Contents lists available at ScienceDirect

Genomics Data

Data in Brief

Genomic-wide transcriptional profiling in primary myoblasts reveals Runx1-regulated genes in muscle regeneration

Kfir Baruch Umansky [a], Ester Feldmesser [b], Yoram Groner [a,*]

[a] Department of Molecular Genetics, The Weizmann Institute of Science, Rehovot 76100, Israel
[b] The Grand Israel National Center for Personalized Medicine (G-INCPM), The Weizmann Institute of Science, Rehovot 76100, Israel

ARTICLE INFO

ABSTRACT

In response to muscle damage the muscle adult stem cells are activated and differentiate into myoblasts that regenerate the damaged tissue. We have recently showed that following myopathic damage the level of the Runx1 transcription factor (TF) is elevated and that during muscle regeneration this TF regulates the balance between myoblast proliferation and differentiation (Umansky et al.). We employed Runx1-dependent gene expression, Chromatin Immunoprecipitation sequencing (ChIP-seq), Assay for Transposase-Accessible Chromatin with high-throughput sequencing (ATAC-seq) and histone H3K4me1/H3K27ac modification analyses to identify a subset of Runx1-regulated genes that are co-occupied by the TFs MyoD and c-Jun and are involved in muscle regeneration (Umansky et al.). The data is available at the GEO database under the superseries accession number GSE56131.

Specifications	

Organism/cell line/tissue	*In vitro* experiments:
	    Microarray, ChIP and ATAC-sequencing: primary myoblasts (PMs)
	In vivo experiments:
	    RNA-sequencing: Soleus muscles

Sex	PM and Soleus muscle samples were derived from female and male mice, respectably.
Sequencer or array type	Illumina HiSeq 2500; Affymetrix Mouse Gene 1.0 ST microarrays

Data format	Microarray:
	    Raw: CEL files
	ChIP and ATAC-sequencing:
	    Raw: SRA files
	    Analyzed: bed files (peak files)
	RNA-sequencing:
	    Raw: SRA files
	    Analyzed: csv file (DESeq2 differential expression)

Experimental factors	Transcriptome using microarrays: 4 cultures of $Runx1^{L/L}$ [1] and $Runx1^{f/f}$ derived PMs were used for RNA purification and transcriptome analysis. ChIP- and ATAC-sequencing: 2 wild type PM cultures were used for specific antibodies (Abs) and relevant controls (non-immune serum or specie-specific IgG). RNA-sequencing: 4 Soleus muscle samples were extracted from P60 $mdx/Runx1^{L/L}$ and $mdx/Runx1^{f/f}$ mice [1].

(continued)

Specifications	
Experimental features	Transcriptome analysis was performed comparing $Runx1^{f/f}$ PMs to $Runx1^{L/L}$ control *in vitro*, and $mdx/Runx1^{f/f}$ vs. $mdx/Runx1^{L/L}$ *in vivo*; TF (Runx1, MyoD and c-Jun) binding and enhancer markers were examined in WT PMs.
Consent	The experiments were conducted in strict accordance with the recommendations of the US National Institutes of Health Guide for the Care and Use of Laboratory Animals [1].
Sample source location	N.A.

1. Direct link to deposited data

The data is available at the GEO database under:
http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE56131.

2. Experimental design, materials and methods

2.1. Experimental groups and design

To elucidate the Runx1-mediated myoblast transcriptional program during muscle regeneration, we employed genetically modified $Runx1^{f/f}$ and $mdx/Runx1^{f/f}$ mice [1]. The experimental design is schematically described in Fig. 1: First, we derived the Runx1-responsive genes by comparing the $Runx1^{f/f}$ PM transcriptome to that of $Runx1^{L/L}$ PMs. Next, we defined the Runx1-regulated gene subset by cross-analyzing the Runx1-responsive gene subset with genome-wide Runx1 ChIP-seq

* Corresponding author.
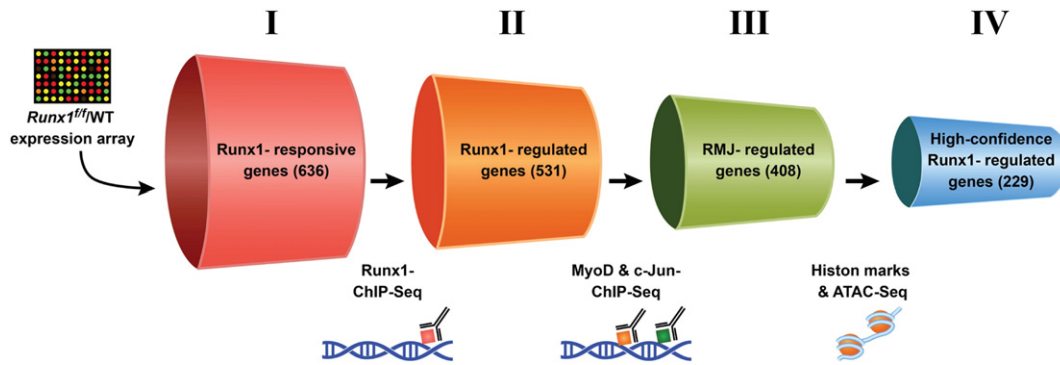   E-mail address: Yoram.groner@weizmann.ac.il (Y. Groner).

Fig. 1. Experimental design. Schematic representation of the selection procedures used to identify high-confidence Runx1-regulated genes (adapted from Umansky et al. [1]). Each cylinder represents a gene subset, with the gene number given in brackets. I — Runx1-responsive genes were derived from *Runx1^{L/L} vs. Runx1^{f/f}* PM microarray expression data [1]. II — Runx1-regulated genes were derived by cross-analysis of the Runx1-responsive gene dataset and Runx1 ChIP-seq data. This gene subset represents Runx1-responsive genes that are also occupied by Runx1. III — RMJ-regulated genes are Runx1-responsive genes that are co-occupied by Runx1, MyoD and c-Jun. IV — High-confidence Runx1-regulated gene subset is RMJ-regulated genes that were also marked as having adjacent active regulatory elements by both anti-histone modifications (H3K4me1 & H3K27ac) ChIP-seq and ATAC-seq analysis.

data in wild type PMs. To further characterize this Runx1-regulated gene subset we singled out Runx1-bound gene loci that were co-bound by Runx1 transcriptional collaborators MyoD and c-Jun (Fig. 1). Finally, we characterized the open/active chromatin by genome-wide mapping of active enhancer markers (H3K4me1, H3K27Ac) and ATAC-seq analyses. The combination of this comprehensive analysis generated a list of high-confidence Runx1-regulated genes. The expression profile of this high-confidence was validated *in vivo* using RNA-seq of RNA derived from muscles of *mdx/Runx1^{f/f} vs. mdx/Runx1^{L/L}* mice [1].

2.2. RNA purification

PM cultures were established as previously described [1,2]. For transcriptome analysis, $1e^6$–$5e^6$ PMs were collected after three stages of myoblast enrichment using pre-plating, a total of 6 days post-muscle extraction. To avoid RNA degradation, the cells were washed twice with cold PBS and then subjected to flash-freezing in liquid nitrogen. RNA was isolated by the PerfectPure RNA tissue kit (# 2302410, 5 PRIME, Germany) according to the manufacturer's instructions (cell culture protocol), using a rotor–stator as the disruption method (Omni-TH 02, Omni international, USA). For RNA-seq, Soleus muscles were harvested from 2 month old mice, and RNA was isolated as described above (tissue protocol, Proteinase K added).

2.3. Transcriptome analysis

For microarray analysis, purified total RNA was reverse-transcribed, amplified, and labeled with the Affymetrix GeneChip whole transcript sense target labeling kit. Labeled cDNA was analyzed using Affymetrix Mouse Gene 1.0 ST microarrays, according to the manufacturer's instructions. Microarray data were analyzed using the Partek Genomic Suite software. CEL files (containing raw expression measurements) were imported and data was preprocessed and normalized using the Robust Multichip Average (RMA) algorithm [3]. To identify differentially expressed genes ANOVA was applied and gene fold-changes were calculated. For RNA-seq analysis, purified total RNA was subjected to Illumina TruSeq®. RNA Sample Preparation v2 was used according to the manufacturer's instructions. Indexed samples were sequenced in an Illumina HiSeq 2500 machine in a single read mode. The obtained reads, 50 bp long, were mapped to the mm9 mouse genome assembly using TopHat2 [4] version 2.0.12.0.10 with default options. Expression at the gene level was quantified by HTSeq (version 0.6.1) [5], and using the known genes from the UCSC browser in General Feature Format (GFT) as annotation. Differential expression was calculated utilizing the DESeq2 software (version 1.2.10) [6].

2.4. ChIP-seq analysis

For ChIP, we used WT PM cultures similar to those described above in Section 2.2. Crosse-linked chromatin from $1.2e^8$ cells (Runx1 ChIP), $6e^7$ cells (MyoD and c-Jun ChIP) and $1e^7$ cells (H3K4me1 and H3K27Ac ChIP) was prepared and fragmented to an average size of approximately 200 bp by 20–35 cycles of sonication (30 s each) in 15-ml tubes using the Bioruptor UCD-200 sonicator (Diagenode, USA). Relevant antibodies and controls are described in the Materials and methods section of [1]. DNA was purified using QIAquick spin columns (QIAGEN) and sequencing was performed using Illumina HiSeq 2500. Two biological repeats were conducted and separately sequenced for each ChIP-seq experiment. For ChIP-seq analysis, the reads were aligned to the mouse genome (mm9) allowing one mismatch and using the Bowtie aligner [7]. Reads with a unique best alignment were retained for further processing. Immunoprecipitated samples were compared against the negative control to find binding sites using the MACS2 software with the callpeak function and default parameters. The broad peak setting was used only for the histone marks [8].

2.5. ATAC-seq analysis

ATAC was performed as previously described [9]. Briefly, 5×10^4 PMs were harvested, and underwent the recommended transposition protocol without the lysis stage. The resulting transposed DNA was enhanced using 12 cycles of PCR, as described. The resulting libraries were sequenced using Illumina HiSeq 2500. For ATAC-seq analysis, we obtained paired-end reads of 50 bp length. Adapters were trimmed from the reads, Bowtie2 was run as previously described for ChIP-seq analysis, duplicated reads were removed using the Picard MarkDuplicates function module, and MACS2 was applied as detailed above for ChIP-seq analysis (Section 2.4).

2.6. Functional analysis

The Ingenuity Pathway Analysis tool (https://apps.ingenuity.com/) was used for pathway and bio-function annotation of Runx1-regulated genes and the GREAT software [10] was used for ChIP-seq peak Gene Ontology (GO) analysis. The GSEA software [11] was used to perform gene set enrichment analysis on groups of genes.

3. Results & discussion

Skeletal muscle is a highly regenerative tissue. Upon muscle injury, the adult stem cells the satellite cells (SCs) are activated and differentiate and fuse to create myofibers that regenerate the tissue. While not
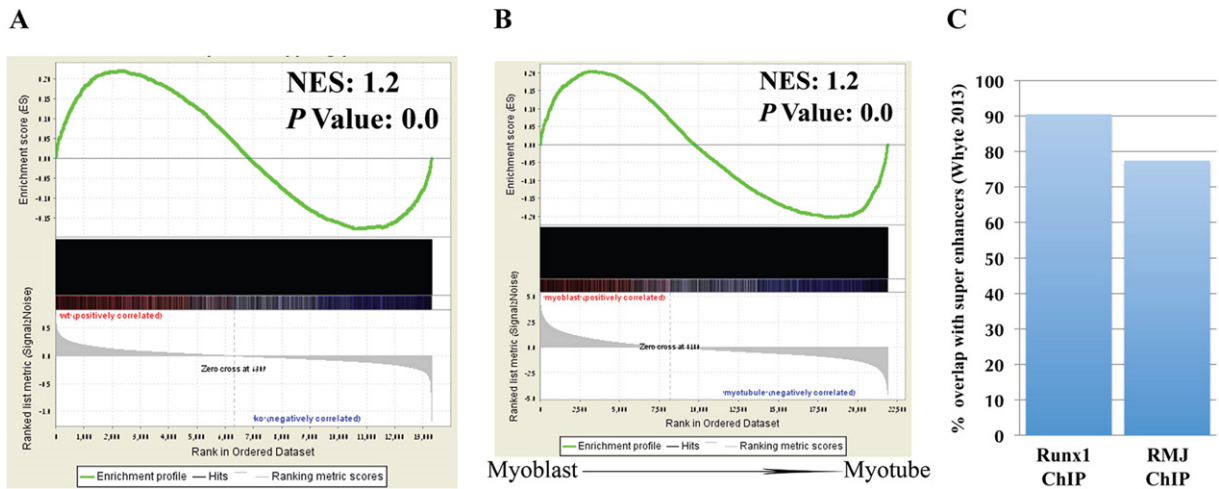
Fig. 2. Correlation with previously-acquired data sets. (A, B) GSEA analysis of genes bound by all three TFs in comparison to Runx1-responsive genes in proliferating PMs (A) or gene expression data of WT myoblasts *vs.* differentiated myotubes (B). (C) The Runx1- and RMJ-bound loci in PM ChIP-seq were compared to published C2C12 cell line derived MyoD bound "Super enhancers" [16] loci. Percent overlapping loci of the relevant dataset are presented. C2C12 MyoD bound "Super-enhancers" compared to Runx1 and RMJ ChIP-seq. Overlap presented as percent of Super-enhancers.

expressed in healthy muscle tissue, expression of TF Runx1 significantly increases in response to various types of muscle damage including neuropathic damage [12,13] and myopathic damage [1]. Specifically, we have recently shown that Runx1 prevents premature differentiation of SC-derived myoblasts during muscle regeneration [1]. In the myoblastic cell line C2C12 Runx1 was co-bound with the major myogenic factor, MyoD [14]. Using *in vivo* loss- and gain-of-function models and genome-wide transcriptional analyses we identify a subset of Runx1-regulateed genes that participate in muscle regeneration. Significantly, Runx1 transcriptome and ChIP-seq analysis show good correlation (Fig. 2A). Our ChIP-seq data also showed correlation with myoblast-specific transcriptional-profiling data. For example, we found that the proximity of myoblast-specific expressed genes [15] was highly enriched for co-binding of Runx1, MyoD and c-Jun (RMJ) (Fig. 2B). In addition, analysis of myoblast-specific Super-enhancers [16], revealed a strong overlap with our ChIP-seq data in sites regulating myoblast fate (Fig. 2C). Interestingly, Runx1 is involved in cell fate decision of additional adult stem cells populations, including hair follicle stem cells [17], and differentiation of mesenchymal stem cells into myofibroblasts [18]. It is conceivable that common transcription regulatory mechanisms are shared by the three systems. In summary, our data suggests that Runx1 regulates the core transcriptional program that determines myoblastic cell fate and facilitates muscle regeneration by preventing premature differentiation of proliferating myoblasts. The comprehensive data set provides a useful recourse for deciphering the molecular mechanisms underlying muscle regeneration.

References

[1] K.B. Umansky, Y. Gruenbaum-Cohen, M. Tsoory, E. Feldmesser, D. Goldenberg, et al., Runx1 transcription factor is required for myoblasts proliferation during muscle regeneration. PLoS Genet. 11 (2015) e1005457.

[2] Y. Gruenbaum-Cohen, I. Harel, K.B. Umansky, E. Tzahor, S.B. Snapper, et al., The actin regulator N-WASp is required for muscle-cell fusion in mice. Proc. Natl. Acad. Sci. U. S. A. 109 (2012) 11211–11216.

[3] R.A. Irizarry, B. Hobbs, F. Collin, Y.D. Beazer-Barclay, K.J. Antonellis, et al., Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4 (2003) 249–264.

[4] D. Kim, G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, et al., TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 14 (2013) R36.

[5] S. Anders, P.T. Pyl, W. Huber, HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31 (2) (2015 Jan 15) 166–169.

[6] M.I.H.W. Love, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. bioRxiv (2014), http://dx.doi.org/10.1101/002832,2014.

[7] B. Langmead, C. Trapnell, M. Pop, S.L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10 (2009) R25.

[8] Y. Zhang, T. Liu, C.A. Meyer, J. Eeckhoute, D.S. Johnson, et al., Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9 (2008) R137.

[9] J.D. Buenrostro, P.G. Giresi, L.C. Zaba, H.Y. Chang, W.J. Greenleaf, Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods 10 (2013) 1213–1218.

[10] C.Y. McLean, D. Bristor, M. Hiller, S.L. Clarke, B.T. Schaar, et al., GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. 28 (2010) 495–501.

[11] A. Subramanian, P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, et al., Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 15545–15550.

[12] X. Wang, C. Blagden, J. Fan, S.J. Nowak, I. Taniuchi, et al., Runx1 prevents wasting, myofibrillar disorganization, and autophagy of skeletal muscle. Genes Dev. 19 (2005) 1715–1722.

[13] X. Zhu, J.E. Yeadon, S.J. Burden, AML1 is expressed in skeletal muscle and is regulated by innervation. Mol. Cell. Biol. 14 (1994) 8051–8057.

[14] R. Blum, V. Vethantham, C. Bowman, M. Rudnicki, B.D. Dynlacht, Genome-wide identification of enhancers in skeletal muscle: the role of MyoD1. Genes Dev. 26 (2012) 2763–2779.

[15] M.S. Huh, M.H. Parker, A. Scime, R. Parks, M.A. Rudnicki, Rb is required for progression through myogenic differentiation but not maintenance of terminal differentiation. J. Cell Biol. 166 (2004) 865–876.

[16] W.A. Whyte, D.A. Orlando, D. Hnisz, B.J. Abraham, C.Y. Lin, et al., Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell 153 (2013) 307–319.

[17] C.S. Hoi, S.E. Lee, S.Y. Lu, D.J. McDermitt, K.M. Osorio, et al., Runx1 directly promotes proliferation of hair follicle stem cells and epithelial tumor formation in mouse skin. Mol. Cell. Biol. 30 (2010) 2518–2536.

[18] W. Kim, D.A. Barron, R. San Martin, K.S. Chan, L.L. Tran, et al., RUNX1 is essential for mesenchymal stem cell proliferation and myofibroblast differentiation. Proc. Natl. Acad. Sci. U. S. A. 111 (2014) 16389–16394.