

Comparison of CytoSure™ Constitutional NGS with Microarrays for CNV detection

Introduction

For the detection of copy number variation (CNV), DNA microarrays are viewed as the 'gold standard'. The use of array comparative genomic hybridisation (aCGH) has been particularly effective in detecting CNVs within DNA samples from individuals with intellectual disability (ID) and developmental delay (DD). This has resulted in the detection of novel syndromes which were previously undetectable¹. In order to further facilitate cytogenetics research, specific microarrays, such as the CytoSure Constitutional v3 array, have been developed. These microarrays have an enhanced probe coverage for genes of interest as well as probes regularly spaced throughout the genome to detect larger CNVs on the genomic backbone.

Next Generation Sequencing (NGS) has become a revolutionary technology for the analysis of ID and DD samples, having the advantage of being able to detect single nucleotide variations (SNVs) and small insertions and deletions (indels) simultaneously within many genes. However, the ability of NGS to call small CNVs is still not routine and robust. In order to facilitate the transition of small (as low as single exon) CNV calling to NGS from arrays, appropriate capture panel design and improved software are both required.

OGT, with its world-leading probe design and software capabilities, is well placed to meet this need. We have designed an NGS assay and analysis software that is able to detect large and small CNVs with the same precision and sensitivity as microarrays, as well as providing SNV and Indel calling.

The CytoSure Constitutional NGS assay is a hybridisation-based method to capture specific regions of the genome. These regions are:

- Over 700 genes implicated in ID and DD at the single exon level – to detect SNV/Indels within the genes and UTRs as well as small intragenic CNVs.
- Nucleotides flanking target genes, up to 35 base pairs from the exons to capture splice site variants
- Extensive backbone baits spaced throughout the genome – to detect large CNVs and stretches of loss of heterozygosity (LOH)

Like OGT's CytoSure Constitutional arrays, the gene list has been developed with input from the ClinGen database and the Deciphering Developmental Disorders project, as well as leading Cytogeneticists. In addition to the targeted gene/exon regions, there are 28,641 backbone baits which are spaced according to the priority regions in the genome and provide an estimated CNV resolution of 189kb in high priority areas and a LOH resolution of 5Mb in the non-targeted region.

This technical note compares performance of CytoSure Constitutional NGS and a range of microarrays for 255 research samples processed in three independent laboratories and OGT.

Experimental outline

OGT's comprehensive library preparation kit provides all the components required for the CytoSure NGS workflow (Figure 1). The initial library preparation step involves the ligation of adaptors to fragmented DNA followed by a PCR step. Following the first stage of library preparation, samples are pooled into sets of 8 to allow better handling, and the baits are hybridised to the libraries and then washed to remove non-specific DNA, leaving just the DNA of interest. After a second PCR, the prepared libraries are ready for sequencing. Samples can then be sequenced, either 24 on the Illumina® High Output NextSeq or larger sample numbers on the NovaSeq.

The resulting FASTQ files generated from the sequencer are aligned using the OGT Interpret software. As well as providing comprehensive QC metrics on the performance of the run, the software provides the complete analytical pipeline to investigate CNVs, LOH, Indels and SNVs.

In this study OGT investigated samples from three independent collaborators to determine the performance of the CytoSure NGS assay in CNV, LOH, SNV and Indel detection. The results were compared to those obtained with alternative technologies e.g. microarrays, NGS targeted panels and whole exome panels.

Data Analysis

OGT's Interpret software is a state-of-the-art analytical pipeline that includes a proprietary CNV calling algorithm. This analytical pipeline enables the calling of CNVs, LOH, SNVs and Indels and presents the information in a user-friendly interface. The software allows the user to set a customisable analysis protocol, but is provided with a default protocol as standard. In this study the default protocol within the software was used. Interpret has been designed to allow a great deal of flexibility and dynamic filtering - the filtering protocol employed in this study is illustrated in Figure 2.

In order to call CNVs, the OGT software requires a set of samples that do not have large CNVs to be used as a reference. Interpret calculates the ratio between the sample read counts and the reference read counts and a threshold is applied to call the CNV.

The selection of an appropriate reference set is important, Interpret is flexible in that it allows the user to select the reference data and format. Examples of reference formats for the samples to be analysed are as follows:

- All samples in the same NGS run (intra-run).
- A subset of samples in the same NGS run or from other runs.

OGT recommends employing intra-run referencing using every other sample on the run as the reference for screening purposes, this provides the best quality data.

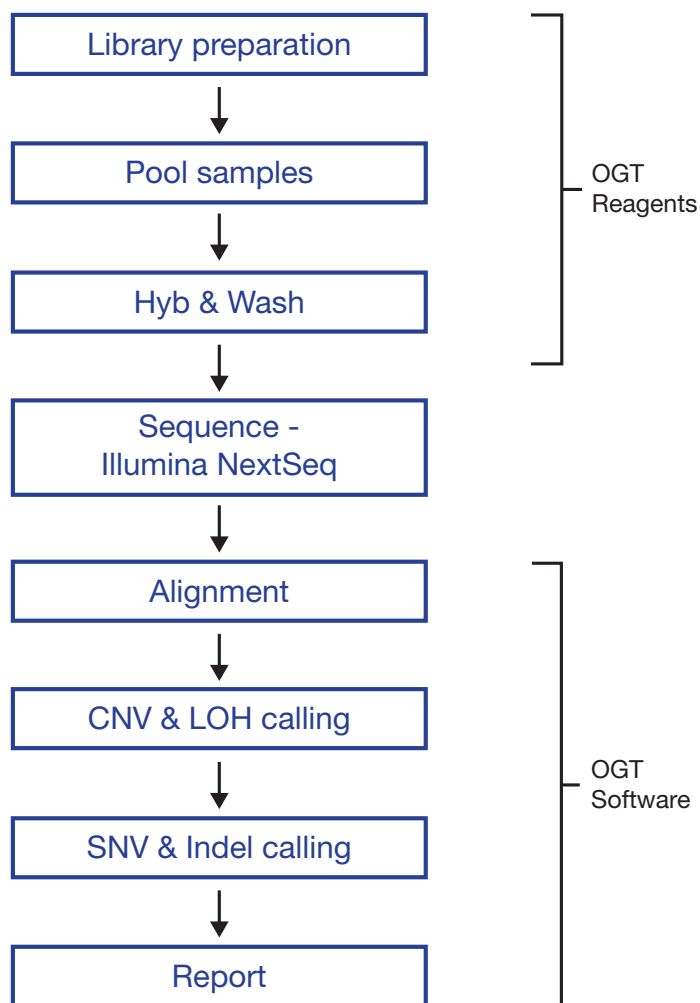


Figure 1: Overview of the laboratory and data analysis workflow

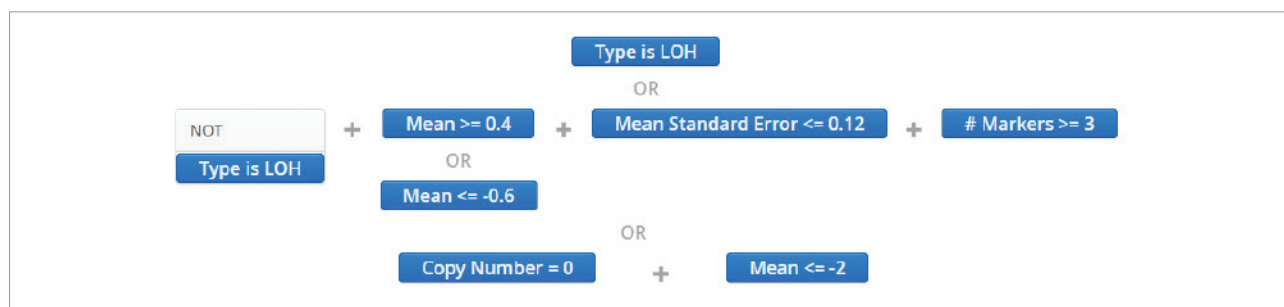


Figure 2: Protocol filter used in the software

Results

The study

To test the detection performance of the CytoSure Constitutional NGS reagents and software, a total of 255 samples were tested. The assay was carried out in four laboratories, and the samples consisted of the following:

- Coriell samples with an identified pathogenic CNV, SNV or stretch of LOH
- Control samples – with no pathogenic aberration
- Research samples from intellectual disability and developmental delay patients with previously identified pathogenic aberrations – provided by the independent laboratories.

The aberrations were initially detected using a range of microarrays (CNVs and LOH), including CytoSure Constitutional v2 and v3 arrays, Cytoscan™ arrays, and custom arrays – SNVs and Indels were sequenced using Illumina based sequencing platforms.

QC metrics – mean target coverage (MTC)

The CytoSure Constitutional NGS assay targets more than 11,268 exons and UTRs with an even spread of baits across the genome. The mean target coverage is a measure of the number of reads across the desired target regions. For the samples in this study the median MTC was 378 (Figure 3). For robust CNV calling, achieving an MTC of over 150 is recommended.

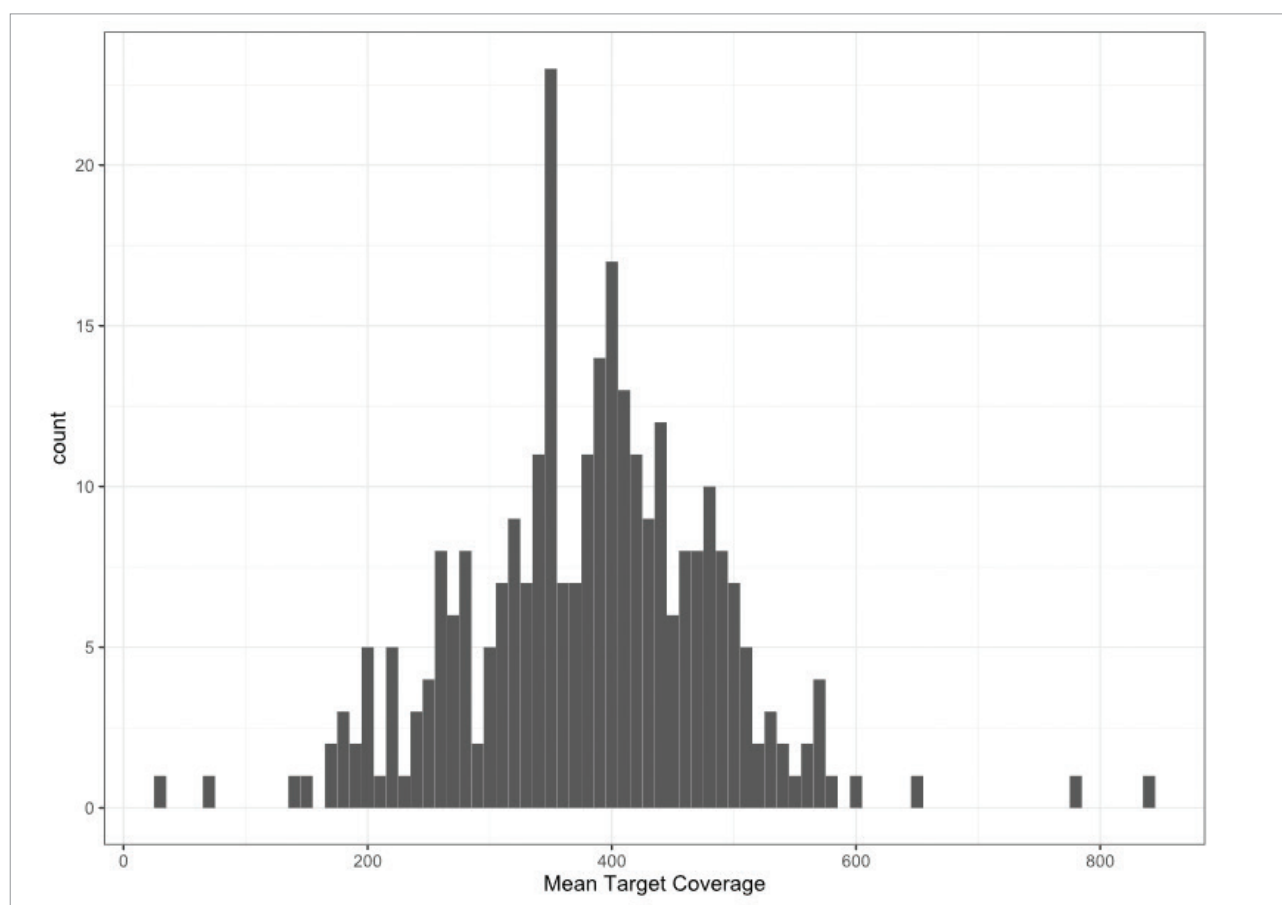


Figure 3: Showing the mean target coverage for the samples in the study

Calling Results

SNV/ Indel performance

One of the major advantages of CytoSure Constitutional NGS over an array is the ability to detect SNVs. In this study 41 samples were included (Table 1) which contained a total of 47 pathogenic SNV/Indels within the target regions, all of them were correctly detected by Interpret.

	Number expected	Number called	Accuracy (%)
Pathogenic SNVs/ Indels	47	47	100

Table 1: Accuracy of SNV calling with CytoSure Constitutional NGS.

SNV and Indel Calling data examples



Figure 4: An insertion within JAG1.

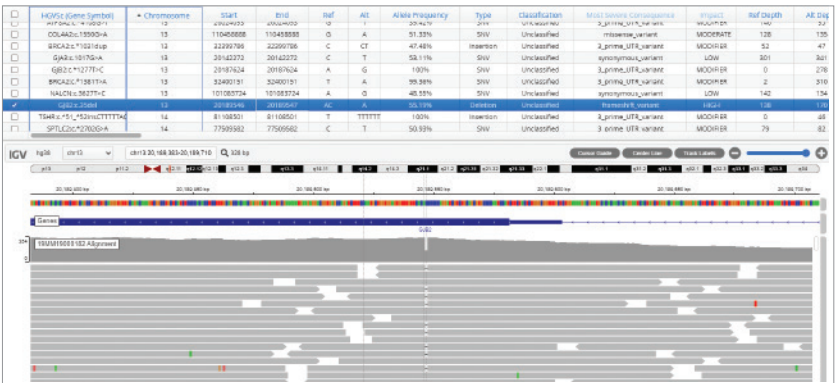


Figure 5: A frameshift variant within GJB2.

In addition, the performance of the assay for SNV calling was also benchmarked using 11 genome-in-a-bottle (GIAB) samples whose data was generated in all four participating laboratories (three independent labs and OGT).

NA12878	True Positive	False Positive	False Negative	Precision (%)	Sensitivity (%)	F-Measure (%)
SNV/Indel	2153	15	49	99.33	97.81	98.56

Table 2: Summary of GIAB results.

NA12878	True Positive	False Positive	False Negative	Precision (%)	Sensitivity (%)	F-Measure (%)
SNV	1826	19	8	98.98	99.59	99.28
Indel	333	25	40	93.09	89.26	91.12

Table 3: SNV and Indel split of the GIAB samples.

The SNV/Indel variants were also split according to their mutation (Table 3) demonstrating a slightly higher precision for the SNV detection over the Indel calling, in-line with what has been reported in the literature².

Many of the false negatives can be attributed to highly repetitive regions (e.g. homopolymer regions) and appear to be artifacts present in the GIAB WGS. These kinds of regions are highlighted in the software. Precision is defined as $TP/(TP+FP)$ and Sensitivity as $TP/(TP+FN)$ where TP are true positive, FN false negative and FP false positive. F measure is the harmonic mean of Precision and Sensitivity.

CNVs performance

CytoSure Constitutional NGS was developed to give the same high performance CNV calling as the gold-standard aCGH. To determine if we had delivered that level of detection we analysed and compared 101 samples (with 118 known pathogenic CNVs), sourced from our collaborators. The CNVs were a range of sizes, 54 were less than 2Mb and 64 were larger than 2Mb (Table 4), this was to ensure that we could produce the same robust calling across the whole size range. We achieved overall concordance of 96% and an impressive 98% for the smaller CNVs (<2Mb). The CNVs we didn't call were visible on the software but due to the filtering protocol of this study, were precluded and not called. This is also not unusual on array platforms and visual analysis or adjustment of filtering parameters is deployed.

	Number expected	Number called	Accuracy (%)
Pathogenic large CNVs (>2Mb)	64	61	95.3
Pathogenic small CNVs (<2Mb)	54	53	98.1
Pathogenic CNVs (all)	118	114	96.6

Table 4: Accuracy of CNV calling for CytoSure Constitutional NGS.

CNV calling specificity

To assess the specificity of the CytoSure Constitutional NGS assay we investigated the prevalence of false positives reported from 11 GIAB control samples, which should not contain any CNVs at the exon level. The assay covers 11255 exons across 707 genes. We reported only seven false CNVs across all 11 samples, six samples had specificity of 100% and called no CNVs, four samples had one false positive, and one sample had three false positives. This data gives a specificity of 99.99% across the 11 samples, this specificity is more impressive due to the samples being analysed by several labs and from different batches of GIAB.

CNV Calling in Duplication and Deletion Samples

CytoSure Constitutional NGS was able to detect CNV deletions and duplications across a wide size range (Figures 6-9). For the targeted 707 genes associated with ID/DD there is exon level resolution, which allows the assay to pick up very small CNVs (Figure 7), and due to the backbone of baits we are also able to detect larger CNVs across the genome. The Interpret software was designed to present data in a similar format to existing aCHG analytical software, to ease the transition of CNV analysis from array to NGS.

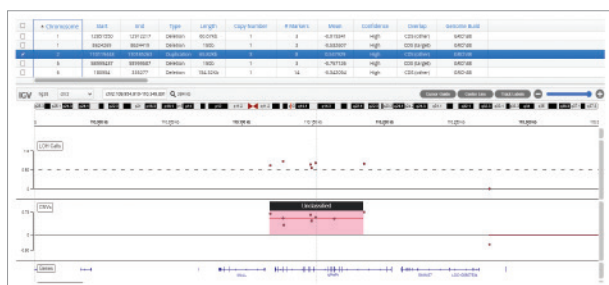


Figure 6: 65.8kb duplication on chromosome 2. Top panel: analysis on CytoSure Constitutional NGS, Bottom panel: microarray analysis.

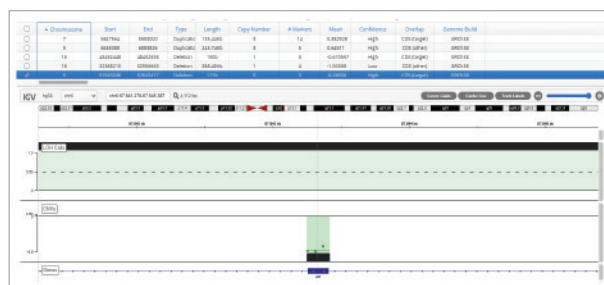


Figure 7: 150bp deletion in Androgen Receptor gene on chromosome X. Top panel: analysis on CytoSure Constitutional NGS, Bottom panel: microarray analysis.

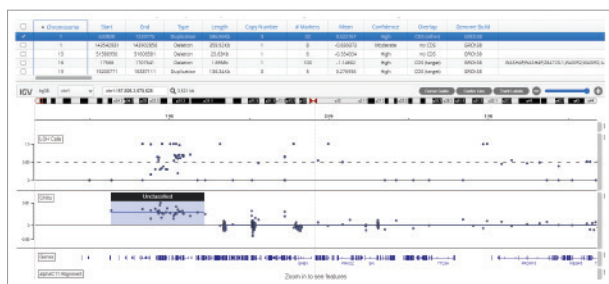


Figure 8: 586kb duplication on chromosome 1. Top panel: analysis on CytoSure Constitutional NGS, Bottom panel: microarray analysis.

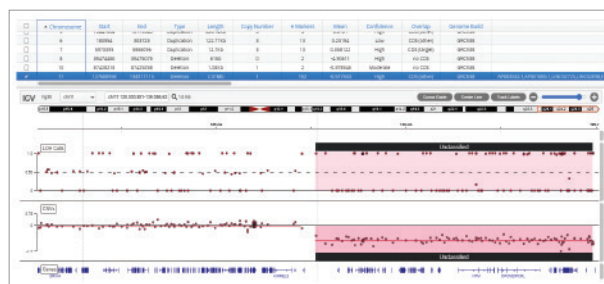


Figure 9: A 7.3MB deletion on chromosome 11 with a stretch of LOH. Top panel: analysis on CytoSure Constitutional NGS, Bottom panel: microarray analysis.

LOH performance

There were seven samples with known stretches of LOH (Figures 10-11). Some of the samples, being consanguineous had multiple regions of LOH. CytoSure Constitutional NGS is able to detect LOH calls of 5Mb and greater. In the seven samples, there were a total of 51 reported LOH regions over 5Mb and the CytoSure NGS assay analytical software, Interpret, called all of them (Table 5).

Number of LOH over 5Mb	Detected by CytoSure NGS	%
51	51	100

Table 5: Accuracy of LOH detection in CytoSure Constitutional NGS.



Figure 10: 10.39Mb LOH and a 213.75Kb duplication on chromosome 7. Top panel: analysis on CytoSure Constitutional NGS, Bottom panel: analysis on CytoSure ISCA v2 array.

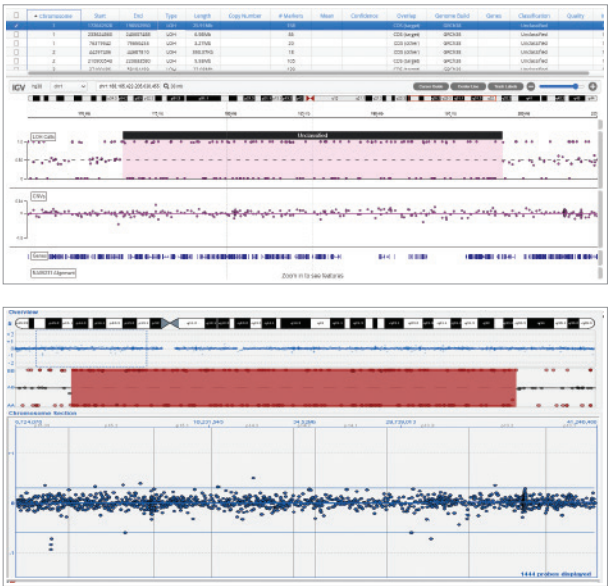


Figure 11: 25.91MB LOH on chromosome 1. Top panel: analysis on CytoSure Constitutional NGS, Bottom panel: analysis on CytoSure ISCA v2 array.

Mosaic samples

In addition to the standard CNV research samples, we received a range of mosaic samples (not included in Table 4), differing in degree of mosaicism and size of aberration (Table 6). Of the eight samples, Interpret was able to automatically call five of them, including a 7Mb deletion with 50% mosaicism (Figure 12). The three aberrations that weren't called automatically were visible on inspection of the IGV viewer, but were not automatically called due to the protocol settings. The thresholds can be eased by the user to better call these mosaic mutations. The ability of CytoSure Constitutional NGS and Interpret to detect mosaic samples demonstrates the versatility and sensitivity of the assay.

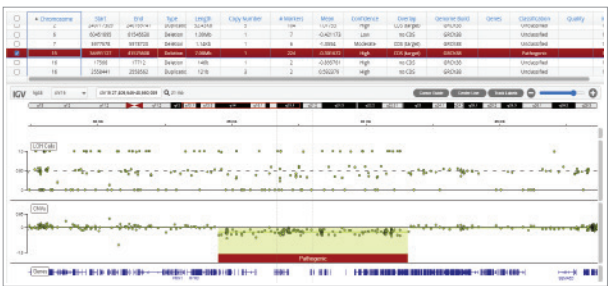


Figure 12: A 7Mb deletion on chromosome 15 in a mosaic sample.

Mosaic Samples	Size (Mb)	Mosaicism (%)	Called	Visible
chrX:1-156040895	156	>50%	Yes	Yes
chr9:1-138394717	138	<10%	No	Yes
chr11:107621153-135062054	27	>50%	Yes	Yes
chr1:143200000-165500000	22	>50%	Yes	Yes
chrY:10863-17452498	17	>50%	No	Yes
chr15:34622140-41637368	7	>50%	Yes	Yes
chr11:44113922-46371104	2	>50%	Yes	Yes
chr16:2046053-2062016	<1	>50%	No	Yes

Table 6: Summary of mosaic samples in the study.

Conclusions

This study, with over 200 samples, has shown that the CytoSure Constitutional NGS assay is as effective as microarrays in calling CNVs and LOH, with the additional ability to be able to detect SNVs and Indels. The ability to detect CNVs, SNVs, Indels and LOH in a single robust assay reduces analytical cost and burden as well as reducing the overall time taken to deliver a result for a given sample.

The CytoSure Constitutional NGS solution includes everything you have come to rely on with the well-established CytoSure microarray brand from Oxford Gene Technology (OGT), namely, the most up-to-date ID/DD content, expert panel design, class-leading complimentary software and unparalleled support. It enables the seamless transition from microarrays to NGS, delivering a significant increase in information obtained from a single assay without extensive analysis time and costly data generation and storage.

Acknowledgements

Centre Hospitalier Universitaire de Sherbrooke, Quebec, Canada

All Wales Medical Genetics Service, Cardiff, UK

EGL Genetics, Georgia, USA

References

1. Sharp *et al.*, Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. *Nature genetics*. 2006 August; 38: 1038-1042
2. Krusche *et al.* Best practices for benchmarking germline small-variant calls in human genomes. *Nat Biotechnol*. 2019 May;37(5):555-560.



A Sysmex Group
Company

Oxford Gene Technology, Begbroke Science Park, Woodstock Road, Begbroke, Oxfordshire, OX5 1PF, UK

T: +44(0)1865 856800 (US: 914-467-5285) E: products@ogt.com W: www.ogt.com

CytoSure: For Research Use Only; Not for Diagnostic Procedures.

This document and its contents are © Oxford Gene Technology IP Limited – 2020. All rights reserved. OGT™ and CytoSure™ are trademarks of Oxford Gene Technology IP Limited.