by the RAG complex. *Nature* 428:88-93.
- Rahman, N.S., Godderz, L.J., Stray, S.J., Capra, J.D. & Rodgers, K.K. (2006). DNA cleavage of a cryptic recombination signal sequence by RAG1 and RAG2. Implications for partial V(H) gene replacement. *J Biol Chem* 281(18):12370-12380.
- Rajewsky, K. (1996). Clonal selection and learning in the antibody system. *Nature* 381(6585):751-758.
- Ramon-Maiques, S., Kuo, A.J., Carney, D., Matthews, A.G., Oettinger, M.A., Gozani, O. & Yang, W. (2007). The plant homeodomain finger of RAG2 recognizes histone H3 methylated at both lysine-4 and arginine-2. *Proc Natl*

Acad Sci U S A 104:18993-18998.

Ramsden, D.A., & Baetz, K. & Wu, G.E. (1994). Conservation of Sequence in Recombination Signal Sequence Spacers. *Nucleic Acids Res* 22(10):1785-1796.

Ramsden, D.A. & Gellert, M. (1995). Formation and resolution of double-strand break intermediates in V(D)J rearrangement. *Genes Dev* 9(19):2409-2420.

Ramsden, D.A., McBlane, J.F., van Gent, D.C. & Gellert, M. (1996). Distinct DNA sequence and structure requirements for the two steps of V(D)J recombination signal cleavage. *EMBO J* 15:3197-3206.

Ramsden, D.A., Weed, B.D. & Reddy, Y.V. (2010). V(D)J recombination: Born to be wild. *Semin Cancer Biol* 20(4):254-260.

Reddy, Y.V., Perkins, E.J. & Ramsden, D.A. (2006). Genomic instability due to V(D)J recombination-associated transposition. *Genes Dev* 20(12):1575-1582.

Reeves, R., Gorman, C.M. & Howard, B. (1985). Minichromosome assembly of non-integrated plasmid DNA transfected into mammalian cells. *Nucleic Acids Res* 13:3599-3615.

Reilly, E.B. Frackelton, A.R. & Eisen, H.N. (1982). Synthesis of lambda 1, lambda 2, and lambda 3 light chains by mouse spleen B cells. *Eur J Immunol* 12(7):552-557.

Reynaud, C.A., Anquez, V., Dahan, A. & Weill, J.C. (1985). A single rearrangement event generates most of the chicken immunoglobulin light chain diversity. *Cell* 40(2):283-291.

Rhoades, R.A. & Pflanzner, R.G. (2002). Human physiology. 4th Edition. Thomson learning. 584.

Robbiani, D.F. & Nussenzweig, M.C. (2013). Chromosome translocation, B cell lymphoma, and activation-induced cytidine deaminase. *Annu Rev Pathol* 8:79-103.

Rodgers, K.K., Bu, Z., Fleming, K.G., Schatz, D.G., Engelman, D.M. & Coleman, J.E. (1996). A zinc-binding domain involved in the dimerization of RAG1. *J Mol Biol* 260:70-84.

Rogakou, E.P., Pilch, D.R., Orr, A.H., Ivanova, V.S. & Bonner, W.M. (1998). DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J Biol Chem* 273:5858-5868.

Roldan, E., Fuxa, M., Chong, W., Martinez, D., Novatchkova, M., Busslinger, M. & Skok, J.A. (2005). Locus "decontraction" and centromeric recruitment contribute to allelic exclusion of the immunoglobulin heavy-chain gene. *Nat Immunol* 6 (1):31-41.

- Rommel, P.C., Oliveria, T.Y., Nussenzweig, M.C. & Robbiani, D.F. (2017). RAG1/2 induces genomic insertions by mobilizing DNA into RAG1/2-independent breaks. *J Exp Med* 214(3):815-831.
- Rosenfeld, C., Goutner, A., Venuat, A.M., Choquet, C., Pico, J.L., Dore, J.F., Liabeuf, A., Durandy, A., Desgrange, G. & De The, G. (1977). An effect human leukaemic cell line: Reh. *Eur J Cancer* 13(4-5):377-379.
- Roth, D.B. (2003). Restraining the V(D)J recombinase. *Nat Rev Immunol* 3(8):656-666.
- Roth, D.B. (2014). V(D)J recombination: mechanism, errors, and fidelity. *Microbiol Spectr* 2(6): MDNA3-0041-2014.
- Roth, D.B., Menetski, J.P., Nakajima, P.B., Bosma, M.J. & Gellert, M. (1992). V(D)J recombination: broken DNA molecules with covalently sealed (hairpin) coding ends in scid mouse thymocytes. *Cell* 70(6):983-991.
- Roth, D.B. & Roth, S.Y. (2000). Unequal access: regulating V(D)J recombination through chromatin remodeling. *Cell* 103:699-702.
- Ru, H., Chambers, M.G., Fu, T.-M., Tong, A.B., Liao, M. & Wu, H. (2015). Molecular mechanism of V(D)J recombination from synaptic RAG1-RAG2 complex structures. *Cell* 163(5):1138-1152.
- Ru, H., Mi, W., Zhang, P., Alt, F.W., Schatz, D.G., Liao, M. & Wu, H. (2018). DNA melting initiates the RAG catalytic pathway. *Nat Struct Mol Biol* 25(8):732-742.
- Sadofsky, M.J., Hesse, J.E., McBlane, J.F. & Gellert, M. (1994). Expression and V(D)J recombination activity of mutated RAG-1 proteins. *Nucleic Acids Res* 22:550.
- Sadofsky, M.J., Hesse, J.E., McBlane, J.F. & Gellert, M. (1993). Expression and V(D)J recombination activity of mutated RAG-1 proteins. *Nucleic Acids Res* 21 (24):5644-5650.
- Sadofsky, M.J., Hesse, J.E., van Gent, D.C. & Gellert, M. (1995). RAG-1 mutations that affect the target specificity of V(D)J recombination: a possible direct role of RAG-1 in site recognition. *Genes Dev* 9(17):2193-2199.
- Saito, H., Kranz, D.M., Takagaki, Y., Hayday, A.C., Eisen, H.N. & Tonegawa, S. (1984). A third rearranged and expressed gene in a clone of cytotoxic T lymphocytes. *Nature* 312:36-40.
- Sakano, H., Maki, R., Kurosawa, Y., Roeder, W. & Tonegawa, S. (1980). Two types of somatic recombination are necessary for the generation of complete immunoglobulin heavy-chain genes. *Nature* 286(5774):676-683.
- Sanchez, P., Nadel, B. & Cazenave, P.A. (1991). V lambda-J lambda rearrangements are restricted within a V-J-C recombination unit in the

mouse. *Eur J Immunol* 21(4):907-911.

Sawchuk, D.J., Weis-Garcia, F., Malik, S., Besmer, E., Bustin, M., Nussenzweig, M.C. & Cortes, P. (1997). V(D)J recombination: modulation of RAG1 and RAG2 cleavage activity on 12/23 substrates by whole cell extract and DNA-bending proteins. *J Exp Med* 185:2025-2032.

Schatz, D.G. & Ji, Y. (2011). Recombination centres and the orchestration of V(D)J recombination. *Nat Rev Immunol* 11:251-263.

Schatz, D.G., Oettinger, M.A. & Baltimore, D. (1989). The V(D)J recombination activating gene, RAG-1. *Cell* 59(6):1035-1048.

Schatz, D.G., Oettinger, M.A. & Schlissel, M.S. (1992). V(D)J recombination: molecular biology and regulation. *Annu Rev Immunol* 10 (1):359-383.

Schatz, D. G. & Spanopoulou, E. (2005). Biochemistry of V(D)J recombination. *Curr Top Microbiol Immunol* 290:49-85.

Schatz, D.G. & Swanson, P.C. (2011). V(D)J recombination: mechanisms of initiation. *Annu Rev Genet* 45:167-202.

Schlimgen, R.J., Reddy, K.L., Singh, H. & Krangel, M.S. (2008). Initiation of allelic exclusion by stochastic interaction of Tcrb alleles with repressive nuclear compartments. *Nat Immunol* 9 (7):802-809.

Schlissel, M. (2002). Allelic exclusion of immunoglobulin gene rearrangement and expression: why and how? *Semin Immunol* 14(3):207-212.

Schlissel, M., Constantinescu, A., Morrow, T., Baxter, M. & Peng, A. (1993). Double-strand signal sequence breaks in V(D)J recombination are blunt, 5'-phosphorylated, RAG-dependent, and cell cycle regulated. *Genes Dev* 7(12B):2529-2532.

Schlissel, M.S. & Baltimore, D. (1989). Activation of immunoglobulin kappa gene rearrangement correlates with induction of germline kappa gene transcription. *Cell* 58:1001-1007.

Schlissel, M.S., Kaffer, C.R. & Curry, J.D. (2006). Leukemia and lymphoma: a cost of doing business for adaptive immunity. *Genes Dev* 20(12):1539-1544.

Schroeder, H.W. & Cavacini, L. (2010). Structure and function of immunoglobulins. *J Allergy Clin Immunol* 125(2 Suppl 2): S41-52.

Shedlock, D.J. & Shen, H. (2003). Requirement for CD4 T cell help in generating functional CD8 T cell memory. *Science* 300(5617):337-339.

Shimazaki, N., Askary, A., Swanson, P.C. & Lieber, M.R. (2012). Mechanistic basis for RAG discrimination between recombination sites and the off-target sites of human lymphomas. *Mol Cell Biol* 32:365-375.

Shimazaki, N. & Lieber, M.R. (2014). Histone methylation and V(D)J recombination. *Int J Hematol* 100(3):230-237.

Shinkai, Y., Rathbun, G., Lam, K.P., Oltz, E.M., Stewart, V., Mendelsohn, M., Charron, J., Datta, M., Young, .F, Stall, A.M. & Alt, F.W. (1992). RAG-2-deficient mice lack mature lymphocytes owing to inability to initiate V(D)J rearrangement. *Cell* 68(5):855-867.

Showe, L.C. & Croce, C.M. (1987). The role of chromosomal translocations in B- and T-cell neoplasia. *Annu Rev Immunol* 5:253-277.

Sikes, M.L., Meade, A., Tripathi, R., Krangel, M.S. & Oltz, E.M. (2002). Regulation of V(D)J recombination: a dominant role for promoter positioning in gene segment accessibility. *Proc Natl Acad Sci U S A* 99:12309-12314.

Silva, J., Chang, K., Hannon, G.J. & Rivas, F.V. (2004). RNA-interference-based functional genomics in mammalian cells: reverse genetics coming of age. *Oncogene* 23:8401-8409.

Silver, D.P., Spanopoulou, E., Mulligan, R.C. & Baltimore, D. (1993). Dispensable sequence motifs in the RAG-1 and RAG-2 genes for plasmid V(D)J recombination. *Proc Natl Acad Sci U S A* 90:6100-6104.

Siminovitch, K.A., Bakhshi, A., Goldman, P. & Korsmeyer, S.J. (1985). A uniform deleting element mediates the loss of kappa genes in human B cells. *Nature* 316(6025):260-262.

Simkus, C., Anand, P., Bhattacharyya, A. & Jones, J.M. (2007). Biochemical and folding defects in a RAG1 variant associated with Omenn syndrome. *J Immunol* 179: 8332-8340.

Simkus, C., Makiya, M. & Jones, J.M. (2009). Karyopherin alpha 1 is a putative substrate of the RAG1 ubiquitin ligase. *Mol Immunol* 46:1319-1325.

Sleckman, B.P., Bassing, C.H., Hughes, M.M. Okada, A., D'Auteuil, M., Wehrly, T.D., Woodman, B.B., Davidson, L., Chen, J. & Alt, F.W. (2000). *Proc Natl Acad Sci U S A* 97(14):7975-7980.

Sobacchi, C., Marrella, V., Rucci, F., Vezzoni, P. & Villa, A. (2006). RAG-dependent primary immunodeficiencies. *Hum Mutat* 27:1174-1184.

Sodora, D.L., Douek, D.C., Silvestri, G., Montgomery, L., Rosenzweig, M., Igarashi, T., Bernacky, B., Johnson, R.P., Feinberg, M.B., Martin, M.A. & Koup, R.A. (2000). Quantification of thymic function by measuring T cell receptor excision circles within peripheral blood and lymphoid tissues in monkeys. *Eur J Immunol* 30(4):1145-1153.

Sonoda, E., Pewzner-Jung, Y., Schwers, S., Taki, S., Jung, S., Eilat, D. & Rajewsky, K. (1997). B cell development under the condition of allelic inclusion. *Immunity* 6 (3):225-233.

Spanopoulou, E., Zaitseva, F., Wang, F.H., Santagata, S., Baltimore, D. & Panayotou, G. (1996). The homeodomain region of Rag-1 reveals the parallel mechanisms of bacterial and V(D)J recombination. *Cell* 87(2):263-276.

Srivastava, M., Nambiar, M., Sharma, S., Karki, S.S., Goldsmith, G., Hedge, M., Kumar, S., Pandey, M., Singh, R.K., Ray, P., Natarajan, R., Kelkar, M., De, A., Choudhary, B. & Raghavan, S.C. (2012). An inhibitor of nonhomologous end-joining abrogates double-strand break repair and impedes cancer progression. *Cell* 151(7):1474-1487.

Stanhope-Baker, P., Hudson, K.M., Shaffer, A.L., Constantinescu, A. & Schlissel, M.S. (1996). Cell type-specific chromatin structure determines the targeting of V(D)J recombinase activity in vitro. *Cell* 85(6):887-897.

Stanton, M.L. & Brodeur, P.H. (2005). Stat5 mediates the IL-7-induced accessibility of a representative D-Distal VH gene. *J Immunol* 174(6):3164-3168.

Stavnezer, J., Guikema, J.E.J. & Schrader, C.E. (2008). Mechanism and regulation of class switch recombination. *Annu Rev Immunol* 26:261-292.

Stavnezer, J. & Schrader, C.E. (2006). Mismatch repair converts AID-instigated nicks to double-strand breaks for antibody class-switch recombination. *Trends Genet* 22(1): 23-28.

Steen, S.B., Gomelsky, L. & Roth, D.B. (1996). The 12/23 rule is enforced at the cleavage step of V(D)J recombination in vivo. *Genes Cells* 1:543-553.

Steen, S.B., Gomelsky, L., Speidel, S.L. & Roth, D.B. (1997). Initiation of V(D)J recombination in vivo: role of recombination signal sequences in formation of single and paired double-strand breaks. *EMBO J* 16(10):2656-2664.

Steen, S.B. Han, J.O., Mundy, C., Oettinger, M.A. & Roth, D.B. (1999). Role of the "dispensable" portions of RAG-1 and RAG-2 in V(D)J recombination. *Mol Cell Biol* 19(4):3010-3017.

Stevens, T. L., Bossie, A., Sanders, V. M., Fernandez-Botran, R., Coffman, R. L., Mosmann, T. R. & Vitetta, E. S. (1988). Regulation of antibody isotype secretion by subsets of antigen-specific helper T cells. *Nature* 334(6179):255-258.

Strob, U. Haasch, D. Arp, B. Sanchez, P., Cazenave, P.A. & Miller, J. (1989). Physical linkage of mouse lambda genes by pulsed-field gel electrophoresis suggests that the rearrangement process favors proximate target sequences. *Mol Cell Biol* 9(2):711-718.

Stros, M. (1998). DNA bending by the chromosomal protein HMG1 and its high mobility group box domains. Effect of flanking sequences. *J Biol Chem* 273:10355-10361.

Struhl, K. (2007). Transcriptional noise and the fidelity of initiation by RNA

polymerase II. *Nat Struct Mol Biol* 14(2):103-105.

Subrahmanyam, R., Du, H., Ivanova, I., Chakraborty, T., Ji, Y., Zhang, Y., Alt, F.W., Schatz, D.G. & Sen, R. (2012). Localized epigenetic changes induced by DH recombination restricts recombinase to DJH junctions. *Nat Immunol* 13(12):1205-1212.

Subrahmanyam, R. & Sen, R. (2010). RAGs' eye view of the immunoglobulin heavy chain gene locus. *Semin Immunol* 22(6):337-345.

Sun, J.C., Williams, M.A. & Bevan, M.J. (2004). CD4⁺ T cells are required for the maintenance, not programming, of memory CD8⁺ T cells after acute infection. *Nat Immunol* 5(9):927-933.

Sun, Y., Liu, Z., Ren, L., Wei, Z., Wang, P., Li, N. & Zhao, Y. (2012). Immunoglobulin genes and diversity: what we have learned from domestic animals. *J Anim Sci Biotechnol* 3(1):18.

Swanson, P.C. (2001). The DDE motif in RAG-1 is contributed in trans to a single active site that catalyzes the nicking and transesterification steps of V(D)J recombination. *Mol Cell Biol* 21:449-458.

Swanson, P.C. (2002a). A RAG-1/RAG-2 tetramer supports 12/23-regulated synapsis, cleavage, and transposition of V(D)J recombination signals. *Mol Cell Biol* 22:7790-7801.

Swanson, P.C. (2002b). Fine structure and activity of discrete RAG-HMG complexes on V(D)J recombination signals. *Mol Cell Biol* 22:1340-1351.

Swanson, P.C. (2004). The bounty of RAGs: recombination signal complexes and reaction outcomes. *Immunol Rev* 200:90-114.

Swanson, P.C. & Desiderio, S. (1998). V(D)J recombination signal recognition: distinct, overlapping DNA-protein contacts in complexes containing RAG1 with and without RAG2. *Immunity* 9:115-125.

Swanson, P.C. & Desiderio, S. (1999). RAG-2 promotes heptamer occupancy by RAG-1 in the assembly of a V(D)J initiation complex. *Mol Cell Biol* 19:3674-3683.

Swanson, P.C., Kumar, S. & Raval, P. (2009). Early steps of V(D)J rearrangement: insights from biochemical studies of RAG-RSS complexes. *Adv Exp Med Biol* 650: 1-15.

Swanson, P.C., Volkmer, D. & Wang, L. (2004). Full-length RAG-2, and not full-length RAG-1, specifically suppresses RAG-mediated transposition but not hybrid joint formation or disintegration. *J Biol Chem* 279(6):4034-4044.

Tallmadge, R.L., Tseng, C.T. & Felipe, M.J. (2014). Diversity of immunoglobulin lambda light chain gene usage over developmental stages in the horse. *Dev Comp Immunol* 46(2):171-179.

- Talukder, S.R., Dudley, D.D., Alt, F.W., Takahama, Y. & Akamatsu, Y. (2004). Increased frequency of aberrant V(D)J recombination products in core RAG-expressing mice. *Nucleic Acids Res* 32:4539-4549.
- Taylor, B., Cobb, B.S., Bruno, L., Webster, Z., Fisher, A.G. & Merckenschlager, M. (2009). A reappraisal of evidence for probabilistic models of allelic exclusion. *Proc Natl Acad Sci U S A* 106(2):516-521.
- Teng, G., Maman, Y., Resch, W., Kim, M., Yamane, A., Qian, J., Kieffer-Kwon, K.R., Mandal, M., Ji, Y., Meffre, E., Clark, M.R., Cowell, L.G., Casellas, R. & Schatz, D.G. (2015). RAG represents a widespread threat to the lymphocyte genome. *Cell* 162:751-765.
- Tevelev, A. & Schatz, D.G. (2000). Intermolecular V(D)J recombination. *J Biol Chem* 275: 83418348.
- Thompson, C.B. (1995). New insights into V(D)J recombination and its role in the evolution of the immune system. *Immunity* 3:531-539.
- Tiegs, S.L., Russell, D.M. & Nemazee, D. (1993). Receptor editing in self-reactive bone marrow B cells. *J Exp Med* 177(4):1009-1020.
- Till J.E. & McCulloch, E.A. (1961). A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. *Radiat Res* 14:213-222.
- Tillman, R. E., Wooley, A. L., Hughes, M. M., Khor, B. & Sleckman, B. P. (2004). Regulation of T-cell receptor beta-chain gene assembly by recombination signals: the beyond 12/23 restriction. *Immunol Rev* 200:36-43.
- Tillman, R. E., Wooley, A. L., Khor, B., Wehrly, T. D., Little, C. A. & Sleckman, B. P. (2003). Targeting of V beta to D beta rearrangement by RSSs can be mediated by the V(D)J recombinase in the absence of additional lymphoid-specific factors. *J Immunol* 170:5-9.
- Toledo, F. & Wahl, G.M. (2006). Regulating the p53 pathway: in vitro hypotheses, in vivo veritas. *Nat Rev Cancer* 6(12):909-923.
- Tonegawa, S. (1983). Somatic generation of antibody diversity. *Nature* 302:575-581.
- Torres, T.T. Metta, M., Ottenwalder, B. & Schlotterer, C. (2008). Gene expression profiling by massively parallel sequencing. *Genome Res* 18(1):172-177.
- Tripathi, R., Jackson, A. & Krangel, M.S. (2002). A change in the structure of Vbeta chromatin associated with TCR beta allelic exclusion. *J Immunol* 168:2316-2324.
- Tsai, A.G., Lu, H., Raghavan, S.C., Muschen, M., Hsieh, C.L. & Lieber, M.R. (2008). Human chromosomal translocations at CpG sites and a theoretical

basis for their lineage and stage specificity. *Cell* 135(6):1130-1142.

Tsai, C.L., Drejer, A.H. & Schatz, D.G. (2002). Evidence of a critical architectural function for the RAG proteins in end processing, protection, and joining in V(D)J recombination. *Genes Dev* 16(15):1934-1949.

Tsai, M.C., Manor, O., Wan, Y., Mosammamaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E. & Chang, H.Y. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. *Science* 329(5992):689-693.

Tsujimoto, Y., Finger, L.R., Yunis, J., Nowell, P.C. & Croce, C.M. (1984). Cloning of the chromosome breakpoint of neoplastic B cells with the t(14;18) chromosome translocation. *Science* 226(4678):1097-1099.

Tsujimoto, Y., Gorham, J., Cossman, J., Jaffe, E. & Croce, C.M. (1985). The t(14;18) chromosome translocations involved in B-cell neoplasms result from mistakes in VDJ joining. *Science* 229:1390-1393.

Tsung, A., Tohme, S. & Billiar, T.R. (2014). High-mobility group box-1 in sterile inflammation. *J Intern Med* 276(5):425-443.

Tsuzuki, S., Seto, M., Greaves, M. & Enver, T. (2004). Modeling first-hit functions of the t(12;21) TEL-AML1 translocation in mice. *Proc Natl Acad Sci U S A* 101:8443-8448.

Turka, L.A., Schatz, D.G., Oettinger, M.A., Chun, J.J., Gorka, C., Lee, K., McCormack, W.T. & Thompson, C.B. (1991). Thymocyte expression of RAG-1 and RAG-2: termination by T cell receptor cross-linking. *Science* 253 (5021):778-781.

Turner, D.L. & Weintraub, H. (1994). Expression of achaete-scute homolog 3 in *Xenopus* embryos converts ectodermal cells to a neural fate. *Genes Dev* 8(12):1434-1447.

Tycko, B., Reynolds, T.C., Smith, S.D. & Sklar, J. (1989). Consistent breakage between consensus recombinase heptamers of chromosome 9 DNA in a recurrent chromosomal translocation of human T cell leukemia. *J Exp Med* 169(2):369-377.

Tycko, B. & Sklar, J. (1990). Chromosomal translocations in lymphoid neoplasia: a reappraisal of the recombinase model. *Cancer Cells* 2(1):1-8.

Uhler, J.P., Hertel, C. & Svejstrup, J.Q. (2007). A role for noncoding transcription in activation of the yeast PHO5 gene. *Proc Natl Acad Sci U S A* 104(19):8011-8016.

Um, J.H., Brown, A.L., Singh, S.K., Chen, Y., Gucek, M., Lee, B.S., Luckey, M.A., Kim, M.K., Park, J.H. & Sleckman, B.P. (2013). Metabolic sensor AMPK directly phosphorylates RAG1 protein and regulates V(D)J recombination. *Proc Natl Acad Sci U S A* 110(24):9873-9878.

van Gent, D.C., Ramsden, D.A. & Gellert, M. (1996). The RAG1 and RAG2 proteins establish the 12/23 rule in V(D)J recombination. *Cell* 85:107-113.

van Gent, D.C., Hiom, K., Paull, T.T. & Gellert, M. (1997). Stimulation of V(D)J cleavage by high mobility group proteins. *EMBO J* 16:2665-2670.

van Gent, D.C. & van der Burg, M. (2007). Non-homologous end-joining, a sticky affair. *Oncogene* 26 (56):7731-7740.

van Zelm, M.C., van de Burg, M., de Ridder, D., Barendregt, B.H., de Haas, E.F., Reinders, M.J., Lnkester, A.C. Revesz, T., Staal F.J. & van Dongen J.J. (2005). Ig gene rearrangement steps are initiated in early human precursor B cell subsets and correlate with specific transcription factor expression. *J Immunol* 175(9):5912-5922.

Vanura, K., Montpellier, B., Le, T., Spicuglia, S., Navarro, J.M., Cabaud, O., Roulland, S., Vachez, E., Prinz, I., Ferrier, P. Marculescu, R., Jager, U. & Nadel, B. (2007). In vivo reinsertion of excised episomes by the V(D)J recombinase: a potential threat to genomic stability. *PLoS Biol* 5:e43.

Vettermann, C. & Schlissel, M.S. (2010). Allelic exclusion of immunoglobulin genes: models and mechanisms. *Immunol Rev* 237(1):22-42.

Vousden, K.H. & Lu, X. (2002). Live or let die: the cell's response to p53. *Nat Rev Cancer* 2(8):594-604.

Wabl, M. & Steinberg, C. (1992). Allelic exclusion model questioned. *Nature* 359 (6394):371-372.

Wang, G., Dhar, K., Swanson, P.C., Levitus, M. & Chang, Y. (2012). Real-time monitoring of RAG-catalyzed DNA cleavage unveils dynamic changes in coding end association with the coding end complex. *Nucleic Acids Res* 40:6082-6096.

Wang, X., Xiao, G., Zhang, Y., Wen, X., Gao, X., Okada, S. & Liu, X. (2008). Regulation of Tcrb recombination ordering by c-Fos-dependent RAG deposition. *Nat Immunol* 9:794-801.

Weir, H.M., Kraulis, P.J., Hill, C.S., Raine, A.R., Laue, E.D. & Thomas, J.O. (1993). Structure of the HMG box motif in the B-domain of HMG1. *EMBO J* 12:1311-1319.

Welzel, N., Le, T., Marculescu, R., Mitterbauer, G., Chott, A., Pott, C., Kneba, M., Du, M.Q., Kusec, R., Drach, J. Raderer, M., Mannhalter, C., Lechner, K., Nadel, B. & Jaeger, U. (2001). Templated nucleotide addition and immunoglobulin JH-gene utilization in t(11;14) junctions: implications for the mechanism of translocation and the origin of mantle cell lymphoma. *Cancer Res* 61:1629-1636.

West, R.B. & Lieber, M.R. (1998). The RAG-HMG1 complex enforces the 12/23 rule of V(D)J recombination specifically at the double-hairpin formation step. *Mol Cell Biol* 18:6408-6415.

- West, K.L., Singha, N.C., De Ioannes, P., Lacomis, L., Erdjument-Bromage, H., Tempst, P. & Cortes, P. (2005). A direct interaction between the RAG2 C terminus and the core histones is required for efficient V(D)J recombination. *Immunity* 23: 203-212.
- Wiederschain, D., Wee, S., Chen, L., Loo, A., Yang, G., Huang, A., Chen, Y., Caponigro, G., Yao, Y.M., Lengauer, C., Sellers, W.R. & Benson, J.D. (2009). Single-vector inducible lentiviral RNAi system for oncology target validation. *Cell Cycle* 8:498-504.
- Williams, M.A. & Bevan, M.J. (2007). Effector and memory CTL differentiation. *Annu Rev Immunol* 25(1):171-192.
- Wilson, A., Held, W. & MacDonald, H.R. (1994). Two waves of recombinase gene expression in developing thymocytes. *J Exp Med* 179:1355-1360.
- Wu, C., Bassing, C. H., Jung, D., Woodman, B. B., Foy, D. & Alt, F. W. (2003). Dramatically increased rearrangement and peripheral representation of V beta 14 driven by the 3' D beta 1 recombination signal sequence. *Immunity* 18:75-85.
- Wu, C., Ranganath, S., Gleason, M., Woodman, B. B., Borjeson, T. M., Alt, F. W. & Bassing, C. H. (2007). Restriction of endogenous T cell antigen receptor beta rearrangements to Vbeta14 through selective recombination signal sequence modifications. *Proc Natl Acad Sci U S A* 104:4002-4007.
- Xu, C.R. & Feeney, A.J. (2009). The epigenetic profile of Ig genes is dynamically regulated during B cell differentiation and is modulated by pre-B cell receptor signaling. *J Immunol* 182:1362-1369.
- Xu, C.R., Schaffer, L., Head, S.R. & Feeney, A.J. (2008). Sustained activation of STAT5 is essential for chromatin remodeling and maintenance of mammary-specific function. *Proc Natl Acad Sci U S A* 105(25):8685-8690.
- Xu, R., Nelson, C.M., Muschler, J.L., Veiseh, M., Vonderhaar, B.K. & Bissell, M.J. (2009). Sustained activation of STAT5 is essential for chromatin remodeling and maintenance of mammary-specific function. *J Cell Biol* 184(1):57-66.
- Yamane, H. & Paul, W.E. (2013). Early signaling events that underlie fate decisions of naive CD4(+) T cells toward distinct T-helper cell subsets. *Immunol Re* 252 (1):12-23.
- Yancopoulos, G.D. & Alt, F.W. (1985). Developmentally controlled and tissue-specific expression of unrearranged VH gene segments. *Cell* 40:271-281.
- Yancopoulos, G.D. & Alt, F.W. (1986). Regulation of the assembly and expression of variable-region genes. *Annu Rev Immunol* 4:339-368.
- Yarnell Schultz, H., Landree, M.A., Qiu, J.X., Kale, S.B. & Roth, D.B. (2001).

Joining-deficient RAG1 mutants block V(D)J recombination in vivo and hairpin opening in vitro. *Mol Cell* 7:65-75.

Ye, S.K., Agata, Y., Lee, H.C., Kurooka, H., Kitamura, T., Shimizu, A., Honjo, T. & Ikuta, K. (2001). The IL-7 receptor controls the accessibility of the TCRgamma locus by Stat5 and histone acetylation. *Immunity* 15:813-823.

Yin, F.F., Bailey, S., Innis, C.A., Ciubotaru, M., Kamtekar, S., Steitz, T.A. & Schatz, D.G. (2009). Structure of the RAG1 nonamer binding domain with DNA reveals a dimer that mediates DNA synapsis. *Nat Struct Mol Biol* 16 (5):499-508.

Yoshida, T., Tsuboi, A., Ishiguro, K., Nagawa, F. & Sakano, H. (2000). The DNA-bending protein, HMG1, is required for correct cleavage of 23 bp recombination signal sequences by recombination activating gene proteins in vitro. *Int Immunol* 12: 721-729.

Yu, K. & Lieber, M.R. (1999). Mechanistic basis for coding end sequence effects in the initiation of V(D)J recombination. *Mol Cell Biol* 19: 8094-8102.

Zachau, H.G. (1993). The immunoglobulin kappa locus-or-what has been learned from looking closely at one-tenth of a percent of the human genome. *Gene*. 135(1-2):167-173.

Zagelbaum, J., Shimazaki, N., Esguerra, Z.A., Watanabe, G., Lieber, M.R. & Rothenberg, E. (2016). Real-time analysis of RAG complex activity in V(D)J recombination. *Pro Natl Acad Sci U S A* 113(42):11853-11858.

Zhang, J., Ding, L., Holmfeldt, L., Wu, G., Heatley, S.L., Payne-Turner, D., Easton, J., Chen, X., Wang, J. & Rusch, M. et al. (2012). The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. *Nature* 481:1575163.

Zhang, L., Reynolds, T.L., Shan, X. & Desiderio, S. (2011). Coupling of V(D)J recombination to the cell cycle suppresses genomic instability and lymphoid tumorigenesis. *Immunity* 34:163-174.

Zhang, M. & Swanson, P.C. (2008). V(D)J recombinase binding and cleavage of cryptic recombination signal sequences identified from lymphoid malignancies. *J Biol Chem* 283(11):6717-6727.

Zhao, S., Gwyn, L.M., De, P. & Rodgers, K.K. (2009). A non-sequence-specific DNA binding mode of RAG1 is inhibited by RAG2. *J Mol Biol* 387:744-758.

Zhu, C., Mills, K.D., Ferguson, D.O., Lee, C., Manis, J. Fleming, J., Gao, Y., Morton, C.C. & Alt, F.W. (2002). Unrepaired DNA breaks in p53-deficient cells lead to oncogenic gene amplification subsequent to translocations. *Cell* 109(7):811-821.

Zhu, C., Bogue, M.A. & Roth, D.B. (1996). Thymocyte differentiation in gamma-irradiated severe-combined immunodeficient mice: characterization of intermediates and products of V(D)J recombination at the T cell receptor alpha

locus. *Eur J Immunol* 26:2859-2865.

Zubakov, D., Liu, F., van Zelm, M.C., Vermeulen, J., Oostra, B.A., van Duijn, C.M., Driessen, G.J., van Dongen, J.J., Kayser, M. & Langerak, A.W. (2010). Estimating human age from T-cell DNA rearrangements. *Curr Biol* 20:R970-971.