ISSN: 0976-3031

Available Online at http://www.recentscientific.com

International Journal of Recent Scientific Research
Vol. 7, Issue, 11, pp. 14181-14188, November, 2016

International Journal of Recent Scientific Research

Research Article

ANALYSIS OF GENE EXPRESSIONS IN BOTH MELANOMA AND MELANOCYTE CELL LINES BY CHIP-SEQ TECHNOLOGY

Bilwashri.H and Seema J Patel

Department of Biotechnology, GMIT College, Davangere

ARTICLE INFO

Article History:

Received 17th August, 2016
Received in revised form 21th September, 2016
Accepted 28th October, 2016
Published online 28th November, 2016

Key Words:

Melanoma and melanocyte cell lines, histone methylation, gene expressions, chip-seq technology

ABSTRACT

The melanoma and melanocyte cell lines have irreversible gene variations that constantly change DNA sequences, chromosomal alterations and gene mutations. DNA binding proteins such as transcription factors or modified histone proteins and gene mutations within melanoma and melanocyte cell lines are identified using genome-wide Chip-Seq method. Different Chip-Seq algorithms are used to read the whole genome sequencing data to align reads and find the differential gene expressions, gene-protein expression patterns, gene alterations and gene interaction that helps to provide novel prognostic and potential therapeutic targets. Here, HEMn_H3K4me3 of reference genome build hg18 were mapped with Illumina genome analyzer. The aligned reads of gap size was set at FDR ≤5% was estimated with E-value divided by number of identified candidate domains. DNA methylation and histone modifications of genes such as HOXD29, COLA2, HSPB6 and MT1G were validated by potential epigenetic targets and are used as potential drug targets.

INTRODUCTION

Our basic thoughtful of the genetic sources for human disease has only just begun to change as an outcome of the entire decoding of the human genome in 2001, led by the revolutionary efforts of a number of research groups [1-4]. The unique findings have enlightened and stimulated researchers, to re-work our current sample of cancer development and metastatic progression to benefit of changing our way to the level of gene function and regulation. It is fascinating to imagine of the outstanding technologies, which have occurred over the last 10 years, such as complexity, the sheer beauty, the intricate nature of the human genome and how it relates to the development of human malignant transformation genes of melanoma in this process have not been fully described [6-11].

Nowadays, the clinical and histopathological criteria are standard use to define the prognosis and predicting the clinical outcome of melanoma patients. Unfortunately, failed in utilizing gene expression profiling to predict which patients have adjuvant therapy for stage III and IV disease and, also which patients will respond to therapy based on their melanoma genes. The development of prognostic test is greatly anticipated to allow clinical treatments for improved predictors of response, as shown with gene microarray profiling of melanoma [13].

The permanent changes are occurring inside the human DNA sequence, including gene mutations, chromosomal deletions and amplifications; these are implicated in the improvement and progression of melanoma. The biological understanding of methylation of DNA and modification patterns of histone in NHEM and melanoma cells are matching to gene expression patterns and genetic alterations, also providing novel diagnostic markers and therapeutic targets. This area of research is important for the future treatment options of cancer patients [13].

Advanced technologies are available such as DMH(differential methylation hybridization),restriction landmark genomic scanning (RLGS),epigenomic reactivation screening, methyl-DIP(methylated DNA immunoprecipitation) for DNA methylation profiling, and ChIP-on-chip arrays for profiling histone modifications in the field of cancer epigenetic to be considered at a genome-wide level [29]. We are finding more genes that are modified and its impact on functions in melanoma progression.

Next-generation sequencing (NGS) is most important technologies in the genetic sciences of the preceding 30 years. It refers to non-Sanger-based high-throughput DNA sequencing technologies.

ChIP-seq is the first applications of NGS which has significantly improved to understand the transcriptional regulation and interactions of DNA-proteins on global mapping. This can be obtained by chromatin immunoprecipitation (ChIP) along with sequencing (ChIP-seq).

Corresponding author: **Bilwashri.H*
Department of Biotechnology, GMIT College, Davangere

Chromatin Immunoprecipitation (ChIP) is a type of immunoprecipitation and investigational techniques which is used to interaction between DNA and proteins in the cell. It is able to recognize the binding sites of some DNA-associated proteins on a genome-wide scale [14].

This technique is helpful for understanding the expression of gene analysis, modification of histone methylation and finding the binding motif.

METHODOLOGY

Data Search

To test histone methylation of melanocytes and melanoma in human cell lines, ChIP-seq data were retrieved from two public domain databases such as Array Express and SRA. The Chip-Seq data consists of 14 samples of which methylation of DNA and two histone modifications (H3K4me3 and H3K27me3), in two types of normal melanocytes (HEMn and HEMa) and melanoma cell lines SK-MEL-28. Using this data, to analyze the relationship between epigenetic factors and gene expression status in both normal and melanoma cells (SRP009470). ChIP-seq of H3K4me3 and H3K27me3 in two types of normal melanocytes (HEMn and HEMa) and melanoma cell lines (SK-MEL-28) datasets (GSE33930) each with single replicates of both normal and melanoma cell lines. The FASTQ sequences files were taken from EBI website (SRP009470). The ChIP-seq datasets consists of 12 libraries from different cell lines (SK-MEL-28) each cell line has single replicates with 2.6 million pairs in a library. The Chip-Chip sequence annotations were prepared using GPL9115 Illumina Genome Analyzer II platform to annotate the differential gene expression and classification of genes in both DNA methylation in melanoma and melanocytes cell lines. After collecting the data, we are using Galaxy online bioinformatics tool which is very helpful to understand the Next-generation sequencing platforms.

Read quality assessment

FastQC is a software package which is available in galaxy online tool which is very easy to check the overall sequence quality, for each sample. It gives the reports of per-base and per-read quality and also the level of duplication content and possible contaminations in sequence. It gives the result of mapping rate and also improves this by reducing low-quality bases or adaptor sequences for each sample.

Read mapping algorithms

There are more number of mapping software packages are available which is analyze to short map reads with reference genome. But in this work, we are using BWA (burrows wheel aligner) tool which is available in Galaxy.

Post-processing of mapped reads

After alignment, the mapped reads are using to improved peak calling specificity and sensitivity. This step addressed the issues of duplicate reads, poorly mapped reads, and mapping reads to multiple locations.

Reads with low mapping quality

It is a general processes to remove reads with low mapping quality. So using SAM tools to remove the reads based on a mapping quality score. This tool does not contain the pairing information while performing mapping quality depends on reducing paired-end reads. Therefore, we are using BWA tool to removing pairs of reads that have at least one or two ends of the mapping quality. After the filtering or reducing, the pairing information will be maintained.

Duplicate reads

In this experiment, the sequencing library is generated from a smaller amount of DNA compared to DNA or RNA sequencing. Duplicate reads are frequently present in ChIP-Seq datasets. Using advanced applications are helpful to remove this reads as well as removed artifacts. But in this work, we want large number of duplicate reads to identification of more number of transcriptional factors and binding sites in peak. In chip-seq experiment it is not necessary to remove the duplicate reads.

Reads mapping to multiple genomic locations

In these methods, mapping reads with multiple genomic locations are frequently removed. It depends on epigenetic mark. For example, a substantial fraction of the H3K9me3 modification occurs in regions consists of repetitive DNA sequences. We take data from ENCODE project (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/ website),in the survey of 12 H3K9me3 ChIP-Seq datasets , between 16% and 28% of the mapped reads have more matches in the genome.

Peak calling

After mapping reads with reference genome, this method is helpful to identify the transcriptional factors in peak regions and also gives the enrichment level of histone modification. It is useful to define the individual binding domains and profiles. We are using MACS (Model-based Analysis of ChIP-Seq) to identify binding sites with its high sensitivity and specificity. It also scans the genome for enriched regions and overlapping regions into the peaks and captures the signal variation.

Differential Binding Analysis

The DiffBind packages are used to recognition of genomic loci which is differentially between two conditions. It was developed more algorithms used for differential gene expression analysis by RNA-seq. These methods gives the results depends on normalization of read counts in ChIP-seq datasets.

Motif Analysis

The important aims of ChIP-seq methods are to discover the sequence of motifs in binding proteins with genome. The Multiple EM for Motif Elicitation (MEME) algorithm is the common tool for discovery of motif. It is founds more number of motifs for each sample and also analysis motifs on low quality ChIP-seq data.

Chromatin State

Next useful analysis of ChIP-seq methods are using with ENCODE project which is helpful to describe genomic regions based on modification of histone. Several histone modifications are tested using antibodies in ChIP-seq data. To obtained a profile of that histone mark in sample. Their ongoing protocol

is to cover antibody validation, sequencing depth, experimental replication, metadata reporting, and quality assessment of data.

Data visualization

We are using Integrative Genomics Viewer (IGV) which is genomic browser tool that easily visualize genomic features. CEAS tool input format is Wig file and other tools like MACS version 2 input format is not in Wig file and SICER also gets in a this file with a large size (200 bp), and also designed IGV tools are found in bedGraph, Wig and tiled data format (TDF) files for visualization of data.

In bedGraph file, reduced/filtered reads are obtained in BAM format. Single-end reads are completed by the fragment length in library (default 200 bp). In paired-end reads, the first end and completed by the fragment length from mapping positions of the two ends, instead by the average fragment length in library. The bed file is used to achieve a bedGraph file from the genome Coverage Bed command in BED Tools.

The Wig file is developed from the bedGraph file, using to figure out the coverage read (default: 20 bp). This read is normalized to a library size of one million mapped reads, and converted into the TDF format using the TDF command from the IGVtools package. The normalized coverage in TDF format and recognized peaks in bed format that is able to visualize by transferring files to IGV, or provided igv_session.xml file in IGV.

Peak and binding profile annotation module

ChIP-Seq data consists of three tools to annotate peaks and binding profiles. These are MEME to recognizing the TF binding motif; Cis-regulatory Element Annotation System(CEAS) for developing binding profiles with genomic features and also predicting possible g regulated genes by cis-regulatory elements; and another tool for calculating enrichment in Gene ontology (GO) used for peak-associated genes.

RESULTS AND DISCUSSION

In these studies, we have taken 4 histone modification datasets, 2 normal melanocytes and 3 melanoma cell lines to analyze the DNA methylated proteins involved in histone methylation human melanocytes and melanoma cell lines. The data were produced in Illumina consists of 36 base paired-end reads with the insert length 200.

The selected datasets is used for Chip-Seq Data analysis pipelines to analyze the data in 4 different steps.

1. Quality Reads/Preprocessing
2. Peak calling
3. Differential binding Analysis
4. Validation and downstream analysis of Motif Analysis, Annotation, integrating binding and expression data

Chip-Seq Data Selection

We performed comprehensive DNA methylation profiling of H3K4me3 and H3K27me3 in two types of normal melanocytes (HEMn and HEMa) and melanoma cell lines (SK-MEL-28) (Table 1). The datasets contains 14 replicates of which different spots and large number of bases. The average length of all probes has 36 bases that aligned in ASCII values.

We are directly compared the epigenomic of melanoma cell lines SK-MEL-28 to normal epidermal melanocytes isolated from adults (HEMa) and neonates (HEMn). Briefly, we applied Digital Restriction Enzyme Analysis of Methylation (DREAM) to quantitatively map DNA methylation on a genome-wide scale, and ChIP-seq analysis to map the H3K4me3 pattern.

Quality analysis

FastQC tools are using to predict raw sequence quality, using basic statistical mean median calculations to analyze base quality. The base quality shows the base call that is expressed in Phred scale where log10 is taken of the probability that the bases are wrong and multiplied by -10. The quality values of all the datasets has typically range from 0 to 40. Instead of numbers, they are encoded with ASCII characters in FASTQ files in order to save space (Table:.2). The FastQC output file is found in html files and igv_session.xml files for visualization of data.

Sequence quality in base

The sequence quality in base indicates the quality of base call; it is expressed in Phred scale, where log10 is taken of the probability that the base is wrong and multiplied by -10. The overall quality of all datasets has constant quality at the range of 36-37%, that shows the quality of distribution of based is constant is overall datasets.

Table1 Chip-seq Data selection of histone methylation of normal melanocytes and melanoma cell lines selected from SRA database

Library Name	Run	spots	bases	avgLength	size_MB
GSM838927: HEMn_H3K4me3	SRR385628	25450458	916216488	36	518
GSM838928: HEMn_H3K27me3	SRR385629	24525343	882912348	36	487
GSM838929: HEMn_H3	SRR385630	21074238	758672568	36	426
GSM838930: HEMa_H3K4me3	SRR385631	26413372	950881392	36	552
GSM838931: HEMa_H3K27me3	SRR385632	26171432	942171552	36	548
GSM838932: HEMa_H3	SRR385633	22292328	802523808	36	434
GSM838933: SK-MEL-28_H3K4me3	SRR385634	12369630	445306680	36	325
GSM838934: SK-MEL-28_H3K27me3	SRR385635	12832737	461978532	36	334
GSM838934: SK-MEL-28_H3K27me3	SRR385636	40229444	1448259984	36	954
GSM838935: SK-MEL-28_H3	SRR385637	12399427	446379372	36	315
GSM838936: LOXIMVI_H3K4me3	SRR385638	12293978	442583208	36	311
GSM838937: LOXIMVI_H3K27me3	SRR385639	12301664	442859904	36	315
GSM838937: LOXIMVI_H3K27me3	SRR385640	41145481	1481237316	36	960
GSM838938: LOXIMVI_H3	SRR385641	12559043	452125548	36	347

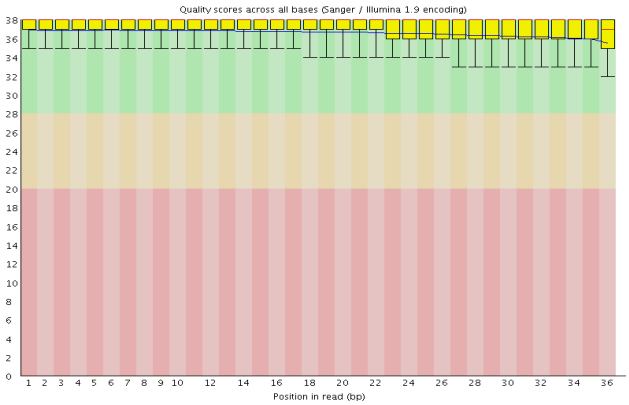

Fig.1 Output image PBS

Sequence quality in tile

It appears only in Illumina library base quality reads. The quality scores from each tile across all bases. The colors are on cold to hot scale; the overall blue color plot represents high quality of per base tile sequence quality. Our result shows at the position of 28. Hotter color represents that worse quality and remaining replicates shows good quality.

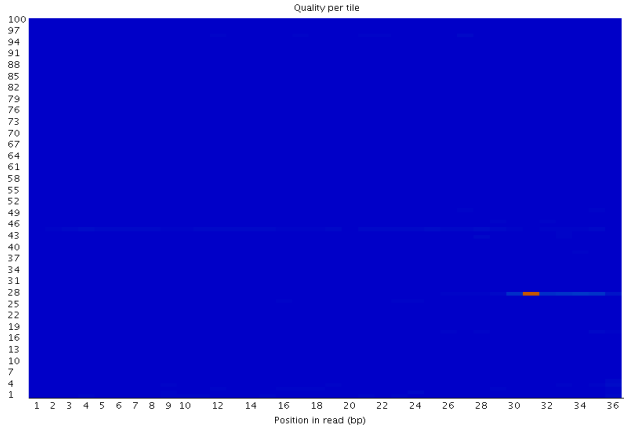

Fig.2 output image PTS

Sequence quality score

The sequence quality score represents overall quality of all sequences is in between the quality of mean sequence quality which represents in Phred score, the overall average quality of sequence is in 37 that shows good quality of bases.

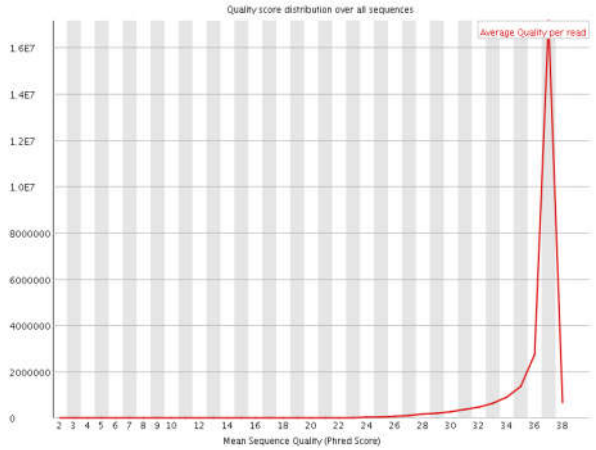

Fig.10 Output image PSQ

Base sequence content

The per base sequence content represents the proportion of each base position and depends on ATGC. Our result shows warning that represent in the A / T or G / C is greater than 10% in any position of bases.

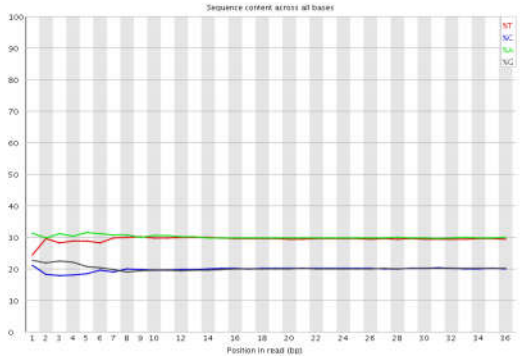

Fig. 4 output image PBSC

GC content

The GC content must follow a normal distribution and center on the GC content of the organism. In FASTQC the GC content per base position, which should produce a flat line at the level of the source of the genome's GC content. Different GC content at certain base positions indicated a presence of an overrepresented sequence in the library. The previously discussed sequence-specific bias shows also in the GC content plot.

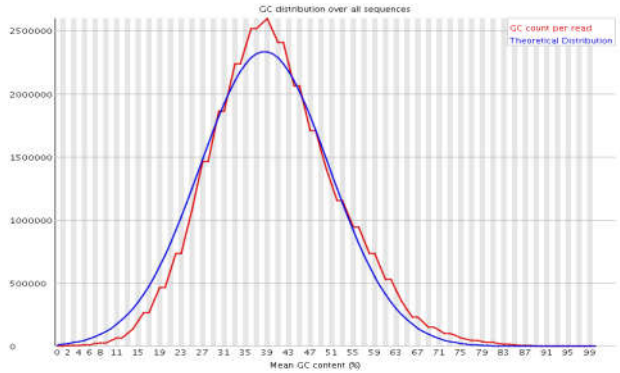

Fig.5 Output image GC%

N content

The percentage of base calls at each position for which an N calls appearing in a sequence, especially nearer the end of a sequence. This proportion raises few percent that analysis which was unable to interpret the data well enough to make valid base calls.

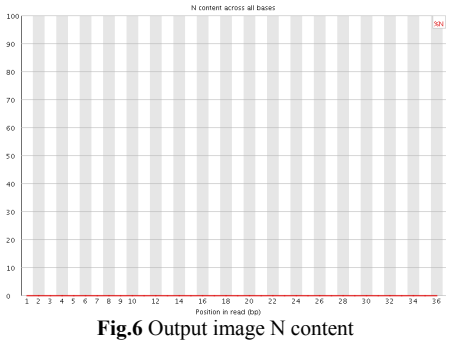

Fig.6 Output image N content

Length distribution in Sequence

The overall length of sequence is generated based on sequence fragment of 36 bases. Even within uniform length libraries of trim sequence to remove poor quality base calls from the end.

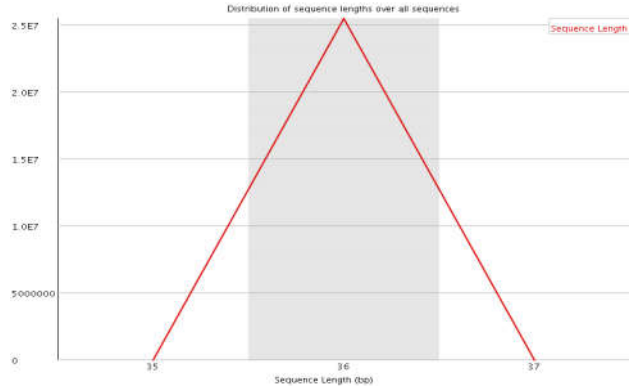

Fig.7 Output image SLD

The module generates a distribution of fragment sizes in the analyzed files.

Duplicate level in Sequence

Most sequences occurs once only in the final set. A less level of duplication indicates a more level of the coverage in the target sequence, but a high level of duplication is indicates some kind of enrichment bias. A high level of identical reads can indicate PCR over amplification, but in the context of RNA-seq duplicates are often a natural consequence of sequencing highly expressed transcripts. For differential expression analysis, it is not recommended to remove duplicates, because they would flatten the dynamic range and read counts would not be proportional to the expression level any more. However, if a sparsely covered transcript has a tower of reads in one position, this is likely to indicate a PCR artifact.

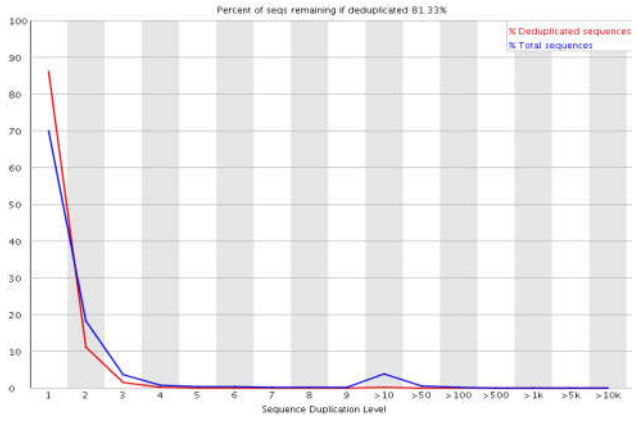

Fig.8 Output image %SLD

Adapter content

Illumina techniques are used to design an adapter which needs to trim end of the sequences at particular length. Also other tags such as multiplexing identifiers and primers need to be removed. The problem with this "read-through" situation is that the adapter in the 3′ end can be partial and therefore difficult to recognize. Finally, if the data come from a public database, the adapter sequence information might not even be available.

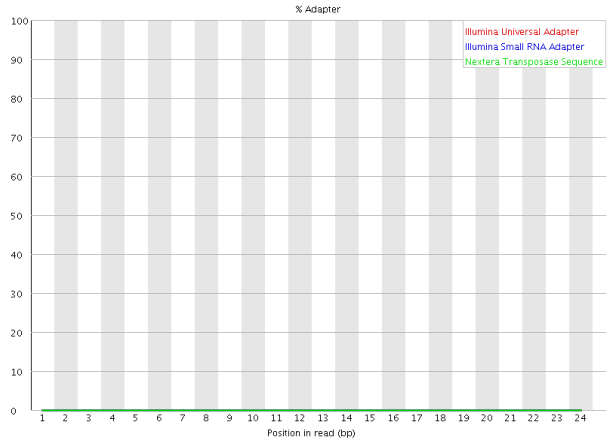

Fig.16 output image Adapter

K-mers

K-mer represents the Brujin graph associated with (k-1)-mer. Two nodes are connected i.e. A and B, if the there is a k-mer whose prefix is the (k 1)-mer of the node A and the suffix is the (k 1)-mer of the node B. It consists of simply extracting all k-mers from reads and connecting the nodes representing the (k 1)-mers.

Table 2 Quality control of chip-seq data using fastqc server

Parameters	1	2	3	4	5	6	7	8	9	10
PBSQ	37	37	37	37	37	37	37	37	37	37
PTSQ%	28	100	74	100	100	100	100	100	100	100
PSQS	37	37	37	37	37	37	37	37	37	37
PBSC (AT/GC)	30/20	30/20	30/20	30/20	30/20	30/20	30/30	30/20	30/20	30/30
GC%	40	40	39	39	39	40	39	39	38	39
PBNC	0-36	0-36	0-36	0-36	0-36	0-36	0-36	0-36	0-36	0-36
SLD	36	36	36	36	36	36	36	36	36	36
SDL%	81.33	76.77	89.51	75.33	57.23	90.62	92.1	94.33	71.69	93.56
OS	0	0	0	0	0	0	0	0	0	0
Adapter	1-24	1-24	1-24	1-24	1-24	1-24	1-24	1-24	1-24	1-24
K-mer	0	11.56	7.33	14.11	21.21	7.49	0	6.49	5.38	0

Chip-Seq Mapping

Using Mapping algorithm such as BWA of melanoma and melanocytes cell lines data is mapped with reference genome hg18 whole genome sequence, the most accurate of short read aligners of datasets from transcriptional factors (TFs) and MACS peak calling functions which containing motifs were identified from BWA-mapped reads. However the mapping quality is assigned for each read is mapped unambiguously, but its mate falls in tandom repeats shows 52 % quality of alignments. In BWA different parameters is used such as –n, -q, -k and -1, the overall results is predicted in table: .3.The accuracy of alignments is highly repetitive that helps to find functional motifs of bases. A base of each quality is paired with high end model of unmapped reads with a high error rate and high quality anomalous pairs to fix potential alignment errors.

Table 3 overall results predicted from BWA

Library	Type	Total (millions)	BWA		
			A	B	C
GSM838927	Melanoma cell lines neonatal	26	10	500	0.427
GSM838928	Human epidermal melanocytes, Adult	25	16	500	0.182
GSM838929	Melanoma cell lines neonatal	21	16	488	0.841
GSM838930	Human epidermal melanocytes, Adult	27	10	321	0.125
GSM838931	Human epidermal melanocytes, Adult	26	18	452	0.632
GSM838932	Human epidermal melanocytes, Adult	22	14	128	0.185
GSM838933	Melanocyte cell lines	14	14	421	0.188
GSM838934	Melanocyte cell lines	16	16	441	0.425
GSM838934	Melanocyte cell lines	42	12	237	0.424
GSM838935	Melanocyte cell lines	14	12	451	0.162
GSM838936	Melanoma cell lines	12	13	508	0.751
GSM838937	Melanoma cell lines	13	13	566	0.475
GSM838937	Melanoma cell lines	16	16	95	0.471
GSM838938	Melanoma cell lines	22	16	192	0.455

A: percentage of pairs with at least one uniquely mapped end; B: percentage of pairs with multiple mapping locations; C: others, including unmapped pairs, pairs with only one mapped end and improperly mapped pairs with small insertion size or wrong orientation.

Peak calling

Peak calling methods helps to identification of protein-DNA binding of sequence reads enriched in the genome after mapping in a particular region. Here, relatively unbiased reads that accumulate at sites of protein binding faster than in background region. The tools BWA is used to align the millions of sequence read in this datasets with reference genome. This helps in identification of mismatches allowed and affects the percentage of sequence which can be placed by the reads to multiple locations of motif repeat regions that mask true binding events. The MACS tools used in identification of transcriptional factor peaks callers and also did the work of removing redundant reads and read-shifting. It uses control samples and local statistics to reduce the bias and find out an empirical FDR.

Table 4 detailed information the output results

chr	start	end	length	summit	tags	#NAME?	fold_enrichment	FDR(%)
chr1	17616	17729	114	38	7	68.54	21.55	21.41
chr1	17979	18435	457	268	34	267.29	42.42	2.06
chr1	18451	18548	98	37	11	106.2	41.69	6.22
chr1	18578	18689	112	38	8	70.15	38.92	20.28
chr1	703609	703713	105	49	8	80.66	29.78	12.83
chr1	704470	704611	142	79	9	92.6	39.67	8.77
chr1	704635	704705	71	36	6	60.85	39.67	32.96
chr1	751349	751461	113	77	6	63.02	19.83	29.22
chr1	751501	751596	96	70	6	56.93	18.15	40.68
chr1	751676	751773	98	32	7	67.12	26.45	22.85
chr1	751831	752133	303	178	22	197.61	39.67	2.29
chr1	752238	752349	112	86	6	63.27	26.45	28.64
chr1	883622	883725	104	37	8	91.69	39.67	9

Read Shifting

Only one-end of fragment of the aligned reads of 36bp length sequences from single-end sequencing is called read. They aligned to sense or antisense strands of DNA fragments which is pulled down in the immunoprecipitation. They are shifted and data from combination of both strands to find out most likely bases involved in binding proteins.

Peak Identification

A peak is to find out the number of reads which exceeds a pre-determined potential value of minimum enrichment compared to background signal, frequently in a sliding window across the genome. We must determine if less number of high-quality peaks is preferred over low quality peaks.

Significance Analysis

P value for called peaks is computed by the number of peak callers where as others use the height of the peaks over background to rank peaks, but these do not give the accurate statistical values. The false discovery rate (FDR) is used to provide a truer peak list and can be computed from the P values.

Duplicate removal of TF in peak calling

Using MACS tool is used to identify Chip-seq peaks identified which is found in three ER libraries. Before duplicate removal the minimum errors is 486 and maximum Differential values is 670 peaks. There are 362 peaks to 1121 peaks only from reads after duplication is filtered with noise that likely associated with ER binding sites. MEME are used to identify transcriptional motifs those bindings with ER binding of transcriptional regulation sites. There are 1676 ER binding motifs has unique peaks from reads without duplication removal.

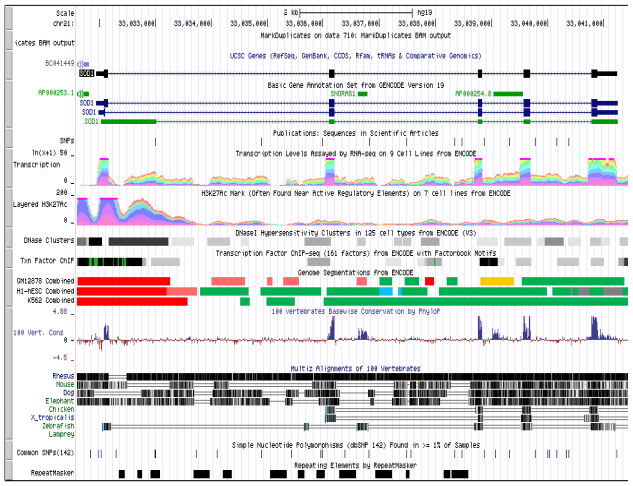

Fig.10 UCSC Image

There are different approaches that differentially expressed in melanoma and melanocytes cell lines of H3K4Me4_HEMn cell types, HEMn_H3K4me3, HEMn_H3, HEMa_H3K4me3, HEMa_H3K27me3, HEMa_H3, SK-MEL-28_H3K4me3, SK-MEL-28_H3K27me3, SK-MEL-28_H3, L, of three different libraries. We further predicts the distribution of duplicate peak versus non-peak region, we extended 100bp of single pair that remains duplicate with two regions.

Novel Motifs identification

Using MEME-Chip web server to identify sequence motifs that computationally predicted from genomic region. A FASTA formatted sequence having 100bp length is helps to discover the motifs and also helps to enrichment analysis to trim each sequence to identify motifs. A DREME algorithm predicts 18144 motifs based on p-value. The motif occurrence is the probability of a random sequence with the same length as motif matching of the sequence with better score. The know motifs in their peaks datasets from TF binding experiment pipeline annotates with peaks with information from Ensembl and RefSeq genes, repeating sequence elements, conservation

scores, and GO terms. The combined results from multiple TF-binding experiments are provided by Boolean algebra. The widely used tool to identify de novo motif are, Weeder, MEME, and ChIPMunk, and a scanning tool to motifs that is Paster.

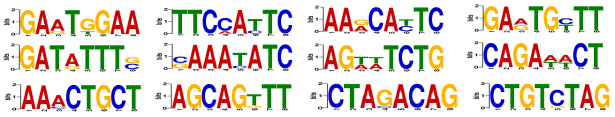

Fig .11 output images of motifs

Table 5 Detailed information about output images

Consensus Sequences	Cout	Sig Values	Coverage
atnnna	39160	1068.9	88.70%
cagag	2239	1064.9	37.20%
cactc	2174	1064.9	34.50%
ctgtc	2189	1064.9	34.00%
ctnagaa	1784	1063	26.30%
cannncac	2665	1061.7	39.80%
ccannnnnctc	1469	1059.7	21.80%
ccanncga	1272	1059.7	12.70%
caanngga	1332	1059.7	29.00%
cagnnnngaa	1451	1059.7	29.80%
ccanncca	7767	1059.7	20.80%
cacnnnnntcc	1066	1059.7	12.00%
atgnnnncat	1263	1059.7	20.50%
actnncag	1173	1059.7	26.80%
atgnnatc	1816	1059.7	16.50%
acannnaga	1610	1059.7	26.80%
atcnnnnngaa	1811	1059.7	26.30%
agannnnttc	2649	1059.7	35.50%

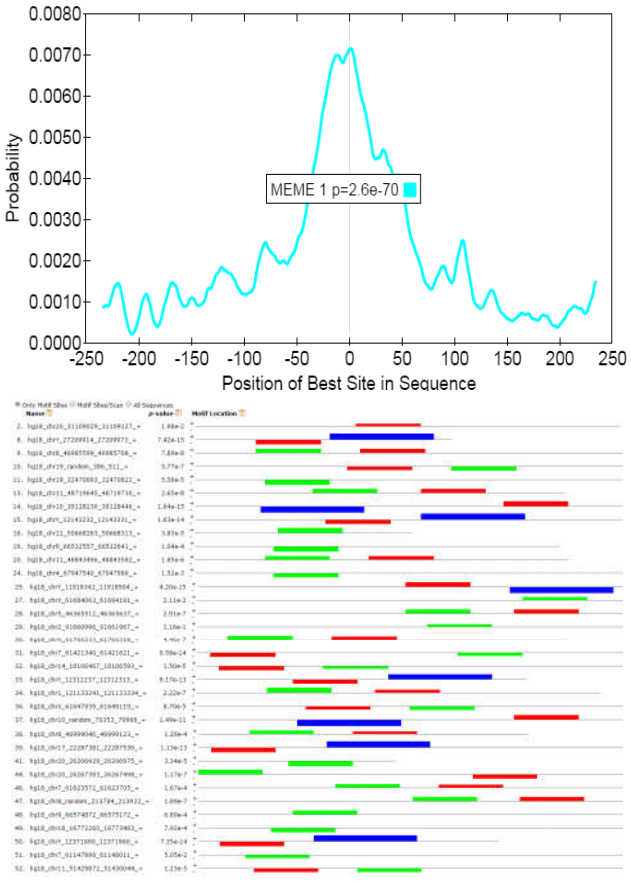

Fig .12 output images of MEME

Expression patterns of H3K4me3 and H3K27me3 of melanocytes cell lines

The genome wide associations of epigenetic features during immune cells such as CD4+ receptor within T cell differentiation are analyzed to gene expression. In the results of short reads we have aligned 296 millions of all 14 datasets of genome hg18. We have identified characteristic genes that significantly associated with TF binding sites is listed in table.

Table 6 gene expressions are found in both melanocyte cell lines

H3K4me3 TF	H3K27me3 TF	Cell lines	Base over	Z-score
RAD21	CTCF	H3K4me3	0.901119955	501.8198328
RAD21	CTCF	H3K4me3	0.786580707	499.9208537
RAD21	CTCF	H3K4me3	0.764023986	474.8276646
CTCF	RAD21	H3K4me3	0.680706638	455.0569546
CTCF	RAD21	H3K27me3	0.630203579	419.5486292
RAD21	CTCF	H3K27me3	0.980694828	415.6684443
USF2	USF1	H3K27me3	0.776649205	415.4739101
CTCF	RAD21	H3K27me3	0.521009954	414.9364309
SMC3	RAD21	H3K27me3	0.648169546	370.2196073
SMC3	RAD21	H3K27me3	0.501175033	366.9276815
USF1	USF2	H3K4me3	0.309495006	364.6448006
RAD21	CTCF	H3K4me3	0.722715121	357.4661918
JUN	FOS	H3K4me3	0.536575244	345.8424252
FOS	FOSL1	H3K4me3	0.495241377	345.2163727
RAD21	SMC3	H3K4me3	0.846764621	339.4776239
FOSL1	FOS	H3K4me3	0.54822508	338.8380125
SMC3	CTCF	H3K27me3	0.708379235	336.2395811
FOS	JUN	H3K27me3	0.519990815	333.4275419
CTCF	RAD21	H3K27me3	0.556853898	303.7776133
SMC3	CTCF	H3K27me3	0.887045357	303.5921456
GATA1	GATA2	H3K27me3	0.887774419	297.2367316

We have identified the protein biding region by mapping the binding locations of 119 DNA-binding motifs of TRIM28, SETDB1, ZNF274RTN4, DNMT3A, G1ME and FNIP1 genes is predicted in promoter regions. The other genes such as KAP1, SETDB1 and SUZ1 are involved in histone methylation which is complexes with H3K4me3 and H3K27me3 cell lines. The CHD2 of ATP-dependent chromatin complexes is differentially expressed in melanoma cell lines.

Expression patterns of SK-MEL-28 of melanoma cell lines

The expression of SK-MEL-28 in melanoma cell lines is helps to identify differential expression of metalloproteinase binds integrin receptors of tumor cells. To evaluate the molecular effects of SK-Mel-28 cells that express in fibroblasts of integrins, cadherins, caspases, and TP53. There are 8 integrin binding expression profiles that upregulate in SK-MEL-28 cell lines such as CASP3, CASP6, CASP8, CDKN2A, CDKN1A, TP53 and E-CDH. Other genes are found such as ITGB6, ITGB7, CASP3, TP53, and CDKN1B in fibroblasts.

Finally, we find out differentially expressed genes which are present in both melanocytes and melanoma human cell lines by using chip-seq technology.

CONCLUSION

Here we study Chip-Seq data such as H3K4me3 and H3K27me3 in two types of normal melanocytes and melanoma cell lines. We have done mapping read, peak calling for binding sites and functional analysis. By applying Galaxy of singled read, we identified the filtering of duplicates which increases the sensitivity of peak calling using MACS. This result suggest the essential of providing duplicate filtering for

TF peak calling and using all mapped reads for the estimation and identification of differential binding sites. On the other hand, it has less impact on peak calling and showing broad binding profile like H3K27me3.In the current data of mapping H3K4me3 and H3K27me7 of human epidermal melanocytes; Sk_MEL_28_H3K4me3 and Sk_MEL_28_H3K27me3 of melanoma cell lines of melanoma cell lines is mapped with human genome hg18 have predicted 272 genes that significantly associated with epigenetic function. This result helps future researchers to predict markers of different cancer types of diagnosis and treatment.

References

1. Venter JC, Adams MD, Myers EW (2001), The sequence of the human genome. Science, vol291(5507), pp: 1304-1351.
2. Lander ES, Linton LM, Birren B (2001) Initial sequencing and analysis of the human genome. Nature; vol 409(6822), pp:860-921,
3. Birney E,Stamatoyannopoulos JA, Dutta A (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature; vol447(7146), pp:799-8.
4. Levy S, Sutton G, Ng PC (2007) The diploid genome sequence of an individual human. PLoS Biol., vol 5(10), pp: e254.
5. Jemal A, Siegel R, Ward E. Cancer statistics, (2007) CA Cancer J Clin. 2007; vol 57(1), pp: 43-66.
6. Hoek K, Rimm DL, Williams KR (2004) Expression profiling reveals novel pathways in the transformation of melanocytes to melanomas. Cancer Res, vol 64(15), pp: 5270-5282.
7. Smith AP, Hoek K, Becker D (2005) Whole-genome expression profiling of the melanoma progression pathway reveals marked molecular differences between nevi/melanoma in situ and advanced-stage melanomas. Cancer BiolTher.vol 4(9),pp: 1018-1029.
8. Haqq C, Nosrati M, Sudilovsky D (2005) The gene expression signatures of melanoma progression.Proc Natl Acad Sci U S A. vol 102(17),pp: 6092-6097.
9. Talantov D, Mazumder A, Yu JX, (2005) Novel genes associated with malignant melanoma but not benign melanocytic lesions. Clin Cancer Res. vol 11(20), pp:7234-7242.
10. Winnepenninckx V, Lazar V, Michiels S (2006) Gene expression profiling of primary cutaneous melanoma and clinical outcome. *J Natl Cancer Inst*. vol 98(7),pp: 472-482.
11. Jaeger J, Koczan D, Thiesen HJ (2007) Gene expression signatures for tumor progression, tumor subtype, and tumor thickness in laser-microdissected melanoma tissues. Clin Cancer Res. vol 13(3),pp:806-815.
12. Haddad FF, Stall A, Messina J. (1999) The progression of melanoma nodal metastasis is dependent on tumor thickness of the primary lesion. Ann Surg Oncol. vol6(2), pp:144-149.
13. Rao J, Seligson D, Hemstreet GP (2002) Protein expression analysis using quantitative fluorescence image analysis on tissue microarray slides. Biotechniques.vol32(4), pp:924-926, 928-930, 932.
14. Cliff Meldrum, Maria A Doyle and Richard W Tothill, Next-Generation Sequencing for Cancer Diagnostics: a Practical Perspective, (2011) Clin Biochem Rev. vol: 32, pp: 177– 195.
15. Leiter U, Eigentler T, Garbe C,Epidemiology of skin cancer, (2014) Adv Exp Med Biol. vol 810, pp120-40.
16. Aleksandar Sekulic, MD, PhD, Paul Haluska, Jr, MD, PhD, and Svetomir N. Markovic, MD, PhD, (2008) Malignant Melanoma in the 21st Century: The Emerging Molecular Landscape, Mayo Clin Proc. vol 83, pp 825–846.

****

How to cite this article:

Bilwashri.H and Seema J Patel.2016, Analysis of Gene Expressions in Both Melanoma and Melanocyte cell Lines by Chip-Seq Technology. *Int J Recent Sci Res.* 7(11), pp. 14181-14188.