Summer 2016

Selfish Mutations: the Genetic Basis of the Paternal Age Effect

Eoin C. Whelan
Old Dominion University, ewhel001@odu.edu

SELFISH MUTATIONS:

THE GENETIC BASIS OF THE PATERNAL AGE EFFECT

by

Eoin C. Whelan
B.Sc. August 2003, University College London, United Kingdom

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfilment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

BIOMEDICAL SCIENCES

OLD DOMINION UNIVERSITY
August 2016

Approved by:

Christopher Osgood (Director)

Michael Stacey (Member)

Stephan Olariu (Member)

ABSTRACT

SELFISH MUTATIONS:
THE GENETIC BASIS OF THE PATERNAL AGE EFFECT

Eoin C. Whelan
Old Dominion University, 2016
Director: Dr. Christopher Osgood

As the mean age of childrearing grows, the effect of parental age on genetic disease and child health becomes ever more important. A number of autosomal dominant disorders show a dramatic paternal age effect due to selfish mutations: substitutions that grant spermatogonial stem cells (SSCs) a selective advantage in the testes of the father but have a deleterious effect in offspring.

I present a mathematical model to analyse the normal function of the stem cell compartment, which provides a framework for SSC renewal and accommodates differences between animal systems. In order to model the SSC mutation accumulation, a Markov chain was used to model the probabilities of mutation and positive selection with cell divisions. This model provided average numbers of mutant sperm produced with increasing paternal age. The proportions of mutant to wildtype cells with increasing paternal age was used to generate a simulated population and observed/expected curves. These were then fitted against existing disease and sequencing data. The parameter for the probability of positive selection per division of a mutant cell was estimated. Incidence of the diseases was predicted closely for most disorders and was influenced by the site-specific mutation rate caused by hypermutable CpG sites and the number of mutable alleles. The incidence of disease was explained satisfactorily only when a combination of positive selection and the site-specific mutation rate were included in the analysis.

To provide experimental evidence for the hyposthesis that paternal age effect mutations present a selective advantage, I selected the mutation in the RET (*RE*arranged during *T*ransfection) gene that causes multiple endocrine neoplasia type 2B. SSCs were created by inducing differentiation to spermatogonia of induced pluripotent stem cells. Wildtype and mutant SSCs were generated by transfection with a plasmid containing the normal RET gene and the gene containing the disease mutation, respectively. Mutant SSCs showed increased proliferation in culture. This effect was counteracted when the mutant receptors were saturated with their ligand, GDNF (glial-derived neurotropic factor).

This research demonstrated theoretical and experimental evidence for positive selection in SSCs for multiple endocrine neoplasia type 2B and other paternal age effect syndromes.

This thesis is dedicated to my family, without whose
love and support none of this would have been possible.

ACKNOWLEDGMENTS

Many people have guided me and contributed to the completion of this dissertation. I extend many thanks to my major advisor, Chris Osgood and my committee members (past and present) for their guidance on my research and editing this manuscript. Dr Osgood has provided support for my project and guidance throughout. Thanks to Mike Stacey for practical advice and the use of his lab space and thanks to Ellen Jing for help in the core lab. Stephan Olariu provided invaluable contributions to the mathematical modelling of stem cell niches. I very much appreciate the contributions of Alex Nwala and Samiur Arif to the computational areas of my work and Holly Gaff for introducing me to coding and to Dr Olariu's group. Thanks to Patrick Sachs and Pete Mollica for kindly sharing their iPSCs and help throughout the differentiation process. Thanks to Loreé Heller for her patience and guidance creating plasmids and R. James Swanson for providing invaluable experience with mouse handling and physiology. Special thanks to Anthony Asmar for help and discussion throughout my research.

ABBREVIATIONS

Cell Types

MEF	Mouse Embryonic Fibroblast
iPSCs	induced Pluripotent Stem Cell
SC	Stem Cell
SSC	Spermatogonial Stem Cells
STO	SIM (Sandos Inbred Mice) Thioguanine/Ouabain-resistant mouse fibroblast cell line, used as a feeder layer.
TA cells	Transit Amplifying cells

Diseases

ACH	Achondroplasia
FOP	Fibrodysplasia Ossificans Progressiva
MEN2B	Multiple Endocrine Neoplasia Type 2B
TD	Thanatophoric Dysplasia

Genes and Genetic Elements

CpG	Dinucleotide sequence of cytosine followed by guanine
DDX4	DEAD-box helicase 4, a marker for spermatogonia
FGFR	Fibroblast Growth Factor Receptor, the membrane receptor for FGF
FGF	Fibroblast Growth Factor, a growth factor
GDNF	Glial Derived Neurotropic Factor, a growth factor
PLZF	Promyelocytic leukaemia zinc finger protein, a marker for spermatogonia
RET	REarranged during Transfection, the membrane receptor for GDNF

Model Variables

α	Male-to-female mutation rate ratio
$\alpha_1/\alpha_2/\alpha_3$	Probability of a stem cell undergoing symmetric division (2 SCs) / asymmetric division / symmetric division (2 TAs)
$\beta_1/\beta_2/\beta_3$	Probability of a TA cell undergoing symmetric division (2 TAs) / asymmetric division / symmetric division (2 differentiated cells)
μ	Mutation rate, per nucleotide per generation
n	Niche size (cells)
p	Mutation probability, per nucleotide, per cell division
q	Probability no mutation occurs, per nucleotide, per cell division ($1 - p$)
r	Probability of positive selection (i.e. symmetric renewing division)
p_i	Probability of the niche gaining one net mutant cell after a cell division
q_i	Probability that the niche gains no net mutant cells after a cell division ($1 - p_i$)
M	Average number of mutant cells per niche
S	Simulated number of mutant births
O/E	Observed mutant births divided by the Expected mutant births
v_0/v_1	The rate at which SC divide (v_0) / TA cells divide (v_1)
$\gamma_0/\gamma_1/\gamma_2$	The apoptosis rate of stem cells (γ_0) / TA cells (γ_1) / differentiated cells (γ_2)
$X_0/X_1/X_2$	The size of the SC/TA/differentiated cell compartments

TABLE OF CONTENTS

Page

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION TO THE PATERNAL AGE EFFECT

1.1 Paternal Age

As the average age of parenthood becomes more delayed, understanding the effect of parental age on child health becomes more important. (Bray *et al.* 2006). The effect of maternal age has long been acknowledged (Hook 1981) but in recent years the effect of paternal age has been the subject of a great deal of study. An association between incidence of genetic disease and increasing paternal age was first noted by Weinberg (1912) while studying risk factors for achondroplasia. Advanced paternal age is now associated with a number of mutations that cause dominant disorders and X-linked diseases (Vogel 1975, Risch *et al*. 1987, Glaser & Jabs 2004). It has also been linked to degeneration of other polygenic traits in human offspring; reduction in longevity (Gavrilov 1997), decrease of telomere length (Unryn *et al.* 2005), reduced IQ (Malaspina *et al.* 2005), increased sporadic incidence of polygenic diseases such as schizophrenia (Sipos *et al.* 2004, Malaspina *et al.* 2002), autism (Reichenberg *et al.* 2006) and multiple sclerosis (Montgomery *et al.* 2004). Congenital defects, cancer predisposition disorders, bipolar disorder, and Alzheimer's disease have also been linked to father's age (reviewed in Paul and Robaire 2013).

1.2 Copy-error

Due to the larger number of male germline cell divisions compared with the female germline, males produce 3-6 times as many mutations than females throughout evolution (Li *et al.* 1996).

Sperm are produced by a continually-dividing population of stem cells and each division represents a chance for replication errors to happen. Haldane (1935, 1947) noted a male-bias in the incidence of haemophilia A and suggested that this could be accounted for by fidelity errors in replication. Penrose (1955) proposed that replication errors also provided an explanation for the observed increased incidence of mutation with advanced paternal age. In many cases the influence of paternal age is relatively subtle compared with the large scale chromosomal abnormalities characteristic of maternal age effect because point mutations typically have small or no effect on phenotype. However certain substitutions can have devastating effect on those who carry the allele.

1.3 Positive Selection

Diseases that show a strong paternal age effect, however, are not explained purely by Penrose's copy-error hypothesis. These diseases show an exponential increase in incidence with father's age. The mutations responsible typically display a very specific spectrum of mutations, often caused by missense substitutions at very specific sites. Mutations accumulate faster than the raw mutation rate can account for. Mutations in these disorders present in clumps, indicating a positive selective mechanism of mutation accumulation as opposed to a high mutation rate or "hot spot" model. Such evidence has so far presented for achondroplasia (Shinde *et al.* 2013), Apert's syndrome (Qin *et al.* 2007, Choi *et al.* 2008), Costello syndrome (Ginnoulatou *et al.* 2013) and Noonan Syndrome (Yoon *et al.* 2013). There is a parallel with the intestinal crypt where mutant cells colonise their niche through selective advantage conferred by their new phenotype (Bozic & Nowak 2013). It is also relevant to cancer etiology as paternal age effect mutations are typically found in tumors (Maher *et al.* 2014).

1.4 Outline

In this thesis I present multiple avenues of research into this phenomenon. First I introduce a mathematical model of the stem cell niche. Chapter 3 is a mathematical and computational analysis of the paternal age effect and premeiotic selection, comparing the model to existing disease and mutation data. Finally I present *in vitro* analysis of one particular mutation, the M918T mutation in the *RET* gene that produces the disease Multiple Endocrine Neoplasia type 2B and show that spermatogonial stem cells carrying this mutation do indeed show a selective advantage that would explain the exponential paternal age effect.

CHAPTER 2

PATERNAL AGE AND THE MUTATION RATE

2.1 Paternal age and cell divisions.

In contrast to the human female germ cell line, which undergoes 22 divisions by birth (after which oogenesis is complete and the oocytes persist until the final two meiotic divisions [Vogel & Rathenberg 1975]), the gonocytes of the male germ cell line continue to divide after birth. By puberty, the stem cells have divided ~30 times and continue to divide at an estimated rate of one division every 16 days (de Rooij & Russel 2000, de Rooij & van Beek 2013) although this is likely to decrease as the male ages (Vogel & Motulsky 1997). With no reduction in cell division rate with age, a man's germ cells will have divided 600 times by age 40 (Vogel & Rathernberg 1975). This disparity of cell division number has a dramatic effect on the number and types of mutation between the sexes and accumulate with age.

2.2 Types of mutation

Particular types of mutation are associated with paternal transmission, primarily point mutations and other mitotic errors, due to the much greater number of cell divisions in the male germ line (Miyata *et al.* 1987). According to this principle, continued cell division in the male germ-cell line translate into a build up of replication-dependent mutations, thus the risk of disease caused by these errors rises with age (Glaser & Jabs 2004). On the other hand, mutations associated with maternal transmission, such as chromosomal aneuploidies or deletions, are not expected to be associated with paternal age (Glaser & Jabs 2004). For any given syndrome one would expect the strength of the paternal age effect to be governed by the type of mutation: those

that are maternally-inherited or show no parent-of-origin effect may dilute or mask the cumulative excess of replication-dependent mutations from the male germ line with age.

2.2.1 Point mutations

The majority of human syndromes that show the strongest paternal age effect are autosomal dominant disorders characterised by point mutations (see Fig. 1 and Glaser & Jabs 2004). In two disorders nearly all instances of disease are caused by one transition at a single position within the gene (achondroplasia [Bellus *et al.* 1995, Wilkin *et al.* 1998] and fibrodysplasia ossificans progressive [Shore *et al.* 2006]), while others can be caused of a variety of point mutations at different sites (*e.g.* Crouzon and Pfeiffer syndromes [Passos-Bueno *et al.* 1999]). Approximately half of human paternal age effect syndromes that are known to be caused exclusively by point mutations occur in the context of CpG dinucleotides (*i.e.* cytosine and guanine nucleotides present adjacent within the genome, see *Table 1*).

There are syndromes that do not appear to fit this pattern, however. Alexander disease would seem to be a prime candidate for a paternal age effect: almost all mutations are simple missense substitutions which display a pronounced paternal origin of mutation. Yet curiously, while the average age for the parents of the proband (i.e. affected individual) is higher than the average for the control population, this was not significant ($P>0.5$, *T*-test, Li *et al.* 2006).

The paternal-age effect is also apparent in X-linked diseases. A novel mutation may appear to occur in the mother's germ line, or the mutation may have originated in a previous generation. Assuming most mutations occur in males, the most likely event is that the mutation arose on an X chromosome produced by the maternal grandfather (due to the greater number of cell divisions in his embryonic stem cell line compared with the germ lines of the grandmother or

mother) and that the proband's mother would be a carrier. Becker *et al.* (1996) found a significant grandpaternal age effect apparent in individuals with haemophilia A, when only those cases that originated from *de novo* point mutations were considered. When all types of mutation were analysed together (including deletions acquired from the mother), there was no significant effect of grandpaternal age, consistent with the theory that maternally-derived mutations conceal the paternal age effect. A grandpaternal age effect has also been suggested in regard to Lesch-Nyhan disease (Franke 1976).

Syndrome	CpG	CpG transition/ transversion	Reference
Achondroplasia	Yes	Transition	Bellus *et al.* 1995, Wilkin *et al.* 1998
Apert	Yes	Transversion	
Costello	Yes	Transition	
Crouzon	No		
Fibrodysplasia ossificans progressiva	Yes	Transition	Shore *et al.* 2006
Muenke-type craniosynostosis	Yes	Transversion	
Multiple endocrine neoplasia 2A	No		
Multiple endocrine neoplasia 2B	No		
Noonan	No		
Pfeiffer	No		
Progeria	Yes	Transition	Eriksson *et al.* 2003, Cao & Hegele 2003
Thanatophoric dysplasia	Yes	Transition	Tavormina *et al.* 1995

Table 1 - Syndromes with paternal age effect caused by point mutations.
For those syndromes (n=7) where the majority of point mutations are at CpG dinucleotides, the type of substitution (transition/ transversion) is noted. Data from (Online Mendelian Inheritance in Man, 2016) except where noted.

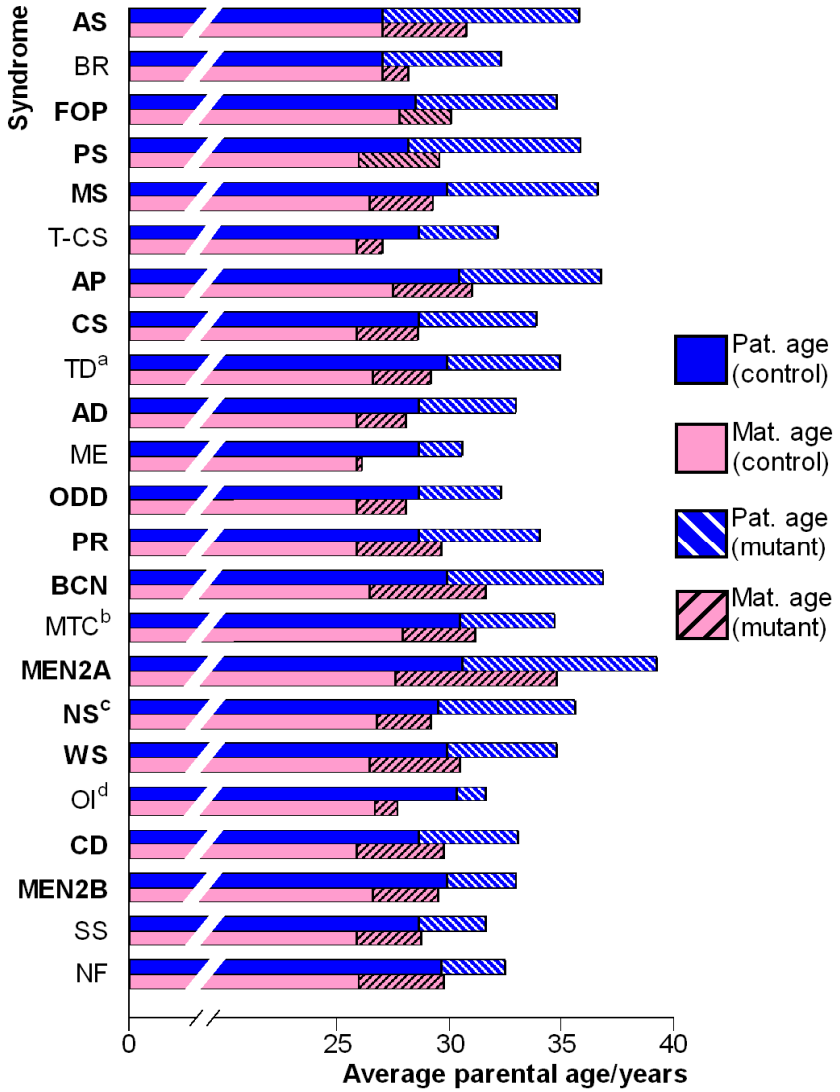


Fig. 1 - Paternal and maternal ages for human syndromes demonstrating a paternal age effect. Average age for the parents of an affected child is shown (mutant) with the average maternal and paternal ages of the control population census (control). Diseases shown in bold have been shown to have an exponential paternal age effects (Risch *et al.* 1987, Glaser & Jabs 2004) and have been ordered by decending ratio of male:female parental age. Data adapted from Glaser & Jabs (2004), with additions from: (a) Lurie 1994, control data from the 1990 census of Japan, Department of Statistics, Ministry of Internal Affairs and Communications, Japan (note that due to the variety of patients in Lurie 1994, control data is intended for comparative basis only) (b) Orioli *et al* 1995 (combined IPIMC and ECLAMC data), (c) Rannan-Eliya *et al* 2004, (d) Tartaglia *et al* (2004) (e) Orioli *et al* 1995 (combined IPIMC and ECLAMC data). (c) and (e) did not attain formal statistical significance in the samples cited, but showed strong paternal age effect trends (p=0.06 for both Muenke-type craniosynostosis and the ECLAMC cohort of osteogenesis imperfecta, Rannan-Eliya *et al* 2004, Orioli et al 1995).

BR=bilateral retinoblastoma; T-CS=Treacher Collins syndrome; AP=achondroplasia; AD=acrodsostosis; ME=multiple exostoses; ODD=oculodentodigital dysplasia; BCN=basal cell nevus; MTC=Muenke-type craniosynostosis; MEN2A=multiple endocrine neoplasia type 2A; WS=Waardenburg syndrome; OI=osteogenesis imperfecta; CD=cleidocranial dysostosis; NF=neurofibromatosis.

2.2.2 Insertions/deletions

Deletions are not associated with parent-of-origin effects since these occur most commonly during meiosis (Glaser & Jabs 2004). While some of the paternal age effect syndromes in Fig. 1 are associated with a range of mutational types (including insertions, deletions and rearrangements as well as point mutations), but these are for the most part, syndromes designated by Risch *et al.* (1987) to be only weakly associated with paternal age. Thus a weak association indicates that a variety of mutational processes are at work and that age-independent insertions/deletions are mitigating or masking the paternal age effect caused by replicative mutations.

2.2.3 Tandem repeats

Tandem repeat mutations are postulated to be associated primarily with replication slippage (Ellegran 2000) and as such can be expected to show a paternal age effect. Huntingdon's disease, caused by an expanded $(CAG)_n$ repeat, shows a pronounced paternal age effect in the probability of repeat expansion, and children of older fathers with the premutation allele are more likely to inherit increased number of triplet repeats with a greater change in repeat number. Huntingdon's disease shows a paternal age effect regardless of which parent passes on the mutant allele, indicating that genomic imprinting is involved in repeat expansion (Evans-Galea *et al.* 2013, Farrer *et al.* 1992). Similarly, fragile X syndrome (an X-linked disorder associated with permutation transmission of $[CGG]_n$ triplet repeats in the *FMR1* gene) demonstrates a correlation of increasing paternal age with a greater magnitude in repeat number change: older fathers are more likely to pass on a mutated allele involving a large expansion than younger fathers (Ashley-Koch *et al.* 1998). Human microsatellites, in general, demonstrates a

paternal age effect (Nikitina & Nazarenkon 2004), wherein a male bias of 17:3 is seen in the frequency of mutational events (Brinkmann *et al.* 1998).

2.2.4 Chromosomal abnormalities:

Historically, an association of advanced male age with chromosomal aneuploides has been postulated (Rives *et al.* 2002, Sartorelli *et al.* 2001), but more recent data suggests that there is little or no association between older paternal age and chromosomal aneuploidies (Fonseka & Griffin, 2011, Jung *et al.* 2003), with the exception of disomic sex-chromosome incidence (Kühnert & Nieschlag 2004) such as the frequency of XY sperm causing Klinefelter's syndrome (Lowe *et al*. 2001).

Column A			**Column B**			**Column C**	
Autosomal dominant disease with strong paternal age effect	**α**	**n**	**X-linked disease**	**α**	**n**	**Higher primate comparative studies**	**α**
Achondroplasia	∞	40	Haemophilia A[c]	9.4	44	Huang et al. 1997	5.14
Apert	∞	57	Haemophilia B[d]	8.6	42	Shimmin et al. 1993	6.26
Crouzon & Pfeiffer	∞	22	Pelizaeus-Merzbacher	4	5	Chang et al. 1996	4.2
MEN2A	∞	10	Rett	13.5	29	Makova & Li 2002	5.25
MEN2B	∞	25				Ebersberger 2002	3
Muenke[a]	∞	10	**Other autosomal diseases**			Anagnostopoulos 1999	3.5
Noonan[b]	∞	14	NF2	1.3	23	Bohossian et al 2000	1.7
			Von Hippel-Lindau disease	1.3	7	Erlandsoon et al 2000	2.5
			FAP[e]	3.0	16	Agulnik et al 1997	4
			Alexander disease[f]	6	28		
			Townes-Brocks[g]	7	16		
		Total			Total		
		178			210		
Weighted mean	∞			7.4			4.0

Table 2 - Human male/female mutation ratio estimates.
Data from Li *et al.* 2002 unless noted otherwise. (a) Rannan-Eliya *et al.* 2004. (b) Tartaglia *et al*. 2004 (PTPN11 cohort only). (c) Becker *et al.* 1996 (α value listed is the mean of those derived from direct and indirect estimates). (d) Green *et al.* 1999. (e) Shore *et al.* 2006. (f) Li *et al.* 2006. (g) Böhm *et al.* 2006.

2.3 Male-to-female mutation rate

The ratio of male-to-female mutation rates (α) has been estimated in humans through two methods. Evolutionary biologists have analyzed molecular sequence variation between closely-related species and measured sequence divergence between sex chromosomes and autosomes. While autosomes are equally divided between the sexes, at any given time two thirds of the X-chromosomes will be found in females and Y-chromosomes are only found in males. The substitution rate (μ) can be inferred from sequence divergence of homologous chromosomes between taxa. Assuming the female mutation rate to be proportional to 1, then the mutation rate of the Y-chromosome will be α, and the X-chromosome mutation rate will be (α+2)/3, as these are found with a frequency of 2/3 in females. Thus the sex differential in mutation rate within taxa can be calculated from the ratio of substitution rates between the sex chromosomes: $\mu_Y/\mu_X=3\alpha/(2+\alpha)$ (Miyata *et al.* 1987). The second method has been used by medical researchers looking at parental origin of *de novo* autosomal dominant and X-linked recessive disorders ($\alpha = \mu_m/\mu_f$). A summary of data derived by both methods is shown in *Table 2*. Column A shows those autosomal dominant disorders associated with an exponential increase in paternal age for which parent-of-origin data is known. This subset of the data shows an exclusively paternal origin of mutation. For those disorders that are not associated with a strong paternal age effect, column B, the male:female ratio average does not differ significantly from that calculated via sequence analysis for primates over evolutionary time, column C ($p>0.10$ paired *T*-test). Both results correspond to the average difference in germ line division between the sexes. Assuming a generous male-to-female mutation ratio (α=15), the probability that of all the mutations listed in Column A (n=178) not one originated in the maternal germ line can be calculated as p = (α-1/α)^n = (14/15)^178 = 0.000005. Clearly there is an underlying mechanism that has skewed the

parent-of-origin effects of those disorders characterised by a strong paternal age effect far beyond the normal male-to-female ratio.

2.4 Patterns within the data

While mutations associated with the paternal age effect broadly fall within the classical spectrum of replication errors (Glaser & Jabs 2004) there are a number of anomalies that are not explained by the simple copy-error hypothesis. (a) The male-to-female mutation rate in these syndromes, particularly those strongly associated with paternal age, appears to be highly skewed toward paternal origin of disease alleles (see *Table 2*), far more than would be expected from the difference in germ cell divisions. (b) There are a number of genetic diseases which are predominantly caused by point mutations that do not show any paternal age effect – for example, Alexander's disease (OMIM: 203450, Li *et al.* 2006) and Townes-Brocks syndrome (OMIM: 107480, Böhm *et al.* 2006). (c) Over half of the strongly age-associated syndromes are caused by mutations at CpG dinucleotides, which are not commonly understood to be replication mediated (*Table 1*). (d) Risch *et al.* (1987) found an exponential increase in incidence of disease that is positively associated with paternal age for the majority of disorders (those designated as having a strong effect in Fig. 1). This casts doubt on the copy-error hypothesis, since, if single-gene mutations were accruing solely due to cell division one would expect a linear increase with age (Risch *et al.* 1987). The exponential effect may instead be explained by differential methylation, degradation of repair enzymes, replication fidelity senescence or positive selection.

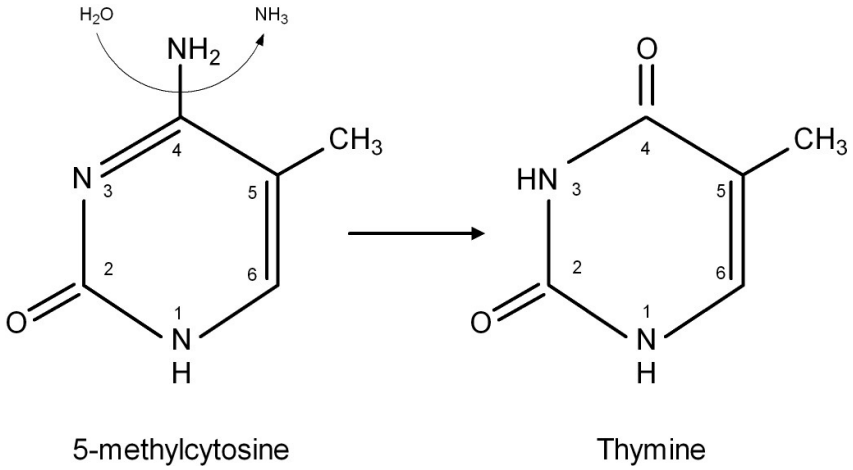


Fig. 2 - Spontaneous deamination of 5-methylcytosine to thymine.
Hydrophilic attack at C-4 of 5-methylcytosine substitutes a hydroxyl group for the amino group, which via tautomeric shift becomes a double-bonded oxygen and forms thymine (Walsh & Xu, 2006, Cooper & Krawczak 1993, Vogel & Motulsky 1997)

2.5 Methylation effects

CG couplet sequences form mutational 'hotspots' in mammals. Cytosines within CpG dinucleotide sites are usually subject to methylation and the increased substitution rate is due to the fact that 5-methylcytosine can spontaneously deaminate to form thymine (see Fig. 2), which is commonly thought to be replication-independent. Transitions at CpG dinucleotides are reported to be the most common type of point mutation in humans (Antonarakis 2000), constituting approximately 23% of human coding gene mutations (Krawczak *et al.* 1998, Krawczak & Cooper 1997). As a consequence, CpG sites are rare throughout mammalian genomes in comparison to the expected number assuming a random distribution. (Subramanian & Kumar 2003). While cytosine is prone to deamination to uracil, this change is quickly and efficiently corrected by repair enzymes such as uracil DNA glycosylase since uracil is not

normally found in DNA (Walsh & Xu 2006). Deamination of 5-methylcytosine is much harder for mismatch repair enzymes to recognise since thymine is a normal component of DNA. When this occurs, it leaves a T=G mispair. If not corrected by mismatch repair enzymes, this mistake will be incorporated as a point mutation in one of the daughter strands when the cell divides (Fig. 3).

Transitions at CpG sites are the cause of many human diseases, including several dominant disorders with pronounced paternal age effect and paternal origin of mutation (see *Table 1*). This is surprising because if transitions at CpG dinucleotides are purely time-dependent, there should be no difference in mutation rate between the sexes.

There is some evidence that deamination is enhanced by replication. The mismatch-repair enzymes which function to remove the thymine from the T=G mispair – such as methyl-CpG-binding proteins or thymine DNA glycosylase – are much less efficient than the corresponding uracil-excision enzymes (Pfeifer 2006, Lindhal 1993). One can hypothesise that if deamination of 5-methylcytosine occurs randomly, one would expect more fixation of mutation in dividing cells: if the DNA replication process occurs before the mismatch-repair process, the mutation will be fixed in one of the daughter cells (see Fig. 3). In non-dividing cells, however, the mismatch repair enzymes have almost unlimited time to repair the mismatch. Methylated CpG sites in non-dividing *E. coli* cells have been described as stable and not mutational hotspots, consistent with this hypothesis (Lieb and Rehmat 1997).

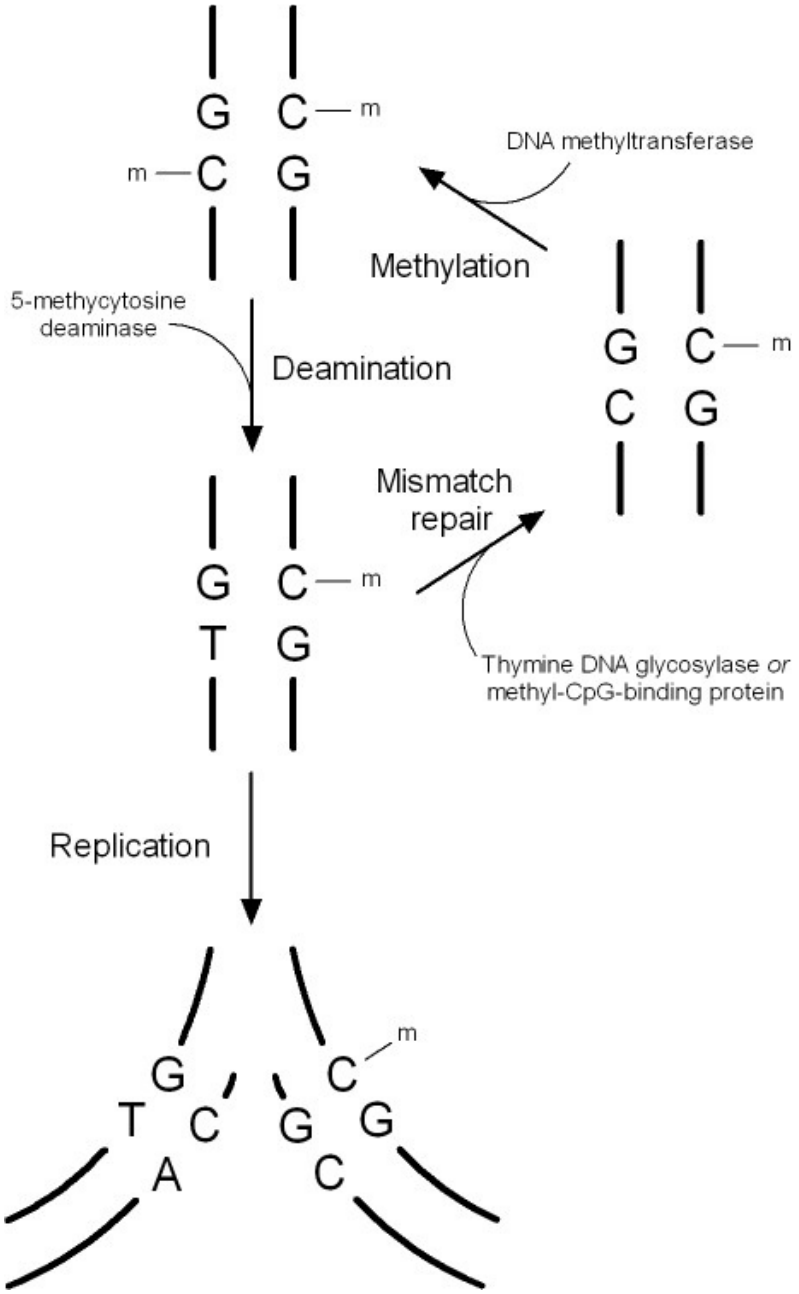


Fig. 3 - Deamination mutation and repair.
(a) Methylated CpG dinucleotide. Spontaneous deamination (Vogel & Motulsky 1997) or action via deaminase enzymatic activity (Morgan *et al.* 2004) leads to a T=G mismatch (b). Mismatch repair enzymes such as thymine DNA glycosylase (Walsh & Xu 2006) or methyl-CpG-binding proteins (Brero *et al.* 2006) excise the out-of-sequence thymine and replace with an unmethylated cytosine (c). Methylation is restored via DNA methyltransferase (a) (Bird 1999). If repair of the T=G mismatch (b) does not occur before the next DNA replication event, the thymine base will be incorporated in one of the daughter strands as a new point mutation (d) (Lindahl 1993).

CpG dinucleotides may be particularly vulnerable to the deamination of 5-methyl-cytosine during the *de novo* methylation process itself. DNA methyltransferase binds to the C6 position of the cytosine ring, forming a dihydropyrimidine intermediate that is prone to deamination (Lindahl 1993), and dividing cells will have to remethylate half their genomes with each replication. Replication might also enhance mutation as the DNA is in a single-stranded form for longer, which is more vulnerable to deamination (Impellizzeri 1991).

DNA deaminases are expressed in germ cells, possibly for epigenetic reprogramming, and could pose risk as a mutagen if mistargetted (Morgan *et al.* 2004), although these may normally be active only in non-dividing stages of the cell cycle, as a preventative measure. Furthermore, it has been suggested that deaminases are catalysts of demethylation. Deaminases work by causing a transitional mutation of the 5-methylcytosine, which, after repair of the mismatch, leaves the cytosine unmethylated (see Fig. 3). Failure to repair such deamination would be a source of mutation, particularly in times of extensive demethylation, such as the paternal genome undergoes, shortly after fertilisation. In this case it would be most likely to cause mosaicism of mutation as the zygote consists of multiple cells.

The relative male to female mutation rates (α) can be estimated by relative comparison of chromosome mutation rates derived from sequence homology alignments of closely-related species. Anagnostopoulos *et al.* (1999) found many more CpG transitions on the Y-chromosome relative to the X-chromosome. However, Nachman & Crowell (2000) and Smith & Hurst (1999) found no significant difference in the number of transitions between autosomes and the X-chromosome, which would be predicted if transitions were replication-dependent. Taylor *et al.* (2005) found that in human-chimp comparisons, CpGs sites demonstrate weak male mutation bias, far lower than would be expected if transitions were purely replication-dependent. This is

consistent with the molecular data indicating that transitions at CpG dinucleotides should not be considered totally replication-independent, but are chiefly random events that occur only moderately more in mammalian males than females, unlike other point mutations.

Differential methylation of maternal and paternal chromosomes would also contribute to the skew in mutational origins. As female imprinting patterns are established in the oocyte prior to fertilisation and male imprints are established at birth and persist throughout life, differential methylation of gonocytes could contribute to the paternal age effect (Schaefer *et al.* 2007). However, the paternal germ line has more hypomethylated regions than somatic tissue does (Biermann & Steger 2007), and paternally imprinted genes are usually controlled by other regulatory elements such as noncoding RNAs. This dimorphism may be due to selective pressures to reduce deamination mutations within the male germ line (Bourc'his & Bestor 2006). Although data is scant for most syndromes, El-Maarri *et al.* (1998) showed equal methylation in both maternal and paternal germ lines for the *Factor VIII* and *FGFR3* genes, indicating that discrepancy in methylation is not a factor in the incidence of diseases associated with these genes.

2.6 Germ-line replication fidelity with age

The paternal age effect has been observed in mice, which demonstrate a linear increase in mutation frequency with age (Walter *et al.* 2004, 1998, Ono *et al.* 2000). Only the mutation rate within the germ-line is of relevance to disease in the progeny, and the mutation rate in the germ-line is consistently lower than other tissue types (Ono *et al.* 2000, Hill *et al.* 2004, Nishino *et al.* 1996), and base excision repair genes show the highest levels of expression in the testes (Intano *et al*. 2001). Decreased expression of base excision repair genes has been described as associated

with both age and mutational frequency in older mice (Huamani *et al*. 2004); however other studies have found DNA repair to remain relatively well conserved with age (Intano *et al*. 2001) with little change evident in the male germline mutation rate with age (Hill *et al*. 2004, Nishino *et al.* 2004). A decrease in germ-line replication fidelity also doesn't explain the large heterogeneity in paternal age effect between syndromes with paternal origin of mutation. There is little evidence for a dramatic decrease in germ line repair system efficiency being responsible for the dramatic increase in mutation rate with age.

2.7 Protein-driven positive selection

One mechanism that could explain the observed increase of disease with paternal age is selective advantage of mutant cells, either through premeiotic clonal expansion within the germ line or via selective advantage for sperm. Gain-of-function advantage to sperm was first proposed as an explanation for the high incidence of the G-to-A transition of *FGFR3* within a CpG dinucleotide at base pair 1138 in achondroplasia (Tiemann-Boege 2002, Hurst & Ellegren 2002). Empirical evidence for protein-driven positive selection was found in an analysis of *FGFR2* mutations which cause Apert's syndrome. Goriely *et al.* (2005, 2003) found that the C → G transversion mutation at the mutant CpG dinucleotide is more common than the transition C → T, despite transitions being generally more common, suggesting a greater selective advantage for the transversion, which is corroborated as the variance of this mutation is greater. Neighbouring nucleotide mutation rate was not elevated, nor was the C → T mutation at the same site (which produces a stop codon). Goriely and colleagues found several examples of double mutants, which would be very improbable without some form of selection operating.

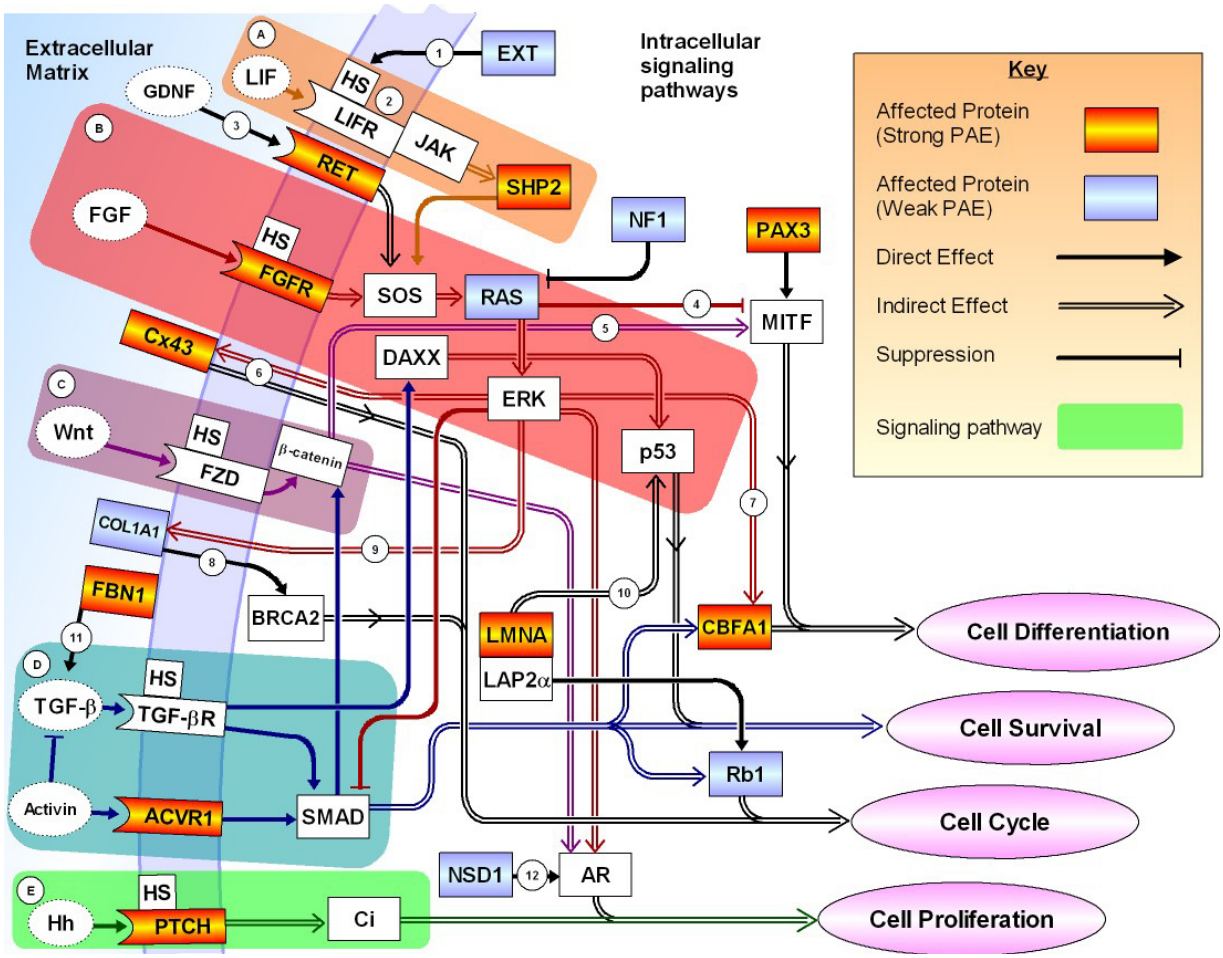


Fig. 4 - Signalling pathways which involve paternal age effect mutations.
Pathways adapted from KEGG pathways resource (http://www.genome.jp/kegg/, Kanehisa *et al* 2006) with additional references below. Groups of proteins belonging to major signaling pathways are highlighted above: (A) JAK-STAT pathway; (B) MAPK pathway; (C) Wnt pathway; (D) TGF-β pathway; (E) Hedgehog pathway.

1) Hilton *et al.* 2005, Bernfield 1999
2) Pradervand *et al* 2004
3) Takahashi 2001
4) Wellbrock & Marais 2005, Wellbrock & Arozarena 2016,
5) Tachibana *et al.* 2000
6) Mograbi *et al.* 2003
7) Franceschi *et al.* 2003
8) Moro *et al.* 2005
9) Tharaux *et al.* 2000
10) Dorner *et al.* 2006, Varela *et al.* 2005
11) Byers 2004, Pradervand et al 2004
12) Bakin *et al.* 2003
13) Melcon *el al.* 2006
14) Yang *et al.* 2002

Screening for the same mutations in blood showed no elevated mutation rate for any of the mutations and the C → G mutation rate did not show the abnormal elevation for a CpG site (Goriely *et al.* 2005, 2003). The results of an increase in sperm mutation frequency with age for achondroplasia has been replicated, but only one of two cohorts displayed an age-enhanced rate for Apert's syndrome substitutions (Wyrobek *et al.* 2006).

2.7.1 Positive selection and strong paternal age effect syndromes

Fig. 4 shows a representation of the signalling pathway network involving gene products mutated in disorders associated with paternal age effect syndromes. Most of the proteins fall within pathways that are canonical requirements of cancerous cells evolving constitutive growth and resistance to antigrowth signals (Kelleher *et al.* 2006, Hanahan & Weinberg 2000). It is therefore unlikely to be a coincidence that advanced paternal age has been associated with increased risk of cancer in progeny (Paul & Robaire 2013, Yip *et al.* 2006, Choi *et al.* 2005). Indeed, almost all of the genes which cause disease associated with paternal age are also known to contribute to cancer in somatic tissue (see *Table 3*).

This raises the possibility that the gross increase of incidence with paternal age of all strongly-designated syndromes – especially those that show mutations at CpG couplets, encode receptors or regulatory proteins such as transcription factors (which may be gain-of-function mutations) or are involved with cell division regulation (oncogenes/tumour suppressors) – is due to the selective advantages of these mutations rather than specific mutability of the genes in question.

Since positive selection has been found to be a factor in Apert's syndrome incidence, it may potentially play a role in other *FGFR2* diseases such as Crouzon and Pfeiffer syndromes. Basal cell nevus mutations affect the sonic hedgehog protein which, amongst other things,

governs stem cell division (Katoh & Katoh 2006). In fibrodysplasia ossificans progressiva, a single mutational hotspot occurs within a CpG dinucleotide in the *ACVR1* gene and it encodes a *BMP* receptor in the TGF-β signalling pathway. Taken together this is suggestive of positive selection. Multiple endocrine neoplasia 2B shows a mutational hotspot (creating a CpG site) at codon 918 affecting the catalytic core of the tyrosine kinase domain of *RET* (Carlson *et al.* 1994) and positive selection for *RET* mutations has been demonstrated in cancerous cells (Babenko *et al*. 2006). Hutchinson-Gilford's progeria is predominantly caused by a transition at a single CpG site, affecting the regulatory lamin A protein. This disrupts the *Rb1* cell cycle control via interaction with the *LAP2α* protein (Dorner *et al.* 2006), but mutations in *LMNA* can also affect structural nature of the nuclear envelope which in turn results in accumulation of DNA damage and activation of the p53 signalling pathway (Cadiñanos *et al.* 2005). Fibrillin is a structural protein forming the lattice that traps *TGF-β* and, when mutated, induces overstimulation of the pathway leading to cell proliferation (Chaudhry *et al.* 2007). Finally, *PTPN11*, the mutant gene frequently responsible for Noonan's syndrome, is a phosphatase that is important to cell division and cell survival through simulation of the MAPK pathway (Gelb & Tartaglia 2006).

2.7.2 Positive selection and weak paternal age effect

Those disorders characterised with a linear rather than exponential incidence increase with age are hypothesised to show a weaker effect due to the masking effect of maternally-derived mutations (Glaser & Jabs 2004). The relatively weak effect of these syndromes can be contrasted in many cases with that of other syndromes with predominant paternal origin of mutation yet which show little or no measurable paternal age effect at all (*e.g.* haemophilia A [Becker 1996] and Alexander disease [Li *et al.* 2006]).

The retinoblastoma and neurofibromatosis proteins are inhibitors of cell proliferation, and mutations are usually passed on from the paternal side (Glaser & Jabs 2004, Dryja *et al.* 1989, Jadayel *et al.* 1990), although deletions are common. The mutant genes responsible for multiple extostoses (*EXT1-3*) are involved in the synthesis of heparan sulphate, which is important for the correct functioning of many cell-surface receptors including fibroblast growth factor receptors, TGF-β pathway receptors and *PTCH* (Hilton *et al.* 2005, Bernfield 1999). Lastly, *NSD1*, the gene product mutated in Sotos syndrome, is a regulator of androgen, the testosterone receptor: the male-specific hormone which is critically involved in spermatogenesis and preventing germ-cell apoptosis (Dohle *et al.* 2003). It is also a key factor in prostrate cancer (Yang *et al.* 2002).

2.8 Conclusions

Many of the human genetic diseases with a paternal age association display an effect that increases faster than would be expected by the copy error hypothesis alone (Goriely & Wilkie 2012, Risch *et al.* 1987). Over half of the most dramatic paternal age effects occur in syndromes where the majority of mutations are transitions at CpG dinucleotides (*Table 1)*, which are associated with being time-dependent and would not be expected to show a paternal excess, let alone the extreme mutational hotspots and association with paternal age that these syndromes display. The diseases which show a strong increase in incidence with paternal age are also associated with an exclusively paternal origin of mutation (see *Table 2*)

While spontaneous deamination of 5-methylcytosine should not be considered totally replication-independent, there is little evidence of a drastic replication-connected mechanism that would account for the exponential increase in mutations seen with paternal age, or the large variance in the paternal age effect within disorders that are caused primarily by point mutations.

Equally there is little evidence that differential methylation between male and female germ cells is causing this severe age effect. The germ-line repair enzymes are relatively protected from senescence and while accumulation of both methylated-related and replication-mediated mutations due to deficiency in repair undoubtedly occurs, again it is not a convincing candidate for explaining the paternal age effect. The large variance in influence of paternal age on differing syndromes with similar mutational origins strongly suggests that a general mechanism such as breakdown in repair enzymes or mutational clusters is not the cause: this would be expected to have a broadly similar effect on different point mutations.

Protein-driven positive selection of mutant germ cells has been evaluated for Apert's syndrome (Yoon *et al.* 2009, Goriely *et al*. 2005). There is a striking pattern within the set of paternal age effect syndromes of mutated gene-products being involved with cell cycle and proliferation (*Table 3* and Fig. 4). This suggests that mutations in the germ line causes premeiotic expansion of the affected stem cell, causing the observed exponential increase in mutation frequency with age, especially since direct measurement of mutation in germ line cells in mice show a linear increase (Ono *et al.* 2000). The weakly-associated syndromes may show less pronounced paternal age effects due to one or more of the following: a) stronger selection masked by concealing effects such as maternally-derived cytogenetic mutations, b) weaker selection, or c) true copy-error mutational accumulation. Given their importance in cell cycle control pathways, bilateral retinoblastoma and neurofibromatosis may be more easily explained as a case of large deletions which do not show a paternal age effect masking a stronger positive selection effect (deletions in NF1, for example, are usually maternally-derived, while point-mutations are from the father's side). Weak effects, such as those seen in haemophilia A and Treacher-Collins syndrome, are probably due to real copy-error effects. Haemophilia A in

particular shows a paternal age effect specific to point mutations, consistent with the copy-error hypothesis.

The hypothesis that positive selection is the primary factor in the paternal age effect also provides an explanation for syndromes with a marked paternal origin of mutation, but no observed effect of father's age, such as Alexander disease (Li *et al.* 2006) and Townes-Brocks syndrome (Böhm *et al.* 2006). Without exacerbation of the paternal age effect, any increase in male mutation rate due to the copy error mutation accumulation might be too weak to measure significantly, especially using the limited sample sizes characteristic of human disease. In the case of haemophilia A, the paternal origin of mutation has been observed, but a paternal age effect is so weak it is only observable when deletions are excluded from the analysis (Becker *et al.* 1996).

In future study of the paternal age effect, tests for protein-driven positive selection can be conducted, as this appears to be the major contributing factor. For those researchers attempting to calculate male-to-female mutation ratios, those estimates of α derived from human genetic disease that display a paternal age effect are likely heavily skewed by positive selection. Predictions based on the hypothesis that the strong association of certain mutations with paternal age is due to positive selection can be made, for example, changes to the SOS (son of sevenless) gene has been characterised within hereditary gingival fibromatosis and a form of Noonan's syndrome, both inherited primarily as autosomal dominant disorders. The key positioning of SOS within the MAPK cascade, and the fact that it mediates signals between other proteins mutant in strong paternal age effect syndromes – such as the FGF receptors, SHP2 and RAS – makes the possibility that certain mutations associated with this gene could provoke a strong positive selective effect on mutation rates tantalizing, but as yet this has not been studied.

Syndrome	OMIM	Mutant Gene	Protein function
Acrodysotosis	101800	(unknown)	-
Achondroplasia	100800	FGFR3	Growth factor receptor – regulates cell growth, proliferation, differentiation & survival.
Apert's syndrome	101200	FGFR2	Growth factor receptor – regulates cell growth, proliferation, differentiation & survival.
Basal cell nevus	109400	PTCH	Sonic Hedgehog receptor – key role in essential developmental processes including regulating stem cell division.
Cleidocranial dysostosis	119600	CBFA1	A transcription factor of the runt domain gene family – regulates osteoblast and chondrocyte differentiation & migration.
Crouzon syndrome	123500	FGFR2	Growth factor receptor – regulates cell growth, proliferation, differentiation & survival.
FOP	135100	ACVR1	BMP receptor - important roles during osteogenesis, chrondogenesis, neurogenesis & haematopoiesis
Marfan syndrome	154700	FBN1	Constituent of microfilbrils in extracellular matrix.
MEN2A	171400	RET	Receptor tyrosine kinases – interior cell-surface, transduce signals for cell growth and differentiation.
MEN2B	162300	RET	Receptor tyrosine kinases – interior cell-surface, transduce signals for cell growth and differentiation.
Noonan's syndrome	163950	PTPN11/SHP2	Nonreceptor protein tyrosine phosphatase – important for signal response and cell division.
Oculodentodigital dysplasia	164200	Cx-43	Connexin– transmembrane gap junction protein, which forms intercellular channels.
Pfeiffer syndrome	101600	FGFR2	Growth factor receptor – regulates cell growth, proliferation, differentiation & survival.
Progeria	176670	LMNA	Lamin A - structural protein forming the meshwork lining of the nuclear envelope
Waardenburg syndrome	193500	PAX3	Transcription factor expressed in neural crest cells important for the migration and differentiation of melanocytes.
Bilateral retinoblastoma	180200	RB1	Inhibitor of cell cycle progression.
Costello syndrome	218040	HRAS	Guanosine triphosphatase – acts as molecular switch that relays growth signals.
Multiple exstoses	133700	EXT1, EXT2, EXT3	Endoplasmic reticulum-localised glycoprotein – tumor suppressor, enhances the synthesis of cell surface heparan sulphate.
Muenke-type craniosynostosis	602849	FGFR3	Growth factor receptor – regulates cell growth, proliferation, differentiation & survival.
Neurofibromatosis 1	162200	NF1	Key inhibitor of the RAS proto-oncogene.
Osteogenesis imperfecta	166200	COL1A1 or COL1A2	Collagen proteins – major extracellular structural protein.
Sotos syndrome	117550	NSD1	Co-regulator of androgen receptor.
Thanatophoric dysplasia	187600	FGFR3	Growth factor receptor – regulates cell growth, proliferation, differentiation & survival.
Treacher Collins syndrome	154500	TCOF1	Involved in ribosomal DNA gene transcription

Table 3 - Paternal age effect syndromes in humans with corresponding mutated genes.
A brief overview of protein function listed. Syndromes marked in bold have "strong" paternal age effects (Risch *et al.* 1987, Glaser & Jabs 2004).

Penrose's copy-error hypothesis is an important factor in paternal age effect disorders. However, the human diseases grouped together as those for which increasing paternal age has a strong effect on the incidence rate apparently correspond to another category: cancer-like selfish mutations which propagate themselves within the male germ-line.

CHAPTER 3

HOMEOSTASIS IN STEM CELL LINEAGES

3.1 Introduction

The concept of the stem cell niche is one that is now pivotal to our understanding of how most tissues are organised and maintained. Complex feedback networks allow tissues to self-organise and prevent the disorderly growth of cells.

Chapter 1 introduced the paternal age effect and the spectrum of human mutations. The motivation for this chapter is to provide theoretical underpinning of normal, unmutated stem cell division and to elucidate the dynamics that must exist in the feedback mechanisms that maintain the homeostasis. Chapters 4 and 5 will discuss some of these signalling pathways involved and how gain-of-function mutations can perturb the normal cell dynamic.

The goal of this chapter is to offer a mathematical model of tissue-level homeostasis within a framework of a general stem cell system. While the focus is on germline stem cells, we will discuss specific stem cell niches and the difference between them. Our model assumes cells within a tissue belong to one of three compartments, the stem cell (SC) compartment comprised of the stem cells themselves, the transit amplifying (TA) compartment, comprised of intermediate differentiating cells that divide rapidly, typically with a limited lifetime, and the differentiated cell compartment, which we assume contains the functional tissue cells that are post-mitotic. We illustrate these compartments as physically distinct chambers but there is no requirement for the compartments to be physically separated in the tissue. The dynamics of individual compartments is dominated by three parameters, the rate at which cells divide, the rate of apoptosis and the cell fate of the daughter cells. Our model reveals the intricate interplay

between local and global equilibrium required for homeostasis of the tissue. These conditions for homeostasis must be reflected in the feedback mechanisms in order for the net gain/loss of cells to be constant.

Our main results include:

- conditions required for homeostasis of the stem cell (SC) compartment;
- conditions required for homeostasis of the transit amplifying (TA) cell compartment;
- conditions required for homeostasis of the differentiated cell (post-mitotic) compartment
- a proof that the TA compartment can only reach a homeostatic state if the stem cell compartment is also in homeostasis.

We assume the existence of an integrated feedback mechanism responsible for maintaining homeostasis at the tissue level. The local equilibrium equations spell out conditions that the feedback mechanism must address. This model does not outline the feedback mechanisms required to manage cell numbers but provides boundary conditions required for homeostasis. We also assume that the signalling by which the feedback mechanism regulates the biological processes in the tissue compartments occurs without appreciable delay. In the time intervals between these pulses, the various compartment parameters remain constant.

3.2 The stem cell compartment

\

Let $X_0(t)$ denote the size of the SC compartment at time $t \geq 0$, with $X_0(t) = n_0 \geq 1$, where n_0 is the starting number of stem cells in the compartment. For $h > 0$, we let:

- $\nu_0(h)$ denote the fraction of the stem cells that divide in the interval $(t,t+h]$;
- $\gamma_0(h)$ denote the fraction of the stem cells that undergo apoptosis in the interval $(t,t+h]$.

We assume that the two limits $\lim_{h\to 0} \frac{\nu_0(h)}{h}$ and $\lim_{h\to 0} \frac{\gamma_0(h)}{h}$ exist and are finite. ν_0 represents the rate at which stem cells divide. Similarly, γ_0 represents the rate at which stem cells

are lost to apoptosis (here we include all cell death or cell loss other than differentiation under the umbrella of apoptosis). As mentioned in the previous section, we assume that over longer time intervals under normal niche maintenance both ν_0 and γ_0 are constants. Since the following ratio will appear quite often in our derivations, we find it convenient to write:

$$\theta_0 = \frac{\gamma_0}{\nu_0} \tag{1}$$

θ_0 represents the rate of change of the cell numbers in the niche. In a simple system with only stem cell symmetric divisions and apoptosis can occur, homeostasis requires $\theta_0 = 1$, higher and the numbers will shrink, lower and the numbers will increase.

Stem cells, when they divide, have the possibility of dividing symmetrically and either producing two stem cells or two differentiating cells, or asymmetrically and one of each daughter cell is produced. At the niche level, we assume this process to be one that is stochastic for individual cells but tightly controlled overall. For a dividing stem cell we let

- α_1 denote the probability that the cell produces two daughter stem cells;
- α_2 denote the probability that the cell produces one daughter stem cell and one TA cell;
- α_3 denote the probability that the cell produces two daughter TA cells;

Clearly:

$$\alpha_1 + \alpha_2 + \alpha_3 = 1 \tag{2}$$

For later reference we observe that $\alpha_1 \neq 1$ for normal tissue organisation, or no differentiated cells can be made.

In the above notation, $\nu_0(h)X_0(t)$ and $\gamma_0(h)X_0(t)$ denote, respectively, the expected number of stem cells that divide and those that undergo apoptosis in the time interval $(t,t+h]$. The size, $X_0(t+h)$, of the SC compartment at time $t+h$ can be expressed as:

$$\begin{aligned} X_0(t+h) &= X_0(t) - \nu_0(h)X_0(t) - \gamma_0(h)X_0(t) + \nu_0(h)X_0(t)\left[2\alpha_1 + \alpha_2\right] \\ &= X_0(t) + \nu_0(h)X_0(t)\left[2\alpha_1 + \alpha_2 - 1\right] - \gamma_0(h)X_0(t) \\ &= X_0(t) + \nu_0(h)X_0(t)\left[\alpha_1 - \alpha_3\right] - \gamma_0(h)X_0(t) \text{ [by (2)]} \end{aligned} \tag{3}$$

After transposing $X_0(t)$ and dividing both sides by h we obtain:

$$\frac{X_0(t+h) - X_0(t)}{h} = \frac{\nu_0(h)}{h} X_0(t)\left[\alpha_1 - \alpha_3\right] - \frac{\gamma_0(h)}{h} X_0(t) \tag{4}$$

which, upon taking limits as $h \to 0$, yields the differential equation:

$$\begin{aligned} \frac{dX_0(t)}{dt} &= \lim_{h\to 0} \frac{\nu_0(h)}{h} X_0(t)\left[\alpha_1 - \alpha_3\right] - \lim_{h\to 0} \frac{\gamma_0(h)}{h} X_0(t) \\ &= X_0(t)\left[\nu_0\left(\alpha_1 - \alpha_3\right) - \gamma_0\right] \end{aligned} \tag{5}$$

with the boundary condition $X_0(0) = n_0$. By solving (5) for $X_0(t)$ we obtain

$$X_0(t) = n_0 e^{\left[\nu_0(\alpha_1 - \alpha_3) - \gamma_0\right]t} \tag{6}$$

We now take note of a conceptually useful result implied by (6).

Corollary 3.2.1: Unless $\theta_0 < 1$, the SC compartment cannot be homeostatic.

Proof: Since both α_1 and α_3 are probabilities, and since $\alpha_{1 \neq 1,}$ the difference $|\alpha_1 - \alpha_3|$ is strictly less than 1. As a consequence, $\gamma_{0 \geq} \nu_0$ implies $\gamma_{0 \geq} \nu_0|\alpha_1 - \alpha_3|$ which, by (6), guarantees that $X_0(t)$ is a decreasing function of time.

In turn, Corollary 3.2.1 confirms the intuitive feeling that if the SC compartment is to be homeostatic then the rate at which stem cells are lost to apoptosis must be strictly smaller than the rate at which they divide. However, this condition $\theta_0 < 1$ might be violated for short, transient, periods of time in the wider context of time-dependent feedback mechanisms.

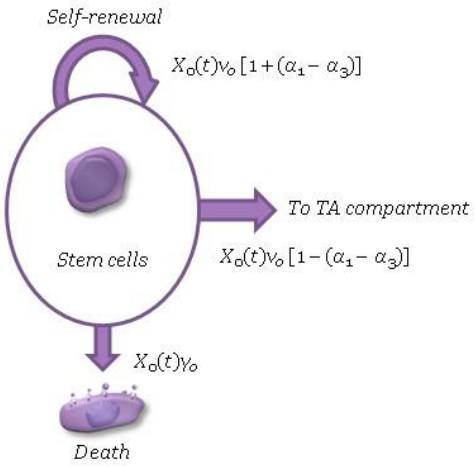


Fig. 5 - Illustrating the dynamics of the SC compartment

The dynamics of the stem SC compartment are illustrated in Fig. 5. We can determine that:

- new stem cells are being produced as a result of cell divisions at the rate of $X_0(t)v_0(2\alpha_2+\alpha_3) = X_0(t)v_0[1+(\alpha_1-\alpha_3)]$;
- stem cells are being lost to apoptosis at the rate of $X_0(t)\gamma_0$;
- TA cells are being produced at the rate of $X_0(t)v_0(\alpha_2 + 2\alpha_3) = X_0(t)v_0[1-(\alpha_1-\alpha_3)]$.

We are now in a position to state and prove the following fundamental result implied by (6).

Theorem 3.2.2. A necessary and sufficient condition for homeostasis of the SC compartment:

$$\alpha_1 - \alpha_3 = \frac{\gamma_0}{v_0} = \theta_0 \tag{7}$$

Proof: If (7) holds true, then $v_0(\alpha_1-\alpha_3)-\gamma_0 = 0$ and so, by virtue of (6), $X_0(t) = n_0$, independent of t, indicating that the SC compartment is homeostatic. Conversely, suppose the SC compartment is homeostatic and therefore $X_0(t) = n_0$. (6) implies that $v_0(\alpha_1-\alpha_3)-\gamma_0 = 0$ so (7) is validated. This also makes intuitive sense. In a homeostatic state, the loss of stem cells to cell death must be countered by symmetric stem cell division.

Theorem 3.2.2 indicates that $\theta_0 = \frac{\gamma_0}{v_0}$ is a critical value as far as homeostasis of the SC compartment is concerned. Indeed, if $\alpha_1 - \alpha_3 > \frac{\gamma_0}{v_0}$ the SC compartment grows exponentially. On the other hand, if $\alpha_1 - \alpha_3 < \frac{\gamma_0}{v_0}$ then the SC compartment decreases exponentially.

At this point we observe that:

- equation (7) is independent of n_0, the original number of cells in the SC compartment. In turn, this seems to suggest that any feedback mechanism that maintains SC compartment homeostasis must act on v_0, γ_0 or indeed the probabilities α_1 and α_3 subject to (7);

- as will be discussed later in some detail, the rate at which TA cells are produces by cell division in the stem cell compartment depends on $\alpha_1 - \alpha_3$. The lower $\alpha_1 - \alpha_3$ is, the more TA cells are being produced per unit time;
- The probabilities α_1 and α_3 only occur in (7) through the expression $\alpha_1 - \alpha_3$. This implies that as long as the probabilities of symmetric divisions are shifted up or down by equal amounts, homeostasis is preserved. This is reasonable: providing the requirements of differentiating cell production and stem cell attrition replacement are met, it doesn't matter if the majority of stem cells divide with α_2 asymmetrical divisions such as in the drosophila germ cell niche (Chen *et al.* 2016) or a balance of α_1 and α_3 divisions as may occur in the mammailian spermatogonial stem cell niche (de Rooij & Griswold 2012), the balance of differentiating to stem cells is what matters. The feasible ranges for these probabilities are investigated below.

Lemma 3.2.3 For every value of α_1 in the range $\left[\theta_0, \frac{1+\theta_0}{2}\right]$ *there exist feasible probabilities of α_2 and α_3 that satisfy both (2) and (7).*

Proof: We begin by justifying the stated range for α_1. For this purpose, recall that by (2), $2\alpha_1 + \alpha_2 = 1 + (\alpha_1 - \alpha_3)$. Since $\alpha_2 \geq 0$, (7) leads to:

$$\alpha_1 \leq \frac{1+\theta_0}{2} \tag{8}$$

Combined with the fact that $\alpha_1 \geq \theta_0$:

$$\theta_0 \leq \alpha_1 \leq \frac{1+\theta_0}{2} \tag{9}$$

To complete the proof, we need to show that for each value of α_1 in the range $\left[\theta_0, \frac{1+\theta_0}{2}\right]$ there exist probabilities α_2 and α_3 that satisfy both (2) and (7). To see that this is the case, we can assign $\alpha_1 = u$ for an arbitrary u in the interval $\left[\theta_0, \frac{1+\theta_0}{2}\right]$.

By (7), the expression of α_3 must be $\alpha_3 = u - \theta_0$. We can therefore observe:

$$0 \le \alpha_3 \le \frac{1+\theta_0}{2} \tag{10}$$

It also follows that $\alpha_2 = 1 + \theta_0 - 2u$. By our choice of u, then:

$$0 \le \alpha_2 \le 1-\theta_0 \tag{11}$$

The expressions for α_1, α_2, and α_3 obtained above satisfy (2) and (7) and the proof is complete.

Lemma 3.2.3 confirms our intuition that homeostasis of the SC compartment can occur for a large number of values of α_1 and, consequently, of α_2 and α_3. Moreover, as we shall show, as long as the criteria for (7) are met, the rate at which new TA cells arise as a result of divisions in the SC compartment is independent of the actual values of α_1, α_2, and α_3. It is likely that feedback mechanisms that keeps the SC compartment homeostatic favour some of these values over others, there are undoubtedly other factors beyond purely the maintenance of homeostasis that feature into the decision.

Next, we turn our attention to the rate at which TA cells are being produced within the SC compartment..

Lemma 3.2.4 Under homeostatic conditions, the rate at which TA cells are being produced in the SC compartment is $n_0 (v_0 - \gamma_0)$.

Proof: Recall that the rate at which TA cells are being produced by cell divisions in the SC compartment is $X_0(t)\nu_0(\alpha_2 + 2\alpha_3)$. Consequently under homeostasis, we can write:

$$\begin{aligned} X_0(t)\nu_0(\alpha_2+2\alpha_3) &= n_0\nu_0(\alpha_2+2\alpha_3) \\ &= n_0\nu_0\left[1-(\alpha_1-\alpha_3)\right] \\ &= n_0\nu_0(1-\theta_0) \ \text{ [by (7)]} \\ &= n_0(\nu_0-\gamma_0) \ \text{ [by (1)]} \end{aligned} \tag{12}$$

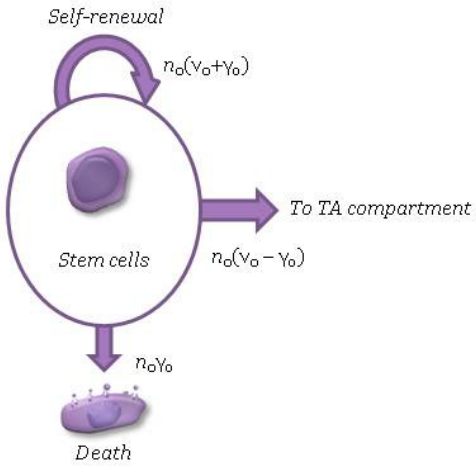


Fig. 6 - Illustrating the dynamics of a homeostatic SC compartment.

New stem cells are produced at the rate of $n_0(\nu_0 + \gamma_0)$, cells are lost to apoptosis at a rate of $n_0\gamma_0$ and TA cells are being produced at the rate of $n_0(\nu_0 - \gamma_0)$. Once created, TA cells join the dynamics of the TA compartment described in section 3.3.

3.3 The transit amplifying cell compartment

Let $X_1(t)$ denote the size of the TA cell compartment at time $t \geq 0$, with $X_1(t) = n_1 \geq 1$. For $h>0$, we let:

- $\nu_1(h)$ denote the fraction of TA cells that divide in the interval $(t,t+h]$;

• $\gamma_1(h)$ denote the fraction of TA cells that undergo apoptosis in the interval $(t,t+h]$.

We assume that the two limits $\lim_{h\to 0} \frac{\nu_1(h)}{h}$ and $\lim_{h\to 0} \frac{\gamma_1(h)}{h}$ exist and are finite. In this notation, ν_1 represents the rate at which TA cells divide, while γ_1 represents the rate at which TA cells undergo apoptosis.

For a given dividing TA cell, we let:

- β_1 denote the probability that the cell produces two daughter TA cells;
- β_2 denote the probability that the cell produces one daughter TA cell and one terminally differentiated cell;
- β_3 denote the probability that the cell produces two terminally differentiated cells.

As with the SC compartment:

$$\beta_1 + \beta_2 + \beta_3 = 1 \tag{13}$$

For some small $h>0$, let $X_1(t + h)$ denote the number of TA cells at time $t+h$. $X_1(t + h)$ involves the following components:

- $X_1(t) - \nu_1(h)X_1(t) - \gamma_1(h)X_1(t)$: the TA cells in existence at time t that have undergone neither division nor apoptosis in $(t,t+h]$;
- $X_0(t)\nu_0(h)[\alpha_2 + 2\alpha_3]$: the new TA cells generated in $(t,t+h]$ by divisions in the SC compartment;
- $X_1(t)\nu_1(h)[2\beta_1 + \beta_2]$: the new TA cells generated in $(t,t+h]$ by divisions of TA cells.

Therefore we can write these together as:

$$\begin{aligned} X_1(t+h) &= X_1(t) - \nu_1(h)X_1(t) - \gamma_1(h)X_1(t) + X_0(t)\nu_0(h)\left[2\alpha_2 + \alpha_3\right] + X_1(t)\nu_1(h)\left[2\beta_1 + \beta_2\right] \\ &= X_1(t) + X_1(t)\left[\nu_1(h)\left(\beta_1 - \beta_3\right) - \gamma_1(h)\right] + X_0(t)\nu_0(h)\left[\alpha_2 + 2\alpha_3\right] \end{aligned} \tag{14}$$

After algebraic manipulations of the above expression we obtain:

$$\frac{X_1(t+h) - X_1(t)}{h} = X_1(t)\left[\frac{\nu_1(h)}{h}\left(\beta_1 - \beta_3\right) - \frac{\gamma_1(h)}{h}\right] + X_0(t)\frac{\nu_0(h)}{h}\left[\alpha_2 + 2\alpha_3\right] \tag{15}$$

which, upon taking limits as $h \to 0$, yields the differential equation:

$$\frac{dX_1(t)}{dt} = X_1(t)\left[\nu_1\left(\beta_1 - \beta_3\right) - \gamma_1\right] + X_0(t)\nu_0\left[\alpha_2 + 2\alpha_3\right] \tag{16}$$

with boundary condition $X_1(0) = n_1$.

By (6), the differential equation (16) with boundary condition $X_1(0) = n_1$ can be rewritten as:

$$\frac{dX_1(t)}{dt} - AX_1(t) = \phi(t) \tag{17}$$

where

- $A = \nu_1(\beta_1 - \beta_3) - \gamma_1$, and
- $\phi(t) = n_0\nu_0(\alpha_2 + 2\alpha_3)e^{[\nu 0(\alpha 1 - \alpha 3) - \gamma 0]t} = n_0\nu_0[1 - (\alpha 1 - \alpha 3)]e^{[\nu 0(\alpha 1 - \alpha 3) - \gamma 0]t}$

Using standard techniques, the solution of (17) turns out to be

$$X_1(t) = e^{At}\left[n_1 + \int_0^t e^{-Au}\phi(u)du \right] \tag{18}$$

which, upon evaluating the integral, becomes

$$X_1(t) = n_1 e^{At} + n_0 v_0 \left[1-\left(\alpha_1 - \alpha_3\right)\right] \frac{e^{\left[v_0(\alpha_1-\alpha_3)-\gamma_0\right]t} - e^{At}}{v_0\left(\alpha_1-\alpha_3\right)-\gamma_0 - A} \tag{19}$$

where, recall, $A = v_1(\beta_1 - \beta_3) - \gamma_1$.

We now state and prove a technical lemma that reveals the complexity of the feedback mechanism at work in tissue lineages.

Lemma 3.3.1. If $A = v_0(\alpha_1 - \alpha_3) - \gamma_0$ then the TA compartment cannot be homeostatic.

Proof: Suppose that this is not the case. Lemma A.2 in Appendix A guarantees that as $A \to v_0(\alpha_1 - \alpha_3) - \gamma_0$

$$\frac{e^{\left[v_0(\alpha_1-\alpha_3)-\gamma_0\right]t} - e^{At}}{v_0\left(\alpha_1-\alpha_3\right)-\gamma_0 - A} \to te^{At} \tag{20}$$

so the expression of $X_1(t)$ in (19) becomes:

$$X_1(t) = e^{At}\left(n_1 + n_0 v_0\left[1-\left(\alpha_1-\alpha_3\right)\right]t\right) \tag{21}$$

However this latter expression shows that $X_1(t)$ is a function of t. This remains true even if $A = 0$ in which case $X_1(t) = n_1 + n_0v_0[1 - (\alpha_1 - \alpha_3)]t$ grows linearly with t. This prohibits homeostasis and completes the proof.

Observe that the conclusion of Lemma 3.1 is independent of whether or not the SC compartment is homeostatic. The proof of Lemma 3.1 gives us a hint of the growth regimens seen by the TA compartment if it is not in equilibrium. Refer to the Appendix for details.

The following useful corollary follows directly from the proof of Lemma 3.3.1.

Corollary 3.3.2 If the TA compartment is homeostatic then A≠0.

Next, recalling that $A = v_1(\beta_1 - \beta_3) - \gamma_1$, Lemma 3.3.1 can be rephrased as follows:

Corollary 3.3.3 If

$$\beta_1 - \beta_3 = \frac{\gamma_1}{v_1} + \frac{v_0(\alpha_1 - \alpha_3) - \gamma_0}{v_1} \tag{22}$$

then the TA compartment cannot be homeostatic regardless of whether or not the SC compartment is.

The dynamics of the TA compartment are illustrated in Fig. 7. Note that:

- new TA cells are being produced as a result of TA cell divisions at the rate of $X_1(t)v_1[2\beta_1 + \beta_2] = X_1(t)v_1[1 + (\beta_1 - \beta_3)]$;
- new TA cells are being produced by divisions in the SC compartment at a rate of $X_0(t)v_0[1 - (\alpha_1 - \alpha_3)]$;
- TA cells are being lost to apoptosis at the rate of $X_1(t)\gamma_1$;
- differentiated cells are being produced at the rate of $X_1(t)v_1[\beta_2 + 2\beta_3] = X_1(t)v_1[1 - (\beta_1 - \beta_3)]$;

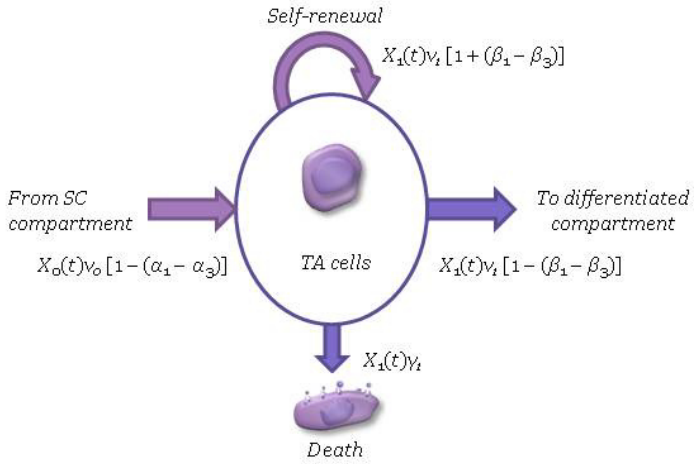


Fig. 7 - Illustrating the dynamics of the TA cell compartment

As pointed out by several authors (Johnston *et al.* 2007, Lander *et al.* 2009, Tomlinson & Bodmer 1995), accumulated empirical evidence suggests that if homeostasis is not present at the SC compartment, then neither can the TA compartment be homeostatic. To our knowledge, this phenomenon has not previously been proven.

Theorem 3.3.4 In order for the TA compartment to be homeostatic then the SC compartment must also be homeostatic.

Proof: If the TA compartment is homeostatic, then $X_1(t) = n_1$ for all $t \geq 0$. With algebraic manipulation, (19) can be re-written as:

$$n_1\left[1-e^{At}\right]=\frac{n_0\nu_0\left[1-\left(\alpha_1-\alpha_3\right)\right]}{\nu_0\left(\alpha_1-\alpha_3\right)-\gamma_0-A}\left(e^{\left[\nu_0(\alpha_1-\alpha_3)-\gamma_0\right]t}-e^{At}\right) \tag{23}$$

where, as a direct consequence of Corollary 3.3.2

$$1-e^{At}\neq 0 \tag{24}$$

After dividing (23) by $1 - e^{At}$ and some simple algebra, n_1 can be written as:

$$\begin{aligned} n_1 &= \frac{n_0 v_0\left[\alpha_2+2\alpha_3\right]}{v_0\left(\alpha_1-\alpha_3\right)-\gamma_0-A}\cdot\frac{e^{\left[v_0(\alpha_1-\alpha_3)-\gamma_0\right]t}-e^{At}}{1-e^{At}} \\ &= \frac{n_0 v_0\left[1-\left(\alpha_1-\alpha_3\right)\right]}{v_0\left(\alpha_1-\alpha_3\right)-\gamma_0-A}\cdot\left[1-\frac{1-e^{\left[v_0(\alpha_1-\alpha_3)-\gamma_0\right]t}}{1-e^{At}}\right] \end{aligned} \tag{25}$$

Since the TA compartment is homeostatic, the right-hand side of (25) must be independent of t. Observe that:

$$\frac{n_0 v_0\left[1-\left(\alpha_1-\alpha_3\right)\right]}{v_0\left(\alpha_1-\alpha_3\right)-\gamma_0-A} \tag{26}$$

is a constant and therefore, in order for the right-hand side of (25) to be a constant, the expression

$$1-\frac{1-e^{\left[v_0(\alpha_1-\alpha_3)-\gamma_0\right]t}}{1-e^{At}} \tag{27}$$

must be a constant. By Lemma A.1 in Appendix A this happens if and only if $v_0(\alpha_1–\alpha_3)–\gamma_0 = 0$. However, by Theorem 3.2.2 this guarantees the SC must be homeostatic.

One last point needs clarification, namely that the conditions of Lemma A.1 in Appendix A are verified. To confirm this with $a = v_0(\alpha_1–\alpha_3)–\gamma_0$ and $b = A$ we can see that:

- by Corollary 3.3.2, $1 - e^{At} \neq 0$ and so $b \neq 0$,
- by Lemma 3.3.1, $A \neq v_0(\alpha_1 - \alpha_3) - \gamma_0$ and so $a \neq b$,

confirming that Lemma A.1 does indeed apply. This completes the proof of Theorem 3.3.4.

We are now in a position to offer a necessary and sufficient condition for the homeostasis of the TA compartment.

Theorem 3.3.5 The TA compartment is homeostatic if and only if the SC compartment is homeostatic and, in addition,

$$\beta_1 - \beta_3 = \frac{n_1\gamma_1 - n_0\left(\nu_0 - \gamma_0\right)}{n_1\nu_1} \tag{28}$$

In biological terms this means that the net gain of TA cells must equal rate of loss minus the rate of influx per cell division.

Proof: Firstly, if the TA compartment is homeostatic, then by Theorem 3.3.4, the SC compartment is also homeostatic. Recall that by (7), $\alpha_1 - \alpha_3 = \gamma_0/v_0$. Thus (25) can be re-written as:

$$\begin{aligned} n_1 &= \frac{n_0 v_0\left[1-\left(\alpha_1-\alpha_3\right)\right]}{v_0\left(\alpha_1-\alpha_3\right)-\gamma_0-A}\cdot\left[1-\frac{1-e^{\left[v_0(\alpha_1-\alpha_3)-\gamma_0\right]t}}{1-e^{At}}\right] \\ &= \frac{n_0 v_0\left[1-\dfrac{\gamma_0}{v_0}\right]}{-A} \\ &= \frac{-n_0\left(v_0-\gamma_0\right)}{v_1\left(\beta_1-\beta_3\right)-\gamma_1} \end{aligned} \tag{29}$$

With algebraic manipulation we can confirm that (29) yields (28).

Conversely, we can assume that the stem cell compartment is homeostatic and that (28) holds true. We therefore need to show that the TA compartment is also homeostatic. Since the stem cell compartment is homeostatic, (7) along with Lemma A.1 (i.e. $v_0\,[\alpha_1 - \alpha_3] - \gamma_0 = 0$) guarantees the following:

$$n_0 v_0 \left(\alpha_2 + 2\alpha_3\right) = n_0 \left(v_0 - \gamma_0\right) \tag{30}$$

$$A = v_1 \left(\beta_1 - \beta_3\right) - \gamma_1 = \frac{-n_0 \left(v_0 - \gamma_0\right)}{n_1} \tag{31}$$

$$v_0 \left(\alpha_1 - \alpha_3\right) - \gamma_0 - A = -A = \frac{n_0 \left(v_0 - \gamma_0\right)}{n_1} \tag{32}$$

$$e^{\left[v_0(\alpha_1-\alpha_3)-\gamma_0\right]t} - e^{At} = 1 - e^{At} \tag{33}$$

Upon replacing the values of these expressions back into (19), we obtain:

$$X_1(t) = n_1 e^{At} + n_1 \left[1 - e^{At}\right] = n_1 \tag{34}$$

confirming that the TA compartment is homeostatic. This completes the proof of the theorem.

This leads to the following observations:

- The probabilities β_1 and β_3 only occur in (28) through the expression $\beta_1 - \beta_3$. This implies that, provided the SC compartment is homeostatic, as long as the probabilities of symmetric divisions are shifted up or down by equal amounts, homeostasis of the TA compartment is preserved;
- (28) relies on knowledge of the rate $n_0(v_0 - \gamma_0)$ at which TA cells are being produced in the SC compartment;
- In contrast to the SC compartment, (28) depends on the size, n_1, of the TA compartment. In practice, this is achieved by a finite space in the niche architecture, so that TA cells are limited by the presence of neighbours if the niche becomes full or by feedback mechanisms (for example, Aoki *et al.* 2016)
- (28) suggests that homeostasis of the TA compartment may occur in two mutually exclusive ways, depending on the sign of the expression $\beta_1 - \beta_3$:
    - **Case 1:** $\beta_1 - \beta_3 \geq 0$, in which case the probability that a dividing TA cell produces two daughter TA cells is larger than or equals the probability of producing two differentiated cells;

- **Case 2:** $\beta_1 - \beta_3 < 0$, in which case the probability that a dividing TA cell produces two daughter TA cells is strictly smaller than the probability that it produces two differentiated cells.

We assume that Case 2 is the normal circumstance of stem cell systems as Case 1 bypasses the need for stem cells altogether as transit amplifying cells would self-renew indefinitely. However as discussed in the conclusions to this chapter, there is an argument for Case 1 to occur in some stem cell systems.

For later reference, we now prove the following technical result.

Lemma 3.3.6 Under homeostatic conditions in the TA compartment

$$X_1(t)v_1\left(\beta_2+2\beta_3\right)=n_0\left(v_0-\gamma_0\right)+n_1\left(v_1-\gamma_1\right) \tag{35}$$

Proof: Under homeostasis, $X_1(t) = n_1$ independent of t. Thus we can write:

$$\begin{aligned} X_1(t)v_1\left(\beta_2+2\beta_3\right) &= n_1v_1\left(\beta_2+2\beta_3\right) \\ &= n_1v_1\left(1-\beta_1+\beta_3\right) \\ &= \left(n_0v_0+n_1v_1\right)-\left(n_0\gamma_0+n_1\gamma_1\right) \text{ [by Theorem 3.3.5]} \\ &= n_0\left(v_0-\gamma_0\right)+n_1\left(v_1-\gamma_1\right) \end{aligned} \tag{36}$$

completing the proof.

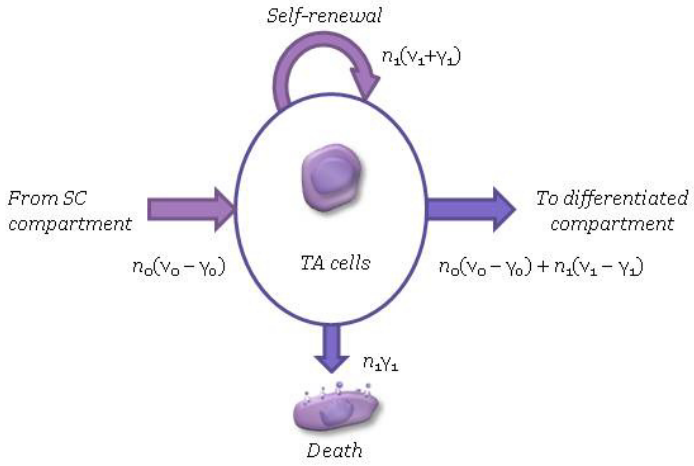


Fig. 8 - Illustrating the dynamics of the TA cell compartment under homeostasis

Fig. 8 captures the dynamics of the TA cell compartment under homeostatic conditions. Indeed, with n_1 standing for the steady state number of TA cells:

- new TA cells are produced by divisions in the TA compartment at the rate of $n_1(\nu_1 + \gamma_1)$;
- new TA cells are produced by divisions in the SC compartment at the rate of $n_0(\nu_0 - \gamma_0)$
- TA cells are being lost to apoptosis at the rate of $n_1\gamma_1$;
- terminally differentiated cells are being produced at the rate of $n_0(\nu_0 - \gamma_0) + n_1(\nu_1 - \gamma_1)$.

3.4 The differentiated cell compartment

Let $X_2(t)$ denote the size at time $t \geq 0$ of the terminally differentiated (i.e. post-mitotic) cell compartment. We assume $X_2(t) = n_2 \geq 0$.

For $h>0$, let $\gamma_2(h)$ denote the fraction of the differentiated cell compartment that dies and is shed in $(t,t+h]$. We assume that the limit $\lim_{h\to 0} \frac{\gamma_2(h)}{h}$ exists and is finite. γ_2 represents the rate at which differentiated cells die.

Referring to Fig. 9, let us consider what might happen in the time interval $(t,t+h]$:

- terminally differentiated cells do not divide and cannot self-renew;
- $X_1(t)v_1[\beta_2 + 2\beta_3]$ new differentiated cells are added from the TA cells compartment;
- $X_2(t)\gamma_2(h)$ differentiated cells die and are shed.

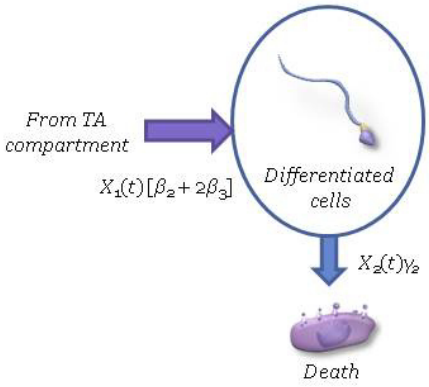


Fig. 9 - Illustrating the dynamics of the differentiated cell compartment.

$X_2(t + h)$ can be broken into the following components

$$X_2(t+h) = X_2(t) - X_2(t)\gamma_2(h) + X_1(t)v_1(h)\left[\beta_2 + 2\beta_3\right] \tag{37}$$

which can be written as

$$\frac{X_2(t+h)X_2(t)}{h} = \frac{\gamma_2(h)}{h} X_2(t) + \frac{v_1(h)}{h} X_1(t)\left[\beta_2 + 2\beta_3\right] \tag{38}$$

Upon taking limits as $h \to 0$ and using Lemma 3.3.6 we obtain the differential equation

$$\frac{dX_2(t)}{dt} = \gamma_2 X_2(t) = \left(n_0 v_0 + n_1 v_1\right) - \left(n_0 \gamma_0 + n_1 \gamma_1\right) \tag{39}$$

with boundary condition $X_2(0) = n_2$.

Using standard techniques the solution of this differential equation turns out to be

$$X_2(t) = e^{-\gamma_2 t}\left[n_2 - \frac{\left(n_0 v_0 + n_1 v_1\right) - \left(n_0 \gamma_0 + n_1 \gamma_1\right)}{\gamma_2}\right] + \frac{\left(n_0 v_0 + n_1 v_1\right) - \left(n_0 \gamma_0 + n_1 \gamma_1\right)}{\gamma_2} \tag{40}$$

3.4.1 Homeostasis of the differential cell compartment

We observe that there cannot exist homeostasis of the differentiated cell compartment unless both the stem cell and TA compartments are homeostatic.

If we are to see homeostasis in the differentiated cell compartment, it must be that $X_2(t) = n_2$ independent of t. In particular, this observation along with (40) allow us to write

$$n_2 = e^{-\gamma_2 t}\left[n_2 - \frac{\left(n_0 v_0 + n_1 v_1\right) - \left(n_0 \gamma_0 + n_1 \gamma_1\right)}{\gamma_2}\right] + \frac{\left(n_0 v_0 + n_1 v_1\right) - \left(n_0 \gamma_0 + n_1 \gamma_1\right)}{\gamma_2} \tag{41}$$

which, after suitable manipulations, yields

$$n_2\left[1 - e^{-\gamma_2 t}\right] = \frac{\left(n_0 v_0 + n_1 v_1\right) - \left(n_0 \gamma_0 + n_1 \gamma_1\right)}{\gamma_2}\left[1 - e^{-\gamma_2 t}\right] \tag{42}$$

Since $\gamma_2 \neq 0$, we can simplify the above to:

$$n_2 = \frac{(n_0 v_0 + n_1 v_1) - (n_0 \gamma_0 + n_1 \gamma_1)}{\gamma_2} \tag{43}$$

or, equivalently,

$$n_0 v_0 + n_1 v_1 = n_0 \gamma_0 + n_1 \gamma_1 + n_2 \gamma_2 \tag{44}$$

Observe that equation (47) tells us that the differentiated cell compartment is homeostatic only if the expected number of stems cells and TA cells produced by division per unit time matches the expected number of cells that die (in all compartments) per unit time.

3.5 Conclusions

The stem cell compartment is governed by the critical value $\theta_0 = \frac{\gamma_0}{\nu_0}$. This condition must be met for homeostasis to occur within the compartment. In the special case where $\gamma_0 = 0$ and only symmetric divisions occur, that is $\alpha_2 = 0$, has been studied extensively in the literature (Colijn & Mackey 2005, Johnston *et al.* 2007, Lander *et al.* 2009, Tomlinson & Bodmer 1995). As several of these authors pointed out, in that particular case, homeostasis of the SC compartment hinges on the very strong condition $\alpha_2 = \alpha_2 = ½$ which shows that the system is inherently unstable.

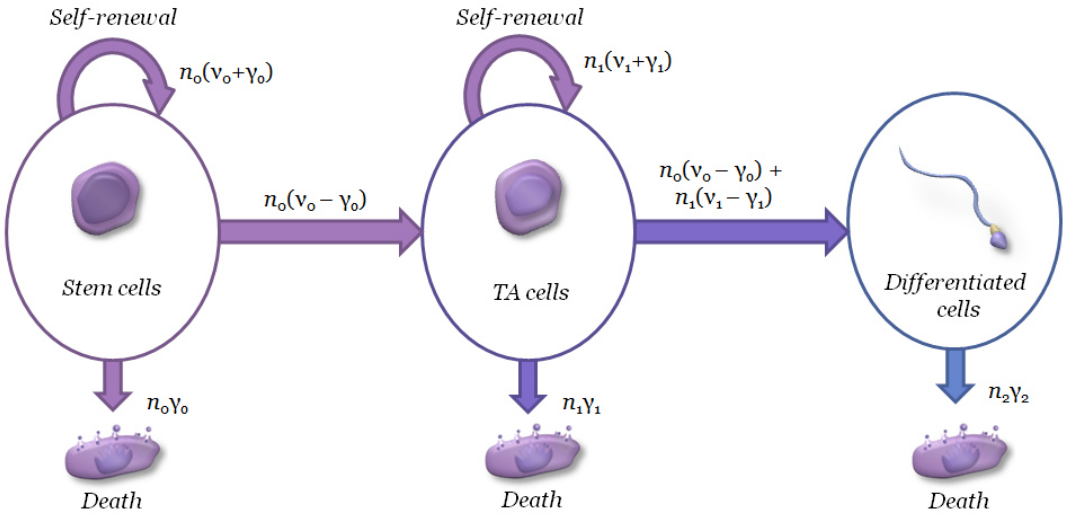


Fig. 10 - Summary of the model

It is noteworthy that this critical value is irrespective of the number of cells within the niche and is controlled by the net gain/loss of new SCs under the balance of symmetric divisions α_1–α_3 and there can exist a range of feasible values for α_1 and α_3 that satisfy the conditions for homeostasis (see Lemma 3.2.3).

In order for the transit amplifying compartment to be homeostatic, the stem cell compartment must also be homeostatic (Theorem 3.3.4). This has implications for modelling developing stem cell systems as well as systems returning from a perturbation (e.g. wound healing). Starting from a position of exponential growth, the system must reach equilibrium in the stem cell compartment before (or alongside) the transit amplifying compartment. Similarly to the SC compartment, the TA compartment must balance self-renewing divisions against the influx from the SC compartment minus the loss to apoptosis. See (28). As might be expected,

homeostasis in the TA compartment relies on the niche balancing the rate of production of TA cells in the SC compartment. Unlike the SC compartment, the balance of differentiating vs. self renewing symmetric divisions in the TA compartment depends not only on the number of TA cells but the number of SC cells as well.

The dynamics of the TA compartment can be in equilibrium in one of two mutually exclusive arrangements, depending on the balance of symmetrical self renewing divisions (β_1) to symmetrical differentiating divisions (β_3). **Case 1:** $\beta_1 - \beta_3 \geq 0$ and **Case 2:** $\beta_1 - \beta_3 < 0$.

Case 2 seems to be the normal circumstance by which the stem cell system operates (e.g. colonic crypt), as Case 1 would mean that the stem cells do not divide and are not required for TA compartment homeostasis unless the compartment's balance is disturbed in some fashion. However, while it may seem counter-intuitive that you might have a stem cell system without stem cell divisions, Case 1 may occur in some systems, for example the spermatogonial stem cell nice. In the scenario of Case 2, stem cells would not divide, but given a disruption in the TA compartment such that $\beta_1 - \beta_3 < 0$ no longer held true, then stem cells would recommence division. Within the spermatogonial stem cell niche, there are two populations of stem cells, A_{dark} cells that are quiescent and A_{pale} stem cells that divide continuously and give rise to differentiating progeny (Yoon *et al.* 2009). When the A_{pale} population is reduced through radiation, the A_{dark} cells activate and become A_{pale} cells (van Alphen *et al.* 1988). Within the context of our model, the A_{pale} cells, although commonly regarded as the spermatogonial stem cells, are what we would term as the long-lived TA cells in the unusual Case 2 scenario and the A_{dark} cells are the true stem cells.

The differentiated stem cell compartment is fairly straightforward. The TA compartment must produce enough differentiated cells to offset loss in the differentiated compartment.

However an interesting result per equation (47) is that homeostasis in the differentiated compartment requires that the total number of stem cells and transit amplifying cells produced at any given time must offset the loss in all compartments to cell death.

The model presented in this chapter is a novel model for a generic stem cell system and details boundary conditions for homeostasis that must be adhered to via cell feedback mechanisms in order for equilibrium to be reached. Unlike previous work this model does not assume any arbitrary determination of cell fate after division and allows for both symmetric and asymmetric cell division. It is noteworthy that this model allows for a great variety of cell divisions patterns and symmetric stem cell divisions (α_2) are unimportant for homeostasis as long as the boundary conditions are met for homeostasis. This allows the number of α_2 divisions to vary wildly and indeed in practice we can see in nature, even within the spermatogonial stem cell niche, systems that seem to be chiefly symmetric like germ-line stem cell divisions in drosophila (Chen *et al.* 2016) to largely symmetric divisions such as the mammalian stem cell niche (Klein *et al.* 2010).

CHAPTER 4

SELECTIVE MUTATION ACCUMULATION:

A COMPUTATIONAL MODEL OF THE PATERNAL AGE EFFECT

4.1 Introduction to the model

This chapter presents a mathematical model to simulate the accumulation of mutations in the spermatogonial stem cell (SSC) niche due to replication errors and premeiotic positive selection.

4.1.1 The Spermatogonial Stem Cell Niche

SSCs reside on the basal lamina on the outer edge of the seminiferous tubules within the testes. Spermatogonia are surrounded by much larger Sertoli cells that form the microenvironment for the cells. The spermatogonia divide in cyclical waves and progeny of the stem cells migrate as they divide and differentiate towards the hollow centre of the seminiferous tubule (de Rooij & Russel 2000). In humans the active SSCs are comprised of type A_{pale} spermatogonia (so named to differentiate between those spermatogonia that cycle continuously and the A_{dark} undifferentiated spermatogonia that are quiescent, although both sets seem to have stem cell properties).

Certain stem cell systems like the colonic crypt have specific arrangements of cells with strictly limited numbers of stem cells (Humphries & Wright 2008). The SSC niche on the other hand lacks obvious repeating structures. However, SSCs are observed to localise to specific areas of the seminiferous epithelium (Yoshida 2008). While spermatogonia can repopulate whole seminiferous tubules that have been depleted by radiation (Shinohara *et al.* 2001), studies of live

imaging of stem cells indicate limited migrational capabilities (Klein *et al.* 2010). Additionally, cells that migrate away from the niche are likely to differentiate (de Rooij & van Beek 2013), likely because GDNF distribution is patch-like (Sato *et al.* 2011), creating effective niche limits.

4.1.2 Motivation & Predictions

The objective of this model was to simulate the accumulation of mutations through positive selection. The magnitude of the selective effect (*r*) has not been determined experimentally, although it has been estimated to be 0.014 for the FGFR2 mutations causing Apert's syndrome by Yoon and colleagues in earlier models based upon spatial arrangement of mutations (Yoon *et al.* 2009, Choi *et al.* 2008, Qin *et al.* 2007). This corresponds to an expansion of mutant stem cells over wildtype occurring with a probability of 1.4% with every division of a mutant cell.

Our primary aim was to estimate the *r* value for a range of disease causing mutations. We hypothesised that diseases with a more exponential increase with paternal age would have larger *r* values. We also aimed to determine if differences in incidence rate between alleles of a single gene or between mutations that cause different disorders is due to the underlying mutation rate or to the selective advantage of the particular alleles. Additionally, where different mutations affect the same gene, if there is a stronger activating mutation, we expected a higher *r* value.

The SSC niche has been the subject of some computer simulations of the normal stem cell niche (de Rooij & van Beek 2013, Ray *et al.* 2014) and also in terms of mutation accumulation by positive selection for Apert syndrome (Yoon *et al.* 2009) but no models have mathematically modeled the accumulation of mutation or looked at mutation accumulation

across different genes and disorders. This paper presents a new model based on a Markov chain system along with simulation and comparison to data from available data.

4.2 Methods

4.2.1 Model

We assume that the stem cell niche represents a closed system, with n cells contained within. Of the n cells, each can be in one of two states, mutant or wildtype. With each cell division, a stem cell acquires the mutation with probability p. Stem cells are assumed to divide asymmetrically, one daughter differentiating and ultimately being lost and the other remaining in the niche. With normal divisions, therefore, the number of stem cells will not change regardless of divisions. Mutant cells, however, have a positive selective probability of r and if a selection event happens, will divide symmetrically to produce two mutant cells. This will increase the niche size above the maximum. To correct this, a random cell is ejected from the niche and lost (all cells are eligible for ejection, including the newly-formed mutant cell).

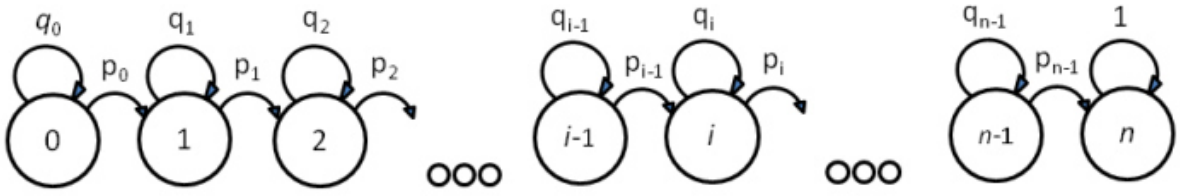


Fig. 11 - Representation of the probabilities associated with a niche of *n* stem cells.

The niche, represented by circles, starts with 0 mutant cells out of n. After a stem cell divides in the niche, the probability p_0 indicates the chance that a mutation occurs and brings the system into state 1 (i.e. one mutant cell in the niche) and q_0 that after the cell division the niche remains in state 0. Once in state 1, the probability of advancing to state 2 has changed to p_1 as the chance of mutation remains but the mutant cells may expand with a selective event. Correspondingly, the chance that a cell division occurs that maintains the niche in state 1 is now q_1. This continues until, ultimately, all of the n cells are mutant, at which point the system remains in state n.

We have modeled the system as a Markov chain (Fig. 11). The chain has n+1 states, where state 0 represents the niche comprised entirely of wildtype cells and in the final state, n, the cells are entirely mutant. Cells are selected at random from the niche to divide sequentially. State i represents the niche with i mutant cells and n-i wildtype cells. In state i there is within the niche a probability p_i that after a cell division the number of mutant cells will increase by one, and a probability q_i that they will remain the same. The probability of a mutation and a subsequent reversal at the same site is sufficiently small as to be ignored.

Let us imagine that the niche is in state i and a random cell is selected to divide. If the cell selected is wildtype, then with probability p it is transformed into a mutant and returned to the niche, otherwise it is returned as a wildtype. If a mutant cell is selected, it is simply returned to the niche unless a selection event happens with probability r, in which case two mutant cells are returned and subsequently one random cell is lost (all cells including the returned cells are eligible to be lost).

In order to calculate q_i, the probability that after a cell division the system remains in state i, there are therefore three mutually exclusive possibilities with the following probabilities:

(1) A wildtype cell is selected for division, but no mutation occurs:

$$\frac{n-i}{n}(1-p) \tag{45}$$

(2) A mutant cell is selected for division and no selection event happens:

$$\frac{i}{n}(1-r) \tag{46}$$

(3) A mutant cell divides, a selection event happens and a mutant cell is lost from the niche.

$$\frac{i}{n} r \frac{i+1}{n+1} \tag{47}$$

The combined probability q_i is therefore:

$$q_i = \frac{n-i}{n}(1-p) + \frac{i}{n}(1-r) + \frac{i}{n} r \frac{i+1}{n+1} \tag{48}$$

With some rearrangement:

$$q_i = 1 - p + \frac{i}{n}\left[p - \frac{r(n-i)}{n+1} \right] \tag{49}$$

Since $p_i = 1 - q_i$, this can be rewritten as:

$$p_i = \frac{n-i}{n}\left[p + \frac{ir}{n+1} \right] \tag{50}$$

The Markov chain produces a matrix, T, with dimensions of $(n+1) \times (n+1)$, where rows indicate the starting state and column denote the final state (i.e. the initial and final number of mutant cells).

$$T = \begin{bmatrix} q_0 & p_0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & q_1 & p_1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & q_2 & p_2 & \cdots & 0 & 0 \\ 0 & 0 & 0 & q_3 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & q_{n-1} & p_{n-1} \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \tag{51}$$

This matrix (51) gives the probability of moving from one state to another in one step i.e. one cell division within the stem cell niche. The element T_{ij} gives the probability of starting in state i (i.e. a niche containing i mutant cells) and after one cell division in the niche ending in state j with j mutant cells. In order to model the progression of multiple cell divisions within the niche, the matrix can simply be raised to the power of the number of cell divisions. T^2 will give a matrix that provides probabilities for two steps (i.e. two cell divisions within the niche) and T^K will provide probabilities for traversing in K steps. In the final matrix, therefore, the element $T^K{}_{ij}$ represents the probability that a niche starting with i mutant stem cells will end with j mutant stem cells after K cell divisions. Note that K is total cell divisions occurring amongst any of the cells in the niche, not the average number of divisions per cell, which would be K/n.

For an individual, the number of steps required, K, is a factor of the rate of cell division d (divisions per year per cell), the age of the individual a in years and the number of cells per niche n.

$$K = nda \tag{52}$$

The matrix T^K was for any given age solved computationally, by generating a matrix T and calculating values for each cell and then raising the matrix to the power K. Since the model assumes every individual starts with 0 mutant cells, the only relevant part of the final solved matrix is the top row.

$$\begin{array}{cc} State & \begin{array}{ccccccc} 0 & 1 & 2 & 3 & \dots & n-1 & n \end{array} \\ P_S & \begin{bmatrix} P_0 & P_1 & P_2 & P_3 & \dots & P_{n-1} & P_n \end{bmatrix} \end{array} \tag{53}$$

For a given state s, the value P_s denotes the probability that starting with 0 mutant cells, after K cell divisions, the niche will have s mutant cells. The average number of mutants per

niche (assuming sufficient number of replications), *M*, can be calculated by summing all of the final state values multiplied by the corresponding probabilities.

$$M = \sum_{s=0}^{n} sP_s \tag{54}$$

M/n gives us a single proportion of mutant to wildtype cells at a given age *a.* Within an individual person many niches will deviate dramatically from the average value, but since the number of niches within a single individual can be assumed to be high (see section 2.3), but all contribute sperm equally, we can assume that the overall proportion of mutant sperm to wildtype sperm as a whole will tend towards to *M/n.*

This can be proven as follows. There are *N* niches each with *n* stem cells. Each niche has a different number of mutant cells, u_1, u_2, etc. Each stem cell contributes an equal number of sperm, for simplicity we assume one sperm per stem cell but the following is true for any value of sperm produced per stem cell division as the proportion will remain constant.

The total number of mutant cells *U* is:

$$U = \sum_{i=0}^{N} u_i \tag{55}$$

Dividing by the total number of stem cells over all the niches (*Nn*) gives the proportion of mutant stem cells over total cells, which is equivalent to the mean proportion of mutant cells per niche:

$$\frac{U}{Nn} = \frac{\overline{X}_U}{n} \tag{56}$$

Where $\overline{X}_U$ is the mean number of mutant cells per niche for a given number of niches *N*. Taken to the limit:

$$\lim_{N\to\infty}\left(\frac{\overline{X}_U}{n}\right)=\frac{M}{n} \tag{57}$$

Therefore using the ideal average value *M/n* from a single niche is an accurate measure of mutant sperm proportion for the entire individual providing *N* is large and all niches contribute sperm equally.

4.2.2 Confirmation of Model Design by Simulation

To test the mathematical model, we designed a simulation alongside it to emulate the progression of a single stem cell niche. Early versions of our simulation were much more complex and attempted to simulate all the SSCs within an individual, but simulation of a single niche is sufficient to test the model, particularly averaged over a large number of repeats (see equation 2.13). Simulation of the stem cell niche was designed in C++. Script design simulated a stem cell population at a niche level independent of the Markov chain model (supplementary algorithm 1).

The simulation progressed by selecting a random cell. If the selected cell was wildtype it became mutant with probability *p* and if the selected cell was mutant it underwent a selective event with probability *r*. Selective events represented a symmetric division and added a mutant cell to the niche. As a consequence a random cell was then ejected and lost from the niche. This process was repeated and the simulation was allowed to run for a specific number of cell divisions sufficient to represent a human reproductive lifespan ($n\times d\times 80$ years). The results from each run were then averaged over a hundred thousand repeats of the simulation.

The simulation was tested against the mathematical expression by applying equation 2.10 with the same parameters and age values as the simulation. The simulation tended towards the

values provided by the model and showed perfect agreement with sufficient replication (supplementary Fig. B.2).

4.2.3 Parameters

Mutation probability, *p*. Rahbari *et al.* (2016) estimated the mutation frequency per nucleotide per germline cell division at 4×10^{-11} calculated by sequencing multi-sibling families. This mutation rate is close to that calculated with phylogenetic data (Lynch 2010) and point mutations on the Y-chromosome (Helgason *et al.* 2015). Rahabari and colleagues also noted little variation of mutation rate with paternal age, which allowed us to assume p is a constant value regardless of age. This is the baseline mutation probability per site, before accounting for elevated mutational frequency due to CpG sites or multiple disease-causing alleles within a single gene.

Stem cells per niche, *n*. SSCs are not organised in regular repeating structures with defined cell numbers like the colonic crypt (Humphries & Wright, 2008) and this makes estimating the niche size difficult. While initial models of the SSC niche did not have discrete compartmentalization, recent research has shown preferential clustering of SSCs to specific regions of the seminiferous basal lamina (de Rooij & Griswold 2012, Yoshida *et al.* 2007). To estimate the number of SSCs per niche, we turned to studies in mice, where spermatogenesis has been reconstituted in sterile mice by transplantation of SSCs bearing a reporter gene. From Shinohara *et al.* (2001), adult mice generated at least 108 colonies per testis. Given 35,000 stem cells per mouse testis (Tegelenbosch & de Rooij DG 1993), this amounts to 324 stem cells per niche. Scaling up to human testes by weight, assuming the same number of stem cells per niche,

gives us approximately 3 million individual niches, which fulfills the requirement of equation 2.13. The model also assumes the number of SSCs remains constant throughout life. In reality, the number of stem cells declines with age (Paul & Robaire 2013). Assuming attrition occurs evenly among mutant and wildtype stem cells, this stochastic loss will not affect the proportion of mutant to wildtype sperm over many niches. The caveat to this is that it is possible that the mutant cells, rather than a proliferative advantage, are granted some form of resistance to the age-based attrition. Finally, as hypothesised by Yoon and colleagues (Yoon *et al.* 2009), non-proliferating A_{dark} spermatogonia may activate as reserve stem cells and replace losses (including lost mutant cells) with wildtype cells. In Chapter 1, this would represent a Case 1 system, see equation (28) in Theorem 3.3.5. However with cell death of the long-lived transit amplifying A_{pale} cells, the balance of $(\beta_1 - \beta_3)$ divisions would shift to a Case 2 dynamic and necessitate the activation of A_{dark} cell division to replace them. This would cause a "dip" in the mutation frequency, irrespective of gene as fresh wildtype cells are introduced into the system. This is beyond the scope of this model, however the implications are discussed in Appendix B.

Stem cell divisions per year, *d*. Human spermatogenesis results in one stem cell divisions per spermatogenic cycle of the A_{pale} spermatogonia, so once every 16 days (de Rooij & Russel, 2000), although lower estimates exist (Tomasetti & Vogelstein, 2015). The model selects cells to divide randomly rather than in waves, however the odds of the same cell being selected repeatedly is low and the results are averaged over a large number of niches, the effect of this is negligible and allows us to avoid tracking individual cells.

Selection pressure, *r*. This is the probability that a mutant stem cell will divide symmetrically and self-renew. As this value has not been determined experimentally, the model fits a best-fit curve for the optimal *r* value to fit the data. Note that in normal steady-state

division, SSCs may divide symmetrically into A_{paired} spermatogonia where both daughter cells remain stem cells or both differentiate (de Rooij & Griswold 2012) for which there is some evidence (Klein *et al.* 2010). For simplicity we have assumed, as earlier models have done (Yoon *et al.* 2009), that each stem cell divides asymmetrically in normal homeostatic cell division rather than a balance of differentiating and self-renewing divisions.

4.2.4 Fitting the Model to Mutation Data

In order to match our model to existing paternal age effect data, we started with birth incidence of various genetic diseases. The larger number of younger parents versus older ones is accounted for by looking at Observed/Expected values, the number of births for a given age category divided by the expected number of births assuming the total number of disease-affected children were distributed to each age category proportional to births in that population.

Using census data from 1966 USA birth data (Vital Statistics of the United States, 1966, U.S. Department of Health, Education and Welfare) to estimate a number of births per age category, C_a, (as per Risch *et al.* 1987). 1966 most closely matched the birth data used. The fraction, M/n, of mutant-to-wildtype sperm for the given age category was used to derive a number of disease-affected births for that category, simulated mutant births (S).

$$S = \frac{M}{n} C_a \tag{58}$$

By dividing the number of fathers in the age category by the total population and then multiplying this proportion by the total number of simulated mutant births, we can calculate the predicted mutant births (P) assuming simulated mutant births are distributed proportional to the paternal age distribution.

$$P = \frac{C_a}{C_t} \sum S \tag{59}$$

Simulated/Predicted is therefore directly comparable to the Observed/Expected data.

In order to fit the S/P data to the existing O/E values, the strength of selection, *r*, had to be empirically determined. A script was generated in R (R Development Core Team, 2008) that would match the O/E values according to the following algorithm:

(1) For a given value of *r,* calculate a single S/P value for the median age of the following age categories: 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54. In each case the average age for the category was used to generate the S/P value.

(2) Calculate the difference between each S/P value and the corresponding O/E value. Take the sum of squares (SOS) of these differences. .

(3) The process was optimised for *r* using R's *optim* function by searching for the *r* value with the lowest SOS score.

We also compared our model to high-throughput sequencing data. In this case, proportional numbers of mutant sperm were compared directly to the model's predicted proportion of mutant stem cells to wildtype cells. The same procedure as above was used, except instead of SOS of O/E-S/P, the number of mutant cells per 10^6 cells was used directly and the SOS between experimental and calculated number of mutants was generated.

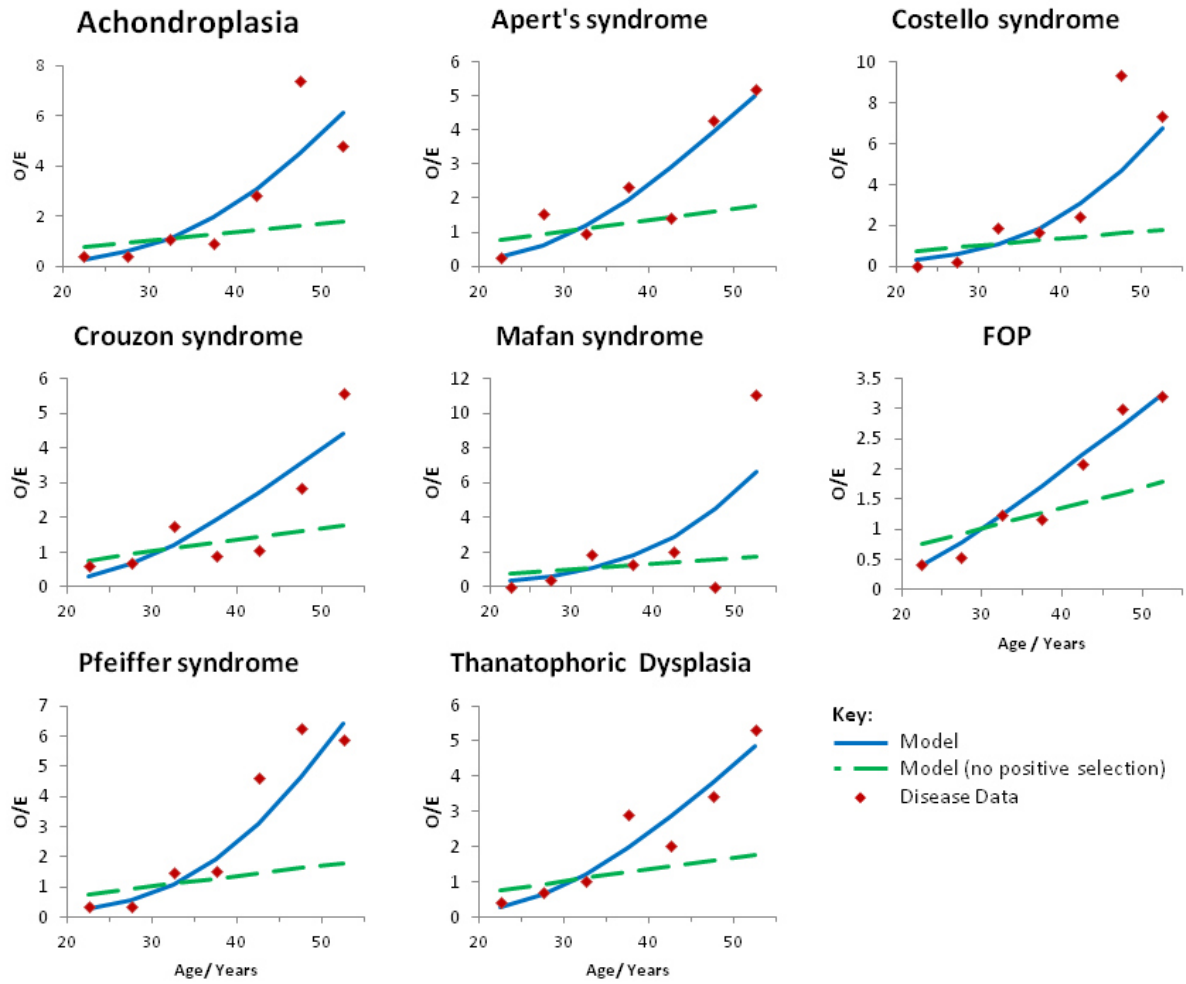


Fig. 12 - Simulation of observed/expected birth numbers.

FOP – fibrodysplasia ossificans progressiva, TD – thanatophoric dysplasia.

O/E disease data from Risch *et al*. (1987), except for Costello syndrome (O/E ratios calculated from Lurie 1994), achondroplasia and thanatophoric dysplasia from Orioli *et al*. (1995).

4.3 Results

Fig. 11 shows the matched model and disease data graphs for 8 disorders that show a strong paternal age effect. Excluding FOP as an outlier (see discussion), the remaining predicted incidence values correlated significantly with the actual incidence values (Pearson's correlation

coefficient = 0.91, $p<0.05$). With those disorders where sequence data is available (Costello, Apert and thanatophoric dysplasia syndromes), *r* values can be compared directly between disease data and sperm mutation rates and show close agreement (TOST equivalence test, $\varepsilon=0.0053$, $p<0.05$). The probabilities of positive selection varied from 0.5% to 1.5% with a mean *r* value of 0.0083 for *r* values from birth data and 0.0094 from sequence data.

The predicted incidence rate is shown in table 2. The raw incidence assumes the baseline mutation rate *p* of 4×10^{-11}. However, mutation rate varies by location in the genome and the sequence in question and of particular interest are CpG sites. These are particularly mutable as cytosine in CpG sites is a methylation site and can spontaneously deaminate to thymine (Lynch 2010). In order to account for the increased mutation chance, the probability of CpG-specific alleles mutating was multiplied by a factor of 15 for a transition or 5 for a transversion (Nachman & Crowell 2000). Additionally, a number of paternal age effect disorders are caused by multiple mutations at a variety of loci and by looking at disease incidence of the disorder as a metric, we include all mutations that contribute to that disease phenotype. In order to simulate this in the model, the aggregate probability of mutation, p_a, is the probability of any of the mutations occurring that give rise to the phenotype, with the formula:

$$p_a = 1 - \left((1 - p)^a \right) \tag{60}$$

Where *p* is the baseline mutation probability and *a* is the number of potential mutation sites that can produce mutant alleles.

Disease	Gene	*r* value (sequencing data)[a]	*r* value (birth data)[b]
Achondroplasia	FGFR3	-[c]	0.00741
Apert's syndrome	FGFR2	(C755G) 0.0124 (C758G) 0.0126	0.00888
Costello syndrome	HRAS	(G34A) 0.00526	0.00606
Crouzon syndrome	FGFR2	-	0.00997
FOP	ACVR1	-	0.0135
Marfan syndrome	FBN1	-	0.00517
Pfeiffer syndrome	FGFR2	-	0.00668
TD	FGFR3	(A1948G) 0.0105	0.00937

Table 4 - Strength of positive selection (r) for 8 diseases
FOP – fibrodysplasia ossificans progressiva, TD – thanatophoric dysplasia.
[a]Calculated by directly matching mutation incidence to that from sperm DNA sequencing (see Fig. 12), specific mutation is shown in parentheses.
[b]Calculated from birth incidence rates by making the best fit of O/E curves, with adjusted mutation rates (see Fig. 11).
[c]Several studies have estimated the mutation rate of achondroplasia in sperm but have been omitted due to concerns of the methodology (Maher *et al.* 2014).

4.4 Discussion

We hypothesised that one of the two Apert's syndrome-causing mutations (C755G) has a higher incidence than the other (C758G) because the former occurs at a CpG dinucleotide, which has a higher mutability due to spontaneous deamination. Our results support this hypothesis as both of the Apert mutations have a very similar *r* value with 0.0124 for C755G compared with

0.0126 for C758G. These two distinct amino-acid substitutions (S252W and P253R) appear therefore to have the same selective effect on the cell and the increased incidence of S252W is purely because of increased mutability at this site. The projected incidence rate is very sensitive to the model parameters and particularly the mutation rate. For example, achondroplasia has a high incidence rate relative to the other diseases (1 in 27,000, see table 2). The computed *r* value on the other hand was middle of the range, which failed to account for the high incidence rate when the baseline value for *p* was used. Once the value of *p* was increased to the level of a C→T transition, the predicted incidence agreed well. So the site-specific mutation rate accounted for the relatively high incidence rate of this disease.

The predicted incidence rates, after accounting for the number of alleles and the mutation rate, present close to the actual values, with the exception of fibrodysplasia ossificans progressiva. This disease is anomalous as it is a very rare disease (one in 2 million births) but it is predicted to have a high incidence rate as it is caused by a transition at a single CpG site (Shore *et al*. 2006). While the rates of substitution vary by location in the genome, including the rates at CpG sites (Mugal & Ellegren 2011, Fryxell & Moon 2005), it is unlikely the substitution rate could be low enough at this point, even if unmethylated. It might be explained by a very low selective advantage but the projected *r* value is high (0.0135), producing an O/E curve similar to achondroplasia. The low birth prevalence is also not explained by low survival to term of affected offspring as FOP does not show severe symptoms until later in life or by any variation of expression of the mutant allele as it shows complete penetrance (Petrie *et al.* 2009) so the low incidence rate remains unexplained.

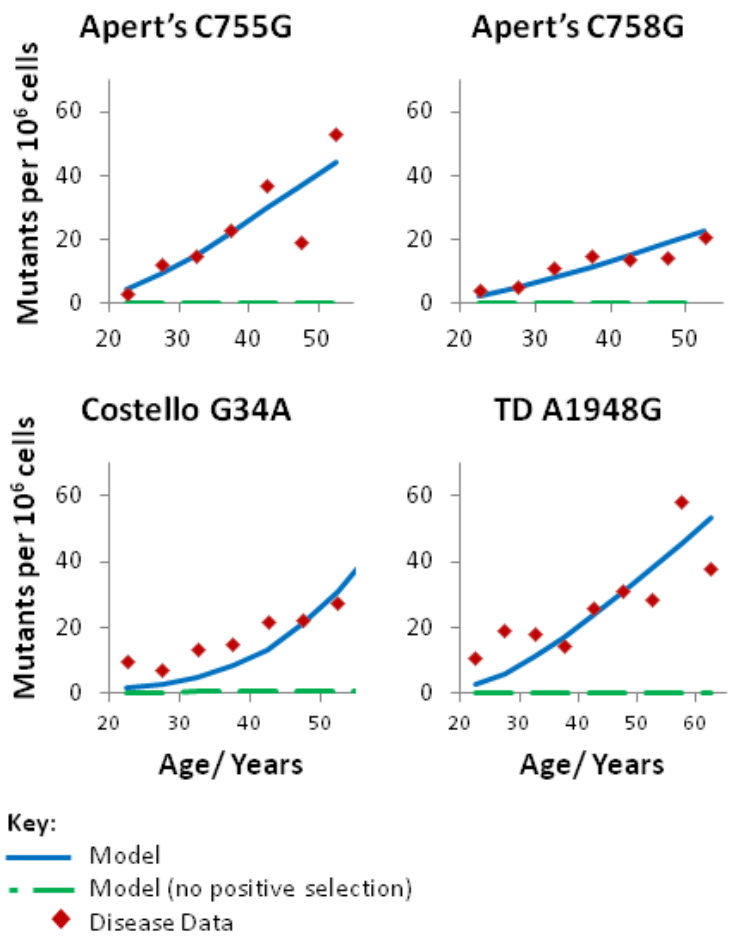


Fig. 13 - Rates of mutants per million sperm with age.

Apert syndrome (Yoon *et al.* 2009), Costello syndrome (Giannoulatou *et al.* 2013), Thanatophoric dysplasia (TD) adapted from Goriely *et al.* (2009).

The mutations causing thanatophoric dysplasia and achondroplasia both cause constitutive activation of the *FGFR3* but TD mutations activate the receptor more strongly, leading to a more severe phenotype (Naski *et al.* 1996, Bonaventure *et al.* 2007). We can therefore predict the *r* value for TD to be higher than that for achondroplasia, which is confirmed by our model (TD $r = 0.0105$, achondroplasia $r = 0.00741$). Note that both sets of data were taken from one study (Orioli *et al.* 1995) in order to ensure that they are comparable.

Disease	Raw Predicted Incidence Rate	Adjusted Predicted Incidence Rate	Literature Incidence Rate	Reference	Alleles
ACH	1 in 400,000	1 in 27,000	1 in 26,000	[1]	1
Apert	1 in 700,000	1 in 130,000	1 in 100,000	[2]	2
Costello	1 in 2,300,000	1 in 160,000	1 in 300,000	[3]	14
Crouzon	1 in 600,000	1 in 37,000	1 in 60,000	[4]	16
FOP	1 in 340,000	1 in 23,000	1 in 2,000,000	[5]	1
Marfan	1 in 3,400,000	1 in 68,000	1 in 70,000	[6]	50
Pfeiffer	1 in 1,600,000	1 in 130,000	1 in 100,000	[7]	12
TD	1 in 720,000	1 in 60,000	1 in 40,000	[8]	12

Table 5 - Incidence Rates of 8 Diseases
ACH – achondroplasia, FOP – fibrodysplasia ossificans progressiva, TD – thanatophoric dysplasia.

Raw predicted incidence rates calculated with a baseline mutation rate of 4×10^{-11}. Adjusted rates account for variation in mutation rate at CpG dinucleotides and the number of mutable alleles that cause the disease phenotype. Alleles denote the number of most common genetic variants that comprise at least 95% of cases of the disease. (Online Mendelian Inheritance in Man, 2016). Incidence rates of ACH is a mean value between 0.36 and 0.6 per 10,000 after accounting for 20% of ACH cases being inherited from an affected parent. Sources: [1] Faruqi *et al.* 2014, [2] Blank 1960, [3] Lurie 1994, [4] Helman *et al*. 2014, [5] Hüning & Gillessen-Kaesbach 2014, [6] Lynas 1958, [7] Vogels & Fryns 2006, [8] Connor *et al*. 1985.

In a previous model, Yoon and colleagues estimated r to be 0.014 (Yoon *et al.* 2009) and our model presents a value close to that based on their sequencing data ($r = 0.0125$), although the value from birth data was lower ($r = 0.00888$). It is noteworthy that these values for Apert's syndrome were higher than the average value for the other disorders.

The model presented in this chapter provides a mathematical understanding of the accumulation of selfish disease-causing mutations. We have successfully predicted the incidence rates of different diseases based on O/E curves and information of the molecular nature of the mutations and estimates for the strength of selection. The selective advantage granted by these mutations is the most important factor in terms of the exponential increase over time but the site-specific mutation rate and the number of mutable sites plays a key role in how common the disease is at the population level.

CHAPTER 5

INVESTIGATION INTO THE PROLIFERATIVE ADVANTAGE OF RET M918T MUTATIONS IN HUMAN SPERMATOGONIAL STEM CELLS

5.1 *In Vitro* Analysis of the Paternal Age Effect

Investigations into the cause of the paternal age effect have focussed on two diseases, achondroplasia and Apert's syndrome, both caused by single SNPs in the *FGFR2* and *FGFR3* genes respectively, and have concentrated in analysing mutation distribution patterns in human testes (Qin et al. 2007, Yoon *et al.* 2009, Dakouane Giudicelli *et al.* 2007). These studies rely on sampling sperm or testes cells and extrapolating a causative nature to the mutations. No *in vitro* or animal model of premeiotic germ cell selection has been created, and while many other candidate genes exist, no others have been studied.

The hypothesis that mutations expand in the germline prior to mitosis requires a functional effect of the mutation on stem cell division. Spermatogonial stem cell division is tightly regulated, typically dividing into two daughter cells with different fates. Traditional stem cell theory holds that one remains a stem cell, maintaining the stem cell population within the niche; the other divides and differentiates into spermatocytes. This asymmetric cell division is controlled by extracellular signals that activate certain signalling cascades. More recent research, however, suggests that some of the early differentiating forms can retain stem-cell capacity and may be able to recolonise a stem cell niche. (Barroca *et al.* 2009). Recurrent paternal age effect mutations are often the same as found in cancer (see Fig. 14).

While the full details of the cues and controls determining spermatogonial stem cell fate are a subject of active research, glial cell derived neurotrophic factor (GDNF) is undoubtedly a

key player that dictates whether any given stem cell will differentiate, divide asymmetrically or divide symmetrically to produce two new stem cells (Sariola & Saarma 2003). Produced by the Sertoli cells surrounding the spermatogonial stem cells, GDNF is essential for maintaining normal stem cell division: high GDNF dosage is associated with spermatogonial stem cells dividing symmetrically (Sariola & Saarma 2003). Mice overexpressing GDNF show accumulation of undifferentiated spermatogonia and eventually display nonmetastatic testicular tumours (Meng *et al.* 2000).

The gene of interest was selected on the basis of the following criteria; (a) strong paternal age effect and connection to canonical growth signalling pathway, (b) disease phenotype caused by highly recurrent (>90% of disease incidence) single base pair mutation and (c) connection to known factors associated with stem cell renewal. Prime candidates were *FGFR2* C755G (Apert's syndrome), *FGFR3* G1128A (achondroplasia), *RET* T2943C (multiple endocrine neoplasia type 2B) and *ACVR1* G617A (fibrodysplasia ossificans progressiva), which fulfill all criteria. SSCs themselves represent an invaluable stem cell model system given their ease of access *in vivo* and as germline cells their DNA will ultimately be passed on to the next generation, crucial for both natural and bioengineered hereditary traits.

This investigation's results will have impact in understanding the specific causes of the high *de novo* incidence of MEN2B, the paternal age effect in general as well as broader questions concerning spermatogonial stem cell disregulation.

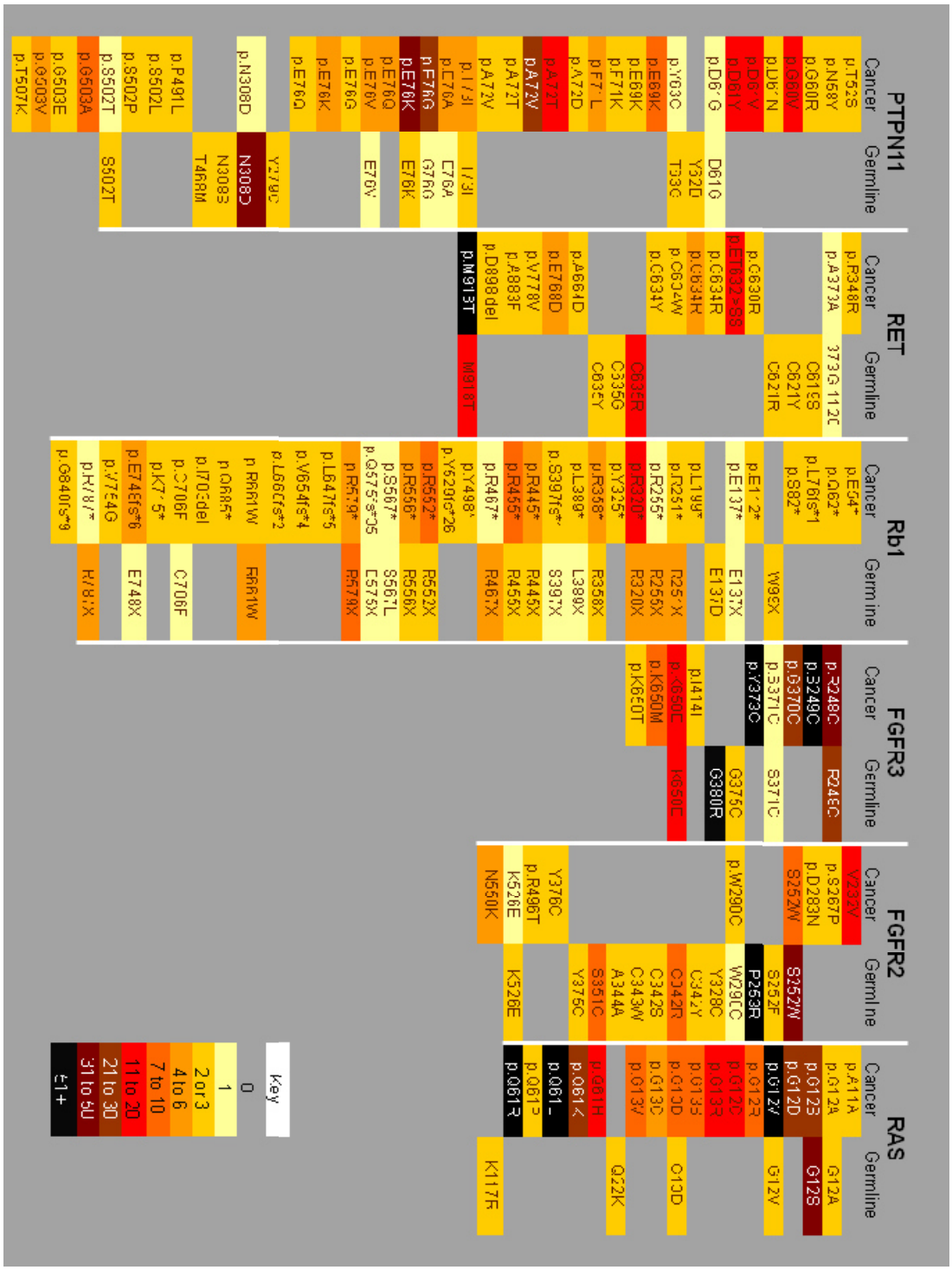


Fig. 14 - Incidence rates in four genes of mutations in 6 genes
Each incidence of a particular mutation is noted and colour-coded. Data was compiled from OMIM (germline data - http://www.omim.org/) and COSMIC (cancer data: http://cancer.sanger.ac.uk/cosmic)

5.2 Methods

5.2.1 Induced Pluripotent Stem Cell (iPSC) creation

iPSCs were a kind gift from Patrick Sachs and Peter Mollica (per established techniques, e.g. Fusaki *et al.* 2009). To briefly summarise their work in manufacturing the iPSCs, they were created using the CytoTune-iPS 2.0 Sendai Reprogramming Kit (Invitrogen). Briefly, human BJ cells were transiently transfected with the proprietary construct that did not integrate (IBC #13-021). Tests were performed to ensure (1) they were endogenously expressing pluripotency markers, (2) all of the virus was removed from the cells, (3) they differentiated into each dermal lineage and (4) they expressed TRA-1-81 and positive for alkaline phosphatase activity. For iPSC-BJ passaging, initially manual passaging was conducted to eliminate MEF contamination, and subsequently we used enzymatic passaging with dispase. iPSCs were maintained on MEF (mouse embryonic fibroblast) feeder layers and supplemented with KSOR media.

5.2.2 Differentiation into Spermatogonial Stem Cells

iPSCs were differentiated by applying SSC media, changed daily, to iPSC colonies for 10 days before passaging, per Easley *et al.* (2012). SSCs were mainted on STO feeders. STO feeders (SNL 76/7, mitomycin C treated, Applied StemCell Inc) were plated at a concentration of 5×10^4 cells/cm^2 (typically plated at a concentration of 2×10^5 cells/ml). STOs were cultured in DMEM supplemented with 15% FBS for at least 24 hours before plating SSCs. SSCs were cultured in 6-well plates in SSC media changed every 1 day initially and every 2 days once established.

Reagent	Manufacturer	Concentration
α-MEM	GIbco	Basal media
Penicillin	Sigma	50 units/ml
Streptomycin	Sigma	50 units/ml
Bovine serum albumin (BSA)	Calibiochem	0.2%
Transferrin	Sigma/Gibco	10 µg/ml
Linolenic acid	Sigma	5.6 mM
Oleic acid	Sigma	13.4 mM
Palmitoleic acid	Sigma	2.3 mM
Linoleic acid	Sigma	35.6 mM
Palmitic acid	Sigma	31.0 mM
Stearic acid	Sigma	11.6 mM
Na_2SeO_3	Sigma	3×10^{-8} M
L-glutamine	Sigma	2 mM
Insulin	Gibco	25 µg/ml
HEPES	J. T. Baker	10 mM
Putrescine	Sigma	120 µM
2-mercaptoethanol	Sigma	120 µM
Recombinant human GDNF	Gibco	20 ng/ml
rat GFRα1	Gibco	150 ng/ml
human bFGF	Gibco	20 ng/ml

Table 6 - Standard SSC Media

5.2.3 Immunocytofluorescence.

DDX4 (DEAD-Box Helicase 4) and PLZF (Promyelocytic Leukaemia Zinc Finger protein) were chosen as markes for SSCs, both show enrichment for SSC identity and good antibodies exist for these proteins (Easley *et al.* 2012). Cells were fixed in 4% paraformaldehyde (except for the Tra-1 stain) and then blocked for 1 hour at 37°C. Blocking buffer used was 5% BSA and 5% Normal Goat Serum (NGS), 0.25% Triton-X in PHEM. DDX4 was used in primary incubation of 1:200, PLZF at 1:50 with secondary antibodies anti-DDX4 (LifeTech Alexafluor 594 goat anti-mouse) at 1:1000 and anti-PLZF (LifeTech alexafluor 488 goat anti-rabbit) at 1:500. Primary incubation was overnight at 4°C, followed by three washes of PHEM with 0.25% Triton-X of 5 minutes each. Secondary incubation was 2 hours at 37°C followed by three washes as above. CD9 antibody was conjugated so only a single incubation overnight at 4°C. Hoesch staining (Sigma) was performed at 37°C for 8 minutes at a concentration of 1:800. Tra-1-81 (Stemgent, StainAlive) was used diluted to 5 μg/mL in SSC media for 30 minutes at 37°C in 5% CO_2 and washed twice with media.

iPSC and SSC colonies were stained with Tra-1-81 (an indicator of pluripotency) and DDX4 (a marker for spermatogonia), see Figs. 15 and 16. SSC colonies stained for DDX4 but had lost the pluripotency marker. To further confirm the SSC phenotype, more antibody assays were done with PLZF, another marker for spermatogonia, and CD9, a marker for stem cells, and the SSC colonies showed specific binding for all three, see Fig. 17.

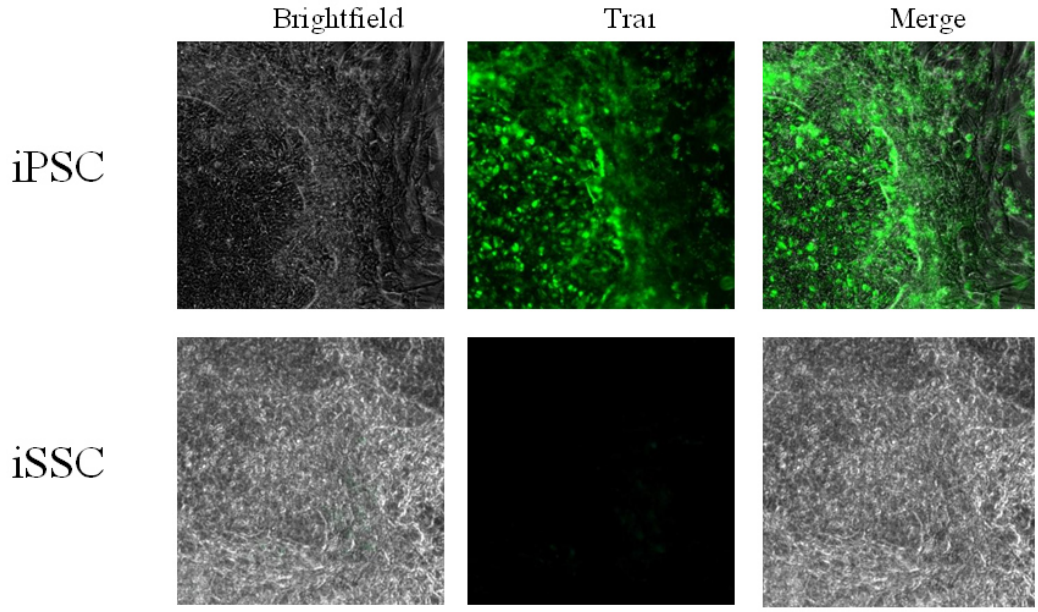


Fig. 15 - Tra-1-81 stains for pluripotency.
iPSC and differentiated SSC colonies (iSSC) following 10 days in culture in KSOR and SSC media respectively, stained with Tra-1-81 without fixing.

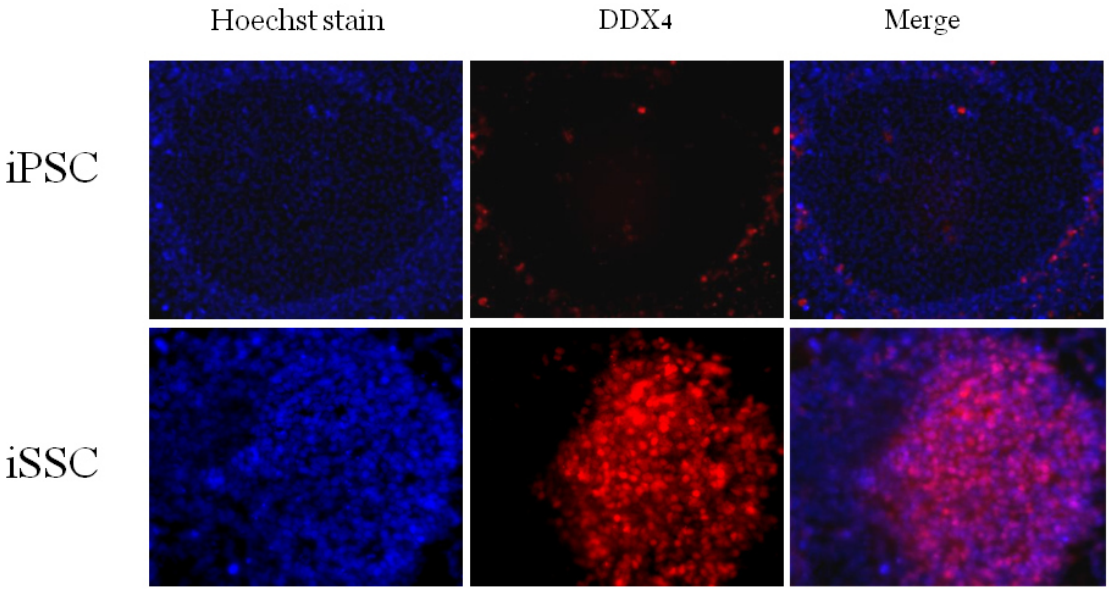


Fig. 16 - DDX4 stains for spermatogonia.
iPSC and differentiated SSC colonies were fixed in 4% paramaformaldehyde and stained for hoechst and DDX4.

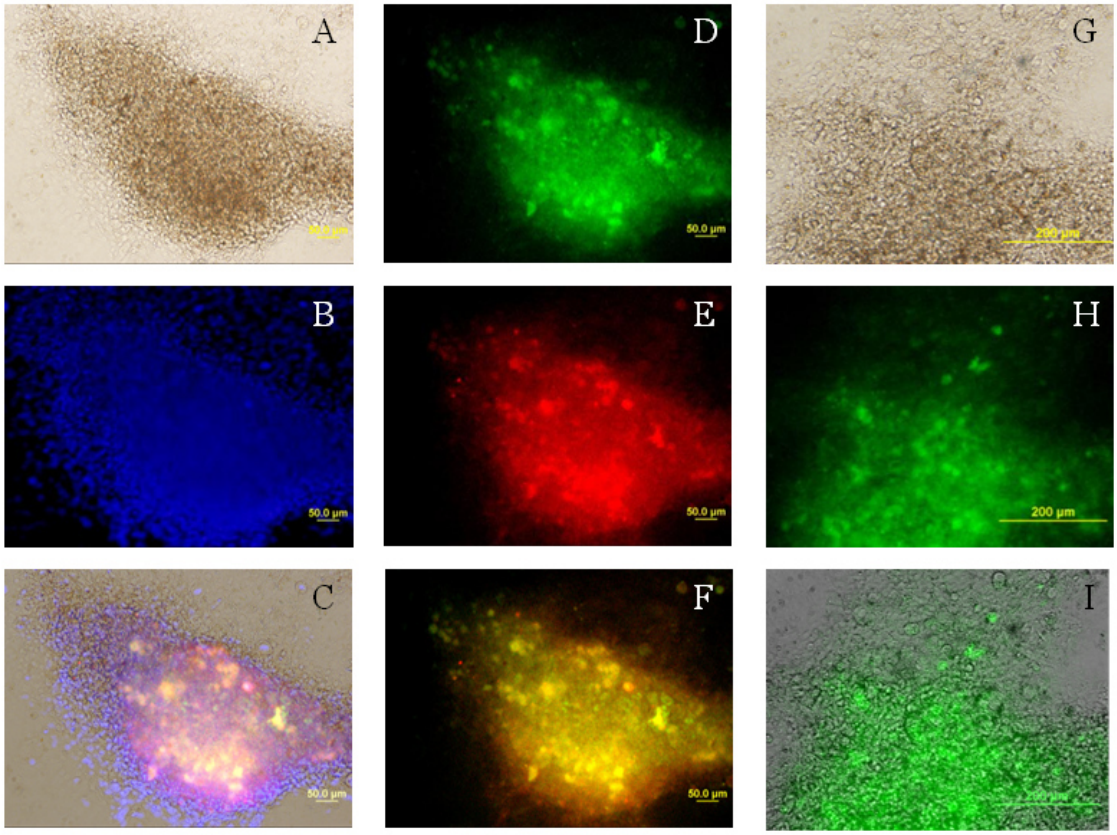


Fig. 17 - Staining of differentiated SSCs for spermatogonial markers
A-F Shown is a SSC colony, passage 11 after differentiation, 3 weeks post-transfection. **A.** Brightfield. **B.** Hoesch stain for DNA. **C.** Merge of images A-F. **D.** Anti-PLZF stain, **E.** Anti-DDX4 stain, **F.** merge of PLZF and DDX4. **G-I** Shown is a different SSC colony. **G.** Brightfield. **H.** Anti-CD9 stain **I.** Merge of H and I.

5.2.4 Plasmid Design and Transfection

Plasmids were constructed as follows: RET9 genes (wt and 2B mutant) supplied as kind gift by Lois Mulligan (Queen's University, Ontario, Canada) (Gujral et al. 2006). The cDNAs were supplied within the vector pcDNA3.1. These were excised by NotI/HindIII digestion and subsequently blunted, phosphatased and ligated into SmaI-digested pIRES-hrGFP-1a. The

orientation of insert into vector was determined by SacI digestion of recombinant clones. See Fig. 19.

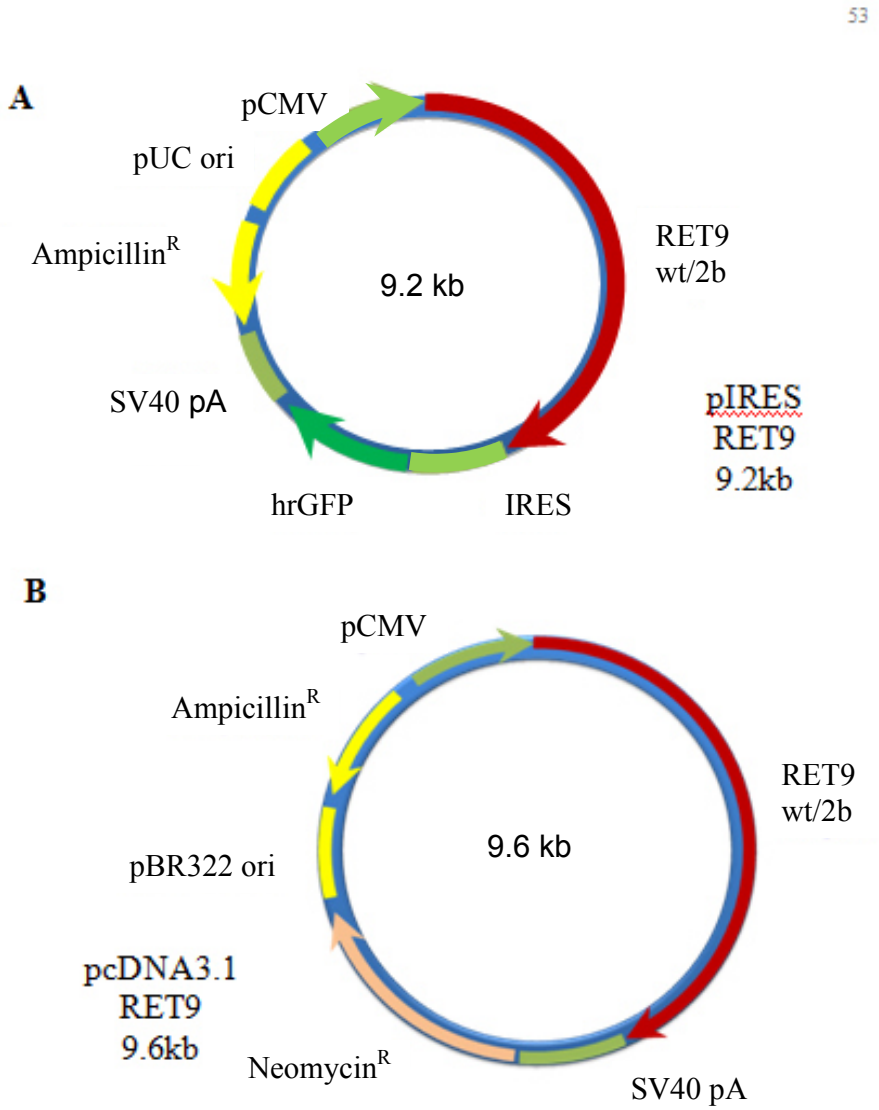


Fig. 18 - Map of the RET9 plasmids.
A) the pIRES hrGFP 1a RET9 plasmids, one containing the RET9 isoform of the wildtype RET gene and the other containing the MEN2B T2943C SNP change.
B) the pcDNA3.1 RET9 plasmids, again a pair of plasmids one with the wildtype RET gene and the other mutant.

To test the plasmids, they were first transfected into an easy-to-grow cell line. Mouse fibroblasts (3T3/NIH) were grown under standard cell culture conditions with both wt and mutant plasmids separately (described below). See figure 7. A proliferative advantage was confirmed for the mutant-expressing plasmid.

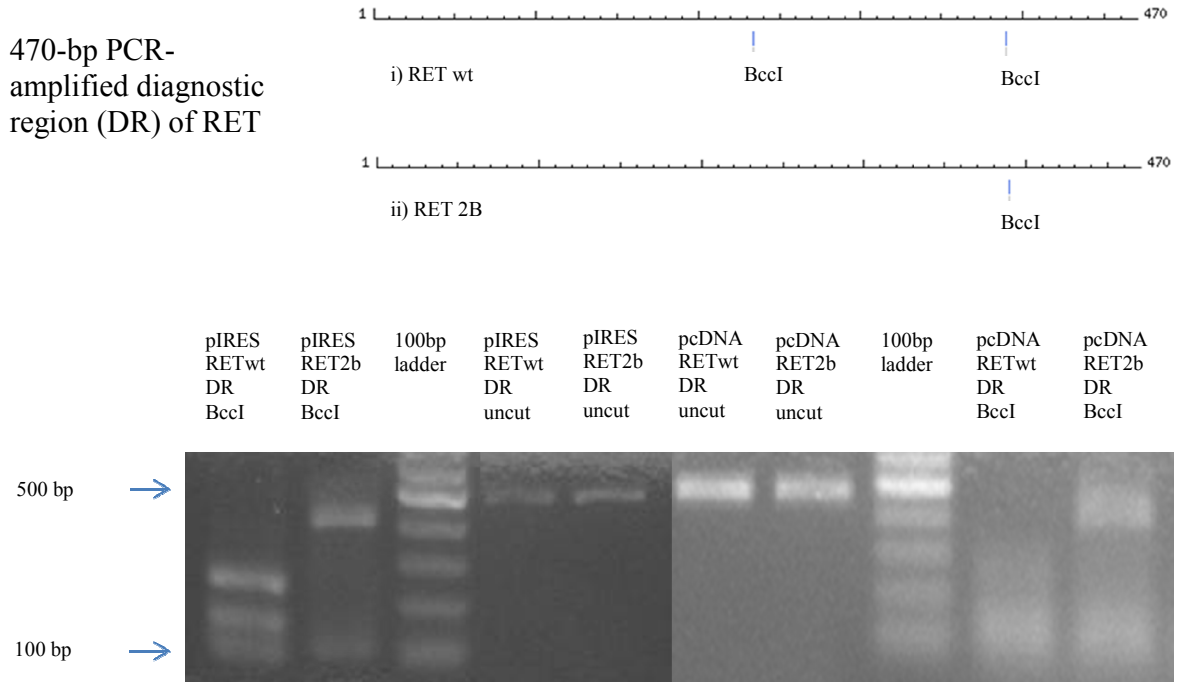


Fig. 19 - Diagnostic restriction digest of amplified region of RET.
The T2943C SNP mutation in the *RET* gene is contained within a *BccI* restriction diagnostic region (DR) amplified by PCR. Restriction using the *BccI* restriction endonuclease cleaves the wildtype gene at this sequence but not the mutant sequence. **(A)** diagram of *BccI* restriction sites within the SNP-containing amplified region. **(B)** gel shows the wildtype (wt) amplified region cleaved whereas the mutant sequence (2b) is not.

To linearlise the plasmids prior to transfection, plasmids were cut with SSPI restriction enzyme at 37°C for one hour and inactivated at 65°C for 20 minutes. Transfection was achieved

with Lipofectamine 3000 (Thermo Fisher Scientific) using 500ng of plasmid DNA in a 24 well plate and 1.5µL of P3000 reagent per well. Each sample was split into two halves – one received the wildtype human RET9 cDNA spliceform and the other received the M918 mutant. 4 days post-transfection, wells were cultured in SSC media supplemented with 500mg/ml G4180 Sulfate (Affymetrix) for 3 weeks to ensure only cells with linear integrated geneticin resistance gene were selected for (IACUC #12-028).

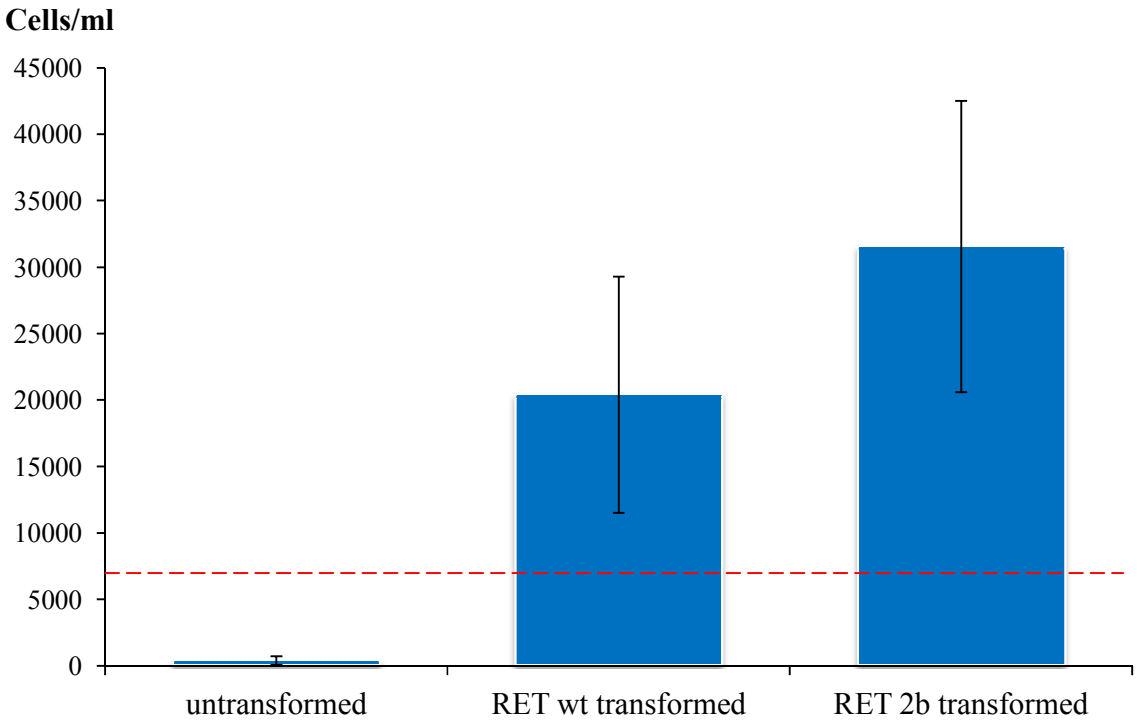


Fig. 20 - Effect of RET mutant on NIH/3T3 cell growth.
Preliminary data using NIH/3T3 cells transformed with pcDNA3.1/RET under G418 selection (200µg/ml). Significant difference between all three treatment groups (difference between wt & 2b *t*-test, p=0.03, n=9). Error bars show standard deviation. Dashed red line shows starting number before 8 days of growth (untransformed cells died due to G418).

5.2.5 Cell proliferation assay

Cells were seeded at a density of 10,000 cells/well in a 96-well plate. Cells were supplied with standard SSC media. Each treatment was seeded in triplicate and every 24 hours, three wells from each treatment were removed with trypsin and counted using a haemocytometer.

5.3 Results

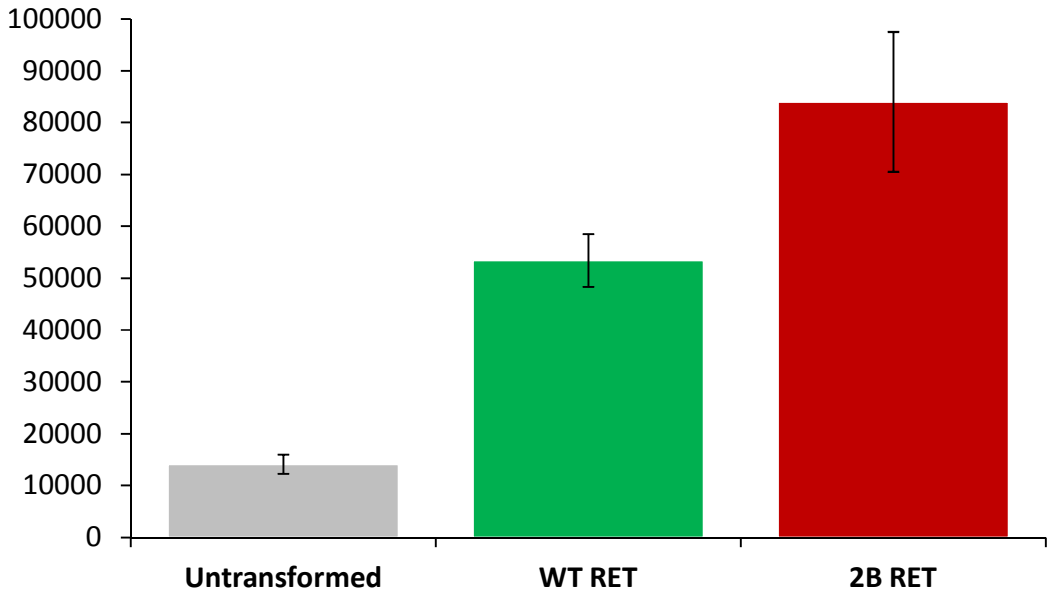


Fig. 21 - cell counts after 4 days of culture of transformed SSCs
Cultured in SSC media with 500 µg/ml G418. Error bars show standard deviation. Significant difference between all 3 groups (ANOVA, $p<0.05$)

The SSCs expressing the mutant showed a significant (repeated measures ANOVA, $p<0.05$) selective advantage in vitro (see Fig. 23). Both sets of cells grew until cells started to approach confluence and tapered off but the mutant cells not only grew faster but also appear to have a higher carrying capacity or tolerance for being at a higher density, which leads to

speculation that mutant cells may have the capability to exist within the niche in vivo at a higher density than wildtype cells.

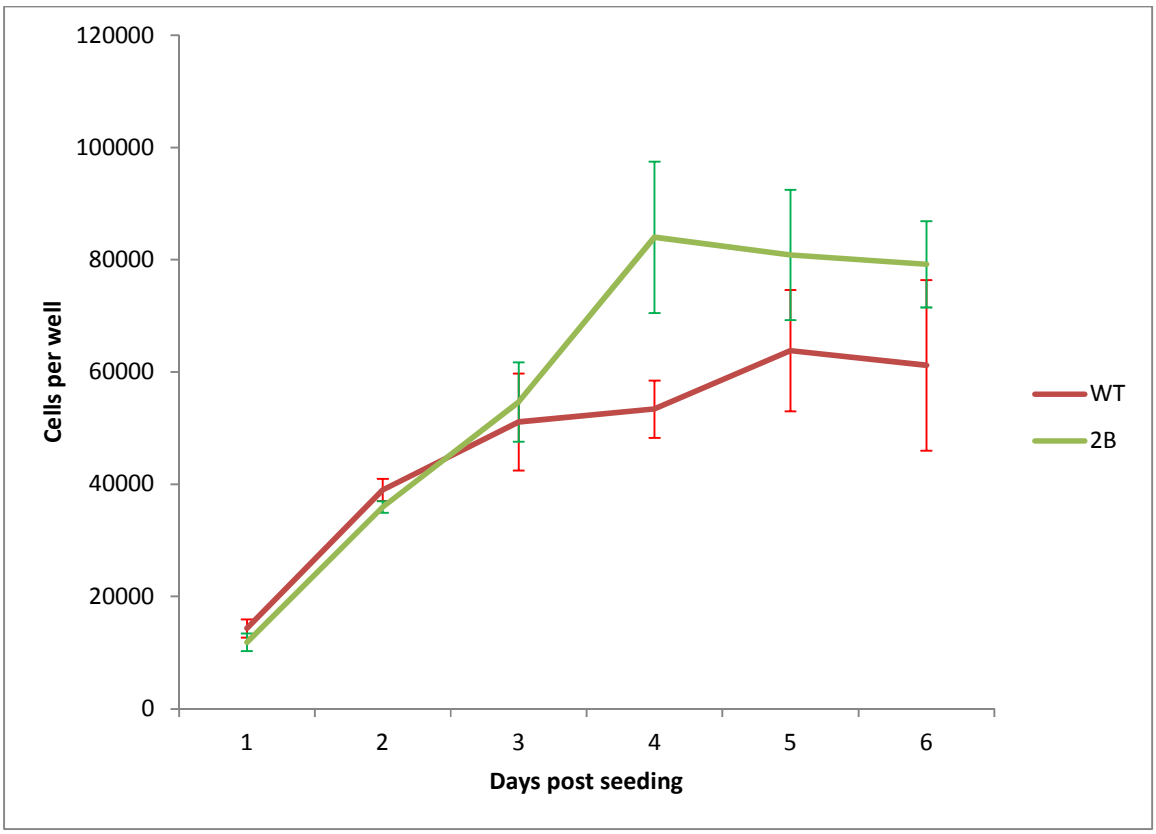


Fig. 22 - Cell growth of mutant and wildtype RET-carrying spermatogonial stem cells
Error bars show standard deviation. Populations show significant difference by day 4 ($p<0.05$, repeated measures ANOVA).

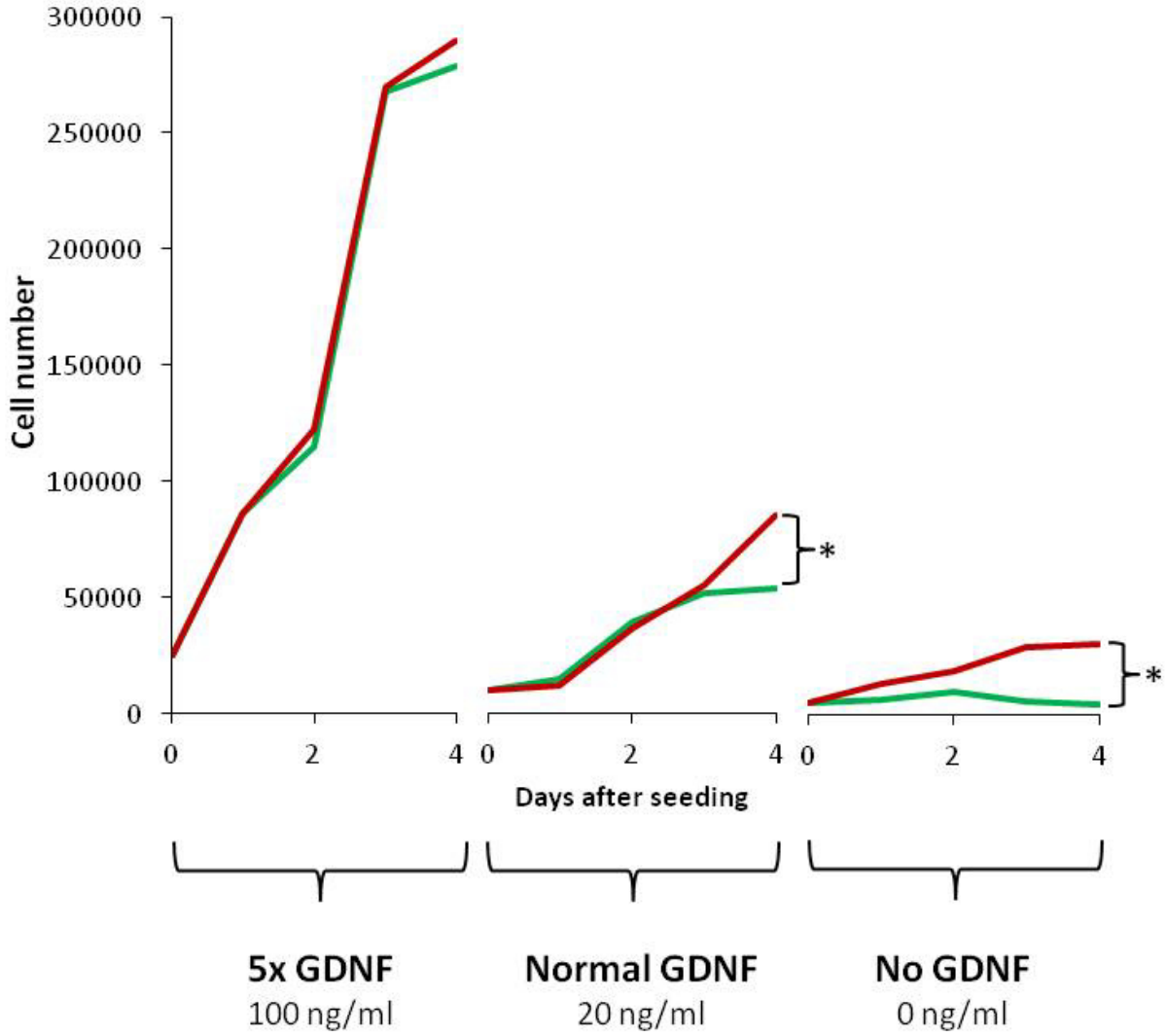


Fig. 23 - Growth of mutant and wildtype tranformed SSCs with varied GDNF
Higher exposure to GDNF promoted faster cell growth to both treatments. The difference between groups after 4 days of growth was more pronounced with lower GDNF levels. No significant difference at 100ng/ml, but mutant transfected populations had a significantly higher number of cells at lower concentrations (t-test, $p<0.05$).

5.4 Conclusions

The differentiation procedure transformed iPSCs to SSC-like cells no longer showing Tra-1-81, but become positive for CD9, PLZF and DDX4, markers consistent with spermatogonial stem cells. These results need to be treated with some caution: while they show differentiation into spermatogonia, without a functional assay to confirm stem cell status, there may be some differences between these cell populations and true spermatogonial stem cells. Future work will assay additional markers in order to confirm SSC identity. ID4 (Inhibitor of DNA expression 4) in particular appears to be a very promising marker for SSCs (Sun *et al.* 2015, Oatley *et al.* 2011).

This study experimentally confirmed that MEN2B mutations show a selective advantage for spermatogonia in culture. This effect is most likely due to constitutive downstream activation of the MAPK pathway (see Fig. 4), and further evidence is provided by the strength of the effect being dependent on the level of GDNF present (Fig. 23), as the receptor can be saturated by the ligand.

This chapter shows direct experimental evidence that MEN2B mutations underpin the strong paternal age effect. This provides mechanistic understanding of the cause of the high sporadic incidence of this disease and validate the theoretical positive selection model.

In terms of refining this line of inquiry, using the CRISPR/CAS9 gene editing system to directly edit the RET gene in SSCs would be an improvement as it would be under normal transcriptional control. Different genes related to other paternal age effect syndromes could be edited (FGFR2, FGFR3, etc). Moving to an *in vivo* system in mice would be the next logical step (discussed more in Chapter 6).

CHAPTER 6

CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

As we saw in Chapter 2, paternal age can have a modest effect on a range of substitutions and, over evolutionary time, males represent the source of the majority of substitutions that occur. However, for strong paternal age effect syndromes, pure copy-error accumulation presents an inadequate explanation for the observed incidence rates and for the distribution of mutations observed. This thesis was intended to mathematically model the SSC niche, to understand the accumulation of mutation and to experimentally show the effect of positive selection in cell culture.

In Chapter 3 we model normal stem cell homeostasis, which has some interesting comparisons for spermatogonial stem cell dynamics compared between species. The equations presented show the conditions required for stem cell homeostasis but allow for the spectrum between the drosophila germline stem cell system, which has classical asymmetric cell division that are in constant use and mammalian systems where divisions are likely comprised of symmetrical divisions with a subsection of cells quiescent. Both systems represent solutions to the equations that arise from simple initial starting conditions.

In order to incorporate mutation accumulation, we used a Markov chain to model the probabilities of mutation and positive selection with cell divisions (Chapter 4). The proportions of mutant to wildtype cells with increasing paternal age was used to generate a simulated population and observed/expected curves. The parameter for the probability of positive selection per cell division was estimated by fitting the simulated population model to available data on

disease incidence (observed/expected) and also the direct mutation estimates to sequencing assays of sperm donors. Additionally the simulated populations was used to generate a predicted incidence rates overall of the disorders with given mutation and positive selection values. Strength of selective advantage is presented for a range of disorders including Apert's syndrome and achondroplasia and showed significant agreement between molecular and incidence data. Incidence of the diseases was predicted closely for most disorders and was heavily influenced by the site-specific mutation rate caused by hypermutable CpG sites and the number of mutable alleles. Both positive selection and the rate of copy-error mutations are important in adequately explaining the paternal age effect. For the two mutations that cause Apert's syndrome the difference in the incidence rates of the two alleles is accounted for by the mutation rate alone and both mutations have a similar positive effect. On the other hand, comparison of the mutations in the Fibroblast Growth Factor Receptor 3 gene that cause achondroplasia and thanatophoric dysplasia shows a stronger selective advantage for more strongly activating gain-of-function mutation, as predicted. Overall, the model shows good agreement with existing data on paternal age effect syndromes. It provides a mathematical understanding of the accumulation of disease-causing mutations and provides evidence that disorders strongly associated with paternal age are linked by positive selective advantage.

SSCs were created using induced pluripotent stem cells derived from BJ fibroblasts and differentiating them with SSC media including the growth factors GDNF, bFGF and GFRa1 over a period of 10 days. SSC identity compared with the iPSCs was confirmed by immunohistochemistry with the SSCs staining positive for CD9, PLZF and DDX4 and negative for the pluripotency marker Tra-1-81. Wildtype and mutant SSCs were generated by transfection with a pcDNA3.1 plasmid containing the normal RET gene and the gene containing the M918T

mutation, respectively. Plasmid integration was achieved through treatment with gentamicin and a restriction-enzyme assay for the mutation was developed. Mutant SSCs showed increased proliferation in culture. This effect was magnified by decreasing GDNF concentration in culture and reduced when the GDNF concentration was increased, saturating the receptors. This research demonstrated experimental evidence for the positive selection effect of the M918T mutation in RET and validating the hypothesis that this selection effect is to blame for the marked paternal age effect of multiple endocrine neoplasia 2B. This research is the first direct evidence that specific gain of function mutations can positively select for stem cells that produce sperm that carry the devastating disease alleles.

6.2 Implications

The phenomenon of positive selection is one that is a cornerstone of evolution and can be found throughout nature. For higher organisms, a component of their cells gaining a positive selective advantage over others must be resisted, for fear of cancer and other deregulated cell growth. Understanding how deleterious selective mechanisms occur and are counteracted is vital in understanding human mutational load in somatic and germline tissues.

An observation from the experiment in Chapter 5 is that when the SSCs have unrestricted growth, there is no significant difference between wildtype and mutant cells. This suggests that the cells may have less of a response to space limitations and/or antigrowth factors from other cells. It would be informative to test the growth advantages of the mutant vs wildtype in circumstances of low and high cell density and also if both populations can be marked with different fluorescent tags, to coculture together and observe competition between the two groups of cells directly.

There are also implications for technologies designed to culture a patient's cells and then differentiate them into tissues. In particular, culture of human cells (SSCs and iPSCs) with the intention of using them to recover spermatogenesis in patients that have undergone gonadotoxic therapies (Poirot & Schubert 2011) may very well be susceptible to this sort of positive selection *in vitro*. As divisions occur in culture, culture flasks may be overtaken by rare mutants. Should these cells be used to recover spermatogenesis, these mutations would be passed on to offspring.

The implications of this research is important both for understanding the normal regulation of the stem cell niche, the mechanisms that surround the most extreme cases of paternally derived genetic disease and the role that delayed reproduction has on the incidence of *de novo* inherited disorders.

6.3 Future directions

The model presented in Chapter 2 does not take into accounts the specifics of feedback mechanisms. A continuation of this work could look specifically at the feedback loops of the spermatogonial niche and spatial arrangement. A model or simulation of the SSC niche incorporating the signalling and feedback could be used to confirm the boundary conditions for homeostasis in Chapter 2.

Future directions will expand the arena of research on this subject to other genes involved in the same pathways that all share a paternal age effect pattern of disease incidence as well as an *in vivo* model involving injecting the plasmids into mouse testes with a marker gene and observing the kinetics of the stem cells within the normal cell niche.

Electroporation of the whole testis could be used to provide random integration of the plasmid into a subset of the native SSCs. Alternatively, removal and culture of SSCs from one

testis could be edited with CRISPR/Cas9 and then marked with a reporter gene and reinjected into the mouse's other testis to do lineage tracing of the injected cells. The mathematical expressions of Chapter 4 could be confirmed with more precise estimation of the positive selection advantage numerically *in vivo*.

REFERENCES

Agulnik AI, Bishop CE, Lerner JL, Agulnik SI, Solovyev VV. (1997) Analysis of mutation rates in the SMCY/SMCX genes shows that mammalian evolution is male driven. *Mamm Genome.* **8**:134-8.

van Alphen MM, van de Kant HJ, de Rooij DG. (1988) Repopulation of the seminiferous epithelium of the rhesus monkey after X irradiation. *Radiat Res.* **3**:487-500.

Anagnostopoulos T, Green PM, Rowley, G, Lewis, CM, Giannelli, F (1999). DNA Variation in a 5-Mb Region of the X Chromosome and Estimates of Sex-Specific/Type-Specific Mutation Rates. *Am J Hum Genet* **64**:508-517

Antonarakis, SE, Krawczak, M, Cooper, DN (2000). Disease-causing mutations in the human genome. *Eur J Pediatr* **159**:S173-8

Aoki R, Shoshkes-Carmel M, Gao N, Shin S, May CL, Golson ML, Zahm AM, Ray M, Wiser CL, Wright CV, Kaestner KH. (2016) Foxl1-expressing mesenchymal cells constitute the intestinal stem cell niche. *Cell Mol Gastroenterol Hepatol.* **2**(2):175-188.

Arnheim N, Calabrese P. Understanding what determines the frequency and pattern of human germline mutations. (2009) *Nat Rev Genet*.; **10**(7):478-88

Ashley-Koch AE, Robinson H, Glicksman AE, Nolin SL, Schwartz CE, Brown WT, Turner G, Sherman SL. (1998) Examination of factors associated with instability of the FMR1 CGG repeat. *Am J Hum Genet* **63**:776-85.

Axelsson E, Smith NG, Sundström H, Berlin S, Ellegren H. (2004) Male-biased mutation rate and divergence in autosomal, z-linked and w-linked introns of chicken and Turkey. *Mol Biol Evol* **21**:1538-47.

Babenko VN, Basu MK, Kondrashov FA, Rogozin IB, Koonin EV (2006) Signs of positive selection of somatic mutations in human cancers detected by EST sequence analysis. *BMC Cancer* **6**:36.

Barroca V, Lassalle B, Coureuil M, Louis JP, Le Page F, Testart J, Allemand I, Riou L, Fouchet P. (2009) Mouse differentiating spermatogonia can generate germinal stem cells in vivo. *Nat Cell Biol*; **11**(2):190-6.

Becker J, Schwaab R, Moller-Taube A, Schwaab U, Schmidt W, Brackmann HH, Grimm T, Olek K, Oldenburg J. (1996) Characterization of the factor VIII defect in 147 patients with sporadic hemophilia A: family studies indicate a mutation type-dependent sex ratio of mutation frequencies. *Am J Hum Genet* **58**:657-70.

Bellus GA, Hefferon TW, Ortiz de Luna RI, Hecht JT, Horton WA, Machado M, Kaitila I, McIntosh I, Francomano CA. (1995) Achondroplasia is defined by recurrent G380R mutations of FGFR3. *Am J Hum Genet* **56**:368-73.

Biermann K, Steger K (2007) Eipgenetics in male germ cells. *J Androl* **28**:466-80

Bird A (1999) DNA Methylation de Novo. *Science* **286**:2287-2288.

Blank CE. (1960) Apert's syndrome (a type of acrocephalosyndactyly)-observations on a British series of thirty-nine cases. *Ann Hum Genet.* **24**:151-64.

Böhm J, Munk-Schulenburg S, Felscher S, Kohlhase J. 2006. SALL1 mutations in sporadic Townes-Brocks syndrome are of predominantly paternal origin without obvious paternal age effect. *Am J Med Genet A* **140**:1904-8.

Bohossian HB, Skaletsky H, Page DC. (2000) Unexpectedly similar rates of nucleotide substitution found in male and female hominids. *Nature* **406**:622-5.

Bonaventure J, Horne WC, Baron R. (2007) The localization of FGFR3 mutations causing thanatophoric dysplasia type I differentially affects phosphorylation, processing and ubiquitylation of the receptor. *FEBS J.* **274**(12):3078-93.

Bourc'his D, Bestor, TH. (2006) Origins of extreme sexual dimorphism in genomic imprinting. *Cytogenet Genome Res* **113**:36-40.

Bozic I, Nowak MA. 2013 Cancer. Unwanted evolution. *Science.* **22**;342(6161):938-9

Bray I, Gunnell D, Davey Smith G. (2006) Advanced paternal age: how old is too old? *J Epidemiol Community Health.* **60**(10):851-3.

Brero A, Leonhardt H, Cardoso. (2006) Replication and Translation of Epigenetic Information. *Curr Top Microbiol Immunol* **301**:24-44

Brinkmann B, Klintschar M, Neuhuber F, Huhne J, Rolf B. (1998) Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *Am J Hum Genet* **62**:1408-15.

Cadiñanos J, Varela I, López-Otín C, Freije JM. (2005) From immature lamin to premature aging: molecular pathways and therapeutic opportunities. *Cell Cycle* **4**:1732-5

Cao H, Hegele RA. (2003) LMNA is mutated in Hutchinson-Gilford progeria (MIM 176670) but not in Wiedemann-Rautenstrauch progeroid syndrome (MIM 264090). *J Hum Genet* **48**:271-4.

Carlson KM, Bracamontes J, Jackson CE, Clark R, Lacroix A, Wells SA Jr, Goodfellow PJ. (1994) Parent-of-origin effects in multiple endocrine neoplasia type 2B. *Am J Hum Genet* **55**:1076-82.

Chang BH, Hewett-Emmett D, Li WH. (1995) Male-to-female ratios of mutation rate in higher primates estimated from intron sequences. *Zool Studies* **35**:36-48

Chaudhry SS, Cain SA, Morgan A, Dallas SL, Shuttleworth CA, Kielty CM. (2007) Fibrillin-1 regulates the bioavailability of TGFβ1. *J Cell Biol* **176**:355-67.

Chen C, Fingerhut JM, Yamashita YM. (2016) The ins(ide) and outs(ide) of asymmetric stem cell division. *Curr Opin Cell Biol.* **43**:1-6.

Choi SK, Yoon SR, Calabrese P, Arnheim N. (2008) A germ-line-selective advantage rather than an increased mutation rate can explain some unexpectedly common human disease mutations. *Proc Natl Acad Sci U S A.* **105**: 10143–10148.

Colijn C and Mackey CM. (2005) A mathematical model of hematopoiesis – i. periodic chronic myelogenous leukemia. *Journal of Theoretical Biology*, **237**:117–132.

Connor, J. M., Connor, R. A. C., Sweet, E. M., Gibson, A. A. M., Patrick, W. J. A., McNay, M. B., Redford, D. H. A. (1985) Lethal neonatal chondrodysplasias in the West of Scotland, 1970-1983, with a description of a thanatophoric, dysplasialike, autosomal recessive disorder, Glasgow variant. *Am. J. Med. Genet.* **22**: 243-253.

Cooper DN, Krawczak M. (1990) The mutational spectrum of single base-pair substitutions causing human genetic disease: patterns and predictions. *Hum Genet* **85**:55-74.

Crow JF. (2000) The origins, patterns and implications of human spontaneous mutation. *Nat Rev Genet* **1**:40-47.

Dakouane Giudicelli M, Serazin V, Le Sciellour CR, Albert M, Selva J, Giudicelli Y (2007) Increased achondroplasia mutation frequency with advanced age and evidence for G1138A mosaicism in human testis biopsies. *Fertil Steril.* **89**(6):1651-6.

Dhup S, Majumdar SS. (2008) Transgenesis via permanent integration of genes in repopulating spermatogonial cells in vivo. *Nat Methods.* **5**(7):601-3.

Dohle GR, Smit M, Weber RF. (2003) Androgens and male fertility. *World J Urol* **21**:341-5.

Dorner D, Vlcek S, Foeger N, Gajewski A, Makolm C, Gotzmann J, Hutchison CJ, Foisner R. (2006) Lamina-associated polypeptide 2alpha regulates cell cycle progression and differentiation via the retinoblastoma-E2F pathway. *J Cell Biol* **173**:83-93.

Drake JW, Charlesworth B, Charlesworth D, Crow JF. (1998) Rates of spontaneous mutation. *Genetics* **148**:1667-1686.

Dryja TP, Mukai S, Petersen R, Rapaport JM, Walton D, Yandell DW. (1989) Parental origin of mutations of the retinoblastoma gene. *Nature* **339**:556-8.

Easley CA 4th, Phillips BT, McGuire MM, Barringer JM, Valli H, Hermann BP, Simerly CR, Rajkovic A, Miki T, Orwig KE, Schatten GP. (2012) Direct differentiation of human pluripotent stem cells into haploid spermatogenic cells. *Cell Rep.* **2**(3):440-6.

Ebersberger I, Metzler D, Schwarz C, Paabo S. (2002) Genomewide comparison of DNA sequences between humans and chimpanzees. *Am J Hum Genet* **70**:1490-1497.

Ellegren H. (2007) Characteristics, causes and evolutionary consequences of male-biased mutation. *Proc Biol Sci* **274**:1-10.

Ellegren H. (2000) Microsatellite mutations in the germline: implications for evolutionary inference. *Trends Genet* **16**:551-558.

Ellegren H, Fridolfsson AK. (2003) Sex-specific mutation rates in salmonid fish. *J Mol Evol* **56**:458-63.

El-Maarri O, Olek A, Balaban B, Montag M, van der Ven H, Urman B, Olek K, Caglayan SH, Walter J, Oldenburg J. (1998) Methylation levels at selected CpG sites in the factor VIII and FGFR3 genes, in mature female and male germ cells: implications for male-driven evolution. *Am J Hum Genet* **63**:1001-1008

Eriksson M, Brown WT, Gordon LB, Glynn MW, Singer J, Scott L, Erdos MR, Robbins CM, Moses TY, Berglund P, Dutra A, Pak E, Durkin S, Csoka AB, Boehnke M, Glover TW, Collins FS. (2003) Recurrent de novo point mutations in lamin A cause Hutchinson-Gilford progeria syndrome. *Nature* **423**:293-298.

Erlandsson R, Wilson JF, Paabo S. (2000) Sex chromosomal transposable element accumulation and male-driven substitutional evolution in humans. *Mol Biol Evol* **17**:804-12.

Evans-Galea MV, Hannan AJ, Carrodus N, Delatycki MB, Saffery R. (2013) Epigenetic modifications in trinucleotide repeat diseases. Trends Mol *Med*. **19**(11):655-63

Faruqi T, Dhawan N, Bahl J, Gupta V, Vohra S, Tu K, Abdelmagid SM. (2014) Molecular, phenotypic aspects and therapeutic horizons of rare genetic bone disorders. *Biomed Res Int.* **2014**:670842.

Farrer LA, Cupples LA, Kiely DK, Conneally PM, Myers RH. (1992) Inverse relationship between age at onset of Huntington disease and paternal age suggests involvement of genetic imprinting. *Am J Hum Genet* **50**:528-35.

Fonseka KG, Griffin DK. (2011) Is there a paternal age effect for aneuploidy? *Cytogenet Genome Res.* **133**(2-4):280-91.

Francke U, Felsenstein J, Gartler SM, Migeon BR, Dancis J, Seegmiller JE, Bakay F, Nyhan WL. (1976) The occurrence of new mutants in the X-linked recessive Lesch-Nyhan disease. *Am J Hum Genet* **28**:123-37.

Frank-Raue K, Raue F. (2009) Multiple endocrine neoplasia type 2 (MEN 2). *Eur J Cancer*. **45** Suppl 1:267-73.

Fryxell KJ1, Moon WJ. (2005) CpG mutation rates in the human genome are highly dependent on local GC content. *Mol Biol Evol.* **22**(3):650-8.

Fusaki, N., Ban, H., Nishiyama, A., Saeki, K., & Hasegawa, M. (2009). Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome. *Proceedings of the Japan Academy. Series B, Physical and Biological Sciences,* **85**(8), 348–362.

Gavrilov LA, Gavrilova NS, Kroutko VN, Evdokushkina GN, Semyonova VG, Gavrilova AL, Lapshin EV, Evdokushkina NN, Kushnareva YE. (1997) Mutation load and human longevity. *Mutat Res* **377**:61-2.

Gelb BD, Tartaglia M. (2006) Noonan syndrome and related disorders: dysregulated RAS-mitogen activated protein kinase signal transduction. *Hum Mol Genet* **15**:R220-R226

Giannoulatou E, McVean G, Taylor IB, McGowan SJ, Maher GJ, Iqbal Z, Pfeifer SP, Turner I, Burkitt Wright EM, Shorto J, Itani A, Turner K, Gregory L, Buck D, Rajpert-De Meyts E, Looijenga LH, Kerr B, Wilkie AO, Goriely A. (2013) Contributions of intrinsic mutation rate and selfish selection to levels of de novo HRAS mutations in the paternal germline. *Proc Natl Acad Sci U S A*. **110**(50):20152-7.

Glaser RL, Broman KW, Schulman RL, Eskenazi B, Wyrobek AJ, Jabs EW. (2003) The Paternal-Age Effect in Apert Syndrome Is Due, In Part, to the Increased Frequency of Mutations in Sperm. *Am J Hum Genet* **73**:939-947

Glaser RL, Jabs EW. (2004). Dear old dad: paternal age and the origin of spontaneous mutations in humans [review]. *Sci Aging Knowledge Envion* **3**:1-11

González-González E, López-Casas PP, Del Mazo J. (2008) Gene silencing by RNAi in mouse Sertoli cells. *Reprod Biol Endocrinol*. **6**:29.

Goriely A, Wilkie AO. (2012) Paternal age effect mutations and selfish spermatogonial selection: causes and consequences for human disease. *Am J Hum Genet.* **90**(2):175-200.

Goriely A, Hansen RM, Taylor IB, Olesen IA, Jacobsen GK, McGowan SJ, Pfeifer SP, McVean GA, Rajpert-De Meyts E, Wilkie AO. (2009) Activating mutations in FGFR3 and HRAS reveal a shared genetic origin for congenital disorders and testicular tumors. *Nat Genet.* **41**(11):1247-52.

Goriely A, McVean GAT, van Pelt AMM, O'Rourke AW, Wall SA, de Rooij DG, Wilkie AOM. (2005) Gain-of-function amino acid substitutions drive positive selection of FGFR2 mutations in human spermatogonia. *Proc Natl Acad Sci U S A* **102**:6051-6056

Goriely A, McVean GA, Rojmyr M, Ingemarsson B, Wilkie AO (2003) Evidence for selective advantage of pathogenic FGFR2 mutations in the male germ line. *Science* **301**:643-6.

Guan K, Wolf F, Becker A, Engel W, Nayernia K, Hasenfuss G. (2009) Isolation and cultivation of stem cells from adult mouse testes. *Nat Protoc*. **4**(2):143-54.

Gujral TS, Singh VK, Jia Z, Mulligan LM. (2006) Molecular mechanisms of RET receptor-mediated oncogenesis in multiple endocrine neoplasia 2B. *Cancer Res.* **66**(22):10741-9.

Haldane JBS. (1935) The rate of spontaneous mutation of a human gene. *J Genet* **31**:317-326.

Haldane JBS. (1947) The mutation rate of the gene for haemophilia and its segregation ratios in males and females. *Ann Eugen* **13**:262-271.

Hamra FK, Chapman KM, Wu Z, Garbers DL. (2008) Isolating highly pure rat spermatogonial stem cells in culture. *Methods Mol Biol.* **450**:163-79.

Hanahan D, Weinberg RA. (2000) The hallmarks of cancer. *Cell* **100**:57-70.

Helgason A, Einarsson AW, Guðmundsdóttir VB, Sigurðsson Á, Gunnarsdóttir ED, Jagadeesan A, Ebenesersdóttir SS, Kong A, Stefánsson K. (2015) The Y-chromosome point mutation rate in humans. *Nat Genet.* **47**(5):453-7.

Helman SN, Badhey A, Kadakia S, Myers E. (2014) Revisiting Crouzon syndrome: reviewing the background and management of a multifaceted disease. *Oral Maxillofac Surg.* **18**(4):373-9.

Hill KA, Buettner VL, Halangoda A, Kunishige M, Moore SR, Longmate J, Scaringe WA, Sommer SS. (2004) Spontaneous mutation in Big Blue mice from fetus to old age: tissue-specific time courses of mutation frequency but similar mutation types. *Environ Mol Mutagen* **43**:110-20.

Hilton MJ, Gutierrez L, Martinez DA, Wells DE. (2005) EXT1 regulates chondrocyte proliferation and differentiation during endochondral bone development. *Bone* **36**:379-86.

Hofmann MC. (2008) Gdnf signaling pathways within the mammalian spermatogonial stem cell niche. *Mol Cell Endocrinol.* **288**(1-2):95-103.

Hook EB. (1981) Rates of chromosome abnormalities at different maternal ages. *Obstet Gynecol.* **58**(3):282-5.

Huamani J, McMahan CA, Herbert DC, Reddick R, McCarrey JR, MacInnes MI, Chen DJ, Walter CA. (2004) Spontaneous mutagenesis is enhanced in Apex heterozygous mice. *Mol Cell Biol* **24**:8145-53.

Huang W, Chang BH, Gu X, Hewett-Emmett D, Li W. (1997) Sex differences in mutation rate in higher primates estimated from AMG intron sequences. *J Mol Evol* **44**:463-5.

Hüning I, Gillessen-Kaesbach G. (2014) Fibrodysplasia ossificans progressiva: clinical course, genetic mutations and genotype-phenotype correlation. *Mol Syndromol.* **5**(5):201-11.

Humphries A, Wright NA. (2008) Colonic crypt organization and tumorigenesis. *Nat Rev Cancer.* **8**(6):415-24.

Hurst LD, Ellegren H. (2002) Mystery of the mutagenic male. *Nature* **420**:365-366.

Hurst LD, Ellegren H. (1998) Sex biases in the mutation rate. *Trends Genet* **14**:446-52.

Huttley GA, Jakobsen IB, Wilson SR, Easteal S. (2000) How important is DNA replication for mutagenesis? *Mol Biol Evol* **17**:929-37.

Impellizzeri KJ, Anderson B, Burgers PM. (1991) The spectrum of spontaneous mutations in a Saccharomyces cerevisiae uracil-DNA-glycosylase mutant limits the function of this enzyme to cytosine deamination repair. *J Bacteriol* **173**:6807-6810.

Intano GW, McMahan CA, Walter RB, McCarrey JR, Walter CA. (2001) Mixed spermatogenic germ cell nuclear extracts exhibit high base excision repair activity. *Nucleic Acids Res* **29**:1366-72.

Jadayel D, Fain P, Upadhyaya M, Ponder MA, Huson SM, Carey J, Fryer A, Mathew CG, Barker DF, Ponder BA. (1990) Paternal origin of new mutations in von Recklinghausen neurofibromatosis. *Nature* **343**:558-9.

Johnston MD, Edwards CM, Bodmer WF, Maini PK, and Chapman SJ. (2007) Examples of mathematical modelling – tales from the crypt. *Cell Cycle*, 6(17):2106–2112,

Jung A, Schuppe HC, Schill, WB. (2003) Are children of older fathers at risk for genetic disorders? *Andrologia* **35**:191-199

Kanatsu-Shinohara M, Miki H, Inoue K, Ogonuki N, Toyokuni S, Ogura A, Shinohara T. (2005) Long-term culture of mouse male germline stem cells under serum-or feeder-free conditions. *Biol Reprod.* **72**(4):985-91.

Katoh Y, Katoh M. (2006) Hedgehog signaling pathway and gastrointestinal stem cell signaling network (review). *Int J Mol Med* **18**:1019-23.

Kelleher FC, Fennelly D, Rafferty M. (2006) Common critical pathways in embryogenesis and cancer. *Acta Oncol* **45**:375-88.

Klein AM, Nakagawa T, Ichikawa R, Yoshida S, Simons BD. (2010) Mouse germ line stem cells undergo rapid and stochastic turnover. *Cell Stem Cell.* **6**;7(2):214-24.

Krawczak M, Cooper DN. (1997) The Human Gene Mutation Database. *Trends Genet* **13**:121-122.

Krawczak M, Ball EV, Cooper DN. (1998) Neighboring-Nucleotide Effects on the Rates of Germ-Line Single-Base-Pair Substitution in Human Genes. *Am J Hum Genet* **63**:474-488

Kühnert B, Nieschlag E. (2004) Reproductive functions of the ageing male. *Hum Reprod Update* **10**:327-39.

Lander AD, Gokoffski KK, Wan FYM, Nie Q, and Calof A. Cell lineages and the logic of proliferative control. *PLoS Biol*, 7(1), 2009.

Li R, Johnson AB, Salomons GS, van der Knaap MS, Rodriguez D, Boespflug-Tanguy O, Gorospe JR, Goldman JE, Messing A, Brenner M. (2006) Propensity for paternal inheritance of de novo mutations in Alexander disease. *Hum Genet* **119**:137-44.

Li WH, Yi S, Makova K. (2002) Male-driven evolution. *Curr Opin Genet Dev* **12**:650-656.

Li WH, Ellsworth DL, Krushkal J, Chang BH, Hewett-Emmett D. (1996) Rates of nucleotide substitution in primates and rodents and the generation-time effect hypothesis. *Mol Phylogenet Evol.* **5**(1):182-7.

Lieb M, Rehmat S. (1997) 5-Methylcytosine is not a mutation hot spot in nondividing Escherichia coli. *Proc Natl Acad Sci U S A* **94**:940-945.

Lindahl T. (1993) Instability and decay of the primary structure of DNA. *Nature* **362**:709-715.

Long TA, Pischedda A. (2005) Do female Drosophila melanogaster adaptively bias offspring sex ratios in relation to the age of their mate? *Proc Biol Sci* **272**:1781-7.

Lowe X, Eskenazi B, Nelson DO, Kidd S, Alme A, Wyrobek AJ. (2001) Frequency of XY sperm increases with age in fathers of boys with Klinefelter syndrome. *Am J Hum Genet* **69**:1046-54.

Lurie IW. (1994) Genetics of the Costello syndrome. *Am J Med Genet.* **52**(3):358-9.

Lynas MA. (1958) Marfan's syndrome in Northern Ireland; an account of thirteen families. *Ann Hum Genet.* **22**(4):289-309.

Lynch M. Rate, molecular spectrum, and consequences of human mutation. (2010) *Proc Natl Acad Sci U S A*. **107**(3):961-8.

Maher GJ, Goriely A, Wilkie AO. (2014) Cellular evidence for selfish spermatogonial selection in aged human testes. *Andrology*. **2**(3):304-14.

Makova KD, Li WH. (2002) Strong male-driven evolution of DNA sequences in humans and apes. *Nature* **416**:624-626.

Malaspina D, Corcoran C, Fahim C, Berman A, Harkavy-Friedman J, Yale S, Goetz D, Goetz R, Harlap S, Gorman J. (2002) Paternal age and sporadic schizophrenia: evidence for de novo mutations. *Am J Med Genet* **114**:299-303.

Malaspina D, Reichenberg A, Weiser M, Fennig S, Davidson M, Harlap S, Wolitzky R, Rabinowitz J, Susser E, Knobler HY. (2005) Paternal age and intelligence: implications for age-related genomic changes in male germ cells. *Psychiatr Genet* **15**:117-25.

Meng X, Lindahl M, Hyvönen ME, Parvinen M, de Rooij DG, Hess MW, Raatikainen-Ahokas A, Sainio K, Rauvala H, Lakso M, Pichel JG, Westphal H, Saarma M, Sariola H. (2000)

Regulation of cell fate decision of undifferentiated spermatogonia by GDNF. *Science.* **287**:1489-93.

Miyata T, Hayashida H, Kuma K, Mitsuyasu K, Yasunaga T. (1987) Male-driven molecular evolution: a model and nucleotide sequence analysis. *Cold Spring Harb Symp Quant Biol* **52**:863-7.

Mologni L, Sala E, Cazzaniga S, Rostagno R, Kuoni T, Puttini M, Bain J, Cleris L, Redaelli S, Riva B, Formelli F, Scapozza L, Gambacorti-Passerini C. (2006) Inhibition of RET tyrosine kinase by SU5416. *J Mol Endocrinol.* **37**(2):199-212.

Montgomery SM, Lambe M, Olsson T, Ekbom A (2004) Parental age, family size, and risk of multiple sclerosis. *Epidemiology* **15**:717-23.

Morgan HD, Dean W, Coker HA, Reik W, Petersen-Mahrt SK (2004) Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. *J Biol Chem* **279**:52353-52360.

Mugal CF, Ellegren H. (2011) Substitution rate variation at human CpG sites correlates with non-CpG divergence, methylation level and GC content. *Genome Biol.* **12**(6):R58.

Muramatsu T, Shibata O, Ryoki S, Ohmori Y, Okumura J. (1997) Foreign gene expression in the mouse testis by localized in vivo gene transfer. *Biochem Biophys Res Commun.* **233**(1):45-9.

Nachman MW, Crowell SL. (2000) Estimate of the mutation rate per nucleotide in humans. *Genetics*. **156**(1):297-304.

Naski MC, Wang Q, Xu J, Ornitz DM. (1996) Graded activation of fibroblast growth factor receptor 3 by mutations causing achondroplasia and thanatophoric dysplasia. *Nat Genet.***13**(2):233-7.

Nikitina TV, Nazarenko SA. (2004) [Human microsatellites: mutation and evolution]. *Genetika* **40**:1301-18.

Nishino H, Buettner VL, Haavik J, Schaid DJ, Sommer SS. (1996) Spontaneous mutation in Big Blue transgenic mice: analysis of age, gender, and tissue type. *Environ Mol Mutagen* **28**:299-312.

Oatley MJ1, Kaucher AV, Racicot KE, Oatley JM. (2011) Inhibitor of DNA binding 4 is expressed selectively by single spermatogonia in the male germline and regulates the self-renewal of spermatogonial stem cells in mice. *Biol Reprod.* **85**(2):347-56.

Oatley JM, Brinster RL. Regulation of spermatogonial stem cell self-renewal in mammals. (2008) *Annu Rev Cell Dev Biol.* **24**:263-86.

Oldenburg J, Schwaab R, Grimm T, Zerres K, Hakenberg P, Brackmann HH, Olek K. (1993) Direct and indirect estimation of the sex ratio of mutation frequencies in hemophilia A. *Am J Hum Genet* **53**:1229-38.

Online Mendelian Inheritance in Man, OMIM®. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD), 2016. World Wide Web URL: http://omim.org/

Ono T, Ikehata H, Nakamura S, Saito Y, Hosoi Y, Takai Y, Yamada S, Onodera J, Yamamoto K. (2000) Age-associated increase of spontaneous mutant frequency and molecular nature of mutation in newborn and old lacZ-transgenic mouse. *Mutat Res* **447**:165-77.

Orioli IM, Castilla EE, Scarano G, Mastroiacovo P. (1995) Effect of paternal age in achondroplasia, thanatophoric dysplasia, and osteogenesis imperfecta. *Am J Med Genet.* **6**;59(2):29-17.

Paul C, Robaire B. (2013) Ageing of the male germ line. *Nat Rev Urol.* **10**(4):227-34

Passos-Bueno MR, Wilcox WR, Jabs EW, Sertie AL, Alonso LG, Kitoh H. (1999) Clinical spectrum of fibroblast growth factor receptor mutations. *Hum Mutat* **14**:115-25.

Penrose LS. (1955) Parental age and mutation. *Lancet.* **269**:312-313.

Petrie KA, Lee WH, Bullock AN, Pointon JJ, Smith R, Russell RG, Brown MA, Wordsworth BP, Triffitt JT. (2009) Novel mutations in ACVR1 result in atypical features in two fibrodysplasia ossificans progressiva patients. *PLoS One.* **4**(3):e5005.

Poirot C, Schubert B. (2011) [Fertility preservation in prepubertal children]. Bull Cancer. **98**(5):489-99.

Pfeifer, GP. (2006) Mutagenesis at Methylated CpG Sequences. *Curr Top Microbiol Immunol* **301**:259-281

Qin J, Calabrese P, Tiemann-Boege I, Shinde DN, Yoon SR, Gelfand D, Bauer K, Arnheim N. (2007) The molecular anatomy of spontaneous germline mutations in human testes. PLoS Biol. **5**(9):e224.

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Rahbari R, Wuster A, Lindsay SJ, Hardwick RJ, Alexandrov LB, Al Turki S, Dominiczak A, Morris A, Porteous D, Smith B, Stratton MR; UK10K Consortium, Hurles ME. (2016) Timing, rates and spectra of human germline mutation. *Nat Genet.* **48**(2):126-33.

Rannan-Eliya SV, Taylor IB, De Heer IM, Van Den Ouweland AM, Wall SA, Wilkie AO. (2004) Paternal origin of FGFR3 mutations in Muenke-type craniosynostosis. *Hum Genet* **115**:200-207

Raue F, Frank-Raue K. (2010) Update multiple endocrine neoplasia type 2. *Fam Cancer*. **9**(3):449-57

Ray D, Pitts PB, Hogarth CA, Whitmore LS, Griswold MD, Ye P. (2014) Computer simulations of the mouse spermatogenic cycle. *Biol Open.* **4**(1):1-12

de Rooij DG, Russell LD. (2000) All you wanted to know about spermatogonia but were afraid to ask. *J Androl.* **21**(6):776-98.

de Rooij DG, Griswold MD. (2012) Questions about spermatogonia posed and answered since 2000. *J Androl.* **33**(6):1085-95.

de Rooij DG, van Beek ME. (2013) Computer simulation of the rodent spermatogonial stem cell niche. *Biol Reprod.* **88**(5):131

Rosendaal FR, Brocker-Vriends AH, van Houwelingen JC, Smit C, Varekamp I, van Dijck H, Suurmeijer TP, Vandenbroucke JP, Briet E. (1990) Sex ratio of the mutation frequencies in haemophilia A: estimation and meta-analysis. *Hum Genet* **86**:139-46.

Risch N, Reich EW, Wishnick MM, McCarthy JG. (1987) Spontaneous mutation and parental age in humans. *Am J Hum Genet*. **41**(2):218-48.

Rives N, Langlois G, Bordes A, Simeon N, Mace B. (2002) Cytogenetic analysis of spermatozoa from males aged between 47 and 71 years. *J Med Genet* **39**:E63.

Sandstedt SA, Tucker PK. (2005) Male-driven evolution in closely related species of the mouse genus Mus. *J Mol Evol* **61**:138-44

Sarabipour S, Hristova K. (2016) Mechanism of FGF receptor dimerization and activation. *Nat Commun.* **7**:10262.

Sariola H, Saarma M. (2003) Novel functions and signalling pathways for GDNF. *J Cell Sci.* **116**:3855-62.

Sartorelli EM, Mazzucatto LF, de Pina-Neto JM. (2001) Effect of paternal age on human sperm chromosomes. *Fertil Steril* **76**:1119-23.

Sato T, Aiyama Y, Ishii-Inagaki M, Hara K, Tsunekawa N, Harikae K, Uemura-Kamata M, Shinomura M, Zhu XB, Maeda S, Kuwahara-Otani S, Kudo A, Kawakami H, Kanai-Azuma M, Fujiwara M, Miyamae Y, Yoshida S, Seki M, Kurohmaru M, Kanai Y. (2011) Cyclical and patch-like GDNF distribution along the basal surface of Sertoli cells in mouse and hamster testes. *PLoS One.* **6**(12):e28367.

Schaefer CB, Ooi SKT, Bestor TH, Bourc'his D. (2007) Epigenetic Decisions in Mammalian Germ Cells. *Science* **316**: 398-399

Shi W, Shi T, Chen Z, Lin J, Jia X, Wang J, Shi H. (2010) Generation of sp3111 transgenic RNAi mice via permanent integration of small hairpin RNAs in repopulating spermatogonial cells in vivo. *Acta Biochim Biophys Sin (Shanghai).* **42**(2):116-221.

Shimmin LC, Chang BH, Li WH. (1993) Male-driven evolution of DNA sequences. *Nature* **362**:745-7.

Shinde DN, Elmer DP, Calabrese P, Boulanger J, Arnheim N, Tiemann-Boege I. (2013) New evidence for positive selection helps explain the paternal age effect observed in achondroplasia. Hum Mol Genet. **22**(20):4117-26.

Shinohara T, Orwig KE, Avarbock MR, Brinster RL. (2001) Remodeling of the postnatal mouse testis is accompanied by dramatic changes in stem cell number and niche accessibility. *Proc Natl Acad Sci U S A.* **98**(11):6186-91.

Shinohara T, Avarbock MR, Brinster RL. (1999) beta1- and alpha6-integrin are surface markers on mouse spermatogonial stem cells. *Proc Natl Acad Sci U S A.* **96**(10):5504-9.

Shore EM, Xu M, Feldman GJ, Fenstermacher DA, Cho TJ, Choi IH, Connor JM, Delai P, Glaser DL, LeMerrer M, Morhart R, Rogers JG, Smith R, Triffitt JT, Urtizberea JA, Zasloff M, Brown MA, Kaplan FS. (2006) A recurrent mutation in the BMP type I receptor ACVR1 causes inherited and sporadic fibrodysplasia ossificans progressiva. *Nat Genet.* 38(5):525-7.

Sipos A, Rasmussen F, Harrison G, Tynelius P, Lewis G, Leon DA, Gunnell D. (2004) Paternal age and schizophrenia: a population based cohort study *BMJ*. **329**:1070

Smith NG, Hurst LD. (1999) The causes of synonymous rate variation in the rodent genome. Can substitution rates be used to estimate the sex bias in mutation rate? *Genetics* **152**:661-73.

Stukenborg JB, Wistuba J, Luetjens CM, Elhija MA, Huleihel M, Lunenfeld E, Gromoll J, Nieschlag E, Schlatt S. (2008) Coculture of spermatogonia with somatic cells in a novel three-dimensional soft-agar-culture-system. *J Androl.*; **29**(3):312-29.

Subramanian S, Kumar S. (2003) Neutral substitutions occur at a faster rate in exons than in noncoding DNA in primate genomes. *Genome Res* **13**:838-44.

Sun F, Xu Q, Zhao D, Chen CD. (2015) Id4 Marks Spermatogonial Stem Cells in the Mouse Testis. *Sci Rep.* **5**:17594.

Sundström H, Webster MT, Ellegren H. (2003) Is the rate of insertion and deletion mutation male biased?: Molecular evolutionary analysis of avian and primate sex chromosome sequences. *Genetics* **164**:259-268

Taylor J, Tyekucheva S, Zody M, Chiaromonte F, Makova KD. (2006) Strong and weak male mutation bias at different sites in the primate genomes: insights from the human-chimpanzee comparison. *Mol Biol Evol* **23**:565-73.

Tartaglia M, Cordeddu V, Chang H, Shaw A, Kalidas K, Crosby A, Patton MA, Sorcini M, van der Burgt I, Jeffery S, Gelb BD. (2004) Paternal Germline Origin and Sex-Ratio

Distortion in Transmission of PTPN11 Mutations in Noonan Syndrome. *Am J Hum Genet* **75**:492-497

Tavormina PL, Shiang R, Thompson LM, Zhu YZ, Wilkin DJ, Lachman RS, Wilcox WR, Rimoin DL, Cohn DH, Wasmuth JJ. (1995) Thanatophoric dysplasia (types I and II) caused by distinct mutations in fibroblast growth factor receptor 3. *Nat Genet* **9**:321-8.

Tegelenbosch RA, de Rooij DG. (1993) A quantitative study of spermatogonial multiplication and stem cell renewal in the C3H/101 F1 hybrid mouse. *Mutat Res.* **290**(2):193-200.

Tiemann-Boege I, Navidi W, Grewal R, Cohn D, Eskenazi B, Wyrobek AJ, Arnheim N. (2002) The observed human sperm mutation frequency cannot explain the achondroplasia paternal age effect. *Proc Natl Acad Sci U S A* **99**:14952-14957

Tomasetti C, Vogelstein B. (2015) Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science.* **347**(6217):78-81.

Tomlinson PM and Bodmer WF. (1995) Failure of programmed cell death and differentiation as causes of tumors: Some simple mathematical models. *Proc. of the National Academy of Sciences*, **92**:11130–11134

Unryn BM, Cook LS, Riabowol KT. (2005) Paternal age is positively linked to telomere length of children. *Aging Cell* **4**:97-101.

Vogel F, Motulsky AG. (1997) Human genetics: problems and approaches, 3rd edn. Springer, Berlin Heidelberg New York

Vogel F, Rathenberg R. (1975) Spontaneous mutation in man. *Adv Hum Genet.* **5**:223-318.

Vogels A, Fryns JP. Pfeiffer syndrome. (2006) *Orphanet J Rare Dis.* **1**:19.

Walsh CP, Xu GL. (2006) Cytosine methylation and DNA repair. *Curr Top Microbiol Immunol* **301**:283-315.

Walter CA, Intano GW, McCarrey JR, McMahan CA, Walter RB. (1998) Mutation frequency declines during spermatogenesis in young mice but increases in old mice. *Proc Natl Acad Sci U S A* **95**:10015-9.

Walter CA, Intano GW, McMahan CA, Kelner K, McCarrey JR, Walter RB. (2004) Mutation spectral changes in spermatogenic cells obtained from old mice. *DNA Repair* **3**:495-504.

Weinberg W. (1912) Zur Vererbung des Zergwuchses. *Arch Rassen-u Gesell Biol* **9**:710-718.

Wellbrock C, Arozarena I. (2016) The Complexity of the ERK/MAP-Kinase Pathway and the Treatment of Melanoma Skin Cancer. *Front Cell Dev Biol.* **4**:33.

Wilkin DJ, Szabo JK, Cameron R, Henderson S, Bellus GA, Mack ML, Kaitila I, Loughlin J, Munnich A, Sykes B, Bonaventure J, Francomano CA. (1998) Mutations in fibroblast growth-factor receptor 3 in sporadic cases of achondroplasia occur exclusively on the paternally derived chromosome. *Am J Hum Genet* **63**:711-6.

Wu Z, Luby-Phelps K, Bugde A, Molyneux LA, Denard B, Li WH, Süel GM, Garbers DL. (2009) Capacity for stochastic self-renewal and differentiation in mammalian spermatogonial stem cells. *J Cell Biol.* **187**(4):513-24.

Wyrobek AJ, Eskenazi B, Young S, Arnheim N, Tiemann-Boege I, Jabs EW, Glaser RL, Pearson FS, Evenson D. (2006) Advancing age has differential effects on DNA damage, chromatin integrity, gene mutations, and aneuploidies in sperm. *Proc Natl Acad Sci U S A* **103**:9601-6.

Yomgogida K. (2008) Mammalian testis: a target of in vivo electroporation. *Dev Growth Differ.* **50**(6):513-5.

Yomogida K, Yagura Y, Nishimune Y. (2002) Electroporated transgene-rescued spermatogenesis in infertile mutant mice with a sertoli cell defect. *Biol Reprod.* **67**(3):712-7.

Yoon SR, Qin J, Glaser RL, Jabs EW, Wexler NS, Sokol R, Arnheim N, Calabrese P. (2009) The ups and downs of mutation frequencies during aging can account for the Apert syndrome paternal age effect. PLoS Genet. **5**(7)

Yoon SR, Choi SK, Eboreime J, Gelb BD, Calabrese P, Arnheim N. (2013) Age-dependent germline mosaicism of the most common noonan syndrome mutation shows the signature of germline selection. *Am J Hum Genet.* **92**(6):917-26.

Yoshidaf S, Sukeno M, Nabeshima Y. (2007) A vasculature-associated niche for undifferentiated spermatogonia in the mouse testis. *Science.* **317**(5845):1722-6.

Yoshida S. (2008) Spermatogenic Stem Cell System in the Mouse Testis. *Cold Spring Harb Symp Quant Biol* **73**: 25-32

APPENDIX A

ADDITIONAL EQUATIONS FOR CHAPTER 3

For $t>0$, define the function ψ such that

$$\psi(t)=\frac{1-e^{at}}{1-e^{bt}} \tag{61}$$

where $a\neq b$, $b\neq 0$.

Lemma A.1. A necessary and sufficient condition for $\psi(t)$ to be a constant is that $a=0$.

Proof: First, if $a = 0$ then (61) guarantees that $\psi(t) = 0$ for all $t > 0$.

Conversely, assume $\psi(t)$ to be a constant independent of t. Thus, in particular, $\psi(1) = \psi(2)$, which amounts to:

$$\frac{1-e^{a}}{1-e^{b}}=\frac{1-e^{2a}}{1-e^{2b}} \tag{62}$$

However:

$$\frac{1-e^{2a}}{1-e^{2b}}=\frac{1-e^{a}}{1-e^{b}}.\frac{1+e^{a}}{1+e^{b}} \tag{63}$$

and so

$$\frac{1-e^{a}}{1-e^{b}}=\frac{1-e^{a}}{1-e^{b}}.\frac{1+e^{a}}{1+e^{b}}\,. \tag{64}$$

Since $b \neq 0$, the above is equivalent to:

$$1-e^{a}=\left(1-e^{a}\right).\frac{1+e^{a}}{1+e^{b}}\,. \tag{65}$$

If $a \neq 0$ then $1 - e^{a} \neq 0$ and we obtain:

$$1 = \frac{1+e^a}{1+e^b}. \tag{66}$$

which implies that $a = b$, a contradiction. Therefore $a = 0$, as claimed. This completes the proof of the lemma.

Lemma A.2 If x and y are finite, then for all $t \in \mathbb{R}$

$$\lim_{x \to y} \frac{e^{xt} - e^{yt}}{x - y} = te^{yt}. \tag{67}$$

Proof: We write:

$$\frac{e^{xt} - e^{yt}}{x - y} = e^{yt} \frac{e^{(x-y)t} - 1}{x - y}. \tag{68}$$

and so

$$\begin{aligned} \lim_{x \to y} \frac{e^{xt} - e^{yt}}{x - y} &= \lim_{x \to y} \left(e^{yt} \frac{e^{(x-y)t} - 1}{x - y} \right) \\ &= e^{yt} \lim_{x \to y} \left(\frac{e^{(x-y)t} - 1}{x - y} \right) \\ &= e^{yt} \lim_{z \to 0} \left(\frac{e^{zt} - 1}{z} \right) \text{ [where } z = x - y] \\ &= te^{yt} \end{aligned} \quad . \tag{69}$$

as claimed.

APPENDIX B

SIMULATION OF SPERMATOGONIAL STEM CELL NICHE

This appendix contains supplementary data to Chapter 4.

Source code: C++/R source codes and documentation including compilation instructions are available under GNU license at https://github.com/anwala/NicheSimulation.

number of cells ***n***,
cell divisions ***divisions***,
experiment runs (or replications) ***Runs***,
probabilities ***p*** and ***r***,
mutant cell count ***mutCount***,
m × 1 matrix ***M*** (runs matrix), where |***M***| = ***Runs***,
and m × 1 matrix ***P*** (average runs matrix), where |***P***| = ***e***,

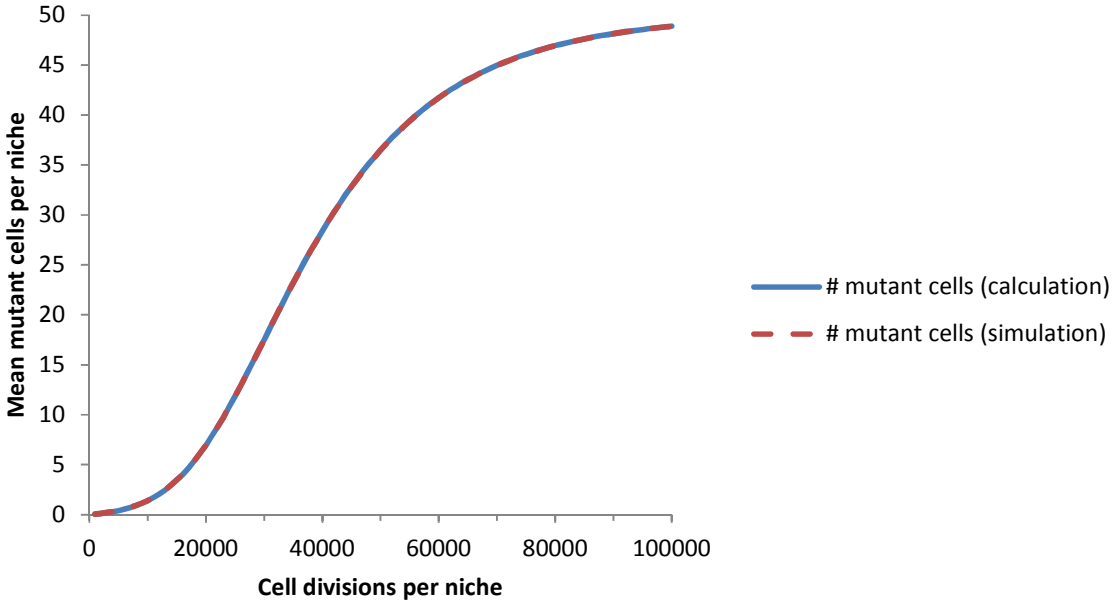
given X ~ *U([0,1]), X is a random variable uniformly distributed on [0, 1]*

```
function runNExperiments(divisions, Runs)

        for i = 0 to divisions - 1 do
                totalRedBallCount = 0;
                //this is for a single division instance run multiple independent times
                for j = 0 to Runs do
                        mutCount = Mj;
                        Mj = nicheSimulation(n, divisions, p, r, mutCount)
                        totalMutCount = totalMutCount + Mj;
                endFor
                //average for a single niche over all the runs
                Pi = totalMutCount / Runs;

        endFor
return P
```

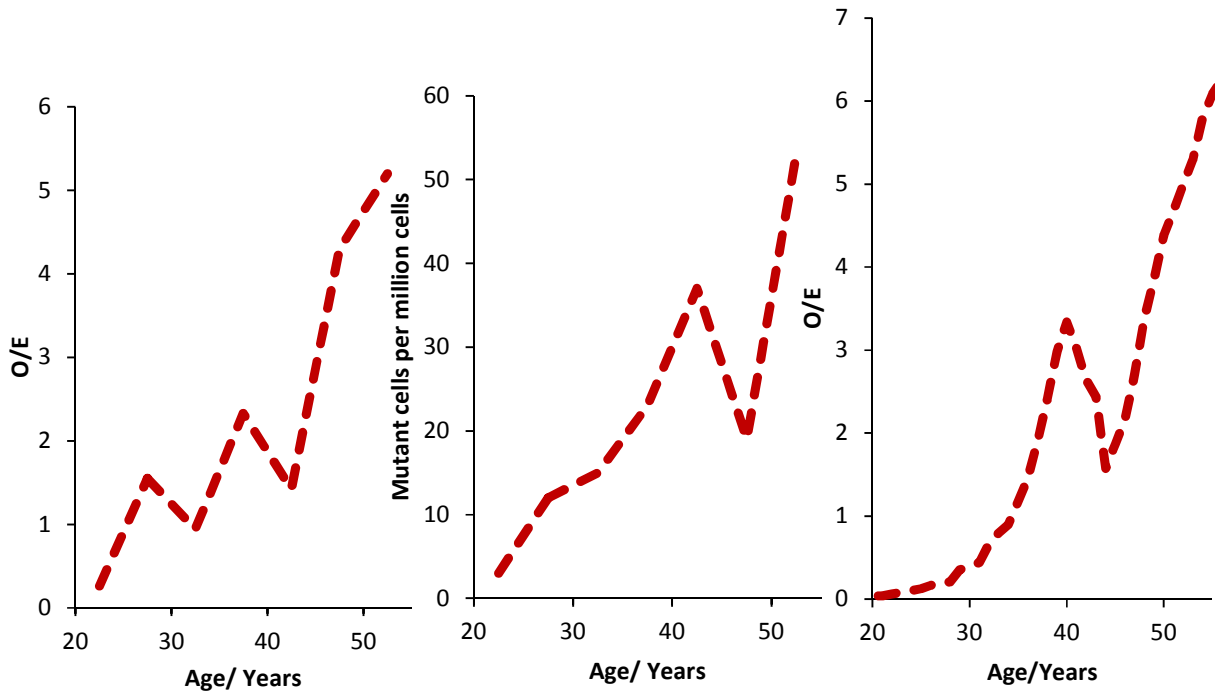
Supplementary Algorithm B.1: Algorithm for simulation of niche division.



Supplementary Fig. B.2 Comparison of mathematical model with computer simulation. Parameters used are arbitrary and intended to test the model. $p=5\times10^{-5}$, $r=0.01$, $n=50$, $N=10^5$. Simulation repeated 10^5 times and mean values reported.

APPENDIX C

SIMULATION OF APOPTOSIS IN THE SC NICHE



Supplementary Fig. C.3. Distribution of Apert's syndrome mutation incidences with age. A. from Yoon *et al.* (2009) shows O/E from birth data. B. also from Yoon *et al.* (2009) shows mutational assay of sperm. C. shows my simulation.

Fig. C.3 shows comparison of simulation to sequence and birth data. The simulation was constructed based off the simulations in Appendix B, with the exception that an equal number of A_{dark} cells were granted to the system as A_{pale} SSCs. Up to age 40 the apoptosis rate is 0, but after 40 the apoptosis rate goes to 0.01 per cell division of A_{pale} SSCs. Upon an apoptotic event, an

A_{pale} cell is lost (whether mutant or wildtype), then a A_{dark} SSC converts to a wildtype A_{pale}. In this model A_{dark} cells that convert are not replenished.

While this simulation did match the sequence data from Yoon and colleagues (2009) and corresponded to their simulation, this result needs to be treated with some scepticism. Firstly, the evidence that A_{dark} cells operate purely as reserve cells is limited and they may cycle in-and-out of active division (K Orwig, personal communication, June 2016). Additionally, this model requires cells to undergo apoptosis starting at age 40 – a continuous low-level apoptosis rate does not produce this pattern. Finally, if this is the mechanism for this cell replacement, one would expect this to happen to all paternal age effect syndromes. It's possible there is a similar "dip" in other syndromes (see Fig. 12) and sequencing data (Fig 14), there is far from a clear pattern across all genes, although it may be obscured by stochastic noise or small sample sizes.

VITA

Eoin C. Whelan
Department of Biological Science
110 Mills, Godwin Life Sciences Bldg, Old Dominion University, Norfolk, VA 23529

Educational Background

Bachelor of Science in Genetics, University College London (2003)
Doctor of Philosophy in Biomedical Sciences, Old Dominion University (2016)

Positions held

Research Assistant, Department of Genetics, Cambridge University, UK (summer 2003)
Instructor for Genetics 303, Department of Biology, Old Dominion University, USA (summer 2010 & 2011)
Laboratory Teaching Assistant, Old Dominion University (2009-2015)

Awards

ODU University Fellowship (2011-2012)
ODU S5 Fellowship (summer 2009)
ODU Dominion Scholarship (2005-2009)

Grants

Mary Louise Andrews Award for Cancer Research (2013)
Sigma Xi Grants-in-aid-of-research (2012)

Papers

Mahadevan K, Patthipati VS, Han S, Swanson RJ, Whelan EC, Osgood C, Balasubramanian R. One-pot synthesis and functionalization of nanocapsules by thiolene reaction: Highly fluorescent resorcinarene nanocapsules with efficient renal clearance for biomedical imaging. *Nanotechnology.* 2016 Aug 19;27(33):33510

Whelan EC, Nwala AC, Osgood C, Olariu S. Selective Mutation Accumulation: A Computational Model of the Paternal Age Effect. *Bioinformatics.* 2016 (accepted pending publication)

Whelan EC, Nwala AC, Osgood C, Olarius S. Reasoning About Homeostasis in Tissue Lineages. *Journal of Mathematical Biology*. In review.

Presentations

Whelan EC, Stacey MW, Osgood C. Positive selection of spermatogonial stem cells. Virginia Academy of Sciences. VCU, Richmond, VA. May 15, 2014.

Whelan EC, Nwala AC, Osgood C, Olariu S. Selective Mutation Accumulation: A Computational Model of the Paternal Age Effect. Germline Stem Cells Conference, June 2016, San Francisco USA.

Whelan EC, Stacey MW, Osgood C. In Vitro Investigation into the Proliferative Advantage of RET M918T Mutations in Spermatogonial Stem Cells. International Society for Stem Cell Research, June 2016, San Francisco USA.