Chromatin state distribution of residue-specific histone acetylation in early myoblast differentiation

Qiao Li (✉ qli@uottawa.ca)

University of Ottawa

Yuan Li

University of Ottawa

Saadia Khilji

University of Ottawa

Jihong Chen

University of Ottawa

Short Report

Keywords: Epigenetics, gene regulation, histone acetyltransferase, histone acetylation, stem cell differentiation, myogenesis

Posted Date: February 21st, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1374891/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Chromatin state distribution of residue-specific histone acetylation in early myoblast differentiation

Yuan Li¹, Saadia Khilji¹, Jihong Chen², Qiao Li^{1,2*}

¹Department of Cellular and Molecular Medicine Faculty of Medicine, University of Ottawa, Ottawa, Ontario, Canada

²Department of Pathology and Laboratory Medicine, Faculty of Medicine, University of Ottawa, Ottawa, Ontario, Canada

* Correspondence:

Qiao Li

qli@uottawa.ca

Keywords:

Epigenetics, gene regulation, histone acetyltransferase, histone acetylation, stem cell differentiation, myogenesis

Abstract

Dynamic changes in epigenetic landscape reflect a critical command of lineage-specific gene expression. In an effort to discern the epigenetic regulatory networks of myogenic differentiation, we have used systematic and integrated approaches to explore multi-omics datasets on global myogenic gene expression, histone acetylation and acetyltransferase occupancy in view of distinct chromatin states. In this brief report, we present a comprehensive assessment regarding the experimental design, quality control, filtering steps and data processing. We also define a gene-level overlap between RNA-seq and ChIP-seq datasets through integrated analyses to offer strategies for future use of the data. Furthermore, our analyses provide a blueprint of chromatin states associated with residue-specific histone acetylation, histone acetyltransferase p300-concomitant epigenetic landscape in committed skeletal myoblasts and differential histone acetylation signatures in early myoblast differentiation. These datasets can be further utilized to delineate the function of muscle-specific regulatory elements governed by other muscle myogenic regulators or signaling molecules.

27 Introduction

28 Epigenetic regulatory mechanisms affect the patterns of global gene expression underpinning lineage
 29 commitment and differentiation of stem cells, and thus impact tissue homeostasis including skeletal
 30 muscle [1–4]. Recent advances in large-scale functional genomics permit the use of omics
 31 approaches to delineate the regulatory roles of epigenetic modifiers in cell fate decision [5]. As such,
 32 we have conducted ChIP-seq of different histone acetylation marks and histone acetyltransferase
 33 (HAT) p300 with condition matching RNA-seq in differentiating and proliferating myoblasts to
 34 examine the coupling of epigenetic landscapes with myogenic transcriptome in early myoblast
 35 differentiation [6–8].

36 Different regulatory elements are marked by defined patterns of chromatin modifications, in
 37 that H3K9ac, H3K18ac, and H3K27ac are often found near transcription start sites (TSS), while
 38 H4K8ac is mostly associated with promoters and gene bodies that are actively transcribed [9–11]. In
 39 addition, H3K4me1 together with H3K27ac are generally considered as marks of active enhancers,
 40 while H3K4me1 alone or in accompany with H3K27me3 indicates a poised enhancer [12, 13].
 41 Besides histone modifications, enhancers are marked also by the occupancy of p300 which is a
 42 transcriptional co-activator containing an intrinsic HAT activity [14, 15]. While p300 occupancy is
 43 associated to both promoters and enhancers [6, 7, 16], it is well recognized as a signature of
 44 enhancers targeting mainly H3K18 and H3K27 for acetylation [17–19].

45 Partaking in the control of expression of a vast number of genes, p300 is known to be
 46 associated with the function of muscle-specific transcription factors and essential for myogenic
 47 differentiation [20, 21]. Underscoring a critical role of p300 in myogenic fate decision is that
 48 embryonic stem cells derived from p300 knockout mice do not express muscle regulatory factors
 49 MyoD and Myf5 and knockdown of p300 impedes myogenic differentiation [22, 23].

50 Cell type specific chromatin states can be defined by large-scale analyses of co-occurrence of
 51 different histone marks [24, 25]. We have previously established a 14-state chromatin model based
 52 on a genome-wide co-occurrence of various histone modifications to study the regulatory
 53 mechanisms of myogenic differentiation [8]. To this end, we used differentiating myoblasts with
 54 proliferating myoblasts to obtain H4K8ac, H3K9ac, H3K18ac, H3K27ac and p300 ChIP-seq datasets
 55 for survey of global epigenetic landscapes in early myoblast differentiation. These ChIP-seq datasets
 56 were generated in matching conditions of our RNA-seq to allow coupling of epigenetic landscape
 57 with global gene expression profiles. Through comprehensive analyses of the omic datasets and
 58 integration with myoblast chromatin states, we have determined the impact of p300 and muscle

master regulator MyoD on myogenic enhancers and defined a loci-specific enrichment of H4K8 and H3K9 acetylation [7]. In addition, we have identified the chromatin state signature associated with myogenin function at myogenic promoters [6]. Our analyses reveal models of histone acetylation augmentation at distinctive regulatory elements albeit a global decrease in histone acetylation at the onset of myoblast differentiation.

Since epigenetic modifiers and transcriptional regulators are essential for myogenic differentiation and their dysfunction is implicated in pathogenesis of skeletal muscles, a better grasp of epigenetic regulatory mechanisms of the myoblast genome is paramount. In this report, we provide technical validation and comprehensive analyses of 14 genomic datasets which profile global gene expression, histone acetylation, and p300 occupancy for an unbiased and genome-wide view of chromatin states associated with myoblast differentiation. For an in-depth usage of the datasets, we provide detailed information on the experimental design, quality control filtration, and integrated data analyses. Our objective is to offer strategies for future integrative analyses of the datasets to advance the knowledge of epigenetic regulatory networks of the myoblast genome in a specific signaling pathway or pathological context. Moreover, comprehensive analyses of epigenomes and transcriptomes in differentiating myoblasts also provide a platform to identify additional histone acetylation signatures of myoblasts genome.

RNA-seq Data Processing

Total RNA was extracted from C2C12 myoblasts using RNeasy kit (Qiagen) according to the manufacturer's instruction [8]. RNA samples from two biological repeats were processed by the McGill University Genome Quebec Innovation Centre for transcriptome library construction and sequenced using an Illumina HiSeq 2000 sequencer as single-end 50 base pair reads. Salmon was used to perform mapping-based estimation of transcript abundance guided by transcripts from the NCBI37.67 database. Gene level counts were subsequently obtained through the tximport package which estimates the counts and transcript lengths for downstream gene-level analysis.

Count tables were then imported into integrated Differential Expression and Pathway (iDEP) [26], where genes with less than 0.5 counts per million (CPM) in all samples were filtered out. Of the 30,099 Ensembl gene IDs, 14,484 genes passed the filter and were processed via a regularized log (rlog) transformation which minimizes the differences between samples with small counts and normalizes with respect to library size. K-means clustering and gene ontology enrichment of each cluster were also performed by iDEP. Gene lists with expression values were subsequently uploaded into Ingenuity Pathway Analysis (IPA - Qiagen) for upstream regulator analysis. This tool integrates

upstream regulator results with downstream differential gene expression in the dataset to predict significant causal interactions of modifiers responsible for particular phenotypic or functional outcomes.

ChIP-seq Data Processing

Since transcriptomes are regulated by epigenetic regulatory networks, we thus analyzed ChIP-seq of histone marks including H4K8ac, H3K9ac, H3K18ac, and H3K27ac. We also examined the genomic association of p300 as it is a critical HAT in myogenic differentiation [21, 22, 27]. ChIP-seq library preparation and sequencing were performed by the McGill University Genome Quebec Innovation Centre with Illumina HiSeq 2000 as single-end 50 bp reads [8]. Sequence reads were aligned to the mm9 reference genome using Bowtie2, with subsequent removal of duplicate reads by Picard tools. Bedgraph files were created following read extension by 125 bp at the 3' end with the BEDTools suite. Hypergeometric Optimization of Motif EnRichment (HOMER) was used to call peaks with options -region and -style factor for p300 or -style histone for histone acetylation marks.

Resulting peak files were used as input into HOMER for gene annotation and *de novo* motif discovery with input sequences of 200 bps surrounding the peak center. The annotatePeaks tool from HOMER was utilized to plot motif enrichment across input loci. ChIP enrichment at TSSs and p300-associated sites was calculated with ngs.plot, which calculates the coverage vectors for each region based on specified ChIP-seq alignment files. Following normalization and transformation on the coverage, a heatmap as well as an average profile is plotted from the mean of all regions, normalized to the total number of mapped reads in millions. The average PhastCons score plots were generated with the Conservation Plot tool as part of the Cistrome Analysis Pipeline and Integrative Genomics Viewer (IGV) was used for visualization of ChIP-seq signals.

Over View of Experimental Design

In this study, we examined myogenic expression and histone acetylation profiles on a genome-wide scale using C2C12 myoblasts as a system of skeletal myoblast differentiation. These myoblasts are myogenic precursors of mesenchymal origin and represent a well-established model of myogenic differentiation. They switch from a proliferation to differentiation state under low mitogen conditions, followed by a fusion process to form multinucleated myotubes and have been a model of choice for many genome-wide studies due to a high correlation of their transcription profiles with primary myoblasts [28, 29]. They are also less prone to spontaneous differentiation than primary myoblasts, and thus offer a more homogenous population for large-scale studies [4, 30].

For our study, the C2C12 myoblasts were differentiated for 24 hours (Fig. 1) to capture early transcriptional dynamics, since thousands of genes are differentially expressed within 24 hours of differentiation [8, 31, 32]. Additionally, our analyses focused upon epigenetic changes at the onset of differentiation compared to other studies that examined that of terminal differentiation [4]. Since histone modifications are known to be transient [33, 34], an earlier time point permits the coupling of epigenetic modifiers with differential gene expression to reveal regulatory networks critical to differentiation initiation. As such, we performed ChIP-seq of H4K8ac, H3K9ac, H3K18ac, H3K27ac, H3K27me3, and p300 in matching condition of RNA-seq to capture epigenetic changes of the myoblast genome (Fig. 1). Analyses of these datasets in view of other signaling pathways or transcription factor association should allow new biological insights into the epigenetic dynamics of myogenic differentiation. Here, we provide the characterization of the datasets with respect to general read quality, genome alignment, ChIP enrichment, replicate correlation, and integrated analyse.

RNA-seq Quality Control

To assess the quality of RNA-seq datasets for differentiating and proliferating myoblasts, base call scores were analyzed using the FastQC program (Fig. 2A). The quality scores across reads for all data sets fell within the high confidence range (base quality score of 30-40). Moreover, an inspection of the average GC content across reads revealed an overall normal distribution representative of a standard random library (Fig. 2B). We also assessed the percentage of unique and duplicate reads across all datasets (Fig. 2C). Theoretically, the number of reads mapping to a certain transcript reflect the number of molecules of the transcript in the sample and thus duplicate reads in RNA-seq can represent highly-expressed genes where deduplication of samples can lead to bias in true expression measurements [35]. As shown in figure 2C, the average duplication rate across all samples was approximately 40%, below the normal duplication rate cut-off of 50%, indicating sufficient starting material and library complexity.

Next, the RSeQC gene body coverage tool was used to examine uniformity of read coverage across the gene body for indications of any 5'/3' biases. Figure 2D exhibits relatively even coverage of reads across the gene body with approximately equal percentage of reads that map near the 5' and 3' end. Previous studies have demonstrated an increased 3' mapping bias with decreasing RNA integrity [36, 37]. Thus, uniform gene body coverage in all samples is one indicator of suitable quality of the starting RNA material. Interestingly, despite a decrease in the number of reads in differentiating myoblasts compared to proliferating myoblasts, we observed even gene coverage across all conditions (Fig. 2D).

In RNA-seq, distribution of reads across known gene features is an additional important measure of experimental validity. Most reads are expected to map to exons compared to introns and intergenic regions [38]. As shown in supplementary figure 1E, the exonic rate across all samples was similar at about 60% with an intronic rate of about 35% (Table 1). The relatively high number of intronic reads are likely a result of the library construction method of using random priming rather than poly-A selection [38]. Nonetheless, the majority of reads mapped to exonic regions, warranting for applications of differential expression analyses.

In addition to detecting novel transcripts, a principal benefit of RNA-sequencing has been replicability between samples, which was measured as the correlation of gene expression between samples (Fig. 2F). Spearman correlation analysis revealed greater similarity between replicates for differentiating and proliferating myoblasts than across conditions. Additionally, the correlation between replicates was approximately 0.95, indicating highly satisfactory technical reproducibility. Overall, our RNA-seq datasets are good resources to guide the validation and functional characterization of early myogenic gene expression and to identify regulatory networks associated with specific signaling pathways.

ChIP-seq Quality Control

The average base quality scores for all ChIP-seq datasets in the present study fell within the high confidence range (Fig. 3A) and displayed an overall normal distribution for the average GC content of reads (Fig. 3B). Furthermore, raw sequence reads were aligned to the mm9 build of the mouse genome with an average mapping rate of 96% (Fig. 3C), representing an average of 36 million total aligned reads across samples (Table 2). Subsequent ChIP-seq analysis was performed with uniquely aligned reads to eliminate the obscurity of reads that align to multiple locations. An average of 30 million uniquely aligned reads were retained across all ChIP-seq samples, displaying a mean unique alignment rate of 83% (Table 2).

Following the confirmation of satisfactory alignment metrics, we also examined genome coverage and ChIP enrichment simultaneously through fingerprint plot analysis (Fig. 3D). Intersection of the x-axis at approximately 0.1 by the fingerprint plot traces of all samples indicates an absence of read coverage across 10% of the genome. Given that the rightward deflection of a trace indicates the extent of ChIP enrichment, all datasets reflect ChIP enrichment in comparison to input. As an example, following the trace for H3K27ac, the intersection of the dashed lines indicates that 10% of the genome is enriched with approximately 70% of all uniquely aligned reads, suggesting

185 strong enrichment in the dataset. Fingerprint plots for other samples showed similar patterns of
186 genome coverage and ChIP enrichment.

187 **Chromatin State Model**

188 We have previously utilized multiple marks of histone acetylation, histone methylation and RNA-Pol
189 II to establish a model of 14 chromatin states in committed myoblasts [8], by using a hidden Markov
190 model-based method [24]. These chromatin states are categorized as promoters, enhancers, and
191 “others” that include genomic regions which are either actively transcribed, or inactive/repressed.
192 The enhancer states are classified by the presence of H3K4me1 and absence of H3K4me3 and
193 H3K36me3, and further divided into active or poised in accordance with the levels of histone
194 acetylation. The promoter states are identified by enrichments of H3K4me3, transcription start sites
195 (TSSs), RNA Pol II occupancy and levels of histone acetylation. This model was used in this report
196 to index loci of histone acetylation and p300 localization into distinct chromatin states of committed
197 myoblasts.

198 **Chromatin States Associated with Residue-Specific Histone Acetylation**

199 To delineate histone acetylation signature associated with myogenic expression, we sought to
200 integrate RNA-seq and ChIP-seq datasets after quality verification of the raw sequencing data. First,
201 PCA plots were used to assess the clustering of histone acetylation marks into their respective
202 biological groups for chromatin modification and differentiation status (Fig. 4A). The plot showed
203 close clustering of H3K18ac and H3K27ac signals in comparison to H4K8ac and H3K9ac which
204 were distinct from one another. Since H3K18ac and H3K27ac enrichment in addition to p300
205 occupancy are key features that denote enhancers [13, 18], the close correlation of these marks
206 suggests a co-regulation of myogenic enhancers during the differentiation process. Furthermore, the
207 inter-modification differences were greater than the variation between proliferation and
208 differentiation myoblasts as expected.

209 Next, we examined the chromatin state distribution of residue-specific histone acetylation in
210 early differentiation utilizing our established chromatin state model. As shown in figure 4B,
211 H3K18ac and H3K27ac had a more pronounced association with the enhancers (69% and 61%,
212 respectively) in proliferating myoblasts, whereas H4K8ac and H3K9ac displayed a clear association
213 with poised promoters (11% and 9%, respectively). More interestingly, there was an evident increase
214 in the distribution of acetylation marks to poised enhancers in differentiating myoblasts compared to
215 proliferating myoblasts, with the largest increase particularly for H3K18ac and H3K27ac signals, to

16% and 9%, respectively (Fig. 4B). Nonetheless, the distribution of H4K8ac and H3K9ac to promoters were similar in both proliferating and differentiation myoblasts (Fig. 4B). Established in the progenitors, poised enhancers are marked by distinctive histone modification and important for the control of gene expression required for lineage specific differentiation [39]. Thus, an increased distribution of histone acetylation to poised enhancers reflects the involvement of residue-specific histone acetylation in the activation of discrete gene programs.

Histone acetylation at p300-associated loci in early myoblast differentiation.

To delineate epigenetic networks involved in p300-mediated gene expression and the contribution of p300 HAT activity to myogenic differentiation, we also examined histone acetylation profiles associated with p300 peaks at large (Fig. 5A). Interestingly, p300-associated sites were generally coupled with an enrichment of H3K18ac and H3K27ac in differentiating myoblasts as compared to proliferating myoblasts (Fig. 5A). Using k-means clustering, we correlated the levels of histone acetylation to 5 distinct clusters centered at the p300 peaks in differentiation myoblasts (Fig. 5B). An increase in histone acetylation during the transition from proliferation to differentiation was parallel to a significant increase in p300 enrichment signals in p300 cluster 4 (Fig. 5B, C). Similarly, the expression of gene associated with this cluster were significantly increased in differentiating myoblasts compared to proliferating myoblasts as determined by RNA-seq analysis (Fig. 5D). Interestingly, the p300 clusters exhibited a distinct pattern of chromatin state distribution where p300 peaks in cluster 4 and 5 were clearly associated with the poised enhancers (Fig. 5E). Moreover, the correlation of an increased level of histone acetylation with parallel p300 enrichment signals in differentiating myoblasts is well exhibited by an IGV snapshot (Fig. 5F) at the putative regulatory locus of *Tbc1d1*, a glucose regulator in skeletal muscle [40].

Conclusion

Integrated analyses of p300 ChIP-seq data with chromatin state association and corresponding transcriptome offer an invaluable avenue for identifying chromatin state distribution of residue-specific histone acetylation, and discerning the distinct function of HATs in epigenetic landscape and gene regulatory networks in a specific biological milieu. The high throughput sequencing datasets presented here are biologically validated and in high quality, and can be used for future studies to determine the roles of muscle regulatory factors, reissue-specific histone acetylation and p300 function in early myogenic differentiation, particularly in terms of their impact on chromatin dynamics.

247 **Declarations:**

248 **Ethics approval and consent to participate**

249 Not applicable.

250 **Consent for publication**

251 Not applicable.

252 **Availability of data and materials**

253 RNA-seq FPKM_tracking files as well as ChIP-seq bedgraph files for histone acetylation marks have
254 been deposited in the NCBI Gene Expression Omnibus (GEO) SuperSeries under accession number
255 GSE94561. Bedgraph and peak files for p300 have been deposited under accession number
256 GSE109636. GEO linked FASTQ and bam files were also deposited in the Sequence Read Archive
257 (SRA). Tables 1 and 2 contain the associated database identification numbers for each dataset.

258 **Competing interests**

259 The authors declare that the research was conducted in the absence of any commercial or financial
260 relationships that could be construed as a potential conflict of interest.

261 **Funding**

262 This research was supported by an Operating Grant from the Natural Sciences and Engineering
263 Research Council of Canada to Q.L. (NSERC #250174).

264 **Author contributions**

265 Y.L. and S.K, data analysis and manuscript preparation. J.C., and Q. L. conception and finalizing the
266 manuscript. All authors reviewed the manuscript.

267 **Acknowledgements**

268 We thank our colleagues for a very supportive research environment. Y.L is supported by an
269 Exchange Study Scholarship from Southeast University of China. S.K. is a recipient of Canada
270 Graduate Scholarship from the NSERC.

References

1. Robinson DCL, Dilworth FJ (2018) Epigenetic Regulation of Adult Myogenesis. pp 235–284
2. Sincennes M-C, Brun CE, Rudnicki MA (2016) Concise Review: Epigenetic Regulation of Myogenesis in Health and Disease. *STEM CELLS Translational Medicine* 5:282–290
3. Barreiro E, Tajbakhsh S (2017) Epigenetic regulation of muscle development. *Journal of Muscle Research and Cell Motility* 38:31–35
4. Asp P, Blum R, Vethantham V, Parisi F, Micsinai M, Cheng J, Bowman C, Kluger Y, Dynlacht BD (2011) Genome-wide remodeling of the epigenetic landscape during myogenic differentiation. *Proceedings of the National Academy of Sciences*.
<https://doi.org/10.1073/pnas.1102223108>
5. Gan Q, Yoshida T, McDonald OG, Owens GK (2007) Concise Review: Epigenetic Mechanisms Contribute to Pluripotency and Cell Lineage Determination of Embryonic Stem Cells. *STEM CELLS* 25:2–9
6. Khilji S, Hamed M, Chen J, Li Q (2020) Dissecting myogenin-mediated retinoid X receptor signaling in myogenic differentiation. *Communications Biology*.
<https://doi.org/10.1038/s42003-020-1043-9>
7. Khilji S, Hamed M, Chen J, Li Q (2018) Loci-specific histone acetylation profiles associated with transcriptional coactivator p300 during early myoblast differentiation. *Epigenetics*.
<https://doi.org/10.1080/15592294.2018.1489659>
8. Hamed M, Khilji S, Dixon K, Blais A, Ioshikhes I, Chen J, Li Q (2017) Insights into interplay between retinoid signaling and myogenic regulatory factor-associated chromatin state in myogenic differentiation. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkx800>
9. Wang Z, Zang C, Rosenfeld JA, et al (2008) Combinatorial patterns of histone acetylations and methylations in the human genome. *Nature Genetics* 40:897–903
10. Li B, Carey M, Workman JL (2007) The Role of Chromatin during Transcription. *Cell* 128:707–719
11. Rajagopal N, Ernst J, Ray P, Wu J, Zhang M, Kellis M, Ren B (2014) Distinct and Predictive Histone Lysine Acetylation Patterns at Promoters, Enhancers, and Gene Bodies. *G3 Genes|Genomes|Genetics*. <https://doi.org/10.1534/g3.114.013565>
12. Creyghton MP, Cheng AW, Welstead GG, et al (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences* 107:21931–21936
13. Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J (2011) A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* 470:279–283
14. Ait-Si-Ali S, Polesskaya A, Filleur S, Ferreira R, Duquet A, Robin P, Vervish A, Trouche D, Cabon F, Harel-Bellan A (2000) CBP/p300 histone acetyl-transferase activity is important for the G1/S transition. *Oncogene* 19:2430–2437
15. Kung AL, Rebel VI, Bronson RT, Ch'ng LE, Sieff CA, Livingston DM, Yao TP (2000) Gene dose-dependent control of hematopoiesis and hematologic tumor suppression by CBP. *Genes & development* 14:272–7
16. Wang Z, Zang C, Cui K, Schones DE, Barski A, Peng W, Zhao K (2009) Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. *Cell* 138:1019–31
17. Heintzman ND, Stuart RK, Hon G, et al (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genetics*.
<https://doi.org/10.1038/ng1966>

- 318 18. Visel A, Blow MJ, Li Z, et al (2009) ChIP-seq accurately predicts tissue-specific activity of
319 enhancers. *Nature* 457:854–858
- 320 19. Jin Q, Yu L-R, Wang L, Zhang Z, Kasper LH, Lee J-E, Wang C, Brindle PK, Dent SYR, Ge K
321 (2011) Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac
322 in nuclear receptor transactivation. *The EMBO journal* 30:249–62
- 323 20. Yao TP, Oh SP, Fuchs M, Zhou ND, Ch'ng LE, Newsome D, Bronson RT, Li E, Livingston
324 DM, Eckner R (1998) Gene dosage-dependent embryonic development and proliferation
325 defects in mice lacking the transcriptional integrator p300. *Cell* 93:361–372
- 326 21. Puri PL, Avantaggiati ML, Balsano C, Sang N, Graessmann A, Giordano A, Levrero M (1997)
327 p300 is required for MyoD-dependent cell cycle arrest and muscle-specific gene transcription.
328 *The EMBO journal* 16:369–383
- 329 22. Roth JF, Shikama N, Henzen C, Desbaillets I, Lutz W, Marino S, Wittwer J, Schorle H,
330 Gassmann M, Eckner R (2003) Differential role of p300 and CBP acetyltransferase during
331 myogenesis: p300 acts upstream of MyoD and Myf5. *EMBO Journal*.
332 <https://doi.org/10.1093/emboj/cdg473>
- 333 23. Chen J, Wang Y, Hamed M, Lacroix N, Li Q (2015) Molecular Basis for the Regulation of
334 Transcriptional Coactivator p300 in Myogenic Differentiation. *Scientific Reports*.
335 <https://doi.org/10.1038/srep13727>
- 336 24. Ernst J, Kellis M (2012) ChromHMM: automating chromatin-state discovery and
337 characterization. *Nature methods* 9:215–6
- 338 25. Ernst J, Kheradpour P, Mikkelsen TS, et al (2011) Mapping and analysis of chromatin state
339 dynamics in nine human cell types. *Nature* 473:43–49
- 340 26. Ge SX, Son EW, Yao R (2018) iDEP: an integrated web application for differential expression
341 and pathway analysis of RNA-Seq data. *BMC Bioinformatics* 19:534
- 342 27. Polesskaya A (2001) CBP/p300 and muscle differentiation: no HAT, no muscle. *The EMBO*
343 *Journal* 20:6816–6825
- 344 28. Blais A, Tsikitis M, Acosta-Alvear D, Sharan R, Kluger Y, Dynlacht BD (2005) An initial
345 blueprint for myogenic differentiation. *Genes & development* 19:553–69
- 346 29. Cao Y, Yao Z, Sarkar D, et al (2010) Genome-wide MyoD Binding in Skeletal Muscle Cells:
347 A Potential for Broad Cellular Reprogramming. *Developmental Cell* 18:662–674
- 348 30. Burattini S, Ferri P, Battistelli M, Curci R, Luchetti F, Falcieri E C2C12 murine myoblasts as
349 a model of skeletal muscle development: morpho-functional characterization. *European*
350 *journal of histochemistry : EJH* 48:223–33
- 351 31. Doynova MD, Markworth JF, Cameron-Smith D, Vickers MH, O'Sullivan JM (2017)
352 Linkages between changes in the 3D organization of the genome and transcription during
353 myotube differentiation in vitro. *Skeletal Muscle* 7:5
- 354 32. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold
355 BJ, Pachter L (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated
356 transcripts and isoform switching during cell differentiation. *Nature Biotechnology* 28:511–
357 515
- 358 33. Levenson JM, O'Riordan KJ, Brown KD, Trinh MA, Molfese DL, Sweatt JD (2004)
359 Regulation of Histone Acetylation during Memory Formation in the Hippocampus. *Journal of*
360 *Biological Chemistry* 279:40545–40559
- 361 34. Halder R, Hennion M, Vidal RO, et al (2016) DNA methylation changes in plasticity genes
362 accompany the formation and maintenance of memory. *Nature Neuroscience* 19:102–110
- 363 35. Parekh S, Ziegenhain C, Vieth B, Enard W, Hellmann I (2016) The impact of amplification on
364 differential expression analyses by RNA-seq. *Scientific Reports* 6:25533
- 365 36. Choi I, Bao H, Kommadath A, et al (2014) Increasing gene discovery and coverage using
366 RNA-seq of globin RNA reduced porcine blood samples. *BMC Genomics* 15:954

- 367 37. Sigurgeirsson B, Emanuelsson O, Lundeberg J (2014) Sequencing Degraded RNA Addressed
368 by 3' Tag Counting. PLoS ONE 9:e91851
- 369 38. Hrdlickova R, Toloue M, Tian B (2017) RNA-Seq methods for transcriptome analysis. Wiley
370 interdisciplinary reviews RNA. <https://doi.org/10.1002/wrna.1364>
- 371 39. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H,
372 Glass CK (2010) Simple Combinations of Lineage-Determining Transcription Factors Prime
373 cis-Regulatory Elements Required for Macrophage and B Cell Identities. Molecular Cell
374 38:576–589
- 375 40. An D, Toyoda T, Taylor EB, Yu H, Fujii N, Hirshman MF, Goodyear LJ (2010) TBC1D1
376 Regulates Insulin- and Contraction-Induced Glucose Transport in Mouse Skeletal Muscle.
377 Diabetes 59:1358–1365
378
379

380 **Figure legends**

381 **Figure 1. Overview of experimental design and computational analysis workflow.**

382 The diagram displays the procedures used for RNA-seq and ChIP-seq analyses in C2C12 myoblasts
 383 differentiated for 24 hours with proliferating myoblasts as controls. The analyses include quality
 384 control, correlation between samples, differential enrichment or expression analyses, and integration
 385 of gene expression with localized ChIP enrichment, as described in methods.

386 **Figure 2. Quality control of RNA-seq datasets.**

387 (A) distribution of base quality scores is shown for two biological replicates of RNA-seq of C2C12
 388 myoblasts following 24 hours of differentiation (DM1 and DM2) with proliferating myoblasts as
 389 controls (GM1 and GM2). (B) The distribution of GC content across the reads. (C) The percentage
 390 of unique and duplicate reads from all mapped reads in each of the dataset. (D) The distribution of
 391 read coverage across the gene body. (E) The alignment of reads to known gene features in the
 392 reference genome (CDS, coding sequence; TSS, transcription start site; TES, transcription end site).
 393 (F) Heatmap of Spearman correlation coefficients calculated between $\log_2$ - transformed, nonzero
 394 FPKM expression values in both replicates of proliferating and differentiating myoblasts.

395 **Figure 3. Quality control of ChIP-seq datasets.**

396 (A) The distribution of base quality scores for ChIP-seq of the histone acetylation marks in C2C12
 397 myoblasts differentiated for 24 hours (DM) and proliferating myoblasts (GM). (B) The average GC
 398 content of reads. (C) The percentage of reads that mapped to the reference genome are displayed.
 399 (D) ChIP-enrichment analysis plot that displays the cumulative percentage of total reads found in a
 400 given percentage of the mappable reference genome.

401 **Figure 4. Chromatin state associated with residue specific histone acetylation.**

402 (A) C2C12 myoblasts were differentiated for 24 hours (DM) and subjected to ChIP-seq with
 403 proliferating myoblasts (GM) as control. Shown is the PCA plot for histone acetylation ChIP-seq
 404 datasets. (B) The distribution of H4K8ac, H3K9ac, H3K18ac, and H3K27ac enrichment signals to
 405 distinct chromatin states.

406 **Figure 5. Histone acetylation associated with p300 peaks in early myoblast differentiation.**

407 (A) C2C12 myoblasts were differentiated for 24 hours (DM) and subjected to ChIP-seq with
 408 proliferating myoblasts (GM) as control. Shown are the profiles of average ChIP-seq enrichment

409 signals for p300 and histone acetylation at the sites of p300 occupancy. (B) The p300 sites in
410 differentiating myoblasts were used to classify the enrichment of p300 and accompanying histone
411 marks into 5 clusters with k-means clustering. The read density is presented $\pm$ 1 kb from the peak
412 center. (C) Quantification of p300 enrichment signals was plotted as $\log_2$ -fold change in
413 differentiating myoblasts relative to proliferating myoblasts. (D) Gene expression levels from the
414 RNA-seq are presented as FPKM (Fragments Per Kilobase of transcript per Million mapped reads)
415 for genes associated to each p300 cluster. (E) The p300 sites in each cluster were classified into
416 distinct chromatin states. (F) Genome browser snapshot displays the read density of p300 occupancy
417 and histone acetylation at the *Tbc1d1* locus. Blue bars show the Refseq gene position and the colours
418 of ChromHMM track below correspond to that of chromatin states designated in panel b.

Figure 1

Figure 2

Figure 3

Figure 4

Figure 5

Tables

Table 1. RNA-seq quality metrics

RNA-seq Sample	SRA ID	Mapped Reads	Mapping Rate (%)	Exonic Rate (%)	Intronic Rate (%)	Intergenic Rate (%)	rRNA Rate (%)	GEO Accession Number
GM1	SRX2540692	37,330,414	96.3	66.68	28.8	4.6	1.9	GSM2478318
GM2	SRX2540693	32,115,371	96.2	60.2	35.2	4.6	1.6	GSM2478319
DM1	SRX2540698	25,537,075	96.2	59.8	35.5	4.7	1.5	GSM2478324
DM2	SRX2540699	30,280,640	96.1	54.2	41.3	4.6	1.0	GSM2478325

Table 2. Sequencing reads alignment statistics for ChIP-seq datasets

ChIP-seq target	Antibody	Condition	SRA ID	Total Mapped Reads	Uniquely Mapped Reads	Unique Mapping Rate (%)	GEO Accession Number
p300	sc-584 (Santa Cruz)	GM	SRX3599537	47,518,797	36,874,586	77.6	GSM2947737
		DM	SRX3599538	42,017,222	31,849,054	75.8	GSM2947738
H4K8ac	ab15823 (Abcam)	GM	SRX2540676	34,321,318	29,825,225	86.9	GSM2478290
		DM	SRX2540677	34,787,095	29,986,476	86.2	GSM2478291
H3K9ac	ab4441 (Abcam)	GM	SRX2540679	31,129,778	24,748,174	79.5	GSM2478293
		DM	SRX2540680	38,841,418	30,684,720	79.0	GSM2478294
H3K18ac	ab1191 (Abcam)	GM	SRX2540682	36,763,380	32,131,194	87.4	GSM2478296
		DM	SRX2540683	27,544,482	24,239,144	88.0	GSM2478297
H3K27ac	ab4729 (Abcam)	GM	SRX2540685	35,542,983	30,815,766	86.7	GSM2478299
		DM	SRX2540686	30,247,017	26,375,399	87.2	GSM2478300