

EPIGENETIC REGULATIONS, MOTIF AND PATHWAY IDENTIFICATION OF MYELOFIBROSIS CHIP SEQUENCES

Sravani Sane1*, Shylesh Murthy IA2 and Preenon Bagchi3

^{1*} Padmashree Institute of Management and Sciences, Bengaluru, India. Corresponding author: Email: sanesravani836@gmail.com
 ²Junior Researcher, Vasishth Academy of Advanced Studies and Research, Bengaluru, India.

³Institute of Biosciences and Technology, MGM University, India.

Abstract: Next generation sequence technique that is Chromatin immunoprecipitation (ChIP) on the ChIP–Seq of Myelofibrosis. Myelofibrosis is a type of blood cancer which is also considered as a form of chronic leukemia. It is two types primary myelofibrosis and secondary myelofibrosis its leads to complication of an autoimmune disease, additional disease features include hepatosplenomegaly, extramedullary hematopoiesis (EMH) etc., In this work, I have retrieved its chip-Seq from SRA database, removed the PCR duplicates and finally identified the genome enriched region using the MAC2 call peak tool This annotated peak table was used to identify the motifs present in the chip-Seq of Myelofibrosis. Peak table was annotated to identify the genes and the corresponding pathways was identified from the KEGG pathway.

Keywords: Chromatin immunoprecipitation (ChIP), Myelofibrosis, Next Generation sequencing, PCR duplicates, MACS2 Callpeak, Motif, KEGG pathway.

1 INTRODUCTION

ChIP sequencing (ChIP-Seq) is a powerful method for identifying genome-wide DNA binding sites for transcription factors and other proteins. "Next generation" genome sequences technology can provide one – two orders of magnitude increase in the amount of sequence that will be cost effective and new technology ChIP – Seq directly provide whole genome of mammalian Protein – DNA interactions [1, 2]. It starts with crosslinking of DNA-protein complexes and then fragmented, they are treated with an exonuclease to cut unbound oligonucleotides. Protein-specific antibodies are used to immunoprecipitation the DNA-protein complex. DNA is extracted and sequenced, giving high-resolution sequences of the protein-binding sites. ChIP Seq has many advantages such as, it captures DNA targets for transcription factors or histone modifications across the entire genome of any organism and also it defines transcription factor binding sites [3]. ChIP–Seq could be the way for genome-wide profiling of DNA-binding proteins, or nucleosomes and attributable to the tremendous progress in next-generation sequencing technology [4].

Myelofibrosis is a type of blood cancer which is also considered as a form of chronic leukemia [5, 6]. In this disease, the bone marrow is replaced by fibrous scar tissue. It is two types primary myelofibrosis (occurs on its own) and secondary myelofibrosis (oc-

179

curs as the result of separate disease) its leads to complication of an autoimmune disease. The bone marrow develops three types of cells white blood cells, red blood cells and platelets [7]. When mutation occurs in a single cell of DNA, it passes on to new cells and leading to defective of cell divides. Characteristic of myelofibrosis like megakaryocytes (over production of giant cells) [8], decreases production of red blood cells – may occur anemia and thrombocytopenia (deficiency of platelet production). Overall world in about 12% cases, primary myelofibrosis will progress to acute myeloid leukemia fastly growing of blood cancer [9, 10].

2. MATERIALS AND METHODS

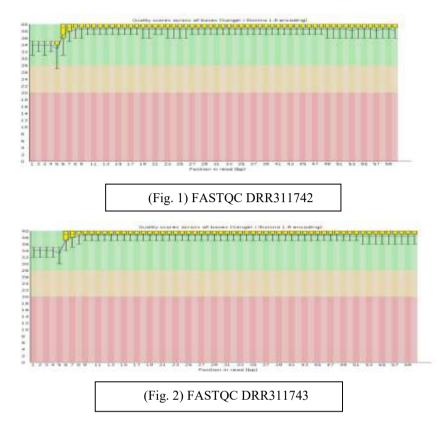
The primary aim of the current research is to provide a detailed overview of epigenetic regulations of Myelofibrosis and throw a light on the pathways involved [11]. Galaxy tutorial by Lauren Mills, Analyzing ChIP-Seq Data in Galaxy is used to analyze *Mus musculus* Myelofibrosis fastq sequences having SRA accession number DRR311742 and DRR311743 were retrived from SRA database. We mapped the reads using Bowtie2 [12]. Using Collect Alignment Summary Metrics tool we take the summary of our alignment. Next using RmDup we remove PCR duplicates [13] and Collect Alignment Summary Metrics tool was re-run. Finally, using MACS2 Callpeak we identify peaks from alignment results [14, 15]. Using Peak calling we identify areas in our genome that have been enriched with our aligned reads, these areas are those where protein interacts with DNA.

Next, we annotate our peaks table to take top 100 most significant peaks and identify the genes overlapping with these peaks and Genes were selected model using Swiss prot with Ramachandran plot [16,17,18] was done. We identify the motifs using SeqPos motif analysis tool. Biological sequence motifs are short conserved sequence pattern associated with distinct functions that usually represents important structural or functional features [19]. Finally, the pathways of these genes were identified from KEGG pathways [20, 21].

3. RESULTS AND DISCUSSION

3.1 FASTQC quality reports

FASTQC quality reports was given quality control to *Mus musculus* chip-sequences with SRA accession number DRR311742 and DRR311743. FASTQC results were given in fig. 2 DRR311742 and fig.3 DRR311743.



3.2 Bowtie2

Mus musculus chip-sequences with SRA accession number DRR311742 and DRR311743 were mapped using *Mus musculus* ref Seq using Bowtie2.

Mapping output of DRR311742:

Concession of the local division of the loca	Distance in the local	10440	100		Contraction of the		100 B	
		-947			20240	The state of the second s		
						And the second of the second of the second s		
						a second s		
						and the second se	12 PERMIT	THE CONTRACTOR OF A DOWN
								HE COLLEGE TYPES
						Automatical and the second second and the second se	and the second se	THE CONTRACT OF A DAY
			1.000			A Design of the other sector is a sector of the sector is a sector is a sector of the sector is a sector is a sector of the sector is a sector of the sector is a sector of the sector is a sector is a sector of the sector is a sector is a sector of the sector is a sector		
					+			
						ADDRESS OF THE OWNER		Desk C.C. Browshark & Fig. Ch.
						The state of the second s		
						Section of the sectio	And a second sec	the set of straight watches
	1.10.000		1.000			- 35 Designation of the state o	And a local distance of the second	
						Contraction of the second s		
the second state of the se	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1		100 100			Approduction in Researching on Dates Table & Contribution the Lother of Researching	 MARKAGE CONTRACTOR CONTRACTOR STRUCTURE 	
	10 mm		10 Aug			The P. Devise of the later in the later is t	Contraction of the second se	the state of the s
1111 (pt) ((p pt) ()) (pp) (pt)	the second	-	the second second			NAME AND ADDRESS OF A DESCRIPTION OF A D	and and a second s	the state of the second s
	1.000		Acres 1000			When the second se	and a second sec	Control of the second state of the second se
THE MUSIC WALLS	11.141.0001	the second se	100 March 100 Ma			 Constraints and the second state of the second state	Children of the second s	- 000-00-00-00-000000 Person III
						The second	110110000000000000000000000000000000000	STATISTICS IN CONTRACTOR
	100000		10 March 100			- Add & specification in THE & More and property fitting frames and the research in the fit	Contracting and the second se second second seco	
the second		-	- 10 Aug			the statement of the second by the second second statement of the second s		terror with a start of the order of the second start
the second state of the second second second second second	CONTRACT.	-	100.000		1.0			CONTRACTOR AND ADDRESS OF ADDRESS OF ADDRESS ADDRESS
the second secon						and the second second in press appell and the second secon	1 New York Call Composed on the Advantage of the Composed of t	THE STATE COMPOSITION AND THE REPORT OF
Contract of the later was a series of the	and the second second	- innerit-				 With Million and Annual and Particle Workshops and a Physics and Art Particle Providence. 	The second s	and the party of the second second
Security of the second second						A REAL PROPERTY AND A REAL		PROPERTY AND A DESCRIPTION OF A DESCRIPR
And I ARRANGED AND A DESCRIPTION OF THE OWNER.		- market					C I DATE OF THE REPORT OF THE	Call State of the
and the second se		-						THE REPORT OF THE REPORT OF THE PARTY OF THE
and other from the last state of the last				- internet				And the set of the set
		1000						And a second sec
and the second s	-		-	_		And the second s		CALL CONTRACTOR

Mapping output of DRR311743:

	20	100	10000	Page 10	1000	100	1000	0	and .	
		12-1	-44							
	. +	1840	_					· · · · · · · · · · · · · · · · · · ·		and in the second secon
Carbon and Parket State and Annual State State	- 16-	1400		10.4	in	-	- #-	 Additional south for president south for the second se second second sec		OTHER DESIGNATION OF THE OTHER
install in this way hit out much he		1945			ia) (1	a (18.	 Provide the second secon	The second	Address of the set of
the state of the second	100	1400		-	4		1.81	 Total Design and the Design and Design and	Interest interest in the second statement of the secon	den and a state of the
The set of the		1865	-					a line was and a suppression of a superior was the		Date of each of his of which it
Control of the Local Distance of the Local D	- 10	-	-					A DESCRIPTION OF THE PROPERTY OF A DESCRIPTION OF A DESCR	provide the terminal control of the set of the strugglost theory in the set	Source which on this or when a
And States of the second states which a summarized states when		-	and a		ia .		+	 The new Designation successive for each or has been been been been been been been bee	Contraction of the second	And the state of the state of the state
And the second s	- *	-	-			•	14	- House & concerns to be a sub-		NAME OF TAXABLE PARTY AND ADDRESS OF TAXABLE PARTY.
Interest of the last star when the composition of the stars.		1000	-			-	10	A Thickness comment was in Research and Research & Rear Works and Re-		Species and a rest of sections.
And the second s	- 8	-	-				1.9	 manufacture of the last test to construct the last test test test to construct the last test test test test test test test t	·	The state of the second second
And the second state and the second second		-440					1.0	The second se	and the second	TALLS NAME AND ADDRESS OF
Name of Street or Street o	-	1					14	· OTHER PROPERTY AND INCOME.	· and a second state in a second state of the	State State of Landson
And the second residence of the second secon	1.8		and -	1.4	-			· provident and the second contract of the second s		And the second second second second
NAMES OF TAXABLE PARTY.	1.0	100	-		- 1		14	· · · · · · · · · · · · · · · · · · ·		Water and the second
1000 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1010 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 10000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1	-	-		1.1	-		1.4	 II		All Control of the Party of the
		-			- 1	• ;	1	· The second state in the second state of the state of the second		And the other strength in the
parate in the last first production and sold		-	-	1.7		-		·	· · · · · · · · · · · · · · · · · · ·	ACTUAL AND ADDRESS THE AN

3.3 Collect alignment before:

We use the tool Collect Alignment Summary Metrics tool take the summary of our mapping done above. Table 1 contains the alignment summary for DRR311742and Table 2 for DRR311743.

Table: 1 Alignment summary for DRR311742

## METRICS CLASSpicard.analysis.AlignmentSummaryMetrics							
CATEGORY TOTAL_READS PF_READS PCT_PF_READS PF_NOISE_READS							
PF_READS_ALIGNED PCT_PF_READS_ALIGNED							
PF_ALIGNED_BASES PF_HQ_ALIGNED_READS PF_HQ_ALIGNED_BASES							
PF_HQ_ALIGNED_Q20_BASES PF_HQ_MEDIAN_MISMATCHES							
PF_MISMATCH_RATE PF_HQ_ERROR_RATE PF_INDEL_RATE							
MEAN_READ_LENGTH READS_ALIGNED_IN_PAIRS							
PCT_READS_ALIGNED_IN_PAIRS PF_READS_IMPROPER_PAIRS							
PCT_PF_READS_IMPROPER_PAIRS BAD_CYCLES							
STRAND_BALANCE PCT_CHIMERAS PCT_ADAPTER SAMPLE							
LIBRARY READ_GROUP							
UNPAIRED 20249149 20249149 1 41 20088298 0.992056							
572908355 18350815 519991244 512894155 0 0.750071							
0.74941 0.000289 59.260725 0 0 0 0 0							
0.499535 0 0.000015							

Table: 2 Alignment summary for DRR311743

CATEGORY TOTAL_REAL				PF_1	NOISE_R	EAD
PF_READS_ALIGNE						
PF_ALIGNED_BASE:	S PF_HQ_AL	IGNED_REAL	S PF_1	IQ AL	IGNED_I	BASE
PF_HQ_ALIGNED_Q	20_BASES PF_I	HQ_MEDIAN_	MISM.	ATCHE	35	
PF_MISMATCH_RAT	E PF HO ER	ROR RATE	PF 1	NDEL	RATE	
MEAN READ LENG					The second s	
PCT_READS_ALIGN	ED IN PAIRS	PF READS	IMPR	OPER	PAIRS	
PCT_PF_READS_IMI					2 - 2 (C) (C)	
STRAND BALANCE				TER	SAMP	LE
LIBRARY READ_				1000	100000000	177.57
UNPAIRED 19632035	9632035 1	36 1948	9580	0.99	2744	
558276194 1784323	52 508531897	501967250	0	0.74	9773	
0.749075 0.00030	3 59.337699	0 0	0	o	0	
0.499767 0 0						

Next, we removed the PCR duplicates using RmDup. Table 3 & 4 contains the Alignment summary for DRR311742 and DRR311743 and removing PCR duplicates.

Table: 3 Alignment summary for DRR311742 post RmDup

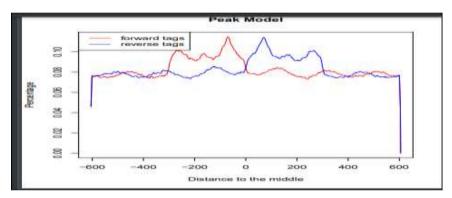
## METRICS CLASSpics	ard analysis .	AlignmentSu	ımmaryN	Aetrics			
CATEGORY TOTAL_R PF_READS_ALIC	NED F	PCT_PF_RE.	ADS_AI	IGNED			ISE_READS
PF_ALIGNED_BA PF HQ ALIGNEI							NED_BASES
PF_MISMATCH_	RATE F	F_HQ_ERR	TOR_RA	TE I	PF_IND		ATE
MEAN_READ_LI PCT_READS_AL						ED DA	TRS
PCT_PF_READS_	IMPROPER	PAIRS	BAD_	CYCLES		- 10 C	
STRAND_BALAN LIBRARY REA			RAS	PCT_A	DAPTE	R	SAMPLE
UNPAIRED 20249149	2024914	9 1	41	2008829	28 0	.9920:	56
572908355 183	50815 5	19991244	512894	155 (o c	0.7500	71
0.74941 0.00	00289 5	9.260725	0	0 0) ()	0
0.499535 0	0.00001:	5					

Table: 4 Alignment summary for DRR311743 post RmDup

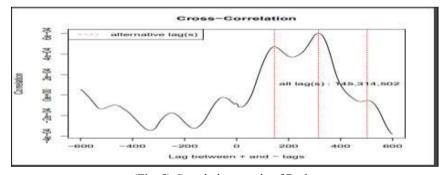
## METRICS CLAS	Spicard.analy	sis AlignmentS	ummaryMetri	cs.		
CATEGORY TOTA			the Constraint of the Constraint of the Constraint of the		PF_I	NOISE_READS
		PCT_PF_RE PF_HQ_ALI			HQ_AL	IGNED_BASE
		BASES PF_H PF_HQ_ERF				
MEAN_REA	D_LENGTH	READS_AL	IGNED_IN_P	AIRS		
PCT_PF_RE.	ADS_IMPRO	IN_PAIRS PER_PAIRS	BAD_CYC	LES		
	ALANCE READ GRO	PCT_CHIMI	ERAS PCT	_ADAI	PTER	SAMPLE
UNPAIRED 19632	035 1963	2035 1	36 1948	39580	0.99	2744
558276194	17843252	508531897	501967250	0	0.74	9773
0.749075	0.000303	59.337699	0 0	0	0	0
0.499767	0 0.000	0015	043795 043795	545895	545795	545795

3.4 MACS2 Callpeak:

As per the alignment summary (Table 3, 4), we see that the reads are less post RmDup which implies that the duplicate reads are removed. Next, we use MAC2 call peak tool to identify areas in the genome that are enriched with the aligned reads. Model-based Analysis of ChIP-Seq (MACS) is a commonly used tool for identifying transcription factor binding sites. The algorithm confines the influence of genome complexity to evaluate the significance of enriched ChIP regions. This tool improves the spatial resolution of binding sites by combining the information of both sequencing tag position and orientation. Here, MACS is used along with a control sample (DRR311742) which increases specificity of the peak calls (Fig. iii). MACS2 models the distance between the paired forward and reverse strand peaks and uses 1000 enriched regions to model the distance between the forward and reverse strand peaks.



(Fig. 4) peak model in graphical format



(Fig. 5) Correlation metric of Peak

3.5 Motif analysis:

We identify the motifs present in our Myelofibrosis genome ChIP-Seq. We used SeqPos motif analysis tool. The file, top 100 most significant peaks in bed format was selected for motif identification (Table 5).

lusters	collapsed_id	Factor	DNA binding domain	hitz	cutoff	zacore	-10*log(pval)	similarity to top	mean_position
1	M01231	GLIS3		46	2.86	-3.046	67.608		-0.152
	MA0024	E2F1	E2F	48	0.965	-2.849	61.239		-0.143
20	M00050	E2F1 E2f1	Fork head / winged helix	54	0.858	-2.598	53.613	0.999	-0.127
4:	M00740	E2F1		36	1.1.02	-2.395	47.896	0.983	-0.146
	M00738	E2F4::TFDP1		51	0.78	-2.338	46.367	0.976	-0.122
з.	MA0145	Tefep211	CP2	52	2.579	-2.712	56,997		-0.133
4	hPDI021	RFXANK		38	4.327	-2.664	55.56		-0.152
5.	M01715	MACI		27	2.364	-2.48	50,257		-0.171
6.	M00469	AP-2alpha/TFAP2A		31	4.772	-2.35	46,672		-0.155
0,	MA0003	TFAP2A	Helix-Loop-Helix	31	4.772	-2.35	46.672	1.0	-0.155

Table –	5	Motif	Ang	lucie
	2	with	лпа	19515.

3.6 SWISS PORT

Genes were selected from the Myelofibrosis genome, model using Swiss prot with Ramachandran plot.

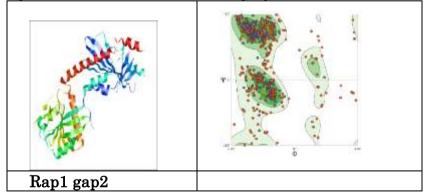
Sl. No	Genes	Name	Accession
			number
1.	ENSMUST00000116495	Rap1gap2	NP_001015046.1
2.	ENSMUST00000183873	Lrrc1	NP_001139520.1
3.	ENSMUST0000058725	Antxr1	NP_766396.2
4.	ENSMUST0000025862	Smarca2	NP_035546.2
5.	ENSMUST00000113531	Lnx1	NP_001346003.1

Table: 6 Genes are involved in Myelofibrosis with Accession number:

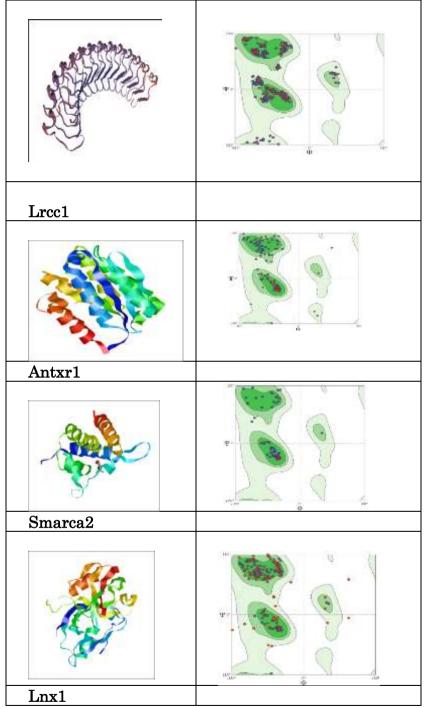
Abbreviations of genes:

- 1. Rap1gap2: Rap1 GTPase activating protein-2
- 2. Lrrc1: Leucine rich repeat containing 1
- 3. Antxr1: Anthrax toxin receptor like
- 4. Smarca2: Sw1/SNF related, matrix associated actin depended of chromatin subfamily
- 5. Lnx1: Ligand of numb protein x1

The receptor model and corresponding Ramachandran plot results are given in Fig.6. Accession number used for modeling is given in table 6.



185



(Fig.6) Swiss-model generated receptor models with their Ramachandran plot.

4. KEGG pathway analysis:

Finally, from the annotation results our peaks table by taking top 100 most significant peaks, we identify the genes overlapping with these peaks which are given in Table 7. The pathways of these genes were identified from KEGG pathways names in Table: 7

Table: 7 Identified Genes and pathways

Sl. No.	GENES	PATHWAYS
1.	NP_001139520 to NM_001146048.1	Tight junctions (TJs)
2.	NP_001015046.1 to NM_001015046.	RAP1 signaling
3.	NP_766396.2 to NM_172808	cGMP - PKG signaling
4.	NP_035546.2 to XM_036161667.1	Thermogenesis
5.	NP_001346003.1 to XM_036164835.1	Tight junctions (TJs)

4.1 Tight Junctions (TJs) Pathway:

Tight junctions (TJs) are essential for establishing a selectively permeable barrier to diffusion through the paracellular space between neighboring cells. TJs are composed of at least three types of transmembrane protein -occluding, claudin and junctional adhesion molecules (JAMs)- and a cytoplasmic 'plaque' consisting of many different proteins that form large complexes. These are proposed to be involved in junction assembly, barrier regulation, cell polarity, gene transcription, and other pathways.

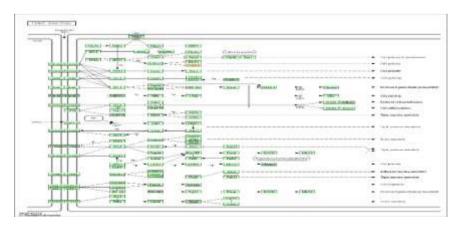


Fig. 7 Tight Junction Pathway

4.2 RAP1 signaling Pathway:

Rap1 is a small GTPase that controls diverse processes, such as cell adhesion, cellcell junction formation and cell polarity. Like all G proteins, Rap1 cycles between an inactive GDP-bound and an active GTP-bound conformation. A variety of extracellular signals control this cycle through the regulation of several unique guanine nucleotide exchange factors (GEFs) and GTPase activating proteins (GAPs). Rap1 plays a dominant role in the control of cell-cell and cell-matrix interactions by regulating the function of integrin's and other adhesion molecules in various cell types. Rap1 also regulates MAP kinase (MAPK) activity in a manner highly dependent on the context of cell types.

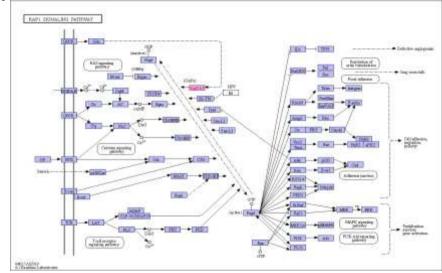


Fig.8 RAP1 signaling Pathway

4.3 cGMP - PKG signaling Pathway:

Cyclic GMP (cGMP) is the intracellular second messenger that mediates the action of nitric oxide (NO) and natriuretic peptides (NPs), regulating a broad array of physiologic processes. The elevated intracellular cGMP level exerts its physiological action through two forms of cGMP-dependent protein kinase (PKG), cGMP-regulated phosphodiesterase's (PDE2, PDE3) and cGMP-gated cation channels, among which PKGs might be the primary mediator. PKG1 isoform-specific activation of established substrates leads to reduction of cytosolic calcium concentration and/or decrease in the sensitivity of myofilaments to Ca2+ (Ca2+-desensitization), resulting in smooth muscle relaxation. In cardiac myocyte, PKG directly phosphorylates a member of the transient potential receptor canonical channel family, TRPC6, suppressing this nonselective ion channel's Ca2+ conductance, G-alpha-q agonist-induced NFAT activation, and myocyte hypertrophic responses. PKG also opens mitochondrial ATP-sensitive K+ (mito-KATP) channels and subsequent release of ROS triggers cardio protection.

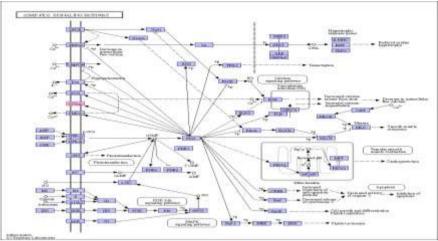


Fig. 10 cGMP - PKG signaling Pathway

4.4 Thermogenesis Pathway:

Thermogenesis is essential for warm-blooded animals, ensuring normal cellular and physiological function under conditions of environmental challenge. Thermogenesis in brown and beige adipose tissue is mainly controlled by norepinephrine, which is released from sympathetic nervous system in response to cold or dietary stimuli. The mitochondrial uncoupling protein 1 (UCP1) is responsible for the process whereby chemical energy is converted into heat in these adipocytes. Activation of these adipocytes leads to an increase in calorie consumption and is expected to improve overweight conditions, providing a potential strategy for treating obesity and its related metabolic disorders

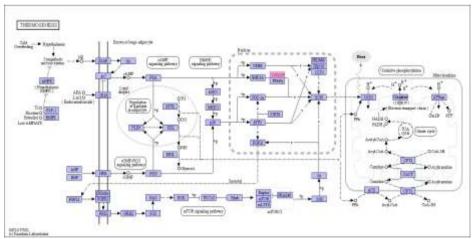


Fig. 10 Thermogenesis Pathway

4.5 Tight Junction (TJs) Pathway:

Tight junctions (TJs) are essential for establishing a selectively permeable barrier to diffusion through the paracellular space between neighboring cells. TJs are composed of at least three types of transmembrane protein -occluding, claudin and junctional adhesion molecules (JAMs)- and a cytoplasmic 'plaque' consisting of many different proteins that form large complexes. These are proposed to be involved in junction assembly, barrier regulation, cell polarity, gene transcription, and other pathways.

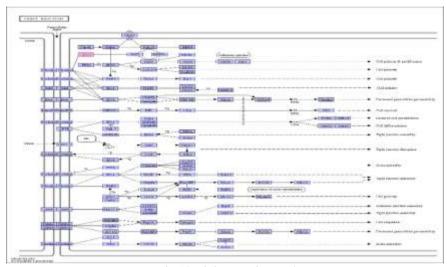


Fig.12 Tight Junction Pathway

1. CONCLUSION

A high-quality immunoprecipitation ChIP peak is detected in our results which imply high concentration of DNA binding proteins. Further, motifs were detected from the ChIP peak. Further, top DNA binding protein genes were identified and corresponding pathways was identified from KEGG pathway. Tight junctions (TJs) pathway, RAP1 signaling pathway, cGMP - PKG signaling pathway, Thermogenesis pathway and Tight junctions (TJs).

6. REFERENCES

1. Deliard S, Zhao J, Xia Q, Grant SF. Generation of high quality chromatin immunoprecipitation DNA template for high-throughput sequencing (ChIP-seq). J Vis Exp. 2013 Apr 19;(74):50286.

2. Furey TS. ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. Nat Rev Genet. 2012;13(12):840-852.

3. Pavesi, Giulio (2016). [Advances in Biochemical Engineering/Biotechnology] || ChIP-Seq Data Analysis to Define Transcriptional Regulatory Networks., (Chapter 43), -. doi:10.1007/10_2016_43.

4. Meyer CA, Liu XS. Identifying and mitigating bias in next-generation sequencing methods for chromatin biology. Nat Rev Genet. 2014;15(11):709-721.

5. Rumi E, Trotti C, Vanni D, Casetti IC, Pietra D, Sant'Antonio E. The Genetic Basis of Primary Myelofibrosis and Its Clinical Relevance. Int J Mol Sci. 2020 Nov 24;21(23):8885. doi: 10.3390/ijms21238885. PMID: 33255170; PMCID: PMC7727658.

6. Mughal TI, Vaddi K, Sarlis NJ, Verstovsek S. Myelofibrosis-associated complications: pathogenesis, clinical manifestations, and effects on outcomes. Int J Gen Med. 2014;7:89-101. Published 2014 Jan 29. doi:10.2147/IJGM.S51800.

7. Zahr AA, Salama ME, Carreau N, Tremblay D, Verstovsek S, Mesa R, Hoffman R, Mascarenhas J. Bone marrow fibrosis in myelofibrosis: pathogenesis, prognosis and targeted strategies. Haematologica. 2016 Jun;101(6):660-71. doi: 10.3324/haematol.2015.141283. PMID: 27252511; PMCID: PMC5013940.

8. J. T. Reilly; D. Barnett; G. Dolan; P. Forrest; J. Eastham; A. Smith (1993). Characterization of an acute micromegakaryocytic leukaemia: evidence for the pathogenesis of myelofibrosis. , 83(1), 58–62. doi:10.1111/j.1365-2141.1993.tb04631.x

9. Tefferi A. Primary myelofibrosis: 2019 update on diagnosis, risk-stratification and management. Am J Hematol. 2018 Dec;93(12):1551-1560. doi: 10.1002/ajh.25230. Epub 2018 Oct 26. PMID: 30039550.

10. ().The Korean Journal of Internal Medicine, 33(4), -. doi:10.3904/kjim.2018.033 11. Ambrosini, G., Dreos, R., Kumar, S. et al. The ChIP-Seq tools and web server: a resource for analyzing ChIP-seq and other types of genomic data. BMC Genomics 2016, 17, 938.

12. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012 Mar 4;9(4):357-9. doi: 10.1038/nmeth.1923. PMID: 22388286; PMCID: PMC3322381.

13. Mark T. W. Ebbert, Mark E. Wadsworth, Lyndsay A. Staley... (2016). Evaluating the necessity of PCR duplicate removal from next-generation sequencing data and a comparison of approaches. BMC Bioinformatics, 17(7 Supplement), –. doi:10.1186/s12859-016-1097-3

14. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. (2008) Model-based Analysis of ChIP-Seq (MACS), Genome Biology, 2008;9 (9):R137.

15. Baxevanis, Andreas D.; Petsko, Gregory A.; Stein, Lincoln D.; Stormo, Gary D. (2002). Current Protocols in Bioinformatics || Using MACS to Identify Peaks from ChIP-Seq Data., (), -. doi:10.1002/0471250953.bi0214s34

16. Bairoch A, Apweiler R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res. 2000 Jan 1;28(1):45-8. doi: 10.1093/nar/28.1.45. PMID: 10592178; PMCID: PMC102476.

17. Hollingsworth SA, Karplus PA. A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins. Biomol Concepts. 2010 Oct;1(3-4):271-283. doi: 10.1515/BMC.2010.022. PMID: 21436958; PMCID: PMC3061398.

18. Mannige, R. V., Kundu, J., & Whitelam, S. (2016). *The Ramachandran Number:* An Order Parameter for Protein Geometry. PLOS ONE, 11(8), e0160023. doi:10.1371/journal.pone.0160023

19. Salma Aouled El Haj Mohamed, Mourad Elloumi and Julie D. Thompson (December 14th 2016). Motif Discovery in Protein Sequences, Pattern Recognition - Analysis and Applications, S. Ramakrishnan, IntechOpen, DOI: 10.5772/65441

20. Altermann, E., Klaenhammer, T.R. PathwayVoyager: pathway mapping using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. BMC Genomics 6, 60 (2005). https://doi.org/10.1186/1471-2164-6-60.

21. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27-30. doi:10.1093/nar/28.1.27

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

