



THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.



THE UNIVERSITY
of EDINBURGH

**Investigating PAX6 and SOX2 dynamic
interactions at the single molecule level in live
cells**

Ioannis Christodoulou

Thesis submitted for the degree of Doctor of Philosophy at the
University of Edinburgh

2021

Declaration

I am the sole author of this thesis and all work has been completed by myself unless otherwise stated. None of this work has been submitted for any other degree or qualification.

Signed... ..Date: 30/11/2021.....

Acknowledgements

Firstly, I would like to thank my supervisor Dr Dimitrios Papadopoulos for mentoring and believing in me throughout my PhD. I would also like to thank my secondary supervisors Professor Wendy Bickmore and Professor David FitzPatrick, as well as my remaining thesis committee, chair Professor Nick Gilbert and external chair Dr Andrew Wood. Your help and guidance has been greatly influential these past four years.

Secondly I would like to thank Julia Auer for the endless discussions concerning science, or not, and for always helping me in times of need. I would also like to thank everyone in Professor FitzPatrick's lab for all their help and especially Dr Hemant Bengani and Dr Nikki Hall.

Most importantly though I would like to thank my friends and family for this would have not been possible without them. To my friends, the ones here but also the ones thousands of miles away you have been a pillar of support. I would namely like to thank Andreas, Nicholas, Stylianos, Alexander and especially Anthony! To my parents Panos and Maria, my sister Ersi and grandmothers Ersi and Paraskevou, this would not have been possible without your constant support and drive. I am extremely grateful for all your support and belief in me throughout these years. None of this would have been possible without your words of encouragement. Last but not least I would like to thank Ana for all her help and support over the last four years. I am certain without her, my PhD life would have been very different. Lastly I would like to dedicate this thesis to my grandfather Andreas. I miss you.

Abstract

The abundance of transcription factor (TF) molecules in the nuclei of eukaryotic cells are in the range of thousands. However, the functional binding sites of most TFs lie in the range of hundreds. This suggests that there is a surplus of the number of molecules for many TFs, relative to their binding sites at any given time. Nevertheless, precise TF levels are instrumental for normal development and maintenance, with haploinsufficiency (namely lowering the dosage of a TF by half) being a hallmark of many TF-related human developmental disorders. Qualitative methods assessing TF binding such as chromatin immunoprecipitation, provide static information, from fixed cell populations and so fail to provide insight into TF dynamic behaviour. Live-cell imaging methodologies such as Fluorescence Correlation Spectroscopy (FCS) offer the ability to measure kinetics of binding to chromatin, protein-protein interactions, absolute concentrations of molecules and the underlying cell-to-cell variability.

SOX2 and PAX6 TFs exhibit haploinsufficiency in humans. Heterozygous point mutations, deletions or insertions in these genes can lead to a plethora of abnormal ocular developmental disorders (e.g. coloboma, aniridia, microphthalmia, anophthalmia). SOX2 encodes a high-mobility group (HMG) domain-containing TF, essential for maintaining self-renewal of embryonic stem cells and is expressed in proliferating central nervous system (CNS) progenitors. PAX6 contains two DNA binding domains; a PAIRED domain (PD) and a homeodomain (HD). Both DNA binding domains present in PAX6 (PD and HD) can function either jointly, or separately, to regulate a plethora of genes implicated in the development and maintenance of the CNS, the eye and the pancreas. Despite existing genetic and phenotypic evidence, it remains unclear how PAX6 and SOX2 influence each other at the molecular level and how sensitive their stoichiometry is during ocular development.

In this thesis I investigated the dynamic interplay between PAX6/SOX2 and chromatin in live cells, at the molecular level. I compared wild-type protein function with pathogenic missense variants using advanced fluorescence

microscopy techniques and assessed how these mutations quantitatively and qualitatively affected molecular behaviour. My results showed that both SOX2 and PAX6 pathogenic missense mutants display differential subnuclear localisation, as well as altered protein-protein and protein-chromatin interactions, linking molecular diffusion to pathogenic phenotype in humans. More importantly, I identified a novel role of SOX2 in stabilising PAX6-chromatin complexes in live cells, providing further insight into the complex and dynamic relation of PAX6 and SOX2 in ocular tissue specification, maintenance and development.

Lay Summary

The human body consists of roughly 30 trillion cells. Inside each cell, there is a nucleus that acts as the cell's 'brain' and contains all the genetic material and instructions required for the cell to function correctly. The genetic material is made up of a chemical called DNA (deoxyribonucleic acid) which consists of two long molecules that are arranged in a spiral, like a twisted ladder. This is called the double helix structure. DNA contains all the information that a living organism needs to grow, reproduce and function. Genes are short sections of DNA that carry information for characteristics such as hair and eye colour. There are roughly 20,000 genes in humans and each gene has specific information, which the cell uses to make a specific protein. Most genetic diseases in humans are caused because certain proteins are not made at all, or they are made faulty. *PAX6* and *SOX2* are transcription factor proteins. Transcription factors are proteins which directly interact with DNA at specific regions and influence when a gene is turned on or off, ensuring that the correct proteins are made in the correct cell at the correct time.

One of the key functions of *PAX6* and *SOX2* is to ensure correct eye development. When these proteins are made incorrectly, this leads to eye diseases. These diseases are usually aniridia or anophthalmia when *PAX6* or *SOX2* respectively, are faulty. Aniridia occurs when the iris, the coloured part of the eye, does not fully form causing variable vision impairment and increased sensitivity to light. Patients with aniridia, usually develop additional eye problems such as increased pressure in the eye (glaucoma) and clouding of the lens (cataract), which further impair vision. Anophthalmia is the medical term used for the absence of one or both eyes.

In my PhD, I wanted to investigate how faulty *PAX6* and *SOX2* proteins interact with DNA and each other and how this may contribute to eye diseases observed in humans. By using advanced microscopy methodologies such as FCS (Fluorescence Correlation Spectroscopy), I was able to link the strength of protein to DNA interaction with eye disease severity. The weaker the interaction of the protein with DNA, the more severe the eye disease observed

in human patients. In addition, SOX2 and PAX6 faulty proteins had altered localisation in the nucleus when compared to the normal proteins, potentially further contributing to abnormal eye development.

Previous studies have shown that *PAX6* and *SOX2* can work together to turn other genes on and off, making their interaction very important. Through my experiments, I was able to show for the first time in live cells, that SOX2 is responsible for the stabilisation of the SOX2-PAX6-DNA complex, providing further understanding into the complicated relation of these two proteins in gene regulation and ultimately, normal eye development.

Table of Contents

Chapter 1 Introduction	20
1.1 The Evolution of Vision	20
1.2 Human eye development.....	21
1.3 Anatomy of the developed human eye.....	24
1.3.1 Outer Layer	25
1.3.2 Middle Layer.....	26
1.3.3 Inner Layer	27
1.4 <i>PAX6</i> Gene and Protein Architecture	29
1.4.1 <i>PAX6</i> Gene Architecture	29
1.4.2 <i>PAX6</i> Protein Architecture.....	30
1.5 <i>PAX6</i> Expression and Function	34
1.5.1 <i>PAX6</i> in the Pancreas	34
1.5.2 <i>PAX6</i> in the Brain	35
1.5.3 <i>PAX6</i> in the Eye	38
1.5.4 <i>PAX6</i> in the developing lens	39
1.5.5 <i>PAX6</i> Pathogenic Mutations.....	42
1.6 <i>SOX2</i> Gene and Protein Architecture	44
1.6.1 <i>SOX2</i> Gene Architecture.....	44
1.6.2 <i>SOX2</i> Protein Architecture	45
1.6.3 <i>SOX2</i> Posttranslational Modifications	47
1.6.4 <i>SOX2</i> Interaction with RNA	49
1.7 <i>SOX2</i> Expression and Function.....	51
1.7.1 Embryonic stem cells	51
1.7.2 Central Nervous System	53

1.8	Molecular mechanisms of Transcription Factor function	57
1.8.1	Transcription Factor Haploinsufficiency	57
1.8.2	TF Intranuclear Concentration and Binding Site Availability	60
1.8.3	Protein Phase Separation	62
1.9	Overview of advanced microscopy methodologies	66
1.9.1	Fluorescence Correlation Spectroscopy (FCS)	66
1.9.2	Fluorescence Cross-Correlation Spectroscopy (FCCS)	68
1.9.3	Single-Particle Tracking (SPT)	70
1.9.4	Fluorescence Recovery after Photobleaching (FRAP)	71
1.9.5	Förster Resonance Energy Transfer (FRET)	72
1.9.6	Stochastic Optical Reconstruction Microscopy (STORM)	73
1.10	Overview of Fluorescence Correlation Spectroscopy	74
1.10.1	FCS instrumentation and basic working principles	74
1.10.2	FCS applications in biological systems	79
1.11	Thesis Aims	80
Chapter 2 Materials and Methods		82
2.1	Cell culture	82
2.1.1	HEK293 Cell line maintenance	82
2.1.2	mESC Cell line maintenance	82
2.1.3	Sub-culturing/passaging cells	83
2.1.4	Thawing/Freezing cells	83
2.1.5	Transfection	84
2.1.6	DNA extraction	84
2.1.7	Immunofluorescence	84
2.1.8	Co-immunoprecipitation	86

2.2	Plasmid Construction	88
2.2.1	Electrocompetent cell preparation	88
2.2.2	TOPO Cloning	89
2.2.3	Colony PCR and Plasmid Purification	90
2.2.4	Site-directed mutagenesis	91
2.2.5	Primer Sequences	92
2.3	HEK293 Stable Cell line generation.....	98
2.4	FCS.....	101
2.4.1	FCS measurements	101
2.4.2	FCS analysis	102
2.4.3	FCCS measurements.....	104
2.5	Condensate assays	106
2.5.1	Alexa fluorophore protein conjugation	106
2.5.2	Droplet assay setup.....	106
2.5.3	Condensate dissolution in live cells.....	107
2.6	PONDR Score Predictor and CIDER webserver analysis.....	107
2.7	Genome editing using CRISPR-Cas9	108
2.7.1	mESCs	108
2.7.2	Zebrafish	109
2.8	Confocal Imaging	109
Chapter 3 The molecular mechanism of PAX6 and SOX2 missense mutations in ocular diseases		110
3.1	Introduction	110
3.2	Results.....	114
3.2.1	Fluorophore position influences the molecular behaviour of PAX6.....	114

3.2.2 PAX6 mutant variants exhibit weaker chromatin binding dynamics and altered subnuclear localisation	127
3.3 Discussion	140
Chapter 4 Study of the regulation of PAX6 function and nuclear concentration by biomolecular condensates.....	144
4.1 Introduction	144
4.2 Results.....	146
4.2.1 PAX6 forms condensates in a PD dependent manner	146
4.2.2 PAX6 foci are dynamic and can be dissolved by 1-6 hexanediol.....	152
4.2.3 PAX6 forms condensates <i>in vitro</i>	159
4.3 Discussion	163
Chapter 5 Missense mutations alter the interactions of PAX6 and SOX2..	166
5.1 Introduction	166
5.2 Results.....	168
5.2.1 Investigating the effect of fluorophores on FCS measurements.....	168
5.2.2 PAX6 missense variants influence homo-complex formation efficiency	172
5.2.3 The trans-interaction of PD with HD is altered in the presence of PAX6 mutants	177
5.2.4 PAX6 and SOX2 mutations reduce heterocomplex formation efficiency	183
5.2.5 SOX2 influences PAX6-chromatin binding kinetics in a non-reciprocal fashion	Error! Bookmark not defined.
5.2.6 A reporter assay investigating PAX6-SOX2 stoichiometry on transcription.....	191

5.3 Discussion	200
Chapter 6 Summary and Conclusions	203
6.1 Linking the molecular diffusion of PAX6 and SOX2 mutants to pathogenic phenotype in humans	203
6.1.1 PAX6 mutant variants exhibit weaker chromatin binding dynamics and altered subnuclear localisation	203
6.1.2 SOX2 mutant variants exhibit weaker chromatin binding dynamics and altered subnuclear localisation	204
6.2 PAX6 forms biomolecular condensates via a PD dependant mechanism	204
6.3 PAX6 missense variants influence homo-complex formation efficiency.....	205
6.4 PD-HD trans-interaction is altered in the presence of PAX6 mutants.....	205
6.5 PAX6 and SOX2 mutations reduce heterocomplex formation efficiency.....	206
6.6 SOX2 influences PAX6-chromatin binding kinetics in a non-reciprocal fashion	206
Chapter 7 Bibliography	208

List of Figures

Figure 1-1: Schematic representation of early ocular morphogenesis in humans	23
Figure 1-2: Schematic illustration of the human eye	24
Figure 1-3: Schematic illustration of the human cornea	26
Figure 1-4: Schematic illustration of the retina	28
Figure 1-5: Map of the human <i>PAX6</i> locus	30
Figure 1-6: Schematic representation of the three <i>PAX6</i> human isoforms ..	31
Figure 1-7: Schematic representation of <i>PAX6</i> -DNA complex.....	31
Figure 1-8: Expression pattern and function of <i>PAX6</i> in the murine brain....	36
Figure 1-9: Transverse section through the developing forebrain.....	Error!
	Bookmark not defined.
Figure 1-10: <i>SOX2</i> and <i>PAX6</i> expression in a E12.5 mouse eye	Error!
	Bookmark not defined.
Figure 1-11: Schematic representation of auto- and cross- regulation of <i>PAX6</i> and <i>SOX2</i> during optic field induction and development.	Error! Bookmark not defined.
Figure 1-12: Schematic representation of how post translational modifications of Sox proteins may influence their function	48
Figure 1-14: Schematic representation of the mechanism through which <i>SOX2</i> cooperates with <i>linc1614</i> to recruit the <i>PRC2</i> complex.....	49
Figure 1-15: A schematic illustration summarising the differentiation outcomes upon alteration of <i>SOX2</i> , <i>NANOG</i> and <i>OCT4</i> protein levels in embryonic stem cells.....	51
Figure 1-16: An illustration of the retinal anatomy including the major retinal cell types and their organization in the retina	55
Figure 1-17: Quantitative mechanisms that affect TF function	59

Figure 1-18: Protein Phase Separation Characteristics	62
Figure 1-19: Super enhancer phase separation schematic.....	64
Figure 1-20: Schematic representation of FCS	67
Figure 1-21: Schematic representation of FCCS	69
Figure 1-22: Schematic representation of SPT	70
Figure 1-23: Schematic representation of FRAP	71
Figure 1-24: Schematic representation of FRET	72
Figure 1-25: Schematic representation of STORM	73
Figure 1-26: A representation of a typical FCS setup in combination with a confocal microscope	75
Figure 1-27: Schematic illustration of FCS imaging system and its detection volume	76
Figure 2-1: Schematic of PCR amplicons and restriction enzymes used for generation of final constructs	94
Figure 2-2: Illustration of pEF1a-IRES-Neo plasmid map	94
Figure 2-3: Maps of the plasmids used for generating the stable GFP-PAX6 TET HEK293 cell line	100
Figure 2-4: Representation of key parameters during measurement analysis using the SymphoTime 64 software A	103
Figure 2-5: Schematic representation of FCS and FCCS	104
Figure 2-6: Schematic of the experimental setup for the <i>in vitro</i> experiments of recombinant PAX6	106
Figure 3-1: Schematic representation of the selected PAX6 variants on the PD in relation to DNA	110
Figure 3-2: Comparison of PD(WT) when labelled at either N or C terminus with mCitrine.	116

Figure 3-3: Comparison of PD(S54R) when labelled at either N or C terminus with mCitrine	118
Figure 3-4: Comparison of PD(R92Q) when labelled at either N or C terminus with mCitrine	120
Figure 3-5: Comparison of PD(S121L) when labelled at either N or C terminus with mCitrine	122
Figure 3-6: Comparison of PD(N124K) when labelled at either N or C terminus with mCitrine	124
Figure 3-7 Comparison of PD(C52R) when labelled at either N or C terminus with mCitrine	126
Figure 3-8: Comparison of PD-mCit between WT and missense variants .	129
Figure 3-9: Comparison of mCit-PD between WT and missense variants .	131
Figure 3-10: Comparison of SOX2-mCherry between WT and missense variants	138
Figure 3-11: Comparison of SOX2 protein localisation with DNA	139
Figure 4-1: Graphical representation of PONDR score analysis of the WT and mutant PAX6 PD	147
Figure 4-2: Analysis of the WT and mutant PAX6 PD sequences using CIDER	148
Figure 4-3: Comparison of PAX6 isoforms.....	151
Figure 4-4: HEK293 cells expressing mCitrine-PAX6 24 hours post-transfection, treated with 1M arginine	152
Figure 4-5: HEK293 cells expressing mCitrine-PAX6 24 hours post-transfection, treated with 20mM mitoxantrone	153
Figure 4-6: HEK293 cells expressing mCitrine-PAX6 24 hours post-transfection treated with 10% 1,6-hexanediol	153
Figure 4-7: Stable HEK293 cell line expressing GFP-PAX6 24 hours post tetracycline treatment.....	154

Figure 4-8: HEK293 cells expressing mCitrine-PAX6 fixed and stained for RNA PolII (Ser5).....	156
Figure 4-9: HEK293 cells expressing mCitrine-PAX6 fixed and stained for HP1 α	157
Figure 4-10: In vitro condensate experiments of recombinant PAX6 covalently labelled with Alexa 488	159
Figure 4-11: In vitro experiments of PD-Alexa 488	162
Figure 4-12: In vitro experiment of PD-Alexa 488 condensate recovery after FRAP	Error! Bookmark not defined.
Figure 5-1: Comparison of PD(WT) when labelled at the N terminus with either mCitrine or mCherry.....	170
Figure 5-2: Comparison of PAX6(WT) when labelled at the N terminus with either mCitrine or mCherry.....	171
Figure 5-3: Confocal images (40x lens) of HEK293 cells co-expressing WT and/or mutant mCitrine-PAX6 and mCherry-PAX6	174
Figure 5-4: Cross-correlation curves compared against the negative control (GFP + mCherry-PAX6	175
Figure 5-5: Schematic representation of PAX6 trans-interactions.....	176
Figure 5-6: Confocal images (40x lens) of HEK293 cells co-expressing mCherry-PAX6(δ PD) and WT or mutant mCitrine-PD	179
Figure 5-7: Cross-correlation curve comparison between PAX6(δ PD) and PD	180
Figure 5-8: Mean mCitrine-PD autocorrelation curve comparison in the absence or presence of PAX6(δ PD).....	181
Figure 5-9: Schematic representation of PD-HD trans-interactions	183
Figure 5-10: FCCS analysis of PAX6 and SOX2.	186
Figure 5-11: Co-IP analysis of the PAX6 and SOX2 interaction in HEK293 cells	187

Figure 5-12: Autocorrelation curves of SOX2 in the absense or presense of PAX6.....	188
Figure 5-13: Schematic representation of PAX6-SOX2 heterocomplex formation on chromatin	190
Figure 5-14: Autocorrelation curve of PAX6 in the absense or presense of SOX2	189
Figure 5-15: Illustration of DC5 reporter cassette	192
Figure 5-16: HELA cells co-expressing mCitrine-PAX6 and SOX2-mCherry and mCerulean-NES from the transfected DC5 reporter plasmid	194
Figure 5-17: Endogenously labelled SOX2-mCherry in mESCs.....	195
Figure 5-18: Comparison of SOX2-mCherry (human) in HEK293 versus endogenously labelled SOX2-mCherry in mESC.....	197
Figure 5-19: FCS experiments in endogenously labelled Pax6 in 12 hpf zebrafish embryo and mouse optic cup organoid.....	199

List of Tables

Table 1-1: Human Pax gene family	34
Table 2-1: mESC maintenance media	82
Table 2-2: The amount of each Effectene reagent which was used for cell transfection in a 6 well plate.....	84
Table 2-3: Antibodies used for immunofluorescent experiments	85
Table 2-4: Amount of each Effectene reagent used for cell transfection in a 10cm dish	86
Table 2-5: Adenine overhang mixture	89
Table 2-6: Reaction mixture and PCR conditions for site directed mutagenesis of single nucleotides	91
Table 2-7: PCR conditions for site directed mutagenesis for addition or deletion up to 40bp	92
Table 2-8: Table containing primers used for cDNA amplification.....	93
Table 2-9: Table containing primers used for site directed mutagenesis for the generation of the PAX6 and SOX2 variants.....	95
Table 2-10: Table containing the primers used for generating the PAX6 isoforms	96
Table 2-11: PCR mixture for site-directed insertion of the 5 α sequence	96
Table 2-12: PCR parameters for the site-directed insertion of the 5 α sequence in PAX6	97
Table 2-13. Table containing the guide RNA primer sequence and the primers used for genotyping mESCs for mCherry knock-in at the C terminus of genomic SOX2	108
Table 3-1: Summary of phenotypes identified in patients with PAX6 missense mutations	111
Table 3-2: Summary of phenotypes identified in patients with PAX6 missense mutations	113

Table 3-3: Table summarising the comparison of the molecular behaviour of PAX PD variants when tagging them with mCitrine at the N- or C- terminus. 127

Table 3-4: A summary of differences between the PAX6 PD WT versus the mutants with N-or C-terminally tagged mCitrine..... 132

Table 3-5: In silico analysis of the pathogenic *PAX6* variants using PolyPhen-2 and SIFT softwares..... 132

Table 3-6: In silico analysis of the pathogenic *SOX2* variants using PolyPhen-2 and SIFT softwares..... 137

List of Abbreviations

TF	transcription factor
ChIP-seq	chromatin immunoprecipitation sequencing
Co-IP	co-immunoprecipitation
FCS	fluorescence correlation spectroscopy
FCCS	fluorescence cross-correlation spectroscopy
HMG	high-mobility group
CNS	central nervous system
PD	PAIRED
HD	homeodomain
PST	proline-serine-threonine
lncRNA	long non-coding RNA
DBD	DNA-binding domain
PE3	pancreas cis-regulatory element 3
PE4	pancreas cis-regulatory element 4
ESCs	embryonic stem cells
PSPB	pallial-subpallial boundary
NR	neural retina
CE	ciliary body epithelium
Rax	Retinal homeobox
OTX2	Orthodenticle homeobox 2
SOX2	SRY-box 2
DC5	δ -crystallin 5
ADHD	attention deficit hyperactivity disorders
MAC	microphthalmia, anophthalmia, coloboma

SEE	super-enhancer element
SOX2OT	SOX2 overlapping transcript
NLS	nuclear localisation signals
IMP α	importin- α
PTMs	post-translational modifications
CDK	cyclin-dependent kinase
lincRNAs	large-intergenic noncoding RNAs
PRC2	polycomb repressive complex 2
RMST	rhabdomyosarcoma 2-associated transcript
NPC	neural progenitor cells
SVZ	subventricular zone
RPC	retinal progenitor cells
P-TEFb	positive transcription elongation factor
SPT	single-particle tracking
APD	avalanche photodetector
EMSA	electrophoretic mobility shift assays
PBS	phosphate-buffered saline
O/N	overnight
κ	confocal detection volume
cpm	counts per molecule
gRNA	guideRNA
cpms	counts per molecule second
GFP	green fluorescent protein
LLPS	liquid-liquid phase-separation
IDRs	intrinsically disordered regions

FCS	Fetal Calf Serum
PONDR	Predictor of Natural Disordered Regions
HP1	heterochromatin protein 1
CPC	chromosome passenger complex
FRAP	fluorescence recovery after photobleaching

Chapter 1 Introduction

1.1 The Evolution of Vision

The eye is one of the most fascinating organs whose study has been a main theme in evolutionary biology. Charles Darwin was puzzled how an eagle's eye with all its perfection could have simply been derived by random variation and selection (*The Origin of Species*). This was followed up by a proposition that complex, highly specified eyes such as the eagle's, originated from a simple prototypic organ consisting of an optic nerve (photoreceptor cell) surrounded by pigment cells and covered by translucent skin (not including a lens or any refractive body). Such a prototypic eye would enable determining light direction and hence provide a selective advantage over organisms which could only distinguish between light and dark. The development and centralisation of the central nervous system, allowing for the perception of light, provided directional movement and as a result an increased avoidance of predators which was an enormous advantage in terms of survival. At the time, Darwin was not aware that such basic ancient organs are still present in nature - the planarian *Polycelis auricularia* has multiple eyes in the head region, which consist of one photoreceptor and one pigment cell (Gehring, 2005). An alternate theory assumed that the various eye-types found in different phyla arose independently, judging from the sheer differences in morphology, physiology and morphogenesis. However, genetic experiments indicate that the various metazoan eye types are controlled by the same set of transcription factors (TF). In particular, *PAX6* serves as a master TF for eye morphogenesis in taxa as different as insects and mammals (Halder, Callaerts and Gehring, 1989; Chow *et al.*, 1999; Xu *et al.*, 1999), suggesting that *PAX6* was already functioning as a master TF in the last common ancestor of arthropods and chordates. *PAX6* has been found to be associated with eye development in all creatures with two eyes studied so far, strongly supporting the hypothesis of a monophyletic origin of the bilaterian eye.

1.2 Human eye development

Eye development in humans begins at approximately three weeks into gestation and is not completed until several months after birth. It is a tightly controlled developmental process that is regulated by spatiotemporal gene expression patterns and physical interactions. The eye is initially derived from the neuroectoderm, the surface ectoderm and mesenchyme. The neuroectoderm gives rise to the neural retina (NR), retinal pigmented epithelium (RPE), optic nerve, iris dilator and ciliary body. The surface ectoderm contributes to lens, conjunctival and cornea epithelia formation. The mesenchyme, which is derived from the mesoderm and neural crest cells, forms the corneal epithelium, stroma, sclera and vasculature.

Eye development begins by the formation of a pair of optic vesicles from the anterior neural plate lateral of the forebrain (Eintracht, Toms and Moosajee, 2020), which is characterised by expression of early eye field TFs such as *PAX6*, *OTX2*, *SIX3*, *SIX6* and *RAX* (Chen *et al.*, 2017). Sonic Hedgehog (*SHH*) and Fibroblast Growth Factors (*FGF*) signalling pathways which are regulated by Transforming Growth Factor (*TGF*) signalling, initiate eye field splitting and the subsequent posterior to anterior migration of cells to drive optic vesicle evagination (Cardozo, Almuedo-Castillo and Bovolenta, 2020). Failure of the eye-field to split, results in single central eye formation known as cyclopia, which is associated with *SHH* mutations (Placzek and Briscoe, 2018).

Formation and growth of the optic vesicles form in the fourth week of embryonic development and are connected to the developing forebrain by the optic stalk which later develops into the optic nerve. Key TFs such as *PAX6*, *OTX2* and *SOX2* are expressed in the optic vesicle and are required for cell fate-determining pathways (Matsushima *et al.*, 2011). Towards the end of the fourth week of gestation, the optic vesicle physically contacts the overlying surface ectoderm, inducing a cascade of transcriptional and ultimately physical changes (invagination), which are essential for further eye development. Optic vesicle and surface ectoderm invagination form the optic cup and lens pit respectively. The invagination of the optic vesicle to form the bi-layered optic

cup is stimulated by Bone Morphogenic Protein 4 (*BMP4*) and retinoic acid (Harding and Moosajee, 2019). The inner layer of the optic cup gives rise to the neural retina (NR) and the outer layer becomes the retinal pigmented epithelium (RPE). The interface of these two domains, the peripheral optic cup margin, gives rise to the non-pigmented ciliary body epithelium and the inner iris epithelium. Progenitor cells at the boundary between the potential NR and ciliary body epithelium (CE) undergo differentiation into either neurogenic or non-neurogenic cell fates (Matsushima *et al.*, 2011). Several TFs have been associated with influencing this cell fate decision, such as *Rax*, *OTX2*, *SOX2* and *PAX6* (Hever, Williamson and van Heyningen, 2006; Matsushima *et al.*, 2011; Chen *et al.*, 2017).

During the fifth week of ocular development, the optic fissure develops along the ventral surface of the optic cup extending to the optic stalk (Plaisancié *et al.*, 2019). This morphological change allows for the vasculature to enter and supply the developing eye (Wang *et al.*, 2019). Following optic fissure fusion, the formation of the major eye structures is largely complete. Subsequently, this is followed by overall ocular tissue maturation. At around the sixth week mark, physical interactions between the lens and optic cup induce the formation of the cornea from the surface ectoderm. During the seventh week of embryonic development, periocular mesenchymal and neural crest cells migrate into the space between the surface ectoderm and lens vesicle in three distinct waves to form the corneal stroma, epithelium and endothelium. In addition, the third wave of mesenchymal cell migration also contributes to iris formation (Eghrari, Riazuddin and Gottsch, 2015).

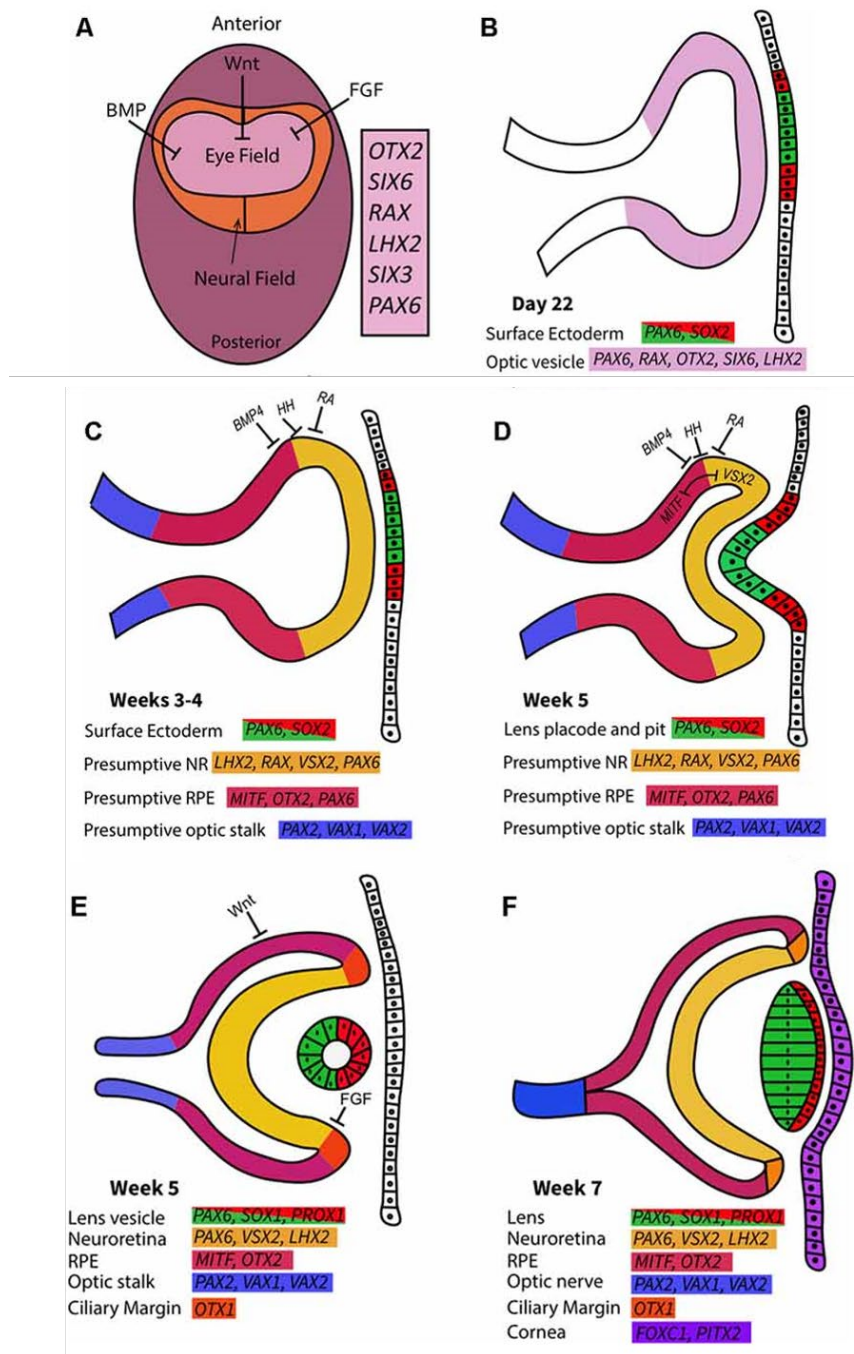


Figure 1-1: Schematic representation of early ocular morphogenesis in humans. **A)** The specified region known as the ‘eye-field’ is created by the upregulation of eye-field TFs in the anterior neural plate. Upregulation of these TFs is driven by developmental pathways such as Wnt, BMP and fibroblast growth factor (FGF). **B)** Three weeks post-conception, evagination of the optic vesicle occurs whereby all eye-field TFs are ubiquitously expressed **C)** During the third and fourth week of gestation, the combinatorial action of signalling pathways determines presumptive regions in the optic vesicle. **D)** Physical interactions between the optic vesicle, surface ectoderm and extraocular mesenchyme cause the invagination of the optic cup at around five weeks. **E)** Following optic cup formation, Wnt and FGF pathways drive neural retina (NR) differentiation and define these region through ciliary margin formation. Simultaneously, as the lens vesicle forms, it detaches from the surface ectoderm. **F)** By the seventh week, the lens is formed from the lens vesicle and the cornea is formed from the overlying surface ectoderm. The NR and RPE are clearly defined, while the optic nerve forms from the convergence of the optic stalk. Adapted from *Eintracht, Toms and Moosajee, 2020*

1.3 Anatomy of the developed human eye

The eyes are one of the most complex organs in the human body. The developed human eye, consists of three distinguishable layers which are described below.

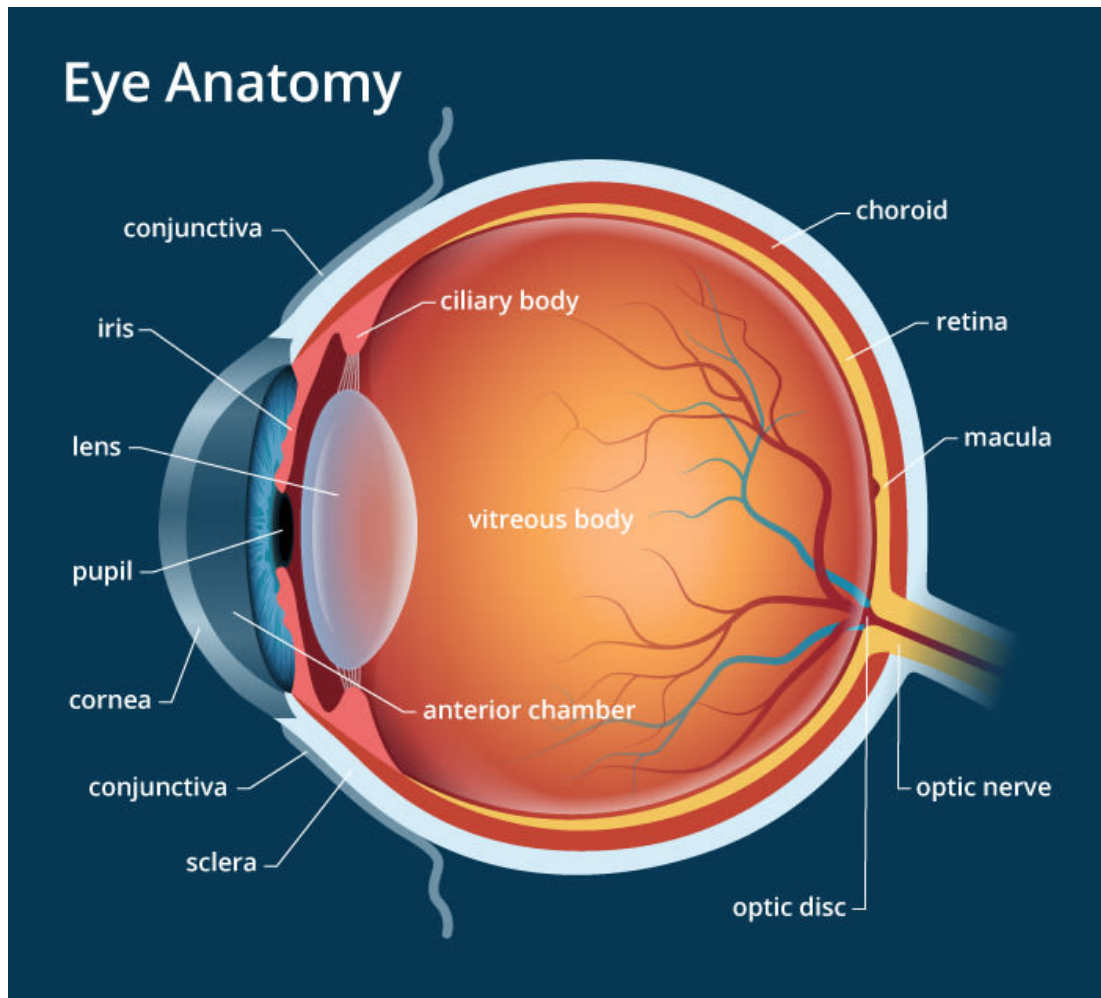


Figure 1-2: Schematic illustration of the human eye. Light is refracted and focused by the cornea. The iris is able to adjust the size of the pupil which regulates the amount of light that passes through the lens. The focused light reaches the retina which converts the light into electrical signals. These signals are transmitted via the optic nerve to the visual cortex of the brain. Adapted from Gary Heiting, OD www.allaboutvision.com

1.3.1 Outer Layer

The outer region of the eye consists of the cornea and the sclera. The cornea is a transparent dome-shaped structure that covers the front of the eye and accounts for nearly two-thirds of the total refractive power of the eye. Even though the cornea is an avascular structure except at its margins, it is highly innervated making it very sensitive to pain and touch. The cornea consists of six layers; epithelium, Bowman's membrane, stroma, Dua's layer, Descemet's membrane and endothelium (DeMonte and Kim, 2011).

The epithelium is the outmost layer of the cornea and plays a crucial role in maintaining a smooth refractive surface and acts as a mechanical barrier to all external pathogens. Bowman's Layer is situated between the epithelium and the stroma and is mainly composed of randomly arranged collagen fibers (type 1 and type 3). The layer helps maintain the concave shape of the cornea and is resistant to trauma (Wilson, 2020). The stroma contributes to almost 90% of corneal thickness and is made of collagen fibers (mainly type 1) arranged in parallel which run from limbus to limbus. Dua's Layer is a tough, acellular layer composed of tightly packed collagen bundles which run in longitudinal, transverse and oblique directions. Descemet's Membrane is secreted by the endothelial cells and is primarily composed of collagen types 4, 8 and laminin. It is a tough and elastic layer that is strongly attached to the post surface of the stroma. The endothelium is a single layer of hexagonal cells that maintains corneal transparency by keeping the stroma dehydrated (DeMonte and Kim, 2011).

The sclera is the white part of the eye and is a tough, non-transparent tissue that shapes, supports and protects the eye and prevents light from reaching the retina other than through the cornea. It wraps around the majority of the eye extending from the cornea in the front to the optic nerve in the back (Downie *et al.*, 2021).

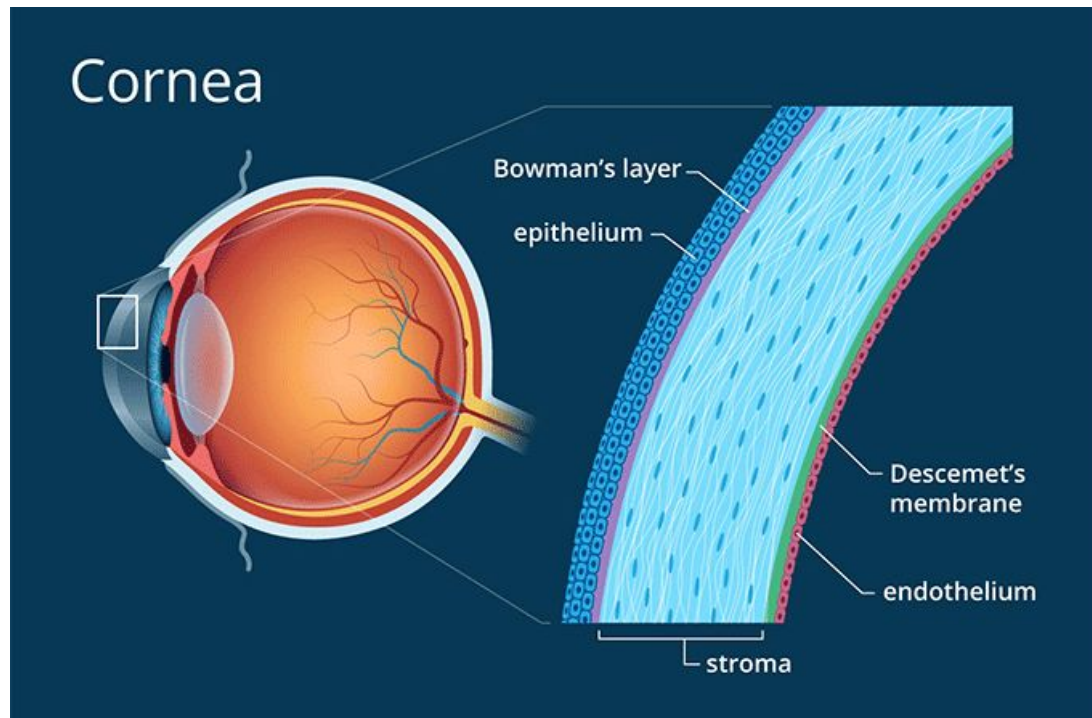


Figure 1-3: Schematic illustration of the human cornea. The cornea consists of six layers. These are the epithelium, Bowman's layer, stroma, Descemet's membrane and endothelium. Adapted from Gary Heiting, OD www.allaboutvision.com

1.3.2 Middle Layer

The middle layer of the eye also known as uvea is composed of the iris, the ciliary body and the choroid body. The iris is the coloured part of the eye and is located between the cornea and the lens. It controls the size of the pupil and therefore the amount of light passing through the lens. The lens is a transparent biconvex structure which along with the cornea, helps to refract and focus light on the retina. The ciliary body controls the power and shape of the lens and is the site of aqueous production. Aqueous humor is found in the anterior and posterior chambers of the eye. The anterior chamber is located between the cornea and the iris and the posterior chamber is found between the iris and the front of the lens (Downie *et al.*, 2021). The aqueous humor provides nourishment to surrounding eye structures and removes waste from the eye. In addition, it is responsible for maintaining balanced pressure within

the eye. The vitreous body is a transparent, colourless, gelatinous mass that fills the space between the retina and the lens. It is surrounded by a layer of collagen called the vitreous membrane and makes up about 80% of the eyeball volume. The choroid body is located between the sclera and the retina and is made up almost entirely of blood vessels, which provide nutrients and oxygen to the surrounding tissues, making it invaluable for eye function (Nickla and Wallman, 2010).

1.3.3 Inner Layer

The inner layer of the eye consists of the retina which is a complex, layered structure of neurons that captures and processes light. The retina lines the inner surface of the eye, surrounding the vitreous cavity. It is protected and held in the appropriate position by the surrounding sclera. The major cellular components of the retina include the retinal pigmented epithelium (RPE), the photoreceptor cells, bipolar cells, horizontal cells, amacrine cells, ganglion cells and Müller glia cells (Hoon *et al.*, 2014).

The RPE is a single layer of hexanocuboidal cells, which form the outer blood-retina barrier. It regulates the transport of nutrients and waste products to and from the retina and is essential for the maintenance and survival of the overlying photoreceptor cells (Strauss, 2005).

The photoreceptor cells are a specialised type of neuroepithelial cell which are capable of visual phototransduction (conversion of light into electrical signal). There are three known types of photoreceptor cells. These are the rods, the cones and the photosensitive retinal ganglion cells. The rods primarily contribute to night-time vision (scotopic conditions) whereas cones primarily contribute to colour and day-time vision (photopic conditions). The photosensitive retinal ganglion cells do not directly contribute to sight but have been shown to have a role in circadian rhythm and pupillary reflex (Berson, 2007).

The density of rods and cones varies between different regions of the retina. In humans, about 50% of cones are found in the central part of the retina,

roughly corresponding with the macula. The fovea in the macula, which is a central space in the retina, has the highest concentration of cones but has no rods. This area of the retina is responsible for sharp central vision, with visual acuity been the highest in the foveola, the thin bottom of the fovea. Bipolar, horizontal and amacrine cells are found in the inner nuclear retina layer and form complex neuroretinal circuitries that process the photoreceptor signal and transmit the signal to the ganglion cell layer (Hoon *et al.*, 2014). Ganglion cells are responsible for transmitting visual information from the retina to the brain via the optic nerve.

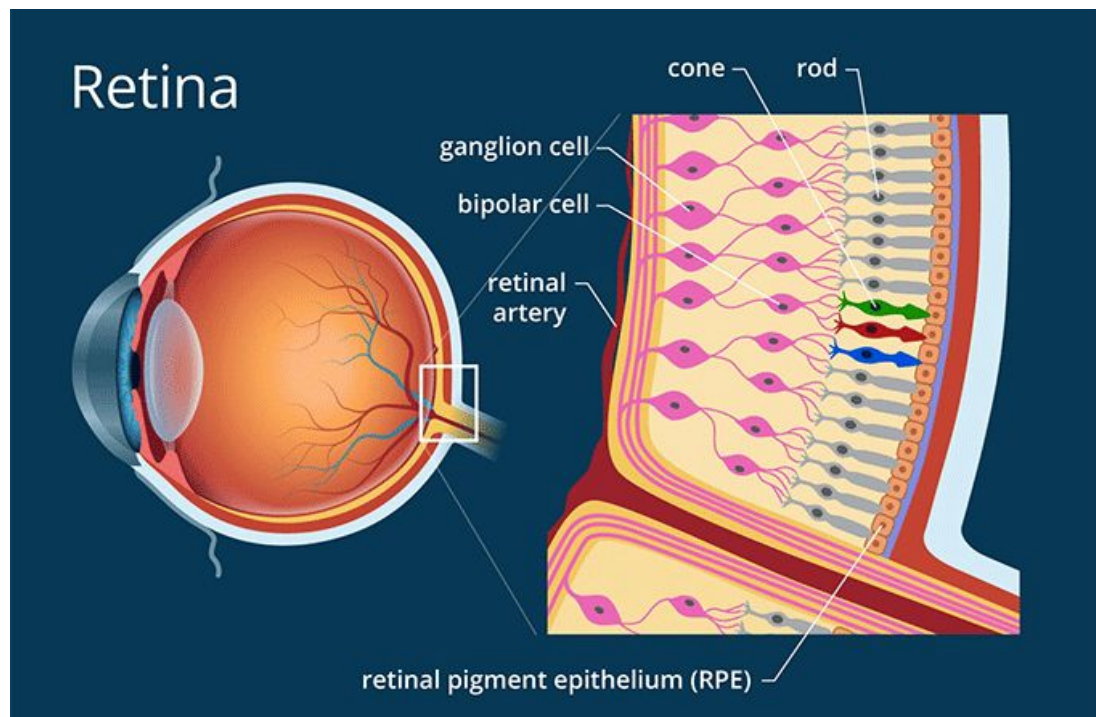


Figure 1-4: Schematic illustration of the retina. The retina is responsible for capturing and processing light. The major components of the retina are the retinal pigmented epithelium, rods, cones, bipolar cells and ganglion cells. Adapted from Gary Heiting, OD www.allaboutvision.com

1.4 PAX6 Gene and Protein Architecture

1.4.1 PAX6 Gene Architecture

The *PAX6* gene was first identified as a member of the paired box family in mice (Walther *et al.*, 1991). Heterozygous loss of function *Pax6* mutant mice and rats, develop small eyes with a distinct small eye syndrome when compared to wild type animals (Quinn, West and Hill, 1996). Heterozygous *PAX6* mutations likewise cause aniridia in humans (Tzoulaki, White and Hanson, 2005; Hever, Williamson and van Heyningen, 2006; Hingorani *et al.*, 2009). Aniridia is a rare genetic disorder characterized by total or partial absence of the iris, associated with visual impairment which deteriorates with age (Landsend, Lagali and Utheim, 2021). Homozygous *PAX6* mutant mouse embryos, completely lack eye and nose development and die soon after birth (Hanson *et al.*, 1994). Initial experiments to understanding *PAX6* function were conducted in *Drosophila melanogaster*. Interestingly a link between mice and *Drosophila* was established when ectopic expression of mouse Pax6 protein (100% amino acid conservation to human *PAX6*) in *Drosophila* triggered development of morphologically normal ectopic eyes (Halder, Callaerts and Gehring, 1995). Additional experiments demonstrated that eye development was rescued in *eyeless* (the fly *PAX6* orthologue) homozygous mutant flies when the full length protein was expressed (Clements *et al.*, 2009) and that ectopic *PAX6* expression in *Drosophila* lead to morphologically normal eye development in the legs, antennae and wings (Halder, Callaerts and Gehring, 1989). Ultimately, these experiments were able to demonstrate the requirement, sufficiency and functional conservation of *PAX6* in eye development.

Human *PAX6* (OMIM 607108) is found in chromosome 11 and consists of 14 exons. *PAX6* coding exons span approximately 20kb of genomic DNA whereas the *PAX6* cis-regulatory regions can span 450kb (Figure 1-5) (Bhatia *et al.*, 2014; Cunha *et al.*, 2019). As well as a multitude of tissue specific enhancers, the long non-coding RNAs (lncRNA) *Paupar* and *PAX6OS1* have also been identified at the *PAX6* locus.

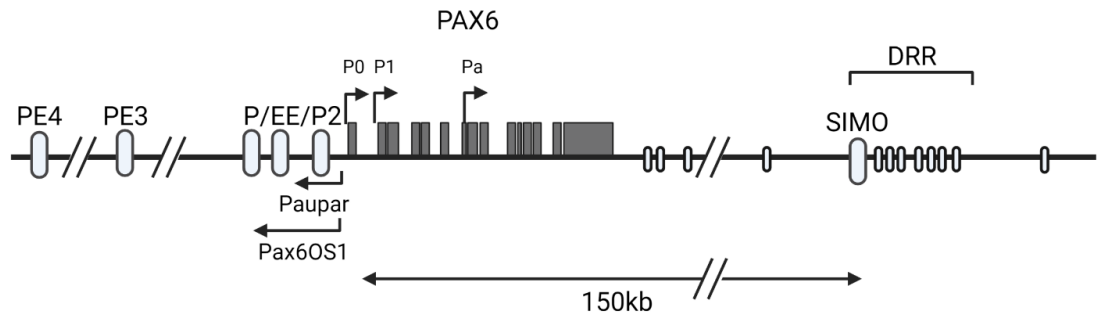


Figure 1-5: Map of the human *PAX6* locus (exons marked by grey boxes), *PAX6* promoters (P0, P1, Pa) and nearby transcriptional units (*Paupar* and *Pax6OS1*). Blue ellipsoids represent cis-regulatory elements, including the lens-specific ectodermal enhancer (EE), SMO and the distal regulatory region (DRR). DRR is located 150kb downstream from the P1 promoter and is a collection of regulatory elements.

Paupar is expressed in the brain, is located about 25kb from the P0 promoter, and is reported to regulate the expression of *PAX6* as well as that of other genes located in distinct chromosomes such as the histone methyl-transferase NSD1 (chromosome 5) which plays a critical role in epigenetic regulation of hESC cortical differentiation (Xu *et al.*, 2021). Knockdown of *Paupar* disrupts the cell cycle of neuroblastoma cells and induces neural differentiation, in part through physical association with the *PAX6* protein (Vance *et al.*, 2014). Although its biological significance remains ill-defined, *PAX6OS1* has recently been linked with pancreatic β -cell function whereby increased expression of *PAX6OS1* at high glucose levels, contributes to β -cell dedifferentiation and failure in some forms of type 2 diabetes (Lopez-noriega *et al.*, 2020).

1.4.2 *PAX6* Protein Architecture

Three distinct *PAX6* promoters have been identified, termed P0, P1 and Pa. Each promoter has been linked to each of the three *PAX6* isoforms; P0 for the canonical *PAX6*; P1 for *PAX6* (5 α) and Pa for *PAX6* (δ PD). The human *PAX6* gene can transcribe three protein isoforms. These are the canonical *PAX6* (422 amino acids), the *PAX6*(5 α) (436 amino acids) and the *PAX6*(δ PD) (280 amino acids) (Figure 1-6) (Lakowski, Majumder and Lauderdale, 2007).

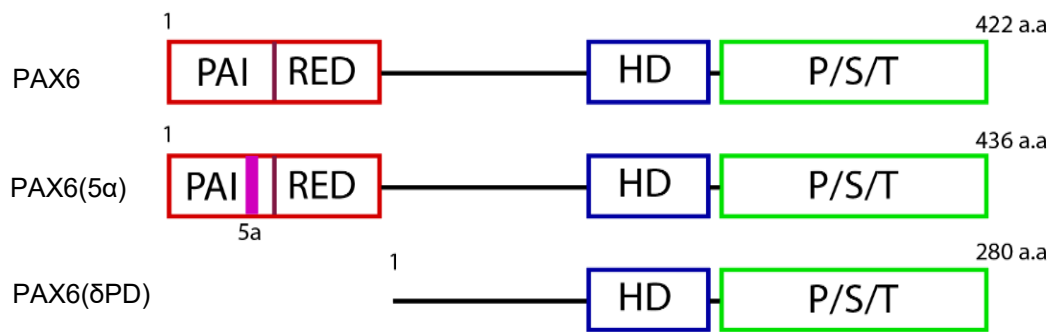


Figure 1-6: Schematic representation of the three PAX6 human isoforms. HD: homeodomain P/S/T: proline serine threonine transactivation domain

TFs are grouped into five classes based on their DNA-binding domain (DBD) family. These are helix-turn-helix, basic helix-loop-helix, basic leucine zipper, zinc fingers and nuclear hormone (steroid) receptors. The canonical PAX6 protein contains two helix-turn-helix DNA binding domains; a PAIRED domain (PD) and a homeodomain (HD). The

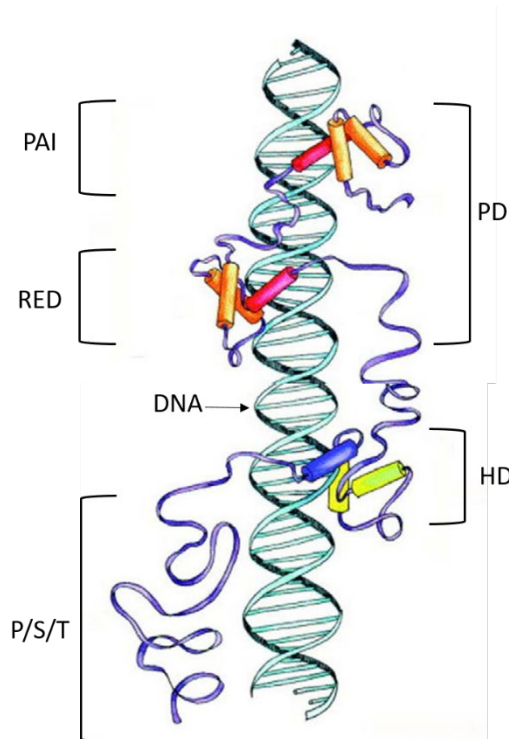


Figure 1-7: Schematic representation of PAX6-DNA complex. Cylinders represent α -helices. Blue lines represent linker regions and the P/S/T transactivation domain which are primarily disordered. Schematic adapted from Blake and Ziman, 2014

PD is composed of two subdomains called PAI (N-terminus; residues 1-60) and RED (C-terminus; residues 77-133) which are separated by a small linker which binds to the minor groove near the centre of the site. The linker sequence makes multiple contacts with the sugar phosphate backbone and DNA bases over an 8-bp region. Each subdomain contains three α -helices that fold like a homeodomain. The PAI subdomain in addition, has an N-terminal β -sheet (Figure 1-7)(Epstein *et al.*, 1994) that spans the minor groove of the DNA and contacts the sugar phosphate backbone of both DNA strands (Eric Xu *et al.*, 1999). The

PAI subdomain is highly conserved among PAX proteins and uses a helix-turn-helix unit to dock against the major groove at one end of the binding site. The RED subdomain also uses a helix-turn-helix unit to dock against the major groove in the distal portion of the Pax6 binding site, whereby helix 6 (the 'recognition helix' of the RED subdomain) fits directly into the major groove. This interaction is stabilised by the phosphate contacts from the amino-terminal portion of helix 5 and from the carboxy-terminal portion of the linker (Eric Xu *et al.*, 1999). Both the PD and HD bind to specific DNA target sites and influence each other's binding to DNA (Mishra, Ivan P. Gorlov, *et al.*, 2002; Haubst *et al.*, 2004). The two DBDs are linked by a glycine rich linker and the C-terminus of the protein consists of a proline-serine-threonine (PST) rich transactivation domain. PAX6 nuclear localization signals are located in the PD and N-terminal region of the HD (Blake and Ziman, 2014).

PAX6 (5a) contains an alternative splice variant in comparison to the canonical isoform, whereby a 14 amino acid sequence is incorporated in the PAI subdomain of the PD. This disrupts the PAI subdomain resulting in altered DNA-binding properties of the DBD (Haubst *et al.*, 2004). This has been shown to differentially influence gene expression when compared to the canonical isoform (Walcher *et al.*, 2013). The stoichiometry between the canonical PAX6 protein and the 5a isoform has been shown to be critical in gene regulation and development. For example, the ratio between the two isoforms in embryonic mouse tissues is 8:1, which subsequently becomes 1:1 in adult ocular tissues such as the cornea, iris and lens (Pinson *et al.*, 2005). Altering the ratio of the two isoforms, leads to altered gene expression and regulation. Mouse experiments linked PAX6 (5a) function to promote the development of the neural retinal structure, whereby deletion of exon 5a disrupted predominantly iris formation (Singh *et al.*, 2002).

The third and least studied isoform, PAX6 (δ PD), completely lacks the PD. In the developing eyes of mice, PAX6 (δ PD) is present in a small cell population of GABAergic-only amacrine cells, whereby the stoichiometry between the canonical PAX6 and PAX6 (δ PD) is again believed to be crucial for both cornea

and lens development (Lakowski, Majumder and Lauderdale, 2007; Kim and Lauderdale, 2008).

PAX6 proteins bound to regulatory elements participate in a range of molecular interactions whose overall output is transcriptional activation or repression of target genes. *In vivo*, these interactions can act in combination or alone and vary from the dynamic interaction of PAX6 with DNA in the context of chromatin (Xie *et al.*, 2013), recruitment of transcriptional co-activators and co-repressors (Yang *et al.*, 2004; Elvenes *et al.*, 2011), indirect histone post-translational modifications (Sun *et al.*, 2016), heterocomplexes with other TFs (Kamachi *et al.*, 2001; Ninkovic *et al.*, 2013) and direct or indirect interactions with RNA Polymerase II and its subunits.

The DNA-binding sites of PAX6 proteins have been investigated since the early '90s, resulting in the identification of a variety of similar DNA-binding sequences. Using SELEX, a consensus binding site for the PAX6 PD was obtained, termed P6CON (Epstein *et al.*, 1994). The identified binding site was further investigated for other Pax family members, Pax2 and Pax5, leading to the identification of the two subdomains of the PD. Notably, the DNA-binding site of PAX6(5a) was found to be different from the PD consensus sequence, since the amino acid extension modified the folding of the PAI subdomain (Chauhan *et al.*, 2004b). Investigation of the interactions and cooperative binding of the PAX6 PD/HD and PD(5a)/HD, identified sites which are recognised by both HD and PAI subdomain but also established a novel optimal site for PAX6 (5a) (Xie and Cvekl, 2011; Xie *et al.*, 2014). Crystallographic studies of PAX6 PD and PAX6CON complexes have provided invaluable information on the docking structure and geometry of the two PD subdomains but more importantly have provided a structural framework for understanding the pathogenicity of missense mutations in human patients (Eric Xu *et al.*, 1999). The full length PAX6 protein though, is yet to be studied by crystallography, making our understanding of the complex incomplete.

1.5 PAX6 Expression and Function

Pax genes encode TFs which are predominantly involved in development and specification of the central nervous system and its organs in both vertebrates and invertebrates. In humans, there are 9 *Pax* genes, which are grouped into six different classes (Table 1-1) (Chi and Epstein, 2002). Classification is based on the presence of gene products containing the obligatory PD and complete or partial HD. Among all *Pax* genes, PAX6 is the first one to be expressed in the developing central nervous system, eye and olfactory primordia.

Gene	Structural domains				Target organs/tissues	Chromosome	Associated human disease
	PD	OP	HD1	HD2/3			
PAX1	X	X			Skeleton,thymus	20p11	Oligodontia
PAX2	X	X	X		CNS,kidney,eye	10q25	Renal-coloboma syndrome
PAX3	X	X	X	X	CNS,neural crest,skeletal muscle	2q35	Waardenburg syndrome, rhabdomyosarcoma
PAX4	X		X	X	Pancreas	7q	Type 2 diabetes mellitus
PAX5	X	X	X		CNS, B cells	9p13	Large cell lymphoma
PAX6	X		X	X	CNS,eye,pancreas	11p13	Aniridia
PAX7	X	X	X	X	CNS, cranio-facial,skeletal muscle	1p36.2	Rhabdomyosarcoma
PAX8	X	X	X		CNS,kidney,thyroid	2q12-q14	Thyroid dysplasia, thyroid follicular carcinoma
PAX9	X	X			Skeleton,cranio-facial,tooth	14q12-q13	Oligodontia

Table 1-1: Human Pax gene family. Structural domains include PAIRED domain (PD), octapeptide motif (OP), first helix of homeodomain (HD1) and the helix-turn-helix motif (HD2/3). X denotes presence of motif. Table adapted from Chi and Epstein, 2002

1.5.1 PAX6 in the Pancreas

PAX6 is required for correct development of the endocrine pancreas and specification of hormone producing endocrine cell types. *PAX6* is initially expressed in the early pancreatic bud and is necessary for insulin homeostasis in the adult pancreas. The core functional unit of the endocrine pancreas is the islet of Langerhans. Islets are found within the exocrine tissue of the pancreas and consist of four cell types: α -, β -, γ -, and δ - cells. Insulin is produced from β -cells, forming the core of the islet. α -, δ - and γ - cells are arranged in the periphery of the islets and produce glucagon, somatostatin and pancreatic

polypeptide respectively. *PAX6* is required for the correct development and spatial organisation of all islet cell precursors and subsequent differentiation of α -cells. *PAX6* expression in the pancreas is primarily driven by the pancreas and surface ectoderm enhancer (P/EE; Figure 1-6) cluster (Kammandel *et al.*, 1999). Heterozygous pathogenic mutations in human *PAX6* lead to impaired glucose tolerance and are associated with a diabetic phenotype and a decrease in insulin-producing cell number (Buckle *et al.*, 2018). Genome wide association studies (GWAS) have revealed that single nucleotide polymorphisms located in regulatory elements of pancreatic TFs are linked to diabetes, including those functionally linked to *PAX6*. For example, pancreas cis-regulatory element 3 (PE3) and PE4 (Figure 1-6) are located 50 and 100kb upstream and interact with different parts of the *PAX6* promoter and other non-coding RNAs (Buckle *et al.*, 2018). They have been shown to drive expression in the developing pancreas and interact with pancreas-related TFs such as *PAX4*. A common variant present in the PE4 affects *PAX4* binding, which is a known pancreatic regulator, leading to misregulation of *PAX6* gene expression and elevated risk of diabetes (Buckle *et al.*, 2018).

1.5.2 *PAX6* in the Brain

In mammals, *PAX6* specifies the neuroectodermal fate from pluripotent embryonic stem cells (ESCs) and produces new neurons from neural stem/progenitor cells (neurogenesis). Additionally, it is crucial for forebrain patterning and specifically, for telencephalic and diencephalic regionalization and in defining the boundary between adjacent embryonic brain domains (Haubst *et al.*, 2004). Murine *PAX6* expression starts as early as embryonic day 8 in the neuroepithelium, at the time when the neural tube is closing. As neuronal development progresses, *PAX6* is expressed in the developing brain from the mitotic germinal zone, and gradually expands in the embryonic forebrain, hindbrain midbrain and eventually adult brain (Figure 1-8) (Manuel *et al.*, 2015).

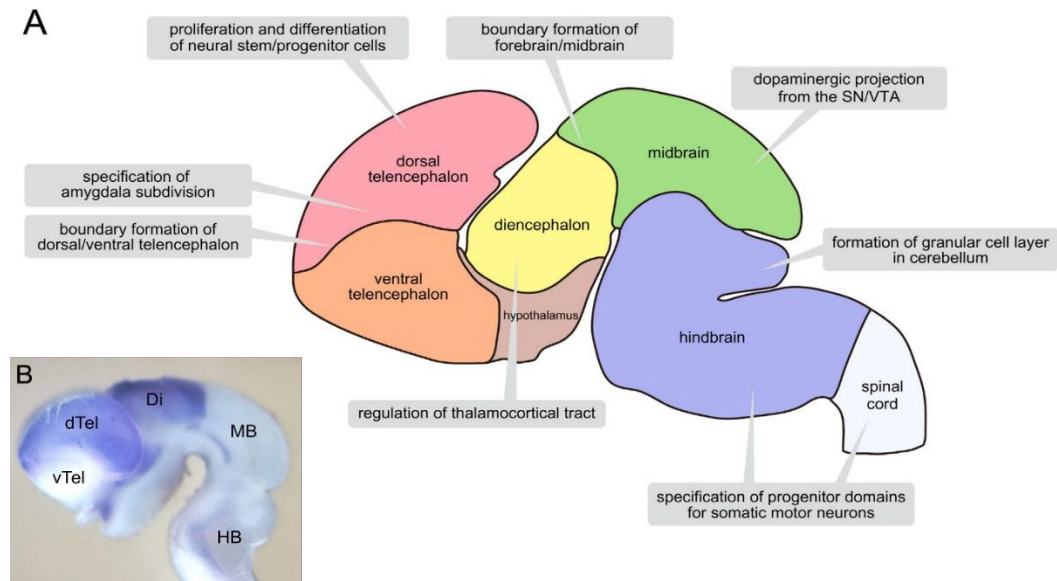


Figure 1-8: Expression pattern and function of PAX6 in the murine brain. A) Summary representation of PAX6 functions in specific brain areas during embryogenesis. B) Whole-mount in situ hybridization of an E13.5 rat neural tube. Di, diencephalon; dTel, dorsal telencephalon; HB, hindbrain; MB, midbrain; vTel, ventral telencephalon. Figure adapted from Kikkawa *et al.*, 2019

In the developing central nervous system, *PAX6* plays an instrumental role in multiple processes such as, patterning of the neural tube, cell proliferation, migration of neurons and formation of neural circuits (Kikkawa *et al.*, 2019). Homozygous loss of function *PAX6* mice, display severe defects in dorsal-ventral patterning of the telencephalon and die soon after birth. The boundary between dorsal and ventral telencephalon, known as pallial-subpallial boundary (PSPB), is severely impacted, by loss of *PAX6* (Stoykova *et al.*, 1996; Georgala, Carr and Price, 2011). Misregulation of adhesion molecules such as R-cadherin, further contribute in the loss of a mechanical palisade at the PSPB formed by radial glial cells (Price *et al.*, 1997). This palisade, acts as a physical barrier to prevent mixing of the dorsal and ventral telencephalic progenitors and in its absence, an increased dorsal migration of ventral telencephalic cells is observed. Loss of *PAX6* in mice has also been linked with thinner cortices due to *PAX6*'s function in increasing cortical progenitor proliferation in a dosage-dependent manner (Jones *et al.*, 2002). In addition, an increase in post-mitotic cortical progenitors at early stages leads to a premature depletion of the progenitor cell pool, shifting the balance between

proliferation and differentiation of the system towards the latter. Though studies of *PAX6* haploinsufficiency in humans have predominantly focused on eye malformations, some studies have identified additional brain malformations. These include the absence or hypoplasia of the anterior commissure and a reduced sense of smell (Njajou *et al.*, 2001). Even though the exact function of the anterior commissure remains elusive, it has been linked with colour perception and attention.

1.5.3 *PAX6* in the Eye

Even though *PAX6* is expressed in all tissues forming the eye (e.g. lens, retina, cornea and iris), its mode of action is highly diverse and complex. Prior studies have indicated that tissue-specific functions of *PAX6* arise from its interactions with other TFs, but also its ability to regulate expression of multiple TFs in a tissue-specific manner. For example, at the optic cup/optic stalk boundary, reciprocal inhibition between *PAX6* and *PAX2*, establishes proximal/distal patterning of the optic vesicle neuroepithelium. *PAX2* can repress *PAX6* expression by binding to the *PAX6* retina enhancer, whereas *PAX6* can repress *PAX2* by binding to its upstream control region (Schwarz *et al.*, 2000). Furthermore, *PAX6* is expressed throughout the boundary between NR and CE and acts as a regulator of boundary determination. *SOX2* is primarily expressed in the NR during this stage and antagonizes *PAX6* in boundary determination (Figure 1-9) (Matsushima *et al.*, 2011).

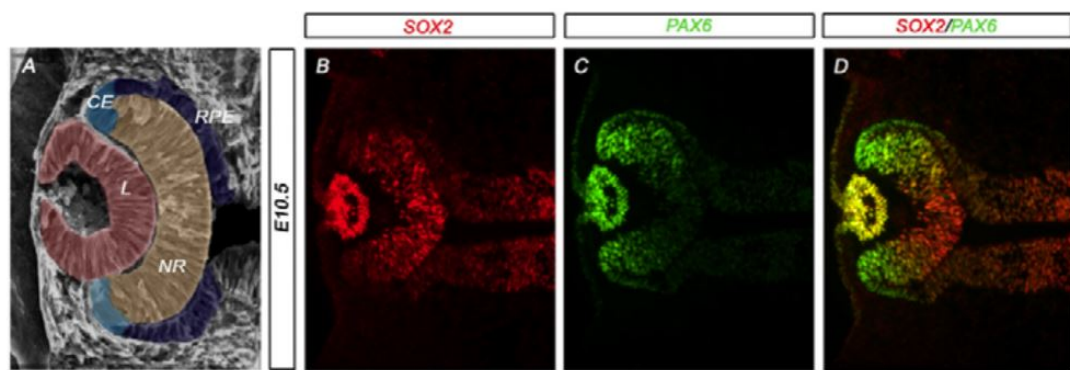


Figure 1-9: The neural retina and optic cup margin are defined by an inverse gradient of *SOX2* and *PAX6*. A) A schematic of an E12.5 eye showing the boundaries of the mouse neural retina (NR), lens (L), prospective ciliary body epithelium (CE) and retinal pigment epithelium (RPE). B-D) Immunofluorescence on horizontal sections of wild-type embryos. High *SOX2* expression (red) in the central optic cup and high *PAX6* expression (green) in the peripheral optic cup margin. Scale bar: 100μm. Figure has been adapted from Matsushima *et al.*, 2011.

Both TFs are essential for the maintenance of pluripotent retinal progenitor cells and normal eye development and maintenance. Changes in *PAX6* or *SOX2* protein levels results in ocular defects such as aniridia and anophthalmia. For example, upon *SOX2* protein level reduction, there is an

immediate increase in *PAX6* expression, causing a bias towards a neurogenic to non-neurogenic cell fate conversion. Conversely, *PAX6* protein level reduction leads to loss of CE precursors and a distal shift in the boundary between prospective NR and CE (Smith *et al.*, 2009). The ability of both TFs to mutually repress each other leads to the refinement of the portion of the ciliary epithelium (high *PAX6*, low *SOX2* levels) at the distal end of the neural retina (high *SOX2*, low *PAX6* levels) (Matsushima *et al.*, 2011).

In the corneal epithelium, *ALDH3a1* is jointly regulated by *PAX6* and *OCT1* whereby reduction in *ALDH3a1* correlates with reduced *PAX6* in corneas of heterozygous Small eye mice (Davis *et al.*, 2008). *ALDH3a1* can constitute up to 50% of the water-soluble proteins present in the corneal epithelial cells of most mammals (Nees *et al.*, 2002). *ALDH3a1* has been implicated in direct absorption of UV light (Pappa *et al.*, 2003), metabolism of toxic aldehydes produced by light-induced lipid peroxidation (Estey *et al.*, 2007), acting as a chaperone (Manzer *et al.*, 2003) and even lengthening the cell cycle.

1.5.4 *PAX6* in the developing lens

The initial 3D eye primordium is comprised of the lens vesicle, optic cup and surface ectoderm. The key signalling pathways involved in the initial formation of the eye primordium are FGF (fibroblast growth factor), retinoic acid, TGF- β (transforming growth factor-beta) and Wnt signalling (Huang *et al.*, 2015; Jidigam *et al.*, 2015) (Betters *et al.*, 2010; Huang *et al.*, 2015; Jidigam *et al.*, 2015).

Lens formation is dependent on correct development of the retinal neuroectoderm which is located beneath the head surface ectoderm. Suppression of Wnt/ β -catenin signalling in the surface ectoderm is one of the prerequisites for lens development. *PAX6* can directly control expression of several Wnt inhibitors such as *SFRP1*, *SFRP2* and *DKK1* in the presumptive lens which in turn decrease Wnt/ β -catenin signalling allowing for lens induction (Machon *et al.*, 2010). Absence of *PAX6* function, leads to aberrant canonical Wnt activity in the presumptive lens leading to abnormal lens development.

Crystallins are the essential, long-lived proteins of the eye lens. In humans, the major crystallins are classed in two separate superfamilies: the heat-shock proteins (α -crystallins) and the $\beta\gamma$ -crystallins. Crystallins are found in terminally differentiated fibre cells of the lens and survive without turnover throughout life, while ensuring that their molecular organisation maintains the refractive properties of the lens. *PAX6* is a key regulator of crystallin gene expression. In the developing lens, *PAX6* can autoregulate its own expression through an evolutionary conserved enhancer region termed ectodermal enhancer (EE), which is upstream of the P0 promoter (Williams *et al.*, 1998). Aside from *PAX6* binding sites, the EE also contains multiple *MEIS1/2*, *OCT1*, *SIX3* and *SOX1/2* binding sites which also influence *PAX6* expression. It is believed that *PAX6* autoregulation in the developing lens is very important, as this potentially ensures the maintenance of lens cell type memory following cell divisions (Cvekl and Ashery-Padan, 2014). Furthermore, *PAX6* can repress *TGFB2* in the lens (Sun *et al.*, 2015), while TGF- β signalling through SMAD3 binding to *PAX6* is able to repress *PAX6* autoregulation at the P1 promoter (Grocott *et al.*, 2007).

During lens placode formation, *PAX6* forms a molecular complex with *SOX2* on lens-specific enhancers such as the chicken δ -crystallin (DC5) enhancer (Kamachi *et al.*, 2001). *PAX6* and *SOX2* form a tertiary complex on these enhancers which is stabilized by both protein-protein and protein-DNA interactions. Not only can these two TFs cross-regulate each other and form complexes to regulate target genes, they also have the ability to auto-regulate, making the control of protein levels even more intricate (Figure 1-11). For example, the cis-element enhancer SIMO, which is situated 150kb downstream of the *PAX6* P1 promoter (Figure 1-5), can activate *PAX6* transcription when bound by *PAX6* (Bhatia *et al.*, 2013). Point mutations in this regulatory element have been identified in patients with aniridia (Bhatia *et al.*, 2013). *SOX2* expression in the lens placode, is induced jointly by *PAX6* and *SOX2* on the N3 *SOX2* enhancer (Nishimura *et al.*, 2012), whereas the synergistic interaction of *PAX6* and *SOX2* on the Le9 enhancer positively regulates *PAX6* expression (Aota *et al.*, 2003). Overall, this complex

relationship between the two TFs is a highly regulated, dosage sensitive process which is centered on the stoichiometry of *PAX6* and *SOX2* protein levels for correct ocular development.

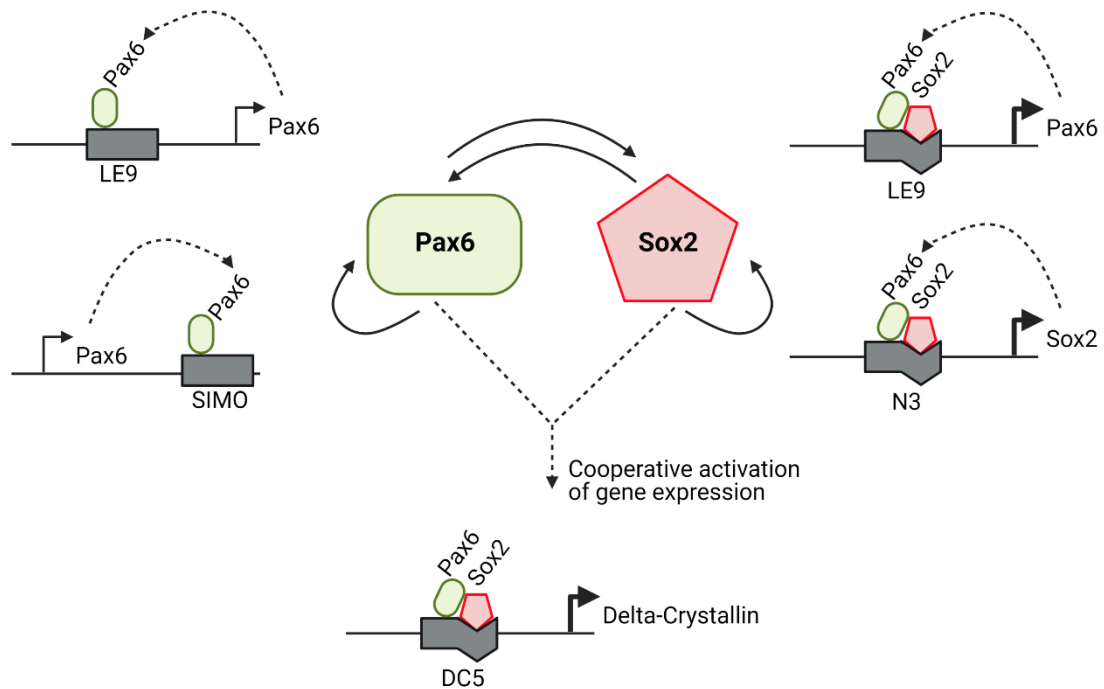


Figure 1-10: Schematic representation of auto- and cross- regulation of *PAX6* and *SOX2* during optic field induction and development. SIMO, is situated 150kb downstream of the *PAX6* P1 promoter and activates *PAX6* transcription when bound by *PAX6*. The N3 enhancer is a *SOX2* enhancer which is activated by the synergistic interaction of both *PAX6* and *SOX2* in the lens placode. The Le9 enhancer is a *PAX6* enhancer which is regulated by the combinatorial interaction of *PAX6* and *SOX2*. N3 and Le9 can activate gene transcription when bound by *SOX2* and *PAX6* respectively, but not to the same extent when activation is synergistic. For the δ -crystallin (DC5) enhancer, the synergistic interaction of both *PAX6* and *SOX2* is required. *PAX6* and *SOX2* form a tertiary complex on the enhancer which is stabilized by both protein-protein and protein-DNA interactions.

1.5.5 *PAX6* Pathogenic Mutations

A database of all reported *PAX6* mutations in humans was first assembled in 1998 (Brown *et al.*, 1998) with a total of 90 mutations. Since then, the total number has exceeded 500, which have been described to affect *PAX6* and its regulatory regions (Cunha *et al.*, 2019). These mutations range from insertions, deletions and duplications to single nucleotide substitutions, resulting in frameshift, missense and nonsense molecular consequences. Over 95% of variants in the current '*PAX6* Mutation Database' are intragenic mutations (Brown *et al.*, 1998), while the remaining variants are gene deletions or mutations identified in 5' and 3' regulatory elements. The large majority of intragenic mutations are located within exons 5, 6 and 9, which encode the PD (exon 5 and 6) and HD (exon 9). When considering the mutation type, the most common intragenic mutations are nonsense (30%) followed by frameshifts (27%), splice site alterations (15%) and missense (12%) (Tzoulaki, White and Hanson, 2005). Even though it is easy to understand how nonsense, frameshift and splice site alterations can be pathogenic, it is of great interest to us how some missense mutations in the PD (protein can be made) may still cause severe ocular malformations such as bilateral microphthalmia (i.e. p.Ser54Arg and p.Asn124Lys) (Williamson *et al.*, 2020).

The aforementioned result in a broad range of phenotypes with the most common being aniridia, which is a pan-ocular disorder that is primarily characterised by the absence or hypoplasia of the iris (Landsend, Lagali and Utheim, 2021). Aniridia is estimated to occur in the general population in 1 in 70 000 in individuals independently of race or sex (Richardson *et al.*, 2016). Classical aniridia is almost exclusively caused by genomic variants that result in *PAX6* haploinsufficiency such as gene deletions, gene disruptive variants (e.g. frameshift mutations leading to premature stop codons and NMD) and even cis-regulatory variants (SIMO enhancer). *PAX6* missense variants often mimic haploinsufficiency, but are usually associated with milder phenotypes, such as partial aniridia, optic disk malformations and isolated foveal hypoplasia. Other ocular phenotypes can include severe bilateral microphthalmia (Williamson *et al.*, 2020), optic nerve anomalies (i.e.

hypoplasia) or anterior segment dysgenesis to name a few (Hingorani *et al.*, 2009). PAX6 missense mutations p.Ser54Arg and p.Asn124Lys for example, have been termed as ‘worse-than-null’ (gain-of-function), with distinguishable severe bilateral microphthalmia, coloboma, congenital corneal opacification, retinal detachment and lens defects comprising primary aphakia, reduced size and subluxation (Williamson *et al.*, 2020).

As previously mentioned, *PAX6* is also expressed in the brain and thus such mutations can also cause neurodevelopmental abnormalities, like autism (Maekawa *et al.*, 2009) and attention deficit hyperactivity (ADHD) disorders (Kikkawa *et al.*, 2019), language impairment and in some cases the absence or malformations of the pineal and pituitary gland (*PAX6* expression driven by RB enhancer) (Mitchell *et al.*, 2003; Kleinjan *et al.*, 2006; Hanish *et al.*, 2016). Additionally, *PAX6* mutations have been associated with obesity and diabetes mellitus due to its role in development of the endocrine pancreas and specification of hormone producing endocrine cell types (Boese *et al.*, 2020).

The pathogenic *PAX6* mutations observed are heterozygous, sporadic or familial autosomal dominant with notable variability in phenotypic abnormalities (Tarilonte *et al.*, 2018). Recent studies have reported that the frequency of mosaicism could be as high as 17.5% among apparent *de novo* cases for different autosomal dominant disorders (Qin *et al.*, 2016) and there have been several reports attributing the phenotypic variability observed in some sporadic *PAX6* patients (identical mutations) to mosaicism.

The majority of pathogenic missense *PAX6* mutations are found in the PD (Williamson *et al.*, 2020) and have been linked with MAC (microphthalmia, anophthalmia and coloboma) phenotypes. MAC is a group of eye developmental disorders which is characterised by reduced size or absence of the ocular globe and it is caused by mutations in more than 90 genes, including *PAX6* and *SOX2* (Harding and Moosajee, 2019). MAC phenotypes are primarily associated with mutations that likely disrupt PAX6-DNA interaction, chiefly via the PD. *SOX2* interacts with PAX6 (mainly PD) on DNA to regulate key genes during eye development. Mutations in *SOX2* (primarily HMG

domain) are the most common cause of bilateral anophthalmia and severe microphthalmia.

1.6 SOX2 Gene and Protein Architecture

1.6.1 SOX2 Gene Architecture

SOX2 belongs to the SOX family of transcription factors that contain a 79 amino acid HMG DNA-binding domain. SOX1, SOX2 and SOX3 belong to the B1 subfamily and are expressed in various stages of embryonic development and cell differentiation, playing critical roles from the earliest stages of development. SOX2 is a single-exon gene found on chromosome 3 whose expression is regulated by more than twenty different enhancers distributed over a 200kb genomic region with distinct spatial and temporal activity, ranging from pluripotency of ES cells, CNS differentiation to optic field induction.

The majority of these enhancers are conserved in DNA sequences across vertebrate species (Nishimura *et al.*, 2012). For example the N2 enhancer, is

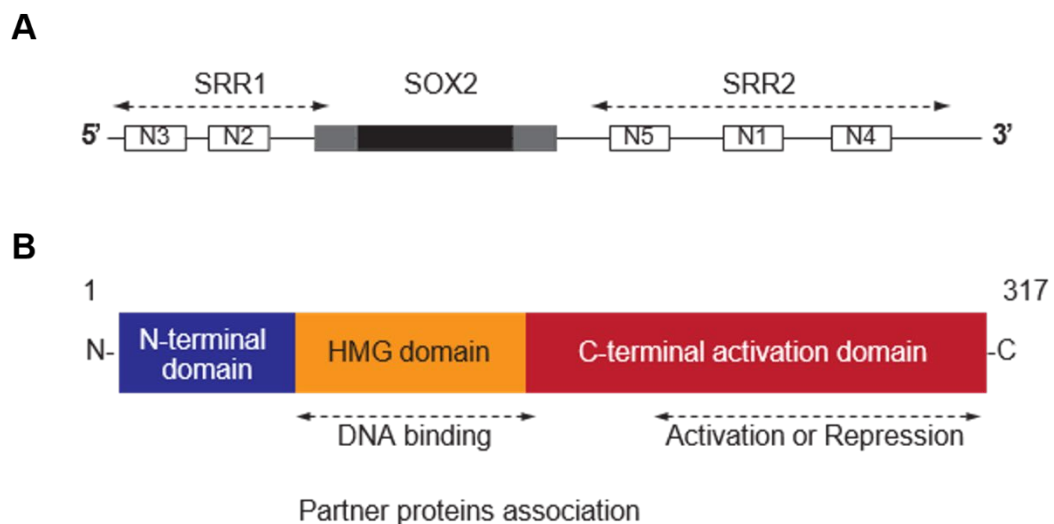


Figure 1-11: SOX2 protein and gene structure **A)** Gene structure of human SOX2. SOX2 coding and untranslated regions are coded in black and grey boxes respectively. SOX2 is an intron-less single open reading frame gene with conserved cis-regulatory elements, named N1 to N5 respectively. These are found in the SRR1 and SRR2 (SOX2 regulatory regions) at the 5' and 3' untranslated regions **B)** The domain structure of human SOX2 protein. Schematic adapted from *Shimojaki, 2017*

responsible for SOX2 activation in the epiblast and early anterior neural plate and is regulated by Zic, Pou and Otx TFs (Iwafuchi-Doi *et al.*, 2011). Conversely the N1 enhancer is activated in the caudal lateral epiblast adjacent to the primitive streak (Kondoh and Takemoto, 2012), whereby its activation is dependent upon Wnt and Fgf signals. During lens placode formation, the SOX2 N3 enhancer is activated by the combinatorial actions of both PAX6 and SOX2, providing evidence for SOX2 autoregulation. Additionally, a 13kb-long distal super-enhancer element (SEE) located 100kb downstream of the mouse SOX2 gene locus was identified (Li *et al.*, 2014). This SEE was shown to be occupied by Oct4, Sox2 and Nanog, physically interact with the Sox2 gene locus via DNA looping, and be responsible for over 90% of Sox2 expression in mESCs.

SOX2 lies in an intron of a long multiexon (10 exons) lncRNA that is transcribed in the same orientation, termed SOX2 overlapping transcript (SOX2OT) (Fantes *et al.*, 2003). SOX2OT has several transcription start sites and can be differentially spliced, generating several SOX2OT isoforms. Even though SOX2OT's function is not completely understood, it has been shown to negatively regulate neural progenitor proliferation by interacting with the epigenetic regulator YY1 in the developing cerebral cortex. In addition the authors identified that YY1 binding to CpG islands at the SOX2 locus is SOX2OT dependent, leading to transcriptional repression of SOX2 (Messemaker *et al.*, 2018). A SNP (rs9842371) present in SOX2OT has been associated with the risk of developing age-related nuclear cataract (Yonova-Doing *et al.*, 2020).

1.6.2 SOX2 Protein Architecture

TFs bind DNA at specific sites to regulate transcription. Nevertheless, TF target sequences can often be inaccessible due to nucleosomes. The core structure of a nucleosome is comprised of two copies of four different histone proteins (H2A, H2B, H3 and H4) that form an octameric complex which is wrapped by about 150 DNA base pairs for ~ 1.7 turns (Cutter and Hayes,

2015), known as nucleosomal DNA, which is inaccessible to most TFs, with some exceptions such as the pioneer TFs (Magnani, Eeckhoutte and Lupien, 2011). The ability of pioneer TFs such as SOX2, to recognize and bind their target sequences even within nucleosomal DNA, is of critical importance especially for initial cell programming or reprogramming in embryonic development. SOX2 is one of the key TFs required and able to reprogram differentiated cells to become pluripotent (Soufi, Donahue and Zaret, 2012). Recent cryo-electron microscopy experiments have shown that SOX2 can bind and locally distort DNA and detach terminal nucleosomal DNA from the histone octamer, thereby increasing DNA accessibility (Jagga *et al.*, 2021).

SOX2 protein (317 amino acids) is comprised of three main domains (Figure 8). These include the N-terminal, HMG (DNA binding domain) and a transactivation domain. SOX2 nuclear localisation is regulated by two nuclear localisation signals (NLS) positioned within the two extremities of the HMG domain, conserved across all Sox family members. Protein transport to the nucleus via the classical import pathway is an active process and requires the presence of an NLS to be recognized by importin- α (IMP α) (Goldfarb *et al.*, 2004) and more precisely for SOX2, the IMP3 α (Jagga *et al.*, 2021). Mutations within these regions can impair nuclear localisation which has been associated with poor prognosis in cancer (Li *et al.*, 2001; Sumita *et al.*, 2018).

Even though SOX2 can bind to enhancers and regulate gene expression as a monomer, it mostly operates as a heterocomplex. The most well-studied binding partner of SOX2 is OCT4 (another stemness/Pioneer TF factor) and together they control self-renewal and pluripotency in embryonic stem cells. The stoichiometry of OCT4 and SOX2 is important for their function as increasing levels of OCT4 or SOX2 in ESCs leads to differentiation to either primitive endoderm or mesoderm lineages whereas knocking down either of the two leads to trophectoderm-like differentiation (Rizzino, 2013). OCT4 and SOX2 can cooperatively bind to two adjacent cis-regulatory elements known as the octamer motif; ATGC(A/T)AAT and the SOX-binding motif; C(T/A)TTGTT (Chen *et al.*, 2008). Molecular simulations have indicated that SOX2 influences the orientation and geometry of OCT4-DNA complexes

(Merino, Bouvier and Cojocaru, 2015). A single-molecule imaging study proposed a model in which SOX2 interacts with chromatin and primes the binding site of OCT4 to facilitate OCT4-DNA heterocomplex formation at composite recognition sites (Chen *et al.*, 2014).

1.6.3 SOX2 Posttranslational Modifications

Post-translational modifications (PTMs) are a dynamic, reversible processes adapted to rapidly transmit signalling information. PTMs are driven by enzymes which can catalyse the addition and removal of these sites. PTMs are broadly be separated into chemical modifications such as phosphorylation, methylation, acetylation and small protein modifications such as ubiquitination and SUMOylation. SOX2 PTMs have been reported throughout SOX2 (Figure 1-12) (Williams, Soufi and Pollard, 2020). One which has directly been linked with TF-DNA binding dynamics is phosphorylation. Phosphorylation of proteins is the addition of a phosphate group (negatively charged) to either a tyrosine, threonine or serine amino acid. Eukaryotic chromatin is a negatively charged long polymer primarily composed of genomic DNA and histones. Therefore, it is logical that the majority of proteins interacting with chromatin such as TFs are to have a net positive charge. The isoelectric point (pI) of a protein is the pH at which it carries no net electrical charge. Proteins found at a pH lower than their pI, carry a net positive charge and vice versa. The average pH of a cell nucleus is around 7.3 and the pI of SOX2 is 9.74. Phosphorylation of SOX2 changes its charge and thus its mode of interaction with DNA and other proteins. For example, phosphorylation of threonine at position 116 (T116) in the N-terminal portion of the HMG domain of SOX2 by Akt surprisingly

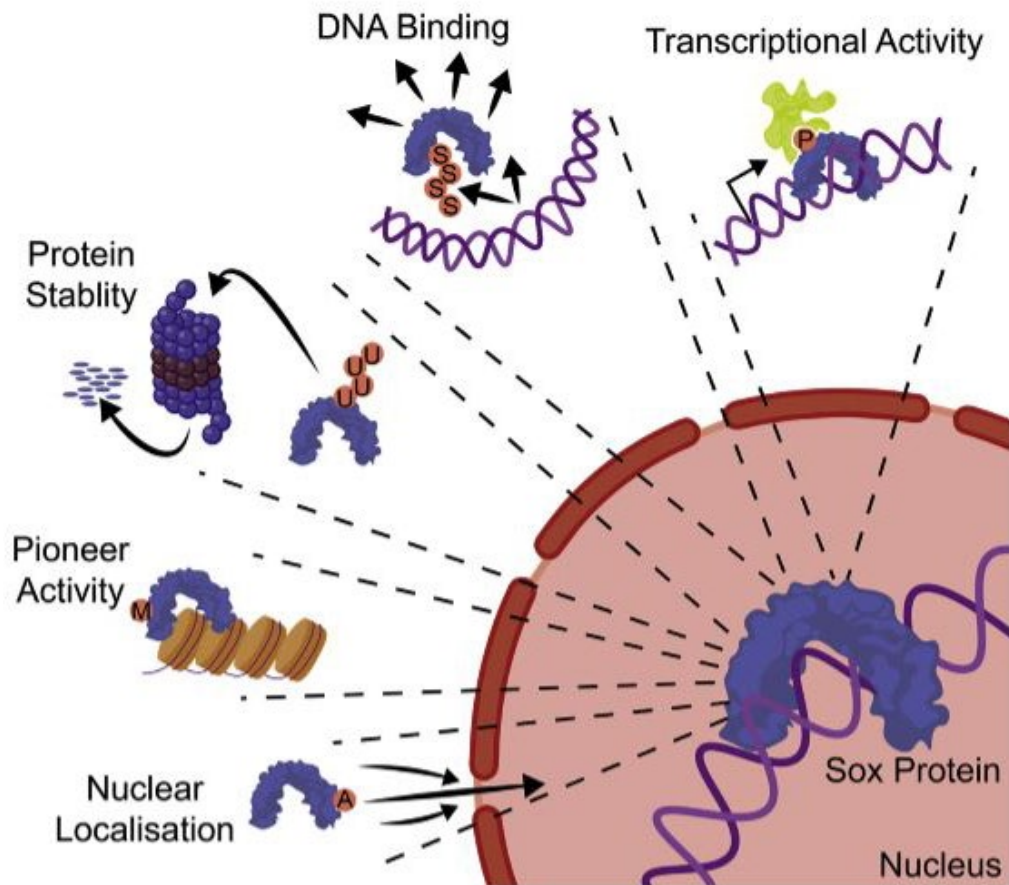


Figure 1-9: Schematic representation of how post translational modifications of Sox proteins may influence their function such as transcriptional activity, DNA-binding, protein stability, pioneer activity and nuclear localisation. Schematic adapted from Williams, Soufi and Pollard, 2020.

enhances its DNA-binding affinity and has been shown to enhance the activity of SOX2 in reprogramming of mouse embryonic fibroblasts to induced pluripotent stem cells (Jeong *et al.*, 2010; Justilien *et al.*, 2014). SOX2 is also phosphorylated by Cyclin-dependent kinase (CDK) at serine residues 39 and 253 (S39, S253) with proposed roles in reprogramming to pluripotency (Ouyang *et al.*, 2015).

Both S39 and T116 flank the HMG domain which could influence conformational changes and/or promote interaction with transcriptional regulators. SUMOylation is reported to play a role in repression of transcriptional activity. For example, SUMOylation of SOX2 lysine 245 (K245) reduces SOX2 DNA-binding affinity to the Fgf4 enhancer (Tsuruzoe *et al.*,

2006). SOX2 methylation at arginine 115 (R115) has been reported in ESCs to enhance oligomerization, possibly through the formation of dimers or larger protein complexes, whereas SOX2 retention on chromatin restricts its methylation levels (Zhao *et al.*, 2011). Although multiple SOX2 PTMs have been identified which influence almost every aspect of SOX2 function, the precise molecular mechanisms of these PTMs are still not fully understood.

1.6.4 SOX2 Interaction with RNA

Interactions of TFs with RNA molecules can functionally regulate gene expression. There have been multiple studies linking SOX2 with several RNA molecules and implicating these interactions with specific functional outcomes (Ng *et al.*, 2013; Messemaker *et al.*, 2018). Large-intergenic noncoding RNAs (lincRNAs), make up the majority of lncRNAs. Even though they do not code for proteins, these transcripts have an exon-intron-exon structure similar to protein-coding transcripts.

More than 8000 lincRNAs have so far been identified, however their functions remain largely unknown. With what is known so far, these RNA molecules can tune gene regulation by affecting nuclear architecture and sequestering intracellular molecules. Additionally they are involved in chromatin topology regulation by influencing transcriptional activation and in protein and RNA scaffolding. Linc1614, has been identified as a potential partner of SOX2 in maintaining pluripotency in ESCs, by primarily mediating the function of SOX2 in gene repression by co-operating with the Ezh2, an essential subunit of the polycomb repressive complex 2 (PRC2) (Figure 1-14) (Guo *et al.*, 2018).



Figure 1-10: Schematic representation of the mechanism through which SOX2 cooperates with linc1614 to recruit the PRC2 complex and repress expression of developmental genes in pluripotent cells. Schematic adapted from Guo *et al.*, 2018

In another example, during neurogenesis the lncRNA, rhabdomyosarcoma 2-associated transcript (RMST) is a transcriptional coregulator of SOX2 whereby it influences the regulation of neural stem cell fate. More precisely, RNA interference and genome-wide SOX2 binding studies, found that RMST is required for the binding of SOX2 to promoter regions of neurogenic TFs (Ng *et al.*, 2013). More recently, analyses of the protein interactome of SOX2 in mESCs identified a number of RNA-binding protein partners, suggesting of the broader capability of SOX2 to interact with RNA molecules (Holmes *et al.*, 2020). In all aforementioned cases though, there were no conclusive experiments to show whether the interaction of SOX2 with RNA molecules was direct and/or indirect (Mallam *et al.*, 2019; Samudiyata *et al.*, 2019). Recent *in vitro* experiments investigated the possibility of SOX2 directly interacting with RNA molecules (Holmes *et al.*, 2020). They were able to demonstrate that human SOX2 directly interacts with lncRNA with high affinity through its HMG DNA-binding domain. The interactions were primarily with double-stranded RNA in a non-sequence specific fashion, ultimately proposing a SOX2 high affinity/low specificity for RNA binding both *in vitro* and *in vivo* (Holmes *et al.*, 2020). It is established that the interactions of TFs and RNA molecules exist (Baltz *et al.*, 2012; He *et al.*, 2016; Mullari *et al.*, 2017), but the developmental significance and mechanisms remain poorly understood.

1.7 SOX2 Expression and Function

1.7.1 Embryonic stem cells

ESCs are pluripotent cells derived from the inner cell mass of blastocysts. They have the ability to self-renew and the potential to differentiate into all cell types. This is primarily attributed to the tightly regulated expression of *SOX2*, *OCT4* and *NANOG*. Altering the levels of any of these TFs leads to different differentiation outcomes as illustrated in Figure 1-15 (Chambers *et al.*, 2003; Babaie *et al.*, 2007; Rodriguez *et al.*, 2007; Kopp *et al.*, 2008).

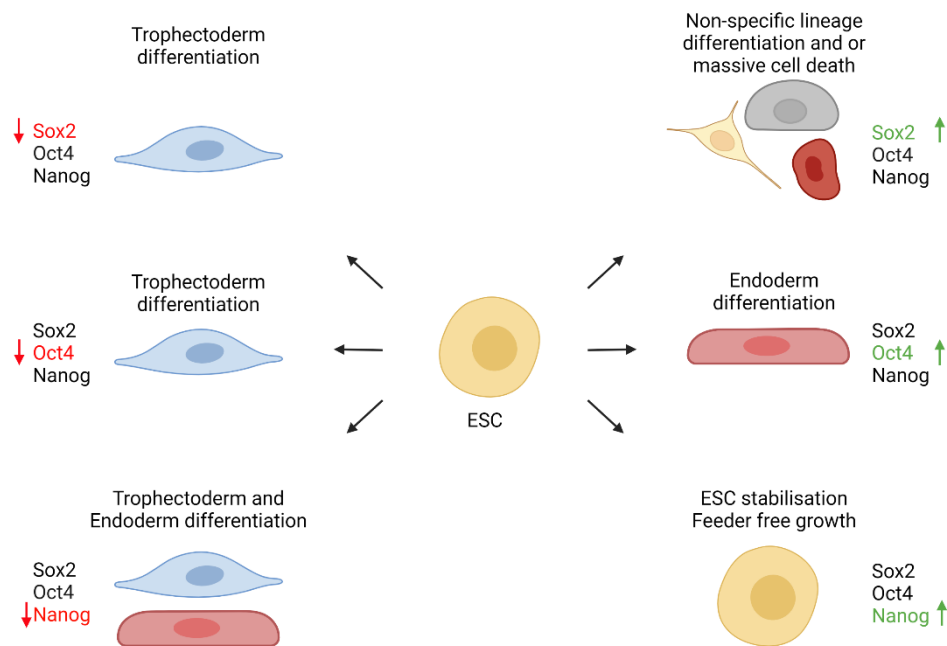


Figure 1-11: A schematic illustration summarising the differentiation outcomes upon alteration of *SOX2*, *NANOG* and *OCT4* protein levels in embryonic stem cells. *SOX2* is required for epiblast maintenance whereas *SOX2*-depleted mESCs differentiate into trophectoderm like-cells. *SOX2* overexpression in mESCs induces non-specific lineage differentiation and/or massive cell death. *OCT4* is required for the maintenance of the pluripotent state in vitro. Decrease in *OCT4* protein levels in mESCs leads to rapid loss of self-renewal and induction of trophectoderm differentiation. Conversely, *OCT4* overexpression triggers endoderm differentiation. *NANOG* controls epiblast versus primitive endoderm decision in the blastocyst. Downregulation of *NANOG* in mESCs triggers differentiation into trophectoderm and endoderm cell lineages. *NANOG* overexpression in ESCs blocks differentiation of ESCs into extra-embryonic endoderm and enables feeder-free growth in hESCs. Overall, *SOX2*, *NANOG* and *OCT4* are the core components in maintaining the pluripotent state in ESC

SOX2 is highly expressed in ESCs, driven by the SOX2 super-enhancer, and SOX2 depletion either by gene-knockout or RNA interference considerably compromises pluripotent marker expression in both mouse and human ESCs. SOX2 and OCT4 co-occupy a large number of enhancers/promoters and regulate the expression levels of multiple target genes required for the maintenance of pluripotency (Boyer *et al.*, 2005; Rodda *et al.*, 2005; Chen *et al.*, 2014). During differentiation, SOX2 controls ESC specification to neuroectoderm while OCT4 and NANOG promote differentiation to mesendoderm (Thomson *et al.*, 2011). SOX2 induces neural induction and enhances neural differentiation by repressing key regulators of other lineage fates (Zhao *et al.*, 2004; Wang *et al.*, 2012). In development, SOX2 expression is initially detected at the morula stage and more specifically at the inner cell mass of blastocyst and epiblast during the latter stages. Zygotic deletion of SOX2 is embryonic lethal, attributed to the failure in forming an epiblast (Avilion *et al.*, 2003).

1.7.2 Central Nervous System

In addition to its role in embryonic development and neural induction, SOX2 also functions to maintain self-renewal of neural progenitor stem cells. It is highly expressed in proliferating neural progenitor cells (NPC), whereby its downregulation leads to differentiation to post-mitotic neuronal and glial cells (Bylund *et al.*, 2003). By reducing SOX2 levels in NPCs, it hinders their self-renewal and proliferation properties and promotes cell-cycle exit and terminal differentiation. Conversely, ectopic expression of SOX2 prevents NPC differentiation into neurons and glia cells (Graham *et al.*, 2003).

Neurogenesis in the adult brain continues in two germinal layers; the subventricular zone (SVZ) of the lateral ventricle and the dentate gyrus subgranular zone of the hippocampus. In both regions SOX2 is expressed in proliferating precursor cells and glial-like cells that are believed to represent stem cells (Ellis *et al.*, 2004).

Despite its peripheral location, the optic nerve of the eye, is actually considered a part of the central nervous system. Like *PAX6*, SOX2 plays an instrumental role in ocular development and maintenance. SOX2 is expressed from the earliest stages of eye development in both lens and retinal tissues (Kamachi *et al.*, 1998). The head ectoderm initially expresses low levels of SOX2 with a ventral to dorsal gradient. Upon contact with the optic vesicle, this leads to an increase in SOX2 expression at the contact region which continues to the lens placode. Invagination of the optic vesicle results in generation of the inner and outer layer of the optic cup (covered in previous sections). Retinal neuronal cells begin to differentiate in the inner layer, whereby the neuroblastic retinal progenitor cells (RPC) become the main SOX2 expressing cells. In addition, among the differentiated cells, ganglion cells (cholinergic), amacrine cells and Müller cells also express SOX2 (Figure 1-15) (Lin *et al.*, 2009). Conditional inactivation of SOX2 in RPCs, results in loss of cell proliferation and differentiation that normally occurs in a cell-autonomous manner (Taranova *et al.*, 2006). A key mechanism in RPC regulation is the direct activation of *Notch1* by SOX2, whereby *Notch* signalling promotes RPC maintenance and

Müller cell differentiation (Taranova *et al.*, 2006). Müller cells are the major type of glia cells found in the vertebrate retina. They are specialized radial glial cells which span the entire thickness of the neural retina. They support the correct function and metabolism of retinal neurons and are essential for normal retinal function. Radial glial cells are bipolar-shaped progenitor cells that are responsible for the production of neurons in the cerebral cortex. Unlike cortical radial glia cells, Müller glia do not appear in the retina until after the first rounds of neurogenesis have occurred. The main property which sets Müller cells apart from radial glia in the other areas of the brain is their optical properties. Müller cells minimize light scattering and act as living optical fibers that are able to guide light through the inner retinal layers towards the photoreceptors (Franze *et al.*, 2007). They are remarkably resilient to damage and retain their RPC potential. Upon retinal tissue injury, they continue cell proliferation and give rise to multiple neuronal cell types (Fischer and Reh, 2001; Bringmann *et al.*, 2009). Consistent with the aforementioned notion, indeed ablation of SOX2 in Müller cells, results in their depletion and overall retinal degeneration (Surzenko *et al.*, 2013).

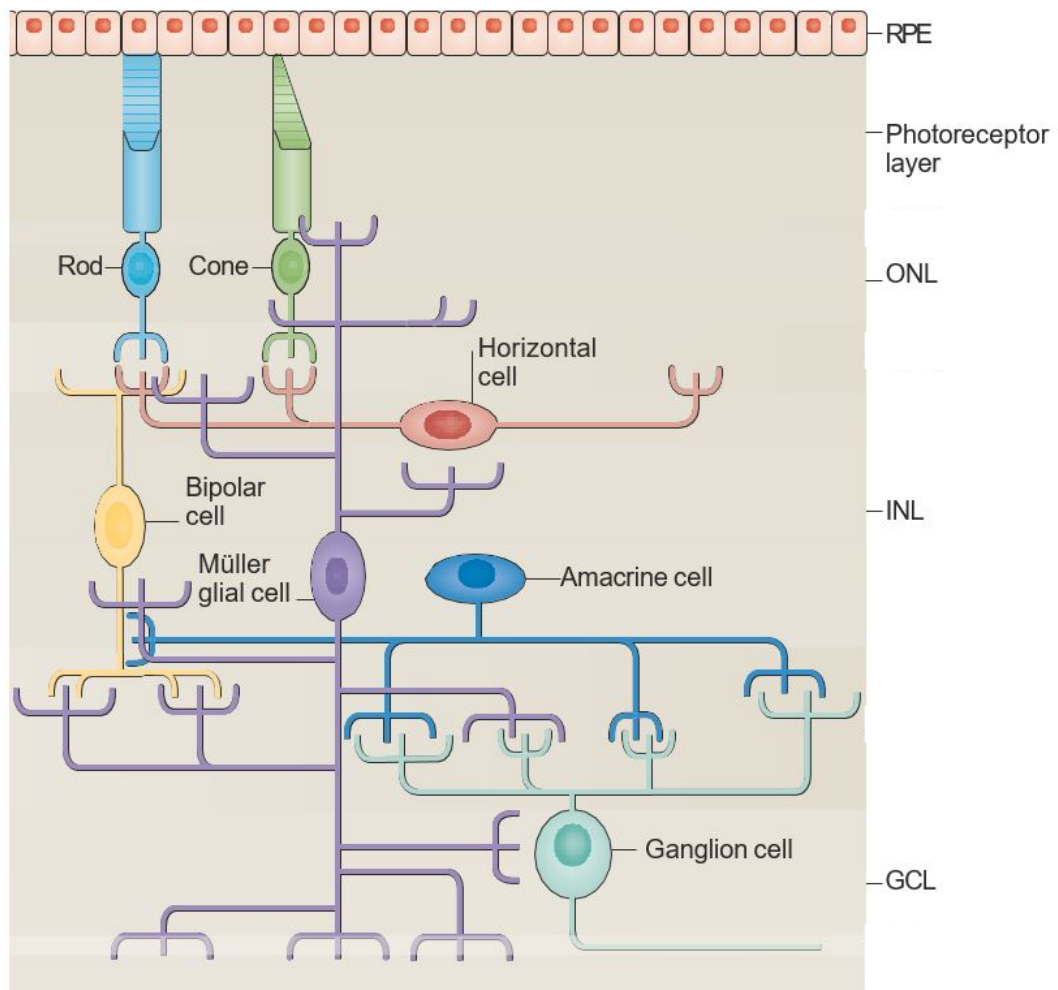


Figure 1-12: An illustration of the retinal anatomy including the major retinal cell types and their organization in the retina. The retina is divided into three laminal layers below the retinal pigmented epithelium (RPE); the outer nuclear layer (ONL), the inner nuclear layer (INL) and the ganglion cell layer (GCL). Schematic adapted from *Goldman, 2014*

The function of SOX2 as a transcriptional activator primarily involves its combinatorial effect with other partner TFs binding to nearby genomic sites such as *PAX6* and *OCT4*. This mechanism has been described as the SOX-partner code (Kamachi and Kondoh, 2013). Retina and anterior neural fold homeobox gene, *Rax*, belongs to the paired-like homeodomain family and is among the earliest genes expressed in the developing eye. *Rax* expression initiates in the optic vesicle and is subsequently localized to the neural retina. *Rax*-deficient mutant mouse embryos and other animal models such as

Xenopus, display defects in optic vesicle development (Bailey *et al.*, 2004). Rax expression is regulated by the combinatorial effect of SOX2 and OTX2, which bind to a conserved regulatory element containing juxtaposed binding sequences of the respective TFs (Danno *et al.*, 2008). Both SOX2 and OTX2 are uniformly expressed from the stage of the prevesicle eye field to the stage immediately before contact with the surface ectoderm (Danno *et al.*, 2008). Considering the lack of optic vesicle formation in Rax-deficient mouse embryos, SOX2-OTX2 interactions are essential for the initiation of eye development. Overall, SOX2 and its interacting partners have an instrumental role in embryogenesis and early development, but also in the maintenance of pluripotent cell populations especially in the CNS.

1.8 Molecular mechanisms of Transcription Factor function

1.8.1 Transcription Factor Haploinsufficiency

For the majority of human genes, a single copy is enough to support normal function. A small subset of genes though exhibit sensitivity to dosage. These genes are characterised as haploinsufficient, and as the word insinuates, this is when decreased amounts of the protein are not sufficient for normal growth and development. Taking into consideration haploinsufficiencies in organismal fitness (physiology, homeostasis, maintenance) one may postulate as to why these genes have not been modulated over evolutionary time. Even though the answer to this question remains elusive, it is believed that these genes may be evolutionarily 'stuck' (Morrill and Amon, 2019). In the way that while decreased protein levels cause an aberrant phenotype, it is believed - and has indeed been shown for some genes such as *PAX6* - that overexpression can also cause a pathogenic phenotype (Schedl *et al.*, 1996), *PKD1* (Thivierge *et al.*, 2006) and *RAI1* (Mullegama *et al.*, 2017). Gene regulation though is not just activating or repressing transcription in a linear fashion. It is a complex, multi-layered and precise process. Gene expression is regulated through the integrated action of multiple cis-regulatory elements such as promoters, enhancers and insulators. Among these elements, enhancers and the TFs that bind them have a major role in initiating gene expression. TFs have the ability to directly interact with DNA and influence gene expression/repression by indirectly promoting or blocking recruitment of RNA polymerase. Quantitative proteomic approaches have shown that the intracellular abundance of TF molecules in eukaryotic cell nuclei lie in the range of thousands of molecules (Simicevic *et al.*, 2013). However, the specific binding sites of most TFs lie in the range of hundreds (MacQuarrie *et al.*, 2011), suggesting that for many TFs a surplus of TF molecules exists over their binding sites at any given time. Yet, the function of many TFs is strongly dependent on dosage with haploinsufficiency being the hallmark of many human developmental disorders. Together with transcription or repression of target genes being a complex process, multiple kinetically distinct events are required for normal TF

function. These events can include, TFs finding and binding their enhancers and overcoming transient non-specific interactions with chromatin, recruitment and interaction with cofactors and chromatin architecture (Figure 1-17).

Taking this into account, it becomes evident why a surplus amount of TF molecules is required as to their binding sites and why TF function is concentration-dependent for gene expression to proceed accurately and timely during development. At the molecular level, TF levels alter functionality by interfering with the dynamic behaviour of TFs in cells. Due to the complex nature of TF functionality, there may therefore be several molecular and/or functional properties of TFs, which may underlie TF haploinsufficiency in a tissue-specific manner and that need to be investigated. These include; the absolute TF numbers, their stoichiometry with interacting proteins, the availability and binding kinetics to their target sites and their biophysical properties such as their ability to form condensates (Auer *et al.*, 2021). In all cases, reduced TF levels can have a profound effect on one or more of the aforementioned properties. Exactly how these parameters vary simultaneously between or within cells, tissues or entire organisms adds to the sheer complexity in characterising the causes of dosage sensitivity and haploinsufficiency in humans.

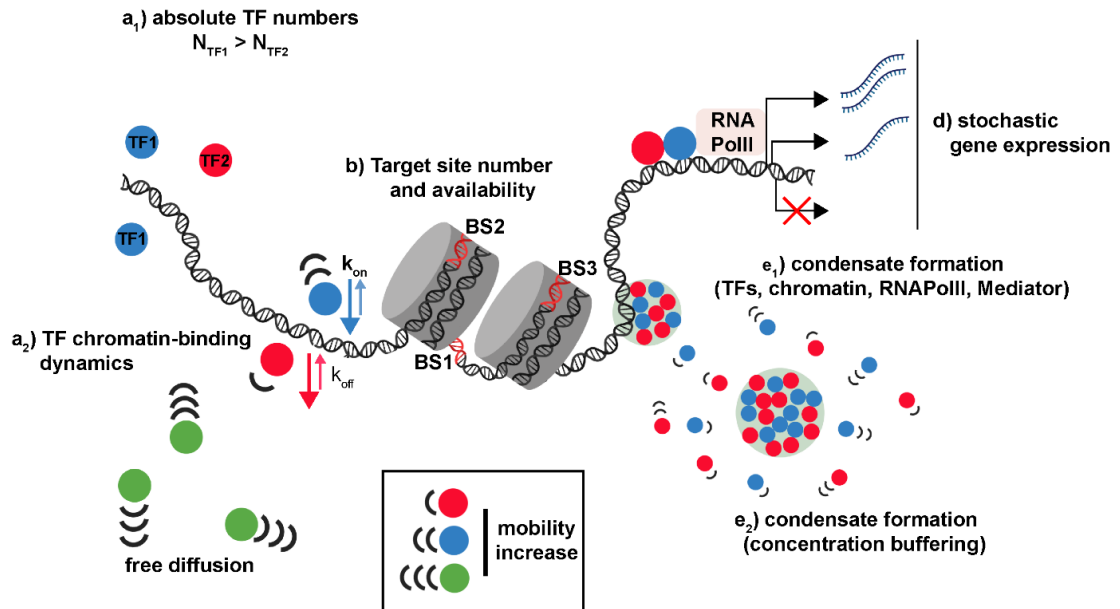


Figure 1-13: Quantitative mechanisms that affect TF function. (A) a₁: N , absolute numbers of TF molecules. a₂: Quantitative TF-chromatin binding dynamics. (B) Number and availability of TF-binding sites on chromatin (red DNA stretches represent TF-binding sites). (C) Stoichiometry of TF-interacting proteins. (D) Stochastic gene expression. (E) e₁: TF co-condensate formation with chromatin, Mediator and RNA Pol II. e₂: Intracellular buffering of TF concentration caused by formation condensate. BS, binding site. Blue, red and green circles represent different TFs; grey cylinders represent nucleosomes; circles with a light green background represent condensates. Schematic adapted from Auer *et al.*, 2021

1.8.2 TF Intranuclear Concentration and Binding Site Availability

Spatiotemporal precision in TF level regulation is paramount in driving cell specification and differentiation. Particularly in early development, different levels of stemness TFs control the fate of individual cells. For example in mouse blastocysts, SOX2 and OCT4 distribution between cells confers an initial differentiation bias at the 4-cell stage (Goolam *et al.*, 2016). In addition, the intrinsically fluctuating levels of OCT4 and SOX2 in stem cell populations contributes to lineage commitment during differentiation (Strebinger *et al.*, 2019). The dependence of TF concentration for correct function, explains why abnormal alterations in TF gene dosage can lead to detrimental developmental effects. TF dosage sensitivity is not the same in all tissues. Changes to the required concentrations of TFs at different spatiotemporal developmental points, can elicit diverse effects in gene regulation, resulting in pathogenic phenotypes of variable severity. For example both *PAX6* and *SOX2* are expressed in tissues other than the eye, but most often the most severe phenotype of haploinsufficiency is restricted to the eye, possibly due to the extensive interactome of the two proteins and how their precise protein levels influence cell-type specific differentiation.

At the subnuclear level, local TF concentrations can influence gene expression by affecting the kinetics of TF binding to chromatin, RNAPolIII recruitment and even transcriptional bursting (Nelson *et al.*, 2004; Senecal *et al.*, 2014). Single-molecule tracking experiments revealed that different TFs can have distinct target-search strategies. For example c-Myc has been described as a ‘global explorer’ of the nucleus, whereas P-TEFb (positive transcription elongation factor) is able to reach its targets in a position-dependent manner, implying that TFs-binding site exploration geometry and identification is restrained by their interactions with nuclear structures (Izeddin *et al.*, 2014). In mESCs clustering of SOX2-bound enhancers influences the search for local SOX2 target sites and facilitates gene transcription (Liu *et al.*, 2014). TF subnuclear localisation can also vary between cell cycle stages and many TFs become enriched on mitotic chromosomes (including SOX2), termed as mitotic bookmarking (Deluz *et al.*, 2017; Dufourt *et al.*, 2018) whereby TF binding is

maintained through cell divisions, conferring transcriptional memory (Palozola, Lerner and Zaret, 2019). TFs rarely act alone and predominantly bind regulatory elements jointly with other TFs or cofactors to regulate gene expression (e.g. SOX2-OCT4, SOX2-PAX6) (Golan-Lagziel *et al.*, 2018). TFs stoichiometry with interacting partners depends largely on concentration whereby lack of or excess amounts may impede their molecular function (Wuebben *et al.*, 2012; Gao *et al.*, 2017). As a whole, the absolute number of TF molecules is an important determinant influencing gene expression by means of various mechanisms.

Cell specific sensitivity to TF levels also depends on the number of the TF binding sites on chromatin, their accessibility (subject to developmental control) and the overall binding dynamics. TF binding affinity to their respective DNA-binding sites is of great importance in terms of normal gene regulation (Levine, 2010; Vukojević *et al.*, 2010). This is chiefly dependent on binding site sequence, which can determine the ‘strength’ with which the TF binds. For example, low-affinity binding sites are less likely to be bound by a TF when its concentration is low and may therefore require either increased amounts and/or cofactors for transcriptional regulation. Binding-site affinity and efficacy in conjunction with TF concentration can serve as an additional mechanism in controlling TF spatiotemporal function. The accessibility of these binding sites depends on chromatin conformation and nuclear organisation. Chromatin conformation within nuclei is dynamic and cell-type specific whereby extensive research in its effect on gene expression has been conducted. Some examples include, chromatin compaction (Akhtar *et al.*, 2013), nucleosome positioning (Li *et al.*, 2020) and topologically associated domains (Tsai *et al.*, 2017). Furthermore, at the whole genome level, histone modification signatures and DNA methylation can also directly influence the ability of a TF to bind to its binding site. As such, TF binding site accessibility, or lack of, can trigger differential gene expression and differentiation biases between seemingly identical cells, even when the responsible TFs are expressed in all cells at the same levels (Plachta *et al.*, 2011; White *et al.*, 2016).

1.8.3 Protein Phase Separation

Phase separation is the creation of two distinct phases from a single homogeneous mixture. Protein phase separation has gained substantial attention in biological research whereby in the past, has been associated with multiple cellular functions, from membrane-less organelles such as the nucleolus (Feric *et al.*, 2016), to normal and abnormal variants of widely studied proteins, such as *FUS* (Amyotrophic lateral sclerosis), *G3BP1* (stress granule assembly) and BRD4 (chromatin and TF organization) (Han *et al.*, 2020; Niaki *et al.*, 2020; Riback *et al.*, 2020). The ability of TFs to phase separate and aid in nuclear compartmentalisation and gene regulation (Boija *et al.*, 2018) is an exciting concept. The ability of proteins to phase separate depends on physico-chemical conditions such as protein concentration,

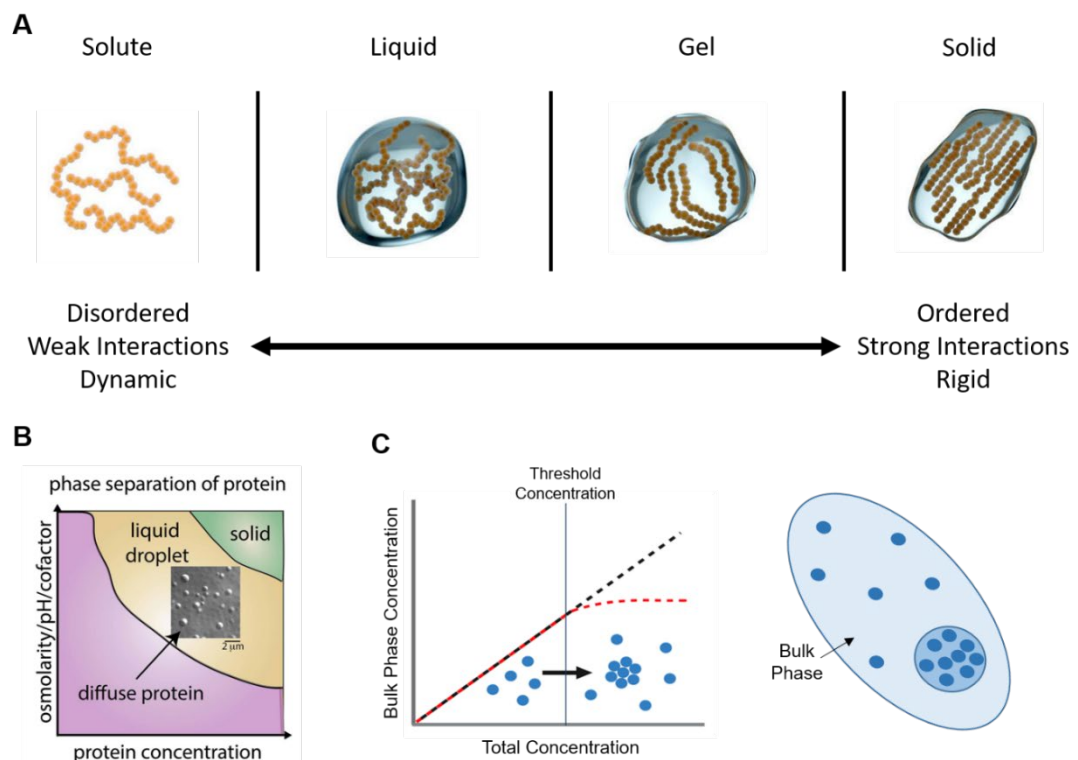


Figure 1-14: Protein Phase Separation Characteristics. A) Material state and dynamics can vary in a wide range from liquid-like to solid phase separated states. B) Phase separation of proteins is largely concentration driven but can be greatly influenced by pH and other cofactors in biological systems. C) The threshold concentration is the concentration in which when surpassed the protein can form phase separated compartments and as a result could potentially buffer the bulk phase concentration upon further increase in concentration. Blue dots depict protein molecules. Part A was adapted from Boeynaems *et al.*, 2018

molecular charge, 3D structure and environmental pH conditions (Figure 1-18) (Taratuta *et al.*, 1990; Wang *et al.*, 2018). The thermodynamics of phase-separated systems predicts that the protein concentration within the condensates is higher than the surrounding dilute phase (Klosin *et al.*, 2020). Protein phase separation can possess liquid, gel or solid-like properties, but the biological significance of these entities remains ill-defined (Strom *et al.*, 2017; Boeynaems *et al.*, 2018; Alberti and Dormann, 2019). Liquid-liquid phase separation (LLPS) condensates exhibit rapid protein exchange with the surrounding dilute phase, movement within the dense phase, including fission and fusion phenomena during formation and maturation (Phair and Misteli, 2000; Hoege *et al.*, 2009). Even though the precise biophysical mechanism of condensate formation remains elusive, weak multivalent interactions predominantly in intrinsically disordered regions (usually found within linker sequences and transactivation domains) are believed to be the main driver (Dzuricky *et al.*, 2020). Since condensate formation has the ability to buffer concentration and functionally compartmentalise the nucleus (Klosin *et al.*, 2020), then it is possible for condensates to control local TF concentrations.

1.8.3.1 Super-enhancers

Classically, enhancers have been defined as elements that can increase transcription from a target gene promoter when inserted in either orientation at various distances upstream or downstream of the promoter (Khoury and Gruss, 1983). Super-enhancers are occupied by an unusually high density of interacting factors, they are able to drive higher levels of transcription than typical enhancers, and are exceptionally vulnerable to perturbation of components commonly associated with most enhancers (Chapuy *et al.*, 2013). Super-enhancers are formed by the binding of TFs at each component enhancer, leading to recruitment of unusually high densities of coactivators, such as Mediator and BRD4. Mediator is a large multisubunit complex that has multiple roles in transcriptional regulation, like the bridging of interactions between TFs and RNA Pol II (Hnisz *et al.*, 2013). BRD4 enables the release of RNA Pol II from the site of transcription initiation. The presence of both MED1 (subunit of Mediator) and BRD4 can be used to define super-enhancers

(Sabari *et al.*, 2018). The ability of these co-activator proteins to form phase-separated condensates at super-enhancers, is believed to be a key mechanism in compartmentalizing and concentrating the transcription apparatus at key cell-identity genes. These properties may also explain why cancer cells acquire large super-enhancers at driver oncogenes and why SEs that facilitate transcriptional dysregulation in disease can be especially sensitive to transcriptional inhibitors. Such molecular crowding brings regulatory sequences and promoters into close proximity (Hnisz *et al.*, 2013) and such reversible local networks favour gene expression. This process can depend on nucleation events triggered by physiological (e.g. NF- κ B in inflammation; Nair *et al.*, 2019), developmental (e.g. Prospero in neural differentiation; Liu *et al.*, 2020) or molecular processes, such as depletion of proteins (e.g. Mediator complex and BRD4 (Sabari *et al.*, 2018)).

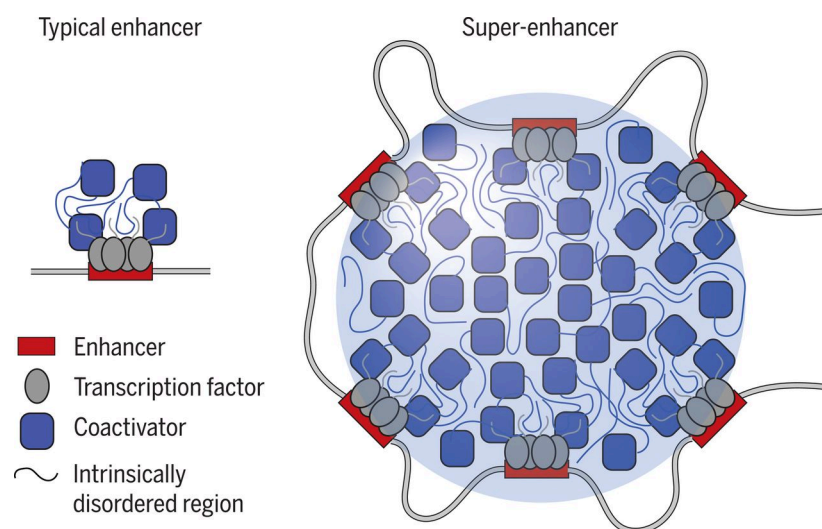


Figure 1-15: Enhancers are gene regulatory elements bound by TFs that recruit coactivators and the transcription apparatus to regulate gene expression. Super-enhancers are large clusters of enhancers with unusually high concentrations of coactivators and transcription apparatus proteins that drive robust expression of genes responsible for key roles in cell identity. This ability of super-enhancers is believed to be largely achieved by protein phase separation, which is driven in part by non-specific interactions of intrinsically disordered domains. Schematic was adapted by Sabari *et al.*, 2018

TFs such as *SOX2*, *OCT4* and *NANOG* (Boija *et al.*, 2018), co-activators such as *BRD4* and *MED1* (Sabari *et al.*, 2018) and RNAPolIII (Boehning *et al.*, 2018) have all been shown to form condensates *in vitro* and/or *in vivo*. Nevertheless, to what extent this depends on condensate formation alone or on clustered DNA-binding interactions, remains under investigation (Li *et al.*, 2020). The majority of work on TF phase separation has focused on linking phase separation with transcriptional output (Sabari *et al.*, 2018). Condensates though have recently been proposed to be involved in buffering the concentration of TFs in the dilute phase (Klosin *et al.*, 2020). Although the universality of such functions remains to be experimentally verified, it is exciting if one relates this to the stringency of TF concentration for normal function, such as in the case of *PAX6* and *SOX2* in ocular development. Therefore, TF condensate formation and how they influence their function, is something that warrants further investigation.

1.9 Overview of advanced microscopy methodologies

The ability to study TF concentration and function mechanisms is becoming increasingly feasible thanks to the advancement of multiple microscopy methodologies. Each methodology has its own advantages and disadvantages and their specific application largely depends on the scientific question that needs to be addressed. For more information regarding methodology application, please refer to Auer et al., 2021 (co-author). Below is a short summary of the most common microscopy methodologies that are currently used to study TF concentration, localisation, stoichiometry and dynamic behaviour.

1.9.1 Fluorescence Correlation Spectroscopy (FCS)

FCS is able to exploit the optical arrangement in a confocal microscope, to detect fluorescent signal in a tiny volume (0.2-2 fl). Fluorescently labelled molecules move in and out of the detection volume causing fluorescence intensity fluctuations, which are recorded over time. These fluctuations can be detected with single-photon sensitivity and sub-microsecond temporal resolution. Subsequently, the recorded fluctuations are subjected to autocorrelation analysis whereby absolute concentrations, mobility and binding properties of the investigated molecules can be derived. This methodology can quantitatively characterise in live cells, tissues and whole organisms, TF concentration, mobility and chromatin binding dynamics (Vukojević *et al.*, 2010; Papadopoulos *et al.*, 2019). The advantages of using FCS are that it has single-molecule sensitivity, it is quantitative and non-destructive, which allows it to be used in live specimens. Due to its high temporal resolution, FCS is ideally suited for the study of fast dynamical processes such as the diffusion of fast small molecules and fast reaction kinetics. Conversely, FCS is not ideal for studying slow reactions, as it does not 'see' immobile molecules (i.e. histones), nor molecules with high concentrations (>1 mM) as this cannot be reliably measured.

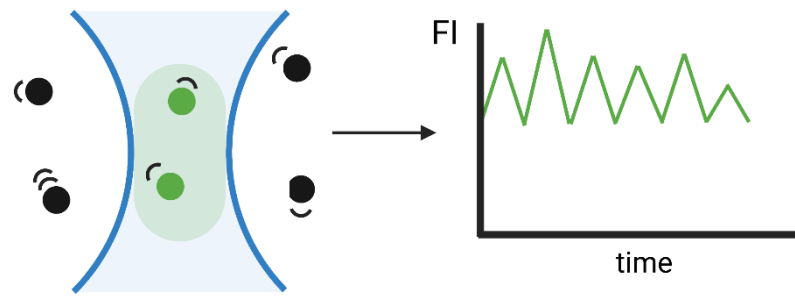


Figure 1-16: Schematic representation of FCS. Fluorescently labelled molecules pass through the detection volume where they are excited by a laser. This leads to photon release which is recorded over time. As molecules move in and out of the detection volume over time, a fluorescence intensity time trace is recorded which is subsequently used for autocorrelation analysis. Black Spheres; Fluorescent molecules in dark state, Green Spheres; Excited molecules inside the detection volume, emitting photons, Light green ellipsoid; Detection volume, FI; fluorescence intensity

1.9.2 Fluorescence Cross-Correlation Spectroscopy (FCCS)

FCCS is often used to investigate interactions between two proteins which are labelled with spectrally distinct fluorophores. Separate detectors enable the simultaneous recording of fluorescence intensity fluctuations from both fluorophores. The fluorophores often used for these experiments are GFP and mCherry (including their derivatives). The recorded fluctuations from each fluorophore are analysed separately. The covariance in fluorescence intensity fluctuations from both fluorophores is subsequently analysed to assess the co-diffusion of the investigated molecules (Haustein and Schwille, 2007; Papadopoulos *et al.*, 2015).

This methodology can be used to study the efficiency and stoichiometry of protein complex formation, such as hetero- and homo-dimers. FCCS is able to provide information on protein-protein interactions with no requirement for proximity and orientation of individual fluorophores. In addition, information regarding the concentrations and mobility of each labelled protein can be obtained, without the need for separate measurements. A main disadvantage of FCCS is that false-positive results may be obtained in terms of protein-protein interaction. This is because sometimes, signal bleed-through (cross-talk) from one detector to the other occurs. For example, this could be caused by large concentration differences between the two investigated proteins or when the system alignment is not optimal. In both cases though, there are established controls which are used prior to experimental measurements.

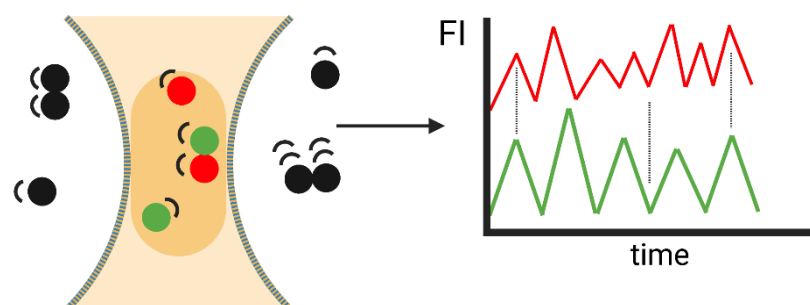


Figure 1-17: Schematic representation of FCCS. Spectrally distinct fluorescently labelled molecules pass through the detection volume where they are excited by their respective laser. This leads to photon release which is recorded over time. As molecules move in and out of the detection volume over time, a fluorescence intensity time trace is recorded for the two different populations of molecules which is subsequently used for autocorrelation and cross-correlation analysis. Black Spheres; Fluorescent molecules in dark state, Green Spheres; molecules excited from the green inside the detection volume, Red Spheres; molecules excited by the red laser inside the detection volume Dark orange ellipsoid; common detection volume, FI; fluorescence intensity

1.9.3 Single-Particle Tracking (SPT)

This methodology is able to track individual fluorescently-labelled particles in order to obtain information on their dynamic behaviour over time. By tracking individual molecules, the trajectories are visualised, and displacements of molecules can be measured. Several models can be used to derive information regarding target molecule kinetics (Chen *et al.*, 2014; Chong *et al.*, 2018). Two key advantages of SPT when compared to FCS are that slow-moving molecules can be recorded and concentration is not a limiting factor. The spatiotemporal resolution of this method is limited by the resolution and image acquisition speed of the microscope setup (Mazza *et al.*, 2012). Analysis relies heavily on which mathematical models are used to fit the data, which are often accompanied by complex statistics. Overall, SPT can require more expensive hardware and much more complex data analysis (statistics and modelling) than FCS.

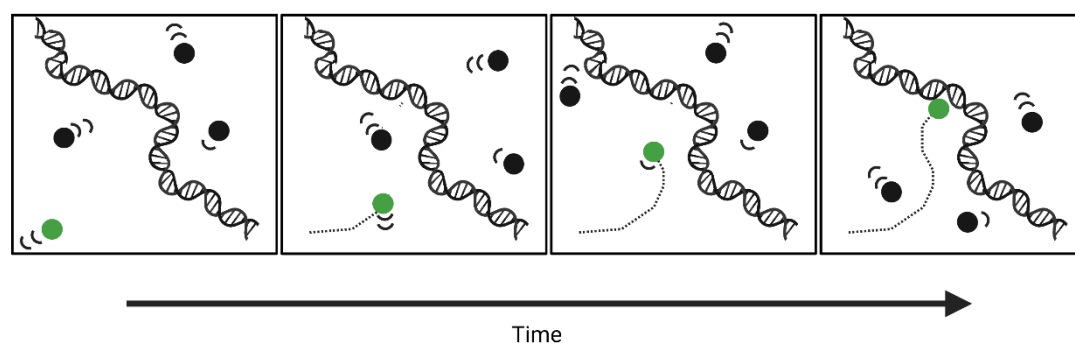


Figure 1-18: Schematic representation of SPT, The green ball represents the fluorescently labelled molecule that is tracked over time. The black balls depict fluorescently labelled molecules which are not tracked. Each image is a representation of a specific time point. Green ball; fluorescently labelled molecule that is tracked over time, Black ball; fluorescently labelled molecules of the same population that are not tracked

1.9.4 Fluorescence Recovery after Photobleaching (FRAP)

FRAP allows for the measurement of target molecule diffusion and in the case of TFs, one is able to derive freely diffusing versus chromatin-bound fractions (Mazza *et al.*, 2012; Liu *et al.*, 2020). Within a small selected region of a sample, intense laser light renders the fluorescent molecules present unable to fluoresce. The rate of fluorescence recovery in the photobleached region is subsequently recorded, by using a low-intensity laser beam to excite the not-bleached molecules. This rate of recovery is then fitted with suitable mathematical models to derive diffusion. FRAP is not limited by target molecule concentration and can therefore be used when higher concentrations are present such as in the case of phase-separated condensates. It is an easy methodology to perform, understand and teach to people who are not so familiar with advanced microscopy. Due to the initial high intensity laser light at the beginning, photochemical and/or thermal cell injuries may be caused. In addition, incomplete fluorescence recovery due to obstacle presence and reversibly photobleached fluorophores may lead to false interpretations.

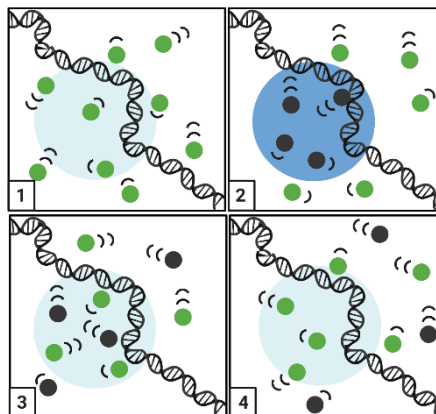


Figure 1-19: Schematic representation of FRAP. A region depicted by the light blue circle in image 1 represents the pre-selected region which will be targeted by the intense laser light. Image 2 represents the time point immediately after the targeting of the region with the intense laser light. The dark blue region represents the region which was targeted by the intense laser light. The black circles represent fluorescently labelled molecules which have been rendered unable to fluoresce. Image 3 represents a short time point after, in which all molecules have moved in random directions (Brownian motion). Some fluorescently labelled molecules which were not damaged by the intense laser light move into the targeted region. After more time has passed (image 4) the pre-targeted region is predominantly made up of excitable fluorescently labelled molecules. Throughout the experiment, the fluorescence intensity of the pre-selected region is recorded (rate of fluorescence recovery), which can be used to calculate target molecule diffusion.

1.9.5 Förster Resonance Energy Transfer (FRET)

This method is often used to study binding and complex-formation between different molecules, such as the formation of hetero- and homo-dimers (Mivelaz *et al.*, 2020). Interacting molecules are fluorescently labelled with distinct fluorophores so that the emission of one fluorophore (donor) overlaps with the absorption spectrum of the other fluorophore (acceptor). The transfer of an energy from an excited donor to an acceptor molecule leads to excitation of an acceptor molecule and subsequent emission of fluorescence which is recorded. This allows for the measurement of close proximity interactions (<10nm) to be measured. It is relatively simple to configure on wide-field fluorescence microscopes that are available in most labs. Due to the spectral crosstalk between donor and acceptor fluorophores though, it is fairly difficult to obtain true FRET data. As a result, considerable control experiments and image processing are required in order to subtract crosstalk components.

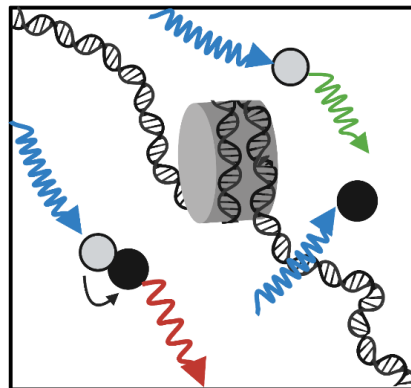


Figure 1-20: Schematic representation of FRET. Donor (light grey ball) and acceptor (black ball) molecules are labelled with spectrally different fluorophores. Upon excitation of the molecules with a blue light (blue arrows) the donor molecule (grey ball) is excited and emits photons which can be collected in the green channel (green arrow). The acceptor molecule (black ball) is not excited by the blue light and does not emit any photons. When the acceptor and donor molecules though are in close proximity, the excited donor molecule, excites the acceptor molecule which emits photons that can be detected in the red spectrum, ultimately suggesting close interaction of the two molecules (<10 nm). Grey cylinder; nucleosome

1.9.6 Stochastic Optical Reconstruction Microscopy (STORM)

This is a super-resolution fluorescence microscopy methodology which relies on the use of photo-switchable fluorophores that undergo reversible changes in their intrinsic properties through the use of light as an external stimulus. At the beginning of the imaging process, all fluorophores are in the 'off' state (non-fluorescent). A small subset of molecules is then stochastically switched 'on' through photoactivation. After precise recording of their location, those molecules switch back to the 'dark' state and a new subset of molecules is switched 'on'. By reading out the location of the stochastically switched 'on' molecules in combination with computational methods and analysis, a high-resolution image can be constructed. This method can locate proteins with a precision in the tens of nanometers range (~ 50 nm). STORM provides the highest spatial resolution of all optical microscopy methodologies and also has quantitative properties as it can detect single molecules and count their numbers. Although multi-colour imaging is possible it is extremely complex. This method requires particular fluorescent labelling such as photo-switchable fluorophores or dyes. Reconstruction of the images is time consuming and requires many on-off cycles ($>10,000$).

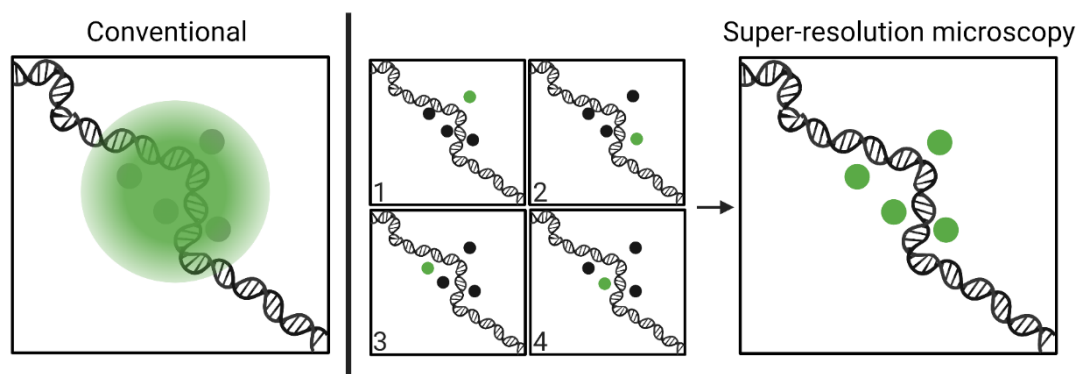


Figure 1-21: Schematic representation of STORM. Conventional confocal microscopes are able to detect fluorescence with a good resolution but not with single molecule sensitivity. As shown by the four numbered squares, thousands of 'on' and 'off' cycles are carried out, recording thousands of images. These images are subsequently analysed and put together, providing an image which can depict single molecules with a precision in the 10s of nanometers (~ 50 nm)

1.10 Overview of Fluorescence Correlation Spectroscopy

Investigating the molecular dynamics of *PAX6* and *SOX2* in live systems, requires methods that can follow the dynamics of individual protein molecules within cells. One way of doing this is single-particle tracking (SPT) which tracks individual fluorescently-labelled particles over time, obtaining information on their dynamic behaviour such as their chromatin-binding kinetics (Izeddin *et al.*, 2014). Another methodology is Fluorescence Correlation Spectroscopy (FCS). FCS allows for study of the spatiotemporal properties of fluorescently tagged molecules in solution, by performing temporal autocorrelation analysis of fluorescence intensity fluctuations (Haustein and Schwille, 2007). Using this approach, one is able to quantitatively study protein-DNA and protein-protein interactions as well as absolute protein concentrations through the analysis of their diffusion behaviour in live cells and tissues. FCS was the chosen method in this thesis.

1.10.1 FCS instrumentation and basic working principles

The principles of FCS were formulated over 50 years ago based on a physical method called fluctuation correlation analysis. In this method, statistical analysis of the time course and amplitudes of spontaneous fluctuations in the number of molecules occurring in a very small detection volume of a system over time was performed, deriving conventional diffusion transport and chemical rate coefficients. During its introduction in the seventies, FCS consisted of large excitation volumes and long correlation times. The combinatorial effect of these two parameters, made the studied molecules prone to photodestruction (Richards, Pope and Widom, 1972; Elson and Magde, 1974). Following the introduction of technological advancements, FCS acquired its present properties, including single-molecule detection sensitivity and short measurement times (Rigler *et al.*, 1993; Widengren, Rigler and Mets, 1994). This paved the way for the study of fast dynamic processes in living cells even when reporter molecules were present at low levels (nanomolar concentrations) (Bacia and Schwille, 2003).

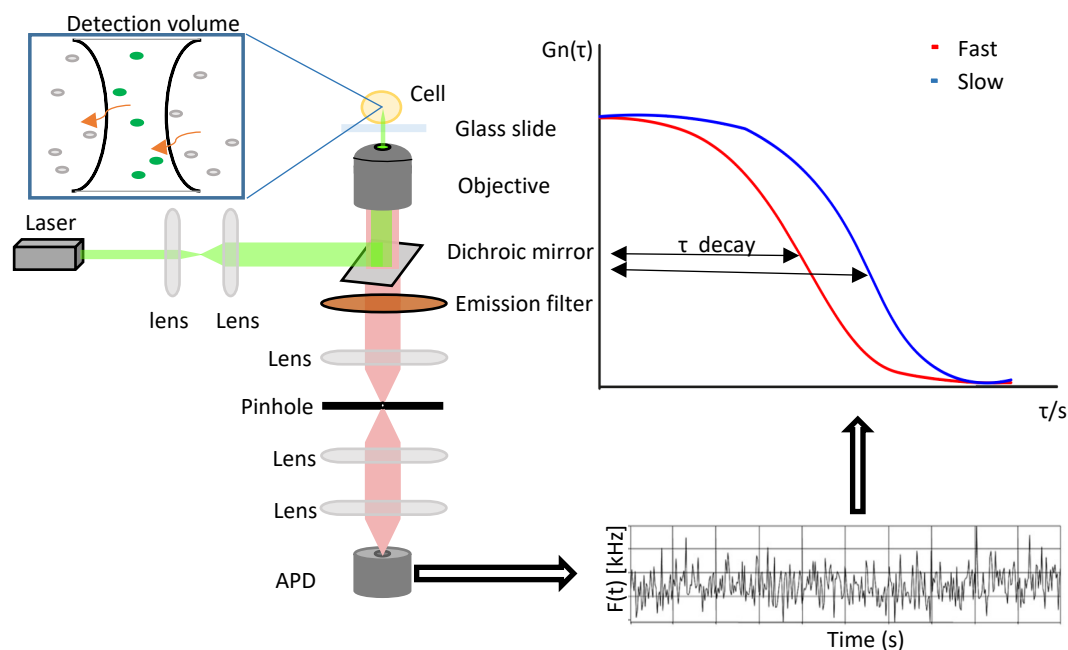


Figure 1-22: A representation of a typical FCS setup in combination with a confocal microscope. An Avalanche Photodiode (APD) records photon counts from the detection volume, which are used to construct a photon count histogram. The photon count histogram is then analysed to obtain an autocorrelation curve, which is subsequently fitted with pre-existing mathematical diffusion models.

Commercial FCS instrumentation is nowadays available from several manufacturers. A schematic overview of a typical FCS setup and subsequent analysis, is shown in (Figure 1-26). The typical FCS setup consists of a laser typically ranging from 405-633 nm (visible spectrum). The laser is reflected into a microscope objective by a dichroic mirror. A dichroic mirror can selectively reflect certain wavelengths and allow other wavelengths to pass. For example a dichroic mirror may allow red laser light to pass through it, whereas it may reflect green laser light. This ultimately allows for the separation of fluorescence light which contains the image information. The laser is focused on a specific area of the sample and the fluorescent molecules that cross the focal volume of the laser become excited and fluoresce. The emitted fluorescence goes back through the objective, passes through the dichroic mirror and is eventually detected by an avalanche photodiode (APD) (Figure 1-27).

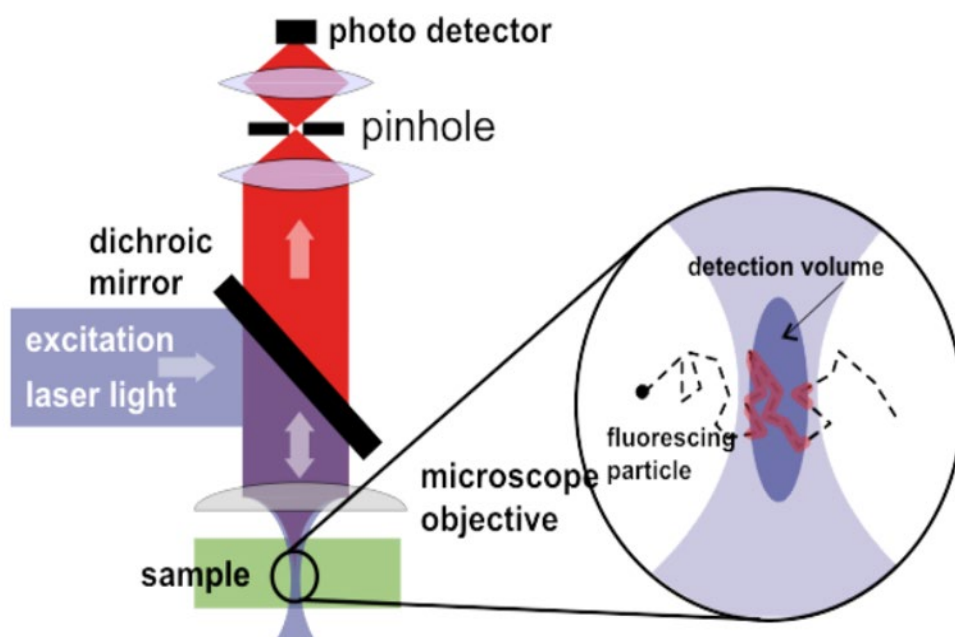


Figure 1-23: Schematic illustration of FCS imaging system and its detection volume. Laser light is focused within a certain volume which makes up the detection volume. When fluorescently labelled molecules move into the detection volume, their emitted fluorescence is recorded by a photo detector. Over time, as molecules move in and out of the detection volume a fluorescence intensity timetrace is recorded.

The APD responds with an electric pulse to each detected photon. The number of pulses derived from the detected photons, is recorded over a defined time interval and correspond to the measured light intensity. Fluorescence intensity fluctuations may be caused, among other reasons, by alterations in the concentration of fluorescent particles moving through the detection volume over time. The recorded fluorescence intensity fluctuations are typically further analysed with correlation analysis.

1.10.1.1 Correlation Analysis

Correlation analysis is used when one wants to observe whether two variables are related to each other. The relationship between two variables can be referred to as the correlation coefficient. Correlation coefficient values can range between -1 and +1 which are considered as the perfect negative/positive

correlations respectively. When the value is 0, it suggests that there is no correlation and no relationship between the two variables.

Autocorrelation is the correlation between a variable and a delayed copy of itself. In other words it is the correlation between two values of the same variable over successive time points. In FCS, fluorescence autocorrelation measures the fluorescence intensity that was recorded, at different time points. Unless the photo-physical properties of the fluorophore change due to processes not related to diffusion (such as chemical reactions, variation in local pH, oxidation, etc.) the fluorescence intensity over time is directly proportional to the number of fluorescent molecules and therefore both parameters can be represented in the following equations:

$$N(t) = [N] + \delta N$$

$$I(t) = [I] + \delta I(t)$$

The overall particle number, $N(t)$, or overall fluorescence intensity, $I(t)$ is derived from the average particle number, $[N]$, or average fluorescence intensity, $[I]$, plus the fluctuations, $\delta N, \delta I(t)$ around the average $[N], [I]$, respectively. A temporal autocorrelation analysis uses this measured signal, $\delta I(t)$ to extract how fast the particles move in the sample. The autocorrelation function is mathematically defined as:

$$G(\tau) = 1 + \frac{\langle \delta I(t) \cdot \delta I(t + \tau) \rangle}{\langle I(t) \rangle^2}$$

This function measures the self-similarity of the fluctuations $\delta I(t)$, if compared to itself, a time τ (lag time) later. To derive information about molecular numbers and corresponding diffusion times, experimentally obtained autocorrelation curves are subsequently fitted with mathematical models describing free three-dimensional (3D) diffusion, and for our experiments, with two diffusional components and a triplet state.

1.10.1.2 Mathematical Model of 3D diffusion

The two diffusion components represent the population of molecules which are freely diffusing and in the case of TFs the population which is interacting with chromatin (Vukojević *et al.*, 2010; Papadopoulos *et al.*, 2015, 2019).

$$G(\tau) = 1 + \frac{1}{N} \left(\frac{1-y}{\left(1 + \frac{\tau}{\tau_{D1}}\right) \cdot \sqrt{1 + \frac{w_{xy}^2 \tau}{w_z^2 \tau_{D1}}}} + \frac{y}{\left(1 + \frac{\tau}{\tau_{D2}}\right) \cdot \sqrt{1 + \frac{w_{xy}^2 \tau}{w_z^2 \tau_{D2}}}} \right) \cdot \left(1 + \frac{T}{1-T} \cdot e^{-\frac{\tau}{\tau_T}} \right)$$

In the above equation, N is the average number of molecules in the detection volume; y is the fraction of the slowly moving molecules; τ_{D1} is the diffusion time of the free molecules; τ_{D2} is the diffusion time of the molecules undergoing nonspecific interactions with DNA; w_{xy} and w_z are radial and axial parameters, respectively, related to spatial properties of the detection volume; T is the average equilibrium fraction of molecules in the triplet state; and τ_T the triplet correlation time related to rate constants for intersystem crossing and the triplet decay. The triplet state function is included, to account for the population of excited particles which change into a long-living (several microseconds) dark triplet state. From there the excited particle relaxes down to the ground state after an average lifetime, τ_T . This leads to blinking (switching off of the fluorophore) on a characteristic timescale τ_T .

1.10.2 FCS applications in biological systems

Qualitative methods assessing TF binding such as chromatin immunoprecipitation and ChIP-seq, provide static information derived from fixed cell populations (Valouev *et al.*, 2008). Despite being informative, these analysis fail to provide insight into dynamic protein behaviour in real-time. Live-cell methodologies such as FCS offer the ability to measure kinetics (e.g. binding to chromatin, cell membrane docking), absolute concentrations of molecules and underlying cell-to-cell variability.

Due to its submicrometer spatial resolution (typical detection volume $> 0.5\mu\text{m} \times 0.5\mu\text{m} \times 2\mu\text{m}$), FCS can be performed on precisely defined locations and can be exploited to detect individual macromolecules absorbed or bound to surfaces and even track biological function in real-time. For example, this strategy was applied to study conformational dynamics of DNA oligomers (Wennmalm, Edman and Rigler, 1997), horseradish peroxidase (Edman *et al.*, 1999) and flavin reductase (Yang *et al.*, 2003) at the single-molecule level in model systems. FCS has been applied in numerous studies over the past 15 years to investigate the dynamic behaviour of TF binding to chromatin *in vivo*. Some examples include, the binding dynamics of SOX2-OCT4 complexes on chromatin (Chen *et al.*, 2014), TetR (Normanno *et al.*, 2015), the specific and non-specific binding of the Hox TF Sex combs reduced (Scr) (Papadopoulos *et al.*, 2015) and the variable dynamic binding behaviour of MYC within nuclei (Rosales *et al.*, 2013). Additionally, FCS has been employed to study TF concentration and variability and how these parameters influence differential developmental fates. Some examples include the Senseless TF in the *Drosophila* wing imaginal discs (Giri *et al.*, 2020) and the transcriptional co-activator Yorkie in different subcellular compartments of the developing *Drosophila* airways (Skouloudaki *et al.*, 2019). Taken together, these studies demonstrate how the diverse applications of FCS can provide information on TF dynamic function and how this can complement qualitative methods such as chromatin immunoprecipitation and ChIP-seq.

Even though extensive genetic and phenotypic studies have been carried out regarding *PAX6* and *SOX2* pathogenic missense mutations, the dynamic behaviour at the molecular level *in vivo* (e.g. kinetics of binding to chromatin, subnuclear distribution, complex formation) remains unclear. Investigating the molecular behaviour of *PAX6* and *SOX2*, protein-protein and protein-chromatin interactions, and how pathogenic missense variants impede or influence this was a key aim in my PhD.

1.11 Thesis Aims

Extensive genetic and phenotypic studies for both *PAX6*, *SOX2* and their analogous mutants have been thoroughly characterised. Electrophoretic mobility shift assays (EMSA) have shown that both TFs can form tertiary complexes with DNA to regulate expression of key ocular genes such as the δ -crystallin gene during lens placode formation. Furthermore, qualitative methods such as chromatin immunoprecipitation and ChIP-seq have been able to identify and map the protein-chromatin interactome of both TFs at different developmental stages and link them with the regulation of a plethora of genes.

However, the dynamic behaviour between these two TFs at the molecular level *in vivo* (e.g. chromatin binding dynamics, subnuclear distribution, protein-protein interactions) and how human pathogenic missense mutations influence their molecular behaviour remains ill-defined. Therefore the overall aim of this thesis was to investigate in live samples, at the molecular level, the functional effect of human pathogenic missense mutations of both *PAX6* and *SOX2*, two TFs integral for normal eye development and maintenance. To accomplish this, FCS methodology was used, which allows for the detection of fluorescently labelled molecules at a single-molecule sensitivity. Due to its high temporal resolution, FCS is suitable for the study of fast dynamical processes such as protein-chromatin and protein-protein interactions.

The key objectives of this thesis were to:

1. Ensure that the fluorescently fused proteins were a good approximation as to their normal biochemistry

2. Investigate the biophysical properties of *PAX6* and *SOX2* pathogenic missense mutations in live cells and how this could link with the observed phenotypes in human patients
3. Assess the potential of *PAX6* in forming condensates and investigate how this could be linked with protein function
4. Investigate how missense mutations in *PAX6* and *SOX2* influence *PAX6/SOX2* complex formation and how this could contribute to the observed phenotypes in human patients
5. Generate an endogenously fluorescently labelled *PAX6* (GFP) and *SOX2* (mCherry) mESC line by using a CRISPR-Cas9 approach to investigate TF function in developing optic cup organoids

Chapter 2 Materials and Methods

2.1 Cell culture

2.1.1 HEK293 Cell line maintenance

HEK293 cell line was maintained in DMEM-F12 media (ThermoFisher), supplemented with 10% fetal calf serum (FCS) and 1% penicillin/streptomycin at 37°C and 5% CO₂ in T75 flasks. Fresh culture medium was added to cells every 3 days. Culture medium replacement was carried out gently, as HEK293 cells detached easily from the flask. Upon removal of old media, cells were carefully rinsed with 5ml PBS (phosphate-buffered saline) and 10ml of pre-warmed (in 37°C water bath) culture medium was added.

2.1.2 mESC Cell line maintenance

Mouse embryonic stem cells were maintained at 37 °C and 5% CO₂. The maintenance media composition can be seen in Table 2-1. A similar approach was conducted in terms of passaging, freezing and thawing mESC cells as with HEK293 cells.

500ml	GMEM
5.1ml	Non-essential amino acids (NEAA)
5.1ml	Sodium Pyruvate
5.1ul	2-b-mercaptoethanol
58ml	Knockout Serum Replacement (KOSR)
5.8ml	Fetal calf serum (FCS)
1 ml (1vial)	Leukaemia inhibitory factor (LIF)
5ml	Pen/Strep (optional)

Table 2-1: mESC maintenance media

2.1.3 Sub-culturing/passaging cells

HEK293 cells were passaged when 80% confluent at a ratio of 1:10 which was once a week. The culture medium was carefully aspirated using an aspiration pump. The cells were subsequently rinsed with 5ml PBS. The cells were then detached using 2.5 ml of Trypsin-EDTA solution (Sigma; T4174) and placed at 37°C for 2 minutes. Subsequently, 10ml of pre-warmed media were added to the flask and gently pipetted up and down, to obtain a single cell suspension. The cell suspension was placed in a 15ml falcon tube and centrifuged for 3 minutes at 1200 x g. The suspension was aspirated and replaced with 5ml of fresh media which was used to dissolve the cell pellet. Ultimately, 500µl of the suspended cells were added to a new T75 flask and supplemented with 9.5ml of fresh media.

2.1.4 Thawing/Freezing cells

Cells were recovered from liquid nitrogen storage and placed in a 37°C water bath until nearly (~80%) thawed. Before completely thawing, 1 ml of pre-warmed media was added and gently pipetted up and down. The suspension was then added to a 15ml falcon tube containing 10ml of pre-warmed media and centrifuged for 3 minutes at 1200 x g to pellet cells. Supernatant was then removed and replaced with 10ml of fresh pre-warmed media. Suspension was then added in a T75 flask and placed in an incubator at 37°C with 5% CO₂. The medium was replaced after 24 hours and cells were maintained as previously mentioned.

For long term storage of cells are frozen in liquid nitrogen. Cells from a confluent (80%) T75 flask are detached as previously described, centrifuged at 1200 x g for 3 minutes and the supernatant is removed. The cell pellet is resuspended in 6 ml of freezing mix (90% FCS, 10% DMSO) and 1 ml is added in each 2ml cryovial (Merck; BR114831). The cryovials are labelled with cell name, passage number and date and are placed in a Styrofoam™ rack. The cryovials are placed at -80 °C and 24hrs later are moved to liquid nitrogen for long term storage.

2.1.5 Transfection

The day before transfecting cells, a confluent (80%) T75 flask was split in 6-well plates at 70% confluence. The following day (when cells had attached), Effectene Transfection Reagent (QIAGEN; 301425) was used as per manufacturer's instructions.

Reagent	Quantity
Plasmid DNA	400ng
Effectene	10ul
Enhancer	3.2ul
EC Buffer	100ul
F12-DMEM media	2.6 ml

Table 2-2: The amount of each Effectene reagent which was used for cell transfection in a 6 well plate.

The transfection mix was left for 24 hours in an incubator at 37°C with 5% CO₂ at which point it was replaced with fresh media or split in an μ -Slide 8 well chambered coverslip (ibidi; 80826) at an 80% cell confluence (this was applied when conducting FCS experiments the following day).

2.1.6 DNA extraction

Genomic DNA from cells was extracted using the PicoPure DNA extraction kit (ThermoFisher; KIT0103) as per manufacturer's instructions.

2.1.7 Immunofluorescence

Confluent HEK293 cells were detached from a T75 flask, washed and resuspended in warm media. Coverslips were placed in a 10cm dish, sprayed with 70% ethanol and allowed to air dry in a tissue culture hood. Subsequently, one coverslip was carefully placed in each well of a six well plate and polylysine (ThermoFisher; A3890401) added (to cover the coverslip) for 30 seconds before being removed. PBS was then added (2ml) to prevent the coverslips from drying and placed into an incubator until the cells were detached.

Following cell detachment by trypsinisation, PBS was removed and replaced with 2ml pre-warmed media. 10^5 cells were added in each well and the plate was placed in an incubator O/N. The following day, cells were transfected using the Effectene Transfection Reagent and placed in the incubator O/N. The medium was removed and the cells were covered with 4% PFA diluted in PBS and incubated for 15 minutes at RT. The PFA was then removed and the cells were washed 3 times, for 3 minutes each, with PBS (carefully not to detach cells). The cells were then blocked for 60 minutes at RT (blocking buffer: 4% horse serum, 0.1% TritonX-100 in PBS). The blocking solution was removed and primary antibodies (1:500), diluted in blocking solution were added.

Antibody	Code
Donkey anti-mouse (Alexa 488)	abcam; ab150105
Donkey anti-rabbit (Alexa 568)	abcam; ab175470
Donkey anti-goat (Alexa 568)	abcam; ab175704

Antibody	Code
GFP (mouse)	Merck; G6539
Phospo-RNA Polymerase II (Ser5) (rabbit)	ThermoFisher; A304408AM
HP1 α (goat)	abcam; ab77256

Table 2-3: Antibodies used for immunofluorescent experiments

The cells were incubated O/N at 4°C on a shaker. The following day, the cells were washed 3 times with PBS for 3 minutes each time. Secondary antibodies (1:1000) were subsequently added and the cells were incubated for 1hr at RT on a shaker in the dark. The wells were washed 5 times with PBS for 10 minutes each on a shaker at RT. The coverslips were subsequently carefully removed using forceps, dried on tissue paper and mounted on a slide using 20 μ l of Vectashield (Vector Laboratories; H1000-10). For samples in which DAPI was used, it was diluted in PBS 1:10000 and added with the last PBS wash after the removal of the secondary antibody, then washed 3 further times.

2.1.8 Co-immunoprecipitation

Cells were transfected when 70% confluent in 10cm dishes. The following day, the transfection efficiency and relative fluorescent intensity was compared between samples.

Reagent	Quantity
Total Plasmid DNA	6000ng
Effectene	50ul
Enhancer	16ul
EC Buffer	500ul
F12-DMEM media	13 ml

Table 2-4: Amount of each Effectene reagent used for cell transfection in a 10cm dish

The growth media was subsequently removed and cells were gently washed with cold PBS. This was followed by addition of 3ml of cold PBS and cells were detached using a cell scraper (ThermoFisher; 179693). Cells were collected in 15ml falcon tubes and the volume was made to 10ml by adding cold PBS. The cells were spun for 3 minutes at 1200 x g and the supernatant was discarded. Cells were resuspended in 1 ml cold PBS and transferred to a 1.5 ml eppendorf tube and spun for 3 minutes at 1200 x g. For the preparation of the lysis buffer, protease inhibitor (1x) (Cell Signalling; 5871) and phosphatase inhibitor (1x) (Cell Signalling; 5870) were added to the Pierce IP lysis buffer (ThermoFisher; 87787). The PBS was removed from the cell pellet and resuspended with 500 µl of lysis buffer. The samples were subsequently placed on a shaker at 4°C for 15 minutes. The samples were then spun for 20 minutes at 13000 x g and 4°C. The supernatant was transferred to a new 1.5 ml tube and 50 µl from each sample were transferred to another tube (input) and stored at -20°C.

For immunoprecipitation, GFP-trap agarose beads (chromotek; gta-10) were carefully resuspended and 20 µl of bead slurry were added in a 1.5 ml eppendorf tube. This was followed by addition of 500 µl of Pierce IP lysis buffer and centrifugation at 2500 x g for 5 minutes at 4°C. The supernatant was

removed and replaced by the lysate and rotated end-over O/N at 4°C. The following day, the beads were sedimented by a 5 minute centrifugation at 2500 x g at 4°C and the supernatant was discarded. The beads were then resuspended in 500 µl of Pierce IP lysis buffer and again spun down at 2500 x g for 5 minutes at 4°C. This step was repeated twice. After the final centrifugation, the supernatant was removed up to the point where the beads were still covered by the lysis buffer and 25 µl of 4x LDS sample buffer (ThermoFisher; NP0007) and 5 µl of 1M DTT (ThermoFisher; R0861) were added. The sample was placed in a heat block at 70°C for 10 minutes. The sample was then placed on ice for 1 minute and centrifuged for 1 minute at 13000 x g. The supernatant was removed and transferred to a new tube and stored at -20°C.

2.1.8.1 SDS-PAGE

Samples were thawed on ice and centrifuged at 13000 x g for 1 minute and loaded on a NuPage 4 to 12%, Bis-Tris, 1.0 mm gel (ThermoFisher; NP0321PK2). Gels were run for 1 hour at 150V in SDS containing running buffer (ThermoFisher; LC2675).

2.1.8.2 Wet Transfer

A PVDF membrane was activated in 100% methanol for 1 minute. The membrane and gel were added in a sandwich format of filter paper which was soaked in transfer buffer (ThermoFisher; NP00061). The transfer was run at 30V for 1 hour. The membrane was washed with PBST (0.1% Tween) and blocked on a shaker at RT for 1 hour in 5% milk dissolved in PBST.

2.1.8.3 Protein Detection

After blocking, membranes were incubated on a shaker, O/N at 4°C with a primary antibody dissolved in BSA-PBS-Azide. The following day, membranes were washed with PBST (3x, 5 min on shaker, RT) and incubated with horseradish peroxidase (HRP) conjugated secondary antibodies in PBST on a shaker, for 1 hour at RT. This was followed by a 1 hour wash with PBST (changed buffer every 15 minutes). Western blotting luminal agent (Santa Cruz

Technologies; SC-2048) was used to detect proteins by chemifluorescence. The membrane was incubated with the agent for 1 minute at RT and signal was detected on Amersham hyperfilm ECL (ThermoFisher; 10534205).

2.2 Plasmid Construction

2.2.1 Electrocompetent cell preparation

The day before, 4L of sterile LB medium and 6L of Milli-Q H₂O (kept at 4°C) were prepared. One TOP10 bacteria colony was obtained from a streptomycin plate and inoculated in 100ml of LB streptomycin (50µg/ml) overnight at a 37°C shaker.

The following day two 1L Erlenhymer flasks were used to inoculate 500ml of LB with no antibiotic in each flask. A 1:80 dilution of the fresh overnight culture was added in each flask and placed in a shaker at 37°C for 4 hours or until the OD₆₀₀ was close but strictly below 0.5. During culture incubation, the Avanti JE centrifuge was pre-chilled at 4°C. In addition, 3L of 10% glycerol solution, 250 eppendorf tubes (1.5ml), 4 eppendorf storing boxes (-80°C), pipette tips (10ml and 20ml) and 250ml autoclaved centrifuge tubes were obtained and stored at 4°C. Following optimal culture density, flasks were placed on ice for 20 minutes. The culture solution was transferred to the 250ml centrifuge tubes and centrifuged for 15 minutes at 3000 x g and 4°C. Supernatant was subsequently discarded and the remaining culture solution was added and the step was repeated. The cell pellet was gently resuspended in pre-chilled 10% glycerol and centrifuged for 10 minutes at 3000 x g and 4°C. The supernatant was then removed and the step was repeated two more times. The cell pellet was then resuspended in 200ml of 10% glycerol and centrifuged one final time at 3000 x g and 4°C. Following supernatant removal, cell pellet was resuspended in 10ml of 10% glycerol solution and placed in 1.5 ml eppendorf tubes (40µl in each tube). This was conducted at 4°C (cold room). The tubes were flash frozen in liquid nitrogen and immediately transferred and stored at -80°C.

2.2.2 TOPO Cloning

Phusion High-Fidelity PCR Master Mix (ThermoFisher; F531L) was used for amplification of target sequences based on manufacturer's instructions. Primers were designed and ordered via Sigma-Aldrich. PCR product was mixed as per manufacturer's instructions with DNA Gel Loading Dye 6x (ThermoFisher; R0611) and loaded on a 1% agarose gel. The gel was run for 1 hour at 130 volts. A 1kb Plus DNA ladder (NEB; N3200L) was used as size reference. Following electrophoresis, the correct DNA band was excised using a scalpel and an open UV box. The gel fragment was placed in a 1.5 ml tube and weighed. Gel purification was carried out using the QIAquick Gel Extraction kit (QIAGEN; 28706X4) as per manufacturer's instructions.

Adenine overhangs were then added to the purified PCR fragment. Conditions are indicated below. The mixture was incubated in a thermocycler at 72°C for 20 minutes.

Purified PCR fragment	5 µl
Taq Polymerase (MERCK; D1806)	0.25 µl
dATP (5µM) (ThermoFisher; R0141)	0.25 µl
Taq Polymerase 10x buffer (MERCK; D1806)	0.5 µl

Table 2-5: Adenine overhang mixture

The mixture was placed on ice for 1 minute. For the ligation step, the TOPO TA cloning kit (ThermoFisher; K450002) was used based on manufacturer's instructions. The mixture was subsequently desalted using MF-Millipore Membrane Filters with a 0.025µm pore size (MERCK; VSWP02500). Milli-Q water was added in a bowl and the filter paper was placed on top. The ligation mixture was carefully added on the filter paper and left at RT for 45 minutes. Meanwhile, 32 µl (50mg/mL) of X-gal (ThermoFisher; B1690) was added on kanamycin agar plates and placed at 37°C. Electrocompetent cells were removed from -80°C and slowly thawed on ice. The desalted solution was then

mixed with the thawed electrocompetent bacteria and placed in a 0.1cm electroporation cuvette (ThermoFisher; P41050) on ice. A MicroPulser Electroporator was used at 1350V, 25mFarads and 600 ohms. Following electroporation, bacteria were rescued with 1ml pre-warmed LB broth and transferred to a 1.5 ml tube. The rescued bacteria were then placed in a shaking incubator at 37°C for 1hour. They were then plated on the X-gal, kanamycin agar plates and left O/N at a 37°C incubator.

2.2.3 Colony PCR and Plasmid Purification

Single, white, bacterial colonies were isolated using a pipette tip and transferred to a new, LB agar plate. Green Taq Master Mix (ThermoFisher; K1081) was used for the PCR reactions whereby 10µl were added in each 0.2 ml PCR tube. The pipette tip following the transfer of the bacterial colony to a new plate, was added in the PCR tube. The PCR mixture and thermocycler conditions were as per manufacturer's instructions. The positive control was remnants from the ligation product and the negative control was an empty pCR2.1 TOPO plasmid. The PCR products were subsequently loaded on a 1% agarose gel for standard gel electrophoresis. Colonies with positive PCR bands were grown overnight in 5ml LB at a 37°C shaking incubator. Plasmid purification was carried out using PureLink™ Quick Plasmid Miniprep Kit (ThermoFisher; K210010) as per manufacturer's instructions. Correct plasmids were further verified by restriction enzyme digestions and subsequently were sent for Sanger sequencing. Glycerol stocks (0.3ml 80% glycerol and 0.7ml LB culture) were made for correct plasmids, and stored at -80°C.

For subsequent plasmid constructions, restriction enzyme digestions were used to transfer sequenced inserts to the pEF1a-IRES NEO vector (28019#) whereby a similar approach was followed as previously described.

2.2.4 Site-directed mutagenesis

For single nucleotide changes 40bp complementary primers were ordered, whereby the changed nucleotide was in the middle of the primer. The following PCR protocol was used. Amplification was carried out for 24 cycles.

Volume	Reagent	Temperature	Time
0.5 ul	Forward Primer	98	3:00
0.5ul	Reverse Primer	98	0:30
1ul	Plasmid DNA	65	0:30
18ul	H ₂ O	72	4:00
20ul	Master Mix (Phusion HF)	72	5:00

Table 2-6: Reaction mixture and PCR conditions for site directed mutagenesis of single nucleotides

Following amplification, the PCR product was purified using QIAquick PCR purification kit (QIAGEN; 28104), DpnI digested per manufacturer's instructions (NEB), desalted and transformed in appropriate antibiotic LB agar plates. Colonies were cultured in 5ml LB (containing appropriate antibiotic), followed by plasmid purification and sent for Sanger sequencing.

When mutagenesis of more than one nucleotide, either deletion or addition of up to 40 nucleotides, a different PCR protocol was used. Complementary primers were again used. Overhangs of >30bp on either side were ordered when nucleotides had to be added whereas the primers for deletion of sequences were with 25bp homology arms on either side of the deletion. Due to large size of primers, all primers were HPLC purified. The first step was to set up two reaction tubes, whereby 500ng of template was added in each PCR tube. The forward primer was added in one reaction and the reverse primer in the other reaction tube. Phusion® High-Fidelity PCR Master Mix was used as per manufacturer's instructions in terms of PCR mixture. The two reaction tubes were placed in a thermocycler and amplified for 3 cycles. PCR conditions are shown in the table below.

STEP	TEMP	TIME
Initial Denaturation	98°C	30 seconds
(3 or 17) Cycles	98°C 65°C 72°C	10 seconds 30 seconds 4min
Final Extension	72°C	5min
Hold	4°C	

Table 2-7: PCR conditions for site directed mutagenesis for addition or deletion up to 40bp

The two PCR tubes were added together and 1µl of Phusion® High-Fidelity PCR Master Mix was added. The tube was placed in a thermocycler for a second round of amplification (17 cycles). PCR conditions are shown in the table above. The PCR product was subsequently digested with DpnI (NEB) as per manufacturer's instructions and subsequently purified using QIAquick PCR purification kit (QIAGEN; 28104). Template plasmid was also digested with DpnI to be used as a negative control for transformation. Following purification the same steps as previously mentioned were followed for transformation. The following day, if there was at least a 10x surplus of colonies when compared to the negative control, colony PCR was conducted to identify positive colonies. These were subsequently grown, purified and sent for Sanger sequencing.

2.2.5 Primer Sequences

The primers in the below table were used to construct mCitrine-PD, PD-mCitrine, mCitrine-PAX6 and SOX2-mCherry.

Oligo Name	Tm (°C)	Sequence
mCitrine forward (for mCit-PD)	79	actagtatggtgagcaagggcgaggagctgttca
mCitrine reverse (for mCit-PD)	74	ctcgagttttggacgcgccttgtagcgtc
mCitrine forward (for PD-mCit)	76.1	ctcgagatggtgagcaagggcgaggagctgttca
mCitrine reverse (for PD-mCit)	81.5	gcggccgcttaacgcgccttgtagcgtcgtccatgccg
PD forward (for mCit-PD)	74.4	ctcgagatgcagaacagtcacagcggagtgaaac
PD reverse (for mCit-PD)	76.2	gcggccgctcactgttgcctttcgtagccaggttg
PD forward (for PD-mCit)	75	actagtatgcagaacagtcacagcggagtgaaac
PD reverse (for PD-mCit)	76.3	ctcgagtcactgttgcctttcgtagccaggttg
PAX6 full length forward	79.9	ctactcgagatgcagaacagtcacagcggagtgaaac
PAX6 full length reverse	71.4	gatgcggccgcttactgtaattctggccagtattgag
mCherry Forward (for SOX2-mChe)	81.3	gaactcgagatggtgagcaagggcgaggaggata
mCherry Reverse (for SOX2-mChe)	85.8	gaagcggccgcttactgtacagctcgtccatgccg
SOX2 Forward	77.1	gaagctagcatgtacaacatgatggagacggagc
SOX2 Reverse	85.7	gaactcgagcatgtgtgagaggggcagtggtccg

Table 2-8: Table containing primers used for cDNA amplification. These primers were used to construct mCitrine-PD, PD-mCitrine, mCitrine-PAX6 and SOX2-mCherry

The PCR amplicons were ligated into pCR2.1 TOPO plasmid and transformed into electrocompetent cells. Colony-PCR was carried out to identify correct inserts, which were subsequently tested by restriction enzyme test digest (DreamTaq protocol). Correct digestions were confirmed by Sanger sequencing and the final constructs were stored as glycerol stocks at -80°C. Subsequently inserts were subcloned in the pEF1a-IRES NEO vector (Figure 2-2) and sequence verified. Transformed strains containing the correct sequence were stored as glycerol stocks at -80°C.

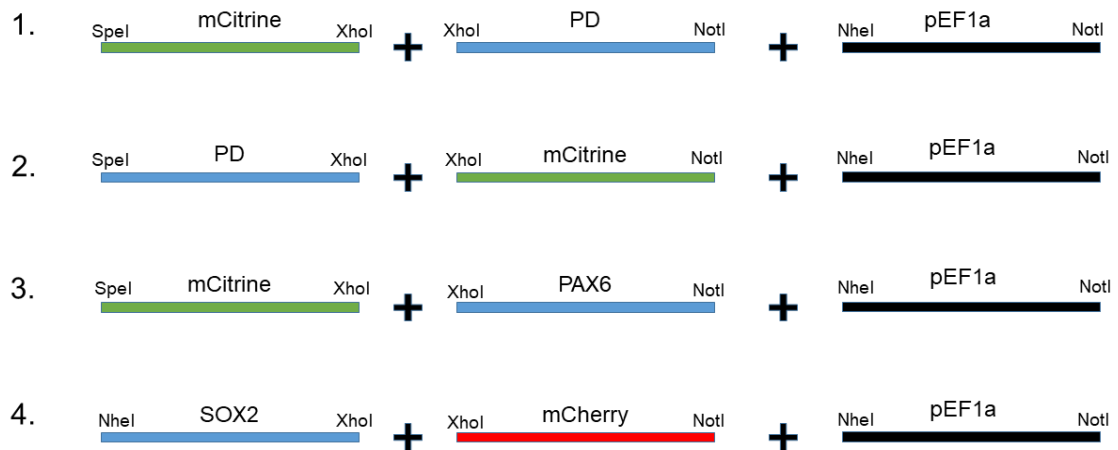


Figure 2-1: Schematic of PCR amplicons and restriction enzymes used for generation of final constructs

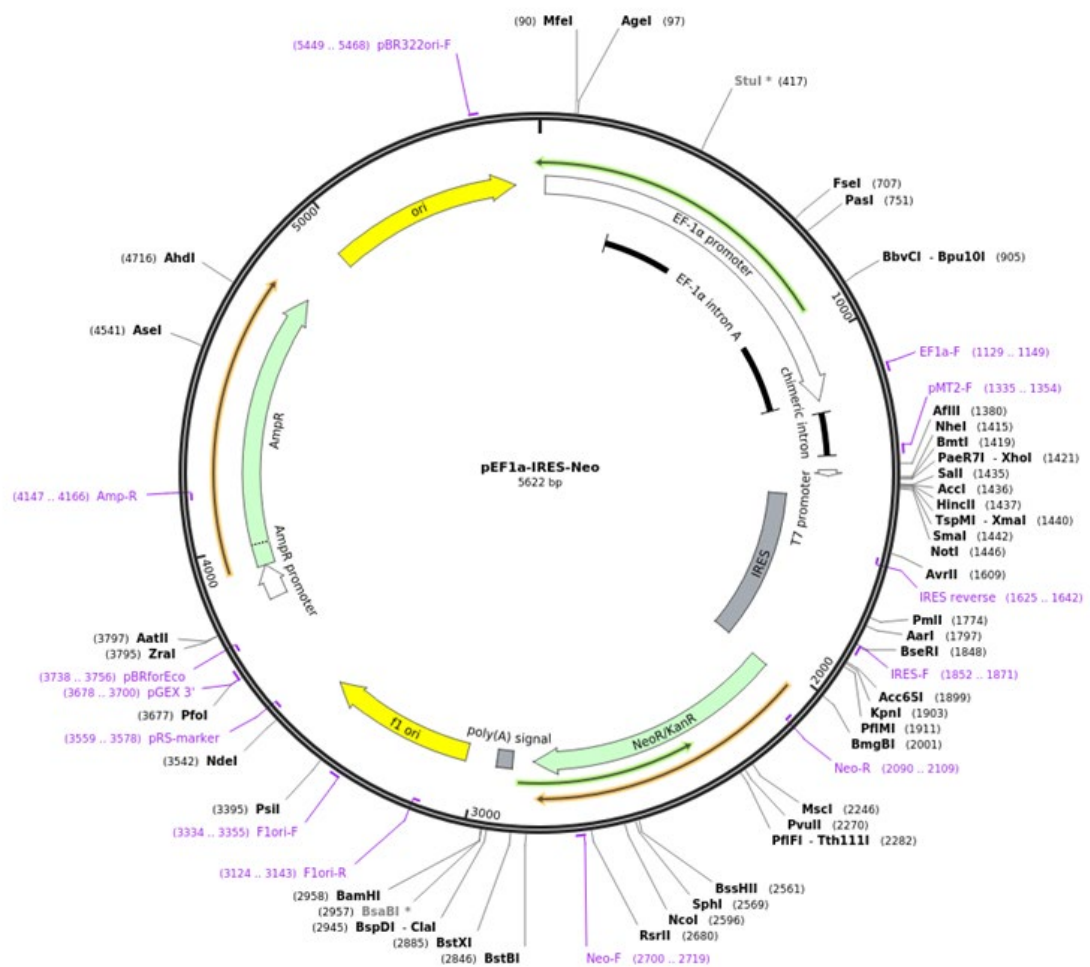


Figure 2-2: Illustration of pEF1a-IRES-Neo plasmid map. Adapted from Addgene 28019#

Subsequently, site directed mutagenesis was carried out using the following primers. These primers were used to generate the five PAX6 variants and the four SOX2 variants.

Oligo Name	Tm (°C)	Sequence
R74P Forward	89.8	caactcggagatcagcaagcccctggcgccgagtg
R74P Reverse	89.8	cactcggcgcccaggggcttgctgatctccgagttg
N46K Forward	84	caagcggcccatgaaggccttcattggtgtg
N46K Reverse	84	cacaccatgaaggccttcattggccgcttg
L97P Forward	87.6	gaggctaagcggcccgagcgctgcacatg
L97P Reverse	87.6	catgtgcagcgctcgcggccgcttagcctc
G130A Forward	97.3	gtacacgctgcccgccgctgctggccccggcggaatag
G130A Reverse	97.3	ctattgccgccccggggccagcagcgccgggcagcgtgtac
C52R Forward	78.1	cagaattttactcacacgtgtgagtaaaattctg
C52R Reverse	78.1	tgcaggtgtccaacggacgtgtgagtaaaattctg
S54R Forward	78.1	tacctgcccagaattttctcacacatccgttga
S54R Reverse	78.1	tccaacggatgtgtgaggaaaattctgggcaggtg
R92Q Forward	78.7	ggacgggcactcctgcttatactgggct
R92Q Reverse	78.7	agcccagtataagcaggagtgcccgcc
S121I Forward	78.7	cgaagaactctgtttattgataacacgcttggtatgtatcgt
S121I Reverse	78.7	acgataacataccaagcgtgttatcaataaacagagttcttcg
N124K Forward	78.4	tgcgaagaactctctttattgatgacacgcttggtatgt
N124K Reverse	78.4	acataccaagcgtgtcatcaataaagagagttcttcgca

Table 2-9: Table containing primers used for site directed mutagenesis for the generation of the PAX6 and SOX2 variants.

For generating the PAX6 isoforms, the following primers were used.

Oligo Name	Tm (°C)	Sequence
PAX6(5a) Forward	96.1	cgtgcgacattcccgaattctgcagacccatgcagatgcaaaagt ccaagtgtggaacaatcaaaacgtgtccaacggatgtgtgagtaa aattg
PAX6(5a) Reverse	96.1	gaattttactcacacatccgttggaacggtttgattgtccagcacttg gacttttgcactctgcatgggtctgcagaattcgggaaatgtcgc
PAX6(δ PD) Forward	81.6	gaactcgagatgggagcgagacggcatgtatgata
PAX6(δ PD) Reverse	82.3	gaagcggccgcctgtaatcttgccagattgagac
mCherry Forward	80.9	gaagctagcatggtagcaagggcgaggaggata
mCherryReverse	84.7	ttcctcgagctgtacagctcgtccatgccgcc

Table 2-10: Table containing the primers used for generating the PAX6 isoforms

For PAX6 (δ PD), the primers were used to amplify the cDNA from the pre-existing PAX6 construct. For PAX6 (5 α), a different approach was implemented. The pEF1 α mCitrine-PAX6 construct was used as an initial vector, to which the 5 α sequence was added by site-directed mutagenesis. The PCR mixture is shown in the table below.

Reagent	Volume(μ l)
Forward Primer	1
H ₂ O	18
Master Mix (Phusion HF)	20
Template	1 (500ng)

Reagent	Volume(μ l)
Reverse Primer	1
H ₂ O	18
Master Mix (Phusion HF)	20
Template	1 (500ng)

Table 2-11: PCR mixture for site-directed insertion of the 5 α sequence

Two reactions, one with forward primer and the other with the reverse primer, were set up. Three PCR cycles were carried out using parameters from Table 2-12. The two reactions were subsequently mixed and 2µl of Phusion HF Master mix was added. The initial amplicon was subsequently amplified for a further 18 cycles using the same PCR parameters.

PCR Step	Tm (°C)	Time (s)
Denaturation	95	60
Denaturation	95	5
Annealing	65	45
Elongation	72	180

Table 2-12: PCR parameters for the site-directed insertion of the 5α sequence in PAX6

Following amplification, the PCR mixture was incubated with DpnI (NEB; R0176S) for 2 hours at 37°C. The reaction was column purified (QIAGEN; 28104) following the manufacturer's instructions and transformed in TOP10 E.coli cells.

2.3 HEK293 Stable Cell line generation

Stable cell lines were generated by Flp recombinase-mediated integration using HEK293-Flp-In T-REX host cells (provided by Dr Hemant Bengani). Prior to stable cell line generation HEK293-Flp-In T-REX cells were maintained in F12-DMEM media (GIBCO) with 10% FCS, 1% Pen/Strep, 100µg/ml Zeocin (ThermoFisher; R25001), 15µg/ml Blasticidin (ThermoFisher; A1113903) and at 37°C and 5% CO₂ in T75 flasks.

PAX6 cDNA was amplified using primers containing attb1 sequences (underlined).

PAX6 Forward primer with attb1 site:

ggggacaagtttgtacaaaaaagcaggctacatgcagaacagtcacagcggagtgatcag

PAX6 Reverse primer with attb1 site:

ggggaccactttgtacaagaaagctgggtattactgtaatcttgccagttattgagac

The PCR product was run on a 1% agarose gel, excised, gel-purified and used in a BP Clonase reaction with pDONR221 (kindly provided by Dr Hemant Bengani). For the BP reaction, Gateway BP Clonase II Enzyme Mix (ThermoFisher; 11789100) was used following manufacturer's instructions. The reaction mixture was subsequently transformed into TOP10 E.coli bacteria and grown overnight on kanamycin (50µg/ml) agar plates. Following correct plasmid identification (Methods), an LR reaction was carried out. The Gateway LR Clonase II Enzyme mix (ThermoFisher; 11791020) was used following manufacturer's instructions. The destination plasmid used was pcDNA5/FRT/TO/GFP.Dest (kindly provided by Dr Hemant Bengani). The product of the recombination reaction was subsequently transformed in TOP10 E.coli bacteria and grown overnight on ampicillin (100µg/ml) agar plates.

When HEK293-Flp-In T-REX were 25% confluent (500,000 cells) in 10cm dishes, the medium was removed and replaced with medium without Zeocin. Effectene Transfection Reagent was used to cotransfect pOG44 plasmid (1.8µg) with pcDNA5/FRT/TO/GFP_PAX6 (200ng) at a ratio of 9:1. After 48 hours, the medium was changed and replaced with media containing 15µg/ml Blasticidin and 100µg/ml Hygromycin (ThermoFisher; 10687010). Cells were

maintained in an incubator at 37°C and 5% CO₂ and media was changed every 72 hours. Two weeks after transfection 1 µg/ml of tetracycline was added to the cells. The following day, cells were imaged using a fluorescent microscope to identify GFP positive cell colonies. Following colony identification, colonies were carefully removed using a P10 pipette tip and trypsin, transferred to a 6 well plate, and expanded.

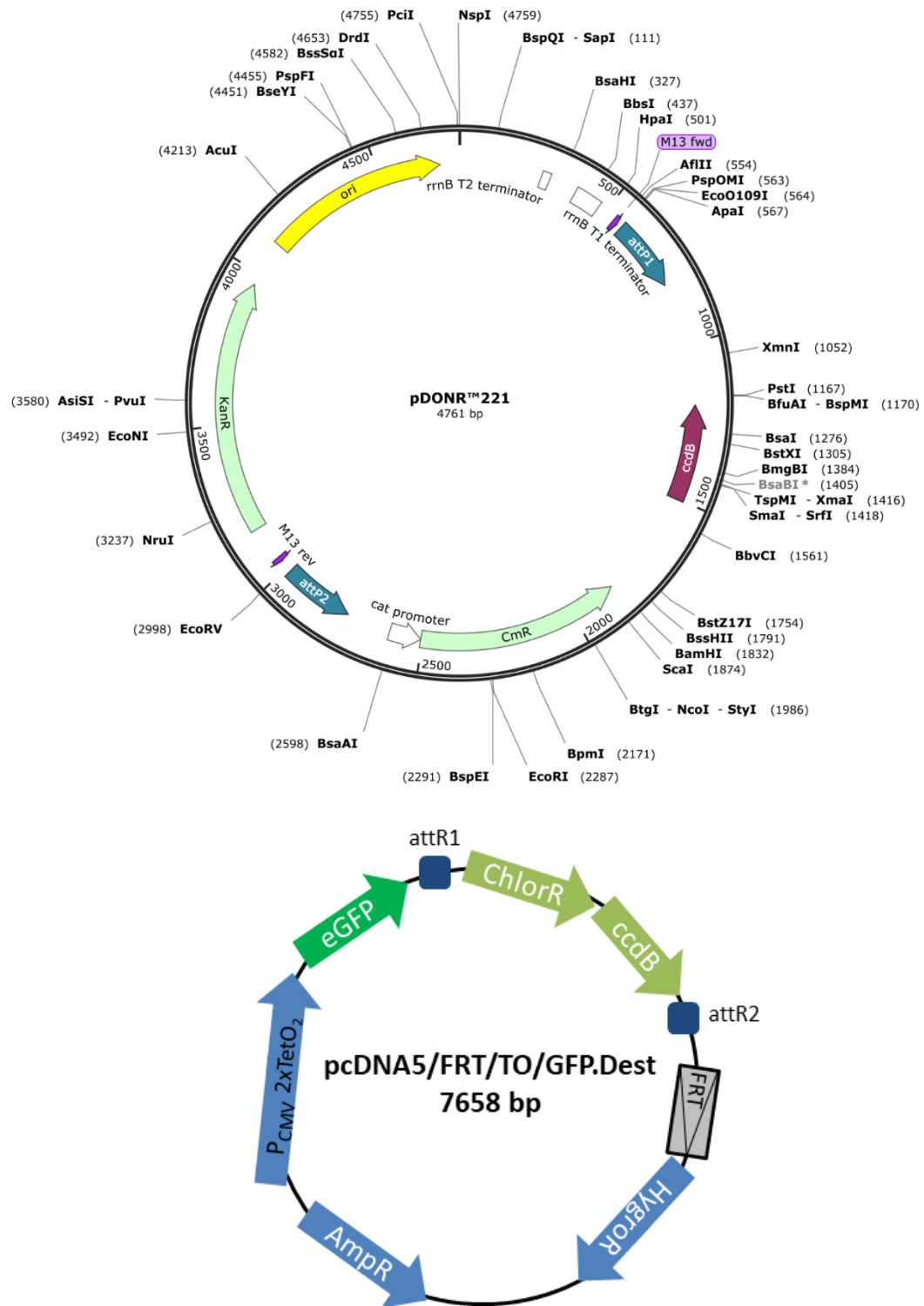


Figure 2-3: Maps of the plasmids used for generating the stable GFP-PAX6 TET HEK293 cell line

2.4 FCS

2.4.1 FCS measurements

Modified PicoQuant Leica SP5 microscope (40x water objective) was used for all experiments. The presence of Avalanche PhotoDiodes in this setup enabled close to single-photon detection. All live cell experiments were conducted at 37°C and 5% CO₂.

Cells were seeded the previous day in μ -Slide 8 well chambered coverslips (ibidi; 80826) at an 80% cell confluence. FCS measurements were performed by recording fluorescence intensity fluctuations in an ellipsoid observation volume element ($\sim 0.2\mu\text{m}$ wide and $1\mu\text{m}$ long) that is generated by focusing laser light through the microscope objective and collecting the fluorescence light through the same objective using a pinhole in front of the detector to block out-of-focus light. The Leica SP5 microscope with the 40x water objective was used for all experiments. All experiments were conducted at 37°C. Alexa 488 10nM dye solution was used to assess the detection volume prior to data acquisition. The expected counts per molecule second (cpms) obtained from an Alexa488, 10nM measurement at 20% argon laser power was approximately 40000 with the number of molecules being about 8. The autocorrelation was subsequently fitted using a triple extended (3D) fitting model with 1 diffusional and 1 triplet state component. The diffusion coefficient of Alexa488 is $435\ \mu\text{m}^2\text{s}^{-1}$ and a value as close to this was expected. Subsequently, the derived structural parameter of the confocal volume (κ) was recorded to be later used during data analysis.

In each measurement, fluorescence intensity fluctuations were recorded for 100 seconds. Measurements with excessive bleaching (i.e. a constant decrease in overall fluorescent intensity over 100 seconds) and/or low counts per molecule (cpms) (less than 1000) were not used for analysis. Variability between independent measurements reflects variability between cells, rather than imprecision of FCS measurements. All FCS measurements were carried out in the nucleoplasmic background avoiding any regions of enriched

fluorescence as shown by the red marker in Figure 3-4. Additionally, cells expressing the fluorophore in low amounts were selected for the measurements to avoid excessive photobleaching.

2.4.2 FCS analysis

The timetrace measurements were recorded using the SymphoTime 64 software. For measurement analysis, an intensity-fluctuations timetrace devoid of obvious photobleaching occurred was selected. The minimum timetrace window selected was 20 seconds. For autocorrelation curve calculation, 500 sample points were selected with a maximum lag time of 3000ms. Autocorrelation curves were fitted using a triple extended (3D) fitting model with 2 diffusional and 1 triplet state component. The structural parameter (κ) obtained from the Alexa 488nm measurement was kept constant in all fits of the same experiment. For each experimental sample, multiple measurements were recorded and then averaged to produce mean autocorrelation curves. One measurement was taken per cell. Measurements near the nuclear envelope were avoided. After fitting the autocorrelation curves, data was exported to Microsoft Excel where the autocorrelation curves, diffusion coefficients and the fractions of the total N molecules undergoing slow diffusion were averaged.

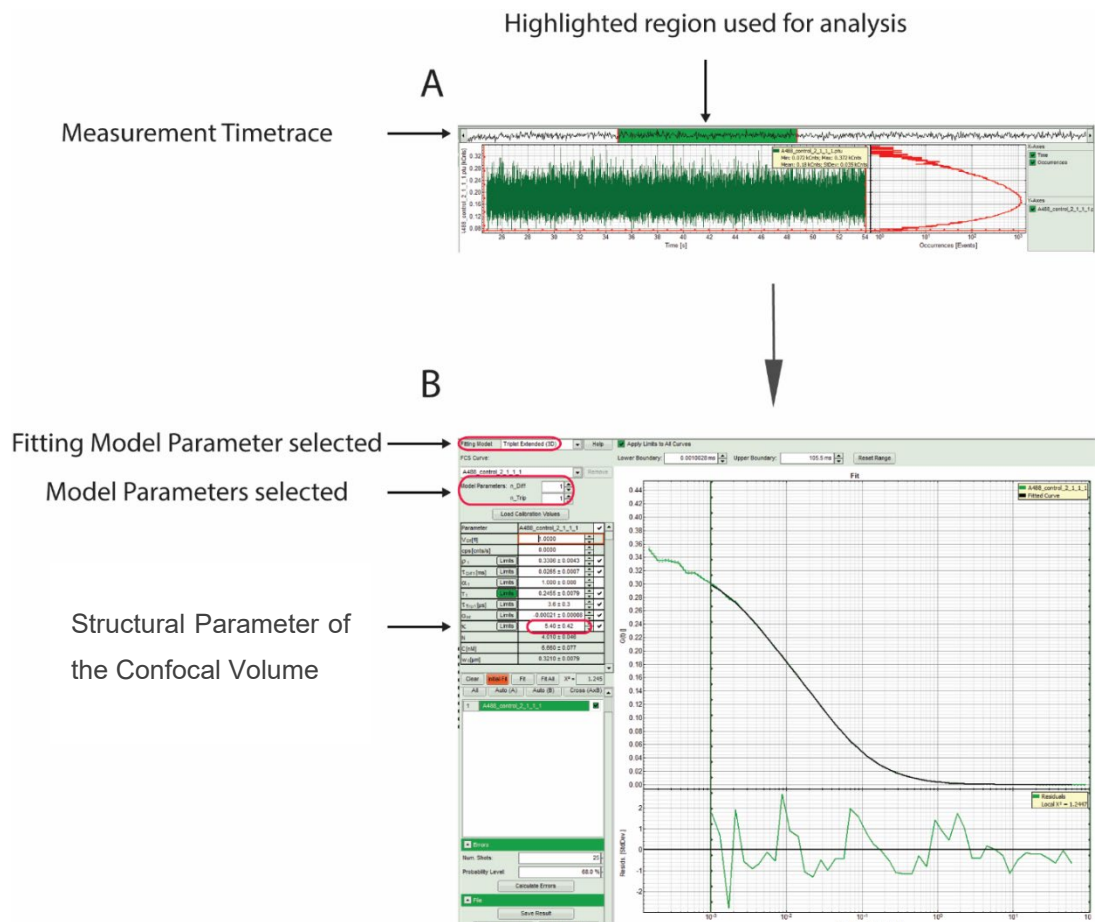


Figure 2-4: Representation of key parameters during measurement analysis using the SymphoTime 64 software A. Highlighting the region from the measurement timetrace used for analysis B. Fitting parameter values derived from fitting the calculated autocorrelation curves

2.4.3 FCCS measurements

The Leica SP5 microscope with the 40x water objective was used for all experiments. All experiments were conducted at 37°C. Alexa 488 10nM and Alexa 568 20nM dye solutions were used to assess the detection volume prior to data acquisition. The autocorrelation curves were subsequently fitted using a triple extended (3D) fitting model with 1 diffusional and 1 triplet state component. The diffusion coefficient of Alexa488 is $390 \mu\text{m}^2\text{s}^{-1}$ and Alexa568 is $400 \mu\text{m}^2\text{s}^{-1}$ and therefore a value as close to this was expected. Subsequently, the derived structural parameter of the confocal volume (κ) was recorded to be later used during data analysis. For the generation of the cross-correlation curves, the average structural parameter of the confocal volume (κ) of the two dyes was used. For fitting cross-correlation curves, a triple extended (3D) fitting model was used, with 2 diffusional and 0 triplet state components.

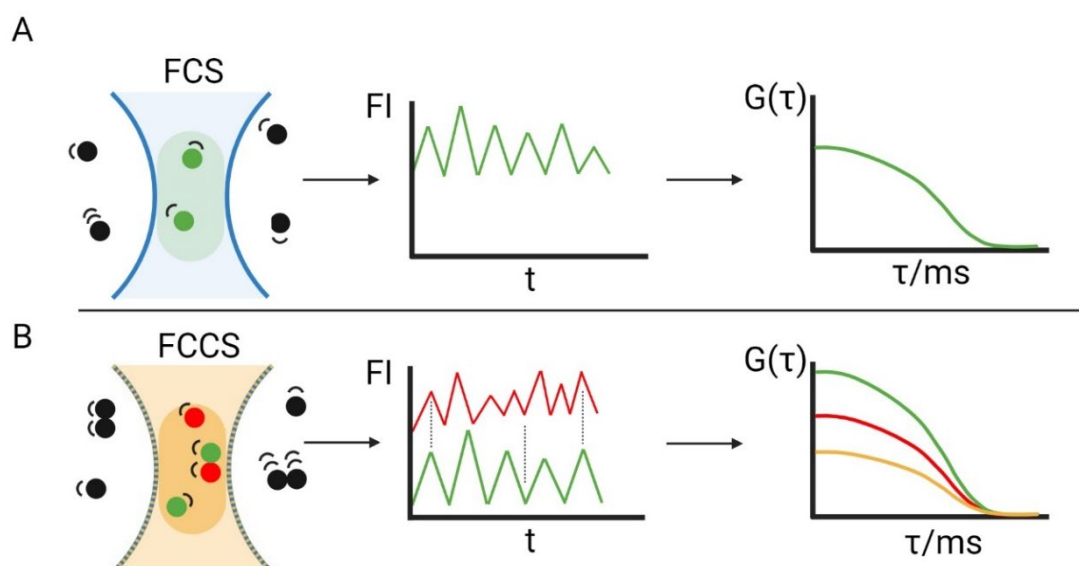


Figure 2-5: Schematic representation of FCS (A) and FCCS (B) measurements. Measurements are recorded as fluorescent intensity fluctuations over time and are subsequently plotted as autocorrelation curves. The orange curve represents the cross-correlation curve

In each measurement, fluorescence intensity fluctuations were recorded for 100 seconds. Measurements with excessive bleaching (i.e. a constant decrease in overall fluorescent intensity over 100 seconds) and/or low counts per molecule (cpms) (less than 1000) were not used for analysis. Variability between independent measurements reflects variability between cells, rather than imprecision of FCS measurements. All FCS measurements were carried out in the nucleoplasmic background avoiding any regions of enriched fluorescence. Additionally, cells expressing the fluorophores in low and similar concentrations were selected for the measurements to avoid excessive photobleaching and cross-correlation curve artefacts.

2.5 Condensate assays

2.5.1 Alexa fluorophore protein conjugation

Recombinant PAX6 was purchased (HTF-0109, LD BioPharma) at a concentration of 10.8 μ M and the recombinant PAX6 PD was kindly provided by Dr Nikki Hall. The protein buffer was replaced with 10 μ l PBS KCl (150mM) using Zeba MicroSpin Desalting Columns (ThermoFisher; 89877) following manufacturer's instructions. The recombinant protein was subsequently labelled with Alexa488 using the Alexa Fluor 488 Microscale Protein Labeling Kit (Invitrogen; A30007) following the manufacturer's instructions.

2.5.2 Droplet assay setup

The fluorescently labelled recombinant proteins were centrifuged at 20,000x g for 1 minute and subsequently mixed with 10% PEG-8000 (used as crowding agent) and PBS KCl (500Mm) in a final salt concentration of 150mM KCl and a total volume of 10 μ l in a 0.2 ml PCR tube. The protein solution was mixed by pipetting up and down multiple times and immediately loaded onto a homemade chamber comprising a glass slide with a coverslip attached by two parallel strips of double-sided tape. Slides were imaged using a Leica SP5 microscope with a 40x water objective. Following mounting, 1-6 hexanediol (40% stock concentration in PBS) was added using a syringe to a final concentration of 10%.

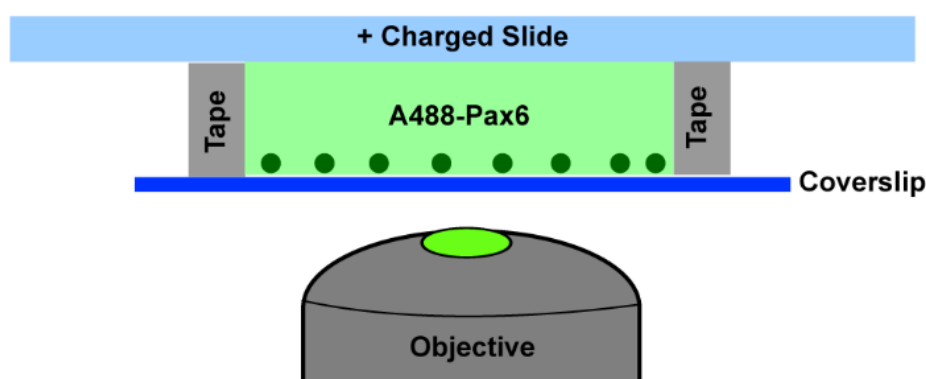


Figure 2-6: Schematic of the experimental setup for the *in vitro* experiments of recombinant PAX6 protein covalently conjugated to Alexa488

2.5.3 Condensate dissolution in live cells

Cells were transfected with 400ng of a plasmid expressing mCitrine-PAX6 (WT) using the Effectene Transfection Reagent. The following day, cells were split (100,000) in an 8-well glass slide (ibidi; 80826) at 50% confluence and experiments were carried out after 24 hours. The following day, the 8-well glass slide (ibidi; 80826) was mounted on a Leica SP5 microscope at 37°C and 5% CO₂. The medium was carefully removed using a P200 pipette and replaced with 200µl fresh pre-warmed medium containing the crowding and dissolving agent (agent and final concentration are stated in Results). Images were subsequently recorded every 60s.

2.6 PONDR Score Predictor and CIDER webserver analysis

Predictor of Natural Disordered Regions (PONDR) is an open-source bioinformatics software tool, which, based on the amino acid sequence of a protein, produces a predictive score of disordered regions. The CIDER webserver provides information on several parameters associated with disordered protein sequences. The PD amino acid sequences were uploaded on www.pondr.com and www.pappulab.wustl.edu/CIDER.com and graphs were exported.

2.7 Genome editing using CRISPR-Cas9

2.7.1 mESCs

The CRISPR-Cas9 system was used to genetically engineer mESCs. Custom synthetic crRNAs, tracrRNA and recombinant purified Cas9 were manufactured by Integrated DNA technologies (IDT). The Alt-R CRISPR-Cas9 system protocol from IDT was followed as per manufacturer's instructions. The RNA backbone and ends were chemically modified for protection against cellular RNases. The 36-mer crRNA contains a variable gene-specific 20-nt target sequence followed by a 16-nt sequence complementary to crRNA. The lyophilised crRNA and tracrRNA pellets were resuspended in Duplex buffer (IDT) at 100µM and stored in 20µl aliquots at -80°C. For sequence incorporation of a fluorophore protein (GFP or mCherry), 500 bp homology arms on either side were added and the construct was ordered from TWIST Bioscience in a form of a plasmid with ampicillin resistance. The guideRNA sequences were obtained from CHOPCHOP.

Upon RNP complex formation, mESCs were transfected as per manufacturer's instructions (Alt-R CRISPR-Cas9 system protocol; IDT). Upon detection of fluorescent nuclei (SOX2-mCherry) colonies were isolated and DNA was extracted. PCR using an internal primer (inside the fluorophore) and an external primer (outside homology arms) was used to verify the correct insertion of the fluorophore. Two external primers were used to validate if the insert was homozygous.

Oligo name	Tm (°C)	Sequence
SOX2 crRNA	64.6	gtgagggctggactgcgaac
SOX2 genomic mouse forward	59.8	gataagtacacgcttcccg
SOX2 genomic mouse reverse	72.5	gcacccctcccaattcccttg
mCherry reverse internal	69.3	gttcacacgcgctccac

Table 2-13. Table containing the guide RNA primer sequence and the primers used for genotyping mESCs for mCherry knock-in at the C terminus of genomic SOX2

2.7.2 Zebrafish

The CRISPR-Cas9 system was used to genetically engineer zebrafish embryos. All plasmid constructs, genotyping, husbandry and injections were carried out by the lab technician Jack Stoddart in which the GFP protein was fused to the N-terminus of the Pax6a cDNA. The gRNA sequences used can be found below:

Pax6a gRNA forward: tactataaccgggccacg

Pax6a gRNA reverse: cgtggcccgggtatagta

2.8 Confocal Imaging

Confocal images were acquired using the Leica SP5 microscope with the 40x water objective. Images were z-projections which were analyzed using ImageJ software.

Chapter 3 The molecular mechanism of PAX6 and SOX2 missense mutations in ocular diseases

3.1 Introduction

Over the past 20 years extensive work has been carried out to characterise the *PAX6* mutations which cause eye abnormalities such as aniridia, microphthalmia and/or coloboma in humans (Brown *et al.*, 1998). The majority of these studies has primarily focused on clinical characteristics observed in humans, however the molecular function/underpinnings of such mutations remain ill-defined. In this chapter, I have been able to link dynamic behaviour of TF molecules with phenotype severity in humans, focussing on 5 missense variants, all present in the PAX6 PD. The position of these five variants

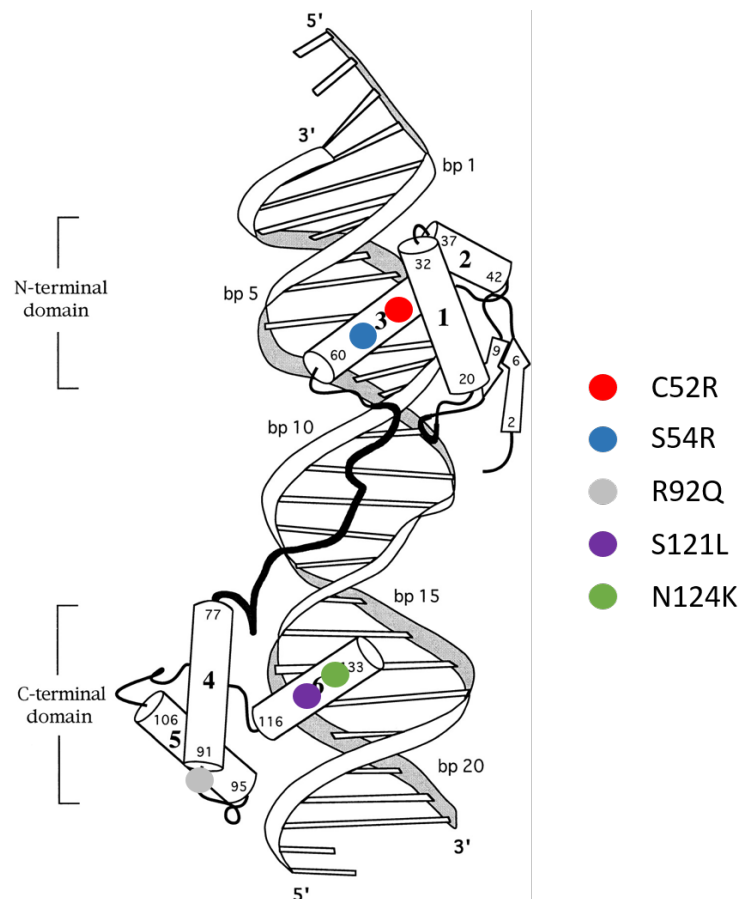


Figure 3-1: Schematic representation of the selected PAX6 variants on the PD in relation to DNA. The N-terminal domain is PAI and the C-terminal domain is RED.

(p.Ser54Arg, p.Asn124Lys, p.Cys52Arg, p.Ser121Leu and p.Arg92Gln) on the PD and how this relates to DNA interactions are shown in Figure 3-1.

The first four variants have been found in human patients suffering from ocular abnormalities such as bilateral microphthalmia and are in regions of the PD that directly interact with DNA. Therefore, it is likely that these four variants lead to a pathogenic phenotype due to alterations in PD-DNA binding dynamics. The only variant which does not appear to directly interact with the DNA backbone is R92Q. This variant (rs769095184) was initially assumed to be a neutral variant because it was present in eight East Asian individuals with an allele frequency of 0.0004349. Through structural predictions (FoldX) and EMSA characterization, R92Q was shown to moderately perturb the interaction with DNA and thus potentially be pathogenic (Williamson *et al.*, 2020). More recently, the variant was associated for the first time with classical aniridia in a Chinese family with four affected individuals, although limited segregation analysis where only *PAX6* cDNA sequencing was performed (Xiao *et al.*, 2019). The R92Q variant was initially aimed to be used as a neutral variant for FCS experiments, but based on the aforementioned publications it was no longer the case. Since R92Q was the only variant though which was predicted to not directly interact with DNA, it remained interesting to examine how it could potentially alter the overall *PAX6* chromatin-binding dynamics.

PAX6 Variant	Eye Phenotype	Other Phenotype	Reference
p.Ser54Arg	<ul style="list-style-type: none"> Bilateral microphthalmia Sclerocornea Right Congenital Aphakia 	<ul style="list-style-type: none"> Mild mental retardation Microcephaly Protruding ears Downward slanting palpebral fissures 	Williamson <i>et al.</i> , 2020
p.Asn124Lys	<ul style="list-style-type: none"> Bilateral microphthalmia Coloboma Congenital corneal opacification Retinal detachment 	<ul style="list-style-type: none"> Mild mental retardation Mild microcephaly Autism Epilepsy Congenital hyperinsulinism 	Williamson <i>et al.</i> , 2020
p.Arg92Gln	<ul style="list-style-type: none"> Congenital aniridia Cataract Horizontal Nystagmus 	<ul style="list-style-type: none"> None identified 	Xiao <i>et al.</i> , 2019
p.Ser121Leu	<ul style="list-style-type: none"> Aniridia Congenital ptosis 	<ul style="list-style-type: none"> None specified 	Chong <i>et al.</i> , 2018
p.Cys52Arg	<ul style="list-style-type: none"> Aniridia Cataract 	<ul style="list-style-type: none"> None specified 	Chao <i>et al.</i> , 2003

Table 3-1: Summary of phenotypes identified in patients with *PAX6* missense mutations

In this chapter, the chromatin-binding dynamics and subnuclear localisation of these 5 variants in live cells was investigated and compared to WT PAX6. Additionally, each of these missense variants in human patients demonstrate various degrees of pathogenicity implying differences in the importance of each amino acid residue for normal *PAX6* function (Table 3-1). For example, S54R and N124K have, so far, always been associated with bilateral microphthalmia (Williamson *et al.*, 2020), a condition considered more severe than classical aniridia (often observed in heterozygous *PAX6* deletions). Human patients carrying the aforementioned variants may also have iris defects, coloboma, congenital corneal opacification, retinal detachment, or primary aphakia (Williamson *et al.*, 2020). Both PD subdomains contain 3 α -helices whereby the 3rd helix is in direct contact with the major groove of the DNA. What is intriguing about these two missense variants is that they are both identified at approximately equivalent positions (in the 3rd helix within their subdomains (PAI and RED respectively) and both make direct contacts with the major groove of the DNA in similar relative positioning in the 5' and 3' regions of the DNA target binding site. The S121L variant was first identified in a patient with aniridia and multiple bioinformatic tools predicted the damaging effect on *PAX6* function (Chong *et al.*, 2018). The final variant, C52R, was first identified in 2003 in a patient with classical aniridia and cataract (Chao *et al.*, 2003).

SOX2 is also a key TF required for ocular development. Even though SOX2 can bind to enhancers and regulate gene expression as a monomer (Holmes *et al.*, 2020) it mostly operates as a heterocomplex. Its most studied binding partner is OCT4, with which it controls self-renewal and pluripotency (Goolam *et al.*, 2016; White *et al.*, 2016). During ocular development though, OCT4 is replaced by PAX6. SOX2 heterozygous mutations in humans also lead to aberrant ocular phenotypes with the most severe been anophthalmia (Faivre *et al.*, 2006). In this chapter, the dynamic behaviour of four SOX2 missense mutations, found in the HMG domain was investigated, and compared to WT SOX2. These were the N46K, R74P, L97P and G130A with the last variant found immediately outside of the HMG domain (Table 3-2).

SOX2 Variant	Eye Phenotype	Other Phenotype	Reference
p.Asn46Lys	• Anophthalmia	Severe hydrocephaly (fetus)	Faivre <i>et al.</i> , 2006
p.Arg74Pro	• Severe bilateral microphthalmia	Oesophageal atresia	Williamson <i>et al.</i> , 2006
p.Leu97Pro	• Microphthalmia • Coloboma • Sclerocornea	None specified	Ragge <i>et al.</i> , 2005
p.Gly130Ala	• Mild Iris Hypoplasia	None specified	Kelberman <i>et al.</i> , 2006

Table 3-2: Summary of phenotypes identified in patients with PAX6 missense mutations

The R74P mutation was first identified in a patient with severe bilateral microphthalmia (Williamson *et al.*, 2006). The arginine at position 74 is conserved among all SOX group B genes and plays a critical role in protein-DNA interaction (Williamson *et al.*, 2006). The N46K variant was first identified in a mosaic mother with bilateral anophthalmia although its presence in the first affected fetus could not be confirmed as no genetic material was obtained (Faivre *et al.*, 2006). Nevertheless, bioinformatic analysis predicted that the arginine at this position plays a critical role in SOX2 function. The L97P variant has been identified in a patient with microphthalmia (Ragge *et al.*, 2005), although once again, no further molecular analysis was carried out as to the reason of its pathogenicity. The G130A variant (rs121918652) is located immediately outside the HMG domain. The amino acid residue is conserved among humans, mice, zebrafish and chicken and was first identified in a patient with mild iris hypoplasia. In gel-shift assays, the G130A variant binds SOX2 binding sites with the same affinity as the WT SOX2 protein and as a result it was used as a control in our experiments (Kelberman *et al.*, 2006).

3.2 Results

3.2.1 Fluorophore position influences the molecular behaviour of PAX6

In order to use FCS, one must fluorescently label the target protein with either a fluorescent protein such as GFP, or with a fluorescent dye such as Alexa 488nm. In these experiments, proteins were fluorescently labelled with mCitrine, a fluorescent protein derived from GFP. Previous studies have demonstrated that the position at which a fluorophore is fused to a target protein may greatly influence the protein's function. To address potential disruptive effects on protein function, the PD was labelled at both the N-terminus and C-terminus and the molecular behaviour and subnuclear localisation was compared. These experiments were carried out to select which fusion protein was less damaging, in order to be used for future experiments. Prior to any measurements in cells, Alexa dye solution (10nM for Alexa 488 and 20nM for Alexa 568) was used as a calibration control to determine the structural parameters of the confocal detection volume (κ). Each FCS measurement was recorded for 100 seconds. Measurements with excessive bleaching (i.e. constant decrease in overall fluorescence intensity) and/or low counts per molecule (cpm), typically less than 1000, were not used for analysis. For all FCS experiments, each data point represented one measurement in one cell. Cells were selected based on their morphology and expression levels of the fusion protein. Cells with high fusion protein concentration ($1\text{mM} <$), were avoided as this could not be reliably measured by FCS.

3.2.1.1 PD (WT)

The average slow diffusion coefficients (Figure 3-2A) derived for the two tagged proteins were roughly equal, in which mCit-PD was $0.855 \mu\text{m}^2\text{s}^{-1}$ and PD-mCit was $0.837 \mu\text{m}^2\text{s}^{-1}$, suggesting that the position of mCitrine, did not influence the overall PD (WT) interaction with chromatin. The fraction of slow diffusing molecules though (Figure 3-2B) was significantly higher for PD-mCit. The fraction of slow diffusing molecules for mCit-PD and PD-mCit was 43.7% and 58.1% respectively. This difference was reflected in the normalised autocorrelation curve (Figure 3-2C), whereby a slight shift of the PD-mCit autocorrelation curve to shorter characteristic times was present. The normalised autocorrelation curve for each construct was derived by averaging all FCS measurements. For mCit-PD there were 12 FCS measurements from 12 cells and for PD-mCit 11 measurements from 11 cells respectively. Even though there was a difference in the fraction of the slow diffusing molecules, there was no significant difference in the slow diffusion coefficient nor the subnuclear localisation (Figure 3-2D) indicating that the location of the fluorophore did not significantly alter the overall molecular diffusion of the PD (WT). A similar comparison was carried out for the remaining PD variants to investigate whether they behaved similarly to the WT protein.

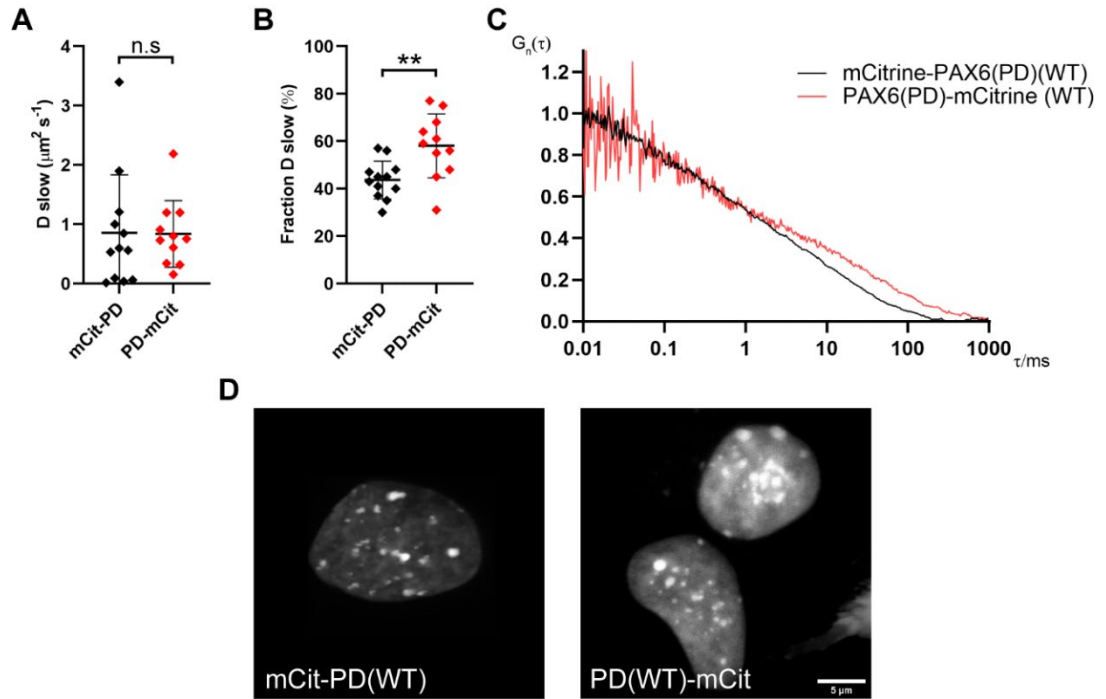


Figure 3-2: Comparison of PD (WT) when labelled at either N or C terminus with mCitrine. The autocorrelation function data were fitted with an established 3D diffusion mathematical model for obtaining the fractions and diffusion coefficients of the free moving (fast) and chromatin-bound (slow) molecules **A**. Slow diffusion coefficient comparison, as derived by fitting **B**. Fraction of the total N molecules undergoing slow diffusion **C**. Autocorrelation where $G(\tau)$ is normalised to 1 at the zeroth (0^{th}) lag time. **D**. Confocal images of the nuclear distribution of the PD of PAX6 in HEK293 cells 24 hours post-transfection. The image is representative of cells that were used for FCS measurements. mCit-PD $N_{cells}=12$, PD-mCit $N_{cells}=11$. One measurement per cell was carried out to minimize photodamage and areas of elevated concentration (foci) were avoided to prevent excessive photobleaching. Student's T-test ; p-value $<0.05=*$; p-value $<0.01=**$; p-value $<0.001=***$

PD (S54R)

The S54R mutant displayed both a different diffusion coefficient and fraction of slowly diffusing molecules. The diffusion coefficient of the slow mCit-PD (S54R) ($1.016 \mu\text{m}^2\text{s}^{-1}$) FCS component was significantly lower ($p\text{-value} < 0.01$) than the PD (S54R)-mCit ($1.861 \mu\text{m}^2\text{s}^{-1}$) suggesting that the fluorophore may influence the overall dynamic behaviour of the PD variant. Additionally, the fraction of the slow diffusing molecules was also significantly different ($p\text{-value} < 0.01$) between the two fusion proteins. For mCit-PD (S54R) the fraction was 53.8% whereas for PD (S54R)-mCit it was 42.3%. These results suggest that the mCit-PD (S54R) interaction with chromatin is stronger and/or longer-lasting, but also the number of chromatin-interacting molecules (fraction) is higher. The observed differences between the two constructs could only be attributed to the mCitrine position as this was the only difference between them. The normalised averaged autocorrelation curves (Figure 3-3C) reflected the observed differences in fraction and diffusion of the chromatin-interacting molecules as there was a pronounced shift of the mCit-PD (S54R) autocorrelation curve to longer characteristic decay times. The averaged autocorrelation curves were derived from the mean of all FCS measurements. For PD (S54R)-mCit there were 16 measurements from 16 cells and for mCit-PD (S54R) there were 14 measurements from 14 cells. In terms of subnuclear localisation, both variants demonstrated similar distribution with multiple foci of elevated concentration inside the nucleus. When compared to the PD (WT), no HEK293 cells expressing the PD (S54R) were identified that had these foci of elevated concentration.

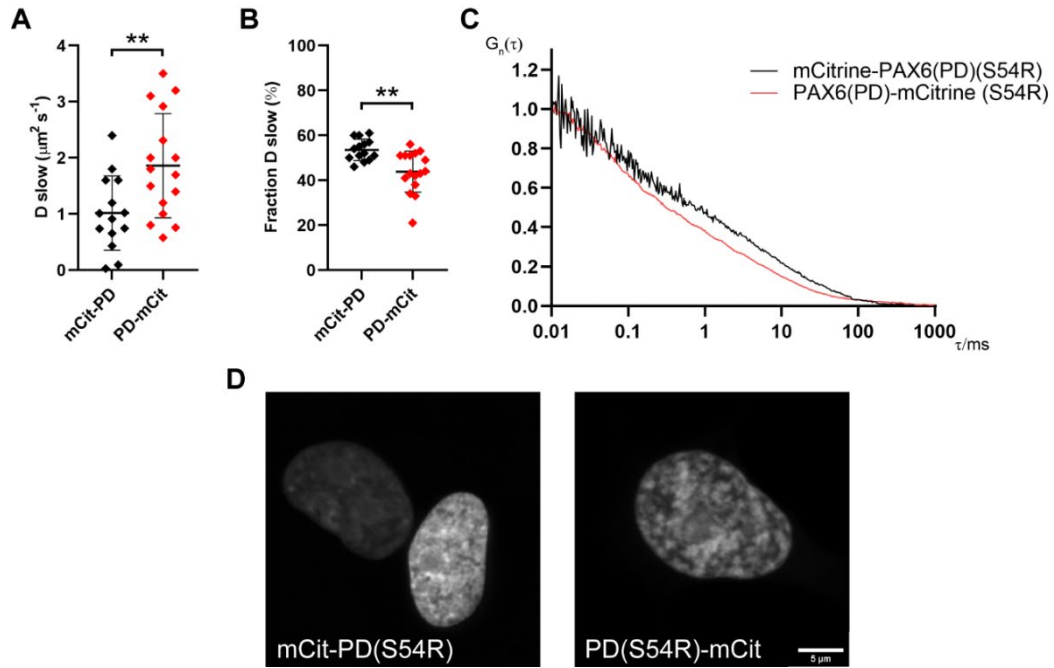


Figure 3-3: Comparison of PD (S54R) when labelled at either N or C terminus with mCitrine. The autocorrelation function data were fitted with an established 3D diffusion mathematical model for obtaining the fractions and diffusion coefficients of the free moving (fast) and chromatin-bound (slow) molecules **A**. Slow diffusion coefficient comparison, as derived by fitting **B**. Fraction of the total N molecules undergoing slow diffusion **C**. Autocorrelation where $G(\tau)$ is normalised to 1 at the zeroth (0^{th}) lag time. **D**. Confocal images of the nuclear distribution of PD (S54R) in HEK293 cells 24 hours post-transfection. The image is representative of cells that were used for FCS measurements. mCit-PD (S54R) $N_{\text{cells}}=14$, PD (S54R)-mCit $N_{\text{cells}}=16$. One measurement per cell was carried out to minimize photodamage Student's T-test ; p-value $<0.05= *$; p-value $<0.01= **$; p-value $<0.001= ***$

PD (R92Q)

The R92Q variant (rs769095184) was initially assumed to be a neutral variant as it was present in eight East Asian individuals with no recorded eye malformations. More recently though, the variant was associated for the first time with classical aniridia in a Chinese family with four affected individuals, although limited segregation analysis was conducted, where only *PAX6* cDNA sequencing was performed (Xiao *et al.*, 2019). Nevertheless, the variant was the only one which was not predicted to directly interact with DNA and was therefore still interesting to examine whether any diffusional differences could be observed when compared to the WT protein. Prior to comparing R92Q with the WT protein, a comparison between N and C terminal tagging of R92Q was carried out. For both, PD (R92Q)-mCit and mCit-PD (R92Q), 11 FCS measurements in 11 cells were used for the final analysis. When the diffusion coefficient of the slow moving molecules was compared, no major differences were observed, suggesting that the fluorophore position did not preferentially alter the diffusional profile of the chromatin-interacting molecules. Conversely a significant difference ($p\text{-value} < 0.001$) in the fraction of the slow moving molecules was observed when the two fusion proteins were compared (Figure 3-4B). The fraction for mCit-PD (R92Q) was 53.6% whereas for PD (R92Q)-mCit was 42.2%. Even though there was a significant difference between the two constructs, the normalised mean autocorrelation curves did not reveal any major differences between them (Figure 3-4C). In fact, the two autocorrelation curves overlapped, indicating similar behaviours, reflected by the comparable decay. The subnuclear localisation in HEK293 expressing cells was similar with the WT PD, which was not present in the other pathogenic variants.

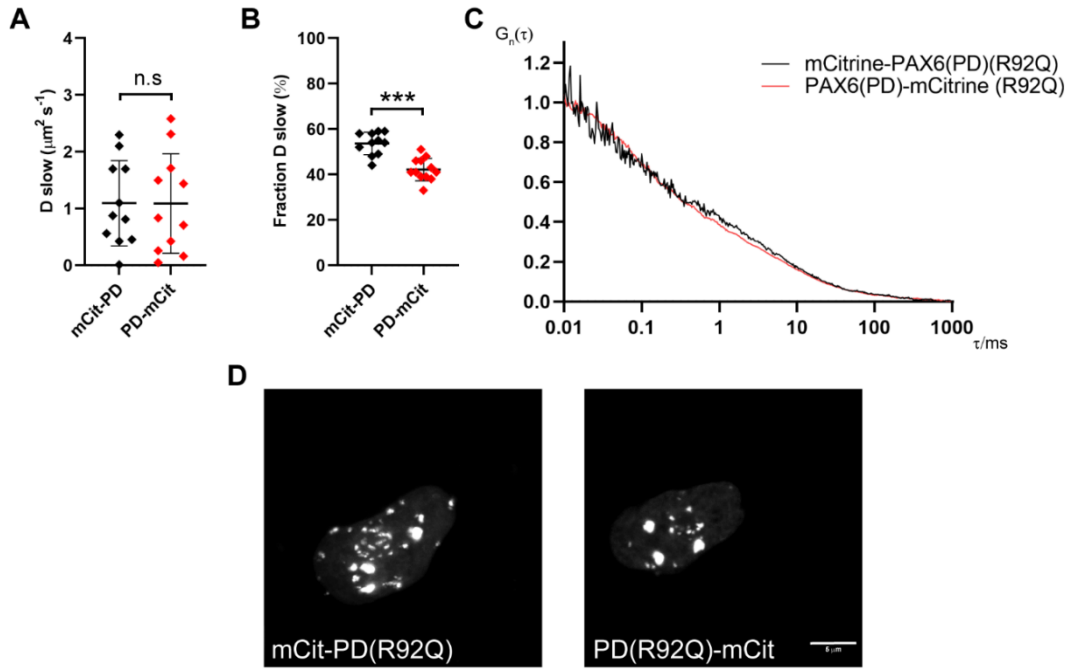


Figure 3-4: Comparison of PD(R92Q) when labelled at either N or C terminus with mCitrine. The autocorrelation function data were fitted with an established 3D diffusion mathematical model for obtaining the fractions and diffusion coefficients of the free moving (fast) and chromatin-bound (slow) molecules **A**. Slow diffusion coefficient comparison, as derived by fitting **B**. Fraction of the total N molecules undergoing slow diffusion **C**. Autocorrelation where $G(\tau)$ is normalised to 1 at the zeroth (0^{th}) lag time. **D**. Confocal images of the nuclear distribution of PD (R92Q) in HEK293 cells 24 hours post-transfection. The image is representative of cells that were used for FCS measurements. mCit-PD (R92Q) $N_{\text{cells}}=11$, PD (R92Q)-mCit $N_{\text{cells}}=11$. One measurement per cell was carried out to minimize photodamage and areas of elevated concentration (foci) were avoided to prevent excessive photobleaching. Student's T-test ; p-value $<0.05= *$; p-value $<0.01= **$; p-value $<0.001= ***$

PD (S121L)

The S121L variant is found in the 3rd α -helix of the RED subdomain, it is predicted to directly interact with DNA and has been associated with classical aniridia. Previous EMSA experiments showed that this variant completely disrupted binding to the SIMO enhancer, including a near 90% loss in binding to the Le9 enhancer (Williamson *et al.*, 2020). When comparing the overall molecular behaviour of the variant when fused either N- or C- terminally with mCitrine, no major differences were identified (Figure 3-5), suggesting that the position of the fluorophore did not preferentially influence the overall biophysical behaviour of the protein.

For mCit-PD (S121L), 10 FCS measurements in 10 cells were used for the final autocorrelation analysis and for PD (S121L)-mCit, 11 measurements in cells. Even though the standard deviation of the diffusion coefficient of the slow moving molecules (chromatin-interacting) was large (PD (S121L)-mCit $sd=3.23$, mCit-PD (S121L) $sd=3.22$) this was present for both. The observed variability between the independent measurements was a reflection of the variability between cells, rather than imprecision of the FCS measurements. There was no significant difference in the fraction of the slow-diffusing molecules when the two fused proteins were compared (Figure 3-5B) although the variability was once again substantial (PD (S121L)-mCit $sd=15.4$, mCit-PD (S121L) $sd=22.3$). Comparison of the mean normalised autocorrelation curves showed no major differences although for PD (S121L)-mCit, the autocorrelation curve was slightly shifted to the right, suggesting an overall slower diffusion. Considering though that no significant differences in fraction and diffusion coefficient of the slow-moving molecules were found, this difference was not considered to be major.

In terms of protein localisation, both variants were exclusively present in the nucleus and formed large foci. These foci did not appear similar to the WT TF, as they were fewer in number and considerably larger in size. Even though this was not experimentally shown, it was likely that this foci enrichment was in nucleoli.

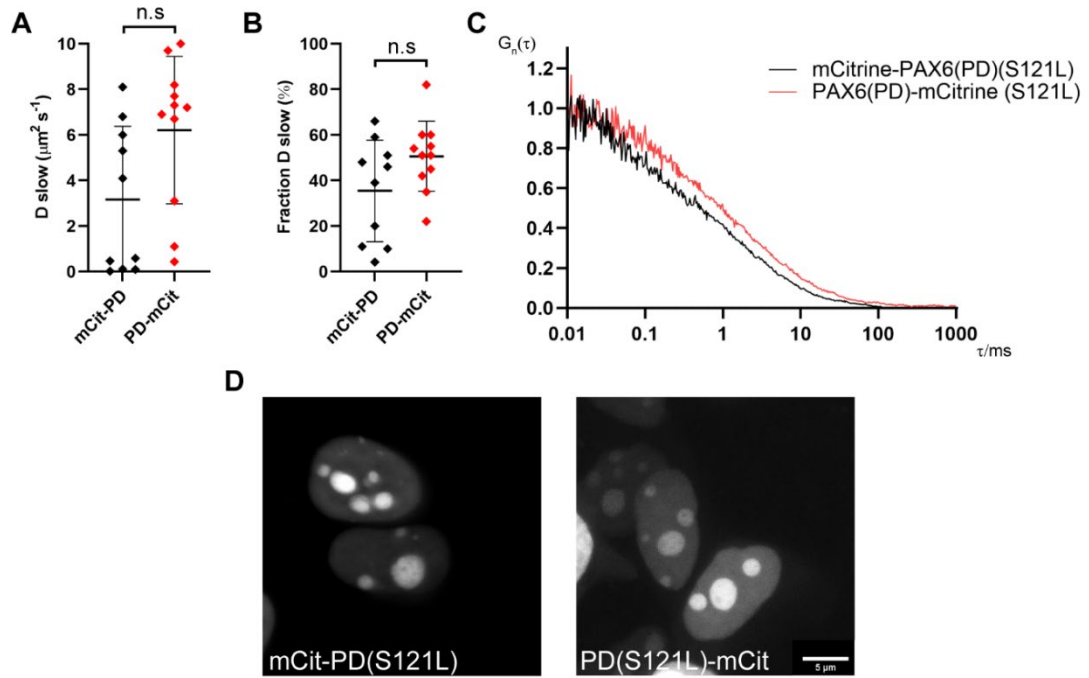


Figure 3-5: Comparison of PD (S121L) when labelled at either N or C terminus with mCitrine. The autocorrelation function data were fitted with an established 3D diffusion mathematical model for obtaining the fractions and diffusion coefficients of the free moving (fast) and chromatin-bound (slow) molecules **A**. Slow diffusion coefficient comparison, as derived by fitting **B**. Fraction of the total N molecules undergoing slow diffusion **C**. Autocorrelation where $G(\tau)$ is normalised to 1 at the zeroth (0^{th}) lag time. **D**. Confocal images of the nuclear distribution of PD (S121L) in HEK293 cells 24 hours post-transfection. The image is representative of cells that were used for FCS measurements. mCit-PD (S121L) $N_{\text{cells}}=10$, PD (S121L)-mCit $N_{\text{cells}}=11$. One measurement per cell was carried out to minimize photodamage and foci of elevated protein concentration were avoided to prevent excessive photobleaching. Student's T-test; p-value $<0.05= *$; p-value $<0.01= **$; p-value $<0.001= ***$

PD (N124K)

The N124K variant has previously been associated with severe bilateral microphthalmia, a phenotype considered to be worse-than-null (Williamson *et al.*, 2020). Initial comparison between N- and C- terminally fused mCitrine to PD (N124K) displayed striking differences in protein localisation. The mCit-PD (N124K) expressing HEK293 cells had similar subnuclear localisation as the S121L variant (enrichment in nucleoli). PD (N124K)-mCit though, appeared to be excluded from nucleoli, and enriched in the cytoplasm. Cytoplasmic localisation was not expected since the PD of PAX6 has two nuclear localisation signals (NLS) which ensure the protein is transported and maintained in the nucleus. For this reason, it was likely that when mCitrine was fused to the C-terminus of N124K, this potentially impaired the NLS functionality.

For mCit-PD (N124K), 14 FCS measurements in 14 cells were used for the final autocorrelation analysis and for PD (N124K)-mCit, 10 measurements in 10 cells. Although the N124K variants exhibited similar diffusion coefficients of the slow-moving molecules, the mCit-PD (N124K) had a significantly higher (p-value < 0.012) portion of these molecules. For mCit-PD (N124K) the percentage of molecules experiencing slow diffusion was 42.7%, whereas for PD (N124K)-mCit it was 19.6%. The observed difference was supported by a slight shift to longer characteristic decay times of the normalised averaged autocorrelation curve of mCit-PD (N124K) (Figure 3-6C).

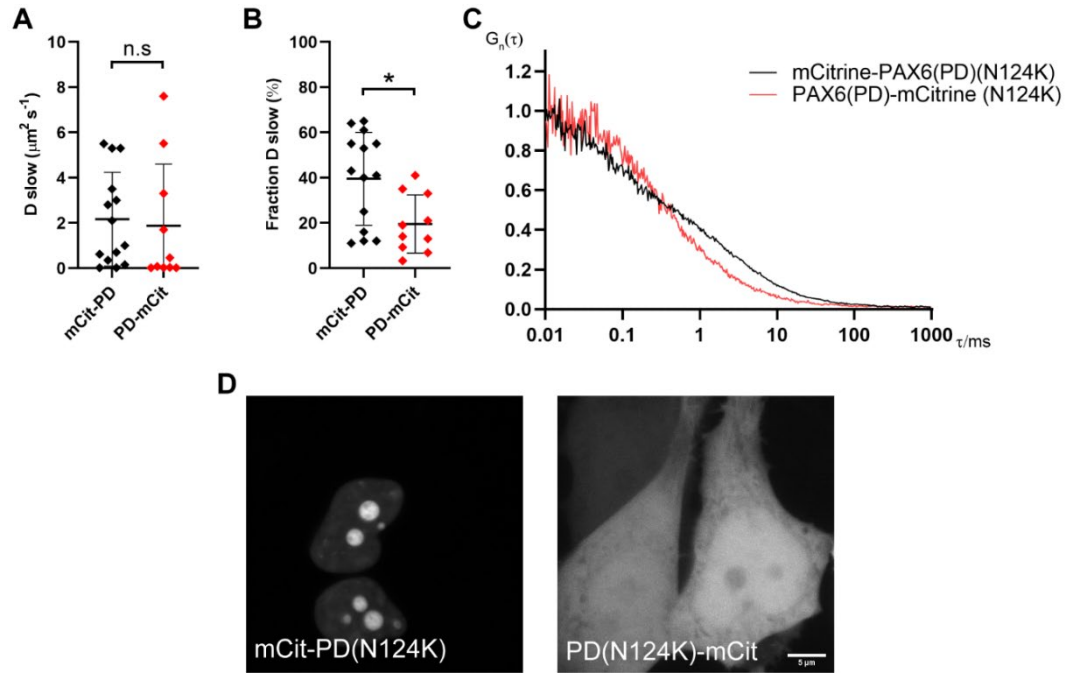


Figure 3-6: Comparison of PD(N124K) when labelled at either N or C terminus with mCitrine. The autocorrelation function data were fitted with an established 3D diffusion mathematical model for obtaining the fractions and diffusion coefficients of the free moving (fast) and chromatin-bound (slow) molecules **A**. Slow diffusion coefficient comparison, as derived by fitting **B**. Fraction of the total N molecules undergoing slow diffusion **C**. Autocorrelation where $G(\tau)$ is normalised to 1 at the zeroth (0^{th}) lag time. **D**. Confocal images of the nuclear distribution of PD (N124K) in HEK293 cells 24 hours post-transfection. The image is representative of cells that were used for FCS measurements. mCit-PD (N124K) $N_{\text{cells}}=14$, PD (N124K)-mCit $N_{\text{cells}}=10$. One measurement per cell was carried out to minimize photodamage and foci of elevated protein concentration were avoided to prevent excessive photobleaching. Student's T-test ; p-value $<0.05= *$; p-value $<0.01= **$; p-value $<0.001= ***$

PD (C52R)

Similar to S121L, the C52R variant has previously been characterized as a classical aniridia variant, whereby previous EMSA experiments have shown largely disrupted binding to PAX6 enhancers such as SIMO and LE9 (Williamson *et al.*, 2020).

Comparison between N- and C-terminally fused mCitrine to PD (C52R), there were some differences in protein localisation. PD (C52R)-mCit for example, was enriched in the nucleoli and was also present in the cytoplasm whereas mCit- PD(mCit) was not. In both cases though, a large number of small foci of elevated protein concentration were present in the nucleus, but these appeared different than those of the WT TF.

For mCit-PD (C52R), 12 FCS measurements in 12 cells were used for the final autocorrelation analysis and for PD (C52R)-mCit, 14 measurements in 14 cells. Although the C52R variants exhibited similar diffusion coefficients of the slow-moving molecules, the mCit-PD (C52R) had a significantly (p -value < 0.007) higher portion of these molecules. For mCit-PD (C52R), the percentage of molecules experiencing slow diffusion was 45.3%, whereas for PD (C52R)-mCit it was 26.7%. The observed difference was supported by a shift to longer characteristic decay times of the normalised averaged autocorrelation curve of mCit-PD (C52R).

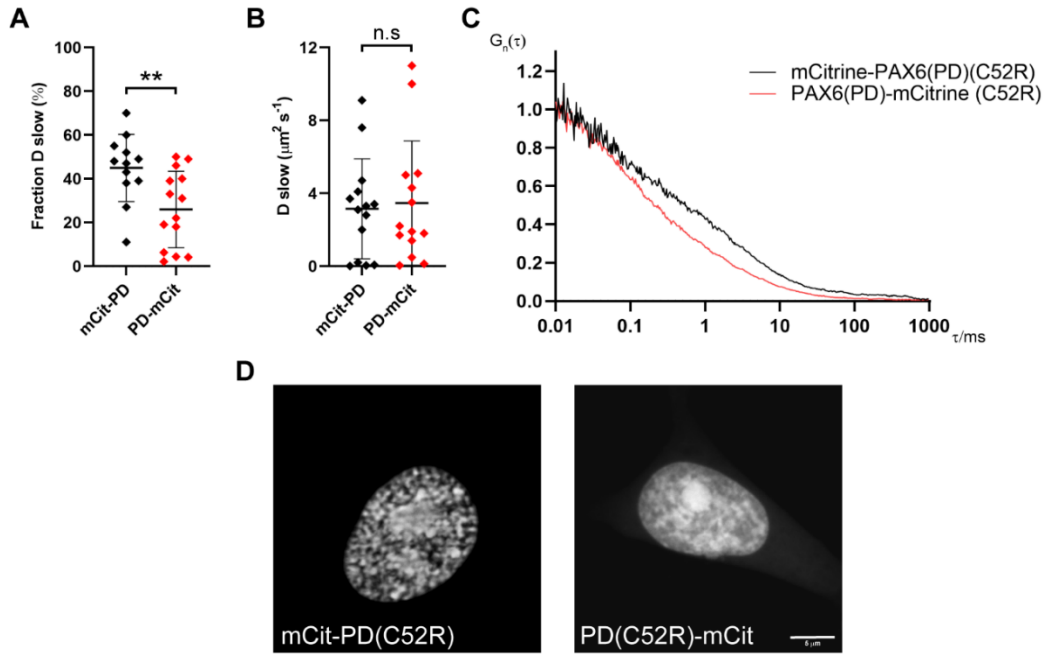


Figure 3-7: Comparison of PD (C52R) when labelled at either N or C terminus with mCitrine. The autocorrelation function data were fitted with an established 3D diffusion mathematical model for obtaining the fractions and diffusion coefficients of the free moving (fast) and chromatin-bound (slow) molecules **A**. Slow diffusion coefficient comparison, as derived by fitting **B**. Fraction of the total N molecules undergoing slow diffusion **C**. Autocorrelation where $G(\tau)$ is normalised to 1 at the zeroth (0^{th}) lag time. **D**. Confocal images of the nuclear distribution of PD (C52R) in HEK293 cells 24 hours post-transfection. The image is representative of cells that were used for FCS measurements. mCit-PD (C52R) $N_{\text{cells}}=12$, PD (C52R)-mCit $N_{\text{cells}}=14$. One measurement per cell was carried out to minimize photodamage and foci of elevated protein concentration were avoided to prevent excessive photobleaching. Student's T-test; $p\text{-value} < 0.05 = *$; $p\text{-value} < 0.01 = **$; $p\text{-value} < 0.001 = ***$

Overall, the aforementioned comparisons of the PD variants show that the position of the fluorophore influences the molecular behaviour, but not majorly. The subnuclear localisation of some variants though appeared to be notably influenced, namely of C52R and N124K in which mCitrine was fused at the C-terminus of the PD. A summary of the observations from each PAX6 variant can be seen in the table below.

	mCitrine N- vs C-Terminus Comparison		
PAX6 PD	Protein Localisation	Slow Diffusion Coefficient	Fraction of Slow Diffusing Molecules
WT	Similar	Similar	Significant Difference
N124K	Different	Similar	Significant Difference
R92Q	Similar	Similar	Significant Difference
C52R	Different	Similar	Significant Difference
S54R	Similar	Significant Difference	Significant Difference
S121L	Similar	Similar	Similar

Table 3-3: Table summarising the comparison of the molecular behaviour of PAX PD variants when tagging them with mCitrine at the N- or C- terminus.

3.2.2 PAX6 mutant variants exhibit weaker chromatin binding dynamics and altered subnuclear localisation

The variants were subsequently compared against the WT PD as to their subnuclear localisation and molecular behaviour. Since these variants are pathogenic in humans and mutations are found in key regions of the PAX6 PD (DBD) I hypothesised that the overall mobility of the variants would be faster based on the fact that they would have shorter-lived and weaker interactions with chromatin.

PD-mCit WT vs mutants

A striking difference in both the subnuclear localisation (Figure 3-8B) and the overall diffusion was observed upon comparing the variants with the WT PD (Figure 3-8C-F). The constructs were all expressed at fairly similar concentrations, as can be seen in Figure 3-8D, whereby concentration is inversely proportional to the amplitude of the autocorrelation at the 0th lag time. The fluorescence was solely observed in the nucleus, as expected, but for C52R and N124K there was substantial fluorescence found also in the cytoplasm. The foci observed in both the WT and R92Q variant could possibly play a role in the correct function of the PD since the pathogenic variants did not appear to form them.

The normalised mean autocorrelation curves (Figure 3-8C) showed that the WT PAX6 PD displayed the longest decay time, suggesting that it formed the strongest and longest-lived complexes with chromatin when compared to the other variants. Even though the R92Q mutation shared a similarly slow diffusion coefficient with the WT (Figure 3-8F), the relative fraction of slowly diffusing molecules was significantly lower (Figure 3-8E), potentially explaining the slight but obvious shift of the autocorrelation curve to shorter characteristic times.

In terms of its molecular behaviour, N124K, showed a somewhat slow molecular diffusion indicative of chromatin binding. This may indicate why a worse than loss of function phenotype is observed in humans. In order to verify this assumption a ChIPSeq experiment investigating the PAX6 binding sites could possibly provide further evidence. The S54R variant showed significantly faster molecular diffusion when compared to the WT but slower than S121L and C52R, indicative of some form of chromatin binding. In both cases though, the slow FCS fraction of both variants was notably lower than the WT. The S121L and C52R mutant variants demonstrated significantly faster diffusion and with a smaller fraction of bound molecules, indicative of markedly weaker and shorter-lived chromatin interactions in comparison with the WT.

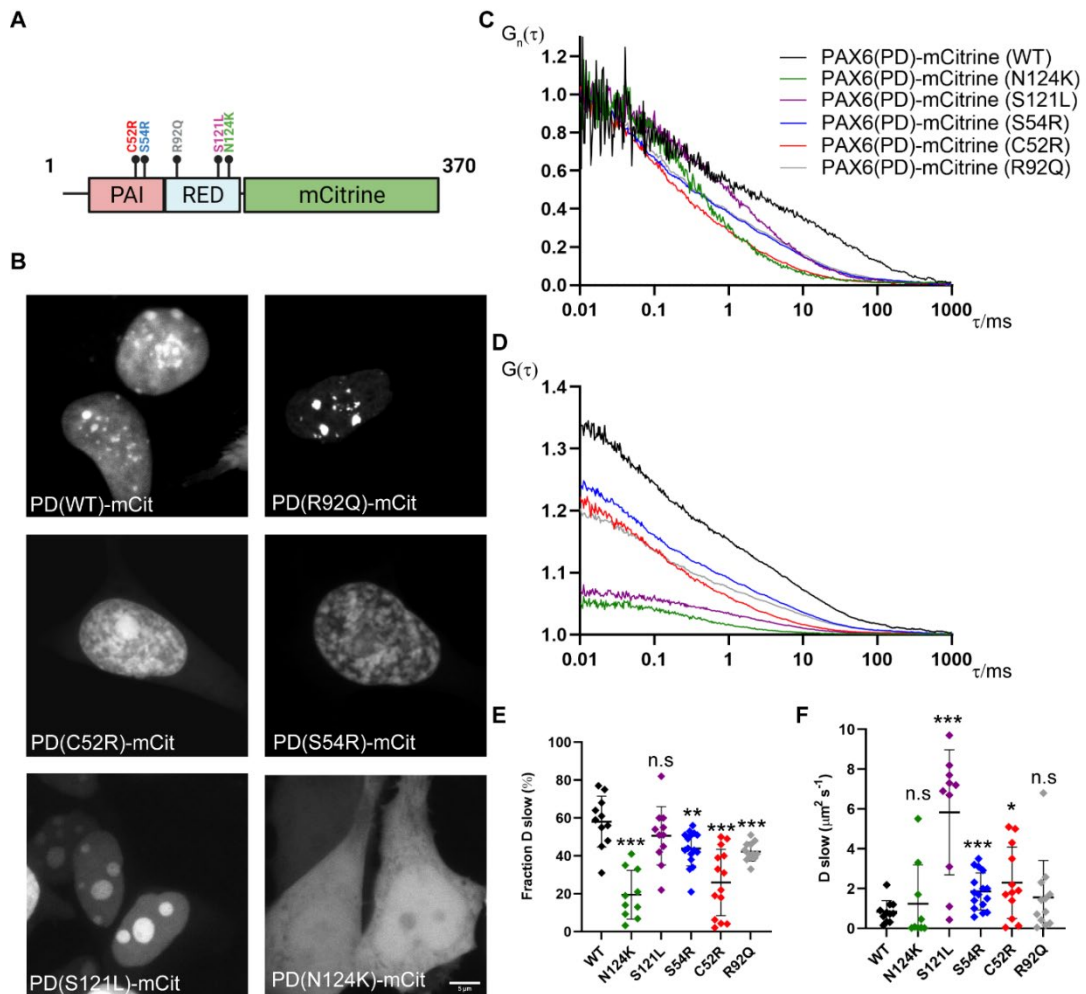


Figure 3-8: Comparison of PD-mCit between WT and missense variants. The autocorrelation function data were fitted with an established 3D diffusion mathematical model for obtaining the fractions and diffusion coefficients of the free moving (fast) and chromatin-bound (slow) molecules **A**. Schematic illustration of fused PD with mCitrine and variant annotation **B**. Confocal images of the subnuclear localisation of the PD fused protein in HEK293 cells 24 post-transfection **C**. Mean autocorrelation curve whereby $G(\tau)$ is normalised to 1 at the zeroth (0th) lag time **D**. Mean autocorrelation curves derived from FCS measurements **E**. Scatter plot comparing the fraction of the slow diffusion coefficient **F**. Scatter plot comparing the slow diffusion coefficient ; The images are representative of cells that were used for FCS measurements. PD (WT)-mCit $N_{\text{cells}}=11$, PD (N124K)-mCit $N_{\text{cells}}=10$, PD (S121L)-mCit $N_{\text{cells}}=11$, PD (S54R)-mCit $N_{\text{cells}}=16$, PD (R92Q)-mCit $N_{\text{cells}}=11$, PD (C52R)-mCit $N_{\text{cells}}=14$. One measurement per cell was carried out to minimize photodamage and foci of elevated protein concentration were avoided to prevent excessive photobleaching. Student's T-test versus PD (WT); p-value <0.05 = *; p-value <0.01 = **; p-value <0.001 = ***

A similar trend was observed when the variants were labelled with mCitrine at the N terminus although the differences between them were less pronounced (Figure 3-9). The subnuclear localisation of R92Q and the WT were again very similar, with higher fluorescence intensity foci being present. In the case of the other pathogenic mutants, two distinct types of behaviour were observed. Mutants N124K and S121L showed a higher fluorescence intensity in large foci, as opposed to the rest of the nucleus, and mutants C52R and S54R showed a 'grainy' pattern evenly distributed throughout the nucleus. In all cases no pathogenic mutant protein had a similar subnuclear localisation to the WT, which may indicate lack of functionality. The major difference between N- and C-terminal tagging of mCitrine was observed for the N124K mutant. Upon mCitrine fusion to the C terminus, there was a large amount of the protein residing in the cytoplasm and the nucleolus showed a lower fluorescence than the rest of the nucleus (Figure 3-6).

Although N-terminally tagging provided similar subnuclear localisation signals, the autocorrelation curves did not display as pronounced differences between the mutants and WT (Figure 3-9C). When the values of the slow diffusion coefficient were plotted (Figure 3-9F) there was a significant difference between the WT and the three mutants S121L, N124K and C52R. The relative amplitude/fractions of slow diffusing molecules did not display as great a difference as before. Only S54R and R92Q variants were significantly different when compared to the WT and, surprisingly, formed a higher proportion of chromatin bound complexes than the WT, possibly owed to the substantially smaller standard deviation of S54R and R92Q. The FCS curves overlapped for short characteristic times, suggesting similar free diffusion between the mutants and the WT (0.01-0.8ms range; Figure 3-9C). In the case of C terminus tagging though this was not the case.

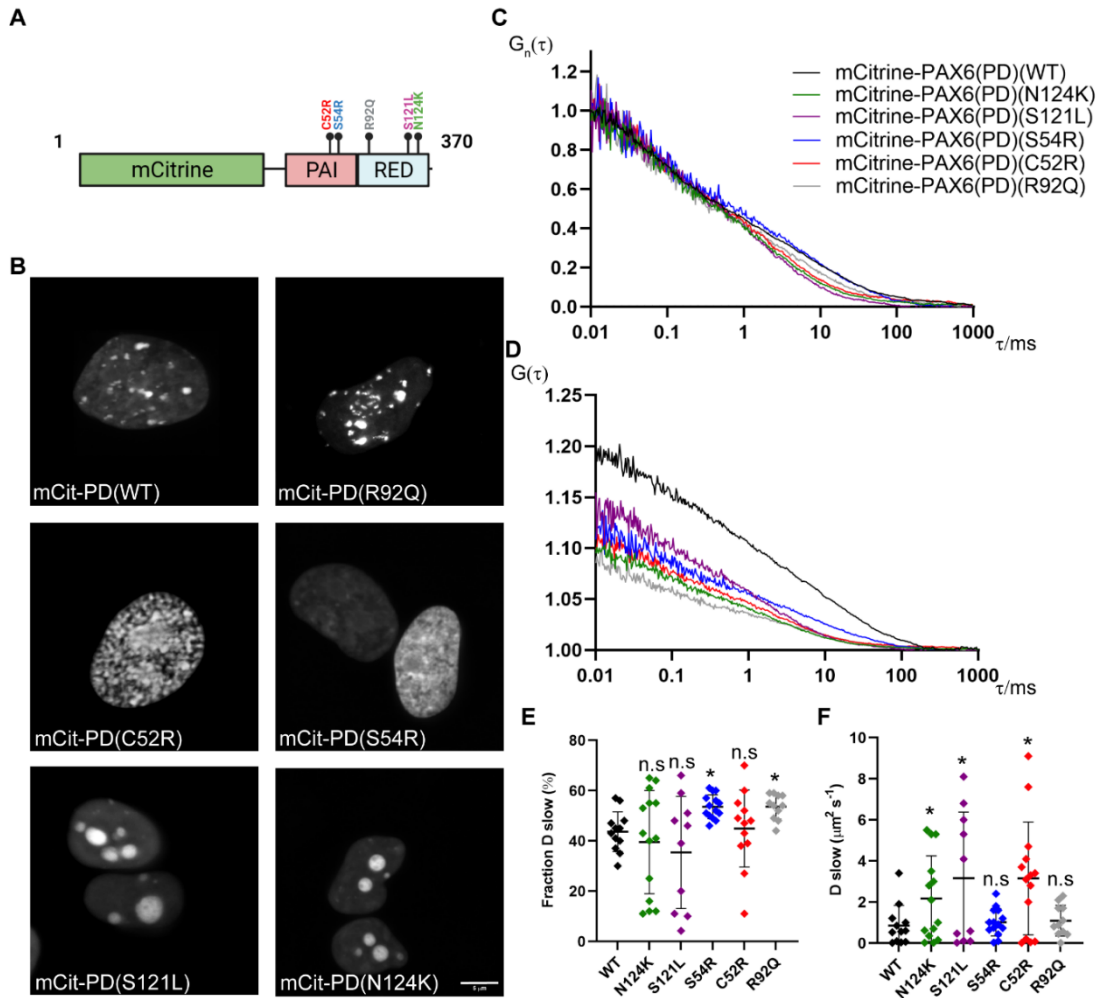


Figure 3-9: Comparison of mCit-PD between WT and missense variants. The autocorrelation function data were fitted with an established 3D diffusion mathematical model for obtaining the fractions and diffusion coefficients of the free moving (fast) and chromatin-bound (slow) molecules **A**. Schematic illustration of fused PD with mCitrine and variant annotation **B**. Confocal images of the subnuclear localisation of the PAX6 PD fused protein in HEK293 cells 24 post-transfection **C**. Mean autocorrelation curve whereby $G(\tau)$ is normalised to 1 at the zeroth (0th) lag time **D**. Mean autocorrelation curves derived from FCS measurements **E**. Scatter plot comparing the fraction of the slow diffusion coefficient **F**. Scatter plot comparing the slow diffusion coefficient ; The images are representative of cells that were used for FCS measurements. mCit-PD (WT) $N_{cells}=12$, mCit-PD (N124K) $N_{cells}=14$, mCit-PD (S121L) $N_{cells}=10$, mCit-PD (S54R) $N_{cells}=14$, mCit-PD (R92Q) $N_{cells}=11$, mCit-PD (C52R) $N_{cells}=12$. One measurement per cell was carried out to minimize photodamage and foci of elevated protein concentration were avoided to prevent excessive photobleaching. Student's T-test versus PD(WT); p-value $<0.05= *$; p-value $<0.01= **$; p-value $<0.001= ***$

The comparisons of PD WT versus the mutants have been summarised in the table below.

	mCit-PD (WT) vs mCit-Mutant			PD mCit-(PD) vs Mutant-mCit		
PD mutant	Protein Localisation	Slow Diffusion Coefficient	Fraction of Slow Diffusing Molecules	Protein Localisation	Slow Diffusion Coefficient	Fraction of Slow Diffusing Molecules
N124K	Different	Significant Difference	Similar	Different	Similar	Significant Difference
R92Q	Similar	Similar	Significant Difference	Similar	Similar	Significant Difference
C52R	Different	Significant Difference	Similar	Different	Significant Difference	Significant Difference
S54R	Different	Similar	Significant Difference	Different	Significant Difference	Significant Difference
S121L	Different	Significant Difference	Similar	Different	Significant Difference	Similar
	WT=S54R>R92Q>C52R>N124K>S121L			WT=R92Q>S121L>S54R>C52R>N124K		

Table 3-4: A summary of differences between the PAX6 PD WT versus the mutants with N-or C-terminally tagged mCitrine. At the bottom of the table, the constructs have been ordered based on the mean normalised autocorrelation curve decay.

Furthermore, in silico analysis of the investigated PAX6 variants predicted with high confidence that all variants except R92Q, had a negative effect on protein function (Table 3-5). Regarding R92Q, PolyPhen-2 software predicted with high confidence the variant to be damaging, whereas the SIFT software predicted that the variant could be tolerated.

PAX6 variant	C52R	S54R	R92Q	S121L	N124K
PolyPhen-2 score	Probably_damaging (1)	Probably_damaging (1)	Probably_damaging (0.99)	Possibly_damaging (0.9)	Probably_damaging (0.99)
SIFT score	Affect_protein_function (0)	Affect_protein_function (0)	Tolerated (0.26)	Affect_protein_function (0)	Affect_protein_function (0)

Table 3-5: In silico analysis of the pathogenic PAX6 variants using PolyPhen-2 and SIFT softwares. The SIFT score ranges from 0.0 (deleterious) to 1 (tolerated). Variants with scores between 0.0 and 0.05 are considered deleterious with a high confidence. Variants from 0.05 to 1 are predicted to be tolerated. The closer the score to 1 the higher the confidence that it is tolerated (benign). PolyPhen-2 score ranges from 0.0 to 1 and predicts the possible impact of an amino acid substitution on the structure and function of a human protein. If a variant has a value between 0.0 and 0.15 it is predicted to be benign. If it has a value between 0.15 and 1.0 it is predicted to be possibly damaging. The higher the value the higher the confidence.

The overall conclusions from these two experiments, were that there was a clear difference in subnuclear localisation between the WT and the pathogenic variants and that the pathogenic variants excluding the S54R variant displayed a faster decay, suggesting weaker and shorter-lived interactions with chromatin.

mCitrine-PAX6 (FL) WT vs mutants

The PAX6 PD protein in its self, does not exist in humans (it is always a portion of a protein). I hence aimed to corroborate these experiments using the full-length PAX6 protein with a mCitrine N-terminal fusion.

The protein localisation of all PAX6 variants was found exclusively in the nucleus. Interestingly, the full-length WT PAX6 and R92Q TFs showed similar distribution as their PD counterparts. The remaining pathogenic variants displayed a relatively similar subnuclear localisation pattern, all distinct from the WT or R92Q variant. The consistency observed in subnuclear localisation at the PD and full-length PAX6 protein level when compared with the pathogenic variants suggested a potential hallmark of their abnormal function. Normalised autocorrelation curve comparisons between the WT and mutant variants displayed similar patterns as mCit-PD (Figure 3-9), whereby the WT TF had the longest decay, followed by R92Q and then the pathogenic variants. Once again, the autocorrelation curve was similar between all constructs suggesting similar free diffusion amongst the proteins. Unlike the PD experiments, no significant differences amongst the constructs were observed in the fraction of the slow diffusion coefficient. Therefore the observed differences in the mean autocorrelation curves was attributed to the slow diffusion coefficient. Significant differences were only observed between the WT and the pathogenic variants but not with R92Q variant.

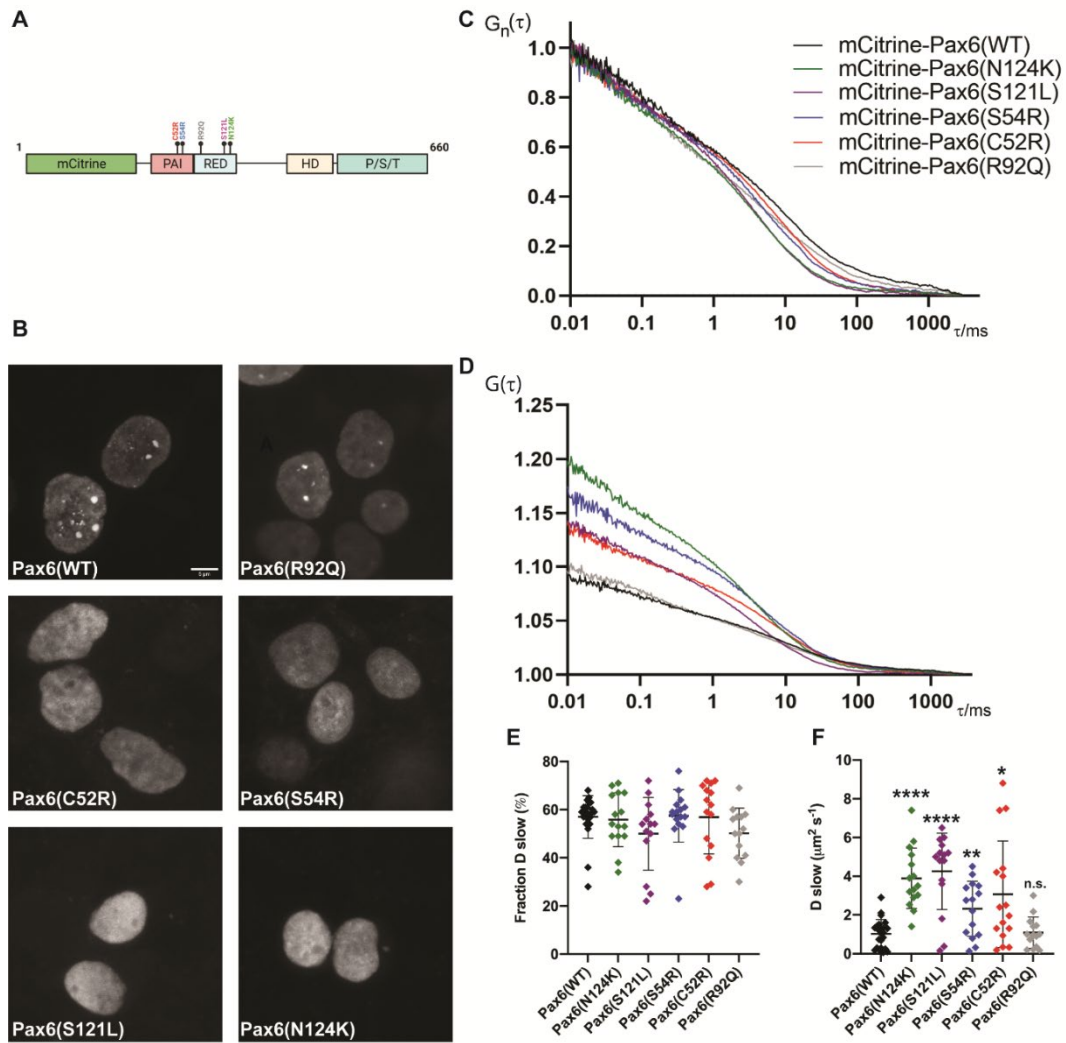


Figure 3-10: Comparison of mCit-PAX6 between WT and missense variants. The autocorrelation function data were fitted with an established 3D diffusion mathematical model for obtaining the fractions and diffusion coefficients of the free moving (fast) and chromatin-bound (slow) molecules **A**. Schematic illustration of fused PD with mCitrine and variant annotation **B**. Confocal images of the subnuclear localisation of the PAX6 PD fused protein in HEK293 cells 24 post-transfection **C**. Mean autocorrelation curve whereby $G(\tau)$ is normalised to 1 at the zeroth (0th) lag time **D**. Mean autocorrelation curves derived from FCS measurements **E**. Scatter plot comparing the fraction of the slow diffusion coefficient **F**. Scatter plot comparing the slow diffusion coefficient ; The images are representative of cells that were used for FCS measurements. mCit-PAX6 (WT) $N_{\text{cells}}=16$, mCit-PAX6 (N124K) $N_{\text{cells}}=15$, mCit-PAX6 (S121L) $N_{\text{cells}}=14$, mCit-PAX6 (S54R) $N_{\text{cells}}=16$, mCit-PAX6 (R92Q) $N_{\text{cells}}=13$, mCit-PAX6 (C52R) $N_{\text{cells}}=15$. One measurement per cell was carried out to minimize photodamage and foci of elevated protein concentration were avoided to prevent excessive photobleaching. Student's T-test versus PAX6 (WT); p-value $<0.05= *$; p-value $<0.01= **$; p-value $<0.001= ***$

SOX2-mCherry (FL) WT vs mutants

Unlike PAX6, SOX2 has been successfully labelled endogenously with a fluorescent protein in mESCs (Bressan *et al.*, 2017). For this reason, the SOX2 TF was also fused to mCherry at the C-terminus as previous studies indicated that this largely retained protein functionality. In contrast to PAX6, SOX2 has only one DBD and the majority of pathogenic missense mutations are found there (HMG domain). For this study, four variants present in human patients were chosen. Apart from phenotypic and genetic analysis, no further molecular experiments have previously been conducted with these variants. Therefore, their subnuclear localisation and chromatin binding dynamics with the WT TF was compared. When the normalised mean autocorrelation curves of the variants were compared against the WT SOX2 (Figure 3-11C), a difference was observed for all except for the G130A variant. Statistical analysis on the slow diffusion coefficient and the fraction of slow diffusing molecules, indicated that all variants except G130A had a significantly faster slow diffusion coefficient than the WT. There were no observed differences in the fraction of the slow diffusing molecules, suggesting that the observed differences in the normalised autocorrelation curves could be attributed to the impaired chromatin binding capabilities. The subnuclear localisation of all variants and WT TF was exclusively in the nucleus, as expected. No major differences in protein localisation were observed between the WT protein and the variants. The SOX2 variants (excluding G130A) were evenly distributed throughout the nucleus with no regions of accumulation (i.e. no preference for a specific locus). On the contrary, the WT TF displayed regions of preferential accumulation in the nucleus. Since only the WT and G130A (intact HMG domain) demonstrated this phenomenon, it is possible that this was directly linked with the ability to interact with chromatin. In order to further test this hypothesis, SOX2 cDNAs were co-expressed and DNA was counterstained with a nuclear-permeant dye (NucBlue; ThermoFisher; R37605) (Figure 3-12). Overall, the variants present in the HMG domain did not show a preference towards chromatin localisation (Figure 3-12) nor did they interact with chromatin to the same extent as the WT SOX2 or G130A. These results

agreed with the observed phenotypes in human patients, in which the G130A variant exerts the least severe effect. Furthermore, *in silico* analysis of the SOX2 variants indicated that the G130A variant is likely to be tolerated (Table 3-6).

SOX2 variant	N46K	R74P	L97P	G130A
PolyPhen-2 score	Probably_damaging (1)	Possibly_damaging (0.91)	Probably_damaging (0.99)	Benign (0.42)
SIFT score	Affect_protein_function (0)	Affect_protein_function (0)	Affect_protein_function (0)	Tolerated (0.28)

Table 3-6: In silico analysis of the pathogenic SOX2 variants using PolyPhen-2 and SIFT softwares. The SIFT score ranges from 0.0 (deleterious) to 1 (tolerated). Variants with scores between 0.0 and 0.05 are considered deleterious with a high confidence. Variants from 0.05 to 1 are predicted to be tolerated. The closer the score to 1 the higher the confidence that it is tolerated (benign). PolyPhen-2 score ranges from 0.0 to 1 and predicts the possible impact of an amino acid substitution on the structure and function of a human protein. If a variant has a value between 0.0 and 0.15 it is predicted to be benign. If it has a value between 0.15 and 1.0 it is predicted to be possibly damaging. The higher the value the higher the confidence.

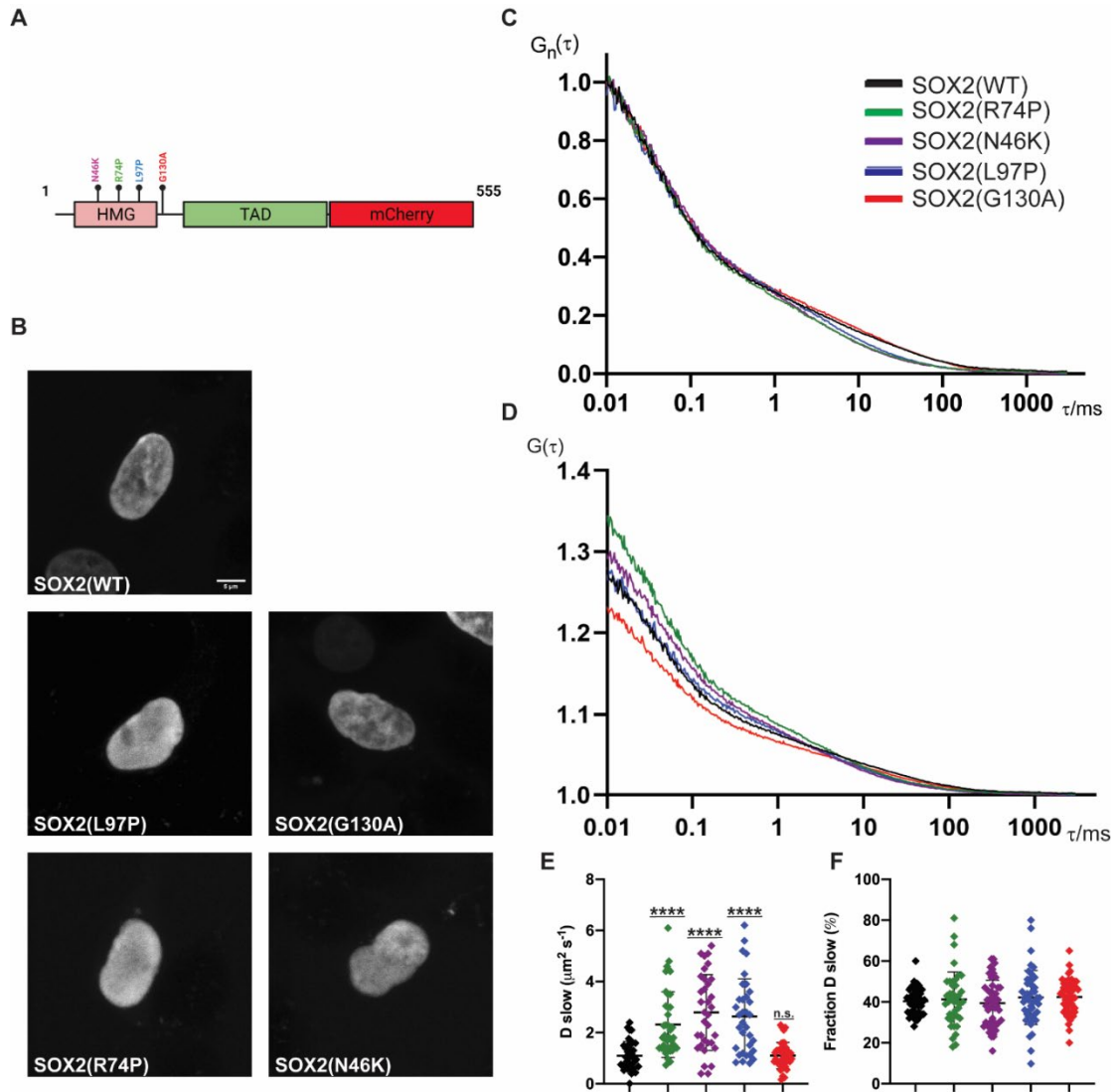


Figure 3-11: Comparison of SOX2-mCherry between WT and missense variants. The autocorrelation function data were fitted with an established 3D diffusion mathematical model for obtaining the fractions and diffusion coefficients of the free moving (fast) and chromatin-bound (slow) molecules. **A.** Schematic illustration of SOX2 with mCherry and variant annotation. **B.** Confocal images of the subnuclear localisation of the SOX2 fused protein in HEK293 cells 24 post-transfection. **C.** Autocorrelation curve whereby $G(\tau)$ is normalised to 1 at the zeroth (0th) lag time. **D.** Mean autocorrelation curves derived from FCS measurements. **E.** Scatter plot comparing the fraction of the slow diffusion coefficient. **F.** Scatter plot comparing the slow diffusion coefficient. The images are representative of cells that were used for FCS measurements. FCS measurements are the sum of three biological replicates; SOX2-mCherry (WT) $N_{\text{cells}}=55$, SOX2-mCherry (R74P) $N_{\text{cells}}=41$, SOX2-mCherry (N46K) $N_{\text{cells}}=52$, SOX2-mCherry (G130A) $N_{\text{cells}}=56$, SOX2-mCherry (L97P) $N_{\text{cells}}=47$. One measurement per cell was carried out to minimize photodamage and foci of elevated protein concentration were avoided to prevent excessive photobleaching. ; Student's T-test versus PD(WT); p-value $<0.05= *$; p-value $<0.01= **$; p-value $<0.001= ***$

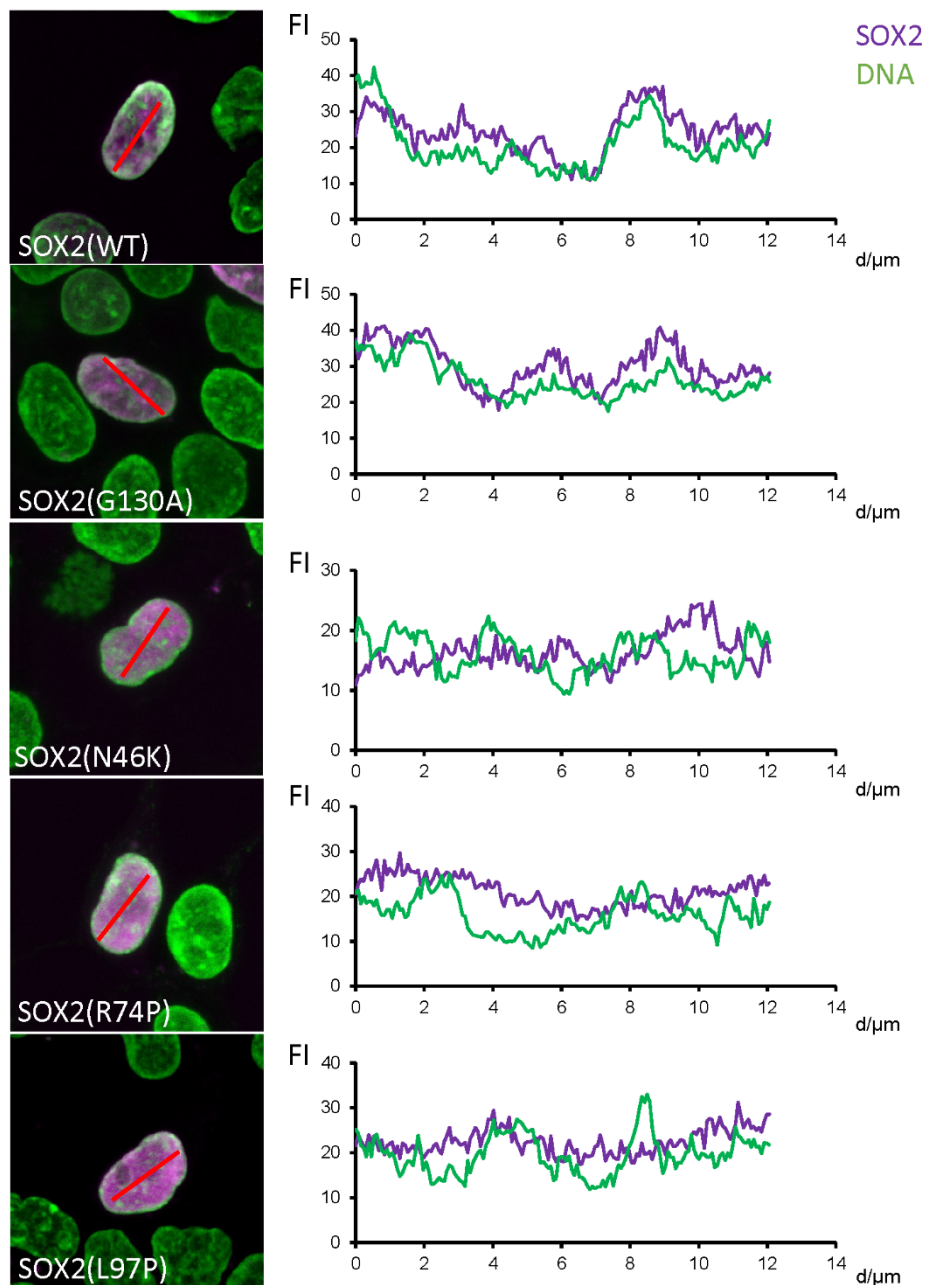


Figure 3-12: Comparison of SOX2-mCherry protein localisation with DNA. The green colour represents DNA and purple represents SOX2. The graphs depict the fluorescence intensity across the red line depicted in the images. Fluorescence intensity analysis was carried out using imageJ.

3.3 Discussion

This chapter aimed to investigate the biophysical properties of pathogenic missense variants of *SOX2* and *PAX6* and their role in phenotype severity in human patients. Although FCS experiments have previously been used to study *SOX2*, it is the first time that *PAX6* has been investigated. Furthermore, it is the first time, in which the biophysical properties of pathogenic variants for both *PAX6* and *SOX2* have been compared to the WT TFs by using FCS.

Qualitative methods assessing TF binding such as chromatin immunoprecipitation and ChIP-seq, provide static information derived from fixed cell population (Valouev *et al.*, 2008). Despite being informative, these analysis fail to provide insight into the dynamic behaviour of TF molecules in real-time. Live-cell methodologies such as FCS, offer the ability to measure protein chromatin-binding dynamics and concentration, without extensively damaging the cell.

FCS has previously been used to study TF concentration, protein variability and chromatin-binding dynamics in live cells (Vukojević *et al.*, 2005), *Drosophila* (Vukojević *et al.*, 2010) and zebrafish embryos (Shi *et al.*, 2009), making it a versatile methodology with an established pipeline for the biophysical analysis of TF proteins. A key disadvantage of FCS which is shared by the large majority of advanced microscopy methodologies, is that the target protein is required to be fluorescently labelled. This more often than not, may have a detrimental effect in protein function which must be considered and investigated. Although multiple key developmental TFs have been fluorescently endogenously labelled, such as *SOX2* (Bressan *et al.*, 2017), *OCT4* (Strebingner *et al.*, 2019) and *NANOG* (Heo *et al.*, 2015), *PAX6* remains elusive. It is widely believed that *PAX6* is both a difficult protein to purify but also to investigate biochemically. In fact, the majority of biochemical experiments such as EMSAs have only used the PD of *PAX6* (Kamachi *et al.*, 2001; Williamson *et al.*, 2020). Therefore, knowing that fusing a protein with a fluorescent protein may be damaging, I investigated which labelling approach would be less damaging.

To address potential disruptive effects on protein function, the PD of PAX6 was initially labelled at the N-terminus and C-terminus respectively and the molecular behaviour and subnuclear localisation was compared, in an effort to use the least damaging tagging approach. Overall, the comparisons between N- and C-terminally fused mCitrine to PD showed that the position of the fluorophore influenced the overall molecular behaviour, but was not considered to be major. The subnuclear localisation of some variants though was notably influenced, namely C52R and N124K in which mCitrine was fused at the C-terminus of the PD (summary of differences can be seen in Table 3-3). All PAX6 PD variants were exclusively localised in the nucleus when fused N-terminally with mCitrine and the free diffusion of all variants including the WT, was similar. Free TF diffusion is often in the region of 0.2-0.8 ms whereby you expect similar diffusion for similar sized proteins (Vukojević *et al.*, 2010). This was only observed when the PD was labelled at the N-terminus (Figure3-9C), but not at the C-terminus (Figure3-8C). Taking this into consideration, future PAX6 experiments were done with an N-terminus mCitrine tag.

Following the assessment of the tagging strategy, the chromatin-binding dynamics and subnuclear localisation of five *PAX6* and four *SOX2* pathogenic variants was investigated in HEK293 cells and compared against the WT proteins in the context of the full-length proteins (Figure 3-10, Figure 3-11). All investigated pathogenic variants except *SOX2* (G130A), are found inside a DBD, the PD domain for *PAX6* and HMG domain for *SOX2*, which are integral for normal protein function. Additionally, each of these missense variants in human patients demonstrate various degrees of pathogenicity implying differences in the importance of each amino acid residue for normal *PAX6* (Table 3-1) and *SOX2* function (Table 3-2). Heterozygous loss-of-function *PAX6* variants are associated with typical aniridia (Cunha *et al.*, 2019) whereas *SOX2*, with anophthalmia (Fantes *et al.*, 2003).

The results obtained in all three experiments (Figure 3-8, Figure 3-9, Figure 3-10) consistently supported the idea that WT *PAX6* exhibited stronger and longer lasting interactions with chromatin than the pathogenic variants. In addition, the WT and R92Q variants consistently bore foci of high TF

concentration, whereas the mutants did not. In conclusion, impairment of chromatin interaction and altered subnuclear localisation were identified as key aberrations of PAX6 pathogenic variants. It is very likely that these two molecular properties may be the underlying reason of their pathogenic nature in humans. Something which was unexpected, was the clear subnuclear localisation of the WT PAX6 when compared to the pathogenic variants. In all cases (mCit-PD, PD-mCit, mCit-PAX6), the WT PAX6 protein, formed foci of increased concentration which were not present in the pathogenic variants. In addition, the R92Q variant had similar expression pattern as the WT. Considering the clear difference between the WT and the pathogenic variants, this may possibly be linked to their pathogenicity. TF phase separation has recently been associated with gene regulation (Boija *et al.*, 2018) and the observed foci may likely be phase-separated, a phenomenon which will be investigated in the next chapter.

Although EMSAs are typically used for the molecular investigation of PAX6 missense variants, the results in this chapter suggest that FCS can be an alternative methodology for the study and comparison of overall protein binding dynamics. No peptides are required to be purified for FCS which can often be a time-consuming step of EMSAs. On the contrary, FCS is a methodology which measures average diffusion over time and therefore it cannot definitively identify whether a peptide interacts with a specific DNA sequence. In conclusion, FCS may not necessarily be used as a replacement to EMSAs for studying mutant protein-DNA binding, but rather be used in combination.

For SOX2, no fluorophore tagging assessment was necessary, as SOX2 has previously been endogenously fluorescently labelled with mCherry at the C-terminus, in mESCs. Experiments were able to show that SOX2 largely retained its functionality in these cells (Bressan *et al.*, 2017), although to definitively state that it is not damaging, work at an organismal level would be required. The mobility and subnuclear localisation of the SOX2 variants demonstrated altered behaviour dynamics when compared to the WT TF, with

their molecular behaviour been analogous to the severity of their phenotype in patients (Figure 3-11).

The experiments in this chapter utilised, for the first time, FCS for the investigation and comparison of pathogenic missense variants of both PAX6 and SOX2 in live cells. Even though HEK293 cells are shown in Human Cell Atlas (www.humancellatlas.org) to express both PAX6 and SOX2, a more relevant cell line should be used in the future such as neural progenitor cells (NPCs). Additionally, another experimental limitation was that all investigated proteins were transiently transfected and thus were not at endogenous levels. Although large differences in protein concentration can influence FCS measurements and TF-chromatin binding dynamics, this was controlled as much as possible whereby all measurements were carried out in cells of similar expression levels.

The results in this chapter demonstrated that in all cases, the molecular diffusion of each variant was analogous to the severity of its phenotype in patients. The impaired chromatin interaction and altered subnuclear localisation were identified as key aberrations of *PAX6* and *SOX2* pathogenic variants. It is very likely that these two molecular properties may be the underlying reason of their pathogenic nature in humans.

Chapter 4 Study of the regulation of PAX6 function and nuclear concentration by biomolecular condensates

4.1 Introduction

Phase separation of substances is a physical phenomenon that refers to different phases that form in a homogeneous mixture. In cells, formation of biomolecular condensates has been implicated in a plethora of cellular functions. These range from the constitution of membrane-less organelles such as the nucleolus to the normal (BRD4, PSPC1) and abnormal function (FUS, TDP43) of proteins implicated in disease (Han *et al.*, 2020; Niaki *et al.*, 2020; Riback *et al.*, 2020). However, in-depth understanding of the biological meaning of condensate formation if any, particularly by TFs, is lacking. Several TFs have the ability to form condensates, aiding in nuclear compartmentalisation and potential gene regulation, although how widespread this is still remains unclear. Since condensate formation has the ability to buffer concentration and functionally compartmentalise the nucleus, then it is possible for condensates to control local TF concentrations. One example determined by this phenomenon is super-enhancers which are clusters of regulatory elements that accumulate components of the transcriptional machinery. TF condensates favour the emergence of local reaction-diffusion compartments which assist/favour transcriptional events without excluding additional functions (Boija *et al.*, 2018; Cho *et al.*, 2018). Such molecular crowding may also bring regulatory sequences and promoters into close proximity and such reversible networks can favour gene expression. This process can depend on nucleation events triggered by physiological (e.g. NF- κ B in inflammation; Nair, developmental (e.g. Prospero in neural differentiation; Liu *et al.*, 2020) or molecular processes, such as depletion of proteins (e.g. Mediator complex and BRD4; Boija *et al.*, 2018). Originally, TF condensates have been considered to be the functional forms of TFs on chromatin. A recent study has proposed that TF condensates may also be involved in buffering the concentration of TFs in the dilute phase (Klosin *et al.*,

2020). This though, remains to be experimentally verified at endogenous levels.

Protein phase separation depends on physico-chemical conditions including protein concentration, charge, 3D structure and environmental pH. Condensates may have liquid, gel or solid-like properties, although the biological consequences of such entities and whether they participate in cellular function or not, remain elusive (Strom *et al.*, 2017; Boeynaems *et al.*, 2018; Alberti and Dormann, 2019). The thermodynamics of liquid-liquid phase-separated (LLPS) systems predicts that the concentration inside the condensate is higher than the surrounding phase, whereby a further increase in the total concentration of the PS-substance leads to an increase in condensate size/number and the maintenance of the same dilute phase concentration. LLPS systems exhibit rapid protein movement into, and within, the dense phase and can exhibit fusion and fission behaviour during their formation and maturation (Phair and Misteli, 2000; Hoege *et al.*, 2009). Even though the precise biophysical mechanism of condensate formation remains elusive, weak multivalent interactions between intrinsically disordered regions (IDRs) of proteins are believed to be one of the primary drivers. Coactivator proteins MED1 and BRD4 have been shown to form condensates at super-enhancers. In addition, several TFs such as SOX2, OCT4 and NANOG have been shown to form condensates with the Mediator complex *in vitro* (Boija *et al.*, 2018).

In this chapter the biophysics and function of the WT PAX6 condensates observed in the previous chapter were investigated. Additionally, the contribution of the PD to condensate formation was investigated.

4.2 Results

4.2.1 PAX6 forms condensates in a PD dependent manner

Many studies have linked the efficiency of protein phase separation to the presence of intrinsically disordered regions, as well as the enrichment of specific amino acids such as arginine and lysine. Foci were observed with both the PD and the full-length *PAX6* TFs but not with the pathogenic variants. Therefore I initially analyzed the amino acid sequence of the PD and compared the PONDR scores between the WT and mutants. The missense mutation R92Q, which was also able to form similar foci to the WT, yielded the highest 'ordered' prediction score amongst all mutant TFs (Figure 4-1). For the remaining missense mutations, none bore substantially altered PONDR scores, which would explain their inability to form foci. If foci formation was solely dependent on the degree of 'disorderiness' of the IDRs, then one would expect the pathogenic missense variants to have an overall more 'ordered' structure. Following this rationale, the C52R variant which is predicted to bear a more 'disordered' region would have been expected to be able to form foci. Considering the inability of the pathogenic mutant variants to form foci and the minor changes in PONDR scores, when compared to the WT, it is likely that the IDRs are not the sole contributor to the observed foci formed by WT *PAX6*. Additionally, analysis using the CIDER tool (Figure 4-2), did not identify any major differences in the values of the parameters known to influence phase separation, such as hydrophobicity, net charge and charged amino acid sequence distribution within proteins (Campen *et al.*, 2008).

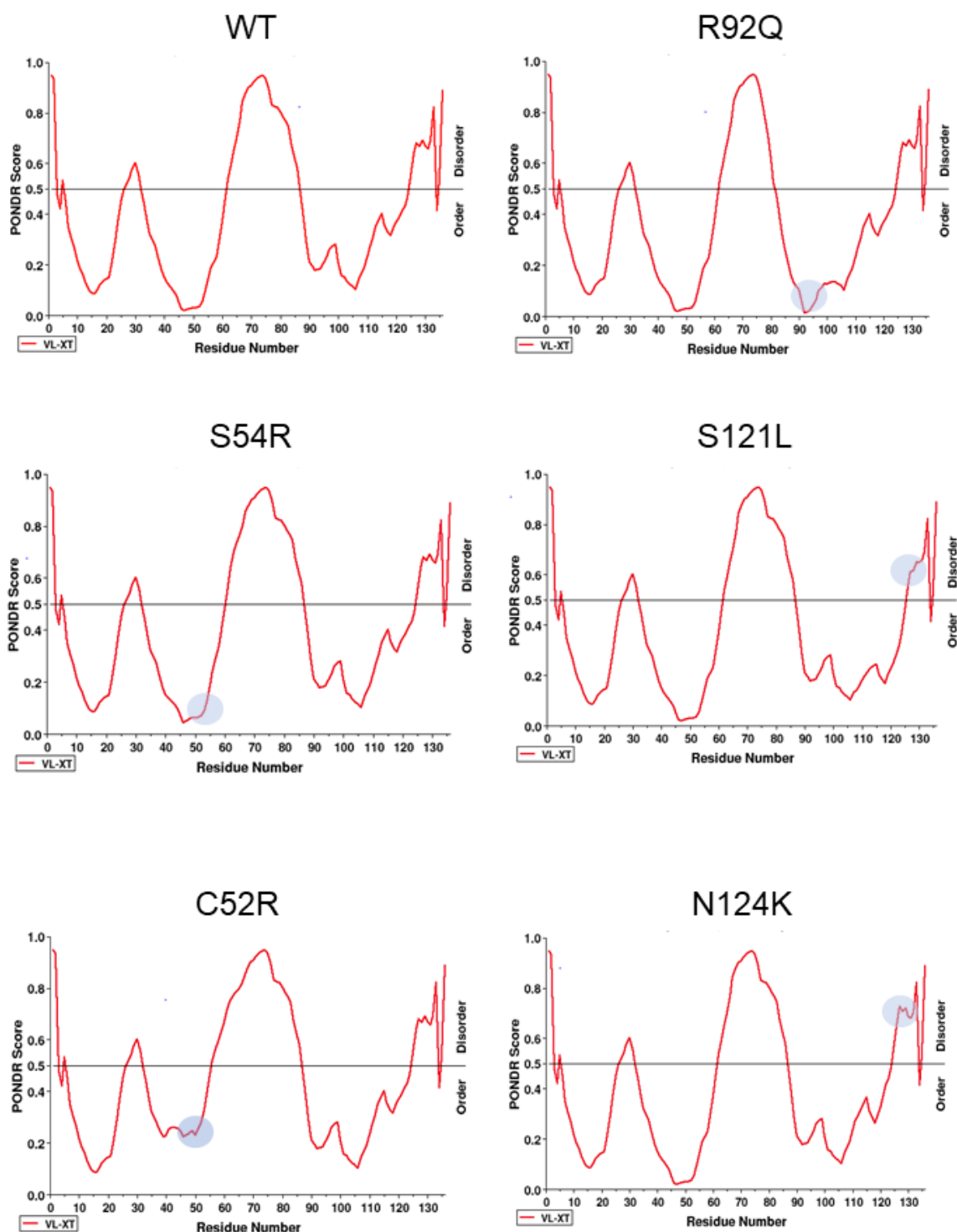
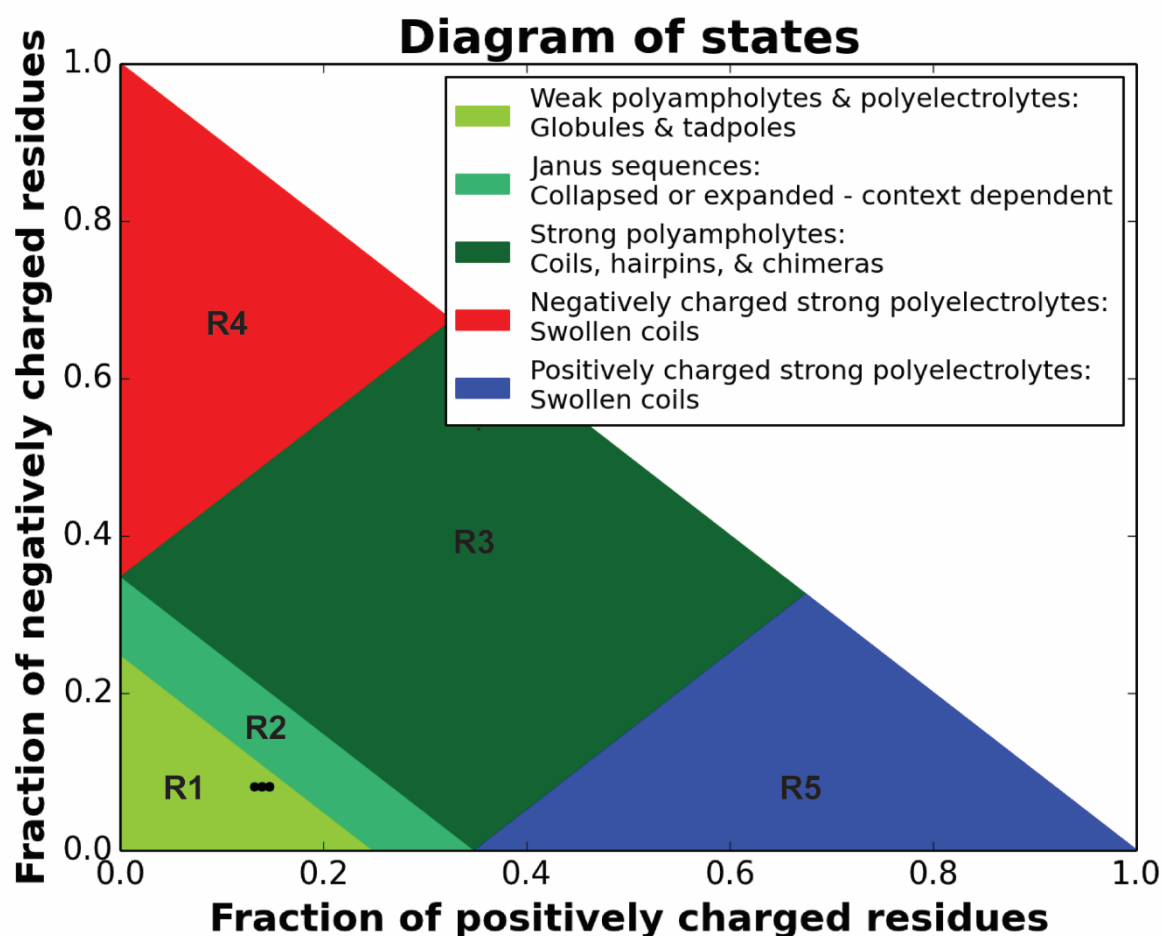


Figure 4-1: Graphical representation of PONDR (predictor of natural disordered regions) score analysis of the WT and mutant PAX6 PD. The blue circles found in each of the pathogenic variants represent the position of the missense mutation. Regions with a PONDR score higher than 0.5 are considered to be disordered regions and as a result have a higher probability in contributing to protein phase separation. Analysis was conducted using www.pondr.com

A.



B.

PD	Length (aa)	κ	FCR	NPCR	Hydropathy	Disorder Promoting	Plot Region
WT	136	0.14	0.22	0.059	4.113	0.625	1
C52R	136	0.14	0.23	0.066	4.061	0.632	1
S54R	136	0.15	0.23	0.066	4.085	0.625	1
R92Q	136	0.14	0.21	0.051	4.12	0.625	1
S121L	136	0.14	0.22	0.059	4.146	0.618	1
N124K	136	0.15	0.23	0.066	4.11	0.632	1

Figure 4-2: Analysis of the WT and mutant PAX6 PD sequences using CIDER. **A.** R1-R5 illustrate five groups of proteins based on the ratio of positively and negatively charged residues. All 6 peptides analysed are found in group R1 with minimal differences. **B.** Comparison of WT and mutant PAX6 PD parameters influencing protein phase-separation. These are the κ (charged amino acid mixing), FCR (fraction of charged residues), NPCR (net charge per residue), extent of overall hydrophobicity and value of overall disorder. Overall, there are no major electrostatic differences between the WT and pathogenic variants which can explain why the WT protein is able to form foci whereas the mutants are not.

PAX6 encodes three main isoforms, each bearing differences in the PD (Figure 4-3). PAX6 (5a) contains an additional 14 amino acids in the PAI subdomain which alters the binding specificity and affinity of the PD. The PAX6 (δ PD) isoform lacks a PD completely. Since foci also form when the PD is expressed alone, I wanted to investigate whether foci formation required the presence of a WT PD. To investigate this, I expressed the three isoforms in HEK293 cells and observed their subnuclear localisation. The canonical and PAX6 (5a) isoforms formed foci whereas the PAX6 (δ PD) did not, suggesting that the mechanism of foci formation is PD-dependent.

I used FCS to investigate the diffusion properties of the three isoforms. The normalised mean autocorrelation curves (Figure 4-3C) showed that the canonical PAX6 isoform displayed the longest decay time of the slow component, suggesting that it formed the strongest and longest-lived complexes with chromatin when compared against the other two isoforms. Similarly, the PAX6 (5a) mean autocorrelation curve displayed longer decay times than that of PAX6 (δ PD). These differences can also be observed in both the relative fraction of slowly diffusing molecules (Figure 4-3E) and their diffusion coefficient (Figure 4-3F). The canonical PAX6 isoform had a significantly lower diffusion coefficient of the slow FCS component when compared to the other two isoforms. In terms of the relative fraction of bound and free molecules, the canonical PAX6 isoform displayed a significantly higher fraction than the PAX6 (δ PD) but not the PAX6 (5a). There was no significant difference when comparing the diffusion coefficients of the slow component of PAX6 (5a) against the one of PAX6 (δ PD). Previous studies have demonstrated that the 5a insertion impedes binding of the PAI subdomain to chromatin (Mishra, Ivan P Gorlov, *et al.*, 2002). My results demonstrated that the 5a insertion alters the overall behaviour of the protein such that its overall diffusion is similar to the one of PAX6 (δ PD). Next to the characterised differences in PAX6 binding site recognition between the canonical PAX6 and the PAX6 (5a) TFs, FCS was able to demonstrate macroscopic differences in PAX6-chromatin interactions, which are expected

to mediate its ability to find its target sites in the nucleus. Overall, the PONDR scores suggested that the IDRs are not the main driver of PAX6 foci and the isoform comparison suggests that the foci are likely PD dependent.

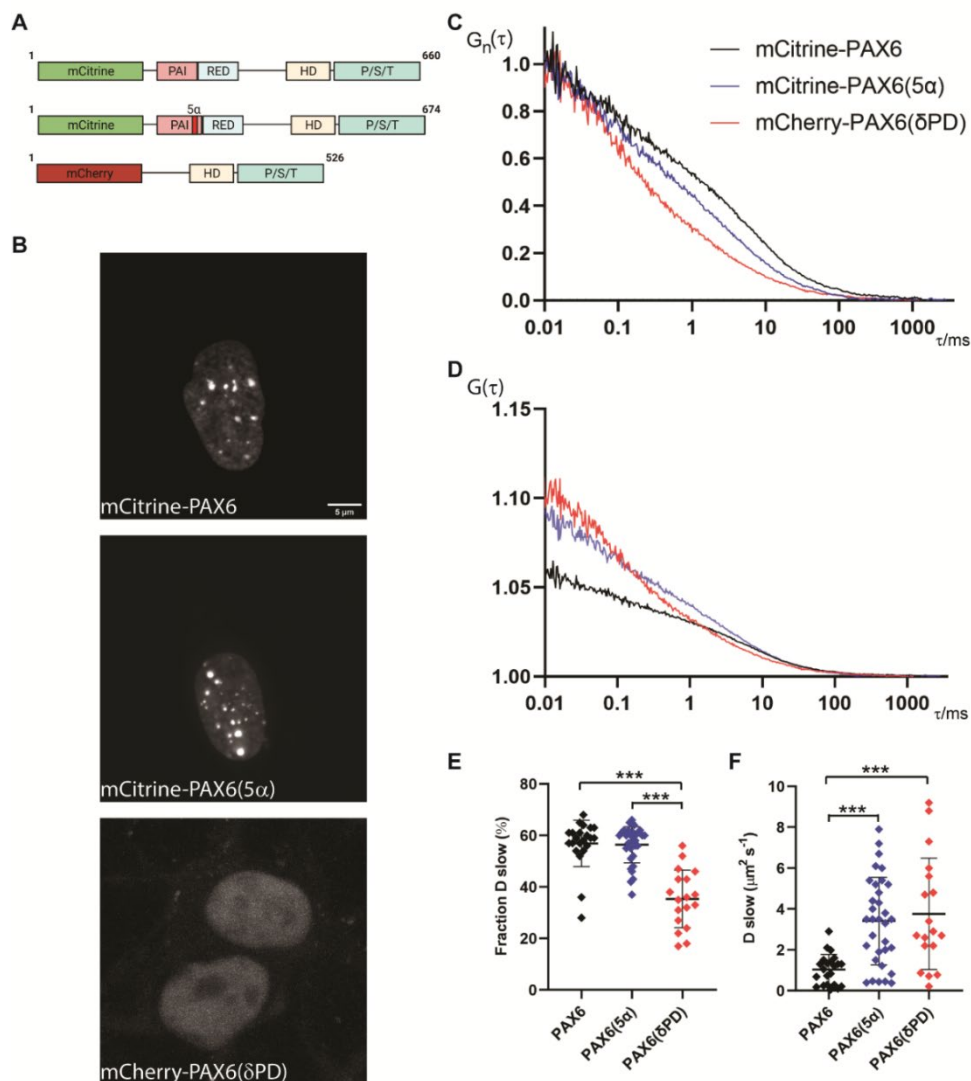


Figure 4-3: Comparison of PAX6 isoforms. The autocorrelation function data were fitted with an established 3D diffusion mathematical model for obtaining the fractions and diffusion coefficients of the free moving (fast) and chromatin-bound (slow) molecules **A**. Schematic representation of fused PAX6 isoforms with either mCititrine or mCherry **B**. Confocal images of the subnuclear localisation of the tagged proteins in HEK293 cells 24 hours post-transfection **C**. Normalised autocorrelation curves at the zeroth (0th) lag time ($G_n(\tau)$) **D**. Mean autocorrelation curves derived from FCS measurements **E**. Scatter plot comparing the fraction of the slow diffusing molecules **F**. Scatter plot comparing the diffusion coefficient of the slow FCS component; The images are representative of cells that were used for FCS measurements. mCit-PAX6 (WT) $N_{cells}=22$, mCit-PAX6 (5α) $N_{cells}=33$, mCherry-PAX6 (δPD) $N_{cells}=18$. One measurement per cell was carried out to minimize photodamage and foci of elevated protein concentration were avoided to prevent excessive photobleaching. Student's t-test p-value $<0.05= *$; $<0.01= **$; $<0.001= ***$

4.2.2 PAX6 foci are dynamic and can be dissolved by 1-6 hexanediol

Although the majority of assays to characterise phase-separation of proteins are conducted *in vitro* using recombinant proteins, some methods can also be used in live cells. The underlying physical and chemical interactions that drive phase separation in IDRs is not fully understood but electrostatic (Pak *et al.*, 2016), cation- π , π - π interactions (Kim *et al.*, 2016) and the overall hydrophobic effects (Yeo, Keeley and Weiss, 2011) have all been proposed to contribute to IDR phase separation. Arginine and glutamine are two abundant amino acids of many phase-separating proteins (Wang *et al.*, 2018). Addition of surplus arginine is believed to distort the microenvironment of these phase-separated systems leading to their dissolution. I therefore, expressed mCitrine-PAX6 in HEK293 cells and incubated them with 1M arginine for 10 minutes (Figure 4-4). No major differences in PAX6 foci number or size were observed.

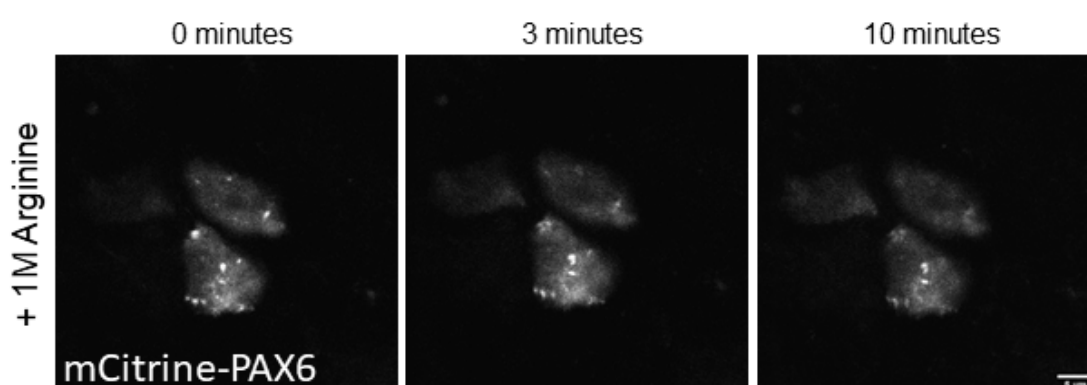


Figure 4-4: HEK293 cells expressing mCitrine-PAX6 24 hours post-transfection. Treatment with 1M arginine was carried out by replacing media. Images are of a single z plane representing the brightest foci. Scale bar= 5 μ m

Although IDRs play a major role in protein phase-separation other protein modules cannot be excluded. Considering that the PAX6 foci were also observed when the PD alone was expressed, PAX6 IDRs outside of the PD are dispensable for foci formation.

I therefore repeated the experiment, but this time using mitoxantrone. Mitoxantrone is a type II topoisomerase inhibitor leading to disruption of DNA

synthesis and repair by intercalation (Kreft *et al.*, 2018). Previous studies have used mitoxantrone as a dissolving agent for stress granules (Wheeler *et al.*, 2019; Basu *et al.*, 2020), but in my experiments, no evident dissolution of PAX6 condensates was observed.

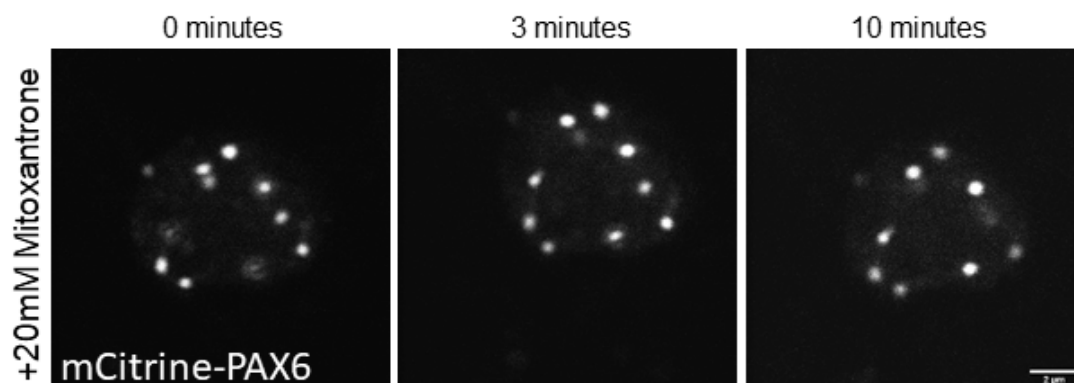


Figure 4-5: HEK293 cells expressing mCitrine-PAX6 24 hours post-transfection. Treatment with 20mM mitoxantrone was carried out by replacing the medium. Images are single plane representations of the brightest foci.

Lastly, I repeated the experiment using 1-6-hexanediol. This aliphatic alcohol interferes with weak hydrophobic interactions of IDRs and is often used for dissolving protein condensates *in vitro* and in cells (Chong *et al.*, 2018; Liu *et al.*, 2020). When 10% 1-6-hexanediol was used in HEK293 cells expressing mCitrine-PAX6, dissolution of PAX6 condensates was observed within 10 minutes of incubation (Figure 4-6). Compared to previous studies of other proteins, this rate of dissolution is considered relatively slow, suggesting the

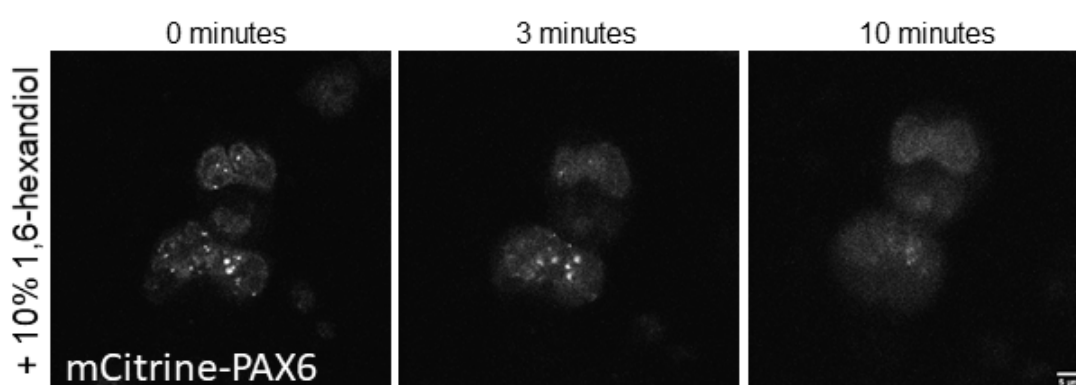


Figure 4-6: HEK293 cells expressing mCitrine-PAX6 24 hours post-transfection. Treatment with 10% 1,6-hexanediol was carried out by replacing media. Images are of a single z plane representing the brightest foci.

PAX6 condensates have more of a gel-like consistency rather than a liquid one as previously indicated with the FUS protein (Wang *et al.*, 2018).

To further explore the dynamic nature of the PAX6 condensates, a stable HEK293 cell line expressing GFP-PAX6 was created using the TET-ON system. Tetracycline inducible GFP-PAX6 HEK293 cells were incubated with 1 μ M tetracycline O/N and the following day were imaged for 6 hours in 20 minute intervals. The ability of condensates to form (Figure 4-7B) and dissolve (Figure 4-7A) without any external input, further suggested that the observed foci were dynamic PAX6 condensates. The dynamic nature of these condensates in the absence of any external input suggests that they respond to changes in cellular biochemistry.

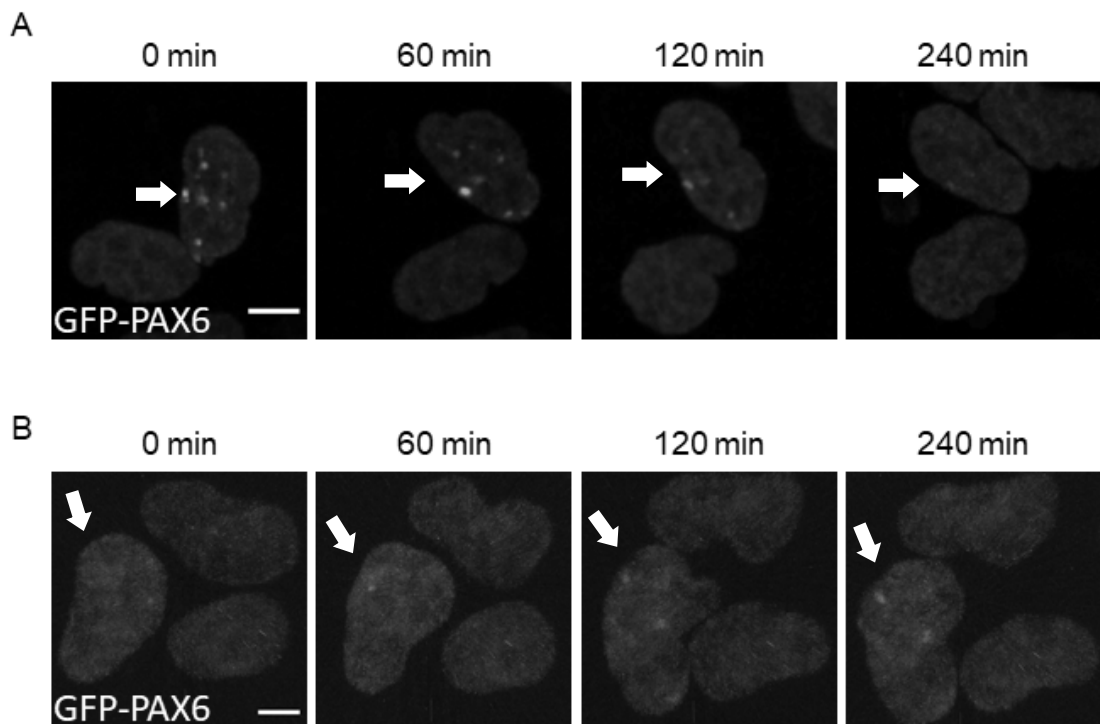


Figure 4-7: Stable HEK293 cell line expressing GFP-PAX6 24 hours post tetracycline treatment. Images (z-stacks) were acquired every 60 minutes. PAX6 condensates in different cells dissolved (A) and formed (B) without any external input over time

Previous studies have shown that TF condensates often colocalise with components of the transcriptional machinery including; RNA PolII, the co-activator BRD4 and the Mediator complex (Boija *et al.*, 2018). I wished to determine whether PAX6 condensates colocalise with any transcriptional machinery complexes. HEK293 cells expressing mCitrine-PAX6 were fixed and stained with an RNA Polymerase II Serine 5 phosphorylation (S5P CTD PolII) antibody. S5P CTD PolII is necessary for transcription initiation and is therefore often found at active promoters (Nojima *et al.*, 2018). Upon comparison of PAX6 condensates with PolII Ser5, no significant colocalisation was observed (Figure 4-8). Some PAX6 foci did not colocalise with PolII Ser5, but with chromatin dense regions which are usually transcriptionally silent. Previous studies have shown TF foci colocalisation with chromatin dense regions, namely SOX2, although the biological significance remains largely unknown (Veneri *et al.*, 2020). Additionally, other Pax family members (PAX3 and PAX9) have been shown to form similar foci at major satellite repeats by associating with DNA within pericentric heterochromatin (Bulut-Karslioglu *et al.*, 2012). Nevertheless, this was attributed to a consensus homeodomain binding site present in satellite repeats. As previously shown, PAX6 foci are PD dependant (Figure 4-3), although this does not exclude the possibility of a similar mechanism (i.e PD binding site).

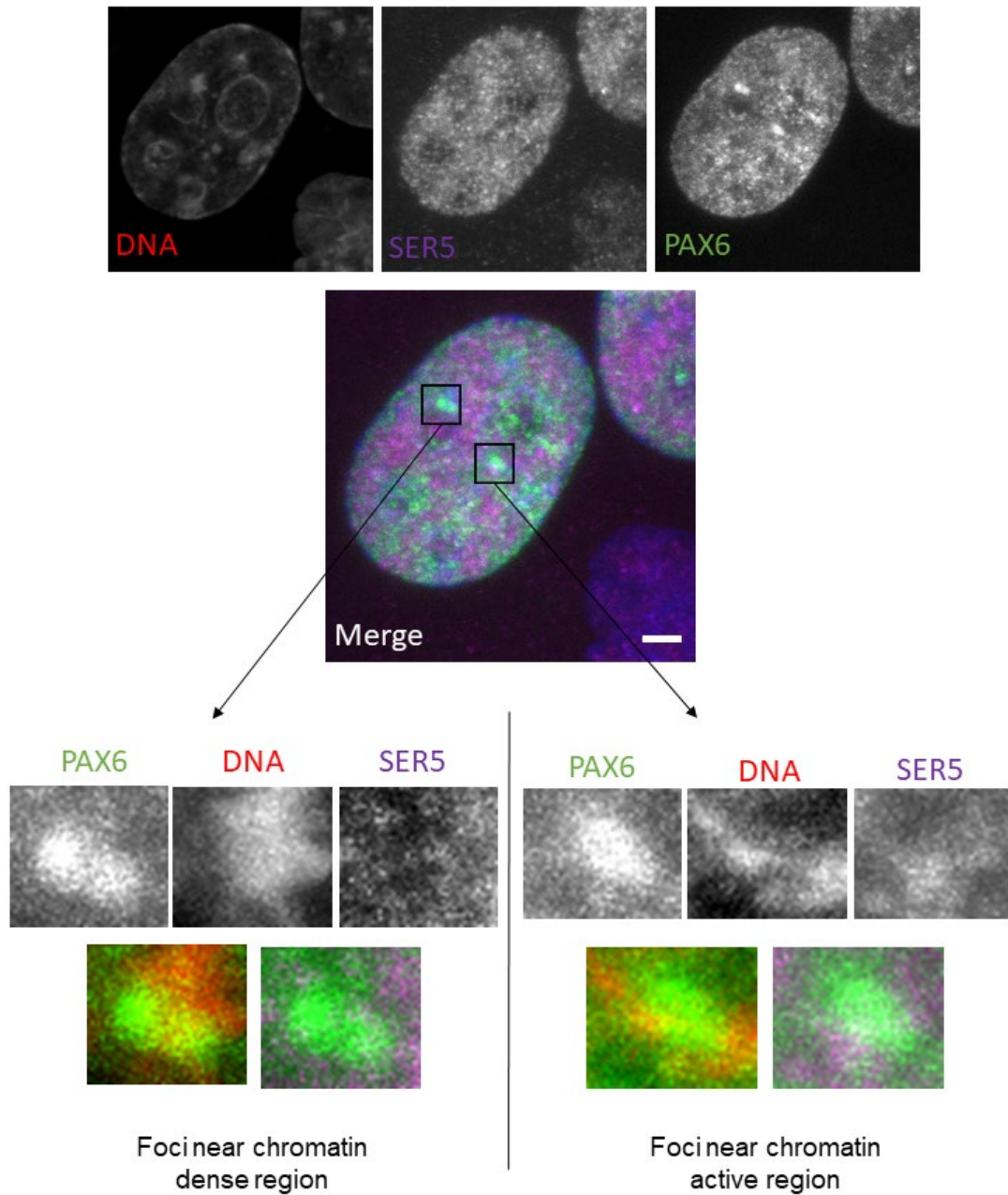


Figure 4-8: HEK293 cells expressing mCitrine-PAX6 fixed and stained for RNA PolII (Ser5). No exclusive colocalisation between PAX6 condensates and PolII Ser5, was observed. There was no exclusive colocalisation between PAX6 and DNA dense regions.

Heterochromatin protein 1 (HP1) plays a role in gene silencing by heterochromatin. It is proposed that HP1 proteins spread across large regions of the genome compacting the underlying chromatin and recruiting diverse ligands (Kang *et al.*, 2011; Larson *et al.*, 2017). Human HP1a has been identified to undergo LLPS in which components such as nucleosomes and DNA preferentially partition into these droplets (Larson *et al.*, 2017). Colocalization was observed between expressed mCitrine-PAX6 and endogenously labelled HP1a protein in HEK293 cells (Figure 4-9).

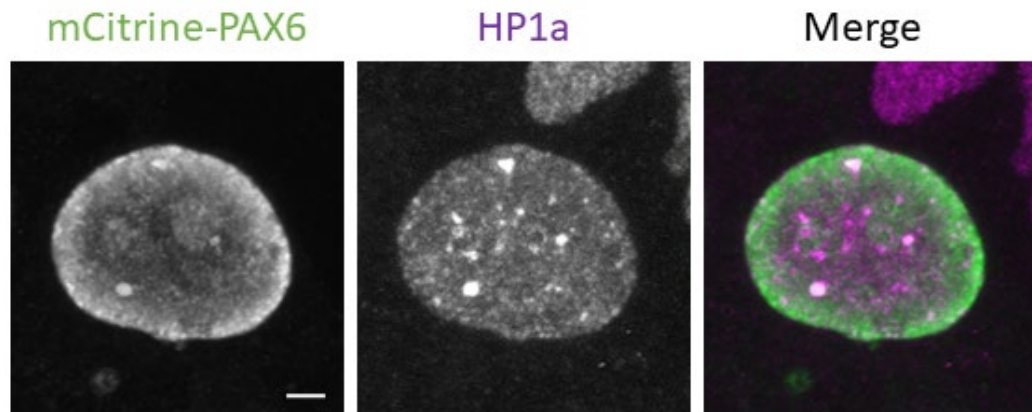


Figure 4-9: HEK293 cells expressing mCitrine-PAX6 fixed and stained for HP1 α

A previous study had suggested a possible role of *PAX6* in cellular proliferation and more specifically, its localisation in pericentromeric areas during mitosis which may influence mitotic progression (Zaccarini *et al.*, 2007). HP1a is targeted to mitotic centromeres by INCENP, which is a subunit of the chromosome passenger complex (CPC) (Kang *et al.*, 2011). Therefore, considering the above independent studies, the colocalization of HP1a and PAX6 which is observed in my experiment, is likely to be pericentric heterochromatin as previously shown for other Pax family members (Bulut-Karslioglu *et al.*, 2012) found on centromeres. Although the biological function of HP1a in centromeres has been investigated (Yi *et al.*, 2018), further work is needed for PAX6. PAX6 foci association with pericentric heterochromatin has been suggested to have a detrimental effect in cell cycle (Zaccarini *et al.*, 2007). More recent work though, has linked Pax family members with a beneficial role in major satellite transcript repression, heterochromatic mark maintenance and chromosome segregation (Bulut-Karslioglu *et al.*, 2012)

Even though the condensates I have observed may be due to overexpression, the differences between the WT and the pathogenic mutants are consistent. Additionally, the fact that endogenous PAX6 condensates were not observed, does not exclude the possibility for PAX6 to form them. I therefore wanted to further investigate this by carrying out *in vitro* experiments using recombinant PD and full length PAX6 protein.

4.2.3 PAX6 forms condensates *in vitro*

Recombinant full length PAX6 protein was covalently labelled with Alexa 488 and a mounting setup (Figure 4-10A) adapted from Boija *et al.*, 2018 was constructed. No condensates were originally observed, but upon the addition of 10% PEG, condensates with a spherical shape that were also visible in brightfield were formed (Figure 4-10B). PEG is a hydrophilic nonionic polymer, commonly used as a crowding agent for liquid-liquid partitioning and precipitation of biomacromolecules (Annunziata *et al.*, 2002). FCS measurements before and after the formation of condensates demonstrate a pronounced increase in decay time, indicative of larger and slower moving molecular complexes (condensates) (Figure 4-10D).

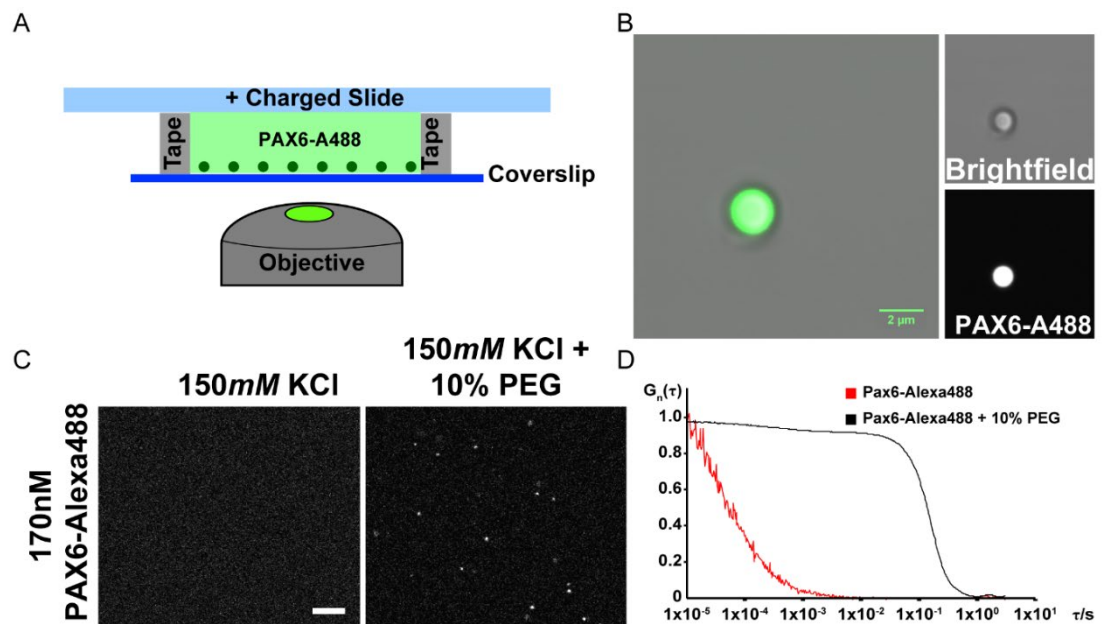


Figure 4-10: In vitro condensate experiments of recombinant PAX6 covalently labelled with Alexa 488 **A**. Schematic of the experimental setup **B**. Confocal imaging (40x) of phase separated PAX6-Alexa 488 upon addition of 10% PEG (solution of 150mM KCl) **C**. Confocal imaging (40x) of PAX6-Alexa 488 before and after addition of 10% PEG (scale bar = 20μm) **D**. Normalised autocorrelation curves at the zeroth (0th) lag time of PAX6 before and after addition of 10%PEG

A similar setup was used for investigating the recombinant peptide of PD alone, which was kindly provided by Nikki Hall. In cells expressing *PAX6*, condensates were dissolved with 10% 1-6-hexanediol (Figure 4-6). I therefore repeated this experiment *in vitro* and performed FCS measurements in three distinct conditions, as depicted in Figure 4-11A. Initially, the concentration was calculated from FCS measurements of PD-Alexa 488. The initial concentration was calculated to be 500nM and the estimated concentration measured by FCS measurements was 395nM. The differences may be attributed to incomplete labelling efficiency since the FCS measurements reflect Alexa 488 concentration. Nevertheless, no condensates were observed when no crowding agent was present. Condensates were formed upon the addition of 10% PEG (Figure 4-11B). FCS measurements were then recorded away from any visible condensates (in bulk phase). The calculated concentration from these FCS measurements was 85nM. Although the total volume of the system had increased (due to addition of PEG) and therefore a reduction in concentration was expected, the total volume had only increased by 10% whereas the bulk phase concentration had decreased by 80%. This suggests that the reduction in bulk phase concentration is likely attributed to condensate formation. Lastly, addition of 10% 1-6-hexanediol partially dissolved the condensates (Figure 4-11B), similarly to what was observed in cells. FCS measurements were again recorded, and the average calculated concentration was 110 nM. If 1-6-hexanediol had no effect on *PAX6* condensates, the total *PAX6* concentration should have further decreased (due to volume increase). The opposite was observed (Figure 4-11D), indicating that 1-6-hexanediol partially dissolves the condensates. If complete dissolution of condensates had occurred, the final total concentration would have been close to the concentration recorded prior to PEG addition.

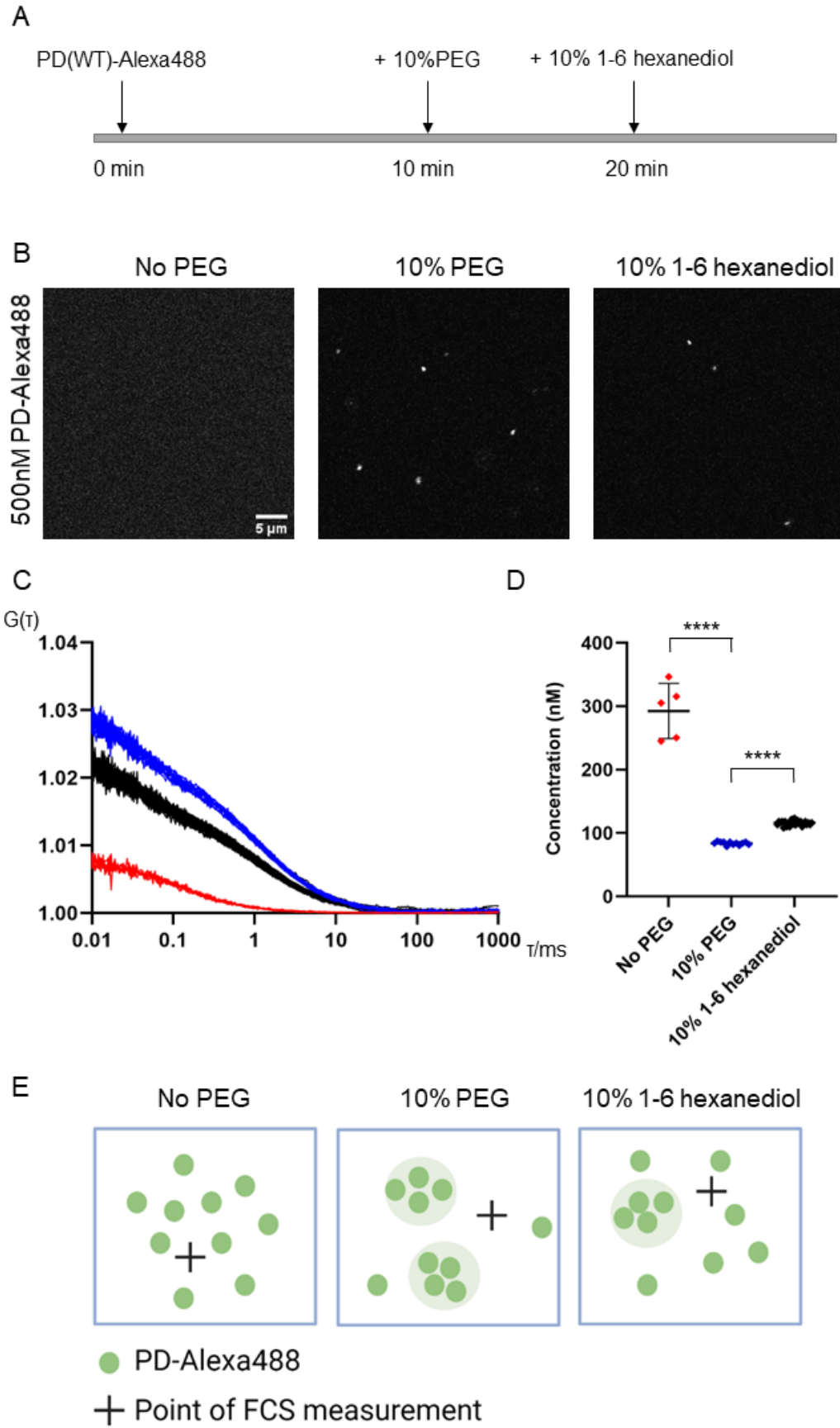


Figure 4-11: In vitro experiments of PD-Alexa 488 **A.** Schematic of the experimental procedure with FCS measurements recorded in each condition **B.** Confocal imaging (40x) of PD-Alexa 488 with no PEG, 10% PEG and 10% 1-6 hexanediol **C.** Autocorrelation curves derived from the measurements in each condition . All measurements recorded for each condition are displayed. **D.** Calculated concentration derived from FCS measurements **E.** Schematic representation of predicted behaviour of PD-Alexa 488 upon addition of 10% PEG and 10% 1-6 hexanediol; p-value <0.05= *; <0.01= **; <0.001= ***

Lastly, FRAP experiments are often used to investigate the physical properties of condensates. FRAP experiments on PD condensates, showed nearly full recovery after photo-bleaching, although the rate of recovery ($t_{1/2} = 49\text{s}$) when compared to other proteins undergoing LLPS was very slow (BRD4S $t_{1/2} = 4\text{s}$; Han *et al.*, 2020;) (HOXD13 IDR $t_{1/2} = 12\text{s}$; Basu *et al.*, 2020). This suggests that the condensates observed are likely liquid-droplets due to near full FRAP recovery, with reduced dynamic properties, suggesting possible maturation into more of a gel-like condensate as in the case of Tau ($t_{1/2} = 56\text{s}$; Kanaan *et al.*, 2020).

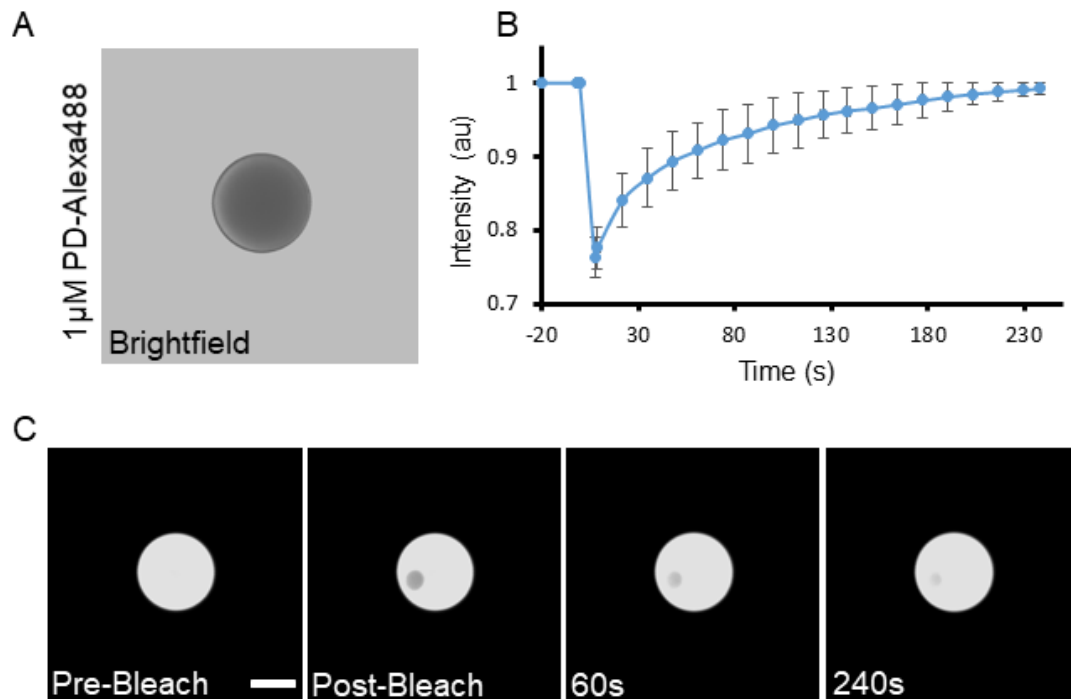


Figure 4-12: In vitro experiment of PD-Alexa 488. **A.** Brightfield image of PD-Alexa 488 condensate following addition of 10% PEG **B.** Graph illustrating the FRAP of a ROI inside the PD-Alexa 488 condensate and confocal images are illustrated in **C**; scale bar = 1 μm

4.3 Discussion

This chapter aimed to investigate the nature of the WT PAX6 foci observed in Chapter 3 and to explore the potential for PAX6 to phase-separate *in vitro*. Results from this chapter indicate for the first time that both recombinant full length PAX6 and the PD alone can form condensates *in vitro* upon the addition of a crowding agent (PEG) and can be dissolved by addition of 10% 1-6 hexanediol, an alcohol often used for the dissolution of protein condensates (Chong *et al.*, 2018; Liu *et al.*, 2020). Furthermore, results suggest that the WT PAX6 foci observed in HEK293 cells, are condensates which colocalize with HP1a and that these condensates are largely dependent on the presence of an intact PD.

HP1 plays a key role in gene silencing by heterochromatin. It is proposed that HP1 proteins spread across large regions of the genome, compacting the underlying chromatin and recruiting diverse ligands (Kang *et al.*, 2011; Larson *et al.*, 2017). Human HP1a has been identified to undergo LLPS in which components such as nucleosomes and DNA preferentially partition into these droplets (Larson *et al.*, 2017). A previous study had suggested a possible role of PAX6 in cellular proliferation and more specifically, its localisation in pericentromeric areas during mitosis, potentially influencing mitotic progression (Zaccarini *et al.*, 2007). Therefore, considering the above independent studies, the observed colocalization of HP1a and PAX6, was likely to be pericentric heterochromatin which has been shown to occur for other Pax family members (Bulut-Karslioglu *et al.*, 2012). Even though the biological function of HP1a in centromeres has been investigated (Yi *et al.*, 2018), further work is needed for PAX6. PAX6 association with pericentric heterochromatin has been suggested to have a detrimental effect in cell cycle (Zaccarini *et al.*, 2007), suggesting it as a putative mechanism for the pathogenic phenotype observed in mice when PAX6 is overexpressed (Schedl *et al.*, 1996). On the contrary, association of Pax family members with pericentric heterochromatin has been linked with a beneficial role in major satellite transcript repression, heterochromatic mark maintenance and chromosome segregation (Bulut-Karslioglu *et al.*, 2012). Just recently, a study

has linked PAX6 foci with mitotic chromosome retention (Lan *et al.*, 2021). Furthermore, pathogenic missense variants present in the PD of PAX6 had a diminished ability for mitotic chromosome retention, which was similar to what was observed in my experiments. To take it one step further, the Yang group, were also able to observe endogenous PAX6 foci in adult retinal pigmented epithelial cells (ARPE-19) throughout all stages of mitosis (Lan *et al.*, 2021), suggesting that the foci observed in HEK293 cells were not an artefact of overexpression or cell type.

Results show that the ability of PAX6 to form these condensates is largely dependent on the presence of an intact PD, potentially explaining the reason why pathogenic missense variants do not appear to form them to the same extent. In accordance to this, results published by the Yang group, indicated that for full mitotic chromosome retention of PAX6, the presence of intact DNA-binding domains was required (Lan *et al.*, 2021).

For in vitro experiments, recombinant full length PAX6 protein was purchased from LD BioPharma (HTF-0109) and a major limitation would have been to ensure that the recombinant protein was functional and possessed similar properties and 3D structure to the endogenous protein. The same peptide was used in previous studies to demonstrate the functionality of the recombinant protein, by performing rescue experiments in heterozygous PAX6 mutant limbal epithelial stem cells (Roux *et al.*, 2018). Investigating the ability of proteins to phase-separate *in vitro* is a rapidly emerging field that has not been clearly standardised yet. Although these experiments can provide invaluable information on the biophysical properties of phase-separated proteins, *in vitro* assays are designed in such a way as to provide the ideal environment for a protein to phase-separate. Phase separation is exquisitely sensitive to changes in physicochemical conditions. Even small differences in temperature, protein, nucleic acid or salt concentration can lead to different outcomes and it is therefore imperative for these parameters to be controlled as much as possible. For the experiments in this chapter, the protein and salt concentration were tightly controlled and the temperature was maintained constant at 37°C. Both recombinant full length PAX6 and the PD were able to

form condensates *in vitro* upon the addition of a crowding agent (PEG). Furthermore, condensate dissolution by 1-6-hexanediol was also observed *in vitro*, in agreement with the results observed in cells expressing mCitrine-PAX6. Although only the WT PAX6 protein was investigated *in vitro*, it would be interesting in the future to repeat these experiments using the PAX6 pathogenic missense mutants.

In summary, results from this chapter suggest that the observed PAX6 foci in HEK293 cells are likely to be condensates which appear to colocalize with HP1a. Additionally, the ability of PAX6 to form these condensates is largely dependent on the presence of an intact PD. The biological significance of PAX6 forming condensates remains unknown, but considering the sensitivity of PAX6 concentration for correct eye development in humans, the ability to form these condensates may be a mechanism to control concentration, the ratios between the isoforms: PAX6 & PAX6 (5a) vs PAX6 (δ PD), or even cell identity maintenance by mitotic chromosome retention (Lan *et al.*, 2021).

Chapter 5 Missense mutations alter the interactions of PAX6 and SOX2

5.1 Introduction

PAX6 and SOX2 TFs are essential for normal eye development and maintenance. Aberrant changes in PAX6 and SOX2 protein level result in ocular defects such as microphthalmia (*PAX6*) and anophthalmia (*SOX2*). More often than not, these changes are attributed to heterozygous insertions, deletions, duplications and single nucleotide substitutions, resulting in frameshift, missense and nonsense molecular consequences. So far, I have shown that the investigated missense mutations in both TFs have been shown to result in altered behaviour dynamics compared to the WT proteins, with their molecular behaviour been analogous to severity of their phenotype in patients (Chapter 3). Previous studies have shown that PAX6 and SOX2 can cross- and auto-regulate each other, but also form complexes to regulate target genes. For example, the ability of both TFs to mutually repress each other leads to refinement of the portion of the ciliary epithelium (high PAX6, low SOX2 levels) at the distal end of the neural retina (high SOX2, low PAX6 levels) (Matsushima *et al.*, 2011). Conversely, during lens placode formation, PAX6 forms a molecular complex with SOX2 on lens-specific enhancers such as the chicken δ -crystallin (DC5) enhancer (Kamachi *et al.*, 2001). PAX6 and SOX2 form a tertiary complex on these enhancers which is stabilised by both protein-protein and protein-DNA interactions. This complex relationship between the two TFs is a highly regulated, dosage sensitive process which is centred on the stoichiometry of PAX6 and SOX2 protein levels for correct ocular development. Therefore, in this chapter, I wanted to investigate how missense mutations alter heterocomplex formation and whether one TF influences the molecular behaviour of the other.

PAX6 consists of two DNA binding domains, a PD and a HD. In addition, *PAX6* encodes a paired-less isoform of PAX6 (*PAX6*(δ PD)) from an internal start codon found downstream of the PD, and *PAX6* (5 α) which contains an

alternative splice variant, whereby a 14 amino acid sequence is incorporated in the PAI subdomain of the PD. Previous studies have shown that the full-length PAX6 isoform can physically interact with the PAX6 (δ PD) leading to transactivation activity enhancement of PAX6 from paired domain binding sites (Mikkola *et al.*, 2001). Additionally, RAX, a homeobox-containing TF that functions in early eye development, has also been shown to interact with the full-length PAX6 isoform (Mikkola *et al.*, 2001). Furthermore, the ratio between the canonical PAX6 protein and PAX6 (5 α) has been shown to be critical in gene regulation and development. For example, the ratio between the two isoforms in embryonic mouse tissues is 8:1, which subsequently becomes 1:1 in adult ocular tissues such as the cornea, iris and lens (Pinson *et al.*, 2005). Altering the ratio of the two isoforms, leads to altered gene expression and regulation. Mouse experiments linked PAX6 (5a) function to promoting the development of the neural retinal structure, whereby deletion of exon 5a disrupted predominantly iris formation (Singh *et al.*, 2002). All pathogenic *PAX6* mutations in humans are heterozygous, sporadic or familial autosomal dominant with notable variability in phenotypic abnormalities. Therefore, all cells which express *PAX6*, express both the WT and mutant *PAX6* (if possible). Considering the direct or indirect influences which PAX6 molecules can have on each other, I wanted to investigate the effect of PAX6 missense mutations on these interactions. So far, all experiments regarding interactions of PAX6-SOX2 and PAX6-PAX6 have been carried out from protein extracts (Co-IP and EMSA)(Kamachi *et al.*, 2001; Grapp *et al.*, 2009). In this chapter, I investigate for the first time these interactions in live-cells and investigate molecular behaviour using FCCS and confocal microscopy.

5.2 Results

5.2.1 Investigating the effect of fluorophores on FCS measurements

FCCS is often used to investigate interactions between proteins such as efficiency and stoichiometry of TF complex formation (e.g. hetero-/homo-dimers) as well as chromatin-binding behaviour of these complexes (Bacia and Schwille, 2003; Papadopoulos *et al.*, 2015). The investigated proteins are each labelled with spectrally distinct fluorophores, such as mCitrine and mCherry. Separate detectors enable simultaneous recording of temporal fluctuations of fluorescent intensity from both fluorophores. Temporal autocorrelation analysis is performed for each signal separately and covariance in fluorescence intensity fluctuations from the two different fluorophores is analysed to assess the co-diffusion of the investigated molecules (used to plot the cross-correlation curve)(Bacia and Schwille, 2003; Haustein and Schwille, 2007).

Before carrying out any FCCS experiments, I first wanted to investigate the two fluorophores (mCitrine and mCherry). I therefore labelled the N-terminus of the PAX6 PD with either mCitrine or mCherry. The size of the two fluorophores is very similar and therefore no diffusional differences would be expected due to protein size. HEK293 cells were co-transfected and expressed both mCitrine-PD and mCherry-PD (Figure 5-1A). No differences in the subnuclear localisation of the two constructs was observed. The diffusion coefficient of the slow diffusing molecules of most mCitrine-PD were significantly smaller than the mCherry-PD (Figure 5-1B). Similarly, the fraction of the slow diffusing molecules was significantly greater in mCitrine-PD (Figure 5-1C). Both of these parameters suggest that mCherry, may lead to weaker and shorter-lived chromatin interactions of the PD. The two mean autocorrelation curves were normalised and compared (Figure 5-1E). The mCitrine-PD autocorrelation had a pronounced larger decay time when compared to mCherry-PD. The shape of the autocorrelation curves though had a very different decay. The autocorrelation of mCherry had a pronounced

decay (up to 0.3 τ /ms) which the mCitrine autocorrelation did not have. This can be either due to an extended triplet state or the detector itself. All experiments conducted in our lab, even for proteins not shown in this thesis that used mCherry, had the same distinct initial decay. It is therefore likely that the underlying reason may in fact be in the detector although not verified.

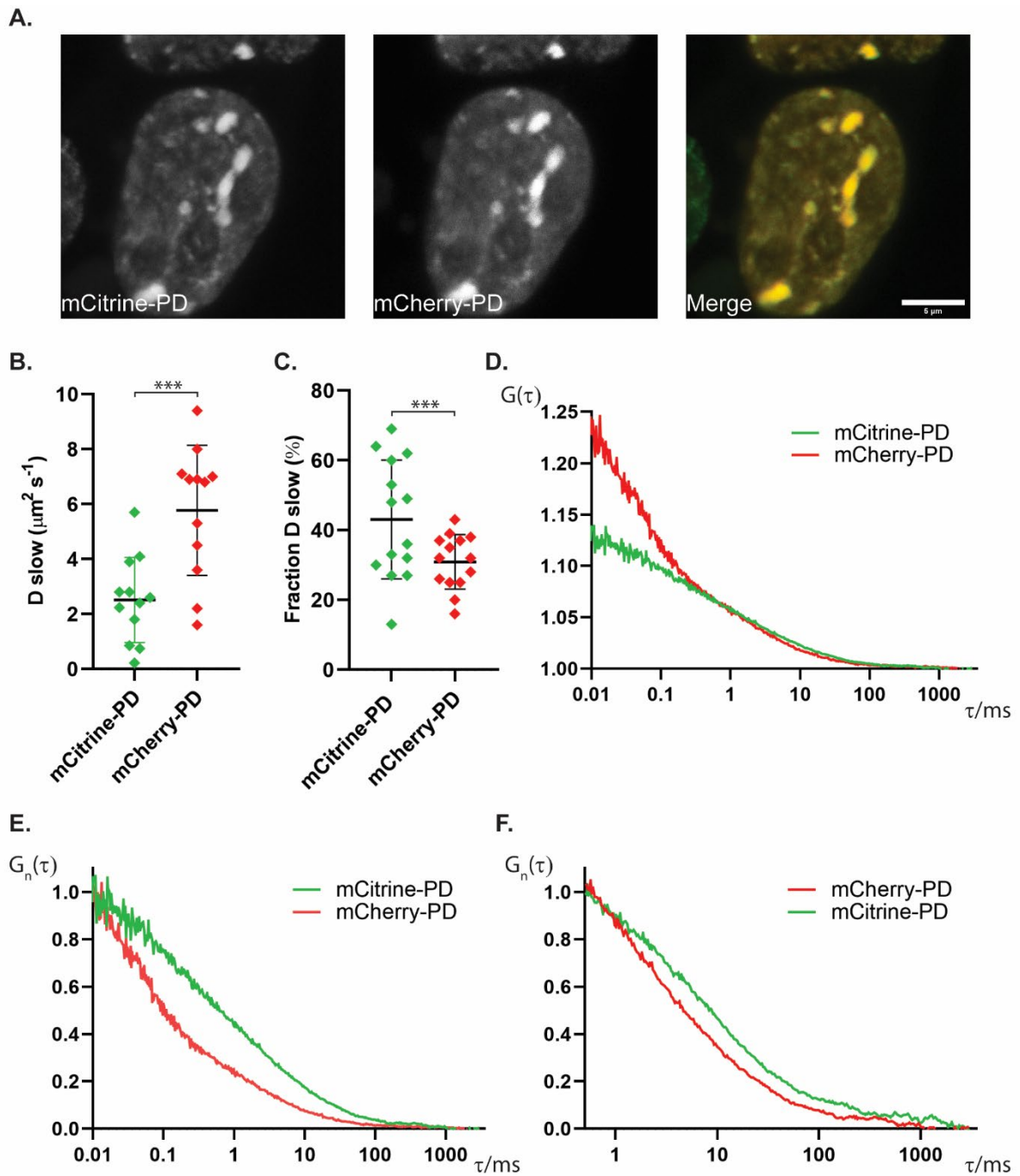


Figure 5-1: Comparison of PD (WT) when labelled at the N terminus with either mCitrine or mCherry. The autocorrelation function data were fitted with an established 3D diffusion mathematical model for obtaining the fractions and diffusion coefficients of the free moving (fast) and chromatin-bound (slow) molecules **A**. Confocal images of the nuclear distribution of co-expressed mCitrine-PD and mCherry-PD in HEK293 cells 24 hours post-transfection **B**. Comparison of the diffusion coefficient of the slow moving PD molecules, as derived by autocorrelation fitting **C**. Comparison of the fraction of the total N molecules undergoing slow diffusion **D**. Mean autocorrelation curve comparison between mCitrine-PD and mCherry-PD **E**. Mean autocorrelation curve comparison where $G(\tau)$ is normalised to 1 at zeroth (0^{th}) lag time **F**. Mean autocorrelation curve comparison where $G(\tau)$ is normalised to 1 at $0.5 \tau/\text{ms}$; The images are representative of cells that were used for FCS measurements; mCitrine-PD $N_{\text{cells}}=14$, mCherry-PD $N_{\text{cells}}=14$; Student's t-test p-value $<0.05= *$; $<0.01= **$; $<0.001= ***$

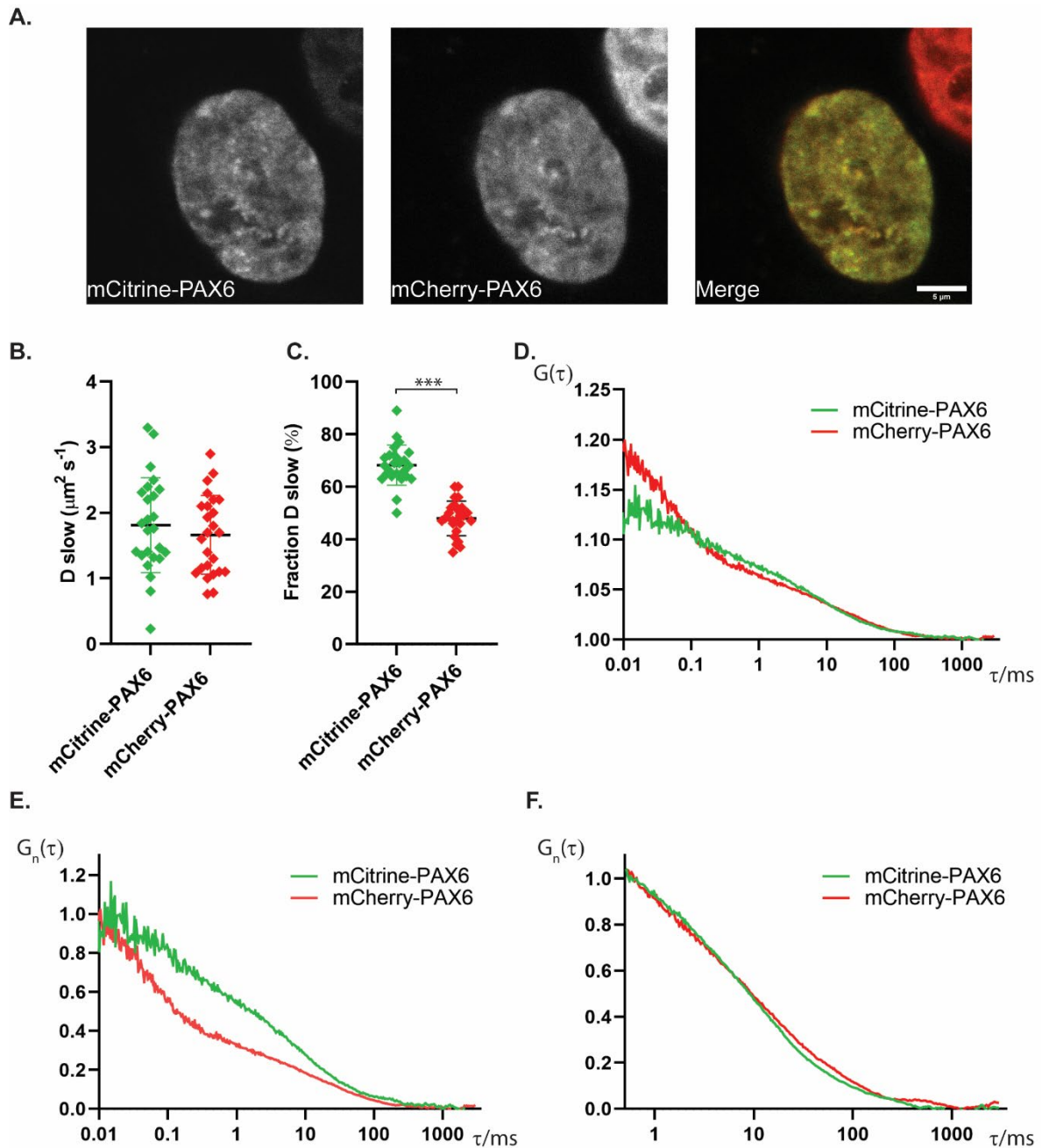


Figure 5-2: Comparison of PAX6 (WT) when labelled at the N terminus with either mCitrine or mCherry. The autocorrelation function data were fitted with an established 3D diffusion mathematical model for obtaining the fractions and diffusion coefficients of the free moving (fast) and chromatin-bound (slow) molecules **A**. Confocal images of the nuclear distribution of co-expressed mCitrine-PAX6 and mCherry-PAX6 in HEK293 cells 24 hours post-transfection **B**. Comparison of the diffusion coefficient of the slow moving PAX6 molecules, as derived by autocorrelation fitting **C**. Comparison of the fraction of the total N molecules undergoing slow diffusion **D**. Mean autocorrelation curve comparison between mCitrine-PAX6 and mCherry-PAX6 **E**. Mean autocorrelation curve comparison where $G(\tau)$ is normalised to 1 at zeroth (0^{th}) lag time **F**. Mean autocorrelation curve comparison where $G(\tau)$ is normalised to 1 at 0.5 τ/ms ; The images are representative of cells that were used for FCS measurements; mCitrine-PAX6 $N_{\text{cells}}=25$, mCherry-PAX6 $N_{\text{cells}}=26$; Student's t-test p-value $<0.05= *$; $<0.01= **$; $<0.001= ***$

I then performed the same experiment but using the full-length PAX6. Similar to the PD (Figure 5-1A) no subnuclear localisation differences between mCitrine-PAX6 and mCherry-PAX6 were observed. Upon normalisation of the two mean autocorrelation curves (Figure 5-2E) a pronounced difference in curve decay was observed, suggesting stronger and longer-lived mCitrine-PAX6 interactions with chromatin. The observed differences in autocorrelation curve decay though, were not evident when comparing the diffusion coefficient of slow moving molecules and their fraction (Figure 5-2B-C). When the mean autocorrelation curves were normalised at 0.5 τ /ms though (Figure 5-2F), no major differences were observed. Normalisation at 0.5 τ /ms excludes the triplet state and the majority portion of fast diffusing molecules. It is therefore likely that the observed differences are not due to the diffusional behaviour of the protein but rather due to the detection and interpretation differences acquired from the two detectors. Based on these results, no further comparisons were made between constructs that were fluorescently labelled with different fluorophores.

5.2.2 PAX6 missense variants influence homo-complex formation efficiency

All human pathogenic PAX6 missense mutations are heterozygous, meaning that all cells which express PAX6 are made up of both the WT and mutant protein. Considering that previous studies have shown that PAX6 has the ability to interact with other PAX6 molecules, I wanted to investigate the effect that the mutant variants may have on this interaction and whether they influence the molecular behaviour of the WT protein.

WT and mutant PAX6, labelled with either mCitrine or mCherry, were expressed in HEK293 cells (Figure 5-3). When both constructs were WT PAX6 (mCherry and mCitrine), condensates were observed in a fraction of cells (~20% of transfected cells). When either of the expressed proteins was a mutant variant (S121L or C52R) no WT PAX6 condensates were observed. It is therefore possible, that the mutant PAX6 variants may have an inhibitory

effect on WT PAX6 condensate formation. Overall, the subnuclear localisation between WT and mutant PAX6 proteins was very similar (Figure 5-3). FCCS measurements were carried out in HEK293 cells co-expressing WT and/or mutant PAX6 at similar concentrations (Figure 5-4B) to avoid excessive signal bleed-through (cross-talk) from one detector to the other (predominantly green to red) which would give false-positive results (overestimation of cross correlation). Cross-correlation curve amplitude is proportional to concentration. Therefore the higher the amplitude the higher the frequency of complex formation. Initially, the cross-correlation curves of PAX6 WT+PAX6 WT; PAX6 WT+PAX6 C52R; PAX6 WT+PAX6 S121L were compared (Figure 5-4A). In all cases, all three cross-correlation curves had a greater amplitude than the PAX6 WT+GFP (negative control) indicating that in all three conditions protein co-diffusion was present, implying complex formation. PAX6 has previously been shown, by Co-IP, to interact with itself via PD-HD trans-interactions (Chauhan *et al.*, 2004a), but the results presented here are the first time this has been observed in live cells. Normalisation of the mean cross-correlation curves (Figure 5-4C) ($G(\tau) = 1$ at the 0th lag time) indicated no differences in decay time of the complexes, suggesting that the particular mutants of PAX6 tested form complexes less efficiently than the WT, but such complexes bind chromatin equally strongly. Comparison of the mean normalised autocorrelation curves of mCherry-PAX6 (WT) (Figure 5-4D) showed no pronounced differences suggesting that interaction of PAX6 (WT) with mutant PAX6 does not influence its overall diffusion.

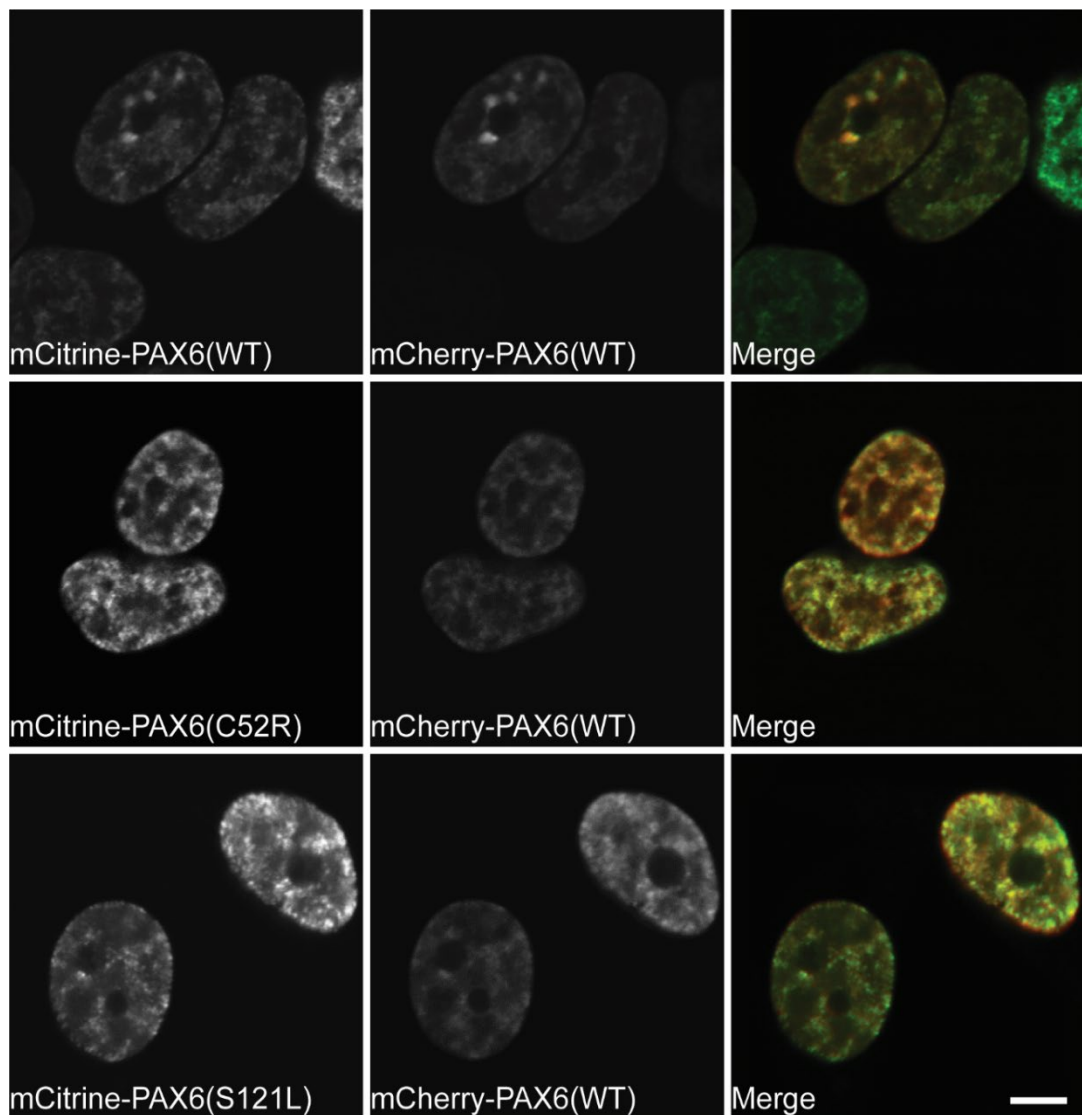


Figure 5-3: Confocal images (40x lens) of HEK293 cells co-expressing WT and/or mutant mCitrine-PAX6 and mCherry-PAX6. The images are representative of cells that were used for FCS measurements; Scale bar = 5 μ m

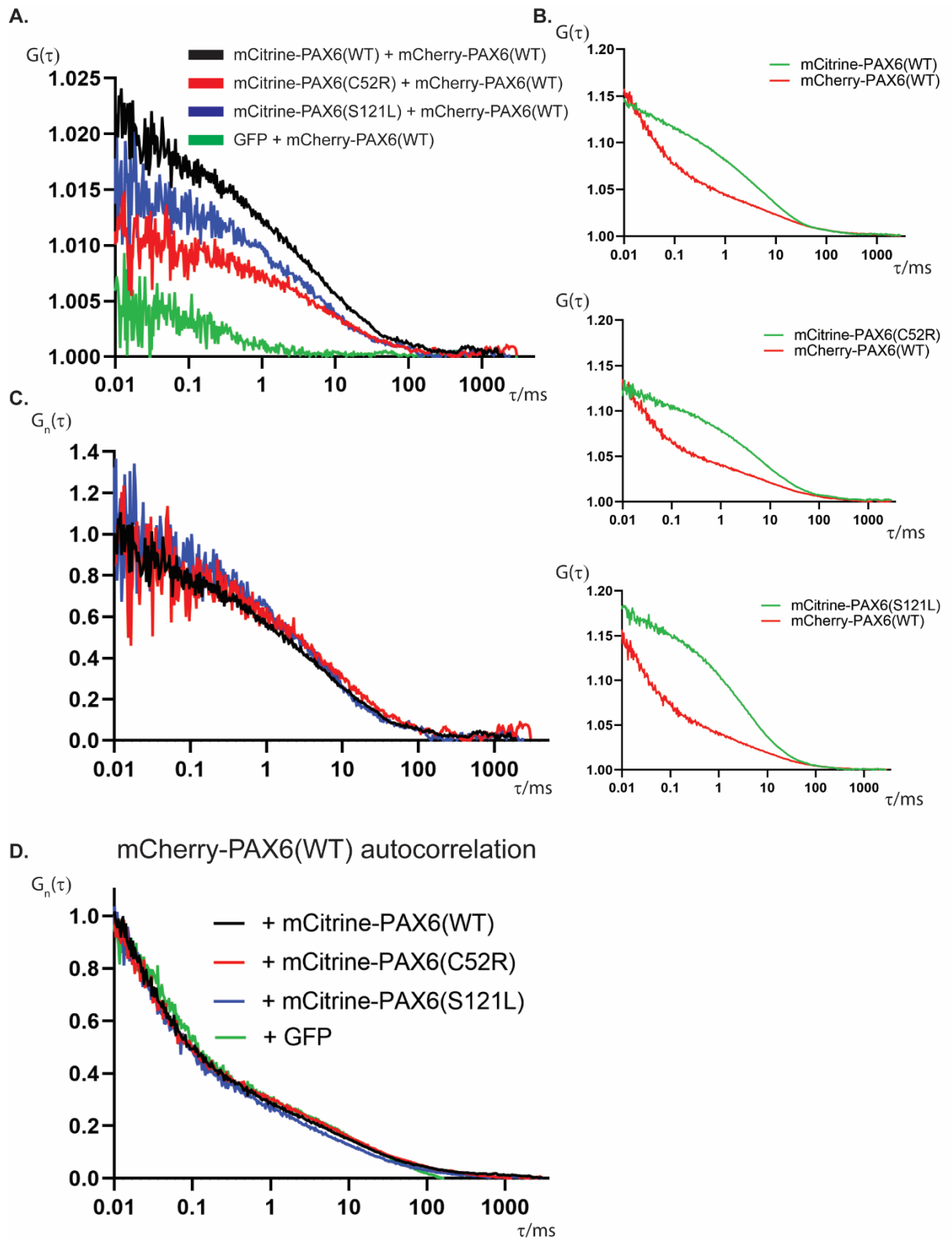


Figure 5-4: PAX6 missense variants influence homo-complex formation efficiency. The autocorrelation function data were fitted with an established 3D diffusion mathematical model for obtaining the fractions and diffusion coefficients of the free moving (fast) and chromatin-bound (slow) molecules **A.** Mean cross-correlation curves compared against the negative control (GFP + mCherry-PAX6) **B.** Mean autocorrelation curves showing similar concentrations between co-expressed proteins **C.** Mean cross-correlation curves whereby $G(\tau)$ is normalised to 1 at zeroth (0^{th}) lag time **D.** Mean mCherry-PAX6(WT) autocorrelation curve comparison, whereby $G(\tau)$ is normalised to 1 at the 0th lag time. mCherry PAX6 (WT) +GFP $N_{cells}=12$, mCherry-PAX6 (WT) + mCitrine-PAX6 (WT) $N_{cells}=30$, mCherry-PAX6 (WT) + mCitrine-PAX6 (C52R) $N_{cells}=27$, mCherry-PAX6 (WT) + mCitrine-PAX6 (S121L) $N_{cells}=19$

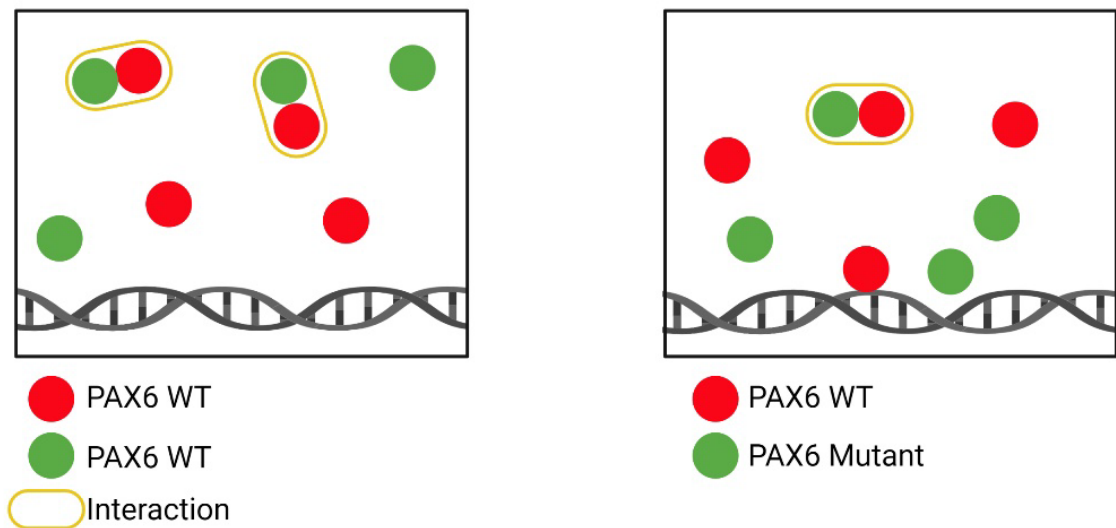


Figure 5-5: Schematic representation of PAX6 trans-interactions. The majority of interactions appear to occur away from chromatin whereby PAX6 mutants interact less efficiently

In summary, the above experiments show that PAX6 molecules can interact in trans (Figure 5-5). The investigated PAX6 mutants can also form these trans-interactions, although not as efficiently as the WT protein. WT PAX6 condensates were not observed when PAX6 mutants were co-expressed (Figure 5-3). Protein condensate formation depends on protein-protein interactions. It is therefore possible that the reduced interaction efficiency of mutant-WT PAX6, negatively influences condensate formation.

5.2.3 The trans-interaction of PD with HD is altered in the presence of PAX6 mutants

PAX6 has complex DNA-binding properties because of its multiple DNA-binding subdomains (PAI, RED, HD) which can be used in varying combinations when binding to DNA (Grapp *et al.*, 2009). It has been shown that canonical PAX6 can interact with PAX6 (δ PD) possibly via trans PD-HD interactions leading to gene regulation (Chauhan *et al.*, 2004a). I therefore wanted to investigate this interaction in the context of PAX6 pathogenic variants. I expressed WT or mutant PD and PAX6 (δ PD) labelled with either mCitrine or mCherry in HEK293 cells (Figure 5-6). When WT PD and PAX6 (δ PD) were co-expressed, PAX6 (δ PD) did not partition to the PD condensates, in line with the PD showing enhanced condensate formation, and suggesting PD intermolecular interactions are the driving force for condensate formation. In the case of PD mutants (S121L and C52R), large foci were present which again did not appear to partition with PAX6 (δ PD). In all instances, PAX6 (δ PD) subnuclear localisation was not greatly influenced when either WT or mutant PD were co-expressed. Subsequently, FCCS measurements were carried out in HEK293 cells co-expressing WT or mutant PD in the presence of PAX6 (δ PD). All mean cross-correlations curves had an amplitude higher than the negative control (GFP + PAX6 δ PD) implying trans-interactions between the PD and HD (Figure 5-7A). The amplitude of all cross-correlation curves was similar, implying that the efficiency of complex formation is comparable between the WT and mutant proteins. When the mean cross-correlation curves of the WT were compared against S121L and C52R mutants (Figure 5-7C-D), pronounced differences in decay times were observed. The observed differences suggest that even though complex formation efficiency is similar (similar amplitude), the molecular behaviour of the complex is different. The longer characteristic decay times of the WT mean cross-correlation curve, suggests that the complex has longer-lived and stronger chromatin interactions than in the presence of the two mutants. When the WT mean cross-correlation curve was compared to the R92Q variant (predicted non-disruptive) though (Figure 5-7E), the differences in decay times were not as

pronounced. These results suggest that these particular mutants of PAX6 form complexes with PAX6 (δ PD) equally efficiently, but such complexes bind chromatin less strongly than the WT PD. To further corroborate these observations, comparisons of PD autocorrelations in the presence and absence of the PAX6 (δ PD) were made (Figure 5-8).

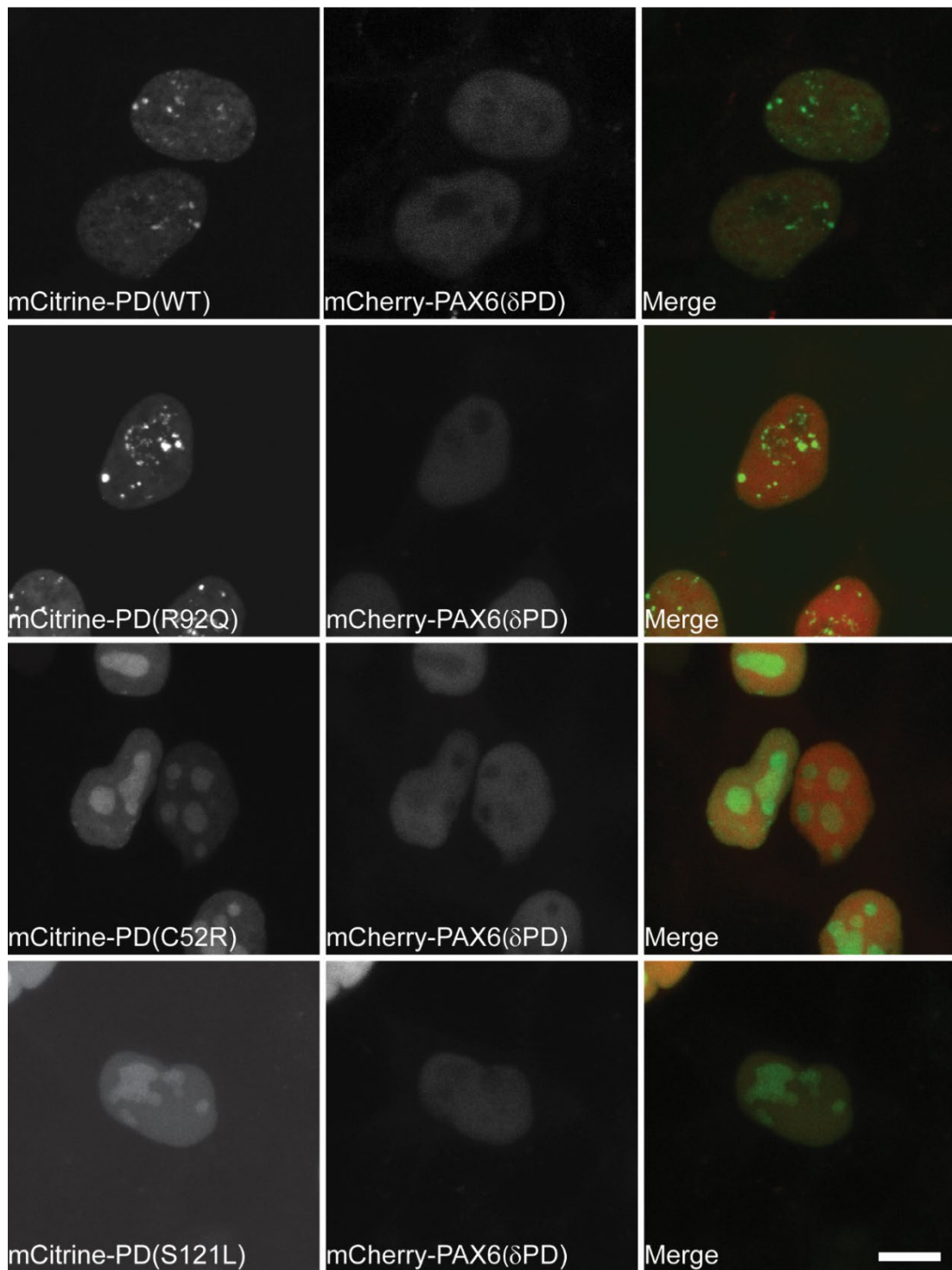


Figure 5-6: Confocal images (40x lens) of HEK293 cells co-expressing mCherry-PAX6(δ PD) and WT or mutant mCitrine-PD. The images are representative of cells that were used for FCS measurements; Scale bar = 10 μ m

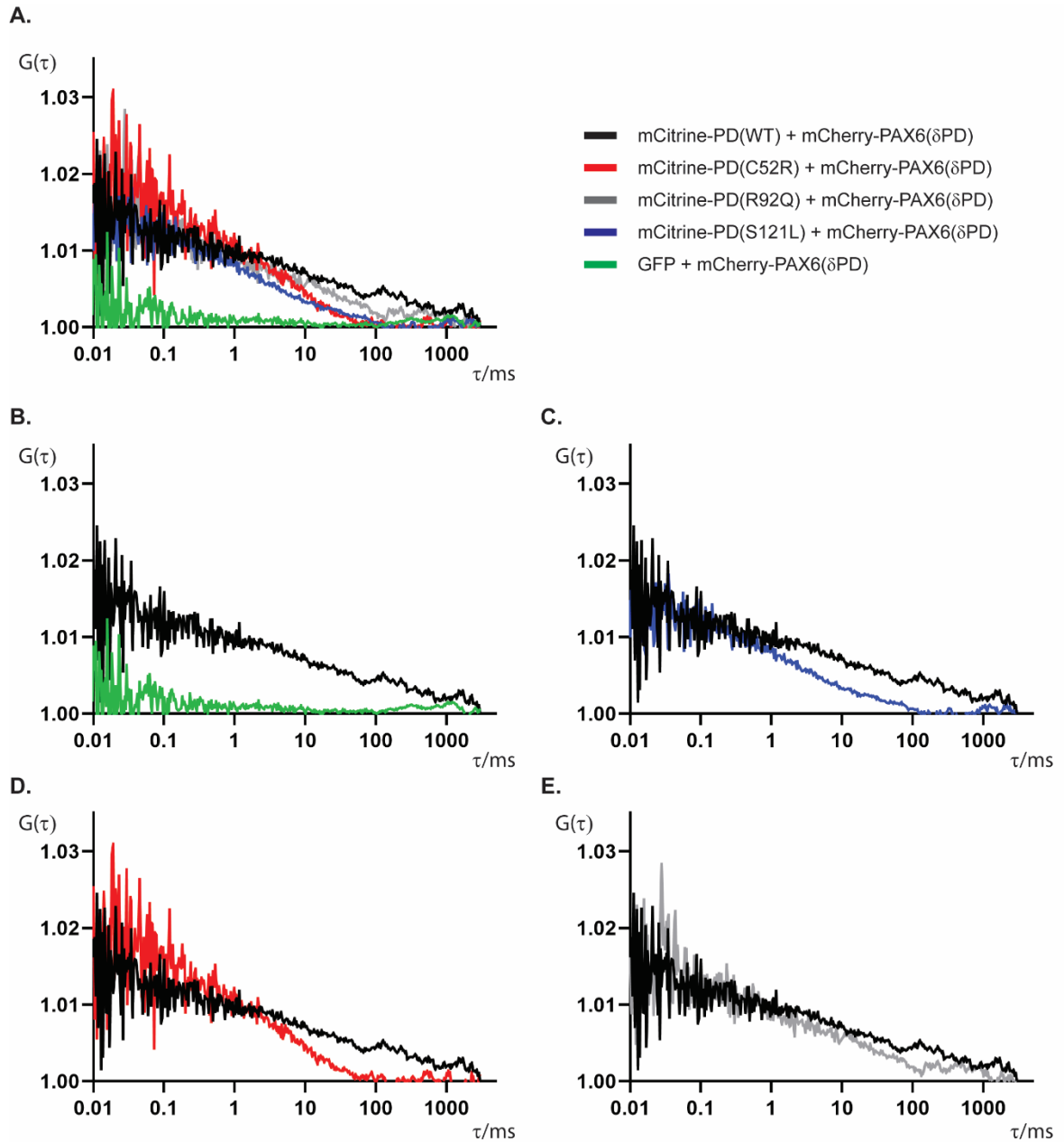


Figure 5-7: The trans-interaction of PD with HD is altered in the presence of PAX6 mutants. Cross-correlation curve comparison between PAX6 (δ PD) and PD. The autocorrelation function data were fitted with an established 3D diffusion mathematical model for obtaining the fractions and diffusion coefficients of the free moving (fast) and chromatin-bound (slow) molecules **A.** Mean cross-correlation curve comparison between PAX6(δ PD) and WT or mutant PD **B.** Mean cross-correlation comparison between PAX6(δ PD) and PD(WT) vs PAX6(δ PD) and GFP **C.** Mean cross-correlation comparison between PAX6(δ PD) and PD(WT) vs PAX6(δ PD) and PD(S121L) **D.** Mean cross-correlation comparison between PAX6(δ PD) and PD(WT) vs PAX6(δ PD) and PD(C52R) **E.** Mean cross-correlation comparison between PAX6(δ PD) and PD(WT) vs PAX6(δ PD) and PD(R92Q); mCitrine-PD (WT) + mCherry-PAX6 (δ PD) $N_{\text{cells}}=13$; mCitrine-PD (C52R) + mCherry-PAX6 (δ PD) $N_{\text{cells}}=13$; mCitrine-PD (S121L) + mCherry-PAX6 (δ PD) $N_{\text{cells}}=10$; mCitrine-PD (R92Q) + mCherry-PAX6 (δ PD) $N_{\text{cells}}=11$; GFP + mCherry-PAX6 (δ PD) $N_{\text{cells}}=11$

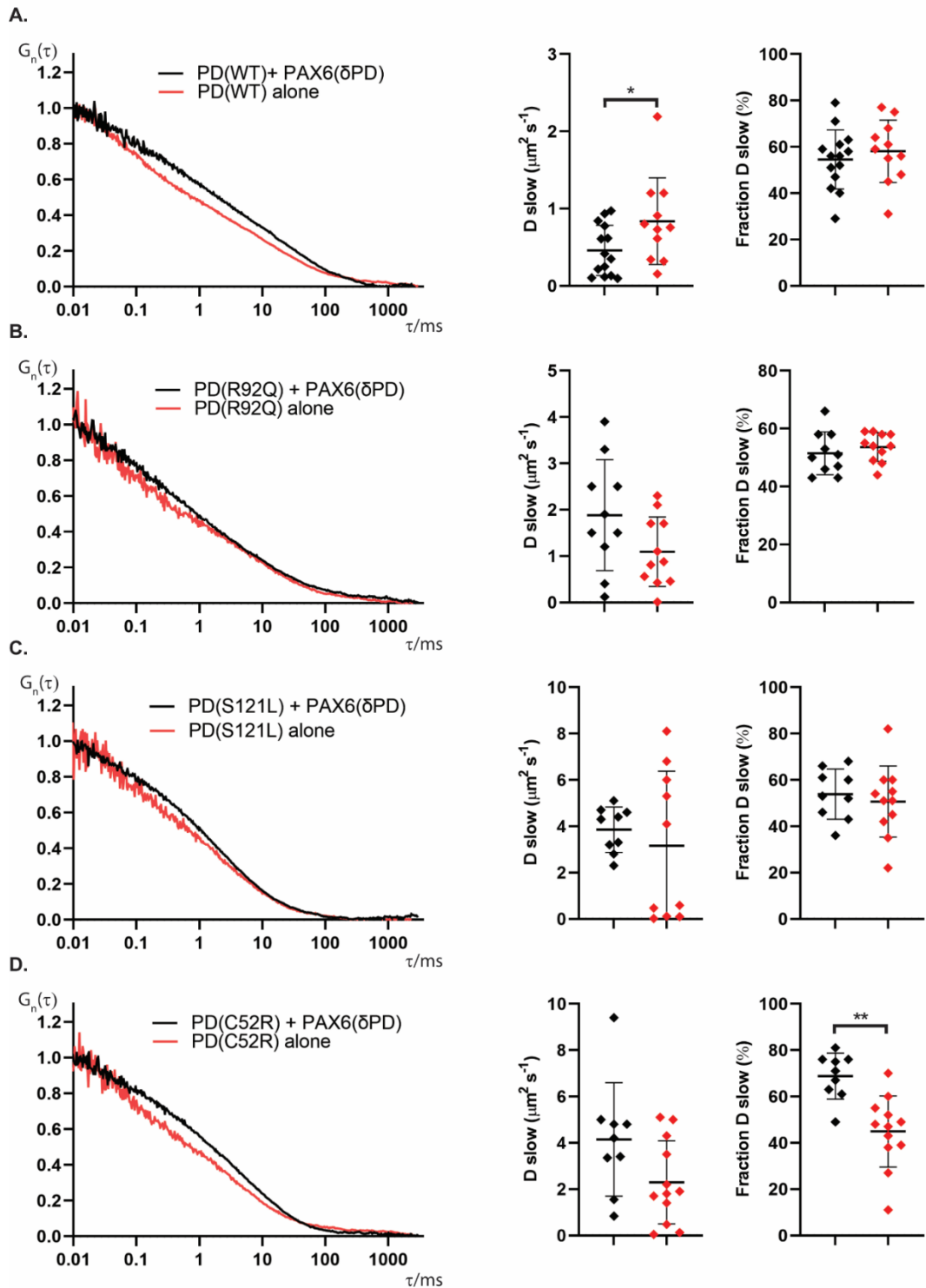


Figure 5-8: Comparison of PD (WT), PD (R92Q), PD (S121L) and PD (C52R) in the absence or presence of PAX6 (δPD). The autocorrelation function data were fitted with an established 3D diffusion mathematical model for obtaining the fractions and diffusion coefficients of the free moving (fast) and chromatin-bound (slow) molecules **A.** Mean mCitrine-PD (WT) autocorrelation curve comparison in the absence or presence of PAX6(δPD), whereby $G(\tau)$ is normalised to 1 at the 0th lag time. Comparison of the diffusion coefficient of the slow moving PD(WT) molecules, as derived by autocorrelation fitting and comparison of the fraction of the total N molecules undergoing slow diffusion; +PAX6 (δPD) $N_{\text{cells}}=14$, alone $N_{\text{cells}} = 11$ **B.** Comparison of PD (R92Q); +PAX6 (δPD) $N_{\text{cells}}=10$, alone $N_{\text{cells}} = 11$ **C.** Comparison of PD (S121L); +PAX6 (δPD) $N_{\text{cells}}=9$, alone $N_{\text{cells}} = 10$ **D.** Comparison of PD (C52R); +PAX6 (δPD) $N_{\text{cells}}=9$, alone $N_{\text{cells}} = 12$; p-value $<0.05= *$; $<0.01= **$; $<0.001= ***$

Comparison of mean autocorrelation curves of WT PD in the absence or presence of PAX6 (δ PD) demonstrated differences in decay times, suggesting that PAX6 (δ PD) increases chromatin binding properties of the PD. In fact, the diffusion coefficient of the slow-moving PD molecules was significantly lower when PAX6 (δ PD) was present. These differences were not observed when S121L and C52R mean autocorrelations curves were compared in the presence or absence of PAX6 (δ PD) (Figure 5-8C-D). Additionally, even though the mean cross-correlation curve of R92Q was relatively similar to the WT (Figure 5-7E), no significant differences were observed upon autocorrelation comparison (Figure 5-8B). The overall differences observed when autocorrelation curves were compared in the presence or absence of PAX6 (δ PD) (Figure 5-8), agree with the differences observed in mean cross-correlation curve comparison (Figure 5-7). PAX6 (δ PD) presence decreases the diffusion coefficient of the slow PD molecules, suggesting stronger and longer-lived chromatin interactions, but does not significantly influence the overall nucleoplasmic diffusion of the mutants. These results suggest that the PD and HD can interact away and on chromatin. The PAX6 mutants, retain their ability to interact with PAX6 (δ PD) with the same efficiency but this interaction is likely away from chromatin (Figure 5-9).

These findings demonstrate the ability of trans-interactions of PAX6 DBDs but also uncover an additional potential explanation as to the pathogenic nature of these variants.

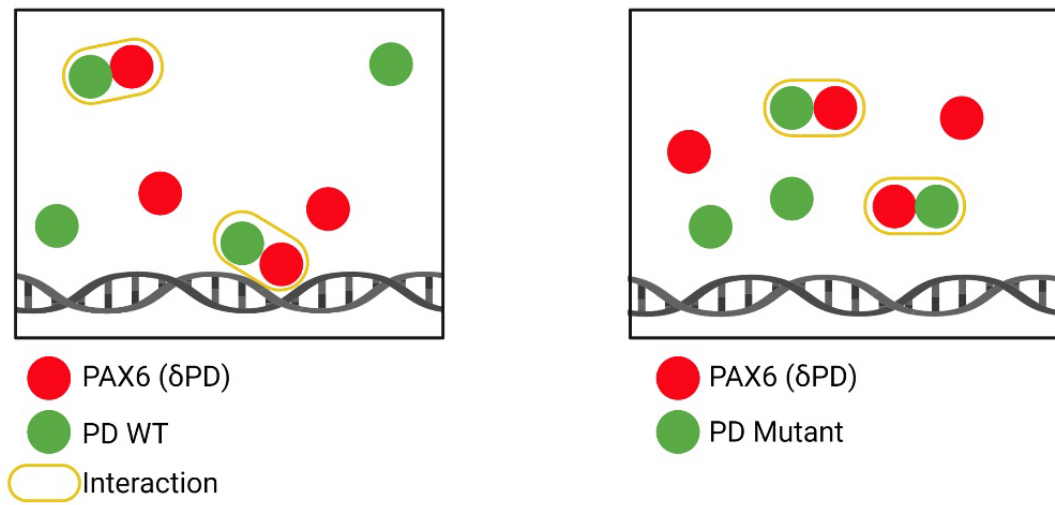


Figure 5-9: Schematic representation of PD-HD trans-interactions. Investigated PD mutants do not influence heterocomplex efficiency but unlike the WT, the majority of interactions are away from chromatin

5.2.4 PAX6 and SOX2 mutations reduce heterocomplex formation efficiency

PAX6 and SOX2 are two key TFs in ocular development. Previous studies have shown that these TFs have the ability to cross-regulate each other and to form complexes to regulate target genes. It has previously been suggested that PAX6 and SOX2 form tertiary complexes on target enhancers which are stabilised by both protein-protein and protein-DNA interactions. For example, during lens placode formation, PAX6 forms a molecular complex with SOX2 on lens-specific enhancers such as the chicken δ -crystallin (DC5) enhancer (Kamachi *et al.*, 2001). Another example includes the PAX6 Le9 enhancer where the synergistic interaction of SOX2 and PAX6 positively regulates PAX6 expression (Aota *et al.*, 2003). Qualitative methods assessing PAX6 and SOX2 binding such as chromatin immunoprecipitation and ChIP-seq (Thakurela, Tiwari, Schick, Garding, Ivanek, Berninger and Vijay K Tiwari, 2016) provide static information, derived from fixed cell populations (Valouev *et al.*, 2008). Despite being informative, these analysis fails to provide insight into TF dynamic behaviour in real-time. It is unclear how PAX6 and SOX2 influence each other at the molecular level and how this is altered when either TF bears a pathogenic mutation. To investigate this, FCCS experiments were carried out in HEK293 cells co-expressing WT and/or mutant mCitrine-PAX6 and SOX2-mCherry at similar concentrations (Figure 5-10B). PAX6 (WT) was still able to form condensates when WT or mutant (R74P; HMG box) SOX2 protein was co-expressed, but SOX2 did not particularly partition in these condensates (Figure 5-10A). All mean cross-correlation curves had amplitudes higher than the negative control (GFP + SOX2 (WT)) suggesting that complexes were formed in all three conditions (PAX6 (WT) + SOX2 (WT); PAX6 (WT) + SOX2 (R74P); PAX6 (C52R) + SOX2 (WT)). The highest occurrence of complex formation was obtained when both protein were wild type, whereas when one of the two proteins was mutated this was reduced by approximately half (Figure 5-10C). Normalisation of the cross correlation curves ($G(\tau) = 1$ at the 0th lag time) indicated no differences in decay time of the complexes, suggesting that these particular mutants of PAX6 and SOX2 form TF-TF complexes less

efficiently, but such complexes bind chromatin equally strongly (Figure 5-10D). Overall, these results suggest that the investigated PAX6 or SOX2 mutants will form roughly half the number of heterocomplexes when expressed in equal amounts to the WT proteins. Considering the importance of TF molecule number, this may have an important role in influencing gene expression. For example, low-affinity binding sites are less likely to be bound by a TF or TF complex when its concentration is low, providing a further molecular explanation of the pathogenic nature of the investigated mutants.

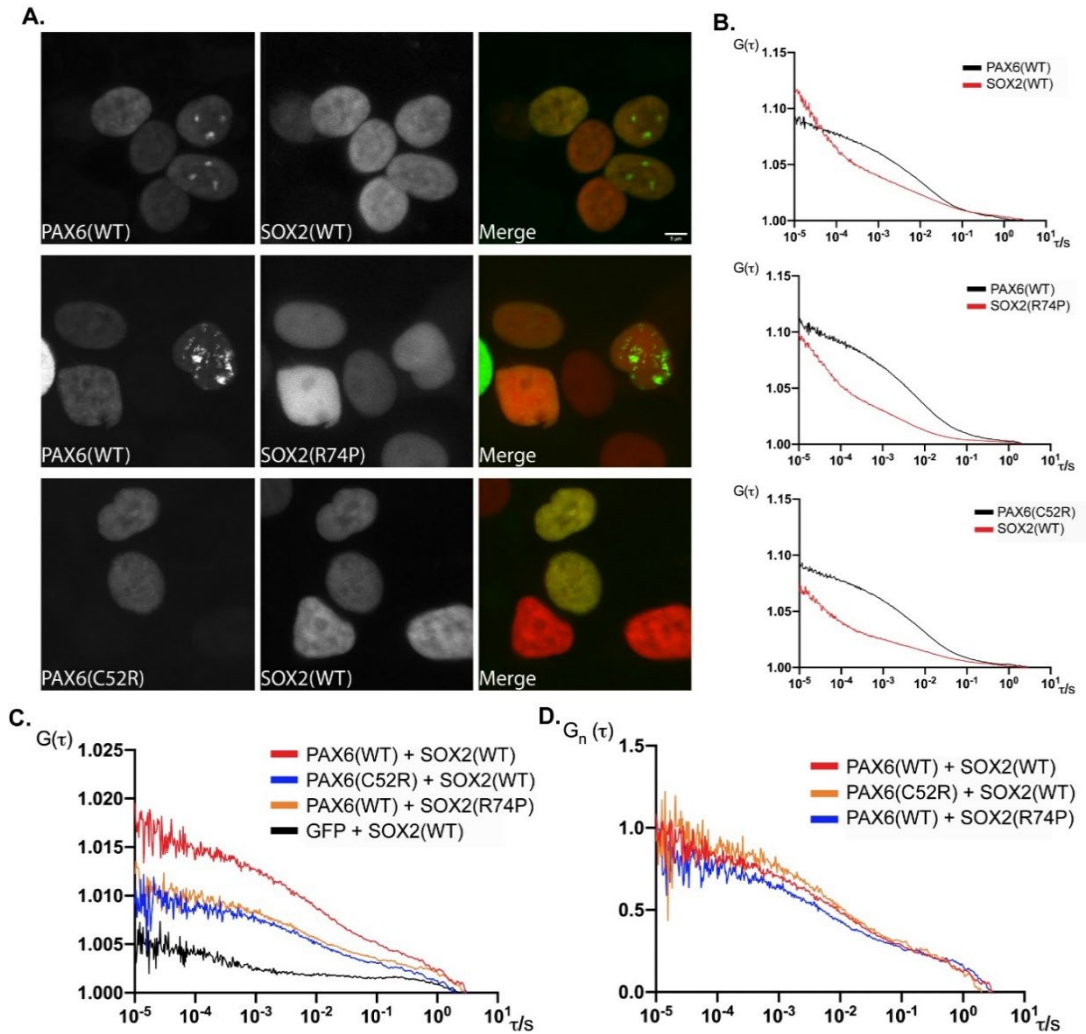


Figure 5-10: PAX6 and SOX2 mutations reduce heterocomplex formation. The autocorrelation function data were fitted with an established 3D diffusion mathematical model for obtaining the fractions and diffusion coefficients of the free moving (fast) and chromatin-bound (slow) molecules **A**. Confocal imaging (40x) of HEK293 cells co-expressing mCitrine-PAX6 and SOX2-mCherry. The images are representative of cells that were used for FCS measurements **B**. Mean autocorrelation curves showing similar concentrations of PAX6 and SOX2 **C**. Mean cross-correlation curves compared to the negative control. Cross-correlation curve amplitude is directly proportional to the concentration of co-diffusing complexes. **D**. Normalised mean cross-correlation curves where $G(\tau)$ is normalised to 1 at the zeroth (0^{th}) lag time. The experiment was conducted in three separate occasions and all measurements were averaged together; PAX6 (WT) + SOX2 (WT) $N_{cells}= 34$; PAX6 (C52R) + SOX2 (WT) $N_{cells}= 34$; PAX6 (WT) + SOX2 (R74P) $N_{cells}= 35$; GFP + SOX2 (WT) $N_{cells}= 16$

Although PAX6-SOX2 heterocomplex formation has previously been suggested to occur on chromatin, no conclusive experiments have previously been carried out to prove this. I therefore carried out Co-IP experiments using HEK293 cells co-expressing mCitrine-PAX6 and SOX2-mCherry, in which the interaction between PAX6 and SOX2 had been observed (Figure 5-11). Upon DNA degradation using DNase I, no interaction between PAX6 and SOX2 was detected, indicating that PAX6-SOX2 predominantly interact on chromatin as a tertiary complex (TF-TF-DNA).

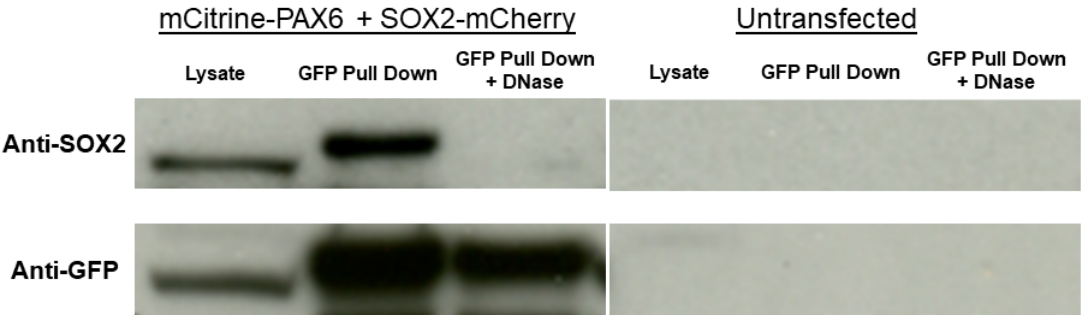


Figure 5-11: HEK293 cells co-expressed mCitrine-PAX6 and SOX2-mCherry. Cell lysate was divided in two equal aliquots in which one aliquot was DNase treated prior to GFP-trap agarose bead immunoprecipitation. mCitrine-PAX6 and SOX2-mCherry protein levels were detected via Western blot analysis

5.2.5 SOX2 influences PAX6-chromatin binding kinetics in a non-reciprocal fashion

Since PAX6-SOX2 heterocomplexes are largely formed on chromatin, I aimed to understand if this was driven/stabilised by predominantly one of the two TFs. I therefore compared the binding kinetics of each TF, when alone or in the presence of the other TF. Normalised SOX2 (WT) autocorrelation curves showed no significant differences in diffusion when PAX6 (WT) or PAX6 (C52R) were present as opposed to SOX2 (WT) alone, suggesting that PAX6 does not influence the binding kinetics of SOX2 (Figure 5-12).

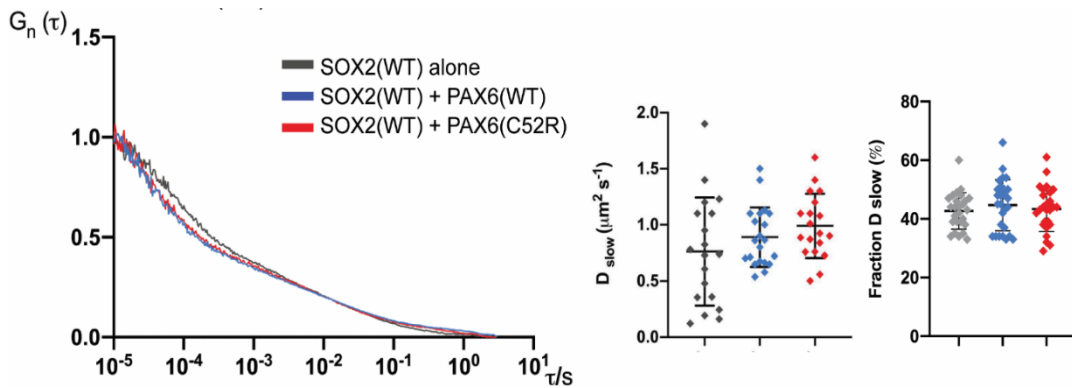


Figure 5-12: Autocorrelation curve of SOX2 (WT) whereby $G(\tau)$ is normalised to 1 at the 0th lag time. The autocorrelation function data were fitted with an established 3D diffusion mathematical model for obtaining the fractions and diffusion coefficients of the free moving (fast) and chromatin-bound (slow) molecules. Scatter plot of the diffusion coefficient of the slow FCS component, as derived by fitting; Scatter plot of the fraction of the slow diffusing molecules of SOX2 (WT); The experiment was conducted in three separate occasions and all measurements were averaged together; PAX6 (WT) + SOX2 (WT) $N_{cells}= 34$; PAX6 (C52R) + SOX2 (WT) $N_{cells}= 34$; SOX2 (WT) alone $N_{cells}= 26$; One-way ANOVA with post-hoc Tukey Test was performed; $p\text{-value} < 0.05 = *$; $< 0.01 = **$; $< 0.001 = ***$

Conversely, normalised PAX6 (WT) autocorrelation curves displayed a shift to longer characteristic decay times when SOX2 (WT) was present, as opposed to PAX6 (WT) alone, suggesting stronger and longer-lived chromatin interactions. No significant differences were observed in the diffusion coefficient of the slow moving molecules of PAX6 (WT) when SOX2 (R74P) was present, suggesting that SOX2 stabilizes the PAX6-SOX2 heterocomplex on chromatin via a DBD dependent manner. In addition, no significant changes in the fraction of slow-moving PAX6 molecules were observed, suggesting that the shift of the autocorrelation curve to longer characteristic decay times was largely attributed to stronger and longer lived chromatin interactions, rather than an increase in PAX6-chromatin interactions. These results suggest that a generic principle of PAX6-SOX2 complex stabilisation applies in widespread chromatin locations, as evidenced by the repeatedly observed slow diffusion of PAX6 when in the presence of SOX2 (WT).

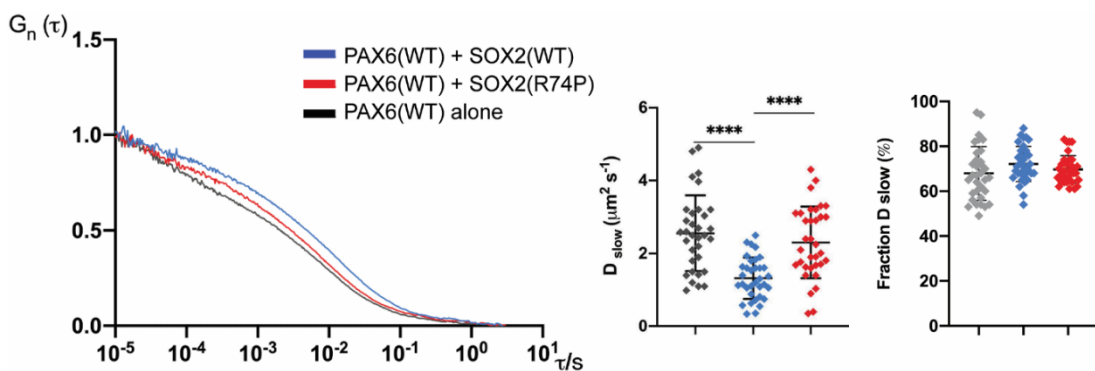


Figure 5-13: Autocorrelation curve of PAX6 (WT) whereby $G(\tau)$ is normalised to 1 at the 0th lag time. The autocorrelation function data were fitted with an established 3D diffusion mathematical model for obtaining the fractions and diffusion coefficients of the free moving (fast) and chromatin-bound (slow) molecules. Scatter plot of the diffusion coefficient of the slow FCS component, as derived by fitting; Scatter plot of the fraction of the slow diffusing molecules of PAX6 (WT); The experiment was conducted in three separate occasions and all measurements were averaged together; PAX6 (WT) + SOX2 (WT) $N_{cells}= 34$; PAX6 (C52R) + SOX2 (WT) $N_{cells}= 34$; PAX6 (WT) alone $N_{cells}= 40$; One-way ANOVA with post-hoc Tukey Test was performed; p-value < 0.05 = *; < 0.01 = **; < 0.001 = ***

Considering that PAX6 and SOX2 interaction is largely on chromatin (Figure 5-11) and that SOX2 influences the binding kinetics of PAX6 in a non-reciprocal fashion via DBD dependent manner (Figure 5-12; Figure 5-14), a possible mechanism illustrated in Figure 5-13 has been proposed. It is suggested that SOX2 initially interacts with chromatin (importance of intact DBD; Figure 5-14), followed by PAX6 recruitment and binding which would not have occurred to the same degree if SOX2 was absent (overall increase in binding kinetics observed for PAX6; Figure 5-14).

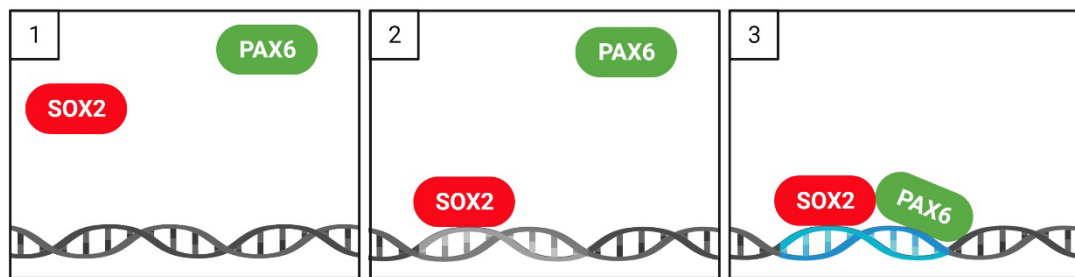


Figure 5-14: Schematic representation of PAX6-SOX2 heterocomplex formation on chromatin. PAX6 and SOX2 largely interact on chromatin whereby it is likely that SOX2 firstly interacts with chromatin, allowing for PAX6 to bind

5.2.6 A reporter assay investigating PAX6-SOX2 stoichiometry on transcription

DC5 is an enhancer of the δ -crystalline gene required for lens induction. Qualitative methods identified that PAX6 and SOX2 form a tertiary complex on this enhancer which is stabilised by both protein-protein and protein-DNA interactions (Kamachi *et al.*, 2001). I therefore wanted to create a reporter assay whereby PAX6 and SOX2 stoichiometry would be compared to transcriptional output. To carry this out, I created a plasmid in which 28 sequence repeats of the DC5 enhancer were placed upstream of a minimal promoter. mCerulean fused to 3x nuclear export signal (NES) would be transcribed and the protein would translocate to the cytoplasm. Activation of the reporter would occur upon co-expression of mCitrine-PAX6 and SOX2-mCherry and binding to the DC5 enhancer. By designing this assay the concentrations of PAX6, SOX2 and heterocomplex efficiency could be measured by FCCS and compared to transcriptional output of the reporter measured by FCS (mCerulean).

To assess if the reporter plasmid was responding as predicted, it was transfected into HeLa cells together with co-expression of mCitrine-PAX6 and SOX2-mCherry (Figure 5-15). HeLa cells were used to conduct pilot experiments as they are an easy cell line to use, has high transfection efficiency and plasmids can readily integrate into the genome. mCerulean localised in the cytoplasm as predicted and its fluorescent signal was much greater when compared to no co-expression of the two TFs suggesting that the assay was responding as predicted. Expression of mCerulean was still observed when no TFs were co-expressed possibly due to a partially leaky system. A similar mCerulean expression was observed when the PD was expressed as opposed to the full length PAX6 protein which further suggested that the reporter was responding since DC5 has a PD binding motif (Figure 5-15).

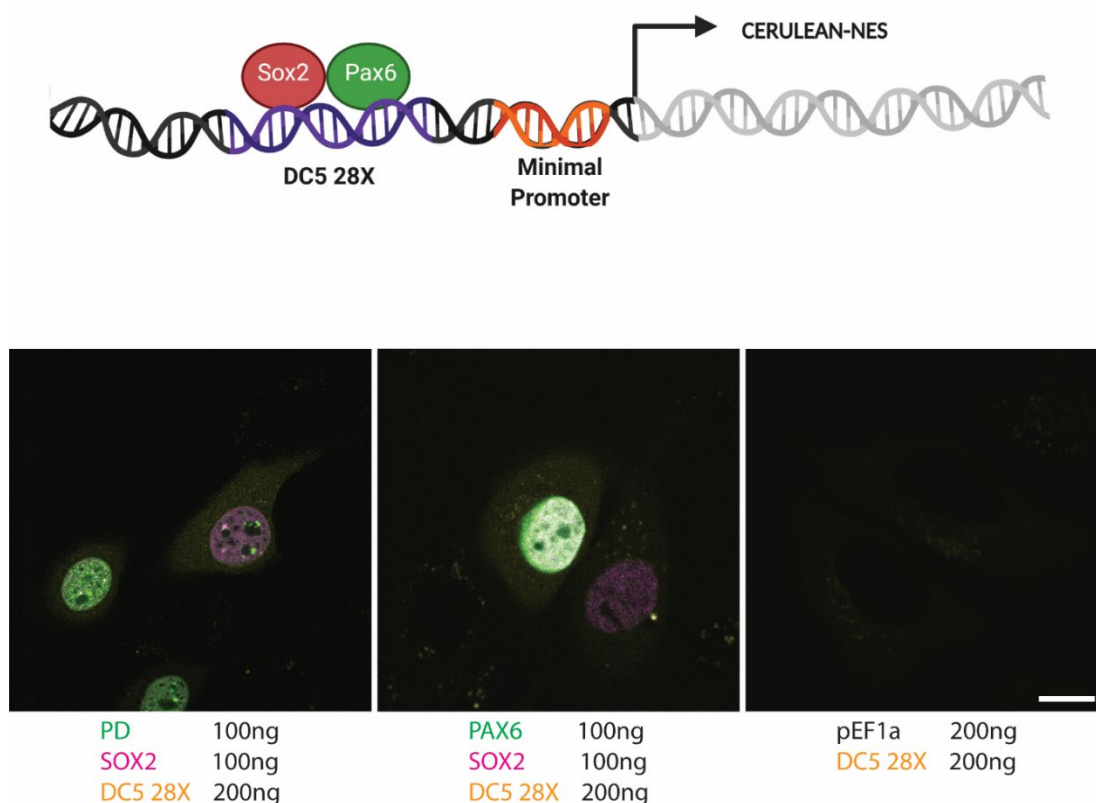


Figure 5-15: Illustration of DC5 reporter cassette. HELa cells co-expressing PD, SOX2 or PAX6. The concentrations indicate amounts of each transfected plasmid

Subsequently, I wanted to measure the diffusion and concentration of mCerulean to ensure that the APD detectors could record fluorescence fluctuations without any interference from the two other fluorophores. A mixture of cells expressing mCerulean at different concentrations were measured as depicted by the red crosses in Figure 5-16. mCerulean was only visible in cells which were co-expressing both mCitrine-PAX6 and SOX2-mCherry. The mean diffusion coefficient of mCerulean was $12\mu\text{m}^2\text{s}^{-1}$ which is within the range of cytoplasmic free GFP diffusion (Gura Sadovsky *et al.*, 2017). The average calculated concentration varied from 17nM to 81nM with a mean of 49nM falling well within the detection range of FCS. Following reporter validation, I aimed to create a stable cell line containing equal copies of the reporter sequence to circumvent the inherent variability of transient transfection. Transfected HeLa cells were selected against blasticidin (15 $\mu\text{g}/\text{ml}$) for 2 weeks

(to select for DC5 plasmid integration), but surviving cells did not express mCerulean upon co-expression of mCitrine-PAX6 and SOX2-mCherry. Unfortunately due to COVID-19 restrictions there was not sufficient time to further troubleshoot and try a more relevant cell line like NPCs and aRPE-19. Nevertheless it remains an enticing approach to quantitatively, qualitatively and functionally investigate TF stoichiometry in relation to function at the live cell stage.

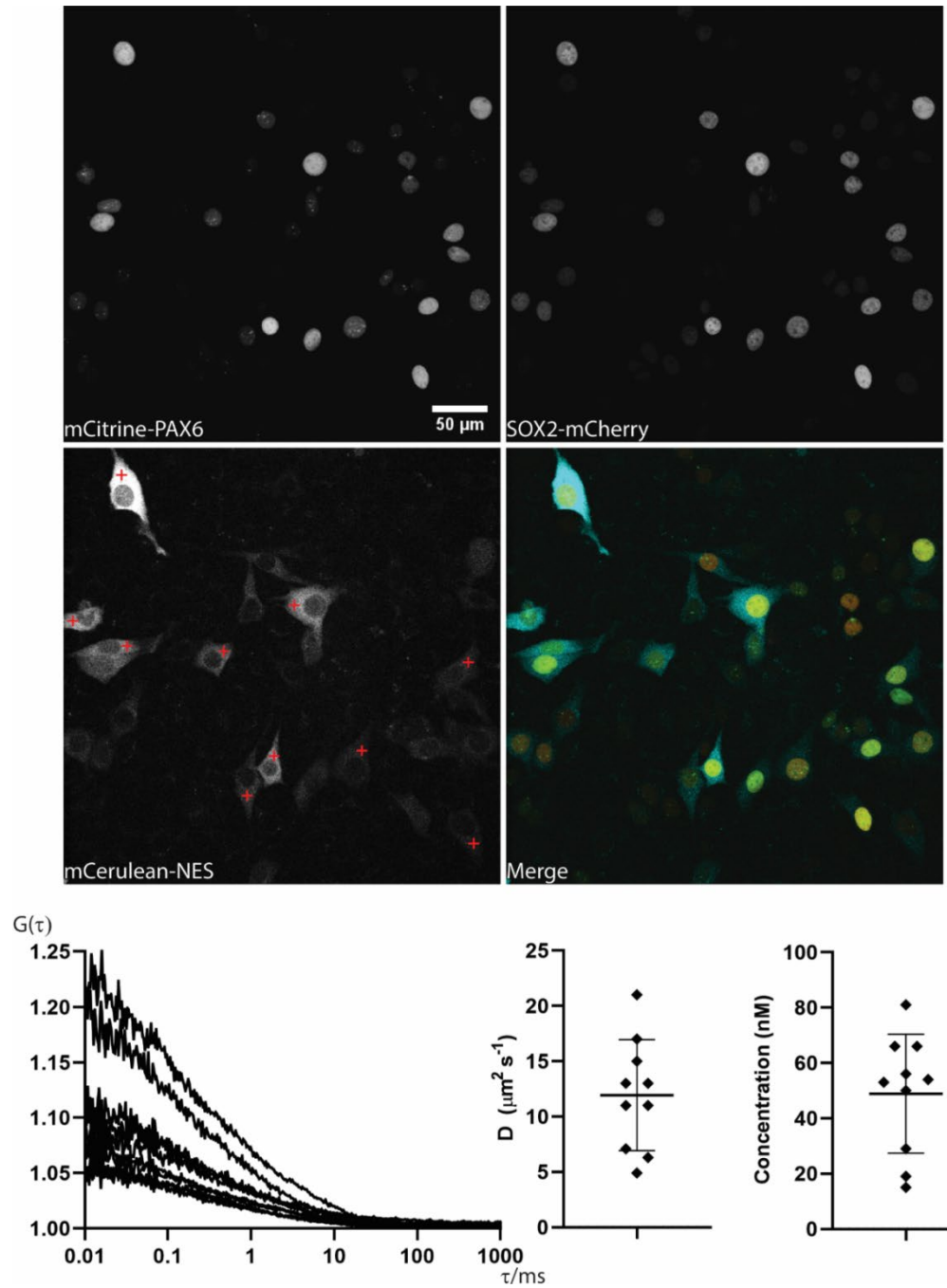


Figure 5-16: HELA cells co-expressing mCitrine-PAX6 and SOX2-mCherry and mCerulean-NES from the transfected DC5 reporter plasmid. Red crosses indicate regions where FCS measurements were taken. The autocorrelation function data were fitted with an established 3D diffusion mathematical model for obtaining the fractions and diffusion coefficients of the free moving (fast) and chromatin-bound (slow) molecules. Autocorrelation curves from each FCS measurement are indicated below. The mean diffusion coefficient and concentration were subsequently calculated.

5.2.7 Endogenous labelling of SOX2 with mCherry in mESCs

In order to differentiate mESCs into optic cup organoids, thus allowing us to investigate SOX2-PAX6 interaction at the endogenous level in a developmentally relevant background, I generated an endogenously labelled SOX2-mCherry mESCs containing a Rx:GFP reporter (retina homeobox gene required for retinal cell fate determination). This cell line was generated by using a CRISPR-Cas9 ribonucleoprotein with a linear double stranded DNA repair template. Initial FCS experiments were carried out, in which an average concentration was obtained through the fitting of the autocorrelation curves with an established 3D diffusion mathematical model consisting of 1 triplet state and 2 diffusion populations. The mean calculated concentration of SOX2 was 112 nM (sd=34) which falls within the range of previously calculated SOX2 concentrations (Strebinger *et al.*, 2019).

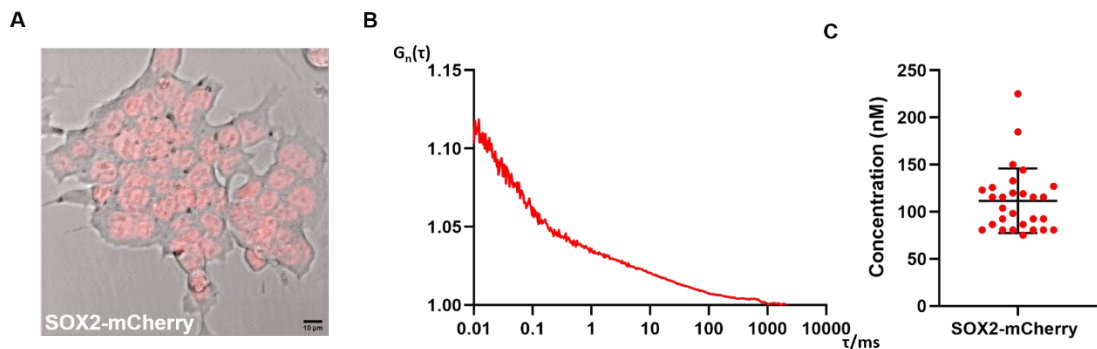


Figure 5-17: Endogenously labelled SOX2 with mCherry at the C-terminus in mESCs (heterozygous) **A.** Confocal imaging (10x) of endogenously labelled SOX2-mCherry in mESC. The cells are representative of cells that were used for FCS measurements. One measurement per cell was carried out to minimize photodamage. **B.** Mean autocorrelation curve of SOX2-mCherry derived from FCS measurements $N_{\text{cells}} = 28$ **C.** Nuclear concentration of endogenous SOX2-mCherry in mESCs. The autocorrelation function data were fitted with an established 3D diffusion mathematical model for obtaining the nuclear concentration. Each data point represents one cell nucleus.

Subsequently, FCS measurements between transfected SOX2-mCherry in HEK293 cells and endogenously labelled SOX2-mCherry in mESCs were compared. Even though the human SOX2 protein was transfected in HEK293 cells, SOX2 protein sequence similarity between human and mouse is 98%, with the HMG domain been largely identical. When the normalised mean autocorrelation curves were compared, no major differences were observed (Figure 5-18 C). In addition, no significant differences were observed when the diffusion coefficient of the slow moving molecules was compared, suggesting that the dynamic binding behaviour of SOX2 with chromatin was similar in both systems. A significant difference was observed in the fraction of the slow moving molecules (Figure 5-18 B), which may partly be due to the completely different cell types and ultimately, the overall chromatin architecture. Nevertheless, this experiment suggests that the biophysical behaviour of overexpressed SOX2 in HEK293 cells is similar to the endogenous one.

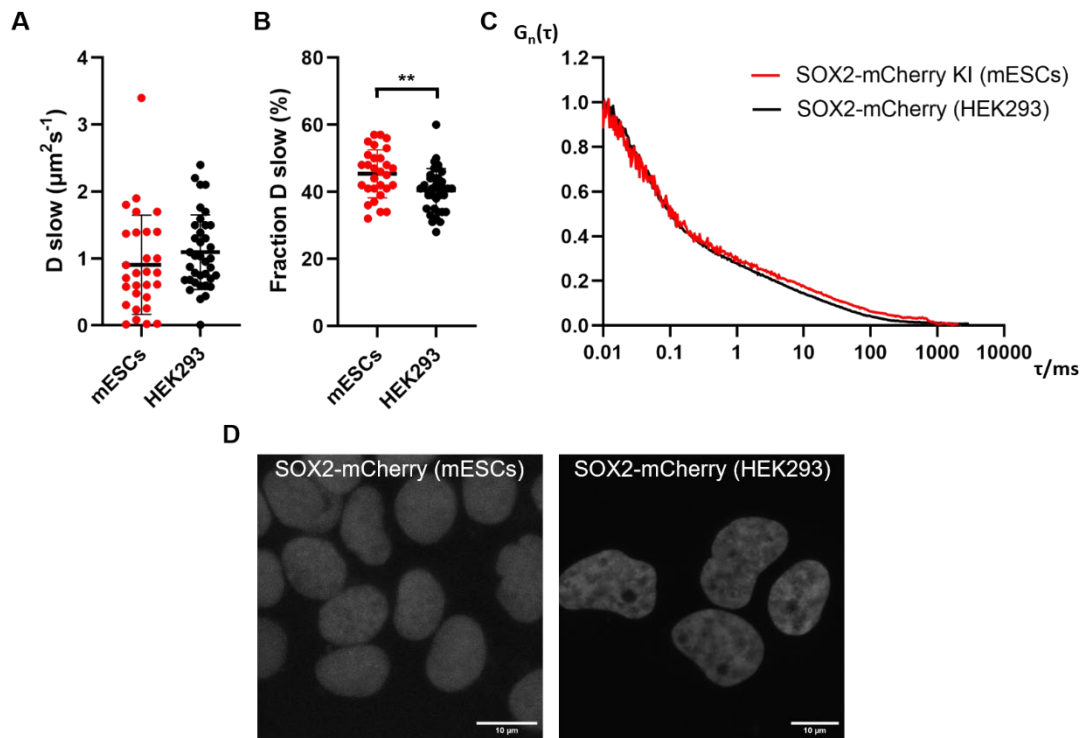


Figure 5-18: Comparison of SOX2-mCherry (human) overexpressed in HEK293 cells against endogenously labelled SOX2-mCherry (mouse) in mESCs. The autocorrelation function data were fitted with an established 3D diffusion mathematical model for obtaining the fractions and diffusion coefficients of the free moving (fast) and chromatin-bound (slow) molecules **A**. Slow diffusion coefficient comparison of SOX2-mCherry, as derived by fitting **B**. Fraction of the total N molecules undergoing slow diffusion **C**. Mean autocorrelation where $G(\tau)$ is normalised to 1 at the zeroth (0^{th}) lag time **D**. Confocal images of the nuclear distribution of SOX2-mCherry in HEK293 and mESCs. The image is representative of cells that were used for FCS measurements. SOX2-mCherry (mESCs) $N_{\text{cells}} = 28$, SOX2-mCherry $N_{\text{cells}} = 37$. One measurement per cell was carried out to minimize photodamage. Student's T-test; p-value $< 0.05 = *$; p-value $< 0.01 = **$; p-value $< 0.001 = ***$

5.2.8 FCS analysis of endogenously labelled PAX6 in mESC optic cup organoids and 12 hpf zebrafish embryo

Over the past year, our lab has been able to generate a confirmed endogenously labelled heterozygous mCitrine-Pax6 α zebrafish line (Jack Stoddart). Currently, the zebrafish line is in the process of breeding a homozygous line to further assess if the tag is non-disruptive, although this is proving to be challenging. Nevertheless, initial FCS experiments were carried out to characterise the dynamics of Pax6 α behaviour.

Similarly, in collaboration with Brianda Mohran from Professor David FitzPatrick's lab, we were able to generate a heterozygous SOX2-mCherry, homozygous HALO-PAX6 double knock-in in mESCs. These cells were differentiated into optic cup organoids although consistency, expression and HALO dye labelling have proved to be challenging. No optic cup organoid containing both endogenously labelled TFs was able to be successfully differentiated. Due to great inconsistencies in optic cup organoid differentiation, including the WT cells, we were not able to definitively state that this was due to the endogenous labelling. Nevertheless, we were able to generate an optic cup organoid using mESCs that were endogenously labelled with HALO-PAX6. Using FCS, I was also able to characterise the dynamics of PAX6 behaviour in mouse optic cup organoids. In both cases, I was able to confirm the existence of differential chromatin-binding behaviour and high variability in Pax6 cell-to-cell concentrations in both systems.

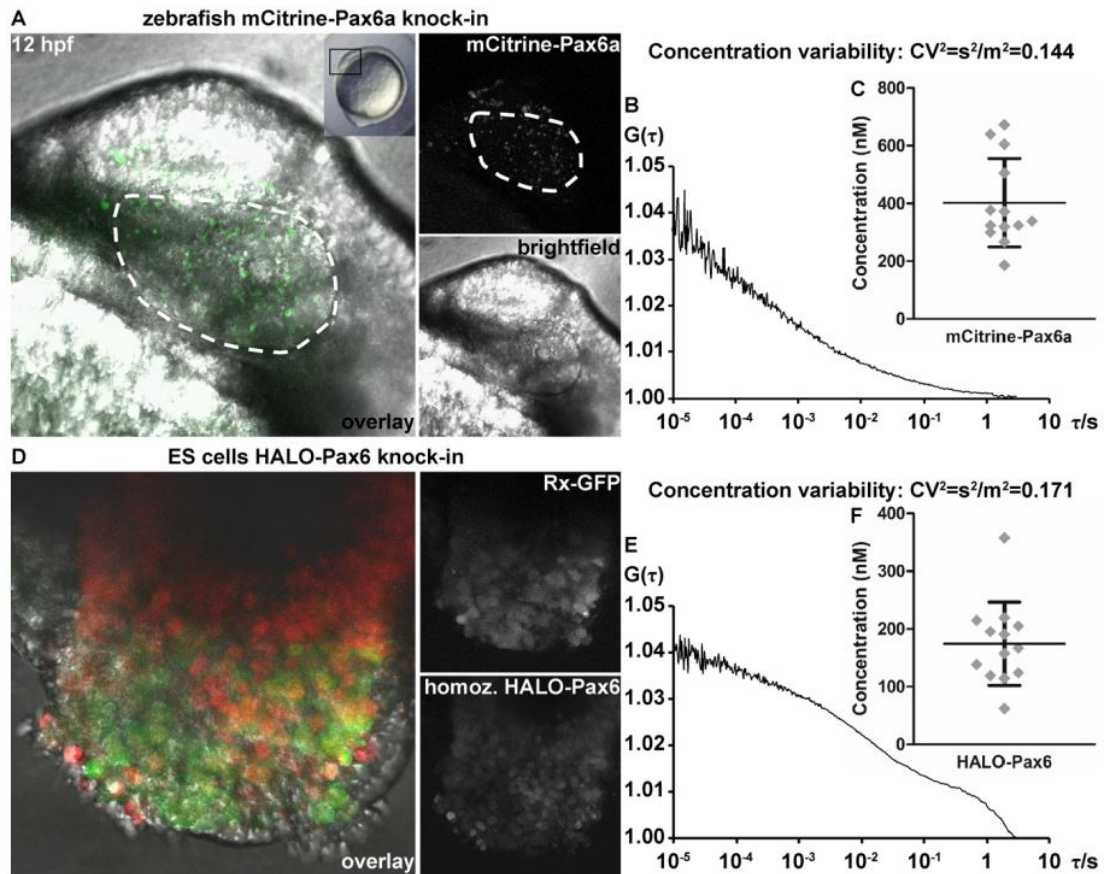


Figure 5-19: FCS experiments in endogenously labelled PAX6 in 12 hpf zebrafish embryo and mouse optic cup organoid **A**. Confocal imaging of the head of a 12hpf endogenously labelled mCitrine-Pax6a heterozygous zebrafish **B**. FCS measurements were performed to determine the absolute concentration of mCitrine-Pax6a TF molecules $N_{\text{cells}} = 12$. The mean autocorrelation curve is shown **C**. Plot of mCitrine-Pax6a internuclear concentration indicated a mean concentration of 402 nM, whereas square coefficient of variability, indicated high variability in Pax6a numbers among cells **D**. Confocal imaging (40x) of an optic cup organoid derived from mESCs at day 6 after optic cup induction, mESCs have been endogenously and homozygously tagged with HALO at the N-terminus of the PAX6 locus and have been stained with HALO TMR ligand (excitation 555nm); a GFP reporter of the endogenous Rx gene was used as a marker of ocular fate induction **E**. FCS measurements as in B, of HALO-PAX6 positive nuclei were acquired to derive the absolute concentrations of PAX6 **F**. A mean concentration of 174 nM was derived through the fitting of the autocorrelation curves with a mathematical model for 3D diffusion. This was accompanied by a high cell-to-cell variability of PAX6 concentration

5.3 Discussion

This chapter aimed to investigate the dynamic interactions between PAX6 and SOX2 as well as the interactions between the two DBDs of PAX6 (PD and HD). Furthermore, I aimed to assess how pathogenic missense mutations may influence these interactions.

Considering that all PAX6 missense mutations are heterozygous, I wanted to investigate if these mutations altered PAX6 homo-complex formation. Results obtained through FCCS experiments indicated that the PAX6 mutants were also able to form homo-complexes with WT PAX6, but not to the same extent. In addition, no WT PAX6 condensates were observed when co-expressed with mutant PAX6 protein, suggesting a possible inhibitory role in WT PAX6 condensate formation. The two DBDs of PAX6 have previously been shown to interact and co-regulate target genes (Mishra, Ivan P. Gorlov, *et al.*, 2002; Dames *et al.*, 2010; Sun *et al.*, 2015). Previous studies have shown that the canonical PAX6 isoform can physically interact with the PAX6 (δ PD) and lead to transactivation activity enhancement of PAX6 from PD binding sites (Mikkola *et al.*, 2001). FCCS experiments that were carried out in this chapter indicated that the PD and HD trans-interactions can occur on and away from chromatin. Even though PAX6 variants were able to retain their ability to interact with PAX6 (δ PD) with the same efficiency as the WT TF, the interaction was largely away from chromatin. This could potentially compromise the transactivation activity enhancement of PAX6 by PAX6 (δ PD) and as a consequence have an effect on the expression of downstream targets. Although these results provide key information on the trans-interactions of the PAX6 DBDs, it would be beneficial to repeat these experiments in a more relevant cell line with a possible read-out such as a luciferase assay.

The complex interplay between PAX6 and SOX2 has previously been investigated. Nevertheless, the majority of work has been through qualitative methods such as Co-IP and ChIP-seq (Kamachi *et al.*, 2001; Valouev *et al.*, 2008; Thakurela, Tiwari, Schick, Garding, Ivanek, Berninger and Vijay K Tiwari, 2016). Despite being informative, these analysis fail to provide insight

into the dynamic behaviour of these TFs in real-time. In addition, most work has solely focused on investigating the interaction between the two WT TFs. To address this, I carried out FCCS experiments, investigating the interaction of PAX6 and SOX2 in live cells at the single molecule level and compared my results when either TF was mutated. The results suggest that the investigated PAX6 and SOX2 mutants, C52R and R74P respectively, formed roughly half the number of heterocomplexes when expressed in equal amounts to the WT proteins. Considering the importance of TF molecule number for correct function and especially the dosage sensitivity of PAX6 in ocular development (Schedl *et al.*, 1996; Kim and Lauderdale, 2008; Matsushima *et al.*, 2011), these results suggest a further molecular explanation as to the pathogenic nature of the investigated mutants.

For the first time, a novel function of SOX2 in supporting PAX6 chromatin complexes is uncovered in this chapter. Initially, Co-IP experiments have shown that PAX6-SOX2 interactions are predominantly found on chromatin. Subsequently, FCS experiments suggested that SOX2 is able to stabilize the PAX6-SOX2 heterocomplex on chromatin via DBD dependent manner. This was initially suggested by Kamachi *et al.*, 2001, although only the DC5 enhancer was investigated. Since FCS measurements record an average diffusional behaviour over time, the complex stabilisation by SOX2 is believed to be a generic principle of the PAX6-SOX2 complex that applies in widespread chromatin locations. For future experiments it will be key to carry out ChIP-seq experiments to further validate this hypothesis. It is imperative though to conduct these experiments in a more relevant cell type such as aRPE-19 cells and ideally at the endogenous levels.

Previous work has shown that the optic tissue morphogenesis in optic cup organoids faithfully mimics the cellular architecture and developmental events which occur *in vivo* (Eiraku *et al.*, 2011), making them a suitable system for studying how ocular TF concentrations and chromatin binding dynamics influence cell induction and differentiation during development. In order to study the aforementioned TF properties of SOX2 and PAX6, I endogenously labelled these TFs in mESCs using the CRISPR-Cas9 system, in collaboration

with Brianda Mohran (Professor David FitzPatrick' lab). Optic cup induction was proven to be challenging even when WT mESCs were used which was a problem across the whole unit. As a result we were not able to state if the endogenous labelling of the two TFs was responsible for the lack of optic cup organoid induction. Nevertheless, an optic cup organoid was able to be induced from mESCs that were endogenously labelled with HALO-PAX6. FCS measurements confirmed the existence of differential chromatin-binding behaviour and high variability in PAX6 cell-to-cell concentrations which is hypothesised to be integral for correct ocular development. Due to the limitations and difficulties in optic cup generation, no definitive conclusions were able to be made. Further work would be required to validate how disruptive the tags are in terms of protein function and a consistent ability to induce optic cup organoids would need to be addressed. If both limitations are addressed, then the variability in TF numbers among different cells during optic cup inductions and differentiation could be investigated. In collaboration with Professor Vladana Vukojevic (Karolinska Institute, Stockholm, Sweden), an expert in FCS, we would be able to use a uniquely developed system, which can perform multi-point FCS measurements. This system uses a matrix of detectors which can simultaneously record 1024 FCS measurements to provide 'heat maps' of molecular diffusion, chromatin-binding dynamics and cell-to-cell variability of the concentrations of fluorescent molecules (Oasa *et al.*, 2021). As a result, this tunable system could potentially be used to investigate the complex spatiotemporal stoichiometry of PAX6 and SOX2 in early optic cup induction and development, providing us with information which currently remains elusive.

Chapter 6 Summary and Conclusions

Extensive genetic and phenotypic studies for both PAX6, SOX2 and their analogous mutants have been thoroughly characterised. Electrophoretic mobility shift assays (EMSA) have shown that both TFs can form tertiary complexes with DNA to regulate expression of key ocular genes such as δ -crystallin during lens placode formation (Kamachi *et al.*, 2001). Furthermore, qualitative methods such as chromatin immunoprecipitation have been able to identify and map the protein-chromatin interactome of both TFs at different developmental stages and link them with target gene regulation (Xie and Cvekl, 2011; Chen *et al.*, 2014). However, the dynamic behaviour of these two TFs at the molecular level *in vivo* (e.g. chromatin binding dynamics, subnuclear distribution, protein-protein interactions) and how human pathogenic missense mutations influence their molecular behaviour remains ill-defined. To investigate this, I exploited the advantages of FCS methodology, which allows for the detection of fluorescently labelled molecules with a single-molecule sensitivity. Due to its high temporal resolution, FCS is suitable for the study of fast dynamical processes such as protein-chromatin and protein-protein interactions.

6.1 Linking the molecular diffusion of PAX6 and SOX2 mutants to pathogenic phenotype in humans

6.1.1 PAX6 mutant variants exhibit weaker chromatin binding dynamics and altered subnuclear localisation

In Chapter 3 I investigated the chromatin-binding dynamics and subnuclear localisation of five PAX6 and four SOX2 pathogenic missense mutants *ex vivo*, and compared them to the WT TFs. Previous phenotypic studies in human patients demonstrated varying degrees of pathogenicity implying functional differences in the importance of each amino acid residue for normal TF function (Table 3-1; Table 3-2).

Human patients carrying the investigated PAX6 variants can have a broad range of phenotypes ranging from iris defects and coloboma to aniridia and

microphthalmia (Williamson *et al.*, 2020). Varying levels of molecular diffusion were obtained for each PAX6 mutant from FCS measurements. In all cases, the variants demonstrated overall faster molecular diffusion, indicative of weaker and shorter-lived chromatin interactions than WT protein. In addition, the subnuclear localisation of all pathogenic PAX6 mutants was different when compared to the WT protein. The WT and R92Q variant (predicted benign) were able to form distinct foci (condensates) which the pathogenic variants did not appear to form. In conclusion, impaired chromatin interaction and altered subnuclear localisation were identified as key aberrations of PAX6 pathogenic variants. It is very likely that these two molecular properties may be the underlying reason of their pathogenic nature in humans.

6.1.2 SOX2 mutant variants exhibit weaker chromatin binding dynamics and altered subnuclear localisation

SOX2 was fluorescently labelled with mCherry at the C-terminus, as previous studies indicated that this retains protein functionality (Bressan *et al.*, 2017). Three of the four investigated pathogenic missense mutants are in the HMG domain and no further molecular experiments had previously been conducted with these variants. FCS and confocal microscopy, showed that the variants in the HMG domain showed no inclination towards chromatin localisation (co-localisation analysis) nor did they interact with chromatin to the same extent as WT or G130A SOX2. These results agree with the phenotypes in human patients, in which the G130A variant exerts the least severe effect.

Overall, missense mutations in both TFs demonstrated altered dynamic behaviour compared to the WT TFs, with their molecular behaviour being analogous to phenotypic severity in patients.

6.2 PAX6 forms biomolecular condensates via a PD dependant mechanism

TF molecules have the ability to form condensates, aiding in nuclear compartmentalisation and potential gene regulation, although this still remains a novel concept (Boija *et al.*, 2018; Cho *et al.*, 2018). Considering the

consistently observed subnuclear localisation differences of WT PAX6 when compared to the pathogenic mutants, in Chapter 4, I wanted to investigate the nature and function of these foci. *PAX6* encodes three main isoforms in which the PD varies in each isoform. The PAX6(δ PD) isoform lacks the PD completely and the PAX6(5a) contains an additional 14 amino acids in the PAI subdomain which alters the binding specificity and affinity of the PD. When the three isoforms were expressed in HEK293, no foci were observed in cells expressing PAX6(δ PD). Considering that the PD alone was also able to form the observed foci, this suggested a PD dependent mechanism. Experiments using dissolution agents such as 1-6 hexanediol and *in vitro* protein aggregation assays suggested that the observed foci were likely condensates. In addition, these condensates were found to colocalise with endogenous HP1a, most likely in pericentric heterochromatin as previously shown for other Pax family members (Bulut-Karslioglu *et al.*, 2012). The association of Pax family members with these heterochromatic regions has been associated with major satellite transcript repression, heterochromatic mark maintenance and chromosome segregation.

6.3 PAX6 missense variants influence homo-complex formation efficiency

Considering that all PAX6 missense mutations are heterozygous, I wanted to investigate the effect that mutant variants may have on PAX6 homo-complex formation. Results obtained through FCCS experiments indicated that the PAX6 mutants can also form homo-complexes with WT PAX6, but not as efficiently. In addition, no WT PAX6 condensates were observed when co-expressed with mutant PAX6 protein, suggesting a possible inhibitory role in WT PAX6 condensate formation.

6.4 PD-HD trans-interaction is altered in the presence of PAX6 mutants

Previous studies have demonstrated the ability of the canonical PAX6 isoform to physically interact with the PAX6(δ PD) leading to transactivation activity

enhancement of PAX6 from PD binding sites (Mikkola *et al.*, 2001). FCCS experiments investigating PD - PAX6(δ PD) trans-interactions suggest that the PD and HD can interact away from and on chromatin. Although PAX6 mutants can retain their ability to interact with PAX6(δ PD) with the same efficiency as the WT protein, this interaction is largely away from chromatin and therefore possibly compromises the transactivation activity enhancement of PAX6 by PAX6(δ PD).

6.5 PAX6 and SOX2 mutations reduce heterocomplex formation efficiency

PAX6 and SOX2 have the ability to cross-regulate each other and form tertiary complexes on target enhancers which are stabilised by protein-protein and protein-DNA interactions (Kamachi *et al.*, 2001). The majority of work investigating the complex interplay between PAX6 and SOX2 has been through qualitative methods such as Co-IP and ChIP-seq (Valouev *et al.*, 2008; Thakurela, Tiwari, Schick, Garding, Ivanek, Berninger and Vijay K. Tiwari, 2016). Despite being very informative, these analysis fail to provide insight into the dynamic behaviour of these TFs in real-time. To address this, I carried out FCCS experiments in which I investigated the interaction of PAX6 and SOX2 at the molecular level and compared results when either TF was mutated. The results suggested that the investigated PAX6 and SOX2 mutants will form roughly half the number of heterocomplexes when expressed in equal amounts to the WT proteins. Considering the importance of TF molecule number, this may have an important influence on gene expression. For example, low-affinity binding sites are less likely to be bound by a TF or TF complex when its concentration is low, providing a further molecular explanation of the pathogenic nature of the investigated mutants.

6.6 SOX2 influences PAX6-chromatin binding kinetics in a non-reciprocal fashion

Co-IP experiments indicated that PAX6-SOX2 interactions are predominantly found on chromatin. Following this observation, I investigated whether this was

driven/stabilised by predominantly one of the two TFs. FCS suggested that SOX2 stabilizes the PAX6-SOX2 heterocomplex on chromatin via a DBD dependent manner. From the obtained results, this is believed to be a generic principle of PAX6-SOX2 complex stabilisation that applies in widespread chromatin locations. In the future, ChIP-seq experiments may be carried out, to validate this suggestion.

The next steps of this project would be to validate the functionality of the endogenously labelled PAX6 and SOX2 mESC line in order to study endogenous TF concentrations, chromatin-binding dynamics and cell-to-cell variability in a developmentally relevant system such as the optic cup organoids. By using mpFCS, we would be able to efficiently investigate the complex spatiotemporal stoichiometry of PAX6 and SOX2 in early optic cup induction and investigate how precise TF concentrations and their chromatin binding dynamics influence cell identity. Furthermore, this system could be used to investigate and developmentally follow how missense variants cause abnormal ocular phenotypes in humans.

Chapter 7 Bibliography

Akhtar, W. *et al.* (2013) 'XChromatin position effects assayed by thousands of reporters integrated in parallel', *Cell*, 154(4), pp. 914–927. doi:10.1016/j.cell.2013.07.018.

Alberti, S. and Dormann, D. (2019) 'Liquid-Liquid Phase Separation in Disease', *Annual Review of Genetics*, 53, pp. 171–194. doi:10.1146/annurev-genet-112618-043527.

Annunziata, O. *et al.* (2002) 'Effect of polyethylene glycol on the liquid-liquid phase transition in aqueous protein solutions', *Proceedings of the National Academy of Sciences of the United States of America*, 99(22), pp. 14165–14170. doi:10.1073/pnas.212507199.

Aota, S.I. *et al.* (2003) 'Pax6 autoregulation mediated by direct interaction of Pax6 protein with the head surface ectoderm-specific enhancer of the mouse Pax6 gene', *Developmental Biology*, 257(1), pp. 1–13. doi:10.1016/S0012-1606(03)00058-7.

Auer, J.M.T. *et al.* (2021) 'Of numbers and movement – Understanding transcription factor pathogenesis by advanced microscopy', *DMM Disease Models and Mechanisms*, 13(12). doi:10.1242/DMM.046516.

Avilion, A.A. *et al.* (2003) 'Multipotent cell lineages in early mouse development depend on SOX2 function', *Genes and Development*, 17(1), pp. 126–140. doi:10.1101/gad.224503.

Babaie, Y. *et al.* (2007) 'Analysis of Oct4-Dependent Transcriptional Networks Regulating Self-Renewal and Pluripotency in Human Embryonic Stem Cells', *Stem Cells*, 25(2), pp. 500–510. doi:10.1634/stemcells.2006-0426.

Bacia, K. and Schwille, P. (2003) 'A dynamic view of cellular processes by in vivo fluorescence auto- and cross-correlation spectroscopy', *Methods*, 29(1), pp. 74–85. doi:10.1016/S1046-2023(02)00291-8.

Bailey, T.J. *et al.* (2004) 'Regulation of development by Rx genes', *International Journal of Developmental Biology*, 48(8–9), pp. 761–770.

doi:10.1387/ijdb.041878tb.

Baltz, A.G. *et al.* (2012) 'The mRNA-Bound Proteome and Its Global Occupancy Profile on Protein-Coding Transcripts', *Molecular Cell*, 46(5), pp. 674–690. doi:10.1016/j.molcel.2012.05.021.

Basu, S. *et al.* (2020) 'Unblending of Transcriptional Condensates in Human Repeat Expansion Disease', *Cell*, 181(5), pp. 1062-1079.e30. doi:10.1016/j.cell.2020.04.018.

Berson, D.M. (2007) 'Phototransduction in ganglion-cell photoreceptors', *Pflugers Archiv European Journal of Physiology*, 454(5), pp. 849–855. doi:10.1007/s00424-007-0242-2.

Bettters, E. *et al.* (2010) 'Analysis of early human neural crest development', *Developmental Biology*, 344(2), pp. 578–592. doi:10.1016/j.ydbio.2010.05.012.

Bhatia, S. *et al.* (2013) 'Disruption of autoregulatory feedback by a mutation in a remote, ultraconserved PAX6 enhancer causes aniridia', *American Journal of Human Genetics*, 93(6), pp. 1126–1134. doi:10.1016/j.ajhg.2013.10.028.

Bhatia, S. *et al.* (2014) 'A survey of ancient conserved non-coding elements in the PAX6 locus reveals a landscape of interdigitated cis-regulatory archipelagos', *Developmental Biology*, 387(2), pp. 214–228. doi:10.1016/j.ydbio.2014.01.007.

Blake, J.A. and Ziman, M.R. (2014) 'Pax genes: Regulators of lineage specification and progenitor cell maintenance', *Development (Cambridge)*, 141(4), pp. 737–751. doi:10.1242/dev.091785.

Boehning, M. *et al.* (2018) 'RNA polymerase II clustering through carboxy-terminal domain phase separation', *Nature Structural and Molecular Biology*, 25(9), pp. 833–840. doi:10.1038/s41594-018-0112-y.

Boese, E.A. *et al.* (2020) 'Novel Intragenic PAX6 Deletion in a Pedigree with Aniridia, Morbid Obesity, and Diabetes', *Current Eye Research*, 45(1), pp. 91–96. doi:10.1080/02713683.2019.1649704.

- Boeynaems, S. *et al.* (2018) 'Protein Phase Separation: A New Phase in Cell Biology'. doi:10.1016/j.tcb.2018.02.004.
- Boija, A. *et al.* (2018) 'Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains', *Cell*, 175(7), pp. 1842–1855.e16. doi:10.1016/j.cell.2018.10.042.
- Boyer, L.A. *et al.* (2005) 'Core transcriptional regulatory circuitry in human embryonic stem cells', *Cell*, 122(6), pp. 947–956. doi:10.1016/j.cell.2005.08.020.
- Bressan, R.B. *et al.* (2017) 'Efficient CRISPR/cas9-assisted gene targeting enables rapid and precise genetic manipulation of mammalian neural stem cells', *Development (Cambridge)*, 144(4), pp. 635–648. doi:10.1242/dev.140855.
- Bringmann, A. *et al.* (2009) 'Cellular signaling and factors involved in Müller cell gliosis: Neuroprotective and detrimental effects', *Progress in Retinal and Eye Research*, 28(6), pp. 423–451. doi:10.1016/j.preteyeres.2009.07.001.
- Brown, A. *et al.* (1998) 'The human PAX6 Mutation Database', *Nucleic Acids Research*, 26(1), pp. 259–264. doi:10.1093/nar/26.1.259.
- Buckle, A. *et al.* (2018) 'Functional characteristics of novel pancreatic Pax6 regulatory elements', *Human Molecular Genetics*, 27(19), pp. 3434–3448. doi:10.1093/hmg/ddy255.
- Bulut-Karslioglu, A. *et al.* (2012) 'A transcription factor-based mechanism for mouse heterochromatin formation', *Nature Structural and Molecular Biology*, 19(10), pp. 1023–1032. doi:10.1038/nsmb.2382.
- Bylund, M. *et al.* (2003) 'Vertebrate neurogenesis is counteracted by Sox1-3 activity', *Nature Neuroscience*, 6(11), pp. 1162–1168. doi:10.1038/nn1131.
- Campen, A. *et al.* (2008) 'TOP-IDP-Scale: A New Amino Acid Scale Measuring Propensity for Intrinsic Disorder', *Protein & Peptide Letters*, 15(9), pp. 956–963. doi:10.2174/092986608785849164.
- Cardozo, M.J., Almuedo-Castillo, M. and Bovolenta, P. (2020) 'Patterning the

Vertebrate Retina with Morphogenetic Signaling Pathways', *Neuroscientist*, 26(2), pp. 185–196. doi:10.1177/1073858419874016.

Chambers, I. *et al.* (2003) 'Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells', *Cell*, 113(5), pp. 643–655. doi:10.1016/S0092-8674(03)00392-1.

Chao, L.Y. *et al.* (2003) 'Missense mutations in the DNA-binding region and termination codon in PAX6', *Human Mutation*, 21(2), pp. 138–145. doi:10.1002/humu.10163.

Chapuy, B. *et al.* (2013) 'Discovery and Characterization of Super-Enhancer-Associated Dependencies in Diffuse Large B Cell Lymphoma', *Cancer Cell*, 24(6), pp. 777–790. doi:10.1016/j.ccr.2013.11.003.

Chauhan, B.K. *et al.* (2004a) 'Functional interactions between alternatively spliced forms of Pax6 in crystallin gene regulation and in haploinsufficiency', *Nucleic Acids Research*, 32(5), pp. 1696–1709. doi:10.1093/nar/gkh334.

Chauhan, B.K. *et al.* (2004b) 'Functional Properties of Natural Human PAX6 and PAX6(5a) Mutants', *Investigative Ophthalmology and Visual Science*, 45(2), pp. 385–392. doi:10.1167/iovs.03-0968.

Chen, J. *et al.* (2014) 'Single-molecule dynamics of enhanceosome assembly in embryonic stem cells', *Cell*, 156(6), pp. 1274–1285. doi:10.1016/j.cell.2014.01.062.

Chen, J. *et al.* (2017) 'Analysis of expression of transcription factors in early human retina', *International Journal of Developmental Neuroscience*, 60(January), pp. 94–102. doi:10.1016/j.ijdevneu.2017.01.015.

Chen, X. *et al.* (2008) 'Integration of External Signaling Pathways with the Core Transcriptional Network in Embryonic Stem Cells', *Cell*, 133(6), pp. 1106–1117. doi:10.1016/j.cell.2008.04.043.

Chi, N. and Epstein, J.A. (2002) 'Getting your Pax straight: Pax proteins in development and disease', *Trends in Genetics*, 18(1), pp. 41–47. doi:10.1016/S0168-9525(01)02594-X.

Cho, W.-K. *et al.* (2018) *MOLECULAR BIOLOGY Mediator and RNA polymerase II clusters associate in transcription-dependent condensates*, *Science*. Available at: <http://science.sciencemag.org/> (Accessed: 7 September 2018).

Chong, S. *et al.* (2018) 'Imaging dynamic and selective low-complexity domain interactions that control gene transcription Downloaded from'. doi:10.1126/science.aar2555.

Chow, R.L. *et al.* (1999) 'Pax6 induces ectopic eyes in a vertebrate', *Development*, 126(19), pp. 4213–4222. doi:10.1242/dev.126.19.4213.

Clements, J. *et al.* (2009) 'Mutational analysis of the eyeless gene and phenotypic rescue reveal that an intact Eyeless protein is necessary for normal eye and brain development in *Drosophila*', *Developmental Biology*, 334(2), pp. 503–512. doi:10.1016/j.ydbio.2009.08.003.

Cunha, D.L. *et al.* (2019) 'The spectrum of PAX6 mutations and genotype-phenotype correlations in the eye', *Genes*, 10(12). doi:10.3390/genes10121050.

Cutter, A.R. and Hayes, J.J. (2015) 'A brief review of nucleosome structure', *FEBS Letters*, 589(20), pp. 2914–2922. doi:10.1016/j.febslet.2015.05.016.

Cvekl, A. and Ashery-Padan, R. (2014) 'The cellular and molecular mechanisms of vertebrate lens development', *Development (Cambridge)*, 141(23), pp. 4432–4447. doi:10.1242/dev.107953.

Dames, P. *et al.* (2010) *Relative roles of the different Pax6 domains for pancreatic alpha cell development*. Available at: <http://www.biomedcentral.com/1471-213X/10/39> (Accessed: 23 April 2020).

Danno, H. *et al.* (2008) 'Molecular links among the causative genes for ocular malformation: Otx2 and Sox2 coregulate Rax expression', *Proceedings of the National Academy of Sciences of the United States of America*, 105(14), pp. 5408–5413. doi:10.1073/pnas.0710954105.

Davis, J. *et al.* (2008) 'Gene expression of the mouse corneal crystallin Aldh3a1: Activation by Pax6, Oct1, and p300', *Investigative Ophthalmology*

and Visual Science, 49(5), pp. 1814–1826. doi:10.1167/iov.07-1057.

DelMonte, D.W. and Kim, T. (2011) 'Anatomy and physiology of the cornea', *Journal of Cataract and Refractive Surgery*, 37(3), pp. 588–598. doi:10.1016/j.jcrs.2010.12.037.

Deluz, C. *et al.* (2017) 'The elusive role of mitotic bookmarking in transcriptional regulation: Insights from Sox2', *Cell Cycle*, 16(7), pp. 601–606. doi:10.1080/15384101.2017.1288332.

Downie, L.E. *et al.* (2021) 'CLEAR - Anatomy and physiology of the anterior eye', *Contact Lens and Anterior Eye*, 44(2), pp. 132–156. doi:10.1016/j.clae.2021.02.009.

Dufourt, J. *et al.* (2018) 'Temporal control of gene expression by the pioneer factor Zelda through transient interactions in hubs', *Nature Communications*, 9(1), pp. 1–13. doi:10.1038/s41467-018-07613-z.

Dzuricky, M. *et al.* (2020) 'De novo engineering of intracellular condensates using artificial disordered proteins', *Nature Chemistry*, 12(9), pp. 814–825. doi:10.1038/s41557-020-0511-7.

Edman, L. *et al.* (1999) 'The fluctuating enzyme: A single molecule approach', *Chemical Physics*, 247(1), pp. 11–22. doi:10.1016/S0301-0104(99)00098-1.

Eghrari, A.O., Riazuddin, S.A. and Gottsch, J.D. (2015) *Overview of the Cornea: Structure, Function, and Development*. 1st edn, *Progress in Molecular Biology and Translational Science*. 1st edn. Elsevier Inc. doi:10.1016/bs.pmbts.2015.04.001.

Eintracht, J., Toms, M. and Moosajee, M. (2020) 'The Use of Induced Pluripotent Stem Cells as a Model for Developmental Eye Disorders', *Frontiers in Cellular Neuroscience*, 14(August), pp. 1–15. doi:10.3389/fncel.2020.00265.

Eiraku, M. *et al.* (2011) 'Self-organizing optic-cup morphogenesis in three-dimensional culture', *Nature*, 472(7341), pp. 51–58. doi:10.1038/nature09941.

Ellis, P. *et al.* (2004) 'SOX2, a persistent marker for multipotential neural stem

cells derived from embryonic stem cells, the embryo or the adult', *Developmental Neuroscience*, 26(2–4), pp. 148–165. doi:10.1159/000082134.

Elson, E.L. and Magde, D. (1974) 'Fluorescence correlation spectroscopy. I. Conceptual basis and theory', *Biopolymers*, 13(1), pp. 1–27. doi:10.1002/bip.1974.360130102.

Elvenes, J. *et al.* (2011) 'Pax6 represses androgen receptor-mediated transactivation by inhibiting recruitment of the coactivator SPBP', *PLoS ONE*, 6(9). doi:10.1371/journal.pone.0024659.

Epstein, J.A. *et al.* (1994) 'Two independent and interactive DNA-binding subdomains of the Pax6 paired domain are regulated by alternative splicing.', *Genes & development*, 8(17), pp. 2022–34. doi:10.1101/GAD.8.17.2022.

Eric Xu, H. *et al.* (1999) 'Crystal structure of the human Pax6 paired domain-DNA complex reveals specific roles for the linker region and carboxy-terminal subdomain in DNA binding', *Genes and Development*, 13(10), pp. 1263–1275. doi:10.1101/gad.13.10.1263.

Estey, T. *et al.* (2007) 'ALDH3A1: a corneal crystallin with diverse functions', *Experimental Eye Research*, 84(1), pp. 3–12. doi:10.1016/j.exer.2006.04.010.

Faivre, L. *et al.* (2006) 'Recurrence of SOX2 anophthalmia syndrome with gonosomal mosaicism in a phenotypically normal mother [2]', *American Journal of Medical Genetics*, 140 A(6), pp. 636–639. doi:10.1002/ajmg.a.31114.

Fantes, J. *et al.* (2003) 'Mutations in SOX2 cause anophthalmia', *Nature Genetics*, 33(4), pp. 461–463. doi:10.1038/ng1120.

Feric, M. *et al.* (2016) 'Coexisting Liquid Phases Underlie Nucleolar Subcompartments', *Cell*, 165(7), pp. 1686–1697. doi:10.1016/j.cell.2016.04.047.

Fischer, A.J. and Reh, T.A. (2001) 'Müller glia are a potential source of neural regeneration in the postnatal chicken retina', *Nature Neuroscience*, 4(3), pp. 247–252. doi:10.1038/85090.

- Franze, K. *et al.* (2007) 'Müller cells are living optical fibers in the vertebrate retina', *Proceedings of the National Academy of Sciences of the United States of America*, 104(20), pp. 8287–8292. doi:10.1073/pnas.0611180104.
- Gao, J. *et al.* (2017) 'Cell contact and pressure control of YAP localization and clustering revealed by super-resolution imaging', *Nanoscale*, 9(43), pp. 16993–17003. doi:10.1039/c7nr05818g.
- Gehring, W.J. (2005) 'New perspectives on eye development and the evolution of eyes and photoreceptors', *Journal of Heredity*, 96(3), pp. 171–184. doi:10.1093/jhered/esi027.
- Georgala, P.A., Carr, C.B. and Price, D.J. (2011) 'The role of Pax6 in forebrain development', *Developmental Neurobiology*, 71(8), pp. 690–709. doi:10.1002/dneu.20895.
- Giri, R. *et al.* (2020) 'Ordered patterning of the sensory system is susceptible to stochastic features of gene expression', *eLife*, 9, pp. 1–34. doi:10.7554/eLife.53638.
- Golan-Lagziel, T. *et al.* (2018) 'Analysis of rat cardiac myocytes and fibroblasts identifies combinatorial enhancer organization and transcription factor families', *Journal of Molecular and Cellular Cardiology*, 116(February), pp. 91–105. doi:10.1016/j.yjmcc.2018.02.003.
- Goldfarb, D.S. *et al.* (2004) 'Importin α : A multipurpose nuclear-transport receptor', *Trends in Cell Biology*, 14(9), pp. 505–514. doi:10.1016/j.tcb.2004.07.016.
- Goldman, D. (2014) 'Müller glial cell reprogramming and retina regeneration', *Nature Reviews Neuroscience*, 15(7), pp. 431–442. doi:10.1038/nrn3723.
- Goolam, M. *et al.* (2016) 'Heterogeneity in Oct4 and Sox2 Targets Biases Cell Fate in 4-Cell Mouse Embryos', *Cell*, 165(1), pp. 61–74. doi:10.1016/j.cell.2016.01.047.
- Graham, V. *et al.* (2003) 'SOX2 functions to maintain neural progenitor identity', *Neuron*, 39(5), pp. 749–765. doi:10.1016/S0896-6273(03)00497-5.

Grapp, M. *et al.* (2009) 'The homeodomain of PAX6 is essential for PAX6-dependent activation of the rat glucagon gene promoter: Evidence for a PH0-like binding that induces an active conformation', *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1789(5), pp. 403–412. doi:10.1016/j.bbagr.2009.02.001.

Grocott, T. *et al.* (2007) 'The MH1 domain of Smad3 interacts with Pax6 and represses autoregulation of the Pax6 P1 promoter', *Nucleic Acids Research*, 35(3), pp. 890–901. doi:10.1093/nar/gkl1105.

Guo, X. *et al.* (2018) 'LincRNA-1614 coordinates Sox2/PRC2-mediated repression of developmental genes in pluripotency maintenance', *Journal of Molecular Cell Biology*, 10(2), pp. 118–129. doi:10.1093/jmcb/mjx041.

Gura Sadovsky, R. *et al.* (2017) 'Measurement of Rapid Protein Diffusion in the Cytoplasm by Photo-Converted Intensity Profile Expansion', *Cell Reports*, 18(11), pp. 2795–2806. doi:10.1016/j.celrep.2017.02.063.

Halder, G., Callaerts, P. and Gehring, W.J. (1995) 'New perspectives on eye evolution', *Current Opinion in Genetics and Development*, 5(5), pp. 602–609. doi:10.1016/0959-437X(95)80029-8.

Halder, G., Callaerts, P. and Gehring, W.J. (1989) *Induction of Ectopic Eyes by Targeted Expression of the eyeless Gene in Drosophila* Downloaded from, *Clin. Exp. Rheumatol.* Available at: <http://science.sciencemag.org/> (Accessed: 25 November 2018).

Han, X. *et al.* (2020) 'Roles of the BRD4 short isoform in phase separation and active gene transcription', *Nature Structural and Molecular Biology*, 27(4), pp. 333–341. doi:10.1038/s41594-020-0394-8.

Hanish, A.E. *et al.* (2016) 'Pineal hypoplasia, reduced melatonin and sleep disturbance in patients with PAX6 haploinsufficiency', *Journal of Sleep Research*, 25(1), pp. 16–22. doi:10.1111/jsr.12345.

Hanson, I.M. *et al.* (1994) 'Peters' anomaly', 6(february).

Harding, P. and Moosajee, M. (2019) *The molecular basis of human anophthalmia and microphthalmia*, *Journal of Developmental Biology*.

doi:10.3390/jdb7030016.

Haubst, N. *et al.* (2004) 'Molecular dissection of Pax6 function: The specific roles of the paired domain and homeodomain in brain development', *Development*, 131(24), pp. 6131–6140. doi:10.1242/dev.01524.

Haustein, E. and Schwille, P. (2007) 'Fluorescence correlation spectroscopy: Novel variations of an established technique', *Annual Review of Biophysics and Biomolecular Structure*, 36, pp. 151–169. doi:10.1146/annurev.biophys.36.040306.132612.

He, C. *et al.* (2016) 'High-Resolution Mapping of RNA-Binding Regions in the Nuclear Proteome of Embryonic Stem Cells', *Molecular Cell*, 64(2), pp. 416–430. doi:10.1016/j.molcel.2016.09.034.

Heo, Y.T. *et al.* (2015) 'CRISPR/Cas9 nuclease-mediated gene knock-in in bovine-induced pluripotent cells', *Stem Cells and Development*, 24(3), pp. 393–402. doi:10.1089/scd.2014.0278.

Hever, A.M., Williamson, K.A. and van Heyningen, V. (2006) 'Developmental malformations of the eye: The role of PAX6, SOX2 and OTX2', *Clinical Genetics*, 69(6), pp. 459–470. doi:10.1111/j.1399-0004.2006.00619.x.

Hingorani, M. *et al.* (2009) 'Detailed ophthalmologic evaluation of 43 individuals with PAX6 mutations', *Investigative Ophthalmology and Visual Science*, 50(6), pp. 2581–2590. doi:10.1167/iovs.08-2827.

Hnisz, D. *et al.* (2013) 'XSuper-enhancers in the control of cell identity and disease', *Cell*, 155(4), p. 934. doi:10.1016/j.cell.2013.09.053.

Hoege, C. *et al.* (2009) 'Dissolution / Condensation', 5(June).

Holmes, Z.E. *et al.* (2020) 'The Sox2 transcription factor binds RNA', *Nature Communications*, 11(1). doi:10.1038/s41467-020-15571-8.

Hoon, M. *et al.* (2014) 'Functional architecture of the retina: Development and disease', *Progress in Retinal and Eye Research*, 42(i), pp. 44–84. doi:10.1016/j.preteyeres.2014.06.003.

Huang, J. *et al.* (2015) 'Negative and positive auto-regulation of BMP

expression in early eye development', *Developmental Biology*, 407(2), pp. 256–264. doi:10.1016/j.ydbio.2015.09.009.

Iwafuchi-Doi, M. *et al.* (2011) 'The Pou5f1/Pou3f-dependent but SoxB-independent regulation of conserved enhancer N2 initiates Sox2 expression during epiblast to neural plate stages in vertebrates', *Developmental Biology*, 352(2), pp. 354–366. doi:10.1016/j.ydbio.2010.12.027.

Izeddin, I. *et al.* (2014) 'Single-molecule tracking in live cells reveals distinct target-search strategies of transcription factors in the nucleus', *eLife*, 2014(3), pp. 1–27. doi:10.7554/eLife.02230.

Jagga, B. *et al.* (2021) 'Structural basis for nuclear import selectivity of pioneer transcription factor SOX2', *Nature Communications*, 12(1), pp. 1–11. doi:10.1038/s41467-020-20194-0.

Jeong, C.H. *et al.* (2010) 'Phosphorylation of Sox2 cooperates in reprogramming to pluripotent stem cells', *Stem Cells*, 28(12), pp. 2141–2150. doi:10.1002/stem.540.

Jidigam, V.K. *et al.* (2015) 'Apical constriction and epithelial invagination are regulated by BMP activity', *Biology Open*, 4(12), pp. 1782–1791. doi:10.1242/bio.015263.

Jones, L. *et al.* (2002) 'Pax6 is required for the normal development of the forebrain axonal connections', *Development*, 129(21), pp. 5041–5052. doi:10.1242/dev.129.21.5041.

Justilien, V. *et al.* (2014) 'The PRKCI and SOX2 Oncogenes Are Coamplified and Cooperate to Activate Hedgehog Signaling in Lung Squamous Cell Carcinoma', *Cancer Cell*, 25(2), pp. 139–151. doi:10.1016/j.ccr.2014.01.008.

Kamachi, Y. *et al.* (1998) 'Kamachi Y, 1998.pdf', *Development*, 125(13), pp. 2521–2532.

Kamachi, Y. *et al.* (2001) 'Pax6 and SOX2 form a co-DNA-binding partner complex that regulates initiation of lens development', *Genes and Development*, 15(10), pp. 1272–1286. doi:10.1101/gad.887101.

- Kamachi, Y. and Kondoh, H. (2013) 'Sox proteins: Regulators of cell fate specification and differentiation', *Development (Cambridge)*, 140(20), pp. 4129–4144. doi:10.1242/dev.091793.
- Kammandel, B. *et al.* (1999) 'Distinct cis-essential modules direct the time-space pattern of the Pax6 gene activity', *Developmental Biology*, 205(1), pp. 79–97. doi:10.1006/dbio.1998.9128.
- Kanaan, N.M. *et al.* (2020) 'Liquid-liquid phase separation induces pathogenic tau conformations in vitro', *Nature Communications*, 11(1). doi:10.1038/s41467-020-16580-3.
- Kang, J. *et al.* (2011) 'Mitotic centromeric targeting of HP1 and its binding to Sgo1 are dispensable for sister-chromatid cohesion in human cells', *Molecular Biology of the Cell*, 22(8), pp. 1181–1190. doi:10.1091/mbc.E11-01-0009.
- Kelberman, D. *et al.* (2006) 'With Abnormalities in the Hypothalamo- Pituitary- Gonadal Axis in Mice and Humans', *The Journal of Clinical Investigation*, 116(9), pp. 2442–2455. doi:10.1172/JCI28658.2442.
- Khoury, G. and Gruss, P. (1983) 'Enhancer elements', *Cell*, 33(2), pp. 313–314. doi:10.1016/0092-8674(83)90410-5.
- Kikkawa, T. *et al.* (2019) 'The role of Pax6 in brain development and its impact on pathogenesis of autism spectrum disorder', *Brain Research*, 1705, pp. 95–103. doi:10.1016/j.brainres.2018.02.041.
- Kim, J. and Lauderdale, J.D. (2008) 'Overexpression of pairedless Pax6 in the retina disrupts corneal development and affects lens cell survival', *Developmental Biology*, 313(1), pp. 434–454. doi:10.1016/j.ydbio.2007.10.043.
- Kim, S. *et al.* (2016) 'Complexation and coacervation of like-charged polyelectrolytes inspired by mussels', *Proceedings of the National Academy of Sciences of the United States of America*, 113(7), pp. E847–E853. doi:10.1073/pnas.1521521113.
- Kleinjan, D.A. *et al.* (2006) 'Long-range downstream enhancers are essential for Pax6 expression', *Developmental Biology*, 299(2), pp. 563–581.

doi:10.1016/j.ydbio.2006.08.060.

Klosin, A. *et al.* (2020) 'Phase separation provides a mechanism to reduce noise in cells', *Science*, 367(6476), pp. 464–468. doi:10.1126/science.aav6691.

Kondoh, H. and Takemoto, T. (2012) 'Axial stem cells deriving both posterior neural and mesodermal tissues during gastrulation', *Current Opinion in Genetics and Development*, 22(4), pp. 374–380. doi:10.1016/j.gde.2012.03.006.

Kopp, J.L. *et al.* (2008) 'Small Increases in the Level of Sox2 Trigger the Differentiation of Mouse Embryonic Stem Cells', *Stem Cells*, 26(4), pp. 903–911. doi:10.1634/stemcells.2007-0951.

Kreft, D. *et al.* (2018) 'Binding mechanism of anti-cancer chemotherapeutic drug mitoxantrone to DNA characterized by magnetic tweezers', *Journal of Nanobiotechnology*, 16(1), pp. 1–7. doi:10.1186/s12951-018-0381-y.

Lakowski, J., Majumder, A. and Lauderdale, J.D. (2007) 'Mechanisms controlling Pax6 isoform expression in the retina have been conserved between teleosts and mammals', *Developmental Biology*, 307(2), pp. 498–520. doi:10.1016/J.YDBIO.2007.04.015.

Lan, H.C. *et al.* (2021) 'Ocular disease-associated mutations diminish the mitotic chromosome retention ability of PAX6', *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1864(11–12), p. 194751. doi:10.1016/j.bbagr.2021.194751.

Landsend, E.C.S., Lagali, N. and Utheim, T.P. (2021) 'Congenital aniridia – A comprehensive review of clinical features and therapeutic approaches', *Survey of Ophthalmology* [Preprint]. doi:10.1016/j.survophthal.2021.02.011.

Larson, A.G. *et al.* (2017) 'Liquid droplet formation by HP1 α suggests a role for phase separation in heterochromatin', *Nature*, 547(7662), pp. 236–240. doi:10.1038/nature22822.

Levine, M. (2010) 'Transcriptional enhancers in animal development and evolution', *Current Biology*, 20(17), pp. R754–R763.

doi:10.1016/j.cub.2010.06.070.

Li, B. *et al.* (2001) 'Human sex reversal due to impaired nuclear localization of SRY: A clinical correlation', *Journal of Biological Chemistry*, 276(49), pp. 46480–46484. doi:10.1074/jbc.C100388200.

Li, J. *et al.* (2020) 'Single-gene imaging links genome topology, promoter–enhancer communication and transcription control', *Nature Structural and Molecular Biology*, 27(11), pp. 1032–1040. doi:10.1038/s41594-020-0493-6.

Li, Y. *et al.* (2014) 'CRISPR reveals a distal super-enhancer required for Sox2 expression in mouse embryonic stem cells', *PLoS ONE*, 9(12), pp. 1–17. doi:10.1371/journal.pone.0114485.

Lin, Y.P. *et al.* (2009) 'Sox2 plays a role in the induction of amacrine and müller glial cells in mouse retinal progenitor cells', *Investigative Ophthalmology and Visual Science*, 50(1), pp. 68–74. doi:10.1167/iovs.07-1619.

Liu, X. *et al.* (2020) 'Mitotic Implantation of the Transcription Factor Prospero via Phase Separation Drives Terminal Neuronal Differentiation', *Developmental Cell*, 52(3), pp. 277-293.e8. doi:10.1016/j.devcel.2019.11.019.

Liu, Z. *et al.* (2014) '3D imaging of Sox2 enhancer clusters in embryonic stem cells', *eLife*, 3, p. e04236. doi:10.7554/eLife.04236.

Lopez-noriega, L. *et al.* (2020) 'The long non-coding RNA Pax6os1 / PAX6-AS1 modulates pancreatic β -cell identity and function . * Equal contributions
† Correspondence to Guy . A Rutter , g.rutter@imperial.ac.uk and Timothy J . Pullen , Key words : Long non-coding RNAs , Pax6os1 / PAX6-A'.

Machon, O. *et al.* (2010) 'Lens morphogenesis is dependent on Pax6-mediated inhibition of the canonical Wnt/Beta-catenin signaling in the lens surface ectoderm', *Genesis*, 48(2), pp. 86–95. doi:10.1002/dvg.20583.

MacQuarrie, K.L. *et al.* (2011) 'Genome-wide transcription factor binding: Beyond direct target regulation', *Trends in Genetics*, 27(4), pp. 141–148. doi:10.1016/j.tig.2011.01.001.

Maekawa, M. *et al.* (2009) 'A novel missense mutation (Leu46Val) of PAX6

found in an autistic patient', *Neuroscience Letters*, 462(3), pp. 267–271. doi:10.1016/j.neulet.2009.07.021.

Magnani, L., Eeckhoute, J. and Lupien, M. (2011) 'Pioneer factors: Directing transcriptional regulators within the chromatin environment', *Trends in Genetics*, 27(11), pp. 465–474. doi:10.1016/j.tig.2011.07.002.

Mallam, A.L. *et al.* (2019) 'Systematic Discovery of Endogenous Human Ribonucleoprotein Complexes', *Cell Reports*, 29(5), pp. 1351–1368.e5. doi:10.1016/j.celrep.2019.09.060.

Manuel, M.N. *et al.* (2015) 'Regulation of cerebral cortical neurogenesis by the Pax6 transcription factor', *Frontiers in Cellular Neuroscience*, 9(March), pp. 1–21. doi:10.3389/fncel.2015.00070.

Manzer, R. *et al.* (2003) 'Ultraviolet radiation decreases expression and induces aggregation of corneal ALDH3A1', *Chemico-Biological Interactions*, 143–144, pp. 45–53. doi:10.1016/S0009-2797(02)00171-0.

Matsushima, D. *et al.* (2011) 'Combinatorial regulation of optic cup progenitor cell fate by SOX2 and PAX6.', *Development (Cambridge, England)*, 138(3), pp. 443–54. doi:10.1242/dev.055178.

Mazza, D. *et al.* (2012) 'A benchmark for chromatin binding measurements in live cells', *Nucleic Acids Research*, 40(15), pp. 1–13. doi:10.1093/nar/gks701.

Merino, F., Bouvier, B. and Cojocaru, V. (2015) 'Cooperative DNA Recognition Modulated by an Interplay between Protein-Protein Interactions and DNA-Mediated Allostery', *PLoS Computational Biology*, 11(6), pp. 1–23. doi:10.1371/journal.pcbi.1004287.

Messemaker, T.C. *et al.* (2018) 'Allele-specific repression of Sox2 through the long non-coding RNA Sox2ot', *Scientific Reports*, 8(1), pp. 1–13. doi:10.1038/s41598-017-18649-4.

Mikkola, I. *et al.* (2001) 'Superactivation of Pax6-mediated Transactivation from Paired Domain-binding Sites by DNA-independent Recruitment of Different Homeodomain Proteins', *Journal of Biological Chemistry*, 276(6), pp. 4109–4118. doi:10.1074/jbc.M008882200.

- Mishra, R., Gorlov, Ivan P, *et al.* (2002) 'PAX6, paired domain influences sequence recognition by the homeodomain.', *The Journal of biological chemistry*, 277(51), pp. 49488–94. doi:10.1074/jbc.M206478200.
- Mishra, R., Gorlov, Ivan P., *et al.* (2002) 'PAX6, paired domain influences sequence recognition by the homeodomain', *Journal of Biological Chemistry*, 277(51), pp. 49488–49494. doi:10.1074/jbc.M206478200.
- Mitchell, T.N. *et al.* (2003) 'Polymicrogyria and absence of pineal gland due to PAX6 mutation', *Annals of Neurology*, 53(5), pp. 658–663. doi:10.1002/ana.10576.
- Mivelaz, M. *et al.* (2020) 'Chromatin Fiber Invasion and Nucleosome Displacement by the Rap1 Transcription Factor', *Molecular Cell*, 77(3), pp. 488-500.e9. doi:10.1016/j.molcel.2019.10.025.
- Morrill, S.A. and Amon, A. (2019) 'Why haploinsufficiency persists', *Proceedings of the National Academy of Sciences of the United States of America*, 116(24), pp. 11866–11871. doi:10.1073/pnas.1900437116.
- Mullari, M. *et al.* (2017) 'Specifying RNA-Binding Regions in Proteins by Peptide Cross-Linking and Affinity Purification', *Journal of Proteome Research*, 16(8), pp. 2762–2772. doi:10.1021/acs.jproteome.7b00042.
- Mullegama, S. *et al.* (2017) 'RAI1 Overexpression Promotes Altered Circadian Gene Expression and Dyssomnia in Potocki–Lupski Syndrome', *Journal of Pediatric Genetics*, 06(03), pp. 155–164. doi:10.1055/s-0037-1599147.
- Nair, S.J. *et al.* (2019) 'Phase separation of ligand-activated enhancers licenses cooperative chromosomal enhancer assembly', *Nature Structural and Molecular Biology*, 26(3), pp. 193–203. doi:10.1038/s41594-019-0190-5.
- Nees, D.W. *et al.* (2002) 'Structurally Normal Corneas in Aldehyde Dehydrogenase 3a1-Deficient Mice', *Molecular and Cellular Biology*, 22(3), pp. 849–855. doi:10.1128/mcb.22.3.849-855.2002.
- Nelson, D.E. *et al.* (2004) 'Oscillations in NF- κ B signaling control the dynamics of gene expression', *Science*, 306(5696), pp. 704–708. doi:10.1126/science.1099962.

- Ng, S.Y. *et al.* (2013) 'The long noncoding RNA RMST interacts with SOX2 to regulate neurogenesis', *Molecular Cell*, 51(3), pp. 349–359. doi:10.1016/j.molcel.2013.07.017.
- Niaki, A.G. *et al.* (2020) 'Loss of Dynamic RNA Interaction and Aberrant Phase Separation Induced by Two Distinct Types of ALS/FTD-Linked FUS Mutations', *Molecular Cell*, 77(1), pp. 82–94.e4. doi:10.1016/j.molcel.2019.09.022.
- Nickla, D.L. and Wallman, J. (2010) 'The multifunctional choroid', *Progress in Retinal and Eye Research*, 29(2), pp. 144–168. doi:10.1016/j.preteyeres.2009.12.002.
- Ninkovic, J. *et al.* (2013) 'The BAF complex interacts with Pax6 in adult neural progenitors to establish a neurogenic cross-regulatory transcriptional network', *Cell Stem Cell*, 13(4), pp. 403–418. doi:10.1016/j.stem.2013.07.002.
- Nishimura, N. *et al.* (2012) 'A systematic survey and characterization of enhancers that regulate Sox3 in neuro-sensory development in comparison with Sox2 enhancers', *Biology*, 1(3), pp. 714–735. doi:10.3390/biology1030714.
- Njajou, O.T. *et al.* (2001) 'PAX6 haploinsufficiency causes', 28(july), pp. 3–5.
- Nojima, T. *et al.* (2018) 'RNA Polymerase II Phosphorylated on CTD Serine 5 Interacts with the Spliceosome during Co-transcriptional Splicing', *Molecular Cell*, 72(2), pp. 369–379.e4. doi:10.1016/j.molcel.2018.09.004.
- Normanno, D. *et al.* (2015) 'Probing the target search of DNA-binding proteins in mammalian cells using TetR as model searcher', *Nature Communications*, 6. doi:10.1038/ncomms8357.
- Oasa, S. *et al.* (2021) 'Dynamic Cellular Cartography: Mapping the Local Determinants of Oligodendrocyte Transcription Factor 2 (OLIG2) Function in Live Cells Using Massively Parallel Fluorescence Correlation Spectroscopy Integrated with Fluorescence Lifetime Imaging Microscopy ('), *Analytical Chemistry*, 93(35), pp. 12011–12021. doi:10.1021/acs.analchem.1c02144.
- Ouyang, J. *et al.* (2015) 'Cyclin-dependent kinase-mediated Sox2

phosphorylation enhances the ability of Sox2 to establish the pluripotent state', *Journal of Biological Chemistry*, 290(37), pp. 22782–22794. doi:10.1074/jbc.M115.658195.

Pak, C.W. *et al.* (2016) 'Sequence Determinants of Intracellular Phase Separation by Complex Coacervation of a Disordered Protein', *Molecular Cell*, 63(1), pp. 72–85. doi:10.1016/j.molcel.2016.05.042.

Palozola, K.C., Lerner, J. and Zaret, K.S. (2019) 'A changing paradigm of transcriptional memory propagation through mitosis', *Nature Reviews Molecular Cell Biology*, 20(1), pp. 55–64. doi:10.1038/s41580-018-0077-z.

Papadopoulos, D.K. *et al.* (2015) 'Probing the kinetic landscape of Hox transcription factor-DNA binding in live cells by massively parallel Fluorescence Correlation Spectroscopy', *Mechanisms of Development*, 138, pp. 218–225. doi:10.1016/j.mod.2015.09.004.

Papadopoulos, D.K. *et al.* (2019) 'Control of hox transcription factor concentration and cell-to-cell variability by an auto-regulatory switch', *Development (Cambridge)*, 146(12). doi:10.1242/dev.168179.

Pappa, A. *et al.* (2003) 'ALDH3A1 protects human corneal epithelial cells from ultraviolet- and 4-hydroxy-2-nonenal-induced oxidative damage', *Free Radical Biology and Medicine*, 34(9), pp. 1178–1189. doi:10.1016/S0891-5849(03)00070-4.

Phair, R.D. and Misteli, T. (2000) '35007077', 404(April).

Pinson, J. *et al.* (2005) 'Regulation of the Pax6 : Pax6(5a) mRNA ratio in the developing mammalian brain'. doi:10.1186/1471-213X-5-13.

Plachta, N. *et al.* (2011) 'Oct4 kinetics predict cell lineage patterning in the early mammalian embryo', *Nature Cell Biology*, 13(2), pp. 117–123. doi:10.1038/ncb2154.

Placzek, M. and Briscoe, J. (2018) 'Sonic hedgehog in vertebrate neural tube development', *International Journal of Developmental Biology*, 62(1–3), pp. 221–230. doi:10.1387/ijdb.170293jb.

Plaisancié, J. *et al.* (2019) *Genetics of anophthalmia and microphthalmia. Part 1: Non-syndromic anophthalmia/microphthalmia, Human Genetics*. Springer Berlin Heidelberg. doi:10.1007/s00439-019-01977-y.

Price, J. *et al.* (1997) 'Pax6-dependent regulation of adhesive patterning, R-cadherin expression and boundary formation in developing forebrain', *Development*, 124(19), pp. 3765–77. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9367432.

Qin, L. *et al.* (2016) 'Detection and Quantification of Mosaic Mutations in Disease Genes by Next-Generation Sequencing', *Journal of Molecular Diagnostics*, 18(3), pp. 446–453. doi:10.1016/j.jmoldx.2016.01.002.

Quinn, J.C., West, J.D. and Hill, R.E. (1996) 'Multiple functions for Pax6 in mouse eye and nasal development', *Genes and Development*, 10(4), pp. 435–446. doi:10.1101/gad.10.4.435.

Ragge, N.K. *et al.* (2005) 'SOX2 anophthalmia syndrome', *American Journal of Medical Genetics*, 135 A(1), pp. 1–7. doi:10.1002/ajmg.a.30642.

Riback, J.A. *et al.* (2020) 'Composition-dependent thermodynamics of intracellular phase separation', *Nature*, 581(7807), pp. 209–214. doi:10.1038/s41586-020-2256-2.

Richards, M.G., Pope, J. and Widom, A. (1972) 'Evidence for isotopic impurities in solid helium', *Physical Review Letters*, 29(11), pp. 708–711. doi:10.1103/PhysRevLett.29.708.

Richardson, R. *et al.* (2016) 'Clinical utility gene card for: Aniridia', *European Journal of Human Genetics*, 24(11), pp. 1–4. doi:10.1038/ejhg.2016.73.

Rigler, R. *et al.* (1993) 'Fluorescence correlation spectroscopy with high count rate and low background: analysis of translational diffusion', *European Biophysics Journal*, 22(3), pp. 169–175. doi:10.1007/BF00185777.

Rizzino, A. (2013) 'Interdependent Network Integrated At Multiple Levels', *Stem Cells*, 31(6), pp. 1033–1039. doi:10.1002/stem.1352.The.

- Rodda, D.J. *et al.* (2005) 'Transcriptional regulation of Nanog by OCT4 and SOX2', *Journal of Biological Chemistry*, 280(26), pp. 24731–24737. doi:10.1074/jbc.M502573200.
- Rodriguez, R.T. *et al.* (2007) 'Manipulation of OCT4 levels in human embryonic stem cells results in induction of differential cell types', *Experimental Biology and Medicine*, 232(10), pp. 1368–1380. doi:10.3181/0703-RM-63.
- Rosales, T. *et al.* (2013) 'Partition of Myc into immobile vs. Mobile complexes within nuclei', *Scientific Reports*, 3. doi:10.1038/srep01953.
- Roux, L.N. *et al.* (2018) 'Modeling of Aniridia-Related Keratopathy by CRISPR/Cas9 Genome Editing of Human Limbal Epithelial Cells and Rescue by Recombinant PAX6 Protein', *STEM CELLS*, 36(9), pp. 1421–1429. doi:10.1002/stem.2858.
- Sabari, B.R. *et al.* (no date) 'Coactivator condensation at super-enhancers links phase separation and gene control'. doi:10.1126/science.aar3958.
- Samudiyata *et al.* (2019) 'Interaction of Sox2 with RNA binding proteins in mouse embryonic stem cells', *Experimental Cell Research*, 381(1), pp. 129–138. doi:10.1016/j.yexcr.2019.05.006.
- Schedl, A. *et al.* (1996) 'Influence of PAX6 gene dosage on development: Overexpression causes severe eye abnormalities', *Cell*, 86(1), pp. 71–82. doi:10.1016/S0092-8674(00)80078-1.
- Schwarz, M. *et al.* (2000) 'Spatial specification of mammalian eye territories by reciprocal transcriptional repression of Pax2 and Pax6', *Development*, 127(20), pp. 4325–4334. doi:10.1242/dev.127.20.4325.
- Senecal, A. *et al.* (2014) 'Transcription factors modulate c-Fos transcriptional bursts', *Cell Reports*, 8(1), pp. 75–83. doi:10.1016/j.celrep.2014.05.053.
- Shi, X. *et al.* (2009) 'Probing events with single molecule sensitivity in zebrafish and Drosophila embryos by fluorescence correlation spectroscopy', *Developmental Dynamics*, 238(12), pp. 3156–3167. doi:10.1002/dvdy.22140.
- Simicevic, J. *et al.* (2013) 'Absolute quantification of transcription factors during

cellular differentiation using multiplexed targeted proteomics', *Nature Methods*, 10(6), pp. 570–576. doi:10.1038/nmeth.2441.

Singh, S. *et al.* (2002) 'Iris hypoplasia in mice that lack the alternatively spliced Pax6(5a) isoform', *Proceedings of the National Academy of Sciences of the United States of America*, 99(10), pp. 6812–6815. doi:10.1073/pnas.102691299.

Skouloudaki, K. *et al.* (2019) 'Yorkie controls tube length and apical barrier integrity during airway development', *Journal of Cell Biology*, 218(8), pp. 2762–2781. doi:10.1083/JCB.201809121.

Smith, A.N. *et al.* (2009) 'Erratum: Stage-dependent modes of Pax6-Sox2 epistasis regulate lens development and eye morphogenesis (Development (2009) vol. 136 (2977-2985))', *Development*, 136(19), p. 3377. doi:10.1242/dev.043802.

Soufi, A., Donahue, G. and Zaret, K.S. (2012) 'Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome', *Cell*, 151(5), pp. 994–1004. doi:10.1016/j.cell.2012.09.045.

Stoykova, A. *et al.* (1996) 'Forebrain patterning defects in Small eye mutant mice', *Development*, 122(11), pp. 3453–3465. doi:10.1242/dev.122.11.3453.

Strauss, O. (2005) 'The retinal pigment epithelium in visual function', *Physiological Reviews*, 85(3), pp. 845–881. doi:10.1152/physrev.00021.2004.

Strebinger, D. *et al.* (2019) 'Endogenous fluctuations of OCT 4 and SOX 2 bias pluripotent cell fate decisions', *Molecular Systems Biology*, 15(9), pp. 1–19. doi:10.15252/msb.20199002.

Strom, A.R. *et al.* (2017) 'Phase separation drives heterochromatin domain formation', *Nature*, 547(7662), pp. 241–245. doi:10.1038/nature22989.

Sumita, Y. *et al.* (2018) 'Cytoplasmic expression of SOX9 as a poor prognostic factor for oral squamous cell carcinoma', *Oncology Reports*, 40(5), pp. 2487–2496. doi:10.3892/or.2018.6665.

Sun, J. *et al.* (2015) 'Identification of in vivo DNA-binding mechanisms of Pax6

and reconstruction of Pax6-dependent gene regulatory networks during forebrain and lens development', *Nucleic Acids Research*, 43(14), pp. 6827–6846. doi:10.1093/nar/gkv589.

Sun, J. *et al.* (2016) 'Pax6 associates with H3K4-specific histone methyltransferases Mll1, Mll2, and Set1a and regulates H3K4 methylation at promoters and enhancers', *Epigenetics and Chromatin*, 9(1), pp. 1–18. doi:10.1186/s13072-016-0087-z.

Surzenko, N. *et al.* (2013) 'SOX2 maintains the quiescent progenitor cell state of postnatal retinal Müller glia', *Development (Cambridge)*, 140(7), pp. 1445–1456. doi:10.1242/dev.071878.

Taranova, O. V. *et al.* (2006) 'SOX2 is a dose-dependent regulator of retinal neural progenitor competence', *Genes and Development*, 20(9), pp. 1187–1202. doi:10.1101/gad.1407906.

Taratuta, V.G. *et al.* (1990) 'Liquid-liquid phase separation of aqueous lysozyme solutions: Effects of pH and salt identity', *Journal of Physical Chemistry*, 94(5), pp. 2140–2144. doi:10.1021/j100368a074.

Tarilonte, M. *et al.* (2018) 'Parental mosaicism in PAX6 causes intra-familial variability: Implications for genetic counseling of congenital aniridia and microphthalmia', *Frontiers in Genetics*, 9(OCT). doi:10.3389/fgene.2018.00479.

Thakurela, S., Tiwari, N., Schick, S., Garding, A., Ivanek, R., Berninger, B. and Tiwari, Vijay K (2016) 'Mapping gene regulatory circuitry of Pax6 during neurogenesis', *Cell Discovery*, 2(1), p. 15045. doi:10.1038/celldisc.2015.45.

Thakurela, S., Tiwari, N., Schick, S., Garding, A., Ivanek, R., Berninger, B. and Tiwari, Vijay K. (2016) 'Mapping gene regulatory circuitry of Pax6 during neurogenesis', *Cell Discovery*, 2. doi:10.1038/celldisc.2015.45.

Thivierge, C. *et al.* (2006) 'Overexpression of PKD1 Causes Polycystic Kidney Disease', *Molecular and Cellular Biology*, 26(4), pp. 1538–1548. doi:10.1128/mcb.26.4.1538-1548.2006.

Thomson, M. *et al.* (2011) 'Pluripotency factors in embryonic stem cells

regulate differentiation into germ layers', *Cell*, 145(6), pp. 875–889. doi:10.1016/j.cell.2011.05.017.

Tsai, A. *et al.* (2017) 'Nuclear microenvironments modulate transcription from low-affinity enhancers', *eLife*, 6, pp. 1–18. doi:10.7554/eLife.28975.

Tsuruzoe, S. *et al.* (2006) 'Inhibition of DNA binding of Sox2 by the SUMO conjugation', *Biochemical and Biophysical Research Communications*, 351(4), pp. 920–926. doi:10.1016/j.bbrc.2006.10.130.

Tzoulaki, I., White, I.M.S. and Hanson, I.M. (2005) 'PAX6 mutations: Genotype-phenotype correlations', *BMC Genetics*, 6(Mim 106210), pp. 1–12. doi:10.1186/1471-2156-6-27.

Valouev, A. *et al.* (2008) 'Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data', *Nature Methods*, 5(9), pp. 829–834. doi:10.1038/nmeth.1246.

Vance, K.W. *et al.* (2014) 'The long non-coding RNA paupar regulates the expression of both local and distal genes', *EMBO Journal*, 33(4), pp. 296–311. doi:10.1002/emboj.201386225.

Veneri, P. *et al.* (2020) 'Dynamical reorganization of the pluripotency transcription factors Oct4 and Sox2 during early differentiation of embryonic stem cells', *Scientific Reports*, 10(1), pp. 1–12. doi:10.1038/s41598-020-62235-0.

Vukojević, V. *et al.* (2005) 'Study of molecular events in cells by fluorescence correlation spectroscopy', *Cellular and Molecular Life Sciences*, 62(5), pp. 535–550. doi:10.1007/s00018-004-4305-7.

Vukojević, V. *et al.* (2010) 'Quantitative study of synthetic Hox transcription factor-DNA interactions in live cells', *Proceedings of the National Academy of Sciences of the United States of America*, 107(9), pp. 4093–4098. doi:10.1073/pnas.0914612107.

Walcher, T. *et al.* (2013) 'Functional dissection of the paired domain of Pax6 reveals molecular mechanisms of coordinating neurogenesis and proliferation', *Development*, 140(5), pp. 1123–1136. doi:10.1242/dev.082875.

- Walther, C. *et al.* (1991) 'Pax: A murine multigene family of paired box-containing genes', *Genomics*, 11(2), pp. 424–434. doi:10.1016/0888-7543(91)90151-4.
- Wang, J. *et al.* (2018) 'A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins', *Cell*, 174(3), pp. 688–699.e16. doi:10.1016/j.cell.2018.06.006.
- Wang, Z. *et al.* (2012) 'Distinct lineage specification roles for NANOG, OCT4, and SOX2 in human embryonic stem cells', *Cell Stem Cell*, 10(4), pp. 440–454. doi:10.1016/j.stem.2012.02.016.
- Wang, Z. *et al.* (2019) 'Wnt Signaling in vascular eye diseases', *Progress in Retinal and Eye Research*, 70(July 2018), pp. 110–133. doi:10.1016/j.preteyeres.2018.11.008.
- Wennmalm, S., Edman, L. and Rigler, R. (1997) 'Conformational fluctuations in single DNA molecules', *Proceedings of the National Academy of Sciences of the United States of America*, 94(20), pp. 10641–10646. doi:10.1073/pnas.94.20.10641.
- Wheeler, R.J. *et al.* (2019) 'Small molecules for modulating protein driven liquid-liquid phase separation in treating neurodegenerative disease', *bioRxiv*, pp. 1–48. doi:10.1101/721001.
- White, M.D. *et al.* (2016) 'Long-Lived Binding of Sox2 to DNA Predicts Cell Fate in the Four-Cell Mouse Embryo', *Cell*, 165(1), pp. 75–87. doi:10.1016/j.cell.2016.02.032.
- Widengren, J., Rigler, R. and Mets, Ü. (1994) 'Triplet-state monitoring by fluorescence correlation spectroscopy', *Journal of Fluorescence*, 4(3), pp. 255–258. doi:10.1007/BF01878460.
- Williams, C.A.C., Soufi, A. and Pollard, S.M. (2020) 'Post-translational modification of SOX family proteins: Key biochemical targets in cancer?', *Seminars in Cancer Biology*, 67(August 2019), pp. 30–38. doi:10.1016/j.semcancer.2019.09.009.
- Williams, S.C. *et al.* (1998) 'A highly conserved lens transcriptional control

element from the Pax-6 gene', *Mechanisms of Development*, 73(2), pp. 225–229. doi:10.1016/S0925-4773(98)00057-4.

Williamson, K.A. *et al.* (2006) 'Mutations in SOX2 cause anophthalmia-esophageal-genital (AEG) syndrome', *Human Molecular Genetics*, 15(9), pp. 1413–1422. doi:10.1093/hmg/ddl064.

Williamson, K.A. *et al.* (2020) 'Recurrent heterozygous PAX6 missense variants cause severe bilateral microphthalmia via predictable effects on DNA–protein interaction', *Genetics in Medicine*, 22(3), pp. 598–609. doi:10.1038/s41436-019-0685-9.

Wilson, S.E. (2020) 'Bowman's layer in the cornea– structure and function and regeneration', *Experimental Eye Research*, 195(March), p. 108033. doi:10.1016/j.exer.2020.108033.

Wuebben, E.L. *et al.* (2012) 'Musashi2 is required for the self-renewal and pluripotency of embryonic stem cells', *PLoS ONE*, 7(4). doi:10.1371/journal.pone.0034827.

Xiao, Y. *et al.* (2019) 'A Novel PAX6 Heterozygous Mutation Found in a Chinese Family with Congenital Aniridia', *Genetic Testing and Molecular Biomarkers*, 23(7), pp. 495–500. doi:10.1089/gtmb.2018.0334.

Xie, Q. *et al.* (2013) 'Pax6 Interactions with Chromatin and Identification of Its Novel Direct Target Genes in Lens and Forebrain', *PLoS ONE*, 8(1). doi:10.1371/journal.pone.0054507.

Xie, Q. *et al.* (2014) 'Gene regulation by PAX6: Structural-functional correlations of missense mutants and transcriptional control of Trpm3/miR-204', *Molecular Vision*, 20(November 2013), pp. 270–282.

Xie, Q. and Cvekl, A. (2011) 'The orchestration of mammalian tissue morphogenesis through a series of coherent feed-forward loops', *Journal of Biological Chemistry*, 286(50), pp. 43259–43271. doi:10.1074/jbc.M111.264580.

Xu, P.X. *et al.* (1999) 'Regulation of Pax6 expression is conserved between mice and flies', *Development*, 126(2), pp. 383–395.

doi:10.1242/dev.126.2.383.

Xu, Y. *et al.* (2021) 'PAUPAR and PAX6 sequentially regulate human embryonic stem cell cortical differentiation', *Nucleic acids research*, 49(4), pp. 1935–1950. doi:10.1093/nar/gkab030.

Yang, H. *et al.* (2003) 'Protein conformational dynamics probed by single-molecule electron transfer', *Science*, 302(5643), pp. 262–266. doi:10.1126/science.1086911.

Yang, Y. *et al.* (2004) 'Transcriptional regulation of mouse α B- and γ F-crystallin genes in lens: Opposite promoter-specific interactions between Pax6 and large Maf transcription factors', *Journal of Molecular Biology*, 344(2), pp. 351–368. doi:10.1016/j.jmb.2004.07.102.

Yeo, G.C., Keeley, F.W. and Weiss, A.S. (2011) 'Coacervation of tropoelastin', *Advances in Colloid and Interface Science*, 167(1–2), pp. 94–103. doi:10.1016/j.cis.2010.10.003.

Yi, Q. *et al.* (2018) 'HP 1 links centromeric heterochromatin to centromere cohesion in mammals', *EMBO reports*, 19(4), pp. 1–13. doi:10.15252/embr.201745484.

Yonova-Doing, E. *et al.* (2020) 'Common variants in SOX-2 and congenital cataract genes contribute to age-related nuclear cataract', *Communications Biology*, 3(1). doi:10.1038/s42003-020-01421-2.

Zaccarini, R. *et al.* (2007) 'PAX6 P46 binds chromosomes in the pericentromeric region and induces a mitosis defect when overexpressed', *Investigative Ophthalmology and Visual Science*, 48(12), pp. 5408–5419. doi:10.1167/iovs.07-0413.

Zhao, H. yong *et al.* (2011) 'CARM1 mediates modulation of Sox2', *PLoS ONE*, 6(10). doi:10.1371/journal.pone.0027026.

Zhao, S. *et al.* (2004) 'SoxB transcription factors specify neuroectodermal lineage choice in ES cells', *Molecular and Cellular Neuroscience*, 27(3), pp. 332–342. doi:10.1016/j.mcn.2004.08.002.

