Blueprint Genetics

# Validation of low-coverage whole genome sequencing for detection of copy number aberrations in inherited disorders

**Juha Koskenvuo**[1], Pertteli Salmenperä[1], Miko Valori[1], Ilari Scheinin[2,3], Massimiliano Gentile[1], Samuel Myllykangas[1]

[1]Blueprint Genetics, Biomedicum 2U, Tukholmankatu 8, Helsinki, 00290, Finland
[2]Department of Pathology, VU University Medical Center, Amsterdam, The Netherlands
[3]Department of Pathology, Haartman Institute and HUSLAB, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland

## Detection of Copy Number Aberrations using low-coverage whole genome sequencing

Detection of copy number aberrations in clinical diagnostics has been limited to few genes in MLPA and low resolution in microarrays. We have developed and validated a low-coverage whole genome sequencing assay for genome-wide and high-resolution detection of copy number aberrations (CNAs) from inherited disorders.

Low-coverage whole genome sequencing and a segmentation analysis were applied to detect CNAs (**Figure 1**). Illumina NextSeq500 sequencing system was applied to generate paired-end 40 base reads that were divided into 5 kb bins. Difficult to sequence regions (bins representing 5.8% of the genome) were filtered out. Read counts were corrected for GC content and average mappability of each bin. Segmentation and calling algorithms (ODNAseq, DNAcopy and CGHcall) were applied to detect CNAs.

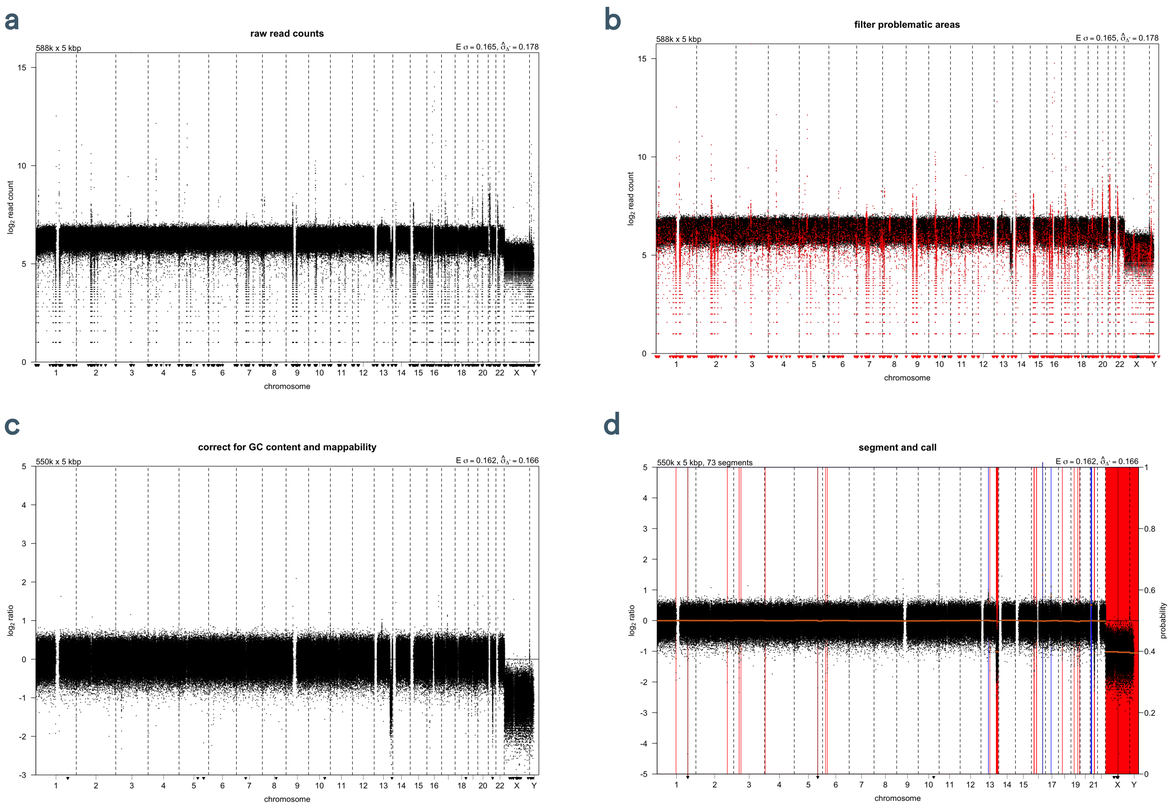


**Figure 1** Methods for detecting copy number aberrations using low-coverage whole genome sequencing. a Sequencing and binning, b filtering, c GC content and mappability and d segmentation and calling.

## Confirmation of CNA findings using quantitative PCR

Quantitative PCR is applied to confirm pathogenic and likely pathogenic Del/Dup variants from probands as well as for screening of disease-causing mutations in the family members. Validation of Del/Dup confirmation assay was performed using 15 well-characterized reference samples and 32 Del/Dup assays (**Table 3**).

**Table 3** Analytic validation of quantitative PCR assay for CNA confirmation.

| Performance metric | Value | Measurements |
|---|---|---|
| Sensitivity | 0.954 | TP: 21 |
| Specificity | 0.920 | TN: 23 |
| Positive predictive value | 0.913 | FP: 2 |
| Accuracy | 0.936 | FN:1 |
| Repeatability | 1.000 | |
| Reproducibility | 1.000 | |

## In-process quality control

A well-characterized golden-standard reference sample (NA12878) is added into each sample processing, sequencing and data analysis batch to control for quality of the testing and to assess the sensitivity of the performed assays (**Table 4**). For each clinical sample and in-process quality control sample, number of mapped reads, mapping rate and predicted >25kb are measured (**Figure 4**) to ensure that each analysis fulfills the criteria derived from the analytic validation studies.

**Table 4** Quality metrics and acceptance criteria for clinical testing

| Quality metrics | Acceptance criteria | Actual* |
|---|---|---|
| In-process reference sample sensitivity (>50 kb) | >0.8 | 1.0 |
| Number of mapped reads | >39M | 58.2M |

* 3-month average.

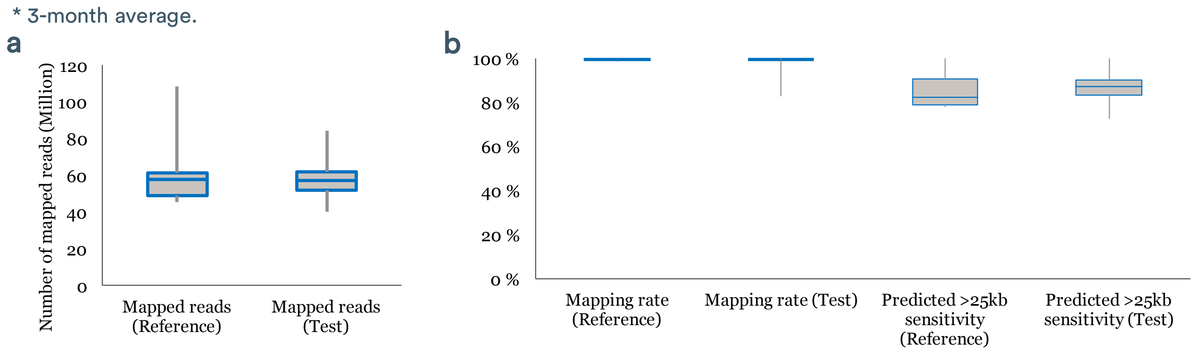


**Figure 4** In-process quality control of low-coverage CNA detection during a 3-month period. **a** Number of mapped reads, **b** mapping rate and predicted >25kb sensitivity are measured for reference samples and test samples.

## Analytic validation of CNA detection

Blueprint Genetics performs analytic validation of all laboratory and data analysis assays per ACMG guidelines *(1)*. Metrics included in the analytic validation of Blueprint Genetics diagnostics assays are described in **Table 1**.

**Table 1** Metrics applied in the analytic validation of CNAs in inherited disorders.

| Validation metric |
|---|
| PRECISION is a statistical measure of the performance of the assay to generate correct test result. For estimation of the precision, true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) are measured and sensitivity, specificity, positive predictive value and accuracy are calculated. |
| CLINICAL SENSITIVITY provides a statistical measure of the assays clinical performance. It reflects the assay's ability to provide a diagnosis in specific clinical use cases. Clinical sensitivity is measured by calculating the extent in which all clinical diagnosis scenarios are met. |
| REPORTABLE RANGE is the functional range of an assay over which the analyte can be analyzed. |
| REPEATABILITY is the technical variation in measurements taken by a single person on the same instrument, on the same experiment, under the same conditions, and in a short time. |
| REPRODUCIBILITY is the ability of a test result to be duplicated under all variable conditions (different users, between reagent lots, using different instruments and different testing times). |

Blueprint Genetics applies independent, publicly available sample materials and data sets as the reference materials in all validation studies to ensure full traceability of the validation results. The CNA calls from the reference samples were compared to Affymetrix Genome-Wide Human SNP Array 6.0, G-banded karyotyping analysis and fluorescence in situ hybridization (FISH) results (**Figure 2**). 28 reference samples with 34 confirmed chromosomal aberrations and the golden standard reference sample (NA12878) were applied in the validation (**Table 2**). Sensitivity to detect >25kb Del/Dups was estimated based on diversity of the sequencing library (**Figure 3**).

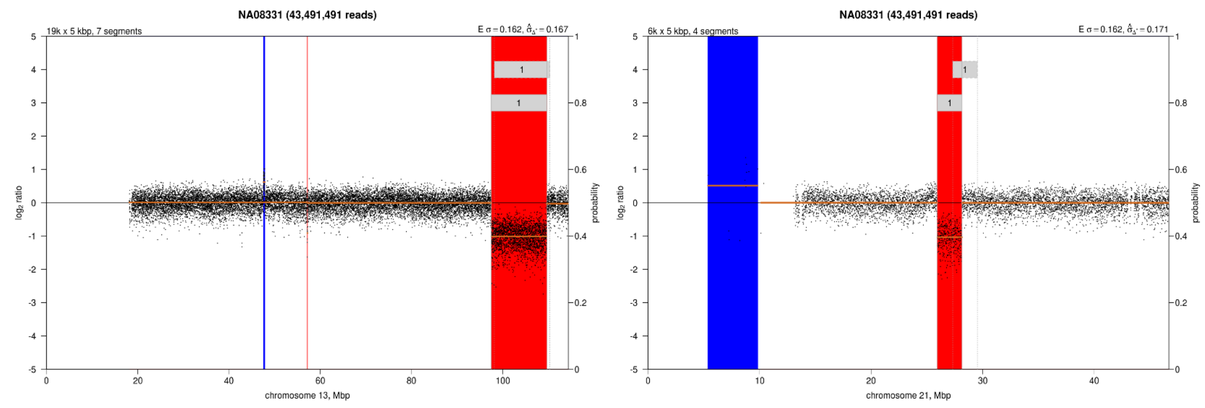


**Figure 2** Known deletions in chromosomes 13 and 21. CNA calls aligned with microarray data showing 1 copy in different genome coordinate systems (hg19/hg39). Red bars depict deletions and blue bars depict duplications.

**Table 2** Analytic validation of detection of CNAs.

| Performance metric | Value | Measurements |
|---|---|---|
| Sensitivity (chromosomal aberrations, >100 kb) | 0.97 | TP/FN: 33 / 1 |
| Sensitivity (Del/Dups) | | |
| 10-25 kb | 0.06 | TP/FN: 16 / 273 |
| 25-50 kb | 0.76 | TP/FN: 143 / 44 |
| >50 kb | 0.99 | TP/FN: 101 / 1 |
| Clinical sensitivity (% genes that are analyzed*) | | 130 284 out of 138 259 segments used |
| Clinical sensitivity (% genome that is analyzed**) | 94.2 % | 527 553 out of 619 150 segments used |
| Reportable range | 10 kb and larger | |
| Repeatability | 0.86 | |
| Reproducibility | 0.85 | |

* Coding regions of the human genes contain regions that are affected by pseudogenes, repeats and extreme GC-content that are not reproducibly analyzed using short read sequencing.

** Human genome contains difficult to sequence regions such as centromeres and telomeres.

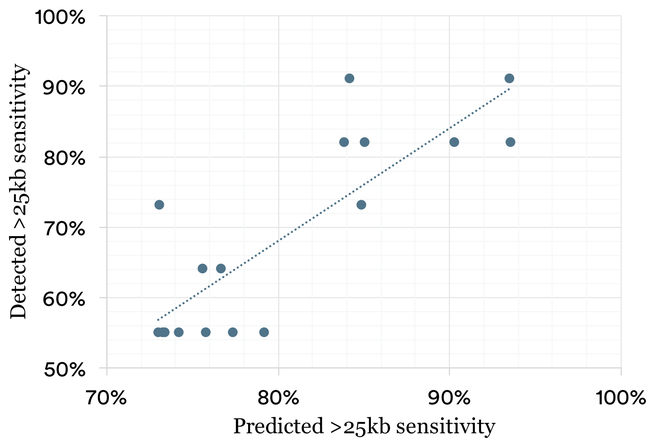


**Figure 3** Correlation ($R^2$=0.68) between predicted and detected >25kb Del/Dup sensitivity.

## References

(1) Rehm HL, Bale SJ, Bayrak-Toydemir P, Berg JS, Brown KK, Deignan JL, Friez MJ, Funke BH, Hegde MR, Lyon E; Working Group of the American College of Medical Genetics and Genomics Laboratory Quality Assurance Commitee. ACMG clinical laboratory standards for next-generation sequencing. Genet Med. 2013 Sep;15(9):733-47