Genomics xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Genomic features of human limb specific enhancers

Shahid Ali [a], Bibi Amina [a], Saneela Anwar [b], Rashid Minhas [a], Nazia Parveen [a], Uzma Nawaz [c], Syed Sikandar Azam [b,*], Amir Ali Abbasi [a,*]

[a] National Center for Bioinformatics, Program of Comparative and Evolutionary Genomics, Faculty of Biological Sciences, Quaid-i-Azam University, Islamabad 45320, Pakistan
[b] National Center for Bioinformatics, Computational Biology Lab, Faculty of Biological Sciences, Quaid-i-Azam University, Islamabad 45320, Pakistan
[c] Department of Statistics, The Women University, Multan 60000, Pakistan

ARTICLE INFO

Article history:
Received 16 March 2016
Received in revised form 8 August 2016
Accepted 27 August 2016
Available online xxxx

Keywords:
Limbs
Enhancers
Genomics
Regulatory networks
Transcriptional factor
Development, disease

ABSTRACT

To elucidate important cellular and molecular interactions that regulate patterning and skeletal development, vertebrate limbs served as a model organ. A growing body of evidence from detailed studies on a subset of limb regulators like the *HOXD* cluster or *SHH*, reveals the importance of enhancers in limb related developmental and disease processes. Exploiting the recent genome-wide availability of functionally confirmed enhancer dataset, this study establishes regulatory interactions for dozens of human limb developmental genes. From these data, it appears that the long-range regulatory interactions are fairly common during limb development. This observation highlights the significance of chromosomal breaks/translocations in human limb deformities. Transcriptional factor (TF) analysis predicts that the differentiation of early nascent limb-bud into future territories entail distinct TF interaction networks. Conclusively, an important motivation for annotating the human limb specific regulatory networks is to pave way for the systematic exploration of their role in disease and evolution.

1. Introduction

Vertebrate fins and limbs are homologous structures, as they share striking conservation of developmental mechanisms [1–3]. Identification of many genes critical for limb development and patterning have affirmed the idea that despite of morphological and functional diversification from fish fin to tetrapod limb, vertebrate appendicular architecture is built upon a fairly similar repertoire of regulatory genes [4]. The tetrapod limb skeleton consists of three major segments, i.e. stylopod (humorous/femur), zeugopod (radius/ulna and tibia/fibula) and autopod (carpel/tarsal). Fossil records and comparative developmental studies suggest that the proximal limb regions (stylopod and zeugopod) have homologs in fish fin, but the origin of the most distal limb region, (autopod with digits) is an evolutionary novelty of tetrapods [2,5].

A growing body of evidence from detailed studies on a subset of limb regulators like the *HOXD* cluster or *SHH*, suggests that the major transformations in limb morphology during vertebrate history might entail *cis*-acting regulatory innovations [6]. However, the picture is still unclear, as *cis*-acting regulatory control of many crucial genes involved in limb patterning and development remains unknown [7]. The detection and functional analysis of *cis*-acting regulatory networks for key limb developmental regulators is considered to be a prerequisite for a better understanding of molecular evolutionary events that shaped the amazing architectural and functional diversification of limbs between tetrapod and fish lineages and within tetrapods [8]. Furthermore, the limb specific *cis*-regulatory catalog will possibly contribute in understanding the genetic basis of those limb related human birth defects where coding interval of concerned gene bodies are unaltered. However, this task remains difficult due to lack of knowledge of the vocabulary controlling gene regulation and the vast genomic search space [9–12]. Transcriptional process of a typical animal gene is governed by integrated action of the multiple distinct enhancers which can be positioned in 5′ and 3′ genomic regions, as well as within intronic intervals [13,14]. Many of such *cis*-regulators are remotely positioned from their target gene bodies [11,15]. Furthermore, the Metazoan *cis*-regulatory networks are often modular with each enhancer usually mediate expression within a specific tissue/cell type or developmental phase/domain. Enhancers are typically up to 500 bp long and contain binding sites for sequence specific several distinct TFs [13].

Recently the evolutionary conservation metric and ChIP-seq technique was employed to discern thousands of putative *cis*-regulatory elements in human genome [16]. Subset of these human genomic segments is analyzed in transgenic mice assay to verify their *in vivo* function and to find their tissue specificity. This dataset comprised of ~1000 functionally confirmed enhancers, directing reproducibly the

* Corresponding authors.
E-mail addresses: Shahidbsb28@gmail.com (S. Ali), bbamina120@gmail.com (B. Amina), amirsanila@yahoo.com (S. Anwar), minhas61@gmail.com (R. Minhas), naziaahmad.nm@gmail.com (N. Parveen), uzm_69@yahoo.com (U. Nawaz), ssazam@qau.edu.pk (S.S. Azam), abbasiam@qau.edu.pk (A.A. Abbasi).

http://dx.doi.org/10.1016/j.ygeno.2016.08.003
0888-7543/

reporter gene expression in diverse set of body tissues at mouse embryonic day 11.5 [16]. This large-scale availability of *in vivo* characterized enhancer dataset provides an unprecedented opportunity to reveal the significance of human *cis*-regulators in development, physiology and medicine. Furthermore, this dataset can help understand the genomic aspects of enhancer action and to decipher the transcriptional factor vocabulary of tissue specific gene regulatory networks.

The present study attempts to extract and interpret the genomic features of limb specific enhancers. For the purpose, 61/1154 functionally confirmed *cis*-regulatory regions with *in vivo* activity restricted to developing limb were short-listed. Using comparative synteny and comparing pattern of reporter expression induced by enhancers with the reported endogenous expression pattern of syntenically conserved genes, probable target gene bodies were identified for each subject. These enhancer-target gene associations were used as training dataset to gauge the range of action of limb specific enhancers. In addition, the study defines the cooperativity among distinct TFs during early limb growth and patterning.

2. Methods

2.1. Dataset collection

Experimentally validated human limb specific enhancers are retrieved from VISTA Enhancer Browser (http://enhancer.lbl.gov/) [16]. The detailed information on the data available at VISTA Enhancer Browser is briefly described previously [17]. Currently this database contains information on 2192 *in vivo* tested elements and 1154 of these elements show enhancer activity at embryonic day 11.5. This study is limited to 61 limb specific enhancers (Table 1, for detailed list see Supplementary Table 1). The degree of evolutionary sequence conservation of these elements varies from teleost fishes to afrotherians (Fig. 1). For each of the selected subset of enhancers the genomic sequence, vista id, enhancer coordinates, name of flanking genes, tissue specificity and image data were retrieved from VISTA Enhancer Browser.

2.2. Allocating target genes to the human limb specific enhancers

Association of human limb specific enhancers with their suitable target genes is based on two filters such as comparative genomics approach and endogenous expression pattern analysis as described previously [17] (Supplementary Table 1 and Supplementary Fig. 1).

After assigning target gene bodies to large numbers of limb specific enhancers, we next sought to estimate the genomic range of limb enhancer activity. For this purpose we calculated the distance between limb enhancers and transcriptional start site of their predicted target genes and then examined the distribution of distances. The range of limb enhancer action is partitioned as, enhancers embedded within intronic intervals of target gene (intragenic) and enhancers whose target gene lies within the ranges, e.g. 0–250 kb, 251–500 kb, 501–750, 751–1000 kb and >1 Mb (Supplementary Fig. 2).

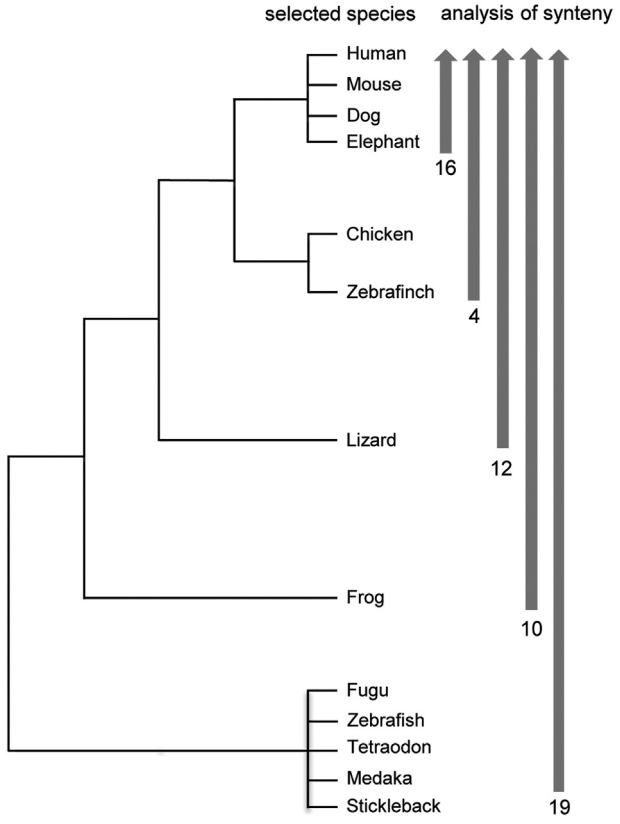

Fig. 1. Evolutionary relatedness of species used in comparative synteny analysis. Comparative analysis revealed varying depth of evolutionary conservation for functionally characterized limb specific enhancers. Majority of human limb enhancers were tetrapod specific (42/61) whereas only 19 enhancers were conserved down to teleost lineage. Vertical arrows depict the evolutionary depth of synteny analysis and number of enhancers falling in each category.

2.3. Transcription factor analysis

To establish the limb specific transcription factor code, 61 human enhancers that showed expressions exclusively in the developing mouse limb bud were opted (Supplementary Table 1). Mouse orthologs of human limb specific enhancers were acquired through BLAST based similarity searches. Human and mouse limb-specific enhancers were submitted to P-MATCH to predict the putative transcription factor binding sites. The P-MATCH combines pattern matching and weight matrix approaches to predict TFBSs (transcription factor binding sites) with high accuracy. Furthermore, this tool employs the TF binding sites library of TRANSFAC database and therefore provides the possibility to search for a large variety of distinct transcription factor binding sites (TFBSs) [18]. P-MATCH searches for putative TFBSs were performed against vertebrate matrices from the Transfac library with default parameters [18]. Binding sites for 111 distinct transcription factors were

Table 1
Association of human limb-specific enhancers with their putative target genes by employing comparative synteny analysis and expression pattern comparisons.

Sr. No	No. of enhancers	No. of target genes	Depth of synteny comparisons used to build enhancer-target gene association					Minimal evidence for association		
			Fish	Amphibians	Reptiles	Birds	Mammals	Orthology mapping (synteny)	Only synteny	Synteny & expression (MGI)
1.	39	1	16	5	8	2	8	✓	✓	✓
2.	14	2	2	4	2	2	4	✓	–	✓
3.	3	3	–	–	1	–	2	✓	–	✓
4.	5	>3	1	1	1	–	2	✓	–	✓

Sr. No: Serial Number.
MGI: Mouse Genome Informatics.