
8

Mutations in Swi/SNF chromatin remodeling complex gene ARID1B cause Coffin-Siris syndrome

Gijs W.E. Santen, Emmelien Aten, Yu Sun, Rowida Almomani, Christian Gilissen, Maartje Nielsen, Sarina G. Kant, Irina N. Snoeck, Els A.J. Peeters, Yvonne Hilhorst-Hofstee, Marja W. Wessels, Nicolette S. den Hollander, Claudia A.L. Ruivenkamp, Gert-Jan B. van Ommen, Martijn H. Breuning, Johan T. den Dunnen, Arie van Haeringen, Marjolein Kriek

Nature Genetics. 2012;44(4):379-80

Abstract

We identified *de novo* truncating mutations in ARID1B in three individuals with Coffin-Siris syndrome (CSS) by exome sequencing. Array-based copy-number variation (CNV) analysis in 2,000 individuals with intellectual disability revealed deletions encompassing ARID1B in three subjects with phenotypes partially overlapping that of CSS. Together with published data, these results indicate that haploinsufficiency of the ARID1B gene, which encodes an epigenetic modifier of chromatin structure, is an important cause of CSS and is potentially a common cause of intellectual disability and speech impairment.

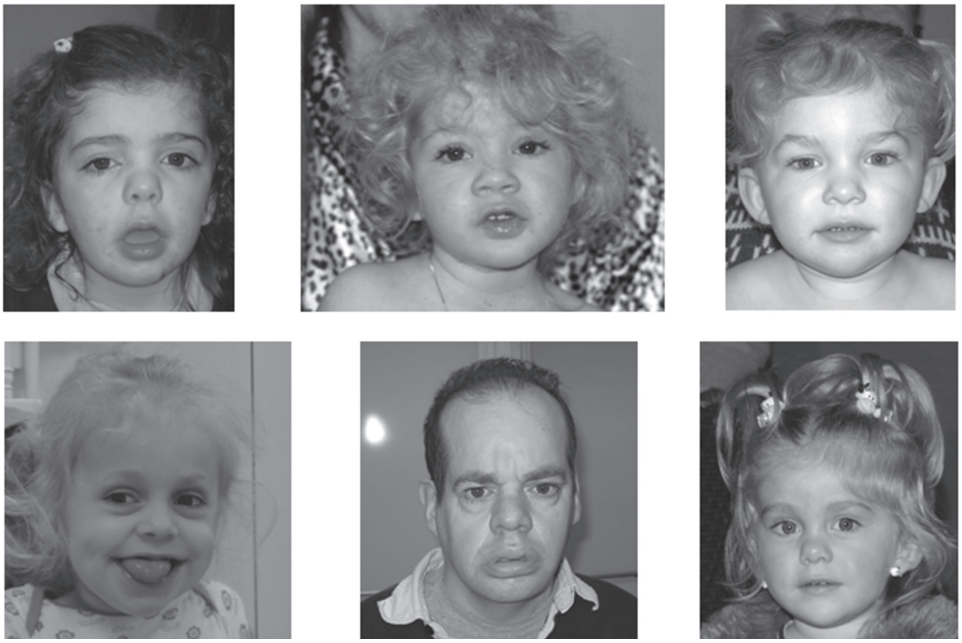
Main text

Coffin-Siris syndrome (OMIM %135900, CSS)[1] is characterized by developmental delay, severe speech impairment, coarse facial features, hypertrichosis, hypoplastic or absent fifth finger- or toe nails[2] and agenesis of the corpus callosum (Supplementary Table 1). Few published patients fulfill the complete spectrum of the CSS phenotype and it is debated whether all CSS patients have the same syndrome. CSS is generally assumed to display autosomal recessive inheritance, although autosomal dominant inheritance has not been formally excluded[3, 4].

To identify the genetic cause of CSS we performed whole exome sequencing in one patient-parent trio and two sporadic patients with a clinical CSS diagnosis all diagnosed in one hospital by the same clinical geneticist (Fig. 1, Supplementary note, Supplementary Fig. 1, and Supplementary Tables 1 and 2, Supplementary methods). Exome sequencing data are available upon request. Using the GATK sequence analysis pipeline[5, 6] we identified 12,722- 14,642 exonic/splice site variants per individual. Filtering steps using variant databases (dbSNP132 and the 1000 genomes project database) and a selection for coding regions, revealed variants in 34 genes that were shared by all three patients. Filtering for recessive inheritance (discarding all genes with only one heterozygous variant) showed that no gene in agreement with a recessive inheritance model was shared between the patients. Accepting dominant inheritance, we queried heterozygous and *de novo* variants and identified *ARID1B* as the only possible affected gene (Supplementary Table 3). All variants truncate the *ARID1B* reading frame (two nonsense variants -c.5329A>T,p.Lys1777X and c.3223C>T,p.Arg1075X), one frameshift -c.4619_4628del,p.Gln1541ArgfsX35-, Table 1). The mutations were validated using Sanger sequencing and shown to occur *de novo* in all three patients (Supplementary Fig. 2). Since thus far an autosomal recessive inheritance could not be ruled out, the parents of a CSS patient received a recurrence risk of 10%[7]. The identification of *de novo* mutations in *ARID1B* in CSS patients, allowed us to reduce this risk to 1-2%[8].

When we queried our in-house database for patients with potential copy number variants (CNVs) including *ARID1B* we identified three patients with a deletion in 2000 subjects screened for intellectual disability (Fig.1, Supplementary Fig. 1 and 2, Supplementary Table 1). This cohort consists of patients with intellectual disability and/or congenital malformations (syndromic and non-syndromic) referred for array-based CNV analysis. In comparison, we found six patients with the relatively frequent 22q11.2 duplication in this cohort. Patient four has a *de novo* 2.72 Mb deletion of band 6q25 encompassing *ARID1B*, *C6orf35*, *ZDHHC14*, *SNX9* and *SYNJ2*.

Figure 1: Facial features of all patients. Top, left to right: Patient 1 at 4.5 years, patient 2 at 2,5 years, patient 3 at 3 years. Bottom, left to right: Patient 4 at 3.5 years , patient 5 at 45 years, patient 6 at 3,5 years. All patients share coarse facial features, thick eyebrows and broad nasal tips. For further details see Supplementary Table 1. The parents or legal guardians of all affected individuals gave consent for publication of the clinical photographs.



Patient five has a *de novo* 0.73 Mb deletion encompassing *ARID1B, C6orf35* and *ZDHHC14*. Patient six has a *de novo* deletion of 0.88 Mb encompassing the same genes as patient five. Analysis of the phenotypes of these patients shows that, similar to our CSS patients, they have moderate to severe intellectual disability and severe speech delay. These patients also share facial similarities with the CSS patients (Figure 1) but lack the typical CSS abnormalities, such as hypoplastic or absent finger- or toenails. The CSS diagnosis was considered only in patient four because of her hypoplastic fingernails (Supplementary Table 1). Halgren *et al.*[9], describe eight patients with haploinsufficiency of *ARID1B* (in one subject the gene was disrupted by a reciprocal translocation). An additional disruption of *ARID1B* as well as a *de novo* intragenic deletion have been reported in patients with either corpus callosum agenesis or autism[9, 10]. Although the sizes of the published deletions range

Table 1: Overview of the detected variants and CNVs affecting ARID1B.

Patient (DECIPHER ID)	Chr	Position	Size of deletion	cDNA position	Protein Position	Exon	phyloP score
1	6	g.157527604A>T	-	c.5329A>T	p.Lys1777X	20	2.9
2	6	g.157502190C>T	-	c.3223C>T	p.Arg1075X	12	1.9
3	6	g.157522346_157522356del	-	c.4619_4628del	p.Gln1541ArgfsX35	18	3.0
4 (248472)	6	155,797,565 - 158,517,307	2.72 Mb	-	-		-
5 (250455)	6	157,079,676 - 157,806,675	0.73 Mb	-	-		-
6 (257917)	6	157,144,644 - 158,028,969	0.88 Mb	-	-		-

Variants are mapped to the hg19 reference genome. Deletion size represents the maximum deletion. All variants occurred *de novo*. Chr.,chromosome. Subjects 4 and 5 were published previously [9].

from <1 Mb to > 14 Mb, the phenotypes of the patients largely overlap with those of our CSS patients, with the exception of the typical fifth finger CSS abnormalities. Disruption of *ARID1B* therefore seems to be the main driver of the observed phenotype and CSS should be considered in all patients with intellectual disability and speech impairment, particularly in combination with agenesis of the corpus callosum. Based on these findings we conclude that haploinsufficiency of *ARID1B* is likely to be an important cause of CSS.

The public database DECIPHER (http://decipher.sanger.ac.uk) contains 12 patients with haploinsufficiency of *ARID1B* (including the three described here and those described by Halgren *et al.*). Since all deletions and mutations published thus far have been *de novo*, disease penetrance is expected to be high. No truncating or splice-site mutations are reported in the 1000 genomes project, nor in the ~ 5400 exomes on the Exome Variant Server (http://evs.gs.washington.edu/EVS/). However, the Database of Genomic Variants harbors two deletions including exon 1 and exon 2-20 of the *ARID1B* respectively. This could signify reduced penetrance, but an alternative explanation would be that these deletions represent a technical artifact. Both deletions were not validated using alternative methods.

Together with *ARID1A*, *ARID1B* encodes for ARID (AT rich interactive domain), a subunit of the BAF complex. BAF (Brahma associated factor) is one of the two main components of the Switch/sucrose nonfermentable (SWI/SNF)-like chromatin remodeling complex. They act as epigenetic modifiers by altering the structure of chromatin to facilitate access of

transcription factors to DNA. ARID1A and ARID1B proteins have antagonistic functions and they are both important for the regulation of the cell cycle. Although *ARID1B* is predominantly expressed in differentiated cell-types, it has also been suggested to be involved in early development of the brain[11]. *ARID1A* is more abundantly expressed in embryonic tissue and somatic mutations have recently been related to gastric cancer[12]. Recently, Gilissen *et al*[13] remarked that histone modifying proteins seem to have a dual role in developmental disorders and malignancies. One could hypothesize that activating mutations in *ARID1A* might give a clinical phenotype similar to CSS.
In conclusion, CSS can be added to a growing list of syndromes characterized by congenital malformations, and intellectual disability that are caused by mutations in epigenetic genes that encode modifiers of chromatin structure[14, 15].

Accession code. Data for *ARID1B* is deposited in RefSeq under the accession code NM_020732.3.

Acknowledgements

We thank the patients and their parents for participation in this study. We would like to acknowledge Alexander Hoischen for his useful suggestions regarding the interpretation of exome sequencing data. We received funding from the EU FP7 framework program agreements 223026 (NMD-chip) and 223143 (TechGene).

Author contributions

G.W.E.S. analyzed the data and wrote the manuscript. E.A., G.J.v.O, M.H.B., J.T.d.D, A.v.H and M.K. conceived and designed the experiments. E.A., Y.S. and R.a.M. performed the experiments. C.G. contributed analysis tools. M.N., S.K., I.S., E.A.J.P, M.W.W., N.S.d.H, Y.H-H and A.v.H clinically characterized the patients. C.A.L.R. analysed SNP-array data. A.v.H. selected the patients for sequencing. A.v.H. and M.K. jointly supervised the research. All authors contributed to the final manuscript.

Competing financial interests

The authors declare no competing financial interests.

1. Coffin,G.S. & Siris,E. *Am. J. Dis. Child.* **119**, 433-439 (1970).

2. Fleck,B.J., Pandya,A., Vanner,L., Kerkering,K., & Bodurtha,J. *Am. J. Med. Genet.* **99**, 1-7 (2001).

3. Carey,J.C. & Hall,B.D. *Am. J. Dis. Child.* **132**, 667-671 (1978).

4. Haspeslagh,M., Fryns,J.P., & van den Berghe,H. *Clin. Genet.* **26**, 374-378 (1984).

5. DePristo,M.A. *et al. Nat. Genet.* **43**, 491-498 (2011).

6. McKenna,A. *et al. Genome Res.* **20**, 1297-1303 (2010).

7. Levy,P. & Baraitser,M. *J. Med. Genet.* **28**, 338-341 (1991).

8. Harper,P.S. *Practical Genetic Counseling*(Edward Arnold (Publishers) Ltd.,2004).

9. Halgren,C. *et al. Clin. Genet.* **DOI:10.1111/j.1399-0004.2011.01755.x**, (2011).

10. Backx,L., Seuntjens,E., Devriendt,K., Vermeesch,J., & Van,E.H. *Cytogenet. Genome Res.* **132**, 135-143 (2011).

11. Flores-Alcantar,A., Gonzalez-Sandoval,A., Escalante-Alcalde,D., & Lomeli,H. *Cell Tissue Res.* **345**, 137-148 (2011).

12. Wang,K. *et al. Nat. Genet.* **43**, 1219-1223 (2011).

13. Gilissen,C., Hoischen,A., Brunner,H.G., & Veltman,J.A. *Genome Biol.* **12**, 228 (2011).

14. Hargreaves,D.C. & Crabtree,G.R. *Cell Res.* **21**, 396-420 (2011).

15. Bokhoven,v.H. *Annu. Rev. Genet.* **45**, 81-104 (2011).

Supplementary information

The parents of patients 1-3 gave informed consent to participate in this study. The parents or legal guardians of all patients gave consent for publication of the clinical photographs. All DNA samples were isolated from peripheral blood leucocytes according to standard techniques.

Samples for next generation sequencing were prepared according to the manufacturer's instructions. Target enrichment was performed using Agilent's 50 Mb Sureselect v2 exome capture kit. Paired-end reads with a length of 100 nucleotides were sequenced on Illumina's Hiseq 2000. Sequencing statistics are summarized in Supplementary Table 2. BWA[1] was used to align the short reads to the reference genome (hg19), followed by GATK[2] base quality score recalibration, INDEL realignment and duplicate removal. SNP and INDEL discovery and genotyping for each sample were performed separately using standard hard filtering parameters for INDELS and variant quality score recalibration[3] for SNP calls as indicated on the GATK website (http://www.broadinstitute.org/gsa/wiki/index.php/Best_Practice_Variant_Detection_with_the_GATK_v3). Because of the low coverage of our data two modifications were made: (1) the minimum quality for variant calling was set at 20 (default 30) and (2) the minimum number of reads supporting an indel was set at 3 (default 6). Variants were annotated using Seattleseq (http://gvs.gs.washington.edu/) and in-house developed software.

→

Supplementary Table 1: Patient characteristics. Present +, Absent -, not tested/observed: ?. Patients 1-3 fulfill the diagnostic criteria for CSS since they have either fifth ray abnormalities or hypoplastic phalanges in their toes. Patients with deletions of *ARID1B* were not noticed to have these abnormalities, the main reason why they were not diagnosed with CSS. In retrospect a subtle degree of fifth finger/nail hypoplasia is discernable in patient 5.

Patient number	Truncating mutation			Deletion		
	1	2	3	4	5	6
Gender	female	female	female	female	male	female
Age at presentation	4.5 years	6 years	3 years	2 years	46 years	3.5 years
Mutation/CNV *ARID1B*	mutation	mutation	mutation	deletion	deletion	deletion
Inheritance of mutation	*de novo*	*de novo*	*de novo*	*de novo*	*de novo*	*de novo*
Birth (gestational age)	42	40	39	39+4	At term	41+1
Birth weight (g)	3370	3720	normal	2770	3800	2722
OFC at birth (cm)	?	36.9	?	?	?	35
Intellectual disability (mild, moderate, severe)	severe	severe	moderate	severe	severe	moderate
Speech delay (mild, moderate, severe)	severe	severe	severe	no speech	no speech	moderate
Coarse facial features	+	+	+	+	+	-
Thick eyebrows	+	+	+	+	+	+
Low frontal hairline	+	+	+	-	-	+
Hypertrichosis	+	+	+	-	-	-
Joint laxity	+	+	-	-	+	?
Brachydactyly	+	+	-	-	-	-
Brachydactyly fifth finger	+	+	-	-	-	-
Hypoplastic nails	-	-	-	+	-	-
Hypoplastic nail fifth finger	-	+	+	-	-	-
X-ray hand/foot performed	+	+	+	-	-	-
Missing/hypoplastic phalanx of toes	-	+	+	?	?	?
MRI performed	+	+	+	-	-	+
Agenesis of the corpus callosum	+	partial	+	?	?	+
Height (SDS) at last visit	-2,5	-1	0	-1,5	-3,5	-2.6
Weight (SDS) at last visit	0	0	+1	-3,0	0	0.3
OFC (SDS) at last visit	0	+1	+1	0	-0,75	-0.2
Other remarkable features	colpocephaly; autism; strabismus	fetal finger pads	Colpocephaly; strabismus	atrial septum defect; sparse hair; no nail growth; unilateral single palmar crease; nephrocalcinosis at 4 years	anal atresia; double ureters.	pectus excavatum; abnormal palmar creases; striped toe nails

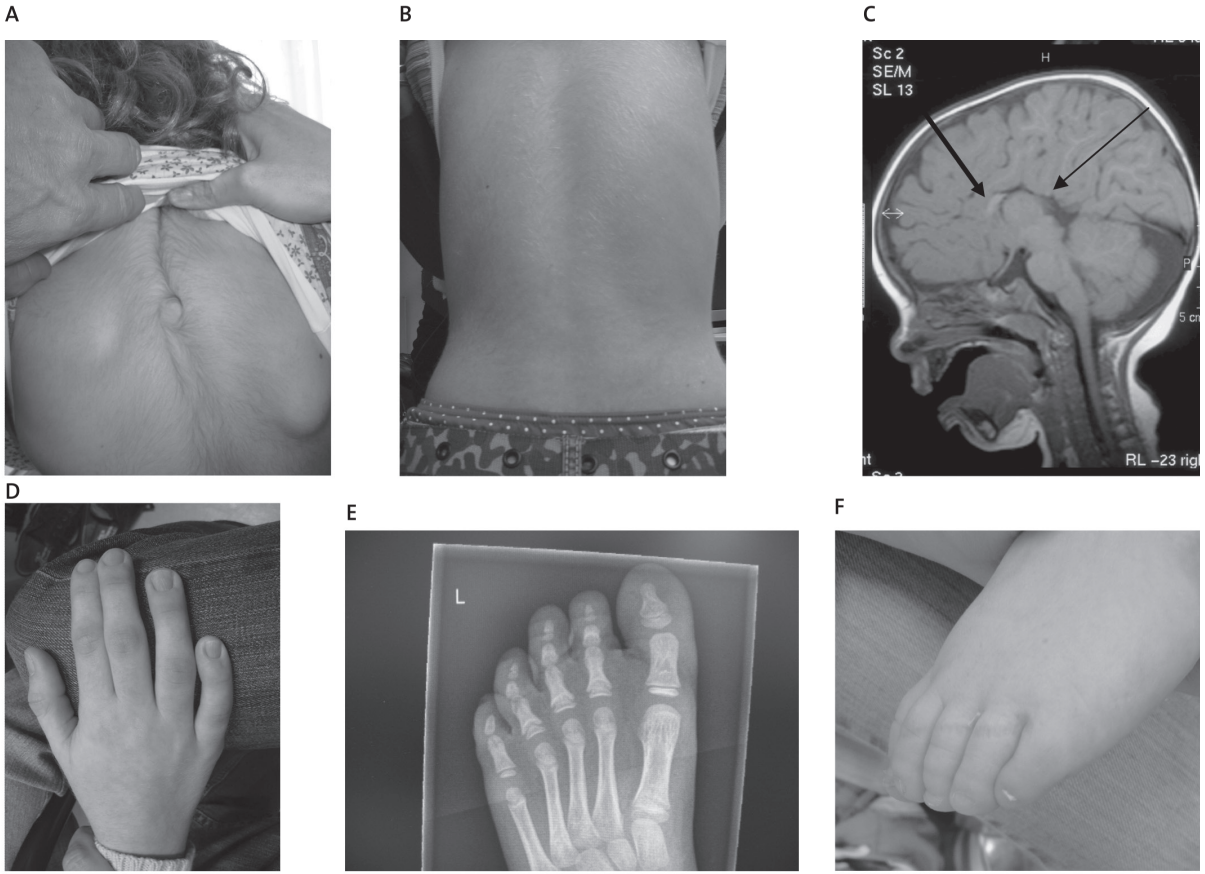
Supplementary Table 2: Sequencing statistics (patients 1, 2 and the parents of patient 1 were sequenced twice).

Patient	1	2	3	Father patient 1	Mother patient 1
Total number of sequenced reads	67,491,568	119,409,964	36,045,692	108,724,122	74,764,116
Aligned reads	64,567,654	114,400,816	35,140,431	104,347,767	71,874,327
Properly paired reads	63,859,108	112,053,996	34,476,740	103,155,322	71,167,452
Mean target coverage	19.5x	34.6x	14.8x	27.7x	21x
% Target bases covered >10x	56	61	57	61	57

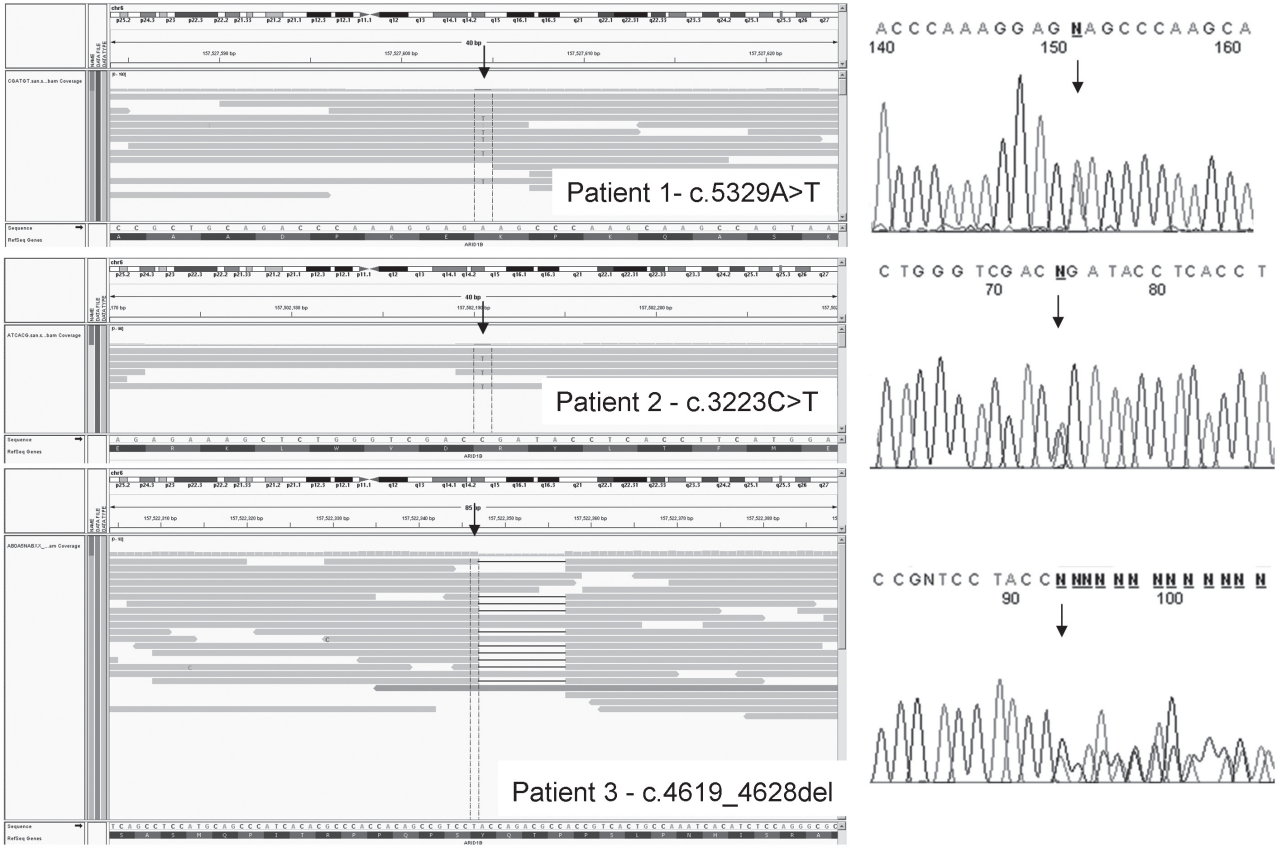
Supplementary Table 3. Filtering steps in 3 CSS patients. Filtering for recessive inheritance did not result in any genes shared by the patients. The number of *de novo* variants in patient one appears high; for six of these variants the coverage in either parent was low (<8). Two variants appear to be inherited (but a slightly different insertion was called in the parents). One 24 bp deletion is called in *SPRR3* but is present in only 3/14 reads. The only compelling variant remaining is the nonsense variant in *ARID1B*.

Filter	Patient 1	Patient 2	Patient 3	# genes with mutation in all patients
Number of variants	147,120	275,502	71,594	
Not in DBSNP/1000G	6,244	12,292	3,089	1,185
Only exonic / splice-site	213	256	313	34
De novo	10			
Heterozygous		223	285	1

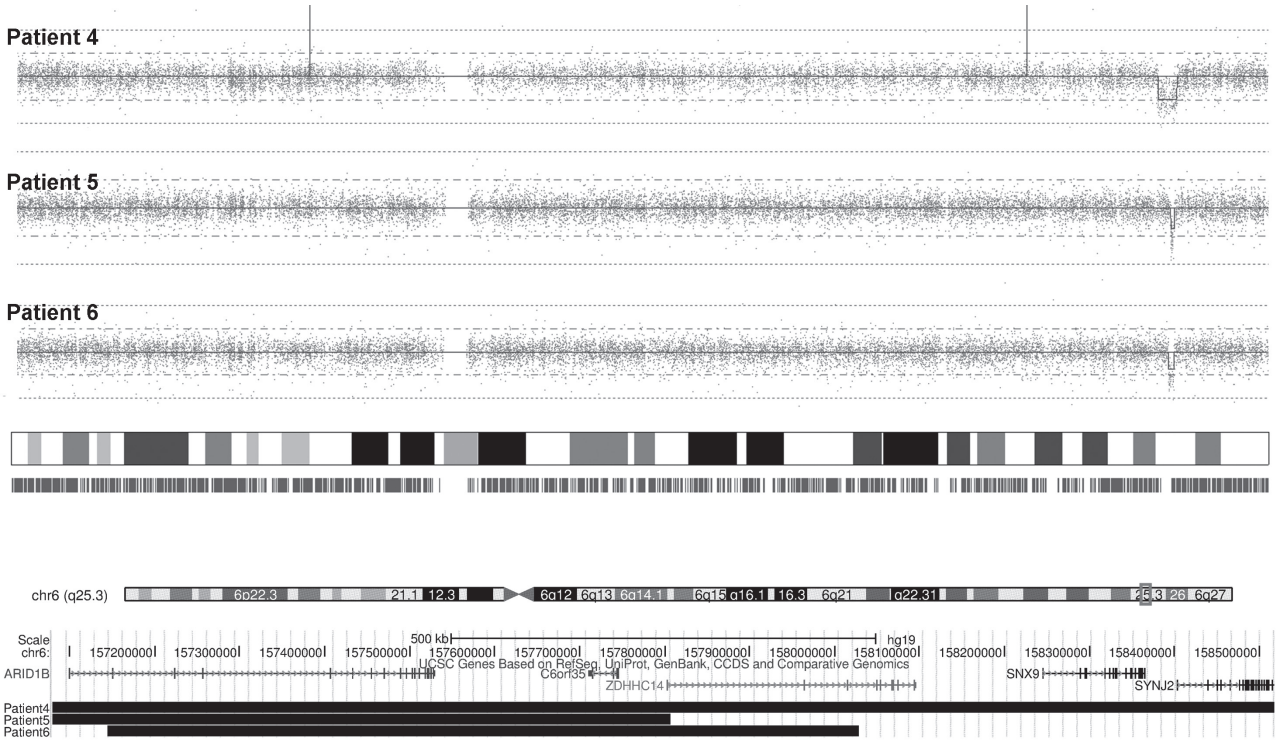
Supplementary Figure 1: Additional features in the CSS patients. Hypertrichosis in patients 1 (A) and 2 (B). Agenesis of the corpus callosum in patient 1 (thin arrow, thick arrow points to a small part of the corpus callosum that is present) (C). Brachydactyly of the fifth finger in patient 2 (D), missing terminal phalanx of the fifth toe in patient 3 (E), hypoplastic nail in patient 4 (F.



Supplementary Figure 2: On the left the raw data around the pathogenic variant in the Integrative Genomics Viewer[3] f or each of the 3 patients. Arrows indicate the location of the variant. On the right the Sanger sequencing validation results.



Supplementary Figure 3: Top: Plot of the genomic deletions detected in patients 4 to 6. Red dots represent the raw data, the blue line represents the calculated CNV score. All deletions are *de novo* (data of parents not shown). No other possibly pathogenic CNVs were found in these patients. Bottom: graph of the size and location of each of the deletion.



1. Li,H. & Durbin,R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. **25**, 1754-1760 (2009).

2. McKenna,A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297-1303 (2010).

3. Robinson,J.T. et al. Integrative genomics viewer. Nat. Biotechnol. **29**, 24-26 (2011).