Duplication of *C7orf58, WNT16* and *FAM3C* in an Obese Female with a t(7;22)(q32.1;q11.2) Chromosomal Translocation and Clinical Features Resembling Coffin-Siris Syndrome

Jun Zhu¹, Jun Qiu¹, Gregg Magrane¹, Malak Abedalthagafi², Andrea Zanko³, Mahin Golabi³, Farid F. Chehab¹*  

¹ Department of Laboratory Medicine, University of California San Francisco, San Francisco, California, United States of America, ² Department of Pathology, University of California San Francisco, San Francisco, California, United States of America, ³ Department of Pediatrics, University of California San Francisco, San Francisco, California, United States of America

Abstract

We characterized the t(7;22)(q32;q11.2) chromosomal translocation in an obese female with coarse features, short stature, developmental delay and a hypoplastic fifth digit. While these clinical features suggest Coffin-Siris Syndrome (CSS), we excluded a CSS diagnosis by exome sequencing based on the absence of deleterious mutations in six chromatin-remodeling genes recently shown to cause CSS. Thus, molecular characterization of her translocation could delineate genes that underlie other syndromes resembling CSS. Comparative genomic hybridization microarrays revealed on chromosome 7 the duplication of a 434,682 bp region that included the tail end of an uncharacterized gene termed *C7orf58* (also called *CPED1*) and spanned the entire *WNT16* and *FAM3C* genes. Because the translocation breakpoint on chromosome 22 did not disrupt any apparent gene, her disorder was deemed to result from the rearrangement on chromosome 7. Mapping of yeast and bacterial artificial chromosome clones by fluorescent in situ hybridization on chromosome spreads from this patient showed that the duplicated region and all three genes within it were located on both derivative chromosomes 7 and 22. Furthermore, DNA sequencing of exons and splice junctional regions from the duplicated region and all three genes within it on both derivative chromosomes 7 and 22 revealed the presence of potential splice site and promoter mutations, thereby augmenting the detrimental effect of the duplicated genes. Hence, dysregulation and/or disruptions of *C7orf58*, *WNT16* and *FAM3C* underlie the phenotype of this patient, serve as candidate genes for other individuals with similar clinical features and could provide insights into the physiological role of the novel gene *C7orf58*.

Introduction

The uncovering of genes that cause clinically severe genetic disorders often reveals a critical physiological role for these genes. Recently, mutations that cause Coffin-Siris syndrome (CSS) [MIM 135900] [1,2], a rare congenital disorder, were unveiled and occurred in chromatin remodeling genes of the SW1/SNF complex [3,4,5]. Other congenital anomalies that overlap with clinical features of CSS, such as the brachymorphism-onychodysplasia-dysphalangism syndrome [6,7,8], Cornelia de Lange syndrome [9,10], Mabry syndrome, Nicolaides–Baraitser syndrome (NCBRS), DOOR syndrome (deafness, onychodystrophy, osteodystrophy, mental retardation, fetal alcohol syndrome, fetal hydantoin/phenytoin embryopathy and trisomy 9 [5], complicate the differential diagnosis of these disorders. To help in their differential diagnosis, clinical algorithms were proposed [5,11] without however, any molecular basis. To begin the identification of genes that share clinical features with CSS and related disorders, we characterized the translocation breakpoints in a female with a *de novo* t(7;22)(q32;q11.2) chromosomal translocation [12], who presented with clinical features that resemble CSS [11]. We found that a duplication encompassing the *WNT16* and *FAM3C* genes and the tail end of a novel gene, termed *C7orf58*, underlie the clinical phenotype of this patient. The identification of mutations in these genes among similarly affected patients could provide insights into their biological roles and may help resolve their ambiguous diagnoses.

Methods

Ethics Statement, DNA Samples and Leptin Radioimmunoassay

Blood samples were obtained from the index patient, following informed written consent and approval by an institutional review
Average coverage of 100 billion reads and 7.56 GigaBases, representing an overall coverage of 100×. The raw sequences were aligned to the human genome build hg19 using Burrows-Wheeler Aligner (BWA) software and the resulting SAM files analyzed for single-nucleotide polymorphisms (SNPs) and insertions/deletions (INDELs) with PICARD, SAMTOOLS and GATK. SNPs and INDELs were further analyzed and annotated with ANNOVAR.

**Exome Sequencing and Bioinformatics Analysis**

Next Generation Sequencing (NGS) was performed on 3 μg of peripheral blood DNA using an Illumina HiSeq following library preparation and exome enrichment with Agilent SureSelect Human All Exon v4 51 Mb kit. Paired-end readings resulted in further analyzed and annotated with ANNOVAR.

**Array-Based Comparative Genomic Hybridization**

Array CGH microarray was performed at Baylor College of Medicine. Briefly, normal and patient DNA samples were labeled with Cy3 and Cy5 fluorescent dyes and hybridized to a 60-mer oligonucleotide human genome microarray platform (24K Agilent, Santa Clara, CA) at a probe density of 5–6 kb. The results were analyzed with the Agilent DNA Analytics software. Copy number variations were predicted by NetPhos [22]. MicroRNA binding sites were queried for 15 nucleotides on either side of the SNP against the mature miRNAs database using the miRBase search engine [23] set at maximum E-value of 10 and using the BLASTN search method.

**Results**

**Clinical Synopsis**

The clinical findings of the index female patient reported in this study were initially reported in [12]. She is a Caucasian girl with a 46, XX karyotype and a de novo apparently balanced chromosome translocation, 46, XX, t(7;22)(q32.1;q11.2). Her parents had normal karyotypes. She attended a special educational resources classroom throughout her schooling. A developmental assessment at 9 years chronological age showed a performance below 6.5-year level and a delay of about three years. Her strengths were in verbal reasoning and her weaknesses in concentration, visual-motor pursuits, fine motor development and poor gross motor with unsteady gait. In earlier childhood, she experienced frequent bouts of respiratory infections and otitis media, but did not require hospitalization. At 9 years, it was noted that her hair was of normal consistency and growth pattern. She had experienced two grand mal seizures with normal EEG and normal MRI. She had two congenital melanocytic nevi surgically excised, one from the right posterior scalp and the other from the left clavicle. An ulcerated pyogenic granuloma was excised from the left mid thoracic spine. At 11 years of age, she was in 4th grade with special educational resources and had exhibited considerable weight gain and a decline in gross motor abilities. At that time, her height was 129.5 cm (5–10%), weight 57.7 Kg (>97%) head circumference 54.25 cm (75%). Her features and clinical history were not felt to be consistent with Prader-Willi Syndrome (PWS) and FISH/methylation studies for PWS were normal. She had a BMI of 34.4 and her leptin levels were 31.5 ng/ml. Endocrine evaluations

**Determination of Allele Frequencies in SNPs**

Allele frequencies were determined using 20–40 anonymous DNA samples from Caucasian individuals, who requested cystic fibrosis carrier screening at the UCSF Molecular Diagnostics Laboratory. The SNPs and number of individuals tested were in C7orf58: rs78456119 (n = 20), rs3990520 (n = 25), rs6817311 (n = 23), INDEL GAAA (n = 21), in WNT16: rs75556099 (n = 40), rs3832519 (n = 32), and in FAM3C: INDEL AAAAACTT (n = 20), rs3837124 (n = 20).

**Bioinformatics Mutation Predictions**

Analysis of the transcription factor binding sites in the putative promoter region was determined by the PROMO [15] and TESS [16] tools. Analysis of missense mutations was performed using the bioinformatic tools SIFT [17], PolyPhen2 [18] and Mutation Taster [19]. Putative mutations affecting splice donor, splice acceptor and branch point sites were analyzed with Human Splicing Finder [20] and Spliceport [21]. Phosphorylation sites were predicted by NetPhos [22]. Micro RNA binding sites were queried for 15 nucleotides on either side of the SNP against the mature miRNAs database using the miRBase search engine [23].
found no etiology for her clinical presentation and her hearing was within normal limits. Her vision was 20/60 with corrective glasses prescribed. While the primary concern remained her obesity, she had no self-restraint and food intake was to be continuously monitored. She also manifested some obsessive-compulsive behaviors. Review of multiple photographs at 9, 10 and 11 years of age likely excluded a CSS diagnosis based on atypical features of facial coarseness, sparse scalp hair, excessive amount of hair in eyebrow/eyelash, dysplasia and aplastic nails (Bryan Hall, personal communication). At 22 years old, she weighed 88.9 Kg, measured 137.2 cm and had a BMI of 47.2. She only gained 7.6 cm in height over the past 10 years. Hyperphagia, obesity and reduced cognition remain the predominant issues.

Exome Sequencing

There were no deleterious mutations such as stops or frameshift mutations in any of the chromatin remodeling genes SMARCA2, SMARCA4, SMARCE1, SMARCB1, ARID1A and ARID1B, recently reported to be mutated in CSS [3,4]. However, synonymous and non-synonymous single nucleotide variations were found in SMARCA2, SMARCA4, SMARCB1 and ARID1B and in other members of the SMARCA and ARID gene families, SMARCA5, SMARCAD1, ARID3A, ARID3C and ARID4A (Table 1). All these variants had common minor allele frequencies (MAF), which preclude them from underlying this patient's rare phenotype. Furthermore, no deleterious mutations were found in genes known to cause morbid obesity such as leptin, leptin receptor, melanocortin 4 receptor, POMC, adiponectin and adiponectin receptor.

Characterization of the Chromosomal Translocation by Array CGH

Array CGH analysis of this patient's DNA revealed copy number variations (CNVs) on chromosomes 1, 4, 15 and 16, shown in Figure 1 and annotated in Table 2. On chromosome 7, a unique duplication of 434,682 bp was revealed and defined by hg18 (Human Genome Build 36.3) between coordinates 120,724,175–121,158,857 (Fig. 2A) and encompassed the WNT16, FAM3C and part of C7orf58. One boundary of the duplication was located in intron 22 of C7orf58, while the other was approximately 44 Kb upstream of the protein tyrosine phosphatase gene (PTPRZ1). This duplication was not previously reported and occurs at 7q31.31.

On chromosome 22, a 47,394 bp region was amplified at multiple copies (Fig. 2B) around the translocation breakpoint region on chromosome 22q11.2 (Fig. 2B) but was previously reported as variation 39,980 in the DGV database. This CNV maps between coordinates 22,677,959 and 22,725,353 in hg18. It was also reported in the Watson genome [24] and in 24 of 30 [25] and 22 of 39 individuals in the HapMap study [26]. Therefore, based on its common occurrence among normal individuals, this amplification of DNA sequences on chromosome 22q11.2 is unlikely to contribute to the phenotype of this patient. Therefore, the duplication on chromosome 7 represents her most significant molecular lesion.

Table 1. SNPs revealed by exome sequencing in chromatin remodeling genes of the index patient.

| Gene     | Nucleotide | Protein | dbSNP | Genotype | SNV   | MAF |
|----------|------------|---------|-------|----------|-------|-----|
| SMARCA2  | c.G3672A   | p.E1224E| rs6601 | 0/1 syn  | 0.203|
| SMARCA4  | c.T1524C   | p.H508H | rs7935 | 0/1 syn  | 0.312|
| SMARCA4  | c.T4887C   | p.D1629D| rs7275 | 0/1 syn  | 0.22 |
| SMARCA5  | c.T2424C   | p.N808N | rs13139128 | 1/1 syn   | 0.996|
| SMARCA1  | c.T1839T   | p.D613D | rs6823404 | 0/1 syn  | 0.432|
| SMARCA1  | c.T902C    | p.V301A | rs7439689 | 0/1 nonsyn| 0.287|
| SMARCB1  | c.C707G    | p.P236R | rs75919464 | 1/1 nonsyn| 0.228|
| ARID1B   | c.G1959A   | p.A653A | rs3734441 | 1/1 syn  | 0.499|
| ARID3A   | c.A225G    | p.P775P | rs1799595 | 1/1 syn  | 0.115|
| ARID3A   | c.G150A    | p.E50E | rs3826948 | 0/1 syn  | 0.496|
| ARID3A   | c.T1161C   | p.N387N | rs12608658| 1/1 syn  | 0.056|
| ARID3A   | c.C1320T   | p.A440A | rs6510986 | 1/1 syn  | 0.295|
| ARID3A   | c.A1650G   | p.G550G | rs1051504 | 1/1 syn  | 0.374|
| ARID3A   | c.G1666A   | p.G556S | rs1051505 | 1/1 nonsyn| 0.403|
| ARID3C   | c.G72A     | p.P24P | rs13283357 | 0/1 syn  | 0.234|
| ARID3C   | c.T1003G   | p.C335G | rs3808869 | 0/1 nonsyn| 0.435|
| ARID3C   | c.A1061G   | p.D354G | N/A | 0/1 nonsyn| N/A  |
| ARID4A   | c.A2171G   | p.N724S | rs2230098 | 0/1 nonsyn| 0.035|
| ARID4A   | c.A2335G   | p.T779A | rs1051858 | 0/1 nonsyn| 0.384|

Bolded are the genes mutated in Coffin- Siris Syndrome. Standard nomenclatures for cDNA and protein variations (hg19) with their dbSNP reference numbers are listed. Genotypes variations from the reference hg19 genome sequence are indicated as 0/1 and 1/1 for heterozygous and homozygous variants, respectively. The effect of each SNP, denoted as single nucleotide variation (SNV), is shown as either synonymous (syn) or non-synonymous (nonsyn). Minor allele frequencies (MAF) are taken from the 1000 Genomes Project. N/A denotes non-available information.

doi:10.1371/journal.pone.0052353.t001
Figure 1. CNVs on chromosomes 1, 4, 15 and 16 of the index patient revealed by Agilent 244K human genome CGH microarray hybridization. The profiles shown represent the averaged combined hybridization data of patient vs. reference DNA. Regions of gain and losses are boxed and depicted as positive or negative log2 ratios relative to the midline that denotes no gain/no loss of DNA sequences.
doi:10.1371/journal.pone.0052353.g001

| Chr. | Genomic position hg18 | Variation No. | Gain (bp) | Loss (bp) | Frequency (n) |
|------|------------------------|---------------|-----------|-----------|---------------|
| 1    | 25,482,036–25,536,138  | 38960         | 54,102    | –         | Watson        |
| 1    | 150,823,073–150,862,088| 38961         | –         | 29,382    | Watson        |
| 1    | 246,795,223–246,862,088| 66541         | –         | 66,865    | 182/450       |
| 4    | 69,057,735–69,165,872  | 38966         | 108,137   | –         | Watson        |
| 15   | 19,805,960–20,220,475  | 2182          | 414,515   | –         | 81/269        |
| 15   | 32,517,513–32,594,948  | 1960          | –         | 77,435    | 48/50         |
| 16   | 32,105,104–33,539,082  | 38978         | 1,433,978 | –         | 28/43         |
| 22   | 22,677,959–22,725,353  | 39,980        | 47,394    | –         | 22/39, 24/40  |

Table 2. Copy number variation (CNV) identified by array CGH in the patient.

Shown are the CNV coordinates in hg18 (Build 36.3) for the gains and losses of DNA sequences on the listed chromosomes (Chr.) The number of individuals or occurrence in the Watson genome are indicated in the frequency data, which are derived from the Toronto Database of Genomic Variants (DGV).
doi:10.1371/journal.pone.0052353.t002
FISH Mapping of the Translocation Breakpoint and the Duplicated Region Encompassing C7orf58, WNT16 and FAM3C

Various yeast and bacterial artificial clones from chromosome 7 were mapped by fluorescent in situ hybridization (FISH) onto chromosome spreads from this patient. The aim of such a mapping was to sublocalize the duplicated region by determining whether a specific DNA clone hybridizes onto her chromosomes 7 and der(7), 7 and der(22) or 7, der(7) and der(22). A representative FISH from each pattern is shown in Figure 3B–D. Towards these ends, we derived YAC clones from a 12.9 Mb chromosomal map region [27] extending from 7q31.2–7q32.1 and defined by the cystic fibrosis transmembrane regulator (CFTR) and carboxypeptidase 1 (CPA1) genes. Figure 4A displays a map of three regions depicted each by a FISH hybridization pattern. One region displayed hybridization to six YACs (857F2, yWSS4875, yWSS2618, 755A9, 915C4, 823H10) onto chromosomes 7 and der(22), indicating that their corresponding DNA sequences from chromosome 7q were translocated to der(22). Another region characterized by its hybridization patterns on chromosomes 7 and der(7) included 4 YACs (E146, 851C4, 910B4, 824H1) and represented DNA sequences that were retained on der(7). A third region that lied between the two previous regions was spanned by five YACs (1047E14, 769P3) located within the last intron of C7orf58 and outside of the duplicated region hybridized to chromosomes 7 and der(7) and der(22), suggesting that these clones span the translocation breakpoint on chromosome 7.

To refine the FISH map within the region of interest that showed hybridizations to chromosomes 7, der(7) and der(22), we mapped BAC clones within and around the duplicated region (Fig. 4B). Two clones (1047E14, 769P3) located within the last intron of C7orf58 and outside of the duplicated region hybridized to chromosomes 7 and der(7). However, all other clones located in the duplicated region hybridized to chromosomes 7, der(7) and der(22), showing that the duplication extends to the segment of chromosome 7 that has been translocated to der(22) and was thus not tandem duplicated on der(7). Incidentally, BAC 146J04 spans

Figure 2. Array CGH array of the index patient showing duplication and amplification of DNA sequences on chromosomes 7 and 22. (A) Chromosome 7 and 22 plots displaying gains of DNA sequences, shown in boxes, in the 7q31.3 and 7q22 regions. (B, C) Detailed plots of oligonucleotide probes from the microarray along each region of interest denoting the genes and their chromosomal coordinates (hg18) in the duplicated and amplified regions of chromosomes 7 and 22, respectively.
doi:10.1371/journal.pone.0052353.g002
the tail end of C7orf58 and all of WNT16 and FAM3C. Therefore, we mapped restriction fragments DNA probes from BAC 146J04 to confirm that those genes had indeed been translocated onto der(22). The map in Figure 4C shows that the 11,085 bp and 16,617 bp Bam HI fragments, the 17,923 bp, Hind III fragment and the 27,092 bp Bam HI fragment from the opposite end, all map to chromosome 7, der(7) and der(22), thus corroborating their location on der(7) and translocation to der(22).

C7orf58, WNT16 and FAM3C

The three genes identified within the duplicated region of chromosome 7 are C7orf58 (also called FLJ21986 and CPED1), WNT16 and FAM3C (also called GS3786). The closest two genes on either side of the duplicated region are ING3, 13 kb upstream of C7orf58, and PTPRZ1, 475 kb downstream of FAM3C. Information about the expression and structure of C7orf58, WNT16, and FAM3C was derived from NCBI gene resources (http://www.ncbi.nlm.nih.gov/gene). The novel C7orf58 gene spans approximately 306,038 bp from the initiation to termination codons of its longest isoform and generates at least three protein isoforms of 117.6 kD, 89.2 kD and 22.6 kD that are encoded by exons 1–23 (excluding exon 18), 1–18 and 1–3A, respectively. Another 64.6 kD isoform, which originates from an alternate promoter is encoded by exons 5–18. The WNT16 gene spans approximately 16 kb from the initiation to termination codons. It encodes two protein isoforms of 40.5 kD, 40.7 kD, based on an alternate initiation site of translation that encompasses exon 1A or exon 1B. The two resulting isoforms, termed WNT16a and WNT16b differ by 30 amino acids at their amino terminus. FAM3C, a member of a gene family with sequence similarity, encodes a ubiquitously expressed 227 amino acid secretory cytokine-like protein [28]. It spans approximately 32.5 kb and encodes a polypeptide chain of 24.7 kD, encompassing 9 exons. All three genes are expressed in multiple tissues except for the WNT16a isoform, which is restricted to the pancreas [29].

Mutation Analysis of C7orf58, WNT16 and FAM3C

We found 21 SNPs in C7orf58, 5 SNPs in WNT16 and 4 SNPs in FAM3C (Table S1). Two previously unreported SNPs, INDEL GAAAA and INDEL AAAACTT were located in intron 13 of C7orf58 and intron 7 of FAM3C, respectively. Allelic frequencies of 8 SNPs (rs78458119, rs3390520, rs68173311, INDEL GAAAA, rs75556099, rs3832519, INDEL AAAACTT, rs3837124) occurred at fairly common frequencies (range 0.065–0.5) to justify their clinical significance as the sole cause of a rare disorder. However, their individual or combined effects may add to the effect of the duplication. Table S1 also lists the potential effects of SNPs that were analyzed with various online bioinformatics tools. Overall, among the 30 SNPs, there were 11 potentially functional SNPs (pfSNPs). The most significant change was rs2110277 in the alternate

Figure 3. Representative fluorescent in situ hybridization (FISH) patterns of DNA probes onto chromosomes 7, 22 der(7) and der(22) from this patient’s chromosome spreads. (A) Schematic drawing of normal chromosomes 7 and 22 and derivative chromosomes der(7) and der(22), showing the translocated chromosomal regions. The three panels below depict each the typical hybridization of a probe located (B) on chromosomes 7, der(7), (C) on chromosomes 7, der(7), der(22) and (D) chromosomes 7 and der(22).

doi:10.1371/journal.pone.0052353.g003
promoter region of the 64.6 kD isoform, which is predicted to differentially bind transcription factors (Fig. S1). The A allele constitutes a potential binding site for the glucocorticoid receptor beta (GR-β), the yin-yang (YY1) activator/repressor and the CCAAT-enhancer-binding proteins (CEBPα, CEBPβ), while the G allele eliminates these DNA binding sites and generates instead a DNA sequence recognized by the early growth response (EGR) transcription factors 1–3. No SNP variation was found in the promoter region upstream of the 117.6 kD isoform. Cryptic acceptor splice sites were likely to be generated by rs55666906, rs798949 and rs68173311 while reduced splice acceptor efficiencies were predicted by rs55666906 and rs111663168. One SNP (rs61128227), located just 7 nucleotides downstream of the intron 12 splice donor site, creates a new splice donor site, which when used, would cause the insertion of three in-frame amino acids, producing an elongated protein. Furthermore, two SNPs (rs41281692, rs10953934) that cause missense mutations (A551G, D554N) were found to constitute conservative changes and are unlikely to be functionally significant. Two other SNPs (rs798948, rs10241888) were predicted to bind mature miRNAs (hsa-miR-449c-5p and hsa-miR-4733-3p) even though rs798948 is a synonymous amino acid change. Previous studies have shown that a synonymous variant could cause a significant change in miRNA binding site [30].

In WNT16, the most deleterious effect was the homozygous CCAAT DEL allele (rs755560) at codons 6–7, which produces a
truncated WNT16a polypeptide of 46 amino acids instead of the normal 361 amino acids. Also, the missense mutation G72R (rs2908004) was found to have a deleterious effect (by SIFT and SNPeffect), whereas the T253I (rs2707466) was a tolerant change by SIFT. On the other hand, the threonine residue anticipated to be phosphorylated by NetPhos (score 0.833) would be lost by the substitution to isoleucine and the proband is heterozygous. In contrast, rs3832519 has a lower probability of substitution to isoleucine and the proband is homozygous for the INS allele of the WNT16a isoform of 64.6 kD isofrom of C7orf58, which is predicted to differentially bind transcription factors and the DEL protein truncating mutation (rs75556099) at codon 6 of the WNT16a isoform that produces an aberrant polypeptide chain of 46 amino acids. The differential binding of transcription factors could alter the expression of C7orf58 and cause ectopic expression. As for the truncation of the WNT16a isoform, its restricted expression to the pancreas [29] coupled with its allelic frequency of 0.312 makes it more likely to be associated with a frequent disorder such as obesity, rather than the rare CSS-related disorders. In this vein, the relationship of WNT proteins to adipose tissue is exemplified by WNT10a, which suppresses adipocytes differentiation [42] and inhibits obesity in genetically obese mice [43]. Hence, it is conceivable that wild-type WNT16a could play a similar role, whereas its deletion mutant allele (rs75556099) could reverse this effect by promoting adipocytes differentiation and causing obesity. While the early onset obesity and hyperphagia of this patient are not associated with endocrine abnormalities, mutations in leptin and leptin receptor were ruled out based on her elevated leptin levels, which correlated with her BMI of 34.4 [44] and with the observation that heterozygous leptin null mutations result in significantly lower than normal leptin levels [45]. Furthermore, exome sequencing did not reveal a deleterious mutation in any of the known obesity genes. Yet another possibility is a translocation-mediated change with bone mineral density [34], consistent with the short height of this patient. Another deleterious effect could result from the disruption of C7orf58 at its duplicated breakpoint in the last intron, which would truncate its longest isoform. As the pathogenesis of C7orf58 remains unknown, it is interesting to note that a patient with mental retardation, anxiety disorder and autistic features was reported to have a complex 7q rearrangement that also appeared to disrupt C7orf58 [35]. While our patient did not exhibit the major types of symptoms as the patient reported by Dauwerse et al. [35], it is unclear what is the phenotypic contribution of a disrupted C7orf58 gene in the context of a large chromosomal rearrangement or a smaller duplication that encompasses multiple genes. Yet another possibility is a translocation-mediated change in chromatin, which could impact on the expression of the nearby gene, ING3 that is located just 13 Kb upstream of C7orf58. Most interestingly, the protein encoded by ING3 contains a PHD-finger domain, a common motif involved in chromatin remodeling genes [36] that are increasingly being recognized as target genes for developmental disorders, such as recently in Potocki-Shaffer syndrome, which is caused by disruption of a BRAF-histone deacetylase gene [37]. While no potentially functional SNPs were found by exome sequencing in ING3, chromatin disruption could interfere with the regulated expression of its allele near the duplicated region. Interestingly, the association of allelic loss and reduced expression of ING3 with head and neck cancers [38] is reminiscent of a similar association in chromatin remodeling genes of the SWI/SNF complex in CSS [3,4] with tumorigenesis and gastric cancer [39,40,41].

It is also possible that functional SNPs could independently and further contribute to the deleterious effect of the duplication. Exonic and junctional intronic SNPs in C7orf58, WNT16 and FAM3C suggested potentially deleterious SNPs in these three genes. Among these, the two most significant SNPs were rs2110277 in the alternate promoter region of the 64.6 kD isofrom of C7orf58, which is predicted to differentially bind transcription factors and the DEL protein truncating mutation (rs75556099) at codon 6 of the WNT16a isoform that produces an aberrant polypeptide chain of 46 amino acids. The differential binding of transcription factors could alter the expression of C7orf58 and cause ectopic expression. As for the truncation of the WNT16a isoform, its restricted expression to the pancreas [29] coupled with its allelic frequency of 0.312 makes it more likely to be associated with a frequent disorder such as obesity, rather than the rare CSS-related disorders. In this vein, the relationship of WNT proteins to adipose tissue is exemplified by WNT10a, which suppresses adipocytes differentiation [42] and inhibits obesity in genetically obese mice [43]. Hence, it is conceivable that wild-type WNT16a could play a similar role, whereas its deletion mutant allele (rs75556099) could reverse this effect by promoting adipocytes differentiation and causing obesity. While the early onset obesity and hyperphagia of this patient are not associated with endocrine abnormalities, mutations in leptin and leptin receptor were ruled out based on her elevated leptin levels, which correlated with her BMI of 34.4 [44] and with the observation that heterozygous leptin null mutations result in significantly lower than normal leptin levels [45]. Furthermore, exome sequencing did not reveal a deleterious mutation in any of the known obesity genes, suggesting that, if obesity was related to this translocation/duplication, the culprit region would likely fall on any or a combination of genes involved in the duplication. Overall, the molecular basis of disease in this patient appears to be confined to the three-gene cluster, C7orf58, WNT16 and FAM3C. The molecular characterization of these genes in patients with a phenotype resembling CSS and overlapping with its clinical features will vindicate their clinical associations. Eventually, to unveil the biological functions of these genes and to delineate their contributions to the various clinical phenotypes will require their targeted disruption, singly and in combination, in animal models.

Supporting Information

Figure S1. Effect of rs2110277 (A/G) on transcription factors binding sites in the C7orf58 alternative promoter. The PROMO computer prediction software shows that the A and G alleles differ in their binding of the transcription factors early growth response 1–3 (EGR), glucocorticoid receptor beta (GR-β), yin-yang (YY1) activator/repressor, CCAAT-enhancer-binding proteins (CEBPα, CEBPβ). (TIF)

Table S1. Amplicon sizes and DNA sequences of PCR primers used for the amplification and sequencing of
exons, surrounding introns, untranslated sequences and immediate promoter regions of the WNT16, FAM3C and C7orf58 genes. The locations of the primers within each gene are donated by their closest exon (Ex), promoter (P) or untranslated (UTR) regions.

References

1. Coffin GS, Siris E (1970) Mental retardation with absent fifth frenum and terminal phalanx. Am J Dis Child 119: 433–439.
2. Tunnissen WW Jr, McMillan JA, Levin MB (1978) The Coffin-Siris syndrome. Am J Dis Child 132: 393–395.
3. Santen GW, Aten E, Sun Y, Alimomani R, Gilson C, et al. (2012) Mutations in SWI/SNF chromatin remodeling complex gene ARID1B cause Coffin-Siris syndrome. Nat Gen 44: 379–380.
4. Tsurusaki Y, Okamoto N, Ohashi H, Kosho T, Imai Y, et al. (2012) Mutations affecting components of the SWI/SNF complex cause Coffin-Siris syndrome. Nat Gen 44: 376–378.
5. Schirrer SA, Bodurtha JN, Burton B, Chudley AE, Chong MA, et al. (2012) The Coffin-Siris syndrome: A proposed diagnostic approach and assessment of 15 overlapping cases. Am J Med Genet A 158A: 1085–1087.
6. Brautbar A, Ragsdale J, Shinawi M (2009) Is this the Coffin-Siris syndrome or the BOD syndrome? Am J Med Genet A 149A: 559–562.
7. Elliott AM, Teebi AS (2000) New autosomal dominant syndrome reminiscent of Coffin-Siris syndrome and Brachymorphism-Onychodysplasia-Dysphalangism syndrome. Clin Dys 9: 15–19.
8. Senior B (1971) Impaired growth and onychodysplasia. Short children with tiny toenails. Am J Dis Child 122: 7–9.
9. Fryns JP (1986) On the nosology of the Cornelia de Lange and Coffin-Siris syndromes. Hum Genet 72: 523–530.
10. Musio A, Selicorni A, Focarelli ML, Gervasoni C, Milani D, et al. (2006) X-linked Cornelia de Lange syndrome owing to SHC1L1 mutations. Nat Gen 38: 520–523.
11. Fleck BJ, Pandya A, Vanner L, Kerkering K, Bodurtha J (2001) Coffin-Siris syndrome: review and presentation of new cases from a questionnaire study. Am J Med Genet A 99: 1–7.
12. McGhee EM, Klump CJ, Bitts SM, Cotter PD, Lamber EJ (2000) Candidate region for Coffin-Siris syndrome at 7q32.3–34. Am J Hum Genet A 93: 241–243.
13. Yonis JJ (1976) High resolution of human chromosomes. Science 191: 1268–1270.
14. Kallioniemi OP, Kallioniemi A, Kurisu W, Thor A, Chen LC, et al. (1992) ERBB2 amplification in breast cancer analyzed by fluorescence in situ hybridization. Proc Natl Acad Sci U S A 89: 5321–5325.
15. Massuger NC, Escudero R, Farré D, Núñez O, Martinez J, et al. (2002) PROMO: detection of known transcription regulatory elements using species-tailored searches. Bioinformatics 18: 333–334.
16. Schug J (2008) Using TESS to predict transcription factor binding sites in DNA sequence. Current protocols in bioinformatics/editorial board, Andrews D Baxevanis, et al. Chapter 2: Unit 2 6.
17. Kumar P, Henikoff S, Ng PC (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 4: 1073–1081.
18. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, et al. (2010) A method and server for predicting damaging missense mutations. Nat Meth 7: 249–249.
19. Schwarz JM, Rodelsperger C, Schauel M, Seelow D (2010) MutationTaster evaluates disease-causing potential of sequence alterations. Nat Meth 7: 575–576.
20. Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustrès M, et al. (2009) Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucl Acids Res 37: e67.
21. Dogan R, Gotoh T, Willum WJ, Mount SM (2007) SplicePort-an interactive splice-site analysis tool. Nucl Acids Res 35: W295–298.
22. Blom N, Gammeltoft S, Brunak S (1999) Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. J of molecular biology 294: 1351–1362.
23. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ (2008) miRBase: tools for microRNA genomics. Nucl Acids Res 36: D154–158.
24. Wheeler DA, Stinarvas M, Egholm M, Shen Y, Chen L, et al. (2008) The complete genome of an individual by massively parallel DNA sequencing. Nature 452: 872–876.

Acknowledgments

We thank Drs. Arthur Beaudet and Craig Chinault for their help in the analysis of the CGH microarray and Dr. Bryan Hall for clinical advice.

Author Contributions

Conceived and designed the experiments: FCC. Performed the experiments: JZ GM JQ. Analyzed the data: FCC JZ GM JQ MA. Wrote the paper: FCC JZ. Clinical Geneticist: MG. Genetic Counselor: AZ. Pathologist: MA.