

RESEARCH

Open Access

Mammalian-wide interspersed repeat (MIR)-derived enhancers and the regulation of human gene expression

Daudi Jjingo¹, Andrew B Conley¹, Jianrong Wang¹, Leonardo Mariño-Ramírez^{2,3}, Victoria V Lunyak^{3,4} and I King Jordan^{1,3*}

Abstract

Background: Mammalian-wide interspersed repeats (MIRs) are the most ancient family of transposable elements (TEs) in the human genome. The deep conservation of MIRs initially suggested the possibility that they had been exapted to play functional roles for their host genomes. MIRs also happen to be the only TEs whose presence in-and-around human genes is positively correlated to tissue-specific gene expression. Similar associations of enhancer prevalence within genes and tissue-specific expression, along with MIRs' previous implication as providing regulatory sequences, suggested a possible link between MIRs and enhancers.

Results: To test the possibility that MIRs contribute functional enhancers to the human genome, we evaluated the relationship between MIRs and human tissue-specific enhancers in terms of genomic location, chromatin environment, regulatory function, and mechanistic attributes. This analysis revealed MIRs to be highly concentrated in enhancers of the K562 and HeLa human cell-types. Significantly more enhancers were found to be linked to MIRs than would be expected by chance, and putative MIR-derived enhancers are characterized by a chromatin environment highly similar to that of canonical enhancers. MIR-derived enhancers show strong associations with gene expression levels, tissue-specific gene expression and tissue-specific cellular functions, including a number of biological processes related to erythropoiesis. MIR-derived enhancers were found to be a rich source of transcription factor binding sites, underscoring one possible mechanistic route for the element sequences co-option as enhancers. There is also tentative evidence to suggest that MIR-enhancer function is related to the transcriptional activity of non-coding RNAs.

Conclusions: Taken together, these data reveal enhancers to be an important *cis*-regulatory platform from which MIRs can exercise a regulatory function in the human genome and help to resolve a long-standing conundrum as to the reason for MIRs' deep evolutionary conservation.

Background

Transposable elements (TEs) are abundant in eukaryotic genomes, particularly mammalian genomes. Indeed, at least 45% of the human genome is made up of TE-derived sequences [1,2], which are non-randomly distributed across the genome. For example, human Alu short interspersed elements (SINEs) are predominantly found in GC- and gene-rich regions, whereas L1 long interspersed elements (LINEs) are most prevalent in low-GC and gene-poor

regions [1,3]. Transposable elements have also been shown to affect the expression of host genes via the provisioning of a variety of regulatory sequences [4]. The non-random genomic distribution of human TEs, considered together with their regulatory potential, initially suggested the possibility that the TE environment of human genes might affect the way that they are expressed.

In fact, a number of associations between the TE environment in-and-around human genes and their expression levels and functional patterns have subsequently been observed. Weakly expressed genes generally contain low SINE and high LINE densities, while the most highly expressed human genes are enriched for SINEs (Alu) [5] and depleted

* Correspondence: king.jordan@biology.gatech.edu

¹School of Biology, Georgia Institute of Technology, Atlanta, GA, USA

³PanAmerican Bioinformatics Institute, Santa Marta, Magdalena, Colombia

Full list of author information is available at the end of the article

in L1 elements [6]. Additionally, Alu elements are significantly associated with the breadth of gene expression across tissues [7,8]. Thus, highly and broadly expressed house-keeping genes are identifiable by their TE-content, which is rich in Alus and poor in L1s [9]. Functionally, TEs have recently been demonstrated to have been exapted during the evolution of novel phenotypic characteristics, such as mammalian pregnancy [10,11]. Mammalian-wide interspersed repeats (MIRs) are the only TEs that show a positive association between their prevalence in-and-around genes and tissue-specific gene expression [8,12].

MIR elements are an ancient family of tRNA-derived SINEs [13,14], whose anomalous sequence-conservation levels among mammalian genomes were initially taken as evidence that they encode some unknown regulatory function [15]. Succeeding studies demonstrated that, in a number of individual cases, MIRs do in fact donate transcription-factor binding sites [16-20], enhancers [18,21,22], microRNAs [23,24] and *cis* natural antisense transcripts [25] to the human genome. The association of MIRs with tissue-specific expression, along with their propensity to be exapted as regulatory sequences, suggests to us the possibility that they might provide numerous tissue-specific regulatory sequences across the human genome [8].

Enhancers are regulatory elements that are most highly associated with tissue-specific expression [26,27]. They are also characterized by a unique chromatin environment made up of a specific combination of histone modifications [26-29]. Consistent with their role as tissue-specific regulatory elements, the enhancer chromatin environment is highly variable across cell-types, compared to other classes of regulatory sequences [26,27,29]. We hypothesized that the global coincident association of both MIRs and enhancers to tissue-specific gene expression is at least in part a consequence of MIR sequences frequently acting either as enhancers and/or constituting fragments of enhancer sequences. This would be consistent with previously reported individual cases of TE-derived enhancers [21,30-32]. We also reasoned that the enhancer-characteristic chromatin environment could serve as a useful proxy to identify putative MIR-derived enhancers.

To test our hypothesis, we performed a genome-wide assessment of the relative prevalence of MIRs within enhancer sequences and explored the potential mechanistic bases and functional consequences of this relationship. We found that not only are MIRs highly concentrated in predicted enhancers, but they also constitute a significant part of the core of genic enhancers; this analysis identified many more putative MIR-derived enhancers than previously reported [22,33]. These MIR-derived enhancers have cell-type specific chromatin profiles that are highly similar to those seen for canonical enhancers. Furthermore, we report MIRs to be major donors of transcription-factor

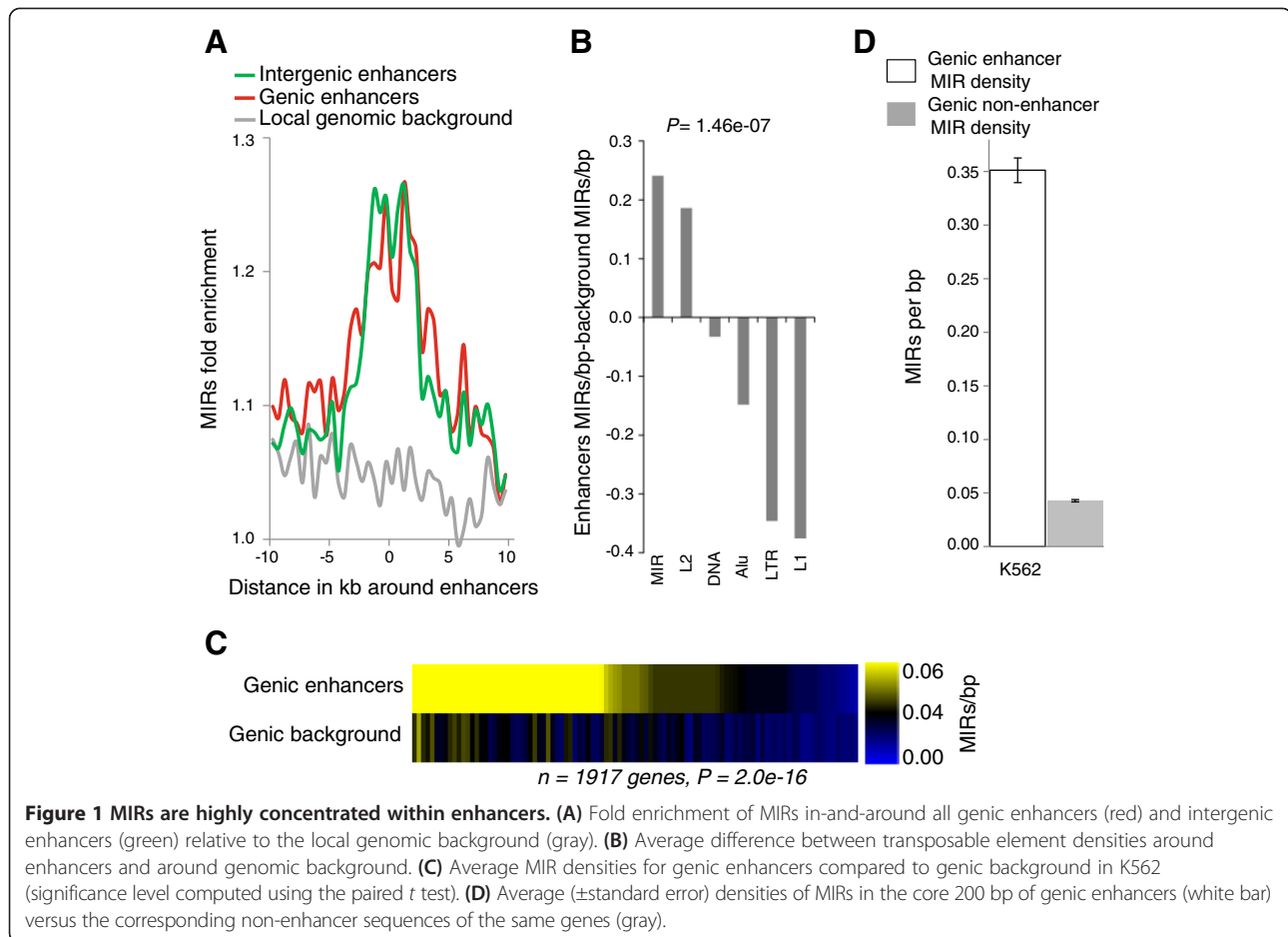
binding sites (TFBSs) within enhancers, and show that MIR-derived enhancers affect both the level and tissue-specificity of gene expression. Using the erythroid K562 cell-line as an example, we show that MIR-enhancers are involved in the modulation of several developmentally-specific biological processes related to erythropoiesis.

Results and discussion

MIRs are highly concentrated in enhancers

As noted in the introduction, MIRs are the only TEs that show a positive association with tissue-specific gene expression [8]. Similarly, unlike other *cis*-regulatory elements, enhancers are marked with highly cell-type specific histone modification patterns [26] and are accordingly also highly related to tissue-specific gene expression [26,27]. We thus sought to test our working hypothesis that these apparent similarities between MIRs and enhancers are largely a consequence of MIR sequences either frequently acting as enhancers and/or constituting fragments of enhancer sequences.

The genomic coordinates of 24,538 and 36,550 putative transcriptional enhancers in the K562 and HeLa cell-lines, respectively, [26] were intersected with those of all 593,339 MIRs in the genome. For all genomic enhancers, we computed the fraction of MIRs in and around 20 kb loci centered on all genomic enhancers ($n = 24,538$ and $36,550$ for K562 and HeLa cell-lines respectively) and compared it with MIR enrichment in the local genomic background. The results reveal MIRs to be highly enriched within all enhancers genome-wide, with up to $\approx 26\%$ and $\approx 27\%$ more MIRs within enhancers than in the genomic background for K562 and HeLa cell-lines ($\chi^2 = 4592$, $P < 1.0 \times 10^{-308}$, K562; $\chi^2 = 7470$, $P < 1.0 \times 10^{-308}$, HeLa; Figure 1A and Additional file 1: Figure S1A). The overall distribution of MIRs within enhancers reveals more than 74% of enhancers to contain MIRs (Additional file 1: Figure S1D). Furthermore, MIRs show a significantly higher concentration within enhancers relative to all other TEs (Figure 1B). Additionally, the locations of 19,536 Refseq gene annotations [8] were also intersected with those of enhancers. This yielded 1,917 and 2,090 genes with predicted enhancers in their gene bodies in the K562 and HeLa cell-lines. For each of these genes, its resident enhancers and its non-enhancer sequences were intersected with a set of all genomic MIRs from the University of California Santa Cruz (UCSC) Genome Browser [34,35], yielding MIR densities within both genic enhancers and genic non-enhancer regions. Within gene bodies, MIRs show significantly higher densities in enhancers than in non-enhancer sequences (Figure 1C and Additional file 1: Figure S1B). Furthermore, MIRs are vastly overrepresented components of the core 200 bp of genic enhancers, showing eight- and seven-fold enrichment compared to their presence



in non-enhancer sequence regions from the same genes in K562 and HeLa cell-lines (Figure 1D and Additional file 1: Figure S1C).

Thus, while MIRs have been previously reported to be enriched within introns [12], our data clearly reveal this genic enrichment to be strongly biased towards enhancers. And while MIRs are known to donate enhancers in a number of individual cases [21,22,33], these data show an even deeper relationship, namely that MIRs are substantially concentrated in enhancers genome-wide.

Numerous MIRs provide enhancers or are linked to enhancers

Finding MIRs to be highly concentrated within enhancers, we sought to establish the numbers and locations of MIRs that provide core enhancer sequences themselves (MIR-enhancers) as well as those that lie within enhancer regions (enhancer-MIRs). We found 934 and 1,429 MIRs to be MIR-enhancers in K562 and HeLa cell-lines (genomic locations in Additional file 2). This is in contrast to the 669 and 996 MIRs that would be expected to be enhancers in the two cell-lines if MIRs were randomly distributed among enhancers ($\chi^2 = 105$, $P = 1.2 \times 10^{-24}$,

K562; $\chi^2 = 188$, $P = 1.0 \times 10^{-42}$, HeLa). Furthermore, the numbers of MIR-derived enhancers identified here is far greater than has been previously reported [22,33], owing to the availability of more enhancer annotations and a more accurate estimate of the size of enhancers, which we deduced from the span of characteristic enhancer histone-modifications at enhancer sites. When this analysis was expanded to include all enhancer-linked MIRs (that is, enhancer-MIRs), the extent to which enhancers are connected to MIRs became even more apparent. We found 16,144 and 26,520 enhancers to be linked to MIRs in K562 and HeLa celllines respectively, compared with the 6,559 and 9,320 enhancer-MIRs that would be expected by chance alone ($\chi^2 = 14,007$, $P < 1.0 \times 10^{-308}$, K562; $\chi^2 = 31,742$, $P < 1.0 \times 10^{-308}$, HeLa).

We compared the chromatin (histone modification) environment of MIRenhancers and enhancer-MIRs with canonical predicted enhancers to further evaluate their status as enhancers and their regulatory potential. The histone modifications H3K4me1 and H3K27Ac have been shown to be enriched at experimentally characterized enhancers in a number of studies [26,29,36,37]. We found both MIR-enhancers and enhancer-MIRs to

have enrichment profiles of these two modifications similar to those of canonical predicted enhancers in K562 and HeLa cells (Figure 2 and Additional file 1: Figure S2A-E). However, the order of histone modification congruity is tissue-specific, with H3K4me1 showing the highest congruity in K562 (Figure 2D and Additional file 1: Figure S3) and H3K27ac with the highest congruity in HeLa (Additional file 1: Figure S3). Also, the repressive mark H3K27me3 shows the least congruity (Additional file 1: Figure S3A) and is, in fact, deleted at both MIR-enhancers and enhancer-MIRs (Additional file 1: Figure S2E, F). As expected, enhancer-MIRs show a somewhat diminished enrichment and congruity of these two modifications, since this category includes enhancer-linked MIRs rather than MIR-enhancers that lie at the core of enhancers. Interestingly, MIR-enhancers show a significantly stronger enrichment of the enhancer distinguishing modifications H3K4me1 and H3K27Ac than the canonical predicted

enhancers ($P = 6.9 \times 10^{-14}$, K562; 9.6×10^{-24} , HeLa, paired t test for the two modifications).

MIRs are enriched for TFBSs

Enhancers are known to boost gene expression by recruiting transcription factors (TFs), which in turn interact with promoters to recruit RNA polymerase II, facilitating the initiation of transcription [38]. Accordingly, a plausible evolutionary route for the exaptation of MIRs into enhancers would be that MIRs offered a rich source of TFBSs compared with random genomic sequences. We investigated this possibility by exploring the contribution of enhancer-associated MIRs (MIR-enhancers and enhancer-MIRs) to TFBS sequence motifs corresponding to several TFs that are known to be active in K562-specific cellular processes: C-JUN, NF-E2, and ZNF274 [39-42]. TFBS sequence motifs for C-JUN and NF-E2 are found in significantly higher copy numbers in enhancer-associated

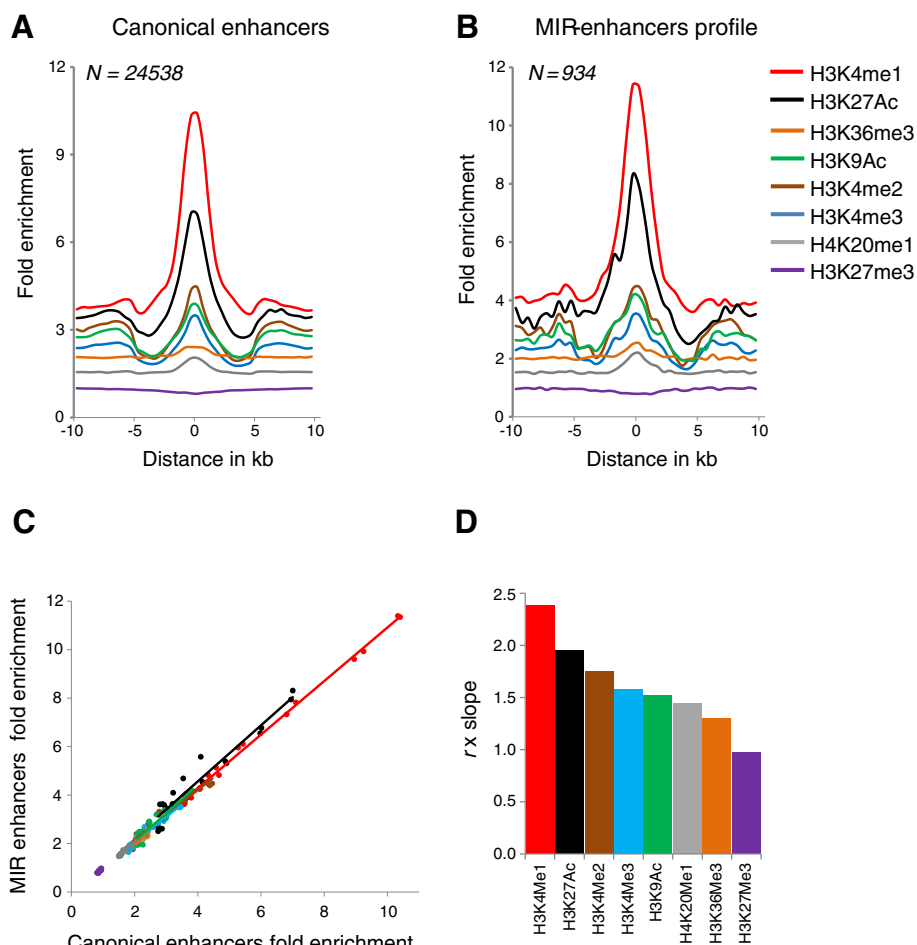


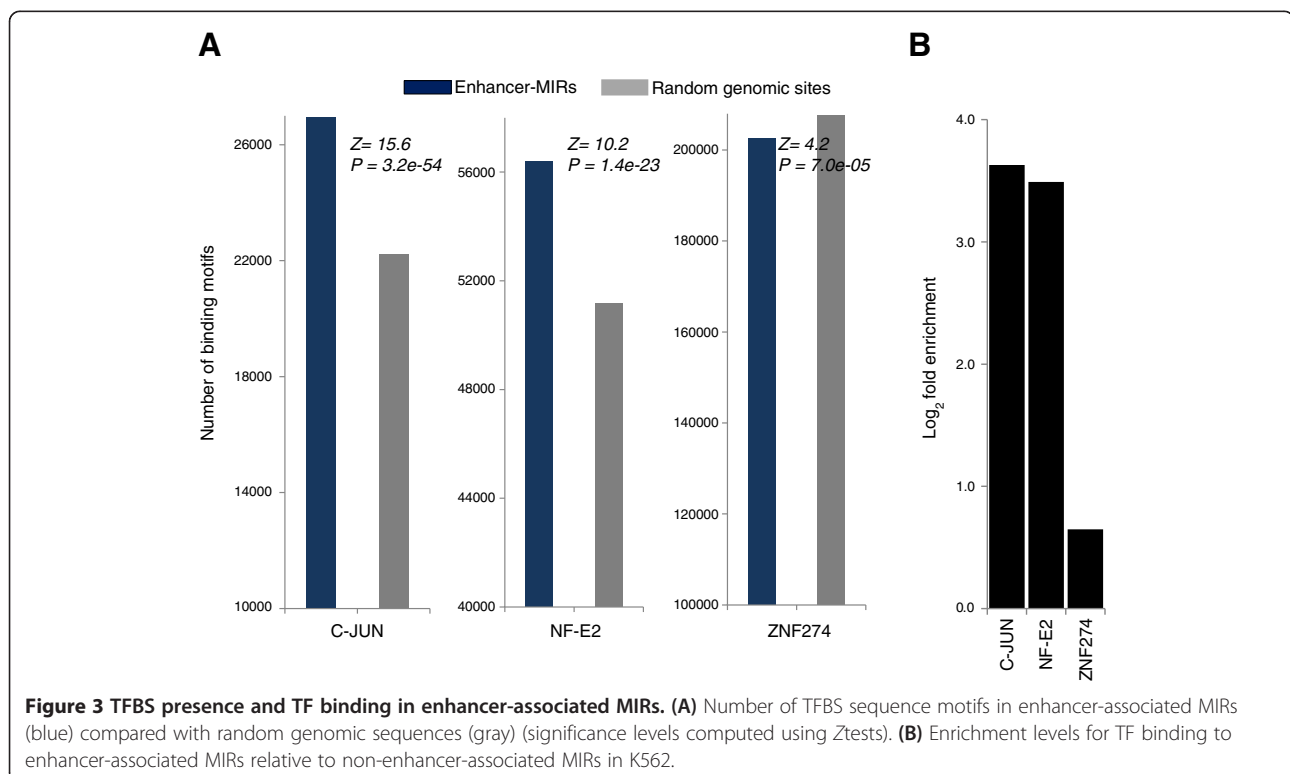
Figure 2 The chromatin environment of MIR-enhancers is similar to that of canonical enhancers. Fold enrichment of histone modifications centered on (A) canonical enhancers and (B) MIR-enhancers. (C) Congruence of histone modification fold enrichment levels between canonical enhancers and MIR-enhancers. (D) Rank order of correlations of histone modification fold enrichments between canonical enhancers and MIR-enhancers weighted by slope.

MIRs than seen for random genomic sites adding up to the same fraction of the genome. This is the case when motifs are counted using position weight matrices (PWMs) of the TFBS [43] (Figure 3A) or the regular expression representations of the TFBS (Additional file 1: Figure S4A and Table S1A). The PWM method [44] additionally yielded factors Elf-1, SF-1, and LRH-1 as enriched in enhancer-MIRs (P values of 2.7×10^{-2} , 7.2×10^{-3} , and 1.3×10^{-3} respectively). We also surveyed experimentally characterized occupancy of enhancer-associated MIRs by these same TFs using ChIP-seq (chromatin immunoprecipitation followed by high-throughput sequencing) data from the ENCODE project. Transcription-factor binding characterized in this way confirms that C-JUN and NF-E2 occupancy levels are > nine-fold higher, while ZNF274 is also enriched, albeit marginally, within enhancer-associated MIRs relative to non-enhancer-associated MIRs in the K562 cell-line (Figure 3B). Similar analyses conducted with additional TFBSs corresponding to TFs with binding experimentally characterized in K562 also revealed enrichments of TFBS sequence motifs and TF bound sites in enhancer-associated MIRs. Enrichment of TF binding at enhancer-associated MIRs was observed for 37/39 and 37/44 bound TFs in K562 and HeLa cells, respectively (Additional file 1: Figure S4B). Finally, these two data types were integrated by evaluating the presence of canonical TFBS sequence motifs among the set of enhancer-associated MIRs experimentally characterized to

be bound by the TFs C-JUN, NF-E2, and ZNF274. Consistent with previous ChIP-seq studies [45], not all enhancer-associated MIRs that are experimentally characterized to be bound by TFs show canonical TFBS sequence motifs. Nevertheless, numerous MIR-derived enhancer sequences can be seen to provision TFBS sequence motifs for these TFs in K562 (Figure 4). Taken together, these data are consistent with the notion that the evolutionary co-option of MIRs into enhancers is due in part to their relatively large and functionally relevant repertoire of TFBSs.

Enhancer-associated MIRs influence gene expression levels and tissue-specificity

To check whether the observed prevalence and TF binding capacity of enhancer-associated MIRs translates into genome-wide regulatory effects, we related enhancer-associated MIR densities to two gene expression parameters: gene expression level and tissue-specificity. For both K562 and HeLa cell-lines, the density of enhancer-associated MIRs in and around genes is significantly related to gene expression levels (Figure 5A and Additional file 1: Figure S5A). Similarly, there are significant relationships between enhancer-associated MIR densities in both K562 and HeLa cell-lines and tissue-specific expression across 6 ENCODE cell-lines (Figure 5B and Additional file 1: Figure S5B,D). The relationship between enhancer-associated MIR density and tissue-specific expression is even more apparent when tissue-specificity indices are calculated across 79



A

TFs	TF-bound enhancer-MIRs	MEME-derived motifs in enhancer-MIRs	Number of sites with MEME motifs	Canonical motif	Canonical motifs on positive strand	Similarity of canonical and MEME motifs	Total binding sites
C-JUN	45		4	TGA(CIG)TC A	2	86%	5
NF-E2	26		11	TCA(TIC)	19	100%	19
ZNF274	5		5	(GIA)A(AIG) TG(TIG)	1	83%	6

B

C-JUN				ZNF274			
Strand	Start	P-value	Site	Strand	Start	P-value	Site
+	47	2.32e-04	TCTTGCTAG CTGAGTC ACCTTAGACA	+	51	3.84e-04	CTTGCTGTG AAAATGA AGAGAATACT
-	16	2.32e-04	CTTCTTATGT CTGAGTC TCACTCTTCA	+	123	3.84e-04	CAGAGGTGTC AAAATGA CCTCTCCAAG
-	8	2.32e-04	CACCTAATTG CTGAGTC CTCAGGC	+	56	3.84e-04	CTCCTGTGA AAAATGA GGGTGATAAG
-	195	2.32e-04	GTCTGACCAT CTGAGTC AACATCCTGG	-	51	3.84e-04	CCTAAGAATT AAAATGA AAACAATATT
				-	3	6.83e-04	AGGTCTGTG AAAATGA GG

Figure 4 TFBS sequence motifs for TF bound enhancer-associated MIRs. (A) Data for TFBS sequence motifs present in TF bound enhancer-MIRs. **(B)** Examples of TFBS sequence motifs for C-JUN and ZNF274 present in individual TF bound enhancer-associated MIR sequences. The *P* value calculated by the program MEME is the probability that the TFBS exists within the sequence by chance.

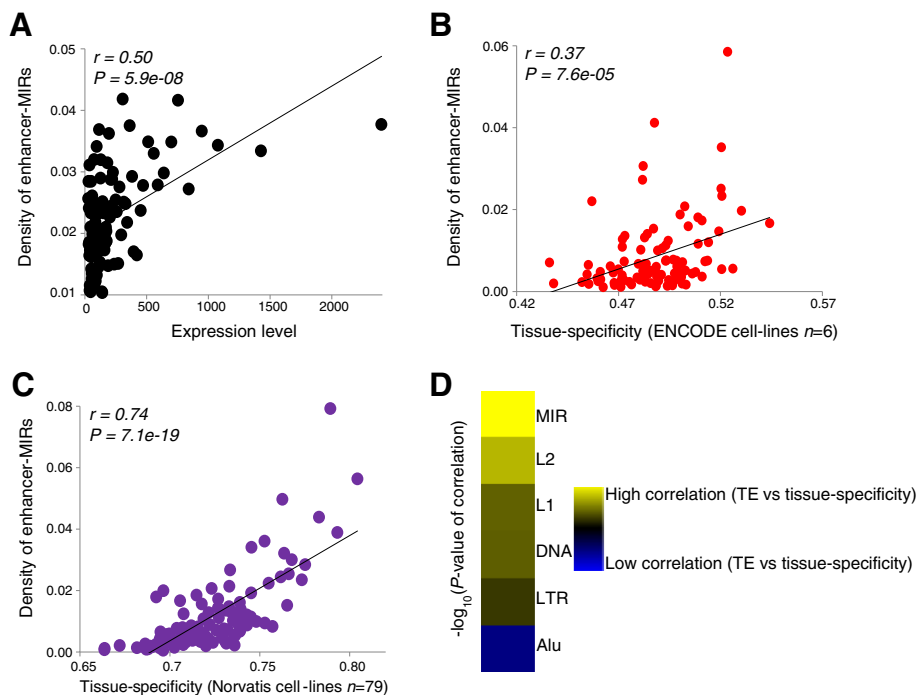


Figure 5 Effect of enhancer-associated MIRs on gene expression. (A) Relationship between density of enhancer-MIRs and gene expression levels in K562. **(B)** Relationship between density of K562 enhancer-MIRs and tissue-specificity of gene expression across six ENCODE cell-lines. **(C)** Relationship between density of K562 enhancer-MIRs and tissue-specificity of gene expression across 79 tissues from the Norvatis Gene Expression Atlas. Pearson correlation coefficient values (*r*) along with their significance values (*P*) are shown for all pairwise regressions. **(D)** Relative strength of correlation between the densities of different TE types around genes and the tissue-specificity of those genes.

different tissues using the Yanai tissue-specificity equation [46] (Figure 5C and Additional file 1: Figure S5C,E). That relationship holds even when tissue-specificity is calculated using Shannon entropy [47] (Additional file 1: Figure S5D,E) and is substantially stronger for MIRs relative to other TEs (Figure 5D). Moreover, this relationship with tissue-specificity is independent of local genomic context as measured by GC content of the TEs (Additional file 1: Table S1C). Furthermore, the number of MIR-derived TFBS in the 5% most tissue-specific genes was $\approx 25\%$ higher than that for the 5% least tissue-specific genes. Taken together, these data reveal enhancer-associated MIRs to have a significant association with the genome-wide patterns of gene expression levels from the cell-lines in which the enhancers were identified, as well as the overall tissue-specificity measured across multiple cell-lines and tissues.

Functional significance of enhancer-associated MIRs

Since enhancer-associated MIRs are related to tissue-specific gene expression, it is reasonable to expect that there are some tissue-specific biological functions that they may help to regulate. We examined this prospect in the K562 cell-line by assessing the functional roles of genes within 100 kb of tissue-specific enhancer-associated MIRs. Of 19,538 non-overlapping Refseq genes, we found 3,798 (19.5%) to be associated with such K562 predicted enhancer-associated MIRs. We tested for relative enrichment of those genes within a set of 350 genes that have been shown to be highly regulated in erythroids across four stages of erythropoiesis [48]. Of the 3,798 enhancer-associated MIR associated genes, 202 overlapped the set of 350 genes that are highly regulated in erythropoiesis or their close homologs. This overlap is highly significant (hypergeometric test, $P = 2.1 \times 10^{-57}$) and suggests that enhancer-associated MIRs might have an impact on erythropoietic regulation. We therefore broadened the analysis to include other biological processes related to erythropoiesis. We tested for enrichment of enhancer-MIR associated genes in gene sets of nine erythroid biological functions obtained from the Broad Institute's molecular signatures database (MSigDB). Gene sets for eight out of the nine erythroid-related biological functions are significantly enriched among enhancer-associated MIR-linked genes (Figure 6A and Additional file 1: Table S2). These results were further supported by functional analysis of our enhancer-associated MIRs using the GREAT tool [49]. The biological processes it identified are highly similar to those identified previously, including erythropoiesis and its related functions, such as myeloid cell differentiation, interphase of mitotic cell cycle and homeostasis of a number of cells.

To further understand the impact that enhancer-associated MIRs might have on K562 cell-type specific

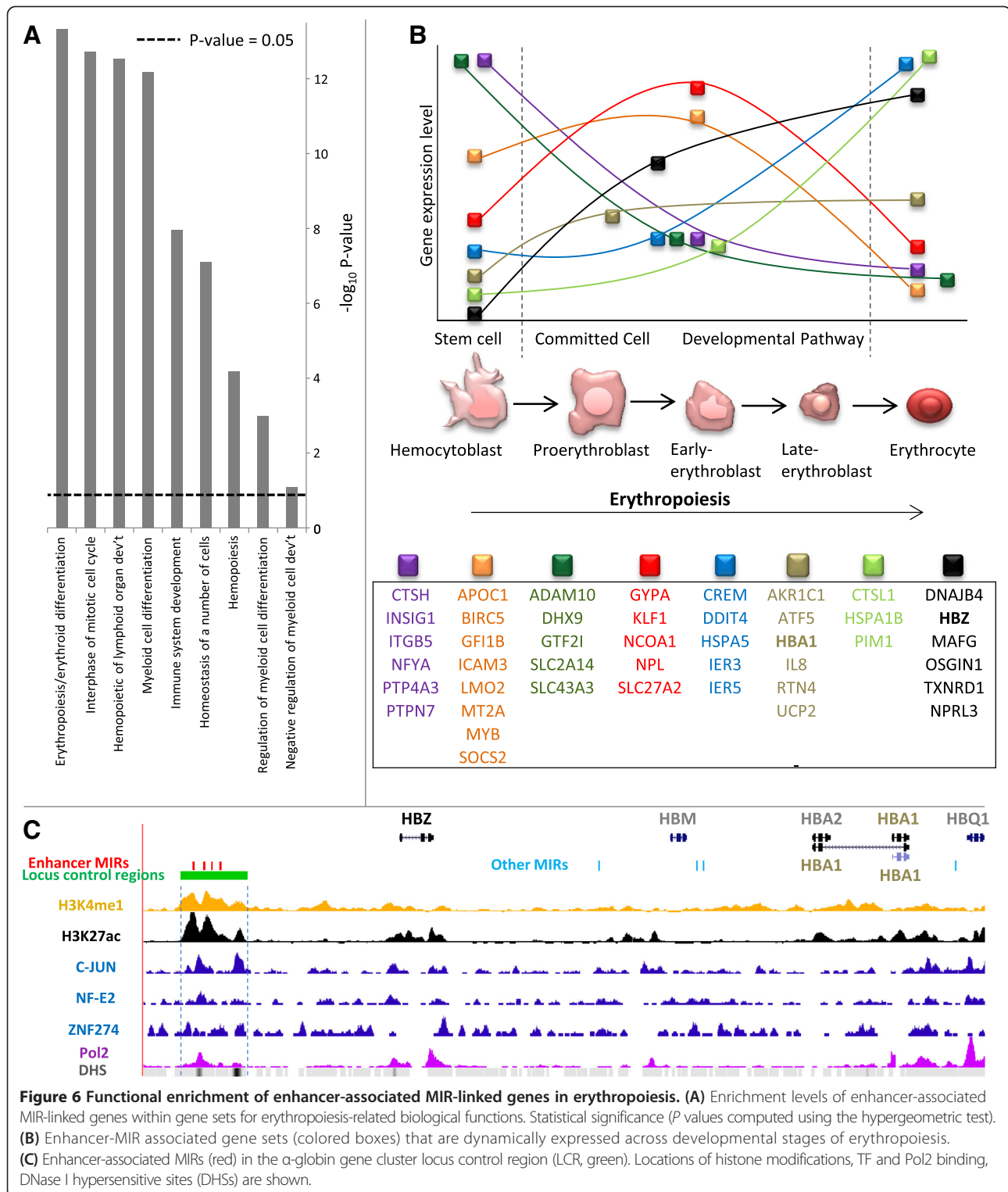
biological functions, we focused on the erythropoiesis biological function, whose gene set has the most significant overlap with enhancer-MIR associated genes (Figure 6A). This gene set contains genes that have been implicated in various aspects of erythrocyte function and development [50]. We compared the expression levels of enhancer-MIR associated genes in this gene set with developmental stages of erythropoiesis and found them to have highly divergent expression levels across development, an indicator that enhancer-associated MIRs might be involved in their regulation during erythropoiesis (Figure 6B and Additional file 1: Table S3).

Interestingly, this erythropoiesis gene set and its related enhancer-associated MIRs include four distinct MIR sequences that were previously implicated as being involved in the regulation of the α -globin gene cluster by virtue of their co-location with an experimentally characterized 'locus control region' [51]. This cluster contains a number of globin genes, including HBZ and HBA1, both of which are enhancer-MIR associated and differentially expressed across the various stages of erythropoiesis (Figure 6B,C). Furthermore, the enhancer-associated MIRs in the locus control region can be seen to be marked by the enhancer-related histone modifications H3K4me1 and H3K27ac, to recruit the TFs C-JUN, NF-E2 and ZNF274, and to reside in a relatively open chromatin environment, as characterized by DNase I hypersensitive sites (DHSs) (Figure 6C). In addition, one of the four locus control region co-located MIRs appears to recruit RNA polymerase II (Pol2) transcriptional machinery. This suggests the possibility that enhancer-associated MIRs might also exert their regulatory activity by virtue of the transcriptional activity of non-coding RNAs, as has been observed for a number of TE- and tRNA-derived regulatory sequences [52-54]. Consistent with this possibility, non-coding RNA transcriptional activity has been suggested as a widespread mechanism underlying enhancer function in mammalian genomes [55-58].

Considered together, these results suggest that K562-predicted enhancer-associated MIRs are active in the cell-type specific and developmental regulation of genes involved in a number of biological processes related to K562 functions in general, and erythropoiesis in particular. Furthermore, the exaptation of MIRs as enhancers might be predicated upon the recruitment of specific TFs, as previously discussed, as well as the transcriptional activity of non-coding RNAs.

Conclusions

A number of previous studies have found different classes and families of TEs and TE-derived sequences to have distinct and substantial effects on genome regulation [4]. In one such study [8], our group identified MIRs to be the only TE-derived sequences whose presence shows a



positive correlation to tissue-specific gene expression. Here, we provide evidence that this correlation is likely to be due to the propensity of MIRs to be exapted as enhancers that regulate cell-type and developmental-stage specific gene expression. We show that MIR-related

enhancer activity is functionally relevant and may be related to TF binding, as well as the transcriptional activity of non-coding RNAs. The widespread exaptation of previously selfish MIR sequences as enhancers resolves a long-standing conundrum arising from the observation

that MIR-derived sequences are far too evolutionarily conserved to be simply non-functional or 'junk' DNA [15].

Methods

Identification of MIR-related enhancers

We used two sets of 24,538 and 36,550 putative transcriptional enhancers, predicted in the K562 and HeLa cell-lines, respectively [26]. These enhancers were predicted as ENCODE regions that bear specific chromatin histone modification (H3K4me1 and H3K27ac) profiles that are similar to those seen for genomic regions bound by the coactivator protein p300, which is known to colocalize at active enhancers [59]. The p300 binding sites were themselves located using a chromatin immunoprecipitation-based microarray method (ChIP-chip) and the histone modification data were taken from the ENCODE project [36,60]. We considered the span of enhancers to be the ± 4 kb region around the predicted enhancer midpoints, which corresponds roughly to the empirically determined range of the characteristic chromatin pattern found at predicted enhancers (Figure 2A). We intersected the coordinates of these enhancers with the RepeatMasker [3] annotations of MIR elements as identified by the Repbase classification system [61,62] to predict MIR-related enhancers. These MIR annotations on the human genome assembly (NCBI build 36.1; UCSC hg18) were downloaded from the UCSC Genome Browser [34,35]. Each putative enhancer analyzed here was originally predicted to be anchored around a single base-pair position [26]. If this core basepair was located in a MIR, then such a MIR was classified as a core enhancer. Such cases are considered *MIR-enhancers* for the purposes of analysis. On the other hand, some MIRs do not donate the core enhancer locus but are instead located within the typical ± 4 kb span for enhancers. These MIRs were considered enhancer-linked MIRs and are referred to as *enhancer-MIRs*. There are thus two categories of MIR-related enhancers analyzed here: *MIR-enhancers* and *enhancer-MIRs*. However, at both the locational and functional levels, both categories of MIRs are part of the enhancer body, which we determined using the span of its chromatin profile.

A set of 19,536 non-overlapping transcriptional units derived from Refseq gene annotations, as defined in reference [8], was used to assess MIR densities within genes. For both the K562 and HeLa cell-lines, regions of overlap between MIR genomic coordinates and four different types of genomic elements or regions were determined: (i) genic enhancers, (ii) genic non-enhancer regions, (iii) non-genic enhancers, and (iv) the core 200 bp region around predicted enhancer midpoints. For each region, the density of MIRs was computed either as the fraction of the length of each region in basepairs that was occupied by MIRs or their fold enrichment within

the regions relative to the local genomic background. The local genomic backgrounds were compiled as regions randomly sampled 100 kb downstream of each locus of interest. Their enrichment in terms of MIRs or other ChIP-seq datasets was then divided by the average densities of those datasets over the entire genome to obtain their normalized signal values.

Expected numbers of MIR-enhancers were computed as the average genome-wide density of enhancers (enhancers/bp) multiplied by the total length in basepairs of all genomic MIRs. Expected numbers of enhancers with MIRs were simulated by mapping random genomic sites equivalent to MIRs (in number and size) to enhancer regions and then counting the number of enhancers that were overlapped.

Histone modification profile analysis

Genome-wide ChIP-seq [63] data for eight histone modifications (H3K4me1, H3K27ac, H3K36me3, H3K9ac, H3K4me2, H3K4me3, H4K20me1, and H3K27me3) in the K562 and HeLa-S3 cell-lines was taken from the 'ENCODE histone modification tracks' of the UCSC Genome Browser. Genomic loci of 20 kb centered on canonical enhancers (all predicted enhancers), MIR-enhancers, and enhancer-MIRs were evaluated for enrichment of the histone modifications. Counts of each histone modification within 500 bp windows across the 20 kb region were then computed and their profiles represented as fold enrichments relative to average counts per 500 bp in the genomic background. The congruence of histone modification profiles between canonical enhancers and MIR-related enhancers was assessed using rank correlations between modification enrichments, which were weighted by the slope of their line-of-best-fit to establish the relative order of histone mark enrichment congruence.

Transcription-factor binding site analysis

TFBS sequence motifs within MIR-related enhancers were identified using both PWMs and regular expression representations of sequence motifs corresponding to K562-related TFs as gleaned from the experimental literature and TF databases [45,64-68]. Counts of TFBS sequence motifs in MIR-related enhancer sequences were compared with counts of the same motifs in random genomic sequences of the same number and size. For each transcription factor, we obtained the expected number by counting the number of motifs in 1,000 random samples, each being of the same size as the enhancer-MIR set (50,599 sequences). The medians of those distributions were then considered the *expected* number for each motif. *P* values were then empirically determined using a *Z* test on the distributions with μ the median and χ the observed number of motifs in enhancer-MIRs, which is the same number as our set of enhancer-associated MIRs.

Experimentally characterized TF binding sites within MIR-related enhancer sequences were identified using ChIP-seq data from the 'ENCODE transcription-factor binding tracks' of the UCSC Genome Browser. Enrichment values of TF occupancy levels for MIR-related enhancer sequences were computed using ChIP-seq tag counts within MIR-related sequences normalized by the genome average ChIP-seq tag counts for non-enhancer-MIRs. The presence of TFBS sequence motifs within ChIP-seq characterized TF bound regions was evaluated using regular expressions, as described, along with the motif finding program MEME [69].

Gene expression analysis

Two sets of gene expression data were used here. The first dataset of exon microarray data for six ENCODE cell-lines (K562, HeLa-S3, GM12878, HepG2, H7Hesc, and HUVEC) was taken from the 'ENCODE exon array' track of the UCSC Genome Browser. Exon array data were converted into gene expression levels for 18,654 genes as outlined in reference [70]. The second dataset of Affymetrix microarray expression data from 79 tissues and cell-lines was taken from the Norvatis Gene Expression Atlas [71]. Signal intensity values for individual probes from this dataset were normalized and associated with 15,658 genes, as previously outlined [8]. For both datasets, a tissue-specificity index (*TS*) for each gene was computed using a previously described formula [46]:

$$TS = \frac{\sum_{i=1}^N (1-x_i)}{N-1}$$

where N is the number of tissues and x_i represents a gene's signal intensity value in each tissue i divided by the maximum signal intensity value of the gene across all tissues. For both datasets, tissue-specificity was also calculated based on the entropy among gene expression levels between tissues using Shannon entropy [47] in R's Bioconductor package.

For each gene, the density of enhancer-associated MIRs in and around the gene (from 10 kb upstream to 10 kb downstream) was computed by dividing the number of enhancer-associated MIRs in that genomic range by the number of base pairs in the range. The density values of the enhancer-associated MIRs were then divided into 100 equal bins whose average densities were regressed against the respective average expression levels or *TS* of the associated genes.

Functional association analysis

The functional effects of enhancer-associated MIRs were evaluated using erythroid (K562)-specific enhancer-associated MIRs (defined as enhancer-associated MIRs present in K562 and absent in HeLa). Genes were associated

with K562-specific enhancer-associated MIRs within 100 kb. The hypergeometric test was used to check for enrichment of enhancer-MIR associated genes within (i) a set of 350 genes previously shown to be developmentally regulated in erythrocytes across four stages of erythropoiesis [48], and (ii) gene sets for nine erythroid-related cellular functions (Additional file 1: Table S1) taken from the Broad Institute's molecular signatures database (MSigDB). Developmental regulation of enhancer-MIR associated erythropoiesis-related genes was assessed using gene expression data for five stages of erythrocyte development [50].

Additional files

Additional file 1: Figures S1-S5 and Tables S1-S3. MIRs regulate human gene expression and function predominantly via enhancers.

Additional file 2: Genomic loci of MIR-derived enhancers in K562 and HeLa cell-lines.

Abbreviations

ChIP: chromatin immunoprecipitation; ChIP-seq: chromatin immunoprecipitation followed by high-throughput sequencing; DHS: DNase I hypersensitive site; ENCODE: Encyclopedia of DNA elements; LINE: long interspersed element; MIR: mammalian-wide interspersed repeat; MSigDB: molecular signatures database; PWM: position weight matrix; SINE: short interspersed element; TE: transposable element; TF: transcription factor; TFBS: transcription-factor binding site; *TS*: tissue-specificity index; UCSC: University of California Santa Cruz.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

DJ, VL, and IKJ conceived and designed the study, performed computational and statistical analyses, and wrote up the results. ABC, JW, and LMR provided technical expertise and assistance for dataset acquisition, curation, and analysis. All authors read and approved the final manuscript.

Acknowledgements

This work was supported in part by the Fulbright foundation through a PhD scholarship to DJ, the School of Biology, Georgia Institute of Technology (to IKJ, DJ, ABC, and JW), an Alfred P Sloan Research Fellowship in Computational and Evolutionary Molecular Biology (BR-4839) to IKJ, and NIH R21 AG043921 to VL. This research was supported in part by the Intramural Research Program of the NIH, NLM, NCBI. No funding bodies played any role in the design, collection, analysis, and interpretation of the data or the writing and submission of the manuscript. The authors would like to acknowledge members of the Jordan laboratory for helpful discussions. The authors are also grateful to the ENCODE consortium for making their data freely accessible and acknowledge the use of data generated by the ENCODE institutions and groups below: Broad Institute and Massachusetts General Hospital-Harvard Medical School/Bernstein Laboratory (histone modifications data). Stanford University/Michael Snyder laboratory, Yale University/Mark Gerstein and Sherman Weissman laboratories, University of Southern California/Peggy Farnham Laboratory and Harvard University/Kevin Struhl Laboratory (transcription-factor binding data) and University of Washington/Sandstrom laboratory (expression data).

Author details

¹School of Biology, Georgia Institute of Technology, Atlanta, GA, USA.

²National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA. ³PanAmerican

Bioinformatics Institute, Santa Marta, Magdalena, Colombia. ⁴Buck Institute for Research on Aging, Novato, CA, USA.

Received: 30 August 2013 Accepted: 10 April 2014
Published: 5 May 2014

References

- Lander ES, Linton LM, Birren B, Nussbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, et al: **Initial sequencing and analysis of the human genome.** *Nature* 2001, **409**(6822):860–921.
- Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, et al: **The sequence of the human genome.** *Science* 2001, **291**(5507):1304–1351.
- RepeatMasker. <http://www.repeatmasker.org>.
- Feschotte C: **Opinion - transposable elements and the evolution of regulatory networks.** *Nat Rev Genet* 2008, **9**(5):397–405.
- Versteeg R, van Schaik BDC, van Batenburg MF, Roos M, Monajemi R, Caron H, Bussemaker HJ, van Kampen AHC: **The human transcriptome map reveals extremes in gene density, intron length, GC content, and repeat pattern for domains of highly and weakly expressed genes.** *Genome Res* 2003, **13**(9):1998–2004.
- Han JS, Szak ST, Boeke JD: **Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes.** *Nature* 2004, **429**(6989):268–274.
- Kim TM, Jung YC, Rhyu MG: **Alu and L1 retroelements are correlated with the tissue extent and peak rate of gene expression, respectively.** *J Korean Med Sci* 2004, **19**(6):783–792.
- Jjingto D, Huda A, Gundapuneni M, Marino-Ramirez L, Jordan IK: **Effect of the transposable element environment of human genes on gene length and expression.** *Genome Biol Evol* 2011, **3**:259–271.
- Eller CD, Regelson M, Merriman B, Nelson S, Horvath S, Marahrens Y: **Repetitive sequence environment distinguishes housekeeping genes.** *Gene* 2007, **390**(1–2):153–165.
- Emera D, Wagner GP: **Transposable element recruitments in the mammalian placenta: impacts and mechanisms.** *Brief Funct Genomics* 2012, **11**(4):267–276.
- Lynch VJ, Leclerc RD, May G, Wagner GP: **Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals.** *Nat Genet* 2011, **43**(11):1154–1158.
- Sironi M, Menozzi G, Comi GP, Cereda M, Cagliani R, Bresolin N, Pozzoli U: **Gene function and expression level influence the insertion/fixation dynamics of distinct transposon families in mammalian introns.** *Genome Biol* 2006, **7**(12):R120.
- Jurka J, Zietkiewicz E, Labuda D: **Ubiquitous mammalian-wide interspersed repeats (MIRs) are molecular fossils from the mesozoic era.** *Nucleic Acids Res* 1995, **23**(1):170–175.
- Smit AF, Riggs AD: **MIRs are classic, tRNA-derived SINEs that amplified before the mammalian radiation.** *Nucleic Acids Res* 1995, **23**(1):98–102.
- Silva JC, Shabalina SA, Harris DG, Spouge JL, Kondrashov AS: **Conserved fragments of transposable elements in intergenic regions: evidence for widespread recruitment of MIR- and L2-derived sequences within the mouse and human genomes.** *Genet Res* 2003, **82**(1):1–18.
- Polavarapu N, Marino-Ramirez L, Landsman D, McDonald JF, Jordan IK: **Evolutionary rates and patterns for human transcription factor binding sites derived from repetitive DNA.** *BMC Genomics* 2008, **9**:226–236.
- Wang J, Bowen NJ, Marino-Ramirez L, Jordan IK: **A c-Myc regulatory subnetwork from human transposable element sequences.** *Molecular Biosyst* 2009, **5**(12):1831–1839.
- Thornburg BG, Gotea V, Makalowski W: **Transposable elements as a significant source of transcription regulating signals.** *Gene* 2006, **365**:104–110.
- Polak P, Domany E: **Alu elements contain many binding sites for transcription factors and may play a role in regulation of developmental processes.** *BMC Genomics* 2006, **7**:133.
- Cui F, Sirotnik MV, Zhurkin VB: **Impact of Alu repeats on the evolution of human p53 binding sites.** *Biol Direct* 2011, **6**:2.
- Marino-Ramirez L, Jordan IK: **Transposable element derived DnaseI-hypersensitive sites in the human genome.** *Biol Direct* 2006, **1**:20.
- Teng L, Firpi HA, Tan K: **Enhancers in embryonic stem cells are enriched for transposable elements and genetic variations associated with cancers.** *Nucleic Acids Res* 2011, **39**(17):7371–7379.
- Piriyapongsa J, Marino-Ramirez L, Jordan IK: **Origin and evolution of human microRNAs from transposable elements.** *Genetics* 2007, **176**(2):1323–1337.
- Kuang HH, Padmanabhan C, Li F, Kamei A, Bhaskar PB, Shu OY, Jiang JM, Buell CR, Baker B: **Identification of miniature inverted-repeat transposable elements (MITEs) and biogenesis of their siRNAs in the Solanaceae: new functional implications for MITEs.** *Genome Res* 2009, **19**(1):42–56.
- Conley AB, Miller WJ, Jordan IK: **Human cis natural antisense transcripts initiated by transposable elements.** *Trends Genet* 2008, **24**(2):53–56.
- Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW, Ching K, Antosiewicz-Bourget JE, Liu H, Zhang X, Green RD, Lobanenkov VV, Stewart R, Thomson JA, Crawford GE, Kellis M, Ren B: **Histone modifications at human enhancers reflect global cell-type-specific gene expression.** *Nature* 2009, **459**(7243):108–112.
- Bulger M, Groudine M: **Enhancers: the abundance and function of regulatory sequences beyond promoters.** *Dev Biol* 2010, **339**(2):250–257.
- Calo E, Wysocka J: **Modification of enhancer chromatin: what, how, and why?** *Mol Cell* 2013, **49**(5):825–837.
- Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, Boyer LA, Young RA, Jaenisch R: **Histone H3K27ac separates active from poised enhancers and predicts developmental state.** *Proc Natl Acad Sci USA* 2010, **107**(50):21931–21936.
- Ferretti V, Poitras C, Bergeron D, Coulombe B, Robert F, Blanchette M: **PreMod: a database of genome-wide mammalian cis-regulatory module predictions.** *Nucleic Acids Res* 2007, **35**:D122–D126.
- Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM, Edsall L, Antosiewicz-Bourget J, Stewart R, Ruotti V, Millar AH, Thomson JA, Ren B, Ecker JR: **Human DNA methylomes at base resolution show widespread epigenomic differences.** *Nature* 2009, **462**(7271):315–322.
- Mayshar Y, Ben-David U, Lavon N, Biancotti JC, Yakir B, Clark AT, Plath K, Lowry WE, Benvenisty N: **Identification and classification of chromosomal aberrations in human induced pluripotent stem cells.** *Cell Stem Cell* 2010, **7**(4):521–531.
- Huda A, Tyagi E, Marino-Ramirez L, Bowen NJ, Jjingto D, Jordan IK: **Prediction of transposable element derived enhancers using chromatin modification profiles.** *PLoS One* 2011, **6**(11):e27513.
- Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, Kent WJ: **The UCSC table browser data retrieval tool.** *Nucleic Acids Res* 2004, **32**:D493–D496.
- Rhead B, Karolchik D, Kuhn RM, Hinrichs AS, Zweig AS, Fujita PA, Diekhans M, Smith KE, Rosenbloom KR, Raney BJ, Pohl A, Pheasant M, Meyer LR, Learned K, Hsu F, Hillman-Jackson J, Harte RA, Giardine B, Dreszer TR, Clawson H, Barber GP, Haussler D, Kent WJ: **The UCSC Genome Browser database: update 2010.** *Nucleic Acids Res* 2010, **38**:D613–D619.
- Heintzman ND, Stuart RK, Hon G, Fu YT, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu CX, Ching KA, Wang W, Weng Z, Green RD, Crawford GE, Ren B: **Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome.** *Nat Genet* 2007, **39**(3):311–318.
- Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J: **A unique chromatin signature uncovers early developmental enhancers in humans.** *Nature* 2011, **470**(7333):279–283.
- Maston GA, Evans SK, Green MR: **Transcriptional regulatory elements in the human genome.** *Annu Rev Genom Hum G* 2006, **7**:29–59.
- Adunyah SE, Chander R, Barner VK, Cooper RS: **Regulation of c-jun mRNA expression by hydroxyurea in human K562 cells during erythroid differentiation.** *Biochim Biophys Acta* 1995, **1263**:123–132. A published erratum appears in *Biochim Biophys Acta* 1995, **1264**(3):409.
- Andrews NC: **The NF-E2 transcription factor.** *Int J Biochem Cell B* 1998, **30**(4):429–432.
- Chen RL, Chou YC, Lan YJ, Huang TS, Shen CK: **Developmental silencing of human ζ -globin gene expression is mediated by the transcriptional repressor RREB1.** *J Biol Chem* 2010, **285**(14):10189–10197.
- Hilberg F, Aguzzi A, Howells N, Wagner EF: **c-jun is essential for normal mouse development and hepatogenesis.** *Nature* 1993, **365**:179–181. A published erratum appears in *Nature* 1993, **366**(6453):368.
- Pages PA H, Gentleman R, DebRoy S: **Biostrings: string objects representing biological sequences, and matching algorithms.** <http://bioc.ism.ac.jp/2.6/bioc/html/Biostrings.html>.

44. Stormo GD, Zhao Y: **Determining the specificity of protein-DNA interactions.** *Nat Rev Genet* 2010, **11**(11):751–760.
45. Wang J, Zhuang JL, Iyer S, Lin XY, Greven MC, Kim BH, Moore J, Pierce BG, Dong XJ, Virgil D, Birney E, Hung JH, Weng Z: **Factorbook.org: a Wiki-based database for transcription factor-binding data generated by the ENCODE consortium.** *Nucleic Acids Res* 2013, **41**(D1):D171–D176.
46. Yanai I, Benjamin H, Shmush M, Chalifa-Caspi V, Shklar M, Ophir R, Bar-Even A, Horn-Saban S, Safran M, Domany E, Lancet D, Shmueli O: **Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification.** *Bioinformatics* 2005, **21**(5):650–659.
47. Schug J, Schuller WP, Kappen C, Salbaum JM, Bucan M, Stoeckert CJ: **Promoter features related to tissue specificity as measured by Shannon entropy.** *Genome Biol* 2005, **6**(4):R33.
48. Merryweather-Clarke AT, Atzberger A, Soneji S: **Global gene expression analysis of human erythroid progenitors.** *Blood* 2011, **117**:e96–e108. A published erratum appears in *Blood* 2011, **118**(26): 6993.
49. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G: **GREAT improves functional interpretation of cis-regulatory regions.** *Nat Biotechnol* 2010, **28**(5):495–U155.
50. Addya S, Keller MA, Delgrosso K, Ponte CM, Vadigepalli R, Gonye GE, Surrey S: **Erythroid-induced commitment of K562 cells results in clusters of differentially expressed genes enriched for specific transcription regulatory elements.** *Physiol Genomics* 2004, **19**(1):117–130.
51. Jordan IK, Rogozin IB, Glazko GV, Koonin EV: **Origin of a substantial fraction of human regulatory sequences from transposable elements.** *Trends Genet* 2003, **19**(2):68–72.
52. Lunyak VV, Prefontaine GG, Nunez E, Cramer T, Ju BG, Ohgi KA, Hutt K, Roy R, Garcia-Diaz A, Zhu X, Yung Y, Montolieu L, Glass CK, Rosenfeld MG: **Developmentally regulated activation of a SINE B2 repeat as a domain boundary in organogenesis.** *Science* 2007, **317**(5835):248–251.
53. Wang JR, Lunyak VV, Jordan IK: **Genome-wide prediction and analysis of human chromatin boundary elements.** *Nucleic Acids Res* 2012, **40**(2):511–529.
54. Lee YS, Shibata Y, Malhotra A, Dutta A: **A novel class of small RNAs: tRNA-derived RNA fragments (tRFs).** *Gene Dev* 2009, **23**(22):2639–2649.
55. Ren B: **Transcription enhancers make non-coding RNA.** *Nature* 2010, **465**(7295):173–174.
56. Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, Harmin DA, Laptewicz M, Barbara-Haley K, Kuersten S, Markenscoff-Papadimitriou E, Kuhl D, Bito H, Worley PF, Kreiman G, Greenberg ME: **Widespread transcription at neuronal activity-regulated enhancers.** *Nature* 2010, **465**(7295):182–187.
57. Li WB, Notani D, Ma Q, Tanasa B, Nunez E, Chen AY, Merkurjev D, Zhang J, Ohgi K, Song XY, Oh S, Kim HS, Glass CK, Rosenfeld MG: **Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation.** *Nature* 2013, **498**(7455):516–520.
58. Lam MTY, Cho H, Lesch HP, Gosselin D, Heinz S, Tanaka-Oishi Y, Benner C, Kaikkonen MU, Kim AS, Kosaka M, Lee CY, Watt A, Grossman TR, Rosenfeld MG, Evans RM, Glass CK: **Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription.** *Nature* 2013, **498**(7455):511–515.
59. Wang QB, Carroll JS, Brown M: **Spatial and temporal recruitment of androgen receptor and its coactivators involves chromosomal looping and polymerase tracking.** *Mol Cell* 2005, **19**(5):631–642.
60. Kim TH, Barrera LO, Zheng M, Qu CX, Singer MA, Richmond TA, Wu YN, Green RD, Ren B: **A high-resolution map of active promoters in the human genome.** *Nature* 2005, **436**(7052):876–880.
61. Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J: **Repbase update, a database of eukaryotic repetitive elements.** *Cytogenet Genome Res* 2005, **110**(1–4):462–467.
62. Kohany O, Gentles AJ, Hankus L, Jurka J: **Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor.** *BMC Bioinforma* 2006, **7**:474.
63. Johnson DS, Mortazavi A, Myers RM, Wold B: **Genome-wide mapping of in vivo protein-DNA interactions.** *Science* 2007, **316**(5830):1497–1502.
64. Frieze S, O'Geen H, Blahnik KR, Jin VX, Farnham PJ: **ZNF274 recruits the histone methyltransferase SETDB1 to the 3' ends of ZNF genes.** *PLoS One* 2010, **5**(12):e15082.
65. Hartman SE, Bertone P, Nath AK, Royce TE, Gerstein M, Weissman S, Snyder M: **Global changes in STAT target selection and transcription regulation upon interferon treatments.** *Gene Dev* 2005, **19**(24):2953–2968.
66. Letoile ND, Fahnestock ML, Shen YH, Aebersold R, Berk AJ: **Human transcription factor-IIIc box B binding subunit.** *Proc Natl Acad Sci USA* 1994, **91**(5):1652–1656.
67. Li ML, Ge QY, Wang W, Wang JK, Lu ZH: **c-jun binding site identification in K562 cells.** *J Genet Genomics* 2011, **38**(6):235–242.
68. Bryne JC, Valen E, Tang MHE, Marstrand T, Winther O, da Piedade I, Krogh A, Lenhard B, Sandelin A: **JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update.** *Nucleic Acids Res* 2008, **36**:D102–D106.
69. Bailey TL, Williams N, Misleh C, Li WW: **MEME: discovering and analyzing DNA and protein sequence motifs.** *Nucleic Acids Res* 2006, **34**:W369–W373.
70. Jjingo D, Conley AB, Yi SV, Lunyak VV, Jordan IK: **On the presence and role of human gene-body DNA methylation.** *Oncotarget* 2012, **3**(4):462–474.
71. Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, Zhang J, Soden R, Hayakawa M, Kreiman G, Cooke MP, Walker JR, Hogenesch JB: **A gene atlas of the mouse and human protein-encoding transcriptomes.** *Proc Natl Acad Sci USA* 2004, **101**(16):6062–6067.

doi:10.1186/1759-8753-5-14

Cite this article as: Jjingo et al.: Mammalian-wide interspersed repeat (MIR)-derived enhancers and the regulation of human gene expression. *Mobile DNA* 2014 **5**:14.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

