<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Sacral agenesis: a pilot whole exome sequencing and copy number study</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Porsch, RM; Merello, E; De Marco, P; Cheng, G; Rodriguez, L; So, MT; Sham, PC; Tam, PKH; Capra, V; Cherny, SS; Garcia-Barcelo, MM; Campbell, DD</td>
</tr>
<tr>
<td><strong>Citation</strong></td>
<td>BMC Medical Genetics, 2016, v. 17 n. 1, p. 98:1-10</td>
</tr>
<tr>
<td><strong>Issued Date</strong></td>
<td>2016</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://hdl.handle.net/10722/238724">http://hdl.handle.net/10722/238724</a></td>
</tr>
<tr>
<td><strong>Rights</strong></td>
<td>BMC Medical Genetics. Copyright © BioMed Central Ltd.; This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.</td>
</tr>
</tbody>
</table>
Sacral agenesis: a pilot whole exome sequencing and copy number study

Robert M. Porsch, Elisa Merello, Patrizia De Marco, Guo Cheng, Laura Rodriguez, Manting So, Pak C. Sham, Paul K. Tam, Valeria Capra, Stacey S. Cherny, Maria-Mercè Garcia-Barcelo and Desmond D. Campbell

Abstract

Background: Caudal regression syndrome (CRS) or sacral agenesis is a rare congenital disorder characterized by a constellation of congenital caudal anomalies affecting the caudal spine and spinal cord, the hindgut, the urogenital system, and the lower limbs. CRS is a complex condition, attributed to an abnormal development of the caudal mesoderm, likely caused by the effect of interacting genetic and environmental factors. A well-known risk factor is maternal type 1 diabetes.

Method: Whole exome sequencing and copy number variation (CNV) analyses were conducted on 4 Caucasian trios to identify de novo and inherited rare mutations.

Results: In this pilot study, exome sequencing and copy number variation (CNV) analyses implicate a number of candidate genes, including SPTBN5, MORN1, ZNF330, CLTCL1 and PDZD2. De novo mutations were found in SPTBN5, MORN1 and ZNF330 and inherited predicted damaging mutations in PDZD2 (homozygous) and CLTCL1 (compound heterozygous). Importantly, predicted damaging mutations in PTEN (heterozygous), in its direct regulator GLTSCR2 (compound heterozygous) and in VANGL1 (heterozygous) were identified. These genes had previously been linked with the CRS phenotype.

Two CNV deletions, one de novo (chr3q13.13) and one homozygous (chr8p23.2), were detected in one of our CRS patients. These deletions overlapped with CNVs previously reported in patients with similar phenotype.

Conclusion: Despite the genetic diversity and the complexity of the phenotype, this pilot study identified genetic features common across CRS patients.

Keywords: Sacral agenesis, Caudal regression, Copy-number variation, Whole exome sequencing

Background

Caudal Regression Syndrome (CRS; Caudal Dysgenesis Syndrome, Caudal Dysplasia Sequence, Congenital Sacral Agenesis; OMIM 600145) is a rare (1 in 7,500–100,000 births [1, 2]) congenital disorder characterized by varying degrees of spinal column agenesis. Associated with it are anomalies of central nervous, genito-urinary, cardiac, respiratory and gastro-intestinal systems [3] with anorectal malformations (ARMs) being the most common.

CRS has been attributed to abnormal fetal development of the caudal mesoderm before the fourth week of gestation [4]. During the abnormal gastrulation, prospective notochordal cells, that are wrongly specified in terms of their rostrocaudal positional encoding, are eliminated. Eventually, fewer or even no cells will be available to form the notochord at a given abnormal segmental level. The consequences of such segmental notochordal paucity are manifold and affect the development of the spinal column and spinal cord as well as other organs that rely on the notochord as their inductor. If the prospective notochord is depleted a wide array of segmental vertebral malformations may develop...
including segmentation defects, indeterminate or block vertebrae, or absence of several vertebrae. Because of lack of neural induction and absence of a floor plate, fewer prospective neuroectodermal cells will be induced to form the neural tube. The resulting malformation essentially depends on the segmental level and the extent of the abnormality along the longitudinal embryonic axis, with subsequent interference on the processes of primary and/or secondary neurulation [5]. However, what triggers such abnormal events is not known.

Caudal spinal abnormalities are the defining characteristics of CRS. Cama et al. [6] and Pang et al. [7] classified the disorder into 5 categories according to the degree of caudal spine involvement: Type I) total sacral agenesis with normal or short transverse pelvic diameter and some lumbar vertebrae possibly missing; Type II) total sacral agenesis without involvement of lumbar vertebrae; Type III) subtotal sacral agenesis or sacral hypodevelopment; Type IV) hemisacrum and Type V coccygeal agenesis.

Maternal type 1 diabetes is a risk factor for CRS, as it is for many other congenital disorders [8]. Maternal type 1 diabetes confers a higher relative risk (252) for CRS than for any congenital disorder [9]. The exact mechanism by which maternal diabetes affects fetal development in humans remains unclear [10]. While animal studies have shown that embryos exposed to higher levels of glucose develop growth anomalies, hyperglycemia has not been associated with abnormal fetal development in humans [11]. During normal pregnancies insulin sensitivity is reduced at the start of the third trimester in order to provide metabolic fuel for both mother and the developing fetus. However, since insulin is unable to cross the placenta, the fetus starts producing its own insulin in order to metabolize nutrition. It has been suggested that insulin, antibodies to insulin, or some other abnormality of carbohydrate metabolism could affect the development of a genetically susceptible fetus [12, 13].

Evidence for a genetic cause is provided by the existence of familial segregation as well as animal models. While the most severe forms of CRS present sporadically, milder CRS forms can be transmitted within families in a dominant manner with reduced penetrance and phenotypic variability [4]. Currarino syndrome (CS) is characterized by sacral agenesis type IV, presacral mass, and ARM. CS has been associated with mutations in the MNX1 gene [14–19]. Yet MNX1 mutations account for only 50% of sporadic and 90% of familial cases [19]. Although private mutations in genes such as VANGL1 [20], HOXD13 [21] and PTEN [22] have been described in sporadic cases with caudal dysgenesis and/or vertebral anomalies, no firm genetic association has been established.

A CRS-like phenotype can be induced by administration in animals of retinoic acid (RA), lithium, cadmium, sulphanamide, or organic solvents [23, 24]. Several mutated genes including Cyp26a1, Hoxd13 [25], Wnt-3a [26], Acd, Ptfla, and Pesk5 underlie a CRS-like phenotype in mice [10, 27], yet mutations in the human orthologs have never been identified in CRS patients. Interestingly, the reverse is also true; Mnx1 (formerly Hxlb9) mutant mice do not present Currarino syndrome features [28]. These exceptions to human-mouse phenotypic correlation suggest differences in genetic etiology between humans and experimental organisms [10].

In order to search for genetic risk factors for CRS the exomes of five sporadic CRS cases and their respective healthy parents were sequenced. Due to the sporadic nature of the disease we have focused on de novo or recessive inherited damaging genetic variants. In addition we also used a SNP chip assay in order to identify rare and de novo CNVs.

Methods
Subjects
The records of patients treated between 1995–2010 at the Neurosurgery Department of Giannina (Genoa, Italy) and at the AbaCid-Genética, Grupo HM Hospitales (Madrid, Spain) for congenital anomalies of the spine were reviewed. For all patients, family history, cardiac, respiratory and endocrine data were collected. This included family history for diabetes. Neurological, neuro-physiological (Somatosensory evoked potential, SEP), radiological, neuroradiologic (MRI), orthopaedic, physical, urological (urodynamic, cystography) and surgical assessments were performed for each case. For this pilot study, we selected four Italian trios (CR5, CR17, CR41, CR46) as well as one trio from Spain (CURRE20). We only selected cases within this period who had sporadic lower spine agenesis. Patients were also affected with additional anomalies of axial skeleton and internal organs. One child had a mother with diabetes type I. The local ethical committees approved the study and written informed consent was obtained from all patients and parents. Subjects are identified by the trio ID suffixed by either A, B or C indicating father, mother or child respectively.

Bioinformatics processing
Capture, alignment and base-calling
Whole exome sequencing (WES) was performed at the Centre of Genomic Sciences of the University of Hong Kong, Hong Kong. The exomes of all five trios were sequenced using Illumina HiSeq PE100 and captured with TruSeq Exome Enrichment kit (FC–121–1024, Illumina Inc.). The exome sequences were alignment against Human Genome HG-19 using BWA MEM [29].
Duplicated reads were flagged with Picard-tools [30]. The GATK tool set was used to realign indels, perform base recalibration, remove duplicates, perform indel and SNP calling, and for genotype refinement to improve accuracy of genotype calls [31]. Data quality for each variant was scored and a hard threshold used to remove low data quality variants. We used the GATK recommended criteria for this (see Additional file 1). Relatedness of our participants was investigated using PLINK [32]. We then assessed variants for their potential pathogenicity and frequency, retaining for further analysis only variants that were rare. We considered a variant to be rare if its minor allele frequency was ≤1% in each of several public databases (see Additional file 1). We considered a variant to be potentially deleterious according to a score obtained from KGGSeq [33]. KGGSeq’s prediction algorithm makes use of available biological information (the mutation’s effect on the gene, i.e. stop gain or loss, frameshift, splice site, missense), as well as scoring from other publicly available prediction algorithms (PolyPhen-2, SIFT and others). KGGSeq scores were only used as an informative instrument, variants were not removed from the list of possible disease relevant candidates based on KGGSeq score alone.

**De novo, homozygous and compound heterozygous mutations**

**Single nucleotide variants (SNVs) and small indels**

Subsequent analysis of de novo and compound heterozygous, as well as homozygous, mutations was performed using KGGSeq [33]. For recessive disease models (homozygous and compound heterozygous) we only considered variants with a minor allele frequency (MAF) of less than 1%. We defined a de novo mutation as a first time genetic alteration of a specific locus in a proband. Compound heterozygous mutations were defined as the co-occurrence of two nonsynonymous alleles, one paternal, the other maternal, within a gene. The probability of de novo mutations in each gene was estimated using the framework of Samocha et al. [34]. These probabilities were used to guide the assessment of de novo mutations and not to filter variants. Since a similar framework was not available for compound heterozygous mutations we made use of the only large control trio dataset publicly available, the Genome of the Netherlands (GoNL) [35]. The GoNL is a population dataset containing 250 unaffected parents-offspring trios. We estimated the background compound heterozygous mutation rate per gene from the GoNL dataset. We prioritized compound heterozygous mutations found in our CRS cases in genes where the background rate was low. Thus we classed as a candidate risk locus any gene for which a recessive or de novo model could be constructed in any of our trios using the set of rare potentially deleterious variants we had identified. Detected de novo, compound heterozygous and homozygous mutations were validated via Sanger sequencing of trio DNA (i.e. genotypes were validated in both parents and child).

Analysis of kinship revealed misattributed paternity within one family (CR46). Hence the family CR46 was excluded from all family based analyses (de novo, compound heterozygous and homozygous mutation analysis).

**Copy number variation**

We investigated copy number variation (CNV) in the families CR5, CR17 and CR41 with Illumina’s HumanCoreExome-24 beadchip. Quality control of the assayed genotypes was performed using GenomeStudio (Illumina Inc.) using the default settings. CURR20 was also genotyped using CytoScan® HD Array, but failed initially quality control and was therefore excluded from the analysis. CNV calling and de novo CNV detection was performed using PennCNV [36]. We identified potentially disease associating CNVs as follows. We retained for further analysis, CNVs which allowed construction of a recessive disease model for any gene in any of our trios. We also retained de novo CNVs and rare CNVs. We deemed a CNV to be rare if it did not overlap with any CNV detected in the 1000 Genome Project. De novo CNVs were validated by quantitative real-time PCR (ABI Prism 7900 Sequence Detection System; Applied Biosystems) using TaqMan® Copy Number Assay (Catalog #: 4400291). Ensembl’s genome browser was used to determine genes or regulatory elements affected by the CNVs [37]. Initially we attempted to call CNVs from the exome sequencing data on our trios using three programs (EXCAVATOR [38], CoNIFER [39], and CONTRA [40]), but found no consistency between used tools. Similar previous studies have demonstrated limited power for CNV detection from exome sequencing data tools [41] (Table 1).

**Results**

After extensive quality control and MAF (MAF ≤ 1%) filtering we retained 229,849 variants of which 92.4% were known in dbSNP137. Hence we only retained variants below the frequency of 1% and those which are novel. Out of these, 7,442 missense, 184 frameshift, 227 nonframe-shift, 150 splicing, 173 stop-gain and 5 stop-loss variants in 3,872 different genes were analyzed in respect to de novo, compound heterozygous and homozygous mutations. Of these variants 236 were either heterozygous or homozygous in all analyzed cases (see Additional file 2). Out of these, 226 were missense and 10 stop-gain variants distributed across 35 different genes. A further 65 of these variants were predicted to be damaging by KGGSeq. We identified two rare mutations in the known CRS related genes PTEN (rs202004587, missense,
<table>
<thead>
<tr>
<th>Subjects</th>
<th>Sacral agenesis and vertebral malformations</th>
<th>Ribs/Limbs anomalies</th>
<th>Genitourinary</th>
<th>Neural tube</th>
<th>ARM</th>
<th>Cardiac</th>
<th>Other</th>
<th>Maternal phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>CR5/M</td>
<td>Type II Hemivertebra T7-T8</td>
<td>Additional 13th right rib</td>
<td>Hydrenephrosis, Hydroureter, Bladder-exstrophy</td>
<td>Lipoma Low-lying conus medullaris</td>
<td></td>
<td>CPT (II) carrier</td>
<td>CPT (II) deficiency</td>
<td></td>
</tr>
<tr>
<td>CR17/F</td>
<td>CRS Type II</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Omphalocele Twisted teeth</td>
<td>Diabetes type I</td>
<td></td>
</tr>
<tr>
<td>CR41/F</td>
<td>CRS Type I Deformed T7-T8-T9</td>
<td>Fusion of 5th and 6th left ribs. Additional 13th left rib Club feet</td>
<td>Incontinence</td>
<td>Lipoma Blunt ending conus medullaris (T11)</td>
<td>Pulmonary vein atresia, Inter-ventricular septal defect</td>
<td>Congenital hip dislocation, Motor delay</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CR46/M</td>
<td>CRS Type II Lumbar kyphosis</td>
<td>Club feet</td>
<td></td>
<td>Anal Stenosis</td>
<td>Intra-ventricular septal defect</td>
<td>Congenital bilateral hip dislocation. Short neck.</td>
<td>Hydrocephalus</td>
<td></td>
</tr>
<tr>
<td>CURR20/F</td>
<td>CRS Type V</td>
<td></td>
<td></td>
<td>Anal Stenosis</td>
<td></td>
<td>Teratoma</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

M male, F female, T thoracic vertebra, S sacral vertebra, ARM anorectal malformations, CPT Carnitine palmitoyl transferase

*According to Cama et al. (8) and Pang et al. (7) classification of sacral agenesis
p.A79T) and VANGLI (rs74117015, stop-gain, p.Ser338Ter) in CR5C inherited from the mother and father respectively (Table 2). Both were predicted to be damaging.

**De novo variants**
In total we identified three *de novo* mutations, two missense and one frameshift mutations in three different genes: MORN1 (p.Gly107Arg), SPTBN5 (p.Glu25Lys), and ZNF330 (p.Lys3fs) in patients CR41C, CR5C, and CURR20C respectively (Table 2). MORN1 encodes MORN (membrane occupation and recognition nexus) repeats [42]. The exact function of this gene is not known, however, in *Toxoplasma gondii* it is known to be involved in nuclear cell division [43]. Furthermore MORN repeats are known to be part of a number of genes, including junctophilins [44] which are involved in cardiomyopathy [45]. Notably, MORN1 was reported to be produced by insulin producing cells (IPCs) derived from pancreatic stem cells [46]. The estimated probability for a *de novo* mutation to occur in MORN1 is 0.8%, but 59% of all analyzed genes have a lower probability [34]. Pathogenicity analysis by KGGSeq suggests that the *de novo* mutation is damaging.

SPTBN5 (OMIM: 605916) is a beta-spectrin encoding protein. It plays an important role in linking proteins, lipids, and cytosolic factors of the cell membrane to the cytoskeletal filament systems of the cell [47]. SPTBN5 is expressed in the cerebellum, pancreas, kidney, and bladder, as well as in a number of other systems. The gene has not been associated with any disease or disorder. The estimated gene-specific probability of *de novo* mutation is 1.8% and 99% of genes have a lower probability of having a *de novo* mutation making this gene less likely to be causally related. Further, KGGSeq’s pathogenic prediction algorithm suggests that this variant is benign.

ZNF330 (OMIM: 609550) is a zinc finger protein with no known disease association and is mainly present in the nucleus during interphase as well as at the centro-meres during mitosis [48]. Interestingly, this gene is differentially expressed in pancreatic Islets of Langerhans and in peripheral blood mononuclear cells [49]. The estimated gene-based *de novo* mutation probability is 0.6%, relatively low but still within the 28th percentile of all genes. KGGSeq was unable to predict a pathogenic score for this variant.

Thus, given the pathogenic nature of the two *de novo* variants and their expression pattern in pancreatic cells, MORN1 and ZNF330 are candidate CRS genes.

We detected one *de novo* CNV deletion on 3q13.13 in CR5C (Table 3). The deletion does not seem to encompass any gene or functional element, yet it overlaps with CNVs previously reported in patients with a similar phenotype. In particular, a documented *de novo* deletion in a Japanese patient with OEIS (omphalocele, exstrophy of the cloaca, imperforate anus, spinal defects) complex who also had a sacrum malformation (DECIPHER: 971) overlaps with the *de novo* CNV identified in CR5C.

**Homozygous and compound heterozygous mutations**
In total we identified 8 compound missense heterozygous and one homozygous missense mutations (PDZD2) which passed the described filtering criteria (detailed in Table 2). Strikingly, mutations in genes related to diabetes were detected in two patients. None of the affected genes were recurrent. The two genes associated with diabetes were PDZD2 and CLTCL1 and were found mutated in CR5C and CR17C respectively.

PDZD2 (p.Ser1106Phe) has been shown to be an important promoter of fetal pancreatic progenitor cell proliferation [50, 51]. Ma et al. [52] showed that expression of PDZD2 is specific to pancreatic beta cells. Furthermore, higher concentrations of secreted PDZD2 in rat islet cell lines were correlated with higher rates of cell proliferation and inhibited transcription of *INS*, an insulin promoter.

CLTCLI (p.Arg1620His, p.Val44Phe) is involved in the intracellular trafficking of glucose transporter GLUT4. Intracellular trafficking of the glucose transporter GLUT4 from storage compartments to the plasma membrane is triggered in muscle and fat during the body’s response to insulin [53].

A compound heterozygous mutation in GLTSCR2 (Glioma Tumor Suppresser Candidate Region Gene 2) was identified in patient CR5C (p.Arg190Trp, p.Thr284Met). GLTSCR2, is expressed at high levels in pancreas and heart, is a tumor suppressor gene and a direct regulator of *PTEN*. Mutations of *PTEN* have been previously identified in a patient affected with VACTERL (Vertebral anomalies, Anal atresia, Cardiac defects, Tracheoesophageal fistula and/or Esophageal atresia, Renal & Radial anomalies and Limb defects) [22] which has commonalities with CRS [27]. This is especially interesting considering that patient CR5C is also harbouring a rare inherited mutation in *PTEN* itself.

A compound heterozygous mutation (p.Ala1616Thr, p.Thr3620Leu) in DNAS10, an inner arm dynein heavy chain, was identified in CR41C. Dynein proteins are implicated in many disorders such as motor neuropathies, cortical development diseases, as well as congenital malformations such as heterotaxia and situs inversus. Moreover, cytoplasmic Dyneins have been reported to interact with Kinesin (KIF1A, mutated in patient CR17C) for interkinetic nuclear migration in neural stem cells [54].

We detected a homozygous CNV deletion encompassing part of chromosome 8p23.2 in patient CR5C (Table 3). The CNV does not overlap with known genes...
<table>
<thead>
<tr>
<th>Subjects</th>
<th>Genes</th>
<th>Nucleotide variants</th>
<th>RsID</th>
<th>MAF (in %)</th>
<th>Aminoacidic variants</th>
<th>OMIM Associated Disease</th>
<th>Functional Role</th>
<th>Mutation status</th>
</tr>
</thead>
<tbody>
<tr>
<td>CR5C</td>
<td>SPTBN5</td>
<td>c.73G &gt; A</td>
<td>–</td>
<td>–</td>
<td>p.(Glu25Lys)</td>
<td>interaction of cytoskeletal filament with other components of the cell [47]</td>
<td>De novo</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PDZD2</td>
<td>c.3317C &gt; T</td>
<td>rs34748216</td>
<td>0.763</td>
<td>p.(Ser1106Phe)</td>
<td>insulin regulation [52]</td>
<td>H</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PKHD1L1</td>
<td>c.8291A &gt; C, c.11969G &gt; A</td>
<td>rs118074609/ rs146831382</td>
<td>0.713/0.081</td>
<td>p.(Asn2764Thr)/ p.(Gly3990Glu)</td>
<td>cellular immunity [65]</td>
<td>CH</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GLUTSCR2</td>
<td>c.568C &gt; T, c.851C &gt; T</td>
<td>rs34462252/ rs200463741</td>
<td>0.163/0.704</td>
<td>p.(Arg190Trp)/ p.(Thr284Met)</td>
<td>PTEN regulation [66]</td>
<td>CH</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PTEN</td>
<td>c.235G &gt; A</td>
<td>rs20204587</td>
<td>–</td>
<td>p.(Ala45Thr)</td>
<td>VATER association with macrocephaly and ventriculomegaly</td>
<td>HET</td>
<td></td>
</tr>
<tr>
<td></td>
<td>VANGL1</td>
<td>c.1013C &gt; A</td>
<td>rs74117015</td>
<td>–</td>
<td>p.(Ser388Ter)</td>
<td>Regulation of growth of human hepatoma cells [20]</td>
<td>HET</td>
<td></td>
</tr>
<tr>
<td>CR17C</td>
<td>ARHGEF16</td>
<td>c.784A &gt; G, c.1477C &gt; T</td>
<td>–/–</td>
<td>–/–</td>
<td>p.(Thr262Ala)/ p.(Leu493Phe)</td>
<td>guanyl-nucleotide exchange factor</td>
<td>CH</td>
<td></td>
</tr>
<tr>
<td></td>
<td>KIF1A</td>
<td>c.4781C &gt; T, c.2522A &gt; T</td>
<td>rs201825284/–</td>
<td>0.011/–</td>
<td>p.(Ser1594Leu)/ p.(Asn841Leu)</td>
<td>mental retardation, spastic paraplegia-30, neuropathy</td>
<td>CH</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CLTC1L</td>
<td>c.4859G &gt; A, c.130G &gt; A</td>
<td>rs5748024/ rs34869740</td>
<td>0.787/0.855</td>
<td>p.(Arg1620His)/ p.(Val44Phe)</td>
<td>intracellular trafficking of the glucose transporter GLUT4 [68]</td>
<td>CH</td>
<td></td>
</tr>
<tr>
<td>CR41C</td>
<td>MORN1</td>
<td>c.319G &gt; A</td>
<td>–</td>
<td>–</td>
<td>p.(Gly107Arg)</td>
<td>cardiomyopathy, hypertrophic-17</td>
<td>De novo</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DNAH10</td>
<td>c.4846G &gt; A, c.10859C &gt; T</td>
<td>rs376989344/ rs20063832</td>
<td>0.012/0.421</td>
<td>p.(Ala161Thr)/ p.(Thr3620Leu)</td>
<td>primary ciliary dyskinesia</td>
<td>CH</td>
<td></td>
</tr>
<tr>
<td>CURR20C</td>
<td>ZNF330</td>
<td>c.6_7insT</td>
<td>–</td>
<td>–</td>
<td>p.(Lys3fs)</td>
<td>inner arm dynein heavy chain [70]</td>
<td>De novo</td>
<td></td>
</tr>
<tr>
<td></td>
<td>VPS18</td>
<td>c.1697A &gt; G, c.1823G &gt; A</td>
<td>rs373608134/ rs34865555</td>
<td>0.023/0.605</td>
<td>p.(Tyr566Cys)/ p.(Arg508His)</td>
<td>inner arm dynein heavy chain [70]</td>
<td>CH</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PKD1L2</td>
<td>c.6241_6242ins19ntb, c.706_707delAA</td>
<td>–/–</td>
<td>–/–</td>
<td>p.(Thr2081fs)/ p.(Asn236fs)</td>
<td>protein transportation to the vacuole [71]</td>
<td>CH</td>
<td></td>
</tr>
</tbody>
</table>

CH compound heterozygous, H homozygous, HET heterozygous

aMutation information are given for the long form of the transcript (NCBI reference: NM_052892)
b19nt: GCTTTCCCCAGGCTTGGCAGTA
but is contained within a duplication found in a patient with abnormal sacrum (DECIPHER: 271204). This documented patient, while also harboring another deletion (7q34–7q36.3), displayed a great variety of phenotypes including central hypertonia, hypermetropia, long thorax, narrow mouth, seizures, strabismus, and deep set eyes. Additional detected rare CNVs overlapped with a number of other genes, however, none were known to be related to CRS (see Additional file 1).

Discussion

Within this pilot study we have identified a number of novel risk loci potentially connected to CRS. We have also found a number of mutations for already known genetic risk factors [10]. Here we highlight these preliminary findings and discuss their relevance for future studies.

Foremost, all four patients were affected by a homozygous, compound heterozygous mutations or de novo variant in a diabetes-relevant (CLTCL1 and PDZD2) or pancreatic expressed (MORNI and ZNF330) gene. While these results are not definitive it is in line with the increased CRS risk for children born to diabetic mothers. In addition, one de novo (chr3q13.13) and one homozygous CNV (chr8p23.2) overlap with CNVs reported in patients with similar phenotype. Identification of overlapping CNVs in patients with similar phenotype is the central aim of DECIPHER [55]. Since many patients with rare diseases harbor novel or extremely rare variants, it is crucial to accumulate evidence across patients in order to foster understanding of the disease. Furthermore, we identified a heterozygous mutation in GLTSCR2, a direct regulator of PTEN. PTEN has been previously associated with CRS-like phenotypes (VACTERL) [10]. Interestingly, the same subject (CR5C) has also an inherited mutation within PTEN and the CRS-related gene VANGL1. These results further strengthen the role of PTEN and VANGL1 in CRS. Likewise, the identification of compound heterozygous mutations in DNAH10 and KIF1A could suggest an involvement of ciliary proteins.

There are, a number of limitations to our study. Our sample size is small, but to be expected given the disease prevalence and the costs of the genetic assays used. We were not able to identify recurrent affected genes across different patients. We did not look for mutations assuming other than recessive inheritance because the yield of true to false positives would be poor. Many types of genetic variation were not assayed, for instance we only assayed exomic SNVs. Lastly, models involving environment, such as gene-environment interactions within the utero, could not be investigated.

The diversity of identified potential disease mechanisms matches that of previous studies [10, 56–58] and also reflects the phenotypic diversity associated with CRS [56]. We previously showed that if a disorder is genetically complex, one should expect large genetic heterogeneity across patients [59]. Thus the number of candidate genes identified is not surprising and is similar to that reported for other complex rare genetic disorders [60]. While the presence of several possible disease mechanisms does not necessary suggest a multigenic model, the large number of candidate genes identified within this study, as well as those reported by others [23, 27, 61–64] suggests that CRS might be caused by a multitude of private genetic risk factors. This makes identification of a common underlying genetic architecture challenging. Furthermore, differences in the genetic etiology between humans and experimental organisms makes it difficult to investigate the exact causal mechanism. Many aspects of the disease are still poorly characterized, for example, disease prevalence. While some studies have estimated that 1 in 7,700 children might be affected [1], others suggest it might be as rare as 1 in 100,000 births [2]. This further complicates estimation of the number of disease causing mechanisms [59].

Table 3  De novo and homozygous CNVs

<table>
<thead>
<tr>
<th>Patients</th>
<th>Chromosome</th>
<th>Start Position</th>
<th>End Position</th>
<th>Length</th>
<th>Type</th>
<th>Genes or regulatory elements</th>
<th>Patients with related symptoms listed in DECIPHER (type of CNV, patient phenotype)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CR5C</td>
<td>3q13.13</td>
<td>109489534</td>
<td>109510473</td>
<td>20939</td>
<td>De novo/deletion</td>
<td>--</td>
<td>971, deletion; abnormality of the sacrum, Abnormality of the small intestine, Anal atresia, Cloacal extrophy, Omphalocele, Spina bifida occulta.</td>
</tr>
<tr>
<td></td>
<td>8p23.2</td>
<td>5599399</td>
<td>5605087</td>
<td>5688</td>
<td>homozygous/deletion</td>
<td>--</td>
<td>271204, duplication; Abnormality of the sacrum, Central hypotonia, Deeply set eye, Hypermetropia, Long thorax, Narrow mouth, Nasogastric tube feeding in infancy, Seizures, Strabismus.</td>
</tr>
</tbody>
</table>
Conclusions
Despite the complexity of the phenotype, we were able to identify common genetic characteristics across patients, potentially causally related to the known risk factors and supposed disease etiology. Our data, although limited to a small group of patients, support a multigenic model for CRS. Future studies should consider larger accumulated samples across multiple centers in order to identify common genetic characteristics via whole genome or whole exome sequencing.

Additional files

**Additional file 1:** Supplementary Methods and Material. (DOCX 51 kb)
**Additional file 2:** All called variants after QC and MAF. (XLSX 1329 kb)

Abbreviations
ARM: Anorectal malformation; CNV: Copy number variation; CRS: Caudal regression; HH: Hedgehog protein; IPC: Insulin producing cells; MAF: Minor allele frequency; QC: Quality control; SEP: Somatosensory evoked potential; SNP: Single nucleotide polymorphism; SNV: Single nucleotide variation; WES: Whole exome sequencing

Acknowledgements
We thank all participants who made this study possible. We also thank the A.S.B.I. (Associazione Spina Bifida Italia).

Funding
Seed Funding Programme for Basic Research, University of Hong Kong, Project Code:20141015002 to MMGB. Small Project Funding, University of Hong Kong, Project Code: 1209176012 to DC. EM; PDM, VC would like to acknowledge Ricerca Corrente Ministero della Salute Italia 5X Mille, Aletti- Volpatti Trust Onlus and private funding resources.

Availability of data and materials
All called variants are included in the supplementary material.

Authors’ contributions
RMP, DDC, SSC, GC, MS and MMGB analyzed the data and drafted the paper. PCS and PKT reviewed the study proposal, provided feedback on the study progress and manuscript. MMGB proposed the study idea. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Consent for publication
Not applicable.

Ethics approval and consent to participate
The study was approved by the Ethical Committee Istituto Giannina Gaslini (IGG-VACA 213/2013), the CEdIN HM Hospitals, and the institutional review board of The University of Hong Kong together with the Hong Kong Hospital Authority (Institutional Review Board: UW 10-399). Written informed consent was obtained from all patients and parents.

Author details
1Department of Psychiatry, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, Hong Kong SAR. 2Department of Surgery, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, Hong Kong SAR. 3Istituto Giannina Gaslini, Genoa, Italy. 4AbACid-Genetica. Grupo HM Hospitals, Madrid, Spain. 5Centre for Reproduction, Development, and Growth, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, Hong Kong SAR. 6Centre for Genomic Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, Hong Kong SAR. 7State Key Laboratory of Brain and Cognitive Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, Hong Kong SAR. 8The Hong Kong Jockey Club Building for Interdisciplinary Research, 5 Sassoon Road, Pokfulam, Hong Kong, People’s Republic of China.

Received: 25 May 2016 Accepted: 2 December 2016
Published online: 22 December 2016

References

**Page 8 of 10**


