UCSF
UC San Francisco Electronic Theses and Dissertations

Title
Analysis of the function of the Deleted-in-AZoospermia-Like gene in germ cell development

Permalink
https://escholarship.org/uc/item/9dx0k53m

Author
Tung, Joyce Yeh-hong

Publication Date
2005

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Analysis of the function of the Deleted-in-AZoospermia-Like gene in germ cell development

by

Joyce Yeh-hong Tung

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Genetics

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Date University Librarian

Degree Conferred:...

Copyright 2005

by

Joyce Yeh-hong Tung

Dedication and Acknowledgements

I would like to dedicate this work to the friends and family who have made the completion of this thesis possible. My friends and colleagues in the Reijo Pera lab have provided me with so much valuable advice, technical assistance, and emotional support throughout the years, for which I am very grateful. I want to thank the members of my thesis committee, Drs. Gail Martin, Barbara Panning, and Cynthia Kenyon, for their conscientious and wise guidance in my project. I would especially like to acknowledge my advisor, Dr. Renee Reijo Pera, whose encouragement and mentorship have made my graduate experience a uniquely positive one. I sincerely appreciate the time and effort she put into making my career a successful one.

I also wish to acknowledge my parents and my brother who, thoughout my life, have encouraged me and given me anything I needed to prosper. They pushed me ever higher, but were always there to support me if I fell.

Finally, I would like to thank my husband, Eddie, who thought that going to graduate school was a good idea, even when he found out that I would not make enough to buy him a Porsche. He has always believed that I could do anything I put my mind to, but would love me even if I failed.

I am a very lucky person indeed.

Chapter 2 was originally published in the journal *Human Genetics*, and Chapter 5 in the journal *Development, Genes & Evolution*. Published here with kind permission of Springer Science and Business Media.

Abstract

Germ cell development is key to the propagation of all sexually reproducing species, yet few genes have been conclusively linked to human germ cell development. One family clearly linked to mammalian germ cell development is the *Deleted in AZoospermia* (*DAZ*) gene family. Deletion of the founding member, Y-chromosome *DAZ*, is associated with azoospermia/oligozoospermia in many infertile men, while disruption of the autosomal gene, DAZ-Like (*Dazl*), in mouse results in germ cell loss and infertility in both sexes. At the onset of this research many questions remained regarding the role, if any, of *DAZL* in human germ cell development and the timing and trigger for germ cell loss in the mouse *Dazl* mutant. To determine the role of *DAZL* in both male and female human germ cell development, I examined the relationship of sequence variants with measures of reproductive capacity. I identified and characterized common polymorphisms associated with age at menopause/ovarian failure and sperm count, and rare, amino-acid changing mutations that appear to be restricted to individuals with impaired fertility. I additionally analyzed the embryonic development of mouse *Dazl* mutants to determine when germ cells are lost in both sexes and which genes are expressed in germ cells of different genotypes. I found that in the *Dazl* mutant, germ cell numbers were reduced before e14.5, earlier than previously described. Many genes important for germ cell development also are expressed aberrantly, indicating that *Dazl* is required for proper fetal germ cell differentiation and potentially mRNA expression. Finally, during these studies, I observed differences in sequence variation in different members of the *DAZ* gene family, prompting me to examine the impact of the introduction of Y-chromosome *DAZ* into the primate lineage on the evolution of the other members of the *DAZ* family,

DAZL and *BOULE*. Though *DAZ* is 90% identical to *DAZL*, it does not appear to relieve the functional constraint on *DAZL* and *BOULE*, most likely because females do not possess *DAZ* and cannot rely on its function.

In summary, this work clearly leads to our understanding of the functional role of mammalian *DAZL* in the differentiation and development of germ cells.

Table of Contents

Chapter	**Page**
1: Introduction	1
2: Variants in *Deleted in AZoospermia-Like* (*DAZL*) are correlated with reproductive parameters in men and women	17
3: Additional sequence variants found in the screen of *Deleted in AZoospermia-Like* (*DAZL*)	53
4: *Dazl* has a meiosis-independent function in germ cell development and results in germ cell loss by e14.5 in the mouse	61
5: Evolutionary comparison of the reproductive genes, *DAZL* and *BOULE,* in primates with and without *DAZ*	79
6: Conclusions	109
References	120

List of Tables

Table	Page
Chapter 1, Table 1 Genes required for germ cell development	11
Chapter 1, Table 2 Homology of DAZ family genes: % identity to human DAZL	14
Chapter 2, Table 1 Baseline characteristics for each study population	34
Chapter 2, Table 2 Primer sequences for amplifying and sequencing the 11 *DAZL* Exons	35
Chapter 2, Table 3 Genotype frequencies for each SNP in Caucasians by population	36
Chapter 2, Table 4 Average age at premature/ovarian failure/menopause, total sperm count, and total motile count by genotype for the 12 *DAZL* SNPs in Caucasians	37
Chapter 2, Table 5 The impact of the 12 *DAZL* SNPs on reproductive parameters	38
Chapter 2, Supplementary Table 1 Genotype frequencies for each SNP in the Coriell population	46
Chapter 3, Table 1 Missense mutations identified in the *DAZL* gene	59
Chapter 4, Table 1 Primers used for real time PCR	74
Chapter 5, Table 1 Comparison of branch lengths from the common ancestor of platyrrhines and catarrhines	100
Chapter 5, Table 2 Pairwise comparison of evolutionary rates using the Tajima relative rate test	101

Chapter 5, Table 3 102
Rate of non-synonymous relative to synonymous change (K_a/K_s ratio) for *DAZL* and *BOULE* across all species examined

Chapter 5, Table 4 103
Intralineage synonymous substitution rates (K_S)

Chapter 6, Table 1 117
Estimated number of doublings in wild type and *Dazl* mutant mouse germ cells

List of Figures

Table	**Page**
Chapter 1, Figure 1 Model of evolutionary history of the *DAZ* gene family	16
Chapter 2, Figure 1 Structure and SNPs in *DAZL* mRNA	42
Chapter 2, Figure 2 SNP characteristics and allele frequencies	42
Chapter 2, Figure 3 Dot plot alignment of mouse and human 3'UTR	43
Chapter 2, Figure 4 Results of haplotype linear regression analysis	44
Chapter 2, Figure 5 Model for how the combined strength of multiple common variants can lead to a spectrum of phenotypes	44
Chapter 2, Figure 6 Opposing effects of the same allele in men and women	45
Chapter 3, Figure 1 Diagram of sequence variants found in *DAZL* screen	60
Chapter 4, Figure 1 GFP expression from Oct4ΔPE promoter in the germ cells of wild type e14.5 testis and ovary	75
Chapter 4, Figure 2 Germ cell counts in e12.5, e14.5, and e16.5 embryos	76
Chapter 4, Figure 3 Germ cell gene expression in male and female germ cells	77
Chapter 4, Figure 4 Model for *Dazl* function in germ cell development	78
Chapter 5, Figure 1 Model of evolutionary history of the *DAZ* gene family	104

Chapter 1

Introduction

Reproduction is essential for the propagation of any species. Yet, in humans, 10-15% of couples are infertile [1]. In approximately 40% of infertility cases, a female factor is identified, in another 40% a male factor is identified and in the remaining 20% of cases, a combined male/female problem is suspected (http://www.medscape.com/viewarticle/480429). In many cases, the underlying etiology of the infertility is not known. For example, 40% of male infertility and at least 20% of female infertility is idiopathic (http://www.andrologyaustralia.org/pdfs/Causes_MaleInfertility.pdf, http://www.medscape.com/viewarticle/480429). Thus in many of these cases, there may be as yet undetermined genetic factor(s) that cause the infertility. As a result, while the study of human reproduction has traditionally focused on hormonal regulation, the genetics of germ cell development has recently become the focus of increasing study.

The pathway of germ cell development

Germ cell development is a key part of sexual reproduction—individuals that do not make germ cells are sterile. Accordingly, the gross features of germ cell development appear to be widely conserved. In species as diverse as *Drosophila melanogaster, Xenopus laevis, Mus musculus,* and *Homo sapiens*, germ cells follow a similar developmental pathway. Germ cells are allocated apart from somatic cells; the primordial germ cells (PGCs) must then migrate to the gonads. Early germ cell populations mitotically proliferate, then undergo meiosis, and finally differentiate into mature sperm or eggs [2, 3]. Through all this, the germ cells maintain totipotency—the ability to produce all the cell types of the embryo under appropriate conditions.

Despite gross similarities, however, there is scant evidence for conservation of the molecular mechanisms underlying germ cell development across species. A substance called germ plasm—embryonic cytoplasm that contains mitochondria, RNAs, and proteins crucial for germ cell development, and visible as an electron dense material under the electron microscope—is found under different names in many species, including P granules in *Caenorhabditis elegans*, pole plasm in *D. melanogaster*, and germ plasm in *X. laevis*. In these species, germ plasm plays a determinative role—cells that inherit germ plasm develop as germ cells; in the absence of germ plasm, germ cells do not develop. Mammals, however, appear to lack germ plasm. Instead, extracellular signals such as BMP4 and BMP8 are required to induce allocation of the germ cell lineage from a pool of equipotent cells [4]. Although these two methods of allocation seem very different, recent data suggests that there may be more similarities than previously thought. Many genes required for germ cell development have been found, and a number of them are shared between organisms that have germ plasm and those that do not [3]. A few of the increasing numbers of homologs of genes important for germ cell development in both vertebrates and invertebrates will be reviewed later.

Human genetics and fertility

Though human reproduction and fertility have been studied for many years, few genes have been identified that contribute to human germ cell production. Several studies, however, have demonstrated that the age at onset of menopause has a significant genetic component. Family history is a significant predictor of early menopause (menopause at age < 47 years), increasing a woman's risk of experiencing an early menopause by 6x if a family member was affected [5, 6]. Sibling studies have estimated the

heritability of menopause to be high, 63% in one study and 85% in another [7, 8]. Analysis of pedigrees from families with premature ovarian failure (POF, menopause at age < 40 years) suggests that POF can be passed paternally or maternally, and appears to be dominant autosomal or X-linked, suggesting that more than one gene is involved [9].

There have been several association studies that examined the relationships between different genes and early menopause or POF. The *FOXL2* gene is clearly associated with a syndrome called blepharophimosis ptosis epicanthus inversus syndrome (BPES), some types of which are also associated with POF; however, only a 30bp deletion in the gene was shown to segregate with idiopathic POF in a few cases [10-13]. Similarly, *EIF2B* was also shown to be associated with POF when coincident with a neurological disorder, but not with POF alone [14, 15]. A missense mutation in the follicle-stimulating hormone receptor gene, *FSHR*, was shown to be associated with POF in the Finnish population, but has not been replicated in any other studies and is thought to be specific to that ethnic group [16-20]. There is a missense mutation in the *inhibin*α gene that has been associated with POF in some ethnic groups (Asian Indian, Italian), but not in others (Korean), so the contribution of this gene to the phenotype is also unclear [21-23]. The strongest genetic associations described thus far seem to be with the X-chromosome. There is a relatively high incidence of the Fragile X premutation, which is an expansion of 61-200 CGG repeats in the 5'UTR of the *FMR-1* gene, in women with POF (16%) [24]. In another study, this premutation was found in 13% of women with familial POF and 3% of women with spontaneous POF [25]. The *POF1* and *POF2* loci on the X chromosome have also been strongly linked to ovarian failure in many studies but it is not clear whether the effects of these loci are linked to specific genes (reviewed in [26]).

Similar to female reproductive failure, there are several studies that have investigated the genetic component to sperm production. In these studies, the most common genetic lesions associated with spermatogenic defects are deletions of the Y chromosome, including deletions that encompass the *DAZ* gene, which are associated with azoospermia (no sperm in the ejaculate) and oligozoospermia (< 20 million sperm/mL of ejaculate) [27-29]. Rare point mutations have also been found in several other genes such as the *SYCP3* and *Protamine-2* genes [30, 31]. There have also been many conflicting reports on associations between the length of CAG and GGC repeats in the androgen receptor gene and male infertility, suggesting that any effect seen could be population dependent [32-36].

Genetics from model organisms

Despite the relative dearth of genetic data from humans, a number of genes required for germ cell development have been identified and characterized in model organisms. Four of the most well-characterized gene families in invertebrates and vertebrates—*Vasa*, *Pumilio*, *Nanos*, and *DAZ*—are described here.

Vasa

Vasa encodes a DEAD-box helicase that is germ cell specific in every species studied to date, making it an excellent germ cell marker. *Drosophila vasa* was first characterized as a maternal-effect gene required for abdomen formation and pole cell (*Drosophila* PGC) formation [37]. Later studies showed that it functions in oogenesis as well [38]. Studies in other species support a role for *vasa* in gametogenesis. However, the

mutant phenotypes vary from pre-pachytene meiotic arrest in mouse spermatocytes to a reduction in developing germline cysts in female flies to pachytene arrest in nematode oocytes [38-40]. These differences in phenotype may be a result of the evolution of the *vasa* gene, or there may be other genes that are the true *vasa* homologs in these organisms. For instance, while the worm *glh* genes have a DEAD-box motif and glycine-rich repeats like *Drosophila vasa*, their repeats contain no charged amino acids and both *glh-1* and *glh-2* contain additional multiple CCHC zinc fingers [39, 41].

Pumilio and *Nanos*

The *pumilio* and *nanos* genes have been most extensively characterized in the fly, where they work together to ensure proper migration and silencing of zygotic genes in pole cells, and to repress translation of both *hunchback* mRNA to promote embryonic abdomen development, and *cyclin B* mRNA in migrating pole cells [42-45]. *Pumilio* and *nanos* both also have roles in oogenesis, although *nanos* appears to function slightly later than *pumilio* indicating these two genes may not act together in the germline [46]. *Pumilio* and two *C. elegans* homologs, *fbf-1* and *fbf-2*, are the founding members of their subfamily, the Puf (PUmilio and Fbf) family [47]. Puf family members include eight additional genes in *C. elegans* and at least three nanos homologs, *nos-1*, *nos-2*, and *nos-3* [48] (www.wormbase.org). FBF-1 and FBF-2 appear to interact with the NOS proteins to effect the spermatogenesis to oogenesis switch in the worm, and numerous *puf* and *nos* mutants show various germ cell defects, including a failure of the germ cells to incorporate into the gonad in the *nos-2* mutant [48, 49].

Although vertebrate homologs of these genes have not been as extensively studied, two *pumilio* homologs have been identified in mice (*Pum1*, *Pum2*) and humans (*PUM1*, *PUM2*). Of the three sequences in the GenBank database for mouse *Pum1*, however, the protein sequence of one (accession #AY027917) is different from the other two (accession #NM_030722, AF321909), which raises questions about the existence of a third *pumilio* homolog. As for human *PUM2*, there are two splicing variants, which differ by one exon (accession #AF315591, XM_015812, NM_015317 vs. XM_037651, D87078). In mice, three *nanos* genes have been identified: *nanos1*, *nanos2*, and *nanos3* [50]. Disruption of *nanos1* does not affect germ cell development, but both *nanos2* and *nanos3* are expressed in the gonads, with *nanos2* expressed primarily developing male PGCs and *nanos3* expressed in both male and female PGCs starting at e9.5 [50, 51]. Accordingly, loss of *nanos2* function results in loss of male germ cells but not female germ cells, while loss of *nanos3* results in impaired maintenance of PGCs during migration in both sexes [50]. A human Nanos protein, NOS1, that interacts with PUM2 by coimmunoprecipitation has also been identified [52]. Other known vertebrate *nanos* homologs include *Xcat-2* in *Xenopus* and *nos1* and *nos2* in zebrafish [53, 54]. Zebrafish *nos1* is required for PGC migration and survival [53].

DAZ

Deleted in AZoospermia (*DAZ*), the founding member of the *DAZ* family, was discovered in 1995 in a screen for Y-chromosome genes that cause azoospermia (no mature germ cells in the ejaculate) when deleted in men [27]. The autosomal homolog of *DAZ*, *Deleted in AZoospermia-Like* (*DAZL*), was found shortly afterwards, and is located

in chromosome 3 in humans and chromosome 17 in mice (Table 1) [55, 56]. Lineage analysis showed that *boule*, a meiotic regulator, is the ancestor of this family and is found in both invertebrates and vertebrates. *DAZL* most likely arose from a duplication event in the vertebrate lineage, while *DAZ* arose from multiple duplications of *DAZL* after the split of Old World monkeys from New World monkeys (Fig. 1) [57, 58].

All DAZ proteins contain a highly-conserved RNA-binding domain that is part of the RNA Recognition Motif (RRM) family (Tables 1, 2). Substantial evidence indicates that DAZ proteins function as RNA-binding proteins [59-63]. In addition, experiments in *Xenopus* demonstrated that DAZL can promote translation of mRNAs, including that of the canonical germ cell marker, *Vasa*, and interacts with a canonical translation initiation factor, the poly(A)-binding protein [64, 65]. In addition to binding PABP, DAZ, DAZL, and BOULE can all bind each other as homodimers and heterodimers [58, 66]. Many other proteins required for germ cell development are involved in post-transcriptional regulation, including Pumilio and Nanos, and as DAZL has also been shown to interact with the Pumilio protein, these proteins may work together as a large RNA binding complex to control the protein expression of important germ cell genes [67].

Gene disruption experiments in the frog and mouse demonstrated that both *Xdazl* and *Dazl* are required for early germ cell development at the PGC stage, suggesting that these genes are crucial for early germ cell development [60, 68]. In frogs, in the absence of *Xdazl*, PGCs are formed normally but do not migrate successfully to the gonads, resulting in male and female infertility [60]. The initial analysis of the mouse *Dazl* knockout described germ cell loss in both sexes by 19dpc [68]. Further studies focused on male germ cell development, and showed that *Dazl* knockout testes have some germ cells, but

reduced in number compared to wild type and unable to progress past the leptotene stage of meiosis [69]. Studies of *DAZL* function in humans have been restricted to looking for associations between two *DAZL* single nucleotide polymorphisms (SNPs)—T54A and T12A—in various populations. A positive association was observed between T54A and infertility in a male Taiwanese population, but other groups have not been able to replicate this finding populations of different ethnicities [70-73].

For my thesis, I have chosen to focus on understanding the function of the autosomal *DAZL* gene in mammalian germ cell development, specifically in human and mouse. The Y-chromosome *DAZ* gene has an established function in human male fertility; however, even though *DAZL* is expressed only in germ cells in both men and women, little is known of its function. In mice, several studies have examined the role of *Dazl* postnatally in male mice but there are very few data on embryonic germ cell development in *Dazl* mutants. As a result, since much of oocyte development occurs before birth, there is also relatively little information about the effect of the mouse *Dazl* knockout on female germ cells. Finally, while mice only have *Dazl*, humans have *DAZL* and the very closely-related Y-chromosome *DAZ* gene. It has been postulated that *DAZ* has no function because the exons and the introns were evolving at the same rate [74]. This claim was subsequently refuted [75]. It is unclear, however, since *DAZ* is functional, what effect its introduction had on the rest of the *DAZ* gene family. In the following chapters, I describe experiments that I conducted with the goal of addressing the following questions:

- What is the function of *DAZL* in humans? Is it required for germ cell development and maintenance in humans?

- When does the germ cell defect first occur in *Dazl* knockout males and females, and what is the nature of this defect?
- How did the introduction of the *DAZ* gene in the Old World monkey and hominid lineage affect the evolution of the other *DAZ* family members? Did *DAZ* relax the functional constraint on the *DAZL* and *BOULE* genes?

Thus, my specific aims were to:

- **Determine whether *DAZL* is required for germ cell development in humans. I sequenced the *DAZL* gene in patient populations of women with early menopause or premature ovarian failure and of men with oligozoospermia or azoospermia, and searched for sequence variants that were associated with these conditions.**
- **Determine the onset and nature of the germ cell defect in the *Dazl* mutant mouse. I examined embryonic germ cell number and germ cell gene expression profiles in both sexes in mutant mice.**
- **Determine whether *DAZ* had an effect on the evolution of the *DAZL* and *BOULE* genes. I analyzed the sequence evolution of these genes in primates with and without Y-chromosome *DAZ*.**

Table 1 Genes required for germ cell development*

	Species	Gene name	Phenotype of Mutant	References**
Vasa	RGG motif; Glycine rich region; CCHC motif; RNA helicase domain			
	Human (*Homo sapiens*)	*VASA*	not yet determined	76
	Mouse (*Mus musculus*)	*Mvh*	pre-pachytene arrest in male germ cells; females are normal	77 40
	Frog (*Xenopus laevis*)	*XVLG1*	failure of PGC differentiation at the tadpole stage	78 79
	Zebrafish (*Danio rerio*)	*vasa*	not yet determined	80
	Chick (*Gallus gallus*)	*Cvh*	not yet determined	81
	Fly (*Drosophila melanogaster*)	*vasa*	disruption of D/V, A/P axes, mislocalization of mRNAs, no *gurken*, *oskar*, *nanos* mRNA translation, abnormal degenerating egg chambers, fewer developing germline cysts, no abdomen development, no pole plasm or germ cells	82 83 38 84 85 37 37
	Nematode (*Caenorhabditis elegans*)	*glh-1*, *glh-2*, *glh-3*, *glh-4*	*glh-1*, *glh-2*, and *glh-4* mutations all cause abnormal gonads, underproliferated germlines, no oocyte production. *glh-1/glh-4(RNAi)* oocytes arrest in pachytene, *glh-1/glh-4(RNAi)* sperm are defective	41 39
	Planarian (*Dugesia japonica*)	*DjvlgA*, *DjvlgB*	not yet determined	86
	Ascidian (*Ciona instestinalis*)	*Ci-DEAD1*	not yet determined	87
Pumilio	Csp1a/b; PUF repeats; Csp2			
	Human (*Homo sapiens*)	*PUM1*, *PUM2*	not yet determined	Genbank accession # *PUM1*: XM_015324, XM_049327, NM_014676, AF315592, D43951 *PUM2*: AF315591, XM_015812, NM_015317, XM_037651, D87078
	Mouse	*Pum1*, *Pum2*	not yet determined	Genbank accession #: *Pum1*:

Family / Organism	Gene	Phenotype	Reference
(*Mus musculus*)			NM_030722, AF321909, AY021917 *Pum2*: NM_030722, AF315590 88
Frog (*Xenopus laevis*)	*Xpum*	not yet determined	89
Fly (*Drosophila melanogaster*)	*pumilio*	no abdomen development; pole cells do not migrate to gonad, ectopic expression of zygotic markers, pole cells proliferate during migration; females: failure to establish or maintain germline stem cells, ovarian somatic cell tumors, abnormal numbers of PGCs, irregular germ cell morphology, oogenesis defects	90, 91, 46, 92, 42
Nematode (*Caenorhabditis elegans*)	*fbf-1, fbf-2, puf-3, puf-4 puf-5, puf-6 puf-7, puf-8 puf-9, puf-10*	*fbf-1, fbf-2*: animals produce excess sperm and no oocytes, small germ lines; *fbf-1/ fbf-2/puf-6/puf-7/puf-8* mutant: PGCs outside of gonad, premature proliferation of PGCs under starvation conditions, germ cell death; *puf-8*: 1° spermatocytes undergo mitosis instead of meiosis, fail in sperm-oocyte switch; *fbf(RNAi) him-5* males Produce fewer and less viable embryos	93, 48, 94, 95, 96, http://www.wormbase.org
Nanos	CCHC Zn finger domain		
Human (*Homo sapiens*)	*NANOS1*	not yet determined	52
Mouse (*Mus musculus*)	*nanos1, nanos2, nanos3*	*nanos1*: no detectable defect; *nanos2*: loss of male germ cells beginning at e15.5; *nanos3*: loss of male and female PGCs beginning during migration	50, 51
Frog (*Xenopus laevis*)	*Xcat-2*	not yet determined	54
Zebrafish	*nos1, nos2*	Most PGCs fail to migrate to the gonads, ectopic PGCs undergo apoptosis	*nos1*: 53 *nos2*: Genbank accession #ai585000
Fly (*Drosophila melanogaster*)	*nanos*	no abdomen development, pole cells do not migrate to gonad, irregular pole cell morphology, ectopic expression of *Sxl*, *ftz*, *eve* in pole cells, pole cells continue to divide after cellular blastoderm stage;	97, 46, 44, 43, 98

		females: pole cells do not incorporate into the gonad, failure to establish and maintain activity of germline stem cells; males: overproliferation of 1° spermatocytes with very few sperm bundles	
Nematode (*Caenorhabditis elegans*)	*nos-1, nos-2* *nos-3*	*nos-1/nos-2* mutants: reduced germ cell proliferation but premature proliferation of PGCs under starvation conditions. *Nos-2* mutant PGCs do not incorporate efficiently into gonad. *nos-1/nos-2/nos-3* mutants fail to switch from spermatogenesis to oogenesis	48 49

DAZ

RNA Recognition Motif — DAZ repeats

DAZL/BOULE

RNA Recognition Motif — DAZ repeat

Human (*Homo sapiens*)	*DAZ*	oligozoospermia, azoospermia	27 28
Human (*Homo sapiens*)	*DAZL*	not yet determined	56
Mouse (*Mus musculus*)	*Dazl*	failure of germ cell maintenance and maturation in both male and female	55 68
Frog (*Xenopus laevis*)	*Xdazl*	PGCs fail to migrate to gonad from ventral mesoderm; both males and females sterile	60 99
Zebrafish (*Danio rerio*)	*Zdazl*	not yet determined	100
Salamander (Abystoma mexicanum)	*Axdazl*	not yet determined	101
Human (*Homo sapiens*)	*BOULE*	not yet determined	58
Mouse (*Mus musculus*)	*Boule*	not yet determined	58
Fly (*Drosophila melanogaster*)	*boule*	arrest in meiosis I of spermatogenesis	102 103
Nematode (*Caenorhabditis elegans*)	*daz-1*	arrest in meiosis I of oogenesis	104

* Not all homologs of these genes have been listed

** References included here were chosen to provide an entry into the literature and are not comprehensive. We regret the omission of those not listed.

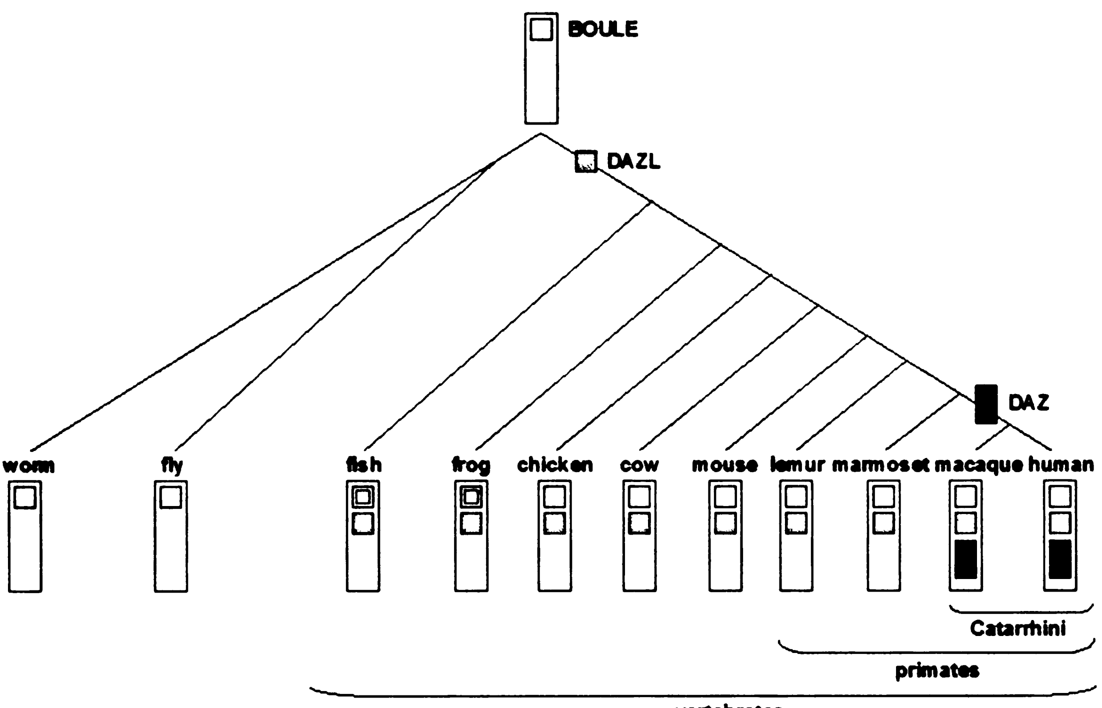
Table 2 Homology of DAZ family genes: % identity to human DAZL

Species	Gene	RRM	DAZ repeat	Overall
Caenorhabditis elegans	*Daz-1*	43%	25%	35%
Drosophila melanogaster	*Boule*	42%	33%	22%
Xenopus laevis	*Xdazl*	85%	50%	60%
Mus musculus	*Dazl*	98%	92%	86%
Homo sapiens	*DAZ*	90%	75%	90%

Figure Legends

Fig. 1 Model of evolutionary history of the *DAZ* gene family. *BOULE* is the ancestral gene and is found in invertebrates and vertebrates. *DAZL*, which has been found in all vertebrate species studied thus far, probably arose by duplication of *BOULE* in the vertebrate lineage, while *DAZ* arose by multiple duplications of *DAZL* in the catarrhine lineage and can only be found in Old World monkeys and hominids. An open box indicates the inferred presence of a *BOULE* homolog that has yet to be identified.

Fig. 1



Chapter 2

Variants in *Deleted in AZoospermia-Like* (*DAZL*) are correlated with reproductive parameters in men and women

Originally published in *Human Genetics*
Published here with kind permission of Springer Science and Business Media

Variants in *Deleted in AZoospermia-Like* (*DAZL*) are correlated with reproductive parameters in men and women

Joyce Y. Tung,*[1] Mitchell P. Rosen,*[1] Lawrence M. Nelson,[2] Paul J. Turek,[1] John S. Witte,[3] Daniel W. Cramer,[4] Marcelle I. Cedars,[1] and Renee A. Reijo Pera[1]

J. Y. Tung, M. P. Rosen, P. J. Turek, M.I. Cedars, R. A. Reijo Pera
Program in Human Embryonic Stem Cell Biology, Center for Reproductive Sciences
Department of Obstetrics, Gynecology, and Reproductive Sciences, Department of Urology
University of California at San Francisco
San Francisco, CA 94143, USA

L. M. Nelson
Developmental Endocrinology Branch
National Institute of Child Health and Human Development
National Institutes of Health
Bethesda, MD 20892, USA

J. S. Witte
Department of Epidemiology and Biostatistics
University of California, San Francisco
San Francisco, CA 94143

D. W. Cramer
OB/gyn Epidemiology Center
Brigham and Women's Hospital
Boston, MA 02115

Address for correspondence and reprints: Dr. Renee A. Reijo Pera, Box 0556, University of California San Francisco, San Francisco, CA 94143. E-mail: reijo@itsa.ucsf.edu. Tel: +1 415 476 3178. Fax: +1 415 476 3121.

*These authors contributed equally.

Keywords: *DAZL*, menopause, spermatogenesis, reproduction, germ cells

Abstract Qualitative and quantitative defects in human germ cell production that result in infertility are common and determined at least in part by genetic factors [1]. Yet, very few genes that are associated with germ cell defects in humans have been identified. In this study, we examined whether variants of the *Deleted in AZoospermia-Like* (*DAZL*) gene are associated with measures of germ cell production in three distinct populations of men and women. We identified 95 sequence variants in *DAZL* and further analyzed twelve single nucleotide polymorphisms (SNPs) that were present across ethnicities. We found that seven of the twelve SNPs were associated with at least one of the parameters studied (age at premature ovarian failure or menopause, total sperm count, or total motile sperm count). Surprisingly, many alleles exhibited opposing effects in men and women, which may be a result of different genetic requirements in male and female germ cells. Single SNP and haplotype analysis suggested that SNPs in the *DAZL* gene may act jointly to affect common reproductive characteristics in the human population.

Introduction

In humans, as in other mammals, male and female germ cells initially develop along a common pathway [105]. Then, after an initial mitotic expansion of premeiotic germ cells, the development of female and male germ cells diverges. The consequences of this divergence are observed postnatally when oocytes are lost through atresia (apoptosis) and to a lesser extent, through ovulation. With no stem cells remaining and the oocyte population depleted, menopause ensues [106]. In contrast, in men, a reserve of germ line stem cells remains throughout life and mature gametes can be produced continuously [105]. One gene that may be required for fertility in both men and women is the *DAZL* (*Deleted in AZoospermia-Like*) gene. *DAZL* is an autosomal homolog of *Deleted in AZoospermia (DAZ)*, a Y chromosome gene deleted in approximately 10% of infertile men [27, 56]. In all animals studied thus far, *DAZL* expression is specific to germ cells, and mutations in *DAZL* homologs are limited to quantitative and qualitative defects in development of the germ cell lineage in one or both sexes [55, 60, 100, 102, 104]. To date, few genes other than *DAZ* gene family members have been identified that have the wealth of information linking their function to germ cell development across such broad evolutionary groups. Thus, we hypothesized that sequence variants in *DAZL* may impact measures of germ cell numbers in men and women, namely sperm count and age at premature ovarian failure or menopause.

Materials and Methods

Study Populations

Sequence variants were identified by direct sequencing of the *DAZL* gene in three study populations (table 1). DNA samples from the first group were collected from women with premature ovarian failure (age range: 13-41 years). In this group, age at menopause was defined as amenorrhea for at least four months, with two serum FSH (follicle stimulating hormone) levels above 40 IU/L one month apart. Causes of ovarian failure such as X chromosome abnormalities were excluded. DNA samples from the second group were collected from women with ovarian failure (menopause) prior to or after the age of 46, as described [5]. By survey, one third of the women in this population reported an early menopause with a family history of early menopause (prior to age 46), one third reported an early menopause and no family history, and the remaining third reported menopause after the age of 46 or were still menstruating at age 46. Samples were obtained only from the women surveyed. DNA samples from the third group were collected from infertile men; each man in this group had a complete history, physical exam, karyotype, semen analysis performed according to WHO criteria, and testicular biopsy when appropriate [107]. In the case of multiple semen analyses with different reported values, the best semen sample was included in statistical models.

200 DNA samples from the DNA Polymorphism Discovery Resource of mixed ethnicity were obtained from the Coriell Institute for Medical Research (Camden, NJ) and used as a sampling of the general population. Although the approximate proportion of each ethnicity within this population is known, the ethnicity of each individual is not known.

DNA extraction and genotyping

DNA extraction was performed as described [108-110]. The entire coding region of *DAZL* was sequenced, including all exons and flanking regions. Primers, primer concentrations, and PCR conditions are as indicated (table 2).

Samples were amplified initially as follows: samples were diluted down to 25ng/uL and amplified with primers designed with Primer3 that were specific to the 11 exons of human *DAZL* [111]. Each PCR reaction contained 2µL of template DNA, 1µL of each primer at 20µM, 2µL of 1mM dNTPs, 2µL of 10X PCR buffer supplied with the enzyme, 1 unit of Taq polymerase (Promega, Madison, WI), and ddH_2O up to 20µL. The PCR program on the Hybaid Omn-E thermal cycler was as follows: 95°C for 10min, 8 cycles of 95°C for 30s, 59°C for 60s, decreasing 0.5°C every cycle, and 72°C for 90s, then 30 cycles of 95°C for 30s, 55°C for 60s, 72°C for 90s, ending with 72°C for 10min. Excess PCR primers and nucleotides were removed by adding 1 unit each of shrimp alkaline phosphatase and exonuclease I and incubating at 37ºC for 1 hour, followed by denaturation of the enzymes by incubation at 90ºC for 15 minutes. Sequencing of the PCR products was performed in 96-well plates with a 10µL reaction: 2.2µL of PCR reaction, 1µL of primer at 10µM, 2µL of Big Dye II sequencing reagent (Applied Biosystems, Foster City, CA), 2µL of HalfBD (formerly GenPak, now Genetix, Boston, MA), and 2.8µL of ddH_2O. The plate was heated in a Hybaid Omn-E thermal cycler for 25 cycles of 95°C for 10s, 50°C for 15s, and 60°C for 4min. The sequencing reactions were then spun through Edge Filter Blocks (Edge Biosystems, Gaithersburg, MD) into optical 96-well plates (Applied Biosystems) to clean out any unincorporated dNTPs.

Filtered sequencing reactions were performed by the University of California San Francisco (UCSF) Genomics Core Facility (GCF) (details below).

Confirmatory amplification and all sequencing of samples was performed at the UCSF GCF. PCR primers were designed using Primer3 with the human repeat mispriming library. Exon primers were designed using *DAZL* sequence aligned with *DAZ* to guarantee *DAZL*-specific amplification; in addition, in some case, such as for Exon 8, internal sequencing primers were required to avoid PCR induced sequence artifacts from poly T stretches flanking the exon. Finally, exon 11 was amplified as a single long range (2077bp) PCR fragment and sequenced using sequencing primers that generated overlapping sequencing reads. Primers, primer concentrations, and PCR conditions are as listed (table 2). One of three PCR reagent cocktails was used. The first used Platinum Taq polymerase (Invitrogen, Carlsbad, CA), containing 20ng DNA template, 50mM KCl, 20mM Tris-Hcl (pH 8.4), 200μM dNTPs, 1.5mM $MgCl_2$, 1.5M Betaine, and 0.25 units Platinum Taq DNA polymerase. The second used AmpliTaq Gold, 20ng DNA template, and 3μl AmpliTaq Gold Master Mix (2X) (Applied Biosystems), and the third used JumpStart AccuTaq LA DNA polymerase mix (Sigma, St Louis, MO) (0.5units), 20ng DNA template, 1μl AccuTaq LA buffer (500 mM Tris-HCl, 150 mM ammonium sulfate (pH 9.3, adjusted with NH_4OH), 25 mM $MgCl_2$, 1% Tween), 200uM dNTPs. Reactions were cycled in GeneAmp 9700 thermal cyclers (Applied Biosystems). All amplicons could be amplified with one of three protocols: 1) "Short touchdown" 5 minutes at 95°C, 10 cycles of 94°C for 20s, 61°C for 20s, decreasing 0.5°C every cycle, and 72°C for 45s, then 35 cycles of 94°C for 20s, 56°C for 20s, and 72°C for 45s, ending with 72°C for 10min; 2) "Long touchdown" 5 minutes at 95°C, 14 cycles of 94°C for 20s, 63°C for 20s,

decreasing 0.5°C every cycle, 72°C for 45s, then 35 cycles of 94°C for 20s, 56°C for 20s, 72°C for 45s, ending with 68°C for 10min; or 3) "LA touchdown" 5 minutes at 95°C, 14 cycles of 94°C for 20s, 63°C for 20s, decreasing 0.5°C every cycle, 68°C for 2min, then 35 cycles of 94°C for 20s, 56°C for 20s, 68°C for 2min, ending with 68°C for 10min. Sequencing was carried out as previously described [108]. Briefly, excess PCR primers and nucleotides were removed enzymatically with shrimp alkaline phosphatase and exonuclease I as described above. The sequencing was carried out in 5μl reactions in a 384 well GeneAmp 9700 thermal cycler using BigDye Terminators v3.1 Cycle Sequencing Kit (Applied Biosystems) and analyzed on ABI 3700 or 3730xl DNA Analyzer capillary electrophoresis platforms (Applied Biosystems). Sequencher v4.1.4 (Gene Codes, Ann Arbor, MI), Seqman (DNAStar Inc., Madison, WI) and Mutation Surveyor v2.1 (SoftGenetics, State College, PA) were used to align sequences and to identify polymorphic bases.

Statistical Analysis

Initial descriptive analyses using χ^2 and t-tests were performed for each SNP to assess allele frequencies and to determine whether the distribution patterns were in Hardy-Weinberg equilibrium. Average age at premature ovarian failure or menopause, average total sperm count (= sperm concentration x volume), and average total motile count (= total sperm count x % motility) were calculated for all genotypes at each SNP. Each group was then analyzed separately by linear regression, controlling for ethnicity. Each SNP was initially coded as co-dominant to minimize assumptions with respect to mode of

inheritance. Subsequently SNPs were coded based on the risk best reflected from the co-dominant model to evaluate the data in the most parsimonious fashion.

To quantify the degree of linkage disequilibrium (LD) between all pairs of SNPs, the linkage disequilibrium coefficient, D', was calculated using the Haploview program [112]. Estimated haplotype frequencies were inferred using Phase v2.1 [113, 114]. Linear regression analyses were used to evaluate the association of the estimated haplotypes with the age at menopause, and total sperm and total motile counts. Only haplotypes with population frequencies > 1% were included and each analysis was controlled for ethnicity.

All tests were two-sided, and considered statistically significant when the corresponding p-value < 0.05. The statistical analyses were performed using Stata version 7.0 (Stata Corporation, College Station, TX).

Sequence analysis

Alignment of mouse and human *DAZL* 3'UTRs and the dot plot were performed using MegAlign (DNASTAR, Madison, WI, USA). For the dot plot, the minimum percentage match was set to 80%, and the window size was set to 20 nucleotides.

Results

Initial identification and analysis of DAZL SNPs

We directly sequenced the *DAZL* gene in four populations: three distinct study populations and one panel of DNA from the general population. The first study population contained 93 women diagnosed with premature ovarian failure (average age = 29.7 years), while the second study population contained 324 women with early menopause (average age = 42.2 years) [5, 115]. The third study population contained 102 infertile men with few or no sperm (oligozoospermia, < 20 million sperm/ml), and/or immotile sperm (asthenozoospermia, < 50% motile sperm) (infertile male group). The fourth population, from the Coriell Institute's Polymorphism Discovery Resource, which is designed to reflect the diversity in the human population, consisted of individuals from the general population and was only used for sequence variant discovery as the ethnicity of each individual was not known. In these groups, we identified 99 sequence variants within the eleven exons and flanking regions of *DAZL* (fig. 1A). A comparison of all the variants revealed twelve SNPs (fig. 1B) that were in Hardy-Weinberg equilibrium, had allele frequencies > 1%, and were found across ethnic groups (data not shown). Allele and genotype frequencies for each population are shown (fig. 2, table 3, supplementary table 1). Of the twelve SNPs, one results in a non-synonymous amino acid substitution, six alter nucleotides of the 3'UTR, and five map to introns.

As allele frequencies differed across ethnic groups (fig. 2), we first focused our analysis on the largest ethnic group in this study, Caucasians. In addition to genotype frequencies, we calculated the average age at ovarian failure/menopause, total sperm count, and total motile count in the study groups with respect to SNP genotype and found evidence that several SNPs were associated with these reproductive parameters (table 3, table 4). To control for ethnicity, regression analyses that included all ethnic groups were

used to determine the strength of the associations (table 5). Each population was considered separately and each SNP was analyzed to determine if it was an independent predictor for age at ovarian failure/menopause and total sperm count and total motile count, while controlling for ethnicity. Notably, seven of the twelve SNPs analyzed were significantly associated with at least one of the parameters studied. Remarkably, these were the seven SNPs present in mature *DAZL* mRNA, whereas the five SNPs that had no significant associations ($p > 0.05$) were the five that mapped to the introns. Note however that two of the intronic SNPs, located at positions 8496 and 14586, are marginally associated with lower total motile count in the infertile male group (9.5 million decrease in total motile count, $p = 0.059$) and an earlier age at ovarian failure in the most severely affected female group (3.6 year decrease in age at ovarian failure, $p = .077$), respectively—findings that might become statistically significant with larger sample sizes. Here we will concentrate on the SNPs with significant associations as defined in Materials and Methods ($p <= 0.05$).

A *DAZL* coding SNP

First, an a→g transition that results in a Thr→Ala change just upstream of the RNA binding domain in exon 2 was identified. Homozygosity for the 8175g allele was associated with an earlier age at ovarian failure in the most severely affected female group, and conversely with an increase in both total sperm count and total motile count in men, with the coefficients indicating the magnitude of association relative to the reference allele (table 4, table 5). The 8175g allele was significantly associated with a 4-

year decrease in the age at ovarian failure in the premature ovarian failure group (p = 0.024), and an increase of 21 million in total sperm count (p = 0.026) and 10 million in total motile count (p = 0.03) in the infertile male group. The co-dominant model revealed that the magnitude of the degree of association was similar for a heterozygous or homozygous change (data not shown).

DAZL SNPs in the 3'UTR

Second, three SNPs located in the 3' UTR, located at positions 18929 (t→a), 19218 (c insertion), and 19936 (g→a), were associated with an increase in total sperm count and total motile count in the infertile male group but were not associated with age at ovarian failure/menopause in females (table 4). The 18929a, 19218ins_c, and 19936c alleles are almost in complete linkage disequilibrium as evidenced by allele dropout due to colinearity, and are associated with an increase of 15 million in total sperm count (p = 0.015), and an increase of 6 million in total motile count (marginal significance, p = 0.05, 0.06, and 0.06). The strength of association is increased with homozygosity, consistent with a recessive model of inheritance (data not shown). In addition, a fourth SNP, the t18932_ SNP, also located in the 3'UTR, demonstrated an association with sperm parameters similar to those of the co-segregrating alleles described above. It was associated with an increase in total sperm count and total motile count in Caucasians, and in all ethnic groups, a 14.7 million increase in total motile count (p = 0.042) and a corresponding 27 million increase in total sperm count (p = 0.058) was observed.

Finally, in Caucasians, two additional 3'UTR SNPs at positions 18645 (a→g) and 18976 (c→t) were associated with increased total sperm count and total motile count, but not age at ovarian failure (table 4). The regression analysis demonstrated, however, that the association observed in Caucasians is not evident when the entire study group is analyzed. Instead, contrary to observations in Caucasians, an association with age at ovarian failure was observed. The 18645g and 18976t alleles corresponded to a four-year decrease in the age at ovarian failure in the premature ovarian failure group ($p=0.040$, $p = 0.018$). The co-dominant model suggests that these alleles may be acting in a recessive manner: a significant association is only observed when 18645g and 18976t are homozygous (data not shown).

As SNPs in the 3'UTR do not change the amino acid sequence coded by the mRNA, it is often difficult to determine the function of these SNPs. Nucleotides that are conserved across species are more likely to play an important functional role, so we compared the 3'UTRs of human and mouse *DAZL* to look for conserved sequences (fig. 3). Overall, the mouse and human *DAZL* 3'UTRs exhibit 63% similarity. All six 3'UTR SNPs analyzed in this study were found in regions of high conservation and for each SNP, one of the two alleles was identical to the equivalent nucleotide in the mouse sequence. This data suggests that these 3'UTR SNPs may be functional and could be responsible for the phenotypes observed.

Haplotype analysis

To extend our analysis, we next determined the extent of linkage disequilibrium (LD) between the SNPs in the *DAZL* gene. All of the SNPs within *DAZL* exhibited some LD, with D' values ranging from 0.1 to 1.0 and varying between each study population; however, SNPs t18929a, _19218c, and g19936c were in nearly complete disequilibrium (data not shown). Thus, we were able to use SNP g19936c to represent SNPs t18929a and _19218c when constructing haplotypes. Examining haplotypes, or groups of closely linked genetic markers which tend to be inherited together, allowed us to evaluate combined effects of multiple SNPs. Only haplotypes with a population frequency > 1%, which accounted for 27-50% of the estimated haplotypes, depending on the population being analyzed, were considered when analyzing the association with age at ovarian failure/menopause, total sperm count, and total motile count. The reference haplotype was chosen because of its high frequency in the population and its combination of reference alleles. Then, eleven, ten and nine haplotypes were modeled for the premature ovarian failure, ovarian failure/menopause, and infertile male groups, respectively. Significant haplotype associations were observed in the premature ovarian failure and infertile male groups (fig. 4). In the premature ovarian failure group, haplotype H2 had a population frequency of 11.8%, and showed a strong association with lower age at ovarian failure, decreasing it by 5.6 years ($p = 0.002$). Haplotype H4 (population frequency 8.6%) was also associated with a decrease in the age at ovarian failure in the premature ovarian failure group (4.9 years, $p = 0.024$). In the infertile male group, haplotypes H2 and H1 were associated with an increase of 8.6 million and 9.5 million in total motile count, respectively ($p = 0.038$ and 0.004). Similar results were observed for total sperm count (data not shown).

Discussion

Our results show that polymorphisms in the *DAZL* gene may contribute to variation in age at ovarian failure/menopause and total sperm and total motile counts in clinical populations. Significant associations were found in more than one population, and in both sexes, lending strength to this conclusion. Previous studies of polymorphisms in the *DAZL* gene only examined two SNPs, one of which showed an association with male infertility but has only been found in the Taiwanese population [70-72]. In our study we have identified seven SNPs that are associated with reproductive parameters and found in multiple ethnic groups. Notably, six of the seven significant SNPs were located in the 3'UTR and were found in regions of high conservation from mouse to human. As the expression of many genes required for germ cell development is regulated through the 3'UTRs of their mRNAs, these six *DAZL* SNPs may affect the binding of regulatory proteins to *DAZL* mRNA, thus changing mRNA stability or translation.

Past work on the genetics of human reproduction has generally focused on rare, amino acid changing mutations that severely impair protein function but are not a common cause of reproductive failure [11, 16, 30, 116]. In contrast to these rare variants, common variants are expected to confer small but real effects on reproductive parameters. Consequently, two women carrying the same allele may experience menopause at different ages, depending on their genetic backgrounds. Thus, to uncover common variant-common phenotype relationships, it is advantageous to examine variation over the entire spectrum of phenotypes, rather than in artificially-constructed categories of

broad phenotypic groups. In this way, small associations may be less easily overlooked (fig. 5).

Initially, we hypothesized that variants in *DAZL* would affect male and female reproduction similarly: either increasing sperm count and age at ovarian failure/menopause, or decreasing both. Surprisingly, however, the non-reference alleles of several SNPs were associated with increased sperm counts in males but decreased ages at ovarian failure/menopause (fewer oocytes) in females (fig. 6). Since mutation or knockdown of *DAZL* in many organisms leads to loss of both sperm and oocytes, the effect of these human SNPs must be more complex. The stability or translation of *DAZL* mRNA may be affected by the different complements of proteins in male and female germ cells. Alternatively, as DAZL is also known to interact with several protein partners and help initiate translation in conjunction with the poly(A)-binding protein, sex-specific effects of *DAZL* SNPs may be due to quantitative/qualitative differences in translation initiation complexes including DAZL in males and females [64, 66, 67, 117]. Regardless, one might have expected that alleles that potentially reduce the number of offspring a female produces would be eliminated from the population. However, only in recent history has women's life expectancy extended substantially beyond the reproductive window [118]. Interestingly, if women continue the current trend of postponing childbearing, alleles that decrease age at ovarian failure/menopause may impact the number of children born and reduce the prevalence of those alleles in the population. This study opens the door for further analysis by identifying common *DAZL* SNPs and emphasizing the importance of fully characterizing common haplotypes in complex traits.

Acknowledgments We thank Anne Chang, Nicole Morison, and Vladimir Bakalov for technical assistance, Vanessa Angeles, Rhoda Chang, Anthony Dobson, Mark Fox, Joanna Gonsalves, Kehkooi Kee, Fred Moore, and Cory Nicholas for helpful comments and criticisms, and the UCSF GCF personnel for assistance with sequencing and analysis. This work was supported by fellowships from the National Science Foundation and Achievement Rewards for College Scientists to JYT, and by grants from the Sandler Family Foundation, the Searle Scholar Program and the National Institutes of Child Health and Human Development (R01 HD37095 and R01 HD047721-01A1) to RARP.

Table 1 Baseline characteristics for each study population

	premature ovarian failure group (N=93)	Ovarian failure/menopause group (N=324)	infertile male group (N=102)
Age range (avg)	13-41 y/o (29.7)	28-55 y/o (42.2)	22-51 y/o (36.4)
Ethnicity			
% Caucasians	76%	98.4%	65%
% Latinos	13%	0%	9%
% Asians	6.5%	1%	21%
% African Americans	4.5%	0.6%	5%
Karyotype	XX	Not tested	XY
Origin of population	NIH [110]	Massachusetts [5]	UCSF [108]

Table 2 Primer sequences for amplifying and sequencing the 11 *DAZL* exons

PCR PRIMER SEQUENCES				PCR CONDITIONS		
EXON	*FORWARD*	*REVERSE*	*SIZE (bp)*	*PCR CYCLE*	*PCR RXN MIX*	*PRIMER CONC.*
1	GATGTGTGTGTCTGTGGGTCTA	GCCGAGGATGACTTCACTTT	547	LONG TD	Platinum Taq	332 nM
2	CAAGGGCGAGACTCTCTCAA	TGGGTCAAATGTAAAACCAATTCT	494	LONG TD	Amplitaq GOLD	166nM
3	CTACTGGTCATTTGGGGATAAA	AAGTGTCAGAATACCAATTTTGAAG	294	SHORT TD	Amplitaq GOLD	166nM
4	TCAGATGAAGAAAAAGCAGTATTTG	TAGGTGCCCAGTCAAAAATAAA	397	SHORT TD	Amplitaq GOLD	166nM
5	GAAGACTTTAGGTGTTCATCCAAG	GGTGGAGGAGGATGATTAAAAA	294	SHORT TD	Amplitaq GOLD	166nM
6	TTGTTACTGTTTTATTTCAGTCACAG	AATTCCACAGAAGGTACGATGA	386	SHORT TD	Amplitaq GOLD	166nM
7f[a]	TGCATTTAGAATTCTGCATTGTT	GCTATTTGACAGCAAAGATGAAA	494	SHORT TD	Amplitaq GOLD	166nM
7m[a]	TCTGGGTGGGAATAAGTACACA	GCTATTTGACAGCAAAGATGAAA	441	LONG TD	Amplitaq GOLD	166nM
8	CAAGAACATCAGGGATTAGAAACATA	CCCATGGGAGTTTCTAACTGG	397	SHORT TD	Amplitaq GOLD	166nM
9	TGATACCAGCTTAGCTTTTTGAA	ACAAATTTCTGCTGAAGGATGA	299	SHORT TD	Amplitaq GOLD	166nM
10	CCACTCCCAGTCTCTTTTAGCAT	CCAGGTAAAACCCACTCTGGT	396	SHORT TD	Amplitaq GOLD	166nM
11	CCAAAACAAATGGTTTAAGTTTTT	CACAGTTCCCACCCAACTCC	2077	LA TD	JumpStart AccuTaq LA	166nM

SEQUENCING PRIMER SEQUENCES			
EXON	*SEQUENCE*	*DIRECTION*	*AMPLICON*
8	AAAAAAATACTTTAAAATAA	R	8
11	CCAAAACAAATGGTTTAAGTTTTT	F	11T1[b]
11	AATCTGTTTTGCTGGGTTTAGA	R	11T1
11	TGGCAGATCAAATGGTACTGAT	F	11T2[b]
11	CCTATTTTGAAAGGTCAGAA	F	11T2
11	TAAACCCAGCAAAACAGATTCA	R	11T2
11	GCAGTTCAAGGGCCAAACC	F	11T3[b]
11	CACAGTTCCCACCCAACTCC	R	11T3

[a]Initial primers for exon 7 (7f) were used on samples from women, and additional primers (7m) that were *DAZL*-specific (no spurious amplification of the Y-chromosome *DAZ* genes) were subsequently designed specifically for use in men.

[b]Due to its length, the amplicon for exon 11 was sequenced in three parts, denoted 11T1, 11T2, and 11T3.

Table 3 Genotype frequencies for each SNP in Caucasians by population

SNP genotype	*Premature ovarian failure group*	*Ovarian failure/menopause group*	*Infertile male group*
8175 a/a	71.4%	79.4%	71.4%
a/g	22.9%	20.3%	19.6%
g/g	5.7%	0.3%	8.9%
8496 a/a	15.9%	19.9%	22.2%
a/c	42.9%	44.7%	44.4%
c/c	41.3%	35.4%	33.3%
9287 a/a	80.0%	75.2%	86.0%
a/c	18.5%	23.5%	12.3%
c/c	1.5%	1.3%	1.8%
9783 t/t	79.7%	75.0%	83.9%
t/c	20.3%	23.6%	14.3%
c/c	0.0%	1.4%	1.8%
11465 a/a	100.0%	74.7%	3.3%
a/g	0.0%	24.2%	13.0%
g/g	0.0%	1.0%	3.7%
14586 g/g	81.5%	79.8%	87.9%
g/a	18.5%	19.2%	12.1%
a/a	0.0%	1.0%	0.0%
18645 g/g	83.3%	76.6%	90.0%
g/a	14.6%	20.6%	7.5%
a/a	2.1%	2.8%	2.5%
18929 t/t	13.1%	11.8%	13.2%
t/a	39.3%	40.3%	43.4%
a/a	47.5%	47.9%	43.4%
18932 t/t	78.1%	83.2%	86.8%
t/_	21.9%	15.8%	13.2%
_/_	0.0%	1.0%	0.0%
18976 c/c	79.4%	77.7%	88.2%
c/t	15.9%	20.4%	9.8%
t/t	4.8%	1.8%	2.0%
19218 _/_	12.5%	11.3%	13.2%
_/c	35.9%	40.3%	41.5%
c/c	51.6%	48.3%	45.3%
19936 g/g	11.9%	10.9%	13.2%
g/c	37.3%	36.8%	41.5%
c/c	50.7%	52.3%	45.3%

Table 4 Average age at premature/ovarian failure/menopause, total sperm count, and total motile count by genotype for the 12 *DAZL* SNPs in Caucasians

	Premature ovarian failure group	Ovarian failure/menopause group	Infertile male group	
SNP genotype	Age at ovarian failure (yrs)	Age at ovarian failure/menopause (yrs)	Total sperm count (millions)	Total motile count (millions)
8175 a/a	28.3 (71%)[a]	44.7 (77%)	14.7 (71%)	4.7
a/g	26.2 (23%)	44.3 (23%)	14.8 (20%)	2.9
g/g	20.3 (6%)	None (0%)	57.6 (9%)	28.4
8496 a/a	25.7 (16%)	44.6 (20%)	17.3 (22%)	3.8
a/c	25.2 (43%)	44.7 (45%)	22.2 (44%)	6.6
c/c	28.5 (41%)	44.7 (35%)	26.0 (33%)	10.9
9287 a/a	28.7 (80%)	44.8 (75%)	21.2 (86%)	7.4
a/c	20.2 (18%)	44.5 (24%)	32.1 (12%)	8.1
c/c	25.0 (2%)	43.5 1%)	42.0 (2%)	11.0
9783 t/t	28.7 (80%)	44.8 (75%)	20.5 (84%)	7.1
t/c	20.5 (20%)	44.6 (24%)	31.6 (14%)	8.9
c/c	None (0%)	43.5 (1%)	42.0 (2%)	11.0
11465 a/a	27.4 (100%)	44.7 (75%)	16.1 (83%)	6.1
a/g	None (0%)	44.7 (24%)	30.7 (13%)	8.1
g/g	None (0%)	43.7 (1%)	40.0 (4%)	12.5
14586 g/g	29.3 (82%)	44.8 (80%)	15.9 (88%)	5.9
g/a	19.6 (18%)	44.2 (19%)	32.1 (12%)	8.1
a/a	None (0%)	42.3 (1%)	None (0%)	None
18645 g/g	31.0 (83%)	44.7 (77%)	8.1 (90%)	2.4
g/a	17.0 (15%)	44.2 (21%)	48.0 (8%)	13.3
a/a	29.0 (2%)	43.3 (3%)	42.0 (2%)	11.0
18929 t/t	29.1 (13%)	45.2 (12%)	0.1 (13%)	0.0
t/a	27.0 (39%)	44.7 (40%)	10.5 (43%)	3.8
a/a	28.1 (38%)	44.5 (48%)	30.2 (43%)	11.1
18932 t/t	27.3 (78%)	44.6 (83%)	12.3 (87%)	3.7
t/_	27.1 (22%)	44.8 (16%)	52.6 (13%)	24.9
_/_	None (0%)	45.7 (1%)	None (0%)	None
18976 c/c	29.1 (79%)	45.0 (78%)	16.5 (88%)	6.4
c/t	17.3 (16%)	44.3 (20%)	29.4 (10%)	8.0
t/t	27.0 (5%)	43.6 (2%)	42.0 (2%)	11.0
19218 _/_	29.1 (12%)	45.3 (11%)	0.1 (13%)	0.0
_/c	26.8 (36%)	44.9 (40%)	10.1 (42%)	3.8
c/c	27.2 (52%)	44.4 (48%)	29.8 (45%)	10.8
19936 g/g	29.1 (12%)	44.7 (11%)	0.1 (13%)	0.0
g/c	26.8 (37%)	44.7 (37%)	10.1 (42%)	3.8
c/c	27.2 (51%)	44.4 (52%)	29.8 (45%)	10.8

[a]The percentage of patients in each group with the specified genotype is indicated in parentheses.

Table 5 The impact of the 12 *DAZL* SNPs on reproductive parameters.

	premature ovarian failure group		***ovarian failure/menopause group***		***infertile male group***			
	age at ovarian failure (yrs)		*age at ovarian failure/menopause (yrs)*		*total sperm count (millions)*		*total motile count (millions)*	
SNP	Assoc. Coeff.[a]	p-value[b]	Assoc. Coeff.[a]	p-value[b]	Assoc. coeff.[a]	p-value[b]	Assoc coeff.[a]	p-value[b]
8175	-3.944	0.024	-0.469	0.457	21.741	0.026	10.595	0.030
8496	1.185	0.464	-0.117	0.739	-15.150	0.192	-9.537	[illegible]
9287	-2.278	0.241	-0.229	0.670	15.745	0.597	0.387	0.977
9783	-2.329	0.239	-0.297	0.590	18.085	0.535	1.618	0.899
11465	6.541	0.192	-0.227	0.717	10.751	0.361	2.394	0.689
14586	-3.569	[illegible]	-0.064	0.975	13.638	0.224	2.780	0.625
18645	4.717	0.040	0.669	0.708	-13.920	0.285	-1.437	0.825
18929	-3.116	0.218	-0.141	0.873	15.635	0.015	6.367	[illegible]
18932	1.020	0.614	0.984	0.719	27.223	[illegible]	14.724	0.042
18976	4.561	0.018	-0.039	0.983	-20.536	0.400	-2.217	0.858
19218	-3.327	0.202	-0.213	0.813	15.557	0.015	6.244	[illegible]
19936	-3.431	0.180	-0.084	0.874	15.557	0.015	6.244	[illegible]

[a]Association coefficients indicate the magnitude of association of the allele shown to each reproductive parameter, relative to the reference allele.

[b]P-values < 0.05 are highlighted in black, p-values > 0.05 and < 0.1 are highlighted in gray. Analysis was adjusted for ethnicity.

Figure Legends

Fig. 1 Structure and SNPs in *DAZL* mRNA. (a) Structure of *DAZL* mRNA. The *DAZL* mRNA consists of 11 exons (red arrows), with exons 2-10 making up the majority of the coding region (orange arrow). Within the coding region, there are two conserved domains: the RNA Recognition Motif (RRM), and the DAZ repeat (DAZ rpt) (green arrows). (b) Locations of SNPs in the *DAZL* gene. The open box represents the open reading frame, with solid arrows corresponding to the 5' and 3' untranslated regions. Coding exons are depicted by solid orange boxes. Each SNP is labeled with the exon it is located in or closest to, and its nucleotide position relative to a reference *DAZL* sequence (Supplementary Note), with exonic SNPs labeled in black, intronic SNPs in blue.

Fig. 2 SNP characteristics and allele frequencies. For each SNP, its location within the gene, the nucleotide change involved, the effect of that change on the protein, and its SNP allele frequency in each population by ethnicity, is tabulated. For the description of the position of each SNP, its nucleotide position in (**bold**), before (-), or after (+) the nearest exon is noted in parentheses. '*' indicates that the zero allele frequency is a result of zero individuals of that ethnicity in that population. POF = premature ovarian failure group, IM = infertile male group, EM = ovarian failure/early menopause group.

Fig. 3 Dot plot alignment of mouse and human 3'UTR. The 3'UTRs of the mouse (accession #BC099940) and human *DAZL* were aligned and displayed as a dot plot. The human sequence is on the x-axis and the mouse on the y-axis, with the scale representing

nucleotide position. Diagonal lines showing regions of conservation, with red lines denoting higher alignment scores and blue lines lower scores. The positions of the six 3'UTR SNPs are indicated by the white arrows. From upper left to lower right, the arrows point to: SNP 18645, SNP 18976, SNPs 18929 and 18932 (these two SNPs are too close to each other to distinguish on the graph), SNP 19218, and SNP 19936.

Fig. 4 Results of haplotype linear regression analysis. Haplotypes that were significantly associated with age at menopause in the premature ovarian failure group (a) and total motile count in the infertile male group (b) are described. Allele differences between the significant haplotypes and the reference haplotype are in bold. The p-value associated with the strength of association follows the coefficient in parentheses. For the infertile male group, similar results were found for total sperm count.

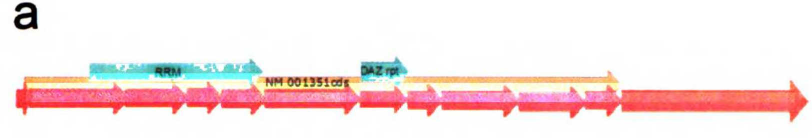
Fig. 5 Model for how the combined strength of multiple common variants can lead to a spectrum of phenotypes. Many phenotypes, such as male and female fertility, are measured by continuous variables like sperm concentration and age at ovarian failure/menopause. These measurements are continuums, even if arbitrary cutoffs of disease have been established. Instead of looking solely for rare variants that cause serious disease and are not found in "unaffecteds", it may be informative to consider each characteristic and disease as a spectrum on which every individual must fall. The combined effects of common alleles from multiple genes can push an individual back and forth on that spectrum. For example, the combination of three negative alleles and one

positive allele could adjust age a woman's age at menopause from the average age to cause premature ovarian failure at age 36.

Fig. 6 Opposing effects of the same allele in men and women. Several *DAZL* SNP alleles that are potentially beneficial to one sex (higher sperm counts, later age at menopause) are simultaneously detrimental to the other sex (earlier age at menopause, lower sperm counts). While higher sperm counts have clear evolutionary advantage for males, earlier ages at menopause should be a disadvantage for females. One possibility is that these female reproductive defects were not selected against until more recently, when dramatically lengthened lifespans and the ability to postpone pregnancy has allowed women not only to live past menopause but also to push the limits of the reproductive window.

Fig. 1

a



b

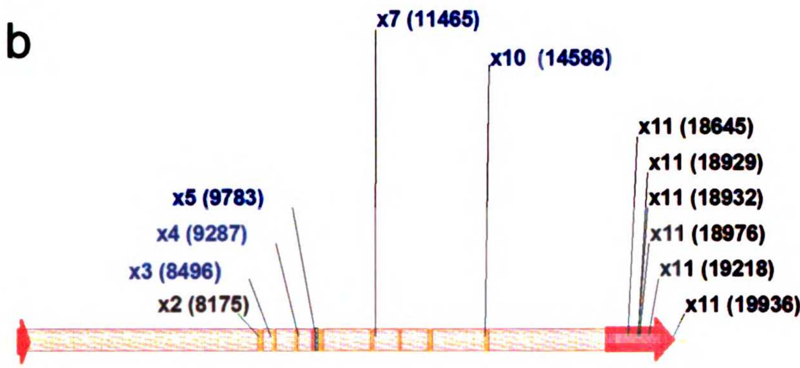


Fig. 2

SNP CHARACTERISTICS					ALLELE FREQUENCY (%)											
					POF	EM	IM	POF	EM	IM	POF	EM	IM	POF	EM	IM
SNP	Exon/Intron	Position	Change	Type	Caucasian			African-Amer.			Hispanic			Asian		
8175	exon2	exon2 (34)	a>g	Thr-Ala	17.1	10.5	18.1	4.2	0	20	0	0*	16.7	0	0	14.7
8496	intron	exon3 (-)69	a>c	intron	62.7	57.7	55.6	63.6	74.5	60	62.5	0*	50	10	16.7	54
9287	intron	exon4 (+)34	a>c	intron	10.8	13.1	2.9	13.6	0	10	12.5	0*	33.3	80	66.7	8.3
9783	intron	exon5 (+)28	t>c	intron	10.2	13.2	8.9	13.6	0	10	12.5	0*	33.3	75	66.7	10
11465	intron	exon7 (+)36	a>g	intron	0	13.1	10.2	0	0	12.5	0	0*	37.5	40	66.7	11.1
14586	intron	exon10 (-)9	g>a	intron	9.2	10.6	6	12.5	0	0	25	0*	22.2	75	50	13.2
18645	exon11	exon11 (1444)	g>a	3'UTR	9.4	13.1	6.2	15	0	10	33.3	0*	33.3	80	66.7	7.1
18929	exon11	exon11 (1728)	t>a	3'UTR	67.2	68.1	65.1	95.8	75	80	83.3	0*	87.5	91.7	83.3	64.7
18932	exon11	exon11 (1731)	t>_	3'UTR	10.9	8.9	6.6	4.2	0	0	0	0*	6.3	0	0	0
18976	exon11	exon11 (1775)	c>t	3'UTR	12.7	12	6.9	12.5	0	10	33.3	0*	31.2	75	66.7	8.8
19218	exon11	exon11 (2017)	_>c	3'UTR	69.5	68.5	66	95.8	75	80	83.3	0*	87.5	91.7	75	64.7
19936	exon11	exon11 (2735)	g>c	3'UTR	69.5	70.7	66	95.8	75	80	83.3	0*	87.5	91.7	66.7	64.7
				N=	71	319	66	12	2	5	4	0	9	6	3	22

Fig. 3

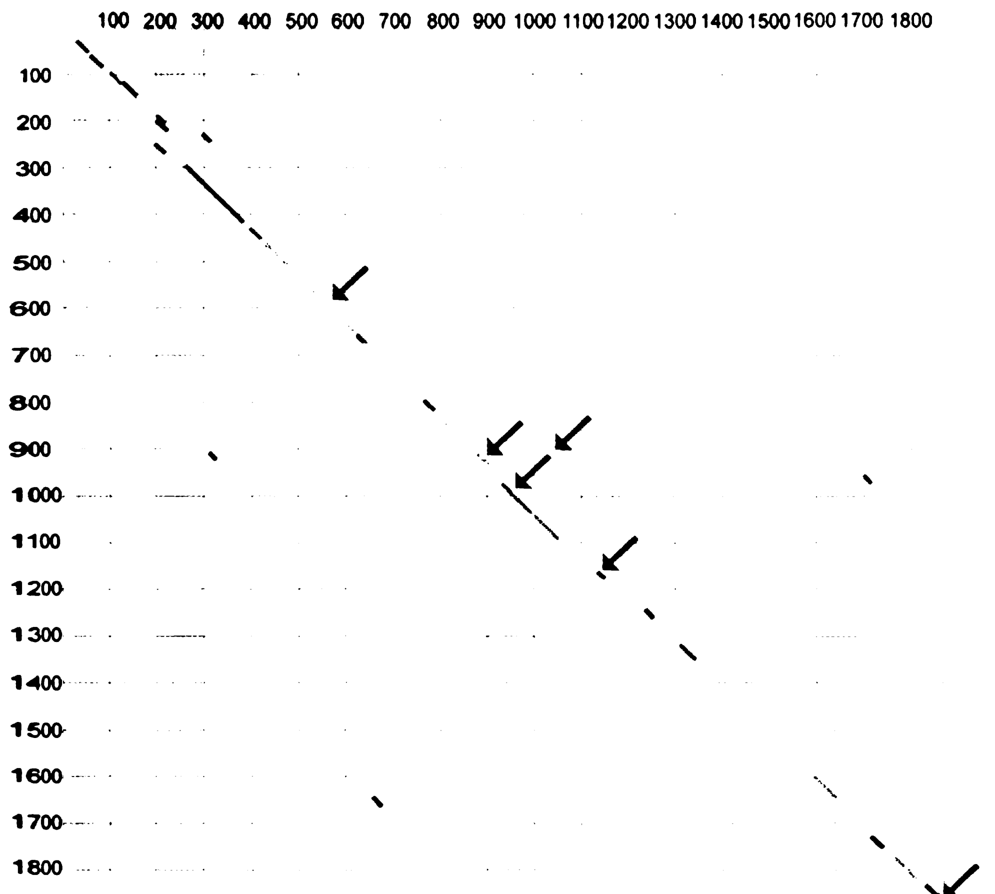


Fig. 4

a

Haplotype	8175	8496	9287	9783	11465	14586	18645	18932	18976	19936	Population frequency	Strength of Association
Reference	a	a	a	t	a	g	g	t	c	g	20 4%	--
H2	g	c	a	t	a	g	g	t	c	c	11 8%	-5 6 (0 002)
H4	a	a	c	c	a	a	a	t	t	c	8 6%	-4 9 (0 024)
Other	-	-	-	-	-	-	-	-	-	-	59 2%	not significant

b

Haplotype	8175	8496	9287	9783	11465	14586	18645	18932	18976	19936	Population frequency	Strength of Association
Reference	a	a	a	t	a	g	g	t	c	g	32 2%	--
H1	a	c	a	t	a	g	g	t	c	c	40 3%	9 5 (0 004)*
H2	g	c	a	t	a	g	g	t	c	c	8 7%	8 6 (0 038)*
Other	-	-	-	-	-	-	-	-	-	-	18 8%	not significant

Fig. 5

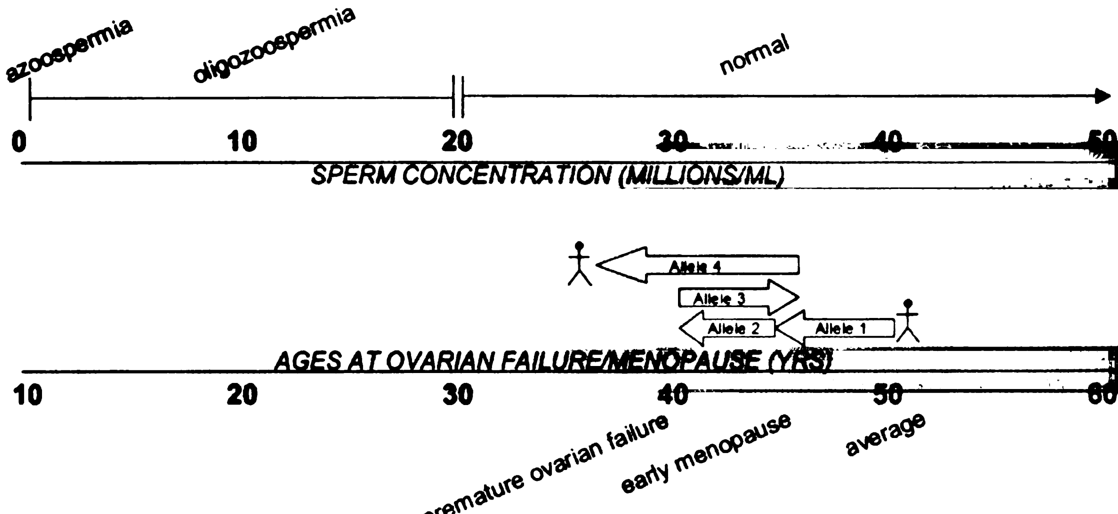
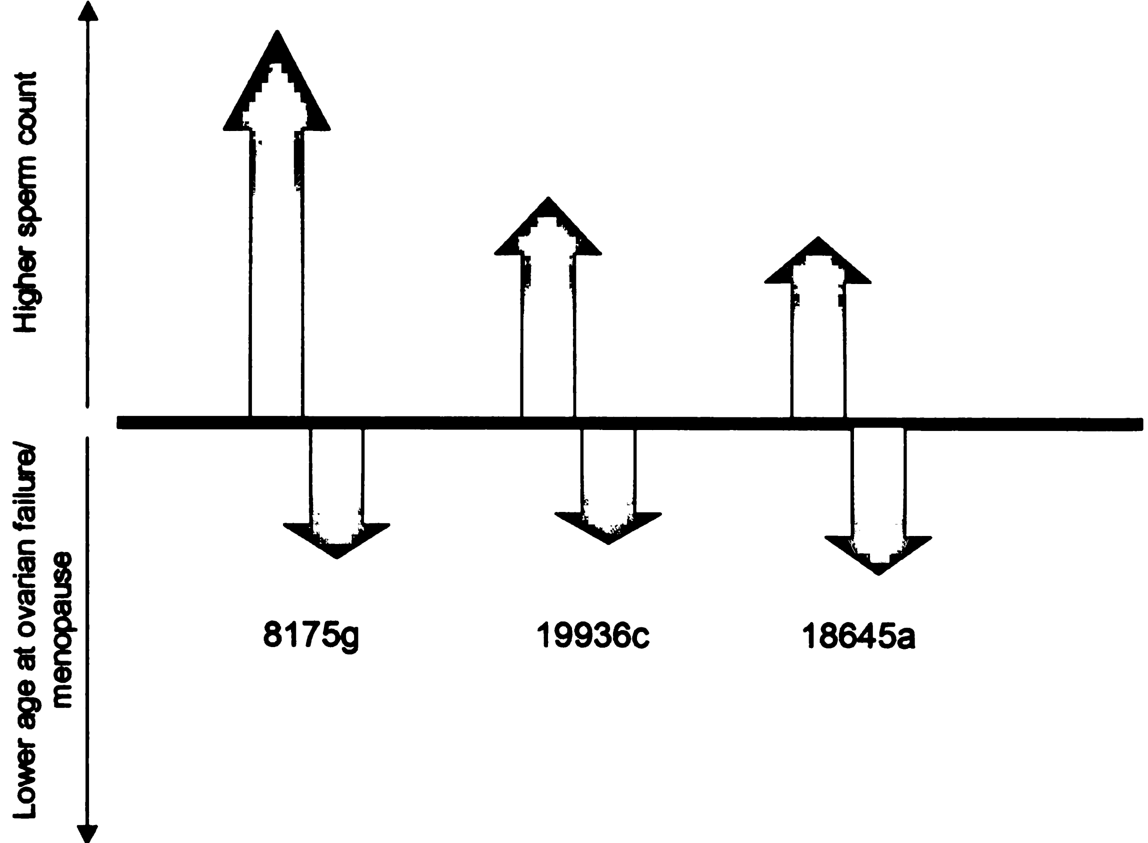


Fig. 6



Supplementary Table 1 Genotype frequencies for each SNP in the Coriell population

SNP genotype	*Coriell panel*
8175 a/a	84.5%
a/g	2.7%
g/g	12.8%
8496 a/a	20.9%
a/c	49.0%
c/c	30.0%
9287 a/a	66.7%
a/c	28.5%
c/c	4.8%
9783 t/t	67.8%
t/c	27.3%
c/c	4.9%
11465 a/a	68.2%
a/g	28.5%
g/g	3.4%
14586 g/g	71.5%
g/a	24.4%
a/a	4.1%
18645 g/g	64.3%
g/a	28.6%
a/a	7.1%
18929 t/t	4.0%
t/a	36.4%
a/a	59.7%
18932 t/t	96.7%
t/_	3.3%
_/_	0.0%
18976 c/c	70.2%
c/t	25.4%
t/t	4.4%
19218 _/_	2.8%
_/c	35.6%
c/c	61.6%
19936 g/g	3.9%
g/c	34.8%
c/c	61.2%

Supplementary Note

Reference *DAZL* sequence used for identifying locations of SNPs. This sequence was originally obtained from Genbank (accession #NT_005927), but it is no longer available through NCBI as the record was removed as a result of standard genome annotation processing.

```
atttagggacaatggcgttttgcagagtactcagttccagctcccaccctgagcccttacctcctaccac
taaacatccagatctcaagttacagagagtaacaacactttgcccacacctaggatgaccaatccttctgg
tttccttagggacgtaggaatttctcagtgctaaaatcatgtaagtcctgggcaaactagtatgactggct
gccctacagcccagcccccatatcccgggcaggagaagaatcagttgaataagtcattgctctggttttca
gtttattttggtacttggaatttcccttcttttcttatgggtttagctgcatattgaaaataaaacgttt
tattttaatcccgcattactaggtgttagtaatgagatactgctcacgttatagaagtccaccgttttgcc
agctcaaaagtaacattttaatgctttccgtcagtacagaagtgaagtttgtggataaaggtcatagcta
cttagccctgtgtgtatagctaaggtgctgcagtggttcccgaagttacgtgatgggctccaccaactggc
ccacagcttgctcctgctctacgcccgtattctgaccttaagaatgttctagaattctctaggcaggccta
gcgccttcctgtaaggcctggaaggtgaattctcccattaaaatgtaagatccctgagcaaacacatgttc
tgttattctgttcactcctgcatctgctgtacctggtaatcacaggcactcagtaactatttcttgaaaga
acaaacatcctgaatgtcaaactggcatttcctcccatccaaccccttctatctcaaacttgacagaaagg
agtgcaagtttctggcctgttgatgtctccttctgttttccacagacttgcttcgatttatttctgtaggc
atttaaaaagattttggggtgagtgtgaagctaaaagtgatggcgctaaccctgtgccttagagaagcca
tttggaaactgcagcctagacacattagattatcattttgtaaactataccatacatttcagtatgcata
cttattttaggggtgttttgggggtgatgtgtgtgtctgtgggtctatgtgagcgcggcgccttcctggga
ctgaggcaccggacctgcgcacacgcgctgcgtaccctacgcccccgccgcgtgggcgccgtcgtccccg
cagccccccacgcaggggcggggcgctgccacccattggctgaggccgataccacgcgccccgataccc
ggcgcaggagccacctccctgagccccgcggaccacgcctcagtccgcctgcgctcctcagcctgacggtc
cgcctttcggggctcctcagccttgtcacccgctcttggttttccttttctcttcatctttggctcctttg
accactcgaagccgcgcagcgggttccagcggacctcacagcagccccagaagtggtgcgccaagcacagc
ctctgctcctcctggagccggtcgggaactgctgcctgccgccatcatggtgagttgagggagaggcccga
ggggcaaggctggcggagtcttcactcgtgggtgaaactcggctctgcagggtcggaaagtgaagtcatcc
tcggctttcctgtttgggggcacctgcctggacgcaccacagccatggcccacagagttgggagaggcgg
cctggcagtggcgcggcctacacagagaggcaggcgcgggacgccaagggcggcaaggtgggggtgggccc
cgaggacgcgctgcctcgccggccacgtgcaagggccacggccttcttgaggcacccatttcccggttcgg
gttctctactggcagaggtttggggtgcgggttccccacccccgactgccccaccccgcccccactcccac
gcgctcccccactcacgcgcctccgcccctcccccactcccacgcgcctccgccccgccagcctcatttct
ggcgcgctctccccgtggctccgcaagatggcggcgggcctgctgagggacttgccttcttaaacttcgag
ggccggagcttccttaggaagtggttccctcctggtccctgtacttggtggggtgggcggggttcgtgggg
gcctcagcgtggggtgtggggcgtcctgaaggcatggccctgtagctgagggggctgggccggaagtgcct
ctggtcccctcgcgcctgtcagctgcgggggacttgctgatggcggctccctcgtgtggcggcgacagaaa
agctcgcctgacgccatctttgccgcacgaggttggtttcagcagttaccattaagaattttagaatttt
aaattctgggtacggtttttccctctcgataaaccaggatcctttgacaaaatgtccatttggtagatag
cgttgggtgtgcgtttttgtgagtttggtaaatgcctagcgttgcttcagcgccaccacaagatgcagctc
tgaacacgcagaccaaccactgagtcacgaacgcgcgtattacatgctgtgtgtgtgtctgcattgataat
tctcagtctcttcagggaggaaatttatacacagtttatattaccatgtataaacattaaaataaaaataa
acacctattatgtgtaattgtgcaacatatagttatccatgctaataaaaagtaatcaccttgagttgtaa
gtaattatttattataacacttttcatttttagagttaaatccgaattgtttcattggattatttttctga
ttataaaagtatcatatttactcatttgaaaacaatcgctgtaagaaaatagcaatccatcttaatcccaa
cacttataaccgttaaatttggtacatgtatgaccttctagattatttttatgcatgtattgcatataga
tgggccaacttttacaagagttacaccgtgctgcataatactattttacaacttgcatctttcaattgatg
tgggcattgttctgaagataattttaattttcagtatatgttaaaaattataccattattttaccgaattc
gcttatgatgaatactttttcaatattaaaacaatactatcatggctgggcgcggtggctcacgcctgta
atcaatctctttaggaggctgaggcgggtggatcacttgaggtcaggagttcgagaccagcctggctaata
```

cggcgaaaccccatcgctactaaaaacagaaaaatttgccaggcgtggtggtgcgcacttgtaatcccagc
tactcgggaggctgaggcaggagaatcacttgaacctgtgaggcaggttgcagtgagctgagatcatgcca
ctgcactccagcctaggcaacaaagaggggaggctcagtctcaaaaaaaaaaaaatgctatgaatatcctt
gtgtaaatatcttggcatgcttgtattattaattccttgagatgaattgctgtaagtggaattgctgggtg
aaagcgctttaccccttggtaacaattttaagttacacccctcattagtatagtgatgagtatcctcgcc
tatcaaattatgatacatcttttccaagctactcttggtgtaaagaatgtatcagtattttcccactgatg
gtgtgtagagtgatctcttttccacaccctttccaatgaaaggtattaggatttttgaattgtcaaccta
atagctggaaaatatttactatgttgatgtgcacttttaaaaatcatcaatgaagttgatcgatcattag
ctttttaaacttccttttattttatttgtattgattcataaagactgtattaatgttcacctttgtcat
acatagtggaaaaaaatacatagtagtaataaggaggtgcccttaccttattcctaatttcaatggagatg
actctagtatttcactttaagcttaatgatagggtagaattagattgatagaaaataagatgtctctattt
ctgttttacttggatatttagataatttaaaaatgaatgctgtattttgtcaaatgctttttcagtggtga
gttaaagtaacaatatcctgtgaagtattcttgcatttccaatataaactatattctttatatactgctgg
aattttgcatttgtgttgagggcagggatggctcatcttgatgaggctttgttacttatatgctgttagct
ttgtaaaatgcgttgagaaatactttttcctctttgaaaattgaagaaaagtgtttgaatttttaatct
cctcttctgtgttttctgtctttatttctggtgtctgatgtgtatgtgcacctctctctgacacaaacata
cacccaagttatctatttctttacaagttgaaaaaactcccacggaaactgagccaggtatcattaggtgt
tttttctcaagattgttttgtaaaaatttaaatttctctacaagtttgtcatgtcgttgatttccaaattt
atttctatagagtttatgcatagcattgttttataatctgtggtcataacccatttctcaggtctcatata
tccttttgctttcttgcctttcttgcttagttttgccaggattgtttattgcttgaatttcatcctcccc
acccccaactccccaagaatctactcttggatttacttaaagttttgattctgattttcatgtcatatttt
cttaattttagcattctctttattcgttccttcctattggtgtgaacgtaactttgataactacaaattat
agccttttcctggtcagtttatataaatctcagtttgctttcatgttcacatagtaattgaattagataa
tactgtgaaagaagctgacatcccaggagttacataataatttttttccagcttttagctgtctttgtcact
tggttgtccaaatataacaaggaatatgtttgctataaataacaatgagagttgtgtatatgatggtcttg
aagctcacccgtttactattccaacatttctcatttcaaaattcaccaactcacttgaaggtatgattaag
aaaattacttctttaatgcagtttgttttcaggatgctattgacagttacgatggtactttctttaaaatt
aagcttttgcttatcctaagcattcatttatatttttggccacggtttcaccaaagtttgtctattgattg
tggattttatatatacattataattatgtataaatacagtaggaatttaacaacccttgctaatcaagcc
gtggtaaaaatgaagtacattgcttactttgttagctttgagcgccacctgctggctgagaactaggacat
gatggtgcctccttaatttaagatgttattacgctattaaagtgatcctatgaattttatatttacgattc
cagttaagcaggcagccagtttgtaagatccaagccaaagtcagacttcctaataatatcaaatgcctctt
ttcctctttaccaaactactgttgtcgccaagtatccgttttactcttgtatagggtatgagaattagaag
agacaccagagatcttttccaaccattcattttcaagcaaggaaactgaggtcctgaaaagacttacttta
ctgaacaatggaaatcaggttcttcccaaacaataattttcctcctcaattcttgtggcactgtcacaagg
tttcttaaccacaccctattcaaatttggaccaaatagttccttgttgtaggaagctttctttgcattgt
aggatgtgcagtagtattcctggccttcagcccatcagatgccaggaatacacctcagttgtgacaagcaa
tctctaaacgttgttaaatgcccccagggctcaaaaatcttccccagttgagaatcaatgcattatcatga
cctcttataatctgacacagttttctgccagatgttacacttctaaattaaaatacatccaaacactgtga
tgcagagacaatactgtatatatagcttgtttatcaaagaaaaacatatccaaagttcatccacaaatcaa
gggggaattgaaaaggagaaaaaactacaagaatttcagaatttattttgttttgttttattctgtttggg
atttctgaatgaactgagaagggattggtaaattgtctgtagggggggaaatctaatcaggttcctaatagt
tcaagatacgcagcttaatgtcatttagaaaaacaacgtaagacatttttagaaaggaaaaactgtagaaa
tagtttttaaatttttcttccacatcattgtcatgaaataccactattaaaagaaagaattaggaattatt
tacaaacccaaaatagcaagtataagcacaaacgaatagagaaaagcaataaactcagctttagttgattt
gacaatttagatttcatctctagtaaagctgtgtagaatcagtaacttctagcatctaattaacaaagcag
gaactctactttcttattaattcctataaaattactgtgcttttcaaaaagataacagttgctttcttttg
tcttgttttctttcagacaggtatatgtgtaatagctttctctttgcagcttttgtagttacctttgt
agtcaacaggcacttcaccaatgatttttaaattaatgatgtttcatagtttcctgatgtttcaattcgtt
cttaattaaaagtatcttcaagataagtcagctaaaatagtaaagccaaaaaagaaacttgaaaaaaaatc
tgaaaataaattcaaataacaggacactaacttatattgatataccactttatactttacaaaatgaattc
acagatgttatctcatttccttctcaaagcagtttgcccagagatacacggctggcaagtaagtggcttgt
aaccacttacctgtaagggacataaacctagatctcctgactaaattaatttttgtttcactctgctatat
tgcttctccaacatatatattacttacaaattctacttcagaaaatttacattagccttagagtctactg
cggaaagaataaagacagttctgatcaaaatccagaaaaaatgagttatatgtgtgtttctttatctgaaa
ataaagatggtataatttatatctgcctactgtgataattagaattatcagtgggactgatttggtgttag
gaagaaaaatgttatttcaaaaaatgatttgaacaccctagcattagggaagaggattataatcctttttt

tagaaggagctgtgacatgtgagcaaataatgtttccatgatccatttccaaagtaaagcaggcatactaa
cacagcgttaaaattgtaaaaattcatgttaattttggtgaaaaaggcgaagcaggacaaattggtttaaa
actttgaggcttatttctttttgatgtggcttttggagtttttaatatttgtcatccttttattagt
gtactcagtaaaatgaaaagttgtttatgaatatagaacaggatgattttaattttgatttttacaaatct
gaactttccaattttgggtttgttttttttttaagtaattcacatagtattttcacaaatactaagtattc
ttccacagcattgtgcaagactcggtaatgagcaatactgggtttaagaaagtactgtatcctggctgcgc
gtggtagctcacgcctgtaatcccagcactttgggaggccaaggcgggtggatcacctgaggtcgggagtt
cgagaccagcctggccaacatggagaaaccccatctctactaaaaatacaaaattagccgggcgtggtgtc
gggcgcatgtaatcccagctactcaggaggctgaggcaggagaatcgcttgaacctgggagacagaggttg
cagtgagccgagatcacgccattgcactccagcctgggtgacaagggcgagactctctcaaaaaaaaaaaa
aaagtactatatcctttcattaaaatttcatgctgatctactctgataagtttaaatcctgagcctgaact
aacttagaatgaaagtaatttatataaaatatttatacttaaaatacaaaaattatatttcacattatgt
ataggttttcctgtgtatctaattatgatgtctactttccaaattacagtctactgcaaatcctgaaact
ccaaactcaaccatctccagagaggccagcacccagtcctcatcagctgcaaccagccaaggctatatttt
accagaaggcaaaatcatgccaaacactgttttttgttggaggaattgatgttagggtattgtattcatacc
tcatttttaccttgaaatgcattatgaataatgggatttgggccctgttacaaacttaaggtttttttgta
cttcatggaggtttagaattggttttacatttgacccataggtactaaaaatatctttgacaaagagctac
tggtcatttggggataaatgggggagaaattgtcacatcatcgaaccttcttttagtaaaattaaaatttt
tgaatgctgaattttttactcttgaagttcaattcttttccatagatggatgaaactgagattagaagcttc
tttgctagatatggttcagtgaaagaagtgaagataatcactgatcgaactggtgtgtccaaagggtgagt
aattttatcaaaaatatctgaactctagccacgtatagagtatcagagaagacttcaaaattggtattctg
acacttaacataaatttgttgctgtgttaatttctttcatgtagaggataaaagtttattgccagctcttt
taatcattttttctaatgtttgtttttgaatatcttcaatctttatttcatacaaatgaattaatcagtttt
ctcaaagaactgttttcttcatacattgcacagtatcttaaattttaaccaccttgtcttagatagtaagt
taaattcaggttcatggatatgaattctttgttaatcaatgaagttttcacactaccctaatcttagcaca
ttttgattgacatagttcagatgaagaaaaagcagtatttgtagaggatctatcatgtacatcttaaatat
ttagcatagtatattattcatattttagtcatgatcacttccgtatatagtagaagttctgaaccatgtac
tgtatgatggtgattttatacttcatttatctgcctttatagctatggatttgtttcatttttttaatgacg
tggatgtgcagaagatagtagaagtaagtaatctaatagaaaaatctcatttgtttaattgatacaatgtt
tagtgtcaagtgatatactcagtcttgtgtaaaatttggaagacgatacactttctggtcaaaaaaatcca
aacttagaggaatcttacataacttgttagaacctgtttattttgactgggcacctaggttcatgaacta
cagacaggaagggttggagacagggcagggtaatgaaaagttttgatcaactttcacttgatgcctcttg
acactgattagagaaataagggtaaggtagcttcatgatgacaaattttaatttggtgtgtagttatcact
gatcttctatgataataggaatttagaagactttaggtgttcatccaagtcttggaagtaaagacttgaa
aattgattctagttttgttactgttttattttcagtcacagataaatttccatggtaaaaagctgaagctg
ggccctgcaatcaggaaacaaaatttatgtgagtacaaaaagaaattgttctttttaaacataaaagaaat
tgtgtagcttttcaaagaactaaaaataggccttttctttttgcttttcaaaaaggtgcttatcatgtgca
gccacgtcctttggttttaatcatcctcctccaccacagtttcagaatgtctggactaatccaaacactg
aaacttatatgcagcccacaaccacgatgaatcctataactcagtatgttcaggtaagaattgcttaacgt
tcctattctcttgtttattgtagtcatcgtaccttctgtggaattacatccaatttctgtagtaagtggta
atggcatccctgtttgaatactttgagtgatgggaaatttattacttttgttagaaatttcttattcttag
tgttgttctttatatgagctaaagatcttctgtatactttgtctttatcttagaaatgacctctgtagaca
catgaagaaatctcttcctttttccccccatataactagttcctaaagcatttgaaagcagctatcgttat
gtgtctgtctagtatattcttctctaaggttagcaaatcatctagctattcttatttgcaatgatttcca
gatgcctcctcatataaattgctgacttctggatatattctggttctggaatgggtagatttctgatgtgt
tttactatgtaaatcctgtgagtttctggcatataatttctctgatcttggttactttgatatttaaagta
ggatttgacatactatcacttactggtggtaaataacatttcttcttagttcattttatttaacgattta
gtttaaaagacattgtctttgctggaaaataaagtagcaaaacaggagtgaaatagttcttcagtgtcttt
cattcattgacattttccatgtacttgaaatgtgtagggtgtacctcctcttctttctccttctctgaaca
atggctagaaaaaaaaaagccctacttgtttctaacatttactgtgagtcattactgcatctgggtgtatt
cgtgtatgctgccacctatatgttttcaatcagtagctatttattgaaaaatataagacattatactgttt
cttttccagttttggattatatacagcccttagttcttcgaaatgaagtacagaaaaagccatagcatctg
taggaggactacatattaccctataatattgtcaaacacaaaactgtctagaagtatttgacaaagaat
agcaaatgtattaatttaacttacattgaagtctgtctgaatagagccttatcaccagtataaaaaataac
ttctgggtgggaataagtacacagtataaatatgataaactttgcctgttgaaatagtcacttcttttgtc
atttgtctgttccccctccccacccaaagggtaacacttgacagagaatatttctttcttcataaagtcag
tcatgcatttagaattctgcattgttgtatatagaaaaatattttaaaagttttcatattttgttatatt

gggaataatatttctaattttaaaaaatgttttatattattcattctttctgtaaacttgattttcaggca
tatcctacttacccaaattcaccagttcaggtcatcactggatatcagttgcctgtatataattatcaggt
aatttaaaagggagtaaaatgatttactttcagattttattgaggcctttaacttgtttatacaaattgtc
tgaatggtttgtcattttaaactagtgaaatgtacctaaaatttaagaaaaaaattagtctagaattaaga
cctctttattatttagaagtaatggaataatatttgacagggatatacttagcaataactttctctaca
acagttttatgagattcggtgtccccttctgtatttcagcatgtatttttcatctttgctgtcaaatagc
cgaaacagccagacggactttcatcaatttttagggagatagagtgaaataaaattatccaattcttaga
gcacagaattcaaattgtattttattttagctgactgcttcatgatagtagttctttgagactctttaca
tagatatgactgtatctgtgacccataatcatatctatggtaataaattcaaggaactaatatctctgaga
tttccacaatgccaactccagaaaattgggaaaaaggtgaggttttatatatggaagtaacaagaacatca
gggattagaaacataaagtactttttttttcattctgtttcttttattataacaacaaaggagccagcatga
taagtacttcaatattgtgtatctcgtgtgtttttgaaaatttgtaggaatatttgaataattttggtttc
cttttttttttttttaagatgccaccacagtggcctgttggggagcaaaggagctatgttgtacctccggt
aaagtgaattagccaaacatataattcctgttattttaaagtatttttttttgtttaactatatttcttga
ttgttttccatgtcatatatgcctatattttaaatagtttttgtattaatgtgttacattttgttactt
tctttttaacccagttagaaactcccatgggagcaacagtgccttcttctctcaggtttttgtgtgcttaa
gcaatggctggtccacataatgataagtgttcagttacttgttgatagactatataattcaggaaaggtag
tattaatgtggctttagaattagcatgtatctgcctagactctgcctctggctttaccagccataaaaaa
actgttgaagaggaacagaaatgttttgctgttaattactgttaaataaaaataggaataaaacaagagta
ttacgtctaaaacacccaagctgctgtcttttatcaggattggaaatttgaaggatataataagtgttaaa
attctcaaacactctcttactacattagtttcttagagttcttctacttctgattatgttgatgccttaaa
gacattctaaaacagaaggctgccttactgtatttcaactgttcatttaaaacaaagtttttgaaataata
taattgaatttcaacacaacctacattgaaacttttgataccagcttagctttttgaagaataagtggctt
ttaaatatatctgtatatctgtttaattacactttcattattttaaatataggcttattcagctgttaact
accactgtaatgaagttgatccaggagctgaagttgtgccaaatgaatgttcagttcatgaagctactcca
ccctctggaaatggcccacaaaaggcaaacatctaattttgtttatatatatttcatatttatttttcta
gttttaataatggtttttgagaagcaatcatccttcagcagaaatttgtaattgaaaaggttaactagag
aagagttagttgactggcttgaccaaatagtaaaagaaaattttagatacagaaagcagatcttggctggg
tgcagtggctcacgcctgtaatcccagcacttttggggctgagatgggtggattgcttgagctcaggagt
tcgagaccaccctgggcaagataggagccataaattatagatcttaaataattgactaatattcagcagtt
aatgtaaaggttggtgaaatttcagatagcccaaattttcaatgtatacataaagtttctgattcagcagt
cctttcctatattccagttccactaatttttaaaaaccatacttcaataaatactatattaataggatttg
gcagaatgttatgggaaggtttcctcaagaattctactcttcaaaagaggaattggtacgaattacatgta
ccttttcttttataattttgatatttacttaaacagtgaaacacattacttacggcattctctctgtaaca
ttatatatggcagtagttcccaacagtggtgaagtcagtaataattattcaaatattgaattaggcagggt
tctttcttcttttatcattagagcaaatttccataataatactatcactcttgaatcacatatgttctct
taaatgagcgggagtgtagaagtagttgatgtgttggtaatatgtataacactgaagtcccattggagtgt
aaattccttgatttgatactcgattttaaaatgcgaataaatattaaaatagcttacagtgtaattttcag
gttttgtcctgaatatttttttcttgaaacattggagttcacttagggatttaacaaattcagcttttaaa
ccagtattctatcactaaggttctaaaatctttgtcagaaaacttgcatattgagtgatacgtggtaagaa
ttatgaatcacattttatgaatttctttttttttttttttccccaagacagagtctcgctgtgttgccca
ggctggagtgcagtggtgtgatcttggctcactgcaatttccacctcctggattcaagcaattctccctgc
ctcagcctcctgagtagctgagattacaggcacccgccactatgcccagctgatatttgtattttttagt
agagatggggtttcaccatattggccaggctggtcttgaattcctgacctcaggaggaggcaggaatcctc
ctgccttggcctcctaaagtgctgggattacaggcgtgagccaccactcccagtctcttttagcattttt
gcatttctttagaaataaagtaatatattcattaaactatcaaaaaaaaacccttactaagagtgagtgaa
agagtcgtctttacattactgaaaacttctgtgtttcagaaatctgtggaccgaagcatacaaacggtggt
atcttgtctgtttaatccagagaacagactgagaaactctgttgttactcaagatgactacttcaaggtat
gaatataacagttacgcacaattaaaaagcacacttgttagctttaaagtatctgttttccttgtggtata
tgtattttgagatttcttagagcatttaattaactgttggcatttgactataacacagtaaaccagagtgg
gttttacctggcagtatattttctgctgctgaaccttgacataatgtagttatctttagggaagaatcctt
cagcagaaatttgtaattgaaagggttaactagagaagagttagttgactggcttgaccaaatagtaaaag
aaaattttagatacagaaagcagatcttggctgggtgcagtggctcacgcctgtaatcccagcacttttgg
gggctgagatgggtggattgcttgagctcaggagttcgagaccaccctgggcaacatggcaaaccccatc
tctacaaaaaatacaaaacttagccagttgtgatggtgcacgcctgtagtcccagctacttcagaggagg
ctgaggtaggaggattgcttgagcctgggaagttgagtctgcattgagccatgattgtgccactgtactcc
agcctgggcaacagagtgagaccttgtctcaaaaaaaaaaaaaaatacatgcatattggactatagagaaga

aaagaaatggtttactcagaagatatacctgaacagtgtgaagggagaaaaggggtaaagtgaagcagtaa
aatgttgagtagaaagaattggaggttgattcagacagagttggtaaagtggaagagaatgtgggtagttg
aattccagaaaaatctgatttctgatcccgccgtctatccatgttggatagataaatcttattaagactcc
agtttttacaagtctaaaatgagaaggtacaggactaaaggtttctgggtccctgtagttctaagtctatg
aataggaaaaagaactaacttggtcagtccaatgggagagaaatattacggttagtaaagggaaggtgttt
tttaaataataggtttattgaaatatataattgagataccagtaatacaatttacctatttaaagtttgca
tttcactggtttttcatatattcaaagttgtgtaaccatgaccacaatcaattttagaatacttaaatcac
cccaaaaatcaccccccctaccttagcagtcacctgctattttccccatcctgtgcagccctaggcaacca
ctaatttacttactttctctaaggattttcctgtcctggacatttcatgtatatggaatcatacataatgt
ggccttttgtgaccggtttttcagtcagcataatattttcaaggttcatcaatattctagcacgtatcag
aacttcatttctttttatttgtggctattattcattctgtttatccattcatctgttaaagacattgggat
tatttcctcttttaactgttagaaataatgctgtgaacattcatgtacaagttattacgtggacatatgt
tcttatttctcttgcgtatatacttgggaatggaattgctaagtcatatttaaccttcagtggaactgcca
gaatttgtcaaaactggctacacactttacattcaaaaggaaatgtttaaccatcactttgtgtcttaaaa
cagaaagactagctttttttcatctgtgaatggatataggatgaagcttaagccttttaaggggttatt
attatggatctcctgtataatgtagaagagtagagccgtatagcagaattaagttcttaacatcttcgcaa
cagggagtaaacatacttaaagttgacatttgttctcttgttgcttcattctacatagatagtataattta
gaaaagaaggaactgaaattgtatatcagcttactttgcccaaagttctgatgattacataggtgtctaca
gtagtacttaaatgatttcaaagcagaagatagtcttctgtggtttttgtttttttgaggtgacctcat
ttagtcacctaggctggattgcagtgtaacaatcacagcttactacaacctcaaactcctgacgcaaggga
tccttctgcctcagcctcccaaatagttaggactacagacatgcaccgctacacttggctagttaaaaaga
tttttttttttttttaagagacagaatcttactatattgcccctggctggtcttgaactcctgggctcaa
gtgatcctcctacctcagcctcccaaagtgctgggattacaggcgtgagccaccatgcccagcccagaagg
agttttgaaacttgagttctcacctggtattagacaaaagactccctttaaactcccagagttttctgctt
gtttgggagacgaatcagaagttggcatcctgtggttgtctgacatctagacctattttgattgactgcat
gttgttgtattgaatttgaatgcatgaggtatttttgaatgtatttactatttccatagctaattccactc
ttcactgtcttgattctgtagtcccttcataatcttgtttcctgcgtgacctgtgaagatacatgttggtg
accagttttctagattttaacttaaataggatatcactcttttgacagattaggtaacttccagaagccac
ttttatttatatgacagtgaaactgaggtcctaggaaaagggcacagttaactttgtagaaaataaaatgc
tattatgttattactcatttgcttagtcccatttctccttttaaggcctctcactttttctctctgcaca
tctgtaggcaacatagagtgaaaagaaagtttgcatgtatttaaatttttctctttctttctaaagaat
aatacatcttcacaggttaatgatatgtctttaaatgccaaaacttacatcttttaacctaaaaacatgaa
atttcagattggagagctgttcacaagccgtggttcctattagatgcagttcagtcagtgaaaacagtatt
ttttagaattacatttctaccagctgtctttgggacattactgcaaaatcattaattaagaagtacataa
aatgatattgagtctaagtcctgttatttctgagtttaatgggatttcttttttttaattgatttctttt
ttaactaaactgtttaataaaactagccatcctggtatatatgttatcccagtgttcaagaatgcttctca
aaaagaataatctttttctcattatttataatgtttaaacccaaaacaaatggtttaagtttttgacaac
tttcagatctatagtagtaatcagaaattttcagtaaagtaaaaggactctttctgtcttttccaggataa
aagagtgcatcactttagaagaagtcgggcaatgcttaaatctgtttgatcctcctggcttatctagttac
atgggaagttgctggttttgaatattaagctaaaaggtttccactattatagaaattctgaattttggtaa
atcacactcaaactttgtgtataagttgtattattagactctctagttttatcttaaactgttcttcatta
gatgtttatttagaaactggttctgtgttgaaatatagttgaaagtaaaaaaataattgagactgaaagaa
actaagatttatctgcaaggatttttaaaaattggcattttaagtgtttaaaagcaaatactgatttca
aaaaaatgtttttaaaaacctattttgaaaggtcagaatttgttggtctgaatacaaacatttcacttct
ccaacaagtacctgtgaacagtacagtatttacagtattgagctttgcatttatgatttctccagaaattt
accacaaaagcaaaatttttaaaactgcatttttaatcagtggaactcaatatatagttagctttattgaa
gtcttcttatctaaacccagcaaaacagattcaaagcgaacagtccaatcagtgggtcatatgtttattca
aaatattttatcttttagctagaatccacacatatatatcctatttgattagggtagtaattaggataact
aaaattctgggcctaatttttaaagaatccaagacaaactaaactttactaggtacataagcttctcaat
gagtcaccattcttcttttttgtaaaaactttttctttgaaatgctaaacttggctgtatgtcaaattgt
gcaaaatattggtattaaagaatgctgcaactttttatgtctcttagaggttaatcagagtatctgaagg
gaattgttttataaaaacattgaaatattagttacttgctataaatagatttagtctgttatatttcctt
ttgtaaagtaaaatatgtccagaagagtcaaagtagttagttttggttatttctaaaccacaaaagttgtt
taataagtatatcttaagaatgtgctagagttaaaagttagcattgtttctagattagctggtgtcttcat
tttacattgtgacaaacagctagagcatcagagccctttgctataccacagtctttcgtttccagccttt
gtcactagtctttgaggaggtttgctcctagaactggtgatataaagaatgaaagtagctgtatgagcaga
gcagttcaagggccaaaccctggaacggtagcaatgggatataatacctttctaagggaaaagttgtatc

agtaccatttgatctgccatggacatgagtttaaagcggctttctggcccttctttcagtgacttcttccc
taaaatgtagaaattctaacttaatgtagttactgtgagccatattactagtgcccccctagggtctataat
tccttaaaattttcattctgaatctgaaggagagagtcttttaactttagaattcccaagagggctttatt
acacctcagaaattgaaagcactatgaatttgtccatttaaaaatgatctgtagtttttttggtgctataa
cattctgacacatatcattctgtgattaaatctccagcttactataaatgatatctatattctaaagagct
acttctaattattccaatatgaccttaaggaaaagtaagggaataaattttttgtctttgttggagtgaagc
atttaaaagagtaagggtaaaaaagataaagtcctgaacctttcaaaatggaaaattaattctaaacttag
aaatatgcttctgcctattgctgatactgtctttgcatacatgaataaaaataaagtttttttcttcaaaa
gtgtttttagcattctgatgtaataatctaaaatggtggggagttgggtgggaactgtgtaacaagattta
gattcttaaaatgcaaaagtataaagttcaatttactgatctggcaaagttaggtcataaaagcaaattca
ggtacaaagagaaatttaagagatgcatgaacagcagtgccgagtaggaatgctgatgaacacatctgact
ctgctatctcatggctaaggtccctctcaatttggaccctatgaaacattttgtctactgtacgctttggg
cctaatctctagacctttttatttcggggtattgcggttgcctaatagagcttaaatttttttgtattgcatg
tacttcttatggatgcaatcaaaaaaaaatcagtacttactacctcttccagctgtagttctagtactccc
aattctataagctgagcccagtggagaggaaattctcccctcagacactgcttctatatttcatcatttag
taagcttactgattcataaccagaaagttgactccaaggttctgcaggataccaaacagtgctttctgcat
caccaaagattaaattgtgatgtttaatgtctaataataggctaaaaattagtaattcttaaatgtgtata
tgtgtatatatgtatacatatgtgtgtctgtatatgcgtatatctatggttacgtatataccaaccattac
ccagaatatttggttatagcacaagctcagtgagtttatggtattttcagaccccaaaactagacttacat
atgccttaagatagttgctttacaccagcttgttatctcaagcatgtttgtggtttagtattaaattctta
acacaaaatcaaatctttggaattgaaaaatactgttcagcactttttttgtaaaaagttcaagttatggca
aaatcaagcaactctaaaaaggttggtcacctccataagtgtattctgcatgtttgtttttttctttttcta
aa

Chapter 3

Additional sequence variants found in the screen of *Deleted in AZoospermia-Like* (*DAZL*)

Introduction

In our screen for sequence variants in the *DAZL* gene as described in Chapter 2, we found a total of 99 sequence variants, of which twelve common single nucleotide polymorphisms (SNPs) were described in detail (Fig. 1). The requirements for a sequence variant to be included in our previous analysis were: 1) that the variant be a polymorphism, i.e. the major allele did not occur in more than 99% of the population, 2) that the variant be found in all three patient populations, and 3) that the variant be in Hardy-Weinberg equilibrium. There were, however, a number of sequence variants that did not meet the requirements for our previous analysis, due to a low prevalence in the population, that were still interesting because they were missense mutations. In total, five missense mutations were identified, one of which, SNP 8175, was common enough to be included in the analysis described in Chapter 2. The other four, listed in Table 1, were found only one or two individuals in all three populations. Three were located in exon 2 and one in exon 5. These four missense mutations are described in more detail.

Results

Pro6 → His6

This sequence variation changes a proline to a histidine in the N-terminus of the protein, before the beginning of the highly-conserved RNA Recognition Motif (RRM). Although proline is usually considered a hydrophobic amino acid whose unique cyclic structure causes it to influence protein architecture, its secondary amino group means that it is not averse to being on the surface of the protein exposed to water. Histidine is also a bulky

amino acid that can either be neutral or positively charged, depending on its environment. Overall, since both amino acids have the ability to sit on the surface of the protein and are large, this amino acid change is most likely to affect the structure of the protein, as proline is very rigid and histidine has an extra carbon that allows its imidazole ring to rotate. We identified one woman who was heterozygous for this mutation, and reached an early menopause at age 45, but had no family history of the condition. She was able to produce three children before experiencing menopause, so this mutation is compatible with producing offspring.

Asn10 → Cys10

This mutation was identified in two patients: a male homozygote with azoospermia and a female heterozygote with early menopause at the age of 44. Asparagine is a negatively charged hydrophilic residue, while cysteine contains the hydrophobic and highly reactive sulfhydryl group. Although this particular change lies in exon 2 outside of the RRM, switching a hydrophilic for a hydrophobic residue can have serious consequences for protein folding. In this case, it appears that this mutation may act in a dose-dependent manner: the female heterozygote has a less severe phenotype than the male homozygote.

Ile37 → Ala37

This isoleucine is part of the RRM, five amino acids upstream of the RNP2 sequence which is one of the most highly conserved parts of this motif. Isoleucine and alanine are both classified as hydrophobic amino acids although alanine, with its shorter side chain, is not as hydrophobic as isoleucine. The structure of the RRM in another protein has

UCSF LIBRARY

been deduced, and in that structure, the amino acid corresponding to the isoleucine forms a hydrogen bond with another amino acid in the structure [119]. The shorter side chain of alanine would most likely disrupt this hydrogen bond. Hydrogen bonds can vary in strength from 2 kcal/mol to 7 kcal/mol and are weaker over larger distances, so to determine the exact effect of this amino acid change would require the structure of the RRM of the DAZL protein. We identified this mutation in a woman who experienced a spontaneous early menopause at age 43. She was heterozygous for the mutation and produced one child.

Arg115 → Gly115

Arg115 is the near the end of the RRM. In this mutation, a basic, hydrophilic amino acid is swapped for the simplest amino acid, glycine. By sequence comparison to the RRM of the HuD protein, which has been crystallized bound to RNA, it appears that this amino acid may be key to making side-chain and main-chain connections to the RNA and the other RRM in this protein [120]. As DAZL is also known to homodimerize with itself, this amino acid may be key for both the RNA binding and protein binding functions of DAZL. Mutating the basic side chain of arginine to the simple hydrogen atom of glycine should severely disrupt the interactions this amino acid normally makes. Accordingly, the one patient we identified with this mutation was homozygous for this change, and experienced premature ovarian failure at age 34, having borne no children.

Discussion

Our screen of the *DAZL* gene for sequence variations that are associated with male and female fertility produced not only a number of interesting common SNPs, but also intriguing rare missense mutations. None of these mutations were found in women that had a normal menopause or were still menstruating at age 46, so they may be associated with impaired germ cell development and fertility. It would be very useful to screen a group of male and female control chromosomes to determine whether these mutations can be found in individuals with "normal" fertility. These mutations were also clustered in exons that form part of the RRM, though only one/two of the variants were actually part of the conserved region. It is possible that this region of the protein is simply highly variable, but it is also possible that our population is enriched for mutations in a functionally important part of the protein. In other RRM-containing proteins, many of the residues surrounding the motif are involved in RNA binding specificity, so even those mutations that are not part of the RRM may be required for proper protein function [121]. Recently, a number of potential RNA targets for DAZL have been found, allowing us to test whether these mutations affect the ability of the protein to bind its targets [59, 61, 62, 65]. In addition, as this region is known to be involved in DAZL homodimerization (see Chapter 5) and binding to other proteins, it would also be interesting to determine whether these mutations, like Arg115Gly for example, affect the ability of DAZL to interact with its protein partners.

Interestingly, we saw that possessing a missense mutation, especially a heterozygous missense mutation, was compatible with producing offspring. Each patient that was heterozygous for a missense mutation was able to produce at least one offspring. In contrast, the patient that was homozygous for the Arg115Gly mutation experienced

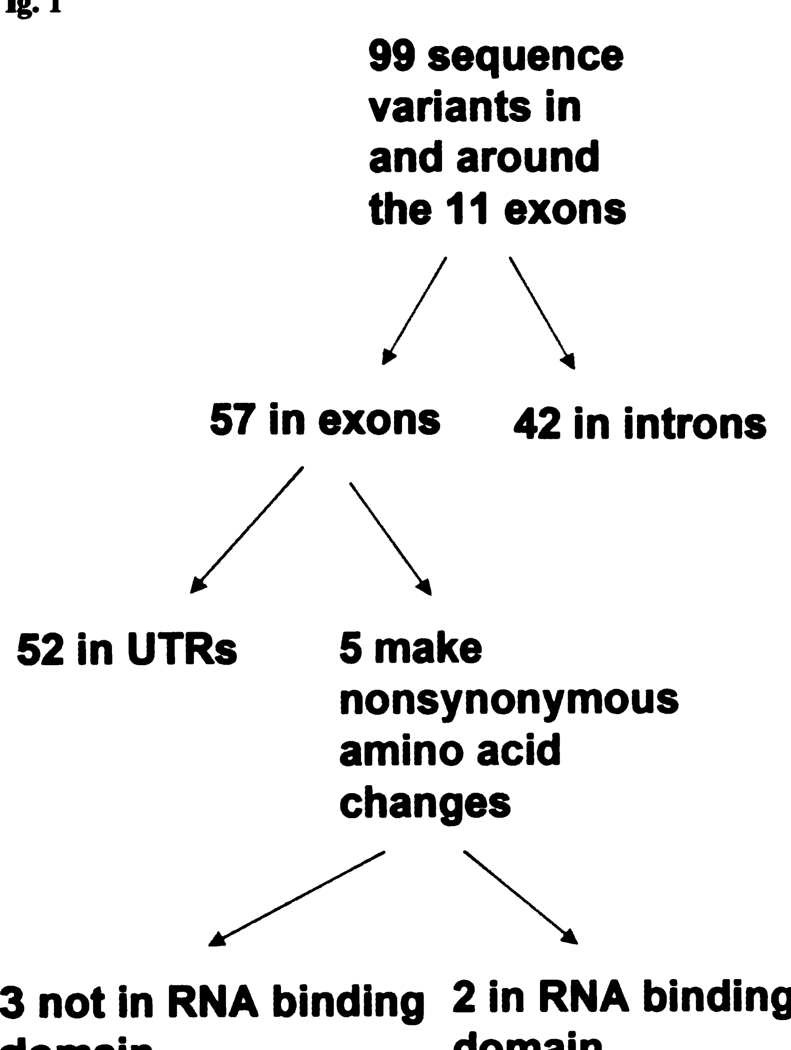
ovarian failure at age 34 and did not bear any children, while the male patient that was homozygous for the Asn10Cys mutation did not produce any sperm at all. This is consistent with the data from mouse, where mice heterozygous for a *Dazl* knockout allele are fertile but homozygous mutant knockouts are infertile. We might expect that individuals heterozygous for a hypomorphic allele would experience slightly impaired fertility, in the form of oligozoospermia or early menopause, whereas individuals homozygous for the hypomorphic allele would experience a more severe phenotype, such as azoospermia or premature ovarian failure. Future work on how these mutations affect RNA and protein partner binding could provide key insights into which residues are key for the function of DAZL.

Table 1 Missense mutations identified in the *DAZL* gene

Position	Exon	Change	Occurrence		
			Sex (Population)	Genotype	Phenotype
8169	2	Pro6 → His6	Female (early menopause/ovarian failure group)	Heterozygous	Spontaneous early menopause at age 45. 3 children previously.
8182	2	Asn10 → Cys10	Male (infertile male group)	Homozygous	Azoospermia
			Female (early menopause/ovarian failure group)	Heterozygous	Familial early menopause at age 44. 4 children previously.
8262	2	Ile37 → Ala37	Female (early menopause/ovarian failure group)	Heterozygous	Spontaneous early menopause at age 43. 1 child previously.
9740	5	Arg115 → Gly115	Female (POF group)	Homozygous	Spontaneous premature ovarian failure at age 34. 0 children previously.

For each mutation, its position in the *DAZL* gene (as defined in Chapter 2), exon, amino acid change, and each occurrence of the mutation is described.

Fig. 1



Chapter 4

Dazl has a meiosis-independent function in germ cell development and results in germ cell loss by e14.5 in the mouse

Introduction

Germ cells are required for the propagation of all species. One gene family that is known to be important for germ cell development across diverse species is the *Deleted in AZoospermia (DAZ)* gene family, which encodes a group of RNA-binding proteins. The family contains three members: *Deleted in AZoospermia* (*DAZ*), *Boule*, and *DAZ-Like* (*DAZL*). All three genes encode RNA-binding proteins with a conserved RNA Recognition Motif and DAZ repeat. The founding member of this family is *DAZ*, a Y-chromosome gene found in Old World monkeys and hominids, whose deletion is associated with azoospermia and oligozoospermia [27, 122]. *DAZ* deletions occur in roughly 13% of azoospermic men and 6% of oligozoospermic men [123]. *Boule* is an autosomal gene that was first described as the *Drosophila* homolog of *DAZ*, and promotes the G2/M transition in meiosis by enhancing the translation of the cell-cycle regulator *twine* [102, 103]. In worms and flies, disruption of *boule* results in loss of oocytes and sperm, respectively [102, 104]. Though *Boule* has been strongly conserved from flies to humans, the lack of variation in the human *BOULE* gene has made it difficult to determine whether it is required for human germ cell development [58, 108, 124].

DAZL is also an autosomal gene that maps to chromosome 3 in humans and chromosome 17 in mice [55, 56]. Recently, DAZL has been shown to bind to a number of different mRNAs in both mice and humans, and to promote translation initiation of mRNAs in *Xenopus* oocytes, potentially through interaction with the poly(A)-binding protein [59, 61, 64]. In zebrafish, zDAZL protein was able to activate translation in a 3' untranslated region (3'UTR)-dependent manner [125]. Work using mouse Dazl protein in

Xenopus oocytes demonstrated that Dazl can promote translation of mouse *vasa* (*Mvh*) through its 3' untranslated region (3'UTR) [65]. Together, this data suggests that DAZL protein functions by regulating mRNA expression via the 3'UTR.

The germ plasm, an electron-dense region of cytoplasm containing RNAs and proteins required for germ cell specification, of many species contains *DAZL* mRNA and/or protein, suggesting that *DAZL* plays a role in germ cell development in those species [60, 100, 101]. Knock down of the *Xdazl* gene in frogs results in male and female infertility due to a primordial germ cell (PGC) migration defect which prevents the germ cells from reaching the gonads [99]. Disruption of the mouse *Dazl* gene also results in male and female sterility through loss of the germ cell populations, most of which are gone by birth [68]. Female germ cells in the mutant appear normal at e15.5, but many of the oocytes appear atretic by e17.5, and by day 4 after birth, no germ cells were detected [69]. In the male, *Dazl* mutant testes exhibit germ cell loss by birth, although few and variable numbers of spermatogonia and leptotene spermatocytes are observed in the tubules [69]. This data suggests that *Dazl* may play a role at meiosis.

However, as germ cells were reduced in number in the males even before birth, and meiosis in male mice does not begin until puberty, it is clear that *Dazl* must also have a role in germ cell development independent from meiosis. Additionally, while male *Dazl* heterozygotes are fertile, they exhibit reduced sperm counts and high numbers of abnormal sperm [68]. It is unknown whether female *Dazl* heterozygotes have a similar phenotype. In this study, by counting germ cells and analyzing gene expression in different *Dazl* genotypes at a spectrum of embryonic stages, we sought to understand the

meiosis-independent role of *Dazl* and to ascertain whether *Dazl* has a dose-dependent effect in females.

Materials and methods

Mice

The Oct4ΔPE-GFP transgene has been shown to be specifically expressed only in the germ cell lineage; thus, mice carrying the Oct4ΔPE-GFP transgene on a FVB background were mated to C57/BL6 mice that were heterozygous for the *Dazl* knockout allele to produce mice that were doubly heterozygous for the transgene and the knockout allele. Male and female double heterozygotes were mated to produce the litters to be dissected. The morning that the copulation plug was observed was considered 0.5 days post coitum (dpc). At different embryonic stages, the mothers were sacrificed and the embryos removed from the uterus and placed in phosphate-buffered saline (PBS) until dissection.

Embryonic gonad dissociation

Gonads were dissected out of the embryos and placed in DMEM/F-12 medium with 20% fetal bovine serum (DMEM/F-12 + FBS), then pelleted at 2000 rpm for 3 minutes and the media removed. The gonads were first enzymatically dissociated with 1 mg/mL of collagenase/dispase (Roche Molecular Biochemicals, Mannheim, Germany) in PBS for 30 minutes at 37°C, which was removed by aspiration after pelleting the gonads, then with 0.25% trypsin, .02% Versene (University of California San Francisco Cell Culture Facility, San Francisco, CA) for 20 minutes at 37°C, followed by DNase I (Roche Molecular Biochemicals) at a final concentration of 1mg/mL for 10 minutes at 37°C.

The cells were mechanically dissociated to a single-cell suspension by pipetting. DMEM/F-12 + FBS was added to inactivate the trypsin, then the cells were pelleted, and washed once with DMEM/F-12 + FBS.

Fluorescence-Activated Cell Sorting (FACS)

To count germ cells, dissociated germ cells were resuspended in PBS + 1% bovine serum albumin (BSA) with propidium iodide (Roche Molecular Biochemicals) at a final concentration of 10% to mark necrotic cells and analyzed on a CyAn ADP analyzer (Dako, Fort Collins, CO). To collect germ cells for later analysis, cells were prepared similarly as for counting, but collected on a MoFlo High-Performance Cell Sorter (Dako), then frozen immediately at -80°C for storage.

RNA Extraction, cDNA construction, and Real Time PCR

RNA was extracted from previously sorted germ cells using the Picopure RNA Extraction kit (Arcturus Bioscience, Mountain View, CA) according to the manufacturer's protocol. cDNA was then constructed using either the iScribe cDNA Synthesis Kit (Bio-Rad Laboratories, Hercules, CA) according to the manufacturer's protocol, or using SuperScript II (Invitrogen, Carlsbad, CA) according to the manufacturer's protocol.

For real time PCR, each reaction contained 2 uL cDNA, 1 uL of each primer, 10 uL of iQ SYBR Green Supermix (Bio-Rad), and 6 uL ddH2O. The reactions were run and analyzed on a MyIQ Single-Color Real-Time PCR Detection System (Bio-Rad) with the

following protocol: 95°C x 5 min, followed by 45 cycles of 95°C x 30 sec, 60°C x 1 min, 72°C x 1 min, followed by 72°C x 5 min. All reactions were also checked by gel electrophoresis for the presence of primer dimers. The expression of each gene was normalized to *GAPDH* expression using the following formula: $2^{-(\text{gene CT value} - \text{GAPDH CT value})}$. For primers, see Table 1.

Statistics

The data for germ cell counts was fitted to a negative binomial regression model using Stata v6.0 (Stata Corporation, College Station, TX). Error bars for gene expression were calculated in Excel using the standard deviation function.

Results

Dazl is required for germ cell maintenance prior to e14.5

To mark germ cells for counting by fluorescence-activated cell sorting, mice heterozygous for the *Dazl* knockout allele were mated to mice carrying an Oct4ΔPE-GFP transgene that is only expressed in the germline [126] (Fig. 1). Counts were taken at embryonic days 12.5, 14.5, and 16.5. There was variation in germ cell number between litters which may be due to differences in genetic background and the timing of germ cell proliferation and has been observed before [127-130]. To test the significance of differences in germ cell number amongst genotypes, we fitted the data to a negative binomial regression and stratified by litter. At e12.5, there were no significant differences in germ cell number amongst the different genotypes in either sex, though the *Dazl* mutant mice may have a slight reduction in germ cells (Fig. 2A, B). By e14.5, however, there significant differences in germ cell number were observed between the mutants and the heterozygotes, and between the mutants and the wild type embryos in both sexes (Fig. 2C, D). At e16.5 as well, we saw a significant deficiency in the number of germ cells in the mutant compared to the other two genotypes (Fig. 2E, F). The heterozygotes at this stage appear to have reduced germ cell counts relative to the wild type as well, though these differences were not statistically significant. This data demonstrates that there is a requirement for *Dazl* in both sexes for germ cell development prior to e14.5.

Dazl mutants have aberrant germ cell gene expression patterns

To help determine how loss of *Dazl* affects the developmental progression of germ cells, we analyzed the expression of a number of genes known to be involved in germ cell development in isolated germ cells of different *Dazl* genotypes. *Pumilio-2*, *Nanos2*, *Nanos3*, *Stella*, *Oct-4*, and *c-kit* are all expressed in primordial germ cells before they reach the presumptive gonad, while *Vasa*, and the two meiotic genes, *SYCP3* and *SYCP1*, are expressed in germ cells after they reach the gonad [131]. The results of this analysis are shown in Fig. 3. We observed a significant amount of variation in gene expression, similar to the variation in germ cell number. Most likely, it is also attributable to differences in genetic background and the exact timing and maturity of the germ cells in each litter. In male germ cells at e12.5, no large differences were observed across genotypes, which correlates with the statistically non-significant differences in the number of germ cells across genotypes observed at that time point. Beginning at e13.5 and extending through e15.5, however, the *Dazl* mutant male germ cells exhibit decreases in expression in early germ cell markers like *Pumilio-2*, *Stella*, and *Oct-4*, and even larger decreases in the meiotic markers *SYCP3* and *SYCP1*, relative to the wild type cells. There was also a slight increase in *c-kit* expression in the mutant cells relative to the wild type, although in general *c-kit* expression appeared to stay fairly level in all genotypes over the course of this experiment.

Female germ cells, similar to the male germ cells, showed no large differences in gene expression at e12.5, which also correlates with the absence of statistically significant differences in germ cell number at this time. Also similar to the male, beginning at e13.5, mutant germ cells show drastically decreased expression of *SYCP3* and *SYCP1* relative to the wild type, although the females generally expressed higher

levels of these genes, which is not surprising since female germ cells enter meiosis during this developmental window. In addition, the mutant female germ cells expressed higher levels of a few of the early markers, including *Pumilio-2*, *Stella, Oct-4*, and *c-kit* at e13.5 and e14.5, and to some extent at e12.5, with high *Stella* expression persisting until e15.5. Mutant *Pumilio-2* expression went down significantly at e15.5. *Vasa* expression in the mutant, however, decreased over time from e12.5, down to almost negligible levels at e15.5.

Discussion

Previous work suggested that the initial germ cell loss in the mouse *Dazl* mutant occurs somewhere between e15.5 and e17.5. Our work, however, shows that both male and female mutants exhibit significant germ cell loss by e14.5, and that there may already be some germ cell loss in the mutants by e12.5. At e16.5, the heterozygotes of both sexes appear to have fewer germ cells than the wild type embryos, though not statistically fewer. These results are paralleled by what we observed when analyzing expression of germ cell genes in sorted germ cells of different *Dazl* genotypes. At e12.5, we did not observe large differences in gene expression across genotypes in either male or female germ cells. Starting at e13.5, there were larger differences in the expression of certain genes between wild type and mutant germ cells. Mutant male germ cells showed relatively low expression of early germ cell markers such as *Pumilio-2*, *Stella* and *Oct-4*, while mutant female germ cells had relatively high expression of these same markers. Expression of markers that are turned on after the germ cells reach the gonad, such as

Vasa, *SYCP3*, and *SYCP1* were low in mutants of both sexes. The low expression of *Vasa* is consistent with work that showed that *Dazl* mutants produce reduced quantities of mouse Vasa protein [65]. This same work showed that *SYCP3* mRNA was enriched in a Dazl immunoprecipitation, suggesting that it is a potential target for Dazl action, and we have shown here that *SYCP3* levels are indeed reduced in the *Dazl* mutant [65]. As most of the migrating germ cells should have arrived at the gonad by e12.5, this data suggests that loss of *Dazl* may result in defects in germ cell differentiation as the germ cell are arriving at the gonad, with major germ cell loss manifesting itself after gonocyte populations should be established. Loss of germ cells could occur through increased apoptosis or decreased proliferation, or a combination of the two. It is unclear, however, whether germ cell loss is a direct or indirect effect of disruption of *Dazl*.

Based on the assumption that *Dazl* has a role at meiosis, we might expect that major germ cell loss in the *Dazl* mutant should occur at different times in male and females, since male germ cells begin to undergo meiosis at puberty while female germ cells begin and arrest in meiosis before birth. We observed, however, that significant germ cell loss in both sexes at e14.5, which is around the time that female gonocytes enter meiosis and male gonocytes mitotically arrest. We also observed aberrant expression of a number of germ cell genes in the mutant that are turned on long before meiosis begins. These data strongly suggest that *Dazl* has a meiosis-independent role in germ cell development. As the Dazl protein has the ability to bind the 3'UTRs of many different genes, and has also been shown to promote translation, one possible explanation of these results is that in the absence of Dazl, many important genes are not translated, which could be read by our real time PCR since lack of translation may affect the

stability of the mRNA. Without the expression of key genes, germ cells may not mature along the proper timecourse. In females, for instance, the high expression of early germ cell genes and low expression of *Vasa* and meiotic markers in the mutant may represent an inability of the germ cells to progress past the mitotic gonocyte stage. The low expression of almost all the germ cell markers in male mutant cells also suggests incomplete differentiation. This may affect other germ cell development milestones, such as erasure of genomic imprints, which also occurs during this window. Incomplete germ cell differentiation could trigger germ cell checkpoints, such as the pachytene checkpoint in meiosis, and lead to apoptosis. It may seem surprising to see *SYCP3* and *SYCP1* expression in embryonic male germ cells, but though male germ cells do not enter meiosis until puberty, they have also been shown to express SYCP3 and SYCP1 protein embryonically for a short time at e13.5 [132]. This may reflect a germ cell intrinsic property to begin meiosis that is suppressed by the testicular environment, and another potential trigger for a developmental checkpoint. This model is summarized in Fig. 4.

The fact that germ cell loss in the *Dazl* mutant, as measured quantitatively by FACS, occurs earlier than previously observed by histology, and that we were able to see a decrease in germ cell number in female heterozygotes, underscores the utility and importance of this technique for understanding germ cell phenotypes. This method allows for the rapid analysis of a large number of embryos which can identify smaller differences not readily observed by eye. Moreover, germ cells can be isolated and collected, and studied biochemically, or molecularly as described here. By examining a relatively pure population of germ cells, we can learn more about what is happening to the germ cells themselves, in addition to how they affect the whole animal. In this case,

we can see that Dazl dose-dependent effect in females, as well as in males. We can also see that not only does disruption of the *Dazl* gene lead to infertility in both sexes, but that this infertility stems from an early loss of germ cells shortly after they arrive at the gonad, and that the loss itself may arise from incomplete germ cell differentiation triggering developmental checkpoints. Future studies examining cell proliferation, apoptosis, and imprinting should help us understand more about how any why germ cells are lost.

Acknowledgements

We would like to thank Chris Wylie for the Oct4ΔPE-GFP mice, Howard Cooke for the *Dazl* knockout mice, and Shuwei Jiang for excellent assistance in cell sorting. This work was supported by fellowships from the National Science Foundation and Achievement Rewards for College Scientists to JYT.

Figure Legends

Fig. 1 GFP expression from Oct4ΔPE promoter in the germ cells of wild type e14.5 (A) testis and (B) ovary

Fig. 2 Germ cell counts in (A, B) e12.5, (C, D) e14.5, and (E, F) e16.5 embryos. Counts for testes are in the left column, counts for ovaries in the right. At e12.5, each bar represents the germ cell count for one animal, whereas at e14.5 and e16.5, each bar represents the germ cell count for one gonad. The number of germ cells is on the y-axis. '*' denotes a statistically significant difference in germ cell count with the mutant with

$p<0.05$. '**' denotes a statistically significant difference in germ cell count with the mutant with $p<0.001$.

Fig. 3 Germ cell gene expression in male (A) and female (B) germ cells. Expression profiles are shown at e12.5, e13.5, e14.5, and e15.5 and genes are divided into those that are expressed during migrating PGCs and those that are expressed after the germ cells reach the gonad. Each bar on the graph represents the average of expression of sorted germ cells from three to seventeen embryos. Expression levels on the y-axis are normalized to a housekeeping gene, *GAPDH.*

Fig. 4 Model for *Dazl* function in germ cell development. Dazl protein is expressed from migrating primordial germ cells all the way to mature sperm and oocytes. In the mutant, expression of *Pumilio-2*, *Oct-4*, *Stella*, *c-kit*, *Vasa*, *SYCP3* and *SYCP1* are all affected (normal expression patterns noted by gray lines, as in [131]), which suggests that Dazl may directly or indirectly regulate the expression of these genes, most likely translationally. The absence of *Dazl* results in incomplete germ cell differentiation, which may trigger developmental checkpoints at the female transition to meiosis, at the male mitotic arrest, and at the male transition to meiosis (checkmarks). It may also affect other developmental milestones in this window, such as erasure of genomic imprints.

Table 1 Primers used for real time PCR

Gene	Sequence
Pumilio2	Forward: TATCAATTAATGACAGATGTTTTTGGA
	Reverse: CTGATCAGAAGAAATAGATTCTAACGC
Nanos2	Forward: GGAATAAGAGGAAGGTACAGGAACTA
	Reverse: TATATTGGATGGGTAGAAGAGAGAGAA
Nanos3	Forward: CTACCTTCGTCTACTGCTACACCAC
	Reverse: ACTTTTGGAACCTGCATAGACACCT
Vasa	Forward: CTAGGAAGACCAAATAGTGAATCTGAC
	Reverse: TCCAGAACCTGTTACTACTTCTTCATT
Stella	Forward: CTTTTCAAAGACTAAGCAATCTTGTTC
	Reverse: ATGACCTTTACTAGTGTTTCTGGTTGT
Oct-4	Forward: AGTCTGGAGACCATGTTTCTGAAGT
	Reverse: TACTCTTCTCGTTGGGAATACTCAATA
c-kit	Forward: AGAATATTGTTGCTATGGTGATCTTTT
	Reverse: CATGTCCATATATTCATTTGAACTGTC
SYCP3	Forward: AGAAATGTATACCAAAGCTTCTTTCAA
	Reverse: TTAGATAGTTTTTCTCCTTGTTCCTCA
SYCP1	Forward: AAGTTTGATTCTAAAACAACTCCTTCA
	Reverse: ACTCTTTTTAGTTGGTGTCTTCACTGT
GAPDH	Forward: TTCACCACCATGGAGAAGGC
	Reverse: GGCATGGACTGTGGTCATGA

Fig. 1

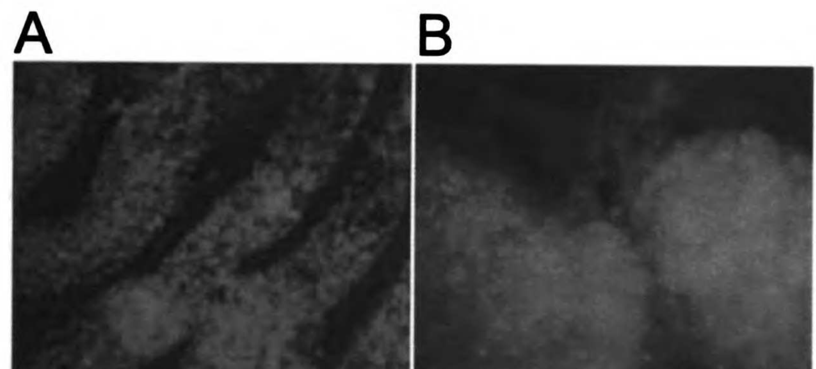


Fig. 2

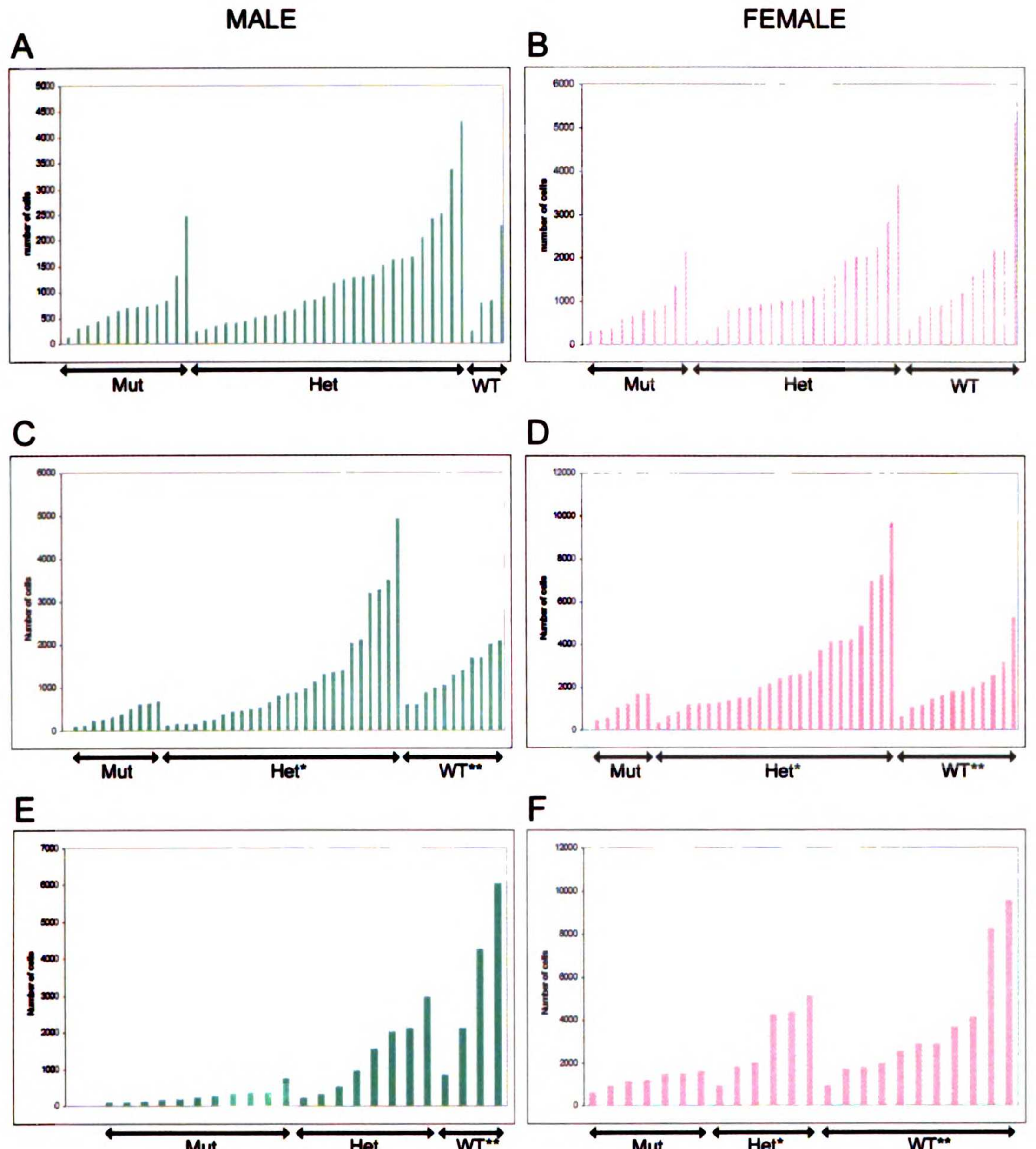


Fig. 3

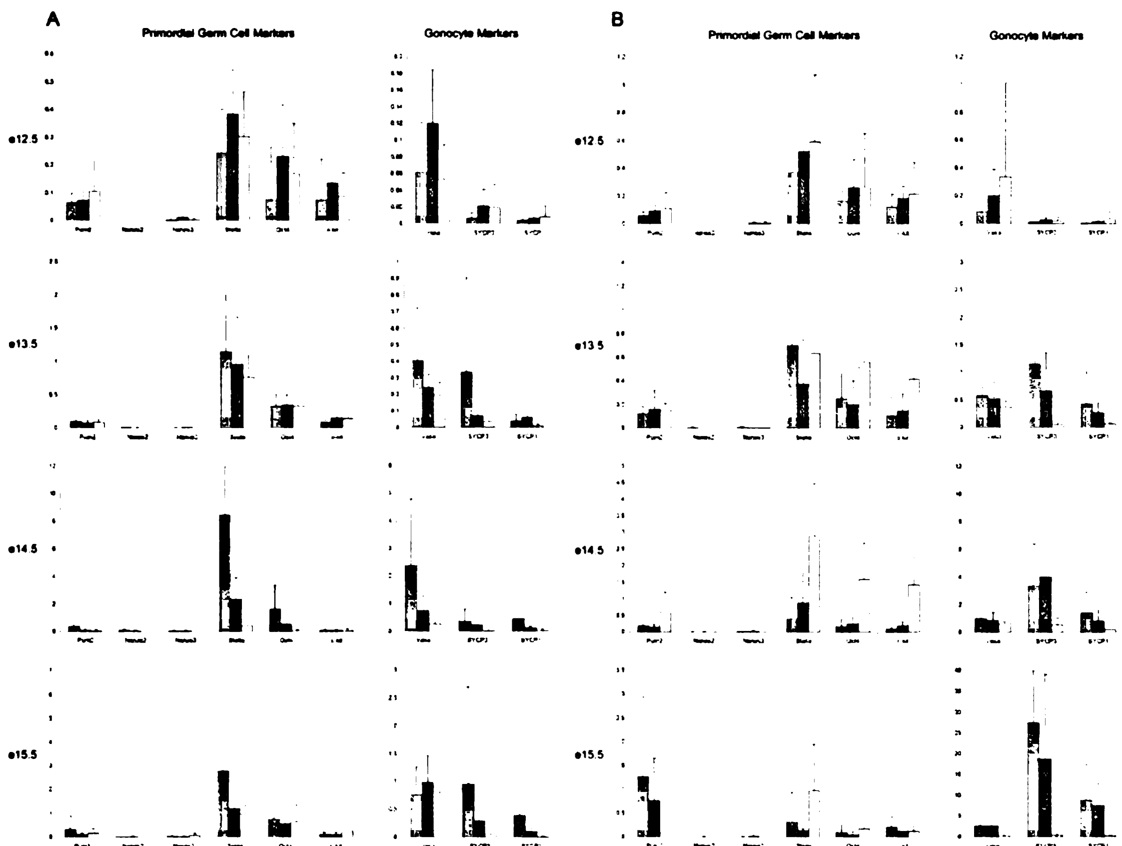
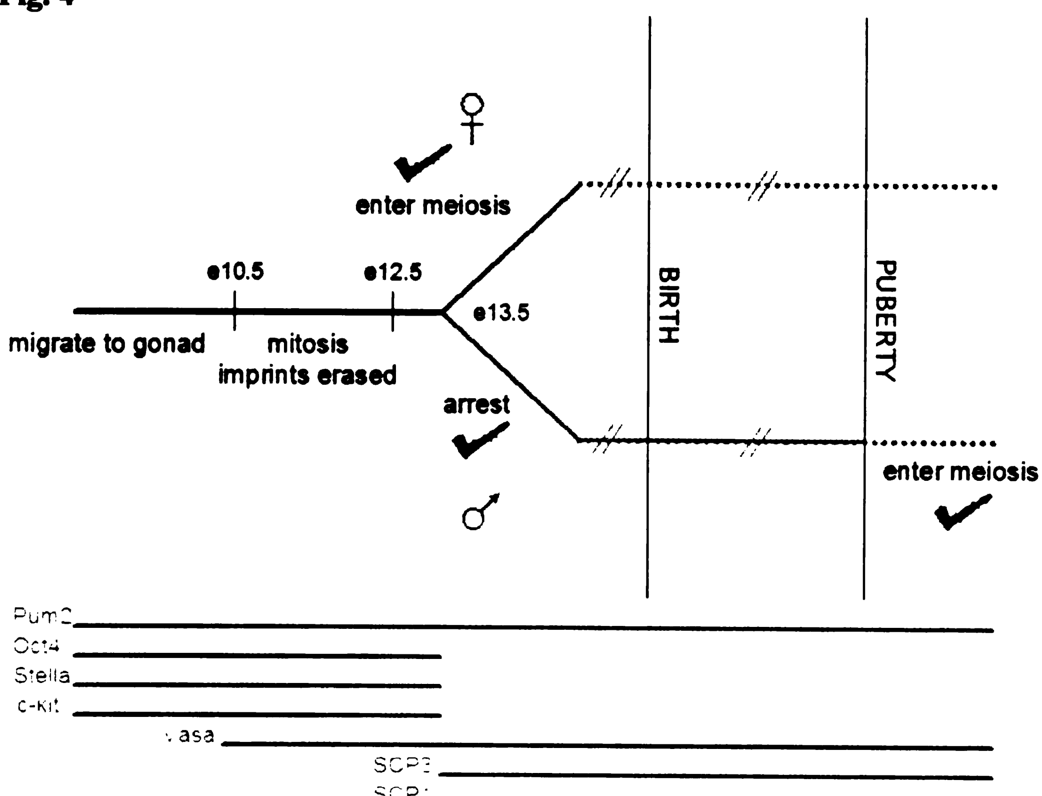


Fig. 4



Chapter 5

Evolutionary comparison of the reproductive genes, *DAZL* and *BOULE*, in primates with and without *DAZ*

Originally published in *Development, Genes, & Evolution*
Published here with kind permission of Springer Science and Business Media

Evolutionary comparison of the reproductive genes, *DAZL* and *BOULE,* in primates with and without *DAZ*

Joyce Y. Tung, C. Marc Luetjens, Joachim Wistuba, Eugene Y. Xu, Renee A. Reijo Pera, Jörg Gromoll

Affiliations:
J. Tung, E. Xu, R. Reijo Pera
Center for Reproductive Sciences; Department of Obstetrics, Gynecology and Reproductive Sciences; Departments of Physiology and Urology; Programs in Human Genetics, Cancer Genetics and Development, and Stem Cell Biology
University of California at San Francisco
Box 0556
San Francisco, CA 94143-0556, USA

C. Luetjens, J. Wistuba, J. Gromoll
Institute of Reproductive Medicine, Westphalian Wilhelms-University
D-48129 Muenster, Germany

Correspondence: Renee A. Reijo Pera, University of California San Francisco, Box 0556, San Francisco, CA 94143-0556. FAX: 415 476 3121. E-mail: reijo@itsa.ucsf.edu

Abstract Genes of the *DAZ (Deleted in AZoospermia*) gene family, *DAZ, DAZL (DAZ-Like)* and *BOULE,* encode closely-related RNA-binding proteins that are required for fertility in numerous organisms, yet the genomes of different organisms possess different complements of *DAZ* family genes. Thus, invertebrates such as flies and worms contain just a single *DAZ* homolog, *boule*, while genomes of vertebrates, other than catarrhine primates (Old World monkeys and hominids), possess both *Boule* and *Dazl* genes. Finally, catarrhine primates possess *BOULE, DAZL* and *DAZ* genes. Since the *DAZ* genes arose recently in evolution in the catarrhine lineage, we sought to examine how the sequences and expression of this gene family may have changed, after introduction of a new member, *DAZ*. Based on previous results, we hypothesized that introduction of a new member of the *DAZ* gene family into catarrhines could reduce functional constraint on *DAZL*. Surprisingly, however, we found that platyrrhine *DAZL* demonstrated significantly more sequence divergence than catarrhine *DAZL* ($p = 0.0006$ for nucleotide and $p = 0.05$ for amino acid sequence); however, comparison of K_a/K_s ratios suggests that the *DAZL* and *BOULE* genes are under similar functional constraints regardless of lineage. Thus, our data is most consistent with the hypothesis that the introduction of *DAZ* did not affect the evolution of *DAZL* or *BOULE*, and that a higher neutral mutation rate in platyrrhines than in catarrhines, along with the greater tolerance of *DAZL* for variation relative to *BOULE,* may be the foundation for the observed differences in sequence divergence in this gene family.

Keywords Reproduction, Evolution, Primates, *DAZ, BOULE*

Introduction

Every gene functions not only as part of the entire genome, but also more directly, as part of a gene family. Therefore it is critical to examine how genes change and develop in the context of their families, especially when a new family member is introduced. A particularly interesting reproductive gene family is the *DAZ (Deleted in AZoospermia)* gene family (Fig. 1). Evidence from studies of infertile men indicates that the *DAZ* genes may function early in germ line stem cells. Men with deletions encompassing the Y chromosome *DAZ* gene cluster have defects in spermatogenesis beginning in the stem cell populations; these men frequently lack all germ cells and only somatic cells are present. In addition there are two other members of the *DAZ* gene family in humans—*DAZL* and *BOULE*—which are ancestral to *DAZ*. All of these genes are only expressed in germ cells in humans and where homologs of the *DAZ* gene family have been identified throughout metazoans, they are inevitably expressed only in the germ cell lineage (Fig. 1). They have also been shown to be required for germ cell development in many different species [68, 99-102, 104]. Thus, it is curious that genes that are apparently essential for reproduction have evolved into such a diverse family with different species containing different complements of *DAZ* genes. Note that in some cases, authors have suggested that the human Y chromosome genes may not be required as the genes have diverged significantly through time; the authors did not, however, discount positive selection [74]. Indeed, other authors have suggested that each of the genes of this family may be required for optimal fertility in men [133].

We and others recently sequenced the human *BOULE* gene and observed that there were few sequence variants in a population of 200 individuals [108, 124]. In contrast, sequence databases and sequence analysis record the high frequency of variants in human *DAZL* with common variants approximately 1/100 basepairs or less (http://genome.ucsc.edu, see Chapter 2; [70, 72, 134]. This contrast was surprising since *BOULE* and *DAZL* are members of the same gene family, share a highly-conserved functional RNA-binding domain, and have homologs that are required for germ cell development in diverse model organisms [68, 99, 102, 104]. We reasoned that greater sequence variation in *DAZL* than in *BOULE* might be linked to the introduction of the Y chromosome *DAZ* genes, which are most closely related to *DAZL*, into the catarrhine (Old World monkey and hominid) lineage. In other words, we reasoned that the introduction of the Y chromosome *DAZ* genes may have provided a gene environment conducive to the accumulation of more changes in *DAZL* than in *BOULE*, by accommodating part of the functional burden of *DAZL* in males and relaxing the functional constraint on *DAZL*. Indeed, the duplication-degeneration-complementation model suggests that duplication and degeneration of a gene may be tolerated so long as together the duplicated genes maintain the function of the ancestral gene [135]. In this study, we compared *BOULE* and *DAZL* sequences and expression patterns in closely related lineages in the presence (catarrhini) or absence (platyrrhini, or New World monkeys) of the Y-chromosomal *DAZ* genes in order to examine changes that may have occurred in each lineage, including evidence for loss of functional constraint on the *DAZL* or *BOULE* genes when in the presence of *DAZ*.

Materials and Methods

Tissue collection

Testes samples were obtained from different primate species as part of a separate project funded by the German Research Foundation [136] and were snap-frozen in liquid nitrogen. Testes were obtained from the following species: Rodent: *Mus musculus*; Strepsirhini (prosimians): *Microcebus murinus* (gray mouse lemur); Platyrrhini (New World monkeys), *Callithrix jacchus* (common marmoset), *Saguinus oedipus* (cotton top tamarin), *Saimiri sciureus* (squirrel monkey); Catarrhini: Cercopithecidae (Old World monkeys): *Macaca fascicularis* (long-tailed, crab-eating or cynomologus monkey) and Hominoidea (Great Apes): *Pan paniscus* (bonobo, pygmy chimpanzee), *Homo sapiens sapiens* (human). Primate tissues were obtained from various primate research centers facilities via the German Primate Center (Göttingen, Germany). Testis tissues from the cynomologus monkey, *Macaca fascicularis*, and the marmoset monkey, *Callithrix jacchus*, were obtained from the primate colonies maintained at the Institute of Reproductive Medicine. Human testis samples were obtained from prostate cancer patients and were provided via Dr. T.G. Cooper from the Institute of Reproductive Medicine. All men provided informed consent according to protocols approved by the local ethics committee. Testes samples from these patients had normal spermatogenesis. Analysis of other primate samples was confined to samples from reproductively active males and demonstrated normal spermatogenesis. When necessary, tissues were fixed in Bouin´s solution and embedded in paraffin and prepared for immunohistochemistry analysis and assessment of germ cell associations, as previously reported [58, 67, 137, 138].

Oligonucleotides

Oligonucleotides for RT-PCR reactions were designed as consensus sequences from highly conserved nucleotide sequence stretches in the 5' and 3' ends of the open reading frames of human and mouse *BOULE* and *DAZL*. Sequences were: *BOULE:* 5′-CAT CAA ACC AGA TGC AAA CAG ATT C-3′, 5′-GCT GGT TCG TTG AAG CTG GAT CTC-3′. *DAZL*: 5′-CGC GGA TCC ATG TCT GCT GCA AAT CCT GAA AC-3′, 5′-CCG CTC GAG CAT AGC CAG GAG GAT CAA ACA G-3′

RT-PCR and DNA sequencing

RNA isolation from testes tissues was performed using Ultraspec (AMS Biotechnology, Germany), followed by DNAse digestion with the DNA-free Kit (Ambion, Austin, TX, USA). Five μg total RNA were reverse transcribed using the appropriate reverse primer and Superscript Transcriptase (Invitrogen, Karlsruhe, Germany). One to two μg of cDNA were used for each of the subsequent PCR reactions. PCR cycle conditions were: 1 cycle of 94°C, 2 min; 30 cycles of 94°C, 50 sec, 58°C, 40 sec, and 72°C, 60 sec; 1 cycle of 72°C 10 min. Reaction products were electrophoresed on 2% agarose. PCR-amplified DNA was purified using the High Pure Kit (Roche, Penzburg, Germany) and either directly sequenced or cloned into the pGEM-T Easy Vector and subsequently sequenced. A minimum of 5 different clones per amplicon were sequenced to minimize sequence errors due to PCR. DNA sequences (Fig. 3) indicate consensus sequences obtained from these clones using Sequencher software (Genecodes, Ann Arbor, MI). Corresponding

cDNA sequences that were generated as part of this study were deposited in the EMBL database under the following accession numbers: AJ717405, AJ717406, AJ717407, AJ717408, AJ717409, AJ717410, and AJ717411.

Immunohistochemistry

Immunohistochemistry was performed as described [58, 137, 139]. Testes sections were incubated with primary antibodies in blocking buffer for 60 min, rinsed and incubated with the DAKO-LSAB 2 System (DAKO Diagnostika, Hamburg, Germany), 30 min, then rinsed thoroughly and incubated with DAB (DAKO Diagnostika, Hamburg, Germany) to develop color, 20 minutes. Controls were performed by omitting the primary antibody on adjacent sections. Sections were counterstained with hematoxylin for 10 seconds, mounted and photographed (Axioskop, Zeiss, Oberkochem, Germany; objectives 25x and 40x). Digital images of equal exposure were acquired with a CCD camera (Axiocam, Gottingen, Germany) controlled by image software (Axiovision, Gottingen, Germany). For testes of all species, nonspecific staining in peritubular cells, blood capillaries, and some Leydig cells was observed with the BOULE antisera.

Phylogenetic analysis

Sequences were aligned with MegAlign (DNASTAR, Madison, WI, USA). Phylogenetic trees and branch lengths were calculated and drawn using the DnaML, ProML, Retree, and Drawtree programs from PHYLIP (Phylogeny Inference Package, version 3.6a3, distributed by J. Felsenstein, Department of Genetics, University of Washington, Seattle),

with a transition/transversion ratio of 2.0, empirical base frequencies, unweighted sites, and no rate variation among sites. K_a and K_s values were determined using K-Estimator 6.0 as described [140]. Statistics from the relative rate test, as developed by Fumio Tajima, were calculated using MEGA (Molecular Evolutionary Genetics Analysis) 2.1 [141, 142].

Results and Discussion

Conserved and variable amino acids in DAZ homologs

We compared substitutions in the protein sequences of DAZ homologs, in the known functional domains, in order to gain insight into which amino acids are most conserved and which are most variable. Species with homologs that were examined included three catarrhine primates: *Homo sapiens* (human), *Pan paniscus* (bonobo), and *Macaca fascicularis* (crab-eating macaque); and three platyrrhine primates: *Saimiri sciureus* (squirrel monkey), *Saguinus oedipus* (cotton-top tamarin), and *Callithrix jacchus* (common marmoset) (see Fig. 3a for phylogeny). Alignment of catarrhine and platyrrhine DAZL protein sequences with the sequence of their common ancestor (deduced by PHYLIP's maximum likelihood program) is as shown (Fig. 2a). We observed that there were only two amino acids changed within the two well-characterized functional domains, the RNA-binding RNA Recognition Motif (RRM) and the DAZ repeat. Within the RRM, a hydrophilic serine at position 48 has been substituted with a small, hydrophobic glycine in the Callitrichidae primates, *C. jacchus* and *S. oedipus*. In the DAZ repeat, Pro170 has been substituted with a threonine in the three catarrhine

primates. In contrast, a number of substitutions were scattered throughout the linker region of DAZL: seven in the platyrrhine primates, but only two in the catarrhines.

A similar alignment of simian BOULE sequences demonstrated that the majority of amino acid substitutions in this homolog also occur outside the RRM and DAZ repeat (Fig. 2b). However, we also observed that in general, there were very few amino acid substitutions even outside these domains, during the evolution of BOULE, and that the changes that did occur were conservative. For example, the only substitution that may potentially alter protein function was observed in *M. fascicularis*; Thr187 is substituted for an alanine.

Unrooted phylogenetic trees

We next examined the rate of sequence variation of *DAZL* and *BOULE* in the primate lineages described above, and in addition, a lemur from the strepsirhine lineage, *Microcebus murinus* (gray mouse lemur) (Fig. 3a). Using programs that implemented the maximum likelihood method, we deduced unrooted phylogenetic trees for the coding sequences of *DAZL* and *BOULE*, using mouse (*Mus musculus*) sequences as an outgroup (Fig. 3b and 3c). In unrooted phylogenetic trees, branch lengths are proportional to the expected number of substitutions per site. Thus, we predicted that the introduction of the *DAZ* genes into the catarrhine lineages would be reflected by longer branch lengths in catarrhine *DAZL* genes than in platyrrhine *DAZL* genes, relative to their common ancestor. However, when we examined the topology of the tree, we observed the opposite.

To test the significance of this observation, the rate of evolution along branches was examined using the Tajima relative rate test [141]. The results of this test are

summarized in Table 2. Sequences from two species and an outgroup were compared, and the values m1, m2, and m3 denote the number of unique nucleotides in Species 1, Species 2, and the outgroup, respectively. As we used *M. musculus* as the outgroup, this test compares the rate of evolution of Species 1 and Species 2 since their divergence from their common ancestor with the mouse. Table 2 shows that *H. sapiens* and *P. paniscus* *DAZL* have evolved significantly slower than *S. oedipus* and *S. sciureus DAZL*. The same trend was observed for *M. fascicularis* and *C. jacchus*, though the differences were not significant. In contrast, the *BOULE* gene shows no evidence of greater sequence variation (or faster evolution) in either the catarrhine or the platyrrhine lineage; in fact, the comparison of *P. paniscus* with *C. jacchus* is the only instance in which one of the two lineages has evolved significantly more quickly. Within the platyrrhine lineage, however, *S. oedipus* appears to have evolved more quickly than the other two species. This test demonstrated that the difference in evolutionary rate observed between platyrrhine and catarrhine *DAZL* was significant, while there was no significant difference between the platyrrhine and catarrhine *BOULE* genes.

Possible explanations for accelerated evolution

We next considered possible explanations for the accelerated evolution of platyrrhine *DAZL* relative to catarrhine *DAZL*. One possibility is that *DAZL* has undergone positive selection in platyrrhine primates. In that case, we might suppose that changes in *DAZL* that increase fertility, for instance, through an increase in the number of germ cells, would increase individual fitness and be maintained. However, when we investigated whether *DAZL* and *BOULE* underwent positive selection by calculating the rate of

protein sequence evolution scaled to the neutral mutation rate (the K_a/K_s ratio), we found no evidence of positive selection (Table 3). To test for positive selection in individual domains, we calculated K_a/K_s ratios for five regions in *DAZL* and *BOULE*: the N-terminus up to the RNA Recognition Motif (RRM), the RRM, the linker between the RRM and the DAZ repeat, and the C-terminus after the DAZ repeat (see Fig. 2). No evidence of positive selection was observed in any of these domains (data not shown).

Given that there was no evidence for positive selection in the *DAZL* gene, another possibility for the observed differences in evolutionary rates is that the base neutral mutation rate in platyrrhines is higher than in catarrhines, and that the relative conservation of *BOULE* compared to *DAZL* simply makes it difficult to observe such a difference between lineages in the *BOULE* gene. We tested this hypothesis by comparing the synonymous substitution rates (K_S), as an approximation of the neutral mutation rate, between different primates in each lineage for both the *BOULE* and *DAZL* genes (Table 4). We observed that both *BOULE* and *DAZL* exhibit similar synonymous substitution rates in catarrhine primates, whereas both genes exhibit higher K_s values in platyrrhine primates, even though the K_s values of *DAZL* are larger than those of *BOULE*. These observations strongly suggested that the neutral mutation rate may also be higher in the platyrrhines.

Variation in expression of DAZ homologs

As the DAZ protein is only expressed in males, we performed immunohistochemistry on testes sections to examine DAZL and BOULE protein expression patterns in the catarrhine and platyrrhine lineages. In the testes of the mouse, primates, and humans, we

observed that expression of BOULE protein was generally restricted to cells in the first meiotic division (prophase) beginning at the zygotene spermatocyte stage and reaching its highest level in pachytene spermatocytes (Fig. 4). The only exception to this expression pattern was observed in the cebid platyrrhine, *S. sciureus,* in which the onset of expression of BOULE protein was shifted to more advanced stages of pachytene spermatocytes. In postmeiotic cells, significant variation in expression was observed in different species, independent of taxonomic group; for example, BOULE protein expression in five species, *P. paniscus*, *M. fascicularis*, *S. sciureus*, *S. oedipus*, and *M. murinus*, was detected in secondary spermatocytes and in very early round spermatids but was not detected in more advanced stages of germ cells (Fig. 4). In contrast, in the neotropical platyrrhine *C. jacchus*, BOULE expression was confined to pachytene spermatocytes only (Fig. 4d).

In several species, expression of DAZL overlapped that of BOULE protein. In humans and the prosimian, *M. murinus*, DAZL is detected in all germ cell types with higher levels of expression in some cell types (as previously reported in humans [137], and Fig. 5). Thus, in contrast to BOULE, DAZL expression is not restricted to meiotic and/or postmeiotic stages. Some quantitative differences in DAZL expression were also observed in histological sections. Sections from the testes of *P. paniscus*, *S. sciureus*, *C. jacchus*, and *M. murinus*, had little (or even no) expression of DAZL prior to spermatid elongation, whereas, expression persisted in elongating spermatids in the human, *M. fascicularis* and *S. oedipus* (Figs. 4c2, 4g, 5). Most notably in this analysis, however, was the observation that there are no significant differences in the expression patterns that we

observed in sections from catarrhine primates compared to those from platyrrhine primates (Fig. 4a2-c2).

The introduction of *DAZ* did not affect *DAZL* and *BOULE* evolution

The *DAZ* gene family contains three members that encode RNA-binding proteins that possess distinct functions even though all family members arose from a common ancestor. For example, even though *DAZL* and *DAZ* are more than 90% identical, neither can completely compensate for the loss of the other, and loss of function of these genes is not compensated for by the presence of *BOULE* [27, 68]. Thus, men with Y chromosome deletions that encompass the *DAZ* genes make few or no sperm even though presumably functional *DAZL* and *BOULE* genes are present in the genome [27]. In addition, null mutations in the mouse *Dazl* gene result in infertility in both male and female mice even though a functional *Boule* gene remains in the genome [68]. These observations suggest that this gene family has actively evolved to encompass multiple functions in gametogenesis.

We hypothesized that introduction of *DAZ* into catarrhine primates might be accompanied by relaxed functional constraints on *DAZL* in that lineage, as it has been observed that the *DAZL* gene is highly variant in humans (http://genome.ucsc.edu; see Chapter 2; [70, 72, 134]. Though some have argued that the four copies of DAZ on the Y chromosome are of limited function, there is ample evidence that DAZ does function in germ cell development [74]. At least three of the four *DAZ* copies are expressed, and maximum likelihood analysis by Bielawski et al. indicated that both *DAZL* and *DAZ* experienced increased rates of nonsynonymous substitution after a copy of *DAZL* was

translocated to the Y chromosome to form *DAZ* [75, 143]. In addition, introduction of a human *DAZ* transgene in the mouse *Dazl* mutant resulted in partial rescue [144]. To our surprise, however, our results suggested that platyrrhine *DAZL* evolved more quickly than catarrhine *DAZL*. By comparing branch lengths and using the Tajima relative rate test, we demonstrated that there were significant differences between the rates of evolution of catarrhine and platyrrhine *DAZL*, but not of catarrhine and platyrrhine *BOULE*. The Tajima test compared sequences in a pairwise fashion, which showed that all catarrhine *DAZL* sequences evolved more slowly than all platyrrhine *DAZL* sequences, though only *H. sapiens* and *P. paniscus DAZL* evolved significantly more slowly. For *BOULE*, neither lineage appeared to evolve more quickly; there are significant differences in the rate of evolution within the platyrrhine *BOULE* lineage, but the differences across lineages were not significant.

As the *DAZL* gene is likely to be required either for the allocation or the maintenance of the germ cell population, a second possibility is that substitutions in platyrrhine *DAZL* have led to an improvement in germ cell production (e.g. an increase in sperm count in males) that has been selected for in the platyrrhine lineage [68, 99]. A previous study of spermatogenic efficiency and the structural organization of the seminiferous epithelium, however, argues against this hypothesis [136]. We speculated that positive selection may have driven the evolution of *DAZL* in the platyrrhine lineage; however, we found no evidence of positive selection.

A third possibility is that the neutral mutation rate in platyrrhines is higher than in catarrhines but that the difference was not apparent for *BOULE* because its sequence is more constrained, as noted previously. Analysis of the synonymous substitution rate in

each lineage shows that there are clearly more synonymous changes in the platyrrhine lineage for both genes, which supports this hypothesis. K_a/K_s ratios are generally similar for both genes across lineages, but the majority are slightly lower for *BOULE* in intralineage comparisons, consistent with the protein sequence comparison results. This suggests that each gene is under similar functional constraints in both catarrhine and platyrrhine primates, but that *DAZL* is more tolerant of variation than *BOULE* even after evolving a different function. We cannot exclude the possibility that *BOULE* has also gained new functions, but in a previous study, a human *BOULE* transgene expressed in *boule* mutant flies exhibited the same degree rescue as a fly *boule* transgene, providing evidence that the function of *BOULE* has been strongly conserved across diverse species [108]. In addition, embryonic analysis of the mouse *Dazl* phenotype suggests that *Dazl* has developed a function independent of meiosis in that species, while the expression pattern of *BOULE* suggests that it only functions at meiosis (see Chapter 4; [58]). As our data shows no relaxation of functional constraints in either lineage, it appears that the introduction of *DAZ* did not affect the rate of *DAZL* or *BOULE* evolution. Although males in the catarrhine lineage express both Y-chromosome *DAZ* and autosomal *DAZL*, the females must still rely on the *DAZL* gene exclusively. Data from model organisms shows that *DAZL* homologs are required for female fertility [68, 99]. Thus, the functional constraints on *DAZL* in females may have limited the number of amino acid changes *DAZL* could tolerate despite the introduction of *DAZ*, a gene likely of similar function, in males.

Why would there be more nucleotide changes in the platyrrhine lineage than in the catrrhine lineage? There are several papers that support a hominoid rate-slowdown

hypothesis, in which the evolutionary rate in the hominid lineage has slowed down considerably since the split with Old World monkeys, while Old World monkeys and New World monkeys appear to evolve at a similar rate [145-147]. This is supported by the generation time effect hypothesis, which states that the shorter the generation time, the more mutations can be passed on in a fixed amount of time, thus increasing the evolutionary rate [148]. Old World monkeys and New World monkeys have short generation times (macaque 3-5 years, marmoset 1.5 years) while great apes and humans have longer generation times (chimpanzee 8.7 years, human 17 years) [149, 150]. This hypothesis has proved controversial, and contradicting evidence has been published [151]. In our analysis, the New World monkeys (*S. sciureus*, *S. oedipus*, *C. jacchus*) did exhibit higher substitution rates than the hominids (*H. sapiens*, *P. paniscus*), while the Old World monkey (*M. fascicularis*) demonstrated an intermediate substitution rate reflective of its intermediate generation time, consistent with the generation time effect hypothesis.

In summary, the evolution of the *DAZ* gene family is intriguing to consider. However, simple explanations for its current structure and function are unlikely to be uncovered and additional theoretical and functional analyses are merited.

Acknowledgements The authors thank Nicole Terwort and Jutta Salzig for excellent technical assistance, and Joanna Gonsalves for helpful comments. This work was supported by a National Science Foundation Fellowship (J.Y.T.), an Achievement Rewards for College Scientists Fellowship (J.Y.T.), by grants from the Fritz-Thyssen Foundation and the IMF (GR 110311, to J.G.), by grants from the German Research

Figure Legends

Fig. 1 Model of evolutionary history of the *DAZ* gene family. *BOULE* is the ancestral gene and is found in invertebrates and vertebrates. *DAZL*, which has been found in all vertebrate species studied thus far, probably arose by duplication of *BOULE* in the vertebrate lineage, while *DAZ* arose by multiple duplications of *DAZL* in the catarrhine lineage and can only be found in Old World monkeys and hominids. An open box indicates the inferred presence of a *BOULE* homolog that has yet to be identified.

Fig. 2 Alignment of simian protein sequences with ancestral sequence, as deduced by the maximum likelihood method. The RNA Recognition Motif is highlighted in pink, the DAZ repeat in blue, and amino acid changes from the ancestral sequence in yellow. The highly conserved RNP2 and RNP1 motifs in the RRM are boxed, and the table shows how many amino acid changes are found in each domain. The alignment for DAZL (**a**) also shows the regions required for homodimerization and interaction with the PUM2 protein [66, 67]. "*" denotes that the change lies in a region where the PUM2 interacting domain and the DAZ repeat overlap. The alignment for BOULE (**b**) also shows the region required for homodimerization and interaction with the DAZL protein [152]. "*" denotes that the change lies in a region where the homodimerization/DAZL interacting domain and the DAZ repeat overlap.

Fig. 3 Phylogeny of the primates used in this study (**a**). Unrooted phylogenetic trees were constructed for (**b**) *DAZL* and (**c**) *BOULE* using primate coding sequences. The arrow indicates the most recent common ancestor of the catarrhine and platyrrhine lineages.

Branch lengths are proportional to the expected number of changes per site. *DAZL* sequence accession numbers: *M. musculus* NM_010021, *M. murinus* AJ746580, *S. sciureus* AJ717411, *S. oedipus* AJ717410, *C. jacchus* AF144131, *P. paniscus* AJ717409, *M. fascicularis* AF144132, *H. sapiens* NM_001351. *BOULE* sequence accession numbers: *M. musculus* AF272859, *M. murinus* AJ746579, *S. sciureus* AJ717408, *S. oedipus* AJ717406, *C. jacchus* AJ717407, *P. paniscus* AJ717405, *M. fascicularis* AB074454, *H. sapiens* NM_033030.

Fig. 4 Immunohistochemical staining of BOULE (left column) and DAZL (right column) in testes sections from seven different species *Pan paniscus* (**a1-a2**), *Saimiri sciureus* (**b1-b2**), *Microcebus murinus* (**c1-c2**), *Callithrix jacchus* (**d**), *Saguinus oedipus* (**e**), *Mus musculus* (**f**), and *Macaca fascicularis* (**g**). For comparison purposes, we only considered strong protein expression. Scale bar = 100μm.

Fig. 5 A model showing all spermatogenic cell types demonstrating strong BOULE (light gray) and DAZ/DAZL (dark gray) expression for eight species: *Homo sapiens sapiens*, *Pan paniscus, Macaca fascicularis, Saimiri sciureus, Callithrix jacchus*, *Saguinus oedipus*, *Microcebus murinus,* and *Mus musculus.* Dark gray lines denote DAZ/DAZL expression as described in previous studies [137]. BOULE and DAZL expression begins in meiotic stages (mainly pachytene) or earlier, and persists in some species until the spermiation phase, after which any persistent expression in maturing germ cells is species specific. In catarrhine primates, both the cytoplasm and the nuclei were stained for DAZL. The expression window for BOULE in maturing germ cells is generally shorter then that

for DAZL. In most of the primates examined here, DAZL expression reaches up to the mid-spermiation phase, while BOULE expression is limited to late spermatocytes. Cell types: type A spermatogonia (A), type B spermatogonia (B), preleptotene (Pl), leptotene (L), zygotene (Z), pachytene (P), secondary spermatocyte (S), round spermatids (R), elongating spermatids (E), and spermatids (L).

Table 1 Comparison of branch lengths from the common ancestor of platyrrhines and catarrhines

DAZL CODING NUCLEOTIDE			
Catarrhini		**Platyrrhini**	
H. sapiens	0.0138[a]	*S. sciureus*	0.0273
P. paniscus	0.0150	*S. oedipus*	0.0267
M. fascicularis	0.0174	*C. jacchus*	0.0289
***DAZL* PROTEIN**			
Catarrhini		**Platyrrhini**	
H. sapiens	1.992[b]	*S. sciureus*	3.657
P. paniscus	1.993	*S. oedipus*	2.326
M. fascicularis	1.664	*C. jacchus*	2.987
***BOULE* CODING NUCLEOTIDE**			
Catarrhini		**Platyrrhini**	
H. sapiens	0.0147	*S. sciureus*	0.0123
P. paniscus	0.0135	*S. oedipus*	0.0206
M. fascicularis	0.0127	*C. jacchus*	0.0185
***BOULE* PROTEIN**			
Catarrhini		**Platyrrhini**	
H. sapiens	1.857	*S. sciureus*	1.450
P. paniscus	2.216	*S. oedipus*	2.215
M. fascicularis	2.099	*C. jacchus*	2.207

[a] Branch lengths for both *DAZL* and *BOULE* were calculated for platyrrhine and catarrhine primates from their most recent common ancestor using the maximum likelihood method. For nucleotide sequences, branch lengths are proportional to the expected number of substitutions per site.

[b] For protein sequences, branch lengths are proportional to the expected number of substitutions per 100 sites.

Table 2 Pairwise comparison of evolutionary rates using the Tajima relative rate test

DAZL coding nucleotide						
Species 1	**Species 2**	**M1**	**m2**	**m3**	**chi-square**	**p-value**
Catarrhine vs. Catarrhine						
H. sapiens	*P. paniscus*	1	2	91	0.33	0.564
H. sapiens	*M. fascicularis*	4	8	88	1.33	0.248
P. paniscus	*M. fascicularis*	5	8	88	0.69	0.405
Platyrrhine vs. Platyrrhine						
S. sciureus	*S. oedipus*	11	12	91	0.04	0.835
S. sciureus	*C. jacchus*	14	12	89	0.15	0.695
S. oedipus	*C. jacchus*	8	5	95	0.69	0.405
Catarrhine vs. Platyrrhine						
H. sapiens	*S. sciureus*	11	24	81	4.83	**0.028***
H. sapiens	*S. oedipus*	10	24	80	5.76	**0.016***
H. sapiens	*C. jacchus*	13	24	78	3.27	0.071
P. paniscus	*S. sciureus*	12	24	81	4	**0.046***
P. paniscus	*S. oedipus*	11	24	80	4.83	**0.028***
P. paniscus	*C. jacchus*	14	24	78	2.63	0.105
M. fascicularis	*S. sciureus*	14	23	82	2.19	0.139
M. fascicularis	*S. oedipus*	13	23	81	2.78	0.096
M. fascicularis	*C. jacchus*	15	22	80	1.32	0.25
BOULE coding nucleotide						
Species 1	**Species 2**	**m1**	**m2**	**m3**	**chi-square**	**p-value**
Catarrhine vs. Catarrhine						
H. sapiens	*P. paniscus*	0	3	70	3	0.083
H. sapiens	*M. fascicularis*	5	6	65	0.09	0.763
P. paniscus	*M. fascicularis*	6	4	67	0.4	0.527
Platyrrhine vs. Platyrrhine						
S. sciureus	*S. oedipus*	4	12	67	4	**0.046***
S. sciureus	*C. jacchus*	4	10	67	2.57	0.109
S. oedipus	*C. jacchus*	3	1	75	1	0.317
Catarrhine vs. Platyrrhine						
H. sapiens	*S. sciureus*	10	11	59	0.05	0.827
H. sapiens	*S. oedipus*	9	18	60	3	0.083
H. sapiens	*C. jacchus*	8	15	61	2.13	0.144
P. paniscus	*S. sciureus*	11	9	61	0.2	0.655
P. paniscus	*S. oedipus*	10	16	62	1.38	0.239
P. paniscus	*C. jacchus*	9	13	63	0.73	0.394
M. fascicularis	*S. sciureus*	10	10	60	0	1
M. fascicularis	*S. oedipus*	9	17	61	2.46	0.117
M. fascicularis	*C. jacchus*	8	14	62	1.64	0.201

[a] m1, m2, and m3 represent the number of changes unique to Species 1, Species 2, and the outgroup *M. musculus,* respectively. See the original description of the test by Tajima for further discussion of the theory [141].

Table 3 Rate of non-synonymous relative to synonymous change (K_a/K_s ratio) for *DAZL* and *BOULE* across all species examined

DAZL

Strepsirhini	Platyrrhini			Catarrhini			
Microcebus murinus	*Callithrix jacchus*	*Saguinus oedipus*	*Saimiri sciureus*	*Macaca fascicularis*	*Pan paniscus*	*Homo sapiens*	
0.2731							*C. jacchus*
0.2410	0.2212						*S. oedipus*
0.3003	0.3067	0.2473					*S. sciureus*
0.2202	0.2199	0.1698	0.3229				*M. fascicularis*
0.2548	0.2378	0.2002	0.3854	0.2344			*P. paniscus*
0.2685	0.2529	0.2129	0.4209	0.2842	0.5763		*H. sapiens*
0.1335	0.2063	0.1858	0.2070	0.1904	0.2084	0.2136	*M. musculus*

BOULE

Strepsirhini	Platyrrhini			Catarrhini			
Microcebus murinus	*Callithrix jacchus*	*Saguinus oedipus*	*Saimiri sciureus*	*Homo sapiens*	*Pan paniscus*	*Macaca fasciculari s*	
0.1081							*C. jacchus*
0.1179	0.0038						*S. oedipus*
0.1015	0.0413	0.0495					*S. sciureus*
0.1065	0.0620	0.0708	0.0558				*H. sapiens*
0.0964	0.0535	0.0619	0.0473	0.0076			*P. paniscus*
0.0965	0.0536	0.0620	0.0474	0.0353	0.0273		*M. fascicularis*
0.2790	0.2935	0.3073	0.2702	0.2554	0.2683	0.2687	*M. musculus*

Table 4 Intralineage synonymous substitution rates (K_s)

DAZL CODING NUCLEOTIDE					
Catarrhine vs. Catarrhine			Platyrrhine vs. Platyrrhine		
Species 1	**Species 2**	**Ks**	**Species 1**	**Species 2**	**Ks**
H. sapiens	*P. paniscus*	0.0059	*S. sciureus*	*S. oedipus*	0.0655
H. sapiens	*M. fascicularis*	0.0278	*S. sciureus*	*C. jacchus*	0.0639
P. paniscus	*M. fascicularis*	0.0337	*S. oedipus*	*C. jacchus*	0.0434
***BOL* CODING NUCLEOTIDE**					
Catarrhine vs. Catarrhine			Platyrrhine vs. Platyrrhine		
Species 1	**Species 2**	**Ks**	**Species 1**	**Species 2**	**Ks**
H. sapiens	*P. paniscus*	0.0076	*S. sciureus*	*S. oedipus*	0.0495
H. sapiens	*M. fascicularis*	0.0353	*S. sciureus*	*C. jacchus*	0.0413
P. paniscus	*M. fascicularis*	0.0273	*S. oedipus*	*C. jacchus*	0.0038

Fig. 1

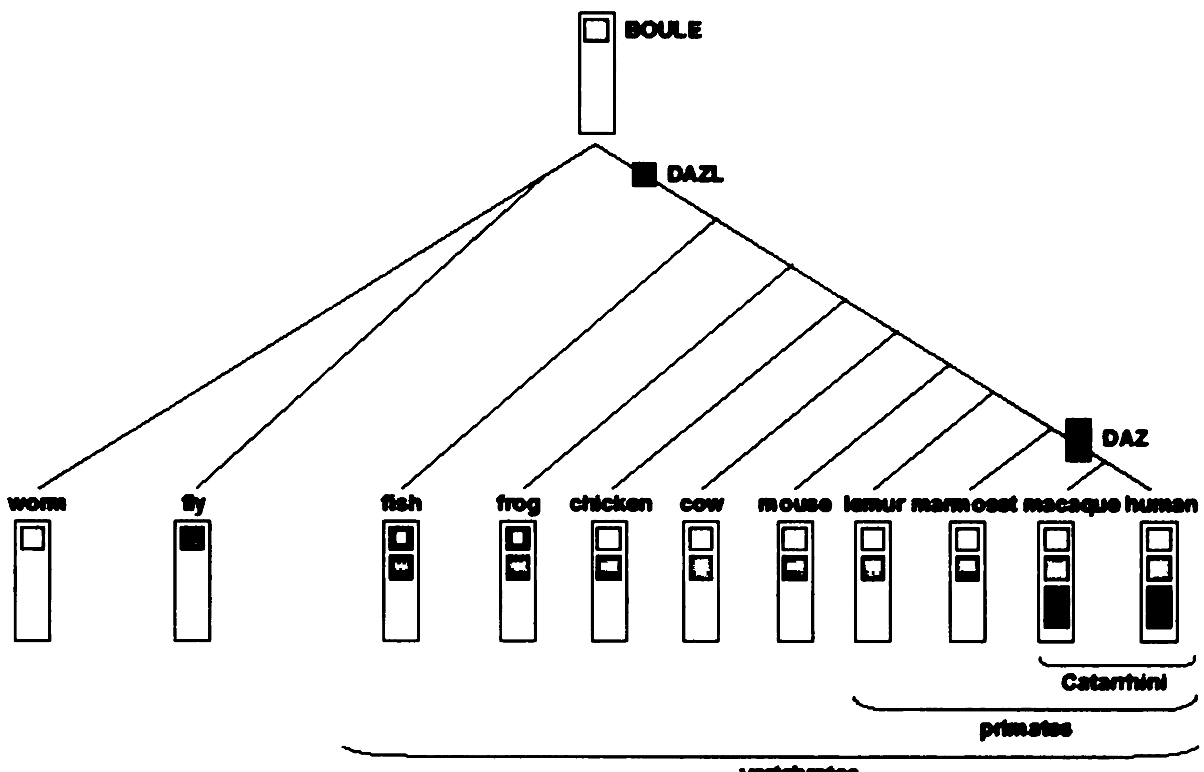


Fig. 2

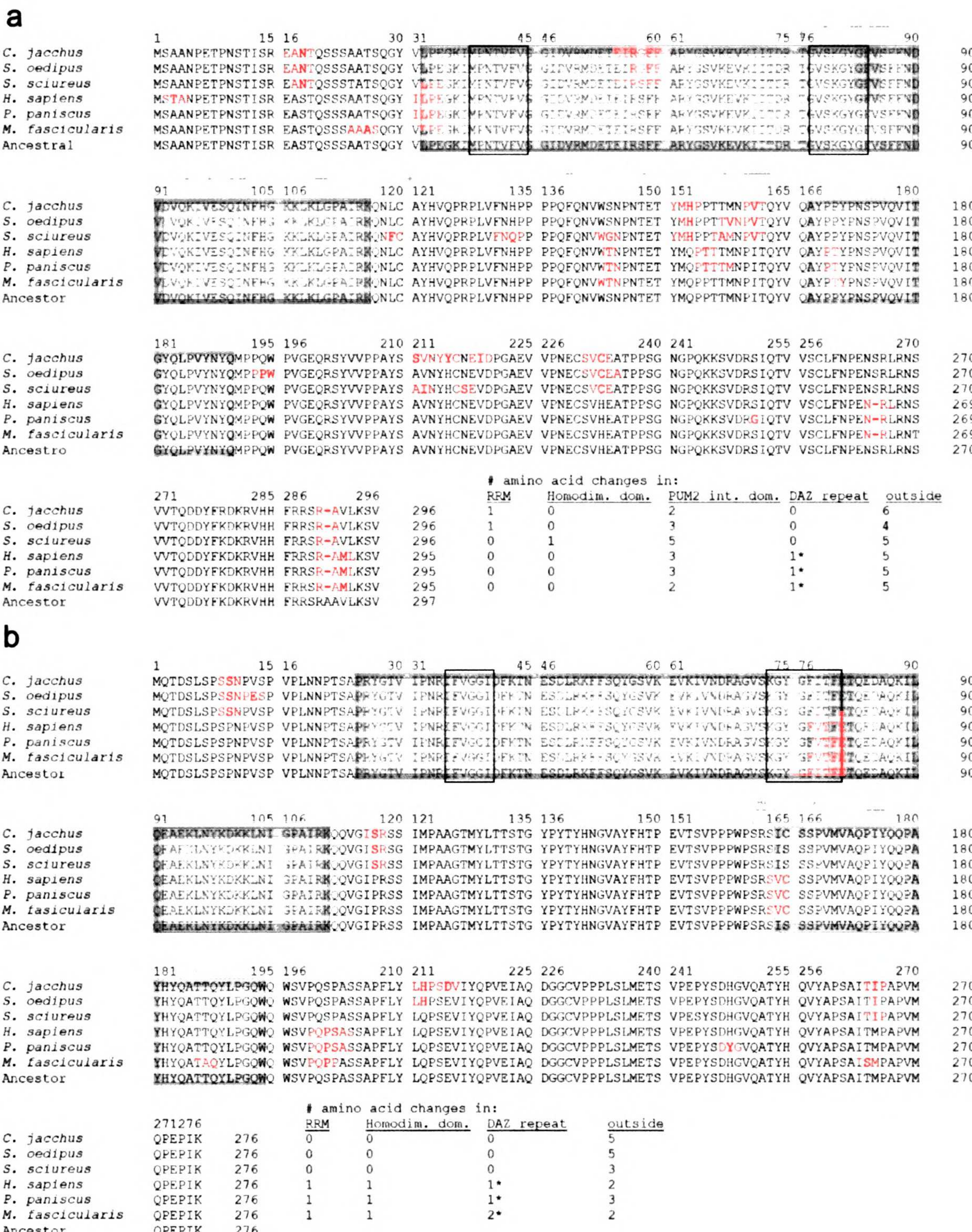
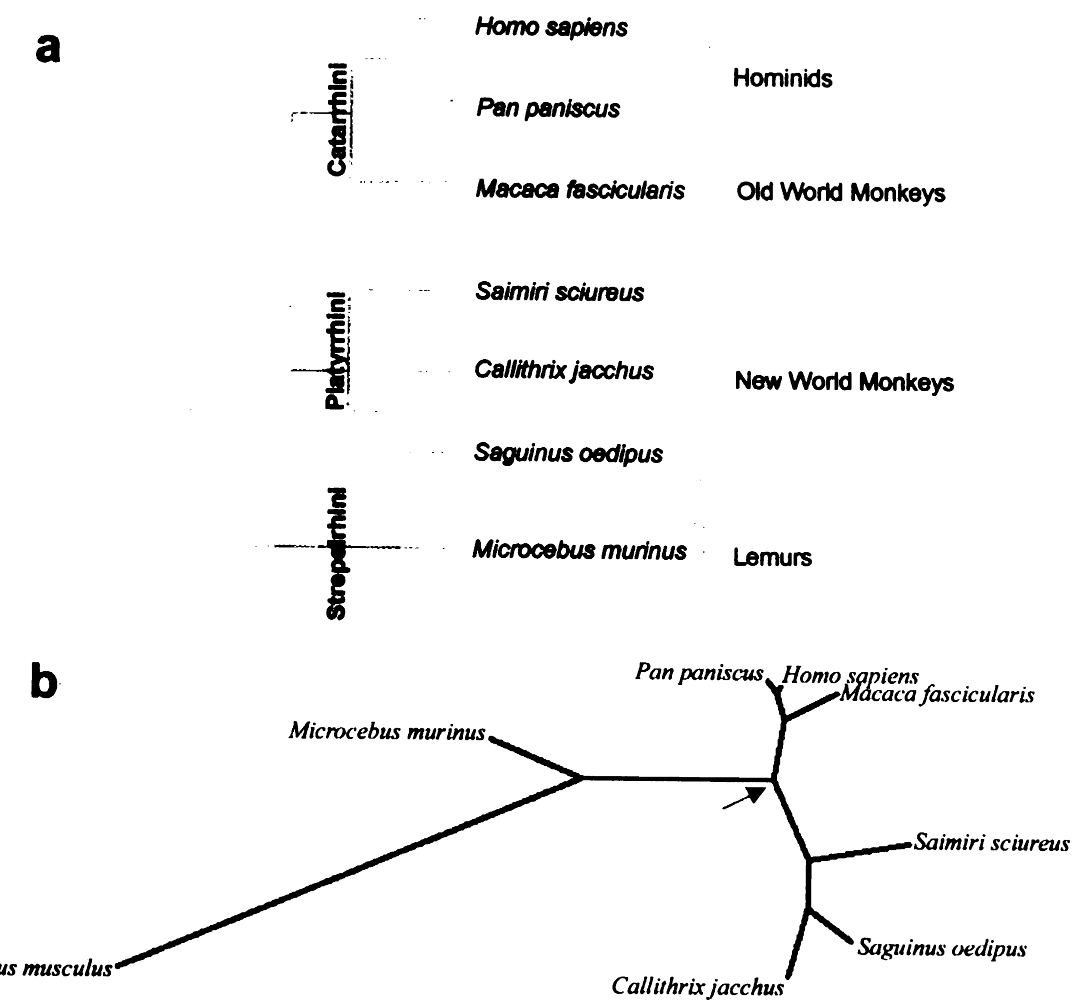


Fig. 3



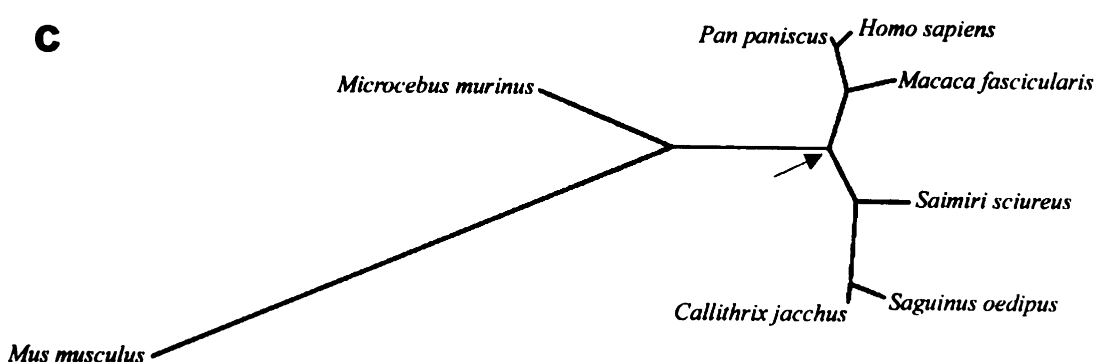


Fig. 4

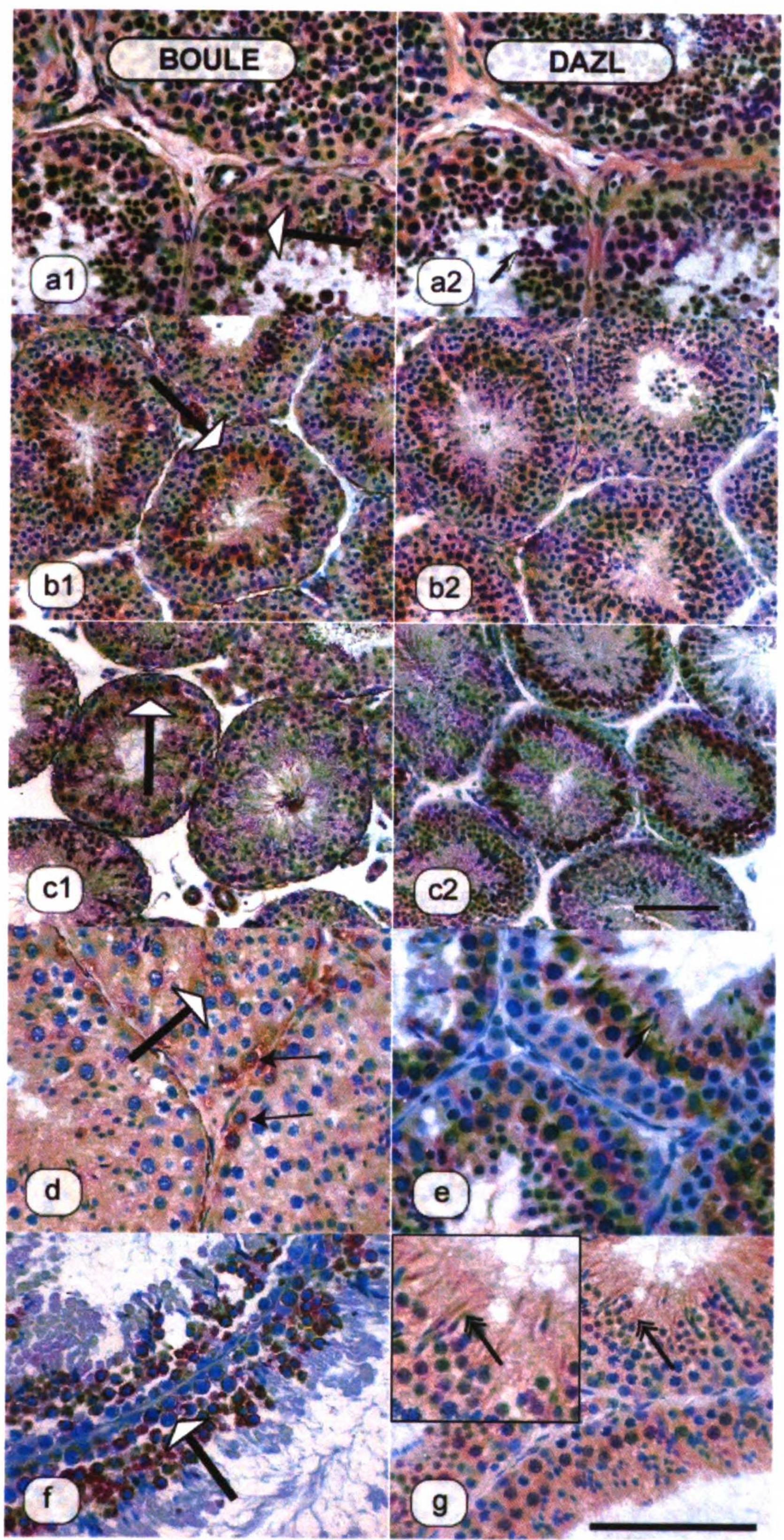
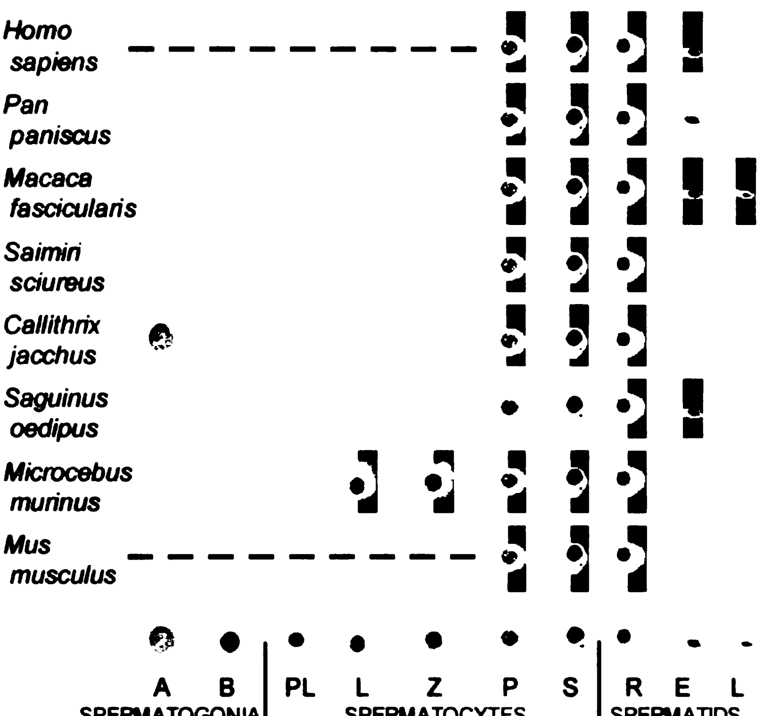


Fig. 5



Chapter 6

Conclusions

The major findings of my thesis can be summarized as follows:

- Single nucleotide polymorphisms (SNPs) in the human *DAZL* gene are associated with two reproductive parameters: age of onset of menopause/ovarian failure and sperm count.
- The associated SNPs have sex-specific interactions with each other and opposite effects in men and women.
- The *Dazl* gene is required prior to e14.5 for germ cell differentiation and maintenance in the mouse in both sexes.
- Mouse *Dazl* has a dose-dependent effect in both males and females.
- The evolution of the *DAZ* family members, *DAZL* and *BOULE*, was not affected by the introduction of *DAZ* into the primate lineage.

I discuss these in more detail in the conclusions sections below.

DAZL is a conserved requirement for germ cell development in mammals that acts in a dose-dependent manner

By examining the function of the *DAZL* gene in both humans and mice, this work has shown that *DAZL* has a conserved function in germ cell development in mammals. Initially, the function of *Dazl* in the mouse appeared to be different than the function of *Xdazl* in frog, as knockdown of *Xdazl* by morpholinos resulted in a PGC migration defect whereas disruption of mouse *Dazl* resulted in a meiotic defect [68, 99]. The authors of the paper on *Xdazl* function suggested that the migration defect they observed was a result of improper early PGC differentiation [99]. Based on the gene expression data we acquired in this study, we suggest that the defect observed in the mouse *Dazl* knockout is also a result

of defective germ cell differentiation, with germ cell loss occurring at developmental checkpoints. The fact that the germ cell loss occurs later in mouse may be due to a number of different factors. First, only the last exon of the RNA Recognition Motif was removed when making the *Dazl* knockout, which may mean it has some residual function [68]. Second, the point at which lack of expression of important germ cell genes triggers germ cell loss may differ between frogs and mice. Third, the genes required for migration may differ between frogs and mice, thus *Dazl* may not be as critical for mouse PGC migration as it is for frog PGC migration. In mice, the *c-kit* gene is required for mouse primordial germ cell migration, but no functional homolog of *c-kit* has been identified in frogs [153].

Until this study, it was hypothesized that human *DAZL* was likely to be required for human germ cell development and fertility, but there was no clear data to support that hypothesis. As mentioned previously, only one study had shown an association between a *DAZL* SNP and male infertility, but this finding could not be replicated in different ethnic groups. In contrast, the work presented here identified a number of *DAZL* SNPs that are associated with measures of fertility in both men and women and across ethnic groups. We also described multiple haplotypes with even stronger associations with sperm count and age at ovarian failure. As sperm count and age at menopause/ovarian failure are indicators of germ cell number, this data strongly suggests that *DAZL* is required for human germ cell development and that defects in *DAZL* result in reduced germ cell counts. It is also clear from this data that patients that were heterozygous for a given SNP also showed a phenotype, usually a phenotype intermediate between the two homozygous genotypes. This is consistent with our data, in which we showed that *Dazl*

has a dose-dependent effect in both sexes. Male and female *Dazl* heterozygotes had fewer germ cells than their wild type littermates, but more germ cells than their mutant littermates. In addition, the original mouse *Dazl* knockout showed that males heterozygous for the *Dazl* knockout allele produced a high percentage of abnormal sperm [68].

DAZL may be involved in a translational cascade

3'UTR regulation is a major theme in germ cell development. Not only are many genes required for germ cell development regulated through their 3'UTRs, but germ cell genes such as *Pumilio* and *Nanos* are known to repress mRNA translation through binding sites in the 3'UTRs of other genes [42, 154-156]. *DAZL* is no exception to this rule. In our human association study, we identified six SNPs located in the 3'UTR that were statistically significantly associated with reproductive parameters and part of regions that were strongly conserved from mouse to human. Most likely, the 3'UTR of *DAZL* is important for the regulation of protein expression and these SNPs affect how much DAZL protein is produced. In turn, several studies have identified a number of mRNA targets for both human and mouse DAZL and shown in both mouse and zebrafish that DAZL is capable of promoting translation through 3'UTR sequence elements [59, 61, 64, 65, 125]. The potential targets include genes such as *Vasa* and *SYCP3*, which we have shown to have reduced expression in mouse *Dazl* mutants [65]. This same study found that *Dazl* mRNA was enriched in a Dazl immunoprecipitation in wild type versus *Dazl* mutant mice, suggesting that Dazl may also regulate its own production. DAZL has been shown to bind to the PUMILIO-2 protein, which, similar to DAZL, is able to bind a key

regulator of translational repression and activation, the cytoplasmic polyadenylation element binding protein [67, 89]. Thus, *DAZL* is likely to be an integral part of a translational cascade required for germ cell development, in which the timing and levels of protein production are regulated at the post-transcriptional level (Fig. 1).

The effect of sex on DAZ family members

Some of the most intriguing aspects of studying *DAZL* function are the differences between the sexes in *DAZ* family gene complements, interactions between human *DAZL* SNPs, phenotype of human *DAZL* SNPs, and gene expression in mouse *Dazl* mutants. In Old World monkeys and hominids, the males possess *BOULE*, *DAZL*, and *DAZ*, while the females possess only *BOULE* and *DAZL*. This imposes interesting evolutionary constraints on *DAZL* that are not imposed on *DAZ*. As *DAZ* is only expressed in males, any sequence changes that promote male fertility will be maintained. *DAZL* is expressed in both males and females; therefore, any sequence changes that promote fertility in one sex cannot be too detrimental to fertility in the other sex. This principle was likely in play when *DAZ* was introduced in the primate lineage. Though *DAZ* is very closely related to *DAZL*, we saw that the introduction of *DAZ* did not significantly relieve the functional constraint on *DAZL* because females must still rely on *DAZL* alone. Similarly, though there are common human *DAZL* alleles that improve sperm production at the expense of oocyte production, the decrease in female fertility is only on the order of a few years, not enough for there to be strong evolutionary pressure against the decrease. This line of reasoning suggests that *DAZL* is conserved as a critical determinant of germ cell development. At the same time, relative to its family member,

BOULE, *DAZL* is highly variable in sequence. Our work in the mouse shows that *DAZL* acts in a dose-dependent fashion; therefore, a decrease in *DAZL* levels leads to a concomitant decrease in the number of germ cells. As mouse *Dazl* heterozygotes are still fertile, most organisms can probably still function with a small decrease in the levels or functionality of the DAZL protein. This means that the *DAZL* gene can tolerate a certain amount of sequence variation over time. The expression of *BOULE*, on the other hand, must be more tightly controlled, leaving no room for much sequence variation to change its function or protein levels.

The sex-specific interactions between and phenotypes of the human DAZL SNPs are also very interesting. Through haplotype analysis, we showed that there were additive interactions between the SNPs in females, but not in males. Surprisingly, we also found that for many of our SNPs, alleles that increased sperm count also decreased the age of onset of menopause. The cellular environment in which *DAZL* is functioning must be different between males and females, allowing SNPs to have different effects. Not only do male germ cells produce male-specific proteins like *DAZ* and *CDY*, but signals from the testes and ovary most likely induce different proteins in their respective germ cells. For instance, the male genital ridge at e12 is postulated to produce a factor that inhibits the entry of male germ cells into meiosis [157].

Still, the fact that so many of the *DAZL* SNPs identified in this study displayed opposite effects in men and women is baffling. Clearly, severe disruption of the gene led to germ cell loss in both sexes in frogs and mice [68, 99]. Why, then, should minor disruptions of the gene lead to opposite effects the two sexes in humans? One possibility is that there are one or two specific SNPs that are the most critical in determining the

phenotype, while the rest of the SNPs we identified are simply in linkage disequilibrium with the critical SNPs. These critical SNPs may interact with a protein complex that promotes *DAZL* function in one sex but cannot interact with a slightly different protein complex in the other sex.

Another possibility is that a slight decrease in the amount of *DAZL* produced, while it may decrease sperm production, may actually increase oocyte production. In our work with the mouse *Dazl* mutant, we observed an increase in the levels of early germ cell markers, but a decrease in the levels of late germ cell markers, in the mutant compared to the wild type. From this we concluded that in the female *Dazl* mutant, the germ cells may be remaining in a more primitive gonocyte state rather than progressing through to meiosis. In fact, if the average number of germ cells at e12.5 and e14.5 in wild type versus mutant mice are compared, we can see that the *Dazl* mutant germ cells actually appear to go through more doublings in that period of time than their wild type counterparts (Table 1). Later on, between e14.5 and e16.5, the failure to differentiate causes the number of mutant germ cells to go down in this time period, probably through apoptosis. In humans, if a SNP results in reduced levels of *DAZL* or impairs its function, the female gonocytes may differentiate more slowly and undergo a few extra rounds of proliferation before they accumulate enough DAZL protein to differentiate and begin meiosis. The male mouse *Dazl* mutant expressed decreased levels of almost all the genes tested and showed a decrease in the amount of proliferation before mitotic arrest (Table 1). Men with hypomorphic levels of *DAZL* would simply have fewer germ cells.

With the advent of embryonic stem cells, recent advances in assisted reproductive techniques, and improved human genetics, the fertility field is now rapidly changing.

Genes from the *DAZ* gene family are at the forefront of research on mammalian germ cell development. From this work, we now know that *DAZL*, not just *DAZ*, is important for human health, and may one day be a diagnostic marker for early menopause or oligozoospermia/azoospermia. We also now know from the mouse that *Dazl* is required for germ cell differentiation and the proper expression of a large number of important germ cell genes. Over the course of this project, I have concluded that *DAZL* has a conserved function as an essential regulator of mammalian germ cell development. Hopefully, future work will define *DAZL*'s precise role in the cascade of germ cell regulation and determine how *DAZL* variation can be used to manage human infertility.

Table 1 Estimated number of doublings in wild type and *Dazl* mutant mouse germ cells

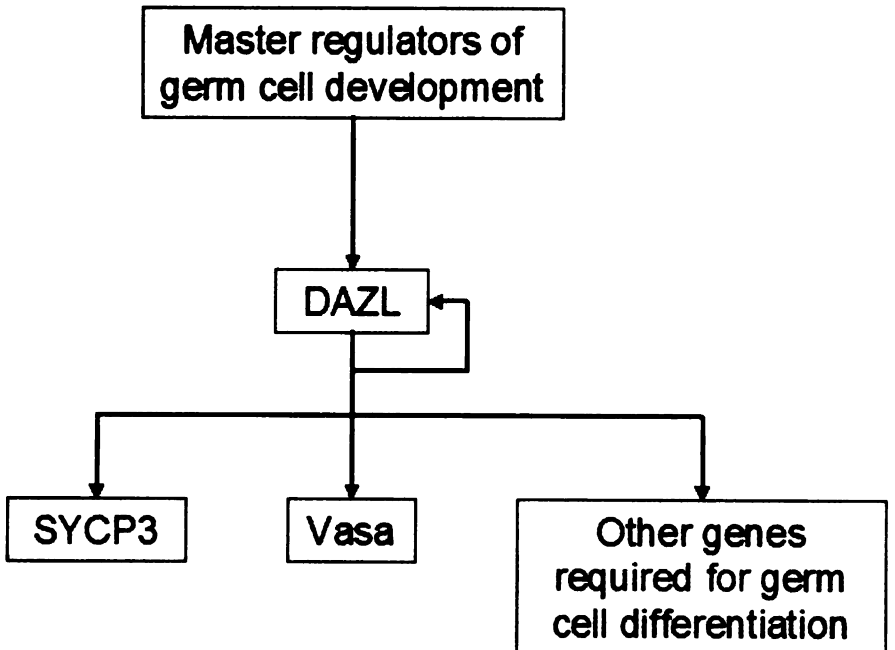
	Males		**Females**	
Time interval	**e12.5-e14.5**	**e14.5-e16.5**	**e12.5-e14.5**	**e14.5-e16.5**
Dazl+/+	2.53	1.71	2.67	0.89
Dazl-/-	1.5	Negative [a]	3.43	Negative [a]

[a] "Negative" indicates that there were fewer germ cells at the end of the time internal than at the beginning

Figure Legends

Fig. 1 Model for translational cascade involving *DAZL*. *DAZL* is involved in a cascade of translational regulation that is required for proper germ cell development. One or more master regulators, as yet unknown, control DAZL protein expression, most likely through its 3'UTR. In turn, DAZL controls the production of other proteins required for germ cell development and differentiation, including Vasa and SYCP3, through the 3'UTRs of their mRNAs.

Fig. 1



References

1 Matzuk, M. M. and Lamb, D. J. Genetic dissection of mammalian fertility pathways. *Nat Cell Biol* **4 Suppl**, s41-49 (2002).

2 Starz-Gaiano, M. and Lehmann, R. Moving towards the next generation. *Mech Dev* **105**, 5-18. (2001).

3 Wylie, C. Germ cells. *Cell* **96**, 165-174. (1999).

4 Lawson, K. A., Dunn, N. R., Roelen, B. A., Zeinstra, L. M., Davis, A. M., Wright, C. V., Korving, J. P. and Hogan, B. L. Bmp4 is required for the generation of primordial germ cells in the mouse embryo. *Genes Dev* **13**, 424-436. (1999).

5 Cramer, D. W., Xu, H. and Harlow, B. L. Family history as a predictor of early menopause. *Fertil Steril* **64**, 740-745 (1995).

6 Torgerson, D. J., Thomas, R. E. and Reid, D. M. Mothers and daughters menopausal ages: is there a link? *Eur J Obstet Gynecol Reprod Biol* **74**, 63-66 (1997).

7 Snieder, H., MacGregor, A. J. and Spector, T. D. Genes control the cessation of a woman's reproductive life: a twin study of hysterectomy and age at menopause. *J Clin Endocrinol Metab* **83**, 1875-1880 (1998).

8 de Bruin, J. P., Bovenhuis, H., van Noord, P. A., Pearson, P. L., van Arendonk, J. A., te Velde, E. R., Kuurman, W. W. and Dorland, M. The role of genetic factors in age at natural menopause. *Hum Reprod* **16**, 2014-2018 (2001).

9 Vegetti, W., Grazia Tibiletti, M., Testa, G., de Lauretis, Y., Alagna, F., Castoldi, E., Taborelli, M., Motta, T., Bolis, P. F., Dalpra, L. and Crosignani, P. G.

Inheritance in idiopathic premature ovarian failure: analysis of 71 cases. *Hum Reprod* **13**, 1796-1800. (1998).

10 Crisponi, L., Deiana, M., Loi, A., Chiappe, F., Uda, M., Amati, P., Bisceglia, L., Zelante, L., Nagaraja, R., Porcu, S., Ristaldi, M. S., Marzella, R., Rocchi, M., Nicolino, M., Lienhardt-Roussie, A., Nivelon, A., Verloes, A., Schlessinger, D., Gasparini, P., Bonneau, D., Cao, A. and Pilia, G. The putative forkhead transcription factor *FOXL2* is mutated in blepharophimosis/ptosis/epicanthus inversus syndrome. *Nat Genet* **27**, 159-166. (2001).

11 De Baere, E., Dixon, M. J., Small, K. W., Jabs, E. W., Leroy, B. P., Devriendt, K., Gillerot, Y., Mortier, G., Meire, F., Van Maldergem, L., Courtens, W., Hjalgrim, H., Huang, S., Liebaers, I., Van Regemorter, N., Touraine, P., Praphanphoj, V., Verloes, A., Udar, N., Yellore, V., Chalukya, M., Yelchits, S., De Paepe, A., Kuttenn, F., Fellous, M., Veitia, R. and Messiaen, L. Spectrum of *FOXL2* gene mutations in blepharophimosis-ptosis-epicanthus inversus (BPES) families demonstrates a genotype--phenotype correlation. *Hum Mol Genet* **10**, 1591-1600. (2001).

12 Gersak, K., Harris, S. E., Smale, W. J. and Shelling, A. N. A novel 30 bp deletion in the *FOXL2* gene in a phenotypically normal woman with primary amenorrhoea: case report. *Hum Reprod* **19**, 2767-2770 (2004).

13 Harris, S. E., Chand, A. L., Winship, I. M., Gersak, K., Aittomaki, K. and Shelling, A. N. Identification of novel mutations in *FOXL2* associated with premature ovarian failure. *Mol Hum Reprod* **8**, 729-733 (2002).

14 Fogli, A., Gauthier-Barichard, F., Schiffmann, R., Vanderhoof, V. H., Bakalov, V. K., Nelson, L. M. and Boespflug-Tanguy, O. Screening for known mutations in *EIF2B* genes in a large panel of patients with premature ovarian failure. *BMC Womens Health* **4**, 8 (2004).

15 Fogli, A., Rodriguez, D., Eymard-Pierre, E., Bouhour, F., Labauge, P., Meaney, B. F., Zeesman, S., Kaneski, C. R., Schiffmann, R. and Boespflug-Tanguy, O. Ovarian failure related to eukaryotic initiation factor 2B mutations. *Am J Hum Genet* **72**, 1544-1550 (2003).

16 Aittomaki, K., Lucena, J. L., Pakarinen, P., Sistonen, P., Tapanainen, J., Gromoll, J., Kaskikari, R., Sankila, E. M., Lehvaslaiho, H., Engel, A. R. and et al. Mutation in the follicle-stimulating hormone receptor gene causes hereditary hypergonadotropic ovarian failure. *Cell* **82**, 959-968 (1995).

17 Kok, H. S., van Asselt, K. M., Peeters, P. H., van der Schouw, Y. T., Grobbee, D. E., Pearson, P. L. and Wijmenga, C. Age at natural menopause is not linked with the follicle-stimulating hormone receptor region: a sib-pair study. *Fertil Steril* **81**, 611-616 (2004).

18 Sundblad, V., Chiauzzi, V. A., Escobar, M. E., Dain, L. and Charreau, E. H. Screening of FSH receptor gene in Argentine women with premature ovarian failure (POF). *Mol Cell Endocrinol* **222**, 53-59 (2004).

19 Conway, G. S., Conway, E., Walker, C., Hoppner, W., Gromoll, J. and Simoni, M. Mutation screening and isoform prevalence of the follicle stimulating hormone receptor gene in women with premature ovarian failure, resistant ovary syndrome and polycystic ovary syndrome. *Clin Endocrinol (Oxf)* **51**, 97-99 (1999).

20 Whitney, E. A., Layman, L. C., Chan, P. J., Lee, A., Peak, D. B. and McDonough, P. G. The follicle-stimulating hormone receptor gene is polymorphic in premature ovarian failure and normal controls. *Fertil Steril* **64**, 518-524 (1995).

21 Marozzi, A., Porta, C., Vegetti, W., Crosignani, P. G., Tibiletti, M. G., Dalpra, L. and Ginelli, E. Mutation analysis of the inhibin alpha gene in a cohort of Italian women affected by ovarian failure. *Hum Reprod* **17**, 1741-1745 (2002).

22 Jeong, H. J., Cho, S. W., Kim, H. A., Lee, S. H., Cho, J. H., Choi, D. H., Kwon, H., Cha, W. T., Han, J. E. and Cha, K. Y. G769A variation of inhibin alpha-gene in korean women with premature ovarian failure. *Yonsei Med J* **45**, 479-482 (2004).

23 Dixit, H., Deendayal, M. and Singh, L. Mutational analysis of the mature peptide region of inhibin genes in Indian women with ovarian failure. *Hum Reprod* **19**, 1760-1764 (2004).

24 Allingham-Hawkins, D. J., Babul-Hirji, R., Chitayat, D., Holden, J. J., Yang, K. T., Lee, C., Hudson, R., Gorwill, H., Nolin, S. L., Glicksman, A., Jenkins, E. C., Brown, W. T., Howard-Peebles, P. N., Becchi, C., Cummings, E., Fallon, L., Seitz, S., Black, S. H., Vianna-Morgante, A. M., Costa, S. S., Otto, P. A., Mingroni-Netto, R. C., Murray, A., Webb, J., Vieri, F. and et al. Fragile X premutation is a significant risk factor for premature ovarian failure: the International Collaborative POF in Fragile X study--preliminary data. *Am J Med Genet* **83**, 322-325 (1999).

25 Conway, G. S., Payne, N. N., Webb, J., Murray, A. and Jacobs, P. A. Fragile X premutation screening in women with premature ovarian failure. *Hum Reprod* **13**, 1184-1187 (1998).

26 Davison, R. M., Davis, C. J. and Conway, G. S. The X chromosome and ovarian failure. *Clin Endocrinol (Oxf)* **51**, 673-679. (1999).

27 Reijo, R., Lee, T. Y., Salo, P., Alagappan, R., Brown, L. G., Rosenberg, M., Rozen, S., Jaffe, T., Straus, D., Hovatta, O. and et al. Diverse spermatogenic defects in humans caused by Y chromosome deletions encompassing a novel RNA-binding protein gene. *Nat Genet* **10**, 383-393. (1995).

28 Reijo, R., Alagappan, R. K., Patrizio, P. and Page, D. C. Severe oligozoospermia resulting from deletions of azoospermia factor gene on Y chromosome. *Lancet* **347**, 1290-1293. (1996).

29 Hargreave, T. B. Genetic basis of male fertility. *Br Med Bull* **56**, 650-671 (2000).

30 Miyamoto, T., Hasuike, S., Yogev, L., Maduro, M. R., Ishikawa, M., Westphal, H. and Lamb, D. J. Azoospermia in patients heterozygous for a mutation in *SYCP3*. *Lancet* **362**, 1714-1719 (2003).

31 Tanaka, H., Miyagawa, Y., Tsujimura, A., Matsumiya, K., Okuyama, A. and Nishimune, Y. Single nucleotide polymorphisms in the *protamine-1* and *-2* genes of fertile and infertile human male populations. *Mol Hum Reprod* **9**, 69-73 (2003).

32 Dowsing, A. T., Yong, E. L., Clark, M., McLachlan, R. I., de Kretser, D. M. and Trounson, A. O. Linkage between male infertility and trinucleotide repeat expansion in the androgen-receptor gene. *Lancet* **354**, 640-643 (1999).

33 Lund, A., Tapanainen, J. S., Lahdetie, J., Savontaus, M. L. and Aittomaki, K. Long CAG repeats in the AR gene are not associated with infertility in Finnish males. *Acta Obstet Gynecol Scand* **82**, 162-166 (2003).

34 Asatiani, K., von Eckardstein, S., Simoni, M., Gromoll, J. and Nieschlag, E. CAG repeat length in the androgen receptor gene affects the risk of male infertility. *Int J Androl* **26**, 255-261 (2003).

35 Ferlin, A., Bartoloni, L., Rizzo, G., Roverato, A., Garolla, A. and Foresta, C. Androgen receptor gene CAG and GGC repeat lengths in idiopathic male infertility. *Mol Hum Reprod* **10**, 417-421 (2004).

36 Ruhayel, Y., Lundin, K., Giwercman, Y., Hallden, C., Willen, M. and Giwercman, A. Androgen receptor gene GGN and CAG polymorphisms among severely oligozoospermic and azoospermic Swedish men. *Hum Reprod* **19**, 2076-2083 (2004).

37 Schupbach, T. and Wieschaus, E. Germline autonomy of maternal-effect mutations altering the embryonic body pattern of *Drosophila*. *Dev Biol* **113**, 443-448. (1986).

38 Styhler, S., Nakamura, A., Swan, A., Suter, B. and Lasko, P. *vasa* is required for GURKEN accumulation in the oocyte, and is involved in oocyte differentiation and germline cyst development. *Development* **125**, 1569-1578. (1998).

39 Kuznicki, K. A., Smith, P. A., Leung-Chiu, W. M., Estevez, A. O., Scott, H. C. and Bennett, K. L. Combinatorial RNA interference indicates *GLH-4* can compensate for *GLH- 1*; these two P granule components are critical for fertility in *C. elegans*. *Development* **127**, 2907-2916. (2000).

40 Tanaka, S. S., Toyooka, Y., Akasu, R., Katoh-Fukui, Y., Nakahara, Y., Suzuki, R., Yokoyama, M. and Noce, T. The mouse homolog of *Drosophila Vasa* is required for the development of male germ cells. *Genes Dev* **14**, 841-853. (2000).

41 Gruidl, M. E., Smith, P. A., Kuznicki, K. A., McCrone, J. S., Kirchner, J., Roussell, D. L., Strome, S. and Bennett, K. L. Multiple potential germ-line helicases are components of the germ-line- specific P granules of *Caenorhabditis elegans*. *Proc Natl Acad Sci U S A* **93**, 13837-13842. (1996).

42 Asaoka-Taguchi, M., Yamada, M., Nakamura, A., Hanyu, K. and Kobayashi, S. Maternal *Pumilio* acts together with *Nanos* in germline development in *Drosophila* embryos. *Nat Cell Biol* **1**, 431-437. (1999).

43 Deshpande, G., Calhoun, G., Yanowitz, J. L. and Schedl, P. D. Novel functions of *nanos* in downregulating mitosis and transcription during the development of the *Drosophila* germline. *Cell* **99**, 271-281. (1999).

44 Kobayashi, S., Yamada, M., Asaoka, M. and Kitamura, T. Essential role of the posterior morphogen *nanos* for germline development in *Drosophila*. *Nature* **380**, 708-711. (1996).

45 Wreden, C., Verrotti, A. C., Schisa, J. A., Lieberfarb, M. E. and Strickland, S. *Nanos* and *pumilio* establish embryonic polarity in *Drosophila* by promoting posterior deadenylation of *hunchback* mRNA. *Development* **124**, 3015-3023. (1997).

46 Forbes, A. and Lehmann, R. *Nanos* and *Pumilio* have critical roles in the development and function of *Drosophila* germline stem cells. *Development* **125**, 679-690. (1998).

47 Wickens, M., Bernstein, D. S., Kimble, J. and Parker, R. A PUF family portrait: 3'UTR regulation as a way of life. *Trends Genet* **18**, 150-157. (2002).

48 Subramaniam, K. and Seydoux, G. *nos-1* and *nos-2*, two genes related to *Drosophila nanos*, regulate primordial germ cell development and survival in *Caenorhabditis elegans*. *Development* **126**, 4861-4871. (1999).

49 Kraemer, B., Crittenden, S., Gallegos, M., Moulder, G., Barstead, R., Kimble, J. and Wickens, M. NANOS-3 and FBF proteins physically interact to control the sperm- oocyte switch in *Caenorhabditis elegans*. *Curr Biol* **9**, 1009-1018. (1999).

50 Tsuda, M., Sasaoka, Y., Kiso, M., Abe, K., Haraguchi, S., Kobayashi, S. and Saga, Y. Conserved role of nanos proteins in germ cell development. *Science* **301**, 1239-1241 (2003).

51 Haraguchi, S., Tsuda, M., Kitajima, S., Sasaoka, Y., Nomura-Kitabayashid, A., Kurokawa, K. and Saga, Y. *nanos1*: a mouse *nanos* gene expressed in the central nervous system is dispensable for normal development. *Mech Dev* **120**, 721-731 (2003).

52 Jaruzelska, J., Kotecki, M., Kusz, K., Spik, A., Firpo, M. and Reijo Pera, R. A. Conservation of a Pumilio-Nanos complex from *Drosophila* germ plasm to human germ cells. *Dev Genes Evol* **213**, 120-126 (2003).

53 Koprunner, M., Thisse, C., Thisse, B. and Raz, E. A zebrafish *nanos*-related gene is essential for the development of primordial germ cells. *Genes Dev* **15**, 2877-2885. (2001).

54 Mosquera, L., Forristall, C., Zhou, Y. and King, M. L. A mRNA localized to the vegetal cortex of *Xenopus* oocytes encodes a protein with a *nanos*-like zinc finger domain. *Development* **117**, 377-386. (1993).

55 Cooke, H. J., Lee, M., Kerr, S. and Ruggiu, M. A murine homologue of the human *DAZ* gene is autosomal and expressed only in male and female gonads. *Hum Mol Genet* **5**, 513-516 (1996).

56 Yen, P. H., Chai, N. N. and Salido, E. C. The human autosomal gene *DAZLA*: testis specificity and a candidate for male infertility. *Hum Mol Genet* **5**, 2013-2017. (1996).

57 Saxena, R., Brown, L. G., Hawkins, T., Alagappan, R. K., Skaletsky, H., Reeve, M. P., Reijo, R., Rozen, S., Dinulos, M. B., Disteche, C. M. and Page, D. C. The *DAZ* gene cluster on the human Y chromosome arose from an autosomal gene that was transposed, repeatedly amplified and pruned. *Nat Genet* **14**, 292-299 (1996).

58 Xu, E. Y., Moore, F. L. and Pera, R. A. From the Cover: A gene family required for human germ cell development evolved from an ancient meiotic gene conserved in metazoans. *Proc Natl Acad Sci U S A* **98**, 7414-7419. (2001).

59 Fox, M., Urano, J. and Reijo Pera, R. A. Identification and characterization of RNA sequences to which human PUMILIO-2 (PUM2) and deleted in Azoospermia-like (DAZL) bind. *Genomics* **85**, 92-105 (2005).

60 Houston, D. W., Zhang, J., Maines, J. Z., Wasserman, S. A. and King, M. L. A *Xenopus DAZ-like* gene encodes an RNA component of germ plasm and is a functional homologue of *Drosophila boule*. *Development* **125**, 171-180 (1998).

61 Jiao, X., Trifillis, P. and Kiledjian, M. Identification of Target Messenger RNA Substrates for the Murine Deleted in Azoospermia-Like RNA-Binding Protein. *Biol Reprod* **66**, 475-485. (2002).

62 Tsui, S., Dai, T., Warren, S. T., Salido, E. C. and Yen, P. H. Association of the mouse infertility factor DAZL1 with actively translating polyribosomes. *Biol Reprod* **62**, 1655-1660 (2000).

63 Venables, J. P., Ruggiu, M. and Cooke, H. J. The RNA-binding specificity of the mouse Dazl protein. *Nucleic Acids Res* **29**, 2479-2483. (2001).

64 Collier, B., Gorgoni, B., Loveridge, C., Cooke, H. J. and Gray, N. K. The DAZL family proteins are PABP-binding proteins that regulate translation in germ cells. *Embo J* (2005).

65 Reynolds, N., Collier, B., Maratou, K., Bingham, V., Speed, R. M., Taggart, M., Semple, C. A., Gray, N. K. and Cooke, H. J. Dazl binds in vivo to specific transcripts and can regulate the pre-meiotic translation of Mvh in germ cells. *Hum Mol Genet* (2005).

66 Ruggiu, M. and Cooke, H. J. In vivo and in vitro analysis of homodimerisation activity of the mouse Dazl1 protein. *Gene* **252**, 119-126 (2000).

67 Moore, F. L., Jaruzelska, J., Fox, M. S., Urano, J., Firpo, M. T., Turek, P. J., Dorfman, D. M. and Pera, R. A. Human Pumilio-2 is expressed in embryonic stem cells and germ cells and interacts with DAZ (Deleted in AZoospermia) and DAZ-like proteins. *Proc Natl Acad Sci U S A* **100**, 538-543 (2003).

68 Ruggiu, M., Speed, R., Taggart, M., McKay, S. J., Kilanowski, F., Saunders, P., Dorin, J. and Cooke, H. J. The mouse *Dazla* gene encodes a cytoplasmic protein essential for gametogenesis. *Nature* **389**, 73-77 (1997).

69 Saunders, P. T., Turner, J. M., Ruggiu, M., Taggart, M., Burgoyne, P. S., Elliott, D. and Cooke, H. J. Absence of *mDazl* produces a final block on germ cell development at meiosis. *Reproduction* **126**, 589-597 (2003).

70 Teng, Y. N., Lin, Y. M., Lin, Y. H., Tsao, S. Y., Hsu, C. C., Lin, S. J., Tsai, W. C. and Kuo, P. L. Association of a single-nucleotide polymorphism of the *deleted-in-azoospermia-like* gene with susceptibility to spermatogenic failure. *J Clin Endocrinol Metab* **87**, 5258-5264 (2002).

71 Bartoloni, L., Cazzadore, C., Ferlin, A., Garolla, A. and Foresta, C. Lack of the T54A polymorphism of the *DAZL* gene in infertile Italian patients. *Mol Hum Reprod* **10**, 613-615 (2004).

72 Tschanter, P., Kostova, E., Luetjens, C. M., Cooper, T. G., Nieschlag, E. and Gromoll, J. No association of the A260G and A386G *DAZL* single nucleotide polymorphisms with male infertility in a Caucasian population. *Hum Reprod* **19**, 2771-2776 (2004).

73 Yang, X. J., Shinka, T., Nozawa, S., Yan, H. T., Yoshiike, M., Umeno, M., Sato, Y., Chen, G., Iwamoto, T. and Nakahori, Y. Survey of the two polymorphisms in *DAZL*, an autosomal candidate for the azoospermic factor, in Japanese infertile men and implications for male infertility. *Mol Hum Reprod* **11**, 513-515 (2005).

74 Agulnik, A. I., Zharkikh, A., Boettger-Tong, H., Bourgeron, T., McElreavey, K. and Bishop, C. E. Evolution of the *DAZ* gene family suggests that Y-linked *DAZ*

plays little, or a limited, role in spermatogenesis but underlines a recent African origin for human populations. *Hum Mol Genet* **7**, 1371-1377. (1998).

75 Bielawski, J. P. and Yang, Z. Positive and negative selection in the *DAZ* gene family. *Mol Biol Evol* **18**, 523-529. (2001).

76 Castrillon, D. H., Quade, B. J., Wang, T. Y., Quigley, C. and Crum, C. P. The human *VASA* gene is specifically expressed in the germ cell lineage. *Proc Natl Acad Sci U S A* **97**, 9585-9590. (2000).

77 Fujiwara, Y., Komiya, T., Kawabata, H., Sato, M., Fujimoto, H., Furusawa, M. and Noce, T. Isolation of a DEAD-family protein gene that encodes a murine homolog of *Drosophila vasa* and its specific expression in germ cell lineage. *Proc Natl Acad Sci U S A* **91**, 12258-12262. (1994).

78 Komiya, T., Itoh, K., Ikenishi, K. and Furusawa, M. Isolation and characterization of a novel gene of the DEAD box protein family which is specifically expressed in germ cells of *Xenopus laevis*. *Dev Biol* **162**, 354-363. (1994).

79 Ikenishi, K. and Tanaka, T. S. Involvement of the protein of *Xenopus* vasa homolog (*Xenopus vasa-like gene 1*, *XVLG1*) in the differentiation of primordial germ cells. *Dev Growth Differ* **39**, 625-633. (1997).

80 Yoon, C., Kawakami, K. and Hopkins, N. Zebrafish *vasa* homologue RNA is localized to the cleavage planes of 2- and 4-cell-stage embryos and is expressed in the primordial germ cells. *Development* **124**, 3157-3165. (1997).

81 Tsunekawa, N., Naito, M., Sakai, Y., Nishida, T. and Noce, T. Isolation of chicken *vasa* homolog gene and tracing the origin of primordial germ cells. *Development* **127**, 2741-2750. (2000).

82 Hay, B., Ackerman, L., Barbel, S., Jan, L. Y. and Jan, Y. N. Identification of a component of *Drosophila* polar granules. *Development* **103**, 625-640. (1988).

83 Hay, B., Jan, L. Y. and Jan, Y. N. A protein component of *Drosophila* polar granules is encoded by *vasa* and has extensive sequence similarity to ATP-dependent helicases. *Cell* **55**, 577-587. (1988).

84 Markussen, F. H., Michon, A. M., Breitwieser, W. and Ephrussi, A. Translational control of *oskar* generates short OSK, the isoform that induces pole plasma assembly. *Development* **121**, 3723-3732. (1995).

85 Gavis, E. R., Lunsford, L., Bergsten, S. E. and Lehmann, R. A conserved 90 nucleotide element mediates translational repression of *nanos* RNA. *Development* **122**, 2791-2800. (1996).

86 Shibata, N., Umesono, Y., Orii, H., Sakurai, T., Watanabe, K. and Agata, K. Expression of *vasa*(*vas*)-related genes in germline cells and totipotent somatic stem cells of planarians. *Dev Biol* **206**, 73-87. (1999).

87 Fujimura, M. and Takamura, K. Characterization of an ascidian DEAD-box gene, *Ci-DEAD1*: specific expression in the germ cells and its mRNA localization in the posterior- most blastomeres in early embryos. *Dev Genes Evol* **210**, 64-72. (2000).

88 White, E. K., Moore-Jarrett, T. and Ruley, H. E. PUM2, a novel murine puf protein, and its consensus RNA-binding site. *Rna* **7**, 1855-1866. (2001).

89 Nakahata, S., Katsu, Y., Mita, K., Inoue, K., Nagahama, Y. and Yamashita, M. Biochemical identification of *Xenopus* Pumilio as a sequence-specific cyclin B1 mRNA-binding protein that physically interacts with a Nanos homolog, Xcat-2,

and a cytoplasmic polyadenylation element-binding protein. *J Biol Chem* **276**, 20945-20953. (2001).

90 Macdonald, P. M. The *Drosophila pumilio* gene: an unusually long transcription unit and an unusual protein. *Development* **114**, 221-232. (1992).

91 Lehmann, R. and Nusslein-Volhard, C. Involvement of the *pumilio* gene in the transport of an abdominal signal in the *Drosophila* embryo. *Nature* **329**, 167-170 (1987).

92 Parisi, M. and Lin, H. The *Drosophila pumilio* gene encodes two functional protein isoforms that play multiple roles in germline development, gonadogenesis, oogenesis and embryogenesis. *Genetics* **153**, 235-250. (1999).

93 Zhang, B., Gallegos, M., Puoti, A., Durkin, E., Fields, S., Kimble, J. and Wickens, M. P. A conserved RNA-binding protein that regulates sexual fates in the *C. elegans* hermaphrodite germ line. *Nature* **390**, 477-484. (1997).

94 Subramaniam, K. and Seydoux, G. (International *C. elegans* Meeting, 2001).

95 Wickens, M., Guenthner, B., Kimble, J. and Bernstein, D. (Midwest Worm Meeting, 2000).

96 Luitjens, C., Gallegos, M., Kraemer, B., Kimble, J. and Wickens, M. CPEB proteins control two key steps in spermatogenesis in *C. elegans*. *Genes Dev* **14**, 2596-2609. (2000).

97 Nusslein-Volhard, C., Frohnhofer, H. G. and Lehmann, R. Determination of anteroposterior polarity in *Drosophila*. *Science* **238**, 1675-1681. (1987).

98 Bhat, K. M. The posterior determinant gene *nanos* is required for the maintenance of the adult germline stem cells during *Drosophila* oogenesis. *Genetics* **151**, 1479-1492. (1999).

99 Houston, D. W. and King, M. L. A critical role for *Xdazl*, a germ plasm-localized RNA, in the differentiation of primordial germ cells in *Xenopus*. *Development* **127**, 447-456 (2000).

100 Maegawa, S., Yasuda, K. and Inoue, K. Maternal mRNA localization of zebrafish *DAZ-like* gene. *Mech Dev* **81**, 223-226. (1999).

101 Johnson, A. D., Bachvarova, R. F., Drum, M. and Masi, T. Expression of axolotl *dazl* rna, a marker of germ plasm: widespread maternal rna and onset of expression in germ cells approaching the gonad. *Dev Biol* **234**, 402-415. (2001).

102 Eberhart, C. G., Maines, J. Z. and Wasserman, S. A. Meiotic cell cycle requirement for a fly homologue of human *Deleted in Azoospermia*. *Nature* **381**, 783-785. (1996).

103 Maines, J. Z. and Wasserman, S. A. Post-transcriptional regulation of the meiotic Cdc25 protein Twine by the *Dazl* orthologue *Boule*. *Nat Cell Biol* **1**, 171-174. (1999).

104 Karashima, T., Sugimoto, A. and Yamamoto, M. *Caenorhabditis elegans* homologue of the human azoospermia factor *DAZ* is required for oogenesis but not for spermatogenesis. *Development* **127**, 1069-1079. (2000).

105 Larsen, W. J. *Human embryology* (Churchill Livingstone, New York, 1997).

106 Richardson, S. J., Senikas, V. and Nelson, J. F. Follicular depletion during the menopausal transition: evidence for accelerated loss and ultimate exhaustion. *J Clin Endocrinol Metab* **65**, 1231-1237. (1987).

107 World Health Organization. *WHO laboratory manual for the examination of manual for the examination of human semen and sperm-cervical mucus interaction, 4th edition* (Cambridge University Press, 1999).

108 Xu, E. Y., Lee, D. F., Klebes, A., Turek, P. J., Kornberg, T. B. and Reijo Pera, R. A. Human *BOULE* gene rescues meiotic defects in infertile flies. *Hum Mol Genet* **12**, 169-175 (2003).

109 Kenneson, A., Cramer, D. W. and Warren, S. T. Fragile X premutations are not a major cause of early menopause. *Am J Hum Genet* **61**, 1362-1369 (1997).

110 Shibanuma, K., Tong, Z. B., Vanderhoof, V. H., Vanevski, K. and Nelson, L. M. Investigation of *KIT* gene mutations in women with 46,XX spontaneous premature ovarian failure. *BMC Womens Health* **2**, 8 (2002).

111 Rozen, S. and Skaletsky, H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* **132**, 365-386 (2000).

112 Barrett, J. C., Fry, B., Maller, J. and Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263-265 (2005).

113 Stephens, M., Smith, N. J. and Donnelly, P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* **68**, 978-989 (2001).

114 Stephens, M. and Donnelly, P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet* **73**, 1162-1169 (2003).

115 Nelson, L. M., Covington, S. N. and Rebar, R. W. An update: spontaneous premature ovarian failure is not an early menopause. *Fertil Steril* **83**, 1327-1332 (2005).

116 Bione, S., Sala, C., Manzini, C., Arrigo, G., Zuffardi, O., Banfi, S., Borsani, G., Jonveaux, P., Philippe, C., Zuccotti, M., Ballabio, A. and Toniolo, D. A human homologue of the *Drosophila melanogaster diaphanous* gene is disrupted in a patient with premature ovarian failure: evidence for conserved function in oogenesis and implications for human sterility. *Am J Hum Genet* **62**, 533-541. (1998).

117 Tsui, S., Dai, T., Roettger, S., Schempp, W., Salido, E. C. and Yen, P. H. Identification of two novel proteins that interact with germ-cell- specific RNA-binding proteins DAZ and DAZL1. *Genomics* **65**, 266-273 (2000).

118 MacLennan, W. J. and Sellers, W. I. Ageing through the ages. *Proc R Coll Physicians Edinb* **29**, 71-75 (1999).

119 Ding, J., Hayashi, M. K., Zhang, Y., Manche, L., Krainer, A. R. and Xu, R. M. Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. *Genes Dev* **13**, 1102-1115. (1999).

120 Wang, X. and Tanaka Hall, T. M. Structural basis for recognition of AU-rich element RNA by the HuD protein. *Nat Struct Biol* **8**, 141-145 (2001).

121 Burd, C. G. and Dreyfuss, G. Conserved structures and diversity of functions of RNA-binding proteins. *Science* **265**, 615-621. (1994).

122 Najmabadi, H., Huang, V., Yen, P., Subbarao, M. N., Bhasin, D., Banaag, L., Naseeruddin, S., de Kretser, D. M., Baker, H. W., McLachlan, R. I. and et al. Substantial prevalence of microdeletions of the Y-chromosome in infertile men with idiopathic azoospermia and oligozoospermia detected using a sequence-tagged site-based mapping strategy. *J Clin Endocrinol Metab* **81**, 1347-1352 (1996).

123 Kostiner, D. R., Turek, P. J. and Reijo, R. A. Male infertility: analysis of the markers and genes on the human Y chromosome. *Hum Reprod* **13**, 3032-3038 (1998).

124 Westerveld, G. H., Repping, S., Leschot, N. J., van der Veen, F. and Lombardi, M. P. Mutations in the human *BOULE* gene are not a major cause of impaired spermatogenesis. *Fertil Steril* **83**, 513-515 (2005).

125 Maegawa, S., Yamashita, M., Yasuda, K. and Inoue, K. Zebrafish DAZ-like protein controls translation via the sequence 'GUUC'. *Genes Cells* **7**, 971-984 (2002).

126 Yoshimizu, T., Sugiyama, N., De Felice, M., Yeom, Y., Ohbo, K., Masuko, K., Obinata, M., Abe, K., Scholer, H., Matsui, Y. Germline-specific expression of the Oct-4/green fluorescent protein (GFP) transgene in mice. *Development Growth & Differentiation* **41**, 675-684 (1999).

127 Tilly, J. L. Ovarian follicle counts - not as simple as 1, 2, 3. *Reprod Biol Endocrinol* **1**, 11 (2003).

128 Ohinata, Y., Payer, B., O'Carroll, D., Ancelin, K., Ono, Y., Sano, M., Barton, S. C., Obukhanych, T., Nussenzweig, M., Tarakhovsky, A., Saitou, M. and Surani, M. A. *Blimp1* is a critical determinant of the germ cell lineage in mice. *Nature* **436**, 207-213 (2005).

129 Bendsen, E., Byskov, A. G., Laursen, S. B., Larsen, H. P., Andersen, C. Y. and Westergaard, L. G. Number of germ cells and somatic cells in human fetal testes during the first weeks after sex differentiation. *Hum Reprod* **18**, 13-18 (2003).

130 Bendsen, E., Byskov, A. G., Andersen, C. Y. and Westergaard, L. G. Number of germ cells and somatic cells in human fetal ovaries during the first weeks after sex differentiation. *Hum Reprod* (2005).

131 Clark, A. T., Bodnar, M. S., Fox, M., Rodriquez, R. T., Abeyta, M. J., Firpo, M. T. and Pera, R. A. Spontaneous differentiation of germ cells from human embryonic stem cells in vitro. *Hum Mol Genet* **13**, 727-739 (2004).

132 Di Carlo, A. D., Travia, G. and De Felici, M. The meiotic specific synaptonemal complex protein SCP3 is expressed by female and male primordial germ cells of the mouse embryo. *Int J Dev Biol* **44**, 241-244 (2000).

133 Bienvenu, T., Patrat, C., McElreavey, K., de Almeida, M. and Jouannet, P. Reduction in the *DAZ* gene copy number in two infertile men with impaired spermatogenesis. *Ann Genet* **44**, 125-128 (2001).

134 Becherini, L., Guarducci, E., Degl'Innocenti, S., Rotondi, M., Forti, G. and Krausz, C. *DAZL* polymorphisms and susceptibility to spermatogenic failure: an example of remarkable ethnic differences. *Int J Androl* **27**, 375-381 (2004).

135 Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. L. and Postlethwait, J. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**, 1531-1545 (1999).

136 Wistuba, J., Schrod, A., Greve, B., Hodges, J. K., Aslam, H., Weinbauer, G. F. and Luetjens, C. M. Organization of seminiferous epithelium in primates: relationship to spermatogenic efficiency, phylogeny, and mating system. *Biol Reprod* **69**, 582-591 (2003).

137 Reijo, R. A., Dorfman, D. M., Slee, R., Renshaw, A. A., Loughlin, K. R., Cooke, H. and Page, D. C. DAZ family proteins exist throughout male germ cell development and transit from nucleus to cytoplasm at meiosis in humans and mice. *Biol Reprod* **63**, 1490-1496. (2000).

138 Moore, F. L., Jaruzelska, J., Dorfman, D. M. and Reijo-Pera, R. A. Identification of a novel gene, *DZIP* (*DAZ-interacting protein*), that encodes a protein that interacts with DAZ (deleted in azoospermia) and is expressed in embryonic stem cells and germ cells. *Genomics* **83**, 834-843 (2004).

139 Gromoll, J., Weinbauer, G. F., Skaletsky, H., Schlatt, S., Rocchietti-March, M., Page, D. C. and Nieschlag, E. The Old World monkey *DAZ* (*Deleted in AZoospermia*) gene yields insights into the evolution of the *DAZ* gene cluster on the human Y chromosome. *Hum Mol Genet* **8**, 2017-2024 (1999).

140 Comeron, J. M. K-Estimator: calculation of the number of nucleotide substitutions per site and the confidence intervals. *Bioinformatics* **15**, 763-764 (1999).

141 Tajima, F. Simple methods for testing the molecular evolutionary clock hypothesis. *Genetics* **135**, 599-607 (1993).

142 Kumar, S., Tamura, K., Jakobsen, I. B. and Nei, M. MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17**, 1244-1245 (2001).

143 Saxena, R., de Vries, J. W., Repping, S., Alagappan, R. K., Skaletsky, H., Brown, L. G., Ma, P., Chen, E., Hoovers, J. M. and Page, D. C. Four *DAZ* genes in two clusters found in the AZFc region of the human Y chromosome. *Genomics* **67**, 256-267. (2000).

144 Slee, R., Grimes, B., Speed, R. M., Taggart, M., Maguire, S. M., Ross, A., McGill, N. I., Saunders, P. T. and Cooke, H. J. A human *DAZ* transgene confers partial rescue of the mouse *Dazl* null phenotype. *Proc Natl Acad Sci U S A* **96**, 8040-8045. (1999).

145 Ellsworth, D. L., Hewett-Emmett, D. and Li, W. H. Insulin-like growth factor II intron sequences support the hominoid rate-slowdown hypothesis. *Mol Phylogenet Evol* **2**, 315-321 (1993).

146 Li, W. H. and Tanimura, M. The molecular clock runs more slowly in man than in apes and monkeys. *Nature* **326**, 93-96 (1987).

147 Koop, B. F., Tagle, D. A., Goodman, M. and Slightom, J. L. A molecular view of primate phylogeny and important systematic and evolutionary questions. *Mol Biol Evol* **6**, 580-612 (1989).

148 Li, W. H., Ellsworth, D. L., Krushkal, J., Chang, B. H. and Hewett-Emmett, D. Rates of nucleotide substitution in primates and rodents and the generation-time effect hypothesis. *Mol Phylogenet Evol* **5**, 182-187 (1996).

149 Marshall, V. S., Browne, M. A., Knowles, L., Golos, T. G. and Thomson, J. A. Ovarian stimulation of marmoset monkeys (*Callithrix jacchus*) using recombinant human follicle stimulating hormone. *J Med Primatol* **32**, 57-66 (2003).

150 Martin, A. P. and Palumbi, S. R. Body size, metabolic rate, generation time, and the molecular clock. *Proc Natl Acad Sci U S A* **90**, 4087-4091 (1993).

151 Kawamura, S., Tanabe, H., Watanabe, Y., Kurosaki, K., Saitou, N. and Ueda, S. Evolutionary rate of immunoglobulin alpha noncoding region is greater in hominoids than in Old World monkeys. *Mol Biol Evol* **8**, 743-752 (1991).

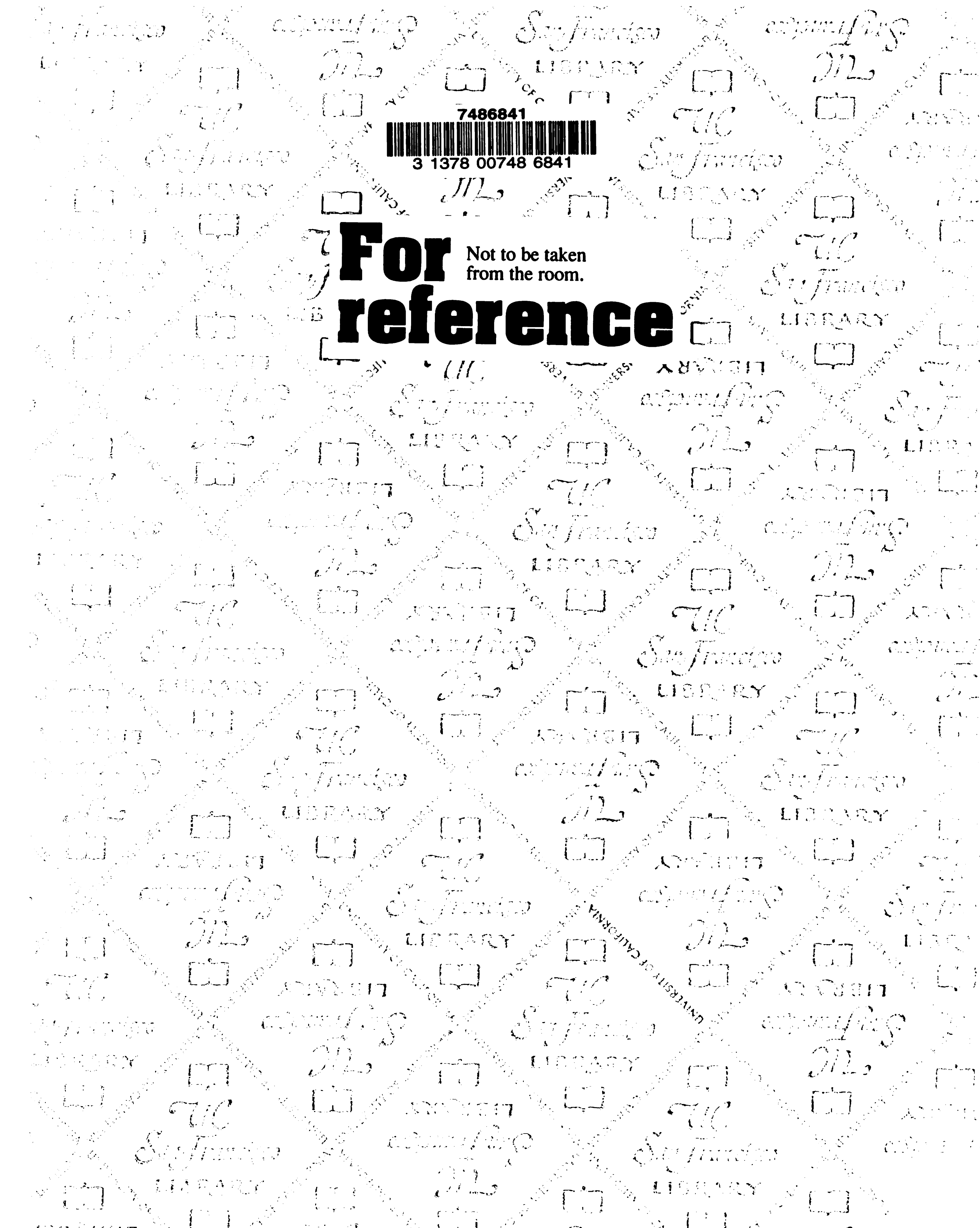
152 Urano, J., Fox, M. S. and Reijo Pera, R. A. Interaction of the conserved meiotic regulators, BOULE (BOL) and PUMILIO-2 (PUM2). *Mol Reprod Dev* **71**, 290-298 (2005).

153 Chabot, B., Stephenson, D. A., Chapman, V. M., Besmer, P. and Bernstein, A. The proto-oncogene *c-kit* encoding a transmembrane tyrosine kinase receptor maps to the mouse W locus. *Nature* **335**, 88-89 (1988).

154 Dalby, B. and Glover, D. M. Discrete sequence elements control posterior pole accumulation and translational repression of maternal *cyclin B* RNA in *Drosophila*. *Embo J* **12**, 1219-1227 (1993).

155 Gavis, E. R. and Lehmann, R. Translational regulation of *nanos* by RNA localization. *Nature* **369**, 315-318 (1994).

156 Kim-Ha, J., Kerr, K. and Macdonald, P. M. Translational regulation of *oskar* mRNA by bruno, an ovarian RNA-binding protein, is essential. *Cell* **81**, 403-412. (1995).


7486841
3 1378 00748 6841
For reference
Not to be taken from the room.