Switching on sex: transcriptional regulation of the testis-determining gene Sry

Christian Larney, Timothy L. Bailey and Peter Koopman*

ABSTRACT
Mammalian sex determination hinges on the development of ovaries or testes, with testis fate being triggered by the expression of the transcription factor sex-determining region Y (Sry). Reduced or delayed Sry expression impairs testis development, highlighting the importance of its accurate spatiotemporal regulation and implying a potential role for SRY dysregulation in human intersex disorders. Several epigenetic modifiers, transcription factors and kinases are implicated in regulating Sry transcription, but it remains unclear whether or how this farrago of factors acts co-ordinately. Here we review our current understanding of Sry regulation and provide a model that assembles all known regulators into three modules, each converging on a single transcription factor that binds to the Sry promoter. We also discuss potential future avenues for discovering the cis-elements and trans-factors required for Sry regulation.

KEY WORDS: Disorders of sex development, Sex determination, Sry, Testis development, Transcriptional regulation

Introduction
Sex in mammals is usually determined by the presence or absence of Sry, a small, single-exon gene located in the male-specific region of the Y chromosome (Gubbay et al., 1990; Sinclair et al., 1990). The expression of Sry in the genital ridges typically results in their development into testes, whereas the absence or dysfunction of Sry leads to the development of ovaries (Fig. 1A). Sry is the only gene from the Y chromosome required for testis determination, as a 14 kb genomic DNA fragment containing Sry and no other genes is able to direct male development in XX transgenic mice (Koopman et al., 1991).

Sry encodes a transcription factor that acts by binding to and activating the testis-specific enhancer (TES) of the related gene Sox9 (Sekido and Lovell-Badge, 2008). A network of gene activity downstream of Sox9 then promotes male development while simultaneously impeding the gene network that drives ovarian development (reviewed by Warr and Greenfield, 2012) (Fig. 1B). These two alternative gene regulatory networks give the bipotential gonad its unique ability to differentiate into two morphologically and functionally distinct organs.

Even though Sry triggers a cascade that influences sex-specific development throughout the organism and has lifelong consequences, its expression is limited to only a few cells and, at least in mice, to a narrow time window (Fig. 1C,D). Sry is expressed exclusively within the supporting cells of the genital ridges, directing them towards the fate of a Sertoli cell (Albrecht and Eicher, 2001), which are the primary constituents of testis cords and which sequester primordial germ cells and facilitate spermatogenesis. Other testicular cell types do not express Sry and adopt a male fate in response to molecular signals arising from Sertoli cells. Sry expression is detectable in mice from 10.5 days post-coitum (dpc) (Koopman et al., 1990; Hacker et al., 1995; Jeske et al., 1996) and is initially restricted to the central region of the gonad (Bullejos and Koopman, 2001). This region of Sry expression then rapidly expands to encompass the entire gonad at ~11.5 dpc (Fig. 1D) before being extinguished, first in the central region and then outwards towards the poles. Expression is undetectable by 12.5 dpc. This expression profile appears to place tests determination on a knife edge, with slight decreases in Sry expression levels, or delays of peak expression by as little as a few hours, leading to failure of the testis-determining pathway and malapropos development of either ovotestes or ovaries (Hiramatsu et al., 2009; Wilhelm et al., 2009).

Accurate transcriptional upregulation of Sry is, therefore, essential for triggering male development, and malfunction of this mechanism is likely to be responsible for at least some of the many unexplained cases of human XY disorders of sexual development (DSD). Many attempts have been made to understand how Sry expression is so precisely upregulated, but owing to a unique combination of circumstances, a full explanation has not been forthcoming (see Box 1). Even less is known about how Sry expression is extinguished (see Box 2). Despite these obstacles, the clear requirement for timely, localised expression of Sry in the genital ridges during sex determination demands that a robust regulatory system must exist. In this Review, we summarise our current understanding of the molecular mechanisms contributing to the regulation of Sry transcription. Although epigenetic modifications within the Sry locus have been identified and implicated in regulating Sry expression (see Box 3), we focus here on transcription factor pathways and propose a model in which more than a dozen different factors fall into three modules responsible for Sry upregulation. We also suggest potential avenues for closing the remaining gaps.

Transcription factors and proximal cis-elements that regulate Sry: an integrative model of Sry regulation
Biochemical analyses and/or studies in transgenic mice have implicated upwards of a dozen factors in regulating Sry. However, a coherent model to explain the organisation of these factors has been lacking. Our model proposes that three key transcription factors bind to the Sry locus, and that each is the focus of a regulatory module comprising various and overlapping combinations of other factors, which function as partner proteins, post-translational modifiers and upstream regulators of the primary factors (Fig. 2). Evidence for this model is discussed in the following sections.

Module 1: pathways converging on GATA4
The expression of Sry is known to be dependent on the presence of the transcription factor GATA4 (Tevosian et al., 2002). Gata4 knockout induced at 10.5 dpc, but not at later times, results in sex reversal,
Mutations in the related protein MAP3K1 also lead to human XY DSD (Pearlman et al., 2010), which evidently results from downregulation of testis-inducing genes, including SRY, and upregulation of ovary-inducing genes (Loke et al., 2013). However, homozygosity for a non-functional allele of Map3k1 leads to only minor testis abnormalities in XY mice (Warr et al., 2011), leaving its role in regulating Sry unclear. It is possible that different members of the MAP3K family are responsible for regulating Sry in different species, a possibility that is given weight by the lack of MAP3K4 mutations in human DSD. Alternatively, because human DSD phenotypes result from single nucleotide changes in MAP3K1, whereas complete loss of function has little

Box 1. Understanding Sry regulation: why has it been so difficult?

A number of idiosyncrasies of Sry have combined to render the question of its regulation impenetrable to date.

(1) **Sry is located on the Y chromosome.** This location excludes it from repair mechanisms available to autosomal or X-linked genes, so potential regulatory sequence motifs are easily degraded. Moreover, XX rather than XY samples are used in most genome sequencing projects (Hughes and Rozen, 2012), such that relatively few Y chromosome sequences are available, thus further impeding studies.

(2) **Mouse Sry is embedded in a large inverted repeat.** As such, any potential regulatory element will normally reside at a similar distance both 5' and 3' of the Sry ORF (Gubbay et al., 1992), complicating loss-of-function analyses.

(3) **Mouse Sry has more than one promoter.** In addition to its principal linear mRNA transcript, Sry is able to generate transcripts that arise from a second, more distal 5' promoter in the inverted repeat region. The RNA generated forms a stem-loop structure that can be spliced to form circular transcripts, which are not loaded onto polysomes and are mostly produced in germ cells of the adult testis, so a role in sex determination is unlikely (Capel et al., 1993a; Dolci et al., 1997). The question of how this second promoter is regulated has received little attention.

(4) **Sry is amplified in several species.** Rats have 11 non-identical copies of Sry (Prokop et al., 2013), at least six of which are expressed (Turner et al., 2007); rabbits (Geralses and Ferrand, 2006), pigs (Groen et al., 2012) and dogs (Li et al., 2013) also have multiple Sry genes. It is not known which copy (or copies) carries the regulatory elements required for tests determination, confounding efforts to identify these sequences.

(5) **Sry expression in mice is limited in time and space.** Only a limited quantity of Sry-expressing tissue can be collected, making it difficult to study Sry regulation experimentally. Moreover, regulatory models must account not only for Sry activation and silencing, but also for the asynchronous, centre-to-pole wave of expression that occurs.

(6) **Sry regulation is intertwined with cell proliferation in the genital ridges.** A number of mutations in candidate Sry-regulating genes result in decreased Sry expression (Bradford et al., 2009; Fujimoto et al., 2013; Katoh-Fukui et al., 2012). In these situations, it is necessary to determine whether reduced expression is a result of lower Sry expression in each cell, and hence supportive of a defect in Sry regulation, or simply reflects a reduced population of Sry-expressing cells.

(7) **Sry expression is variable between species.** Compared with mice, Sry expression is more widespread in humans and wallabies and more durable in sheep and rabbits. This implies that the regulatory sequences are not necessarily conserved in evolution, further limiting the usefulness of comparative studies between species.

(8) **A suitable cell culture model is lacking.** Cell lines corresponding to endogenous Sry-expressing pre-Sertoli cells remain unavailable, and primary cell lines are quick to extinguish Sry. These limitations hamper biochemical analyses of transcription factor binding to putative regulatory sequences flanking Sry.

indicating that GATA4 is important for the earliest steps of male sex determination (Manuylov et al., 2011). Co-transfection experiments in HeLa cells indicate that GATA4 transactivates Sry promoter constructs from both mouse and pig, but not human (Miyamoto et al., 2008). In addition, PCR amplification of sequences obtained through chromatin immunoprecipitation (ChIP) has confirmed that GATA4 binds to DNA in two regions (Fig. 2) of the mouse Sry promoter (Gierl et al., 2012). Interaction between GATA4 and its co-factor FOG2 (also known as ZFPM2) is also required for Sry expression. Transgenic mice with a null FOG2 allele, or those with a targeted mutation in Foga4 that eliminates its interaction with FOG2, display reduced levels of Sry expression (Tevosian et al., 2002). FOG2 expression, in turn, appears to be regulated by the homeobox transcription factors SIX1 and SIX4. Accordingly, SIX1/SIX4 double-knockout mice express significantly lower levels of both FOG2 and SRY, and exhibit XY sex reversal (Fujimoto et al., 2013).

Recent studies have also shown that GATA4 binds to the Sry promoter only after the protein has been phosphorylated (Gierl et al., 2012). GATA4 phosphorylation is evidently carried out by p38 MAP kinases, as transgenic mice with a p38β (also known as Mapk11) null allele and a p38α (Mapk14) allele that is conditionally deleted in epiblast-derived tissues exhibit reduced Sry expression leading to sex reversal (Warr et al., 2012). Upstream components of the p38 MAPK cascade have also been implicated in regulating Sry expression; the gene encoding mitogen activated protein kinase kinase kinase 4 (MAP3K4) was found in a forward genetic screen in mice to be mutated in XY gonadal sex reversal (Bogani et al., 2009).

A premature stop codon in Map3k4 in these mice resulted in a protein that lacked the kinase domain, rendering it non-functional. Quantitative RT-PCR showed a significant reduction in gonadwide Sry expression in these mutant mice, and immunofluorescence microscopy confirmed that this was partly due to reduced expression per cell.
**Box 2. Switching off Sry expression**

In mice, Sry has a ‘cameo role’, with expression in the genital ridges disappearing only 2 days after the gene is activated. What switches Sry off and why? Because Sry is downregulated just as SOX9 expression reaches its peak, it is commonly assumed that SOX9 represses Sry transcription. Prolonged expression of Sry in mice conditionally null for Sox9 supports this concept (Barrionuevo et al., 2006; Chaboissier et al., 2004), although the observation that Sry is also downregulated in the ovarian portion of ovoocytes, where SOX9 is not expressed, suggests that other factors are in play (Wilhelm et al., 2009). If SOX9 does contribute to Sry downregulation, this factor must be capable of acting as both a transcriptional activator and repressor in the same cell type. Nonetheless, two observations suggest that this overall phenomenon is not important for sex determination. First, the rapid downregulation of Sry is observed only in mice; in several other mammalian species, Sry expression has been shown to continue for much longer, even into adult life. Second, in transgenic mouse models in which Sry is upregulated correctly but fails to be downregulated on time or at all, male development can proceed normally (KidoKoro et al., 2005). For these reasons, the cis-sequences and trans-acting factors involved in Sry downregulation in mice have not been characterised.

In effect in mice, it is possible that the human mutations cause a gain-of-function or even a dominant-negative effect that interferes with other steps in the regulatory pathway. Even more uncertainty surrounds the related MAP kinase kinase (MAP2K) family, the participating member(s) of which remain to be identified. Regardless, available evidence supports the MAPK cascade as a crucial regulator of GATA4 activity and, hence, Sry regulation (Fig. 2).

The activity of the MAPK cascade, in turn, is triggered by the stress-response protein GADD45G, which is able to bind to the N-terminus of MAP3K4, thereby disrupting an autoinhibitory state and instead activating the enzymatic activity of MAP3K4 (Miyake et al., 2007). Gadd45g is required for normal Sry expression (Gierl et al., 2012) and its loss leads to sex reversal (Johnen et al., 2013; Warr et al., 2012). Importantly, Gadd45g expression in mice proceeds in a centre-to-pole wave that foreshadows the expression of Sry by ~5 h (Warr et al., 2012). GADD45G also appears to link the insulin receptor family to the MAPK pathway upstream of Sry. The expression of both Gadd45g and Sry is perturbed in mice that are null for members of the insulin receptor tyrosine kinase family (Nef et al., 2003; Pitetti et al., 2013). For example, a peak in Sry expression in insulin receptor (Insr)/insulin-like growth factor 1 receptor (Igf1r) double-knockout mice is not reached until 13.5 dpc, 2 days later than usual, and greatly reduced levels of Gadd45g are observed at 11.5 dpc, highlighting that GADD45G and insulin signalling may be linked. In support of this, insulin signalling has been shown to be capable of inducing another member of the Gadd45 family, Gadd45b (Nakamura et al., 2004). These sites are largely hypermethylated again by 15.5 dpc, presumably required for these processes are recruited to the Sry promoter at the same stage in gonadal cells, but not in cells of the adjacent mesonephros, where Sry is not expressed. How JMJD1A and the other histone modifiers function (Katoh-Fukui et al., 2012). In these mice, Gadd45g expression is required for normal Sry expression (Gierl et al., 2012) (Fig. 2). Also, it is not clear whether it is the number of Sry-expressing cells or the level of Sry expression per cell (or indeed both) that is reduced in Cbx2 knockout gonads (see Box 1). In addition, given that other genes with roles in gonad development, including Gata4 and Lhx9, are also affected in Cbx2 knockout mice, Cbx2 might well have multiple roles in this system. It is also important to note that Polycomb repressive complex 1, of which CBX2 is a member, is known to have activating potential, for example of SF-1 and steroidogenesis (Morohashi et al., 1992). Studies using an electrophoretic mobility shift assay (EMSA) indicate that NR5A1 binds to specific sites in the pig and human Sry promoters (Pilon et al., 2003; De Santa Barbara et al., 2001). Directed mutagenesis of these sites leads to reduced Sry expression in cell-based reporter assays, although verifying these results *in vivo* is difficult because the sites in pig and human are not found in the corresponding position in mouse. Instead, sequences resembling the short NR5A1 core binding motif are found at many positions in the mouse Sry 5’ flanking sequence. It is assumed that NR5A1 binds somewhere in this region, but it is not clear where.

The expression of NR5A1, in turn, appears to be regulated by the Polycomb group protein CBX2. XY sex reversal occurs in humans carrying CBX2 mutations (Biason-Lauber et al., 2009) and in mice lacking Cbx2 function (Katoh-Fukui et al., 2012). In these mice, Sry is expressed at negligible levels and far fewer cells express Nr5a1 compared with wild-type gonads, suggesting that, rather than directly regulating Sry, CBX2 might mediate its activities through the regulation of Nr5a1 (Katoh-Fukui et al., 2012) (Fig. 2). Also, it is not clear whether it is the number of Sry-expressing cells or the level of Sry expression per cell (or indeed both) that is reduced in Cbx2 knockout gonads (see Box 1). In addition, given that other genes with roles in gonad development, including Gata4 and Lhx9, are also affected in Cbx2 knockout mice, CBX2 might well have multiple roles in this system. It is also important to note that Polycomb repressive complex 1, of which CBX2 is a member, is known to have activating potential, for example of Nr5a1 during adrenal development (Katoh-Fukui et al., 2005), despite more commonly being considered a transcriptional silencer.

A second factor likely to interact with Nr5a1 to regulate Sry is CITED2 (Cbp/p300-interacting transactivator with Glu/Asp-rich carboxy-terminal domain 2). The expression of both Nr5a1 and Sry...
expression per cell is affected. The molecular details of how SIX1/SIX4 might act in this system are currently unknown.

Module 3: gene regulation via WT1

The third major regulatory module is provided directly by WT1 (Fig. 2). In vitro cell co-transfection experiments in HeLa cells indicate that WT1 is able to transactivate Sry promoter constructs from mouse, pig and human (Miyamoto et al., 2008). Subsequent mutational analyses confirmed a requirement for direct DNA binding in this regulation in both mice and humans (Miyamoto et al., 2008; Hossain and Saunders, 2001). The importance of WT1 has also been shown in vivo, as XY mice lacking the +KTS isoform express Sry at only 25% of wild-type levels and fail to develop testes (Hammes et al., 2001). Reduced Sry expression is caused both by lower expression levels of Sry per cell and by a reduction in the number of Sry-expressing cells, which occurs as a result of reduced cell proliferation (Bradford et al., 2009). This reduced cell proliferation is not observed in similarly mutated XX embryos, implying that it occurs downstream of reduced Sry expression in addition to being part of the cause.

It has been speculated that CITED2 participates in the direct regulation of Sry by WT1, but compelling evidence remains lacking (Buaas et al., 2009). By contrast, co-transfection of Sry promoter constructs with both WT1 and GATA4 shows synergistic effects in mice, pigs and human (Miyamoto et al., 2008), suggesting that GATA4 and WT1 might physically interact to regulate Sry.

The search for distal cis-regulatory elements

Above, we have proposed that various regulators converge on three distinct factors—GATA4, NR5A1 and WT1—to directly regulate the expression of Sry. Although binding sites that might allow both WT1 and GATA4 to regulate Sry have been identified (Gierl et al., 2012; Miyamoto et al., 2008), these sites are restricted to the proximal promoter, within just a few hundred base pairs of the transcription start site (TSS) (Fig. 2). The presumed binding site(s) for NR5A1 have not been identified, and might be proximal or more distal in position. In the following sections, we consider evidence from transgenic mice, chromosomal rearrangements, and comparative analyses of flanking sequences that suggests the presence of other cis-regulatory elements at more distal positions than those currently known.

Sry transgene constructs that induce XX sex reversal

Sry was initially confirmed as the testis-determining factor by making XX sex-reversed transgenic mice using a 14.6 kb genomic DNA fragment known as L741 (Fig. 3A) that contains the Sry gene, ~8 kb of 5′ flanking sequence and ~3 kb of 3′ flanking sequence (Koopman et al., 1991). As a result, the assumption that all regulatory sequences required for Sry function in sex determination reside within that fragment has become the cornerstone for attempts to understand Sry regulation. More specifically, it has been assumed that regulatory sequences will lie in the 5′ region of the construct, with little consideration of possible regulatory sequences located 3′. This assumption is supported by the fact that the 3 kb of 3′ flanking sequence in L741 duplicates sequences in the 5′ flank (Gubbay et al., 1992). Subsequent studies showed that the same 8 kb of 5′ regulatory sequence driving an enhanced green fluorescent protein (EGFP) reporter transgene (Fig. 3B) accurately recapitulates the endogenous Sry centre-to-pole expression profile described previously (Albrecht and Eicher, 2001). Although original reports suggested that the expression of this reporter transgene continued after 12.5 dpc, a more recent study found it to be sex-specifically extinguished in XY embryos at the correct time (Jameson et al., 2012). Together, these
studies strongly suggest that the sequences required to correctly regulate mouse Sry expression are located in the 8 kb immediately 5′ to the Sry ORF.

Shorter Sry transgenic constructs that use an endogenous ORF and nested deletions of 5′ flanking sequence are also able to induce sex reversal in mice with reasonable efficiency (Fig. 3C,D) (Bowles et al., 1999; Koopman, 2002). Although this ability might be linked to the presence of multiple copies of the transgene in the genome of these animals, an alternative explanation for the efficacy of these constructs is that they contain the recognised WT1 binding site (Miyamoto et al., 2008) (Fig. 2). However, the pattern of expression that they generate has not been detailed.

Another putative regulatory region within the mouse Sry locus has been indicated by experiments using Sry/Cre fusion constructs to drive lacZ expression (Ito et al., 2005). Echoing the above results obtained using transgenes containing the endogenous Sry ORF, these fusion constructs were found to be expressed in the gonad regardless of the length of 5′ sequence that they contain. However, whereas a construct with 400 bp of 5′ sequence (Fig. 3E) was highly expressed in cultured gonadal cells at 11.5 and 13.5 dpc, as well as in samples from liver and brain, a construct with 500 bp of 5′ sequence (Fig. 3F) was exclusively expressed in 11.5 dpc gonadal cells, mirroring endogenous Sry expression. Although expression of the longer construct occurred at much lower levels than that of the shorter construct, the different stage- and tissue-specific patterns that they give rise to suggest that an important cell-specific regulatory element is present between positions −484 and −371 (Fig. 3, orange box between E and F). A function for this region might be restricted to the mouse, given that it is not conserved in any other species studied (Ito et al., 2005).

Also in mouse, a 58 bp region ∼2.4 kb 5′ to Sry was found to be bound by nuclear extracts from 11.5 dpc gonads when assayed by EMSA (Fig. 3G) (Yokouchi et al., 2003). However, extracts from both 12.5 dpc and 13.5 dpc gonads failed to bind to this region, suggesting stage-specific regulation during the sex-determining period.

The regulatory regions within human, pig and goat Sry loci have also been examined. For example, it was shown that human SRY regulatory regions remain functional in mice. When driven by 5 kb of human 5′ sequence (Fig. 3H), a reporter transgene is expressed in...
the murine gonad, but this expression is absent when using a short 3.3 kb construct (Fig. 3I) (Boyer et al., 2006), suggesting the presence of an important element in the −3.1 to −4.8 kb region (Fig. 3, orange box between H and I). Pig Sry regulatory regions are also functional in mice. Fluorescent reporters driven by either 1.6 kb (Fig. 3J) or 2.6 kb of pig 5′ sequence were shown to be expressed in the genital ridge of mice (Boyer et al., 2006). Furthermore, pig 5′ sequences of 4.5 kb and 1.4 kb (Fig. 3K) can drive robust reporter gene expression in porcine genital ridge (PGR) 9E11 cells (Pilon et al., 2003). Reporter expression was reduced by 40% when a shorter construct of 906 bp (leaving just 245 bp before the TSS) was used (Fig. 3L). In addition, removal of the S′ UTR from the reporter construct virtually eliminated expression, even if the remainder of the 4.5 kb construct was included (Fig. 3M,N). Elements that are essential for Sry expression are thus likely to be found in the S′ UTR between positions +47 and +660 (Fig. 3, orange box between L, M and N), while an additional element located somewhere between −245 and −765 (Fig. 3, orange box between K and L) appears to enhance the levels of transcription. Finally, it has been shown that goat Sry also has some function in mice, as XX transgenic mice with a 90 kb insert containing goat Sry, some 80 kb of 5′ sequence and 6 kb of 3′ sequence developed testes in two out of six cases (Fig. 3O) (Pannetier et al., 2006). However, 1.8 kb of a goat Sry promoter was unable to drive a Cre/loxP/ lacZ reporter when transfected into cells from 11.5 dpc genital ridges (Fig. 3P) (Ito et al., 2005), suggesting the presence of regulatory sequences in the goat beyond this region.

In summary, these results argue for the presence of regulatory regions up to several kilobases from the TSS (Fig. 3, orange bands). However, as discussed below, large-scale rearrangements of the Y chromosome involving deletions and inversions suggest either that even longer range regulation is occurring or that Sry is susceptible to position effects.

Deletions and rearrangements that result in XY sex reversal

Deletions in the regions flanking human SRY (Fig. 4) are associated with sex reversal but how they mediate this reversal is poorly understood. A large deletion, from ~1.8 kb 5′ of SRY to between 23 kb and 50 kb 5′ of the gene (McElreavey et al., 1992), has been noted in a patient (NV) with streak gonads and external female appearance (Fig. 4). This deletion potentially removes almost the entire sequence between SRY and the gene encoding RPS4Y1 (ribosomal protein S4, Y-linked 1), which is the nearest neighbour to SRY in the 5′ direction. Another patient (SC) with dygenetic testes and external female appearance was reported to have a 2.5-7 kb deletion (Fig. 4) beginning between 2 and 3 kb 5′ of SRY and extending into the pseudoautosomal region (McElreavey et al., 1996). Although the lack of corroborating patients makes proving a causal link between these mutations and the associated phenotypes difficult, it is known that both unaffected fathers lack similar mutations, which are thus de novo in the affected individuals.

Large deletions of the Y chromosome, estimated to be 3–4 Mb in size (Mahadeviah et al., 1998), also cause sex reversal in mice, but in this case the deletions are further away from the gene and are thought to act through a generalised position effect, bringing Sry closer to a heterochromatic region that prevents its expression (Capel et al., 1993b; Laval et al., 1995). They are, therefore, not instructive in locating potentially important regulatory intervals.

Two recent reports discuss XY sex reversal resulting from pericentric inversion of the Y chromosome (Fig. 5). In the first case, the breakpoint on the short arm was located ~93 kb 3′ of SRY, while the long arm breakpoint was located in the heterochromatin of the q11.2 cytoband (Gimelli et al., 2006) (Fig. 5A). This inversion places heterochromatin relatively close to the 3′ end of SRY, apparently silencing the gene completely. Accordingly, CpG sites located in the promoter of SRY in this patient were more highly methylated than the same sites in the patient’s unaffected father, suggesting that, as in the mouse, DNA methylation might be involved in SRY regulation in humans. In the second case, the inversion breakpoints were located in p11.2 and q11.2 (Fig. 5B). The short arm breakpoint was located some 350 kb 5′ of SRY, distinguishing this case from the previous one and suggesting some form of long-range 5′ regulation. RT-PCR of the patient’s streak gonads revealed SRY expression (Mitsuhashi et al., 2010), suggesting a more subtle misregulation than in the first case. Although both of these cases illustrate the sensitivity of SRY expression to its chromosomal position, it remains unclear how they might further inform the search for regulating factors and cis-regulatory regions.

Finally, it should be noted that a number of single nucleotide polymorphisms (SNPs) have been associated with XY sex reversal phenotypes (Kwok et al., 1996; Poulat et al., 1998; Veitia et al., 1997a). However, ambiguous results and small sample sizes make it difficult to assess whether these mutations are causative of sex reversal or simply a coincidence in otherwise affected individuals.

Insights from bioinformatic comparative analyses

A common strategy used to identify potential regulatory sites is comparative analysis of a gene’s flanking sequences. Unfortunately, the peculiar difficulties associated with Sry (Box 1) mean that these studies have been limited by the availability of only short sequences from the 5′ flank, with little consideration given to the possibility of regulatory sequences located in the 3′ region, even in non-mouse species in which the two sequences are independent. In addition, the use of different sets of species in each study – only four species appear in more than two studies – makes comparisons difficult. Finally, the use of pairwise, rather than multiple, alignments in a number of studies makes it difficult to assess just how well conserved the sequences are across a broad spectrum of species. Nonetheless, a number of comparative studies have revealed conserved sequences adjacent to the Sry coding sequence (Fig. 6, Table 1).

The region around the TSS has been found in all comparative studies to be somewhat conserved (Hacker et al., 1995; Margit et al., 1998; Pilon et al., 2003; Ross et al., 2008; Veitia et al., 1997a). This region includes putative binding sites for the transcription factor SP1 in all species studied, albeit with varying degrees of conservation. The function of two of these SP1 sites has been verified in the human NT2-D1 cell line (Descoleau et al., 1998). Although the role of SP1 as a ubiquitous transcription factor...
rules it out as the primary driver of the stage- and tissue-specific regulation of Sry, the presence of these sites suggests that the factors that control these aspects need to be competent to interact with SP1. Motifs for additional transcription factor families, including GATA and Oct factors, are also found in this region (Margarit et al., 1998), although the function of these motifs has not been reported.

In this same region around the TSS, Hacker et al. (1995) identified a 146 bp region that is conserved between humans (−58 to +88) and mice (−132 to +14; Fig. 6). Independently, Veitia et al. (1997b) aligned sequences from human, four other primate species, cow and pig and found a short conserved interval, which they labelled Motif C, that is located between −57 and −43 in human (Fig. 6). The fact that the regions identified in these two studies overlap suggests strong conservation of this region across a range of species. The only species found to lack Motif C to date has been sheep (Margarit et al., 1998), and a comparison of this motif with known transcription factor binding motifs suggests that it is competent to bind aryl hydrocarbon receptor nuclear translocator (ARNT), or possibly SOX9.

Veitia et al. (1997b) also identified two more distally located conserved motifs, which they termed Motif A and Motif B (Fig. 6). In human, these motifs are located at −373 to −362 and −349 to −340, respectively. Whereas Motif A is considered competent to bind HinfA or Oct factors, no transcription factor motifs could be associated with Motif B, and neither motif has been identified in other comparative studies.

Moving further away from Sry, the next significantly conserved region is a cluster of five transcription factor consensus binding motifs located within a region of just 90 bp (Ross et al., 2008) (Fig. 6). Located between positions −907 and −817 in human, this cluster is conserved between cow, human, goat and pig and is particularly remarkable because the five motifs are so densely clustered yet only nine putative transcription factor binding sites were found to be conserved between all four species elsewhere in the entire 5 kb of sequence that was analysed (Ross et al., 2008). The clustering of transcription factor motifs in this manner is often used to predict functional cis-regulatory elements (Spitz and Furlong, 2012). Of the five clustered sites, two could potentially bind caudal type homeobox (CDX) factors. Another is competent to bind a member of the zinc fingers, C2H2-type with BTB/POZ domain (ZBTB) family, and a further motif is predicted to bind NF-κB. The final conserved element is a serum-response element capable of binding serum response factor (SRF). However, there are no reports to date of any of these factors being involved in Sry regulation, and the functional significance of this cluster remains to be investigated.

Further 5′ of Sry, Hacker et al. (1995) recognised a second, ~100 bp region of conservation between human and mouse sequences (Fig. 6). This region lies within the 5′ arm of the
Ross et al. (2008) found that positions to the human sequence between 2202 and 1453. Independently, Ross et al. (2008) found that positions −1502 to −1452 of the human sequence are conserved with the cow sequence, and roughly the same portion of the human sequence has been found to share homology with the pig sequence (Pilon et al., 2003). Thus, a number of studies point to a region that is somewhat conserved between human, cow, pig and mouse (Fig. 6, leftmost green rectangle), although the functionality of this region, as for other potentially conserved regions, remains untested.

Conclusions and future directions

The various approaches described above, aimed at identifying and locating sequence motifs and factors important for Sry regulation, have yielded only modest progress to date. Genetic studies have identified three pivotal transcription factors – GATA4, WT1 and NR5A1 – that play important roles, each assisted by a suite of accessory factors arranged into proposed modules. For each module, it is clear that additional components might come to light through further studies, and suggested roles for the various components do not preclude the possibility of other roles for them in Sry regulation.

In addition, the specific binding sites occupied by these three factors, and the functionality of these sites, have yet to be confirmed with in vivo evidence. The evidence for DNA binding is, instead, restricted to regions of 100 bp or more (GATA4; Fig. 2), or else is based only on in vitro evidence (WT1; Fig. 2). Alignment-based comparative analysis has also only been able to provide hints of conserved regions that might indicate likely regulatory cis-elements. As a result, it is not known whether GATA4, WT1 and NR5A1 each bind single critical sites, nor whether the binding sites for these factors are clustered or dispersed. Furthermore, the putative regulatory sites for GATA4 and WT1 are located adjacent to the TSS, so all that is currently understood about Sry regulation is restricted to the proximal promoter region.

The existence of further distal regulatory elements seems likely from transgenesis and DNA rearrangement studies, even though the location of these elements remains unknown. Findings from alignment-based comparative analyses suggest that potential regulatory intervals might not be conserved between species, or at least might not occupy a similar position in the species being compared. Particularly problematic, given the role of the mouse as a primary model organism, are the different expression profiles of Sry and the high level of sequence variability between mouse and human. Furthermore, the presence of large inverted repeats surrounding the mouse Sry locus suggests that mouse Sry might have been transposed from a different position on the chromosome at some point in the past. Taken together, these observations suggest that there are some differences between Sry regulation in humans and mice, even though the primary function of the gene has clearly been conserved. Such differences may limit the usefulness of the mouse model as a means of identifying regulatory motifs that can be examined in detail for mutations in human XY gonadal dysgenesis cases.

Future studies of Sry regulation must both identify and functionally confirm putative regulatory sites. The comparative analysis of genomic DNA sequences from different species is likely to yield insights when longer Sry flanking sequences from a greater range of species become available. Such analyses would also benefit from the availability of Sry flanking sequences in a broader variety of mouse strains than the B6 sample currently used as the mouse reference genome. The sequencing of Y chromosomes from a variety of mouse strains is therefore a clear priority.

In addition to identifying regulatory sequences solely through analysis of genomic sequences, it is now also possible to directly assay cell-specific regulatory activity. Perhaps the most widely utilised of these methods is ChIP followed by sequencing (ChIP-seq) for histone modifications associated with active regulatory regions, in particular histone H3 monomethylation at lysine 4 (H3K4me1) and acetylation at lysine 27 (H3K27ac) (Creighton et al., 2010). Alternatively, the more recent finding that regulatory regions can be transcribed, producing so-called enhancer-derived RNAs (eRNAs) (Andersson et al., 2014; Hah et al., 2013; Melgar et al., 2011), shows that some gene expression assays, in particular global run-on sequencing (GRO-seq) (Core et al., 2008) and cap analysis of gene expression (CAGE) (Kodzius et al., 2006), can also predict the locations of cis-regulatory elements. Finally, promoter-enhancer interactions can be captured in situ through the use of chromosome conformation capture methods such as Hi-C (Lieberman-Aiden et al., 2009). Whereas ChIP-seq experiments continue to require in the order of 10⁶ cells (Adli and Bernstein, 2011), presenting a formidable target when working with embryonic gonads, protocols exist to perform CAGE (Islam et al., 2012) and Hi-C (Nagano et al., 2013) on single cells. Not only does this scale make these strategies more amenable to the small samples

---

Table 1. Species, sequence lengths and alignment methods used in comparative studies of Sry regulatory regions

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Length of sequences</td>
<td>Several (~8) kb</td>
<td>205 bp to 4.5 kb</td>
<td>346 bp to 4.5 kb</td>
<td>4.5 kb</td>
<td>4.6-17 kb</td>
</tr>
<tr>
<td>Alignment method</td>
<td>Pairwise</td>
<td>Multiple (Clustal W)</td>
<td>Multiple (Clustal)</td>
<td>Pairwise</td>
<td>Pairwise (AVID, LAGAN)</td>
</tr>
<tr>
<td>Human</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Pig</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Mouse</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Cow</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Chimpanzee</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Sheep</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Goat</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Rhesus macaque</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Olive baboon</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Green monkey</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Gorilla</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Gazelle</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Rat</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Guinea pig</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>
available from nascent organs, but, if applied to groups of cells from different parts of the developing gonad, it also provides the potential for understanding intra-tissue transcriptional dynamics, such as the wave pattern of Sry expression.

Whatever methods are used to identify putative regulatory regions, verifying the function of these regions remains a challenge. The availability of genome editing techniques, such as transcription activator-like effector nuclease (TALEN) (Ding et al., 2013) or clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR associated (Cas) (Wang et al., 2013a) technology, is likely to speed this effort. The utility of these techniques for manipulating the Y chromosome has already been proven (Kato et al., 2013; Wang et al., 2013b). Ultimately, identifying which cis-elements and trans-factors contribute to the correct regulation of Sry might have both immediate medical utility for those living with variations or disorders of sex development and an enduring relevance for basic biology.

Acknowledgements
We thank Dagmar Wilhelm for the image used in Fig. 1D and the anonymous reviewers for valuable suggestions.

Competing interests
The authors declare no competing financial interests.

Funding
This work was supported by research grants from the Australian Research Council (ARC), the National Health and Medical Research Council of Australia (NHMRC), and the US National Institutes of Health. C.L. is the recipient of an Australian Postgraduate Award and a University of Queensland Scholarship. P.K. is a Senior Principal Research Fellow of the NHMRC. Deposited in PMC for release after 12 months.

References


expression in mice require direct interaction between transcription partners GATA4 and FOG2. Development 129, 4627-4634.


