

# Sequencing and microarrays for genome analysis: complementary rather than competing?

Simon Hughes, Richard Capper, Sandra Lam and Nicole Sparkes

## Introduction

The human genome is comprised of more than 3 billion base pairs with greater than 99% being identical between any two unrelated people. The remaining 1% contains a mixture of sequence variants that range in size from a single base (single nucleotide polymorphisms or SNPs) to indels (insertions/deletions) >1bp – <1kb and copy number variations (CNVs) comprising duplications or deletions greater than 1kb in length<sup>1</sup>. Many sequence variants have no associated disease phenotype whilst others, which include inherited and *de novo* changes, can predispose people to diseases such as autoimmune disease<sup>2</sup>, asthma<sup>3</sup>, schizophrenia<sup>4</sup>, obesity<sup>5</sup> as well as a variety of cancers<sup>6-8</sup>.

The development of genome analysis technologies such as DNA microarrays and next generation sequencing (NGS) has provided the researcher with the unique ability to screen for sequence variants of clinical relevance. Although DNA microarrays and NGS might be viewed as competing platforms, this article examines how exome and targeted sequencing is being implemented in biomedical laboratories and how NGS and microarrays could complement each other to address a range of biological questions.

## Applications of genome analysis

Array comparative genomic hybridisation (aCGH) platforms, which allow the detection of known and *de novo* CNVs present in a cell or tissue, play an important role in genome analysis and have had a major impact on the diagnosis of genetic disorders, accelerating CNV discovery for many diseases<sup>9</sup>. Genome-wide association studies (GWAS) using oligonucleotide aCGH are considered the gold-standard for CNV detection<sup>10</sup>. In a clinical setting they have been key for identifying novel disease

loci<sup>11</sup> and recent data suggest that they will have a pivotal role in prenatal diagnosis<sup>12</sup>. NGS offers an alternative approach for genome analysis, providing single base resolution that has permitted the successful identification of causal mutations for a number of monogenic disorders<sup>13-15</sup> as well as for cancer<sup>16,17</sup>.

## Choice of platform for genome analysis

As both microarrays and NGS offer the potential for discovery of sequence variants, when selecting a platform it is important to consider some initial questions.

1. Which platform is the best fit for addressing a particular research objective?
2. Which platform is most appropriate with respect to sample throughput and cost?
3. What data analysis is required?

*Which platform is the best fit for addressing a particular research objective?*

Microarrays are an established technology and can routinely detect aneuploidy, unbalanced chromosomal rearrangements, subchromosomal deletions or duplications, loss of heterozygosity and SNPs (Table 1).

The power of microarrays for detecting such variants comes from the density, coverage and genomic distribution of oligonucleotides on the array. This is of particular relevance clinically and can be addressed by utilising high-density genome-wide array designs or designs combining probes for specific focus regions with lower density probes covering the genomic backbone.

NGS offers the ability to detect the sequence variants outlined above but can also be used to screen for copy-neutral variants (e.g. balanced chromosomal

inversions or translocations), indels or single base variants (e.g. point mutations) (Table 1). NGS provides the user with the capability to scan for disease causing variants without *a priori* sequence information. When planning a NGS experiment there are several components to consider, including protocol (whole exome, custom sequencing, etc.), library preparation, capture method, instrumentation, coverage, etc., which are all essential to ensure that the desired information is obtained from a sequencing run.

**Table 1: DNA sequence variants detected by microarrays and NGS**

Sequence variant	Microarray	Whole exome/ Custom sequencing
Aneuploidy	✓✓	✓
Balanced chromosomal rearrangements		✓
Unbalanced chromosomal rearrangements	✓✓	✓
Sub chromosomal deletions/duplications	✓✓	✓
Loss of heterozygosity	✓✓	✓
SNPs	✓	✓✓✓
Indels		✓✓

*Which platform is most appropriate with respect to sample throughput and cost?*

The numbers of samples that can be processed using either microarrays or NGS vary; however, the most important points to consider are the labour, time and cost involved in sample analysis.

Microarrays enable parallel analysis of large numbers of samples and thus offer the potential to classify patient cohorts. As an example of the throughput achievable using microarrays, Oxford Gene Technology (OGT) recently used its high-throughput technology to process 20,000 samples in 20 weeks as part of the Wellcome Trust Case

Control Consortium (WTCCC) CNV study<sup>18</sup>. A major factor determining the cost of microarray processing is the array format utilised. The facility for parallel processing of samples on a single microarray slide allows significant time and cost savings to be made. In contrast, whole genome sequencing can be more costly with a long turn-around time. To address this, more focussed sequencing approaches can be applied:

- Whole exome sequencing, which focuses on just the 1.5% of the human genome corresponding to gene encoding regions that contain approximately 85% of disease-causing mutations<sup>19</sup>.
- Custom sequencing, to target specific region(s) of interest (ROI) ranging from 0.2 – 34 megabases. Focusing in on one or more ROI will enable increased depth of coverage for those regions and increased confidence for detecting causal mutations.

These approaches present an attractive area for NGS diagnostic development and offer the advantage of shorter turn-around time, reduced sequencing costs, multiplexing and the potential to study larger numbers of patients. However, when considering NGS an awareness of hidden costs is important, as the data analysis hardware and infrastructure requirements as well as experienced bioinformaticians can significantly increase the overall price tag<sup>20</sup>.

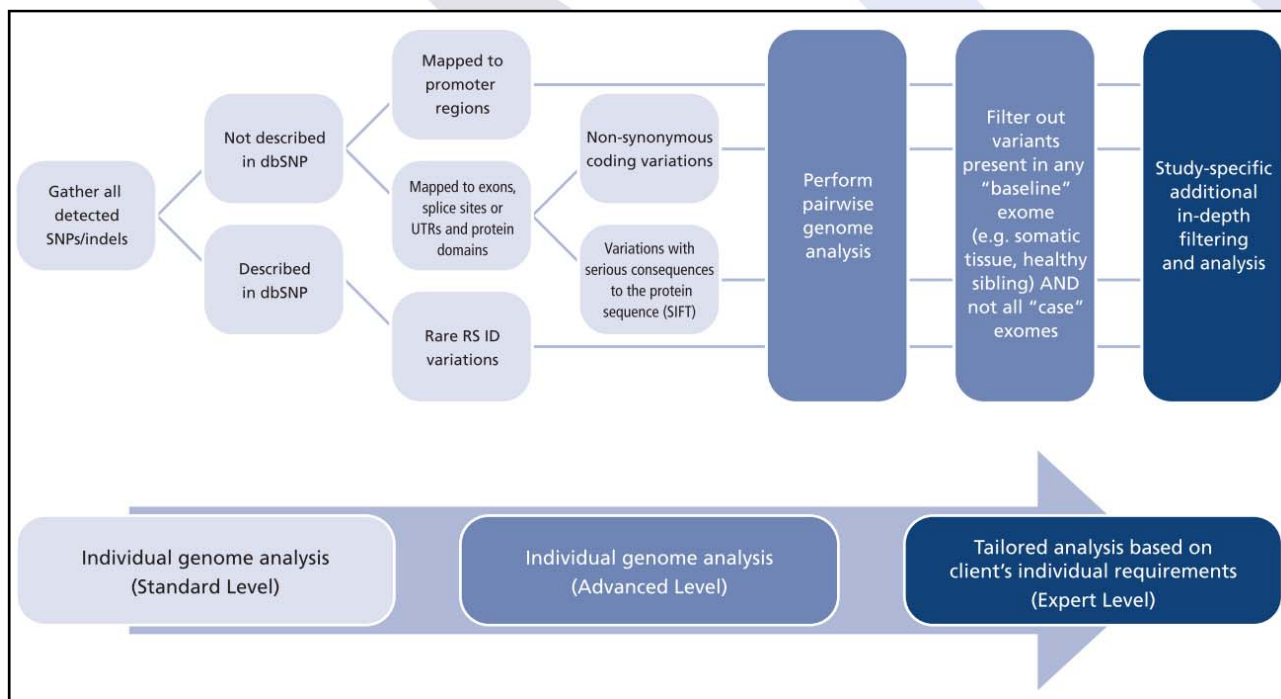
*What data-analysis is required?*

Not all sequence variants detected in a microarray or NGS experiment are necessarily relevant, as the presence of sequence variants in the healthy population indicates that many of these will be benign rather than disease causing<sup>21</sup>. The analysis tools for distinguishing relevant from irrelevant variants from microarray data are well developed and some service providers (e.g. OGT) make these tools available. However, for NGS the identification of important sequence variants can represent a serious obstacle for researchers due to; (i) the volume of data generated and (ii) relatively slow development of standardised “off-the-shelf” data analysis tools. This is of critical importance as incorrect data

analysis, due to problems with base calling, alignment and assembly, may result in clinically relevant information being missed. A robust analytical pathway for data analysis, such as that developed by OGT (Figure 1), can circumvent these problems and accelerate progress in understanding the biological meaning of complex NGS data.

variants with biomedical relevance<sup>15</sup>. This information could then be used to generate new diagnostic arrays or add additional content to existing diagnostic arrays.

The correct combination is essential to ensure that the most information is obtained with a careful balance needed between cost and information required.



**Figure 1:** Translating data into information with OGT's advanced analysis pipeline

## Microarrays or Sequencing?

At present no single platform, either microarrays or NGS, can identify all sequence variants within the genome. Although both platforms function perfectly well in isolation each offers complementary qualities that can, in combination, be used to identify and screen for known or *de novo* sequence variants. The exact order in which the platforms are used depends on the types of questions that need to be answered.

1. If the requirement is to screen a large number of samples to identify a particular subset or genomic region<sup>7</sup> for more comprehensive analysis, microarrays will be more effective for screening followed by sequencing.
2. If the goal is discovery, sequencing could be used to identify sequence

## About OGT

Oxford Gene Technology (OGT) has a proven track record in providing high-quality genomic analysis technology and services, which are designed to lead the researcher all the way from project conception to high-quality results. To enable this, we offer our expertise and experience to help implement the most effective solution for your study and budget.

Our Genefficiency™ high-throughput array service provides researchers with the choice of standard or custom arrays to achieve maximum resolution over either the whole genome or particular regions of interest, while our Genefficiency™ NGS service offers a tailored approach to targeted sequencing, providing the correct combination of components to deliver high-quality results.



Our team of highly experienced project scientists have a strong scientific background and therefore understand the importance of the biology underlying each experiment. By leveraging the skills available at OGT, researchers can focus on the biology rather than the technical aspects and data analysis.

**To find out how we can help advance your genomic research, visit [www.ogt.co.uk/geneficiency](http://www.ogt.co.uk/geneficiency) or contact us on +44 (0)1865 856826.**

## References

1. Gökçümen, O. and Lee, C. (2009) Copy number variants (CNVs) in primate species using array-based comparative genomic hybridization. *Methods* 49, 18-25
2. Fanciulli, M. *et al* (2007) FCGR3B copy number variation is associated with susceptibility to systemic, but not organ-specific, autoimmunity. *Nature Genetics* 39, 721-723
3. Brasch-Andersen, C. *et al* (2004) Possible gene dosage effect of glutathione-S-transferases on atopic asthma: using real-time PCR for quantification of GSTM1 and GSTT1 gene copy numbers. *Human Mutation* 24, 208-214
4. Walsh, T. *et al* (2008) Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* 320, 539-543
5. Walters, R.G. *et al* (2010) A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. *Nature* 463, 671-675
6. Hughes, S. *et al* (2006) The use of whole genome amplification to study chromosomal changes in prostate cancer: insights into genome-wide signature of preneoplasia associated with cancer progression. *BMC Genomics* 7, 65
7. Ernst, T. *et al* (2010) Transcription factor mutations in myelodysplastic/myeloproliferative neoplasms. *Haematologica* 95, 1473-1480
8. Dyrso, T. *et al* (2011) Identification of chromosome aberrations in sporadic microsatellite stable and unstable colorectal cancers using array comparative genomic hybridization. *Cancer Genetics* 204, 84-95
9. Shaffer, L.G. *et al* (2007) The identification of microdeletion syndromes and other chromosome abnormalities: cytogenetic methods of the past, new technologies for the future. *American Journal of Medical Genetics* 145C, 335-345
10. Carter, N.P. (2007) Methods and strategies for analyzing copy number variation using DNA microarrays. *Nature Genetics* 39 (7 suppl), S16-S21
11. Slavotinek, A.M. (2008) Novel microdeletion syndromes detected by chromosome microarrays. *Human Genetics* 124, 1-17
12. Kleeman, L. *et al* (2009) Use of array comparative genomic hybridization for prenatal diagnosis of fetuses with sonographic anomalies and normal metaphase karyotype. *Prenatal Diagnosis* 29, 1213-1217
13. Choi, M. *et al* (2009) Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proceedings of the National Academy of Sciences of the United States of America* 106, 19096-19101
14. Ng, S.B. *et al* (2010) Exome sequencing identifies the cause of a mendelian disorder. *Nature Genetics* 42, 30-35
15. Ng, S.B. *et al* (2009) Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 461, 272-276
16. Wei, X. *et al* (2011) Exome sequencing identifies GRIN2A as frequently mutated in melanoma. *Nature Genetics* 43, 442-446
17. Yan, X.J. *et al* (2011) Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. *Nature Genetics* 43, 309-315

18. Conrad, D.F. *et al* (2010) Origins and functional impact of copy number variation in the human genome. *Nature* 464, 704-712

19. Choi, M. *et al* (2010) Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proceedings of the National Academy of Sciences of the United States of America* 106, 19096-19101.

20. McPherson, J.D. (2009) Next-generation gap. *Nature Methods Supplement* 6, S2-S5.

21. MacArthur, D.G. and Tyler-Smith C. (2010) Loss-of-function variants in the genomes of healthy humans. *Human Molecular Genetics* 19, R125-R130.

**Oxford Gene Technology**

Begbroke Science Park,  
Sandy Lane, Yarnton,  
Oxford OX5 1PF  
United Kingdom

T: +44 (0)1865 856826

F: +44 (0)1865 848684

[www.ogt.co.uk](http://www.ogt.co.uk)

