Author/s:
Eggers, S;Smith, KR;Bahlo, M;Looijenga, LHJ;Drop, SLS;Juniarto, ZA;Harley, VR;Koopman, P;Faradz, SMH;Sinclair, AH

Title:
Whole exome sequencing combined with linkage analysis identifies a novel 3 bp deletion in NR5A1

Date:
2015-04-14

Citation:
Eggers, S., Smith, K. R., Bahlo, M., Looijenga, L. H. J., Drop, S. L. S., Juniarto, Z. A., Harley, V. R., Koopman, P., Faradz, S. M. H. & Sinclair, A. H. (2015). Whole exome sequencing combined with linkage analysis identifies a novel 3 bp deletion in NR5A1. European Journal of Human Genetics, 23 (4), pp.486-493. https://doi.org/10.1038/ejhg.2014.130.

Persistent Link:
https://hdl.handle.net/11343/263658

European Journal of Human Genetics (2015) 23, 486–493

www.nature.com/ejhg

EJHG Open

ARTICLE

Whole exome sequencing combined with linkage analysis identifies a novel 3 bp deletion in *NR5A1*

Stefanie Eggers[1,2], Katherine R Smith[3,4], Melanie Bahlo[3,5], Leendert HJ Looijenga[6], Stenvert LS Drop[7], Zulfa A Juniarto[8], Vincent R Harley[9], Peter Koopman[10], Sultana MH Faradz[8] and Andrew H Sinclair*[,1,2]

Disorders of sex development (DSDs) encompass a broad spectrum of conditions affecting the development of the gonads and genitalia. The underlying causes for DSDs include gain or loss of function variants in genes responsible for gonad development or steroidogenesis. Most patients with DSD have an unknown genetic etiology and cannot be given an accurate diagnosis. We used whole exome capture and massively parallel sequencing to analyse a large family with 46,XY DSD and 46,XX premature ovarian insufficiency. In addition, we used a recently developed method for linkage analysis using genotypes extracted from the MPS data. This approach identified a unique linkage peak on chromosome 9 and a novel, 3 bp, in-frame deletion in exon six of *NR5A1* (steroidogenic factor-1 or SF1) in all affected individuals. We confirmed that the variant disrupts the SF1 protein and its ability to bind and regulate downstream genes. *NR5A1* has key roles at multiple points in gonad development and steroidogenic pathways. The variant described here affects the function of SF1 in early testis development and later ovarian function, ultimately leading to the 46,XY DSD and 46,XX premature ovarian insufficiency phenotypes, respectively. This study shows that even at low coverage, whole exome sequencing, when combined with linkage analysis, can be a powerful tool to identify rapidly the disease-causing variant in large pedigrees.

European Journal of Human Genetics (2015) **23,** 486–493; doi:10.1038/ejhg.2014.130; published online 6 August 2014

INTRODUCTION

Disorders of sex development (DSDs) cover a wide spectrum of phenotypes, ranging from complete sex reversal to ambiguous genitalia, the latter affecting 1 in 4500 births.[1]

Patients with 46,XY partial or complete gonadal dysgenesis have either underdeveloped testes, ovotestes (partial gonadal dysgenesis) or streak gonads (complete gonadal dysgenesis). Disruption of testis development leads to either undervirilization of the external genitalia, ambiguous genitalia or a female external phenotype. The clinical management of these complex conditions is made difficult because up to 70% of 46,XY gonadal dysgenesis cases lack a diagnosis and the underlying molecular cause remains unknown.[2] However, point variants, deletions or duplications of genes such as sex-determining region Y (*SRY*), SRY-related HMG-box, gene 9 (*SOX9*), nuclear receptor subfamily 5, group A, member 1 (*NRA51*), and mitogen-activated protein kinase kinase kinase 1, E3 ubiquitin protein ligase (*MAP3K1*) have been shown to be responsible for 30% of 46,XY DSD cases.[3]

Premature ovarian insufficiency (POI) in 46,XX individuals is characterized by early failure of ovarian function, affecting 1% of women.[4] In the majority of 46,XX POI cases, the underlying genetic cause remains unknown. Few genes such as autoimmune regulator (*AIRE*; causing autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED) syndrome) and forkhead box L2 (*FOXL2*) have been associated with 46,XX POI in a minority of patients.[5]

Studies in mouse have identified *Nr5a1* as a key gene involved in gonadal sex determination, differentiation, and maintenance.[3] *Nr5a1* encodes steroidogenic factor 1 (SF1), an orphan nuclear receptor expressed early in the bipotential gonad, and later in the developing and mature testis and ovary. *Nr5a1*-null XY mice show streak gonads and feminization, which mimics the phenotype of patients affected by 46,XY complete gonadal dysgenesis. Genes downstream of *NR5A1* in the testis and ovary have been identified by analysing different mouse strains and human fetal gonads. These include genes required for testis and ovary development, such as *SOX9*, anti-Müllerian hormone (*AMH*), and nuclear receptor subfamily 0, group b, member 1 (*Nr0b1* or *Dax1*), as well as, in particular, gonadal steroidogenesis.[6–8]

Variants in and around *NR5A1* are causative for 46,XY gonadal dysgenesis.[9,10] A recent study analysed four families with individuals exhibiting 46,XY gonadal dysgenesis and female relatives affected by 46,XX POI.[11] *NR5A1* was found to be the causative gene for both phenotypes in all four families, representing the first link of *NR5A1* to 46,XX POI. The authors also found variants in *NR5A1* in two sporadic 46,XX POI cases. Another study identified 10 novel variants in *NR5A1* in a cohort of sporadic cases of 46,XY gonadal dysgenesis and 46,XX POI.[12] Here we describe the analysis of a multigeneration family affected by 46,XY DSD and 46,XX POI, using low-coverage whole exome sequencing combined with linkage analysis. This approach was extremely powerful in quickly refining the search area within the exome down to the nucleotide level, allowing rapid identification of a

[1]Murdoch Children's Research Institute, Royal Children's Hospital, Melbourne, VIC, Australia; [2]Department of Paediatrics, The University of Melbourne, Melbourne, VIC, Australia; [3]The Walter and Eliza Hall Institute of Medical Research, Melbourne, VIC, Australia; [4]Department of Medical Biology, The University of Melbourne, Melbourne, VIC, Australia; [5]Department of Mathematics and Statistics, The University of Melbourne, Melbourne, VIC, Australia; [6]Department of Pathology, Erasmus MC-University Medical Center Rotterdam, Josephine Nefkens Institute, Rotterdam, The Netherlands; [7]Department of Paediatrics, Division of Pediatric Endocrinology, Erasmus MC – Sophia Children's Hospital Rotterdam, Rotterdam, The Netherlands; [8]Center for Biomedical Research Faculty of Medicine Diponegoro University (FMDU), Semarang, Indonesia; [9]Prince Henry's Institute of Medical Research, Melbourne, VIC, Australia; [10]Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD, Australia
*Correspondence: Professor A Sinclair, Murdoch Children's Research Institute, Royal Children's Hospital, Flemington Road, Melbourne, 3052 VIC, Australia. Tel: +61 3 8341 6424; Fax: +61 3 8341 6429; E-mail: andrew.sinclair@mcri.edu.au
Received 23 December 2013; revised 3 April 2014; accepted 3 June 2014; published online 6 August 2014

novel 3-bp deletion in *NR5A1* as the causative variant in this family. In addition, we analysed the function of the mutant SF1 protein to investigate the biologic consequences of the lesion.

MATERIALS AND METHODS

Clinical data: family members with 46,XY DSD

Case II:8: The 46,XY DSD patient II:8 (later referred to as 'unclassified 46,XY DSD') presented with ambiguous genitalia and gynaecomastia, severe penoscrotal hypospadias, bifid scrotum, micropenis, small bilateral testis (2–3 ml), and primary infertility. Owing to the mainly male external genitalia, he was raised as a male.

Case III:1: This 46,XY gonadal dysgeneis patient presented with atrophy of testis (bilateral small inguinal palpable testis; <1 ml), with labia major (fusion of labia major, which resembles a scrotum and is darker in colour) and small labia minor, blind ending vagina, and an enlarged clitoris. The child was raised as a girl. Abdominal testes were removed and the patient will have lifelong hormonal treatment. Pathology results from the removed testis revealed low number of Leydig cells and atrophy of seminiferous tubules. No malignant cells and spermatozoa, but few spermatid cells were detected. Low testosterone levels (8 ng/dl before and <20 ng/dl after gonadectomy) were measured.

Case III:2: Presented with 46,XY gonadal dysgenesis, micropenis (phallus length 5.4 cm at the age of 10 years), Quigley stage 2, Tanner stage V, severe scrotal hypospadias (meatus at glans), bifid scrotum, severe chordee, and small bilateral testis (both in scrotum and both 5 ml volume). The patient underwent plastic surgery three times to repair the hypospadias and was raised as a male. No uterus or ovary could be detected.

Clinical data: family members with 46,XX POI

The four female relatives who presented with 46,XX POI (I:2, II:2, II:4, and II:7) were all diagnosed at different ages (II:4 onset at 35, and II:7 at 24 years of age; for I:2 and II:2 the age of onset is unknown), by onset of secondary amenorrhoea.

Patient DNA samples

DNA from the 15 family members was extracted from peripheral EDTA blood samples according to a standard protocol (desalting method). Approval for this study was obtained by the Human Ethics Committee of the Faculty of Medicine, Diponegoro University, Semarang, Indonesia. All patients and family members gave their informed consent before taking part in this study. This study also received approval from the Human Ethics Committee at the Royal Children's Hospital, Melbourne, Victoria, Australia.

Exome capture, sequencing, and bioinformatic analysis

Exome capture (TruSeq Exome Enrichment Kit; Illumina Inc., San Diego, CA, USA) and sequencing (Illumina HiSeq 2000; Illumina) were performed by the Australian Genome Research Facility (Melbourne, VIC, Australia). The 100 bp paired-end reads were aligned to UCSC hg19 using Novoalign version 2.07.09 in paired-end mode (http://www.novocraft.com). Reads mapping to multiple locations were discarded; presumed PCR duplicates were discarded using the MarkDuplicates utility from Picard (http://picard.sourceforge.net/). Variants were detected using the mpileup and bcftools view commands from SAMtools version 0.1.17,[13,14] specifying parameters −C50 and −q13. Low confidence variants were discarded using the vcfutils.pl varFilter script from the same program. Variants were annotated against the UCSC KnownGene annotation, dbSNP132, and the May 2011 release of the 1000 Genomes Project[15] using ANNOVAR.[16] *In silico* analysis of the functional effects of coding non-synonymous variants was performed with SIFT[17] and HumVar-trained PolyPhen-2 v.2.1.0.[18] We used SAMtools to infer, from the exome sequence data, genotypes at the location of HapMap Phase II and III SNPs, specifying parameters −cg and −t0.5.[19] Parametric multipoint linkage analysis was performed using MERLIN[20] under a fully penetrant autosomal dominant model with a 0% phenocopy rate and a disease allele frequency of 0.00001. SNP allele frequencies were obtained from HapMap CEU (Utah residents with ancestry from northern and western Europe from the CEPH collection) genotypes.

In silico prediction on the wild-type and mutant protein

To predict changes between the wild-type and mutant protein structure, an *in silico* prediction of the two protein structures was performed using the protein comparative modelling server 3D-JIGSAW (http://bmm.cancerresearchuk.org/~3djigsaw/).

Plasmids

The mammalian expression vector pCMV6-Entry-*hNR5A1* (RC207577; OriGene Technologies Inc., Rockville, MD, USA) containing the human cDNA ORF of *NR5A1* was used to obtain the expression vector containing the mutant *NR5A1* cDNA (pCMV6-Entry-*hNR5A1*-p.(Lys372del)). Site-directed mutagenesis was performed using the QuickChange II XL Site-directed Mutagenesis Kit (Agilent Technologies Inc., Santa Clara, CA, USA) according to the manufacturer's instructions (for primer sequences see Supplementary Table S5). To create the pGL4.10-*hNR0B1* construct for dual-luciferase reporter assays, the human *NR0B1* (*hNR0B1*) promoter was PCR amplified (for PCR primer sequences see Supplementary Table S5) and cloned into the pGL4.10 vector (Promega Corporation, Madison, WI, USA) using the *Xho*I and *Hind*III restriction sites within the multiple cloning site.

Cell culture, transfections, and preparation of nuclear extracts

COS7 (monkey kidney cell line), KK-1 (mouse Granulosa cell line), 15P-1 (mouse Sertoli cell line), and Hek293 (human embryonic kidney) cells were maintained in D-MEM supplemented with 10% FBS and 1.8 g/l $NaHCO_3$ and incubated at 37 °C. For nuclear extracts, cells were transfected with 30 μg of either the wild-type or mutant (pCMV6-Entry-*hNR5A1* or pCMV6-Entry-*hNR5A1*-p.(Lys372del)) construct. The transfection reagent Lipofectamine 2000 (Invitrogen, Thermo Fisher Scientific Inc., Waltham, MA, USA) was used for transfections. Cells were harvested 48 h posttransfection and nuclear extracts were prepared using the NE-PER Nuclear and Cytoplasmic Extraction Reagents (Thermo Fisher Scientific Inc.). Protein concentrations were determined using the Pierce BCA Protein Assay Kit (Thermo Fisher Scientific Inc.). Absorbance at 562 nm was measured using the Infinite M200 Pro plate reader (Tecan, Männedorf, Switzerland).

SDS-PAGE and western blot

Fifteen micrograms of nuclear extract was diluted to 30 μl with H_2O, 10 μl of 4 × loading buffer (50% (v/v) glycerol, 0.16 M Tris-HCl (pH 6.8), 2% (w/v) SDS, 0.01% (w/v) bromophenol blue, and 350 mM DTT) was added, and the samples were denatured at 95 °C for 15 min. Total protein of 7.5 μg and 5 μl of Precision Plus Protein Dual Color Standard (Bio-Rad Laboratories, Hercules, CA, USA) were loaded onto a NuPAGE 10% Bis-Tris gel. SDS-PAGE was performed in 1 × NuPAGE MOPS running buffer (Invitrogen) at 100 V for 2 h. After electrophoresis, proteins were blotted onto a PVDF membrane (GE Healthcare Bio-Sciences, Pittsburgh, PA, USA) using 1 × NuPAGE Transfer buffer (Invitrogen) with 20% methanol overnight at 4 °C at 100 mA. Blocking and incubation with primary (1:1000 dilutions) and secondary antibodies (donkey anti-rabbit AlexaFlour 488 and donkey anti-goat AlexaFluor 488; Invitrogen) was performed as described previously. Fluorescence signals were visualized using a Typhoon Trio Variable Mode Imager (GE, Fairfield, CT, USA).

Electrophoretic-mobility shift assays

Fluorescently labelled oligonucleotide (0.5 nmol each) and the corresponding unlabelled synthetic complementary strand, containing the mouse *Nr0b1* SF1-responsive element (other SF1-responsive elements were also tested – data not shown; for oligosequences see Supplementary Table S5), were annealed and diluted to 10 ng/μl in 1 × TE buffer. Fifteen micrograms of either the wild-type or mutant SF1 overexpression nuclear extracts were incubated in 2.5 × binding buffer[21] for 10 min at RT. One microgram of poly(dI–dC) (Roche Diagnostics Pty Ltd, Rotkreuz, Switzerland), 5 μl of a rabbit anti-FLAG antibody (for supershifts) (Sigma-Aldrich, Castle Hill, NSW, Australia), and water were added to a final volume of 48 μl and incubated for 30 min. Two microliters (20 ng) of double-stranded probes were added to a final volume of 50 μl. The samples were incubated for 30 min and the reaction stopped by adding 5 μl of 6 × loading buffer (250 mM Tris-HCl, 40% (v/v) glycerol, and 0.01% (w/v) Orange G). Twenty microliters of each reaction was loaded onto a

7.5% Mini-Protean TGX gel (Bio-Rad) and DNA–protein complexes were separated by electrophoresis at 100 V for 90 min. Fluorescence signals were visualized using a Typhoon Trio Variable Mode Imager (GE).

Dual-luciferase reporter assays

Cells were split onto 96-well plates (Nunc) 12–24 h before transfection and transfected with 75 ng of the reporter construct, pGL4.10-*hNR0B1*, and 15 ng of the *Renilla* construct (pRL-TK) for normalization of transfection efficiency. For the transactivation assay, cells were also co-transfected with 50 ng of either wild-type or mutant pCMV6-Entry-*hNR5A1* construct. To adjust for differences in the amount of DNA added to each reaction, an independent, non-active plasmid (pSLAX) was used to adjust total DNA amounts to 240 ng for each reaction. Lipofectamine 2000 (Invitrogen) was used as a transfection reagent. Cells were harvested 48 h posttransfection and luciferase activity was measured using the Dual-Luciferase Reporter 1000 Assay System Kit (Promega) according to the manufacturer's instructions. Chemiluminescence was measured on an Infinite M200 Pro plate reader (Tecan). Each experiment was performed in biologic triplicate.

RESULTS

Large families are rare in DSD research, because most DSDs result in infertility. As large pedigrees are an extremely powerful tool for genetic linkage analysis, we decided to sequence the entire exome of

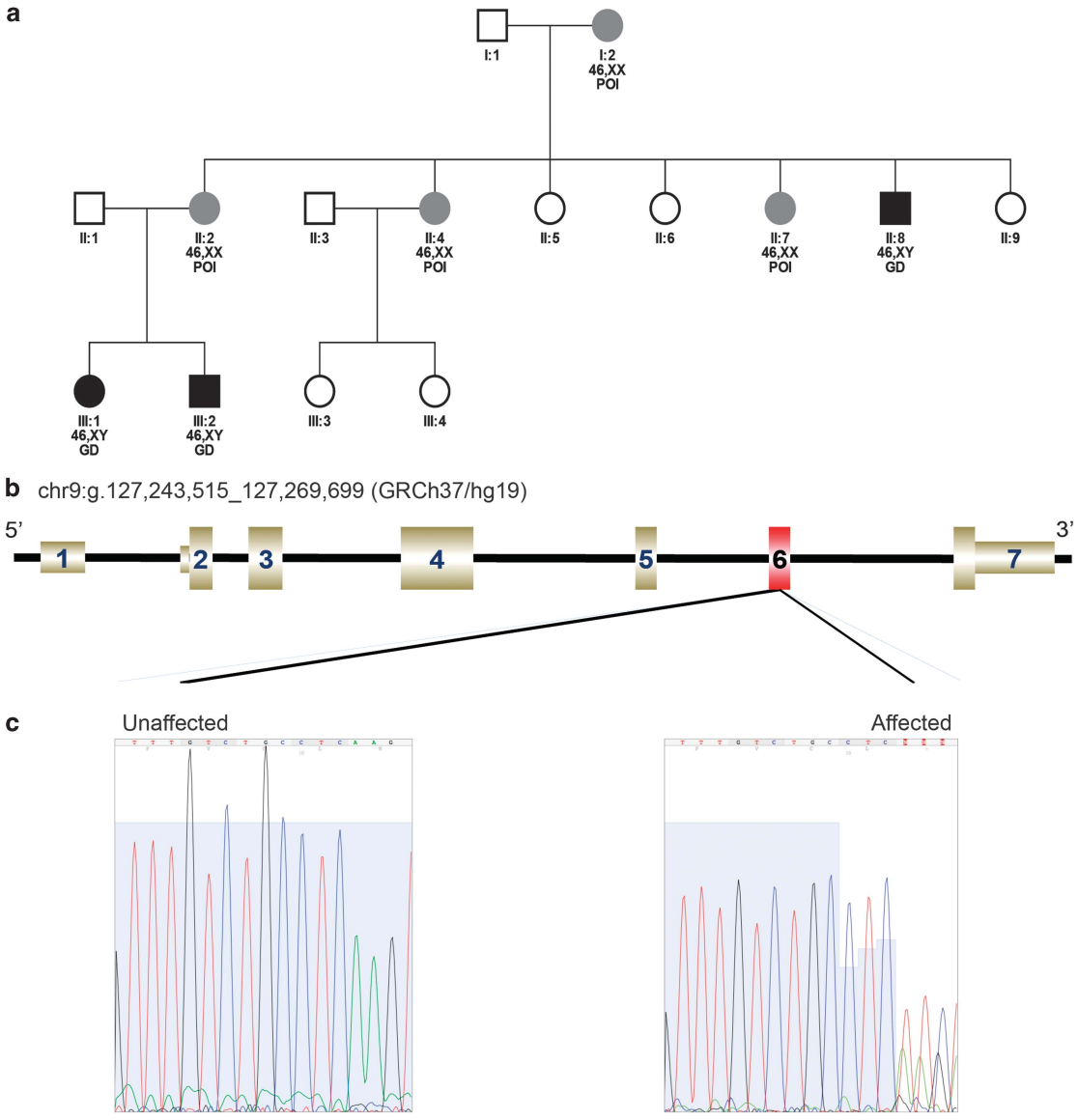

Figure 1 A novel 3-bp deletion in exon 6 of *NR5A1* (exons are numbered as in NG_008176.1) was identified in a multigeneration Indonesian family to be associated with 46,XY DSD and 46,XX POI. (**a**) Pedigree of the family, showing all 15 individuals who were available for this study. Squares and circles represent males and females, respectively. Open symbols indicate unaffected individuals, and filled squares and circles highlight affected individuals.The 46,XY gonadal dysgenesis (46,XY GD) patients were assigned male or female symbols according to their phenotype and assigned gender. The 46,XX POI patients are labelled with 46,XX POI below the symbol. The pedigree has been reduced to the individuals who were available for the study for simplicity and for the sake of confidentiality. (**b**) Schematic view of the *NR5A1* genomic region on chr9:g.127 243 515–127 269 699 (GRCh37/hg19) including the exon–intron structure (*NR5A1* is located on the reverse strand, but shown here in its 5′–3′ orientation from left to right). Exon 6 is highlighted in red. (**c**) Sanger sequencing of an unaffected and affected individual, showing the wild-type exon 6 sequence (unaffected) and the heterozygous, 3-bp deletion (c.1114_1116del) in one affected individual.

all 15 individuals from a multigeneration family from Indonesia (Figure 1a).

Exome capture, sequencing, and bioinformatic analysis

Picard MarkDuplicates software estimated rates of PCR duplication to be very high, ranging from 83.3 to 96.7% of uniquely aligned reads (Supplementary Table S1). As a result, coverage of targeted bases following duplicate removal was very low (Supplementary Table S2), with median coverage ranging between 2 and 9 (median 3).

We attempted to infer genotypes at the locations of 4 071 900 HapMap Phase II and III SNPs. Genotypes were set to missing, unless an SNP was covered by ≥ 5 reads with base and mapping qualities ≥ 13. We discarded SNPs where genotypes were missing for more than 10 of the 15 exomes, resulting in 57 666 SNPs being retained. We then discarded 414 SNPs where genotypes were inconsistent with Mendelian inheritance and 37 685 SNPs where non-missing genotypes did not vary between individuals. Finally, we chose a subset of 4814 SNPs in approximate linkage equilibrium for linkage analysis;[22] on average, CEU individuals are heterozygous for 38% of these markers. Individual genotyping rates for these SNPs ranged between 0.8 and 96.7% (median 21.5%) (Supplementary Table S3).

Linkage analysis identified two peaks on chromosomes 9 and 18 achieving maximum LOD scores of 2.69 and 2.42 (Figure 2). Positive LOD scores were obtained between 128.4 and 159.6 cM (rs1930713–rs4842247) on chromosome 9 and between 29.1 and 48.0 cM (rs767300–rs17448282) on chromosome 18. Given the high number of genotypes missing, we examined the robustness of our results by using MERLIN's – sample option to sample 10 sets of haplotypes according to their likelihood (eg Supplementary Figure S1). This allows assessment of the uncertainty of the haplotype inference; the less an individual's inferred haplotypes vary across samples, the more confidence we have in the haplotyping. For the chromosome 9 peak, all 10 sets of haplotypes were consistent with the autosomal dominant model for 14 of the family members (eg Supplementary Figure S1), indicating high confidence in the results. Inference could not be performed for the remaining individual, affected II:7, for whom all genotypes within the linkage region were missing. For the chromosome 18 peak, only two haplotype sets accorded with the genetic model for 14 family members, indicating that we cannot be very confident of linkage. Unaffected II-3 was missing all genotypes within the linkage region.

Of the 300 591 variants detected in at least one of the 15 exomes, 4453 variants were located within one of the two linkage peaks; 1148 of these were rare (not present or present at a frequency ≤ 0.01 in the 1000 Genomes Project). Of these, 45 variants were predicted to affect protein sequence and/or splicing. Given the low coverage, we did not apply a strict genotype filter; instead, we retained variants that were not detected in the eight unaffected exomes, but were detected in at least two affected exomes. This reduced the list of candidate variants to five (Supplementary Table S4). One of these variants, a 3-bp deletion removing a single amino acid, is located within exon 6 of *NR5A1* (NM_004959.4:c.1114_1116del [p.(Lys372del)]; data have been submitted to the *NR5A1* gene variant database at http://www.LOVD.nl/NR5A1, patient ID 0063642.) (see Figure 1b).

Supplementary Figure S2 shows read alignments at the location of the *NR5A1* variant for affected and unaffected family members with coverage at this location. Sanger sequencing indicated that all seven affected family members were heterozygous for the deletion, whereas the six unaffected family members were all wild type, as were the two girls considered at risk of developing 46,XX POI (III:3 and III:4; see Figure 1b and c).

In silico prediction on the wild-type and mutant SF1 protein

The heterozygous, 3-bp in-frame c.1114_1116del in exon 6 of *NR5A1* results in deletion of the lysine at position 372 (p.(Lys372del)) of SF1, within the predicted ligand-binding domain (LBD; see Figure 3a). Although recent publications have described phosphatidylinositols, sphingosine, and phospholipids as potential ligands for SF1, their biologic role remains to be elucidated.[23–25] To investigate the potential biologic relevance of the p.(Lys372del) variant, we performed an *in silico* prediction on the wild-type and mutant proteins using the online protein comparative modelling server 3D-JIGSAW (http://bmm.cancerresearchuk.org/~3djigsaw/). Direct comparison of the two predicted protein structures (Figure 3a–c) revealed truncation of helix 8 of the LBD at position 372 in the mutant protein. Furthermore, the loss of the β-pleated sheets in the LBD and the DNA-binding domain (DBD) indicated a potential loss of the DNA-binding ability of the mutant protein compared with the wild-type protein. Also, the tertiary structure of the whole protein was predicted to change slightly, indicating a crucial role of Lys372 in the wild-type SF1 protein structure.

Functional assays of SF1 activity

We assessed whether the mutant SF1 protein was correctly translocated into the nucleus. Using immunofluorescence, we confirmed

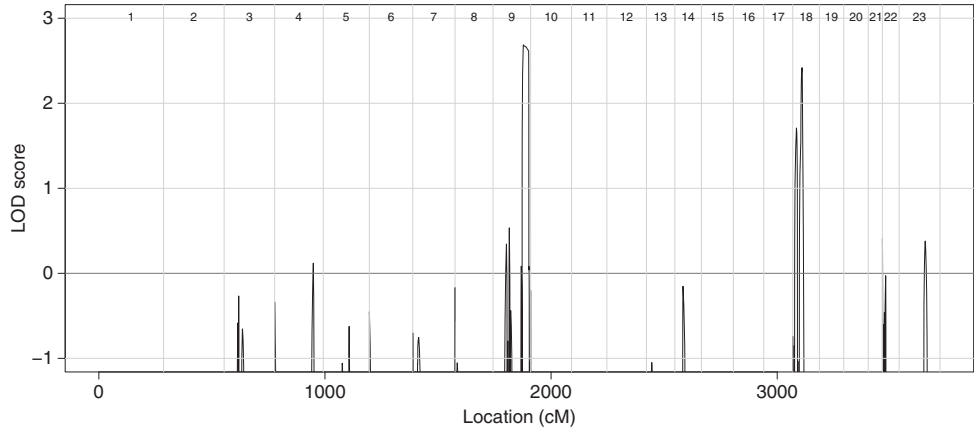

Figure 2 Genome-wide parametric LOD scores under a fully penetrant autosomal dominant inheritance model for the Indonesian family.

that import of the mutant SF1 protein into the nucleus was not affected by the p.(Lys372del) variant (data not shown).

To investigate the potential effect of the p.(Lys372del) variant on the DNA-binding ability of SF1, as predicted by 3D-JIGSAW, we performed electrophoretic mobility shift assays (EMSAs) and dual-luciferase reporter assays using known SF1-responsive elements. Figure 4b shows the results for the wild-type and mutant SF1 proteins, using the *mNr0b1* (*mDax1*) SF1-responsive element.[21] Western blot analysis was used to ensure similar levels of wild-type and mutant SF1 protein, nuclear extracts were loaded (see Figure 4a). The binding of wild-type SF1 to the *mNr0b1* SF1-responsive element yielded two signals as the probe contains two overlapping SF1-binding sites, one on each of the DNA strands. The lower band represents one molecule of SF1 and the second higher band represents two SF1 molecules bound to the probe (see Figure 4b). To confirm that both signals were caused by SF1 binding, we performed supershifts, using an anti-FLAG antibody, which recognizes both the wild-type and mutant SF1 proteins (expressed via the

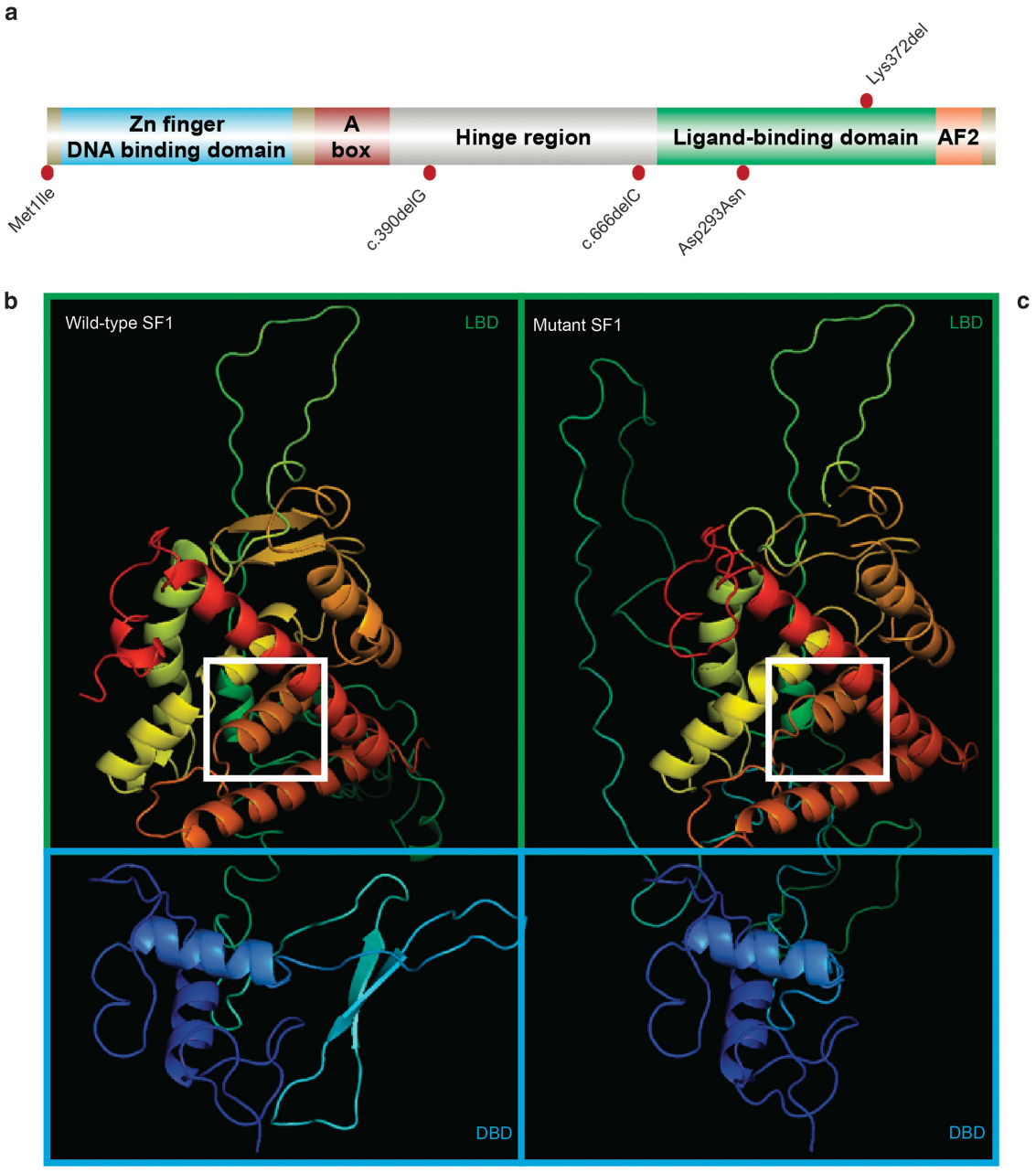

Figure 3 Protein structure of SF1. (**a**) Schematic view of the different domains of the SF1 protein. The approximate locations of previously published variants within *NR5A1* in families affected by 46,XY DSD and 46,XX POI and sporadic cases of 46,XX POI are shown below the protein structure as resulting protein changes.[11,12] The novel 3-bp deletion in exon 6 of *NR5A1* (p.(Lys372del)), located within the predicted LBD, is shown above the protein structure. (**b**) *In silico* predictions of the wild-type and (**c**) the p.(Lys372del) mutant SF1 proteins. The LBD is highlighted by green boxes, whereas the DBD is highlighted by blue boxes. Comparison of the wild-type and mutant SF1 structures indicates that the deletion of Lys372 results in the truncation of helix 8 (highlighted with white boxes) of the LBD at this position. Further changes in the secondary structure of the protein include the loss of the β-pleated sheets in both the LBD and the DBD, indicating a potential loss of the DNA-binding ability of the mutant compared with the wild-type SF1 protein.

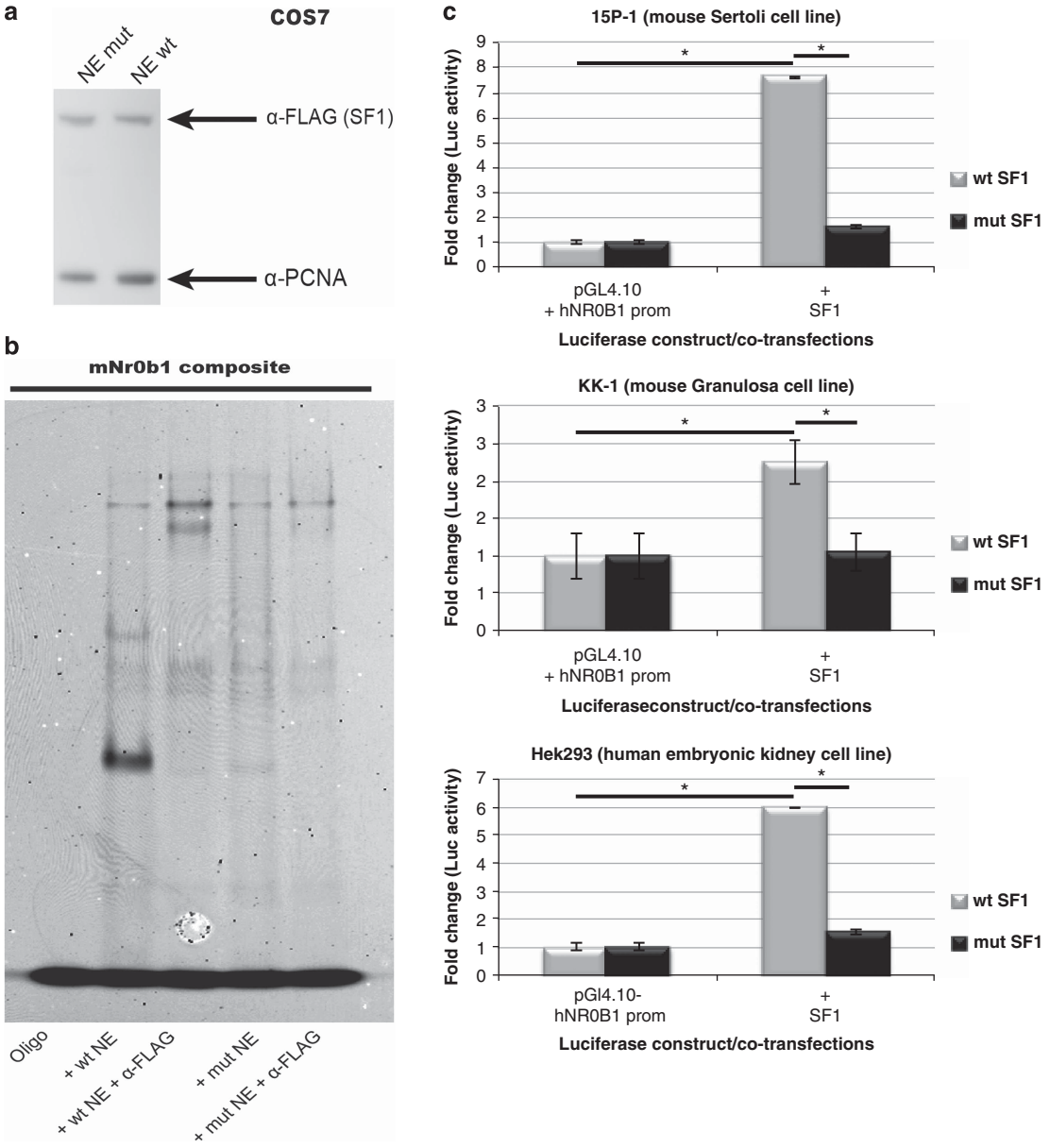

Figure 4 Functional assays on SF1 activity using the *mNr0b1* and *hNR0B1* SF1-responsive element and promoter. (**a**) Western blot of the wild-type and mutant SF1 overexpression nuclear extracts (NE mut and NE wt, respectively) showing similar expression levels of the two proteins. Proliferating cell nuclear antigen (PCNA) was used as loading control. (**b**) EMSA using the mouse *Nr0b1* (*mDax1*) SF1-responsive element. Lane one shows the signal for the probe, and lane two shows the signals obtained by adding the wild-type nuclear extract (wt NE). Two signals are detected owing to the composite SF1-binding site within this element. Lane three shows the supershifts, using an anti-FLAG antibody. Lanes 4 and 5 show the results for the mutant nuclear extract (mut NE). Both signals and supershifts, owing to SF1 binding to the probe, which were observed with the wild-type nuclear extract, are almost completely lost when the mutant nuclear extract is added. (**c**) Dual-luciferase reporter assays using the *hNR0B1*promoter. In all three cell lines (15P-1, KK-1, and Hek293), an increase of luciferase expression can be detected when the wild-type SF1 expression vector is added to the reporter vector (pGL4.10-hNR0B1), whereas addition of the mutant construct results in little or no change in luciferase activity.

pCMV6-Entry-*hNR5A1* vector, which contains C-terminal DDK and c-Myc tags). We obtained two supershift signals, caused by one or two SF1 molecules bound to the responsive element. The shift and supershift experiments with the mutant SF1 overexpression nuclear extracts produced one very weak shift and supershift signal (representing one SF1 molecule bound), and loss of the second signal (two SF1 molecules bound) (see Figure 4b). This indicates that the p.(Lys372del) mutant protein bound with a lower efficiency than the wild-type protein to SF1-responsive elements *in vitro*.

To confirm these results, we performed dual-luciferase reporter assays using the *hNR0B1* (*hDAX1*) promoter (Figure 4c). When co-transfecting the pGL4.10-*hNR0B1* promoter vector with the wild-type pCMV6-Entry-*hNR5A1* overexpression vector, we found a 2- to 7.5-fold increase in luciferase activity compared with the reporter construct-only transfections. The fold increase in the luciferase activity varied between the three cell lines (15P-1, KK-1, and Hek293) tested. When transfecting with the mutant overexpression construct (pCMV6-Entry-*hNR5A1*-p.(Lys372del)), luciferase activity

remained at similar levels to the promoter construct-only transfections. These data support the results from the EMSAs, where the mutant SF1 protein did not bind SF1-responsive elements *in vitro*. Furthermore, the luciferase data showed that the mutant SF1 protein failed to activate SF1-responsive promoters *in vitro*.

DISCUSSION

In this study, we identified a novel, 3-bp, in-frame deletion located in exon 6 of *NR5A1* in all seven affected family members (46,XY DSD (unknown and gonadal dysgenesis) and 46,XX POI). As *NR5A1* had been previously described to cause both phenotypes[11,12] and was inherited in the proposed autosomal dominant manner, we focussed on this variant as the most likely explanation for these clinical phenotypes.

In silico prediction of the wild-type and mutant predicted SF1 protein structures suggested that the mutated lysine 372 has an important structural role. Deletion of this single amino acid is predicted to result in the truncation of helix 8 of the LBD and the loss of all β-pleated sheets, in the LBD, as well as in the DBD. It is also predicted to affect the entire tertiary protein structure, potentially explaining the observed loss of the DNA-binding capability of the mutant SF1 protein. The EMSAs and dual-luciferase reporter assays with known SF1-responsive elements of downstream targets of SF1, such as *mNr0b1/hNR0B1* (*mDax1/hDAX1*), confirmed the *in silico* prediction, showing a marked reduction in the DNA-binding and transactivation ability of the mutant SF1 protein compared with the wild-type SF1 protein *in vitro*. This indicates that the mutant SF1 protein present in the affected family members fails to bind and activate downstream targets of SF1.

Three other variants within the LBD of *NR5A1* have been described in sporadic cases of 46,XX POI (c.691_699delCTGCAGCTG (p.(Leu231_Leu233del)) and c.704C>T (p.(Pro235Leu))) or familial cases of 46,XY DSD and 46,XX POI (c.877G>A (p.(Asp293Asn))).[11,12] A direct comparison of the three other published variants and our variant is difficult, as different SF1-responsive promoters and cell lines were used; however, it appears that single amino-acid exchanges within the LBD lead to a moderate (c.877G>A (p.(Asp293Asn)) (*Cyp11a1* and *Cyp19a1* promoters); c.704C>T (p.(Pro235Leu)) (*Cyp17a1* promoter)) or no significant reduction (c.704C>T (p.(Pro235Leu)) (*CYP11A1* and *HSD3B2* promoters) of the activation of SF1-responsive promoters. By contrast, the effect on the activation of SF1-responsive promoters of the single amino-acid deletion c.1114_1115del (p.(Lys372del)) described here corresponds to the effect seen with larger deletions c.691_699delCTGCAGCTG (p.(Leu231_Leu233) (*Cyp11a1* and *Cyp19a1* promoters)) or variants found in the DBD. This further supports the *in silico* prediction and the proposed structural importance of lysine 372.

The two DSD conditions (46,XY DSD (unknown and gonadal dysgenesis) and 46,XX POI) caused by the same variant can be explained by the different roles of SF1 in the developing testis and ovary. In testes, SF1 is expressed in Leydig and Sertoli cells and is required to upregulate *SRY* and thus to establish the male pathway, resulting in the bipotential gonad developing into a testis.[26] SF1 has also been shown to synergize with SRY to upregulate *SOX9* and with *SOX9* to upregulate *Cyp26b1*.[6,27] In addition, SF1 is required to upregulate *AMH* expression and causing the regression of the Müllerian ducts in XY male embryos.[28] In the ovary, *NR5A1* is expressed in theca and granulosa cells and its expression can be detected in the embryonic, postnatal, prepubertal and adult ovary.[29,30] SF1 has been shown to regulate key genes of ovarian steroidogenesis. SF1 clearly acts at multiple points during gonad development, differentiation, and in crucial steroidogenic pathways even into adulthood. Given this, it is not surprising that the same variant in *NR5A1* can cause testis failure and 46,XY DSD (including gonadal dysgenesis) in male family members, whereas in female relatives it causes 46,XX POI.

We also observed that the same variant caused different external genitalia in the three individuals with 46,XY DSD. II:8 and III:2 had predominantly male genitalia and were raised as boys, whereas III:1 had predominantly female external genitalia at birth and was raised as a girl. This variation is most likely explained either by incomplete penetrance and/or by SNPs or variants in modifying genetic factors that have yet to be identified. MPS studies on larger numbers of DSD patients will help identify the missing genes and networks, which, when disrupted, cause DSDs. Our findings in this family are entirely consistent with earlier studies, which associated variants in *NR5A1* with 46,XY gonadal dysgenesis and 46,XX POI in families. High variability in the phenotype of the external genitalia in familial cases of 46,XY DSD has also been reported.[31]

This study has shown the particular utility of using low-coverage MPS and linkage analysis to identify rapidly the causative variant. In an effort to reduce costs, many groups select a few individuals from large pedigrees for MPS at high coverage. As an alternative, the approach used here involved sequencing more individuals to ensure sufficient power for linkage analysis, combined with cheaper low-coverage MPS to identify the disease-causing genes at a similar cost. The identification of the molecular cause and the mode of inheritance of 46,XY DSD (including cases of 46,XY gonadal dysgenesis) and 46,XX POI in families such as this allows for the provision of genetic counselling to female family members under or at a reproductive age, as they may carry a dominant variant, but have not presented with secondary amenorrhoea. The two girls (III:3 and III:4) in this family are in this situation. Although their mother (II:4) carried the variant in *NR5A1* and was affected by 46,XX POI, both her daughters tested negative for the variant and were not at risk for POI.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank the family for their participation in this study and the staff at the Center for Biomedical Research in Semarang, Indonesia, for sample and data collection. We thank Craig Smith for critical proof reading of the manuscript and the Ian Potter Centre for Personalised Genomics for use of equipment. This work was supported by the Australian National Health and Medical Research Council (Program grant number 546517 to AS, PK VRH and IRIISS, Program grant number 490037 to MB), the Helen Macpherson Smith Trust (Partnership grant 6846 to AS), the Australian Research Council (Future Fellowship 100100764 to MB), The Pratt Foundation (KRS), The University of Melbourne (MIRS to SE), The Australian Government, Department of Innovation, Industry, Science and Research (IPRS to SE), and the Victorian Government's Operational Infrastructure Support Program.

1 Hughes IA, Houk C, Ahmed SF, Lee PA: Consensus statement on management of intersex disorders. *J Pediatr Urol* 2006; **2**: 148–162.
2 Ono M, Harley VR: Disorders of sex development: new genes, new concepts. *Nature reviews. Endocrinology* 2013; **9**: 79–91.
3 Eggers S, Sinclair A: Mammalian sex determination – insights from humans and mice. *Chromosome Res* 2012; **20**: 215–238.
4 Nelson LM: Clinical practice. Primary ovarian insufficiency. *N Engl J Med* 2009; **360**: 606–614.
5 Ferrarini E, Russo L, Fruzzetti F *et al*: Clinical characteristics and genetic analysis in women with premature ovarian insufficiency. *Maturitas* 2013; **74**: 61–67.

6 Sekido R, Lovell-Badge R: Sex determination involves synergistic action of SRY and SF1 on a specific Sox9 enhancer. *Nature* 2008; **453**: 930–934.
7 Hoyle C, Narvaez V, Alldus G, Lovell-Badge R, Swain A: Dax1 expression is dependent on steroidogenic factor 1 in the developing gonad. *Mol Endocrinol* 2002; **16**: 747–756.
8 Shen WH, Moore CC, Ikeda Y, Parker KL, Ingraham HA: Nuclear receptor steroidogenic factor 1 regulates the mullerian inhibiting substance gene: a link to the sex determination cascade. *Cell* 1994; **77**: 651–661.
9 Achermann JC, 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**: 125–126.
10 Paliwal P, Sharma A, Birla S, Kriplani A, Khadgawat R: Identification of novel SRY mutations and SF1 (NR5A1) changes in patients with pure gonadal dysgenesis and 46,XY karyotype. *Mol Hum Reprod* 2011; **17**: 372–378.
11 Lourenço D, Brauner R, Lin L *et al*: Mutations in NR5A1 associated with ovarian insufficiency. *N Engl J Med* 2009; **360**: 1200–1210.
12 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–E1306.
13 Li H: A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 2011; **27**: 2987–2993.
14 Li H, Ruan J, Durbin R: Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res* 2008; **18**: 1851–1858.
15 The 1000 Genomes Project ConsortiumAbecasis GR, Auton A, Brooks LD *et al*: An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012; **491**: 56–65.
16 Wang K, Li M, Hakonarson H: ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010; **38**: e164.
17 Kumar P, Henikoff S, Ng PC, Kumar P, Henikoff S, Ng PC: Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protocols* 2009; **4**: 1073–1081.
18 Adzhubei IA, Schmidt S, Peshkin L *et al*: A method and server for predicting damaging missense mutations. *Nat Methods* 2010; **7**: 248–249.
19 Smith KR, Bromhead CJ, Hildebrand MS *et al*: Reducing the exome search space for Mendelian diseases using genetic linkage analysis of exome genotypes. *Genome Biol* 2011; **12**: R85.
20 Abecasis GR, Cherny SS, Cookson WO, Cardon LR: Merlin – rapid analysis of dense genetic maps using sparse gene flow trees [see comment]. *Nat Genet* 2002; **30**: 97–101.
21 Ito M, Achermann JC, Jameson JL: A naturally occurring steroidogenic factor-1 mutation exhibits differential binding and activation of target genes. *J Biol Chem* 2000; **275**: 31708–31714.
22 Bahlo M, Bromhead CJ: Generating linkage mapping files from Affymetrix SNP chip data. *Bioinformatics* 2009; **25**: 1961–1962.
23 Krylova IN, Sablin EP, Moore J *et al*: Structural analyses reveal phosphatidyl inositols as ligands for the NR5 orphan receptors SF-1 and LRH-1. *Cell* 2005; **120**: 343–355.
24 Sablin EP, Blind RD, Krylova IN *et al*: Structure of SF-1 bound by different phospholipids: evidence for regulatory ligands. *Mol Endocrinol* 2009; **23**: 25–34.
25 Urs AN, Dammer E, Sewer MB: Sphingosine regulates the transcription of CYP17 by binding to steroidogenic factor-1. *Endocrinology* 2006; **147**: 5249–5258.
26 de Santa Barbara P, Mejean C, Moniot B, Malcles MH, Berta P, Boizet-Bonhoure B: Steroidogenic factor-1 contributes to the cyclic-adenosine monophosphate down-regulation of human SRY gene expression. *Biol Reprod* 2001; **64**: 775–783.
27 Kashimada K, Svingen T, Feng CW *et al*: Antagonistic regulation of Cyp26b1 by transcription factors SOX9/SF1 and FOXL2 during gonadal development in mice. *FASEB J* 2011; **25**: 3561–3569.
28 Arango NA, Lovell-Badge R, Behringer RR: Targeted mutagenesis of the endogenous mouse Mis gene promoter: *in vivo* definition of genetic pathways of vertebrate sexual development. *Cell* 1999; **99**: 409–419.
29 Luo X, Ikeda Y, Parker KL: A cell-specific nuclear receptor is essential for adrenal and gonadal development and sexual differentiation. *Cell* 1994; **77**: 481–490.
30 Jeyasuria P, Ikeda Y, Jamin SP *et al*: Cell-specific knockout of steroidogenic factor 1 reveals its essential roles in gonadal function. *Mol Endocrinol* 2004; **18**: 1610–1619.
31 Ciaccio M, Costanzo M, Guercio G *et al*: Preserved fertility in a patient with a 46,XY disorder of sex development due to a new heterozygous mutation in the NR5A1/SF-1 gene: evidence of 46,XY and 46,XX gonadal dysgenesis phenotype variability in multiple members of an affected kindred. *Hormone Res Paediatr* 2012; **78**: 119–126.
32 Thiele H, Nürnberg P: HaploPainter: a tool for drawing pedigrees with complex haplotypes. *Bioinformatics* 2005; **21**: 1730–1732.
33 Robinson JT, Thorvaldsdottir H, Winckler W *et al*: Integrative genomics viewer. *Nat Biotechnol* 2011; **29**: 24–26.

Supplementary Information accompanies this paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)