INFERRING DNA METHYLATION FROM LOW-PASS WGS OF CELL-FREE DNA USING FRAGMENTATION PATTERNS IN OVARIAN CANCER SAMPLES

Péter Hunyadi[1], Abigél Balla[1,2], Szabolcs Máté[3], Ágnes Égető[3], János Rigó Jr.[2,3], Jakub Styk[4], Silvia Bokorová[4], Tatiana Sedláčková[4], Tomas Szemes[4], Orsolya Biró[1,2]

[1]Clinomics Europe Ltd.; [2]Semmelweis University Doctoral School, Surgical Medicine Division; [3]Semmelweis University, Faculty of Medicine, Dept. of Obstetrics and Gynecology; [4]Comenius University Science Park Contact: peter.hunyadi@clinomicseurope.com

Background

- Cell-free DNA (cfDNA) represents extracellular DNA shed into the bloodstream in fragmented form. A portion of cfDNA may originate from cancer cells and its methylation status carries information on progression and tissue-of-origin of cancer. The gold standard method for methylation analysis, whole-genome bisulfite sequencing (**WGBS**), damages cfDNA molecules severely.
- cfDNA fragmentation patterns can be used to circumvent this problem, as these patterns can be observed from whole genome sequencing (WGS) without bisulfite treatment. It has been shown that cfDNA fragmentation holds useful information not only for genetics but for epigenetics applications as well [1].
- Low-pass WGS, also known as shallow WGS (**sWGS**) is a cost-effective technology, especially convenient for cfDNA analysis, to assess widespread information of the genome. On the negative, not point mutations, but only large alterations can be reliably analysed.
- In our study, we aimed to analyze the correlation between cfDNA fragmentome and methylome profile of cfDNA samples from high-grade serous **ovarian cancer** (HGSC) patients.

Methods

- We used samples with matched sWGS and WGBS data – 13 HGSC & 8 control samples
- cfDNA was isolated from plasma samples of HGSC patients and age-matched controls. For the WGS analysis, **TruSeqDNA Nano kit** (Illumina) libraries were prepared and sequenced with low coverage.
- To analyze genome-wide methylation profile, the same set of cfDNA was addressed to bisulfite conversion and library preparation was performed using the **QIAseq Methyl Library kit** (Qiagen).
- **FinaleMe** software was used to derive methylation signature from WGS data and compare case-control groups [2].
- The whole genome data has been filtered down to CpG islands and adjacent regions (CpG island shores), following the methodology of Liu *et al*.

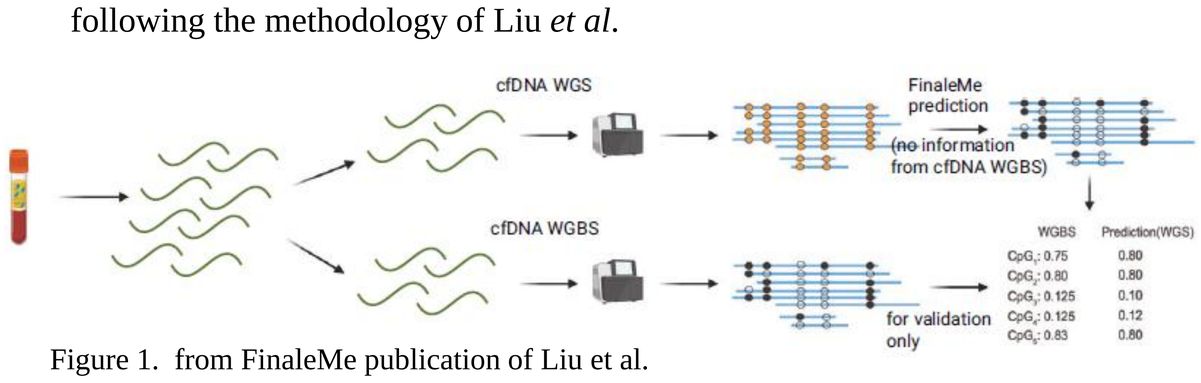

Figure 1. from FinaleMe publication of Liu et al.

Methylation data generation of matched samples happened by two different methods: by prediction from shallow WGS fragmentation with FinaleMe tool and extracted from whole genome bisulphite sequencing.

Results

- To assess the accuracy of the predicted methylation from the fragmentomics pattern, WGBS data was the validation tool. The degree of similarity was determined between inferred methylation and WGBS data the Pearson correlation coefficient metric is utilized. For each sample a correlation value is calculated, which, for our samples, lies between 0,47-0,67 (Figure 2.), indicating the expected postive correlation. In this regard, some difference was noticed between disease-healthy sample groups.
- In the next step, we were curious, if the correlation value increases when we filter the CpG list to sites with larger read coverage. The data reflected this expectation, as with even the smallest incremental coverage increase at CpG sites, a correlation increase tendecy followed cohort-wise (Figure 3.).
- Another way to summarize the accuracy of methylation prediction can be seen in Table 1. In the combined data from all samples, the CpG methylation status (defined with three states) was correctly identified in ~72% of CpGs.
- Finally, to reflect on the broadly known fact that methylation in cancer is deeply perturbed in cancer compared to the healthy state, we could show that even unsupervised methods, like PCA, can distinguish the methylation patterns of samples with cancer and healthy samples (Figure 4.). To this end, average methylation values for 1kbp long segments respectively occuring in both data types were used. This group-wise separation was not only recognizable in the bisulphite data, but also in the inferred methylation from cfDNA cleavage signatures.

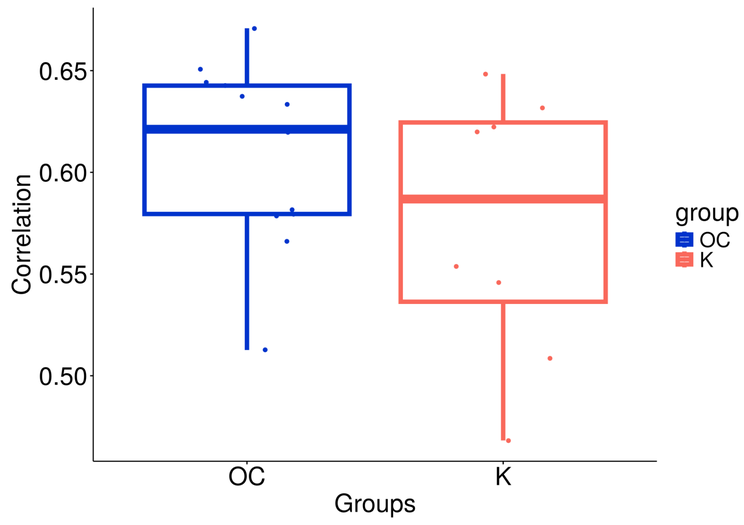

Figure 2. Pearson correlation coefficient of predicted methylation from sWGS and methylation of WGBS. Samples separated by cancer and healthy groups. OC: ovarian cancer; K: healthy control

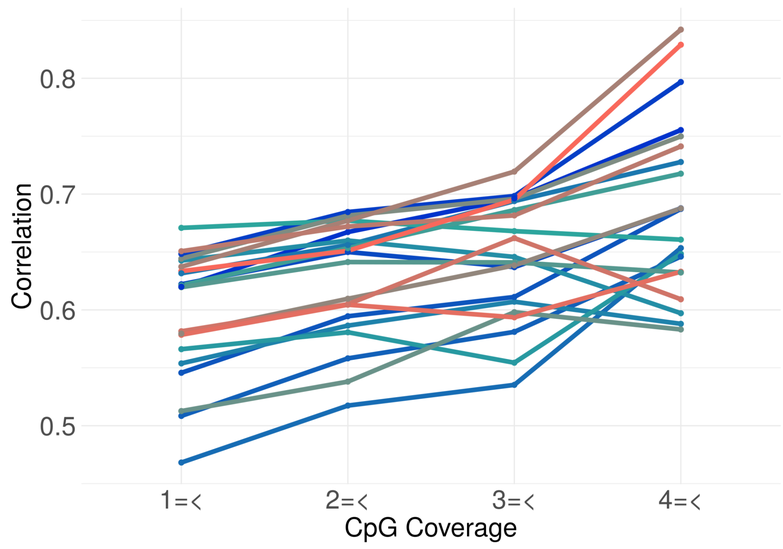

Figure 3. Correlation coefficient shown for each sample separately by selecting CpGs with higher and higher coverage. Columns from left to right: All CpGs retained; CpGs with at least 2 covering reads in both data type retained; same criteria but with 3 covering reads; same criteria but with 4 covering reads

sWGS \ WGBS	methylated	partially methyl.	unmethylated
methylated	32,25%	3,48%	4,57%
partially methyl.	2,01%	0,37%	0,69%
unmethylated	12,73%	4,15%	39,74%

Table 1. CpG methylation status compared between both data types, numbers pooled from all samples. Methylated: Over 80% methylated. Unmethylated: Under 20% methylated. Partially methylated: 20%-80%

Summary & Conclusions

- With higher mean coverage of studied genomic regions, the predicted methylation resembles „ground truth” methylation from WGBS more accurately.
- Methylation is a robust property to distinguish cancer and healthy samples.
- **In the future, analysing the methylome, with its vast amount of use-cases, could be performed through assessing the cfDNA fragmentome, eliminating the disadvantages of bisulfite methodology.**

References:
[1] Zhou, Q. *et al*. Epigenetic analysis of cell-free DNA by fragmentomic profiling. *Proc Natl Acad Sci U S A*. 2022;119(44):e2209852119.
[2] Liu, Y. *et al.* FinaleMe: Predicting DNA methylation by the fragmentation patterns of plasma cell-free DNA. *Nat Commun* **15**, 2790 (2024).

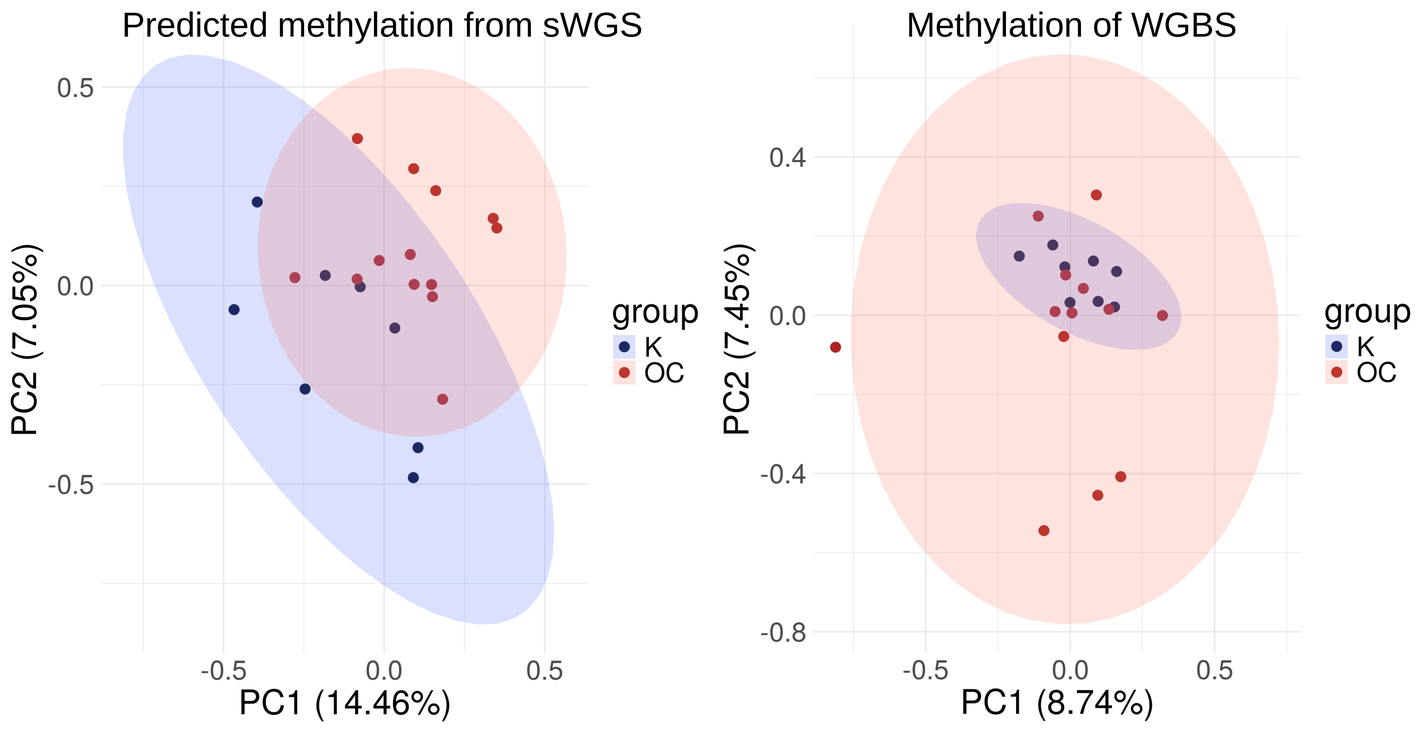

Figure 4. PCA based on average methylation of 1kb long segments. OC: ovarian cancer; K: healthy control