

# Mutations in Epigenetic Regulation Genes Are a Major Cause of Overgrowth with Intellectual Disability

Katrina Tatton-Brown,<sup>1,2</sup> Chey Loveday,<sup>1</sup> Shawn Yost,<sup>1</sup> Matthew Clarke,<sup>1</sup> Emma Ramsay,<sup>1</sup> Anna Zachariou,<sup>1</sup> Anna Elliott,<sup>1</sup> Harriet Wylie,<sup>1</sup> Anna Ardissonne,<sup>3</sup> Olaf Rittinger,<sup>4</sup> Fiona Stewart,<sup>5</sup> I. Karen Temple,<sup>6,7</sup> Trevor Cole,<sup>8</sup> Childhood Overgrowth Collaboration, Shazia Mahamdallie,<sup>1</sup> Sheila Seal,<sup>1</sup> Elise Ruark,<sup>1</sup> and Nazneen Rahman<sup>1,9,10,\*</sup>

To explore the genetic architecture of human overgrowth syndromes and human growth control, we performed experimental and bioinformatic analyses of 710 individuals with overgrowth (height and/or head circumference  $\geq +2$  SD) and intellectual disability (OGID). We identified a causal mutation in 1 of 14 genes in 50% (353/710). This includes *HIST1H1E*, encoding histone H1.4, which has not been associated with a developmental disorder previously. The pathogenic *HIST1H1E* mutations are predicted to result in a product that is less effective in neutralizing negatively charged linker DNA because it has a reduced net charge, and in DNA binding and protein-protein interactions because key residues are truncated. Functional network analyses demonstrated that epigenetic regulation is a prominent biological process dysregulated in individuals with OGID. Mutations in six epigenetic regulation genes—*NSD1*, *EZH2*, *DNMT3A*, *CHD8*, *HIST1H1E*, and *EED*—accounted for 44% of individuals (311/710). There was significant overlap between the 14 genes involved in OGID and 611 genes in regions identified in GWASs to be associated with height ( $p = 6.84 \times 10^{-8}$ ), suggesting that a common variation impacting function of genes involved in OGID influences height at a population level. Increased cellular growth is a hallmark of cancer and there was striking overlap between the genes involved in OGID and 260 somatically mutated cancer driver genes ( $p = 1.75 \times 10^{-14}$ ). However, the mutation spectra of genes involved in OGID and cancer differ, suggesting complex genotype-phenotype relationships. These data reveal insights into the genetic control of human growth and demonstrate that exome sequencing in OGID has a high diagnostic yield.

## Introduction

Human growth control, at the organismal and cellular level, is a complex process essential for health and dysregulated in many developmental disorders and cancers. The mechanistic control of cell size and proliferation has been studied, by diverse approaches, in many different species.<sup>1,2</sup> However, the control of overall size of an organism has been relatively understudied and is still poorly understood. The study of human growth disorders therefore not only improves diagnosis and management of human disease, it also offers an opportunity to enhance knowledge about the fundamental processes governing control of human size.

Human overgrowth syndromes are a nebulous group of conditions defined as having height and/or head circumference  $\geq 2$  SD above the mean, together with additional phenotypic abnormalities, the most common of which is intellectual disability.<sup>3</sup> Overgrowth syndromes usually occur sporadically within a family and can be

caused by several different mechanisms, including gene mutations, imprinting disruption, and chromosome dosage abnormalities.<sup>3,4</sup>

Single-gene disorders associated with overgrowth and intellectual disability (OGID) are well recognized; Sotos syndrome (MIM: 117550) and Weaver syndrome (MIM: 277590) are prototypic examples, due to *NSD1* (MIM: 606681) and *EZH2* (MIM: 601573) mutations, respectively (see GeneReviews by Tatton-Brown et al. in [Web Resources](#)).<sup>5</sup> OGID syndromes have been increasingly identified over the last decade.<sup>3,4</sup> The advent of next-generation sequencing has been the foremost reason for this progress and has allowed elucidation of the genetic causes of clinically established syndromes and the delineation of new syndromes.<sup>5–12</sup>

Despite these advances, many individuals with OGID remain without a genetic diagnosis. In addition, the relative contribution of the different genes to OGID is unknown. To better characterize the genetic landscape of OGID, we have here studied 710 affected individuals including 323 parent-proband trios ([Table S1](#)).

<sup>1</sup>Division of Genetics and Epidemiology, Institute of Cancer Research, 15 Cotswold Road, London SM2 5NG, UK; <sup>2</sup>South West Thames Regional Genetics Service, St George's University Hospitals NHS Foundation Trust, London SW17 0QT, UK; <sup>3</sup>Child Neurology Unit, Fondazione IRCCS C Besta Neurological Institute, Milan 20133, Italy; <sup>4</sup>Landeskrankenanstalten Salzburg, Kinderklinik Department of Pediatrics, Klinische Genetik, Salzburg 5020, Austria; <sup>5</sup>Northern Ireland Regional Genetics Service, Belfast City Hospital, Belfast BT9 7AB, Northern Ireland; <sup>6</sup>Human Development and Health Academic Unit, Faculty of Medicine, University of Southampton, Southampton SO17 1BJ, UK; <sup>7</sup>Wessex Clinical Genetics Service, University Hospital Southampton NHS Trust, Southampton SO16 6YD, UK; <sup>8</sup>West Midlands Regional Genetics Service, Birmingham Women's Hospital NHS Foundation Trust and University of Birmingham, Birmingham Health Partners, Birmingham B15 2TG, UK; <sup>9</sup>Cancer Genetics Unit, Royal Marsden NHS Foundation Trust, London SW3 6JJ, UK

<sup>10</sup>Twitter: @rahman\_nazneen

\*Correspondence: rahmanlab@icr.ac.uk

<http://dx.doi.org/10.1016/j.ajhg.2017.03.010>

© 2017 The Author(s). This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## Subjects and Methods

### Subjects

We recruited participants through the Childhood Overgrowth (COG) Study, which began recruitment in 2005, approved by the London Multicenter Ethics Committee (05/MRE02/17). Informed consent was obtained from all participants and/or parents, as appropriate. Individuals were eligible for this study if they had height and/or head circumference at least two standard deviations above the mean ( $\geq +2$  SD, UK90 growth data)<sup>13</sup> at some point in childhood, together with intellectual disability. We have termed this condition OGID (overgrowth + intellectual disability). Overgrowth phenotypes that are not associated with intellectual disability, such as Beckwith Wiedemann syndrome (MIM: 130650) or Marfan syndrome (MIM: 154700), were not included. Regional or asymmetric overgrowth phenotypes (e.g., hemihypertrophy) in the absence of increased height or head circumference were not included.

710 individuals with OGID were included. 97% (693) were recruited to the study from clinical genetics departments. For 323 individuals, samples from both parents were also available and included. 205 probands had both height and head circumference  $\geq +2$  SD, termed “head+height” in Table S1. 138 had height  $\geq +2$  SD with OFC  $< 2$  SD, termed “height only” and 109 had OFC  $\geq +2$  SD and height  $< 2$  SD, termed “head only.” For the remaining 258 individuals, the child was recruited to the study because they had overgrowth, but measurements for both height and head were not provided. The overgrowth category is termed “unspecified” for these case subjects in Table S1. Intellectual disability was classified by the referring clinician as severe (77 case subjects), moderate (228 case subjects), or mild (229 case subjects). The referrer did not state the severity of the OGID for 176 individuals (termed “unspecified” in Table S1).

### Control Data

We used the Exome Aggregation Consortium (ExAC) data v.3 accessed on 13/11/2015 (excluding the TCGA samples)<sup>14</sup> and the ICR1000 UK exome series<sup>15</sup> as reference data. We generated and analyzed the ICR1000 UK exome series data using the same sequencing and analysis pipeline described for the OGID samples.

### Targeted Gene Analyses

We previously reported mutations in *NSD1*, *EZH2*, *DNMT3A* (MIM: 602769), and *PPP2R5D* (MIM: 601646) in 198 case subjects. The relevant references are in Table S1. Intragenic mutations in these genes were detected with Sanger sequencing. *NSD1* is unusual among the 14 OGID genes included in this study in being prone to deletion by a 2 Mb 5q35 microdeletion, mediated by flanking low-copy repeats.<sup>16</sup> We used MLPA to identify 5q35 microdeletions encompassing *NSD1*.<sup>17</sup> *NSD1* MLPA is

also capable of detecting exon CNVs that account for ~5% of *NSD1* mutations.<sup>17</sup> Microdeletions and exon CNVs in the other genes were not sought, but are unlikely to be a major contributor because the surrounding sequence architecture and/or mechanism of pathogenicity make it much less likely that such events will cause OGID.

### Exome Sequencing

We performed exome sequencing in all probands in whom no mutation had been identified by targeted gene analyses and in parental samples where available. We performed exome sequencing using the Nextera Rapid Capture Exome Kit (Illumina). We prepared libraries from 50 ng genomic DNA using the Nextera DNA Sample Preparation Kit (Illumina). On average 33M reads mapped to the pull-down and 86% of targeted bases had  $\geq 15\times$  coverage. The captured libraries were PCR amplified using the supplied paired-end PCR primers. Exome sequencing in 57 samples was performed before the Nextera Exome Kit was available using the TruSeq Exome Enrichment Kit, which includes the 14 genes involved in OGID. When converting our exome pipeline from TruSeq to Nextera, we undertook in-house evaluation and validation to ensure that the performance was equivalent. Sequencing was performed on an Illumina HiSeq 2000 or HiSeq 2500 (high output mode) using v3 chemistry and generating  $2 \times 101$  bp reads.

### Variant Calling

We used the OpEx v1.0 pipeline to perform variant calling.<sup>18</sup> We converted raw data to FASTQs using CASAVA v.1.8.2 with default settings. The OpEx v1.0 pipeline uses Stampy<sup>19</sup> to map to the human reference genome, Picard to flag duplicates, Platypus<sup>20</sup> to call variants, and CAVA<sup>21</sup> to provide consistent annotation of variants with the HGVS-compliant CSN (Clinical Sequencing Notation) standard v1.0.<sup>21</sup> The transcript information for variant annotation for the 14 relevant genes are given in Table 1.

### Variant Prioritization and Validation

We excluded variants with MAF  $> 0.5\%$  in either the Exome Aggregation Consortium (ExAC) and/or the ICR1000 UK exome series. For the de novo analyses, we identified and validated any high-quality (as defined by OpEx<sup>18</sup>) variant in the child that was not present in either parent. We evaluated and validated all rare variants identified in the 14 genes.

We confirmed all small variants in Table S1 that were called in exomes via Sanger sequencing of M13-tagged PCR products generated from genomic DNA. We performed PCR using the QIAGEN Multiplex PCR Kit according to the manufacturer's instructions. We sequenced PCR products using M13 sequencing primers, the BigDye Terminator Cycle Sequencing Kit, and an ABI 3730 Genetic Analyzer (Applied Biosystems). We analyzed sequences using Mutation Surveyor software v.3.20 (SoftGenetics)

**Table 1. Gene and Transcript Information for 14 Genes Involved in OGID**

Gene	MIM Number	HGNC ID	Ensembl Transcript	RefSeq Transcript
<i>AKT3</i>	611223	HGNC:393	ENST00000366539	NM_005465
<i>BRWD3</i>	300553	HGNC:17342	ENST00000373275	NM_153252
<i>CHD8</i>	610528	HGNC:20153	ENST00000399982	NM_001170629
<i>DNMT3A</i>	602769	HGNC:2978	ENST00000264709	NM_175629
<i>EED</i>	605984	HGNC:3188	ENST00000263360	NM_003797
<i>EZH2</i>	601573	HGNC:3527	ENST00000320356	NM_004456
<i>GPC3</i>	300037	HGNC:4451	ENST00000370818	NM_004484
<i>HIST1H1E</i>	142220	HGNC:4718	ENST00000304218	NM_005321
<i>MTOR</i>	601231	HGNC:3942	ENST00000361445	NM_004958
<i>NFIX</i>	164005	HGNC:7788	ENST00000360105	NM_002501
<i>NSD1</i>	606681	HGNC:14234	ENST00000439151	NM_022455
<i>PIK3CA</i>	171834	HGNC:8975	ENST00000263967	NM_006218
<i>PPP2R5D</i>	601646	HGNC:9312	ENST00000485511	NM_006245
<i>PTEN</i>	601728	HGNC:9588	ENST00000371953	NM_000314

and verified the outputs by manual inspection by two individuals, independently.

### Pathogenic Mutation Determination

Apart from *HIST1H1E* (MIM: 142220), we considered a variant in the other 13 genes to be pathogenic if it fulfilled one or more of the following criteria. (1) It was a de novo mutation in a gene for which such de novo mutations were already proven to cause OGID. (2) The inheritance was unknown, because parental samples were unavailable, but it had been previously identified as a pathogenic de novo mutation in OGID. (3) It was a protein-truncating variant ([PTV] frameshifting indels, stop-gain, or essential splice-site variants) in a gene in which truncating mutations have been proven to be pathogenic. (4) There was clear evidence from the literature that it was pathogenic. The evidence for *HIST1H1E* mutations being pathogenic is provided in the [Results](#).

### *HIST1H1E* Statistical Analyses

We used the methods described in the DDD study<sup>22</sup> to calculate the probability of identifying four de novo frameshift mutations in *HIST1H1E* using the gene-specific mutation rates from Samocha et al.<sup>23</sup> The frameshift mutation rate in *HIST1H1E* ( $4.18 \times 10^{-7}$ ) was multiplied by twice the number of case subjects in this study (710) in order to get the expected number of frameshift mutations. We calculated the probability of observing four or more de novo frameshift mutations in *HIST1H1E* given the expected number of frameshift mutations via the ppois function in R.

We modeled the significance of mutation clustering in *HIST1H1E* under a binomial distribution where the proba-

bility of observing a mutation in a 12 bp region, which comprises 1.8% of the coding sequence, was 0.018.

### Protein Net Charge Calculation

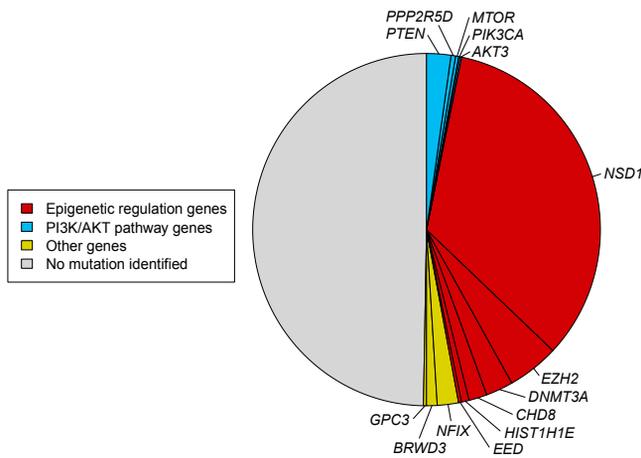
We obtained wild-type *HIST1H1E* cDNA (frame 1) sequence from Ensembl (ENST00000304218.5). We generated the *HIST1H1E* cDNA sequences edited with OGID mutations (frame 2). We used the variant c.430delG to generate the other possible alternative reading frame in *HIST1H1E* (frame 3). We translated the cDNA sequences using the Translate Tool at ExPASy. We calculated the net charge of the carboxy-terminal domain, from p.Lys110 onward, at neutral pH using the Peptide Property Calculator at the Innovagen website.

### Functional Network Analyses

We performed functional enrichment analysis using g:Profiler (v.r1665\_e85\_eg32).<sup>24</sup> We used the 14 genes in [Table 1](#) as our query set. We looked for enrichment among Gene Ontology molecular function terms and KEGG pathway gene sets, requiring the size of the functional category to be between 5 and 500 genes and using the Benjamini-Hochberg false discovery rate as the significance threshold. The FDR q values presented are the Benjamini-Hochberg critical values.

### Phenotypic Analyses

We tested for significant difference in the diagnostic yields between different phenotypic groups using the prop.test function in R. We calculated the significance of association between an individual having macrocephaly and their mutation status (either a mutation in a PI3K/AKT pathway gene or a mutation in an epigenetic regulation gene) using a Fisher's exact test, which we implemented with the



**Figure 1. Causal Mutation Identified in 50% of OGID Probands**  
Proportion of pathogenic mutations identified in 710 individuals with OGID. Epigenetic regulation genes (red), including *NSD1* which is the predominant gene, constitute the major gene set. PI3K/AKT pathway genes (blue) also significantly contribute to OGID.

fisher.test function in R. We calculated the significance of association between an individual having macrocephaly in the absence of increased height and their mutation status, and the significance of association between an individual having increased height in the absence of macrocephaly and their mutation status in the same way. We tested for significant difference in the proportion of individuals with mild intellectual disability for those with a mutation in a PI3K/AKT pathway OGID gene and those with a mutation in an epigenetic regulation OGID gene using the prop.test function in R.

### Height GWAS Gene and Cancer Driver Gene Comparisons

We obtained the list of 611 genes located in regions associated with human height through GWASs from Table S1 of Wood et al.<sup>25</sup> We obtained a list of 260 somatically mutated cancer genes from Table S2 of Lawrence et al.<sup>26</sup> and the somatic mutations from the tumor portal website.

We calculated the probability of seeing the observed overlap of the OGID gene set with the GWAS gene set under a hypergeometric probability distribution assuming a total hypothetical size of 20,000 protein-coding genes in the exome using the phyper function in R. We calculated the probability of seeing the observed overlap of OGID gene set with the cancer driver gene set in the same way.

## Results

### Contribution of Gene Mutations to OGID

Using exome or targeted gene analyses, we identified a pathogenic mutation in one of 14 genes in 357 individuals with OGID, giving a diagnostic yield of 50% (Figure 1).

By far the most common cause was a mutation in *NSD1* (240 cases, 34%), followed by *EZH2* (34, 4.8%), *DNMT3A* (18, 2.5%), *PTEN* (MIM: 601728) (16, 2.3%), *NFIX* (MIM: 164005) (14, 2.0%), *CHD8* (MIM: 610528) (12, 1.7%), *BRWD3* (MIM: 300553) (7, 1.0%), *HIST1H1E* (5, 0.7%), *PPP2R5D* (3, 0.4%), (2 cases each) *EED* (MIM: 605984), *GPC3* (MIM: 300037), and *MTOR* (MIM: 601231), and (1 case each) *AKT3* (MIM: 611223) and *PIK3CA* (MIM: 171834) (Table S1). Among the 323 parent-proband trios, we identified a cause in 191 (59%) of which 179 were de novo mutations and 12 were inherited.

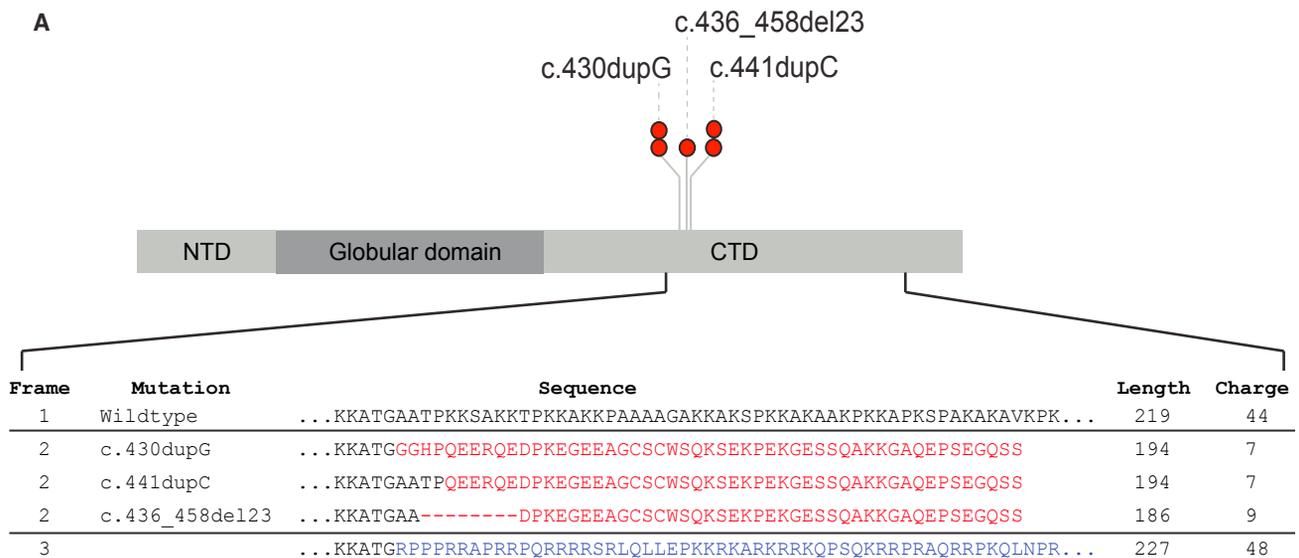
Our data allow confirmation that *EED* mutations cause OGID. Two case reports of individuals with a characteristic phenotype that includes overgrowth have been published.<sup>10,27</sup> We here present two additional cases with a de novo *EED* mutation. The individuals have the same facial phenotype to each other and to previously reported case subjects, with long, narrow palpebral fissures, telecanthus, and retrognathia. Notably, *EED* is a direct binding partner of *EZH2*,<sup>28</sup> which has an established role in causing OGID.<sup>29</sup> Some role in overgrowth was either known, or has been proposed, for the remainder of these, apart from *HIST1H1E* (see GeneReviews by Eng in Web Resources).<sup>6,9,10,12,29–35</sup>

### *HIST1H1E* Mutations Cause OGID

We present here data showing that certain *HIST1H1E* mutations cause OGID. Through exome sequencing we identified five unrelated probands—COG0405, COG0412, COG0552, COG1739, and COG1832—with heterozygous *HIST1H1E* protein truncating variants (PTVs) (Figure 2, Tables 1 and S1). In four probands the PTV had arisen de novo. Parental samples were not available for the fifth child, but she carried the same mutation as one of the children with a de novo mutation. The detection of four de novo *HIST1H1E* mutations in 710 individuals is highly unlikely to have occurred by chance, as determined from gene-specific de novo mutation rates ( $p = 5.17 \times 10^{-15}$ ). None of the mutations are present in the ExAC dataset, nor in 11,677 exomes analyzed in-house with similar pipelines. These results strongly support *HIST1H1E* mutations as a cause of OGID.

*HIST1H1E* encodes histone H1.4. In humans, H1.4 is one of 11 H1 linker histones that mediate the formation of higher-order chromatin structures and regulate the accessibility of regulatory proteins, chromatin remodelling factors, and histone-modifying enzymes to their target sites.<sup>36,37</sup> The five mutations we identified cluster significantly ( $p = 2.0 \times 10^{-9}$ ) to a 12-bp region in the carboxy-terminal domain (CTD) that is involved in chromatin binding and protein-protein interactions (Figure 2A).<sup>36</sup> PTVs in the intronless histones have been shown to evade nonsense-mediated mRNA decay.<sup>38</sup> Thus the OGID-causing mutations are predicted to generate a truncated product.

The CTD of linker histones regulate higher-order chromatin structure through neutralization of negatively



**B COG0405**



1.5 years

13 years

**c COG0412**



1.5 years

15.5 years

**D COG1832**



4 years

**Figure 2. HIST1H1E Mutations Cause OGID**

(A) *HIST1H1E* mutations cluster within 12 bp region in the carboxy-terminal domain (CTD) and have a similar predicted impact on protein function. The three different frameshift mutations generate the same open reading frame (frame 2), which is predicted to reduce the length and net charge (at pH 7) of the CTD compared to the wild-type (frame 1). The other possible alternate reading frame (frame 3) increases the protein length and net charge. Abbreviations: CTD, carboxy-terminal domain; NTD, amino-terminal domain. (B–D) Facial images of three individuals with *HIST1H1E* mutations showing full cheeks and high hairline.

charged linker DNA.<sup>36</sup> The pathogenic *HIST1H1E* mutations all result in the same shift in the reading frame and are predicted to generate similar truncated proteins, with a reduced net charge of 7–9 (compared to 44 for the wild-type protein) (Figure 2A). The mutant protein is thus likely to be less effective in neutralizing negatively charged linker DNA. Moreover, the truncation of the C-terminus likely impedes DNA binding and protein-protein interactions. It is also noteworthy that the other possible alteration in reading frame would reduce neither the net charge nor the length of the protein (Figure 2A). Taken together, these data suggest that specific *HIST1H1E* mutations, restricted in position and type, cause human overgrowth.

#### ***HIST1H1E* Clinical Phenotype**

Individuals with *HIST1H1E* mutations had similar facial appearance in childhood with full cheeks, high hairline, and telecanthus (Figures 2B–2D). Height, head circumfer-

ence, and degree of intellectual disability were variable, as were the additional clinical features. It is currently unclear whether these additional features are *HIST1H1E* associations or coincidental findings. Individual case descriptions are below.

COG0405, a female individual, was born at term with a weight of 3.58 kg (+0.1 SD) and a length of 53 cm (+1.5 SD). She was floppy in the neonatal period. A brain MRI scan at 4 months demonstrated mild ventricular dilatation but no other abnormalities. Her bone age at chronological age of 7 months was advanced to 18–24 months. By 19 months, her length was 87 cm (+2.0 SD) with a weight of 13.4 kg (+1.8 SD) and she had developed a strabismus. At 13 years of age, the individual was noted to have normal growth with a height of 150.8 cm (–0.6 SD), a head circumference of 55.8 cm (–0.5 SD), and a weight of 48.85 kg (+0.4 SD). She has developed a severe kyphoscoliosis for which she required surgery and has a mild intellectual disability.

COG0412, a male individual, was born at 1 week after term following an uncomplicated pregnancy and delivery. He weighed 4.75 kg (+2.4 SD). In the neonatal period he was noted to be floppy; he had poor feeding and undescended testes. At 1.5 years he was very tall at 105 cm (+8.3 SD) with a weight of 18.8 kg (+4.6 SD) and a head circumference of 52.5 cm (+2.6 SD). He was reported to have multiple nevi and redundant skin on the palms of his hands. He had a moderate intellectual disability and no behavioral issues at that time. When he was reviewed at 15.5 years, he was no longer tall with a height of 166.5 cm (−0.6 SD). His head circumference was 58.7 cm (+1.4 SD). By this age he had developed an anxiety disorder that was refractory to medical treatment. He had also developed phobias. In addition, he had major dental problems with crumbling teeth and he had dry, flaky nails.

COG0552, a female individual, was born at term with a weight of 4.79 kg (+2.5 SD) and length of 57 cm (+3.6 SD). She was floppy in the neonatal period with poor feeding. She developed no new medical problems in childhood. At the age of 4.2 years she was reported to be delayed in her development. She had a height of 108 cm (+1.2 SD), head circumference of 55 cm (+3.2 SD), and weight of 24 kg (+2.7 SD).

COG1739, a female individual, was initially thought clinically to have Weaver syndrome. She was born at 37 weeks after an uncomplicated pregnancy and labor with a weight of 3.25 kg (+0.8 SD), length of 49 cm (+0.7 SD), and head circumference of 37 cm (+3.3 SD). She was hypoglycemic and hypertonic in the neonatal period, and was also noted to have camptodactyly. At 1.9 years she was diagnosed with a moderate intellectual disability and had a height of 85 cm (mean), head circumference of 51 cm (+1.8 SD), and weight of 12 kg (−0.3 SD).

COG1832, a male individual, was born at 1 week after term weighing 3.74 kg (+0.4 SD). The pregnancy had been complicated by exposure to chicken pox. At birth, COG1832 was noted to have talipes equinovarus and later in the neonatal period was diagnosed with delayed visual maturation. A brain MRI scan showed a slender corpus callosum and unusual ventricular outline, possibly indicative of a periventricular leukomalacia. At 8.5 years, height was 133.2 cm (+0.5 SD) with a weight of 33 kg (+1.2 SD). The head circumference at 6.3 years was 59 cm (+3.7 SD). He has limited speech but with verbal comprehension markedly ahead of this ability to express himself. He has left amblyopia and astigmatism. His hearing is normal. He suffers from constipation. At times his behavior is challenging.

### Functional Network Analyses

To investigate the biological processes abrogated by OGID pathogenic mutations, we performed functional enrichment analysis using the GO molecular function terms and KEGG pathway gene sets in g:Profiler.<sup>24</sup> The chromatin binding (FDR  $q$  value =  $1.58 \times 10^{-6}$ ) and PI3K/AKT signaling

pathway (FDR  $q$  value =  $6.80 \times 10^{-5}$ ) gene sets were significantly enriched.

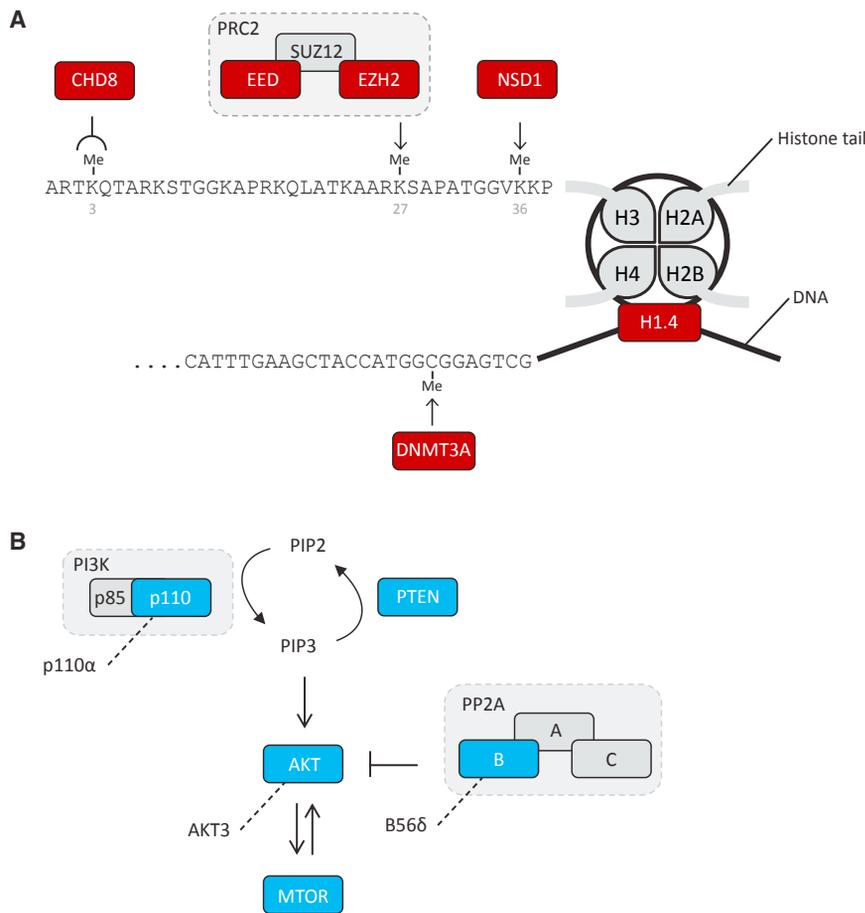
Six genes—*NSD1*, *EZH2*, *DNMT3A*, *EED*, *CHD8*, and *HIST1H1E*—were in the chromatin binding gene set. All encode proteins involved in epigenetic regulation (Figure 3A). *NSD1* is a histone methyltransferase that catalyzes methylation of H3K36, and to lesser extent H4K20, and is primarily associated with transcriptional activation.<sup>39</sup> *EZH2* and *EED* are key components of the polycomb repressive complex 2 (PRC2), which catalyzes methylation of H3K27, resulting in transcriptional repression of target genes.<sup>28</sup> *DNMT3A* is a DNA methyltransferase crucial for the establishment of new methylation marks during early embryogenesis and the sex-dependent methylation of imprinted genes.<sup>40,41</sup> *CHD8* encodes an ATP-dependent chromatin remodeler that binds to methylated H3K4, a key histone modification at active promoters.<sup>35</sup> As noted above, H1.4 binds to linker DNA between nucleosomes and has key roles in chromatin compaction and regulation of gene expression.<sup>37</sup> Together, mutations in these six genes accounted for 311 (44%) of our series. Disruption of epigenetic regulation is therefore a prominent molecular mechanism underlying OGID (Figure 1).

Five of the genes—*PTEN*, *AKT3*, *PIK3CA* (which encodes p110 $\alpha$ , the catalytic domain of the heterodimeric PI3K lipid kinase), *MTOR*, and *PPP2R5D* (which encodes B56 $\delta$  a regulatory subunit of the heterotrimeric PP2A protein phosphatase)—are in the PI3K/AKT pathway, which plays a key role in the regulation of growth (Figure 3B). Activation of the PI3K/AKT pathway results in cellular growth promotion through increased cell metabolism, cell survival, cell turnover, and protein synthesis.<sup>42</sup> Together mutations in these genes made only a minor contribution to our OGID series (23 case subjects, 3.2%). In part this is because individuals with mutations in these genes are more often diagnosed with other conditions, such as Cowden syndrome (MIM: 158350), megalencephaly-capillary malformation syndrome (MIM: 602501), or regional overgrowth (see GeneReviews by Eng in Web Resources).<sup>34</sup>

The remaining three genes—*NFIX*, *GPC3*, and *BRWD3*—encode a transcription factor, a proteoglycan, and a bromodomain-containing protein, respectively<sup>6,31,32</sup> (23 case subjects, 3.2%). There is currently no clear functional link between these genes and the other genes we report here. However, it is possible that *BRWD3* mutations also cause overgrowth through epigenetic regulation dysfunction, as there are data suggesting it is involved in histone H3.3 regulation.<sup>43</sup>

### Phenotype Analyses

There was enrichment of mutations in individuals with both increased height and head circumference, compared to individuals in whom only one growth parameter was increased, as would be expected. Specifically the diagnostic yield in individuals with both macrocephaly and increased height was 59% (120/205), significantly higher than the



### Figure 3. Schematic of Key Biological Processes Impacted in OGID

(A) Epigenetic regulation. NSD1, EED, and EZH2 directly methylate specific histone tail lysine residues. DNMT3A is a de novo DNA methyltransferase and CHD8 is a chromatin remodeling complex protein that binds methylated lysine 4 of histone H3. H1.4 (encoded by *HIST1H1E*) stabilizes higher-order chromatin structures. (B) All OGID mutations are predicted to lead to reduced function PI3K/AKT pathway. The PI3K/AKT pathway positively regulates growth. AKT3, MTOR, and p110 $\alpha$  (encoded by *PIK3CA*) are pathway activators. PTEN and B56 $\delta$  (encoded by *PPP2R5D*) are pathway suppressors. OGID mutations in *AKT3*, *MTOR*, and *PIK3CA* are activating, whereas OGID mutations in *PTEN* and *PPP2R5D* are inactivating.

diagnostic yields in individuals with only macrocephaly (43%, 47/109,  $p = 0.006$ ) or only increased height (45%, 62/138,  $p = 0.009$ ). There was no significant difference between the diagnostic yields in individuals with only macrocephaly and in those with only increased height ( $p = 0.146$ ). There was also no significant difference between the diagnostic yield in individuals with unspecified growth parameters (50%, 130/258) and any other group.

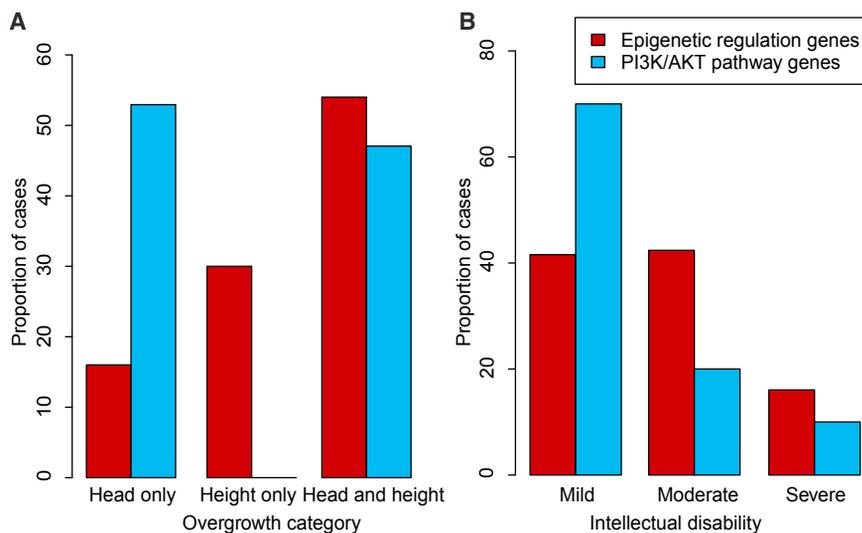
To further explore the phenotypic spectrum of OGID, we compared the growth and intellectual disability severity of the individuals due to mutations in the epigenetic regulation genes and those involved in the PI3K/AKT pathway, using case subjects for which the relevant phenotypic information was available (217 individuals with complete growth data and 263 individuals with intellectual disability severity information) (Figure 4). Macrocephaly (i.e., head circumference  $\geq 2$  SD above the mean) occurred more frequently in individuals with PI3K/AKT pathway gene mutations; all 17 had macrocephaly, compared with 140/200 individuals with OGID due to epigenetic regulation gene mutations ( $p = 4.1 \times 10^{-3}$ ; Figure 4A). Furthermore, 9/17 of the PI3K/AKT pathway case subjects had macrocephaly without increased height compared with 32/200 of the epigenetic regulation pathway cases ( $p = 1.0 \times 10^{-3}$ ; Figure 4A). The remaining 60/200 had increased height without macrocephaly, a combination not present in OGID due to PI3K/AKT pathway gene mutations ( $p = 4.1 \times 10^{-3}$ ; Figure 4A).

Varying severity of intellectual disability was a feature of both groups, but mild intellectual disability was more common in OGID due to PI3K/AKT pathway gene mutations (14/20) than OGID due to epigenetic regulation gene mutations (101/243;  $p = 0.01$ ) (Figure 4B).

The risk of childhood cancer is one of the most controversial areas of OGID management. 8/710 OGID-affected individuals in this study developed cancer in childhood (Table S1). This includes 4/357 with an identified genetic cause, three of whom had an *EZH2* mutation. COG1724 developed neuroblastoma at 46 months, COG0285 developed T cell non-hodgkins lymphoma at 13 years, and COG1521 was diagnosed with both neuroblastoma and acute lymphoblastic leukemia at 13 months. The childhood cancer incidence for *EZH2* mutation carriers in this study was thus 9% (3/34). The remaining child had an *NSD1* microdeletion and T cell non-hodgkins lymphoma. This information will be useful in family discussions about childhood cancer risk, particularly in relation to surveillance strategies, which are generally of unproven benefit and can be associated with appreciable false positive rates.<sup>44</sup>

### Height GWAS Loci Comparative Analyses

We next explored the overlap between the 14 genes and 611 genes implicated through genome-wide association studies (GWASs) to be involved in the control of human height.<sup>25</sup> There was significant overlap; six genes involved in OGID were also present in height GWAS regions ( $p = 6.8 \times 10^{-8}$ ) (Figure S1). The overlap is primarily through the epigenetic regulation genes, all of which (except *EED*) were represented in height GWAS regions. Two separate intronic SNPs in each of *NSD1* and *DNMT3A* were independently associated with height in the GWAS and



**Figure 4. Phenotypic Differences between OGID due to Mutations in Epigenetic Regulation Genes Compared to PI3K/AKT Pathway Genes**

Comparison of the distribution of (A) overgrowth categories and (B) degree of intellectual disability in case subjects with epigenetic regulation gene mutations (red) compared with PI3K/AKT pathway gene mutations (blue).

there were no other genes within the linkage disequilibrium (LD) blocks of association. This strongly suggests that *NSD1* and *DNMT3A* functional impact underlie the height association in these regions (Figure S1). Single SNPs in intron 5 of *CHD8*, intron 9 of *MTOR*, 1 kb downstream of *HIST1H1E*, and 48 kb upstream of *EZH2* were also associated with height.<sup>25</sup> For *HIST1H1E* and *EZH2*, there were no other genes in the LD block of association. For *MTOR* the variant associated with a *cis*-eQTL affecting *MTOR* expression, though the association was better accounted for by an upstream variant (rs2295080) in the *MTOR* promoter region that was in LD with the height SNP (LD  $r^2 = 0.85$ ).<sup>25</sup> Although the causal SNPs and mechanisms of association are not fully elucidated, these data suggest that common variation in some genes involved in OGID also influence height at a population level.

#### Cancer Somatic Driver Mutation Comparative Analyses

Dysregulated cellular growth is a hallmark of cancer, and certain human conditions are associated with both overgrowth and increased cancer risk (see GeneReviews by Eng in Web Resources).<sup>45</sup> We therefore next sought to investigate the overlap between the 14 genes and 260 somatically mutated cancer driver genes reported by Lawrence et al.<sup>26</sup> There was significant overlap; 8/14 genes involved in OGID were somatically mutated in a diverse range of cancers (*NSD1*, *EZH2*, *DNMT3A*, *PTEN*, *CHD8*, *HIST1H1E*, *MTOR*, *PIK3CA*;  $p = 1.7 \times 10^{-14}$ ). For the PI3K/AKT pathway genes, the mutation spectra are similar in OGID and cancer.<sup>34</sup> By contrast, for the epigenetic regulation genes, the mutation spectra in OGID and cancer have substantial, distinctive differences.

Somatic mutations in *HIST1H1E*, *EZH2*, and *DNMT3A* occur in hematological malignancies.<sup>26,46–50</sup> *HIST1H1E* and *EZH2* mutations are each present in ~20% of B cell lymphomas.<sup>48,49</sup> Somatic *HIST1H1E* mutations are nonsynonymous mutations throughout the gene and do not include the clustered PTVs that cause OGID (Figure 5).

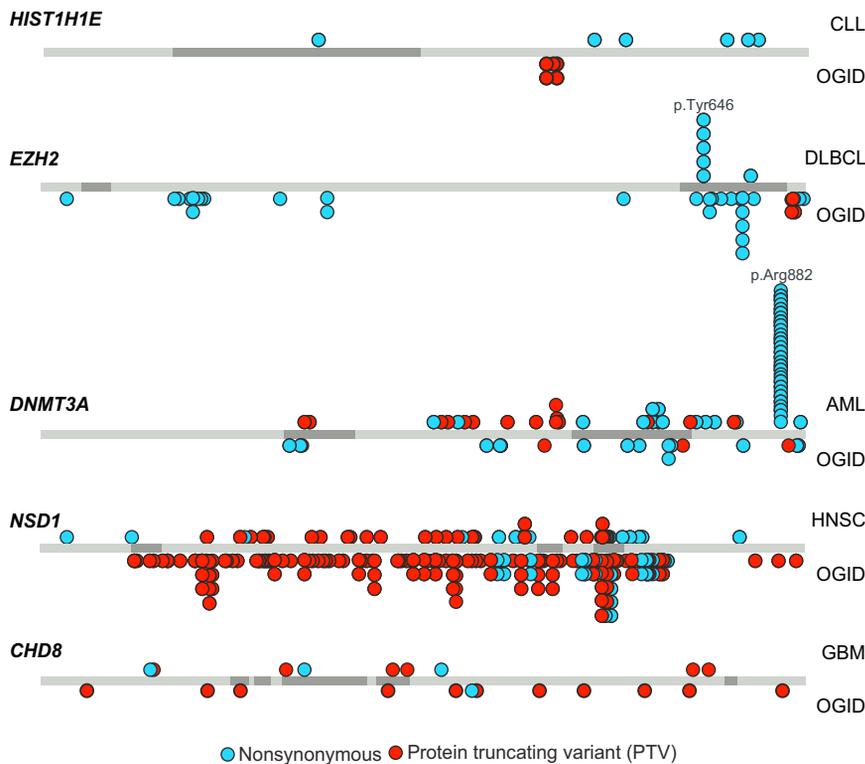
gesting that germline *EZH2* mutations altering p.Tyr646 are not compatible with life (Figure 5). Inactivating *EZH2* mutations are present in myeloid malignancies and in T-ALL.<sup>46–48</sup> A proportion of these latter mutations overlap with *EZH2* mutations in OGID.

*DNMT3A* is one of the most frequently mutated genes in AML and mutations also occur less frequently in other hematological malignancies.<sup>26,50</sup> The majority target a single residue, p.Arg882, with the remainder being nonsynonymous variants and PTVs scattered through the gene. Mutations at p.Arg882 have not thus far been reported in OGID (Figure 5). Protein modeling suggests that the somatic mutations primarily impact DNA binding, whereas the mutations in OGID are more likely to impact histone binding.<sup>12</sup>

Somatic *NSD1* mutations are seen in ~10% of head and neck squamous cell carcinomas<sup>26,51</sup> and somatic *CHD8* mutations are present in ~3% of glioblastoma multiforme (GBM).<sup>26</sup> For these cancers the mutation pattern is similar to that observed in OGID, with PTVs being the most frequent mutation type (Figure 5).<sup>30</sup> Interestingly, Lawrence et al. found *NSD1* and *CHD8* to each be significant in their pan-cancer analysis, present in 2% of cancers.<sup>26</sup> However, the pan-cancer mutation spectra for each gene was different to that observed in OGID, with most being nonsynonymous mutations scattered throughout the gene (Figure 5).

#### Discussion

We present here the largest genetic study of overgrowth and intellectual disability performed to date, including 710 affected individuals and 636 parents. We show that OGID is a highly heterogeneous condition, involving at least 14 genes. Perturbation of epigenetic regulation is a prominent mechanism causing OGID and can be caused by mutations in at least six different genes. *NSD1* mutation is by far the most frequent cause of OGID, accounting for



**Figure 5. Mutations in Epigenetic Regulation Genes in OGID and Cancers**

Protein schematics showing the position of mutations in *HIST1H1E*, *EZH2*, *DNMT3A*, *NSD1*, and *CHD8* in OGID (below the gene) and specific cancers (above the gene). The somatic cancer driver mutations are from Lawrence et al.<sup>26</sup> Abbreviations are as follows: AML, acute myeloid leukemia; CLL, chronic lymphocytic leukemia; DLBCL, diffuse large B-cell lymphoma; GBM, glioblastoma multiforme; HNSC, head and neck squamous cell carcinoma; OGID, overgrowth-intellectual disability.

advance functional and mechanistic understanding of the genes.

One of the most striking results of this study is the high diagnostic yield of genetic testing in OGID; a genetic cause was identified in 50% (357/710) of case subjects. This is likely to be an underestimate as we have been conservative in attributing pathogenicity to OGID gene variants and additional OGID genes almost

240 (34%) of our series. Notably, *NSD1* is within a 2 Mb region flanked by low-copy repeats that mediate a microdeletion, which is one of the commonest causes of Sotos syndrome<sup>16</sup> and was present in 29 individuals. Furthermore, exon deletions or duplications (exon CNVs) are reported in ~5% of case subjects<sup>17</sup> and were present in 9 individuals. We analyzed *NSD1* for these types of mutations, using MLPA, as they are not robustly identifiable in our exome data. We did not examine the other genes for microdeletions or exon CNVs. However, they are not known to be a major contributor to pathogenic mutations in the other genes. Even after excluding microdeletions and exon CNVs, *NSD1* is still the most common cause of OGID, accounting for 202 (28%) of our series.

The comparative analyses of genes involved in OGID with GWAS height loci and with cancer driver genes highlight intriguing similarities and differences. Our data strongly suggest that common variation impacting epigenetic regulation of gene function influences height at a population level. Further investigation of these GWAS loci would be of considerable interest, particularly in relation to advancing knowledge on how, and why, epigenetic regulation dysfunction impacts human growth.

Several genes involved in OGID are somatically mutated in a diverse range of cancers, but the spectra of mutations, particularly in the epigenetic regulation genes, is different in OGID and cancer. The underlying reasons for these differences will be complex and may include embryonic lethality of certain oncogenic mutations when they occur in the germline. Integration of germline and somatic mutational data in future research will be useful, and will likely

certainly exist. Indeed, among the 132 trios in whom a definitive cause was not found, a de novo mutation possibly associated with their phenotype was present in 28; for example, two had de novo nonsynonymous variants in *XRN1*.

The diagnostic yield in our OGID series is higher than exome-sequencing studies in other phenotypes that include intellectual disability, which ranged from 13% to 35%.<sup>22,52–56</sup> The studies are not directly comparable, as most other exome studies included case subjects in which prior genetic testing was negative. Our study recruitment started prior to the discovery and clinical testing of most of the genes we report here, which allows us to provide a much better estimate of the overall contribution of rare gene mutations to this phenotype.

Given the high success rate, strong consideration should be given to using exome sequencing as a first-line diagnostic test in OGID. Height and head circumference can be easily measured and intellectual disability is readily diagnosable. Therefore, implementation of exome sequencing in OGID should be straightforward. Gene testing would provide important diagnostic and recurrence risk information to many families. Furthermore, it would increase genotype-phenotype data, which are urgently required to improve prognostic information. Of equal importance, exome sequencing in OGID would lead to the identification of new genes and new mutations in known genes. In turn, this will stimulate and facilitate scientific research, enhancing knowledge of basic biological processes controlling growth and the diverse pathologies in which human growth control is dysfunctional.

## Supplemental Data

Supplemental Data include one figure, one table, and Supplemental Note and can be found with this article online at <http://dx.doi.org/10.1016/j.ajhg.2017.03.010>.

## Acknowledgments

We thank the families for their participation and the clinicians that recruited them. The full list of collaborators is in the Supplemental Note. We are grateful to Margaret Warren-Perry for assistance in recruitment. We are grateful to Sandra Hanks, Silvana Powell, Imran Uddin, and Ann Strydom for technical and administrative support and Tara Mills for assistance with the GWAS analyses. We acknowledge support from the NIHR RM/ICR Biomedical Research Centre and Wessex NIHR clinical research network. K.T.-B. is supported by funding from the Child Growth Foundation (GR01/13). This work was supported by Wellcome Trust Award 100210/Z/12/Z.

Received: December 7, 2016

Accepted: March 24, 2017

Published: April 27, 2017

## Web Resources

ExAC Browser, <http://exac.broadinstitute.org/>  
ExPASy Translate Tool, <http://web.expasy.org/translate/>  
g:Profiler, <http://biit.cs.ut.ee/gprofiler/>  
GenBank, <http://www.ncbi.nlm.nih.gov/genbank/>  
GeneReviews, Eng, C. (1993). PTEN hamartoma tumor syndrome. <https://www.ncbi.nlm.nih.gov/books/NBK1488/>  
GeneReviews, Tatton-Brown, K., Cole, T.R.P., and Rahman, N. (1993). Sotos syndrome. <https://www.ncbi.nlm.nih.gov/books/NBK1479/>  
ICR1000 UK Exome Series, <http://www.icr.ac.uk/icr1000exomes>  
OMIM, <http://www.omim.org/>  
OpEx NGS Pipeline, <http://www.icr.ac.uk/opex>  
Picard, <http://broadinstitute.github.io/picard/>  
Protein calculator, <http://pepcalc.com/protein-calculator.php>  
TumorPortal, <http://www.tumorportal.org/>

## References

1. Stocker, H., and Hafen, E. (2000). Genetic control of cell size. *Curr. Opin. Genet. Dev.* *10*, 529–535.
2. Saucedo, L.J., and Edgar, B.A. (2002). Why size matters: altering cell size. *Curr. Opin. Genet. Dev.* *12*, 565–571.
3. Tatton-Brown, K., and Weksberg, R. (2013). Molecular mechanisms of childhood overgrowth. *Am. J. Med. Genet. C. Semin. Med. Genet.* *163C*, 71–75.
4. Edmondson, A.C., and Kalish, J.M. (2015). Overgrowth syndromes. *J. Pediatr. Genet.* *4*, 136–143.
5. Tatton-Brown, K., Murray, A., Hanks, S., Douglas, J., Armstrong, R., Banka, S., Bird, L.M., Clericuzio, C.L., Cormier-Daire, V., Cushing, T., et al.; Childhood Overgrowth Consortium (2013). Weaver syndrome and EZH2 mutations: Clarifying the clinical phenotype. *Am. J. Med. Genet. A.* *161A*, 2972–2980.
6. Malan, V., Rajan, D., Thomas, S., Shaw, A.C., Louis Dit Picard, H., Layet, V., Till, M., van Haeringen, A., Mortier, G., Nampoothiri, S., et al. (2010). Distinct effects of allelic NF1X mutations on nonsense-mediated mRNA decay engender either a Sotos-like or a Marshall-Smith syndrome. *Am. J. Hum. Genet.* *87*, 189–198.
7. Gibson, W.T., Hood, R.L., Zhan, S.H., Bulman, D.E., Fejes, A.P., Moore, R., Mungall, A.J., Eydoux, P., Babul-Hirji, R., An, J., et al.; FORGE Canada Consortium (2012). Mutations in EZH2 cause Weaver syndrome. *Am. J. Hum. Genet.* *90*, 110–118.
8. Cordeddu, V., Redeker, B., Stellacci, E., Jongejan, A., Fragale, A., Bradley, T.E., Anselmi, M., Ciolfi, A., Cecchetti, S., Muto, V., et al. (2014). Mutations in ZBTB20 cause Primrose syndrome. *Nat. Genet.* *46*, 815–817.
9. Loveday, C., Tatton-Brown, K., Clarke, M., Westwood, I., Renwick, A., Ramsay, E., Nemeth, A., Campbell, J., Joss, S., Gardner, M., et al.; Childhood Overgrowth Collaboration (2015). Mutations in the PP2A regulatory subunit B family genes PPP2R5B, PPP2R5C and PPP2R5D cause human overgrowth. *Hum. Mol. Genet.* *24*, 4775–4779.
10. Cohen, A.S., and Gibson, W.T. (2016). EED-associated overgrowth in a second male patient. *J. Hum. Genet.* *61*, 831–834.
11. Baynam, G., Overkov, A., Davis, M., Mina, K., Schofield, L., Allcock, R., Laing, N., Cook, M., Dawkins, H., and Goldblatt, J. (2015). A germline MTOR mutation in Aboriginal Australian siblings with intellectual disability, dysmorphism, macrocephaly, and small thoraces. *Am. J. Med. Genet. A.* *167*, 1659–1667.
12. Tatton-Brown, K., Seal, S., Ruark, E., Harmer, J., Ramsay, E., Del Vecchio Duarte, S., Zachariou, A., Hanks, S., O'Brien, E., Akglaede, L., et al.; Childhood Overgrowth Consortium (2014). Mutations in the DNA methyltransferase gene DNMT3A cause an overgrowth syndrome with intellectual disability. *Nat. Genet.* *46*, 385–388.
13. Freeman, J.V., Cole, T.J., Chinn, S., Jones, P.R., White, E.M., and Preece, M.A. (1995). Cross sectional stature and weight reference curves for the UK, 1990. *Arch. Dis. Child.* *73*, 17–24.
14. Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al.; Exome Aggregation Consortium (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature* *536*, 285–291.
15. Ruark, E., Münz, M., Renwick, A., Clarke, M., Ramsay, E., Hanks, S., Mahamdallie, S., Elliott, A., Seal, S., Strydom, A., et al. (2015). The ICR1000 UK exome series: a resource of gene variation in an outbred population. *F1000Res.* *4*, 883.
16. Kurotaki, N., Stankiewicz, P., Wakui, K., Niikawa, N., and Lupski, J.R. (2005). Sotos syndrome common deletion is mediated by directly oriented subunits within inverted Sos-REP low-copy repeats. *Hum. Mol. Genet.* *14*, 535–542.
17. Douglas, J., Tatton-Brown, K., Coleman, K., Guerrero, S., Berg, J., Cole, T.R., Fitzpatrick, D., Gillerot, Y., Hughes, H.E., Pilz, D., et al. (2005). Partial NSD1 deletions cause 5% of Sotos syndrome and are readily identifiable by multiplex ligation dependent probe amplification. *J. Med. Genet.* *42*, e56.
18. Ruark, E., Münz, M., Clarke, M., Renwick, A., Ramsay, E., Elliott, A., Seal, S., Lunter, G., and Rahman, N. (2016). OpEx - a validated, automated pipeline optimised for clinical exome sequence analysis. *Sci. Rep.* *6*, 31029.
19. Lunter, G., and Goodson, M. (2011). Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res.* *21*, 936–939.
20. Rimmer, A., Phan, H., Mathieson, I., Iqbal, Z., Twigg, S.R., Wilkie, A.O., McVean, G., Lunter, G.; and WGS500

- Consortium (2014). Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nat. Genet.* *46*, 912–918.
21. Münz, M., Ruark, E., Renwick, A., Ramsay, E., Clarke, M., Mahamdallie, S., Cloke, V., Seal, S., Strydom, A., Lunter, G., and Rahman, N. (2015). CSN and CAVA: variant annotation tools for rapid, robust next-generation sequencing analysis in the clinical setting. *Genome Med.* *7*, 76.
  22. Deciphering Developmental Disorders Study (2015). Large-scale discovery of novel genetic causes of developmental disorders. *Nature* *519*, 223–228.
  23. Samocha, K.E., Robinson, E.B., Sanders, S.J., Stevens, C., Sabo, A., McGrath, L.M., Kosmicki, J.A., Rehnström, K., Mallick, S., Kirby, A., et al. (2014). A framework for the interpretation of de novo mutation in human disease. *Nat. Genet.* *46*, 944–950.
  24. Reimand, J., Arak, T., Adler, P., Kolberg, L., Reisberg, S., Peterson, H., and Vilo, J. (2016). g:Profiler—a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Res.* *44* (W1), W83–W89.
  25. Wood, A.R., Esko, T., Yang, J., Vedantam, S., Pers, T.H., Gustafsson, S., Chu, A.Y., Estrada, K., Luan, J., Kutalik, Z., et al.; Electronic Medical Records and Genomics (eMEMERGE) Consortium; MIGen Consortium; PAGEGE Consortium; and LifeLines Cohort Study (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* *46*, 1173–1186.
  26. Lawrence, M.S., Stojanov, P., Mermel, C.H., Robinson, J.T., Garraway, L.A., Golub, T.R., Meyerson, M., Gabriel, S.B., Lander, E.S., and Getz, G. (2014). Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* *505*, 495–501.
  27. Cohen, A.S., Tuysuz, B., Shen, Y., Bhalla, S.K., Jones, S.J., and Gibson, W.T. (2015). A novel mutation in EED associated with overgrowth. *J. Hum. Genet.* *60*, 339–342.
  28. Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S., and Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* *298*, 1039–1043.
  29. Tatton-Brown, K., Hanks, S., Ruark, E., Zachariou, A., Duarte, S., Ramsay, E., Snape, K., Murray, A., Perdeaux, E.R., Seal, S., et al.; Childhood Overgrowth Collaboration (2011). Germline mutations in the oncogene EZH2 cause Weaver syndrome and increased human height. *Oncotarget* *2*, 1127–1133.
  30. Tatton-Brown, K., Douglas, J., Coleman, K., Baujat, G., Cole, T.R., Das, S., Horn, D., Hughes, H.E., Temple, I.K., Faravelli, F., et al.; Childhood Overgrowth Collaboration (2005). Genotype-phenotype associations in Sotos syndrome: an analysis of 266 individuals with NSD1 aberrations. *Am. J. Hum. Genet.* *77*, 193–204.
  31. Field, M., Tarpey, P.S., Smith, R., Edkins, S., O'Meara, S., Stevens, C., Tofts, C., Teague, J., Butler, A., Dicks, E., et al. (2007). Mutations in the BRWD3 gene cause X-linked mental retardation associated with macrocephaly. *Am. J. Hum. Genet.* *81*, 367–374.
  32. Cottreau, E., Mortemousque, I., Moizard, M.P., Bürglen, L., Lacombe, D., Gilbert-Dussardier, B., Sigaudy, S., Boute, O., David, A., Faivre, L., et al. (2013). Phenotypic spectrum of Simpson-Golabi-Behmel syndrome in a series of 42 cases with a mutation in GPC3 and review of the literature. *Am. J. Med. Genet. C. Semin. Med. Genet.* *163C*, 92–105.
  33. Saxena, A., and Sampson, J.R. (2014). Phenotypes associated with inherited and developmental somatic mutations in genes encoding mTOR pathway components. *Semin. Cell Dev. Biol.* *36*, 140–146.
  34. Mirzaa, G., Timms, A.E., Conti, V., Boyle, E.A., Girisha, K.M., Martin, B., Kircher, M., Olds, C., Juusola, J., Collins, S., et al. (2016). PIK3CA-associated developmental disorders exhibit distinct classes of mutations with variable expression and tissue distribution. *JCI insight* *1*.
  35. Barnard, R.A., Pomaville, M.B., and O'Roak, B.J. (2015). Mutations and modeling of the chromatin Remodeler CHD8 define an emerging autism etiology. *Front. Neurosci.* *9*, 477.
  36. Harshman, S.W., Young, N.L., Parthun, M.R., and Freitas, M.A. (2013). H1 histones: current perspectives and challenges. *Nucleic Acids Res.* *41*, 9593–9609.
  37. Kalashnikova, A.A., Rogge, R.A., and Hansen, J.C. (2016). Linker histone H1 and protein-protein interactions. *Biochim. Biophys. Acta* *1859*, 455–461.
  38. Maquat, L.E., and Li, X. (2001). Mammalian heat shock p70 and histone H4 transcripts, which derive from naturally intronless genes, are immune to nonsense-mediated decay. *RNA* *7*, 445–456.
  39. Qiao, Q., Li, Y., Chen, Z., Wang, M., Reinberg, D., and Xu, R.M. (2011). The structure of NSD1 reveals an autoregulatory mechanism underlying histone H3K36 methylation. *J. Biol. Chem.* *286*, 8361–8368.
  40. Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* *99*, 247–257.
  41. Kaneda, M., Okano, M., Hata, K., Sado, T., Tsujimoto, N., Li, E., and Sasaki, H. (2004). Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature* *429*, 900–903.
  42. Engelman, J.A., Luo, J., and Cantley, L.C. (2006). The evolution of phosphatidylinositol 3-kinases as regulators of growth and metabolism. *Nat. Rev. Genet.* *7*, 606–619.
  43. Chen, W.Y., Shih, H.T., Liu, K.Y., Shih, Z.S., Chen, L.K., Tsai, T.H., Chen, M.J., Liu, H., Tan, B.C., Chen, C.Y., et al. (2015). Intellectual disability-associated dBRWD3 regulates gene expression through inhibition of HIRA/YEM-mediated chromatin deposition of histone H3.3. *EMBO Rep.* *16*, 528–538.
  44. Katanoda, K. (2016). Neuroblastoma mass screening—what can we learn from it? *J. Epidemiol.* *26*, 163–165.
  45. Lapunzina, P. (2005). Risk of tumorigenesis in overgrowth syndromes: a comprehensive review. *Am. J. Med. Genet. C. Semin. Med. Genet.* *137C*, 53–71.
  46. Ernst, T., Chase, A.J., Score, J., Hidalgo-Curtis, C.E., Bryant, C., Jones, A.V., Waghorn, K., Zoi, K., Ross, F.M., Reiter, A., et al. (2010). Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. *Nat. Genet.* *42*, 722–726.
  47. Ntziachristos, P., Tsiganos, A., Van Vlierberghe, P., Nedjic, J., Trimarchi, T., Flaherty, M.S., Ferres-Marco, D., da Ros, V., Tang, Z., Siegle, J., et al. (2012). Genetic inactivation of the polycomb repressive complex 2 in T cell acute lymphoblastic leukemia. *Nat. Med.* *18*, 298–301.
  48. Bödör, C., Grossmann, V., Popov, N., Okosun, J., O'Riain, C., Tan, K., Marzec, J., Araf, S., Wang, J., Lee, A.M., et al. (2013). EZH2 mutations are frequent and represent an early event in follicular lymphoma. *Blood* *122*, 3165–3168.
  49. Okosun, J., Bödör, C., Wang, J., Araf, S., Yang, C.Y., Pan, C., Boller, S., Cittaro, D., Bozek, M., Iqbal, S., et al. (2014).

- Integrated genomic analysis identifies recurrent mutations and evolution patterns driving the initiation and progression of follicular lymphoma. *Nat. Genet.* *46*, 176–181.
50. Yang, L., Rau, R., and Goodell, M.A. (2015). DNMT3A in haematological malignancies. *Nat. Rev. Cancer* *15*, 152–165.
51. Cancer Genome Atlas Network (2015). Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature* *517*, 576–582.
52. Rauch, A., Wiczorek, D., Graf, E., Wieland, T., Endeke, S., Schwarzmayr, T., Albrecht, B., Bartholdi, D., Beygo, J., Di Donato, N., et al. (2012). Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet* *380*, 1674–1682.
53. de Ligt, J., Willemsen, M.H., van Bon, B.W., Kleefstra, T., Yntema, H.G., Kroes, T., Vulto-van Silfhout, A.T., Koolen, D.A., de Vries, P., Gilissen, C., et al. (2012). Diagnostic exome sequencing in persons with severe intellectual disability. *N. Engl. J. Med.* *367*, 1921–1929.
54. Gilissen, C., Hehir-Kwa, J.Y., Thung, D.T., van de Vorst, M., van Bon, B.W., Willemsen, M.H., Kwint, M., Janssen, I.M., Hoischen, A., Schenck, A., et al. (2014). Genome sequencing identifies major causes of severe intellectual disability. *Nature* *511*, 344–347.
55. Vissers, L.E., Gilissen, C., and Veltman, J.A. (2016). Genetic studies in intellectual disability and related disorders. *Nat. Rev. Genet.* *17*, 9–18.
56. Martinez, F., Caro-Llopis, A., Rosello, M., Oltra, S., Mayo, S., Monfort, S., and Orellana, C. (2017). High diagnostic yield of syndromic intellectual disability by targeted next-generation sequencing. *J. Med. Genet.* *54*, 87–92.