European Journal of Human Genetics (2018) 26:1329–1338
https://doi.org/10.1038/s41431-018-0202-7

ARTICLE

Broad phenotypes in heterozygous *NR5A1* 46,XY patients with a disorder of sex development: an oligogenic origin?

Núria Camats[1,2] · Mónica Fernández-Cancio[2] · Laura Audí[2] · André Schaller[3] · Christa E. Flück[1]

Received: 27 June 2017 / Revised: 23 May 2018 / Accepted: 26 May 2018 / Published online: 11 June 2018

Abstract

SF-1/*NR5A1* is a transcriptional regulator of adrenal and gonadal development. *NR5A1* disease-causing variants cause disorders of sex development (DSD) and adrenal failure, but most affected individuals show a broad DSD/reproductive phenotype only. Most *NR5A1* variants show in vitro pathogenic effects, but not when tested in heterozygote state together with wild-type *NR5A1* as usually seen in patients. Thus, the genotype−phenotype correlation for *NR5A1* variants remains an unsolved question. We analyzed heterozygous 46,XY SF-1/*NR5A1* patients by whole exome sequencing and used an algorithm for data analysis based on selected project-specific DSD- and SF-1-related genes. The variants detected were evaluated for their significance in literature, databases and checked in silico using webtools. We identified 19 potentially deleterious variants (one to seven per patient) in 18 genes in four 46,XY DSD subjects carrying heterozygous *NR5A1* disease-causing variants. We constructed a scheme of all these hits within the landscape of currently known genes involved in male sex determination and differentiation. Our results suggest that the broad phenotype in these heterozygous *NR5A1* 46, XY DSD subjects may well be explained by an oligogenic mode of inheritance, in which multiple hits, individually non-deleterious, may contribute to a DSD phenotype unique to each heterozygous SF-1/*NR5A1* individual.

Introduction

Steroidogenic factor 1 (SF-1/*NR5A1*) was first described as a transcription factor binding to a consensus *cis*-element in three steroid enzyme promoter regions and soon thereafter it was recognized as the human homolog of mouse fushi tarazu factor 1 [1]. Newborn *Ftzf1* null mice lack adrenal glands, have a female sexual phenotype irrespective of their chromosomal sex and show abnormalities in their hypothalamus as well as pituitary gonadotropes [2]. The first human homolog was identified in 1999 [3]: a phenotypically female patient harboring a heterozygous *NR5A1* p. Gly35Glu disease-causing variant presented with primary adrenal failure soon after birth and was found to have a 46, XY disorder of sex development (DSD) with complete sex reversal. Meanwhile, numerous patients with genetic variations in the *NR5A1* gene have been described presenting with an extraordinary broad phenotypic spectra, which remains so far a scientific and medical conundrum [4, 5]. While the first patient presented with an adrenal and a 46, XY DSD phenotype, most individuals identified thereafter (>100) presented with a reproductive phenotype only [4, 5]. In 46,XY, the reproductive phenotype may be male infertility, hypospadias or 46,XY DSD with complete sex reversal and persisting Müllerian structures and/or streak-like gonads [4]. In 46,XX, heterozygous *NR5A1* disease-causing variants have been found in women with familial or sporadic forms of primary or premature ovarian failure (POF) [4, 6]. More recently, the specific, heterozygous *NR5A1* disease-causing variant p.Arg92Trp has been found to affect gonadal determination and differentiation in both

Electronic supplementary material The online version of this article (https://doi.org/10.1038/s41431-018-0202-7) contains supplementary material, which is available to authorized users.

✉ Núria Camats
nuria.camats@vhir.org

[1] Pediatric Endocrinology and Diabetology, Department of Pediatrics and Department of Biomedical Research, University Children's Hospital Bern, 3010 Bern, Switzerland

[2] Growth and Development Research Unit, Vall d'Hebron Research Institute (VHIR), Center for Biomedical Research on Rare Diseases (CIBERER), Instituto de Salud Carlos III, 08035 Barcelona, Catalonia, Spain

[3] Division of Human Genetics, Department of Pediatrics and Department of Biomedical Research, Inselspital, Bern University Hospital, University of Bern, 3010 Bern, Switzerland

Table 1 Patients' characteristics

Patient[a]	Patient in Camats et al. [4]	Disease-causing variant: gene	Disease-causing variant: protein	Karyotype, assigned sex	Phenotype	Gonadal/reproductive function	Same/similar disease-causing variants reported in other individuals
Patient 1	1	*c.58G>C*	p.(Val20Leu)	46,XY DSD, male	Scrotal hypospadias, bilateral cryptorchidism, no Müllerian ducts	Normal baseline gonadal function (1 d and 1 yr)	–
Father 1	–	*c.58G>C*	p.(Val20Leu)	46,XY, male	Normal carrier	Fathered patient 1	–
Patient 2	3	*c.70delC*	p.(His24Thrfs*51)	46,XY DSD, female-to-male	Ambiguous genitalia, progressive virilization at puberty, no Müllerian structures	Abnormal gonadal function (normal T and precursors, high LH and FSH) (18 yrs)	c.70C>T [4]
Patient 3	5	*c.268G>C*	p.(Gly90Arg)	46,XY DSD, female	Female external genitalia, bilateral inguinal hernia, no Müllerian structures	NP	–
Patient 4	7	*c.614_615insC*	p.(Gln206Thrfs*20)	46,XY DSD, female	Female external genitalia, gonads in labia, no Müllerian ducts	Abnormal gonadal function (no T response to hCG) (14 d)	c.614_615insC [12]

All patients and families are Spanish Europeans of Iberian-Peninsula ethnicity.

NP gonadal function testing not performed

[a]Selected patients from Camats et al. [4]. All patients presented one heterozygous variant in *NR5A1*

chromosomal sexes: affected 46,XY individuals present with dysgenetic testes [7], whereas 46,XX individuals may present with testes, dysgenetic testes or ovotestes [8]. By contrast, there are only four subjects reported with heterozygous *NR5A1* disease-causing variants and adrenal failure [3, 9, 10], one of the 46,XX without accompanying gonadal failure so far [10].

NR5A1 disease-causing variants are mostly found in heterozygous state and can be missense, nonsense, frameshift, insertions, deletions or even complex variants. They are found scattered throughout the whole gene without apparent hot spots and a genotype−phenotype correlation remains unsolved. While functional in vitro studies unequivocally show pathogenic effects of identified *NR5A1* variants on known targets, they did not reveal a dominant negative mechanism of action for heterozygous disease-causing variants [4]. Likewise, haploinsufficiency seems not to explain the highly variable phenotype as even subjects harboring identical *NR5A1* disease-causing variants may present with completely different phenotypes [11]. In the extreme, we found a healthy, fertile father carrying the heterozygous *NR5A1* p.Val20Leu disease-causing variant while his heterozygous 46,XY DSD son presents with severe hypospadias and bilateral cryptorchidism [4]. Similarly, heterozygous *NR5A1* p.Arg255Leu/Cys variants were detected in a 46,XX female with adrenal failure, but intact ovarian function [10] and a 46,XX female with normal adrenal function but POF [6].

Thus, the lack of genotype−phenotype correlation for genetic *NR5A1* variants awaits further elucidation. Oligogenic modulators, epigenetic factors, imbalanced transcriptional *cis*-regulation, developmental switches, and environmental factors have been suggested as possible explanations [5, 11]. In fact, a digenic inheritance of gonadal dysgenesis has recently been suggested in a 46,XY DSD patient heterozygous for *NR5A1* and *MAP3K1* variants [12], and in a family harboring heterozygous *NR5A1* mutations manifesting as 46,XY DSD in males and 46,XX POF in females, in whom an additional variant in the *TBX2* gene was found in the females [13].

Oligogenic inheritance is currently discovered for several disorders by next generation sequencing (NGS). For instance, in congenital hypogonadotropic hypogonadism (HH) more than 25 causative genes are now considered to explain around 50% of the cases, and in at least 20% of cases disease-causing variants in two or more genes have been identified [14].

In search for a second genetic hit in heterozygous *NR5A1* patients, we recently tested the liver receptor homolog 1 (LRH-1/*NR5A2*), a close family member of nuclear receptor SF-1/*NR5A1* [15]. Although in vitro studies revealed that LRH-1 may compensate for SF-1 deficiency, we found no potentially disease-causing variants in

NR5A2 in 14 studied subjects. In the present work, we further pursued the hypothesis of possible oligogenic mode of inheritance and performed whole exome sequencing (WES) in five selected subjects harboring a heterozygous *NR5A1* disease-causing variant. For specific data analysis, we developed a DSD- and SF-1-specific data-filtering algorithm. Using this approach, we found up to seven additional potentially disease-causing variants in genes with reported SF-1 interaction in four subjects with a 46,XY DSD phenotype. Our findings suggest that the broad phenotypic spectrum of SF-1/*NR5A1* 46,XY DSD subjects may at least partially be caused by oligogenic inheritance.

Patients and methods

Patients

The study was approved by the Ethics Committee of Hospital Universitari Vall d'Hebron (CEIC), Barcelona, Spain (PR(IR)23/2016). Four 46,XY DSD patients carrying heterozygous *NR5A1* disease-causing variants and one 46,XY related normal carrier were analyzed using WES. The clinical and genetic characteristics of these patients were previously reported in great detail [4] and are summarized in Table 1.

DNA extraction, WES, and bioinformatic analysis

DNA was extracted from blood leukocytes using QiaCube (Qiagen, Hilden, Germany) or manually using a DNA isolation kit (Qiagen). WES was performed by Oxford Gene Technologies (OGT, Begbroke, UK). Putative candidate variants were confirmed by Sanger sequencing.

The genomic datasets were annotated and filtered with VariantStudio v2.2 (Illumina, San Diego, CA, USA), visualized and explored in Integrative Genomics Viewer (IGV, Broad Institute, Cambridge, MA, USA; https://www.broadinstitute.org/igv/) Frequencies of variants of relevant candidate genes were obtained from the Exome Aggregation Consortium (ExAC; Cambridge, MA, USA; http://exac.broadinstitute.org; February, 2016) and the Collaborative Spanish Variant Server (CSVS; CIBERER BIER, Valencia, Spain; http://csvs.babelomics.org/; December 2017). ExAC's dataset comprises more than 60,000 exomes of unrelated individuals from various large-scale sequencing projects. CSVS database includes, among others, exomes from a population of 267 healthy unrelated subjects [16]. We searched for reported (potentially) disease-causing variants with the Human Gene Mutation Database (HGMD® Professional 2016.4, http://www.biobase-international.com/product/hgmd; Biobase) and checked for polymorphisms in dbSNP (http://www.ncbi.nlm.nih.gov/snp/). We used SIFT (*Scale-invariant feature transform*; http://sift.jcvi.org/), PolyPhen-2 (Polymorphism Phenotyping v2; http://genetics.bwh.harvard.edu/pph2/index.shtml), Provean (http://provean.jcvi.org), MutationAssessor (http://mutationassessor.org/r3/), and Mutation Taster (http://www.mutationtaster.org/) to predict the possible impact of amino acid substitutions on the structure and function of corresponding human proteins. GERP++RS scores from dbNSFP database [17] and CADD (Combined Annotation Dependent Depletion; http://cadd.gs.washington.edu/), that scores the deleteriousness of single nucleotide variants as well as insertion/deletion variants in the human genome [18], were accessed through ANNOVAR [19] annotation.

In addition, we generated project-specific filters for DSD-related genes and for SF-1/*NR5A1*-related genes by searching in published literature. For the search for functional human partners of SF-1, the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING, http://string-db.org/), developed at the Center for Protein Research (CPR), EMBL, Swiss Institute of Bioinformatics (SIB), University of Copenhagen (KU), Technical University of Dresden (TUD), and University of Zurich (UZH), was used. The Biological General Repository for Interaction Datasets (BioGRID, thebiogrid.org), developed at Princeton University, University of Montreal, University of Edinburgh, and Mount Sinai Hospital, is a public database that was used to search for protein interactions with SF-1. Gene database GeneCards (https://www.genecards.org/) provided genomic information of human candidate genes in our patients. We also consulted Mouse Genomic Informatics database (MGI, http://www.informatics.jax.org/) and OMIM (https://www.omim.org) for further data analysis. We submitted our data to dbSNP (https://www.ncbi.nlm.nih.gov/snp/; November 2017) [20]. Submission SNP numbers (ss) are included in Table 2. These data will be publicly available when the next dbSNP Build (B152) is released (planned for summer 2018).

Design of a DSD- and SF-1/NR5A1-specific data-filtering algorithm

We prepared two gene lists for filtering our data and catch candidate genes and variants of interest (Fig. 1). Available information of each gene was collected to decide if it could be considered a candidate gene variant for DSD. For that, related literature and databases (GeneCards, STRING (human protein connections) and BioGRID) were consulted and searched for associated phenotypes in humans, mice (KO, microarray studies) and rats (microarray studies), and for related information from basic studies of cell lines from gonadal tissues or related to sex development. We also searched for genes related to SF-1 overexpression and knockdown [21].

Table 2 Identified genes and variants per heterozygous *NR5A1*/SF-1 patient after specific filtering

Patient	Gene	Name	Locus	Coordinates	Variant	Type/ Consequence	HGVSc	HGVSp	Prediction tests[a]							dbSNP ID	Frequency in ExAC browser[b]	Frequency in CSVS database[c]	dbSNP submission[d]
									SIFT	PolyPhen	Provean	Mutation Taster	Mutation Assessor	CADD	GERP				
1	*INHA*	Inhibin A	2q35	220439822	T>T/G	snv/missense	NM_002191.3: c.675T>G	NP_002182.1:p. (Ser225Arg)	Tolerated	Possibly damaging	Neutral	Polymorphism	Medium	2.54	-1.02	rs139608664	0.0013	ND	3626006439
2	*AKR1C3*	Aldo-keto reductase family 1, member C3	10p15-p14	5141619	A>A/G	snv/missense	NM_001253908.1: c.548A>G	NP_001240837.1:p. (Lys183Arg)	Tolerated	Benign	Deleterious	Disease causing	Neutral	20.0	2.67	rs61730879	0.002203	0.007	3626006450
2	*DOCK8*	Dedicator of cytokinesis 8	9p24.3	334238	T>T/C	snv/missense	NM_203447.3: c.1139T>C	NP_982272.2:p. (Ile380Thr)	Damaging	Benign	Neutral	Disease causing	Medium	23.2	4.74		0.000045	ND	3626006448
2	*FSHR*	Follicle-stimulating hormone receptor	2p21-p16	49190428	T>T/C	snv/missense	NM_000145.3: c.1532A>G	NP_000136.2:p. (Tyr511Cys)	Damaging	Probably damaging	Deleterious	Disease causing	High	24.9	5.35		ND	ND	3626006437
2	*NCOR1*	Nuclear receptor corepressor 1	17p11.2	15942948	G>G/A	snv/missense	NM_006311.3: c.6754C>T	NP_006302.2:p. (His2252Tyr)	Damaging	Probably damaging	Deleterious	Disease causing	Low	28.5	5.69	rs61755986	0.001211	0.002	3626006454
2	*NCOR1*	*Nuclear receptor corepressor 1*	17p11.2	15950400	*C>C/T*	*snv/missense*	*NM_006311.3: c.6544G>A*	*NP_006302.2:p. (Ala2182Thr)*	*Damaging*	*Benign*	*Neutral*	*Polymorphism*	*Neutral*	*24.0*	*4.53*	*rs61753149*	*0.01183*	*0.004*	*3626006455*
2	*POR*	P450 (cytochrome) oxidoreductase	7q11.2	75614391	T>T/G	snv/missense	NM_000941.2: c.1264T>G	NP_000932.3:p. (Trp422Gly)	Damaging	Probably damaging	Deleterious	Disease causing	High	24.6	5.14		0.000048	ND	3626006445
3	*CACNG4*	Calcium Channel, Voltage-Dependent, Gamma Subunit 4	17q24	65026851	C>C/T	snv/missense	NM_014405.3: c.715C>T	NP_055220.1:p. (Arg239Trp)	Damaging	Probably damaging	Deleterious	Disease causing	Medium	33.0	2.63		0.000015	ND	3626006456
3	*FBLN2*	Fibulin 2	3p25.1	13612240	G>G/A	snv/missense	NM_001004019.1: c.385G>A	NP_001004019.1:p. (Asp129Asn)	Damaging	Probably damaging	Neutral	Disease causing	Medium	23.8	5.05		0.00002945	ND	3626006440
3	*NAV1*	Neuron Navigator 1	1q32.3	201755657	C>C/A	snv/missense	NM_020443.4: c.2947C>A	NP_065176.3:p. (Pro983Thr)	Tolerated	Probably damaging	Neutral	Disease causing	Neutral	20.4	5.2	rs145865304	0.0007799	ND	2137543826
3	*SMAD6*	SMAD Family Member 6	15q22.31	67073833	G>G/GC	insertion/ frameshift elongation	NM_005585.4: c.1455dupC	NP_005576.3:p. (Cys486LeufsTer79)	NA	NA	NA	Disease causing	NA	NA	NA		ND	ND	3626006452
3	*SRA1*	Steroid Receptor RNA Activator	5q31.3	139936825	G>G/C	snv/missense	NM_001035235.3: c.94C>G	NP_001030312.2:p. (Gln32Glu)	Tolerated	Probably damaging	Neutral	Disease causing	NA	27.0	5.01	rs35610885	0.009544	0.01	3626006443
3	*ZDHHC11*	Zinc Finger, DHHC-Type Containing 11	5p15.33	840718	C>C/T	snv/missense	NM_024786.2: c.676G>A	NP_079062.1:p. (Val226Met)	Tolerated	Possibly damaging	Neutral	Polymorphism	Medium	15.88	0.711		0.00001502	ND	3626006441
3	*FOG2, ZFPM2*	Zinc Finger Protein, FOG Family Member 2	8q23	106573591	G>G/A	snv/ missense, splice region	NM_012082.3: c.302G>A	NP_036214.2:p. (Gly101Glu)	Tolerated	Possibly damaging	Neutral	Disease causing	Neutral	14.48	5.91	rs199605561	0.0001063	ND	3626006447
4	*CHD7*	Chromodomain Helicase DNA Binding Protein 7	8q12.2	61769418	A>A/C	snv/missense	NM_017780.3: c.7579A>C	NP_060250.2:p. (Met2527Leu)	Tolerated	Benign	Neutral	Disease causing	Low	7.5	1.97	rs192129249	0.003176	ND	3626006446
4	*DENND1A*	*DENN/MADD Domain Containing 1A*	*9q33.3*	*126144390*	*G>G/T*	*snv/missense*	*NM_020946.1: c.2351C>A*	*NP_065997.1:p. (Ala784Asp)*	*Damaging*	*Probably damaging*	*Neutral*	*Polymorphism*	*Medium*	*18.79*	*3.5*	*rs189947178*	*0.0101*	*ND*	3626006449
4	*GDNF*	Glial Cell Derived Neurotrophic Factor	5p13.1-p12	37816112	G>G/A	snv/missense	NM_001190468.1: c.328C>T	NP_001177397.1:p. (Arg110Trp)	Damaging	Probably damaging	Neutral	Disease causing	Low	28.8	4.96	rs36119840	0.003567	ND	3626006442
4	*GLI2*	GLI Family Zinc Finger 2	2q14	121747823	C>C/T	snv/missense	NM_005270.4: c.4333C>T	NP_005261.2:p. (Leu1445Phe)	Damaging	Possibly damaging	Neutral	Disease causing	Medium	23.7	2.53	rs146207623	0.002536	ND	3626006438
4	*SOX30*		5q33	157078632		snv/missense			Damaging	Benign	Neutral	Polymorphism	low	11.41	2.25	rs13181859	0.003701	0.007	3626006444

Table 2 (continued)

Patient	Gene	Name	Locus	Coordinates	Variant	Type/ Consequence	HGVSc	HGVSp	Prediction tests[a]						dbSNP ID	Frequency in ExAC browser[b]	Frequency in CSVS database[c]	dbSNP submission[d]
									SIFT	PolyPhen	Provean	Mutation Taster	Mutation Assessor	CADD	GERP			
		SRY (Sex Determining Region Y)-Box 30			G>G/A		NM_178424.1: c.455C>T	NP_848511.1:p.(Pro152Leu)										

All patients were heterozygote for these variants and were checked and confirmed through IGV software (alignment with human genome hg19/grch37; https://www.broadinstitute.org/igv/, Broad Institute, Cambridge, MA, USA)

NA not analyzed by prediction test, *ND* variant not detected in all exomes (ExAC, out of 65,000 individuals; CSVS out of 267 unrelated healthy individuals); in italics: allele frequency >1% in ExAC population

[a]Prediction tests: SIFT, PolyPhen-2, Provean, Mutation Taster and MutationAssessor (r3), CADD (CADD_phred: ≥18: considered deleterious for DSD, ≥20 consensus deleterious), GERP (GERP ++: positive scores indicate that a site is a functional element under evolutionary constraint). Bioinformatic analysis was done in February 2016 and November 2017

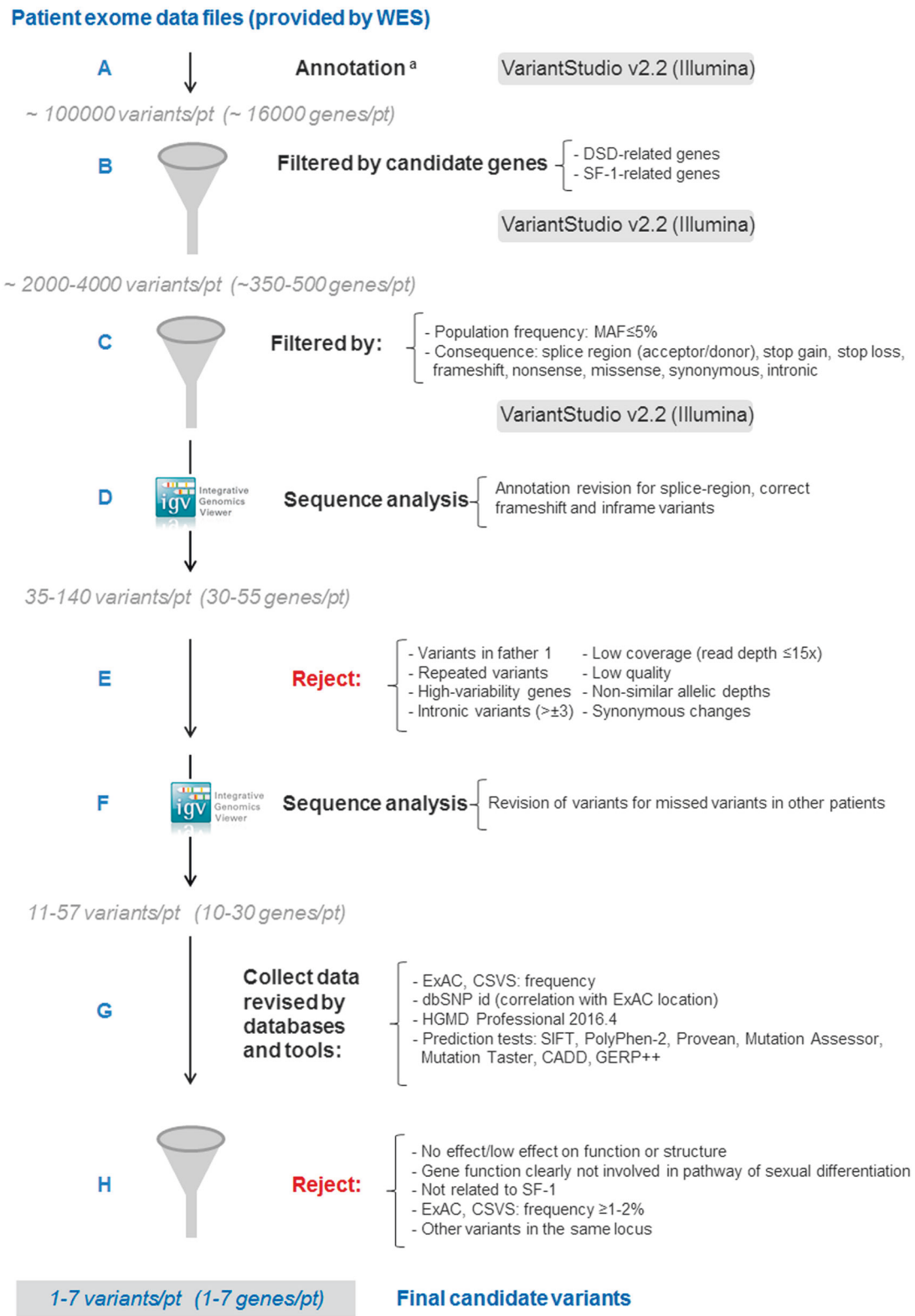
[b]ExAC, Feb2016: frequencies (European non-Finnish) in database given out of 1

[c]CSVS: Collaborative Spanish Variation Server (http://csvs.babelomics.org/), Dec2017: frequencies in database given out of 1

[d]dbSNP submission on Nov 2017. ss: submitted SNP number. These data will be publicly available when the next dbSNP Build (B152) is released, summer, 2018

Our DSD-gene list ($N = 479$) included (a) genes with reported (potentially) deleterious variants in patients with DSD (both XX or XY karyotypes), (b) genes with reported (potentially) disease-causing variants in syndromic patients with involvement of sex development, (c) genes in KO/mutant animal models (mice and rats), which caused/seemed to cause a DSD condition, and (d) overexpressed, upregulated or downregulated genes in rodent embrionic gonadal cells (source: literature). Our SF-1-related gene list ($N = 632$) included genes regulating and/or modulating SF-1 function at the protein, RNA or gene level, and SF-1 targeted genes (sources: literature and String and BioGrid databases). Both lists shared common genes ($N = 97$). Shared genes in DSD- and SF-1-gene lists are included in Table S1.

For patient analysis, we first filtered all genomic datasets separately for each patient using both gene lists (Table S2, Fig. 1, step B). Then, we kept the resulting variants with an MAF (minor allele frequency) ≤5% and the predicted consequences of the variants (Fig. 1, step C), as described [22, 23]. We confirmed the correct annotation and location of variants (splice-region variants, frameshift and inframe variants (deletions and insertions)) by checking their alignment data in IGV (alignment with human genome hg19/grch37) (data not shown) (Fig. 1, step D). In step E (Fig. 1) we excluded variants that were considered non-relevant for our study, e.g.: (1) variants from patient 1 present in his father, (2) in all patients, those variants present in more than one patient, (3) variants from genes with high variability, (4) variants with low coverage and/or low quality, (5) variants with non-similar allelic depths, (6) synonymous changes and (7) intronic variants further than ±3 nucleotides from exon. In step F, we excluded the possibility that some variants may have been missed in the previous annotation steps or may also be present in more than one patient by comparing alignments using IGV (data not shown). In step G, the variant frequencies were checked in the Exome Aggregation Consortium (ExAC). We assessed restrictive low variant frequency in our reference population (European non-Finnish population, MAF ≤ 0.01 −0.02, ExAC Browser, Feb 2016), thus more plausible to be a DSD-causing variant. We searched in the Human Gene Mutation Database (HGMD® Professional 2016.4) to check if previously described, in which case, the reference number (rs) was checked for location in ExAC Browser and dbSNP. Finally, we predicted the possible effect of the identified potentially disease-causing variant (all amino acid substitutions, some deletions and insertions) on its protein function using SIFT, *PolyPhen-2*, Provean, MutationAssessor, Mutation Taster, CADD and GERP++. Variants were also crosschecked with a healthy cohort of the Spanish population (CSVS: 267 unrelated healthy controls) [16].



In summary (step H), variant inclusion criteria were: (1) low frequency (MAF ≤ 0.01−0.02, ExAC Browser and CSVS databases), (2) involved in pathway or with function related to sexual determination, differentiation and development, (3) in relation/interaction with SF-1, and/or (4) at least one of the prediction tests giving an effect on function.

◀ **Fig. 1** Algorithm used for data analysis after whole exome sequencing (WES) of patients harboring heterozygous *NR5A1*/SF-1 disease-causing variants. Number of variants and genes retrieved after each filtering step of the analysis are indicated. Short information on filtering steps is also provided. Capital letters A−H identify the analysis steps. pt patient, a annotation per variant: gene, transcript, protein, change (nucleotide, amino acid, codon (HGVS coding sequence name)), position (chromosome, coordinate, exonic/intronic), genotype (heterozygote, homozygote, hemizygote), type (snv, deletion, insertion), consequence (splice region (acceptor/donor), stop gain, stop loss, frameshift, nonsense, missense, synonymous, intronic), dbSNP id, read depth, filter pass, quality control, allele freq global minor (*minor allele frequency* (MAF)), frequency in EVS, Cosmic, ClinVar, etc., prediction of impact (PolyPhen-2, SIFT)

Results

WES performed in five subjects harboring heterozygous *NR5A1* disease-causing variants revealed a total of about 100,000 variants in ~16,000 genes (Fig. 1, step A; Table S2).

Identified genes were filtered by candidate gene lists (step B) and resulted in 2272−4205 variants based on the DSD-related gene list and 2850–3194 variants based on the SF-1-related gene list. After step C, which filtered for MAF ≤ 0.05 and deleterious consequences, 35–63 DSD-related variants and 694–862 SF-1-related variants were left. By detailed sequence reanalysis (step D) and rejection of weak and unlikely disease-causing variants (see list in step E), 11 and 57 variants, respectively, were left for the final evaluation. These variants were reevaluated in steps G and H and 18 variants in 16 genes were finally rated non-deleterious (Fig. 1): *AMH, CDH3, FBN2, FGFR2, FRZB, IGF1R, IRS1, LEPROT, NCOR2, PEG3, RET, RYR2, SFRP2, TAGLN, TGFBI* and *WWOX*. Details of these rejected variants are listed in Table S3 and corresponding information from the literature is provided in Table S4.

Identification of an oligogenic DSD etiology in 46,XY individuals with heterozygous SF-1/*NR5A1* disease-causing variants

We identified a total of 19 potentially deleterious variants in 18 genes in 4 heterozygous *NR5A1* patients (Table 2). One variant was detected in patient 1, 6 variants in patient 2, 7 variants in patient 3 and 5 variants in patient 4. Nine variants originated from the DSD list, while 12 variants came from the SF-1-related gene list. Patients with more than one variant had variants related to both DSD and SF-1. None of the variants present in patients 2−4 were detected in the father of patient 1.

In patient 1, one variant in the *INHA* gene, c.675T>G [p.(Ser255Arg)], was predicted to be deleterious by the programs PolyPhen-2 and MutationAssessor.

In patient 2, six missense variants were picked up in five genes: *AKR1C3* [c.548A>G, p.(Lys183Arg)], *DOCK8* [c.1139T>C, p.(Ile380Thr)], *FSHR* [c.1532A>G, p.(Tyr511Cys)], *NCOR1* [c.6544G>A, p.(His2252Tyr)] and [c.6754C>T, p.(Ala2182Thr)] and *POR* [c.1264T>G, p.(Trp422Gly)]. The sequence variations in *FSHR*, *NCOR1* (c.6754C>T) and *POR* were judged deleterious by most prediction programs.

Seven heterozygous variants in different genes were found in patient 3: *CACNG4* [c.715C>T, p.(Arg239Trp)], *FBLN2* [c.385G>A, p.(Asp129Asn)], *NAV1* [c.2947C>A, p.(Pro983Thr)], *SMAD6* [c.1455dupC, p.(Cys486LeufsTer79)], *SRA1* [c.94C>G, p.(Gln32Glu)], *ZDHHC11* [c.676G>A, p.(Val226Met)] and *FOG2/ZFPM2* [c.302G>A, p.(Gly101Glu)]. Each of the six missense variants was rated deleterious by at least one of the applied prediction tests.

Finally, we detected five heterozygous missense variants in five genes in patient 4. These were *CHD7* [c.7579A>C, p.(Met2527Leu)], *DENND1A* [c.2351C>A, p.(Ala784Asp)], *GDNF* [c.328C>T, p.(Arg110Trp)], *GLI2* [c.4333C>T, p.(Leu1445Phe)] and *SOX30* [c.455C>T, p.(Pro152Leu)]. Each of the five missense variants was rated deleterious by at least one of the applied prediction tests.

We reviewed the published databases and literature to solve the question whether any potentially deleterious or confirmed disease-causing variants in the identified genes are known in humans or whether, at least, a mouse phenotype has been described (Table 2 and Table S4). Besides the information on control exomes from ExAC, we checked the candidate variants in a cohort of healthy Spanish Population (CSVS: 267 healthy controls; http://csvs.babelomics.org/) [16] (Table 2). Fourteen variants were not present, four had MAF < 0.01 (*AKR1C3*, 2 in *NCOR1* and *SOX30*) and three had MAF > 0.01 (*WWOX*, *RET*, and *SRA1*).

Furthermore, information on genotype−phenotype correlation for the identified genes, as well as current knowledge from research on their involvement in sex determination and differentiation and their relation to SF-1 has been collected and is summarized in Table S4. Finally, we used all this information to draw a scheme to provide the genetic landscape of potential oligogenic hits identified in our 46,XY DSD heterozygous SF-1 patients in perspective to the current view of genetic interactions in gonadal male sex determination and differentiation (Fig. 2).

Discussion

All four studied SF-1 patients harbored at least one other gene variant possibly contributing to a DSD phenotype besides the *NR5A1* disease-causing variant. A summary of

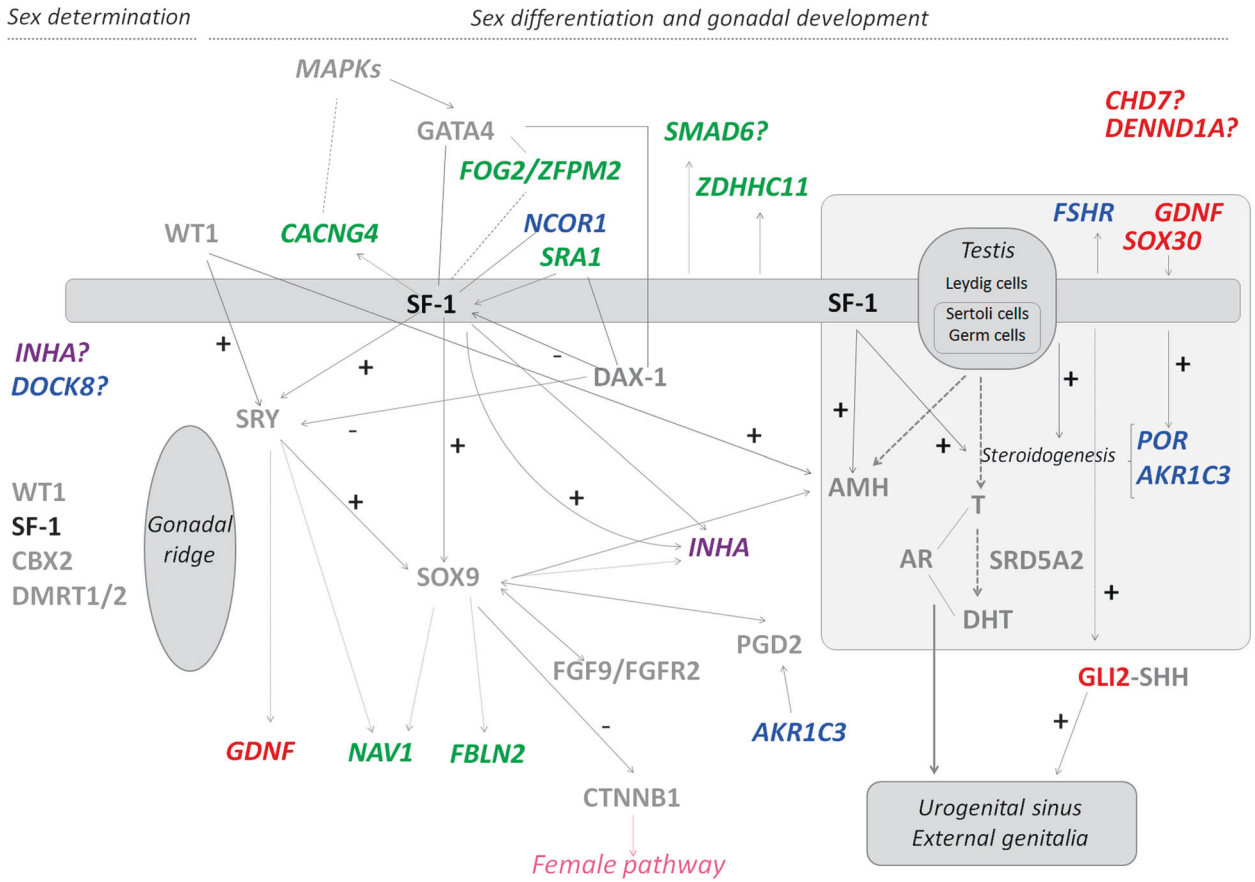


Fig. 2 Additional, likely disease-causing genetic variants identified in four 46,XY patients with disorder of sex development harboring heterozygous *NR5A1*/SF-1 disease-causing variants depicted with respect to the known pathway of male sex determination and differentiation. The scheme shows an overview of involved genes and their interrelationship. It emphasizes on SF-1, which seems to play an important role throughout all developmental processes (indicated by a thick line). Genetic variants identified by whole exome sequencing in the studied patients are given in specific colors. In violet: candidate gene in patient 1; in blue: candidate genes in patient 2; in green: candidate genes in patient 3; in red: candidate genes in patient 4; in gray: known genes involved in sexual development. Interrogation mark (?): function/timing/location is not clear; arrows: regulation/co-activation; dotted arrows: gene with binding regions for SF-1, SRY, and/or SOX9; lines: interaction/partnership; dashed lines: related genes, but thus far unclear how exactly; thick dashed arrows: hormone production

all identified hits discovered in these SF-1 patients is depicted in Fig. 2 showing most of the currently known genes involved in sex development. This underlines the complexity of sex development and visualizes that multiple genetic hits, which may not be deleterious alone, may contribute to a DSD phenotype that may be unique to each heterozygous SF-1/*NR5A1* individual. If so, the genotype −phenotype correlation may greatly depend on the nature of the secondary hit, which would be a plausible explanation for the broad phenotype spectrum.

In our study, we selected for potentially deleterious variants by establishing lists for DSD-related and/or SF-1-related genes from literature and databases. These variants and our rationale why they may contribute to the DSD phenotype is discussed in greater detail in the supplementary information (Table S4).

Among the 18 genes in which we detected variants in our SF-1 patients, eight had been previously reported as DSD-causing in humans (Table 2 and Table S4). Most of them are related to 46,XY DSD/HH (*CHD7, FOG2/ZFPM2*, and *SRA1*) and to 46,XX DSD/POF/HH (*CHD7, DENND1A, FSHR, GLI2, INHA, POR*, and *SRA1*). In other seven genes (*AKR1C3, DOCK8, NCOR1, FBLN2, NAV1, SMAD6*, and *GDNF*), the described potentially deleterious variants have not been related to sex development or gonadal function yet. However, the *NAV1* variant is a plausible candidate as it is present in another 46,XY DSD patient (new cohort currently under study), has an MAF < 0.01 in ExAC, and is not detected in 534 healthy Spanish chromosomes. To be noted that, at the time of the present analysis, we did not consider to include repeated variants (in more than one patient) because our cohort was small and we were focused on very rare variants. However, as our cohort increases, we intend to consider heterozygous repeated variants because the co-occurrence of one specific variant in more than one patient would strengthen its potential causal

involvement, as seen with the previously cited *NAV1* variant. In addition, the three remaining genes with identified variants (*CACNG4, ZDHHC11*, and *SOX30*), and no entry in the HGMD database so far, have been previously discussed as strong DSD candidates [24, 25].

With respect to gene interactions with SF-1, we assessed whether the identified genes were modulated when SF-1 was overexpressed or knocked-down in steroidogenic NCI-H295R cells [21], but none of them were. To further analyze the relationship between identified gene variations and SF-1/*NR5A1*, we used STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) and BioGRID (Biological General Repository for Interaction Datasets), and reviewed the published literature for information on gene−protein and protein−protein interactions (Table S4). Nine identified genes have been previously shown to interact with SF-1/*NR5A1* in functional studies [24, 26–29]. Six of them (*FSHR, CACNG4, GLI2, INHA, SMAD6*, and *ZDHHC11*) are targets of SF-1 [24, 26, 27] and other two (*NCOR1* and *SOX30*) interact at the protein−protein level to regulate SF-1 [26]. In contrast, SRA1 interacts with SF-1 by (non-coding-)RNA−protein interaction [28]. Of note, *INHA* is both SF-1 target and regulator [26].

In summary, our study lends support to the concept that the broad range of DSD phenotypes in heterozygous SF-1/*NR5A1* patients may be due to additional variants in related genes. Thus, we (and others) propose that the broad DSD phenotype in *NR5A1* patients might be caused by oligogenic inheritance as seen in similar disorders such as HH. We should point out, however, that since the filtering protocol is likely to have enriched for variants in the genes described here, some of the potentially deleterious variants now identified can ultimately not be shown to affect sex development. A more extensive study including other DSD cohorts is under way to assess which of the presently identified variants are actually demonstrably involved in modifying DSD, and whether including these variants might provide a better genotype−phenotype correlation. Use of NGS approach for genetic work-up of DSD patients will reveal further insight into more complex genetic traits than thought of today.

Acknowledgements We acknowledge the patients, families and their primary physicians for sharing their data for our study. We also thank Ida Paramonov for her help in the bioinformatic analysis.

Funding This work was supported by grants of the Swiss National Science Foundation (http://www.snf.ch) (320030-146127) to CEF, the Instituto de Salud Carlos III (www.isciii.es/; Madrid, Spain) Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER, http://www.ciberer.es/) U-712 to MF-C, the Agency for Management of University and Research Grants (AGAUR; agaur.gencat.cat), Barcelona, Spain (2009SGR31) to LA, and by the Beatriu de Pinós Fellowship 2014 BP-B 00145 (AGAUR, Catalonia, Spain), the Instituto de Salud Carlos III (www.isciii.es/; Madrid, Spain) Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER; http://www.ciberer.es/) U-712 and the private Foundation Bangerter-Rhyner (http://www.bangerter-stiftung.ch), Basel, Switzerland to NC.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Lala DS, Rice DA, Parker KL. Steroidogenic factor I, a key regulator of steroidogenic enzyme expression, is the mouse homolog of fushi tarazu-factor I. Mol Endocrinol. 1992;6:1249–58.
2. Luo X, Ikeda Y, Parker KL. A cell-specific nuclear receptor is essential for adrenal and gonadal development and sexual differentiation. Cell. 1994;77:481–90.
3. 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;22:125–6.
4. Camats N, Pandey AV, Fernandez-Cancio M, et al. Ten novel mutations in the NR5A1 gene cause disordered sex development in 46,XY and ovarian insufficiency in 46,XX individuals. J Clin Endocrinol Metab. 2012;97:E1294–306.
5. Domenice S, Zamboni Machado A, Moraes Ferreira F, et al. Wide spectrum of NR5A1-related phenotypes in 46,XY and 46,XX individuals. Birth Defects Res C Embryo Today. 2016;108:309–20.
6. Lourenco D, Brauner R, Lin L, et al. Mutations in NR5A1 associated with ovarian insufficiency. N Engl J Med. 2009;360:1200–10.
7. Bashamboo A, Donohoue PA, Vilain E, et al. A recurrent p. Arg92Trp variant in steroidogenic factor-1 (NR5A1) can act as a molecular switch in human sex development. Hum Mol Genet. 2016;25:5286.
8. Baetens D, Stoop H, Peelman F, et al. NR5A1 is a novel disease gene for 46,XX testicular and ovotesticular disorders of sex development. Genet Med. 2017;19:367–76.
9. Malikova J, Camats N, Fernandez-Cancio M, et al. Human NR5A1/SF-1 mutations show decreased activity on BDNF (brain-derived neurotrophic factor), an important regulator of energy balance: testing impact of novel SF-1 mutations beyond steroidogenesis. PLoS ONE. 2014;9:e104838.
10. Biason-Lauber A, Schoenle EJ. Apparently normal ovarian differentiation in a prepubertal girl with transcriptionally inactive steroidogenic factor 1 (NR5A1/SF-1) and adrenocortical insufficiency. Am J Hum Genet. 2000;67:1563–8.
11. Ferraz-de-Souza B, Lin L, Achermann JC. Steroidogenic factor-1 (SF-1, NR5A1) and human disease. Mol Cell Endocrinol. 2011;336:198–205.
12. Mazen I, Abdel-Hamid M, Mekkawy M, et al. Identification of NR5A1 mutations and possible digenic inheritance in 46,XY gonadal dysgenesis. Sex Dev. 2016;10:147–51.
13. Werner R, Mönig I, Lünstedt R, et al. New NR5A1 mutations and phenotypic variations of gonadal dysgenesis. PLoS ONE. 2017;12. https://doi.org/10.1371/journal.pone.0176720.
14. Boehm U, Bouloux PM, Dattani MT, et al. Expert consensus document: European Consensus Statement on congenital hypogonadotropic hypogonadism—pathogenesis, diagnosis and treatment. Nat Rev Endocrinol. 2015;11:547–64.
15. Camats N, Audi L, Fernandez-Cancio M, et al. LRH-1 may rescue SF-1 deficiency for steroidogenesis: an in vitro and in vivo study. Sex Dev. 2015;9:144–54.

16. Dopazo J, Amadoz A, Bleda M, et al. 267 Spanish exomes reveal population-specific differences in disease-related genetic variation. Mol Biol Evol. 2016;33:1205–18.
17. Liu X, Wu C, Li C, Boerwinkle E. dbNSFPv3.0: a one-stop database of functional predictions and annotations for human nonsynonymous and splice-site SNVs. Hum Mutat. 2016; 37:235–41.
18. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014; 46:310–5.
19. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38:e164.
20. Sherry ST. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001;29:308–11.
21. Doghman M, Figueiredo BC, Volante M, Papotti M, Lalli E. Integrative analysis of SF-1 transcription factor dosage impact on genome-wide binding and gene expression regulation. Nucleic Acids Res. 2013;41:8896–907.
22. Lee H, Deignan JL, Dorrani N, et al. Clinical exome sequencing for genetic identification of rare Mendelian disorders. JAMA. 2014;312:1880–7.
23. Baxter RM, Arboleda VA, Lee H, et al. Exome sequencing for the diagnosis of 46,XY disorders of sex development. J Clin Endocrinol Metab. 2015;100:E333–44.
24. Ferraz-de-Souza B, Lin L, Shah S, et al. ChIP-on-chip analysis reveals angiopoietin 2 (Ang2, ANGPT2) as a novel target of steroidogenic factor-1 (SF-1, NR5A1) in the human adrenal gland. FASEB J. 2011;25:1166–75.
25. Han F, Dong Y, Liu W, et al. Epigenetic regulation of sox30 is associated with testis development in mice. PLoS ONE. 2014;9: e97203.
26. Schimmer BP, White PC. Minireview: steroidogenic factor 1: its roles in differentiation, development, and disease. Mol Endocrinol. 2010;24:1322–37.
27. Hoivik EA, Lewis AE, Aumo L, Bakke M. Molecular aspects of steroidogenic factor 1 (SF-1). Mol Cell Endocrinol. 2010; 315:27–39.
28. Xu B, Yang WH, Gerin I, Hu CD, Hammer GD, Koenig RJ. Dax-1 and steroid receptor RNA activator (SRA) function as transcriptional coactivators for steroidogenic factor 1 in steroidogenesis. Mol Cell Biol. 2009;29:1719–34.
29. Kelly VR,Hammer GD, LRH-1 and Nanog regulate Dax1 transcription in mouse embryonic stem cells. Mol Cell Endocrinol. 2011;332:116–24.