

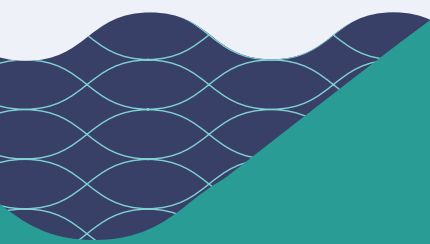
CytoSure



A Sysmex Group Company

Application Note

The impact of microarray probe design on detecting copy number variants at exon-resolution



Introduction

The analysis of structural variants, such as copy number variants (CNVs), is an important aspect of clinical genetics research. Whilst many technologies are used for determining CNVs in human DNA, array comparative genomic hybridisation (aCGH) is now established as the gold standard for detection of CNVs across the entire genome, and is used not just in research but also in clinical applications¹.

As microarray technology has evolved, the resolution at which CNVs can be detected has steadily increased. Despite this, the design strategy behind many microarray designs has made it difficult or impossible to find aberrations smaller than ~30kb², despite the fact that smaller aberrations have been demonstrated to be relevant in Mendelian disorders³. In order to tackle this problem, sophisticated probe design approaches have been developed over the last few years, and have made it possible to increase the resolution of arrays much further – targeting important genetic loci in such a way that CNVs can be found even at the exon-level⁴.

However, in order to develop arrays that can robustly analyse CNVs at this size, a number of factors need to be taken into account during the design process.

Probe targeting

Due to the defined/fixed number of probes on any array, an important aspect of development is to utilise these as effectively as possible. The ability of an array to detect CNVs is defined by the number of probes targeting any given region – these regions might be high-priority regions which need highly clustered probes to deliver desired resolution, or lower priority “backbone” regions which will have less clustered probes. It is therefore important to strike the right balance between probes used across the backbone regions and highly targeted regions to suit the microarray’s purpose. To do this, OGT’s design algorithms work to find the best possible genomic targets for each probe in order to achieve the desired resolution, without compromising on overall content.

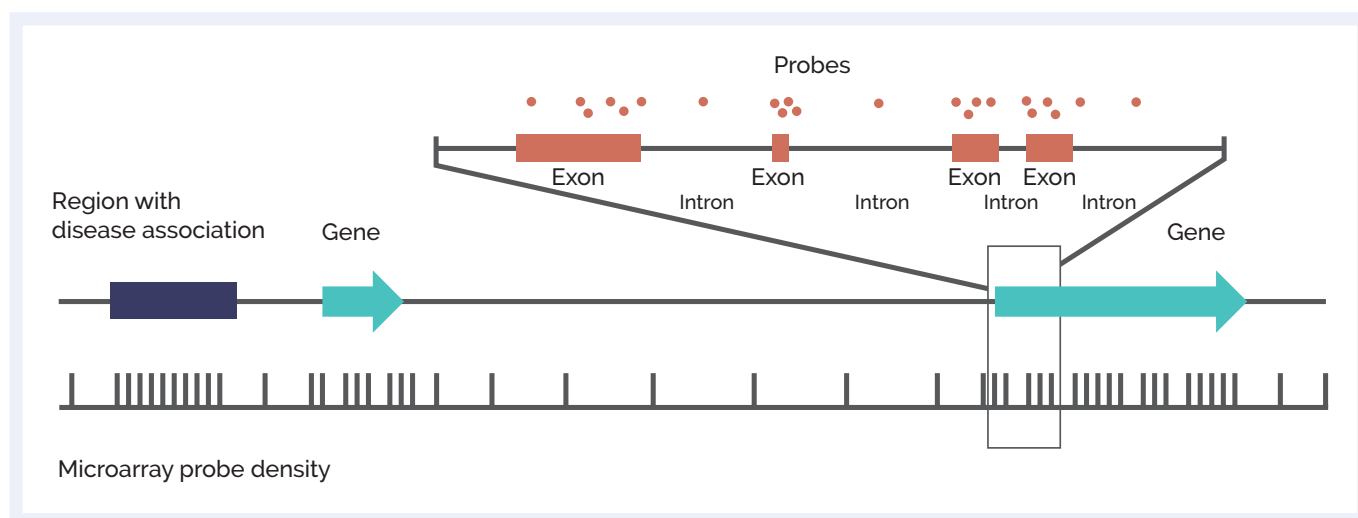


Figure 1: Schematic of microarray probe density across different genetic elements. Probes can be focused across exons to give ultra-high resolution at the exon-level, enabling single exon CNV discovery. High probe density can also be targeted at non-genic regions of importance. Backbone probes, dispersed throughout the genome, allow for discovery of larger variants at standard resolution in other areas.

Factors affecting probe performance

The key to producing a microarray able to robustly detect CNV at any desired resolution is in the design of the probes themselves and the way they function. To understand what makes a good probe, we need to consider the following:

Probes need to target loci in a specific manner (i.e. should not provide false positives). This can be affected by the following factors:

- Cross-hybridisation to other targets — usually very common on genes sitting on segmental duplication regions and pseudogenes
- Non-specific binding (e.g. poly-G stretches or repeats)
- GC content — the higher the GC, the stickier the probe, which tends to cause non-specific binding and noninformative signals (probe fluorescence remains high regardless of copy number change)

Probes need to target loci in a sensitive manner (i.e. should not provide false negatives). This can be affected by the following factors:

- Secondary structures in the probe — can cause the active sequence of the probe to become unavailable for hybridisation
- Secondary structures in the target — secondary structures on the target fragments might hinder hybridisation
- GC content — the lower the GC, the less sticky the probe is, which tends to cause a lack of binding and non-informative signals (probe fluorescence remains low regardless of copy number change)

Probes need to function under relatively similar Isothermal conditions:

- Similar hybridisation conditions — similar characteristics mean that probes will perform more uniformly in any given set of hybridisation conditions
- The more similar the hybridisation performance across probes, the less noise is created in the final datasets

Probe optimisation

With all of these factors in mind, a sophisticated approach to probe design needs to be adopted to ensure probes provide the most informative signals and consequently the best possible resolution across the regions of interest. OGT's design workflow includes all of the following *in silico* steps to design and identify the best possible probes to target any given region:

1. Analysis of target sequences, including gathering sequence metadata (e.g. intronic vs exonic)
2. Identify repetitive and homologous regions
3. Generate all possible probes
4. Analyse the physicochemical properties of the probes
5. Rank probes depending on most desirable properties
6. Choose most suitable probes for target regions

In addition to all the *in silico* steps taken during probe design, all OGT catalogue microarrays also undergo an empirical optimisation process, to ensure that all probes are working at peak performance. At times, selection of the best possible probes from a number of potential options is necessary, especially for regions that are difficult to target due to the nature of the sequence. By testing a number of different probes across a great number of repeats, selection of the best possible probes to target any desired loci is possible (Figure 2). All of this design work leads to better probe specificity, reduced noise, and importantly more accurate results (Figure 3).

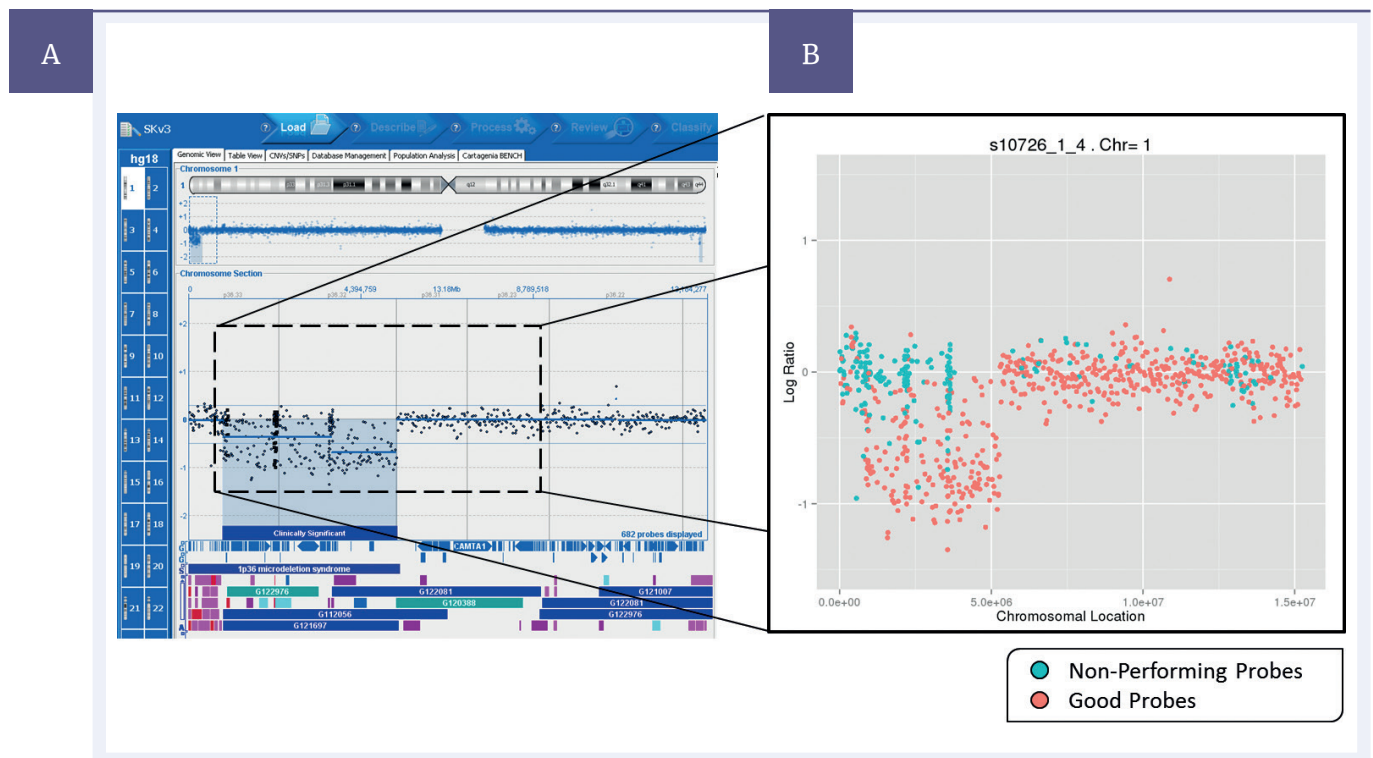


Figure 2: **A** Data from a single experiment with a deletion present on chromosome 1, shown on CytoSure® Interpret Software. **B** average probe performance across 4000 repeats. "Non-performing probes" (blue) defined as those which gave an incorrect result more often than a correct result, "Good probes" (pink) defined as those which gave a correct result more often than an incorrect result.

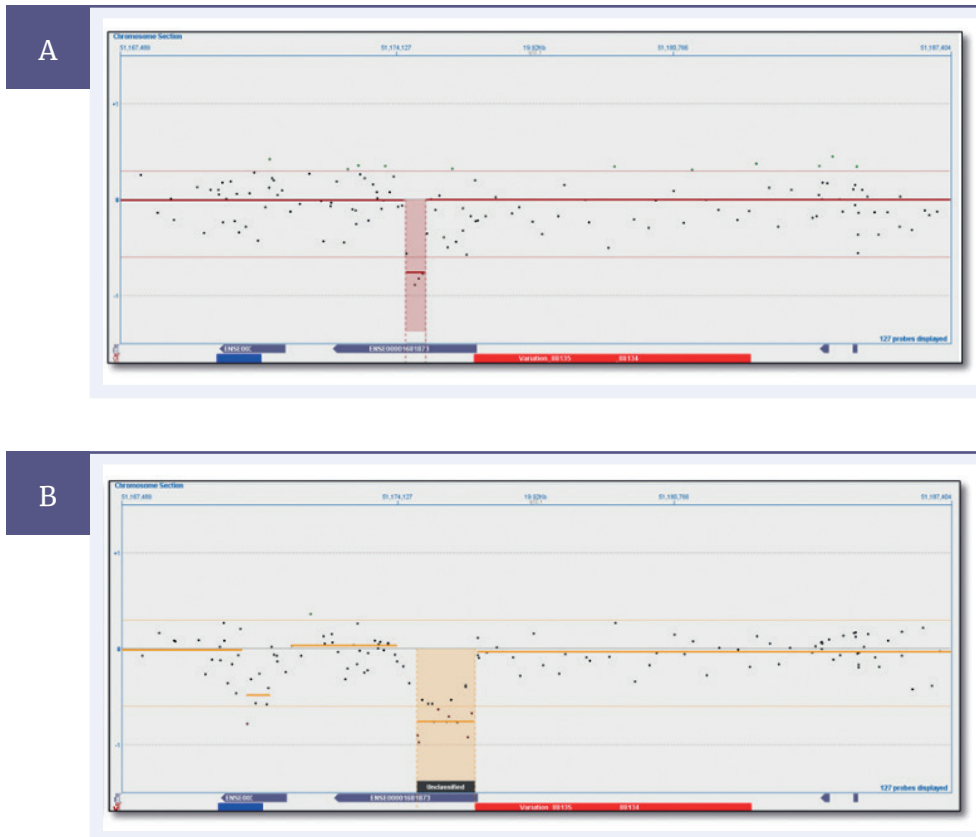


Figure 3: Separate experiments on the same sample with a confirmed ~1600bp deletion. **A** experiment run using non-optimised array, with a ~400bp deletion found. **B** experiment run using optimised array, with a ~1600bp deletion found.

To aid with the design process, OGT has developed a database of probes called the Oligome™, which includes probes across the entire genome that are in silico or empirically optimised, allowing for the rapid development of microarray designs. Many hundreds of design projects have helped to continuously improve the database, (which now includes over 20 million probes) with the data from these projects improving the design process and algorithms in an iterative fashion.

Conclusion

A number of different factors must be taken into account when designing arrays, especially those that offer exon resolution. OGT is able to offer high-resolution arrays due to the design process developed as a result of many years of experience designing bespoke microarrays.

Moreover, the entire CytoSure workflow is optimised for robust analysis of CNV – to complement the excellent performance of the microarrays, CytoSure Labelling kits offer very low DLRS (derivative log ratio spread – a measure of noise in array data); and CytoSure Interpret software offers a robust, feature rich and user-friendly platform for the analysis of microarray data.

Ordering information

UK +44 (0) 1865 856800

US +1 914 467 5285

contact@ogt.com

ogt.com

Product	Contents	Cat. No.
CytoSure Constitutional v3 (8x60k)	Microarray with four arrays of 60,000 spots; CytoSure Interpret Software	020045
CytoSure Constitutional v3 (4x180k)	Microarray with four arrays of 180,000 spots; CytoSure Interpret Software	020046
CytoSure Constitutional v3 +LOH (4x180k)	Microarray with four arrays of 180,000 spots; CytoSure Interpret Software	020047
CytoSure Genomic DNA Labelling Kit	24 reactions, clean-up columns, dyes, nucleotide mix random primers, enzyme, collection tubes	020020
CytoSure HT Genomic DNA Labelling Kit	96 reactions, 2 purification plates, dyes, nucleotide mix, random primers, enzyme	500040
CytoSure Sample Tracking Spike-ins A – H	Sample Tracking Probes sufficient for 12 reactions supplied in three aliquots	500066

References

1. Kearny, H.M. *et al.*, (2011) American College of Medical Genetics recommendations for the design and performance expectations for clinical genomic copy number microarrays intended for use in the postnatal setting for detection of constitutional abnormalities. *Genome in Medicine* 13 p676–685
2. Poultney C.S. *et al.*, (2013) Identification of Small Exonic CNV from Whole-Exome Sequence Data and Application to Autism Spectrum Disorder. *AJHG* Vol 93(4), 607–619
3. Aradhya S. *et al.*, (2012) Exon-level array CGH in a large clinical cohort demonstrates increased sensitivity of diagnostic testing for Mendelian disorders. *Genetics IN Medicine* 14, 594–603
4. Askree S.H. *et al.*, (2013) Detection limit of intragenic deletions with targeted array comparative genomic hybridization. *BMC Genetics* 14:116



A Sysmex Group Company

**What binds us,
makes us.**

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

CytoSure: This product is provided under an agreement between Agilent Technologies, Inc. and OGT. The manufacture, use, sale or import of this product may be subject to one or more of U.S. patents, pending applications, and corresponding international equivalents, owned by Agilent Technologies, Inc. The purchaser has the non-transferable right to use and consume the product for RESEARCH USE ONLY AND NOT for DIAGNOSTICS PROCEDURES. It is not intended for use, and should not be used, for the diagnosis, prevention, monitoring, treatment or alleviation of any disease or condition, or for the investigation of any physiological process, in any identifiable human, or for any other medical purpose. This document and its contents are © Oxford Gene Technology IP Limited – 2021. All rights reserved. OGT™, Oligome™ and CytoSure® are trademarks of Oxford Gene Technology IP Limited.