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

Gillian M. Blue[1,2,3] and David S. Winlaw[1,2,3*]

[1]*Kids Heart Research, The Children's Hospital at Westmead, Sydney, Australia*
[2]*Heart Centre for Children, The Children's Hospital at Westmead, Sydney, Australia*
[3]*Sydney 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.

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

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

Study	CHD type	Sample size	Suspected inheritance	Analysis type	Likely causal variant identified	Novel/ Known variant	Segregation with disease	In silico predictions
Arrington et al.	ASD, VSD, Coarc	2	Dominant	ES	MYH6 (p.Ala290Pro)	Known	No	Damaging
Liu et al.	ASD	1	Dominant	ES and CHD gene filter	TBX20 (p.Asp176Asn)	Novel	Yes	Damaging
Ta-Shma et al.	TA	1	Recessive	ES and homozygosity mapping	PLXND1 (p.Arg1299Cys)	Novel	Yes*	Damaging
Greenway et al.	ASD	2	Dominant	ES	ACTC1 (p.Met178Leu)	Novel	Yes	Damaging and Benign
Martin et al.	BAV, VSD, AVSD	17	Dominant	ES	No	n/a	No	n/a
Tariq et al.	L-TGA, heterotaxy	1	Dominant	ES and homozygosity mapping	SHROOM3 (p.Gly60Val)	Novel	Unknown#	Damaging

*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 unique set of advantages and disadvantages. For this reason, the type of NGS technology applied to a genetic condition, needs to be carefully evaluated both in the context of the disease characteristics, as well as the purpose of the analysis i.e. whether it is clinical/diagnostic or research driven.

References

1. Hoffman JI, Kaplan S (2002) The incidence of congenital heart disease. J Am Coll Cardiol 39: 1890-1900.

2. Marelli AJ, Ionescu-Ittu R, Mackie AS, Guo L, Dendukuri N, et al. (2014) Lifetime Prevalence of Congenital Heart Disease in the General Population From 2000 to 2010. Circulation130: 749-756.

3. Lawoko S, Soares JJ (2002) Distress and hopelessness among parents of children with congenital heart disease, parents of children with other diseases, and parents of healthy children. J Psychosom Res 52: 193-208.

4. Kasparian NA, Fidock B, Sholler GF, Camphausen C, Murphy DN, et al. (2014) Parents' perceptions of genetics services for congenital heart disease: the role of demographic, clinical, and psychological factors in determining service attendance. Genet Med 16: 460-468.

5. Blue GM, Kasparian NA, Sholler GF, Kirk EP, Winlaw DS (2015) Genetic counselling in parents of children with congenital heart disease significantly improves knowledge about causation and enhances psychosocial functioning. Int J Cardiol178: 124-130.

6. Blue GM, Kirk EP, Sholler GF, Harvey RP, Winlaw DS (2012) Congenital heart disease: current knowledge about causes and inheritance. Med J Aust 197: 155-159.

7. Garg V, Kathiriya IS, Barnes R, Schluterman MK, King IN, et al. (2003) GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. Nature 424: 443-447.

8. Schott JJ, Benson DW, Basson CT, Pease W, Silberbach GM, et al. (1998) Congenital Heart Disease Caused by Mutations in the Transcription Factor NKX2-5. Science 281: 108-111.

9. Bonnet D, Pelet A, Legeai-Mallet L, Sidi D, Mathieu M, et al. (1994) A gene for Holt-Oram syndrome maps to the distal long arm of chromosome 12. Nat Genet 6: 405-408.

10. Basson CT, Cowley GS, Solomon SD, Weissman B, Poznanski AK, et al. (1994) The Clinical and Genetic Spectrum of the Holt-Oram Syndrome (Heart-Hand Syndrome). N Engl J Med 330: 885-891.

11. Stevens KN, Hakonarson H, Kim CE, Doevendans PA, Koeleman BPC, et al. (2010) Common Variation in ISL1 Confers Genetic Susceptibility for Human Congenital Heart Disease. PLoS ONE 5: e10855.

12. Cordell HJ, Bentham J, Topf A, Zelenika D, Heath S, et al. (2013) Genome-wide association study of multiple congenital heart disease phenotypes identifies a susceptibility locus for atrial septal defect at chromosome 4p16. Nat Genet 45: 822-824.

13. Cordell HJ, Töpf A, Mamasoula C, Postma AV, Bentham J, et al. (2013) Genome-wide association study identifies loci on 12q24 and 13q32 associated with Tetralogy of Fallot. Hum Mol Genet 22: 1473-1481.

14. Zaidi S, Choi M, Wakimoto H, Ma L, Jiang J, et al. (2013) De novo mutations in histone-modifying genes in congenital heart disease. Nature 498: 220-223.

15. Arrington CB, Bleyl SB, Matsunami N, Bonnell GD, Otterud BEM, et al. (2012) Exome Analysis of a Family With Pleiotropic Congenital Heart Disease / Clinical Perspective. Circulation: Cardiovascular Genetics 5: 175-182.

16. Liu C, Shen A, Li X, Jiao W, Zhang X, et al. (2008) T-box transcription factor TBX20 mutations in Chinese patients with congenital heart disease. Eur J Med Genet 51: 580-587.

17. Ta-Shma A, Pierri CL, Stepensky P, Shaag A, Zenvirt S, et al. (2013) Isolated truncus arteriosus associated with a mutation in the plexin-D1 gene. American Journal of Medical Genetics Part A. 161: 3115-3120.

18. Greenway SC, McLeod R, Hume S, Roslin NM, Alvarez N, et al. (2014) Exome Sequencing Identifies a Novel Variant in ACTC1 Associated With Familial Atrial Septal Defect. Can J Cardiol. 30: 181-187.

19. Martin LJP, Pilipenko VP, Kaufman KMP, Cripe LMD, Kottyan LCP, et al. (2014) Whole Exome Sequencing for Familial Bicuspid Aortic Valve Identifies Putative Variants. Circulation: Cardiovascular Genetics 7: 677-683.

20. Tariq M, Belmont J, Lalani S, Smolarek T, Ware S (2011) SHROOM3 is a novel candidate for heterotaxy identified by whole exome sequencing. Genome Biol 12: R91.

21. Kirk EP, Sunde M, Costa MW, Rankin SA, Wolstein O, et al. (2007) Mutations in Cardiac T-Box Factor Gene TBX20 Are Associated with Diverse Cardiac Pathologies, Including Defects of Septation and Valvulogenesis and Cardiomyopathy. The American Journal of Human Genetics. 81: 280-291.

22. Gilissen C, Hoischen A, Brunner HG, Veltman JA (2012) Disease gene identification strategies for exome sequencing. Eur J Hum Genet 20: 490-497.

23. Grody WW, Thompson BH, Hudgins L (2013) Whole-Exome/Genome Sequencing and Genomics. Pediatrics 132: S211-S5.

24. Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, et al. (2013) ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet Med. 15: 565-574.

25. Manase D, D'Alessandro LC, Manickaraj AK, Al Turki S, Hurles ME, et al. (2017) High throughput exome coverage of clinically relevant cardiac genes. BMC Med Genomics 7: 10.

26. Sikkema-Raddatz B, Johansson LF, de Boer EN, Almomani R, Boven LG, et al. (2013) Targeted Next-Generation Sequencing can Replace Sanger Sequencing in Clinical Diagnostics. Hum Mutat 34: 1035-1042.

27. Meder B, Haas J, Keller A, Heid C, Just S, et al. (2011) Targeted Next-Generation Sequencing for the Molecular Genetic Diagnostics of Cardiomyopathies / Clinical Perspective. Circulation: Cardiovascular Genetics 4: 110-122.

28. Rehm HL (2013) Disease-targeted sequencing: a cornerstone in the clinic. Nat Rev Genet 14: 295-300.

29. Blue GM, Kirk EP, Giannoulatou E, Dunwoodie SL, Ho JWK, et al. (2014) Targeted Next-Generation Sequencing Identifies Pathogenic Variants in Familial Congenital Heart Disease. J Am Coll Cardiol 64: 2498-2506.