Streamlined identification of clinically and functionally relevant genetic regulators of lower-tract urogenital development

Meade Haller[a] (ID), Yan Yin[a] (ID), Gabe Haller[b], Tian Li[a], Qiufang Li[a], Laura E. Lamb[c], and Liang Ma[a,1]

Congenital anomalies of the lower genitourinary (LGU) tract are frequently comorbid due to genetically linked developmental pathways, and are among the most common yet most socially stigmatized congenital phenotypes. Genes involved in sexual differentiation are prime candidates for developmental anomalies of multiple LGU organs, but insufficient prospective screening tools have prevented the rapid identification of causative genes. Androgen signaling is among the most influential modulators of LGU development. The present study uses SpDamID technology in vivo to generate a comprehensive map of the pathways actively regulated by the androgen receptor (AR) in the genitalia in the presence of the p300 coactivator, identifying wingless/integrated (WNT) signaling as a highly enriched AR-regulated pathway in the genitalia. Transcription factor (TF) hits were then assayed for sexually dimorphic expression at two critical time points and also cross-referenced to a database of clinically relevant copy number variations to identify 252 TFs exhibiting copy variation in patients with LGU phenotypes. A subset of 54 TFs was identified for which LGU phenotypes are statistically overrepresented as a proportion of total observed phenotypes. The 252 TF hitlist was then subjected to a functional screen to identify hits whose silencing affects genital mesenchymal growth rates. Overlap of these datasets results in a refined list of 133 TFs of both functional and clinical relevance to LGU development, 31 of which are top priority candidates, including the well-documented renal progenitor regulator, Sall1. Loss of Sall1 was examined in vivo and confirmed to be a powerful regulator of LGU development.

androgen receptor | genitalia development | Sall1 | SpDamID

The reproductive and lower genitourinary (LGU) tracts are not only physically entwined but developmentally linked by master genetic regulators (1). Congenital structural anomalies of the reproductive and LGU systems comprise some of the most common (2) yet most socially stigmatized congenital phenotypes (3). Cryptorchidism, failure of testicular descent into the scrotum, affects an estimated 5.9% of newborn males (4). Hypospadias, proximal displacement of the penile urethral opening, affects approximately 1 in 240 American males (5). Congenital micropenis, inhibited penile distal outgrowth resulting in penile length 2.5 or more SDs below average, affects 0.015% of American newborn males (6). Posterior placement of the urethral valve affects an estimated 1 in 5,000 live newborns (7) and results in harmful retrograde flow of urine termed vesicoureteral reflux, which has a higher incidence of 1% of the general population due to varying etiologies (8). Less common LGU phenotypes include various anomalies of the paralogous labia and scrotum, gonadal dysgenesis, clitoral hypertrophy, ambiguous genitalia, and bladder (BL) exstrophy, among many others (1). During midgestation, the urogenital tract develops rapidly and concordantly in the posterior part of the embryo. Due to its rapid morphogenesis and the tight interconnection among the different urogenital structures, small perturbations in the development of one tissue can lead to developmental deregulation affecting the entire urogenital system. Thus, congenital LGU structural anomalies are quite often closely linked. For example, variations in external genitalia presentation such as hypospadias, chordee, and micropenis are often linked to cloacal phenotypes such as persistent cloaca and imperforated anus (1). Because of this close connection among different urogenital structures, their patterning and morphogenesis share many common signaling pathways, and perturbation of one signaling pathway or transcription factor (TF) often leads to comorbid developmental anomalies in multiple LGU structures or even multiple organ systems (9).

While a host of individual gene knockout mouse models has allowed greater understanding of the complex genetic network regulating LGU formation (1, 10–12), due to their high financial burden, justification for the generation of single-gene knockout mice strains relies on intensive preliminary findings and thus leaves many favorable hypotheses untested. Despite being valuable to the field, there has been a relative paucity of studies attempting to map the genetic network regulating LGU development as a whole,

Significance

The present study describes a tool, in the form of a mouse model, which can screen for likely regulators of embryonic sexual development. In this model, the androgen receptor adds a permanent mark to DNA near the genes that it is regulating. These marks accumulate and can be read with sequencing to identify nearby genes. We chose to use this model to first identify genes involved in genitalia masculinization. Hits classified as transcription factors, which are critical during development because they regulate the expression of other genes, were then analyzed by extensive validation at the levels of gene expression, functional relevance, and clinical relevance. Thirty-one top-tier candidates were identified, including Sall1, which was verified via knockout mouse.

Author affiliations: [a]Division of Dermatology, Department of Medicine, Washington University School of Medicine, St. Louis, MO 63110; [b]Department of Neurosurgery, Washington University School of Medicine, St. Louis, MO 63110; and [c]Department of Urology, William Beaumont School of Medicine, Oakland University, Rochester, MI 48309

particularly in vivo. Forward genetic screening on patients has greatly helped to identify clinically relevant genomic hotspots (12), and while extremely useful in pointing toward the most powerful upstream and most dosage-sensitive regulators, integral downstream mediators have remained largely out of reach due to the lack of prospective genetic screening models.

A recent development in the field of protein-to-DNA complex mapping is DamID and its cousin, SplitDamID (13). In DamID technology, the enzyme DNA adenine methyltransferase (Dam) is fused a DNA-interacting protein, often a TF. Whenever the fusion protein complexes with DNA, the enzyme methylates the DNA on adenine. Adenine methylation being otherwise undetectable or exceedingly rare in mammalian cells (14), and being a covalent cumulative modification, this creates a permanent map of that protein's DNA-interaction history that can be analyzed by sequencing. In SplitDamID (hereafter termed SpDam), the Dam enzyme is split into two parts, each fused to a different protein. Only when both proteins are complexed with DNA will the Dam activity be reconstituted such that methyl marks are placed (13). This variation on the technology makes it more physiologically relevant than both chromatin immunoprecipitation-sequencing (ChIP-seq) (which only captures one time point in fixed tissue) and DamID (which also captures more passive or stochastic binding events) since the energy requirement of two factors cocomplexed with DNA makes it more likely they are performing a relevant physiological function.

Among the most undeniably powerful known active regulators of LGU development is the androgen signaling pathway, which is required to generate the phenotypic presentation of the genetic male relative to the female. It is required for penile outgrowth, testicular formation and descent, scrotal formation, spermatogenesis, and positive feedback for testosterone levels (15–21). Androgen signaling has additional roles across variant tissues in order to generate secondary sex characteristics, both strictly anatomical as well as neurological/behavioral (22). Additionally, ectopic androgen signaling in females results in masculinization phenotypes such as clitoral hypertrophy (23).

The primary goal of this study is to generate a genomic map of urogenital developmental hotspots by utilizing SpDam technology in vivo, identifying major pathways actively regulated by androgen during LGU development, and refining the hit list to identify top candidate genes with high functional and clinical relevance. Due to its reliable history in the field of urogenital development and its clinically translational similarities to human genetics and development, the mouse model was chosen as the in vivo host for SpDam technology as the first step in this streamlined approach. In this model, when androgen receptor (AR) and p300 are complexed together with DNA, the methyltransferase activity of the Dam enzyme is reconstituted, and it proceeds to covalently modify adjacent DNA, creating a readable historical map of AR/p300/DNA-interaction in the given tissue. For the current study, we chose to first investigate the mouse genital tubercle (GT). Due to the delicate sensitivity of the genitalia to AR signaling, we anticipate that genomic hotspots identified by this system in the GT will overlap with those in other LGU tissues. If this is true, cross-referencing of clinical databases will indicate enrichment of various LGU phenotypes, including but not limited to penile phenotypes, in patients with variation at the hotspots identified herein.

Results

Generation of An In Vivo Model Utilizing SpDam Technology to Map the Genital AR Network.
Ideally, the murine model would be both amenable to tissue-specific and temporally specific

induction, with the understanding that with little to no native adenine methylation, leakiness in the system makes the signal simply broader but no less genuine. To generate the in vivo murine model, a bidirectional tetracycline-responsive promoter was used to control V5-tagged human AR fused to the primary fragment (D) of DNA adenine methyl transferase, and murine p300 fused to the secondary fragment (AM) of DNA adenine methyl transferase with human estrogen receptor (ERT2) fusion to enable tamoxifen-induced nuclear translocation (Fig. 1A). Because the methylation induced by Dam activity are covalent modifications, they are cumulative and permanent. We therefore wanted to use a genetic layout that offers multiple layers of specificity so that ideally all but at least the majority of the signal is coming from the tissue and time intended.

First, tissue specificity is controlled by the choice of rtTA murine line crossed into the lineage. Second, temporal induction of transgene expression is controlled by treatment of the animals with doxycycline. Third, translocation of the p300 component into the nucleus is controlled by treatment of the animals with tamoxifen. Fourth, the built-in split component of SpDam technology necessitates that both AR and p300 be simultaneously present at the DNA in order for methyltransferase activity to be reconstituted. A fifth layer of specificity is offered by physical dissection of the intended tissue. Linearized plasmid was used to generate a classic random insertion transgenic mouse line from a single F_0 founder. Genotyping from tail clips of F_1 pups showed successful genomic integration of the transgene biTetO::AR^D/p300$^{AM-ERT2}$ (Fig. 1B). Msx2::rtTA was chosen to drive the system based on its expression in formative regions and times during GT development (24) (Fig. 1C).

Dependency of expression on rtTA was tested by comparing levels of both fusion gene transcripts from E18.5 male GTs between specimens with and without biTetO::AR^D/p300$^{AM-ERT2}$, Msx2::rtTA, and doxycycline treatment. This experiment showed some modest expression in the absence of rtTA from both directions of the promoter, with an increase in both transcripts in Msx2::rtTA positive pups, and maximally induced transcription in the presence of doxycycline. This experiment indicates that there is some modest baseline leaking of both fusion gene transcripts from the bidirectional promoter in the absence of rtTA; thus, there will be some methylation in rtTA-negative cells such that careful dissection of the intended tissue is required. Additionally, rtTA is moderately activating the promoter in the absence of doxycycline, particularly the AR fusion gene. Treatment with doxycycline increases expression of both transcripts to their full level (Fig. 1D). Because AR and p300 must both be present on the DNA to reconstitute Dam activity, and because p300 nuclear translocation is controlled by tamoxifen (25), the expression leaking in the absence of doxycycline does not cause issue with temporal induction. To show that translocation of the p300$^{AM-ERT2}$ component is dependent on tamoxifen induction, male GTme cells were cotransfected with Actin::rtTA and biTetO::AR^D/p300$^{AM-ERT2}$ plasmids and let rest overnight, then they received 500 ng/mL doxycycline with or without 2 μg/mL tamoxifen for 48 h. Cells were fixed in paraformaldehyde and stained with anti-estrogen-receptor-alpha (ER-alpha) antibody to detect both endogenous ER-alpha and exogenous p300$^{AM-ERT2}$ via its affinity to ERT2. Cells were costained with Alexa 488–phalloidin. There was minimal signal detected from endogenous ER-alpha in male GTme cells. An increase in staining with transfection and treatment with doxycycline shows the ER-alpha antibody staining the ERT2 component of p300$^{AM-ERT2}$, and as expected this staining is largely excluded from the nuclei. Upon addition of tamoxifen, a clear shift of p300$^{AM-ERT2}$ into the nuclei is observed, indicating that the temporal induction of the system
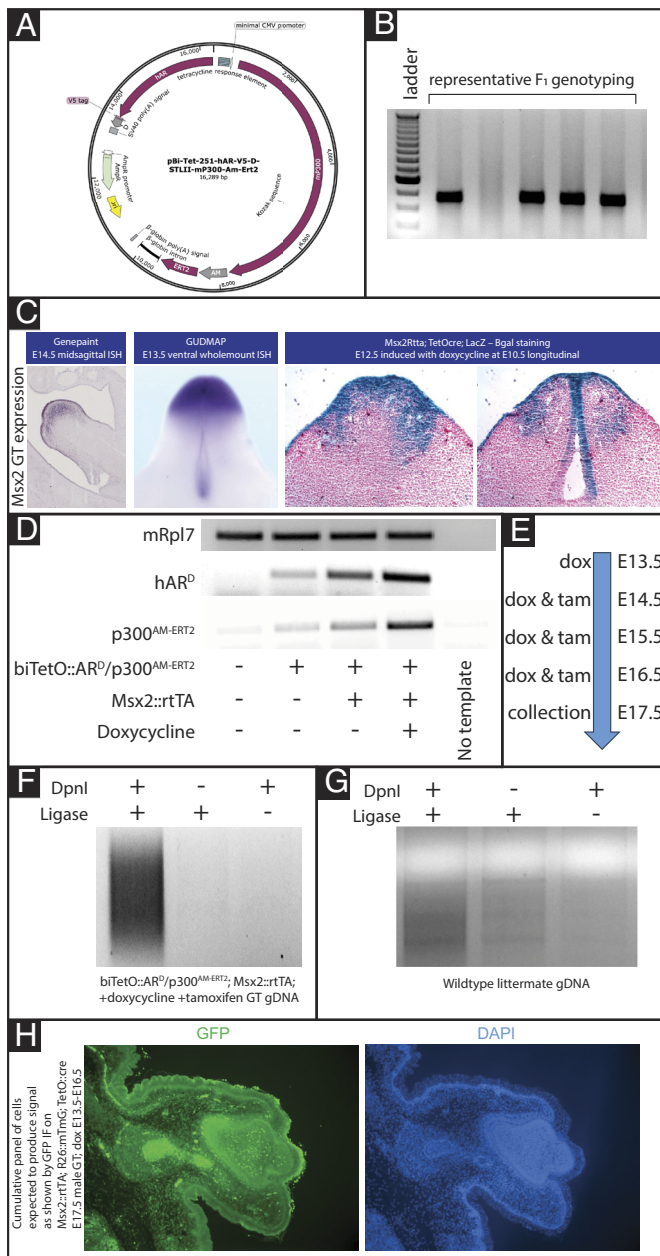
Fig. 1. Derivation and testing of the biTetO::AR^D/p300^AM-ERT2 transgenic mouse. Bidirectional tetracycline-responsive promoter controls V5-tagged hAR fused to the primary fragment of DNA adenine methyl transferase, and murine p300 fused to the secondary fragment of DNA adenine methyl transferase with tamoxifen-responsive translocation via ERT2 (*A*). Linearized plasmid was used to generate a classic random insertion transgenic mouse line from a single F_0 founding stud with genotyping from tail clips of F_1 pups (*B*). Expression pattern of *Msx2* transcript in the murine GT from genepaint (https://gp3.mpg.de/), the GenitoUrinary Development Molecular Anatomy Project (GUDMAP) (https://gudmap.org), and X-gal staining for reporter *LacZ* expression from Msx2::rtTA; TetO::cre; R26R embryos (*C*). Expression of the transgene from both directions of the bidirectional promoter was tested by comparing embryonic E18.5 male GT transcript levels in the presence and absence of biTetO::AR^D/p300^AM-ERT2, Msx2::rtTA, and doxycycline (*D*). Graphic showing the timeline of embryonic induction (*E*). Representative gel showing that smear of SpDam PCR library is detectable when generated from GTs of E17 embryos treated from E13.5 to E16.5 with doxycycline and E14.5 to E16.5 with tamoxifen to cause nuclear translocation of p300^AM and dependence of signal on both DpnI methylation–specific cleavage and ligation of adapters (*F*). Representative gel showing minor but detectable background signal in wild-type GTs prepared using the same methods (*G*). IF staining for GFP on male E17.5 Msx2::rtTA; R26::mTmG; TetO::cre embryos treated from E13.5 to E16.6 with doxycycline shows the panel of GT cells expected to contribute to the SpDam signal (*H*).

is indeed dependent on tamoxifen (*SI Appendix,* Fig. S1). As a critical control, we also show that administration of tamoxifen does not affect the expression of known AR target genes in male embryos (*SI Appendix,* Fig. S2). Together, these data indicate that background signal will be present in uninduced biTetO::AR^D/p300^AM-ERT2 animals from activation outside of the rtTA-guided tissue intended. This signal would be difficult to properly subtract due to its heavy overlap with the intended signal. For this reason, in the present initial study, we chose to use the tissue induction components of the model only as a means of amplifying the signal and to treat the model as a full-body transgenic with tissue isolation being the primary method of signal tissue specificity. The controls used are therefore littermate wild types and the signal is analyzed as a cumulative signal from carefully dissected GTs, with the understanding that the majority of the signal is from Msx2::rtTA expressing tissue during the time period of doxycycline and tamoxifen treatment.

To estimate the levels of the fusion transcripts relative to the endogenous transcripts, qRT-PCR was performed on whole E17.5 GTs using primers specific for endogenous transcript and comparing to primers that will amplify both endogenous and exogenous transcript on whole E17.5 GTs. The results indicate that there are approximately equal amounts of endogenous and exogenous transcripts present in whole GT extracts (*SI Appendix,* Fig. S3) making it unlikely that the fusion proteins are participating in binding events at supraphysiologic levels. SpDam PCR library is easily detectable when generated from GTs of E17 double-transgene-positive embryos treated from E13.5 to E16.5 with doxycycline and E14.5 to E16.5 with tamoxifen (Fig. 1*E*, induction timeline graphic), and this signal is dependent on both DpnI methylation–specific cleavage and ligation of adapters to the DpnI generated fragments (Fig. 1*F*). There is minor but detectable background signal in wild-type GTs prepared using the same methods (Fig. 1*G*), likely due to exceedingly low but nonetheless detectable (14) levels of endogenous murine adenine methylation, which according to a recent review (26) is considered sporadic and unlikely to interfere with peak analysis. Wild-type littermate DNA was therefore prepared as a negative control background signal for comparison. To show the panel of cells from which the bulk of the signal is emerging, Immunofluorescence (IF) staining for green fluorescent protein (GFP) was performed on Msx2::rtTA; TetO::cre; R26::mTmG embryos treated from E13.5 to E16.5 with doxycycline. As the embryos develop, the daily dosing of doxycycline from E13.5 to E16.5 results in a relatively ubiquitous/expanded expression pattern than is seen at any of the individual time points (Fig. 1*H*). Any cell expressing Msx2::rtTA at any of the dosed time points, as well as any of its daughter cells up until the point of collection, will contribute to the SpDam signal once tamoxifen is administered, giving us a wide capture of where AR is active across the tissue and time points in question. Potential methods that can be used to refine the signal to reflect smaller populations of cells (if desired) are described in *Discussion* section.

Sequencing of biTetO::AR^D/p300^AM-ERT2; Msx2::rtTA GT Libraries Reveals 3.5 K Genetic Hotspots. Sequencing of GT SpDam libraries from seven male biTetO::AR^D/p300^AM-ERT2; Msx2::rtTA and five wild-type controls (all doxycycline- and tamoxifen-treated) with wild-type signal averaged and subtracted as background shows the expected significant enrichments of regulatory regions and significant depletions of nonregulatory regions (Fig. 2*A*). Each experimental specimen was individually compared to the average of the controls, and an average of 14,000 peaks were identified per
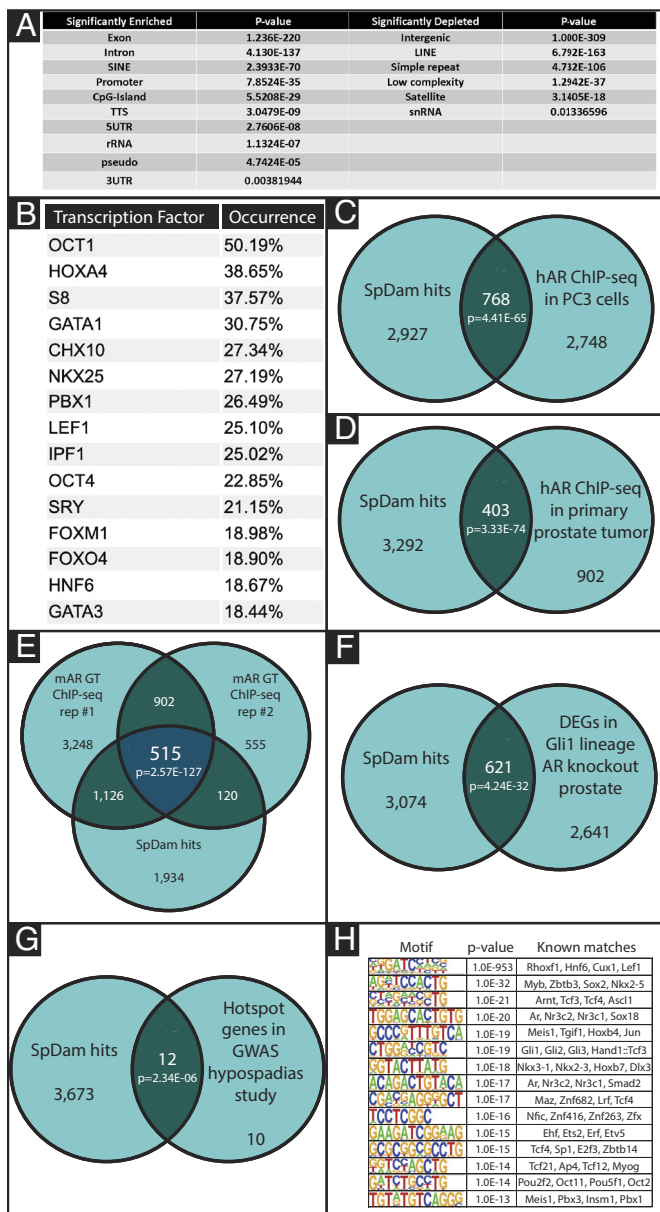
A

Significantly Enriched	P-value	Significantly Depleted	P-value
Exon	1.236E-220	Intergenic	1.000E-309
Intron	4.130E-137	LINE	6.792E-163
SINE	2.3933E-70	Simple repeat	4.712E-106
Promoter	7.8524E-35	Low complexity	1.2942E-37
CpG-Island	5.5208E-29	Satellite	3.1405E-18
TTS	3.0479E-09	snRNA	0.01336596
5UTR	2.7606E-08		
rRNA	1.1324E-07		
pseudo	4.7424E-05		
3UTR	0.00381944		

B

Transcription Factor	Occurrence
OCT1	50.19%
HOXA4	38.65%
S8	37.57%
GATA1	30.75%
CHX10	27.34%
NKX25	27.19%
PBX1	26.49%
LEF1	25.10%
IPF1	25.02%
OCT4	22.85%
SRY	21.15%
FOXM1	18.98%
FOXO4	18.90%
HNF6	18.67%
GATA3	18.44%

C

SpDam hits 2,927 — 768 (p=4.41E-65) — hAR ChIP-seq in PC3 cells 2,748

D

SpDam hits 3,292 — 403 (p=3.33E-74) — hAR ChIP-seq in primary prostate tumor 902

E

mAR GT ChIP-seq rep #1 3,248; 902; mAR GT ChIP-seq rep #2 555; 515 (p=2.57E-127); 1,126; 120; SpDam hits 1,934

F

SpDam hits 3,074 — 621 (p=4.24E-32) — DEGs in Gli1 lineage AR knockout prostate 2,641

G

SpDam hits 3,673 — 12 (p=2.34E-06) — Hotspot genes in GWAS hypospadias study 10

H

Motif	p-value	Known matches
	1.0E-953	Rhoxf1, Hnf6, Cux1, Lef1
	1.0E-32	Myb, Zbtb3, Sox2, Nkx2-5
	1.0E-21	Arnt, Tcf3, Tcf4, Ascl1
	1.0E-20	Ar, Nr3c2, Nr3c1, Sox18
	1.0E-19	Meis1, Tgif1, Hoxb4, Jun
	1.0E-19	Gli1, Gli2, Gli3, Hand1::Tcf3
	1.0E-18	Nkx3-1, Nkx2-3, Hoxb7, Dlx3
	1.0E-17	Ar, Nr3c2, Nr3c1, Smad2
	1.0E-17	Maz, Znf682, Lrf, Tcf4
	1.0E-16	Nfic, Znf416, Znf263, Zfx
	1.0E-15	Ehf, Ets2, Erf, Etv5
	1.0E-15	Tcf4, Sp1, E2f3, Zbtb14
	1.0E-14	Tcf21, Ap4, Tcf12, Myog
	1.0E-14	Pou2f2, Oct11, Pou5f1, Oct2
	1.0E-13	Meis1, Pbx3, Insm1, Pbx1

Fig. 2. Sequencing of biTetO::AR^D/p300$^{AM-ERT2}$; Msx2::rtTA GT libraries. Sequencing of GT SpDam libraries with wild-type signal treated as background shows the expected significant enrichments of regulatory regions and significant depletions of nonregulatory regions (A). Analysis of all hits for common distant regulatory elements (B). There is significant overlap with genes bound by AR in human AR-PC3 cells (28) (C), with genes bound by AR in human primary prostate tumors (29) (D), with genes bound by AR in the murine genitalia (30) (E), with DEGs in Gli1-lineage AR knockout prostates (31) (F), and with hotspots for hypospadias in a GWAS of 1,006 surgery-confirmed patients (32) (G). The most enriched consensus sites found in the reads generated from sequencing and the TFs that bind them (H).

specimen (Datasets S1–S7, corresponding to experimental samples 1 through 7). Nearest gene data was extracted for each peak, and if at least five of the seven experimental specimens shared the same nearest gene, this gene was considered a hit. Using these criteria, 3,695 gene hits were identified (Dataset S8). Analysis of all hits for common distant regulatory elements (27) (permalink—https://dire.dcode.org/index.php?id=0913210358552510) revealed a list of potential cofactors of AR during GT differentiation that includes OCT1, LEF1, and SRY (Fig. 2B). In further efforts to determine the reliability of these findings, the hit list was compared to several datasets with which it is expected to overlap. There is statistically significant overlap with the nearest genes to hAR

peaks in human AR-PC3 cells as identified by ChIP-seq (28) (Fig. 2C) and even more significant overlap with the nearest genes to hAR ChIP-seq peaks in human primary prostate tumors (29) (Fig. 2D). Most importantly, dramatic overlap is present between the nearest genes to mAR ChIP-seq peaks in the mouse GT (30) (Fig. 2E) (Dataset S9). To determine whether hit genes tend to be differentially expressed in tissues where AR is missing, we overlapped our dataset to differentially expressed genes (DEG) in Gli1-lineage AR knockout prostate cells and found significant overlap (31) (Fig. 2F) (Dataset S9). Additionally, 55% of the genes (12 of 22) identified in a genome-wide association study (GWAS) of 1,006 surgery-confirmed hypospadias patients were also SpDam hits in our study, indicating a high likelihood of translational relevance (32) (Fig. 2G) (Dataset S9). The top-most enriched de novo sequences point toward additional likely comembers of AR/p300 regulatory complexes (Fig. 2H) and include multiple members of the Wnt signaling network.

Protein ANalysis THrough Evolutionary Relationships (PANTHER) ontology analysis (33) (https://pantherdb.org) revealed that the molecular functions of hits are most enriched for binding, catalytic activity, transcription regulator activity, and signaling receptor activity (Fig. 3A). The binding subset is most enriched for organic cyclic compound binding, heterocyclic compound binding, protein binding, and ion binding (Fig. 3B). The catalytic activity subset is most enriched for hydrolases, transferases, and enzymes acting on proteins (Fig. 3C). The transcription regulator subset is heavily enriched for DNA-binding TFs (Fig. 3D). The subset of transporter activity is most enriched for genes with transmembrane transporter activity (Fig. 3E) and the subset of molecular function regulators is most enriched for genes with enzyme regulator activity (Fig. 3F) respectively. Unsurprisingly, developmental and anatomical structural processes are among the most statistically enriched biological processes (Fig. 3G) while the gonadotropin hormone-releasing hormone (GnRH), Wnt, and Cadherin signaling pathways are among the most statistically overrepresented signaling pathways (Fig. 3H).

TFs Are Overrepresented in the Dataset and Overwhelmingly Transcriptionally Sensitive to the Presence or Absence of Androgen and WNT Signaling in the Developing Murine Genitalia. A combination of the PANTHER database (https://www.pantherdb.org/) and the Riken TF database (http://gerg.gsc.riken.jp/TFdb/) was used to identify known TFs in the hitlist (N = 298 or 8% of hits) (Dataset S10) which is a dramatic statistical enrichment over the expected random representation of TFs in the genome (approximately 1,600 out of 30,000 genes or 5% of the genome) by $P = 0.0001$ by the chi square test. The TF hits were then analyzed using the search tool for recurring instances of neighbouring genes (STRING) network (34) (https://string-db.org) to identify the overrepresented reactome pathways (Fig. 4A). Among the enriched reactome pathways are WNT-mediated, FOXO-mediated, runt-related transcription factor (RUNX) mediated, SMAD-mediated, and TFAP2-mediated transcription, as well as regulation of transcription by SUMOylation of TFs and intracellular receptors. To estimate the proportion of TF hits that are transcriptionally sensitive to the presence of AR, qRT-PCR was performed on all TF hits in technical triplicate from N ≥ 6 whole GTs of E14.5 and E17.5 males and females to reveal that 69% exhibit sexually dimorphic expression in the murine GT at one or both time points (Fig. 4B) (Dataset S11). While E14.5 may seem too early for dimorphic expression secondary to circulating androgens, previous studies indicate there are already vast differences in the GT transcriptome between the sexes at E14.5, likely from a combination of androgen-independent and
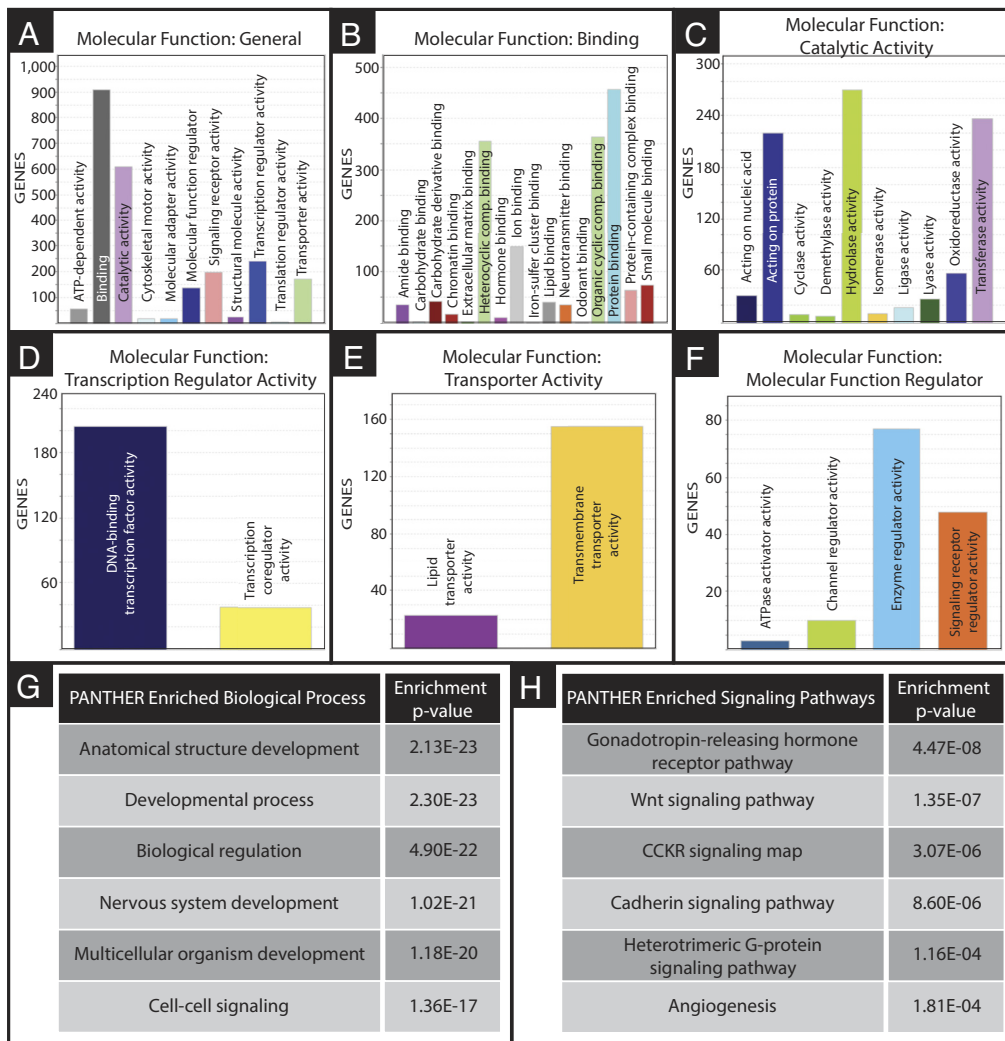
Fig. 3. Ontology analysis of biTetO::AR^D/p300^AM-ERT2; Msx2::rtTA GT hits. PANTHER ontology analysis revealed that the molecular functions of hits are most enriched for binding, catalytic activity, and molecular function regulators (A). The binding subset is most enriched for heterocyclic compound binding, organic cyclic compound binding, and protein binding (B). The catalytic activity subset is most enriched for hydrolases, transferases, and enzymes acting on proteins (C). The transcription regulator subset is most enriched for DNA-binding TFs (D). The transporter subset is most enriched for transmembrane transporters (E). The molecular function regulator subset is most enriched for enzyme and signaling receptor regulators (F). Developmental and anatomical structural processes are among the most statistically enriched biological processes (G) while the GnRH and WNT pathways are the most statistically enriched signaling pathways (H).

low-level ligand-dependent AR signaling (35). Additionally, because WNT signaling was revealed to be one of the top hit pathways, we performed qRT-PCR on a random subset of 148 TF hits in technical triplicate from N ≥ 6 whole GTs of E17.5 β-catenin fl/fl; Msx2::rtTA; TetO::cre and cre-negative littermate control males and females (all embryos treated with doxycycline by single-pulse gavage of dam at E12.5) to reveal that 66% are transcriptionally sensitive to deletion of β-catenin in Msx2-expressing cells in one or both sexes (Fig. 4C) (Dataset S11). In fact, when the full hitlist was cross-referenced to the currently known genetic network of 179 β-catenin interacting proteins (https://hagrid.dbmi.pitt.edu/schizo-pi/index.php/gene/view/1499), there is statistically significant overlap (59 genes, or 33% of the list, *P* = 4.17E–13) (Dataset S12).

DECIPHER Database Mining Reveals 54 TFs for Which Copy Variation Is Disproportionately Associated with LGU Phenotypes. TF genes cobound by AR and p300 in the genitalia are likely also involved in general urogenital development due to the overarching developmental role of AR. To validate and estimate the clinical relevance of our hit list, the comprehensive list of TF hits was cross-referenced with the Database of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources (DECIPHER) database (36) of clinically relevant copy number variations (CNVs) (https://decipher.sanger.ac.uk) to look for associations of hits with LGU phenotypes. For every gene, each specific LGU phenotype in patients with CNVs covering that gene was documented, summed, and calculated as a proportion of

total phenotype observations for all patients with CNV covering the given gene. This observed rate of LGU phenotypes per gene was then compared to the expected rate of LGU phenotypes (the baseline rate in the general DECIPHER cohort) by the Fisher exact test. For 200 out of 292 TF hits with CNV data (68%), at least one LGU phenotype was present in at least one patient with CNV covering the gene. For 54 out of the 292 TF hits with CNV data (18%), LGU phenotypes were found to be present at a statistically higher than expected rate in patients with CNVs covering these genes (Fig. 5A) (Dataset S11). The remaining 38 TF hits had CNV patients recorded in DECIPHER, but none of these patients exhibited LGU phenotypes (13%). It is important to note that the general DECIPHER cohort utilized as background phenotype rate for this analysis is already heavily enriched for patients with clinically relevant CNVs, and therefore this is a likely underestimation of the enrichment that would be observed if compared to the general population. If the genes that are not differentially expressed between the sexes during the two tested time points of murine GT development are excluded from the list, 140 TFs (70%) have at least one CNV patient with an LGU phenotype, 40 TFs (20%) show LGU phenotypes at statistically higher than expected rates, and 21 TFs (10%) have CNV patients in DECIPHER but do not exhibit LGU phenotypes (Fig. 5B) (Dataset S11). The list of DECIPHER coded phenotypes that were considered to be LGU phenotypes and their relative prevalence per patient, as well as their percentage of total phenotype observations can be found in Dataset S13.
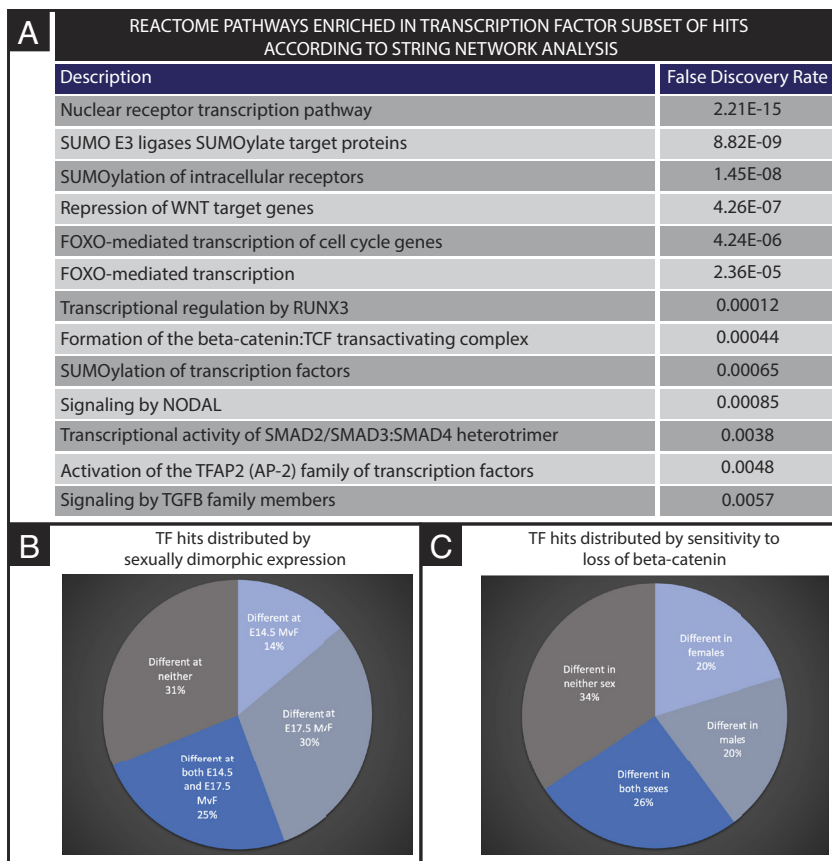
A

REACTOME PATHWAYS ENRICHED IN TRANSCRIPTION FACTOR SUBSET OF HITS ACCORDING TO STRING NETWORK ANALYSIS	
Description	False Discovery Rate
Nuclear receptor transcription pathway	2.21E-15
SUMO E3 ligases SUMOylate target proteins	8.82E-09
SUMOylation of intracellular receptors	1.45E-08
Repression of WNT target genes	4.26E-07
FOXO-mediated transcription of cell cycle genes	4.24E-06
FOXO-mediated transcription	2.36E-05
Transcriptional regulation by RUNX3	0.00012
Formation of the beta-catenin:TCF transactivating complex	0.00044
SUMOylation of transcription factors	0.00065
Signaling by NODAL	0.00085
Transcriptional activity of SMAD2/SMAD3:SMAD4 heterotrimer	0.0038
Activation of the TFAP2 (AP-2) family of transcription factors	0.0048
Signaling by TGFB family members	0.0057

B TF hits distributed by sexually dimorphic expression

C TF hits distributed by sensitivity to loss of beta-catenin

Fig. 4. Analysis of TF hits. A combination of the PANTHER database and the Riken TF database was used to identify all TF hits, which were then analyzed using the STRING network (https://string-db.org) to identify overrepresented reactome pathways (*A*). qRT-PCR was performed on all TF hits in technical triplicate from N ≥ 6 whole GTs of E14.5 and E17.5 males and females to reveal that 69% exhibit sexually dimorphic expression in the GT at one or both time points (*B*). Because of WNT signaling being a major pathway identified, qRT-PCR was also performed on a random subset of TF hits in technical triplicate from N ≥ 6 whole GTs of E17.5 β-catenin fl/fl; Msx2::rtTA; TetO::cre and cre-negative control males and females (all embryos treated with doxycycline at E12.5) to reveal that 66% are transcriptionally sensitive to deletion of β-catenin in Msx2-expressing cells in one or both sexes (*C*).

Functional siRNA Screening in Isolated Mesenchymal GT Cells Further Validates a Subset of TF Hits As Regulators of LGU Development. We recently developed a high-throughput functional screen methodology to identify growth and differentiation modulators of isolated murine genital mesenchymal (GTme) cells (37). In this assay, Vimentin-expressing cells turn on the GFP reporter of the stoplight (mTmG) allele (38), while cells not expressing Vimentin remain red fluorescent. Briefly, GFP-positive GTme cells are isolated using fluorescence-assisted cell sorting (FACS) from doxycycline-treated triple-transgenic Vim::rtTA; TetO::cre; mTmG embryos, treated with putative growth or differentiation modulators, and imaged after several days in culture to calculate the effect of each treatment. To functionally validate a subset of our hitlist, the TFs for which at least one LGU phenotypic CNV patient was identified in the DECIPHER database were subjected to GTme growth screening (regardless of their result in the expression screening) using small interfering RNA (siRNA) knockdown. Of the 252 TFs functionally tested, 196 (77.7%) showed differential GTme growth rates in response to siRNA knockdown in one or both sexes (Fig. 5*C*) with a heavy bias toward impairment of proliferation (Fig. 5*D*) (Dataset S14) (Table 1, representative subset), and 133 of these are also differentially expressed between the sexes at the tested time points (Dataset S15). In order to identify genes ideal for further investigation, we defined a top-tier candidate SpDam hit as a gene meeting all four of the following requirements: 1) the gene encodes a TF; 2) the transcript exhibits sexually dimorphic GT expression in at least one of the two tested time points; 3) the genetic locus has a DECIPHER profile of LGU phenotypes statistically overrepresented as a fraction of total phenotypes in patients with CNVs covering the locus; and 4) GTme cells show proliferative differences when the transcript is silenced. Thirty-one hits fit these criteria (Fig. 5*E*) (Dataset S16), including the well-documented renal progenitor regulator, Sall1. Visualization of peak overlap for several

of the top-tier hits as compared to AR ChIP-seq in murine GT is shown in *SI Appendix*, Fig. S4.

TF Hit Sall1 Is Required for Normal GT Development. *Sall1* exhibits an interesting transcriptional profile in the GT. It is highly expressed in early GT development at E14.5 with an astonishingly low cycle threshold (CT) value averaging 19.8 (relative to internal control *Rpl7* averaging 17) and it is then 80% down-regulated by E17.5. This energy expenditure followed by energy reserve suggests a specifically early purpose for Sall1 in GT development. As expected based on qRT-PCR, RNAscope in situ hybridization (ISH) on midsagittal wild-type E14.5 embryos shows that *Sall1* is strongly expressed in the developing male GT (Fig. 6*A*, blue arrow) in particular all along the ventral aspect enclosing the urethra, the entire distal region, as well as the distal-most portion of the dorsal aspect. (Fig. 6*B*, zoom). It exhibits much lower expression in most other embryonic tissues, with the notable exception of the kidney for which it has well-documented roles (39), the proximal-dorsal aspect of the tail, and certain areas of the brain. RNAscope ISH was performed for *Sall1* levels in β-catenin conditional knockout male GTs at E14.5, when *Sall1* seems to have its most impactful role (Fig. 6*C*). This experiment shows downregulation of *Sall1* at E14.5 in response to the absence of β-catenin in Msx2::rtTA expressing cells. Together our findings put *Sall1* downstream of both AR and of β-catenin.

With access to the conditional allele of *Sall1* through generous colleagues, we endeavored to confirm that *Sall1* is required for normal GT development. Conditional deletion of *Sall1* using the Msx2::rtTA driver with doxycycline induction by daily gavage from E9.5 to E16.5 (*Sall1* cKO) and collection at E17.5 shows dramatically underdeveloped GT with complete failure of urethral tubularization in transverse sections of males (Fig. 6*D*, red arrows showing tubular urethra in controls and ventrally exposed urethral plate in cKO) (additional images *SI Appendix*, Fig. S5). Furthermore, the
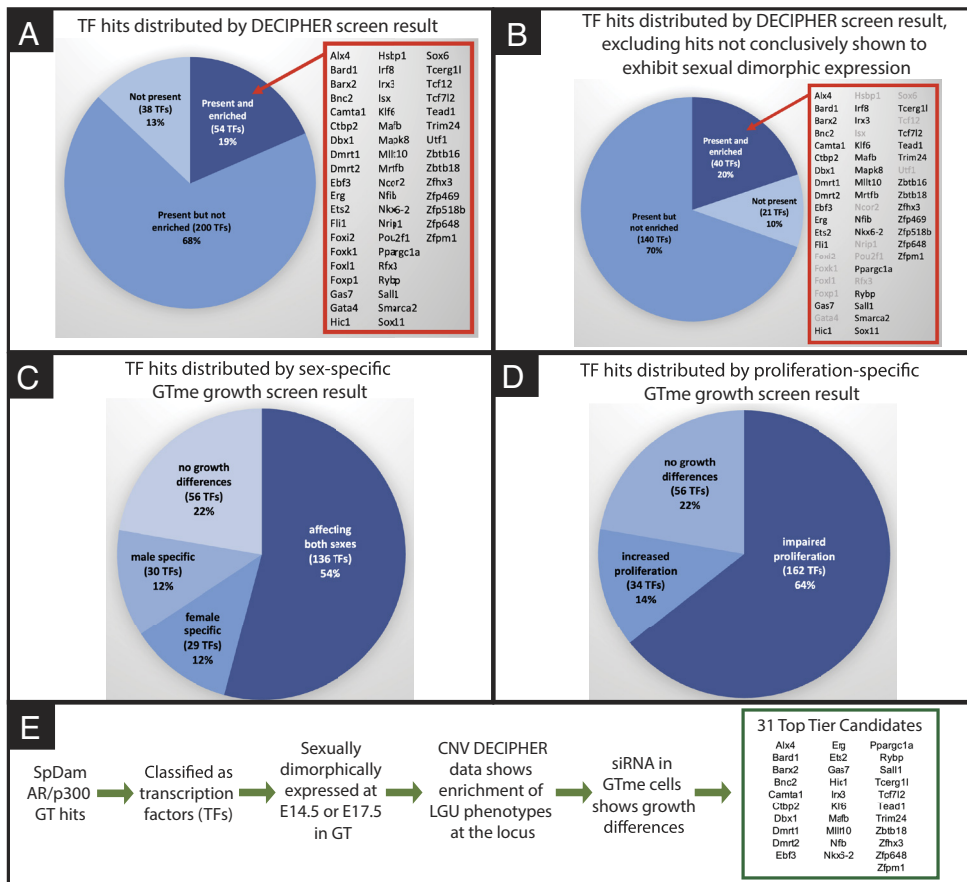
Fig. 5. Clinical and functional relevance of findings. Proportion of TF hits for which CNV patients in DECIPHER exhibit at least one LGU phenotype (present but not enriched), a statistically significant enrichment of LGU phenotypes as a percentage of all phenotype observations (present and enriched), or do not exhibit LGU phenotypes (not present) (*A*), gene names of enriched subset listed in the red box. The same measures if TFs that do not exhibit sexually dimorphic GT expression in the mouse are excluded (*B*), excluded genes from enriched subset grayed out. Functional siRNA GTme screening result of all TF hits with at least one LGU patient listed in the DECIPHER database, distributions based on sex specificity (*C*) and effects on proliferation (*D*). Flowchart of streamlined approach used to systematically identify top-tier candidate genes of sexual differentiation (*E*).

internal urethra, which can be shown connecting clearly to the BL in midsagittal sections (Fig. 6*E*), fails to diverge from the rectum (Fig. 6*F*) with the two converged prior to exiting the body as a single external opening (red arrow) in an apparent congenital distal recto-urethral fistula (additional images *SI Appendix*, Fig. S6). The range of urethral phenotypes as seen in whole-mount fixed specimens was relatively wide, with some male cKO showing a mild distal displacement and others showing a fully open ventral cleft (Fig. 6*G*). IF proliferation staining and quantification of E16.5 *Sall1* cKO (induced daily from E9.5) against phosphorylated histone H3 shows that loss of *Sall1* in Msx2-expressing cells diminishes the proliferative capacity of genital mesenchymal tissues in both sexes (Fig. 6*H*).

One of the documented ways by which the GT mesenchyme can exhibit decreased proliferative capacity is by aberrant upregulation of Bmp4. Variation in the human BMP4 gene is associated with hypospadias (40). In murine models, addition of exogenous Bmp4 results in downregulation of Fgf8 and Wnt5a with concomitant reduced proliferation. Additionally, if the Bmp4 receptor, *Bmpr1a*, is conditionally deleted, an opposite phenotype of hyperplasia is observed (41). Similar studies in birds have indicated that the same signaling pathways are at play (42). Bmp4 is therefore integral to the mesenchymal proliferative capacity in the GT during the outgrowth phase of development and its dose must be balanced via genetic regulation to achieve appropriate levels of proliferation. Sall1 is considered both an activator and a repressor, depending on context. When *Bmp4* levels were assayed by both ISH and qRT-PCR in E16.5 *Sall1* cKO (induced daily from E9.5), it was evident that *Bmp4* is aberrantly highly up-regulated in male mutants compared to controls as much as 17-fold (Fig. 6*I*), suggesting that a key function of *Sall1* in the development of the GT is to titrate the expression of *Bmp4* via fine-tuned repression. Additionally, *Wnt5a* is down-regulated in male *Sall1* cKO (Fig. 6*I*). This investigation of

Sall1 places it firmly within the existing known network of GT regulators (Fig. 6*J*, proposed mechanism) and represents a proof of principle for our streamlined approach to identify candidate LGU regulators, arguing in favor of the investigation of additional hits.

Discussion

TFs are undeniably at the forefront of developmental regulation, and our understanding of their roles for many years naively portrayed them as independent rather than cooperative, complex forming, and interdependent. SpDam technology is a user-friendly, antibody-independent system that allows prospective mapping of the DNA-interaction sites of protein duos, helping to filter out nonproductive binding events and better delineate genetic networks downstream of master regulators. The present study utilizes SpDam technology in vivo in order to map the network of genes downstream of active AR by necessitating it be cobound to p300 to produce signal. We investigated the developing urogenital system using the mouse GT as a representative organ. As a proof of concept, the present study shows that SpDam is amenable for use as an in vivo transgenic system with dramatic clinically translational potential, but that follow-up validation is critical due to the noise produced from modest endogenous Dam activity in the mouse and issues with baseline expression leakage. SpDam mouse models with any desired TF pair can be generated relatively easily using homologous recombination-based CRISPR technology, and they can be driven by any desired rtTA, making it quite versatile. The specificity of the SpDam transgene expression can be further refined in future models, and careful isolation of the target tissue can always be used if induction is less than perfectly specific. If paired with a reporter such as mTmG (and requisite TetO::cre), one can envision an experiment where cells expressing the rtTA after doxycycline treatment

Table 1. Sexual dimorphic expression and GTme screening for representative TFs

GENE	Male vs. female at E14.5		Male vs. female at E17.5		Male cell count normalized to scramble = 100		Female cell count normalized to scramble = 100	
	Fold change	*P*-value	Fold change	*P*-value	Average	*P*-value	Average	*P*-value
Alx4	0.87009	**0.01546**	1.45502	**0.00005**	40.72	**5.25E−11**	50.63	**3.52E−08**
Dbx1	1.44493	**0.00479**	1.07478	0.45385	64.75	**4.80E−02**	79.02	7.30E−02
Dmrt1	0.87736	0.41236	0.48761	**0.01402**	25.16	**7.81E−12**	18.69	**1.13E−13**
Dmrt2	1.23524	**0.01039**	1.56204	**0.00205**	55.72	**1.30E−07**	72.24	**5.50E−04**
Esr1	0.88266	**0.00324**	0.35234	**0.00012**	21.64	**3.00E−14**	29.01	**4.74E−12**
Ets2	0.96077	0.56886	0.77859	**0.02946**	80.25	**3.90E−03**	102.90	6.95E−01
Gli3	0.99074	0.70116	1.30386	**0.00030**	12.05	**3.00E−03**	99.47	9.00E−01
Irf8	0.91159	0.14491	0.82675	**0.03433**	98.61	8.40E−01	94.26	3.44E−01
Irx3	0.86306	**0.00761**	0.73423	**0.00100**	56.38	**6.39E−08**	94.05	3.93E−01
Klf6	1.01709	0.75623	1.25655	**0.00015**	66.32	**4.10E−02**	52.67	**1.60E−02**
Lef1	0.94902	0.59324	0.98741	0.90009	31.77	**4.21E−10**	49.06	**3.91E−08**
Mafb	1.30086	**0.01169**	0.75291	**0.00038**	22.13	**6.06E−14**	64.02	**6.09E−04**
Nkx6-2	0.95625	0.76405	1.36951	**0.04777**	15.58	**4.25E−15**	45.57	**3.70E−09**
Sall1	0.90013	**0.04589**	0.87708	0.18729	62.64	**1.76E−04**	90.91	1.41E−01
Sox11	1.11247	**0.03430**	1.29371	**0.00063**	95.25	1.04E−01	86.77	5.80E−02
Sp8	1.18155	**0.04515**	0.77524	**0.00060**	93.33	6.36E−01	89.46	9.77E−01
Tcf7l2	0.94831	0.08434	1.70985	**0.00004**	88.95	1.37E−01	81.99	**2.33E−02**
Tead1	0.79930	**0.04740**	1.18808	0.09613	29.84	**6.81E−13**	49.78	**2.24E−08**

Significant *P* values are in bold.

are also expressing GFP and can be isolated by FACS, better refining the signal and cleanly putting tissue specificity under control of doxycycline while temporal specificity remains under the control of tamoxifen. Such quadruple transgenic animals would be time consuming to produce, but ideal for the cleanest signal. Additionally, if the goal is to get a smaller snapshot rather than a collective bulk signal across a developmental range, then the time between drug treatment and collection can be shortened such that fewer daughter cells are included in the signal.

The present study identifies an average of ~14,000 peaks per experimental specimen and 3,695 nearest genes in common in five of seven specimens, which is an average of 3.8 peaks per nearest gene. The reads were enriched for regulatory elements and depleted for nonregulatory elements, confirming that the distribution of reads is far from random. Potential LEF1 binding sites are highly enriched both in de novo reads as well as in the distant regulatory elements of hit genes, and the Wnt signaling pathway is the second most enriched signaling pathway upon ontology analysis of hits. Potential AR consensus sites are enriched in the de novo reads as expected, and there is dramatic overlap between our hitlist and nearest gene hits of multiple AR ChIP-seq datasets, including one specific to murine GT (28–30). The SpDam assay can be likened to ChIP-seq in the sense that not all ChIP-seq peaks for a given TF will have associated previously characterized consensus sequences for that TF because the TF need only be a member of the complex, not necessarily a DNA-binding member (43); this does not mean the peak is any less real or less important. We suspect that there are multiple regulatory complexes involving both AR and p300 and that our generalized hit list comprises a comprehensive view of the binding events of these complexes during GT development. Ontology analysis of our hits revealed enrichment for cyclic compound binding proteins, hydrolases, transferases, TFs, extracellular matrix components, and transmembrane transporters, among other functional groups. Predictably the most enriched biological processes were developmental and morphogenesis pathways, and the most enriched signaling pathways were

WNT, GnRH, and Cadherin signaling among others. β-catenin complexes with cadherins (44), so in a sense this is also pointing directly to the WNT pathway.

The analysis of distal regulatory element consensus sites, and consensus sites present in the reads, points toward likely comembers of these complexes. HOXA4 and NKX2-5 binding sites were present at even higher rates to LEF1 binding sites in distal regulatory regions of hit genes, while RHOXF1, MYB, MEIS1, GLI3, and Myc-associated zinc-finger protein (MAZ) were among the most enriched consensus sites present in the sequencing reads. Mutations in HOXA4 are associated with human hypospadias (40). NKX2-5 is pivotal in heart formation and a known modulator of WNT signaling (45, 46), but its potential roles in urogenital development have not been studied. The X-linked reproductive homeobox (RHOX) gene cluster is involved in gametogenesis in both sexes (47). MYB overexpression can override cell cycle arrest caused by androgen depletion in prostate cancer cells (48). MEIS1 is linked to AR activity in prostate cancer (49). GLI3 is a well-documented integrator of androgen signaling and hedgehog signaling during masculinization (50, 51). MAZ is a dosage-sensitive regulator of urogenital development within the commonly copy-variant 16p11.2 genomic locus, and its depletion results in upper (renal) and lower (BL) urogenital phenotypes, deregulation of WNT signaling mediators, and deregulation of cell cycle progression (11, 52).

Some of TFs that bind the enriched distal regulatory element consensus sites also happen to be TF hits in our screen, for example, LEF1, MEIS1, and GLI3—this may be an indication of autoregulation of the various complexes containing AR and p300. *Lef1* transcript did not show differential expression between the sexes at the tested time points, but did, as expected, show reduced expression in response to loss of β-catenin in Msx2-expressing cells in both sexes. Knockdown of *Lef1* in GTme cells also dramatically decreased the growth potential of the cells in both sexes, by 70% in males and 50% in females. *Meis1* transcript does not show differential expression between the sexes at E14.5 but has 25% higher
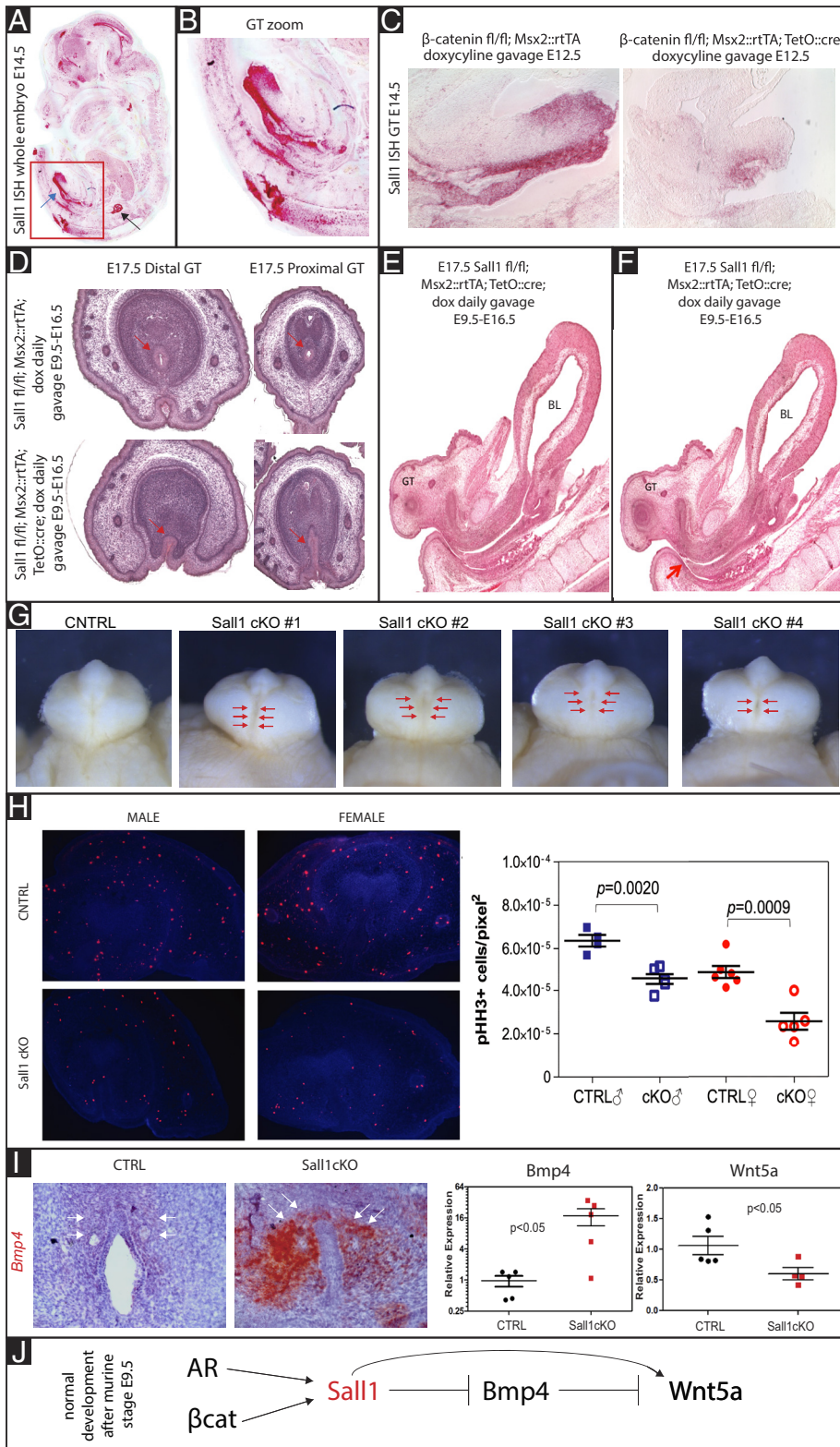
Fig. 6. TF hit *Sall1* is required for normal GT development. RNAscope ISH on midsagittal wild-type male E14.5 embryos shows that *Sall1* transcript is heavily expressed in the developing GT (*A*, blue arrow) (*B*, zoom) as well as the kidney (*A*, black arrow) and dorsal tail relative to other embryonic tissues. While not dramatically transcriptionally sensitive to *β-catenin* conditional deletion (doxycycline at E12.5 and collection at E17.5) by qRT-PCR, *Sall1* transcript levels are clearly dependent on β-catenin levels in the male E14.5 GT as shown by midsagittal RNAscope ISH (doxycycline at E12.5 and collection at E14.5) (*C*). Conditional deletion of *Sall1* in males using the Msx2::rtTA driver with doxycycline induction (*Sall1* cKO) by daily gavage from E9.5 to E16.5 and collection at E17.5 shows dramatically underdeveloped GT with complete failure of urethral tubularization in transverse sections (*D*, red arrows showing tubular urethra in controls and ventrally exposed urethral plate in cKO) and furthermore a failure of the internal urethra, which can be shown connecting clearly to the BL in midsagittal sections (*E*), to diverge from the rectum (*F*) with the two converged prior to exiting the body as a single external opening (red arrow). Whole-mount images of E18.5 fixed specimens show the range of urethral phenotypes observed in male *Sall1* cKO (*G*). IF staining against pHH3 shows decreased proliferation in the mesenchymal compartment of *Sall1* cKO E16.5 GTs (induced with doxycycline daily from E9.5) from both sexes (*H*). *Bmp4* expression is dramatically aberrantly increased in *Sall1* cKO E16.5 male GTs (induced with doxycycline daily from E9.5) by ISH (red staining, white arrows) and qRT-PCR and decreased *Wnt5a* expression by qRT-PCR (*I*). Proposed placement of Sall1 within existing known GT developmental pathways based on findings (*J*).

expression in male than female GTs at E17.5 and appears to be repressed by β-catenin, since loss of β-catenin in Msx2-expressing cells leads to up-regulation of the transcript in the GTs of both sexes (50% in males and 30% in females). Knockdown of *Meis1* in GTme cells also decreased the growth potential of the cells in both sexes by ~40%. *Gli3* transcript also does not show differential expression between the sexes at E14.5 but has 30% higher expression in males than females at E17.5. Knockdown of *Gli3*, as expected, showed male-specific decrease in growth potential by

88%. Lef1 and Gli3 have previously described roles in GT development (discussed later).

To discuss the need to use streamlined validation to refine hit lists such as ours, we must discuss what factors can produce a false-negative or a false-positive hit. First, the nearest gene analysis identifies the nearest gene to a peak, and while this is considered a standard to identify regulated genes near binding sites, the nearest gene is not always the gene regulated by a particular binding event and can therefore produce a false-positive hit. Additionally, nonproductive binding

is possible even in multiprotein complexes, particularly if necessary complex members are absent; this can also produce a false-positive hit. On the other hand, filters tend to be overly strict. For this reason, we use them largely to identify top candidates, while maintaining that lower-tier candidates could also be major players. Sexual dimorphic expression of Sall1 is quite modest at the time points tested, and this gene could have easily slipped through our top candidate filter. The requirement of Sall1 in GT development highlights the need to treat every hit produced by this assay as worthy of exploration and tiered by priority. One caveat of our DECIPHER analysis is that virtually all CNVs in the database cover more than one gene. While the hit gene itself may very well be the LGU-relevant gene in question, this analysis can be viewed more generally to confirm genomic hotspots of urogenital development based on our peaks rather than specific genes, and in this sense, the genes surrounding the hit gene are also potentially important.

The TF hit DBX1, typically considered an exclusively neurodevelopmental patterning gene, has 40% higher expression in the male E14.5 GT than the female but by E17.5 expression is overall higher but not sexually dimorphic. This suggests an unexpected early role of DBX1 in GT masculinization. In the GTme screen, siRNA against DBX1 exhibited male-specific reduction of growth potential by 35%. In fact, neurogenesis genes were unexpectedly enriched in the overall hits (PANTHER go-term "neurogenesis" enrichment $P = 1.68E{-37}$) implicating significant involvement of neurogenesis genes in LGU formation and function, possibly for their underlying patterning roles. Other neurogenesis hits, while not TFs and not tested in our expression or GTme siRNA screenings, include the ephrin type A receptors EPHA 2/3/4/5/6/7, axon guidance receptor genes ROBO 1/2, as well as neurogenesis associated ligands like glial-derived growth factor. Additionally, several interesting intercellular signaling molecule hits were identified, both neurogenesis associated and not, that likely orchestrate LGU patterning. Several of these have multiple gene family members represented in the hits, for example WNT morphogens (WNT 3/5A/5B/11), semaphorins (SEMA 3A/3C/3D/3E/5A/5B/6A/6D), and fibroblast growth factors (FGF 5/6/9/12/13/14/18).

Some TF hits with previously confirmed roles in LGU development include but are not limited to V-maf musculoaponeurotic fibrosarcoma oncogene homolog B (MAFB) (53, 54), GLI2 (50, 55), ESR1 (12), FOXA2 (56), GLI3 (50, 51), LEF1 (57, 58), and SP8 (59). MAFB is a dimorphically expressed, AR-regulated protein during genitalia formation, and knockout male GTs exhibit hypospadias but have normal outgrowth (54). Sonic hedgehog (SHH) (60–63) signaling from the urethral plate epithelium is mediated by GLI2 in the mesenchyme, and both are instrumental for the dimorphic expression of downstream genes in coordination with AR during genitalia formation (55). ESR1, while at first thought to be dispensable to male development, has since been proven to be required in mice (64) and rats (65) for normal testicular and prostatic development and function, and many sexually dimorphically expressed genes have both androgen and estrogen response elements (66). The finding that ESR1 is a target for AR in developing male GTs is consistent with previous findings showing sexually dimorphic expression in the mouse (21) and human genitalia (67) and is consistent with our qRT-PCR data showing threefold more *Esr1* transcript in female E17.5 GTs relative to E17.5 male GTs, and relative to E14.5 GTs of either sex. The presence of so many TFs in our hit list with previously confirmed roles in LGU development lends credence to the likelihood that TF hits without previously reported LGU roles are fundamentally involved. With the use of murine models, this was confirmed to be the case for the TF SALL1, for which previous studies have focused largely on the kidney and upper tract function while not assessing the lower tract.

Last, it is imperative to discuss the clinical and functional significance of these findings. Forty sexually dimorphically expressed TF hits are associated with disproportionately high rates of LGU phenotypes in patients with CNVs covering these genes. As a positive control, this list includes some previously documented urogenital master regulators such as ALX4 (68), MAFB (53, 54), and doublesex-mab3-related transcription factor (DMRT) family genes (69). Thirty-one, or 77.5% of these TFs were also shown to be required for appropriate GTme proliferation during key developmental stages using our functional proliferation screening. In this study, we focused almost entirely on TFs. If the same streamlined validation approach is undertaken for families of hits other than TFs, such as intercellular signaling molecules, hydrolases or transferases, or transmembrane transporters, we would expect a similar percentage of hits to end up in the top-tier of candidacy. The present study justifies generation of conditional ablation in vivo models for unexpected hit genes that otherwise would not be explored for their potential roles in LGU formation and function, and it serves as a template for the versatile and clinically translational uses of SpDam technology, exploring other TF duos and other tissues as they develop, differentiate, age, become diseased, repair, or sustain homeostasis.

Methods

Generation of Transgenic Line biTetO::AR^D/p300$^{AM-ERT2}$. Standard molecular cloning was performed to generate the biTetO::AR^D/p300$^{AM-ERT2}$ plasmid. The Mouse Genetics Core at Washington University in St. Louis used the linearized plasmid for zygote pronuclear microinjection to generate mosaic blastocysts which were implanted into pseudopregnant females. Sixty-three resultant F$_0$ pups were genotyped from tail DNA, and four potential founders were identified. One founder was sickly and infertile, a second did not have germline transfer, and another founder did not exhibit expression of the transgene in the presence of doxycycline. The final founder successfully expressed the transgene and exhibited F$_1$ pups positive for the transgene; thus all animals in this transgenic line have the same insertion point, although the location is random and unknown. Importantly, we note that the chromosome of insertion must be the same as the chromosome of insertion for Msx2::rtTA (also unknown) because of the inability to get double-positive or double-negative pups when a double-positive parent is crossed with a wild-type. Therefore, in order to generate double-positive embryos, Msx2::rtTA females were crossed to biTetO::AR^D/p300$^{AM-ERT2}$ males. Females were used that possess only one copy of Msx2::rtTA due to minor phenotypes present in homozygotes that could interfere with the experiment. Single-copy status was assigned to any Msx2::rtTA female with one parent wild-type for the transgene. Although it is unlikely to be an issue, we avoided using the reverse cross of biTetO::AR^D/p300$^{AM-ERT2}$ females to Msx2::rtTA males in order to avoid contamination of biTetO::AR^D/p300$^{AM-ERT2}$ maternal DNA in the fetal specimen DNA, as the promoter was proven to leak especially in the presence of doxycycline, which the mother received as gavage in order to dose the embryos. Genotyping primers are 5′-TAGGGCTGGGAAGGGTCTAC-3′ and 5′-GGCAGCTGAGTCATCCTCGT-3′. Primers to detect expression of transgene are 5′-TCCTCGGTCTCGATTCTACG-3′ and 5′-ATCGGCAAGGATGTAACGAG-3′ for AR^D and 5′-CAGGATCCGTTCCGTCGGGC-3′ and 5′-AATAGACGACGGATGCATC-3′ for p300$^{AM-ERT2}$. For Rpl7 internal control, primers are 5′-GGCTACGGCAAAATCAACAA-3′ and 5′-TTTCTTCATCCCACCTCGTG-3′. Genotyping and RT-PCR were performed using GoTaq-G2 polymerase (Promega cat. # M7848) with 55 °C annealing temperature according to the manufacturer's protocol. IF against Era/ERT2 on GTme cells was performed as previously described (70) using antibody (MC-20) santa cruz sc-542. IF against GFP was performed on E17.5 GTs as previously described (71) using antibody A11122 from Life Technologies.

Additional Animal Models and Use. All animal husbandry and experimentation was carried out in compliance with Institutional Animal Care and Use Committee and institutional protocols. β-catenin flox mice were purchased from Jackson Laboratories stock #004152. LacZ reporter R26R strain with flox neo cassette upstream of LacZ was purchased from Jackson Laboratories stock #003309. Stoplight reporter mTmG strain was purchased from Jackson Laboratories stock #007676. TetO::cre strain (72) was purchased from Jackson Laboratories stock #006234. Msx2::rtTA transgenic mice (24) and Vim::rtTA (37) were generated previously by our group.

Preparation of SpDam Libraries from the Murine Embryonic Genitalia. Msx2::rtTA females were time mated to biTetO::AR^D/p300$^{AM-ERT2}$ males with the morning of plug considered to be 0.5 d postcoitum. Doxycycline hyclate (Sigma cat. # D9891-10G) was prepared in water at a concentration of 25 mg/mL and given to the females at a concentration of 100 mg/kg by daily gavage from E13.5 to E16.5. Similarly, tamoxifen (Sigma cat. #T5648-1G) was prepared in corn oil at a concentration of 25 mg/mL and given to the females at a concentration of 200 mg/kg by daily gavage from E14.5 to E16.5. Embryos were harvested at E17.5, and GTs were carefully dissected for downstream preparation. Tail clippings were collected in parallel for embryonic genotyping. Male embryos double-positive for the transgenes were used for experimental preparations while male embryos negative for to biTetO::AR^D/p300$^{AM-ERT2}$ were used for background signal preparations. Freshly collected, never-frozen tissue followed by gentle enzymatic (nonmechanical) digestion is imperative to minimize background signal due to DNA breakage prior to DpnI methyl-specific blunt-end digestion. Freshly collected GTs were subjected to digestion in 500 μL proteinase-K lysis solution (100 mM Tris-Cl pH 8.5, 5 mM ethylenediaminetetraacetic acid (EDTA), 2% SDS, and 20 mM NaCl, with 0.5 μL proteinase K per specimen, incubated at 52 to 55 °C overnight with shaking) followed by phenol–chloroform DNA extraction using Gel Lock Gel™ Heavy according to the manufacturer's protocol. Following precipitation, resuspension, and treatment with RNaseA, 2.5 μg genomic DNA per specimen was subjected to DpnI digestion (New England Biolabs cat. #R0176S) at 37 °C overnight in a total volume of 10 μL and 20U DpnI enzyme per reaction, with heat inactivation at 80 °C for 20 min. In the same tube, samples were immediately subjected to blunt-end ligation of double-stranded adaptors (preannealed oligos 5′-CTAATACGACTCACTATAGGGCAGCGTGGTCGCGGCCGAGGA-3′ and 5′-TCCTCGGCCG-3′ using T4 ligase (New England Biolabs cat. #M0202S) overnight at 16 °C with inactivation at 65 °C for 10 min. Immediately following adapter ligation, 3 μL of DNA was subjected to adapter-specific PCR amplification using GoTaq-G2 polymerase (One 50 μL reaction = 3 μL adapter-ligated genomic DNA, 10 μL 5× GoTaq buffer, 2 μL 25× deoxynucleotide triphosphate (dNTPs), 1.25 μL 50 μM primer 5′phospho-GGTCGCGGCCGAGGATC-3′, 33.5 μL water, and 0.25 μL GoTaq enzyme) under the cycling conditions pre-PCR 72 °C 10 min → pre-PCR 94 °C 1 min → pre-PCR 55 °C 5 min → pre-PCR 72 °C 15 min → six cycles of 94 °C 1 min followed by 55 °C 1 min followed by 72 °C 10 min, → 25 cycles of 94 °C 1 min followed by 55 °C 1 min followed by 72 °C 2 min → 4 °C hold. The use of GoTaq-G2 negates the need for A-tailing prior to sequencing preparation due to the presence of single 3′-A overhangs in >80% of PCR products. At least eight PCR reactions per specimen were pooled for PCR cleanup (Qiagen cat. #28104) to maximize library yield, with one additional reaction being used to run to agarose gel for quality control alongside untreated negative controls. Each DNA specimen was then prepared for sequencing by T4-ligation of preannealed Illumina index adapters (3 μL at 50 μM per 50 μL reaction) in a reaction containing 5 μg DNA according to the manufacturer's protocol (New England Biolabs cat. #M0202S) overnight at 16 °C with inactivation at 65 °C for 10 min. Samples then underwent a second PCR cleanup, followed by size-exclusion by gel purification to avoid adapter dimers (Qiagen cat. #28704), followed by a third PCR cleanup to minimize buffer carryover. Samples then underwent an enrichment PCR prior to sequencing, again performing eight or more reactions per sample to increase yield and pooling during purification. Phusion enrichment PCR (New England Biolabs cat. #M0530S) was performed using 250 ng DNA per reaction and Y-adapter specific primers 5′- AATGATACGGCGACCACCGAGATCTACAC-3′ and 5′-CAAGCAGAAGACGGCATACGAGAT-3′ with cycling conditions 98 °C 30 s → 20 cycles of 98 °C 40 s followed by 60 °C 30 s followed by 72 °C 45 s → 72 °C 5 min → 4 °C hold. This was followed by PCR cleanup, size-exclusion by gel extraction, and a final PCR cleanup to minimize buffer carryover from gel extraction. Indexed samples were pooled at equimolar ratios for Illumina sequencing.

Sequencing and Analysis. Experimental and control samples were indexed and sequenced on Illumina NextSeq flowcell at an average of 19 M reads per sample. Reads were aligned to the mm10 mouse reference genome sequence using the Burrows–Wheeler Aligner (BWA) (73) using the mem algorithm. Per base read counts were obtained using the Samtools (74) mpileup algorithm. BedGraph files were created using a custom script using mpileup files as input. BedGraph files were averaged between biological replicates, and averages were compared between sample conditions using a custom script. Analysis was performed pairwise between each experimental sample and its pooled littermate controls. Nearest mouse genes were identified using a custom script and a reference file of human-mouse ortholog gene

locations within the mm10 mouse reference sequence. HOMER (https://homer.ucsd.edu/homer/ngs/) (75) was used to call and annotate SpDam peaks. AR ChIP-Seq peaks were obtained as bed files from the Gene Expression Omnibus (GEO) database for each referenced dataset, and nearest genes for each peak were obtained using a custom script and a reference file of human genes with the hg19 reference sequence. Raw and analyzed files are available in GEO–accession GSE242628.

Expression and Functional Screening. Tissue was collected and stored at −80 °C, parallel tail clippings were genotyped, and GTs of the same age and genotype were pooled at N ≥ 6 for tissue dounce homogenization and RNA extraction according to the manufacturer's protocol (Qiagen cat. #74104) followed by reverse transcription cDNA library preparation according to the manufacturer's protocol (Thermo cat. #4368814). qRT-PCR was performed using SYBR green (Thermo cat. #A25742) and Viaa7 instrument according to the manufacturer's protocol. To detect expression of transgene AR transcripts relative to endogenous transcripts, the endogenous-specific transcript primers are 5′-TAAAGGGGCTACCCAGACCA-3′ and 5′-TCTGCCAATGGGTATGCTGG-3′, and the primers that bind both endogenous and exogenous AR are 5′-ACTTTCCACCCCAGAAGACC-3′ and 5′-CCTTCAGCGGCTCTTTTG-3′. Timed matings were used to collect E14.5 or E17.5 GTs from wild types and β-catenin fl/fl; Msx2::rtTA; TetO::cre embryos treated with a single maternal pulse gavage of 100 mg/kg doxycycline at E12.5. To assess sexually dimorphic expression, E14.5 and E17.5 male and female specimens were compared with normalization to female for each transcript and time point. To assess differences dependent on β-catenin levels, wild types were compared to conditionally ablated specimens of each sex and normalized to wild-type specimens for each transcript. All specimens were tested in technical triplicate at minimum. Primer sequences can be found in Dataset S10. Student's *t* test of relative triplicate measures between specimens with $P < 0.05$ was considered significant. The in vitro siRNA assay was performed as previously described (37) using the Vimentin-based FACS isolation method.

Characterization of the Sall1 Conditional Knockout Genitalia. RNAscope ISH was performed according to the manufacturer's protocol (ACDbio, probe reference 469661) on E14.5 wild-type, β-catenin fl/fl; Msx2::rtTA; TetO::cre, and cre-negative control littermates to assess Sall1 expression. To assess morphology of Sall1 fl/fl; Msx2::rtTA; TetO::cre males relative to cre-negative controls (all embryos treated by maternal doxycycline 100 mg/kg daily gavage from E9.5 to E16.5), GTs were collected at E17.5 and fixed in Bouin's solution for 4 h at room temperature followed by serial ethanol dehydration to xylene, and paraffin embedding for transverse or sagittal 5 μm sectioning followed by standard hematoxylin and eosin staining. While all conditionally ablated male Sall1 GTs (>25 analyzed) showed some combination of tubularization defects and proliferative defects, the severity varied slightly between specimens. IF against pHH3 was performed on E16.5 GTs as previously described (71) using antibody Cell Signaling Cat #9701 and quantified using ImageJ software. RNAscope ISH was performed according to the manufacturer's protocol (ACDbio, probe reference 425011) on E16.5 Sall1 fl/fl; Msx2::rtTA; TetO::cre, and cre-negative control littermates to assess Bmp4 expression. Bmp4 primer sequences for qRT-PCR are 5′-GAACAGGGCTTCCACCGTAT-3′ forward and 5′-GTGTCTGGTGGAGGTGAGTC-3′ reverse. Wnt5a primer sequences for qRT-PCR are 5′-TTCATGGGCTTGGTAGCCTT-3′ forward and 5′-CTCCACTCTGCTCATAAACATGAA-3′ reverse.

1. M. Haller, L. Ma, Temporal, spatial, and genetic regulation of external genitalia development. *Differentiation* **110**, 1–7 (2019).
2. K. Oztarhan *et al.*, Prevalence and distribution of congenital abnormalities in Turkey: Differences between the prenatal and postnatal periods. *Congenit. Anom. (Kyoto)* **50**, 221–225 (2010).
3. A. M. Rolston, M. Gardner, E. Vilain, D. E. Sandberg, Parental reports of stigma associated with child's disorder of sex development. *Int. J. Endocrinol.* **2015**, 980121 (2015).
4. C. L. Acerini, H. L. Miles, D. B. Dunger, K. K. Ong, I. A. Hughes, The descriptive epidemiology of congenital and acquired cryptorchidism in a UK infant cohort. *Arch. Dis. Child.* **94**, 868–872 (2009).
5. A. Springer, M. van den Heijkant, S. Baumann, Worldwide prevalence of hypospadias. *J. Pediatr. Urol.* **12**, e151–e157 (2016).
6. N. Hatipoğlu, S. Kurtoğlu, Micropenis: Etiology, diagnosis and treatment approaches. *J. Clin. Res. Pediatr. Endocrinol.* **5**, 217–223 (2013).
7. A. Krishnan, A. de Souza, R. Konijeti, L. S. Baskin, The anatomy and embryology of posterior urethral valves. *J. Urol.* **175**, 1214–1220 (2006).
8. I. J. Murawski, I. R. Gupta, Vesicoureteric reflux and renal malformations: A developmental problem. *Clin. Genet.* **69**, 105–117 (2006).
9. J. Shi *et al.*, Status of comorbid congenital anomalies and their influence on resource use in pediatric inpatients: A serial cross-sectional study in Shanghai, China. *Front. Public Health* **8**, 580664 (2020).
10. M. Haller, Q. Mo, A. Imamoto, D. J. Lamb, Murine model indicates 22q11.2 signaling adaptor CRKL is a dosage-sensitive regulator of genitourinary development. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 4981–4986 (2017).
11. M. Haller, J. Au, M. O'Neill, D. J. Lamb, 16p11.2 transcription factor MAZ is a dosage-sensitive regulator of genitourinary development. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E1849–E1858 (2018).
12. M. Tannour-Louet *et al.*, Increased gene copy number of VAMP7 disrupts human male urogenital development through altered estrogen action. *Nat. Med.* **20**, 715–724 (2014).
13. M. R. Hass *et al.*, SpDamID: Marking DNA bound by protein complexes identifies notch-dimer responsive enhancers. *Mol. Cell* **64**, 213 (2016).
14. T. P. Wu *et al.*, DNA methylation on N(6)-adenine in mammalian embryonic stem cells. *Nature* **532**, 329–333 (2016).
15. A. Avendaño, I. Paradisi, F. Cammarata-Scalisi, M. Callea, 5-α-reductase type 2 deficiency: Is there a genotype-phenotype correlation? A review. *Hormones (Athens)* **17**, 197–204 (2018).
16. P. Chauhan, A. Rani, S. K. Singh, A. K. Rai, Complete androgen insensitivity syndrome due to mutations in the DNA-binding domain of the human androgen receptor gene. *Sex. Dev.* **12**, 269–274 (2018).
17. C. L. Jordan, L. Doncarlos, Androgens in health and disease: An overview. *Horm. Behav.* **53**, 589–595 (2008).
18. K. A. Mason, M. J. Schoelwer, A. D. Rogol, Androgens during infancy, childhood, and adolescence: Physiology and use in clinical practice. *Endocr. Rev.* **41**, bnaa003 (2020).
19. G. C. Shukla, A. R. Plaga, E. Shankar, S. Gupta, Androgen receptor-related diseases: What do we know? *Andrology* **4**, 366–381 (2016).
20. C. A. Quigley *et al.*, Androgen receptor defects: Historical, clinical, and molecular perspectives. *Endocr. Rev.* **16**, 271–321 (1995).
21. Z. Zheng, B. A. Armfield, M. J. Cohn, Timing of androgen receptor disruption and estrogen exposure underlies a spectrum of congenital penile anomalies. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E7194–E7203 (2015).
22. T. Sato *et al.*, Brain masculinization requires androgen receptor function. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 1673–1678 (2004).
23. D. El-Maouche, W. Arlt, D. P. Merke, Congenital adrenal hyperplasia. *Lancet* **390**, 2194–2210 (2017).
24. C. Lin *et al.*, Construction and characterization of a doxycycline-inducible transgenic system in Msx2 expressing cells. *Genesis* **47**, 352–359 (2009).
25. R. Feil, J. Wagner, D. Metzger, P. Chambon, Regulation of Cre recombinase activity by mutated estrogen receptor ligand-binding domains. *Biochem. Biophys. Res. Commun.* **237**, 752–757 (1997).
26. X. Feng, C. He, Mammalian DNA N6-methyladenosine: Challenges and new insights. *Mol. Cell* **83**, 343–351 (2023).
27. V. Gotea, I. Ovcharenko, DiRE: Identifying distant regulatory elements of co-expressed genes. *Nucleic Acids Res.* **36**, W133–W139 (2008).
28. B. Lin *et al.*, Integrated expression profiling and ChIP-seq analyses of the growth inhibition response program of the androgen receptor. *PLoS One* **4**, e6589 (2009).
29. E. Nevedomskaya *et al.*, Androgen receptor DNA binding and chromatin accessibility profiling in prostate cancer. *Genom. Data* **7**, 124–126 (2016).
30. D. Kajioka *et al.*, Sexual fate of murine external genitalia development: Conserved transcriptional competency for male-biased identity in both sexes. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2024067118 (2021).
31. A. W. Olson *et al.*, Stromal androgen and hedgehog signaling regulates stem cell niches in pubertal prostate development. *Development* **148**, dev199738 (2021).
32. F. Geller *et al.*, Genome-wide association analyses identify variants in developmental genes associated with hypospadias. *Nat. Genet.* **46**, 957–963 (2014).
33. H. Mi *et al.*, PANTHER version 16: A revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Res.* **49**, D394–D403 (2021).
34. D. Szklarczyk *et al.*, STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).
35. B. A. Armfield, M. J. Cohn, Single cell transcriptomic analysis of external genitalia reveals complex and sexually dimorphic cell populations in the early genital tubercle. *Dev. Biol.* **477**, 145–154 (2021).
36. H. V. Firth *et al.*, DECIPHER: Database of chromosomal imbalance and phenotype in humans using ensembl resources. *Am. J. Hum. Genet.* **84**, 524–533 (2009).
37. Y. Yin, M. Haller, T. Li, L. Ma, Development of an in-vitro high-throughput screening system to identify modulators of genitalia development. *PNAS Nexus* **2**, pgac300 (2023).
38. M. D. Muzumdar, B. Tasic, K. Miyamichi, L. Li, L. Luo, A global double-fluorescent Cre reporter mouse. *Genesis* **45**, 593–605 (2007).
39. S. M. Kiefer *et al.*, Sall1-dependent signals affect Wnt signaling and ureter tip fate to initiate kidney development. *Development* **137**, 3099–3106 (2010).
40. T. Chen *et al.*, Mutation screening of BMP4, BMP7, HOXA4 and HOXB6 genes in Chinese patients with hypospadias. *Eur. J. Hum. Genet.* **15**, 23–28 (2007).
41. K. Suzuki *et al.*, Regulation of outgrowth and apoptosis for the terminal appendage: External genitalia development by concerted actions of BMP signaling [corrected]. *Development* **130**, 6209–6220 (2003).
42. A. M. Herrera, S. G. Shuster, C. L. Perriton, M. J. Cohn, Developmental basis of phallus reduction during bird evolution. *Curr. Biol.* **23**, 1065–1074 (2013).
43. G. Robertson *et al.*, Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat. Methods* **4**, 651–657 (2007).
44. S. A. Nichols, B. W. Roberts, D. J. Richter, S. R. Fairclough, N. King, Origin of metazoan cadherin diversity and the antiquity of the classical cadherin/β-catenin complex. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 13046–13051 (2012).
45. M. B. Furtado *et al.*, Point mutations in murine Nkx2-5 phenocopy human congenital heart disease and induce pathogenic Wnt signaling. *JCI Insight* **2**, e88271 (2017).
46. L. Cambier, M. Plate, H. M. Sucov, M. Pashmforoush, Nkx2-5 regulates cardiac growth through modulation of Wnt signaling by R-spondin3. *Development* **141**, 2959–2971 (2014).
47. H. W. Song *et al.*, The RHOX homeobox gene cluster is selectively expressed in human oocytes and male germ cells. *Hum. Reprod.* **28**, 1635–1646 (2013).
48. S. K. Srivastava *et al.*, Myb overexpression overrides androgen depletion-induced cell cycle arrest and apoptosis in prostate cancer cells, and confers aggressive malignant traits: Potential role in castration resistance. *Carcinogenesis* **33**, 1149–1157 (2012).
49. N. C. Whitlock *et al.*, MEIS1 down-regulation by MYC mediates prostate cancer development through elevated HOXB13 expression and AR activity. *Oncogene* **39**, 5663–5674 (2020).
50. F. He *et al.*, Adult Gli2+/−;Gli3Δ699/+ male and female mice display a spectrum of genital malformation. *PLoS One* **11**, e0165958 (2016).
51. A. Kothandapani *et al.*, GLI3 resides at the intersection of hedgehog and androgen action to promote male sex differentiation. *PLoS Genet.* **16**, e1008810 (2020).
52. O. Medina-Martinez *et al.*, The transcription factor Maz is essential for normal eye development. *Dis. Model. Mech.* **13**, dmm044412 (2020).
53. S. Matsushita *et al.*, Androgen regulates Mafb expression through its 3'UTR during mouse urethral masculinization. *Endocrinology* **157**, 844–857 (2016).
54. K. Suzuki *et al.*, Sexually dimorphic expression of Mafb regulates masculinization of the embryonic urethral formation. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 16407–16412 (2014).
55. S. Miyagawa *et al.*, The role of sonic hedgehog-Gli2 pathway in the masculinization of external genitalia. *Endocrinology* **152**, 2894–2903 (2011).
56. M. L. Gredler, S. E. Patterson, A. W. Seifert, M. J. Cohn, Foxa1 and Foxa2 orchestrate development of the urethral tube and division of the embryonic cloaca through an autoregulatory loop with Shh. *Dev. Biol.* **465**, 23–30 (2020).
57. D. N. Shelton *et al.*, The role of LEF1 in endometrial gland formation and carcinogenesis. *PLoS One* **7**, e40312 (2012).
58. R. D. Mullen, R. R. Behringer, Molecular genetics of Müllerian duct formation, regression and differentiation. *Sex. Dev.* **8**, 281–296 (2014).
59. C. Lin *et al.*, Delineating a conserved genetic cassette promoting outgrowth of body appendages. *PLoS Genet.* **9**, e1003231 (2013).
60. S. Miyagawa *et al.*, Dosage-dependent hedgehog signals integrated with Wnt/beta-catenin signaling regulate external genitalia formation as an appendicular program. *Development* **136**, 3969–3978 (2009).
61. C. L. Perriton, N. Powles, C. Chiang, M. K. Maconochie, M. J. Cohn, Sonic hedgehog signaling from the urethral epithelium controls external genital development. *Dev. Biol.* **247**, 26–46 (2002).
62. A. W. Seifert, C. M. Bouldin, K. S. Choi, B. D. Harfe, M. J. Cohn, Multiphasic and tissue-specific roles of sonic hedgehog in cloacal septation and external genitalia development. *Development* **136**, 3949–3957 (2009).
63. A. W. Seifert, Z. Zheng, B. K. Ormerod, M. J. Cohn, Sonic hedgehog controls growth of external genitalia by regulating cell cycle kinetics. *Nat. Commun.* **1**, 23 (2010).
64. M. Chen *et al.*, Defects of prostate development and reproductive system in the estrogen receptor-alpha null male mice. *Endocrinology* **150**, 251–259 (2009).
65. M. A. Rumi *et al.*, Generation of Esr1-knockout rats using zinc finger nuclease-mediated genome editing. *Endocrinology* **155**, 1991–1999 (2014).
66. R. A. Hess, P. S. Cooke, Estrogen in the male: A historical perspective. *Biol. Reprod.* **99**, 27–44 (2018).
67. N. B. Kalloo, J. P. Gearhart, E. R. Barrack, Sexually dimorphic expression of estrogen receptors, but not of androgen receptors in human fetal external genitalia. *J. Clin. Endocrinol. Metab.* **77**, 692–698 (1993).
68. D. Matsumaru *et al.*, Genetic analysis of the role of Alx4 in the coordination of lower body and external genitalia formation. *Eur. J. Hum. Genet.* **22**, 350–357 (2014).
69. K. Ounap *et al.*, Three patients with 9p deletions including DMRT1 and DMRT2: A girl with XY complement, bilateral ovotestes, and extreme growth retardation, and two XX females with normal pubertal development. *Am. J. Med. Genet. A* **130A**, 415–423 (2004).
70. Y. Yin, C. Lin, L. Ma, MSX2 promotes vaginal epithelial differentiation and wolffian duct regression and dampens the vaginal response to diethylstilbestrol. *Mol. Endocrinol.* **20**, 1535–1546 (2006).
71. Y. Yin, M. E. Haller, S. B. Chadchan, R. Kommagani, L. Ma, Signaling through retinoic acid receptors is essential for mammalian uterine receptivity and decidualization. *JCI Insight* **6**, e150254 (2021).
72. A. K. Perl, S. E. Wert, A. Nagy, C. G. Lobe, J. A. Whitsett, Early restriction of peripheral and proximal cell lineages during formation of the lung. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 10482–10487 (2002).
73. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
74. H. Li *et al.*, The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
75. S. Heinz *et al.*, Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
76. M. Haller *et al.*, Data from "Streamlined identification of clinically and functionally relevant genetic regulators of lower-tract urogenital development." Gene Expression Omnibus (GEO). https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE242628. Deposited 7 September 2023.