UNIVERSITY OF LEEDS

Regulation of gene expression in the immune system and in virally-transformed cells

Irene Bassano

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

June 2014

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

Acknowledgments

I am deeply indebted to my supervisors, Prof Eric Blair, Dr Graham Cook and Dr Lucy Stead for their invaluable help and training throughout my PhD. Particularly, I would like to express my gratitude to Eric for allowing me to freely explore my scientific ideas and to grow as an independent research scientist.

Thanks to all the members of the Blair's group, Dr David Orchard-Webb, Dr James Findlay, Dr Laura White, Adam Lee, Richard Hogg and the current members of the Virology group for their support, help and friendship.

I will always be grateful to Dr Lawrence Banks for hosting me in his lab and for his priceless advice on my project. Thanks also to his lab members for making my time in Trieste an intellectually rewarding experience; in particular I am indebted to Dr Paola Massimi for her patience, guidance and invaluable help with the project.

Thanks to my friends who made the time in Leeds an enjoyable experience. In particular, Fr Peter Kravos from the Leeds Catholic Chaplaincy for helping me in the hardest time of my PhD and to my friend Girish for being there from the very first day of my PhD, listening to all my lab experiences. Thank you for entertaining me with your stories over lovely dinner times.

I am deeply thankful to Julian without whom there will be no PhD. Thank you for your patience, unconditional love and understanding throughout these years and for encouraging me to never give up.

Special thanks to my family: Clelia, Andrea, Elena, Francesco and Maria for their constant encouragement to strive towards my goals. Thanks also to Emiliana, Lamberto, Simona and Sara for their support ever since I can remember and for making me truly part of their family.

Finally, thanks with all my heart to my lovely mum Abeba to whom I dedicate this work.

Abstract

The correct development and functioning of the immune system is critical for the defence of the host organism against pathogens and cancers. V(D)J recombination generates diversity of immunoglobulin (Ig) and T cell receptor (TCR) genes by the regulated joining of variable (V), diversity (D) and joining (J) gene segments. Tissue-specific enhancers in the DNA genome activate these genes to undergo recombination by triggering non-coding transcription through the recombining gene segments, following interaction with the respective promoters. How this is achieved is unknown.

The specificity of enhancer/promoter interactions was examined using the murine Igλ chain locus. The transcription factors that bind to the three main promoters were identified by DNase I footprinting. Of these, a factor termed E47 was shown to interact with IRF4 by co-immunoprecipitation experiments. The importance of these interactions was confirmed by mutagenesis where it was shown that mutations of any of the binding sites in DNA for the transcription factors or mutations in the amino acids involved in protein-protein interactions decreased the rate of transcription. Together, these studies suggest that IRF4/E47 interactions may play a key role in triggering locus activation.

RNA-Seq data from HPV-positive samples and cell lines were analysed to identify putative biomarkers for cervical cancer. Infection with HPVs is the main cause for cervical cancer accounting for 10-15% of cancer-related deaths in women worldwide. It is established that HPVs escape the immune response over decades to establish tumorigenesis but the specific mechanism is unknown. Virus integration into the host genome and deregulation of several genes may play a key role in promoting cancer; of particular interest are those transcripts that form the “surfacesome”. Among these, particular interest was given to connexin 26 (Cx26), which is classified as cancer-predisposition gene and was found to be commonly down regulated in all samples analysed.

Recombinant adenoviruses expressing the two HPV16 oncogenes were generated and employed to transduce HaCaT cells to analyse Cx26 mRNA and protein levels coupled with dye transfer assays to study the structural behaviour of connexins. The data presented showed that E6 and E7 alter Cx26 protein expression by relocating Cx26 within the cytoplasm from the membrane-bound form. This was confirmed in the dye transfer assay where cell-cell communications were lost.

Table of contents

List of Tables

List of Figures

Abbreviations

5-Aza-CdR	5-Aza-2'-deoxycytidine
5-CMFDA	Chloromethylfluorescein Diacetate
ADCC	Antibody dependent cellular cytotoxicity
Ads	Adenoviruses
AHR	Aromatic hydrocarbon receptor
AID	Activation-induced cytidine deaminase
AML	Acute myeloid leukaemia
AP-1	Activator protein 1
APN	Aminopeptidase N
APS	Ammonium persulphate
ASCUS	Atypical squamous cells of undetermined significance
ATM	Ataxia telangiectasia mutated
ATM-ATR	Ataxia telangiectasia-mutated and RAD3-related DNA damage response
ATP	Adenine Triphosphate
BACs	Bacterial artificial chromosomes
BAK	Bcl2 antagonist killer
BCR	B-cell receptor
BER	Base excision repair
bHLH	Basic helix loop helix
BMQC	Bromomethyl derivative of coumarin
BPV	Bovine papilloma virus
BSA	Bovine Serum Albumin
Btk	Bruton's Tyrosin kinase
cAMP	Cyclic adenosine monophosphate
CBP	CREB-binding protein
CD	Cluster of differentiation
CDK	Cyclin dependent kinases
CDKi	Cyclin dependent kinases inhibitors
cDNA	Complementary DNA
CDR	Complementarity-determining regions
CDS	Coding sequence
CFP	Cyan fluorescent protein
ChIP	Chromatin immunoprecipitation
CIN	Cervical intraepithelial neoplasia
CIP	Calf intestinal alkaline phosphatase
CKII	Casein kinase II
CLPs	Common lymphoid progenitors
CMV	Cytomegalovirus
CNV	Copy number variation

CR	Conserved regions
cRSSs	Cryptic recombination signal sequences
CSR	Class-switch recombination
CTCF	CCCTC-binding factor
Cx	Connexin
CyPB	Cyclophilin B
DAPI	4',6-diamidino-2-phenylindole
DBP	DNA-binding protein
DEPC	Diethylpyrocarbonate
DMEM	Dulbecco's Minimal Eagle Medium
DNA	Deoxyribose nucleic acid
DNMTs	DNA methyltransferases
dNTPs	Deoxynucleoside triphosphates
dsDNA	Double stranded DNA
E6AP	E6-associated protein
EBF	Early B cell factor
ECM	Extracellular matrix
EDTA	Ethylenediaminetetra acetic acid
EGF	Epidermal Growth Factor
EGFR	Epidermal growth Factor receptor
EGFR	Epidermal growth factor receptor
EHD	E protein homeobox domain
EMSA	Electrophoretic shift assay
ER	Estrogen receptor
Fab	Fragment antigen binding
Fc	Fragment crystallisable
FCS	Foetal calf serum
FDR	False discovery rate
FITC	Fluorescein isothiocyanate
FOG-1	Friend of GATA-1
FPKM	Fragments per kilobase of exon per million mapped fragments
GAPDH	Glyceraldehyde 3-Phosphate Dehydrogenase
GC	Germinal centre
GFP	Green-fluorescent protein
GJICs	Gap-junction intracellular channels
GJs	Gap junctions
GTF	General transcription factor
HAT	Histone Acetyltransferase
HBS	Hepes-buffered saline
HD	Homeodomain
HDAC	Histone deacetylases

hDlg	Human disks large homolog 1
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPV	Human papilloma virus
HR-HPV	High risk HPV
hScribble	Human homologue of *Drosophila* Scribble
HSPG	Heparan sulphate proteoglycan
HSs	Hypersensitive sites
HSV	Herpes Simplex Virus
hTERT	Human telomerase reverse transcriptase
HYM	Hypermutation domain
Id	Inhibitors of differentiation
IF	Immunofluorescence
IFN	Interferon
IPTG	Isopropyl-ß-D-thio-galactopyranoside
IRESs	Internal ribosome entry sites
ITAMs	Immunoreceptor tyrosine based activation motifs
kDa	Daltons
KGFR	Keratinocyte growth factor receptor
LAP	Leucine-rich and PDZ domain
LCR	Locus control region
Lef-1	Lymphoid Enhancer Factor-1
LIM	Lin11, Isl-1 & Mec-3
LZ	Leucine zipper motif
MADS	MCM1, Agamous and Deficiens,SRF
MAGI-1	Membrane associated guanylate kinase 1
MAGUK	Membrane-associated guanylate kinase homologue
MAR	Matrix attachment region
MCM1	Minichromosome maintenance protein 1
MFI	Mean fluorescence intensity
MHC	Major Histocompatibility Complex
MLV	Murine Leukaemia Virus
MPPs	Multipotent progenitors
mRNA	Messenger Ribonucleic Acid
MWCO	Molecular weight cut off
ncLCR	Non-coding long control region
NGS	Next generation sequencing
NHK	Normal human keratinocytes
Oct	Octamer
OD	Optical density
ORF	Open reading frame
PAMPs	Pathogen-associated molecular patterns

Pax5	Paired box 5
PBS	Phosphate Buffered Saline
PCAF	P300/CPB-associated Factor
PCR	Polymerase Chain Reaction
PCV	Packed cell volume
PDZ	PSD-95/Dlg/ZO-1
PEST	Proline, glutamic acid, serine, threonine
PFU	Plaque Forming Unit
PI3K	Phosphoinositide 3-kinase
PIC	Pre-initiation complex
PIP	PU.1 interacting partner
PK	Protein Kinase
PLB	Passive lysis buffer
PNK	Polynucleotide kinase
Pol	Polymerase
PolydI-dC	Poly(deoxyinosinic-deoxycytidylic) acid
PPRs	Pattern recognition receptors
PTK	Protein tyrosine kinase
RAG	Recombination activating gene
Rb	Retinoblastoma
RIG-I	Retinoic acid--inducible gene I
RIPA	Radio immunoprecipitation assay
RLU	Relative luminescence units
RNA	Ribo Nucleic Acid
RNA Pol II	RNA polymerase II
RNA-Seq	RNA Sequencing
RPA	Replication protein A
RPM	Revolution per minute
RPMI	Roswell Park Memorial Institute
RSSs	Recombination signal sequences
RT-PCR	Reverse-transcriptase Polymerase Chain Reaction
SAGE	Serial analysis of gene expression
SAR	Scaffold attachment regions
SD	Standard deviation
SEM	Standard error mean
SNPs	Single nucleotide polymorphisms
SRF	Serum-response factor
ssDNA	Single stranded DNA
sst2	Antiproliferative somatostatin receptor 2
STR	Short tandem repeat
SV40	Simian virus 40

TAD	Transactivation domain
TAF	TFIID associated factors
TBE	Tris Borate EDTA
TBP	TATA binding protein
TBS	Tris buffered saline
TBST	TBS-Tween
TCR	T- cell receptor
TdT	Terminal-deoxynucleotidyl transferase
TESS	Transcription Element Search System
TFII	Transcription factor II
TGS	Tris Glycine SDS
TJ	Tight junction
TSS	Transcription start site
UTR	Untranslated region
VLPs	Virus-like particles
WB	Western blotting
YACs	Yeast artificial chromosomes
ZAP-70	Zeta-associated protein
ZO-1	Zonula occludens 1

Chapter 1. INTRODUCTION

1.1 Overview of the immune system: innate and adaptive immunity

Mammals are constantly exposed to a variety of harmful pathogens and microbes, which can invade the host through the skin, via food ingestion, inhalation, or through the genitourinary tract. The immune system protects the body from pathogens in two ways: first by posing physical barriers such as the skin and the mucosa or secreting specific molecules (lysozyme produced in the eyes, stomach acids or skin oils); secondly by producing antigen-specific antibodies and T cell receptors. The first line of defence represents the innate immune system, while the second falls within the adaptive system (Figure 1-1). When a pathogen overcomes these physical or chemical barriers, then more specialised molecules and cell types intervene, from both the innate and adaptive immune system.

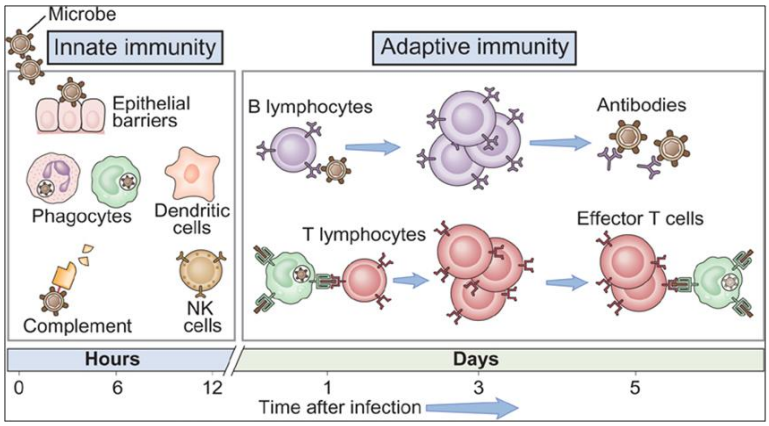


Figure 1-1 Differences between innate and adaptive immunity.

The first line of defence involves epithelial barriers such as skin as well as a cellular component represented by phagocytes, dendritic cells and NK cells. In adaptive immunity, which is more specific, the immune response is mainly mediated by B and T cells through generation of antibodies by B cells and T cell recognition of infected and transformed cells (Abbas, 2010).

1.1.1 Cells of the immune system

Both innate and adaptive immune responses have a wide range of cells that intervene in the case of an infection. All cells originate from haemapoietic stem cells (HSCs) that give rise to the myeloid and lymphoid lineage. The myeloid lineage gives rise to granulocytes (which includes neutrophils, basophils, eosinophils and mast cells) as well as the monocyte/dendritic cell progenitor from which macrophages and dendritic cells originate (Liu & Nussenzweig, 2010). The lymphoid lineage includes T cells (helper, cytotoxic and suppressor), B cells (plasma cells and memory cells) and natural killer (NK) cells (Figure 1-2). B lymphocytes originate and mature in the bone marrow and they are capable of developing into antibody-secreting plasma cells following stimulation by antigens; after a first infection, they are also able to differentiate into memory B cells, which retain an immunological memory of the antigen. In addition, B cells also contain the membrane-bound form of the antibody, namely the B cell receptor (BCR). T cells originate in the bone marrow but mature in the thymus; they are divided into T helper (T_H, $CD4^+$) or cytotoxic (T_C, $CD8^+$) and they both play a key role in coordinating the immune response. T helper cells "see" peptide antigen in the context of major histocompatibilty complex (MHC) class II molecules on antigen presenting cells (APCs) and stimulate the secretion of certain lymphokines which in turn trigger responses from other cells of the immune system (such as B cells and T_C), while T_C are involved in killing infected or tumour cells following presentation of the antigen by MHC class I molecules. Both B and T cells undergo a stringent mechanism of selection that eliminates self-reacting cells.

The same lymphoid progenitor from which B and T cells originate also gives rise to NK cells. These cells kill virus-infected cells as well as tumour cells, in a similar manner to T_C cells, however they do not require recognition of antigen in association with MHC molecules.

In the adaptive immune response, macrophages and dendritic cells play a key role. Both cells are APCs as they present on their surface epitopes of the ingested pathogen following phagocytosis.

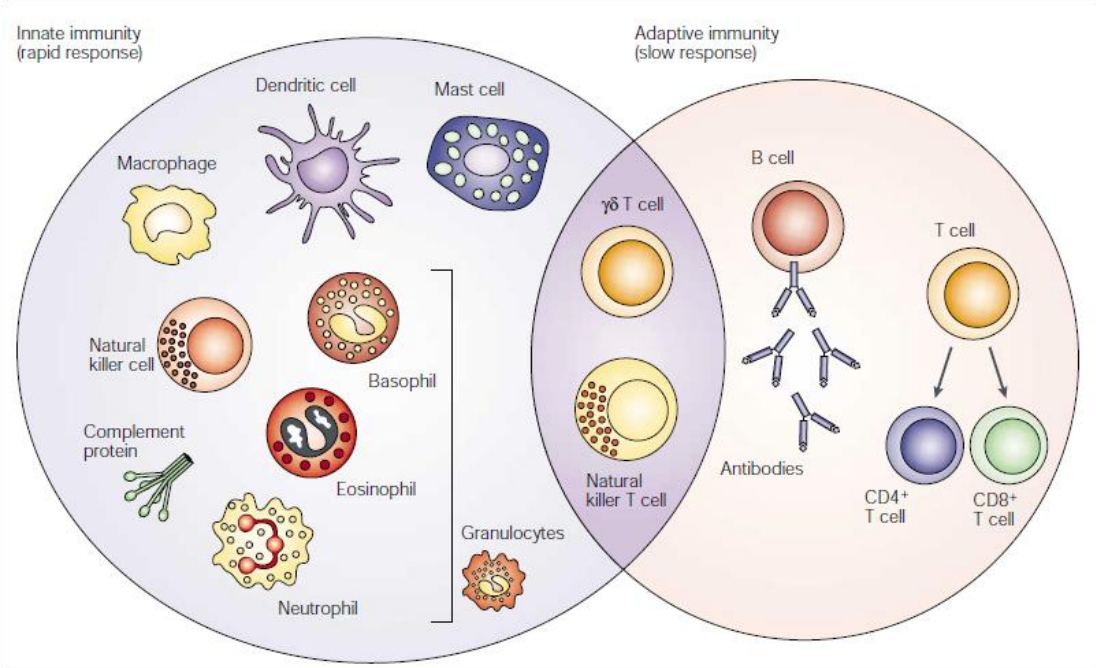


Figure 1-2 Cells of the innate and adaptive immune response.

A lymphoid and a myeloid progenitor give rise to the vast number of cells involved in the immune system (Dranoff, 2004).

1.1.2 Innate immunity

The innate immune system represents the first line of defence against pathogens including bacteria, viruses, fungi and parasites. Physical barriers, such as the skin (epithelial cells) and the linings of the body's tubular structures (mucosal epithelial containing glycoproteins such as mucins) help prevent and reduce infections. In addition, the human body produces tears and saliva, which contain lysozymes and phospholipase that act against the cell wall of many bacteria, to protect the mouth and eyes. The low pH environment that prevents growth of many bacteria protects the gastrointestinal tract. In case of a microtrauma however, pathogens will enter the body and, at that site where infection has occurred, the innate response will

intervene with the recruitment phagocytic cells, as well the innate humoral response which activates proteins of the complement system. Activation of the complement system triggers increased vascular permeability, recruitment of phagocytic cells, and lysis and opsonization of bacteria. In addition, the serum provides other factors such as lactoferrin and transferrin which limit bacteria growth by depriving them of iron, as well as interferons (INFs), which are specific for response against viruses (Abbas, 2010).

1.1.2.1 Phagocytosis in the innate response

Cells of the innate immune system are able to destroy and kill pathogens by phagocytosis. Mainly dendritic cells, neutrophils and macrophages perform this. Following the entry of the pathogen, these cells recognize pathogen-associated molecular patterns (PAMPs) which are absent from the host cells and are found in both pathogenic and non-pathogenic micro-organisms (Medzhitov, 2007). PAMPs include LPS (lipopolysaccharide) and other components of bacterial cell walls, unmethylated CpG islands or double strand RNA (dsRNA). Upon recognition via Pattern Recognition Receptors (PRRs), the pathogen is endocytosed and digested in a phagosome (Figure 1-3). This digestion releases peptides that can be presented at the plasma membrane by MHC class II molecules. PRRs also include the large family of Toll Like Receptors (TLRs); for example, TLR-4 recognise PAMPS (LPS) and activate NF-kB dependent transcription of genes encoding inflammatory cytokines such as TNF-α (tumour necrosis factor-α); TNF is pro-inflammatory and, as well as activating innate immunity, it prepares the immune system to initiate adaptive responses by delivering maturation signals to dendritic cells. Other factors such as, chemokines and Interleukin-1 (IL-1) serve to attract other cell types to the site of the infection (Fauriat et al, 2010) .

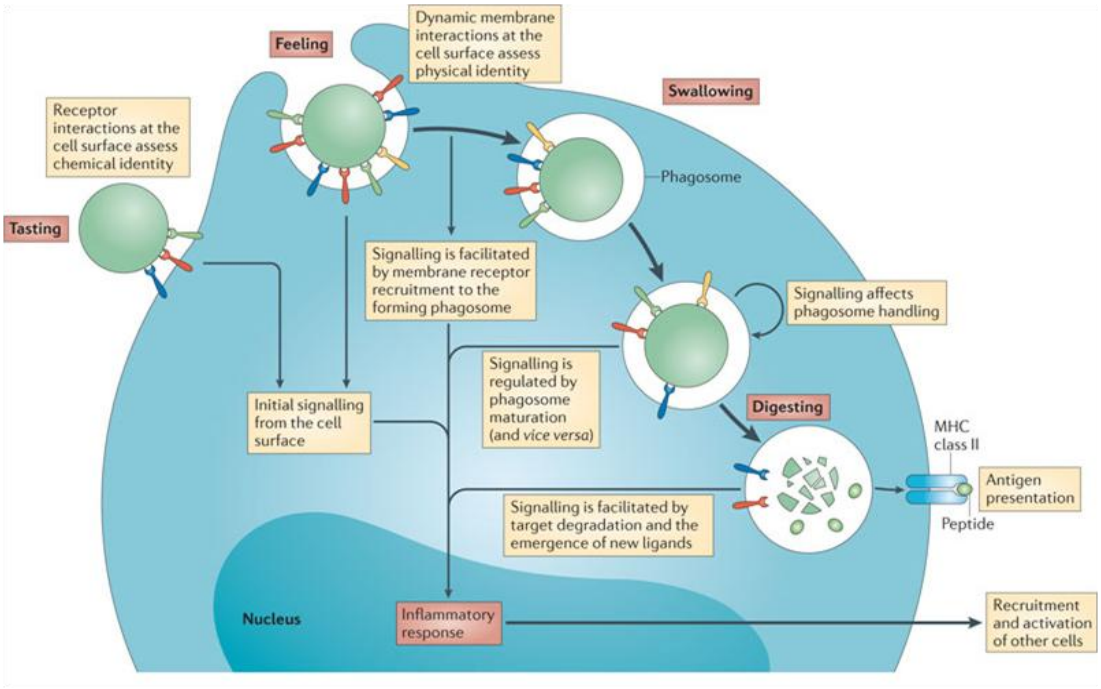


Figure 1-3 Mechanism of phagocytosis.

Phagocytes are attracted to microorganisms via chemotaxis; macrophages have cell-surface receptors that recognize molecules on the surface of various pathogens. The phagocytes will engulf the microorganism and enclose it in a phagocytic vesicle to complete ingestion. Most of the ingested microorganisms are killed by lysosomal enzymes and oxidizing agents (Underhill & Goodridge, 2012).

1.1.2.2 Role of NK cells in the innate immune response

NK cells develop in the foetal liver and, after birth in the bone marrow. NK cells play a key role in the innate immune response as they are responsible for killing infected, as well as malignant cells by producing granzymes and perforin which are the major effector proteins responsible for killing infected cells and tumours (Chiang et al, 2013) by both NK and T_C cells.

NK cells can kill target cells by two main mechanisms; antibody dependent cellular cytotoxicity (ADCC) and via the detection of altered expression of host proteins on the target cell surface (Tyler et al, 1989). For ADCC, IgG coated target cells are recognized by the Fc receptor CD16. For altered self, NK cells express an array of cell surface receptors that detect infection and tumour-induced proteins (such as NKG2D ligands) and deliver activating signals to the NK cell and inhibitory receptors

for MHC class I that, when not engaged, fail to counteract these activating signals (the missing self-model; (Karre et al, 1986)). A net balance of cellular activation leads to the release of granzymes and perforin onto the target cells and the release of cytokines such as TNF-α and IFNγ. NK cells also receive activating signals from DCs, coupling them to other pathways of innate immune activation (Amadei et al, 2010; Narni-Mancinelli et al, 2013).

1.1.2.3 Immune response to viral infection

Viruses are intracellular parasites that use the host cell machinery for replication. Infections are first counteracted by the innate immunity that stops viral invasion and replication. NK cells and the production of type I interferons (IFNα and IFNβ) therefore play a major role. More specifically, viral RNA or DNA is recognized by endosomal TLRs such as TLR-7, -8, -9 which recognize ssRNA and dsDNA, while TLR-3 recognizes dsRNA (Meylan et al, 2006). These are found on plasma and endosomal membrane of various cell types such as B cells, phagocytes and endothelial cells. When viral particles are encountered in the cytosol compartment, such as following replication, viral nucleic acid is recognized by receptors such as RIG-like receptors, RLRs (i.e. RIG-I and MDA-5). Ultimately, both pathways, TLR or RLR-mediated, lead to the activation of protein kinases, which in turn activate the IRF transcription factors in the nuclei to stimulate type I interferon gene transcription (Takaoka et al, 2005). The latter includes IFN-α isoforms, a single IFN-β, and other IFN family members (IFN-ε, -κ -ω and many others) (Stetson & Medzhitov, 2006).

However, when the infection is persistent and the innate immune system fails to stop the infection, higher virus titre activate the adaptive immunity with the production of antibodies and T_C cells aimed to kill infected cells. Antibodies bind to viral envelope where present or capsid antigens to neutralize particles and prevent virus attachment and entry into host cells (ADCC system, classical pathway of complement activation, phagocytosis). However, CD8+ TC cells that recognize cytosolic viral peptides presented by MHC class I molecules mainly carry out elimination of an in-

fected cell. Infected cells could be either APCs or a tissue cell. In the latter case, dendritic cells that process the viral antigens and present them to naive CD8+ TC cells phagocytise these cells. Activated T_C cells differentiate into effector CTLs, which can kill any infected nucleated cell, and in addition are also able to secrete IFN-γ, potent antiviral cytokine.

However, viruses have evolved numerous mechanisms for evading host immunity. A good example, as described in Chapter 1.4, is exemplified by the human papilloma virus (HPV).

1.1.3 Adaptive immunity

Adaptive immunity, or the acquired immune response, represents the second line of defence, when the less specific, innate immunity fails. The adaptive immune response is carried out by B and T lymphocytes: these are able to recognize a seemingly limitless repertoire of antigens and humans can produce over 10^7 lymphocytes, each specific for a particular antigen. Once an antigen or more precisely an epitope from that antigen binds to its receptor on the surface of B and/or T cells, clonal expansion of lymphocytes will begin, producing a vast number of that type of cell with its associated antigen specificity. While these cells are highly specific against that antigen, the response to completely eliminate the pathogen will take few days. On the other hand, since lymphocytes are able to retain an immunological "memory", upon a second infection with the same pathogen, the response will be much quicker (Figure 1-4) (Abbas, 2010).

Adaptive immunity can be divided into branches: humoral immunity, which mediated by B cells via production of antibodies, and cell-mediated immunity, which involves T cells. When the innate immune system fails to prevent an infection, antigens are transported to the lymphoid organs where they will be in contact with naïve B and T cells. The contact will trigger activation of the lymphocytes which differentiate into effector cells. Epitopes of the antigen are recognized by the B cell re-

ceptor (BCR) and T cell receptor (TCR) on the surface of B and T cells, respectively. While B cells can bind directly to the epitopes, T cells require the antigen (mainly short peptides) to be presented on the surface of APCs such as macrophages, dendritic cells and B lymphocytes via MHC molecules.

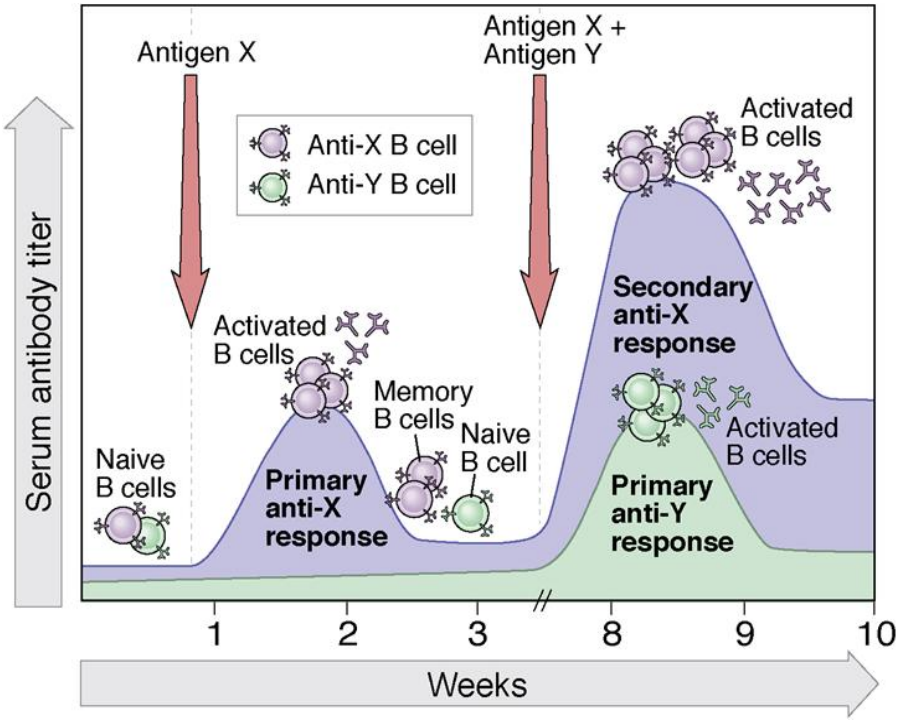


Figure 1-4 Primary and secondary immune response.

The graph shows specificity and memory of adaptive immunity compared to the innate immune response (Abbas, 2010).

MHC molecules regulate both the innate and adaptive immune responses where they bind to epitopes of the processed antigen and present it on the surface of the cells for recognition by T cells (as well as NK cells). MHC molecules are divided into two classes: class I and class II and they present antigens to $CD8^+$ T and $CD4^+$ T lymphocytes, respectively. MHC class I molecules are expressed on almost all nucleated cells of the body, while MHC class II molecules are mainly expressed on the surface of APCs (such as dendritic cells, macrophage and B cells) which have trapped exogenous antigens from the extracellular fluid and are able to interact with helper T cells.

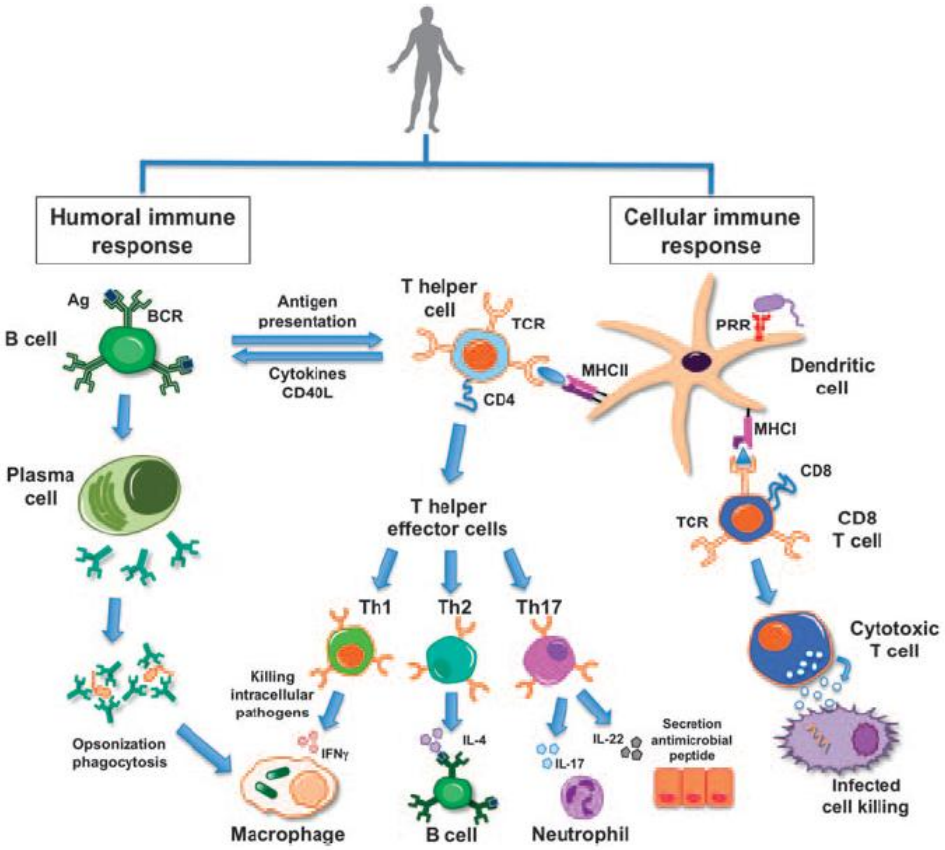


Figure 1-5 Type of adaptive immune response.

In the humoral response, B cells are activated by T helper cells into antibody-secreting plasma cells that will trigger opsonisation of the microorganisms. In the cellular immune response, interaction with the APCs can either trigger activation of T cells into T_H cells or, in case of interaction with MHC class I molecules on the surface of the APCs, activation of T_C cells which will kill the infected cells (Mortellaro & Ricciardi-Castagnoli, 2011).

Since MHC molecules are highly polygenic and polymorphic, it is unlikely that a pathogen will easily evade the immune response, both innate and adaptive (Alberts et al, 2002). Indeed, being polygenic confers these molecules the ability to produce a vast array of peptides allowing them to recognize different epitopes. Secondly, being polymorphic, each copy of the genes can exists in different isoform, consequently generating different forms of MHC molecules, increasing the ability to recognize and bind different antigens for presentation (Alberts et al, 2002). Dendritic cells express both classes of MHCs and following binding of the foreign antigen, they present it to naïve $CD8^+$ or $CD4^+$ T cells, triggering their activation into effector

cells. Macrophages express MHC II and they present antigens on their surface for recognition by $CD4^+$ T cells: following this interaction, macrophages are activated by the release of cytokines by T cells to kill intracellular pathogens and chemokines to attract them to the site of infection. In addition, B cells express MHC class II molecules and they present the antigen to $CD4^+$ T cells that will in turn activate B cell into antibody- secreting plasma cells. Thus B cells and T cells interact to orchestrate both humoral and cell mediated immunity. The key to their specificity is the presence of clonally expressed antigen receptor molecules, immunoglobulin and the T cell receptor.

1.1.4 Lymphocyte development

Lymphocytes are white blood cells that originate in the bone marrow (B lymphocytes) and in the thymus (T cells). While B cells generate and mature in the bone marrow, T cells migrate to the thymus from the bone marrow to complete the maturation process. Both B and T cells then migrate to the secondary lymphoid organs such as spleen and lymph nodes to encounter the antigen and be eventually activated. Humans have an average of $2x10^{12}$ B and T lymphocytes. They are the only cells capable of undergoing V(D)J recombination, the process by which B and T cell receptors are generated (Chapter 1.2). Lymphocyte development occurs in a number of different stages. It is an ordered process that parallels the progressive assembly and expression of the antigen receptor genes on the surface o B and T cells.

Lymphocytes originate from pluripotent HSCs. HSCs give rise to multipotent progenitors (MPPs) that in turn develop into early lymphocyte progenitors (ELPs) (Traver et al, 2000). ELPs up-regulate the lymphoid-specific recombinase genes Rag1 and Rag2, the only proteins required to initiate V(D)J recombination, and differentiate into common lymphoid progenitors (CLPs) that are restricted to B, T and NK cell development (Kondo et al, 1997) (Figure 1-6).

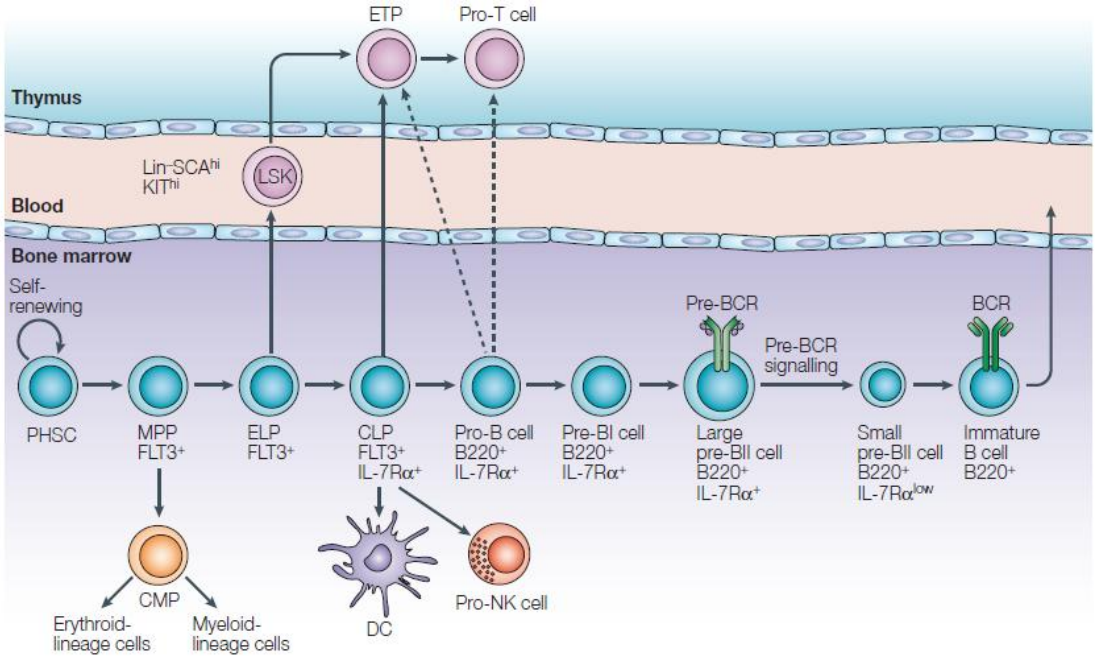


Figure 1-6 Overview of lymphocyte development from human stem cells.

Both B and T cells orginate in the bone marrow from human stem cells. B cells mature in the bone marrow, while T cells migrate to the thymus to complete their maturation (Matthias & Rolink, 2005).

Pluripotent haematopoietic stem cells (HSCs) give rise to all blood cells in foetal liver and bone marrow. The lymphoid lineage, B, T and NK cells, derive from multi-potent progenitors (MPPs). MPPs differentiate into early lymphocyte progenitors (ELPs), which in turn differentiate into common lymphoid progenitors (CLPs) that are restricted to B, T, NK and DC cell development. A fraction of CLPs enter the B cell differentiation pathway and those cells that successfully complete DH-JH rearrangements will constitute the early pro-B cell stage (Brass et al, 1996; Hardy & Hayakawa, 1991; Li et al, 1993). Following productive VH-DJH rearrangement during the late pro-B cell stage, the Igμ chain is expressed on the surface of the cell together with the surrogate light chains, lambda 5 and V pre-B (Melchers, 2005). This complex is known as the pre-B cell receptor (pre-BCR). The pre-BCR represents an important checkpoint in B-cell development. Signals through the pre-BCR are required for B cell proliferation and subsequent recombination of the κ or λ light chain gene. The latter triggers differentiation of pre-B cells to immature B cells.

1.1.4.1 B cell development

B cells development initiates in the foetal liver and continues in the bone marrow throughout the life. During an on going infection, B cells migrate to the secondary lymphoid organs where they will be activated following interaction with a peptide-derived antigen. B cell development parallels the generation of the antigen receptor loci: only when a mature BCR is produced, will naïve B cells migrate to the periphery where they will encounter the antigen and will be activated into effector B cells. Development begins when the lymphoid progenitors start expressing the recombinase activating genes, RAG1 and RAG2, in $CD34^+$ cells (for more details about V(D)J recombination refer to Chapter 1.2). These cells first undergo D-J joining on the heavy chain and also express CD45 (B220) and MHC class II molecules, which defines early pro-B cells, with limited self-renewal capacity. V-to-DJ joining on the heavy chain completes the late pro-B cell stage defined by the expression of the μ heavy chain (Packard & Cambier, 2013); at this stage, heavy chain rearrangement is therefore stopped to promote light chain recombination: these cells express the surrogate light chains lambda 5 and V pre-B (pre-BCR) and are characterised by the expression of signal transduction molecules IgαIgβ from the earlier stage; since the heavy chain is not capable of transmitting signals to the cytoplasm due to its short tail, this is carried by the signal transduction molecules which have ITAM motifs (Immunoreceptor Tyrosine Activation Motifs) (Niiro & Clark, 2002). This triggers a series of cascade reactions which promotes large pre-B cell proliferation and development to the next stage, into small pre-B cell (Miosge & Goodnow, 2005). When these cells have successfully rearranged the light chain, lambda or kappa, a complete immunoglobulin, IgM, is produced and expressed on the cell surface. These cells are defined as immature B cells and they are still found in the bone marrow. Their complete maturation occurs in the peripheral lymphoid organs once they encounter a peptide-derived antigen (Figure 1-7).

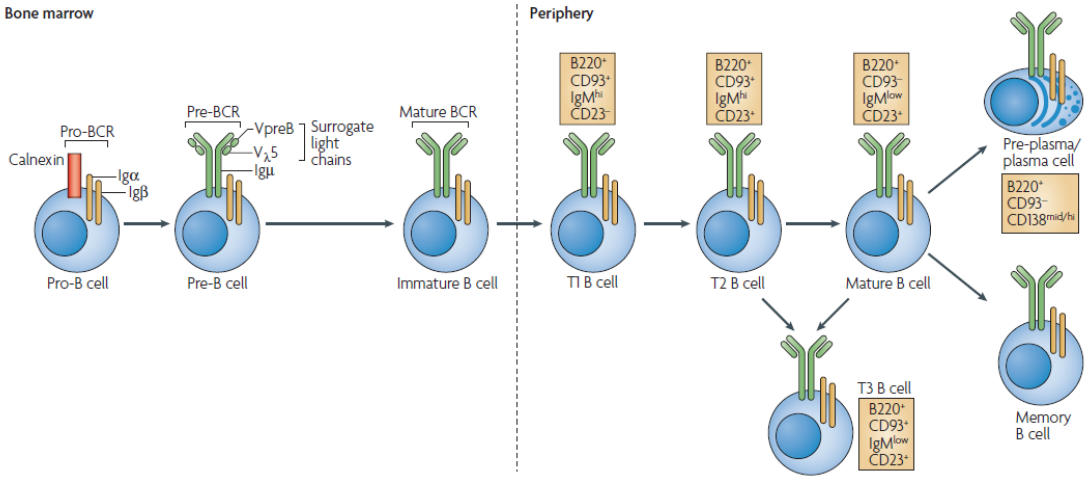


Figure 1-7 B cell development in the bone marrow.

B cells originate and mature in the bone marrow where parallel generation of the antibody receptors occur. In the bone marrow B cells proliferate from pro-B cells to immature B cells which express a complete B cell receptor; however, only migration to secondary lymphoid organs and interaction with an antigen will complete the maturation process into plasma or memory cells (Cambier et al, 2007).

1.1.4.1.1 Transcriptional regulation of B lymphocytes development

B cell development is regulated by the combination of different transcription factors including Ikaros, PU.1, early B-cell factor (EBF), E2A (E12 and E47), Pax5, nuclear factor (NF)-κB, interferon regulatory factor 4 (IRF4) and Oct2 (Schebesta et al, 2002). Mice deficient for E2A, PU.1, EBF or Pax5 have B cell development arrested prior to rearrangement and expression of the Ig heavy chain genes (at the pro-B cell stage (Bain et al, 1994; Georgopoulos et al, 1994; Scott et al, 1994; Zhuang et al, 2004; Zhuang et al, 1994)).

Ikaros and PU.1 are important for the generation of lymphoid progenitors: Ikaros-/- mice fail to generate any B, T or NK cells (Georgopoulos et al, 1994; Kirstetter et al, 2002; Wu et al, 1997), while PU.1-/- mice die around birth and lack B, T and myeloid cells, suggesting that PU.1 is involved in lymphoid and macrophage lineage determination (McKercher et al, 1996; Scott et al, 1994). E2A, EBF, and Pax5 control the development of CLPs into committed pro-B cells (Schebesta et al, 2002). Mice deficient for the E2A gene arrest B cell development at the very first stages of B cell

development before rearrangement of the immunoglobulin heavy chain (IgH) (Bain et al, 1994; Bain et al, 1997; Zhuang et al, 1994). These mice fail to undergo DH-JH rearrangement because Rag1 is not expressed and thus V(D)J recombination fails to initiate.

E proteins, to which E2A belongs to, were first discovered due to their ability to bind sequences within the IgH and Igκ loci, specifically the IgH intronic enhancer (μE5) and the κE3' and κEi sites of the Igκ enhancers. Cells lacking these sites, and thus inability of E proteins to bind and exert their effect, developed severe recombination defects (Afshar et al, 2006; Inlay & Xu, 2003; Schlissel, 2004). Studies on the Igκ locus have shown that E47 can bind the κEi enhancer and moreover can also bind the Igκ 3' enhancer upon interaction with interferon regulatory factor 4 (IRF4) (Lazorchak et al, 2006; Nagulapalli & Atchison, 1998). These studies within the Igκ locus show that in E2A-deficient pre-B cells, the B cell stage at which Igκ rearrangement occurs, Igκ germ line transcription is severely impaired and recombination is blocked in the absence of E47/E12. Ectopic expression of E47/E12 and therefore recruitment to κEi and κE3' correlates with activation of Igκ. Since at the κE3' enhancer E47 recruitment is dependent on IRF4, inhibition of IRF-4 expression leads to a reduction of Igκ germ line transcription (Nagulapalli & Atchison, 1998) and E12/E47 alone, or IRF4 expression alone is not sufficient to promote Igκ sterile transcription (transcription in the absence of protein synthesis) and subsequent recombination. While E2A proteins play a crucial role in Igκ locus rearrangement, recombination of the IgH locus does not necessarily depend on their expression. In fact, forced expression of the EBF1 transcription factors in E2A-deficient mice is sufficient to induce DH to JH and DJH to VH recombination (Seet et al, 2004), more likely because the other two E proteins (HEB and E2-2) can compensate for loss of E2A.

E47 is involved in lymphocyte differentiation, lineage commitment, proliferation, and survival as well as immunoglobulin rearrangement (Bain et al, 1994). E2A knock-out mouse strains display a complete block in B-cell development and altered T-cell development (Bain et al, 1994; Zhuang et al, 1996; Zhuang et al, 1994). These

mice do not express CD19 and other B-cell-lineage-associated genes, such as components of the pre-B-cell receptor (pre-BCR composed of Vpre-B, Vλ5 and an Igμ) and its signalling complex (Igα and Igβ), the paired box protein 5 (Pax5) and RAG1/2 genes (Massari & Murre, 2000; Murre, 2005). As a consequence, no D to J or V to DJ rearrangements at the heavy chain locus are found in the foetal liver or bone marrow (Bain et al, 1994; Zhuang et al, 2004; Zhuang et al, 1994), confirming that in B cells, E2A products regulate both activation of early B cells genes and commitment to the B lineage by expression of the Igs. Ectopic/transgenic expression of either E12 or E47 rescues B cell development, suggesting that E2A proteins are functionally redundant in this process (Bain et al, 1997).

However, other studies have shown that E2A-deficient progenitor cells (obtained by E12 knock-out,) can be isolated (Ikawa et al, 2004): analysis of these cells revealed that they are indeed able to carry low levels of IgH DJ rearrangement and they also possess some pro-B characteristics such as expression of RAG-1, EBF, Pax5 and several BCR-associated proteins. Since E47 production was affected by E12 knock-out, it is thought that the low level of E47 allows the observed rearrangement.

Similarly to E2A, EBF null mice fail to undergo to VH-DJH rearrangement (Lin & Grosschedl, 1995). EBF is also expressed from the HSC stage and works in concert with E2A to regulate expression of Rag1 and Rag2 genes (O'Riordan & Grosschedl, 1999). Commitment to the B-lymphoid lineage depends on the transcription factor Pax5 (Enver, 1999; Nutt et al, 1999). Pax5 can be either an activator or repressor depending on its interacting proteins. It can repress lineage-inappropriate genes and simultaneously activate B-cell-specific genes (Busslinger & Urbanek, 1995). It is expressed from the pro-B to the mature B cell stage and $Pax5^{-/-}$ mice have B cell development arrested at the early pro-B cell stage (Adams et al, 1992; Urbanek et al, 1994).

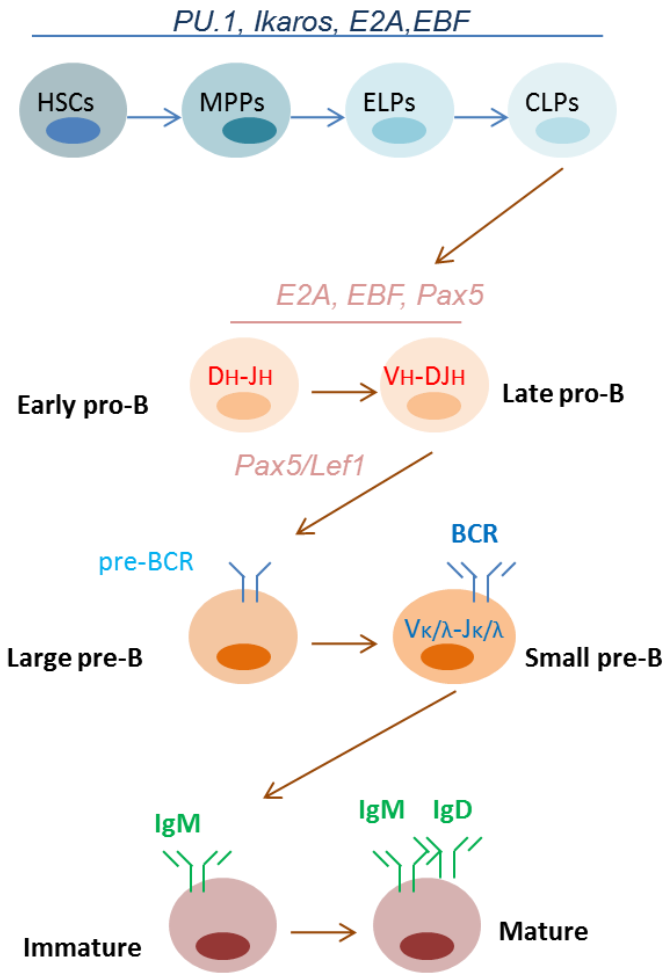


Figure 1-8 Role of transcription factors in B cell development.

Table 1-1 B cell development and sequential expression of the antigen receptor loci.

Stages in B Cell Development							
	stem cell	**early pro-B cell**	**late pro-B cell**	**large pre-B cell**	**small pre-B cell**	**immature B cell**	**mature B cell**
H chain genes	Germline	D-J joining	V-DJ joining	VDJ rearranged	VDJ rearranged	VDJ rearranged	VDJ rearranged
L chain genes	Germline	germline	germline	germline	V-J joining	VJ rearranged	VJ rearranged
Surface Ig	None	None	None	μ chain in pre-B receptor	μchain in cyto-plasm and on sur-face	Membrane IgM	Membrane IgM and IgD
RAG, TdT expression	No	Yes	Yes	No	Yes	Yes	No
Surrogate L chain expression	No	Yes	Yes	Yes	No	No	No
Ig αβ expression	No	Yes	Yes	Yes	Yes	Yes	Yes
Btk *(Bruton's Tyrosin kinsase)*	No	Little	Yes	Yes	Yes	Yes	Yes
Membrane markers	CD34	CD34 CD45 (B220) Class II	CD45R Class II CD19 CD40	CD45R Class II pre-B-R CD19 CD40	CD45R Class II pre-B-R CD19 CD40	CD45R Class II IgM CD19 CD40	CD45R Class II IgM IgD CD19 CD21 CD40

Table adapted from http://microvet.arizona.edu

1.1.4.2 T cell development

T cells originate in the bone marrow from the common lymphoid progenitor but migrate to the thymus where they undergo a strict selection before maturation known as positive and negative selection. This selection leaves only about 2-4% of the total T cells that enter the thymus from the bone marrow. Since the size of the thymus decreases over the years and rapidly from puberty, the whole T cell repertoire will be decreasing with aging. Positive selection promotes the survival of only T cells whose TCRs recognize self-MHC molecules, while negative selection eliminates those T cells that have a strong affinity for self-MHC or self-MHC in presence of self-peptides. Both classes of MHC molecules are indeed highly expressed on the surface of thymic cells, making the selection process effective.

During the first stages of proliferation, which takes place in the cortex region of the thymus, T cells are identified as double negative (DN), since developing thymocytes lack the expression of the co-receptors CD4 and CD8, identifiers of helper and cytotoxic cells, respectively. Change in cell-surface protein expression is crucial within the DN state. This is divided into four stages, DN1 to DN4, each characterised by the expression of specific cell-surface molecules: DN1 are c-Kit+CD44+CD25-, this stage also defines the germline configuration of the antigen receptor loci for T cells; in DN2, cells are CD44+CD25+ and levels of c-Kit are reduced. Expression of the recombination activating genes, RAG1 and RAG2, essential for V(D)J recombination, occurs in stage 2; DN3 are CD44-CD25+ and at this stage a rearrangement of the β chain of the TCR occurs which pairs up with a surrogate α chain to form a pre-TCR (pTα); DN4 T cells are CD44-CD25. T cells which fail the rearrangement of the β chain do not progress to the later stages of development; on the other hand, successful rearrangement leads to cell proliferation and expression of CD4 and CD8 on the cell surface. These cells are identified as double positive T cells (CD4+ CD8+) and will proliferate and express high levels of the TCR only if they recognize self MHC (positive selection), otherwise they will die. Double positive T cells will then start

expressing only one of the two co-receptors, becoming either CD4+ or CD8+ (single positive thymocytes). Negative selection also occurs in double positive thymocytes and selects for those cells that recognize MHC bound to self-antigen. These cells will be ultimately eliminated and, of those that survive, only 2% will migrate to the peripheral lymphoid organs (Figure 1-9).

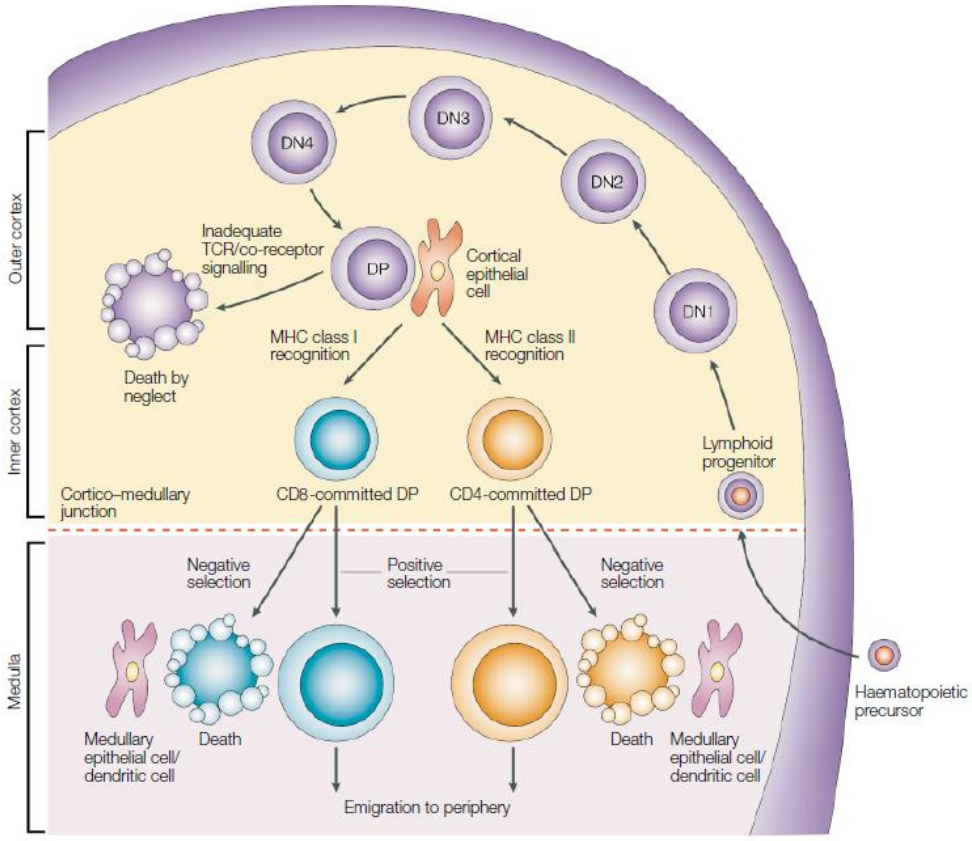


Figure 1-9 T cell development stages in the thymus.

Committed lymphoid progenitors arise in the bone marrow and migrate to the thymus. T cells at this stage are identified as double negative since they lack expression of CD4 and CD8 co-receptors. In the last stage of the DN status, T cells express a pre-TCR composed of the non-rearranging pre-Tα chain and a rearranged TCRβ-chain. Successful rearrangement of the TCR triggers transition to double positive T cells (DP) which also express a complete αβTCR (Germain, 2002).

1.1.4.3 Signalling through the BCR and TCR triggers induction of new gene expression

Both the BCRs and TCRc are associated with co-receptors that transmits information following binding of the antigens-derived peptides, in case of TCRs presented on the

surface of APCs by MHC molecules. BCRs are associated with Igα and Igβ, while TCRs with the CD3 complex and CD3ξ. These co-receptors contain ITAMs motifs whose phosphorylation by the Src protein kinase family triggers a cascade of events, which, through phophorylation of downstream proteins, triggers ultimately to the activation of transcription factors in the nucleus, which in turn will regulate gene expression.

During BCR signalling, Src phosphorylate SyK protein kinase, while in TCR signalling, Src phophorylate Lck (which is associated with CD4 and CD8 molecules) and Fin (associated with CD3ξ and CDε), with the aim to phophorylate ZAP-70, the main target in TCR activation. The signals are then propagated onwards, with ultimately activation of MAP kinase family. The different activation pathways of this family will determine specific gene expression.

1.2 Mechanism of V(D)J recombination

A key characteristic of the adaptive immune system is its ability to recognise and destroy a large variety of pathogens by generating a vast array of immunoglobulin (Ig) and T cell receptor genes. V(D)J recombination generates Ig and TCR gene diversity by assembling germline variable (V), diversity (D) and joining (J) gene segments. The TCR consists of either α and β chains or γ and δ chains, whilst immunoglobulins are composed of a heavy chain and light chain, where the latter can be of two isotypes, kappa (κ) or lambda (λ) (Figure 1-10).

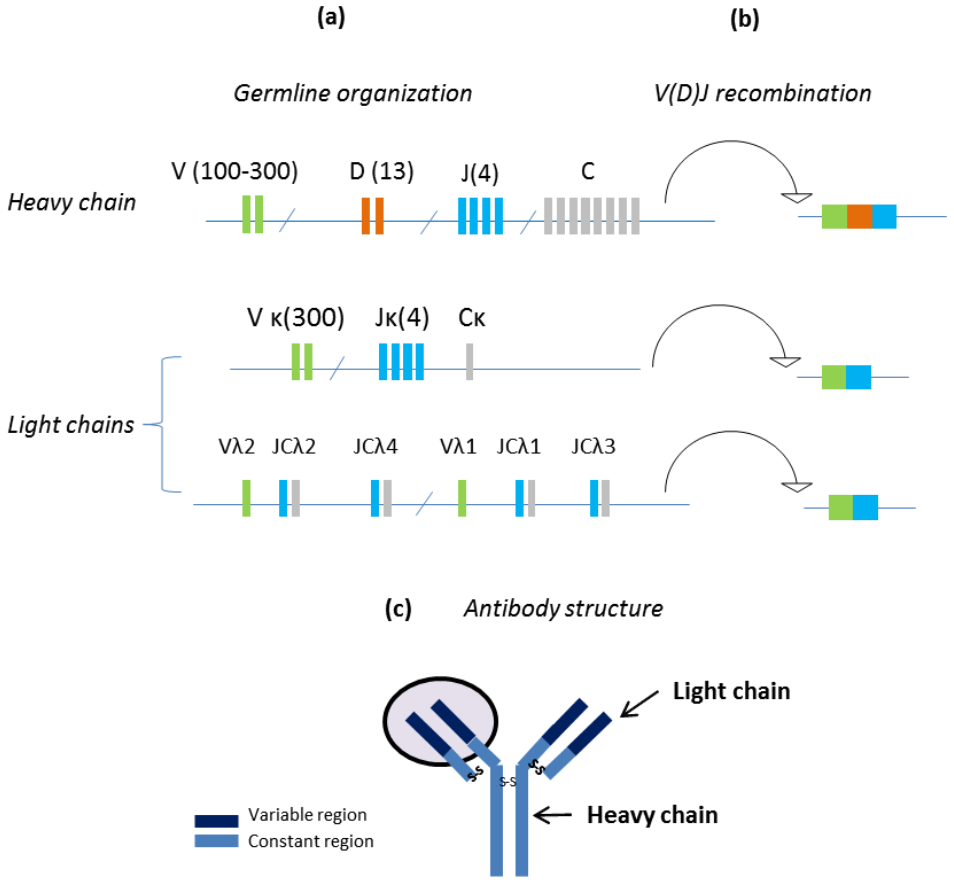


Figure 1-10 Schematic representation of V(D)J recombination in B cells.

In germline, V, D and J gene segments are separated. Following V(D)J recombination, specific gene fragments are selected to form the heavy and light chains of the immunoglobulins.

V(D)J recombination occurs in B and T cells at early stages of development: in pro-B and pre-B cells in the bone marrow and pro-T and pre-T cells in the thymus. Only two lymphoid specific proteins, RAG1 and RAG2, that are encoded by the recombination activating genes Rag1 and Rag2 initiate the process. These proteins introduce double strand DNA breaks (DSBs) within the highly conserved recombination signal sequences (RSSs), which flank each V, D and J gene segment. The RSSs consist of a conserved heptamer and nonamer separated by 12 or 23 bp of non-conserved DNA spacer. Efficient recombination occurs only between functional gene segments abutted by spacers of different length, the 12-RSS and the 23-RSS (Gillies et al, 1983).This restriction is known as the 12/23 rule (Eastman et al, 1996; Gillies et al, 1983). Following the cleavage reaction, two structures are generated (McBlane et al, 1995): a coding end with a hairpin structure and a blunt signal end (Figure 1-11). The coding end and signal end are joined by the non-homologous end-joining (NHEJ) machinery that includes the Ku70/80 protein complex, XRCC4 (X-ray cross complementing group 4), Artemis and DNA ligase IV. The Ku70/80 protein complex initiates the NHEJ pathway. First, this heterodimer binds to each of the broken ends of the double strand DNA and recruits the DNA-dependent protein kinase catalytic subunit (DNA-PKCS) to the DSB forming a protein complex where both DNA ends have been brought in close proximity. Nucleases and polymerases then process the non-compatible DNA. The hairpin structure of the coding ends is opened by the endonuclease Artemis (Bassing et al, 2002) and next the DNA ligase IV/XRCC4 complex catalyses ligation of the processed DNA ends. During the repair process, antibody diversity is greatly increased by a process called junctional diversity: whilst the signal ends are precisely joined, repair of the coding ends involves the addition or deletion of nucleotides. The cleavage of the hairpin by Artemis is random and this generates a short single strand of nucleotides. Therefore, complementary nucleotides are added to generate a palindromic sequence at the end of the coding end (Lewis, 1994). These are called P nucleotides. After this reaction, the

terminal deoxynucleotidyl transferase, (TdT) fills the gap with non-templated nucleotides (N nucleotides) (Desiderio et al, 1984).

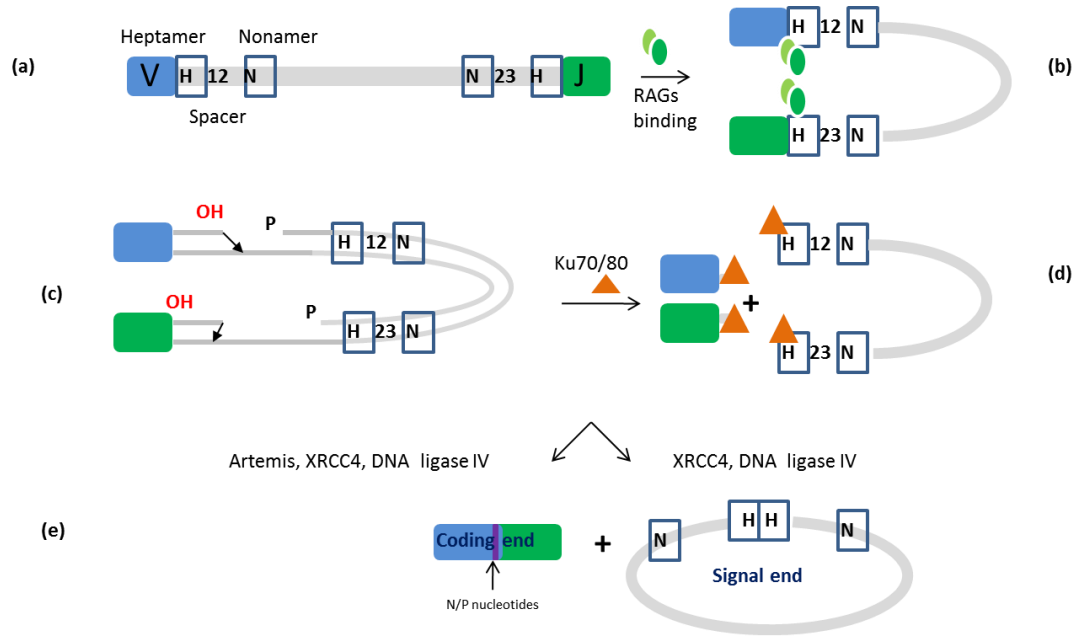


Figure 1-11 Generation of coding and signal end following RAG cutting.

A. Rag proteins recognize and bind a DNA sequence flanking the V, D and J gene segments characterised by the presence of a heptamer sequence separated from a nonamer sequence by a spacer of 12 or 23 nucleotides. B/C. Following binding, the two fragments are brought in close proximity and double strand DNA breaks are introduced. E/D. The non-homologous end-joining machinery joins the coding and signal ends. H, heptamer; N, nonamer; 12/23, spacer.

1.2.1 Structure of the B cell receptor

B cell receptors, the membrane-bound and secreted form of the immunoglobulins (Igs) are "Y" shaped proteins composed of two heavy chains of 50-55 KDa bound by a disulphide bond to two light chains of 25 KDa. There are five classes of Ig: IgG (γ), IgA (α), IgD (δ), IgM (μ) and IgE (ε) (Figure 1-12 B). Both heavy and light chains are composed of a variable region (product of the V(D)J recombination process) which contains the three complementary determining regions, CDRs, where antigens bind and a constant region (C). Structurally, Igs can be divided into two different portions when digested with papain: the Fc (fragment crystallizable) and Fab (fragment antigen binding) (Figure 1-12 A). These are connected by a hinge region that is present

in all class of immunoglobulin with the exception of IgM and IgE. Treatment of immunoglobulins with pepsin instead results in cleavage of the heavy chain after interchain disulfide bonds generating a fragment that contains the variable region to which antigens bind. This fragment is called $F(ab')_2$ because it is divalent. The $F(ab')_2$ binds antigen but without mediating the effector functions (Figure 1-12 A).

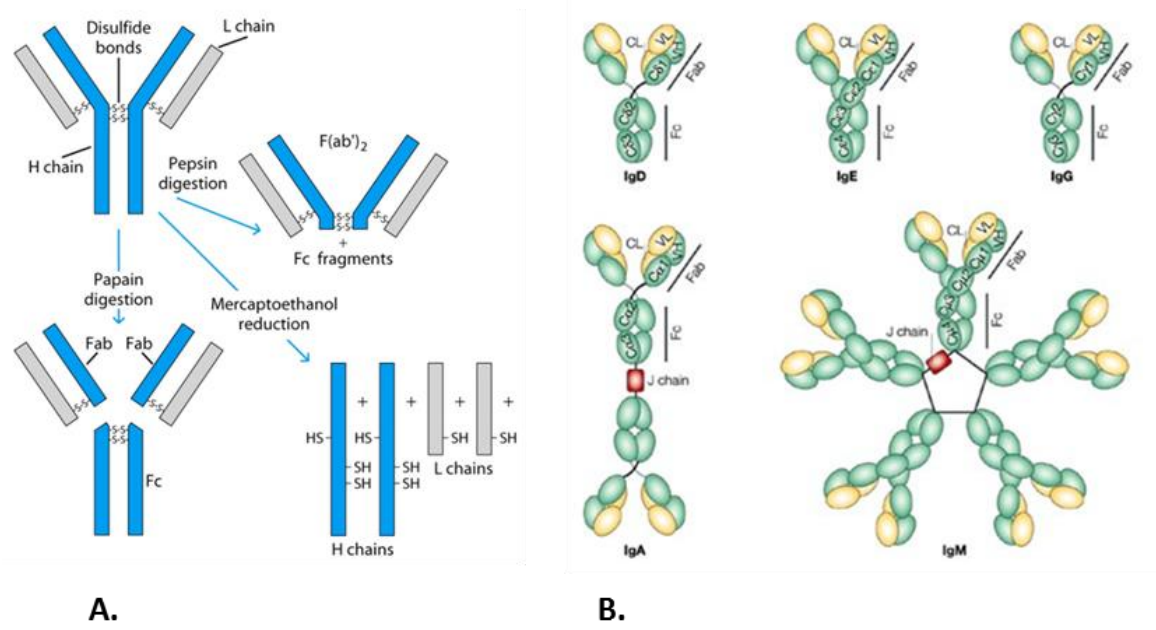


Figure 1-12 Structure of the B cell receptor and classes of Igs.

A. Immunoglobulin molecules can be divided into two portions: the Fab portion is generated by digestion with papain. This digestion generates a fragment that contains both heavy and light chains as well as the region to which the antigen will bind. The other fragment, Fc, contains the rest of the two heavy chains and the constant region. Treatment of immunoglobulins with pepsin results in cleavage of the heavy chain resulting in a fragment that contains both antigen binding sites. This fragment is known as F(ab')2 (Hewitt, 2006). B. Structural classification of the immunoglobulins. Immunoglobulins are classified into five groups: IgD, IgG, IgE, IgM and IgA. IgG exists as a monomer and is the most abundant with the longest half-life (20-22 days). IgGs are involved mainly in increasing the mechanisms of phagocytosis, neutralizing viruses and toxins. IgM represents a small population of total immunoglobulins, up to 10%. It is both secreted and membrane-bound on the surface of B cells. IgA is produced as dimers and it represents about 10-15%. IgA is mainly found in secretions. IgE is a monomer, it comprises only 0.002% of the total Igs and is involved mainly in the allergic reactions. IgD is not well-represented in the blood and has a short half-life of only a few days. Its role is still under investigation. Adapted from (Rojas & Apodaca, 2002).

1.2.2 Structure of the T cell receptor

The main function of T-cell receptors (TCRs) is to recognize peptides derived from proteolysis of extracellular antigens in endosomal-type compartments presented by MHC class II molecules on an antigen-presenting cell (APC) or peptides derived from intracellular degradation of proteins in the cytosol and presented by MHC class I molecules; in addition both classes of MHC molecules must bind to one of the two co-receptors, CD4 or CD8 (Wang & Reinherz, 2002).

TCRs consist of two disulphide-linked polypeptide chains, either a αβ heterodimer or a γδ heterodimer, with the αβ heterodimers being more common than the γδ heterodimers (the latter are known to respond mainly to bacterial and parasitic infections); nevertheless, only one type is expressed on the T cell surface, non-covalently bound to the signal-transducing CD3 subunits (CD3εγ, CD3εδ or CD3ζζ) (Kim et al, 2012; Pitcher & van Oers, 2003) (Figure 1-13). Similar to Ig molecules, both T α and β chains possess variable and constant Ig-like domains, although they are not subjected to somatic hypermutation (Chapter 1.2.3). In addition, they contain a transmembrane domain and a short cytoplasmic tail.

TCR interaction with peptide–MHC complexes triggers a cascade of signals that promotes T-cell differentiation and proliferation. This cascade is initiated by activation of the Src-family kinases which phosphorylate two tyrosine residues within the ITAMs of the TCR ζ and CD3-γ, -δ and ε subunits, rendering the ITAMs high-affinity docking sites for the recruitment of the tandem Src homology 2 (SH2) domain-containing ZAP-70 protein tyrosine kinase (PTK) (Veillette et al, 2002). ZAP-70, in turn, initiates the phosphorylation of several substrates which include: LAT (linker for activation of T cells), SLP76 (SH2-domain-containing leukocyte protein of 76 kDa) and PLC-γ-1 (phospholipase C- γ-1) (Samelson, 2002).

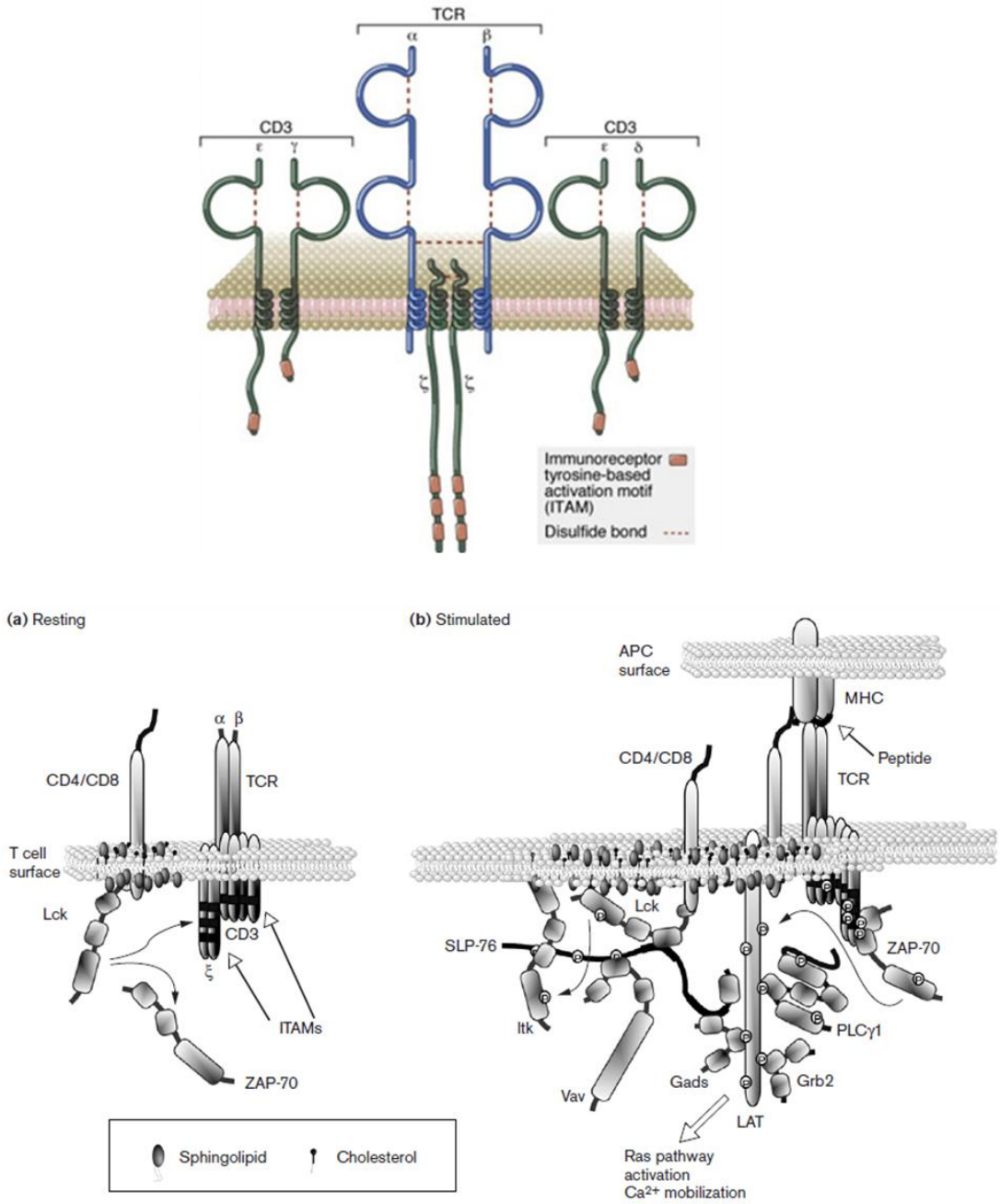


Figure 1-13 Components of the TCR activation machinery.

Top panel: components of the TCR complex (http://www.focisnet.org/). Bottom panel: proximal events in TCR signalling. (a) The elements involved in the early stages of T cell signalling in resting cells are shown; curved arrows indicate phosphorylation of the tyrosine residues within the ITAMs of the CD3 and ζ chains as described in the text. (b) Recognition of MHC–peptide, presented by APC, and illustrates molecules involved after ZAP-70 activation; a cascade of phosphorylation events triggers the initiation of downstream signalling events leading to T cells activation (Kane et al, 2000).

1.2.3 Class switch recombination and somatic hypermutation

Antigen-activated B cells can further diversify their immunoglobulin genes by two mechanisms: class-switch recombination (CSR), which occurs only in the IgH chain and somatic hypermutation (SHM) that can take place on both the heavy and light chains. Both mechanisms greatly increase the variability and affinity of the antibodies repertoire and take place in B cells residing in the germinal centre (GC), a site in the secondary lymphoid organs where B cells actively proliferate.

CSR occurs in B cells which express on their surface a functional IgM receptor, thus it take place only in B cells which have completed V(D)J recombination and which have migrated to the secondary lymphoid organs where they encounter the antigen. Via CSR, B cells are able to switch between the IgM/IgD and the other class of Igs, IgA, IgE or IgG. This recombination event, which consists of a deletion, occurs within G-rich tandem repeated DNA sequences found in the switch regions (S), upstream of each constant gene segment (Figure 1-14) (Cory et al, 1980; Kataoka et al, 1980). The reaction is stimulated by secretion of cytokines by activated T cells and requires also interaction between CD40 on the surface of B cells and CD40 ligand found on T cells. The reaction it is catalysed by the enzyme activation-induced cytidine deaminase (AID) which deaminates cytidines to uridines within the S regions and following deletion, DNA breaks are repaired by the NHEJ repair system (Manis et al, 1998; Muramatsu et al, 2000; Muramatsu et al, 1999).

SHM occurs in the rearranged V(D)J segment of both heavy and light. SHM introduces mutations at high frequency (10^{-3}- 10^{-5} pair bases per generation) and as in the CSR, it is catalysed by AID which deaminates cytidines to uridines at V region (V) genes to produce C-to-T mutations at the same frequency during transcription of the Ig loci. Mutations take place within a specific site called hypermutation domain (HYM) found within the promoter of the rearranged VDJ gene segments and encompassing the rearranged receptor locus and the intron region. Preferred sites or

"hotspots" for mutations are known to reside within the complementarity-determining regions (CDRs) of the variable segments, the portion of the V region where contacts with the antigen occur; indeed, mutation rates are higher in the V region and tend to decrease as the JC region is reached. It is now established that SHM correlates to transcription and that the intronic enhancer (Ei) and promoter (P) are required for SHM (Figure 1-15) (Betz et al, 1994; Fukita et al, 1998). At the end of the process, mismatch repair and base excision repair (BER) process uridines in an error-prone fashion, leading to significant diversification of the Ig locus.

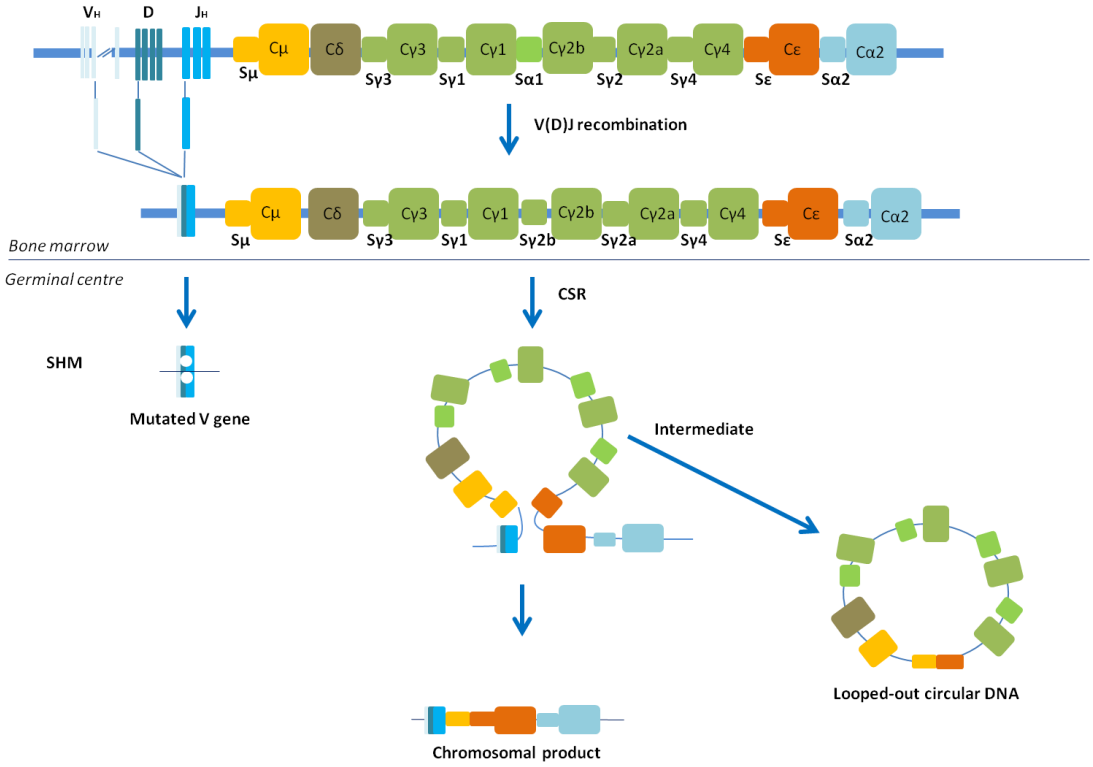


Figure 1-14 Mechanism of CSR and SHM.

CSR and SHM take place in the peripheral lymphoid tissues. SHM introduces mutations in the rearranged V exon in both the heavy and light chains. CSR occurs only within the constant region of the heavy chain. It switches the Cμ region with another C_H region, bringing that constant region close the V region. This reaction occurs in the "switch" regions located upstream of each constant region (except for Cδ). Deleted DNA is released as a circular DNA. S, switch region; C, constant region. Adapted from (Kinoshita & Honjo, 2001).

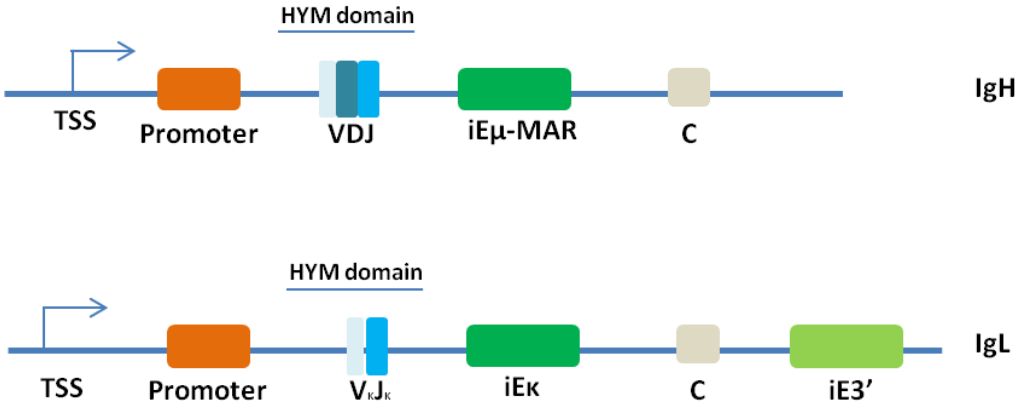


Figure 1-15 Regions in the Ig heavy chain involved in SHM.

SHM is linked to transcription and efficient loading of RNA polymerases at the immunoglobulin promoter is supported by the Ei-MAR region found downstream of the rearranged antigen receptor locus. Mutation frequency is maximal over the V(D)J coding exon and decays at the 3' end approx. 1.5kb downstream from the transcription start site. C, constant region; MAR, matrix attachment region; Ei, intronic enhancer; E3', 3' enhancer. Adapted from (Jacobs & Bross, 2001).

1.2.4 Biochemistry of V(D)J recombination

V(D)J recombination undergoes a meticulous regulation at different levels to ensure that only the right cells will initiate recombination at the right time. The process occurs first at the heavy chain, with the rearrangement of the D_H to J_H gene segments. This occurs on both alleles, while, later rearrangement of D_H J_H to V_H segments occurs on one allele alone. The first recombination event defines the pro-B cell stage, while the second characterise the late pro-B cell stage. If successful, this rearrangement produces a μ chain and the cells progresses to the pre-B cell stage, otherwise cells are lost at the earlier stage.

The μ chain forms a complex together with other two proteins to generate what is known as the pre-B cell receptor. This complex is formed by the μ chain, λ5 and VpreB ("surrogate light chain"). As described earlier (Chapter 1.2), this complex also included the heterodimers IgαIgβ, essential to transmits signals from the membrane to the cytoplasm. Successful expression of the pre-B cell receptor defines also the transition from large pre-B cells to small pre-B cells. At this stage, rearrangement of the light chain is completed and signals through various receptors, such Bruton's tyrosin kinases, alt further recombination events. A successful light chain recombi-

nation event leads to the expression of a complete immunoglobulin M (IgM) on the surface of B cells together with the IgαIgβ heterodimer. Self-reacting B cells are eliminated at this stage, while the rest will proceed to complete their maturation.
The whole process is highly regulated at three main levels: lineage and stage specificity, allelic exclusion and cell cycle regulation.

1.2.4.1 Stage and lineage specificity

Stage-specific regulation of V(D)J recombination ensures that Ig heavy chain rearranges first at the pro-B cell stage followed by Igκ or Igλ light chain at the small pre-B cells (Ehlich et al, 1993; Yancopoulos & Alt, 1986). Lineage specificity ensures that the immunoglobulin loci rearrange only in B cells, while the T cell receptor loci only in T cells. These two processes are governed by changes in the accessibility of the gene segments to recombinase proteins RAG-1 and RAG-2. RAG1 and RAG2 proteins are the only lymphoid-specific factors required for V(D)J recombination: co-expression of these two proteins triggers recombination of extra-chromosomal substrates in non-lymphoid cells (Oettinger et al, 1990; Schatz et al, 1992) and their expression is strictly limited to B and T cells. However, the RSSs at each of the rearranging loci in B and T cells are recognized by the same recombinase machinery. Therefore a key factor determining expression of the antigen receptor loci either in B or T cells is the accessibility of the recombinase to a specific rearranging locus. This is known as the accessibility hypothesis which was formulated by Yancopoulos and Alt (Yancopoulos & Alt, 1985): in the absence of recombination, the antigen receptor loci are not accessible to the recombinase machinery. Following chromatin changes, loci will be accessible to the RAG proteins in order to start rearrangement. Consistent with these findings, *in vitro* studies showed that when nuclei from RAG-deficient B and T cells at different developmental stages were incubated with exogenous RAG proteins, cleavage at the RSSs was dependent on the lineage and developmental stage of the cells from which the nuclei derived (Stanhope-Baker et al, 1996). This suggested that accessibility of the antigen receptor loci to the recombi-

nase machinery in chromatin would determine which segments undergo recombination. Later studies showed that the basic unit of chromatin structure itself inhibit binding of the RAG proteins to the DNA (Golding et al, 1999; Kwon et al, 1998; McBlane & Boyes, 2000). Different factors regulate accessibility to RAG1/2, among which histone post-translational modifications and the activity of the chromatin remodelling factors which can either promote chromatin accessibility or rendering it inaccessible by promoting higher level of organization which make chromatin inaccessible to the recombinase machinery (Inlay & Xu, 2003; Johnson et al, 2003; Maes et al, 2001; McBlane & Boyes, 2000).

1.2.4.2 Mechanism of allelic exclusion

Allelic exclusion ensures that each lymphocyte expresses a single receptor encoded by only one of its Ig heavy and light chain alleles. Activation of one of the two alleles could be a pre-determined regulated process (Yancopoulos & Alt, 1986) or could be via a stochastic mechanism (Coleclough, 1983).

In the first case it is thought that, following productive rearrangement, the pre-BCR signals to stop further rearrangements through a feedback inhibition mechanism (Chowdhury & Sen, 2004). This model proposes that at the pro-B cell stage, both the Ig heavy and Ig kappa light chain loci are repositioned from the periphery to the center of the nucleus. At this stage of development, only the heavy chain rearranges and IgH contracts. Successful rearrangement at the heavy locus triggers down-regulation of the RAG proteins by signals sent through the pre-BCR. At the pre-B cells stage, when the light chain recombines, the IgH de-contracts and the Ig kappa contracts. Because RAGs are re-expressed to allow recombination of the light chains, additionally, one allele from both heavy and light loci is sequestered by the repressive centromeric heterochromatin, making them inaccessible to the recombination machinery and leaving only one Ig kappa allele available to rearrange (Roldan et al, 2005) . This recruitment coincides with down-regulation of IL-7 signalling in pre-B cells which triggers deacetylation of the VH locus (Chowdhury & Sen,

2003; Chowdhury & Sen, 2004) which in turn stops further rearrangements at the heavy chain (Figure 1-16).

The stochastic model instead proposes that rearrangement of both alleles of the same gene is a rare event: functional V-to-DJ rearrangement is an event that occurs with low probabilities, therefore the probability of functional rearrangement on both alleles would be even lower (Liang et al, 2004).

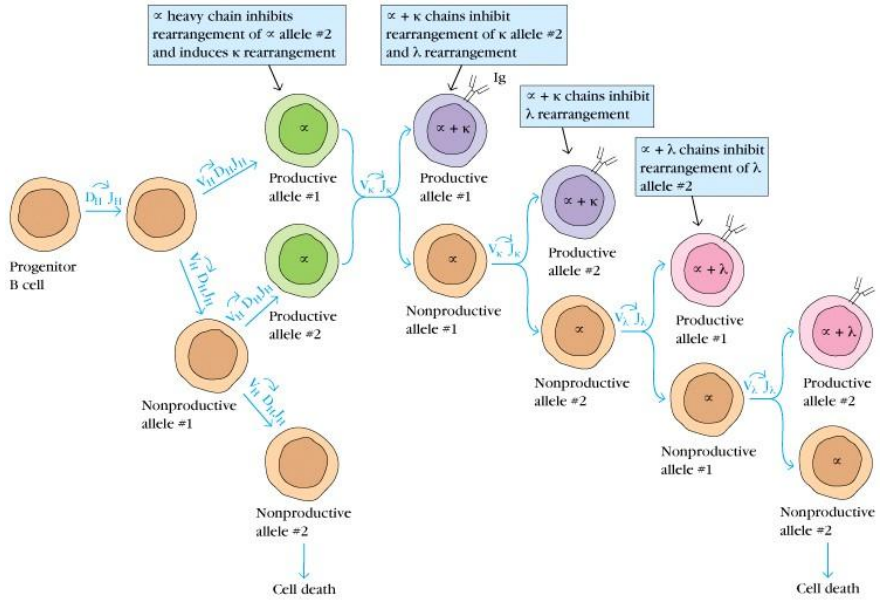


Figure 1-16 Mechanism of allelic exclusion.

Heavy chain genes rearrange first and if a productive chain gene rearrangement takes place, rearrangement of the other heavy-chain allele is prevented while simultaneously light-chain gene rearrangement is initiated. In mice the k light chain genes rearrangement normally precedes the λ genes. In humans, however, either κ or λ rearrangement might take place if a productive heavy-chain rearrangement has occurred. Generation of a complete immunoglobulin inhibits further light-chain gene rearrangement. If a non-productive rearrangement occurs for one allele, then the cell attempts rearrangement of the other allele (Yancopoulos & Alt, 1986).

1.2.4.3 Cell cycle regulation

V(D)J recombination is also regulated at the level of the cell cycle. Notably, it starts with the introduction of double strand breaks (DSBs) at the RSSs. DSBs cause a delay in the transition between the G1 to S phase of the cell cycle to allow repair of the DNA before entering the S phase. Therefore it was suggested that V(D)J recombination is restricted to a specific phase of the cell cycle, namely G0/G1 (Figure 1-17).

Because the RAG proteins are the only factors required to initiate V(D)J recombination, regulation of their synthesis or their degradation could prevent recombination in the wrong phase of the cell cycle. Earlier mutational studies on RAG2 suggested that, following accumulation in G0/G1 phase, phosphorylation of residue Thr490 targets RAG2 for degradation in G1/ early S phase via the ubiquitin-proteasome pathway (Jiang et al, 2005; Li et al, 1996; Lin & Desiderio, 1993), Therefore, it was suggested that V(D)J recombination occurs preferentially during the G0/G1 phase of the cell cycle (Lin & Desiderio, 1994).

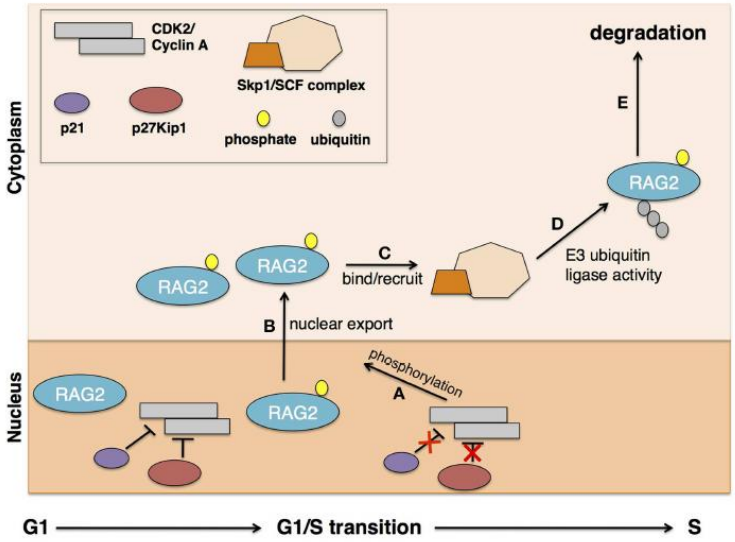


Figure 1-17 Cell cycle regulation of V(D)J recombination.

Cyclin A/CKD2 activity control RAG2 turnover during the G1/S transition phase. In the G1 phase of cell cycle, p27Kip1 and p21 bind to cyclin A/CKD2 inhibiting their action, thus rendering RAG2 stable in the nucleus. A. When cells enter the G1/S transition phase, p27Kip1 and p21 are degraded, therefore cyclin A/CKD2 are able to phosphorylate RAG2 at Thr-490 residue. B/C. Phosphorylation of RAG2 triggers its nuclear export to the cytoplasm where it can bind Skp2 to recruit the SCF complex that retains an E3-ubiquitin ligase activity. D/E. The SCF complex ubiquitinates RAG2 triggering its degradation by the 26S proteasome (Chao et al, 2014).

1.2.5 Organization of the immunoglobulin light chains genes in mice

Immunoglobulin light chain genes can be either κ or λ chain. The ratio between κ and λ differs between species. In humans it is 6:4 while in mice it is 10:1.

The murine κ locus has clustered variable (V), joining (J) and only a single constant (Cκ) gene segment, while the λ locus is bipartite as a result of evolutionary gene du-

plication. It has four JC genes and two V genes (Bernard et al, 1978; Blomberg et al, 1981; Miller & Tang, 2009): Vλ1 rearranges with Jλ1-Cλ1 or Jλ3-Cλ3, while Vλ2 rearranges with Jλ2-Cλ2. The Jλ4-Cλ4 is thought to be a non-functional pseudogene because of an inactive RNA splice site and is thus not fully functional during rearrangement (Carson & Wu, 1989). Overall V1 to J1 accounts for 60% of the total recombination, while V2 to J2 for 30% and V3 to J3 for 10% of the recombination process.

In mice, immunoglobulin light chain expression is under the control of enhancers. The kappa locus has two enhancers: the intronic enhancer, κEi, and the 3' enhancer, κE3'. Mutational analysis showed that both enhancers are essential for locus activation at the pre-B cell stage. The lambda locus has two B cell specific enhancers, Eλ2-4 and Eλ3-1 which are 90% homologous (Eisenbeis et al, 1993; Hagman et al, 1990) (Figure 1-18 B).

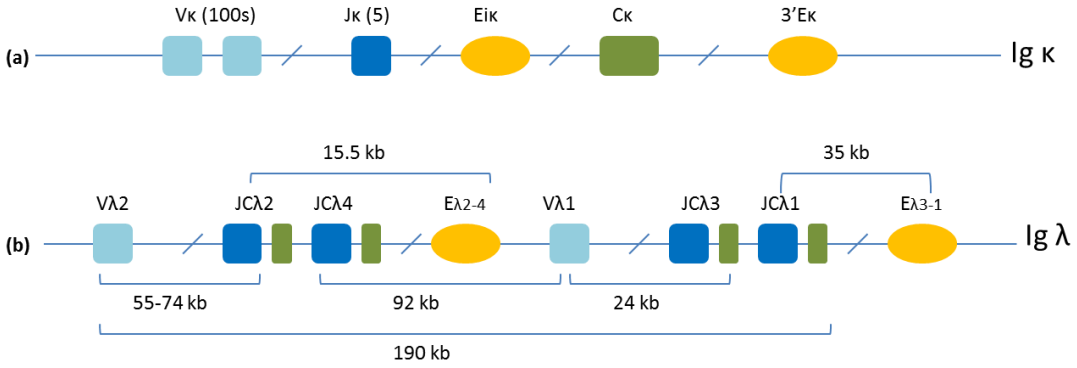


Figure 1-18 Organization of the immunoglobulin light chains in mouse.

A.The immunoglobulin light chain can be either κ or λ. The murine κ locus consists of several V gene segments and five J gene segments. Two enhancers regulate the locus: an intronic enhancer, Eiκ, and one at the 3' end, 3'Eκ. (b) The murine λ light chain results from an evolutionary gene duplication event. It has two V gene segments and four JC gene segments. Two enhancers regulate gene expression: Eλ2-4 found 15.5 kb downstream of JCλ2 and Eλ3-1, 35 kb downstream of JCλ1. At the lambda locus, Vλ1 rearranges mostly with JCλ1 while Vλ2 with JCλ2. This is also consistent with the large distance between the gene segments: there is about 190 kb between Vλ2 and JCλ3, while 55-75 kb separates Vλ2 from JCλ2 and 24 kb separates Vλ1 from JCλ1.

1.2.5.1 Immunoglobulin lambda light chain expression in mice

The lambda locus is a duplicated locus that is mainly regulated by two enhancers, Eλ2-4 and Eλ3-1. Initial analyses identified four hypersensitive sites downstream of JCλ1 (Eccles et al, 1990): HS1, HS2, HS3 and HS4 where HS1 is equivalent to Eλ3-1. Insertion of the 3' half of the lambda locus into a bacterial artificial chromosome (BAC) and subsequent deletion of the hypersensitive sites showed that only one of them, HS1, is required for functional recombination (Haque et al, 2013). Different transcription factors bind to these enhancers to regulate VJ recombination. Earlier footprinting studies of Eλ2-4 enhancer identified four protected regions: λA, λB, Eλ1 and Eλ2.

Mutational analysis showed that two of them, Eλ1 and Eλ2, confer optimal activity, while the other two, λA and λB, are essential for enhancer activity (Eisenbeis et al, 1993; Rudin & Storb, 1992). Subsequent EMSA studies identified these factors. Of these, Eλ1 and Eλ2 were E-box proteins and λA was identified as a member of the myocyte enhancer factor 2 (Mef2) protein family. Mef2 protein members belong to the MADS (MCM1, Agamous and Deficiens and serum-response factor, SRF), a family of transcription factors that also regulates myogenesis through interactions with other transcription factors (Shore & Sharrocks, 1995). Vertebrates have four mef2 genes: mef2a, mef2b, mef2c, and mef2d located on different chromosomes (Pollock & Treisman, 1991). EMSA studies using extracts from the cell line J558L showed that more than one member binds to the λA motif (Satyaraj & Storb, 1998). The two complexes were identified as JA1 and JA2 and are likely to correspond to Mef2C and and Mef2D, respectively. Mef2C is mainly expressed in brain, spleen and skeletal muscle while Mef2D is ubiquitously expressed (Martin et al, 1994; Martin et al, 1993).

Whilst E boxes and Mef proteins are ubiquitously expressed, factors binding to λB were identified as B cell-specific factors (Eisenbeis et al, 1993). This is a protein complex formed by PU.1, a member of the Ets family (Petterson & Schaffner, 1987), and PIP (PU.1-interacting partner, also known as IRF4) (Eisenbeis et al, 1995). Be-

cause the Eλ2-4 and Eλ3-1 enhancers share 90% homology and the binding site for this complex is conserved between the two enhancers, it can be suggested that the PU.1/IRF4 protein complex also binds to the Eλ3-1 enhancer (Rudin & Storb, 1992). IRF4 is expressed in B and T cells (Brass et al, 1996; Eisenbeis et al, 1995) and its DNA binding domain is highly homologous to the DNA binding domain of the interferon regulatory factors (IRF) family. IRF4 is directly recruited by PU.1 to form a protein complex. However, it is not able to bind the λB motif on its own (Eisenbeis et al, 1993; Pongubala et al, 1992): IRF4 is recruited only when interacting with PU.1 and only when PU.1 is phosphorylated at serine 148 of the PEST domain (Pongubala et al, 1993). The C-terminal regulatory domain of IRF4 is required for ternary complex formation with PU.1 and, when IRF4 is not bound to PU.1, the C-terminus functions as an auto-inhibitory domain preventing IRF4 binding to DNA (Brass et al, 1996).

1.3 Eukaryotic transcription

The exact number of human genes is still unknown, however it is has been established that approx. 20-25,000 genes (less than 2%) are expressed as proteins, with the rest of the genome containing non-coding RNA genes, regulatory sequences, introns, and non-coding DNA (International Human Genome Sequencing, 2004; Pertea & Salzberg, 2010) . Proteins-coding genes are regulated at several levels including transcription initiation and elongation, mRNA processing, transport, translation and mRNA stability.

Two different families of cis-acting DNA elements regulate proteins-coding genes by controlling gene expression: a promoter together with the proximal promoter elements, and distal regulatory element such as enhancers and locus control regions (LCRs), silencers and insulators. These DNA sequences contain binding siteS for *trans*-acting DNA binding transcription factors whose function is either to enhance or repress transcription.

1.3.1 Chromatin organization

Eukaryotic genomes are packaged by histone proteins into chromatin, which is composed of condensed fibres to fit 2m of DNA into the nucleus, which has a diameter of just 5-10µm. While DNA must be highly packed, this packaging must also be able to allow different processes like replication, repair and transcription to occur and to allow the binding of different factors required for these processes.

When isolated chromatin is examined by electron microscopy it appears like a "beads on string" of about 10 nm diameter, where the "beads" represent the single nucleosomes and the "string" the DNA linker (Finch et al, 1975). Further micrococcal nuclease digestion and cross-linking studies have shown that the nucleosome is the basic repeating unit of chromatin and it is formed by an octamer of histones around which is wrapped approx. 147bp of DNA. A DNA linker joins individual nucleosomes (Luger et al, 1997). The histone proteins involved are: H1, H2A, H2B, H3

and H4 (Felsenfeld, 1978; Kornberg, 1977) (Figure 1-19). The octamer is formed by a tetramer of two copies each of H3-H4 as well as two H2A/H2B dimers. Histone H1 is not part of the main core, it is less conserved compared to the other histones and it associates with the nucleosome outside of the main core. It binds to both DNA and the core histone octamer and its function is believed to be in the formation of 30 nm fibres (Robinson & Rhodes, 2006).

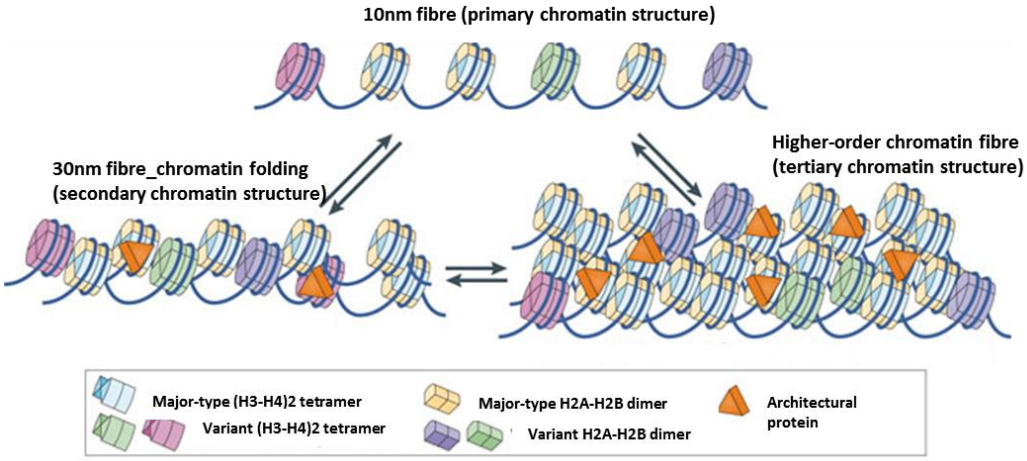


Figure 1-19 Levels of chromatin organization.

DNA is first wrapped around nucleosomes to form the first level of structural compaction, then further association of nucleosomes gives the 30nm fibre and higher levels of organization allow the generation of the final structure found in chromosomes. Adapted from (Luger et al, 2012).

The 30 nm fibre represents a higher level of organization and is formed by the association of nucleosomes. How nucleosomes are organized to form the latter level of compaction is poorly understood and two main models have been proposed. The one start helix (also called the solenoid model) where nucleosomes form a helical structure with 6 nucleosomes per turn, and the two start model, where nucleosomes follow a zig-zag pathway and the linker between nucleosomes is straight. (Dorigo et al, 2004).

Higher levels of compaction are still being studied. Earlier studies showed that by removing histones from metaphase chromosomes, DNA is organized into loop domains (Marsden & Laemmli, 1979; Paulson & Laemmli, 1977). These loops are at-

tached to a scaffold of non-histone chromosomal proteins called scaffold attachment regions (SAR). Similar studies were also conducted on histone-depleted interphase chromosomes and the DNA was organized in loop domains anchored to the nuclear matrix. This scaffold is referred to as the matrix attachment region (MAR) (Heng et al, 2004).

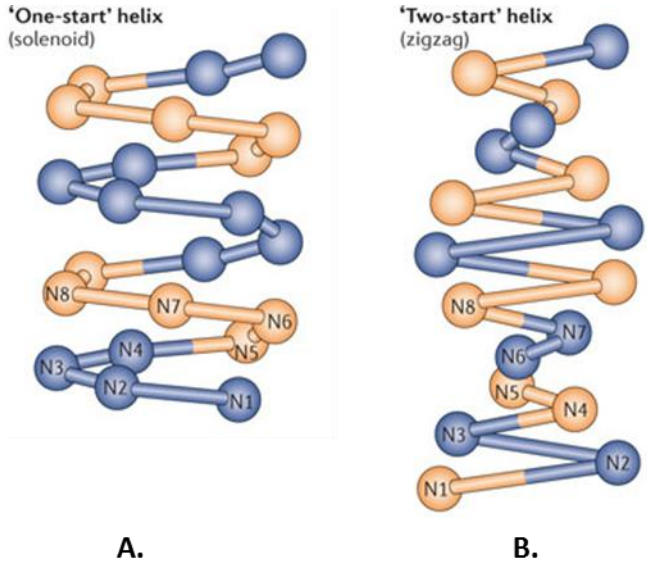


Figure 1-20 The solenoid and the zig-zag models.

A.The solenoid model is characterized by interactions between consecutive nucleosomes: the 30 nm chromatin fibre a nucleosome in the fibre interacts with its fifth and sixth neighbour nucleosomes. C. In zigzag model the interactions occur between alternate nucleosomes (N1 to N8): the chromatin fibre is a two-start helix in which nucleosomes are arranged in a zigzag manner such that a nucleosome in the fibre binds to the second neighbour nucleosome. Adapted from (Li & Reinberg, 2011).

1.3.1.1 Activation of chromatin: chromatin modifications

During transcription, the binding of different proteins to specific DNA sequences is essential to ensure correct initiation. This process is however influenced by the presence of nucleosomes which create a repressive environment due to the inaccessibility of the DNA (nucleosome positioning), as well as some post translational modifications, to which histones are subjected, which can create a repressive chromatin environment (Berger, 2007; de la Cruz et al, 2005; Strahl & Allis, 2000). Post-translational modifications occur in some amino acids in the N-terminal tails of his-

tones and involve acetylation of lysine residues, methylation of lysine and arginine residues, phosphorylation, ubiquitination or ADP-ribosylation (Berger, 2002). Specific enzymes catalyse these reactions that can either create a repressive chromatin or an open chromatin structure. Moreover, once these tails are modified, histones are able to recruit further proteins that can affect chromatin structure (Jenuwein & Allis, 2001; Strahl & Allis, 2000). The combination of the different modifications is known as the "histone code" (Figure 1-21).

Histone acetylation occurs at lysine residues and correlates with gene activation. It can affect chromatin structure at two main levels. It can either trigger 30nm fibre opening by decreasing interactions between histone tails and adjacent nucleosomes or it can help octamer eviction from promoters (Reinke & Horz, 2003). Indeed, acetylation of lysine residue 16 in H4 (H4K16ac) promotes an open chromatin status since it inhibits formation of the 30nm fibre (Robinson et al, 2008; Shogren-Knaak et al, 2006).

In contrast, histone methylation is associated with either promotion or repression of transcription, depending on the lysine modified. Trimethylation of lysine residue 4 in H3 (H3K4me3) is associated with gene expression, but trimethylation of lysine residue 9 and/or lysine residue 27 in H3 (H3K9me3 and H3K27me3) is associated with repression of gene expression. Moreover, the same tail can be subjected to one or more modifications on different residues, thus gaining new functions (Berger, 2007; Strahl & Allis, 2000). In addition to post-translational modifications, chromatin remodelling complexes are required and recruited at the specific site to relax chromatin structure. These complexes use the energy derived from ATP hydrolysis to modify the interaction between histones and DNA, by displacing the histone octamer or altering the histone core to unwrap the DNA (Clapier & Cairns, 2009; Vignali et al, 2000).

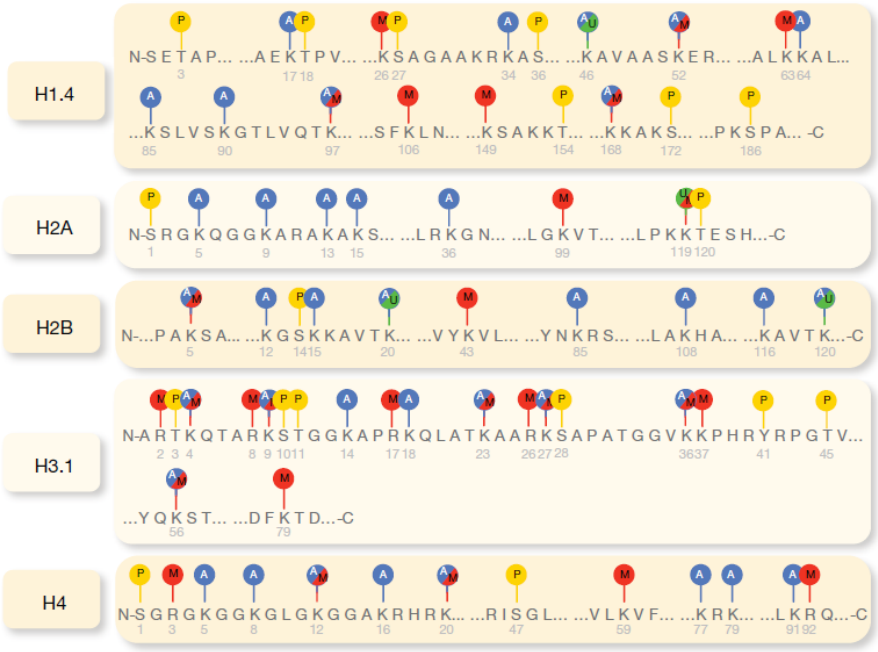


Figure 1-21 Modifications occurring on the histones.

Histone tails are subject to post-translational modification, namely acetylation, methylation, phosphorylation. Acetylation (blue), methylation (red), phosphorylation (yellow) and ubiquitination (green). The number in grey under each amino acid represents its position in the sequence (Portela & Esteller, 2010).

1.3.2 Promoters and the initiation of transcription

Promoters are DNA regulatory elements located near and typically upstream of the genes they regulate. Eukaryotic promoters are composed of two parts: the core promoter (or basal promoter), which includes the TATA box (AT-rich region) for binding of RNA Pol II and the transcriptional start sites (TSSs), and the upstream promoter element (one or more) which regulate transcription (Figure 1-22). A typical promoter also contains short sequence elements which represent binding sites for different transcription factors; these are typically found 200-300bp upstream of the start site. The importance of the TATA box was demonstrated in mutagenesis studies where mutations usually reduced or abolished the activity of the promoters, although it was also shown that not all promoters necessarily have a TATA box

(Breathnach & Chambon, 1981; Farrell). These include housekeeping genes (expressed constitutively) and developmentally-regulated genes (Lodish, 2000).

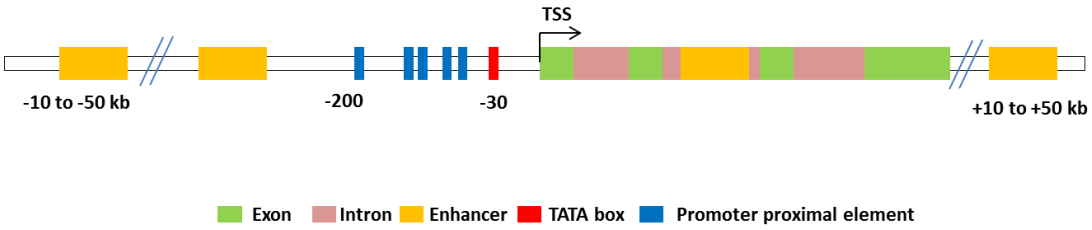


Figure 1-22 Structure of eukaryotic promoters.

To allow binding of RNA Pol II to the promoter, several transcription factors have to find their place within the promoter. These are known as general or basal transcription factors (GTFs): TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH and are known to bind in an ordered, sequential manner. TFIID is the largest of the transcription factors as it consists of the TATA box – binding protein (TBP) and eleven TBP-associated factors (TAFs) (Lodish, 2000).

The first factor to bind to the promoter is TFIID (in complex with TAFs and TBP) followed by TFIIB and a newly formed complex between Pol II and TFIIF (Figure 1-23). One major role played by TFIIF is also to help maintain chromatin at the promoter in an uncondensed state due to the histone acetylase activity. Then, TFIIE binds allowing entrance of the next transcription factor, TFIIH (Alberts et al, 2002). This complex is known as the pre-initiation complex or PIC.

During eukaryotic transcription, a major role is also played by the C terminal domain (CTD) tail on the largest subunit of the RNA Pol II which consists of multiple, conserved repeats of a heptapeptide with the consensus sequence Tyr-Ser-Pro-Thr-Ser-Pro-Ser and serves as a binding site for several transcription factors. Which one of these factors will bind depends on the phosphorylation state of the CTD during gene transcription.

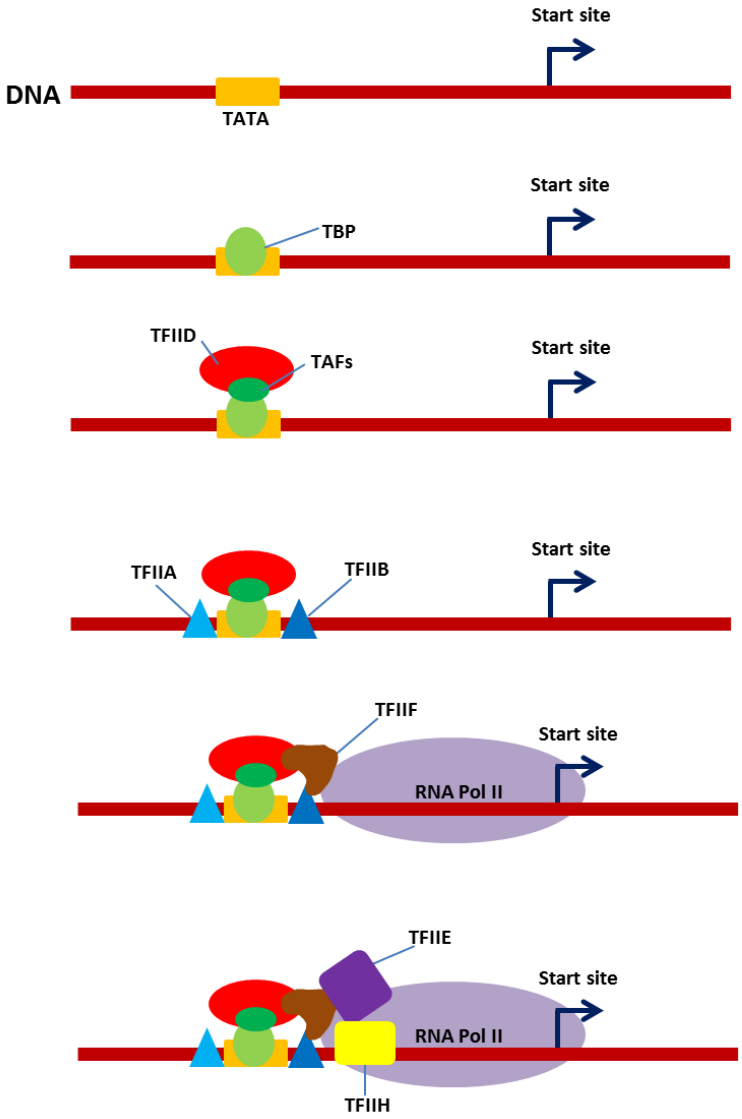


Figure 1-23 Assembly of the pre-initiation complex in eukaryotic cells.

The diagram depicts the ordered assembly of the general transcription factors during eukaryotic transcription.

Phosphorylation occurs mainly on Serine 2 and Serine 5 of the repeats, with the latter being catalysed by TFIIH that has a protein kinase activity (Figure 1-24). *In vivo*, it was shown that RNA Pol II requires the association of a mediator and a large protein complex that includes TFIIB, TFIIH and TFIIF; this association occurs at the CTD of unphosphorylated RNA Pol II.

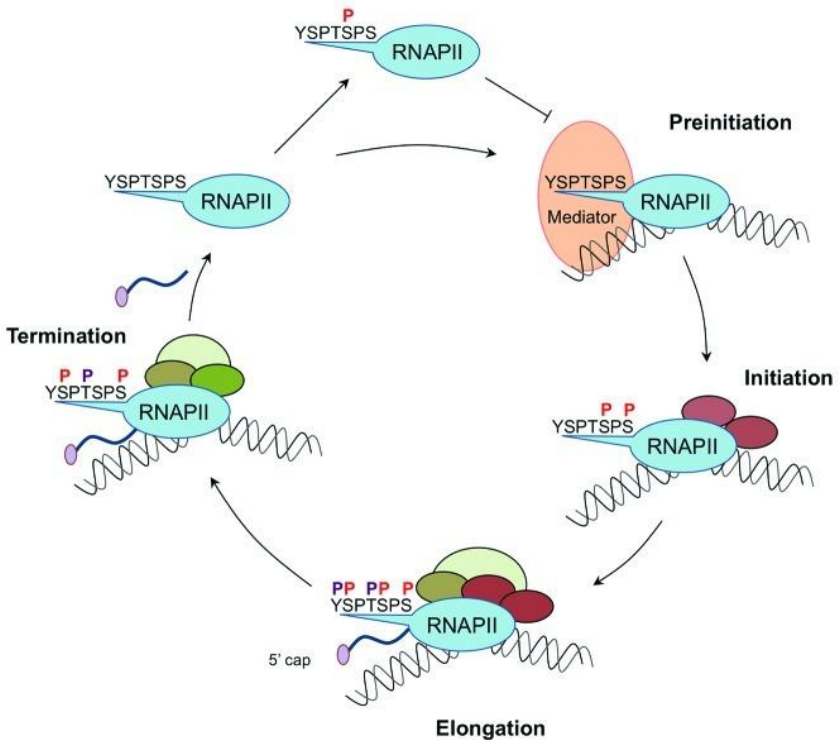


Figure 1-24 CTD phosphorylation patterns during the RNAPolII transcription cycle.

RNA Polymerase II (RNAP II) with hypophosphorylated CTD forms together with general transcription factors (GTFs) and the mediator a pre-initiation complex (PIC). During initiation, elongation and termination various CTD phosphorylation marks (P) are successively introduced and erased. These CTD signatures determine the interaction capacity of RNAP II machinery for different transcription and RNA processing factors (schematically presented by colored ovals). (Heidemann & Eick, 2012).

1.3.3 Enhancers

Enhancers are relatively short DNA sequences characterised for the first time over 30 years ago while analysing the β-globin locus using transient reporter gene assays in cultured cell lines (Banerji et al, 1981). The activity associated with these sequences was first described in viruses and subsequently also in metazoans (Banerji et al, 1983; Gillies et al, 1983). These regulatory sequences contain binding sites for transcription factors and they can activate their target genes over long distances without regard to orientation or location (Bulger & Groudine, 2011). It was also reported that in some cases enhancers can regulate expression of genes located on different chromosomes (Williams et al, 2010). Different mechanisms were de-

scribed to explain how such long regulatory elements control gene expression; however the predominant model envisages enhancers looping out the intervening DNA, to touch and activate the target promoters (Bulger & Groudine, 1999).

While localization of promoter sequences is relatively easy to perform (by sequencing of the 5'-end of its mRNA), identification of DNA sequences corresponding to enhancers is more challenging. However, a common feature of most enhancers sequences is the exposure of the local chromatin structure to DNase I digestion *in vitro*. Such regulatory sequences contain sites for transcription factors that can both exclude nucleosomes and bend DNA and are therefore marked by such nuclease hypersensitivity (Elgin, 1988; Gross & Garrard, 1988; Wu, 1980). DNase I hypersensitivity is therefore considered to be a hallmark of enhancer function in a given DNA sequence. With the advent of microarrays and high-throughput sequencing, further studies have also allowed the genome-wide mapping of nuclease hypersensitivity (Crawford et al, 2006). Another feature of enhancers is that their non-coding sequences are highly conserved (implying their regulatory function) compared to less conserved sequences which instead tend to accumulate mutations and are therefore unlikely to be regulatory elements. A typical example is that of the β-globin locus where expression of all the genes is under the control of a potent enhancer, known as LCR located 20-30 kb upstream of the gene cluster: indeed, the sequences of the LCR are well conserved among mammalian genomes (Visel et al, 2009). In addition, common to sequences identified as enhancers are some "chromatin signature" recently discovered by genome-wide studies. These include H3K4Me1 and the absence of H3K4Me3 (Birney et al, 2007). Interestingly, studies utilizing embryonic stem (ES) cells and multiple primary cell types have suggested that acetylation of histone H3K27 in combination with H3K4Me1 is correlated with active enhancers, whereas H3K4Me1 alone appeared to mark either inactive or ''poised'' enhancers (Creyghton et al, 2010; Rada-Iglesias et al, 2011). These data, together with genome-wide mapping of nuclease hypersensitive sites (HSs), have suggested that the human genome might harbour as many as $1x10^6$ enhancers (Heintzman et al, 2009).

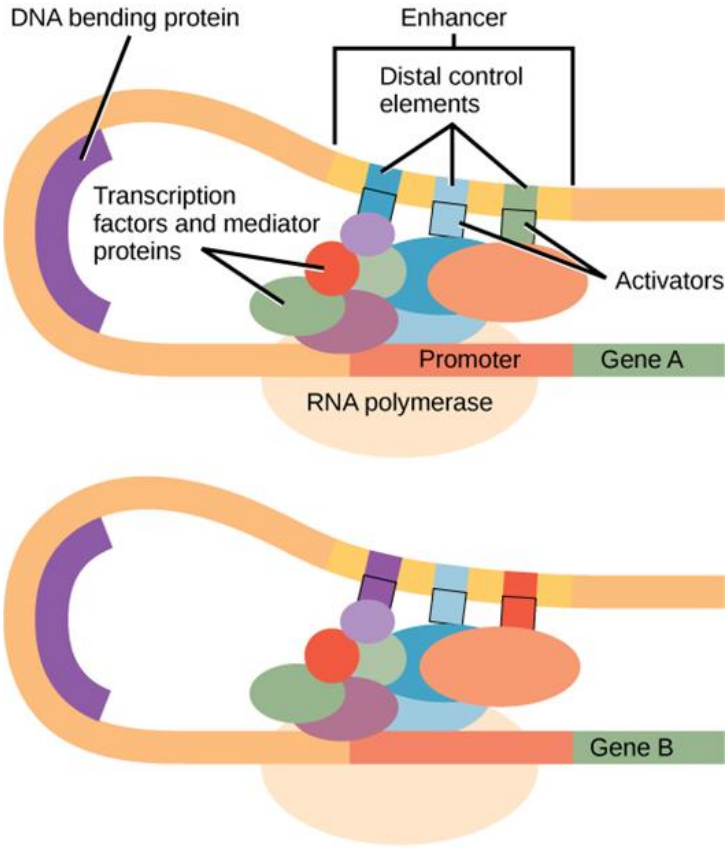


Figure 1-25 Long range chromatin interactions

Distant enhancers cause DNA to loop and touch the cognate promoters following intervention of several transcription factors (https://www.boundless.com).

1.3.4 Locus control regions (LCRs)

LCRs are *cis*-acting regulatory DNA sequences that enable high-level and tissue-specific gene expression.

It is known that transcribed genes are more sensitive to nuclease digestion compared to non-transcribed genes. This is referred to as general sensitivity to differentiate it from the DNase I hypersensitivity found at active promoters and enhancers (Forrester et al, 1986). One of the well-characterized LCRs is the human β-globin, which was the first to be identified (Grosveld et al, 1987). In the β-globin cluster these hypersensitive sites at the LCRs are required for the correct activation of the

genes and they bind a number of different transcription factors (Grosveld et al, 1993; Orkin, 1995).

1.3.4.1 The β-globin LCR

The β-globin LCR is located 6 to 22 kb 5' to the first embryonic globin gene (ε) in the locus which consists of five developmentally regulated genes (ε, Gγ, Aγ,δ, β) organized and expressed at different times during development. The ε gene is the first to be expressed, it is then silenced and expression of the foetal γ genes takes place. At birth, expression of γ genes starts to decrease and δ and β genes are activated (Dean, 2004) (Figure 1-26). The LCR has four erythroid, tissue-specific DNase I–hypersensitive sites, HS1 to HS4 (starting from the 5'-end), while in non-erythroid cells, there are three additional HSs, HS5 to HS7 (Dean, 2006; Li & Stamatoyannopoulos, 1994). Moreover, it was shown that enhancer activity of the β-globin LCR depends on HS2, 3, and 4, but not on HS1 or 5 (Fraser & Grosveld, 1998). Experiments in which the cores of HS2 and HS3 were mutated or deleted respectively, showed a decreased level of expression of the ε-globin as well as the β-globin gene suggesting that these sites are essential for full expression of the locus, while deletion of HS2-5 resulted in complete loss of gene expression (Bungert et al, 1999; Reik et al, 1998). However the contribution of individual HSs for full gene activation is still poorly understood and debate about the function of a single HS is still open. In the absence of the LCR, transcription of the human β-globin gene represents less than 1% of the endogenous murine β-globin mRNA in transgenic mice (Magram et al, 1985; Townes et al, 1985), suggesting that the LCR has strong enhancer activity.

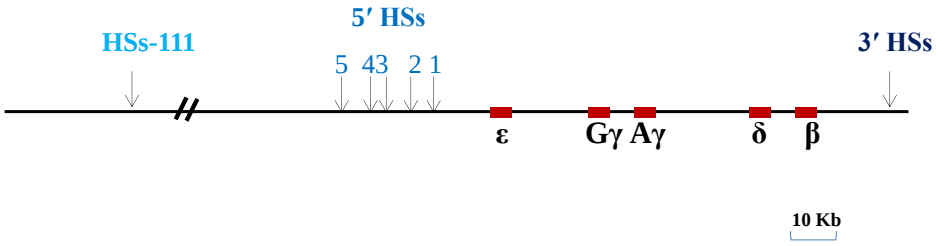


Figure 1-26 Organization of the human β-globin locus.

The human β-globin locus has five genes arranged in order of their developmental expression, 5'-ε-Gγ-Aγ-δ-β-3'. Expression is under the control of the LCR located 5' between 6 and 22kb upstream the distal ε gene. The LCR is composed of five hypersensitive sites (HS1 to HS5) required for high level expression of the genes. The first gene to be expressed is the ε-globin gene in the yolk sac; this is then silenced and the Gγ- andAγ-globin genes are expressed in the foetal liver. After birth the γ-globin genes are silenced and the β-globin gene and δ-globin gene are activated in the bone marrow. The HSs 5'HS1 through 5'HS7 are located −6, −11, −15, −18, −22, −28, and −35 kb relative to the ε-globin gene, respectively.

1.3.5 Insulators can block enhancer-promoter interaction

Specific elements are able to block the interaction between enhancers and promoters. An insulator is a DNA sequence that is able to interfere with the interaction between enhancer and promoter, thus preventing inappropriate gene expression. The best characterized insulators are Gypsy in Drosophila and the chicken β-globin 5'-HS4 (cHS4), both associated with specific proteins which mediate their activity: Su(Hw) and CTCF (CCCTC-binding proten), respectively (Brasset & Vaury, 2005). Two mechanisms describe their action. An insulator can be an enhancer-blocking insulator found between the enhancer and promoter preventing any association or it can act as a barrier insulator by preventing the spread of repressive heterochromatin. In the first mechanism it is thought that the insulators might form chromatin loops to physically separate the enhancer and promoter. The exact mechanism by which this happens is poorly understood; however, it was proposed that an insulator could block specific signals from the enhancer from reaching the promoter (Gaszner & Felsenfeld, 2006; West et al, 2002; Zhao & Dean, 2004). The action of these elements relies on the models proposed to explain long-range enhancer-promoter interactions. Historically three different models were proposed to explain this interaction: the linking, tracking and looping models (Figure 1-27).

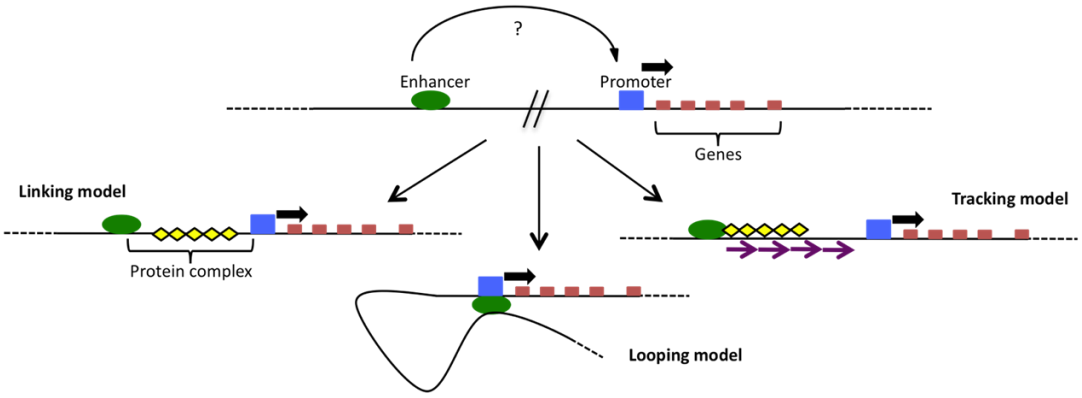


Figure 1-27 Models to describe long-range chromatin interactions.

In the linking model, interaction between distant elements is mediated by a protein complex that forms a link between the distant elements and the promoter. The tracking model, also called the scanning model, is based on the ability of enhancer and protein complexes bound to it to scan the DNA to find a promoter. If the right promoter is found, then a more stable complex is formed. In the looping model, generation of chromatin loops brings the distant enhancer close to its cognate promoter.

1.3.6 Linking, tracking and scanning models

Long-range interactions between distant regulatory elements and promoters have been documented in different systems. The following models explain how these interactions may occur.

The linking model is based on the hypothesis that interaction between distant elements is mediated by a protein complex that forms a link between the distant elements and the promoter (Bulger & Groudine, 1999; Dorsett, 1999; Morcillo et al, 1996). The idea for this hypothesis comes from the activity of some insulators. Because these elements block the signal between enhancer and promoter, it was proposed that what an insulator is blocking is the signal transmitted by specific proteins dispersed along the chromatin, between the distant regulatory element and the promoter (Morcillo et al, 1997). Mutation studies conducted on Drosophila Cut gene, which is regulated by the Chip protein, suggest that Chip could be essential for enhancer-promoter communication. The same is believed to happen for its vertebrate homologues interacting with the LIM (**L**in11, **I**sl-1 & **M**ec-3) domains of nuclear proteins. The LIM domains are zinc-finger motifs involved in different protein-

protein interactions and proteins with these motifs also possess a homeodomain DNA binding motif (HD); therefore it was proposed that the Chip protein (as well as the vertebrate homologues) could promote interactions between the LIM proteins located at the enhancer and HD proteins found along the chromatin between enhancer and promoter generating in this way a sort of chain of proteins that connects the two distant elements (Bulger & Groudine, 1999; Morcillo et al, 1996). However to date, there is no strong experimental evidence to support this model.

The tracking or scanning model is based on the ability of enhancer and protein complexes bound to the DNA and their ability to to scan the DNA to find a promoter. If the right promoter is found, then a more stable complex is formed (Hatzis & Talianidis, 2002; Travers, 1999). Like the linking model, the evidence for this idea comes from the activity of some insulators. Since these elements are able to block signals between enhancer and promoter, the idea is that these elements interrupt the communication by blocking the complex from reaching the promoter during the scanning of the DNA (Figure 1-27). Good evidence for this model comes from the protein CTCF in vertebrates with its binding site for the insulator cHS4 in the chicken β-globin genes. Using minichromosome constructions carrying the ε globin gene and HS2, it was found that this insulator was able to interrupt the communication between HS2 and promoter by altering the pattern of chromatin modification at the promoter (Zhao & Dean, 2004). Positioning of cHS4 between HS2 and the ε globin gene blocked the spreading of histone acetylation that was attributed to the activity of CBP and p300, two proteins required for haematopoietic differentiation (in the case of CBP, acetylation decreased when the insulator was between, outside or inside the enhancer- gene unit). These results were confirmed by creating a cHS4 locus mutated in the sequence that corresponds to the CTCF binding. In this construct, transcription was restored, confirming the role played by this insulator and the two proteins involved (Zhao & Dean, 2004).

The looping model is based on the theory that the generation of loops is the main mechanism by which a distant enhancer reaches its own promoter (Figure 1-27).

This model is supported by results obtained using the Chromosome Conformation Capture (3C) technique (Dekker et al, 2002). The 3C is a powerful technique to determine whether interactions occur *in vivo*; this methodology is based on the ability of formaldehyde-induced cross-links to detect links between regions of chromatin that are interacting with each other (Figure 1-28). Following digestion with a restriction enzyme, the ends of the cross-linked fragments are ligated. Following reversal of the cross-links, PCR is used to analyse ligation frequencies of the fragments (Dekker, 2006). Using the 3C technique to analyse interactions in the human β globin cluster, several loops were generated, collectively identified as "active chromatin hub", a complex network of different loops coming in and out. However a limiting factor for the 3C technique is that it does not explain how the two distant elements "find each other" since it is supposed that the two elements find each other by collision (Bulger & Groudine, 1999; Engel & Tanimoto, 2000). Thus, it is possible that both looping and scanning models are valid.

Two other techniques have been developed to analyse chromosome interactions based on the same theory of 3C, the 4C (Circular chromosome conformation capture or 3C-on chip) and the 5C (Chromosome conformation capture carbon copy). Both techniques detect interaction between genomic regions without prior knowledge of the fragments involved. The 4C technique includes a circularization step to analyse chromatin interaction of all genomic regions with a given gene, while the 5C copies and amplifies parts of the 3C library to obtain many copies of a unique ligation product (Dostie & Dekker, 2007; Ohlsson & Gondor, 2007; Simonis et al, 2007; Zhao et al, 2006). Thus these techniques are useful to detect previously unpredicted interactions.

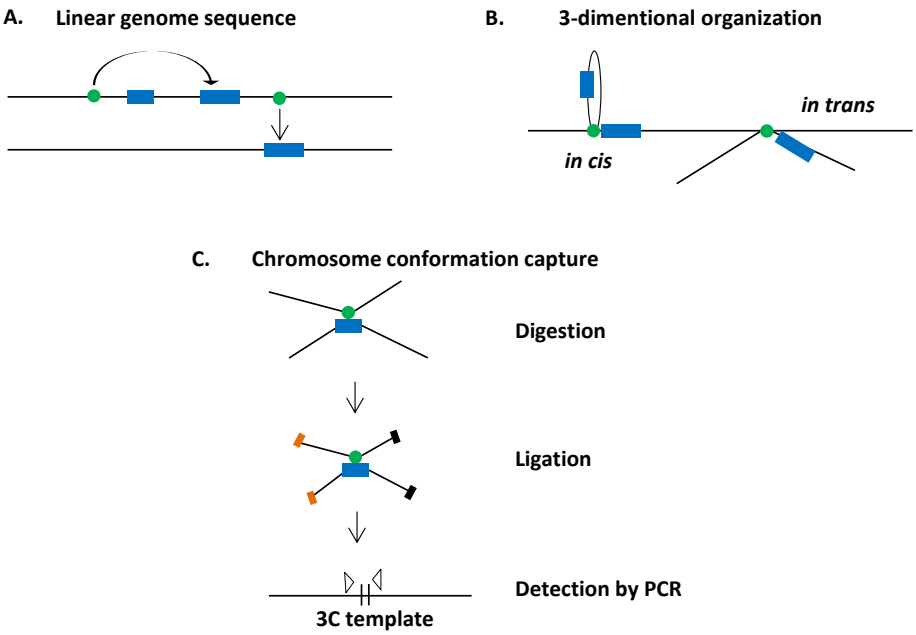


Figure 1-28 Chromosome conformation capture technique.

a) Genes (blue rectangles) and regulatory elements (red circles) are linearly organized along chromosomes (top), but as a result of specific interactions between elements (indicated by arrows, both *in cis* and *in trans*) a complex three-dimensional network is formed inside the cell (bottom). (b) Schematic representation of the 3C assay. Chromatin is cross-linked, digested with a restriction enzyme and then ligated. Specific ligation products can be detected by PCR. Adapted from (Dekker, 2006).

1.3.7 Role of transcription factors in mediating long range chromatin interactions

Although different studies showed that enhancer and promoters interact, they do not show how the specificity of this interaction is achieved. Different studies were conducted on the globin genes and evidence was found in favour of specific transcription factors in mediating long range interactions between regulatory elements and promoters. An example is given by GATA-1, a transcription factor required for erythroid cell development and which activity depends on FOG-1 (friend of GATA-1). A mouse cell line was generated that expressed GATA-1 fused to the oestrogen receptor ligand-binding domain (GATA-1-ER). This hybrid protein, in the presence of either oestrogen or tamoxifen (the results were the same) was able to bind to DNA allowing careful control of GATA-1 binding to the LCR. 3C analysis revealed that in

the absence of the ligand, loop formation between the genes and the LCR was decreased. However, subsequent introduction of the hormone allowed recruitment of GATA-1 to the hypersensitive sites (HS2 and HS3) of the LCR and interaction of the LCR with the promoter by looping, confirming that it is only GATA-1 activity (with the co-factor FOG-1) that promotes this close interaction. Moreover, it was observed that loop formation coincided with RNA Pol II recruitment to the promoter site, suggesting that GATA-1 is also responsible for the recruitment.

The involvement of transcription factors was also investigated in the α-globin locus. In this study Vernimmen et al. used cells at different stages of erythropoiesis to investigate the order in which GTFs, RNApol II, mediator and elongation factors were recruited at the HS regulatory elements (and consequently to the promoter site) and the role played by some transcription factors in regulating this process. The hypothesis is that these genes are kept in a "poised" state (starting from the multipotent haematopoietic cells stage) until full activation, which occurs in mature erythroblasts. During this time, specific transcription factors are recruited by the HSs to start transcription (Anguita et al, 2004; Vernimmen et al, 2007).

The transcription factors involved include GATA-1, GATA-2 and Sp/X-KLF factors. They are bound to the HSs sites at the pro-erythroblast stage by forming a complex, which was then able to recruit the pre-initiation complex. Indeed, 3C analysis confirmed the presence of chromatin looping which was connecting to the distant regulatory elements to the promoter of the α globin genes and also that the PIC complex was recruited at the HS site by looping at the promoter site, but only in the final stages of cell development (primary erythroblasts). Therefore the role of these transcription factors, which are highly specific, is also to ensure that the PIC complex is recruited at the promoter site at the right time of cell development. In the absence of these transcription factors bound at the HSs, PIC is not recruited at these sites, since only the interaction between HSs and these transcription factors facilitate the recruitment. In addition, chromatin modifications are also associated with recruitment of these transcription factors. For example GATA-1 recruits histone ace-

tyltransferase (HAT) whose activity is related to the generation of open chromatin, which in turn facilitate binding of other different proteins (Letting et al, 2003).

More evidence for the role of enhancers in mediating long range interactions comes from studies conducted on the protein p300, a transcriptional co-activator as well as an acetyltransferase also required for correct haematopoietic development (Zhao & Dean, 2004). In this study, ChIP analysis and sequencing (ChIP-seq) of the genome distribution of p300 associated with the enhancer, showed that the presence of this protein (referred as "peaks") in some mouse embryonic forebrain, midbrain and limb was an evidence for enhancer activity in these tissues (Visel et al, 2009); in contrast, the absence of p300 was found to be correlated with a lack of enhancer activity. Therefore p300 binds to the enhancer to drive the interaction with the promoter. Since this protein is widely distributed and associated with enhancer activity, this could explain how enhancers deliver the modification to promoters.

1.3.8 bHLH proteins in Ig genes rearrangement

Strong evidence exists that both rearrangement at a particular locus and Ig class switch recombination are preceded by transcriptional activity of the respective germ-line gene segments (Engel et al, 1999; Lennon & Perry, 1990; Schlissel & Baltimore, 1989). It has been established that *cis*-regulatory elements physically contact the promoter to drive transcription of the target genes, but how the specificity of this interaction is achieved is poorly understood. Different factors may influence this interaction: chromatin modifications (Chapter 1.3.1.1), the interactions between specific transcription factors and boundary elements that interfere with the interaction between enhancers and promoters.

In the next paragraph, the role of transcription factors in mediating long-range interactions in the immunoglobulin chain genes is examined.

At the immunoglobulin lambda light chain locus, enhancer-promoter interactions are driven by B cell specific transcription factors. Of these, the helix loop helix protein, E47 is of particular interest.

The helix-loop-helix transcription factor family are key regulators of different developmental processes such as neurogenesis, myogenesis, and haematopoiesis (Bain et al, 1994; Zhuang et al, 1994). These proteins recognize specific, conserved DNA sequences termed E-box sites. E-box sites were first discovered at the promoter and enhancer regions of both B and T lineage specific genes by *in vivo* methylation protection assays that identified a number of sites which were specifically protected. The protected DNA sequence consisted of the core hexanucleotide sequence, CANNTG (in which N denotes any nucleotide), and was named E box binding protein (Ephrussi et al, 1985). This sequence was subsequently identified at the enhancers of the immunoglobulin heavy chain locus (IgH: μE1, μE2, μE3, μE4, and μE5.), at the immunoglobulin kappa light chain locus (Igκ: κE1, κE2, and κE3), the T cell receptor α and β loci, the promoters of the mb-1, λ5, and pre-Tα genes (Ahmad, 1995; Ernst & Smale, 1995; Ho et al, 1989; Kee & Murre, 2001; Nelsen et al, 1993; Reizis & Leder, 1999; Sigvardsson et al, 1997). Subsequent studies have also identified E-box elements in promoter and enhancer elements that control muscle-, neuron-, and pancreas-specific gene expression (Buskin & Hauschka, 1989; Chaudhary et al, 1997).

1.3.8.1 E-box proteins

Based on structural characteristics, dimerization properties and tissue distribution, E-box binding proteins have been classified into 6 classes (I to VI) (Murre et al, 1994). E12, E47, HEB, E2-2 belong to the Class I HLH proteins. These transcription factors regulate gene expression in different tissues and can form either homo- or heterodimers (Murre et al, 1989b).

Class II HLH proteins are tissue-restricted and include: NeuroD/BETA2, MyoD, myogenin and the achaete-scute complex. Compared to class I proteins, these transcription factors are usually unable to form homodimers, preferring instead heterodimerization with other E proteins. Class III HLH proteins include the Myc family of transcription factors, TFE3, SREBP-1, and the microphthalmia-associated transcription factor, Mi. Transcription factors belonging to this group are characterized by the presence of a leucine zipper motif (LZ) adjacent to the HLH motif. Class IV HLH proteins include Mad, Max, and Mxi that are capable of dimerizing with the Myc proteins. Class V HLH proteins includes proteins that lack the basic region which are called Id proteins, inhibitor of differentiation. Id proteins heterodimerize with E proteins from class I and class II, however, because they lack the DNA binding domain, they regulate E proteins activity by sequestering them as non-functional heterodimers (Benezra et al, 1990). Class VI HLH proteins are characterized by the presence of a proline in the DNA-binding domain and include the Drosophila proteins Hairy and Enhancer of split (Davis & Turner, 2001; Klambt et al, 1989). Class VII is characterised by the presence of the bHLH-PAS domain; members of this class include the aromatic hydrocarbon receptor (AHR), the AHR nuclear-translocator (Arnt), hypoxia-inducible factor 1α, and the *Drosophila* Single-minded and Period proteins (Crews, 1998).

E proteins are broadly expressed and many of them from the different classes can bind DNA as homo- or heterodimers. In B cells, the predominant E box binding complex is comprised of E47 homodimers: E47 is one of the two products of the E2A gene and it plays a key role during B cell development. In thymocytes E47/HEB heterodimers are found predominately (Bain et al, 1993; Murre et al, 1991; Sawada & Littman, 1993; Shen & Kadesch, 1995). In non-B cells, E47 forms heterodimers with other members of the E protein family proteins (class II) such as the muscle-specific bHLH protein MyoD (Lassar et al, 1991; Wendt et al, 1998) and NeuroD1, a key regulator of neuronal development (Lassar et al, 1991; Longo et al, 2008; Mehmood et al, 2009; Murre et al, 1989b; Naya et al, 1995).

1.3.8.2 Structural properties of the E-box binding protein E47

The transcription factor E2A is encoded by the Tcfe2a gene and belongs to the class I family of HLH transcription factors (Murre, 2005). E2A gives rise to E47 and E12 by differential pre-mRNA splicing of the exon encoding both the DNA binding and dimerization domains (Murre et al, 1989a). E12 and E47 share 80% sequence identity within the bHLH domains and they only differ in usage of the exon encoding the bHLH domain (basic helix-loop-helix), however E12 has a lower affinity for DNA due to an inhibitory region nearby the bHLH domain (Sun & Baltimore, 1991). This could explain the major use of the E47 homodimer in B cells, rather than the E12.

The helix-loop-helix motif of the E47 is highly conserved and consists of two amphipathic α-helices separated by a loop structure (Murre et al, 1989b). This domain is common to all E-box binding proteins and is responsible for dimerization. The E47 dimer forms a parallel, four-helix bundle (two dimeric alpha-helical structures) which allows the basic region to contact the major groove of the DNA. In addition to the basic region, residues in the loop and helix 2 also make contact with DNA (Ellenberger et al, 1994).

Structurally, alongside the DNA-binding and the dimerization domains, E12 and E47 also share two conserved transcriptional activation domains known as the activation domain 1 (AD1) and loop-helix (LH) or AD2 domains (Aronheim et al, 1993; Massari et al, 1996; Quong et al, 1993). Both activation domains are helical structures; AD1 is found at the N terminus within the first 99 amino acids (Massari et al, 1996), while AD2 is downstream of the N terminus within residues 349-406 (Figure 1-29). Further analysis of these two transcription factors region identified additional highly conserved regions named E proteins homology domains (EHD1, EHD2 and EHD3). Mutational studies within both the AD and EHD domains have shown that transcriptional activity of E47/E12 decreases remarkably. In particular, point mutations in either AD1 or AD2 cause about 50% reduction of transcriptional activity. EHD1 and EHD2 mutation led to 30-40% decrease, while deletion of the EHD3 domain trigger only slight decreases in transcriptional activity (Massari et al, 1996).

Further studies confirmed that this domain is responsible for E47 regulation by the Notch family via MAPK phosphorylation of the residues within the EHD3 domain (Massari et al, 1996).

Figure 1-29 Domain structure of the E12/47 proteins.

AD1, activation domain 1; AD2, activation domain 2; bHLH, basic helix-loop-helix, EHD, E protein homology protein.

1.3.8.3 Regulation of E47 activity

E47 is widely expressed and together with other E-proteins plays key roles in promoting commitment to and differentiation of the B- and T-cell lineages from the early stages to the generation of the immunoglobulins and TCRs. To this end, its activity is highly regulated. E47 is regulated by association with Id proteins and, among others, by p300 (adenoviral E1A-associated protein, 300 kDa) while its turnover is controlled by Notch/MAPK signalling pathway. P300 is a histone acetyltransferase, which acetylates the ε-amino group of lysine residues in different proteins by regulating their functions, and it works in association with CREB (cAMP-response-element binding protein)-binding protein (CBP). It is now known that E47 contains an acetylation site for p300/CBP within the AD1 domain and that this interaction is responsible for E2A transcription activity and localization in the nuclear compartment (Bradney et al, 2003).

1.3.8.3.1 Regulation of E47 activity by Id proteins

E47 and Id proteins are widely expressed throughout the haematopoietic system and they have crucial roles in cell fate decisions, differentiation and proliferation of lymphocytes as well as in other tissues (de Pooter & Kee, 2010; Engel & Murre, 2001; Greenbaum & Zhuang, 2002; Massari & Murre, 2000). In B lymphocytes E

protein activity is strictly regulated by association with Id proteins (Rivera & Murre, 2001). Id proteins are class V HLH proteins, however they lack the basic DNA binding domain, therefore when they dimerize with E47 and other E proteins, they prevent DNA binding and thus inhibit E protein function, acting as dominant-negative HLH proteins (Benezra et al, 1990; Ghil et al, 2002). There are four known Id proteins (Id1, Id2, Id3 and Id4) and a recent study has showed that E47 interacts with all 4 classes (Teachenor et al, 2012a). In the absence of E-box binding proteins, Id proteins are distributed in the cytoplasm/perinuclear region. Although Id proteins are very small proteins (all four Id proteins range from 13 to 18 kDa) and this implies that they can enter and exit the nucleus by passive diffusion, numerous studies have suggested that other pathways may regulate their subcellular localization. Upon binding with the E proteins, Id proteins relocate to the nuclear compartment (Deed et al, 1996). Because Id proteins lack a nuclear localisation signal, E47 and other E proteins, which are nuclear proteins and therefore possess a classical nuclear localization signal, function as nuclear chaperones (Deed et al, 1996). Moreover, association of Id/E47 protein complex regulates the half-life of both proteins: Id proteins have a half-life of about 1h, while E47 turnover is slower, about 24 hours. Following formation of the dimer, the half-life of Id proteins is increased to 3 hours, while E47 is decreased to 16 hours (Deed et al, 1996).

1.3.9 Importance of E47- IRF4 interaction

The PU.1/IRF4 complex binds the Eλ3-1 enhancer and an interesting question is what mediates the interaction with the promoter of sterile transcription upon chromatin looping. While transcription factor binding to the enhancer site has been widely characterized (such as the E2A protein complex), the transcription factors that bind to the promoter have not been determined. Engel et al. (1999) mapped the promoters of sterile transcription for JCλ1, JCλ2 and JCλ3 (Engel et al, 2001). Initiation of transcription was observed at 293 bp, 116 bp and 79 bp upstream of the Jλ gene segment, respectively. Previous studies showed that IRF4 interacts with E47

at the κE3' enhancer and that this interaction enhanced E2A (E12 and E47) DNA binding and resulted in a strong E2A/IRF4 transcriptional synergy (Nagulapalli & Atchison, 1998; Nagulapalli et al, 2002). However, in this study, the two transcription factors lay immediately adjacent to each other. Therefore it is not known if these factors interact when bound to separate sites on the chromosome. However, by performing an internet database search for E2A binding sites, it was possible to highlight E2A binding sites at the Jλ1, Jλ2 and Jλ3 promoters as well. These were 60bp upstream and 35bp downstream of the start site of sterile transcription of the Jλ1 promoter, 110bp upstream and 50bp downstream of the Jλ2 promoter and 15bp downstream of the Jλ3 promoter. Therefore, a possible E2A-IRF4 interaction was investigated in this study.

1.4 Human papillomaviruses: general considerations

Human papillomaviruses (HPVs) are small, non-enveloped, DNA viruses that belong to the *Papillomaviridae* family which display a particular tropism for the stratified squamous or cutaneous epithelium (zur Hausen, 2009). To date, at least 180 HPV genotypes have been cloned from clinical lesions and they have been numbered sequentially (Bernard et al, 2010). These have been categorized into five genera based on their genomic sequence and subdivided based on the ability to infect the skin (cutaneous) or the mucosal epithelium: α, β, γ, μ and ν (Bernard, 2013; de Villiers et al, 2004). HPVs infect the basal cells and their infectious life cycle is closely related to the differentiation of the stratified epithelium. HPVs that infect the genital mucosa are divided into groups: high risk (HR), such as HPV16, 18, 31 and 33 and low risk (LR), such as HPV6 and 11 that cause benign warts that do not progress to malignancy (Table 1-2). HPV16 and HPV18 are the major causative agents of cervical cancer, accounting for 50% and 20% of all cervical cancer cases, respectively. Cutaneous HR-HPVs that infect the skin include HPV5 and 8 that cause flat lesions of the skin. HPVs are also the major causative agents of head and neck carcinomas.

Cervical cancer is the third-most common cancer in females worldwide, with approximately 275,000 deaths each year (Ferlay et al, 2010; 2008) and the second most common cancer in women worldwide, with over 450 000 new cases of cervical cancer being diagnosed annually (WHO, 2008). In most cases, the infection is cleared by the immune system after a period of 12-24 months (Richardson et al, 2003; Steben & Duarte-Franco, 2007), however, in a few cases (at least 10%), some virus particles will persist and keep multiplying and the persistence of this infection, together with viral genome integration, can potentially lead to cancer (Munger et al, 2004; Stanley, 2008).

Transformation of normal stratified squamous or cutaneous epithelium to cancer occurs in multiple steps. In cervical cancer these are: carcinoma in situ (stage 0, strictly confined to the cervix), and stage I, II, III and IV, with the latter being carci-

noma that has extended to different organs surrounding the cervix (pelvis, uterus). The immune response is quite weak in the cervix, and in addition, HPVs use different strategies to evade immune recognition. $CD4^+$ T helper cells are found mainly during regression of cervical intraepithelial neoplasia (CIN) lesions, while $CD8^+$ cytotoxic T cells are dominant in invasive carcinoma (Stanley, 2009).

Table 1-2 HPV type and disease association.

Disease HPV	Type *
Plantar warts	**1, 2**, 4, 63
Common warts	**2, 1, 7**, 4, 26, 27, 29, 41, 57, 65, 77, 1, 3, 4, 10, 28
Flat warts	**3, 10**, 26, 27, 28, 38, 41, 49, 75, 76
Other cutaneous lesions (e.g., epidermoid cysts, laryngeal carcinoma)	6, 11, 16, 30, 33, 36, 37, 38, 41, 48, 60, 72, 73
Epidermodysplasia verruciformis	**2, 3, 10, 5, 8, 9, 12, 14, 15, 17**, 19, 20, 21, 22, 23, 24, 25, 36, 37, 38, 47, 50
Recurrent respiratory papillomatosis	**6, 11**
Focal epithelial hyperplasia of Heck	**13, 32**
Conjunctival papillomas/carcinomas	**6, 11, 16**
Condyloma acuminata (genital warts)	**6, 11**, 30, 42, 43, 45, 51, 54, 55, 70
Cervical intraepithelial neoplasia:	
Unspecified	30, 34, 39, 40, 53, 57, 59, 61, 62, 64, 66, 67, 68, 69
Low risk	**6, 11,** 16, 18, 31, 33, 35, 42, 43, 44, 45, 51, 52, 74
High risk	**16, 18,** 6, 11, 31, 34, 33, 35, 39, 42, 44, 45, 51, 52, 56, 58, 66
Cervical carcinoma	**16, 18,** 31, 45, 33, 35, 39, 51, 52, 56, 58, 66, 68, 70

***Orders indicate relative frequency; bold: most frequent associations. Adapted from(Burd, 2003).**

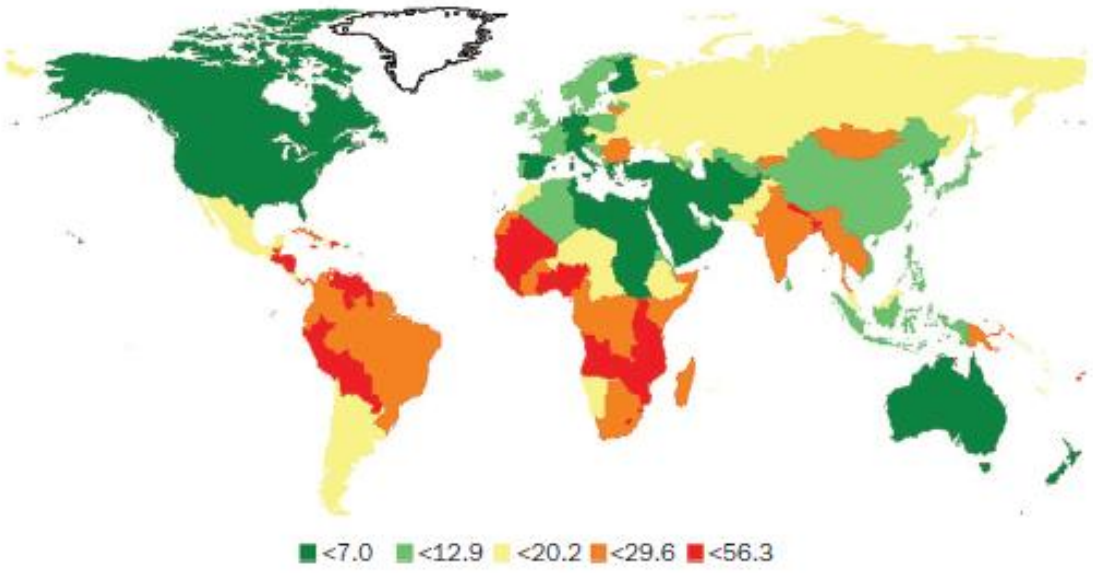


Figure 1-30 Worldwide cervix cancer incidence rate.

The diagram depicts worldwide cervical cancer incidence. Colours indicate the relative percentage distribution within the countries taken in consideration for the analysis, with red indicating the highest rate and dark green the lowest incidence rate (Cobo, 2012).

1.4.1 HPV Structure

HPV is a small, non-enveloped, double stranded DNA virus of 55 nm diameter. The encapsidated circular viral genome of only 8kb is associated with cellular histones to form a minichromosome (Danos et al, 1982; Favre, 1975).

The outer surface of the capsid is comprised of 360 molecules of the major capsid protein L1 organized into 72 capsomeres, each made up of a pentamer of L1 proteins which is important to the structure of the HPV capsid (Baker et al, 1991) (Figure 1-31). Each capsomere is composed of β-strands whose connecting loops were shown to be different in various HPVs (Ludmerer et al, 1996; Ludmerer et al, 1997; Roth et al, 2006). This is taken into consideration when formulating types of vaccines (Bishop et al, 2007; Carter et al, 2006). On the other hand, the minor capsid protein L2 varies from strain to strain suggesting that it is not involved in the determination of the capsid structure of HPV (Bordeaux et al, 2006; Florin et al, 2002).

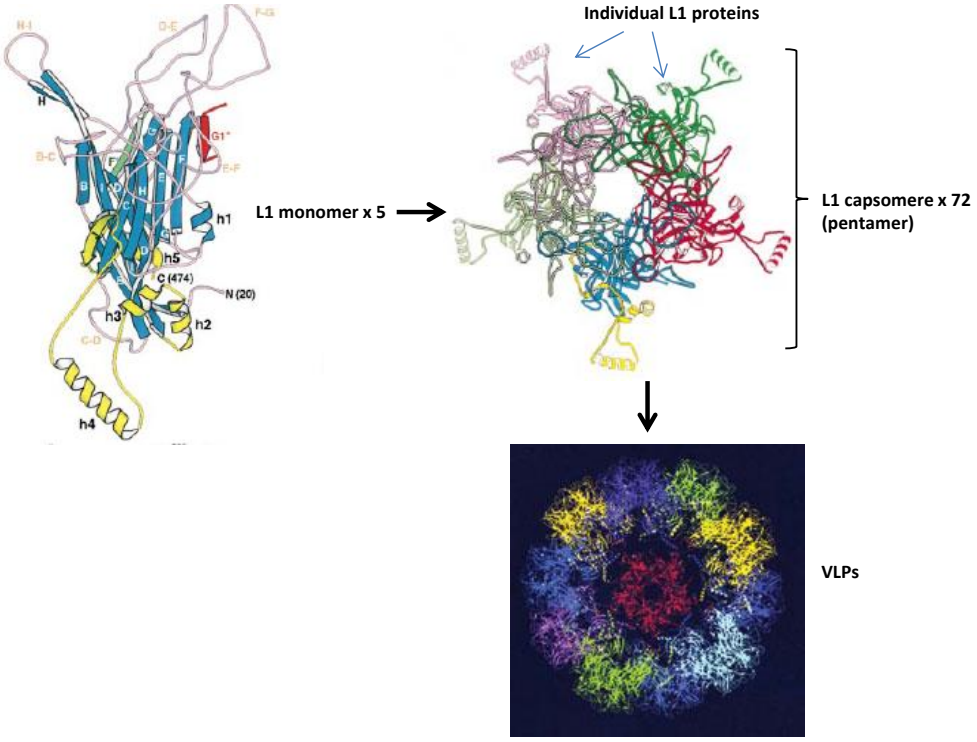


Figure 1-31 Structure of VLPs assembled from the L1 protein

The picture represents the case of HPV16. Adapted from (Chen et al, 2000).

1.4.2 HPV genes: transcription from early and late promoters

HPVs replicate and assemble in the nucleus of keratinocytes where genes are expressed in a temporal and highly regulated manner which is tightly linked to keratinocytes differentiation (Howley & Lowy, 2006). The genome is divided into three regions: a 4 kb early (E) region that encodes non-structural proteins (E1, E2, E4, E5, E6, E7) and a 3-kb late (L) region that encodes the two capsid proteins, L1 and L2 involved in virus particles assembly and capsid generation; these two regions are separated by two polyadenylation sites (pA): early and late pA sites. Both early and late ORFs are encoded on the same DNA strand. A third region encompasses a 400–1000 bp non-coding long control region (ncLCR) that contains the p97 core promoter and different *cis*-regulatory elements important for viral replication and gene expression (Figure 1-32).

Transcription is initiated from different promoters: transcription from the early promoter p97 generates mRNA containing all six early ORFs and, via alternative splicing, gives rise to the different forms of mRNA (Rosenberger et al, 2010; Stacey et al, 1995; Tan et al, 1994). As the viral life cycle progresses, the late promoter p670 is activated in terminally differentiated cells and generates a polycistronic RNA with three exons and two introns which, gives rise to L1 and L2 by RNA splicing (Schwartz, 2013). In addition to promoter switch, transcription of the later genes requires a change of polyA signal and derepression of two alternative splice sites (Figure 1-32).

The expression of early and late genes is regulated at both the level of primary transcription and RNA processing (Bernard, 2002; Graham, 2008; Thierry, 2009; Zheng & Baker, 2006). The six non-structural early viral regulatory proteins are transcribed in undifferentiated or intermediately differentiated keratinocytes while the two genes encoding the viral capsid proteins are transcribed only in keratinocytes undergoing terminal differentiation. An exception is the early protein E4 that is continuously expressed and is also synthetized in terminally differentiated cells (Peh et al, 2002).

Functions of early and late viral proteins

E1 is a protein of 73 kDa that plays a key role in viral DNA replication. This protein binds to the origin of DNA replication (*ori*) in the 3'-LCR as a double hexamer, one for each strand. The E1 protein has a helicase activity (3'-5' directionality) that unwinds the DNA in an ATP-dependent manner (Hughes & Romanos, 1993); this function is supported by the E2 protein, which binding sites are also found within the 3'-LCR. E2 proteins main function is to recruit E1 to the origin of replication site, acting as a molecular tether. The resultant ternary DNA-protein complex, E1-E2-*ori*, has the helicase activity necessary to unwind the supercoiled viral DNA. To stabilize the newly generated single-stranded DNA, E1 interacts with the replication protein A (RPA) (Loo & Melendy, 2004).

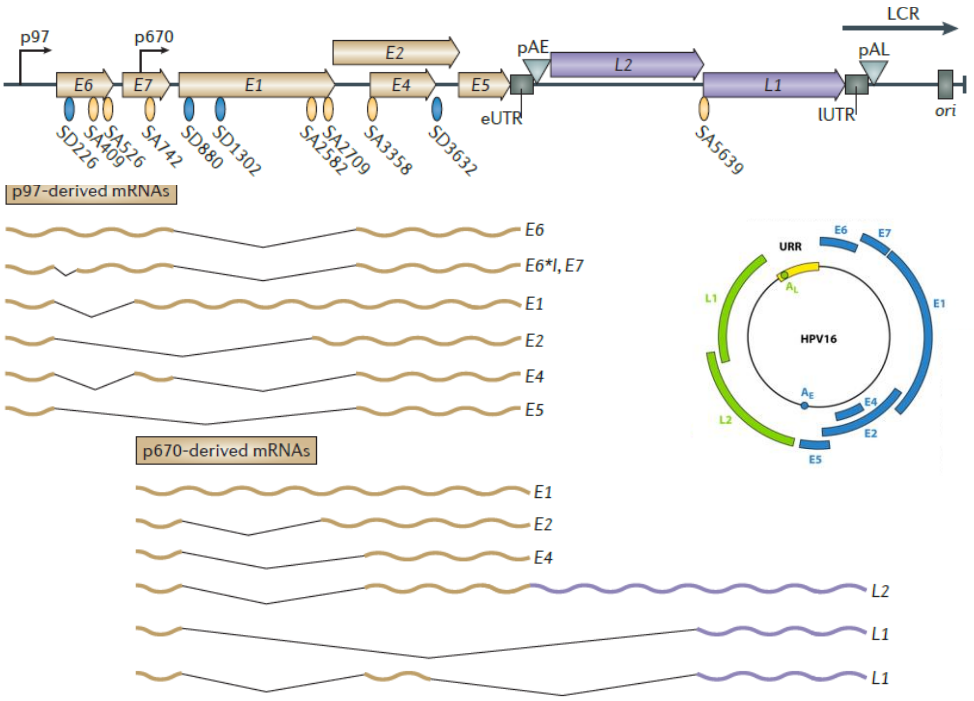


Figure 1-32 Organization of the HPV16 genome and regulation of gene transcription.

Control of early gene transcription and replication is mediated by the upstream regulatory region (URR), which contains promoter and enhancer elements as well as the viral origin of replication; early and late regions are separated by two polyadenylation (A) sites: early, AE and late AL sites. Two promoters, p97 and p670 drive the expression of early and late viral proteins via alternative splicing. Yellow and blue circles represent splice factors-binding sites. Adapted from (Johansson & Schwartz, 2013; Stanley, 2012).

E2 is a 40-45 kDa DNA-binding protein that plays a key role during replication of the viral genome by loading the E1 protein onto the origin of replication in the viral DNA sequence. E2 is the main transcriptional regulator of viral genome, as such it interacts with a broad range of proteins, and among these is Brd4, a chromatin remodelling protein. The interaction, which occurs between the amino terminal transactivation domain (TAD) of E2 and the carboxy-terminal domain of Brd4, is known to be crucial in tethering of viral episomes to cellular chromosomes during mitosis, maintaining viral genomes in an active conformation for transcription as well as transcriptional repression (Jang et al, 2009; Wu et al, 2006; You et al, 2004). In cells in which the viral genome is integrated, the E2 gene is frequently disrupted with con-

sequent loss of the protein (Hamid et al, 2009) and unregulated transcription of E6 and E7 in cervical carcinoma cells (Schwarz et al, 1985).

E4 is translated from a spliced transcript as a fusion protein with E1 (the first 5 amino acids) generating a fusion protein called E1^E4 (Chow et al, 1987; Doorbar et al, 1990) which has little sequence homology between HPV types. E1^E4 proteins are known to associate with the keratin intermediate filaments and in HPV16 they induce collapses of the keratin networks. Keratins constitute the major structural proteins in epithelial cells. This occurs via a domain at the amino terminus of the protein that contains a conserved sequence motif, LLXLL (Roberts et al, 2008). In addition, a proline-rich region is located at the C-terminus which confers upon E1^E4 the ability to interfere with G2-M progression of the cell cycle (Davy & Doorbar, 2007; Knight et al, 2004).

E5 is a small membrane-bound oncoprotein of only 83 amino acids (HPV16). Although HR-HPV E6 and E7 are known to play major roles in malignant transformation of cervical cells, several studies in HPV16 and BPV1 have shown that E5 also is an important mediator for cellular transformation. This occurs via different mechanisms which include up-regulation of EGFR signalling pathway (Pim et al, 1992; Rodriguez et al, 2000), MAPK activation via tyrosin kinase-mediated and protein kinase C (PKC)-dependent pathways modulating cell proliferation, angiogenesis, and anti-apoptosis (Chen et al, 1996; Crusius et al, 1997; Gu & Matlashewski, 1995), disruption of cell-cell communication via down-regulation of Cx26 (BPV1) and Cx43 (HPV16) (Oelze et al, 1995), immune evasion through down-regulation of MHC class molecules (Ashrafi et al, 2005; Zhang et al, 2003).

L1 and L2 are the major and minor capsid proteins, respectively. They are translated from the late promoter and play a key role during the infectious process. L1 is a protein of 55 kDa which self-assembles into virion-like particles (VLP), potent immunogens currently used to formulate VLPs-based vaccine. L1 proteins are highly conserved and form 72 pentamers, each composed of a pentameric L1 capsomer (360 L1 proteins). The minor capsid protein L2, (55 kDa) is mainly located inside the cap-

sid beneath the L1 pentamers and it is not capable to form VLPS like L1. It has been estimated there are approx. 72 L2 proteins/capsid (1:5 L2 to L1 proteins ratio) (Buck & Trus, 2012). Because of L2 location, during virus entry into the host cells, interaction of HPVs with heparan sulphate proteoglycans triggers a conformational change in L2 necessary to expose the amino terminal portion of the protein for cleavage by the pro-convertase enzyme furin which is required for endosome escape (Day & Schiller, 2009; Richards et al, 2006). Both proteins are detected in the nuclei of the terminally differentiated cells and they are required for efficient assembly of infectious virions in the upper layer of the epithelium. During active infection, they both relocate from the cytoplasm to the nucleus and assemble into icosahedral capsids into which the viral genomes are packaged (Conway & Meyers, 2009).

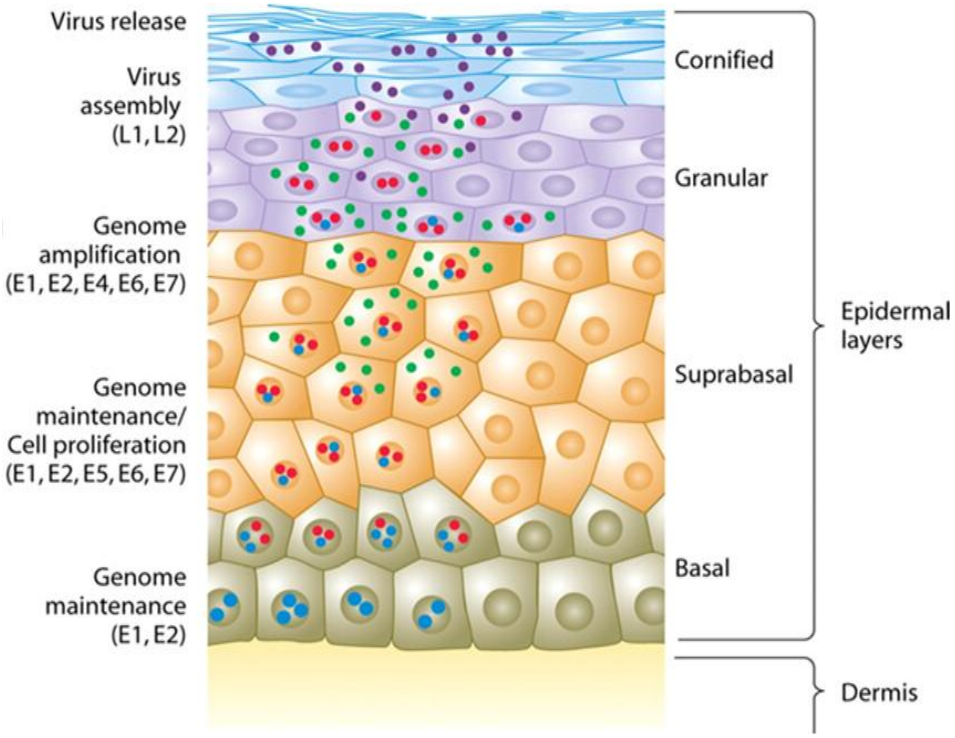


Figure 1-33 Spatial distribution of viral proteins in stratified epithelium.

The viral genomes are replicated in synchrony with cellular DNA replication. Differentiation of HPV-positive cells induces the productive phase of the viral life cycle, which requires cellular DNA synthesis machinery. Blu circles: E1, E2; Red circles: E6, E7; green circles: E4; violet circles: L1 and L2. Adapted from (Lazarczyk et al, 2009).

1.4.3 E6

HPV E6 molecules are small proteins of only 150 amino acids. They are characterised by the presence of two zinc-finger motifs (Cys-X-X-Cys), which are well conserved among various types of HPVs (Howie et al, 2009) (Figure 1-34). The main cellular target of both low and high risk HPV E6 is p53, a tumour suppressor protein involved in maintaining cellular stability by inducing apoptosis in those cells that cannot transit into the S-phase.

When cell cycle regulators are affected by E7, levels of the tumour suppressor protein p53 tend to increase and induce apoptosis in E7-expressing cells that the virus has to counteract in order to maintain its replication in those cells. To this end, high risk HPV E6 proteins recruit a cellular E3 ubiquitin ligase, the E6-associated protein (E6AP) to form a trimeric complex with p53, triggering ubiquitination and proteasomal degradation of p53, thus inhibiting programmed cell death (Demers et al, 1994; Jones et al, 1997; Lechner & Laimins, 1994).

While p53 degradation represents one of the main targets of E6, other cellular mechanisms triggered by E6 proteins impact in the maintenance of cellular stability, such as the ability of E6 to activate the catalytic subunit of the telomerase, hTERT, a key determinant of the enzymatic activity of human telomerase (Klingelhutz et al, 1996). Normal cells undergo only a limited number of cell divisions, until they eventually stop dividing. This process is known as cellular senescence and it is a consequence of the loss of telomeres associated with each round of DNA replication. However, malignant cells are able to induce cell division by activation of telomerase, the enzyme that synthesizes the telomeric repeat DNA (which is lost during division of normal cells). In HPV-positive human keratinocytes cells, the HPV16 E6 protein up-regulates the *hTERT* promoter activity inducing telomerase activity in these cells and it is thought that E6AP may also be involved in this process (Gewin & Galloway, 2001).

E6 proteins from HR-HPVs have been extensively studied and their role in cellular transformation has been established; however, very little is known about the role of LR-HPVs E6 and why they fail to cause malignant disease, triggering only non-invasive lesions. E6 from LR-HPV are known to interact with p53, however they are unable to trigger its degradation; instead it was shown that interaction with p53 (to a different binding site from the one used by HR-HPVs E6) triggers cytoplasmic translocation of p53, binding of p53 to BAK (Bcl2 antagonist killer), leading ultimately to apoptosis (Sun et al, 2008).

p53-independent activities of HR-HPVs E6 (in particular HPV16) that are important for immortalization of keratinocytes are characterised by their interaction with PDZ (PSD-95/Dlg/ZO-1) domain proteins which are known to be involved in the maintenance of cell polarity in epithelial cells (Thomas et al, 2008). Apico-basal and planar cell polarity is regulated at many levels and several proteins are involved, namely the Scribble, the partitioning defective (PAR) and Crumbs (CRBS) complexes; these complexes ensure correct cell-cell junctions development (Assemat et al, 2008). However, the C terminus of HR-HPV16 and 18 E6 proteins contains a well conserved PDZ domain motif (PBM) which interacts with the PDZ domain proteins with deleterious consequences on cell polarity establishment. Among the well-established interactions are those occurring between hDlg (human homologue of *Drosophila* Disks large) MAGI-I (membrane-associated guanylate kinase), ZO-I (zonula occludens I), MUPP-l (multi-PDZ domain protein-I) and hScrib (Javier, 2008).

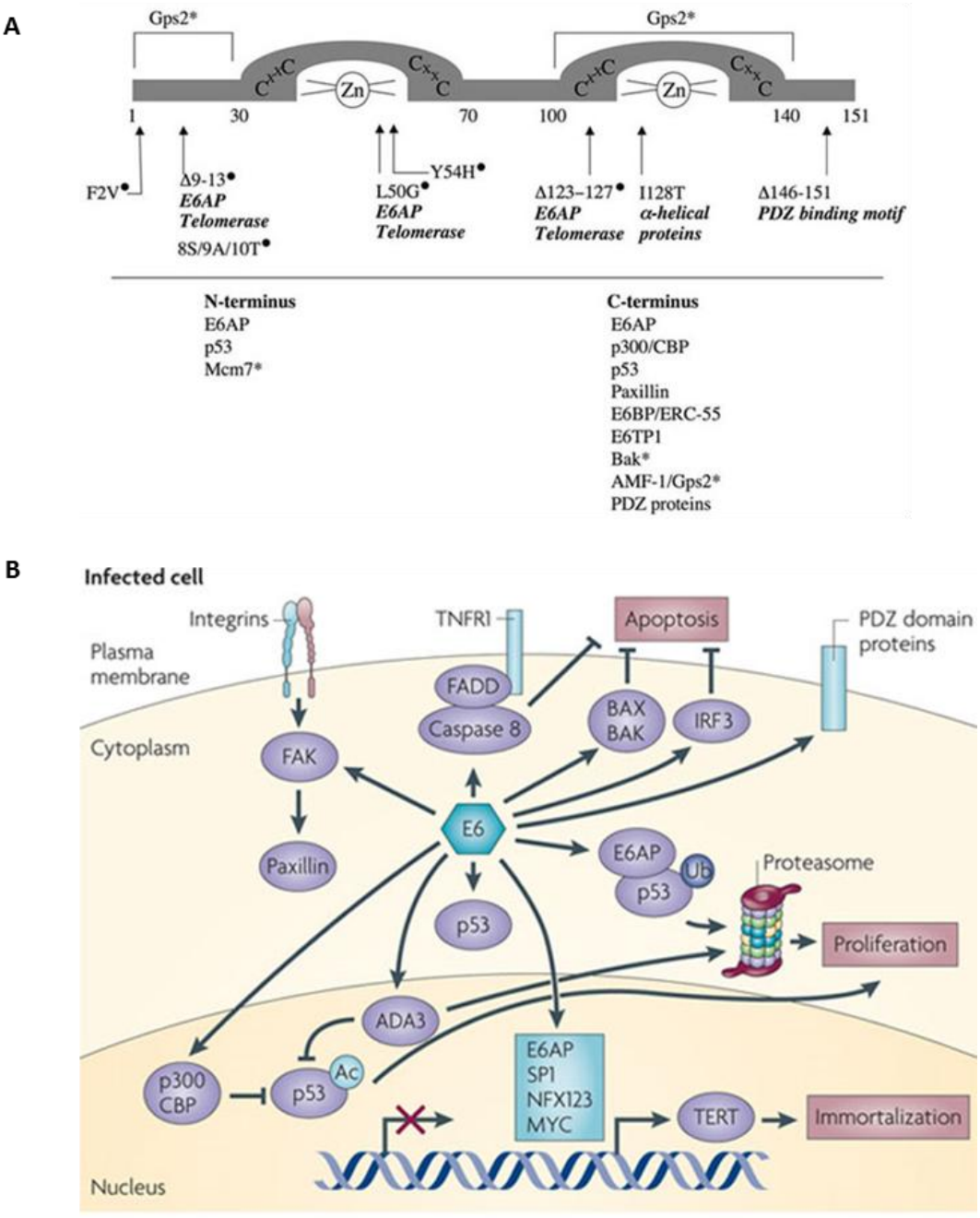


Figure 1-34 Schematic representation of the actions of HPV16 E6.

A.HPV E6 conserved regions and their target proteins (Wise-Draper & Wells, 2008). B. High-risk E6 proteins inhibit p53-dependent growth arrest and apoptosis in response to aberrant proliferation through several mechanisms, resulting in the induction of genomic instability and the accumulation of cellular mutations. Formation of an E6–E6-associated protein (E6AP)–p53 trimeric complex results in p53 degradation, and the interaction of E6 with the histone acetyltransferases p300, CREB binding protein (CBP) and ADA3 prevents p53 acetylation (Ac), inhibiting the transcription of p53-responsive genes (Moody & Laimins, 2010).

1.4.4 E7

E7 protein is a small protein of only 100 amino acids in length (18 kDa) characterised by the presence of three conserved regions (CR) called CR1, CR2 and CR3. These conserved regions are crucial for the oncogenic activities of HR-HPVs. In particular, CR2 is known to contain a conserved motif, namely LXCXE, important for viral maintenance and binding to pRb, the main target of E7 proteins (Munger et al, 1989b). In normal cells, pRb exists in a complex with E2F and its main function is to repress cell cycle progression into S-phase. The G1/S-phase transition is under the control of cyclins, cyclin dependent kinases (cdks), and the inhibitors of cdk (CDKi), which are known to negatively regulate the activity of the cdks. When cells are ready to progress to the S phase following extracellular signals, pRb is phosphorylated by cdks, dissociating it from E2F and thus it is free from the complex and cells are able to enter the S-phase (Dyson, 1998; Zerfass et al, 1995). During HPV infection, E7 proteins bind to the hypo-phosphorylated form of pRb, sequestering it from E2F and triggering cells to enter the S-phase. This interaction is coupled with pRb degradation by the calcium-activated cysteine protease, calpain, which cleaves pRb, as well as other Rb members which are targeted by E7 for degradation via the ubiquitin proteasome family (Boyer et al, 1996). Although all E7 proteins bind to pRb, those from the HR-HPVs have a higher binding affinity (Longworth & Laimins, 2004). Figure 1-35 also shows additional mechanisms adopted by E7 to induce genomic instability in infected cells, such as the ability of E7 to trigger abnormal centrosome synthesis by interacting with γ-tubulin and to induce DNA damage and activation of the ATM–ATR pathway (ataxia telangiectasia-mutated–ATM and RAD3-related DNA damage response) (Moody & Laimins, 2010). In addition, it has been shown that E7 proteins can induce cells to enter the S-phase independently of pRb interaction. This is achieved by E7 interaction with HDAC (histone deacetylase) to deacetylate E2F factors. This interaction is known to occur via a site on E7 different from that used by pRb, which also interacts with HDAC in normal cells to repress the transcription of E2F genes (Harbour & Dean, 2000).

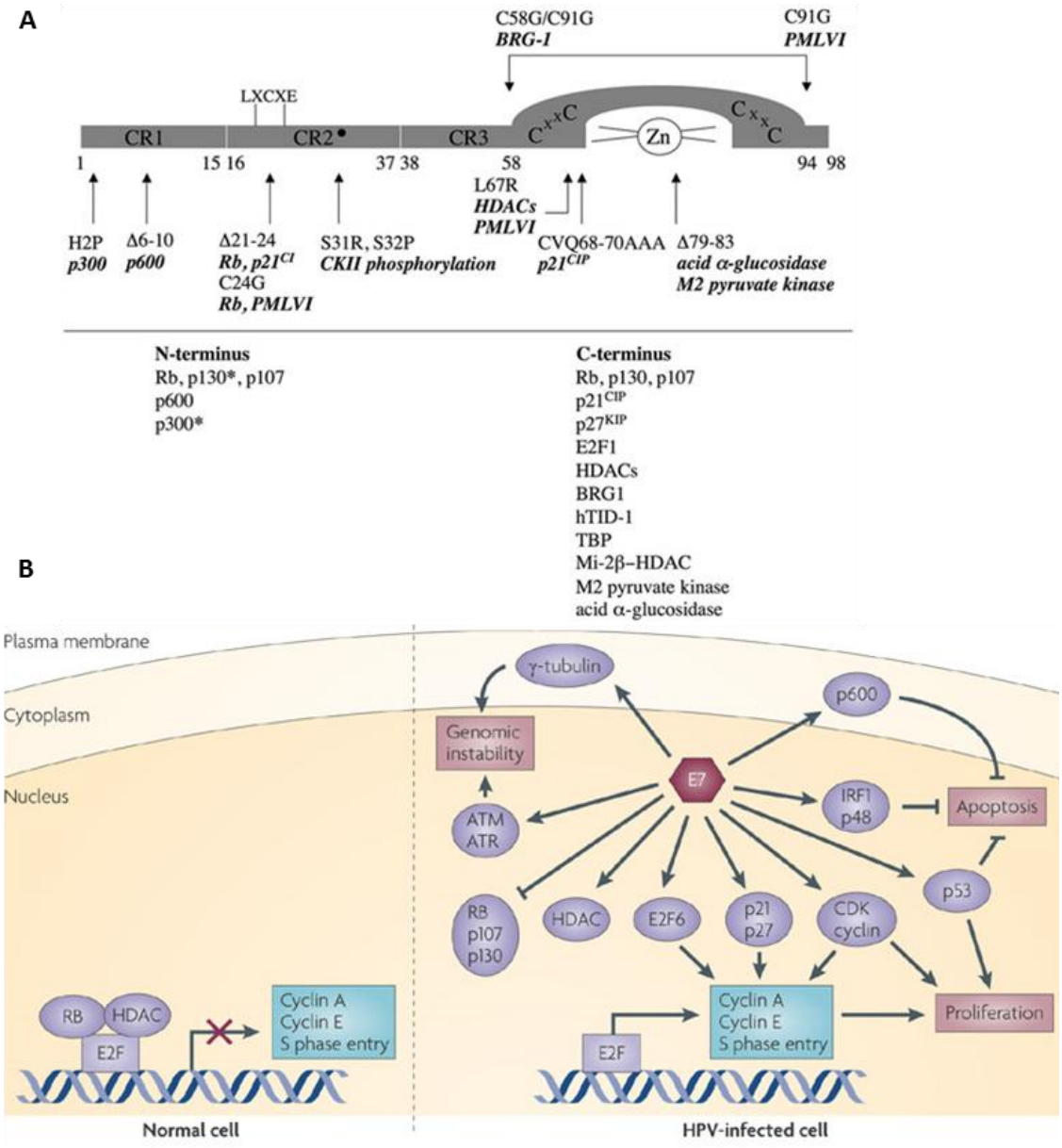


Figure 1-35 Schematic representation of the actions of HPV16 E7.

A. HPV E7 conserved regions and their target proteins (Wise-Draper & Wells, 2008). B. High-risk human papillomavirus (HPV) E7 proteins subvert G1–S arrest and induce hyperproliferation through inhibition of retinoblastoma (Rb) family members and constitutive activation of E2F-responsive genes. E7 also affects cellular gene expression through interaction with histone deacetylases (HDACs) and E2F6. E7 further deregulates cell cycle control through inhibition of cyclin-dependent kinase inhibitors (such as p21 and p27), stimulation of cyclins and through direct activation of cyclin-dependent kinase 2 (CDK2) (Moody & Laimins, 2010).

1.4.5 HPV life cycle: virus/host cell interactions

Being an obligatory intracellular parasite, HPV must be able to transfer its own genetic material into the host cell in order to replicate using the host cell replication machinery (Pelkmans & Helenius, 2003). To follow these steps in experimental studies has been difficult due to the fact that the replication of papillomaviruses is linked to the differentiation programme of keratinocytes, making propagation of the virus difficult: to date, HPVs have been propagated using xenograft and raft culture systems developed in the early 1990's (Kreider et al, 1985; Meyers et al, 1992) followed by subsequent use of DNA-free virus-like particles and pseudovirions (Roden et al, 1996; Rossi et al, 2000; Unckell et al, 1997), both of which have been produced in large scale using packaging cell lines such as 911 and 293 (Buck et al, 2004). Further studies also suggested that pseudovirions needed to be activated in order to infect organotypic raft cultures or primary keratinocytes *in vitro* (Day et al, 2008).

HPV-host cell interactions have been studied using virus-like particles (VLPs) which are produced by synthesis of the L1 and L2 capsid proteins (Kirnbauer et al, 1992; Rose et al, 1993; Volpers et al, 1994; Zhou et al, 1991) since HPVs do not grow in standard submerged cell cultures. Using this system it was established that HPV may use either cell surface heparan sulphate proteoglycans (HSPGs) or α-integrins as their primary attachment receptor (Evander et al, 1997; Giroglou et al, 2001; Joyce et al, 1999; Patterson et al, 2005; Shafti-Keramat et al, 2003). It was also reported in *in vitro* studies that laminin-5 may be involved in the initial interaction with the host cell (Culp et al, 2006).

HSPGs are cell surface and extracellular matrix molecules that are ubiquitously expressed. They are involved in cell adhesion, migration, proliferation, and differentiation and they serve as major attachment molecules for many viruses (Knappe et al, 2007). HSPGs include sydecans and glypicans which are molecules with attached glycosaminoglycan chains, mainly heparan sulphate (Bernfield et al, 1992; Fransson, 2003). These can be subjected to post-translational modifications such as sul-

phation and acetylation, generating a large variety of molecules (Knappe et al, 2007). Evidence that heparan sulphate plays a critical role in HPV-host cell surface interaction comes from elegant studies where removal of HSPG by addition of heparanase or heparinitase resulted in significant reduction of HPV11 VLPs binding (Joyce et al, 1999). Heparin was also effective in inhibiting HPV16 and HPV33 by pseudovirus infection. Similarly, the ability of different HPV VLPs such as HPV16, 18, 31 (classified as high risk) to enter COS-7 cells was successfully inhibited with heparin (Combita et al, 2001). Among the different HSPGs, syndecan-1 is believed to be the main attachment molecule for HPV since it is highly expressed in epithelial tissue and it is up-regulated during wound healing in target cells (Elenius et al, 1991; Gallo et al, 1994), suggesting that in the presence of a microtrauma, these molecules would allow the uptake of virus particles (Sapp & Day, 2009; Selinka et al, 2002).

However, not all HPVs use HSPGs as attachment molecules, such as HPV31 (Patterson et al, 2005). As mentioned above, HPVs are also known to interact with integrins to enter host cells. Integrin are cell-surface heterodimeric proteins which span the phospholipid bilayer and are composed of α and β subunits (Hu & Luo, 2013). HPVs have been shown to interact mainly with the α-integrin subunit (Letian & Tianyu, 2010; McMillan et al, 1999; Yoon et al, 2001). Even though several studies focused on the role of integrins in virus interaction and entry, whether the latter can be considered a primary or secondary receptor is still controversial. Nevertheless, several studies have shown that efficient interaction requires co-receptors and that it is unlikely that the virus will enter the host cell only following interaction with HSPGs.

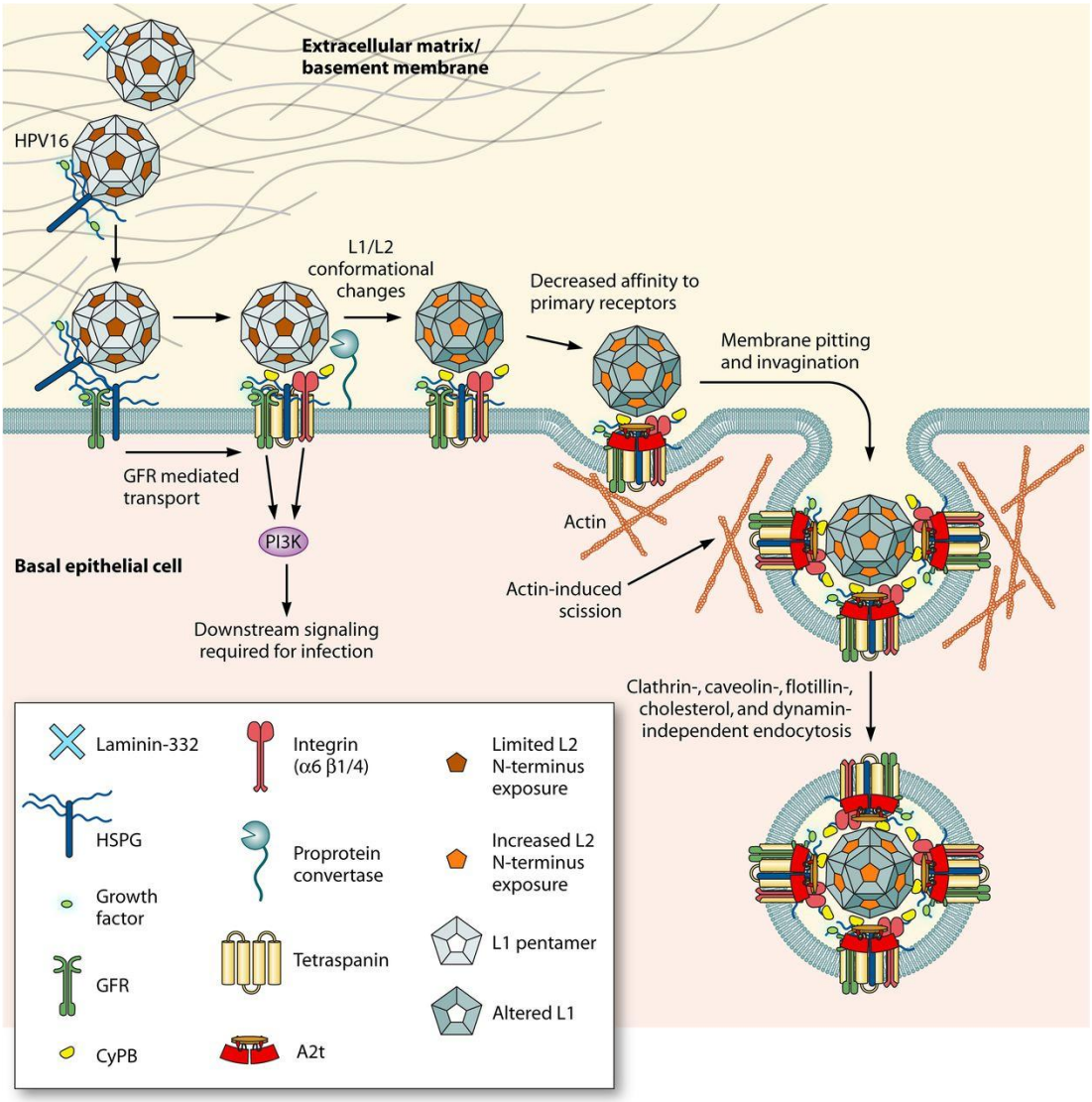


Figure 1-36 HPV/host cell interactions.

The ECM consists of collagens, elastins, fibronectins and laminins. HPV16 interacts with HSPGs either through laminin 332 (formerly laminin 5), the basement membrane or the epithelial cell surface. Following this interaction, the virion undergoes a conformational change facilitated by cyclophilin B (CyPB); therefore, HPV16 binds to α6 integrin, which in turn triggers a second intracellular signalling. Following a series of conformational changes and signaling, the HPV16 capsid binds to A2, which triggers clathrin-, caveolin-, lipid raft-, flotillin-, cholesterol-, and dynamin-independent endocytosis of HPV16; vesicle closure and actin scission is promoted by PI3K. Following the initial attachment to to HSPGs, the epidermal growth factor receptor (EGFR) and/or keratinocyte growth factor receptor (KGFR) may also become activated and may initiate intracellular signalling cascades that include the activation of the phosphoinositide 3-kinase (PI3K) pathway (Raff et al, 2013).

1.4.6 HPV life cycle: mechanism of virus entry

HPV L1/L2 virus-like particles will not bind or infect primary cultured epithelial cells or epithelial tissue *in vivo* (Roberts et al, 2007). The primary site of HPV infection *in vivo* is the basement membrane which is the site with actively dividing cells and the virus will only enter after the tissue is damaged/micro-wounded (Sapp & Bienkowska-Haba, 2009). During mitosis of basal cells, one will become a new basal cell while the other daughter cell will migrate away from the basal layer to complete differentiation: this cell will exit the cell cycle and eventually die by apoptosis (desquamation). Because the basement membrane is the primary site of infection, it was suggested that a secreted form HSPG must be involved (Roberts et al, 2007). The entry mechanism requires L1 interaction with HSPG-1, triggering a conformational change in L1 that allows interaction with a secondary receptor, HSPG-2. These two interactions are responsible for the conformational change observed in L1 and L2 capsid proteins: the L2 N-terminus (amino acids 17–36) becomes accessible and prone to cleavage by furin (Day & Schiller, 2009; Richards et al, 2006). This cleavage step is essential for successful infection, suggesting that the N-terminus of L2 is important to lead the necessary conformational changes observed. Consequently, the virus will be transferred and associated stably with a receptor on the epithelial cell surface. Following interaction, the virus is internalized either via endocytosis, clathrin-or cavelolin- mediated (or both) (Hindmarsh & Laimins, 2007; Laniosz et al, 2008; Smith et al, 2007). It has been shown that certain HPVs might be internalised by a new mechanism involving tetraspanin-enriched domains (Spoden et al, 2008).

1.4.7 Viral genome: episomal and integrated

Following internalization, HPVs capsids are thought to be lost due to their size, which will not allow the virus entry through the nuclear pores (Nelson et al, 2000). It is known that L2 is not essential for the uncoating, but mainly to free the viral DNA from the endosomes (Kamper et al, 2006).

Once in the nucleus (by mechanisms which are not well understood), the virus genome can exist as episome (circular extra chromosomal element) or integrate into the host cell genome. The latter constitutes a key step in cancer progression after infection by HR-HPVs.

In the nucleus, the virus is replicated and amplified through three different stages: establishment, maintenance, and productive stages. Initially, the virus will replicate in the host cell to establish around 100 episomal copies/cell (Moody & Laimins, 2010). The E2 proteins that ensure that episomes are not lost during cell division maintain this function. In the following stages, the virus will replicate as the basal layer keratinocytes enter the cell cycle. E2 also cooperates with E1 to allow virus replication, while E6 and E7 modulate host cell cycle regulators to accommodate long term replication. In the last stage, episomes are amplified exponentially in terminally differentiating keratinocytes and finally packaged into new progeny virions.

In basal cells that are not undergoing differentiation, overall HPV proteins are expressed at very low levels. When the cells start differentiating, viral proteins are expressed at very high levels allowing the virus to escape the immune response that is lower in the differentiating cells destined to die (Frazer, 2009). By doing so, the virus ensures a long-term infection, however since it does not encode its own replication enzymes and the cells undergoing differentiation withdraw from the cell cycle, it will not replicate. To face this problem, HPVs force the cells to re-enter the cell cycle: to this end, a key role is played by the E7 proteins that bind and inactivate pRb. By producing high levels of viral proteins, the virus is able to increase the genome copy number from the initial to up 100 thousands of copies per cell (Bedell et al, 1991; Fehrmann et al, 2003; Ozbun & Meyers, 1998). Because the viral particles are released as consequence of desquamation, in the absence of lysis/necrosis, there will be no sign of evident inflammation (i.e. no immune response).

1.4.8 Immune response against HPVs

Human papilloma viruses (HPVs) attempt to escape the immune system by limiting viral protein synthesis in terminally differentiated cells, and due to the absence of a cell lysis event, they reduce their exposure to the immune response (Stanley, 2009). The presence of HPV should be detected by intraepithelial dendritic cells (DCs); following entry of most viruses into host cells, the capsid proteins are recognised by DCs which are then activated, however it seems that DCs are not activated by the uptake of HPV capsids, suggesting a limited role in the host's response to HPV infection (Fausch et al, 2003; Fausch et al, 2002). Indeed in 2009, Einstein MH stated that "…these escape mechanisms have enabled HPV to become one of the most common sexually transmitted infections worldwide" (Einstein et al, 2009)". Despite this, the immune system, both innate and adaptive, does act in order to limit the infection since most of the infections are cleared within a couple of years (Steben & Duarte-Franco, 2007). In particular, the key role played by T cells, and thus cell-mediated immune response, is clear in suppressed patients such as HIV-positive patients, where clearance of HPV infection is significantly compromised (Koshiol et al, 2006; Scott et al, 2001).

1.4.8.1 From infection to cancer of the cervix

HPV-infected patients are able to clear the infection within a period of 24 months; however, persistence of this infection and viral genome integration probably near fragile sites, leading to genomic instability, can cause abnormal cell growth and changes in the cervix; these changes are identified through a pap smear test which is interpreted as atypical squamous cells of undetermined significance (ASCUS). This process of transformation of the cervical tissue takes usually 10-20 years and suggests that HPVs have evolved a mechanism to evade the immune system and induce cancer (Feller et al, 2010; Koutsky et al, 2002) (Refer to Chapter 1.4 for detailed information about HPVs). It is widely accepted that HPV infection is linked to cancer of the cervix (Walboomers et al, 1999). However, currently there are no therapeutic

vaccine strategies available to defeat the infection and eradicate cervical cancer. The closest achievement has been the development of virus-like particles (VLP) composed of the L1 viral capsid protein being engineered as a tool for new prophylactic vaccination strategies to prevent HPV infection (Koutsky et al, 2002). In recent years, the development of two FDA-approved drugs, Cervarix and Gardasil, have been also administrated to girls as vaccine prevention against HPV infection. Both vaccines are highly effective in preventing infections with HPV types 16 and 18, the two major high risk HPVs, which cause nearly 70% of all the HPV infections. However, the current vaccines do not protect against all HPV infections that might cause cancer. The first drug, Cervarix is produced by GlaxoSmithKline (GSK) and in the UK, its administration started in 2008. Cervarix was used in females' aged 9 to 25 to prevent cervical cancer caused by HPV types 16 and 18. However, after a 4 year trial and 4 million doses administrated, it was replaced by Gardasil in 2012, due to adverse reactions registered among the patients (MHRA, 2010). Gardasil is manufactured by Merck & Co., Inc. and it is a quadrivalent vaccine as it protects against HPV6, 11, 16, and 18. Gardasil is administrated to girls aged 9-26.

HPVs infect cells in the basal layer of stratified squamous epithelia which are exposed to the virus following micro wounds; synthesis of new virions following HPV infections occurs after these cells undergo mitosis and one of the infected daughter cells has differentiated. Basal cells are the only proliferating cells in normal epithelia, while the differentiated cells in the suprabasal layers have exited the cell cycle. Initially, the virus genome is found in episome copies (extrachromosomal elements). To this end, one of the virus proteins, E2, regulates episome copy number by repressing gene expression of the early proteins (Stubenrauch et al, 1998). If the immune system does not clear persistent HPV infections, these can trigger the development of cervical cancer over several decades. The route from infection to tumour initiation involves not only evasion from the immune system and persistent infection. In addition, integration of the viral genome disrupts the E2 gene and its expression, consequently up-regulating expression of the early genes, E6 and E7. Ac-

cordingly, most HPV16- and 18-positive malignancies contain integrated HPV genomes, suggesting that integration may be a crucial event in cancer development and progression. Furthermore it has been shown that cells that have integrated copies of the virus have a selective growth advantage over cells that harbour only episomal copies of the viral genome (Jeon et al, 1995; Jeon & Lambert, 1995). It has also been suggested that the coexistence of HPV episomes with integrated copies might be a crucial step in carcinogenesis (Kadaja et al, 2009).

1.4.8.2 Role of E6 and E7 in cancer development

The major viral proteins involved in the initiation and progression of cervical cancer are the high risk HPV (HR-HPVs) E6 and E7 (McLaughlin-Drubin & Munger, 2009), two small proteins which mainly localize in the nucleus, although a recent study has shown that high cell confluence or density might result in E7 and E6 localization in the cytoplasm (Laurson & Raj, 2011). Both proteins are able to interact with cellular transcription factor complexes, thus altering their activities, but the most important targets are represented by the cellular tumour suppressors pRb and p53. Interestingly, only E7 proteins from the HR-HPVs such as HPV16 and HPV18 can immortalize human keratinocytes on their own, while HR-HPV E6 does not have such an activity. In addition, transformation of cells with E7 and E6 was found to be more efficient than the two oncogenes alone, suggesting that cell immortalization requires the activity of both viral proteins (Hawley-Nelson et al, 1989; Munger et al, 1989a). Elegant experiments using both organotypic raft cultures and transgenic mice have shown that while E7 alone can immortalize cells and cause high grade cervical dysplasia, co-expression of E6 leads to more extensive cervical cancers (McCance et al, 1988; Riley et al, 2003). However, since most people infected with HPV do not necessarily develop cancer, co-expression on the two oncogenes may be necessary but not sufficient for tumour progression, suggesting that other oncogenes or factors might be involved. Genome instability plays a crucial role in tumour progression and maintenance as shown by the number of chromosomal abnormalities (such as those

involving centromeres) found in HPV-positive patients. The two oncogenes can also induce genomic instability by inducing DNA damage via ATM pathway activation (Duensing et al, 2000; Moody & Laimins, 2009).

1.5 Role of gap junctions in cell communication

Cell-cell communication is essential to maintain homeostasis and coordination within the cellular environment; in eukaryotic cells tight junctions (TJs) and gap junctions (GJs) are responsible for the correct functioning of cell-cell communication processes. This chapter will primarily focus on GJs as they are of particular interest for the results presented in Chapter 5.

GJs are a group of ubiquitously expressed proteins with a half-life of only a few hours. They are involved in direct intercellular exchange of small molecules (< 1 kDa) and metabolites, second messengers, electrical signals and ions and they play a key role in the homeostasis and controlled growth and differentiation of the cells (Laird, 2006). GJs are composed of connexin proteins and so far 21 connexins have been discovered in humans and 20 in mice (named based on their molecular weight), whose combinations give rise to a vast array of GJs (Sohl & Willecke, 2004). The basic unit of a GJ is a connexin protein that forms hexamers with other connexins to form a connexon (Figure 1-37). All connexins share a common structure containing four transmembrane domains, an N-terminus and C-terminus intercellular domains, and 3 extracellular loops. Their cytoplasmic domain is often subject to post-translational modifications, such as phosphorylation by protein kinases (PKA, PKC, PKG and mitogen-activated protein kinase among others) with the exception of Cx26 that does not have a long cytoplasmic tail and is thus less of a target for post-translational modifications.

Lower levels or lack of expression of connexin within a tissue is considered to be a sign of tumorigenesis and cell transformation (Czyz, 2008; Naus, 2002). Connexins are therefore considered to be tumour suppressor genes, despite their role in cell transformation being still unclear. In experimental studies, transfection of a con-

nexin is known to down-regulate cell growth (Fujimoto et al, 2005; Fujimoto et al, 2007). For instance, Cx26 was shown to be a tumour suppressor gene in HeLa cells and breast cancer cells as well as in lung cancer cells (Hirschi et al, 1996; McLachlan et al, 2006; Mesnil et al, 1995; Plante et al, 2011) , although deafness is the most frequent disorder due to mutations in this gene. Tumorigenesis might be driven either because of loss of connexin gene transcription or aberrant trafficking of connexin proteins to the plasma membrane.

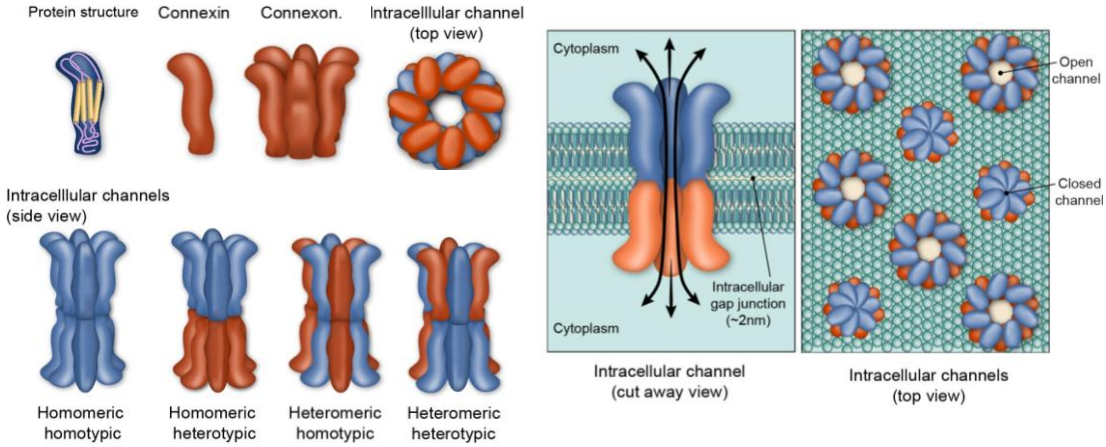


Figure 1-37 Structural organization of gap junction proteins.

Adapted from (Maeda et al, 2009).

1.5.1 Regulation of connexin protein expression

Connexins are expressed in a tissue- or cell-specific manner and several organs or cell types express more than one connexin; for example, keratinocytes express Cx26, Cx30, Cx30.3, Cx31, Cx31.1 and Cx43 (Goliger & Paul, 1994; Wiszniewski et al, 2000). Studies on the expression of connexins during keratinocyte differentiation have shown that Cx26 is expressed in both the basal and spinous layers, while Cx43, one of the most abundant connexins, is also found in the granular layer (Figure 1-38). This pattern of expression is controlled via regulation of connexin proteins at different levels.

Connexin expression can be regulated at the transcriptional, post-transcriptional and translational levels. Knowledge of their gene structures is essential to under-

stand these mechanisms. All connexin proteins share a basic gene structure comprising a 5'-untranslated region (5'-UTR) on exon 1 separated by the coding region and a 3'-UTR. The coding region is generally found within the second exon where the 3'-UTR is also found (Figure 1-39). The basal promoter is approx. 300 bp upstream of the TSS (which is located in exon 1) and contains the binding sites for the TATA binding protein and transcription factors such as Sp1/Sp3 and AP-1, which have been described to be important for the transcriptional activity of connexin genes, as is the case for Cx26, although AP-1 binding sites have not yet been found (Tu & Kiang, 1998). While the activity of several other factors is undoubtedly essential for cell- and tissue-specific expression of the connexin proteins, an interesting aspect of transcriptional control is represented by methylation of connexin promoters that triggers gene silencing. Epigenetic inactivation through promoter hypermethylation was shown for several connexin proteins, included Cx26. Tan et al. showed that the Cx26 promoter was frequently methylated in breast cancer cells and treatment with 5'-aza-2'-deoxycytidine resulted in re-expression of Cx26 mRNA (Tan et al, 2002), although other groups failed to find a correlation between Cx26 promoter hypermethylation and protein expression as observed in mammary epithelial and oesophageal cancer cell lines (Loncarek et al, 2003; Singal et al, 2000).

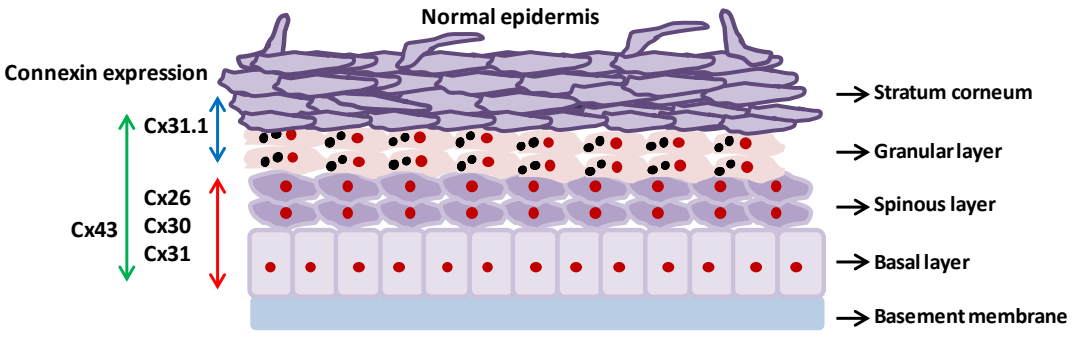


Figure 1-38 Pattern of expression of connexin proteins in the epidermis.
Adapted from (Martin et al, 2014).

In addition to transcriptional control, connexin protein expression was found to be modulated by post-translational events. Earlier studies on Cx26 and Cx43 have iden-

tified internal ribosome entry sites (IRESs) within the 5'-UTR of the transcript. Translation of most eukaryotic cells proceeds from a translational start codon at the capped 5'-UTR and ends at a stop codon found near the beginning of the 3'-UTR. However, in some genes, such as Cx26 and Cx43, a cap-independent translation mechanism was discovered, where translation occurred using an IRES rather than the canonical 5'-cap structure, which means that higher levels of proteins were produced. In an elegant experiment, Lahlou and colleagues showed that increased levels of Cx26 and Cx43 proteins in density-inhibited human pancreatic cancer cells (BxPc-3) were dependent on translation from the IRES and that subsequent generation of GJs was essential to counteract the density-inhibition of BxPC-3 cells triggered by the anti-proliferative somatostatin receptor 2 (sst2)(Lahlou et al, 2005). This shows that, while transcriptional controls regulate mRNA levels that in turn might reflect low or high protein levels, the latter can be achieved via alternative translational routes.

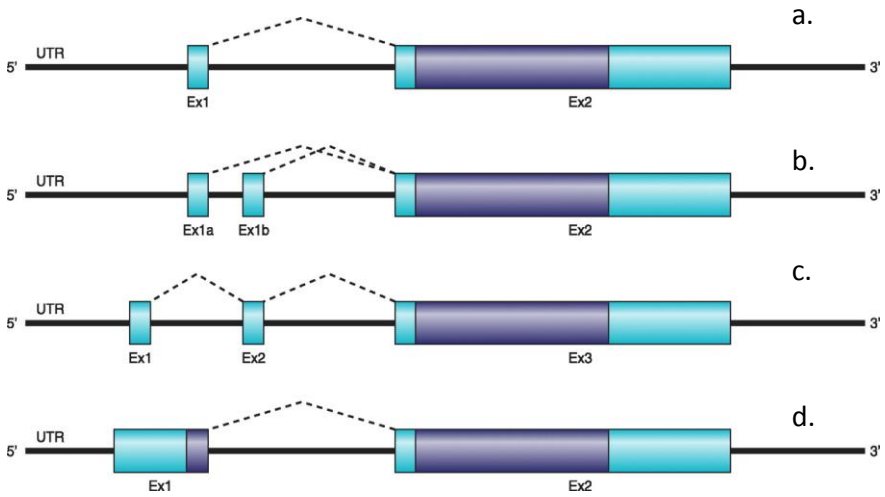


Figure 1-39 Splicing patterns of connexin gene transcripts.

Connexin genes have a similar structure, which comprises between 1 and 3 exons. The coding region is generally confined within exon 2 (or 3 where present; violet), although some connexin genes might have the coding region in two separated exons (d), have exons transcribed from different promoters that generate alternative first exons (b) or have splice variants (c). UTR, untranslated region; Ex, exon. Adapted from (Bosco et al, 2011).

1.6 DNA sequencing

DNA sequencing has become an invaluable tool in molecular biology allowing scientists to sequence long genomes in a relatively short period of time and reduced costs.

The first DNA sequencing methodology was developed by Sanger in 1975 (Sanger & Coulson, 1975). This method called “plus and minus” was a simple method based on the comparison between polynucleotide lengths extended with (plus) or without (minus) specific nucleic acids (A, C, T, or G) in order then to reconstruct the original sequence. However it was only possible to determine about 50 bases per reaction. In the following years Maxam and Gilbert in Harvard, and around the same period Sanger in Cambridge, developed two methodologies for DNA sequencing (Maxam & Gilbert, 1977; Sanger et al, 1977) (Figure 1-40). The Maxam-Gilbert method was based on the chemical modification of DNA and subsequent cleavage at specific bases but due to the extensive use of radioactivity it was subsequently modified by Sanger who introduced the chain termination method. This new method still uses toxic chemicals and radioactivity, but in lower amount (this is known today as “first generation sequencing”). The Sanger methodology synthesized a new DNA molecule on a template strand with its synthesis being stopped when one of four dideoxy nucleotides (lacking a 3’ hydroxyl group) was incorporated, thus stopping the overall reaction of chain growth. These molecules were then run on an electrophoresis gel and based on their size they were put in order, reconstituting the original molecule by reading the last (known) nucleotide added.

The Sanger method represents the basic methodology used in later sequencing technologies. Rather than radiolabelled nucleotides, the DNA molecule are fluorescently labelled, with each of the nucleotides represented by a different colour and then separated based on their size using glass capillaries where the fluorescent dye is detected and recorded following laser excitation.

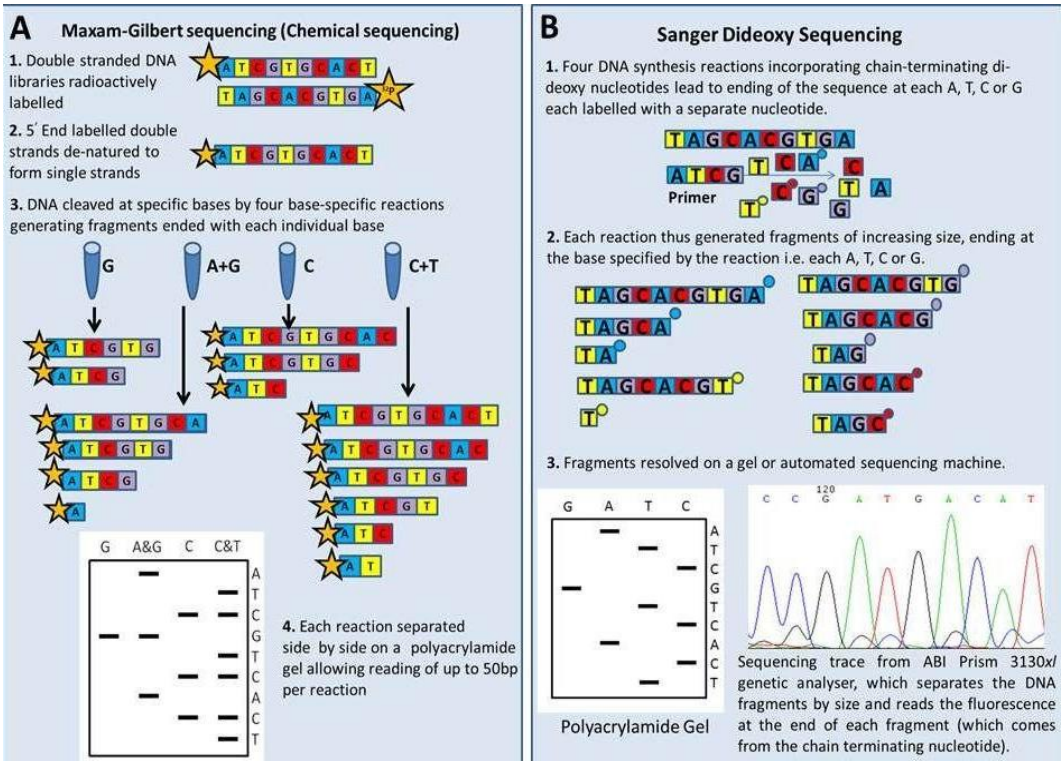


Figure 1-40 Maxam-Gilbert and the Sanger sequencing methods.

The picture depicts the main difference between the two methodologies (www.oxbridgebiotech.com).

1.6.1 Microarrays and other transcriptome sequencing assays

Microarrays made their entry into the market in the mid-1990s, with the first report appearing in Science in 1995 (Schena et al, 1995). The technology is based on the principle of DNA and RNA hybridization: a target sample (DNA or cDNA) is hybridised against a large set of probes attached to a solid support (Miller & Tang, 2009). Microarrays are used to measure changes in gene expression levels, i.e. between a normal and a cancer cell, to analyze chromosomal aberrations, to observe genomic gains and losses via microarray Comparative Genomic Hybridization (CGH) or analyze mutation in DNA samples.

In brief, a plate made of glass (or silicon) is prepared so that it contains imprinted all the probes for the genes (these are represented as “spots”). Target samples are properly prepared depending on the downstream application. In case of gene pro-

file expression, mRNA will be isolated and reverse transcription will be performed to generate cDNA; a mix containing oligoT, primers and fluorescently labeled oligonucleotides will be used (green for the control sample, red for the test sample, for example) (Figure 1-41). In this way, the complementary strand being generated will incorporate the labeled oligonucleotides. This cDNA is hybridized with the probes (ssDNA) on the microarray plate. Following laser excitation, the hybrid bonds will fluoresce and a camera connected to a computer will record the images (Schena et al, 1998). Results are then analyzed: given the colors used, if the spot is green it means the control sample hybridized more than the test sample, if it is red the test sample hybridized more than the control sample and if is yellow then both samples hybridized equally to the target DNA. Some area will show up black meaning neither sample hybridized to the target DNA.

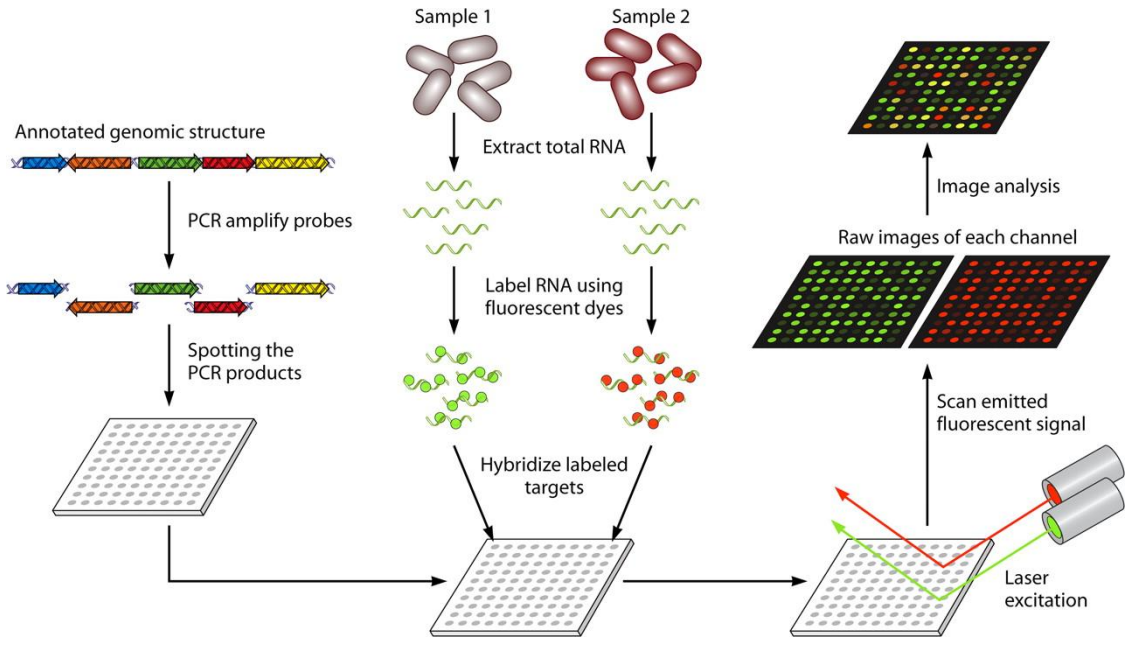


Figure 1-41 Overview of microarrays technology.

For gene expression analysis, probes are prepared by PCR amplification of the target genomic region; for transcript expression levels total RNA is isolated and reverse transcribed; cDNA is labelled using fluorescent dyes; two different colours can be used for the two conditions being studied. The labelled target is placed onto a DNA microarray where it will hybridise to the DNA probe previously synthetized. The microarray is then scanned to quantitate the signal and differences between the samples analysed (Miller & Tang, 2009).

Microarrays were a breakthrough in biology. The simplicity of the assay and the possibility to analyse thousands of genes on a single chip simultaneously made microarrays a powerful tool for over two decades, representing the main tool for several studies in cancer research. Microarrays revealed information about genes that were expressed in certain conditions and their relative magnitude, reflecting the cellular response to the condition being studied.

One of the first studies of microarrays applied to human gene expression levels was published in 1999, investigating fibroblasts response to serum: since addition of FBS triggers proliferation of the fibroblasts, microarray were the ideal tool to study the parallel change of gene expression (Iyer et al, 1999). Since then, microarrays have been used in a vast number of studies including gene expression in lymphomas (Alizadeh et al, 2000; Alizadeh & Staudt, 2000), breast cancer (Chou et al, 2013), colorectal cancer (Hegde et al, 2001; Koga et al, 2014; Moura Franco et al, 2013), oral cancer (Alevizos et al, 2001; Sumino et al, 2013), prostate cancer (Luo et al, 2001), melanoma (Clark et al, 2000) as well as cervical cancer (Jang et al, 2011; Srivastava et al, 2014) and ovarian carcinomas (Bayani et al, 2002; Mok et al, 2001), reaching over 25,000 publications since their first appearance in 1995. In addition, DNA microarrays have been used for the detection of single nucleotide polymorphisms (SNPs), microRNAs analysis as well as CpG islands of target genes.

In recent years, however, next generation sequencing (NGS) technologies have begun to rival microarrays. There are several reasons for this, e.g the higher sensitivity and specificity for NGS, and the fact that microarrays are dependent on previous knowledge of the gene probes, thus covering fewer genes than NGS. Yet, both technologies need robust bioinformatics and statistical approaches to analyse the large amount of information generated, a problem still persisting in both technologies.

1.6.2 Next generation sequencing

Next generation sequencing is a term used to describe technologies that have the ability to massively parallel sequence millions of DNA templates. "Second-generation" and "third-generation sequencing" are terms used to differentiate the newly developed sequencing technologies from the Sanger methodology defined as "first-generation sequencing". Next generation sequencing gives a more complete analysis of the genome by generating larger amounts of data, within a comparatively shorter period of time and at lower costs (compared to the first generation sequencing technologies), being more comprehensive than microarrays (Ledford, 2008). Most of the second generation platforms were developed during the Human Genome Project, but since then, the applications have been numerous, going beyond simple DNA sequencing. Based initially on the use of bacterial artificial chromosomes (BAC), the complete genome was assembled by "pasting" together the clones based on known transcripts and overlaps. Due to the difficulty encountered in assembling the sequence, several genome annotations were developed such as the UCSC Genome, NCBI Map viewer and Ensembl Genome browser (Hubbard et al, 2002; Kent et al, 2002; Wheeler et al, 2001). Today, these tools are still in use when analysing sequencing data, and have evolved over time. BACs are no longer in use and in the following years several companies launched their sequencing platforms starting with Roche/45 FLX in 2004, followed by the Illumina/ Solexa Genome Analyzer in 2006, Applied Biosystems (ABi) in 2007 (Table 1-3). Developed initially for small genome analysis they were soon competing with microarrays and other sequencing technologies, giving a broader range of use which went beyond simple sequencing, such as ChIP-Seq (Chromatin Immunoprecipitation Sequencing) to look at genome-wide measurements of protein-DNA interactions, CNV-Seq (Copy number variation) and RNA-Seq. The latter gives the possibility not only to look at the expression levels of transcripts within a cell, but also splicing events, alleles expression and fusion events.

Table 1-3 Comparison between various sequencing platforms.

	Illumina HiSeq	Illumina MiSeq	454	PacBio	Ion Torrent (PGM)
PCR / Single Molecule	Bridge PCR	Bridge PCR	Emulsion PCR	Single molecule	Emulsion PCR
Read length	50/100 bp	50/150 bp	Mode 500 bp	Mean 2500 bp, 95 percentile is at 5500 bp	100/200 bp
Single read (SR) / Paired end (PE)	SR and PE	SR and PE	SR and PE	Circular	SR
Multiplexing in one run	8 lanes and multiplexing within each lane	1 lane with multiplexing possible	2, 4 or 8 lanes, up to 150 barcodes for multiplexing per lane	Probably possible, not yet released	32 samples with standard kit
Yield/time (representative example) k: kilo, M: mega, G: giga	1G frags @ 100 bp PE in 10 days	5M frags @ 150 bp PE in 1 day	1M frags @500 bp SE in 10 hrs	75k frags @ 2500 bp SE or 500 bp circular in 1.5 hours	3M frags (318 chip) @ 200 bp SR in 5 hours
Announced improvements	HighSeq 2500 machine soon to be launched by Illumina	15M frags @ 250 bp PE in 2 days - expected from Q3 2012	1M reads @700 bp SE in 24 hrs, in testing @ NSC	None at the moment	3M frags @ 200bp PE - during 2012
Raw base accuracy (approximate)	99.99% falling to 99% at 100th		99.5%, falling to 98.9 % @ 500th base	87% single pass, or 99.9% circular consensus	99.8% falling to 99% @100th base
Problems caused by GC content	Possible bias in bridge PCR	No data, but likely to be the same as for the HiSeq	Some bias against high GC	None	?
Homopolymer indel problems	minor problem	minor problem	problem	minor problem	problem
Typical applications	Re-sequencing: Variant calling (whole genome and whole exome), gene expression, chipSeq		De novo: genome and transcriptome assembly, amplicons	De novo: genome assembly, amplicons, SNP validation	Bacterial genome sequencing, amplicon sequencing
Status at NSC	**In production**	In testing (1 test run performed)	**In production**	In testing	In testing (4 test runs performed)

http://www.sequencing.uio.no

1.6.3 Advantages of RNA-Seq over other transcriptome methodologies

Whole transcriptome analysis by total RNA-Seq is used to analyse gene and transcripts expression changes in both coding and non-coding RNA species. In addition, RNA-Seq can be used for transcript discovery and annotation, allele specific expression (SNPs or mutations), mutation discovery and fusion-gene detection. Identifying patterns of gene expression and molecular pathways, in particular for certain tumors, can enable tailored therapies to be developed, based on the transcriptome profile. RNA-Seq analyses the expression level of mRNAs, thus the different transcripts and isoforms of a protein.

RNA-Seq seems to advance on microarray technology. Microarrays are less sensitive; they give information only about the selected probes and being based on hybridization, they are prone to generate non-specific signals, part of the reasons why other technologies such as DNA and RNA-Seq were developed. To overcome the limitations of microarrays, initially several groups developed other methodologies such as serial analysis of gene expression (SAGE) where tagged cDNA fragments are cloned and sequenced. SAGE relies on the tag being present uniquely within the transcript of interest, allowing easy identification of the target. The tag sequences are ideally linked together to form serial long molecules and then sequenced: the number of times the tag is sequenced would provide an estimate of the transcript abundance (Velculescu et al, 1997; Zhang et al, 1997).

RNA-Seq presents considerable advantages also against SAGE as well as whole genome sequencing. First of all, while the whole genome is known to be constant certain conditions might have an effect on gene expression (drugs for example), thus only RNA-Seq might reveal these differences. Secondly, isoforms and fusion transcripts, together with RNA editing, can only be observed if the RNA is studied since these are not characteristic of the DNA molecule. However, modifications occurring at the transcript level do not allow mapping back to the original sequence, which can cause issues with downstream analysis.

Among the various available platforms, RNA-Seq on the Illumina platform is highly reproducible, with virtually no need for technical replicates (Marioni et al, 2008); biological replicates will be however required to confirm the robustness of the results and to quantify noise in the data for biologically relevant signals and be ascertained. In addition, fold changes observed in RNA-Seq are similar to those observed in arrays, validating the new methodology (Marioni et al, 2008).

1.7 Aims of the project

1. To the test the hypothesis that specific transcription factors drive the interaction between enhancer and promoter at the mouse immunoglobulin lambda light chain locus:
    - To identify via bioinformatic searches B cell-specific transcription factors that bind the mouse immunoglobulin lambda light chain locus;
    - To test if any of the above transcription factors bind any of the promoters analysed, notably Jλ1, Jλ2, and Jλ3, via DNase I footprinting and EMSA. Particular focus will be given to E47;
    - To test the hypothesis that the promoter-bound E47 interacts with the enhancer-bound IRF4 to drive enhancer/promoter looping, triggering sterile transcription through the J lambda gene segments;
    - To test if particular residues are crucial for this interaction via mutagenesis studies.
2. To analyse RNA-Seq data from seven HPV-positive tumour biopsies and three HPV-positive cell lines to identify possible biomarkers for cervical cancer:
    - To identify differentially expressed cell surface transcripts between the samples analysed;
    - To generate recombinant adenoviruses expressing either the HPV16 E6 or E7 oncoproteins to be used to test any putative candidate as biomarker;
    - To test whether one of the differentially-expressed cellular transcripts, connexin 26, might be considered a putative biomarker by studying expression levels in normal HaCaT keratinocytes cells transduced with adenovirus expressing HPV16 E6 or E7 oncoproteins.

Chapter 2. MATERIALS AND METHODS

2.1 Chemicals and equipment

0.2 cm electroporation cuvettes	Cat.Z7006086-50EA, Sigma-Aldrich (Dorset, UK).
100mM dNTPs	Cat.10297018, Invitrogen (Paisley, UK).
1kb Plus DNA ladder	Cat.10787-018, Invitrogen (Paisley, UK).
2' deoxygalactose	Cat.D4407, Sigma-Aldrich (Dorset, UK).
2-log DNA	Cat.N3200S, NEB (Hertfordshire, UK).
6 well dishes	Cat. BC010, Appleton Woods (Birmingham, UK)
Acrylamide 30%	Cat.20210010, Thistle Scientific (Uddigston Glasgow, UK)
Agar bacteriological	Cat.LP0011, Oxoid Ltd (Hampshire, UK).
Agarose electrophoresis grade	Cat.MB1200, Melford (Ipswich UK).
Ampicillin	Cat.A5354, Sigma-Aldrich (Dorset, UK).
BioRad DC Protein assay (B)	Cat.500-0114, BioRad (Hertfordshire, UK).
BioRad DC Protein assay (C)	Cat.500-0113, BioRad (Hertfordshire, UK).
BioRad DC Protein assay (S)	Cat.500-0115, BioRad (Hertfordshire, UK).
BSA	Cat.A4503, Sigma-Aldrich (Dorset, UK).
Cell scrapers	Cat.BC323, Appleton Woods (Birmingham, UK).
CellBIND® flasks	Corning (Corning, USA).
Cesium chloride	Cat.15507-023, Invitrogen (Paisley, UK).
Chloramphenicol	Cat.C0378, Sigma-Aldrich (Dorset, UK)
Custom primers	Life Technologies (Paisley, UK).
DAPI	Cat.D9542, Sigma -Aldrich (Dorset, UK).
D-biotin	Cat. L8912, Sigma-Aldrich (Dorset, UK).
D-galactose	Cat.1287700, Sigma-Aldrich (Dorset, UK).
DH5α cells	Cat.18263-012, Invitrogen (Paisley, UK).

Dialysis tubing	Cat.D100, Biodesign Inc.(Carmel, NY, USA).
DMEM	Cat. D6546, Sigma -Aldrich (Dorset, UK).
DNase/RNase-free water	Cat.10977035, Invitrogen (Paisley, UK).
ECL Western blot detection reagents	Cat.RPN2235, GE Healthcare (Buckinghamshire, UK).
Expand HiFi DNA Polymerase	Cat. 11732650001, Roche (Burgess Hill, UK).
Foetal calf serum	GE Healthcare (Buckinghamshire, UK).
Formalin	Cat.HT501128-42, Sigma-Aldrich (Dorset, UK).
Glass capillaries tubing	Cat. 640766, OD 1.0 nm, ID 0.5 mm, G-100-4
Glycogen	Cat.FQ-R0561, Thermo Scientific Pierce (Rockford, USA)
Halt proteinase inhibitor cocktail	Cat.78429, Thermo Scientific Pierce (Rockford, USA)
IPTG	Cat.MB1008, Melford (Ipswich, UK).
L-glutamine solution	Cat. 67513, Sigma-Aldrich (Dorset, UK).
Lipofectamine 2000	Cat.11668027, Life Technologies (Paisley, UK).
L-Leucine	Cat.L8000, Sigma-Aldrich (Dorset, UK).
Lucifer yellow CH dilithium salt	Cat.L0259, Sigma -Aldrich (Dorset, UK).
Magic marker	Cat.LC5602, Invitrogen (Paisley, UK).
Mini slide-A-lyzer	Cat.69550, Thermo Scientific Pierce (Rockford, USA).
Nitrocellulose membrane 0.45 µM	Cat.1620094, BioRad (Hertfordshire, UK).
NucleoBond BAC 100 DNA	Cat.N2740579, Thermo Scientific Pierce (Rockford, USA)
Optimem-1	Cat.11058021, Invitrogen (Paisley, UK).
PBS tablets	Cat.BR0014, Oxoid Ltd (Hampshire, UK).
PCR and gel extraction kits	Qiagen (Hilden, Germany).
Penicillin/Streptomycin	Cat.4458, Sigma-Aldrich (Dorset, UK).
Petri dishes 92X16 mm	Cat. 821473, CAT. 62.547.254, Sarstedt (Leices-

	ter, UK)
Phenol-Chlorophorm	Cat.77617, Sigma-Aldrich (Dorset, UK).
Phusion DNA polymerase	NEB (Hertfordshire, UK).
Plasmid purification kits	Qiagen (Hilden, Germany).
Prestained protein ladder	Cat.10748-010, Life Technologies (Paisley, UK).
Protein G agarose	Cat. 20398, Thermo Scientific Pierce (Rockford, USA)
Proteinase inhibitor tablets	Cat.88665, Thermo Scientific Pierce (Rockford, USA)
Proteinase K	Cat.P2308, Sigma-Aldrich (Dorset, UK).
PVDF microporous membrane 0.45 µM	Cat.IPVH00010, Merk (New York, US)
Restriction enzymes	NEB (Hertfordshire, UK).
Riboruler HR RNA ladder	Cat.SM1821, Thermo Scientific Pierce (Rockford, USA).
RPMI-1640	Cat.R0083, Sigma -Aldrich (Dorset, UK).
RQ1 RNase-free DNase	Cat.M6101, Promoega (Madison, USA)
Subcellular proteome extraction	Cat.539790-1KIT, Calbiochem (Darmstadt, Germany).
SYBR green	Cat.59430, SLS (East Riding of Yorkshire, UK)
T4 polynucleotide kinase	NEB (Hertfordshire, UK).
Taq DNA polymerases	NEB (Hertfordshire, UK).
Thick blot paper	Cat.1703960, BioRad (Hertfordshire, UK).
Transcriptor first strand cDNA Synthesis	Cat.048968-66001, Roche (Burgess Hill, UK).
Transfection reagents	Cat.11668030, Life technologies (Paisley, UK).
Trizol	Cat.15596-018, Invitrogen (Paisley, UK).
Trypsin-EDTA	Cat.25200056, Invitrogen (Paisley, UK).
Ultraclear centrifuge tubes, 13x51 mm	Cat.344060, Beckman Coulter VWR (High Wy-

	combe, UK)
Ultraclear centrifuge tubes, 14x95 mm	Cat.344057, Beckman Coulter VWR (High Wycombe, UK)
Vectashield mounting medium	Cat.H1000, Vector Laboratories (California, USA)
X-gal	Cat.MB1001, Melford (Ipswich, UK).
Zero Blunt cloning kit	Cat.K275040, Invitrogen (Paisley, UK).

All other chemicals used to prepare buffers and solutions, unless stated, were purchased from Sigma-Aldrich (Dorset, UK).

2.2 Buffers and solutions

2.2.1 Agarose gel electrophoresis

- 10x TAE buffer, 1 l: 48.4 g Tris base, 17.4 M glacial acetic acid, 3.7 g EDTA, disodium salt.
- 10x TBE buffer, 1 l : 108 g Tris base, 55 g Boric acid, 40 ml 0.5 M EDTA (pH 8.0).
- 10x Orange Gel Loading Buffer (20 ml): 5 g Ficoll 400, 2 ml 1 M Tris-HCl (pH 7.4), 4 ml 1 M EDTA (pH 8.0), 0.04 g Orange G.
- DNA loading buffer: 2.5% ficoll, 0.025% bromophenol blue, 0.025% xylene cyanol.
- Ethidium bromide solution: 5g/ml in TE buffer.
- SYBR green: 1/10 000 dilution of stock solution in water.
- TE: 1 mM EDTA, 10 mM Tris-HCl pH 7.5.
- PBS: 1 tablet/500 ml sterile water. Autoclaved before use.

2.2.2 Protein extraction buffers

- Buffer A: 10 mM Hepes-KOH pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM DTT.

- Buffer C: 20 mM Hepes-KOH pH 7.9, 25% glycerol, 0.42 M NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 0.5 mM PMSF, 0.5 mM DTT.
- Buffer D: 20 mM Hepes- KOH pH 7.9, 20% glycerol, 0.1 M KCl, 0.2 mM EDTA, 0.5 mM DTT.
- Lysis buffer (whole cell extract) : 20 mM Hepes- KOH, 0.4 M KCl, 2 mM DTT, 20% glycerol, 0.5 mM PMSF (PMSF was prepared in isopropanol).

2.2.3 DNase I footprinting buffers

- Binding buffer: 10 mM Tris-HCl pH 7.5, 80 mM NaCl, 1 mM DTT, 1 mM EDTA and 5% glycerol.
- Formamide loading buffer: 80% formamide, 10 mM NaOH, 1 mM EDTA, 0.1% xylene cyanol, 1% bromophenol blue.
- Polyacrylamide sequencing gel: 7 M urea, 8% 19:1 acrylamide, 250 µl 10% APS, 50 µl TEMED, 5 ml 10x TBE.
- Stop buffer: 0.3 M sodium acetate pH 5.2, 0.1 mM EDTA, 100 µg/ml yeast tRNA.

2.2.4 Western blot and Co-Immunoprecipitation solutions

- 10x Tris-Gycine-SDS buffer (TGS): 25 mM Tris, 192 mM glycine, 0.1% SDS.
- 10x TBS: 60.6g Tris, 87.6g NaCl (1l, adjust pH to 7.5 with concentrated HCl), bring up the volume to 1L with ddH2O.
- Transfer buffer: 25mM Tris, 192 mM glycine, 20% (v/v) methanol, adjust pH to 8.3.
- TBST: 1x TBS in 0.1% Tween-20.
- TBST (for E6 antibody): 1x TBS in 0.1% Triton-X100
- Blocking solution: 5% dry milk in TBST (10% milk for Cx26).
- Blocking solution for HPV16 E6: 1x TBST (E6), 2.5% dry milk, 1% BSA.

- Stripping solution: 10 ml 20% SDS, 6.25 ml 1M Tris pH 6.8, 700 µl β-Mercaptoethanol, increase volume to 100 ml with ddH_2O.
- SDS-PAGE, resolving gel: 1.5 M Tris-HCl, pH 8.8, 10% APS, 10% SDS, TEMED, acrylamide (amount change depending on gel concentration).
- SDS-PAGE, stacking gel: 1 M Tris-HCl, pH 6.8, 10% APS, 10% SDS, TEMED, acrylamide (amount change depending on gel concentration).
- Co-Ip buffer : 10 mM Hepes-KOH, pH 7.9, 10 mM KCl, 1.5 mM MgCl2, 0.1% NP-40, 150 mM NaCl.
- Buffer E: 10 mM Hepes pH 7.9, 10 mM KCl, 1.5 mM $MgCl_2$, 0.1% NP40
- Buffer F: 10 mM Hepes pH 7.9.
- RIPA buffer: 50 mM Tris-HCl pH8, 150 mM NaCl, 1% NP-40, 0.5% sodium de-oxycholate, 0.1% SDS. Supplement with protease inhibitor.
- Supplemental lysis buffer to RIPA: 5% SDS, 0.15 M Tris-HCl, pH 6.7, 30% glycerol. Supplement with protease inhibitor.
- Lysis buffer for mass spectrometry samples: 50 mM Hepes-NaOH, pH 7.4 (4°C), 150 mM NaCl, 50 mM NaF, 1 mM EDTA, 0.1-0.5% NP40.
- 2x SDS sample buffer: 100mM Tris-HCl (pH 6.8), 200mM DTT, 4% SDS, 20% glycerol, 0.2% bromophenol blue.

2.2.5 Buffers for bacterial and virus experiments

- DNase I solution: 20 mM Tris-HCl (pH 7.4), 50 mM NaCl, 1 mM DTT, 0.1 mg/ml BSA, 10 mg/ml bovine pancreatic DNase I, 50% glycerol.
- CsCl gradient solutions for virus purification:
    - 1.5d – 90.8 g CsCl, 109.2 g 10 mM Tris-HCl pH 8.0.
    - 1.35d – 70.4 g CsCl, 129.6 g 10 mM Tris-HCl pH 8.0.
    - 1.25d – 54.0 g CsCl, 146.0 g 10 mM Tris-HCl pH 8.0.
- Luria-Bertani (LB) medium: 10 g/l tryptone, 5 g/l yeast extract, 10 g/l salt, volume was adjusted to 1L with ddH_2O.

- Low salt LB: 10g/L tryptone, 5 g/L yeast extract, 5 g/l salt. Volume was adjusted to 1 l with ddH_2O.
- LB-Agar: 10 g/l tryptone, 5 g/l yeast extract, 10g/l salt, 15g/l agar. Volume was adjusted to 1 l with ddH_2O.
- M63 minimal media (5X): 10 g/l $(NH_4)_2SO_4$, 68 g/l, 2.5 mg/l $FeSO_4.7H_20$ volume was adjusted to 1 l with ddH_2O and pH to 7 with KOH.
- M9 salts (2X): 25.6 g/l Na_2HPO_4, 6 g/l KH_2PO_4, 2 g/l NH_4Cl, 1 g/l NaCl. Volume was adjusted to 1 l with ddH_2O.
- GalK positive selection agar plates: 1X M63 medium, 1 mM $MgS0_4$, 1 mg/l D-biotin , 45 mg/l L-leucine, 0.2% galactose, 12.5 µg/ml Chloramphenicol, 15 g/l agar dissolved in SQ H_2O.
- GalK negative selection agar plates: 1X M63 medium, 1mM $MgS0_4$, 1 mg/l D-biotin, 45 mg/l L-leucine, 0.2% 2´-deoxgalactose, 0.2% glycerol, 12.5 µg/ml chloramphenicol, 15 g/l agar dissolved in SQ H_2O.
- AdZ selection cassette plates: 10 g/l tryptone, 5 g/l Yeast extract, 50 g/l Sucrose,12.5 µg/ml Chloramphenicol, 80 µg/ml X-gal, 0.2mM IPTG, 15 g/l agar dissolved in SQ H_2O.
- Antibiotic concentrations used: Ampicillin (50 µg/µl), Chloramphenicol (12.5 mg/ml), X-gal (1:500 of 40 mg/ml stock), IPTG (1:500 of 100 mM stock).

2.2.6 Calcium phosphate transfection reagents

- 10x Hepes-buffered saline (HBS): 8.18% (w/v) NaCl, 5.94% (w/v) Hepes, 0.2% (w/v) Na_2HPO_4. A working solution of 2x HBS was made; the pH was adjusted to 7.12 with 1 M NaOH and stored at 4°C.
- 2 M $CaCl_2$: this was filter-sterilised and stored at 4°C.
- 15% glycerol/HBS: 30 ml 50% glycerol (w/v), 50 ml 2x HBS, (pH 7.12), 20 ml ddH_2O. This was filter-sterilised and stored at 4˚C.

2.3 Cell lines

- 122-1 cell line is a mouse pro-B cell line transformed with A-MuLV and was used to make cell extracts for both DNase I footprinting and EMSA experiments. These cells were cultured in RPMI 1640 medium supplemented with 10% heat inactivated FCS (foetal calf serum), 5,000 U/ml penicillin, 5,000 µg/ml streptomycin, 2 mM L-glutamine and 0.05 mM 2-β-mercaptoethanol and maintained at 37°C, in a humidified atmosphere containing 5% CO_2.
- Cos-7 cells are a simian virus 40-transformed African green monkey epithelial cell line. They were used for transfection experiments to study enhancer-promoter interactions (Jensen et al, 1964).
- CaSki and SiHa cells are human cervical cancer cell lines containing integrated copies of HPV16 (Friedl et al, 1970; Pattillo et al, 1977) while HeLa is an HPV18 positive cervical cancer adenocarcinoma cell line (Scherer et al, 1953).
- 911 is a human retinoblastoma cell line used to amplify the recombinant adenovirus. The cell line provides the Adenovirus 5 E1 region *in trans*, necessary for adenovirus replication (Fallaux et al, 1996).
- Normal human foreskin keratinocytes (NHK) were a kind gift of Ms Rosie Doble, (University of Leeds, Leeds, UK). These were cultured in keratinocyte growth medium (Promo Cell) and maintained at 37°C, in a humidified atmosphere containing 5% CO_2.
- HaCaT is an immortalised human keratinocyte cell line (Boukamp et al, 1988). These cells were transduced with the recombinant adenovirus to investigate the expression of Cx26. Extracts derived from HaCaT cell lines in which hDlg and hScribble were ablated were kind gifts of Dr Paola Massimi, ICGEB, Trieste, Italy.

122-1 pro-B cells and Cos-7 cells were cultured in RPMI-1640 using the same conditions as described above. CaSki, SiHa, HeLa, 911 and HaCaT cell lines were grown in

DMEM supplemented with 10% heat-inactivated FCS, 5,000 U/ml penicillin, 5,000 µg/ml streptomycin, 2 mM L-glutamine and maintained at 37°C, in a humidified atmosphere containing 5% CO_2. The HaCaT cell lines in which hdlg and Scribble were ablated were maintained at 37°C in a humidified atmosphere containing 10% CO_2. The identities of CaSki, SiHa, HeLa and HaCat cells were established by STR (finger-printing) analysis (Dr Claire Taylor, CRUK Genomics facility, SJUH, Leeds).

2.4 Bacterial strains

- *E. coli* DH5α: non-expression vector used for propagation and storage of plasmids.
- DY380 cells: these cells contained the 3'-half of the duplicated lambda locus which was used for V(D)J recombination studies.
- SW102: DY380 cells previously modified to contain a full galactose operon, which have had a deletion of the *galK* gene. Genotype: as DY380 plus *gal+ΔgalK.* These were used for recombineering studies and were a kind gift of Dr James Findlay, University of Leeds, UK.

2.5 Adenovirus constructs

- AdZ-5F35-CV5: E1-and-E3-deleted Ad5 vector with a selection cassette in place of the E1 region under the control of a CMV promoter with the Ad5 fibre replaced with the Ad35 fibre. A kind gift of Dr James Findlay, University of Leeds, Leeds, UK.
- AdZ-5F35-ECFP: an AdZ-5F35-CV5 vector with an ECFP transgene under the control of the CMV promoter in place of the E3 region. Ampicillin and chloramphenicol resistant. This construct was made in this study.
- AdZ-5F35-ECFP/E6: an AdZ-5F35-CV5 vector with an ECFP transgene in place of the E3 region and HPV16 E6 N-terminal HA-tagged ORF in place of the E1

region, both under control of the CMV promoter. Chloramphenicol resistant. This construct was made in this study.

- AdZ-5F35-ECFP/E7: an AdZ-5F35-CV5 vector with an ECFP transgene in place of the E3 region and HPV16 E7 C-terminal Myc-tagged ORF in place of the E1 region, both under control of the CMV promoter. Chloramphenicol resistant. This construct was made in this study.

2.6 Plasmids

- pUC19: plasmid cloning vector (New England Biolabs, Hertfordshire, UK).
- pBluescriptK+: plasmid cloning vector used to clone the mutants constructs used in Chapter 4.
- pEF3-99 (pEF-HPV16-W12E): plasmid containing the entire HPV16 genome, including the E6 and E7 open reading frames (ORFs) (Flores et al, 1999).
- pGL3-Basic: promoter-less firefly luciferase reporter gene. A kind gift of Dr Ian Wood.
- pCMV-RL: *Renilla* luciferase reporter gene under the control of the Cytomegalovirus promoter (CMV). A kind gift of Dr Ian Wood.
- pGL3/Jλ1: pGL3-Basic vector containing the Jλ1 promoter within the Xho I and HindIII restriction sites. This construct was made in this study.
- pGL3/Jλ1/Eλ3-1: pGL3-Basic vector containing the Jλ1 promoter within the XhoI/HindIII restriction sites and the Eλ3-1 enhancer within the SalI/BamHI restriction sites. This construct was made in this study.
- pCS2Myc/E2A_Myc: plasmid containing the E2A ORF N-terminallly Myc tagged. A kind gift of Dr Sarah Bevington.
- pEF-cx/E2A_Myc: plasmid containing the E2A ORF N-terminallly Myc tagged. This construct was made in this study and used in mutational studies described in Chapter 4.

- pCMV/IRF4: plasmid containing the IRF4 ORF. A kind gift of Dr Sarah Bevington.
- pEF-PU.1_Flag: plasmid containing the PU.1 ORF N-terminallly Flag tagged. A kind gift of Dr Sarah Bevington.

2.7 Protein methods

2.7.1 Preparation of nuclear extracts

Nuclear extracts were prepared according to the method of Dignam et al (Dignam et al, 1983). 122-1 pro-B cells (to reach up to 20-30 pellets from T175 flasks) were harvested by centrifugation at 2000 rpm for 10 minutes at 4°C and washed twice with ice-cold PBS to remove the medium. The packed cell volume (PCV) of the pellet was measured. It was then re-suspended in five PCVs of hypotonic buffer A and the cells were allowed to swell on ice for 20 minutes. The cell lysate was centrifuged at 2000 rpm for 10 minutes at 4°C and the pellet was re-suspended in two PCV of buffer A. The cells were homogenised using a Dounce homogeniser with a type B (loose) pestle with 10 strokes to lyse the cells. This was checked using a phase contrast microscope and the nuclei were recovered by centrifugation in a small Sorvall plastic centrifuge tube at 2500 rpm, at 4˚C for 10 minutes. The nuclear pellet was re-suspended in 1.3 volumes buffer C (to remove proteins bound to DNA) by homogenisation in a Dounce homogeniser type C (tight) pestle (10-20 strokes) to lyse nuclei. The nuclear homogenate was transferred into a beaker and stirred slowly in a cold room for 30 minutes. The homogenate was centrifuged at 15000 rpm at 4°C for 30 minutes and the supernatant was transferred into a fresh sterile beaker. Ground $(NH_4)_2SO_4$ (0.33 g/ml) was added slowly to precipitate proteins, followed by neutralization of the pH with 40 µl of 1 M NaOH per g $(NH_4)2SO_4$. The extract was left stirring in the cold room for 1 hour to “salt out” the proteins. The extract was centrifuged at 15000 rpm for 30 minutes at 4°C and the pellet was re-suspended in 0.1 volumes of buffer D. The nuclear extract was dialysed against two changes of

buffer D for 1.5 hours. Following a final centrifugation at 10000 rpm for 10 minutes at 4°C, the extract was divided into aliquots in siliconised Eppendorf tubes and stored at -80°C. Protein concentration was determined by the Bradford assay (Bio-Rad) according to the manufacturer's instructions. Approximately 100 µg of extract were used per lane in footprinting reactions.

2.7.2 Preparation of whole cell extracts

Whole cell extracts were prepared using the method developed by Kumar and Chambon (Kumar & Chambon, 1988). The cells were harvested by centrifugation for 6 minutes at 1800 rpm. The cell pellet was washed twice in 50 ml ice cold PBS to remove traces of the medium and re-suspended in 200 µl lysis buffer per $2x10^7$ cells supplemented with protease inhibitors (Roche, complete cocktail). The cell suspension was immediately placed on dry ice and allowed to thaw on ice to break cell membranes. The cell lysate was centrifuged at 14000 rpm in a bench top centrifuge for 15 minutes at 4˚C and the supernatant divided into aliquots and stored at -80°C. Protein concentrations were determined by the Bradford assay according to the manufacturer's instructions and approximately 20-30 µg protein was used per lane in EMSA experiments.

2.7.3 Co-Immunoprecipitation assays

Co-Immunoprecipitation (Co-Ip) assays were performed to investigate protein-protein interactions *in vitro*.

To investigate interactions between E47, IRF4 and PU.1, 122-1 pro-B cells ($1x10^8$ cells) were centrifuged at 1800 rpm for 2 minutes and the pellet was washed in 500 µl ice-cold PBS. The pellet was re-suspended in buffer E, supplemented with protease inhibitors (Roche, complete cocktail). An 18 µl aliquot of 5M NaCl was added and the sample vortexed for about 5-10 seconds. Following addition of 25 µl 40% glycerol, the samples were left rotating for 20 minutes in the cold room. The cell lysate was centrifuged for 15 minutes at 14000 rpm in a bench top centrifuge at 4˚C

and the supernatant removed. Following addition of 300 µl of buffer F and 15 µl protein G beads to pre-clear the lysate, the samples were left rotating in the cold room for 20 minutes. Following a brief centrifugation at 2400 rpm for 2 minutes, the clarified supernatant was divided into aliquots in siliconised Eppendorfs and 1/10 of the volume was retained as input. Antibodies against E47 or IRF4 (5 µl) were added and the samples were left rotating overnight in the cold room. Following addition of 20 µl protein G beads, the samples were rotated for 2 hours at 4°C and the immunoprecipitation complex was washed by centrifugation three times with Co-Ip buffer. The pellets were re-suspended in protein loading buffer, boiled for 2 minutes, centrifuged for 10 seconds and the supernatant loaded on a 10% SDS polyacrylamide gel. The gel was subjected to electrophoresis at 125V for 2 hours in 1X TGS. To prevent E47 degradation, two inhibitors were used: MEK inhibitor and MG132 (a proteasome inhibitor), each at a final concentration of 10 µM. The MEK inhibitor was added 3 hours prior to harvesting the cells and MG132 was added 6 to 8 hours prior harvesting. Both inhibitors were also added to buffers E and F.

2.7.4 Sample preparation for mass spectrometry

HaCaT cells were transduced as described in 2.9.7 and samples prepared for mass spectrometry to confirm the expression of HPV16 E6, due to the difficulty found in Western blots. The reagents and protocol was kindly provided by Dr Lawrence Banks. In brief, the HA-conjugated beads (EZnew Red αHa affinity gel, Sigma SLBG7153) were prepared by washing them in the lysis buffer and left rotating at 4˚C. After 45 minutes the antibody-conjugated beads were washed twice in 1ml lysis buffer and left in ice until use. HaCaT cells (confluent 10 cm dishes) were washed twice with cold PBS and lysed with 0.5 ml of lysis buffer supplemented with 0.1 mg/ml dextran, directly on the dish, on ice. Cells were scraped, the lysate was transferred into a cold microcentrifuge tube and centrifuged for 10 minutes at 13,000 rpm. An aliquot was removed to be kept as input to run a co-immunoprecipitation parallel to the mass spectrometry analysis. The remaining lysate was added to the

washed beads and incubated for 3 hours at 4˚C. The complex was washed three times with the lysis buffer (without dextran) and re-suspended in the same lysis buffer and transferred into a new cold fresh micro centrifuge tube. Because the samples were used for co-immunoprecipitation, to improve the stringency of the IP, a glycerol step gradient was performed. In brief, an aliquot of the lysis buffer was supplemented with glycerol to 20% and 200 µl was transferred at the bottom of a microcentrifuge tube. The beads were placed on top of the lysis buffer-glycerol mix and centrifuged for 30 seconds at 13,000 rpm. The supernatant was carefully removed and the beads washed twice with 1 ml of lysis buffer. The pellet was washed three times with PBS and samples split equally for immunoprecipitation and mass spectrometry, respectively. Samples were kept at -80˚C until further use.

2.7.5 Determination of protein concentration

Protein concentration was determined using the BioRad DC assay system (Bio-Rad Inc.) according to the manufacturer‘s instructions. In brief, 20 µl of reagent S were added to each ml of reagent A (this mix will be called A'). The required dilutions of a protein standard (BSA, range 0.2-1 mg/ml) were prepared and 100 µl of standards and samples were pipetted into a clean test tubes. 500 µl of reagent A' was added to each of the tube and quickly vortexed, followed by addition of 4.0 ml reagent B into each test tube and vortexed immediately. Following an incubation of 15 minutes, absorbances were read at 750 nm.

2.7.6 Western blot analysis

Cells of interest were grown in Petri dishes or flasks depending on the experiments. The medium was discarded and cells washed twice with cold PBS prior to lysis in RIPA buffer. The lysate was left on ice for 30 minutes, centrifuged for 10 minutes at 13,000 rpm and the supernatant loaded for Western blot analysis. Where required, protein concentration was determined using either the Bradford assay as described in above or a Nanodrop UV Spectrophotometer with absorbance set at Abs_{280}

(Thermo Scientific, Nanodrop 1000). Following electrophoresis, proteins were transferred to polyvinylidine fluoride (PVDF) membrane that had been pre-soaked in 100% methanol and equilibrated in transfer buffer, except for detection of HPV16 E7 and Westerns from the cell fractionation assays where proteins were blotted on 0.22µ nitrocellulose membranes. The proteins were transferred for 1 hour at 0.068 mA/cm^2 or at $5mA/cm^2$ using the Biometra semidry blotting apparatus (Biometra Fastblot B33/34). For detection of proteins following the cell fractionation assay, wet transfer of gels was with the following settings: 250 mA constant current for 3.5 hours or overnight at 120 mA at room temperature with constant stirring. For detection of HPV16 E7, following transfer, nitrocellulose membranes were fixed in 0.5% glutaraldehyde in 1XTBS for 30 minutes at room temperature. The membrane was washed extensively with ddH_2O before blocking. Membranes were blocked for 1 hour at room temperature with 5% non-fat dried milk (Marvel) in TBST or in 10% milk in TBST for 30 minutes at 37˚C for proteins derived from the cell fractionation assay, and incubated with the primary antibody for 1 hour at room temperature or overnight in the cold room (Table 2-2). The membrane was washed in TBST for 1 hour and incubated with the secondary antibody for 1 hour at room temperature or overnight in the cold room. Following another wash for 1 hour in TBST, the membranes were exposed to a chemiluminescent detection reagent according to the manufacturer's instructions, followed by autoradiography or detection using the Raytek Fujifilm LAS-3000 Imager, according to the manufacturer's instructions.

Table 2-1 List of primary antibodies used in this study.

Antibody	Specificity	Species and Isotype	Source
N-649	Mouse E47	Rabbit polyclonal IgG	Santa Cruz
G-2	Mouse E2A	Mouse monoclonal IgG_1	Santa Cruz
N-18	Mouse IRF4	Goat polyclonal IgG	Santa Cruz
B-9 Spi	Mouse PU.1	Mouse monoclonal IgG_1	Santa Cruz
NM2	HPV16 E7	Mouse monoclonal IgG_{2a}	Santa Cruz
HA-7	HA-tag	Mouse monoclonal IgG_1	Sigma
sc-807	Flag-tag	Rabbit polyclonal IgG	Santa Cruz
BS1715-R	Human Cx26	Rabbit polyclonal IgG	Bioss
Z-Z8	Human Cx26	Rabbit polyclonal IgG	Zymed
CX-1E8	Human Cx26	Mouse monoclonal IgG2a-kappa	Zymed
C65	Human GAPDH	Mouse monoclonal IgG_1	Calbiochem
GTU-99	Human γ-tubulin	Mouse monoclonal IgG_1	Sigma
TUB2.1	Human β-tubulin	Mouse monoclonal IgG_1	Sigma
H-2	Human α-actinin	Mouse monoclonal IgG_1	Santa Cruz

Table 2-2 Antibodies concentration for Western blots (WB) or immunofluorescence (IF).

Primary antibody	Dilution	Secondary antibody	Dilution
E47	1:500 (WB)	anti-rabbit	1:10000
IRF4	1:500 (WB)	anti-goat	1:10000
PU.1	1:500 (WB)	anti-mouse	1:10000
HPV16 E7	1:50 (IF) 1:100 (WB)	anti-mouse anti-mouse	1:1000 1:10000
HA-tag	1:200 (IF)	anti-mouse	1:1000
Cx26 (Bioss) Cx26 (Zymed) Cx26 (Zymed)	1:200 (IF) 1:200 (WB) 1:50 (WB)	anti-rabbit anti-rabbit anti-mouse	1:3000 1:10000 1:10000
GAPDH	1:5000 (WB)	anti-mouse	1:2000
Tubulin	1:5000 (WB)	anti-mouse	1:10000
α-actinin	1:1000 (WB)	anti-rabbit	1:10000
β-galacotsidase	1:2000 (WB)	anti-mouse	1:10000

2.7.7 Immunofluorescence microscopy

HaCaT cells ($1x10^5$) were grown on glass coverslips for 24 hours, transduced with the recombinant adenovirus as described in Chapter 2.9.7 and incubated for 24

hours to allow expression of viral proteins. Cells were washed three times with PBS and fixed with 10% formalin for 10 minutes at room temperature. Cells were washed three times with PBS and treated for 1 hour at room temperature with 2% fish skin gelatin (Sigma, a kind gift of Gareth Howell) to block non-specific protein binding. The cells were washed with PBS and incubated overnight with the primary antibody in 0.1% Triton in PBS. Coverslips were washed with PBS and incubated with the secondary antibody the species-specific Alexa Fluor-labelled secondary antibody for 1 hour at room temperature, in the dark. The secondary antibody was removed with three washes of PBS and cells were stained with DAPI (4-6-diamidino-2-phenylindole 1:5000 of 5 mg/ml) to stain the nuclei for at least 1 minute. Cells were washed again three times with PBS and left to dry before mounting with Vectashield hard-set mounting medium. The coverslips were viewed using an AxioPlan 2 Imaging Fluorescent microscope equipped with an AxioCam MR camera and images processed using the AxioVision 4.3 or Zen 2011 software.

Table 2-3 List of secondary antibodies used in this study.

Antibody	Specificity	Species	Label	Source
A-11070	Rabbit IgG	Goat	Alexa Fluor-488	Invitrogen
A-21206	Rabbit IgG	Donkey	Alexa Fluor-488	Invitrogen
A-11020	Mouse IgG	Goat	Alexa Fluor-594	Invitrogen
A-21203	Mouse IgG	Donkey	Alexa Fluor-594	Invitrogen
A-6782	Mouse IgG	Sheep	HRP	Sigma Aldrich
A-6154	Rabbit IgG	Goat	HRP	Sigma Aldrich

2.8 DNA methods

2.8.1 Standard PCR reactions

Amplification of DNA segments of interests was carried out using Taq Polymerase in a final volume of 25 or 50 µl. The following components were mixed in PCR tube: 1x Taq Buffer, 0.2 mM dNTPs, 0.1 µM primers, 0.625 U Taq DNA polymerase, DNA template, DEPC H_2O to 25 or 50 µl. PCR cycling conditions and primers used for all PCR reactions are listed in the following tables:

Table 2-4 Cycling conditions used for standard PCR reactions.

Temperature (˚C)	Time (seconds)	Cycles
95	300	1
95	30	35-40
55	30	
72	30/kb	
72	600	1

All PCR reactions were performed with heated lid at 110˚C prior to beginning of the PCR cycles.

Table 2-5 Cycling conditions used for Pfx DNA polymerase PCR reactions.

Temperature (˚C)	Time (seconds)	Cycles
98	300	1
98	30	35
55	30/kb	
72	60/kb	
72	300	1

Table 2-6 Cycling conditions used for site directed mutagenesis PCR reactions.

Temperature (˚C)	Time (seconds)	Cycles
95	60	1
95	50	18
60	50	
68	60/kb	
68	420	1

Table 2-7 Cycling conditions used for HiFi^{PLUS} PCR reactions.

Temperature (˚C)	Time (seconds)	Cycles
94	120	1
94	15	10
55	30	
72	60	
94	15	20
55	30	
72	60/kb ((+ 5 seconds/cycle)	
72	420	1

Table 2-8 Primers used to generate the recombinant adenoviruses.

DNA fragment	Primer sequence (5' to 3')
E3 GalK hr F	CTGCTAGTTGAGCGGGACAGGGGACCCTGTGTTCTCACTGTGATTTGCAACTG TCCTAACCTTGGATTACATCCTGTTGACAATTAATCATCGGCA
E3 GalK hr R	TAACTGATTTTAAGTAAGTGATGCTTTATTATTTTTTTTTATTAGTTAAAG
E7 hr F (CterMyc)	TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACCG GGACCGATCCAGCCTGGATCCACCATGCATGGAGATACACCTACAT
E7 hr R (CterMyc)	TTATTGAGTAGGATTACAGAGTATAACATAGAGTATAATATAGAGTATACAATA GTGACGTGGGATCCTTACAGATCTTCTTCAGAAATAAGTTTTTGTT
E6hr F (NterHA)	GTCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC GGGACCGATCCAGCCTGGATCCCCTACCATGTACCCATACGACGTC
E6 hr R(NterHA)	GCGTGACACGTTTATTGAGTAGGATTACAGAGTATAACATAGAGTATAATATA GAGTATACAATAGTGACGTGGGATCCTTACAGCTGGGTTTCTCTACG
ECFP hr F	CTGCTAGTTGAGCGGGACAGGGGACCCTGTGTTCTCACTGTGATTTGCAACTG TCCTAACCTTGGATTACATTAGTTATTAATAG TAATCAATTACGG
ECFP hr R	TGATTTTAAGTAAGTGATGCTTTATTATTTTTTTTTATTAGTTAAAGGGAATAAG ATCTTTGAGACACAGGGTTAAGATACATT GATGAGTTTGGAC

***hr** = homology region*

Table 2-9 Primers used to generate the PCR products for DNase I footprinting reactions.

DNA fragment	Primers (5' to 3')
J_λ1 BamHI F	CGGATCCGCTGATCCCAGAATTTATATCTTGTTAG
J_λ1 HindIII R	GGAAGCTTCCACCAGCTGTGTAAAGTCTATGC
J_λ2 BamHI F	CGGATCCCACTCAGCCTGTAAAATCAAGCC
J_λ2 HindIII R	GGAAGCTTTGTTCTCTCACAGCTGAGCTGAATCA
J_λ3 BamHI F	CCATTTGTTTCATAACTTCTGTTAGAGTCATTGT
J_λ3 HindIII R	GAGCTGTGACCTCACTTGAAGTG

Table 2-10 Primers used for standard PCR for amplification of genes of interest.

DNA fragment	Primer sequence (5' to 3')
$J_\lambda 1$ F $_{BamHI}$	CGGATCCGCAACACCAACACACATTTTTACCTGG
$J_\lambda 1$ R $_{HindIII}$	GGAAGCTTCCACCAGCTGTGTAAAGTCTATGC
$J_\lambda 2$ F $_{BamHI}$	CGGATCCCACTCAGCCTGTAAAATCAAGCC
$J_\lambda 2$ R $_{HindIII}$	GGAAGCTTGTTCTCTCACAGCTGAGCTGAATCA
$J_\lambda 3$ F $_{BamHI}$	CGGATCCCATTTGTTTCATAACTTCTGTTAGAGTCATTG
$J_\lambda 3$ R $_{HindIII}$	GGAAGCTTGAGCTGTCAGCTCACTTGAGTG
$E_\lambda 3$-1 PU.1 F	GAGAAATAATAGGAACTGCAACCAAGTCCATTAGCAGCAAGGC
$E_\lambda 3$-1 PU.1 R	GCCTTGCTGCTAATGGACTTGGTTGCAGTTCCTATTATTTCTC
$E_\lambda 3$-1 IRF4 F	GAGAAATAATAGGAACTGCAACCAAGTCCATTAGCAGCAAGGC
$E_\lambda 3$-1 IRF4 R	GCCTTGCTGCTAATGGACTTGGTTGCAGTTCCTATTATTTCTC
HPV16 E6 F	CACCAACCGAGAACTGCAATG
HPV16 E6 R	CAGCTGGGTTTCTCTACG
HPV16 E7 F	ATGCATGGAGATACACCTAC
HPV16 E7 R	TTATGGTTTCTGAGAACAGA
ECFP F	TAGTTATTAATAGTAATCAATTACGG
ECFP R	TAAGATACATTGATGAGTTTGGAC
HPV16 E6 N_ter HA F	ACCATGTACCCATACGACGTCCCAGACTACGCTATGCAC
HPV16 E6 N_ter HA R	TTACAGCTGGGTTTCTCTACGTGTTCTTGA
HPV16 E7 C_ter Myc F	ACCATGCATGGAGATACACCTACATTG
HPV16 E7 C_ter Myc R	TTACAGATCTTCTTCAGAAATAAGTTT TTGTTCTGGTTT
Cx26 F	TTCCTCCCGACGCAGAGCAA
Cx26 R	GGGCAATGCGTTAAACTGGC
GAPDH F	CTCAACTACATGGTTTACATGTTC
GAPDH R	GCAAATGAGCCCCAGCCTT

Table 2-11 Oligonucleotide sequences used for gel retardation assays.

Probe	Sequences (5' to 3')
E47 ($J_\lambda1$) F	GTGTTCTCCAACTGCTCTCTCCT
E47 ($J_\lambda1$) R	AGGAGAGAGCAGTTGGAGAACAC
E47($J_\lambda2$) F	ATCTACTACACAGCTGTCAG
E47($J_\lambda2$) R	CTGACAGCTGTGTAGTAGAT
E47($J_\lambda3$) F	TTCAGTGCAGCTGTGAGA
E47($J_\lambda3$) R	AAGTCACGTCGACACTCT
Pou1F1a ($J_\lambda1$) F	GATCCTTCAGTGTTTAGTATAATA
Pou1F1a ($J_\lambda1$) R	TATTATACTAAACACTGAAGGATC
IRF1/IRF2 ($J_\lambda1$) F	AGTTTTCTTCAATAACTAGTGAAGATAAAG
IRF1/IRF2 ($J_\lambda1$) R	CTTTATCTTCACTAGTTATTGAAGAAAACT

Table 2-12 Unlabelled competitors used for gel retardation assays.

Unlabelled competitor	Sequences (5' to 3')
AML F	GAATTCACCACATCTAGA
AML R	TCTAGATGTGGTGAATTC
EBF F	TCTAGAATTCCCNNGGGAATGAATTC
EBF R	GAATCCATTCCCNNGGGAATTCTAGA
Oct factor F	TCTAGAATGCAAATGAATTC
Oct factor R	GAATCCATTTGCATTCTAGA
E47 F	CAAACACCACCTGGGTAATC
E47 R	GATTACCCAGGTGGTGTTTG
E2A F	CCTTAGGCACATCTGTTGCTTTCG
E2A R	CGAAAGCAACAGATGTGCCTA AGG
PU.1/IRF4 F	CTTTGAGGAACTGAAAACAGAACCT
PU.1/IRF4 R	AGGTTCTGTTTTCAGTTCCTCAAAG

Table 2-13 Primers used for site directed mutagenesis.

DNA fragment	Primer sequence (5' to 3')
Jλ1 E47mut F	TGACTGTGTATGTGTTCTCTGACTGCTCTCTCCTGAAGTGCCT
Jλ1 E47mut R	AGGCACTTCAGGAGAGAGCAGTCAGAGAACACATACACAGTCA
Eλ3-1 /IRF4 mut F	GAGAAATAATAGGAACTGCAACCAAGTCCATTAGCAGCAAGGC
Eλ3-1/IRF4 mut R	GCCTTGCTGCTAATGGACTTGGTTGCAGTTCCTATTATTTCTC
Internal primer 1	AATTTCTCACCTGCCCCCTCAACGC
Internal primer 2	GCGTTGAGGGGGCAGGTGAG
E47 NheI primer	TGGCACCGCTAGCGATCTCCATGGGC
E47 Acc65I primer	GGTCGCGGCGCTTCCGGAGCAA
IRF4 L368Ala R	ATGGTGAGCAAACACTTGCGCCTCTGAAAACTGCTGTGT
IRF4 L368Ala F	ACACAGCAGTTTCTATCAGAGGCGCAAGTGTTTGCTCACCAT
IRF4 K399Ala F	GACCCTCAGAGACAGAGGGCGCTCATCACAGCTCAT
IRF4 K399Ala R	ATGAGCTGTGATGAGCGCCCTCTGTCTCTGAGGGTC
Eλ3-1 IRF4/PU.1 mut F	AAGAGAAATAATACCGACTGCAACCAAGTCCATTAGCAGCAAGGC
Eλ3-1 IRF4/PU.1 mut R	GCCTTGCTGCTAATGGACTTGGTTGCAGTCGGTATTATTTCTCTT

2.8.2 Recombineering PCR reactions

All primers used for recombineering studies are listed in Table 2-8.

To generate fragments to be used for recombineering reactions the following components and PCR conditions were used. The following components were mixed in a PCR tube: 1x Pfx Buffer, 0.3 mM dNTPs, 0.3 µM primers, 2 mM MgCl2, 1.25 units Pfx DNA polymerase, DNA template, DEPC H_2O to 25 µl. The PCR products were resolved by electrophoresis on a 1% agarose gel and the band of interest excised and gel purified using the Qiagen PCR purification kit according to the manufacturer's instructions. The settings in Table 2-5 were used to run the Pfx DNA polymerase PCR reactions. HiFi DNA Polymerase was used to generate the *GalK*, ECFP, HPV16 E6 or E7 fragments with the homology regions for recombineering reactions. The following component were mixed: 1x Expand HiFi^{PLUS}, 10x reaction buffer 2, 0.2 mM dNTPs, 0.4 µM primers, DNA template, 1.25 units HiFi^{PLUS} DNA polymerase, DEPC H_2O to 25 ul. The cycling conditions are listed in Table 2-7. The resulting PCR products were resolved by electrophoresis on a 1% agarose gel and the band of interest

extracted and purified using the Qiagen PCR purification kit according to the manufacturer's instructions.

2.8.3 Site-directed mutagenesis

All mutants except E47 were prepared using the site-directed mutagenesis kit according to the manufacturer's instructions (Agilent Technologies). The Jλ1 promoter was mutated in the E47 binding site that was found to be protected in DNase I footprinting experiments. The Eλ3-1 enhancer was mutated in the PU.1 and IRF4 binding sites. The IRF4 expression construct was mutated in the amino acids involved in the interaction with E47 and PU.1. In brief, for the mutant strand synthesis reaction two complementary oligonucleotides containing the desired mutation were synthesized. These primers were gel purified and two reactions, a control and a sample containing the mutated primers were prepared. The control reaction contained the following components: 5 µl of 10× reaction buffer, 2 µl (10 ng) of pWhitescript 4.5-kb control plasmid (5 ng/µl), 1.25 µl (125 ng) of oligonucleotide control primer #1 [34-mer (100 ng/µl)] , 1.25 µl (125 ng) of oligonucleotide control primer #2 [34-mer (100 ng/µl)], 1 µl of dNTP mix, 3 µl of QuikSolution reagent, 36.5 µl of double-distilled water (ddH2O) to a final volume of 50 µl, 1 µl of PfuUltra HF DNA polymerase (2.5 U/µl), 10 QuikChange II XL Site-Directed Mutagenesis Kit. For the sample reaction containing the primers with the desired mutations the following reaction was prepared in a final volume of 50 µl: 5 µl of 10× reaction buffer, X µl (10 ng) of dsDNA template, X µl (125 ng) of oligonucleotide primer #1 , X µl (125 ng) of oligonucleotide primer #2 , 1 µl of dNTP mix , 3 µl of QuikSolution , 1 µl of PfuUltra HF DNA polymerase (2.5 U/µl), where X detonates the amount of DNA template or primers required for the specific site directed mutagenesis reaction. The PCR condition in Table 2-6 were applied for both control and sample reactions. The reactions were placed on ice for 2 minutes and 1 µl of the Dpn I restriction enzyme (10 U/µl) was added directly to each amplification reaction and mixed by pipetting the solution up and down several times and incubated at 37°C for 1 hour to digest the pa-

rental (i.e., the non-mutated) supercoiled dsDNA. To proceed with the transformation into competent cells, 45 µl of XL10-Gold Ultracompetent Cells (tetracycline and chloramphenicol-resistant) were aliquoted into a pre-chilled 14-ml BD Falcon polypropylene round-bottom tube and 2 µl of the β-ME mix was added to the cells. Cells were incubated on ice for 10 minutes, swirled gently every 2 minutes and 2 µl of the Dpn I-treated DNA from each control and sample reaction were mixed with the competent cells and the reaction was incubated on ice for 30 minutes. The samples were heat-pulsed in a 42°C water bath for 30 seconds and the tubes were incubate on ice for 2 minutes to recover the cells. 0.5 ml of preheated (at 42°C) NZY+ broth was added to each tube followed by incubation at 37°C for 1 hour with shaking at 225–250 rpm. The reaction was plated into agar plates containing the appropriate antibiotic for the plasmid vector and incubated at 37°C for >16 hours. The resulting clone was isolated and the fragment of interest subjected to PCR and sequenced to confirm the identity of the mutations.

2.8.4 Overlap extension PCR

E47 mutants were prepared by overlap extension PCR. A set of primers (internal primer 1 and internal primer 2) complementary to the sequences containing the mutation of interest were designed. A second set of primers (Acc65I primer and NheI primer) were designed upstream and downstream of the internal primers 1 and 2, respectively. A first round of PCR was carried out by using internal primer 2 and Acc65I, while a second PCR reaction used internal primer 1 and NheI. pCS2Myc/E2A was used as a template (Table 2-13). Each PCR product was sequenced to verify the accuracy of the fragment generated, gel-purified by electrophoresis on a 1.5% agarose gel and the PCR products were mixed. This mixture was used as a template for a further PCR using Acc65I and NheI as primers. The resulting PCR product was sequenced and gel-purified. Because E47 contained several GC-rich regions, the PCR reaction was performed in the presence of 5% DMSO using a high fidelity Taq polymerase (HiFi). The final PCR products were then digested with

Acc65I/NheI and cloned into the Acc65I and NheI sites of pCS2Myc/E2A, previously double-digested and CIP-treated according to the manufacturer's instructions (NEB). The resulting clone was eventually sequenced to confirm the identity of the mutation.

2.8.5 Gel electrophoresis and DNA elution

PCR fragments were separated by agarose gel electrophoresis for 30- to 45 minutes at 75V constant in 1x TAE or 1X TBE buffer. For purification of PCR products, the gel containing the band of interest was cut and purified either using the Quiagen gel purification kit according to the manufacturer's instructions or electroeluted. For the latter, DNA fragments were subjected to electrophoresis in the presence of the intercalating agent ethidium bromide and the band of interest visualised by brief exposure to ultraviolet light. The band was excised using a scalpel blade and electroeluted at 100V for 1 hour using dialysis tubing with a molecular weight cut off of 8 kDa in the presence of 1 ml 1x TAE. Electreluted DNA was cleaned by two phenol chlorophorm extractions and precipitated in 2 volumes 100% ethanol and 1/10th volume of 3M sodium acetate (pH 5.2), followed by two washes in 70% ethanol.

2.8.6 RNA extraction

Cells were grown in T75 flasks and washed twice in cold PBS and scraped. Following a centrifugation at 1800 rpm for 3 minutes, the pellet was washed twice with cold PBS. One ml of TRIzol reagent was added to the pellet and samples mixed vigorously with additional vortexing. The cell suspension was incubated for 8 minutes at room temperature and centrifuged at 11500 rpm for 15 minutes at 4˚C. The aqueous phase was transferred to a fresh Eppendorf tube and RNA was precipitated by addition of 0.5 ml isopropanol. Samples were mixed by gentle inversion and incubated for 10 minutes at room temperature. Following centrifugation at 11500 rpm for 15 minutes at 4˚C, the supernatant was discarded and the pellet re-suspended in 400 µl TE and RNA was precipitated by addition of 40 µl 3M sodium acetate pH 5.2 and

880 µl 100% ethanol. After incubation at room temperature for 10 minutes, samples were centrifuged at 11500 rpm for 10 minutes at 4˚C and pellets washed with ice-cold 80% ethanol. Samples were centrifuged at 9500 rpm for 5 minutes at 4˚C and the pellets left to air dry on ice and re-suspended in 50 µl ddH_2O. For certain further analyses, the RNA was treated using RQ1 RNase-Free DNase I according to the manufacturer's instructions (Promega).

2.8.7 Generation of cDNA

To generate cDNA (complementary DNA), the Roche transcriptor first strand cDNA synthesis kit was used according to the manufacturer's instructions (Roche). In brief, 2 µg of extracted RNA was mixed with 2 µl Random hexamer primers to a final volume of 13 µl. This was subjected to a PCR cycle for 10 minutes at 65˚C, with heated lid. After 10 minutes, the following components were added to the reaction: 4 µl 5x buffer, 2 µl dNTP (10 mM stock), 0.5 µl reverse transcriptase (20 U/µl) and 0.5 µl RNA Protect (40 U/µl) to a final volume of 20 µl. The reaction was subjected to another PCR cycle: 55ºC for 30 minutes, 85ºC for 5 minutes and 4ºC for 4 minutes. The cDNA was then analysed by electrophoresis gel (1-1.5% depending on the size of the cDNA segment of interest) to identify the band of interest.

2.8.8 DNA Ligation

PCR products of the gene of interest were generated and digested with the appropriate restriction enzyme. Plasmid vector DNA was digested with the same restriction enzymes. PCR products were ligated to the vector using T4 DNA ligase enzyme (NEB) using a 3: 1 ratio of insert to vector. In brief, in a microcentrifuge tube the following were mixed: vector DNA, PCR product, 1x ligase Buffer, 0.5-1µL T4 DNA ligase and ddH_20 to a total of 10µL. The PCR-plasmid mix was incubated for 1 hour (or overnight) at 16°C in the cold room and used for bacterial transformation.

2.8.9 Transformation of *E. coli* DH5α cells

In a microcentrifuge tube, 1 to 5 µl of DNA (usually 10pg to 100ng) was mixed with a 20-50 µl aliquot of chemically-competent *E. coli* DH5α cells which had previously been thawed on ice. Following incubation for 30 minutes on ice, cells were heat-shocked at 42˚C for 45 seconds to allow uptake of the plasmid DNA containing the insert of interest into the competent cells. Cells were cooled on ice for at least two minutes and SOC or LB medium at room temperature was added to the cells which were incubated at 37˚C for one hour in a shaking incubator and subsequently plated on LB agar plates supplemented with the required antibiotic. Plates were incubated at 37˚C overnight or until single colonies were visible. To identify plasmids containing the required insert, colonies were shaken overnight in 5 ml LB medium supplemented with the required antibiotic and plasmid isolated as described below (Chapter 2.8.10). A restriction digest was performed on the isolated plasmid to determine the size of the insert.

2.8.10 Plasmid purification

Colonies containing the desired plasmid were grown in a larger volume, typically 5-10 ml (miniprep) or 250-500 ml (maxiprep) of LB medium supplemented with the required antibiotic, at 37˚C overnight and plasmids purified using Qiagen Miniprep Spin Kit or NucleoBond BAC DNA kits according to the manufacturer's instructions.

2.8.11 Determination of nucleic acid concentration

To determine the concentration of DNA and RNA a conventional Nanodrop (Nanodrop ND-1000, Thermo Scientific, UK) or spectrophotometer were used according to the manufacturer's instructions. Quantification was performed measuring the optical density (OD) at 260nm and 280nm wavelengths (OD_{260nm} and OD_{260nm}) according to the Lambert-Beer's law that relates absorbance to concentration using the path length of the measurement and an extinction coefficient:

$$A = \varepsilon c l$$

Where A is the absorbance, ε the molar extinction coefficient, *c* the concentration of the solution and *l* the length in cm of the solution the light passes through (using a standard 10 x 10 cuvette, the light path length is fixed at 1 cm). For 1 cm path lenght standard the extinction coefficient will correspond to: 50 (dsDNA), 40 (ssRNA) or 33 (ssDNA). The OD_{260}/OD_{280} ratio is an indication of the purity of the sample: if the ratio of DNA OD_{260}/OD_{280} is between 1.8 and 2.0, the DNA purity (meaning free from protein contaminants) is approx. 90% or better. If the ratio of RNA OD_{260}/OD_{280} is approx. 2.0, the RNA purity is approx. 90% or better.

2.8.12 Transient transfections

Cos-7 cells were transfected by the calcium phosphate co-precipitation method. In brief, 250 µl 2x HBS were transferred into polystyrene tubes and in a second tube the plasmid DNA of interest was mixed at the concentration required with 20 µl of 2M $CaCl_2$. The mixture was added to the 2x HBS and incubated for 20 minutes at room temperature. In the meantime, cells were washed twice with PBS and, following incubation, the mix was added drop-wise to the cells. Cells were covered with complete medium and incubated at 37˚C for 4 hours. Cells were washed with serum-free DMEM and 500 µl 15% glycerol/HBS was added to the cells and quickly removed. Cells were washed twice with serum-free DMEM and incubated with complete medium for at least 48 hours prior to harvesting.

911 and HaCaT cells were transfected with the recombinant adenovirus constructs using Lipofectamine 2000 (Invitrogen) for generation of viruses according to the manufacturer's instructions.

2.8.13 Cx26 siRNA transfection

siRNA for Cx26 was purchased from ThermoScientific (L-019285-00-005, 5nmol stock concentration, SMART pool on target). siRNA transfection were performed according to the manufacturer's instructions (Life Technologies). In brief, for each

siRNA transfection, 125 µl of serum- and antibiotic-free DMEM was mixed with 4 µl of siRNA and in a second tube, 125 µl of serum- and antibiotic-free DMEM was mixed with 8 µl of RNAMax transfection reagent. The two solutions were mixed and incubated for 20 minutes at room temperature. HaCaT ells were washed twice with PBS and the mixture added carefully to the cells and volume increased to 5 ml with serum- and antibiotic-free DMEM. After 5 hours, the medium was changed with antibiotic-free DMEM in the presence of serum. Cells were incubated for 72 hours prior to harvesting for further analysis.

2.8.14 Probe preparation for Deoxyribonuclease I (DNase I) footprinting reactions

A DNA fragment of approx. 300-400 bp encompassing the region around the start site of sterile transcription of Jλ1, Jλ2 and Jλ3 was amplified by PCR. The bacterial artificial chromosome Rp23-24i11 covering the 3́ half of the lambda locus was used as a template to amplify the Jλ1-3 gene segments. Mouse kidney genomic DNA was used as a template for Jλ2. The PCR products and the vector were digested with HindIII and BamHI and ligated into pBluscript expression plasmid. The ligation reaction was prepared on ice in a final volume of 10 µl using T4 DNA ligase and incubated at 16°C in the cold room using a 3:1 molar ratio of insert to vector. Half of the ligation products were chemically transformed into competent DH5α *E.coli* cells and shaken in LB at 37°C overnight. The cultures were plated on LB agar plates supplemented with ampicillin and incubated for 1-2 days at 37°C. Discrete colonies (5-10) were picked and grown in 2 ml of LB supplemented with ampicillin. Following plasmid isolation, a test digest was performed to confirm the presence of the insert. Positive clones were expanded by growing in 500 ml cultures in the presence of ampicillin in a shaking incubator at 37°C overnight. Maxipreps were performed according to the alkaline lysis procedure using QIAGEN columns, according to the manufacturer's instructions. The final DNA was re-suspended in 500 µl TE. The concentration of plasmid DNA was determined by spectrophotometry (OD260) and 100

µg of DNA was double-digested with SacI/ClaI to generate the Jλ1, Jλ2 and Jλ3 lower strands, and with XbaI/KpnI to generate the Jλ1, Jλ3 upper strands, and XbaI/ApaI to generate the Jλ2 upper strand. Following digestion, samples were separated by electrophoresis in 1% agarose gels at 90V for 1 hour, the DNA fragment was excised and electroeluted at 90V for 1 hour using dialysis tubing with a molecular weight cut off of 6-8 kDa in the presence of 1-1.5 ml of TE. The fragment was purified by two phenol/chloroform extractions and re-suspended in 100 µl TE followed by precipitation in the presence of 10 µl 3 M sodium acetate (pH 5.2) and 200 µl 100% ethanol. The DNA was further purified with two washes of 70% ethanol to remove the salts, re-suspended in TE and quantified by spectrophotometry as described above.

2.8.15 DNA labelling with the Klenow fragment

Purified DNA fragments (1.3 µg) were digested with the appropriate restriction enzymes and labelled using [α-^{32}P]-dCTP and Klenow DNA polymerase by filling in of 3'-recessed ends. The reaction was prepared in a final volume of 50-60 µl depending on the concentration of the DNA. DNA fragments, 10x *E.coli* polymerase buffer, 30 µCi [α-^{32}P]-dCTP and 10 U of Klenow DNA polymerase were incubated for 15 minutes at 12°C. Unlabelled deoxyribonucleoside triphosphatess (dNTPs, 0.2 mM) were added and incubated at room temperature for 10 minutes. To remove unincorporated dNTPs, the reaction was purified using a G-25 spin column (GE Healthcare) according to the manufacturer's instructions. Eluted [α-^{32}P]-dCTP -containing fractions were increased to 200 µl by addition of TE and the samples incubated at 56°C for 30 minutes in presence of 5 µl 20% SDS and 5 µl 0.4 mg/ml proteinase K (PK) to inactivate the Klenow enzyme. The probes were then extracted with two phenol chloroform extractions and ethanol precipitated as described above. The pellets were re-suspended in 100 µl TE and counted using a scintillation counter. Approximately 10000-30000 cpm (counts per minute) were used for each lane on a footprinting gel and for each G+A ladder.

2.8.16 DNase I footprinting

Reactions were performed in a final volume of 40 µl in which the binding buffer, 0.5-1 mg/ml polydI-dC, 100 µg/lane of extract and 10000-30000cpm of the [^{32}P]- α-dCTP labelled probe were incubated for 20 minutes on ice. After incubation, 3.3 µl of 30 mM $MgCl_2$ and 10 mM $CaCl_2$ were added to the samples. Between 0.5 µl -1.0 µl of 0.005 U DNase I were added to the free extract sample used as a control, while 0.8 µl to 3.0 µl 0.25 U DNase I were used in the presence of the extract. In both cases, reactions were incubated for two minutes at room temperature. The cleavage reaction was stopped by addition of 5 µl of 100 mM EDTA and then the volume was increased to 100 µl by addition of 10 mM EDTA. Following two to four phenol-chloroform extractions to remove proteins and ethanol precipitation, the pellets were re-suspended in formamide loading buffer and boiled for 2 minutes. The samples were loaded on a pre-run 8% polyacrylamide urea sequencing gel next to a size marker and a G+A ladder (Chapter 2.8.17) for 4-5 hours at 1800V/38mA in 1x TBE. The gels were dried on Whatman 3MM paper using a vacuum assisted gel drier prior to autoradiography using a phosphorimager LAS 5000. Captured images were analysed using the AIDA software.

2.8.17 Guanine plus adenosine (G+A) ladder.

A G+A ladder generated by Maxam-Gilbert chemical sequencing was applied next to the DNase I footprinting samples to identify the footprinted sequences. The radiolabelled probe was incubated with 25 µl formic acid for 3 minutes to catalyze base-specific modification. The reaction was stopped by addition of 200 µl hydrazine stop buffer and 750 µl 100% ice cold ethanol. The samples were left on dry ice for 5 minutes and centrifuged at 14000 rpm for 5 minutes at 4˚C. Following two washes with 70% ethanol, the samples were dissolved in 200 µl dH_2O and precipitated in the presence of 20 µl 3M sodium acetate (pH 5.2) and 400 µl 100% ethanol. After washing the samples in 70% ethanol to remove traces of formic acid, the samples were incubated with 70 µl 10% piperidine at 90°C for 30 minutes to allow the strand

scission reaction. The samples were then dried in a Speedvac evaporator and the pellet was washed twice in 30 µl and 20 µl dH_2O respectively, to remove traces of piperidine. The pellets were re-suspended in formamide loading buffer, boiled at 90°C for 2-3 minutes and loaded on a pre-run 8% polyacrylamide urea sequencing gel and separated by electrophoresis at 1800V/38mA as described above (Chapter 2.8.16).

2.8.18 Electrophretic shift assay (EMSA)

EMSA is based on the retardation of DNA in protein/DNA complexes when separated by electrophoresis under non-denaturing conditions. One strand of the duplex oligonucleotide probe was end-labelled with [γ-^{32}P] ATP by T4 Polynucleotide kinase (PNK) and annealed to the complementary unlabelled strand. Labelling of one strand was performed by incubation of 100 pmol of the oligonucleotide with 1x T4 PNK buffer, 15 µCi [γ-^{32}P] ATP and 10 U of T4 PNK in a 15 µl reaction for 30' at 37˚C. Unincorporated radiolabelled ATP was removed using a G-25 spin column (GE Healthcare) according to the manufacturer's instructions. The end-labelled DNA was annealed with 120 fmol/µl of the complementary strand in the presence of 2 µl of 5 M NaCl in a final volume of 100 µl. The samples were boiled for 2 minutes to allow denaturation of the DNA and left to cool overnight to anneal the strands. Unlabelled double-stranded DNAs used in competition assays were prepared as follows: 100 pmol of each strand, 2 µl 5M NaCl, 28 µl TE were combined in a final volume of 100 µl and boiled for 2 minutes. The samples were cooled overnight to allow annealing of the strands. A 1000- to 4000-fold excess of each competitor DNA was used in each competition assay. For EMSA experiments, the reactions containing binding buffer, 0.5-2 µg/µl poly(dI-dC,) 20-30 µg protein extract and 10 fmol ^{32}P-labelled probe, were incubated for 10-30 minutes on ice in a final volume of 25-40 µl. The samples were separated by electrophoresis on a native polyacrylamide gels for 2 hours at 150V in 0.25x TBE (Table 2-14). The gels were dried onto Whatman 3MM paper using a vacuum assisted gel drier prior to autoradiography. For anti-

body competition experiments, 3 to 10 µl of antibody was added to the binding reactions for 10 minutes on ice prior to addition of the extract, while for competition assays 1000 to 4000 fold excess of unlabelled competitor DNA was added prior to the addition of extract. For relative affinity experiments, between 30- to 1000-fold excess of unlabelled E2A competitor DNA was used prior to addition of the extract.

Table 2-14 Incubation and electrophoresis conditions for gel retardation assays.

Probe	1X Binding buffer	% gel	Incubation
E47	25 mM Hepes-KOH (pH 7.9), 1 mM $MgCl_2$, 1 mM DTT, 1 mM EDTA, 10% glycerol	7% for Jλ1 6% for Jλ2	10 minutes on ice
IRF1/2	10 mM Tris-HCl (pH 7.5), 1 mM $MgCl_2$, 50 mM NaCl, 0.5 mM EDTA, 20% glycerol	6%	20 minutes on ice
Pou1F1a	20 mM Hepes-KOH (pH 7.9), 10 mM $MgCl_2$, 50 mM KCl, 1 mM DTT, 4% Ficoll	5%	30 minutes on ice

2.8.19 Luciferase assay

For luciferase assays, Cos-7 cells were seeded in 6-well plates at a density of $1.2x10^5$ cells/well one day prior to transfection. Constructs in which either the Jλ1 promoter or the Eλ3-1 enhancer drove the expression of the luciferase gene were used. Both constructs were cloned into the pGL3-basic reporter plasmid (promoter-less). The activity of both the promoter and the enhancer was determined in the presence of plasmid vectors expressing E47, PIP and PU.1 (a kind gift of Dr SL Grange). The pRL-TK *Renilla* vector (a kind gift of Dr Ian Wood) was used as an internal control to monitor transfection efficiency. Cos-7 cells were harvested after 48 hours of transfection and treated with passive lysis buffer (PLB), 500 µl per well, using the Dual-Luciferase Reporter Assay System (Promega) according to the manufacturer's instructions. Relative luciferase units were measured in a luminometer (model FB12; Zylux, Maryville, TN, USA.) with a delay time of 2 seconds and integration time of 10 seconds. Firefly luciferase activity was measured by mixing 20 µl of cell lysate with luciferase assay reagent II (100 µl) followed by the *Renilla* reading using Stop&Glo Reagent (100 µl) according to the manufacturer's instructions. The relative luciferase activity was calculated as the ratio of Firefly Luciferase to *Renilla*

Luciferase activity and normalized against average values given by transfection of the pGL3 basic empty vector. The results shown are means of standard deviations for at least three separate transfections performed in duplicate. Luciferase activity is expressed as relative lights units (RLU). Raw data and calculation methods are shown in Appendix 1 Table 7-2.

2.9 Virus methods

For recombineering experiments a BAC clone termed AdZ-5F35-CV5 adenovirus, a kind gift of Dr James Findlay, was used (Findlay, 2012). This adenovirus contains the backbone of the Ad5 vector, which contains a selection cassette in place of the E1 region under the control of a CMV promoter (Figure 2-1). Furthermore, the Ad5 fibre was replaced with the Ad35 fibre. Since the genome has the E1 and E3 regions deleted, the adenovirus is effectively replication deficient. These regions were replaced with the following genes of interest: ECFP (a kind gift of Gareth Howell) within the E3 region and HPV16 E6 or HPV16 E7 within the E1 region.

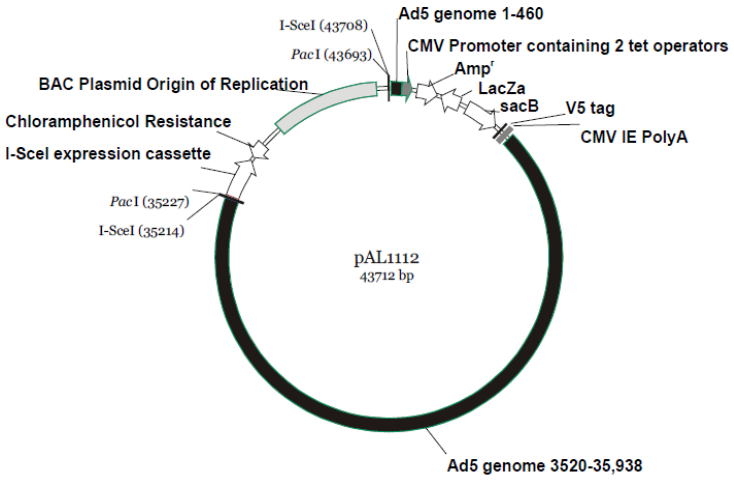


Figure 2-1 Schematic representation of the AdZ-5 vector.

The AdZ-5 vector vectors contain the Ad5 vector genome, deleted for E1 & E3 regions (rendering the vector non-replicative) in a single copy vector. It contains the CMV promoter and the PolyA signal sequence; the cassette encoding ampicillin resistance, lacZα & SacB is found within the MCS. SacB encodes for a gene giving sensitivity to sucrose, lacZα provides blue/white screening. *em7-galK* recombineering.

2.9.1 Generation of recombinant adenovirus expressing GalK (positive selectoin)

The *em7-galK* cassette (a kind gift of Dr James Findlay, University of Leeds, UK) was generated by PCR using the pUC19-em7-galK plasmid as a template and settings as described in Chapter 2.8.2. For positive selection, the *em7-galK* was inserted within the E3 region of the AdZ-5F35-CV5 and SW102 cells were employed as they have a defective phage which expresses the lambda red genes necessary to mediate homologous recombination between DNA fragments. These genes are under temperature sensitive control and therefore they are turned off at 32°C and they can be turned on by growing the bacteria at 42°C for 15 minutes. Therefore *E. coli* SW102 cells containing the AdZ-5F35-CV5 BAC were grown at 32˚C overnight in 5 ml of half-salt LB medium (supplemented with 12.5 µg/ml chloramphenicol and 50 µg/ml ampicillin). The following day, 0.5 ml aliquot of the overnight culture was used to inoculate 25 ml half-salt LB (supplemented with 12.5 µg/ml chloramphenicol and 50 µg/ml ampicillin) and grown in a 32˚C shaking incubator until OD_{600} of approx. 0.6 is reached. The culture was transferred into a 50 ml Falcon tube and incubated for 15 minutes at 42˚C to induce the lambda red proteins. The falcon tube was inverted every 5 minutes and the culture was cooled on ice for 20 minutes and centrifuged at 4000 rpm for 5 minutes at 4˚C (Thermo Scientific IEC CL31R centrifuge). The supernatant was discarded and 1 ml ice-cold sterile water was added to the pellet. This was gently re-suspended by rotation on ice-slurry. Water was added to make up the volume to 20 ml and cells were centrifuged at 4000 rpm for 5 minutes, 4˚C. The latter step was repeated at least 3 times. The re-suspended pellet was then transferred into ice cold 0.5 ml Eppendorf tubes and 3µl of the purified em7-galK selection cassette PCR product (250 ng) was added to 25 µl induced bacteria and incubated on ice for 5 minutes. The mixture was transferred into chilled 0.2 cm electroporation cuvettes and electroporated at 2.5 kV for 5 ms using the Biorad Gene Pulsar X cell electroporator. Cells were recovered in 1 ml half salt LB for 1 hour in a 32˚C shaking incubator and then centrifuged at 17,000 x g for 30 seconds.

The pellets were washed twice with M9 salts and re-suspended in 1 ml M9 salts and three different dilutions (1:1, 1:10, 1:100) plated onto galK positive selection plates. Plates were incubated at 32˚C for 4-6 days. Colonies were selected and streaked into single colonies onto half salt agar plates supplemented with 12.5 μg/ml chloramphenicol and 50 μg/ml ampicillin and incubated at 32˚C for 2-3 days. Colonies were subject to colony PCR (standard PCR reaction using Taq DNA Polymerase) where, as DNA template, a small aliquot of each colony was taken using a micropipette tip. The PCR products were resolved by electrophoresis on a 1% agarose gel and colonies which resulted positive for the insert were grown in 5 ml half salt LB (supplemented with 12.5 μg/ml chloramphenicol and 50 μg/ml ampicillin) at 32˚C overnight. Following purification, BAC DNA was digested with Mlu I for 1.5 hours at 37˚C to select positive colonies. If positive, these were grown in 500 ml half salt LB supplemented with 12.5 μg/ml chloramphenicol and 50 μg/ml ampicillin, purified and used for insertion of the ECFP cassette within the E3 region.

2.9.2 Generation of recombinant adenovirus expressing ECFP (negative selection)

For insertion of ECFP into the AdZ-5F35-CV5 E3 region, *E. coli* SW102 cells containing the AdZ-5F35-CV5 BAC with the *GalK* selection cassette were grown in 5 ml of half-salt LB medium (supplemented with 12.5 μg/ml chloramphenicol and 50 μg/ml ampicillin) in a 32˚C shaking incubator overnight and recombineering performed as described above, using the ECFP PCR product generated previously to replace the *GalK* cassette. Recombineering products were plated onto GalK negative selection agar plates and incubated for 3-4 days at 32˚C until colonies were visible. A number of colonies were grown in half salt LB (supplemented with 12.5 μg/ml chloramphenicol and 50 μg/ml ampicillin) and a colony PCR performed to verify the presence of the insert. Positive colonies were grown in half salt LB at 32˚C overnight and following miniprep, the plasmid DNA was used as a template for a PCR. The PCR products were resolved by electrophoresis on a 1% agarose gel and the band of in-

terest excised and gel purified using the Qiagen PCR purification kit according to the manufacturer's instructions and sent for sequencing to verify correctness of the insert. ECFP expression was also confirmed by transfection of 911 cells.

2.9.3 Generation of recombinant adenovirus expressing HPV16 E6 or E7

For insertion of the HPV16 E6 or E7 within the E1 region, *E. coli* SW102 cells containing the AdZ-5F35-CV5 BAC with ECFP in the E3 region were grown in 5 ml of half-salt LB medium (supplemented with 12.5 µg/ml chloramphenicol and 50 µg/ml ampicillin) in a 32˚C shaking incubator overnight The following day, 0.5ml of the overnight culture was inoculated into 25ml half salt LB and incubated at 32 ˚C until the OD_{600} reached approx. 0.6. The culture was transferred into a 50 ml Falcon tube and incubated for 15 minutes at 42˚C with tube inversion every 5 minutes. The culture was cooled on ice for 20 minutes and centrifuged at 4000 rpm for 5 minutes at 4˚C (Thermo Scientific IEC CL31R centrifuge). The supernatant was discarded and 1 ml ice-cold sterile water was added to the pellet. This was gently re-suspended by rotation on ice-slurry. Water was added to make up the volume to 20 ml and cells were centrifuged at 4000 rpm for 5 minutes, 4˚C. The latter step was repeated at least 3 times. The pellet was re-suspended in the remaining water by gently swirling (20 ml final volume). The re-suspended pellet was then transferred into ice cold 0.5 ml Eppendorf tubes and 3µl of the PCR product of the gene of interest to be inserted was added (HPV16 E7 C-terminus Myc or HPV16 E6 N-terminus HA). The mixture was transferred into chilled 0.2 cm electroporation cuvettes and electroporated at 2.5 kV for 5 ms using the Biorad Gene Pulsar X cell electroporator. Cells were recovered in 1 ml half salt LB, transferred into 15 ml Falcon tubes and the volume increased to 5ml by addition of low salt LB. Cells were incubated for at least 4 hours in a shaking incubator at 32˚C prior to plating on sucrose IPTG/Xgal plates. The suspensions were stored at 4˚C for up to 48 hours before plating to reduce the number of negative and false positive colonies. Plates were incubated for at least for 2 days, after which period several colonies were visible: for a gene giving sensitivity to sucrose,

lacZα provides blue/white screening, however these were discarded as they did not contain the insert, and thus only white colonies were tested for positivity for the insert. These were grown in half salt LB supplemented with 12.5 µg/ml chloramphenicol and following miniprep, a BamHI test digest was performed to confirm the presence of the insert. Positive colonies would lose two bands running at 1.7 kb and 2.5 kb. Once a colony was found to be positive, the insert was amplified by PCR and the products sent for sequencing to verify correctness of the sequence (Chapter 0). This was expanded and the BAC DNA purified for transfection of 911 cells.

2.9.4 Growth of recombinant adenoviruses in 911 cells

911 cells provide the products of the E1 region necessary for virus replication *in trans*. The E1 region was replaced either with HPV16 E6 or E7 coding regions as describe above. Following transfection with Lipofectamine reagents according to the manifacturer's instructions, the cells were incubated at 37 ˚C until they detached, approx. 4-5 days post-transfection. To verify the efficiency of the transfection cells were observed for expression of CFP using a fluorescent inverted microscope (Nikon Eclipse TS100) equipped with Nikon B-A fluorescent filter and Nikon digital Sight DS-5M camera. Images were captured using the AxioVision 4.3 software. Cells were expanded in T175 flasks (Corning). After 3-4 days, cells were collected in 50 ml Falcons and centrifuged at 1200 rpm for 3 minutes and the pellets were washed two times in ice cold PBS. The pellets were subjected to three cycles of snap-freeze in liquid nitrogen, thawed at 37˚C in a water bath and sonicated 45 seconds in a water bath sonicator (Kerry Ultrasonic bath sonicator). This lysate was used to infect a fresh T175 flask of 911 at 70-80% confluence. Ideally, lysate from one T175 flask would infect ten T175 flasks. This procedure was repeated until 30 T175 flasks of infected 911 cells had been generated. Once all the pellets were obtained, virus was purified as described below.

2.9.5 Virus purification

Virus was purified from cell pellet derived from 30xT175 flasks using CsCl density centrifugation. In brief, the cell pellets were transferred into 50 ml falcon tubes and re-suspended in 10 ml sterile 0.1M Tris-HCl (pH 8.0) and 1 ml of 10% sodium deoxycholate was added to a final concentration of 5% (w/v). The re-suspended pellet was left at room temperature for 30 minutes, until it appeared viscous. Then, 100 µl of 2 M $MgCl_2$ and 50 µl DNase I (10mg/ml) were added and the solution incubated at 37˚C for 30 minutes with periodic inversion, until the viscosity was reduced. The lysate was centrifuged at 3000 rpm for 15 minutes at 4ºC and the supernatant transferred into a fresh tube and kept on ice until ready to be loaded on the CsCl gradient. one gradient was prepared for every 5ml of virus lysate. Therefore, two gradients were prepared inSW40 ultraclear Beckman tubes (14x95mm, Cat. 344060) each containing: a bottom layer with 1 ml 1.5d CsCl, a middle layer with 2.5ml of 1.35d CsCl and a top layer with 2.5ml of 1.25d CsCl. The virus lysate was added on top of the gradients and tubes were balanced with 0.1M Tris-HCl (pH 8.0). To prevent diffusion between the gradients the gradients were centrifuged as soon as possible. The tubes were centrifuged for 1 hour in a Beckman's Optima L-80 ultracentrifuge using a SW40 rotor at 35000 rpm at 10ºC at the slowest acceleration and deceleration settings (deceleration was set to 1). After 1 hour, the virus band of interest was located between the 1.35d and 1.25d interface (Figure 2-2). This band is carefully removed from the tubes and kept on ice until a further CsCl gradient was prepared. The purified virus was gently loaded on top of 2ml of 1.35d CsCl using a SW55ti ultracentrifuge tube (ultraclear 13x51mm, cat: 344057). Where required, to achieve layering, bands were diluted with 0.1 M Tris-HCl (pH 8.0). Following centrifugation at 35000rpm, 4˚C for 12 hours using a SW50 rotor (with minimum acceleration, no brake), virus particles were detected in the upper of the two visible bands (Figure 2-2 C). both bands were collected and extensively dialysed against 3 changes of 10mM Hepes-KOH buffer (pH 8.0) using Thermo Scientific Slide-A-lyzer

MINI Dialysis Devices, 3.5K MWCO (molecular weight cut off) 0.1ml (cat: 69550) over the course of a day at 4˚C. The purified virus is collected and sterile glycerol added to 10% (v/v) is. Aliquots were prepared and stored at -80˚C.

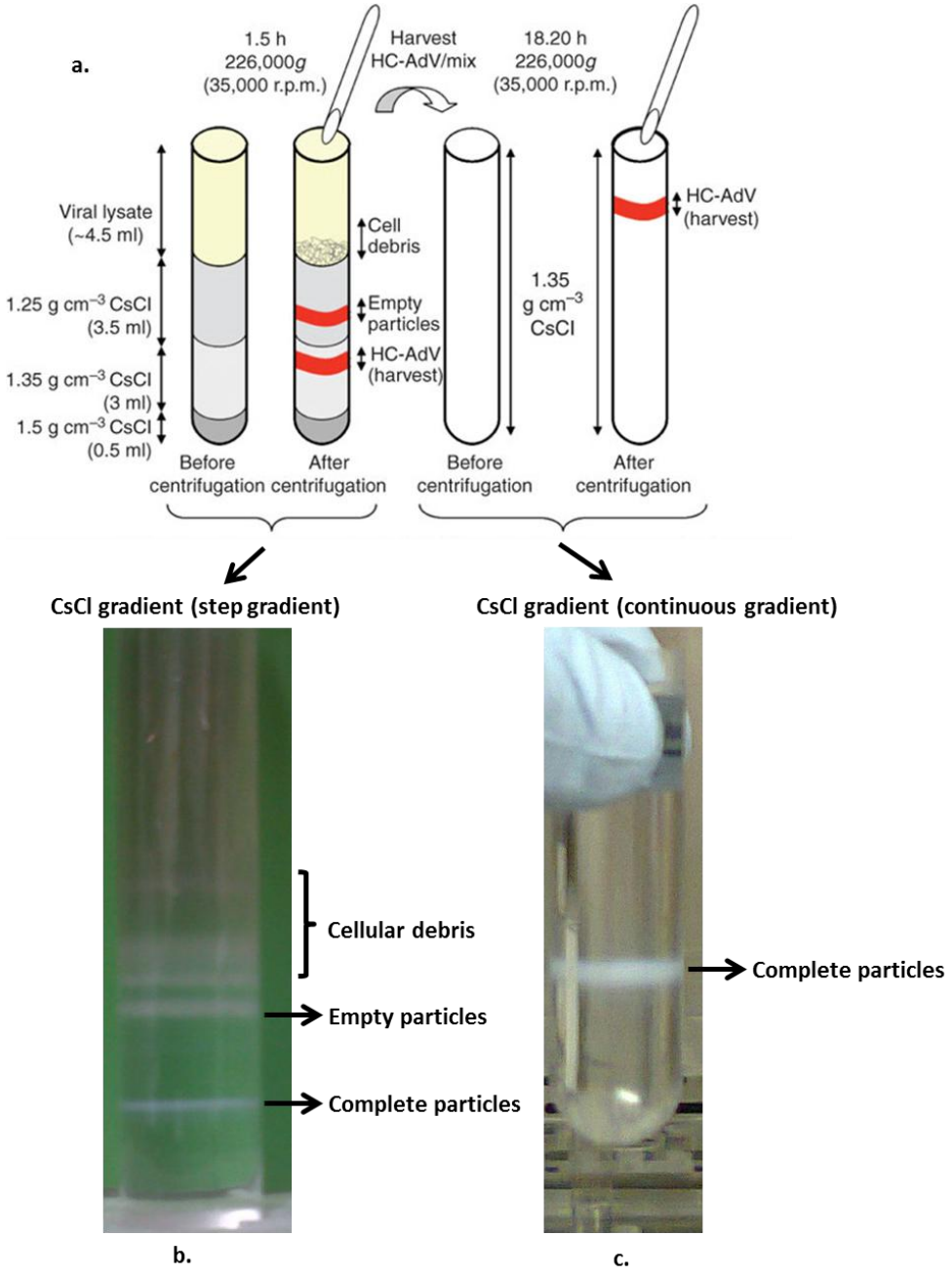


Figure 2-2 CsCl gradient virus purification.

Top panel, a schematic representation of the stages of virus purification (Adapted from (Jager et al, 2009)). B-C: purified bands following the first spin (B) and the overnight centrifuge (C), showing the bands containing complete virus particles. HC-AdV = High-capacity adenoviral vectors.

2.9.6 Virus quantification by the fluorescent focus unit (FFU) assay

To determine the titre of the purified virus, 911 cells were seeded in a 24-well plate at a density of $1.2x10^5$ cells/well. The following day, a series of 10-fold dilution of the virus were prepared for each well, in triplicate (Table 2-15). Cells were washed twice with cold PBS, virus particles diluted in serum-free medium (DMEM) and added to the cells. Following incubation for 1 hour at 37˚C, the medium was changed and complete DMEM was added and cells incubated at 37˚C for 24 hours. On the following day, the cells were washed twice with PBS and incubated with trypsin for 5 minutes. The cells were re-suspended in 500 µl complete DMEM and transferred into 1.5 ml Eppendorfs and centrifuged at 400g for 3 minutes at 4˚C. The pellet was re-suspended in 100 µl 2% PBS buffered paraformaldehyde and incubated for 10 minutes at room temperature. Cells were centrifuged for 3 minutes at 2000g, 4˚C and pellets re-suspended in PBS and analysed by flow cytometry for CFP expression (FACS Fortessa Diva). To calculate FFU from the flow cytometric data, the dilution at which 50% of the cells were CFP positive (X) was multiplied by the number of cells plated at the beginning of the experiment (Y= $1.2x10^5$). This was divided by the dilution factor (i.e. Z= $5x10^{-4}$) and multiplied by 2 to give the number of FFU/ml. To calculate the volume of purified virus to use in the experiments, the following calculation was performed:

$$\mu l \ o\mathbf{f\ purified\ virus} = Yx1000/Z$$

which represents the amount of purified virus which infects 100% of the cells in 24 hours (since cells were left in the presence of the purified virus for 24 hours prior to analysis by flow cytometry).

Table 2-15 Values for virus quantification by the fluorescent focus unit assay.

Tube	Serial 10-fold dilutions	Virus	Volume virus (µl)	Volume media (µl)
1	$5x10^{-3}$	Stock solution	20	1980
2	$5x10^{-4}$	Tube 1	200	1800
3	$5x10^{-5}$	Tube 2	200	1800
4	$5x10^{-6}$	Tube 3	200	1800
5	$5x10^{-7}$	Tube 4	200	1800
6	0	None	0	2000

2.9.7 Transduction of HaCaT cells

To analyse expression of HPV16 E6, E7 and Cx26, HaCaT cells were transduced with an appropriate concentration of virus particles. Cells were washed twice with cold PBS and virus diluted in serum-free DMEM and added directly to the cells in a final volume of 5 ml (T75 flasks) or 1 ml (6-well plates). Incubation was performed for up to 3 hours before the medium was removed and replaced with complete DMEM. For Western blots, cells were incubated at 37˚C for a further 48 hours, for immunofluorescence and dye transfer assays, cells were incubated for 24 hours at 37˚C.

2.9.8 Lyoplate assay

The lyoplate assay was performed on normal human keratinocytes (NHK), CaSki and SiHa cells in collaboration with Dr Erica Wilson (Leeds Institute of Cancer and Pathology, Leeds) according to the manufacturer's instructions (BD Lyoplate™ Screening Panels, BD Bioscience). In brief, the three cell types were cultured until they reached 80-90% confluence and washed with 1x PBS, treated with trypsin and centrifuged at 300g for 5 minutes; to remove any clumps, cells were passed through a BD Falcon™ 40 or 70 µm cell strainer and cells concentration determined. Pellets were re-suspended in 1x PBS with calcium and magnesium with the addition of 100 units/ml DNase at 10 million cells per ml to mitigate cell clumping. Following incu-

bation at room temperature for 15 minutes, cells were washed in two to four volumes of 1X PBS and centrifuged at 300g for 5 minutes. NHK, CaSki and SiHa cells were fluorescently labelled with 0.4µM cell tracker™ green- 5-Chloromethylfluorescein Diacetate (CMFDA, $C_{25}H_{17}ClO_7$), 2µM orange-CMRA ($C_{30}H_{25}Cl_2NO$), 2µM violet-BMQC ($C_{16}H_{16}BrNO_2$), respectively and incubated for 1 hour at 37˚C. Cells were washed 3 times PBS/0.5%BSA and re-suspended in 5 ml BD Pharmingen Stain Buffer (275 ml of BD Pharmingen Stain Buffer (FBS) with the addition of 5 mM EDTA (final concentration)) and 100 µl of cell solution was aliquoted to three fresh, round-bottom 96-well plates at a density of $6x10^5$/well. To reconstitute the antibody, the BD plates were centrifuged at 300 g for 5 minutes. Using a multi-channel pipette, the lyophilized antibodies were reconstituted in 110 µl of 1X sterile PBS. This resulted in an antibody solution that contains five tests (20 µl/test). Reconstituted antibodies were stored at 4˚C until further use. The reconstituted antibody was mixed using a multi-channel pipette, by pipetting carefully up and down 2-3 times and 20 µl were added to corresponding wells of sample plate previously prepared. Cells were incubated on ice for 30 minutes and washed with 100 µl of BD Pharmingen Stain Buffer/ EDTA per well. Following the incubation, cells were centrifuged at 300 X g for 5 minutes and the supernatant carefully removed. Cells were washed with an additional 200 µl of BD Pharmingen Stain Buffer /EDTA and centrifuged at 300 X g for 5 minutes. The supernatant was discarded and 100 µl of the appropriate secondary antibody were added directly to cells in each well and incubate for 30 minutes on ice in the dark. The secondary antibody was diluted in BD Pharmingen Stain Buffer+ EDTA to a final concentration of 1.25 µg/ml (1:200 dilution). Following incubation, cells were washed with 100 µl of BD Pharmingen Stain Buffer /EDTA and centrifuge at 300 X g for 5 minutes. The supernatant was discarded and cells washed with an additional 200µl of BD Pharmingen Stain Buffer + EDTA. Cells were centrifuged at 300 x g for 5 minutes and the pellets re-suspended in 100 µl of live cell viability dye (Zombie NIR) and incubated at room temperature for 30 minutes. Cells were centrifuged at 300 X g for 5 minutes and re-suspended in

100 µl BD Pharmingen Stain Buffer/ EDTA. Cells were centrifuged at 300 X g for 5 minutes and re-suspended in 150 µl BD Pharmingen Stain Buffer/ EDTA and analysed. Flow cytometry data analysis was done with FlowJo or FACSDiva.

2.9.9 Lucifer yellow dye transfer assay

The Lucifer yellow dye transfer assay was carried out in collaboration with Dr Paola Massimi (ICGEB, Trieste, Italy). This was done to assess the expression of gap junction proteins in HaCaT cells transduced with recombinant adenoviruses to investigate the role of E6 in connexin expression. In brief, glass capillaries tubes were prepared for microinjections using a heating apparatus set at a current of 17-18 mA at 613-617 ˚C (BB.CH, Mekanex, Geneva). For the dye transfer, HaCaT cells were seeded in 3 cm dishes at a low density (approx. 25% confluent) a day before being transduced with the recombinant adenovirus as described above. The following day, the medium was replaced with PBS for microinjection which was performed using an Eppendorf Femtojet automated pressure microinjector, at an average pressure of 350 hPa, depending on the cell to be injected. To prepare the dye, 1 M lithium chloride was prepared in water and 400 µl were mixed with 16 µl Lucifer yellow dye (25 mg/ml) to obtain a final concentration of 4% dye. This was centrifuged for 1 minute at maximum speed and kept at 4˚C protected from light, until use. For visual guidance to identify cells for injection a white light and interference contrast optics connected to a screen (Sony Trinitron) was used; this was also connected to the microinjectror. A confocal microscope was used to follow diffusion of the dye and take pictures at selected intervals (Zeiss LSM510). The percentage of cells that passed dye to at least one neighbouring cell was counted. For each sample, 10 to 15 injections were performed.

2.10 Next generation sequencing

Samples used for high throughput RNA-Seq analysis of cell-surface transcripts had been previously prepared and sequenced by Dr Joanne Morgan at the sequencing facility in St James University Hospital, Leeds, UK (Watherston, 2010). In brief, following total RNA extraction and cDNA generation (Chapter 2.8.6 and 2.8.7), full-length cDNA was sheared using a Covaris S-series instrument (USA) to generate fragments of approximately 200–400 bp. The paired-end sequence sample preparation kit (Illumina, USA) was used to perform end repair to create blunt ends in the cDNA as well as for addition of Adenine to the blunt-ended 3' ends of the 200-400 bp cDNA fragments catalysed by Klenow-exo enzyme; 5' and 3' overhangs were removed using a 3' exonuclease and a 5' DNA polymerase. These fragments were purified using the Qiagen PCR purification MinElute column kit (Qiagen, UK) according to the manufacturer's instructions. Following purification, the Adenine-tagged DNA fragments were ligated to the flow cell using DNA ligase and PE adapter oligo-mix for 15 minutes at 20°C. The samples were purified using the Qiagen PCR purification MinElute column kit. Unligated adapter molecules were removed by agarose gel electrophoresis and the correct size templates for the high-throughput next generation sequencing were selected. RNA-Seq of cDNA was performed using the Illumina Genome Analyser II (Illumina Inc., USA). In brief, the nucleotide bases were modified by addition of a fluorophore and they were incorporated into oligo-primed DNA fragments (a sequence-by-synthesis approach). The fluorophore and 3'-OH group of the nucleotide bases, which has been chemically modified to prevent polymerisation, were removed to wash the flow cell before sequencing precedes the next addition of a nucleotide. The number of times the flow cells is washed depends on the read length. For this analysis, read were 75 bp long, thus, this step was repeated 75 times. Reads were saved as fastq formats and ready to be analysed (Figure 2-3). These reads are defined as single-end reads or paired-end reads depending on

whether the sequencing was performed from one strand or both ends of the cDNA molecule, respectively. For this study reads were sequenced from one end only.

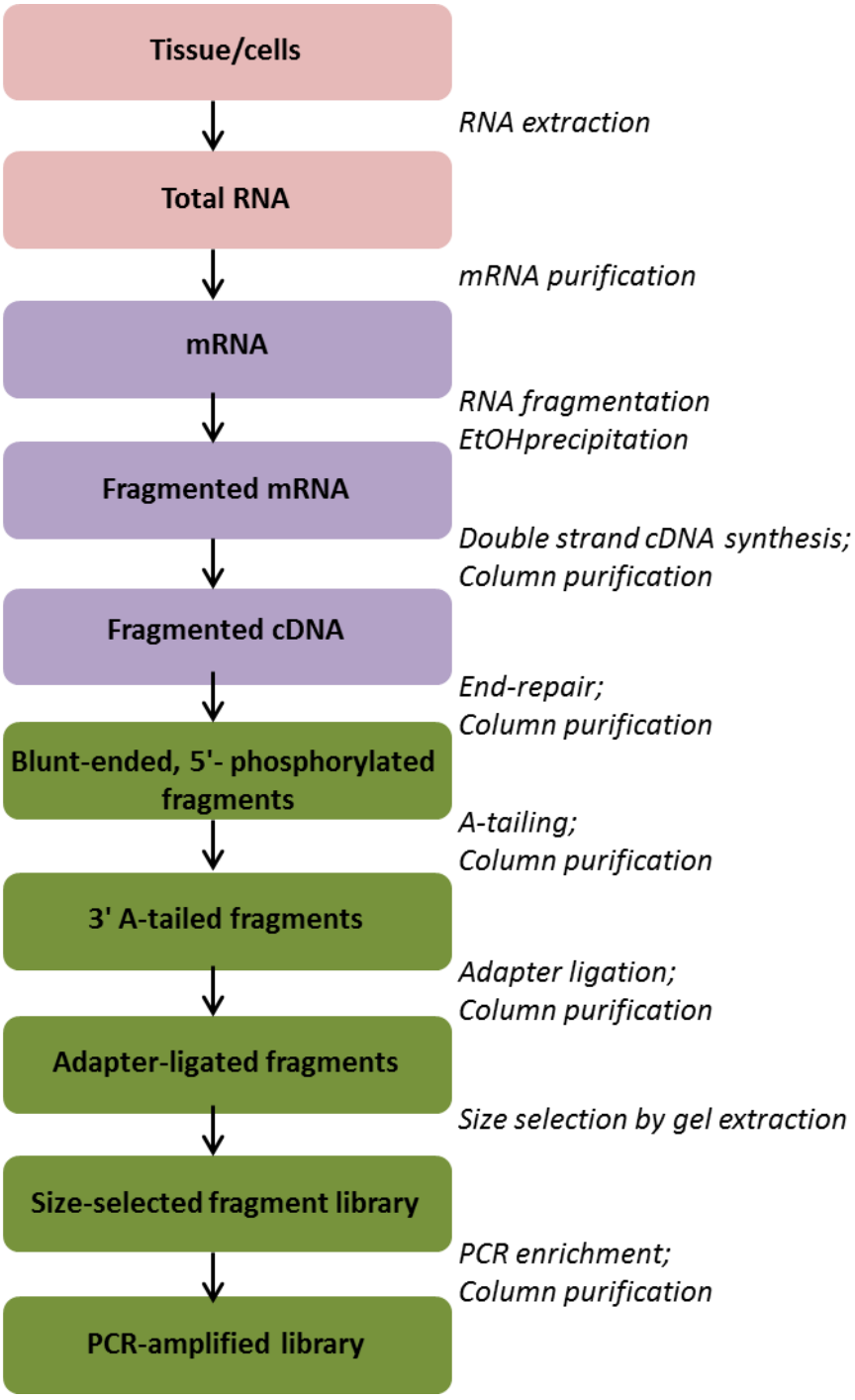


Figure 2-3 Library preparation for Illumina RNA-Seq platform. Adapted from (Kumar et al, 2012).

2.10.1 Analysis of next generation data: overview of the TopHat/Cufflinks pipeline

Bioinformatics analysis was performed to interpret the sequencing results using the Galaxy platform. The Tophat-Cufflinks pipeline was used (Trapnell et al, 2012). In brief, the sequencing reads were first pre-processed to eliminate low quality bases which could interfere with the alignment (Figure 2-4). This was done using the FastQC kit analysis tool. The sequences were then aligned to a reference genome (hg19, 2009) using TopHat and transcript abundance was estimated using Cufflinks. Cuffdiff was used to evaluate differentially expressed transcripts between the samples.

For this experiment, there were no replicates, either biological (i.e. replicates samples of the same tissue) or technical (replicates of the same RNA sample). Therefore, to consolidate the results generated by Cuffdiff, the same samples were also run using a program designed to deal with non-replicated data: GFOLD (Feng et al, 2012). GFOLD ranks the genes by expression level differences between the samples compared (Figure 2-4). Unless specified, TopHat, Cufflinks and Cuffdiff were ran using the default parameters as set on the Galaxy platform (Appendix 4: NGS/RNA-Seq analysis parameters). CummeRbund, an R based program, was used to visualize Cuffdiff outputs; the latter and GFOLD were both run using a UNIX operating system according to the instructions and commands provided in the respective manuals (Feng et al, 2012).

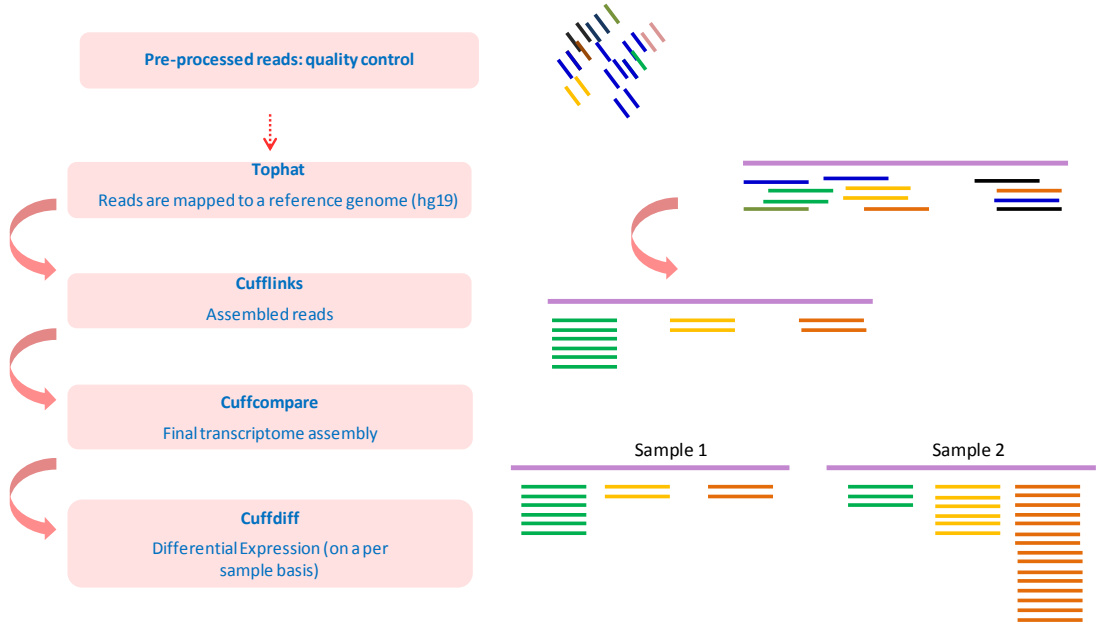


Figure 2-4 RNA-Seq pipeline adopted in this study.

2.10.2 Quality reads analysis

High throughput sequencing pipelines generate raw sequence data which contain low quality bases which must be trimmed off from the reads. In addition, contaminating adapter sequences must be removed, as well as reads that are too short to accurately align. The FastQC tool analyses all the sequences and gives a graphical view of the results. Results show basic statistics, per base sequence quality, per sequence quality scores, per base sequence content, per base GC content, per sequence GC content, per base N content, sequence length distribution, duplicate sequences, and over-expressed sequences. FastQC was run on all the samples, a quality score for each base was calculated and low quality bases trimmed off using FastQC Trimmer. Box and whisker plots of per-base sequence quality were used to assess the base quality. This output gives an overview of the quality values across all the bases in the sequence being analysed, at each position. Therefore, if the lower quartile for any base is <10 (or the median for any base <25), the output will be a "warning"; if the lower quartile for any base is <5 (or the median for any base is <20), the output will be a "fail". In the latter case, by looking at the box and whisker

plot, low quality bases will be trimmed off. Similarly, all the other jobs run by FastQC gave a “pass”, “warning” or “fail” output, depending on which additional modification to the reads were done before aligning the sequences to the reference genome.

2.10.3 How does TopHat work?

TopHat is a program used to align RNA-Seq reads to a reference genome by finding exon-exon splice junctions (Trapnell et al, 2009). TopHat is a wrapper for a program called Bowtie, which aligns sequenced reads one at the time (http://bowtie-bio.sourceforge.net/bowtie2/manual.shtml). Bowtie is one of the earliest and most frequently used alignment tools. It is based on a computational program which “indexes” the reference genome to speed up the process of mapping and use less space. Indexing is based on a technique for compressing large files known as the Burrows-Wheeler. Briefly, Bowtie will align the sequences one at the time to the Burrows-Wheeler compressed genome; as the alignment progresses, the mapping positions will be narrowed, since Bowtie will build up on pre-existing aligned sequences. If Bowtie cannot align the sequence will try again by considering one mismatch; if this fails, then it will try with two mismatches. The number of mismatches allowed has to be set accordingly as it depends on the read length being analysed (Cullum et al, 2011). As a general rule, 2 mismatches are considered standard. TopHat additionally adds another feature: the ability to align exon-exon junctions. Aligning sequences which belong to the same exon requires fewer algorithms than those belonging to several exons; therefore Tophat will use Bowtie to align the sequences back to known exons in the genome, and the remaining sequences will be aligned to junctions between junctions. This is a very important feature since it can potentially predict new exons and thus new transcripts (Trapnell et al, 2009).

2.10.4 How does Cufflinks work?

Following mapping, reads belonging to the same gene/transcripts have to be quantified and normalised by taking into account gene length as well as the sequencing depth. Both are necessary to enable us to compare the samples analysed. By following the pipeline, the next step is to use TopHat output (SAM/BAM format) to run Cufflinks; Cufflinks quantifies transcript levels in units of reads per kilobase per million mapped reads, or RPKM (Mortazavi et al, 2008). In paired end reads it will calculate fragments per kilobase per million reads (FPKM) (Trapnell et al, 2010). FPKM is defined as:

$$\mathbf{FPKM = 10^9 xC/NL}$$

Where: C represents the number of reads mapped on to the gene's exons, N the total number of reads in the experiment and L the sum of the exons in base pairs. This is a crucial step as the "RPKM/FPKM normalizes a transcript read count by both its length and the total number of mapped reads in the sample" (Mortazavi et al, 2008). Default parameters from the Galaxy platform were used to analyse the data.
Having calculated the variance of a transcript abundance in each condition (this is done also at the gene level), Cufflinks will use the Student t Test to find transcripts that are differentially expressed. Eventual changes in abundance are calculated by Cuffdiff using the square root of the Jensen-Shannon divergence. As output, Cufflinks generates data which, together with TopHat outputs, is used to make a statistical estimate of abundance of transcripts. The latter operation is performed using Cuffdiff which uses the square root of the divergence generated above to estimate the abundance between samples and assigning p and q values.

2.10.5 Statistical analysis with Cuffdiff

One of the biggest challenges of next generation sequencing is the analysis of the large amount of data. Although over the years costs and time of sequencing have considerably decreased, analysis has become more challenging. Different software

packages have been developed for the statistical analysis of next generation sequencing data. For the dataset used in this study, Cuffdiff was employed, which is part of the Cufflinks package. Cuffdiff is designed to search for significant changes in transcript expression, splicing, and promoter usage as well as levels of gene expression. It firstly determines for each transcript/gene the mean and variance of the fragment count within a certain condition and then it adopts the t test as a statistical method to detect differentially expressed (DE) transcripts and genes (Student's T-test). It does so by assigning a log ratio between two conditions being compared. It assigns a cut off above which a transcript/gene can be called DE. It then converts the t test into p values: below a certain set value, that change in transcript/gene level will be considered significant. However, having had to deal with multiple comparisons (thousands of transcripts/genes) there is a high chance to incur type I errors, which is when a transcript/gene is detected as a DE when it is actually not altered. For this reason Cuffdiff, rather than using the p values as a tool to assign significance, adopts the q value, which is the p value FDR-adjusted (FDR, false discovery rate) (Benjamini et al, 2001). The FDR is defined as "the proportion of false positives among all of the genes identified as being DE, which is among the entire rejected null hypothesis" (with the null hypothesis being that there would not be a difference in expression between two conditions) (Noble, 2009). For the dataset used in this experiment, an FDR = 0.05 was adopted, which means approximately 5% of the genes/transcripts considered by Cuffdiff as being statistically significant after the correction are actually found by chance and are false positives.

Chapter 3. Analysis of transcription factor binding to the promoters of sterile transcription

3.1 INTRODUCTION

Immunoglobulin gene rearrangement during V(D)J recombination depends on higher-order chromatin structure. *Cis*-regulatory elements such as enhancers and promoters contain binding sites for transcription factors whose interaction drives sterile transcription through the antigen recombining genes. Locus contraction via looping of these chromatin domains triggers long-range interaction and supports chromatin accessibility.

Earlier studies already identified cell type specific transcription factors which bind to the enhancers of sterile transcription of both heavy and light (kappa) chains such as CTCF, Pax5, E2A, YY1, OBF-1, Ikaros and IRF4. Ikaros, YY1 and Pax5 are involved in IgH locus contraction and in their absence locus rearrangement is reduced (Fuxa et al, 2004; Meldrum, 1991; Reynaud et al, 2008), suggesting that cooperation between transcription factors facilitates long range interactions necessary for both VH-DH and Vκ-Jκ rearrangement (Jhunjhunwala et al, 2009).

An important role is played by CTCF whose main function is to prevent inappropriate interactions between regulatory elements. This was identified for example at the imprinted H19-Igf2 locus and the mouse β-globin locus (Phillips & Corces, 2009; Splinter et al, 2006). Indeed, CTCF binding sites were mapped across the immunoglobulin loci (Degner et al, 2011; Partin et al, 1990). Specifically, two CTCF binding sites at the Igh VH-DH locus were identified and implicated in the control of lineage specific V(D)J recombination and also in organizing the VH and DH regions into distinct chromatin domains. As found for the other transcription factors, knockdown of CTCF resulted in a decrease of Igh locus contraction (Degner et al, 2011).

At the Igκ locus, V to J recombination occurs at the pre-B cell stage under the control of two enhancers, the intronic enhancer κEi and the 3' enhancer, κE3'. At this

locus, CTCF limits the interactions between the κ enhancers and the proximal Vκ genes and also prevents inappropriate contacts with elements outside the Igκ locus (Soler et al). Earlier studies also identified E2A to be crucial for the activation of this locus: in E2A-deficient mice, Igκ germline transcription is blocked and consequently accessibility to the locus by the recombinase machinery is severely reduced (Lazorchak et al, 2006). Since E2A interacts with IRF4, inhibition of IRF4 activity also leads to reduced germline transcription and accessibility and neither of the two transcription factors was able to support adequate levels of sterile transcription alone, pointing to the importance of this interaction during Igκ locus activation.

Previous studies also identified transcription factors that bound to the two enhancers at the Igλ locus, Eλ3-1 and Eλ2-4 (Eisenbeis et al, 1995). The B cell specific transcription factors PU.1 and IRF4 were found to bind not only the lambda enhancers but also the κE3' enhancer. The results also showed that IRF4 was present in the complex and could bind the DNA only in presence of DNA-bound PU.1 when this factor was phosphorylated at serine 148 (Eisenbeis et al, 1993; Eisenbeis et al, 1995; Pongubala et al, 1993).

While the transcription factors binding to the enhancers have been elucidated for both heavy and light chains, the promoters of sterile transcription of the Igλ light chain loci have not been extensively studied. In particular, the transcription factors binding to the promoters of the J lambda genes are not known. The murine lambda locus is a good model since it only has four J gene segments whose transcription is under the control of the Eλ3-1 and Eλ2-4 enhancers. Of the four gene segments, 60% of recombination involves the Jλ1 gene segment, followed by Jλ2 and Jλ3. Jλ1 recombines mainly with Vλ1 (Reilly et al, 1984). Sterile transcription through these gene segments will open up the tight chromatin structure to allow binding of the RAG proteins. To this end, interaction between the promoter and the enhancer is crucial to allow delivery of RNA pol II to the promoters; however, what drives the specificity of this interaction is not known. A possible hypothesis is that binding of

transcription factors to the enhancers and the promoters specifically drives this interaction to trigger sterile transcription through the gene segments.

3.2 RESULTS

3.2.1 Bioinformatics analysis of the transcription factor binding sites at the promoters of sterile transcription

A bioinformatics search analysis using the TESS program (Transcription Element Search System, www.cbil.unpenn.edu) was carried out to identify the different transcription factors which potentially bind to the Jλ1, Jλ2 and Jλ3 promoters of sterile transcription. The sequences selected were submitted for analysis and only B cell-specific transcription factors were considered which were expressed and active at the pre-B cell stage (stage at which light chain recombination occurs) (Figure 3-2). To calculate the length of the sequence to be analysed for each promoter the previously mapped transcription start site (Engel et al, 2001) was used as a reference. The mouse transcription initiation sites for the J-C germ line transcripts had previously been characterized by rapid amplification of the 5' cDNA ends (RACE) (Engel et al, 2001). This showed transcripts initiating at 293, 116 and 73 bp upstream of the Jλ1, Jλ2 and Jλ3 gene segments, respectively. Taking into account these start sites, a region 260 bp upstream and 40 bp downstream of the start site of sterile transcription of the Jλ1 gene segment, 230 bp upstream and 70 bp downstream of the Jλ2 gene segment and 200 bp upstream and 40 downstream of the Jλ3 gene segment was the focus of the bioinformatics search and further analysis.

The bioinformatics search identified several B cell transcription factors such as E47, E12/47, IRF1/2, LEF-1, PU.1, and general transcription factors such as TBP, TFIID, TFIIA, TFIII (Figure 3-2). However, for further analysis only a few were considered, mainly based on data from previous research. For example the transcription factor LEF-1 was excluded since it is degraded at the pre-B cell stage, thus it could not trigger enhancer-promoter interactions at the lambda locus. Instead, particular attention was given to E47 that had previously been shown to interact with IRF4 at the kappa locus (Figure 3-1). While initial analysis focused on E47, none of the other transcription factors were excluded.

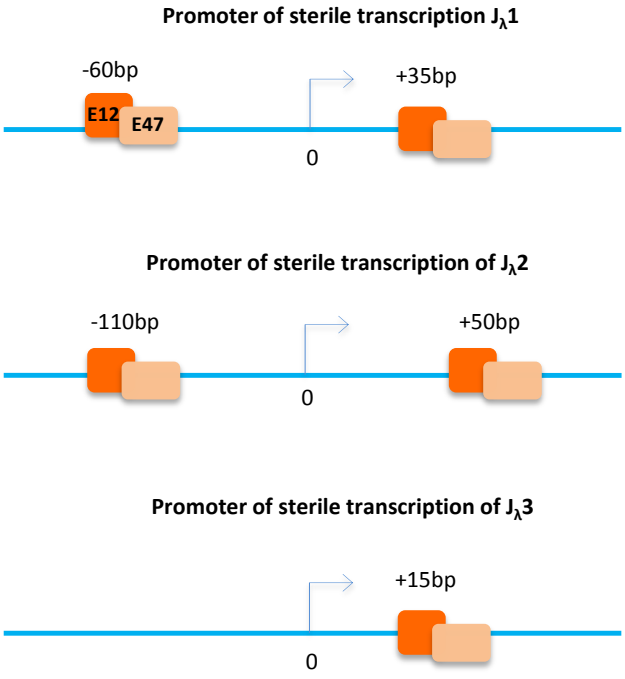


Figure 3-1 Position of E47 putative binding sites relative to the start site.

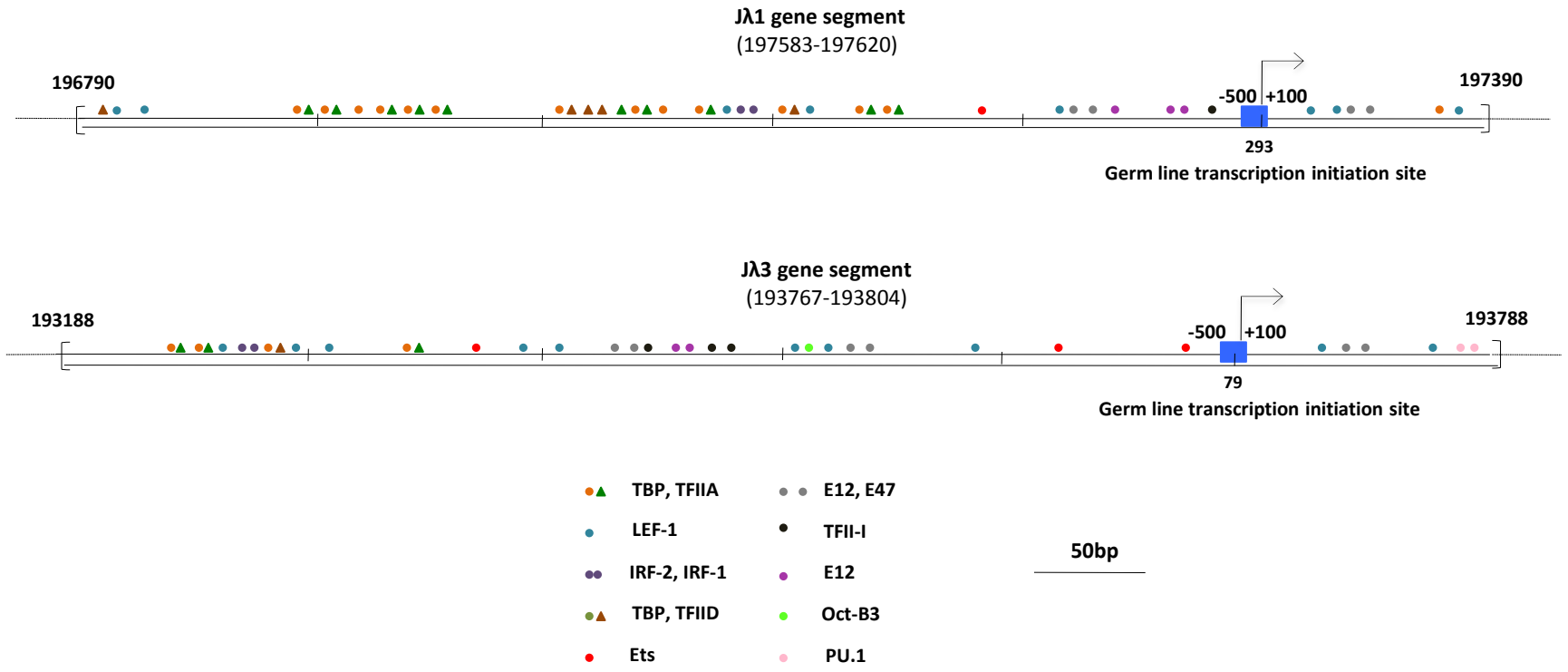


Figure 3-2 Distribution of transcription factor binding sites on the Jλ1 and Jλ3 promoters.

The numbers in brackets indicate the region of the gene segments that were taken in consideration for TESS analysis: starting from the transcription start site, 500 bp upstream and 100 bp downstream were considered. These sequences were taken from NG004051.1, the number in brackets represent the range of sequence taken in consideration for analysis.

3.2.2 Analysis of the Jλ1 promoter by DNase I footprinting

DNase I footprinting is an *in vitro* technique for locating the binding sites of proteins to specific regions of DNA. The technique is based on the property that the binding of a protein will protect the phosphodiester backbone of the DNA from enzymatic cleavage by the nuclease DNase I, generating a "footprint". The fragmented DNA is subject to an electrophoresis in denaturing gel and the binding sites are detected by autoradiography. By analysing the position of the footprint it is then possible to determine the DNA sequences to which the proteins potentially bind.

A DNA fragment encompassing 300bp spanning the start site of sterile transcription of the Jλ1 promoter was incubated with a protein extract derived from 122-1 cells and DNase I footprinting was performed as described in Chapter 2.8.16. Five protected regions were identified on the Jλ1 promoter (Table 3-1). The footprinting reactions were carried out on both strands and the position of the protected regions were confirmed by comparison with the G+A ladder (Figure 3-3). The sequences were submitted to the TESS internet database search and different transcription factors were found to bind to the same protected region. These were located 220bp, 190bp, 160bp, 60bp and 46bp, respectively upstream of the start site of sterile transcription of the Jλ1 promoter.

Table 3-1 Protected regions on the Jλ1 promoter by DNase I footprinting.

Protected region (TESS)	Protected sequence (5'-3')
IRF1/IRF2/AP-1	AGTTTTCTTCAATAACTAGTGAAGATAAAG
Lef-1	TTGCTGACTGTTTTATGAAG
PouF1a	TATTATACTAAACACTGAAGGATC
E47/E12*	GTGTTCTCCAACTGCTCTCTCCT
E12*	GCCTCAGATGTTTC

*** The putative E47/E12 binding site (CANNTG) is shown in bold.**

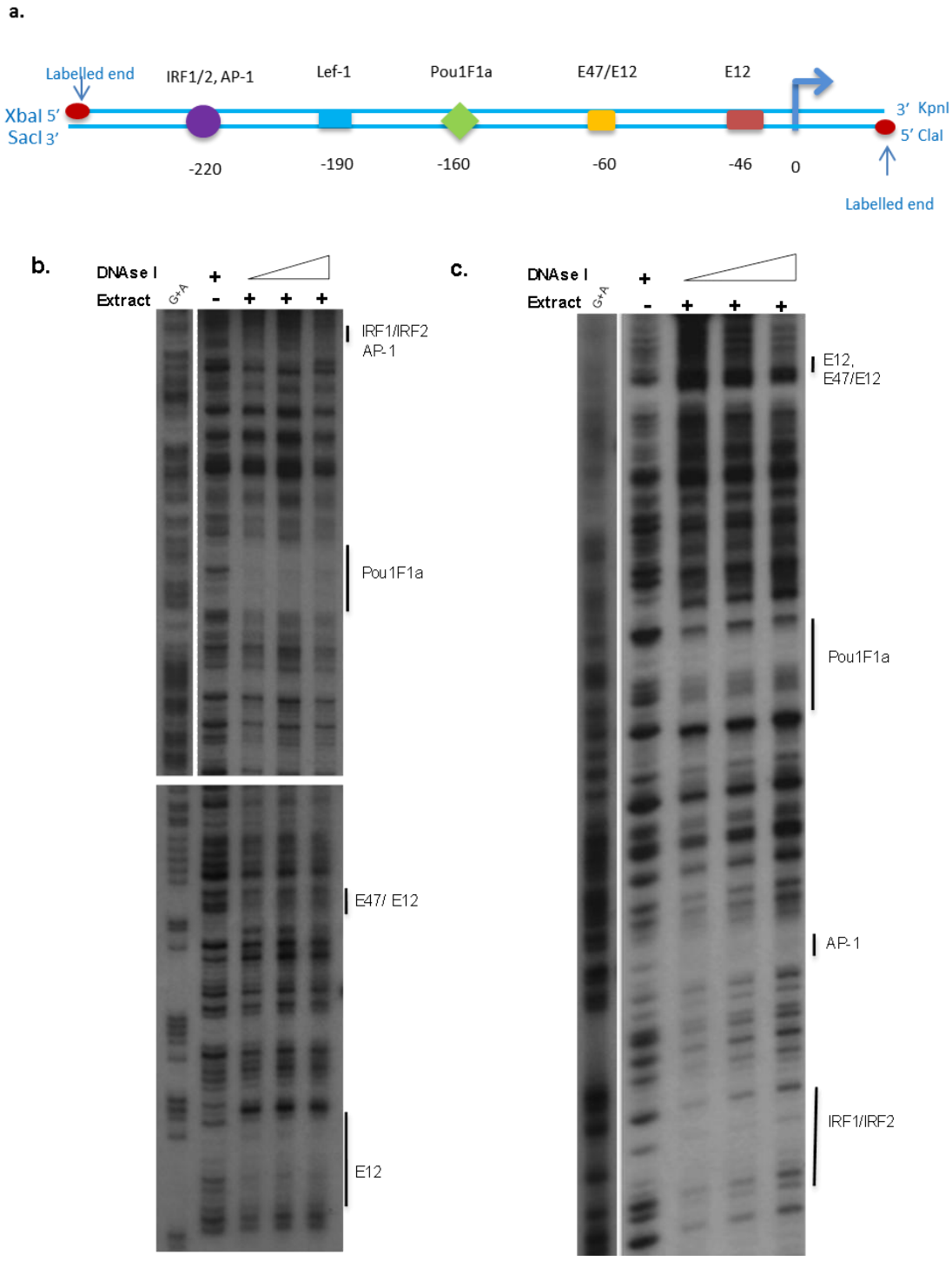


Figure 3-3 DNase I footprinting on the Jλ1 promoter.

(a) Position of the footprints in relation to the start site of sterile transcription. (b) DNase I footprinting of the lower strand of Jλ1 (500bp fragment spanning the promoter). (c) DNase I footprinting of the upper strand of Jλ1 showing the position of the same transcription factors as in the lower strand. The open arrowheads indicate increasing amounts of DNase I; G+A: guanine-adenine ladder used to identify the sequences (prepared as described in Chapter 2.8.17).

3.2.3 Analysis of the Jλ2 promoter of sterile transcription by DNase I footprinting

Although 60% of recombination at the immunoglobulin lambda light chain involves the JCλ1 gene segment, analysis of the other two J gene segments was also performed in order to understand if the same transcription factors were implicated in enhancer/promoter interaction or if different proteins were involved. In the latter case, this could explain the lower rate of recombination of JCλ2 (30%) and JCλ3 (10%) compared to JCλ1.

DNase I footprinting was performed over a region of 300 bp on the Jλ2 promoter of sterile transcription. This region contained different transcription factor binding sites according to the TESS bioinformatics search, however, only a few were taken in consideration for further analysis. The reactions were performed on both strands and the protected regions were identified by comparison with the G+A ladder generated by the Maxam-Gilbert sequencing method. Four protected regions were found on one strand of Jλ2 promoter containing binding sites for more than one transcription factor (Table 3-2).

Table 3-2 Protected regions on the Jλ2 promoter by DNase I footprinting.

Protected region (TESS)	Protected sequence (5'-3')
IRF1/2, Nk x2-5, twi	CTTCTCAAGTGAG
AP-1	TGGTGTTGAATAA
E12/E47*	**CAGCTG**TCAGCA
E12,Pou1F1a,Pou3F2d*	**CACCTG**AATTAATAT

*** The putative E47/E12 binding site (CANNTG) is shown in bold.**

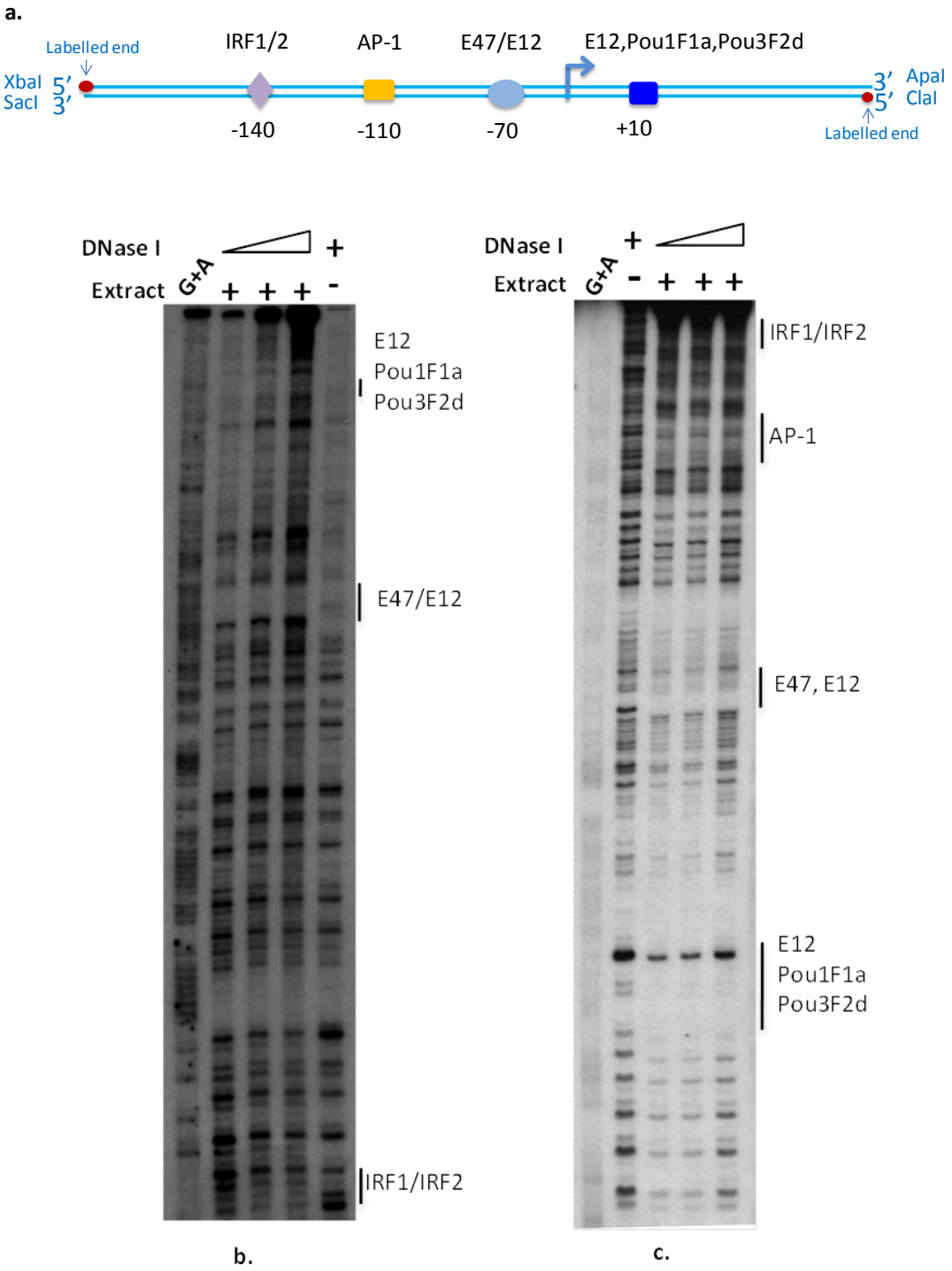


Figure 3-4 DNase I footprinting on the Jλ2 promoter.

(a) Position of the footprints relative to the start site of sterile transcription. (b) DNase I footprinting of the upper strand of Jλ2 (300bp fragment spanning the promoter). (c) DNase I footprinting of the lower strand of Jλ2 showing the position of the same transcription factors found on the lower strand. The open arrowheads indicate increasing amounts of DNase I; G+A: guanine-adenine ladder was used to identify the sequences (prepared as described in Chapter 2.8.17).

3.2.4 Analysis of the Jλ3 promoter by DNase I footprinting

The Jλ3 promoter recombines with Vλ3 gene segment and it accounts for at least 10% of the total recombination occurring at the Igλ locus. Sterile transcription at this locus is under the control of the same enhancer as at the Jλ1 (Eλ3-1), therefore it might be expected that the same transcription factors control enhancer/promoter interaction, but with a lower affinity. Initial DNase I footprinting analysis was carried out of a region of 500bp (Figure 3-2), but this was restricted to a 240 bp fragment to detect the observed footprints more clearly.

Table 3-3 Protected regions on the Jλ3 promoter by DNase I footprinting.

Protected region (TESS)	Protected sequence (5'-3')
AP-1, IRF1/2	CCCACTTCAAGTGAGGTCAC
AP-1, E12/E47*	ATTCAGTG**CAGCTG**TGAGA

*** The putative E47/E12 binding site (CANNTG) is shown in bold.**

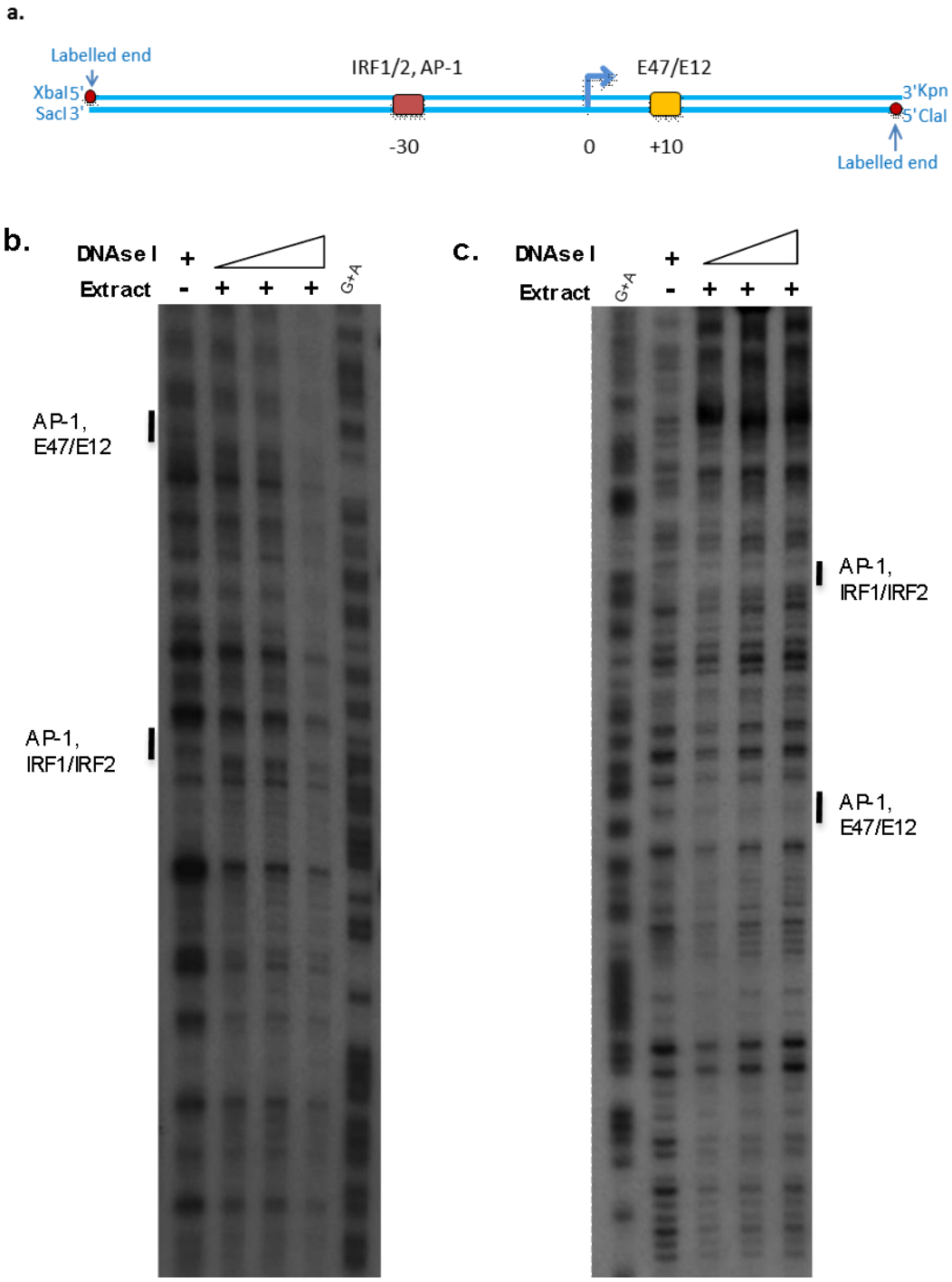


Figure 3-5 DNase I footprinting on the Jλ3 promoter.

a) Position of the footprints relative to the start site of sterile transcription b) DNase I footprinting of the lower strand of Jλ3 (240 bp spanning the promoter). (c) DNase I footprinting of upper strand of Jλ3 showing the positions of the same transcription factors found on the lower strand. The G+A sequencing ladder is indicated as are incubations with (+) and without (-) extract. The open arrow-heads indicate increasing amounts of DNase I; G+A: guanine-adenine ladder used to identify the sequences prepared as described in Chapter 2.8.17.

3.2.5 E47 binds to the Jλ1, Jλ2 and Jλ3 promoters of sterile transcription

To identify and confirm which proteins bound to these regions, radiolabelled oligonucleotides were prepared spanning the footprinted regions and used in gel retardation assays (EMSA, electrophoretic shift assay). The DNase I footprinting experiments showed that E47 is potentially able to bind to all three promoters analysed. Since E47 has been shown to interact with IRF4, it is a good candidate to mediate enhancer/promoter interactions. Therefore, of the different footprints found, oligonucleotides spanning the putative E47/E12 binding site were initially used in gel retardation assay. To test that the same factor binds to the Jλ1, Jλ2 and Jλ3 promoters, a cross-competition experiment was performed. The Jλ1 probe was incubated with a 500- to 4000-fold excess of unlabelled Jλ3 competitor and Jλ3 probe was incubated with the same concentration of Jλ1 unlabelled competitor. As shown in Figure 3-6, each oligonucleotide competed for binding to the other, suggesting that a common factor may bind to both promoters.

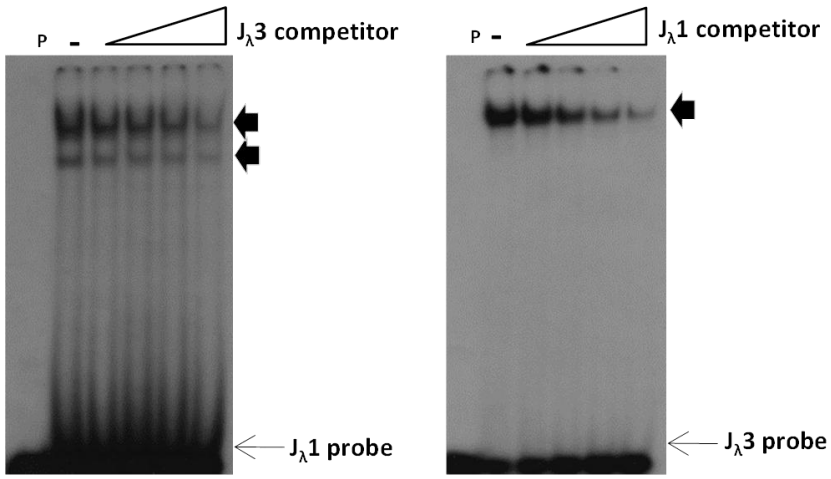


Figure 3-6 Cross competition between Jλ1 and Jλ3 promoters.

A 500 to 4000 fold excess of Jλ3 unlabelled competitor was incubated together with the radiolabeled Jλ1 probe and vice-versa. P= probe without extract added. Open arrowheads indicate increasing amount of competitor. ⬅ indicates position of specific retarded DNA: probe complexes.

Similarly, cross-competition was performed using the Jλ2 as a probe to determine whether E47 might also bind to this promoter, although it does not recombine as

extensively as the Jλ1 and Jλ3 gene segments. The data obtained showed that a similar factor did indeed bind to the Jλ2 gene segment (Figure 3-7).

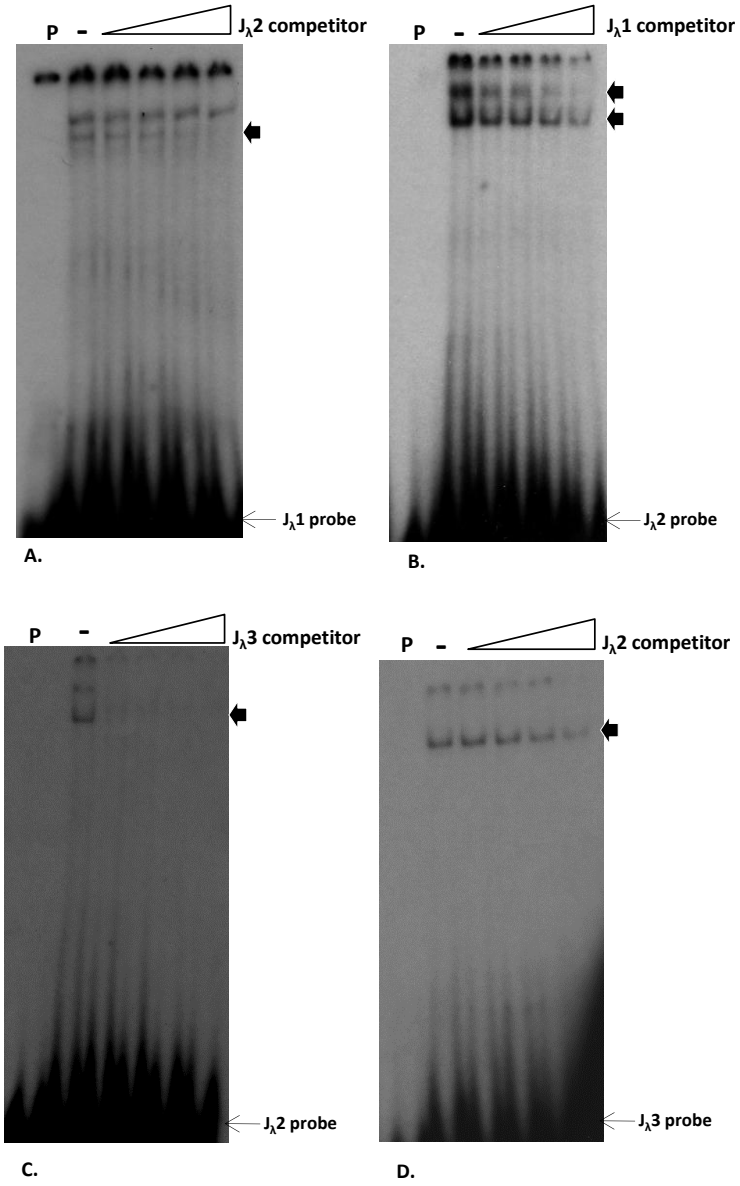


Figure 3-7 Cross-competition between the three promoters.

Same conditions as for the previous cross competition were applied to show that the same factor, more likely E47, binds to the Jλ2 promoter. P= probe without extract added. (A-B): Jλ2 and Jλ1 cross competition; C-D Jλ2 and Jλ3 cross competition. Open arrowheads indicate increasing amounts of competitor. ⬅ indicates position of specific retarded DNA: probe complexes.

3.2.5.1 Competition assay using the consensus site for E47

To identify the bound factor more precisely, competition assays were performed using specific oligonucleotides added at 1000- to 4000-fold excess to both the Jλ1 and Jλ3 promoter fragments (Figure 3-8). Unlabelled competitors containing a consensus site for E47 and E2A were used as specific competitors, while oligonucleotides containing consensus sites for EBF (early B cell factor) and AML (acute myeloid leukaemia) were used as non-specific competitors.

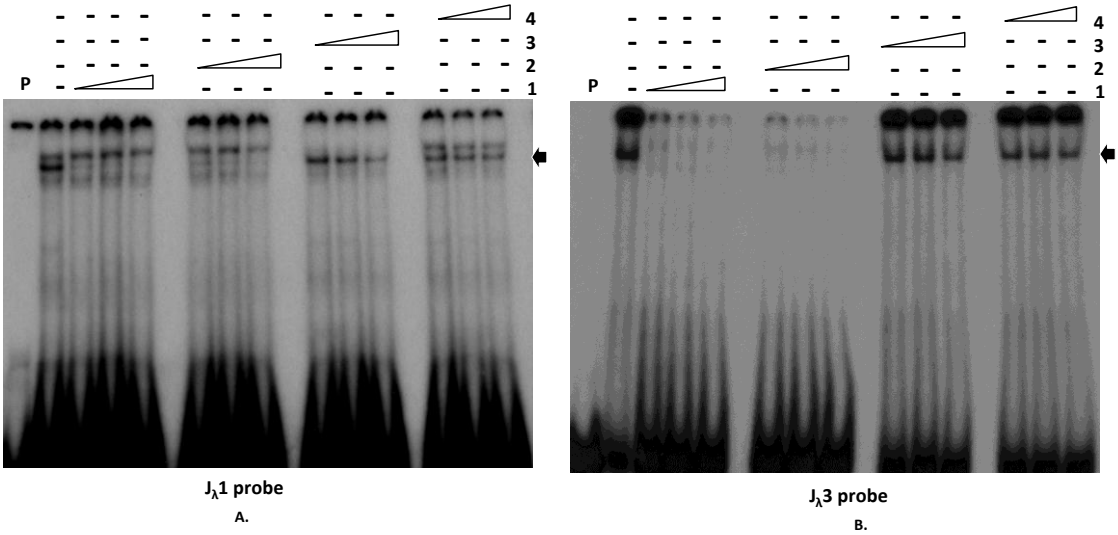


Figure 3-8 Competition assay on the Jλ1 and Jλ3 promoters.

A 1000-, 2000-, 4000-fold excess of unlabelled competitor was used to identify the factor binding to the Jλ1 and Jλ3 fragments. E47 and E2A unlabelled competitors both have a consensus binding site for E47 while EBF and AML do not. The middle band on Jλ1 (A) is competed by addition of E47 and E2A unlabelled competitors, while addition of non-specific unlabelled competitor did not affect the binding. The same was observed for the Jλ3 probe, where the only detectable band competed in the presence of E47 and E2A (B). 1= E47, 2=E2A, 3=AML, 4=EBF. Open arrowheads indicate increasing amounts of competitor. ⬅ indicates position of specific retarded DNA: probe complexes.

Figure 3-8 shows that addition of the specific competitor caused a reduction in the intensity of the middle (Jλ1) and bottom (Jλ3) complexes within the fragments analysed, suggesting that these complexes correspond to E47. In contrast, addition of the non-specific competitor did not change the intensity of the complexes, although the upper and lower complexes generated with the Jλ1 probe appeared to be bound by AML and EBF, respectively. Similarly, addition of the non-specific competitors did not affect the binding of the putative E47.

Competition with specific and non-specific oligonucleotides was also performed using the Jλ2 promoter as a probe. As shown in Figure 3-9, the top and bottom complexes were competed by the specific oligonucleotides, suggesting that two binding sites might be targeted by E2A or E47 alone.

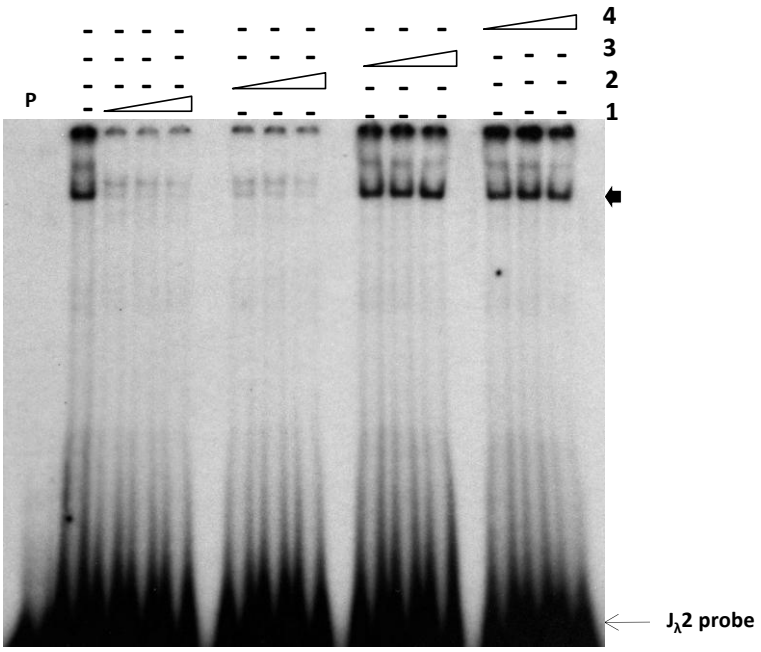


Figure 3-9 Competition assay on Jλ2 promoter.

The same conditions as for Jλ1 and Jλ3 were applied for Jλ2. 1= E47, 2=E2A, 3=AML, 4=EBF. Open arrowheads indicate increasing amounts of competitor. ⬅ indicates position of specific retarded DNA: probe complexes.

3.2.5.2 Antibody competition confirms E47 binding to all three promoter regions

To test if the complex competed by the unlabelled oligonucleotides was due to the binding of E47, an antibody competition was performed (Figure 3-10). Anti-E47-Antibody was added in increasing concentration prior to addition of the extract and incubated for 10 minutes on ice. The middle complex on Jλ1 and the top complex on Jλ2 were lost in the presence of the antibody, suggesting that these complexes are indeed due to E47 binding to the Jλ1 and Jλ2 promoters. Similar results were obtained using the Jλ3 promoter where the putative E47 candidate also seemed to be competed for binding by the antibody, although at a higher antibody concentration compared to the other two promoters.

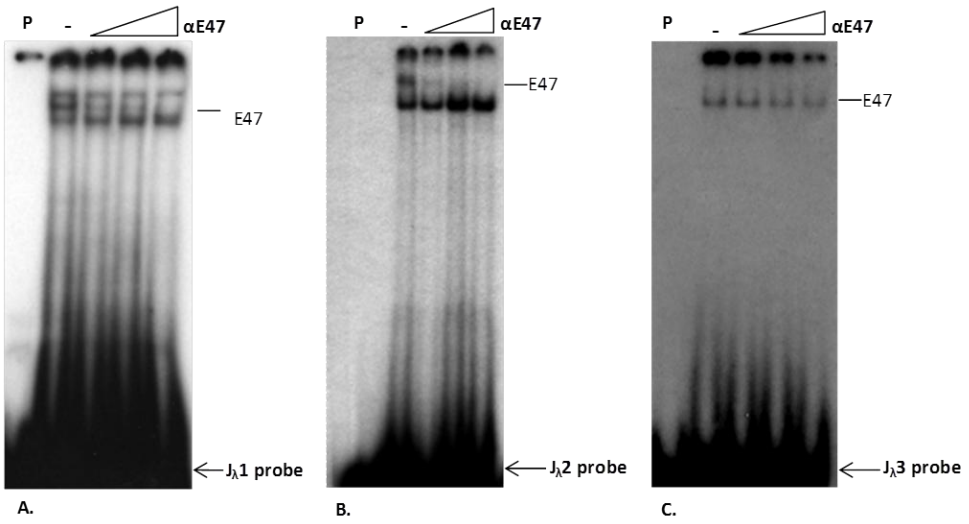


Figure 3-10 Antibody competition using the Jλ1, Jλ2 and Jλ3 promoters.

The probes were incubated for 10 minutes on ice with 2.5, 5 or 10 µl of E47 antibody (αE47) prior addition of the protein extract. The middle complex on Jλ1 (A), the top complex of Jλ2 (B) and the single complex observed on Jλ3 (C) are lost. P= probe without extract added. Open arrowheads indicate increasing concentration of anti-E47 antibody.

3.2.5.3 E47 binds with higher affinity to Jλ1

At the Igλ locus, 60% of rearrangements occur between Vλ1 and Jλ1, whereas only 30% occur between Vλ2 and Jλ2, while only 10% between Vλ3 and Jλ3. To determine if this difference in recombination frequency is related to the activation of the Jλ1 and Jλ2 promoters, the relative affinity of E47 for the two promoter regions was measured. To this end, a competition assay was performed with E2A unlabelled competitor at 30- to 1000-fold excess. As shown in Figure 3-11 A and C, the complex corresponding to E47 was competed from the Jλ3 promoter by low concentrations of unlabelled E2A, while binding persisted at the Jλ1 promoter at the same concentrations. This suggests that E47 has a higher affinity for Jλ1 than Jλ3.

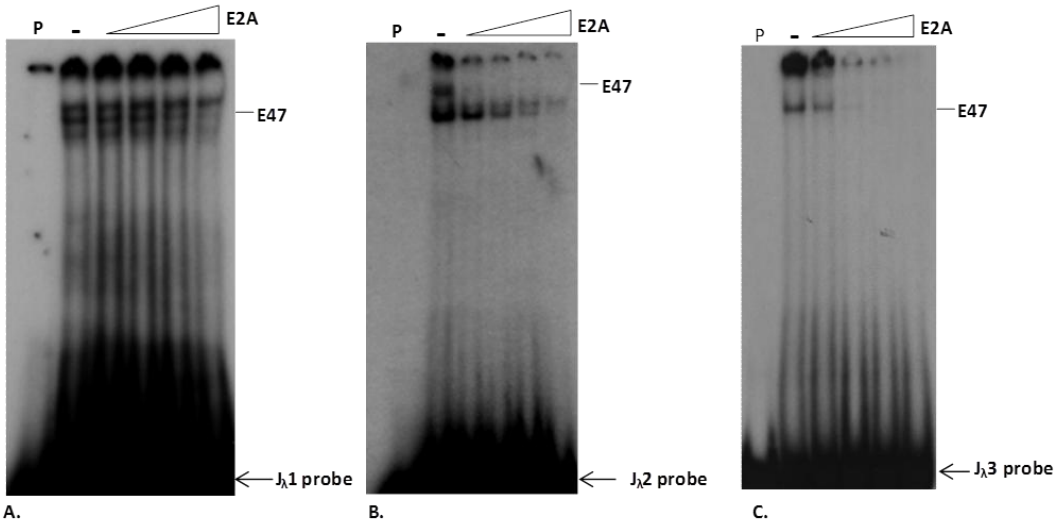


Figure 3-11 Relative affinity of E47 for the three promoters.

E2A (E12/E47) unlabelled competitor was added at 30- to 1000-fold excess to determine the relative affinity of E47 for the three promoters analysed. The complex corresponding to E47 is competed off at the Jλ2 and Jλ3 probes (B-C) at a lower concentration of E2A compared to the Jλ1 probe (A). Open arrowheads indicate increasing concentration of E2A competitor.

Similarly, a relative affinity assay was performed on the Jλ2 promoter, although this is less used during V(D)J recombination compared to Jλ1 and Jλ3. The results showed that the top complex (which in the antibody competition experiment was removed) disappeared at the lowest concentrations, as observed for Jλ3, suggesting that E47 has a higher affinity for the other two promoters analysed (Figure 3-11 B).

3.2.6 E47 interacts with IRF4 and PU.1 as a complex

The *in vitro* binding studies identified putative transcription factor binding sites near the start sites of sterile transcription and further identified E47 as a prime candidate to mediate enhancer/promoter interactions. A question of particular interest was whether this interaction is driven via IRF4. Previous studies have shown this interaction for the kappa enhancer and analysis of transgenic mice over-expressing IRF4 showed that this transcription factor alone can trigger premature V(D)J recombination of the light chain at the pro-B cell stage at the same level as in the pre-B cell stage, suggesting an important role played by IRF4 (Bevington, 2009). Therefore the next step was to test whether E47 interacts with IRF4 *in vitro*. Co-

immunoprecipitation experiments (Co-Ip) with antibodies against E47 and IRF4 were performed to identify interaction between the two transcription factors.

Initially, Co-Ip was performed by using anti-E47 antibodies in the immunoprecipitation reaction followed by immunoblotting with anti-IRF4 antibodies. As shown in Figure 3-12, an interaction between these factors was detected. Because E47 is degraded by the ubiquitin-proteasome pathway in response to phosphorylation by the MAP kinase signalling experiments were also performed to test if inhibition of E47 degradation increased the interaction with IRF4 (Figure 3-13). In the presence of both a MEK inhibitor (U0126, 10µM final concentration) and a proteasome inhibitor (MG132, 10µM final concentration), both the amount of E47 protein in the input and the immunoprecipitate was increased. This suggests that inhibition of MAPK signalling and proteasome-mediated degradation increased either IRF4 or E47, or both (data not shown).

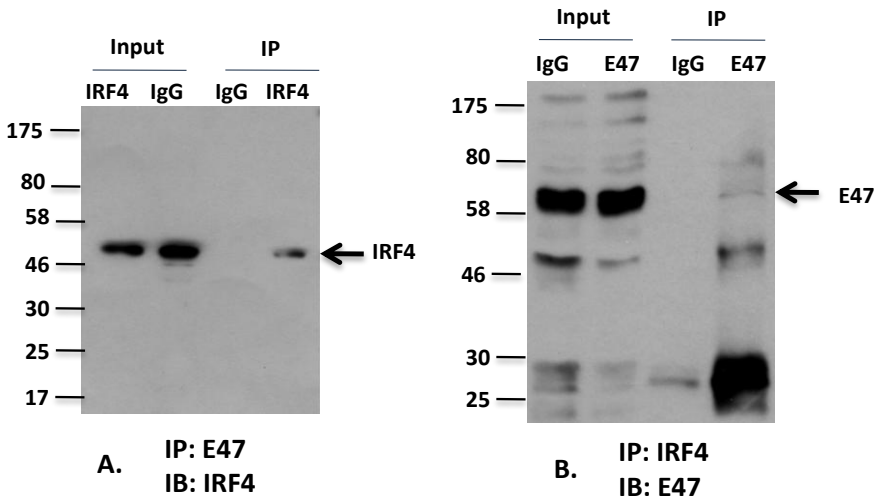


Figure 3-12 Complex formation between E47 and IRF4.

(A) Co-Ip using anti-E47 antibody as a primary immunoprecipitating antibody and immunoblotting with anti-IRF4 antibody was performed as described in the text. Interaction of the two proteins is indicated by a protein migrating at 51 kDa corresponding to IRF4. (B) Co-Ip using anti-IRF4 antibody in the immunoprecipitation reaction and immunoblotting with anti-E47 antibody. A faint band of 67 kDa corresponding to E47 was detected. Normal mouse IgG antibody was used as a negative control. In= input, IP= immunoprecipitate, IB= immunoblot.

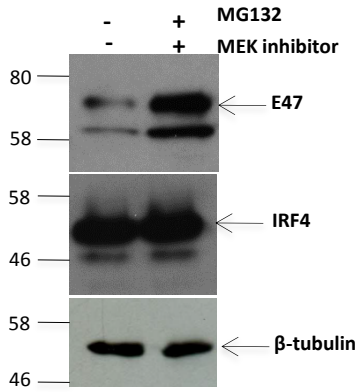


Figure 3-13 Degradation of E47 is following phosphorylation by MAP kinases.

122-1 cells were treated with the MEK inhibitor UO126, for two hours before harvesting in the presence of the proteasome inhibitor, MG132. Western blotting for E47 level and IRF4 were performed. A loading control was performed by stripping the membrane and re-probing for β-tubulin.

A reverse Co-Ip was also performed to validate the IRF4/E47 interaction. However, when probing with IRF4 and immunoblotting with E47, a very faint E47 band was detected (Figure 3-12 B). This may be due to the rapid degradation of E47 which is known to have a half-life of 55-60 minutes (Kho et al, 1997). To overcome the problem of degradation, cells were incubated with proteasome inhibitor for 8 hours and with MEK inhibitor for 3 hours before harvesting. Moreover, the same inhibitors were also added at increasing concentrations in buffers E and F used in the Co-Ip. In addition, shorter times of immunoprecipitation with the first antibody were tested (3 hours instead of overnight incubation) to overcome degradation of the protein. Unfortunately, in all cases, protein degradation was still present (data not shown).

Finally, because IRF4 forms a complex with PU.1 (Eisenbeis et al, 1993) (Chapter 1.2.5.1), a Co-IP was performed to test whether E47 also interacts with PU.1. This interaction was also confirmed (Figure 3-14), suggesting all three proteins form a protein-complex when the enhancer loops over to interact with the promoter.

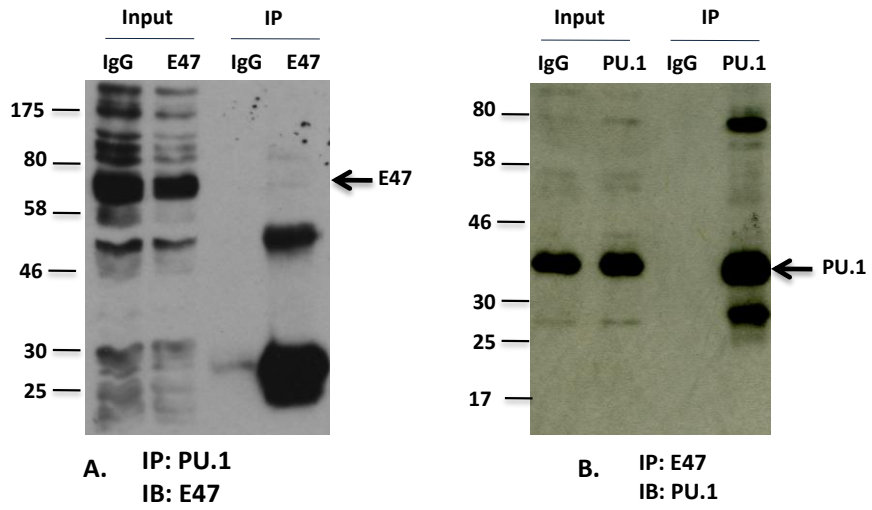


Figure 3-14 Co-Immunoprecipitation between E47 and PU.1.

(A) Co-Ip using anti-PU.1 antibody as a primary immunoprecipitating antibody and immunoblotting with anti-E47 antibody was performed as described in the text. Interaction of the two proteins is indicated by a faint band migrating at 67 kDa corresponding to E47. (B) Co-Ip using anti-E47 antibody in the immunoprecipitation reaction and immunoblotting with anti-PU.1 antibody. A strong band of 31-35 kDa corresponding to PU.1 was detected. Normal mouse IgG antibody was used as a negative control. In= input, IP= immunoprecipitate, IB= immunoblot.

3.3 DISCUSSION

The data presented in this Chapter show that E47 binds to the promoters of Jλ1, Jλ2 and Jλ3 of sterile transcription *in vitro* and also that E47 can interact with IRF4 and PU.1 *in vitro*. These results suggest that the protein complex formed between E47 and IRF4 could mediate interactions between the enhancer and the Jλ promoters of sterile transcription as previously shown for the immunoglobulin kappa locus. Similar experiments such as electrophoretic mobility shift assay where performed to identify factors binding to the region of interest (Haque, 2004; Haque et al, 2013).

The Eλ3-1 enhancer and the Jλ promoters are separated by several kilobases, thus there is a major question of how this interaction is regulated. The IRF4-E47 interaction has been previously investigated and found to be responsible for activation of the immunoglobulin κ germ line sterile transcripts (Lazorchak et al, 2006). Furthermore, IRF4 interaction with E47 enhanced E2A DNA binding and also resulted in strong E2A-IRF4 transcriptional synergy and the amino acids necessary for this interaction were identified (Nagulapalli & Atchison, 1998). Using protein-DNA models to look at both transcription factors when bound to the DNA as well by mutagenesis experiments, it was found that critical amino acids involved in the interaction were Arg 567 or Asp 568 in the HLH domain of E47 and Leu 24 at the N-terminus of IRF4. A possible mechanism was proposed to explain this interaction, namely that IRF4 induces conformational changes in E47 to enhance E47 DNA binding. Once E47 is bound, it triggers a conformational change in IRF4 (Nagulapalli et al, 2002). However, as shown in this study, the E2A and IRF4 binding sites lay immediately adjacent to each other. Therefore, a key question is how this interaction occurs when the proteins are bound to sites that are far apart: in the case of the Eλ3-1 enhancer and the Jλ promoters, the enhancer and the promoter must be brought together over distances of 35-40 kb. A major finding reported in this Chapter is that these proteins interact with each other *in vitro*. Further studies should use transfection

experiments to determine whether these proteins trigger enhancer/promoter interactions *in vitro* and whether the same regions, as identified above, are involved.

The data presented in this Chapter support previous studies that showed E47 and E12 play a role in the regulation of sterile transcription of the Jλ promoters (Bain et al, 1997). Studies in pro-B cells from E12-/+ and E47-/+ mice showed that Vλ1-Jλ1 rearrangements were severely affected by reduction in these proteins, suggesting that both transcription factors are essential to promote V-to-J rearrangement and sterile transcription at the Igλ locus (Beck et al, 2009; Quong et al, 2004). Further analysis revealed that E47 and E12 regulate Igλ locus rearrangement by modulating locus accessibility to the recombinase machinery. In particular, it was suggested that E2A proteins might modulate the level of methylation of lysine 4 of histone 3 (H3K4me3) which is essential for the recruitment of RAG2 via its PHD domain (Beck et al, 2009).

Regulation of the Igλ locus at the pre-B cell stage by E47-IRF4 interaction is also consistent with the levels of these two transcription factors during B cell development: IRF4 levels start to increase during the pro-B to pre-B cell transition (Lu et al, 2003). Conversely, at the pro-B cell stage, E47 levels are high, but the protein is either sequestered by the Id proteins (Inhibitors of differentiation) to form inactive heterodimers (Quong et al, 1993) or is degraded via the ubiquitin-proteasome pathway (Nie et al, 2003). At the pre-B cell stage, Id proteins levels decrease and E47 levels increase (Beck et al, 2009). This suggests that at the pro-B cell stage there is no interaction because of the lower IRF4 levels and the high levels of Id proteins. However, at the pre-B cell stage, both transcription factors may be free to interact and promote Igλ locus rearrangement. Consistent with this hypothesis, knock-out of E47 at the pro-B cell stage leads to a decrease in Igλ locus rearrangement (Beck et al, 2009), while, as shown using transgenic mice, over-expression of IRF4 at the pro-B cell stage stimulates premature lambda light chain recombination, even in the presence of high levels of Id proteins (Bevington, 2009). This suggests an interesting hypothesis in which IRF4 and E47 might be crucial to promote Igλ locus activation at

the correct stage of B cell development. This might be accomplished either by IRF4 sequestering Id proteins or by increasing E47 proteins levels.

3.3.1 Regulation of E2A protein stability

E2A proteins are known to bind only as a homodimer in B cells, although they can form heterodimers with other E-box binding proteins, for example with MyoD to regulate muscle differentiation in other tissues (Lassar et al, 1991; Shen & Kadesch, 1995). B cell restricted DNA binding of E47 proteins as homodimers depends on the phosphorylation state of the protein. Previous studies identified two serine residues at the N terminus of E47 as targets for phosphorylation: hypophosphorylation would allow E47 binding as a homodimer, while hyperphosphorylation would lead to heterodimer formation (Morcillo et al, 1996). Because they are ubiquitously expressed, E12 and E47 must be regulated mainly by post-translational mechanisms which include degradation of E2A proteins by the ubiquitin-proteasome pathway, sequestration of E2A proteins by the Id proteins to form non-DNA-binding dimers and phosphorylation of E47 upstream of its the basic region to inhibit DNA binding.

Pulse-chase analysis revealed that E12/E47 proteins have a half-life of only 55-60 minutes *in vivo* (Kho et al, 1997). Consistent with this, a previous study identified a ubiquitin-conjugating enzyme, UbcE2A, to be an interacting partner of E2A proteins. UbcE2A specifically interacts with a 54-amino acid region (residues 477-530) upstream of the bHLH region of E47 and E12. This interaction leads to protein degradation via the ubiquitin proteasome pathway (Kho et al, 1997). Two highly conserved regions are involved in the degradation: residues 479-494 and 505-513, where the latter contains a PEST sequence rich proline (P), glutamic acid (E), serine (S), and threonine (T) (Huggins et al, 1999). In fact, deletion of this region greatly increased E12/E47 stability. Consistent with these findings, treatment of cells with a proteasome inhibitor, MG132, reduced E12/E47 degradation.

In this study however, reverse co-immunoprecipitation, also in presence of the proteasome inhibitor, did not give similar results, suggesting that other mechanisms

may operate to regulate E47 protein turnover. Because the half-life of these proteins is only 1 hour, further experiments could determine if incubation times shorter than 1 hour might prevent rapid degradation of the proteins or if there are other mechanisms that can be considered to prevent protein degradation.

Notch signalling has been found to be important in targeting E47 protein for degradation via the ubiquitin-proteasome pathway (Ordentlich et al, 1998). This pathway is dependent on phosphorylation of the E-box protein by the p42/p44 MAP kinase (Nie et al, 2003). The target of this phosphorylation resides between the residues found in the EHD3 domain (E protein homeobox domain), upstream of the bHLH (residues 345-367). These residues have been mutated and will be tested in the future to establish their role in E47 stability.

3.3.2 Other factors bind to the promoter region of the Jλ gene segments

DNase I footprinting also identified a protected region in the Jλ1 promoter that could be bound by other factors for example, LEF-1, Pou1F1a, IRF1/IRF2 and E12. LEF-1 was not investigated further since its levels fall dramatically during the pro-to-pre-B cell transition (Schebesta et al, 2002). Hence it is unlikely to trigger enhancer/promoter interactions to activate recombination of light chains that take place at the pre-B cell stage.

Gel retardation experiments using unlabelled competitor DNAs showed that the Pou1F1a binding site may be bound by the octamer binding protein (Figure 7-1). Octamer binding proteins have been found in most Ig promoters suggesting a role of these proteins in determining B cell specificity or regulation of heavy or light chain transcription since their deletion led to promoter inactivity (Eaton & Calame, 1987; Falkner & Zachau, 1984). Among these factors, Oct-2 is mainly expressed in B lymphocytes. Therefore, this is a strong candidate to activate sterile transcription at the Jλ promoters. To date, interactions between Oct-2 and IRF4 have not been described; this could be tested in future work.

IRF1/2 sequences were also found to be protected in DNase I footprinting experiments in all three promoters analysed. However, when these protected regions were tested in gel retardation experiments, competition assays failed to reveal the identity of the bound factor (data not shown). While cross-competition confirmed that the factor binding the three promoters might be indeed the same, antibody competition assays did not yield convincing results. This suggests that other, non-B cell specific transcription factors may bind to this footprinted region and it is possible that the factor that binds to this region is not IRF1/2 (Figure 7-3;Figure 7-2).

3.3.3 E47 undergoes conformational changes following interaction with PU.1/IRF4

Interestingly, co-immunoprecipitation experiments in which was immunoprecipitated either IRF4 or PU.1 followed by Western blotting E47 showed a very faint band corresponding to E47 (67 kDa), suggesting instead some degree of degradation. In addition, two bands detected corresponding to approx. 28 and 48 kDa which could potentially correspond to the light and heavy chain of the immunoglobulins were detected; they might mask E7. To test whether E47 was degraded, the assay was performed on a non-denaturing gel, which would preserve the native conformation of the E47 protein, ruling out the possibility of degradation. A Co-Ip was performed between E47 and p300 (histone acetyltransferase) and the results showed that the E-box binding protein was not degraded, instead, E47 might undergo some conformational changes when interacting specifically with IRF4. Preliminary data also showed that in non-denaturing conditions E47 interacts with IRF4 without any evidence of degradation (data not shown).

Chapter 4. Effect of mutations on enhancer/promoter interactions

4.1 INTRODUCTION

Coordinated recruitment of the basal transcription machinery is a key step during transcription of genes in eukaryotic cell, and the role played by promoters, enhancers and locus control regions is now well established. However, when analysing the transcription of immunoglobulin genes, other factors have to be taken in consideration. As stated in Chapter 3, while enhancer/promoter interactions are crucial for sterile transcription of the loci, particular emphasis is given to the transcription factors driving this interaction and any play a key role in promoting enhancer looping towards the promoter (Lieberman-Aiden et al, 2009).

The advent of the 3C technique has increased our knowledge of long-range interactions and modification of the 3C, notably 4C and 5C, have confirmed the nature of several interactions occurring within the human genome (Dekker et al, 2002), including the immunoglobulin loci (Ju et al, 2007). Within the immunoglobulin lambda light chain, long range interactions have been confirmed by 3C (Bevington, 2009) and earlier studies on the immunoglobulin kappa light chain locus have also analysed a series of mutations within the transcription factors involved in mediating these interactions (Brass et al, 1999).

In this Chapter the importance of specific residues within the enhancer and the promoter of sterile transcription as well as residues within the transcription factors involved in the interaction between these two regulatory elements were investigated to understand to what extent interaction between the specific transcription factors is required to trigger transcription.

It has already been reported that specific residues involved in the interaction between E47, PU.1 and IRF4 transcription factors are crucial for different Ig loci: single amino acid mutations in IRF4 can affect the interaction with E47 (IRF4L24 and IRF4 L368) or PU.1 (IRF4 K399 as well as IRF4 L368) (Brass et al, 1999; Meraro et al,

1999). IRF4 L368 and IRF4 K399 are found at the C-terminal region of the protein, within the IRF4 association domain (IAD) (Figure 4-1). These residues were reported to be essential in mediating the interaction with PU.1 at the λB element and with other members of the IRF family. In addition, the IRF4 L368A mutant affected interaction with E47, driving lower levels of transcription compared to wild type IRF4. The L24 residue of IRF4 was described to be important at the kappa locus in mediating contacts with E47 (Nagulapalli et al, 2002).

The major aim of the work described in this Chapter is to test these residues to investigate their importance in mediating long range interactions at the Igλ locus.

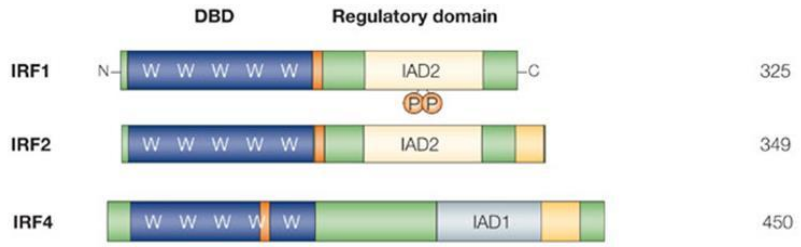


Figure 4-1 Schematic representation of the structural domains of IRF1, IRF2 and IRF4.

All interferon-regulatory factors (IRFs) are composed of a DNA-binding domain (DBD; blue) and a regulatory domain (green). For all IRFs, the DBD is defined by 5 tryptophan (W) residues that are each separated by 10–18 amino acids. Most IRFs also contain an IRF-association domain (IAD) of either type 1 (grey) or type 2 (pale yellow). Some IRFs contain a repression domain(s) (yellow) and a nuclear-localization signal(s) (orange). For IRF1 the activity depends on phosphorylation, as shown. The size of each IRF in number of amino acids is also indicated. C, carboxyl terminus; N, amino terminus. Adapted from (Lohoff & Mak, 2005).

4.2 RESULTS

4.2.1 Transcription levels in presence of Jλ1 and Eλ3-1

In Chapter 3 the protein complex driving the interaction between the Eλ3-1 enhancer and the Jλ1 promoter was identified. It was then necessary to determine to what extent this interaction could affect sterile transcription levels through the Jλ1 promoter. The other two promoters previously analysed, Jλ2 and Jλ3, were not tested in mutagenesis studies since they account only for 30% and 10% respectively of the recombination process.

A plasmid DNA construct was made in which the Jλ1 promoter drives expression of the luciferase reporter gene (Clone I). A second construct had the Eλ3-1 enhancer cloned downstream of the luciferase reporter gene and the Jλ1 promoter upstream of the luciferase gene in order to test if, following interaction between these two regulatory elements, transcription levels increased, compared to the levels observed in the presence of the promoter alone (Clone II) (Figure 4-2).

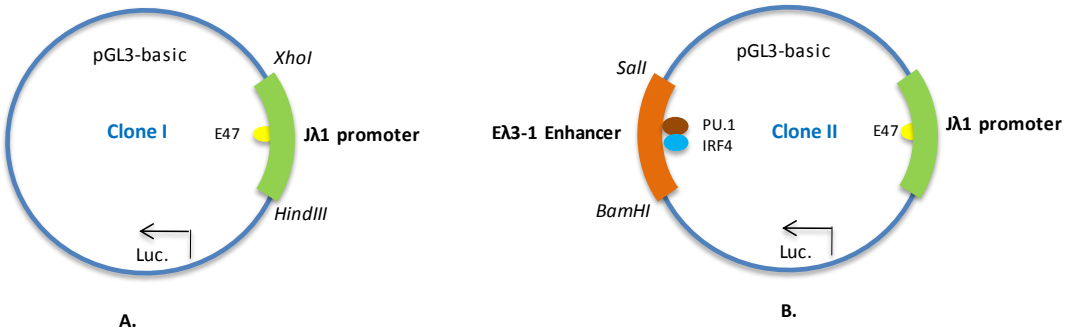


Figure 4-2 Luciferase reporter gene constructs.

A. The Jλ1 promoter containing the E47 binding site identified by DNase I footprinting reactions was cloned upstream of the luciferase reporter gene within the XhoI and Hind III restriction sites. This construct was employed to test promoter activity in the absence of the enhancer. B. The second construct contains the Jλ1 promoter cloned as described in A as well as the Eλ3-1 enhancer containing the PU.1/IRF4 binding sites cloned downstream of the luciferase reporter gene within the BamHI and SalI restriction sites. This construct was used to test enhancer activity in presence of E47, IRF4 and PU.1.

The two constructs were transfected independently into Cos-7 cells together with plasmids designed to express E47, PU.1, and IRF4. Following transfection (three independent experiments per construct analysed, each performed in duplicates), the average relative luminescence units (RLU) were measured (Original values shown in Table 7-2). Initially, it was aimed to detect differences in transcription levels with and without the Eλ3-1 enhancer.

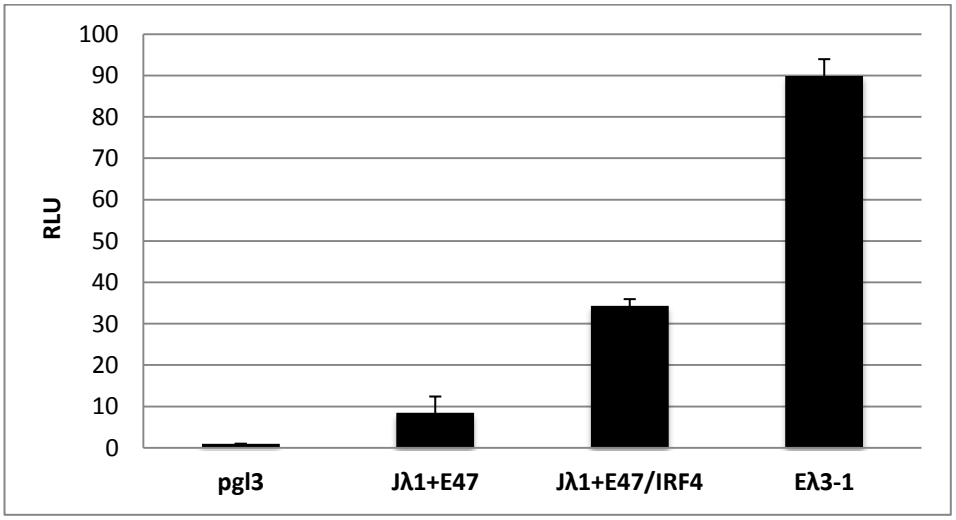


Figure 4-3 Activity of the Jλ1 in presence of the E47 and IRF4 transcription factors.

Cos-7 cells were transfected with Clone I (as described in Figure 4-2) and plasmid expressing either E47 or IRF4. Luciferase activity was measured as described in Chapter 2.8.19. The Eλ3-1 plasmid (Clone II) was always transfected in the presence of both E47 and IRF4/PU.1 protein complex. Values were divided by the co-transfected Renilla luciferase and normalised to the control group (pgl3). All experiments were repeated three times, each performed in duplicate. Expression data are shown as SEM (mean standard error) (Appendix Table 7-1)

As expected, transcription driven by the Jλ1 promoter in the absence of the enhancer did not yield high values with only a 4 fold increase compared to the promoter in the presence of both the transcription factors, E47 and IRF4 (Jλ1+ E47/IRF4) (Table 7-1). Instead, when the promoter and the enhancer were transfected in the presence of the protein complex (E47 and PU.1/IRF4) transcription levels increased 10 fold, suggesting that the enhancer element had a positive effect on the rate of transcription (Figure 4-3).

Since there are no binding sites for PU.1 on the promoter and these studies were mainly focused on the E47/IRF4 interaction, transfection of the promoter construct in the presence of PU.1 was not performed.

4.2.2 Effect of mutations in the E47 binding site on Jλ1

The data presented above show that E47 binds to the promoter and data from Chapter 3 also showed it interacts with both PU.1 and IRF4. But does it need to bind the promoter to drive this interaction? Therefore, to test if the increase in transcription observed was a consequence of the interaction between the enhancer-bound PU.1/IRF4 and the promoter-bound E47, the E47 binding site in the promoter (Eλ3-1/Jλ1 E47mt) was mutated in the construct that carried both promoter and enhancer (Clone II). In this way it was possible to test if the lack of E47 binding to the promoter still allowed long-range interactions and therefore triggering of sterile transcription through the gene (Figure 4-4).

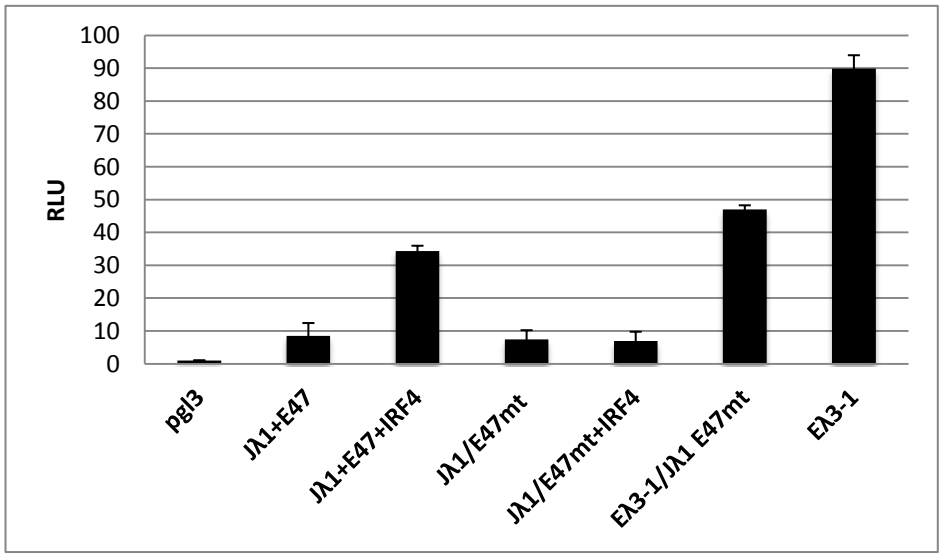


Figure 4-4 Transcription in presence of mutations within the promoter.

Cos-7 cells were transfected with Clone I as described in Figure 4-2 and a plasmid with a mutation in the E47 binding site within the promoter. Luciferase activity was measured as described in Chapter 2.8.19. The Eλ3-1 plasmid (Clone II) was always transfected in the presence of both E47 and IRF4/PU.1 protein complex. Values were divided by the co-transfected Renilla luciferase and normalised to the control group (pgl3). All experiments were repeated three times, each performed in duplicate. Expression data are shown as SEM (Appendix Table 7-1).

While no significant differences between the wild type and mutated promoters in the presence of E47, a 4.6 fold decrease was observed when comparing the wild type promoter with the mutated promoter in the presence of the E47/IRF4 complex. This suggests that, although not bound to the enhancer, IRF4 still interacts with the promoter, either through E47 or other transcription factors. As expected, transfection of the Eλ3-1 enhancer together with the mutated promoter triggered a two-fold reduction in transcription levels.

Overall, this data suggests that the interaction between IRF4 and E47 is critical to maintain adequate levels of transcription and binding of E47 to its site on the promoter is crucial to trigger first the interaction with IRF4 and secondly transcription through the Jλ1 promoter. However, since the reduction in transcription levels was not as high as expected, it may be proposed that other factors are equally important in driving this interaction.

4.2.3 The effect of mutations in the IRF4/PU.1 binding sites in the enhancer region

Having defined the role played by E47, it was important to investigate the role of the enhancer-bound PU.1 and IRF4. Therefore, the binding site for IRF4 (Eλ3-1/IRF4mt) and PU.1 (Eλ3-1/IRF4mt+PU.1mt) on the enhancer were also mutated and further investigated in luciferase assays (Figure 4-5).

The data obtained showed that sterile transcription decreased considerably when comparing the wild type Eλ3-1 enhancer with one in which mutations had been introduced. A 2.1- and 2.4-fold decrease for the Eλ3-1/IRF4^{mt} and Eλ3-1/IRF4^{mt}+PU.1^{mt} were found, respectively. These results suggest that IRF4 needs to bind to its binding site on the Eλ3-1 enhancer to trigger adequate levels of transcription. Since it known that IRF4 is the PU.1 interacting partner, it is expected that mutations in the PU.1 binding site would affect the rate of transcription.

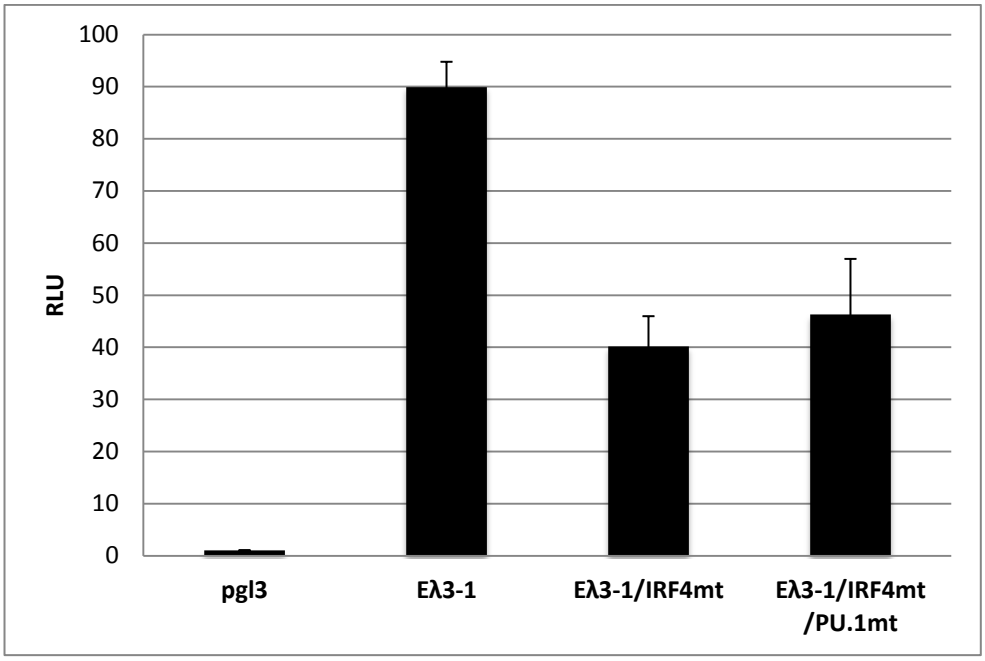


Figure 4-5 The effect of mutations in the IRF4 and PU.1 binding sites in the enhancer region.

Cos-7 cells were transfected with Clone II as described in Figure 4-2 with mutations in the IRF4 or PU.1 binding site in the Eλ3-1 enhancer. Luciferase activity was measured as described in Chapter 2.8.19. The Eλ3-1 plasmid (Clone II) was always transfected in the presence of both E47 and IRF4/PU.1 protein complex. Values were divided by the co-transfected Renilla luciferase and normalised to the control group (pgl3). All experiments were repeated three times, each performed in duplicate. Expression data are shown as SEM (Appendix Table 7-1).

However, no significant decrease in transcription following mutation of the PU.1 binding site within the same construct carrying the Eλ3-1/IRF4^{mt} mutation was detected, as it might have been expected. Previous studies have shown that IRF4 can bind to the DNA only in the presence of bound PU.1 (Brass et al, 1999). Therefore, when PU.1 is unable to bind DNA, it should not recruit IRF4, and thus transcription should decrease to lower levels than those observed when IRF4 is not bound. The data obtained, however suggested that IRF4 binding is independent of PU.1, but is essential to mediate interaction with E47 to trigger transcription of the Jλ1 gene segment. As outlined above, other transcription factors are believed to be essential for this interaction.

4.2.4 The IRF4 L24A, IRF4 L368A and IRF4 K399A mutants do not affect protein-protein interaction but affect sterile transcription levels

Specific residues within the transcription factor IRF4 are known to be essential for productive interaction with both E47 and PU.1. To test if this was the case within the Jλ1 promoter, mutations were prepared and tested in a luciferase assay as well as in a co-immunoprecipitation assay.

Initial results showed that none of these mutations decreased protein-protein interaction when transfected in Cos-7 cells followed by co-immunoprecipitation (Figure 4-6), instead it seemed as one of these mutation (L368) stabilised the interaction; only IRF4 L24A and IRF4 L368A were tested in co-immunoprecipitation assays since IRF4 K399 is known to be important in mediating interactions with PU.1 rather than E47. The immunoprecipiation was performed by transfecting either E47 or IRF4 alone as well as the respective IRF4 mutants (IRF4 L24A and L368A).
On the other hand, when analysing transcription levels, they had a considerable effect compared to wild type IRF4 (Figure 4-7). Transfection of IRF4 L24A or IRF4 L368A mutants in the presence of the enhancer-binding site led to a 2.4 fold decrease in transcription levels for both mutants, suggesting that these two residues in IRF4 are important in making contacts with PU.1 and /or E47 to trigger transcription (Table 7-1).

IRF4 K399 is one of the residues that make electrostatic contacts with PU. 1. This mutation led to only a 1.4-fold decrease compared to the wild type IRF4, suggesting that the other residues may be important in mediating interaction with PU.1.

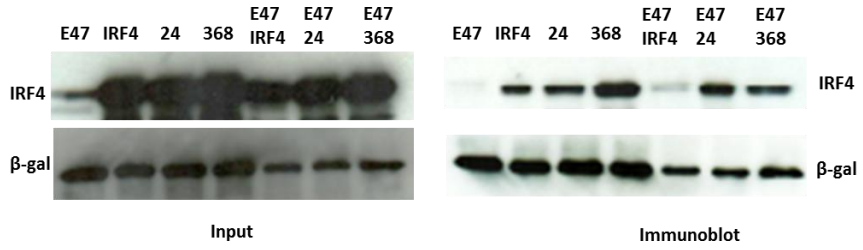


Figure 4-6 Complex formation between E47 and IRF4 mutants.

Co-immunoprecipitation was performed in Cos-7 cells transfected with E47, IRF4 or both, with or without the mutants. As loading control and to monitor transfection efficiency, β-gal was transfected together with the construct. 24 and 368 are the residues in IRF4 that were mutated into alanine. The first four lanes show the transcription factors transfected singularly. The last four lanes show IRF4 and E47 co-transfected together (IRF4 wild type and mutant forms).

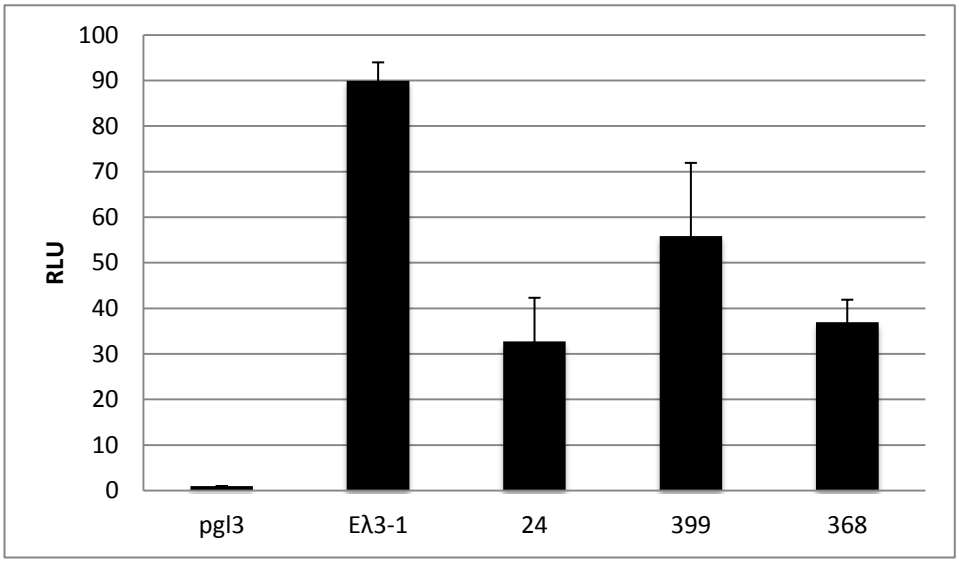


Figure 4-7 Mutations on IRF4 at residues 24, 399 and 368.

Cos-7 cells were transfected with Clone II as described in Figure 4-2. Transfections were performed using three different mutants for IRF4, namely IRF4 L24, IRF4 K399 or IRF4 L368. Luciferase activity was measured as described in Chapter 2.8.19. The Eλ3-1 plasmid (Clone II) was always transfected in the presence of both E47 and IRF4/PU.1 protein complex. The promoter-less cloning vector pGL3-Basic was used as a negative control. . Values were divided by the co-transfected Renilla luciferase and normalised to the control group (pgl3). All experiments were repeated three times, each performed in duplicate. Expression data are shown as SEM (Appendix Table 7-1)

4.2.5 Effects of mutating residues in IRF4 on promoter activity

Having tested the IRF4 mutations within the enhancer, the effect of these mutations was tested on the promoter activity in the absence of the enhancer, since ini-

tial data showed that IRF4 had some effect on the promoter, although in the absence of the enhancer as shown in Figure 4-4.

When comparing the levels of transcription in the presence of E47/IRF4 wild type versus E47/IRF4 L368A or E47/IRF4 L24A, initial data suggested a 2- and 2.6-fold decrease respectively, suggesting that IRF4 is still contacting E47 to trigger higher levels of transcription, presumably by binding within another site on the promoter and then, interacting with the enhancer-bound PU.1 (Figure 4-8, Table 7-1).

This supports the hypothesis that enhancer-promoter interactions are mediating the interaction of E47 with IRF4/PU.1. However, since Co-IP results show that the interaction in presence of the mutants is still occurring, it can be proposed that other residues may be crucial for the interaction and nevertheless, these mutants still allow the interaction to occur.

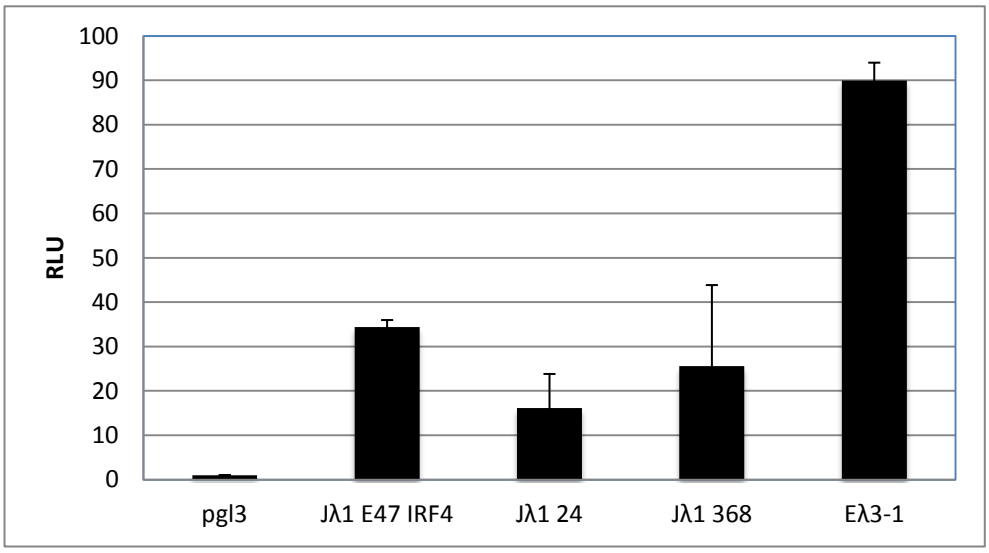


Figure 4-8 Effects of IRF4 mutations on the promoter activity.

Cos-7 cells were transfected with Clone I or Clone II as described in Figure 4-2. Transfections were performed using two different mutants for IRF4, namely IRF4 L24 or IRF4 L368. Luciferase activity was measured as described in Chapter 2.8.19. The Eλ3-1 plasmid (Clone II) was always transfected in the presence of both E47 and IRF4/PU.1 protein complex. The promoter-less cloning vector pGL3-Basic was used as a negative control. Values were divided by the co-transfected Renilla luciferase and normalised to the control group (pgl3). All experiments were repeated three times, each performed in duplicate. Expression data are shown as SEM (Appendix Table 7-1).

4.3 DISCUSSION

The data presented in this Chapter show that E47-PU.1/IRF4 interaction is necessary, but not sufficient to trigger full promoter activation and that both mutations of the E47 binding site on the Jλ1 promoter and the PU.1/IRF4 binding site on the Eλ3-1 enhancer decrease transcription levels by approx. 2- and 2.4-fold respectively. However, promoter activity is still detectable at significant levels in the presence of these mutants, suggesting that other factors might be involved in stabilizing this interaction or simply interacting with E47 at the promoter to stabilize the protein itself. Indeed, footprinting analyses of the Jλ1 promoter showed binding of other factors, among which AP-1 (in particular c-Jun and c-Fos as found in the bioinformatic search). AP-1 has already been described to be essential for transcription at the immunoglobulin kappa locus, confirming the notion, consistent with the current data, that it is unlikely that only one factor mediates these long range interactions. The binding sites for AP-1 were verified by EMSA, but antibody supershift assays using a c-Jun antibody failed to reveal the identity of the factor analysed, suggesting that other transcription factors or c-Fos might bind this region (data not shown).
In addition, mutation of the residue IRF4 K399 led only to a 1.4 fold decrease in promoter activity compared to that in the presence of wild type IRF4 (Figure 4-7). This suggests that the other residues which are known to be important in making contacts with PU.1 might be required. Previous studies, using deletion analysis and secondary structure predictions, suggested that different amino acids within the putative α-helical region of IRF4 (amino acids 399–413) at the C-terminus are important for ternary complex formation with PU.1 (Brass et al, 1999). These other residues should be studied in the future to test this hypothesis.

Interestingly, preliminary co-immunoprecipitation assay using protein extracts prepared from both a pro- and pre-B cell line (122-1 and 103 Bcl) showed that E47 is cleaved when interacting with PU.1 and IRF4 (data not shown). This suggests a novel mechanism which might regulate enhancer-promoter interactions at the

lambda locus. This cleavage reaction could either be required to activate E47 or it could be a specific “signal” to activate the promoter itself. E47 cleavage upon interaction with other proteins has not been previously described before. The E2A proteins are ubiquitously expressed and several interacting partners have been identified, such as MyoD and NeuroD1 (Lluis et al, 2005; Longo et al, 2008), however, a possible clipping reaction involving E47 has not been described, suggesting that this mechanism might be restricted only to B cells or to this specific interaction.

Nevertheless, the mechanism of this cleavage reaction is not yet clear. Initial analyses to predict the site of cleavage in E47 based on co-immunprecipitation analyses in Cos-7 cells where a N-terminally Myc-tagged E47 was used, suggested a possible Caspase 1 cleavage site at position 219 (in mouse E47), which would generate the two products observed in the Western blots. These results are consistent with the hypothesis formulated by Nagulapalli et al. (1998), who proposed that “part of the synergy mechanism between E47 and IRF4 at Ig(k) 3’ enhancer involves the ability of IRF4 to enhance DNA binding by E47, possibly by inducing a conformational change in the E47 protein”, which they also suggested is more favourable for DNA binding (Nagulapalli & Atchison, 1998). Future work should be directed towards elucidating the mechanism of this cleavage reaction to improve understanding of the role of E47 in the activation of of the J lambda gene promoters, in particular Jλ1.

Chapter 5. High throughput RNA sequencing (RNA-Seq) analysis of HPV-positive cervical cancer cell lines and tissues: search for cervical cancer biomarkers

5.1 INTRODUCTION

Viruses are associated with approx. 20% of human cancers. Cervical cancer is the second most common cancer in women, causing over 250000 deaths each year world-wide (Arbyn et al, 2011; WHO, 2008).

More than 100 different HPV types have been identified and infection with high risk HPVs (HR- HPVs) such as HPV16 and 18 is directly linked to the development of cervical cancer (Walboomers et al, 1999). Over 80% of cervical cancer cases are known to be HPV16 or HPV18 positive. Since cancer develops decades after the initial infection, it has been suggested that the virus had evolved unique mechanisms to evade the immune system, probably by altering expression of a number of cellular genes involved in immune regulation or recognition. While it is widely accepted that the two viral oncogenes, E6 and E7, alter the levels of the cellular p53 and pRb tumor suppressor proteins, respectively, de-regulation of cellular transcripts (and possible downstream effects of this) is poorly understood.

Cancer is a complex multi-step disease, which involves deregulation of multiple cellular pathways and transcripts and understanding what is directly linked to the virus infection and what is a consequence of infection is challenging (Floor et al, 2012). In addition, tumours have historically been classified and studied based on morphological appearance, as well as epidemiology. However, similar histological appearance does not necessarily mean similarity at the molecular level. Analyses of gene expression profiles, as well as identification of genetic markers can undoubtedly improve the accuracy of a diagnosis and thus treatment of the cancer.

In the past two decades, the development of microarrays and next generation sequencing (NGS) has improved our knowledge of how the expression of cellular transcript change in malignant, compared to normal cells. DNA microarrays permit the genome-wide assessment of the expression of many thousands of genes at any given time point in normal and malignant cells at any stage of development or under timed experimental conditions in cultured cells. Microarrays may be more limited than NGS technologies, but have been recognized as potential diagnostic tools in cancer.

In this study, using high throughput RNA sequencing (RNA-Seq) of HPV-positive cervical cancer tissues and cell lines (Table 5-1), differential expression of several cellular transcripts that encode putative cell-surface proteins, was detected between normal and HPV-positive cancer cells and/or tissues. Transcripts encoding cell-surface proteins were selected for detailed study because they may be involved in immune regulation or recognition.

Table 5-1 Source of the patients' data used in this study.

Patient	Age	Grade	LVSI	Stage	Nodal status	Histology
1	43	3	Y	2B	negative	SCC
2	42	2	Y	3B	positive	SCC
3	29	3	Y	2B	positive	SCC
4	59	2	N	2B	positive	SCC
5	55	2	N	2B	positive	AdenoSCC
6	57	2	Y	2B	negative	SCC
7	32	3	?	1B2	negative	SCC
CaSki	40	-	-	-	N/A	SCC
SiHa	55	2	-	-	N/A	SCC
HeLa	31	-	-	1	N/A	AdenoSCC

SCC: squamous cell carcinoma; LVSI: lymphovascular space involvement.

Up-regulated transcripts included Bst2, NKD2, GBP2, ENO2, CA9, CDH2, CDH24, while the down-regulated transcripts included FGFBP1, AMIGO2, FEZ1, GJB2, GJA1, GJB4 and GJB5. Among these, attention was focussed on connexin 26 known as GJB2 (Cx26), a gap junction (GJ) protein that mediates intercellular communication. GJs are intercellular channels that allow the free passage of ions and small molecules (<1 kDa) between cells and they are important in maintaining cell homeostasis (Goodenough & Paul, 2009; Pointis et al, 2007). GJ proteins such as Cx43, Cx26 and Cx30 are down-regulated in HR-HPV cervical samples compared to normal cervix and loss of GJs correlates with cell transformation (Aasen et al, 2005; McNutt & Weinstein, 1969; Mesnil et al, 2005). However, the mechanism behind its relationship to HPV oncoproteins expression remains largely unknown. Low levels of connexin expression on the cell surface permits the tumour cell to grow without receiving inhibitory, localisation or differentiation signals: in HPV-positive cells, this could help the virus evade the immune response. Furthermore, GJs allow antigenic peptides to move between polarised cells in the epithelium. Loss of GJs may confer a survival advantage to cancer cells by restricting immune recognition by cytolytic killer cells of the immune system (Ernesto Oviedo-Orta, 2013).

Down-regulation of Cx26 transcripts was further investigated in HPV-transformed cells and in human keratinocytes (HaCaT cells) at both the structural level (i.e. by analysing levels of the Cx26 protein by Western blotting and immunofluorescence) and at the functional level, by analysing the function of gap junctions in normal and HPV-positive epithelial cells using dye microinjection into living cells. These studies were further expanded by attempting to define the possible role of either (or both) the viral E6 and E7 oncoproteins in regulating Cx26 protein expression and gap junction function using HaCaT cells transduced by recombinant human adenoviruses designed to express the E6 or E7 proteins.

5.2 RESULTS

5.2.1 Analysis of transcript expression

Bioinformatic analysis was performed using the Galaxy platform and the TopHat-Cufflinks pipeline on samples that had been previously prepared (Watherston, 2010). To analyse differentially-expressed transcripts, the Cuffdiff program was employed. Cuffdiff estimates differentially-expressed genes and transcripts based on transcript quantification performed by Cufflinks. Cuffdiff calculates gene and transcript expression levels in different conditions and it tests them for significant differences. A number of output files containing results for changes in expression at the level of transcripts and genes are generated. In addition, it estimates changes in the relative abundance of transcripts which share a common transcription start site (TSS) (to study changes in splicing events) as well as in the relative abundances of the primary transcripts of each gene (relative promoter usage within a gene). It generates several results data that can be divided in 3 groups:

1. *FPKM tracking files*: the FPKM (expression metric) value of each transcript, primary transcript and gene in each sample.
    - gene FPKM tracking
    - CDS (coding sequence) FPKM tracking
    - transcript FPKM tracking
    - TSS (transcription start site) groups FPKM tracking
2. *Count tracking files*: the raw data i.e. non-normalised number of fragments that originated from each transcript, primary transcript, and gene in each sample.
    - CDS overloading differential expression testing
    - CDS FPKM differential expression testing
    - TSS groups differential expression testing

3. *Differential expression tests*: the results of differential expression testing between samples for spliced transcripts, primary transcripts, genes, and coding sequences
    - gene differential expression testing
    - splicing differential expression testing
    - **transcript differential expression testing**
    - promoters differential expression testing

In this study only the "transcript differential expression testing" output for each of the samples analysed has been considered. Results were sorted so that only statistically significant data were considered. Thus, of a total of over 40000 transcripts, those whose changes in transcript abundance were not statistically significant were discarded (unless they were considered for comparison with samples which had the same transcripts significantly deregulated). Heatmaps were generated to show the overall expression of all the cellular transcripts between the samples analysed (Figure 5-1, Figure 5-2).

Since attention was focussed on transcripts encoding cell-surface proteins, the results were narrowed down by filtering for encoded cell-surface proteins only (gene ontology, GO: 005886,). This gave a total of about 100 transcripts per sample analysed. These were sorted by expression level ($\log_2$[fold change]) and the top 20 and bottom 20 were chosen as representatives of up-regulated and down-regulated transcripts, respectively (Appendix 6: Significantly deregulated cell surface transcripts). Only transcripts that were consistently deregulated between all samples were considered for further analysis (Figure 5-3, Figure 5-4 and Figure 5-5).

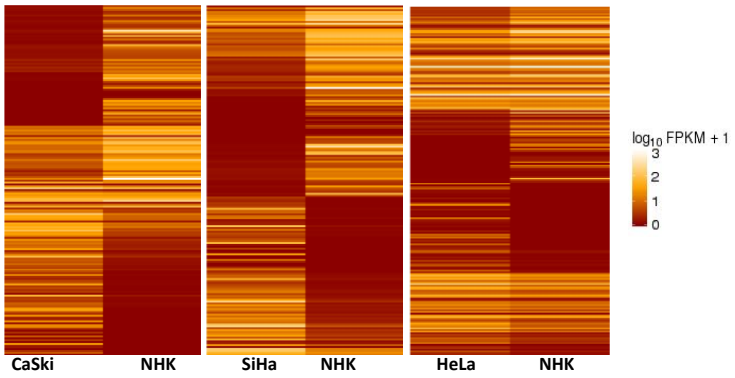


Figure 5-1 Transcript expression levels between the cell lines and NHK.

Values obtained from Cuffdiff output "transcript differential expression testing" were used to compare expression levels between the samples. The transcripts expression levels are represented by the log_{10} FPKM+1 that corresponds to a certain intensity of the colours indicated in the Heat-map. For the Heatmap, all the transcripts generated by Cuffdiff were used, whether the difference was significant or not. NHK=normal human keratinocytes.

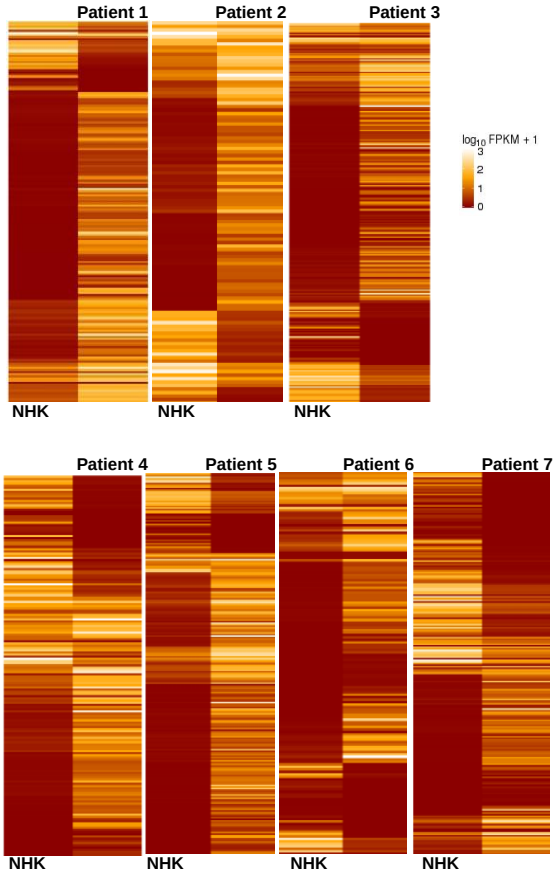


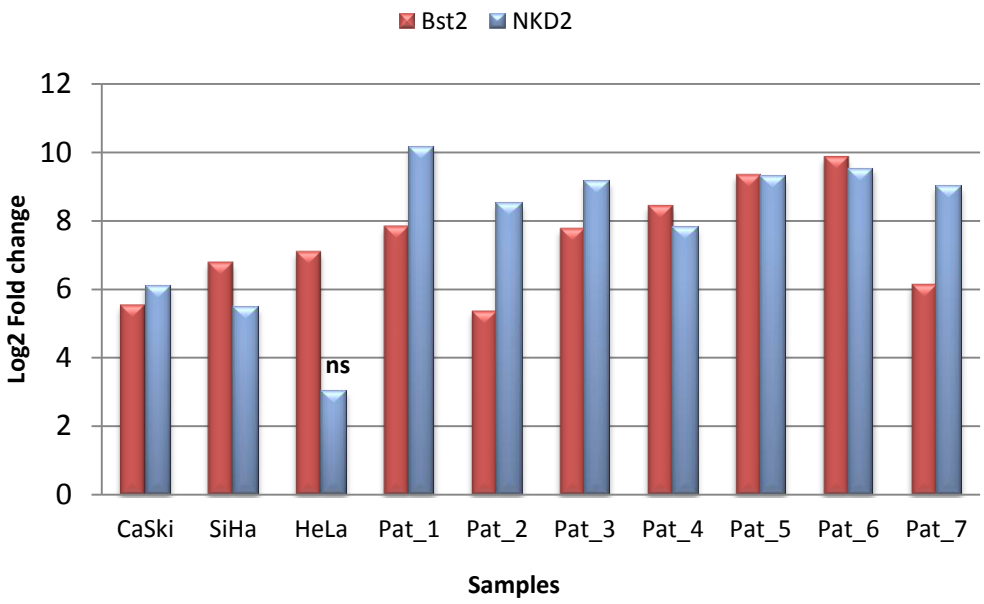
Figure 5-2 Transcript expression levels between the patient cervical cancer tissue samples and NHK.

The heatmaps were generated as described in Figure 5-1.

Since the cell lines were used as standard examples where expression of all transcripts should not be affected by the history of the specific patient, common transcripts were first analysed between CaSki, SiHa and HeLa cells. These were then analysed in the patient samples to determine whether similar expression could be observed.

Among the up-regulated transcripts encoding cell-surface proteins, Bst2, NKD2, GBP, ENO2 and CA9 were found to be commonly deregulated in all samples analysed (Figure 5-3). Literature surveys did not report a connection between these transcripts and HPV infection or cervical cancer, making them an interesting target to follow up for further studies. Bst2, also known as tetherin (CD317) has been recently investigated for its role in the immune response and anti-viral activity against enveloped viruses (Hotter et al, 2013), in particular retroviruses (HIV) (Kuhl et al, 2010; Tokarev et al, 2009). HPVs are not enveloped viruses and they are DNA viruses, nevertheless, Bst2 was a good candidate for further investigation.

Down-regulated transcripts included FGFBP1 (fibroblast growth factor binding factor 1), AMIGO2 (adhesion molecule with Ig-like domain 2), FEZ1 (fasciculation and elongation protein Z-1) and 4 members of the connexin family: GJB (gap junction protein) 2, GJB4, GJB5 and GJA1 (Figure 5-4, Figure 5-5). Of the latter, only one connexin, GJB2 was commonly down-regulated in all samples analysed, while the other connexins could be not analysed in Cuffdiff due to low values assigned during the analysis. This occurred for many transcripts whose expression values were too low to be considered for statistical analysis by Cuffdiff.



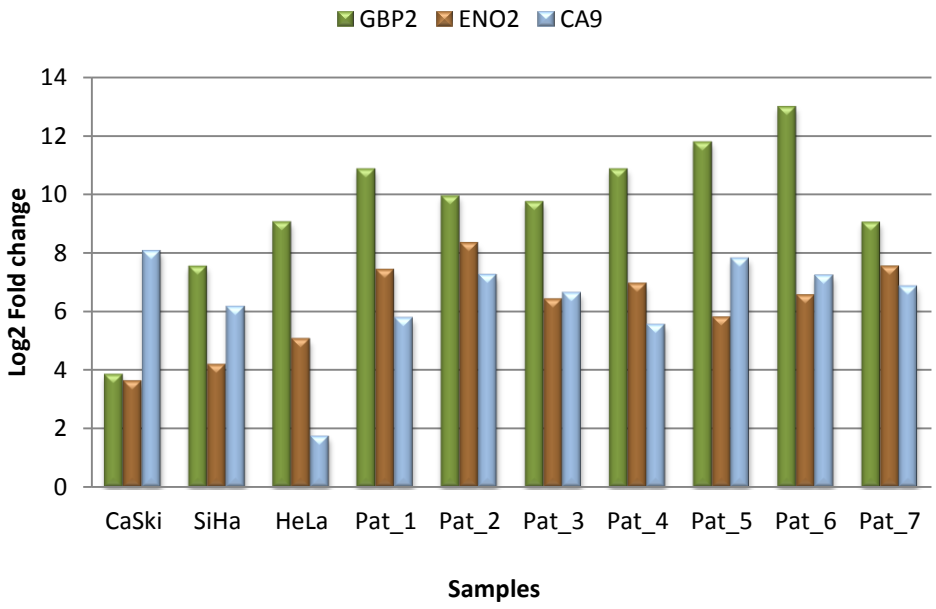


Figure 5-3 Up-regulated cell surface transcripts within the samples analysed.

Unless otherwise stated only statistically significant values were considered for the comparison. (ns= non-significant, Pat_= patient).

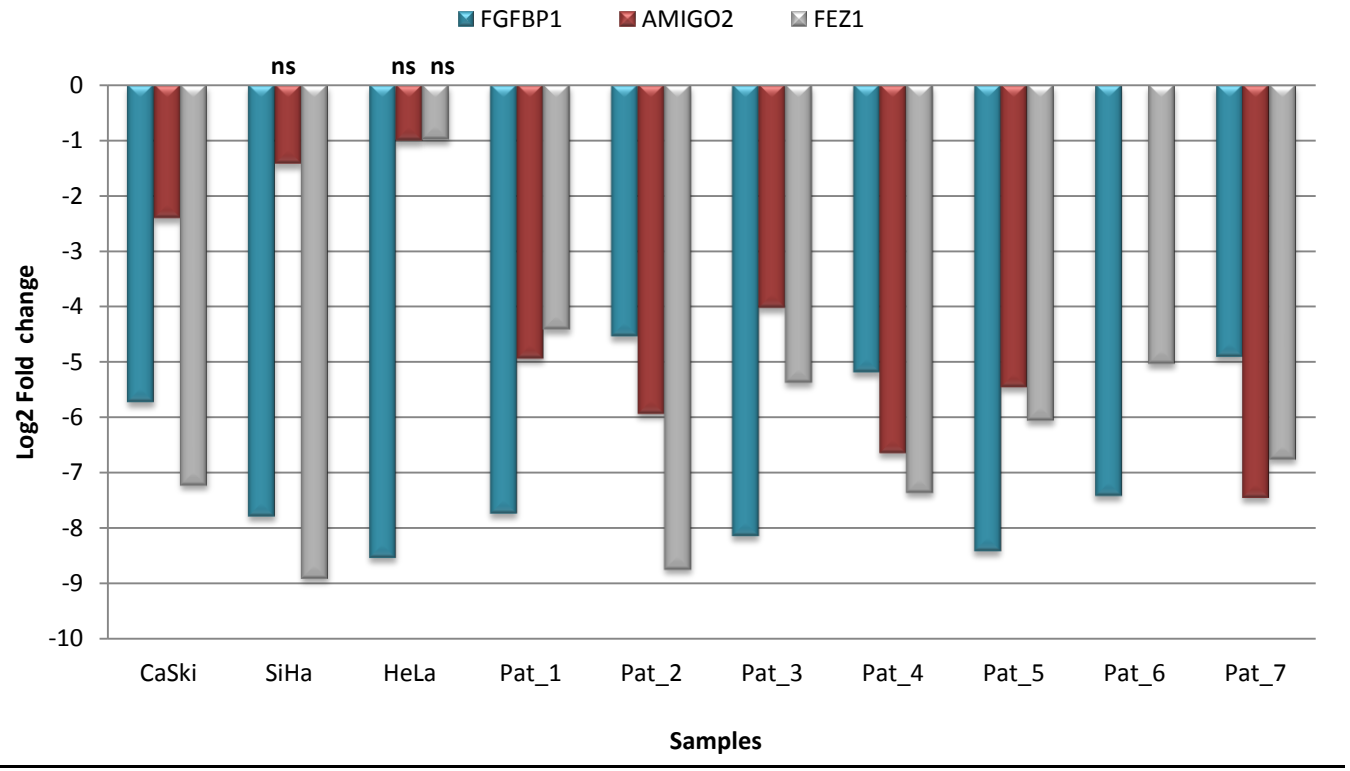


Figure 5-4 Down-regulated transcripts between the samples analysed.

Unless otherwise stated, only statistically significant values were considered for the comparison. (ns= non-significant, Pat_= patient).

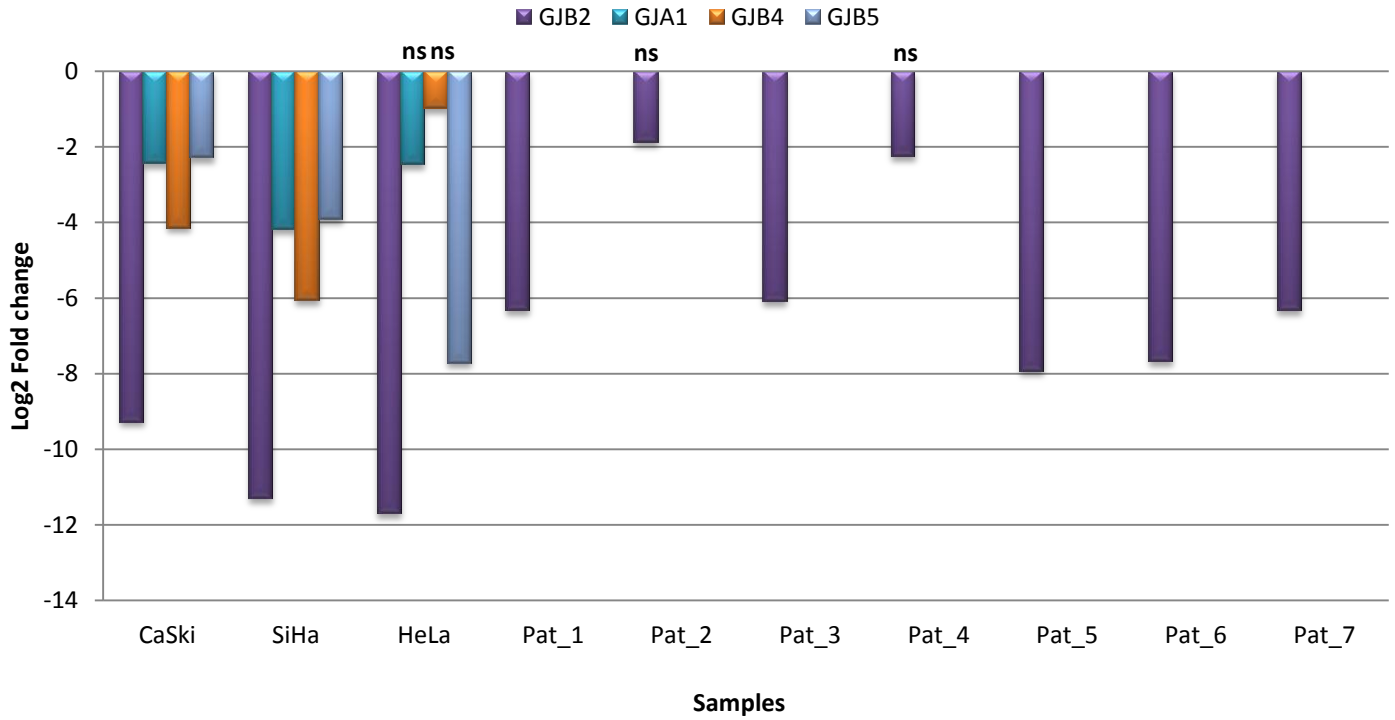


Figure 5-5 Down-regulated connexin transcripts between the samples analysed.

GJA1, GJB4 and GJB5 did not yield values which Cuffdiff could use for the statistical analysis in the patient samples (infinite number). For comparison purposes GFOLD values have been used which, despite being without statistical analysis, are consistent with the fold changes observed in Cuffdiff. Unless otherwise stated, only statistically significant values were considered for the comparison. (ns= non-significant, Pat_= patient).

5.2.2 Data comparison with a lyoplate assay

The lyoplate assay was performed in collaboration with Dr Erica Wilson (Leeds Institute of Cancer and Pathology, Leeds), according to the manufacturer's instructions (BD Bioscience). The experiment was performed to analyse the expression of proteins belonging to the CD family (cluster of differentiation) and then compare the results with the RNA-Seq data (Table 5-2). While the assay only covers a small fraction of the repertoire of cell-surface molecules, it can give an estimate of how mRNA levels calculated by the RNA-Seq reflect the protein levels, as well as giving an indication of which molecules might have a role in the immune system during HPV infection and transformation.

Following flow cytometry, CD antigens/proteins were filtered based on the Mean Fluorescence Intensity (MFI) and the highest and lowest values were considered for further analysis and comparison with the sequencing data (Figure 5-6).

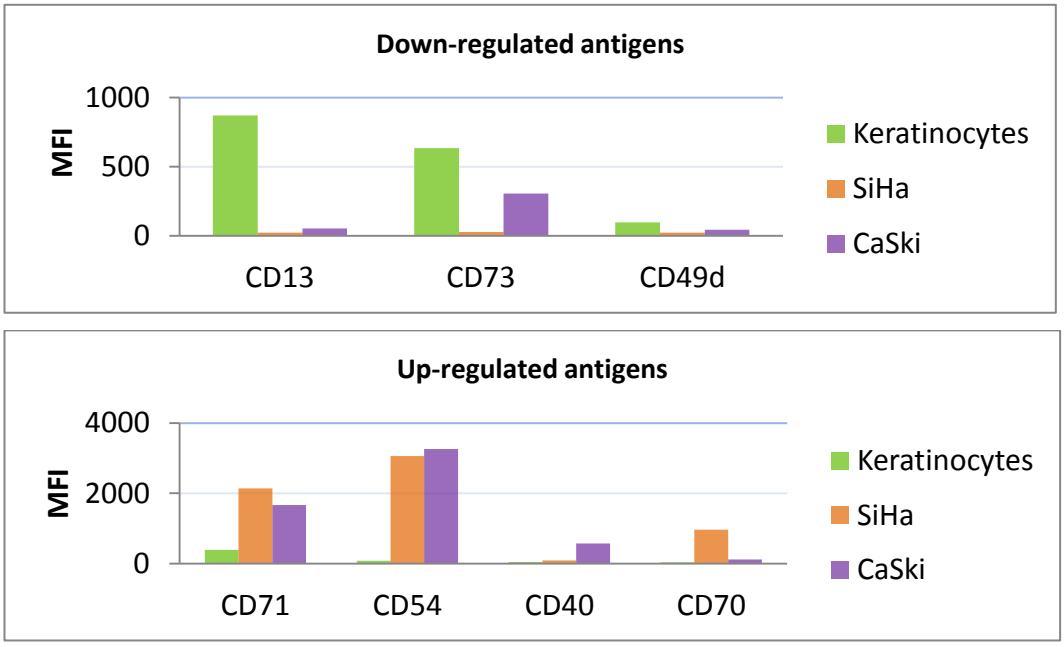


Figure 5-6 Identification of differentially-regulated cell-surface molecules using the lyoplate assay.

Top, cell surface antigens with the lowest MFI in NHK compare to HPV-transformed cells, bottom, cell-surface antigens with the highest MFI. MFI = mean fluorescence intensity.

Deregulated proteins were compared with the RNA-Seq respective mRNA levels. The results showed that some transcripts were similar deregulated at the protein and mRNA level: CD13 (Anpep, an aminopeptidase N), CD73 (NT5E, a 5'-nucleotidase) and CD49d (α subunit of α4β1) were significantly down-regulated in CaSki and SiHa cells compared to normal primary human keratinocytes, while in the patients' samples they were down-regulated but not significantly, except for patients 4 and 7 (data not shown). The up-regulated CD antigens/proteins were also consistent with the mRNA levels and this was also observed in the patients' data, despite not being statistically significant. In addition, some patient data showed down-regulation of these transcripts, such as CD70, which is down-regulated in patient 3, 4, and 8, although not significantly (Table 5-2).

Table 5-2 Cuffdiff values for the transcripts compared with the lyoplate assay.

Bold = up-regulated transcripts, Italic = down-regulated transcripts

		Cuffdiff values						
locus	CaSki	Keratinocytes	test_stat	p_value	q_value	significant	gene	log2(fold_change)
chr15:90328125-90358072	**0.257686**	**53.2823**	**-8.09445**	**6.66E-16**	**5.19E-13**	**yes**	**CD13**	**7.6919**
chr6:86159301-86205509	**10.3605**	**67.8694**	**-3.76306**	**0.000167845**	**0.0053035**	**yes**	**CD73**	**2.71167**
chr2:182321618-182521834	**0**	**1.79769e+308**	**1.79769e+308**	**0.00324421**	**0.0455868**	**yes**	**CD49d**	**4.34436**
chr3:195776154-195809032	*10.6992*	*-3.41721*	*4.4854*	*7.28E-06*	*0.00039*	*yes*	*CD71*	*-3.41721*
chr19:10381516-10397291	*12.2275*	*0.758473*	*4.58318*	*4.58E-06*	*0.00027*	*yes*	*CD54*	*-4.01089*
chr20:44746905-44758384	*23.9708*	*2.35439*	*3.85716*	*0.000114711*	*0.003945*	*yes*	*CD40*	*-3.34785*
chr19:6585849-6591163	*2.03915*	*0.11152*	*3.29818*	*0.00097312*	*0.020684*	*yes*	*CD70*	*-4.1926*
	SiHa	Keratinocytes						
chr15:90328125-90358072	**0.0755584**	**54.3254**	**-7.99128**	**1.33E-15**	**3.82E-12**	**yes**	**CD13**	**9.48982**
chr6:86159301-86205509	**5.42239**	**69.138**	**-4.18021**	**2.91E-05**	**0.0015122**	**yes**	**CD73**	**3.67248**
chr2:182321618-182521834	**0.0123287**	**4.42555**	**-5.29041**	**1.22E-07**	**1.68E-05**	**yes**	**CD49d**	**8.4877**
chr3:195776154-195809032	*107.074*	*10.8989*	*3.34888*	*0.0008114*	*0.020717*	*yes*	*CD71*	*-3.29635*
chr19:10381516-10397291	*23.5672*	*0.773006*	*4.87786*	*1.07E-06*	*0.000108*	*yes*	*CD54*	*-4.93016*
chr20:44746905-44758384	*25.1908*	*2.40188*	*3.36635*	*0.000761709*	*0.0196*	*yes*	*CD40*	*-3.39066*
chr19:6585849-6591163	*52.2431*	*0.113678*	*7.63902*	*2.20E-14*	*4.02E-11*	*yes*	*CD70*	*-8.84415*

5.2.3 Expression of HPV16 E6 and E7 by recombinant adenoviruses

The HPV16 E6 and E7 open reading frames (ORFs) were generated by PCR using the pEF-399 HPV16 plasmid as template (NCBI reference AF125673.1). The PCR samples were ligated into PCR Blunt and this was used as a template to generate the PCR products used for the recombineering reactions, where the primers included the

homology regions for the E1 region of the adenovirus (for insertion of E6 and E7) or the E3 region (for insertion of CFP). To test for positive colonies, a colony PCR was performed (Figure 5-7A and Figure 5-8A) and where found, colonies were subjected to BamHI digest. If the insert was effectively within the E1 region, then two bands of 1.7 kb and 2.5 kb were lost. This can be compared with the adenovirus expressing only CFP, where the two bands are still visible (Figure 5-7B and Figure 5-8B). Provided the colony was positive, it was sent for sequencing and 911 cells transfected to expand the virus. Following virus purification, HaCaT were transduced with the recombinant adenovirus. To eventually confirm that the construct was working, RNA was extracted and cDNA generated, to be used as a template for PCR reactions. Purified virus was also quantitated for number of infective particles as described in Chapter 2.9.6.

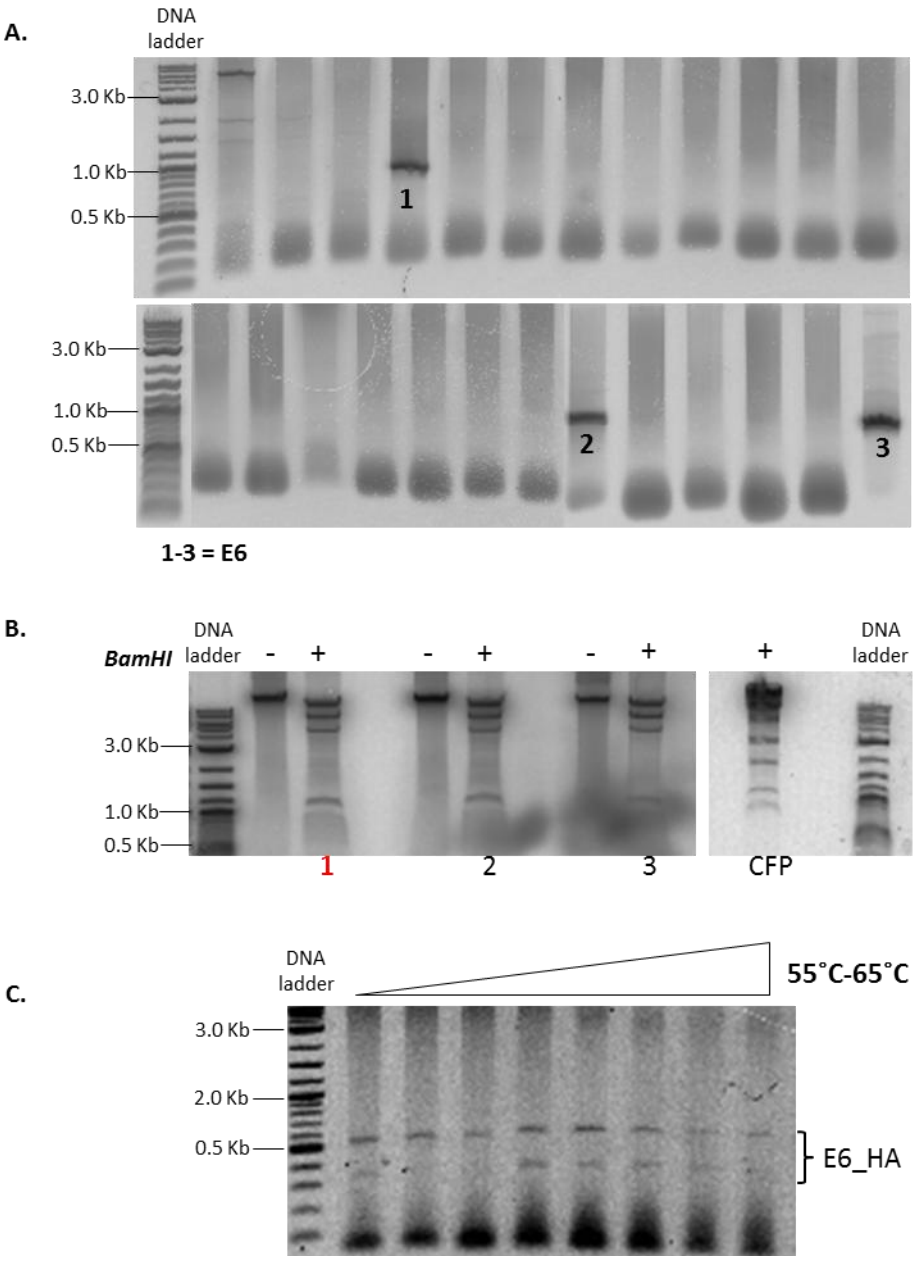
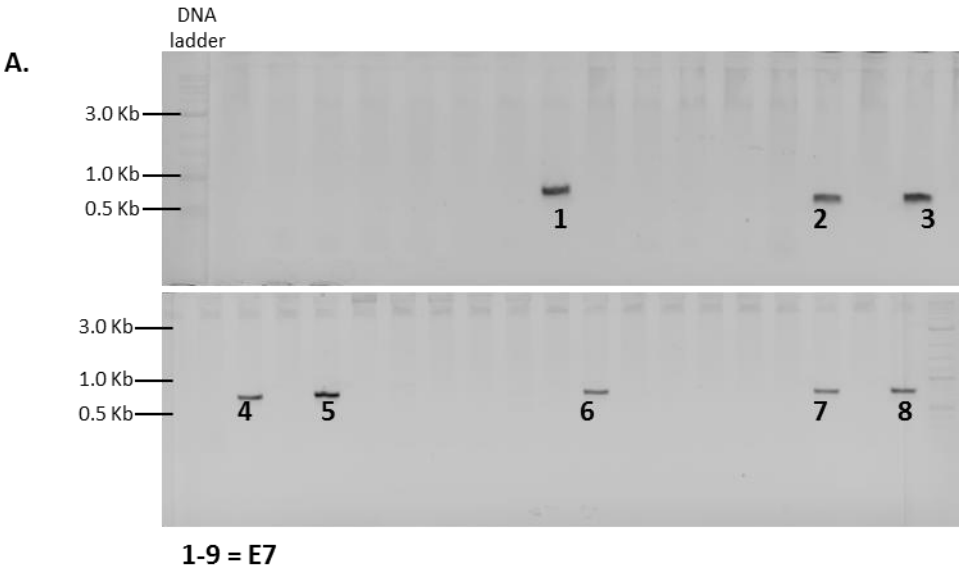
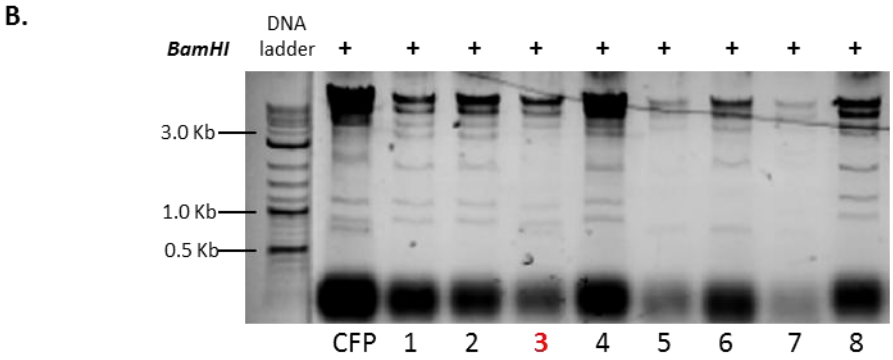


Figure 5-7 Generation of a recombinant adenovirus expressing HPV16 E6.

A. A colony PCR was performed to identify which colony had incorporated the E6 ORF following recombineering. B. *Bam*HI restriction digest of the three putative positive colonies was performed to confirm the presence of the insert within the E1 region. Colony N.1 (red) contained the insert and thus expanded for transfection in 911 cells C. cDNA from transduced HaCaT cells was used as an RT-PCR template to generate an RT-PCR product (PCR gradient 54 to 65°C) to confirm that mRNA is made within the cells.





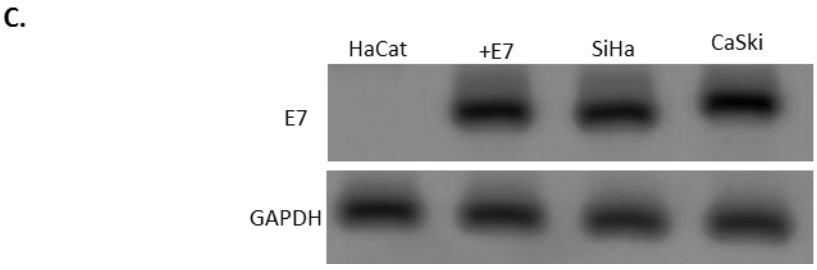


Figure 5-8 Generation of recombinant adenovirus expressing HPV16 E7

A. A. A colony PCR was performed to identify which colony had incorporated the E7 ORF following recombineering. B. *Bam*HI restriction digest of the eight putative positive colonies was performed to confirm the presence of the insert within the E1 region. Colony N.3 (red) contained the insert and thus expanded for transfection in 911 cells C. cDNA from transduced HaCaT cells was used as a RT-PCR template to generate an RT-PCR product. CaSki and SiHa cells were used as a positive control, GAPDH as a loading control.

5.2.4 Studies on the expression of Cx26

Analysis of the RNA-Seq data showed that among the down-regulated transcripts, GJB2 (Cx26) showed a particular expression pattern compared to the other connexins, which were down regulated only in the cervical cancer cell lines. To investigate the expression of Cx26 further, RT-PCR was performed to validate the expression levels detected by RNA-Seq as well as to test the hypothesis that E6 or E7 might down-regulate Cx26 mRNA levels. The results showed that Cx26 mRNA levels were decreased in the presence of E6 (6 -fold decrease) and E7 (2.7-fold decrease) while it was recorded a larger decrease in the presence of E6/E7 (65.4-fold decrease). The figure also shows that the control CFP construct did affect the levels of Cx26 transcripts with 1.4-fold decrease (although in other repeats of the RT-PCR levels were equal to non-transduced HaCat cells; data not shown). Finally, the data shown in Figure 5-9 confirmed that Cx26 mRNA levels were greatly reduced in CaSki, SiHa and HeLa cells compared to HaCaT cells. This also suggests that different factors other than E6 and E7 are involved in regulating Cx26 mRNA levels.

Immunofluorescent antibody labelling was performed to test whether mRNA levels observed in the cell lines reflected protein levels. The data showed that HaCaT cells had a different pattern of Cx26 protein expression to that in CaSki and SiHa cells, suggesting that the low mRNA levels observed reflected lower levels of protein expression on the surface of the cells. Previous literature reports that HeLa cells do not express connexins were confirmed in these experiments (Figure 5-10).

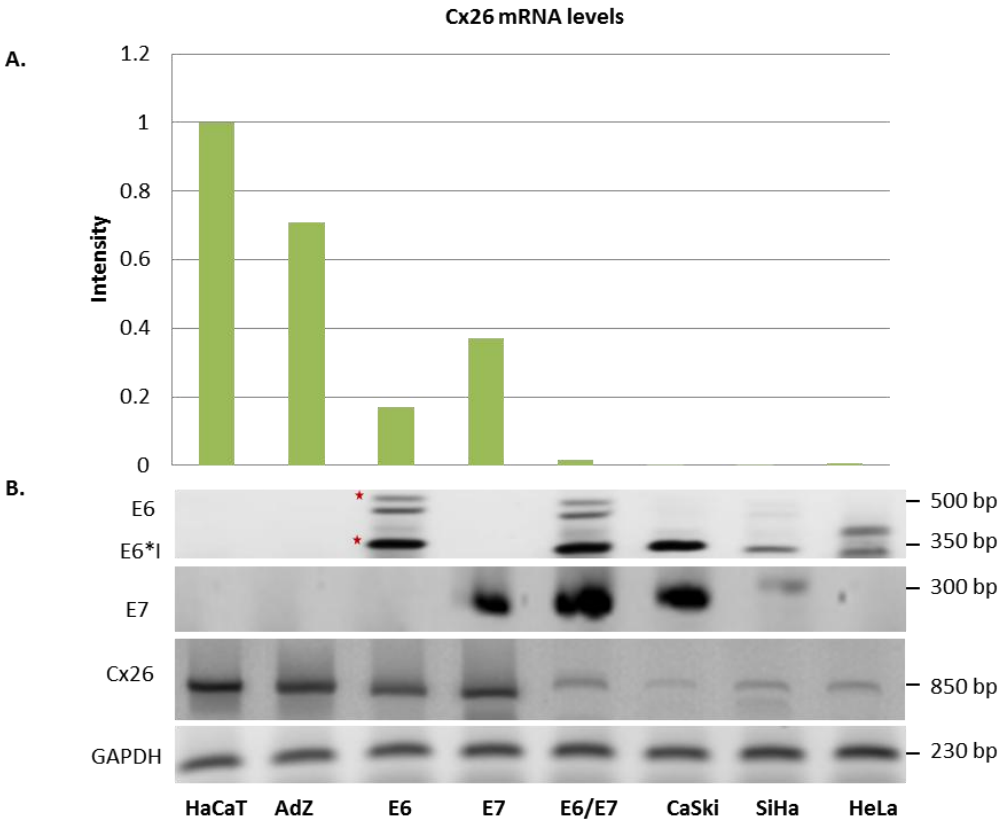


Figure 5-9 Expression of Cx26, HPV16 E6 and E7 mRNAs in cervical cancer (CaSki, SiHa and HeLa) and normal human keratinocytes (HaCaT).

A. Top panel shows a graph with the quantified mRNA values calculated with the Aida Soft Imager which calculates the intensity of the bands. Values normalised against GAPDH and the negative control (HaCaT). Bottom panel (B) shows mRNA for E6 and its variant, E7, Cx26 and GAPDH in HaCaT cells transduced with the recombinant adenoviruses and in the HPV positive cell lines CaSki, SiHa and HeLa. GAPDH was used as a loading control. Red stars indicate the two E6 products.

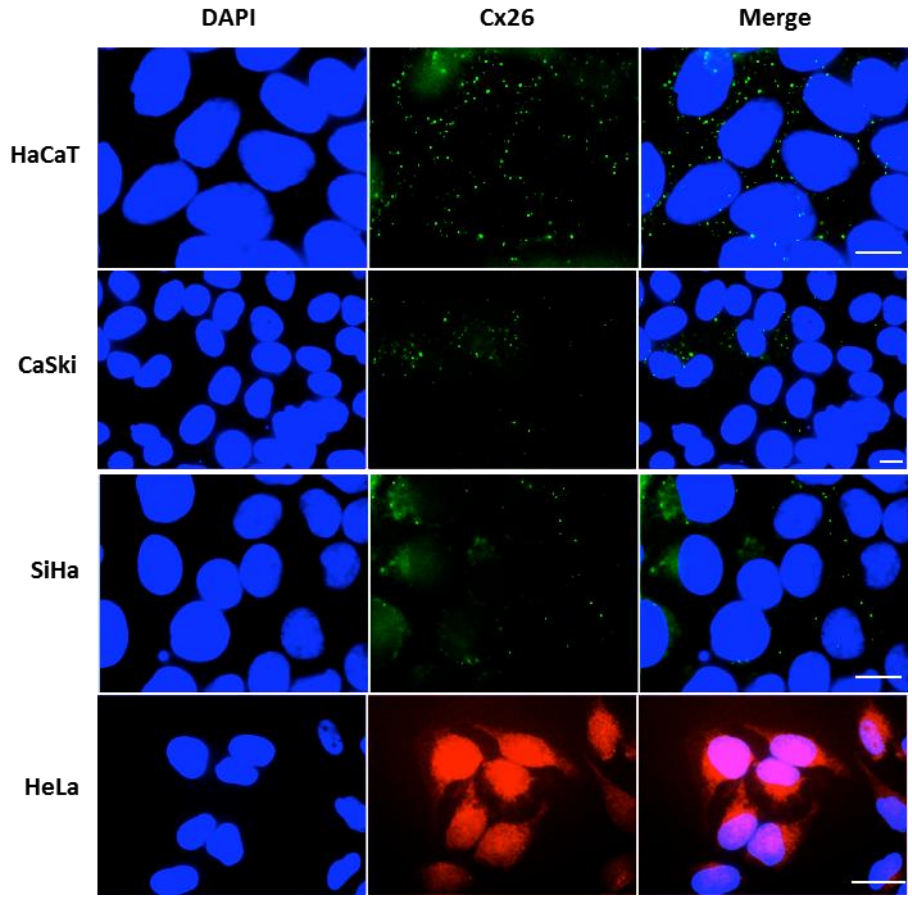


Figure 5-10 Immunofluorescent antibody staining for Cx26 in HaCaT, CaSki, SiHa and HeLa cells.

The picture shows pattern of expression for Cx26 in HaCat, CaSki, SiHa and HeLa cells. Left panel: labelling for nuclei only (DAPI), middle panel: labelling for Cx26; right panel: merged. Scale bar: 10 µm. Pattern of expression confirmed as observed in published literature (Bakirtzis et al, 2003; Gemel et al, 2006; Gilleron et al, 2008).

5.2.4.1 The effect of HPV16 E6 and E7 expression on Cx26 localization

Two adenoviruses that expressed either HPV16 E6 or E7 where prepared as described in Chapter 5.2.3 to test the hypothesis that deregulated mRNA levels observed in the RNA-Seq analysis might be a consequence of the expression of either or both viral oncogenes, E6 and E7. In particular, attention was directed at the pattern of Cx26expression. HaCaT cells were transduced with each construct (or the control Ad vector) and immunofluorescence microscopy performed as previously described (Chapter 2.7.7). CFP-positive cells were selected to identify virally-transduced cells and in those cells the expression of E6 or E7 was confirmed fol-

lowed by analysis of Cx26 expression. While in non-transduced HaCaT cells a distinctive pattern of Cx26 expression was observed, in cells expressing either E6 or E7, Cx26 seemed to be more diffuse and the regular pattern previously observed was lost, suggesting that Cx26 might be accumulating in the cytoplasm and failing to be transported to the cell surface (Figure 5-11).

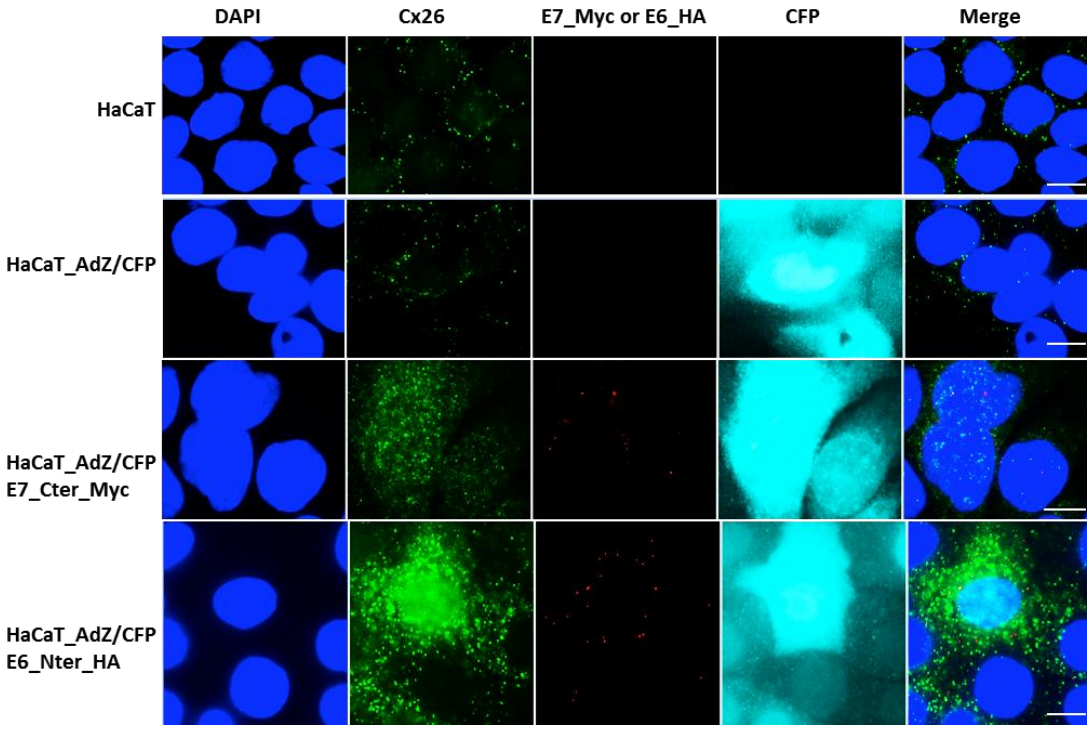


Figure 5-11 Immunofluorescent antibody staining for Cx26 in transduced HaCaT cells.

The picture shows pattern of expression for Cx26 in normal HaCat cells and in transduced HaCaT cells. From the left panel: labelling for nuclei only (DAPI), labelling for Cx26; labelling for E7 Myc tagged or E6 HA tagged; labelling for CFP and far right panel: merged. The first row indicates normal HaCaT cells, the second HaCaT transduced with the control adenovirus, the third HaCaT in presence of E7-Myc-tagged and the last row HaCaT in presence of E6-HA-tagged. Scale bar: 10 µm.

It is worth mentioning that not all CFP-positive cells appeared to co-express either E6 or E7, and it was also observed that some cells that expressed E6 or E7 were at times, not CFP-positive. In addition, CFP-positive cells (in the presence or absence of E6 or E7) seemed to have a bigger nucleus than non-transduced cells, suggesting that the virus might be involved somehow in altering nuclear morphology (data not shown).

5.2.4.2 Cx26 protein levels in transduced HaCaT cells

To provide a more quantitative analysis of the above, HaCaT cells were transduced with the recombinant adenoviruses, whole cell extract was prepared and Western blotting was performed to compare protein levels for Cx26 as well as to confirm the expression of the viral oncoproteins. The results showed that Cx26 protein levels were not affected by expression of either the E6 or the E7 oncogenes, however, when both viral oncoproteins were present, Cx26 protein levels fell dramatically (Figure 5-12).

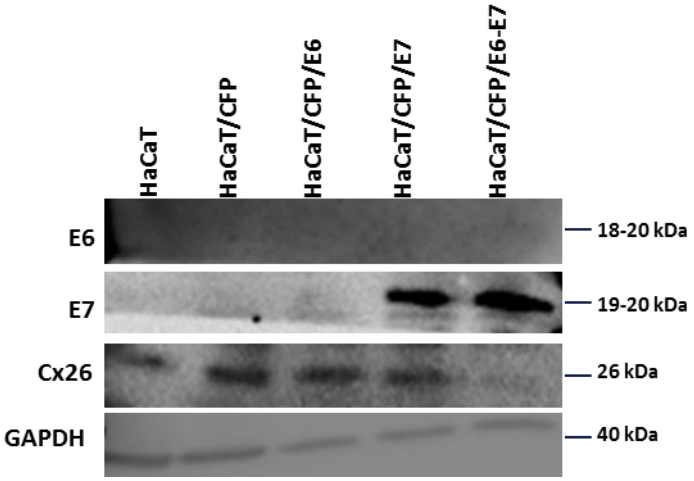


Figure 5-12 Cx26 protein expression in presence of E6 and E7.

Western Blot for Cx26 in HaCaT cells transduced with the recombinant adenovirus expressing E6 and E7.

Moreover, since immunofluorescent antibody staining results showed a re-localization of Cx26 to the cytoplasm from the membrane, cell fractionation was performed to further investigate these findings. Initial results showed that Cx26 re-located to the cytosol in the presence of E6 and E7, while in non-transduced HaCaT or the positive control AdZ, Cx26 does not seem to be expressed in the cytosol (Figure 5-13). As shown in Figure 5-13, analysis of the membrane fractions showed expression of Cx26 protein in both the AdZ negative control and transduced HaCaT cells suggesting that relocation to the cytosol is partial. Knock-down of Cx26 was performed to show that the band was effectively Cx26 since the antibody used rec-

ognized a single band running at 50 kDa, whereas the predicted molecular mass of Cx26 is 25-26 kDa (Figure 5-14).

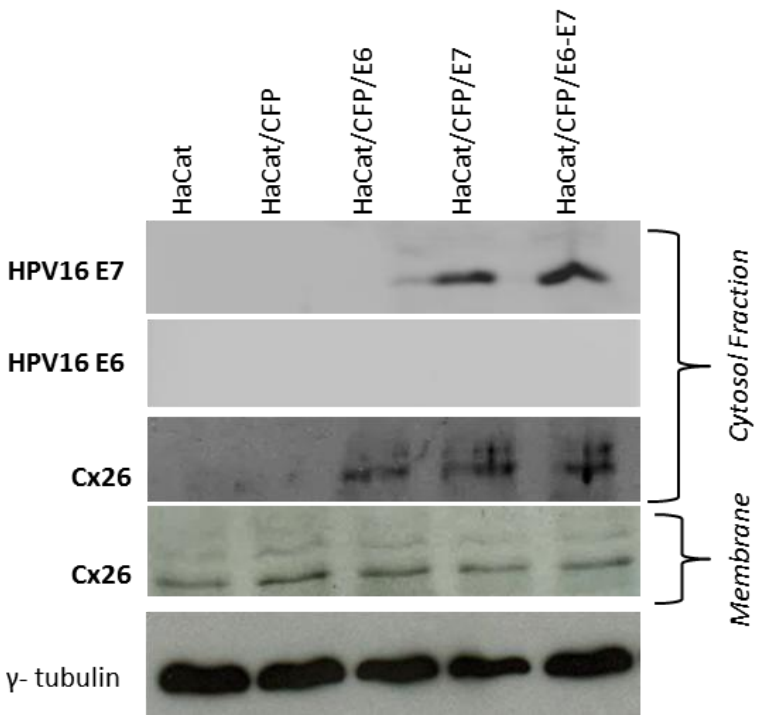


Figure 5-13 Cell fractionation assay on transduced HaCaT cells.

HaCaT cells were transduced with the recombinant adenovirus and assessed for Cx26 expression in the cytosol and membrane fractions.

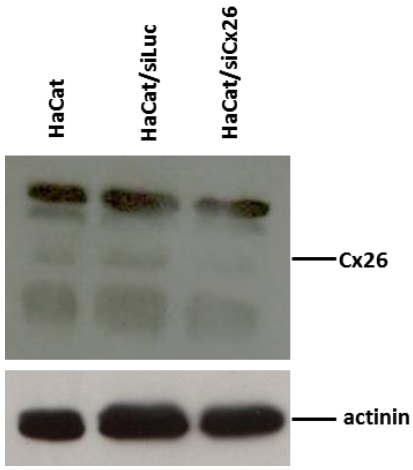


Figure 5-14 Reduction of Cx26 by RNA interference to detect Cx26 protein.

5.2.5 Lucifer yellow dye transfer assay

Lucifer yellow dye transfer assay was performed to monitor functional gap junction expression in HaCaT cells transduced with the recombinant adenovirus expressing E6. Following microinjection of the dye, images were captured between 0 and 5 minutes to follow the dispersion of the dye to the neighbouring cells. This was indicative of the structural identity of the junctions between the cells following transduction by the recombinant adenovirus. An average of 4 injections where performed per cell sample in 3 replicates and spreading of the dye was assessed as the number of cells which became green following dispersion of the dye (Figure 5-15, Figure 5-16).

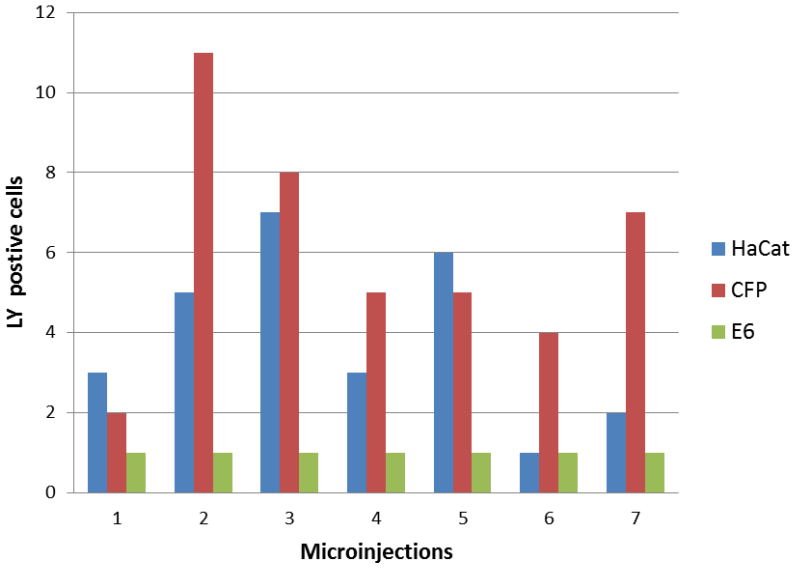


Figure 5-15 Lucifer yellow dye transfer assay in transduced HaCaT cells.

A series of microinjection were performed in HaCaT cells transducer with the recombinant adenovirus and the number of cells in which the Lucifer yellow dye has diffused was counted. The number of cells is reduced in HPV16 E6 transduced cells suggesting that E6 is effectively altering cell-cell communication in HaCaT cells.

The series of injections showed that gap junctions are interrupted in HaCaT cells that were transduced with the recombinant adenovirus expressing E6, while communication is well established in non-transduced HaCaT or the control adenovirus,

suggesting that E6 is somehow involved in altering cell-cell communication in HaCaT cells (Figure 5-15 and Figure 5-16). This is consistent with the immunofluorescence results where HPV16-positive, E6-expressing cells appeared to be deficient in gap junction expression. The construct expressing E7 was not tested for loss of GJs since previous reports have ascribed a role key for the high-risk E6 oncoproteins (Grm & Banks, 2004).

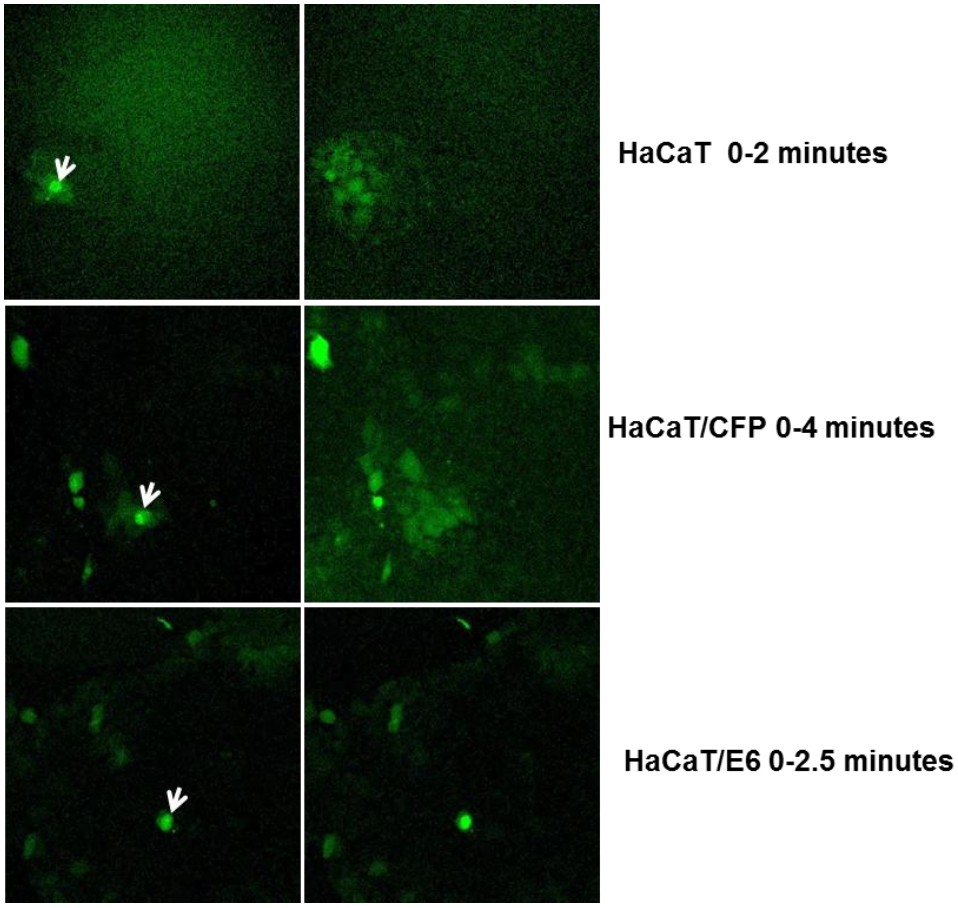


Figure 5-16 Lucifer yellow dye transfer assay in transduced HaCaT cells.

As shown in the picture, while in HaCaT and CFP-transduced cells the dye can diffuse, in E6 positive cells the dye does not diffuse due to the lack of gap junctions. Arrows indicate the injected cell. Cells appearing green other than those indicated by the arrow are CFP-positive cells.

5.3 DISCUSSION

Bioinformatic analysis of RNA-Seq data was performed on three HPV-positive cell lines (CaSki, HeLa, SiHa), seven cervical tumour biopsies and normal human keratinocytes (NHKs). Nine cellular transcripts encoding cell-surface proteins were found to be differentially expressed in all samples analysed. Among these, attention was focused on Cx26, a protein belonging to the gap junction protein family. These results alone cannot provide a direct link between E6 and/or E7 expression and Cx26 deregulation, although both CaSki and SiHa cells showed a reduction in connexin mRNA expression compared to normal keratinocytes, as shown by the RNA-Seq data as well as by RT-PCR analysis. The immunofluorescent antibody staining experiments supported these findings since they showed that Cx26 was expressed, but localized in the cytoplasm, suggesting that additional mechanisms are adopted by the two viral oncoproteins to impair the proteins' localization in the membrane.

Understanding how and why Cx26 is down-regulated in cervical cancer and whether this is a direct or indirect consequence of HPV oncoprotein expression might elucidate a significant aspect of the mechanism of HPV oncogenesis. Previous reports have suggested a correlation between loss of GJ proteins and HPV-related tumour progression; however, specific mechanisms are not well understood. In particular, it is not known whether several or all members of the Cx family are deregulated in cervical cancer or whether deregulation is tissue-specific and restricted to certain connexin family members. Elucidating correlations between transcript and protein levels of Cx26 and the underlining mechanisms may form the basis for future studies on GJ proteins and their role during HPV infection.

It can be proposed that there may be a connection between loss of connexin expression and cervical cancer development as suggested for other types of cancer and as shown by the data presented in this chapter, however at this stage specific mechanisms still need to be elucidated. It can be suggested that, given that kerati-

nocytes are antigen-presenting cells (APC) important in directing the T cell response (Banerjee et al, 2004), GJs can work as immunological junctions to transfer antigenic peptides between infected cells and professional APCs, such as dendritic cells (DCs); DC, in turn present the antigen on their surface and, following activation, migrate to the lymph nodes. HPVs might take advantage of the role played by GJs by down-regulating their expression. HPV16 E5 is known to inhibit gap junction contacts, directly or indirectly by influencing the phosphorylation status of connexin 43 (Oelze et al, 1995), thus increasing the chances of promoting abnormal growth of those cells that have not been specifically recognized by the immune system. In addition it is known that HPV can down-regulate MHC class I molecules (Bottley et al, 2008). These processes might explain how, over a period of 10-20 years, the virus escapes immune surveillance and promotes tumour generation.

Other mechanisms have to be considered to account for loss of GJs. Indeed, elegant studies in Dr Banks' laboratory have shown that a PDZ-domain protein, MAGI-1 (membrane associated guanylate kinase 1), a member of the MAGUK family, is one of the major targets of HPV16 E6 and its degradation leads to disruption of tight junction (TJ), another major component of cell junctions, along with GJs, desmosomes and adherens junctions. It has been also shown that Cx43 establishes a functional interaction with the hDlg protein triggering Cx43 relocation to the cytoplasm, reducing the membrane-bound form and thus reducing GJ intercellular communication. This is consistent with the results presented in this chapter where relocation of Cx26 to the cytoplasm was also observed in presence of HPV16 E6. (Macdonald et al, 2012)

Finally, it has been reported for other tumours that low level of GJs may be a consequence of reduced levels of mRNA due to promoter hypermethylation (Chen et al, 2005; Tan et al, 2002). HPV16 E7 is known to interact with DNA methyltransferases (DNMTs) while HPV16 E6 is known to up-regulate DNMT1 levels following p53 repression , suggesting other possible mechanisms by which HPV controls the levels of

GJ protein expression, namely through regulation of promoter activity and transcript levels (Au Yeung et al, 2010; Burgers et al, 2007). Sequence analysis of the Cx26 gene has indeed revealed a putative CpG island adjacent the promoter (Figure 5-17) and it overlaps the promoter region (within the exon 2 of Cx26) as confirmed by the sequence alignment of the CpG island DNA sequence (UCSC) with the Cx26 gene (NCBI GenBank: JQ595559.1) (Kiang et al, 1997). This suggests that promoter methylation may account for low levels of mRNA observed in the RNA-Seq data. Future experiments, for example, using inhibitors of cytosine methylation such as 5-Aza-2'-Deoxycytidine, might address this hypothesis to test if this is the case.

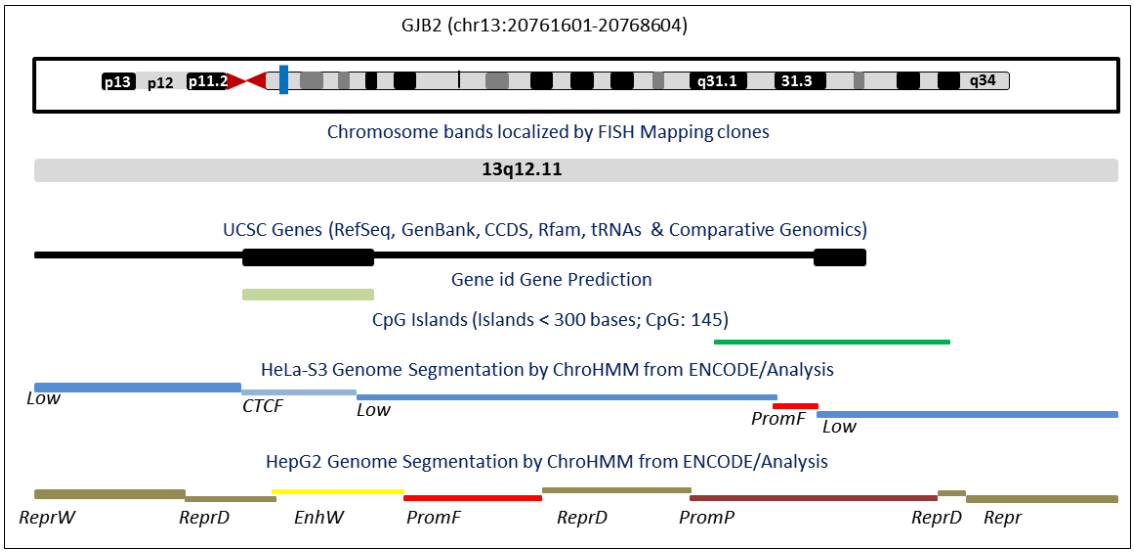


Figure 5-17 Genomic features on the GJB2 gene segment.

The picture was generated using the UCSC genome browser. It shows the presence of the regulatory elements on the GJB2 gene such as promoters and enhancers as well as the CpG island overlapping the promoter (in HeLa cells). CpG islands were predicted by searching the sequence one base at a time, scoring each dinucleotide (+17 for CG and -1 for others) and identifying maximally scoring segments. Each segment was then evaluated for the following criteria: GC content of 50% or greater, length greater than 200 bp, ration greater than 0.6 of observed number of CG dinucleotides to the expected number on the basis of the number of GCs and Cs in the segment. The CpG count (145) is the number of CG dinucleotides in the island. This suggests a possible methylation of the GJB2 promoter which may trigger low mRNA levels as observed in the NGS/RNA-Seq data. *PromF*: promoter flanking, *PromP*: inactive promoter, *Low*: low activity proximal to active sites, *EnhW*: candidate weak enhancer, *CTCF*: distant CTCF/candidate insulator, *Repr/D/W*: Polycomb repressed (http://genome.ucsc.edu; the picture has been adapted from the original).

Chapter 6. General discussion and future considerations

In this study the mechanism by which a remote enhancer communicates with the cognate promoter over long distances was addressed. Using the murine immunoglobulin lambda light chain locus (Igλ), it was shown that E47 forms a protein link between the promoter and the enhancer of sterile transcription of the Jλ1 gene segment by interacting with the enhancer-bound protein complex PU.1/IRF4. The results obtained confirm that the protein link is necessary but not sufficient to mediate productive enhancer-promoter communication in order to trigger sterile transcription of the murine Jλ1 gene segment.

This is in agreement with recent studies that have investigated the presence of additional enhancers at the immunoglobulin heavy and light chains, known as "super-enhancers", in addition to those addressed in this study, suggesting that these other regulatory elements may be targeted for binding by B-cell specific transcription factors such as E47 and IRF4 or that other factors might be involved in this interaction, as shown by the additional factors found in the initial DNase footprinting reactions in Chapter 3 (Predeus et al, 2014). Of these, of particular interest is the transcription factor AP-1 which has been previously shown to be essential at the kappa locus (Schanke et al, 1994) and which should be further studied to determine whether it plays a role during transcription of the antigen receptor light chain loci.

Several human diseases are a consequence of altered action of enhancers' mechanism. To establish how enhancer/promoter interaction occurs during B and T cells development (in the context of the antigen receptor loci transcription) and to investigate the factors involved will be vital considering that mistakes in V(D)J recombination account for 40% of all leukaemia cases (Table 6-1). Illegitimate recombination between Ig or TCRs and non-antigen receptor loci triggers chromosomal translocations that can lead to different forms of leukaemia in both B and T cells (Marculescu et al, 2006). This occurs because the RAG proteins recognize a site that

is similar to the RSSs, namely “cryptic RSSs” (cRSSs), which are found within every 1-2 kb on average (Lewis et al, 1997). Upon cleavage, chromosome translocation is promoted; these translocations are characterised by the ectopic activation of a silent proto-oncogene when it is re-located in the vicinity of an active regulatory element. However, several factors have to be taken in account to explain why such mistakes occur. For example, chromatin accessibility as well as recruitment of co-activators by enhancer-bound transcription factors might play a key role in this process. Taking in consideration the mistakes occurring during V(D)J recombination, the non-antigen receptor loci involved in the recombination reaction must be accessible for RAG cutting, suggesting that, for reasons to be clarified, at that specific moment of B or T cell development, those loci can be accessed by the recombinase machinery.

The advent of 3C and its variants circularized chromosome conformation capture (4C) and genome-wide chromosome conformation captures (Hi-C) have been widely used to determine direct physical contact between separated DNA sequences. These techniques, complemented by studies on histone modification and HAT-binding patterns have helped to predict enhancers within several cell types, suggesting that a similar approach could potentially be applied in B and T cells to study which enhancers are simultaneously active during V(D)J recombination and how distant genes are selected for activation following RAG cutting. In addition, further studies have also shown the presence of stalled RNA polymerases at certain genes suggesting that the main role of enhancers is to promote the transition between transcription initiation and elongation phases (Nechaev & Adelman, 2011). This suggests that a combination of chromatin accessibility factors, enhancer-promoter interaction and transcription factors required to promote this interaction, are necessary at the non-antigen receptor loci to be equally recognized by the recombinase machinery and trigger chromosome translocations. To complement current knowledge of these interactions a combination of 3C with the new high-throughput sequencing methodologies could help discover novel long-range interactions of a spe-

cific locus, such as those involving non-antigen receptor loci. This could be potentially studied in a temporal and spatial manner to investigate loci activation at specific times during development and further understand how certain loci are activated and why, in some cases, they are brought into close proximity of non-antigen loci, promoting chromosomal translocations.

Another prospect is to study these interactions by FISH (fluorescent *in situ* hybridization). FISH as been already applied to study V(D)J recombination events (Daly et al, 2007; Holwerda et al, 2013), and a similar approach could improve the robustness of the data presented in this thesis. For example, using two different probes for enhancer and promoter, long-range interaction could be studied under different conditions. For example, additional mutagenesis studies, such as mutation of the E47 phosphorylation sites as mentioned in Chapter 4, could be performed and this will expand our current knowledge of the importance of these interactions and why the RAG protein complex equally recognizes these cRSSs.

Table 6-1 Representative translocations associated with lymphoid malignancies.

Type	Translocation	Fusion genes	References
TCR RSS to cryptic RSS	t (7;9)	*TCRβ/tal2*	Tycko et al. 1989
IgH RSS to locus without RSS	t (14;18)	*IgH/bcl2*	Tsujimoto et al. 1985
Coding joint	t (1;19) t (9;11)	*E2A/pbx1* *AF-9/ALL-1*	Wiemels et al. 2002 Negrini et al. 1993
Interstitial RSS	Exon 2/3 deletion	*Hprt*	Finette et al. 2002
Deletion	d(1)	*sil/tal1*	Aplan et al. 1990

Indeed, single amino acids mutants have been prepared to test in the hypothesis that phosphorylation might regulate E47 stability, triggering its degradation and that interaction with IRF4 could possibly protect E47, by masking the phosphorylation sites. As discussed in Chapter 1, different molecular mechanisms control the

transcriptional activity of E47, from regulation by the Id proteins to post translational mechanism, such as phosphorylation events. For example, phosphorylation of serine 140 by the p38 MAPK is known to regulate muscle-specific gene transcription, thus MyoD/E47 heterodimerization (Lluis et al, 2005). It is conceivable to suggest that other phosphorylation events might dictate association of E47-interacting partners, such as IRF4 and PU.1 as described in Chapter 3 and as suggested for the Myo/E47 interaction, p38 MAPK could also exert a negative effect by phosphorylating other residues within E47. In light of the various residues that have been prepared, it will not be surprising if analysis of the phosphorylation status by a MAP kinase would result in different outcomes. Moreover, as it was shown in another study, E47 phosphorylation at serine residues found within the N-terminal domain by AKT (in particular S528) were absent in E12, although the two proteins are two splice variant of the *E2A* gene (Teachenor et al, 2012b). In addition to this, because of two distinct acceptor splice sites flanking the E47 exon, two distinct E47 isoforms where generated, of which only one was phosphorylated by AKT site. This suggests another level of regulation of E47 activity by post-transcriptional modification that should be further investigated also in the context of E47 interaction with IRF4 or PU.1.

In the second part of the thesis next generation sequencing was used to analyse HPV-positive cell lines and tumour biopsies. High-throughput sequencing methodologies have been widely used to investigate differences in gene expression between two different conditions, e.g. normal vs. cancerous. As discussed in Chapter 1, the immune system is a very complex system in which interactions between different regulatory elements allows the generation of a diverse array of antibodies and T cell receptors, specific for certain antigens. However, in the case of an HPV infection, it is known that the immune system somehow oversees the infection, for decades, inevitably triggering cervical cancer. Cancer is itself a multifaceted disease

involving more than one pathway and deregulation of several proteins and molecules.

In this study, high-throughput sequencing methodologies were adopted to analyse differences in transcript expression levels between normal human keratinocytes and HPV-positive tumour samples and three cell lines, representative of both HPV16 and 18. Several transcripts were found to be deregulated, among which a member of the gap junction protein family, Cx26. Although the results suggest that the two viral oncogenes E6 and E7 are affecting both Cx26 mRNA and protein levels, this cellular protein is unlikely to solely account for cancer development; therefore, understanding the pathways that trigger its down-regulation and which other molecules might be involved may help in elucidating the mechanism behind the evasion of immunosurveillance by HPVs.

In 2000, a well-known paper from Hanahan and Weinberg stated "…One day, we imagine that cancer biology and treatment—at present, a patchwork quilt of cell biology, genetics, histopathology, biochemistry, immunology, and pharmacology—will become a science with a conceptual structure and logical coherence that rivals that of chemistry or physics."(Hanahan & Weinberg, 2000). At that time, high throughput technologies were less advanced than today; at present researchers are eager to find differences in transcripts/gene expression between tumour samples versus normal tissues that could be used as biomarkers for cancer detection or prevention. Biomarkers may not only help prevent the disease, but at the same time can help elucidate connections between the various altered molecules within a certain condition, to better formulate a diagnosis and a more direct treatment.

Indeed, in addition to Cx26, other transcripts were of particular interest and should be considered for future studies. For example, CD13 (APN, NCBI Ref.NM_001150) is an amino peptidase N that degrades proteins with an N-terminal neutral amino acid. It is described as a "moonlighting ectoenzyme" given the multitude of proc-

esses it is involved in, such as antigen presentation, receptor for some human viruses (coronaviruses), regulation of cell motility and angiogenesis in colon cancer, non-small cell lung cancer and human pancreatic cancer (Ghosh et al, 2012; Hashida et al, 2002; Ikeda et al, 2003; Tokuhara et al, 2006; Yeager et al, 1992). In this study, APN was found to be down regulated, however it was not further analysed as priority was given those transcripts whose expression was significantly deregulated in all samples analysed. However, the fact that APN is involved in antigen presentation in dendritic cells raises an interesting hypothesis that it might be crucial in keratinocytes as well, which are known to be APCs. In this context, HR-HPVs might downregulate APN as a means to limit cell-cell communication between the cells during viral infection, and thus bypass the immune response. This is a hypothesis that should be considered in future work to complement our understanding of viral infection and immune escape by HR-HPVs. To this end, a plasmid expressing APN has been prepared to better understand the role played by this protein.

Future work should also be addressed to understand whether the virus targets the same cellular protein in different cells. HPV is known to be the causative agent of head and neck cancers as well. To this end, head and neck cell carcinoma have been sent for sequencing to analyse their transcriptome. A comparison between the data will allow a better understanding of the transcripts targeted by the virus.

Chapter 7. APPENDICES

7.1 Appendix 1: Supplemental figures

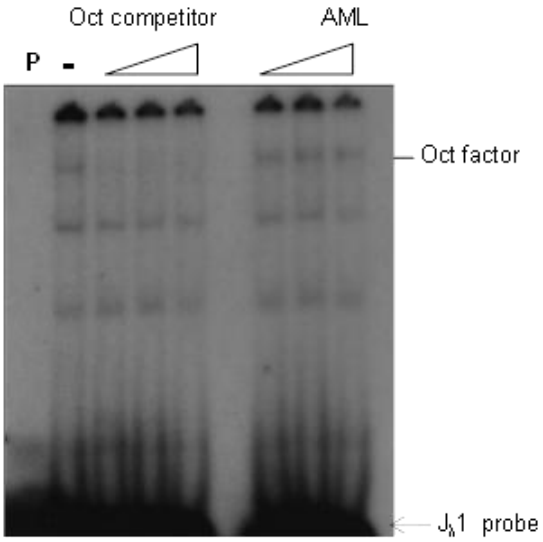


Figure 7-1 Competition assay on the Jλ1 probe.

The competition assay shows that the band to which the extract is binding does correspond to an octamer factor since the band is competed off only in present of the specific oligonucleotide carrying a binding site for an octamer factor.

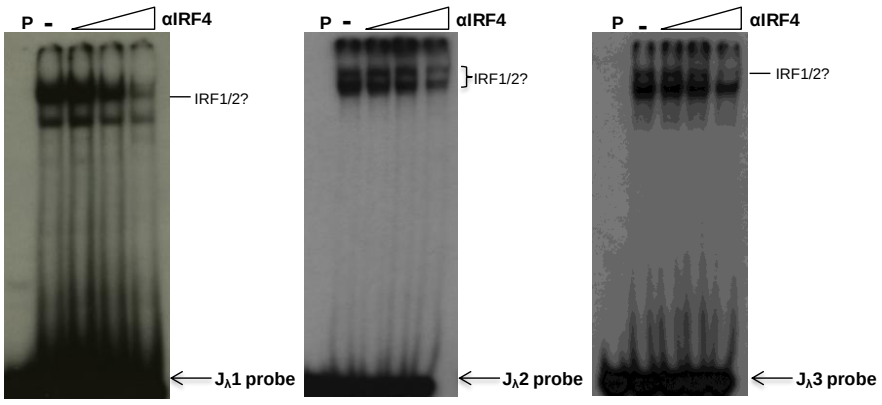


Figure 7-2 Antibody competition to identify IRF1/2 transcription factor.

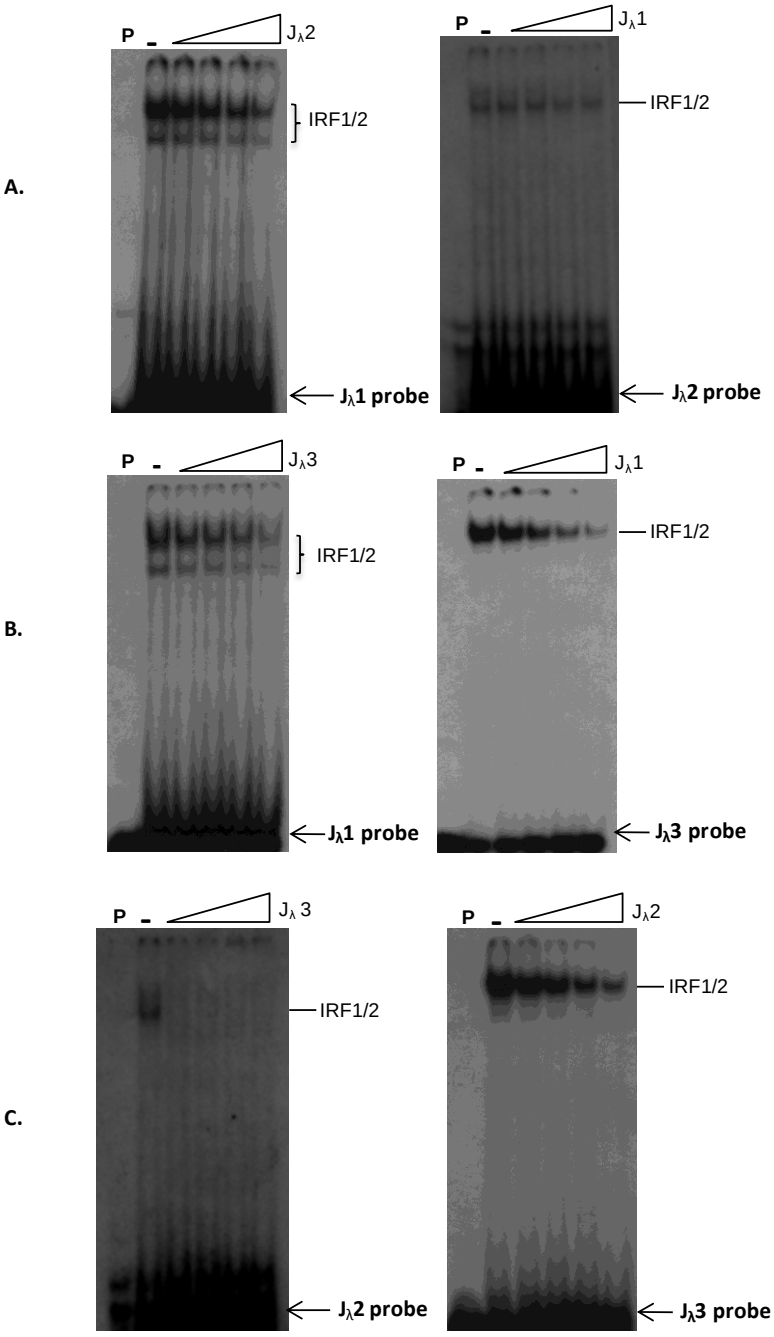


Figure 7-3 Cross competition for IRF1/2.

Although the same factor seems to bind the three promoters, subsequent testing failed to reveal the identity of such factor. Open arrows indicate increasing amount of competitor.

Table 7-1 RLU and standard deviation (SD) values obtained in luciferase assays.

PLASMID	RLU	+/- SD
pgl3	0.99	8.1733E-08
Jλ1+E47	8.5	3.93906815
Jλ1+E47/IRF4	34.37	1.5985352
Eλ3-1	89.92	4.05284523
Jλ1/E47mt	7.46	2.80063692
Jλ1/E47m/IRF4	6.96	2.83318882
Eλ3-1/Jλ1 E47mt	47.02	1.26710013
Eλ3-1/Jλ1 /IRF4mt	40.1696408	5.79311503
Eλ3-1/Jλ1 /IRF4mt /PU.1mt	46.2760251	10.6707235
Eλ3-1/Jλ1/IRF4 L24A	32.7761891	9.544866715
Eλ3-1/Jλ1/IRF4 K399A	55.84231	16.07477412
Eλ3-1/Jλ1/IRF4 L368A	36.9322176	4.967296607
Jλ1/IRF4 L24A	16.1391114	7.68555232
Jλ1/IRF4 L368A	25.6052831	18.251045

Table 7-2 Luciferase assay raw data.

Sample	Luciferase	Renilla	Luciferase/ Renilla ratio	Normalized pGL3/Renilla
pgl3	2472	297307	0.008314638	0.618107888
pgl3	4873	262146	0.018588878	1.381892088
pgl3 average			0.013451758	
2	29273	151179	0.193631391	14.39450447
2	23268	127406	0.182628762	13.57657204
3	44263	162632	0.272166609	20.23279108
3	46294	157287	0.294328203	21.88027785
4	89340	166214	0.53749985	39.95759139
4	81692	142834	0.571936654	42.51761396
5	74580	285580	0.261152742	19.41402319
5	57586	361109	0.159469855	11.85494531

The table shows an example of raw data obtained in the luciferase assays. To obtain the values as shown in the graphs in Chapter 4, Renilla values where divided by Luciferase values (Luciferase/Renilla ratio). These, where further normalised by the average Luciferase/Renilla ratio value calculated as mean of the two repeats for pgl3 (highlighted in yellow). Using this latter value, normalized values were calculated for each sample by dividing Luciferase/Renilla ratio values by the average pgl3 values (yellow). Average of the normalized values, standard deviation (SD) and standard error of the mean (SEM) were calculated. The latter was used to draw error bars. Samples chosen are just an example of early experiments, which were performed only in duplicates for titration purposes. 2: Promoter; 3: Enhancer+ IRF4; 4: Enhancer+IRF4+PU.1; 5: Enhancer+PU.1.

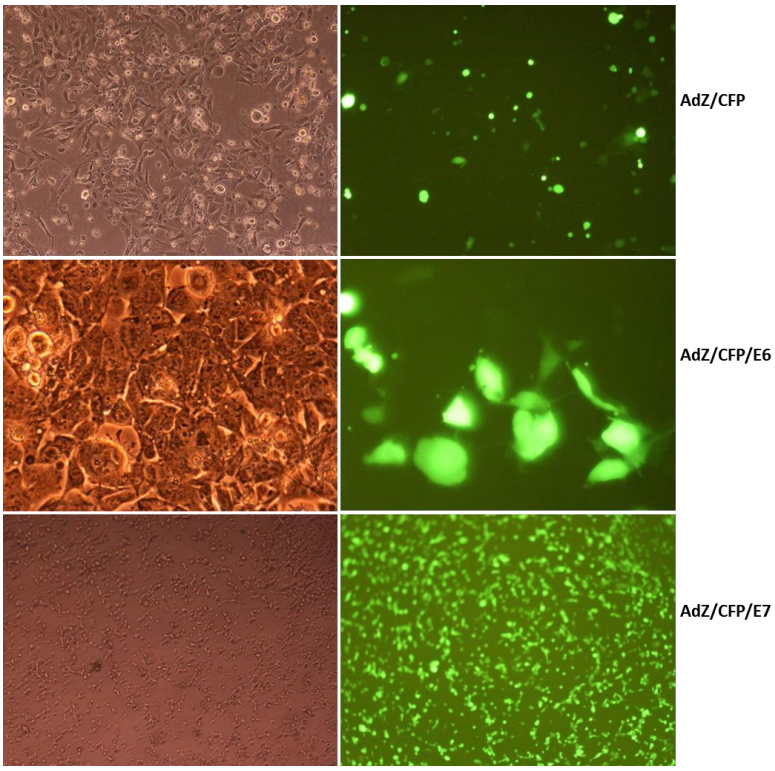


Figure 7-4 911 cells expressing CFP.

911 cells were transfected with the recombinant adenoviruses and 2-3 days later pictures were taken to confirm expression of CFP.

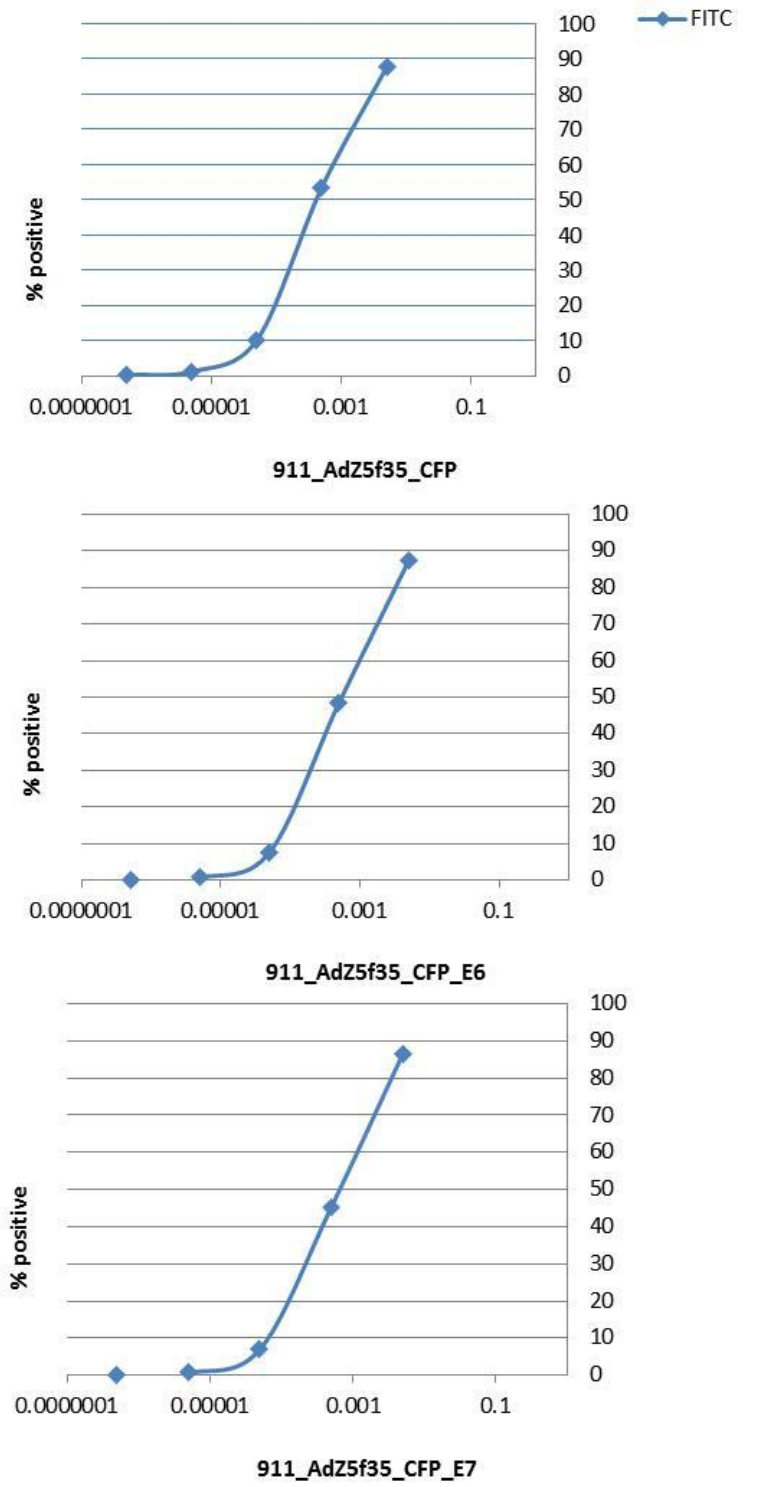


Figure 7-5 Viral infective particles concentration for AdZ/CFP.

The graphs were used to extrapolate the dilution of the virus stock that killed 50% of the infected cells. These values were used to calculate the actual volume of purified virus to be used.

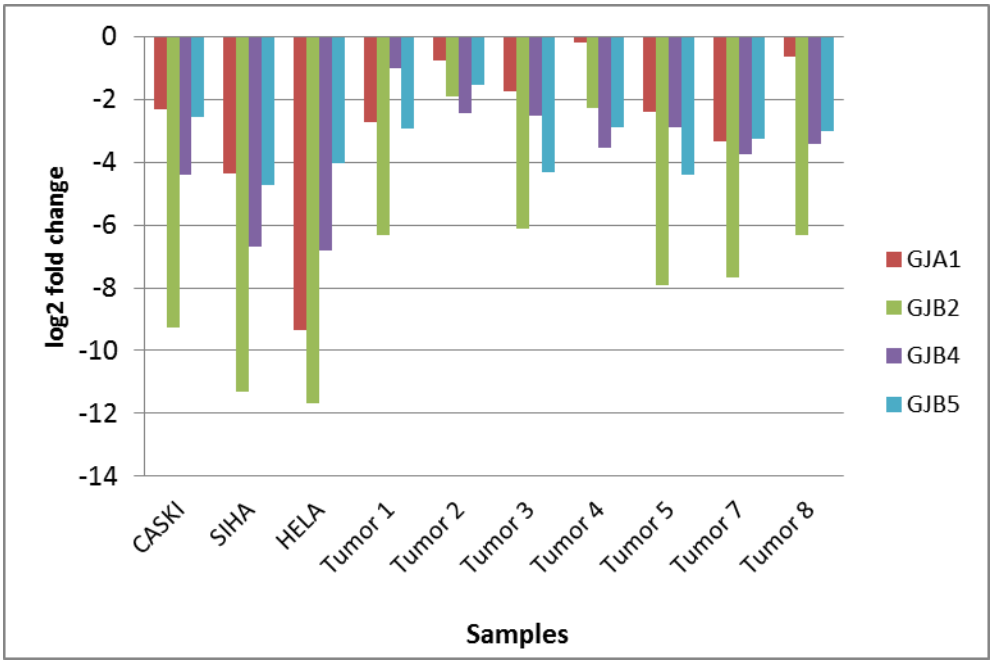


Figure 7-6 Gap junctions fold change values using GFOLD.

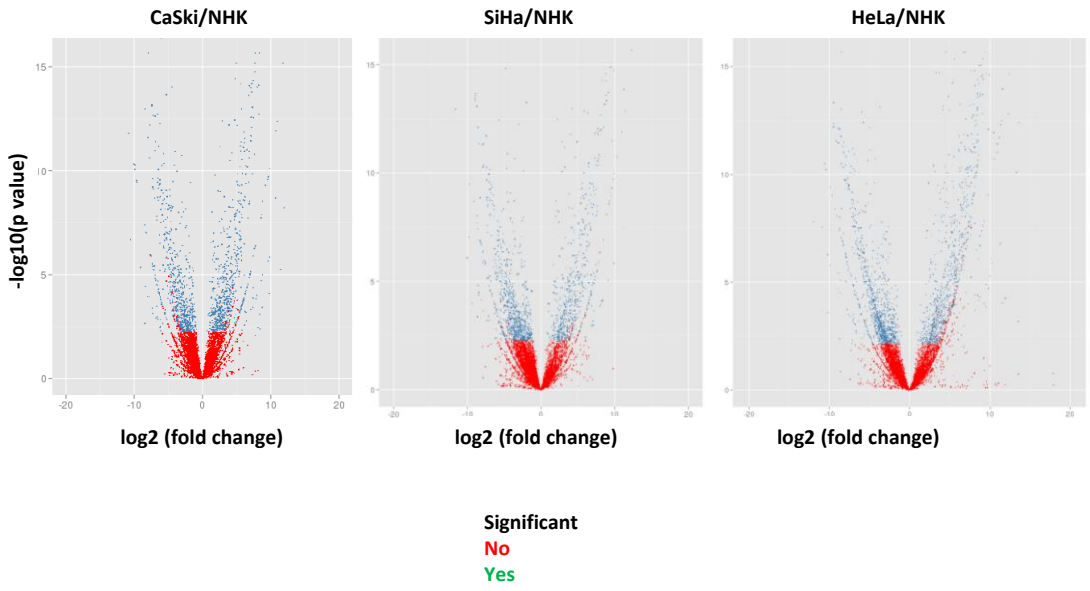


Figure 7-7 Volcano plots for transcripts comparison between HPV-positive cells lines and NHK.

Transcripts expression is shown as log2(fold change) vs log10 (p values) to output the ratio between the significant vs non-significant transcripts The plot was generated to complement the heatmaps used in Chapter 5 (Figure 5-1).

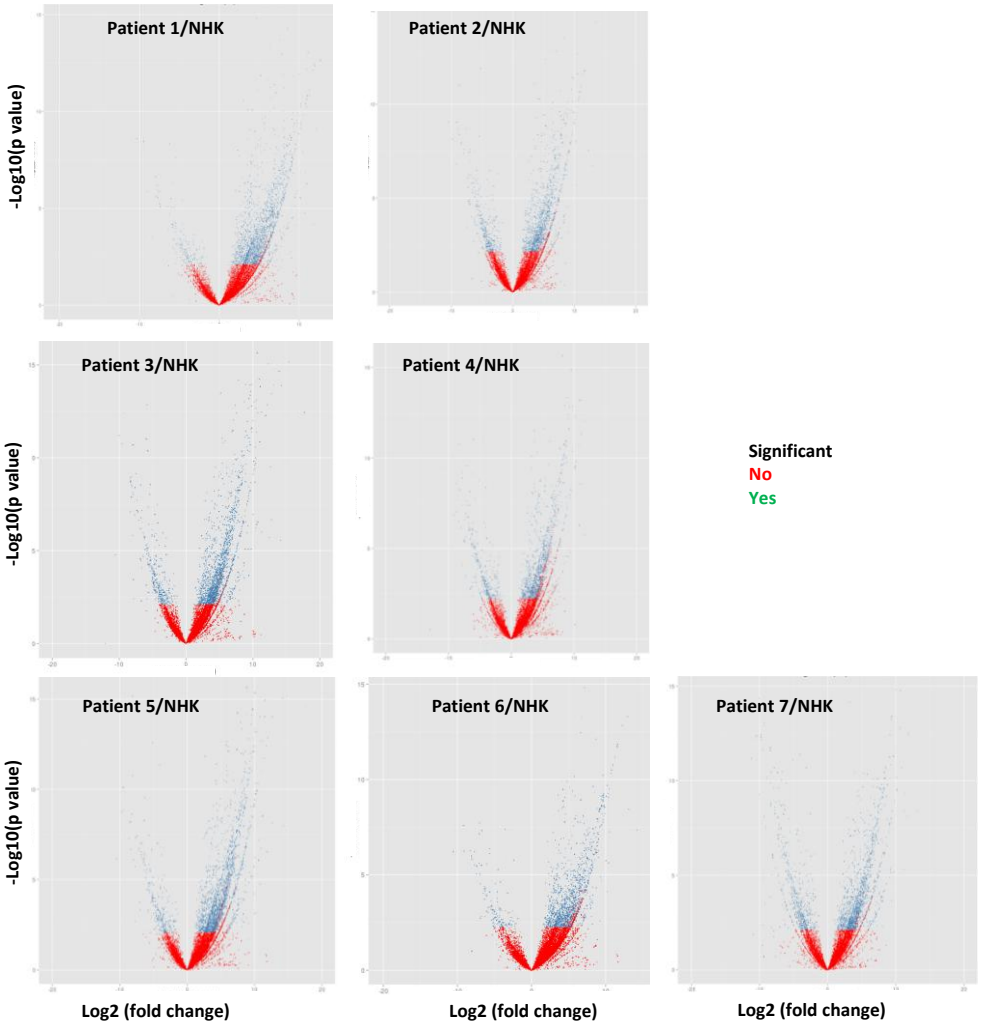


Figure 7-8 Volcano plots for transcripts comparison between patients samples and NHK.

Transcripts expression is shown as Log2(fold change) vs Log10 (p values) to output the ratio between the significant vs non-significant transcripts. The plot was generated to complement the heatmaps used in Chapter 5 (Figure 5-2).

7.2 Appendix 2: Selected analysis tools for NGS

Table 7-3 Selected analysis tools for NGS/RNA-Seq data.

Galaxy	https://usegalaxy.org/
FastQC	http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
TopHat	http://tophat.cbcb.umd.edu/
Bowtie	http://bowtie-bio.sourceforge.net/bowtie2/manual.shtml
Cufflinks	http://cufflinks.cbcb.umd.edu/
Cuffdiff	http://cufflinks.cbcb.umd.edu/howitworks.html/
CummeRbund	http://compbio.mit.edu/cummeRbund/
IGV	http://www.broadinstitute.org/software/igv/
GFOLD	http://web.tongji.edu.cn/~zhanglab/GFOLD/
UCSC	https://genome.ucsc.edu/index.html
R	http://www.bioconductor.org/
Bioinformatics forums	http://seqanswers.com/

7.3 Appendix 3: Sequencing analysis of the recombinant adenoviruses

Green, Kozak sequence; Red, Stop codon; Yellow, Tag sequence. Reverse/Forward = primers used for sequencing.

AdZ/CFP/E6_N_terminus HA sequence

```
(Reverse) ----------------------------------------ganngCCTG
Reference gagatctccctatcagtgatagagagtttagtgaaccgtcagatcgcctg
(Forward) ----------------------------------------------------
(Reverse) GAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACCGGGACCGAT
Reference gagacgccatccacgctgttttgacctccatagaagacaccgggaccgat
(Forward) ----------------------------------------------------
(Reverse) CCAGCCTGGATCCCCTACCATGTACCCATACGACGTCCCAGACTACGCTA
Reference ccagcctggatcccctACCatgTACCCATACGACGTCCCAGACTACGCTa
(Forward) ----------------------------------------------------
(Reverse) TGCACCAAAAGAGAACTGCAATGTTTCAGGACCCACAGGAGCGACCCAGA
Reference tgcaccaaaagagaactgcaatgtttcaggacccacaggagcgacccaga
(Forward) ----------------------------------------------GA
(Reverse) AAGTTACCACAGTTATGCACAGAGCTGCAAACAACTATACATGATATAAT
Reference aagttaccacagttatgcacagagctgcaaacaactatacatgatataat
(Forward) nngTTACCaCAGTTATGCACAGAGCTGCAAACAACTATACATGATATAAT
(Reverse) ATTAGAATGTGTGTACTGCAAGCAACAGTTACTGCGACGTGAGGTATATG
Reference attagaatgtgtgtactgcaagcaacagttactgcgacgtgaggtatatg
(Forward) ATTAGAATGTGTGTACTGCAAGCAACAGTTACTGCGACGTGAGGTATATG
(Reverse) ACTTTGCTTTTCGGGATTTATGTATAGTATATAGAGATGGGAATCCATAT
Reference actttgcttttcgggatttatgtatagtatatagagatgggaatccatat
(Forward) ACTTTGCTTTTCGGGATTTATGTATAGTATATAGAGATGGGAATCCATAT
(Reverse) GCTGTATGTGATAAATGTTTAAAGTTTTATTCTAAAATTAGTGAGTATAG
Reference gctgtatgtgataaatgtttaaagttttattctaaaattagtgagtatag
(Forward) GCTGTATGTGATAAATGTTTAAAGTTTTATTCTAAAATTAGTGAGTATAG
(Reverse) ACATTATTGTTATAGTGTGTATGGAACAACATTAGAACAGCAATACAACA
Reference acattattgttatagtgtgtatggaacaacattagaacagcaatacaaca
(Forward) ACATTATTGTTATAGTGTGTATGGAACAACATTAGAACAGCAATACAACA
(Reverse) AACCGTTGTGTGATTTGTTAATTAGGTGTATTAACTGTCAAAAGCCACTG
Reference aaccgttgtgtgatttgttaattaggtgtattaactgtcaaaagccactg
(Forward) AACCGTTGTGTGATTTGTTAATTAGGTGTATTAACTGTCAAAAGCCACTG
(Reverse) TGTCCTGAAGAAAAGCAAAGACATCTGGACAAAAAGCAAAGATTCCATAA
Reference tgtcctgaagaaaagcaaagacatctggacaaaaagcaaagattccataa
(Forward) TGTCCTGAAGAAAAGCAAAGACATCTGGACAAAAAGCAAAGATTCCATAA
```

```
(Reverse) TATAAGGGGTCGGngGACCGGTCGAngt----------------------
Reference tataaggggtcggtggaccggtcgatgtatgtcttgttgcagatcatcaa
(Forward) TATAAGGGGTCGGTGGACCGGTCGATGTATGTCTTGTTGCAGATCATCAA
(Reverse) ------------------------------------------------
Reference gaacacgtagagaaacccagctgTAAggatcccacgtcactattgtatac
(Forward) GAACACGTAGAGAAACCCAGCTGTAAGGATCCCACGTCACTATTGTATAC
(Reverse) ------------------------------------------------
Reference tctatattatactctatgttatactctgtaatcctactcaataaacgtgt
(Forward) TCTATATTATACTCTATGTTATACTCTGTAATCCTACTcAaTAAACGTgt
(Reverse) ------------------------------------------------
Reference cacgcctgtgaaaccgtactaagtctcccgtgtcttcttatcaccatcag
(Forward) caA---------------------------------------------
```

AdZ/CFP/E7_C_terminus_Myc sequence

```
Reference --------------ttgacgcaaatgggcggtaggcgtgtacggtgggag
(Forward) anaACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAG
Reference gtctatataagcagagctctccctatcagtgatagagatctccctatcag
(Forward) GTCTATATAAGCAGAGCTCTCCCTATCAGTGATAGAGATCTCCCTATCAG
Reference tgatagagagtttagtgaaccgtcagatcgcctggagacgccatccacgc
(Forward) TGATAGAGAGTTTAGTGAACCGTCAGATCGCCTGGAGACGCCATCCACGC
Reference tgttttgacctccatagaagacaccgggaccgatccagcctggatccAcc
(Forward) TGTTTTGACCTCCATAGAAGACACCGGGACCGATCCAGCCTGGATCCACC
Reference ATGcatggagatacacctacattgcatgaatatatgttagatttgcaacc
(Forward) ATGCATGGAGATACACCTACATTGCATGAATATATGTTAGATTTGCAACC
Reference agagacaactgatctctactgttatgagcaattaaatgacagctcagagg
(Forward) AGAGACAACTGATCTCTACTGTTATGAGCAATTAAATGACAGCTCAGAGG
Reference aggaggatgaaatagatggtccagctggacaagcagaaccggacagagcc
(Forward) AGGAGGATGAAATAGATGGTCCAGCTGGACAAGCAGAACCGGACAGAGCC
Reference cattacaatattgtaacctttgttgcaagtgtgactctacgcttcggtt
(Forward) CATTACAATATTGTAACCTTTTGTTGCAAGTGTGACTCTACGCTTCGGTT
Reference gtgcgtacaaagcacacacgtagacattcgtactttggaagacctgttaa
(Forward) GTGCGTACAAAGCACACACGTAGACATTCGTACTTTGGAAGACCTGTTAA
Reference tgggcacactaggaattgtgtgccccatctgttctcagaaaccaGAACAA
(Forward) TGGGCACACTAGGAATTGTGTGCCCCATCTGTTCTCAGAAACCAGAACaA
Reference AAACTTATTTCTGAAGAAGATCTGTAAggatcccacgtcactattgtata
(Forward) AAACTTATTTCTGAAGAAGATCTGTAAGGATCCCacgtCacTAttGTATA
Reference ctctatattatactctatgttatactctgtaatcctactcaataaacgtg
(Forward) CTCTATATTATACTctATGTTATACTCTGTaatCctACtcAATaaAcnt-
Reference tcacgcctgtgaaaccgtactaagtctcccgtgtcttcttatcaccatca
(Forward) ------------------------------------------------
```

7.4 Appendix 4: NGS/RNA-Seq analysis parameters

The parameters were extrapolated from the keratinocytes dataset but they apply to all the samples analysed.

TOOL: TOPHAT FOR ILLUMINA

Name:	Tophat for Illumina on data 1: accepted_hits
Created:	Oct 17, 2012
Filesize:	596.1 MB
Dbkey:	hg19
Format:	bam
Galaxy Tool Version:	1.5.0
Tool Version:	
Tool Standard Output:	stdout
Tool Standard Error:	stderr
Tool Exit Code:	None
API ID:	6651297061e28527

Input Parameter	Value	Note for rerun
RNA-Seq FASTQ file	1: kerat_raw.fq	
Use a built in reference genome or own from your history	indexed	
Select a reference genome	/galaxy/data/hg19/bowtie_index/hg19	
Is this library mate-paired?	single	
TopHat settings to use	full	
Library Type	FR Unstranded	
Anchor length (at least 3)	8	
Maximum number of mismatches that can appear in the anchor region of spliced alignment	0	
The minimum intron length	70	
The maximum intron length	500000	
Allow indel search	Yes	
Max insertion length.	3	
Max deletion length.	3	
Maximum number of alignments to be allowed	20	
Minimum intron length that may be found during split-segment (default) search	50	

TOOL: TOPHAT FOR ILLUMINA

Maximum intron length that may be found during split-segment (default) search	500000
Number of mismatches allowed in the initial read mapping	2
Number of mismatches allowed in each segment alignment for reads mapped independently	2
Minimum length of read segments	25
Use Own Junctions	No
Use Closure Search	No
Use Coverage Search	Yes
Minimum intron length that may be found during coverage search	50
Maximum intron length that may be found during coverage search	20000
Use Microexon Search	No

TOOL: CUFFLINKS

Name:	Cufflinks on data 8 and data 25K: assembled transcripts
Created:	Jul 16, 2012
Filesize:	93.7 MB
Dbkey:	hg19
Format:	gtf
Galaxy Tool Version:	0.0.5
Tool Version:	
Tool Standard Output:	stdout
Tool Standard Error:	stderr
Tool Exit Code:	None
API ID:	7dd06fde8fb56fe3

Input Parameter	Value	Note for rerun
SAM or BAM file of aligned RNA-Seq reads	8: Tophat for Illumina on data 1: accepted_hits	
Max Intron Length	300000	
Min Isoform Fraction	0.1	
Pre MRNA Fraction	0.15	
Perform quartile normalization	Yes	

TOOL: TOPHAT FOR ILLUMINA

Use Reference Annotation	Use reference annotation
Reference Annotation	25: UCSC Main on Human: refGene (genome)
Perform Bias Correction	Yes
Reference sequence data	cached
Use multi-read correct	*not used (parameter was added after this job was run)*
Use effective length correction	*not used (parameter was added after this job was run)*
Global model (for use in Trackster)	No dataset

TOOL: CUFFCOMPARE

Name:	Cuffcompare on data 1, data 4, and others: combined transcripts
Created:	Jul 22, 2012
Filesize:	92.8 MB
Dbkey:	hg19
Format:	gtf
Galaxy Tool Version:	0.0.5
Tool Version:	
Tool Standard Output:	stdout
Tool Standard Error:	stderr
Tool Exit Code:	None
API ID:	09b8046b0fe0fca0

Input Parameter	Value	Note for rerun
GTF file produced by Cufflinks	1: Cufflinks on data 68 and data 79 caski: assembled transcripts	
GTF file produced by Cufflinks	2: Cufflinks on data 8 and data 25 kerat: assembled transcripts	
GTF file produced by Cufflinks	3: Cufflinks on data 70 and data 75 siha: assembled transcripts	
Use Reference Annotation	Yes	
Reference Annotation	4: UCSC Main on Human: refGene (genome)	
Ignore reference transcripts that are not overlapped by any transcript in input files	False	
Use Sequence Data	Yes	
Choose the source for the reference list	cached	

TOOL: CUFFDIFF

TOOL: TOPHAT FOR ILLUMINA

Name:	Cuffdiff on data 12, data 14, and others: transcript differential expression testing
Created:	Jul 29, 2012
Filesize:	5.2 MB
Dbkey:	hg19
Format:	tabular
Galaxy Tool Version:	0.0.5
Tool Version:	
Tool Standard Output:	stdout
Tool Standard Error:	stderr
Tool Exit Code:	None
API ID:	6318041400ee9c4c

Input Parameter	Value	Note for rerun
Transcripts	153: Cuffcompare on data 1, data 4, and others: combined transcripts	
Repeat (Condition)	*not used (parameter was added after this job was run)*	
Library normalization method	*not used (parameter was added after this job was run)*	
Dispersion estimation method	*not used (parameter was added after this job was run)*	
False Discovery Rate	0.05	
Min Alignment Count	10	
Use multi-read correct	*not used (parameter was added after this job was run)*	
Perform Bias Correction	Yes	
Reference sequence data	cached	
Include Read Group Datasets	*not used (parameter was added after this job was run)*	
Set Additional Parameters? (not recommended for paired-end reads)	*not used (parameter was added after this job was run)*	

7.5 Appendix 5: FASTQC results

Keratinocytes

https://usegalaxy.org/datasets/b24ed2066d2f400b/display?to_ext=html/

CaSki

https://usegalaxy.org/datasets/3de7e21bd8faa424/display?to_ext=html/

SiHa

https://usegalaxy.org/datasets/daa5b3d8b71c898f/display?to_ext=html/

Patient 1

https://usegalaxy.org/datasets/5bf80111e06db349/display?to_ext=html/

Patient 2:

https://usegalaxy.org/datasets/8fbe4d77d9eb3275/display?to_ext=html/

Patient 3:

https://usegalaxy.org/datasets/c81eb53a531c4316/display?to_ext=html/

Patient 4:

https://usegalaxy.org/datasets/9ee21b6aa04e50a2/display?to_ext=html/

Patient 5:

https://usegalaxy.org/datasets/debc1daa5ca2d906/display?to_ext=html/

Patient 6:

https://usegalaxy.org/datasets/f9d179e51850f7c2/display?to_ext=html/

Patient 7:

https://usegalaxy.org/datasets/204da6c38255d7ff/display?to_ext=html/

7.6 Appendix 6: Significantly deregulated cell surface transcripts

Figure 7-9 Top 20 and bottom 20 significantly deregulated transcripts in CaSki cell lines.

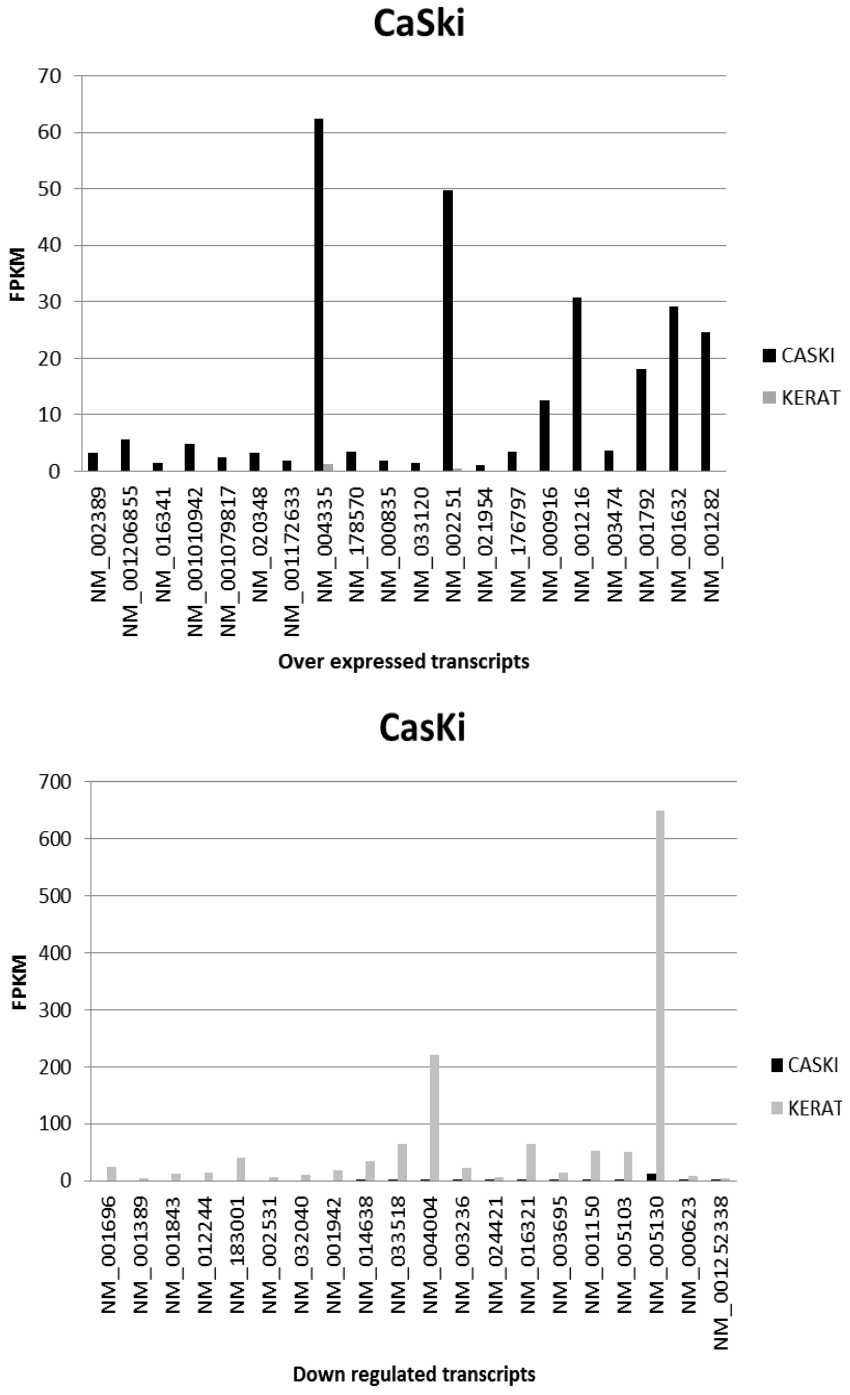


Figure 7-10 Top 20 and bottom 20 significantly deregulated transcripts in SiHa cell lines.

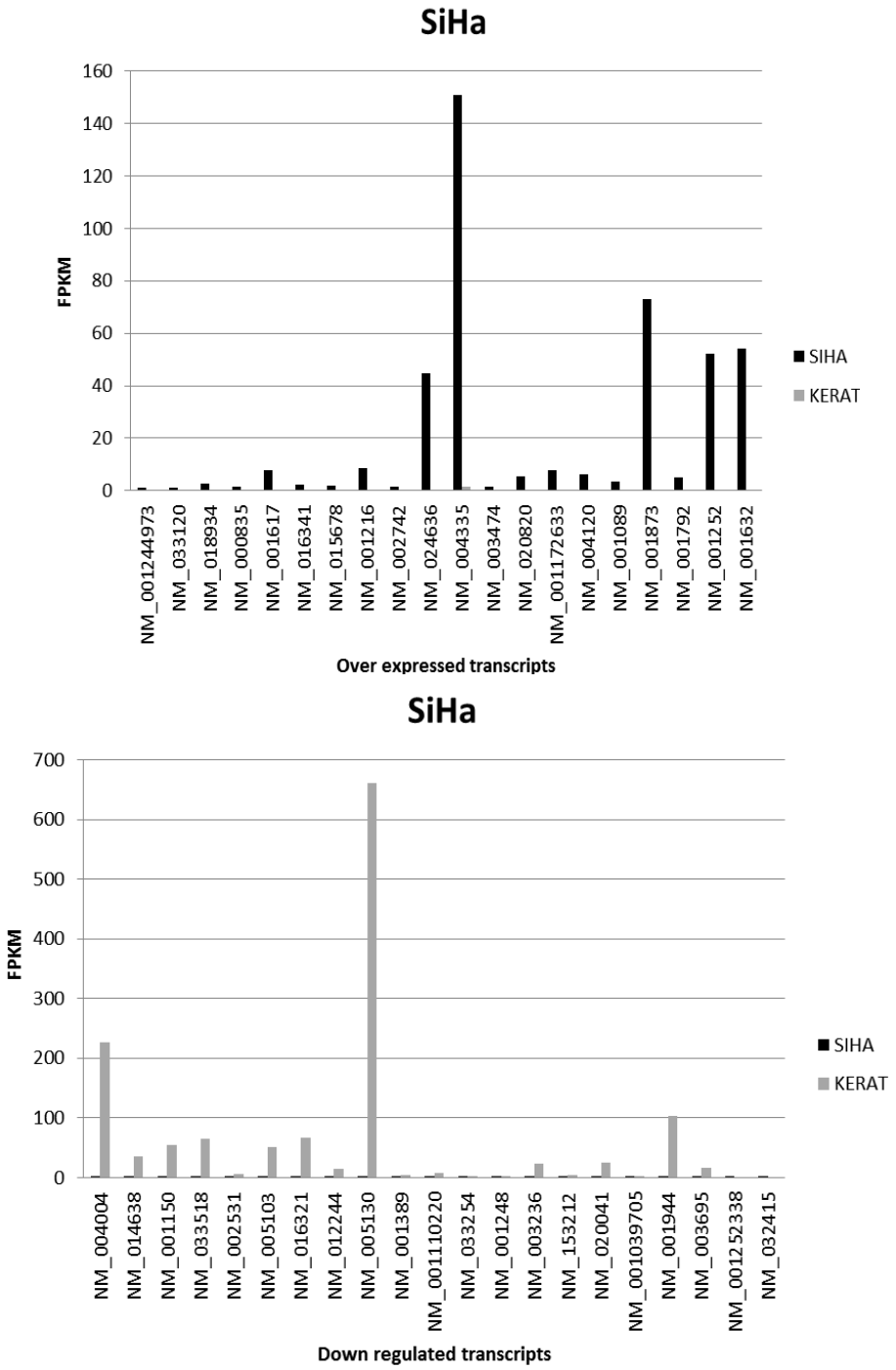
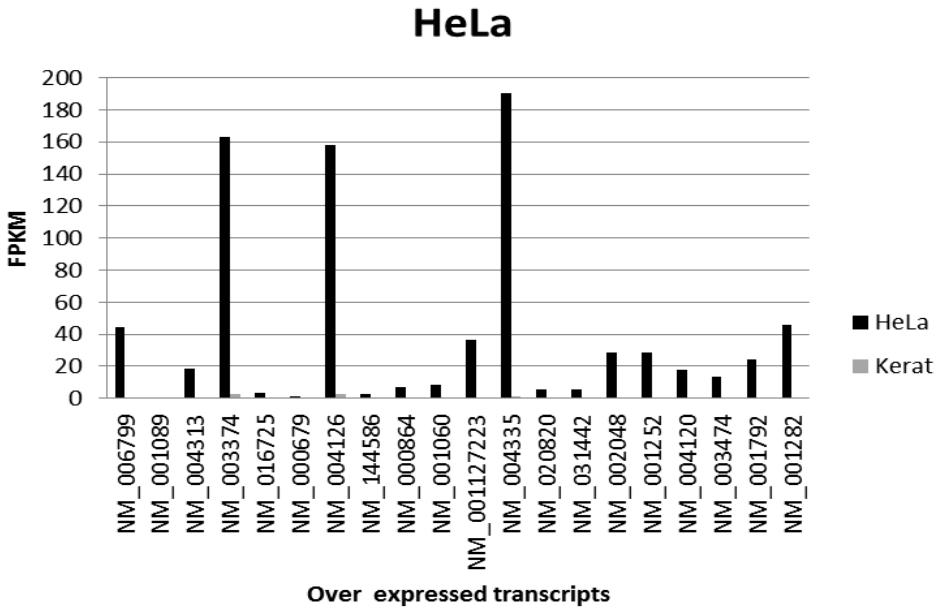


Figure 7-11 Top 20 and bottom 20 significantly deregulated transcripts in HeLa cell lines.



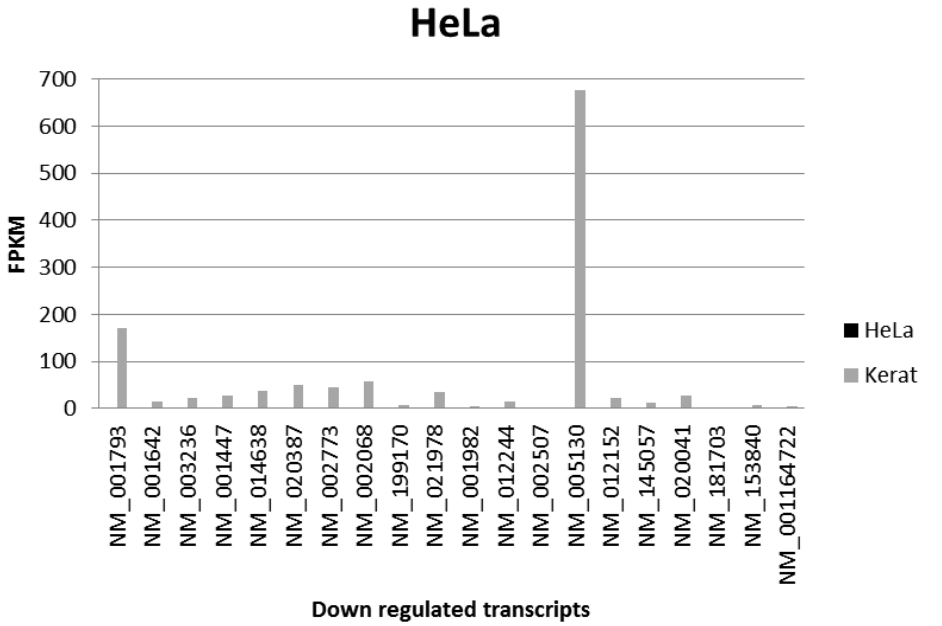
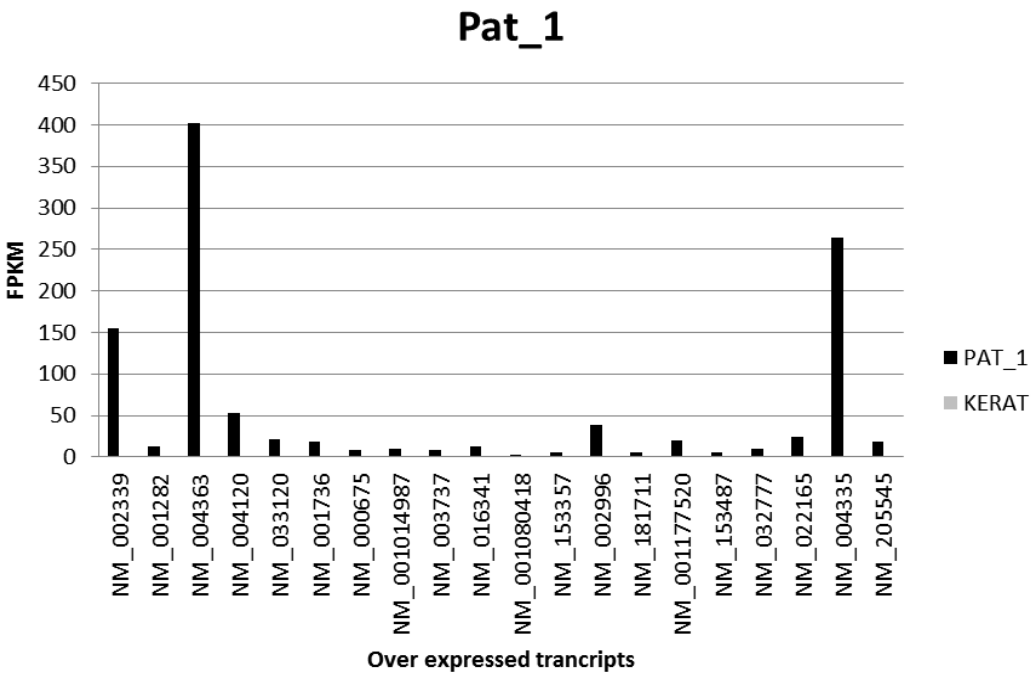


Figure 7-12 Top 20 and bottom 20 significantly deregulated transcripts in patient 1.



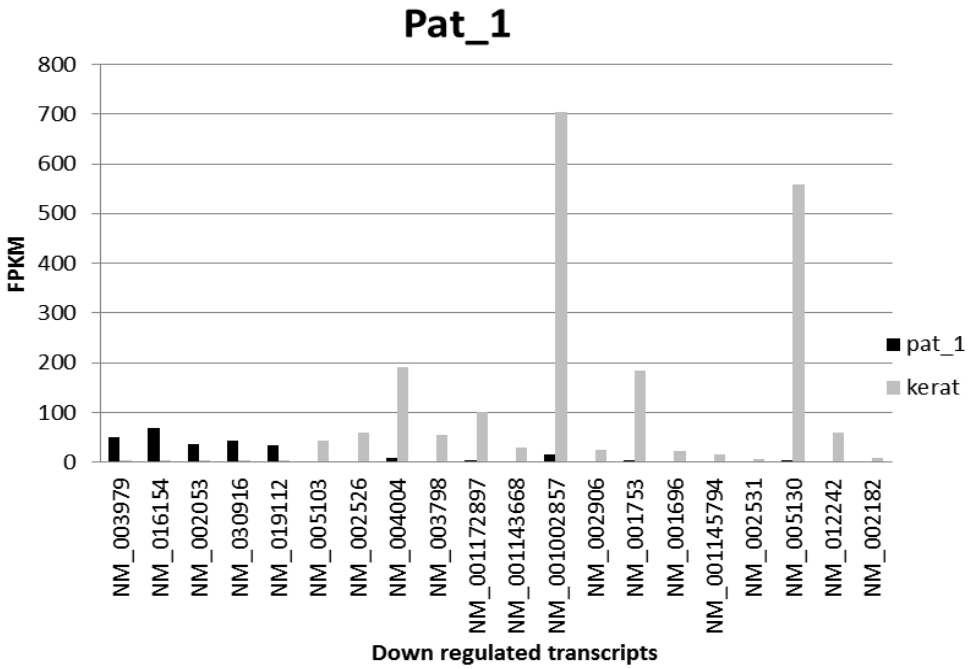
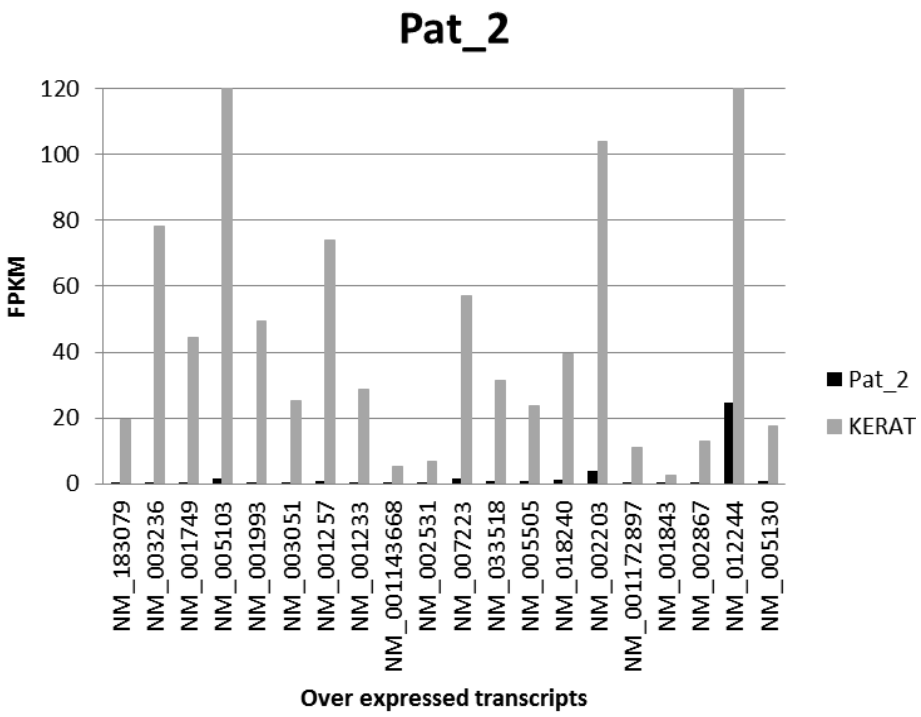


Figure 7-13 Top 20 and bottom 20 significantly deregulated transcripts in patient 2.



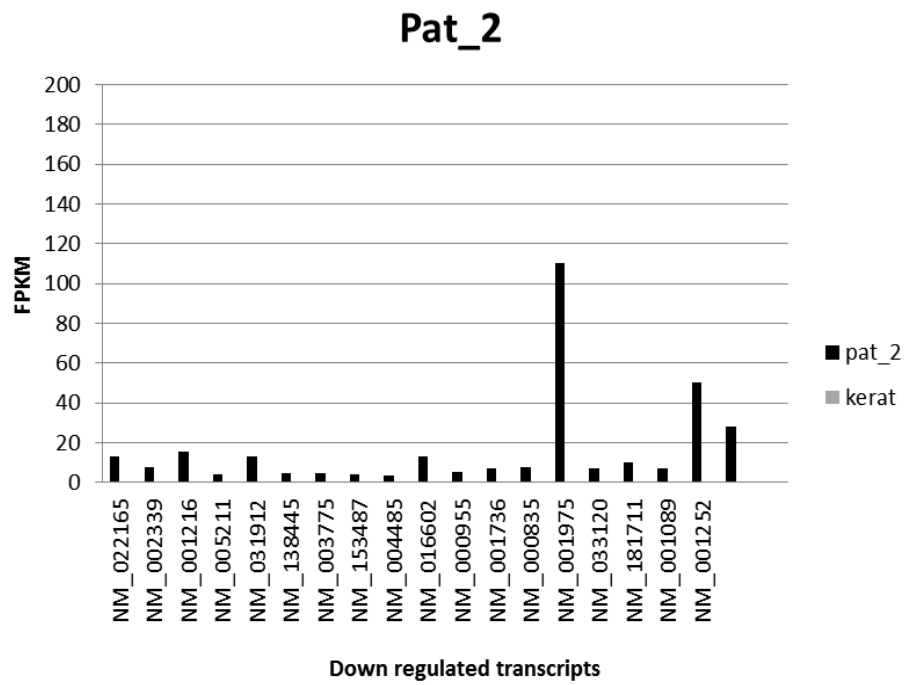
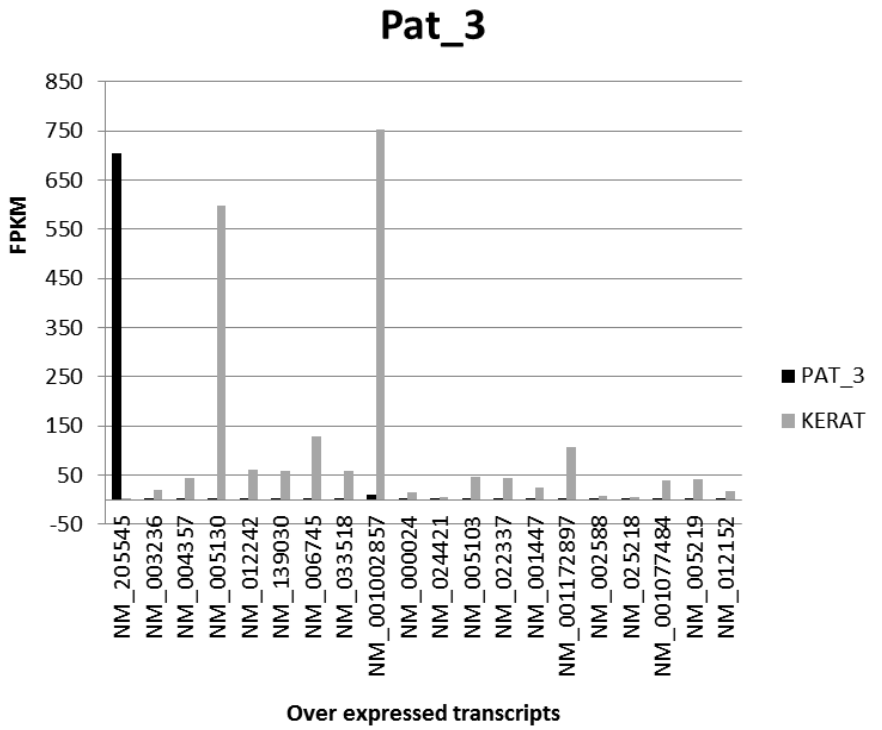


Figure 7-14 Top 20 and bottom 20 significantly deregulated transcripts in patient 3.



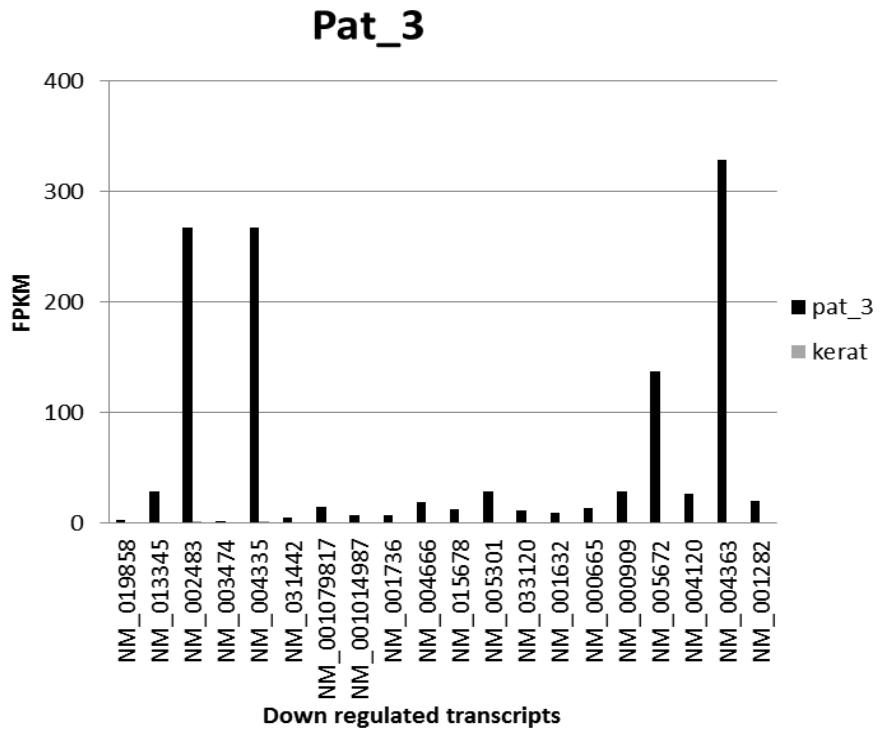
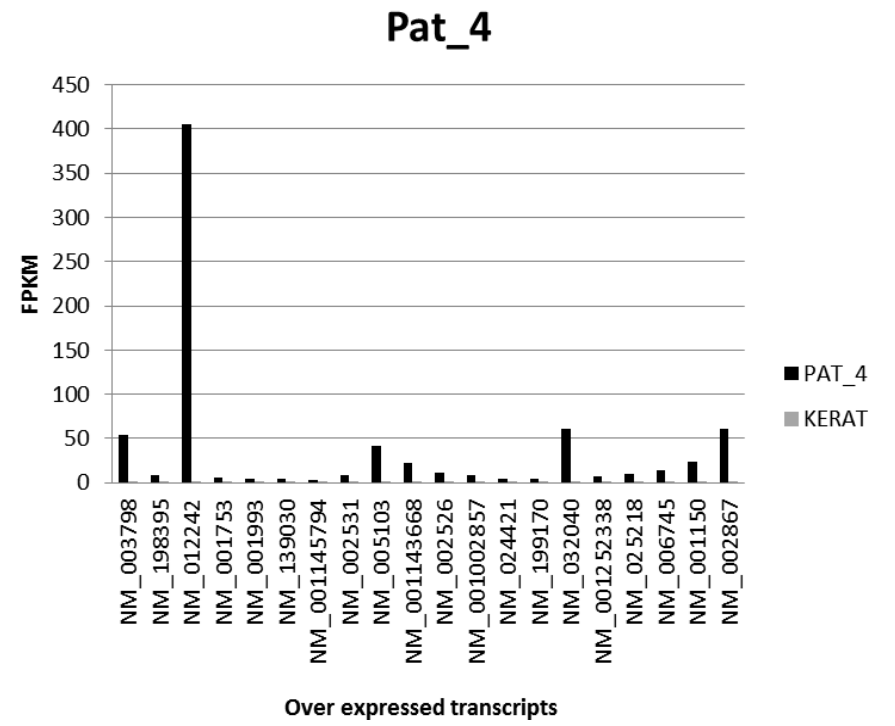


Figure 7-15 Top 20 and bottom 20 significantly deregulated transcripts in patient 4.



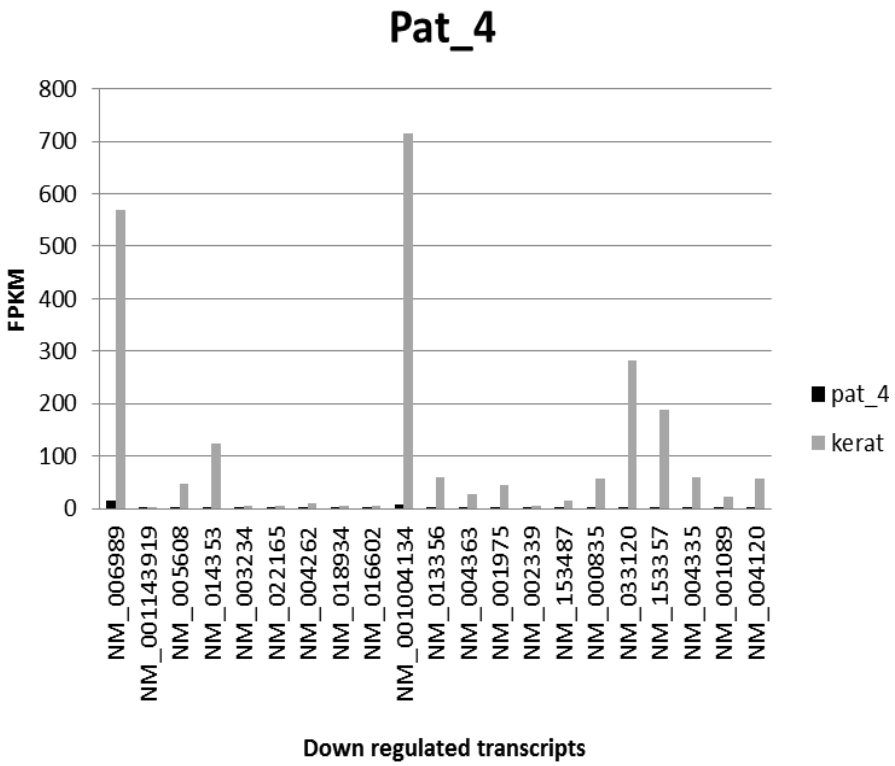
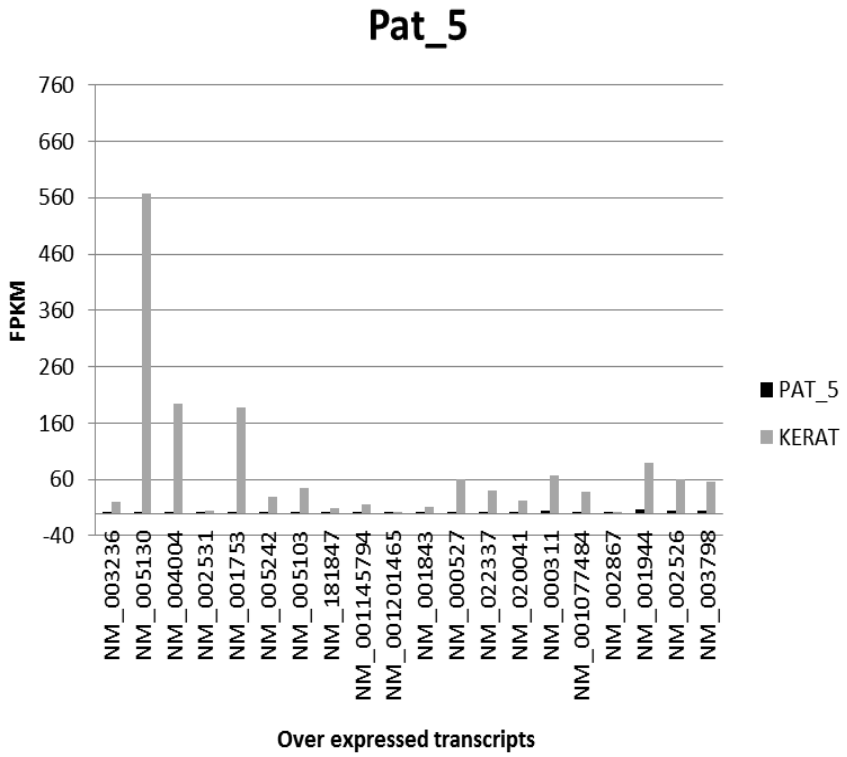


Figure 7-16 Top 20 and bottom 20 significantly deregulated transcripts in patient 5.



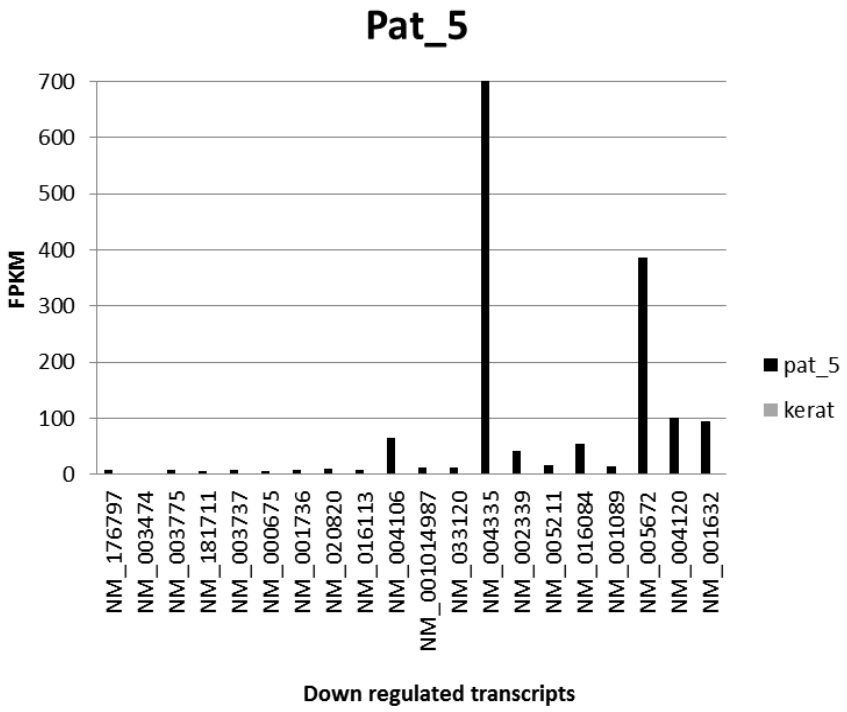
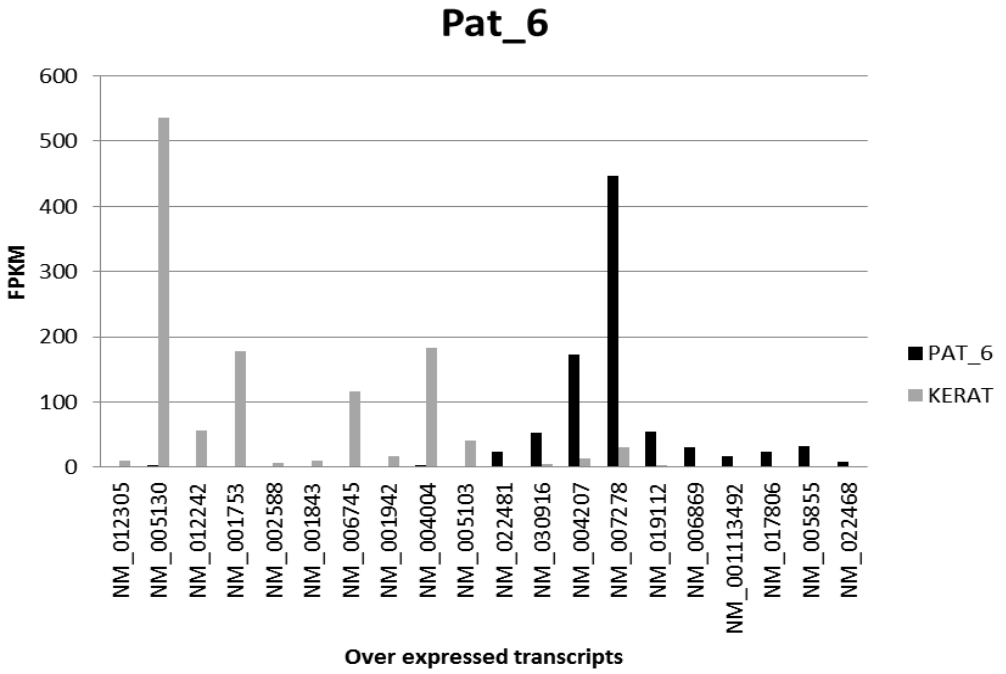


Figure 7-17 Top 20 and bottom 20 significantly deregulated transcripts in patient 6.



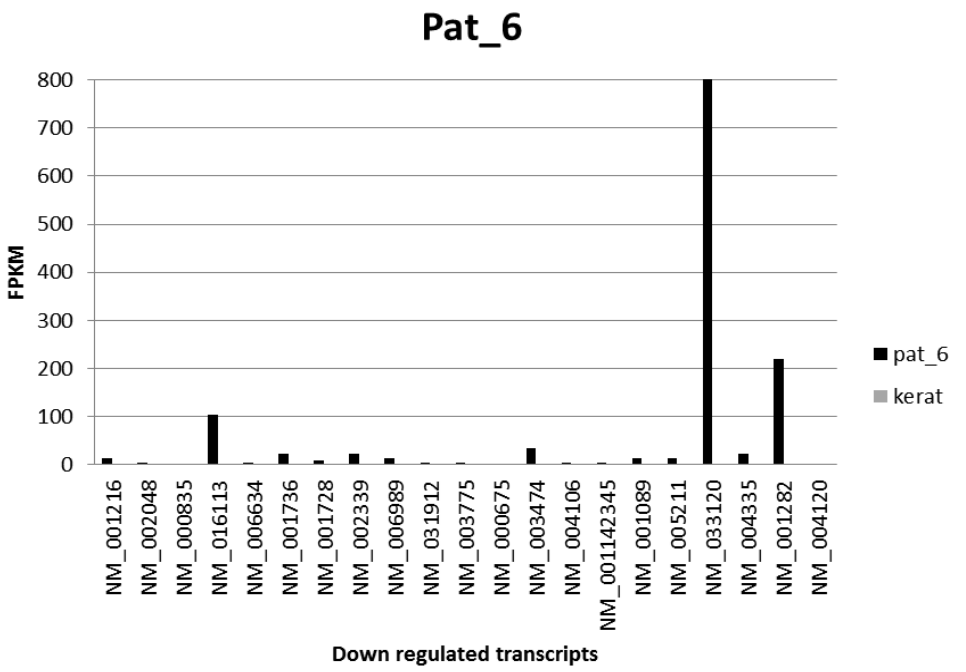
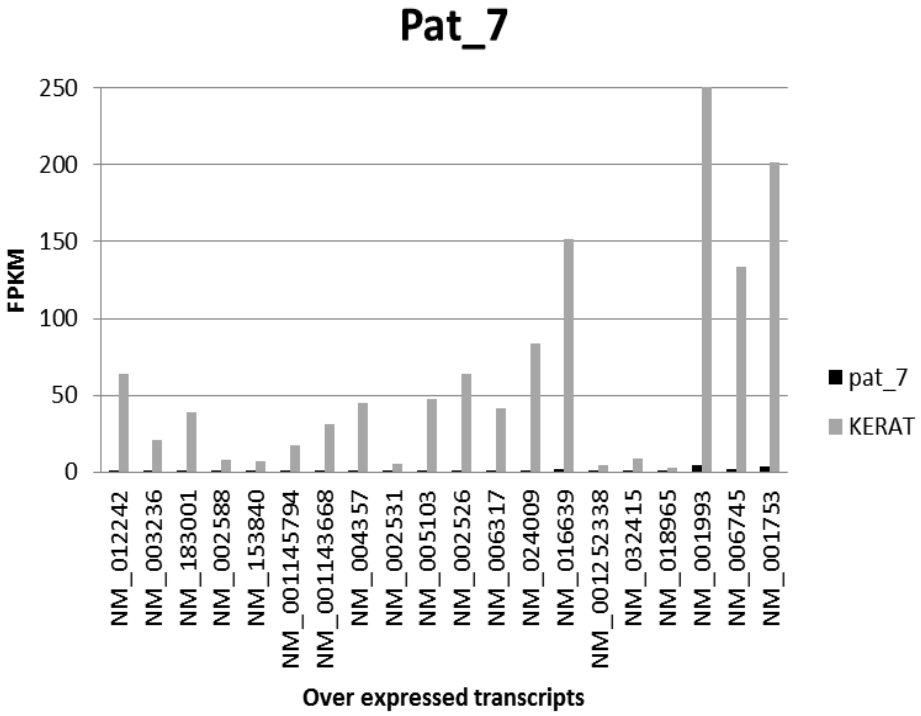
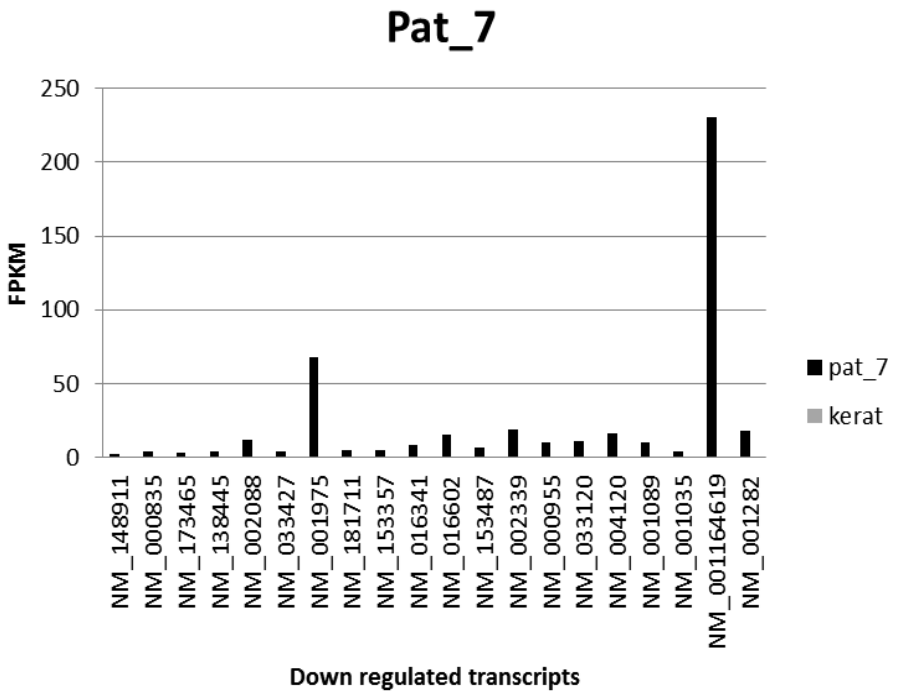


Figure 7-18 Top 20 and bottom 20 significantly deregulated transcripts in patient 7.





7.7 Appendix 7: Mass spectrometry results

Control (Un-transduced HaCaT)

1 -84.96.749.0/16711 108.6ENSP00000346634 gpmDB | psyt | snap [1039/9290] homo (1/1) protein THRAP3:p

2 -30.06.715.0/835 68.1ENSP00000254719 gpmDB | psyt | snap [2590/7319] protein RPA1:p,

3 -29.26.5019/2223 12.3ENSP00000451080 gpmDB | psyt | snap [1063/5488] homo (0/1) protein no protein information available

4 -23.46.5118/2535 16.1ENSP00000221975 gpmDB | psyt | snap [2045/8029] protein RPS19:p,

5 -22.15.832.0/411 76.1ENSP00000349748 gpmDB | psyt | snap protein SFPQ:p, splicing factor proline/glutamine-rich [Source: HGNC 10774] IPR012975 NOPS IPR000504 *(×6)* RRM dom

6 -18.46.289.4/1433 29.5ENSP00000326219 gpmDB | psyt | snap homo (2/17) protein DHRS4:p,

7 -15.75.757.3/1123 38.7ENSP00000341826 gpmDB | psyt | snap homo (0/7) protein HNRNPA1:p,

8 -15.75.952.4/72½3½ 78.3ENSP00000340329 gpmDB | psyt | snap homo (4/6) protein CA-PRIN1:p,

9 -10.86.065.9/1822 53.4ENSP00000254108 gpmDB | psyt | snap homo (4/7) protein FUS:p, fused in sarcoma [Source: HGNC 4010] IPR000504 *(×3)* RRM dom IPR001876 *(×3)* Znf RanBP2

10 -9.46.966.9/922 24.4sp|TRYP_PIG| gpmDB | psyt | snap protein Trypsin; EC 3.4.21.4; Flags: Precursor;

11 -9.25.932.6/423 112.3ENSP00000322542 gpmDB | psyt | snap homo (3/11) protein GTF2I:p,

12 -3.05.822.4/312 42.0ENSP00000355645 gpmDB | psyt | snap homo (32/32) protein ACTA1:p,

13 -2.85.314.0/611 37.4ENSP00000346694 gpmDB | psyt | snap homo (2/2) protein HNRNPA2B1:p, heterogeneous nuclear ribonucleoprotein A2/B1 [Source: HGNC 5033] IPR021662 HnRNPA IPR000504 *(×6)* RRM dom

14 -2.55.601.7/311 54.1ENSP00000254436 gpmDB | psyt | snap homo (1/1) protein TRIM21:p,

15 -2.46.087.4/1013 13.6ENSP00000223129 gpmDB | psyt | snap homo (3/3) protein RPA3:p,

16 -2.46.171.1/213 88.9ENSP00000393151 gpmDB | psyt | snap homo (3/3) protein HNRNPU:p,

17 -2.46.071.2/312 68.4ENSP00000381031 gpmDB | psyt | snap [6110/6343] homo (6/6) protein EWSR1:p,

18 -2.25.371.7/311 54.8ENSP00000372804 gpmDB | psyt | snap [2662/2787] homo (2/2) protein RNMT:p,

19 -2.25.636.4/711 25.2ENSP00000355354 gpmDB | psyt | snap [161/193] homo (1/1) protein SOX15:p,

20 -2.05.874.7/611 17.2ENSP00000265304 gpmDB | psyt | snap [7381/7450] homo (6/6) protein SSBP1:p,

21 -1.85.846.4/811 14.9ENSP00000420311 gpmDB | psyt | snap [8927/8973] homo (1/1) protein RPL23:p,

22 -1.86.201.0/111 119.4ENSP00000264107 gpmDB | psyt | snap [5714/5815] homo (6/6) protein ITGA6:p,

23 -1.65.991.7/311 54.2ENSP00000276079 gpmDB | psyt | snap [12742/12896] homo (6/6) protein NONO:p,

24 -1.25.170.4/111 358.5ENSP00000357643:reversed homo (1/1) protein no protein information available

25 -1.15.782.8/411 33.3ENSP00000336579:reversed homo (4/4) protein no protein information available$\mu_{½}$ = #8#8

CFP (AdZ/CFP empty vector)

1 -167.68.0930/481426 68.1ENSP00000254719 gpmDB | psyt | snap [315/7319] protein RPA1:p,

2 -117.57.8915/271022 90.5ENSP00000283179 gpmDB | psyt | snap [3034/20016] homo (1/3) protein HNRNPU:p, heterogeneous nuclear ribonucleoprotein U (scaffold attachment factor A)[Source: HGNC 5048]IPR001870 B30.2/SPRY IPR008985 ConA-like lec gl sf IPR027417 P-loop NTPase IPR003034 *(×3)* SAP dom IPR003877 SPRY rcpt IPR018355 SPla/RYanodine receptor subgr IPR010488 Zeta toxin domain

3 -105.97.569.8/17812 108.6ENSP00000346634 gpmDB | psyt | snap [759/9290] homo (1/1) protein THRAP3:p,

4 -62.77.8739/53711 16.1ENSP00000221975 gpmDB | psyt | snap [451/8029] protein RPS19:p

5 -61.97.559.6/2648 68.4ENSP00000381031 gpmDB | psyt | snap [289/6343] homo (4/8) protein EWSR1:p,

6 -55.37.8734/50611 17.7ENSP00000393241 gpmDB | psyt | snap [1258/10742] homo (4/11) protein RPS18:p,

7 -43.37.205.7/1034 76.1ENSP00000349748 gpmDB | psyt | snap [4971/14647] protein SFPQ:p, splicing factor proline/glutamine-rich [Source: HGNC 10774] IPR012975 NOPS IPR000504 *(×6)* RRM dom

8 -40.17.6311/3347 53.4ENSP00000254108 gpmDB | psyt | snap [2366/10275] homo (4/7) protein FUS:p, fused in sarcoma [Source: HGNC 4010] IPR000504 *(×3)* RRM dom IPR001876 *(×3)* Znf RanBP2

9 -30.95.929.8/1322 27.5ENSP00000421592 gpmDB | psyt | snap [1781/6514] homo (0/1) protein THOC4:p,

10 -30.86.4913/1922 18.6ENSP00000322887 gpmDB | psyt | snap [685/4385] homo (0/1) protein CIRBP:p,

11 -29.56.6312/1834½ 34.2ENSP00000448617 gpmDB | psyt | snap [7003/12937] homo (7/15) protein heterogeneous nuclear ribonucleoprotein A1 [Source: HGNC 5031]IPR021662 HnRNPA1 IPR000504 *(×6)* RRM dom

12 -25.97.2519/2224 12.3ENSP00000451080 gpmDB | psyt | snap [1303/5488] homo (0/1) protein no protein information available

13 -24.66.998.3/1226 22.9ENSP00000196551 gpmDB | psyt | snap [4252/12186] protein RPS5:p,

14 -24.66.745.1/83360.0ENSP00000252252 gpmDB | psyt | snap [15362/39375] homo (4/14) protein KRT6B:p,

15 -23.47.5415/193½5 24.4sp|TRYP_PIG| gpmDB | psyt | snap [58294/104275] protein Trypsin; EC 3.4.21.4; Flags: Precursor;

16 -22.86.691.4/211 62.3(H) ENSP00000252242 gpmDB | psyt | snap [12850/31245] homo (4/14) protein KRT5:p,

17 -22.76.933.4/534 112.3ENSP00000322542 gpmDB | psyt | snap [3594/8761] homo (4/5) protein GTF2I:p,

18 -20.87.2626/3634 13.6ENSP00000223129 gpmDB | psyt | snap [992/3996] homo (1/3) protein RPA3:p,

19 -17.96.8012/1824 29.2ENSP00000363021 gpmDB | psyt | snap [1285/3821] homo (3/3) protein RPA2:p,

20 -15.26.2315/3123 35.9ENSP00000361626 gpmDB | psyt | snap [7536/13276] homo (2/2) protein YBX1:p,

21 -14.86.967.2/2023 78.3ENSP00000340329 gpmDB | psyt | snap [3981/7872] homo (5/6) protein CAPRIN1:p,

22 -14.06.642.5/311 51.5ENSP00000167586 gpmDB | psyt | snap [15777/30111] homo (4/4) protein KRT14:p,

23 -12.66.541.7/323 140.9ENSP00000356520 gpmDB | psyt | snap [11263/15186] homo (0/1) protein DHX9:p,

24 -10.57.673.5/622 100.2ENSP00000376159 gpmDB | psyt | snap [4186/7031] homo (1/11) protein BCLAF1:p,

25 -10.37.025.3/724 29.6ENSP00000362744 gpmDB | psyt | snap homo (0/1) protein RPS4X:p,
26 -9.97.307.2/1124 29.5ENSP00000326219 gpmDB | psyt | snap homo (2/10) protein DHRS4:p,
27 -9.86.863.7/912 38.4ENSP00000313199 gpmDB | psyt | snap homo (9/9) protein HNRNPD:p,
28 -7.15.605.6/911 24.6ENSP00000358832 gpmDB | psyt | snap homo (4/4) protein RPL10:p,
29 -4.96.2016/3413 7.8ENSP00000397872 gpmDB | psyt | snap protein RPS28:p,
30 -4.96.976.4/712 25.2ENSP00000355354 gpmDB | psyt | snap homo (1/1) protein SOX15:p,
31 -4.86.170.7/111 157.8ENSP00000260665 gpmDB | psyt | snap homo (1/1) protein LRPPRC:p,
32 -4.45.244.4/711 42.4ENSP00000360033 gpmDB | psyt | snap homo (3/3) protein SERBP1:p, SERPINE1 mRNA binding protein 1 [Source: HGNC 17860]IPR006861 HABP4 PAIRBP1-bd
33 -4.36.152.2/312 59.6ENSP00000349428 gpmDB | psyt | snap homo (3/3) protein PTBP1:p, polypyrimidine tract binding protein 1 [Source: HGNC 9583] IPR006536 HnRNP-L PTB IPR000504 *(×11)* RRM dom
34 -3.36.692.5/411 37.4ENSP00000346694 gpmDB | psyt | snap homo (3/3) protein HNRNPA2B1:p, heterogeneous nuclear ribonucleoprotein A2/B1 [Source: HGNC 5033]IPR021662 HnRNPA1 IPR000504 *(×6)* RRM dom
35 -2.75.811.0/211 51.3ENSP00000357980:reversed protein no protein information available
36 -2.76.524.9/612 26.7ENSP00000434643 gpmDB | psyt | snap homo (7/7) protein RPS3:p,
37 -2.66.231.2/21193.6ENSP00000295770:reversed protein
no protein information available
38 -2.56.264.9/811 21.4ENSP00000397798 gpmDB | psyt | snap homo (1/1) protein RPL17:p,
39 -2.56.595.4/611 16.8ENSP00000406181 gpmDB | psyt | snap protein SEPT2:p,
40 -2.55.881.1/211 87.0ENSP00000331678 gpmDB | psyt | snap protein PKP3:p,
41 -2.36.116.2/812 16.4ENSP00000251453 gpmDB | psyt | snap homo (1/1) protein RPS16:p,
42 -2.27.3023/100+11 9.2ENSP00000421635 gpmDB | psyt | snap protein C5orf60:p,
43 -2.26.654.7/612 17.2ENSP00000265304 gpmDB | psyt | snap homo (6/6) protein SSBP1:p,
44 -2.15.8617/3811 10.7ENSP00000370284 gpmDB | psyt | snap protein ZNF561:p,
45 -2.16.991.1/211 157.7ENSP00000362776 gpmDB | psyt | snap homo (3/3) protein COL16A1:p, collagen, type XVI, alpha 1 [Source: HGNC 2193] IPR008160*(×12)*Collagen IPR008985 ConA-like lec gl sf IPR001791 Laminin G 46 -1.86.307.0/1211 13.0ENSP00000348849 gpmDB | psyt | snap homo (2/2) protein RPS26:p,

47 -1.86.135.3/811 27.7ENSP00000238081 gpmDB | psyt | snap homo (2/2) protein YWHAQ:p, tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, theta [Source: HGNC 12854]IPR000308 *(×7)* 14-3-3 IPR023410 *(×3)* 14-3-3 domain

48 -1.76.702.2/311 61.8ENSP00000374461 gpmDB | psyt | snap homo (1/1) protein ACOXL:p,

49 -1.76.651.5/311 50.5ENSP00000446271 gpmDB | psyt | snap homo (2/2) protein no protein information available

50 -1.65.431.3/211 70.6ENSP00000313007 gpmDB | psyt | snap homo (9/9) protein PABPC1:p,

51 -1.55.521.8/311 157.9ENSP00000351875 gpmDB | psyt | snap homo (2/2) protein ANKRD30B:p,

52 -1.56.516.0/1011 26.5ENSP00000418634:reversed homo (4/4) protein no protein information available

53 -1.45.393.6/611 22.6ENSP00000302896 gpmDB | psyt | snap homo (6/6) protein RPS9:p,

54 -1.46.320.7/111 301.6ENSP00000369365:reversed homo (2/2) protein no protein information available

55 -1.45.802.0/311 83.2ENSP00000262843 gpmDB | psyt | snap homo (1/1) protein MID2:p,

56 -1.46.302.5/411 35.9ENSP00000374987 gpmDB|psyt| snap protein IGHG2:p,

57 -1.46.482.9/511 54.7ENSP00000355437:reversed protein no protein information available

58 -1.35.354.8/611 18.9ENSP00000347271 gpmDB | psyt | snap homo (1/1) protein RPS10:p,

59 -1.36.231.8/211 50.1ENSP00000264193 gpmDB | psyt | snap protein CPOX:p,

60 -1.36.431.7/311 54.2ENSP00000276079 gpmDB | psyt | snap homo (6/6) protein NONO:p,

61 -1.34.809.1/1011 32.0ENSP00000392152:reversed homo (1/1) protein no protein information available

62 -1.36.450.7/111 118.0ENSP00000386727 gpmDB| psyt | snap homo (1/1) protein RBM44:p,

63 -1.36.831.3/211 138.0ENSP00000364798 gpmDB| psyt | snap homo (3/3) protein ZBTB40:p,

64 -1.36.951.3/211 92.0ENSP00000345498:reversed homo (6/6) protein no protein information available

65 -1.26.083.6/911 51.8ENSP00000453793 gpmDB| psyt | snap homo (3/3) protein no protein information available

66 -1.27.551.4/211 73.0ENSP00000344233:reversed homo (2/2) protein no protein information available

67 -1.16.571.2/211 47.1ENSP00000393074 gpmDB | psyt | snap homo (2/2) protein ZNF561:p,

68 -1.16.394.9/611 27.8ENSP00000389345 gpmDB | psyt |snap homo (17/17) protein BDNF:p,

69 -1.15.992.6/611 58.2ENSP00000445929:reversed protein no protein information available

70 1.14.881.5/211 236.5ENSP00000376306 gpmDB | psyt | snap homo (1/1) protein CIT:p,

71 -1.17.672.3/411 79.2ENSP00000364867 gpmDB | psyt | snap homo (1/1) protein ROR2:p,

72 -1.16.297.3/1111 15.7ENSP00000342787 gpmDB | psyt | snap homo (9/9) protein RPL28:p,

$\mu_{½}$ = #15#11

E6 (AdZ/CFP/E6)

1 -126.67.4115/261218 108.6ENSP00000346634 gpmDB | psyt | snap[574/9290] homo (1/1) protein THRAP3:p,

2 -104.27.6218/3092068.1ENSP00000254719 gpmDB | psyt | snap [641/7319] protein RPA1:p,

3 -83.87.0713/24713 90.5ENSP00000283179 gpmDB | psyt | snap[4485/20016] homo (0/3) protein HNRNPU:p, heterogeneous nuclear ribonucleoprotein U (scaffold attachment factor A) [Source: HGNC 5048]IPR001870 B30.2/SPRY IPR008985 ConA-like lec gl sf IPR027417 P-loop NTPase IPR003034 *(×3)* SAP dom IPR003877 SPRY rcpt IPR018355 SPla/RYanodine receptor subgr IPR010488 Zeta toxin domain

4 -50.97.139.6/2647 68.4ENSP00000381031 gpmDB | psyt | snap [460/6343] homo (4/6) protein EWSR1:p,

5 -48.66.466.2/1144 76.1ENSP00000349748 gpmDB |psyt | snap [4514/14647] homo (0/1) protein SFPQ:p, splicing factor prline/glutaminerich [Source: HGNC 10774] IPR012975 NOPS IPR000504 *(×6)* RRM dom

6 -40.16.765.5/858 112.3ENSP00000322542 gpmDB | psyt | snap [2454/8761] homo (4/4) protein GTF2I:p,

7 -36.87.2510/31412 53.4ENSP00000254108 gpmDB | psyt | snap [2604/10275] homo (4/7) protein FUS:p, fused in sarcoma [Source: HGNC 4010] IPR000504 *(×3)* RRM dom IPR001876 *(×3)* Znf RanBP2

8 -31.07.0718/2536 16.1ENSP00000221975 gpmDB | psyt | snap [1455/8029] protein RPS19:p,

9 -27.96.264.0/522 51.5ENSP00000167586 gpmDB | psyt | snap [11151/30111] homo (1/14) protein KRT14:p,

10 -27.47.2919/2223$_{½}$ 12.3ENSP00000451080 gpmDB | psyt | snap [1177/5488] homo (0/1) protein no protein information available

11 -26.26.9614/2337 29.2ENSP00000363021 gpmDB |psyt | snap [905/3821] homo (3/3) protein RPA2:p,

12 -26.16.273.3/423 65.4ENSP00000310861 gpmDB | psyt | snap [21457/42389] homo (1/11) protein KRT2:p,

13 -25.96.504.7/133½4 78.3ENSP00000340329 gpmDB | psyt | snap [2546/7872] homo (4/6) protein CAPRIN1:p,

14 -20.66.598.7/1235 29.6ENSP00000362744 gpmDB | psyt | snap[6435/13307] homo (0/3) protein RPS4X:p,

15 -20.06.595.3/833 54.1ENSP00000254436 gpmDB | psyt | snap [663/2205] homo (1/1) protein TRIM21:p,

16 -19.16.7015/2334 17.7ENSP00000393241 gpmDB | psyt | snap [4535/10742] homo (4/11) protein RPS18:p,

17 -17.86.889.4/1435 29.5ENSP00000326219 gpmDB | psyt | snap [346/1642] homo (2/16) protein DHRS4:p,

18 -17.86.221.7/233 157.8ENSP00000260665 gpmDB | psyt | snap [8929/13493] homo (1/4) protein LRPPRC:p,

19 -16.75.961.8/211 48.7(H) sp|K1C15_SHEEP| gpmDB | psyt | snap [17216/43948] homo (3/14) protein Keratin, type I cytoskeletal 15; Cytokeratin-15; CK-15; Keratin-15; K15;

20 15.96.443.9/623 59.6ENSP00000349428 gpmDB | psyt | snap [9083/14463] homo (3/3) protein PTBP1:p, polypyrimidine tract binding protein 1 [Source: HGNC 9583]IPR006536 HnRNP-L PTB IPR000504 *(×11)* RRM dom

21 -15.05.9221/4522 35.9ENSP00000361626 gpmDB | psyt | snap [7573/13276] homo (0/2) protein YBX1:p,

22 -13.65.797.8/1111 22.9ENSP00000196551 gpmDB | psyt | snap [6315/12186] protein RPS5:p,

23 -13.36.416.5/924 37.4ENSP00000346694 gpmDB | psyt | snap [14284/20298] homo (2/3) protein HNRNPA2B1:p, heterogeneous nuclear ribonucleoprotein A2/B1 [Source: HGNC 5033]IPR021662 HnRNPA1 IPR000504 *(×6)* RRM dom

24 -13.06.5919/2124 25.2ENSP00000355354 gpmDB | psyt | snap [41/193] homo (1/1) protein SOX15:p,

25 -11.66.032.4/422 121.4ENSP00000386786 gpmDB | psyt | snap [27250/32039] homo (8/34) protein POTEF:p,

26 -10.56.5313/1823 13.6ENSP00000223129 gpmDB | psyt | snap [1993/3996] homo (1/3) protein RPA3:p,

27 -10.06.254.0/722 54.2ENSP00000276079 gpmDB | psyt | snap [8684/12896] homo (4/7) protein NONO:p,

28 -9.27.126.9/922 24.4sp|TRYP_PIG| gpmDB | psyt | snap [80976/104275] protein Trypsin; EC 3.4.21.4; Flags: Precursor;

29 -7.66.235.8/812 18.6ENSP00000322887 gpmDB | psyt | snap [2662/4385] protein CIRBP:p,

30 -4.86.303.7/612 29.1ENSP00000420588 gpmDB | psyt | snap [2113/3143] homo (1/1) protein TFAM:p,

31 -2.96.543.9/611 28.0ENSP00000378378 gpmDB | psyt | snap [8108/9045] homo (6/6) protein RPL8:p,

32 -2.76.406.4/812 14.9ENSP00000420311 gpmDB | psyt | snap [8513/8973] homo (1/1) protein RPL23:p,

33 -2.55.744.5/911 20.2ENSP00000363676 gpmDB | psyt | snap [9714/10188] protein RPL11:p,

34 2.46.1810/111212.3ENSP00000450909 gpmDB | psyt | snap [3461/3593] homo (1/1) protein no protein information available

35 -2.05.957.0/121113.0ENSP00000348849 gpmDB | psyt | snap [7092/7298] homo (2/2) protein RPS26:p,

36 -2.05.934.7/611 17.2ENSP00000265304 gpmDB | psyt | snap [7381/7450] homo (6/6) protein SSBP1:p,

37 -1.96.850.4/112 245.3ENSP00000407602:reversed protein no protein information available

38 -1.86.312.4/411 96.7ENSP00000017003 gpmDB | psyt |snap [312/350] homo (2/2) protein XYLT2:p,

39 -1.65.801.4/211 62.3ENSP00000252242 gpmDB | psyt | snap[30399/31245] homo (1/1) protein KRT5:p,

40 -1.66.410.7/112227.4ENSP00000267622:reversed homo 2/2 protein; no protein information available

41 -1.55.870.7/111274.2ENSP00000349437:reversed protein no protein information available

42 -1.56.561.4/212 73.7ENSP00000360541: reversed protein no protein information available

43 -1.45.820.4/111,148.3ENSP00000364526:reversed protein
no protein information available

44 -1.46.193.4/511 31.3ENSP00000341885 gpmDB| psyt | snap [12694/12868] homo (4/4) protein RPS2:p,

45 -1.44.792.6/411 65.7ENSP00000382349:reversed protein no protein information available

46 -1.46.800.3/111509.8ENSP00000349892 gpmDB | psyt | snap [3310/3514] homo (2/2) protein MYCBP2:p,

47 -1.25.871.3/211 138.0ENSP00000364798 gpmDB | psyt | snap [434/449] homo (3/3) protein ZBTB40:p,

48 -1.26.621.1/111 116.1ENSP00000182096:reversed homo (2/2) protein no protein information available

49 -1.26.9833/4411 2.2ENSP00000451870 gpmDB | psyt | snap [660/715] protein T cell receptor alpha joining 56 [Source: HGNC 12088]

50 -1.25.951.2/21187.3ENSP00000272313:reversed protein no protein information available

51 -1.15.613.3/411 26.7ENSP00000434643 gpmDB | psyt | snap[11910/11917] homo (15/15) protein RPS3:p, $\mu_{½}$ =

Chapter 8. BIBLIOGRAPHY

Aasen T, Graham SV, Edward M, Hodgins MB (2005) Reduced expression of multiple gap junction proteins is a feature of cervical dysplasia. *Molecular cancer* **4:** 31

Abbas AK, Litchman, A.H.H. (2010) Basic Immunology: function and disorders of the immune system. *Sauders Elseiver Inc*

Adams B, Dorfler P, Aguzzi A, Kozmik Z, Urbanek P, Maurer-Fogy I, Busslinger M (1992) Pax-5 encodes the transcription factor BSAP and is expressed in B lymphocytes, the developing CNS, and adult testis. *Genes & development* **6:** 1589-1607

Afshar R, Pierce S, Bolland DJ, Corcoran A, Oltz EM (2006) Regulation of IgH gene assembly: role of the intronic enhancer and 5'DQ52 region in targeting DHJH recombination. *Journal of immunology* **176:** 2439-2447

Ahmad I (1995) Mash-1 is expressed during ROD photoreceptor differentiation and binds an E-box, E(opsin)-1 in the rat opsin gene. *Brain Res Dev Brain Res* **90:** 184-189

Alberts B, Johnson A, Lewis J, Raff M (2002) Molecular Biology of the Cell.

Alevizos I, Mahadevappa M, Zhang X, Ohyama H, Kohno Y, Posner M, Gallagher GT, Varvares M, Cohen D, Kim D, Kent R, Donoff RB, Todd R, Yung CM, Warrington JA, Wong DT (2001) Oral cancer in vivo gene expression profiling assisted by laser capture microdissection and microarray analysis. *Oncogene* **20:** 6196-6204

Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson J, Jr., Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Levy R, Wilson W, Grever MR, Byrd JC, Botstein D, Brown PO, Staudt LM (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403:** 503-511

Alizadeh AA, Staudt LM (2000) Genomic-scale gene expression profiling of normal and malignant immune cells. *Curr Opin Immunol* **12:** 219-225

Amadei B, Urbani S, Cazaly A, Fisicaro P, Zerbini A, Ahmed P, Missale G, Ferrari C, Khakoo SI (2010) Activation of natural killer cells during acute infection with hepatitis C virus. *Gastroenterology* **138:** 1536-1545

Anguita E, Hughes J, Heyworth C, Blobel GA, Wood WG, Higgs DR (2004) Globin gene activation during haemopoiesis is driven by protein complexes nucleated by GATA-1 and GATA-2. *EMBO J* **23:** 2841-2852

Arbyn M, Castellsague X, de Sanjose S, Bruni L, Saraiya M, Bray F, Ferlay J (2011) Worldwide burden of cervical cancer in 2008. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO* **22:** 2675-2686

Aronheim A, Shiran R, Rosen A, Walker MD (1993) The E2A gene product contains two separable and functionally distinct transcription activation domains. *Proc Natl Acad Sci U S A* **90:** 8063-8067

Ashrafi GH, Haghshenas MR, Marchetti B, O'Brien PM, Campo MS (2005) E5 protein of human papillomavirus type 16 selectively downregulates surface HLA class I. *International journal of cancer Journal international du cancer* **113:** 276-283

Assemat E, Bazellieres E, Pallesi-Pocachard E, Le Bivic A, Massey-Harroche D (2008) Polarity complex proteins. *Biochimica et biophysica acta* **1778:** 614-630

Au Yeung CL, Tsang WP, Tsang TY, Co NN, Yau PL, Kwok TT (2010) HPV-16 E6 upregulation of DNMT1 through repression of tumor suppressor p53. *Oncol Rep* **24:** 1599-1604

Bain G, Gruenwald S, Murre C (1993) E2A and E2-2 are subunits of B-cell-specific E2-box DNA-binding proteins. *Mol Cell Biol* **13:** 3522-3529

Bain G, Maandag EC, Izon DJ, Amsen D, Kruisbeek AM, Weintraub BC, Krop I, Schlissel MS, Feeney AJ, van Roon M, et al. (1994) E2A proteins are required for proper B cell development and initiation of immunoglobulin gene rearrangements. *Cell* **79:** 885-892

Bain G, Robanus Maandag EC, te Riele HP, Feeney AJ, Sheehy A, Schlissel M, Shinton SA, Hardy RR, Murre C (1997) Both E12 and E47 allow commitment to the B cell lineage. *Immunity* **6:** 145-154

Baker TS, Newcomb WW, Olson NH, Cowsert LM, Olson C, Brown JC (1991) Structures of bovine and human papillomaviruses. Analysis by cryoelectron microscopy and three-dimensional image reconstruction. *Biophysical journal* **60:** 1445-1456

Bakirtzis G, Choudhry R, Aasen T, Shore L, Brown K, Bryson S, Forrow S, Tetley L, Finbow M, Greenhalgh D, Hodgins M (2003) Targeted epidermal expression of mutant Connexin 26(D66H) mimics true Vohwinkel syndrome and provides a model

for the pathogenesis of dominant connexin disorders. *Human molecular genetics* **12:** 1737-1744

Banerjee G, Damodaran A, Devi N, Dharmalingam K, Raman G (2004) Role of keratinocytes in antigen presentation and polarization of human T lymphocytes. *Scandinavian journal of immunology* **59:** 385-394

Banerji J, Olson L, Schaffner W (1983) A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. *Cell* **33:** 729-740

Banerji J, Rusconi S, Schaffner W (1981) Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27:** 299-308

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

Bayani J, Brenton JD, Macgregor PF, Beheshti B, Albert M, Nallainathan D, Karaskova J, Rosen B, Murphy J, Laframboise S, Zanke B, Squire JA (2002) Parallel analysis of sporadic primary ovarian carcinomas by spectral karyotyping, comparative genomic hybridization, and expression microarrays. *Cancer research* **62:** 3466-3476

Beck K, Peak MM, Ota T, Nemazee D, Murre C (2009) Distinct roles for E12 and E47 in B cell specification and the sequential rearrangement of immunoglobulin light chain loci. *J Exp Med* **206:** 2271-2284

Bedell MA, Hudson JB, Golub TR, Turyk ME, Hosken M, Wilbanks GD, Laimins LA (1991) Amplification of human papillomavirus genomes in vitro is dependent on epithelial differentiation. *Journal of virology* **65:** 2254-2260

Benezra R, Davis RL, Lockshon D, Turner DL, Weintraub H (1990) The protein Id: a negative regulator of helix-loop-helix DNA binding proteins. *Cell* **61:** 49-59

Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I (2001) Controlling the false discovery rate in behavior genetics research. *Behav Brain Res* **125:** 279-284

Berger SL (2002) Histone modifications in transcriptional regulation. *Curr Opin Genet Dev* **12:** 142-148

Berger SL (2007) The complex language of chromatin regulation during transcription. *Nature* **447:** 407-412

Bernard HU (2002) Gene expression of genital human papillomaviruses and considerations on potential antiviral approaches. *Antiviral therapy* **7:** 219-237

Bernard HU (2013) Taxonomy and phylogeny of papillomaviruses: an overview and recent developments. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* **18:** 357-361

Bernard HU, Burk RD, Chen Z, van Doorslaer K, Hausen H, de Villiers EM (2010) Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments. *Virology* **401:** 70-79

Bernard O, Hozumi N, Tonegawa S (1978) Sequences of mouse immunoglobulin light chain genes before and after somatic changes. *Cell* **15:** 1133-1144

Bernfield M, Kokenyesi R, Kato M, Hinkes MT, Spring J, Gallo RL, Lose EJ (1992) Biology of the syndecans: a family of transmembrane heparan sulfate proteoglycans. *Annual review of cell biology* **8:** 365-393

Betz AG, Milstein C, Gonzalez-Fernandez A, Pannell R, Larson T, Neuberger MS (1994) Elements regulating somatic hypermutation of an immunoglobulin kappa gene: critical role for the intron enhancer/matrix attachment region. *Cell* **77:** 239-248

Bevington SL (2009) The Mechanism of Enhancer Mediated Long-range Chromatin Activation During V(D)J Recombination. *PhD thesis, University of Leeds,*

Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE, Kuehn MS, Taylor CM, Neph S, Koch CM, Asthana S, Malhotra A, Adzhubei I, Greenbaum JA, Andrews RM, Flicek P, Boyle PJ, Cao H, Carter NP, Clelland GK, Davis S, Day N, Dhami P, Dillon SC, Dorschner MO, Fiegler H, Giresi PG, Goldy J, Hawrylycz M, Haydock A, Humbert R, James KD, Johnson BE, Johnson EM, Frum TT, Rosenzweig ER, Karnani N, Lee K, Lefebvre GC, Navas PA, Neri F, Parker SC, Sabo PJ, Sandstrom R, Shafer A, Vetrie D, Weaver M, Wilcox S, Yu M, Collins FS, Dekker J, Lieb JD, Tullius TD, Crawford GE, Sunyaev S, Noble WS, Dunham I, Denoeud F, Reymond A, Kapranov P, Rozowsky J, Zheng D, Castelo R, Frankish A, Harrow J, Ghosh S, Sandelin A, Hofacker IL, Baertsch R, Keefe D, Dike S, Cheng J, Hirsch HA, Sekinger EA, Lagarde J, Abril JF, Shahab A, Flamm C, Fried C, Hackermuller J, Hertel J, Lindemeyer M, Missal K, Tanzer A, Washietl S, Korbel J, Emanuelsson O, Pedersen JS, Holroyd N, Taylor R, Swarbreck D, Matthews N, Dickson MC, Thomas DJ, Weirauch MT, Gilbert J, Drenkow J, Bell I, Zhao X, Srinivasan KG, Sung WK, Ooi HS, Chiu KP, Foissac S, Alioto T, Brent M, Pachter L, Tress ML, Valencia A, Choo SW, Choo CY, Ucla C, Manzano C, Wyss C, Cheung E, Clark TG, Brown JB, Ganesh M, Patel S, Tammana H, Chrast J, Henrichsen

CN, Kai C, Kawai J, Nagalakshmi U, Wu J, Lian Z, Lian J, Newburger P, Zhang X, Bickel P, Mattick JS, Carninci P, Hayashizaki Y, Weissman S, Hubbard T, Myers RM, Rogers J, Stadler PF, Lowe TM, Wei CL, Ruan Y, Struhl K, Gerstein M, Antonarakis SE, Fu Y, Green ED, Karaoz U, Siepel A, Taylor J, Liefer LA, Wetterstrand KA, Good PJ, Feingold EA, Guyer MS, Cooper GM, Asimenos G, Dewey CN, Hou M, Nikolaev S, Montoya-Burgos JI, Loytynoja A, Whelan S, Pardi F, Massingham T, Huang H, Zhang NR, Holmes I, Mullikin JC, Ureta-Vidal A, Paten B, Seringhaus M, Church D, Rosenbloom K, Kent WJ, Stone EA, Batzoglou S, Goldman N, Hardison RC, Haussler D, Miller W, Sidow A, Trinklein ND, Zhang ZD, Barrera L, Stuart R, King DC, Ameur A, Enroth S, Bieda MC, Kim J, Bhinge AA, Jiang N, Liu J, Yao F, Vega VB, Lee CW, Ng P, Yang A, Moqtaderi Z, Zhu Z, Xu X, Squazzo S, Oberley MJ, Inman D, Singer MA, Richmond TA, Munn KJ, Rada-Iglesias A, Wallerman O, Komorowski J, Fowler JC, Couttet P, Bruce AW, Dovey OM, Ellis PD, Langford CF, Nix DA, Euskirchen G, Hartman S, Urban AE, Kraus P, Van Calcar S, Heintzman N, Kim TH, Wang K, Qu C, Hon G, Luna R, Glass CK, Rosenfeld MG, Aldred SF, Cooper SJ, Halees A, Lin JM, Shulha HP, Xu M, Haidar JN, Yu Y, Iyer VR, Green RD, Wadelius C, Farnham PJ, Ren B, Harte RA, Hinrichs AS, Trumbower H, Clawson H, Hillman-Jackson J, Zweig AS, Smith K, Thakkapallayil A, Barber G, Kuhn RM, Karolchik D, Armengol L, Bird CP, de Bakker PI, Kern AD, Lopez-Bigas N, Martin JD, Stranger BE, Woodroffe A, Davydov E, Dimas A, Eyras E, Hallgrimsdottir IB, Huppert J, Zody MC, Abecasis GR, Estivill X, Bouffard GG, Guan X, Hansen NF, Idol JR, Maduro VV, Maskeri B, McDowell JC, Park M, Thomas PJ, Young AC, Blakesley RW, Muzny DM, Sodergren E, Wheeler DA, Worley KC, Jiang H, Weinstock GM, Gibbs RA, Graves T, Fulton R, Mardis ER, Wilson RK, Clamp M, Cuff J, Gnerre S, Jaffe DB, Chang JL, Lindblad-Toh K, Lander ES, Koriabine M, Nefedov M, Osoegawa K, Yoshinaga Y, Zhu B, de Jong PJ (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447:** 799-816

Bishop B, Dasgupta J, Klein M, Garcea RL, Christensen ND, Zhao R, Chen XS (2007) Crystal structures of four types of human papillomavirus L1 capsid proteins: understanding the specificity of neutralizing monoclonal antibodies. *The Journal of biological chemistry* **282:** 31803-31811

Blomberg B, Traunecker A, Eisen H, Tonegawa S (1981) Organization of four mouse lambda light chain immunoglobulin genes. *Proceedings of the National Academy of Sciences of the United States of America* **78:** 3765-3769

Bordeaux J, Forte S, Harding E, Darshan MS, Klucevsek K, Moroianu J (2006) The l2 minor capsid protein of low-risk human papillomavirus type 11 interacts with host nuclear import receptors and viral DNA. *Journal of virology* **80:** 8259-8262

Bosco D, Haefliger JA, Meda P (2011) Connexins: key mediators of endocrine function. *Physiol Rev* **91:** 1393-1445

Bottley G, Watherston OG, Hiew YL, Norrild B, Cook GP, Blair GE (2008) High-risk human papillomavirus E7 expression reduces cell-surface MHC class I molecules and increases susceptibility to natural killer cells. *Oncogene* **27:** 1794-1799

Boukamp P, Petrussevska RT, Breitkreutz D, Hornung J, Markham A, Fusenig NE (1988) Normal keratinization in a spontaneously immortalized aneuploid human keratinocyte cell line. *The Journal of cell biology* **106:** 761-771

Boyer SN, Wazer DE, Band V (1996) E7 protein of human papilloma virus-16 induces degradation of retinoblastoma protein through the ubiquitin-proteasome pathway. *Cancer research* **56:** 4620-4624

Bradney C, Hjelmeland M, Komatsu Y, Yoshida M, Yao TP, Zhuang Y (2003) Regulation of E2A activities by histone acetyltransferases in B lymphocyte development. *The Journal of biological chemistry* **278:** 2370-2376

Brass AL, Kehrli E, Eisenbeis CF, Storb U, Singh H (1996) Pip, a lymphoid-restricted IRF, contains a regulatory domain that is important for autoinhibition and ternary complex formation with the Ets factor PU.1. *Genes & development* **10:** 2335-2347

Brass AL, Zhu AQ, Singh H (1999) Assembly requirements of PU.1-Pip (IRF-4) activator complexes: inhibiting function in vivo using fused dimers. *EMBO J, 18(4): p 977-91*

Brasset E, Vaury C (2005) Insulators are fundamental components of the eukaryotic genomes. *Heredity* **94:** 571-576

Breathnach R, Chambon P (1981) Organization and expression of eucaryotic split genes coding for proteins. *Annual review of biochemistry* **50:** 349-383

Buck CB, Pastrana DV, Lowy DR, Schiller JT (2004) Efficient intracellular assembly of papillomaviral vectors. *Journal of virology* **78:** 751-757

Buck CB, Trus BL (2012) The papillomavirus virion: a machine built to hide molecular Achilles' heels. *Advances in experimental medicine and biology* **726:** 403-422

Bulger M, Groudine M (1999) Looping versus linking: toward a model for long-distance gene activation. *Genes Dev* **13:** 2465-2477

Bulger M, Groudine M (2011) Functional and mechanistic diversity of distal transcription enhancers. *Cell* **144:** 327-339

Bungert J, Tanimoto K, Patel S, Liu Q, Fear M, Engel JD (1999) Hypersensitive site 2 specifies a unique function within the human beta-globin locus control region to stimulate globin gene transcription. *Mol Cell Biol* **19:** 3062-3072

Burd EM (2003) Human papillomavirus and cervical cancer. *Clinical microbiology reviews* **16:** 1-17

Burgers WA, Blanchon L, Pradhan S, de Launoit Y, Kouzarides T, Fuks F (2007) Viral oncoproteins target the DNA methyltransferases. *Oncogene* **26:** 1650-1655

Buskin JN, Hauschka SD (1989) Identification of a myocyte nuclear factor that binds to the muscle-specific enhancer of the mouse muscle creatine kinase gene. *Mol Cell Biol* **9:** 2627-2640

Busslinger M, Urbanek P (1995) The role of BSAP (Pax-5) in B-cell development. *Current opinion in genetics & development* **5:** 595-601

Cambier JC, Gauld SB, Merrell KT, Vilen BJ (2007) B-cell anergy: from transgenic models to naturally occurring anergic B cells? *Nat Rev Immunol* **7:** 633-643

Carson S, Wu GE (1989) A linkage map of the mouse immunoglobulin lambda light chain locus. *Immunogenetics* **29:** 173-179

Carter JJ, Wipf GC, Madeleine MM, Schwartz SM, Koutsky LA, Galloway DA (2006) Identification of human papillomavirus type 16 L1 surface loops required for neutralization by human sera. *Journal of virology* **80:** 4664-4672

Chao J, Rothschild G, Basu U (2014) Ubiquitination Events That Regulate Recombination of Immunoglobulin Loci Gene Segments. *Frontiers in immunology* **5:** 100

Chaudhary J, Cupp AS, Skinner MK (1997) Role of basic-helix-loop-helix transcription factors in Sertoli cell differentiation: identification of an E-box response element in the transferrin promoter. *Endocrinology* **138:** 667-675

Chen SL, Huang CH, Tsai TC, Lu KY, Tsao YP (1996) The regulation mechanism of c-jun and junB by human papillomavirus type 16 E5 oncoprotein. *Archives of virology* **141:** 791-800

Chen XS, Garcea RL, Goldberg I, Casini G, Harrison SC (2000) Structure of small virus-like particles assembled from the L1 protein of human papillomavirus 16. *Molecular cell* **5:** 557-567

Chen Y, Huhn D, Knosel T, Pacyna-Gengelbach M, Deutschmann N, Petersen I (2005) Downregulation of connexin 26 in human lung cancer is related to promoter methylation. *International journal of cancer Journal international du cancer* **113:** 14-21

Chiang SC, Theorell J, Entesarian M, Meeths M, Mastafa M, Al-Herz W, Frisk P, Gilmour KC, Ifversen M, Langenskiold C, Machaczka M, Naqvi A, Payne J, Perez-Martinez A, Sabel M, Unal E, Unal S, Winiarski J, Nordenskjold M, Ljunggren HG, Henter JI, Bryceson YT (2013) Comparison of primary human cytotoxic T-cell and natural killer cell responses reveal similar molecular requirements for lytic granule exocytosis but differences in cytokine production. *Blood* **121:** 1345-1356

Chou HL, Yao CT, Su SL, Lee CY, Hu KY, Terng HJ, Shih YW, Chang YT, Lu YF, Chang CW, Wahlqvist ML, Wetter T, Chu CM (2013) Gene expression profiling of breast cancer survivability by pooled cDNA microarray analysis using logistic regression, artificial neural networks and decision trees. *BMC Bioinformatics* **14:** 100

Chow LT, Reilly SS, Broker TR, Taichman LB (1987) Identification and mapping of human papillomavirus type 1 RNA transcripts recovered from plantar warts and infected epithelial cell cultures. *Journal of virology* **61:** 1913-1918

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:** 229-241

Chowdhury D, Sen R (2004) Mechanisms for feedback inhibition of the immunoglobulin heavy chain locus. *Curr Opin Immunol* **16:** 235-240

Clapier CR, Cairns BR (2009) The biology of chromatin remodeling complexes. *Annu Rev Biochem* **78:** 273-304

Clark EA, Golub TR, Lander ES, Hynes RO (2000) Genomic analysis of metastasis reveals an essential role for RhoC. *Nature* **406:** 532-535

Cobo F (2012) Human papillomavirus infections: from the laboratory to clinical practice.

Coleclough C (1983) Chance, necessity and antibody gene dynamics. *Nature* **303:** 23-26

Combita AL, Touze A, Bousarghin L, Sizaret PY, Munoz N, Coursaget P (2001) Gene transfer using human papillomavirus pseudovirions varies according to virus

genotype and requires cell surface heparan sulfate. *FEMS microbiology letters* **204:** 183-188

Conway MJ, Meyers C (2009) Replication and assembly of human papillomaviruses. *J Dent Res* **88:** 307-317

Cory S, Jackson J, Adams JM (1980) Deletions in the constant region locus can account for switches in immunoglobulin heavy chain expression. *Nature* **285:** 450-456

Crawford GE, Davis S, Scacheri PC, Renaud G, Halawi MJ, Erdos MR, Green R, Meltzer PS, Wolfsberg TG, Collins FS (2006) DNase-chip: a high-resolution method to identify DNase I hypersensitive sites using tiled microarrays. *Nat Methods* **3:** 503-509

Crews ST (1998) Control of cell lineage-specific development and transcription by bHLH-PAS proteins. *Genes & development* **12:** 607-620

Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, Boyer LA, Young RA, Jaenisch R (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* **107:** 21931-21936

Crusius K, Auvinen E, Alonso A (1997) Enhancement of EGF- and PMA-mediated MAP kinase activation in cells expressing the human papillomavirus type 16 E5 protein. *Oncogene* **15:** 1437-1444

Cullum R, Alder O, Hoodless PA (2011) The next generation: using new sequencing technologies to analyse gene regulation. *Respirology* **16:** 210-222

Culp TD, Budgeon LR, Christensen ND (2006) Human papillomaviruses bind a basal extracellular matrix component secreted by keratinocytes which is distinct from a membrane-associated receptor. *Virology* **347:** 147-159

Czyz J (2008) The stage-specific function of gap junctions during tumourigenesis. *Cellular & molecular biology letters* **13:** 92-102

Daly J, Licence S, Nanou A, Morgan G, Martensson IL (2007) Transcription of productive and nonproductive VDJ-recombined alleles after IgH allelic exclusion. *The EMBO journal* **26:** 4273-4282

Danos O, Katinka M, Yaniv M (1982) Human papillomavirus 1a complete DNA sequence: a novel type of genome organization among papovaviridae. *The EMBO journal* **1:** 231-236

Davis RL, Turner DL (2001) Vertebrate hairy and Enhancer of split related proteins: transcriptional repressors regulating cellular differentiation and embryonic patterning. *Oncogene* **20:** 8342-8357

Davy C, Doorbar J (2007) G2/M cell cycle arrest in the life cycle of viruses. *Virology* **368:** 219-226

Day PM, Lowy DR, Schiller JT (2008) Heparan sulfate-independent cell binding and infection with furin-precleaved papillomavirus capsids. *Journal of virology* **82:** 12565-12568

Day PM, Schiller JT (2009) The role of furin in papillomavirus infection. *Future microbiology* **4:** 1255-1262

de la Cruz X, Lois S, Sanchez-Molina S, Martinez-Balbas MA (2005) Do protein motifs read the histone code? *Bioessays* **27:** 164-175

de Pooter RF, Kee BL (2010) E proteins and the regulation of early lymphocyte development. *Immunol Rev* **238:** 93-109

de Villiers EM, Fauquet C, Broker TR, Bernard HU, zur Hausen H (2004) Classification of papillomaviruses. *Virology* **324:** 17-27

Dean A (2004) Chromatin remodelling and the interaction between enhancers and promoters in the beta-globin locus. *Brief Funct Genomic Proteomic* **2:** 344-354

Dean A (2006) On a chromosome far, far away: LCRs and gene expression. *Trends Genet* **22:** 38-45

Deed RW, Armitage S, Norton JD (1996) Nuclear localization and regulation of Id protein through an E protein-mediated chaperone mechanism. *J Biol Chem* **271:** 23603-23606

Degner SC, Verma-Gaur J, Wong TP, Bossen C, Iverson GM, Torkamani A, Vettermann C, Lin YC, Ju Z, Schulz D, Murre CS, Birshtein BK, Schork NJ, Schlissel MS, Riblet R, Murre C, Feeney AJ (2011) CCCTC-binding factor (CTCF) and cohesin influence the genomic architecture of the Igh locus and antisense transcription in pro-B cells. *Proceedings of the National Academy of Sciences of the United States of America* **108:** 9566-9571

Dekker J (2006) The three 'C' s of chromosome conformation capture: controls, controls, controls. *Nat Methods* **3:** 17-21

Dekker J, Rippe K, Dekker M, Kleckner N (2002) Capturing chromosome conformation. *Science* **295:** 1306-1311

Demers GW, Halbert CL, Galloway DA (1994) Elevated wild-type p53 protein levels in human epithelial cell lines immortalized by the human papillomavirus type 16 E7 gene. *Virology* **198:** 169-174

Desiderio SV, Yancopoulos GD, Paskind M, Thomas E, Boss MA, Landau N, Alt FW, Baltimore D (1984) Insertion of N regions into heavy-chain genes is correlated with expression of terminal deoxytransferase in B cells. *Nature* **311:** 752-755

Dignam JD, Lebovitz RM, Roeder RG (1983) Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res* **11:** 1475-1489

Doorbar J, Parton A, Hartley K, Banks L, Crook T, Stanley M, Crawford L (1990) Detection of novel splicing patterns in a HPV16-containing keratinocyte cell line. *Virology* **178:** 254-262

Dorigo B, Schalch T, Kulangara A, Duda S, Schroeder RR, Richmond TJ (2004) Nucleosome arrays reveal the two-start organization of the chromatin fiber. *Science* **306:** 1571-1573

Dorsett D (1999) Distant liaisons: long-range enhancer-promoter interactions in Drosophila. *Curr Opin Genet Dev* **9:** 505-514

Dostie J, Dekker J (2007) Mapping networks of physical interactions between genomic elements using 5C technology. *Nat Protoc* **2:** 988-1002

Dranoff G (2004) Cytokines in cancer pathogenesis and cancer therapy. *Nature reviews Cancer* **4:** 11-22

Duensing S, Lee LY, Duensing A, Basile J, Piboonniyom S, Gonzalez S, Crum CP, Munger K (2000) The human papillomavirus type 16 E6 and E7 oncoproteins cooperate to induce mitotic defects and genomic instability by uncoupling centrosome duplication from the cell division cycle. *Proceedings of the National Academy of Sciences of the United States of America* **97:** 10002-10007

Dyson N (1998) The regulation of E2F by pRB-family proteins. *Genes & development* **12:** 2245-2262

Eastman QM, Leu TM, Schatz DG (1996) Initiation of V(D)J recombination in vitro obeying the 12/23 rule. *Nature* **380:** 85-88

Eaton S, Calame K (1987) Multiple DNA sequence elements are necessary for the function of an immunoglobulin heavy chain promoter. *Proc Natl Acad Sci U S A* **84:** 7634-7638

Eccles S, Sarner N, Vidal M, Cox A, Grosveld F (1990) Enhancer sequences located 3' of the mouse immunoglobulin lambda locus specify high-level expression of an immunoglobulin lambda gene in B cells of transgenic mice. *The New biologist* **2:** 801-811

Ehlich A, Schaal S, Gu H, Kitamura D, Muller W, Rajewsky K (1993) Immunoglobulin heavy and light chain genes rearrange independently at early stages of B cell development. *Cell* **72:** 695-704

Einstein MH, Schiller JT, Viscidi RP, Strickler HD, Coursaget P, Tan T, Halsey N, Jenkins D (2009) Clinician's guide to human papillomavirus immunology: knowns and unknowns. *The Lancet infectious diseases* **9:** 347-356

Eisenbeis CF, Singh H, Storb U (1993) PU.1 is a component of a multiprotein complex which binds an essential site in the murine immunoglobulin lambda 2-4 enhancer. *Molecular and cellular biology* **13:** 6452-6461

Eisenbeis CF, Singh H, Storb U (1995) Pip, a novel IRF family member, is a lymphoid-specific, PU.1-dependent transcriptional activator. *Genes Dev* **9:** 1377-1387

Elenius K, Vainio S, Laato M, Salmivirta M, Thesleff I, Jalkanen M (1991) Induced expression of syndecan in healing wounds. *The Journal of cell biology* **114:** 585-595

Elgin SC (1988) The formation and function of DNase I hypersensitive sites in the process of gene activation. *J Biol Chem* **263:** 19259-19262

Ellenberger T, Fass D, Arnaud M, Harrison SC (1994) Crystal structure of transcription factor E47: E-box recognition by a basic region helix-loop-helix dimer. *Genes Dev* **8:** 970-980

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

Engel H, Ruhl H, Benham CJ, Bode J, Weiss S (2001) Germ-line transcripts of the immunoglobulin lambda J-C clusters in the mouse: characterization of the initiation sites and regulatory elements. *Mol Immunol* **38:** 289-302

Engel I, Murre C (2001) The function of E- and Id proteins in lymphocyte development. *Nat Rev Immunol* **1:** 193-199

Engel JD, Tanimoto K (2000) Looping, linking, and chromatin activity: new insights into beta-globin locus regulation. *Cell* **100:** 499-502

Enver T (1999) B-cell commitment: Pax5 is the deciding factor. *Current biology : CB* **9:** R933-935

Ephrussi A, Church GM, Tonegawa S, Gilbert W (1985) B lineage--specific interactions of an immunoglobulin enhancer with cellular factors in vivo. *Science* **227:** 134-140

Ernesto Oviedo-Orta BRK, William Howard Evans (2013) Connexin Cell Communication Channels: Roles in the Immune System and Immunopathology. 380

Ernst P, Smale ST (1995) Combinatorial regulation of transcription II: The immunoglobulin mu heavy chain gene. *Immunity* **2:** 427-438

Evander M, Frazer IH, Payne E, Qi YM, Hengst K, McMillan NA (1997) Identification of the alpha6 integrin as a candidate receptor for papillomaviruses. *Journal of virology* **71:** 2449-2456

Falkner FG, Zachau HG (1984) Correct transcription of an immunoglobulin kappa gene requires an upstream fragment containing conserved sequence elements. *Nature* **310:** 71-74

Fallaux FJ, Kranenburg O, Cramer SJ, Houweling A, Van Ormondt H, Hoeben RC, Van Der Eb AJ (1996) Characterization of 911: a new helper cell line for the titration and propagation of early region 1-deleted adenoviral vectors. *Human gene therapy* **7:** 215-222

Farrell RE RNA Methodologies: A Laboratory Guide for Isolation and Characterization.

Fauriat C, Long EO, Ljunggren HG, Bryceson YT (2010) Regulation of human NK-cell cytokine and chemokine production by target cell recognition. *Blood* **115:** 2167-2176

Fausch SC, Da Silva DM, Kast WM (2003) Differential uptake and cross-presentation of human papillomavirus virus-like particles by dendritic cells and Langerhans cells. *Cancer research* **63:** 3478-3482

Fausch SC, Da Silva DM, Rudolf MP, Kast WM (2002) Human papillomavirus virus-like particles do not activate Langerhans cells: a possible immune escape mechanism used by human papillomaviruses. *Journal of immunology* **169:** 3242-3249

Favre M (1975) Structural polypeptides of rabbit, bovine, and human papillomaviruses. *Journal of virology* **15:** 1239-1247

Fehrmann F, Klumpp DJ, Laimins LA (2003) Human papillomavirus type 31 E5 protein supports cell cycle progression and activates late viral functions upon epithelial differentiation. *Journal of virology* **77:** 2819-2831

Feller L, Wood NH, Khammissa RA, Chikte UM, Meyerov R, Lemmer J (2010) HPV modulation of host immune responses. *SADJ : journal of the South African Dental Association = tydskrif van die Suid-Afrikaanse Tandheelkundige Vereniging* **65:** 266-268

Felsenfeld G (1978) Chromatin. *Nature* **271:** 115-122

Feng J, Meyer CA, Wang Q, Liu JS, Shirley Liu X, Zhang Y (2012) GFOLD: a generalized fold change for ranking differentially expressed genes from RNA-seq data. *Bioinformatics* **28:** 2782-2788

Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM (2010) Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *International journal of cancer Journal international du cancer* **127:** 2893-2917

Finch JT, Noll M, Kornberg RD (1975) Electron microscopy of defined lengths of chromatin. *Proc Natl Acad Sci U S A* **72:** 3320-3322

Findlay J (2012) The development of an adenovirus vector system to study virus entry and genetic modification of immune cells. *PhD Thesis, University of Leeds*

Floor SL, Dumont JE, Maenhaut C, Raspe E (2012) Hallmarks of cancer: of all cancer cells, all the time? *Trends in molecular medicine* **18:** 509-515

Flores ER, Allen-Hoffmann BL, Lee D, Sattler CA, Lambert PF (1999) Establishment of the human papillomavirus type 16 (HPV-16) life cycle in an immortalized human foreskin keratinocyte cell line. *Virology* **262:** 344-354

Florin L, Sapp C, Streeck RE, Sapp M (2002) Assembly and translocation of papillomavirus capsid proteins. *Journal of virology* **76:** 10009-10014

Forrester WC, Thompson C, Elder JT, Groudine M (1986) A developmentally stable chromatin structure in the human beta-globin gene cluster. *Proc Natl Acad Sci U S A* **83:** 1359-1363

Fransson LA (2003) Glypicans. *The international journal of biochemistry & cell biology* **35:** 125-129

Fraser P, Grosveld F (1998) Locus control regions, chromatin activation and transcription. *Curr Opin Cell Biol* **10:** 361-365

Frazer IH (2009) Interaction of human papillomaviruses with the host immune system: a well evolved relationship. *Virology* **384:** 410-414

Friedl F, Kimura I, Osato T, Ito Y (1970) Studies on a new human cell line (SiHa) derived from carcinoma of uterus. I. Its establishment and morphology. *Proceedings of the Society for Experimental Biology and Medicine Society for Experimental Biology and Medicine* **135:** 543-545

Fujimoto E, Sato H, Shirai S, Nagashima Y, Fukumoto K, Hagiwara H, Negishi E, Ueno K, Omori Y, Yamasaki H, Hagiwara K, Yano T (2005) Connexin32 as a tumor suppressor gene in a metastatic renal cell carcinoma cell line. *Oncogene* **24:** 3684-3690

Fujimoto E, Yano T, Ueno K (2007) [Connexin32 as a tumor suppressor gene in renal cell carcinoma]. *Nihon Yakurigaku Zasshi* **129:** 105-109

Fukita Y, Jacobs H, Rajewsky K (1998) Somatic hypermutation in the heavy chain locus correlates with transcription. *Immunity* **9:** 105-114

Fuxa M, Skok J, Souabni A, Salvagiotto G, Roldan E, Busslinger M (2004) Pax5 induces V-to-DJ rearrangements and locus contraction of the immunoglobulin heavy-chain gene. *Genes Dev* **18:** 411-422

Gallo RL, Ono M, Povsic T, Page C, Eriksson E, Klagsbrun M, Bernfield M (1994) Syndecans, cell surface heparan sulfate proteoglycans, are induced by a proline-rich

antimicrobial peptide from wounds. *Proceedings of the National Academy of Sciences of the United States of America* **91:** 11035-11039

Gaszner M, Felsenfeld G (2006) Insulators: exploiting transcriptional and epigenetic mechanisms. *Nat Rev Genet* **7:** 703-713

Gemel J, Lin X, Veenstra RD, Beyer EC (2006) N-terminal residues in Cx43 and Cx40 determine physiological properties of gap junction channels, but do not influence heteromeric assembly with each other or with Cx26. *Journal of cell science* **119:** 2258-2268

Georgopoulos K, Bigby M, Wang JH, Molnar A, Wu P, Winandy S, Sharpe A (1994) The Ikaros gene is required for the development of all lymphoid lineages. *Cell* **79:** 143-156

Germain RN (2002) T-cell development and the CD4-CD8 lineage decision. *Nat Rev Immunol* **2:** 309-322

Gewin L, Galloway DA (2001) E box-dependent activation of telomerase by human papillomavirus type 16 E6 does not require induction of c-myc. *Journal of virology* **75:** 7198-7201

Ghil SH, Jeon YJ, Suh-Kim H (2002) Inhibition of BETA2/NeuroD by Id2. *Exp Mol Med* **34:** 367-373

Ghosh M, McAuliffe B, Subramani J, Basu S, Shapiro LH (2012) CD13 regulates dendritic cell cross-presentation and T cell responses by inhibiting receptor-mediated antigen uptake. *Journal of immunology* **188:** 5489-5499

Gilleron J, Fiorini C, Carette D, Avondet C, Falk MM, Segretain D, Pointis G (2008) Molecular reorganization of Cx43, Zo-1 and Src complexes during the endocytosis of gap junction plaques in response to a non-genomic carcinogen. *Journal of cell science* **121:** 4069-4078

Gillies SD, Morrison SL, Oi VT, Tonegawa S (1983) A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. *Cell* **33:** 717-728

Giroglou T, Florin L, Schafer F, Streeck RE, Sapp M (2001) Human papillomavirus infection requires cell surface heparan sulfate. *Journal of virology* **75:** 1565-1570

Golding A, Chandler S, Ballestar E, Wolffe AP, Schlissel MS (1999) Nucleosome structure completely inhibits in vitro cleavage by the V(D)J recombinase. *The EMBO journal* **18:** 3712-3723

Goliger JA, Paul DL (1994) Expression of gap junction proteins Cx26, Cx31.1, Cx37, and Cx43 in developing and mature rat epidermis. *Developmental dynamics : an official publication of the American Association of Anatomists* **200:** 1-13

Goodenough DA, Paul DL (2009) Gap junctions. *Cold Spring Harbor perspectives in biology* **1:** a002576

Graham SV (2008) Papillomavirus 3' UTR regulatory elements. *Frontiers in bioscience : a journal and virtual library* **13:** 5646-5663

Greenbaum S, Zhuang Y (2002) Regulation of early lymphocyte development by E2A family proteins. *Semin Immunol* **14:** 405-414

Grm HS, Banks L (2004) Degradation of hDlg and MAGIs by human papillomavirus E6 is E6-AP-independent. *The Journal of general virology* **85:** 2815-2819

Gross DS, Garrard WT (1988) Nuclease hypersensitive sites in chromatin. *Annu Rev Biochem* **57:** 159-197

Grosveld F, Antoniou M, Berry M, De Boer E, Dillon N, Ellis J, Fraser P, Hanscombe O, Hurst J, Imam A, et al. (1993) The regulation of human globin gene switching. *Philos Trans R Soc Lond B Biol Sci* **339:** 183-191

Grosveld F, van Assendelft GB, Greaves DR, Kollias G (1987) Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell* **51:** 975-985

Gu Z, Matlashewski G (1995) Effect of human papillomavirus type 16 oncogenes on MAP kinase activity. *Journal of virology* **69:** 8051-8056

Hagman J, Rudin CM, Haasch D, Chaplin D, Storb U (1990) A novel enhancer in the immunoglobulin lambda locus is duplicated and functionally independent of NF kappa B. *Genes & development* **4:** 978-992

Hamid NA, Brown C, Gaston K (2009) The regulation of cell proliferation by the papillomavirus early proteins. *Cell Mol Life Sci* **66:** 1700-1717

Hanahan D, Weinberg RA (2000) The hallmarks of cancer. *Cell* **100:** 57-70

Haque S (2004) The Role of Transcription Factors in V(D)J recombination. *PhD thesis, University of London*

Haque SF, Bevington SL, Boyes J (2013) The Elambda Enhancer is Essential for V(D)J Recombination of the Murine Immunoglobulin Lambda Light Chain Locus. *Biochemical and biophysical research communications*

Harbour JW, Dean DC (2000) Chromatin remodeling and Rb activity. *Current opinion in cell biology* **12:** 685-689

Hardy RR, Hayakawa K (1991) A developmental switch in B lymphopoiesis. *Proceedings of the National Academy of Sciences of the United States of America* **88:** 11550-11554

Hashida H, Takabayashi A, Kanai M, Adachi M, Kondo K, Kohno N, Yamaoka Y, Miyake M (2002) Aminopeptidase N is involved in cell motility and angiogenesis: its clinical significance in human colon cancer. *Gastroenterology* **122:** 376-386

Hatzis P, Talianidis I (2002) Dynamics of enhancer-promoter communication during differentiation-induced gene activation. *Mol Cell* **10:** 1467-1477

Hawley-Nelson P, Vousden KH, Hubbert NL, Lowy DR, Schiller JT (1989) HPV16 E6 and E7 proteins cooperate to immortalize human foreskin keratinocytes. *The EMBO journal* **8:** 3905-3910

Hegde P, Qi R, Gaspard R, Abernathy K, Dharap S, Earle-Hughes J, Gay C, Nwokekeh NU, Chen T, Saeed AI, Sharov V, Lee NH, Yeatman TJ, Quackenbush J (2001) Identification of tumor markers in models of human colorectal cancer using a 19,200-element complementary DNA microarray. *Cancer research* **61:** 7792-7797

Heidemann M, Eick D (2012) Tyrosine-1 and threonine-4 phosphorylation marks complete the RNA polymerase II CTD phospho-code. *RNA biology* **9:** 1144-1146

Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW, Ching KA, Antosiewicz-Bourget JE, Liu H, Zhang X, Green RD, Lobanenkov VV, Stewart R, Thomson JA, Crawford GE, Kellis M, Ren B (2009) Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459:** 108-112

Heng HH, Goetze S, Ye CJ, Liu G, Stevens JB, Bremer SW, Wykes SM, Bode J, Krawetz SA (2004) Chromatin loops are selectively anchored using scaffold/matrix-attachment regions. *J Cell Sci* **117:** 999-1008

Hewitt CRA (2006) Structural Biology and Functions of Immunoglobulins.

Hindmarsh PL, Laimins LA (2007) Mechanisms regulating expression of the HPV 31 L1 and L2 capsid proteins and pseudovirion entry. *Virology journal* **4:** 19

Hirschi KK, Xu CE, Tsukamoto T, Sager R (1996) Gap junction genes Cx26 and Cx43 individually suppress the cancer phenotype of human mammary carcinoma cells and restore differentiation potential. *Cell Growth Differ* **7:** 861-870

Ho IC, Yang LH, Morle G, Leiden JM (1989) A T-cell-specific transcriptional enhancer element 3' of C alpha in the human T-cell receptor alpha locus. *Proc Natl Acad Sci U S A* **86:** 6714-6718

Holwerda SJ, van de Werken HJ, Ribeiro de Almeida C, Bergen IM, de Bruijn MJ, Verstegen MJ, Simonis M, Splinter E, Wijchers PJ, Hendriks RW, de Laat W (2013) Allelic exclusion of the immunoglobulin heavy chain locus is independent of its nuclear localization in mature B cells. *Nucleic acids research* **41:** 6905-6916

Hotter D, Sauter D, Kirchhoff F (2013) Emerging role of the host restriction factor tetherin in viral immune sensing. *Journal of molecular biology* **425:** 4956-4964

Howie HL, Katzenellenbogen RA, Galloway DA (2009) Papillomavirus E6 proteins. *Virology* **384:** 324-334

Howley PM, Lowy DR (2006) Papillomaviridae. *Virology***:** 2299-2353

Hu P, Luo BH (2013) Integrin bi-directional signaling across the plasma membrane. *Journal of cellular physiology* **228:** 306-312

Hubbard T, Barker D, Birney E, Cameron G, Chen Y, Clark L, Cox T, Cuff J, Curwen V, Down T, Durbin R, Eyras E, Gilbert J, Hammond M, Huminiecki L, Kasprzyk A, Lehvaslaiho H, Lijnzaad P, Melsopp C, Mongin E, Pettett R, Pocock M, Potter S, Rust A, Schmidt E, Searle S, Slater G, Smith J, Spooner W, Stabenau A, Stalker J, Stupka E, Ureta-Vidal A, Vastrik I, Clamp M (2002) The Ensembl genome database project. *Nucleic acids research* **30:** 38-41

Huggins GS, Chin MT, Sibinga NE, Lee SL, Haber E, Lee ME (1999) Characterization of the mUBC9-binding sites required for E2A protein degradation. *J Biol Chem* **274:** 28690-28696

Hughes FJ, Romanos MA (1993) E1 protein of human papillomavirus is a DNA helicase/ATPase. *Nucleic acids research* **21:** 5817-5823

IARC (2008) Cancer incidence in five continents. *IARC scientific publications* **Volume IX:** 1-837

Ikawa T, Kawamoto H, Wright LY, Murre C (2004) Long-term cultured E2A-deficient hematopoietic progenitor cells are pluripotent. *Immunity* **20:** 349-360

Ikeda N, Nakajima Y, Tokuhara T, Hattori N, Sho M, Kanehiro H, Miyake M (2003) Clinical significance of aminopeptidase N/CD13 expression in human pancreatic carcinoma. *Clinical cancer research : an official journal of the American Association for Cancer Research* **9:** 1503-1508

Inlay M, Xu Y (2003) Epigenetic regulation of antigen receptor rearrangement. *Clinical immunology* **109:** 29-36

International Human Genome Sequencing C (2004) Finishing the euchromatic sequence of the human genome. *Nature* **431:** 931-945

Iyer VR, Eisen MB, Ross DT, Schuler G, Moore T, Lee JC, Trent JM, Staudt LM, Hudson J, Jr., Boguski MS, Lashkari D, Shalon D, Botstein D, Brown PO (1999) The transcriptional program in the response of human fibroblasts to serum. *Science* **283:** 83-87

Jacobs H, Bross L (2001) Towards an understanding of somatic hypermutation. *Curr Opin Immunol* **13:** 208-218

Jager L, Hausl MA, Rauschhuber C, Wolf NM, Kay MA, Ehrhardt A (2009) A rapid protocol for construction and production of high-capacity adenoviral vectors. *Nat Protoc* **4:** 547-564

Jang M, Rhee JE, Jang DH, Kim SS (2011) Gene expression profiles are altered in human papillomavirus-16 E6 D25E-expressing cell lines. *Virology journal* **8:** 453

Jang MK, Kwon D, McBride AA (2009) Papillomavirus E2 proteins and the host BRD4 protein associate with transcriptionally active cellular chromatin. *Journal of virology* **83:** 2592-2600

Javier RT (2008) Cell polarity proteins: common targets for tumorigenic human viruses. *Oncogene* **27:** 7031-7046

Jensen FC, Girardi AJ, Gilden RV, Koprowski H (1964) Infection of Human and Simian Tissue Cultures with Rous Sarcoma Virus. *Proceedings of the National Academy of Sciences of the United States of America* **52:** 53-59

Jenuwein T, Allis CD (2001) Translating the histone code. *Science* **293:** 1074-1080

Jeon S, Allen-Hoffmann BL, Lambert PF (1995) Integration of human papillomavirus type 16 into the human genome correlates with a selective growth advantage of cells. *Journal of virology* **69:** 2989-2997

Jeon S, Lambert PF (1995) Integration of human papillomavirus type 16 DNA into the human genome leads to increased stability of E6 and E7 mRNAs: implications for cervical carcinogenesis. *Proceedings of the National Academy of Sciences of the United States of America* **92:** 1654-1658

Jhunjhunwala S, van Zelm MC, Peak MM, Murre C (2009) Chromatin architecture and the generation of antigen receptor diversity. *Cell* **138:** 435-448

Jiang H, Chang FC, Ross AE, Lee J, Nakayama K, Nakayama K, Desiderio S (2005) Ubiquitylation of RAG-2 by Skp2-SCF links destruction of the V(D)J recombinase to the cell cycle. *Molecular cell* **18:** 699-709

Johansson C, Schwartz S (2013) Regulation of human papillomavirus gene expression by splicing and polyadenylation. *Nature reviews Microbiology* **11:** 239-251

Johnson K, Angelin-Duclos C, Park S, Calame KL (2003) Changes in histone acetylation are associated with differences in accessibility of V(H) gene segments to V-DJ recombination during B-cell ontogeny and development. *Molecular and cellular biology* **23:** 2438-2450

Jones DL, Thompson DA, Munger K (1997) Destabilization of the RB tumor suppressor protein and stabilization of p53 contribute to HPV type 16 E7-induced apoptosis. *Virology* **239:** 97-107

Joyce JG, Tung JS, Przysiecki CT, Cook JC, Lehman ED, Sands JA, Jansen KU, Keller PM (1999) The L1 major capsid protein of human papillomavirus type 11 recombinant virus-like particles interacts with heparin and cell-surface glycosaminoglycans on human keratinocytes. *The Journal of biological chemistry* **274:** 5810-5822

Ju Z, Volpi SA, Hassan R, Martinez N, Giannini SL, Gold T, Birshtein BK (2007) Evidence for physical interaction between the immunoglobulin heavy chain variable region and the 3' regulatory region. *The Journal of biological chemistry* **282:** 35169-35178

Kadaja M, Isok-Paas H, Laos T, Ustav E, Ustav M (2009) Mechanism of genomic instability in cells infected with the high-risk human papillomaviruses. *PLoS pathogens* **5:** e1000397

Kamper N, Day PM, Nowak T, Selinka HC, Florin L, Bolscher J, Hilbig L, Schiller JT, Sapp M (2006) A membrane-destabilizing peptide in capsid protein L2 is required for egress of papillomavirus genomes from endosomes. *Journal of virology* **80:** 759-768

Kane LP, Lin J, Weiss A (2000) Signal transduction by the TCR for antigen. *Curr Opin Immunol* **12:** 242-249

Karre K, Ljunggren HG, Piontek G, Kiessling R (1986) Selective rejection of H-2-deficient lymphoma variants suggests alternative immune defence strategy. *Nature* **319:** 675-678

Kataoka T, Kawakami T, Takahashi N, Honjo T (1980) Rearrangement of immunoglobulin gamma 1-chain gene and mechanism for heavy-chain class switch. *Proceedings of the National Academy of Sciences of the United States of America* **77:** 919-923

Kee BL, Murre C (2001) Transcription factor regulation of B lineage commitment. *Curr Opin Immunol* **13:** 180-185

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D (2002) The human genome browser at UCSC. *Genome Res* **12:** 996-1006

Kho CJ, Huggins GS, Endege WO, Hsieh CM, Lee ME, Haber E (1997) Degradation of E2A proteins through a ubiquitin-conjugating enzyme, UbcE2A. *J Biol Chem* **272:** 3845-3851

Kiang DT, Jin N, Tu ZJ, Lin HH (1997) Upstream genomic sequence of the human connexin26 gene. *Gene* **199:** 165-171

Kim ST, Shin Y, Brazin K, Mallis RJ, Sun ZY, Wagner G, Lang MJ, Reinherz EL (2012) TCR Mechanobiology: Torques and Tunable Structures Linked to Early T Cell Signaling. *Frontiers in immunology* **3:** 76

Kinoshita K, Honjo T (2001) Linking class-switch recombination with somatic hypermutation. *Nature reviews Molecular cell biology* **2:** 493-503

Kirnbauer R, Booy F, Cheng N, Lowy DR, Schiller JT (1992) Papillomavirus L1 major capsid protein self-assembles into virus-like particles that are highly immunogenic.

Proceedings of the National Academy of Sciences of the United States of America **89:** 12180-12184

Kirstetter P, Thomas M, Dierich A, Kastner P, Chan S (2002) Ikaros is critical for B cell differentiation and function. *European journal of immunology* **32:** 720-730

Klambt C, Knust E, Tietze K, Campos-Ortega JA (1989) Closely related transcripts encoded by the neurogenic gene complex enhancer of split of Drosophila melanogaster. *EMBO J* **8:** 203-210

Klingelhutz AJ, Foster SA, McDougall JK (1996) Telomerase activation by the E6 gene product of human papillomavirus type 16. *Nature* **380:** 79-82

Knappe M, Bodevin S, Selinka HC, Spillmann D, Streeck RE, Chen XS, Lindahl U, Sapp M (2007) Surface-exposed amino acid residues of HPV16 L1 protein mediating interaction with cell surface heparan sulfate. *The Journal of biological chemistry* **282:** 27913-27922

Knight GL, Grainger JR, Gallimore PH, Roberts S (2004) Cooperation between different forms of the human papillomavirus type 1 E4 protein to block cell cycle progression and cellular DNA synthesis. *Journal of virology* **78:** 13920-13933

Koga Y, Yamazaki N, Takizawa S, Kawauchi J, Nomura O, Yamamoto S, Saito N, Kakugawa Y, Otake Y, Matsumoto M, Matsumura Y (2014) Gene expression analysis using a highly sensitive DNA microarray for colorectal cancer screening. *Anticancer Res* **34:** 169-176

Kondo M, Akashi K, Domen J, Sugamura K, Weissman IL (1997) Bcl-2 rescues T lymphopoiesis, but not B or NK cell development, in common gamma chain-deficient mice. *Immunity* **7:** 155-162

Kornberg RD (1977) Structure of chromatin. *Annu Rev Biochem* **46:** 931-954

Koshiol JE, Schroeder JC, Jamieson DJ, Marshall SW, Duerr A, Heilig CM, Shah KV, Klein RS, Cu-Uvin S, Schuman P, Celentano D, Smith JS (2006) Time to clearance of human papillomavirus infection by type and human immunodeficiency virus serostatus. *International journal of cancer Journal international du cancer* **119:** 1623-1629

Koutsky LA, Ault KA, Wheeler CM, Brown DR, Barr E, Alvarez FB, Chiacchierini LM, Jansen KU, Proof of Principle Study I (2002) A controlled trial of a human papillomavirus type 16 vaccine. *N Engl J Med* **347:** 1645-1651

Kreider JW, Howett MK, Wolfe SA, Bartlett GL, Zaino RJ, Sedlacek T, Mortel R (1985) Morphological transformation in vivo of human uterine cervix with papillomavirus from condylomata acuminata. *Nature* **317:** 639-641

Kuhl BD, Sloan RD, Donahue DA, Bar-Magen T, Liang C, Wainberg MA (2010) Tetherin restricts direct cell-to-cell infection of HIV-1. *Retrovirology* **7:** 115

Kumar R, Ichihashi Y, Kimura S, Chitwood DH, Headland LR, Peng J, Maloof JN, Sinha NR (2012) A High-Throughput Method for Illumina RNA-Seq Library Preparation. *Frontiers in plant science* **3:** 202

Kumar V, Chambon P (1988) The estrogen receptor binds tightly to its responsive element as a ligand-induced homodimer. *Cell* **55:** 145-156

Kwon J, Imbalzano AN, Matthews A, Oettinger MA (1998) Accessibility of nucleosomal DNA to V(D)J cleavage is modulated by RSS positioning and HMG1. *Molecular cell* **2:** 829-839

Lahlou H, Fanjul M, Pradayrol L, Susini C, Pyronnet S (2005) Restoration of functional gap junctions through internal ribosome entry site-dependent synthesis of endogenous connexins in density-inhibited cancer cells. *Molecular and cellular biology* **25:** 4034-4045

Laird DW (2006) Life cycle of connexins in health and disease. *The Biochemical journal* **394:** 527-543

Laniosz V, Holthusen KA, Meneses PI (2008) Bovine papillomavirus type 1: from clathrin to caveolin. *Journal of virology* **82:** 6288-6298

Lassar AB, Davis RL, Wright WE, Kadesch T, Murre C, Voronova A, Baltimore D, Weintraub H (1991) Functional activity of myogenic HLH proteins requires hetero-oligomerization with E12/E47-like proteins in vivo. *Cell* **66:** 305-315

Laurson J, Raj K (2011) Localisation of human papillomavirus 16 E7 oncoprotein changes with cell confluence. *PloS one* **6:** e21501

Lazarczyk M, Cassonnet P, Pons C, Jacob Y, Favre M (2009) The EVER proteins as a natural barrier against papillomaviruses: a new insight into the pathogenesis of human papillomavirus infections. *Microbiol Mol Biol Rev* **73:** 348-370

Lazorchak AS, Schlissel MS, Zhuang Y (2006) E2A and IRF-4/Pip promote chromatin modification and transcription of the immunoglobulin kappa locus in pre-B cells. *Mol Cell Biol* **26:** 810-821

Lechner MS, Laimins LA (1994) Inhibition of p53 DNA binding by human papillomavirus E6 proteins. *Journal of virology* **68:** 4262-4273

Ledford H (2008) The death of microarrays? *Nature* **455:** 847

Lennon GG, Perry RP (1990) The temporal order of appearance of transcripts from unrearranged and rearranged Ig genes in murine fetal liver. *J Immunol* **144:** 1983-1987

Letian T, Tianyu Z (2010) Cellular receptor binding and entry of human papillomavirus. *Virology journal* **7:** 2

Letting DL, Rakowski C, Weiss MJ, Blobel GA (2003) Formation of a tissue-specific histone acetylation pattern by the hematopoietic transcription factor GATA-1. *Mol Cell Biol* **23:** 1334-1340

Lewis SM (1994) P nucleotide insertions and the resolution of hairpin DNA structures in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America* **91:** 1332-1336

Lewis SM, Agard E, Suh S, Czyzyk L (1997) Cryptic signals and the fidelity of V(D)J joining. *Molecular and cellular biology* **17:** 3125-3136

Li G, Reinberg D (2011) Chromatin higher-order structures and gene regulation. *Current opinion in genetics & development* **21:** 175-186

Li Q, Stamatoyannopoulos G (1994) Hypersensitive site 5 of the human beta locus control region functions as a chromatin insulator. *Blood* **84:** 1399-1401

Li YS, Hayakawa K, Hardy RR (1993) The regulated expression of B lineage associated genes during B cell differentiation in bone marrow and fetal liver. *J Exp Med* **178:** 951-960

Li Z, Dordai DI, 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:** 575-589

Liang HE, Hsu LY, Cado D, Schlissel MS (2004) Variegated transcriptional activation of the immunoglobulin kappa locus in pre-b cells contributes to the allelic exclusion of light-chain expression. *Cell* **118:** 19-29

Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326:** 289-293

Lin H, Grosschedl R (1995) Failure of B-cell differentiation in mice lacking the transcription factor EBF. *Nature* **376:** 263-267

Lin WC, Desiderio S (1993) Regulation of V(D)J recombination activator protein RAG-2 by phosphorylation. *Science* **260:** 953-959

Lin WC, Desiderio S (1994) Cell cycle regulation of V(D)J recombination-activating protein RAG-2. *Proceedings of the National Academy of Sciences of the United States of America* **91:** 2733-2737

Liu K, Nussenzweig MC (2010) Origin and development of dendritic cells. *Immunol Rev* **234:** 45-54

Lluis F, Ballestar E, Suelves M, Esteller M, Munoz-Canoves P (2005) E47 phosphorylation by p38 MAPK promotes MyoD/E47 association and muscle-specific gene transcription. *The EMBO journal* **24:** 974-984

Lodish H, Berk, A., Zipursky, S.L., Matsudaira, P., Baltimore, D., and Darnell, J. (2000) Molecular Cell Biology, 4th edition.

Lohoff M, Mak TW (2005) Roles of interferon-regulatory factors in T-helper-cell differentiation. *Nat Rev Immunol* **5:** 125-135

Loncarek J, Yamasaki H, Levillain P, Milinkevitch S, Mesnil M (2003) The expression of the tumor suppressor gene connexin 26 is not mediated by methylation in human esophageal cancer cells. *Molecular carcinogenesis* **36:** 74-81

Longo A, Guanga GP, Rose RB (2008) Crystal structure of E47-NeuroD1/beta2 bHLH domain-DNA complex: heterodimer selectivity and DNA recognition. *Biochemistry* **47:** 218-229

Longworth MS, Laimins LA (2004) Pathogenesis of human papillomaviruses in differentiating epithelia. *Microbiol Mol Biol Rev* **68:** 362-372

Loo YM, Melendy T (2004) Recruitment of replication protein A by the papillomavirus E1 protein and modulation by single-stranded DNA. *Journal of virology* **78:** 1605-1615

Lu R, Medina KL, Lancki DW, Singh H (2003) IRF-4,8 orchestrate the pre-B-to-B transition in lymphocyte development. *Genes Dev* **17:** 1703-1708

Ludmerer SW, Benincasa D, Mark GE, 3rd (1996) Two amino acid residues confer type specificity to a neutralizing, conformationally dependent epitope on human papillomavirus type 11. *Journal of virology* **70:** 4791-4794

Ludmerer SW, Benincasa D, Mark GE, 3rd, Christensen ND (1997) A neutralizing epitope of human papillomavirus type 11 is principally described by a continuous set of residues which overlap a distinct linear, surface-exposed epitope. *Journal of virology* **71:** 3834-3839

Luger K, Dechassa ML, Tremethick DJ (2012) New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nature reviews Molecular cell biology* **13:** 436-447

Luger K, Rechsteiner TJ, Flaus AJ, Waye MM, Richmond TJ (1997) Characterization of nucleosome core particles containing histone proteins made in bacteria. *J Mol Biol* **272:** 301-311

Luo J, Duggan DJ, Chen Y, Sauvageot J, Ewing CM, Bittner ML, Trent JM, Isaacs WB (2001) Human prostate cancer and benign prostatic hyperplasia: molecular dissection by gene expression profiling. *Cancer research* **61:** 4683-4688

Macdonald AI, Sun P, Hernandez-Lopez H, Aasen T, Hodgins MB, Edward M, Roberts S, Massimi P, Thomas M, Banks L, Graham SV (2012) A functional interaction between the MAGUK protein hDlg and the gap junction protein connexin 43 in cervical tumour cells. *The Biochemical journal* **446:** 9-21

Maeda S, Nakagawa S, Suga M, Yamashita E, Oshima A, Fujiyoshi Y, Tsukihara T (2009) Structure of the connexin 26 gap junction channel at 3.5 A resolution. *Nature* **458:** 597-602

Maes J, O'Neill LP, Cavelier P, Turner BM, Rougeon F, Goodhardt M (2001) Chromatin remodeling at the Ig loci prior to V(D)J recombination. *Journal of immunology* **167:** 866-874

Magram J, Chada K, Costantini F (1985) Developmental regulation of a cloned adult beta-globin gene in transgenic mice. *Nature* **315:** 338-340

Manis JP, Gu Y, Lansford R, Sonoda E, Ferrini R, Davidson L, Rajewsky K, Alt FW (1998) Ku70 is required for late B cell development and immunoglobulin heavy chain class switching. *J Exp Med* **187:** 2081-2089

Marculescu R, Vanura K, Montpellier B, Roulland S, Le T, Navarro JM, Jager U, McBlane F, Nadel B (2006) Recombinase, chromosomal translocations and lymphoid neoplasia: targeting mistakes and repair failures. *DNA repair* **5:** 1246-1258

Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y (2008) RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res* **18:** 1509-1517

Marsden MP, Laemmli UK (1979) Metaphase chromosome structure: evidence for a radial loop model. *Cell* **17:** 849-858

Martin JF, Miano JM, Hustad CM, Copeland NG, Jenkins NA, Olson EN (1994) A Mef2 gene that generates a muscle-specific isoform via alternative mRNA splicing. *Molecular and cellular biology* **14:** 1647-1656

Martin JF, Schwarz JJ, Olson EN (1993) Myocyte enhancer factor (MEF) 2C: a tissue-restricted member of the MEF-2 family of transcription factors. *Proceedings of the National Academy of Sciences of the United States of America* **90:** 5282-5286

Martin PE, Easton JA, Hodgins MB, Wright CS (2014) Connexins: sensors of epidermal integrity that are therapeutic targets. *FEBS letters* **588:** 1304-1314

Massari ME, Jennings PA, Murre C (1996) The AD1 transactivation domain of E2A contains a highly conserved helix which is required for its activity in both Saccharomyces cerevisiae and mammalian cells. *Mol Cell Biol* **16:** 121-129

Massari ME, Murre C (2000) Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. *Mol Cell Biol* **20:** 429-440

Matthias P, Rolink AG (2005) Transcriptional networks in developing and mature B cells. *Nat Rev Immunol* **5:** 497-508

Maxam AM, Gilbert W (1977) A new method for sequencing DNA. *Proceedings of the National Academy of Sciences of the United States of America* **74:** 560-564

McBlane F, Boyes J (2000) Stimulation of V(D)J recombination by histone acetylation. *Current biology : CB* **10:** 483-486

McBlane JF, van Gent DC, Ramsden DA, Romeo C, Cuomo CA, Gellert M, Oettinger MA (1995) Cleavage at a V(D)J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps. *Cell* **83:** 387-395

McCance DJ, Kopan R, Fuchs E, Laimins LA (1988) Human papillomavirus type 16 alters human epithelial cell differentiation in vitro. *Proceedings of the National Academy of Sciences of the United States of America* **85:** 7169-7173

McKercher SR, Torbett BE, Anderson KL, Henkel GW, Vestal DJ, Baribault H, Klemsz M, Feeney AJ, Wu GE, Paige CJ, Maki RA (1996) Targeted disruption of the PU.1 gene results in multiple hematopoietic abnormalities. *The EMBO journal* **15:** 5647-5658

McLachlan E, Shao Q, Wang HL, Langlois S, Laird DW (2006) Connexins act as tumor suppressors in three-dimensional mammary cell organoids by regulating differentiation and angiogenesis. *Cancer research* **66:** 9886-9894

McLaughlin-Drubin ME, Munger K (2009) Oncogenic activities of human papillomaviruses. *Virus Res* **143:** 195-208

McMillan NA, Payne E, Frazer IH, Evander M (1999) Expression of the alpha6 integrin confers papillomavirus binding upon receptor-negative B-cells. *Virology* **261:** 271-279

McNutt NS, Weinstein RS (1969) Carcinoma of the cervix: deficiency of nexus intercellular junctions. *Science* **165:** 597-599

Medzhitov R (2007) Recognition of microorganisms and activation of the immune response. *Nature* **449:** 819-826

Mehmood R, Yasuhara N, Oe S, Nagai M, Yoneda Y (2009) Synergistic nuclear import of NeuroD1 and its partner transcription factor, E47, via heterodimerization. *Exp Cell Res* **315:** 1639-1652

Melchers F (2005) The pre-B-cell receptor: selector of fitting immunoglobulin heavy chains for the B-cell repertoire. *Nat Rev Immunol* **5:** 578-584

Meldrum DR (1991) Factors affecting embryo implantation after human in vitro fertilization. *Am J Obstet Gynecol* **165:** 1896-1897

Meraro D, Hashmueli S, Koren B, Azriel A, Oumard A, Kirchhoff S, Hauser H, Nagulapalli S, Atchison ML, Levi BZ (1999) Protein-protein and DNA-protein

interactions affect the activity of lymphoid-specific IFN regulatory factors. *J Immunol* **163:** 6468-6478

Mesnil M, Crespin S, Avanzo JL, Zaidan-Dagli ML (2005) Defective gap junctional intercellular communication in the carcinogenic process. *Biochimica et biophysica acta* **1719:** 125-145

Mesnil M, Krutovskikh V, Piccoli C, Elfgang C, Traub O, Willecke K, Yamasaki H (1995) Negative growth control of HeLa cells by connexin genes: connexin species specificity. *Cancer research* **55:** 629-639

Meyers C, Frattini MG, Hudson JB, Laimins LA (1992) Biosynthesis of human papillomavirus from a continuous cell line upon epithelial differentiation. *Science* **257:** 971-973

Meylan E, Tschopp J, Karin M (2006) Intracellular pattern recognition receptors in the host response. *Nature* **442:** 39-44

MHRA (2010) Suspected Adverse Reaction Analysis CERVARIX Human papillomavirus (HPV) vaccine

Miller MB, Tang YW (2009) Basic concepts of microarrays and potential applications in clinical microbiology. *Clinical microbiology reviews* **22:** 611-633

Miosge LA, Goodnow CC (2005) Genes, pathways and checkpoints in lymphocyte development and homeostasis. *Immunol Cell Biol* **83:** 318-335

Mok SC, Chao J, Skates S, Wong K, Yiu GK, Muto MG, Berkowitz RS, Cramer DW (2001) Prostasin, a potential serum marker for ovarian cancer: identification through microarray technology. *Journal of the National Cancer Institute* **93:** 1458-1464

Moody CA, Laimins LA (2009) Human papillomaviruses activate the ATM DNA damage pathway for viral genome amplification upon differentiation. *PLoS pathogens* **5:** e1000605

Moody CA, Laimins LA (2010) Human papillomavirus oncoproteins: pathways to transformation. *Nature reviews Cancer* **10:** 550-560

Morcillo P, Rosen C, Baylies MK, Dorsett D (1997) Chip, a widely expressed chromosomal protein required for segmentation and activity of a remote wing margin enhancer in Drosophila. *Genes Dev* **11:** 2729-2740

Morcillo P, Rosen C, Dorsett D (1996) Genes regulating the remote wing margin enhancer in the Drosophila cut locus. *Genetics* **144:** 1143-1154

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5:** 621-628

Mortellaro A, Ricciardi-Castagnoli P (2011) From vaccine practice to vaccine science: the contribution of human immunology to the prevention of infectious disease. *Immunol Cell Biol* **89:** 332-339

Moura Franco RM, Linhares MM, Lustosa SS, Silva ID, Souza NC, Matos D (2013) Analysis of differentially expressed genes in colorectal adenocarcinoma with versus without metastasis by three-dimensional oligonucleotide microarray. *Int J Clin Exp Pathol* **7:** 255-263

Munger K, Baldwin A, Edwards KM, Hayakawa H, Nguyen CL, Owens M, Grace M, Huh K (2004) Mechanisms of human papillomavirus-induced oncogenesis. *Journal of virology* **78:** 11451-11460

Munger K, Phelps WC, Bubb V, Howley PM, Schlegel R (1989a) The E6 and E7 genes of the human papillomavirus type 16 together are necessary and sufficient for transformation of primary human keratinocytes. *Journal of virology* **63:** 4417-4421

Munger K, Werness BA, Dyson N, Phelps WC, Harlow E, Howley PM (1989b) Complex formation of human papillomavirus E7 proteins with the retinoblastoma tumor suppressor gene product. *The EMBO journal* **8:** 4099-4105

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:** 553-563

Muramatsu M, Sankaranand VS, Anant S, Sugai M, Kinoshita K, Davidson NO, Honjo T (1999) Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *The Journal of biological chemistry* **274:** 18470-18476

Murre C (2005) Helix-loop-helix proteins and lymphocyte development. *Nat Immunol* **6:** 1079-1086

Murre C, Bain G, van Dijk MA, Engel I, Furnari BA, Massari ME, Matthews JR, Quong MW, Rivera RR, Stuiver MH (1994) Structure and function of helix-loop-helix proteins. *Biochim Biophys Acta* **1218:** 129-135

Murre C, McCaw PS, Baltimore D (1989a) A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. *Cell* **56:** 777-783

Murre C, McCaw PS, Vaessin H, Caudy M, Jan LY, Jan YN, Cabrera CV, Buskin JN, Hauschka SD, Lassar AB, et al. (1989b) Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. *Cell* **58:** 537-544

Murre C, Voronova A, Baltimore D (1991) B-cell- and myocyte-specific E2-box-binding factors contain E12/E47-like subunits. *Mol Cell Biol* **11:** 1156-1160

Nagulapalli S, Atchison ML (1998) Transcription factor Pip can enhance DNA binding by E47, leading to transcriptional synergy involving multiple protein domains. *Mol Cell Biol* **18:** 4639-4650

Nagulapalli S, Goheer A, Pitt L, McIntosh LP, Atchison ML (2002) Mechanism of e47-Pip interaction on DNA resulting in transcriptional synergy and activation of immunoglobulin germ line sterile transcripts. *Mol Cell Biol* **22:** 7337-7350

Narni-Mancinelli E, Ugolini S, Vivier E (2013) Tuning the threshold of natural killer cell responses. *Curr Opin Immunol* **25:** 53-58

Naus CC (2002) Gap junctions and tumour progression. *Can J Physiol Pharmacol* **80:** 136-141

Naya FJ, Stellrecht CM, Tsai MJ (1995) Tissue-specific regulation of the insulin gene by a novel basic helix-loop-helix transcription factor. *Genes Dev* **9:** 1009-1019

Nechaev S, Adelman K (2011) Pol II waiting in the starting gates: Regulating the transition from transcription initiation into productive elongation. *Biochimica et biophysica acta* **1809:** 34-45

Nelsen B, Tian G, Erman B, Gregoire J, Maki R, Graves B, Sen R (1993) Regulation of lymphoid-specific immunoglobulin mu heavy chain gene enhancer by ETS-domain proteins. *Science* **261:** 82-86

Nelson LM, Rose RC, LeRoux L, Lane C, Bruya K, Moroianu J (2000) Nuclear import and DNA binding of human papillomavirus type 45 L1 capsid protein. *Journal of cellular biochemistry* **79:** 225-238

Nie L, Xu M, Vladimirova A, Sun XH (2003) Notch-induced E2A ubiquitination and degradation are controlled by MAP kinase activities. *EMBO J* **22:** 5780-5792

Niiro H, Clark EA (2002) Regulation of B-cell fate by antigen-receptor signals. *Nat Rev Immunol* **2:** 945-956

Noble WS (2009) How does multiple testing correction work? *Nat Biotechnol* **27:** 1135-1137

Nutt SL, Vambrie S, Steinlein P, Kozmik Z, Rolink A, Weith A, Busslinger M (1999) Independent regulation of the two Pax5 alleles during B-cell development. *Nature genetics* **21:** 390-395

O'Riordan M, Grosschedl R (1999) Coordinate regulation of B cell differentiation by the transcription factors EBF and E2A. *Immunity* **11:** 21-31

Oelze I, Kartenbeck J, Crusius K, Alonso A (1995) Human papillomavirus type 16 E5 protein affects cell-cell communication in an epithelial cell line. *Journal of virology* **69:** 4489-4494

Oettinger MA, Schatz DG, Gorka C, Baltimore D (1990) RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. *Science* **248:** 1517-1523

Ohlsson R, Gondor A (2007) The 4C technique: the 'Rosetta stone' for genome biology in 3D? *Curr Opin Cell Biol* **19:** 321-325

Ordentlich P, Lin A, Shen CP, Blaumueller C, Matsuno K, Artavanis-Tsakonas S, Kadesch T (1998) Notch inhibition of E47 supports the existence of a novel signaling pathway. *Mol Cell Biol* **18:** 2230-2239

Orkin SH (1995) Regulation of globin gene expression in erythroid cells. *Eur J Biochem* **231:** 271-281

Ozbun MA, Meyers C (1998) Temporal usage of multiple promoters during the life cycle of human papillomavirus type 31b. *Journal of virology* **72:** 2715-2722

Packard TA, Cambier JC (2013) B lymphocyte antigen receptor signaling: initiation, amplification, and regulation. *F1000Prime Rep* **5:** 40

Partin AW, Walsh AC, Epstein JI, Leventhal BG, Gearhart JP (1990) Nuclear morphometry as a predictor of response to therapy in Wilms tumor: a preliminary report. *J Urol* **144:** 952-954

Patterson NA, Smith JL, Ozbun MA (2005) Human papillomavirus type 31b infection of human keratinocytes does not require heparan sulfate. *Journal of virology* **79:** 6838-6847

Pattillo RA, Hussa RO, Story MT, Ruckert AC, Shalaby MR, Mattingly RF (1977) Tumor antigen and human chorionic gonadotropin in CaSki cells: a new epidermoid cervical cancer cell line. *Science* **196:** 1456-1458

Paulson JR, Laemmli UK (1977) The structure of histone-depleted metaphase chromosomes. *Cell* **12:** 817-828

Peh WL, Middleton K, Christensen N, Nicholls P, Egawa K, Sotlar K, Brandsma J, Percival A, Lewis J, Liu WJ, Doorbar J (2002) Life cycle heterogeneity in animal models of human papillomavirus-associated disease. *Journal of virology* **76:** 10401-10416

Pelkmans L, Helenius A (2003) Insider information: what viruses tell us about endocytosis. *Current opinion in cell biology* **15:** 414-422

Pertea M, Salzberg SL (2010) Between a chicken and a grape: estimating the number of human genes. *Genome biology* **11:** 206

Petterson M, Schaffner W (1987) A purine-rich DNA sequence motif present in SV40 and lymphotropic papovavirus binds a lymphoid-specific factor and contributes to enhancer activity in lymphoid cells. *Genes & development* **1:** 962-972

Phillips JE, Corces VG (2009) CTCF: master weaver of the genome. *Cell* **137:** 1194-1211

Pim D, Collins M, Banks L (1992) Human papillomavirus type 16 E5 gene stimulates the transforming activity of the epidermal growth factor receptor. *Oncogene* **7:** 27-32

Pitcher LA, van Oers NS (2003) T-cell receptor signal transmission: who gives an ITAM? *Trends in immunology* **24:** 554-560

Plante I, Stewart MK, Barr K, Allan AL, Laird DW (2011) Cx43 suppresses mammary tumor metastasis to the lung in a Cx43 mutant mouse model of human disease. *Oncogene* **30:** 1681-1692

Pointis G, Fiorini C, Gilleron J, Carette D, Segretain D (2007) Connexins as precocious markers and molecular targets for chemical and pharmacological agents in carcinogenesis. *Current medicinal chemistry* **14:** 2288-2303

Pollock R, Treisman R (1991) Human SRF-related proteins: DNA-binding properties and potential regulatory targets. *Genes & development* **5:** 2327-2341

Pongubala JM, Nagulapalli S, Klemsz MJ, McKercher SR, Maki RA, Atchison ML (1992) PU.1 recruits a second nuclear factor to a site important for immunoglobulin kappa 3' enhancer activity. *Molecular and cellular biology* **12:** 368-378

Pongubala JM, Van Beveren C, Nagulapalli S, Klemsz MJ, McKercher SR, Maki RA, Atchison ML (1993) Effect of PU.1 phosphorylation on interaction with NF-EM5 and transcriptional activation. *Science* **259:** 1622-1625

Portela A, Esteller M (2010) Epigenetic modifications and human disease. *Nat Biotechnol* **28:** 1057-1068

Predeus AV, Gopalakrishnan S, Huang Y, Tang J, Feeney AJ, Oltz EM, Artyomov MN (2014) Targeted chromatin profiling reveals novel enhancers in Ig H and Ig L chain Loci. *Journal of immunology* **192:** 1064-1070

Quong MW, Martensson A, Langerak AW, Rivera RR, Nemazee D, Murre C (2004) Receptor editing and marginal zone B cell development are regulated by the helix-loop-helix protein, E2A. *J Exp Med* **199:** 1101-1112

Quong MW, Massari ME, Zwart R, Murre C (1993) A new transcriptional-activation motif restricted to a class of helix-loop-helix proteins is functionally conserved in both yeast and mammalian cells. *Mol Cell Biol* **13:** 792-800

Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J (2011) A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470:** 279-283

Raff AB, Woodham AW, Raff LM, Skeate JG, Yan L, Da Silva DM, Schelhaas M, Kast WM (2013) The evolving field of human papillomavirus receptor research: a review of binding and entry. *Journal of virology* **87:** 6062-6072

Reik A, Telling A, Zitnik G, Cimbora D, Epner E, Groudine M (1998) The locus control region is necessary for gene expression in the human beta-globin locus but not the maintenance of an open chromatin structure in erythroid cells. *Mol Cell Biol* **18:** 5992-6000

Reilly EB, Blomberg B, Imanishi-Kari T, Tonegawa S, Eisen HN (1984) Restricted association of V and J-C gene segments for mouse lambda chains. *Proc Natl Acad Sci U S A* **81:** 2484-2488

Reinke H, Horz W (2003) Histones are first hyperacetylated and then lose contact with the activated PHO5 promoter. *Mol Cell* **11:** 1599-1607

Reizis B, Leder P (1999) Expression of the mouse pre-T cell receptor alpha gene is controlled by an upstream region containing a transcriptional enhancer. *J Exp Med* **189:** 1669-1678

Reynaud D, Demarco IA, Reddy KL, Schjerven H, Bertolino E, Chen Z, Smale ST, Winandy S, Singh H (2008) Regulation of B cell fate commitment and immunoglobulin heavy-chain gene rearrangements by Ikaros. *Nat Immunol* **9:** 927-936

Richards RM, Lowy DR, Schiller JT, Day PM (2006) Cleavage of the papillomavirus minor capsid protein, L2, at a furin consensus site is necessary for infection. *Proceedings of the National Academy of Sciences of the United States of America* **103:** 1522-1527

Richardson H, Kelsall G, Tellier P, Voyer H, Abrahamowicz M, Ferenczy A, Coutlee F, Franco EL (2003) The natural history of type-specific human papillomavirus infections in female university students. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **12:** 485-490

Riley RR, Duensing S, Brake T, Munger K, Lambert PF, Arbeit JM (2003) Dissection of human papillomavirus E6 and E7 function in transgenic mouse models of cervical carcinogenesis. *Cancer research* **63:** 4862-4871

Rivera R, Murre C (2001) The regulation and function of the Id proteins in lymphocyte development. *Oncogene* **20:** 8308-8316

Roberts JN, Buck CB, Thompson CD, Kines R, Bernardo M, Choyke PL, Lowy DR, Schiller JT (2007) Genital transmission of HPV in a mouse model is potentiated by nonoxynol-9 and inhibited by carrageenan. *Nature medicine* **13:** 857-861

Roberts S, Kingsbury SR, Stoeber K, Knight GL, Gallimore PH, Williams GH (2008) Identification of an arginine-rich motif in human papillomavirus type 1 E1;E4 protein necessary for E4-mediated inhibition of cellular DNA synthesis in vitro and in cells. *Journal of virology* **82:** 9056-9064

Robinson PJ, An W, Routh A, Martino F, Chapman L, Roeder RG, Rhodes D (2008) 30 nm chromatin fibre decompaction requires both H4-K16 acetylation and linker histone eviction. *J Mol Biol* **381:** 816-825

Robinson PJ, Rhodes D (2006) Structure of the '30 nm' chromatin fibre: a key role for the linker histone. *Curr Opin Struct Biol* **16:** 336-343

Roden RB, Greenstone HL, Kirnbauer R, Booy FP, Jessie J, Lowy DR, Schiller JT (1996) In vitro generation and type-specific neutralization of a human papillomavirus type 16 virion pseudotype. *Journal of virology* **70:** 5875-5883

Rodriguez MI, Finbow ME, Alonso A (2000) Binding of human papillomavirus 16 E5 to the 16 kDa subunit c (proteolipid) of the vacuolar H+-ATPase can be dissociated from the E5-mediated epidermal growth factor receptor overactivation. *Oncogene* **19:** 3727-3732

Rojas R, Apodaca G (2002) Immunoglobulin transport across polarized epithelial cells. *Nature reviews Molecular cell biology* **3:** 944-955

Roldan E, Fuxa M, Chong W, Martinez D, Novatchkova M, Busslinger M, Skok JA (2005) Locus 'decontraction' and centromeric recruitment contribute to allelic exclusion of the immunoglobulin heavy-chain gene. *Nat Immunol* **6:** 31-41

Rose RC, Bonnez W, Reichman RC, Garcea RL (1993) Expression of human papillomavirus type 11 L1 protein in insect cells: in vivo and in vitro assembly of viruslike particles. *Journal of virology* **67:** 1936-1944

Rosenberger S, De-Castro Arce J, Langbein L, Steenbergen RD, Rosl F (2010) Alternative splicing of human papillomavirus type-16 E6/E6* early mRNA is coupled to EGF signaling via Erk1/2 activation. *Proceedings of the National Academy of Sciences of the United States of America* **107:** 7006-7011

Rossi JL, Gissmann L, Jansen K, Muller M (2000) Assembly of human papillomavirus type 16 pseudovirions in Saccharomyces cerevisiae. *Human gene therapy* **11:** 1165-1176

Roth SD, Sapp M, Streeck RE, Selinka HC (2006) Characterization of neutralizing epitopes within the major capsid protein of human papillomavirus type 33. *Virology journal* **3:** 83

Rudin CM, Storb U (1992) Two conserved essential motifs of the murine immunoglobulin lambda enhancers bind B-cell-specific factors. *Molecular and cellular biology* **12:** 309-320

Samelson LE (2002) Signal transduction mediated by the T cell antigen receptor: the role of adapter proteins. *Annu Rev Immunol* **20:** 371-394

Sanger F, Coulson AR (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of molecular biology* **94:** 441-448

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* **74:** 5463-5467

Sapp M, Bienkowska-Haba M (2009) Viral entry mechanisms: human papillomavirus and a long journey from extracellular matrix to the nucleus. *The FEBS journal* **276:** 7206-7216

Sapp M, Day PM (2009) Structure, attachment and entry of polyoma- and papillomaviruses. *Virology* **384:** 400-409

Satyaraj E, Storb U (1998) Mef2 proteins, required for muscle differentiation, bind an essential site in the Ig lambda enhancer. *Journal of immunology* **161:** 4795-4802

Sawada S, Littman DR (1993) A heterodimer of HEB and an E12-related protein interacts with the CD4 enhancer and regulates its activity in T-cell lines. *Mol Cell Biol* **13:** 5620-5628

Schanke JT, Marcuzzi A, Podzorski RP, Van Ness B (1994) An AP1 binding site upstream of the kappa immunoglobulin intron enhancer binds inducible factors and contributes to expression. *Nucleic acids research* **22:** 5425-5432

Schatz DG, Oettinger MA, Schlissel MS (1992) V(D)J recombination: molecular biology and regulation. *Annu Rev Immunol* **10:** 359-383

Schebesta M, Heavey B, Busslinger M (2002) Transcriptional control of B-cell development. *Curr Opin Immunol* **14:** 216-223

Schena M, Heller RA, Theriault TP, Konrad K, Lachenmeier E, Davis RW (1998) Microarrays: biotechnology's discovery platform for functional genomics. *Trends Biotechnol* **16:** 301-306

Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270:** 467-470

Scherer WF, Syverton JT, Gey GO (1953) Studies on the propagation in vitro of poliomyelitis viruses. IV. Viral multiplication in a stable strain of human malignant epithelial cells (strain HeLa) derived from an epidermoid carcinoma of the cervix. *J Exp Med* **97:** 695-710

Schlissel MS (2004) Regulation of activation and recombination of the murine Igkappa locus. *Immunol Rev* **200:** 215-223

Schlissel MS, Baltimore D (1989) Activation of immunoglobulin kappa gene rearrangement correlates with induction of germline kappa gene transcription. *Cell* **58:** 1001-1007

Schwartz S (2013) Papillomavirus transcripts and posttranscriptional regulation. *Virology*

Schwarz E, Freese UK, Gissmann L, Mayer W, Roggenbuck B, Stremlau A, zur Hausen H (1985) Structure and transcription of human papillomavirus sequences in cervical carcinoma cells. *Nature* **314:** 111-114

Scott EW, Simon MC, Anastasi J, Singh H (1994) Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages. *Science* **265:** 1573-1577

Scott M, Nakagawa M, Moscicki AB (2001) Cell-mediated immune response to human papillomavirus infection. *Clinical and diagnostic laboratory immunology* **8:** 209-220

Seet CS, Brumbaugh RL, Kee BL (2004) Early B cell factor promotes B lymphopoiesis with reduced interleukin 7 responsiveness in the absence of E2A. *J Exp Med* **199:** 1689-1700

Selinka HC, Giroglou T, Sapp M (2002) Analysis of the infectious entry pathway of human papillomavirus type 33 pseudovirions. *Virology* **299:** 279-287

Shafti-Keramat S, Handisurya A, Kriehuber E, Meneguzzi G, Slupetzky K, Kirnbauer R (2003) Different heparan sulfate proteoglycans serve as cellular receptors for human papillomaviruses. *Journal of virology* **77:** 13125-13135

Shen CP, Kadesch T (1995) B-cell-specific DNA binding by an E47 homodimer. *Mol Cell Biol* **15:** 4518-4524

Shogren-Knaak M, Ishii H, Sun JM, Pazin MJ, Davie JR, Peterson CL (2006) Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* **311:** 844-847

Shore P, Sharrocks AD (1995) The MADS-box family of transcription factors. *European journal of biochemistry / FEBS* **229:** 1-13

Sigvardsson M, O'Riordan M, Grosschedl R (1997) EBF and E47 collaborate to induce expression of the endogenous immunoglobulin surrogate light chain genes. *Immunity* **7:** 25-36

Simonis M, Kooren J, de Laat W (2007) An evaluation of 3C-based methods to capture DNA interactions. *Nat Methods* **4:** 895-901

Singal R, Tu ZJ, Vanwert JM, Ginder GD, Kiang DT (2000) Modulation of the connexin26 tumor suppressor gene expression through methylation in human mammary epithelial cell lines. *Anticancer Res* **20:** 59-64

Smith JL, Campos SK, Ozbun MA (2007) Human papillomavirus type 31 uses a caveolin 1- and dynamin 2-mediated entry pathway for infection of human keratinocytes. *Journal of virology* **81:** 9922-9931

Sohl G, Willecke K (2004) Gap junctions and the connexin protein family. *Cardiovascular research* **62:** 228-232

Soler E, Andrieu-Soler C, de Boer E, Bryne JC, Thongjuea S, Stadhouders R, Palstra RJ, Stevens M, Kockx C, van Ijcken W, Hou J, Steinhoff C, Rijkers E, Lenhard B, Grosveld F The genome-wide dynamics of the binding of Ldb1 complexes during erythroid differentiation. *Genes Dev* **24:** 277-289

Splinter E, Heath H, Kooren J, Palstra RJ, Klous P, Grosveld F, Galjart N, de Laat W (2006) CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev* **20:** 2349-2354

Spoden G, Freitag K, Husmann M, Boller K, Sapp M, Lambert C, Florin L (2008) Clathrin- and caveolin-independent entry of human papillomavirus type 16--involvement of tetraspanin-enriched microdomains (TEMs). *PloS one* **3:** e3313

Srivastava P, Mangal M, Agarwal SM (2014) Understanding the transcriptional regulation of cervix cancer using microarray gene expression data and promoter sequence analysis of a curated gene set. *Gene* **535:** 233-238

Stacey SN, Jordan D, Snijders PJ, Mackett M, Walboomers JM, Arrand JR (1995) Translation of the human papillomavirus type 16 E7 oncoprotein from bicistronic mRNA is independent of splicing events within the E6 open reading frame. *Journal of virology* **69:** 7023-7031

Stanhope-Baker P, Hudson KM, Shaffer AL, Constantinescu A, Schlissel MS (1996) Cell type-specific chromatin structure determines the targeting of V(D)J recombinase activity in vitro. *Cell* **85:** 887-897

Stanley M (2008) Immunobiology of HPV and HPV vaccines. *Gynecologic oncology* **109:** S15-21

Stanley MA (2009) Immune responses to human papilloma viruses. *The Indian journal of medical research* **130:** 266-276

Stanley MA (2012) Epithelial cell responses to infection with human papillomavirus. *Clinical microbiology reviews* **25:** 215-222

Steben M, Duarte-Franco E (2007) Human papillomavirus infection: epidemiology and pathophysiology. *Gynecologic oncology* **107:** S2-5

Stetson DB, Medzhitov R (2006) Type I interferons in host defense. *Immunity* **25:** 373-381

Strahl BD, Allis CD (2000) The language of covalent histone modifications. *Nature* **403:** 41-45

Stubenrauch F, Lim HB, Laimins LA (1998) Differential requirements for conserved E2 binding sites in the life cycle of oncogenic human papillomavirus type 31. *Journal of virology* **72:** 1071-1077

Sumino J, Uzawa N, Okada N, Miyaguchi K, Mogushi K, Takahashi K, Sato H, Michikawa C, Nakata Y, Tanaka H, Amagasa T (2013) Gene expression changes in initiation and progression of oral squamous cell carcinomas revealed by laser microdissection and oligonucleotide microarray analysis. *International journal of cancer Journal international du cancer* **132:** 540-548

Sun L, Zhang G, Lei T, Huang C, Song T, Si L (2008) Two different HPV-11E6 fusion proteins trap p53 in the cytoplasm and induce apoptosis. *Cancer biology & therapy* **7:** 1909-1915

Sun XH, Baltimore D (1991) An inhibitory domain of E12 transcription factor prevents DNA binding in E12 homodimers but not in E12 heterodimers. *Cell* **64:** 459-470

Takaoka A, Yanai H, Kondo S, Duncan G, Negishi H, Mizutani T, Kano S, Honda K, Ohba Y, Mak TW, Taniguchi T (2005) Integral role of IRF-5 in the gene induction programme activated by Toll-like receptors. *Nature* **434:** 243-249

Tan LW, Bianco T, Dobrovic A (2002) Variable promoter region CpG island methylation of the putative tumor suppressor gene Connexin 26 in breast cancer. *Carcinogenesis* **23:** 231-236

Tan TM, Gloss B, Bernard HU, Ting RC (1994) Mechanism of translation of the bicistronic mRNA encoding human papillomavirus type 16 E6-E7 genes. *The Journal of general virology* **75 (Pt 10):** 2663-2670

Teachenor R, Beck K, Wright LY, Shen Z, Briggs SA, Murre C (2012a) Biochemical and phosphoproteomic analysis of the helix-loop-helix protein E47. *Mol Cell Biol*

Teachenor R, Beck K, Wright LY, Shen Z, Briggs SP, Murre C (2012b) Biochemical and phosphoproteomic analysis of the helix-loop-helix protein E47. *Molecular and cellular biology* **32:** 1671-1682

Thierry F (2009) Transcriptional regulation of the papillomavirus oncogenes by cellular and viral transcription factors in cervical carcinoma. *Virology* **384:** 375-379

Thomas M, Narayan N, Pim D, Tomaic V, Massimi P, Nagasaka K, Kranjec C, Gammoh N, Banks L (2008) Human papillomaviruses, cervical cancer and cell polarity. *Oncogene* **27:** 7018-7030

Tokarev A, Skasko M, Fitzpatrick K, Guatelli J (2009) Antiviral activity of the interferon-induced cellular protein BST-2/tetherin. *AIDS Res Hum Retroviruses* **25:** 1197-1210

Tokuhara T, Hattori N, Ishida H, Hirai T, Higashiyama M, Kodama K, Miyake M (2006) Clinical significance of aminopeptidase N in non-small cell lung cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* **12:** 3971-3978

Townes TM, Lingrel JB, Chen HY, Brinster RL, Palmiter RD (1985) Erythroid-specific expression of human beta-globin genes in transgenic mice. *EMBO J* **4:** 1715-1723

Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25:** 1105-1111

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7:** 562-578

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28:** 511-515

Traver D, Akashi K, Manz M, Merad M, Miyamoto T, Engleman EG, Weissman IL (2000) Development of CD8alpha-positive dendritic cells from a common myeloid progenitor. *Science* **290:** 2152-2154

Travers A (1999) Chromatin modification by DNA tracking. *Proc Natl Acad Sci U S A* **96:** 13634-13637

Tu ZJ, Kiang DT (1998) Mapping and characterization of the basal promoter of the human connexin26 gene. *Biochimica et biophysica acta* **1443:** 169-181

Tyler DS, Lyerly HK, Weinhold KJ (1989) Anti-HIV-1 ADCC. *AIDS Res Hum Retroviruses* **5:** 557-563

Unckell F, Streeck RE, Sapp M (1997) Generation and neutralization of pseudovirions of human papillomavirus type 33. *Journal of virology* **71:** 2934-2939

Underhill DM, Goodridge HS (2012) Information processing during phagocytosis. *Nat Rev Immunol* **12:** 492-502

Urbanek P, Wang ZQ, Fetka I, Wagner EF, Busslinger M (1994) Complete block of early B cell differentiation and altered patterning of the posterior midbrain in mice lacking Pax5/BSAP. *Cell* **79:** 901-912

Veillette A, Latour S, Davidson D (2002) Negative regulation of immunoreceptor signaling. *Annu Rev Immunol* **20:** 669-707

Velculescu VE, Zhang L, Zhou W, Vogelstein J, Basrai MA, Bassett DE, Jr., Hieter P, Vogelstein B, Kinzler KW (1997) Characterization of the yeast transcriptome. *Cell* **88:** 243-251

Vernimmen D, De Gobbi M, Sloane-Stanley JA, Wood WG, Higgs DR (2007) Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression. *EMBO J* **26:** 2041-2051

Vignali M, Hassan AH, Neely KE, Workman JL (2000) ATP-dependent chromatin-remodeling complexes. *Mol Cell Biol* **20:** 1899-1910

Visel A, Rubin EM, Pennacchio LA (2009) Genomic views of distant-acting enhancers. *Nature* **461:** 199-205

Volpers C, Schirmacher P, Streeck RE, Sapp M (1994) Assembly of the major and the minor capsid protein of human papillomavirus type 33 into virus-like particles and tubular structures in insect cells. *Virology* **200:** 504-512

Walboomers JM, Jacobs MV, Manos MM, Bosch FX, Kummer JA, Shah KV, Snijders PJ, Peto J, Meijer CJ, Munoz N (1999) Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *The Journal of pathology* **189:** 12-19

Wang JH, Reinherz EL (2002) Structural basis of T cell recognition of peptides bound to MHC molecules. *Mol Immunol* **38:** 1039-1049

Watherston OG (2010) The effect of small DNA tumour virus oncoproteins on the expression of genes involved in the immune system. *PhD thesis, Leeds*

Wendt H, Thomas RM, Ellenberger T (1998) DNA-mediated folding and assembly of MyoD-E47 heterodimers. *J Biol Chem* **273:** 5735-5743

West AG, Gaszner M, Felsenfeld G (2002) Insulators: many functions, many mechanisms. *Genes Dev* **16:** 271-288

Wheeler DL, Church DM, Lash AE, Leipe DD, Madden TL, Pontius JU, Schuler GD, Schriml LM, Tatusova TA, Wagner L, Rapp BA (2001) Database resources of the National Center for Biotechnology Information. *Nucleic acids research* **29:** 11-16

WHO G (2008) Comprehensive cervical cancer prevention and control: a healthier future for girls and women.

Williams A, Spilianakis CG, Flavell RA (2010) Interchromosomal association and gene regulation in trans. *Trends in genetics : TIG* **26:** 188-197

Wise-Draper TM, Wells SI (2008) Papillomavirus E6 and E7 proteins and their cellular targets. *Frontiers in bioscience : a journal and virtual library* **13:** 1003-1017

Wiszniewski L, Limat A, Saurat JH, Meda P, Salomon D (2000) Differential expression of connexins during stratification of human keratinocytes. *The Journal of investigative dermatology* **115:** 278-285

Wu C (1980) The 5' ends of Drosophila heat shock genes in chromatin are hypersensitive to DNase I. *Nature* **286:** 854-860

Wu L, Nichogiannopoulou A, Shortman K, Georgopoulos K (1997) Cell-autonomous defects in dendritic cell populations of Ikaros mutant mice point to a developmental relationship with the lymphoid lineage. *Immunity* **7:** 483-492

Wu SY, Lee AY, Hou SY, Kemper JK, Erdjument-Bromage H, Tempst P, Chiang CM (2006) Brd4 links chromatin targeting to HPV transcriptional silencing. *Genes & development* **20:** 2383-2396

Yancopoulos GD, Alt FW (1985) Developmentally controlled and tissue-specific expression of unrearranged VH gene segments. *Cell* **40:** 271-281

Yancopoulos GD, Alt FW (1986) Regulation of the assembly and expression of variable-region genes. *Annu Rev Immunol* **4:** 339-368

Yeager CL, Ashmun RA, Williams RK, Cardellichio CB, Shapiro LH, Look AT, Holmes KV (1992) Human aminopeptidase N is a receptor for human coronavirus 229E. *Nature* **357:** 420-422

Yoon CS, Kim KD, Park SN, Cheong SW (2001) alpha(6) Integrin is the main receptor of human papillomavirus type 16 VLP. *Biochemical and biophysical research communications* **283:** 668-673

You J, Croyle JL, Nishimura A, Ozato K, Howley PM (2004) Interaction of the bovine papillomavirus E2 protein with Brd4 tethers the viral DNA to host mitotic chromosomes. *Cell* **117:** 349-360

Zerfass K, Schulze A, Spitkovsky D, Friedman V, Henglein B, Jansen-Durr P (1995) Sequential activation of cyclin E and cyclin A gene expression by human papillomavirus type 16 E7 through sequences necessary for transformation. *Journal of virology* **69:** 6389-6399

Zhang B, Li P, Wang E, Brahmi Z, Dunn KW, Blum JS, Roman A (2003) The E5 protein of human papillomavirus type 16 perturbs MHC class II antigen maturation in human foreskin keratinocytes treated with interferon-gamma. *Virology* **310:** 100-108

Zhang L, Zhou W, Velculescu VE, Kern SE, Hruban RH, Hamilton SR, Vogelstein B, Kinzler KW (1997) Gene expression profiles in normal and cancer cells. *Science* **276:** 1268-1272

Zhao H, Dean A (2004) An insulator blocks spreading of histone acetylation and interferes with RNA polymerase II transfer between an enhancer and gene. *Nucleic Acids Res* **32:** 4903-4919

Zhao Z, Tavoosidana G, Sjolinder M, Gondor A, Mariano P, Wang S, Kanduri C, Lezcano M, Sandhu KS, Singh U, Pant V, Tiwari V, Kurukuti S, Ohlsson R (2006) Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat Genet* **38:** 1341-1347

Zheng ZM, Baker CC (2006) Papillomavirus genome structure, expression, and post-transcriptional regulation. *Frontiers in bioscience : a journal and virtual library* **11:** 2286-2302

Zhou J, Sun XY, Stenzel DJ, Frazer IH (1991) Expression of vaccinia recombinant HPV 16 L1 and L2 ORF proteins in epithelial cells is sufficient for assembly of HPV virion-like particles. *Virology* **185:** 251-257

Zhuang Y, Cheng P, Weintraub H (1996) B-lymphocyte development is regulated by the combined dosage of three basic helix-loop-helix genes, E2A, E2-2, and HEB. *Molecular and cellular biology* **16:** 2898-2905

Zhuang Y, Jackson A, Pan L, Shen K, Dai M (2004) Regulation of E2A gene expression in B-lymphocyte development. *Mol Immunol* **40:** 1165-1177

Zhuang Y, Soriano P, Weintraub H (1994) The helix-loop-helix gene E2A is required for B cell formation. *Cell* **79:** 875-884

zur Hausen H (2009) Papillomaviruses in the causation of human cancers - a brief historical account. *Virology* **384:** 260-265