Design and experimental validation of OPERA_MET-A panel for deep methylation analysis by next generation sequencing

Federico Pio Fabrizio1*, Stefano Castellana2, Flavia Centra3, Angelo Sparaneo1, Mario Mastroianno3, Tommaso Mazza2, Michelinia Coco1, Domenico Trombetta1, Nicola Cingolani4, Antonella Centonza5, Paolo Graziano6, Evaristo Maiello5, Vito Michele Fazio1,6,7 and Lucia Anna Muscarella1*

1Laboratory of Oncology, Fondazione IRCCS, Scientific Institute for Research and Health Care Casa Sollievo della Sofferenza, San Giovanni Rotondo, Italy, 2Unit of Bioinformatics, Fondazione IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo, Italy, 3Scientific Direction, Fondazione IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo, Italy, 4Unit of Pathology, Fondazione IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo, Italy, 5Unit of Oncology, Fondazione IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo, Italy, 6Laboratory of Molecular Medicine and Biotechnology, University Campus Bio-Medico of Rome, Rome, Italy, 7Institute of Translational Pharmacology, National Research Council of Italy (CNR), Rome, Italy

DNA methylation is the most recognized epigenetic mark that leads to a massive distortion in cancer cells. It has been observed that a large number of DNA aberrant methylation events occur simultaneously in a group of genes, thus providing a growth advantage to the cell in promoting cell differentiation and neoplastic transformation. Due to this reason, methylation profiles have been suggested as promising cancer biomarkers. Here, we designed and performed a first step of validation of a novel targeted next generation sequencing (NGS) panel for methylation analysis, which can simultaneously evaluate the methylation levels at CpG sites of multiple cancer-related genes. The OPERA_MET-A methylation panel was designed using the Ion AmpliSeq™ technology to amplify 155 regions with 125-175 bp mean length and covers a total of 1107 CpGs of 18 cancer-related genes. The performance of the panel was assessed by running commercially available fully methylated and unmethylated control human genomic DNA (gDNA) samples and a variable mixture of them. The libraries were run on Ion Torrent platform and the sequencing output was analyzed using the "methylation_analysis" plugin. DNA methylation calls on both Watson (W) and Crick (C) strands and methylated:unmethylated ratio for each CpG site were obtained. Cell lines, fresh frozen and formalin-fixed paraffin-embedded (FFPE) lung cancer tissues were tested. The OPERA_MET-A panel allows to run a minimum of 6 samples/530 chip to reach an observed mean target depth >2,500X (W and C strands) and an average number of mapped reads >750,000/sample. The conversion efficiency, determined by spiking-in unmethylated Lambda DNA into each sample before the bisulfite conversion process, was >97% for all samples. The observed
percentage of global methylation for all CpGs was >95% and <5% for fully methylated and unmethylated gDNA samples, respectively, and the observed results for the variable mixtures were in agreement with what was expected. Methylation-specific NGS analysis represents a feasible method for a fast and multiplexed screening of cancer patients by a high-throughput approach. Moreover, it offers the opportunity to construct a more robust algorithm for disease prediction in cancer patients having a low quantity of biological material available.

KEYWORDS
cancer, driver gene, biomarker, methylation, next generation sequencing

1 Introduction

DNA methylation is one of the most largely investigated epigenetic footprints, due to its link with several diseases as well as cancers and autoimmune or genetic disorders (1, 2). Of particular interest is the role of DNA methylation at cytosine residues by the addition of a methyl group (5-mC), most frequently at the CpG (cytosine-phosphate-guanine) dinucleotide motif of mammalian genomes. The fluctuation of 5-mC level is generally associated with variation in the genes expression levels and contributes in many cases to the definition of clinical phenotypes, as well as representing in some clinical contexts a useful diagnostic biomarker in guiding therapeutical choices (3, 4).

In tumors, the epigenome alteration is linked to the neoplastic transformation, cancer progression and invasion, and encouraged a large number of studies focused on discovering predictive and prognostic power of the methylation status at CpG sites (5). To date, the CpG methylation changes represent an attractive source of biomarkers that could have a significant impact on both early and advanced tumors management. Moreover, methyl CpG sites could arise during therapy resistance in patients, so they could represent a new option for the longitudinal monitoring of the neoplastic evolution. Many scientific findings originate from managing large long-existing available datasets and, in this context, the aberrant methylation of the CpGs located at the promoter regions of tumor suppressor genes is gaining prominence (6–8). By contrast, only few large-scale studies have been focused on the role and prognostic impact on cancer of different methylation patterns across the genes, such as intragenic or gene body methylation, that may have a different role in the transcriptional regulation and efficiency of genes machinery (9, 10).

The translation of all this epigenetic knowledge in clinical practice is not complete, mainly due to the lack of high-throughput and quantitatively accurate approaches that can rapidly profile poor quality and quantity of DNA obtained by tumor tissue biopsies. Pyrosequencing and bisulfite-cloning/sequencing are the most widely used methods for low-cost analyses to measure the methylation level at single CpGs of genes in daily practice with reasonable quantitative accuracy. Although highly useful, the limitations of these techniques include samples consumption, short-read length, and low sample throughput (11).

In this methodology paper, we designed and experimentally validated a customized methylation panel for NGS analysis (OPERA_MET-A panel), to scan relevant CpG sites in 155 regions of 18 cancer-related genes mainly involved in the NRF2/KEAP1 pathway and immunotherapy. gDNA from FFPE, frozen tissues and cell lines were tested to simultaneously evaluate their density and average methylation levels for all or single targeted CpG sites. Finally, using this NGS approach, we obtained details about the strand-specific specificity of CpG methylation in targeted regions of the selected genes.

2 Material and methods

2.1 Biological samples selection

Three different types of biological samples were used to validate the OPERA_MET-A panel: n.3 lung paired tumor/non-neoplastic FFPE tissues (830T/N, 881T/N, 889T/N), n.3 lung tumors OCT-embedded (435T, 475T, 495T), n.2 lung cell lines (tumor A459 and normal MRC5) were used. Tissues were collected from anonymous patients, according to the guidelines of the Local Ethical Committee of IRCCS Casa Sollievo della Sofferenza Hospital, Italy, whereas cell lines were purchased from the American Type Culture Collection.
(ATCC, Manassas, VA, USA). A commercially available fully methylated (>95%, CpGenome Universal Methylated DNA, Millipore, Chemicon) and unmethylated (<5%, CpGenome Universal Unmethylated DNA, Chemicon) genomic gDNA with four mixtures (~25%, ~50%, ~75% and ~90%) were used as positive and negative controls to the optimal DNA conversion and library preparation.

2.2 gDNA extraction and sodium bisulfite conversion

gDNA was extracted from cell lines and fresh OCT embedded tissues by using the standard phenol-chloroform procedure (12), whereas 10 µm FFPE sections were extracted using the GeneRead DNA FFPE kit (Qiagen, MD, USA), following the manufacturer’s instruction. Before starting the bisulfite conversion, 0.33% of Unmethylated Lambda DNA (Promega) was added to each single gDNA in order to estimate bisulfite conversion efficiency during samples treatment. For each sample and control mixture (methylated/unmethylated DNA), a minimum of 250ng of gDNA extracted was treated by using Epitect Bisulfite kit (Qiagen) (13), to ensure a minimum of 100ng of converted gDNA recovery for libraries preparation. The quantification of both extracted gDNA and bisulfite treated gDNA was performed using Qubit® ssDNA Assay Kit on Qubit™ 3.0 Fluorometer (Thermo Fisher, Life Scientific).

2.3 NGS workflow

2.3.1 Ion AmpliSeq™ methylation panel design

The OPERA_MET-A panel was designed in collaboration with the Ion AmpliSeq custom design team through the Ion AmpliSeq White Glove Service (Thermo Fisher Scientific); it includes multiple informative and challenging genes regions also tracked by the Infinium Human Methylation 450 BeadChip (Illumina Inc.) or annotated as having a prognostic value in peer-reviewed scientific papers. Specifically, the OPERA_MET-A panel allows the simultaneous quantification and analysis of 155 genomic regions, each of them covered by amplicons with a size range of 125-175bp length (amplicons designed for both Watson and Crick strands) and covers 1107 CpG sites (14, 15) and the two primer pools of the OPERA_MET-A panel, following the manufacturer’s recommendation for the Ion AmpliSeq Library Kit Plus for Bisulfite methylation library production (Thermo Fisher Scientific). Briefly, 2 µL of 5X Ion AmpliSeq™ HiFi mix, 2 µL of 5X Ion AmpliSeq™ of primers pools and 100ng of bisulfite converted gDNA were mixed and amplified following temperature conditions to achieve DNA target amplification. Then, primer sequences were partially digested by adding 2 µL of FuPa Reagent and loaded in a thermal cycler under user guide conditions. Each library was labeled with a single Ion Xpress™ barcode and Ion P1 adapter (Thermo Fisher Scientific) by adding 4 µL of Switch Solution, 2 µL of diluted barcode and 2 µL of DNA Ligase to the reaction mixture, following the temperature conditions recommended by the manufacturer. Finally, the Agencourt AMPure XP bead (Beckman Coulter, Brea, CA, USA) was used to complete three rounds of purifications on a magnet rack and 50 µL of Low TE was added to elute the library. A 100-fold dilution of purified and amplified libraries was quantified using the Ion Library TaqMan quantitation kit (Thermo Fisher Scientific) in a 7900 Real-Time PCR System (Thermo Fisher Scientific), following standard qPCR cycling. Library profiles were checked by capillary electrophoresis using the High Sensitivity D1000 ScreenTape on Tape Station 2200 (Agilent Technologies), and equimolar concentrations of samples libraries were pooled. 40pM of manually pooled normalized libraries were used for template preparation and Ion 530™ chip loading on the Ion Chef Instrument (Thermo Fisher Scientific) using the Ion 520™&530™ ExT Kit-Chef (Thermo Fisher Scientific). Sequencing was performed on Ion GeneStudio S5 (Thermo Fisher Scientific). A maximum of 6 samples were loaded on a single chip per sequencing run.

2.3.3 Bioinformatic analysis

The methylation analysis was performed using the previously described outline (14, 15) and the Ion Torrent Suite™ Software (version 5.10.1) running on the Torrent Server (Thermo Fisher Scientific) was used to process the sequencing data. The "methylation_analysis" Torrent Suite plugin (Thermo Fisher Scientific) was used to analyze the sequencing output of the OPERA_MET-A panel and annotate the percentage of each targeted CpG site. This analysis plugin performs sequencing read alignment onto the W and C strands of the GRCh37/hg19 reference genome and then assesses the methylation status in a strand-specific manner. Reports and text files were generated for each amplicon, containing the number of methylated and unmethylated reads as well as the percentage of methylation per amplicon in relation to the targeted region/CpG sites (14, 15). A summary report was created for each sample that includes the barcode name, the assigned sample name, the total number of reads that cover the target CpGs, and the percentage of methylated reads. In addition, for each barcode, text files with the number of methylated reads, unmethylated
reads, and percent methylation for each amplicon were generated. Separate text files were created for (i) the target CpGs, (ii) all CpGs in the amplicon insert, and (iii) all non-CpG cytosines in the amplicon insert. Each amplicon may contain zero, one, or more CpG targets (hotspots) of interest.

3 Statistical analysis

A simple linear regression model (Observed methylation – Expected methylation), was used to model the relationship between observed and expected global methylation results using the OPERA MET-A panel. The “Expected values” were % of methylation level of different control mixtures obtained from fully methylated and unmethylated commercial gDNA samples (>95%, ~90%, ~75%, ~50%, ~25%, ~10%, <5%). The Wilcoxon signed-rank test was used to assess significant differences in CpG methylation levels between the W and C strands, as defined by genes and patients. All results were deemed statistically significant when p < 0.05. R Foundation for Statistical Computing was used to perform all statistical analyses and plots (version 4.0, packages: ggplot2, dplyr, devtools, PairedData).

4 Results

4.1 Targeted regions selection

The Ion AmpliSeq™ Design Pipeline considered theoretical CpG genome conversion to perform primer design and selection, amplicon tiling and generation of optimal amplicon pooling. The input targets were a list of both single hotspot CpGs (CpG interspersed) and genomic regions containing CpG sites (CpG Island). The target regions of selected genes ranged from 125 to 175bp. They were chosen for validation among those having a potential translational impact in peer-reviewed scientific papers related to immunotherapy in various cancer types (PD-1, PD-L1) (16–27), oxidative stress (KEAP1, NFE2L2, GPX2, PGD, TXNRD1, GCLC, AKR1C2, SRXN1, ABCC2) (28–33), and other cancer/early-stage prognostic biomarkers (SPARC, HAR1B, TMRPSS4, RASSF1A, HOXA9, Pten, CDKN2A/P16) mainly related to lung cancer (36–59), (Supplemental Table 2). For each gene included in the NGS panel, both island-located and interspersed CpGs were tackled (Table 1).

4.2 Samples type and libraries profile – pre-sequencing quality control

To experimentally validate the performance of our NGS methylation panel, we decided to test three different types of biological samples having different grades of fragmentation and size distribution of amplicons, the higher libraries quality profile was observed for 100 ng of converted gDNA, which was therefore adopted for all the above reported experiments.

4.3 Evaluation of analytical performance of OPERA_MET-A panel

Converted gDNA from cell lines, OCT embedded and FFPE tissues were successfully analyzed. Results about metrics per sample was summarized in Table 2. The mean number of reads for samples was 1034910 ± 101511 (ranging from 329889 to 2077231), the mean number of read length was 76 ± 2 (ranging from 57 to 85 bp); the mean number of mapped reads was 769384 ± 74468 (ranging from 244244 to 1397720), the mean percentage of read on target was 55 ± 3% for W strand (ranging from 33% to 79%), and 53 ± 4% for C strand (ranging from 23% to 81%). A difference in % of unmapped reads was observed between tested mixes of commercially bisulfite treated gDNA samples and bisulfite treated gDNA obtained from biological sample biopsies (<12% vs ≥26%). Anyway, a high intra-samples variability was observed among gDNA from all samples type (Supplemental Table 3). The mean number of read per amplicon was 2862 ± 304 for W strand (ranging from 1035 to 5873), and 3081 ± 345 for C strand (ranging from 333 to 6433), in line with the data previously reported by Luo et al. for the Ion AmpliSeq™ Methylation Panel for Cancer Research (15).

4.4 Estimation of optimal samples per chip number

The empirical calculation of cut-off of minimum required mapped reads per target was made according to the following criteria: >20 methylated read for testing CpG site, >40% of tumor cell content and >10% methylation level per site and was also related to an expected conversion efficiency >99%. We can therefore assume a minimum of 500 total mapped reads per
| Target ID | Gene   | Accession Number | Chromosome | Start (hg19) | End (hg19) | CpG location (Island/interspersed) | Gene location |
|----------|--------|------------------|------------|-------------|-----------|-----------------------------------|---------------|
| cg04909257 | PGD    | NM_001304451.2   | chr1       | 10462497    | 10462499  | interspersed                      | intron        |
| CpG_Island | NFE2L2 | NM_001145412.3   | chr2       | 17812873    | 178129847 | Island/exon                       |               |
| cg11532131 | PD-1/CD279 | NM_005018.3  | chr2       | 24279222    | 242792226 | interspersed                      | exon          |
| cg10057601 |        |                  |            | 24279307    | 242793079 | interspersed                      | exon          |
| cg22235901 |        |                  |            | 242793206   | 242793408 | interspersed                      | exon          |
| CpG_Island-1 |        |                  |            | 242794853   | 242795083 | Island/exon                       |               |
| cg25798782 |        |                  |            | 242795281   | 242795283 | interspersed                      | intron        |
| cg01632474 |        |                  |            | 242799311   | 242799461 | interspersed                      | intron        |
| cg21670983 |        |                  |            | 242799459   | 242799525 | interspersed                      | intron        |
| CpG_Island-2 |        |                  |            | 242800972   | 242800974 | interspersed                      | exon          |
| cg25890838 |        |                  |            | 242801045   | 242801047 | interspersed                      | exon          |
| cg02122525 |        |                  |            | 242801251   | 242801253 | interspersed 5' upstream           |               |
| cg14453145 |        |                  |            | 242801895   | 242801897 | interspersed 5' upstream           |               |
| cg17322655 |        |                  |            | 242802126   | 242802128 | interspersed 5' upstream           |               |
| cg20805133 |        |                  |            | 242802191   | 242802193 | interspersed 5' upstream           |               |
| cg19811994 | RASSF1A | NM_170713.3     | chr3       | 50373640    | 50373642  | interspersed                      | intron        |
| CpG_Island-1 |        |                  |            | 50374264    | 50375629  | Island/exon                       |               |
| cg24049629 |        |                  |            | 50376474    | 50376476  | interspersed                      | intron        |
| CpG_Island-2 |        |                  |            | 50377803    | 50378540  | Island/exon/5' upstream           |               |
| cg10505656 | SPARC  | NM_003118.4     | chr5       | 151051213   | 151051215 | interspersed                      | intron        |
| cg23174201 |        |                  |            | 151054255   | 151054257 | interspersed                      | intron        |
| cg27128761 |        |                  |            | 151055649   | 151055651 | interspersed                      | intron        |
| cg26389330 |        |                  |            | 151057859   | 151057861 | interspersed                      | intron        |
| cg14518209 |        |                  |            | 151066267   | 151066269 | interspersed                      | intron        |
| CpG_Island |        |                  |            | 151066456   | 151066695 | Island/exon                       |               |
| cg25913233 |        |                  |            | 151066682   | 151066684 | interspersed 5' upstream           |               |
| cg22116670 |        |                  |            | 151066729   | 151066731 | interspersed 5' upstream           |               |
| cg07539983 |        |                  |            | 151067340   | 151067342 | interspersed 5' upstream           |               |
| cg02731193 | GCLC   | NM_001498.4     | chr6       | 53407185    | 53407187  | interspersed                      | intron        |
| CpG_Island | HOXA9  | NM_152739.4     | chr7       | 27203915    | 27204642  | Island/exon/5' upstream           |               |
| cg15837913 | PD-L1/CD274 | NM_014143.4   | chr9       | 5448898     | 5448891   | interspersed 5' upstream           |               |
| CpG_Island |        |                  |            | 5450409     | 5450629   | Island/exon/5' upstream           |               |
| cg13474877 |        |                  |            | 5450723     | 5450725   | interspersed                      | intron        |
| cg19724470 |        |                  |            | 5450935     | 5450937   | interspersed                      | intron        |
| cg12840719 | CDKN2A | NM_000077.5     | chr9       | 21968232    | 21968234  | interspersed                      | exon          |
| CpG_Island-1 |        |                  |            | 21968358    | 21968728  | Island/exon                       |               |
| CpG_Island-2 |        |                  |            | 21970913    | 21971190  | Island/exon                       |               |
| CpG_Island-3 |        |                  |            | 21974578    | 21975306  | Island/exon                       |               |
| cg19648686 | AKR1C2 | NM_001354.6     | chr10      | 5044991     | 5044993   | interspersed                      | intron        |
| CpG_Island | PTEN   | NM_000314.8     | chr10      | 89621772    | 89621428  | Island/exon*                      | intron*       |
| cg02307823 |        |                  |            | 89675900    | 89675902  | interspersed                      | intron        |
| cg19378330 | ABCC2  | NM_00392.5      | chr11      | 10165098    | 10165098  | interspersed                      | intron        |
| cg05779518 |        |                  |            | 117947554   | 117947556 | interspersed 5' upstream**        |               |
| cg08634928 |        |                  |            | 117947610   | 117947612 | interspersed 5' upstream**        |               |
| cg27300950 |        |                  |            | 117947627   | 117947629 | interspersed 5' upstream**        |               |
| cg25116503 |        |                  |            | 117947656   | 117947658 | interspersed 5' upstream**        |               |
| cg22957898 |        |                  |            | 117947876   | 117947878 | interspersed                      | exon**        |

(Continued)
site to achieve a successful run/site. Based on the experimental sequencing performance of samples, we assume that a minimum of 6 samples/530 chip can be loaded.

4.5 Performance of bisulfite conversion rate and global methylation detection by OPERA_MET-A panel

Bisulfite conversion of gDNA was employed in order to discriminate methylated versus unmethylated cytosines, where unmethylated cytosines were deaminated to uracil nucleotide; converted DNA template generated thymines during PCR amplification. The bisulfite conversion rate was calculated by evaluating the Lambda control DNA. Once bisulfite conversion is performed, theoretically every C residue in the unmethylated Lambda control DNA should be converted to a T.

The methylation_analysis plugin counted the number of C residues that are present in the sequence to determine the percentage of the sequence that is methylated, assuming that after the bisulfite conversion reaction, the only C residues that remain in the sequence were methylated in the original sample. The value is shown in the percent.ME (percent methylation) column for each sample. The optimal bisulfite conversion rate should be >99%, calculated as the difference from 100% of the average of the observed W and C percent.ME values for the unmethylated Lambda DNA in each sample and control (60). To estimate the conversion efficiency of our samples, replicate commercially gDNA samples of average methylation states across all CpGs of approximately >95% and <5%, and different mixtures of the two methylation states (~90%, ~75%, ~50%, ~25% and ~10%) were used to construct libraries using the custom panel. As per standard operative procedure, unmethylated Lambda DNA was spiked into each sample prior to bisulfite conversion and primers exist in the panel to determine conversion efficiency using the sequencing output files. A high conversion efficiency was obtained for all treated control samples, with a mean of 98.5% ± 0.2 of conversion efficiency (ranging from 97.5% to 99.1%), (Supplemental Figure 2).

The methylation_analysis plugin allowed both the alignment and the methylation status calling. As a result, a high concordance between average percent methylation across all amplicons (target_CpG) of OPERA_MET-A panel between observed and expected % of methylation was obtained for each methylation state (p value<0.0001, Adjusted R2 = 0.95), (Figure 2).

TABLE 1 Continued

| Target ID | Gene  | Accession Number | Chromosome | Start (hg19) | End (hg19) | CpG location (Island/interspersed) | Gene location |
|-----------|-------|------------------|------------|-------------|-----------|-----------------------------------|---------------|
| cg03331715 | TXNRD1 | NM_182729.3      | chr12      | 104689086   | 104689088 | interspersed                       | intron        |
| cg10880599 | GPX2  | NM_002083.4      | chr14      | 65408479    | 65408481  | interspersed                       | intron*       |
| cg09643186 |       | 65409451         |            | 65410144    | 65410146  | interspersed                       | exon**        |
| cg26155983 |       | 10597015         | chr19      | 10597017    | 10597017  | interspersed                       | exon          |
| cg15866432 |       | 10600445         |            | 10602280    | 10602287  | interspersed                       | exon          |
| Cpg_1Island|       |                 |            | 10602280    | 10602287  | Island                            | exon          |
| cg20226327 |       | 10602959         |            | 10611024    | 10611044  | interspersed                       | intron        |
| cg24892871 |       | 10612801         |            | 10613179    | 10613181  | interspersed                       | intron        |
| cg26988016 |       | 10613455         |            | 10613455    | 10613457  | interspersed                       | intron        |
| cg15204119 |       | 10613487         |            | 10613487    | 10613489  | interspersed                       | intron        |
| cg096911149|       | 10613491         |            | 10613491    | 10613493  | interspersed                       | intron        |
| cg25500801 |       | 10613854         |            | 10613854    | 10613856  | interspersed                       | intron        |
| cg04328100 |       | 10614021         |            | 10614021    | 10614023  | interspersed                       | exon          |
| cg25801292 |       | 10614271         |            | 10614273    | 10614273  | 5’ upstream                       | exon          |
| cg12059186 |       | 10615096         |            | 10615098    | 10615098  | 5’ upstream                       | exon          |
| cg03754063 |       | 10615198         |            | 10615200    | 10615200  | 5’ upstream                       | exon          |
| cg14848212 |       | 631460           | chr20      | 631460      | 631462    | interspersed                       | intron        |
| Cpg_1Island| HARIB | NR_003245.1      | chr20      | 61733275    | 61734521  | Island                            | intron/exon    |

* ("intron/exon" for KLLN, NM_001126049.2, C strand); **("intron" for SMIM35, NM_00394164.1, C strand); *** ("intron" for FTNB, NM_001202559.1, W strand).
FIGURE 1
Quality control steps of library construction using the OPERA_MET-A panel for targeted next generation bisulfite sequencing. (A) Representative picture of chromatograms showing a high quality library profile of methylated and unmethylated DNA control samples and eight gDNA samples from cell lines, FFPE and OCT embedded tissues. (B) Representative images from Agilent Tape Station 2200 NGS libraries for all bisulfite treated DNA samples.

TABLE 2 Metrics per sample obtained using the OPERA_MET-A panel for targeted bisulfite NGS.

| Metric                                      | Mean value | Minimum observed | Maximum observed |
|---------------------------------------------|------------|------------------|------------------|
| Number of total reads                       | 1034910    | 329889           | 2077231          |
| Number of mapped reads                      | 769384     | 244244           | 1397720          |
| Read length (base pair)                     | 76bp       | 57bp             | 85bp             |
| Percentage of read on target for W strand   | 55%        | 33%              | 79%              |
| Percentage of read on target for C strand   | 53%        | 23%              | 81%              |
| Number of read per amplicon for W strand    | 2862       | 1035             | 5875             |
| Percentage of read on target for C strand   | 3081       | 333              | 6433             |
Methylation data obtained by NGS from FFPE samples (paired tumor/non-neoplastic samples) were used to evaluate the utility of OPERA_MET-A panel to visualize the CpG methylation distribution along different gene regions (Island and interspersed CpGs). As expected, variations of methylation level of CpGs were observed between paired samples from the same patients and among patients for PD-1/CD279, that was shown as representative model for this specific application (Figure 3). Finally, methylation data obtained by NGS from paired FFPE samples to investigate the CpG methylation distribution between W and C strands. A subset of CpGs mapped in KEAP1, PD-1/CD279 and RASSF1A genes (Supplemental Table 1) was chosen as pilot for this analysis. For each targeted CpG located on both W and C strand, the total number of reads that cover the target CpGs, the number of methylated reads, unmethylated reads, and percent methylation for each amplicon were generated for each targeted CpG site on both W and C strand of the genes. Interestingly, we observed a striking change in global CpG methylation across the RASSF1A gene on both W and C strands in all samples (p<0.01, Wilcoxon signed-rank test), whereas no substantial differences in global CpG methylation distribution between W and C strands were observed in CD279/PD-1 and KEAP1 genes (Figure 4 and Supplemental Figures 3, 4).

5 Discussion

With an increasing number of clinically actionable targets, the optimization of NGS technology ensured a high sensitivity,
specificity and time-saving of analysis by easily assessing every nucleotide change in multiple targets gene/or regions at single-base resolution. More recently, the NGS technology is also enhancing the methylome analysis, thus contributing to expand the knowledge and characterization of differentially methylated gDNA regions in many cancer-related human genes (61). Starting from a relatively low quantity of biological materials, the NGS approach offers in this specific field the great advantage to successfully quantify DNA methylation density and its differences at specific CpG sites of both promoter or intragenic regions (62, 63). The great main translational advantage of this feasible high-throughput approach in a more specific cancer context is the opportunity to construct a more robust and integrated algorithm to predict the disease evolution of patients. Moreover, it could support the selection and a better stratification of surgically-resected patients for follow-up and enhance the development of novel therapeutic approaches, not yet based on an epigenetic profile in daily practice (64, 65).

The design and validation of the OPERA_MET-A panel for NGS analysis were based on this assumption. This panel allows to obtain libraries starting from gDNA extracted from different matrices, such as cell lines, OCT embedded frozen
and FFPE tissues. Considering the recently published data on NGS for methylation analysis using the gene-targeted AmpliSeq technology, very promising and comparable results to those already been published were obtained (15) in terms of mean target depth ≥2,500X (W and C strand), average number of mapped reads >750,000/sample and concordance results between expected and observed % of global methylation for all CpGs. More specifically, the OPERA_MET-A panel primer design pipeline include more amplicons (155 amplicons) than those generated by Ion AmpliSeq™ Methylation Panel for Cancer Research (40 amplicons) that covers 18 non-overlapping genes and can be used to analyze gDNA from cell lines, human FFPE and OCT tissues to reach a comparable performance in terms of time consuming, observed/expected methylation rate concordance and average mapped reads. The same "methylation_analysis" plugin was used to successfully perform

![FIGURE 4](image-url)

Strand specific distributions of CpG methylation in samples using OPERA_MET-A panel. The average percent of global CpG methylation at RASSF1A gene (target_CpGs) in paired non-neoplastic and tumor tissues of FFPE samples 830, 881, 889. Global % CpG methylation at W strands is in light blue boxplots, at C strands in red boxplot (Wilcoxon signed-rank test).
alignment and methylation calling for amplicons on both W and C strands.

Moreover, the values of the obtained run metric parameters confirm that our customized panel allows us to obtain targeted and quantitative information on tumor markers with a high resolution starting from a relatively low DNA quantity and quality input (Figure 5).

The OPERA_MET-A panel covers multiple regions of 18 cancer-driver genes whose methylation profile was previously proven to have a translational impact on lung cancer progression but also in other solid tumors. Among these, the panel allows the epigenetic scanning of poorly investigated CpGs located at the intragenic exon 3 island of the KEAP1 gene, that were linked to KRAS mutant status in NSCLC patients (29) and at the SPARC gene promoter, which has been proposed as an interesting prognostic biomarker in NSCLC with useful application in the squamous early-stage group (40). Moreover, CpGs located at the CD274/PD-L1 promoter region and CD279/PD-1 were also included in the panel design, since they have recently emerged as an independent prognostic factor associated with shorter overall survival in triple negative breast, colorectal, prostate, gastric, papillary thyroid, melanoma and head and neck cancer patients (16–27). Many increasing literature evidences suggest that, not only CpG Islands, but also CpG shores methylation correlate with gene expression (66, 67). Furthermore, additional evidence would suggest methylation regulatory regions can extend into exon1/intron 1 of a gene and are outside the promoter/CpG island locus (68–70). In the presented work, we mainly focused on CpG island loci and selected the candidate gene list and loci based on an existing literature evidences and own research field of interest results (Supplemental Table 2). Further planned analyses will evaluate differences in prognostic impact related to CpGs in single gene or among genes of the presented panel in cancer patients, thus allowing an upgrade of OPERA_MET-A panel.

While the main focus of this study was to investigate methylation on multiple cancer-related genes by using a multigene NGS panel starting from low DNA quality and quantity, our analysis also offers new insight into the heterogeneity of CpG methylation patterns among genes and patients. Using the “methylation_analysis” plugin to manage data obtained from the OPERA_MET-A panel, differentially methylated cytosines can be found along the same gene (i.e. PD-1/CD279) and/or in each gene in both double-stranded (i.e., symmetric, as for PD-1 and KEAPI genes) and single-stranded (i.e., asymmetric, as for RASSF1A) contexts. While single stranded DNA methylation can be more frequently detected in non-CG methylation contexts (71), CpG positions are usually expected to be either fully methylated or fully unmethylated in both DNA strands. Therefore, there should not be a “right” or “wrong” strand when choosing to study conventional methylation patterns and data obtained from one strand can be safely assumed to apply also to the second strand. Double stranded DNA methylation primarily occurs in the case of nucleobase symmetry between sense (CG) and antisense (GC) DNA strands. However, methylation in a CG context is not always double-stranded; it can also occur on just a single strand, as described in solid tumors (72) and should be related to genes function and transcription. By consequence, the ability to determine the exact locations and status of CpG methylation in W and C strands separately could provide interesting experimental evidence for innovative clinical applications.

The main limitation of the OPERA_MET-A panel relies on the limited number of genes and samples to be analyzed in a single run to achieve an acceptable coverage to obtain epigenetic information from degraded samples having variable % tumor cell content.
Further improvements are required, such as extending the number of clinically relevant genes whose methylation has a prognostic value in tumors and additional orthogonal evaluation of the methylation density using conventional approaches to study methylation of genes at single CpG level. It should also keep in mind that, when we worked on different matrices, including fixed tissues, an optimized pre-analytical and analytical workflows are demanded in order to obtain an optimal quality and quantity of DNA and decrease the risk of suffering from a critical bisulfite conversion. As also reported in our small subsets of samples from different matrices (commercial gDNA, cell lines, OCT and FFPE tissues) degraded DNA samples (by cross-linking, deamination and fragmentation) could in fact affect the efficiency of NGS analysis, so their pre-analytical and analytical manipulation represents a critical aspect to evaluate in order to improve the homogeneity and efficiency of bisulfite conversion and high throughput NGS sequencing (73, 74).

In conclusion, considering all together, the obtained NGS performance for OPERA_MET-A panel corroborates the utility NGS approach for methylation pattern analysis among groups, validation of whole approaches, identification of gDNA methylation for different regions of single/multiple genes (promoter and other regulatory regions) or CpG islands in multiple samples aimed at the discovery of biomarkers having clinical relevance.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

Ethics statement

The studies involving human participants were reviewed and approved by Local Ethical Committee of IRCCS Casa Sollievo della Sofferenza Hospital, Italy. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

Conceptualization, supervision, writing—original draft preparation, LAM. Methodology and validation FF, FC and MC. Visualization, formal analysis and software SC, TM and MM. Data curation, FF, AS, LAM. Resources, NC and PG. and and EM. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by the Italian Ministry of Health, Ricerca Corrente 2021-22, by the “5 x1000” voluntary contributions to Fondazione IRCCS Casa Sollievo della Sofferenza.

Acknowledgments

The authors thank Dr. Giorgio Pea for his technical and logistic assistance in the OPERA_MET-A panel design; NGS run parameters and Methylation_analysis plugin optimization. The authors also thank Prof. Andreina Guarnieri for the professional English editing of the manuscript, Teresa Balsamo for her technical assistance in performing NGS analysis and Federica Russo for tissue samples handling.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.968804/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Input of converted DNA comparison for libraries construction using the OPERA_MET-A panel. (A) Representative picture of chromatograms showing library profiles starting from 100ng (D1, E1, F1, G1) and 30ng (H1, A2, B2, C2) of input DNA respectively. (B) Representative images from Agilent Tape Station 2200 NGS libraries for all bisulfite treated DNA samples.

SUPPLEMENTARY FIGURE 2
Conversion efficiency across methylated/unmethylated control mixtures using OPERA_MET-A panel. Commercially control DNA samples with
average methylation of approximately >95% and <5%, and different mixtures of methylation states (~90%, ~75%, ~50%, ~25% and ~10%) were evaluated in triplicate. The bisulfite conversion efficiency was evaluated by using the Lambda control DNA, that was added before starting the bisulfite conversion to each single DNA. R1, replicate 1, R2, replicate 2, R3, replicate 3.

SUPPLEMENTARY FIGURE 3
Strand specific distributions of CpG methylation in samples using OPERA_MET-A panel at CD279/PD-1 gene. The average percent of global CpG methylation (target_CpGs) in paired non-neoplastic and tumor tissues of FFPE samples 830, 881, 889. Global % CpG methylation at W strands is in light blue boxplots, at C strands in red boxplot (Wilcoxon signed-rank test).

SUPPLEMENTARY FIGURE 4
Strand specific distributions of CpG methylation in samples using OPERA_MET-A panel at KEAP1 gene. The average percent of global CpG methylation (target_CpGs) in paired non-neoplastic and tumor tissues of FFPE samples 830, 881, 889. Global % CpG methylation at W strands is in light blue boxplots, at C strands in red boxplot (Wilcoxon signed-rank test).

References
1. Jin Z, Liu Y. DNA Methylation in human diseases. Dis Genet (2018) 5(1):1-8. doi: 10.1016/j.digen.2018.01.002.
2. Mazzone R, Zwerge1 C, Artico M, Taurone S, Balli M, Greco A, et al. The emerging role of epigenetics in human autoimmune disorders. Clin Epigenet (2019) 11(1):34. doi: 10.1186/s13148-019-0632-2.
3. Locke WJ, Guanzon D, Ma C, Liew YJ, Duesing KR, Fung YTC, et al. DNA Methylation cancer biomarkers: Translation to the clinic. Front Genet (2019) 10:1150. doi: 10.3389/fgen.2019.01150.
4. Min JL, Hempfl G, Hannon E, Dekkers KF, Castillo M, Garcia S, Prado E, et al. The timeline of epigenetic drug discovery: From reality to dreams. Clin Epigenet (2019) 11(1):174. doi: 10.1186/s13148-019-0776-0.
5. Duruisseaux M, Esteller M. Lung cancer epigenetics: From knowledge to applications. Semin Cancer Biol (2018) 51:116-28. doi: 10.1016/j.semcancer.2017.09.005.
6. Romero-Garcia S, Prado-Garcia H, Carlos- Reyes A. Role of DNA methylation in the resistance to therapy in solid tumors. Front Oncol (2020) 10:1152. doi: 10.3389/fonc.2020.01152.
7. Tepper MJ, Vaz M, Marrone KA, Brahmer JR, Baylin SB. The emerging role of epigenetic therapeutics in immuno-oncology. Nat Rev Clin Oncol (2020) 17(23):75-90. doi: 10.1038/s41571-019-0266-5.
8. Shenker N, Flanagan JM. Intragenic DNA methylation: Implications of this epigenetic mechanism for cancer research. Br J Cancer (2012) 106(2):248-53. doi: 10.1038/bjc.2011.550.
9. Sun X, Tian Y, Wang J, Sun Z, Zhu Y. Genome-wide analysis reveals the association between alternative splicing and DNA methylation across human solid tumors. BMC Med Genomics (2020) 13(1):14. doi: 10.1186/s12920-019-0654-9.
10. Wani K, Aldape KD. Pd techniques in characterizing DNA methylation. Methods Mol Biol (2016) 1392:177-86. doi: 10.1007/978-1-4939-3360-0_16.
11. Sambrook J, Russell DW. Isolation of high-Molecular-Weight DNA from mammalian cells using new CsCl and phenol. CSH Protoc (2006) 2006(1):pdb.prot4036. doi: 10.1101/pdb.prot4036.
12. Fabrizio FP, Sparanoee A, Centra F, Trombetta D, Storfazzi CT, Graziano P, et al. Methylation density pattern of Keap1 gene in lung cancer cell lines detected by Quantitative Methylation Specific PCR and pyrosequencing. Int J Mol Sci (2019) 20(11):2697. doi: 10.3390/ijms20112697.
13. Xu C, Correces VQ. Resolution of the DNA methylation state of single cpg dyads using in silico strand annealing and wgbs data. Nat Protoc (2019) 14(1):202-16. doi: 10.1038/s41596-018-0090-x.
14. Lao Z, Hatch A, Elwing A, Hyland F, Berman D, Patel P, et al. Custom primer design pipeline and analysis workflow for targeted methylation sequencing using NGS Ion AmpliSeq technology. Cancer Res (2020) 80(16):Supplement:-158. doi: 10.1158/0008-5472.AM2020-158.
15. Marwitz S, Scheufele S, Perner S, Beck M, Ammerpohl O, Dietrich H, et al. Epigenetic modifications of the immune-checkpoint genes CTLA4 and PD1CDI in non-small cell lung cancer results in increased expression. Clin Epigenet (2017) 9:51. doi: 10.1186/s13148-017-0354-2.
16. Goltz D, Gevensleben H, Dietrich J, Schneek F, De vos L, Droege F, et al. Pd1 (CD1 Pd-D1) promoter methylation predicts outcome in head and neck squamous cell carcinoma patients. Oncotarget (2017) 8(25):24110-20. doi: 10.18632/oncotarget.17354.
17. Mishra A, Verma M. Epigenetic and genetic regulation of PDCDI gene in cancer immunology. Rev Methods Mol Biol (2018) 1856:247-54. doi: 10.1007/978-1-4939-8751-1_14.
18. Fabbri FP, Mazza T, Castellana S, Sparanoee A, Muscarella L. Gene code CD274/PD-1: From molecular basis toward cancer immunotherapy. Ther Adv Med Oncol (2018) 10:1758835918815598. doi: 10.1177/1758835918815598.
19. Amini M, Hejazee M, Ghorban K, Mohktarzadeh A, Baradaran B. Identification of functional methylated CpG loci in PD-L1 promoter as the novel epigenetic biomarkers for primary gastric cancer. Gene (2021) 772:145376. doi: 10.1016/j.gene.2020.145376.
20. Fabrizio FP, Sparanoee A, Trombetta D, Muscarella L. Epigenetic versus genetic deregulation of the Keap1/Nrf2 axis in solid tumors: Focus on methylation and noncoding RNAs. Oxid Med Cell Longev (2018) 2018:2492063. doi: 10.1155/2018/2492063.
21. Amini M, Hejazee M, Ghorban K, Mohktarzadeh A, Baradaran B. Identification of functional methylated CpG loci in PD-L1 promoter as the novel epigenetic biomarkers for primary gastric cancer. Gene (2021) 772:145376. doi: 10.1016/j.gene.2020.145376.
22. Fabrizio FP, Sparanoee A, Trombetta D, Muscarella L. Epigenetic versus genetic deregulation of the Keap1/Nrf2 axis in solid tumors: Focus on methylation and noncoding RNAs. Oxid Med Cell Longev (2018) 2018:2492063. doi: 10.1155/2018/2492063.
23. Amini M, Hejazee M, Ghorban K, Mohktarzadeh A, Baradaran B. Identification of functional methylated CpG loci in PD-L1 promoter as the novel epigenetic biomarkers for primary gastric cancer. Gene (2021) 772:145376. doi: 10.1016/j.gene.2020.145376.
24. Fabrizio FP, Mazza T, Castellana S, Sparanoee A, Muscarella L. Epigenetic versus genetic deregulation of the Keap1/Nrf2 axis in solid tumors: Focus on methylation and noncoding RNAs. Oxid Med Cell Longev (2018) 2018:2492063. doi: 10.1155/2018/2492063.
Promoter methylation of PTEN is a significant prognostic factor for breast cancer. J Genet Epidemiol (2017) 19(1):166. doi: 10.1155/2017/14698167

54. Ramadan A, Hashim M, Abouzid A, Swellam M. Clinical impact of PTEN methylation status for a prognostic marker for breast cancer. J Genet Epidemiol (2021) 19(1):166. doi: 10.1155/2017/01281-001-19649-4

55. Vaisiire T, Hung RJ, Zaridze D, Moukheria A, Cuenin C, Fasolo V, et al. Quantitative analysis of DNA methylation profiles in lung cancer identifies aberrant DNA methylation of specific genes and its association with gender and cancer risk factors. Cancer Res 2009; 60:2433-52. doi: 10.1158/0008-5472.CAN-02-0061

56. Shima K, Nosho K, Baba Y, Cantor M, Meyerhardt JA, Giovannucci EL, et al. Prognostic significance of C6CDKN2A (p16) promoter methylation and loss of expression in 902 colorectal cancers: Cohort study and literature review. Int J Cancer 2011; 128(5):1380-94. doi: 10.1002/ijc.25522

57. Carvalho RH, Hou J, Haberle V, Aerts J, Grovold F, Lenhard B, et al. Genomewide DNA methylation analysis identifies novel methylated genes in non-Small-Cell lung carcinomas. J Thorac Oncol (2013) 8(5):562-73. doi: 10.1097/JTO.0b013e318263ed2

58. Li Y, Melnikov AA, Levenson V, Guerra E, Simonee P, Alberta S, et al. A seven–gene CpG island methylation panel predicts breast cancer progression. BMC Cancer (2015) 1:1417. doi:10.1186/s12885-015-1412-9

59. Zhou Y, Wang XB, Qiu XP, Zhang S, Wang C, Zheng F. CDKN2A promoter methylation and hepatocellular carcinoma risk: A meta–analysis. Clin Res Hepatol Gastroenterol (2018) 42(6):529-41. doi:10.1016/j.clinre.2017.07.003

60. User Bulletin TS. Bisulfite methylation library production and analysis using the ion ampliseq™ library kit plus. (2019) 5:1–30. Publication Number MAN017662.

61. Soto J, Rodriguez–Antonio C, Vallespin E, de Castro Carpeno J, Barnez de Caeceres J. The impact of next–generation sequencing on the DNA methylation–based translational cancer research. Transl Regul (2016) 169:1-8. doi:10.1016/j.tran.2015.11.003

62. van den Oord EJ, Buxjarr J, Rudolf G, Nerulla S, McClay JL, Xie Y, et al. Estimation of CpG coverage in whole methylene next–generation sequencing studies. BMC Bioinf (2013) 14:50. doi:10.1186/1471-2105-14-50

63. Barros–Silva D, Marques CJ, Henrique R, Jeronimo C. Profiling DNA methylation based on next–generation sequencing approaches: New insights and clinical applications. Genes (Basel) (2018) 9:39. doi:10.3390/genes9030029

64. Chen X, Gole J, Gore A, He Q, Lu M, Min J, et al. Non–invasive early detection of cancer four years before conventional diagnosis using a blood test. Nat Commun (2020) 11:3475. doi:10.1038/s41467-020-17316-z

65. Klein EA, Richards D, Cohn A, Tummala M, Lapham R, Cosgrove D, et al. Clinical validation of a targeted methylation–based multi–cancer early detection test using an independent validation set. Ann Oncol (2021) 13(9):1167–77. doi:10.1093/annonc/mdab130

66. Irizarry RA, Ladd–Acosta C, Wen B, Wu Z, Montano C, Omyango P, et al. The human colon cancer methylene shows similar hypo– and hypermethylation at conserved tissue–specific CpG island shores. Nat Genet (2009) 41(2):178–86. doi:10.1038/ng.298

67. Doi A, Park IH, Wen B, Burakami P, Ayye MI, Irizarry R, et al. Differential methylation of tissue– and cancer–specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. Nat Genet (2009) 41(12):1350–5. doi:10.1038/ng.471

68. Portela A, Esteller M. Epigenetic modifications and human disease. Nat Biotechnol (2010) 28(10):1057–68. doi:10.1038/nbt.1685

69. Jones PA. Functions of DNA methylation: Islands, start sites, gene bodies and beyond. Nat Rev (2012) 13:478–92. doi:10.1038/nrg3230

70. Yang X, Han H, De Carvalho DD, Lay FD, Jones PA, Liang G. Gene body methylation can alter gene expression and is a therapeutic target in cancer. Cancer Cell (2014) 26(4):577–96. doi:10.1016/j.ccell.2014.07.028

71. Lister R, Pelizzola M, Downen RH, Hawkins RD, Hon G, Toni F, Filippini J, et al. Human DNA methylomes at base resolution show widespread epigenetic differences. Nature (2009) 462(7271):315–22. doi:10.1038/nature08514

72. Jain S, Chang TT, Hamilton JF, Lin SY, Lin YJ, Evans AA, et al. Methylation of the cpg sites only on the sense strand of the APC gene is specific for hepatocellular carcinoma. Plos One (2011) 6(11):e26799. doi:10.1371/journal.pone.0026799

73. Generex DP, Johnson WC, Burden AF, Noger R, Laird CD. Errors in the bisulfite conversion of DNA: modulating inappropriate and failed–conversion frequencies. Nucleic Acids Res (2008) 36:150. doi:10.1093/nar/gkn691

74. Jung M, Uhl B, Kristiansen G, Dietrich D. Bisulfite conversion of DNA from tissues, cell lines, buccal coat, FFPE tissues, microdissected cells, swabs, sputum, aspirates, lavages, effusions, plasma, serum, and urine. Methods Mol Biol (2017) 1589:139–59. doi:10.1007/978-1-536-23524-6_35