The use of ultrasensitive targeted enrichment and sequencing for non-invasive genetic analysis of cell-free circulating DNA of advanced tumors

Anna Wąsowska[1,*], A. Kowalik[2], M. Chłopek[2], M. Krawczyk[1], P. Ogrodnik[1], E. Suchecka[1], M. Jurkowska[1], Anna Boguszewska-Chachulska[1]

[1] Genomed SA, Ponczowa 12, 02-971 Warsaw, Poland; [2] Holy Cross Cancer Centre, Artwińskiego 3, 25-734 Kielce, Poland

* Presenting and contact author, e-mail address: awasowska@genomed.pl

Introduction

Liquid biopsy aims to detect genetic profiles and subclonal variations in the heterogeneous cancers in minimally invasive manner by means of **circulating tumour DNA (ctDNA) sequencing**. Therefore, the current advances in ultrasensitive targeted enrichment methods can be adapted in the analysis of ctDNA for cancer detection, risk stratification and therapy monitoring.

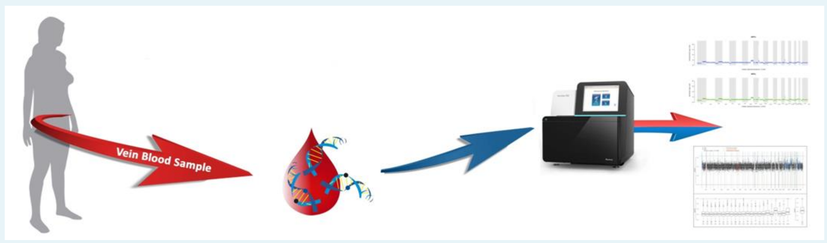

Figure 1 Project concept : cancer – related genes analysis with liquid biopsy

Aim

Development of non-invasive genetic diagnostics and cancer prevention methods, based on the liquid biopsy - **analysis of ctDNA**, using the new generation sequencing technology (NGS) for Polish cancer patients and people at risk. For this purpose, Agilent SureSelect XT HS technology and custom SureSelect bait library was used for targeted enrichment of cancer-related genes in ctDNA samples enabling analysis of changes in target genes.

Materials and methods

- **SureSelect Target Enrichment Bait Library Design** – a custom SureSelect capture probes library of total length of 229kb targeting cancer-related genes was designed in Agilent SureDesign (https://earray.chem.agilent.com/suredesign).
- **Cell-free DNA Isolation** – 3-5 ml plasma samples of 9 breast cancer cases were collected. Cell-free DNAs were isolated using Qiagen Qiacube system and QIAamp Circulating Nucleic Acid Kit. The quantity/quality and size distribution was checked using Agilent Bioanalyzer High Sensitivity DNA chips.
- **Library Preparation** – from 4.1 ng to 39.2 ng of input cfDNA was used for library preparation performed according to manufacturer's protocol. 10 cycles (samples 31cfDNA and 33cfDNA) or 9 cycles of pre-hybridization PCR were performed. The whole volume or 1200 ng (samples 31cfDNA and 33cfDNA) of library product was taken to hybridization followed by post capture washing steps at 70°C. Post-hybridization amplification with 11 cycles (samples 31cfDNA and 33cfDNA) or 12 cycles was performed.
- **Sequencing** – sequencing was performed using Illumina NextSeq500 at Genomed SA in 151 paired-end mode with 10-bp Molecular Barcode (MBC) read.
- **Analysis** – fastq files for paired-end reads and MBC reads were generated using Illumina bcl2fastq Conversion Software. For mapping analysis of data without MBC deduplication trimmimg was conducted using Cutadapt followed by mapping to human genome (hg19) with bwa. For data analysis with deduplication Agilent SureCall 4.0 with default parameters was used. Sequencing metrics were calculated using samtools 1.7, fastx-toolkit, Agilent SureCall 4.0 or in-house written python and R scripts.

Results

cfDNA concentrations ranged from 0.41 ng/µl to 3.92 ng/µl. A total amount of pre-hybridization PCR product varied from 260 ng to 1795 ng. The final molarity of libraries was from 0.96 nM to 4.94 nM. The libraries were sequenced to a total number of raw reads >40 M. giving from 43 M to 96 M total raw reads per sample.

Figure 2 A comparison of number of raw reads, reads mapped to the human genome and reads mapped to targeted regions among samples. Input cfDNA amount for library preparation is indicated in brackets. Regardless the large number of raw reads, only a small proportion was mapped to hg19 and on target.

	29cfDNA	31cfDNA	32cfDNA	33cfDNA	35cfDNA	36cfDNA	38cfDNA	39cfDNA	40cfDNA
Input cfDNA	4.5 ng	39.2 ng	5.8 ng	15.2 ng	4.1 ng	10.0 ng	6.7 ng	4.1 ng	10.0 ng
Raw reads	43 M	90 M	65 M	94 M	52 M	96 M	71 M	68 M	81 M
% on-target before deduplication	74,66%	69.90%	53.77%	68.99%	73.73%	76.41%	71.80%	73.19%	73.58%
% duplicates	91,56%	89,85%	93,19%	91,20%	93,18%	92,36%	94,44%	93,62%	95,44%
Average read depth in target regions before deduplication	6920x	7612x	7158x	7548x	7042x	7420x	7269x	7217x	7252x
% on target after deduplication	40,44%	49,21%	32,49%	40,12%	31,63%	40,31%	26,20%	31,70%	29,38%
Average read depth in target regions after deduplication	419x	1326x	415x	919x	326x	835x	303x	376x	318x
% analyzable reads	**3,00%**	**4,07%**	**2,00%**	**2,86%**	**1,93%**	**2,70%**	**1,31%**	**1,77%**	**1,23%**

Table 1 Sequencing metrics for 9 samples. Comparison of on target coverage and read depth before and after deduplication based on Molecular Barcodes. Samples highlighted in grey were prepared with one more pre-hybridization PCR cycle and one less post-hybridization PCR cycle.

Slightly lower percent of duplicates was observed for samples with the highest input cfDNA amount for library preparation. Samples 31cfDNA and 33cfDNA had the best pre-hybridization PCR yield, what ensured the optimal product amount for hybridization. However, they had a slightly lower percent of reads on target before MBC deduplication. The rest of samples had a suboptimal pre-hybridization library amount. This suggests that the number of PCR cycles may have influence the proportion of on target reads before deduplication. The percent of reads mapped on target after deduplication may be correlated with the input cfDNA amount as the sample with the highest input had the best value. The average coverage is similar among the samples regardless the total number of raw reads. Increasing the number of raw reads impacts the on target average coverage after deduplication.

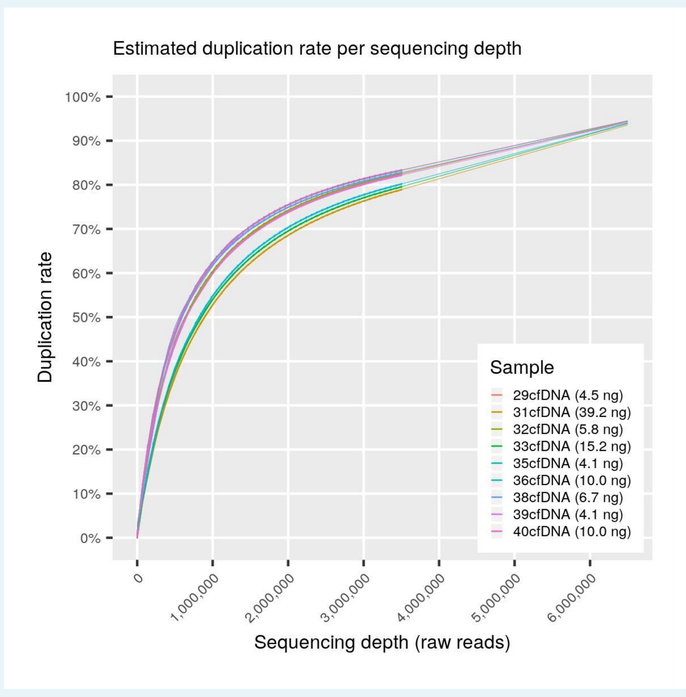

Figure 3 Estimation of duplication rate with increasing sequencing depth. Lines in bold are drawn based on average of randomly subsampled 10 measurements at 51 different sequencing depths in the range 0 - 3.5M of raw MBC reads for each flowcell lane separately. Estimation beyond the range was fitted using generalized additive model. As observed, samples with higher input cfDNA amount have lower duplication rate for the lower sequencing depth, however, for >6M raw reads duplication rate achieves similar level of >90% of duplicates.

Conclusions

The idea of decreasing the false positive rate for low level variants in cancer analysis is very promising, however, limited by the yield of cfDNA isolation and, in consequence, by the amount of cfDNA used for library preparation. Low-input cfDNA libraries result in high percent of MBC duplicates, which are excluded from analysis. Therefore, only a small percentage of sequenced reads is valueable for ctDNA variant analysis focused on low frequency variant detection at high coverage. Achieving the high on target coverage of 5000x-10000x, as recommended, needs compensatory increase of sequencing reads, resulting in a very high sequencing cost. Nevertheless, the large off-target rate may be beneficial for reconstruction of genome-wide copy number profiles in cancer samples.

Acknowledgements

Co-financed from the European Union funds within Regional Operational Programme of the Mazowieckie Voivodeship 2014-2020 - RPMA.01.02.00-14-6209/16.