

RESEARCH HIGHLIGHT

# Making enhancers from spare parts of the genome

David Fredman<sup>1</sup>, Xianjun Dong<sup>2</sup> and Boris Lenhard<sup>3,4\*</sup>

## Abstract

New studies show that novel long-range enhancers of developmental genes can emerge by exaptation of protein-coding sequences with no previous regulatory function.

**Keywords** Comparative genomics, developmental regulation, enhancers, genomic regulatory blocks, whole-genome duplication.

Metazoan genomes contain tens of thousands of regions for which there has been recent experimental evidence for the binding of transcription factors and cofactors [1]. Many of these regions are conserved, sometimes through hundreds of millions of years of evolutionary history, as in the case of enhancers regulating early embryonic development [2]. There are numerous examples that show how new regulatory behavior can appear through modification of existing functional elements. At the same time, most known enhancers, including the most conserved ones, are unique sequences with no similarity to any of the elements that serve other functions, and there is increasing evidence showing that important functions are driven by regulatory elements that are poorly conserved, and showing a lineage-specific accelerated substitution rate in otherwise highly conserved elements. With this apparent dichotomy between turnover and conservation, it is legitimate to ask the question, 'Where and how do new enhancers come into existence?' Some enhancers have emerged from sequences that already had regulatory capacity; for example, duplication of single ancestral regulatory elements. Some others have come from mobile elements with the capacity to bind particular regulatory proteins that have integrated into regions in which nearby genes responded to them.

A new study by Eichenlaub and Ettwiller [3] investigated possible occurrences of an intriguing traceable scenario of *de novo* enhancer creation. By using a

combination of comparative genomics and *in vivo* testing, the study identified sequences that had undergone an apparent shift (exaptation) from an exclusively coding function to an exclusively *cis*-regulatory function, allowing their origin to be traced over hundreds of millions of years of evolutionary history.

## New enhancers recruited from duplicated protein-coding sequences

In principle, any of the sequences in a region from which a target gene is able to receive regulatory input could be exapted into a regulatory element, as long as it does not interfere with its other essential functions. Indeed, tens of thousands of elements overlapping coding exons in mammalian genomes have constrained selection on synonymous sites in their codons; Lin *et al.* [4] managed to assign putative function to 60% of them. The majority of functions, as expected, are splicing related, but other known overlapping functions have been detected: translational initiation, regulation of inclusion of cassette exons and, finally, developmental enhancers. The remaining 40% remain uncharacterized, but since they are enriched within developmental genes, a significant fraction of these are likely to have a regulatory role.

Eichenlaub and Ettwiller [3] explored an evolutionary scenario where an exonic remnant of a copy of a gene that was inactivated (non-functionalized) following teleost whole-genome duplication has acquired a regulatory function, as an enhancer driving part of the expression pattern of a neighboring developmental gene. They first searched for genomic regions in stickleback (*Gasterosteus aculeatus*) that are (1) conserved between human and stickleback; (2) non-coding in stickleback, but whose human ortholog regions are in the coding sequence; (3) near developmental genes. They identified four such exon-turned-enhancers, which they termed recycled regions, in the stickleback genome. The four corresponding human exons belong to the non-developmental genes *TTC29*, *DOCK9*, *CCDC46* and *FAM44B*.

The recycled regions annotated in the stickleback genome were transferred to the medaka genome for experimental validation. Three out of four recycled regions in the medaka genome showed enhancer activity, and each recapitulated part of the expression pattern of a neighboring developmental gene. The authors proceeded to show, for each of those sequences, that the medaka

\*Correspondence: b.lenhard@imperial.ac.uk

<sup>3</sup>Institute of Clinical Sciences, Imperial College London and MRC Clinical Sciences Centre, Hammersmith Hospital Campus, Du Cane Road, London W12 0NN, UK  
Full list of author information is available at the end of the article

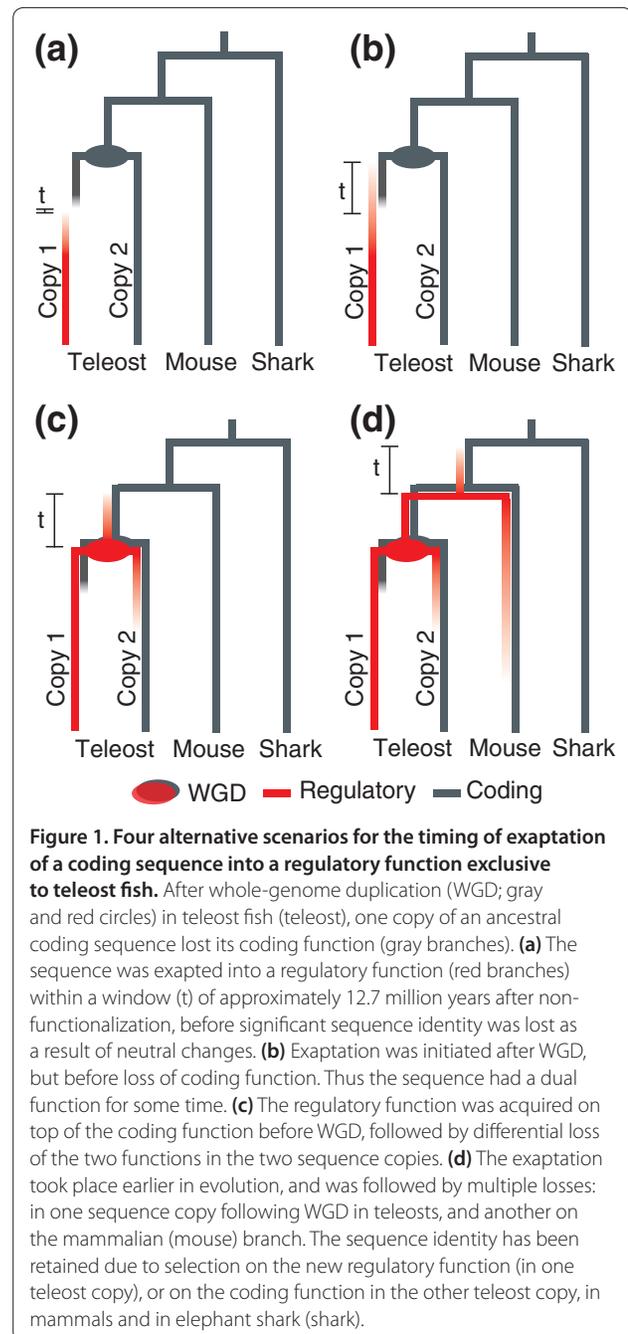
paralog, which is still a coding exon of an active protein-coding gene, does not have enhancer activity, and neither do the orthologous exons in mouse and elephant shark, which represent a sister group (tetrapods) and an out-group (cartilaginous fish), respectively. From this experimental evidence, the authors concluded that the exaptation of new enhancers occurred after whole-genome duplication at the root of teleost fish radiation and after inactivation of the copy of the gene from which the recycled region originated.

The suggested scenario poses some constraints. If the inactivation of the protein-coding gene preceded exaptation, the exaptation should have followed quickly thereafter. Otherwise, the exon sequence conservation would have rapidly decayed beyond recognition by neutral mutation within a relatively narrow window of several million years (Figure 1a). This would make this scenario rare, but not implausible. Indeed, the fact that only four elements were found (three of them in which the exonic remnant itself is required for enhancer function) suggests that this is a rare event.

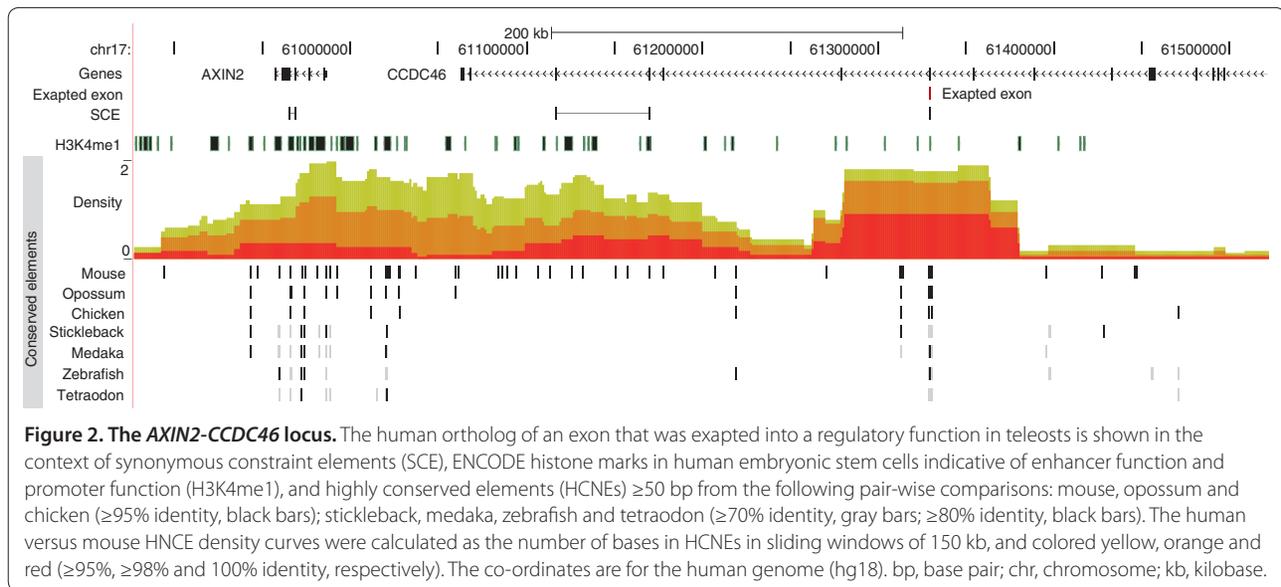
The presented data do not exclude modified or alternative scenarios. For example, the exons could have been co-opted for an enhancer role before the whole-genome duplication (Figure 1b), yielding a dual-function element (enhancer overlapping a functional coding exon) of the kind that has been shown in several other instances [5]. Co-option, in which an additional function is acquired by an existing functional element, could have been followed by the reciprocal loss of enhancer or exon function after the whole-genome duplication [6]. This scenario still fits with the enhancer as newly emerged and teleost specific, but might have the benefit of a significantly longer 'window of opportunity' for emergence without much sequence divergence, because at no time is selective pressure on the element removed.

A slight modification of the scenario depicted in Figure 1b would be that the co-option occurred in the post-whole-genome-duplication period while both copies of the original protein-coding gene were still functional (Figure 1c). Judging from rediploidization events in zebrafish relative to three other teleosts (medaka, stickleback and tetraodon), the post-whole-genome-duplication window of opportunity was also likely to be longer than that prior to whole-genome duplication, although one would assume that selective pressure to retain two copies of the gene was low. Other, more elaborate scenarios, such as that depicted in Figure 1d, would benefit from even longer windows of opportunity, and will only be possible to exclude after additional fish genome sequences become available.

One of the three elements tested by Eichenlaub and Ettwiller [3], the one originating from an exon of *ccdc46*, is shown by the authors to be near a developmental



enhancer that is conserved and functional in mouse, medaka and shark. The *ccdc46* exon sequence from either mouse or elephant shark does not drive expression on its own in their assays, and is not required for the function of the neighboring enhancer in mouse. However, based on analysis of synonymous conservation across coding exons of 29 eutherian mammals [4], the *ccdc46* exon itself overlaps with an element predicted to be still under selection on synonymous sites in eutherian mammals, and bears histone modifications associated with enhancer



**Figure 2. The *AXIN2-CCDC46* locus.** The human ortholog of an exon that was exapted into a regulatory function in teleosts is shown in the context of synonymous constraint elements (SCE), ENCODE histone marks in human embryonic stem cells indicative of enhancer function and promoter function (H3K4me1), and highly conserved elements (HCNEs)  $\geq 50$  bp from the following pair-wise comparisons: mouse, opossum and chicken ( $\geq 95\%$  identity, black bars); stickleback, medaka, zebrafish and tetraodon ( $\geq 70\%$  identity, gray bars;  $\geq 80\%$  identity, black bars). The human versus mouse HNCE density curves were calculated as the number of bases in HCNEs in sliding windows of 150 kb, and colored yellow, orange and red ( $\geq 95\%$ ,  $\geq 98\%$  and 100% identity, respectively). The co-ordinates are for the human genome (hg18). bp, base pair; chr, chromosome; kb, kilobase.

function (H3K4me1) in a subset of ENCODE (Encyclopedia of DNA Elements) cell lines (Figure 2). This indicates that a complex scenario and a contemporary dual role for the exon in mammals cannot be ruled out.

### Turnover and exaptation of enhancers around developmental genes

A key question raised in Eichenlaub and Ettwiller [3] concerns the rate of enhancer turnover and how easy it is to recruit new ones from sequences with no function or another, unrelated function.

Recent work has explored dual-function (coding exon + enhancer) elements [5], and their possible separation by reciprocal loss of the two functions after whole-genome duplication [6]. In each case, the coding exon-turned-enhancer is just one of a multitude of enhancers spanning a broad, often megabase-sized, region from which a target gene is able to receive its regulatory inputs. Such arrangements of elements are frequent around genes encoding key developmental regulators in Metazoa and have been described as genomic regulatory blocks [2].

The four genes (*TTC29*, *DOCK9*, *CCDC46* AND *FAM44B*) whose exonic remnants were exapted into enhancers show non-developmental expression patterns [3]. However, the neighboring target genes of these enhancers (*POU4F2* (*BNB1*), *ZIC2/ZIC5*, *AXIN2* and *NKX2.5*, respectively) are all target genes in genomic regulatory blocks, as evident from their biological function as developmental regulators, and from the distribution of a multitude of highly conserved elements around them (see Figure 2 for the *AXIN2* example). The number of recognizable conserved elements around these genes falls with increasing evolutionary distance, as would be expected with turnover of regulatory sequences. If we

extend the comparison to more distant species, the corresponding orthologous genes in *Drosophila* (*acj6*, *opa*, *axn* and *tin*, respectively) do not have any non-coding elements similar to these at the sequence level, but are instead spanned by their own sets of conserved elements/putative enhancers that can be aligned across drosophilids, but not vertebrate genomes. This means that, while the general regulatory architecture around those genes is similar, their regulatory element turnover since divergence from a common ancestor has been complete, at least in terms of sequence identity (discussed in [7]).

In addition, genomic regulatory blocks in mammals contain thousands of ancient mobile elements that have come under selective pressure [8], and numerous short elements enriched for histone modifications and transcriptional cofactors associated with enhancer functions spanning large regions around their target genes [9]. The conservation across vertebrates of the *ccdc46* exon and its nearby enhancer might also imply the importance of existing proximal regulatory elements in guiding the *de novo* genesis of new ones. The conservation pattern, the slow but steady turnover over long evolutionary times, and the recruitment of numerous elements of recognizable non-regulatory origin all indicate a high susceptibility of sequences within genomic regulatory blocks to recruitment into a regulatory role.

### Relevance and open questions

Eichenlaub and Ettwiller [3] showed how existing functional elements can be recycled for other functions, namely into developmental enhancers. Alongside transcription of non-coding RNAs from regions containing evidence of different past function, the recruitment of

sequences around developmental genes into their enhancer repertoire seems to be the most common and most readily detectable way to reuse an existing genomic sequence. The widespread occurrence of such events could be one of the major mechanisms of evolutionary innovation in Metazoa.

#### Competing interests

The authors declare that they have no competing interests.

#### Acknowledgements

BL acknowledges support from the Bergen Research Foundation (BFS), YFF project 180435 from the Norwegian Research Foundation (NFR), and the Medical Research Council, UK.

#### Author details

<sup>1</sup>Department of Molecular Evolution and Development, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria. <sup>2</sup>Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, 364 Plantation St, Worcester, MA 01605, USA. <sup>3</sup>Institute of Clinical Sciences, Imperial College London and MRC Clinical Sciences Centre, Hammersmith Hospital Campus, Du Cane Road, London W12 0NN, UK. <sup>4</sup>Department of Informatics, University of Bergen, PO Box 7803, N-5020 Bergen, Norway.

Published: 29 December 2011

#### References

1. Lee B-K, Bhinge AA, Battenhouse A, McDaniell RM, Liu Z, Song L, Ni Y, Birney E, Lieb JD, Furey TS, Crawford GE, Iyer VR: **Cell-type specific and combinatorial usage of diverse transcription factors revealed by genome-wide binding studies in multiple human cells.** *Genome Res* 2011, doi: 10.1101/gr.127597.111.

2. Kikuta H, Laplante M, Navratilova P, Komisarczuk AZ, Engstrom PG, Fredman D, Akalin A, Caccamo M, Sealy I, Howe K, Ghislain J, Pezeron G, Mourrain P, Ellingsen S, Oates AC, Thisse C, Thisse B, Foucher I, Adolf B, Geling A, Lenhard B, Becker TS: **Genomic regulatory blocks encompass multiple neighboring genes and maintain conserved synteny in vertebrates.** *Genome Res* 2007, **17**:545-555.
3. Eichenlaub MP, Ettwiller L: **De novo genesis of enhancers in vertebrates.** *PLoS Biol* 2011, **9**:e1001188.
4. Lin MF, Kheradpour P, Washietl S, Parker BJ, Pedersen JS, Kellis M: **Locating protein-coding sequences under selection for additional, overlapping functions in 29 mammalian genomes.** *Genome Res* 2011, **21**:1916-1928.
5. Lin Z, Ma H, Nei M: **Ultraconserved coding regions outside the homeobox of mammalian Hox genes.** *BMC Evol Biol* 2008, **8**:260.
6. Dong X, Navratilova P, Fredman D, Drivenes O, Becker TS, Lenhard B: **Exonic remnants of whole-genome duplication reveal cis-regulatory function of coding exons.** *Nucleic Acids Res* 2010, **38**:1071-1085.
7. Engstrom PG, Ho Sui SJ, Drivenes O, Becker TS, Lenhard B: **Genomic regulatory blocks underlie extensive microsynteny conservation in insects.** *Genome Res* 2007, **17**:1898-1908.
8. Lowe CB, Bejerano G, Haussler D: **Thousands of human mobile element fragments undergo strong purifying selection near developmental genes.** *Proc Natl Acad Sci U S A* 2007, **104**:8005-8010.
9. Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, Afzal V, Ren B, Rubin EM, Pennacchio LA: **ChIP-seq accurately predicts tissue-specific activity of enhancers.** *Nature* 2009, **457**:854-858.

doi:10.1186/gb-2011-12-12-138

**Cite this article as:** Fredman D, *et al.*: Making enhancers from spare parts of the genome. *Genome Biology* 2011, **12**:138.