Comparative Methylation and RNA-Seq Expression Analysis in CpG Context to Identify Genes Involved in Backfat vs Liver Diversification in Nanchukmacdon Pig

Devender Arora
NIAS: National Institute of Animal Science
Jong-Eun Park
NIAS: National Institute of Animal Science
Dajeong Lim
NIAS: National Institute of Animal Science
Bong-Hwan Choi
NIAS: National Institute of Animal Science
In-Cheol Cho
Subtropical Livestock Research Institute, National Institute of Animal Science
Krishnamoorthy Srikanth
Cornell University
Jaebum Kim
Konkuk University
Woncheoul Park (wcpark1982@korea.kr)
National Institute of Animal Science https://orcid.org/0000-0003-3140-5628

Research

Keywords: CpG, DMR, DEGs, Differentiation, Methylation, Motif

Posted Date: February 2nd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-164446/v1

Version of Record: A version of this preprint was published at BMC Genomics on November 7th, 2021.
See the published version at https://doi.org/10.1186/s12864-021-08123-x.

Comparative methylation and RNA-seq expression analysis in CpG context to identify genes involved in Backfat vs Liver diversification in Nanchukmacdon Pig

Devender Arora[1], Jong-Eun Park[1], Dajeong Lim[1], Bong-Hwan Choi[1], In-Cheol Cho[2], Krishnamoorthy Srikanth[1,4], Jaebum Kim[3] and Woncheoul Park[1*]

[1]Animal Genomics and Bioinformatics Division, National Institute of Animal Science, RDA, Wanju 55365, Republic of Korea

[2]Subtropical Livestock Research Institute, National Institute of Animal Science, RDA, Jeju 63242, Korea

[3]Affiliation: Department of Biomedical Science and Engineering, Konkuk University, Seoul 05029, Republic of Korea

[4]Department of Animal Science, Cornell University, Ithaca, NY, United States-14853

[*]Corresponding Author

Abstract

Background: DNA methylation and demethylation at CpG island is one of the main regulatory mechanisms at the transcriptional level that give cells the possibility to respond to different stimuli. These regulatory mechanisms help in developing tissue without affecting the genomic composition or undergone selection. Liver and Backfat play important role in regulating lipid metabolism and control various pathways involved in reproductive performance, meat quality, and immunity. Genes inside these tissue stores plethora of information and their understanding are required to enhance tissue characteristics in the future generation.

Results: In this study, to understand the differentiation mechanism we have performed whole-genome bisulfite sequencing (WGBS) and RNA-seq analysis and identified 16 CpG islands were involved in differentially methylation regions (DMRs) as well differentially expressed genes (DEGs) between liver and backfat. Among the identified differentially-methylated genes (*C7orf50, ACTB, MLC1*) in backfat and (*TNNT3, SIX2, SDK1, CLSTN3, LTBP4, CFAP74, SLC22A23, FOXC1, GMDS, GSC, GATA4, SEMA5A, HOXA5*) in the liver were identified. Motif analysis for DMRs was also performed to understand the major role of methylated motif for tissue-specific differentiation. Gene ontology studies revealed the association with collagen fibril organization, BMP signaling pathway in backfat and Cholesterol biosynthesis, bile acid and bile salt transport, immunity-related pathways in methylated genes expressed in the liver.

Conclusion: Our finding could help in understanding how methylation on certain genes plays an important role and can be used as biomarkers to study tissue specific characteristics.

Keywords: CpG; DMR; DEGs; Differentiation; Methylation; Motif

Background

Despite having the same genome a hidden force is governing the gene expression, development, genome imprinting, diseases, diversification, and has been involved in evolutionary changes in different tissues [1, 2]. A single cell at embryonic stages differentiates to form different tissues which could show contrasting physical characteristics with almost unchanged genomic composition governed by DNA methylation [3, 4] (Figure 1). It involves the transfer of methyl group by the addition of a methyl group to the C5 position of cytosine bases in a heritable fashion to form 5-methylcytosine [5, 6]. Recent advances in high-throughput sequencing technologies integrated with bisulfite treatment enable absolute DNA methylation quantification which decodes answers to the potential role of these hidden forces [7, 8]. Cytosine methylation can be categorized into CG, CHG, and CHH methylation (where H refers to either A, C, or T nucleotides) [9]. In the eukaryotic organism, DNA methylation leads to epigenetic modification which at the promoter site leads to curb the transcription process by binding to regulatory protein and primarily occurs in CpG island that is more abundant in the upstream region of the gene [4, 10, 11]. Comparative analysis of methylation in CpG island majorly focused on cross-species comparative analysis and have revealed intriguing trends in both the conserved and divergent features of DNA methylation in eukaryotic evolution [6, 12, 13]. However, it is still unclear whether methylation profiles can help in identifying tissue-specific genes that have any role in influencing tissue-specific features or involvement in biological functions by directing different pathways. Therefore, this created a void in understanding the tissue specific diversification through methylation and gene regulation pattern. Studying tissue specific DNA methylation is a way forward to better understand the genes involved in these process and that could help in understanding overall regulation mechanism and

[10, 14] such phenomena could ultimately provide us better insight to understand the regulatory mechanism of genes in different tissue controlling biological pathways.

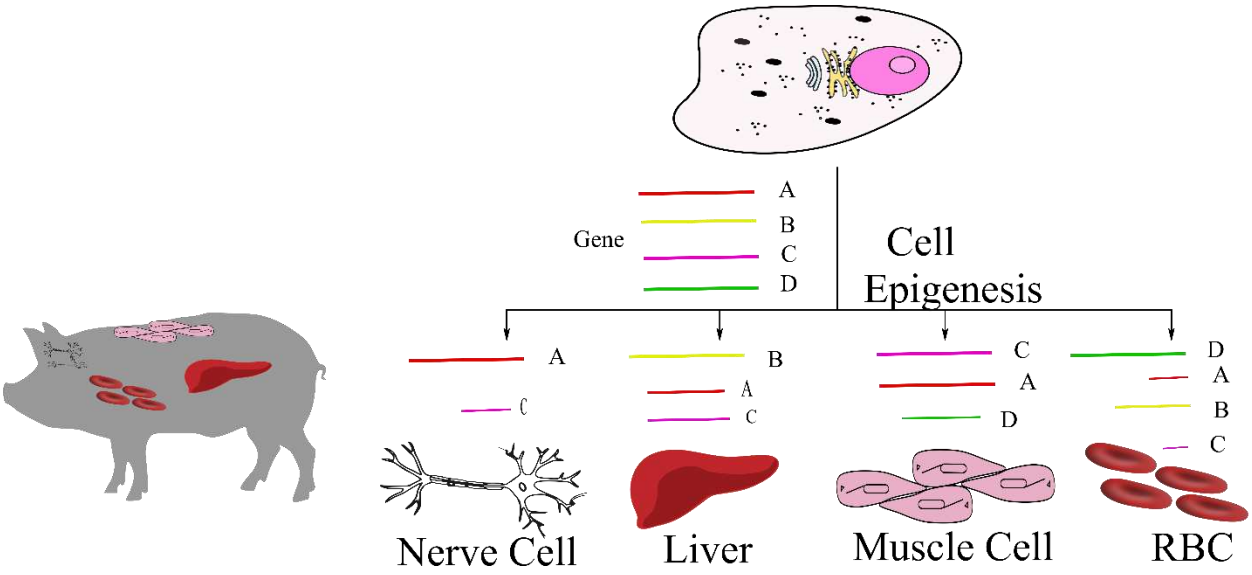


Figure 1: Overview to cell differentiation in to different tissues involving expression of certain genes in one tissues (Highligting gene A,B,C,D) and silent or least expressed in other to govern different pathways required for the development.

Pork is an important food consumed across the world and requires timely effort to monitor and sustain the quality of meat. Several molecular breeding programs are running around the world to understand and to fulfill future requirements with food quality which majorly depends upon consumer preference that ultimately shapes the breeding program by their choice of meat [15, 16]. The Korean peninsula is among one of the high pig-consuming countries and has a huge demand for its Jeju native black pig (JNP) for its superior taste [17, 18]. Due to enhance the taste but low reproduction of JNP a threat of extinction was shadowed over its native JNP breed [19], and to overcome the issue an inbreeding program was conducted to develop a pig breed with a high reproduction rate and sustain the superior taste. In the course of the intensive breeding program

and continued close monitoring using modern biological methods Nanchukmacdon a pig breed was developed with increased fat deposition and metabolism rate and maintained superior characteristics features in generations. The enhanced characteristics displayed by the mixed breed involve the expression of genes and different biological pathways in different tissues that play important role in maintaining the harmony of the cell and development of tissue from single cell [20, 21]. A comparative understanding of tissue diversification is a complex process that involves the expression of certain genes in one tissue while it remains unchanged in another. To understand the hidden forces that led to sustaining such superior characteristics methylation studies in tissue diversification could open a new front in gaining the biological phenomena associated with the new pig breed.

DNA methylation at CpG island does not alter the genomic composition and is one of the main regulatory mechanisms at the transcriptional level that give cells the possibility to respond to different stimuli without going under any mutation and selection [22-24]. These epigenetic mechanisms provide plasticity to the organism and adapt to the different situations by altering the expression pattern of genes to regulates related pathways [6, 11]. While DNA methylation in the mammalian tissue development process is sought to have the conserved process, still understanding of the conversion process at the genome-wide level is at very naïve stages. Understanding these changes requires rigorous analysis at the genome-wide level and recent studies have indicated the role of the methylated region in positively or negatively regulating the gene expression in specific tissue types [25, 26]. Previous studies indicate the role of deposition of backfat is one such aspect associated with growth rate, meat quality, and reproductive performance [27]. Backfat thickness is also considered as one of the main parameters when

97 selecting female pigs into breeding herds since it dominates several reproductive performances
98 [28, 29]. As the liver is also a major organ involved in the regulation of lipid metabolism with
99 fatness and plays a crucial role in animal growth, meat quality, immunity, and reproduction rate.
100 We aimed to understand the tissue-specific methylation in DMRs with emphasis on a parameter
101 such as the hyper-methylated region in a targeted approach to filter out tissue specific
102 diversification and integrated RNA-seq data to gain the understanding of expression pattern in
103 respective methylated region. Along with, we also aimed to evaluate the *de novo* whole genome
104 motif analysis to understand methylated motif and transcription factor binding sites nature in
105 overall changes of tissue and specific pathways.

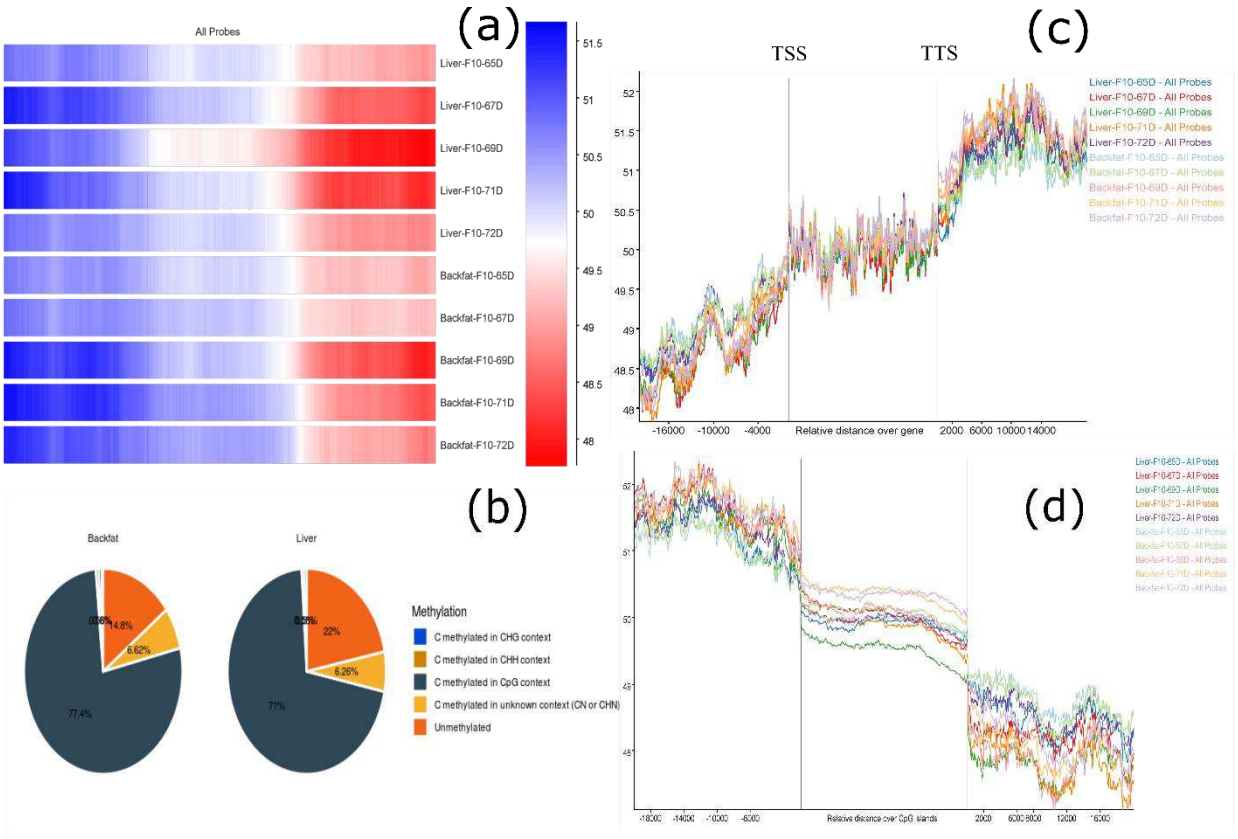
106 In the present work, we reported genes involved in tissue-specific changes at methylation level
107 and the role of gene expression in the regions, we performed WGBS and RNA-seq from (5+5)
108 samples of backfat and liver respectively and integration analysis was undertaken to understand
109 the characteristics tissue. Methylation pattern in CpG island was further studied for their potential
110 role in hyper-methylated region with their respective expression pattern in the specific tissue.
111 RNA-seq studies guided us to decode expression patterns, as well as gene ontology studies, reveals
112 the close association in different biological important pathways that were enriched in different
113 tissue undermethylated conditions.

114 **Results**

115 **WGBS data analysis**

116 WGBS data analysis was performed to compare methylation patterns amongst backfat and liver
117 tissue. Overall mapping of WGBS data on reference genome was ~75% with an average
118 conversion rate in methyl call exceed for reverse and forward (C+T)> 99.4%. Overall methylation

composition was observed inclining towards liver (Figure 2a) with methylation in the CpG context was higher in backfat with 77% and liver with 71% of total methylation (Figure 2b & Additional file: File S1). We have observed a sharp increase at the 2kb region of the TSS region that responsible for the stabilization effect in the relative distance over gene region and again sharply increased and attain stabilization downstream to TTS region Figure 2c). This methylation level remains stable after the promoter region contributing to structural stability and regulation of gene expression. CpG island studies also confirmed and a sharp decrease in methylated CpG level was observed outside of 2kb CpG island (Figure. 2c & 2d). Individual methylation pattern for all the identified genes confirms the pattern of methylation corresponding with the distribution of gene promoters, usually prone to transcription (Additional file: Figures S1). DMR study was to compare the tissue-specific methylation level and de novo motif analysis for TBFS was carried out for backfat vs liver DMRs using the Homer software (Table: 1) (Additional file: Table S1).



131

132 **Figure 2**. **(a)** Heat map was generated for methyl call of each tissue sample and observe the methylation pattern on
133 the overall genome. **(b)** Average methylation composition analysis in context with C methylation in CpG, CHG, CHH,
134 and CN. (H could be A, C, and T nucleotide and N belongs to Unknown) **(c)** Methylation pattern with the relative
135 degree of gene stabilization can be seen and **(d)** sharply increased at TSS region of CpG island and stabilizing
136 afterward.

137

Table 1: Represent the top 5 predicted motif based on rank in the Homer analysis, p-value, %targets, %background, and best match.

Rank	Motif	P-value	% of Targets	% of Background	STD(Bg STD)	Best Match
1		1e-50917	97.64%	73.36%	46.2bp (69.8bp)	AT2G15740(C2H2)
2		1e-2855	13.01%	8.33%	56.2bp (73.5bp)	RFX7
3		1e-1958	10.57%	7.01%	55.8bp (67.3bp)	RAR:RXR(NR)
4		1e-1898	12.13%	8.35%	57.5bp (73.1bp)	RFX3
5		1e-1813	11.64%	8.01%	54.8bp (69.2bp)	MET28

Identification of DEGs, CpG methylation, and Gene ontology

DESeq2 an R package is implemented to identify statistically significant differences in gene expression obtained from featurecount. The overall relationship between backfat and liver was represented in Volcano Plot (Figure. 3a). 2761 in liver and 2375 in backfat DEGs were observed between samples of Nanchukmacdon different tissue (Backfat vs Liver) with Parameter used for DEGs were false discovery rate (FDR) values of ≤ 0.05 and log2FoldChange$\geq \pm 2$.

Lists of DEGs with FDR ≤ 0.05 were compiled and submitted to DAVID v6.8 [30] for functional annotation and enrichment analysis. We divided the dataset into four sets to perform gene ontology studies with hyper-methylated upregulated (729 genes), and downregulated (630 genes) in backfat, hyper-methylated upregulated (792 genes), and downregulated (1032 genes) in liver comprises of total 3183 genes (Additional file: File S2). For each list, enriched gene ontology (GO) Biological Processes (BP), Molecular functions (MF), Cellular Compartments (CC), and KEGG pathway analysis were performed (Additional file: File S3). These terms were then clustered semantically using the ReviGO. Enriched functions throughout the whole transcriptome of Nanchukmacdon with elevated GO-term function and the clustered lower-level GO-terms. The Enriched function with elevated GO term later clustered and corresponds for each GO term found in the treemap (Additional file: Figures S2). We identified the significantly expressed genes related to the KEGG pathway that varies from Metabolic pathway, Fatty acid biosynthesis, ErbB signaling pathway, Adipocytokine signaling pathway, Calcium signaling pathway, and Oxidative phosphorylation are some. CpG island play major role in differentially expression of genes. Methylation at CpG islands have been reported to affect their gene expression. After identification of differentially expressed methylated regions in backfat and liver we retrieved coordinated for all the autosome

chromosomes from UCSC browser and mapped to the identified regions. We have found a total of 16 genes were methylated at CpG island (Table 2).

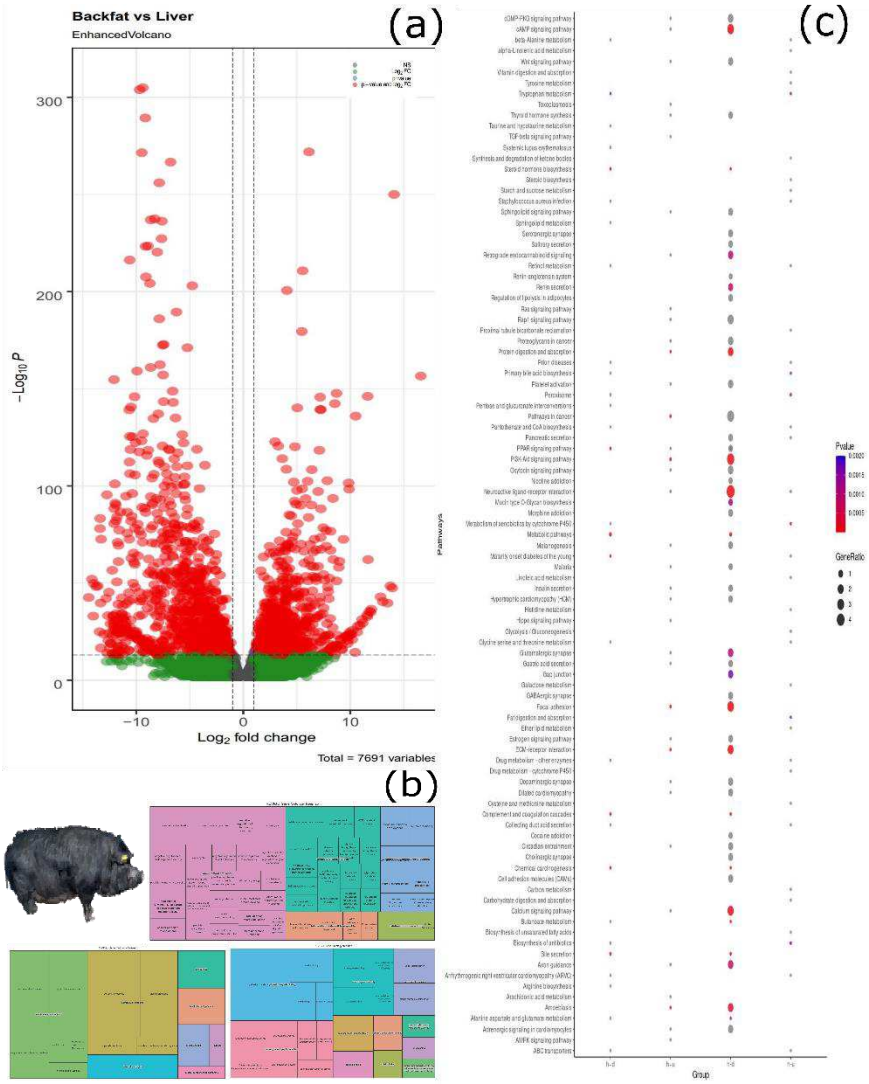


Figure 3. (a) Volcano plot of fold change expression level (y-axis) against $-Log_{10}P$ (x-axis). Each point represents a transcript; those with significant differential expression (FDR ≤ 0.05) are indicated in red. **(b)** Treemap for Gene ontology studies for backfat and liver with BP, MF, and CC. **(c)** KEGG pathway analysis for DEGs with hyper-methylated downregulated liver (h-d), hyper-methylated up-regulated liver (h-u), hyper-methylated downregulated backfat (h+d), and hyper-methylated upregulated backfat (h+u).

Circos plot

Circos plots of all four conditions were generated using CIRCOS tool [31]. The outermost ring represents the 18 autosome chromosomes of *sscrofa*. The second and fourth ring represents the hypermethylated and upregulated genes identified in the DMRs and DEGs for backfat and liver tissues respectively. The third and fifth ring represents the downregulated genes in the methylated regions (Figure 4).

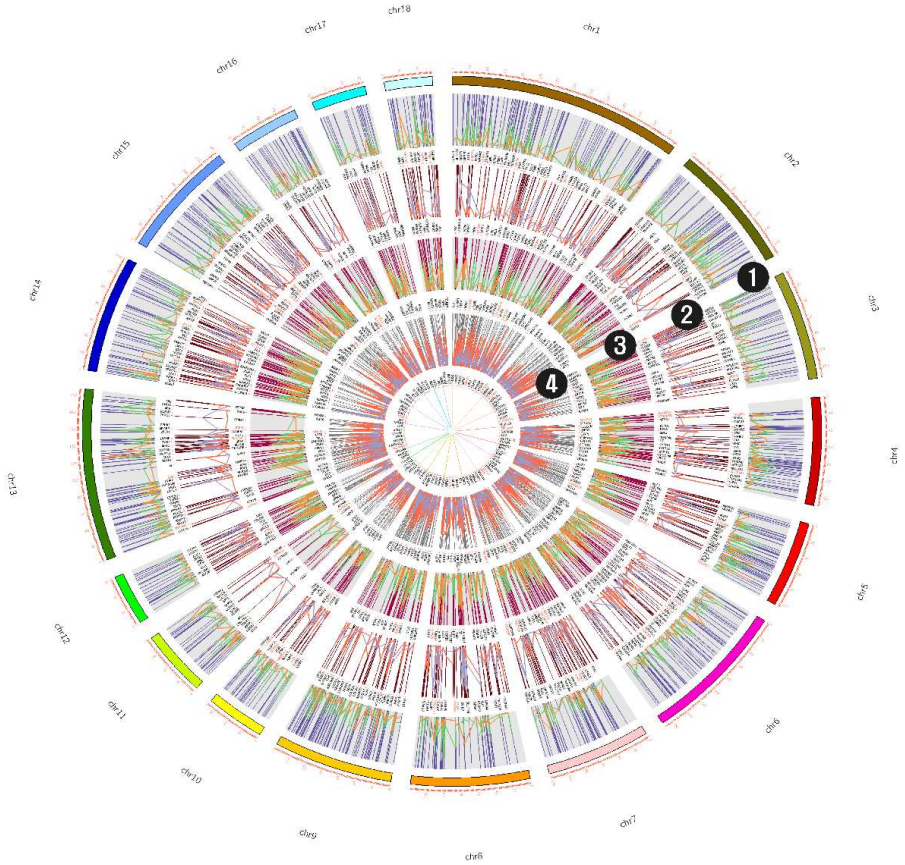


Figure 4: identified regions that were hyper-methylated and gene expression pattern in backfat and liver regions (1 & 3) highlighting hyper-methylation in backfat and liver tissue with their expression pattern. Here green color representing the methylation pattern over the chromosomes and orange represents the upregulated genes in the region and their expression pattern. Similarly, (2 & 4) indicates downregulating genes in backfat and liver hyper-methylated region with dark orange color representing methylated regions and the purple color representing degs belonged in the entire regions.

Table 2: Common genes identified from different conditions.

Ens_Id	chr	CpG	pvalue	padj	meth.diff	log2FoldChange	Gene	Coordinates
ENSSSCG00000032911	2	CpG:_196	1.77E-11	1.3E-10	-30.44324324	-2.326426281	TNNT3	989931-1317600
ENSSSCG00000008446	3	CpG:_73	6.44E-32	1.55E-30	-32.17542336	11.10018784	SIX2	95459937-95464066
ENSSSCG00000007574	3	CpG:_29	1.27E-15	1.28E-14	-27.95608782	4.797926491	SDK1	2814328-3324799
ENSSSCG00000038777	3	CpG:_2584	2.91E-19	3.74E-18	26.54798762	-2.310975194	C7orf50	648140-745331
ENSSSCG00000044546	3	CpG:_268	0.0000031	0.0000134	37.31729323	-2.390040267	ACTB	4091832-4096684
ENSSSCG00000000672	5	CpG:_30	5.58E-19	6.99E-18	-27.95833872	3.098361395	CLSTN3	63572062-63610618
ENSSSCG00000000978	5	CpG:_25	1.09E-10	7.53E-10	40.16694963	-5.798324853	MLC1	571961-591823
ENSSSCG00000033760	6	CpG:_45	2.42E-23	3.81E-22	-41.66461765	3.091738308	LTBP4	48831014-48861507
ENSSSCG00000030513	6	CpG:_22	1.64E-20	2.23E-19	-28.87776243	-3.566656672	CFAP74	63976011-64026767
ENSSSCG00000001004	7	CpG:_113	1.78E-79	1.8E-77	-47.36842105	-5.172667179	SLC22A23	1988695-2131709
ENSSSCG00000039756	7	CpG:_1263	0.000245614	0.000794638	-46.61016949	2.565410851	FOXC1	837171-838805
ENSSSCG00000000994	7	CpG:_48	1.98E-19	2.56E-18	-32.53353973	-2.603381323	GMDS	752239-1285550
ENSSSCG00000002490	7	CpG:_322	7.43E-17	8.2E-16	-26.30769231	4.557359006	GSC	116099047-116100966
ENSSSCG00000022383	14	CpG:_139	0.001072997	0.005378408	-28.33208302	-7.451598322	GATA4	14858159-14939941
ENSSSCG00000017095	16	CpG:_21	5.42E-20	7.2E-19	-29.06597882	3.18365959	SEMA5A	72492516-73329010
ENSSSCG00000016703	18	CpG:_55	3.93E-17	4.43E-16	-41.8356998	4.687258693	HOXA5	45421663-45432885

Discussion

In the present investigation, to understand the role of genes involved in tissue-specific diversification we have presented a comprehensive view with comparative methylation pattern with differentially expressed genes amongst backfat and liver tissue in Nanchukmacdon Pig. Methylation analysis is one of the most promising methods recently evolved used to accurately decode diversification in cross tissue differentiation pattern as well as decode close relationship amongst different tissues. Studying these pattern will ultimately help us in identifying markers that specifically targets breed to enhance tissue of interest. Therefore, we profiled DNA methylation and RNA-seq data for the different tissue and integrated the results to identify genes governing the changes and their involvement in tissue-specific changes led by methylation. Our approach targeted tissue-specific methylation patterns in the CpG context, DMR, and gene expression understanding of each tissue. We have analyzed hyper-methylation differentially expressed regions, motif analysis, and role of CpG island in the DMRs for these changes. Respectively, we performed gene expression analysis and with cutoff FDR$\leq$0.05 and Log2FoldChange $\geq\pm$2, we have identified genes that are expressed in specific tissue types. Finally, we integrated all the data to identify potential genes and regions that are hyper-methylated-upregulated as well as hyper-methylated down-regulated genes in backfat and liver underlying in CpG island and play important role in the tissue-specific diversification. Subsequently, we performed gene ontology studies to gain insight knowledge of the genes involved in each condition.

During tissue-specific comparative analysis, we found C methylation in CpG island of backfat is dominating with 77% and 71% in liver tissue (Figure: 2b) (Additional file: File S1) indicating that the methylation majorly occurred during backfat development which complements by commonly expressed gene and DMRs in the CpG methylation analysis as methylation in CpG

island is necessary to control aberration and in our investigation of comparative analysis common genes in CpG islands with methylation and differentially expressed pattern has limited the total number of genes to 16. Amongst, 13 genes were Hyper-methylated in the liver, and 3 were hyper-methylated in backfat.

We performed DMR analysis for *denovo* methylated regions and found rank 1 motif includes "TATA box" a promoter sequence, which specifies to other molecules where transcription begins and strongly modulates cell- and tissue-specific RANKL expression and osteoclastogenesis process [32]. We have observed a uniform pattern of motif methylation in the highly conserved regulatory factor x genes family which has been reported in the early development and maturation of cells [33] [Table 1]. The top identified motifs were of particular interest, with most motifs were actively involved in upstream binding to transcription factor and regulating cis and epi-cistrome features that regulate DNA landscape [34]. Similarly, the identified motif was found to have a strong association regulatory transcription factor and has been involved in the differentiation process and sought to observe RAR/RXR bound regions are enriched in differentiation regions [35].

Our findings on common genes in CpG islands with methylation and differentially expressed patterns have a limited total number of genes to 16. Amongst, 13 genes were Hyper-methylated in the liver, and 3 were hyper-methylated in backfat. Among the identified genes, *SIX2* is already reported to have involvement in the differentiation process [36]. Methylation in CpG island is necessary to control aberration and to access the impact on gene ontology we have used four different approaches ranges from Hyper-methylated upregulated in backfat and liver, Hyper-methylated down-regulated genes in backfat and liver tissues respectively. KEGG pathway analysis strongly correlated the calcium signaling pathway, fat digestion and absorption, cAMP signaling pathway, etc [Figure 3c] Gene identified downregulated

hypermethylated regions in backfat belongs to complement activation, cholesterol biosynthesis, tissue development, etc. Whereas, the up-regulated genes in hyper-methylated regions were found strongly associated with locomotory behavior, BMP signaling pathways, collagen fibrils development processes. Similarly, genes identified in liver hyper-methylation and upregulated genes were involved in biological important processes that vary from cholesterol biosynthesis, bile acid, and bile salt transport, response to glucose, and immune response mechanism. As well, we have seen, downregulated genes have a role in the embryonic skeletal system, signaling pathways, cell adhesions, etc. Each rectangle in treemap representing a single cluster representative. The representatives are joined into 'superclusters' of loosely related terms, visualized with different colors [Figure 3b & Aditional file 5].

Conclusion: Methylation play important role and understanding gene expression at CpG island in tissue diversification is a potential approach to understand these mechanism. In the present investigation, we have identified common genes highly expressed, and differentially methylated that could be used as potential markers for working in molecular breeding processes and enhancing biologically relevant tissue.

Methods

Preparation of gDNA and Total RNA and Sequencing

We collected tissue samples from the backfat and liver of five Nanchukmacdon pigs. Genomic DNA was isolated using the DNeasy Blood & Tissue Kit (Qiagen, Valencia, CA, USA), and total RNA was isolated using the Trizol method according to manufacturer protocols. The concentrations of DNA and RNA were determined using the Qubit fluorometer (Invitrogen, UK), NanoDrop (Thermo Scientific, USA), and 364 Bioanalyzer (Agilent, UK), and integrity was monitored by agarose gel electrophoresis.

gDNA from Nanchukmacdon backfat and liver was subjected to bisulfite conversion using the fragment size (250bp±25bp), WGBS was performed with MethylMiner Methylated DNA Enrichment kit, and then a sequencing library was constructed using the Illumina Paired-end sequencing on an Illumina, HiSeq2500, 150bpX2. Similarly, RNA-seq data was generated for Nanchukmacdon (N=5) pair-end data after isolation of backfat and liver tissue using TRIzol method following the manufacturer guideline. The sequencing library was constructed using Illumina TruSeq RNA sample preparation kit (Illumina, San Diego, CA, USA).

DMRs and DEGs analysis of WGBS and RNA-seq data

The analysis for WGBS data was performed using reproducible genomics analysis pipeline PiGx-bsseq to understand methylation patterns in identified genes [37]. Where sequence was initially performed for a quality check using trim_galore [38] and alignment were subjected to the filtration of duplicate reads with sam_blaster and sorted using SAMtools [39] afterward mapped to the reference genome of *sscrofa11.1* using Bismark [40]. Bismark methyl extractor was performed to measure the methylation in CpG, CHH, and CHG. Sorting of Bam file was undertaken before running the methylcall with the average conversion rate of >99.4% by applying filters based on a minimum coverage of 10 and a mapping quality of at least 10. Since we were interested in identifying the differential pattern in the respective tissues later performed the DMR studies across backfat and liver using methylKit an R package [41-43]. Logistic regression approach was implemented to model the odd log probability of observing this ratio. False discovery rate ($Q \leq 0.01$) and percent methylation difference larger than 25% were selected and DMRs were extracted.

Similarly, we performed RNA-seq analysis as it becomes the central important feature that enables a comprehensive understanding of the expression pattern of tissue-specific changes in

genes. With statistical advanced tools, we performed the quality check by FastQC to access low-quality pair-end reads [44] and further removed potential adapters using by Trimmomatic tool before sequence alignment [45]. All quality-filtered PE reads were aligned to *Sscrofa* genome (*Sscrofa11.1*) retrieved from the University of California Santa Cruz (UCSC) browser using Hisat2 [46, 47] and reads were counted using FeatureCount [48]. Finally, DESeq2 [49] was used to identify DEGs by setting a cutoff of FDR≤0.05 and log2FoldChange of ±2 for upregulated and downregulated genes.

***De novo* motif discovery**

Hyper-methylated regions were predicted with a cutoff of ±25 in DMRs in backfat and liver. We were interested in understanding the motif for these methylated regions in GC% of CpG island which is found near to transcription start site and performed by findMotifsGenome.pl module of HOMER software at default parameter [50]. Rank-wise motifs were detected with sorted p-value, %target, and %background targets.

Functional enrichment analysis of methylated genes with differentially expressed genes.

After identifying DEGs commonly found in backfat and liver methylated regions with FDR ≤ 0.05 and log2FoldChange ≥±2 were compiled and submitted to DAVID v6.8 [30] for functional annotation and enrichment analysis. For each list, enriched Gene Ontology (GO) studies were performed for Biological Processes, Molecular functions, and Cellular Compartments.These terms were then clustered semantically using the ReviGO server [51] and Clusterprofiler R package [52] were used for summarizing the GO terms.

CpG island and methylation pattern analysis.

Based on DMRs we aimed to identify regions either inclined towards backfat or liver by comparing CpG island coordinates retrieved from UCSC genome browser [53]. A total of

46218 regions were retrieved across the genome by following Table browser with Pig genome of assembly *Sscrofa11.1* as the reference and choose a track for CpG island. The identified island was used to extract DMRs fall in the range and extracted the region of interest that plays a crucial role in tissue diversification.

Supplementary Materials

Additional file: File S1: Cytosine methylation report for backfat and liver.

Additional file: Figure S1: Comparative methylation pattern of identified genes using SeqMonk.

Additional file: Table S1: Output Motif predicted results.

Additional file: File S2: Differentially methylated as well as expressed gene list for backfat and liver.

Additional file: File S3: GO results for Biological process (BP), Molecular function (MF), Cellular compartment, and KEGG pathways.

Additional file: Figure S2: Gene Ontology studies of identified genes in hypermethylation condition w.r.t. backfat and liver.

Abbreviation

WGBS: Whole-Genome Bisulfite Sequencing

DMR: Differentially Methylation Region

DEG: Differentially Expressed Gene

JNP: Jeju Native Black Pig

UCSC: University of California Santa Cruz

GO: Gene Ontology

Conflicts of Interest

The authors declare no conflict of interest.

Availability of data and materials

All data generated or analyzed during this study are included in the supplementary information files or are available from the corresponding author upon request. Statistical Source Data underlying all figures are provided as a separate supplementary files with a tab for each panel generated from source data.

Author's Contributions

D.A. and W.C.P. designed and performed the research, analyzed the data, and wrote the manuscript. J.E.P., D.L., B.H.C., I.C.C., K.S. and J.K. interpreted the results and finalized the manuscript. All authors read and approved the final manuscript.

Funding:

This work was supported by Korea Post-Genome Project (Project title: Deciphering the reference genome and the discovery of trait-associated genes in Nanchukmacdon and mini pigs). Project No. PJ013343 of the National Institute of Animal Science, Rural Development Administration, Republic of Korea.

Ethics approval and consent to participate

The study was approved by National Institute of Animal Science with ethical approval no: **NIAS20181295.**

Consent for publication

Not applicable.

Acknowledgements

NA

References

1. Bartolomei MS, Oakey RJ, Wutz A: **Genomic imprinting: An epigenetic regulatory system**. In.: Public Library of Science San Francisco, CA USA; 2020.
2. Paulsen M, Ferguson-Smith AC: **DNA methylation in genomic imprinting, development, and disease**. *The Journal of pathology* 2001, **195**(1):97-110.
3. Kim M, Costello J: **DNA methylation: an epigenetic mark of cellular memory**. *Experimental & molecular medicine* 2017, **49**(4):e322-e322.
4. Trapnell C: **Defining cell types and states with single-cell genomics**. *Genome research* 2015, **25**(10):1491-1498.
5. Kumar S, Chinnusamy V, Mohapatra T: **Epigenetics of modified DNA bases: 5-methylcytosine and beyond**. *Frontiers in genetics* 2018, **9**:640.
6. Moore LD, Le T, Fan G: **DNA methylation and its basic function**. *Neuropsychopharmacology* 2013, **38**(1):23-38.
7. Li N, Ye M, Li Y, Yan Z, Butcher LM, Sun J, Han X, Chen Q, Wang J: **Whole genome DNA methylation analysis based on high throughput sequencing technology**. *Methods* 2010, **52**(3):203-212.
8. Bibikova M, Lin Z, Zhou L, Chudin E, Garcia EW, Wu B, Doucet D, Thomas NJ, Wang Y, Vollmer E: **High-throughput DNA methylation profiling using universal bead arrays**. *Genome Res* 2006, **16**(3):383-393.
9. Gent JI, Ellis NA, Guo L, Harkess AE, Yao Y, Zhang X, Dawe RK: **CHH islands: de novo DNA methylation in near-gene chromatin regulation in maize**. *Genome research* 2013, **23**(4):628-637.
10. Blake LE, Roux J, Hernando-Herraez I, Banovich NE, Perez RG, Hsiao CJ, Eres I, Cuevas C, Marques-Bonet T, Gilad Y: **A comparison of gene expression and DNA methylation patterns across tissues and species**. *Genome Research* 2020, **30**(2):250-262.
11. Chen X, Schönberger B, Menz J, Ludewig U: **Plasticity of DNA methylation and gene expression under zinc deficiency in Arabidopsis roots**. *Plant and Cell Physiology* 2018, **59**(9):1790-1802.
12. Iguchi-Ariga S, Schaffner W: **CpG methylation of the cAMP-responsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation**. *Genes & development* 1989, **3**(5):612-619.
13. Tate PH, Bird AP: **Effects of DNA methylation on DNA-binding proteins and gene expression**. *Current opinion in genetics & development* 1993, **3**(2):226-231.

14. N'Diaye A, Byrns B, Cory AT, Nilsen KT, Walkowiak S, Sharpe A, Robinson SJ, Pozniak CJ: **Machine learning analyses of methylation profiles uncovers tissue-specific gene expression patterns in wheat**. *The Plant Genome* 2020, **13**(2):e20027.
15. Wallenbeck A, Rydhmer L, Röcklinsberg H, Ljung M, Strandberg E, Ahlman T: **Preferences for pig breeding goals among organic and conventional farmers in Sweden**. *Organic agriculture* 2016, **6**(3):171-182.
16. Xu L, Yang X, Wu L, Chen X, Chen L, Tsai F-S: **Consumers' Willingness to Pay for Food with Information on Animal Welfare, Lean Meat Essence Detection, and Traceability**. *International Journal of Environmental Research and Public Health* 2019, **16**(19):3616.
17. Kim K, Kim D, Min Y, Jeong D, Son YO, Do K: **Myogenic regulatory factors are key players in determining muscle mass and meat quality in Jeju native and Berkshire pigs**. *Veterinary Medicine and Science* 2020.
18. Lee Y-S, Shin D, Won K-H, Kim DC, Lee SC, Song K-D: **Genome-wide scans for detecting the selection signature of the Jeju-island native pig in Korea**. *Asian-Australasian Journal of Animal Sciences* 2020, **33**(4):539.
19. Kim J, Cho S, Caetano-Anolles K, Kim H, Ryu Y-C: **Genome-wide detection and characterization of positive selection in Korean Native Black Pig from Jeju Island**. *BMC genetics* 2015, **16**(1):3.
20. Gonzàlez-Porta M, Frankish A, Rung J, Harrow J, Brazma A: **Transcriptome analysis of human tissues and cell lines reveals one dominant transcript per gene**. *Genome biology* 2013, **14**(7):1-11.
21. Martínez O, Reyes-Valdés MH: **Defining diversity, specialization, and gene specificity in transcriptomes through information theory**. *Proceedings of the National Academy of Sciences* 2008, **105**(28):9709-9714.
22. Deaton AM, Webb S, Kerr AR, Illingworth RS, Guy J, Andrews R, Bird A: **Cell type–specific DNA methylation at intragenic CpG islands in the immune system**. *Genome research* 2011, **21**(7):1074-1086.
23. Herb BR, Shook MS, Fields CJ, Robinson GE: **Defense against territorial intrusion is associated with DNA methylation changes in the honey bee brain**. *BMC genomics* 2018, **19**(1):216.
24. Morales-Nebreda L, McLafferty FS, Singer BD: **DNA methylation as a transcriptional regulator of the immune system**. *Translational Research* 2019, **204**:1-18.

25. Jiang L, Zhang M, Ma K: **Whole-Genome DNA Methylation Associated With Differentially Expressed Genes Regulated Anthocyanin Biosynthesis Within Flower Color Chimera of Ornamental Tree Prunus mume**. *Forests* 2020, **11**(1):90.
26. Wan J, Oliver VF, Wang G, Zhu H, Zack DJ, Merbs SL, Qian J: **Characterization of tissue-specific differential DNA methylation suggests distinct modes of positive and negative gene expression regulation**. *BMC genomics* 2015, **16**(1):1-11.
27. Xing K, Zhu F, Zhai L, Liu H, Wang Z, Hou Z, Wang C: **The liver transcriptome of two full-sibling Songliao black pigs with extreme differences in backfat thickness**. *Journal of animal science and biotechnology* 2014, **5**(1):1-9.
28. Kim JS, Yang X, Pangeni D, Baidoo SK: **Relationship between backfat thickness of sows during late gestation and reproductive efficiency at different parities**. *Acta Agriculturae Scandinavica, Section A—Animal Science* 2015, **65**(1):1-8.
29. Roongsitthichai A, Tummaruk P: **Importance of backfat thickness to reproductive performance in female pigs**. *The Thai Journal of Veterinary Medicine* 2014, **44**(2):171-178.
30. Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA: **DAVID: database for annotation, visualization, and integrated discovery**. *Genome biology* 2003, **4**(9):1-11.
31. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA: **Circos: an information aesthetic for comparative genomics**. *Genome research* 2009, **19**(9):1639-1645.
32. Kitazawa R, Kitazawa S: **Methylation status of a single CpG locus 3 bases upstream of TATA-Box of receptor activator of nuclear factor-κB ligand (RANKL) gene promoter modulates cell-and tissue-specific RANKL expression and osteoclastogenesis**. *Molecular Endocrinology* 2007, **21**(1):148-158.
33. Elkon R, Milon B, Morrison L, Shah M, Vijayakumar S, Racherla M, Leitch CC, Silipino L, Hadi S, Weiss-Gayet M: **RFX transcription factors are essential for hearing in mice**. *Nature communications* 2015, **6**(1):1-14.
34. O'Malley RC, Huang S-sC, Song L, Lewsey MG, Bartlett A, Nery JR, Galli M, Gallavotti A, Ecker JR: **Cistrome and epicistrome features shape the regulatory DNA landscape**. *Cell* 2016, **165**(5):1280-1292.

35. Bourguet W, Vivat V, Wurtz J-M, Chambon P, Gronemeyer H, Moras D: **Crystal structure of a heterodimeric complex of RAR and RXR ligand-binding domains**. *Molecular cell* 2000, **5**(2):289-298.
36. Zhang M, Wang C, Jiang H, Liu M, Yang N, Zhao L, Hou D, Jin Y, Chen Q, Chen Y: **Derivation of novel naive-like porcine embryonic stem cells by a reprogramming factor-assisted strategy**. *The FASEB Journal* 2019, **33**(8):9350-9361.
37. Wurmus R, Uyar B, Osberg B, Franke V, Gosdschan A, Wreczycka K, Ronen J, Akalin A: **PiGx: reproducible genomics analysis pipelines with GNU Guix**. *Gigascience* 2018, **7**(12):giy123.
38. Krueger F: **Trim galore**. *A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files* 2015, **516**:517.
39. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R: **The sequence alignment/map format and SAMtools**. *Bioinformatics* 2009, **25**(16):2078-2079.
40. Krueger F, Andrews SR: **Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications**. *bioinformatics* 2011, **27**(11):1571-1572.
41. Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, Mason CE: **methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles**. *Genome biology* 2012, **13**(10):1-9.
42. Klambauer G, Schwarzbauer K, Mayr A, Clevert D-A, Mitterecker A, Bodenhofer U, Hochreiter S: **cn. MOPS: mixture of Poissons for discovering copy number variations in next-generation sequencing data with a low false discovery rate**. *Nucleic acids research* 2012, **40**(9):e69-e69.
43. Lawrence M, Gentleman R, Carey V: **rtracklayer: an R package for interfacing with genome browsers**. *Bioinformatics* 2009, **25**(14):1841-1842.
44. Andrews S: **FastQC: a quality control tool for high throughput sequence data. 2010**. In.; 2017.
45. Bolger AM, Lohse M, Usadel B: **Trimmomatic: a flexible trimmer for Illumina sequence data**. *Bioinformatics* 2014, **30**(15):2114-2120.
46. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL: **Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype**. *Nature biotechnology* 2019, **37**(8):907-915.
47. Kim D, Langmead B, Salzberg SL: **HISAT: a fast spliced aligner with low memory requirements**. *Nature methods* 2015, **12**(4):357-360.

48. Liao Y, Smyth GK, Shi W: **featureCounts: an efficient general purpose program for assigning sequence reads to genomic features**. *Bioinformatics* 2014, **30**(7):923-930.
49. Love M, Anders S, Huber W: **Differential analysis of count data–the DESeq2 package**. *Genome Biol* 2014, **15**(550):10.1186.
50. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK: **Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities**. *Molecular cell* 2010, **38**(4):576-589.
51. Supek F, Bošnjak M, Škunca N, Šmuc T: **REVIGO summarizes and visualizes long lists of gene ontology terms**. *PloS one* 2011, **6**(7):e21800.
52. Yu G, Wang L-G, Han Y, He Q-Y: **clusterProfiler: an R package for comparing biological themes among gene clusters**. *Omics: a journal of integrative biology* 2012, **16**(5):284-287.
53. Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, Kent WJ: **The UCSC Table Browser data retrieval tool**. *Nucleic acids research* 2004, **32**(suppl_1):D493-D496.

Figures

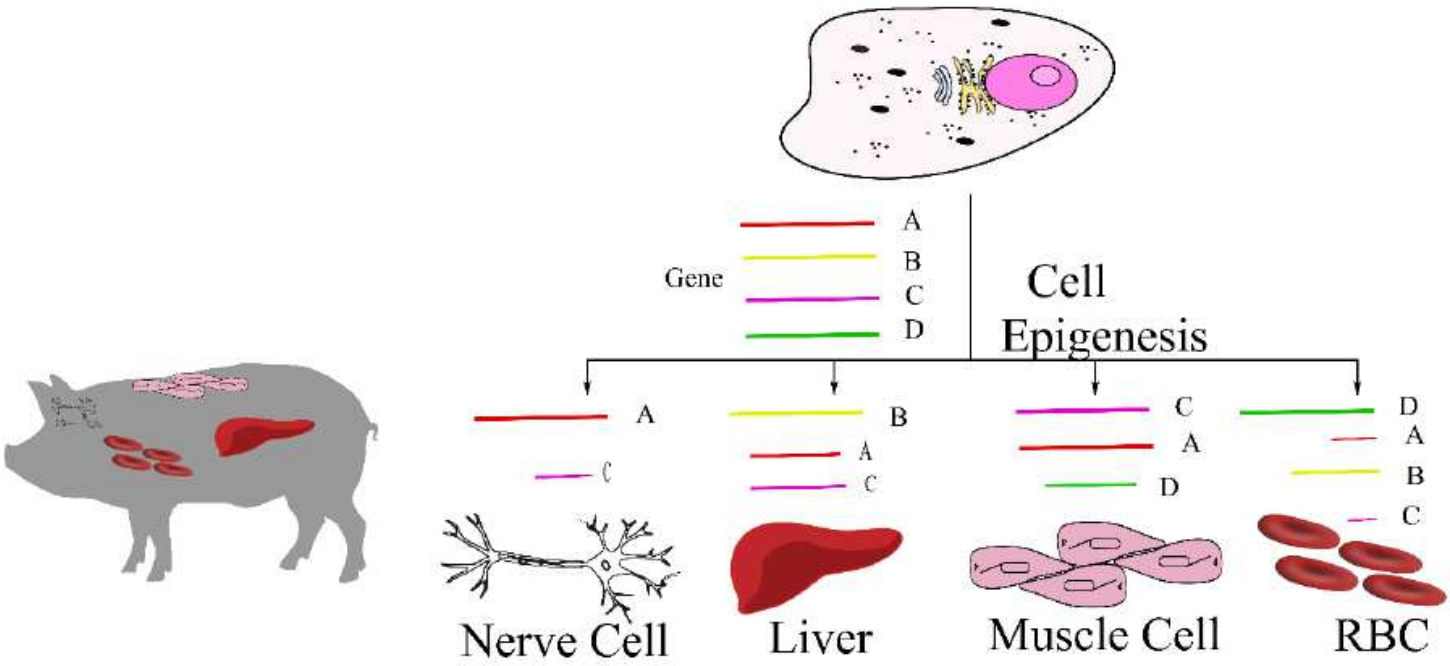


Figure 1

Overview to cell differentiation in to different tissues involving expression of certain genes in one tissues (Highligting gene A,B,C,D) and silent or least expressed in other to govern different pathways required for the development.

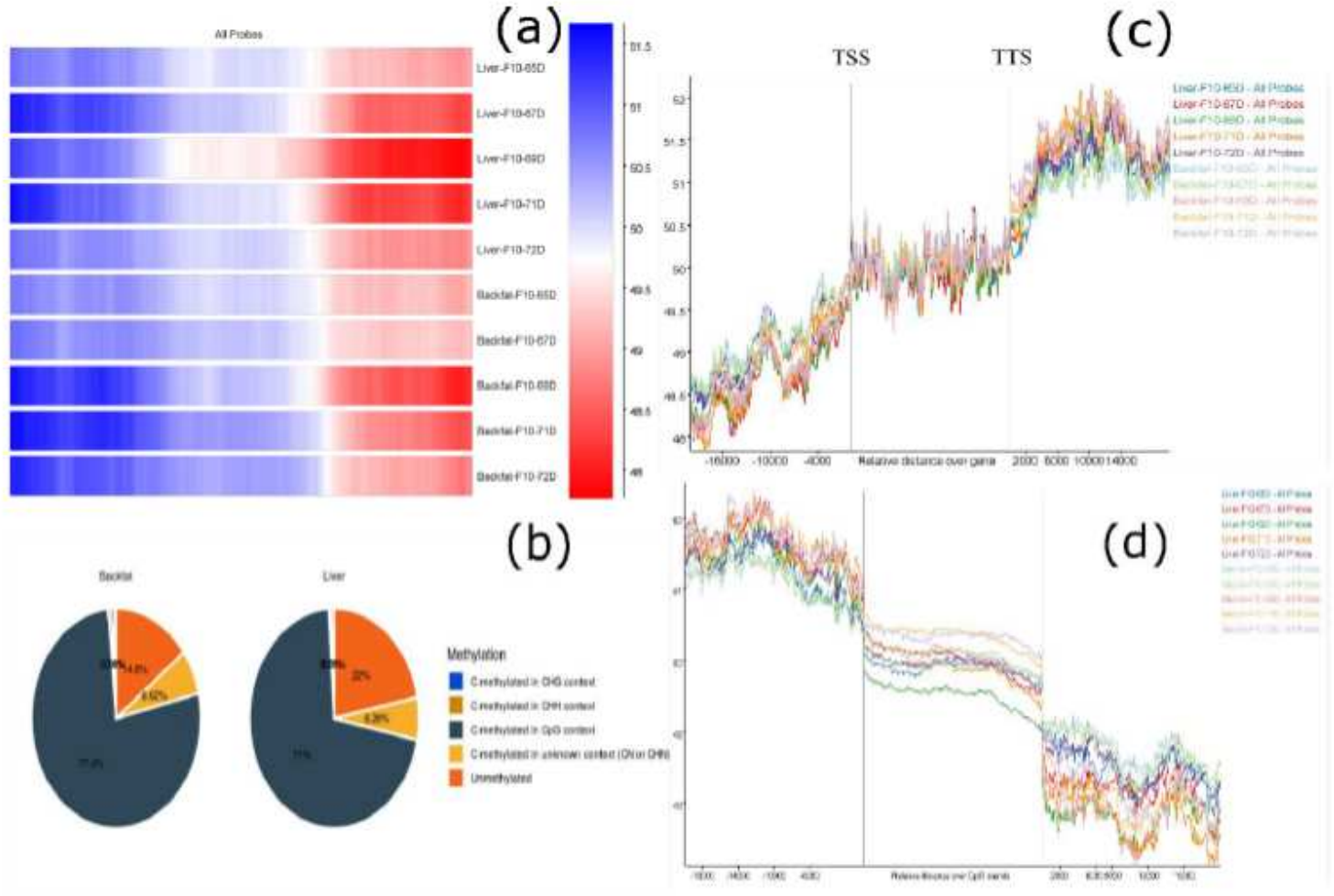


Figure 2

(a) Heat map was generated for methyl call of each tissue sample and observe the methylation pattern on the overall genome. (b) Average methylation composition analysis in context with C methylation in CpG, CHG, CHH, and CN. (H could be A, C, and T nucleotide and N belongs to Unknown) (c) Methylation pattern with the relative degree of gene stabilization can be seen and (d) sharply increased at TSS region of CpG island and stabilizing afterward.

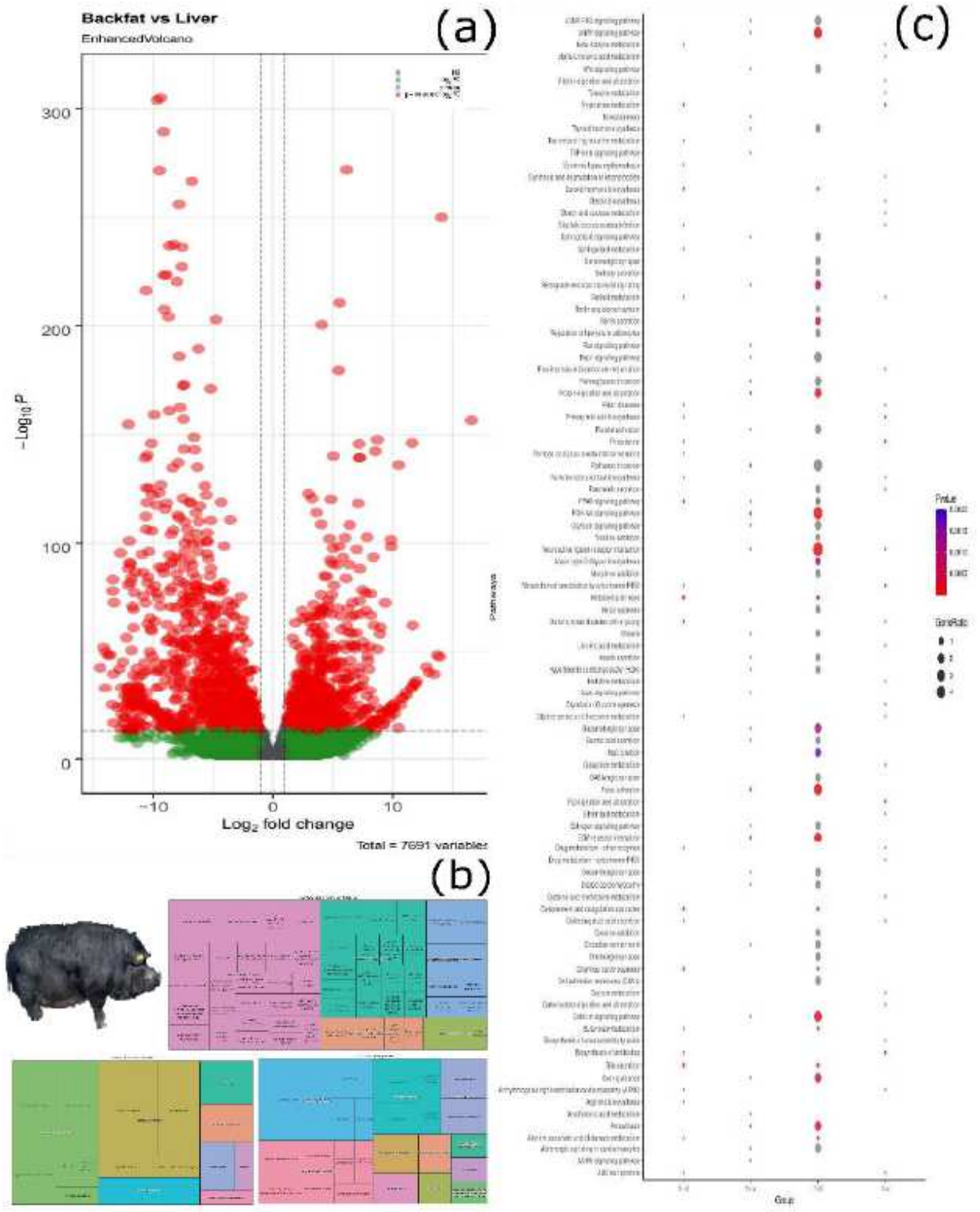


Figure 3

(a) Volcano plot of fold change expression level (y-axis) against –Log10P (x-axis). Each point represents a transcript; those with significant differential expression (FDR ≤ 0.05) are indicated in red. (b) Treemap for Gene ontology studies for backfat and liver with BP, MF, and CC. (c) KEGG pathway analysis for DEGs with hyper-methylated downregulated liver (h-d), hyper-methylated up-regulated liver (h-u), hyper-methylated downregulated backfat (h+d), and hyper-methylated upregulated backfat (h+u).

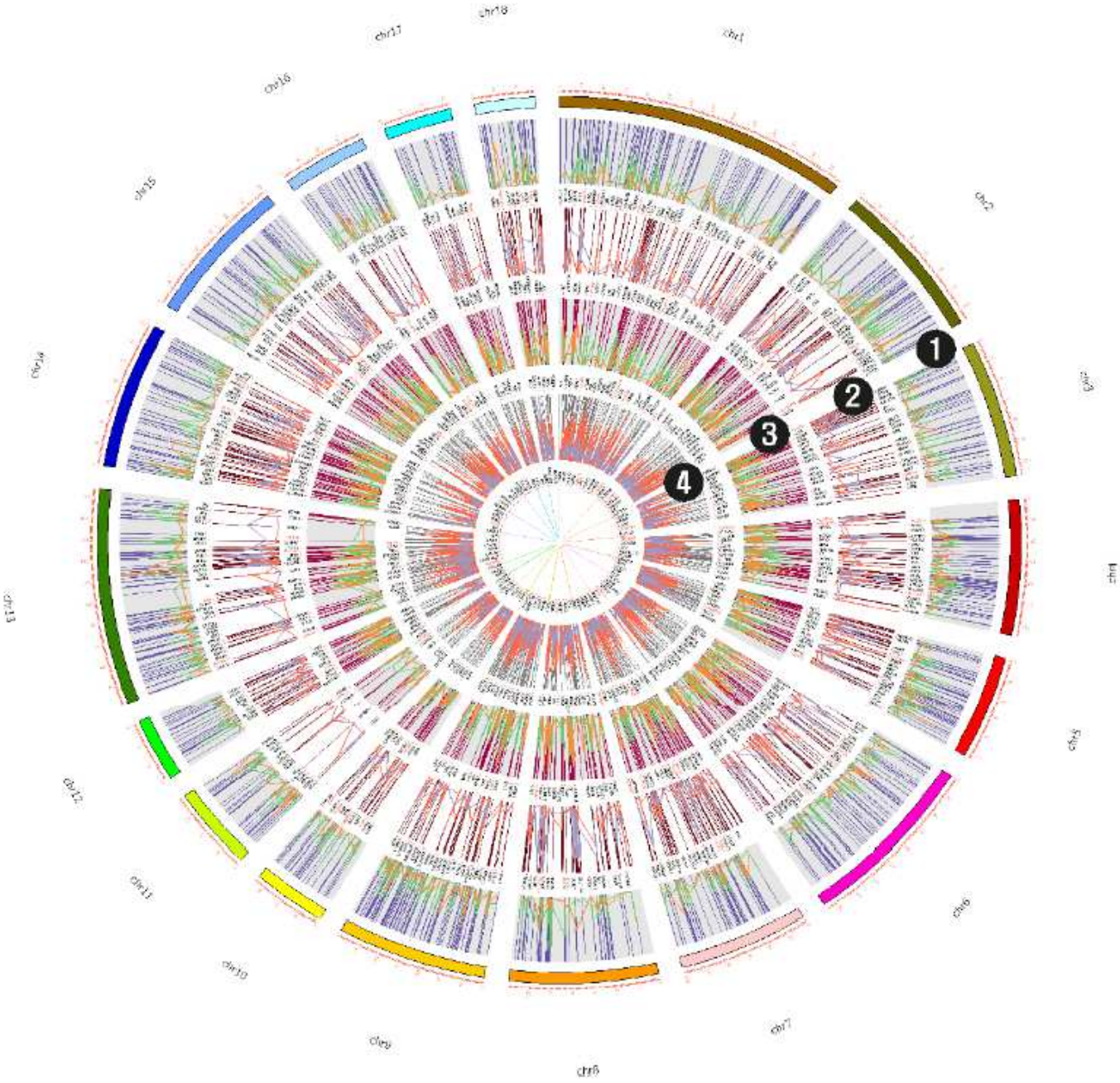


Figure 4

Identified regions that were hyper-methylated and gene expression pattern in backfat and liver regions (1 & 3) highlighting hyper-methylation in backfat and liver tissue with their expression pattern. Here green color representing the methylation pattern over the chromosomes and orange represents the upregulated genes in the region and their expression pattern. Similarly, (2 & 4) indicates downregulating genes in backfat and liver hyper-methylated region with dark orange color representing methylated regions and the purple color representing degs belonged in the entire regions.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AdditionalfileFiguresS1.docx
- AdditionalfileFiguresS2.docx
- AdditionalfileFileS1.xlsx
- AdditionalfileFileS2.xlsx
- AdditionalfileFileS3.xlsx
- AdditionalfileTableS1.docx