Streamlined Preparation of NGS-Plex and NGS-MethylPlex Libraries for DNA Sequence and Methylation Analysis Using Next-Generation Sequencing Platforms

Makarov, V., Kamberov, E., Kurihara, T., Langmore, J., M'Mwirichia, J., Rubicon Genomics,
4355 Varsity Dr, Ste E, Ann Arbor, MI 48108.
e-mail:vlmakarov@yahoo.com; phone: 734 649 8101; web:www.rubicongenomics.com

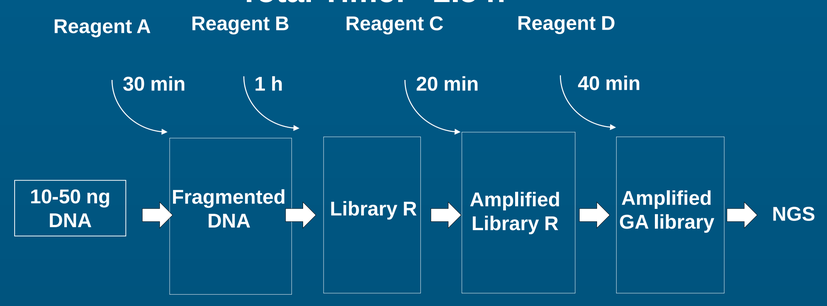
Next-generation DNA sequencers have revolutionized the genetic research by producing millions of DNA sequence reads in a single run. This has enabled researchers to design and perform before-unimaginable genome-wide , ultra-deep sequencing projects such as polymorphism and mutation discovery, highly sensitive gene expression analysis, non-coding RNA profiling, chromatin immune-precipitation, epigenome mapping, etc.

For all mentioned NGS platforms the process of genomic DNA sequencing starts from preparation of a high quality DNA library and this step is critical for successful sequence analysis. Current protocols usually require microgram quantities of DNA and involve multiple steps such as mechanical fragmentation, preparation of DNA ends, ligation of adaptors, gel size fractionation and library amplification plus numerous purification steps which all together consume significant time frequently exceeding 8 h. Preparation of libraries for NGS DNA methylation analysis is even more complicated and requires at least one additional day for isolation of methylated DNA by immuno-precipitation (MeDIP) or magnetic bead binding (MBD) methods.

Using its expertise in whole genome amplification (GenomePlex and EnzyPlex) and whole methylome amplification (MethylPlex) Rubicon has recently developed two streamlined, automatable, protocols for preparation of the NGS-Plex and the NGS-MethylPlex libraries for DNA sequence and methylation analysis for the Illumina Genome AnalyzerTM (GA). Both protocols are substantially simpler, more sensitive, and faster than other sample preparation methods (only 20 - 50 ng of input DNA, and 2.5 or 5 h total preparation time for NGS-Plex or NGS-MethylPlex libraries, respectively).

Shown is the process of NGS-Plex and NGS-MethylPlex library preparation as well as results from bacterial genome sequencing and genome-wide methylation profiling of human prostate primary benign and cancer cell lines (our acknowledgment to Dr. Arul Chinnaiyan and his team from University of Michigan)

Flowchart for Rubicon GA Sample Preparation: Current Protocol Total Time: ~2.5 h



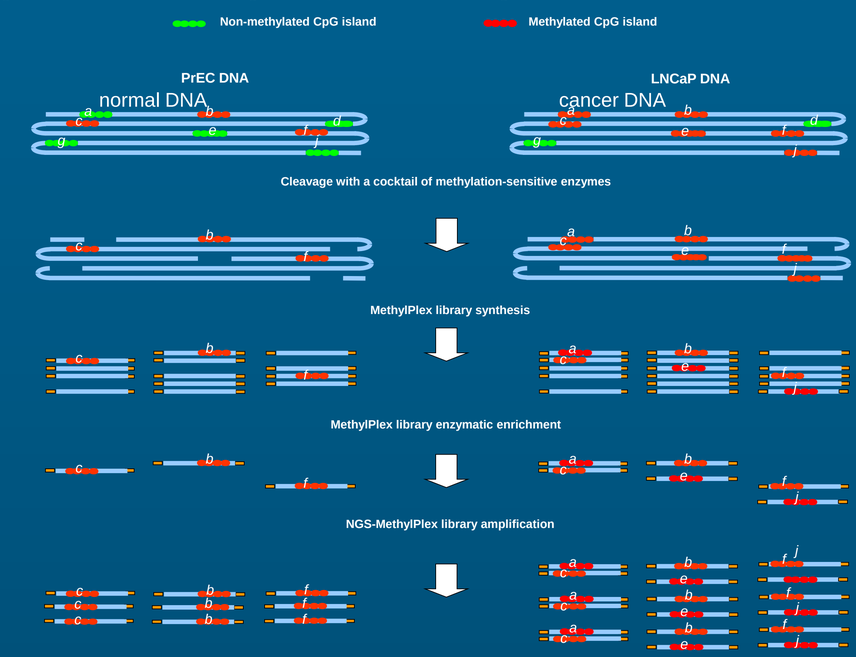
The outlined sample preparation protocol relies on the efficiency of DNA size "fractionation" by PCR, allowing to "filter out" Library R DNA fragments that are smaller than 100 -200 bp. The upper DNA size limit is controlled by DNA fragmentation process.

We employed MethylPlex-NGS Sequencing, a new methodology that utilizes only 50 nanograms of input genomic DNA, to map global DNA methylation patterns in LNCaP prostate cancer cells and PrEC benign prostate epithelial cells. Methylplex libraries were constructed by digesting input DNA with methylation-sensitive restriction enzymes, followed by attachment of universal sequences and subsequent PCR amplification. A second round of enzymatic treatment depleted non-GC rich sequences, followed by an additional amplification to ensure enrichment of highly methylated DNA fragments.

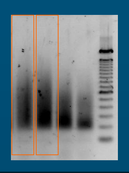
CG dinucleotides were enriched by the Methylplex procedure up to three-fold in mapped reads from M-NGS as compared to previously obtained pan-histone Chip-Seq data. A Hidden Markov Model (HMM)-based algorithm was used to detect enriched regions from mapped reads obtained in each sequencing run. While both cell lines harbored a total of ~55,000 methylated regions each, LNCaP cells exhibited a seven-fold enrichment for promoter CpG island methylation compared to PrEC cells.

Several regions were validated by multiple independent approaches including bisulfite sequencing. Detailed promoter analysis revealed diverse methylation patterns around transcription start sites, including direct methylation of CpG islands, methylation of regions flanking CpG islands, and methylation of sites devoid of CpG islands. Methylated promoters correlated with gene repression, and we observed an enrichment of novel regions methylated in LNCaP cells that were also methylated in prostate cancer tissues. Cancer-specific methylation of selected genes co-occurred with H3K4 trimethylation signatures to mediate transcript isoform switching, which may contribute to cancer progression.

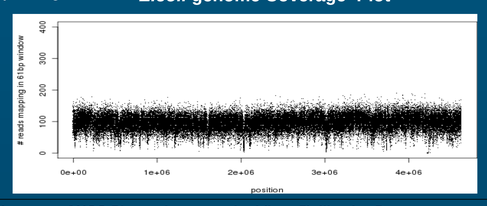
MethylPlex Selectively Amplifies Methylated Promoter Sequences



E. Coli NGS-Plex DNA Used for Sequencing without Size Fractionation



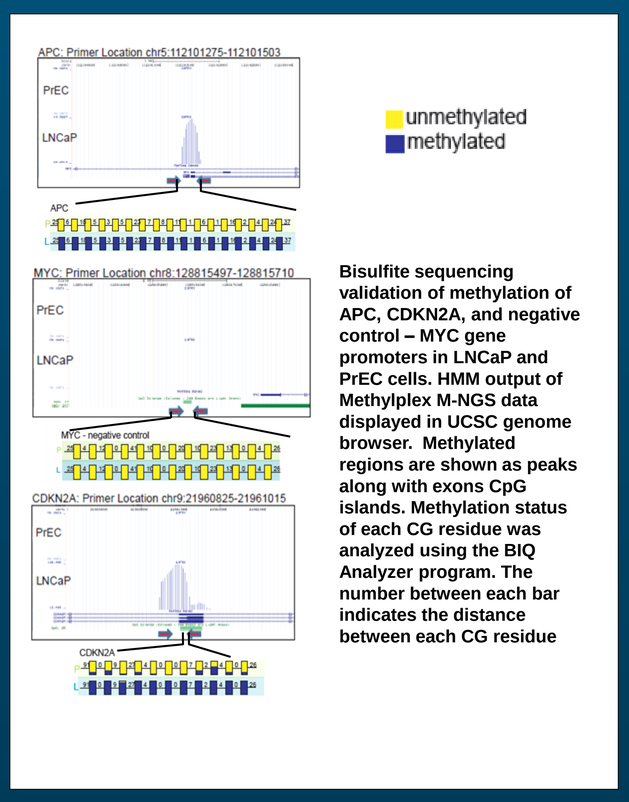
E.coli genome Coverage Plot



Summary of 7 Individual sequencing runs of NGS-Plex and one standard (lane 5) Illumina Prep

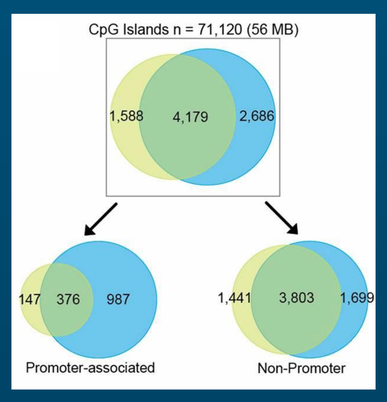
Lane	Length	Lane Yield (kbases)	per tile Clusters (raw)	per tile Clusters (PF)	1st Cycle Int (PF)	% intensity after 20 cycles (PF)	% PF Clusters	% Align (PF)	% Error Rate (PF)	% Phasing	% Prepha sing	Total Aligned Reads per Lane
1	35	263307	186246	146281	108	71.20	78.5	86.0	0.88	0.37	0.53	12580166.000
2	35	285769	210500	158760	99	80.73	75.4	86.0	0.90	0.37	0.53	13653360.000
3	35	283423	200226	157457	115	79.39	78.6	86.1	0.89	0.37	0.53	13557047.700
4	35	264825	194751	147125	82	69.67	75.5	85.1	0.95	0.37	0.53	12520337.500
Std 5	35	181842	110486	101023	138	84.72	91.4	97.9	0.19	0.37	0.53	9890151.700
6	35	286278	211328	159043	155	78.80	75.2	85.6	0.89	0.37	0.53	13614080.800
7	35	286323	205529	159068	144	79.11	77.3	85.7	0.88	0.37	0.53	13632127.600
8	35	259335	202220	144075	104	71.37	71.2	84.7	0.93	0.37	0.53	12203152.500

Validation of differentially methylated regions predicted by MethylPlex-NGS in LNCaP and PrEC cells



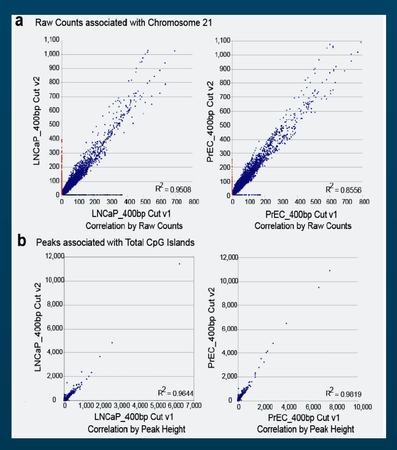
Bisulfite sequencing validation of methylation of APC, CDKN2A, and negative control – MYC gene promoters in LNCaP and PrEC cells. HMM output of Methylplex M-NGS data displayed in UCSC genome browser. Methylated regions are shown as peaks along with exons CpG islands. Methylation status of each CG residue was analyzed using the BIQ Analyzer program. The number between each bar indicates the distance between each CG residue

Venn diagram representing overlap between the regions methylated in LNCaP and PrEC cells



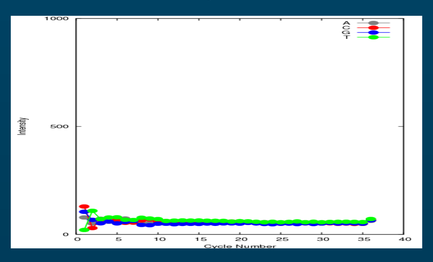
Overlap between methylated regions in LNCaP and PrEC cells is ~ 70%. While there is a 7 fold increase in the methylation of promoter associated CpG islands in LNCaP (blue) compared to PrEC (green) cells, this difference is not seen in non-promoter associated CpG islands.

Regression analysis of MethylPlex-NGS mapped reads and HMM output

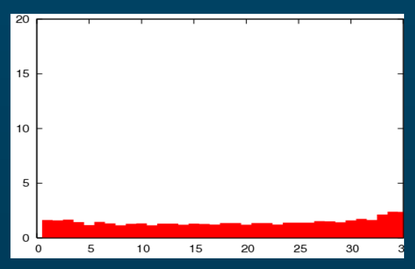


There is high correlation between sequencing runs (a) Reads that mapped to chromosome 21 in LNCaP and PrEC runs are compared using window size of 25bp. In LNCaP samples, a total of 33,627 reads were present at 25 bp windows with R2 value of 0.9508, and in PrEC, 37,406 reads with R2 value of 0.8556 was observed. (b) Linear regression analysis of all DNA methylation that occurred on CGIs showed high correlation (R2 value = 0.9398 and 0.9819, n=5,734 and 4,966, respectively).

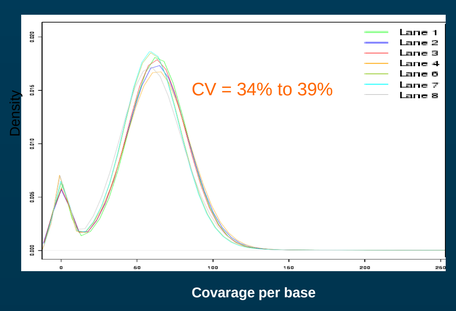
Base Intensity Vs. Position Plots



% Error Rate Plots

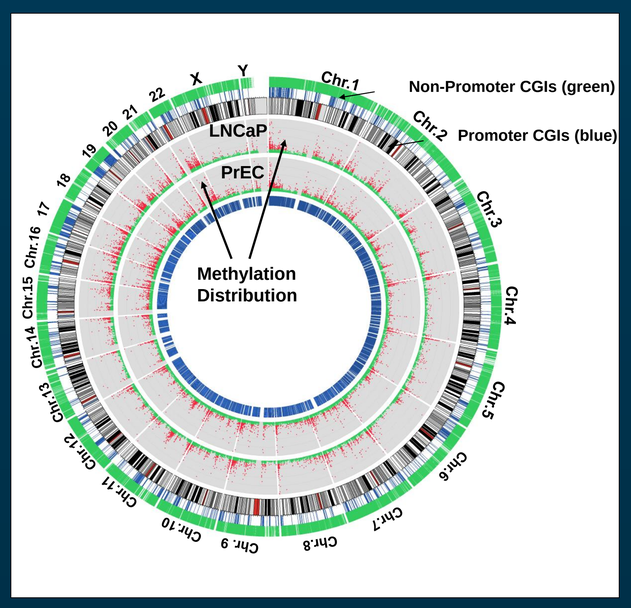


50tiles *E. Coli* covarage density plots (mean covarage normalized to 0 x)



Protocol characteristics	Rubicon NGS-Plex
Sample Preparation Time	2.5 h
Hands-on Time	1 h
Amount of DNA	50 ng or less
Number of Steps	4
Automation	Yes
High Throughput	Yes
Fragmentation	Enzymatic, very small bias
Average genomic insert size	~ 100 – 150 bp
Gel fractionation	No
Multiplexing capabilities	Unlimited

Circos Diagram Summarizing Methylation Detected by MethylPlex-NGS



Conclusions

- New NGS sample preparation protocol developed at Rubicon requires 2.5 hours total time.
- Enzymatic DNA fragmentation has a promising low DNA sequence bias
- The protocol requires <50 ng DNA that is almost 100x less than the current protocol

Conclusions

- While many NGS methodologies to date require several micrograms of input genomic DNA, the very low input requirement in this protocol facilitates the use of finite samples of limited availability.
- Our technology opens up the possibility of carrying out differential global DNA methylation analyses with similar ease as is currently carried out with gene expression technologies.