Short Report

19p13.2 microduplication causes a Sotos syndrome-like phenotype and alters gene expression


Up to 90% of individuals affected by Sotos syndrome have a pathogenic alteration of NSD1 (encodes nuclear receptor-binding Su-var, enhancer of zeste, and trithorax domain protein 1), a histone methyltransferase that functions as both a transcriptional activator and a repressor. Genomic copy number variations may also cause a Sotos-like phenotype. We evaluated a three-generation family segregating a Sotos-like disorder characterized by typical facial features, overgrowth, learning disabilities, and advanced bone age. Affected individuals did not have a detectable NSD1 mutation, but rather were found to have a 1.9 Mb microduplication of 19p13.2 with breakpoints in two highly homologous Alu elements. Because the duplication included the DNA methyltransferase gene (DNMT1), we assessed DNA methylation of peripheral blood and buccal cell DNA and detected no alterations. We also examined peripheral blood gene expression and found evidence for increased expression of genes within the duplicated region. We conclude that microduplication of 19p13.2 is a novel genomic disorder characterized by variable neurocognitive disability, overgrowth, and facial dysmorphism similar to Sotos syndrome. Failed compensation of gene duplication at the transcriptional level, as seen in peripheral blood, supports gene dosage as the cause of this disorder.

Conflict of interest

No authors have relevant conflicts of interests to report.
Sotos syndrome is characterized by early rapid growth, cognitive impairment, a typical facial appearance (a prominent jaw, high-arched palate, sparse and high frontal hairline, and broad forehead), advanced bone age, and occasional visceral malformations (1). With an estimated incidence of 1 in 14,000, it is one of the most common overgrowth syndromes (2). Up to 90% of patients have alterations of NSD1 (1, 3). NSD1 encodes nuclear receptor-binding Su-var, enhancer of zeste, and trithorax domain protein 1, which functions as a histone methyltransferase that both activates and represses transcription through chromatin modification (4). Some individuals with a Sotos-like phenotype have an associated de novo genomic copy number variant (CNV) (5).

The utility of peripheral tissue transcriptome studies for the investigation of genomic disorders of the human central nervous system (CNS) has been minimally explored (6–8), although it is unclear if expression differences in peripheral tissues relate to CNS function. Nonetheless, because access to brain tissue is limited, a peripheral correlate is needed for CNS phenotypes. Peripheral blood may provide an informative correlate as 82% of brain-expressed genes, probably including many that are canonically considered to be ‘brain specific’, have detectable transcripts in circulating blood cells (9).

We report a duplication of chromosome 19p13.2 segregating with a Sotos-like phenotype across three generations. This duplication, which did not show copy number complexity, probably arose by Alu-mediated recombination and was associated with overexpression of many genes within the duplicated region. We hypothesize that failed dosage compensation of gene duplication at the transcriptional level is the cause of the disorder in this family. All participants provided signed informed consent. The study protocol (H07-02142) was approved by the University of British Columbia institutional review board.

Subjects, materials and methods

Clinical reports

Individual III-5 (Fig. 1a,b), the proband, had a birth weight of 4.55 kg (97th percentile), length 54.5 cm (95th percentile), and head circumference (HC) 37 cm (75th percentile). He has globally delayed development. Formal psychometric testing documented a full-scale IQ of 41 and an autism spectrum disorder. At 5 years, he had a high frontal hairline, pointed chin, high-arched palate, partial choanal stenosis, pectus excavatum, advanced bone age (> +2 SD), and fifth finger clinodactyly. His height was 119 cm (97th percentile) and HC was 53 cm (75–95th percentile) (Fig. 1b). Cranial MRI showed a dilated vestibular aqueduct. NSD1 sequencing and deletion analysis did not detect a mutation.

Individual II-4 (Fig. 1b), mother of the proband, had delayed developmental milestones, including walking. She received special educational support. Her HC was 58.5 cm (98th percentile) and her height was 171 cm (90th percentile). She had a broad forehead, long face, mild dolichocephaly, mild down-slanl of the palpebral fissures, and prominent chin.

Individual I-1 (Fig. 1b), the maternal grandfather of the proband, has limited ability to read or write. He had a long face with prognathism, an HC of 58.5 cm (98th percentile), and large ears (8.2 cm, >98th percentile).

Individual II-2 (Fig. 1b), the maternal uncle of the proband, also had learning difficulties, an HC of 60.5 cm (>98th percentile), a height of 185 cm (90th percentile), a long face, a high-arched palate, prognathism, and large ears (7.5 cm, >98th percentile).

Individual III-2 (Fig. 1b), the son of individual II-2, had a birth weight, length and HC of 4.6 kg (97th percentile), 56 cm (>97th percentile), and 37 cm (75th percentile), respectively. He has globally delayed development. An assessment at 3.5 years placed his receptive and expressive language at the 1st percentile and gross and fine motor skills at <1st percentile. At 4.5 years, he was diagnosed with a non-verbal learning disorder (full-scale IQ 83), particularly affecting visuospatial skills, coordination, and communication. At age 9 years, his HC and height were 58 cm (>98th percentile) and 162 cm (>97th percentile), respectively. He had hypotonia, a double hair whorl, mild down-slanl of the palpebral fissures, oblong face, large ears (7.4 cm, >98th percentile), prominent chin, and high-arched palate. A head CT scan was normal except for a pituitary microadenoma that was unassociated with endocrinological abnormalities.

Individual III-4, the unaffected sister of III-5, had a height of 129 cm (5th percentile) at age 10 years and a round, rather than long, face. Individual III-3, the unaffected brother of III-2, had a height of 167 cm (90th percentile) at age 13 years. These siblings have had normal development and above-average academic performance.

Array genomic hybridization

Peripheral blood DNA from individual III-5 was tested with the SignatureChipOS 1.0 oligonucleotide array (Signature Genomics, Spokane, WA),
Fig. 1. (a) Pedigree of the family showing segregation of the duplication with individuals affected by learning disabilities. Individuals III-3 and III-4 have above-average academic performance and do not carry the duplication. Individual III-1, age 4 years, has not been tested for the duplication and has normal developmental milestones. (b) Photographs of individuals with duplication 19p13.2. Despite individual differences in appearance, note the shared features of a high forehead, high frontal hairline, long face, macrocephaly, large ears, and prominent chin. Individual I-1 at ages 10 and 74 years; individual II-2 at ages 6 and 40 years; individual II-4 at ages 6 months, 5, 12, and 45 years; individual III-2 at ages 1 month, 3, 8 and 10 years; and individual III-5 at ages 1 month, 12 months, 3 and 5 years.

which has 105,000 oligonucleotides targeting 1543 loci and with the Affymetrix 6.0 SNP array (Santa Clara, CA), which has 1.8 million probes. Data from the hybridized and scanned Affymetrix array were analyzed using the AFFYMETRIX POWER TOOLS (version 1.6.0) software suite (http://www.
affymetrix.com; Affymetrix, Santa Clara, CA). A
copy number loss or gain relative to gender-
matched controls was determined by a combined
log2 ratio of test/control < -0.2 or > +0.2, respec-
tively, and a p value <0.05. Genomic coordinates
were based on the NCBI Build 36.1 (March 2006)
reference sequence.

Fluorescent in situ hybridization

Fluorescent in situ hybridization (FISH), both stan-
dard and halo, was performed on cultured peripheral
blood T-lymphocytes by Signature Genomics
and by D. C. and P. E. Probes included BACs
RP11-365L4 (19p13.2), RP11-126M21 (19q tel)
and fosmids G248P81532G3 and G248P83309H8.
Probe labeling, hybridization and image capture
were conducted as described (http://international.
abbottmolecular.com).

DNA methylation analysis

DNA methylation was semiquantitatively mea-
sured at 1506 CpG sites within promoter regions of
807 genes in genomic DNA isolated from peripheral
blood mononuclear cells or buccal mucosa. After bisulfite treatment of 500 ng genomic
DNA, the samples were subjected to the Illumina
GoldenGate Methylation Cancer Panel 1 array-
based assay, using reagents and protocols provided
by the manufacturer (Illumina, San Diego, CA).
Subsequent data acquisition and analysis were per-
formed with the ILLUMINA BEADSTUDIO software
(Illumina). A β-value of 0–1 (corresponding to
the percentage of methylation) was calculated for
each CpG site by determining the ratio of methyl-
ated signal intensity to the sum of both methylated
and unmethylated signals for each site and sub-
tracting background. Technically unreliable probes
with a confidence score of <0.05 were excluded
from analysis. To avoid confounding the analysis
with gender-specific DNA methylation signatures,
all CpG sites located on the X chromosome were
excluded from the analysis. The Pearson’s cor-
relation test was applied to test for the presence
or absence of variation between samples for each
promoter.

Whole genome mRNA expression analysis

cDNA from II-4, II-5, III-5, and 77 normal indi-
viduals was hybridized in parallel experiments to
the HumanRef-8 BeadChip (Illumina), which tar-
gets approximately 24,500 RefSeq-annotated tran-
scripts with 50 base pairs (bp) oligonucleotide
probes close to the 3′ end of the transcript.

The expression data on 77 samples have been
previously published and uploaded to Gene
Expression Omnibus (accession no. GSE15180,
were analyzed with BEADSTUDIO software (Illumi-
na). Any gene having one probe with a detection
p value >0.05 in one individual (either control or experimental) was removed from further study.
Comparisons of expression values for the 19p13.2
region from the affected vs control individuals
were made using the Student’s t-test and ANOVA.

Results

The morphological features, overgrowth and learn-
ing disabilities observed among affected indi-
viduals in this family are similar to those seen in
individuals with Sotos syndrome (Fig. 1, Table 1).
The facial features are most suggestive of Sotos
syndrome when the individuals are young and then
lose similarity with maturation. The facial gestalt

<table>
<thead>
<tr>
<th>Features</th>
<th>Number of family members</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facial phenotype</td>
<td></td>
</tr>
<tr>
<td>High forehead</td>
<td>4/5</td>
</tr>
<tr>
<td>Macrocephaly</td>
<td>4/5</td>
</tr>
<tr>
<td>Downslanting</td>
<td>2/5</td>
</tr>
<tr>
<td>palpebral fissures</td>
<td></td>
</tr>
<tr>
<td>Long face/jaw</td>
<td>5/5</td>
</tr>
<tr>
<td>Hypertelorism</td>
<td>0/5</td>
</tr>
<tr>
<td>Pointed chin</td>
<td>4/5</td>
</tr>
<tr>
<td>High palate</td>
<td>5/5</td>
</tr>
<tr>
<td>Developmental delay</td>
<td>5/5</td>
</tr>
<tr>
<td>Early rapid overgrowth</td>
<td>2/2</td>
</tr>
<tr>
<td>Advanced bone age</td>
<td>1/2</td>
</tr>
</tbody>
</table>

*Number of family members = number affected over the number
assessed for the features.
of the proband (III-5) is most similar to that of classical Sotos syndrome. We did not identify an NSD1 alteration in III-5, but did detect a copy number gain within 19p13.2 [arr cgh 19p13.2 (9,109,407–11,068,542 bp) ×3 mat] (Fig. 2a,b). FISH using BAC probes RP11-365L4 (19p13.2) and RP11-126M21 (19q tel) and fosmid clones G248P81532G3 and G248P83309H8 confirmed the duplication and showed that it is a direct repeat (Fig. 2c). The duplication segregates with learning disability through three generations. No other CNVs of interest were detected. Unaffected family members (III-3 and III-4 in Fig. 1a) do not have the duplication and have normal stature and neurodevelopment.

The duplication breakpoint is within an AluY element (9,107,661–9,107,688) and an AluSc5 element (11,094,025–11,094,052, Fig. 2d) that are oriented in the same direction. The 150 bp around the breakpoints within each Alu are 79.2% identical (EMBOSS needle alignment, www.ebi.ac.uk) and have 100% identity at the 28 bp spanning the breakpoint. Breakpoint sequencing in the transmitting mother and her son did not detect instability during meiosis (data not shown).

Within the literature and DECIPHER database (https://decipher.sanger.ac.uk), four individuals have developmental disorders arising from deletions or duplications encompassing this region (11) (Fig. 2e). However, none have breakpoints or clinical features matching those in this family. As assessed by GNF (Genomics Institute of the Novartis Research Foundation) expression atlas 2, the majority of genes within the 19p13.2 duplication show some expression in the brain, complicating discernment of a candidate gene responsible for the intellectual disability. The Database of Genomic Variants (http://projects.tcag.ca/variation; (12)) lists numerous small CNVs within this region among individuals with normal phenotypes (Table S2, Supporting information online).

Compared to controls, the fold expression difference of genes within the duplication vs the remainder of chromosome 19 is increased (1.47-fold vs 1.08-fold; ANOVA; p = 3.5 × 10−15) (Fig. 2f). Thirty-three of the 52 RefSeq-annotated genes within the interval were represented on the array and passed stringency criteria; of these, several showed statistically significant overexpression (Table S3, Supporting information online) and four were validated by independent qRT-PCR experiments: ILF3 (interleukin enhancer-binding factor 3), ZNF266 (zinc finger protein 266), DNMT1 (DNA methyltransferase gene 1), and SMARCA4 (Swi/Snf-related, matrix-associated, actin-dependent regulator of chromatin, subfamily A, member 4) (data not shown).

Because DNMT1 resides in the duplication and was overexpressed, we checked for DNA methylation changes in the buccal mucosal and peripheral blood mononuclear cells (PBMCs) of III-5 and II-4 using the Illumina Golden Gate Cancer Panel I array (Illumina). However, their DNA methylation was comparable to that of unaffected individuals (Fig. 2g,h).

Discussion

19p13.2 duplication syndrome

We report a family with features similar to Sotos syndrome and a duplication of chromosome 19p13.2 segregating through three generations. This syndrome is characterized by overgrowth, a long midface, high forehead, large ears, prominent chin, a high-arched palate, advanced bone age, and variable cognitive disability.

Several observations suggest that the 19p13.2 duplication is pathogenic. First, <1% of control individuals have a CNV larger than 1.5 Mb (12). Second, the syndrome segregates with this CNV and the phenotypic variability is consistent with that reported for other CNV-associated diseases (13). Third, several genes in the region are known to have important roles in brain development or function, including PIN1 (peptidyl–prolyl cis/trans isomerase, NIMA-interacting), DNMT1, and SMARCA4 (14–16). Interestingly, the recently described 19p13.13 deletion syndrome also features overgrowth and intellectual disability, and although one might speculate that cis regulatory elements interact across the 1 Mb between these two regions, our patients did not have repressed expression for any of the 16 genes within the 19p13.13 deletion interval (17).

Mechanism of duplication

Because CNV breakpoints are determined by statistical analysis of continuously variable and somewhat noisy signal intensity data, arrays for CNVs do not define precisely the boundaries of genomic rearrangements. Consistent with this ambiguity, we found a 25 Kb discrepancy between the molecularly defined and array-defined centromeric breakpoint. This recognized limitation of arrays reinforces the importance of characterizing the breakpoint to define precisely the disease-associated elements (18).
19p13.2 microduplication

**Fig. 2.** (a) Affymetrix 6.0 SNP array copy number estimations for the whole of chromosome 19, showing a duplication within the p-arm. The region absent from signals corresponds to the pericentromere. (b) Closer view of the end points of the duplication from microarray data. The centromeric endpoint is clearly less well-defined than the sharper telomeric endpoint. (c) Fluorescent *in situ* hybridization images showing tandem orientation of the duplication using fosmid probes G248P82303H7 (11,066,995–11,05,393; FITC green) and G248P81532G3 (9,287,088–9,326,932; SpO red). The cell nucleus is labeled blue with DAPI. (d) Sequenced junction points of the tandem 19p13.2 duplication. The breakpoint lies within a repetitive element common to both breakpoints. (e) Microdeletions (in dark gray) and microduplications (light gray) in individuals with developmental disorders involving 19p13.2. The copy number of DNA methyltransferase gene 1 is altered in all cases. (f) The difference in fold expression over controls within 19p13.2 vs the remainder of chromosome 19 in affected individuals (ANOVA). (g) Scatter plot showing no significant differences in the methylation status between affected and unaffected individuals as detected in buccal epithelial cells. Pearson’s coefficient = 0.989. (h) Scatter plot showing no significant differences in the methylation status between affected and unaffected individuals as detected in peripheral blood mononuclear cells. Pearson’s coefficient = 0.995. ASD, atrial septal defect; ID, intellectual disability; VSD, ventricular septal defect.
Lehman et al.

We identified the duplication breakpoint for our patients within two Alu elements. Highly similar pairs from the Alu family are known to predispose to non-allelic homologous recombination (NAHR) (19). Several studies have described an excessive presence of Alu elements at or near the breakpoints of human CNVs (20, 21), including CNVs suspected of arising from non-homologous end joining (NHEJ) or from fork stalling and template switching (FoSTes) (21, 22). However, we did not observe the small breakpoint insertion or deletion, or microhomology characteristic of NHEJ for this duplication. Similarly, we did not detect characteristic FoSTes copy number variation (e.g., triplication, deletion) within the duplicated region. Instead, the features of the breakpoint, including same orientation of the repetitive elements, implicate NAHR as the origin of the duplication.

Abnormal gene expression from the duplicated region

Expression array studies in human genomic duplication disorders have shown that some but not all duplicated genes have increased mRNA steady-state levels in multiple tissues such as brain and blood (6, 7, 23). Of genes overexpressed in the brains of fetuses or adults with Down syndrome, 60% and 84%, respectively, are also overexpressed in trisomy 21 lymphoblastoid cells (6, 7). In the case of MECP2 duplication syndrome, increased MECP2 mRNA is observed in peripheral blood cells and hypothesized to be overexpressed in the brain (23). Also, fibroblast cells from a girl with a much larger 19p13.2 duplication and intellectual disability, short stature, and precocious puberty had >1.3-fold overexpression of nearly half of the duplicated genes (11). Large-scale studies of gene expression of affected and unaffected populations will determine if altered gene expression correlates with pathogenicity or if benign CNVs show similar gene expression changes but do not include critical dosage-sensitive genes.

In summary, this 19p13.2 duplication syndrome bears similarity to Sotos syndrome, including the craniofacial dysmorphisms, advanced bone age, greater-than-expected growth, and variable learning disabilities. The duplication leads to overexpression of some genes within the duplicated interval, but further study of individuals with overlapping CNVs and model organisms are needed to delineate the critical region(s) or genes for overgrowth and learning disability.

Supporting Information

The following Supporting information is available for this article:

Table S1. Primer sequences
Table S2. Genes from the duplicated 19p13.2 region previously described in the Database of Genomic Variants as deleted or duplicated in the general population
Table S3. Fold difference in the expression of duplicated 19p13.2 genes in two affected individuals compared to controls (Student’s t-test)

Additional Supporting information may be found in the online version of this article.

Please note: Wiley-Blackwell Publishing is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Acknowledgements

The authors thank the family for their support of this study. The authors thank Drs Jan M. Friedman, Evica Separovic, and Ms Rosemarie Rupps for critical review of this manuscript. The gene expression assays were performed at the Génome Québec Innovation Centre (Montréal, Canada). This study was funded in part by the British Columbia Clinical Genomics Network, a Child and Family Research Institute Establishment Award (C. F. B.), and the Clinical Genomics Platform of the Michael Smith Foundation for Health Research (C. F. B.). C. F. B. and M .S. K are scholars of the Michael Smith Foundation for Health Research.

References

11. Lybaek H, Orstavik KH, Prescott T et al. An 8.9 Mb 19p13 duplication associated with precocious puberty and a sporadic
3.9 Mb 2q23.3q24.1 deletion containing NR4A2 in mentally retarded members of a family with an intrachromosomal 19p-into-19q between-arm insertion. Eur J Hum Genet 2009: 17: 904–910.