Growth Hormone & IGF Research 38 (2018) 24–28

Contents lists available at ScienceDirect

Growth Hormone & IGF Research

journal homepage: www.elsevier.com/locate/ghir

New genetic tools in the diagnosis of growth defects

Andrew Dauber

Cincinnati Center for Growth Disorders, Division of Endocrinology, Cincinnati Children's Hospital Medical Center, Cincinnati, USA
[a] *Division of Endocrinology, Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, MLC 7012, Cincinnati, OH, 45229., United States*

ARTICLE INFO

Keywords:
Growth
Short stature
Genomics
Exome sequencing
Copy number variation

ABSTRACT

Growth is a complex biological process governed by thousands of genes. Genetic defects in a wide array of genes can cause severe growth disorders. Genomic technologies including chromosomal microarrays and whole exome sequencing have revolutionized our ability to diagnose growth disorders. In this brief review, we will discuss each of these technologies and how they have been applied in the field of growth disorders.

Introduction

Short stature is one of the most common reasons for referral to pediatric endocrinologists. The traditional pediatric endocrinologist's evaluation of patients with short stature has focused on hormonal etiologies of growth failure such as hypothyroidism and GH deficiency. Current consensus guidelines [1] regarding the evaluation of children presenting with idiopathic short stature (ISS) include a comprehensive history and physical examination followed by a wide panel of screening laboratory tests looking for signs of renal dysfunction, inflammatory disorders, thyroid problems, celiac disease, and disorders of calcium/bone mineral metabolism. Additionally, detailed investigation of the GH/IGF-I axis is recommended, potentially including GH stimulation testing. Outside of a karyotype in females to exclude Turner Syndrome, genetic testing is only mentioned for clearly identified genetic syndromes with a known etiology (such as Noonan syndrome or Laron Syndrome). Implementation of the consensus guidelines [1] including comprehensive laboratory testing in a large pediatric referral center yielded a diagnosis in only 1% of patients at an estimated cost of approximately US$ 100,000 per diagnosis [2]. Based on this data, one must question whether searching for genetic etiologies of short stature in these patients would lead to a higher diagnostic rate and potentially be more cost effective.

Growth is a highly complex biological process and there are literally thousands of genes involved in the regulation of human growth [3–5]. Many children who present with ISS have mild variants of normal physiology such as mild familial short stature or constitutional delay of growth and puberty. In genetic terms, these children likely have polygenic short stature. This means that they have thousands of genetic polymorphisms which in sum are the major determinants of their final height. These individuals are unlikely to have single genetic variants (i.e. monogenic etiologies) that are having a large effect on their growth. In contrast, there are a subgroup of patients with short stature who do have monogenic causes, and it is those patients who would most benefit from genetic testing. It is often difficult to distinguish between these two groups, but in general, patients with monogenic etiologies are more likely to have severe short stature or other distinctive physical or biochemical features. For a more detailed discussion of the genetic evaluation of short stature, one can refer to other recent reviews on the subject [6,7].

In recent years, there has been tremendous progress in our ability to investigate genetic etiologies of human disease. Two technologies have truly revolutionized the field allowing for rapid assessment of genome-wide copy number changes and sequence variants. These technologies are chromosomal microarrays and next-generation sequencing, and in particular whole exome sequencing. We will briefly review each of these technologies and then discuss their roles in the evaluation of growth disorders.

Copy Number Variation:

Copy number variation refers to areas of the genome where instead of carrying two copies of a specific piece of DNA, an individual carries either more copies (i.e. duplications) or is missing one or both copies (i.e. deletions). According to data from the HapMap project, every person carries on average 3.5 million bases of DNA in which they don't have two copies [8]. This represent ~0.1% of the human genome and approximately one-third of this copy number variation overlaps with protein coding genes [8]. It has now been well established that copy number variants can be responsible for a wide range of human diseases. In order to interrogate genome-wide copy number variation, a number of technologies have been developed. The most commonly used of these is called a chromosomal microarray. There are a number of technical approaches to the design of a microarray with array comparative

E-mail address: andrew.dauber@cchmc.org.

https://doi.org/10.1016/j.ghir.2017.11.002
Received 5 September 2017; Received in revised form 30 October 2017; Accepted 12 November 2017
Available online 13 November 2017

genomic hybridization (CGH) being one of the most frequently used in clinical practice. In array CGH, a chip is used which contains thousands of probes (or even hundreds of thousands). These probes are essentially complimentary pieces of DNA that are designed to cover either the entire genome or specific areas of the genome. The probes are often enriched in areas of known disease genes or genomic disorders. The patient's DNA sample is typically labeled with a fluorophore (i.e. a specific color) while control DNA is labeled with an alternate fluorophore. The patient's and control sample are then hybridized to the array and a digital imager is used to look at the intensity of the two fluorophores at all of the various probes. This data can then be processed and the ratio of fluorophore intensity will equate to the amount of patient DNA versus control DNA at specific regions of the genome. In areas where the patient has a duplication, his or her DNA fluorophore will have increased intensity compared to the control, with the opposite occurring at sites of deletions.

To explore the role of copy number variation in clinical short stature, in 2011, our group examined all patients who had undergone clinical chromosomal microarray testing at Boston Children's Hospital and who had height data available in the electronic medical record system (n = 4411) [9]. We then characterized each patient as having short, tall or normal stature using a height cut-off of plus or minus 2 standard deviations for tall and short stature, respectively. There were 425 individuals with short stature, 196 with tall stature, and 3800 individuals with normal stature. We then calculated the total amount of copy number variation for each individual and subdivided this into deletions and duplication. We further subdivided these based on frequency of the specific copy number variants. We found that individuals with short stature had a larger total burden of copy number variation as well as a larger average size of copy number variants. Upon further stratification, we found that these associations were completely driven by an increase in low frequency (frequency < 5%) and rare (frequency < 1%) deletions but not common deletions or duplications of any frequency. Interestingly, we then extended our findings to a group of population-based cohorts where we found that the global burden of lower frequency deletions was associated with shorter stature in the general population.

Subsequent to this publication, a number of groups have gone on to examine the yield of chromosomal microarray testing in patients referred for the evaluation of short stature. Zahnleiter et al. studied 200 patients presenting with idiopathic short stature, 108 of which were familial [10]. They found likely causal copy number variants in 20 of the patients giving an overall yield of 10%. These were all rare copy number variants, 10 duplications and 10 deletions. In a similar study, van Duyvenvoorde et al. studied 149 families presenting with short stature [11]. A definitive pathologic copy number variant was found in 6 of these families with another 33 families having one or more potentially pathogenic copy number variant. Taken together, these studies clearly demonstrate that rare copy number variants play an important role in the etiology of short stature. Another smaller study showed similar results in a population of patients born small for gestational age without a known cause [12].

Whole Exome Sequencing:

In addition to copy number variants, there are clearly rare genetic mutations that can cause short stature. There are literally hundreds of disorders listed in the Online Mendelian Inheritance in Man (omim.org) catalogue which are associated with short stature. Many of these are due to skeletal defects but, as noted above, there are a multitude of biological processes involved in growth. As a clinician, it can be extremely difficult to know the clinical characteristics of all of these disorders or to identify milder clinical presentations of these disorders. Therefore, it is important to be able to comprehensively evaluate a patient for defects in a wide range of genes and biological pathways. Whole exome sequencing is a powerful tool which provides that opportunity.

There are estimated to be approximately 20,000 protein coding genes in the genome. In the past, a clinician had to decide on a specific gene or small panel of genes to test via standard Sanger sequencing. In cases where a strong candidate gene can be identified based on clinical suspicion, this approach works quite well. However, when one is uncertain of the diagnosis or there are a large number of genes which could lead to a similar phenotype, then this approach can become quite tedious. Whole exome sequencing seeks to examine all of the protein coding genes in the genome simultaneously. To accomplish this goal, a DNA sample is obtained from the patient and preferably both parents if available. The DNA sample is then fragmented into small pieces. The next step is to extract from the entire genomic DNA only the portion that is necessary for sequencing, i.e. the protein coding regions of the genome – all the exons of the protein coding genes (i.e the exome). This is typically accomplished through a hybridization capture approach in which complementary pieces of DNA have been designed to target all of the protein coding exons in the genome. This can either be done on a chip based array or in solution. Regardless of the specific methodology, the selected DNA is then placed in a next generation sequencing machine. Again, there are a number of different technological approaches to next generation sequencing, but the basic idea is that the machine generates millions of short sequences of DNA matching your patient's DNA. Then using a complicated computer algorithm technique called mapping, the individual short sequences of DNA are matched to the appropriate location of the reference human genome. The patient's DNA sequence is then compared to the reference sequence and a list of genetic variants is generated. The length of that list depends on the specifics of the exome capture, the racial and ethnic background of the patient, and the sequencing methodology used, but there are typically tens of thousands of variants identified per individual.

Once the list of variants has been generated, variants need to be filtered and then prioritized based on their likelihood of causing the patient's phenotype. A typical filtering workflow is shown in Fig. 1. The first step is removing variants that are likely artifacts from the sequencing process. Next, variants are eliminated based on their frequency in the general population. There are now large exome sequencing datasets available that can provide reliable population based allele frequency information. For example, the Exome Aggregation Consortium has a publically available website (exac.broadinstitute.org) which provides allele frequency data from over 60,000 individuals. One should be cognizant to use a database that matches your patient's ethnic background. Specific allele frequency thresholds will vary based on the severity of the patient's phenotype and the inheritance pattern in the family. For example, if one is caring for a family in which there is a dominantly inherited pattern of short stature and the individual patients have heights of − 3 standard deviations, then any variant present in a public database with a frequency above 1 in a million is extremely unlikely to be causing this family's phenotype. Once variants above the allele frequency threshold have been removed, the next step is to focus on nonsynonymous variants. These are variants which are predicted to change the protein sequence as opposed to synonymous variants which are “silent” with respect to the protein code. Finally, it is critical to segregate the candidate variants in all available family members to make sure that the genetic data matches the clinical affectation status in the family. Once this is done, there will now be a list of candidate variants in candidate genes. Some of these may be clearly pathogenic variants in a gene which matches the patient's phenotype in which case a firm diagnosis can be made. However, often the results will discover a variant of uncertain significance in a gene with a plausible relationship to the phenotype. These variants and genes will require further translational research to truly understand their effects and to make firm conclusions as to whether they are causing the patient's phenotype.

To date, whole exome sequencing has primarily been used as a means of identifying novel genetic etiologies of growth disorders. However, there are a number of pilot studies examining its use for clinical diagnostics in growth disorders with larger studies underway. In 2013, we first reported the use of exome sequencing to make the