Customizable exon-centric target enrichment strategy for copy number and SNP analysis

Arjun Vadapalli *, Kyeong Soo Jeong*, Ashutosh Ashutosh , Devendra Joshi , Eric Lin, Carlos Pabon, Gilbert Amparo, Jayati Ghosh, Douglas Roberts
Agilent Technologies, Diagnostics and Genomics Group, Santa Clara, California 95051 *Equally contributed

Agilent Technologies

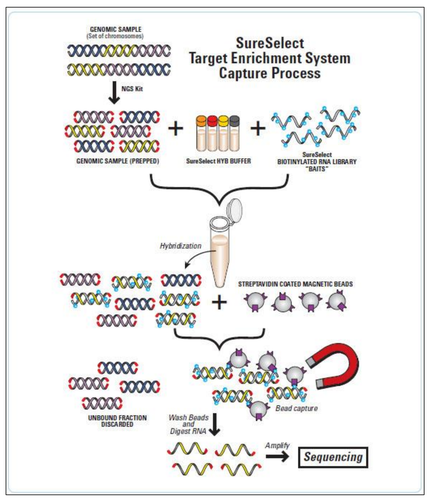
Abstract

Structural variations in the genome can be determined from NGS data with either whole genome sequencing (WGS) or targeted enrichment using exome or gene panels. Copy number variation (CNV) of genomic segments is a large category of structural variation and has been implicated in many Mendelian diseases and complex traits. The impact of CNVs on gene expression is not limited to only the coding regions of genes localized within the CNV but also their flanking regions, an effect that could even extend over the entire length of the chromosome. Genomic coverage of coding regions and their upstream and downstream regions can vary between different platforms that are currently being used to study CNVs. DNA sequencing technologies which allow human genomes to be re-sequenced rapidly and inexpensively, are being used increasingly to detect a comprehensive list of variants relative to the reference genome. Combining DNA sequencing with the ability to selectively capture DNA targets provides additional cost benefits and lower amounts of DNA input per experiment. WGS has the potential to provide a single platform solution for detecting copy number variations (CNVs). However, it is prohibitively expensive in identifying mutations such as single nucleotide polymorphisms (SNPs), and insertions and deletions (INDELs) that require high sequencing read depths.

OneSeq is an NGS target enrichment solution providing simultaneous detection of genome-wide CNVs and copy neutral loss of heterozygosity (cnLOH) in addition to SNPs and INDELs at desired loci. Herein we describe custom OneSeq, enabling higher resolution, exon-centric CN determination in any genes of interest. To this end, we have created a library of high resolution target enrichment probes targeting exon-proximal regions. When deployed in conjunction with existing exonic probes, they improve detection of biologically relevant CNVs in and near exons, while simultaneously detecting exonic SNPs and INDELs. The targeted CN regions can be customized to address only a subset of genes, thus maximizing sequencing efficiency.

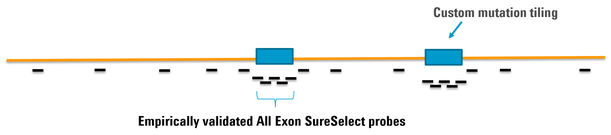
To demonstrate the performance of the custom OneSeq assay, publicly available genomic DNA reference materials for DMD genetic testing are studied for known deletions, duplications and point mutations.

SureSelect Target Enrichment Workflow



OneSeq Custom Design/Analysis

OneSeq probes for exon-level aberration detection in custom regions

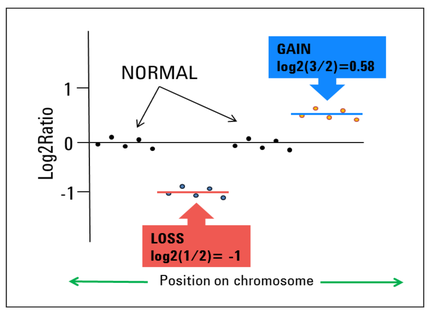


Custom OneSeq Design for the present study

1) *DMD* in chromosome X for exon-level copy number & SNP/INDEL detection
2) Whole chromosome 18 for exon-level copy number detection in 18q deletion

OneSeq data analysis

log2(Number of Reads in sample / Number of Reads in Reference)



Schema of DMD gene

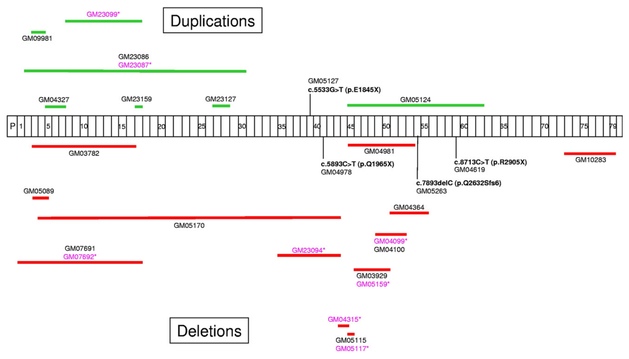


Figure 1. Composition of DMD reference material panel. Schema of the promoter (P) and 79 exons of the DMD gene are indicated by numbered black rectangles. The locations of deletions or duplications of DMD exons in each cell line are indicated by horizontal bars. Duplications are indicated above the exon map, and deletions are shown below. The exonic locations of point mutations are indicated by vertical lines. Coriell numbers (GMXXXXX) are shown for each cell line, and DNA from female donors is indicated with an asterisk.

Exon-level detection in *DMD*

One copy gain in exon 27 and 28 (*DMD* encodes dystrophin)

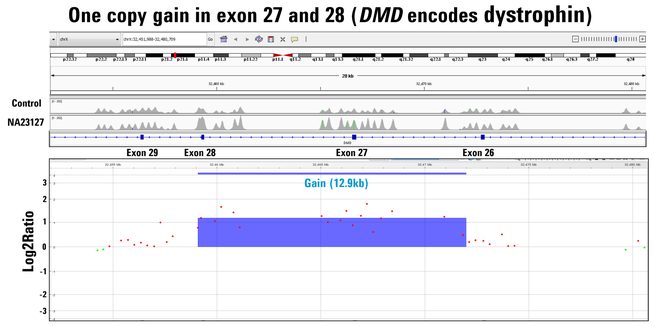
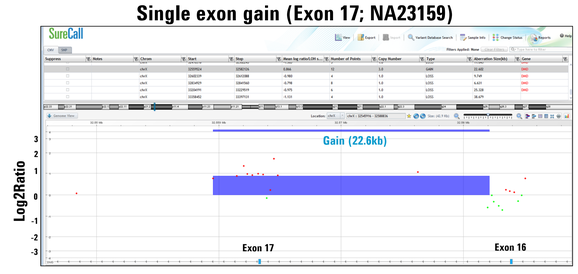
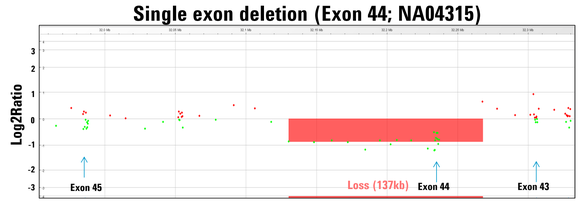


Figure 2. Detection of exon 27 and 28 in *DMD*. 832 probe groups were designed to detect exon-level copy number changes. The standard OneSeq protocol was used with Coriell DNA (NA23127) and Agilent male reference DNA. Sequencing read depths are shown in Integrative Genomics Viewer (top panel). The vertical axis in the bottom panel shows their log ratios. SureCall detects 12.9kb of one copy gain (blue box) across exon 27 and 28.

Single exon gain (Exon 17; NA23159)



Single exon deletion (Exon 44; NA04315)



Single exon deletion (Exon 45; NA05115)

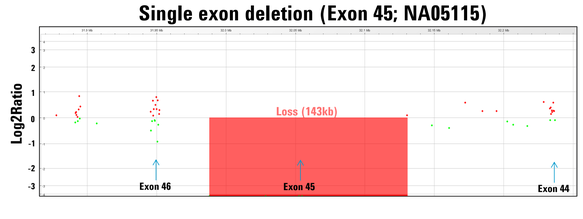


Figure 3. Single exon detection in *DMD*. OneSeq was performed with DNA aberrations in one exon. Baits located in exons, exon-proximal regions, and regional backbone allow for precise detection of aberration boundaries. Each panel shows DNA aberration in one exon and proximal introns.

SNP/INDEL detection in *DMD*

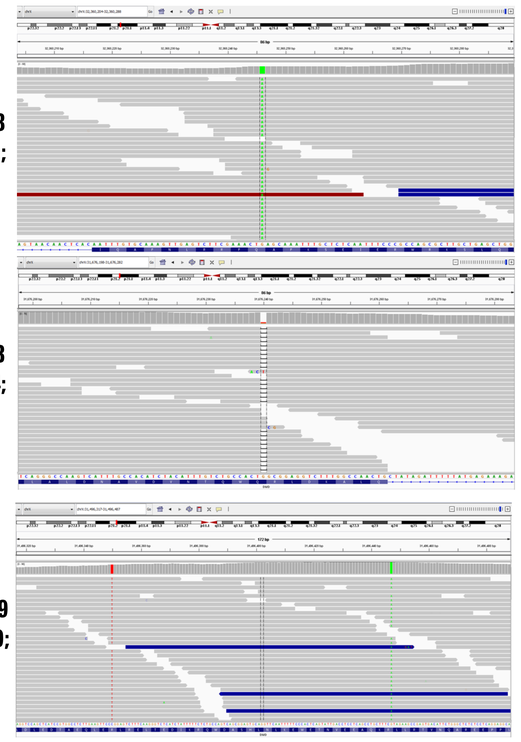


Figure 4. Detection of known SNP/INDEL in *DMD*. Coriell DNA samples were analyzed with custom OneSeq design. All known SNP/INDEL were independently confirmed by microarray in Kalman L. et al.

Exon-level detection in 18q deletion

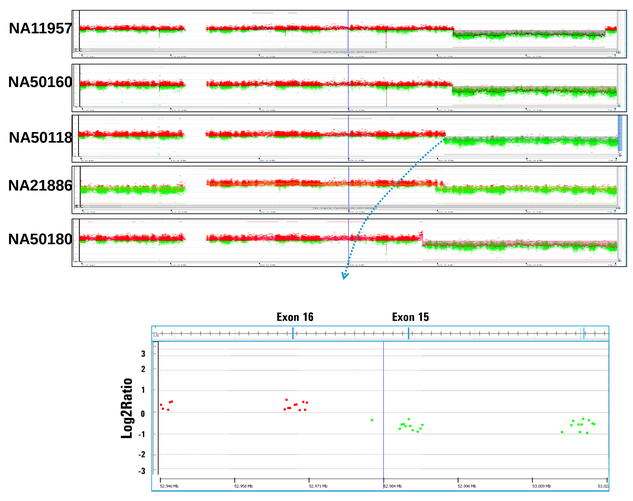


Figure 5. Whole chromosome view for 18q deletion. 23,928 probe groups were designed to include all exons, exon-proximal regions, and chromosome wide backbone in chromosome 18. In NA50118, the boundary of DNA loss is located between exon 15 and 16 in *TCF4* which is linked to Pitt-Hopkins Syndrome.

Conclusions

- **Agilent's Custom OneSeq provides a comprehensive, flexible, and cost-effective means to identify exon-level copy number changes as well as SNP/INDEL in one assay.**
- **Empirically validated probes for exons, exon-proximal regions and genome-wide backbone produce robust performance to detect any combination of CNV, SNP and INDEL.**

Reference

Kalman L. et al., Quality Assurance for Duchenne and Becker Muscular Dystrophy Genetic Testing, The journal of Melecular Diagnostics, 2011, Vol. 13, No. 2
Peippo M. and Ignatius J., Pitt-Hopkins Syndrome, Mol Syndromol, 2012, 2:171-180

For Research Use Only. Not for Use in Diagnostic Procedures.