DIGITAL ACCESS TO
SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY
Office for Scholarly Communication

Whole Exome Sequencing and Large-Scale Targeted Sequencing to Identify Novel Candidate Genes in Idiopathic Short Stature

Citation
Carmichael, Heather. 2015. Whole Exome Sequencing and Large-Scale Targeted Sequencing to Identify Novel Candidate Genes in Idiopathic Short Stature. Doctoral dissertation, Harvard Medical School.

Permanent link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:17295875

Terms of Use
This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story .

Accessibility

TABLE OF CONTENTS

1. *Title Page and Abstract*……………………………………………………………………1

2. *List of Abbreviations*……………………………………………………………………3

3. *Attributions*…………………………………………………………………...……………4

4. *Chapter I: Introduction*………………………………………………………………....7

5. *Chapter II: Whole Exome Sequencing in a Patient with Uniparental Disomy of Chromosome 2 and a Complex Phenotype*…………………………………….………….……14

6. *Chapter III: Large Scale Pooled Next-Generation Sequencing of 1077 Genes to Identify Genetic Causes of Short Stature*……………………………………………..……………31

7. *Chapter IV: Heterozygous mutations in natriuretic peptide receptor-B (NPR2) gene as a cause of short stature*………………………………………………………………………..51

8. *References*……………………………………………………...………………...69

LIST OF ABBREVIATIONS

AMDM	acromesomelic dysplasia, type Maroteaux
BBS	Bardet-Biedl syndrome
CNP	C-type natriuretic peptide
CGH	comparative genomic hybridization
CNV	copy number variant
FACS	fluorescent-activated cell sorting
FISH	fluorescent in-situ hybridization
FHS	Framingham Heart Study
GAD	glutamic acid decarboxylase antibody
GH	growth hormone
GWAS	genome-wide association studies
HGMD	Human Gene Mutation Database
IA2	insulinoma antibody 2
IGF-1	insulin-like growth factor 1
IGF-1R	insulin-like growth factor 1 receptor
IGF-BP3	insulin like growth factor binding protein
ISS	idiopathic short stature
MAF	minor allele frequency
NPR2	natriuretic peptide receptor-B
RP	retinitis pigmentosa
OMIM	Online Mendelian Inheritance in Man
PBMC	peripheral blood mononuclear cells
SDS	standard deviations
SNP	single nucleotide polymorphism
UPD	uniparental disomy

ATTRIBUTIONS

During the summer of 2012, I started working with Dr. Andrew Dauber, a pediatric endocrinologist at Children's Hospital Boston. Most of the chapters of this project are manuscripts of now published papers that developed over the course of that initial summer and which I continued to work on during my clinical years. We collaborated with many members of Dr. Joel Hirschhorn's lab and others at the Broad Institute. We also collaborated extensively with Dr. Vivian Hwa and Dr. Christina Jacobsen. Because much of the work presented in this project was collaborative, I have attempted to delineate my specific contribution to each of these chapters.

Chapter II:

Heather Carmichael: Performed analysis of whole exome sequencing data to select candidate genes. Wrote the initial draft of the manuscript and together with A.D., J.N.H. and Y.S. edited later drafts and submitted for publication.

Andrew Dauber: Recruited and selected patients with idiopathic short stature suitable for whole exome sequencing. Contributed to the writing of the draft and provided extensive feedback and edits to the manuscript.

Jennifer Moon: Performed confirmatory genotyping for select candidate variants.

Thutrang T. Nguyen: Performed confirmatory genotyping for select candidate variants and determination of uniparental disomy.

Yiping Shen: Performed CNV microarray analysis and provided the text on this section of the paper, as well as Figure 1.3.

Joel N. Hirschhorn: Provided extensive feedback and edits to the manuscript.

Chapter III:

Heather Carmichael: Performed analysis of potential pathogenic variants, including analysis and chart reviews for patients with potential mutations in IGF1R. Compared potential pathogenic variants in the cohort to a database (HGMD) of previously reported pathogenic

variants with identification of three subjects with Noonan Syndrome. Wrote the initial draft of the manuscript together with A.D. and S.R.W.

Sophie R. Wang: Performed analysis of pooled sequencing data including assessing false positive and negative rates. Together with H.C. and A.D, wrote the initial manuscript.

Shayne F. Andrew: Performed flow cytometry experiments for determination of IGF1R expression.

Vivian Hwa: Supervised functional studies for potential IGF1R mutations. Contributed significantly to writing these sections of the manuscript.

Andrew Dauber: Recruited and selected patients with idiopathic short stature suitable for inclusion in the targeted sequencing cohort. Together with J.N.H. determined height candidate genes for targeted sequencing. Contributed to the writing of the draft and provided extensive feedback and edits to the manuscript.

Michael A. Derr: Performed immunoblot experiments of IGF1R

Joel N. Hirschhorn: Together with A.D. determined height candidate genes for targeted sequencing. Provided extensive feedback and edits to the manuscript.

Chapter IV:

Heather Carmichael: Performed analysis of potential pathogenic variants in NPR2, including chart reviews and family pedigree analysis. Wrote the initial draft of the manuscript together with A.D., S.R.W., and C.M.J.

Sophie R. Wang: Performed screening and analysis of NPR2 variants in the additional short stature cohorts, including the height extreme cohorts. Together with A.D., H.C., and C.M.J, wrote the manuscript.

Christina M. Jacobsen: Performed functional studies for potential NPR2 mutations. Contributed significantly to writing the manuscript.

Andrew Dauber: Recruited and selected patients with idiopathic short stature suitable for inclusion in the targeted sequencing cohort. Contributed to the writing of the draft and provided extensive feedback and edits to the manuscript.

Aaron B. Edmund: Performed functional NPR2 assays and prepared data for manuscript.

Veronica Mericq: Recruited and phenotyped patients with idiopathic short stature

Timothy. C. Miller: Recruited and phenotyped patients with idiopathic short stature

Jennifer E. Moon: Performed confirmatory sequencing of NPR2 variants

Robert C. Olney: Recruited and phenotyped patients with idiopathic short stature

Lincoln R. Potter: Supervised functional NPR2 assays and revised these sections of the manuscript

Jerid W. Robinson: Performed functional NPR2 assays and prepared data for manuscript.

Matthew Warman: Supervised NPR2 sequencing and functional assays.

CHAPTER I:

Introduction

Idiopathic Short Stature and Skeletal Dysplasia

Growth is an important feature of normal development, and deviation from the norm can be a sign of underlying disease. In the pediatric endocrinology clinic, short stature is a commonly encountered problem. As height has been shown to be more than 90% heritable, it is not surprising that most children with short stature have a family history of short stature as well (1). However, in some cases, children have severe short stature that is out of proportion to parental heights. This severe short stature is often accompanied by other syndromic features, and the underlying biology in many cases is poorly understood. In these cases, children are often diagnosed with idiopathic short stature (ISS), even though some of them may have as yet unrecognized syndromes or skeletal dysplasias that include growth failure as a symptom (2).

Skeletal dysplasia is a general term used to describe abnormal growth or development of the skeleton, usually resulting in disproportionally short stature. Skeletal dysplasias are individually very rare, but collectively occur in 1/5000 live births and represent approximately 5% of all birth defects (3). These diseases cause significant morbidity due to the destruction of bone and cartilage caused by angular deformities, as well as other conditions such as spinal stenosis (4). The past years have shown a dramatic increase in the ability to distinguish between these individual dysplasias on the basis of clinical, radiographic, molecular, and genetic features (5, 6). Understanding the genetic, molecular, and developmental causes of skeletal dysplasias, as well as short stature more generally, represents an important first step in developing molecularly targeted therapies to treat this complex disease.

Indeed, the study of severe, rare monogenic conditions has in the past led to novel therapies that help treat a more widespread condition. For example, in 1939 Carl Muller demonstrated the genetic connection between elevated cholesterol and heart disease through the study of familial hypercholesterolemia, which was more extensively studied by Avedis Khachadurian in the 1960s. This genetic evidence provided in part the impetus for the search for pharmacologic agents designed to target cholesterol synthesis, and eventually led to the development of statins (7).

Genetic Basis of Height

The genetic basis of short stature is not completely understood. Stature is clearly a polygenic trait, with numerous variants within the human genome contributing incrementally to overall height. Early genome-wide association studies (GWAS) identified more than 180 loci associated with stature(8-11), but common variants within these loci explain only 10% of the variation in human height, and only 12% of heritability (1). More recently, the results of a GWAS meta-analysis for human height identified 697 variants within these loci that achieved genome-wide significance, and together explained approximately 20% of heritability (12). Additionally, they found that nearly 60% of heritability could be explained by common variants. These loci include the well-known pathways associated with growth, such as the growth hormone (GH) axis, but also highlight the many other pathways that mediate height (13, 14). Better genetic and molecular understanding of these unexplored loci could lead to new therapies and better understanding of skeletal dysplasias and stature more generally.

As previously described, up to 60% of heritability in height can be explained by common variants. However, there is still a substantial role for rare variants, particularly in patients with extreme short stature. A previous study considered individuals at the extremes of the height distribution (below the 1st percentile or above the 99th percentile) and analyzed known height-associated common variants in these individuals (15). They found that the effects of common variants in these extreme patients are generally similar to their effects in the general population, except in patients of extremely short stature (below the 0.25 percentile). In these patients, the effects of the common variants were much smaller than expected, indicating that additional factors influence height, particularly in those with extremely short stature. We hypothesize that these additional factors are possibly rare genetic variants, and that studying a population of patients with severe short stature (more than 3 standard deviations below the norm) enriches for such rare variants (16). Additionally, there is evidence that copy number variations—genetic deletions or duplications—contribute to short stature at the height extremes (17, 18).

Whole Exome Sequencing and Targeted Sequencing

Recently, the development of next generation DNA sequencing technologies has dramatically changed genetic discovery research (19). Whole exome capture and sequencing of affected individuals (and in some cases, unaffected family members) is one method that can be used to identify genes and mutations associated with rare diseases (20). Most rare, monogenic conditions are caused by mutations in the exons of the genetic code, that is, the regions of the genome that encode for proteins in the human body, and whole exome sequencing captures and sequences these specific sections of the genome.

A significant downside of whole exome sequencing is that it does not account for mutations that occur in non-protein coding regions of the genome, which could affect, among other things, regulation and splicing of gene products (21). However, whole exome sequencing is an efficient first step to identify potentially causal genetic mutations associated with a disease phenotype. The sequencing data can be compared with genomic databases to identify rare genetic variants, and also to other affected individuals or unaffected family members to reduce the number of candidate genes identified. The range of candidate genes can be further narrowed by considering what is already known about the encoded protein in order to develop a plausible story for how the mutated protein could lead to the patient phenotype. By understanding the etiologies of these rare syndromes, we can draw insights about the physiological mechanisms involved in patients with less severe forms of the disease (22).

Targeted sequencing is an additional approach that can be used to search for rare variants that might be causal in short stature. Unlike whole exome sequencing, this method selects for a targeted set of candidate genes prior to exon sequencing. This method relies on the effective selection of a group of candidate genes, as variants in other genes will go undetected. However, this method is currently considerably less expensive than whole exome sequencing, allowing for the study of a larger group of individuals with a disease phenotype, as is described in more detail in Chapter III.

Copy Number Variation

In addition to the recent development of new sequencing technologies, novel genotyping array technology allows for the rapid detection of genomic duplications and deletions, or copy number variants (CNVs). CNVs of chromosome 11q11, for example, have been associated with a phenotype of extreme and early onset obesity (23). GWAS of copy number variation have shown an increased burden of CNVs in populations with both autism and schizophrenia (24, 25). There is evidence for a similar increased burden of CNVs in patients with short stature. Thus, CNV analysis represents another method of genetic discovery that can lead to better understanding of the complex pathways involved in human growth.

Short Stature Cohort

Prior to my involvement in this research project, my research mentor, Dr. Andrew Dauber, recruited a cohort of patients with idiopathic short stature at the Endocrine and Genetics clinic at Children's Hospital Boston. The cohort had 292 patients as of August, 2011, and the research team collected DNA samples (blood and saliva) for these patients. All subjects were brought to the Clinical Translational Science Unit for an initial visit. At this time, the research team collected accurate anthropometric measurements such as height, weight, arm span, seated height, limb measurements, and pubertal staging, as well as a 3-D photograph of the patient. A full history was taken, including background ethnic data and height and pubertal data of all family members. My mentor also collected baseline measurements of IGF-1 (insulin like growth factor 1), IGF-BP3 (insulin like growth factor binding protein), acid labile subunit, and growth hormone binding protein, and also performed thyroid function tests and a bone x-ray for each patient in the study.

Prior to my involvement with this project, Dr. Dauber had selected from this cohort approximately 30 families with cases of severe idiopathic short stature accompanied by other distinctive syndromic features. These families also had pedigrees amenable to genetic investigation (such as consanguinity, or multiple affected and unaffected subjects). This allowed us to use features of the pedigree and the inheritance pattern to narrow the range of

possible variants that might be affecting patient phenotype. The research team performed whole exome sequencing on a subset of these patients. Another subset of this larger cohort was studied in an initial targeted sequencing project, which sequenced 1077 genes with a connection to short stature in these individuals.

Limitations

There are important limitations of the whole exome sequencing and targeted sequencing approaches that must be considered. First, this type of study assumes a priori that rare and previously undiscovered mutations in the exomes of these patients explain the extreme phenotypes observed. As mentioned earlier, this method does not detect variants that occur in non-coding regions of the genome, which can affect, among other things, transcriptional regulation of the gene. We did not consider most intronic variants, meaning we may have missed important variants that could affect transcriptional splicing. Thus, the criteria we used to narrow the large number of genetic variants present in each patient may have by design excluded important variants from further study. Furthermore, the targeted sequencing studies only consider an even smaller subset of genetic information to study for rare variants. However, we believe we used the selection criteria that have the highest likelihood of yielding causal variants, given the extreme nature of patient phenotypes in this cohort and the known association between the genes selected for targeted sequencing and short stature.

Project Goals and Description of Chapters to Follow

The specific aims of my project were to:

1. Define rare sequence variants (allele frequency less than 1%) that could contribute to abnormal growth in patients with severe short stature and syndromic features of unknown etiology

2. When possible, link implicated genes to a plausible developmental and physiological explanation of the patient's phenotype

Chapter II will describe the results of whole exome sequencing in one patient with severe short stature with other syndromic features who was determined to have uniparental disomy of

chromosome 2. Chapter III describes the methods of the pooled targeted sequencing project as well as some initial results including the identification of three patients with known mutations causal of Noonan syndrome as well as a patient with a novel insulin-like growth factor-1 receptor (*IGF1R*) mutation. Chapter IV describes the mutations found in this targeted sequencing cohort in the natriuretic peptide receptor-B (*NPR2*) gene, as well as several other short stature cohort, furthering the evidence that mutations in this receptor may represent a relatively frequent cause of familial idiopathic short stature.

CHAPTER II:

Whole Exome Sequencing in a Patient with Uniparental Disomy of Chromosome 2 and a Complex Phenotype

Heather Carmichael BA[1], Yiping Shen PhD[2,3], Thutrang T Nguyen BA[4,6],
Joel N Hirschhorn MD PhD[4,5,6,7], Andrew Dauber MD MMSc[4,5]

1. Harvard Medical School
2. Department of Laboratory Medicine, Boston Children's Hospital
3. Department of Pathology, Harvard Medical School
4. Division of Endocrinology, Boston Children's Hospital
5. Program in Medical and Population Genetics, Broad Institute
6. Center for Basic and Translational Obesity Research, Boston Children's Hospital
7. Department of Genetics, Harvard Medical School

This is the accepted version of the following article:

Carmichael H, Shen Y, Nguyen TT, Hirschhorn JN, Dauber A. Whole exome sequencing in a patient with uniparental disomy of chromosome 2 and a complex phenotype. Clinical genetics. 2013;84(3):213-22. doi: 10.1111/cge.12064. PubMed PMID: 23167750; PubMed Central PMCID: PMC3996682.

This has been published in final form at:

http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3996682/

Abstract

Whole exome sequencing and chromosomal microarrays are two powerful technologies that have transformed the ability of researchers to search for potentially causal variants in human disease. This study combines these tools to search for causal variants in a patient found to have maternal uniparental isodisomy of chromosome 2. This subject has a complex phenotype including skeletal and renal dysplasia, immune deficiencies, growth failure, retinal degeneration, and ovarian insufficiency. Eighteen nonsynonymous, rare homozygous variants were identified on chromosome 2. Additionally, 5 potential compound heterozygous mutations were detected on other chromosomes that could lead to a disease phenotype independent of the uniparental disomy found in this case. Several genes with potential connection to phenotype are described; no gene stood out as a likely candidate to follow up with functional analysis. This study highlights the potential for detection of a large number of candidate genes using whole exome sequencing complicating interpretation in both the research and clinical settings. Forums must be created for publication and sharing of detailed phenotypic and genotypic reports to facilitate further biological discoveries and clinical counseling.

Introduction

Recently, the development of next generation DNA sequencing as well as technologies to examine genome wide copy number variation have dramatically changed genetic discovery research (26, 27). Two related technologies—whole exome sequencing and array comparative genomic hybridization (CGH) —are also being utilized by clinicians with increasing regularity and will soon be part of daily practice (28). These tools offer great promise in identifying causal mutations and discovering novel biological pathways. However, the amount of information provided by these types of studies can greatly complicate their interpretation, particularly in the clinical setting. Here, we illustrate the challenges of interpreting exome sequencing results in a patient with an undiagnosed and complex medical presentation who was found to have uniparental disomy (UPD) of chromosome 2.

Case Report

The proband is currently a 21 year old female with multiple medical problems. Her parents are nonconsanguineous and are from the Dominican Republic. Severe IUGR was noted at 26 weeks gestation and the patient was born at 35 weeks gestation by Caesarian section with a birth weight of 970 g (-3.57 SDS). Early years were notable for mild global developmental delay; she sat at 12 months, walked at 24 months, had her first words at 18 months, and spoke 2-3 words at 2.5 years. She graduated high school with special education classes and is currently seeking vocational training. She was noted to have a number of dysmorphic features including multiple skeletal anomalies leading to suspicion of a skeletal dysplasia (Figure 2.1). Her hands are notable for brachydactyly, with a palm length of 8.9 cm (50th centile for a 10 year old) and a middle finger length of 5.5 cm (50th centile for a 6 year old). She also has mild clinodactyly of the 2nd and 5th fingers, some soft tissue syndactyly of the 2nd-4th fingers, and an accessory ossification center at the base of her 2nd metacarpals bilaterally. She also has pes planus, widely spaced 1st and 2nd toes, and mild syndactyly of the 2nd and 3rd toes. She carries a diagnosis of Klippel Feil syndrome with fusion of C2 and C3, shortening of C4 and C5, and partial fusion of C6 and C7 (Figure 2.2). Additionally, she has a dysplastic right kidney.

During childhood, she had recurrent sinusitis, upper respiratory infections, and pneumonias and was diagnosed with hypogammaglobulinemia with low serum levels of all immunoglobulins (IgG, IgM and IgA). She receives intravenous immunoglobulin replacement. She also had frequent herpes stomatitis which led to a diagnosis of depressed natural killer cell toxicity on functional testing.

During evaluation for poor growth, at 12 years old, the subject was found to have growth hormone deficiency with a peak growth hormone level of 1.2 ng/mL to insulin induced hypoglycemia and glucagon stimulation. Concurrently, her peak cortisol level was 12.8 ng/mL leading to a diagnosis of central adrenal insufficiency. A brain MRI showed a low volume pituitary. Growth hormone therapy was given for four years with little overall improvement in her growth velocity. She reached a final adult height of 135 cm (SDS -4.26).

At 12 years and immediately following the initiation of growth hormone treatment, the patient developed diabetes mellitus. She was found to have a normal level of insulin antibodies at 20.8 nU/mL (nl <39 nU/mL), negative insulinoma antigen 2 (IA2) antibodies, but a borderline positive glutamic acid decarboxylase antibody (GAD) at 0.136 units (nl <0.1). She was started on insulin therapy which has been continued to date. The exact etiology of her diabetes is unknown.

At 16 years, she had no breast development or onset of menses, and testing revealed high luteinizing hormone and follicle-stimulating hormone levels of 31.05 IU/L and 75.34 IU/L respectively. Estradiol was low at 11.3 pg/mL consistent with premature ovarian insufficiency leading to hypergonadotrophic hypogonadism. She was started on estrogen replacement therapy resulting in pubertal development.

At 10 years, she was noted to have astigmatism but with excellent corrected vision. At 18 years, the patient began to show signs of retinal degeneration, with attenuated retinal vessels bilaterally and an increased cup to disc ratio. At 19 years, she had inconsistent complaints of visual constriction and poor night vision. She has some symptoms consistent with retinitis pigmentosa (RP), including findings of ring scotoma and attenuated vessels bilaterally, rod and cone degeneration, and a bone spicule in the right eye only. However, some important features of RP such as pallor of the optic nerve and peripheral pigment changes are absent, and no definitive diagnosis has been made. There are no signs of diabetic retinopathy.

The patient is overweight as an adult, with a BMI of 29 kg/m^2. At 17 years, she presented with hypercholesterolemia with a total fasting cholesterol of 210 mg/dL and LDL of 140 mg/dL. Her cholesterol levels have remained in the borderline range as an adult. At 19 years, she presented with microalbuminuria with urine protein levels ranging from 89.4 – 411.7 mg/g creatinine (nl <20 mg/g), now managed with lisinopril.

Karyotype was normal female 46, XX. There was no DiGeorge/VCFS deletion detected by FISH. Family history is notable for type 2 diabetes and hypercholesterolemia in her mother but is otherwise unremarkable.

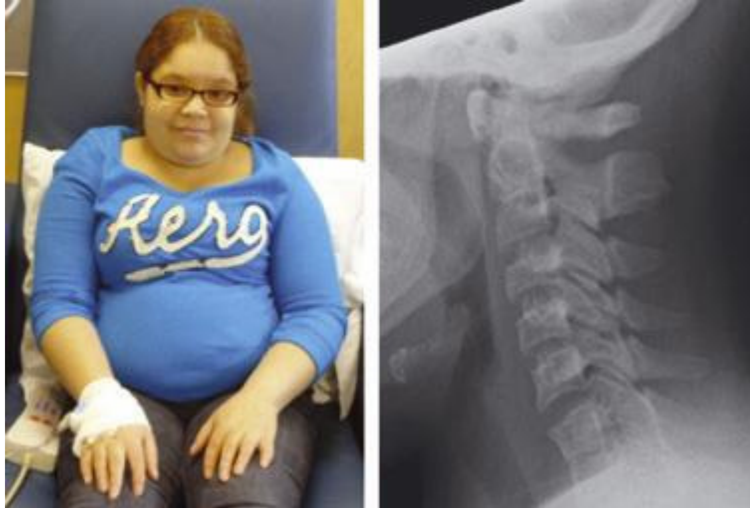


Figure 2.1: Cervical abnormalities. Radiographic image of the cervical spine showing fusion of C2 and C3, shortening of the vertebral bodies of C4 and C5, and partial fusion of C6 and C7. The patient carries a diagnosis of Klippel Feil syndrome.

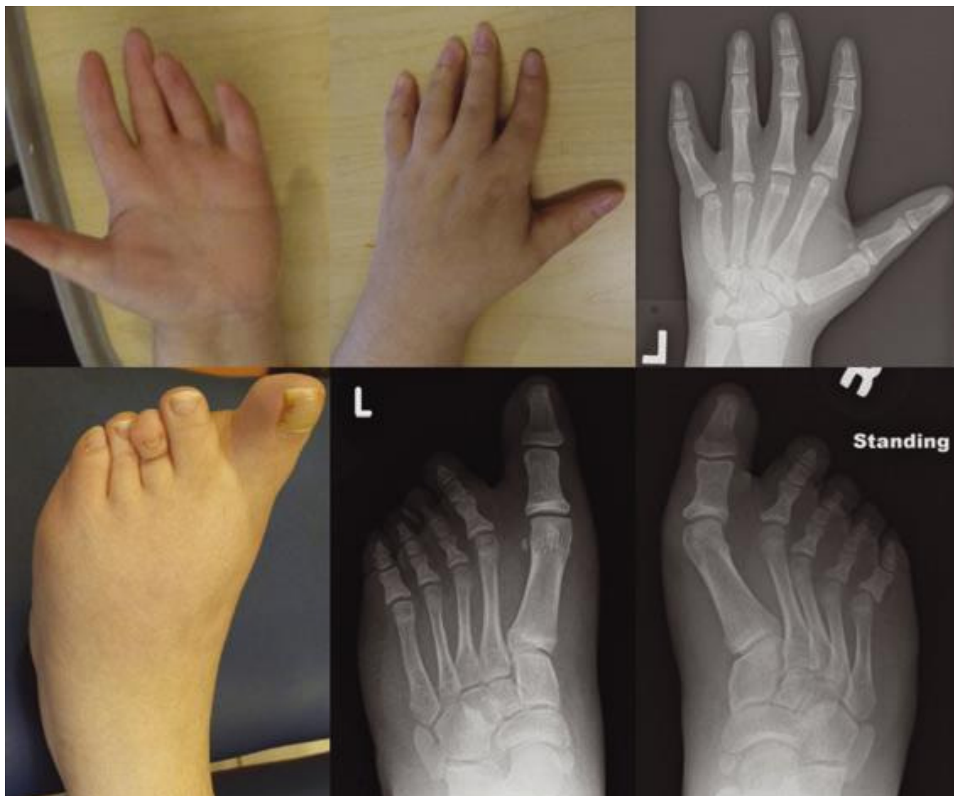


Figure 2.2: Upper and lower digits. Upper panel shows images of the dorsal and palmar surfaces as well as a radiographic image of the left hand. Left hand is notable for brachydactyly and clinodactyly, as well as some soft tissue syndactyly of the 2nd to 4th digits. Lower panel shows left dorsal foot and radiographic images of both feet. The patient has pes planus, widely spaced 1st and 2nd toes, and mild syndactyly of the 2nd and 3rd toes.

Materials and Methods

This study was approved by the institutional review board at Boston Children's Hospital. Written informed consent was obtained from the subject and family members.

The DNA sample was analyzed using Agilent SurePrint G3 custom comparative genomic hybridization and SNP 4x180K microarray (Agilent Technologies, Santa Clara, California), following the manufacturer's instructions. This microarray platform contains 150,000 oligonucleotide probes for the detection of copy number variants (CNVs) and 30,000 probes for the detection of single nucleotide polymorphisms (SNPs). CNV probes were densely populated in genes of interest (detection sensitivity is less than 10 Kb). For the remaining regions of the genome, CNV probe coverage provided detection sensitivity of about 400Kb. SNP probes were evenly distributed across the whole genome; the detection sensitivity for loss of heterozygosity is less than 5 Mb.

Whole exome sequencing of the proband was performed at the Broad Institute. Hybrid selection was performed using Agilent's SureSelect Human All Exon Kit v2 (Agilent Techonologies, Santa Clara, CA). The sample was sequenced using the Illumina HiSeq 2000 platform (Illumina Inc., San Diego, CA). We aligned the resulting reads to the hg19 reference genome with BWA(29), applied GATK(30) base quality score recalibration, indel (insertion-deletion) realignment, and performed SNP (single nucleotide polymorphism) and indel discovery and genotyping using variant quality score recalibration (31). Exome sequencing yielded a total of 72,022,554 reads with 93% passing quality filters. Mean target coverage was 74 reads with 79% of target having 30X coverage, 87% having 20X coverage, and 93% having 10X coverage. Variants were annotated for functional effect using SnpEff 2.0.5 (http://snpeff.sourceforge.net/). Allele frequency data was obtained from the 1000 Genomes project (February 2012 release) (32) and the NHLBI exome variant server (33).

Confirmation of rare variants was performed using individual SNP genotyping or traditional Sanger sequencing in cases of genotyping failure. Individual SNP genotyping was performed at Boston Children's Hospital using Sequenom iPLEX genotyping (Sequenom, Inc, San Diego,

CA, USA). All candidate variants were genotyped in the proband, her parents, and sister as well as 90 HapMap control samples. All passing SNPs had genotype success rates of >90%.

Results

CNV microarray with SNP probes did not reveal any rare copy number variants. However, results suggested that this patient had uniparental disomy of chromosome 2 (UPD2) (Figure 2.3). Data from whole exome sequencing further demonstrates that this patient has near complete UPD of this chromosome. We detected a total of 44,813 variants in the patient's exome. Of these, 37.2% of variants showed homozygosity for the allele present. However, on chromosome 2, 95.0% of variants showed homozygosity for the allele present, demonstrating the ability of exome sequencing to identify regions with potential UPD.

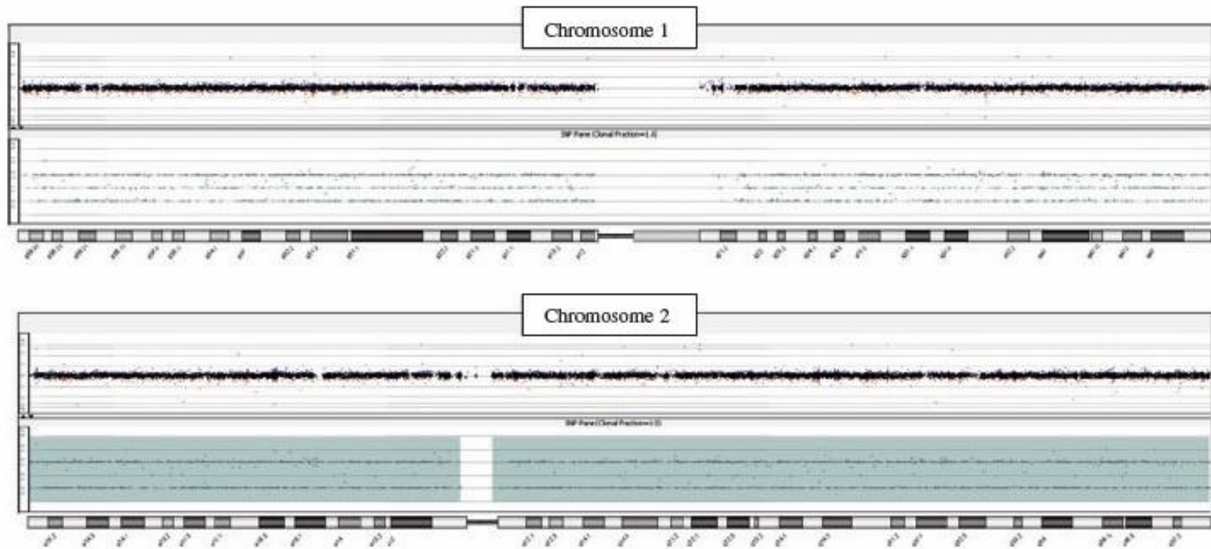


Figure 2.3: Microarray results for chromosome 2. The upper panel displays the CNV scatter plot showing that no CNV was detected on chromosome 2 (black dots indicate Cy5/Cy3 ratio between 0.25 and -0.25; blue dots indicate Cy5/Cy3 ratio > 0.25 and red dots indicate Cy5/Cy3 ratio <-0.25). The bottom panel displays the zygosity plot. In the zygosity plot, the top and bottom tracks indicate homozygous status while the middle track indicates heterozygous status. The paucity of heterozygous SNP across the whole chromosome in this sample indicated the whole chromosome loss of heterozygosity. The absence of loss of heterozygosity in the rest of the genome and the copy neutral status of the chromosome 2 indicate the UPD status of chromosome 2 in this subject.

Given the UPD2, we hypothesized that the subject’s phenotype is most likely due to rare homozygous functional variants on chromosome 2. Whole exome sequencing identified 44,813 variants including both single nucleotide variants (39,744) and indels (5,069). Of these total

variants, there were 16,678 present in the homozygous state. Table 2.1 summarizes the breakdown of total variants detected by functional classification and allele count, as well as a summary of variants located on chromosome 2 only. For analysis of candidate variants, all variants that were present at a minor allele frequency (MAF) greater than or equal to 1% in the 1000 Genomes project, the NHLBI exome variant server, or in 50 HapMap control exomes were excluded leaving 60 rare homozygous variants. Of these, there were 18 nonsynonymous variants in 17 candidate genes, all of which are located on chromosome 2 (Table 2.2). All 18 variants were validated via confirmatory genotyping and were inherited from the proband's mother indicating that she has maternal UPD2. One of these 18 variants is a predicted splice site variant and the others are missense variants. One rare missense variant (MAF = 0.0027), was present in a gene implicated in autosomal recessive retinitis pigmentosa, *FAM161A*, which is a candidate to explain the patient's retinal dystrophy. This variant leads to a L378R substitution in the coded protein which PolyPhen2 (34) predicts is highly likely to be damaging. Although the exact function of this protein remains unknown, truncating mutations in *FAM161A* are a known autosomal recessive cause of retinitis pigmentosa (35), a heterogeneous disease of retinal degeneration characterized by retinal pigment deposits, rod and cone degeneration, and loss of vision (32, 35).

An additional two homozygous variants are located in genes related to kidney structure and function possibly contributing to the patient's history of microabuminuria and dysplastic kidney. We identified a rare missense variant (MAF = 0.0009) in *NAT8* that leads to a W60S substitution which is predicted by PolyPhen2 to likely be deleterious. The subject also has a rare splice donor site variant (MAF = 0.0014) between exons 21 and 22 in *PLA2R1* in a site that is highly evolutionarily conserved.

Table 2.1: Summary of variants detected through exome sequencing.

All Variants		
	Whole Exome	**Chromosome 2**
Total Variants	44813	2101
% Homozygous	37.2%	95.0%
Nonsynonymous SNP	*8980*	*422*
Splice Site	*127*	*3*
Frameshift	*214*	*6*
In-Frame Coding Indel	*328*	*12*
Start Gained	*142*	*7*
Start Lost	*25*	*2*
Nonsense	*59*	*0*
Runthrough	*14*	*0*
Synonymous SNP	*10625*	*481*
Intron	*14711*	*713*
Other Non-Coding	*9588*	*455*

MAF < 0.01, not found in HapMap controls			
	Whole Exome		**Chromosome 2**
All Functional Classes	1707		58
	Homozygous	**Heterozygous**	**Homozygous**
Total Variants	60	1647	54
Nonsynonymous SNP	*17*	*468*	*17*
Splice Site	*1*	*3*	*1*
Frameshift	*0*	*9*	*0*
In-Frame Coding Indel	*0*	*12*	*0*
Start Gained	*0*	*5*	*0*
Start Lost	*0*	*2*	*0*
Nonsense	*0*	*11*	*0*
Runthrough	*0*	*0*	*0*
Synonymous SNP	*12*	*406*	*9*
Intron	*21*	*421*	*19*

Table 2.2: Rare homozygous variants. Variant location is reported using hg19 coordinates. Minor allele frequency data is based on the overall frequency in the 1000 Genomes February 2012 release. Information on gene function was summarized from information available in the Uniprot and Online Mendelian Inheritance in Man (OMIM) databases. OMIM was also used to identify any known disease associations, and mouse knockout phenotypes were collected from the Mouse Genome Informatics (MGI) database. For missense variants, the validated PolyPhen2 prediction tool (34) was used to predict the effect of each amino acid substitution, with 0 being least and 1 being most likely to disrupt structure or function of the final protein.

Gene	Position (hg19)	Ref	Alt	Frequency (1000 Genomes)	Missense	PolyPhen2 Prediction	Associated Disease (OMIM)	Mouse Knockout Phenotype (MGI)	Function (Uniprot)
CAD	2:27445809	G	A	N/A	R238H	0.209			Pyrimidine biosynthesis, cell proliferation
SULT6B1	2:37406661	G	C	0.0018	P157A	0.995			Sulfotransferase involved in thyroxine metabolism
SMEK2	2:55795486	A	T	0.0005	F593L	0.494			Regulatory subunit of serine/threonine protein phosphotase 4 (PP4) involved in hepatic glucose metabolism
FAM161A	2:62067006	A	C	0.0027	L378R	0.985	Retinitis pigmentosa 28		Unknown, expressed in the retina
NAT8	2:73868577	C	G	0.0009	W60S	0.992			Unknown, may play a role in the development and maintenance of normal kidney and liver structure
YSK4 (variant transcript)	2:135740813	T	C	N/A	Q1121R	0.573			Unknown, serine/threonine protein kinase
DARS	2:136673814	C	T	0.0005	G363E	0.347			Aminoacyl-tRNA synthetase
NEB	2:152522801	G	A	N/A	R1612C	0.999	Nemaline myopathy, type 2	Stunted growth, muscle weakness, postnatal death	Large protein component of the sarcomere
ARL6IP6	2:153575472	C	T	0.0009	L112F	0.004			Unknown, interacts with ARL6

Gene	Position (hg19)	Ref	Alt	Frequency (1000 Genomes)	Missense	PolyPhen2 Prediction	Associated Disease (OMIM)	Mouse Knockout Phenotype (MGI)	Function (Uniprot)
PLA2R1	2:160803929	C	T	0.0014	splice site	N/A		Viable and fertile with no overt abnormalities	Receptor for phospholipase A2, plays a role in MAPK-induced cell proliferation
TTN	2:179399863	G	A	N/A	R32186C	0.999	Cardiomyopathy, muscular dystrophy	Cardiac and skeletal defects, growth retardation, muscle weakness, prenatal death	Large protein component of the sarcomere
	2:179638834	C	T	0.0032	R2354H	1.000			
COL5A2	2:189916110	C	G	0.0018	R956P	0.155	Ehlers-Danlos syndrome, type 1	Reduced and abnormal bone growth, fragile skin, perinatal lethality	Collagen V component
SDPR	2:192711244	G	C	0.0009	C136W	0.981			Phospholipid binding protein, PKC substrate
SLC39A10	2:196545348	C	A	0.0032	H194Q	0.010			Unknown, possible zinc influx transporter
HECW2	2:197208381	G	A	0.0032	P134L	0.001			E3 ubiquitin-protein ligase that ubiquitinates and stabilizes TP73
NBEAL1	2:204009459	A	T	0.0027	Y1633F	0.997			Unknown, has structural similarity to neurobeachin
PSMD1	2:231948431	G	A	<0.0001	R559K	0.103			Regulatory subunit of the proteosome

Table 2.3: Potential rare compound heterozygous variants. Annotations are as in Table 2.2. Potential compound heterozygous mutations were genotyped in both parents to determine which parent is a carrier.

Gene	Position (hg19)	Ref	Alt	Frequency (1000 Genomes)	Missense	PolyPhen2 Prediction	Inheritance	Associated Disease (OMIM)	Mouse Knockout Phenotype (MGI)	Function
NPHP4	1:5924483	T	C	0.0023	H1304R	0.006	Paternal	Nephronophthisis type 4, Senior-Loken syndrome type 1	Photoreceptor degeneration, male infertility	Involved in the organization of apical junctions in kidney cells
	1:5965750	G	C	0.0032	Q569E	0.604	Maternal			
	1:6027365	C	T	0.0027	A171T	0.512	Maternal			
SLC6A20	3:45804541	C	A	N/A	A443S	0.428	Maternal	Hyperglycinuria, iminoglycinuria (digenic)		Transporter of small hydrophilic molecules, in kidney and intestine
	3:45821529	G	T	N/A	A109D	0.428	Paternal			
ADAMTS12	5:33624365	C	T	0.0032	R705K	0.000	Paternal		Increased tumor vascularization and invasion	Metalloproteinase that degrades cartilage oligomeric matrix protein
	5:33881583	GCT	G	N/A	N/A	N/A	Maternal			
CD109	6:74440187	A	G	N/A	I133V	0.929	Paternal			Negatively modulates TGF-beta signaling in keratinocytes
	6:74495175	T	C	0.0014	F771S	0.971	Maternal			
OR4D1	17:56232873	A	G	0.0032	Y120C	0.999	N/A			Olfactory receptor
	17:56232977	C	T	0.0041	H1554	1.000	Maternal			

Although the presence of UPD2 is the most obvious potential cause of some or all of this patient's clinical picture, some patients with UPD2 have been reported with normal phenotypes, suggesting that some of the findings in this patient could be attributable to mutations on other chromosomes. Although it is possible that her phenotype was caused by a de novo mutation, determining this would involve whole exome sequencing of both of her parents, which was beyond the scope of the current study. However, we did search for potential compound heterozygous variants on all chromosomes (there were no rare homozygous nonsynonymous variants detected outside of chromosome 2). We identified 20 genes with two or more rare (MAF < 0.01) nonsynonymous heterozygous variants. Eight of these variant pairs are unlikely to be true compound heterozygotes as the minor allele frequencies are nearly identical, suggesting that these variants are in linkage disequilibrium and most likely present on the same chromosome. We performed confirmatory genotyping and familial segregation of the potential compound heterozygous variants in the remaining 12 genes. The proband's unaffected sister was also a carrier of both variants in 7 of these genes ruling them out as candidates. The remaining 5 genes are summarized in Table 2.3. One of these candidate genes, *NPHP4*, is known to cause a ciliopathy, a category of diseases that share some overlap with our subject's phenotype.

Discussion

We report a subject with a complex phenotype including skeletal and renal dysplasia, immune deficiencies, growth failure, retinal degeneration, and ovarian insufficiency. To our knowledge, this constellation of features has not been previously described in the literature. Chromosomal microarray revealed UPD2 which could lead to her phenotype through three primary mechanisms: imprinting, homozygous copy number variants, or unmasking of rare pathogenic sequence variants. There are 19 cases of UPD2 reported in the literature. Importantly, cases of both maternal and paternal UPD2 have been reported in individuals with normal phenotypes, suggesting that imprinting is an unlikely mechanism by which UPD is causing this patient's phenotype (36, 37). Array CGH did not reveal any CNVs on chromosome 2, excluding the potential role of structural variation, to the degree that can be detected by the array (>10 Kb), in explaining the patient's phenotype. Finally, UPD can lead to homozygosity

of rare recessive mutations on the involved chromosome (38). Many cases of UPD2 resulting in the unmasking of rare recessive mutations are reported in the literature, including cases of Crigler-Najjar syndrome, Donnai-Barrow syndrome, severe congenital hypothyroidism, Harlequin ichthyosis, infantile-onset ascending spastic paralysis, long-chain 3-hydroxyacyl-CoA dehydrogenase deficiency, and mitochondrial DNA depletion syndrome (39-44). None of these subjects resemble the phenotype found in our proband, or any of its features, with the exception of one individual with retinal dystrophy believed to be caused by a recessive mutation in *MERTK* (45). Our subject did not have a rare variant in this gene, or any genes identified as candidate genes in prior studies of UPD2. We therefore searched for rare recessive mutations in other genes on chromosome 2.

Exome sequencing highlighted 18 rare nonsynonymous homozygous variants in 17 genes, all of which were on chromosome 2 (Table 2.2). Several candidate genes stood out in our analysis as potentially causal. Our subject has retinal degeneration with some features of retinitis pigmentosa although does not have a definitive diagnosis, and one potential explanation for this phenotype is the variant found in *FAM161A*. Four of the five known pathogenic mutations in this gene are found in exon 3, as is the mutation in our patient. Therefore, it is possible that this variant explains this one aspect of her phenotype. However, all of the known mutations are truncating or frameshift mutations, and it is unclear that our patient's missense variant is truly pathogenic as opposed to an incidental finding of our exome sequencing approach.

Additionally, rare variants were identified in two genes with potential functional effects in the kidney, *NAT8* and *PLA2R1*, which could explain our patient's unilateral dysplastic kidney as well as early onset microalbuminuria. *NAT8* encodes a probable N-acetyltransferase highly expressed in the kidney and liver. It is hypothesized to play a role in the development and maintenance of the structure and function of these tissues, but its biological role is not fully understood (46). *PLA2R1* encodes for a receptor of phospholipase A2 with high levels of expression in the kidney (47). Binding of the ligand leads to activation of MAPK signaling and induces cell proliferation (48). However, it is unclear by what mechanism a defect in *PLA2R1* could cause the phenotype observed in this patient, and mouse knockouts for *PLA2R1* show no

overt abnormalities (49). Again, these variants could be pathogenic, but could also be incidental findings.

Finally, the subject has some features consistent with a ciliopathy, a disease category which is associated with defects in proteins that play a role in the structure and function of cilia. Ciliary dysfunction can lead to a number of clinical syndromes with overlapping features, including Meckel-Gruber syndrome, Bardet-Biedl syndrome, Joubert syndrome, nephronophthisis, Senior-Loken syndrome, and Jeune asphyxiating thoracic dystrophy (50). In particular, the phenotype of our patient resembles Bardet-Biedl syndrome (BBS), which has characteristic features of rod and cone dystrophy, postaxial polydactyly, renal anomalies, obesity (with complications of type II diabetes, hypertension, and hypercholesterolemia), learning disabilities, and hypogonadism (51). While certain aspects of her phenotype are not encompassed by BBS, such as her hypogammaglobulinemia and growth hormone deficiency, it seems plausible that such a diagnosis could at least partially explain the abnormalities we observed.

We found 3 rare mutations (Table 2.3) in one ciliary gene, *NPHP4*, which is implicated in two ciliopathies, Senior-Loken syndrome and nephronophthisis (52). Phasing indicates that one of these three variants was inherited paternally, and her sister is not a carrier. The mother and sister are both carriers of the other two variants. Thus, it is plausible that a combination of two of these variants contributed to the phenotype observed in the patient. Additionally, the patient has a rare missense variant in the heterozygous state in *MKS1*, a gene implicated as causal in Meckel-Gruber syndrome as well as in approximately 4.5% of patients with BBS (51, 53). This gene is located on chromosome 17 and encodes a protein involved in migration of the centrosome to the apical cell surface during ciliogenesis that has been shown to play an important role in regulating the length and number of ciliary structures (54, 55). Finally, we found that the subject has a rare homozygous mutation in *ARL6IP6* (Table 2.2), a gene on chromosome 2 which encodes a membrane protein with no known function. However, this protein was identified because of its interaction with *ARL6*, a causative gene in BBS (51, 56). Ciliopathies like BBS have been proposed to be oligogenic, with evidence for modifying alleles contributing to penetrance and variability in clinical outcome (50, 57, 58). Thus, it could be the

case that this mutation in *ARL6IP6*, in combination with the mutations in *NPHP4* and/or *MKS1*, is contributing to a BBS-like phenotype.

Array CGH and whole exome sequencing are powerful research tools which are revolutionizing researchers' ability to identify candidate genes in disease phenotypes. These tools are increasingly being utilized in the clinical setting, leading to difficulty in the clinical interpretation of complex results. We present a prismatic case in which array CGH and whole exome sequencing were combined to search for causal genes in a complex patient found to have UPD2. The many variants of unknown significance in this case highlight the difficulty that clinicians will experience as genetic technologies gain more widespread use in everyday practice. The large number of candidate genes identified in this study is in part due to the loss of heterozygosity as a consequence of isodisomy. However, in addition to the 18 rare, homozygous variants on chromosome 2, we also identified 12 similarly rare potential compound heterozygous mutations that could lead to clinical effects independent from the patient's UPD.

In contrast to prior studies of UPD2, we were not able to identify a single candidate gene that could explain the subject's complete, complex phenotypic picture. We highlight a number of potentially causal genes including *FAM161A*, a known cause of retinitis pigmentosa, which could explain this patient's retinal degeneration. Additionally, *NAT8* and *PLA2R1* may play a role in her renal dysplasia and microalbuminuria. Finally, many of her features are consistent with a ciliopathy, such as BBS, and she has a number of rare nonsynonymous variants in BBS-related genes. However, we also found rare mutations in many genes that are unlikely to be involved in her phenotype, such as *TTN*, which encodes for a large component of the sarcomere, and *OR4D1*, an olfactory receptor. We additionally failed to locate a likely cause of certain aspects of her phenotype, such as her immune deficiencies and her growth hormone deficit. It is possible that these features of her phenotype are caused by one of the variants we detected, where the causal biology is not yet obvious. Alternatively, these features could be due to differences that are not detected by whole exome sequencing, such as variants in non-coding regulatory regions or epigenetic changes.

The high rates of rare variant identification by whole exome sequencing and microarray analysis will make it difficult to distinguish the causal variants from background rare but non-causal variants for any single case. We believe that it is imperative that researchers and clinicians publish similarly complex cases with detailed phenotypic and genotypic data, even when a causal gene has not yet been established, to permit the identification of causal genes that are implicated in multiple cases with phenotypic overlap. Such reports of genotype-phenotype correlations will empower future biological discoveries as well as assist clinical geneticists and genetic counselors in the interpretation of exome sequencing and array CGH results.

CHAPTER III:

Large Scale Pooled Next-Generation Sequencing of 1077 Genes to Identify Genetic Causes of Short Stature

Sophie R. Wang[1,2,3]*, Heather Carmichael[4]*, Shayne F. Andrew[5], Timothy C. Miller[2], Jennifer E. Moon[2], Michael A. Derr[5], Vivian Hwa[5], Joel N. Hirschhorn[1,2,3], Andrew Dauber[2,3]

1. Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115
2. Division of Endocrinology, Boston Children's Hospital, Boston, Massachusetts 02115
3. Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts 02142
4. Harvard Medical School, Boston, Massachusetts 02115
5. Department of Pediatrics, Oregon Health and Science University, Portland, Oregon 97239

*These authors contributed equally.

This is the accepted version of the following article:

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. The Journal of clinical endocrinology and metabolism. 2013;98(8):E1428-37. doi: 10.1210/jc.2013-1534. PubMed PMID: 23771920; PubMed Central PMCID: PMC3733853.

This has been published in final form at:

http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3733853/

Abstract:

Context: The majority of patients presenting with short stature do not receive a definitive diagnosis. Advances in genetic sequencing allow for large-scale screening of candidate genes potentially leading to genetic diagnoses.

Objectives: To discover genetic variants that contribute to short stature in a cohort of children with no known genetic etiology.

Design: Prospective cohort study of subjects with short stature

Setting: Pediatric endocrinology and genetics clinics at an academic center

Patients: 192 children with short stature with no defined genetic etiology as well as 192 individuals of normal stature from the Framingham Heart Study

Intervention: Pooled targeted sequencing using next-generation DNA sequencing technology of the exons of 1077 candidate genes

Main Outcome Measures: Number of rare nonsynonymous genetic variants found in cases but not controls, known pathogenic variants in cases, and potentially pathogenic variants in *IGF1R*

Results: We identified 4928 genetic variants in 1077 genes that were present in cases but not in controls. Of those, 1349 variants were novel (898 nonsynonymous). False positive rates from pooled sequencing were 4-5% and false negative rate was 0.1% in regions covered well by sequencing. We identified three individuals with known pathogenic variants in *PTPN11* causing undiagnosed Noonan syndrome. There were 9 rare potentially nonsynonymous variants in *IGF1R*, one of which is a novel, likely pathogenic, frameshift mutation. A previously reported pathogenic variant in *IGF1R* was present in a control subject.

Conclusions: Large-scale sequencing efforts have the potential to rapidly identify genetic etiologies of short stature but data interpretation is complex. Noonan Syndrome may be an underdiagnosed cause of short stature.

Introduction

Growth is a fundamental biological process that occurs during childhood. With the exception of diabetes, short stature is one of the most common reasons for referral to a pediatric endocrinologist. In most cases, short stature is familial, consistent with a strong genetic influence on childhood and adult height. In some cases, however, children have severe short stature that is out of proportion to the parental heights, or have short stature associated with syndromic features. Molecular defects associated with these rarer cases have, over the last decade, expanded the list of genes and biological pathways known to influence growth. Multiple mutations, for example, have been found in the growth hormone (GH) pathway, not only in the growth hormone (*GH1*) gene itself, but downstream, within the growth hormone receptor (*GHR*), *STAT5B*, *IGF1*, *IGFALS*, and the *IGF1R* (IGF-1 receptor) genes (5). Many genes underlying severe skeletal dysplasias associated with short stature have also been identified (13). Despite these advances, the molecular causality for the vast majority of patients, including those with severe or syndromic short stature, remain unresolved. Consequently, most affected patients continue to be classified as having idiopathic short stature (ISS).

Genome-wide association (GWA) studies have enabled the identification of common genetic variants (frequencies above 5%) influencing quantitative traits such as height. Indeed, recent GWA studies identified 180 genetic loci with common DNA sequence variants that influence human stature (11). Intriguingly, these common variants are often located in or near genes that underlie syndromes of abnormal skeletal growth. This overlap suggests that rare variants in other genes highlighted by the GWA studies could have significant impacts on growth.

To explore the role of rare genetic variants in short stature, we developed and applied large scale candidate gene sequencing technologies (31) in a cohort of children with short stature of unknown cause. The selected list of 1077 candidates is comprised of genes from identified GWA loci, genes known to cause syndromic short stature, and genes known to be involved in growth plate biology or growth plate signaling. Herein, we report our initial screen and assessment of pooled exonic sequencing on DNA samples from 192 children with short

stature and 192 controls of normal stature. We identified a large number of nonsynonymous variants present in cases but not controls. There are a number of possible analytical approaches to explore this data. First, one can search for variants that have previously been reported to be pathogenic. Second, one can search for novel variants within genes known to cause short stature and then perform further familial segregation and functional studies to validate those variants. Third, one can search for multiple likely deleterious variants in novel genes not previously known to cause short stature. In the current manuscript, we will discuss the first approach, looking for known pathogenic variants, as well as provide a more detailed analysis of rare genetic variants identified in *IGF1R* as an example of the second approach. Haploinsufficiency of *IGF1R* is known to cause significant short stature and our data demonstrate the utility of large scale sequencing and the critical need for careful interpretation of resulting data. Future work will explore the other analytical approaches.

Materials and Methods

Height candidate genes

In this study, we sequenced the exons of 1077 genes (~2Mb total target size). Of these 1077 genes, a third (N=356) were known biological candidates, including genes known to underlie syndromic growth disorders or skeletal dysplasias as well as genes involved in growth plate biology or growth hormone signaling. The remaining two-thirds (N=777) included genes within genomic loci associated with height based on genome-wide association studies; 56 genes belong to both categories (11). For the genes within the GWA loci, we set the genomic boundaries at each height-associated locus using linkage disequilibrium cutoffs (HapMap CEU r^2>0.5) for the top SNP. For loci with ≥2 genes within the genomic boundary, all genes were included. Loci with more than 10 genes were excluded. For SNPs with <2 genes within the genomic boundary, genes beyond the boundary but within the next recombination hotspots were included.

Subjects

This study was approved by the institutional review board at Boston Children's Hospital. All subjects or their legal guardians provided written informed consent. The 192 patients with short stature (>2 SDS below the mean for age and gender (59)) but without defined genetic etiologies, were recruited from the Endocrinology and Genetics clinics at Boston Children's Hospital. As we were searching for rare genetic syndromes, subjects were allowed to have additional medical comorbidities, dysmorphic features, or other hormonal deficiencies as long as these alternate medical problems did not provide a clear explanation for the subject's short stature. Additionally, 192 controls were chosen from Framingham Heart Study (FHS). Controls were chosen from the middle of the FHS height distribution (height Z-scores between -0.7 and +0.7 SDS). Z-scores were calculated by regressing the height phenotype stratifying by gender and adjusting for age.

Sequencing Protocol

DNA samples from multiple subjects were pooled for DNA sequencing using previously described methods available at the Broad Institute (31). In order to identify variants present only in a single individual (hereon referred to as singleton variants), we applied a simple overlapping pooling design. The samples from short stature subjects and control samples were each arranged into a 14x14 matrix of 28 pools, with 13-14 samples in each pool. Four empty "holes" were included in each matrix for assessing false positive rate. Each sample was sequenced in two pools (one row pool and one column pool). Singleton variants appear only in one row pool and one column pool. Therefore, the subject whose DNA sample is present at the intersection of these two pools must be the individual carrying that singleton variant. The targeted exons of the 1077 candidate genes were enriched using a custom Agilent SureSelect hybrid selection system. Sequencing was performed on the Illumina HiSeq platform. There was an average of 12,961,604 reads per pool resulting in mean target coverage of 213 reads (15 reads per subject in a pool of 14 subjects, or 30 total reads per subject, as each subject is present in two pools). Variant calling was performed using Syzygy software (31) and then applied a new likelihood-based secondary calling strategy that integrated the extra information from our matrix design.

Variants were annotated for functional effect using SnpEff 2.0.5 (http://snpeff.sourceforge.net/). Variant allele frequency data was obtained from three publicly available data sets: 1) Integrated variant call set of 1000 Genomes phase 1 samples (60) (February 2012 release); 2) National Heart Lung and Blood Institute Exome Variant Server (33), and 3) ~12,000 sequenced genomes and exomes assembled for exome genotyping chip design (http://genome.sph.umich.edu/wiki/Exome_Chip_Design). Maximal allele frequency from all three sources was used. Novel variants are those not observed in any of these datasets.

Assessing false positive and false negative rate

We estimated the false negative and false positive rates by comparing pooling data with data from exome sequencing previously performed in 6 of the short stature subjects. To start with, we determined the overlapping targets between pooling and exome capture arrays. Then, limiting to sites with ≥10 reads, we assumed that the exome sequencing data reflected the gold standard due to its much greater depth of coverage. False positives were defined as singletons observed in pooling data but not in exome data, while false negatives were those observed in exome data but not in either of the two relevant pools. The false positive rate was also estimated by looking for singleton variants that mapped to one of the empty "holes" in the matrix. The singleton variants that mapped to empty "holes" were false positives, permitting the number of false positive variants per individual to be estimated, and from there we could independently estimate the false positive rate of singleton variants.

IGF1R Functional Studies

Whole blood samples (BD Vacutainer Cell Preparation Tube with sodium heparin, Becton, Dickinson and Company, Franklin Lakes, NJ) were collected from the adopted patient and from the unrelated mother, who served as a normal control. Peripheral blood mononuclear cells (PBMC) were isolated following the manufacturer's protocol. PBMC, in freezing media (RPMI 1640+40% fetal bovine serum+10% DMSO) were stored in liquid nitrogen.

For immunoblot analysis, fresh PBMC ($2x10^6$ per treatment) were re-suspended in serum-free RPMI 1640 media, with or without recombinant IGF-I (100 ng/ml; GroPep Ltd, South Australia, Australia), for 20 min, 37 C, CO2 incubator, prior to pelleting and cell lysed as

previously described for fibroblast cell cultures (21). Western immunoblot analyses were performed as previously described (21).

For Flow cytometry analysis by FACS (Fluorescent-Activated Cell Sorting) of cell surface IGF1R, PBMC, warmed to 37°C from liquid nitrogen storage, were washed twice, aliquoted as $1x10^6$ cells per sample in RPMI 1640+10% FBS, and incubated overnight at 37°C (5% CO2 incubator). Prior to IGF-I treatment, cells were washed twice with serum-free RPMI+0.5% BSA and equilibrated in 0.5ml serum-free RPMI for 4 hours. Cells were treated with or without IGF-I (100ng/ml final concentration) for 1hr, after which cells were washed twice with cold staining media (1X PBS/0.5% BSA/0.1% sodium azide), and incubated with phycoerythrin (PE) conjugated anti-human IGFIR-alpha (CD221, BD Bioscience, San Jose, CA) for 30 minutes at 4°C in the dark. Following antibody staining, cells were washed twice with cold staining media and re-suspended in 200µl staining media/0.25% propridium iodide (PI) and incubated on ice for 10 minutes. 100,000 live PBMC (PI negative, CD221 positive) per sample were acquired via FACs Caliber (BD Bioscience, San Jose, CA) and fluorescence emitted by IGF1R-PE-labeled PBMC analyzed using FCS Express 3 analysis software (De Novo Software, Los Angeles, CA).

Results

Description of Cohort

Participants in this study included 192 subjects, with 106 males and 86 females, 75% of whom were Caucasian. The height Z-scores ranged from -2.05 to -7.01 SDS (Figure 3.1). The ages of these subjects ranged from 3 to 22 years with a mean of 10.3 years. Seventy subjects (36.4%) had begun growth hormone therapy for short stature prior to enrollment in the study. However, only 31 subjects (16%) were diagnosed with growth hormone deficiency; of these, 22 had isolated growth hormone deficiency without additional pituitary hormone defects. For those subjects on growth hormone therapy, only height z-scores prior to initiation of therapy are shown (Figure 3.1). An additional 14 subjects were thought to have known genetic syndromes but clinical diagnostic testing for the suspected syndromes had not identified pathogenic variants. Twenty nine subjects were reported to have developmental delay.

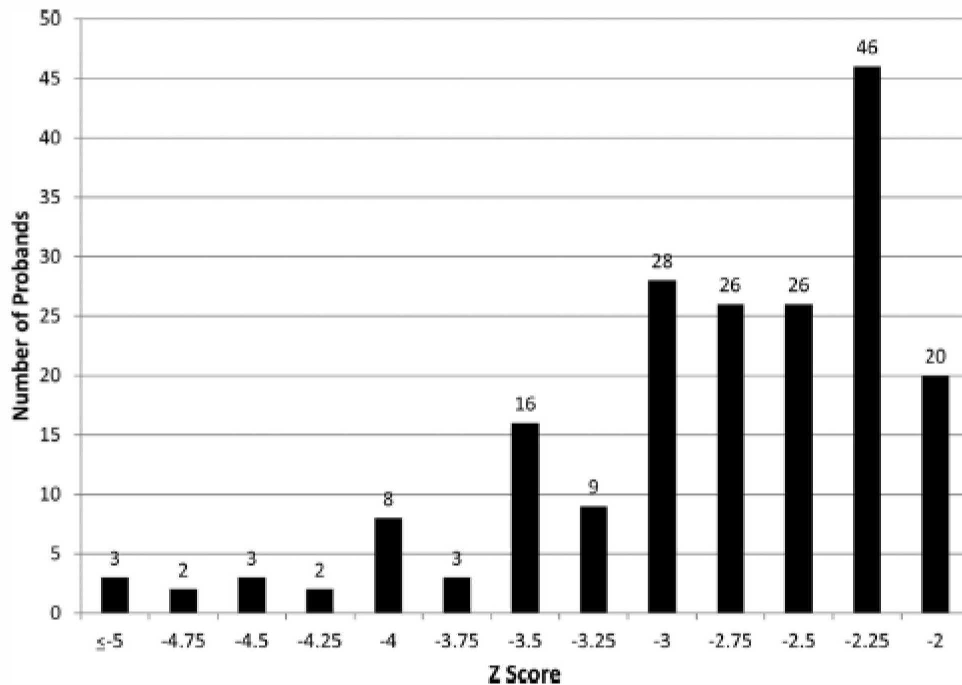


Figure** Error! No text of specified style in document.**1: Height Z-score at enrollment or prior to growth hormone therapy initiation. Each bar represents individuals with a height Z-score less than or equal to the number noted below it on the X-axis but greater than the number below the bar to the left. For example, the right most bar represents individuals with a height z-score $-2.25 < Z \leq -2$.

Validating Pooled Sequencing Results

Using our pooled sequencing design, the false positive rate of singleton variants estimated by comparison to the exome data was 4.8% (Table 3.1). Additionally, a total of 7 singleton variants mapped to the 8 “holes” in the two matrices compared to 7680 singleton variants which mapped to the 384 subjects (patients and controls) resulting in a similar estimation of the false positive rate of 4.2%. These numbers establish the upper bound for the variant false positive rate, as singleton variants are more likely to be false positives compared to variants found in multiple individuals. Out of a total of 6618 variants present in the 6 exome samples within our target region, 7 variants were not identified by the pooled sequencing giving an estimated overall false negative rate of 0.1%. Similar to the false positive rate, the false negative rate for singleton variants is likely to be higher than that of other variants as singleton variants only appear in two pools and are more difficult to identify.

Table 3.1: False positive rate of singletons estimated by comparing with exome sequencing of 6 samples

Sample ID	Number of singletons	False Positive singletons	False Positive rate
1	34	4	11.8%
2	58	1	1.7%
3	55	3	5.5%
4	15	0	0.0%
5	30	0	0.0%
6	16	2	12.5%
Sum	208	10	4.8%

Pathogenicity of Rare Variants

In the 192 short stature cases, we identified a total of 10,819 variants, of which 4,928 were not detected in the control samples. Of these, 1,349 were novel (Table 3.2). In order to screen for possible causal effects of variants found in our cohort, we compared these variants to those found in the Human Gene Mutation Database (HGMD) (61). The database contains 26,995 SNPs or indels located in the 1077 genes in our study where the variant has been reported as associated with a particular clinical phenotype. We identified 66 such SNPs that matched a variant detected in our cases but not in controls. As HGMD is known to have erroneous entries, we eliminated 7 variants with a minor allele frequency of 1% or greater as these are unlikely to be true pathogenic variants. Of the remaining 59 variants, 32 were associated with recessive conditions or predispositions to complex traits, and the clinical pictures of the patients were not consistent with the disease phenotype suggesting that they are unaffected heterozygous carriers. The final 27 variants previously associated with dominantly inherited diseases are listed in a supplementary table. We reviewed the phenotypes of the 27 cases and identified one case of autosomal dominant brachyolmia type 3 and three cases of Noonan syndrome. The remaining 24 case subjects did not have phenotypes consistent with the reported disease associations.

Table 3.2: Variants identified in short stature samples but not in control samples

Variant Type	Known Variants	Novel Variants

	ALL	*MAF≤5%*	*MAF≤1%*	
SILENT	1632	1602	1356	451
MISSENSE	1903	1888	1704	829
SPLICE	11	11	10	8
INDEL	11	11	10	46
NONSENSE	22	22	22	15
TOTAL	3579	3534	3102	1349

Identification of pathological variants associated with brachyolmia and Noonan Syndrome

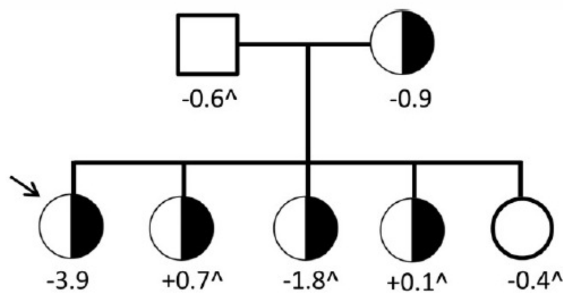
The patient with brachyolmia has a height of -3.88 SDS and platyspondyly of the cervical spine. The mutation in *TRPV4* is a missense mutation (c.1858G>A, V620I) that is a known variant causing the disease (62). Prior to the research results becoming available but subsequent to enrollment in our study, this patient was clinically diagnosed with brachyolmia and clinical testing revealed this mutation.

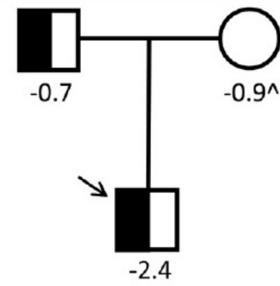
All three patients with Noonan Syndrome carried variants in *PTPN11*, the most common causative gene in this syndrome. Noonan syndrome is an autosomal dominant condition with characteristic dysmorphic facial features as well as short stature, webbed neck, and cardiac abnormalities (63). The first subject is an 11 year old female with a height Z score of -2.7 SDS. She was diagnosed with isolated growth hormone deficiency at age 7 years and had a poor response to growth hormone therapy. Of note, she was born with a transitional atrioventricular canal defect which was repaired at 4 months of age. She had a triangular face with a mildly low posterior hair line and slightly wide spaced eyes. She did not have ptosis or downslanting eyes and ears were normal. She carries the c.188A>G/p.Y63C variant (64). The second subject is an 8 year old female with a height Z-score of -1.7 SDS. She reached a height nadir of -3.0 SDS at age 5 prior to starting growth hormone therapy for an indication of being small for gestational age to which she had a good response. She was evaluated at age 4 by a geneticist for the possibility of Russell-Silver syndrome but no formal diagnosis was made. She has ptosis, epicanthal folds, downslanting eyes, and posteriorly rotated ears. She carries the c.925A>G/p.I309V variant (65) which she inherited from her father whose height is 173 cm (-

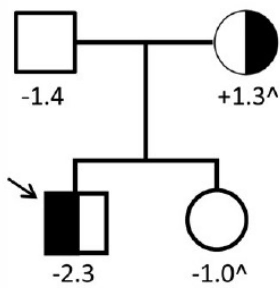
0.5 SDS). The third subject is a 16 year old male with a height Z-score of -2.5 SDS. He reached a height nadir of -3.2 SDS at age 13 prior to being diagnosed with isolated growth hormone deficiency and starting growth hormone therapy. He also initiated testosterone therapy at age 15 for delayed puberty. He has mild learning issues and on exam has a low posterior hair line but no other facial features consistent with Noonan syndrome. He carries the c.853T>C/p.F285L variant (65).

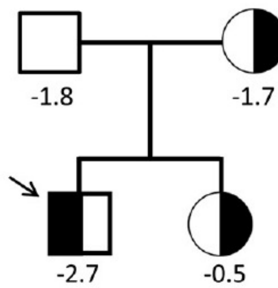
Identification of one pathological IGF1R variant amongst all IGF1R rare variants identified

To demonstrate the utility of a large-scale sequencing approach, and the need for careful interpretation of results, we focused on rare variants in *IGF1R*, a gene for which haploinsufficiency is known to cause significant short stature (66-68). Our approach was to identify nonsynonymous variants present in cases only which segregated with the phenotype of short stature within the families. Variants meeting these criteria would be classified as potentially pathogenic variant requiring further functional validation. In total, our targeted sequencing found 25 unique *IGF1R* variants in both subjects and controls. Of these, 16 were synonymous SNPs, most of which were common (minor allele frequency > 0.01); these were not evaluated further. The remaining 9 variants included 6 missense, one frameshift, and two intronic variants. The intronic variants were both found within 5 base pairs of an exon, which could potentially suggest involvement in splicing and thus were included for further analysis. Five of these variants were present in cases only (Table 3.3). All nine variants were validated via traditional Sanger sequencing and confirmed to be heterozygous. To determine the biological significance of these variants, segregation of variants #2-6 within families was performed (Figure 3.2). There was no correlation between the individual family member's heights and carrier status of the variants suggesting that these variants are unlikely to be major contributors to the patients' short stature, and therefore, we excluded these variants from further consideration as pathogenic variants. Variant #7 was present in multiple cases and controls and was also unlikely to be pathogenic. Of note, one of the two rare missense variants found only in controls in our study (variant #9) was previously reported as pathogenic in the literature (69). This control is of normal stature at -0.4 SDS.









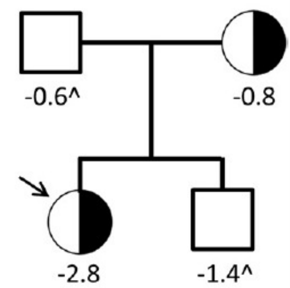


Figure 3.2: Segregation of identified IGF1R nonsynonymous variants in affected families does not correlate with short stature. Numbers below the individuals denote the height Z-scores. ^ indicates that the height was estimated by a family member. All other values were measured. Individuals carrying the heterozygous variants are indicated as black half-filled in circles (females) or squares (males). The arrow points to the affected proband in each family.

Table 3.3: *IGF1R* potentially nonsynonymous variants

Variant	Exon	cDNA	Protein	MAF	Subject	Sex	Height SDS	Birth weight*	IGF-1 (normal range**)
1	2	c.418dupG	p.A140Gfs*5	novel	Case 1	F	-4.1	unknown	389.8 ng/mL (244 - 787)
2	5	c.1247+3A>G	Intron	0.0007	Case 2	F	-3.9	2800 g	26.6 ng/mL (49 - 342)
3	7	c.1463-5C>A	Intron	0.012	Case 3	M	-2.4	3500 g	52.2 ng/mL (49 - 342)
4	6	c.1411C>T	p.R471C	novel	Case 4	M	-2.3	4100 g	34 ng/mL (63 - 279)
5	7	c.1502C>T	p.S501L	novel	Case 5	M	-3.0	3400 g	148 ng/mL (63 - 279)
6	6	c.1336A>G	p.M446V	0.0027	Case 6	F	-2.8	2600 g	97.2 ng/mL (49 - 342)
					Control 6	F	0.0		
7	6	c.1310G>A	p.R437H	0.004	Case 7a	F	-3.3	4200 g	69.9 ng/mL (49 - 342)
					Case 7b	F	-3.1	3800 g	62 ng/mL (63 - 279)
					Control 7a	F	+0.4		
					Control 7b	M	0.0		
					Control 7c	M	-0.4		
8	5	c.1162G>A	p.V388M	0.003	Control 8	M	-0.6		
9	7	c.1532G>A	p.R511Q	0.003	Control 9	M	-0.4		

*All cases were the product of full-term pregnancies (>37 weeks) with the exception of case #8 (36.5 weeks). None of the cases met the definition for IUGR (weight <2500 g at birth for normal gestation). **Normal range for sex and Tanner stage. All IGF-1 values were obtained during a baseline clinical evaluation and were measured when the patient was not on growth hormone therapy.

Variant #1 was a novel frameshift mutation (c.418dupG/p.A140Gfs*5) found in one case in the heterozygous state. The mutation causes severe truncation of the protein with complete abrogation of the transmembrane and intracellular domains and thus was predicted to lead to haploinsufficiency. This patient was adopted from China at 6 years of age and therefore a complete history and familial samples could not be obtained. At the age of 15 years, the patient had Tanner stage 4 breast development with a height of 136 cm (-4.06 SDS), weight of 30.2 kg (-4.87 SDS), and a head circumference of 49.3 cm (-4.4 SDS). She has a history notable for bilateral cleft lip and palate as well as attention deficit disorder and mild developmental delay. Her IGF-1 was normal at 389.8 ng/mL (normal range 244 – 787 ng/ml for Tanner Stage 4 female). Growth hormone stimulation testing with arginine and glucagon demonstrated a peak growth hormone level of 18.8 ng/ml. She had previously been treated with growth hormone therapy with a possible mild increase in growth velocity, although this occurred concurrently with entering puberty.

Variant #1 was the only variant in IGF1R that met our pre-specified criteria for consideration as a potential pathogenic variant. To determine if Variant #1 was causal for the patient's phenotype, we evaluated IGF1R expression and function in primary PBMC derived from the patient compared to control PBMC (procured from the unrelated adoptive mother). Flow cytometric analysis by FACS of live PBMC (Counts, Y-axis, Figure 3.3) indicated fluorescence emitted by IGF1R-PE-labeled PBMC was markedly reduced (Fluorescence Intensity, X-axis, Figure 3.3) in patient PMBC when compared to the normal control PBMC (Figure 3.3A). When the live PBMC were treated with IGF-I, emitted fluorescence was comparably reduced for both control and patient PBMC, suggesting normal internalization of IGF1R upon ligand binding (Figure 3.3B). Immunoblot analysis of cell lysates, furthermore, indicated total IGF1R expression was reduced in the patient PBMC with correlating reductions in IGF-I-induced signaling (Figure 3.3C). Altogether, the results are consistent with the heterozygous *IGF1R c.418dupG* variant inducing a state of IGF1R deficiency and being an excellent candidate to cause the subject's short stature.

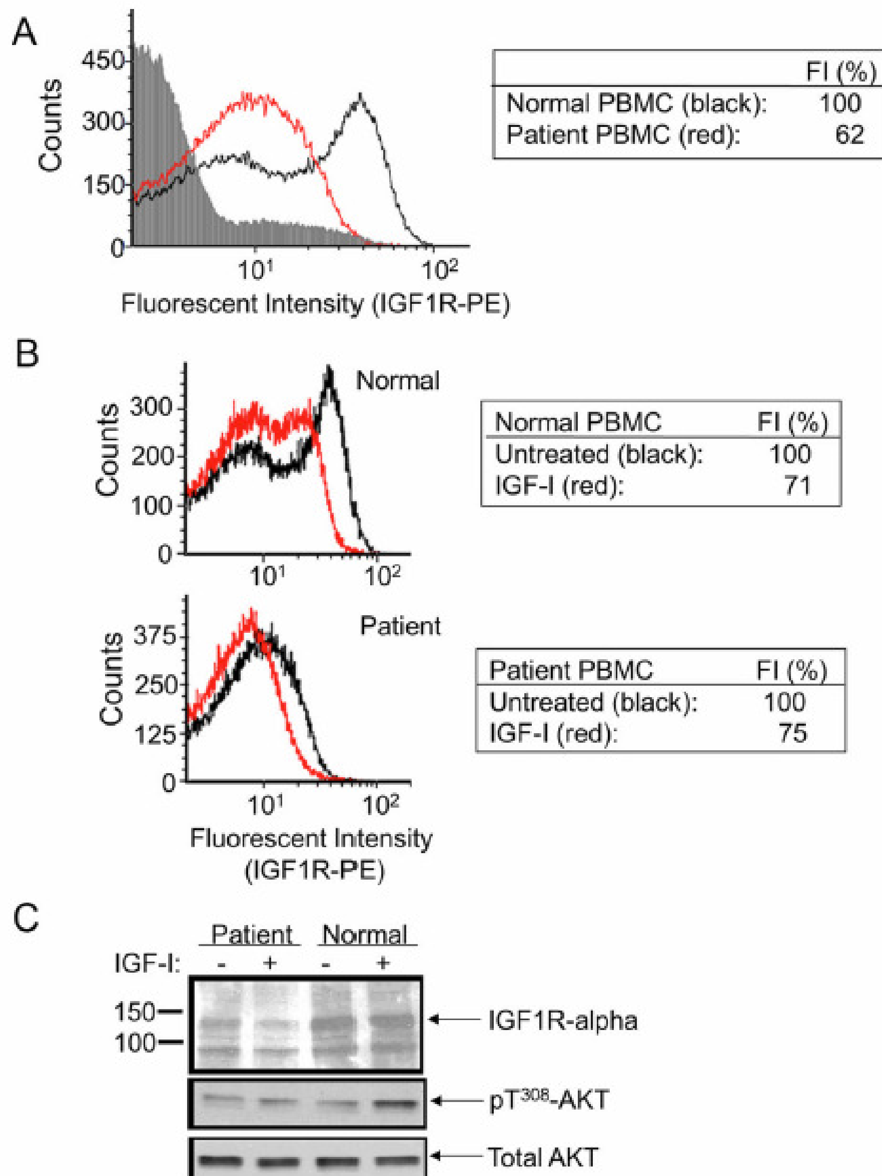


Figure 3.3: IGF1R expression and signaling in primary peripheral blood mononuclear cells (PBMC) of patient carrying heterozygous IGF1R c.418dupG. PBMC were isolated as indicated in Methods. Flow cytometry analysis by FACS was employed to detect IGF1R, labeled by PE-conjugated anti-human IGF1R-alpha antibody (see Methods), on the cell surface of live PBMC. Live PBMC (Counts, Y-axis) and fluorescence emitted by the IGF1R-PE-labeled PBMC, were collated (log scale Fluorescence Intensity, X-axis). (A) Patient (red graph) compared to Normal (black graph) PBMCs. Background fluorescence emitted by unlabeled and untreated PBMC control is shown by the grey shaded region. Geometric mean of the Fluorescent Intensity (FI) detected in Normal PBMC was given an arbitrary unit of 100% (table). (B) The effect of IGF-I treatment (100 ng/ml, 1 hr) on the detection of IGF1R-PE-labeled PBMC from Normal (upper panel) and Patient (lower panel). For each, the geometric mean of the Fluorescent Intensity (FI) detected in untreated PBMC was given an arbitrary unit of 100%. (C) Western immunoblot analysis of total cell lysates from PBMC treated with IGF-I (100ng/ml) versus untreated cells. Molecular weight (kDa) is indicated on the left side of the immunoblots. The intracellular proteins detected are indicated by arrows (on right).

Discussion

Short stature is a common problem confronting pediatric endocrinologists. After exclusion of other chronic diseases or overt hormonal deficiencies, clinicians are often unable to provide a definitive diagnosis for the etiology of an individual patient's short stature. There are a multitude of genetic causes for short stature, but the majority of patients do not fall into a previously identified genetic syndrome. We, therefore, designed and performed a large-scale sequencing project to identify pathogenic rare genetic variants in individuals with short stature. We sequenced a list of 1077 candidate genes including known skeletal dysplasia genes, genes within the growth hormone signaling pathway, genes known to affect growth plate biology, and genes within loci associated with adult height in large GWA studies. Using this approach, we identified four known pathogenic variants causing short stature as well as novel variants in genes known to affect stature.

To facilitate the sequencing of a large number of genes in many subjects, we employed a pooled sequencing design which significantly reduced the cost of such analysis (70). The majority of the cost associated with a targeted next generation sequencing project is typically incurred at the library construction stage, in which targeted regions of DNA are separated from the remainder of the genome for sequencing. In a pooled sequencing design, this process only has to be performed once per pool. Although actual sequencing depth may need to be increased in order to ensure adequate representation of all samples in the pool, the associated cost is relatively minor. Indeed, the cost per sample of our pooled targeted sequencing approach is estimated to be ~ 20% of the cost for individual exome sequencing (i.e. $150 compared to $800 per sample, based on current charges at the Broad Institute). Exome sequencing could also be done in a pooled fashion, in which case, cost differences will depend on the cost of sequencing coverage, a process which is becoming cheaper to perform. While pooled exome sequencing does have the advantage that nearly all genes are evaluated, analysis and interpretation of the data generated would be much more complex due to the large number of novel nonsynonymous variants in genes with no known connection to the phenotype of interest. Our simple matrix pooling design, in contrast, allows for the rapid assessment of low frequency variants in candidate genes, and to identify individuals carrying singleton variants, which are more likely

to be pathogenic compared to variants with a higher minor allele frequency. However, pooling does limit the ability to discern if a single variant is homozygous or heterozygous in an individual subject and follow up confirmatory genotyping is necessary. Using this design, we were able to identify a large number of very rare nonsynonymous variants within our candidate genes with a low false positive and low false negative rate.

We identified four subjects in our cohort who had known pathogenic variants implicated in disease. Notably, three of these subjects have mutations in *PTPN11* that are causal of Noonan syndrome. Noonan syndrome is known to have a wide phenotypic spectrum leading to difficulty in diagnosis (63), and, indeed, one of our subject's fathers carries a proven pathogenic variant in *PTPN11* yet never presented with overt clinical manifestations of Noonan syndrome. While it is true that our subjects may have had features consistent with Noonan syndrome which were unrecognized, such as a cardiac defect or delayed puberty, this retrospective recognition of related features does not eliminate the benefit of genetic screening. The lack of diagnoses in our cohort represents clinical reality, as these subjects were extensively evaluated by experienced pediatric endocrinologists and in one case a geneticist as well. This suggests that a substantial number of patients with Noonan syndrome are designated as having idiopathic short stature or isolated growth hormone deficiency even after clinical evaluation by pediatric subspecialists. Additional research is needed to determine if widespread screening of *PTPN11* or the other Noonan syndrome genes is warranted for patients with short stature of unknown etiology.

It is important to note that, in our cohort, the vast majority of HGMD reported disease-causing dominant mutations did not manifest with the associated clinical phenotype. The classification of these variants as pathogenic is likely erroneous and based on insufficient clinical evidence. However, we cannot rule out the possibility that the variants have variable expressivity and some of our subjects are presenting on the very mild end of the clinical spectrum with short stature as their disease manifestation.

Our focus on rare variants of the *IGF1R* gene illustrates the critical importance of providing supporting familial segregation and functional data when a rare variant has been identified. IGF-1, the primary mediator of growth hormone function, is essential for growth.

Heterozygous and compound heterozygous mutations in *IGF1R* that lead to decreases in quantity or function of the receptor have been described in nearly a dozen human cases (5, 66-68, 71, 72). These patients display variable phenotypes, with shared characteristics that include poor prenatal and postnatal growth, microcephaly, high or normal IGF-1 levels, and developmental delay (5, 66-68, 71, 72).

Our targeted sequencing approach identified 7 unique rare nonsynonymous *IGF1R* variants as well as 2 intronic variants with the potential to affect splicing due to their proximity to exons. Only one of these 9 variants, a novel *c.418dupG* frameshift mutation located in exon 2, was associated with clinical features suggestive of a pathological IGF1R deficiency state (high levels of serum IGF-1, microcephaly, intrauterine growth retardation). Furthermore, in primary cells derived from this patient, significant decreases in both IGF1R expression and IGF-I-induced signaling supported the pathogenicity of the *IGF1R c.418dupG* defect. None of the remaining variants found in cases show convincing evidence of pathogenicity. This example demonstrates that large scale sequencing efforts will identify numerous very rare and novel nonsynonymous variants in candidate genes. The majority of these variants will be missense variants leading to a change in a single amino acid, most of which will not affect protein function and represent incidental findings. Segregation of these variants with the phenotype within families is the first critical step in evaluating potential pathogenicity, highlighting the importance of collecting familial samples at the time of initial DNA collection.

Filtering strategies based on population allele frequency are useful and necessary but the majority of public databases do not provide individual phenotypic data linked to the subject's genotype, thus limiting the ability to determine if a variant is potentially pathogenic. Therefore, simultaneous sequencing of a control cohort with a known phenotype, in this case normal stature, provides additional information about the lack of pathogenicity of rare variants in a gene. This is exemplified by our finding that an *IGF1R* variant previously reported to be pathogenic (c.1532G>A/p.R511Q) (69) was found in the heterozygous state in a control of normal stature. This variant was originally identified in the heterozygous state in a patient and her maternal aunt, both of whom presented with extreme short stature (-6.1 and -5.7 SDS, respectively). It is of note that information regarding the parents of the patient were lacking in

this report. *In vitro* reconstitution studies of the homozygous p.R511Q variant were performed to support the pathogenicity of this variant, although the effect of heterozygosity was unknown. These caveats, together with our identification of the same variant in a normal statured control, strongly suggest that a heterozygous p.R511Q is unlikely to be causal of the previously reported family's extreme short stature. Furthermore, Kansra et al. (73) recently detected the R511Q variant in 6 of 1800 public school students. Indeed, carriers had an average height around the 27th percentile, thus providing additional evidence that this variant does not cause severe short stature. Altogether, these results support the importance of segregation analysis and the need to include primary cells in functional analysis.

Our study has a number of important limitations. We recruited a very heterogeneous cohort, allowing for the inclusion of dysmorphic features, other congenital anomalies, and hormonal deficiencies provided that there was no known genetic etiology for these findings. Thus, subjects in this cohort do not meet a strict definition of idiopathic short stature (2). Nevertheless, we believe that this cohort more accurately represents the diversity of patients who are seen in a referral setting and is likely enriched for individuals with rare genetic variants that may have multisystem effects. Additionally, our hybrid selection strategy only targets the exons of the candidate genes, and thus, any noncoding variation that affects gene expression cannot be detected by our methods. Variants affecting gene expression can play an important role in causing short stature. For example, Russell-Silver Syndrome, an important syndromic form of short stature, is often due to abnormalities in methylation of chromosome 11p15.5 leading to aberrant gene expression (74). Additionally, our current approach does not assess copy number variation (i.e. deletions or duplications of genes) which may also be an important genetic defect leading to short stature. We are currently pursuing copy number analysis of this cohort using a custom chromosomal microarray (data not shown). Furthermore, we do not obtain perfect sequencing coverage of all variants in the targeted region and could miss potentially pathogenic variants in the candidate genes. Finally, due to the large number of rare missense variants in both cases and controls, we have limited power to discover new genes with a statistically significant excess of mutations in cases versus controls. Ongoing work to increase sample size and examine subjects at the extremes of the height distribution will provide additional data to support novel gene discovery.

In conclusion, we present the initial results of a large-scale candidate gene sequencing effort in children with short stature and demonstrate the complexity of data interpretation of such efforts. Three of our 192 subjects were found to have known pathogenic variants in *PTPN11* highlighting the possibility that Noonan syndrome is underdiagnosed in the clinical setting. We report a novel frameshift mutation in *IGF1R* and demonstrate its pathogenicity *in vivo*. Additionally, we provide evidence that a previously reported variant in *IGF1R* is not pathogenic. Analyses of variants identified in the other candidate genes are currently ongoing.

Chapter IV:

Heterozygous mutations in natriuretic peptide receptor-B (*NPR2*) gene as a cause of short stature

Sophie R. Wang[1*], Christina M. Jacobsen[2,3*], Heather Carmichael[4*], Aaron B. Edmund[5], Jerid W. Robinson[6], Robert C. Olney[7], Timothy C. Miller[8], Jennifer E. Moon[2], Veronica Mericq[9], Lincoln R. Potter[5,6], Matthew Warman[3,4,10,11], Joel N. Hirschhorn[2,10,12], Andrew Dauber[2,12,13]

1. Genome Institute of Singapore, Singapore 138672
2. Division of Endocrinology, Boston Children's Hospital, Boston, Massachusetts 02115
3. Orthopaedic Research Laboratories, Department of Orthopaedic Surgery, Boston Children's Hospital, Boston, Massachusetts 02115
4. Harvard Medical School, Boston, Massachusetts 02115
5. Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota, Minneapolis, Minnesota 55455
6. Department of Pharmacology, University of Minnesota, Minneapolis, Minnesota 55455
7. Division of Endocrinology, Nemours Children's Clinic, Jacksonville, Florida 32207
8. Massachusetts College of Pharmacy and Health Sciences, Boston, Massachusetts 02115
9. Institute of Maternal and Child Research, Faculty of Medicine, University of Chile, Santiago, Chile
10. Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115
11. Howard Hughes Medical Institute, Boston, Massachusetts 02115
12. Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts 02142
13. Division of Endocrinology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229

*These authors contributed equally.

This is the accepted version of the following article:

Wang SR, Jacobsen CM, Carmichael H, Edmund AB, Robinson JW, Olney RC, Miller TC, Moon JE, Mericq V, Potter LR, Warman ML, Hirschhorn JN, Dauber A. Heterozygous Mutations in Natriuretic Peptide Receptor-B (NPR2) Gene as a Cause of Short Stature. Hum Mutat. 2015 Feb 23.

This has been published in final form at:

http://onlinelibrary.wiley.com/doi/10.1002/humu.22773/abstract;jsessionid=824910CF7014022225EAC65883AFE1B4.f02t01

Abstract

Based on the observation of reduced stature in relatives of patients with acromesomelic dysplasia, Maroteaux type (AMDM), caused by homozygous or compound heterozygous mutations in natriuretic peptide receptor-B gene (*NPR2*), it has been suggested that heterozygous mutations in this gene could be responsible for the growth impairment observed in some cases of idiopathic short stature (ISS). We enrolled 192 unrelated patients with short stature and 192 controls of normal height and identified seven heterozygous *NPR2* missense or splice site mutations all in the short stature patients, including one *de novo* splice site variant. Three of the six inherited variants segregated with short stature in the family. Nine additional rare nonsynonymous *NPR2* variants were found in three additional cohorts. Functional studies identified eight loss-of-function mutations in short individuals and one gain-of-function mutation in tall individuals. With these data we were able to rigorously verify that *NPR2* functional haploinsufficiency contributes to short stature. We estimate a prevalence of *NPR2* haploinsufficiency of between 0 and 1/26 in people with idiopathic short stature. We suggest that *NPR2* gain of function may be a more common cause of tall stature than previously recognized.

Introduction

C-type natriuretic peptide (CNP, gene natriuretic peptide precursor C, *NPPC*; OMIM 123830) is a small, secreted peptide and a member of the natriuretic peptide family. CNP binds to a homodimeric transmembrane receptor named natriuretic peptide receptor-B (NPR-B, gene *NPR2*; OMIM 108961), which functions as a guanylyl cyclase to generate cGMP in chondrocytes, female reproductive organs, and endothelial cells(75, 76). In the growth plate chondrocytes, binding of CNP to NPR-B stimulates chondrocyte differentiation and hypertrophy as well as increases matrix synthesis. This occurs in part through NPR-B signaling inhibiting MAP kinase signaling by FGFR3 (77-79).

Several lines of evidence indicate that CNP/NPR-B signaling is an important regulator of skeletal growth. CNP-overexpressing mice exhibit excessive growth (80), while defects of the

Nppc (78) or *Npr2 (81)* gene lead to impairment of skeletal development. In humans, biallelic loss-of-function mutations in *NPR2* cause acromesomelic dysplasia, Maroteaux type (AMDM; OMIM 602875). This autosomal recessive skeletal dysplasia is characterized by dwarfism and short limbs (82). On the other hand, overproduction of CNP due to a chromosomal translocation has been reported in association with a skeletal dysplasia characterized by tall stature (83, 84). In addition, gain-of-function mutations of *NPR2* have been identified in several families and are associated with an overgrowth disorder with only mild skeletal features (85-87).

Interestingly, in the first report of biallelic *NPR2* mutations causing AMDM, parents of patients with AMDM (obligate heterozygotes) were noted to be shorter than expected for their population of origin (82), although these individuals came from a wide range of geographic and ethnic backgrounds, which complicated the interpretation of these findings. Another study that evaluated a single family with an AMDM proband showed that the heterozygous carriers had a mean height 1.4 SDS lower than their non-carrier family members(22). In this single-family study, the proband's parents share a common ancestor, so it is possible that the heterozygous carriers share some other mutation causing short stature, and the level of evidence of a heterozygous effect was limited by the size of the family. Based on these two studies, it is presumed that heterozygous *NPR2* loss-of-function mutations can mildly impair long bone growth. It has further been hypothesized that one person in 30 with idiopathic short stature (ISS) will carry a *NPR2* mutation (22, 82).

Recent studies searched for heterozygous *NPR2* mutations in cohorts with ISS (Table 1). One study of 47 independent Brazilian patients identified heterozygous *NPR2* mutations in 6% of patients(88). Another study on 101 unrelated Japanese patients with short stature identified heterozygous *NPR2* mutations in 2% of patients (89). While providing observational data consistent with the hypothesis that a monoallelic *NPR2* mutation could cause short stature, these studies did not include controls without short stature, and were based on a relatively small number of patients. Hence, the hypothesis that *NPR2* haploinsufficiency leads to short stature has not been rigorously verified. Analyses of larger cohorts, including controls, are needed to more clearly define the role of heterozygous *NPR2* defects in patients with ISS.

Table 4.1: Summary of recent studies searching for heterozygous NPR2 mutations in cohorts with ISS

Cohort	# of functional NPR2 variants	# of ISS individuals	Fraction of ISS individuals with functional NPR2 variants	Reference
Brazilian patients with ISS	3	47	6%	(Vasques et al. 2013)
Japanese patients with short stature	2	101	2%	(Amano et al.)

Materials and Methods

Short stature patient cohort and Framingham Heart Study controls

192 patients with short stature (height more than 2 SDS below the mean for age and sex) (59) but without defined etiology, were recruited from the endocrinology and genetics clinics at Boston Children's Hospital. 104 subjects fit strict criteria for ISS with no evidence of growth hormone deficiency, significant medical comorbidity, developmental delay, or suspected syndromic cause for their short stature. Of these 104 patients, height data was available for both parents in 99 cases. 22 of these 99 subjects (22.2%) had at least one parent with a height below -2 SDS and were defined as having familial ISS. In addition, 192 control subjects were chosen from the middle of the Framingham Heart Study (FHS) height distribution (height z scores between −0.7 and +0.7). More detailed description of the cohort is reported in Wang, *et al.* (90).

Second short stature patient cohort

216 patients were taken from the Genetics and Neuroendocrinology of Short Stature International Study (GeNeSIS), sponsored by Eli Lilly and Company and various pediatric endocrinology clinics in Jacksonville, FL and Santiago, Chili. Patients were diagnosed by the treating provider with idiopathic short stature and had height < -2 SDS below the mean for age and sex.

FINRISK height extreme cohort

272 subjects were chosen from the extremes ($<1^{st}$ percentile or $>99^{th}$ percentile) of the FINRISK surveys (FINRISK 1992, FINRISK 1997, FINRISK 2002, FINRISK 2007) height distribution. The FINRISK cohorts comprise the respondents of representative, cross-sectional population surveys that are carried out every 5 years since 1972, to assess the risk factors of chronic diseases and health behavior in the working age population, in five large study areas of Finland (91).

Estonian Biobank height extreme cohort

1000 subjects were selected from the extremes ($<1.25^{th}$ percentile or $>98.75^{th}$ percentile) of the Estonian Biobank height distribution. The Estonian Biobank cohort is a volunteer-based sample of the Estonian resident adult population (age ≥18 years). The age, sex and geographical distribution closely reflect those of the Estonian adult population and encompass close to 5% of the entire adult population of Estonia (92).

Pooled sequencing protocol

For the first short stature cohort and the FINRISK and Estonian cohorts, DNA samples from multiple subjects were pooled for DNA sequencing using previously described methods available at the Broad Institute(93). To identify variants present only in a single individual (hereafter referred to as singleton variants), we applied a simple overlapping pooling design as described in Wang, *et al*. (90).

Sequencing was performed on the Illumina HiSeq platform. Variant calling was performed using a likelihood-based calling strategy we developed as described previously (90). Variants were annotated for functional effect using SnpEff 2.0.5 (http://snpeff.sourceforge.net/). The reference sequence used for numbering nucleotide change and amino acid change was NM_003995.3 and ENSP00000341083. Nucleotide numbering uses +1 as the A of the ATG translation initiation codon in the reference sequence, with the initiation codon as codon 1.Variant allele frequency data was obtained from two publicly available databases: 1) Integrated variant call set of 1000 Genomes samples (94) (May 2013 release); 2) National Heart, Lung, and Blood Institute Exome Variant Server (National Heart, Lung). Maximal allele

frequency from the two sources was used. Novel variants are those not observed in either of these databases. Confirmation of variants found in *NPR2* through pooled sequencing was done via Sanger sequencing or Sequenom MassArray genotyping. Each variant was sequenced in the proband and in all related family members for whom DNA samples were provided. Sanger sequencing of *NPR2* was performed in the second short stature cohort.

In silico prediction of mutation effects

To identify the potential effects of sequence variants identified in *NPR2* on protein function or structure, the wild-type and variant sequences were assessed using PolyPhen-2 (software version 2.2.2, http://genetics.bwh.harvard.edu/pph2) (95). The options used were: Classifier model [HumDiv], Genome aseembly [GRCh37/hg19], Transcripts [CCDS], Annotations [All].

Assaying wild-type and mutant NPR2 activity

Missense mutations in *NPR2* were generated by site-directed mutagenesis kit (Agilent Technologies) using the wild-type rat *NPR2* expression construct pRK5-NPR-B. Activity in HEK 293T cells was measured by at least one of two assays, one on whole cell lysate and another enzyme based assay done on membrane fractions of transfected cells, as described elsewhere (96, 97). Variants showing loss of activity *in vitro* in either assay were characterized as loss-of-function. Variants showing normal *in vitro* activity in the whole cell lysate assay were tested in the more sensitive enzyme based assay. Variants showing clear increases in activity in both assays were characterized as gain-of-function.

Results

Discovery of NPR2 variants in patients with short stature and controls using pooled sequencing

We selected 192 patients with short stature (more than 2 SDS below the mean for age and sex) and 192 controls of matching ancestry from the middle of the Framingham Heart Study height distribution (height z scores between -0.7 and +0.7). Characteristics of the subjects were described previously (90); 104 subjects fit strict criteria for ISS with no evidence of growth

hormone deficiency, significant medical comorbidity, developmental delay, or suspected syndromic cause for their short stature. Pooled targeted sequencing of the exons of 1077 candidate genes was performed (90). We detected 11 variants in *NPR2.* Of these, two were synonymous SNPs found in multiple patients and controls, and two were synonymous variants found only in patients. We focused on the seven potential loss-of-function variants, which included one splice site and six missense variants, all found in patients only (Table 4.2). These seven variants were all validated via traditional Sanger sequencing and confirmed to be heterozygous in each individual carrier.

Family analyses and clinical phenotypes

The splice site mutation is c.1352-1G>A in a highly conserved base pair in the acceptor splice site at the 5' end of exon 7, carried in a single patient (Patient 1, Table 4.2). Sanger sequencing of the proband, his parents, and a brother confirmed that the variant is found in the patient but not in his mother, father or brother. These results were confirmed by a second round of sequencing, and paternity was confirmed by SNP genotyping. There was no family history of short stature (Figure 4.1), consistent with the hypothesis that the *de novo* variant in *NPR2* is contributing to short stature in this patient.

Table 4.2: ***NPR2*** **potentially nonsynonymous variants in short stature patients**

Cohort	Subject	Sex	Variant	Minor Allele Frequency	PolyPhen2 Prediction	Nadir Height SDS	Current Height SDS	GH Therapy	IGF-1 Level (ng/ml)
Short stature patient cohort	Patient 1	M	c.1352-1G>A	Novel	N/A	-3.14	-2.15	Yes	53 @ 7 yr (44-211)
	Patient 2	M	c.142G>T; p.A48S (brachyolmia with known TRPV4 variant)	Novel	1.00	-4.15	-3.91	No	NA
	Patient 3	F	c.1167G>T; p.E389D	Novel	0.00	-3.84	-3.14	Yes	53 @ 6 yr (47-217)
	Patient 4	F	c.1481T>G; p.I494S	Novel	0.997	-3.57	-1.79	Yes	115@ 6 yr (47-217)
	Patient 5	M	c.1645G>A; p.A549T	Novel	0.998	-3.88	-2.70	Yes	34 @ 4 yr (25-157)
	Patient 6	M	rs62637657: c.491C>G; p.A164G	0.02%	0.065	-2.53	-2.16	IGF-1 therapy	72 @ 9 yr (61-252)
	Patient 7	F	rs114147262: c.2359C>T; p.R787W	0.16%	0.998	-2.92	-2.41	No	61 @ 7 yr (55-238)
Second short stature cohort	Patient	NA	c.316G>T; p.A106S	Novel	0.189	N/A	N/A	N/A	N/A
	3 Patients	NA	rs180950551: c.1802G>C; p.R601P	0.02%	1.00	N/A	N/A	N/A	N/A
	Patient	NA	c.2710A>T; p.K904X	Novel	N/A	N/A	N/A	N/A	N/A
	Patient	NA	rs114115939: c.1517G>A; p.R506H	0.06%	0.959	N/A	N/A	N/A	N/A

Abbreviations: F, female; M, male; GH, growth hormone; IGF-1, insulin-like growth factor 1; yr, years old; N/A, not available.
Normal age and gender specific ranges for IGF-1 are provided in parentheses.

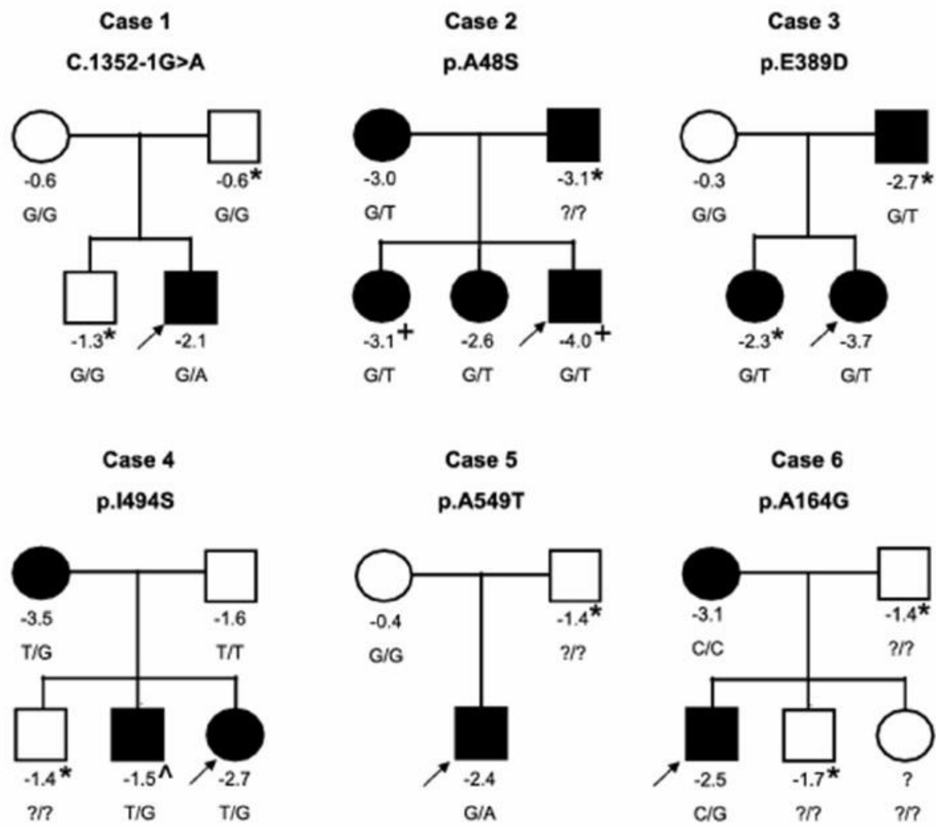


Figure 4.1: Numbers below the individuals denote the height z scores. + indicates family members who also carry the *TRPV4* mutation causal of brachyolmia. ^ indicates that this family member was treated with growth hormone therapy prior to this height measurement and had a nadir height of -2.8 SDS. * indicates that the height was estimated by a family member. All other values were measured. Individuals with short stature are indicated as black circles (females) or squares (males). The arrow points to the affected proband in each family.

We also detected six missense variants in *NPR2*; of these, four were private variants not found in the public databases (Patients 2-5, Table 4.2). Sequencing of relatives demonstrated that three of these novel variants segregated with the short stature phenotype (Patients 2-4, Figure 1). The mother of Patient 5, who did not have short stature, did not carry the *NPR2* variant, but the father's DNA was not available to further confirm the segregation. Of note, one of the three segregating variants (Patient 2) was found in a male patient who was also found to carry a known *TRPV4* mutation (c.1858G>A (p.V620I)) previously reported as causal for brachyolmia type 3 (62). The patient carries a clinical diagnosis of brachyolmia and features

consistent with this disease, including platyspondyly of the cervical spine. We previously reported the presence of this likely pathogenic variant in this patient (90). The *NPR2* variant was present in the patient's mother, who does not carry the *TRPV4* variant, and in two sisters, the elder of which also carries the *TRPV4* variant. The mother and both sisters had short stature with height SDS scores below -2.5, and the sister carrying both variants had a height SDS score of -3.1. Notably the patient's father is deceased but also was reported to have had short stature (-3.1 SDS) and presumably carried the *TRPV4* variant. Thus, this patient and one of his sisters inherited the *NPR2* variant from his mother and the *TRPV4* variant from his father, presumably resulting in severe short stature with skeletal dysplasia.

We also detected two rare missense variants present in the public databases (Patients 6-7). Neither of these variants has been reported to be pathogenic, and both are reported to be relatively rare (minor allele frequency 0.02% and 0.16%). The mother of Patient 6, who also had short stature, did not carry the *NPR2* variant, but DNA was not available from the sibling or father for further segregation analysis. Patient 7 was adopted from China, so the family history of short stature was unknown and testing for segregation of the variant was not possible.

Discovery of NPR2 variants in one additional clinical short stature cohort

To further refine the estimate of the fraction of ISS individuals with functional *NPR2* variants, we screened for rare nonsynonymous *NPR2* variants in a second short stature cohort (n=216) from the Lilly GeNeSIS Study and a number of pediatric endocrinology clinics. Patients were diagnosed by the treating provider with ISS and had height more than 2 SDS below the mean for age. The exons and intron/exon boundaries of the *NPR2* gene were sequenced using Sanger sequencing.

Four nonsynonymous variants in the coding region were identified, one nonsense and three missense variants (Table 4.2). One of the missense variants (rs180950551: c.1802G>C (p.R601P)) was seen in three separate patients. An *in silico* analysis suggested that one variant (c.316G>T (p.A106S)) is benign and the other two (rs180950551: c.1802G>C (p.R601P) and rs114115939: c.1517G>A (p.R506H)) are probably damaging.

Screening for NPR2 mutations in population-based cohorts of individuals from the extremes of the height distribution

We then went on to screen for rare nonsynonymous *NPR2* variants in two additional cohorts. The first cohort of individuals (n=272) was selected from the extremes (<1st percentile or >99th percentile) of height distribution from four FINRISK surveys (~33,000 samples in total). Pooled targeted sequencing of the exons of 1077 candidate genes was performed. We detected three variants in *NPR2.* Of these variants, one was a common synonymous SNP (observed in the short stature patients and FHS controls as well), which we did not pursue further. The other two were missense variants, which we validated by genotyping and confirmed to be heterozygous in tall extremes only (Table 4.3). Neither variant was found in the public databases. An *in silico* analysis suggested that one mutation is benign (c.739A>G (p.N247D)) and the other is probably damaging (c.1685G>A (p.R562Q)).

Table 4.3: *NPR2* potentially nonsynonymous variants in FINRISK and Estonian Biobank height extreme samples

Variant	Observation	Minor Allele Frequency	PolyPhen2 Prediction
c.739A>G; p.N247D	FINRISK tall extreme (also observed in Estonian tall extreme and short extreme)	Novel	0.002
c.1685G>A; p.R562Q	FINRISK tall extreme	Novel	0.995
c.2449G>A; p.E817K	Estonian short extreme	Novel	0.609
c.2338G>A; p.G780R	Estonian short extreme	Novel	0.013
c.1982C>A; p.T661K	Estonian short extreme	Novel	0.999

The second cohort of individuals (n=1,000) was taken from the extremes (<1.25th percentile or >98.75th percentile) of height distribution from the Estonian Biobank (~52,000 samples in total). Pooled targeted sequencing of the exons of four candidate genes was performed. As before, we did not attempt to validate or further evaluate the potential synonymous or intronic variants. One missense variant (c.739A>G (p.N247D)), previously identified in one FINRISK tall extreme sample, was observed in both tall extreme and short extreme samples. The remaining three missense variants (Table 4.3) were each found in single short individuals. These allelic variants were not found in the public databases. An *in silico*

analysis suggested that the three missense mutations are possibly damaging (c.2449G>A (p.E817K)), benign (c.2338G>A (p.G780R)), and probably damaging (c.1982C>A (p.T661K)), respectively.

Functional characterization of NPR2 variants

Overall, across four cohorts, we observed and validated 16 *NPR2* nonsynonymous variants (14 missense, 1 nonsense and 1 splice site), 14 in short stature samples, 1 in a tall extreme sample, 1 which was present in both short and tall extreme samples, and none in normal height controls (Table 4.4). To further examine the functional consequences of the identified *NPR2* nonysnonymous variants, we assessed the CNP-dependent cGMP-producing capacities of the 14 missense variants by transfecting HEK 293T cells, which do not normally express *NPR2,* using at least one of two methods. Relative guanylyl cyclase activities (±SDS) of these variants compared with wild type are listed in Table 4.5. The whole cell lysate method was less sensitive at detecting partial loss of function mutations. The *de novo* splice variant (c.1352-1G>A) and the nonsense variant (c.2710A>T (p.K904*)) were not tested and are presumed to be loss of function variants.

Table 4.4: Observation of *NPR2* nonsynonymous variants in four cohorts

Cohort	Observation of nonsynonymous variants	Note
Short stature patients and FHS controls	7 variants in 192 short stature patients 0 variants in 192 FHS controls	Patient 1: *de novo* mutation Patients 2-4: all heterozygous relatives had short stature (height SDS < -2) Patient 5: variant not present in average height mother and father's DNA unavailable Patient 6: variant not present in short mother and father's DNA unavailable Patient 7: no family data available
Second short stature cohort	4 variants in 216 short stature patients	No control samples
Extremes of FINRISK height distribution	0 variants in 136 short extremes 2 variants in 136 tall extremes	No family data available
Extremes of Estonian Biobank height distribution	3 variants in 500 short extremes 0 variant in 500 tall extremes	No family data available

Table 4.5: Functional characterization of *NPR2* nonsynonymous variants in three cohorts

Variant	Observation	Minor Allele Frequency	Segregation Within Family	Relative Guanylyl Cyclase Activities % (±SD) – Enzyme Assay with 1uM CNP	Relative Guanylyl Cyclase Activities % (±SD) – Whole Cell Lysate
c.1352-1G>A	Patient 1	Novel	*de novo*	N/A	N/A
c.142G>T; p.A48S	Patient 2	Novel	Yes	36.6 (14.9)*	93.5 (15.4)
c.1167G>T; p.E389D	Patient 3	Novel	Yes	57.0 (12.0)*	108.5 (24.6)
c.1481T>G; p.I494S	Patient 4	Novel	Yes	<6*	54.3 (8.6)*
c.1645G>A; p.A549T	Patient 5	Novel	Unknown	68.7 (26.9)	111.2 (9.6)
rs62637657: c.491C>G; p.A164G	Patient 6	0.02%	No	25.9 (10.3)*	130.6 (15.5)*
rs114147262: c.2359C>T; p.R787W	Patient 7	0.16%	N/A	62.8 (36.6)	90.9 (14.0)
c.316G>T; p.A106S	Patient	Novel	N/A	83.9 (32.5)	N/A
rs180950551: c.1802G>C; p.R601P	Patient	0.02%	N/A	64.5 (12.9)	N/A
c.2710A>T; p.K904X	Patient	Novel	N/A	N/A	N/A
rs114115939: c.1517G>A; p.R506H	Patient	0.06%	N/A	<6*	N/A
c.739A>G; p.N247D	FINRISK tall extreme (also observed in Estonian tall extreme and short extreme)	Novel	N/A	69.1 (37.3)	246.9 (66.5)*
c.1685G>A; p.R562Q	FINRISK tall extreme	Novel	N/A	145.7 (31.6)*	711.8 (238.2)*
c.2449G>A; p.E817K	Estonian short extreme	Novel	N/A	N/A	31.1 (0.3)*
c.2338G>A; p.G780R	Estonian short extreme	Novel	N/A	N/A	98.3 (11.9)
c.1982C>A; p.T661K	Estonian short extreme	Novel	N/A	N/A	65.7 (19.1)*

* $p < 0.05$.

To characterize the variants as likely nonpathogenic, variants of unknown significance, and likely pathogenic, we considered both the functional results and the pattern of segregation in the families, where available. For the six missense variants identified in the first short stature patient cohort, three of them segregate with short stature (Patient 2-4). For the variants in Patient 2-4, we observed decreased functional activity in at least one of the assays, strongly suggesting that these are pathogenic variants. The variant from patient 4 also showed decreased expression by Western blot. For the variant in Patient 5, the segregation is unknown and the functional data are negative. The variant in Patient 7, a known rare population variant, also had unchanged functional activity. We conclude that these two variants are unlikely to be pathogenic. The variant in Patient 6, which showed a slight increase in activity in the whole cell lysate assay and decreased activity in the enzyme based assay, was characterized as variant of unknown significance. Interestingly, three of the pathogenic variants were found amongst the 22 cases of familial ISS thus providing a diagnostic yield of 13.6%. However, only a single pathogenic variant, the de novo variant, was found amongst the 77 cases of non-familial ISS leading to a much lower diagnostic yield of 1.3%.

One of three missense variants in the second clinical short stature cohort (rs114115939: c.1517G>A (p.R506H)) decreased NPR-B activity likely secondary to decreased expression of the mutated receptor as seen on Western blot. Two out of three missense variants in the Estonian short extremes decreased NPR-B activity (c.2449G>A (p.E817K) and c.1982C>A (p.T661K)). Interestingly, one of two missense variants identified in FINRISK tall extremes showed increased NPR-B activity in both assays (c.1685G>A (p.R562Q)), strongly suggesting this is a gain of function variant contributing to the individuals' tall stature. The second variant, which showed increased NPR-B activity only in the whole cell lysate assay, was also observed in Estonian short extreme samples, making it even less likely to be pathogenic.

Verifying the role of heterozygous functional NPR2 variants in ISS individuals

To test formally whether there are more loss of function *NPR2* variants in short stature individuals than in controls without short stature, we combined observations from the three cohorts that have control samples (in other words, excluding the second clinical short stature

cohort), considering validated variants seen either only in cases or only in controls. In total, there are six likely loss-of-function mutations (five missense and one splice site) based on segregation and functional evidence, all observed in short samples, and one gain-of-function mutation in tall samples. Assuming that all variants are equally likely to occur in short stature samples and tall extreme samples (or control samples of normal height) under the null hypothesis, this observation gives a p value of 0.008 (one-tailed test, p value is the probability of the observed outcome i.e. 6:0 plus 0:1 and all more extreme). The estimated prevalence of loss of function *NPR2* variants in ISS individuals ranged from 1/26 to 0, and no variants were observed in the cohort of 136 short individuals from Finland (Table 4.6).

Table 4.6: Fraction of ISS individuals with clearly functional *NPR2* variants

Cohort	# of functional *NPR2* variants	# of ISS individuals	Fraction of ISS individuals with functional *NPR2* variants
Short stature patients and FHS controls	4	104	1/26
Second short stature cohort	2	216	1/108
Extremes of FINRISK height distribution	0	136	0
Extremes of Estonian Biobank height distribution	2	500	1/250

Discussion

To explore the role of *NPR2* variation in short stature, we screened for nonsynonymous *NPR2* variants in four different cohorts (Table 4.4). Familial segregation analyses in the short stature patient cohort supported the hypothesis that rare heterozygous *NPR2* loss of function variants contribute strongly to short stature in individuals carrying these variants. We also identified rare nonsynonymous *NPR2* variants in three additional cohorts, where family data are not available. Functional studies of the *NPR2* missense variants identified eight loss-of-function mutations in short stature samples and one gain-of-function mutation in tall stature samples. The frequencies of heterozygous mutation carriers in ISS individuals varied across cohorts from 1/26 to 0, and no variants were observed in the short individuals from Finland. This variability may be due to different ancestries, variable sample selection criteria, and

statistical fluctuation. Of note, for the second short stature cohort, coverage was not 100% for 5 out of 22 exons and exon 3 was not evaluated. This can reduce the observed frequency of allelic variants found in this group of patients.

In contrast with homozygous mutations in *NPR2*, which produces a severe short stature and body disproportion, heterozygous mutations in *NPR2* seem to be associated with mild and variable growth impairment without a distinct skeletal phenotype. In our studies, the severity of short stature, body proportions, and presence of nonspecific skeletal abnormalities varied across individuals, consistent with previous observations (22, 88). This variability is likely due to differences in the nature of *NPR2* mutations carried by the individuals, as well as variable expressivity. Data was available on growth hormone response in three of our patients with pathogenic *NPR2* variants and response was quite variable.

Previously, a study of 47 Brazilian patients identified heterozygous *NPR2* mutations in 6% of patients(88), while another study on 101 short stature Japanese patients identified mutations in 2% of patients(89). Functional analyses were performed in both studies to evaluate the pathogenicity and elucidate the molecular mechanisms of the identified mutations. However, both studies included small numbers of patients and lacked the controls needed to formally show an enrichment of loss of function variants in ISS patients. Our study, including much larger cohorts, controls, family data, and functional analyses, is a more comprehensive assessment of heterozygous mutations in *NPR2* as a potential cause of growth impairment in ISS patients.

Our study does show that different methods of analyzing the function of *NPR2* variants can produce different results. The whole cell lysate method under the conditions specified is less sensitive at detecting partial loss of function mutations compared to a more rigorous enzyme based assay conducted on membrane fractions of transfected cells using known concentrations of substrate. However, the enzyme based assay is performed in crude membranes from broken cell preparations and cannot evaluate the ability of the mutations to affect the concentrations of the receptors that are properly targeted to the cell surface. Therefore, mutations that reduce activity in the enzyme assay are likely to be damaging but this

assay may not detect mutations that reduce activity due to inappropriate plasma membrane targeting or mutations that affect receptor degradation.

The functional assays used do have limitations. Although the rat NPR2 construct is >98% identical to human NPR2 and identical at each of the amino acids mutated in this study, there remains the possibility that non-identical residues located near mutant residues could affect protein function. In addition, the assays do not provide mechanistic data for why these heterozygous mutations reduced NPR2 function. Neither assay is able to differentiate between loss-of-function due to haploinsuffiency and loss of function due to dominant negative effect of the mutant allele.

The frequency of heterozygous carriers of AMDM mutations was previously estimated to be about 0.14% (22). In the NHLBI Exome Sequencing Project, the cumulative frequency of all nonsynonymous *NPR2* variants is approximately 0.4% in ~4000 European American samples. The discrepancy between these two allele frequencies highlights the importance of functional studies, which can help distinguish neutral background variants from true loss of function variants. In our study, out of 14 rare nonsynonymous *NPR2* variants that were functionally tested, only 6 showed decreased functional activity. Although the variants with no *in vitro* functional consequences may still contribute to short stature, our *in vitro* findings are consistent with the 3-fold higher frequency of nonsynonymous variants in the population database compared with expectation based on estimated AMDM carrier rates.

Despite the rarity of AMDM mutations (and nonsynonymous *NPR2* variants in general), these loss-of-function alleles likely have relatively large effect sizes (loss-of-function alleles were previously estimated to reduce height by ~1.8 SDS (22)). Assuming an effect size of 1.8 SDS and a cumulative population allele frequency of 0.14%, these maximum-likelihood estimates predict that ~2.6% of individuals with height below -2 SDS would have a loss of function *NPR2* variant and that ~4.2% of individuals with height below the 1st percentile would carry such a variant. These estimates are somewhat higher than our empirical findings in the two clinical short stature cohorts combined (1.9%), and also substantially higher than our estimates from the Finnish and Estonia cohorts (0.3% of individuals below the ~1st percentile), suggesting that the effect size or the population allele frequencies previously estimated for loss-

of-function *NPR2* variants are inflated. The higher reported rates of *NPR2* mutations in previous studies (88, 89) could also be due to population differences, or more stringent selection criteria in earlier studies that made it more likely to ascertain individuals carrying *NPR2* loss of function mutations (such as high rates of familial short stature). Interestingly, we also identified one gain-of-function *NPR2* variant in tall extremes of FINRISK height distribution, suggesting *NPR2* gain of function may be a more common cause of tall stature than previously recognized.

AMDM patients have been posited to have an abnormality in the GH/IGF-1 system, characterized by low insulin-like growth factor 1 (IGF-1; OMIM 147440) levels, high growth hormone (GH) levels, and lack of a response to GH treatment (22). However, IGF-1 levels were not low in either the carriers of AMDM mutations (22, 88, 89) or in the subjects in our cohort with loss of function mutations for whom data were available. Larger prospective studies are needed to determine whether heterozygous carriers of clearly pathogenic *NPR2* variants will respond to growth hormone. Regardless, *NPR2* loss-of-function variants likely account for approximately 0.4-4% of all patients with short stature (height < -2 SDS), and should be considered in the diagnostic evaluation of patients presenting with ISS. Interestingly, we identified NPR2 loss-of-function variants in 13.6% of familial cases, suggesting that such variants will provide a frequent explanation for patients with familial ISS.

References

1. Dauber A, Hirschhorn JN. Genome-wide association studies in pediatric endocrinology. Hormone research in paediatrics. 2011;75(5):322-8. PubMed PMID: 21464565.

2. Cohen P, Rogol AD, Deal CL, Saenger P, Reiter EO, Ross JL, et al. Consensus statement on the diagnosis and treatment of children with idiopathic short stature: a summary of the Growth Hormone Research Society, the Lawson Wilkins Pediatric Endocrine Society, and the European Society for Paediatric Endocrinology Workshop. The Journal of clinical endocrinology and metabolism. 2008;93(11):4210-7. PubMed PMID: 18782877.

3. Alanay Y, Lachman RS. A review of the principles of radiological assessment of skeletal dysplasias. Journal of clinical research in pediatric endocrinology. 2011;3(4):163-78. PubMed PMID: 22155458; PubMed Central PMCID: PMC3245489.

4. Krakow D, Rimoin DL. The skeletal dysplasias. Genetics in medicine : official journal of the American College of Medical Genetics. 2010;12(6):327-41. PubMed PMID: 20556869.

5. David A, Hwa V, Metherell LA, Netchine I, Camacho-Hubner C, Clark AJ, et al. Evidence for a continuum of genetic, phenotypic, and biochemical abnormalities in children with growth hormone insensitivity. Endocr Rev. 2011;32(4):472-97. Epub 2011/04/29. PubMed PMID: 21525302.

6. Warman ML, Cormier-Daire V, Hall C, Krakow D, Lachman R, LeMerrer M, et al. Nosology and classification of genetic skeletal disorders: 2010 revision. American journal of medical genetics Part A. 2011;155A(5):943-68. PubMed PMID: 21438135; PubMed Central PMCID: PMC3166781.

7. Endo A. A historical perspective on the discovery of statins. Proceedings of the Japan Academy Series B, Physical and biological sciences. 2010;86(5):484-93. PubMed PMID: 20467214; PubMed Central PMCID: PMC3108295.

8. Gudbjartsson DF, Walters GB, Thorleifsson G, Stefansson H, Halldorsson BV, Zusmanovich P, et al. Many sequence variants affecting diversity of adult human height. Nature genetics. 2008;40(5):609-15. PubMed PMID: 18391951.

9. Lettre G, Jackson AU, Gieger C, Schumacher FR, Berndt SI, Sanna S, et al. Identification of ten loci associated with height highlights new biological pathways in human growth. Nature genetics. 2008;40(5):584-91. PubMed PMID: 18391950; PubMed Central PMCID: PMC2687076.

10. Weedon MN, Lango H, Lindgren CM, Wallace C, Evans DM, Mangino M, et al. Genome-wide association analysis identifies 20 loci that influence adult height. Nature genetics. 2008;40(5):575-83. PubMed PMID: 18391952; PubMed Central PMCID: PMC2681221.

11. 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;467(7317):832-8. PubMed PMID: 20881960; PubMed Central PMCID: PMC2955183.

12. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. Nature genetics. 2014;46(11):1173-86. PubMed PMID: 25282103; PubMed Central PMCID: PMC4250049.

13. Rimoin DL, Cohn D, Krakow D, Wilcox W, Lachman RS, Alanay Y. The skeletal dysplasias: clinical-molecular correlations. Annals of the New York Academy of Sciences. 2007;1117:302-9. PubMed PMID: 18056050.
14. Rappold G, Blum WF, Shavrikova EP, Crowe BJ, Roeth R, Quigley CA, et al. Genotypes and phenotypes in children with short stature: clinical indicators of SHOX haploinsufficiency. Journal of medical genetics. 2007;44(5):306-13. PubMed PMID: 17182655; PubMed Central PMCID: PMC2597980.
15. Chan Y, Holmen OL, Dauber A, Vatten L, Havulinna AS, Skorpen F, et al. Common variants show predicted polygenic effects on height in the tails of the distribution, except in extremely short individuals. PLoS genetics. 2011;7(12):e1002439. PubMed PMID: 22242009; PubMed Central PMCID: PMC3248463.
16. Hirschhorn JN, Lettre G. Progress in genome-wide association studies of human height. Hormone research. 2009;71 Suppl 2:5-13. PubMed PMID: 19407491.
17. Zahnleiter D, Uebe S, Ekici AB, Hoyer J, Wiesener A, Wieczorek D, et al. Rare copy number variants are a common cause of short stature. PLoS genetics. 2013;9(3):e1003365. PubMed PMID: 23516380; PubMed Central PMCID: PMC3597495.
18. Dauber A, Yu Y, Turchin MC, Chiang CW, Meng YA, Demerath EW, et al. Genome-wide association of copy-number variation reveals an association between short stature and the presence of low-frequency genomic deletions. Am J Hum Genet. 2011;89(6):751-9. PubMed PMID: 22118881; PubMed Central PMCID: PMC3234379.
19. Kuhlenbaumer G, Hullmann J, Appenzeller S. Novel genomic techniques open new avenues in the analysis of monogenic disorders. Human mutation. 2011;32(2):144-51. PubMed PMID: 21280146.
20. Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, et al. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nature genetics. 2010;42(9):790-3. PubMed PMID: 20711175; PubMed Central PMCID: PMC2930028.
21. 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. The Journal of clinical endocrinology and metabolism. 2009;94(5):1740-7. PubMed PMID: 19240156.
22. Olney RC, Bukulmez H, Bartels CF, Prickett TC, Espiner EA, Potter LR, et al. Heterozygous mutations in natriuretic peptide receptor-B (NPR2) are associated with short stature. The Journal of clinical endocrinology and metabolism. 2006;91(4):1229-32. PubMed PMID: 16384845.
23. Jarick I, Vogel CI, Scherag S, Schafer H, Hebebrand J, Hinney A, et al. Novel common copy number variation for early onset extreme obesity on chromosome 11q11 identified by a genome-wide analysis. Hum Mol Genet. 2011;20(4):840-52. PubMed PMID: 21131291; PubMed Central PMCID: PMC3024044.
24. International Schizophrenia C. Rare chromosomal deletions and duplications increase risk of schizophrenia. Nature. 2008;455(7210):237-41. PubMed PMID: 18668038; PubMed Central PMCID: PMC3912847.
25. Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R, et al. Functional impact of global rare copy number variation in autism spectrum disorders. Nature. 2010;466(7304):368-72. PubMed PMID: 20531469; PubMed Central PMCID: PMC3021798.

26. Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, et al. Exome sequencing as a tool for Mendelian disease gene discovery. Nat Rev Genet. 2011;12(11):745-55. Epub 2011/09/29. PubMed PMID: 21946919.
27. Bejjani BA, Shaffer LG. Clinical utility of contemporary molecular cytogenetics. Annu Rev Genomics Hum Genet. 2008;9:71-86. Epub 2008/10/25. PubMed PMID: 18949852.
28. Need AC, Shashi V, Hitomi Y, Schoch K, Shianna KV, McDonald MT, et al. Clinical application of exome sequencing in undiagnosed genetic conditions. Journal of medical genetics. 2012;49(6):353-61. Epub 2012/05/15. PubMed PMID: 22581936; PubMed Central PMCID: PMC3375064.
29. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754-60. Epub 2009/05/20. PubMed PMID: 19451168; PubMed Central PMCID: PMC2705234.
30. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297-303. Epub 2010/07/21. PubMed PMID: 20644199; PubMed Central PMCID: PMC2928508.
31. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature genetics. 2011;43(5):491-8. Epub 2011/04/12. PubMed PMID: 21478889; PubMed Central PMCID: PMC3083463.
32. Langmann T, Di Gioia SA, Rau I, Stohr H, Maksimovic NS, Corbo JC, et al. Nonsense mutations in FAM161A cause RP28-associated recessive retinitis pigmentosa. Am J Hum Genet. 2010;87(3):376-81. Epub 2010/08/14. PubMed PMID: 20705278; PubMed Central PMCID: PMC2933350.
33. (ESP) NESP. Exome Variant Server Seattle, WA2012 [cited 2012 February]. Available from: http://evs.gs.washington.edu/EVS/.
34. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010;7(4):248-9. Epub 2010/04/01. PubMed PMID: 20354512; PubMed Central PMCID: PMC2855889.
35. Bandah-Rozenfeld D, Mizrahi-Meissonnier L, Farhy C, Obolensky A, Chowers I, Pe'er J, et al. Homozygosity mapping reveals null mutations in FAM161A as a cause of autosomal-recessive retinitis pigmentosa. Am J Hum Genet. 2010;87(3):382-91. Epub 2010/08/14. PubMed PMID: 20705279; PubMed Central PMCID: PMC2933343.
36. Bernasconi F, Karaguzel A, Celep F, Keser I, Luleci G, Dutly F, et al. Normal phenotype with maternal isodisomy in a female with two isochromosomes: i(2p) and i(2q). Am J Hum Genet. 1996;59(5):1114-8. Epub 1996/11/01. PubMed PMID: 8900241; PubMed Central PMCID: PMC1914849.
37. Keller MC, McRae AF, McGaughran JM, Visscher PM, Martin NG, Montgomery GW. Non-pathological paternal isodisomy of chromosome 2 detected from a genome-wide SNP scan. American journal of medical genetics Part A. 2009;149A(8):1823-6. Epub 2009/07/18. PubMed PMID: 19610117.
38. Yamazawa K, Ogata T, Ferguson-Smith AC. Uniparental disomy and human disease: an overview. Am J Med Genet C Semin Med Genet. 2010;154C(3):329-34. Epub 2010/08/31. PubMed PMID: 20803655.

39. Petit FM, Gajdos V, Parisot F, Capel L, Aboura A, Lachaux A, et al. Paternal isodisomy for chromosome 2 as the cause of Crigler-Najjar type I syndrome. Eur J Hum Genet. 2005;13(3):278-82. Epub 2004/12/09. PubMed PMID: 15586176.
40. Kantarci S, Ragge NK, Thomas NS, Robinson DO, Noonan KM, Russell MK, et al. Donnai-Barrow syndrome (DBS/FOAR) in a child with a homozygous LRP2 mutation due to complete chromosome 2 paternal isodisomy. American journal of medical genetics Part A. 2008;146A(14):1842-7. Epub 2008/06/17. PubMed PMID: 18553518; PubMed Central PMCID: PMC2891749.
41. Bakker B, Bikker H, Hennekam RC, Lommen EJ, Schipper MG, Vulsma T, et al. Maternal isodisomy for chromosome 2p causing severe congenital hypothyroidism. The Journal of clinical endocrinology and metabolism. 2001;86(3):1164-8. Epub 2001/03/10. PubMed PMID: 11238503.
42. Castiglia D, Castori M, Pisaneschi E, Sommi M, Covaciu C, Zambruno G, et al. Trisomic rescue causing reduction to homozygosity for a novel ABCA12 mutation in harlequin ichthyosis. Clinical genetics. 2009;76(4):392-7. Epub 2009/08/12. PubMed PMID: 19664001.
43. Herzfeld T, Wolf N, Winter P, Hackstein H, Vater D, Muller U. Maternal uniparental heterodisomy with partial isodisomy of a chromosome 2 carrying a splice acceptor site mutation (IVS9-2A>T) in ALS2 causes infantile-onset ascending spastic paralysis (IAHSP). Neurogenetics. 2009;10(1):59-64. Epub 2008/09/24. PubMed PMID: 18810511.
44. Douglas GV, Wiszniewska J, Lipson MH, Witt DR, McDowell T, Sifry-Platt M, et al. Detection of uniparental isodisomy in autosomal recessive mitochondrial DNA depletion syndrome by high-density SNP array analysis. J Hum Genet. 2011;56(12):834-9. Epub 2011/10/21. PubMed PMID: 22011815.
45. Thompson DA, McHenry CL, Li Y, Richards JE, Othman MI, Schwinger E, et al. Retinal dystrophy due to paternal isodisomy for chromosome 1 or chromosome 2, with homoallelism for mutations in RPE65 or MERTK, respectively. Am J Hum Genet. 2002;70(1):224-9. Epub 2001/12/01. PubMed PMID: 11727200; PubMed Central PMCID: PMC384890.
46. Ozaki K, Fujiwara T, Nakamura Y, Takahashi E. Isolation and mapping of a novel human kidney- and liver-specific gene homologous to the bacterial acetyltransferases. J Hum Genet. 1998;43(4):255-8. Epub 1998/12/16. PubMed PMID: 9852678.
47. Ancian P, Lambeau G, Mattei MG, Lazdunski M. The human 180-kDa receptor for secretory phospholipases A2. Molecular cloning, identification of a secreted soluble form, expression, and chromosomal localization. J Biol Chem. 1995;270(15):8963-70. Epub 1995/04/14. PubMed PMID: 7721806.
48. Granata F, Petraroli A, Boilard E, Bezzine S, Bollinger J, Del Vecchio L, et al. Activation of cytokine production by secreted phospholipase A2 in human lung macrophages expressing the M-type receptor. J Immunol. 2005;174(1):464-74. Epub 2004/12/22. PubMed PMID: 15611272.
49. Hanasaki K, Yokota Y, Ishizaki J, Itoh T, Arita H. Resistance to endotoxic shock in phospholipase A2 receptor-deficient mice. J Biol Chem. 1997;272(52):32792-7. Epub 1998/01/31. PubMed PMID: 9407054.
50. Waters AM, Beales PL. Ciliopathies: an expanding disease spectrum. Pediatr Nephrol. 2011;26(7):1039-56. Epub 2011/01/07. PubMed PMID: 21210154; PubMed Central PMCID: PMC3098370.
51. Forsythe E, Beales PL. Bardet-Biedl syndrome. Eur J Hum Genet. 2012. Epub 2012/06/21. PubMed PMID: 22713813.

52. Fillastre JP, Guenel J, Riberi P, Marx P, Whitworth JA, Kunh JM. Senior-Loken syndrome (nephronophthisis and tapeto-retinal degeneration): a study of 8 cases from 5 families. Clin Nephrol. 1976;5(1):14-9. Epub 1976/01/01. PubMed PMID: 1248184.
53. Leitch CC, Zaghloul NA, Davis EE, Stoetzel C, Diaz-Font A, Rix S, et al. Hypomorphic mutations in syndromic encephalocele genes are associated with Bardet-Biedl syndrome. Nature genetics. 2008;40(4):443-8. Epub 2008/03/11. PubMed PMID: 18327255.
54. Dawe HR, Smith UM, Cullinane AR, Gerrelli D, Cox P, Badano JL, et al. The Meckel-Gruber Syndrome proteins MKS1 and meckelin interact and are required for primary cilium formation. Hum Mol Genet. 2007;16(2):173-86. Epub 2006/12/23. PubMed PMID: 17185389.
55. Tammachote R, Hommerding CJ, Sinders RM, Miller CA, Czarnecki PG, Leightner AC, et al. Ciliary and centrosomal defects associated with mutation and depletion of the Meckel syndrome genes MKS1 and MKS3. Hum Mol Genet. 2009;18(17):3311-23. Epub 2009/06/12. PubMed PMID: 19515853; PubMed Central PMCID: PMC2733821.
56. Zhang Q, Nishimura D, Seo S, Vogel T, Morgan DA, Searby C, et al. Bardet-Biedl syndrome 3 (Bbs3) knockout mouse model reveals common BBS-associated phenotypes and Bbs3 unique phenotypes. Proc Natl Acad Sci U S A. 2011;108(51):20678-83. Epub 2011/12/06. PubMed PMID: 22139371; PubMed Central PMCID: PMC3251145.
57. Vinogradov AV. [Differences in ways of development of myocardial infarct with severe and mild clinical course]. Kardiologiia. 1990;30(9):5-8. Epub 1990/09/01. PubMed PMID: 2273737.
58. Schaefer E, Zaloszyc A, Lauer J, Durand M, Stutzmann F, Perdomo-Trujillo Y, et al. Mutations in SDCCAG8/NPHP10 Cause Bardet-Biedl Syndrome and Are Associated with Penetrant Renal Disease and Absent Polydactyly. Mol Syndromol. 2011;1(6):273-81. Epub 2011/12/23. PubMed PMID: 22190896; PubMed Central PMCID: PMC3214956.
59. Kuczmarski R, Ogden C, Guo S. 2000 CDC growth charts for the United States: Methods and development. National Center for Health Statistics Vital Health Stat. 2002;11:246.

60. A map of human genome variation from population-scale sequencing. Nature. 2010;467(7319):1061-73. Epub 2010/10/29. PubMed PMID: 20981092; PubMed Central PMCID: PMC3042601.
61. Stenson PD, Ball E, Howells K, Phillips A, Mort M, Cooper DN. Human Gene Mutation Database: towards a comprehensive central mutation database. Journal of medical genetics. 2008;45(2):124-6. Epub 2008/02/05. PubMed PMID: 18245393.
62. Rock MJ, Prenen J, Funari VA, Funari TL, Merriman B, Nelson SF, et al. Gain-of-function mutations in TRPV4 cause autosomal dominant brachyolmia. Nature genetics. 2008;40(8):999-1003. Epub 2008/07/01. PubMed PMID: 18587396; PubMed Central PMCID: PMC3525077.
63. Romano AA, Allanson JE, Dahlgren J, Gelb BD, Hall B, Pierpont ME, et al. Noonan syndrome: clinical features, diagnosis, and management guidelines. Pediatrics.126(4):746-59. Epub 2010/09/30. PubMed PMID: 20876176.
64. Tartaglia M, Mehler EL, Goldberg R, Zampino G, Brunner HG, Kremer H, et al. Mutations in PTPN11, encoding the protein tyrosine phosphatase SHP-2, cause Noonan syndrome. Nature genetics. 2001;29(4):465-8. Epub 2001/11/13. PubMed PMID: 11704759.
65. Tartaglia M, Kalidas K, Shaw A, Song X, Musat DL, van der Burgt I, et al. PTPN11 mutations in Noonan syndrome: molecular spectrum, genotype-phenotype correlation, and

phenotypic heterogeneity. Am J Hum Genet. 2002;70(6):1555-63. Epub 2002/05/07. PubMed PMID: 11992261; PubMed Central PMCID: PMC379142.
66. Ester WA, van Duyvenvoorde HA, de Wit CC, Broekman AJ, Ruivenkamp CA, Govaerts LC, et al. Two short children born small for gestational age with insulin-like growth factor 1 receptor haploinsufficiency illustrate the heterogeneity of its phenotype. The Journal of clinical endocrinology and metabolism. 2009;94(12):4717-27. Epub 2009/10/30. PubMed PMID: 19864454.
67. Abuzzahab MJ, Schneider A, Goddard A, Grigorescu F, Lautier C, Keller E, et al. IGF-I receptor mutations resulting in intrauterine and postnatal growth retardation. N Engl J Med. 2003;349(23):2211-22. Epub 2003/12/06. PubMed PMID: 14657428.
68. Kawashima Y, Kanzaki S, Yang F, Kinoshita T, Hanaki K, Nagaishi J, et al. Mutation at cleavage site of insulin-like growth factor receptor in a short-stature child born with intrauterine growth retardation. The Journal of clinical endocrinology and metabolism. 2005;90(8):4679-87. Epub 2005/06/02. PubMed PMID: 15928254.
69. Inagaki K, Tiulpakov A, Rubtsov P, Sverdlova P, Peterkova V, Yakar S, et al. A familial insulin-like growth factor-I receptor mutant leads to short stature: clinical and biochemical characterization. The Journal of clinical endocrinology and metabolism. 2007;92(4):1542-8. Epub 2007/02/01. PubMed PMID: 17264177.
70. Golan D, Erlich Y, Rosset S. Weighted pooling--practical and cost-effective techniques for pooled high-throughput sequencing. Bioinformatics. 2012;28(12):i197-206. Epub 2012/06/13. PubMed PMID: 22689761; PubMed Central PMCID: PMC3371840.
71. 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). The Journal of clinical endocrinology and metabolism. 2011;97(2):E243-7. Epub 2011/12/02. PubMed PMID: 22130793.
72. Walenkamp MJ, van der Kamp HJ, Pereira AM, Kant SG, van Duyvenvoorde HA, Kruithof 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. The Journal of clinical endocrinology and metabolism. 2006;91(8):3062-70. Epub 2006/06/08. PubMed PMID: 16757531.
73. Kansra AR, Dolan LM, Martin LJ, Deka R, Chernausek SD. IGF receptor gene variants in normal adolescents: effect on stature. Eur J Endocrinol.167(6):777-81. Epub 2012/09/14. PubMed PMID: 22972910.
74. Abu-Amero S, Wakeling EL, Preece M, Whittaker J, Stanier P, Moore GE. Epigenetic signatures of Silver-Russell syndrome. Journal of medical genetics. 2010;47(3):150-4. Epub 2010/03/23. PubMed PMID: 20305090.
75. Schulz S. C-type natriuretic peptide and guanylyl cyclase B receptor. Peptides. 2005;26(6):1024-34. PubMed PMID: 15911070.
76. Potter LR, Abbey-Hosch S, Dickey DM. Natriuretic peptides, their receptors, and cyclic guanosine monophosphate-dependent signaling functions. Endocr Rev. 2006;27(1):47-72. PubMed PMID: 16291870.
77. Mericq V, Uyeda JA, Barnes KM, De Luca F, Baron J. Regulation of fetal rat bone growth by C-type natriuretic peptide and cGMP. Pediatric research. 2000;47(2):189-93. PubMed PMID: 10674345.

78. Chusho H, Tamura N, Ogawa Y, Yasoda A, Suda M, Miyazawa T, et al. Dwarfism and early death in mice lacking C-type natriuretic peptide. Proc Natl Acad Sci U S A. 2001;98(7):4016-21. PubMed PMID: 11259675; PubMed Central PMCID: PMC31171.
79. Krejci P, Masri B, Fontaine V, Mekikian PB, Weis M, Prats H, et al. Interaction of fibroblast growth factor and C-natriuretic peptide signaling in regulation of chondrocyte proliferation and extracellular matrix homeostasis. Journal of cell science. 2005;118(Pt 21):5089-100. PubMed PMID: 16234329.
80. Yasoda A, Komatsu Y, Chusho H, Miyazawa T, Ozasa A, Miura M, et al. Overexpression of CNP in chondrocytes rescues achondroplasia through a MAPK-dependent pathway. Nature medicine. 2004;10(1):80-6. PubMed PMID: 14702637.
81. Tamura N, Doolittle LK, Hammer RE, Shelton JM, Richardson JA, Garbers DL. Critical roles of the guanylyl cyclase B receptor in endochondral ossification and development of female reproductive organs. Proc Natl Acad Sci U S A. 2004;101(49):17300-5. PubMed PMID: 15572448; PubMed Central PMCID: PMC534612.
82. Bartels CF, Bukulmez H, Padayatti P, Rhee DK, van Ravenswaaij-Arts C, Pauli RM, et al. Mutations in the transmembrane natriuretic peptide receptor NPR-B impair skeletal growth and cause acromesomelic dysplasia, type Maroteaux. Am J Hum Genet. 2004;75(1):27-34. PubMed PMID: 15146390; PubMed Central PMCID: PMC1182004.
83. Bocciardi R, Giorda R, Buttgereit J, Gimelli S, Divizia MT, Beri S, et al. Overexpression of the C-type natriuretic peptide (CNP) is associated with overgrowth and bone anomalies in an individual with balanced t(2;7) translocation. Human mutation. 2007;28(7):724-31. PubMed PMID: 17373680.
84. Moncla A, Missirian C, Cacciagli P, Balzamo E, Legeai-Mallet L, Jouve JL, et al. A cluster of translocation breakpoints in 2q37 is associated with overexpression of NPPC in patients with a similar overgrowth phenotype. Human mutation. 2007;28(12):1183-8. PubMed PMID: 17676597.
85. Miura K, Kim OH, Lee HR, Namba N, Michigami T, Yoo WJ, et al. Overgrowth syndrome associated with a gain-of-function mutation of the natriuretic peptide receptor 2 (NPR2) gene. American journal of medical genetics Part A. 2014;164A(1):156-63. PubMed PMID: 24259409.
86. Hannema SE, van Duyvenvoorde HA, Premsler T, Yang RB, Mueller TD, Gassner B, et al. An activating mutation in the kinase homology domain of the natriuretic peptide receptor-2 causes extremely tall stature without skeletal deformities. The Journal of clinical endocrinology and metabolism. 2013;98(12):E1988-98. PubMed PMID: 24057292.
87. Robinson JW, Dickey DM, Miura K, Michigami T, Ozono K, Potter LR. A human skeletal overgrowth mutation increases maximal velocity and blocks desensitization of guanylyl cyclase-B. Bone. 2013;56(2):375-82. PubMed PMID: 23827346.
88. Vasques GA, Amano N, Docko AJ, Funari MF, Quedas EP, Nishi MY, et al. Heterozygous mutations in natriuretic peptide receptor-B (NPR2) gene as a cause of short stature in patients initially classified as idiopathic short stature. The Journal of clinical endocrinology and metabolism. 2013;98(10):E1636-44. PubMed PMID: 24001744.
89. Amano N, Mukai T, Ito Y, Narumi S, Tanaka T, Yokoya S, et al. Identification and functional characterization of two novel NPR2 mutations in Japanese patients with short stature. The Journal of clinical endocrinology and metabolism. 2014;99(4):E713-8. PubMed PMID: 24471569.

90. 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. The Journal of clinical endocrinology and metabolism. 2013;98(8):E1428-37. PubMed PMID: 23771920; PubMed Central PMCID: PMC3733853.
91. Vartiainen E, Laatikainen T, Peltonen M, Juolevi A, Mannisto S, Sundvall J, et al. Thirty-five-year trends in cardiovascular risk factors in Finland. International journal of epidemiology. 2010;39(2):504-18. PubMed PMID: 19959603.
92. Leitsalu L, Haller T, Esko T, Tammesoo ML, Alavere H, Snieder H, et al. Cohort Profile: Estonian Biobank of the Estonian Genome Center, University of Tartu. International journal of epidemiology. 2014. PubMed PMID: 24518929.
93. Rivas MA, Beaudoin M, Gardet A, Stevens C, Sharma Y, Zhang CK, et al. Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. Nature genetics. 2011;43(11):1066-73. Epub 2011/10/11. PubMed PMID: 21983784; PubMed Central PMCID: PMC3378381.
94. Genomes Project C, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, et al. A map of human genome variation from population-scale sequencing. Nature. 2010;467(7319):1061-73. PubMed PMID: 20981092; PubMed Central PMCID: PMC3042601.
95. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. Nucleic acids research. 2002;30(17):3894-900. PubMed PMID: 12202775; PubMed Central PMCID: PMC137415.
96. Yoder AR, Kruse AC, Earhart CA, Ohlendorf DH, Potter LR. Reduced ability of C-type natriuretic peptide (CNP) to activate natriuretic peptide receptor B (NPR-B) causes dwarfism in lbab -/- mice. Peptides. 2008;29(9):1575-81. PubMed PMID: 18554750.
97. Robinson JW, Potter LR. ATP potentiates competitive inhibition of guanylyl cyclase A and B by the staurosporine analog, Go6976: reciprocal regulation of ATP and GTP binding. J Biol Chem. 2011;286(39):33841-4. PubMed PMID: 21828054; PubMed Central PMCID: PMC3190780.