Next generation sequencing reveals novel genetic variants (SRY, DMRT1, NR5A1, DHH, DHX37) in adults with 46,XY DSD

Federica Buonocore, Oliver Clifford-Mobley, Tom F.J. King, Niccolò Striglioni, Elim Man, Jenifer P. Suntharalingham, Ignacio del Valle, Lin Lin, Carlos F. Lagos, Gill Rumsby, Gerard S. Conway, John C. Achermann

Journal of the Endocrine Society
Endocrine Society

Submitted: August 15, 2019
Accepted: October 04, 2019
First Online: October 10, 2019

Advance Articles are PDF versions of manuscripts that have been peer reviewed and accepted but not yet copyedited. The manuscripts are published online as soon as possible after acceptance and before the copyedited, typeset articles are published. They are posted "as is" (i.e., as submitted by the authors at the modification stage), and do not reflect editorial changes. No corrections/changes to the PDF manuscripts are accepted. Accordingly, there likely will be differences between the Advance Article manuscripts and the final, typeset articles. The manuscripts remain listed on the Advance Article page until the final, typeset articles are posted. At that point, the manuscripts are removed from the Advance Article page.

DISCLAIMER: These manuscripts are provided "as is" without warranty of any kind, either express or particular purpose, or non-infringement. Changes will be made to these manuscripts before publication. Review and/or use or reliance on these materials is at the discretion and risk of the reader/user. In no event shall the Endocrine Society be liable for damages of any kind arising references to, products or publications do not imply endorsement of that product or publication.

Journal of the Endocrine Society; Copyright 2019 DOI: 10.1210/js.2019-00306

NGS for adult 46,XY DSD

Next generation sequencing reveals novel genetic variants (SRY, DMRT1, NR5A1, DHH, DHX37) in adults with 46,XY DSD

Federica Buonocore,[1] Oliver Clifford-Mobley,[2] Tom F.J. King,[3] Niccolò Striglioni, [1] Elim Man, [1] Jenifer P. Suntharalingham, [1] Ignacio del Valle, [1] Lin Lin, [1] Carlos F. Lagos,[4] Gill Rumsby,[2] Gerard S. Conway,[3]* John C. Achermann[1]*

[1]Genetics & Genomic Medicine, UCL Great Ormond Street Institute of Child Health, University College London, London WC1N 1EH, United Kingdom; [2]Clinical Biochemistry, University College London Hospitals, London W1T 4EU, United Kingdom; [3]Reproductive Medicine Unit, University College London Hospitals, London NW1 2PG, United Kingdom; [4]Chemical Biology & Drug Discovery Lab, Facultad de Medicina y Ciencia, Universidad San Sebastián, Lota 2465 Providencia, Santiago, Chile.

ORCiD numbers:

0000-0002-8274-7241

Buonocore

Federica

0000-0003-1715-2054

Clifford-Mobley

Oliver

0000-0002-0383-3433

King

Tom F.J.

0000-0003-1213-3049

Striglioni

Niccolò

0000-0003-1549-2450

Man

Elim

0000-0002-8897-7697

Suntharalingham

Jenifer P.

0000-0003-2595-3652

del Valle

Ignacio

0000-0002-5432-9747

Lagos

Carlos F.

0000-0002-3223-3752

Rumsby

Gill

0000-0002-3346-5029

Conway

Gerard S.

0000-0001-8787-6272

Achermann

John C

Received 15 August 2019. Accepted 04 October 2019.

* denotes equal contribution

This article has been published under the terms of the Creative Commons Attribution License (CC-BY; https://creativecommons.org/licenses/by/4.0/).

Context
The genetic basis of human sex development is slowly being elucidated and more than 40 different genetic causes of differences (or disorders) of sex development (DSD) have now been reported. However, reaching a specific diagnosis using traditional approaches can be difficult, especially in adults where limited biochemical data may be available.

Objective
We used a targeted next-generation sequencing approach to analyze known and candidate genes for DSD in individuals with no specific molecular diagnosis.

Partcipants and Design
We studied 52 adult 46,XY women attending a single-center adult service, who were part of a larger cohort of 400 individuals. Classic conditions such as17β-hydroxysteroid dehydrogenase deficiency type 3, 5α-reductase deficiency type 2 and androgen insensitivity syndrome were excluded. The study cohort had broad working diagnoses of complete gonadal dysgenesis (CGD) (n=27) and partially-virilised 46,XY DSD (pvDSD) (n=25), a group that included partial gonadal dysgenesis (PGD) and those with a broad ”partial androgen insensitivity syndrome” label. Targetted sequencing of 168 genes was undertaken.

Results
Overall a likely genetic cause was found in 16/52 (30.8%) individuals (22.2% CGD; 40.0% pvDSD). Pathogenic variants were found in SRY (n=3), DMRT1 (n=1), NR5A1/SF-1 (n=1) and DHH (n=1) in the CGD group, and in NR5A1 (n=5), DHH (n=1) and DHX37 (n=4) in the pvDSD group.

Conclusions
Reaching a specific diagnosis can have clinical implications and provides insight into the role of these proteins in sex development. Next-generation sequencing approaches are invaluable, especially in adult populations or where diagnostic biochemistry is not possible.

Keywords: sex determination, DSD, SRY, steroidogenic factor-1, DHX37, desert hedgehog

Introduction

It is now more than 25 years since SRY was identified as the main testis-determining gene in humans and mice and as a cause of 46,XY differences (disorders) of sex development (DSD)[1–3]. Since then, at least 40 other genetic causes of DSD have been reported, but the relative contribution to these genes within the clinical setting is poorly documented[4–6].

DSD is usually considered in three broad categories: sex chromosome DSD (SCDSD), 46,XY DSD and 46,XX DSD [7]. Whilst the diagnosis of specific forms of SCDSD and 46,XX DSD can usually be made with karyotyping or biochemical analysis (for congenital adrenal hyperplasia, CAH), the specific diagnosis of 46,XY DSD is often more challenging.

46,XY DSD is commonly divided into conditions affecting gonad (testis) development (e.g. complete gonadal dysgenesis, also known as Swyer syndrome) or conditions affecting androgen biosynthesis and action. Proximal genetic blocks in androgen biosynthesis are rare and usually affect the adrenal gland as well as the gonad (steroidogenic acute regulatoty protein/*STAR*, P450 side-chain cleavage/*CYP11A1*, 3β-hydroxysteroid dehydrogenase type 3/*HSD3B2*, 17α-hydroxylase/*CYP17A1*, P450 oxidoreducatse/*POR*). More specific defects in testosterone biosynthesis (17β-hydroxysteroid dehydrogenase deficiency type 3, *HSD17B3*), in the conversion of testosterone to dihydrotestosterone (5α-reductase deficiency type 2, *SRD5A2*), or in androgen action (androgen insensitivity syndrome, androgen receptor/*AR*) can sometimes be suspected on clinical or biochemical grounds, but genetic testing is becoming increasingly important given the overlap in clinial features amongst various causes [8,9].

To complicate diagnosis further, each condition can have a spectrum of genital phenotypes depending on the underlying variant and other modifying factors. This situation is especially apparent for changes in steroidogenic factor-1 (SF-1, encoded by *NR5A1*), where clinical and endocrine features ranging from complete gonadal dysgenesis through to hypospadias or male factor infertility can be seen [10,11].

Newer technologies such as high-throughput next generation sequencing (NGS) now allow the analysis of many genes simultaneously in the DNA from one individual. Panel-based approaches and exome sequencing have been reported recently for the diagnosis of several endocrine conditions such as neonatal diabetes mellitus and primary adrenal insufficiency [12,13], as well as for DSD [14–20].

Here, we report the results of a targeted NGS analysis of a relatively large cohort of individuals with 46,XY DSD who had the working diagnosis of complete gonadal dysgenesis (CGD), or partially-virilised 46,XY DSD (pvDSD) (n=25), a group that included partial gonadal dysgenesis (PGD) and those with a broad ”partial androgen insensitivity syndrome (PAIS)” label (androgen receptor negative). We analyzed more than 150 genes that are either established or potential causes of 46,XY DSD, or candidate genes for 46,XY DSD based on our published “Atlas” of differential transcriptomic expression during early human embryonic/fetal gonad development [21].

1. Material and Methods

A. Study cohort

The study was undertaken in a cohort of 52 individuals (46 sporadic, 3 sibling pairs) with 46,XY DSD of unknown molecular etiology attending adult DSD clinics at University College London Hospitals NHS Foundation Trust (Fig. 1). This Adult DSD clinic takes place in the setting of a Department of Women’s Health. Therefore the spectrum of DSD phenotype is predominantly female. Consecutive subjects were approached without clinical selection criteria providing a real life clinical cohort. These individuals were part of a larger cohort of 400 women with 46,XY DSD, where 17β-hydroxysteroid dehydrogenase deficiency type 3 (17β-HSD) (*HSD17B3*) (n=26), 5α-reductase deficiency type 2 (5ARD) (*SRD5A2*) (n=30) and androgen insensitivity syndrome (AIS) (*AR*) (n=170) had been pre-screened (Fig. 1)[5,6].

A total of 27 women had a working diagnosis of CGD and mostly presented with female-typical genitalia and primary amenorrhoea/absent puberty in adolescence ("Swyer syndrome"). They were recruited from a total cohort of 84 women with CGD.

Another 25 individuals presented with atypical genitalia and were brought up as girls with a working diagnosis of PGD or "PAIS". They were recruited from a total of 53 women with this phenotype where the underlying diagnosis was unknown. Historically, many individuals have been labeled under the "umbrella" term "PAIS", without genetic or biochemical evidence of the diagnosis. We have termed this group "partially-virilised 46,XY DSD (pvDSD)".

Individuals with adrenal insufficiency or high steroidogenic blocks/CAH, renal anomalies or obvious syndromic associations of DSD were excluded.

Informed consent was given by all participants. Research Ethics Committee approval was obtained as part of The Reproductive Life Course Project (IRAS project ID 184646, REC reference 16/LO/0682) or study of the genetics of Reproductive Biology (07Q050824).

B. Genetic analysis

DNA was extracted from peripheral blood leukocytes using standard methods and subjected to genetic analysis using the approaches outlined below.

B-1. Sanger Sequencing for NR5A1

Analysis of *NR5A1* was undertaken using standard Sanger sequencing with primers and conditions reported previously [22].

B-2. Haloplex® targeted Next Generation Sequencing panel

A HaloPlex DNA targeted gene enrichment panel (Agilent Technologies, Inc., Santa Clara, CA) was designed using SureDesign software (www.agilent.com/genomics/suredesign). This panel included 168 known DSD-associated genes as well as potential candidate genes (Table 1). The panel captured all coding exons and 100bp of intronic flanking sequence of genes of interest with predicted target coverage >98.87%.

Genomic DNA samples (225ng) were processed for Illumina Sequencing according to the "HaloPlex Target Enrichment System" protocol (Version D.5, Agilent Technologies, Inc.) and as described previously [13]. Resultant libraries were then subjected to NGS using an Illumina MiSeq platform (Illumina, Inc., San Diego, CA). Raw FASTQ files were analysed using SureCall (v3.0.1.4) software (Agilent Technologies, Inc.). Visual inspection of BAM files was also undertaken to ensure deletions of key genes were not overlooked.

B-3. Nonacus® targeted Next Generation Sequencing panel

A further 12 potential known and candidate genes for DSD were analysed using a Cell3 Target enrichment panel for NGS (Nonacus Ltd, Birmingham, UK) (Table 1). This captured all coding exons, including 100 bp intronic flanking sequence, with a predicted target coverage of 97.35%.

Genomic DNA samples (100ng) were prepared for Illumina Sequencing according to the "Cell3™Target Cell Free DNA Target Enrichment for Next Generation Sequencing (Illumina sequencers) v1.2.2" protocol. In brief, DNA was sheared by an enzymatic fragmentation before undergoing end-repair and dA-tailing. This enables the Illumina UMI adapters to be ligated on both 5' and 3' ends. DNA was then purified using Agencourt AMPure beads to remove residual non ligated adapters. DNA was amplified using primers that bind to the ligated adapter. Libraries were hybridized with the customized probes before undergoing a further bead wash to remove non-targeted DNA. Targeted library DNA sequences were then amplified using primers binding specifically to sequences within the Illumina adapters. A final bead clean-up step was carried out before quantification of captured libraries. These were then sequenced on a Miseq platform (Illumina). Raw FASTQ files were analysed using a bioinformatic pipeline provided by Nonacus and consequent variant analysis was carried out in Ingenuity Variant Analysis software (Qiagen, Valencia, CA). Inspection of BAM files was carried out using Integrative Genomics Viewer (IGV) [23].

B-4. Variant validation

Validation of key variants identified was performed using Sanger sequencing. Regions of interest were PCR amplified and sequenced using a Big Dye Terminator v.1.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA) on an ABI3130 sequencer (Applied Biosystems), and visualized using Sequencher v5.2.4 (Gene Codes Corporation, Ann Arbor, MI). Control data for population genomic variation was obtained using the Genome Aggregation Database (gnomAD) (gnomAD, Cambridge, MA, https://gnomad.broadinstitute.org; accessed January 2019) [24].

B-5. Functional prediction of genetic variants

Functional predictions of the effects of specific variants was performed using SIFT, PolyPhen2 and MutationTaster algorithms (http://sift.jcvi.org/; http://genetics.bwh.harvard.edu/pph2/; http://www.mutationtaster.org/). Conservancy mapping was undertaken with UniProt (https://www.uniprot.org). Cartoons of protein domains were constructed using Domain Graph (DOG) version 2.0 [25].

C. Molecular modeling and simulation of DMRT-DNA complexes

The molecular models of wild-type (WT) and mutant DMRT1 (Doublesex and Mab3-related transcription factor 1) proteins were constructed using MODELLER, within the Protein Modeling module of Discovery Studio v2.1 (Accelrys Inc., San Diego, CA) [26]. The human DMRT1 sequence retrieved from the Uniprot database (https://www.uniprot.org/uniprot/Q9Y5R6) was used as a reference and only residues 70-130 were considered for modeling purposes [27]. Wild-type (WT) and p.R80S mutant were modeled using the recently reported crystal structure of human DMRT1 (70-130) in complex with DNA (PDB id 4YJ0) as template [28].

For each protein model, a set of 100 models were constructed and the best one according to the Modeller internal PDF score was selected for simulations steps. The DMRT1 WT and p.R80S mutant in complex with DNA systems were generated using VMD v1.93, by inserting each protein into a 110 × 70 × 80 Å box consisting of a classic TIP3P model for water molecules, neutralized with Na^+ or Cl^- ions [29,30]. Periodic boundary conditions were imposed in all three directions and the Particle Mesh Ewald (PME) method was used to account for full long-range electrostatic interactions within the selected boundary condition within a relative tolerance of $1x10^{-6}$ [31]. The final systems were composed of nearly 60,120 atoms.

Molecular dynamics simulations were carried out with the NAMD v2.13 simulation package, using the CHARMM36 force field parameters for proteins [32–34]. The simulation was started with a brief energy minimization for 5,000 steps, followed by 1 nanosecond (ns) heating with protein backbone sequential release of alpha carbon restraints (force constants were gradually reduced from 10 kcal/mol $Å^2$ to 0 kcal/mol $Å^2$), 4 ns of equilibration and 10 ns production simulation for each protein was performed. The PME method was used for full long-range electrostatics within a relative tolerance of 1×10^{-6}. A 12 Å cut-off was used to compute non-bonded interactions with a smooth switching function applied at a distance of 10 Å. To impose the thermal exchange with an external thermostat, the isobaric-isothermal ensemble (NPT) with constant number of particles N, pressure P and temperature T was used. Constant temperature was maintained by coupling the system to a thermal bath whose temperature is maintained via Langevin dynamics with a friction coefficient of 1 ps^{-1}. Constant pressure was maintained using a Langevin piston at a nominal value of 1 atm [35]. The SHAKE algorithm, with a tolerance of $1x10^{-8}$ Å, was applied to constrain the length of all covalent bonds involving hydrogen, thus allowing the use of a 2 fs integration time. Trajectory analyses and measurements were performed using VMD v1.9.3 [29]. By plotting Cα-root-mean-squared (RMSD) deviation along the molecular dynamic simulation, we assessed the structural equilibration reached by the models and the distances between the terminal heavy atom (Cz in Arg80 or O in the case of Ser80) and the nearest oxygen on the phosphate backbone. Figures were rendered using Pymol v2.3.2 (Schrodinger LLC., https://www.schrodinger.com).

D. Transient gene transcription assays

Expression vectors containing the NR5A1/SF-1 p.G22D or p.L420P missense variants were generated by site-directed mutagenesis (QuikChange, Stratagene, Amsterdam, The Netherlands) using a wild-type (WT) pCMX-NR5A1 template and validated by direct sequencing. Transient transfection studies were performed using lipofectamine 2000 (Invitrogen, Paisley, UK) in 96-well plates and a dual-luciferase reporter assay system (Promega, Madison, WI), as reported previously [22]. Studies were performed in tsa201 human embryonic kidney cells by transfecting empty, WT, or mutant NR5A1/SF-1 expression vectors (2ng/well; p.G22D, p.L420P) with the SF-1-responsive minimal promoter of Cyp11a linked to luciferase (100ng/well). 24h following transfection, cells were lysed and luciferase activity was assayed (Dual Luciferase Reporter Assay system, Promega; FLUOstar Optima, BMG Labtech, Aylesbury, UK), with standardization for Renilla co-expression. Results are shown as the mean±SEM of three independent experiments, each performed in triplicate.

E. Immunohistochemistry

Immunohistochemistry (IHC) for desert hedgehog (DHH) was undertaken in human fetal testis (9 weeks post conception) with ethical approval (REC reference 08/H0712/34+5) and informed consent in collaboration with the Human Developmental Biology Resource (HDBR, www.hdbr.org). In brief, 12µm sections were fixed briefly in 4% PFA in TBS, rinsed in TBS and blocked in 1% BSA in TBS-Tween (0.5% Tween) before incubating overnight with mouse monoclonal anti-human DHH antibody (Santa Cruz Biotechnology, Inc., CA, sc-133244, 1:50 dilution)[36] and rabbit anti-human anti-Mullerian hormone (AMH) antibody (AbCam, Cambridge, UK, ab-103233; 1:200 dilution) [37]. Sections were washed with TBS-Tween (0.5% Tween) then incubated for 1 hr with Alexa488 goat anti-mouse (Invitrogen, A11001; 1:400) [38] and Alexa555 goat anti-rabbit (Invitrogen, Waltham, MA, A21429; 1:400) [39], respectively and counterstained with DAPI (Sigma Aldrich, St Louis, MO). Images were captured on a Zeiss LSM 710 confocal microscope (Carl Zeiss, Oberkochen, Germany) and analyzed using Image J (NIH, Bethesda, MD).

2. Results

Overall a likely pathogenic variant was found in 16 individuals in the cohort (16/52, 30.8%).

A. Complete gonadal dysgenesis

A molecular etiology was attained in 6/27 (22.2%) individuals with a clinical diagnosis of CGD (Figs. 1 and 2; Table 2)

Three individuals had hemizygous pathogenic variants in SRY (sex-determining region Y), with two changes affecting codons in the high mobility group (HMG) box (p.R62P, p.N65D) and one variant being a novel, complex insertion-deletion affecting the native stop codon of SRY (Table 2, Fig. 3A and B).

One individual with CGD was found to have a heterozygous p.R80S variant in the first zinc finger of DMRT1 (Table 2, Fig. 4A). This variant is predicted to be damaging and was not found in the gnomAD database (Table 2). To gain insight into the potential effect this variant has on DMRT1, detailed molecular dynamics simulation of both WT DMRT and the p.R80S mutant was performed. The R80 residue interfaces with DNA, binding through a hydrogen-bond (H-bond) interaction with the phosphate backbone of thymidine 16 (Fig. 4B and C). The p.R80S mutant loses H-bond contacts with this phosphate backbone of DNA, and displays longer distances compared to wild-type (7.63 ± 0.36 v/s 4.34 ± 0.18 Å) during simulation time (Fig. 4C and D). The electrostatic potential of the surface does not change significantly (Fig. 4C).

One woman diagnosed with CGD was found to harbor a heterozygotic p.A280E variant in helix 3 of the ligand-binding domain (LBD) of steroidogenic factor-1 (NR5A1) (Table 2, Fig. 5A).

Another woman with primary amenorrhea and absent uterus was found to have a homozygous p.F242L variant in desert hedgehog (DHH) (Table 2, Fig. 6A).

B. Partially-virilised 46,XY DSD (pvDSD)

A molecular diagnosis was attained in a greater proportion of individuals with a clinical phenotype of pvDSD (10/25; 40.0%) compared to CGD (Figs. 1 and 2; Table 2).

Five pathogenic variants, all in the heterozygous state, were found in *NR5A1* (p.G22D, p.R281C, p.G328R, p.E367Sfs*15, p.L420P) (Table 2, Fig. 5A). These include four missense changes and a frameshift change affecting codons in the LBD. Transient transfection studies confirmed that two of these NR5A1 missense variants (p.G22D and p.L420P) likely cause loss-of-function (Fig. 5B), whereas the other two variants affect codons previously shown to be disrupted in individuals with DSD.

One woman with partial virilisation and absent Müllerian structures was found to have compound heterozygous (p.A227V/p.R245P) variants in DHH, affecting a fairly localized region within the carboxy (C)-terminal domain close to the p.F242 codon described above (Table 2, Fig. 6A). Using IHC we showed strong expression of DHH in interstitial cells of the testis just after the establishment of androgen biosynthesis in the human fetus (9 wpc), with weaker expression in the Sertoli cells of primitive seminiferous tubules (Fig. 6B).

Finally, four individuals had pathogenic heterozygous variants in DHX37 (DEAH-Box Helicase 37). Three of these affect a p.R308Q hotspot, and these have recently been reported separately (Table 2, Fig. 7) [40]. One additional variant (p.T477M) was identified, which affects a highly conserved amino acid in the Rec-A2 motif IV RNA-binding region and is predicted to be damaging and disease causing.

C. Candidate genes for DSD

Analysis of candidate genes for DSD based on our experience of transcriptomic profiling did not reveal any clear pathogenic variants in those genes analyzed (Table 1). No likely pathogenic changes were found in the three sibling pairs studied.

Discussion

Reaching a specific diagnosis in 46,XY DSD can be challenging as there are many different potential causes and historically, these may have been grouped under “umbrella” terms such as Swyer syndrome (CGD) or pvDSD [41].

Whilst many of these patients have classic conditions such as 17β-HSD, 5ARD and AIS, individuals may not have been fully investigated initially, and making a diagnosis can be difficult if the testes have been removed [5,6]. Reaching a specific diagnosis can sometimes have benefits for understanding the natural history of a condition, identifying associated features, defining likely inheritance and chances of other family members being affected, and in the long-term for understanding tumor risk[8]. Therefore, a molecular approach to diagnosis is becoming increasingly important in both research and clinical settings [4,8,20].

Here, we report our analysis of a large cohort of 46,XY women with DSD from a single center. Using targeted genetic analysis and biochemical profiling (e.g. urine steroid profiling by GC-MS-MS), those with 17β-HSD, 5ARD and AIS had been previously screened out, leaving a cohort of more than 50 women who were recruited with a working diagnosis of CGD or pvDSD [5,6]. This is a typical scenario in clinical practice and will be increasingly common as adult DSD services and teams are established to provide better long-term follow up and support in a multidisciplinary setting [42-44].

Using a NGS targeted sequencing approach of known and candidate genes, we were able to reach a likely molecular diagnosis in 30.8% of the cohort as a whole, with a higher proportion of genetic diagnoses being reached in the pvDSD group (40.0%) compared to the CGD group (22.2%). This is similar to the other limited data available, but it is important to note that many studies to date have included individuals with conditions such as 17β-HSD, 5ARD and AIS, so have been less stringent than this current study focusing on a cohort of adults with undiagnosed DSD.

Within our study cohort we found likely pathogenic variants in five different genes (SRY, DMRT1, NR5A1, DHH, DHX37), with some overlap in presentation between the two broad diagnostic groups. Our findings also provide insight into molecular aspects of these conditions.

As expected, hemizygous variants in SRY were found in three women with CGD who had presented with absent puberty and primary amenorrhea in adolescence. The prevalence of 11.5% of CGD is consistent with previous data and as expected, the two missense variants (p.R62P, p.N65D) affect amino acids in the HMG-box of SRY, potentially in the region of a nuclear localization signal or calmodulin-binding motif [45,46]. Of interest, a third individual had a complex insertion-deletion event that removes the native stop codon of SRY and is predicted to result in the translation of a protein with an additional seven amino acids at the carboxyl-terminal end (p.L204fs*211 p.L204PLDKANG*). This alteration was not defined clearly on NGS sequencing and a Sanger sequencing approach was needed to clarify the exact changes (Fig. 3B).

One individual with CGD was found to harbor a heterozygous p.R80S variant in the first zinc finger of DMRT1. DMRT1 has been known for several years to play a key role in sex-determination in different species, but its role in human sex development beyond 10q23 deletion syndrome is less clear [47]. More recently, a variant in DMRT1 (p. R111G) has been reported in association with testicular dysgenesis, that affects the DMRT1 recognition helix [28]. Using a similar approach, it is likely that the p.R80S variant affects interactions between DMRT1 and the minor groove of DNA (Fig. 4 B-D). As well as being located in a key domain, this variant is highly conserved amongst species, is predicted to be damaging and is not present in population databases. Using molecular modeling and simulation, we have shown that substitution of the larger, polar arginine residue with the smaller serine disrupts a H-bond between DMRT1 and DNA. It is expected that an additive effect on the loss of affinity that will depend on the stoichiometry of the DMRT1-DNA complex occurs, resulting in the final phenotypic effect. Of note, Wang et al. recently reported a p.Y84C variant in DMRT1 associated with DSD, but concluded the significance was unknown [14].

Six individuals in our cohort were found to have heterozygous missense (n=5) or frameshift (n=1) changes in steroidogenic factor-1 (NR5A1). SF-1/NR5A1 is a nuclear receptor transcription factor that was originally shown to be important in adrenal and gonad development [48]. However, it has emerged in the past 15 years that variants in SF-1/NR5A1 are one of the most common causes of 46,XY DSD, with phenotypes ranging from a complete gonadal dysgenesis scenario through various degrees of virilization and hypospadias to male factor infertility [10,11,22,49]. Most changes occur *de novo* or in a sex-limited dominant pattern, although primary ovarian insufficiency (POI) can also occur [50].

The prevalence of SF-1/NR5A1 variants in our cohort (6/52, 11.5%) is similar to other studies, with an enrichment seen in the pvDSD sub-group (19.2%) [22]. One of the probands had a family history consistent with sex-limited dominant inheritance. One of the variants discovered (p.G22D) affects the DNA-binding domain (DBD) of SF-1/NR5A1 and was shown to impair SF-1 function, as did the p.L420P variant in the LBD (Fig. 5B). The other missense variants identified (p.A280E, p.R281C, p.G328R) have either been reported previously or affect previously disrupted codons, and are all predicted to be damaging [10,51-53]. Although it is not clear whether SF-1 has a true biological ligand, crystallization has shown that codons 280/281 form a key region of helix 3 that interacts with co-repressors such as NR0B1/NR0B2 (DAX-1, SHP) [54]. Our findings confirm that variants in SF-1/NR5A1 are still the most prevalent cause of 46,XY DSD currently known.

Another key gene emerging as a relatively prevalent cause of DSD is DHH [55-61]. We identified DHH changes in two individuals with a recessive form of DSD who had homozygous (p.F242L) or compound heterozygous (p.A227V/p.R245P) variants, and with differing phenotypes. Disruption of DHH is sometimes associated with minifascicular neuropathy and was thought potentially to affect testicular interstitial/Leydig cell development in mice and humans, as well as Leydig-Sertoli cell interactions [56,59,62]. Furthermore, we have shown by IHC that DHH is expressed predominantly in developing Leydig cells in the human fetal testis at a critical early stage of development just after the onset of steroidogenesis (9 weeks post-conception) (Fig, 6B) [21]. However, DHH may also play a role in Sertoli cells and Leydig-Sertoli interactions, as areas of gonadal dysgenesis or streak gonads have been reported and Dhh is expressed in mouse Sertoli cells in single-cell RNA sequencing, and DHH expression was seen in Sertoli cells, albeit at a lower level (Fig. 6B) [56,59,63,64]. Consistent with this variability, one woman had a CGD phenotype with functional uterus and streak gonads, whereas the other had partially virilised genitalia and no Müllerian structures. The variants identified are located within a fairly localized region of DHH, but relatively little is known about the structural or biological effects of these changes.

Finally, four individuals (7.7% overall; 15.4% of pvDSD) in our cohort had heterozygous disruptive variants in DHX37. DHX37 is an RNA-helicase and predicted ribosomal RNA binding protein, but its exact biological function in testis development is still unclear. Variants in DHX37 are emerging as a relatively prevalent cause of a range of DSD phenotypes, including vanishing testis syndrome [40,65]. Three individuals in our cohort were found to have a recurrent p.R308Q variant in the RecA1 motif, and have been included in a recent series describing the role of DHX37 in 46,XY DSD [40]. The p.R308Q variant is often *de novo* and has been found in diverse ancestral backgrounds. One additional person in our cohort with mild virilization and absent uterus was found to have a p.T477M variant located within the RecA2 motif IV involved in RNA binding (Fig. 7). This variant occurs in a highly conserved codon/region and is a predicted disruptive event. Future studies will be needed to elucidate the role of DHX37 in sex development but the association of variants in this gene with 46,XY DSD is becoming well established.

Whilst we did not expect to discover variants involved in proximal steroidogenic blocks with adrenal or biochemical phenotypes (eg STAR, CYP11A1, HSD3B2, CYP17A1, POR, CYB5A) or associated with specific features (eg SOX9, campomelic dysplasia; GATA4/ZFPM2, cardiac; WT1, renal), we did not find likely-pathogenic variants in other DSD-associated genes such as MAP3K1, SOX8, ESR2 or ZNRF3. Furthermore, no clear pathogenic variants were found in candidate genes based on our transcriptomic studies of genes expressed in early human testis development [21]. In these studies, we used a modeling approach to identify genes that were either upregulated in a time-series dataset with similar patterns to SOX9 (e.g CITED1), or that were differentially expressed fetal testis genes or potential novel components of steroidogenesis. While it might be expected to find variants in some of these candidate factors, in fact no clear causative defects were identified.

This study has certain limitations. Firstly, a panel based approach means that this study is more effective at identifying the relative prevalence of changes in known genes in this cohort. Whole exome and whole genome sequencing approaches are better for gene discovery. As costs reduce, analysis becomes more robust and more accurate counseling can be provided about the risk of incidental findings, these technologies are becoming the first-line strategy in many situations, especially where multiple family members are affect. It remains possible that non-genomic events such as methylation defects or complex gene-environment interactions may cause some forms of DSD, especially complete testicular dysgenesis where the diagnostic rate is still relatively low. Secondly, functional assays are not well established for several of these factors, although we feel the genetic evidence for causality presented here is strong. Finally, studying a predominantly adult cohort means that access to historic data or family members is limited. However, it is increasingly being recognized that adult DSD services need to be established in parallel with multidisciplinary pediatric services, so this work provides useful insight into the range of diagnoses that might be made.

4. Conclusion

Next-generation sequencing approaches are improving the diagnostic yield in individuals with complete and partial forms of gonadal dysgenesis, or in those who have been labeled as having "partial androgen insensitivity" previously. Reaching a specific genetic diagnosis can inform genetic counseling, particularly for families with ongoing consanguinity, and can help to identify associated comorbiditites. Lastly, as we gain more information on the life course of adults with DSD, so we can seek correlations between genotype and phenotype especially with regard to late onset features such as osteoporosis.

Appendix Web Resources

Mutation Taster: http://www.mutationtaster.org/

PolyPhen-2: http://genetics.bwh.harvard.edu/pph2/

Sorting Intolerant from Tolerant (SIFT): https://sift.bii.a-star.edu.sg

The Genome Aggregation Database: http://gnomad.broadinstitute.org/; accessed March 2019

Ingenuity variant analysis: http://www.ingenuity.com/

UniProt: https://www.uniprot.org

Acknowledgements

We are grateful to those subjects who were involved in this study, and to Dianne Gerrelli and the Human Developmental Biology Resource.

Corresponding author: John C. Achermann, MD PhD, Genetics & Genomic Medicine, UCL Great Ormond Street Institute of Child Health, 30 Guilford Street, London WC1N 1EH, United Kingdom. E-mail: j.achermann@ucl.ac.uk

Funding information: J.C.A. is a Wellcome Trust Senior Research Fellow in Clinical Science (grants 098513/Z/12/Z and 209328/Z/17/Z) with research support from Great Ormond Street Hospital Children's Charity (grant V2518) and the National Institute for Health Research, Great Ormond Street Hospital Biomedical Research Centre (grant IS-BRC-1215-20012). The views expressed are those of the authors and not necessarily those of the National Health Service, National Institute for Health Research, or Department of Health. The human fetal material was provided by the Joint MRC/Wellcome Trust (grant# MR/R006237/1) Human

Developmental Biology Resource (http://www.hdbr.org). C.F.L was partly supported by the supercomputing infrastructure of the National Laboratory for High Performance Computing (NLHPC) (ECM-02).

Wellcome Trust http://dx.doi.org/10.13039/100004440, 098513/Z/12/Z, John C Achermann; Wellcome Trust http://dx.doi.org/10.13039/100004440, 209328/Z/17/Z, John C Achermann; Great Ormond Street Hospital Charity http://dx.doi.org/10.13039/501100001279, V2518, John C Achermann; National Institute for Health Research http://dx.doi.org/10.13039/501100000272, IS-BRC-1215-20012, John C Achermann; Medical Research Council/Wellcome Trust, MR/R006237/1, Not Applicable; National Laboratory for High Performance Computing, Chile, ECM-02, Carlos F. Lagos

Disclosure Summary:
The authors have nothing to disclose.

Data Availability
The datasets generated during and/or analyzed during the current study are not publicly available but are available from the corresponding author on reasonable request.

Data availability
Restrictions apply to the availability of data generated or analyzed during this study to preserve patient confidentiality or because they were used under license. The corresponding author will on request detail the restrictions and any conditions under which access to some data may be provided.

References

1. Sinclair AH, Berta P, Palmer MS, Hawkins JR, Griffiths BL, Smith MJ, Foster JW, Frischauf AM, Lovell-Badge R, Goodfellow PN. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature*. 1990;346(6281):240-244.
2. Berta P, Hawkins JB, Sinclair AH, Taylor A, Griffiths BL, Goodfellow PN, Fellous M. Genetic evidence equating SRY and the testis-determining factor. *Nature*. 1990;348(6300):448-450.
3. Koopman P, Münsterberg A, Capel B, Vivian N, Lovell-Badge R. Expression of a candidate sex-determining gene during mouse testis differentiation. *Nature*. 1990;348(6300):450-452.
4. Audí L, Ahmed SF, Krone N, Cools M, McElreavey K, Holterhus PM, Greenfield A, Bashamboo A, Hiort O, Wudy SA, McGowan R; The EU COST Action. GENETICS IN ENDOCRINOLOGY: Approaches to molecular genetic diagnosis in the management of differences/disorders of sex development (DSD): position paper of EU COST Action BM 1303 'DSDnet'. *Eur J Endocrinol*. 2018;179(4):R197--R206.
5. Phelan N, Williams EL, Cardamone S, Lee M, Creighton SM, Rumsby G, Conway GS. Screening for mutations in 17b-hydroxysteroid dehydrogenase and androgen receptor in women presenting with partially virilised 46,XY disorders of sex development. *Eur J Endocrinol*. 2015;172(6):745-751.
6. Berra M, Williams EL, Muroni B, Creighton SM, Honour JW, Rumsby G, Conway GS. Recognition of 5α-reductase-2 deficiency in an adult female 46XY DSD clinic. *Eur J Endocrinol*. 2011;164(6):1019-1025.
7. Hughes IA, Houk C, Ahmed SF, Lee PA; LWPES Consensus Group; ESPE Consensus Group. Consensus statement on management of intersex disorders. *Arch Dis Child*. 2006 Jul;91(7):554-563.

8. Achermann JC, Domenice S, Bachega TA, Nishi MY, Mendonca BB. Disorders of sex development: Effect of molecular diagnostics. *Nat Rev Endocrinol*. 2015;11(8):478-488.

9. Ahmed SF, Achermann JC, Arlt W, Balen A, Conway G, Edwards Z, Elford S, Hughes IA, Izatt L, Krone N, Miles H, O'Toole S, Perry L, Sanders C, Simmonds M, Watt A, Willis D. Society for Endocrinology UK guidance on the initial evaluation of an infant or an adolescent with a suspected disorder of sex development (Revised 2015). *Clin Endocrinol (Oxf)*. 2016;84(5):771-788.

10. Suntharalingham JP, Buonocore F, Duncan AJ, Achermann JC. DAX-1 (NR0B1) and steroidogenic factor-1 (SF-1, NR5A1) in human disease. *Best Pract Res Clin Endocrinol Metab*. 2015;29(4):607-619.

11. Domenice S, Zamboni Machado A, Moraes Ferreira F, Ferraz-de-Souza B, Marcondes Lerario A, Lin L, Yumie Nishi M, Lisboa Gomes N, Evelin da Silva T, Barbosa Silva R, Vieira Correa R, Ribeiro Montenegro L, Narciso A, Maria Frade Costa E, Achermann JC, Bilharinho Mendonca B. Wide spectrum of NR5A1-related phenotypes in 46,XY and 46,XX individuals. *Birth Defects Res Part C - Embryo Today Rev*. 2016;108(4):309-320.

12. De Franco E, Flanagan SE, Houghton JAL, Allen HL, MacKay DJG, Temple IK, Ellard S, Hattersley AT. The effect of early, comprehensive genomic testing on clinical care in neonatal diabetes: An international cohort study. *Lancet*. 2015;386(9997):957-963.

13. Guran T, Buonocore F, Saka N, Ozbek MN, Aycan Z, Bereket A, Bas F, Darcan S, Bideci A, Guven A, Demir K, Akinci A, Buyukinan M, Aydin BK, Turan S, Agladioglu SY, Atay Z, Abali ZY, Tarim O, Catli G, Yuksel B, Akcay T, Yildiz M, Ozen S, Doger E, Demirbilek H, Ucar A, Isik E, Ozhan B, Bolu S, Ozgen IT, Suntharalingham JP, Achermann JC. Rare causes of primary adrenal insufficiency: Genetic and clinical characterization of a large nationwide cohort. *J Clin Endocrinol Metab*. 2016;101(1):284-292.

14. Wang H, Zhang L, Wang N, Zhu H, Han B, Sun F, Yao H, Zhang Q, Zhu W, Cheng T, Cheng K, Liu Y, Zhao S, Song H, Qiao J. Next-generation sequencing reveals genetic landscape in 46, XY disorders of sexual development patients with variable phenotypes. *Hum Genet*. 2018;137(3):265-277.

15. Eggers S, Sadedin S, van den Bergen JA, Robevska G, Ohnesorg T, Hewitt J, Lambeth L, Bouty A, Knarston IM, Tan TY, Cameron F, Werther G, Hutson J, O'Connell M, Grover SR, Heloury Y, Zacharin M, Bergman P, Kimber C, Brown J, Webb N, Hunter MF, Srinivasan S, Titmuss A, Verge CF, Mowat D, Smith G, Smith J, Ewans L, Shalhoub C, Crock P, Cowell C, Leong GM, Ono M, Lafferty AR, Huynh T, Visser U, Choong CS, McKenzie F, Pachter N, Thompson EM, Couper J, Baxendale A, Gecz J, Wheeler BJ, Jefferies C, MacKenzie K, Hofman P, Carter P, King RI, Krausz C, van Ravenswaaij-Arts CMA, Looijenga L, Drop S, Riedl S, Cools M, Dawson A, Juniarto AZ, Khadilkar V, Khadilkar A, Bhatia V, Dũng VC, Atta I, Raza J, thi Diem Chi N, Hao TK, Harley V, Koopman P, Warne G, Faradz S, Oshlack A, Ayers KL, Sinclair AH. Disorders of sex development: insights from targeted gene sequencing of a large international patient cohort. *Genome Biol*. 2016;17(1):243.

16. Hughes LA, McKay Bounford K, Webb E, Dasani P, Clokie S, Chandran H, McCarthy L, Mohamed Z, Kirk JMW, Krone N, Allen S, Cole T. Next generation sequencing (NGS) to improve the diagnosis and management of patients with disorders of sex development (DSD). *Endocr Connect*. 2019;8(2):100.

17. Kim JH, Kang E, Heo SH, Kim G-H, Jang J-H, Cho E-H, Lee BH, Yoo H-W, Choi J-H. Diagnostic yield of targeted gene panel sequencing to identify the genetic etiology of disorders of sex development. *Mol Cell Endocrinol*. 2017;444:19-25.

18. Dong Y, Yi Y, Yao H, Yang Z, Hu H, Liu J, Gao C, Zhang M, Zhou L, Asan, Yi X, Liang Z. Targeted next-generation sequencing identification of mutations in patients with disorders of sex development. *BMC Med Genet*. 2016;17(1):23.
19. Baxter RM, Arboleda VA, Lee H, Barseghyan H, Adam MP, Fechner PY, Bargman R, Keegan C, Travers S, Schelley S, Hudgins L, Mathew RP, Stalker HJ, Zori R, Gordon OK, Ramos-Platt L, Pawlikowska-Haddal A, Eskin A, Nelson SF, Délot E, Vilain E. Exome sequencing for the diagnosis of 46, XY disorders of sex development. *J Clin Endocrinol Metab*. 2015;100(2):E333--
20. Délot EC, Papp JC, Délot EC, Fox M, Grody W, Lee H, Papp JC, Vilain E, Keegan C, Ramsdell L, Green J, Barseghyan H, Dorrani N, Mohnach L, Pearson MA, Diaz J, Leon E, Hopkin RJ, Johnson J, Saal H, Amarillo I, Adam M, Sandberg DE, Vilain E. Genetics of disorders of sex development: The DSD-TRN experience. *Endocrinol Metab Clin North Am*. 2017;46(2):519-537.
21. del Valle I, Buonocore F, Duncan AJ, Lin L, Barenco M, Parnaik R, Shah S, Hubank M, Gerrelli D, Achermann JC. A genomic atlas of human adrenal and gonad development. *Wellcome Open Res*. 2017;2:25.
22. Lin L, Philibert P, Ferraz-de-Souza B, Kelberman D, Homfray T, Albanese A, Molini V, Sebire NJ, Einaudi S, Conway GS, Hughes IA, Jameson JL, Sultan C, Dattani MT, Achermann JC. Heterozygous missense mutations in steroidogenic factor 1 (SF1/Ad4BP, NR5A1) are associated with 46,XY disorders of sex development with normal adrenal function. *J Clin Endocrinol Metab*. 2007;92(3):991-999.
23. Robinson JT, Thorvaldsdóttir H, Wenger AM, Zehir A, Mesirov JP. Variant review with the Integrative Genomics Viewer. *Cancer Res*. 2017;77(21):e31-e34.
24. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, Gauthier LD, Brand H, Solomonson M, Watts NA, Rhodes D, Singer-Berk M, Seaby EG, Kosmicki JA, Walters RK, Tashman K, Farjoun Y, Banks E, Poterba T, Wang A, Seed C, Whiffin N, Chong JX, Samocha KE, Pierce-Hoffman E, Zappala Z, O'Donnell-Luria AH, Minikel EV, Weisburd B, Lek M, Ware JS, Vittal C, Armean IM, Bergelson L, Cibulskis K, Connolly KM, Covarrubias M, Donnelly S, Ferriera S, Gabriel S, Gentry J, Gupta N, Jeandet T, Kaplan D, Llanwarne C, Munshi R, Novod S, Petrillo N, Roazen D, Ruano-Rubio V, Saltzman A, Schleicher M, Soto J, Tibbetts K, Tolonen C, Wade G, Talkowski ME, Consortium TGAD, Neale BM, Daly MJ, MacArthur DG. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv*. January 2019:531210.
25. Ren J, Wen L, Gao X, Jin C, Xue Y, Yao X. DOG 1.0: illustrator of protein domain structures. *Cell Res*. 2009;19(2):271-273.
26. Webb B, Sali A. Protein structure modeling with MODELLER. In: *Methods in Molecular Biology (Clifton, N.J.)*. Vol 1137. ; 2014:1-15.
27. UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Res*. 2015;43(D1):D204--D212.
28. Murphy MW, Lee JK, Rojo S, Gearhart MD, Kurahashi K, Banerjee S, Loeuille GA, Bashamboo A, McElreavey K, Zarkower D, Aihara H, Bardwell VJ. An ancient protein-DNA interaction underlying metazoan sex determination. *Nat Struct Mol Biol*. 2015;22(6):442-451.
29. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. *J Mol Graph*. 1996;14(1):33-38.
30. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys*. 1983;79(2):926-935.
31. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. *J Chem Phys*. 1995;103(19):8577-8593.

32. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L, Schulten K. Scalable Molecular Dynamics with NAMD. *J Comput Chem*. 2005;26:1781-1802.
33. Best RB, Zhu X, Shim J, Lopes PEM, Mittal J, Feig M, MacKerell AD. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain χ(1) and χ(2) dihedral angles. *J Chem Theory Comput*. 2012;8(9):3257-3273.
34. Klauda JB, Venable RM, Freites JA, O'Connor JW, Tobias DJ, Mondragon-Ramirez C, Vorobyov I, MacKerell AD, Pastor RW. Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J Phys Chem B*. 2010;114(23):7830-7843.
35. Feller SE, Zhang Y, Pastor RW, Brooks BR. Constant pressure molecular dynamics simulation: The Langevin piston method. *J Chem Phys*. 1995;103(11):4613-4621.
36. RRID:AB_2091715
37. RRID:AB_10711946
38. RRID:AB_2534069
39. RRID:AB_141761
40. McElreavey K, Jorgensen A, Eozenou C, Merel T, Bignon-Topalovic J, Tan DS, Houzelstein D, Buonocore F, Warr N, Kay RGG, Peycelon M, Siffroi J-P, Mazen I, Achermann JC, Shcherbak Y, Leger J, Sallai A, Carel J-C, Martinerie L, Le Ru R, Conway GS, Mignot B, Van Maldergem L, Bertalan R, Globa E, Brauner R, Jauch R, Nef S, Greenfield A, Bashamboo A. Pathogenic variants in the DEAH-box RNA helicase DHX37 are a frequent cause of 46,XY gonadal dysgenesis and 46,XY testicular regression syndrome. *Genet Med*. 2019;in press. doi:10.1038/s41436-019-0606-y.
41. Minto CL, Crouch NS, Conway GS, Creighton SM. XY females: Revisiting the diagnosis. *BJOG*. 2005;112(10):1407-1410.
42. Cools M, Nordenström A, Robeva R, Hall J, Westerveld P, Flück C, Köhler B, Berra M, Springer A, Schweizer K, Pasterski V. Caring for individuals with a difference of sex development (DSD): A Consensus Statement. *Nat Rev Endocrinol*. 2018;14(7):415-429.
43. Falhammar H, Claahsen-van der Grinten H, Reisch N, Slowikowska-Hilczer J, Nordenström A, Roehle R, Bouvattier C, Kreukels BPC, Köhler B; dsd-LIFE group. Health status in 1040 adults with disorders of sex development (DSD): a European multicenter study. *Endocr Connect*. 2018;7(3):466-478.
44. Rapp M, Mueller-Godeffroy E, Lee P, Roehle R, Kreukels BPC, Köhler B, Nordenström A, Bouvattier C, Thyen U. Multicentre cross-sectional clinical evaluation study about quality of life in adults with disorders/differences of sex development (DSD) compared to country specific reference populations (dsd-LIFE). *Health Qual Life Outcomes*. 2018;16(1):54.
45. Sim H, Argentaro A, Harley VR. Boys, girls and shuttling of SRY and SOX9. *Trends Endocrinol Metab*. 2008;19(6):213-222.
46. Sim H, Argentaro A, Czech DP, Bagheri-Fam S, Sinclair AH, Koopman P, Boizet-Bonhoure B, Poulat F, Harley VR. Inhibition of SRY-calmodulin complex formation induces ectopic expression of ovarian cell markers in developing XY gonads. *Endocrinology*. 2011;152(7):2883-2893.
47. Matson CK, Zarkower D. Sex and the singular DM domain: Insights into sexual regulation, evolution and plasticity. *Nat Rev Genet*. 2012;13(3):163-174.
48. Achermann JC, Ito M, Ito M, Hindmarsh PC, Jameson JL. A mutation in the gene encoding steroidogenic factor-1 causes XY sex reversal and adrenal failure in humans. *Nat Genet*. 1999;22(2):125-126.

49. McElreavey K, Achermann JC. Steroidogenic Factor-1 (SF-1, NR5A1) and 46,XX ovotesticular disorders of sex development: One factor, many phenotypes. *Horm Res Paediatr*. 2017;87(3):189-190.
50. Lourenço D, Brauner R, Lin L, De Perdigo A, Weryha G, Muresan M, Boudjenah R, Guerra-Junior G, Maciel-Guerra AT, Achermann JC, McElreavey K, Bashamboo A. Mutations in NR5A1 associated with ovarian insufficiency. *N Engl J Med*. 2009;360(12):1200-1210.
51. Philibert P, Polak M, Colmenares A, Lortat-Jacob S, Audran F, Poulat F, Sultan C. Predominant Sertoli cell deficiency in a 46,XY disorders of sex development patient with a new NR5A1/SF-1 mutation transmitted by his unaffected father. *Fertil Steril*. 2011;95(5):1788.e5-9.
52. Tantawy S, Lin L, Akkurt I, Borck G, Klingmüller D, Hauffa BP, Krude H, Biebermann H, Achermann JC, Köhler B. Testosterone production during puberty in two 46,XY patients with disorders of sex development and novel NR5A1 (SF-1) mutations. *Eur J Endocrinol*. 2012;167(1):125-130.
53. Rocca MS, Ortolano R, Menabò S, Baronio F, Cassio A, Russo G, Balsamo A, Ferlin A, Baldazzi L. Mutational and functional studies on NR5A1 gene in 46,XY disorders of sex development: identification of six novel loss of function mutations. *Fertil Steril*. 2018;109(6):1105-1113.
54. Krylova IN, Sablin EP, Moore J, Xu RX, Waitt GM, MacKay JA, Juzumiene D, Bynum JM, Madauss K, Montana V, Lebedeva L, Suzawa M, Williams JD, Williams SP, Guy RK, Thornton JW, Fletterick RJ, Willson TM, Ingraham HA. Structural analyses reveal phosphatidyl inositols as ligands for the NR5 orphan receptors SF-1 and LRH-1. *Cell*. 2005;120(3):343-355.
55. Umehara F, Tate G, Itoh K, Yamaguchi N, Douchi T, Mitsuya T, Osame M. A novel mutation of desert hedgehog in a patient with 46, XY partial gonadal dysgenesis accompanied by minifascicular neuropathy. *Am J Hum Genet*. 2000;67(5):1302-1305.
56. Werner R, Merz H, Birnbaum W, Marshall L, Schröder T, Reiz B, Kavran JM, Bäumer T, Capetian P, Hiort O. 46,XY gonadal dysgenesis due to a homozygous mutation in desert hedgehog (DHH) identified by exome sequencing. *J Clin Endocrinol Metab*. 2015;100(7):E1022--E1029.
57. Paris F, Flatters D, Caburet S, Legois B, Servant N, Lefebvre H, Sultan C, Veitia RA. A novel variant of DHH in a familial case of 46,XY disorder of sex development: Insights from molecular dynamics simulations. *Clin Endocrinol (Oxf)*. 2017;87(5):539-544.
58. Baldinotti F, Cavallaro T, Dati E, Baroncelli GI, Bertini V, Valetto A, Massart F, Fabrizi GM, Zanette G, Peroni D, Bertelloni S. Novel familial variant of the desert hedgehog gene: Clinical findings in two Sisters with 46,XY gonadal dysgenesis or 46,XX karyotype and literature review. *Horm Res Paediatr*. 2018;89(3):141-149.
59. Rothacker KM, Ayers KL, Tang D, Joshi K, van den Bergen JA, Robevska G, Samnakay N, Nagarajan L, Francis K, Sinclair AH, Choong CS. A novel, homozygous mutation in desert hedgehog (DHH) in a 46, XY patient with dysgenetic testes presenting with primary amenorrhoea: a case report. *Int J Pediatr Endocrinol*. 2018;2018(1):2.
60. Tajouri A, Kharrat M, Hizem S, Zaghdoudi H, M'rad R, Simic-Schleicher G, Kaiser FJ, Hiort O, Werner R. In vitro functional characterization of the novel DHH mutations p.(Asn337Lysfs*24) and p.(Glu212Lys) associated with gonadal dysgenesis. *Hum Mutat*. 2018;39(12):2097-2109.
61. Ayers K, Van Den Bergen J, Robevska G, Listyasari N, Raza J, Atta I, Riedl S, Rothacker K, Choong C, Faradz SMH, Sinclair A. Functional analysis of novel desert hedgehog gene variants improves the clinical interpretation of genomic data and provides a

more accurate diagnosis for patients with 46,XY differences of sex development. *J Med Genet*. 2019;56(7):434-443.
62. Yao HHC, Whoriskey W, Capel B. Desert Hedgehog/Patched 1 signaling specifies fetal Leydig cell fate in testis organogenesis. *Genes Dev*. 2002;16(11):1433-1440.
63. Canto P, Söderlund D, Reyes E, Méndez JP. Mutations in the *Desert hedgehog* (*DHH*) gene in patients with 46,XY complete pure gonadal dysgenesis. *J Clin Endocrinol Metab*. 2004;89(9):4480-4483.
64. Stévant I, Kühne F, Greenfield A, Chaboissier M-C, Dermitzakis ET, Nef S. Dissecting cell lineage specification and sex fate determination in gonadal somatic cells using single-cell transcriptomics. *Cell Rep*. 2019;26(12):3272--3283.e3.
65. Evilen da Silva T, Gomes NL, Lerário AM, Keegan CE, Nishi MY, Carvalho FM, Vilain E, Barseghyanm H, Martinez-Aguayo A, Forclaz MV, Papazian R, Pedroso de Paula LC, Costa EC, Carvalho LR, Jorge AA, Elias F, Mitchell R, Frade Costa EM, Mendonca BB, Domenice S. Genetic evidence of the association of DEAH-box helicase 37 defects with 46,XY gonadal dysgenesis spectrum. *J Clin Endocrinol Metab*. 2019:in press doi:10.1210/jc.2019-00984.

Figure 1. Overview of the study cohort and summary of findings. CGD, complete gonadal dysgenesis; pvDSD, partially-virilized DSD.

Figure 2. Overview of genetic diagnoses reached. CGD, complete gonadal dysgenesis; pvDSD, partially-virilized DSD.

Figure 3. A) Cartoon of SRY demonstrating the mutations identified, with amino acid conservancy shown below. The position of variants is indicated by red arrowheads. B) Aligned sequence and chromatogram showing the complex in-del variant. The position of the typical stop codon at 205 is shown with a blue arrow. The red asterix indicates where the frameshift ends. HMG, High Mobility Group

Figure 4. A) Cartoon of DMRT1 demonstrating the p.R80S mutation identified, with amino acid conservancy shown below. The position of p.R80 is indicated by a red arrowhead. B) Schematic representation of the DMRT1-DNA complex (PDB id 4YJ0). The DMRT1 protein is shown in white and the R80 residues are depicted as spheres, Zinc atoms are shown in magenta, C). Snapshot of the R80 or S80 vicinity (*upper*) and electrostatic potential of surface of the respective DMRT1-DNA complexes (*lower*). D) Cα-root-mean-squared (RMSD) deviation and distances between Arg or Ser 80 and the phosphate backbone of the DNA along the MD simulation.

Figure 5. A) Cartoon of NR5A1 (SF-1) demonstrating the mutations identified, with amino acid conservancy shown below. The positions of variants is indicated by red arrowheads. B) Transient transfection assays showing activation of a Cyp11a1 promoter by WT SF-1 and impared transcriptional activity by the p.G22D and p.L420P mutants. Results are shown as a percentage of WT SF-1 activity (relative light units). Data are shown as mean ± SEM of three experiments performed in triplicate. Zn, Zinc finger. WT, wild-type.

Figure 6. A) Cartoon of desert hedgehog (DHH) demonstrating the mutations identified, with amino acid conservancy shown below. The positions of variants is indicated by red arrowheads. B) Immunohistochemistry of human fetal testis at 9 weeks post-conception showing expression of DHH (green) predominantly in interstitial Leydig cells but also in Sertoli cells. AMH (red) is expressed strongly in Sertoli cells of primitive seminiferous tubules. Nuclei are stained blue with DAPI. No staining is seen in the peripheral capsular

region at the bottom right of the image. Scale bar, 100 μm NTD, N-terminal domain; CTD, C-terminal domain.

Figure 7. Cartoon of DHX37 demonstrating the mutations identified, with amino acid conservancy shown below. The positions of variants is indicated by red arrowheads. NTD, N-terminal domain; RecA1, ATP binding DEAH box Helicase; RecA2, C-terminal Helicase; HA2, Helicase Associated 2 domain; OB, oligonucleotide/oligosaccharide like domain; CTD, C-terminal domain.

Table 1. Summary of genes included on the targeted panel

Genes implicated in DSD (n=53)
SRY, SOX9, NR5A1, WT1, NR0B1, LHCGR, STAR, CYP11A1, HSD3B2, CYP17A1, POR, HSD17B3, SRD5A2, AR, CYP19A1, CDKN1C, SAMD9
CBX2, MAP3K1, GATA4, MAMLD1, HHAT, DHH, SOX8, CYB5A, AKR1C2, AKR1C4, SOX3, RSPO1, ARX, DHCR7, AMH, AMHR2, POU1F1, *DHX37, MYRF, DMRT1, FGF9, FGFR2, WWOX, ANOS1, CHD7, ESR2, FGF8, FGFR1, PROK2, PROKR2, STARD8, WDR11, ZFPM2, ZNRF3, PBX1 TSPYL1*
Other genes sequenced (n=127)
WNT4, CXCL12, MAP3K15, MAPK15, MAPK4, FOXO4, ABCA10, ABCA9, ACAT2, ADAMTS5, AGPAT9, ALAS1, ALDH1A1, AMIGO2, ANKRD18A, AOC3, APOA1, APOBEC3G, ARAP2, ASPN,, BNC1, BNC2, BRDT, CALB2, CARTPT, DHCR24, CD96, CD99, CILP, CITED1, CITED4, CNTNAP4, CORO2A, CTSV, CUL4B, EGFLAM, ELOVL2, ENTPD3, EPPIN, FATE1, FDX1, FDXR, RMPD1, G6PD, GLI4, GNRHR, GPD1L, GPX1, GRAMD1B, GSTA1, HIST1H2AA, HK2, HOXA10, HPGD, HPRT1, HRASLS5, HSD17B11, HSD17B6, HSD3B1, IDI1, IFI16, ILDR2, INHA, INHBB, INHBE, INSL3, ITGA9, JAM2, KEL, LDLR, MAGEB1, MANEA, MGARP, MOCOS, MRO, MSC, NANOG, NLRP7, NOSTRIN, NOTCH2, NPY, NPY1R, NRK, OPN3, OR10G9, OR4D5, OR8D4, PEG3, PHF24, PLEKHB1, PRND, PRPS2, PTCH1, PTCH2, RASSF2, REEP1, RELN, REN, RGAG1, RMDN2, SCARB1, SCUBE1, SERPINA5, SERPINF1, SESN1, SHH, SLAIN1, SLC16A9, SLC27A3, SLC52A3, SLC8B1, SMOC1, SOAT1, SORCS1, SRA1, SSFA2, STARD4, STK10, SUSD3, TCF21, TFAP2C, TMC5, TSPYL2, VCAM1, ZFAND5, ZNF280B, ZNF676

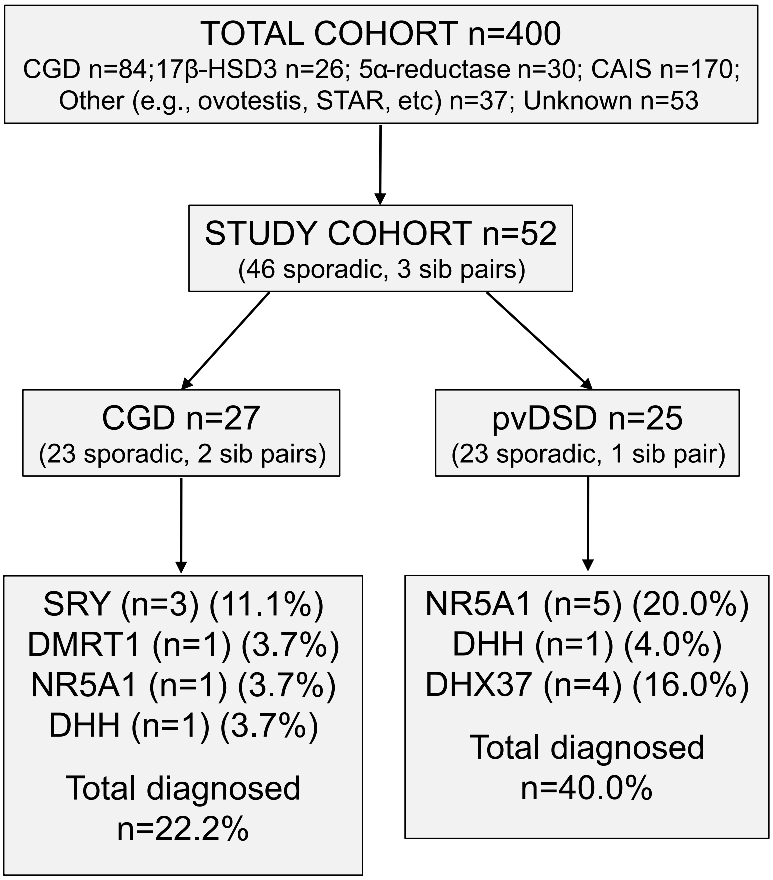
Underlined genes were sequenced using a Nonacus panel

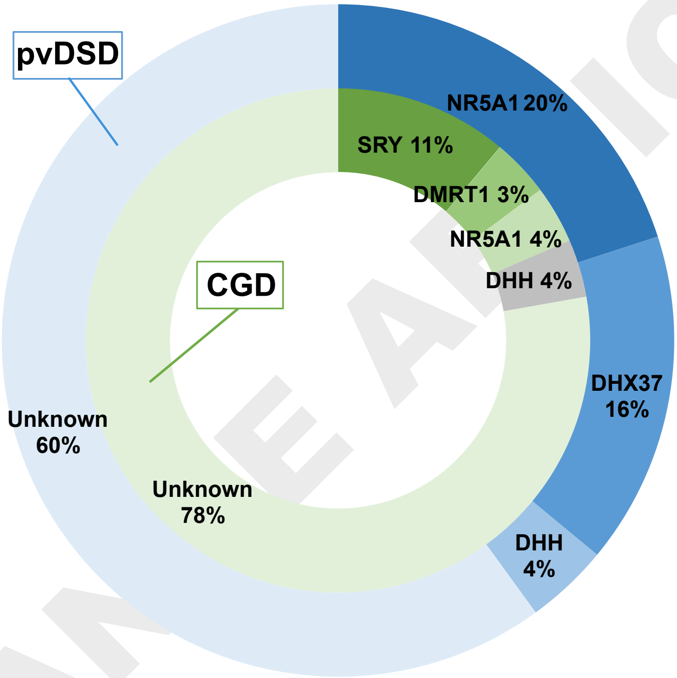
Table 2. Overview of clinical features and pathogenic variants identified in the study cohort.

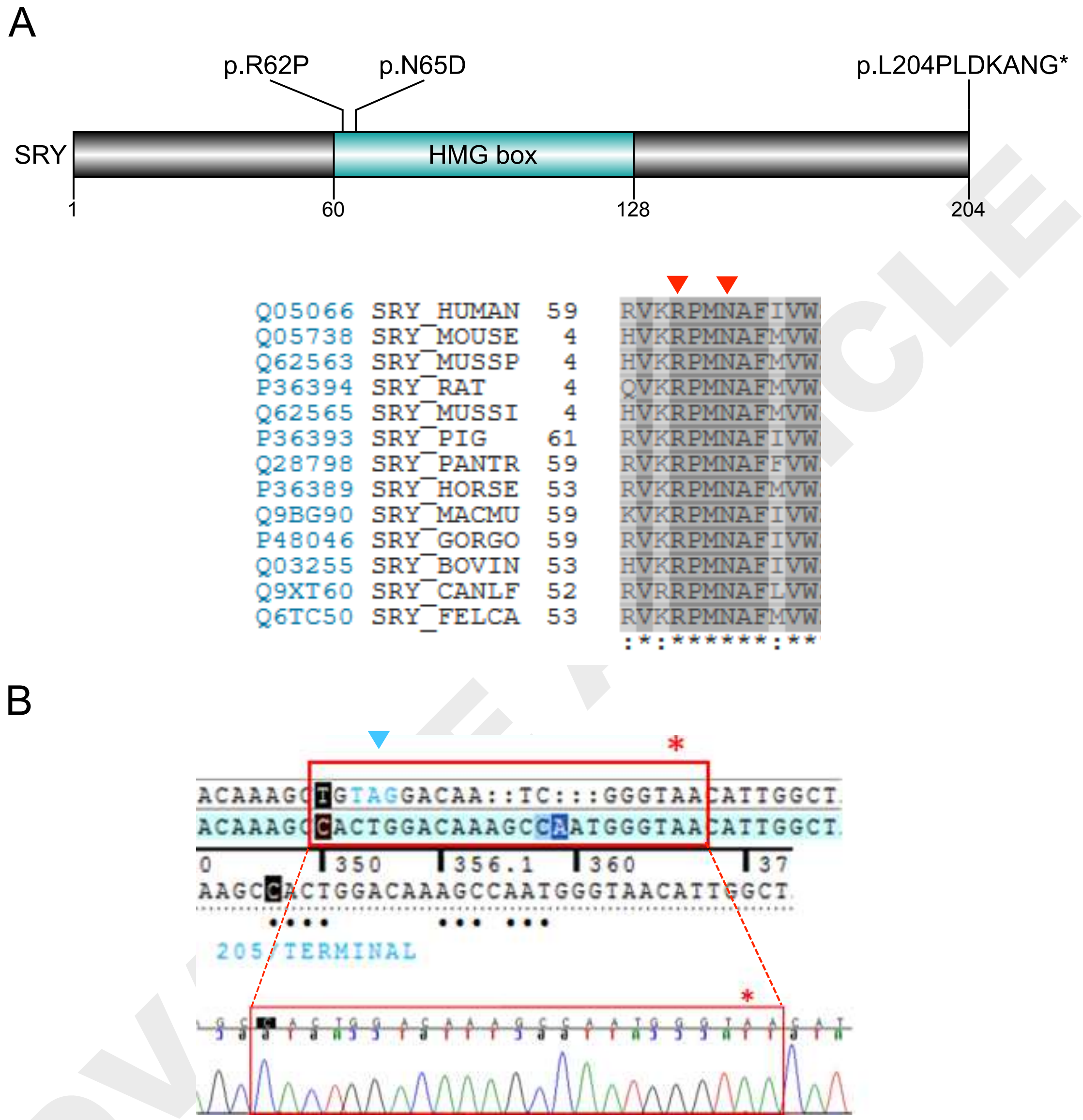
Working diagnosis	**Genital phenotype**	**Müllerian structures**	**Gene**	**Sequence Variation**	**Genotype**	**Inheritance**	**gnomAD allele frequency**	**SIFT**	**PolyPhen-2**	**Mutation taster**
CGD	Female	Present	SRY	p.R62P c.185G>C	Hemizygous	Sporadic	0	Damaging (0)	Probably damaging (1)	Disease causing
CGD	Female	Present	SRY	p.N65D c.193A>G	Hemizygous	Sporadic	0	Damaging (0)	Probably damaging (0.998)	Disease causing
CGD	Female	Present	SRY	p.L204fs*211 p.L204PLDKANG*	Hemizygous	Sporadic	0	N/A	N/A	N/A
CGD	Female	Present	DMRT1	p.R80S c.240G>C	Heterozygous	Sporadic	0	Damaging (0)	Probably damaging (0.997)	Disease causing
CGD	Female	Present	NR5A1	p.A280E c.839C>A	Heterozygous	Sporadic	0	Damaging (0)	Probably damaging (1)	Disease causing
CGD	Female	Present	DHH	P.F242L c.724T>C	Homozygous	Sporadic	0	Damaging (0.014)	Probably damaging (0.997)	Disease causing
pvDSD	Partially virilized	Absent	NR5A1[a]	p.G22D c.65G>A	Heterozygous	Sporadic	0	Damaging (0)	Probably damaging (1)	Disease causing

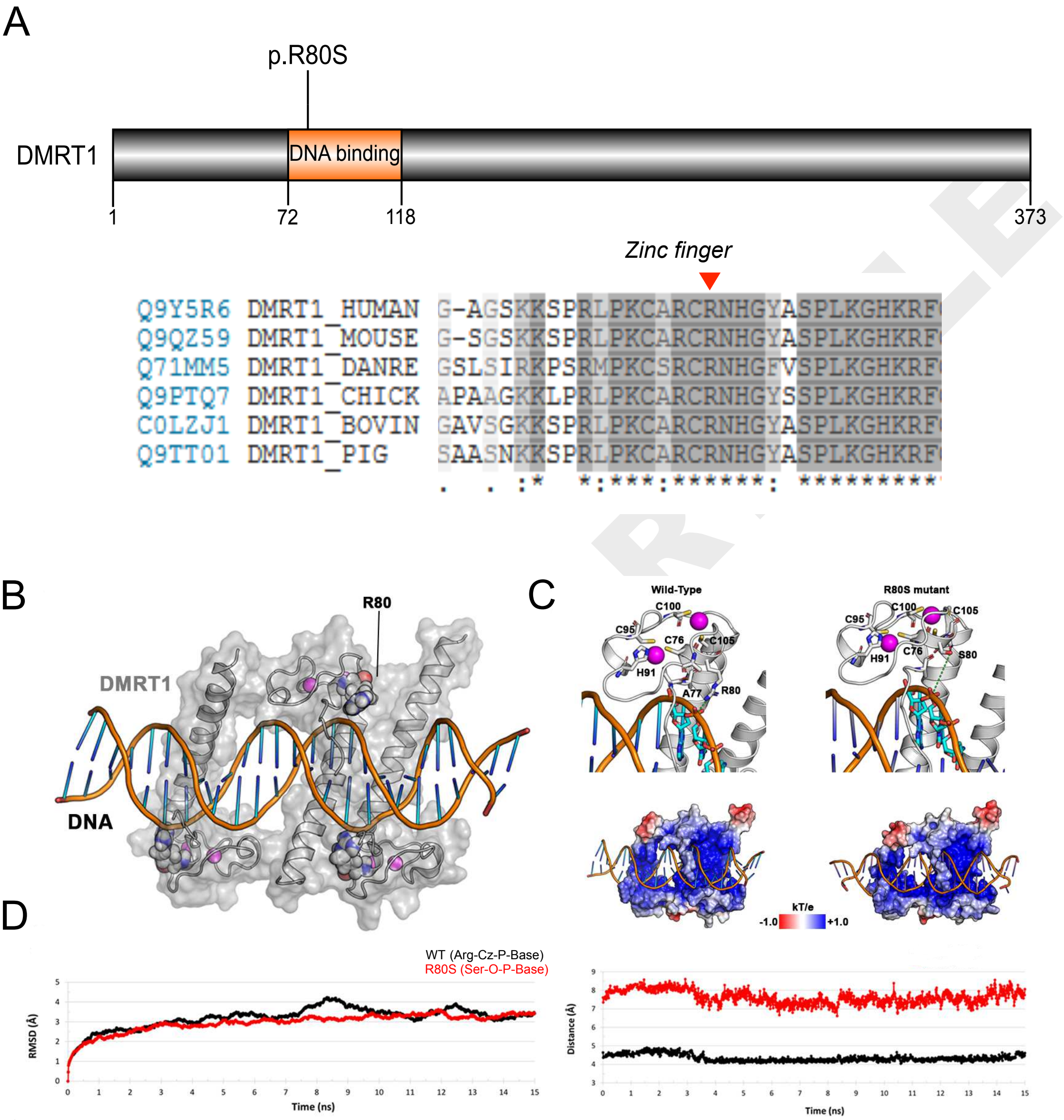
pvDSD	Partially virilized	Absent	NR5A1	p.R281C c.841C>T	Heterozygous	Sporadic	0	Damaging (0)	Probably damaging (1)	Disease causing
pvDSD	Partially virilized	Absent	NR5A1	p.G328R c.982G>C	Heterozygous	Sporadic	0	Damaging (0)	Probably damaging (1)	Disease causing
pvDSD	Partially virilized	Absent	NR5A1	p.E367Sfs*15 c.1099delG	Heterozygous	Sporadic	0	N/A	N/A	N/A
pvDSD	Virilization at puberty	Absent	NR5A1	p.L420P c.1259T>C	Heterozygous	Familial (SLD)	0	Damaging (0)	Probably damaging (1)	Disease causing
pvDSD	Partially virilized, further virilization at puberty	Absent	DHH	p.R245P c.734G>C	Compound Heterozygous	Sporadic	0	Damaging (0.002)	Probably damaging (0.999)	Disease causing
				p.A227V c.680C>T			2/30956	Tolerated (0.101)	Possibly damaging (0.904)	Disease causing
pvDSD	Partially virilized	Absent	DHX37[b]	p.R308Q c.923G>A	Heterozygous	Sporadic	1/30936	Damaging (0)	Probably damaging (1)	Disease causing
pvDSD	Partially virilized	Absent	DHX37[b]	p.R308Q c.923G>A	Heterozygous	Sporadic	1/30936	Damaging (0)	Probably damaging (1)	Disease causing
pvDSD	Partially virilized	Vaginal septum/uterine didelphys	DHX37[b]	p.R308Q c.923G>A	Heterozygous	Sporadic	1/30936	Damaging (0)	Probably damaging (1)	Disease causing
pvDSD	Partially virilized	Absent	DHX37	p.T477M c.1430C>T	Heterozygous	Sporadic	0	Damaging (0)	Probably damaging (1)	Disease causing

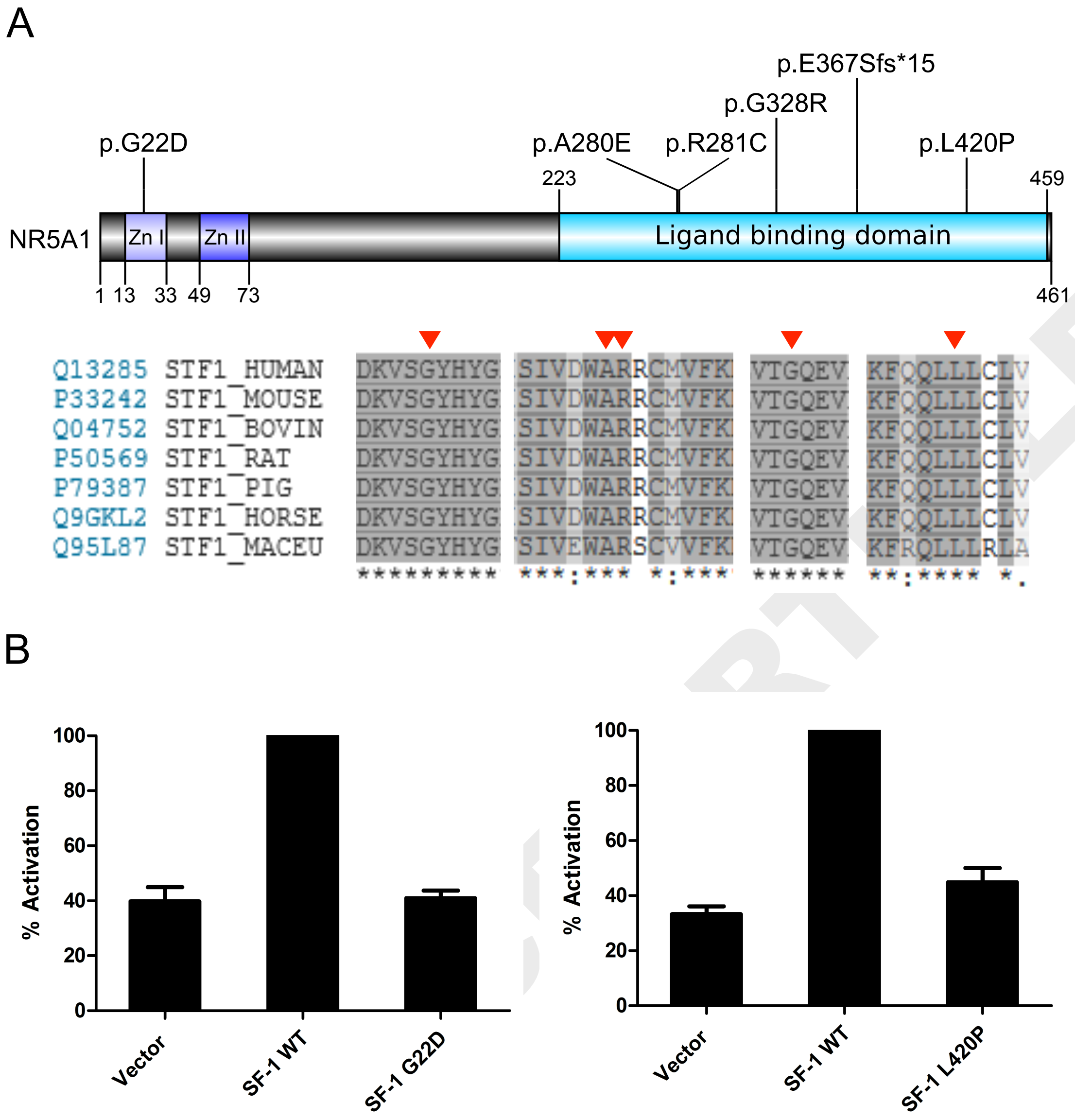
Abbreviations: N/A, not applicable; pvDSD, partially-virilized DSD; SLD, sex-limited dominant. [a] also heterozygous for a c.1305G>T, p.E435D variant that is predicted benign; [b] previously published [40].

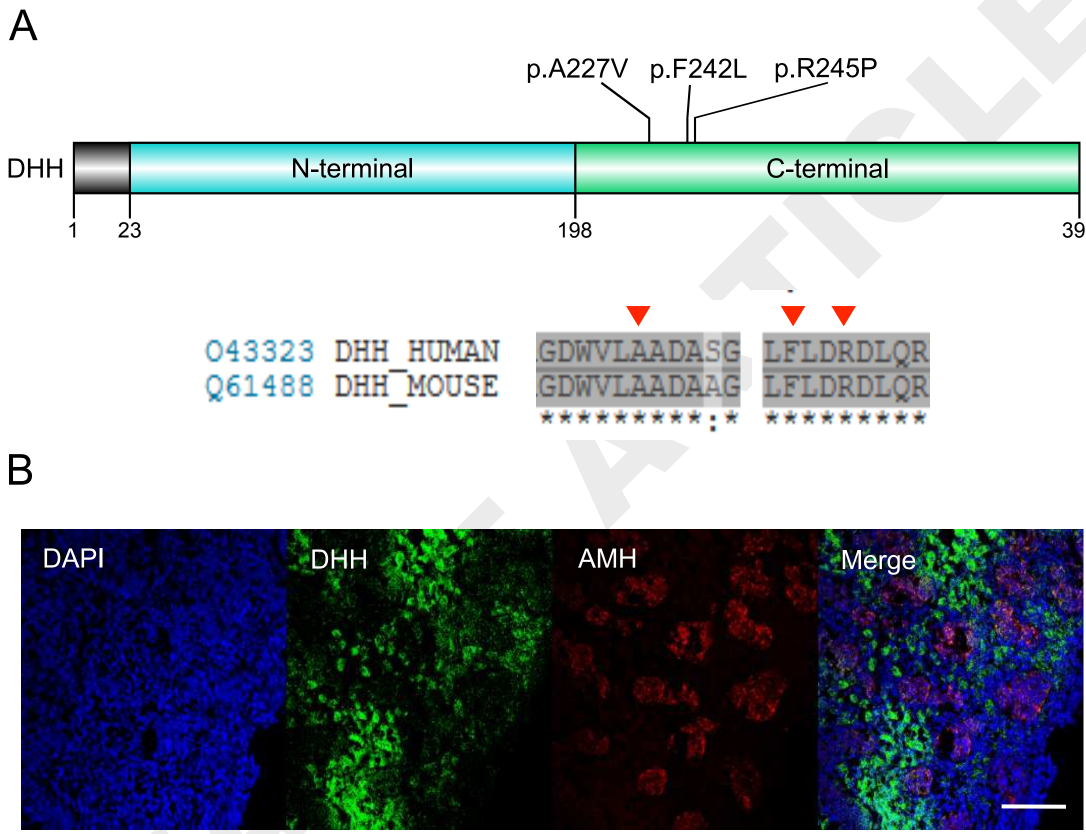

TOTAL COHORT n=400
CGD n=84;17β-HSD3 n=26; 5α-reductase n=30; CAIS n=170;
Other (e.g., ovotestis, STAR, etc) n=37; Unknown n=53
STUDY COHORT n=52
(46 sporadic, 3 sib pairs)
CGD n=27
(23 sporadic, 2 sib pairs)
pvDSD n=25
(23 sporadic, 1 sib pair)
SRY (n=3) (11.1%)
DMRT1 (n=1) (3.7%)
NR5A1 (n=1) (3.7%)
DHH (n=1) (3.7%)
Total diagnosed
n=22.2%
NR5A1 (n=5) (20.0%)
DHH (n=1) (4.0%)
DHX37 (n=4) (16.0%)
Total diagnosed
n=40.0%

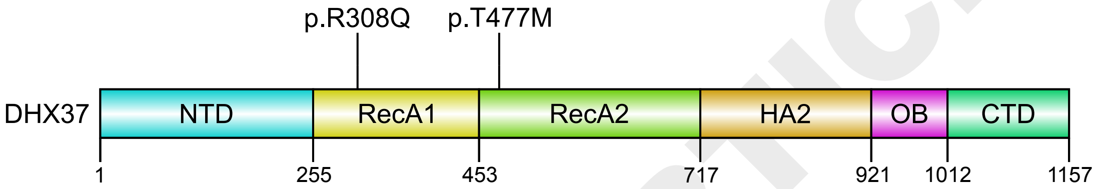

pvDSD
CGD
NR5A1 20%
SRY 11%
DMRT1 3%
NR5A1 4%
DHH 4%
DHX37 16%
DHH 4%
Unknown 60%
Unknown 78%

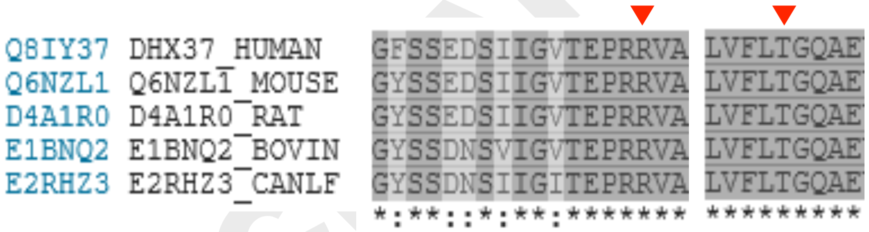

A
p.R62P
p.N65D
p.L204PLDKANG*
SRY
HMG box
1
60
128
204
Q05066 SRY_HUMAN 59 RVKRPMNAFIVW
Q05738 SRY_MOUSE 4 HVKRPMNAFMVW
Q62563 SRY_MUSSP 4 HVKRPMNAFMVW
P36394 SRY_RAT 4 QVKRPMNAFMVW
Q62565 SRY_MUSSI 4 HVKRPMNAFMVW
P36393 SRY_PIG 61 RVKRPMNAFIVW
Q28798 SRY_PANTR 59 RVKRPMNAFFVW
P36389 SRY_HORSE 53 RVKRPMNAFMVW
Q9BG90 SRY_MACMU 59 KVKRPMNAFIVW
P48046 SRY_GORGO 59 RVKRPMNAFIVW
Q03255 SRY_BOVIN 53 HVKRPMNAFIVW
Q9XT60 SRY_CANLF 52 RVRRPMNAFLVW
Q6TC50 SRY_FELCA 53 RVKRPMNAFMVW
:*:******:**
B
*
ACAAAGCTGTAGGACAA::TC:::GGGTAACATTGGCT
ACAAAGCCACTGGACAAAGCCAATGGGTAACATTGGCT
0
350
356.1
360
37
AAGCCACTGGACAAAGCCAATGGGTAACATTGGCT
205 TERMINAL
*


A
p.R80S
DMRT1
DNA binding
1
72
118
373
Zinc finger
Q9Y5R6 DMRT1_HUMAN G-AGSKKSPRLPKCARCRNHGYASPLKGHKRF
Q9QZ59 DMRT1_MOUSE G-SGSKKSPRLPKCARCRNHGYASPLKGHKRF
Q71MM5 DMRT1_DANRE GSLSIRKPSRMPKCSRCRNHGFVSPLKGHKRF
Q9PTQ7 DMRT1_CHICK APAAGKKLPRLPKCARCRNHGYSSPLKGHKRF
C0LZJ1 DMRT1_BOVIN GAVSGKKSPRLPKCARCRNHGYASPLKGHKRF
Q9TT01 DMRT1_PIG SAASNKKSPRLPKCARCRNHGYASPLKGHKRF
. . :* *:***:******: *********
B
R80
DMRT1
DNA
C
Wild-Type
C100
C95
C76
C105
H91
A77
R80
R80S mutant
C100
C105
C95
H91
C76
S80
kT/e
-1.0
+1.0
D
WT (Arg-Cz-P-Base)
R80S (Ser-O-P-Base)
RMSD (Å)
Time (ns)
Distance (Å)
Time (ns)


A
p.G22D
p.A280E
p.R281C
p.G328R
p.E367Sfs*15
p.L420P
223
459
NR5A1
Zn I
Zn II
Ligand binding domain
1
13
33
49
73
461
Q13285 STF1_HUMAN DKVSGYHYG SIVDWARRCMVFKI VTGQEV KFQQLLLCLV
P33242 STF1_MOUSE DKVSGYHYG SIVDWARRCMVFKI VTGQEV KFQQLLLCLV
Q04752 STF1_BOVIN DKVSGYHYG SIVDWARRCMVFKI VTGQEV KFQQLLLCLV
P50569 STF1_RAT DKVSGYHYG SIVDWARRCMVFKI VTGQEV KFQQLLLCLV
P79387 STF1_PIG DKVSGYHYG SIVDWARRCMVFKI VTGQEV KFQQLLLCLV
Q9GKL2 STF1_HORSE DKVSGYHYG SIVDWARRCMVFKI VTGQEV KFQQLLLCLV
Q95L87 STF1_MACEU DKVSGYHYG SIVEWARSCVVFKI VTGQEV KFRQLLLRLA
********* ***:*** *:*** ****** **:**** *.
B
% Activation
100
80
60
40
20
0
Vector
SF-1 WT
SF-1 G22D
% Activation
100
80
60
40
20
0
Vector
SF-1 WT
SF-1 L420P


A
p.A227V
p.F242L
p.R245P
DHH
N-terminal
C-terminal
1
23
198
396
O43323 DHH_HUMAN GDWVLAADASG LFLDRDLQR
Q61488 DHH_MOUSE GDWVLAADAAG LFLDRDLQR
**********:* *********
B
DAPI
DHH
AMH
Merge


p.R308Q
p.T477M
DHX37
NTD
RecA1
RecA2
HA2
OB
CTD
1
255
453
717
921
1012
1157


Q8IY37 DHX37_HUMAN GFSSEDSIIGVTEPRRVA LVFLTGQAE
Q6NZL1 Q6NZL1_MOUSE GYSSEDSIIGVTEPRRVA LVFLTGQAE
D4A1R0 D4A1R0_RAT GYSSEDSIIGVTEPRRVA LVFLTGQAE
E1BNQ2 E1BNQ2_BOVIN GYSSDNSVIGVTEPRRVA LVFLTGQAE
E2RHZ3 E2RHZ3_CANLF GYSSDNSIIGITEPRRVA LVFLTGQAE
*:**::*:**:******* *********