

# Elucidation of the transcription network governing mammalian sex determination by exploiting strain-specific susceptibility to sex reversal

Steven C. Munger,<sup>1</sup> David L. Aylor,<sup>2</sup> Haider Ali Syed,<sup>2</sup> Paul M. Magwene,<sup>3</sup> David W. Threadgill,<sup>2,4</sup> and Blanche Capel<sup>1,5</sup>

<sup>1</sup>Department of Cell Biology, Duke University Medical Center, Durham, North Carolina 27710, USA; <sup>2</sup>Department of Genetics, University of North Carolina, Chapel Hill, North Carolina 27599, USA; <sup>3</sup>Department of Biology, Duke University, Durham, North Carolina 27708, USA; <sup>4</sup>Department of Genetics, North Carolina State University, Raleigh, North Carolina 27695, USA

Despite the identification of some key genes that regulate sex determination, most cases of disorders of sexual development remain unexplained. Evidence suggests that the sexual fate decision in the developing gonad depends on a complex network of interacting factors that converge on a critical threshold. To elucidate the transcriptional network underlying sex determination, we took the first expression quantitative trait loci (eQTL) approach in a developing organ. We identified reproducible differences in the transcriptome of the embryonic day 11.5 (E11.5) XY gonad between C57BL/6J (B6) and 129S1/SvImJ (129S1), indicating that the reported sensitivity of B6 to sex reversal is consistent with a higher expression of a female-like transcriptome in B6. Gene expression is highly variable in F2 XY gonads from B6 and 129S1 intercrosses, yet strong correlations emerged. We estimated the F2 coexpression network and predicted roles for genes of unknown function based on their connectivity and position within the network. A genetic analysis of the F2 population detected autosomal regions that control the expression of many sex-related genes, including *Sry* (sex-determining region of the Y chromosome) and *Sox9* (*Sry*-box containing gene 9), the key regulators of male sex determination. Our results reveal the complex transcription architecture underlying sex determination, and provide a mechanism by which individuals may be sensitized for sex reversal.

[*Keywords:* Sex determination; genetical genomics; systems genetics; eQTL; transcription networks; organogenesis]

Supplemental material is available at <http://www.genesdev.org>.

Received June 24, 2009; revised version accepted September 4, 2009.

Many lines of evidence indicate that developmental pathways are not linear, but rather consist of many parallel inputs that are integrated in a combinatorial fashion to drive differentiation. Within a developing organ, the transcription architecture underlying individual cell fate decisions is initially plastic, and relatively small changes in gene expression can act cumulatively to bias the transcription network and drive cell fate decisions. After the cell fate decision is established, it is reinforced by intracellular and intercellular mechanisms that amplify the selected pathway and repress the alternative fate pathway, resulting in a robust canalized transcription network that is resistant to further perturbations.

The mammalian gonad is unusually plastic at the time of sex determination, which makes it unique among

organ systems and ideal for elucidating the transcription architecture responsible for mediating and reinforcing fate decisions. In mice, the paired gonads arise on embryonic day 10 (E10) as bipotential primordia that are initially competent to differentiate as testes or ovaries irrespective of chromosomal sex (Brennan and Capel 2004). This brief window of plasticity, lasting until approximately E11.5, stems from a balanced network state established by the antagonistic intercellular signals WNT4 (wingless-related MMTV integration site 4) and FGF9 (fibroblast growth factor 9) (Kim et al. 2006). Sexual fate is determined by biasing the balanced network toward the male or female pathway. This decision is subsequently maintained by reinforcing the active pathway and repressing the alternative pathway. In XY gonads, transient expression of the transcription factor *Sry* (sex-determining region of the Y chromosome) in gonadal somatic cells in a precise developmental window (E10.5–E12.5) biases the balanced network toward the

<sup>5</sup>Corresponding author.

E-MAIL [b.capel@cellbio.duke.edu](mailto:b.capel@cellbio.duke.edu); FAX (919) 681-7978.

Article is online at <http://www.genesdev.org/cgi/doi/10.1101/gad.1835809>.

male pathway, resulting proximally in Sertoli cell specification and ultimately in testis differentiation (Gubbay et al. 1990; Koopman et al. 1990; Sinclair et al. 1990; Hiramatsu et al. 2008). SRY directly regulates *Sox9* (Sry-box containing gene 9) (Sekido and Lovell-Badge 2008), which has been shown by mutation in humans and transgenic experiments in mouse to be necessary and sufficient for testis development (Foster et al. 1994; Wagner et al. 1994; Vidal et al. 2001; Chaboissier et al. 2004; Barrionuevo et al. 2006). Mounting evidence implicates the up-regulation of SOX9 expression as the critical factor driving male sex determination (Hiramatsu et al. 2008; Wilhelm et al. 2009). Male-specific *Sox9* expression is reinforced downstream from *Sry* by multiple mechanisms, including a feedforward loop with FGF9 (Kim et al. 2006). In the absence of *Sry*, as in XX gonads, *Sox9* expression is not established, and up-regulation of *Wnt4* biases the balanced network toward the female pathway to drive ovarian differentiation. Thus, differentiation as a testis or an ovary proceeds as a consequence of sexually dimorphic gene expression and downstream feedback loops that disrupt the WNT4/FGF9 balance to amplify or repress *Sox9* expression.

It is becoming clear that sex is not a simple Mendelian trait but rather a complex threshold dichotomy (Palmer and Burgoyne 1991a; Capel 2006; Mittwoch 2006). Under normal conditions, SRY expression is sufficient to establish *Sox9* expression above the critical threshold required to drive Sertoli cell differentiation and downstream male-specific morphogenetic events, as expected of a dominant determinant. However, analyses of sex reversal in human cases and mouse mutants have identified a number of X-linked and autosomal genes that play a role in mediating or canalizing the sexual fate decision (Fleming and Vilain 2005; Wilhelm et al. 2007). Similarly, genetic background differences have been identified in inbred mouse strains that can compromise male sex determination when combined with certain Y-chromosome variants or other genetic perturbations (Eicher et al. 1982, 1996). These modifier loci likely affect *Sox9* expression either in a direct manner independent of *Sry* (Bouma et al. 2005), in an indirect manner by affecting *Sry* expression (Bullejos and Koopman 2005), or by up-regulating components of the alternative pathway to bias the balanced network in favor of the female fate (Eicher et al. 1995). Moreover, recent microarray experiments on gonadal somatic cells have identified hundreds of genes that exhibit sexually dimorphic expression patterns at E11.5 (Nef et al. 2005; Small et al. 2005; Beverdam and Koopman 2006; Bouma et al. 2007; Cory et al. 2007), revealing a large network that is active in establishing and/or canalizing the sexual fate decision. Although *Sry* lies at the top of the hierarchy in therian mammals, it is evident that the combined actions of many transcription factors and signaling pathways are required to amplify the male pathway and repress the competing female pathway to establish the testis fate. The question of how these many pathways are interwoven lies at the center of sex determination, yet few methods have been applied to detail and interpret this complex interplay at a global level.

Expression quantitative trait loci (eQTL) analysis, or "Genetical Genomics" (Jansen and Nap 2001), has been applied previously to adult tissues with static transcriptomes, and holds promise to reveal genetic interactions at a global level in developing organs. This approach treats the natural variation in the expression levels of genes in a population of individuals as quantitative traits to genetically map the positions of loci throughout the genome that control expression levels. Genes that share an eQTL can be related to infer a gene interaction network (Wessel et al. 2007; Zhu et al. 2008). As with any genetic mapping approach, choosing strains that differ in the phenotype of interest (gene expression in the case of an eQTL analysis) is a critical first step in the process. Inbred strain differences in gonad phenotypes have been well-characterized and predict significant underlying expression differences that can be mapped by an eQTL approach. For example, Eicher et al. (1982) first identified the C57BL/6J (B6) background as being particularly sensitive to XY sex reversal in response to the introduction of a *Mus poschiavinus* Y chromosome (Y<sup>POS</sup>). Other strains, including DBA/2J and 129S1/SvImJ (129S1), were found to be completely resistant to Y<sup>POS</sup>-associated sex reversal. Similarly, the B6 genome was later found to be uniquely sensitive to XY sex reversal in response to other genetic perturbations, including *Fgf9* (Colvin et al. 2001), *Dax1* (*Nr0b1*) (Meeks et al. 2003), *SF1* (*Nr5a1*), and *Wt1* (Brennan and Capel 2004). Classical QTL mapping experiments identified multiple loci that conferred some but not all of the observed sensitivity to XY sex reversal in B6 (Eicher et al. 1996; Whitney et al. 2000; Bouma et al. 2005; Nikolova et al. 2008). Importantly, these findings implicate a disrupted global network rather than a few discrete genes as conferring this phenotype. However, sex reversal is never observed in wild-type B6 males, implicating the existence of a compensatory mechanism that normally counteracts this variant network to successfully initiate and canalize the male pathway. As gene expression is a more proximate measure of a QTL's effect than phenotype, an eQTL approach will likely identify novel loci that confer or counteract the sensitive background in B6.

We present the first integrated coexpression and eQTL analysis of a developing mammalian organ, the mouse gonad, during sex determination. We identified reproducible differences in the transcriptome of the developing XY gonad between B6 and 129S1 at the critical time point of E11.5, which suggested that the mechanism for B6 sensitivity to XY sex reversal is overexpression of genes typically associated with female gonads. We extended this analysis to quantify the expression of many sex determination pathway genes in a large panel of F2 XY gonads derived from B6 and 129S1 intercrosses. We found that gene expression is highly variable across the F2 population, yet many genes are coexpressed. We used a first-order conditional independence (FOCI) model to identify the most robust correlations in the F2 expression data and predict important roles for genes of unknown function based on their high degree of connectivity and positions within the network. A genetic analysis of the F2

population detected multiple autosomal regions that control the expression of many sex-related genes. Importantly, we identified autosomal regions that control the expression of *Sry* and *Sox9*, the key regulators of male sex determination. Our results provide insight into the complex transcription architecture underlying cell fate determination, organogenesis, and sex determination, and suggest a mechanism by which individuals may be sensitized to sex reversal.

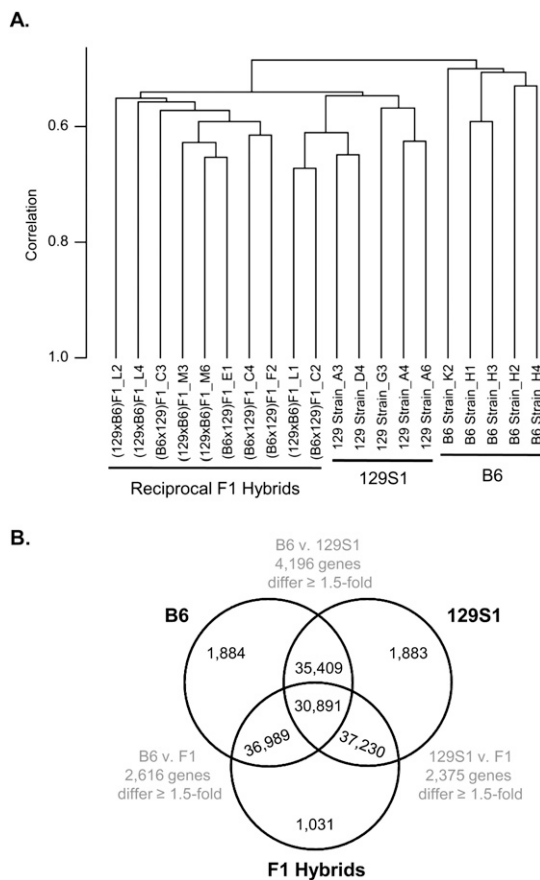
## Results

### *Inbred strain differences in the XY gonad transcriptome at E11.5 are reproducible and significant*

To investigate the genetic differences that underlie observed differences in gonad phenotypes among inbred strains, we analyzed whole-genome transcript abundance using microarrays in individual pairs of XY gonads at the critical time point of sex determination (E11.5). Significant and reproducible differences in gene expression were found between strains, whereas variation among individual XY gonads from the same parental strain was much lower, and very little, if any, litter bias could be detected (Fig. 1A) (mean pairwise Pearson correlation within a strain = 0.84; mean pairwise correlation among parental strains and reciprocal F1 hybrids = 0.76; mean pairwise correlation between parental B6 and 129S1 strains alone =

0.60). Gene expression differences between the parental strains were significant, as all B6 and 129S1 samples clustered in a strain-dependent pattern. In contrast, reciprocal F1 hybrid B6129S1F1 and 129S1B6F1 samples failed to cluster as distinct groups, suggesting that expression differences due to parental origin effects were negligible. Furthermore, expression profiles of F1 hybrid samples were more similar to the expression profiles of the 129S1 parent than the B6 parent irrespective of cross-direction (Fig. 1A). This observation is consistent with earlier studies showing that the 129S1B6F1 hybrid background is resistant to  $Y^{POS}$ -associated sex reversal, like the 129S1 parental strain (Whitney et al. 2000).

We then averaged the transcript abundance profiles of individual gonad pairs within each strain to quantify the total number of genes that were differentially expressed. Because individual expression profiles of the reciprocal F1 hybrids did not cluster into distinct groups, we treated all F1 samples as one group ( $n = 10$  samples) for subsequent analyses. Of the 39,605 genes that passed the filtering criteria, 30,891 showed similar expression levels in all three groups. The differences in gene expression between B6 and 129S1 in E11.5 XY gonads were significant. More



**Figure 1.** Microarray profiles of individual E11.5 XY gonad pairs exhibit high within-strain reproducibility. (A) Clustering dendrogram of individual E11.5 XY samples. Total transcript abundance was profiled for individual pairs of E11.5 XY gonads ( $n = 5$ ) from each of the following strains: the parental strains C57BL/6J (B6) and 129S1/SvImJ (129S1), and the (B6x129S1)F1 and (129S1xB6)F1 progeny from reciprocal crosses. Individual samples were clustered based on the set of genes differentially expressed at twofold or higher in at least 20% of the samples. Individual gonads from the same parental strain were highly correlated, and little or no litter bias was detected (letter following strain designation denotes litter; e.g., samples A3 and A4 are from the same litter). Individual (B6x129S1)F1 and (129S1xB6)F1 samples failed to cluster as distinct groups, suggesting that expression differences due to parental origin effects are negligible. Expression profiles of F1 hybrid gonads more closely resemble those from the 129S1 parental strain, in agreement with phenotypic studies showing that B6129S1F1 hybrid mice are resistant to  $Y^{POS}$ -associated sex reversal. (B) Venn diagram depicting the number of genes differentially expressed >1.5-fold between strains. Total transcript abundance profiles of individual gonad pairs within a strain were averaged ( $n = 5$  XY samples for each strain) to quantify the total number of genes differentially expressed between strain groups. Because they failed to cluster into distinct groups, all reciprocal hybrid F1 XY samples ( $n = 10$ ) were averaged together for purposes of this analysis. A total of 4196 genes are differentially expressed between B6 and 129S1 XY gonads at the 1.5-fold threshold level or higher (35,409 genes are similar). The combined F1 XY group differs from the B6 parental strain for 2616 genes at the 1.5-fold threshold (36,989 genes are similar), and from the 129S1 parental strain for 2375 genes at the 1.5-fold threshold (37,230 genes are similar). A set of 1884 genes are differentially expressed in B6 relative to both the 129S1 strain and F1 hybrid group, while 1883 genes are differentially expressed in 129S1 relative to both the B6 strain and F1 hybrids, and 1031 genes are differentially expressed in the F1 hybrid group relative to both parental strains.

than 10% of all probes tested, representing 4196 genes, varied between the parental strains at a threshold of 1.5-fold or higher (Fig. 1B; gene annotations and fold change values are listed in Supplemental Table 1). At the same threshold, a total of 2616 genes differed between the combined F1 group and the B6 parental strain, and 2375 genes differed between the combined group and the 129S1 parental strain. At the more conservative twofold threshold, the expression levels of 1213 genes differ between B6 and 129S1, 883 genes differ between B6 and the F1 group, and 739 genes differ between 129S1 and the F1 group.

*Expression bias of sexually dimorphic genes between strains predicts differences in sensitivity to XY sex reversal*

Somatic support cells are known to direct the sexual fate decision in the gonad (Palmer and Burgoyne 1991a). Consequently, we predicted that many of the genes differentially expressed between B6 and 129S1 XY gonads that confer the sex reversal phenotype would likely be sexually dimorphic at the critical time point (E11.5) and expressed specifically in the somatic support cell lineage. To focus on genes most likely to play a role in sex determination, we cross-referenced our XY strain-dependent data set to previously published microarray studies that quantified XX and XY differences in *SF1*-EGFP-positive gonadal somatic cells at E11.5 (Nef et al. 2005; Beverdam and Koopman 2006) and *Sry*-EGFP-positive Sertoli cell progenitors at E13.0 (Bouma et al. 2007). Based on the significance criteria established in each previous study, we derived a list of 656 genes expressed higher in XX somatic cells relative to their XY counterparts (hereafter referred to as “female genes”), as well as a list of 770 genes expressed higher in XY somatic cells relative to their XX counterparts (hereafter referred to as “male genes”).

A striking pattern was observed for female genes differentially expressed in our XY strain-dependent data

set: 106 female genes were differentially expressed in B6 and 129S1 E11.5 XY gonads at 1.5-fold or higher, and the vast majority of these genes were up-regulated in B6 relative to 129S1 (83 up in B6 vs. 23 up in 129S1;  $\chi^2$  test of goodness-of-fit  $P = 1.99E-09$ ) (Table 1; Supplemental Table 2). This list includes *Fst* (follistatin; 3.14-fold up in B6) and *Wnt4* (1.53-fold up in B6 XY gonads), which are known early markers of the female pathway (Vainio et al. 1999; Yao et al. 2004). In addition, 81 female genes were differentially expressed between B6 and F1 hybrid gonads, with a similar majority being up-regulated in B6 (65 up in B6, 16 up in F1 hybrid group;  $P = 3.58E-09$ ). This suggests that B6 alleles responsible for the up-regulation of the female pathway are recessive, consistent with previous results (Eicher and Washburn 1986). In contrast to B6, no significant bias in female gene expression is observed between 129S1 and the F1 hybrid group (13 up in 129S1, 18 up in F1 hybrid group;  $P = 0.31$ ). Importantly, the female shift in the B6 XY gonad transcriptome suggests a molecular mechanism responsible for B6 sensitivity to XY sex reversal.

A similar number of male genes were differentially expressed in the XY strain data set: A total of 89 male genes were differentially expressed at 1.5-fold or higher in B6 and 129S1 E11.5 XY gonads (36 up in B6, 53 up in 129S1;  $P = 0.10$ ) (Table 1; Supplemental Table 3). Although *Sry* was not differentially expressed between B6 and 129S1, expression of *Sox9* was up-regulated 1.94-fold in the seemingly female-leaning B6 background. Consequently, some of the male genes up-regulated in B6 relative to 129S1 at E11.5 may be a direct result of the up-regulation in *Sox9* expression and may offset the female shift in the B6 XY transcriptome. In contrast to the lack of bias observed between B6 and 129S1, many male genes were significantly up-regulated in the F1 hybrid group relative to both parental strains (67 up in F1 hybrids relative to B6,  $P = 3.62E-10$ ; 50 up in F1 hybrids relative to 129S1,  $P = 1.13E-07$ ) (Table 1). The robust

**Table 1.** Many sexually dimorphic genes are differentially expressed at >1.5-fold between strains in E11.5 XY gonads

Strain A	Strain B	Gene set	Up in strain A	Up in strain B	P-value
C57BL/6J	129S1/SvImJ	Female genes	83	23	1.99E-09
C57BL/6J	F1 hybrids		66	15	3.58E-09
129S1/SvImJ	F1 hybrids		13	18	0.32
C57BL/6J	129S1/SvImJ	Male genes	36	53	0.10
C57BL/6J	F1 hybrids		10	67	3.62E-10
129S1/SvImJ	F1 hybrids		10	50	1.13E-07

The pairwise lists of total genes differentially expressed >1.5-fold among B6, 129S1, and the reciprocal F1 hybrids were cross-referenced for genes previously identified as being enriched in XX gonad somatic cells relative to their XY counterparts (referred to here as “female genes”) or in XY gonad somatic cells relative to their XX counterparts at E11.5 (referred to here as “male genes”). The B6 strain exhibits a strong bias for the overexpression of female genes relative to 129S1 and the reciprocal hybrid group. A total of 106 female genes (out of a total of 656 female genes identified from previous analyses) are differentially expressed between B6 and 129S1 XY gonads, and 83 out of 106 are up-regulated in B6 at the 1.5-fold threshold ( $P = 1.99E-9$ ,  $\chi^2$  test of goodness-of-fit). In the comparison of B6 and F1 hybrid XY gonads, 66 out of 81 female genes are up-regulated in B6 relative to the F1 hybrids ( $P = 3.58E-9$ ). In contrast, female genes are not significantly biased in 129S1 relative to the F1 hybrids (13 out of 31 up-regulated in 129S1;  $P = 0.32$ ). A total of 89 male genes (out of a total of 770 male genes identified from previous analyses) are differentially expressed between B6 and 129S1 XY gonads, but do not appear to be expressed in a significantly biased manner (53 out of 89 are up-regulated in 129S1 at the 1.5-fold threshold;  $P = 0.10$ ,  $\chi^2$  test of goodness-of-fit). In contrast, male genes are significantly overexpressed in the F1 hybrid XY group relative to either parent strain (67 out of 77 up-regulated in F1 group relative to B6,  $P = 3.62E-10$ ; 50 out of 60 up-regulated in F1 relative to B6,  $P = 1.13E-7$ ).

expression of the male pathway in the F1 hybrid group indicates that multiple transgressive alleles of male pathway genes have segregated in the parental strains.

#### High gene expression variation in F2 E11.5 gonads derived from B6 and 129S1 intercrosses

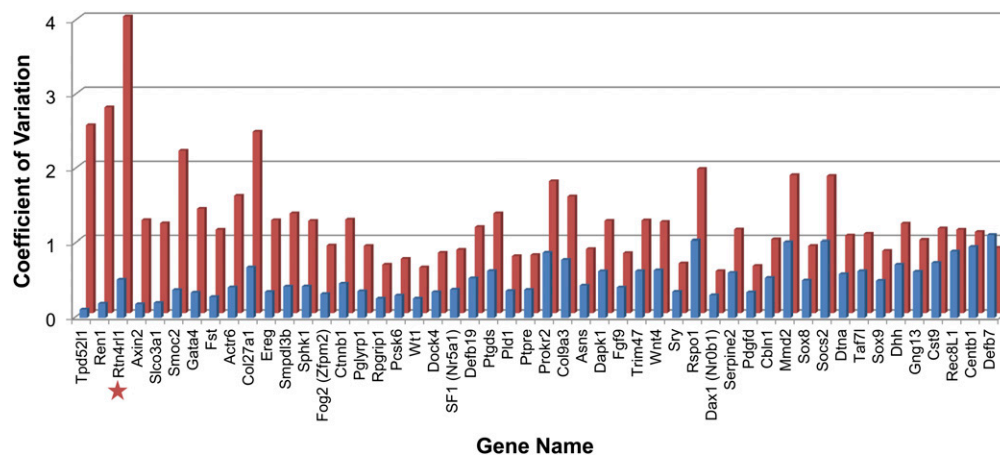
The large number of genes that were differentially expressed between B6 and 129S1 suggest that the transcriptional network controlling gonad organogenesis and male sex determination is highly plastic. To investigate the degree of plasticity in gene expression that can be tolerated in the gonad network, we established a cross between B6 and 129S1 mice, and then intercrossed the resulting F1 progeny to generate a segregating F2 panel of E11.5 embryos. We collected F2 and B6 control XY gonads for this analysis over a narrow window of development, and determined the exact developmental stage of each F2 embryo by counting tail somites (ts) posterior to the hindlimb bud (Hacker et al. 1995). Only embryos that fell in the 17–20 ts developmental window (~9-h window surrounding E11.5) were included in the expression analysis. We chose 54 genes expected to play a major role in strain-related sex reversal for analysis. This list included 44 sexually dimorphic genes that were significantly differentially expressed between B6 and 129S1 (26 male and 18 female genes), and 10 genes that did not differ between the parent strains but were known to be important players in sex determination. In addition, we included four housekeeping genes for normalization controls (gene names and primer sequences are listed in Supplemental Table 4). Transcript abundance was quantified by real-time qPCR.

The variation in gene expression (coefficient of variation, CV = standard deviation/mean) was higher in the F2

XY panel relative to B6 controls for all but one of the genes analyzed that passed a quality filter (excluding three genes in which >50% of the samples were missing data) (Fig. 2). On average, the expression variability observed in F2 samples was 2.5-fold higher than in B6 controls (mean  $CV_{F2} = 1.3$ , mean  $CV_{B6} = 0.52$ ; relative CV =  $CV_{F2}/CV_{B6} = 2.5$ ), with the highest relative CV observed for *Tpd5211* (22.3-fold higher in the F2 panel), and the lowest relative CV observed for *Defb7* (1.26-fold lower in the F2 panel). Any technical variability introduced in the dissection process or expression analysis should be consistent between the F2 panel and B6 control groups. The extreme variability in levels of gene expression observed for this subset of genes likely to play a major role in sex determination suggests that there is a high degree of plasticity in the gonad transcription network underlying sex determination, despite the fact that all F2 XY gonads are expected to develop as testes.

#### Many genes are coexpressed across F2 XY gonads

Despite the high levels of variation in gene expression observed across the F2 XY population, a preliminary analysis revealed that many genes were coexpressed across F2 gonads (Pearson correlation data not shown). To characterize the extent of coexpression relationships in more detail, we used a FOCI model (Magwene and Kim 2004) to estimate a gene coexpression network based on the F2 XY panel expression data. FOCI modeling is a type of graphical modeling approach that is useful for exploring the correlation structure of multivariate data sets, and is particularly appropriate for genomic studies where there are many genes to consider but only a modest number of samples are available (Magwene and Kim 2004; Wille



**Figure 2.** Gene expression variability is high in F2 XY gonads relative to B6 controls. The expression of 50 genes was quantified by nanoliter-scale qPCR (BioTrove OpenArray) in a panel of F2 XY gonads ( $n = 82$ ) and a set of B6 XY controls ( $n = 5$ ). The CV was then calculated and compared in the F2 XY panel relative to B6 controls. For each gene tested, CV is calculated as the standard deviation divided by the mean. CV is plotted on the Y-axis, and the individual genes tested are denoted on the X-axis. Red bars represent the CV in the F2 XY panel, and blue bars represent the CV in B6 XY controls. The F2 CV bar for *Rtn4r11* (starred) was cropped to better observe the CV values of other genes (*Rtn4r11* F2 CV = 6.43). Genes are ordered based on the magnitude of the variation observed in the F2 population relative to the B6 controls (relative CV =  $CV_{F2}/CV_{B6}$ ). Genes on the left of the graph exhibit the most variation in the F2 population relative to the B6 controls. *Tpd5211* exhibits the highest relative expression variation (*Tpd5211*  $CV_{F2}/CV_{B6} = 22.3$ ), while *Defb7* exhibits the lowest relative expression variation (*Defb7*  $CV_{F2}/CV_{B6} = 0.79$ ). Gene expression variability is consistently higher in the F2 XY panel relative to the B6 XY controls (mean  $CV_{F2}/\text{mean } CV_{B6} = 2.54$ ), with the exception of *Defb7*.

et al. 2004; Wille and Buhlmann 2006). Applied to gene expression data, the FOCI model highlights pairs of genes whose association (correlation) cannot be “explained away” by their joint association with a third gene.

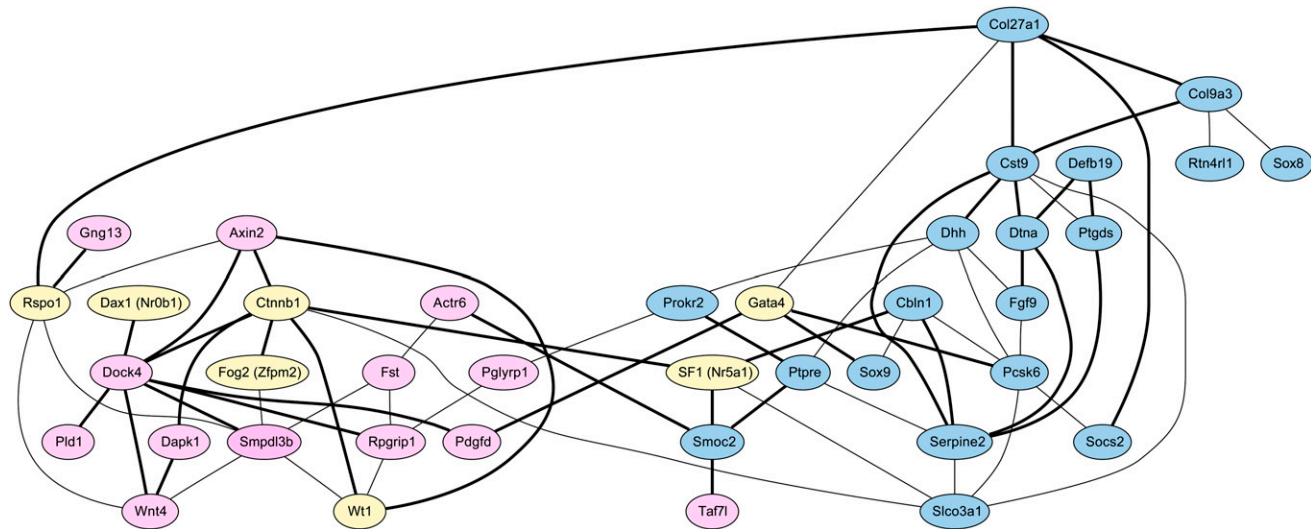
We estimated the FOCI network for 40 genes based on 68 samples from the F2 XY panel data set (we omitted samples and genes from this analysis that had substantial numbers of missing values). We used a modified false discovery rate (FDR) approach (Benjamini and Yekutieli 2001; Drton and Perlman 2007) to control the edgewise FDR for the FOCI graph (FDR = 0.05). Figure 3 shows the FOCI graph estimated from the F2 XY expression data with “male genes” and “female genes” (differentially expressed between parental strains) colored in blue and pink, respectively, and other genes (i.e., genes important for sex determination, but not differentially expressed between parental strains) colored in yellow. We found that 39 of the 40 genes included in this analysis were coexpressed with at least one other gene at a partial correlation threshold >0.25. A total of 68 coexpression relationships (edges) are shared among the 39 genes (nodes) in the FOCI network, and individual genes are coexpressed on average with 3.5 other genes (median = 4). The most highly connected gene in the FOCI network is *Dock4* (dedicator of cytokinesis 4;  $n = 8$  edges), a gene enriched in XX somatic support cells relative to XY at E11.5, but with no known role in sex determination. The robust coexpression of *Dock4* with multiple female pathway genes of known importance, including *Wnt4*, *Axin2*, *Cttnb1*, and *Dax1* (*Nr0b1*), strongly suggests an important role for this gene in the female pathway. Other sexually dimorphic genes with unknown roles in sex determination are found to be highly connected in the FOCI network, including *Smpdl3b* (sphingomyelin phos-

phodiesterase, acid-like 3B;  $n = 6$  edges), *Pcsk6* (proprotein convertase subtilisin/kexin type 6;  $n = 6$  edges), and *Serpine2* (serine peptidase inhibitor, clade E, member 2;  $n = 6$  edges). The coexpression of these genes with multiple other genes of known importance strongly predicts important roles for these genes in sex determination as well.

Two highly connected subgraphs are observed within the FOCI network estimated from the F2 XY panel data set (Fig. 3). The largest connected subgraph includes 36 edges shared among 22 genes. Interestingly, most of the genes (19 out of 22) in this subgraph are “male” genes (enriched in XY vs. XX somatic support cells at E11.5). The second largest connected subgraph in the network includes 26 edges shared among 17 nodes, 12 of which are “female” genes (enriched in XY vs. XX somatic support cells at E11.5).

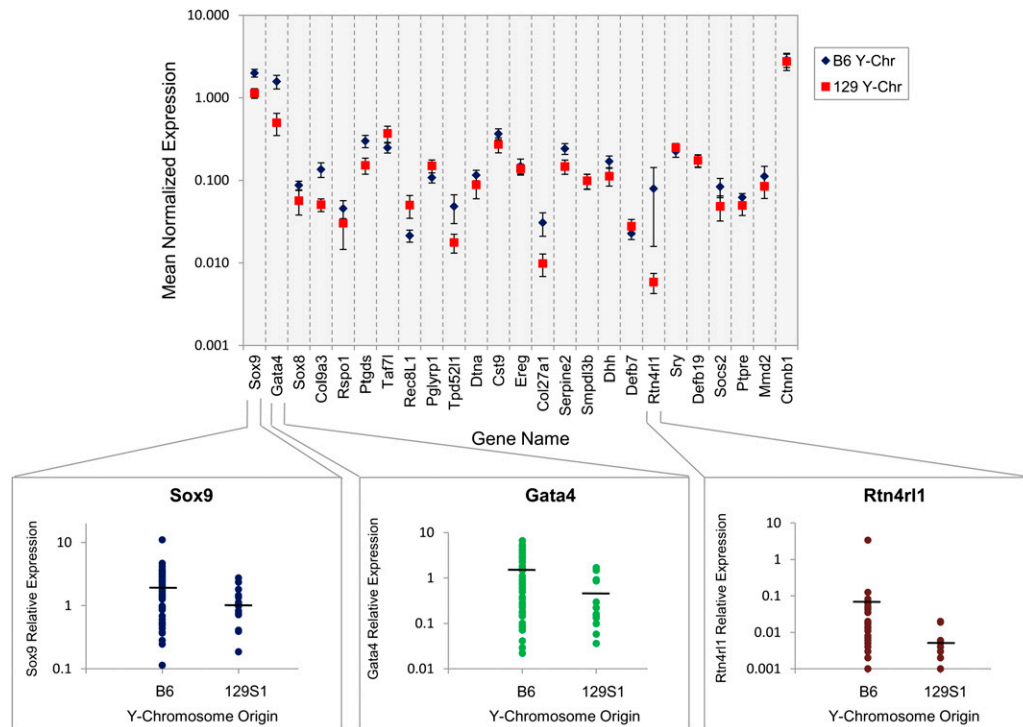
*The source of the Y chromosome is not correlated with gene expression in XY F2 gonads*

To account for any potential effects of the Y chromosome on gene expression in our F2 population, we established a breeding scheme that included both B6 and 129S1 grandfathers. We then grouped the F2 XY progeny based on the strain origin of their Y chromosome, and investigated correlations between the source of the Y chromosome and gene expression for the 51 genes quantified in the F2 panel (Fig. 4). We found that the expression levels of only two genes, *Gata4* and *Sox9*, were correlated with the strain origin of the Y (Mann Whitney U-test; *Gata4*  $P = 0.04$ , *Sox9*  $P = 0.02$ ), and no significant correlations to Y origin are found after adjusting the significance threshold for multiple comparisons by Bonferroni correction.



**Figure 3.** Gene coexpression network based on the F2 XY panel expression data. A FOCI network was estimated for a subset of 40 genes based on 68 samples from the F2 XY panel data set (samples and genes that had substantial numbers of missing values were omitted from this analysis). “Male” genes that are enriched in XY gonads at E11.5 are highlighted in blue, whereas “female”-enriched genes are highlighted in pink, and genes with known roles in sex determination that are not expressed in a sexually dimorphic pattern at E11.5 are highlighted in yellow. Thick black edges represent coexpression relationships having a partial correlation coefficient  $\geq 0.33$ , while thinner edges represent coexpression relationships having a partial correlation coefficient  $\geq 0.25$ .





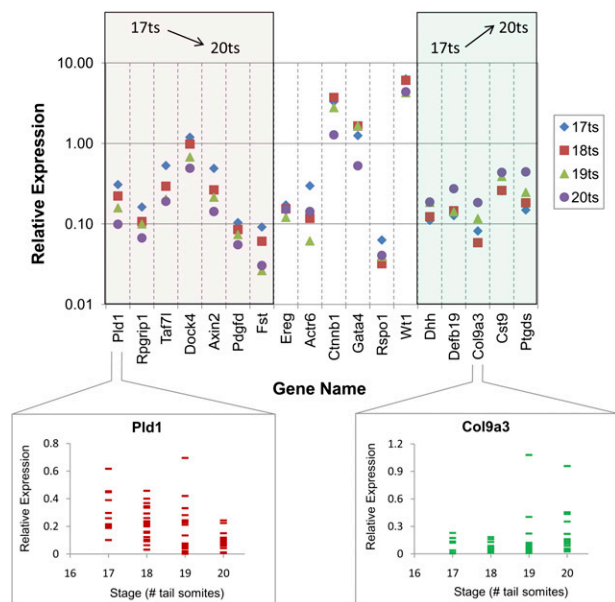
**Figure 4.** The source of the Y chromosome is correlated with very little of the overall gene expression variability observed in XY F2 gonads. The expression of 50 genes was quantified by nanoliter-scale qPCR (BioTrove OpenArray) in a panel of F2 XY gonads ( $n = 82$ ), and individual samples were then grouped based on the origin of their Y chromosome (B6 Y chromosome  $n = 61$ ; 129S1 Y chromosome  $n = 21$ ) to examine effects of the Y on gene expression. Only two genes, *Sox9* and *Gata4*, are found to be affected by the origin of the Y chromosome at a statistically significant level ( $P \leq 0.05$ ) by the Mann-Whitney U-test (*Sox9*  $P = 0.018$ ; *Gata4*  $P = 0.036$ ); however, no significant correlations to Y origin are found after adjusting the significance threshold for multiple comparisons by Bonferroni correction. (Top) Mean normalized gene expression values are plotted based on Y origin for a subset of 25 genes found to have the highest correlation between expression and Y-chromosome origin. Genes are ordered by  $P$ -value from left to right, with the most significant genes being found on the left. Blue diamonds indicate mean expression values for samples having the B6 Y chromosome, and red squares indicate samples with the 129S1 Y chromosome. Mean normalized expression is plotted on the Y-axis (note logarithmic scale), and gene names are denoted on the X-axis. Error bars indicate  $\pm$  one standard error. (Bottom) Individual F2 expression values are plotted based on Y-chromosome origin for *Sox9* (blue circles), *Gata4* (green circles), and *Rtn4r11* (maroon circles). Group mean expression values are denoted by black bars. A high level of expression variability is observed among individuals having the same Y, indicating that most of the expression differences in F2 individuals likely stem from autosomal differences. Mean expression of *Rtn4r11* in the F2 samples with the B6 Y is skewed by one outlier sample; however, expression of *Rtn4r11* is not significantly correlated with the origin of the Y chromosome ( $P = 0.268$ ).

This finding suggests that most of the expression variability observed in the F2 panel is regulated not by the Y, but rather by autosomal differences. Even for *Sox9* and *Gata4*, most of the expression variability observed among F2 individuals is independent of the Y origin (Fig. 4, bottom panels). We expected that the effect of the Y chromosome on *Sox9* expression would most likely stem from differential expression of its direct upstream regulator *Sry*, but we do not find a significant correlation between Y origin and *Sry* expression in this analysis ( $P = 0.30$ ). This is consistent with the observation that *Sry* expression is equivalent in the reciprocal F1 hybrid gonads.

*A small amount of the expression variation in a subset of genes is associated with developmental stage differences among the F2 XY panel*

To account for potential artifacts introduced by small differences in the developmental stage of individual

embryos, F2 gonads were grouped based on their exact stage of development, and stage-specific expression values were compared (Fig. 5). We found that approximately one-third (18 of 51) of the genes we examined showed a correlation between expression and developmental stage (corrected for multiple comparisons using a FDR of 0.05) (Benjamini and Hochberg 1995). Some of these genes are down-regulated from 17 to 20 ts (highlighted in red in Fig. 5), while others are up-regulated from 17 to 20 ts (highlighted in green in Fig. 5). However, there remains a large portion of the variability in these 18 genes that is independent of developmental stage. For example, *Pld1* expression trends downward from 17 to 20 ts, yet the variability observed among individuals within one group (19 ts) matches the variability observed across stage groups (Fig. 5, bottom panel). Moreover, for nearly two-thirds of the genes tested, expression does not appear to be significantly correlated with small differences in developmental stage. We conclude that most of the expression



**Figure 5.** A small amount of the gene expression variability observed in XY F2 gonads is significantly correlated with differences in developmental stage for approximately one-third of the genes tested. F2 XY gonad samples were grouped by developmental stage (17–20 ts) to analyze the effects of small differences in staging on gene expression. For a total of 18 out of 51 genes in this survey, small but significant effects of developmental stage on gene expression were detected (FDR  $q$ -value  $\leq 0.05$ ). Stage-matched mean normalized expression values for the 18 significant genes are plotted on the Y-axis (note logarithmic scale). Gene names are indicated on the X-axis. The red box highlights genes that are down-regulated from 17 to 20 ts, while the green box highlights genes that are up-regulated from 17 to 20 ts. Genes in the middle section exhibit a statistically significant, but nonlinear trend in gene expression from 17 to 20 ts. (Blue diamonds) 17-ts group ( $n = 13$ ); (red squares) 18-ts group ( $n = 24$ ); (green triangles) 19-ts group ( $n = 25$ ); (purple circles) 20-ts group ( $n = 10$ ). (Bottom) Individual F2 relative expression values are plotted grouped by developmental stage for *Pld1* (red dashes) and *Col9a3* (green dashes). *Pld1* expression is down-regulated between 17 and 21 ts, and *Col9a3* expression is up-regulated from 17 to 20 ts. Note the high level of expression variability within a stage. For example, the range for both *Pld1* and *Col9a3* expression observed within the 19-ts group encompasses the entire range observed across all stages.

variability observed in the F2 individuals results from genetic rather than stage differences.

#### Autosomal eQTLs regulate gene expression in XY gonads at E11.5

To begin to identify the genetic determinants of differential gene expression in B6 and 129S1 XY gonads at E11.5, we genotyped 128 single-nucleotide polymorphisms (SNPs) in each F2 XY embryo and searched for correlations between marker genotype and transcript abundance. SNPs were evenly spaced throughout the genome, and were genotyped using a TaqMan-based chemistry on the same platform used for the real-time quantitative PCR (qPCR) analysis (BioTrove OpenArray). Transcript abundance was

then correlated with marker genotype by maximum likelihood interval mapping in J/qtl, a Java-based graphic user interface (GUI) for R/qtl (Broman et al. 2003). Although we previously found that the Y-chromosome and stage effects were minor relative to other genetic differences, we included these variables as additive covariates in our eQTL analysis to account for possible confounding effects.

We found that the expression levels of 44 out of 51 total genes analyzed in this study were linked to at least one area of the genome (eQTL) at a suggestive level or greater ( $\alpha < 0.63$ , 1000 permutation tests for each gene). eQTLs were identified throughout the genome (Supplemental Fig. 1, all eQTL plots), and eight eQTLs are associated with the expression of multiple genes; these regions will hereafter be referred to as *trans*-band eQTLs (these regions are also referred to as putative master regulators of gene expression, or eQTL “hotspots”). *Trans*-band eQTLs were identified on chromosomes (Chrs) 1, 3, 5, 12, 14, 15, 17, and 19 (Supplemental Table 5, list of all linked transcripts and eQTLs).

The most prominent *trans*-band eQTL affecting the expression of the most genes is found at the distal end of Chr 3 (65–76 centimorgans [cM]); this region was linked to the expression of >30% of all genes tested (16 linked transcripts = *Gata4*, peak logarithm of odds (LOD) 3.4 at cM 72.5,  $\alpha = 0.34$ ; *Fog2/Zfp2*, peak LOD 4.2 at cM 72.5,  $\alpha = 0.03$ ; *Wt1*, peak LOD 2.5 at cM 76.2,  $\alpha = 0.45$ ; *SF1/Nr5a1*, peak LOD 4.1 at cM 68.5,  $\alpha = 0.05$ ; *Cttnb1*, peak LOD 5.3 at cM 76.2,  $\alpha = 0.03$ ; *Fgf9*, peak LOD 3.0 at cM 66.5,  $\alpha = 0.27$ ; *Cbln1*, peak LOD 3.7 at cM 64.4,  $\alpha = 0.14$ ; *Col9a3*, peak LOD 2.9 at cM 76.2,  $\alpha = 0.32$ ; *Asns*, peak LOD 2.5 at cM 66.5,  $\alpha = 0.46$ ; *Rspo1*, peak LOD 13.4 at cM 68.5,  $\alpha = 0.28$ ; *Dock4*, peak LOD 3.6 at cM 76.2,  $\alpha = 0.11$ ; *Pld1*, peak LOD 3.9 at cM 76.2,  $\alpha = 0.04$ ; *Dapk1*, peak LOD 4.1 at cM 74.5,  $\alpha = 0.07$ ; *Socs2*, peak LOD 3.0 at cM 70.5,  $\alpha = 0.35$ ; *Smpd3b*, peak LOD 3.4 at cM 68.5,  $\alpha = 0.34$ ; and *Rec8L1/Rec8*, peak LOD 4.8 at cM 68.5,  $\alpha = 0.25$ ). Many of these genes are known to be important during the early stages of gonadogenesis, and all genes exhibit higher expression in F2 embryos homozygous 129S1 in this region (data not shown). These genes as a whole are not expressed at different levels in the parent strains, suggesting the presence of one or more additional transgressive alleles that repress the expression of these genes in the 129S1 parental strain.

We identified a prominent *trans*-band eQTL on Chr 12 (33–49 cM) that affects the expression of nine genes, including seven genes that are known to play a role in the female pathway or are up-regulated in XX relative to XY somatic support cells (*Pdgfd*, peak LOD 2.5 at cM 40.2,  $\alpha = 0.48$ ; *Axin2*, peak LOD 3.2 at cM 35.5,  $\alpha = 0.17$ ; *Dock4*, peak LOD 3.7 at cM 43.0,  $\alpha = 0.09$ ; *Rpgrip1*, peak LOD 5.4 at cM 41.0,  $\alpha = 0.01$ ; *Cttnb1*, peak LOD 2.8 at cM 35.5,  $\alpha = 0.49$ ; *Pld1*, peak LOD 2.3 at cM 28.6,  $\alpha = 0.47$ ; and *Dax1/Nr0b1*, peak LOD 2.3 at cM 45.0,  $\alpha = 0.61$ ), and two genes important during the early stages of gonadogenesis (*Wt1*, peak LOD 3.8 at cM 47.0,  $\alpha = 0.04$ ; and *Fog2/Zfp2*, peak LOD 3.6 at cM 37.0,  $\alpha = 0.10$ ). These genes exhibit higher expression in F2 embryos homozygous 129S1 in this region (data not shown).

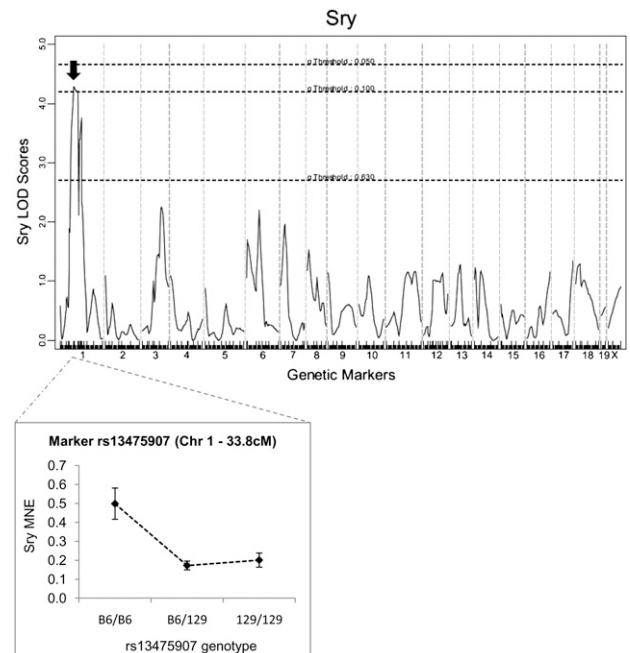


In addition, we identified two *trans*-band eQTLs that influence the expression of multiple male pathway genes. A region of Chr 5 (26–46 cM) controls the expression of six genes that have known roles in the male pathway, or are up-regulated in XY relative to XX somatic support cells at E11.5. These genes include *Dhh* (peak LOD 2.9 at cM 42.0,  $\alpha = 0.46$ ), *Serpine2* (peak LOD 2.4 at cM 42.0,  $\alpha = 0.57$ ), *Sphk1* (peak LOD 6.5 at cM 32.0,  $\alpha = 0.11$ ), *Mmd2* (peak LOD 3.7 at cM 42.0,  $\alpha = 0.27$ ), *Trim47* (peak LOD 2.8 at cM 34.0,  $\alpha = 0.35$ ), and *Tpd52l1* (peak LOD 2.5 at cM 46.0,  $\alpha = 0.49$ ). All of these genes exhibit higher expression in F2 embryos homozygous B6 in this region (data not shown). These genes are not overexpressed in the B6 parental strain relative to 129S1, suggesting the presence of one or more additional transgressive alleles that repress the expression of these genes in the B6 parental strain.

A *trans*-band eQTL on Chr 1 (33–49 cM) controls the expression of five genes, four of which have important roles in the male pathway. Two of these genes, *Sry* (peak LOD 4.3 at cM 33.8,  $\alpha = 0.09$ ) and *Sox9* (peak LOD 2.6 at cM 38.2,  $\alpha = 0.45$ ), are master regulators of male sex determination, while two others, *Fgf9* (peak LOD 2.5 at cM 30.2,  $\alpha = 0.50$ ) and *Ptdgs* (peak LOD 3.1 at cM 36.2,  $\alpha = 0.17$ ), are known to reinforce *Sox9* expression downstream from *Sry*. In addition, one gene of unknown function, *Cbln1* (peak LOD 2.6 at cM 32.3,  $\alpha = 0.50$ ), is regulated by this eQTL and exhibits early male-specific expression patterns. Four of the five genes (*Sox9*, *Fgf9*, *Ptdgs*, and *Cbln1*) exhibit higher expression in F2 XY gonads that are homozygous 129S1 in this region, yet *Sry* appears to be expressed at lower levels when this region is homozygous 129S1 (Fig. 6). As SOX9 may down-regulate *Sry* in a negative feedback loop, the same 129S1 region on Chr 1 correlated with an increase in *Sox9* expression could also be correlated with a consequent decrease in *Sry* expression.

#### *Sox9* expression is controlled by multiple autosomal regions in a complex manner

*Sox9* expression is regulated in a complex fashion by at least four autosomal *trans*-acting eQTLs (Fig. 7). The strain-specific effects of the three most significant eQTLs on *Sox9* expression are detailed in the bottom panels of Figure 7. *Sox9* expression is down-regulated in F2 XY embryos homozygous B6 in the Chr 1 interval, as described above (Fig. 7, bottom left panel). In addition, the eQTLs on Chr 5 (peak LOD 4.1 at cM 10.1,  $\alpha = 0.04$ ) and Chr 11 (peak LOD 2.3 at cM 14.1,  $\alpha = 0.63$ ) up-regulate *Sox9* expression only when they are heterozygous (effects of Chr 5 eQTL displayed in the middle panel in Fig. 7; Chr 11 effects not shown). The eQTL having the most significant effect on *Sox9* expression is found on Chr 6 (peak LOD 6.4 at cM 22.8,  $\alpha = 0.00$ ), and F2 embryos that are homozygous B6 in this region exhibit higher *Sox9* expression levels relative to those that are homozygous 129S1 (Fig. 7, bottom right panel). Multiple transgressive autosomal alleles control the expression of *Sox9*; however, the Chr 6 region appears most influential in causing the observed 1.94-fold increase in *Sox9* expression in the B6

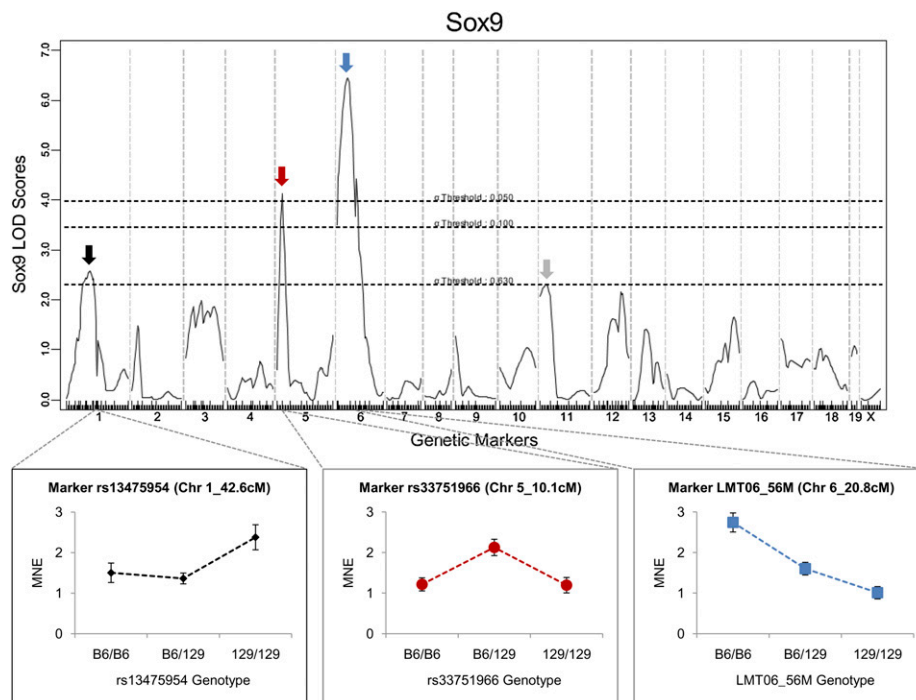


**Figure 6.** *Sry* expression is regulated by an autosomal eQTL on Chr 1. Linkage analysis was performed to find correlations between *Sry* expression in the F2 XY population and the genotype at 128 genome-wide SNP markers. *Sry* expression is significantly correlated with a region of Chr 1 (28–50 cM, closest to marker rs13475907; see arrow). Genetic markers are denoted as ticks on the X-axis relative to their position on Chrs 1–19 and the X chromosome. LOD scores at each marker are plotted on the Y-axis. The eQTL scan was performed in J/qrtl, and individual expression values were permuted 1000 times to obtain empirical significance thresholds. Three significance thresholds are plotted as horizontal dotted lines, and correspond to  $\alpha$  values of 0.05 (top line), 0.10 (middle line), and 0.63 (bottom line). (Bottom) Mean normalized expression (MNE) of *Sry* is graphed for F2 samples grouped by genotype at marker rs13475907. The three possible genotypes at rs13475907 are denoted on the X-axis, and *Sry* MNE is plotted on the Y-axis. Black diamonds denote the mean expression value for each group, and error bars represent  $\pm$ SEM. *Sry* expression is highest when marker rs13475907 is homozygous for the B6 allele, and lowest for those samples that are heterozygous or homozygous for the 129S1 allele.

parental strain, and is likely responsible for counteracting the underlying female-leaning imbalance in the B6 gonad transcription network. Although *Sry* is a mammalian invention, up-regulation of *Sox9* expression is a conserved feature of the testis pathway across vertebrates (Morrish and Sinclair 2002). The important position of *Sox9* in the testis pathway suggests that multiple independent pathways and feedback loops converge on *Sox9* expression. One or more of the *Sox9* regulatory genes residing in these four eQTL intervals could act as the genetic trigger for male sex determination in other species.

#### Discussion

The highly interwoven characteristics of signaling and transcriptional programs demand a means to study organogenesis at the level of the global network, yet few



**Figure 7.** *Sox9* expression is regulated by four eQTLs having antagonistic effects. The expression of *Sox9* was quantified by nanoliter-scale qPCR (BioTrove OpenArray) in a panel of F2 XY gonads ( $n = 79$ ), and individual samples were genotyped at 128 SNP markers located evenly throughout the genome. Linkage analysis was performed to identify markers that were correlated with *Sox9* expression. Four autosomal eQTLs, located on Chrs 1, 5, 6, and 11 (denoted by arrows), are found to regulate *Sox9* expression above a suggestive significance threshold. Genetic markers are denoted as ticks on the X-axis relative to their position on Chrs 1–19 and the X chromosome, and LOD scores are plotted on the Y-axis. The eQTL scan was performed in J/qtl, and individual expression values were permuted 1000 times to obtain empirical significance thresholds. Three significance thresholds are plotted as horizontal dotted lines, and correspond to  $\alpha$  values of 0.05 (top line; significant), 0.10 (middle line), and 0.63 (bottom line; suggestive). (Bottom) Mean normalized expression (MNE) of *Sox9* is graphed for F2 samples grouped by genotype at markers located closest to the three most significant eQTLs; rs13475954 (Chr 1–42.6cM; black diamonds), rs33751966 (Chr 5–10.1cM; red circles), and LMT06\_56M (Chr 6–27.4cM; blue squares). The three possible genotypes at each marker are denoted on the X-axis, and *Sox9* MNE is plotted on the Y-axis. Error bars represent  $\pm$ SEM. The eQTLs in these regions have antagonistic effects on *Sox9* expression. F2 individuals that are homozygous 129S1 at marker rs13475954 exhibit higher *Sox9* expression than those individuals that are homozygous B6 at this marker. The opposite trend is observed at marker LMT06\_56M, where the highest *Sox9* expression is observed for F2 individuals that are homozygous B6 at this marker. F2 individuals that are heterozygous B6/129S1 at marker rs33751966 exhibit higher *Sox9* expression than individuals that are homozygous B6/B6 or 129S1/129S1.

effective methods have surfaced. To identify genetic interactions on a global scale during organogenesis, we took an eQTL approach. Similar to classical QTL mapping, an eQTL analysis considers the expression levels of individual genes as quantitative traits and maps areas of the genome that affect the expression of individual genes in a segregating population. This approach has been successfully applied to analyze the genetics of gene expression in steady-state transcriptional networks. Here, we applied this approach for the first time to a developing organ, the mouse gonad during the critical period of sex determination.

This study provides insight into the complex nature of the transcription architecture underlying sex determination in mammals, and reveals a high degree of plasticity in the gonad transcription network at the critical time point when the bipotential gonad is establishing its sexual fate. Despite the high level of expression variability among samples, significant correlations between genes in the network emerged. We constructed a coex-

pression network based on the most robust correlations in the F2 XY panel, and predict important roles in sex determination for genes of unknown function based on their degree of connectivity and position within the network. By combining expression profiling with linkage analysis, we extended our analysis to identify autosomal regions that regulate the expression of cohorts of genes, as well as regions that affect the expression of important regulators of sex determination, including *Sry* and *Sox9*. The study implicates novel genes and autosomal loci as candidates for causative genes underlying the 75% of human disorders of sexual development (DSDs) that are unexplained.

*The gonad transcription network in C57BL/6J is shifted toward the female fate, but is counteracted by the up-regulation of Sox9 expression*

A strain comparison of the global XY gonad transcriptome at the critical time point of sex determination (E11.5)

revealed significant differences between B6 and 129S1. Cross-referencing this strain data set to previously published XX/XY microarray comparisons showed that a large number of female-enriched genes are up-regulated in B6 relative to 129S1, providing a likely explanation for the sensitivity of the B6 strain to sex reversal. These data predict that the B6 alleles conferring sensitivity to XY sex reversal act by elevating or prematurely up-regulating the female pathway, thereby raising the threshold required to establish the male pathway. These findings provide evidence in support of the “timing mismatch” hypothesis proposed by Eicher (Eicher and Washburn 1986) and Burgoyne (Burgoyne 1988; Palmer and Burgoyne 1991b). Interestingly, *Sox9* expression is up-regulated in the B6 strain, and may act as a compensatory mechanism to counteract this variant network state and ensure testis determination. As a consequence of this imbalanced network state, the transcriptional equilibrium in B6 is particularly vulnerable to the introduction of allelic variants that delay or reduce the expression level of *Sry* or *Sox9* (Albrecht et al. 2003; Bouma et al. 2005; Ballejos and Koopman 2005), or heterozygous mutations in genes that positively regulate *Sry* or *Sox9* (e.g., *SF1*, *Wt1*, and *Fgf9*) (Colvin et al. 2001; Brennan and Capel 2004). An analogous situation may exist in human cases of sex-reversed XY females that carry the same SRY mutation as their fully masculinized and fertile fathers (Dipple and McCabe 2000; Jordan and Vilain 2002). For example, adverse effects from the SRY mutation may be masked in the father by downstream positive modifiers of SOX9 expression, but may be exposed in his XY offspring by the introduction of allelic variants from the mother that compromise this compensatory mechanism or up-regulate the opposing female pathway.

As *Sox9* is found in a region of Chr 11 that is of shared ancestry between B6 and 129S1, a *trans*-acting factor is most likely responsible for the difference in *Sox9* expression observed between B6 and 129S1. The most obvious candidate for a *trans*-acting factor regulating *Sox9* expression is SRY; however, we did not see any differences in the level of *Sry* expression between B6 and 129S1; moreover, the B6 and 129S1 SRY proteins are identical (Albrecht and Eicher 1997). Furthermore, we find little or no influence of Y-chromosome origin on *Sox9* expression, strongly suggesting that the differential expression of *Sox9* in B6 and 129S1 is regulated independently of *Sry* and the Y chromosome by autosomal loci, four of which were identified in the present study.

*Gene expression is highly variable across the F2 XY population, yet large cohorts of genes are coexpressed*

We found a surprising level of variability in gene expression in the F2 population relative to a subset of B6 controls. Despite this high variability in gene expression levels, all XY gonads in this experiment would be expected to become testes. These data likely reveal a complex system of checks and balances buffered by opposing forces. Reciprocal changes and complimentary

adjustments in expression levels occur between strains and among individuals in the F2 population, and this combinatorial control of gene expression levels may buffer the system against mutation and provide a stable equilibrium resistant to perturbation. This system likely has strong evolutionary advantages, as variation can be easily accommodated, and a new equilibrium can be reached by a series of compensatory changes that establish a new balanced state.

Although expression is highly variable across the F2 population, many genes are significantly coexpressed within individual samples. We estimated this coexpression network using a FOCCI model developed by Magwene and Kim (2004). Because the FOCCI model was estimated based on a modest number of samples, this analysis is best considered exploratory. However, the model does suggest important roles for a number of genes with unknown functions in sex determination (including *Dock4*, *Smpd3b*, *Pcsk6*, and *Serpine2*) based on their high degree of connectivity within the FOCCI network and coexpression with genes of known importance. Furthermore, two highly connected subgraphs consisting of primarily male or female genes are apparent in the FOCCI network. The distinct structure of these subgraphs illustrates the ability of the unbiased FOCCI model to reveal male or female modules within the network that are likely to have biological relevance. Significantly, edges linking these two subgraphs (e.g., between *Prokr2* and *Pglyrp1* in Fig. 3) may connect genes important for the bifurcation of a bipotential network structure (i.e., common to both XX and XY gonads) toward a male or female state.

It is unknown whether the high degree of coexpression in this subset of genes is representative of the global gonad transcription network or of other developing systems. As an organ, the gonad is unusual, in that each cell within the gonad is thought to be initially bipotential, and sex determination proceeds as the sum of individual cell fate decisions. Because of this unique plasticity, the transcription architecture underlying cell fate decisions within the gonad may be similar to other multipotent stem cell populations, as these cells make a fate decision to self-renew or differentiate under the influence of multiple converging signals. It will be interesting to obtain a more dynamic view of how the gonad transcriptome approaches this decision point and diverges toward the ovary or testis pathway. Follow-up studies will extend this expression analysis to the level of the whole genome over an extended time course.

*Multiple autosomal regions regulate gene expression in the gonad*

We identified no statistically significant correlation between Y-chromosome origin and gene expression in the F2 XY panel, a finding that is consistent with the data from the reciprocal F1 hybrid groups (Fig. 1A). However, we cannot rule out any role for the Y chromosome in contributing to the phenotypic differences between these strains, as any Y-chromosome differences would likely affect *Sry* and *Sox9* expression before E11.5

and may be masked by opposing effects from the Chr 1 eQTL at E11.5. We did find a significant correlation between developmental stage and gene expression for about one-third of the genes tested, although the effect of developmental stage on variability of gene expression across this narrow developmental range was minor relative to the genetic contribution.

In concert with our F2 expression analysis, we genotyped each F2 individual and used linkage analysis to identify regions correlated with differences in gene expression. In total, we identified one or more eQTLs for nearly all of the genes tested (44 out of 51); moreover, we identified eight *trans*-band eQTLs that controlled the expression of cohorts of genes associated with early gonadogenesis or predominantly with the male or female pathways. The genes underlying these eQTLs are likely to be important nodes in the gonad transcription network that exert their effects by directly up-regulating multiple targets, or by initiating a cascade. Future work will be aimed at fine-mapping these *trans*-bands; however, multiple lines of evidence already implicate *Dock4* as a strong candidate underlying the Chr 12 eQTL: (1) *Dock4* resides immediately adjacent to the 95% confidence interval for this *trans*-band eQTL; (2) regulation of its expression is linked to this region, predicting a *cis*-acting polymorphism; and (3) many SNPs differ between B6 and 129S1 in and around *Dock4*. Furthermore, although nothing is known about its function during mammalian sex determination, *Dock4* has been shown recently in zebrafish to associate with the  $\beta$ -catenin degradation complex and to be essential for Wnt/ $\beta$ -catenin signaling (Upadhyay et al. 2008). Wnt/ $\beta$ -catenin signaling is critical for the establishment of the female pathway in mammalian sex determination (Vainio et al. 1999; Kim et al. 2006; Maatouk et al. 2008).

Importantly, we identified a *trans*-band eQTL on Chr 1 (33–49 cM) that controls the expression of five male pathway genes, including *Sry* and *Sox9*. This is the first autosomal region identified to date that regulates *Sry* expression. The reciprocal effects of this eQTL on *Sry* and *Sox9* expression levels may result from SOX9 down-regulating *Sry* expression in a negative feedback loop. Thus, increasing the relative expression of *Sox9* in animals that are homozygous 129S1 at this locus may decrease levels of *Sry*. In contrast, the other four genes regulated by the Chr 1 eQTL are positively correlated with *Sox9* expression, suggesting that this eQTL (or SOX9 itself) affects them in the same direction. Interestingly, this eQTL colocalizes to a region of Chr 1 identified recently by Eicher's group (Bouma et al. 2005) that modifies the *Dax1*-/Y sex reversal phenotype in B6. XY sex reversal in B6 *Dax1*-/Y gonads results from a failure to up-regulate *Sox9* expression, importantly, despite normal timing and levels of *Sry* expression. The mode of action of the B6 alleles at these Chr 1 loci is consistent, down-regulating *Sox9* expression in a manner independent of *Sry* expression. Fine-mapping studies will determine whether the same gene(s) that modifies the *Dax1*-/Y sex reversal phenotype in B6 underlies this *trans*-band eQTL.

#### *Many loci are likely involved in conferring sensitivity to sex reversal in B6*

Eicher et al. (1996) identified three regions on Chrs 2, 4, and 5 (designated as *tda2*, *tda1*, and *tda3*, respectively) that were likely to confer some of the sensitivity to Y<sup>POS</sup>-associated sex reversal in B6. More recently, Nikolova and colleagues (Whitney et al. 2000; Nikolova et al. 2008) mapped a region of Chr 11 from the 129S1 strain that could partially protect against Y<sup>POS</sup>-associated sex reversal when present in an otherwise B6 background. These results implicate the combined effects of many loci (rather than a single gene) as conferring the sensitivity to sex reversal in B6. If the causative genes underlying these large regions conferred their effects on sex reversal by controlling the expression of one or more of the genes analyzed in the present study, we might expect to replicate these regions in our eQTL analysis.

We identified three eQTLs that colocalize with these previously mapped regions. In addition to the Chr 1 *trans*-band eQTL described above that colocalized with a region associated with *Dax1*-/Y sex reversal, we identified a prominent *trans*-band eQTL on Chr 5 that controlled the expression of six predominantly male pathway genes and colocalized with the *tda3* locus identified by Eicher et al. (1996). In addition, we identified an eQTL on Chr 11 that controlled the expression of *Sox9* and colocalized with the 129S1 protective region identified by Nikolova et al. (2008) (see gray arrow in Fig. 7). *Sox9* is found on Chr 11 but does not reside in this region, implicating a *trans*-acting factor (rather than a polymorphism in the *cis*-regulatory region of *Sox9*) as being responsible for conferring the effects on *Sox9* expression. Finally, we identified one eQTL on Chr 4 that controlled *Smpdl3b* expression at a level approaching the suggestive threshold (peak LOD 2.3 at cM 71.6,  $\alpha = 0.70$ ) and colocalized with the *tda1* locus identified in the study by Eicher et al. (1996). As *Smpdl3b* is located in this interval on Chr 4, the polymorphism underlying this eQTL could lie in the *cis*-regulatory regions of *Smpdl3b*. *Smpdl3b* has no known role in sex determination; however, it is highly coexpressed with six other genes in the FOCI network, including *Wnt4* and *Rspo1*, both of which have key roles in the female pathway. Both *Wnt4* and *Rspo1* also reside in this region of Chr 4, and are also candidate genes underlying *tda1*. The conserved linkage of these three genes in humans and mice could be important for their role in the female pathway. Overall, the considerable overlap between this eQTL analysis and previous phenotype-based mapping studies implies that this novel approach has revealed key transcriptional relationships in the sex determination network that are likely to have important implications for sex reversal.

In conclusion, we uncovered a remarkable level of variability in gene expression during organogenesis of the gonad, both between the two parental strains and among their F2 progeny. These results suggest that it will be productive to regard organogenesis as a balanced web of signaling and transcriptional networks regulated by key nodes where pathways intersect. Through FOCI analysis, we identified key nodes in the network that

extend our understanding of sex determination and represent new candidates for causative genes underlying human DSDs. Future efforts will focus on fine-mapping the eQTLs identified in this study to elucidate the genes underlying these loci and produce a directed network of gene expression. It will be important to extend the expression analysis of F2 individuals using an unbiased microarray approach, and to extend the analysis to a broader range of developmental stages to establish a dynamic picture of expression changes. It will also be interesting to determine whether this approach can be applied productively to other developing systems.

## Materials and methods

### *Mouse strains and crosses*

C57BL/6J (stock no. 000664) and 129S1/SvImJ (stock no. 002448) mice were obtained from The Jackson Laboratory. For the microarray analyses of parental strains and F1 reciprocal hybrids, the following crosses were established: C57BL/6J  $\times$  C57BL/6J, 129S1/SvImJ  $\times$  129S1/SvImJ, C57BL/6J  $\times$  129S1/SvImJ, and 129S1/SvImJ  $\times$  C57BL/6J. For the eQTL analysis, reciprocal crosses of C57BL/6J and 129S1/SvImJ mice were established. The resulting F1 progeny were then intercrossed to produce a segregating F2 population.

### *Dissection, staging, and genotyping of E11.5 XY gonads*

Timed matings were established for the above crosses, and embryos were collected from pregnant mothers at E11.5, where E0.5 is defined as noon on the day a mating plug was detected. Embryos were staged in a more precise manner by counting ts distal to the hindlimbs: E11.5 corresponds to  $\sim$ 18 ts (Hacker et al. 1995). B6, 129S1, and reciprocal F1 hybrid embryos that fell within the 17- to 21-ts window were included in the microarray analysis, and F2 embryos that fell within the 17- to 20-ts window were included in the eQTL analysis. The chromosomal sex of each embryo was determined by PCR on tail or head genomic DNA using primers to detect the *Smcx/Smcy* genes (5'-TGAA GCTTTTGGCTTTGAG-3' and 5'-CCGCTGCCAAATTCTTT GG-3'). Only XY embryos were included in this analysis. Gonads were dissected away from mesonephroi in sterile PBS (Gibco/Invitrogen, catalog no. 1490-144) and were stored in RNAlater RNA stabilization solution (Ambion, catalog no. AM7024) at  $-80^{\circ}\text{C}$ . Sterile technique was used throughout the dissection procedure, and all instruments and surfaces were treated with RNaseZAP RNase decontamination fluid (Ambion, catalog no. AM9780), followed by 70% EtOH in DEPC-treated water.

### *RNA isolation*

For all expression analyses, total RNA was first extracted from E11.5 whole gonads (separated from mesonephroi) using TRIzol reagent (Invitrogen, catalog no. 15596-026) as directed by the manufacturer, with the following change: Glycogen (0.5–1.0  $\mu\text{L}$  of 15 mg/mL stock; Ambion GlycoBlue, catalog no. AM9515) was added to each reaction before the isopropanol precipitation step. Total RNA was eluted with 8  $\mu\text{L}$  of RNase-free water (not DEPC-treated). To prevent genomic DNA contamination, RNA was treated with 1  $\mu\text{L}$  of DNase I and 1  $\mu\text{L}$  of 10 $\times$  DNase I Reaction Buffer (Sigma, catalog no. AMP-D1) for 15 min at room temperature. DNase was inactivated by treatment with 1  $\mu\text{L}$  of DNase I Stop Solution for 10 min

at  $70^{\circ}\text{C}$ . Total RNA quantity and quality was assessed using a NanoDrop ND-1000 or ND-8000 spectrophotometer (NanoDrop Technologies, Inc.). Only samples with  $>100$  ng of total RNA and an A280:A260 ratio of  $>1.6$  were included in the expression analyses.

### *Microarray processing and analysis*

Whole-genome transcript abundance was quantified in individual pairs of E11.5 XY gonads using the Agilent  $4 \times 44,000$  mouse two-color platform (Agilent, catalog no. G4122F) according to protocols established by the manufacturer and optimized by us (Syed and Threadgill 2006), with minor modifications. Briefly, 200–450 ng of total RNA were amplified from each pair of XY gonads ( $n = 5$  pairs tested per cross) into cRNA using the Low RNA Input Linear Amplification kit (Agilent, catalog no. 5184-3523) and labeled with 1.5  $\mu\text{L}$  of Cy5-CTP (Perkin Elmer, catalog no. NEL581). Multiple tissues were pooled to create a reference sample; this pool included XY gonads, XX gonads, XY whole embryos, and XX whole embryos. The reference sample was amplified into cRNA and labeled with 1.5  $\mu\text{L}$  of Cy3-CTP (Perkin Elmer, catalog no. NEL580). To each array, 900 ng each of Cy5-labeled experimental and Cy3-labeled reference cRNA were cohybridized for 17 h at  $65^{\circ}\text{C}$ . Microarrays were scanned using an Agilent scanner, and raw data were collected using feature extraction software (Agilent). Data are available (accession no. GSE16917) at <http://www.ncbi.nlm.nih.gov/geo>.

Probe level data were imported into BRB-ArrayTools software (BRB-ArrayTools version 3.7.0, <http://linus.nci.nih.gov/BRB-ArrayTools.html>) (Simon et al. 2007), and normalized using the Lowess smoother function (Yang et al. 2002). The log ratio of background-adjusted Cy5-to-Cy3 signals was calculated for each spot and used for subsequent analyses. Probes for which  $>50\%$  of the arrays were missing data were excluded from the analysis. For hierarchical clustering of individual samples, genes whose expression differed by at least twofold in at least 20% of the samples were retained, and samples were clustered using a centered correlation with average linkage (genes not centered). To quantify expression differences among the strains and reciprocal F1 hybrids, individual samples within a group were averaged, and genes that differed at  $>1.5$ -fold in pairwise comparisons between groups were identified.

### *Reverse transcription, primer design, and real-time qPCR in F2 samples*

To quantify the expression of 56 genes in each F2 XY gonad, we used a SYBR-based chemistry on the BioTrove OpenArra platform (BioTrove). Total RNA from individual pairs of F2 whole gonads was reverse-transcribed into cDNA using the High-Capacity cDNA Archive kit (Applied Biosystems, catalog no. 4322171) followed by cleavage with Exonuclease I (Amersham Biosciences, catalog no. E70073X), according to BioTrove protocols, with the following changes: For each sample, 700 ng of total RNA were reverse-transcribed into cDNA in a total reaction volume of 11.5  $\mu\text{L}$ , resulting in a final concentration of 61 ng/ $\mu\text{L}$ . cDNA was stored at  $-20^{\circ}\text{C}$  until needed. Primers were designed for real-time qPCR analysis in PerlPrimer software (<http://perlprimer.sourceforge.net>) (Marshall 2004) according to BioTrove specifications. Primer sequences and amplicon lengths are listed in Supplemental Table 4. Primer sequences were submitted to BioTrove for synthesis and application to OpenArray slides.

Experimental and control cDNA samples were applied in replicate to OpenArray slides according to BioTrove protocol.



Slides were cycled and scanned on the OpenArray NT Cyclor (BioTrove), and the resulting amplification curves were analyzed in the accompanying OpenArray real-time qPCR system software (BioTrove) according to the manufacturer's parameters. Threshold cycle values ( $C_t$ ) were obtained for each sample, and replicate values that surpassed a confidence score of 100 were included in the analyses of F2 expression range, Y-chromosome, and stage effects, and coexpression. A more stringent confidence threshold of 300 was set for inclusion in the eQTL analysis to minimize the potential for identifying false positive associations.  $C_t$  values for all experimental genes were normalized to the housekeeping gene *B2m* to obtain relative expression values. Genes in which >50% of the F2 samples failed to meet the  $C_t$  confidence score criteria were excluded from the expression analysis.

*Fgf9* and *Dax1* expression levels in F2 and control cDNA samples were quantified independently on the iCycler Thermal Cyclor (Bio-Rad, catalog no. 170-8720) with MyiQ Single-Color Real-Time Detection System (Bio-Rad, catalog no. 170-9740). Samples were run in duplicate, and reactions included a 2- $\mu$ L sample cDNA, 10  $\mu$ L of 2 $\times$  SensiMix Plus SYBR + Fluorescein mastermix (Quanta, catalog no. QT615-05), 4  $\mu$ L of RNase-free water, and 2  $\mu$ L each of gene-specific forward and reverse primers (200 nM final concentration each). PCR cycling parameters are as follows: 15 min at 95°C (one cycle); 30 sec at 95°C, 30 sec at 59°C, 30 sec at 72°C (40 cycles); and 5 min at 72°C (one cycle).  $C_t$  for replicated samples were averaged and normalized to *B2m*.

#### FOCI network analysis

To estimate the structure of the gene coexpression network from the F2 XY panel expression data, we used a FOCI model (Magwene and Kim 2004). FOCI modeling is a type of graphical modeling approach that is useful for exploring the correlation structure of multivariate data sets. This and related methods are particularly appropriate for genomic studies where there are many genes to consider, but only a modest number of samples are available (Magwene and Kim 2004; Wille et al. 2004; Wille and Buhlmann 2006.) To minimize the potential for spurious coexpression relationships arising from samples or genes with significant amounts of missing data, we removed samples that were missing expression data for greater than one-third of the genes analyzed (12 out of 80 samples removed) and removed genes that were missing data in >20% of the F2 samples (10 out of 50 genes removed). In total, expression data for a reduced subset of 40 genes in 68 XY F2 samples were included in the FOCI analysis. We used a modified FDR approach (Benjamini and Yekutieli 2001; Drton and Perlman 2007) to control the edgewise FDR for the FOCI graph (FDR = 0.05), and edges in the graph were weighted based on partial correlation thresholds of 0.25 and 0.33 (stricter threshold = heavier weight).

#### SNP genotyping and eQTL analysis

For SNP genotyping, genomic DNA was first extracted from individual F2 samples using the Qiagen DNeasy kit (catalog no. 69504) according to the manufacturer's instructions. DNA from each sample (150 ng total) was then applied to BioTrove OpenArray slides according to BioTrove protocol. OpenArray slides were configured to genotype F2 samples at 117 SNPs (located evenly throughout the genome) by TaqMan-based chemistry with FAM- and HEX-labeled probes. Slides were cycled and scanned on the OpenArray NT Cyclor (BioTrove), and genotype clusters were individually determined in the accompanying OpenArray SNP genotyping software (BioTrove). An additional 11 SNPs (plus one positive control) from Chrs 1 and 12 were genotyped independently in each F2 sample using a standard TaqMan reaction on a StepOnePlus thermal cyclor (Applied

Biosystems). In this case, 5 ng total of genomic DNA were analyzed in 10- $\mu$ L reactions, and allele-specific TaqMan probes were conjugated to FAM and VIC dyes. F2 Expression data, SNP identifiers, marker position, and individual sample genotypes at each SNP are listed in Supplemental Table 6.

Genetic linkage analysis was conducted in J/qtl (Broman et al. 2003) to identify regions of the genome (eQTLs) that correlated with differences in gene expression in the F2 population. Y-chromosome origin (B6 or 129S1) and developmental stage (ts number) were included as additive covariates in all scans to account for potential small confounding effects from these variables. For each gene, one QTL genome scan was performed using the expectation maximization (EM) algorithm (Lander and Botstein 1989) with default convergence parameters (Maximum number of iterations = 4000; convergence tolerance = 1E-4). Physical SNP locations were based on NCBI build 37 and converted to genetic map positions using the Mouse Map Converter function within the JAX Center for Genome Dynamics Web site (<http://cgd.jax.org/mousemapconverter>). Significance thresholds were determined empirically by permuting the expression data for each individual gene 1000 times. eQTLs were deemed suggestive if they met or exceeded a 37% genome-wide adjusted threshold, and significant if they met or exceeded a 95% genome-wide adjusted threshold. Bayesian 0.95 credible intervals were determined for each eQTL in J/qtl.

#### Acknowledgments

We are grateful to Dr. Fernando Pardo-Manuel de Villena and Timothy Bell for the use of and assistance with the BioTrove OpenArray platform. We thank the laboratories of Thomas Denny and Victor Dzau at Duke University for their generous loan of multichannel pipettes and time on the StepOnePlus thermocycler. We also thank Lindsey Barske for artistic support with figures, Iordan Batchvarov for assistance with animals, and members of the Capel laboratory for many helpful discussions. Microarray analyses were performed using BRB-ArrayTools version 3.7.0 release developed by Dr. Richard Simon and Amy Peng Lam. This work was funded by a generous seed grant from the Duke University IGSP Center for Systems Biology (to B.C.) and bridge funding from the Duke University Medical Center (to B.C.), for which we are grateful.

#### References

- Albrecht KH, Eicher EM. 1997. DNA sequence analysis of *Sry* alleles (subgenus *Mus*) implicates misregulation as the cause of C57BL/6J-Y<sup>POS</sup> sex reversal and defines the SRY functional unit. *Genetics* **147**: 1267–1277.
- Albrecht KH, Young M, Washburn LL, Eicher EM. 2003. *Sry* expression level and protein isoform differences play a role in abnormal testis development in C57BL/6J mice carrying certain *Sry* alleles. *Genetics* **164**: 277–288.
- Barrionuevo F, Bagheri-Fam S, Klattig J, Kist R, Taketo MM, Englert C, Scherer G. 2006. Homozygous inactivation of *Sox9* causes complete XY sex reversal in mice. *Biol Reprod* **74**: 195–201.
- Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol* **57**: 289–300.
- Benjamini Y, Yekutieli D. 2001. The control of the false discovery rate in multiple testing under dependency. *Ann Stat* **29**: 1165–1188.
- Beverdam A, Koopman P. 2006. Expression profiling of purified mouse gonadal somatic cells during the critical time window

- of sex determination reveals novel candidate genes for human sexual dysgenesis syndromes. *Hum Mol Genet* **15**: 417–431.
- Bouma GJ, Albrecht KH, Washburn LL, Recknagel AK, Churchill GA, Eicher EM. 2005. Gonadal sex reversal in mutant *Dax1* XY mice: A failure to upregulate *Sox9* in pre-Sertoli cells. *Development* **132**: 3045–3054.
- Bouma GJ, Affourtit JP, Bult CJ, Eicher EM. 2007. Transcriptional profile of mouse pre-granulosa and Sertoli cells isolated from early-differentiated fetal gonads. *Gene Expr Patterns* **7**: 113–123.
- Brennan J, Capel B. 2004. One tissue, two fates: Molecular genetic events that underlie testis versus ovary development. *Natl Rev* **5**: 509–521.
- Broman KW, Wu H, Sen S, Churchill GA. 2003. R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**: 889–890.
- Bullejos M, Koopman P. 2005. Delayed *Sry* and *Sox9* expression in developing mouse gonads underlies B6-Y<sup>DOM</sup> sex reversal. *Dev Biol* **278**: 473–481.
- Burgoyne PS. 1988. Role of mammalian Y chromosome in sex determination. *Philos Trans R Soc Lond* **322**: 63–72.
- Capel B. 2006. *R-spondin1* tips the balance in sex determination. *Nat Genet* **38**: 1233–1234.
- Chaboissier MC, Kobayashi A, Vidal VI, Lutzkendorf S, van de Kant HJ, Wegner M, de Rooij DG, Behringer RR, Schedl A. 2004. Functional analysis of *Sox8* and *Sox9* during sex determination in the mouse. *Development* **131**: 1891–1901.
- Colvin JS, Green RP, Schmahl J, Capel B, Ornitz DM. 2001. Male-to-female sex reversal in mice lacking fibroblast growth factor 9. *Cell* **104**: 875–889.
- Cory AT, Boyer A, Pilon N, Lussier JG, Silversides DW. 2007. Presumptive pre-Sertoli cells express genes involved in cell proliferation and cell signalling during a critical window in early testis differentiation. *Mol Reprod Dev* **74**: 1491–1504.
- Dipple KM, McCabe ER. 2000. Modifier genes convert ‘simple’ Mendelian disorders to complex traits. *Mol Genet Metab* **71**: 43–50.
- Drton M, Perlman MD. 2007. Multiple testing and error control in gaussian graphical model selection. *Stat Sci* **22**: 430–449.
- Eicher EM, Washburn LL. 1986. Genetic control of primary sex determination in mice. *Annu Rev Genet* **20**: 327–360.
- Eicher EM, Washburn LL, Whitney JB II, Morrow KE. 1982. *Mus poschiavinus* Y chromosome in the C57BL/6J murine genome causes sex reversal. *Science* **217**: 535–537.
- Eicher EM, Shown EP, Washburn LL. 1995. Sex reversal in C57BL/6J-Y<sup>POS</sup> mice corrected by a *Sry* transgene. *Philos Trans R Soc Lond* **350**: 263–268.
- Eicher EM, Washburn LL, Schork NJ, Lee BK, Shown EP, Xu X, Dredge RD, Pringle MJ, Page DC. 1996. Sex-determining genes on mouse autosomes identified by linkage analysis of C57BL/6J-Y<sup>POS</sup> sex reversal. *Nat Genet* **14**: 206–209.
- Fleming A, Vilain E. 2005. The endless quest for sex determination genes. *Clin Genet* **67**: 15–25.
- Foster JW, Dominguez-Steglich MA, Guioli S, Kowk G, Weller PA, Stevanovic M, Weissenbach J, Mansour S, Young ID, Goodfellow PN, et al. 1994. Campomelic dysplasia and autosomal sex reversal caused by mutations in an SRY-related gene. *Nature* **372**: 525–530.
- Gubbay J, Collignon J, Koopman P, Capel B, Economou A, Munsterberg A, Vivian N, Goodfellow P, Lovell-Badge R. 1990. A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. *Nature* **346**: 245–250.
- Hacker A, Capel B, Goodfellow P, Lovell-Badge R. 1995. Expression of *Sry*, the mouse sex determining gene. *Development* **121**: 1603–1614.
- Hiramatsu R, Matoba S, Kanai-Azuma M, Tsunekawa N, Katoh-Fukui Y, Kurohmaru M, Morohashi KI, Wilhelm D, Koopman P, Kanai Y. 2008. A critical time window of *Sry* action in gonadal sex determination in mice. *Development* **136**: 120–138.
- Jansen RC, Nap JP. 2001. Genetical genomics: The added value from segregation. *Trends Genet* **17**: 388–391.
- Jordan BK, Vilain E. 2002. *Sry* and the genetics of sex determination. *Adv Exp Med Biol* **511**: 1–13.
- Kim Y, Kobayashi A, Sekido R, DiNapoli L, Brennan J, Chaboissier MC, Poulat F, Behringer RR, Lovell-Badge R, Capel B. 2006. *Fgf9* and *Wnt4* act as antagonistic signals to regulate mammalian sex determination. *PLoS Biol* **4**: e187. doi: 10.1371/journal.pbio.0040187.
- Koopman P, Munsterberg A, Capel B, Vivian N, Lovell-Badge R. 1990. Expression of a candidate sex-determining gene during mouse testis differentiation. *Nature* **348**: 450–452.
- Lander ES, Botstein D. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**: 185–199.
- Maatouk DM, DiNapoli L, Alvers A, Parker KL, Taketo MM, Capel B. 2008. Stabilization of  $\beta$ -catenin in XY gonads causes male-to-female sex-reversal. *Hum Mol Genet* **17**: 2949–2955.
- Magwene PM, Kim J. 2004. Estimating genomic coexpression networks using first-order conditional independence. *Genome Biol* **5**: R100. doi: 10.1186/gb-2004-5-12-r100.
- Marshall OJ. 2004. PerlPrimer: Cross-platform, graphical primer design for standard, bisulphite and real-time PCR. *Bioinformatics* **20**: 2471–2472.
- Meeks JJ, Weiss J, Jameson JL. 2003. *Dax1* is required for testis determination. *Nat Genet* **34**: 32–33.
- Mittwoch U. 2006. Sex is a threshold dichotomy mimicking a single gene effect. *Trends Genet* **22**: 96–100.
- Morrish BC, Sinclair AH. 2002. Vertebrate sex determination: Many means to an end. *Reproduction* **124**: 447–457.
- Nef S, Schaad O, Stallings NR, Cederroth CR, Pitetti JL, Schaer G, Malki S, Dubois-Dauphin M, Boizet-Bonhoure B, Descombes P, et al. 2005. Gene expression during sex determination reveals a robust female genetic program at the onset of ovarian development. *Dev Biol* **287**: 361–377.
- Nikolova G, Sinsheimer JS, Eicher EM, Vilain E. 2008. The Chromosome 11 region from strain 129 provides protection from sex reversal in XY<sup>POS</sup> mice. *Genetics* **179**: 419–427.
- Palmer SJ, Burgoyne PS. 1991a. In situ analysis of fetal, prepuberal and adult XX ↔ XY chimaeric mouse testes: Sertoli cells are predominantly, but not exclusively, XY. *Development* **112**: 265–268.
- Palmer SJ, Burgoyne PS. 1991b. The *Mus musculus domesticus* *Tdy* allele acts later than the *Mus musculus musculus* *Tdy* allele: A basis for XY sex-reversal in C57BL/6-Y<sup>POS</sup> mice. *Development* **113**: 709–714.
- Sekido R, Lovell-Badge R. 2008. Sex determination involves synergistic action of SRY and SF1 on a specific *Sox9* enhancer. *Nature* **453**: 930–934.
- Simon RM, Lam A, Li MC, Ngan M, Menendez S, Zhao Y. 2007. Analysis of gene expression data using BRB-ArrayTools. *Cancer Inform* **2**: 11–17.
- Sinclair AH, Berta P, Palmer MS, Hawkins JR, Griffiths BL, Smith MJ, Foster JW, Frischauf AM, Lovell-Badge R, Goodfellow PN. 1990. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature* **346**: 240–244.
- Small CL, Shima JE, Uzumcu M, Skinner MK, Griswold MD. 2005. Profiling gene expression during the differentiation and

- development of the murine embryonic gonad. *Biol Reprod* **72**: 492–501.
- Syed HA, Threadgill DW. 2006. Enhanced oligonucleotide microarray labeling and hybridization. *Biotechniques* **41**: 685–686.
- Upadhyay G, Goessling W, North TE, Xavier R, Zon LI, Yin J, Yin V. 2008. Molecular association between  $\beta$ -catenin degradation complex and Rac guanine exchange factor DOCK4 is essential for Wnt/ $\beta$ -catenin signaling. *Oncogene* **27**: 5845–5855.
- Vainio S, Heikkilä M, Kispert A, Chin N, McMahon AP. 1999. Female development in mammals is regulated by *Wnt-4* signalling. *Nature* **397**: 405–409.
- Vidal VP, Chaboissier MC, de Rooij DG, Schedl A. 2001. *Sox9* induces testis development in XX transgenic mice. *Nat Genet* **28**: 216–217.
- Wagner T, Wirth J, Meyer J, Zabel B, Held M, Zimmer J, Pasantes J, Bricarelli FD, Keutel J, Hustert E, et al. 1994. Autosomal sex reversal and campomelic dysplasia are caused by mutations in and around the *SRY*-related gene *SOX9*. *Cell* **79**: 1111–1120.
- Wessel J, Zapala MA, Schork NJ. 2007. Accommodating pathway information in expression quantitative trait locus analysis. *Genomics* **90**: 132–142.
- Whitney JB, Mills TM, Lewis RW, Wartell R, Abney TO. 2000. A single genetic determinant that prevents sex reversal in C57BL-Y<sup>POS</sup> congenic mice. *Biochem Genet* **38**: 119–137.
- Wilhelm D, Palmer S, Koopman P. 2007. Sex determination and gonadal development in mammals. *Physiol Rev* **87**: 1–28.
- Wilhelm D, Washburn LL, Truong V, Fellous M, Eicher EM, Koopman P. 2009. Antagonism of the testis- and ovary-determining pathways during ovotestis development in mice. *Mech Dev* **126**: 324–336.
- Wille A, Buhlmann P. 2006. Low-order conditional independence graphs for inferring genetic networks. *Stat Appl Genet Mol Biol* **5**: 1–34 Article 1.
- Wille A, Zimmermann P, Vranova E, Furholz A, Laule O, Bleuler S, Hennig L, Prelic A, von Rohr P, Thiele L, et al. 2004. Sparse graphical Gaussian modeling of the isoprenoid gene network in *Arabidopsis thaliana*. *Genome Biol* **5**: R92. doi: 10.1186/gb-2004-5-11-r92.
- Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J, Speed TP. 2002. Normalization for cDNA microarray data: A robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res* **30**: e15. doi: 10.1093/nar/30.4.e15.
- Yao HH, Matzuk MM, Jorgez CJ, Menke DB, Page DC, Swain A, Capel B. 2004. Follistatin operates downstream of *Wnt4* in mammalian ovary organogenesis. *Dev Dyn* **230**: 210–215.
- Zhu J, Zhang B, Smith EN, Drees B, Brem RB, Kruglyak L, Bumgarner RE, Schadt EE. 2008. Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. *Nat Genet* **40**: 854–861.