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

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

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 ≥+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.[1,2] 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 ≥2 SD above the mean, together with additional phenotypic abnormalities, the most common of which is intellectual disability.[3] 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.[3,4]

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).[5] OGID syndromes have been increasingly identified over the last decade.[3,4] 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.[5–12]

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).

[1]Division of Genetics and Epidemiology, Institute of Cancer Research, 15 Cotswold Road, London SM2 5NG, UK; [2]South West Thames Regional Genetics Service, St George's University Hospitals NHS Foundation Trust, London SW17 0QT, UK; [3]Child Neurology Unit, Foundation IRCCS C Besta Neurological Institute, Milan 20133, Italy; [4]Landeskrankenanstalten Salzburg, Kinderklinik Department of Pediatrics, Klinische Genetik, Salzburg 5020, Austria; [5]Northern Ireland Regional Genetics Service, Belfast City Hospital, Belfast BT9 7AB, Northern Ireland; [6]Human Development and Health Academic Unit, Faculty of Medicine, University of Southampton, Southampton SO17 1BJ, UK; [7]Wessex Clinical Genetics Service, University Hospital Southampton NHS Trust, Southampton SO16 6YD, UK; [8]West Midlands Regional Genetics Service, Birmingham Women's Hospital NHS Foundation Trust and University of Birmingham, Birmingham Health Partners, Birmingham B15 2TG, UK; [9]Cancer Genetics Unit, Royal Marsden NHS Foundation Trust, London SW3 6JJ, UK
[10]Twitter: @rahman_nazneen
*Correspondence: rahmanlab@icr.ac.uk
http://dx.doi.org/10.1016/j.ajhg.2017.03.010.

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)[13] 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 syndome (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)[14] and the ICR1000 UK exome series[15] 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.[16] We used MLPA to identify 5q35 microdeletions encompassing *NSD1*.[17] *NSD1* MLPA is also capable of detecting exon CNVs that account for ~5% of *NSD1* mutations.[17] 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 2×101 bp reads.

Variant Calling

We used the OpEx v1.0 pipeline to perform variant calling.[18] We converted raw data to FASTQs using CASAVA v.1.8.2 with default settings. The OpEx v1.0 pipeline uses Stampy[19] to map to the human reference genome, Picard to flag duplicates, Platypus[20] to call variants, and CAVA[21] to provide consistent annotation of variants with the HGVS-compliant CSN (Clinical Sequencing Notation) standard v1.0.[21] 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[18]) 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)