Clinical and genetic analysis of a rare syndrome associated with neoteny

Richard F. Walker, PhD¹, Serban Ciotlos, BS², Qing Mao, PhD², Robert Chin, MS², Snezana Drmanac, MD², Nina Barua, MS², Misha R. Agarwal, PhD², Rebecca Yu Zhang, BS², Zhenyu Li, MS³, Michelle Ka Yan Wu, BS², Kevin Sun, BS², Katharine Lee, BS², Staci Nguyen, BS², Jia Sophie Liu, PhD², Paolo Carnevali, PhD², Radoje Drmanac, PhD^{2,3} and Brock A. Peters, PhD^{2,3}

Purpose: We describe a novel syndrome in seven female patients with extreme developmental delay and neoteny.

Methods: All patients in this study were female, aged 4 to 23 years, were well below the fifth percentile in height and weight, had failed to develop sexually, and lacked the use of language. Karyotype and array chromosome genomic hybridization analysis failed to identify large-scale structural variations. To further understand the underlying cause of disease in these patients, whole-genome sequencing was performed.

Results: In five patients, coding *de novo* mutations (DNMs) were found in five different genes. These genes fell into similar functional categories of transcription regulation and chromatin modification. Comparison to a control population suggested that individuals with neotenic complex syndrome (NCS)—a name that we propose

herein—could have an excess of rare inherited variants in genes associated with developmental delay and autism, although the difference was not significant.

Conclusion: We describe an extreme form of developmental delay, with the defining characteristic of neoteny. In most patients we identified coding DNMs in a set of genes intolerant of haploinsufficiency; however, it is not clear whether these contributed to NCS. Rare inherited variants may also be associated with NCS, but more samples need to be analyzed to achieve statistical significance.

Genet Med advance online publication 21 September 2017

Key Words: *de novo* mutation; developmental delay; long fragment read technology; neoteny; whole-genome sequencing

INTRODUCTION

In 2009 we reported the case of a 16-year-old girl suffering extreme developmental delay associated with multiple organ system dysfunctions.¹ These included structural and functional defects of her nervous, gastrointestinal, respiratory, cardiac, and musculoskeletal systems. Most striking was her immature facies, which was that of a toddler rather than a teenager (**Figure 1**, top left picture). Because of dysmorphic features and her unusually juvenile appearance, associated with multiple congenital anomalies (MCA), her pediatrician felt that the condition had a genetic basis. However, in karyotyping and array chromosome genomic hybridization analysis, she appeared to have the genome of a normal female. Owing to the absence of published reports on similar conditions, her pediatrician felt that her condition was unique and unknown, so he called it “syndrome X.” Subsequently, based on her poorly integrated and significantly slow rate of somatic remodeling during childhood, we proposed that gene(s) affecting the development/aging continuum were mutated. If so, such mutation(s) might be identified by DNA sequencing and thereby provide a better understanding of genetic factors contributing to human development

and aging.^{1,2} However, she died from complications of tracheomalacia before we could sequence her genome. We have since been contacted by many families seeking information about their children experiencing developmental delay, as a result of the press coverage of the first case. While most of these cases were unlike the index patient, seven families presented with children having similar congenital anomalies and all displayed neoteny. In this report we describe, both clinically and genetically, this novel syndrome, and we formally name it the neotenic complex syndrome (NCS).

MATERIALS AND METHODS

Subjects in this study were not recruited through advertisement nor offered any monetary compensation or material incentive to participate. All families contacted us asking to participate in the study after viewing one or more of the media documentaries or articles about “syndrome X.” Those whose children met the inclusion criteria (**Table 1**) for the study signed BGI institutional review board–approved informed-consent documents, pursuant to current regulations governing human subject research.

¹Retired; ²Complete Genomics, Inc., San Jose, California, USA; ³BGI–Shenzhen, Shenzhen, China. Correspondence: Richard F. Walker (drwalker@gmail.com), Radoje Drmanac (rade@completegenomics.com), Brock A. Peters. (bpeters@completegenomics.com)
The first four authors contributed equally to this work.

Submitted 15 December 2016; accepted 20 June 2017; advance online publication 21 September 2017. doi:10.1038/gim.2017.140

Figure 1 Physical appearance of girls diagnosed with neotonic complex syndrome. Chronological ages at the time of these photos from top left to bottom right are 16, 6, 3, 10, 4,13, 23, and 18 years, respectively. The family of subject #44 chose not to provide material for genetic analysis. Consent to publish photographs was obtained from the family of each subject.

Table 1 Study inclusion criteria

1. Delayed physical and functional development, causing them to appear significantly younger than their chronological ages
2. Failure to thrive: retarded growth rate with height more than three standard deviations below the mean for chronological age and sex; and failure of hormonal (because unresponsive to human growth hormone (hGH) or thyroid hormone) and/or nutritional (hypercaloric diet) interventions intended to overcome the deficit
3. Problems in swallowing, gastric reflux, bloating chronic constipation, associated gastrointestinal developmental disorders requiring nasogastric or gastrostomy tube for feeding
4. Minimal cognitive development, lack of language or speech, visual and/or hearing defects
5. Neurological deficits, seizures, and hypotonia or hypertonia
6. Structural brain abnormalities in some subjects, including, but not limited to, agenesis of the corpus callosum, pachygyria, lissencephaly, or polymicrogyria
7. Dysmorphic features
8. Atrial septal defect, patent foramen ovale
9. Clubfeet and hip dislocations
10. Laryngotracheomalacia and reactive airway disease

The first two criteria were required. All subjects met most of the remaining criteria, although differing in severity.

Fourteen of 21 suspected NCS children did not meet the inclusion criteria because they exhibited growth problems of known etiology or did not present with neoteny. However, seven suspected cases met the inclusion criteria and presented no evidence of chromosomal abnormalities according to a karyotype and/or array chromosome genomic hybridization analysis (**Supplementary Table S1** online). We consider these seven female subjects as representing “pure” cases of NCS (**Figure 1**).

The families of these seven subjects provided copies of medical records accumulated over the course of their children’s lives. These, along with personal experiences conveyed during interviews, were carefully reviewed and compared with those of the index patient to determine whether they displayed comparable clinical profiles. Once a relatively confident clinical diagnosis of NCS had been established, blood was collected from the six subjects whose families agreed to undergo genetic analyses

and from as many of each subject's direct bloodline relatives as possible (**Supplementary Table S2**). Descriptions of the methods used to analyze each genome can be found in **Supplementary Materials**.

RESULTS

Clinical characteristics

The defining physical characteristic of those affected with NCS is their retention of juvenile phenotypes, so that they present a biologically younger appearance than that corresponding to chronological age (**Figure 1**). Of course, this trait is not evident at birth, but it can be detected as early as 3 years of age, becoming more obvious with the passage of time. However, in addition to neoteny, there is a spectrum of developmental disorders that precede its emergence and are apparent at birth. A composite description of the most common NCS attributes as well as tables listing phenotypes can be found in the **Supplementary Materials** and **Supplementary Table S3**.

Whole-genome sequence analysis of patients

The genome of each patient whose family consented to analysis, and her unaffected siblings (**Supplementary Table S2**), was sequenced to a depth of $50\times$ and $100\times$ by Complete Genomics' standard (STD)³ and long fragment read (LFR)⁴ technologies, respectively. This enabled high call rates across the genome and exome (**Supplementary Table S4**). Using read coverage we confirmed the clinical labs' reports that no large chromosomal amplifications or deletions existed in these patients (**Supplementary Figure S1**). Copy-number variations (CNVs) and structural variations (SVs) were also analyzed using read coverage and unexpected mate-pair mappings, respectively (**Supplementary Tables S5–8**). On average, there were 1.83 de novo small (<15 kb) SVs discovered in each patient and 0.57 in each sibling (**Supplementary Table S8**). This difference was found to be statistically significant by a *t*-test ($P = 0.042$). There were no large de novo SVs or CNVs discovered. No obvious de novo or rare (not found in the 1,000 Genomes Project (1 KG) or Complete Genomics' internal control genomes) inherited disrupting CNVs and SVs were identified.

Approximately 4 million small variants were identified in each genome (**Supplementary Table S4**, row 41). These numbers were similar to those found in previous studies^{3–5} and did not suggest anything unusual about the genomes of NCS patients or their families. LFR data enabled over 99% of heterozygous SNPs to be placed into long phased contigs with an average N50 of 1.27 Mb (**Supplementary Table S9**). Comparing variant calls between LFR and standard libraries and parent–child trios demonstrated over 96% concordance, suggesting that the sequence data quality for these samples is very high (**Supplementary Figure S2**).

Based on interviews with the families of NCS patients it was determined that a recent history of this syndrome does not exist in any of the families (two families had an extended family member with Down syndrome, but it is not clear how

this would be associated with NCS). As such, the expected cause of NCS would be through DNM. In order to detect DNMs, only positions that were called with high confidence and identified as matching the reference genome in both parents were evaluated (Methods section of **Supplementary Materials**). Candidate DNMs were further filtered by removing those found in 1KG,⁵ the Welllderly Project,⁶ and 200 LFR genomes from the Personal Genome Project (PGP)⁷ databases. We discovered ~76 DNMs per patient and ~82 per sibling and found approximately 0.8 coding DNMs per subject and 1.1 per sibling (**Supplementary Tables S10 and S11**). All coding DNMs were confirmed by Sanger sequencing (Methods section of the **Supplementary Materials** and **Supplementary Figures S3 and S4**). Based on these data, patients did not have a higher burden of DNMs than their siblings and the difference in coding DNMs between siblings and subjects was found to be insignificant (**Supplementary Table S12**). Using LFR allowed the determination of the parent of origin for most DNMs. Plotting the total number of phased DNMs for each parent versus parental age at birth (**Supplementary Figure S5**) resulted in the expected pattern of approximately one additional DNM per paternal year of age, but a much smaller increase in DNMs per maternal year of age.^{8,9} This approximate pattern was observed for both patients and unaffected siblings.

Comparing the genes with coding DNMs in patients to those in their siblings revealed several important differences (**Table 2**). In patients, genes with coding DNMs were found to be highly constrained,¹⁰ that is, intolerant to variation based on data from over 100,000 disease free individuals (exome aggregation consortium database, ExAC¹¹), with an average missense Z-score of 4.46 and a probability of loss of function intolerance (pLI) of 0.96. A missense z-score above 3.09¹⁰ and a pLI score greater than 0.9¹¹ are considered to show significant constraint. By comparison the genes affected by DNMs in the siblings had a much lower average missense z-score and pLI of 1.30 and 0.41, respectively (**Figure 2**). For both scores this difference between subjects and unaffected siblings was statistically significant (missense z-score $P = 0.0374$ and pLI $P = 0.0165$) based on a *t*-test. Genes with DNMs in the patients fell into similar functional categories of histone modification and gene expression, whereas those in the siblings were more broadly distributed across different categories (**Table 2**). In addition, three genes (*DDX3X*, *TLK2*, and *HDAC8*) with DNMs in our patients were shared with genes found in databases of individuals with intellectual disability and developmental delay (ID/DD)¹² and autism spectrum disorder (ASD),¹³ and an additional gene with a DNM found in patient #8 (*TMEM63B*) was recently identified in a large mouse knockout study¹⁴ as likely to result in disease in humans. In the siblings, only one gene with a DNM (*CACNA1D*) was associated with any diseases (**Table 2**). Examination of inactivation of the other allele of these genes through an inherited variant (coding or noncoding, **Supplementary Table S13**), imprinting (**Supplementary Table S14**), or CNV or SV (**Supplementary Table S15**), did not result in any significant findings. Given

Table 2 Coding de novo mutations in subjects and unaffected siblings

Family	Individual	Individual ID	Gene	Chromosomal coordinates	Impact	Nucleotide change	AA change	Missense score	pLI	Functional category
1	Subject	3	DDX3X	chrX:41206198–41206199	Missense	c.2509C > T	p.Pro551Leu	5.13	1	RNA helicase/transcriptional regulation
3	Subject	8	TMEM63B	chr6:44117682–44117683	Missense	c.1564T > G	p.Phe500Val	5.21	1	Chloride channel
7	Subject	29	TLK2	chr17:60678096–60678097	Missense	c.1810C > T	p.Arg513Trp	5.67	1	Kinase/chromatin assembly
8	Subject	36	CCDC101	chr16:28596246–28596247	Missense	c.273G > A	p.Arg29Gln	3.71	0.9	Histone acetyltransferase
9	Subject	39	HDAC8	chrX:71715087–71715088	Missense	c.809A > T	p.Asn155Lys	2.59	0.9	Histone deacetylase
7	Brother 1	30	SPTBN2	chr11:66481588–66481589	Missense	c.700G > A	p.Ala209Val	4.53	1	Glutamate signaling regulator
7	Brother 1	30	TEP1	chr14:20863756–20863757	Missense	c.1819T > C	p.Met593Val	-1.05	0	Telomerase activity
7	Sister 2	31	PUM2	chr2:20497322–20497323	Frameshift	c1022delIT	p.Thr333His	1.88	1	Translational repressor
7	Sister 2	31	CLHC1	chr2:55407767–55407768	Missense	c.1607G > C	p.Ala420Gly	-2.8	0	Vesicle transport
7	Sister 3	32	BRD9	chr5:889214–889215	Missense	c.693A > G	p.Ile175Thr	1.2	0.3	Chromatin regulation
7	Sister 3	32	IFRD2	chr3:50326053–50326054	Missense	c.1432A > C	p.Phe434Cys	-0.28	0	Interferon-related
7	Brother 2	33	CACNA1D	chr3:53835169–53835170	Missense	c.5288G > A	p.Arg1693His	5.57	1	Calcium channel
10	Sister	43	PREX2	chr8:69046466–69046467	Missense	c.3966G > A	p.Ala1313Thr	1.36	0	Guanine nucleotide exchange factor

AA, amino acid; pLI, probability of loss of function intolerance. Genes affected by de novo mutations.

Figure 2 The pLI (probability of loss of function intolerance) and missense z-score of genes affected by de novo mutations (DNMs) in subjects and siblings. The pLI (a) and missense z-scores (b) were plotted for subjects (red circles) and siblings (blue circles) for all genes affected with DNMs in this study. Jitter was used to separate overlapping data points. The average of each score for subjects and siblings is represented by the horizontal black line in each plot. The difference in scores between subjects and siblings was found to be statistically significant using a t-test.

the high constraint scores of these genes, this is not surprising as loss of function of both alleles in highly constrained genes is likely to be embryonic lethal. However, because the DNMs in these genes are missense, a dominant mechanism of action cannot be ruled out.

DNMs in noncoding regions annotated as functionally important for gene expression and regulation from GREAT¹⁵ and Regulome¹⁶ DBs were also examined. No strongly damaging regulatory variants, as determined by Combined Annotation Dependent Depletion scores¹⁷ in these regions were identified; however, in subject #40 a DNM was discovered within a glycine tRNA gene on chromosome 19 (Supplementary Table S10). This DNM occurs in the anticodon stem of the tRNA and would be expected to destabilize the stem structure.¹⁸ Finally, we examined the mitochondrial sequence of each patient, sibling, and mother for potentially damaging DNMs. In one subject (#39) we found a single base change from a G to an A at position 5540 in the mitochondrial tryptophan tRNA gene in approximately 10% of reads from both LFR and STD libraries. Like the DNM found in subject #40, this variant is predicted to destabilize the anticodon stem of the tRNA.

The majority of genes identified with DNMs in NCS patients have previously been found to harbor DNMs in patients with syndromes other than NCS, suggesting that an inherited genetic background risk for developing NCS could exist. To determine whether there was some recent shared ancestry between NCS patients, a PLINK¹⁹ analysis was performed across all parental genomes and genotype data from 1,397 individuals from the HapMap 3 project.²⁰ This analysis found no more factors denoting relationship between the families of NCS patients than there were between

HapMap 3 samples (**Supplementary Figure S6**). In addition, the mitochondrial and Y chromosome haplogroups of all NCS families were determined, and none are shared between families (**Supplementary Table S2**). These data, taken together, suggest that a recent common ancestor of NCS patients does not exist; however, small tracts of DNA (haplotypes) could still be shared between NCS families. LFR enabled the haplotyping of most regions across the genomes of NCS patients and siblings. Comparing the LFR haplotypes of NCS patients to those of 200 healthy volunteers from the PGP allowed for the filtering of most commonly shared haplotypes from each NCS genome and enabled the discovery of a small region (~150 kb) on chromosome X with a rare haplotype found only in patients #8 and #40 (**Supplementary Table S16**). There are no genes within this region, but it appears to have several regulatory sequences, as evidenced by histone acetylation and DNase I hypersensitivity sites determined as part of the Encyclopedia of DNA Elements (ENCODE) project.²¹ In addition, there are several genes (*AP1S2*, *MRX59*, *MRXSF*, *MRXS21*, *MRXS5*, *PGS*) near this region with variants associated with mental retardation.²² This haplotype is also shared with the sister of patient #40 (individual #43) and was inherited from healthy parents for both patients, suggesting that the region alone is not causative, but could contribute to their condition. We also explored the possibility of inherited causative variants shared between families. Our small sample size limited our ability to associate common variants with NCS. Instead, we focused on variants found in 1% or fewer of the subjects from the 1KG and Welllderly studies. Using this criterion did not result in any coding or noncoding variants being found in more than two subjects in or near genes with high pLI or missense z-scores. In addition, we performed pathway analyses in an attempt to identify variants in genes or networks common to these patients, but this did not result in any significant findings when compared to the genomes of the PGP (**Supplementary Table S17**).

Other than a rare haplotype found in two patients, there appears to be no obvious shared genetic background in the families of NCS patients. However, there could be inherited variations, not shared between NCS families, which increase the risk of developing NCS, when combined with DNMs in specific genes. In order to find recessive risk factors of this type, we examined both alleles of a gene for variants acting in a homozygous or compound heterozygous mechanism of inactivation (**Supplementary Table S13**). We also examined variants found in imprinted regions of the genome (**Supplementary Table S14**) and in regions with loss of one allele from CNVs/SVs in that region (**Supplementary Table S15**). These lists contained some interesting genes found in ID/DD and ASD databases. However, for the most part, genes with truncating variants had low pLI scores and those genes found to have detrimental missense variants tended to have low missense z-scores, making it difficult to associate any of these with NCS. We also compared the average number of rare variants found in NCS subjects in different gene groups

and different MAFs with the average found in the unaffected NCS family members and in the set of 1KG samples sequenced by Complete Genomics using the same technology and analysis pipeline as used for our NCS samples (**Figure 3** and **Supplementary Table S18**). While there are significantly more rare variants with MAF less than or equal to 0.001 in the NCS patients in high missense z-score genes, there is also a larger number of total variants in NCS patients at this MAF. For the most part there were more rare variants in the NCS patients than there were unaffected family members, but this reached statistical significance in only 3 of the 40 comparisons (**Supplementary Table S19**). To further explore if these results were meaningful, we randomly selected 20 sets of six samples from each ethnicity and then compared them as if they were NCS patients. For the MAF of 0.001, we found that approximately 9% of the time a random sampling of six samples could result in a significantly higher number of variants in comparison with at least one ethnic group. However, when we restricted the comparison to only genes with a missense z-score of 3.09 or greater, the proportion of significant comparisons dropped to 5%, of which almost all were from the Yoruban group. Finally, when we restricted the comparison to only genes associated with ID/DD with a missense z-score of 3.09 or greater, the proportion of significant comparisons dropped to only 0.6%, of which all were from the Yoruban group (**Supplementary Table S19**).

Finally, we closely investigated the regions of the genome we were unable to confidently call in each sample and determined how many genes in the ID/DD database were not well covered (**Supplementary Tables S20** and **S21**). We also examined the average pLI and missense Z-scores for all of the genes in which 10% or more of the exons were not covered (**Supplementary Table S22**). These scores averaged 0.29 and 1.11 respectively, suggesting that regions of low coverage in our genomic data were not enriched for these types of genes.

DISCUSSION

Owing to the profoundly juvenile characteristics of NCS patients, we initially proposed that they harbored mutations in one or more genes controlling the rate of aging.^{1,2} In this study, our clinical and genetic findings help clarify that the neotenic features seen in NCS patients are caused by changes in development and should be differentiated from slowed aging and extended healthy life span, which are caused mostly by reducing damage to cell components and improving tissue maintenance. We did not identify a single genetic cause for NCS, but instead we discovered DNMs in five different genes in five of the six patients whose genomes were analyzed. These genes are highly intolerant of variation in the human genome. Given the small sample size of our study, however, these DNMs cannot be conclusively linked to NCS. We also observed a few small (<15 kb) de novo SVs in the genomes of each NCS patient, but it is unclear whether these SVs disrupt any important genes, as none of them overlapped with coding sequence. We were

Figure 3 Rare variant burden analysis. The average numbers of rare variants were determined for all NCS subjects, their families, and various ethnic groups from the 1,000 Genomes Project (1KG). Rare variant minor allele frequencies (MAFs) of 0.001, 0.005, 0.01, and 0.05 were examined. The averages were calculated for (a) all genes in the genome, (b) every gene with a missense z-score of 3.09 or greater, (c) every gene found in a large intellectual disability and developmental delay database¹² with a missense z-score of 3.09 or greater, and (d) every gene found in a large autism spectrum disorder database¹³ with a missense z-score of 3.09 or greater. Stars over columns denote *P* values less than 0.05, suggesting a statistically significant difference between the average number of variants in the comparison group and those in the neotonic complex syndrome (NCS) samples as determined by a *t*-test. The complete list of comparisons can be found in **Supplementary Table S19**.

unable to identify any inherited small variants or SV/CNVs that appeared to contribute to this syndrome and were shared between our patients. However, we did find that inherited rare and family-specific single nucleotide variants in highly constrained genes in each family and found an excess of these variants (less than or equal to 0.001 MAF) in genes associated with ID/DD and ASD, compared to control samples. This difference was significant from different ethnic groups sequenced with the same technology; but that could be due to the small number of control samples in our study. Random sampling of each ethnic group and performance of the same analysis as was carried out on NCS samples lends support to the notion that these results could be significant, but more NCS samples and controls are needed to confirm this. We also observed a larger number of rare variants of all categories in the NCS patients than in their healthy family members; however, this result was significant in only a few comparisons. Finally, although there may be rare variants that contribute to NCS, we are unable to identify any specific variants as causative.

While each gene with a DNM differed between NCS patients, there were similarities in their functions. Most of the genes were involved in transcription regulation, primarily through histone modification. There are numerous examples in the literature of serious diseases, many with growth and neurological consequences, which are the result of mutations in genes that modulate chromatin structure;²³ and recent large-scale studies of ASD^{24,25} and ID/DD¹² have found similar groups of genes affected by DNMs. Interestingly, DNMs in one of these genes, *DDX3X*, have been found to be responsible for the majority of severe ID/DD^{12,26} in females, and our patient had the same mutation as individual #37 in the study by Snijders Blok *et al.*²⁶ Like our patients, this particular individual has severe intellectual disability, low weight, and microcephaly, but, given the reported phenotypes, it is not clear whether this is an additional case of NCS. The other 37 individuals in that study had a wide range and varying severity of phenotypes, even though each of them was found to harbor a DNM in only one allele of *DDX3X*. From the available phenotypic information we were not able to confirm any additional NCS cases in the study by Snijders Blok, further suggesting that NCS is very rare. A DNM in *TLK2* was also discovered in one of our patients and has recently been described in several individuals with ID/DD,²⁷ ASD,^{24,28} and schizophrenia.²⁹ We also identified a patient with a DNM in *HDAC8*. De novo mutations in this gene were recently described as causing a small proportion of cases of Cornelia de Lange syndrome (CdLS).³⁰ One of our patients was found to have a DNM in *TMEM63B*, and a recent mouse knockout study¹⁴ suggested that mutations in this gene could be a potential source of human disease. In addition, one NCS patient harbored a single DNM in the mitochondrial genome, at a highly conserved position, that is expected to disrupt the anticodon stem of the tryptophan tRNA. This DNM was found previously in an individual with encephalomyopathy.³¹ Because we sequenced only the blood of each patient it is not

clear how prevalent this mutation is in other parts of the body or how much this mutation will contribute to the phenotypes seen in these NCS patients. Finally, within the gene *CCDC101*, exactly the same DNM as in patient #36 was found in the ExAC database in three healthy individuals. Taken together, these findings suggest that these DNMs alone are probably not sufficient to cause NCS.

While NCS shares some phenotypes with the syndromes associated with DNMs in some of the genes identified in our study, the most important, neoteny, has not been described in other syndromes in the same detail as NCS. Our results agree with recent studies in ASD^{24,25} and ID/DD^{12,27} showing that DNMs in the same genes can result in entirely different neurological and developmental phenotypes in different patients. The cause of this phenomenon is unclear, but could be specific mutations in each gene, differential expression of the remaining unaffected allele, each individual's inherited genetic background, environmental factors, somatic mutations in other unanalyzed tissues, or some combination of all of these. Although we did not identify any other pathogenic variants, it is possible that NCS could result from multiple overlapping genetic diseases caused by inherited or de novo variations, such as have been seen in recent studies.³² It is also important to note that, even though we used two different methods to make sequencing libraries for each patient, it is possible we missed inherited or de novo variations that could contribute to NCS. In addition, there are noncoding de novo and inherited variations whose effects are currently beyond our understanding and could also contribute to this syndrome.

In two patients we identified a shared rare haplotype on chromosome X. This haplotype was not found in the genomes of 184 healthy participants of the PGP⁷ and appears to be extremely rare. It is located in a region with many important developmental genes, although we did not identify any obvious coding or regulatory variants within this haplotype. As it was inherited from a healthy parent in each case, and was shared with the healthy sister of one patient, it is clearly not sufficient to cause NCS, but it is a potentially intriguing risk factor.

We have so far identified NCS only in females. Perhaps it is by chance alone, since the number of confirmed individuals in our study is small. Both *DDX3X* and *HDAC8* are found on the X chromosome and have high constraint scores, making it likely that loss, if these mutations are inactivating, of the single copy in males is lethal. Another possibility is that while certain neoteny traits, such as curiosity and plasticity of behavior, are shared equally between the sexes, human females are more pedomorphic in physical appearance than males. Thus, there may be a genetic tendency towards neoteny in females, which is amplified through mutation in these patients.

Finally, syndrome X is the historic name for metabolic syndrome, which is a group of risk factors that predispose to diabetes and heart disease. We propose that the multiple congenital anomalies described herein be called neotenic complex syndrome (NCS). This name is consistent with

recent trends toward describing medical conditions by symptoms rather than eponyms, and neoteny, which is associated with a specific complex of developmental disorders, is the syndrome's most diagnostic criterion.

SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at <http://www.nature.com/gim>

ACKNOWLEDGMENTS

This work was supported in part by the Shenzhen Municipal Government of China (NO.KQTD20150330171505310). B.A.P. is a recipient of and this work was partially supported by the Research Fund for International Young Scientists, National Natural Science Foundation of China (31550110216). We thank all the NCS patients and their families for participating in this study. We acknowledge the ongoing contributions and support of all Complete Genomics employees, in particular the many highly skilled individuals who work in the libraries, reagents, and sequencing groups that make it possible to generate high-quality whole-genome data.

AUTHOR CONTRIBUTIONS

R.F.W., S.D., J.S.L., B.A.P., and R.D. conceived the study. R.Y.Z. and R.C. isolated genomic DNA and generated all of the sequencing libraries. N.B. managed the sequencing process. P.C. and Z.L. developed algorithms for LFR analysis. R.C., M.W., K.S., K.L., and R.Y.Z. confirmed DNMs through Sanger Sequencing. S.C., Q.M., and S.N. processed the data. S.C., Q.M., M.R.A., Z.L., and B.A.P. performed genome analyses. R.F.W., J.S.L., R.D., and B.A.P. coordinated the study. R.F.W., R.D., and B.A.P. wrote the manuscript. All authors reviewed and edited the manuscript.

ACCESSION NUMBERS

All sequencing data reported in this paper have been deposited in the database of the European Genome-phenome Archive (EGA) under accession number EGAS00001002419.

DISCLOSURE

R.F.W. and Z.L. declare no conflict of interest. The other authors, as employees of Complete Genomics, are shareholders in BGI holdings, and Complete Genomics and BGI derive income from whole-genome sequencing.

REFERENCES

1. Walker RF, Pakula LC, Sutcliffe MJ, Kruk PA, Graakjaer J, Shay JW. A case study of "disorganized development" and its possible relevance to genetic determinants of aging. *Mech Ageing Dev.* 2009;130:350–356.
2. Walker RF. Developmental theory of aging revisited: focus on causal and mechanistic links between development and senescence. *Rejuvenation Res.* 2011;14:429–436.
3. Drmanac R, Sparks AB, Callow MJ, et al. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* 2010;327:78–81.
4. Peters BA, Kermanni BG, Sparks AB, et al. Accurate whole-genome sequencing and haplotyping from 10 to 20 human cells. *Nature* 2012;487:190–195.
5. Genomes Project C, Auton A, Brooks LD, et al. A global reference for human genetic variation. *Nature*. 2015;526:68–74.
6. Erikson GA, Bodian DL, Rueda M, et al. Whole-genome sequencing of a healthy aging cohort. *Cell* 2016;165:1002–1011.
7. Mao Q, Ciotlos S, Zhang RY, et al. The whole genome sequences and experimentally phased haplotypes of over 100 personal genomes. *Gigascience*. 2016;5:42.
8. Kong A, Frigge ML, Masson G, et al. Rate of de novo mutations and the importance of father's age to disease risk. *Nature* 2012;488:471–475.
9. Goldmann JM, Wong WS, Pinelli M, et al. Parent-of-origin-specific signatures of de novo mutations. *Nat Genet* 2016;48:935–939.
10. Samocha KE, Robinson EB, Sanders SJ, et al. A framework for the interpretation of de novo mutation in human disease. *Nat Genet* 2014;46:944–950.
11. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016;536:285–291.
12. Deciphering Developmental Disorders Study. Large-scale discovery of novel genetic causes of developmental disorders. *Nature* 2015;519:223–228.
13. Basu SN, Kollu R, Banerjee-Basu S. AutDB: a gene reference resource for autism research. *Nucleic Acids Res.* 2009;37(database issue):D832–836.
14. Dickinson ME, Flenniken AM, Ji X, et al. High-throughput discovery of novel developmental phenotypes. *Nature* 2016.
15. McLean CY, Bristol D, Hiller M, et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* 2010;28:495–501.
16. Boyle AP, Hong EL, Hariharan M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* 2012;22:1790–1797.
17. Kircher M, Witten DM, Jain P, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014;46:310–315.
18. Parisien M, Wang X, Pan T. Diversity of human tRNA genes from the 1000-genomes project. *RNA Biol* 2013;10:1853–1867.
19. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559–575.
20. International HapMap C, Altshuler DM, Gibbs RA, et al. Integrating common and rare genetic variation in diverse human populations. *Nature* 2010;467:52–58.
21. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012;489:57–74.
22. Online Mendelian Inheritance in Man. <http://omim.org/>. Accessed August 11, 2016.
23. Hendrich B, Bickmore W. Human diseases with underlying defects in chromatin structure and modification. *Hum Mol Genet.* 2001;10:2233–2242.
24. Iossifov I, O'Roak BJ, Sanders SJ, et al. The contribution of de novo coding mutations to autism spectrum disorder. *Nature* 2014;515:216–221.
25. De Rubeis S, He X, Goldberg AP, et al. Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* 2014;515:209–215.
26. Snijders Blok L, Madsen E, Juusola J, et al. Mutations in DDX3X are a common cause of unexplained intellectual disability with gender-specific effects on wnt signaling. *Am J Hum Genet.* 2015;97:343–352.
27. Lelieveld SH, Reijnders MR, Pfundt R, et al. Meta-analysis of 2,104 trios provides support for 10 new genes for intellectual disability. *Nat Neurosci* 2016;19:1194–1196.
28. O'Roak BJ, Deriziotis P, Lee C, et al. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nat Genet* 2011;43:585–589.
29. Gulsuner S, Walsh T, Watts AC, et al. Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. *Cell* 2013;154:518–529.
30. Deardorff MA, Bando M, Nakato R, et al. HDAC8 mutations in Cornelia de Lange syndrome affect the cohesin acetylation cycle. *Nature* 2012;489:313–317.
31. Silvestri G, Mongini T, Odoardi F, et al. A new mtDNA mutation associated with a progressive encephalopathy and cytochrome c oxidase deficiency. *Neurology* 2000;54:1693–1696.
32. Posey JE, Harel T, Liu P, et al. Resolution of disease phenotypes resulting from multilocus genomic variation. *N Engl J Med* 2017;376:21–31.