Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2020 doi:10.20944/preprints202010.0365.v1

Article

Epigenetics of muscle- and brain-specific expression of KLHL family genes

Kenneth C. Ehrlich[1], Carl Baribault[2], Melanie Ehrlich[3]*

[1] Tulane Center for Biomedical Informatics and Genomics, Tulane University Health Sciences Center, New Orleans, LA 70112, USA; kehrlich@tulane.edu
[2] Center for Research and Scientific Computing (CRSC), Tulane University Information Technology, LA 70112, USA; cbaribault@tulane.edu
[3] Tulane Center for Biomedical Informatics and Genomics, Tulane Cancer Center, Hayward Genetics Program, Tulane University Health Sciences Center, New Orleans, LA 70112; ehrlich@tulane.edu
* Correspondence: ehrlich@tulane.edu; Tel.: +1-504-939-0916 (M.E.)

Abstract: *KLHL* genes and the related *KBTBD* genes encode components of the Cullin-E3 ubiquitin ligase complex and typically target tissue-specific proteins for degradation, thereby affecting differentiation, homeostasis, metabolism, cell signaling, and the oxidative stress response. Despite their importance in normal cell function and in disease (*KLHL40, KLHL41, KBTBD13, KEAP1,* and *ENC1*), previous studies that examined epigenetic factors affecting transcription were predominantly limited to promoter DNA methylation. Using diverse tissue and cell culture whole-genome profiles, we examined 17 *KLHL* or *KBTBD* genes preferentially expressed in skeletal muscle or brain to identify tissue-specific enhancer and promoter chromatin, open chromatin (DNaseI hypersensitivity), and DNA hypomethylation. Sixteen of the 17 genes displayed muscle- or brain-specific enhancer chromatin in their gene bodies, and most exhibited specific intergenic enhancer chromatin as well. Seven genes were embedded in a super-enhancer. The enhancer chromatin regions typically displayed foci of DNA hypomethylation at peaks of open chromatin. In addition, we found evidence that gene neighbors (*HHATL* and *FBXO32*) of *KLHL40* and *KLHL38* harbor enhancer chromatin that likely upregulates the adjacent *KLHL* gene. Many *KLHL*/*KBTBD* genes had tissue-specific promoter chromatin at their 5′ ends, but surprisingly, two (*KBTBD11* and *KLHL31*) had constitutively unmethylated promoter chromatin in their 3′ exons that overlaps a retrotransposed *KLHL* gene. Our findings demonstrate the importance of expanding epigenetic analyses beyond the 5′ ends of genes in studies of normal and abnormal gene regulation.

Keywords: DNA methylation; chromatin structure; enhancers; super-enhancers; promoters; retrogene; DNase hypersensitive sites; skeletal muscle; brain; topologically associating domains

1. Introduction

Establishing and maintaining tissue-specific levels of active proteins that help mediate differentiation depends on regulating protein stability as well as transcription, mRNA stability, translation, and post-translational protein modification [1-4]. Kelch-like (*KLHL*) genes encode substrate-specific adapters for cullin E3 ubiquitination that play a major role in targeting specific proteins for degradation and thus affect differentiation and homeostasis [5,6]. KLHL proteins contain five or six Kelch domains, which mediate substrate recruitment; a BTB/POZ domain, which binds to the Cullin E3 ligase; and a BACK domain with uncertain function that links the other two domains. Like the KLHL proteins, KBTBD proteins have Kelch and BTB/POZ domains and, in some proteins, a BACK domain. They also usually function in specific ubiquitin-mediated proteolysis. Both types of proteins influence

© 2020 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2020 doi:10.20944/preprints202010.0365.v1

fundamental cellular processes including cell signaling, transcription, cell division, cytoskeletal remodeling, response to oxidative stress, apoptosis, and autophagy [6,7]. Because of the structural and functional similarities of *KLHL* and *KBTBD* genes, we refer to both as being in the *KLHL* gene family.

Consistent with their wide-ranging roles, some of the KLHL or KBTBD proteins are associated with somatic diseases or linked to hereditary disease [6,7]. Downregulation of *KEAP1* (*KLHL19*), the most well-studied *KLHL* family gene, is implicated in many types of cancer [6,8], and cancer-associated mutation of other *KLHL* family genes has been described [5]. Among the other diseases in which *KLHL* family proteins are implicated are myopathies [9,10], diabetes [11], pseudohypoaldosteronism (a mineral transport disorder) [12], and cardiac dysfunction [13]. Although *KLHL* family genes are usually highly tissue-specific in their transcription, almost one-third of the proteins encoded by the 51 genes in this family (including *KBTBD* genes) are poorly understood as to their biological function(s).

Despite the biological importance of the *KLHL* family genes, there has been only a limited examination of their epigenetics, mostly with respect to methylation of the promoter regions, which is usually defined as a 0.1- to 2 kb region upstream of the transcription start site (TSS) and extending up to 1 kb downstream of the TSS [6]. Most human promoter regions have low or negligible levels of DNA methylation in all tissues and are enriched in CpGs, unlike most of the rest of the genome. Such a region is called a CpG island (CGI) [14]. When a CGI overlapping an active promoter becomes highly methylated, the promoter is almost always silenced. However, the lack of DNA methylation at a promoter and an open chromatin configuration can be seen in some genes that are poised for transcription as well as at those that are actively transcribed.

Enhancers are highly important in establishing tissue-specific expression [4,15]. Enhancer chromatin in genome-wide studies is usually defined as regions enriched in histone H3 lysine 27 acetylation (H3K27ac) and histone H3K4 monomethylation (H3K4me1) while active promoters exhibit H3K27ac enrichment and H3K4 trimethylation (H3K4me3) instead of H3K4me1 and are usually adjacent to the TSS. Enhancers can be immediately upstream of promoters but are more frequently found elsewhere [4,16]. Just as tissue-specific or disease-associated DNA methylation can down-regulate promoters, it can also down-modulate enhancers, mostly by limiting the accessibility and binding of a transcription factor to the enhancer [15,17]. Enhancers are sometimes very far from the promoter that they upregulate. However, they should be in the same higher-order chromatin loop or topologically associating domain (TAD) established by binding of the CCCTC-binding factor (CTCF) at the ends of the loop [18].

In the current study we used publicly available transcription and epigenetic profiles of the 17 *KLHL* family genes that show preferential expression in skeletal muscle (SkM) or brain to compare these genes' transcriptomics and epigenomics in many human tissues and primary cell cultures. Because of the importance of gene neighborhoods to transcription regulation, we extended our analysis to include genes that surround the *KLHL* family genes. Our analysis reinforces the importance of studying the epigenetics of transcription regulatory regions beyond the canonical upstream promoter region to better understand tissue-specific gene expression and to give new insights into well-studied *KLHL* family genes as well as into many little-examined ones.

2. Results

2.1 Many KLHL and KBTBD genes are expressed preferentially in SkM or brain

The three tissues with the most frequent tissue-specific expression of the 51 *KLHL* family genes (with the designation "*KLHL*" or "*KBTBD*"), were SkM, brain and testis (Supplementary Tables S1 and S2) as determined

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2020 doi:10.20944/preprints202010.0365.v1

from the median TPM (transcripts per kilobase million) from hundreds of biological replicates per tissue in the GTEx database [19]. We studied the epigenetics of only SkM- or brain-associated genes because of the availability of many epigenetic profiles for SkM and brain but not for testis. There were 10 *KLHL* family genes preferentially expressed in SkM and 6 in most examined subregions of the brain (Tables 1 and 2, respectively) using as the definition of preferential expression that the median TPM for SkM or brain (prefrontal cortex) is at least 3-fold higher than the median TPM of non-SkM or non-brain tissue types in the GTEx database, which has 52 tissue types. Five genes (*KBTBD6*, *KBTBD12*, *KLHL1*, *KLHL3*, and *KLHL22*) that were preferentially expressed specifically in cerebellum, the most transcriptionally distinct brain subregion [20], were not included in our study because of the lack of available chromatin state segmentation profiles for this brain subregion. Genes that were preferentially expressed in prefrontal cortex usually had preferential expression in the other examined brain regions, with cerebellum being the most frequent exception (Table 2 and Supplemental Table S2). Because *KEAP1* (which had its highest expression in SkM, Supplementary Table S2) is by far the most studied of all the *KLHL* family genes and is involved in various diseases [21], it is included in our study even though its SkM/non-SkM expression ratio was only 2.4 instead of >3.

Table 1. *KLHL* family genes that are preferentially expressed in skeletal muscle[a]

Description	TPM in SkM[b]	TPM ratio: SkM to median of other tissues	FPKM in myoblasts	FPKM ratio: myoblasts to median of heterologous cell cultures[c]	FPKM ratio: myotubes to myoblasts
KBTBD12 (Figure S4)	15	13	0.3	0	2.8
KBTBD13[d] (Figure S6)	1.5	25	0	0	0
KEAP1 (*KLHL19*, Figure 3A)	88	2.4	32	0.7	1.0
KLHL21 (Figure S10)	84	3.3	31	0.6	1.4
KLHL30 (Figure 2A)	110	145	16	18.0	2.8
KLHL31 (Figure S2)	18	59	5.6	5.8	872
KLHL33 (Figure S1)	11	24	0	0	0
KLHL34 (Figure S3)	8.5	113	0	0	0
KLHL38 (Figure 2B)	90	405	0.1	0	19
KLHL40 (Figures 1B; S5)	303	3373	7.7	#DIV/0!	17
KLHL41 (Figure 1A)	3420	1946	104	1613	254

[a]We define preferential expression as genes that display a median TPM ≥ 1 in skeletal muscle (SkM) and a ratio of TPM in SkM to the median TPM in 51 other tissues > 3, the only exception is *KEAP1*. TPM, transcripts per kilobase million; FPKM, fragments per kilobase million, similar to TPM

[b]The RNA-seq expression data for SkM and 51 other tissues (Supplemental Table S2) are from the GTEx Analysis Release V8 [22,23].

[c]FPKM, fragments per kilobase of exon per million reads mapped, are from technical duplicates from the ENCODE database [24] (Supplemental Table S2)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2020 doi:10.20944/preprints202010.0365.v1

[d]Highest expression is in tibial artery (TPM, 2.6)

Table 2. *KLHL* family genes that are preferentially expressed in brain[a]

Description	Brain tissue (median TPM)			Ratio of brain tissue TPM to median TPM of non-brain tissues[b]		
	Cortex[c]	Hippoc.[c]	Cerebel.[c]	Cortex	Hippoc.	Cerebel.
ENC1 (*KLHL37*, Figure 3B)	322	87	4.3	56	15	0.7
KBTBD11 (Figure 4B)	59	27	46	17	7.2	12
KLHL2 (Figure S9A)	47	47	21	3.2	45	39
KLHL32 (Figure 4A)	10	12	11	37	3.9	6.0
KLHL35[d] (Figure S9C)	3.3	1.6	2.5	8.0	13	2.9
KLHL4 (Figure S9B)	3.6	4.5	1.0	11	15	0.7

[a]Criteria for preferential expression is that the ratio of TPM in brain frontal cortex to median TPM of non-brain tissues > 3.

[b]In the GTEx dataset there are 40 non-brain tissues

[c]Cortex, frontal cortex; hippoc, hippocampus; cerebel, cerebellum

[d]Highest expression is in testis (TPM, 10.3)

2.2 Extensive intragenic promoter chromatin and overlapping DNA hypomethylation correlates with the extremely high expression of KLHL41 in skeletal muscle

KLHL41 (formerly, *KBTBD10*), a gene whose recessive coding mutations can cause nemaline myopathy [10], had much higher expression in SkM (TPM, 3420) than any other *KLHL* family gene and much lower expression in other tissues (Tables 1 and S2). It is also preferentially expressed in myoblasts relative to other cell cultures but displays much higher expression in myotubes (Table 1), suggesting that it is strongly upregulated upon myoblast fusion to form multinucleated myotubes. Consistent with its SkM lineage specificity, the SkM-specific transcription factor (TF) MyoD (Figure 1A, three triangles over H3K27ac tracks) was bound to sites in the *KLHL41* upstream region (121 and 1294 bp upstream of the transcription start site, TSS) as well as to a site in intron 1, as determined by chromatin immunoprecipitation coupled with next-gen DNA sequencing in human myoblasts (MyoD ChIP-seq) [25]. These MyoD sites overlap regions of enriched histone acetylation (Figure 1A, H3K27ac, purple box) in myoblasts, myotubes and SkM, which denote either enhancer or promoter chromatin, except that the most distal MyoD site overlaps an ~0.2-kb hole in H3K27ac signal that probably corresponds to a nucleosome-depleted subregion. Such sites are created by especially strong TF binding. Indeed, all three MyoD binding sites in myoblasts overlap tissue-specific peaks of open chromatin (DNase-seq peaks, purple box), which correspond to enhancer chromatin or less frequently, to tissue-specific promoter chromatin [26].

The strong histone acetylation (H3K27ac) signal seen in cells in the SkM lineage extend upstream of the TSS to 5 kb downstream of it and indicate the presence of a super-enhancer (Figure 1A, dotted line over chromatin state tracks and purple box in DNA methylation tracks). A super-enhancer is a long (>3 kb) cluster of enhancer or

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2020 doi:10.20944/preprints202010.0365.v1

promoter chromatin that is associated with strong, differentiation-related expression of the related gene [27,28]. Overlapping most of this super-enhancer was strong and SkM-specific DNA hypomethylation (Figure 1A, DNA meth, purple box) that was confirmed in two additional SkM samples from the Roadmap Project (data not shown). The unusual long region of DNA hypomethylation (Figure 1A, DNA meth, purple box) at mostly promoter-type chromatin (Figure 1A, SkM, chromatin state, red segments) in the super-enhancer may be responsible, in part, for the especially high level of expression of this gene (Figure 1A, red segments in chromatin state segmentation). DNA-hypomethylated promoter chromatin extending far downstream from the TSS and associated with preferential expression in SkM was also seen in *KLHL33* and *KLHL31* (Supplementary Figures. S1 and S2).

Heart had the second highest level of *KLHL41* expression among the 50 examined tissues (left ventricle TPM, 57, and right atrial appendage TPM, 48). Surprisingly, the heart and SkM DNaseI hypersensitivity profiles (Figure 1A) and their H3K4me3 profiles (data not shown) were very similar despite ~60-fold higher expression of *KLHL41* in these two tissues. However, heart lacked the extensive H3K27ac signal, the super-enhancer, and the tissue-specific DNA hypomethylation of SkM (Figure 1A).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2020 doi:10.20944/preprints202010.0365.v1

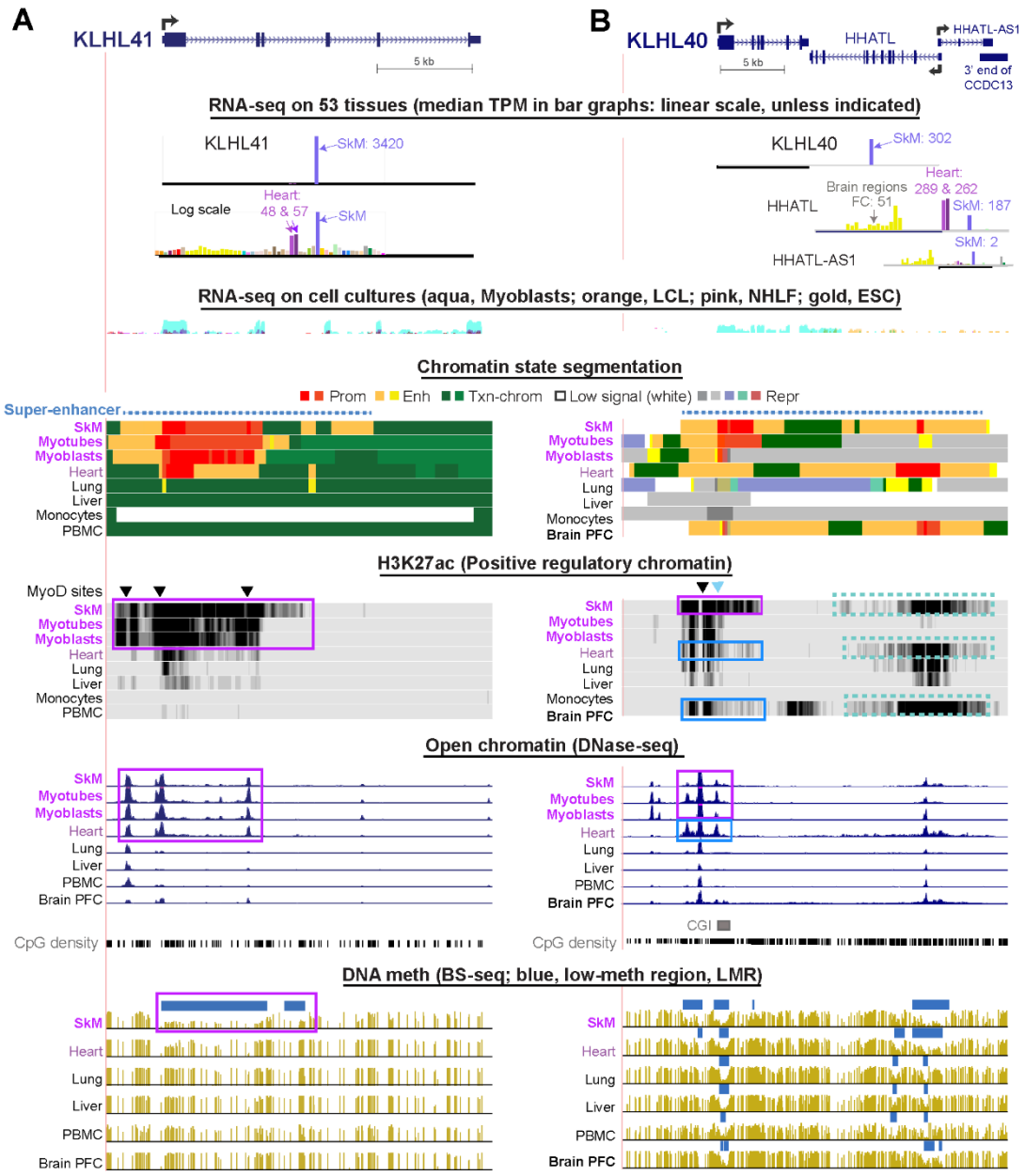

Figure 1. *KLHL40* and *KLHL41*, genes involved in nemaline myopathy, have tissue-specific chromatin and DNA methylation profiles that correlate with their skeletal muscle-specific expression. (A) *KLHL41*, chr2:170,363,256-170,383,339 and **(B)** *KLHL40* and neighboring genes *HHATL* and the 3'-end of *CCDC13*, chr3:42,719,627-42,749,414. RNA-seq profiles for tissues, either linear or $\log_{10}$-transformed TPM bar graphs as indicated; RNA-seq for cell cultures, overlay signal for the indicated four cell cultures; brain FC, frontal cortex; heart, right atrial appendage and left ventricle. Dotted blue line over the chromatin state tracks, SkM super-enhancer; H3K27ac signal, vertical viewing ranges 0-20 and 0-10 for *KLHL41* and *KLHL40*, respectively; DNase hypersensitivity, vertical viewing range 0-20; BS, bisulfite-seq with blue bars indicating regions of low-methylation relative to methylation in the same tissue. All tracks are aligned in this and other figures and are from the UCSC Genome Browser (http://www.genome.ucsc.edu/) hg19 reference genome. Boxed regions are described in the text.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2020 doi:10.20944/preprints202010.0365.v1

KLHL41 displayed H3 lysine-36 trimethylation (H3K36me3; transcription-type chromatin, Txn-chrom, green segments, Figure 1A) in gene-upstream and downstream regions as well as in the gene body in most of the tissue and cell samples. These included samples with negligible RNA-seq levels for *KLHL41* (e.g., liver and leukocytes). This H3K36me3 enrichment is characteristic of actively transcribed chromatin downstream of the 5′ end of the gene body. Such transcription chromatin in tissues without *KLHL41* transcription is probably due to low levels of read-through transcription from adjacent genes, *BBS5* (ENST00000513963/RP11-724O16.1), expressed at highest levels in testis) and *FASTKD1*, whose 3′ ends are only 2 – 3 kb from the 5′ or 3′ ends of *KLHL41* (data not shown).

2.3 Intragenic and intergenic enhancer chromatin at KLHL40 or its neighbor, HHATL may upregulate both genes or one or the other gene depending on the tissue

KLHL40 (formerly, *KBTBD5*), another gene that can contain recessive mutations causing nemaline myopathy [29], is the second most highly expressed *KLHL* gene in SkM (TPM, 302). Like *KLHL41*, it is expressed specifically in SkM and in myotubes and, to a lesser extent, in myoblasts (Table 1). Only SkM and myotubes displayed much promoter chromatin near the TSS (Figure 1B, Chromatin state tracks). This promoter chromatin overlaps a CGI that had only low levels of methylation in all tissues regardless of expression of the associated gene (Figure 1B, DNA meth), like most CGIs at promoters [14]. However, specifically in SkM, there was DNA hypomethylation in enhancer chromatin extending 0.3-kb upward from the TSS as well as in a 1.5-kb enhancer chromatin region a little further upstream. Tissue-specific DNA demethylation extending from a constitutively unmethylated promoter CGI was correlated with tissue-specific expression in eight other *KLHL* family genes, including *KLHL34* and *KBTBD12* (Supplemental Figures S3 and S4). In the promoter region of *KLHL40*, there was a single occupied MyoD site overlapping enhancer chromatin in human myoblasts (Figure 1B, black triangle above H3K27ac tracks) [25]. This is one of the three predicted MyoD sites in the ~1.2 kb region upstream of the *KLHL40* TSS described by Bowlin et al. [30] and shown by them to be important for driving transcription of a reporter gene in mouse myoblast host cells. A second MyoD binding site described by Bowlin et al. was seen only in a ChIP-seq profiles of mouse myoblasts and myotubes [31]. The position of this MyoD site in the mouse/human homologous region is shown in Figure 1B (blue triangle above the H3K27ac tracks).

Both heart and brain displayed enhancer chromatin over the *KLHL40* coding region (Figure 1B, blue boxes) even though they did not express this gene. This heart and brain enhancer chromatin replaced a long segment of TSS-downstream promoter chromatin seen in SkM and is likely to be upregulating the *KLHL40*-upstream gene, *HHATL* (Figure 1B), rather than *KLHL40*, which had little or no promoter chromatin in these tissues. *HHATL* encodes a hedgehog acyltransferase that is involved in SkM maturation ([32]) and is present in the sarcoplasmic reticulum of SkM cells and cardiac myocytes [32]. *HHATL* is expressed at high levels in heart, SkM, and brain. The only study of *HHATL* related to the nervous system reported changes in the levels of HHATL protein in myocardial tissue in a rat epilepsy model [33]. Based on chromatin conformation capture analysis of skin fibroblasts (Micro-C, [34]), *KLHL40* and *HHATL* are likely to be in the same TAD anchored by CTCF sites (Supplemental Figure S5) if the TAD structure in heart and brain cells is like that of fibroblasts. A heart/brain enhancer overlapping *KLHL40* (Figure 1B) could thereby upregulate its *HHATL* neighbor. Interestingly, there is an antisense *HHATL* transcript (*HHATL-AS1*) whose expression is highest in SkM (TPM ratio in SkM to median TPM in 50 tissues = 18) and secondarily in brain, like *HHATL* (Figure 1B, RNA-seq bar graphs [19]). Specific enrichment in H3K27ac is seen in the *HHATL/HHATL-AS1* region in brain and SkM (Figure 1B, dotted boxes in H3K27ac

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2020 doi:10.20944/preprints202010.0365.v1

tracks). Therefore, *HHATL-AS1*, like many long intergenic non-coding RNA genes 5' to a coupled protein-encoding gene, may upregulate expression of *HHATL* specifically in SkM and brain.

A third *KLHL* family gene that is linked to autosomal nemaline myopathy is *KBTBD13*, but its disease mutations are dominant [35]. Unlike the previously described genes, *KBTBD13* is very small (3.1 kb) with no introns (Supplemental Figure S6) and is expressed only at a low levels in SkM (TPM, 1.4) although the expression ratio in SkM to the median of non-SkM tissues is 25 (Table 1). *RASL12*, which encodes a small GTPase superfamily protein, is expressed antisense to *KBTBD13*. It has an isoform (ENST00000434605) with a TSS that is only 0.1 kb upstream of the *KBTBD13* TSS. This isoform and *KBTBD13* have partly overlapping expression profiles, which is not surprising given their shared core promoter region. However, preferential expression in SkM and heart was seen only for *KBTBD13*. Although *KBTBD13* is not expressed in myoblasts, there were several occupied MyoD binding sites downstream of the gene in these cells (Supplemental Figure S6, above H3K27ac tracks), which may be used for activation of this gene in later stages of myogenesis.

2.4 Skeletal muscle-specific expression, promoter chromatin and enhancer chromatin in KLHL30 and in KLHL38, a paralog that contains a retrogene from KLHL30

KLHL30, a gene that has not been the focus of any published articles, is much more highly expressed in SkM than in any other tissue and has higher expression in myoblasts and myotubes than in non-muscle cell cultures (Table 1). Its high expression in SkM is associated with a SkM-specific super-enhancer (Figure 2A, dotted line above chromatin state tracks). This super-enhancer contained strong promoter chromatin interspersed in the gene body, which might represent promoter activity driving an unusually high levels of expression of the unstable noncoding RNAs (ncRNAs) associated with active enhancers [36]. Heart and aorta had considerable expression of this gene, although ~5 – 10 times less than that of SkM and lacked the super-enhancer and this intragenic promoter chromatin. SkM, heart, and aorta had similar open chromatin profiles (Figure 2A). Heart and aorta also lacked two H3K27ac-rich chromatin regions (one within *KLHL30* and one downstream) that overlapped DNA hypomethylation in SkM (Figure 2A, H3K27ac and DNA methylation tracks, dotted boxes). There were two MyoD binding sites within the super-enhancer in myoblasts overlapping SkM-hypomethylated regions (Figure 2A, triangles over H3K27ac tracks).

There is an intriguing relationship between *KLHL30* and *KLHL38*, another little studied gene that may help to reverse muscle atrophy [37]. Most of the open reading frame (ORF) of *KLHL38* is in exon 2 and is largely derived from a retrotransposed copy (retrogene) of coding sequences in exon 2 of *KLHL30* (Figure 2A and B). Both genes have a strong preference for expression in SkM (Table 1). The retrogene from *KLHL30* located in *KLHL38* encodes the BTB/POZ and Back domains and two of the Kelch repeats [6] as determined by protein BLAST (BLASTp). There are eight other *KLHL* family genes containing a retrogene exon from exonic DNA of another *KLHL* gene (Supplemental Table S3). The *KLHL30* retrogene in exon 2 of *KLHL38* is an expressed shuffle retrogene, *i.e.*, a retrogene that is expressed in a pre-existing host gene and contributes part of its ORF to the host gene's ORF [24,38]. Five other *KLHL* family genes contained expressed shuffle retrogenes derived from other *KLHL* family members (Supplementary Table S3). Three additional *KLHL* family genes contained retrogenes derived from another *KLHL* family member but did not meet the definition of an expressed shuffle retrogene. The retrogene in *KLHL38* had an especially high percentage sequence identity to its parent gene, namely, 80%. In comparison, among the 740 host genes in the human genome containing expressed shuffle retrogenes, the median percentage of bases matching the parent gene was only 31% (Supplementary Table S4). These findings are consistent with

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2020 doi:10.20944/preprints202010.0365.v1

KLHL38 and KLHL30 coming from the same clade (Clade 4, Supplemental Figure S7; see [7] for a similar analysis that did not identify clades).

Like *KLHL30*, *KLHL38* was embedded in a super-enhancer in SkM. *KLHL38* also displayed a super-enhancer in heart even though the steady-state *KLHL38* RNA levels in heart were only modest unlike those for SkM (Figure 2B). However, the heart super-enhancer extended only 2 kb upstream of the TSS while the SkM enhancer began 13 kb upstream. Importantly, heart had fewer subregions of DNA hypomethylation and less H3K27ac enrichment in the region extending from the *KLHL38* super-enhancer through a super-enhancer overlapping its downstream neighbor *FBXO32* (Figure 2B, dotted boxes). In the 1-Mb neighborhood of *KLHL38*, *FBXO32* (Atrogin-1) was the only other gene exhibiting strong preferential expression in SkM. *FBXO32* is critically involved in the ubiquitin-driven degradation of proteins in SkM during sarcopenia ([39]. It is expressed most highly in SkM and at an even higher level than *KLHL38*, but at much lower levels in heart (Figure 2B). In the *KLHL38*/*FBXO32* intergenic region, MyoD binding was seen at four sites in human myoblasts (Figure 2B, black triangles over H3K27ac tracks), which express *FBXO32* but not *KLHL38* (Figure 2B and Table 1). This suggests that the MyoD binding sites and enhancer chromatin in the intergenic region between *FBXO32* and *KLHL38* are upregulating just *FBXO32* in these SkM progenitor cells. Similarly, aorta strongly expresses *FBXO32* but weakly expresses *KLHL38* and has extensive enhancer chromatin and subregions of tissue-specific DNA hypomethylation in the intergenic region. Moreover, aorta also exhibited DNA hypomethylation in the gene body of *KLHL38* and the *KLHL38*-upstream region, which probably also upregulates only *FBXO32*. In SkM there might be long-distance promoter-enhancer interactions that upregulate both genes, which are preferentially expressed in this tissue. Findings from chromosome looping analysis are consistent with this hypothesis because the *KLHL38*/*FBXO32* neighborhood is in a TAD of ~225 kb ending at CTCF sites as determined in skin fibroblasts (Micro-C, [34], Supplemental Figure S8).

2.5 Only minor skeletal muscle-associated epigenetic differences occur in KEAP1 consistent with its highest expression in SkM but otherwise broad tissue expression profile

KEAP1 (Kelch Like ECH Associated Protein 1, *KLHL19*) plays a major role in regulation of cellular oxidative stress in SkM and other tissues and contributes to cancer and other diseases [6] (Supplemental Tables S5). KEAP1 protein participates in the oxidative stress response by targeting the transcription factor NFE2L2 (NRF2) for degradation [40]. Given the protective role of KEAP1, it is not surprising that the gene is expressed with a broad tissue distribution. Nonetheless, SkM has a higher level of *KEAP1* RNA compared with other normal tissues and an expression ratio of 2.3 for SkM TPM relative to the median TPM of 51 other tissues (Tables 1, S2, and Figure 3A). Chromatin segmentation, open chromatin, and DNA methylation profiles in and around this gene were similar in all tissues except that two of the three SkM samples (SkM 2 and SkM 3) for which histone modification profiles were available had more enhancer chromatin and H3K27ac signal in the promoter-downstream region than seen in the other samples (Figure 3A, purple boxes in chromatin state and H3K27ac tracks). The main Roadmap sample (SkM 1) is from psoas muscle, but the source of the SkM of these two samples is from unknown regions of the leg. The 803 SkM samples used to determine the median RNA levels for the GTEx database were from gastrocnemius (calf) muscle samples and could account for differences in chromatin state profiles of these

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2020 doi:10.20944/preprints202010.0365.v1

three muscle samples. The exact source of SkM might affect its *KEAP1* expression as it does for some other genes [41] suggesting that for expression of *KEAP1*, SkM 2 and 3 is more like gastrocnemius muscle than like psoas.

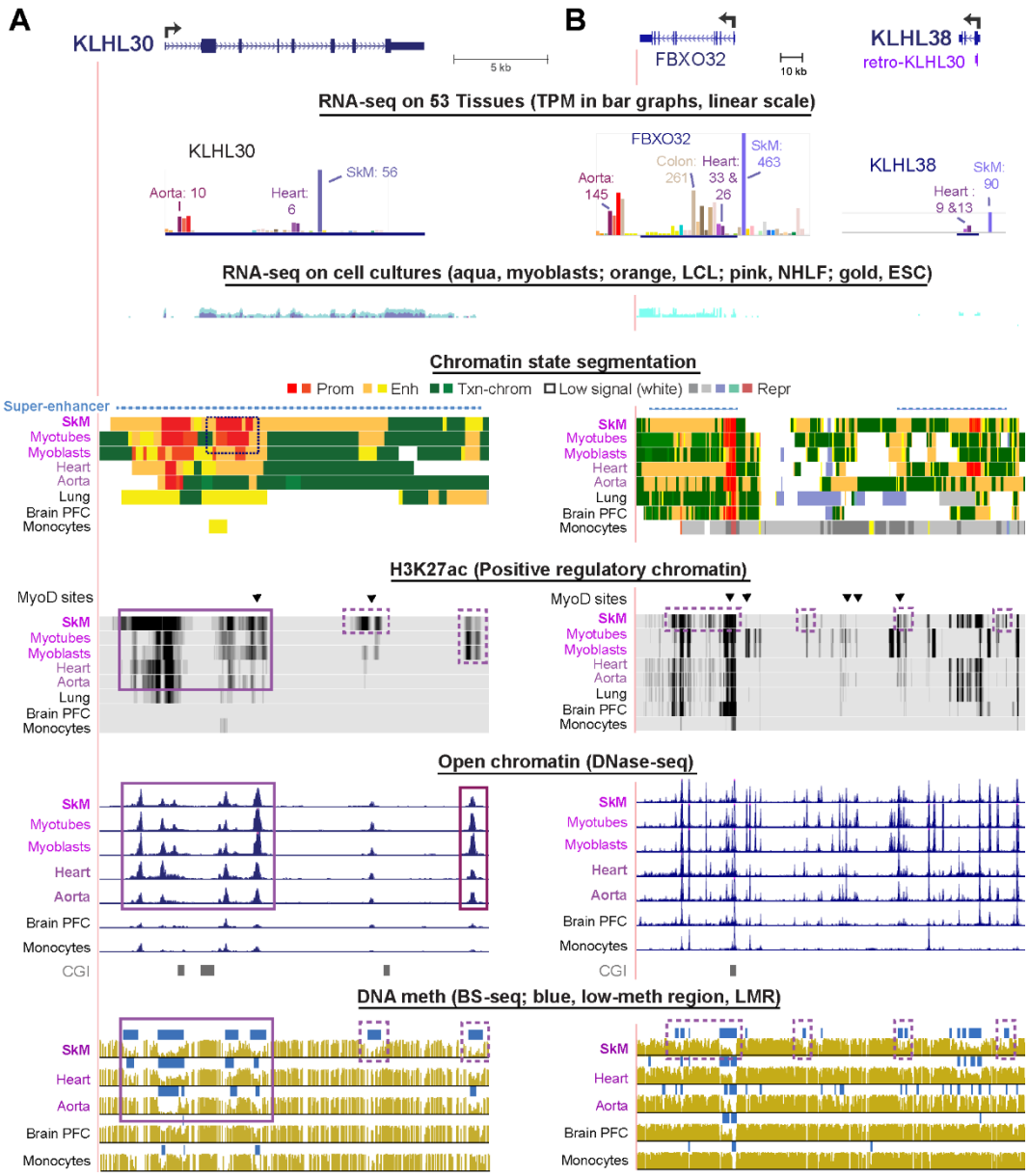

Figure 2. ***KLHL30*** **and** ***KLHL38*****, little studied** ***KLHL*** **genes, have super-enhancer chromatin and DNA hypomethylation that can drive their high, muscle-specific expression. (A)** *KLHL30* (chr2:239,043,816-239,065,093) and **(B)** *KLHL38* and its downstream neighbor, *FBXO32* (chr8:124,508,193-124,686,846). Tracks from UCSC browser are as described in Figure 1. The immediate upstream neighbor of *KLHL38*, *FBXO32*/Atrogen-1, a gene encoding an E3-ubiquitin ligase implicated in muscle atrophy [42], is shown because shared enhancer regions in SkM might coordinately regulate expression of both genes.

2.6 The neurogenesis-associated ENC1 gene exhibited much more enhancer chromatin and DNA hypomethylation in fetal than in adult brain

ENC1 (Ectodermal-Neural Cortex 1; *KLHL37*) is the most highly expressed *KLHL* family gene in the frontal cortex and hippocampus of brain (TPM, 322 and 87, respectively) (Table 2 and Supplemental Table S2,).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2020 doi:10.20944/preprints202010.0365.v1

Like *KEAP1*, *ENC1* helps control the oxidative stress response as a post-translational regulator of the transcription factor Nrf2 [43] and is expressed in all tissues although much more highly in the frontal cortex of brain (Figure 3B). *ENC1* also plays an important, but poorly understood, role in prenatal neural development [44]. The most prominent postnatal brain-specific epigenetic marks were the upstream and downstream intergenic open chromatin, H3K27ac signal, and weak or strong enhancer chromatin (Figure 3B, dotted blue boxes). The higher level of expression of *ENC1* during prenatal brain formation than in adults [44] is matched by fetal brain-associated DNA hypomethylation, strong H3K27ac signal in the gene body, and a super-enhancer spanning the gene, all of which were also seen in NeuN/RBFOX3-antibody isolated postnatal neurons [45] or neuropheres (suspended *in vitro* clusters of neural progenitor cells; Figure 3B, dotted purple boxes). Non-neuronal cells from postnatal brain [45] lacked the prominent and specific DNA hypomethylation of neurons and fetal brain (Figure 4B and data not shown). Importantly, *ENC1* was the only one of the six *KLHL* family genes with brain-preferential expression that displayed a consistent difference in DNA methylation profiles with age, namely, a 35-day postnatal sample from frontal cortex displayed the same DNA hypomethylation seen in a 17-wk fetal cerebral cortex missing in samples of frontal cortex from 2-yo, 12-yo, 16-yo, 53-yo, 55-yo, and 64-yo samples (Figure 3B and data not shown). Two population studies of genetic and epigenetic correlates of neurological function gave evidence for differential methylation of several CpGs 1 to 2 kb upstream of the *ENC1* TSS being associated with cognitive decline or schizophrenia (Figure 3B, orange and blue arrows, positions of CpGs in studies by White et al. [46] and van den Oord et al. [47], respectively). Interestingly, these CpGs are in a region that shows the most enhancer chromatin signal and the lowest DNA methylation in fetal brain (Figure 3B).

A caveat in studies of DNA methylation in brain is that bisulfite-seq, like most DNA methylation analysis techniques, cannot distinguish between 5mC and 5hmC and brain tissue is particularly rich in genomic 5hmC [45]. Other tissues, like SkM, are also enriched in 5-hydroxymethylcytosine (5hmC) at specific loci [48]. Nonetheless, the main attributes of DNA methylation patterns that we observed are tissue-specific differences in the presence of low-methylated regions, which can be detected independently of whether the surrounding DNA is rich in 5mC or 5hmC. In addition, complications like allele-specific differences in DNA methylation [49] will not affect our search for regional DNA hypomethylation.

ENC1 displays moderately high expression and much enhancer chromatin in several primary cultures or progenitor cells unrelated to brain (Table S2). This is in accord with reported additional developmental roles for *ENC1* in non-brain tissues [50]. Expression of *ENC1* is unlikely to be affected by expression of a long isoform of *HEXB* (*ENST00000511181*), which is antisense to *ENC1* and whose TSS overlaps the first exon of *ENC1*, because this RNA is expressed at only very low levels in brain (Fig. 3B).

2.7 The strong brain-specificity of KLHL32 is mirrored by clusters of intragenic enhancer chromatin seen only in brain

The steady-state RNA levels from *KLHL32*, a gene which has received very little attention, ranged from TPM 7.6 to 25.2 in spinal cord or different brain tissues in contrast to 0 to 2.6 for 39 other tissues, and expression was only barely detectable in various cell cultures unrelated to brain (Supplementary Table S2). In accord with its nervous system-related transcription, *KLHL32* exhibited brain-specific enhancer chromatin segments overlapping brain-specific DNase-seq peaks and H3K27ac peaks throughout its long gene body (Figure 4A, chromatin state tracks; H3K27ac and DNase-seq, blue boxes). In addition, all brain samples had a broadened promoter chromatin region at the TSS (Figure 4A). Neurons isolated from two individuals displayed a large broadening of the DNA hypomethylation at the CGI promoter (Figure 4A). This hypomethylation was not seen in nonneuronal cells [45]

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2020 doi:10.20944/preprints202010.0365.v1

from the brains of the same individuals nor in fetal brain. Fetal brain also lacked most of the enhancer chromatin of postnatal brain. The main TSS of *KLHL32* [19] is 25 kb from the TSS of *NDUFAF4*, which encodes a mitochondrial NADH:Ubiquinone oxidoreductase complex assembly factor. *NDUFAF4* is broadly expressed but its RNA levels are highest in cerebellum and frontal cortex (TPM, 31 for both brain regions vs. a median TPM of 2.5 for non-brain tissues). *NDUFAF4* displayed brain-specific enhancer chromatin immediately upstream of its promoter (Figure 4A). Therefore, the brain-specificity of *KLHL32* has its counterpart in the brain-preferential expression and epigenetics of *NDUFAF4*, although it is unclear if they influence each other because in skin fibroblasts, they are in different TADs (data not shown).

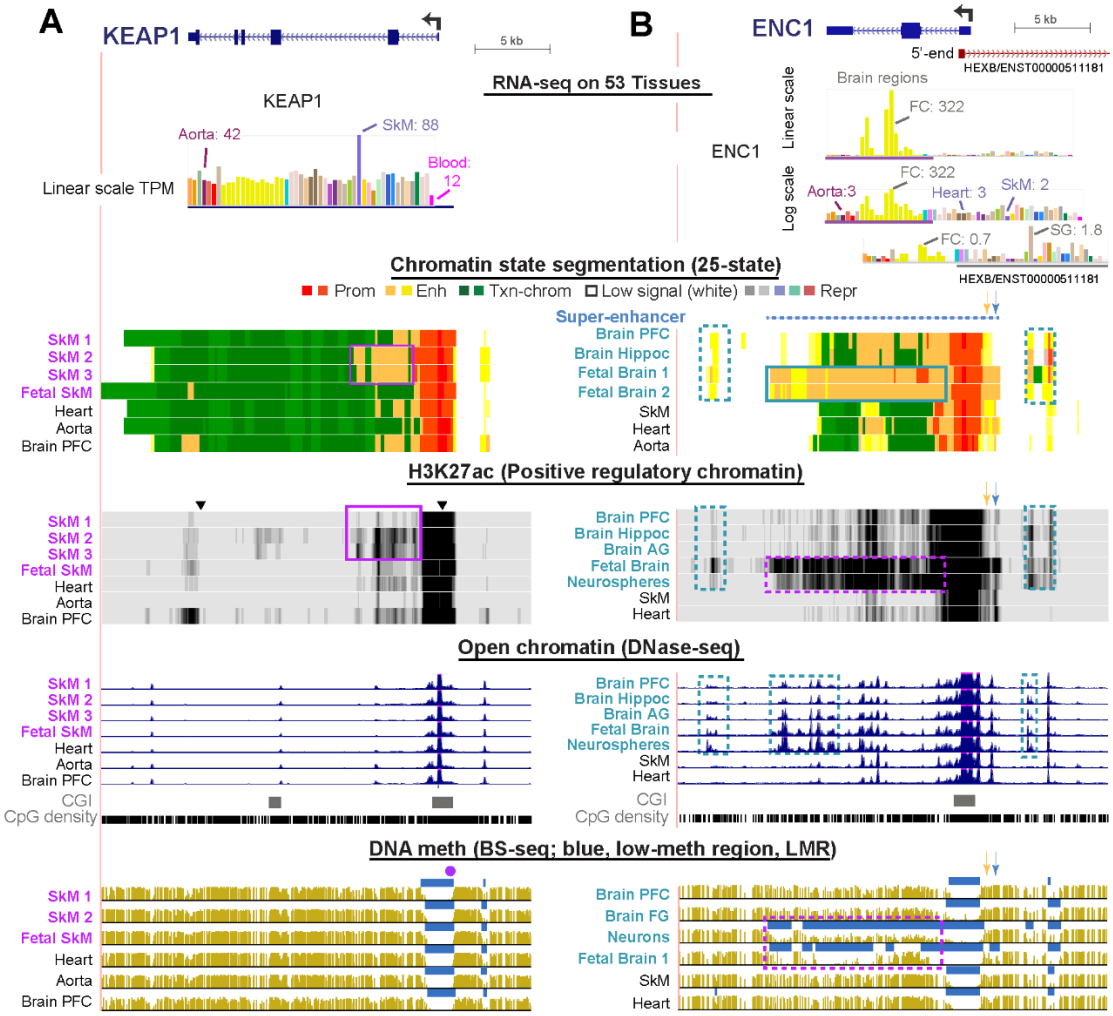

Figure 3. Epigenetic profiles for *KEAP1*/*KLHL19* and *ENC1*/*KLHL37*, which encode critical proteins that bind to NRF2 and regulate adaptation of cells to oxidative stress. (A) *KEAP1* (chr19:10,591,087-10,619,523) and **(B)** *ENC1* (chr5:73,908,815-73,950,517). Tracks are shown as in Figure 1. SkM 1 refers to psoas muscle while SkM 2 and SkM 3 are SkM samples obtained from unspecified leg locations [4,15].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2020 doi:10.20944/preprints202010.0365.v1

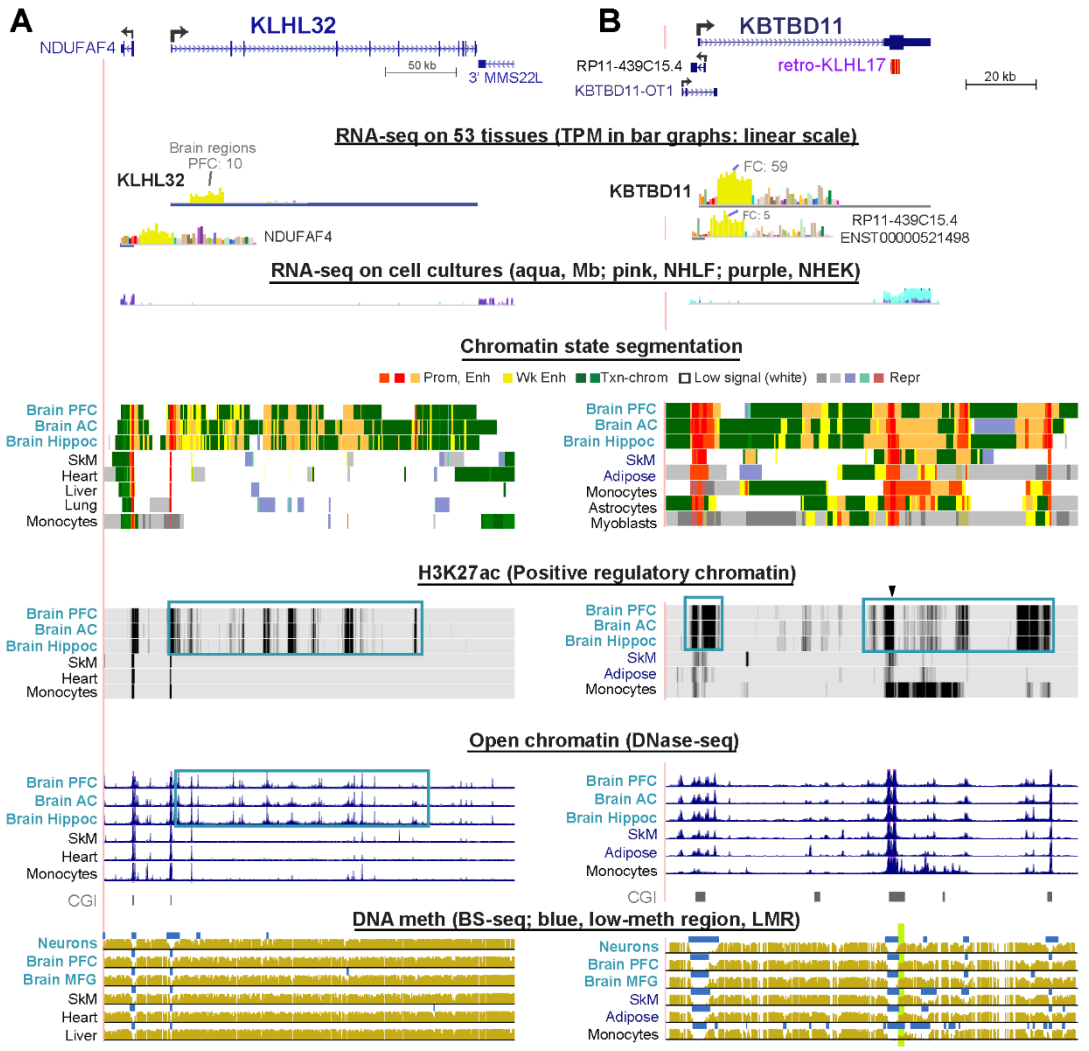

Figure 4. Multiple segments of enhancer chromatin without brain-specific DNA hypomethylation are associated with highly preferential expression in brain of *KLHL32* and *KBTBD11*. (A) *KLHL32* (chr6:97,324,594-97,615,081) and **(B)** *KBTBD11* (chr8:1,917,290-1,976,132). Tracks are as in Figure 1. The GTEx RNA-seq bar graphs are in linear scale but the scale for *KBTBD11* is different from that of its 5′ overlapping antisense transcript RP11-439C15.4 as indicated by the different TPM values for frontal cortex. The green highlighting in the DNA methylation tracks for *KBTBD11* refers to a region with leukemia-subtype-related DNA hypomethylation as previously reported [51] and explained in the Discussion.

2.8 KBTBD11, a brain-specific gene, has a retrogene overlaying a 3′ promoter for a novel noncoding RNA gene

KBTBD11 has been little studied in brain but is implicated in adipocyte differentiation [52]. It has moderately low expression in normal subcutaneous or visceral adipose tissue (TPM, 4 – 6), as in most other non-brain tissues. We found that this gene is preferentially expressed in all brain tissues (TPM, 23 – 59; Supplemental Table S2) and expression correlated with brain-specific enhancer chromatin and H3K27ac enrichment (Figure 4B, blue box). Fetal brain, for which expression data are not available, had much less enhancer chromatin than postnatal brain (data not shown). Even though there was negligible expression of *KBTBD11* mRNA in SkM, there

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2020 doi:10.20944/preprints202010.0365.v1

was moderate and selective expression in myoblasts (Supplemental Table S2) and osteoblasts (UCSC Genome Browser, data not shown). Much of this transcription in myoblasts and osteoblasts initiates from the last exon of *KBTBD11*, in the sense direction, as indicated by analysis of 5′ cap ends of RNA (5′ Cap Analysis Gene Expression, CAGE, not shown [24]) and confirmed by RNA-seq and active promoter chromatin at the end of the gene (Figure 4B, chromatin state). Besides myoblasts and osteoblasts, monocytes displayed especially strong expression of just this 3′ exonic ncRNA, which is consistent with monocytes' unusually long region of promoter chromatin at the 3′ end of the gene (Figure 4B, chromatin state). This 3′ region of *KBTBD11* overlaps a CGI with constitutive unmethylation, and this unmethylated region is broader specifically in monocytes (Figure 4B, bottom). Monocytes displayed clusters of TSSs for the plus strand in this region but not at the 5′ end of *KBTBD11* (CAGE data, not shown). The ORF of *KBTBD11* is fully contained in the final exon of this gene. Curiously, most of this ORF overlaps a *KLHL*-derived retrogene, *retro-KLHL17* (Figure 4B, top). However, the ORF of this retrogene deviates so much from that of its parent gene that it is not classified as an expressed shuffle retrogene (Supplemental Table S3). The above-mentioned 3′ promoter chromatin, which was seen in many types of tissues and cell cultures, overlaps the retrogene. A central CTCF binding site was observed in the retrogene in some samples regardless of their *KBTBD11* expression status (Figure 4B, triangle over H3K27ac tracks [25]).

This undocumented monocyte/myoblast/osteoblast ncRNA transcribed from the 3′ end of *KBTBD11* may have some unknown function, possibly encoding an miRNA sponge [3]), given its cell type-specific transcription which is unrelated to the tissue/cell specificity of *KBTBD11* expression. There is, in addition, an antisense ncRNA gene *RP-11-439C15.4* overlapping the beginning of the first intron of *KBTBD11* (Figure 4B, top). Its tissue-specific expression profile is very similar to that of *KBTBD11*, although it is expressed at lower levels and so probably helps regulate the transcription of its overlapping protein-encoding gene. We could find no evidence of transcription of another ncRNA gene *KBTBD11-OT1* (Figure 4B, top), in the studied tissues or cell lines.

2.9 Overview of epigenetic features associated with the studied KLHF family genes

As shown in Table 3, enhancer chromatin, expanded promoter chromatin at the 5′ end, and/or DNA hypomethylation specific for SkM or brain correlated with the preferential expression of 17 *KLHL* family genes in SkM or brain. SkM or brain DNaseI hypersensitive sites usually, but not always, were seen in the tissue-specific enhancer or promoter chromatin and were frequently at SkM- or brain-specific peaks of H3K27ac signal (Figures 1 – 4). Posttranscriptional regulation was not examined in this study but is, of course, widespread. MicroRNAs (miRNAs) fine-tuning mRNA levels of *KLHL* family genes were previously described for six of these genes examined in our study (*KEAP1*, *ENC1*, *KLHL2*, *KLHL31*, *KBTBD11* and *KBTBD12*, Supplemental Table S5) but despite the importance of miRNAs regulating mRNA stability and translation, the epigenetic contribution to controlling transcription of these *KLHL* family genes was clear.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2020 doi:10.20944/preprints202010.0365.v1

Table 3. Epigenetic patterns associated with tissue-specific expression of *KLHL* genes in skeletal muscle or brain[a]

Epigenetic profile specifically associated with SkM or brain	Genes preferentially expressed in SkM (some genes also in heart)	Genes preferentially expressed in brain
Prom chromatin upstream of the TSS	*KLHL31, 33, 34*	None
Prom chromatin downstream of the TSS	*KLHL30, 31, 33, 38, 41*	*KLHL32, 35, KBTBD11*
Broadening of a constitutively unmethylated region at the TSS	*KLHL21, 33, 34, 40; KBTBD12*	*KLHL2,4, 32* (neurons[c]); *ENC1, KBTBD11* (neurons/fetal brain)
DNA hypometh upstream of the TSS[b]	*KLHL30, 31, 33, 34, 38, 40, KBTBD13*	*KLHL32* (neurons)
DNA hypometh downstream of TSS[b]	*KLHL21, 30, 31, 33, 41; KBTBD12*	*KLHL2, 4, 32* (neurons)
Super-enhancer	*KLHL 21, 30, 31, 38, 40, 41*	*ENC1*
Intragenic enh chromatin	*KLHL34, 30, 31, 33, 34, 38, 40, 41; KEAP1; KBTBD12, 13*	*KLHL2, 4, KBTBD11, ENC1*
Gene-upstream intergenic enh chromatin	*KLHL21, 30, 31, 33, 34, 38, 40, 41*	*KLHL32, KBTBD11, ENC1* (neurons & fetal brain)
Gene-downstream intergenic enh chromatin	*KLHL21, 38, KBTBD12*	*KBTBD11, ENC1* (fetal brain)
DNA hypomethylation in enh chromatin	*KLHL21, 30, 31, 33, 34, 38, 40, 41, KBTBD12, 13*	*KLHL2* (hippocampus, anterior caudate)

[a]The epigenetic features for the listed *KLHL* and *KBTBD* genes are shown in Figures 1-4 and S1-S6, S8-S10. Only epigenetic features correlated with preferential expression in skeletal muscle (SkM; sometimes also in heart) or in brain are shown. Prom, promoter; enh, enhancer; hypometh, hypomethylation; TSS, transcription start site; constitutive unmethylated region, mostly or completely unmethylated in all or almost all the >15 examined tissues

[b]DNA hypomethylation immediately upstream or downstream of the TSS

[c]Enhancer and promoter chromatin states are not available for neurons

3. Discussion

The preferential expression of 17 *KLHL* family genes in SkM or brain (Tables 1 and 2 and Supplemental Tables S1 and S2) is matched by the SkM- or brain-specific intragenic enhancer chromatin seen for almost all of these genes (Table 3). Their epigenetic specificity is consistent with their known tissue-specific roles in targeting certain proteins for degradation [6]. *KLHL35* and *KLHL4,* which are preferentially expressed in brain, exhibited little or no intragenic enhancer chromatin, respectively, but they had only modest levels of RNA in brain (Supplemental Table S2). Most of the examined *KLHL* family genes also displayed SkM- or brain-specific enhancer chromatin upstream or downstream of the gene. The intergenic and intragenic enhancer chromatin usually contained foci of DNA hypomethylation. Tissue-specific enhancer chromatin was more common than tissue-specific promoter chromatin, as has been found for some other gene families [16]. About half of these genes displayed SkM- or brain-associated

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2020 doi:10.20944/preprints202010.0365.v1

demethylation extending upstream and/or downstream from a constitutively unmethylated CGI at the promoter region.

Promoter-enhancer interactions that boost transcription initiation are mediated by looping of the enhancer to its target promoter, which may facilitate their sharing of bound TFs and cis-acting short-lived ncRNAs [36]. Such sharing is facilitated by a local lack of DNA methylation but is interrupted by insulator-type loop boundaries. Therefore, the enhancer-promoter pairs should be in the same topologically associating domain (TAD) for effective enhancer action [53]. We found evidence consistent with long-distance enhancer-promoter interactions that involve an enhancer within one gene and a promoter of another gene, namely, *KLHL40* with *HHATL* and *KLHL38* with *FBXO32* (Figures 1B and 2B; Supplemental Figures S5 and S8). The 3′ end of *KLHL40* is only 0.1 kb from the 3′ end of *HHATL* and both have high and preferential expression in SkM but *HHATL* is also expressed in heart. HHATL (Hedgehog acyltransferase like) is needed for normal postnatal skeletal muscle maturation [32] while mutation of *KLHL40* is linked to Nemaline Myopathy 8 and Severe Congenital Nemaline Myopathy reflecting the role of KLHL40 in regulating skeletal muscle development [29]. The TSS-downstream promoter chromatin of *KLHL40* in SkM appears as enhancer chromatin associated with *HHATL*-upregulation in heart (Figure 1B). This dual function of a region as a promoter or enhancer depending on the tissue type is consistent with recent studies [54]. In addition, we propose that, in SkM, there are cooperative interactions between enhancers overlapping *KLHL40* and *HHATL* to potentiate expression of both genes, just as cooperative enhancer-enhancer interactions have been inferred elsewhere in the human genome [55].

KLHL38 and *FBXO32* (Muscle atrophy F-box Protein) (Figure 2B) are another gene pair that we predict have shared enhancer-enhancer interactions. These interactions could span their shared 103-kb intergenic region that overlaps multiple MyoD sites and SkM-specific DNA hypomethylated regions (Figure 2B). Both proteins encoded by these genes are involved in ubiquitin-mediated protein degradation. FBXO32 is well known for its major role in SkM atrophy [56]. In a rat castration-reversal model of muscle atrophy, *Klhl38* was upregulated while its neighbor *Fbxo32* was downregulated [37]. However, in humans, both genes have their highest expression in normal SkM and their intragenic and intergenic enhancer chromatin could positively interact with each other (Figure 2B, Supplemental Figure 8). Consistent with cooperative regulation of these genes, both *klhl38* (*klhl38b*) and its neighbor *fbxo32* in zebrafish are upregulated by glucocorticoids and may be involved in glucocorticoid-mediated muscle atrophy [56].

The most dramatic enhancer cooperation is seen in super-enhancers, which are large structures consisting of clusters of neighboring enhancers (or enhancers and promoters) that strongly upregulate expression of differentiation-related genes [27], including SkM-associated genes [15,16]. Seven of the 17 SkM- or brain-associated *KLHL* family genes exhibited expression-linked super-enhancers (Table 3). Four of these, *KLHL40, KLHL41, KLHL31,* and *ENC1,* are known to be important for SkM or brain development [29,44,57]. The other three genes, *KLHL21, KLHL30,* and *KLHL38*, have been the subject of only a few studies. The presence of super-enhancers in SkM and the high expression in myoblasts and myotubes or, for *KLHL38,* just in myotubes (Table 1) as well as in postnatal SkM suggests that they have as yet unrecognized roles in muscle development and in SkM homeostasis. The above-postulated *KLHL38-FBXO32* enhancer interactions in SkM would be between two neighboring super-enhancers. Super-enhancers may help to compartmentalize transcription regulatory factors by a type of localized phase separation [58]. Focal cell/tissue-specific DNA hypomethylation is considered to be an important hallmark of super-enhancers (and traditional enhancers [59]) by helping to facilitate the concentration of transcription factors at some of their most influential subregions. We found such SkM-specific hypomethylated

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2020 doi:10.20944/preprints202010.0365.v1

foci in SkM super-enhancers, except for *KLHL41,* which has a super-enhancer that consisted mostly of unusually extensive promoter chromatin that was hypomethylated over most of its length. Both epigenetic features may reflect *KLHL41*'s extremely high expression (Figure 1A).

Surprisingly, a promoter chromatin region overlapped the last exon of *KBTBD11* (Figure 4B). This 3' promoter chromatin likely evolved from a retrotransposed copy (retrogene) of *KLHL17*. Also, *KLHL31* similarly has a *KLHL* retrogene (*retro-KLHL36*) at its 3' end that overlaps an ORF (Supplemental Figure S2). The *KLHL31* retrogene overlaps exonic promoter chromatin in *KLHL31*-expressing and non-expressing tissues. Importantly, both the *KLHL31* and *KBTBD11* intragenic retrogenes are in CGIs that are constitutively unmethylated, and so both retrotransposed sequences are predisposed to forming intragenic promoters [60], and, indeed both had evidence of transcription initiation in some normal cell cultures by CAGE profiling. Of the nine SkM- or brain-associated *KLHL* genes that contained a *KLHL* retrogene in their exonic DNA (Supplemental Table S3), only one other gene, *KLHL21*, had a constitutively unmethylated CGI overlapping the retrogene, but this retrogene was in the first exon rather than at the 3' end of the host gene. Therefore, it is not surprising that the retrogene in *KLHL21* overlapped promoter chromatin in all tissues (Supplemental Figure S10). The three other *KLHL* family genes (*KLHL17, KLHL25*, and *KLHL26*) that contained intragenic CGI-overlapping *KLHL* retrogenes had their retrogenes embedded in transcription-type chromatin (enriched in H3K36me3) that is highly and constitutively methylated in the CGI. This constitutive DNA methylation is consistent with evidence for DNA methylation repressing the latent promoter activity of intragenic CGIs [61], just as hypermethylation of constitutively unmethylated and active promoter regions overlapping CGIs almost invariably leads to repression [62].

The importance of having normal levels of expression of *KLHL* family genes is supported by siRNA knockdown or overexpression studies in which decreases or increases in protein levels of the *KLHL* family genes impacted normal cellular function (e.g., for *KLHL40*, *KLHL41, KEAP1, ENC1* and *KBTBD11* [30,63-66]). Chromatin and DNA epigenetics are frequently involved in setting or maintaining gene expression levels [4,16,62]. In addition, fine-tuning of RNA abundance by downregulation at the post-transcriptional level is often mediated by miRNAs. Such post-transcriptional regulation is attested to by reports of miRNA regulation of *KEAP1, ENC1, KLHL2, KLHL31, KBTBD11*, and *KBTBD12* expression (Supplemental Table S5).

Most epigenetic studies of *KLHL* family genes focused on cancer-associated promoter hypermethylation, especially, hypermethylation of the *KEAP1* CGI-promoter during carcinogenesis or acquisition of drug resistance [6]. Under homeostatic conditions, KEAP1 is a major regulator of antioxidant and metabolic genes by continuously ubiquitinating and targeting for degradation the oxidative-stress responsive transcription factor, NFE2L2/NRF2 [40]. Increases in NRF2 levels can result from KEAP1 inactivation by the reaction of electrophiles or oxidants with cysteine residues in KEAP1 or by disruption of the KEAP1:NRF2 complex [21]. It can also occur by extensive *KEAP1* promoter hypermethylation, as has been described in diverse cancers [67,68]. The resulting increase in NRF2 protein levels can affect tumor initiation, tumor progression, and drug resistance. It has been reported that there is cataract-associated DNA hypomethylation in lens epithelial cells at the TSS-upstream portion of the 5' *KEAP1* CGI [67], which is surprising because the analyzed region described as highly methylated in normal lens had low levels of methylation in more than 20 examined human tissues with bisulfite-seq profiles at the UCSC Genome Browser (Figure 3A, bottom, purple circle; [24]). However, a methylome profile for lens tissue was not available and *KEAP1* methylation might be exceptional in this subregion. What is clear is that cellular levels of KEAP1 must be tightly regulated and the first step in such regulation is by epigenetic control of the gene's transcription.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2020 doi:10.20944/preprints202010.0365.v1

Examination of *KBTBD11* epigenetics gives another example of how using publicly databases can elucidate epigenetic changes in normal and diseased tissues. In one of the few studies of non-promoter epigenetics of *KLHL* family genes, Kachroo et al. [51] found, in a genome-wide study, that a 0.8-kb region in the last exon of *KBTBD11* was significantly hypomethylated in lymphoblasts from untreated pediatric patients with a poor-prognosis subtype of B-cell leukemia relative to those with a good-prognosis leukemia subtype. DNA hypomethylation at this 0.8 kb region in the poor prognosis-subtype correlated with increased expression of *KBTBD11*. We found that this hypomethylated region in the ORF of the last exon of *KBTBD11* was hypomethylated in normal monocytes (Figure 4B, bottom, green highlighting) relative to other studied normal samples (Figure 5B, bottom, green highlighting). Moreover, the monocyte hypomethylation is associated with production of a novel ncRNA from the 3'UTR with a different tissue-specificity from that of *KBTBD11*. Therefore, the epigenetics, transcription, and function of this previously unknown ncRNA gene encoded at the 3' retrocopy promoter of *KBTBD11* should be examined in future studies of normal and cancer samples. Our results indicate the need for follow-up studies of the effects of experimental manipulation of some of the most tissue-specific of the enhancer chromatin regions that we characterized, especially those that are correlated with disease.

4. Methods

4.1. RNA-seq for tissues and cells

TPM values for RNA levels for tissues were from the GTEx RNA-seq database (Supplemental Table S2) [19]. In that database median TPM values are given from analysis of hundreds of samples for each of 52 tissue types. We did not include values for skin fibroblasts and lymphoblast cultures, which are also included in that database. For genes with more than one isoform, except where otherwise noted, only the main transcribed isoform is shown in the figures as deduced from GTEx isoform expression data and, for isoforms with separate promoters, the position of TSS-overlapping promoter chromatin or of the TSS deduced by 5' Cap Analysis Gene Expression (CAGE; RIKEN Omics Science Center [24]). RNA-seq data (FPKM) for six human cell cultures are included in Supplemental Table S2. Data are for myoblasts, GM12878 (a lymphoblastoid cell line), ESC (embryonic stem cells), HUVEC (human umbilical vein endothelial cells), NHEK (normal human epidermal keratinocytes), and NHLF (normal human lung fibroblasts) [24]. For quantitation of cell culture RNA-seq data, we determined the signal for each gene from ENCODE data [24] and from our analyses with Cufflinks [69,70]. For comparisons of RNA levels in myoblasts and myotubes, we used our previously generated data [16].

4.2. Databases and analyses used for epigenetics studies

The 18-state Roadmap Epigenomics chromatin state segmentation analysis (chromHMM, AuxilliaryHMM) [4]) was used for determination of chromatin states (promoter, enhancer, repressed, *etc*.) except for *KEAP1* and *ENC1* in which the 25-state analysis was used because fetal brain was included only in that dataset. The color code for chromatin state segmentation in the figures was slightly simplified from the original as shown in the color keys in the figures. Bisulfite-seq profiles of genome-wide CpG methylation and the DNaseI-hypersensitivity profiles were also from the RoadMap Project [4]. For bisulfite-seq samples of SkM, heart, and brain, several biological replicates were available as follows: SkM, psoas from a 3 yo male, 30 yo female, or 34 yo male; left or right ventricle from the aforementioned 3 yo and 30 yo samples or right atrium from the 34 yo; brain prefrontal or midfrontal cortex from seven female or male donors ranging from fetal to 55 y. Except for a few genes preferentially

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2020 doi:10.20944/preprints202010.0365.v1

transcribed in brain, similar results were obtained from these biological replicates unless otherwise noted. With the exception of brain, and sometimes SkM, generally the same samples that had been used for bisulfite-seq were used for chromatin state segmentation (and the associated H3K27ac signal) and DNase-seq. SkM 1 refers to psoas (combined 3 yo and 34 yo) and SkM 2 and 3 to leg muscle (unspecified) from a 72 yo female and a 54 y male, respectively. Fetal muscle was from a 15-week post gestational male or femaie. In the epigenetic tracks in the figures, SkM and heart with no further designation refer to mixtures of the 3 yo and 34 yo samples of psoas muscle or left ventricle. Further descriptions of other samples used for these epigenetic analyses were given previously [4,48]. Super-enhancers were assessed by dbSUPER [71] and confirmed by looking at the H3K27ac track in the UCSC Genome Browser [24]. MYOD binding sites in myoblasts were determined from the Unibind data track in the UCSC browser [24,25]. Low methylated regions (LMRs) shown in the figures refer to regions with significantly lower DNA methylation than in the rest of the same genome as determined by Song et al. [72]. Transcription associating domains (TADs) were obtained using the Micro-C tracks at the UCSC browser for foreskin fibroblasts (Krietenstein et al., 2020). A high score between two regions suggests that they are probably in proximity in 3D space within the nucleus of a cell as shown by a more intense color in the heatmap [73].

4.3. Alignments and phylogenetic analysis

Alignments of the main isoform for each of the *KLHL* family genes was done using the phylogenetic tree view in COBALT (COnstraint-Based multiple ALignment Tool, with the default parameters: fast minimum evolution., Max seq difference 0.85, distance, Grishin (for protein) [74] https://www.ncbi.nlm.nih.gov/tools/cobalt/cobalt.cgi?LINK_LOC=BlastHomeLink.

5. Conclusions

Analysis of *KLHL* family genes preferentially expressed in SkM or brain showed that differences in the amount of enhancer chromatin are more likely to be indicative of tissue-specific differences in mRNA levels than are promoter chromatin differences. However, DNA hypomethylation extending from unmethylated CpG-rich promoters to immediate upstream and downstream regions was often seen in tissues with the strongest expression of the gene. Occasionally active intragenic promoters were associated with expression of novel ncRNA genes that had a different tissue specificity than that of the host protein-coding gene. Lastly, not only were *KLHL* intragenic and intergenic regions associated with tissue-specific expression, but also, several *KLHL* family genes gave epigenomic/transcriptomic evidence for intragenic and intergenic enhancers in adjacent genes exerting *cis* effects on transcription of their neighbors.

Supplemental Material: Supplemental files can be found at www.mdpi.com/xxx/s1.

Table S1. Disease associations and protein targets of *KLHL* family genes

Table S2. Expression levels of *KLHL* family genes in 52 tissues and six types of cell cultures

Table S3. *KLHL* family genes containing retrotransposed genes (retrogenes) in coding exons derived from other *KLHL* family genes

Table S4. Retrotransposed genes whose open reading frame contributes to the host gene's open reading frame (expressed shuffle retrogenes): Six *KLHL* genes are among the 741 genes containing expressed shuffle retrogenes

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2020 doi:10.20944/preprints202010.0365.v1

Table S5. Previous reports of microRNA regulation of *KLHL* family genes in this study

Figure S1. Preferential expression in skeletal muscle of the little-studied gene, *KLHL33* is associated with spreading of DNA hypomethylation from a CpG island at the transcription start site

Figure S2. Preferential expression in heart and skeletal muscle of *KLHL31*, a gene implicated in congenital myopathy

Figure S3. Expression of *KLHL34*, a little studied gene preferentially expressed in skeletal muscle and heart, mirrors that of its upstream neighbor *SMPX* (small muscle protein X-linked gene)

Figure S4. *KBTBD12* has an intragenic enhancer chromatin region in intron 4 that likely contributes to its preferential expression in skeletal muscle and heart

Figure S5. Genes surrounding *KLHL40* can share a topologically associating domain (TAD)

Figure S6. *KBTBD13*, like *KLHL40* and *KLHL41*, is associated with nemaline myopathy

Figure S7. Phylogenetic tree of full-length proteins encoded by *KLHL* and *KBTBD* genes

Figure S8. *KLHL38* and *FBXO32* can share a topologically associating domain

Figure S9. Epigenetics of three additional *KLHL* genes preferentially expressed in brain (*KLHL2*, *KLHL4* and *KLHL35)*

Figure S10. *KLHL21*, a gene required for cytokinesis, is widely expressed although with highest expression in skeletal muscle and has the most DNA hypomethylation and enhancer chromatin in skeletal muscle

Author Contributions: K.C.E. and M.E conceived the project, did the analyses on epigenetic relationships, and wrote the paper; C.B. did the analyses on retrogenes**.**

Funding: This research was supported in part by grants to ME from the National Institutes of Health (NS04885) and the Louisiana Cancer Center.

Acknowledgments: This research was supported in part using services provided by Information Technology at Tulane University, New Orleans, LA.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

AC	anterior caudate (brain)
AG	angular gyrus (brain)
BACK	BTB and C-terminal Kelch) domain
BS	bisulfite
CAGE	5' Cap Analysis Gene Expression
Cereb	brain cerebellum
CGI	CpG island
CG	brain cingulate gyrus
ChIP-seq	chromatin immunoprecipitation-sequencing
Chrom	chromatin
Enh	enhancer
ESC	embryonic stem cells

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2020 doi:10.20944/preprints202010.0365.v1

FC	frontal cortex (brain)
FG	frontal gyrus (brain)
FPKM	fragments per kilobase of exon per million reads mapped
H3K27ac	Histone H3 lysine 27 acetylation
H3K4me1	histone H3 lysine 4 monomethylation
H3K4me3	histone H3 lysine 4 trimethylation
Hippoc	hippocampus (brain)
KBTBD	Kelch and BTB domain
KLHL	Kelch-like gene family members; BTB, bric-a-brac, tram track broad complex; POZ, poxvirus and zinc finger
LCL	lymphoblastoid cell line
LMR	low methylated region
Micro-C	chromatin capture analysis using micrococcal nuclease
miRNA	microRNA
ncRNA	non-coding RNA
Neurosph	neurospheres (cortex derived);
NHLF	normal human lung fibroblasts
ORF	open reading frame
PBMC	peripheral blood mononuclear cells
PFC	prefrontal cortex (brain)
siRNA	small interfering or silencer RNA
SkM	skeletal muscle
SN	brain substantia nigra
TAD	topologically associating domain
TPM	transcripts per kilobase million
TSS	transcription start site
Txn	transcription
Wk	weak
Yo	year-old

References

1. Guo, Y.; Yu, S.; Zhang, C.; Kong, A.N. Epigenetic regulation of Keap1-Nrf2 signaling. *Free Radic Biol Med* **2015**, *88*, 337-349, doi:10.1016/j.freeradbiomed.2015.06.013.
2. Varshavsky, A. The Ubiquitin System, Autophagy, and Regulated Protein Degradation. *Annu Rev Biochem* **2017**, *86*, 123-128, doi:10.1146/annurev-biochem-061516-044859.
3. Mumtaz, P.T.; Taban, Q.; Dar, M.A.; Mir, S.; Haq, Z.U.; Zargar, S.M.; Shah, R.A.; Ahmad, S.M. Deep Insights in Circular RNAs: from biogenesis to therapeutics. *Biol Proced Online* **2020**, *22*, 10, doi:10.1186/s12575-020-00122-8.
4. Roadmap_Epigenomics_Consortium; Kundaje, A.; Meuleman, W.; Ernst, J.; Bilenky, M.; Yen, A.; Heravi-Moussavi, A.; Kheradpour, P.; Zhang, Z.; Wang, J., et al. Integrative analysis of 111 reference human epigenomes. *Nature* **2015**, *518*, 317-330, doi:10.1038/nature14248.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2020 doi:10.20944/preprints202010.0365.v1

5. Dhanoa, B.S.; Cogliati, T.; Satish, A.G.; Bruford, E.A.; Friedman, J.S. Update on the Kelch-like (KLHL) gene family. *Hum Genomics* **2013**, *7*, 13, doi:10.1186/1479-7364-7-13.
6. Shi, X.; Xiang, S.; Cao, J.; Zhu, H.; Yang, B.; He, Q.; Ying, M. Kelch-like proteins: Physiological functions and relationships with diseases. *Pharmacol Res* **2019**, *148*, 104404, doi:10.1016/j.phrs.2019.104404.
7. Gupta, V.A.; Beggs, A.H. Kelch proteins: emerging roles in skeletal muscle development and diseases. *Skelet Muscle* **2014**, *4*, 11, doi:10.1186/2044-5040-4-11.
8. Elshaer, M.; ElManawy, A.I.; Hammad, A.; Namani, A.; Wang, X.J.; Tang, X. Integrated data analysis reveals significant associations of KEAP1 mutations with DNA methylation alterations in lung adenocarcinomas. *Aging (Albany NY)* **2020**, *12*, 7183-7206, doi:10.18632/aging.103068.
9. Sewry, C.A.; Laitila, J.M.; Wallgren-Pettersson, C. Nemaline myopathies: a current view. *J Muscle Res Cell Motil* **2019**, *40*, 111-126, doi:10.1007/s10974-019-09519-9.
10. Jirka, C.; Pak, J.H.; Grosgogeat, C.A.; Marchetii, M.M.; Gupta, V.A. Dysregulation of NRAP degradation by KLHL41 contributes to pathophysiology in Nemaline Myopathy. *Hum Mol Genet* **2019**, 10.1093/hmg/ddz078, doi:10.1093/hmg/ddz078.
11. Zhang, Z.; Turer, E.; Li, X.; Zhan, X.; Choi, M.; Tang, M.; Press, A.; Smith, S.R.; Divoux, A.; Moresco, E.M., et al. Insulin resistance and diabetes caused by genetic or diet-induced KBTBD2 deficiency in mice. *Proc Natl Acad Sci U S A* **2016**, *113*, E6418-E6426, doi:10.1073/pnas.1614467113.
12. Yoshida, S.; Araki, Y.; Mori, T.; Sasaki, E.; Kasagi, Y.; Isobe, K.; Susa, K.; Inoue, Y.; Bomont, P.; Okado, T., et al. Decreased KLHL3 expression is involved in the pathogenesis of pseudohypoaldosteronism type II caused by cullin 3 mutation in vivo. *Clin Exp Nephrol* **2018**, *22*, 1251-1257, doi:10.1007/s10157-018-1593-z.
13. Hedberg-Oldfors, C.; Abramsson, A.; Osborn, D.P.S.; Danielsson, O.; Fazlinezhad, A.; Nilipour, Y.; Hubbert, L.; Nennesmo, I.; Visuttijai, K.; Bharj, J., et al. Cardiomyopathy with lethal arrhythmias associated with inactivation of KLHL24. *Hum Mol Genet* **2019**, *28*, 1919-1929, doi:10.1093/hmg/ddz032.
14. Deaton, A.M.; Bird, A. CpG islands and the regulation of transcription. *Genes Dev.* **2011**, *25*, 1010-1022.
15. Ehrlich, K.C.; Paterson, H.L.; Lacey, M.; Ehrlich, M. DNA hypomethylation in intragenic and intergenic enhancer chromatin of muscle-specific genes usually correlates with their expression. *Yale J Biol Med* **2016**, *89*, 441-455.
16. Ehrlich, K.C., Lacey, M., Ehrlich, M. Epigenetics of Skeletal Muscle-Associated Genes in the ASB, LRRC, TMEM, and OSBPL Gene Families. *Epigenomes* **2020**, *4*, 1-22, doi:10.3390/epigenomes4010001.
17. Heberle, E.; Bardet, A.F. Sensitivity of transcription factors to DNA methylation. *Essays Biochem* **2019**, *63*, 727-741, doi:10.1042/EBC20190033.
18. Lazaris, C.; Aifantis, I.; Tsirigos, A. On Epigenetic Plasticity and Genome Topology. *Trends Cancer* **2020**, *6*, 177-180, doi:10.1016/j.trecan.2020.01.006.
19. GTEx_Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **2015**, *348*, 648-660, doi:10.1126/science.1262110.
20. Illingworth, R.S.; Gruenewald-Schneider, U.; De Sousa, D.; Webb, S.; Merusi, C.; Kerr, A.R.; James, K.D.; Smith, C.; Walker, R.; Andrews, R., et al. Inter-individual variability contrasts with regional homogeneity in the human brain DNA methylome. *Nucleic Acids Res* **2015**, *43*, 732-744, doi:10.1093/nar/gku1305.
21. Dayalan Naidu, S.; Dinkova-Kostova, A.T. KEAP1, a cysteine-based sensor and a drug target for the prevention and treatment of chronic disease. *Open Biol* **2020**, *10*, 200105, doi:10.1098/rsob.200105.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2020 doi:10.20944/preprints202010.0365.v1

22. Pirinen, M.; Lappalainen, T.; Zaitlen, N.A.; GTEx_Consortium; Dermitzakis, E.T.; Donnelly, P.; McCarthy, M.I.; Rivas, M.A. Assessing allele-specific expression across multiple tissues from RNA-seq read data. *Bioinformatics* **2015**, *31*, 2497-2504, doi:10.1093/bioinformatics/btv074.
23. Searle, B.C.; Gittelman, R.M.; Manor, O.; Akey, J.M. Detecting Sources of Transcriptional Heterogeneity in Large-Scale RNA-Seq Data Sets. *Genetics* **2016**, *204*, 1391-1396, doi:10.1534/genetics.116.193714.
24. Haeussler, M.; Zweig, A.S.; Tyner, C.; Speir, M.L.; Rosenbloom, K.R.; Raney, B.J.; Lee, C.M.; Lee, B.T.; Hinrichs, A.S.; Gonzalez, J.N., et al. The UCSC Genome Browser database: 2019 update. *Nucleic Acids Res* **2019**, *47*, D853-D858, doi:10.1093/nar/gky1095.
25. Gheorghe, M.; Sandve, G.K.; Khan, A.; Cheneby, J.; Ballester, B.; Mathelier, A. A map of direct TF-DNA interactions in the human genome. *Nucleic Acids Res* **2019**, *47*, 7715, doi:10.1093/nar/gkz582.
26. Xi, H.; Shulha, H.P.; Lin, J.M.; Vales, T.R.; Fu, Y.; Bodine, D.M.; McKay, R.D.; Chenoweth, J.G.; Tesar, P.J.; Furey, T.S., et al. Identification and characterization of cell type-specific and ubiquitous chromatin regulatory structures in the human genome. *PLoS Genet* **2007**, *3*, e136.
27. Whyte, W.A.; Orlando, D.A.; Hnisz, D.; Abraham, B.J.; Lin, C.Y.; Kagey, M.H.; Rahl, P.B.; Lee, T.I.; Young, R.A. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **2013**, *153*, 307-319, doi:10.1016/j.cell.2013.03.035.
28. Khan, A.; Zhang, X. dbSUPER: a database of super-enhancers in mouse and human genome. *Nucleic Acids Res* **2016**, *44*, D164-171, doi:10.1093/nar/gkv1002.
29. Blondelle, J.; Tallapaka, K.; Seto, J.T.; Ghassemian, M.; Clark, M.; Laitila, J.M.; Bournazos, A.; Singer, J.D.; Lange, S. Cullin-3 dependent deregulation of ACTN1 represents a new pathogenic mechanism in nemaline myopathy. *JCI Insight* **2019**, *5*, doi:10.1172/jci.insight.125665.
30. Bowlin, K.M.; Embree, L.J.; Garry, M.G.; Garry, D.J.; Shi, X. Kbtbd5 is regulated by MyoD and restricted to the myogenic lineage. *Differentiation* **2013**, *86*, 184-191, doi:10.1016/j.diff.2013.08.002.
31. Cao, Y.; Yao, Z.; Sarkar, D.; Lawrence, M.; Sanchez, G.J.; Parker, M.H.; MacQuarrie, K.L.; Davison, J.; Morgan, M.T.; Ruzzo, W.L., et al. Genome-wide MyoD binding in skeletal muscle cells: a potential for broad cellular reprogramming. *Dev. Cell* **2010**, *18*, 662-674.
32. Van, B.; Nishi, M.; Komazaki, S.; Ichimura, A.; Kakizawa, S.; Nakanaga, K.; Aoki, J.; Park, K.H.; Ma, J.; Ueyama, T., et al. Mitsugumin 56 (hedgehog acyltransferase-like) is a sarcoplasmic reticulum-resident protein essential for postnatal muscle maturation. *FEBS Lett* **2015**, *589*, 1095-1104, doi:10.1016/j.febslet.2015.03.028.
33. Zhang, P.; Zhang, L.; Li, Y.; Zhu, S.; Zhao, M.; Ding, S.; Li, J. Quantitative Proteomic Analysis To Identify Differentially Expressed Proteins in Myocardium of Epilepsy Using iTRAQ Coupled with Nano-LC-MS/MS. *J Proteome Res* **2018**, *17*, 305-314, doi:10.1021/acs.jproteome.7b00579.
34. Hsieh, T.S.; Fudenberg, G.; Goloborodko, A.; Rando, O.J. Micro-C XL: assaying chromosome conformation from the nucleosome to the entire genome. *Nat Methods* **2016**, *13*, 1009-1011, doi:10.1038/nmeth.4025.
35. de Winter, J.M.; Molenaar, J.P.; Yuen, M.; van der Pijl, R.; Shen, S.; Conijn, S.; van de Locht, M.; Willigenburg, M.; Bogaards, S.J.; van Kleef, E.S., et al. KBTBD13 is an actin-binding protein that modulates muscle kinetics. *J Clin Invest* **2020**, *130*, 754-767, doi:10.1172/JCI124000.
36. Sartorelli, V.; Lauberth, S.M. Enhancer RNAs are an important regulatory layer of the epigenome. *Nat Struct Mol Biol* **2020**, *27*, 521-528, doi:10.1038/s41594-020-0446-0.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2020 doi:10.20944/preprints202010.0365.v1

37. de O'Coelho, P.; Guarnier, F.A.; Figueiredo, L.B.; Zaramela, L.S.; Pacini, E.S.A.; Godinho, R.O.; Gomes, M.D. Identification of potential target genes associated with the reversion of androgen-dependent skeletal muscle atrophy. *Arch Biochem Biophys* **2019**, *663*, 173-182, doi:10.1016/j.abb.2019.01.009.
38. Baertsch, R.; Diekhans, M.; Kent, W.J.; Haussler, D.; Brosius, J. Retrocopy contributions to the evolution of the human genome. *BMC Genomics* **2008**, *9*, 466, doi:10.1186/1471-2164-9-466.
39. Sukari, A.; Muqbil, I.; Mohammad, R.M.; Philip, P.A.; Azmi, A.S. F-BOX proteins in cancer cachexia and muscle wasting: Emerging regulators and therapeutic opportunities. *Semin Cancer Biol* **2016**, *36*, 95-104, doi:10.1016/j.semcancer.2016.01.002.
40. Bellezza, I.; Giambanco, I.; Minelli, A.; Donato, R. Nrf2-Keap1 signaling in oxidative and reductive stress. *Biochim Biophys Acta Mol Cell Res* **2018**, *1865*, 721-733, doi:10.1016/j.bbamcr.2018.02.010.
41. van Rooij, E.; Quiat, D.; Johnson, B.A.; Sutherland, L.B.; Qi, X.; Richardson, J.A.; Kelm, R.J., Jr.; Olson, E.N. A family of microRNAs encoded by myosin genes governs myosin expression and muscle performance. *Dev Cell* **2009**, *17*, 662-673, doi:10.1016/j.devcel.2009.10.013.
42. Bodine, S.C.; Baehr, L.M. Skeletal muscle atrophy and the E3 ubiquitin ligases MuRF1 and MAFbx/atrogin-1. *Am J Physiol Endocrinol Metab* **2014**, *307*, E469-484, doi:10.1152/ajpendo.00204.2014.
43. Wang, X.J.; Zhang, D.D. Ectodermal-neural cortex 1 down-regulates Nrf2 at the translational level. *PLoS One* **2009**, *4*, e5492, doi:10.1371/journal.pone.0005492.
44. Mesman, S.; Kruse, S.J.; Smidt, M.P. Expression analyzes of early factors in midbrain differentiation programs. *Gene Expr Patterns* **2018**, *27*, 8-15, doi:10.1016/j.gep.2017.09.001.
45. Lister, R.; Mukamel, E.A.; Nery, J.R.; Urich, M.; Puddifoot, C.A.; Johnson, N.D.; Lucero, J.; Huang, Y.; Dwork, A.J.; Schultz, M.D., et al. Global epigenomic reconfiguration during mammalian brain development. *Science* **2013**, *341*, 1237905, doi:10.1126/science.1237905.
46. White, C.C.; Yang, H.S.; Yu, L.; Chibnik, L.B.; Dawe, R.J.; Yang, J.; Klein, H.U.; Felsky, D.; Ramos-Miguel, A.; Arfanakis, K., et al. Identification of genes associated with dissociation of cognitive performance and neuropathological burden: Multistep analysis of genetic, epigenetic, and transcriptional data. *PLoS Med* **2017**, *14*, e1002287, doi:10.1371/journal.pmed.1002287.
47. van den Oord, E.J.; Clark, S.L.; Xie, L.Y.; Shabalin, A.A.; Dozmorov, M.G.; Kumar, G.; Swedish Schizophrenia, C.; Vladimirov, V.I.; Magnusson, P.K.; Aberg, K.A. A Whole Methylome CpG-SNP Association Study of Psychosis in Blood and Brain Tissue. *Schizophr Bull* **2016**, *42*, 1018-1026, doi:10.1093/schbul/sbv182.
48. Terragni, J.; Zhang, G.; Sun, Z.; Pradhan, S.; Song, L.; Crawford, G.E.; Lacey, M.; Ehrlich, M. Notch signaling genes: myogenic DNA hypomethylation and 5-hydroxymethylcytosine. *Epigenetics* **2014**, *9*, 842-850, doi:10.4161/epi.28597.
49. Do, C.; Lang, C.F.; Lin, J.; Darbary, H.; Krupska, I.; Gaba, A.; Petukhova, L.; Vonsattel, J.P.; Gallagher, M.P.; Goland, R.S., et al. Mechanisms and Disease Associations of Haplotype-Dependent Allele-Specific DNA Methylation. *Am J Hum Genet* **2016**, *98*, 934-955, doi:10.1016/j.ajhg.2016.03.027.
50. Worton, L.E.; Shi, Y.C.; Smith, E.J.; Barry, S.C.; Gonda, T.J.; Whitehead, J.P.; Gardiner, E.M. Ectodermal-Neural Cortex 1 Isoforms Have Contrasting Effects on MC3T3-E1 Osteoblast Mineralization and Gene Expression. *J Cell Biochem* **2017**, *118*, 2141-2150, doi:10.1002/jcb.25851.
51. Kachroo, P.; Szymczak, S.; Heinsen, F.A.; Forster, M.; Bethune, J.; Hemmrich-Stanisak, G.; Baker, L.; Schrappe, M.; Stanulla, M.; Franke, A. NGS-based methylation profiling differentiates TCF3-HLF and

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2020 doi:10.20944/preprints202010.0365.v1

TCF3-PBX1 positive B-cell acute lymphoblastic leukemia. *Epigenomics* **2018**, *10*, 133-147, doi:10.2217/epi-2017-0080.

52. Watanabe, K.; Yokota, K.; Yoshida, K.; Matsumoto, A.; Iwamoto, S. Kbtbd11 contributes to adipocyte homeostasis through the activation of upstream stimulatory factor 1. *Heliyon* **2019**, *5*, e02777, doi:10.1016/j.heliyon.2019.e02777.
53. Szabo, Q.; Bantignies, F.; Cavalli, G. Principles of genome folding into topologically associating domains. *Sci Adv* **2019**, *5*, eaaw1668, doi:10.1126/sciadv.aaw1668.
54. Andersson, R.; Sandelin, A. Determinants of enhancer and promoter activities of regulatory elements. *Nat Rev Genet* **2020**, *21*, 71-87, doi:10.1038/s41576-019-0173-8.
55. Chen, H.; Xiao, J.; Shao, T.; Wang, L.; Bai, J.; Lin, X.; Ding, N.; Qu, Y.; Tian, Y.; Chen, X., et al. Landscape of Enhancer-Enhancer Cooperative Regulation during Human Cardiac Commitment. *Mol Ther Nucleic Acids* **2019**, *17*, 840-851, doi:10.1016/j.omtn.2019.07.015.
56. Chen, Q.; Li, C.; Gong, Z.; Chun Yong Chan, E.; Snyder, S.A.; Lam, S.H. Common deregulated gene expression profiles and morphological changes in developing zebrafish larvae exposed to environmental-relevant high to low concentrations of glucocorticoids. *Chemosphere* **2017**, *172*, 429-439.
57. Papizan, J.B.; Garry, G.A.; Brezprozvannaya, S.; McAnally, J.R.; Bassel-Duby, R.; Liu, N.; Olson, E.N. Deficiency in Kelch protein Klhl31 causes congenital myopathy in mice. *J Clin Invest* **2017**, *127*, 3730-3740, doi:10.1172/JCI93445.
58. Wang, X.; Cairns, M.J.; Yan, J. Super-enhancers in transcriptional regulation and genome organization. *Nucleic Acids Res* **2019**, *47*, 11481-11496, doi:10.1093/nar/gkz1038.
59. Bell, E.; Curry, E.W.; Megchelenbrink, W.; Jouneau, L.; Brochard, V.; Tomaz, R.A.; Mau, K.H.T.; Atlasi, Y.; de Souza, R.A.; Marks, H., et al. Dynamic CpG methylation delineates subregions within super-enhancers selectively decommissioned at the exit from naive pluripotency. *Nat Commun* **2020**, *11*, 1112, doi:10.1038/s41467-020-14916-7.
60. Jeziorska, D.M.; Murray, R.J.S.; De Gobbi, M.; Gaentzsch, R.; Garrick, D.; Ayyub, H.; Chen, T.; Li, E.; Telenius, J.; Lynch, M., et al. DNA methylation of intragenic CpG islands depends on their transcriptional activity during differentiation and disease. *Proc. Natl. Acad. Sci. U S A* **2017**, *114*, E7526-E7535, doi:10.1073/pnas.1703087114.
61. Ponnaluri, V.K.; Ehrlich, K.C.; Zhang, G.; Lacey, M.; Johnston, D.; Pradhan, S.; Ehrlich, M. Association of 5-hydroxymethylation and 5-methylation of DNA cytosine with tissue-specific gene expression. *Epigenetics* **2017**, *12*, 123-138, doi:10.1080/15592294.2016.1265713.
62. Ehrlich, M. DNA hypermethylation in disease: mechanisms and clinical relevance. *Epigenetics* **2019**, *14*, 1141-1163, doi:10.1080/15592294.2019.1638701.
63. Paxton, C.W.; Cosgrove, R.A.; Drozd, A.C.; Wiggins, E.L.; Woodhouse, S.; Watson, R.A.; Spence, H.J.; Ozanne, B.W.; Pell, J.M. BTB-Kelch protein Krp1 regulates proliferation and differentiation of myoblasts. *Am J Physiol Cell Physiol* **2011**, *300*, C1345-1355, doi:10.1152/ajpcell.00321.2010.
64. Xian, S.; Li, J.; Zhang, Z. miR-26b inhibits isoproterenol-induced cardiac fibrosis via the Keap1/Nrf2 signaling pathway. *Exp Ther Med* **2020**, *19*, 2067-2074, doi:10.3892/etm.2020.8455.
65. Zhou, Y.; Tang, X.; Niu, L.; Liu, Y.; Wang, B.; He, J. Ectodermal-neural cortex 1 as a novel biomarker predicts poor prognosis and induces metastasis in breast cancer by promoting Wnt/beta-catenin pathway. *J Cell Mol Med* **2020**, 10.1111/jcmm.15520, doi:10.1111/jcmm.15520.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2020 doi:10.20944/preprints202010.0365.v1

66. Watanabe, K.; Yoshida, K.; Iwamoto, S. Kbtbd11 gene expression in adipose tissue increases in response to feeding and affects adipocyte differentiation. *J Diabetes Investig* **2019**, *10*, 925-932, doi:10.1111/jdi.12995.
67. Li, W.; Pung, D.; Su, Z.Y.; Guo, Y.; Zhang, C.; Yang, A.Y.; Zheng, X.; Du, Z.Y.; Zhang, K.; Kong, A.N. Epigenetics Reactivation of Nrf2 in Prostate TRAMP C1 Cells by Curcumin Analogue FN1. *Chem Res Toxicol* **2016**, *29*, 694-703, doi:10.1021/acs.chemrestox.6b00016.
68. Fabrizio, F.P.; Sparaneo, A.; Centra, F.; Trombetta, D.; Storlazzi, C.T.; Graziano, P.; Maiello, E.; Fazio, V.M.; Muscarella, L.A. Methylation Density Pattern of KEAP1 Gene in Lung Cancer Cell Lines Detected by Quantitative Methylation Specific PCR and Pyrosequencing. *Int J Mol Sci* **2019**, *20*, doi:10.3390/ijms20112697.
69. Trapnell, C.; Roberts, A.; Goff, L.; Pertea, G.; Kim, D.; Kelley, D.R.; Pimentel, H.; Salzberg, S.L.; Rinn, J.L.; Pachter, L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **2012**, *7*, 562-578, doi:10.1038/nprot.2012.016.
70. Tsumagari, K.; Baribault, C.; Terragni, J.; Varley, K.E.; Gertz, J.; Pradhan, S.; Badoo, M.; Crain, C.M.; Song, L.; Crawford, G.E., et al. Early de novo DNA methylation and prolonged demethylation in the muscle lineage. *Epigenetics* **2013**, *8*, 317-332, doi:10.4161/epi.23989.
71. Chen, P.H.; Smith, T.J.; Wu, J.; Siesser, P.F.; Bisnett, B.J.; Khan, F.; Hogue, M.; Soderblom, E.; Tang, F.; Marks, J.R., et al. Glycosylation of KEAP1 links nutrient sensing to redox stress signaling. *EMBO J* **2017**, *36*, 2233-2250, doi:10.15252/embj.201696113.
72. Song, Q.; Decato, B.; Hong, E.E.; Zhou, M.; Fang, F.; Qu, J.; Garvin, T.; Kessler, M.; Zhou, J.; Smith, A.D. A reference methylome database and analysis pipeline to facilitate integrative and comparative epigenomics. *PLoS One* **2013**, *8*, e81148, doi:10.1371/journal.pone.0081148.
73. Krietenstein, N.; Abraham, S.; Venev, S.V.; Abdennur, N.; Gibcus, J.; Hsieh, T.S.; Parsi, K.M.; Yang, L.; Maehr, R.; Mirny, L.A., et al. Ultrastructural Details of Mammalian Chromosome Architecture. *Mol Cell* **2020**, *78*, 554-565 e557, doi:10.1016/j.molcel.2020.03.003.
74. Papadopoulos, J.S.; Agarwala, R. COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics* **2007**, *23*, 1073-1079, doi:10.1093/bioinformatics/btm076.