Genetic profiling and early cancer detection using NGS-based Liquid Biopsy

A. Wąsowska [1*], E. Matczyńska [1], M. Krawczyk [1], P. Łyszkiewicz [1], A. Kowalik [2], M. Chłopek [2], J. Kosakowski [1], M. Jurkowska [1, 1] and A. Boguszewska-Chachulska [1, 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

European Human Genetics Virtual Conference (ESHG 2020.2), June 6-9, 2020

INTRODUCTION

There is an urgent need for early and efficient cancer detection, risk stratification, and therapy monitoring. **Circulating cell-free tumor DNA (ctDNA)** is considered to be a promising cancer development marker, therefore, may provide a sensitive and representative target in NGS-based diagnostic methods.

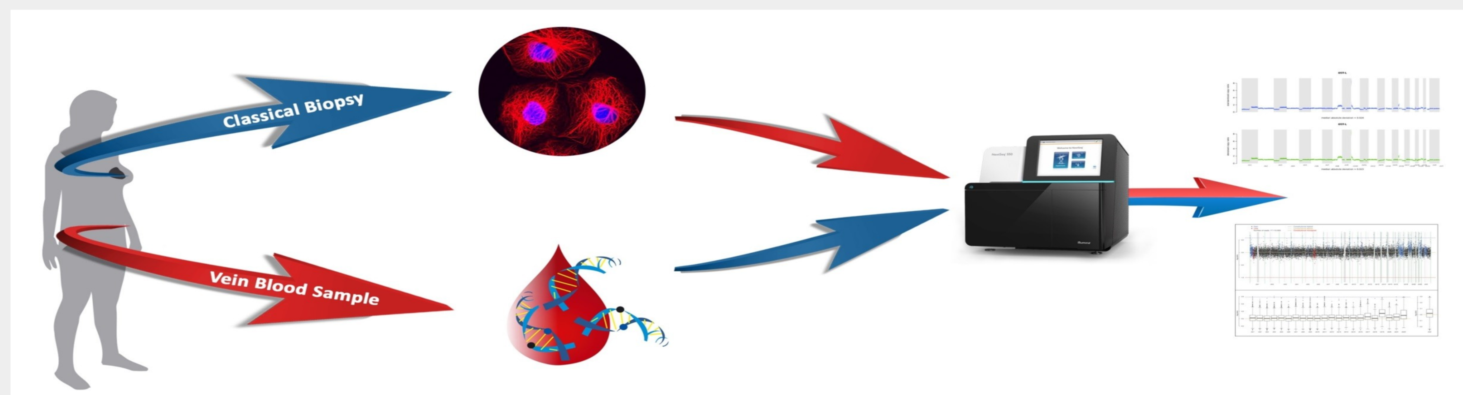


Fig. 1 Project concept: cancer genome analysis, combining deep WGS for tumour tissue with liquid biopsy

AIM

The project aimed at developing a non-invasive genetic diagnostics and monitoring of cancer progression, regression and/or relapse, based on the **liquid biopsy (LB)** for Polish cancer patients and people at risk. For this purpose a combination of **deep Whole Genome Sequencing (dWGS)** of advanced tumour specimens, **standard WGS** of corresponding blood samples was performed to identify potential biomarkers that subsequently served as objective for data analysis by cfDNA **targeted sequencing** of breast cancer-related gene panel. Additionally, **shallow WGS (sWGS)** of cfDNA was performed for CNV analysis.

MATERIALS

The study group of patients was recruited at Holy Cross Cancer Centre (Kielce, Poland). 15 breast and colorectal frozen tumour specimens were collected for 12 patients (2 specimens in 3 cases) together with the corresponding 9 blood samples. Cell-free DNA (cfDNA) of 82 breast, colorectal or lung cancer cases was isolated from plasma.

METHODS

- **cfDNA Isolation** – 3-5 ml plasma samples of 82 patients were collected. Qiagen Qiacube system and QIAamp Circulating Nucleic Acid Kit were used for cell-free DNA isolation. The quantity/quality and size distribution was checked using Agilent Bioanalyzer High Sensitivity DNA chips.The efficiency of isolation (ctDNA copy number per µl of plasma) and different DNA sub-populations were checked with droplet PCR.

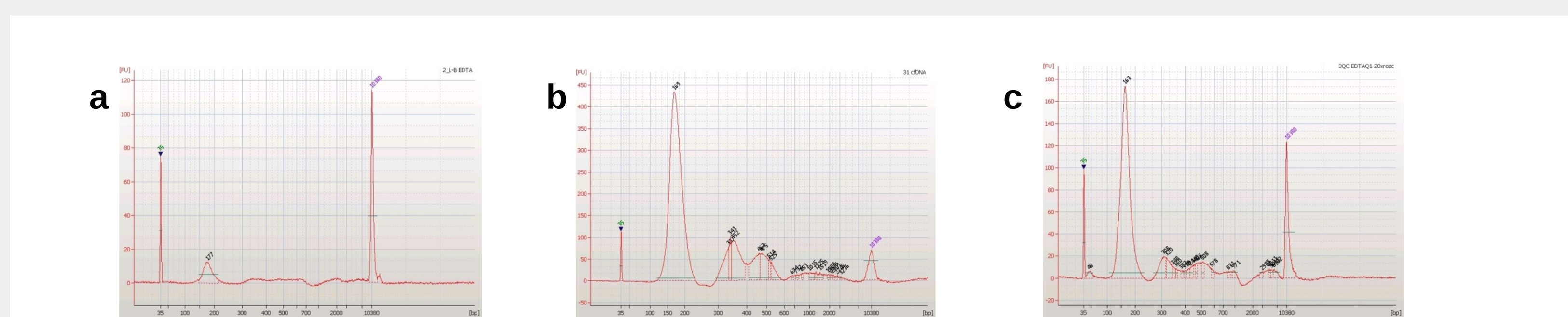


Fig. 2 Differences in cfDNA size distribution for **(a)** healthy control samples, **(b)** early stage and **(c)** metastatic stage breast cancer samples

- **dWGS and WGS** – 9 breast cancer DNA samples with 7 corresponding germline DNA samples met the quality criteria and were chosen for 2x100 paired-end sequencing using BGISEQ-500 sequencers (BGI Tech, Hongkong, China). The expected genome coverage was 90x for tumour samples (dWGS) and 30x for germline samples (standard WGS).
- **sWGS** – 500 ng of cfDNA was used for PCR-free library preparation with NEBNext Ultra Library Prep Kit for Illumina (New England Biolabs, Ipswich, MA, USA).
- **targeted sequencing of breast cancer-related gene panel** – 2.6-10 ng of cfDNA was used for library preparation according to the QIAseq Human Breast Cancer Panel workflow (Qiagen).
- sWGS and QIAseq libraries were qualified using Agilent Bioanalyzer High Sensitivity DNA chips and quantified using qPCR. Sequencing was performed using NextSeq500 sequencer with 2x150 paired-end reads.
- **dWGS and WGS analysis** – fastq files from BGISEQ-500 sequencing were trimmed using Trimmomatic (version 0.38). Mapping to human reference genome (hg38) was conducted with ALT-aware BWA version 0.7.15-r1140. bamsormadup (biobambam2 package), version 2.0.89.was used for deduplication. Somatic SNV and Indel variants were identified using Mutect2 from GATK package, Strelka and Vardict software. Copy ratio alterations were detected by CNVkit. Standardized and denoised read counts in 5kb bins were ploted on the whole-genome plots.
- **QIAseq data analysis** – fastq files generated by Illumina bcl2fastq Conversion Software were uploaded to QIAGEN server. UMI deduplication, mapping to the reference genome and variant detection were performed by implemented in QIAGEN server pipeline for QIAseq Targeted Panel analysis.
- **sWGS** – fastq files were generated using Illumina bcl2fastq Conversion Software. CNV alaysis of sWGS data was conducted using Wisecondor software with several bin sizes. The ctDNA fraction was estimated with ichorDNA package.
- Variant functional annotation and interpretation was reported by Personal Cancer Genome Reporter (PCGR)

RESULTS

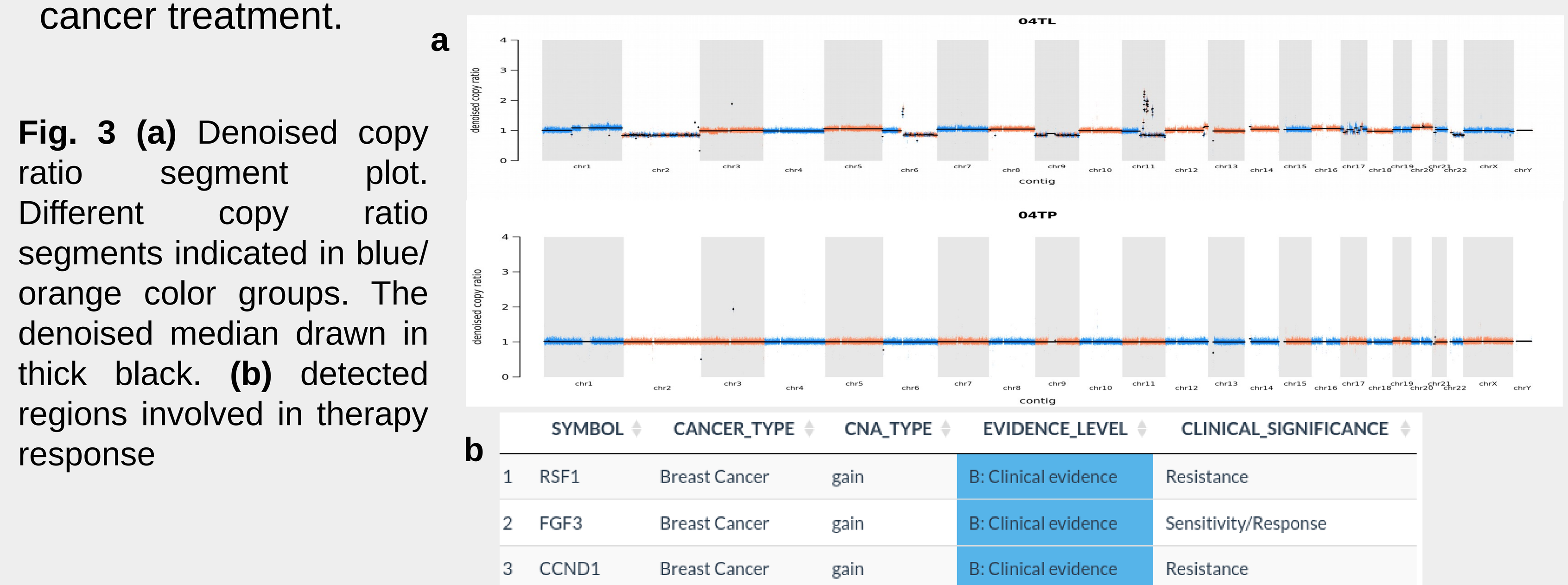
The average coverage was 102x for dWGS tumour samples and 39x for WGS germline samples. The mutational burden for the dWGS tumours samples in the chosen 36Mb ROI was estimated from 0.17 mutations/Mb to 2.36 mutations/Mb, most of which were classified as non-coding variants. For six breast cancer samples the **PIK3CA** somatic variants of strong clinical evidence were identified (**H1047R and E545K**). In one samples a variant of potentional clinical significance was observed in **SF3B1** gene (**K700E**). None of these variants was found in the corresponding QIAseq sequencing data from liquid biopsy samples, while somatic variants identified in ctDNA were missing in the tumour DNA, probably due to subclonal variants. Liquid biopsy identified three variants of strong clinical evidence (**KRAS: G13D, PIK3CA: L1036M, L1036***) in ctDNA of patients with colorectal cancer.

Sample report	variants of strong clinical significance	variants of potential clinical significance
04TL	0	0
04TP	0	0
05TL	PIK3CA (E545K)	0
05TP	PIK3CA (H1047R)	0
07T	0	0
09T	0	0
10T	PIK3CA (H1047R)	0
13T	0	0
21T	PIK3CA (H1047R)	0
22T	PIK3CA (H1047R)	0
27T	PIK3CA (E545K)	SF3B1 (K700E)

Tab. 1 Variants identified in dWGS of tumour samples

CNV analysis

As expected, the CNV analysis performed for tumour samples allowed to distinguish differently advanced tumours for the same patient (Fig. 3). The genome loss/gain profile visualises detected cancer genome instability, covering regions with proto-oncogenes and some regions with clinical evidence of resistance or sensitivity to cancer treatment.



	SYMBOL	CANCER_TYPE	CNA_TYPE	EVIDENCE_LEVEL	CLINICAL_SIGNIFICANCE
1	RSF1	Breast Cancer	gain	B: Clinical evidence	Resistance
2	FGF3	Breast Cancer	gain	B: Clinical evidence	Sensitivity/Response
3	CCND1	Breast Cancer	gain	B: Clinical evidence	Resistance

Fig. 3 **(a)** Denoised copy ratio segment plot. Different copy ratio segments indicated in blue/ orange color groups. The denoised median drawn in thick black. **(b)** detected regions involved in therapy response

The median coverage for sWGS samples was 0.30x and the estimated ctDNA fraction varied from nearly 0% to almost 32%. CNV analysis of the sWGS data also shows genome instability, however, inconsistent with corresponding tumour dWGS results in some cases. (Fig.4).

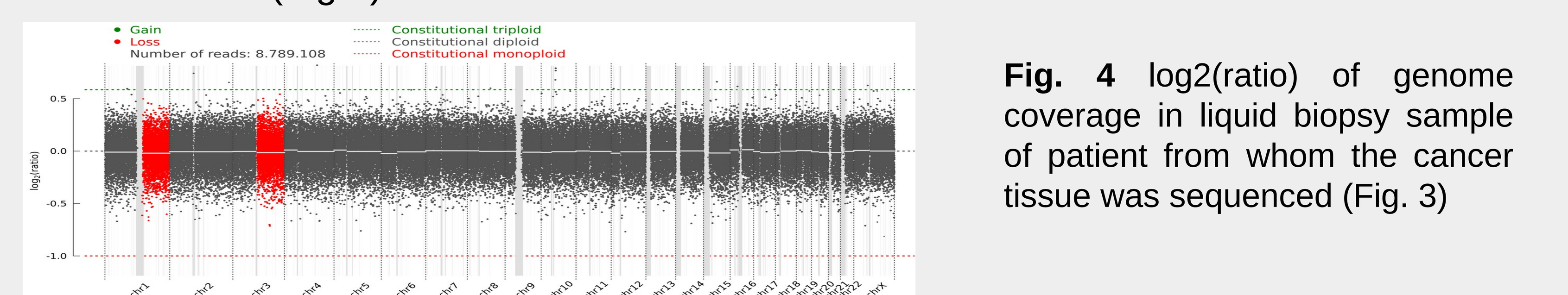


Fig. 4 log2(ratio) of genome coverage in liquid biopsy sample of patient from whom the cancer tissue was sequenced (Fig. 3)

sWGS CNV analysis was more successful in patients with advanced/metastatic cancer, but it also allowed to identify a possible early cancer development in control samples (Fig.5).

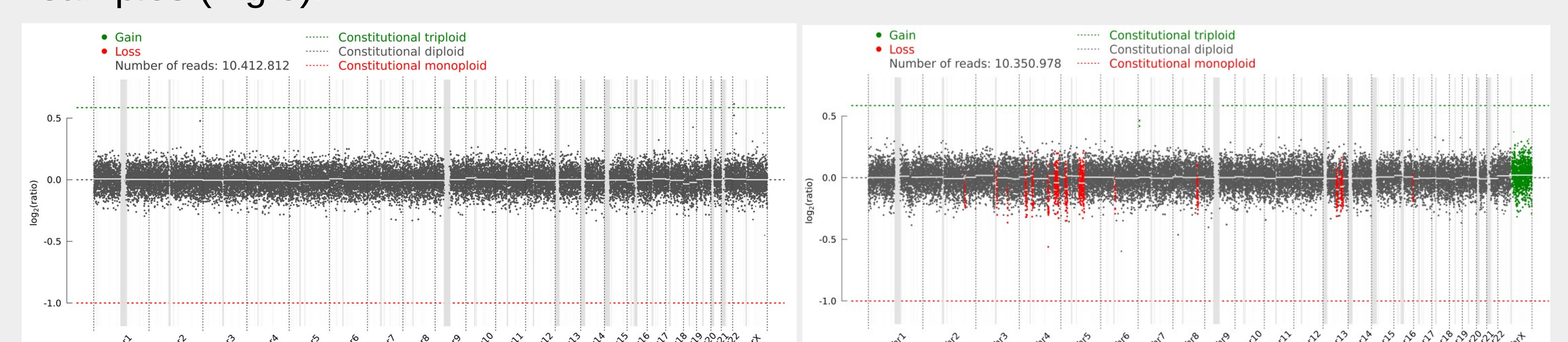


Fig. 5 log2(ratio) of genome coverage in control samples. **Left:** healthy control. **Right:** possible early cancer profile in control sample.

CONCLUSIONS

A method allowing breast cancer detection at a relatively early stage, based on sWGS and CNV analyses, was developed. Discrepancies observed between the tumour and the liquid biopsy results may be partially due to clonal variants in the heterogenous tumour tissues, but it also suggests the need for a further improvement regarding sensitivity of liquid biopsy methods.

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.