© 2014. Published by The Company of Biologists Ltd | Development (2014) 141, 3662-3671 doi:10.1242/dev.113936

RESEARCH ARTICLE

The mammalian Doublesex homolog DMRT6 coordinates the transition between mitotic and meiotic developmental programs during spermatogenesis

Teng Zhang[1,3], Mark W. Murphy[1], Micah D. Gearhart[1], Vivian J. Bardwell[1,2] and David Zarkower[1,2,*]

ABSTRACT

In mammals, a key transition in spermatogenesis is the exit from spermatogonial differentiation and mitotic proliferation and the entry into spermatocyte differentiation and meiosis. Although several genes that regulate this transition have been identified, how it is controlled and coordinated remains poorly understood. Here, we examine the role in male gametogenesis of the *Doublesex*-related gene *Dmrt6* (*Dmrtb1*) in mice and find that *Dmrt6* plays a crucial role in directing germ cells through the mitotic-to-meiotic germ cell transition. DMRT6 protein is expressed in late mitotic spermatogonia. In mice of the C57BL/6J strain, a null mutation in *Dmrt6* disrupts spermatogonial differentiation, causing inappropriate expression of spermatogonial differentiation factors, including SOHLH1, SOHLH2 and DMRT1 as well as the meiotic initiation factor STRA8, and causing most late spermatogonia to undergo apoptosis. In mice of the 129Sv background, most *Dmrt6* mutant germ cells can complete spermatogonial differentiation and enter meiosis, but they show defects in meiotic chromosome pairing, establishment of the XY body and processing of recombination foci, and they mainly arrest in mid-pachynema. mRNA profiling of *Dmrt6* mutant testes together with DMRT6 chromatin immunoprecipitation sequencing suggest that DMRT6 represses genes involved in spermatogonial differentiation and activates genes required for meiotic prophase. Our results indicate that *Dmrt6* plays a key role in coordinating the transition in gametogenic programs from spermatogonial differentiation and mitosis to spermatocyte development and meiosis.

KEY WORDS: DMRT6, DMRT1, Spermatogenesis, Meiosis, Testis, Mouse

INTRODUCTION

Male mammals continuously produce large numbers of sperm from puberty onwards, in some cases for decades. This sustained gametogenesis is supported by a population of undifferentiated type A spermatogonia that includes spermatogonial stem cells, or SSCs (Nakagawa et al., 2010), and is embedded within a seminiferous epithelium formed by Sertoli cells (Griswold, 1998). In the course of developing from single undifferentiated A_s spermatogonia to mature spermatozoa, male germ cells in the mouse undergo a series of nine mitotic and two meiotic cell divisions, accompanied by a number of differentiation steps. This extended series of cell divisions allows an individual A_s cell that undergoes differentiation to produce more than 4000 spermatozoa and involves dramatic changes, that include epigenetic reprogramming, reduction of ploidy from 2N to 1N, repackaging of chromatin by histone-to-protamine replacement and morphological transformation to allow survival and motility in the female reproductive tract (Dada et al., 2012; de Rooij and Russell, 2000; Jan et al., 2012).

Spermatogenesis involves several crucial transitions between distinct developmental programs. During steady-state spermatogenesis, division of A_s spermatogonia produces, by incomplete cytokinesis, chains of two to 16 undifferentiated spermatogonia (A_{pr} and A_{al}) linked by cytoplasmic bridges. Spermatogonial differentiation starts when these undifferentiated spermatogonia form A_1 spermatogonia. This transition occurs in waves that sweep the seminiferous tubules with a species-specific period, which is 8.6 days in mice. A_1 spermatogonia then undergo five additional rounds of mitotic division, coupled with differentiation, to generate $A_{2\text{-}4}$, intermediate (Int) and type B spermatogonia (de Rooij, 1998, 2001; de Rooij and Grootegoed, 1998; de Rooij and Russell, 2000). A_1-B spermatogonia are collectively called differentiating spermatogonia, but they have discrete morphologies and gene expression profiles and are found at distinct stages of the seminiferous epithelial cycle.

When spermatogonial differentiation is complete, germ cells need to make a further transition into meiosis. B spermatogonia divide mitotically to form preleptotene spermatocytes, which undergo a final mitotic division before entering meiosis to eventually form haploid spermatids. Spermatids do not divide further after meiosis but undergo extensive postmeiotic differentiation to form spermatozoa. Because spermatid differentiation requires 35 days but new waves of differentiation initiate every 8.6 days, differentiating germ cells accumulate in layers above the undifferentiated spermatogonia. The cellular composition of these layers varies during the cycle, allowing the cycle to be divided into morphologically distinct stages (de Rooij, 1998, 2001; de Rooij and Grootegoed, 1998; de Rooij and Russell, 2000). Continuously producing large numbers of gametes without depleting the intermediate cell types requires tight and well-integrated control of proliferation and differentiation at many points in the process, all in the context of major transitions in the genome, epigenome and cell biology of the male germ line.

Spermatogonial development is under the control of two prominent signaling pathways. GDNF signaling plays a crucial role in maintenance of SSCs. Mutations in *Gdnf* or its co-receptors *Ret* and *Gfra1* cause progressive germ cell loss indicative of SSC depletion, whereas overexpression of GDNF causes accumulation of undifferentiated A_s cells (Buageaw et al., 2005; Meng et al., 2000; Naughton et al., 2006). Retinoic acid (RA) is required for the initiation of spermatogonial differentiation in the juvenile testis (Mark et al., 2008), for entry of undifferentiated spermatogonia into

[1]Developmental Biology Center, Department of Genetics, Cell Biology, and Development, University of Minnesota, Minneapolis, MN 55455, USA. [2]University of Minnesota Masonic Cancer Center, Minneapolis, MN 55455, USA. [3]Molecular, Cellular, Developmental Biology and Genetics Graduate Program, University of Minnesota, Minneapolis, MN 55455, USA.

*Author for correspondence (zarko001@umn.edu)

Received 5 June 2014; Accepted 1 August 2014

Development (2014) 141, 3662-3671 doi:10.1242/dev.113936

differentiation (the A_{al} to A_1 transition) during steady-state adult spermatogenesis and likely for the initiation of meiosis by preleptotene spermatocytes (Griswold et al., 1989; Hogarth and Griswold, 2010; McCarthy and Cerecedo, 1952; Snyder et al., 2010; Thompson et al., 1964; Van Pelt and de Rooij, 1990). The latter two functions of RA are thought to be mediated by the RA-inducible gene *Stra8* (Anderson et al., 2008; Mark et al., 2008; Oulad-Abdelghani et al., 1996).

A number of transcriptional regulators also have been shown to play essential roles in controlling spermatogonial differentiation, including the basic helix-loop-helix (bHLH) proteins SOHLH1 and SOHLH2, and the DMRT protein DMRT1 (Ballow et al., 2006; Hao et al., 2008; Matson et al., 2010; Suzuki et al., 2012). DMRT proteins are transcription factors that bind DNA via the DM domain, a structurally distinct class of zinc-finger motif (Erdman and Burtis, 1993; Zhu et al., 2000). DMRT proteins occur in virtually all metazoan animals and regulate sexual development in a wide variety of species, ranging from planaria to insects to nematodes to vertebrates (Matson and Zarkower, 2012). In mice, DMRT1 is required in germ cells at several stages of their development, and another DMRT protein, DMRT7 (DMRTC2 – Mouse Genome Informatics), associates with the sex chromosomes of spermatocytes during meiosis and is required for sex chromatin modification (Fahrioglu et al., 2007; Kim et al., 2007b; Krentz et al., 2009; Matson et al., 2010; Raymond et al., 2000).

Here, we examine the role of the DMRT protein DMRT6 (DMRTB1 – Mouse Genome Informatics) in gametogenesis. Previous analysis showed that *Dmrt6* is widely conserved among vertebrates and is expressed strongly in the gonad in mice (Kim et al., 2003; Ottolenghi et al., 2002). We show that DMRT6 protein is expressed in the postnatal mouse testis in differentiating spermatogonia, disappearing as B spermatogonia become preleptotene spermatocytes. Using a *Dmrt6* null allele, we found that *Dmrt6* is crucial for spermatogenesis: loss of *Dmrt6* in C57BL/6J (B6) mice disrupted the transition from A_4 to In and B spermatogonia and caused the extended expression of spermatogonial differentiation factors, such as SOHLH1, SOHLH2 and DMRT1 and the meiotic initiation factor STRA8, into inappropriate cell types. Analysis of *Dmrt6* in mice of the 129Sv genetic background revealed an additional requirement for spermatogonial expression of *Dmrt6*: in these mice most spermatogonia completed differentiation and enter meiosis but showed defects in chromosome pairing, establishment of the XY body and processing of recombination foci, with very few cells progressing beyond mid-pachynema. mRNA profiling of *Dmrt6* mutant testes and DMRT6 chromatin immunoprecipitation sequencing (ChIP-seq) analysis suggested that DMRT6 helps to coordinate the transition from spermatogonial development to meiosis by repressing genes involved in spermatogonial differentiation and by activating genes required for meiotic prophase. Relatively little is known about how differentiation of late-stage spermatogonia is controlled or how spermatogonia make the transition to spermatocytic development. Our results reveal that *Dmrt6* plays a key role in coordinating an orderly transition between gametogenic programs from spermatogonial differentiation and mitosis to spermatocyte development and meiosis, and allow identification of a number of new candidates to mediate this process.

RESULTS

DMRT6 is expressed in intermediate and B spermatogonia

Dmrt6 is one of seven vertebrate DM domain genes (Kim et al., 2003; Ottolenghi et al., 2002). Some analysis of *Dmrt6* expression has previously been reported: in the fetal mouse *Dmrt6* mRNA was detected primarily in the brain (Kim et al., 2003). *Dmrt6* is strongly expressed in adult testis in mouse and human and weakly in human adult ovary and pancreas (Ottolenghi et al., 2002). mRNA expression profiling in the postnatal mouse testis detected *Dmrt6* from P5 onward (www.mrgd.org) (Shima et al., 2004; Su et al., 2004). In the adult mouse, using RT-PCR we detected *Dmrt6* strongly in testis and more weakly in ovary and brain but not in pancreas (supplementary material Fig. S1).

To examine expression of DMRT6 protein we generated a polyclonal antibody directed against a region of DMRT6 C-terminal to the DM domain and confirmed that it specifically recognizes DMRT6 in immunofluorescence (IF) and western blots (Fig. 1; supplementary material Fig. S2A-C). ENSEMBL predicts DMRT6 to have two potential protein isoforms of about 22 kDa and 38 kDa. We detected just one strongly expressed protein of about 47 kDa that was specific to wild-type testes (supplementary material Fig. S2C). The slower than predicted gel mobility is typical of DMRT proteins and may reflect secondary structure or protein modification.

We first compared expression of DMRT6 with that of DMRT1, which is expressed in Sertoli cells and spermatogonia (Matson et al., 2010; Raymond et al., 2000). Double staining of adult testes for DMRT6 and DMRT1 showed that DMRT6 is expressed in a subset of spermatogonia and not in Sertoli cells or in meiotic and postmeiotic germ cells (Fig. 1A). Double staining for DMRT6 and the meiotic marker SYCP3 (Yuan et al., 2000) showed mutually exclusive expression, confirming that DMRT6 is not expressed in germ cells that have initiated meiotic prophase (Fig. 1B).

To determine which spermatogonial cell types express DMRT6 we examined its expression relative to proteins expressed during different stages of spermatogonial development and also assessed during which stages of the cycle of the seminiferous epithelium DMRT6 is expressed. Double staining showed no overlap of DMRT6 expression with the undifferentiated spermatogonial marker PLZF (ZBTB16 – Mouse Genome Informatics) (Ballow et al., 2006; Costoya et al., 2004), indicating that DMRT6 is expressed only in differentiating spermatogonia (Fig. 1C,D). SOHLH1 and SOHLH2 proteins appear together in a subset of A_{al} spermatogonia; SOHLH2 expression then disappears before formation of Int spermatogonia, whereas SOHLH1 disappears later, before formation of type B spermatogonia (Suzuki et al., 2012). In the seminiferous epithelial cycle we found that DMRT6 expression begins during stage I and encompasses SOHLH2-positive differentiating A_4 spermatogonia (Fig. 1E,F) as well as SOHLH1-positive Int and SOHLH1-negative B spermatogonia (Fig. 1G-J) (Ballow et al., 2006; Hao et al., 2008; Suzuki et al., 2012). DMRT6 was absent by stage VII, when preleptotene spermatocytes are formed, indicating that its disappearance is coincident with the completion of spermatogonial differentiation (Fig. 1K,L). A diagram summarizing the dynamics of DMRT6 expression based on these data is shown in Fig. 1M.

Dmrt6 regulates the transition from early to late spermatogonial differentiation

To investigate the role of *Dmrt6* in spermatogonial development, we generated a conditional mutant allele, $Dmrt6^{flox}$, in which the proximal promoter and first exon are flanked by Cre recombinase recognition sites (*loxP* sites; 'floxed') (supplementary material Fig. S2E). Breeding $Dmrt6^{flox}$ to *beta-actin-Cre* mice generated the putative null allele $Dmrt6^{-}$, which was used for all experiments. $Dmrt6^{-}$ lacks both the proximal promoter and sequences encoding the DNA-binding DM domain, which is essential for DMRT protein

Development (2014) 141, 3662-3671 doi:10.1242/dev.113936

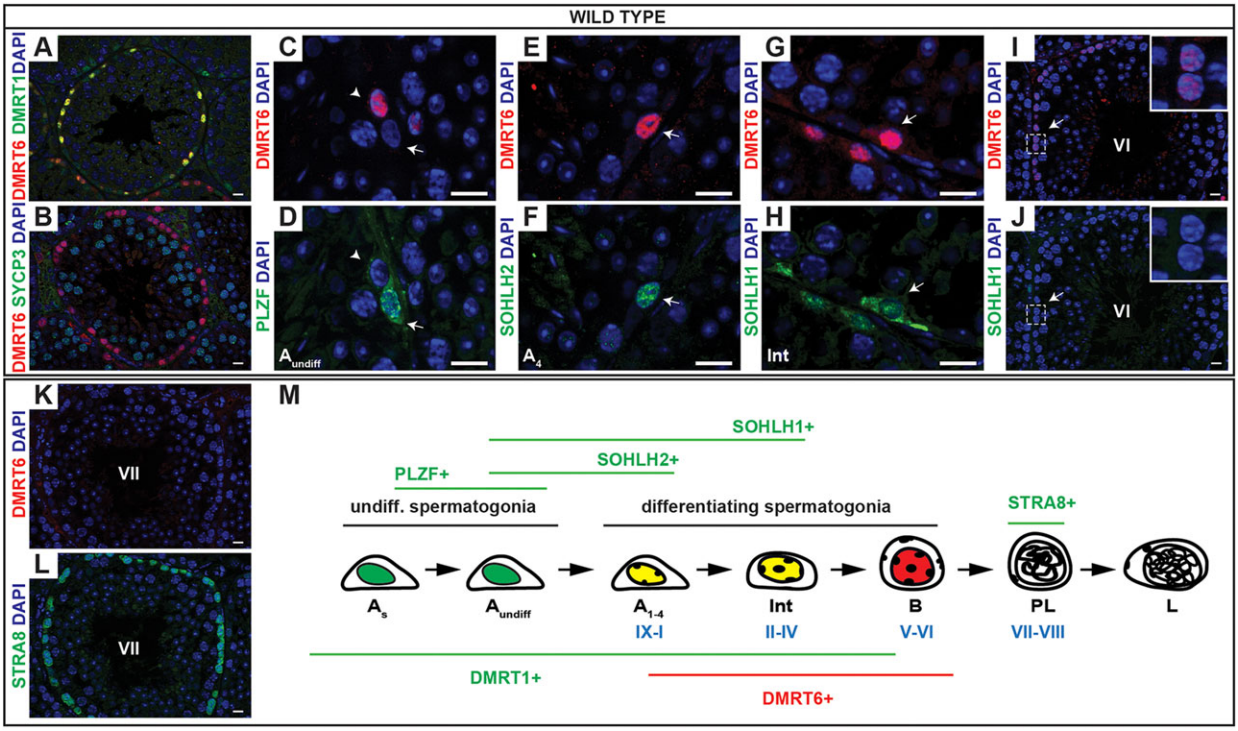

Fig. 1. DMRT6 is expressed in differentiating spermatogonia in the adult mouse testis. IF of adult wild-type testis sections. (A) DMRT6 is expressed in DMRT1-positive spermatogonia but not in DMRT1-positive Sertoli cells. (B) DMRT6 is not expressed in SYCP3-positive primary spermatocytes. (C-L) Paired co-stainings of DMRT6 with other germ cell proteins. (C,D) DMRT6 expression does not overlap with that of the undifferentiated spermatogonial marker PLZF. (E,F) DMRT6 expression overlaps with that of the spermatogonial differentiation marker SOHLH2 in A4 spermatogonia. (G,H) DMRT6 expression overlaps with that of the spermatogonial differentiation marker SOHLH1 in Int spermatogonia. (I,J) DMRT6 is expressed in B spermatogonia at stage VI, whereas SOHLH1 is not. (K,L) Preleptotene spermatocytes at stage VII are negative for DMRT6 and strongly express STRA8. (M) Graphical summary of expression data from A-L. Scale bars: 10 µm.

function, and does not express detectable DMRT6 protein (supplementary material Fig. S2A-C). *Dmrt6*$^{-/-}$ males had severe spermatogenesis defects (described below), but were otherwise normal and healthy. *Dmrt6*$^{-/-}$ females were normal and fertile with no apparent ovarian defects (not shown).

Testes of *Dmrt6*$^{-/-}$ mutants on the C57BL/6J (B6) genetic background were small and deficient in TRA98-positive differentiating germ cells as early as postnatal day 10 (supplementary material Fig. S2D; Fig. 2A-F) but had apparently normal SOX9-positive Sertoli cells (Fig. 2B,D,F). No mature spermatozoa were detected either in the seminiferous tubules or the epididymis in 8-week-old adult mutant males (Fig. 2E-H). Mutant animals were also aged to 8 months and we did not observe any striking progression of the phenotype, suggesting that the spermatogonial stem cell population was not compromised. To further evaluate the basis of the germ cell deficiency we examined expression of markers of major germ cell types. In adult mutant testes, undifferentiated spermatogonia appeared normal and expressed PLZF (not shown), but differentiating spermatogonia had altered expression of the differentiation proteins SOHLH1 and SOHLH2 (Fig. 3A-D). The number of spermatogonia strongly expressing these proteins was greatly increased in *Dmrt6* mutants, suggesting either an accumulation of differentiating spermatogonia or misexpression of these proteins outside their normal cell types. By contrast, the number of strongly STRA8-positive preleptotene spermatocytes was greatly reduced (Fig. 3E,F). Likewise, the number of BC7-positive primary spermatocytes (stages IX-XII) was greatly reduced and SUMO-1 positive XY bodies, which normally form during mid-pachynema (Koshimizu et al., 1995; La Salle et al., 2008), were absent from DMRT6 mutant testes (Fig. 3G,H). Together, these results indicate that loss of *Dmrt6* disrupts spermatogonial differentiation starting at or before the formation of preleptotene spermatocytes and also prevents progression of meiotic prophase beyond early pachynema.

Cell-type-inappropriate gene expression in *Dmrt6* mutant germ cells

As mentioned above, the increased number of cells expressing SOHLH1 and SOHLH2 could reflect either an accumulation of differentiating spermatogonia or the expression of these proteins outside their normal cell types. Consistent with the latter possibility, we observed many SOHLH1- and SOHLH2-positive germ cells in mutant tubules at stages II-VII, when Int and B spermatogonia and preleptotene spermatocytes normally are present and SOHLH2 should not be expressed. To confirm that key regulators were expressed outside their normal cell types, we performed double-staining experiments in adult mutant testes. These revealed that both SOHLH1 and SOHLH2 were ectopically expressed with STRA8 in mutant preleptotene spermatocytes (Fig. 4A-D), and that mutant stage VII preleptotene spermatocytes also ectopically expressed DMRT1 (Fig. 4E,F), which normally is silenced in late B spermatogonia (Matson et al., 2010). STRA8 was also misexpressed: it normally disappears before meiotic entry (Oulad-Abdelghani et al., 1996), but in *Dmrt6* mutants STRA8 persisted into meiotic prophase in stage IX SYCP3-positive leptotene spermatocytes (Fig. 4G,H). Based on these results, we conclude that *Dmrt6* mutant spermatogonia can progress to preleptonema. However, they fail to appropriately downregulate expression of key regulatory genes at several stages of differentiation:

Development (2014) 141, 3662-3671 doi:10.1242/dev.113936

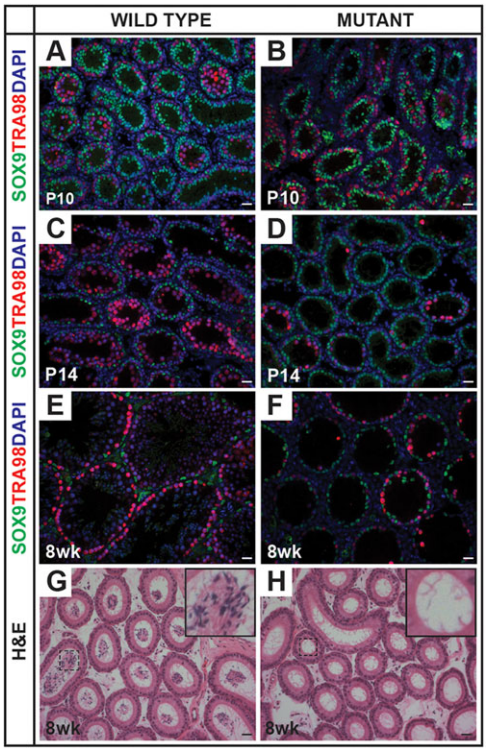

Fig. 2. Germ cell depletion in *Dmrt6* mutant testes. (A,B) By P10, *Dmrt6* mutants are deficient in TRA98-positive differentiating germ cells (primarily differentiating spermatogonia and primary spermatocytes) but have normal expression of the Sertoli cell marker SOX9. (C,D) At P14, a severe deficit in meiotic germ cells, which accumulate in the interior of seminiferous tubules in wild type, is clearly apparent in *Dmrt6* mutants. (E,F) In wild-type adults a ring of differentiating spermatogonia and primary spermatocytes, which have very strong TRA98 expression, is apparent; *Dmrt6* mutants have many fewer cells with strong TRA98 expression and lack spermatids based on DAPI staining. (G,H) Hematoxylin and eosin (H&E) staining detects abundant epididymal sperm in wild type but none in *Dmrt6* mutants. Scale bars: 20 µm.

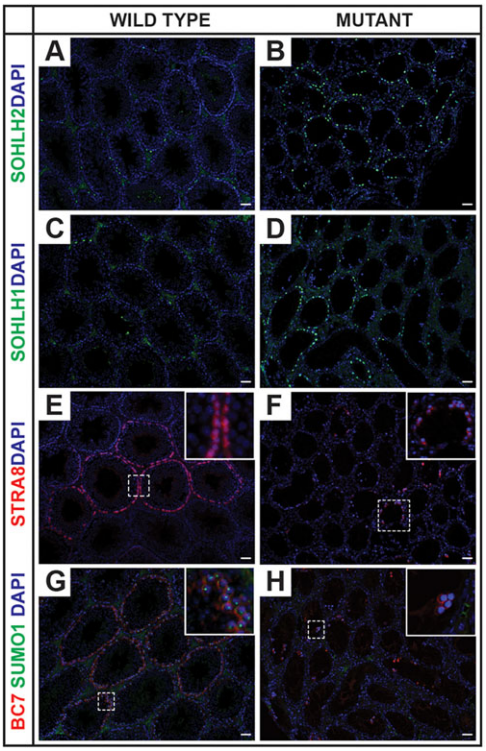

Fig. 3. Disrupted spermatogonial development in *Dmrt6* mutants. IF of adult testes. (A,B) *Dmrt6* mutants accumulate spermatogonia strongly expressing SOHLH2, which is normally expressed mainly in undifferentiated spermatogonia through A_4 spermatogonia. (C,D) *Dmrt6* mutants accumulate spermatogonia strongly expressing SOHLH1, which is normally expressed in undifferentiated spermatogonia through Int spermatogonia. (E,F) *Dmrt6* mutants are severely deficient in strongly STRA8-positive preleptotene spermatocytes and appear to have reduced nuclear STRA8 relative to wild type (insets). (G,H) *Dmrt6* mutants are severely deficient in germ cells expressing the primary spermatocyte marker BC7 and completely lack SUMO1-positive XY bodies that normally form in mid-pachynema (insets). Scale bars: 40 µm.

SOHLH2 when they progress from A4 to Int; SOHLH1 when they progress from Int to B; DMRT1 when they progress from B to preleptotene; and STRA8 when they enter meiotic prophase. The inappropriate combinations of germ cell regulators that result are likely to disrupt the transition from mitosis to meiosis and cause the observed failure of spermatogenesis in *Dmrt6* mutants.

The cell type analysis described above showed that *Dmrt6* mutants have abundant A_4 spermatogonia but become severely deficient in germ cells from late spermatogonial stages onward. Presumably this germ cell deficiency represents reduced proliferation or increased cell death of late spermatogonia, or possibly both. We used bromodeoxyuridine (BrdU) incorporation to assess proliferation and terminal deoxynucleotidyl transferase dUTP nick-end labeling (TUNEL) to assess apoptosis in juveniles at 10 and 14 days and in adults. *Dmrt6* mutant germ cells from all three time points were positive for BrdU in a stage-specific manner as in wild type, but the percentage of BrdU-positive A_4 cells was slightly higher in *Dmrt6* mutant tubules, consistent with our observation of increased A_4 cells in *Dmrt6* mutants and suggesting that their proliferation was not reduced and may indeed have been increased (not shown). Mutants at each time point had TUNEL-positive cells in tubules containing late spermatogonia, unlike the control, where apoptotic cells were rare in these stages (Fig. 4I,J). Together, these results suggest that the reduced abundance of late spermatogonia in *Dmrt6* mutants is largely due to elevated apoptosis.

DMRT6 activity in mitotic spermatogonia is required for progression through meiotic prophase

Dmrt6 mutant germ cells have at least two apparent defects, a disruption of spermatogonial differentiation and a failure to complete meiotic prophase. We were able to partially separate these phenotypes by breeding the *Dmrt6* mutation onto a different genetic background. We found that loss of *Dmrt6* in 129Sv strain mice caused minimal loss of spermatogonia, with greater numbers of germ cells entering meiosis, but mutant cells still arrested in meiotic prophase as in B6 mutants (supplementary material Fig. S3A-D). The genetic separation of these phenotypes suggests that DMRT6 performs two distinct functions during late spermatogonial differentiation, one required at the time of DMRT6 expression for completion of spermatogonial development and one that becomes important later for successful execution of meiotic prophase.

We used the 129Sv mutant strain to more closely examine the meiotic defect in *Dmrt6* mutants by assessing several landmarks of normal meiotic prophase. First, we stained meiotic chromosome spreads with SYCP1 and SYCP3 to examine synaptonemal complex (SC) formation, which normally begins during leptonema and is completed by pachynema (Cohen et al., 2006; Fraune et al., 2012; Yang and Wang, 2009). The SC formed in mutant spermatocytes but frequently was incomplete: 37/50 mutant pachytene spermatocytes had incomplete synapsis compared with 4/50 in wild type (Fig. 5A-F).

Development (2014) 141, 3662-3671 doi:10.1242/dev.113936

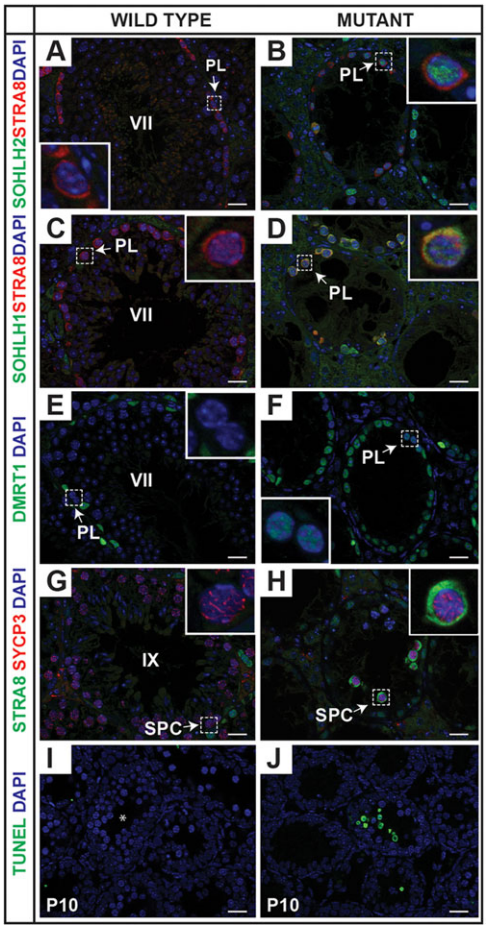

Fig. 4. Inappropriate overlapping expression of germ cell differentiation regulators. (A,B) Double staining for the differentiating A spermatogonial marker SOHLH2 and preleptotene spermatocyte marker STRA8 shows no overlap in wild type at stage VII but preleptotene spermatocytes in *Dmrt6* mutant express both markers (insets). (C,D) Wild-type preleptotene spermatocytes express STRA8 but not SOHLH1, whereas mutant preleptotene spermatocytes express both proteins (insets). (E,F) Preleptotene spermatocytes in wild type do not express DMRT1 but many *Dmrt6* mutants do (insets). (G,H) Wild-type stage IX primary spermatocytes express SYCP3 but not STRA8, whereas *Dmrt6* mutant primary spermatocytes frequently express both (insets). (I,J) TUNEL labeling at P10 detects elevated apoptosis in mutant spermatogonia. Scale bars: 20 µm. PL, preleptotene spermatocyte; SPC, primary spermatocyte.

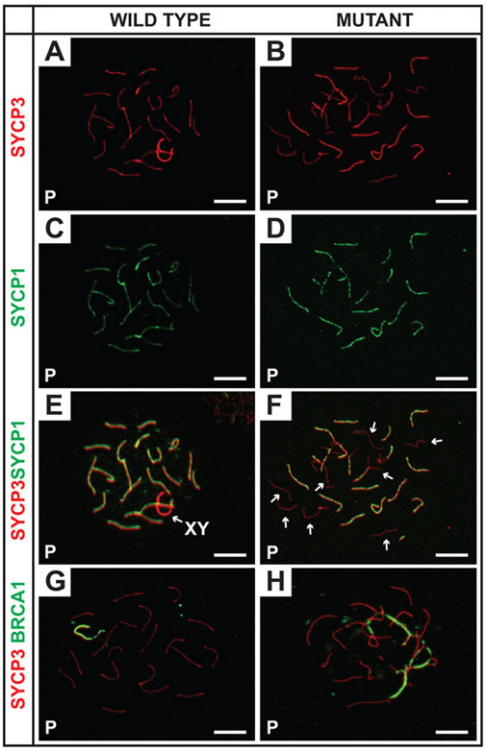

Fig. 5. Incomplete chromosomal synapsis in *Dmrt6* primary spermatocytes. (A-H) IF of chromosome spreads from wild-type or *Dmrt6* adult pachytene spermatocytes. (A,B) The synaptonemal complex lateral element protein SYCP3 accumulates in wild-type and *Dmrt6* spermatocytes, indicating that meiotic pairing does occur in mutants. (C,D) The synaptonemal complex transverse element protein SYCP1 also accumulates in wild-type and *Dmrt6* spermatocytes, indicating that synaptonemal complexes in mutant cells can contain key structural components. (E,F) Merge of SYCP3 and SYCP3 staining shows that whereas synaptonemal complex formation is complete in wild type (except for the non-homologous regions of the sex chromosomes labeled 'XY'), synaptonemal complex formation is incomplete in *Dmrt6* mutant spermatocytes not only on the sex chromosomes but also on a number of autosomes. White arrows indicate unpaired chromosomes. (G,H) In wild-type germ cells, BRCA1 accumulates on the unpaired sex chromosomes, but in mutant spermatocytes there is more extensive BRCA1 accumulation, further illustrating the synapsis defect and confirming that surveillance for unpaired chromosomes is active in mutant cells. Scale bars: 10 µm.

The unpaired chromosomes and chromosome segments accumulated BRCA1 (Fig. 5G,H), indicating that surveillance for unpaired chromosomes was intact in the mutant spermatocytes (Turner et al., 2004).

We next examined meiotic crossing over, which normally occurs during pachynema. Double-strand breaks (DSBs) were initiated and repaired in *Dmrt6* mutants, with RAD51 foci normal in number and distribution except on unpaired regions (supplementary material Fig. S4A-D) (Cohen et al., 2006). Transitional nodules made up of MSH4, MSH5 and other associated proteins normally form at DSBs during zygonema (Baudat and de Massy, 2007; Cohen et al., 2006; Hoffmann and Borts, 2004). In *Dmrt6* mutants we observed a severe decrease in the number of MSH4 foci (supplementary material Fig. S4E,F). During pachynema, a subpopulation of transitional nodules should convert to MLH1-positive recombination nodules, at which crossovers occur (Baker et al., 1996; Cohen et al., 2006). In *Dmrt6* mutant testes no MLH1-positive nodules were detectable (supplementary material Fig. S4G,H), indicating a lack of meiotic recombination, and we did not observe spermatocytes progressing beyond pachynema. Together, these results indicate that loss of *Dmrt6* causes defects in meiotic pairing that lead to arrest and apoptosis during pachynema, before recombination. These defects are present both in B6 mutants, which have reduced numbers of late spermatogonia and early spermatocytes, and in 129Sv and mixed background mice, which have normal numbers of differentiated spermatogonia and early primary spermatocytes. Because DMRT6 is expressed only during spermatogonial differentiation, the meiotic defects must be consequences of improper gene regulation before the initiation of the meiotic program.

DMRT6 activates and represses spermatogonial gene expression

Phenotypic analysis indicated that DMRT6 in differentiating spermatogonia is required both for the completion of the spermatogonial program and transition into preleptonema, as well as to properly preconfigure spermatogonia for subsequent success in meiotic prophase. For more insight into how DMRT6 performs

Development (2014) 141, 3662-3671 doi:10.1242/dev.113936

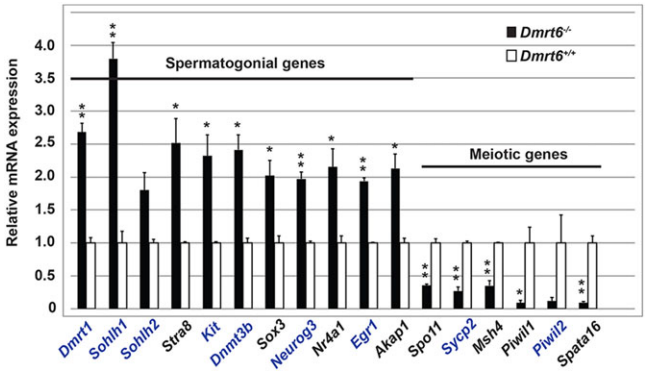

Fig. 6. Misregulation of spermatogonial and spermatocytic mRNAs in *Dmrt6* mutant testes. RT-qPCR comparing mRNA expression in P14 wild-type (white) and *Dmrt6* mutant (black) mice. Expression was normalized to *Hprt* expression. Error bars indicate s.d. **P*<0.05 (Student's *t*-test). ***P*<0.005. Blue font indicates genes that are potential direct target of DMRT6 based on ChIP-seq.

these functions, we profiled mRNA expression in *Dmrt6* mutant testes at P8 on the B6 background. In wild-type testes at this stage the first wave of differentiating DMRT6-positive spermatogonia has formed. In P8 mutants germ cells have not yet arrested and undergone apoptosis, and thus a normal mix of germ cell types is still present, increasing the chance of detecting the proximal changes in gene expression that lead to later phenotypic defects. Because P8 is before the overt spermatogonial phenotype in *Dmrt6* mutants, we should avoid false positives caused by the depletion of late spermatogonia and spermatocytes; however, we would also expect to miss some of the mRNAs misexpressed in those cell types.

We performed RNA sequencing to compare wild-type and *Dmrt6* mutant testes, each in triplicate (data analysis is described in Materials and Methods). This analysis identified 1595 misexpressed genes ($P<0.05$; supplementary material Table S1). As expected, based on the early time point examined, most expression differences between mutant and wild type were modest. We therefore selected a number of genes for validation based on expression in type A and type B spermatagonia and Ingenuity-defined functions in meiosis and testis development (supplementary material Table S2) and assayed expression of 20 of these genes at P14 by RT-qPCR. We also assayed expression of *Dmrt1*, *Sohlh2* and *Stra8* mRNAs at this stage. Finally, because mRNA profiling at P8 was expected to miss meiotic mRNAs activated by DMRT6, we also assayed a number of key meiotic genes by RT-qPCR at P14.

RT-qPCR confirmed the misexpression of many of these mRNAs and also revealed a consistent pattern: genes involved in spermatogonial differentiation or expressed in spermatogonia generally had elevated expression in mutant testes, whereas genes involved in meiosis were reduced in expression (Fig. 6). Among the overexpressed genes were the known regulators of spermatogonial differentiation *Sohlh1* and *Sohlh2*, *Dmrt1*, *Stra8*, *Kit*, *Sox3* and *Neurog3* (Raverot et al., 2005; Schrans-Stassen et al., 1999; Shirakawa et al., 2013; Yoshida et al., 2004), as well as several genes (*Nr4a1*, *Egr1* and *Akap1*) expressed in spermatogonial cells but of unknown function (www.mrgd.org) (McCarrey and Skinner, 1999; Shima et al., 2004). Among the underexpressed genes were meiotic genes, including *spo11*, *sycp2* and *msh4*. Expression of genes involved in epigenetic modification was also altered in *Dmrt6* mutants: *Dnmt3b* is a DNA methyltransferase required for spermatogonial differentiation (Shirakawa et al., 2013), and *Piwil1* (*miwi*) and *Piwil2* (*mili*) are essential regulators of retrotransposon silencing (Di Giacomo et al., 2013; Grivna et al., 2006; Kuramochi-Miyagawa et al., 2004; Vourekas et al., 2012). Collectively, the mRNA profiling and RT-qPCR results suggest that DMRT6 acts in late spermatogonia to coordinate the transition from the mitotic program to the meiotic program by repressing genes that promote spermatogonial differentiation and proliferation and by activating genes that will later be required for meiosis.

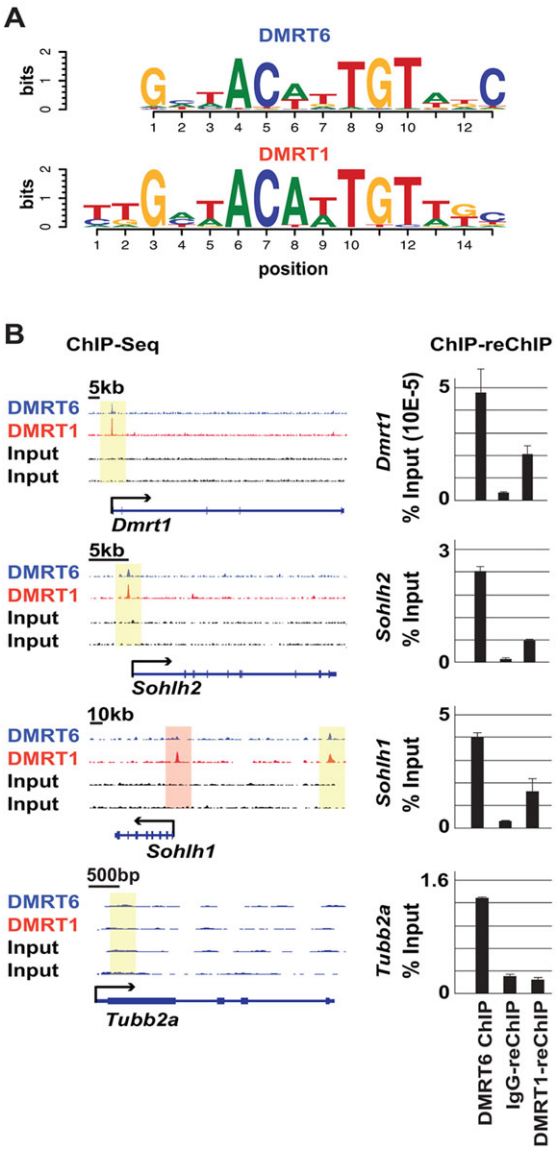

Fig. 7. DMRT6 and DMRT1 bind similar DNA sequences and associate with shared sites *in vivo*. ChIP-Seq was performed on adult testis with antibodies specific to DMRT6 (blue profile) and DMRT1 (red profile). (A) An enriched DNA motif was identified within regions bound by DMRT6 that is similar to the previously described DNA sequence bound by DMRT1 *in vitro*. (B) Left column: binding profiles for both antibodies are shown for the *Dmrt1*, *Sohlh2*, *Sohlh1* and *Tubb2a* (negative control) loci. Yellow filled boxes indicate sites bound comparably by DMRT6 and DMRT1. The red filled box represents preferential association with DMRT1. The data range shown for each binding profile is between 0 and 2 counts per million. Right column: re-ChIP experiments were also conducted for *Dmrt1*, *Sohlh2* and *Sohlh1* loci on sites bound by both DMRT6 and DMRT1 (yellow boxes). ChIP was performed using anti-DMRT6 antibody and then re-ChIP was performed using either protein A purified pre-immune serum (negative control) or anti-DMRT1 antibody. Re-ChIP products were evaluated by qPCR and enrichment reported as percent total input recovered.

Development (2014) 141, 3662-3671 doi:10.1242/dev.113936

To help identify potential direct targets of DMRT6 transcriptional regulation we profiled DMRT6 DNA association in the adult testis using ChIP-Seq. It identified 14,862 peaks associated with 10,363 unique genes (supplementary material Table S3; see Materials and Methods). Binding of DMRT6 at these sites was correlated with binding by DMRT1 (Spearman correlation=0.642). Motif enrichment analysis of the 1724 strongest peaks bound by DMRT6 identified a motif that closely resembled the site bound by DMRT1 (Fig. 7A). Sites bound by both DMRT6 and DMRT1 were more likely to have DMRT1 binding motif than sites specifically enriched for DMRT6 (chi-squared test, P=2.2e-16). In some cases (e.g. *Dmrt1* and *Sohlh2*; Fig. 7B, left column), DMRT1 and DMRT6 binding appeared equivalent, whereas in others (e.g. *Sohlh1*; Fig. 7B, left column), the relative strength of binding of the two proteins appeared to differ at some sites. To determine whether DMRT1 and DMRT6 bind concurrently to the same chromatin interval, we performed re-ChIP analysis, precipitating chromatin first with the DMRT6 antibody and then with the DMRT1 antibody or non-specific pre-immune IgG. Re-ChIP indicated that a number of chromatin sites are associated with both DMRT1 and DMRT6 (Fig. 7B, right column).

In addition to *Dmrt1*, *Sohlh1* and *Sohlh2*, other germ cell regulatory genes that were bound by DMRT6 and misexpressed in *Dmrt6* mutant testes included the spermatogonial regulators *Kit*, *Dnmt3b* and *Neurog3*, and the meiotic genes *Sycp2* and *Piwil2*. The ChIP and RT-qPCR results together suggest that DMRT6 acts as a bifunctional transcription factor to control late spermatogonial development and early meiosis, repressing spermatogonial genes and activating meiotic genes (Fig. 8).

DISCUSSION

Here, we have found that the DM domain protein DMRT6 is expressed in differentiating mammalian spermatogonia and has two distinct functions in male germ cells: first, it helps spermatogonia complete their differentiation program; and second, it is required for spermatocytes to undergo proper meiotic prophase. Based on protein expression data, RNA-seq and ChIP-seq, we suggest DMRT6 helps coordinate the shift from the mitotic spermatogonial program to the meiotic spermatocyte program, one of the major transitions in male gametogenesis, by controlling the transcription of suites of key regulatory genes. DMRT6 ensures that genes promoting spermatogonial proliferation and differentiation are downregulated appropriately during several steps of late spermatogonial development and also activates genes that will be required once meiosis initiates. In this manner, DMRT6 helps bring spermatogonial development to an orderly conclusion, sets the stage for meiosis and ensures the separation of the mitotic and meiotic programs. In *Dmrt6* mutants the timing of gene expression is disrupted, and we suggest that the resulting simultaneous expression of genes normally active during different stages of germ cell differentiation (e.g. Fig. 4) is likely to cause the observed developmental arrest and apoptosis.

Dmrt6, together with *Dmrt1* and *Dmrt7*, is one of three Dmrt genes that have been shown to play essential roles in mammalian spermatogenesis. The three genes act at different steps of germ cell development and have distinct functions. *Dmrt1* regulates proliferation and pluripotency of fetal germ cells, stimulates resumption of mitosis and differentiation in neonatal germ cells, and acts both to stimulate spermatogonial proliferation and differentiation and to inhibit meiotic initiation in adult spermatogonia (Fahrioglu et al., 2007; Krentz et al., 2009; Matson et al., 2010). *Dmrt7* is expressed in primary spermatocytes, preferentially associates with the XY body and is required for formation of postmeiotic sex chromatin and passage through diplonema (Kim et al., 2007a). Here, we have found that *Dmrt6* acts between *Dmrt1* and *Dmrt7*, controlling the transition between the phases of gametogenesis that the other two genes regulate. Thus, *Dmrt* genes participate in a regulatory 'relay' that helps guide germ cells through much of their development. Our observation that DMRT6 acts in spermatogonia to activate transcription of genes that will later be required in meiosis suggests that DMRT6 serves not only to prevent the inappropriate mixed expression of genes from different stages of germ cell development (e.g. STRA8 with SYCP3; Fig. 4), but also helps establish, before meiosis, a 'dowry' of meiotic regulators that will allow successful synapsis and meiotic recombination after meiotic initiation.

DMRT6 is co-expressed with DMRT1 in differentiating spermatogonia, and several factors indicate an antagonistic relationship between the two: each gene affects expression of the other (this study and Murphy et al., 2010); DMRT6 binds *Dmrt1* (Fig. 7); most of the genomic sites bound by DMRT6 also are bound by DMRT1; and we found that DMRT1 and DMRT6 both regulate expression of several key spermatogonial regulators, including *Sohlh1* and *Sohlh2* (this study and Matson et al., 2010).

Examination of target gene expression suggests two ways in which DMRT1 and DMRT6 may functionally intersect. First, by repressing *Dmrt1* expression, DMRT6 can indirectly regulate some of its targets. This might be the case for *Sohlh1*, which shows strong promoter-proximal binding only for DMRT1 but the expression of which is regulated by both genes (Fig. 4). Second, the association of DMRT1 and DMRT6 at many shared sites suggests that they compete for binding or heterodimerize on DNA to somehow modify each other's activity (Murphy et al., 2010). Re-ChIP analysis showed that binding of DMRT1 and DMRT6 is not mutually exclusive; an attractive model is that DMRT1/DMRT6 heterodimers in late spermatogonia may repress genes such as *Sohlh1* that are activated by DMRT1 in early spermatogonia.

The transition from spermatogonial differentiation to meiosis is a crucial step in male gametogenesis and is incompletely

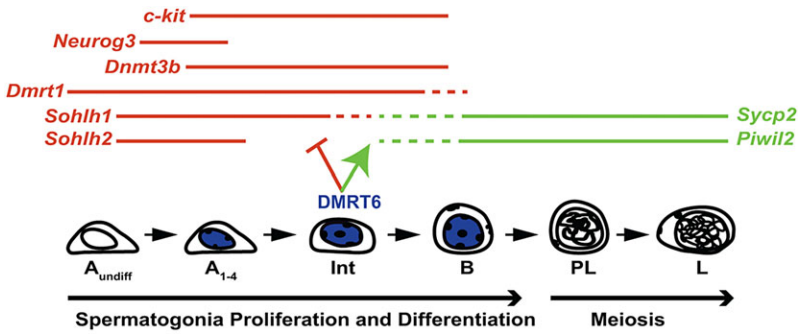

Fig. 8. Model of DMRT6 regulation of mammalian spermatogenesis. Based on mRNA and protein expression together with ChIP-seq analysis, a parsimonious working model is that DMRT6 acts in late spermatogonial cells to shut off spermatogonial differentiation through direct transcriptional repression of key spermatogonial differentiation factors that include *Dmrt1*, *Sohlh1*, *Sohlh2*, *Kit*, *Neurog3* and *Dnmt3b* (red). DMRT6 also promotes meiosis by directly activating transcription of genes including *Sycp2* and *Piwil2* (green). DMRT6-positive cells are indicated in blue.

Development (2014) 141, 3662-3671 doi:10.1242/dev.113936

understood. It has been apparent for some time that RA and its target *Stra8* are likely to play key roles in stimulating the shift to meiosis but is less clear how the spermatogonial program is brought to a conclusion; indeed, DMRT6 is the first transcriptional regulator shown to be essential in B spermatogonia. We have found that DMRT6 helps to 'draw the curtain' on spermatogonial development and set the stage for meiosis. Important future directions will include testing whether DMRT6 also plays this role in other vertebrates, elucidating how the regulatory networks anchored by DMRT1 and DMRT6 and possibly other DMRT proteins functionally intersect, and determining how DMRT proteins activate and repress transcription in a context-dependent manner via the same consensus DNA elements to achieve different outcomes in germline gene regulation.

MATERIALS AND METHODS

Dmrt6 targeting vector construct

We isolated BAC clones containing *Dmrt6* from a 129/SvJ strain BAC library (Source Bioscience) and used these to construct the *Dmrt6* targeting vector by bacterial recombineering (Liu et al., 2003). The left and right homology arms were cloned into the backbone vector PL253 (Liu et al., 2003). The final targeting construct, DM6-TV4, is shown in supplementary material Fig. S2D. To generate *Dmrt6* mutant animals, the targeting vector DM6-TV4 was linearized with *Not*I and transfected into CJ7 ES cells by electroporation (Liu et al., 2003). Three homologous recombinant colonies were identified from 300 G418-resistant colonies by Southern blot hybridization using a DNA probe containing sequences upstream of exon 1 to screen genomic DNA digested with *Eco*RV. Homologous recombination was confirmed on both ends of the targeted region by Southern blot hybridization using probes generated by PCR using primers LKB016/LKB017 (5′ probe) and LKB053/LKB054 (3′ probe), listed below. Two targeted embryonic stem cell clones containing the floxed allele $Dmrt6^{neo}$ were injected into C57BL/6J blastocysts to generate chimeras. Chimeric males were bred with C57BL/6J females to generate heterozygotes carrying $Dmrt6^{neo}$. $Dmrt6^{+/neo}$ females were then bred with male *beta-actin-Cre* transgenic mice (Lewandoski et al., 1997) to delete the floxed sequences and to generate heterozygous $Dmrt6^{+/-}$ animals, which were interbred to generate homozygous $Dmrt6^{-/-}$ mutants. The $Dmrt6^{-}$ allele was used for all experiments in this study. Experimental protocols involving animals were approved by the University of Minnesota Institutional Animal Care and Use Committee.

Genotyping

For genotyping, tail-clip DNA was amplified for 35 cycles with an annealing temperature of 56°C. The *Dmrt6* wild-type and floxed alleles were detected by PCR with DM6F/DM6R1, resulting in 480 bp or 424 bp amplicons, respectively. The *Dmrt6* null allele was detected by PCR with DM6F/DM6R2, resulting in a 309 bp amplicon.

Primers

Primers used for genotyping of ES cells and mice were: LKB016, 5′-CACACACACACGTGTTCA-3′; LKB017, 5′-ATTCCAAGCAGGGATGTGAC-3′; LKA053, 5′-TAACGCTCTGCTAGGGGAAA-3′; LKA054, 5′-GT-GGTGCGTAAGATCCCTGT-3′; DM6F, 5′-TTCATTGGGAGTTTTAAAAC-3′; DM6R1, 5′-CCATGGTTCCGGCACCTAGA-3′; DM6R2, 5′-AC-CATCAGAGCCAGCAAGATGGCTAAAGGCA-3′. Primers used for construction of *Dmrt6* targeting vector by BAC recombineering (Liu et al., 2003) were: A, 5′-GCGGCCGCTAACTTCACTTAGATTCTAC-3′; B, 5′-AAGCTTGGCAAGATCACGGTCTTTTA-3′; C, 5′-AAGCTTTGGTTTCTGCAATGCCTGTA-3′; D, 5′-GAATTCGATATCGTTAACTGTTCCCGCCAAATTGTAAC-3′; E, 5′-GGATCCATTGCAGTTCTCCGGACATT-3′; F, 5′-GCGGCCGCCACAGATGCACTGCTTCCAG-3′; G, 5′-CTCGAGGAGTCTGAGTCGAGGGGATG-3′; H, 5′-GAATTCGAGGCAGGTGGATTTCTGAG-3′; I, 5′-GGATCCGATATCGGAAACCGGTTCTTTCCTTC-3′; J, 5′-GCGGCCGCTAACCCCACTGTCCCTGAAC-3′; Y, 5′-AAGCTTTTTTTAACTTTTAAAAACCA-3′; Z, 5′-ACTAGTCAAAAGAAAGCCCAAGAATT-3′.

DMRT6 antibody

Rabbit polyclonal antibodies to DMRT6 were raised against a purified DMRT6 fusion protein containing glutathione-S-transferase (GST) fused to the C-terminal 148 amino acids of DMRT6. Antibodies to GST were removed by GST-affigel 10 chromatography and the antiserum was then purified by GST-DMRT6-affigel 10 chromatography. For IF, DMRT6 antibody was used at 1:200 dilution with a goat anti-rabbit secondary antibody (Abcam) at 1:500 dilution.

Histological analysis

Dissected testes were fixed 4% PFA overnight at 4°C, progressively dehydrated in graded ethanol series and embedded in paraffin wax. Sections of 5 µm were deparaffinized, rehydrated and stained with hematoxylin and eosin.

Tissue immunofluorescent staining

Slides with paraffin sections were rehydrated and boiled with 10 mM of citric acid (pH 6.0). Slides were blocked with 10% serum (goat or donkey depending on the secondary antibody used) in PBS with 0.1% Triton X-100 at room temperature for 1 h and incubated with primary antibody overnight at room temperature followed by 2 h incubation with secondary antibody. Nuclei were stained with 4′,6-diamidino-2-phenylindole (DAPI). All images were captured with a Zeiss Imager Z1 microscope using a Zeiss MRm camera. Antibodies, suppliers, catalog numbers and dilutions used are listed in supplementary material Table S4.

BrdU incorporation

BrdU (Sigma) was dissolved in phosphate-buffered saline (PBS) and injected intraperitoneally at 50 mg per kg of body weight. Testes were harvested from 4% PFA perfusion-fixed animals at 2 h following BrdU injection. To detect BrdU incorporation, testis sections were stained with an anti-BrdU antibody listed in supplementary material Table S1.

TUNEL staining

Slides with paraffin sections were rehydrated and apoptotic cells were detected using DeadEnd Fluorometric TUNEL System (Promega).

Meiotic spreads

Meiotic chromosome spreads were prepared from testes of 4-week-old mice as described previously (Peters et al., 1997).

RNA-Seq

RNA was isolated from testes of three independent wild-type and three $Dmrt6^{-/-}$ P8 animals using TRIzol reagent (Life Technologies), further purified with the RNeasy MinElute cleanup kit (74204, Qiagen), and RNA quality was verified with Qubit 2.0 (Invitrogen). TruSeq RNA libraries (Illumina) were prepared for sequencing on the Illumina HiSeq 2000 platform.

ChIP-Seq

Chromatin from testes of three adult wild-type B6 and 129Sv mixed genetic background mice was cross-linked with formaldehyde, sheared and immunoprecipitated with anti-DMRT6 antibody as described previously (Krentz et al., 2013).

Re-ChIP analysis

First-round ChIP was performed as for ChIP-Seq. After elution from the first round using 1% sodium dodecyl sulfate (SDS), 100 mM $NaHCO_3$, complexes were diluted 1:11 with DOC RIPA (Murphy et al., 2010). Diluted complexes were subjected to a second round of ChIP using magnetic beads (Dynabeads M-280 Tosylactivated, Invitrogen) covalently attached to either affinity purified anti-DMRT1 antibody or protein A purified pre-immune serum. Re-ChIP products were evaluated by qPCR and enrichment reported as percent total input.

Bioinformatics analysis

Briefly, RNA-Seq reads were mapped to the mm9 mouse genome assembly and differential gene expression was determined by counting reads within

Development (2014) 141, 3662-3671 doi:10.1242/dev.113936

genes defined by the Ensembl release 67 using published protocols (Anders et al., 2013). To identify potentially relevant genes we annotated the gene list based on expression in published microarray data [GSE4193 (Namekawa et al., 2006)]. We also annotated this list to identify genes associated with Ingenuity functional terms related to testis development and meiosis (Development of Genital Organ, Gamet*, Germ Cell, Gonad, Meiosis, Seminiferous, Seminal, Sperm* and Testis). The annotated gene list (supplementary material Table S1) was filtered based on expression in type A or type B spermatagonia and a match to one or more Ingenuity function terms. This filtered gene list (supplementary material Table S2) was further annotated with the number of PubMed publications for each gene and the keyword 'testis' and whether that gene was associated with a peak in the DMRT1 and DMRT6 ChIP-Seq data. ChIP-Seq data were mapped to mm9 and peaks were identified using MACS (Feng et al., 2012) using a *P*-value cutoff of 10^{-5} as described previously (Krentz et al., 2013). ChIP-Seq peaks were annotated with overlapping and nearest start features (supplementary material Table S3). Details can be found in the DMRT6 integrative analysis in the supplementary material. Our analysis data have been deposited at the NCBI Gene Expression Omnibus (GEO) under accession number GSE60440.

Acknowledgements

We thank Drs Hiromitsu Tanaka and Aleksander Radjkovic for antibodies, Drs Clinton Matson and Anthony Krentz for much helpful advice, Connie Corcoran for technical assistance and Dr Satoshi Namekawa for antibodies, technical advice and helpful discussions.

Competing interests

The authors declare no competing financial interests.

Author contributions

T.Z, M.W.M. and M.D.G. performed the experiments; all authors helped design experiments and interpret data; T.Z., V.J.B. and D.Z. wrote the paper.

Funding

This work was supported by the National Institutes of Health [5 RO1 GM59152 and 5 T32 HD060536] and the Minnesota Medical Foundation. Deposited in PMC for release after 12 months.

Supplementary material

Supplementary material available online at
http://dev.biologists.org/lookup/suppl/doi:10.1242/dev.113936/-/DC1

References

Anders, S., McCarthy, D. J., Chen, Y., Okoniewski, M., Smyth, G. K., Huber, W. and Robinson, M. D. (2013). Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nat. Protoc.* **8**, 1765-1786.

Anderson, E. L., Baltus, A. E., Roepers-Gajadien, H. L., Hassold, T. J., de Rooij, D. G., van Pelt, A. M. M. and Page, D. C. (2008). Stra8 and its inducer, retinoic acid, regulate meiotic initiation in both spermatogenesis and oogenesis in mice. *Proc. Natl. Acad. Sci. USA* **105**, 14976-14980.

Baker, S. M., Plug, A. W., Prolla, T. A., Bronner, C. E., Harris, A. C., Yao, X., Christie, D.-M., Monell, C., Arnheim, N., Bradley, A. et al. (1996). Involvement of mouse Mlh1 in DNA mismatch repair and meiotic crossing over. *Nat. Genet.* **13**, 336-342.

Ballow, D., Meistrich, M. L., Matzuk, M. and Rajkovic, A. (2006). Sohlh1 is essential for spermatogonial differentiation. *Dev. Biol.* **294**, 161-167.

Baudat, F. and de Massy, B. (2007). Regulating double-stranded DNA break repair towards crossover or non-crossover during mammalian meiosis. *Chromosome Res.* **15**, 565-577.

Buageaw, A., Sukhwani, M., Ben-Yehudah, A., Ehmcke, J., Rawe, V. Y., Pholpramool, C., Orwig, K. E. and Schlatt, S. (2005). GDNF family receptor alpha1 phenotype of spermatogonial stem cells in immature mouse testes. *Biol. Reprod.* **73**, 1011-1016.

Cohen, P. E., Pollack, S. E. and Pollard, J. W. (2006). Genetic analysis of chromosome pairing, recombination, and cell cycle control during first meiotic prophase in mammals. *Endocr. Rev.* **27**, 398-426.

Costoya, J. A., Hobbs, R. M., Barna, M., Cattoretti, G., Manova, K., Sukhwani, M., Orwig, K. E., Wolgemuth, D. J. and Pandolfi, P. P. (2004). Essential role of Plzf in maintenance of spermatogonial stem cells. *Nat. Genet.* **36**, 653-659.

Dada, R., Kumar, M., Jesudasan, R., Fernández, J. L., Gosálvez, J. and Agarwal, A. (2012). Epigenetics and its role in male infertility. *J. Assist. Reprod. Genet.* **29**, 213-223.

de Rooij, D. G. (1998). Stem cells in the testis. *Int. J. Exp. Pathol.* **79**, 67-80.

de Rooij, D. G. (2001). Proliferation and differentiation of spermatogonial stem cells. *Reproduction* **121**, 347-354.

de Rooij, D. G. and Grootegoed, J. A. (1998). Spermatogonial stem cells. *Curr. Opin. Cell Biol.* **10**, 694-701.

de Rooij, D. G. and Russell, L. D. (2000). All you wanted to know about spermatogonia but were afraid to ask. *J. Androl.* **21**, 776-798.

Di Giacomo, M., Comazzetto, S., Saini, H., De Fazio, S., Carrieri, C., Morgan, M., Vasiliauskaite, L., Benes, V., Enright, A. J. and O'Carroll, D. (2013). Multiple epigenetic mechanisms and the piRNA pathway enforce LINE1 silencing during adult spermatogenesis. *Mol. Cell* **50**, 601-608.

Erdman, S. E. and Burtis, K. C. (1993). The Drosophila doublesex proteins share a novel zinc finger related DNA binding domain. *EMBO J.* **12**, 527-535.

Fahrioglu, U., Murphy, M. W., Zarkower, D. and Bardwell, V. J. (2007). mRNA expression analysis and the molecular basis of neonatal testis defects in *Dmrt1* mutant mice. *Sex Dev.* **1**, 42-58.

Feng, J., Liu, T., Qin, B., Zhang, Y. and Liu, X. S. (2012). Identifying ChIP-seq enrichment using MACS. *Nat. Protoc.* **7**, 1728-1740.

Fraune, J., Schramm, S., Alsheimer, M. and Benavente, R. (2012). The mammalian synaptonemal complex: protein components, assembly and role in meiotic recombination. *Exp. Cell Res.* **318**, 1340-1346.

Griswold, M. D. (1998). The central role of Sertoli cells in spermatogenesis. *Semin. Cell Dev. Biol.* **9**, 411-416.

Griswold, M. D., Bishop, P. D., Kim, K. H., Ping, R., Siiteri, J. E. and Morales, C. (1989). Function of vitamin A in normal and synchronized seminiferous tubules. *Ann. N. Y. Acad. Sci.* **564**, 154-172.

Grivna, S. T., Pyhtila, B. and Lin, H. (2006). MIWI associates with translational machinery and PIWI-interacting RNAs (piRNAs) in regulating spermatogenesis. *Proc. Natl. Acad. Sci. USA* **103**, 13415-13420.

Hao, J., Yamamoto, M., Richardson, T. E., Chapman, K. M., Denard, B. S., Hammer, R. E., Zhao, G. Q. and Hamra, F. K. (2008). Sohlh2 knockout mice are male-sterile because of degeneration of differentiating type A spermatogonia. *Stem Cells* **26**, 1587-1597.

Hoffmann, E. R. and Borts, R. H. (2004). Meiotic recombination intermediates and mismatch repair proteins. *Cytogenet. Genome Res.* **107**, 232-248.

Hogarth, C. A. and Griswold, M. D. (2010). The key role of vitamin A in spermatogenesis. *J. Clin. Invest.* **120**, 956-962.

Jan, S. Z., Hamer, G., Repping, S., de Rooij, D. G., van Pelt, A. M. M. and Vormer, T. L. (2012). Molecular control of rodent spermatogenesis. *Biochim. Biophys. Acta* **1822**, 1838-1850.

Kim, S., Kettlewell, J. R., Anderson, R. C., Bardwell, V. J. and Zarkower, D. (2003). Sexually dimorphic expression of multiple doublesex-related genes in the embryonic mouse gonad. *Gene Expr. Patterns* **3**, 77-82.

Kim, S., Bardwell, V. J. and Zarkower, D. (2007a). Cell type-autonomous and non-autonomous requirements for Dmrt1 in postnatal testis differentiation. *Dev. Biol.* **307**, 314-327.

Kim, S., Namekawa, S. H., Niswander, L. M., Ward, J. O., Lee, J. T., Bardwell, V. J. and Zarkower, D. (2007b). A mammal-specific Doublesex homolog associates with male sex chromatin and is required for male meiosis. *PLoS Genet.* **3**, e62.

Koshimizu, U., Nishioka, H., Watanabe, D., Dohmae, K. and Nishimune, Y. (1995). Characterization of a novel spermatogenic cell antigen specific for early stages of germ cells in mouse testis. *Mol. Reprod. Dev.* **40**, 221-227.

Krentz, A. D., Murphy, M. W., Kim, S., Cook, M. S., Capel, B., Zhu, R., Matin, A., Sarver, A. L., Parker, K. L., Griswold, M. D. et al. (2009). The DM domain protein DMRT1 is a dose-sensitive regulator of fetal germ cell proliferation and pluripotency. *Proc. Natl. Acad. Sci. USA* **106**, 22323-22328.

Krentz, A. D., Murphy, M. W., Zhang, T., Sarver, A. L., Jain, S., Griswold, M. D., Bardwell, V. J. and Zarkower, D. (2013). Interaction between DMRT1 function and genetic background modulates signaling and pluripotency to control tumor susceptibility in the fetal germ line. *Dev. Biol.* **377**, 67–78.

Kuramochi-Miyagawa, S., Kimura, T., Ijiri, T. W., Isobe, T., Asada, N., Fujita, Y., Ikawa, M., Iwai, N., Okabe, M., Deng, W. et al. (2004). Mili, a mammalian member of piwi family gene, is essential for spermatogenesis. *Development* **131**, 839-849.

La Salle, S., Sun, F., Zhang, X.-D., Matunis, M. J. and Handel, M. A. (2008). Developmental control of sumoylation pathway proteins in mouse male germ cells. *Dev. Biol.* **321**, 227-237.

Lewandoski, M., Meyers, E. N. and Martin, G. R. (1997). Analysis of Fgf8 gene function in vertebrate development. *Cold Spring Harb. Symp. Quant. Biol.* **62**, 159-168.

Liu, P., Jenkins, N. A. and Copeland, N. G. (2003). A highly efficient recombineering-based method for generating conditional knockout mutations. *Genome Res.* **13**, 476-484.

Mark, M., Jacobs, H., Oulad-Abdelghani, M., Dennefeld, C., Féret, B., Vernet, N., Codreanu, C.-A., Chambon, P. and Ghyselinck, N. B. (2008). STRA8-deficient spermatocytes initiate, but fail to complete, meiosis and undergo premature chromosome condensation. *J. Cell Sci.* **121**, 3233-3242.

Matson, C. K. and Zarkower, D. (2012). Sex and the singular DM domain: insights into sexual regulation, evolution and plasticity. *Nat. Rev. Genet.* **13**, 163-174.

Matson, C. K., Murphy, M. W., Griswold, M. D., Yoshida, S., Bardwell, V. J. and Zarkower, D. (2010). The mammalian doublesex homolog DMRT1 is a transcriptional gatekeeper that controls the mitosis versus meiosis decision in male germ cells. *Dev. Cell* **19**, 612-624.
McCarrey, J. R. and Skinner, M. K. (1999). Construction and preliminary characterization of a series of mouse and rat testis cDNA libraries. *J. Androl.* **20**, 635-639.
McCarthy, P. T. and Cerecedo, L. R. (1952). Vitamin A deficiency in the mouse two figures. *J. Nutr.* **46**, 361-376.
Meng, X., Lindahl, M., Hyvönen, M. E., Parvinen, M., de Rooij, D. G., Hess, M. W., Raatikainen-Ahokas, A., Sainio, K., Rauvala, H., Lakso, M. et al. (2000). Regulation of cell fate decision of undifferentiated spermatogonia by GDNF. *Science* **287**, 1489-1493.
Murphy, M. W., Sarver, A. L., Rice, D., Hatzi, K., Ye, K., Melnick, A., Heckert, L. L., Zarkower, D. and Bardwell, V. J. (2010). Genome-wide analysis of DNA binding and transcriptional regulation by the mammalian Doublesex homolog DMRT1 in the juvenile testis. *Proc. Natl. Acad. Sci. USA* **107**, 13360-13365.
Nakagawa, T., Sharma, M., Nabeshima, Y.-I., Braun, R. E. and Yoshida, S. (2010). Functional hierarchy and reversibility within the murine spermatogenic stem cell compartment. *Science* **328**, 62-67.
Namekawa, S. H., Park, P. J., Zhang, L.-F., Shima, J. E., McCarrey, J. R., Griswold, M. D. and Lee, J. T. (2006). Postmeiotic sex chromatin in the male germline of mice. *Curr. Biol.* **16**, 660-667.
Naughton, C. K., Jain, S., Strickland, A. M., Gupta, A. and Milbrandt, J. (2006). Glial cell-line derived neurotrophic factor-mediated RET signaling regulates spermatogonial stem cell fate. *Biol. Reprod.* **74**, 314-321.
Ottolenghi, C., Fellous, M., Barbieri, M. and McElreavey, K. (2002). Novel paralogy relations among human chromosomes support a link between the phylogeny of doublesex-related genes and the evolution of sex determination. *Genomics* **79**, 333-343.
Oulad-Abdelghani, M., Bouillet, P., Decimo, D., Gansmuller, A., Heyberger, S., Dolle, P., Bronner, S., Lutz, Y. and Chambon, P. (1996). Characterization of a premeiotic germ cell-specific cytoplasmic protein encoded by Stra8, a novel retinoic acid-responsive gene. *J. Cell Biol.* **135**, 469-477.
Peters, A. H. F. M., Plug, A. W., van Vugt, M. J. and de Boer, P. (1997). A drying-down technique for the spreading of mammalian meiocytes from the male and female germline. *Chromosome Res.* **5**, 66-68.
Raverot, G., Weiss, J., Park, S. Y., Hurley, L. and Jameson, J. L. (2005). Sox3 expression in undifferentiated spermatogonia is required for the progression of spermatogenesis. *Dev. Biol.* **283**, 215-225.
Raymond, C. S., Murphy, M. W., O'Sullivan, M. G., Bardwell, V. J. and Zarkower, D. (2000). Dmrt1, a gene related to worm and fly sexual regulators, is required for mammalian testis differentiation. *Genes Dev.* **14**, 2587-2595.
Schrans-Stassen, B. H. G. J., van de Kant, H. J. G., de Rooij, D. G. and van Pelt, A. M. M. (1999). Differential expression of c-kit in mouse undifferentiated and differentiating type A spermatogonia. *Endocrinology* **140**, 5894-5900.
Shima, J. E., McLean, D. J., McCarrey, J. R. and Griswold, M. D. (2004). The murine testicular transcriptome: characterizing gene expression in the testis during the progression of spermatogenesis. *Biol. Reprod.* **71**, 319-330.
Shirakawa, T., Yaman-Deveci, R., Tomizawa, S.-i., Kamizato, Y., Nakajima, K., Sone, H., Sato, Y., Sharif, J., Yamashita, A. and Takada-Horisawa, Y. (2013). An epigenetic switch is crucial for spermatogonia to exit the undifferentiated state toward a Kit-positive identity. *Development* **140**, 3565-3576.
Snyder, E. M., Small, C. and Griswold, M. D. (2010). Retinoic acid availability drives the asynchronous initiation of spermatogonial differentiation in the mouse. *Biol. Reprod.* **83**, 783-790.
Su, A. I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K. A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G. et al. (2004). A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl. Acad. Sci. USA* **101**, 6062-6067.
Suzuki, H., Ahn, H. W., Chu, T., Bowden, W., Gassei, K., Orwig, K. and Rajkovic, A. (2012). SOHLH1 and SOHLH2 coordinate spermatogonial differentiation. *Dev. Biol.* **361**, 301-312.
Thompson, J. N., Howell, J. M. and Pitt, G. A. J. (1964). Vitamin A and reproduction in rats. *Proc. R. Soc. Lond. B Biol. Sci.* **159**, 510-535.
Turner, J. M. A., Aprelikova, O., Xu, X., Wang, R., Kim, S., Chandramouli, G. V. R., Barrett, J. C., Burgoyne, P. S. and Deng, C.-X. (2004). BRCA1, histone H2AX phosphorylation, and male meiotic sex chromosome inactivation. *Curr. Biol.* **14**, 2135-2142.
Van Pelt, A. M. and de Rooij, D. G. (1990). Synchronization of the seminiferous epithelium after vitamin A replacement in vitamin A-deficient mice. *Biol. Reprod.* **43**, 363-367.
Vourekas, A., Zheng, Q., Alexiou, P., Maragkakis, M., Kirino, Y., Gregory, B. D. and Mourelatos, Z. (2012). Mili and Miwi target RNA repertoire reveals piRNA biogenesis and function of Miwi in spermiogenesis. *Nat. Struct. Mol. Biol.* **19**, 773-781.
Yang, F. and Wang, P. (2009). The mammalian synaptonemal complex: a scaffold and beyond. *Genome Dyn.* **5**, 69-80.
Yoshida, S., Takakura, A., Ohbo, K., Abe, K., Wakabayashi, J., Yamamoto, M., Suda, T. and Nabeshima, Y.-i. (2004). Neurogenin3 delineates the earliest stages of spermatogenesis in the mouse testis. *Dev. Biol.* **269**, 447-458.
Yuan, L., Liu, J.-G., Zhao, J., Brundell, E., Daneholt, B. and Höög, C. (2000). The murine SCP3 gene is required for synaptonemal complex assembly, chromosome synapsis, and male fertility. *Mol. Cell* **5**, 73-83.
Zhu, L., Wilken, J., Phillips, N. B., Narendra, U., Chan, G., Stratton, S. M., Kent, S. B. and Weiss, M. A. (2000). Sexual dimorphism in diverse metazoans is regulated by a novel class of intertwined zinc fingers. *Genes Dev.* **14**, 1750-1764.

DMRT6 Integrative Analysis

Source code for this analysis is available at the UMN github repository: https://github.umn.edu/gearh006/umn-gcd-bioinformatics-dmrt6

Process raw data using STAR, FASTQC, PICARD, SAMTOOLS and IGVTOOLS (Minnesota Supercomputing Institute)

```
dd=/home/zarkowe0/data_release/umgc/hiseq/131125_SN261_0527_AC3540ACXX/Project_Zarkower_Project_013
wd=/home/bardwell/gearhart/dmrt6/
org=mm9

for i in 1663_1_DM6_WT_ATCACG 1663_3_DM6_WT_TTAGGC \
1663_5_DM6_Null_ACAGTG 1663_2_DM6_Null_CGATGT 1663_4_DM6_WT_TGACCA 1665_2_DM6_Null_GCCAAT

#i="£{file%.*}"

do

sf1="${i}_L005_R1_001.fastq"
sf2="${i}_L005_R2_001.fastq"

cat << EOF > $i.star.pbs
#PBS -l mem=32000mb,nodes=1:ppn=4,walltime=10:00:00
#PBS -m a
#PBS -M gearh006@umn.edu
#PBS -q lab
mkdir $wd/$i
cd $wd/$i
/home/bardwell/shared/STAR_2.3.0e/STAR --genomeDir /home/bardwell/shared/STAR_GENOME/$org/ \
--runThreadN 8 --readFilesIn $dd/$sf1 $dd/$sf2

qsub $wd/$i.igv.pbs

EOF

cat << EOF > $i.igv.pbs
#PBS -l mem=8000mb,nodes=1:ppn=1,walltime=08:00:00
#PBS -m a
#PBS -M gearh006@umn.edu
#PBS -q lab
module load samtools

cd $wd

/home/bardwell/shared/FastQC/fastqc -o fastqc $dd/$sf1
/home/bardwell/shared/FastQC/fastqc -o fastqc $dd/$sf2

cd $wd/$i
#convert sam to bam
samtools view -bS -o $i.raw.bam Aligned.out.sam

#sort the bam file
samtools sort $i.raw.bam $i.sort

#remove duplicates
java -Xmx2g -jar /home/bardwell/shared/picard-tools-1.94/MarkDuplicates.jar INPUT=$i.sort.bam OUTPUT=$i.bam RE

#create the index file
samtools index $i.bam
```

```
#igvtools to make a TDF File
java -Xmx2g  -jar /home/bardwell/shared/IGVTools_2/igvtools.jar count -z 5 -w 25 -e 100 $i.bam $i.tdf \
/home/bardwell/shared/IGVTools_2/genomes/$org.genome

rm $i.sort.bam
rm $i.raw.bam

mv $i.bam $wd/
mv $i.bam.bai $wd/
mv $i.tdf $wd/
EOF

qsub $i.star.pbs

done
```

Analyse Reads for differential expression with EdgeR (RNA-SEQ mm9 version)

```
library(Rsamtools)
```

```
## Loading required package: IRanges
## Loading required package: BiocGenerics
## Loading required package: parallel
##
## Attaching package: 'BiocGenerics'
##
## The following objects are masked from 'package:parallel':
##
##     clusterApply, clusterApplyLB, clusterCall,
##     clusterEvalQ, clusterExport, clusterMap,
##     parApply, parCapply, parLapply, parLapplyLB,
##     parRapply, parSapply, parSapplyLB
##
## The following object is masked from 'package:stats':
##
##     xtabs
##
## The following objects are masked from 'package:base':
##
##     Filter, Find, Map, Position, Reduce,
##     anyDuplicated, append, as.data.frame, as.vector,
##     cbind, colnames, do.call, duplicated, eval,
##     evalq, get, intersect, is.unsorted, lapply,
##     mapply, match, mget, order, paste, pmax,
##     pmax.int, pmin, pmin.int, rank, rbind, rep.int,
##     rownames, sapply, setdiff, sort, table, tapply,
##     union, unique, unlist
##
## Loading required package: GenomicRanges
## Loading required package: GenomeInfoDb
## Loading required package: XVector
## Loading required package: Biostrings
```

```
library(GenomicFeatures)
```

```
## Loading required package: AnnotationDbi
## Loading required package: Biobase
## Welcome to Bioconductor
##
```

```
##     Vignettes contain introductory material; view
##     with 'browseVignettes()'. To cite Bioconductor,
##     see 'citation("Biobase")', and for packages
##     'citation("pkgname")'.

library(GenomicRanges)
library(GenomicAlignments)

## Loading required package: BSgenome
##
## Attaching package: 'BSgenome'
##
## The following object is masked from 'package:AnnotationDbi':
##
##     species
##
##
## Attaching package: 'GenomicAlignments'
##
## The following object is masked _by_ '.GlobalEnv':
##
##     last

library(edgeR)

## Loading required package: limma
##
## Attaching package: 'limma'
##
## The following object is masked from 'package:BiocGenerics':
##
##     plotMA

library(qvalue)

# For transcriptDB and annotations
library(biomaRt)

# For Pubmed Lookups
library(XML)

# For microarray
library(GEOquery)

## Setting options('download.file.method.GEOquery'='curl')

library(Biobase)

# For Chip Analysis
library(rtracklayer)
library(ChIPpeakAnno)

## Loading required package: grid
## Loading required package: VennDiagram
## Loading required package: DBI

data(TSS.mouse.NCBIM37)
library(org.Mm.eg.db)
```

```
##
```

```
library(GOstats)
```

```
## Loading required package: Category
## Loading required package: Matrix
##
## Attaching package: 'Matrix'
##
## The following object is masked from 'package:IRanges':
##
##     expand
##
## Loading required package: GO.db
## Loading required package: graph
##
## Attaching package: 'graph'
##
## The following object is masked from 'package:XML':
##
##     addNode
##
## The following object is masked from 'package:Biostrings':
##
##     complement
##
##
## Attaching package: 'GOstats'
##
## The following object is masked from 'package:AnnotationDbi':
##
##     makeGOGraph
```

```
library("GO.db")
```

```
# For Motif Analysis
library(BSgenome.Mmusculus.UCSC.mm9)
library(rGADEM)
```

```
## Loading required package: seqLogo
```

```
library(motifStack)
```

```
## Loading required package: grImport
## Loading required package: MotIV
##
## Attaching package: 'MotIV'
##
## The following object is masked from 'package:rGADEM':
##
##     readPWMfile
##
## The following object is masked from 'package:seqLogo':
##
##     makePWM
##
## The following object is masked from 'package:stats':
##
```

```
##     filter
##
## Loading required package: ade4
##
## Attaching package: 'ade4'
##
## The following object is masked from 'package:rtracklayer':
##
##     score
##
## The following object is masked from 'package:BSgenome':
##
##     score
##
## The following object is masked from 'package:Biostrings':
##
##     score
##
## The following object is masked from 'package:GenomicRanges':
##
##     score
##
## The following object is masked from 'package:IRanges':
##
##     score
```

This section uses a package called biomaRt to download data from Ensembl. We will get a list of all the Ensembl genes in the genome and some annotation information for these genes. Since our data is mapped to mm9 we will use the May 2012 archive of Ensembl (their current release is based on mm10). Ensembl chromosomes are numbered 1-19,X,Y whereas our bam files are references as chr1-chr19,chrX,chrY so we have to do a quick swtich of the chromosome names to use Ensembl genes on USCS mapped data.

```
# use may2012 archive to get mm9 NCBIM37 build (Ensembl
# Release 67)
ensembl = useMart(host = "may2012.archive.ensembl.org", biomart = "ENSEMBL_MART_ENSEMBL",
    dataset = "mmusculus_gene_ensembl")
# ensembl=useMart(biomart='ensembl',dataset='mmusculus_gene_ensembl')
mme <- makeTranscriptDbFromBiomart(host = "may2012.archive.ensembl.org",
    biomart = "ENSEMBL_MART_ENSEMBL", dataset = "mmusculus_gene_ensembl")
exonsByGene <- exonsBy(mme, by = "gene")
chroms <- seqlevels(mme)
chroms[1:21]

# oldSeqLevelsToKeep
oldSeqLevelsToKeep <- as.character(chroms[1:21])
str(oldSeqLevelsToKeep)
oldSeqLevelsToKeep

# Create a named character vector to use hg19 chromosome
# names
chromRename <- paste("chr", as.character(chroms[1:21]), sep = "")
names(chromRename) <- as.character(chroms[1:21])
str(chromRename)
chromRename

exonsByGene[1000:1000]
exonsByGene <- keepSeqlevels(exonsByGene, oldSeqLevelsToKeep)
exonsByGene[1000:1000]
exonsByGene <- renameSeqlevels(exonsByGene, chromRename)
exonsByGene[1000:1000]
```

```
save(exonsByGene, file = "exonsByGene_mm9_biomart_ensembl.rdata")
```

This chunck counts all the reads in the data. Can take a long time so better to do it on the server.

```
#PBS -l mem=32gb,nodes=1:ppn=1,walltime=2:00:00
#PBS -m a
#PBS -M gearh006@umn.edu
#PBS -q lab

cd /home/bardwell/gearhart/dmrt6/

cat << EOF > summarizeOverlaps.r

library(Rsamtools)
load("exonsByGene_mm9_biomart_ensembl.rdata")

fls <- list.files("/home/bardwell/gearhart/dmrt6", pattern="bam$",full=TRUE)
bamlst <- BamFileList(fls)
genehits <- summarizeOverlaps(exonsByGene, bamlst, mode="Union",
                              singleEnd=TRUE, ignore.strand=TRUE)
save(genehits,file= "120313_DMRT6_counts_mm9_biomart_chrRN_ensembl.rdata")
quit(save="no")

EOF

/panfs/roc/groups/10/bardwell/shared/R/R-3.0.1/bin/R --no-save < summarizeOverlaps.r
```

Once this is done, you can just reload in the counts whch are saved in the genehits variable in this file. This section removes all the genes that are not expressed (Total Reads across all samples < 10)

```
load("/mnt/afp/teng/data/120313_DMRT6_counts_mm9_biomart_chrRN_ensembl.rdata")
str(genehits)

## Formal class 'SummarizedExperiment' [package "GenomicRanges"] with 4 slots
##   ..@ exptData:Formal class 'SimpleList' [package "IRanges"] with 4 slots
##   .. .. ..@ listData       : list()
##   .. .. ..@ elementType    : chr "ANY"
##   .. .. ..@ elementMetadata: NULL
##   .. .. ..@ metadata       : list()
##   ..@ rowData :Formal class 'GRangesList' [package "GenomicRanges"] with 5 slots
##   .. .. ..@ unlistData     :Formal class 'GRanges' [package "GenomicRanges"] with 6 slots
##   .. .. .. .. ..@ seqnames       :Formal class 'Rle' [package "IRanges"] with 4 slots
##   .. .. .. .. .. .. ..@ values         : Factor w/ 21 levels "chr1","chr2",..: 3 20 16 7 20 11 6 13 4 9 ...
##   .. .. .. .. .. .. ..@ lengths        : int [1:25890] 9 9 24 15 56 32 7 4 51 5 ...
##   .. .. .. .. .. .. ..@ elementMetadata: NULL
##   .. .. .. .. .. .. ..@ metadata       : list()
##   .. .. .. .. ..@ ranges         :Formal class 'IRanges' [package "IRanges"] with 6 slots
##   .. .. .. .. .. .. ..@ start          : int [1:415076] 107910198 107912321 107914853 107915391 107918681 1
##   .. .. .. .. .. .. ..@ width          : int [1:415076] 2037 210 154 130 129 158 142 43 259 214 ...
##   .. .. .. .. .. .. ..@ NAMES          : NULL
##   .. .. .. .. .. .. ..@ elementType    : chr "integer"
##   .. .. .. .. .. .. ..@ elementMetadata: NULL
##   .. .. .. .. .. .. ..@ metadata       : list()
##   .. .. .. .. ..@ strand         :Formal class 'Rle' [package "IRanges"] with 4 slots
##   .. .. .. .. .. .. ..@ values         : Factor w/ 3 levels "+","-","*": 2 1 2 1 2 1 2 1 2 1 ...
##   .. .. .. .. .. .. ..@ lengths        : int [1:18288] 57 175 15 37 95 115 7 32 28 47 ...
##   .. .. .. .. .. .. ..@ elementMetadata: NULL
##   .. .. .. .. .. .. ..@ metadata       : list()
```

```
##   .. .. .. .. ..@ elementMetadata:Formal class 'DataFrame' [package "IRanges"] with 6 slots
##   .. .. .. .. .. .. ..@ rownames       : NULL
##   .. .. .. .. .. .. ..@ nrows          : int 415076
##   .. .. .. .. .. .. ..@ listData       :List of 2
##   .. .. .. .. .. .. .. ..$ exon_id  : int [1:415076] 82094 82095 82096 82097 82098 82099 82100 82101 82102 
##   .. .. .. .. .. .. .. ..$ exon_name: chr [1:415076] "ENSMUSE00000363317" "ENSMUSE00000404895" "ENSMUSE0000
##   .. .. .. .. .. .. ..@ elementType    : chr "ANY"
##   .. .. .. .. .. .. ..@ elementMetadata: NULL
##   .. .. .. .. .. .. ..@ metadata       : list()
##   .. .. .. .. ..@ seqinfo        :Formal class 'Seqinfo' [package "GenomicRanges"] with 4 slots
##   .. .. .. .. .. .. ..@ seqnames   : chr [1:21] "chr1" "chr2" "chr3" "chr4" ...
##   .. .. .. .. .. .. ..@ seqlengths : int [1:21] NA NA NA NA NA NA NA NA NA NA ...
##   .. .. .. .. .. .. ..@ is_circular: logi [1:21] NA NA NA NA NA NA ...
##   .. .. .. .. .. .. ..@ genome     : chr [1:21] NA NA NA NA ...
##   .. .. .. .. ..@ metadata       : list()
##   .. .. ..@ elementMetadata:Formal class 'DataFrame' [package "IRanges"] with 6 slots
##   .. .. .. .. ..@ rownames       : NULL
##   .. .. .. .. ..@ nrows          : int 37583
##   .. .. .. .. ..@ listData       : Named list()
##   .. .. .. .. ..@ elementType    : chr "ANY"
##   .. .. .. .. ..@ elementMetadata: NULL
##   .. .. .. .. ..@ metadata       : list()
##   .. .. ..@ partitioning   :Formal class 'PartitioningByEnd' [package "IRanges"] with 5 slots
##   .. .. .. .. ..@ end            : int [1:37583] 9 18 42 57 113 129 145 152 156 207 ...
##   .. .. .. .. ..@ NAMES          : chr [1:37583] "ENSMUSG00000000001" "ENSMUSG00000000003" "ENSMUSG0000000
##   .. .. .. .. ..@ elementType    : chr "integer"
##   .. .. .. .. ..@ elementMetadata: NULL
##   .. .. .. .. ..@ metadata       : list()
##   .. .. ..@ elementType    : chr "GRanges"
##   .. .. ..@ metadata       :List of 1
##   .. .. .. .. ..$ genomeInfo:List of 20
##   .. .. .. .. ..$ Db type                                    : chr "TranscriptDb"
##   .. .. .. .. ..$ Supporting package                         : chr "GenomicFeatures"
##   .. .. .. .. ..$ Data source                                : chr "BioMart"
##   .. .. .. .. ..$ Organism                                   : chr "Mus musculus"
##   .. .. .. .. ..$ Resource URL                               : chr "may2012.archive.ensembl.org:80"
##   .. .. .. .. ..$ BioMart database                           : chr "ENSEMBL_MART_ENSEMBL"
##   .. .. .. .. ..$ BioMart database version                   : chr "Ensembl Genes 67"
##   .. .. .. .. ..$ BioMart dataset                            : chr "mmusculus_gene_ensembl"
##   .. .. .. .. ..$ BioMart dataset description                : chr "Mus musculus genes (NCBIM37)"
##   .. .. .. .. ..$ BioMart dataset version                    : chr "NCBIM37"
##   .. .. .. .. ..$ Full dataset                               : chr "yes"
##   .. .. .. .. ..$ miRBase build ID                           : chr NA
##   .. .. .. .. ..$ transcript_nrow                            : chr "97639"
##   .. .. .. .. ..$ exon_nrow                                  : chr "416230"
##   .. .. .. .. ..$ cds_nrow                                   : chr "318339"
##   .. .. .. .. ..$ Db created by                              : chr "GenomicFeatures package from Bioconductor"
##   .. .. .. .. ..$ Creation time                              : chr "2013-09-16 22:47:51 -0500 (Mon, 16 Sep 201
##   .. .. .. .. ..$ GenomicFeatures version at creation time: chr "1.12.3"
##   .. .. .. .. ..$ RSQLite version at creation time        : chr "0.11.4"
##   .. .. .. .. ..$ DBSCHEMAVERSION                         : chr "1.0"
##   ..@ colData :Formal class 'DataFrame' [package "IRanges"] with 6 slots
##   .. .. ..@ rownames       : chr [1:6] "/home/bardwell/gearhart/dmrt6/1663_1_DM6_WT_ATCACG.bam" "/home/bard
##   .. .. ..@ nrows          : int 6
##   .. .. ..@ listData       :List of 1
##   .. .. .. ..$ fileName:Formal class 'BamFileList' [package "Rsamtools"] with 4 slots
##   .. .. .. .. .. ..@ listData       :List of 6
##   .. .. .. .. .. .. ..$ /home/bardwell/gearhart/dmrt6/1663_1_DM6_WT_ATCACG.bam  :Reference class 'BamFile'
##   .. .. .. .. .. .. .. ..$ .extptr  :<externalptr>
##   .. .. .. .. .. .. .. ..$ path     : chr "/home/bardwell/gearhart/dmrt6/1663_1_DM6_WT_ATCACG.bam"
```

```
##   .. .. .. .. .. .. .. ..$ index    : chr "/home/bardwell/gearhart/dmrt6/1663_1_DM6_WT_ATCACG.bam"
##   .. .. .. .. .. .. .. ..$ yieldSize: int NA
##   .. .. .. .. .. .. .. ..$ obeyQname: logi FALSE
##   .. .. .. .. .. .. .. ..and 12 methods,  .. .. .. .. .. .. ..$ /home/bardwell/gearhart/dmrt6/1663_2_DM6_Nu
##   .. .. .. .. .. .. .. ..$ .extptr  :<externalptr>
##   .. .. .. .. .. .. .. ..$ path     : chr "/home/bardwell/gearhart/dmrt6/1663_2_DM6_Null_CGATGT.bam"
##   .. .. .. .. .. .. .. ..$ index    : chr "/home/bardwell/gearhart/dmrt6/1663_2_DM6_Null_CGATGT.bam"
##   .. .. .. .. .. .. .. ..$ yieldSize: int NA
##   .. .. .. .. .. .. .. ..$ obeyQname: logi FALSE
##   .. .. .. .. .. .. .. ..and 12 methods,  .. .. .. .. .. .. ..$ /home/bardwell/gearhart/dmrt6/1663_3_DM6_WT
##   .. .. .. .. .. .. .. ..$ .extptr  :<externalptr>
##   .. .. .. .. .. .. .. ..$ path     : chr "/home/bardwell/gearhart/dmrt6/1663_3_DM6_WT_TTAGGC.bam"
##   .. .. .. .. .. .. .. ..$ index    : chr "/home/bardwell/gearhart/dmrt6/1663_3_DM6_WT_TTAGGC.bam"
##   .. .. .. .. .. .. .. ..$ yieldSize: int NA
##   .. .. .. .. .. .. .. ..$ obeyQname: logi FALSE
##   .. .. .. .. .. .. .. ..and 12 methods,  .. .. .. .. .. .. ..$ /home/bardwell/gearhart/dmrt6/1663_4_DM6_WT
##   .. .. .. .. .. .. .. ..$ .extptr  :<externalptr>
##   .. .. .. .. .. .. .. ..$ path     : chr "/home/bardwell/gearhart/dmrt6/1663_4_DM6_WT_TGACCA.bam"
##   .. .. .. .. .. .. .. ..$ index    : chr "/home/bardwell/gearhart/dmrt6/1663_4_DM6_WT_TGACCA.bam"
##   .. .. .. .. .. .. .. ..$ yieldSize: int NA
##   .. .. .. .. .. .. .. ..$ obeyQname: logi FALSE
##   .. .. .. .. .. .. .. ..and 12 methods,  .. .. .. .. .. .. ..$ /home/bardwell/gearhart/dmrt6/1663_5_DM6_Nu
##   .. .. .. .. .. .. .. ..$ .extptr  :<externalptr>
##   .. .. .. .. .. .. .. ..$ path     : chr "/home/bardwell/gearhart/dmrt6/1663_5_DM6_Null_ACAGTG.bam"
##   .. .. .. .. .. .. .. ..$ index    : chr "/home/bardwell/gearhart/dmrt6/1663_5_DM6_Null_ACAGTG.bam"
##   .. .. .. .. .. .. .. ..$ yieldSize: int NA
##   .. .. .. .. .. .. .. ..$ obeyQname: logi FALSE
##   .. .. .. .. .. .. .. ..and 12 methods,  .. .. .. .. .. .. ..$ /home/bardwell/gearhart/dmrt6/1665_2_DM6_Nu
##   .. .. .. .. .. .. .. ..$ .extptr  :<externalptr>
##   .. .. .. .. .. .. .. ..$ path     : chr "/home/bardwell/gearhart/dmrt6/1665_2_DM6_Null_GCCAAT.bam"
##   .. .. .. .. .. .. .. ..$ index    : chr "/home/bardwell/gearhart/dmrt6/1665_2_DM6_Null_GCCAAT.bam"
##   .. .. .. .. .. .. .. ..$ yieldSize: int NA
##   .. .. .. .. .. .. .. ..$ obeyQname: logi FALSE
##   .. .. .. .. .. .. .. ..and 12 methods,  .. .. .. .. .. ..@ elementType    : chr "BamFile"
##   .. .. .. .. .. .. ..@ elementMetadata: NULL
##   .. .. .. .. .. .. ..@ metadata       : list()
##   .. .. ..@ elementType    : chr "ANY"
##   .. .. ..@ elementMetadata: NULL
##   .. .. ..@ metadata       : list()
##   ..@ assays  :Reference class 'ShallowSimpleListAssays' [package "GenomicRanges"] with 1 fields
##   .. ..$ data:Formal class 'SimpleList' [package "IRanges"] with 4 slots
##   .. .. .. ..@ listData       :List of 1
##   .. .. .. .. ..$ counts: int [1:37583, 1:6] 7056 0 1443 10239 11435 2 1196 944 2018 684 ...
##   .. .. .. ..@ elementType    : chr "ANY"
##   .. .. .. ..@ elementMetadata: NULL
##   .. .. .. ..@ metadata       : list()
##   .. ..and 12 methods,
```

```
temp = assays(genehits)$counts
colnames(temp)
```

```
## [1] "/home/bardwell/gearhart/dmrt6/1663_1_DM6_WT_ATCACG.bam"
## [2] "/home/bardwell/gearhart/dmrt6/1663_2_DM6_Null_CGATGT.bam"
## [3] "/home/bardwell/gearhart/dmrt6/1663_3_DM6_WT_TTAGGC.bam"
## [4] "/home/bardwell/gearhart/dmrt6/1663_4_DM6_WT_TGACCA.bam"
## [5] "/home/bardwell/gearhart/dmrt6/1663_5_DM6_Null_ACAGTG.bam"
## [6] "/home/bardwell/gearhart/dmrt6/1665_2_DM6_Null_GCCAAT.bam"
```

```
colnames(temp) <- c("WT_R1", "Null_R1", "WT_R2", "WT_R3", "Null_R2",
    "Null_R3")
```

```
big10 = apply(temp, 1, sum) > 10
TotalReads = temp[big10, ]
nrow(TotalReads)
```

```
## [1] 22744
```

```
colnames(TotalReads)
```

```
## [1] "WT_R1"   "Null_R1" "WT_R2"   "WT_R3"   "Null_R2"
## [6] "Null_R3"
```

We will also use biomaRt to get annotations for all the mouse Ensembl genes. Namely we want EntrezIDs and MGI data and positions in the genome.

```
ensembl = useMart(host = "may2012.archive.ensembl.org", biomart = "ENSEMBL_MART_ENSEMBL",
    dataset = "mmusculus_gene_ensembl")
# filters = listFilters(ensembl) filters[1:100,] attributes =
# listAttributes(ensembl) attributes[1:100,]

myattributes <- c("ensembl_gene_id", "mgi_id", "mgi_symbol",
    "chromosome_name", "start_position", "end_position", "strand",
    "entrezgene")
# test on a few genes
annot = getBM(attributes = myattributes, filters = "ensembl_gene_id",
    values = c("ENSMUSG00000040363", "ENSMUSG00000017652"), mart = ensembl)
head(annot)
```

```
##      ensembl_gene_id      mgi_id mgi_symbol chromosome_name
## 1 ENSMUSG00000017652   MGI:88336       Cd40               2
## 2 ENSMUSG00000040363 MGI:1918708       Bcor               X
##   start_position end_position strand entrezgene
## 1      164881127    164898448      1      21939
## 2       11613866     11737481     -1      71458
```

Define a function to Extract Mouse Gene Names from Human Entrez IDs which we need for parsing Incomplete Ingenuity Data

```
ensemblHuman = useMart(biomart = "ensembl", dataset = "hsapiens_gene_ensembl")
# filtersHuman = listFilters(ensemblHuman)
# filtersHuman[grep('Entrez',filtersHuman[,2]),]
# filtersHuman[grep('with_homolog',filtersHuman[,1]),]
# filtersHuman[1:10,] attributesHuman =
# listAttributes(ensemblHuman)
# attributesHuman[grep('homolog_ensembl_gene',attributes[,1]),]
myattributesHuman <- c("ensembl_gene_id", "mmusculus_homolog_ensembl_gene")
getBM(attributes = myattributesHuman, filters = c("entrezgene",
    "with_homolog_mmus"), values = list(c("54880"), TRUE), mart = ensemblHuman)
```

```
##   ensembl_gene_id mmusculus_homolog_ensembl_gene
## 1 ENSG00000183337             ENSMUSG00000040363
```

```
# Define a Function to do this on-the-fly below
humanEntrezToMouseEnsemble <- function(xyz) {
    getBM(attributes = myattributesHuman, filters = c("entrezgene",
        "with_homolog_mmus"), values = list(xyz, TRUE), mart = ensemblHuman)
}
```

Create an annotation matrix for genes in Total Reads

```
annot <- getBM(attributes = myattributes, filters = "ensembl_gene_id",
    values = rownames(TotalReads), mart = ensembl)
annot <- annot[!duplicated(annot[, "ensembl_gene_id"]), ]
rownames(annot) <- annot[, "ensembl_gene_id"]
new_annot <- as.data.frame(TotalReads)
new_annot$ensembl_gene_id <- rownames(new_annot)
# annotation has to be in teh same order as TotalReads
new_annot <- merge(new_annot, annot)
rownames(new_annot) <- rownames(TotalReads)
str(new_annot)
```

```
## 'data.frame':    22744 obs. of  14 variables:
##  $ ensembl_gene_id: chr  "ENSMUSG00000000001" "ENSMUSG00000000028" "ENSMUSG00000000031" "ENSMUSG00000000037
##  $ WT_R1          : int  7056 1443 10239 11435 2 1196 944 2018 684 1227 ...
##  $ Null_R1        : int  8128 1830 10646 12437 2 1429 847 2752 812 1534 ...
##  $ WT_R2          : int  9178 2164 11908 15461 4 1544 997 2658 761 1701 ...
##  $ WT_R3          : int  7908 2172 12174 13101 6 1391 817 2340 732 1577 ...
##  $ Null_R2        : int  8126 1856 11199 13838 18 1533 1336 2368 690 1323 ...
##  $ Null_R3        : int  6418 1752 9946 12395 0 1228 638 2153 1131 1344 ...
##  $ mgi_id         : chr  "MGI:95773" "MGI:1338073" "MGI:95891" "MGI:1340042" ...
##  $ mgi_symbol     : chr  "Gnai3" "Cdc45" "H19" "Scml2" ...
##  $ chromosome_name: chr  "3" "16" "7" "X" ...
##  $ start_position : int  107910198 18780540 149761434 157555125 108204668 121098567 17231185 5860735 120077
##  $ end_position   : int  107949064 18812080 149764048 157696145 108275710 121117170 17239115 5869639 120202
##  $ strand         : int  -1 -1 -1 1 1 1 1 1 1 1 ...
##  $ entrezgene     : int  14679 12544 NA 107815 11818 67608 12390 23849 29871 12858 ...
```

Create a function that will take a list of gene symbols and a query term and then return the number of publications in Pubmed and a URL to those publications.

```
pubmedBatchQuery <- function(temp, qt) {
    output = data.frame()
    for (i in 1:length(temp)) {
        # query=paste0(temp[i,'mgi_symbol'],' AND ',qt)
        query = paste0(temp[i], " AND ", qt)
        query = gsub("\\s+", "+", query)
        url = paste0("http://eutils.ncbi.nlm.nih.gov/entrez/eutils/",
            "esearch.fcgi?retmax=50000&db=pubmed&term=", query)
        datafile = tempfile(pattern = "pub")
        try(download.file(url, destfile = datafile, method = "internal",
            mode = "wb", quiet = TRUE), silent = TRUE)
        xml <- xmlTreeParse(datafile, asTree = TRUE)
        nid = xmlValue(xmlElementsByTagName(xmlRoot(xml), "Count")[[1]])
        lid = xmlElementsByTagName(xmlRoot(xml), "IdList", recursive = TRUE)[[1]]
        pid = paste(unlist(lapply(xmlElementsByTagName(lid, "Id"),
            xmlValue)), sep = ":")
        # print(c(hit_list[i],nid,pid))
        output[i, "PubMed"] = nid
        output[i, "URL"] = paste0("http://www.ncbi.nlm.nih.gov/pubmed/?term=",
            query)
    }
    return(output)
}

# Test it out
pubmedBatchQuery(c("Dmrt1", "Sox9"), "Testis")
```

```
##   PubMed
## 1    188
## 2    425
##                                                          URL
## 1 http://www.ncbi.nlm.nih.gov/pubmed/?term=Dmrt1+AND+Testis
## 2  http://www.ncbi.nlm.nih.gov/pubmed/?term=Sox9+AND+Testis
```

Use EdgeR to find differentially expressed genes.

```
group = factor(unlist(strsplit(colnames(TotalReads), "_"))[seq(from = 1,
    to = 2 * length(colnames(TotalReads)), by = 2)])
group
```

```
## [1] WT   Null WT   WT   Null Null
## Levels: Null WT
```

```
d = DGEList(counts = TotalReads, group = group, genes = new_annot)
design <- model.matrix(~0 + group)
design
```

```
##   groupNull groupWT
## 1         0       1
## 2         1       0
## 3         0       1
## 4         0       1
## 5         1       0
## 6         1       0
## attr(,"assign")
## [1] 1 1
## attr(,"contrasts")
## attr(,"contrasts")$group
## [1] "contr.treatment"
```

```
d <- calcNormFactors(d)
d$samples
```

```
##         group lib.size norm.factors
## WT_R1      WT 23768316       1.0026
## Null_R1  Null 26170933       1.0050
## WT_R2      WT 29046494       1.0073
## WT_R3      WT 27493481       0.9986
## Null_R2  Null 27369348       1.0082
## Null_R3  Null 26321829       0.9786
```

```
d <- estimateCommonDisp(d)
d$common.dispersion
```

```
## [1] 0.04091
```

```
d <- estimateTagwiseDisp(d)
et <- exactTest(d, pair = c("WT", "Null"))
summary(de <- decideTestsDGE(et, p = 0.05, adjust = "BH"))
```

```
##    [,1]
## -1    7
## 0 22721
## 1    16
```

```
tt <- topTags(et, n = 20, sort.by = "PValue", adjust.method = "BH")
detags <- rownames(d)[as.logical(de)]
plotSmear(et, de.tags = detags)
abline(h = c(-2, 2), col = "blue")
```

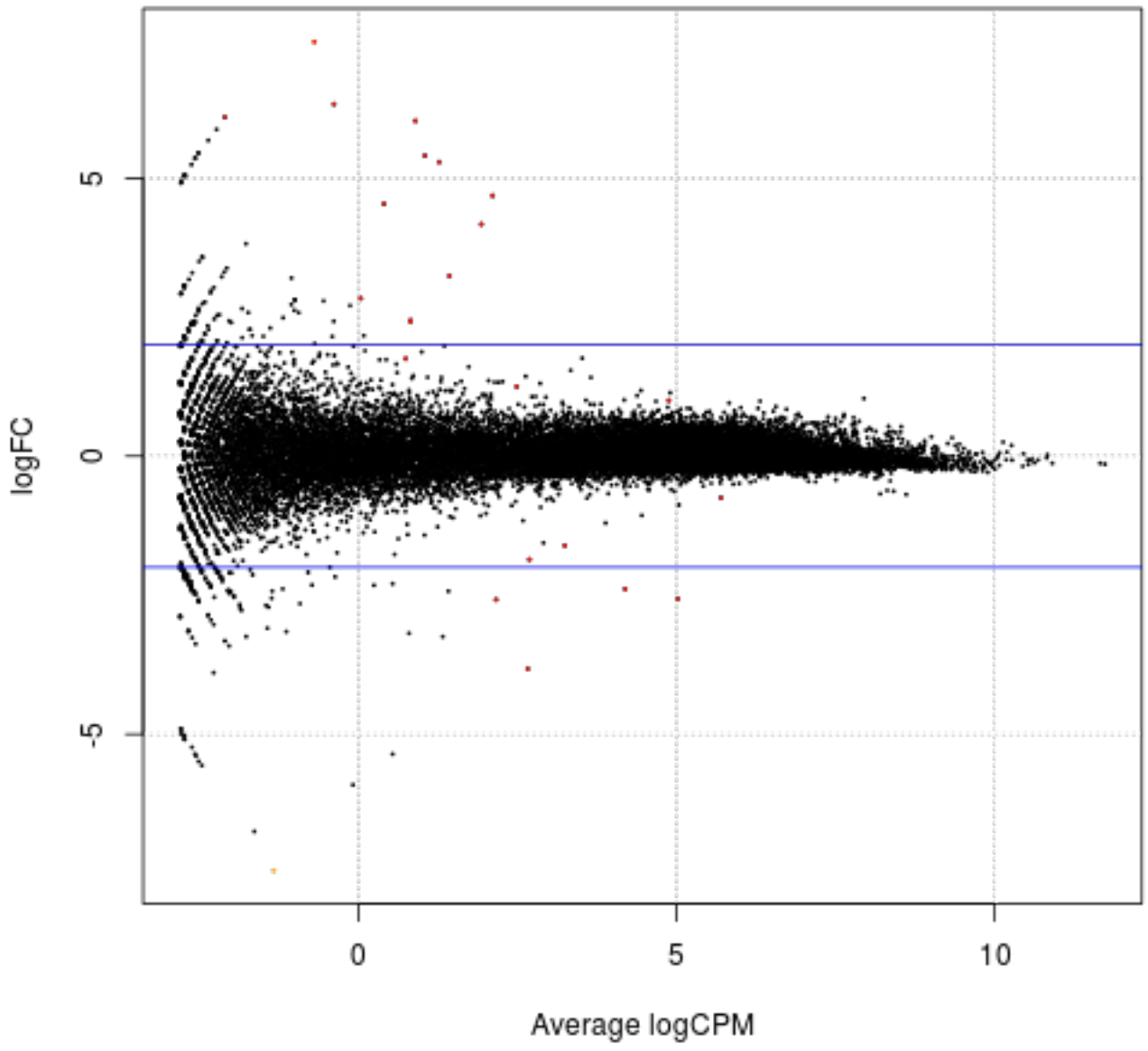

Figure 1: plot of chunk EdgeR

```
keep <- as.logical(de >= 1)
up = d[keep, ]
upt <- exactTest(up, pair = c("WT", "Null"))
uptt <- topTags(upt, n = 200, sort.by = "logFC", adjust.method = "BH")$table
```

Use EdgeR to build a GLM

```
D <- d
D <- estimateGLMCommonDisp(d, design)
```

```
# D <- estimateGLMTrendedDisp(d,design)
D <- estimateGLMTagwiseDisp(d, design)
plot(d$tag, D$tag, xlab = "ordinary dispersion", ylab = "GLM dispersion")
```

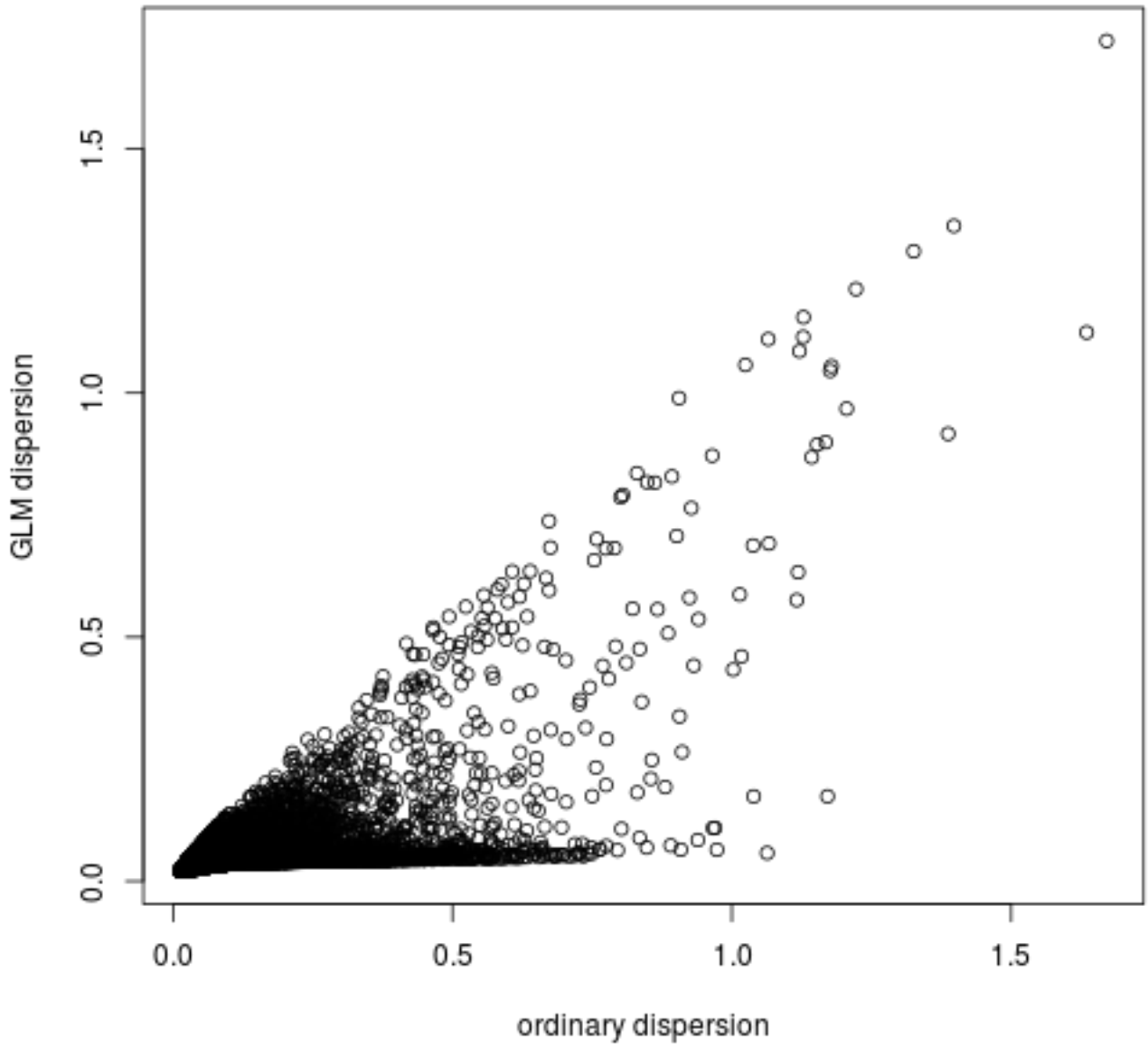

Figure 2: plot of chunk GLM

```
D_fit <- glmFit(D, design)
colnames(design)
```

```
## [1] "groupNull" "groupWT"
```

```
D6 <- c(1, -1)
lrt.D6 = glmLRT(D_fit, contrast = D6)
head(lrt.D6$table)
```

```
##                       logFC logCPM      LR PValue
## ENSMUSG00000000001 -0.07439  8.189 0.15962 0.6895
## ENSMUSG00000000028 -0.06068  6.126 0.09963 0.7523
## ENSMUSG00000000031 -0.09707  8.690 0.31132 0.5769
## ENSMUSG00000000037 -0.02933  8.937 0.02789 0.8674
## ENSMUSG00000000049  0.71888 -1.867 1.04772 0.3060
## ENSMUSG00000000056  0.03653  5.699 0.04014 0.8412
```

```
plotMDS(D)
```

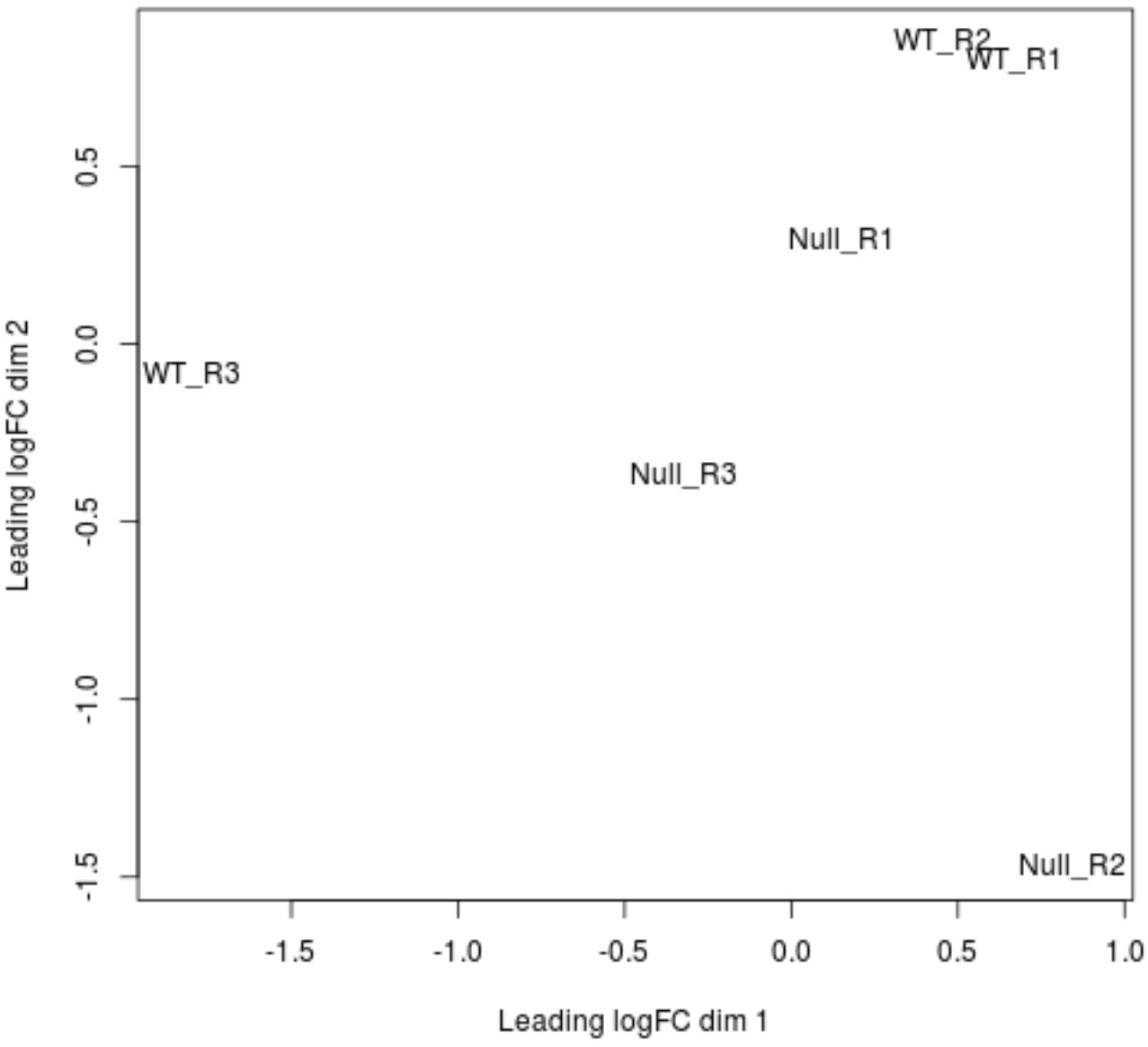

Figure 3: plot of chunk GLM

```
plotBCV(D)
```

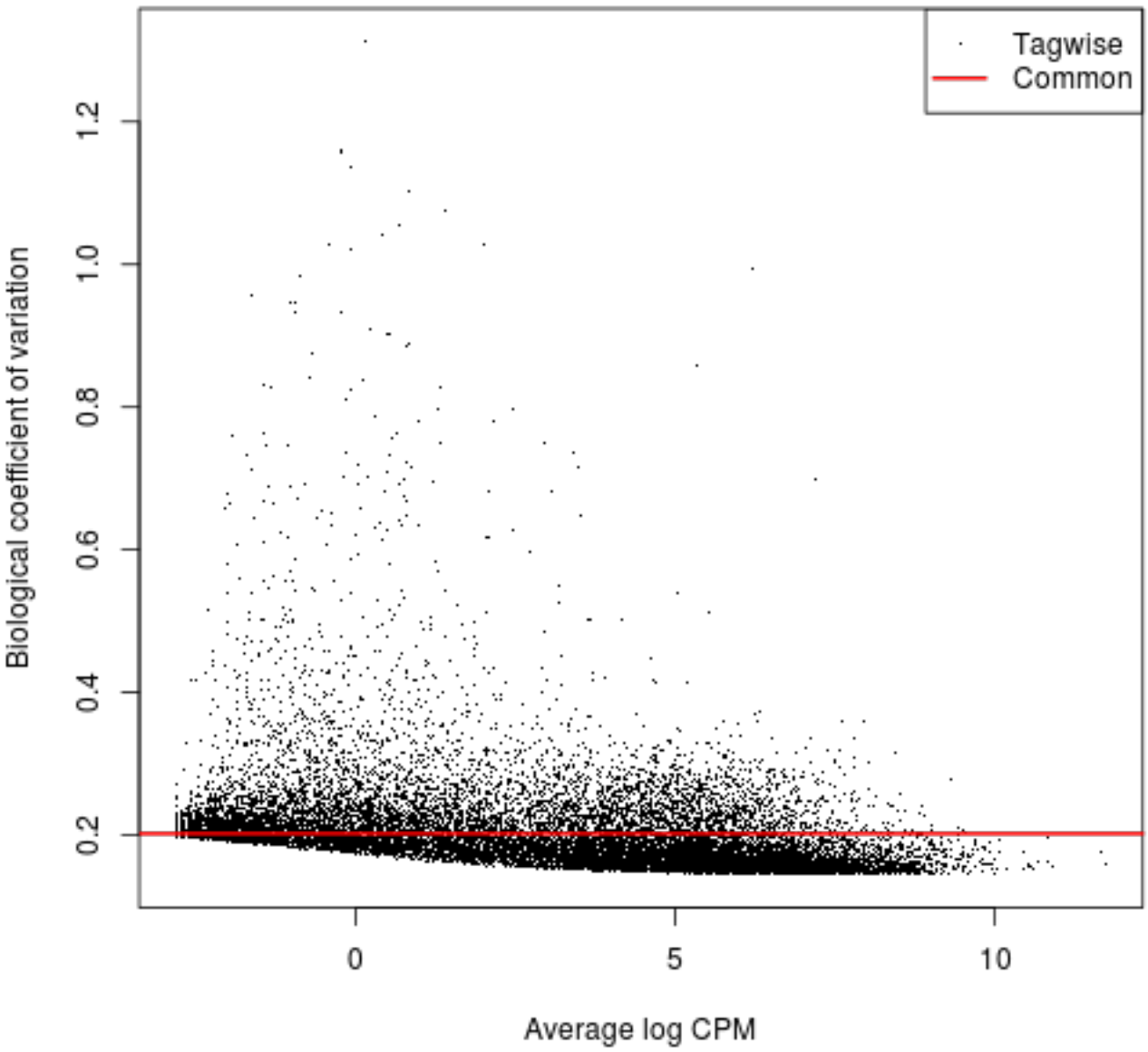

Figure 4: plot of chunk GLM

```
# PlotSmear: LogFC as a function of logCPM
summary(de <- decideTestsDGE(lrt.D6, p = 0.05, adjust = "BH"))
```

```
##    [,1]
## -1    43
## 0  22642
## 1     59
```

```
de.lrt <- rownames(D)[as.logical(de)]
plotSmear(lrt.D6, de.tags = de.lrt)
```

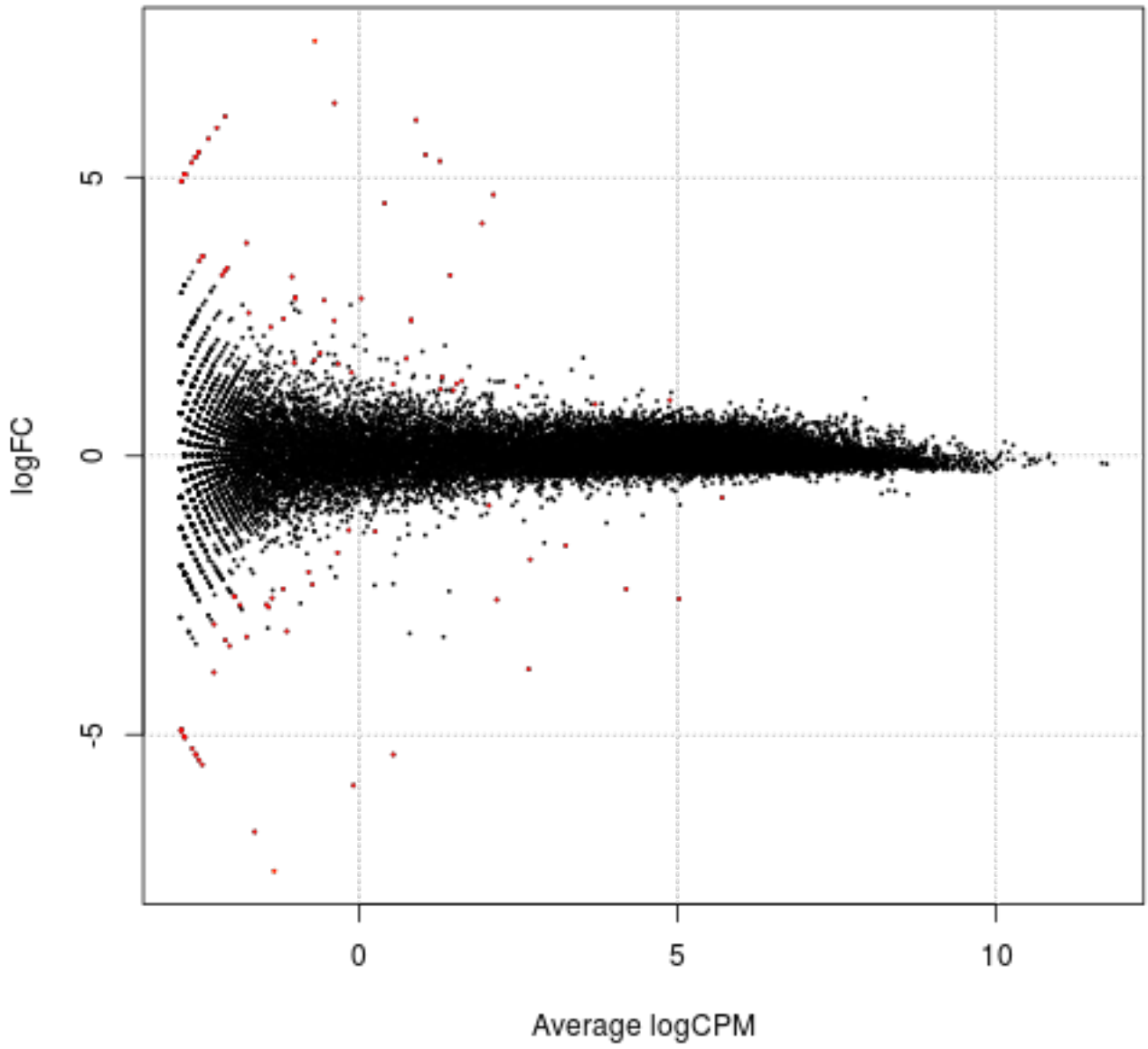

Figure 5: plot of chunk GLM

```
D6tt <- topTags(lrt.D6, n = Inf, sort.by = "none", adjust.method = "BH")$table
hist(D6tt$PValue, main = "PValue Distribution")
```

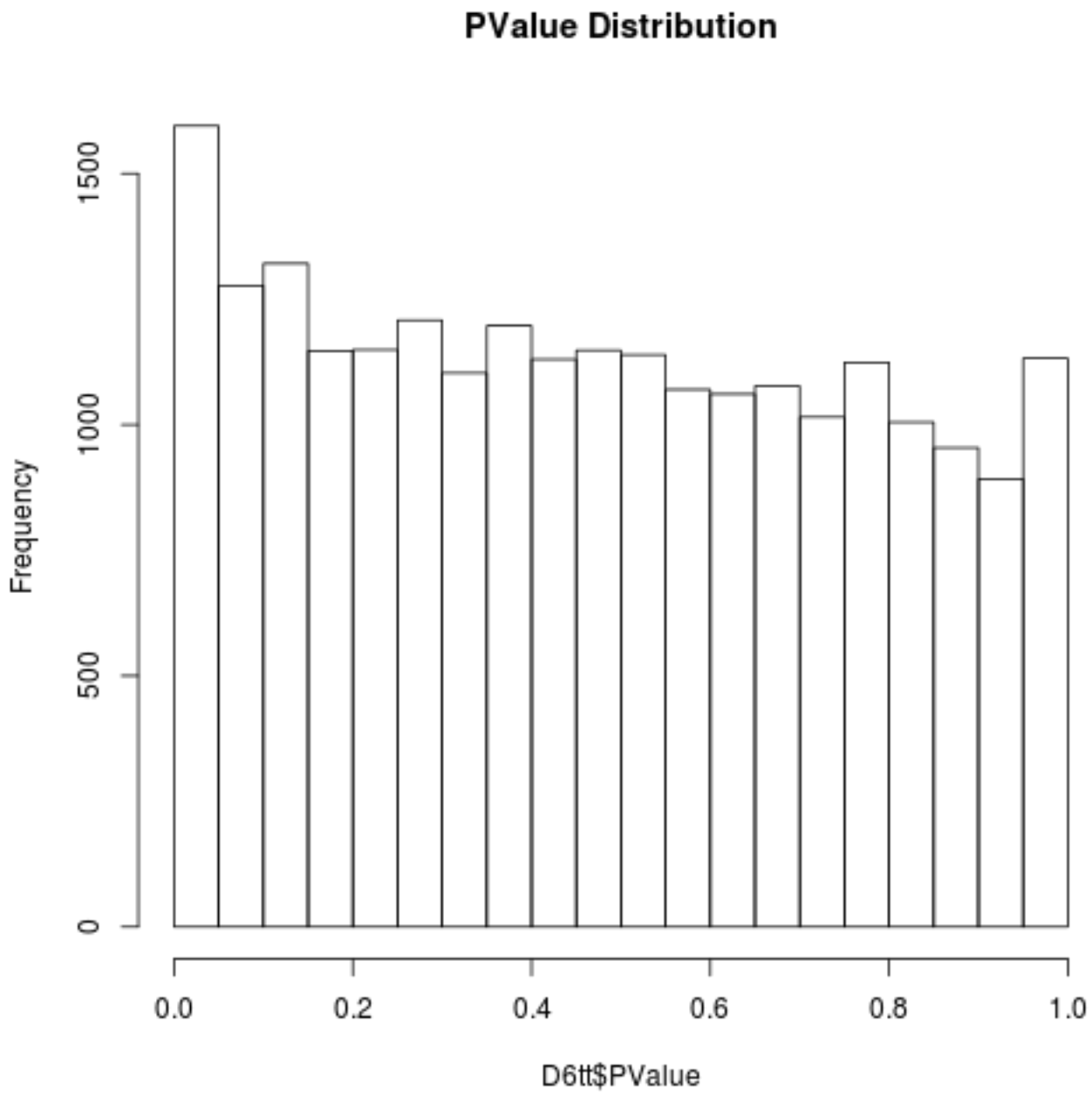

Figure 6: plot of chunk GLM

```
D6tt$qvalue <- qvalue(D6tt$PValue)$q
head(D6tt)
```

```
##                       ensembl_gene_id WT_R1 Null_R1 WT_R2
## ENSMUSG00000000001 ENSMUSG00000000001  7056    8128  9178
## ENSMUSG00000000028 ENSMUSG00000000028  1443    1830  2164
## ENSMUSG00000000031 ENSMUSG00000000031 10239   10646 11908
## ENSMUSG00000000037 ENSMUSG00000000037 11435   12437 15461
## ENSMUSG00000000049 ENSMUSG00000000049     2       2     4
## ENSMUSG00000000056 ENSMUSG00000000056  1196    1429  1544
##                    WT_R3 Null_R2 Null_R3      mgi_id
## ENSMUSG00000000001  7908    8126    6418   MGI:95773
## ENSMUSG00000000028  2172    1856    1752 MGI:1338073
## ENSMUSG00000000031 12174   11199    9946   MGI:95891
```

```
## ENSMUSG00000000037 13101   13838   12395 MGI:1340042
## ENSMUSG00000000049     6      18       0   MGI:88058
## ENSMUSG00000000056  1391    1533    1228 MGI:1914858
##                    mgi_symbol chromosome_name
## ENSMUSG00000000001      Gnai3               3
## ENSMUSG00000000028      Cdc45              16
## ENSMUSG00000000031        H19               7
## ENSMUSG00000000037      Scml2               X
## ENSMUSG00000000049       Apoh              11
## ENSMUSG00000000056       Narf              11
##                    start_position end_position strand
## ENSMUSG00000000001      107910198    107949064     -1
## ENSMUSG00000000028       18780540     18812080     -1
## ENSMUSG00000000031      149761434    149764048     -1
## ENSMUSG00000000037      157555125    157696145      1
## ENSMUSG00000000049      108204668    108275710      1
## ENSMUSG00000000056      121098567    121117170      1
##                    entrezgene    logFC logCPM      LR
## ENSMUSG00000000001      14679 -0.07439  8.189 0.15962
## ENSMUSG00000000028      12544 -0.06068  6.126 0.09963
## ENSMUSG00000000031         NA -0.09707  8.690 0.31132
## ENSMUSG00000000037     107815 -0.02933  8.937 0.02789
## ENSMUSG00000000049      11818  0.71888 -1.867 1.04772
## ENSMUSG00000000056      67608  0.03653  5.699 0.04014
##                    PValue    FDR qvalue
## ENSMUSG00000000001 0.6895 0.9560 0.8376
## ENSMUSG00000000028 0.7523 0.9669 0.8471
## ENSMUSG00000000031 0.5769 0.9366 0.8206
## ENSMUSG00000000037 0.8674 0.9825 0.8608
## ENSMUSG00000000049 0.3060 0.8908 0.7804
## ENSMUSG00000000056 0.8412 0.9766 0.8556
```

```
# Volcano Plot - LogFC vs Pvalue
plot(D6tt$logFC, -1 * log10(D6tt$PValue), cex = 0.5, pch = 19,
    col = ifelse(rownames(D6tt) %in% de.lrt, "red", "black"),
    main = "Dmrt6 Differential Expression")
```

```
# Use one of the following selection criteria
# D6tt<-D6tt[grep('Rhox',D6tt£mgi_symbol),] D6tt<-D6tt[de !=
# 0,] D6tt<-D6tt[D6tt£ensembl_gene_id %in%
# dmrt6Anno£feature,]
D6tt <- D6tt[D6tt$PValue < 0.05, ]
# D6tt<-D6tt[abs(D6tt£logFC)>1,]
```

Use Griswold's Microarray data to look at the expression of these genes through spermatogenesis.

```
gset <- getGEO("GSE4193", destdir = "/mnt/afp/micah/R/dmrt6",
    GSEMatrix = TRUE)
```

```
## ftp://ftp.ncbi.nlm.nih.gov/geo/series/GSE4nnn/GSE4193/matrix/
## Found 1 file(s)
## GSE4193_series_matrix.txt.gz
## Using locally cached version: /mnt/afp/micah/R/dmrt6/GSE4193_series_matrix.txt.gz
## Using locally cached version of GPL1261 found here:
## /mnt/afp/micah/R/dmrt6/GPL1261.soft
```

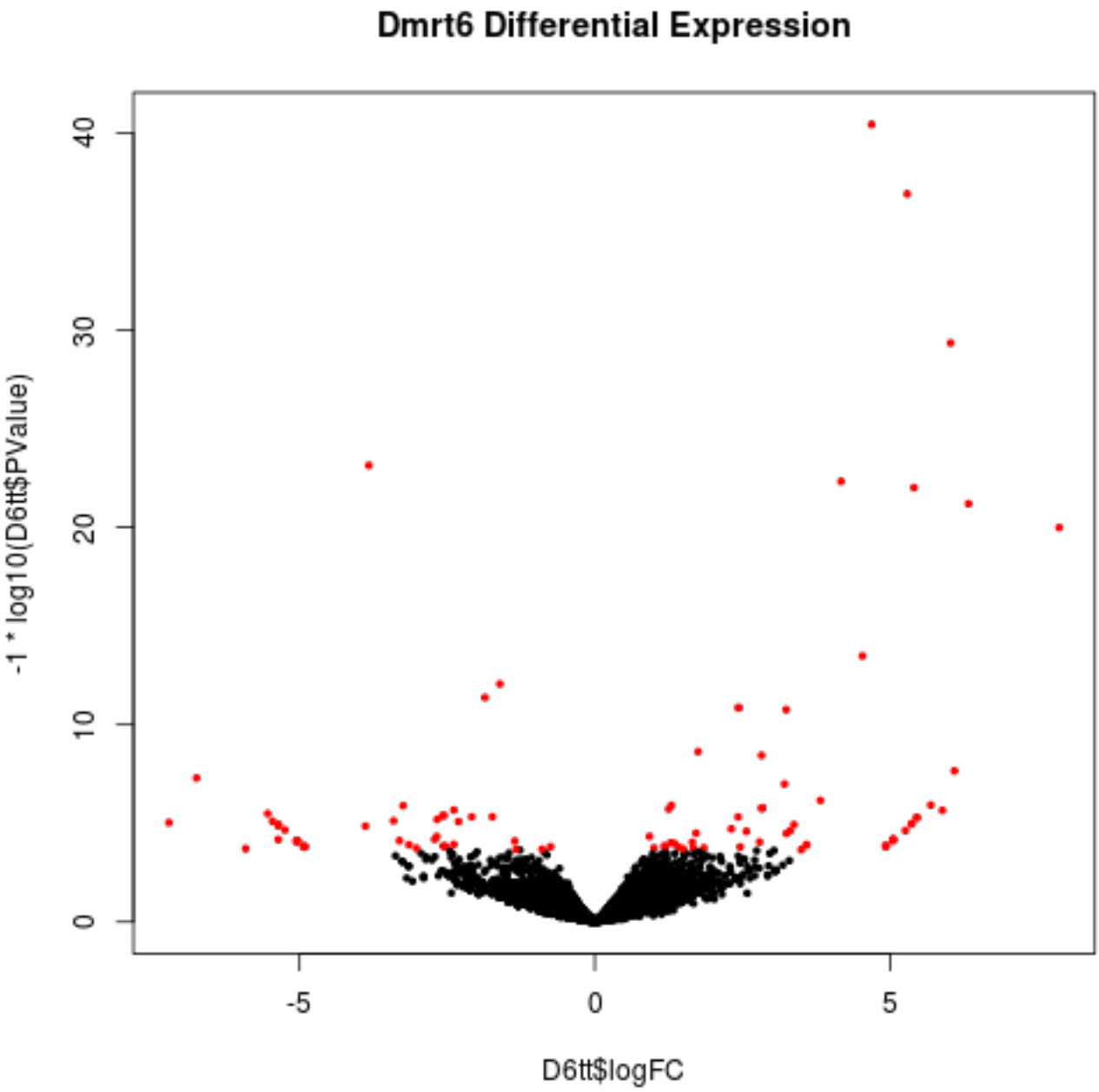

Figure 7: plot of chunk GLM

```
sml <- c("A", "A", "B", "B", "P", "P", "R", "R")
# the March 2014 change to this GSE now returns a list of
# expression sets so I need the [[1]]
ex <- exprs(gset[[1]])[, order(sml)]
ex["1427252_at", ]
```

```
## GSM95928 GSM95929 GSM95930 GSM95947 GSM95948 GSM95949
##    317.5    254.0    374.4    734.6   1640.0   1912.4
## GSM95950 GSM95951
##   2508.1   2607.4
```

```
ex["1460015_at", ]
```

```
## GSM95928 GSM95929 GSM95930 GSM95947 GSM95948 GSM95949
##    374.9    531.4    465.4    266.8     71.8     74.7
## GSM95950 GSM95951
##     91.8     74.7
```

```
gpl <- annotation(gset[[1]])
platf <- getGEO(gpl, AnnotGPL = TRUE, destdir = "/mnt/afp/micah/R/dmrt6")
```

```
## Using locally cached version of GPL1261 found here:
## /mnt/afp/micah/R/dmrt6/GPL1261.annot.gz
```

```
## Warning: seek on a gzfile connection returned an internal error
## Warning: seek on a gzfile connection returned an internal error
## Warning: seek on a gzfile connection returned an internal error
## Warning: seek on a gzfile connection returned an internal error
## Warning: seek on a gzfile connection returned an internal error
## Warning: seek on a gzfile connection returned an internal error
## Warning: seek on a gzfile connection returned an internal error
## Warning: seek on a gzfile connection returned an internal error
## Warning: seek on a gzfile connection returned an internal error
## Warning: seek on a gzfile connection returned an internal error
## Warning: seek on a gzfile connection returned an internal error
## Warning: seek on a gzfile connection returned an internal error
## Warning: seek on a gzfile connection returned an internal error
## Warning: seek on a gzfile connection returned an internal error
## Warning: seek on a gzfile connection returned an internal error
## Warning: seek on a gzfile connection returned an internal error
## Warning: seek on a gzfile connection returned an internal error
## Warning: seek on a gzfile connection returned an internal error
## Warning: seek on a gzfile connection returned an internal error
## Warning: seek on a gzfile connection returned an internal error
## Warning: seek on a gzfile connection returned an internal error
## Warning: seek on a gzfile connection returned an internal error
## Warning: seek on a gzfile connection returned an internal error
## Warning: seek on a gzfile connection returned an internal error
## Warning: seek on a gzfile connection returned an internal error
## Warning: seek on a gzfile connection returned an internal error
## Warning: seek on a gzfile connection returned an internal error
## Warning: seek on a gzfile connection returned an internal error
## Warning: seek on a gzfile connection returned an internal error
## Warning: seek on a gzfile connection returned an internal error
```

```
ncbifd <- data.frame(attr(dataTable(platf), "table"))
colnames(ex) <- c("A_R1", "A_R2", "B_R1", "B_R2", "P_R1", "P_R2",
    "R_R1", "R_R2")
```

```
ex2 <- merge(ex, ncbifd, by.x = 0, by.y = "ID")
ex2 <- subset(ex2, select = c("Row.names", "A_R1", "A_R2", "B_R1",
    "B_R2", "P_R1", "P_R2", "R_R1", "R_R2", "Gene.ID", "Gene.symbol"))
# NCBI Entry got update in March 2014, presumably to replace
# the log value with the raw value
# ex2£A<-0.5*(2^ex2£A_R1+2^ex2£A_R2)
# ex2£B<-0.5*(2^ex2£B_R1+2^ex2£B_R2)
# ex2£P<-0.5*(2^ex2£P_R1+2^ex2£P_R2)
# ex2£R<-0.5*(2^ex2£R_R1+2^ex2£R_R2)
ex2$A <- 0.5 * (ex2$A_R1 + ex2$A_R2)
ex2$B <- 0.5 * (ex2$B_R1 + ex2$B_R2)
ex2$P <- 0.5 * (ex2$P_R1 + ex2$P_R2)
ex2$R <- 0.5 * (ex2$R_R1 + ex2$R_R2)
ex2$sum <- ex2$A + ex2$B
ex2 <- ex2[with(ex2, order(-sum)), ]
ex2$Gene.ID <- as.numeric(as.character(ex2$Gene.ID))

## Warning: NAs introduced by coercion

ex2$Gene.symbol <- as.character(ex2$Gene.symbol)
ex2[grep("Sohlh1", ex2$Gene.symbol), ]

##        Row.names  A_R1  A_R2  B_R1  B_R2 P_R1 P_R2 R_R1
## 44308 1460015_at 374.9 531.4 465.4 266.8 71.8 74.7 91.8
##       R_R2 Gene.ID Gene.symbol     A     B     P     R
## 44308 74.7  227631      Sohlh1 453.1 366.1 73.25 83.25
##         sum
## 44308 819.2

nrow(ex2)

## [1] 45101

sum(!(duplicated(ex2[, "Gene.ID"])) & !is.na(ex2[, "Gene.ID"]))

## [1] 20992

# head(ex2[is.na(ex2[,'Gene.ID']),]£Gene.ID,n=50) rm(ex3)

ex3 <- ex2[!(duplicated(ex2[, "Gene.ID"])) & !is.na(ex2[, "Gene.ID"]),
    ]
head(ex3)

##                         Row.names A_R1 A_R2 B_R1 B_R2 P_R1
## 23165               1438859_x_at 6415 5679 5917 6338 5231
## 35396               1451101_a_at 5765 5053 5407 5737 3894
## 44873               1460581_a_at 5936 4941 5050 5966 3571
## 8941                  1424635_at 5674 5080 5242 5743 5018
## 45078 AFFX-b-ActinMur/M12481_3_at 5828 5243 4865 5502 3436
## 210                 1415879_a_at 5763 4425 5121 5123 1765
##       P_R2 R_R1 R_R2 Gene.ID Gene.symbol    A    B    P
## 23165 5583 5998 6306   20090       Rps29 6047 6128 5407
## 35396 4186 4432 4456   54127       Rps28 5409 5572 4040
## 44873 3948 3206 3467  270106       Rpl13 5439 5508 3759
## 8941  5527 4578 5103   13627      Eef1a1 5377 5492 5272
## 45078 3704 3964 3448   11461        Actb 5536 5184 3570
## 210   2006 2248 2748   67186       Rplp2 5094 5122 1886
##          R   sum
```

```
## 23165 6152 12174
## 35396 4444 10981
## 44873 3337 10947
## 8941  4840 10870
## 45078 3706 10719
## 210   2498 10217
```

```
ex3[grep("Dmrtb1", ex3$Gene.symbol), ]
```

```
##        Row.names  A_R1 A_R2  B_R1  B_R2 P_R1 P_R2 R_R1 R_R2
## 11558 1427252_at 317.5  254 374.4 734.6 1640 1912 2508 2607
##       Gene.ID Gene.symbol     A     B    P    R   sum
## 11558   56296      Dmrtb1 285.8 554.5 1776 2558 840.2
```

```
rownames(ex3) <- ex3$Gene.ID
ex3 <- subset(ex3, select = c("Gene.symbol", "A", "B", "P", "R"))
ex3[grep("Dmrtb1", ex3$Gene.symbol), ]
```

```
##       Gene.symbol     A     B    P    R
## 56296      Dmrtb1 285.8 554.5 1776 2558
```

```
ex3[grep("Sohlh1", ex3$Gene.symbol), ]
```

```
##        Gene.symbol     A     B     P     R
## 227631      Sohlh1 453.1 366.1 73.25 83.25
```

```
# ncbifd[grep('Dmrtb1',ncbifd£Gene.symbol),] Merge D6tt with
# Griswold's data head(D6tt)
D6tt <- merge(D6tt, ex3, by.x = "entrezgene", by.y = 0, all.x = TRUE)
# sum(duplicated(D6tt£ensembl_gene_id))
# rownames(D6tt)<-D6tt£ensembl_gene_id head(D6tt) nrow(ex2)
# ex2<-ex2[!is.na(ex2£Gene.ID),] ex3<-ex2[1:nrow(ex2),]
# rownames(ex3)<-ex3£Gene.ID
```

```
D6tt[(grep("Dmrtb1", D6tt$mgi_symbol)), ]
```

```
##     entrezgene    ensembl_gene_id WT_R1 Null_R1 WT_R2 WT_R3
## 395      56296 ENSMUSG00000028610   194      14   510   254
##     Null_R2 Null_R3      mgi_id mgi_symbol chromosome_name
## 395      25      26 MGI:1927125     Dmrtb1               4
##     start_position end_position strand  logFC logCPM    LR
## 395      107348895    107356835     -1 -3.824  2.657 101.5
##        PValue      FDR    qvalue Gene.symbol     A     B
## 395 7.123e-24 4.05e-20 3.548e-20      Dmrtb1 285.8 554.5
##        P    R
## 395 1776 2558
```

```
D6tt[(grep("Dmrt1", D6tt$mgi_symbol)), ]
```

```
##  [1] entrezgene      ensembl_gene_id WT_R1
##  [4] Null_R1         WT_R2           WT_R3
##  [7] Null_R2         Null_R3         mgi_id
## [10] mgi_symbol      chromosome_name start_position
## [13] end_position    strand          logFC
## [16] logCPM          LR              PValue
## [19] FDR             qvalue          Gene.symbol
## [22] A               B               P
## [25] R
## <0 rows> (or 0-length row.names)
```

Include Chip-Seq Data in D6tt

```
# Run on Server macs14 -t M8W_chip_dedup.bam -c
# M8W_input_dedup.bam -f BAM -s 25 \ -g 1.87e9 -p 1e-05
# --slocal 100 --llocal 1000 -n M8W_dedup_macs14_pe05 macs14
# -t DM6_chip_dedup.bam -c DM6_input_dedup.bam -f BAM -s 25
# \ -g 1.87e9 -p 1e-05 --slocal 100 --llocal 1000 -n
# DM6_dedup_macs14_pe05

# read in MACS Peaks and find overlaps with DMRT1 sites
d1p05 <- import("M8W_dedup_macs14_pe05_peaks.bed")
d6p05 <- import("DM6_dedup_macs14_pe05_peaks.bed")

# find overlaps between
mp05overlap <- findOverlaps(d6p05, d1p05)

grid.newpage()
vennplot <- draw.pairwise.venn(length(d1p05), length(d6p05),
    length(mp05overlap), c("Dmrt1", "Dmrt6"))
grid.draw(vennplot)



# Annotate d6macs peaks
d6macs <- annotatePeakInBatch(as(d6p05, "RangedData"), AnnotationData = TSS.mouse.NCBIM37,
    output = "both")
d6macs <- addGeneIDs(d6macs, "org.Mm.eg.db", c("refseq", "symbol"))

## Adding refseq ... done
## Adding symbol ... done
## prepare output ... done

d1macs <- annotatePeakInBatch(as(d1p05, "RangedData"), AnnotationData = TSS.mouse.NCBIM37,
    output = "both")
d1macs <- addGeneIDs(d1macs, "org.Mm.eg.db", c("refseq", "symbol"))

## Adding refseq ... done
## Adding symbol ... done
## prepare output ... done


# Calculate # of Unique Features in D6
length(unique(d6macs$feature))

## [1] 10363

length(d6p05)

## [1] 14862

length(unique(d1macs$feature))

## [1] 14769


# Annotate Dmrt6 TopTable with Dmrt1 & Dmrt6 Chip Occupancy
D6tt$d6macs <- D6tt$ensembl_gene_id %in% d6macs$feature
D6tt$d1macs <- D6tt$ensembl_gene_id %in% d1macs$feature
```

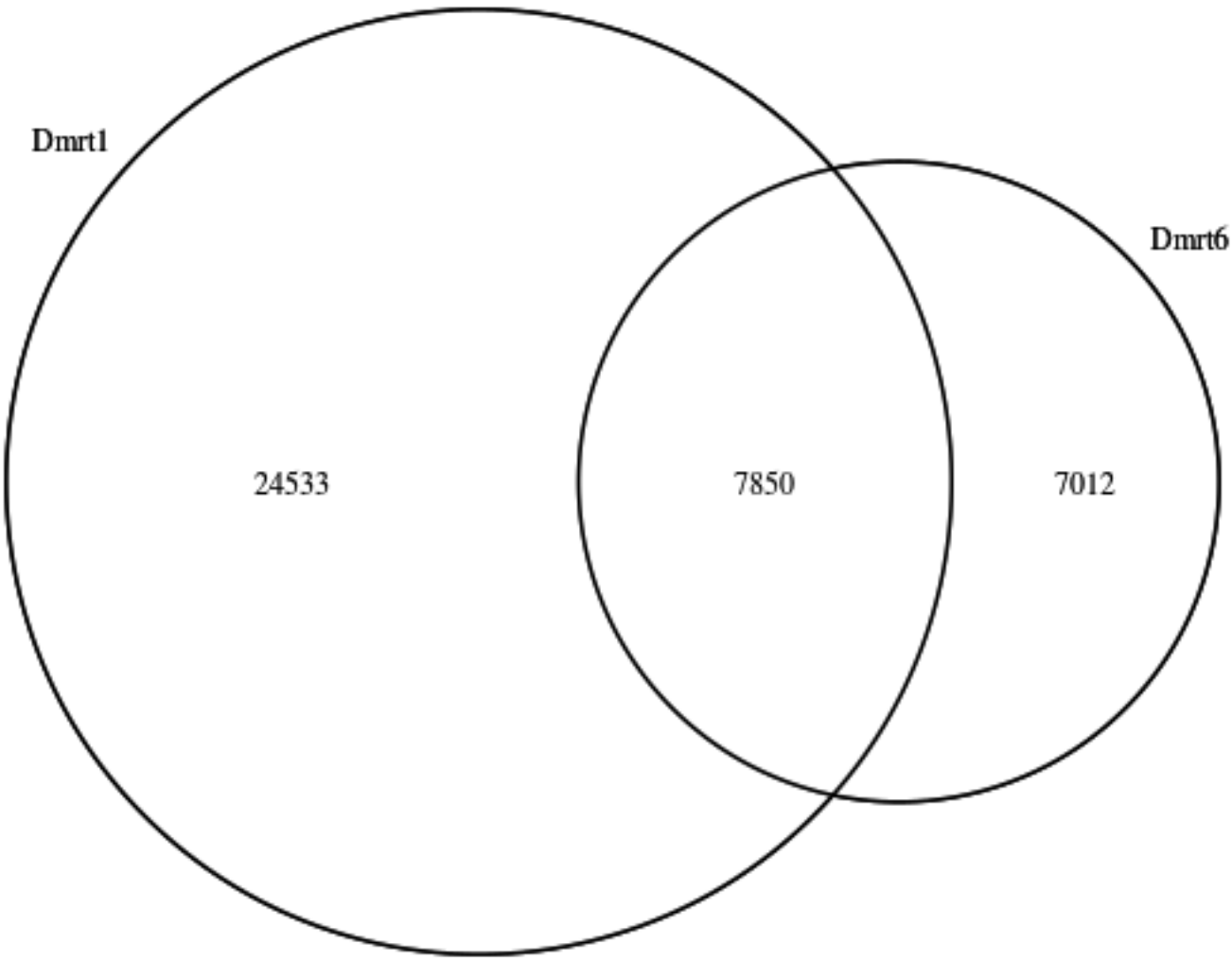

Figure 8: plot of chunk ChipSeq

```
invitro_site <- readDNAStringSet("/mnt/afp/murphy/profit/temp.fa")
pfm_vitro <- consensusMatrix(invitro_site)
pwm_vitro <- PWM(invitro_site)
pfm.vitro <- new("pfm", mat = t(t(pfm_vitro[1:4, ]) * 1/colSums(pfm_vitro[1:4,
    ])), name = "In Vitro DMRT1 Site 2007")
plotMotifLogo(pfm.vitro)
```

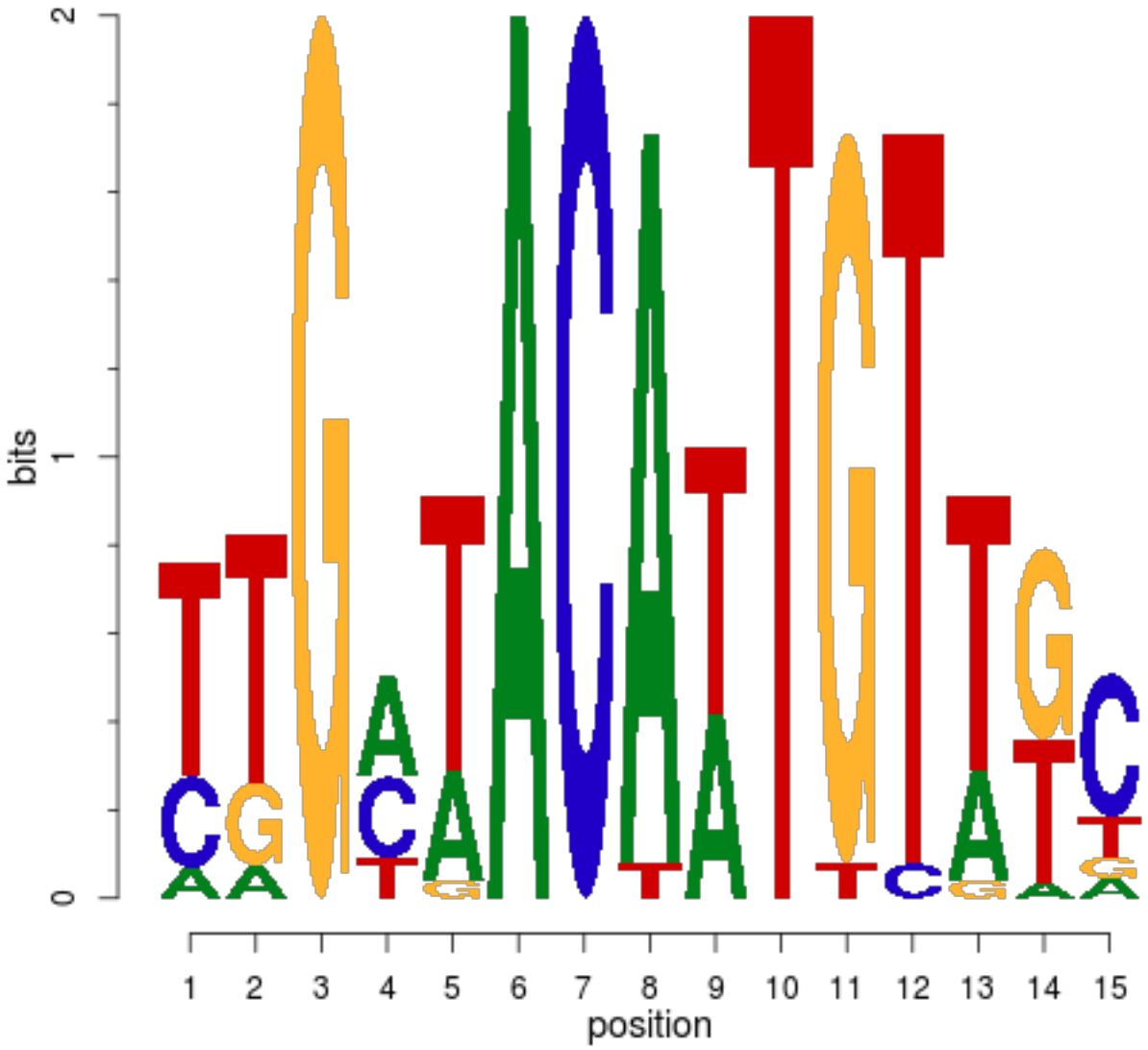

Figure 9: plot of chunk inVitroPwmSearch

```
findPWMinGR <- function(gr, pwm) {
    c <- numeric()
    for (i in 1:length(gr)) {
        peak <- DNAString(Mmusculus[[as.character(seqnames(gr[i])@values)]],
            start = ranges(gr[i])@start, nchar = ranges(gr[i])@width)
```

```
        site <- matchPWM(pwm, peak, min.score = "70%", with.score = TRUE)
        # c[i]<-ifelse(length(site)>0,paste(round(elementMetadata(site)£score,4),collapse=';'),'0')
        if (length(site) > 0) {
            c[i] <- max(elementMetadata(site)$score)
        } else {
            c[i] <- 0
        }
    }
    return(c)
}

# test Genomic Range on Peaks of interest
gr <- d6p05[c(1219, 8236, 8237, 7547, 8688)]
findPWMinGR(gr, pwm_vitro)

## [1] 0.9316 0.0000 0.7475 0.8561 0.0000


# Find DM domain motifs in full macs peak list
d6p05DF <- as.data.frame(d6p05)
system.time(d6p05DF$maxsite <- findPWMinGR(d6p05, pwm_vitro))

##    user  system elapsed
##   791.6    20.3   830.7


# Calculate fraction of peaks that have DM domain binding
# motifs
sum(d6p05DF$maxsite > 0.7)/nrow(d6p05DF)

## [1] 0.7391

# plot(d6p05DF£score,d6p05DF£maxsite,ylim=c(0.7,1),xlim=c(50,3500),cex=0.5,
# pch=19)

# Calculate Correlation, excluding outliers
d6p05_tempDF <- d6p05DF[d6p05DF$maxsite > 0.7 & d6p05DF$score <
    2000, ]
plot(d6p05_tempDF$score, d6p05_tempDF$maxsite, cex = 0.5, pch = 19)


cor(d6p05_tempDF$score, d6p05_tempDF$maxsite)

## [1] 0.2438


# Cummulative Sum of sites as Pvalue decreases (MACS score
# increases)
d6p05_tempDF <- d6p05DF[with(d6p05DF, order(-score)), ]
plot(cumsum(d6p05_tempDF$maxsite > 0.7), cex = 0.5, pch = 19)
abline(0, sum(d6p05DF$maxsite > 0.7)/nrow(d6p05DF), col = "red")
```

Count reads for Adult DMRT1 and DMRT6 ChipSeq data.

```
bamlst <- BamFileList(list.files("/mnt/afp/murphy/data/mm9",
    pattern = glob2rx("M8W_*_dedup.bam"), full = TRUE))
d1counts <- summarizeOverlaps(d6p05, bamlst, mode = "Union",
```

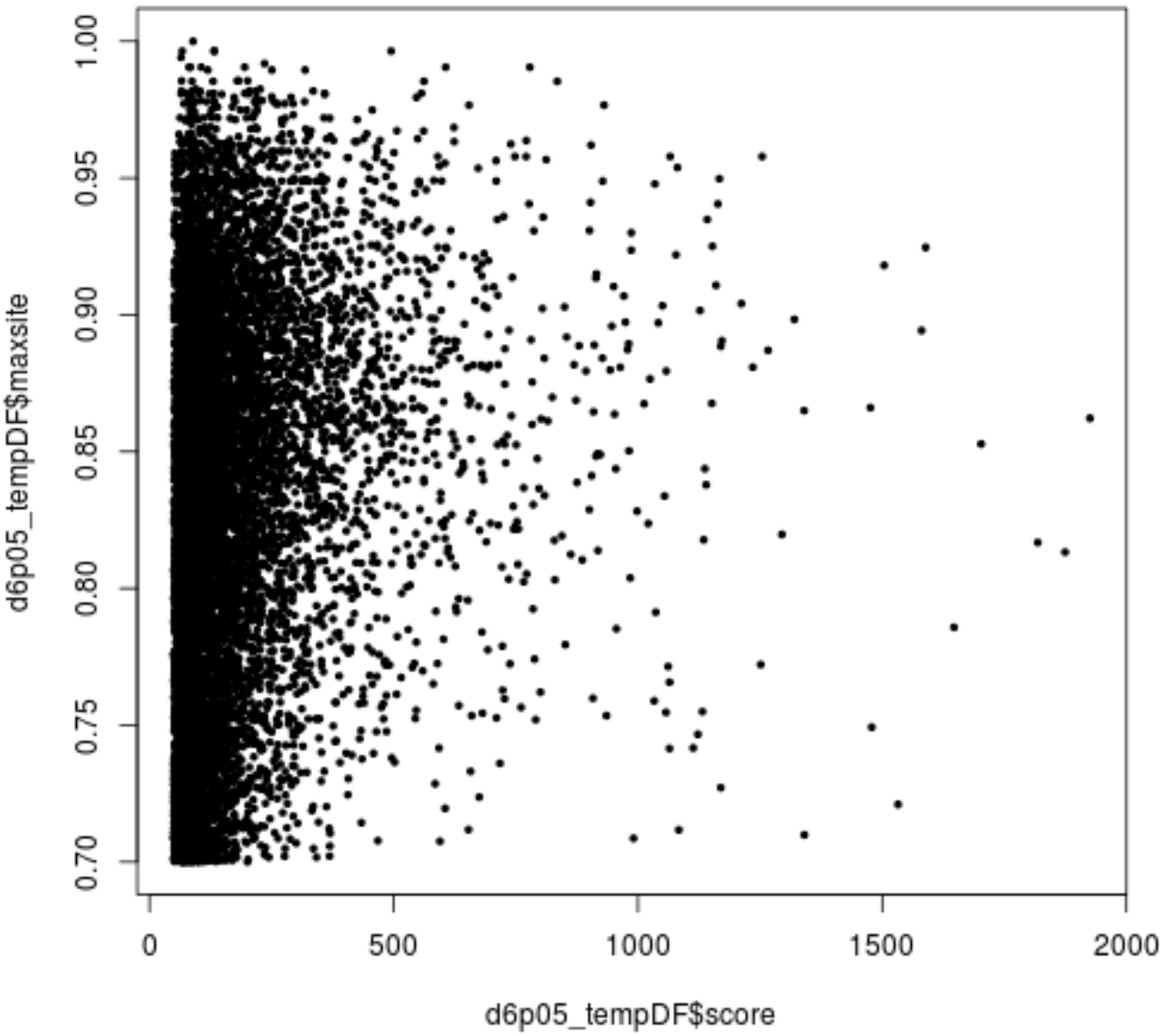

Figure 10: plot of chunk inVitroPwmSearch

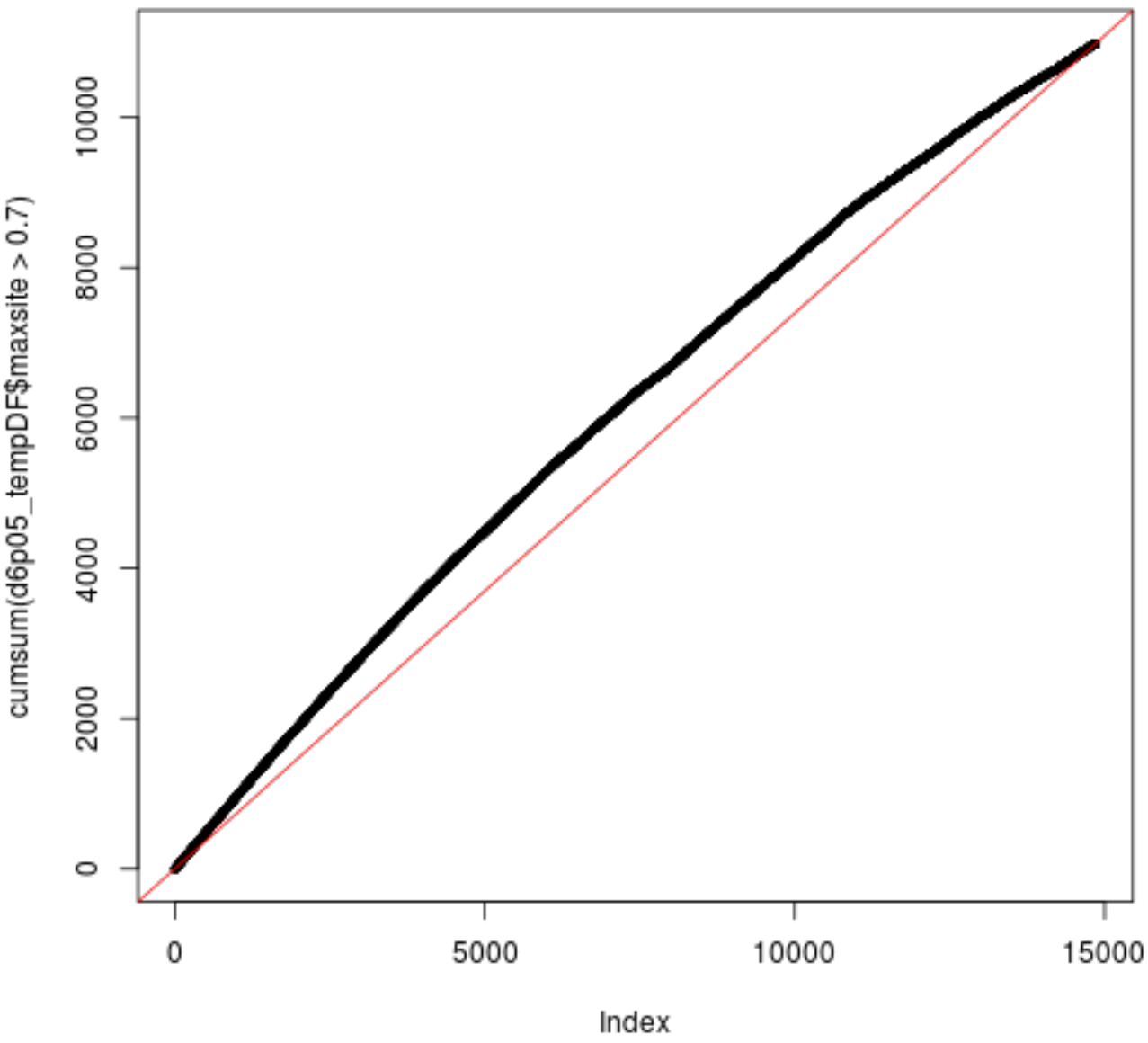

Figure 11: plot of chunk inVitroPwmSearch

```
    singleEnd = TRUE, ignore.strand = TRUE)
d1countsDF <- as.data.frame(assays(d1counts)$counts)

bamlst <- BamFileList(list.files("/mnt/afp/murphy/data/mm9",
    pattern = glob2rx("DM6_*_dedup#.bam"), full = TRUE))
d6counts <- summarizeOverlaps(d6p05, bamlst, mode = "Union",
    singleEnd = TRUE, ignore.strand = TRUE)
d6countsDF <- as.data.frame(assays(d6counts)$counts)
save(d1countsDF, d6countsDF, file = "chip_count_p05.rdata")
```

Analyze ChIP counts to identify Dmrt6 Specific Binding sites.

```
load("chip_count_p05.rdata")

# Normalize to Counts within regions of interest
colnames(d1countsDF) <- c("d1c", "d1i")

# normalize to total counts in genomic intervals
d1Enrichment <- log2(10^6 * d1countsDF[, 1]/sum(d1countsDF[,
    1]))
colnames(d6countsDF) <- c("d6c", "d6i")
d6Enrichment <- log2(10^6 * d6countsDF[, 1]/sum(d6countsDF[,
    1]))
# define logical varible to loosely define 'dmrt6 specific
# Peaks'
subset = d6Enrichment/d1Enrichment > 1.25

plot(d1Enrichment, d6Enrichment, ylim = c(4, 14), pch = 19, cex = 0.5,
    col = ifelse(subset, "red", "black"))



# calculate correlation coefficient for DMRT6 an DMRT1
# binding intensity
cor(d6Enrichment, d1Enrichment, method = "spearman")

## [1] 0.6424


# Output a Table sum(d6Enrichment/d1Enrichment > 1.25)
d6p05DF$d6cpm <- d6countsDF[, "d6c"]
d6p05DF$d1cpm <- d1countsDF[, "d1c"]
d6p05DF$d6Enrichment <- d6Enrichment
d6p05DF$d1Enrichment <- d1Enrichment
d6p05DF$ratio <- d6Enrichment/d1Enrichment

d6macsDF <- as.data.frame(d6macs)
d6macsDF$peak <- as.integer(d6macsDF$peak)
d6macsDF <- d6macsDF[, c("peak", "feature", "symbol", "insideFeature")]

d6out <- merge(d6p05DF, d6macsDF, by.x = 0, by.y = "peak", all = T)
d6out$row <- as.integer(d6out$Row.names)
d6out <- d6out[with(d6out, order(row)), ]
d6out <- d6out[, c("feature", "symbol", "seqnames", "start",
    "end", "width", "score", "maxsite", "name", "d6cpm", "d1cpm",
    "d6Enrichment", "d1Enrichment", "ratio")]
d6out <- d6out[with(d6out, order(-score)), ]
# head(d6out) d6out[grep('Kat6a',d6out$symbol),]
write.csv(d6out, file = "/mnt/afp/teng/data/Supplementary_Table_3.csv",
    quote = F, row.names = F)
```

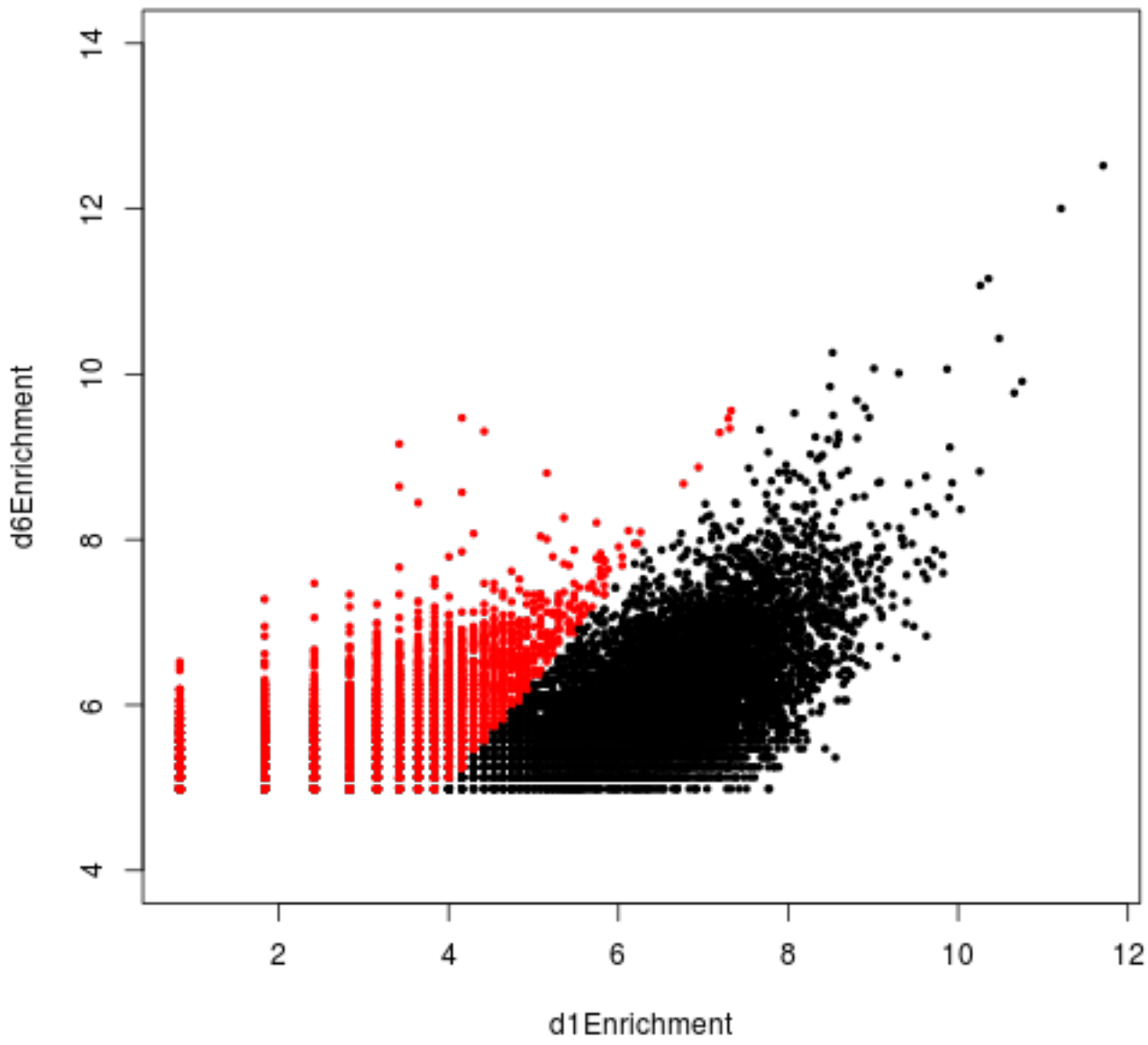

Figure 12: plot of chunk analyzeChipSeqCounts

Quick check for Enriched GO Terms in DMRT6 Specific Peaks

```
# universe<-keys(org.Mm.eg.db,'SYMBOL')
univ1 <- unique(as.character(na.omit(d1macs$symbol)))
univ6 <- unique(as.character(na.omit(d6macs$symbol)))
universe <- unique(c(univ1, univ6))
length(universe)
```

```
## [1] 12171
```

```
selected <- unique(as.character(na.omit(d6macs[subset, ]$symbol)))
length(selected)
```

```
## [1] 4120
```

```
univmap <- select(org.Mm.eg.db, universe, "ENTREZID", "SYMBOL")
genemap <- select(org.Mm.eg.db, selected, "ENTREZID", "SYMBOL")
param <- new("GOHyperGParams", geneIds = genemap, universeGeneIds = univmap,
    annotation = "org.Mm.eg.db", ontology = "BP", pvalueCutoff = 0.01,
    conditional = FALSE, testDirection = "over")
```

```
## Warning: converting geneIds from list to atomic vector via unlist
## Warning: removing duplicate IDs in geneIds
## Warning: converting univ from list to atomic vector via unlist
## Warning: removing duplicate IDs in universeGeneIds
```

```
hyp <- hyperGTest(param)
tt <- head(summary(hyp), 20)
tt
```

```
##        GOBPID    Pvalue OddsRatio ExpCount Count Size
## 1  GO:0031323 4.069e-11     1.330   1054.9  1203 3118
## 2  GO:0050794 2.877e-09     1.259   1794.8  1943 5305
## 3  GO:0080090 3.476e-09     1.293   1026.1  1157 3033
## 4  GO:0060255 5.704e-09     1.294    970.3  1097 2868
## 5  GO:0044260 6.488e-09     1.262   1393.6  1533 4119
## 6  GO:0048519 1.365e-08     1.309    792.0   907 2341
## 7  GO:0019222 1.577e-08     1.266   1173.0  1303 3467
## 8  GO:0051252 1.765e-08     1.327    674.6   782 1994
## 9  GO:0019219 1.825e-08     1.302    813.0   928 2403
## 10 GO:0010468 2.082e-08     1.303    800.1   914 2365
## 11 GO:0051171 2.143e-08     1.299    822.1   937 2430
## 12 GO:0009653 2.613e-08     1.373    492.6   586 1456
## 13 GO:0006355 3.429e-08     1.324    652.3   756 1928
## 14 GO:2001141 3.750e-08     1.322    656.3   760 1940
## 15 GO:0048523 4.384e-08     1.308    718.3   825 2123
## 16 GO:0016070 6.149e-08     1.292    790.7   900 2337
## 17 GO:0032774 6.955e-08     1.313    663.1   765 1960
## 18 GO:0006351 7.487e-08     1.314    657.7   759 1944
## 19 GO:0050789 9.404e-08     1.229   1901.0  2034 5619
## 20 GO:2000112 1.484e-07     1.294    715.9   818 2116
##                                                    Term
## 1                       regulation of cellular metabolic process
## 2                                 regulation of cellular process
## 3                        regulation of primary metabolic process
## 4                    regulation of macromolecule metabolic process
```

```
## 5                         cellular macromolecule metabolic process
## 6                       negative regulation of biological process
## 7                                regulation of metabolic process
## 8                             regulation of RNA metabolic process
## 9  regulation of nucleobase-containing compound metabolic process
## 10                                  regulation of gene expression
## 11              regulation of nitrogen compound metabolic process
## 12                               anatomical structure morphogenesis
## 13                     regulation of transcription, DNA-templated
## 14                         regulation of RNA biosynthetic process
## 15                       negative regulation of cellular process
## 16                                           RNA metabolic process
## 17                                         RNA biosynthetic process
## 18                                      transcription, DNA-templated
## 19                                   regulation of biological process
## 20      regulation of cellular macromolecule biosynthetic process
```

```
# barplot(-log10(tt£Pvalue), names.arg=paste(tt£Term,
# tt£GOBPID), las=2, ylab='-log10 p-value', col='Red')

# try another test for all DMRT6 peaks
selected <- univ6
genemap <- select(org.Mm.eg.db, selected, "ENTREZID", "SYMBOL")
param <- new("GOHyperGParams", geneIds = genemap, universeGeneIds = univmap,
    annotation = "org.Mm.eg.db", ontology = "BP", pvalueCutoff = 0.01,
    conditional = FALSE, testDirection = "over")
```

```
## Warning: converting geneIds from list to atomic vector via unlist
## Warning: removing duplicate IDs in geneIds
## Warning: converting univ from list to atomic vector via unlist
## Warning: removing duplicate IDs in universeGeneIds
```

```
hyp <- hyperGTest(param)
tt <- head(summary(hyp), 20)
tt
```

```
##        GOBPID    Pvalue OddsRatio ExpCount Count Size
## 1  GO:0044260 1.896e-15     1.383     2662  2855 4119
## 2  GO:0031323 1.296e-12     1.368     2015  2174 3118
## 3  GO:0043170 2.645e-12     1.320     2940  3113 4549
## 4  GO:0060255 1.232e-11     1.360     1854  2001 2868
## 5  GO:0006139 2.205e-11     1.334     2150  2303 3327
## 6  GO:0016070 3.602e-11     1.383     1510  1644 2337
## 7  GO:0044237 6.671e-11     1.286     3567  3732 5519
## 8  GO:0019222 1.023e-10     1.316     2241  2390 3467
## 9  GO:0080090 1.530e-10     1.328     1960  2102 3033
## 10 GO:0010467 2.218e-10     1.319     2044  2186 3162
## 11 GO:0046483 2.447e-10     1.310     2201  2346 3405
## 12 GO:0034645 2.468e-10     1.352     1611  1742 2493
## 13 GO:0034641 3.892e-10     1.303     2252  2397 3485
## 14 GO:0010468 4.208e-10     1.354     1528  1655 2365
## 15 GO:0006725 6.894e-10     1.300     2216  2358 3429
## 16 GO:0090304 7.546e-10     1.331     1708  1838 2643
## 17 GO:0051171 1.145e-09     1.339     1570  1695 2430
## 18 GO:0006807 1.501e-09     1.285     2386  2528 3692
## 19 GO:0019219 2.094e-09     1.334     1553  1675 2403
## 20 GO:0009059 2.117e-09     1.325     1653  1778 2558
##                                                          Term
## 1                         cellular macromolecule metabolic process
```

```
## 2                        regulation of cellular metabolic process
## 3                                macromolecule metabolic process
## 4                     regulation of macromolecule metabolic process
## 5                  nucleobase-containing compound metabolic process
## 6                                           RNA metabolic process
## 7                                     cellular metabolic process
## 8                                  regulation of metabolic process
## 9                        regulation of primary metabolic process
## 10                                            gene expression
## 11                                heterocycle metabolic process
## 12                     cellular macromolecule biosynthetic process
## 13                    cellular nitrogen compound metabolic process
## 14                                 regulation of gene expression
## 15                    cellular aromatic compound metabolic process
## 16                                 nucleic acid metabolic process
## 17              regulation of nitrogen compound metabolic process
## 18                               nitrogen compound metabolic process
## 19 regulation of nucleobase-containing compound metabolic process
## 20                               macromolecule biosynthetic process
```

```
# barplot(-log10(tt£Pvalue), names.arg=paste(tt£Term,
# tt£GOBPID), las=2, ylab='-log10 p-value', col='Red')
```

Check to see if there is a DMRT binding site under the DMRT6 Specific Peaks

```
# Use a Chi-Squared test to see how unlikely the distribution
# of sites is
d6ySy <- sum(d6p05DF[subset, "maxsite"] > 0.7)
d6ySn <- sum(subset) - d6ySy
d6nSy <- sum(d6p05DF$maxsite > 0.7) - d6ySy
d6nSn <- nrow(d6p05DF) - d6nSy - d6ySn - d6ySy
contable <- matrix(c(d6ySy, d6nSy, d6ySn, d6nSn), nr = 2, nc = 2)
contable
```

```
##      [,1] [,2]
## [1,] 2770 3235
## [2,] 8215  642
```

```
chisq.test(contable)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity
##  correction
##
## data:  contable
## X-squared = 4032, df = 1, p-value < 2.2e-16
```

```
# Compare In Vivo defined DMRT6 site with In vitro Site
d6summits <- read.table("DM6_dedup_macs14_pe05_summits.bed",
    skip = 0)

# Make 50bp windows around the summit
d6summits <- RangedData(space = d6summits[, 1], IRanges(start = d6summits[,
    2] - 25, end = d6summits[, 3] + 25), strand = "*")

# look for motifs under strong Dmrt6 peaks
sum(d6p05DF$score > 250)
```

```
## [1] 1724

system.time(d6motifs <- GADEM(d6summits[d6p05DF$score > 250,
    ], genome = Mmusculus, weightType = 1, maskR = 1))

##     user   system  elapsed
## 1380.322    2.744  361.488

length(d6motifs@motifList)

## [1] 5

consensus(d6motifs)

## [1] "GmTACwTTGTAkC" "nGGGGGrGGGGn"  "GmwACwGTwrCAr"
## [4] "nrGCwGCTGn"    "TkGCTACAn"

dmrt6.pwm <- getPWM(d6motifs)
pfm.dmrt6 <- new("pfm", mat = dmrt6.pwm[[1]], name = "Dmrt6 Chip-Seq 2014")

plotMotifLogoStack(DNAmotifAlignment(c(pfm.vitro, pfm.dmrt6)))



# look for motifs in DMRT6 peaks that do not have an In vitro
# site sum(d6p05DF£maxsite==0 & d6p05DF£score > 100)
system.time(novel_motifs <- GADEM(d6summits[d6p05DF$maxsite ==
    0 & d6p05DF$score > 100, ], genome = Mmusculus, weightType = 1,
    maskR = 1))

##    user  system elapsed
## 444.928   3.112 118.050


length(novel_motifs@motifList)

## [1] 2

consensus(novel_motifs)

## [1] "rnsrGrrrrrGGGrGGGGGGGGGGGrGGGrrrrrr"
## [2] "yTGyTrCTGTTGCwGy"

novel.pwm <- getPWM(novel_motifs)
novel1.pfm <- new("pfm", mat = novel.pwm[[1]], name = "Novel Site 1")
plotMotifLogo(novel1.pfm)


novel2.pfm <- new("pfm", mat = novel.pwm[[2]], name = "Novel Site 2")
plotMotifLogo(novel2.pfm)
```

Use Ingenuity's Ontology Categories to highlight spermatogenesis genes.

```
# read in ingenuity csv's
fls <- list.files("/mnt/afp/micah/From Vivian to Micah/csv/",
    pattern = "csv$", full = TRUE)
rm(humanEntrez)
```

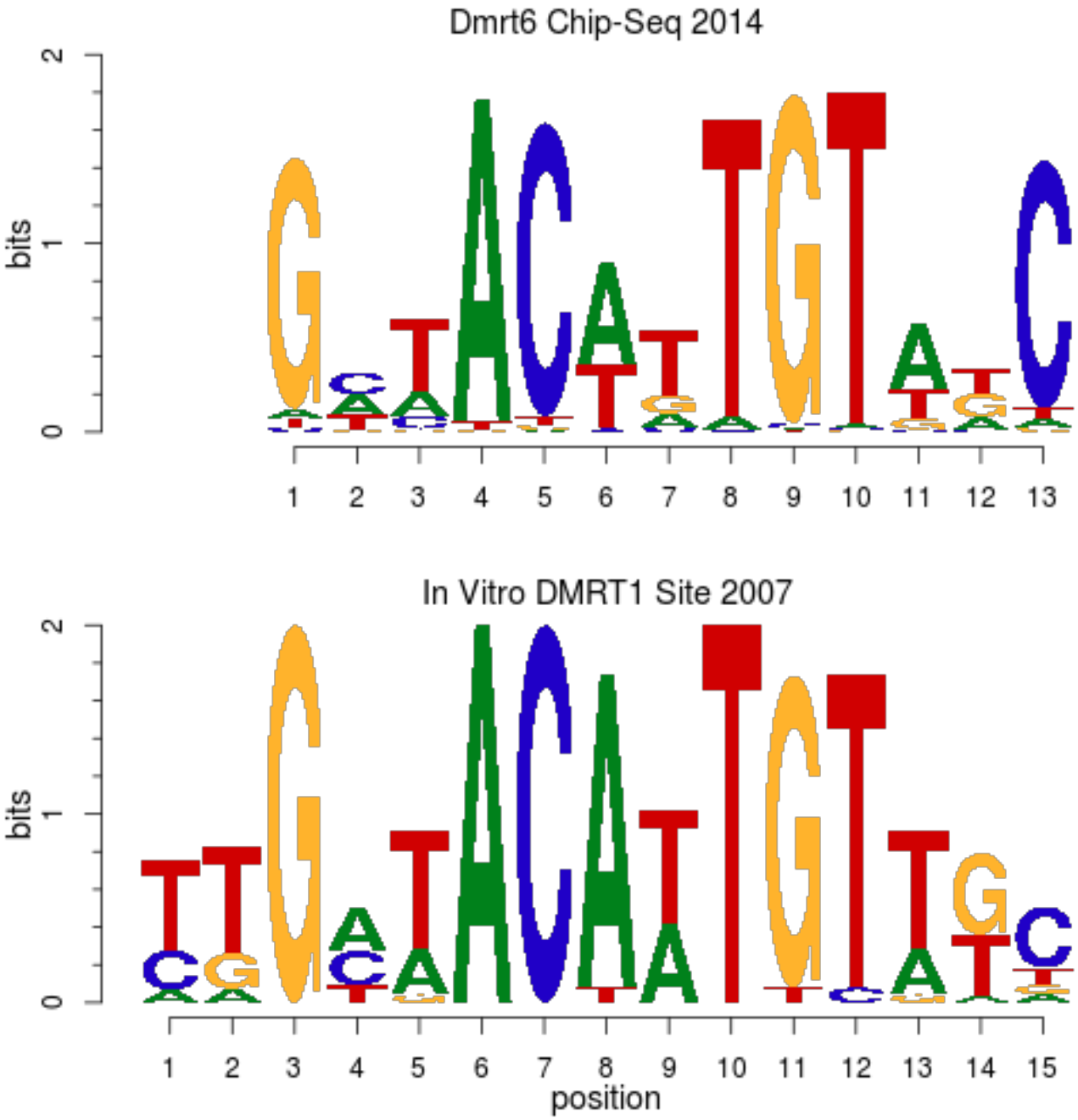

Figure 13: plot of chunk MotifAnalysis

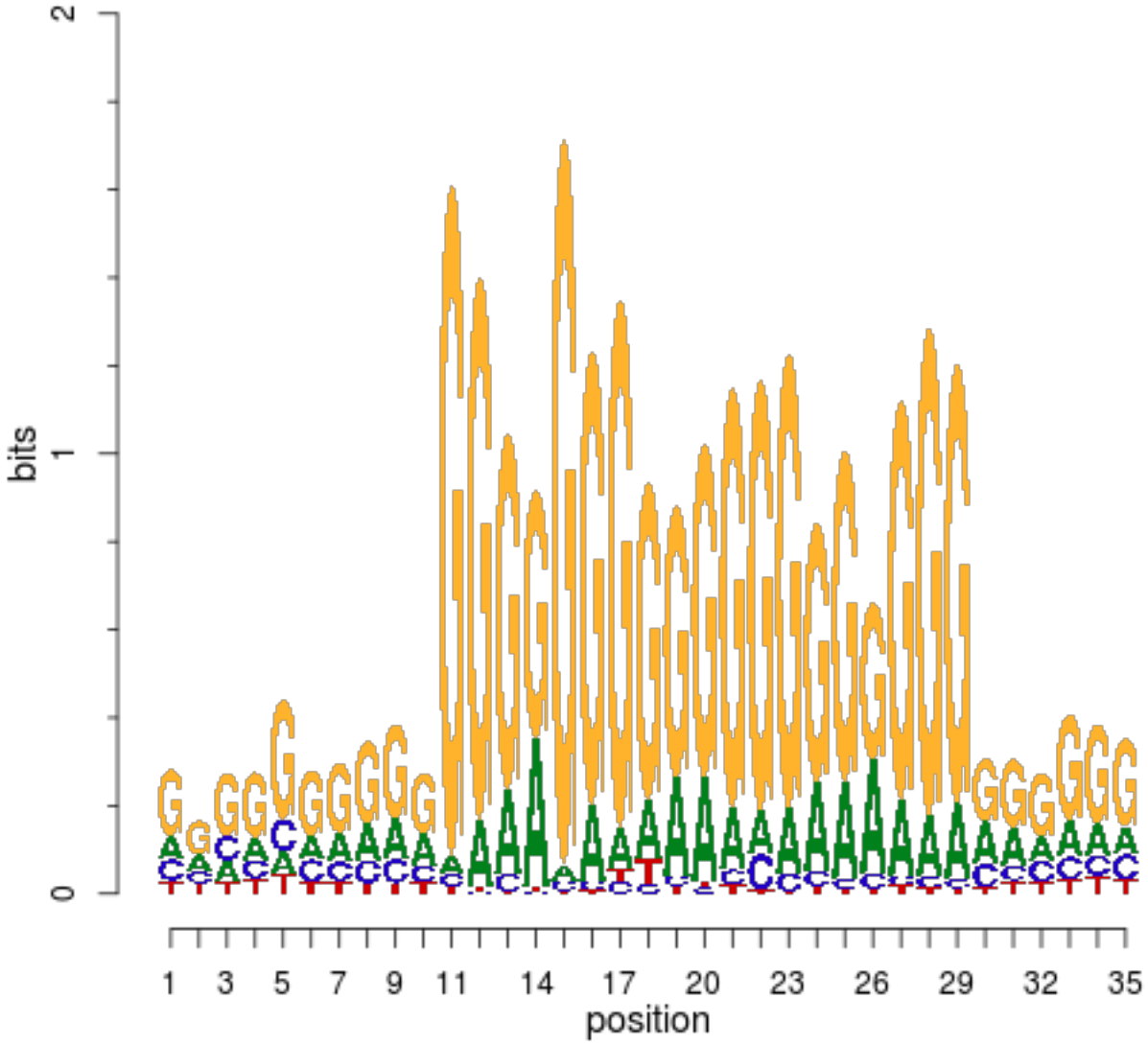

Figure 14: plot of chunk MotifAnalysis

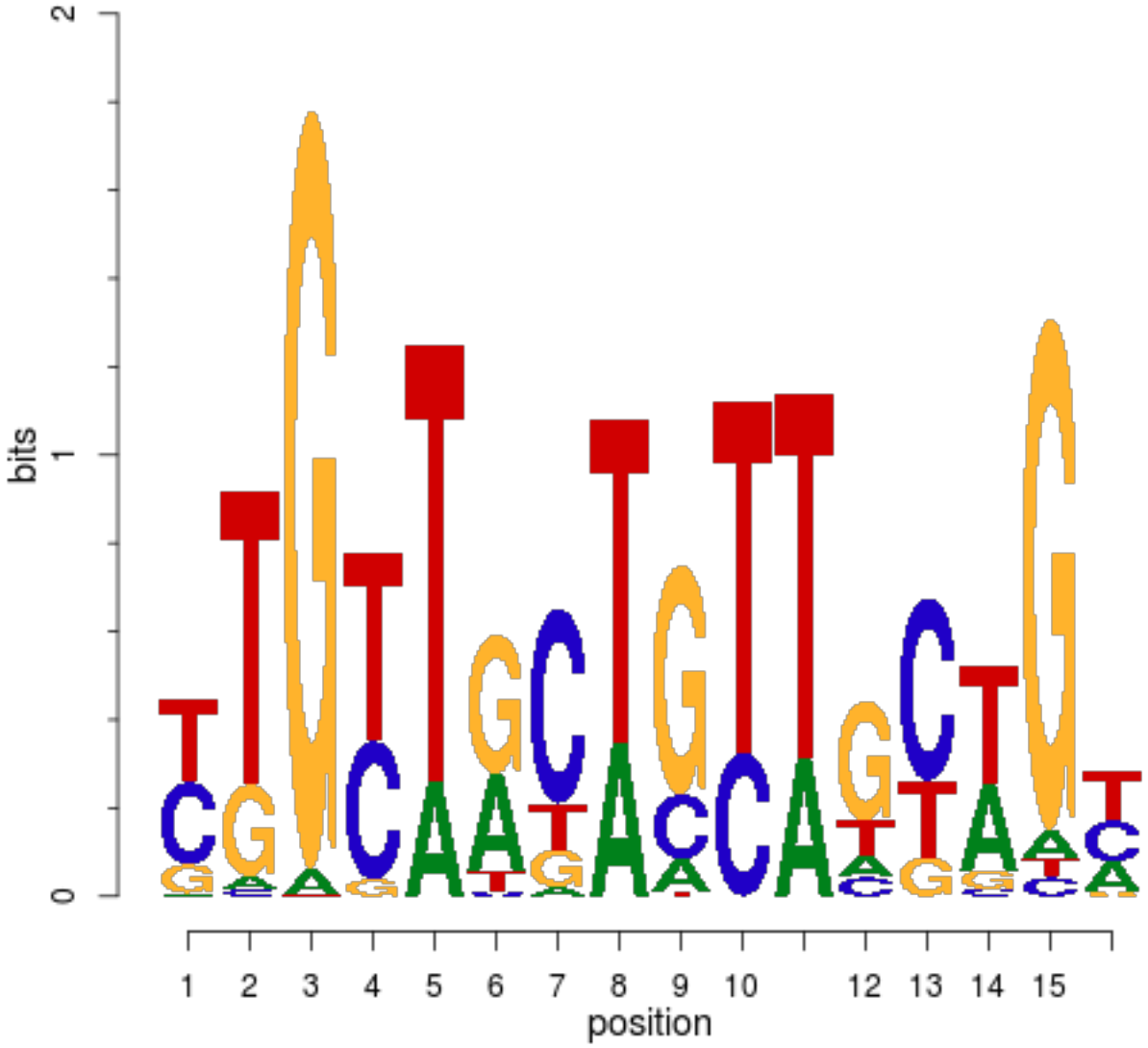

Figure 15: plot of chunk MotifAnalysis

```
## Warning: object 'humanEntrez' not found

rm(mouseEntrez)

## Warning: object 'mouseEntrez' not found

humanEntrez = list()
mouseEntrez = list()
for (i in 1:length(fls)) {
    print(fls[i])
    temp <- read.csv(fls[i], skip = 1, header = T, stringsAsFactors = F)
    human <- temp$Entrez.Gene.ID.for.Human
    human <- human[!is.na(human)]
    human <- unlist(strsplit(as.character(human), "\\|"))
    mouse <- temp$Entrez.Gene.ID.for.Mouse
    mouse <- mouse[!is.na(mouse)]
    mouse <- unlist(strsplit(as.character(mouse), "\\|"))
    humanEntrez[[i]] <- human
    mouseEntrez[[i]] <- mouse
}

## [1] "/mnt/afp/micah/From Vivian to Micah/csv//genes without mouse entrez.csv"
## [1] "/mnt/afp/micah/From Vivian to Micah/csv//ingenuity2 dev of genital organ.csv"
## [1] "/mnt/afp/micah/From Vivian to Micah/csv//ingenuity2 gamet.csv"
## [1] "/mnt/afp/micah/From Vivian to Micah/csv//ingenuity2 germ cell.csv"
## [1] "/mnt/afp/micah/From Vivian to Micah/csv//ingenuity2 gonad.csv"
## [1] "/mnt/afp/micah/From Vivian to Micah/csv//ingenuity2 meiosis.csv"
## [1] "/mnt/afp/micah/From Vivian to Micah/csv//ingenuity2 semiferous.csv"
## [1] "/mnt/afp/micah/From Vivian to Micah/csv//ingenuity2 seminal.csv"
## [1] "/mnt/afp/micah/From Vivian to Micah/csv//ingenuity2 sperm.csv"
## [1] "/mnt/afp/micah/From Vivian to Micah/csv//ingenuity2 testis.csv"

names(humanEntrez) <- c("misc", "dev", "gamet", "germ", "gonad",
    "meiosis", "seminiferous", "seminal", "sperm", "testis")
names(mouseEntrez) <- c("misc", "dev", "gamet", "germ", "gonad",
    "meiosis", "seminiferous", "seminal", "sperm", "testis")

# Add Columns to master tt table names(humanEntrez)
for (i in 1:length(humanEntrez)) {
    print(names(humanEntrez)[i])
    oldcolnames <- colnames(D6tt)
    temp <- D6tt$ensembl_gene_id %in% humanEntrezToMouseEnsemble(humanEntrez[[i]])[,
        2] | D6tt$entrezgene %in% mouseEntrez[[i]]
    D6tt <- cbind(D6tt, temp)
    colnames(D6tt) <- c(oldcolnames, names(humanEntrez[i]))
}

## [1] "misc"
## [1] "dev"
## [1] "gamet"
## [1] "germ"
## [1] "gonad"
## [1] "meiosis"
## [1] "seminiferous"
## [1] "seminal"
## [1] "sperm"
## [1] "testis"
```

Make a table of "Genes of Interest" to validate by QPCR.

```
# Create some Logical variables (decider1-3) to indicate
# whether the gene is 'interesting' Decider1 tells us that it
# is one of the ingenuity categories
decider1 <- D6tt$misc | D6tt$dev | D6tt$gamet | D6tt$germ | D6tt$gonad |
    D6tt$meiosis | D6tt$seminiferous | D6tt$seminal | D6tt$sperm |
    D6tt$testis
sum(decider1)
```

```
## [1] 122
```

```
# decider2 is just the p-value (may be redundant with GLM
# section above)
decider2 <- D6tt$PValue < 0.05
sum(decider2)
```

```
## [1] 1595
```

```
# decider2 <- D6tt£'PValue' <0.05 & !is.na(D6tt£entrezgene)

# We want to only consider genes that are expressed in A's
# and B's or have unknown expression because they weren't on
# the microarray
decider3 <- D6tt$A > 100 | D6tt$B > 100
decider3[is.na(decider3)] <- TRUE
sum(decider3)
```

```
## [1] 903
```

```
# D6tt[D6tt£directTarget & D6tt£germIPA & D6tt£'PValue'
# <0.05,]
D6ttGOI <- D6tt[decider1 & decider2 & decider3, ]
nrow(D6ttGOI)
```

```
## [1] 58
```

```
# run pubmedBatchQuery on interesting genes
D6ttGOI <- cbind(D6ttGOI, pubmedBatchQuery(D6ttGOI$mgi_symbol,
    "Testis"))

D6ttGOI <- D6ttGOI[with(D6ttGOI, order(PValue)), ]
# temp[,c('mgi_symbol','mgi_id','logFC','PValue','A','B','P','R','PubMed')]
```

Output the results

```
D6tt <- D6tt[with(D6tt, order(-logFC)), ]
D6tt[grep("Dmrtb1", D6tt$mgi_symbol), ]
```

```
##     entrezgene    ensembl_gene_id WT_R1 Null_R1 WT_R2 WT_R3
## 395      56296 ENSMUSG00000028610   194      14   510   254
##     Null_R2 Null_R3      mgi_id mgi_symbol chromosome_name
## 395      25      26 MGI:1927125     Dmrtb1               4
##     start_position end_position strand  logFC logCPM    LR
## 395      107348895    107356835     -1 -3.824  2.657 101.5
##        PValue      FDR    qvalue Gene.symbol     A     B
```

```
## 395 7.123e-24 4.05e-20 3.548e-20      Dmrtb1 285.8 554.5
##        P    R d6macs d1macs  misc   dev gamet  germ gonad
## 395 1776 2558   TRUE   TRUE FALSE FALSE FALSE FALSE FALSE
##     meiosis seminiferous seminal sperm testis
## 395   FALSE        FALSE   FALSE FALSE  FALSE

write.table(D6tt, "/mnt/afp/teng/data/Supplementary_Table_1.csv",
    quote = F, row.names = F, sep = ",")
write.table(D6ttGOI, "/mnt/afp/teng/data/Supplementary_Table_2.csv",
    quote = F, row.names = F, sep = ",")
sessionInfo()

## R version 3.1.0 (2014-04-10)
## Platform: x86_64-pc-linux-gnu (64-bit)
##
## locale:
##  [1] LC_CTYPE=en_US.UTF-8 LC_NUMERIC=C
##  [3] LC_TIME=C            LC_COLLATE=C
##  [5] LC_MONETARY=C        LC_MESSAGES=C
##  [7] LC_PAPER=C           LC_NAME=C
##  [9] LC_ADDRESS=C         LC_TELEPHONE=C
## [11] LC_MEASUREMENT=C     LC_IDENTIFICATION=C
##
## attached base packages:
## [1] grid      parallel  stats     graphics  grDevices
## [6] utils     datasets  methods   base
##
## other attached packages:
##  [1] motifStack_1.8.0
##  [2] ade4_1.6-2
##  [3] MotIV_1.20.0
##  [4] grImport_0.9-0
##  [5] rGADEM_2.12.0
##  [6] seqLogo_1.30.0
##  [7] BSgenome.Mmusculus.UCSC.mm9_1.3.99
##  [8] GOstats_2.30.0
##  [9] graph_1.42.0
## [10] Category_2.30.0
## [11] GO.db_2.14.0
## [12] Matrix_1.1-3
## [13] org.Mm.eg.db_2.14.0
## [14] ChIPpeakAnno_2.12.1
## [15] RSQLite_0.11.4
## [16] DBI_0.2-7
## [17] VennDiagram_1.6.5
## [18] rtracklayer_1.24.0
## [19] GEOquery_2.30.0
## [20] XML_3.98-1.1
## [21] biomaRt_2.20.0
## [22] qvalue_1.38.0
## [23] edgeR_3.6.1
## [24] limma_3.20.1
## [25] GenomicAlignments_1.0.1
## [26] BSgenome_1.32.0
## [27] GenomicFeatures_1.16.0
## [28] AnnotationDbi_1.26.0
## [29] Biobase_2.24.0
## [30] Rsamtools_1.16.0
## [31] Biostrings_2.32.0
## [32] XVector_0.4.0
```

```
## [33] GenomicRanges_1.16.3
## [34] GenomeInfoDb_1.0.2
## [35] IRanges_1.22.6
## [36] BiocGenerics_0.10.0
## [37] knitr_1.5
##
## loaded via a namespace (and not attached):
##  [1] AnnotationForge_1.6.1 BBmisc_1.6
##  [3] BatchJobs_1.2         BiocParallel_0.6.0
##  [5] GSEABase_1.26.0       MASS_7.3-33
##  [7] RBGL_1.40.0           RCurl_1.95-4.1
##  [9] Rcpp_0.11.1           annotate_1.42.0
## [11] bitops_1.0-6          brew_1.0-6
## [13] codetools_0.2-8       digest_0.6.4
## [15] evaluate_0.5.5        fail_1.2
## [17] foreach_1.4.2         formatR_0.10
## [19] genefilter_1.46.0     iterators_1.0.7
## [21] lattice_0.20-29       multtest_2.20.0
## [23] plyr_1.8.1            sendmailR_1.1-2
## [25] splines_3.1.0         stats4_3.1.0
## [27] stringr_0.6.2         survival_2.37-7
## [29] tcltk_3.1.0           tools_3.1.0
## [31] xtable_1.7-1          zlibbioc_1.10.0
```

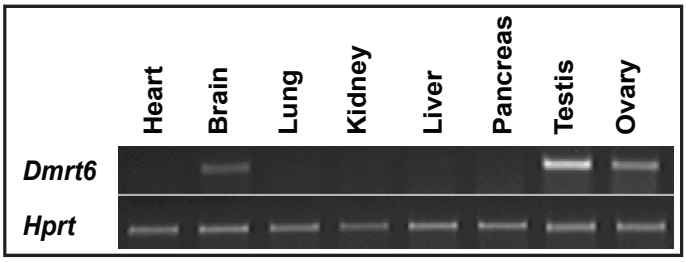

Figure S1 Expression of *Dmrt1* mRNA in adult tissues.
RT-PCR of mRNA from eight tissues detects abundant *Dmrt6* expression in testis and low *Dmrt6* expression in brain and ovary. *Hprt* was used as a positive control.

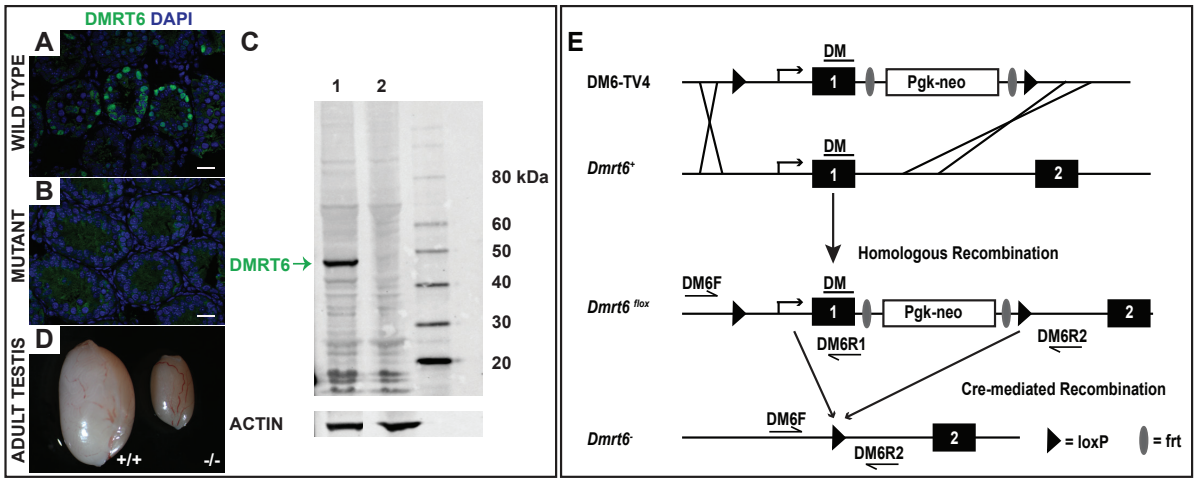

Figure S2 Generation of *Dmrt6* floxed and null alleles.
(A, B) IF staining juvenile (P10) testes with affinity purified DMRT6 polyclonal antibody (see Materials and Methods) showing that *Dmrt6* mutant testes lack detectable DMRT6 protein. Scale bar: 10 μm (C) Western Blot of P14 wild type (lane 1) and *Dmrt6* mutant testes (lane 2). Protein expression is normalized to beta-ACTIN. (D) Seven-week-old wild type and *Dmrt6* mutant testes, showing reduced size of mutant testis. (E) Diagram of gene targeting strategy. Homologous recombination in embryonic stem cells (ES cells) generated the *Dmrt6^{flox}* allele in which exon 1 of *Dmrt6*, which contains the DM DNA binding domain, flanked by *loxP* sites. *Dmrt6^{flox}* also contains a neomycin resistance cassette (*Pgk-neo*) flanked by recognition sites for the Flp recombinase (*frt* sites), allowing its excision. Cre-mediated recombination excised exon 1 and the proximal promoter as well as part of intron 1, generating the putative null allele *Dmrt6^{-}*. We confirmed homologous recombination of both targeting vector arms by Southern blotting (not shown) and also confirmed that mice made from two independently targeted ES cell clones had identical phenotypes. Genotyping primers are indicated.

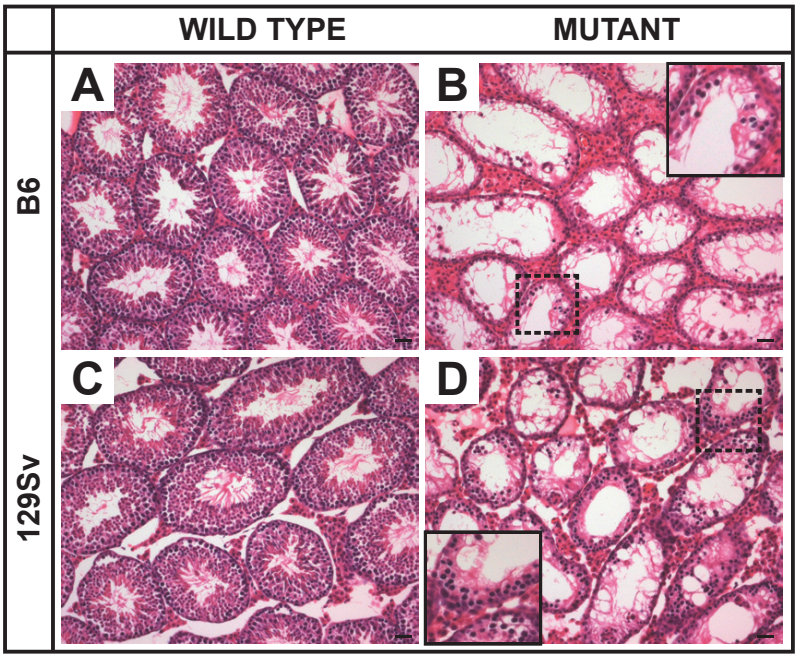

Figure S3 Histology comparison of B6 and 129Sv *Dmrt6* mutant testes.
Hematoxylin/eosin (H&E) staining of seven-week-old wild type (A,C) and *Dmrt6* mutant testes (B,D) on B6 (A,B) or 129Sv (C,D) genetic backgrounds. Insets are higher magnification showing the greater number of meiotic cells in *Dmrt6* mutant testes on the 129Sv background . Scale bars: 20 μm

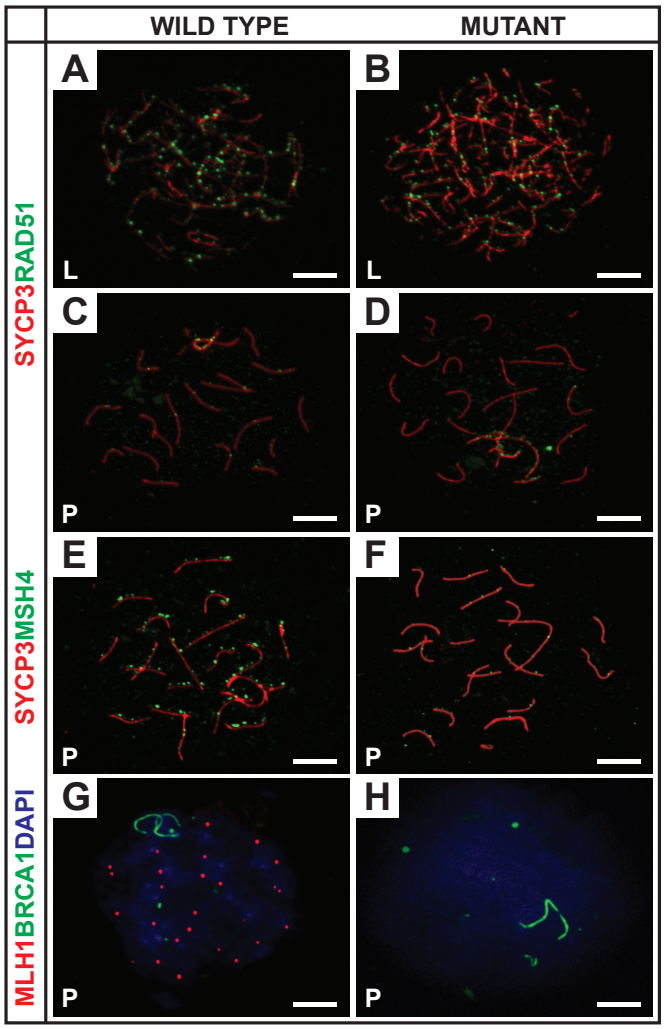

Figure S4 *Dmrt6* mutant spermatocytes can form double strand breaks but do not undergo meiotic recombination.

IF of chromosome spreads from wild type or *Dmrt6* adult leptotene (L) and pachytene (P) spermatocytes. (A-D) Wild type and *Dmrt6* leptotene spermatocytes accumulate foci of RAD51, indicating that double strand DNA breaks form and are repaired normally in *Dmrt6* mutants. (E-H) Wild type but not mutant pachytene spermatocytes accumulate MSH4 positive transitional nodules and MLH1 positive recombination nodules, which are sites of crossing over. BRCA1 marks unpaired sex chromosomes in pachytene spermatocytes. Note that mutant cells with extensive homolog pairing were selected for this figure in order to increase the chance of detecting recombination nodules. Typical mutant cells are shown in Figure 4. Scale bar 10 μm.

Table S1: Genes misexpressed in *Dmrt6* mutant testes.

Download Table S1

Table S2: Genes misexpressed in *Dmrt6* mutant testes selected for spermatogonial expression and function in meiosis or testis development.

Download Table S2

Table S3: DMRT6-associated chromatin regions from ChIP-seq.

Download Table S3

Table S4: Antibodies used in this study.

Primary Antibody		Dilution	Source	Reference
BC7	Rat monoclonal	1:100	Hiromitsu Tanaka	Koshimizu et al. PMID: 7766415
BRCA1	Rabbit polyclonal	1:200	Satoshi Namekawa	Ichijima et al. PMID: 3084029
Brdu	Rat monoclonal (IgG)	1:200	Abcam	Cat. No. ab6326
DMRT1	Goat polyclonal	1:50	Santa Cruz	Cat. No. sc-104885
DMRT6	Rabbit polyclonal	1:200	David Zarkower	This paper
MLH1	Rabbit polyclonal	1:100	Santa Cruz	Cat. No. sc-581
MSH4	Rabbit polyclonal	1:50	Abcam	Cat. No. ab58666
PLZF	Mouse monoclonal (IgG)	1:200	Calbiochem	Cat. No. OP128
RAD51	Rabbit polyclonal	1:200	Calbiochem	Cat. No. PC130
SOHLH1	Guinea pig polyclonal	1:100	Aleksandar Rajkovic	Pangas et al. PMID: 1472434
SOHLH2	Guinea pig polyclonal	1:200	Aleksandar Rajkovic	Ballow et al. PMID: 16564520
STRA8	Rabbit polyclonal	1:200	Abcam	Cat. No. ab15092
SOX9	Rabbit polyclonal	1:200	Millipore	Cat. No. AB5535
SUMO-1	Mouse monoclonal (IgG)	1:200	Zymed Laboratories	Cat. No. 33-2400
SYCP1	Rabbit polyclonal	1:200	Abcam	Cat. No. ab15090
SYCP3	Mouse polyclonal	1:200	Abcam	Cat. No. ab96672
TRA98	Rat monoclonal (IgG)	1:200	Bio Academia	Cat. No. 73-003 PMID: 9568529

Secondary Antibody		Dilution	Source	Reference
Anti Goat IgG Alexa Fluor 594	Donkey polyclonal	1:500	Invitrogen	Cat. No. A11058
Anti Guinea pig Alexa Fluor 488	Goat polyclonal	1:500	Invitrogen	Cat. No. A11073
Anti Mouse IgG Alexa Fluor 568	Goat polyclonal	1:500	Invitrogen	Cat. No. A11004
Anti Rabbit Alexa Fluor 488	Goat polyclonal	1:500	Invitrogen	Cat. No. A11008
Anti Rabbit Alexa Fluor 488	Donkey polyclonal	1:500	Invitrogen	Cat. No. A21206
Anti Rat Alexa Fluor 594	Goat polyclonal	1:500	Invitrogen	Cat. No. A11007