Genome-wide identification of regulatory elements in Sertoli cells


This is a copy of the final article, which is published here by permission of the publisher and which appeared first in:

Development
2017 FEB 15 : 144(4): 720-730
doi: 10.1242/dev.142554
Publisher: The Company of Biologists Ltd.

© 2017 The authors
Techniques and Resources

ABSTRACT

A current goal of molecular biology is to identify transcriptional networks that regulate cell differentiation. However, identifying functional gene regulatory elements has been challenging in the context of developing tissues where material is limited and cell types are mixed. To identify regulatory sites during sex determination, we subjected Sertoli cells from mouse fetal testes to DNase-seq and ChIP-seq for H3K27ac. DNASel-seq identified putative regulatory sites around genes enriched in Sertoli and pregranulosa cells; however, active enhancers marked by H3K27ac were enriched proximal to only Sertoli-enriched genes. Sequence analysis identified putative binding sites of known and novel transcription factors likely controlling Sertoli cell differentiation. As a validation of this approach, we identified a novel Sertoli cell enhancer upstream of Wt1, and used it to drive expression of a transgenic reporter in Sertoli cells. This work furthers our understanding of the complex genetic network that underlies sex determination and identifies regions that potentially harbor non-coding mutations underlying disorders of sexual development.

KEY WORDS: Genome-wide enhancer profiling, Sex determination, Sertoli cell development, Mouse

INTRODUCTION

During development, multipotent progenitor cell commitment to one cell fate versus another is driven by complex gene networks controlled by diverse regulatory factors, including transcription factors (TFs), chromatin remodelers and non-coding RNAs (ncRNAs). Although identification of the cohorts of genes that drive cell fate determination is an important focus of developmental studies, it is equally important to identify the regulatory sites through which these factors operate. Mutations of regulatory sites in enhancers and promoters that alter gene expression have been discovered serendipitously, but until recently, an unbiased genome-wide approach to identify regulatory elements in specific fetal cell types remained difficult. Thus, it has not been clear how the accessibility of chromatin sites and the availability of TFs act together to control cell fate decisions during fetal development from a pluripotent to a committed fate.

The developing gonad is a powerful model system for understanding the gene regulatory mechanisms that drive cell fate determination. The process of sex determination occurs in gonad progenitor cells whose differentiation controls the sexual fate of the organism. Although genetic sex (XY/XX) is established upon fertilization, these innate chromosomal differences do not produce phenotypic differences in male or female early embryos (Eggers and Sinclair, 2012). Instead, sexual dimorphism is initiated in the early gonad supporting cells, a single cell lineage that differentiates into Sertoli (XY) (Sekido et al., 2004) or pregranulosa (XX) cells (Albrecht and Eicher, 2001; Mork et al., 2012). Prior to sex determination, XX and XY supporting cells are bipotential and express nearly all genes at similar levels, with minor differences limited to genes on the X and Y chromosomes (Nef et al., 2003; Munger et al., 2013).

In the mouse, male sex determination is initiated around embryonic day 11.5 (E11.5) by expression of the Y-linked gene Sry, which diverts XY supporting cells to the Sertoli cell fate (Fig. 1A). Sry is transiently expressed and many SRY-binding sites are subsequently bound by its downstream target SOX9 (Li et al., 2014). SRY and SOX9 control the differentiation of Sertoli cells by triggering a dramatic transcriptional reprogramming of bipotential progenitor cells within just 24 h, leading to the upregulation of over 200 genes important for Sertoli cell development and downregulation of ~100 pregranulosa cell-expressed genes that were expressed at the bipotential stage (Munger et al., 2013). In the absence of Sry, XX supporting cells differentiate to pregranulosa cells, but in contrast to the dramatic transcriptional changes that accompany Sertoli cell specification, pregranulosa cell differentiation proceeds with fewer transcriptional changes consistent with the finding that undifferentiated XX and XY supporting cells are initially transcriptionally biased toward ovarian fate (Jameson et al., 2012b).

Although SRY acts as a pivotal switch in the sex-determining pathway, mutations in SRY account for only ~10-15% of XY disorders of sex development (DSDs) (Cameron and Sinclair, 1997). Over 35 other genes have been identified that contribute to DSDs (Arboleda and Vilain, 2011; Arboleda et al., 2013), suggesting that a complex genetic network controls fate commitment in the early gonad. It is evident from human DSD cases where gene dose is altered by duplications or haploinsufficiencies (Foster et al., 1994; Muscatelli et al., 1994; Zanaria et al., 1994; Jordan et al., 2001) that perturbations to the level or timing of individual genes within this network can lead to sex reversals. In the C57BL/6J inbred mouse strain, the altered expression of several sex-determining genes causes unique strain-dependent sex-reversal phenotypes in response to specific mutations (Colvin et al., 2001; Bouna et al., 2005; Kim et al., 2007b; Munger et al., 2009, 2013; Correa et al., 2012; Warr et al., 2012). These studies point to the importance of precise spatial and temporal gene...
regulation. However, our lack of knowledge regarding the locations of crucial gene-regulatory sites limits our ability to study the regulatory mechanisms that control gene expression in Sertoli cells.

To identify putative regulatory sites in Sertoli cells, we performed DNaseI-seq on purified Sertoli cells from fetal testes just after sex determination. Comparison with DNaseI-seq data from a variety of other mouse tissues and cell types revealed thousands of novel Sertoli cell-specific putative regulatory elements, which are enriched near genes expressed in Sertoli cells, as well as genes that are silent in Sertoli cells, but expressed in the alternate pregranulosa cell lineage. This suggests that DNaseI-seq identified sites involved in the activation and repression of genes required for Sertoli cell fate commitment. Using ChIP-seq for H3K27ac, we located active enhancer elements specifically enriched near genes expressed in Sertoli cells, and used this data to identify TF motifs that distinguish active from inactive enhancers. Finally, we validated the enhancer activity of an active regulatory element located 50 kb upstream of Wt1, which encodes a TF essential for early gonad development. This data will enable future experiments to characterize the regulatory network driving Sertoli cell development and may aid in the identification of non-coding human mutations underlying unexplained cases of DSDs, the majority of which cannot be attributed to mutation within the coding region of a known sex-determining gene (Arboleda and Vilain, 2011; Arboleda et al., 2013).

RESULTS
Adapting DNaseI-seq to developmentally limited cell populations

We conducted a proof-of-principle experiment to establish the efficacy of DNaseI-seq in identifying regulatory elements in Sertoli cells that had been purified by fluorescence-activated cell sorting (FACS) followed by long-term storage at −80°C. E15.5 Sertoli cells were isolated by FACS based on the Sertoli cell-specific expression of the Sox9-CFP transgene (Fig. 1A). Significantly reducing the starting material required, we found that 5×10⁶ cells routinely yielded enough chromatin to perform the DNaseI-seq assay. DNaseI-seq identified over 81,000 DNaseI hypersensitive sites (DHSs) in E15.5 Sertoli cells. Using previously published methods...
Identification of DHSs important for gene regulation in Sertoli cells

Given the success of this experiment, we performed DNaseI-seq on E13.5 Sertoli cells, a stage of development for which we have extensive expression data (Nef et al., 2005; Bouma et al., 2007a, Jameson et al., 2012b) (Table S2). Our E13.5 DNaseI-seq data showed similar properties to the data from E15.5 Sertoli cells; ∼71% of the DHSs were shared between both data sets and had nearly identical genomic distributions (Fig. S2A,B).

To identify the DHSs most relevant for gene expression during testis development, we conducted DNaseI-seq on six other mouse samples [primary skin fibroblasts, embryonic stem cells (ESCs), adult kidney, liver, heart and brain]. To identify putative enhancers important for gene expression in Sertoli cells, we used the six other somatic cell types to classify E13.5 and E15.5 Sertoli cell DHSs as unique (in only Sertoli cells), shared (in Sertoli cells and in one to five additional cell types) or common (in all cell types). For example, at the Sox9 promoter, all cell types had a DHS (common DHS); however, the TES enhancer was found only in the E13.5 and E15.5 Sertoli cell datasets (unique DHS) (Fig. 2A). Twenty-seven percent (21,743) of all E15.5 Sertoli cell DHSs and 25% (21,863) of E13.5 Sertoli cell DHSs were common, with promoters accounting for over half of the common sites (Fig. 2B; Fig. S2C). CTCF binding sites are highly conserved across different cell types (Kim et al., 2007a; Xi et al., 2007; Song et al., 2011) and, consistent with this, CTCF-binding sites found in eight other cell types (Shen et al., 2012) were enriched in the common DHSs (Fig. 2C). In fact, of the common sites not overlapping a promoter, ∼75% overlapped a CTCF binding site. The remaining DHSs were unique (E15.5: 29%; 23,548, E13.5: 32%, 28,133) to the Sertoli cell lineage or shared (E15.5: 44%; 36,176, E13.5: 43%, 38,111). A high proportion of unique (85%) and shared (73%) sites localized to intergenic or intronic regions (Fig. 2B; Fig. S2B), and are likely to be important regulators of gene expression in Sertoli cells.

Sertoli cell DHSs associate with Sertoli and pregranulosa cell-expressed genes, whereas active enhancers are enriched near Sertoli cell genes

The DHSs identified by DNaseI-seq correlate with several genomic features, including promoters, enhancers, repressors and insulators (Heintzman et al., 2007, 2009; Boyle et al., 2008a). To focus on regions of the genome driving Sertoli cell-specific gene expression, we conducted chromatin profiling for H3K27ac, a strong predictor of active enhancer activity (Creighton et al., 2010; Rada-Iglesias et al., 2011). H3K27ac ChIP-seq on E13.5 Sertoli cells identified >28,000 H3K27ac Peaks, ~70% of which overlapped a DHS in our E13.5 Sertoli cell dataset (Fig. 3A, Fig. S4).

We predicted that active enhancers would be enriched around Sertoli cell-expressed genes. To test this, the enrichment of DHSs, or H3K27ac-positive DHSs, was examined around Sertoli cell-expressed genes and compared against genes expressed in pregranulosa and germ cells (Jameson et al., 2012b). Using a nearest-neighbor approach (McLean et al., 2010), we assigned

Fig. 2. Identification of DHSs unique to the Sertoli cell lineage. (A) Comparison of the genomic region around Sox9 in E15.5 Sertoli cells versus other cell types (fibroblast, ESC) and tissues (kidney, liver, heart, brain). The Sox9-CFP transgene is driven by the 3 kb TES enhancer (blue bar) located 13 kb upstream from the Sox9 TSS. Enhancer activity was previously narrowed to a subregion of TES: the 1.3 kb TESCO element (Sekido and Lovell-Badge, 2008) (light-blue bar). DNaseI-seq identifies a DHS (black bar) spanning the TESCO element that is unique to the Sertoli cell lineage, consistent with the expression pattern of the transgene. Black bars above the track denote DHSs; Parzen scores are shown to the left of each track. (B) Based on their cell type specificity, E15.5 Sertoli cell DHSs were divided into unique, shared or common. Each group was subcategorized based on genomic location (promoter, gray; intergenic, purple; intronic, green; exonic, orange). (C) The percentage of DHSs overlapping CTCF sites based on the average overlap with eight previously generated CTCF ChIP-seq data sets (Shen et al., 2012).
DHSs to one or two neighboring genes (Fig. 3A) and the number of DHSs (Fig. 3B,C) or H3K27ac-positive DHSs (Fig. 3D) was counted around genes enriched in these cell types. As expected, common DHSs were not preferentially enriched around Sertoli-, pregranulosa- or germ cell-expressed genes (Fig. 3B). However, there was a significant increase in the number of cell type-specific (unique+shared) DHSs neighboring both Sertoli and pregranulosa cell-expressed genes (Fig. 3C). The similar enrichment of DHSs around Sertoli and pregranulosa cell-expressed genes may reflect the shared developmental origin of the supporting cell lineages, which have similar expression profiles just a few days prior to our analysis (E10.5) (Nef et al., 2005; Munger et al., 2013). DHSs around pregranulosa cell-expressed genes may define sites where repression complexes are bound to silence the alternative female pathway upon Sertoli cell differentiation.

H3K27ac-positive DHSs, which are indicative of active enhancers, were significantly enriched specifically in neighboring Sertoli cell-expressed genes (Fig. 3D), as expected. Additionally, we observed no enrichment of cell type-specific DHSs or H3K27ac-positive DHSs near germ cell-expressed genes, consistent with the unrelated origin of this cell lineage. Thus, using DNasel-seq, we identified thousands of regulatory sites that are likely to be critical regulators of gene expression during Sertoli cell differentiation. Additionally, ChIP-seq for H3K27ac classified these into active enhancers (associated with actively-expressed genes; H3K27ac-positive DHSs) or inactive enhancers (DHS-positive only) (Table S2).

Sertoli cell DHSs are enriched for TFs important for supporting cell development

To identify the TFs that bind putative Sertoli cell regulatory sites, we conducted computational analysis of the sequences in the Sertoli cell unique DHSs that did not overlap a promoter. We counted 6-mers identified within Sertoli cell unique DHSs that are highly predictive of whether a region will be open specifically in Sertoli cells compared with flanking regions [area under receiver operating characteristic curve (AuROC)=0.82] for H3K27ac classified these into active enhancers (associated with actively-expressed genes; H3K27ac-positive DHSs) or inactive enhancers (DHS-positive only) (Table S2).

Sertoli cell DHSs are enriched for TFs important for supporting cell development

To identify the TFs that bind putative Sertoli cell regulatory sites, we conducted computational analysis of the sequences in the Sertoli cell unique DHSs that did not overlap a promoter. We counted 6-mers in these DHSs, and used these as features in an L1 logistic regression classifier to classify these regions from length-matched sites flanking these DHSs (Koh et al., 2007) (Fig. 4A). A similar approach was shown to aid in the identification of sequence features underlying enhancer function in other mammalian systems (Lee et al., 2011). Our method is designed to identify enriched 6-mers found in a substantial fraction of DHSs. However, similar to other motif enrichment approaches, it is possible that it does not identify TFs that play a significant role in Sertoli cell biology, but bind to only a few locations.

Our analysis revealed 305 6-mers identified within Sertoli cell unique DHSs that are highly predictive of whether a region will be open specifically in Sertoli cells compared with flanking regions [area under receiver operating characteristic curve (AuROC)=0.82].
To rule out the possibility of high classification accuracy due to repeats in flanking regions, we trained a classifier to distinguish Sertoli cell DHSs from a union set of unique DHSs from other cell types, including ES and fibroblast cells, and heart, kidney, liver and brain tissue. The classifier again showed high classification accuracy (AuROC=0.87). Furthermore, the 6-mers identified are similar between the two classification tasks, indicating that the features identified are robust predictors of Sertoli cell-specific DHSs (Table S3A,D).

We used TOMTOM (Bailey et al., 2009) to identify TFs that match these 6-mer motifs and found 75 6-mers matching known motifs. As multiple TFs have similar motifs, we used Sertoli cell gene expression (Jameson et al., 2012b) to identify the most likely candidates. The 6-mer most predictive of open chromatin in Sertoli cells matched the TF-binding motif for SF1 (Fig. 4C, Table S3), a factor required for early gonad formation in both sexes (Luo et al., 1994; Baba et al., 2014). Other prominent motifs identified include the SOX motif, which reflects the crucial role of SOX9 and SOX8 in the divergence of the Sertoli cell lineage from the undifferentiated progenitor cells (Chaboissier et al., 2004; Barrionuevo and Scherer, 2010), and the GATA motif, likely representing GATA4 (which is involved in early gonadal differentiation in both sexes) (Bouma et al., 2007b; Manuylov et al., 2008). Matches to SF1, SOX9 and GATA4 motifs were also found to be statistically enriched in Sertoli cell-specific DHSs, supporting the validity of our method (Table S3E). Although less predictive of open chromatin, the FOX and TCF/LEF motifs were also identified (Table S3). FOXL2 (Schmidt et al., 2004; Ottolenghi et al., 2005) and Wnt signaling (which modulates gene expression through β-catenin binding with TCF/LEF factors) (Vainio et al., 1999; Chassot et al., 2008; Liu et al., 2009) are female-specific factors that are important for pregranulosa cell differentiation; enrichment of these motifs in Sertoli cell open chromatin may indicate sites requiring repression during Sertoli cell differentiation. Also identified was ELF1, an Ets-related TF. Ets family TFs can act downstream of FGF signaling (Sharrocks, 2001), which is crucial for repressing Wnt signaling to reinforce testis development (Kim et al., 2006; Jameson et al., 2012a). Additionally, we identified 6-mers matching TFs that are as yet not known to play a role in Sertoli cells, including TEAD1 and ZBTB33, as well as 6-mers that show no similarity to any known TF-binding motif (Table S3).

To determine whether distinct sequence features could differentiate active from inactive enhancers, we again used the L1-logistic regression classifier to classify the active enhancers (H3K27ac-positive DHS) from inactive DHSs (H3K27ac-negative DHS). The performance of the classifier was significantly reduced compared (Fig. 4B). To rule out the possibility of high classification accuracy due to repeats in flanking regions, we trained a classifier to distinguish Sertoli cell DHSs from a union set of unique DHSs from other cell types, including ES and fibroblast cells, and heart, kidney, liver and brain tissue. The classifier again showed high classification accuracy (AuROC=0.87). Furthermore, the 6-mers identified are similar between the two classification tasks, indicating that the features identified are robust predictors of Sertoli cell-specific DHSs (Table S3A,D).

We used TOMTOM (Bailey et al., 2009) to identify TFs that match these 6-mer motifs and found 75 6-mers matching known motifs. As multiple TFs have similar motifs, we used Sertoli cell gene expression (Jameson et al., 2012b) to identify the most likely candidates. The 6-mer most predictive of open chromatin in Sertoli cells matched the TF-binding motif for SF1 (Fig. 4C, Table S3), a factor required for early gonad formation in both sexes (Luo et al., 1994; Baba et al., 2014). Other prominent motifs identified include the SOX motif, which reflects the crucial role of SOX9 and SOX8 in the divergence of the Sertoli cell lineage from the undifferentiated progenitor cells (Chaboissier et al., 2004; Barrionuevo and Scherer, 2010), and the GATA motif, likely representing GATA4 (which is involved in early gonadal differentiation in both sexes) (Bouma et al., 2007b; Manuylov et al., 2008). Matches to SF1, SOX9 and GATA4 motifs were also found to be statistically enriched in Sertoli cell-specific DHSs, supporting the validity of our method (Table S3E). Although less predictive of open chromatin, the FOX and TCF/LEF motifs were also identified (Table S3). FOXL2 (Schmidt et al., 2004; Ottolenghi et al., 2005) and Wnt signaling (which modulates gene expression through β-catenin binding with TCF/LEF factors) (Vainio et al., 1999; Chassot et al., 2008; Liu et al., 2009) are female-specific factors that are important for pregranulosa cell differentiation; enrichment of these motifs in Sertoli cell open chromatin may indicate sites requiring repression during Sertoli cell differentiation. Also identified was ELF1, an Ets-related TF. Ets family TFs can act downstream of FGF signaling (Sharrocks, 2001), which is crucial for repressing Wnt signaling to reinforce testis development (Kim et al., 2006; Jameson et al., 2012a). Additionally, we identified 6-mers matching TFs that are as yet not known to play a role in Sertoli cells, including TEAD1 and ZBTB33, as well as 6-mers that show no similarity to any known TF-binding motif (Table S3).

To determine whether distinct sequence features could differentiate active from inactive enhancers, we again used the L1-logistic regression classifier to classify the active enhancers (H3K27ac-positive DHS) from inactive DHSs (H3K27ac-negative DHS). The performance of the classifier was significantly reduced compared...
with our previous classifiers (AuROC=0.67), indicating that, despite the fact that active enhancers drive expression in Sertoli cells, sequence features between active enhancers and inactive DHSs are more similar than those between DHS and flanking regions (Fig. 4B). Inspection of the 6-mers that facilitate prediction of one class versus the other revealed 80 6-mers predictive of active enhancers and 64 6-mers predictive of inactive DHSs, many of which are not associated with a known TF. However, of the 6-mers predictive of an active enhancer, 25 did match a known motif. DBP, GATA4, FOXL1 and other FOX factor motifs were among the most predictive motifs for active enhancers (Fig. 4D). Although GATA4 is a well-known driver of Sertoli cell development (Bouma et al., 2007b), the role of DBP is unknown. Our identification of FOX motifs is consistent with a role of FOXL2 in repression of a Sox9 enhancer (TESCO) in granulosa cells after birth (Uhlenhaut et al., 2009) and suggests that FOXL2 may play a similar role earlier in fetal development (see Discussion). Alternatively, several other FOX factors are expressed in both Sertoli and granulosa cells (Jameson et al., 2012b) that have yet to be characterized and could bind these sites.

In addition, the same analysis identified 20 6-mers with matching motifs that were more predictive of inactive DHSs compared with active enhancers. Although we did not identify motifs associated with known pregranulosa cell-specific TFs, the H3K27ac-negative DHSs likely represent a diversity of open chromatin sites, potentially associated with many genes that are repressed or poised in Sertoli cells, with or without sex-specific functions (Fig. 4E). Interestingly, some of the highest scoring 6-mers in this group were not associated with known binding factors.

The motif for another TF important for Sertoli cell development, DMRT1, did not come up as predictive of open chromatin in this analysis. However, we used available ChIP-seq data from E13.5 whole testes to analyze the enrichment of DMRT1-bound sites in Sertoli cell open chromatin (Matson and Zarkower, 2012). We found that 70% of DMRT1 sites overlapped a DHS in our Sertoli cell dataset. However, the majority of DMRT1 sites overlapped promoters, with only 30% overlapping Sertoli cell unique and shared sites (Fig. S3). The exclusion of promoters from our 6-mer motif analysis may account for the absence of DMRT1 enrichment in DHSs analyzed.

**DNaasel-Seq identifies a novel enhancer of Sertoli cell gene expression near Wt1**

With the exception of the TESCO element, few distal regulatory elements have been identified for sex-determining genes. Our dataset provides an excellent resource with which to identify enhancers driving the expression of genes important in Sertoli cell differentiation. To demonstrate this, we performed a transient transgenic assay with a putative active enhancer of the *Wt1* gene (Fig. 5A,B). *Wt1* is initially expressed in multiple embryonic tissues, including the mesonephros and XX and XY gonadal somatic cells. Mutation of *Wt1/WT1* in mice and humans leads to kidney and genitourinary defects, including disorders of sexual development (Kreidberg et al., 1993; Eggers and Sinclair, 2012). After sex determination, *Wt1* expression becomes restricted to Sertoli cells in the testis and remains active in most somatic cells in XX gonads (Pelletier et al., 1991; Armstrong et al., 1993; Maatouk et al., 2012). We identified a DHS peak located 50 kb upstream of *Wt1* that was unique to E13.5 and E15.5 Sertoli cells compared with the other cell types we examined, and marked by H3K27ac, suggesting that it functions as an active enhancer. Finally, this region contained matches to several of the 6-mers identified in our motif analysis, including binding sites for SF1, SOX and FOX motifs.

Analysis of in vivo enhancer activity using a transgenic β-galactosidase reporter showed that the putative *Wt1* enhancer

---

**Fig. 5. Transient transgenic analysis of a DHS unique to Sertoli cells identifies a new Sertoli cell enhancer upstream of Wt1.** (A) DNaasel-seq and H3K27ac ChIP-Seq data around the *Wt1* locus. Gene names indicate the TSS of the transcript. Nearby genes are indicated in gray. DNaasel-Seq data (Parzen score) and peak calls (boxes above track) are in light blue; H3K27ac ChIP-seq data and peak calls are in dark blue. A box surrounds the DHS chosen for in vivo analysis. (B) Detailed view of the DHS showing the DNaasel-seq Parzen score, the smoothed base counts (light blue) and the H3K27ac data (dark blue). The light- and dark-blue bars indicate the DNaasel-seq and H3K27ac peaks, respectively. The black bar marks the region cloned upstream of an *hsp68-LacZ* reporter cassette. (C) An E13.5 testis from a transgenic embryo showed β-galactosidase expression (*TgWt1*, green) specifically in Sertoli cells. Sertoli cells were labeled by AMH immunostaining (red) and germ cells by CDH1 (blue). The confocal image was taken using a 20× objective. Scale bar: 100 μm. (D) Confocal image of the sample in C taken with a 40× objective. Scale bar: 12.5 μm. All three panels show germ cells (blue); the left panel shows Sertoli cells (red); the middle panel shows β-galactosidase expression (green); the right panel shows the merge (yellow indicates co-expression of β-galactosidase and AMH).
was active in Sertoli cells of the testis (6/7 transient transgenic XY embryos; Fig. 5C, D and Fig. S5), where it exclusively localized in cells expressing the Sertoli cell marker AMH. Expression was also detected in somatic cells of the ovary (11/12 XX embryos; Fig. S5). However, no expression was detected in the mesonephros, embryonic kidney, heart or brain (Fig. 5C, Fig. S5 and data not shown) consistent with the cell-type specificity of this enhancer in our Sertoli cell datasets. These results demonstrate the power of our approach for the identification of distal enhancers that activate transcription in Sertoli cells.

**DISCUSSION**

**Identification of Sertoli cell gene regulatory sites**

Understanding the genetic networks that underlie developmental processes requires a comprehensive understanding of the genes involved as well as the regulatory mechanisms that modulate the expression of these genes. This is among a small number of studies to adapt epigenetic approaches, including DNAseI-seq and ChIP-seq, to address the resolution of cell fate in a cell type isolated from a complex developing tissue.

To discover cis-regulatory sequences during sex determination, we used DNAseI-seq to identify open chromatin sites and H3K27ac ChIP-seq to identify active enhancers in Sertoli cells. Using these data, we discovered TFs that might bind to these sites. We identified a novel active enhancer unique to Sertoli cells upstream of Wt1, and showed that it functions as a specific Sertoli cell enhancer in transgenic animals, as predicted. These data will enable future studies of cis-regulatory elements in Sertoli cells.

Although we were able to reduce the number of cells required for DNAseI-seq to 5 million cells, collection required more than 1 year, considering that a pair of E13.5 fetal testes contains only 10–20,000 Sertoli cells. Unfortunately, our attempts to perform DNAseI-seq on pregranulosa cells failed, possibly owing to the longer storage times required to collect sufficient numbers of cells (our pregranulosa cell reporter enabled purification of only 2–4000 cells per embryo). Rather than identifying sites of open chromatin based on DNAseI-seq, a new technique, ATAC-seq, uses a transpososome-based method where the transposome preferentially integrates into sites of open chromatin (Buenrostro et al., 2013) and is highly correlative with DNAseI-seq data. Working on as few as 500 cells, this method presents an attractive alternative to enable studies of chromatin during cellular differentiation. Future studies will address whether a common chromatin architecture is maintained in differentiated XX and XY supporting cells, or whether these sites are eventually remodeled in a sexually dimorphic manner as differentiation proceeds.

**DHSs adjacent to pregranulosa cell genes might indicate TF-mediated repression in Sertoli cell fate maintenance**

In multipotent progenitor cell types, many developmentally regulated genes exist in an epigenetically poised state. Cis-regulatory elements of poised genes are marked by H3K4me1 and open chromatin (Heintzman et al., 2007). Upon differentiation, regulatory sites for active genes gain H3K27ac (Creyghton et al., 2010; Rada-Iglesias et al., 2011) and retain open chromatin, while repressed genes either remain poised or exhibit a loss of H3K4me1 and open chromatin (Whyte et al., 2012; Lara-Astiaso et al., 2014), perhaps rendering regulatory sites refractory to access by TFs.

In bipotential progenitor cells of the gonad, it is currently unknown whether genes that become sex-specific exist in an epigenetically poised state prior to sex determination. However, we hypothesized that, by E13.5, when sex-specific gene expression patterns are well-established, open chromatin would be enriched near genes actively expressed in Sertoli cells. Although the enrichment of H3K27ac-positive active enhancers was associated with only Sertoli cell-expressed genes as expected, an intriguing finding was the widespread enrichment of DHSs around both Sertoli and pregranulosa cell-expressed genes in E13.5 differentiated Sertoli cells. There is evidence that many genes associated specifically with ovarian development are expressed in both sexes at the bipotential stage, suggestive of lineage priming (Jameson et al., 2012b), and initiation of the male pathway is characterized by extensive upregulation of the testis-pathway genes with simultaneous downregulation of ovarian-pathway genes. In contrast, transcriptional priming of pregranulosa progenitors towards the female pathway results in fewer transcriptional changes upon differentiation (Munger et al., 2009; Jameson et al., 2012b). One possible explanation for the retention of open chromatin near repressed pregranulosa cell genes in Sertoli cells is that loss of open chromatin sites occurs slowly, downstream of faster transcriptional changes. At a later developmental stage, fewer open chromatin sites might be associated with repressed gene loci. During ES cell differentiation, reduced H3K4me1 and open chromatin domains were observed near repressed genes within 48 h of differentiation (Whyte et al., 2012; Buecker et al., 2014). However, in multiple hematopoietic lineage descendants, the loss of open chromatin near repressed genes did not occur until late differentiation stages (Lara-Astiaso et al., 2014). We observed a similar enrichment of DHSs around both Sertoli and pregranulosa cell genes at E13.5 and E15.5 (Fig. S2), up to 4 days after peak Sry expression. Although it remains possible that a loss of open chromatin near repressed pregranulosa cell genes occurs later in adult Sertoli cells, an alternative possibility is that open chromatin regions persist near granulosa-expressed genes, perhaps reflecting the original bipotential progenitor capable of differentiating as a Sertoli or granulosa cell. Open chromatin near granulosa genes could be bound either by TFs that function to preserve a poised chromatin state or bound by repressive TFs that maintain the Sertoli cell fate. It is unclear whether this is a generalizable feature of cell populations that derive from a common bipotential progenitor.

An intriguing characteristic of gonadal supporting cells is that long after sex determination, some mutations cause transdifferentiation to the alternate cell fate. In the testis, Dmr1 is required to prevent postnatal Sertoli to granulosa cell transdifferentiation by repression of several pregranulosa cell genes, including Foxl2, Esr2 and β-catenin (Matson et al., 2011; Minkina et al., 2014). In a reciprocal manner, mutation of Foxl2 causes granulosa-to-Sertoli cell transdifferentiation in the adult ovary (Schmidt et al., 2004; Ottolenghi et al., 2005). Our finding that Sertoli cells maintain open chromatin near many repressed granulosa genes might suggest that continued repression by TFs represents a mechanism for long-term maintenance of the Sertoli cell fate, and that these repressive mechanisms are established early during Sertoli cell development. Consistent with this, the TESCO enhancer of Fox9 is bound by FOXL2 in adult granulosa cells (Uhlenhaut et al., 2009), where Fox9 is silenced; however, it is bound by SRY/Sox9 and Sf1 in Sertoli cells (Sekido and Lovell-Badge, 2008), where Sox9 is active. The use of the same cis-regulatory element to repress expression (via FOXL2 binding) in granulosa cells and activate expression (via SRY/Sox9 and Sf1 binding) in Sertoli cells suggests that at least some cis-regulatory sites serve a dual purpose during sex-specific differentiation.

If the same regulatory sites act as enhancers in one cell type and repressors in the opposing cell type, then motifs for both male-promoting and female-promoting TFs would be located.
within the active enhancers of Sertoli cell genes, available for competitive occupancy by male- or female-promoting TFs. Consistent with this, our motif analysis discriminating active versus inactive enhancers identified both the SOX and FOX motifs as predictive of active enhancers in Sertoli cells (Fig. 4 and Table S3). Furthermore, the lack of high classification accuracy from simple sequence features in active enhancers and inactive DHSs points to a complex and combinatorial regulatory code. Alternatively, differences in protein concentrations or post-translational modifications to regulators might explain the different transcriptional programs. The dual use of enhancers, such as TESCO, may underlie the ability of gonadal supporting cells to switch fates; however, it is currently unclear how extensive this mechanism is. Future studies mapping open chromatin sites in granulosa cells will reveal whether shared cis-regulatory sites represent a subset of gene regulatory mechanisms, or whether open chromatin sites are extensively shared between these functionally and morphologically distinct cell lineages.

**Regulatory sites may harbor non-coding mutations that cause DSDs**

The inability to provide a genetic diagnosis for the majority of individuals with a DSD has led to the development of more robust diagnostic tests to screen large regions of the genome for potential disease-causing mutations, including exome sequencing and copy number variation arrays (White et al., 2011; Arboleda et al., 2013; Baxter et al., 2015). This approach has revealed non-coding mutations upstream of SOX9 that result in DSDs (Kim et al., 2015). However, identifying non-coding mutations on a genome-wide scale remains difficult, and classification of non-coding mutations as potentially disease-causing is challenging, as these mutations do not directly alter the protein sequence. Recent studies have attempted to identify human enhancers for this purpose, using publicly available data from the NIH Roadmap Epigenomics Program (Roadmap Epigenomics Consortium et al., 2015). DNasel-seq datasets for human fetal ovary and testis tissue have been used to identify putative enhancers (Ohnorsg et al., 2016). These tissues were presumably collected well after 4-6 weeks of gestation, the stage at which sex determination occurs in humans, and it is currently unclear whether enhancers that operate at the sex-determining stage are maintained in the late fetal stages. These issues can be addressed in mice; however, future studies will need to determine the extent of functional conservation between mouse and human enhancers. The genome-wide identification of regulatory sites that function during sex differentiation will enable more detailed studies on the regulation of sex-determining genes and will identify candidate regions that might harbor disease-causing mutations in humans with DSDs.

**Materials and Methods**

**Mouse strains, matings and cell/tissue isolation**

To isolate Sertoli cells, Sox9-ECFP homozygous transgenic males (Kim et al., 2007b) were bred to CD-1 (Charles River) females in timed matings to generate E13.5 and E15.5 embryos. Following sorting by FACS, the cell pellet was snap-frozen and stored at −80°C for RNA-seq. For DNasel-seq, cell pellets from the CFP-positive and CFP-negative fractions were resuspended in 250 μl of recovery-cell culture freezing media (Gibco) and slowly frozen to −80°C.

Mouse tissues (kidney, liver, heart, and brain) were collected from adult B6 mice. The B6 strain mouse fibroblast cell line was commercially available (Jackson Labs) and B6 strain ESCs were kindly provided by Ute Hochgeschwender (Duke University). For further details see the supplementary Materials and Methods.

Mice were housed in accordance with National Institutes of Health guidelines, and experimental protocols were approved by the Institutional Animal Care and Use Committee of Duke University Medical Center.

**DNasel-seq assay and data processing**

DNasel-seq assay was performed as previously described (Boyle et al., 2008a; Song and Crawford, 2010) with few modifications. DNasel-seq libraries were prepared and data processed as previously described (Boyle et al., 2008b; Li and Durbin, 2009, Song and Crawford, 2010; Song et al., 2011). For further details see the supplementary Materials and Methods.

**RNA preparation and RNA-seq data processing**

E15.5 Sox9-ECFP-positive cells were pooled into three independent biological replicates containing ~1 million cells. RNA isolation was performed using the RNeasy Micro Kit according to the manufacturer’s protocol (Qiagen). Poly-A enriched mRNA libraries were generated from 0.8-1 μg of RNA using the standard Illumina Tru-Seq V2 protocol and sequencing was carried out by the Duke Genome Sequencing and Analysis Core.

RNA-seq data was processed using Bowtie version 0.12.7 (Langmead et al., 2009) and RSEM version 1.2.0 (Li and Dewey, 2011) as described in the supplementary Materials and Methods. RSEM expected counts were quartile normalized and square-root transformed.

**TSS and regulatory domain assignments**

Transcriptional start and end sites were downloaded from the UCSC Genome browser and used to generate regulatory domains using GREAT (McLean et al., 2010). For further details see the supplementary Materials and Methods.

**Genomic location assignments**

DHSs were assigned to specific genomic locations using previously published methods (Song et al., 2011). For further details see the supplementary Materials and Methods.

**Cell type specificity categorization**

Cell type specificity was determined by comparing DHSs from DNasel-seq datasets from other cell types using the BedTools Software suite (Quinlan et al., 2010). For further details see the supplementary Materials and Methods.

**CTCF overlap**

We used previously published CTCF ChIP-seq datasets (Shen et al., 2012) and compared these against Sertoli-unique, Sertoli-specific and Sertoli-common DHSs using BedTools. For further details see the supplementary Materials and Methods.

**DMRT1-binding site enrichment analysis**

We used previously published DMRT1 ChIP-seq data (Krentz et al., 2013) and assessed overlap with Sertoli DHSs using BedTools. For further details see the supplementary Materials and Methods.

**Enrichment of DHSs in Sertoli cell-expressed genes**

To calculate enrichment, we assessed the number of DHSs that were assigned to each gene’s regulatory domain. DHSs were divided as common, Sertoli-unique, Sertoli-specific and Sertoli-common DHSs using BedTools. For further details see the supplementary Materials and Methods.

**Transgenic methods, immunocytochemistry and imaging**

A putative regulatory region upstream of Wt1 was amplified and cloned into the Hsp68-LacZ reporter vector (Addgene: Plasmid #33515). DNA was prepared for zygote injection, and resuspended in EmbryoMax Injection Buffer (Millipore, MR-095-10F). Pronuclear injections into B6SILF1/J zygotes were performed by the Duke Transgenic Core Facility to generate transgenic embryos. Embryos were dissected at E13.5 and genotyped for the LacZ gene. For immunostaining, fixed gonads were processed as previously described.
ChIP-seq assay and data processing
For ChIP-seq, FACS-purified Sertoli cells were pelleted, resuspended, crosslinked and stored at −80°C. Briefly, 1 million cells were pooled from multiple sorts, washed, lysed, sonicated and incubated with bead-antibody complexes with 2.5 μg antibody (Rabbit-anti-H3K27ac; Abcam ab4729). Library preparation was performed on immunoprecipitated chromatin using the Rubicon TruPLEX FD kit according to the manufacturer’s protocol. Sequencing was performed at Duke’s Genome Sequencing and Analysis core facility.

Sequence analysis to identify predictive 6-mers and matching motifs
Enriched motifs were identified by incorporating previous methods (Lee et al., 2011; Natarajan et al., 2012). 6-mers and their reverse complements, were counted in each region, normalized by the length of the region, and used as features for an L1-norm sparse logistic regression classifier (Koh et al., 2007). Ten randomized iterations of fourfold cross-validation were performed to generate 40 different partitions of the data. 6-mers that showed non-zero regression coefficients in over 75% of the partitions were deemed to be significant and are shown in Table S3A-D. 6-mers were matched to known TF PWMs with TOMTOM (Gupta et al., 2007) using the JASPAR core vertebrate motifs, the UNIPROBE database and those generated by known TF PWMs with TOMTOM (Gupta et al., 2007) using the JASPAR database of transcription factor-binding profiles: new content and tools in the 2008 update. Nucleic Acids Res. 36, D1213-D1218.

Analysis of significance of motif matches
We used FIMO from the MEME Suite (Grant et al., 2011) to scan different regions of the genome. Length normalized motif match scores were used in a one-sided Mann-Whitney U-test to test for significance. For further details see the supplementary Materials and Methods.

Acknowledgements
We are grateful to the following people who assisted with this project: Olivier Fredigo, Lisa Bukovnik and Fangfei Ye at Duke’s Genome Sequencing and Analysis core facility; Chris Futtner, Cheryl Bock and MeiLiang Flowers at the Transgenic Core Facility for their work deriving the transient transgenic lines; and Mike Cook and Lynn Martinek at the Duke Comprehensive Cancer Center Flow Cytometry Shared Resource facility. The labs of Vivian Bardwell and David Zarkower (University of Minnesota) kindly shared DMR1 ChIP-seq data prior to publication. We thank Steve Munger (Jackson Laboratories) for his guidance regarding RNA-seq and Ute Hochgeschwender (Duke University) for providing ESC material. Iordan Batchvarov (Maatouk et al., 2012). Antibodies were used at the following dilutions: rat-anti-CDH1, 1:250 (Zymed, 13-1900); rabbit-anti-β-galactosidase, 1:10,000 (MP Biomedicals, 55796); goat-anti-MIS/AMH, 1:250 (Santa Cruz, sc-6886); Alexa Fluor 488-anti-rat, 1:500 (Molecular Probes, A21208); Cy3-anti-rabbit, 1:500 (Jackson ImmunoResearch Laboratories, 711-165-15); Alexa Fluor 647-anti-goat, 1:500 (Molecular Probes, A21447). For further details see the supplementary Materials and Methods.

References


