Targeted Searches of the Electronic Health Record and Genomics Identify an Etiology in Three Patients with Short Stature and High IGF-I Levels

Catalina Cabrera-Salcedo^{a, b} Colin P. Hawkes^{c, d} Leah Tyzinski^a
Melissa Andrew^{a, e} Guillaume Labilloy^f Diego Campos^g Amalia Feld^h
Annalisa Deodati^{e, i} Vivian Hwa^{a, j} Joel N. Hirschhorn^{h, k} Adda Grimberg^{c, d}
Andrew Dauber^{a, e, l} the Genomics Research and Innovation Network

^aCincinnati Center for Growth Disorders, Division of Endocrinology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA; ^bDivision of Endocrinology, Department of Pediatrics, University of Louisville, Louisville, KY, USA; ^cDivision of Endocrinology and Diabetes, The Children's Hospital of Philadelphia, Philadelphia, PA, USA; ^dDepartment of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; ^eDivision of Endocrinology and Center for Genetic Medicine Research, Children's National Hospital, Washington, DC, USA; ^fDivision of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA; ^gDepartment of Biomedical and Health Informatics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA; ^hDivision of Endocrinology, Boston Children's Hospital, Boston, MA, USA; ⁱDipartimento Pediatrico Universitario Ospedaliero "Bambino Gesù" Children's Hospital-Tor Vergata University, Rome, Italy; ^jDepartment of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA; ^kPrograms in Metabolism and Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA; ^lDepartment of Pediatrics, George Washington University School of Medicine and Health Sciences, Washington, DC, USA

Keywords

Electronic Health Records · Genomics · Short stature · Bioinformatics · Exome sequencing · IGF1R

Abstract

Introduction: Short stature is one of the most common reasons for referral to a pediatric endocrinologist and can result from many etiologies. However, many patients with short stature do not receive a definitive diagnosis. **Objective:** To ascertain whether integrating targeted bioinformatics searches of electronic health records (EHRs) combined with genomic studies could identify patients with previously un-

diagnosed rare genetic etiologies of short stature. We focused on a specific rare phenotypic subgroup: patients with short stature and elevated IGF-I levels. **Methods:** We performed a cross-sectional cohort study at three large academic pediatric healthcare networks. Eligible subjects included children with heights below -2 SD, IGF-I levels >90 th percentile, and no known etiology for short stature. We performed a search of the EHRs to identify eligible patients. Patients were then recruited for phenotyping followed by exome se-

Members of the Genomics Research and Innovation Network are listed in the Supplementary Appendix

quencing and in vitro assays of IGF1R function. **Results:** A total of 234 patients were identified by the bioinformatics algorithm with 39 deemed eligible after manual review (17%). Of those, 9 were successfully recruited. A genetic etiology was identified in 3 of the 9 patients including 2 novel variants in *IGF1R* and a de novo variant in *CHD2*. In vitro studies supported the pathogenicity of the *IGF1R* variants. **Conclusions:** This study provides proof of principle that patients with rare phenotypic subgroups can be identified based on discrete data elements in the EHRs. Although limitations exist to fully automating this approach, these searches may help find patients with previously unidentified rare genetic disorders.

© 2019 S. Karger AG, Basel

Introduction

Concerns about short stature are one of the most common reasons for referral to a pediatric endocrinologist. Human height is influenced by a multitude of biological processes, and molecular variants in a wide array of genes can result in short stature [1]. Consequently, despite adequate clinical and laboratory evaluation, a pathological diagnosis is not identified in the vast majority of children evaluated for short stature [2]. While many of these children have variants of normal physiology, a number have undiagnosed genetic growth disorders. Recent advances in genomic technologies have revolutionized our ability to discover novel or rare genetic etiologies of growth failure. However, it is often quite difficult to determine which of the many patients seen for short stature may have an undiagnosed genetic cause.

The widespread adoption of electronic health records (EHRs) in the US has created a rich asset to facilitate clinical research. EHRs are commonly used as the main data source for safety surveillance, observational, epidemiological, and prospective cohort studies [3, 4]. Most recently, with the increased popularity of EHR-linked biorepositories, EHRs have become a valued tool for the augmentation of clinical characterization to drive discovery of phenotype-genotype associations [5, 6]. EHRs improve our ability to identify cohorts of patients with a specific phenotype of interest by using a combination of diagnosis codes, pharmacy data, narrative text, laboratory values and other clinical features including anthropometric data [5].

The objective of our study was to assess the feasibility of using dense phenotypic information from the EHR to systematically identify a distinct cohort of patients with short stature who have a high probability of harboring a

monogenic etiology for their short stature. As an initial proof of principle, we chose to focus on patients with elevated insulin-like growth factor I (IGF-I) levels, a rare finding in patients with short stature. High IGF-I levels are most commonly seen in patients with heterozygous mutations of the gene encoding the receptor for IGF-I (*IGF1R*). Affected individuals typically present with distinct phenotypic features including prenatal and postnatal growth retardation, microcephaly, and often with serum IGF-I concentrations above the normal ranges [7–9]. However, despite multiple reports in the literature, IGF-I resistance due to *IGF1R* mutations is underdiagnosed in clinical practice, perhaps because of its wide phenotypic spectrum [9–11] as well as laboratory evaluation that can be misinterpreted as reassuring. Our group previously conducted large-scale candidate gene sequencing in a cohort of patients with idiopathic short stature and identified a frameshift mutation in *IGF1R* in a patient with clinical features suggestive of a pathological IGF1R deficiency state [12]. Similarly, Storr et al. [13] reported a case of a novel pathogenic *IGF1R* mutation identified in a cohort of patients with IGF-I resistance. Most recently, Yang et al. [11] reported four novel variants in *IGF1R* in a cohort of patients with prenatal and postnatal growth failure. In addition, numerous other conditions might resemble the phenotype of IGF-I resistance [14].

In the present study, we sought to provide a proof of principle that searches of the EHR could identify a discrete clinical subgroup of patients with short stature who were at high risk of having a particular type of undiagnosed genetic disorder. We executed this algorithm at three leading pediatric hospitals and discuss the challenges in its implementation.

Materials and Methods

Identification of Cohort of Patients with IGF-I Resistance

We performed a targeted bioinformatics search of the EHR at three large academic pediatric endocrine departments (Boston Children's Hospital (BCH), Children's Hospital of Philadelphia (CHOP), and Cincinnati Children's Hospital Medical Center (CCHMC)). CHOP and CCHMC use the EPIC EHR system (Epic, Verona, WI, USA) while BCH uses the Cerner system (Cerner, North Kansas City, MO, USA). These are two of the most widely used EHR systems in the United States. The research study protocol was approved by the institutional review board of each institution. Only subjects seen since 2012 were included as they were the most likely to be available for prospective recruitment. Inclusion criteria were height below -2 SD (adjusted for age and sex, based on CDC reference tables [15]) and an IGF-I level above the 90th percentile (assay specific and adjusted for age and sex [16]). An elevated IGF-I level was required only at a single time point in or-

Table 1. Summary of EHR search and recruitment process

	CCHMC	CHOP	BCH
Total patients identified with bioinformatics EHR search (Ht <−2 SD, IGF-I >90th percentile)	104 patients	68 patients	62 patients
Excluded	87 patients – 63 with Duchenne muscular dystrophy on high-dose steroids – 11 with other known genetic disorder ^a – 9 with other chronic conditions – 3 without persistent IGF-I elevation – 1 with precocious puberty	54 patients – 1 with Duchenne muscular dystrophy on high-dose steroids – 4 on high-dose steroids – 7 with other known genetic disorder ^b – 7 with other chronic condition – 15 without persistent IGF-I elevation – 10 achieved height >−2 SD without treatment – 2 receiving GH treatment but missed by algorithm – 8 with normal IGF-I corrected for pubertal status	54 patients – 1 with Duchenne muscular dystrophy on high-dose steroids – 6 with other known genetic disorder ^c – 9 with other chronic condition – 1 without persistent IGF-I elevation – 21 achieved normal height without treatment – 3 on GH treatment – 3 on rhIGF-I treatment – 10 with precocious puberty
Eligible for recruitment	17 patients – 8 unable to contact – 3 did not show for research visit – 1 refused participation	14 patients – 2 unable to contact – 9 refused participation	8 patients – 6 unable to contact – 1 refused participation
Recruited for study	5 patients	3 patients	1 patient

CCHMC, Cincinnati Children's Hospital Medical Center; CHOP, Children's Hospital of Philadelphia; BCH, Boston Children's Hospital; ^a Chromosome deletion syndrome (2 patients), Williams syndrome, congenital adrenal hyperplasia, trisomy 21 (2 patients), hypochondroplasia, Cockayne syndrome, SHOX deficiency, CHARGE syndrome, FGFR1 mutation. ^b Trisomy 21, Turner syndrome, Angelman syndrome, neurofibromatosis type 1, retinitis pigmentosa, Aicardi Goutier syndrome, Noonan syndrome. ^c Osteogenesis imperfecta, progeria (4 patients), leukocyte adhesion deficiency type I.

der to meet screening criteria to maximize capture rates. At all sites, the bioinformatics algorithm excluded patients who had growth hormone (GH) or IGF-I listed as an active medication at the time of IGF-I measurement. This was done to exclude patients in whom the IGF-I elevation was likely a result of treatment.

Determination of IGF-I percentiles was site-specific. At CCHMC, serum IGF-I was determined by a chemiluminescent immunometric assay (IDS-iSYS; Immunodiagnostic Systems Ltd., Boldon, UK). A single IGF-I assay was used for all the samples during this period. IGF-I percentiles were calculated using the assay's manufacturer datasheets [16]. At CHOP, IGF-I levels were measured by various commercial laboratories, per clinical routine (patient and insurance preference), and some of the laboratories utilized different IGF-I assays during the study period. Representative samples were identified for each of the IGF-I assays for each laboratory using the EHR. The reference ranges listed in each assay were recorded manually for the pubertal stage and/or the age of the patient. Mean and standard deviations were generated from these, and an algorithm was created to query the EHR and generate percentiles for each test according to pubertal stage and/or age. Some assays reported Z scores directly and, where possible, these were extracted and converted to percentiles. In order to validate this approach, we counted the number of samples with IGF-I concentrations above the 90th percentile for each assay to ensure that no assay was overrepresented on initial screen. At BCH, a single IGF-

I assay (Immulite, Siemens Medical Solutions, Tarrytown, NY, USA) was used except for a 6-month period when samples were sent to ARUP. IGF-I percentiles were calculated for these 2 assays using the assay's manufacturer datasheets.

Charts of the patients identified by the bioinformatics screening algorithm were manually reviewed by one of the investigators at each of the participating institutions. Patients with known underlying genetic conditions or other chronic conditions (not a structural birth defect) explaining their short stature, and patients with precocious puberty were excluded from the study (Table 1). Patients who subsequently achieved a normal height or who had a repeat IGF-I level that was not elevated without treatment were excluded as the intention of the study was to only examine patients with a likely genetic etiology for their clinical presentation. All eligible patients were approached by mail or telephone and offered the opportunity to participate in the study. Written informed consent and assent were obtained from all participants and/or their parents/guardians, as age appropriate. The eligible patients and their immediate relatives were invited to attend a single research study visit at each of the participating institutions for standardized phenotyping and sample acquisition (blood, saliva, or both).

Genetic Analysis

Whole exome sequencing was performed either at CCHMC or at the Broad Institute on DNA extracted from whole blood or sa-

Table 2. Clinical information of the patients recruited for genetic testing

Patient	Gender	GA, weeks	BW, g (SD)	Age, years ^a	Height, cm (SD) ^a	IGF-I, ng/mL (normal range)	IGFBP-3, ng/mL (normal range)	Additional clinical data	Mutation data
1	Male	34	1,928 (-0.8)	5.5	94 (-3.7)	279.7 (50–233)	4,290 (1,843–4,968)	Developmental delay, hearing loss, hypercalcemia, cystinuria, dysmorphic features	No mutation identified
2	Female	38	3,062 (0.03)	6	95 (-4.2)	299.9 (42–240)	7,067 (1,945–5,403)	Congenital lymphedema, nephrotic syndrome secondary to mesangial sclerosis	No mutation identified
3	Female	37	2,045 (-1.9)	2.9	81 (-2.9)	139.7 (24–154)	2,775 (1,471–4,357)	Tricuspid atresia, body asymmetry, dysmorphic features, failure to thrive	No mutation identified
4	Female	36	1,559 (-2.7)	10	124 (-2.2)	378 (79–400)	4,650 (2,300–5,801)	Twin; cloacal exstrophy, dextrocardia, dysplastic kidney	No mutation identified
5	Male	40	3,062 (-1.1)	15	145 (-2.8)	535.1 (127–554)	7,677 (2,614–6,306)	Autism, ADHD, intellectual disability, dysmorphic features	CHD2 c.1618G>T, p.Val540Phe
6	Male	40	2,690 (-2.0)	7.5	111.8 (-3.0)	261 (44–211, Z = 3)	3,000 (2,100–4,200)	ADHD, familial short stature	IGF1R c.3037G>T, p.Val1013Phe
7	Female	41	2,778 (-1.8)	8.6	118.6 (-2.6)	255 (64–259, Z = 1.9)	3,400 (2,100–4,200)	Hypotonia, speech and fine motor developmental delay; upslanting palpebral fissures	No mutation identified
8	Female	39	3,700 (0.9)	7.8	119.1 (-2.3)	215 (55–238, Z = 1.6)	2,900 (2,100–4,200)	Sensorineural hearing loss	No mutation identified
9	Female	39	2,438 (-1.8)	8.9	114.2 (-3.1)	310 (64–345)	5,200 (1,600–6,500)	Idiopathic short stature	IGF1R c.80_82del, p.Thr28del

GA, gestational age; BW, birth weight; IGF-I, insulin-like growth factor I; IGFBP-3, insulin-like growth factor-binding protein-3. Assay-specific IGF-I and IGFBP-3 reference data from commercial laboratories presented. IGF-I Z score included when provided by commercial laboratory. Height Z score according to CDC growth charts. ^a At the time of evaluation.

liva samples from 9 patients and their immediate family members. Details of the exome sequencing methods have been previously described [17]. For each patient, we performed a number of analyses. For all subjects, we searched for nonsynonymous, homozygous, or compound heterozygous recessive variants with a minor allele frequency below 0.01 in the 1,000 Genomes and gnomAD databases (<http://browser.1000genomes.org/>, and <http://gnomad.broadinstitute.org/>). Missense variants were only included if they were predicted to be pathogenic by a minimum of 2 of the following in silico prediction programs: Polyphen2, MutationTaster, and SIFT. If DNA was available from both parents, we searched for de novo variants in the proband which were required to be absent from the aforementioned databases (i.e., novel variants). In male subjects, we included a search for potential X-linked variants which were also required to be novel. Finally, if a parent was also affected with short stature, we looked for novel heterozygous variants inherited from the affected parent. In addition to exome sequencing, the subjects had previously undergone clinical testing for copy number variants, as assessed by chromosomal microarray analysis, and were all negative for significant copy number variants.

Functional Studies

Generation of Recombinant IGF1R

Each of the identified *IGF1R* missense variants was regenerated with the QuikChange[®] II site-directed mutagenesis kit (Agilent Technologies Santa Clara, CA, USA) using *IGF1R* cDNA in the pcDNA3.1/AMP expression plasmid, as previously described [18, 19]. The primers used for generating the variants were: Val1013Phe (V1013F) 5'-GGGGTCGTTTGGGATGTTCTATGAAGGAGT-TGC-3' (sense), 5'-GCAACTCCTTCATAGAACATCCCAAA-CGACCCC-3' (antisense); Thr28del: 5'-GCTCTGGCCGAGT-GGAGAAATCT-3' (sense), 5'-ATTTCTCCACTCGGCCAGAGC-3' (antisense). The resultant cDNA constructs carrying the point mutations were verified by Sanger DNA sequencing.

Cell Culture and Transfection

HEK293 cells (ATCC; LGC Standards, Wesel, Germany) were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum, 100 U mL⁻¹ penicillin, and 100 µg mL⁻¹ streptomycin at 37 °C in 5% CO₂. Cells were seeded in 6-well dishes at a density of 30–50,000 cells/well and transfected when approximately 70% confluent. Cells were transiently transfected with pcDNA3.1 vector or plasmids encoding WT or variant

IGF1R for Thr28del and V1013F, using PolyJet In Vitro DNA Transfection Reagent (SignaGen; Rockville, MD, USA) following the manufacturer's instructions. Transfected HEK293 cells were placed in serum starvation media (DMEM supplemented with 0.1% bovine serum albumin (BSA) and penicillin/streptomycin) for 8–24 h followed by fresh starvation media for 90 min. After 90 min, cells were treated with 100 ng of recombinant human IGF-I (Life Technologies, Frederick, MD, USA) in DMEM or a negative treatment of DMEM with phosphate-buffered saline (PBS) for 10 min, washed in cold PBS, and cells solubilized in lysis buffer (1× PBS, 1% Nonidet P-40, 0.5% sodium deoxycholate, 0.1% SDS, 10 mg mL⁻¹ phenylmethylsulfonyl fluoride, 100 nM sodium orthovanadate) and protease inhibitor mixture (complete Mini-EDTA-free; Sigma, St. Louis, MO, USA). Cell debris was removed by centrifugation, and final whole-cell lysates were stored at –20 °C.

Western Immunoblot Analysis

We performed the BioRad DC protein assay according to the manufacturer's protocol (Biorad, Hercules, CA, USA). For immunoblot analysis, 30 µg of whole-cell lysates were separated on 4–20% SDS-polyacrylamide gels, transferred to 0.45-µm nitrocellulose membranes. Membranes were blocked in 5% BSA prior to primary antibody exposure (overnight at 4 °C). The primary antibodies were: rabbit monoclonal IgG against IGF1R β-subunit (D23H3) (dilution 1:1,000), rabbit monoclonal IgG against phospho-IGF1R β-subunit (Tyr 1135/1136) (dilution 1:1,000), and mouse monoclonal IgG against alpha-tubulin (DM1A) from Cell Signaling Technologies (Beverly, MA, USA). Secondary antibodies (anti-mouse IgG and anti-rabbit IgG) conjugated with horseradish peroxidase were obtained from BioRad. Detection of immune-labeled proteins was performed using a commercial chemiluminescent assay (ECL prime; Amersham, UK). Visualization and quantitative measurements were performed on the BioRad ChemiDoc Touch Imaging System, and densitometry was performed using Image J software. All immunoblot data shown are representative of at least three independent experiments. Unpaired *T* tests were used to compare densitometric analyses between wild-type and variant IGF1R.

Results

Identification of Cohort of Patients with IGF-I Resistance

An overview of eligibility assessment and recruitment is provided in Table 1. In total, 234 patients were identified by the algorithm across all three institutions with 39 being deemed eligible after manual review (17% eligibility rate). Of those, 9 were successfully recruited (23% recruitment rate). Interestingly, 89 of a total of 234 patients (38%) flagged by the algorithm had a preidentified genetic condition including 65 patients with Duchenne muscular dystrophy. Additional details of the excluded patients are available in Table 1. The 9 recruited subjects ranged in age from 2 to 15 years and had a mean height SDS at presentation of –2.7 SD (range –2.2 to –4.2 SD)

Fig. 1. a Schematic of human IGF1R structural domains and encoding exons. Selected previously reported mutations are indicated [31]. The two new heterozygous mutations are boxed. L1, L2, leucine-rich domains; CR, cysteine-rich, furin-like domains; FN1, 2, 3, fibronectin type III domains; TM, transmembrane domain; TK, tyrosine kinase catalytic domain, CT, carboxy-terminus tail. **b** IGF1R signal sequence cleavage site. The signal sequence is boxed, with numbering of amino acid residues shown. The 5 critical residues are underlined. Thr28 is the first amino acid of these five critical residues. In wild type (WT), the cleavage site is known and accurately predicted in silico by <http://www.cbs.dtu.dk/services/SignalP/>. The predicted probability of cleavage is indicated for both WT and the p.Thr28del.

(Table 2). The majority of subjects had additional syndromic features, although some were classified as idiopathic short stature. Additional clinical details are provided in Table 2.

Genomic Analysis

Exome analysis was performed in all 9 patients. DNA was available for both parents in 6 of the families, and paternal samples were missing in 3 families (patients 1, 6, and 9). Prioritization of exome sequencing variants was performed as per the methods stated above. Details of all variants identified in the analysis can be found in the online supplementary Table and Excel File (see www.karger.com/doi/10.1159/000504884 for all online suppl. material). We identified a likely genetic etiology in 3 of the 9 cases. Two of the patients had novel, likely pathogenic variants in IGF1R including a novel missense variant (p.Val1013Phe) and a maternally inherited single amino acid deletion (p.Thr28del) (Fig. 1a). Functional validation studies supported the pathogenicity of these 2 variants (see below). In addition, patient 5 was found

to have a novel, de novo missense variant in *CHD2* (p.Val540Phe) which is the likely etiology of his short stature and intellectual disability. Details of these 3 cases are presented below.

Case Reports

Patient 6

Patient 6 is a 7.5-year-old African-American male with short stature (SDS -3) and attention deficit hyperactivity disorder (ADHD) born at 40 weeks of gestation with a birth weight of 2.69 kg (SDS -2.0). His endocrine evaluation at the age of 7.5 years showed elevated serum IGF-I (261 ng/mL, +2.8 SDS) and serum insulin-like growth factor-binding protein-3 (IGFBP-3) of 3 mg/L (normal range 2.1–4.2). On physical exam, he was prepubertal, nondysmorphic, and had proportional body segments. His head circumference was -1.5 SD. Family history revealed that his father had short stature with a height of 160 cm (SDS -2.3), whereas his mother had a normal height of 155 cm (SDS -1.3). The patient, his unaffected mother, and his aunt underwent whole exome sequencing. DNA from his father was not available. Based on the available clinical information, we prioritized heterozygous nonsynonymous variants only present in the patient which could theoretically be de novo or inherited from the affected father. This identified a novel missense variant (c.3037G>T, p.Val1013Phe) in *IGF1R* which was predicted to be damaging by Polyphen2, MutationTaster, and SIFT. The missense variant was validated by Sanger sequencing.

Patient 9

Patient 9 is an 8-year 11-month-old Caucasian female referred to endocrinology for short stature (SDS -3.1). Her birth weight was 2.438 kg (-1.8 SD). On examination, the patient presented with microcephaly with a head circumference of -3.6 SD. At the age of 8 years 5 months, she had an IGF-I level of 310 ng/mL (90–95th percentile) and IGFBP-3 of 5.2 mg/L (normal range 1.6–6.5). Her mother had a normal height just below the 5th percentile (152 cm, SDS -1.7). The patient, her mother, and sister underwent WES. DNA from her father was not available. Given the available clinical information and history of mother's height below the 5th percentile, we prioritized nonsynonymous, heterozygous variants inherited from the mother but not present in the sister whose height was at the 46th percentile. A novel maternally inherited single amino acid deletion (c.80_82del, p.Thr28del) in *IGF1R* was identified and predicted to be damaging. The variant was validated by Sanger sequencing.

Patient 5

Patient 5 is a 15-year-old Caucasian male with short stature (SDS -2.6), intellectual disability, ADHD, and autism who was born at 40 weeks' gestation with a birth weight of 3.062 kg (SDS -1.1) and length of 47.6 cm (SDS -1.21). The patient had a history of microcephaly but his head circumference at birth was not available. In early infancy, he was noted to have mild developmental delay and feeding difficulties. His physical exam revealed multiple dysmorphic features including prominent curved eyebrows, prominent nasal tip, short wide philtrum, full lips, and down turned corners of mouth. He was diagnosed with ADHD and autism during childhood and underwent formal neuropsychological evaluation which revealed an IQ of 72. Parents denied history of seizures. An endocrine evaluation revealed a serum IGF-I of 535.1 ng/mL (90–95th percentile) and IGFBP-3 of 7.7 mg/L (normal range 2.6–6.3). His pubertal development advanced normally. There was no family history of short stature; his mother's height was 155 cm (SDS -1.3) and his father's 175 cm (SDS -0.2). Exome sequencing identified a novel de novo missense variant (c.1618G>T, p.Val540Phe) in *CHD2* which was predicted to be damaging by Polyphen2, MutationTaster, and SIFT. The variant was validated by Sanger sequencing. *CHD2* is involved in chromatin remodeling, and de novo mutations in *CHD2* have been previously shown to cause a wide range of phenotypes including autism, developmental delay, and early-onset seizures [20, 21]. Multiple patients with mutations in *CHD2* have been reported to have short stature [22], and a mouse model of *CHD2* deficiency demonstrated growth retardation [23]. The patient's variant is found in the critical SNF2-related helicase domain, a site of previously identified pathogenic missense mutations [20].

IGF1R Functional Analysis

To investigate the pathogenicity of the novel missense variants in *IGF1R* identified in our patients, plasmids containing wild-type and mutant *IGF1R* were transfected into HEK293 cells. Immunoblot analyses indicated that both the p.V1013F and p.Thr28del variants caused significant decreases in IGF1R expression with concomitant poor responses to IGF-I stimulation (Fig. 2), thus supporting the pathogenic impact of the variants. It is of note that pIGF1R was detected with increasing loading of p.Thr28del cell lysates, but not in response to IGF-I (online suppl. Fig. 1).

Fig. 2. a Western blot analysis of protein cell lysate (30 µg) from transfected HEK293 cells untreated and treated with IGF-I. Antibodies for phosphorylated IGFIRβ (pIGFIRβ), total IGFIRβ (tIGFIRβ), and α-tubulin were employed. Wild-type control has a robust phosphorylation response (pIGFIRβ) to IGF-I treatment.

V1013F and Thr28del show little to no pIGFIRβ expression, and the response to IGF-I is abrogated. tIGFIRβ expression is significantly lower in both mutant variants compared to wild-type control. **b** Densitometric analysis of average tIGFIR expression relative to α-tubulin. * $p < 0.05$, ** $p < 0.001$.

Discussion/Conclusion

Short stature can be caused by numerous genetic etiologies. Often, it can be difficult to ascertain the specific genetic etiology based on clinical grounds alone. In one study, detailed clinical and biochemical evaluation by an endocrinologist and a geneticist yielded an etiology in approximately 14% of patients referred for evaluation of growth retardation or short stature [24]. In the same study, an additional 16.5% of cases were molecularly diagnosed by exome sequencing in a random subset of 200 patients with short stature of no clear etiology [24]. This supports the power of larger-scale genetic sequencing to identify causes of short stature that are missed on clinical grounds alone as many patients present with subtle phenotypic or biochemical abnormalities. However, larger-scale genetic sequencing is too expensive to feasibly perform as a screen for patients presenting with short stature. The purpose of the current study was to leverage an EHR-based algorithm to narrow the scope and focus large-scale genetic screening to identify genetic etiologies in patients with a rare form of short stature. We believe that this could prove a much more efficient approach to identifying patients with rare genetic growth disorders. As a proof of principle, we focused on a distinctive, rare biochemical phenotype, namely patients with elevated IGF-I levels in the setting of short stature. Since IGF-I is a key mediator of GH action, it is unusual for patients with short stature

to present with elevated IGF-I levels. After manual review, we successfully identified a total of 39 patients fitting our criteria, but despite approaching all eligible patients at three large academic centers, only 9 patients were willing to participate in our study. This further underscores the importance of multicenter collaborations for the investigation of rare phenotypic presentations.

The 9 patients in our cohort had a range of phenotypic features including many with dysmorphic features or developmental delay. We identified a genetic etiology in each of 3 patients, two of whom had novel missense variants in *IGF1R*. Notably, these two patients were classified as having idiopathic short stature and did not have additional syndromic features. Both individuals inherited the variants from one of their parents, one of whom had short stature and one of whom had a height in the low normal range (−1.7 SD). Patients who carry pathogenic mutations in *IGF1R* present with a wide range of heights and it is not unexpected to have a parent whose height is in the low-normal range [9]. *IGF1R* is a lead candidate gene for patients with short stature and elevated IGF-I levels, but neither of these patients had clinical testing for mutations in *IGF1R*. One of the patients had a frankly elevated IGF-I level, while serum IGF-I was at the upper end of the normal range for the other 2 patients. The importance of these diagnoses is that GH therapy can increase height in patients with *IGF1R* mutations, although the IGF-I level may need to be titrated to

a higher level [9, 25]. The third patient had a de novo missense variant in *CHD2*, a gene involved in chromatin remodeling. Chromatin dynamics have been identified as playing an important role in overall height biology [26], and defects in this specific gene cause a syndrome marked by developmental delays, autism, early-onset seizures, and often short stature [22]. Of note, our patient did not have seizures, one of the more common manifestations of mutation in *CHD2*, which may be why this gene was not considered as part of the clinical evaluation. It remains mechanistically unclear how *CHD2* perturbs the GH/IGF-I axis. We do not know the reason for the elevation of IGF-I levels in the patients without a molecular diagnosis. It is possible that these patients have mutations outside of the exome, such as in gene regulatory regions or possibly epigenetic changes, which affect the GH/IGF-I axis. Additional research is needed to determine potential etiologies in these patients.

The novel *IGF1R* variants, p.Thr28del and p.V1013F, identified in patient 6 and patient 9 respectively, were likely causal of impaired growth. Both were poorly expressed in reconstitution studies. The p.Thr28del variant, within the secretory signal peptide (amino acid residues 1–30), is close to the proteolytic cleavage site and is the first pathological mutation discovered that involves the signal peptide of *IGF1R*. Appropriate removal of secretory signal peptides is essential for exporting proteins destined for insertion into the plasma membrane or the extracellular milieu. Interestingly, an in silico signal peptide program (<http://www.cbs.dtu.dk/services/SignalP/>) indicates that 5 critical residues, ²⁸ThrSerGlyGluIle³² (Fig. 1b), are important for normal cleavage between Gly30 and Glu31 (predicted high probability of 0.9354 out of 1.0), but loss of residue Thr28 shifts the 5 critical residues to ²⁷ProSerGlyGluIle, which destroys the cleavage site (predicted probability of cleavage dropped to 0.2446). Hence, the poorly immune-detected *IGF1R* p.Thr28del suggests that the aberrant *IGF1R* variant may be targeted for degradation pathways. Previously, several inactivating mutations have been reported within the intracellular tyrosine kinase domain including missense variants p.E1080K [27] and p.G1155A [28]. The *IGF1R* p.V1013F identified in this report lies within the nucleotide binding region of ATP (residues 1,005–1,013), a motif that binds nucleotide phosphates. The missense p.V1013F not only results in significantly reduced protein expression, but the tyrosine kinase activities were also significantly reduced, comparable to the previously described *IGF1R* p.E1080K mutation [26]. Altogether, our studies support *IGF1R* haploinsufficiency in patients 6

and 9 as the likely established cause of short stature consistent with previously reported cases.

This study highlights some of the key difficulties in implementing an EHR-based search for rare phenotypes. First, our algorithm was limited by its ability to only search for data elements that are collected as discrete data fields in the EHR. This allowed us to identify patients based on their height Z scores and IGF-I levels. However, manual review was required and resulted in exclusion of the majority (83%) of screened cases. The major reasons for exclusion were the presence of another chronic condition, IGF-I levels appropriate for pubertal stage, or lack of persistence in IGF-I elevation. Interestingly, 65 out of the 234 (28%) patients flagged by the algorithm had Duchenne muscular dystrophy. Previous reports examining the etiology of short stature in these patients have found IGF-I levels within the normal range [29]. However, the elevated IGF-I levels may be due to concomitant treatment with high-dose steroids which can lead to IGF-I resistance and elevation in circulating IGF-I levels [30]. The large number of patients with this condition is likely due to a specific multidisciplinary clinic at CCHMC for these patients. Diagnostic coding is not particularly reliable in the EHR and thus simple selection of patients with diagnostic codes for short stature or growth failure would not exclude patients with underlying chronic condition. It is possible to create a list of diagnoses to exclude which may be a more reliable approach and would decrease the numbers of charts needed for manual review, but that list would never be comprehensive enough to fully eliminate the need for manual review. Pubertal status is not routinely recorded in our current EHR systems as a discrete data element but is rather often found as text within the clinic note. To further streamline the EHR search, a discrete data field for pubertal status could be created within note templates for growth evaluations, or alternatively, searches that employ natural language processing could “read” the notes and incorporate the pubertal information into the search. Finally, for simplicity of algorithm coding, we did not include a rule requiring serial IGF-I levels to be persistently elevated, but this could be easily incorporated into a future version of the algorithm. In sum, our study provides proof of principle that an EHR algorithm can identify a relevant patient subpopulation among the many patients presenting with short stature, but significant improvements in search strategy are required before this can be fully automated into routine clinical practice. Despite these limitations and the need for manual review, our approach was able to relatively quickly identify a high yield group of individuals from the

thousands of children evaluated for short stature at our three large medical centers.

In conclusion, our study demonstrates the feasibility of identifying rare phenotypic subgroups based on discrete data elements in the EHR. With further improvements in the algorithm, it should now be possible to proactively flag these patients as being at high risk for having an underlying genetic etiology. In our cohort, after significant manual review, we found a genetic etiology in 3 of 9 patients with short stature and high IGF-I levels, and evaluation of the *IGF1R* gene alone would have identified a molecular cause in 2 of the 9 patients. As these are rare genetic etiologies, practicing clinicians may not be aware of the likelihood of identifying a specific genetic etiology or realize that such an etiology exists. Decision support through the EHR can alert clinicians to the possibility of rare disorders whose diagnosis could affect therapeutic decisions. Our study was not designed to examine the yield of an algorithm-based approach versus comprehensive sequencing as an initial approach in the diagnosis of short stature as some have advocated. Further research will be needed to compare these approaches, but we first sought to provide proof of principle that an algorithm-based approach could prove useful in the evaluation of rare subgroups of patients with short stature. Additional research is needed to understand the applicability of this approach to a wider range of phenotypic subgroups, but our model of multi-institutional collaboration will be critical to the success of these endeavors.

Acknowledgements

We would like to thank Casey Gorman and Anna Bartels for assistance with subject recruitment. Research reported in this publication was supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health under Award Number R01HD093622. The

content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Additional funding support was provided by the Genomics Research and Innovation Network.

Statement of Ethics

All studies were conducted in compliance with the Declaration of Helsinki. Institutional review board approval was obtained for all studies, and written informed consent was obtained from subjects or their guardians as appropriate.

Disclosure Statement

A.G. serves on the Steering Committee of the Pfizer International Growth Study Database and is a Consultant for Sandoz Inc. A.D. has received research support from Novo Nordisk and Ipsen and has served as a consultant for Ascendis, OPKO Biologics, Pfizer, Ipsen, Sandoz, and Novo Nordisk. J.N.H. is a member of the scientific advisory board for Camp4 Therapeutics.

Funding Sources

Research reported in this publication was supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health under Award Number R01HD093622. Additional funding support was provided by the Genomics Research and Innovation Network. None of the funding agencies played any role in the study design, conduct, analysis, or writing of the manuscript.

Author Contributions

C.C.S. and A.D. designed the study, performed data analysis, and wrote the manuscript. C.P.H., L.T., A.F., J.N.H., and A.G. recruited patients, phenotyped patients, and participated in data analysis. G.L. and D.C. performed the bioinformatics analysis. M.A., A.D., and V.H. designed and performed the in vitro studies. All authors reviewed and edited the manuscript.

References

- 1 Dauber A, Rosenfeld RG, Hirschhorn JN. Genetic evaluation of short stature. *J Clin Endocrinol Metab*. 2014 Sep;99(9):3080–92.
- 2 Sisley S, Trujillo MV, Khoury J, Backeljauw P. Low incidence of pathology detection and high cost of screening in the evaluation of asymptomatic short children. *J Pediatr*. 2013 Oct;163(4):1045–51.
- 3 Dean BB, Lam J, Natoli JL, Butler Q, Aguilar D, Nordyke RJ. Review: use of electronic medical records for health outcomes research: a literature review. *Med Care Res Rev*. 2009 Dec;66(6):611–38.
- 4 Cowie MR, Blomster JI, Curtis LH, Duclaux S, Ford I, Fritz F, et al. Electronic health records to facilitate clinical research. *Clin Res Cardiol*. 2017 Jan;106(1):1–9.
- 5 Kohane IS. Using electronic health records to drive discovery in disease genomics. *Nat Rev Genet*. 2011 Jun;12(6):417–28.
- 6 Wolford BN, Willer CJ, Surakka I. Electronic health records: the next wave of complex disease genetics. *Hum Mol Genet*. 2018 May;27 R1:R14–21.
- 7 Abuzzahab MJ, Schneider A, Goddard A, Griorescu F, Lautier C, Keller E, et al.; Intrauterine Growth Retardation (IUGR) Study Group. IGF-I receptor mutations resulting in intrauterine and postnatal growth retardation. *N Engl J Med*. 2003 Dec;349(23):2211–22.
- 8 David A, Hwa V, Metherell LA, Netchine I, Camacho-Hübner C, Clark AJ, et al. Evidence for a continuum of genetic, phenotypic, and biochemical abnormalities in children with growth hormone insensitivity. *Endocr Rev*. 2011 Aug;32(4):472–97.

- 9 Walenkamp MJ, Robers JM, Wit JM, Zandwijken GR, van Duyvenvoorde HA, Oostdijk W, et al. Phenotypic Features and Response to GH Treatment of Patients With a Molecular Defect of the IGF-1 Receptor. *J Clin Endocrinol Metab.* 2019 Aug;104(8):3157–71.
- 10 Walenkamp MJ, Losekoot M, Wit JM. Molecular IGF-1 and IGF-1 receptor defects: from genetics to clinical management. *Endocr Dev.* 2013;24:128–37.
- 11 Yang L, Xu DD, Sun CJ, Wu J, Wei HY, Liu Y, et al. IGF1R variants in patients with growth impairment: four novel variants and genotype-phenotype correlations. *J Clin Endocrinol Metab.* 2018 Nov;103(11):3939–44.
- 12 Wang SR, Carmichael H, Andrew SF, Miller TC, Moon JE, Derr MA, et al. Large-scale pooled next-generation sequencing of 1077 genes to identify genetic causes of short stature. *J Clin Endocrinol Metab.* 2013 Aug;98(8):E1428–37.
- 13 Storr HL, Dunkel L, Kowalczyk J, Savage MO, Metherell LA. Genetic characterisation of a cohort of children clinically labelled as GH or IGF1 insensitive: diagnostic value of serum IGF1 and height at presentation. *Eur J Endocrinol.* 2015 Feb;172(2):151–61.
- 14 Finken MJ, van der Steen M, Smeets CC, Walenkamp MJ, de Bruin C, Hokken-Koelega AC, et al. Children born small for gestational age: differential diagnosis, molecular genetic evaluation and implications. *Endocr Rev.* 2018 Dec;39(6):851–94.
- 15 Kuczmarski RJ, Ogden CL, Guo SS, Grummer-Strawn LM, Flegal KM, Mei Z, et al. 2000 CDC Growth Charts for the United States: methods and development. *Vital Health Stat* 11. 2002 May;(246):1–190.
- 16 Bidlingmaier M, Friedrich N, Emeny RT, Spranger J, Wolthers OD, Roswall J, et al. Reference intervals for insulin-like growth factor-1 (igf-i) from birth to senescence: results from a multicenter study using a new automated chemiluminescence IGF-I immunoassay conforming to recent international recommendations. *J Clin Endocrinol Metab.* 2014 May;99(5):1712–21.
- 17 de Bruin C, Mericq V, Andrew SF, van Duyvenvoorde HA, Verkaik NS, Losekoot M, et al. An XRCC4 splice mutation associated with severe short stature, gonadal failure, and early-onset metabolic syndrome. *J Clin Endocrinol Metab.* 2015 May;100(5):E789–98.
- 18 Fang P, Schwartz ID, Johnson BD, Derr MA, Roberts CT Jr, Hwa V, et al. Familial short stature caused by haploinsufficiency of the insulin-like growth factor I receptor due to nonsense-mediated messenger ribonucleic acid decay. *J Clin Endocrinol Metab.* 2009 May;94(5):1740–7.
- 19 Fang P, Cho YH, Derr MA, Rosenfeld RG, Hwa V, Cowell CT. Severe short stature caused by novel compound heterozygous mutations of the insulin-like growth factor 1 receptor (IGF1R). *J Clin Endocrinol Metab.* 2012 Feb;97(2):E243–7.
- 20 Carvill GL, Heavin SB, Yendle SC, McMahon JM, O'Roak BJ, Cook J, et al. Targeted resequencing in epileptic encephalopathies identifies de novo mutations in CHD2 and SYNGAP1. *Nat Genet.* 2013 Jul;45(7):825–30.
- 21 Petersen AK, Streff H, Tokita M, Bostwick BL. The first reported case of an inherited pathogenic CHD2 variant in a clinically affected mother and daughter. *Am J Med Genet A.* 2018 Jul;176(7):1667–9.
- 22 Suls A, Jaehn JA, Kecskés A, Weber Y, Weckhuysen S, Craiu DC, et al.; EuroEPINOMICS RES Consortium. De novo loss-of-function mutations in CHD2 cause a fever-sensitive myoclonic epileptic encephalopathy sharing features with Dravet syndrome. *Am J Hum Genet.* 2013 Nov;93(5):967–75.
- 23 Kulkarni S, Nagarajan P, Wall J, Donovan DJ, Donnell RL, Ligon AH, et al. Disruption of chromodomain helicase DNA binding protein 2 (CHD2) causes scoliosis. *Am J Med Genet A.* 2008 May;146A(9):1117–27.
- 24 Hauer NN, Popp B, Schoeller E, Schuhmann S, Heath KE, Hisado-Oliva A, et al. Clinical relevance of systematic phenotyping and exome sequencing in patients with short stature. *Genet Med.* 2018 Jun;20(6):630–8.
- 25 Walenkamp MJ, de Muinck Keizer-Schrama SM, de Mos M, Kalf ME, van Duyvenvoorde HA, Boot AM, et al. Successful long-term growth hormone therapy in a girl with haploinsufficiency of the insulin-like growth factor-I receptor due to a terminal 15q26.2-[[GT]]qter deletion detected by multiplex ligation probe amplification. *J Clin Endocrinol Metab.* 2008 Jun;93(6):2421–5.
- 26 Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature.* 2010 Oct;467(7317):832–8.
- 27 Walenkamp MJ, van der Kamp HJ, Pereira AM, Kant SG, van Duyvenvoorde HA, Kruit-hof MF, et al. A variable degree of intrauterine and postnatal growth retardation in a family with a missense mutation in the insulin-like growth factor I receptor. *J Clin Endocrinol Metab.* 2006 Aug;91(8):3062–70.
- 28 Kruis T, Klammt J, Galli-Tsinopoulou A, Wallborn T, Schlicke M, Müller E, et al. Heterozygous mutation within a kinase-conserved motif of the insulin-like growth factor I receptor causes intrauterine and postnatal growth retardation. *J Clin Endocrinol Metab.* 2010 Mar;95(3):1137–42.
- 29 Nagel BH, Mortier W, Elmlinger M, Wollmann HA, Schmitt K, Ranke MB. Short stature in Duchenne muscular dystrophy: a study of 34 patients. *Acta Paediatr.* 1999 Jan;88(1):62–5.
- 30 Miell JP, Taylor AM, Jones J, Holly JM, Gail-lard RC, Pralong FP, et al. The effects of dexamethasone treatment on immunoreactive and bioactive insulin-like growth factors (IGFs) and IGF-binding proteins in normal male volunteers. *J Endocrinol.* 1993 Mar;136(3):525–33.
- 31 Rosenfeld RG. Genetic Diagnosis of Growth Failure. Genetic Diagnosis of Endocrine Disorders. 2nd ed. Amsterdam: Elsevier; 2016.