Chromobox homolog 2 (CBX2) is a chromatin modifier that plays an important role in sexual development and its disorders (disorders of sex development [DSD]), yet the exact rank and function of human CBX2 in this pathway remains unclear. Here, we performed large-scale mapping and analysis of in vivo target loci of the protein CBX2 in Sertoli-like NT-2D1 cells, using the DNA adenine methyltransferase identification technique. We identified close to 1600 direct targets for CBX2. Intriguingly, validation of selected candidate genes using qRT-PCR in cells overexpressing CBX2 or in which CBX2 has been knocked down indicated that several CBX2-responsive genes encode proteins that are involved in DSD. We further validated these effects on the candidate genes using a mutated CBX2 causing DSD in human patient. Overall, our findings suggest that CBX2 role in the sex development cascade is to stimulate the male pathway and concurrently inhibit the female pathway. These data provide fundamental insights into potential etiology of DSD.

The process of sexual differentiation is central for reproduction in almost all metazoan and, therefore, for maintenance of practically all multicellular organisms. The critical step in this process is the determination of the bipotential embryonic gonad into either ovaries or testes. Defects in any of the genes involved in either the testicular or ovarian development can result in disorders of sex development (DSD) (1, 2). Although there have been considerable advances in our understanding of the genetic factors involved in sexual differentiation, the fact remains that in most DSD the underlying genetic cause is unknown, in fact, it has been estimated that a molecular diagnosis is made in only about 20% of DSD (3, 4). This highlights the substantial gap in knowledge that needs to be addressed.

Chromobox homolog 2 (CBX2) was recently added as an additional player in human sex development process (5). CBX2 and its mouse homologue M33 are members of the polycomb (PcG) family of proteins. PcG proteins are highly conserved transcriptional regulators that form large multiprotein complexes, which exert their function through the modulation of higher order chromatin structure (6, 7). They are required for stable maintenance of gene expression programs through developmental and cell differentiation processes. Although best known for their role in controlling the expression of Hox genes in Drosophila during development, numerous PcG members have been now implicated in the control of various cellular processes, such as chromosome X-inactivation, cell fate decisions, tumorigenesis, and sexual differentiation (8–12). Human CBX2 exists in 2 isoforms, a 532-amino acid long isoform (NP_005180), called CBX2.1 (hereafter referred to as CBX2), and a second shorter 211-amino acid isoform (NP_116036), named CBX2.2. The importance of M33/CBX2 for sexual development is highlighted by the fact that M33-deficient mice have male-to-female sex reversal (13). Similarly, in humans, we made the intriguing discovery of a CBX2 loss-of-function
double heterozygote mutation state in a 46, XY girl with ovarian tissue at histology, normal uterus, and external genitalia, accidentally diagnosed because of discrepancy between prenatal karyotypes and phenotype at birth. Functional studies demonstrated that the mutated CBX2 does not properly bind to and does not adequately regulate the expression of target genes essential for sex development such as SF1/NR5A1 (5), which was in agreement with other data (14).

Although these data elucidated that CBX2 is involved in the molecular pathogenesis of DSD and that it lies upstream of SRY (Sex determining Region of the Y chromosome) in the sex determination cascade, little is known about the exact rank of CBX2 and its role in this cascade. To gain insight into CBX2 function, we have employed the DNA adenine methyltransferase (Dam) identification (DamID) coupled with next-generation sequencing (NGS) to analyze the human CBX2 localization across the genome and identify its direct targets.

In this study, we identified close to 1600 direct targets of CBX2 in human Sertoli-like cells. Furthermore, we provide evidence that the expression of some of the identified targets is CBX2 dependent, thereby identifying potential genes that could be involved in DSD in humans and setting the ground for future studies.

Materials and Methods

Cell culture

HEK293T (ATCC CRL-11268) and NT2-D1 (ATCC CRL-1973) cells were obtained from the American Type Culture Col-

A

DAPI

Dam-V5-CBX2

Merge

Figure 1. The DamID platform for location analysis of human CBX2. A, Dam-CBX2 localizes to the nucleus. NT2–D1 cells were transfected with the Dam-V5-CBX2; 48 hours later, cells were detergent extracted and fixed. Dam-V5-CBX2 was visualized by indirect immunofluorescence using a V5 antibody, nuclei were visualized by DAPI. Scale bar, 5 μm. B, Immunoblot analysis of Dam-V5-CBX2 expression in NT2–D1 cells. Whole-cell extracts of cells transfected with same construct as in A were analyzed by immunoblotting for the expression of Dam-V5-CBX2 using a monoclonal antibody against V5; tubulin is used as a loading control. C, Identified enriched regions localize relatively close to TSSs. In a 5-kb window around the TSS are 2 peaks in the distribution. The highest peak reflects binding events in the promoter region. Considering all identified putative binding regions there is a general peak close to a TSS.
lection and cultured at 37°C, 5% and 10% CO2, respectively, in DMEM containing 10% fetal bovine serum and supplemented with penicillin-streptomycin.

Small interference RNA (siRNA)

siRNA duplexes were purchased from Microsynth, siRNA sequences are as follows: (si scrambled) (CGUACGCGGAUACUUGCATT) (15), CBX2–145 (GGCUGGUCCUCCAAACAUATT), and CBX2-411 (GGAUGACAGUGAGUUAGAUTT) (5). siRNA duplexes were introduced into cells using Lipofectamine RNAiMAX (Invitrogen) in 2 consecutive rounds at a final concentration of 40nM. Experiments were typically performed 48 hours after the first transfection. For targets that did not show a significant change in expression at 48 hours, experiments were repeated at 72 hours.

Immunofluorescence

Cells grown on cover slips were preextracted for 5 minutes on ice using 25mM HEPES (pH 7.4), 50mM NaCl, 1mM EDTA, 3mM MgCl2, 300mM sucrose, and 0.5% Triton X-100 before fixation in 4% formaldehyde (wt/vol) in PBS for 15 minutes at room temperature. Cover slips were incubated overnight at 4°C with primary antibodies and Alexa Fluor-conjugated secondary antibodies (check Supplemental Information for detailed antibodies information) for 1 hour at room temperature. The cover slips were mounted with Vectashield (Vector Laboratories) containing DAPI (4',6-diamidino-2-phenylindole). Images were acquired using a Nikon eclipse Ni-E microscope system.

Dam identification

DamID was performed using the lentiviral transduction protocol as previously described (16), using pLgw-EcoDam-V5-CBX2 (EcoDam-CBX2) or pLgw-V5-EcoDam. Briefly, EcoDam-CBX2 or pLgw-V5-EcoDam was transfected into HEK239T cells together with the packaging plasmids (pRSV-Rev, pCMV-dR8.2, pMD2.G) using Metafectene (Biontex Laboratories); 48 hours later, the supernatant containing the lentivirus was harvested and added to NT-2D1 cells with polybrene 1 mg/mL (Santa Cruz Biotechnology, Inc), and 2 consecutive rounds of transduction were performed. Two days after the first infection, genomic DNA was isolated using DNeasy blood and tissue kit (QIAGEN). The isolated gDNA was subject to DpnII digestion, ligation of DamID adaptors and DpnII digestion. Samples without addition of DpnII or ligation adaptors served as negative controls. DpnII-digested fragments were used as templates to amplify methylated genomic fragments.

Next-generation sequencing

Library preparation was performed using standard procedures, see Supplemental Methods for a detailed protocol. The sequence alignment to the human genome (hg19) was performed with Bowtie2 using the options “--sensitive” and “--nondeterministic.” To identify peaks from the data, we used MACS (version 1.4) with default options. The subsequent analysis, including the annotation and visualization of peaks, was done with the Bioconductor-packages ChIPpeakAnno and Gviz from R.

Quantitative real-time PCR

Total RNA was extracted using the RNeasy Mini kit (QIAGEN) and reverse transcribed using the Omniscript reverse-transcription system (QIAGEN). Experiments were performed with an ABI StepOnePlus Real-Time PCR system, PCR products were quantified fluorometrically using the KAPA SYBR FAST master mix (KAPA Biosystems). The reference mRNA cyclophilin was used for normalizing the data (primers sequence available upon request). All samples were run in triplicates. Unpaired t test was performed using GraphPad Prism version 5.04 for Windows, GraphPad Software.

Results

Mapping of CBX2 target loci

To gain insights into the function of CBX2 and to identify its targets, we applied the DamID-sequencing
method that couples NGS to DamID assay. DamID is a methodology independent of antibodies, fixation, chromatin shearing, and other technically challenging features of chromatin immunoprecipitation, a technique commonly used to explore DNA/protein interactions. It has proven to be a powerful tool for the mapping of genomic binding sites of chromatin proteins (16). First, we tethered human CBX2 to *Escherichia coli* Dam and expressed the fusion protein in human testis-derived NT-2D1 cells. In eukaryotic cells that lack endogenous Dam proteins, the tethered Dam is capable of methylating adenines in GATC sequences that are in close spatial proximity (17). We confirmed by Western blotting and immunofluorescence that the exogenously introduced CBX2-Dam fusion protein was expressed and properly localized to the nucleus (Figure 1, A and B). Next, we transduced NT-2D1 cells with the lentivirus expressing Dam-CBX2, purified the gDNA from transduced cells, specifically amplified the methylated DNA fragments, and sequenced them. Cells expressing free Dam protein served as a control to

**Figure 3.** Identified peak regions of validated target genes. The coverage profile of the peak region normalized by the control sample is shown for selected validated target genes. The genomic coordinates are indicating the peak region. The distance (d) to the TSS of the target gene is provided.
correct for the unspecific DNA accessibility by Dam alone and for sequencing biases. We prepared 2 replicates for each fusion protein or free Dam and sequenced the DNA libraries separately. All pairs of replicates had highly correlated read densities along the genome (Pearson correlation coefficients \( > 0.90, P < 2 \times 10^{-16} \)) and thus were combined for the follow-up analyses. In total, we obtained 85 million and 17 million reads from NT-2D1 expressing Dam-CBX2 and free Dam, respectively. After peak-calling, we identified 1595 direct targets of CBX2 with a \( P < 1 \times 10^{-5} \) (Supplemental Table 1).

In order to elucidate the detailed distribution of CBX2 binding, we analyzed the identified putative target regions of CBX2. The distance to the closest transcription start site (TSS) showed that most of the called peaks is nearby a TSS (Figure 1C). A detailed view on the 5-kb up- and downstream region around the TSS showed depletion in the actual start site, whereas the largest peak is located in the promoter region of the putative target genes (Figure 1C). Analysis of the putative CBX2 binding sites (Figure 2A) confirms that most peaks are upstream to the closest TSS. The fraction of peaks found in the promoter region defined by 5 kb upstream is rather small (<3%) (Figure 2B). In contrast, many peaks (>65%) are located in intergenic regions (Figure 2B and 3).

**Validation of CBX2 targets**

To independently test the DamID-seq results, we selected a subset of CBX2 targets identified by DamID-seq for further investigation and validation. Target selection was based on their reported putative involvement in sexual differentiation, their biological role in human disease and animal models, or their specific expression in tissues involved in sex development (gonads, sex organs, hypothalamus, and pituitary) (Table 1). Applying these selection criteria, we shortlisted 19 genes. In basal (control) conditions, expression analysis showed that NT2-D1 cells endogenously expressed fifteen of these genes (SOX9, MAML1, EXO1, SOX3, TEX10, TBX2, PBX1, TSNAX, FGFR2, FGFR2, ATRX, WTAP, MTM1, TSPYL4, and FZD1), whereas 4 genes, namely KISS1, ZP4, PAXD1, and IGF1, were not significantly expressed in this particular cell line. (Figures 4 and 5 and data not shown). To study the influence of CBX2 on the expression of the selected genes, we measured their expression levels in NT2-D1 cells under overexpression and upon selectively knocking down endogenous CBX2 using specific siRNAs (5), respectively. The expression level of CBX2 upon its exogenous overexpression or knockdown was confirmed at the protein and mRNA levels by Western blot and qRT-PCR (quantitative Reverse Transcriptase-Polymerase Chain Reaction) analyses, respectively (Supplemental Figure 1).

As a prelude for validating the candidate target genes, we first checked the effect of CBX2 on genes known to be regulated by it such as SOX9, SF1, and SRY (5, 11). SOX9 was among the identified targets in our DamID-seq. Compared with control transfection (empty vector), overexpression of CBX2 enhanced expression of SOX9, SF1/NR5A1, and SRY by 4-, 3-, and 3-fold, respectively, whereas knocking down endogenous CBX2 resulted in down-regulating the expression of the same genes to 0.6-, 0.5-, and 0.6-fold, respectively (Supplemental Figure 2), consistent with previous observations (5).

Next, we examined the effect of CBX2 on the selected candidate genes. The wild-type (WT) CBX2 significantly enhanced the expression of endogenous EXO1, SOX3, MAML1, TBX2, TEX10, FGFR2, WTAP, TSPYL4, and MTM1 (by 2.9-, 2.6-, 2.6-, 2.3-, 2.2-, 2.4-, 1.6-, 2.2-, and 1.6-fold, respectively) but decreased the expression of PBX1 and FZD1 (to 0.5-, and 0.7-fold, respectively) (Figures 4 and 5). The expression of TSNAX, FGFR2, and ATRX seemed not to be affected by CBX2 (Supplemental Figure 3). To further explore the specificity of these observations, we knocked down endogenous CBX2 expression by siRNA oligos. Expression analysis showed that down-regulation of CBX2 significantly reduced the expression of EXO1, SOX3, MAML1, TBX2, TEX10, and TSPYL4 (to 0.7-, 0.6-, 0.6-, 0.6-, 0.7-, and 0.8-fold, respectively) while enhancing the expression of PBX1 and FZD1 (by 1.9-, and 1.4-fold, respectively), compared with control (scrambled siRNA), further supporting the CBX2-dependent expression of these genes. The lack of significant effects on the expression of MTM1, WTAP, and FGFR2 after down-regulating CBX2 is unclear, but it might be attributed to redundant pathways controlling the expression of these particular genes. Expression analysis for MTM1, WTAP, and FGFR2 were reexamined at different time points of CBX2 down-regulation and no significant change in their expression levels were observed (data not shown).

In whole-transcriptome studies in NT2-D1 cells, we observed that CBX2 negatively regulates FOXL2 expression (our unpublished data). The validation of this results showed a slight but significant reduction in the expression of FOXL2 in response to overexpressed CBX2 and, reciprocally, a significant up-regulation of its expression when CBX2 was knocked down (Figure 5).

To study the pathophysiological consequences of CBX2 defects, we took advantage of the mutations that we described as being the molecular basis of 46, XY DSD...
in human patients (5) and repeated the experiments by introducing the mutated forms of CBX2 harboring P98L and R443P mutations into NT2-D1. Immunofluorescence imaging showed that both mutated forms localize to the nucleus, and qRT-PCR analysis suggested that there are no significant differences in expression between WT and mutated CBX2 (Supplemental Figure 4). In contrast to WT CBX2, the mutated CBX2 had no effect on the expression of all the selected targets and behaved essentially as the control (Figures 4 and 5).

**Discussion**

Chromatin regulators are critical integration points wherein biological signals are converted into changes in gene expression. These protein machineries are essential to normal cell function, development, and differentiation (18). Critical for understanding the biology of these regulators is the determination of their binding sites to deduce their actions within the genome and eventually understand how the subsequent cellular responses occur. This is of particular importance in case of DSD, where...
The 2 right most columns indicate the behavior of the candidate target gene in response to CBX2 as up (↑) or down (↓)-regulated. Change in expression are represented as relative to control = 1. EV, empty vector; siScr, si scrambled; n.s., nonsignificant.

although the advance in understanding the genetic basis of human sexual determination and differentiation, in most subjects with DSD the underlying genetic cause is unknown. Although pathogenic mutations have been identified in DSD patients, the phenotype can be highly variable, even within families, suggesting that other factors, including genetic variants, are influencing the expression of the phenotype (3). Unexplained DSD cases may be due to mutations in novel gonad-determining genes, or genomic alterations affecting regulatory regions that lead to atypical gene expression. The identification of new genes involved in sex determination is therefore a goal of the utmost importance to the understanding of DSD and will allow for more accurate diagnosis and clinical management of these complex conditions (2).

To this goal, we chose to generate the first genome-wide binding data for human CBX2, a chromatin modulator that was shown to be involved in sex development in mice (11) and humans (12). To explore its precise position in this developmental cascade, we employed the DamID assay coupled with NGS (DamID-seq) that enabled us to identify 1595 direct targets for CBX2 in the human Sertoli-like cells NT-2D1. To further refine our identification of targets, and to place these genes in the context of the sexual determination pathway, we performed expression profiling experiments on a subset of genes in cells overexpressing CBX2 or in cells, in which endogenous CBX2 has been selectively knocked down.

Among the factors known to play a role in male sex development, we found SOX9 to be among our identified targets of CBX2. SOX9 is the first identified target of SRY, and its expression is essential for testis development (19). SOX9 expression has been linked to CBX2 in mouse studies (11), as well as in human cells (5), yet it remained unclear whether CBX2’s influence on SOX9 is direct or indirect. By identifying SOX9 sequences as targets for CBX2 binding, our finding establishes the first direct link between CBX2 and SOX9. Accordingly, the expression of SOX9 was increased by approximately 4-fold upon overexpressing CBX2, while decreased to 0.6-fold upon knocking down endogenous CBX2, which is agreement with previous data (5).

Numerous genes identified by our DamID-seq were linked to DSD in humans: MAMLD1, SOX3, FGFR2,
In addition, further analysis of the genomic landscape surrounding CBX2 revealed a number of additional gene candidates, such as TEX10, EXO1, TBX2, TSPYL4, WTAP, and MTM1, suggesting an additional role of CBX2 in sexual development. For example, TEX10 has been found to be a downstream binding target of SRY (20). TBX2 is involved in the transcriptional regulation of genes required for mesoderm differentiation, is overexpressed in reproductive system cancers, and has been suggested to play a role in the development of the mammalian reproductive system (21), whereas TSPYL4 was reported to be potentially linked to testicular dysgenesis (22). EXO1 is essential for male and female meiosis, with Exo1−/− mice being sterile (23). WTAP functions by impairing Wilms tumor protein-1 (WT1) DNA binding and inhibits expression of WT1 target genes (24). MTM1 deletion has been associated with myotubular myopathy and abnormal genital development (25).

We were unable to detect any significant changes in expression of FGF2, WTAP, and MTM1 after knockdown of CBX2, despite the fact that their expression levels were significantly increased upon overexpressing CBX2. According to these results, CBX2 may only have a secondary role in the regulation of these genes. However, it should be noted that the siRNA-mediated knockdown of CBX2, although substantial, may have been insufficient or not long enough to cause detectable alteration in gene expression.

Few of our identified genes (TSNAX, FGFR2, and ATRX) showed no significant change in expression levels in response to CBX2 overexpression or knockdown. For ATRX, a previous study reported a similar observation, where the expression of the murine Atrx was not significantly affected in M33/CBX2 knockout mice (11).
gests that additional factors, perhaps coactivators and chromatin state, are involved in their transcriptional regulation.

In the same line of reasoning, we observed that among the identified CBX2 targets are other transcription factors (TBX2 and ATRX). Accordingly, previous reports suggested that CBX2 regulates other genes, particularly SRY in an indirect fashion through the regulation of other genes encoding transcription factors (11). We and others (11) could not find a direct interaction between CBX2 and SRY that could explain the enhanced expression of SRY in response to CBX2. It is intriguing to speculate the association of any of the identified transcription factors in the CBX2-dependent SRY regulation. Supporting this idea, 2 recent studies in rats lacking the Y chromosome and the SRY gene showed no effect on the relative expression of all the target genes compared with control. The data in all graphs are the average of 3 independent experiments, error bars represent SD, and values are expressed as relative to control = 1.

****, P < .0001; ***, P < .001; **, P < .01; *, P < .05.

CBX2 in male sexual development as a factor independent from SRY, although additional work is still needed in this perspective.

We also identified CBX2 binding near a considerable number of genes encoding microRNAs, implying a possible still undiscovered role of microRNAs in human sex development, similarly to what has been shown for microRNAs and testicular proteome in mice (28), as well as in gonadal sex differentiation in chicken embryos (29). Further characterization of the impact of CBX2 activation on microRNAs and their potential involvement in sexual development in humans is essential.

Among the identified targets, there were also factors involved in female sex development, such as on FZD1 and PBX1. Fzd1 encodes a wingless-type MMTV integration site family (WNT) receptor whose expression is markedly induced in both mural granulosa cells and cumulus cells during the preovulatory period, in a manner similar to Wnt4 (30). WNT4 is a secreted protein essential for female sex development, in that it suppresses male sexual differentiation in mice and patients (31, 32). In humans, additional copies of WNT4 have been associated with male-to-female sex reversal and WNT4 loss of function causes DSD in women (33, 34). In a recent mouse study, Fzd1 has been showed to be required for normal female fertility, with Fzd1−/− females sterile or subfertile (30). Precellular leukemia transcription factor 1 (PBX1), which we found to be negatively regulated by CBX2, has been shown to play a role in female internal sex organ differentiation. In mice, the loss of Pbx1 resulted in absence of Müllerian ducts (35). In the same context, we observed that CBX2 negatively regulates FOXL2 expression. FOXL2 is considered to be the female counterpart of SOX9 (36). In this regard, we hypothesize that CBX2 functions, at least in part, by actively repressing ovarian development in XY gonads, which is in agreement with previous data showing that mutations in CBX2 causes male-to-female reversal with ovarian tissue in SRY-positive mice (11) and in humans (5).

Interestingly, when we repeated the experiments by introducing a mutated form of CBX2, we previously de-
scribed as the molecular basis of 46, XY DSD human patient (5), there was no effect on all the selected targets compared with the control, confirming the loss-of-function nature of the mutations and that the observed effects on gene expression are CBX2 specific.

Taken together, our findings represent the first comprehensive genome-wide study of CBX2 binding coupled with the expression profiling that further extends the role of CBX2 in human sexual differentiation, and also unravels the complexity and context dependence of CBX2-modulated responses. In fact, it appears that CBX2 on one hand acts in the testis by stimulating the expression of male-specific genes, SOX9, and simultaneously suppresses the female pathway by negatively regulating FOXL2 and the WNT4 pathway (Figure 6). The mechanisms underlying this bifunctional role are speculation but are most likely due to differences in the composition of regulatory complexes (37) and confirm that, as nicely put by Sekido and Lovell-Badge, “sex determination is a story of opposing forces and crucial alliances. [. . .]. It is a matter of timing (and expression level) that determines which pathway wins” (38). Moreover, our study shed more light on novel and potential pathological mechanisms that could be involved in DSD: further characterization of how these new targets fit into an expanding CBX2-regulated network should reveal how CBX2 activation and suppression can impact our understanding of DSD pathogenesis and ultimately DSD diagnosis and management. Finally, because all these effects were abolished when the mutated CBX2 was tested, the hypothesized role of CBX2 as both a stimulator of male and inhibitor of female pathway might help to explain the presence of ovarian tissue in the gonads of our CBX2-deficient 46, XY DSD patient (5) and shed further light on the possible mechanism of disease.

The future of clinical genetic analysis is doubtless NGS. Nevertheless, whole-exon sequencing identifies around 23%–25% of the genetic causes of diseases (39), and the percentage seems to be even lower in DSD (<10%; personal communication). Several reasons beside technical pitfalls might be accounted for these results. For example, the cause of the disease might be located outside the coding sequences, the disease is not monogenic, and too many candidate variants remained after filtering or part of the mechanism of disease can be ascribed to differences in gene regulation rather than gene structure. Although whole-exon sequencing remains a valuable tool in DSD, the study of regulatory networks is going to complement NGS in the understanding of these conditions. Experiments like ours leading to the discovery and understanding of pathways and networks will help “connecting the dots” in the interpretation of the NGS studies.

Acknowledgments

We thank Dr Bas van Steensel, Netherlands Cancer Institute, for providing the DamID constructs.

Address all correspondence and requests for reprints to: Professor Anna Biason-Lauber, Division of Endocrinology, Department of Medicine, University of Fribourg, Chemin du Musee 5, 1700 Fribourg, Switzerland. E-mail: anna.lauber@unifr.ch.

This work was supported by the Swiss National Science Foundation Grant 320030_130645 (to A.B.-L.).

Disclosure Summary: The authors have nothing to disclose.
References


