Next Generation Sequencing in Congenital Heart Disease: Gene Discovery and Clinical Application

Gillian M. Blue1,2,3 and David S. Winlaw1,2,3*

1Kids Heart Research, The Children’s Hospital at Westmead, Sydney, Australia
2Heart Centre for Children, The Children’s Hospital at Westmead, Sydney, Australia
3Sydney Medical School, University of Sydney, Australia

Abstract

Congenital Heart Disease (CHD) affects 6-8 per 1000 live births. It is the most common cause of death in newborn infants and is a significant burden on society with an increasing adult population. A number of genes have been implicated in disease development; however traditional research techniques such as linkage analysis are generally not suited as they require large families with multiple affected individuals which are rare in CHD. While more recent techniques, such as GWAS, provide insight into population-attributable risks, the information cannot be directly related to the affected individual. In comparison, the information gained from next generation sequencing (NGS) technology can directly impact the treatment and care of patients and their families. For this reason NGS is rapidly becoming the analysis of choice in both clinical and research arenas. The different types of NGS techniques have unique sets of advantages and disadvantages and it is important to assess these in the context of the disease characteristics, as well as the purpose of the analysis, so that the most appropriate technology is selected. Based on our experience and recent findings from research in CHD, we propose that a gene panel-based approach is currently more suited to clinical workflow and the molecular analysis of CHD and that exome sequencing and whole genome sequencing are best suited to a research environment that can provide the necessary support and analyses required for novel gene discovery.

Keywords: Next generation sequencing; Gene panels; Congenital heart disease; Molecular diagnosis

Introduction

Congenital Heart Disease (CHD) affects 6-8 per 1000 live births [1]. It is the most common cause of death in newborn infants in the western world and is also a significant cause of fetal loss. A third of live-born infants will require surgery within the first year of life; some requiring many more surgeries throughout adulthood. Due to improvements in surgical techniques and treatment, it is now generally accepted that adult CHD constitutes the majority of the CHD population in high-income countries [2]. CHD places a significant burden on the affected individual and their families with numerous studies reporting impairments in psychosocial functioning consistent with clinical criteria for depression, anxiety and somatisation [3-5]. CHD also places a significant burden on the community both in terms of providing care and funding for treatment.

Early Research Techniques

The high heritability of CHD (0.6-0.7) suggest a strong genetic component and current research points to both genetic and environmental contributions to disease development in the majority of cases, which tend to be sporadic in nature. CHD is a complex disease in that it often displays genetic heterogeneity, variable penetrance and variable expression. Approximately 20% of cases present as part of a genetic syndrome or chromosomal abnormality and for these individuals the ‘reason’ or ‘cause’ for their CHD is generally understood [6]. A small portion of cases exhibit familial forms of disease with seemingly Mendelian inheritance patterns. Most of the genetic information we know about CHD to date has been obtained from studying familial forms of disease due to their suitability to early research techniques, including linkage analysis and candidate gene studies [7-10]. While linkage analysis is a powerful tool and has successfully uncovered novel disease genes in many genetic conditions, it does have limitations. A prerequisite to using this technique is large families with multiple affected individuals. Linkage also requires the disease to segregate according to Mendelian principles with complete penetrance. Linkage analysis is therefore not ideal for studies on complex diseases such as CHD, as large families with multiple affected individuals, segregating disease according to Mendelian principles are rare.

Recent Research Techniques

More recent techniques, such as genome-wide association studies (GWAS), have provided evidence that common genetic variation can influence the risk of certain types of CHD [11-13]. However, while GWAS provide insight into population-attributable risks, the information is not directly transferable to the individual with a heart defect. Advances in sequencing technology, combined with the steady decline in costs, have resulted in the development of new-age sequencing technologies that are collectively termed next generation sequencing (NGS). The techniques can be targeted (exome sequencing and disease-specific gene panels), or non-targeted (whole genome sequencing (WGS)). While all types of NGS use massively parallel sequencing, the targeted approaches include a prior capture and enrichment step ensuring only targeted regions are sequenced.

NGS is well suited to the study of complex, heterogeneous diseases such as CHD. First and foremost, and in contrast to GWAS, the results of NGS are directly applicable to the affected individual. This information has the potential to provide diagnostic prospects and

*Corresponding author: David S. Winlaw, Heart Centre for Children, The Children’s Hospital at Westmead, Locked Bag 4001, Westmead NSW 2145, Australia, Tel: +61 2 9635 5298; Fax: +612 9845 2163; E-mail: david.winlaw@health.nsw.gov.au

Received March 16, 2015; Accepted April 21, 2015; Published April 23, 2015


Copyright: © 2015 Blue GM, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
directly impact the treatment and care of patients and their families. Secondly, the technology enables the rapid simultaneous interrogation of multiple genes, or even the entire genome in the case of WGS, which is particularly appealing to the study of heterogeneous diseases. Finally, NGS is (in most cases) unbiased in that it permits the discovery of both known and novel variants and disease genes, thereby promoting gene discovery.

**Exome Sequencing in CHD**

A number of studies have successfully applied NGS to CHD, more specifically exome sequencing. Zaidi et al. identified a significant increase in de novo coding variants in genes expressed in the developing heart in over 300 sporadic cases of CHD [14]. The study was important in defining the mechanisms of causation with a higher burden of variation evident in genes involved in cardiac development. However, while these findings contribute significantly towards the overall understanding of CHD and developmental pathways involved, they did not provide a cause of CHD for the majority of families included in this study.

Other studies have focused on individual families with presumed Mendelian forms of CHD [15-20] Table 1. Exome sequencing in families exhibiting Mendelian inheritance assists greatly with analysis as the presumed mode of inheritance can be used as part of the filtering strategy. In this way, exome sequencing successfully identified pathogenic variants in most of these families; however, the evidence on whether the variant identified was solely responsible for the CHD, is questionable in some cases. Together, these studies highlight both the successes as well as the complexities involved in applying exome sequencing to the study of complex, heterogeneous CHD. Despite sophisticated study designs and filtering strategies, reduced penetrance and variable expression present significant hurdles for analysis. The biggest obstacle during any exome analysis is to single out the causal variant from the thousands of variants identified during sequencing. In order to assist with this, strategies are applied to filter out the unlikely variants; however, these strategies are based upon a number of assumptions, including the assumption that the causal variant is fully penetrant. This may not always be the case in CHD, despite the family appearing to be ‘Mendelian’. Familial forms of CHD are known to display variable expression, various ages of presentation as well as variable penetrance [21] which complicate the exome sequencing analysis. Furthermore, in any exome analysis, the filtering process almost always contains a degree of bias in that it is limited to what is known; whether this is filtering against variants in known databases, retaining variants deemed to be pathogenic following in silico analyses or during educated variant curation.

NGS technology, more specifically exome sequencing and WGS, have some disadvantages that complicate their use in general, particularly in the clinical setting. From a technical perspective, issues with coverage, both in terms of depth as well as capture of on-target regions, mean that not all regions of the exome or genome have been covered to an appropriate depth, if at all. It is therefore possible that a causal variant is located in an uncovered, or poorly covered, region and will be missed [22]. Exome sequencing on-target coverage is generally reported in the range of 95%-97% [23] which is a concern for clinicians as causal variants may be missed if they fall in regions of poor, or no, coverage. A recent study evaluated the exome sequencing coverage of clinically relevant cardiac genes, including 31 genes present on the American College of Medical Genetics (ACMG) list of genes for reporting of incidental findings [24] as well as an additional 19 genes related to CHD [25]. They found that there was considerable variation in coverage of the individual genes; with 10 genes, including six from the ACMG list, covered less than the clinically required 30x.

Another concern is the lengthy data analyses and the issues with data storage, particularly in regards to cost and the resources required. These issues hinder efficient translation into a clinical setting.

Finally, there is also the issue of ‘incidental findings’. Sequencing an individual’s entire coding genome means that in some cases variants associated with a disease other than the one in question may be identified. For this reason, lengthy discussions about this possibility need to be conducted with participants or patients to ensure their preferences are met should an ‘incidental finding’ be identified. This can be very time-consuming and may result in significant delays, particularly from a research perspective.

**Gene Panels in CHD**

There is no question about the power that exome sequencing and WGS offer the investigator in terms of novel gene discovery due to their unbiased approach; however issues relating to coverage, analysis and storage of large amounts of data, as well as reporting of incidental findings complicate their use in the clinical setting [26]. In comparison, limiting the capture regions to only those known to be associated with the disease(s) of interest mitigates some of these issues; however this is at the cost of identifying novel disease genes.

Specific tools enable the design of individualised gene panels so that

<table>
<thead>
<tr>
<th>Study</th>
<th>CHD type</th>
<th>Sample size</th>
<th>Suspected inheritance</th>
<th>Analysis type</th>
<th>Likely causal variant identified</th>
<th>Novel/ Known variant</th>
<th>Segregation with disease</th>
<th>In silico predictions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arrington et al.</td>
<td>ASD, VSD, Coarc</td>
<td>2</td>
<td>Dominant</td>
<td>ES</td>
<td>MYH6 (p.Ala290Pro)</td>
<td>Known</td>
<td>No</td>
<td>Damaging</td>
</tr>
<tr>
<td>Liu et al.</td>
<td>ASD</td>
<td>1</td>
<td>Dominant</td>
<td>ES and CHD gene filter</td>
<td>TBX20 (p.Asp176Asn)</td>
<td>Novel</td>
<td>Yes</td>
<td>Damaging</td>
</tr>
<tr>
<td>Ta-Shma et al.</td>
<td>TA</td>
<td>1</td>
<td>Recessive</td>
<td>ES and homozygosity mapping</td>
<td>PLXND1 (p.Arg1299Cys)</td>
<td>Novel</td>
<td>Yes*</td>
<td>Damaging</td>
</tr>
<tr>
<td>Greenway et al.</td>
<td>ASD</td>
<td>2</td>
<td>Dominant</td>
<td>ES</td>
<td>ACTC1 (p.Met178Leu)</td>
<td>Novel</td>
<td>Yes</td>
<td>Damaging and Benign</td>
</tr>
<tr>
<td>Martin et al.</td>
<td>BAV, VSD, AVSD</td>
<td>17</td>
<td>Dominant</td>
<td>ES</td>
<td>No</td>
<td>n/a</td>
<td>No</td>
<td>n/a</td>
</tr>
<tr>
<td>Tariq et al.</td>
<td>L-TGA, heterotaxy</td>
<td>1</td>
<td>Dominant</td>
<td>ES and homozygosity mapping</td>
<td>SHROOM3 (p.Gly90Val)</td>
<td>Novel</td>
<td>Unknown*</td>
<td>Damaging</td>
</tr>
</tbody>
</table>

*No other affected individuals available for testing; # Only proband tested.

ASD: Atrial Septal Defect; AVSD: Atrioventricular Septal Defect; BAV: Bicuspid Aortic Valve; CHD: Congenital Heart Disease; Coarc: Coarctation of the Aorta; ES: Exome Sequencing; L-TGA: Congenitally Corrected Transposition of the Great Arteries; n/a: not applicable; TA: Truncus Arteriosus; VSD: Ventricular Septal Defect

Table 1: Summary of the available literature on exome sequencing in familial forms of CHD.
parameters can be individually adjusted to ensure maximum coverage of on-target regions. The substantial reduction in sequencing of a gene panel also allows for a significantly increased coverage depth compared with exome sequencing. This is usually ten-fold, or more, higher than the average coverage achieved using exome sequencing [26,27]. A higher depth of coverage significantly reduces the false positive rate, resulting in higher confidence in variant calling. In some areas small regions of the target genes may, however, still be missed. This is usually due to probe failure or issues with regions of high GC content. Being able to supplement these small regions with Sanger sequencing further ensures that all target regions have been adequately covered and increases confidence with results.

Gene panels also produce substantially less data for analysis and storage, making them an arguably more suitable approach in the diagnostic arena [28]. Furthermore, issues relating to reporting of incidental findings can be avoided entirely as only disease-relevant genes are screened.

Conclusion

With the steady decline in sequencing costs, exome sequencing has entered the clinical arena and WGS will, no doubt, follow suit in the coming years. We have experience with both exome sequencing (unpublished work) and gene panels in CHD [29] and our findings suggest that at present, a gene panel-based approach is more suited to the clinical and molecular analysis of CHD, at least as an initial screen, followed by exome sequencing in the event that no candidate variants are identified. Exome sequencing and WGS, on the other hand, are currently better suited to a research environment where novel gene discovery can be supported with appropriate functional studies to confirm and substantiate findings. Our experience is; however, limited to CHD, which is a complex disease, so this may not be the case for other genetic diseases with ‘simpler’ inheritances. Exome sequencing analysis is more straightforward in true Mendelian forms of disease where complicating factors such as incomplete penetrance and possible di-, oligo- or polygenic inheritances can be avoided. Furthermore, being able to identify both known and novel variants, using exome sequencing and WGS cannot be underestimated, especially in gene discovery research.

In summary, NGS technologies are revolutionising the field of genetics and genetic medicine. While the different techniques (WGS, exome sequencing and gene panels) use the same sequencing processes, they differ in their capture methods and therefore each present a role of demographic, clinical, and psychological factors in determining service uptake. Exome sequencing and WGS cannot be underestimated, especially in gene discovery and clinical application. Next Generation Sequenc & Applic, an open access journal


