Estimating true-positive rates of CNV detection by WES read depth-based analysis in 434 assays

Fátima Lopes[1], Paulo Silva[1], Susana Sousa[1], Sara Morais[1], Joana Sá[1], Ana Filipa Brandão[1], Ana Lopes[1], Rita Bastos-Ferreira[1], Alexandra Lopes[1], João Parente Freixo[1], Jorge Sequeiros[1,2], Jorge Oliveira[1]

[1]CGPP-IBMC – Centro de Genética Preditiva e Preventiva, Instituto de Biologia Molecular e Celular; i3S – Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Portugal
[2] ICBAS – Insitutto de Ciências Biomédicas Abel Salazar, Universidade do Porto, Portugal

Presenter: Fátima Lopes; fatima.lopes@ibmc.up.pt

Background

Whole-exome sequencing (WES) enables the simultaneous analysis of all coding regions of the human genome, through a single methodology based on next-generation sequencing (NGS). Although NGS started by primarily targeting the detection of single nucleotide variants (SNVs) or small insertion or deletions (InDels), bioinformatics tools have been developed to detect copy number variants (CNVs).

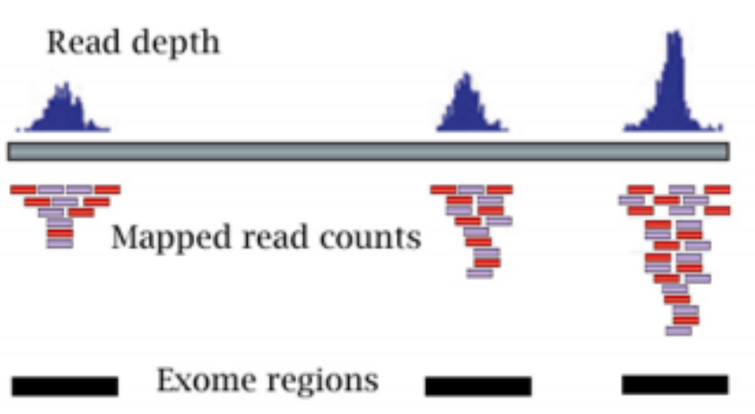

Figure 1 – Schematic representation of a read depth-based approach to detect CNVs using WES data. Adapted from [1]

Aim

To evaluate the efficacy of detecting CNVs through read depth-based analysis in routine diagnostics.

Methods

- Samples: 2,426 samples were analyzed by WES;
- Capture: Agilent's SureSelect Human AllExon or Twist's Human Core Exome Kit;
- Sequencing: Illumina platform (HiSeq4000 or NovaSeq);
- Method: read depth-based analysis;
- Software: VarSeq (Golden Helix);
- CNVs confirmation: 434 CNVs considered likely to contribute to patients' phenotype were confirmed by na orthologue method (qPCR or MLPA).

Results

Table I – True-positive and false-positive rates for the 434 CNVs tested distributed by capture kit.

	Total of 434 CNVs tested			
	Type	**Agilent kit**	**Twist kit**	*TOTAL*
True Positives (TP)	**Loss**	76	44	*100*
	Gain	23	24	*67*
	Total	99 (31%)	68 (59%)	*167*
False Positives (FP)	**Loss**	118	24	*142*
	Gain	101	24	*125*
	Total	219 (69%)	48 (41%)	*267*
TOTAL		**318**	**116**	***434***

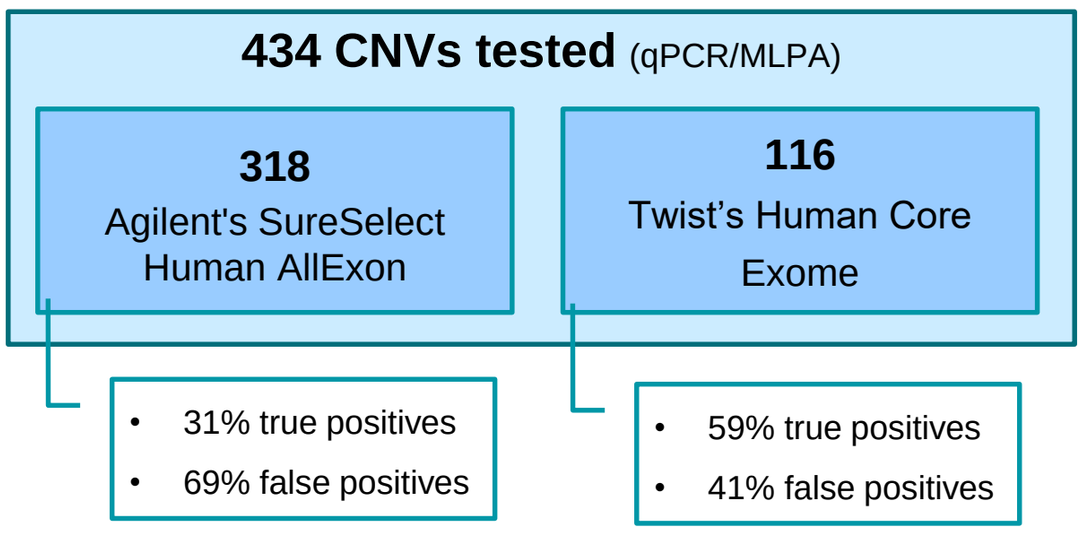

Figure 2 – qPCR and MLPA confirmation of the 434 CNVs studied. The vast majority of the CNVs tested (318) were detected using the Agilent capture kit; only 116 CNVs were detected using the Twist capture kit.

Table II – True-positive and false-positive rates for the 434 CNVs tested distributed by capture kit and by exon number.

	Total of 434 CNVs tested			
	Type	**Exon number**	**Agilent kit**	**Twist kit**
True Positives (TP)	**Loss**	≤ 3 exons	39 (51%)	17 (61%)
		≥ 4 exons	37 (49%)	27 (39%)
	Gain	≤ 3 exons	5 (22%)	4 (17%)
		≥ 4 exons	18 (78%)	20 (83%)
False Positives (FP)	**Loss**	≤ 3 exons	114 (97%)	23 (96%)
		≥ 4 exons	4 (3%)	1 (4%)
	Gain	≤ 3 exons	79 (78%)	18 (75%)
		≥ 4 exons	22 (22%)	6 (25%)

Conclusions

- From our experience:

  - improvements in capture and sequencing technologies contributed for the increase of the true positive rate increase in our laboratory;

  - accumulated internal laboratory experience also played an important role;

- The inclusion of read depth-based CNV detection from NGS data in routine bioinformatics pipelines is a cost-effective add-on for diagnostic laboratories.

- Despite this advantage, there are well-known intrinsic limitations and false-positive rates remain high.

- Considering these limitations in routine practices → depth-based CNV detection requires compulsory confirmation by another methodology.

References

[1]- Zhao M et al, 2013; PMID:24564169