Differential Methylation Analysis of H460 Parent and H460 Knock-in cell lines on 2.1M Nimblegen Array

Nitesh Turaga

February 5, 2014
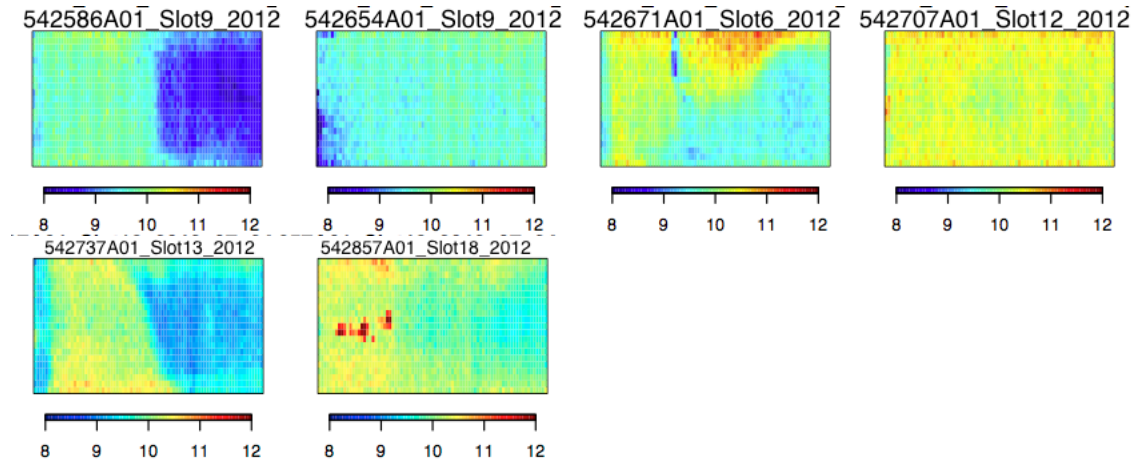
To perform differential methylation analysis of H460 knock-in and H460 parent cell lines.The data was generated by Nimblegen 2.1M microarrays. Raw data are in the form of images (TIF files). Data set included total of 6 samples, 3 for each cell line (H460 knock-in and H460 parent). Array images were processed with DEVA-v1.2 (Nimblegen software for automated feature extraction and data analysis). The TIF files were converted to **XYS** files for analysis. These files report the, **X** - coordinate of the feature on the image, **Y**-coordinate of the feature on the image, and the **S**ignal - the flourescence intensity of the pixels that make the feature. The TIF files were also processed with DEVA using the DNA methylation work flow to identify peaks for each sample.

Preliminary Assessments

General QC-analyses brought to light that the signal intensities and local enrichment at methylated sites were not large. The data quality was assessed by looking at the Enriched channel in the MeDIP array, where we expect every probe to have a signal. Since, the enriched channel has methylated DNA, a successful hybridization would indicate a signal. The array signal is calculated as the average percentile rank of the signal probes among the background probes. The score ranges between 0 to 100, where 100 indicates the ideal scenario or perfect hybridization. This quality score is calculated before any kind of normalization is done on the arrays.

Table 1: Array quality scores

Sample ID	Status	Quality Score
542586A01_Slot9_2012-07-24_H460	H460_parent	59.6
542737A01_Slot13_2012-07-24_H460	H460_parent	64.4
542857A01_Slot18_2012-07-24_H460	H460_parent	69.8
542671A01_Slot6_2012-07-24_H460	H460_knockin	71.2
542654A01_Slot9_2012-07-25_H460	H460_knockin	74.9
542707A01_Slot12_2012-07-25_H460	H460_knockin	76.5

542586A01_Slot9_2012   542654A01_Slot9_2012   542671A01_Slot6_2012   542707A01_Slot12_2012

8   9   10   11   12   8   9   10   11   12   8   9   10   11   12   8   9   10   11   12

542737A01_Slot13_2012   542857A01_Slot18_2012

8   9   10   11   12   8   9   10   11   12

As we can in Table 1, the quality score of the H460 Parent arrays are pretty low.

The CHARM Algorithm

The basic measurement used to quantify methylation is the log-ratio of the intensities observed in the treated and control channels. To detect methylated regions in CHARM, the M-values were normalized and processed using genome-weighted smoothing. The normalization method uses genome sequence information and knowledge of select pseudo-housekeeping probes for which one can assume $M = 0$. Loess is applied to the pseudo-housekeeping genes, to correct M-values for all probes. To obtain a smoothed M-value at any given genomic location, average all the M-values that were within a prespecified distance from the location in question. The interval providing the values that are averaged is referred to as the smoothing window and its length is referred to as the window size. (5)

After estimating the DNA methylation in terms of percentage methylation, we use the regression based DMR-finding approach after correcting for batch effects. No DMRs were found using this method.

The Nimblegen Algorithm

Peaks of enrichment which coincide with methylated regions were found. Peaks near known transcription start sites (TSS) were identified and mapped to overlapping features upto 5000 bp upstream and 1500 downstream in relation to the TSS. In each cell line, the genes which were intersecting among all the samples were identified. To make the two sets independent, all the genes which were identified as common between both cell lines were removed, hence each gene set unique and exclusive for a specific cell line. Both sets of unique and exclusive genes, were ordered by distance from the Transcription start site. 3877 genes were identified in H460 knock-in cell line and 2639 in H460 parent. (Table 2)

Table 2: Number of features in each cell line

Cell line	Feature Track	Number of features
H460 Knock in	transcription start site	3877
H460 Parent	transcription start site	2639

The distribution of the distance from the TSS is also shown for both cell lines. This region encompasses the greater promoter region.

Table 3: Summary of Feature distance from data point

Cell line	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
H460 Knock in	-5000	-3163	-1503	-1599	0	1500
H460 Parent	-4991	-3500	-1900	-1756	0	1499

Table 4: Distribution of Shortest distance from feature to data point
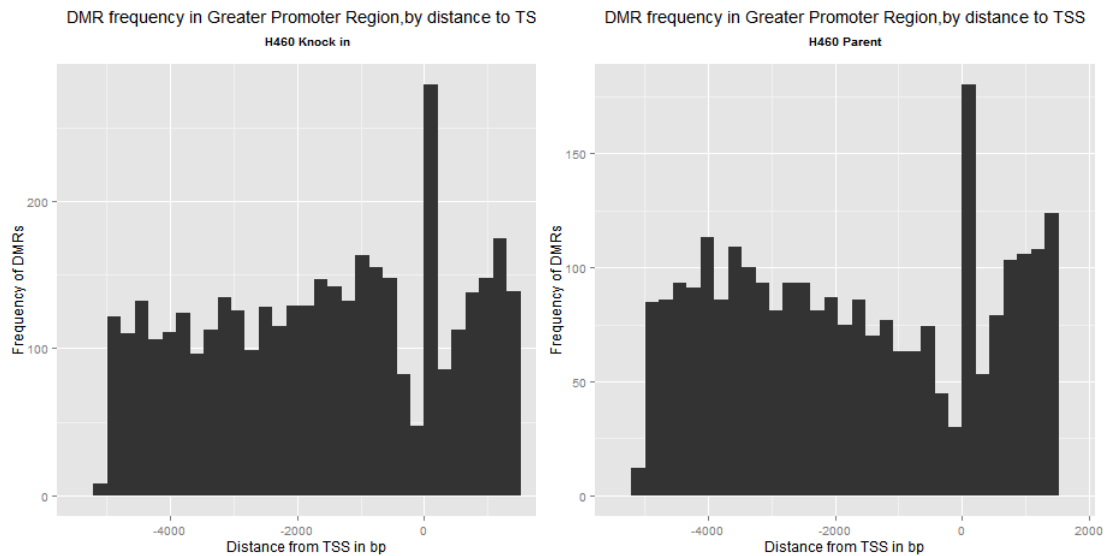
(a) H460 knock-in distance from TSS   **(b)** H460 parent distance from TSS

Pathway Comparisons

Both the list of H460 Knock-in genes and H460 parent genes were compared with a list of selected pathways.

1. Cell Cycle
2. Complete Homeobox (HOX) Genes
3. Complete Human Inflammatory Response and Autoimmunity
4. Complete Human Tumor Suppressor Genes
5. Complete Stem Cell Transcription Factors
6. Complete Stress and Toxicity
7. Cytokine Production
8. DNA Repair
9. Human Epithelial to Mesenchymal Transition (EMT)
10. Human Notch Signaling Pathway
11. Human T-Cell B-Cell Activation Methylation
12. Human T Helper Cell Differentiation
13. Human Tumor Suppressor Genes
14. Inflammatory Response and Autoimmunity
15. Polycomb and Trithorax Complexes
16. Stem Cell Transcription Factors
17. TGF f BMP Signaling Pathway
18. Toll-Like Receptor Signaling Pathway
19. WNT Signaling Pathway

Genes which belong to these pathways matched with the H460 knock-in genes and the H460 parent genes are shown graphically to represent the frequency. The H460 knock-in genes have 189 genes in common with the pathways (partially listed in Table 5)and H460 parent genes have 71 genes in pathways(Table 6). .The following tables shows you the pathways and the genes along with gene description. Supplemental tables have links to the full results.

Table 5: H460 Knock-in gene with corresponding pathways

name	description	pathway
BRCA1	breast cancer 1, early onset	Cell_Cycle
CCNF	cyclin F	Cell_Cycle
RAD9A	RAD9 homolog A (S. pombe)	Cell_Cycle
TP53	tumor protein p53	Cell_Cycle
ALX1	ALX homeobox 1	Complete_Homeobox_(HOX)_Genes
ALX4	ALX homeobox 4	Complete_Homeobox_(HOX)_Genes
CDX2	caudal type homeobox 2	Complete_Homeobox_(HOX)_Genes
DLX1	distal-less homeobox 1	Complete_Homeobox_(HOX)_Genes
DLX5	distal-less homeobox 5	Complete_Homeobox_(HOX)_Genes
DLX6	distal-less homeobox 6	Complete_Homeobox_(HOX)_Genes
EMX1	empty spiracles homeobox 1	Complete_Homeobox_(HOX)_Genes
EN1	engrailed homeobox 1	Complete_Homeobox_(HOX)_Genes
HOXA11	homeobox A11	Complete_Homeobox_(HOX)_Genes
HOXA13	homeobox A13	Complete_Homeobox_(HOX)_Genes
HOXA2	homeobox A2	Complete_Homeobox_(HOX)_Genes
HOXA4	homeobox A4	Complete_Homeobox_(HOX)_Genes
HOXA5	homeobox A5	Complete_Homeobox_(HOX)_Genes
HOXA7	homeobox A7	Complete_Homeobox_(HOX)_Genes
HOXA9	homeobox A9	Complete_Homeobox_(HOX)_Genes
HOXB2	homeobox B2	Complete_Homeobox_(HOX)_Genes
HOXB3	homeobox B3	Complete_Homeobox_(HOX)_Genes
HOXB4	homeobox B4	Complete_Homeobox_(HOX)_Genes
HOXB6	homeobox B6	Complete_Homeobox_(HOX)_Genes
HOXB7	homeobox B7	Complete_Homeobox_(HOX)_Genes
HOXB8	homeobox B8	Complete_Homeobox_(HOX)_Genes
HOXC10	homeobox C10	Complete_Homeobox_(HOX)_Genes
HOXC11	homeobox C11	Complete_Homeobox_(HOX)_Genes
HOXC12	homeobox C12	Complete_Homeobox_(HOX)_Genes
HOXC13	homeobox C13	Complete_Homeobox_(HOX)_Genes
HOXD1	homeobox D1	Complete_Homeobox_(HOX)_Genes
HOXD10	homeobox D10	Complete_Homeobox_(HOX)_Genes
HOXD11	homeobox D11	Complete_Homeobox_(HOX)_Genes
HOXD12	homeobox D12	Complete_Homeobox_(HOX)_Genes
HOXD3	homeobox D3	Complete_Homeobox_(HOX)_Genes
HOXD9	homeobox D9	Complete_Homeobox_(HOX)_Genes
ISL1	ISL LIM homeobox 1	Complete_Homeobox_(HOX)_Genes
LBX1	ladybird homeobox 1	Complete_Homeobox_(HOX)_Genes
LBX2	ladybird homeobox 2	Complete_Homeobox_(HOX)_Genes
MIXL1	Mix1 homeobox-like 1 (Xenopus laevis)	Complete_Homeobox_(HOX)_Genes
MSX2	msh homeobox 2	Complete_Homeobox_(HOX)_Genes
PHOX2A	paired-like homeobox 2a	Complete_Homeobox_(HOX)_Genes
PHOX2B	paired-like homeobox 2b	Complete_Homeobox_(HOX)_Genes
PITX3	paired-like homeodomain 3	Complete_Homeobox_(HOX)_Genes
BCL6	B-cell CLL/lymphoma 6	Complete_Human_Inflammatory_Response_&_Autoimmunity
CD276	CD276 molecule	Complete_Human_Inflammatory_Response_&_Autoimmunity
CXCL5	chemokine (C-X-C motif) ligand 5	Complete_Human_Inflammatory_Response_&_Autoimmunity
IL10RB	interleukin 10 receptor, beta	Complete_Human_Inflammatory_Response_&_Autoimmunity
IL17RA	interleukin 17 receptor A	Complete_Human_Inflammatory_Response_&_Autoimmunity
IL4R	interleukin 4 receptor	Complete_Human_Inflammatory_Response_&_Autoimmunity
IL6ST	interleukin 6 signal transducer (gp130, oncostatin M receptor)	Complete_Human_Inflammatory_Response_&_Autoimmunity
INHBA	inhibin, beta A	Complete_Human_Inflammatory_Response_&_Autoimmunity
IRF1	interferon regulatory factor 1	Complete_Human_Inflammatory_Response_&_Autoimmunity
LTB4R	leukotriene B4 receptor	Complete_Human_Inflammatory_Response_&_Autoimmunity
NOD1	nucleotide-binding oligomerization domain containing 1	Complete_Human_Inflammatory_Response_&_Autoimmunity
PAX1	paired box 1	Complete_Human_Inflammatory_Response_&_Autoimmunity
RELA	v-rel reticuloendotheliosis viral oncogene homolog A (avian)	Complete_Human_Inflammatory_Response_&_Autoimmunity
S1PR3	sphingosine-1-phosphate receptor 3	Complete_Human_Inflammatory_Response_&_Autoimmunity
SOCS1	suppressor of cytokine signaling 1	Complete_Human_Inflammatory_Response_&_Autoimmunity
TH1L	TH1-like (Drosophila)	Complete_Human_Inflammatory_Response_&_Autoimmunity
TIRAP	toll-interleukin 1 receptor (TIR) domain containing adaptor protein	Complete_Human_Inflammatory_Response_&_Autoimmunity

Table 6: H460 parent gene with corresponding pathways

name	description	pathway
BRCA2	breast cancer 2 early onset	Cell_Cycle
CCNE1	cyclin E1	Cell_Cycle
MRE11A	MRE11 meiotic recombination 11 homolog A (S. cerevisiae)	Cell_Cycle
HOPX	HOP homeobox	Complete_Homeobox_(HOX)_Genes
MKX	mohawk homeobox	Complete_Homeobox_(HOX)_Genes
SIX6	SIX homeobox 6	Complete_Homeobox_(HOX)_Genes
ATF2	activating transcription factor 2	Complete_Human_Inflammatory_Response_and_Autoimmunity
CD274	CD274 molecule	Complete_Human_Inflammatory_Response_and_Autoimmunity
CXCL1	chemokine (C-X-C motif) ligand 1 (melanoma growth stimulating activity alpha)	Complete_Human_Inflammatory_Response_and_Autoimmunity
IL12A	interleukin 12A (natural killer cell stimulatory factor 1 cytotoxic lymphocyte maturation factor 1 p35)	Complete_Human_Inflammatory_Response_and_Autoimmunity
IL12B	interleukin 12B (natural killer cell stimulatory factor 2 cytotoxic lymphocyte maturation factor 2 p40)	Complete_Human_Inflammatory_Response_and_Autoimmunity
IL15	interleukin 15	Complete_Human_Inflammatory_Response_and_Autoimmunity
NFKB1	nuclear factor of kappa light polypeptide gene enhancer in B-cells 1	Complete_Human_Inflammatory_Response_and_Autoimmunity
RIPK2	receptor-interacting serine-threonine kinase 2	Complete_Human_Inflammatory_Response_and_Autoimmunity
TLR2	toll-like receptor 2	Complete_Human_Inflammatory_Response_and_Autoimmunity
BRCA2	breast cancer 2 early onset	Complete_Human_Tumor_Suppressor_Genes
CDKN2A	cyclin-dependent kinase inhibitor 2A (melanoma p16 inhibits CDK4)	Complete_Human_Tumor_Suppressor_Genes
ING1	inhibitor of growth family member 1	Complete_Human_Tumor_Suppressor_Genes
MDM2	Mdm2 p53 binding protein homolog (mouse)	Complete_Human_Tumor_Suppressor_Genes
NFKB1	nuclear factor of kappa light polypeptide gene enhancer in B-cells 1	Complete_Human_Tumor_Suppressor_Genes
AR	androgen receptor	Complete_Stem_Cell_Transcription_Factors
DACH1	dachshund homolog 1 (Drosophila)	Complete_Stem_Cell_Transcription_Factors
NFYA	nuclear transcription factor Y alpha	Complete_Stem_Cell_Transcription_Factors
PCNA	proliferating cell nuclear antigen	Complete_Stem_Cell_Transcription_Factors
RUNX2	runt-related transcription factor 2	Complete_Stem_Cell_Transcription_Factors
SMAD2	SMAD family member 2	Complete_Stem_Cell_Transcription_Factors
SOX2	SRY (sex determining region Y)-box 2	Complete_Stem_Cell_Transcription_Factors
WRN	Werner syndrome RecQ helicase-like	Complete_Stem_Cell_Transcription_Factors
ZFPM2	zinc finger protein multitype 2	Complete_Stem_Cell_Transcription_Factors
DNAJB9	DnaJ (Hsp40) homolog subfamily B member 9	Complete_Stress_and_Toxicity
MDM2	Mdm2 p53 binding protein homolog (mouse)	Complete_Stress_and_Toxicity
MRE11A	MRE11 meiotic recombination 11 homolog A (S. cerevisiae)	Complete_Stress_and_Toxicity
NFKB1	nuclear factor of kappa light polypeptide gene enhancer in B-cells 1	Complete_Stress_and_Toxicity
PCNA	proliferating cell nuclear antigen	Complete_Stress_and_Toxicity
RAD17	RAD17 homolog (S. pombe)	Complete_Stress_and_Toxicity
IL12A	interleukin 12A (natural killer cell stimulatory factor 1 cytotoxic lymphocyte maturation factor 1 p35)	Cytokine_Production
TLR2	toll-like receptor 2	Cytokine_Production
BRCA2	breast cancer 2 early onset	DNA_Repair
MLH3	mutL homolog 3 (E. coli)	DNA_Repair
MRE11A	MRE11 meiotic recombination 11 homolog A (S. cerevisiae)	DNA_Repair
CTNNAL1	catenin (cadherin-associated protein) alpha-like 1	Human_Epithelial_to_Mesenchymal_Transition_(EMT)
DSP	desmoplakin	Human_Epithelial_to_Mesenchymal_Transition_(EMT)
TGIF1	TGFB-induced factor homeobox 1	Human_Epithelial_to_Mesenchymal_Transition_(EMT)
YES1	v-yes-1 Yamaguchi sarcoma viral oncogene homolog 1	Human_Epithelial_to_Mesenchymal_Transition_(EMT)
HDAC2	histone deacetylase 2	Human_Notch_Signaling_Pathway
CD274	CD274 molecule	Human_T-Cell_B-Cell_Activation_Methylation
IL12A	interleukin 12A (natural killer cell stimulatory factor 1 cytotoxic lymphocyte maturation factor 1 p35)	Human_T-Cell_B-Cell_Activation_Methylation
RIPK2	receptor-interacting serine-threonine kinase 2	Human_T-Cell_B-Cell_Activation_Methylation
IRF8	interferon regulatory factor 8	Human_T_Helper_Cell_Differentiation
PERP	PERP TP53 apoptosis effector	Human_T_Helper_Cell_Differentiation
TGIF1	TGFB-induced factor homeobox 1	Human_T_Helper_Cell_Differentiation
TNFSF11	tumor necrosis factor (ligand) superfamily member 11	Human_T_Helper_Cell_Differentiation

R-packages Used

List of R-packages used for analysis:

1. Charm
2. Bioconductor
3. BiocGenerics
4. plyr
5. RCircos

References

1. Aryee MJ et al., Accurate genome-scale percentage DNA methylation estimates from microarray data, Biostatistics (2011) 12(2): 197-210

2. Seth Falcon, Benilton Carvalho with contributions by Vince Carey, Matt Settles and Kristof de Beuf. pdInfoBuilder: Platform Design Information Package Builder. R package version 1.24.0.

3. Rafael A. Irizarry, Martin Aryee, Hector Corrada Bravo, Kasper D. Hansen and Harris A. Jaffee (). bumphunter: Bump Hunter. R package version 1.0.0.

4. RCircos: an R package for Circos 2D track plots

5. Irizarry RA, Ladd-Acosta C, Carvalho B, et al. Comprehensive high-throughput arrays for relative methylation (CHARM) Genome Res. 2008;18(5):780790.

6. Martin J. Aryee, Zhijin Wu, Christine Ladd-Acosta, Brian Herb, Andrew P. Feinberg, Srinivasan Yegnasubramanian, and Rafael A. Irizarry. Accurate genome-scale percentage dna methylation estimates from microarray data. Biostatistics, 12(2):197-210, 2011.

Supplementary Tables

The results of the pathway analysis and their links are provided.

1. H460 knock-in genes with corresponding pathways, with changed cutoff of -5000 to +1500.(https://app.box.com/s/7xr01qkv3xx41oo5izew)

2. H460 parent genes with corresponding pathways, with changed cutoff of -5000 to +1500.(https://app.box.com/s/pu0n89gusqlf1umjwwhh)