

Research article

Open Access

## Multi-species sequence comparison reveals dynamic evolution of the elastin gene that has involved purifying selection and lineage-specific insertions/deletions

Helen Piontkivska<sup>1,2</sup>, Yi Zhang<sup>3</sup>, Eric D Green<sup>4</sup>, NISC Comparative Sequencing Program<sup>4</sup> and Laura Elnitski\*<sup>5</sup>

Address: <sup>1</sup>Department of Biology, Pennsylvania State University, University Park, PA 16802, USA, <sup>2</sup>Department of Biological Sciences, University of South Carolina, Columbia, SC 29208, USA, <sup>3</sup>Department of Computer Science and Engineering, Center for Comparative Genomics and Bioinformatics, Pennsylvania State University, University Park, PA 16802, USA, <sup>4</sup>Genome Technology Branch and NIH Intramural Sequencing Center, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA and <sup>5</sup>Department of Computer Science and Engineering, Center for Comparative Genomics and Bioinformatics, Pennsylvania State University, University Park, PA 16802, USA

Email: Helen Piontkivska - elena@biol.sc.edu; Yi Zhang - yuz10@psu.edu; Eric D Green - egreen@nhgri.nih.gov; NISC Comparative Sequencing Program - ; Laura Elnitski\* - LLB111@psu.edu

\* Corresponding author

Published: 18 May 2004

Received: 22 December 2003

*BMC Genomics* 2004, **5**:31

Accepted: 18 May 2004

This article is available from: <http://www.biomedcentral.com/1471-2164/5/31>

© 2004 Piontkivska et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

### Abstract

**Background:** The elastin gene (ELN) is implicated as a factor in both supravalvular aortic stenosis (SVAS) and Williams Beuren Syndrome (WBS), two diseases involving pronounced complications in mental or physical development. Although the complete spectrum of functional roles of the processed gene product remains to be established, these roles are inferred to be analogous in human and mouse. This view is supported by genomic sequence comparison, in which there are no large-scale differences in the ~1.8 Mb sequence block encompassing the common region deleted in WBS, with the exception of an overall reversed physical orientation between human and mouse.

**Results:** Conserved synteny around ELN does not translate to a high level of conservation in the gene itself. In fact, ELN orthologs in mammals show more sequence divergence than expected for a gene with a critical role in development. The pattern of divergence is non-conventional due to an unusually high ratio of gaps to substitutions. Specifically, multi-sequence alignments of eight mammalian sequences reveal numerous non-aligning regions caused by species-specific insertions and deletions, in spite of the fact that the vast majority of aligning sites appear to be conserved and undergoing purifying selection.

**Conclusions:** The pattern of lineage-specific, in-frame insertions/deletions in the coding exons of ELN orthologous genes is unusual and has led to unique features of the gene in each lineage. These differences may indicate that the gene has a slightly different functional mechanism in mammalian lineages, or that the corresponding regions are functionally inert. Identified regions that undergo purifying selection reflect a functional importance associated with evolutionary pressure to retain those features.

## Background

As part of the broader NISC Comparative Sequencing Program (see <http://www.nisc.nih.gov>), we generated sequences of the genomic region encompassing the locus commonly deleted in WBS in multiple mammalian species. The resulting multiple-sequence alignments provide the opportunity to examine properties of orthologous genes (such as ELN) that might clarify a functional role or contribution to human disease.

Elastin is a highly hydrophobic protein that plays a major role providing the property of elastic recoil in dermis, lungs, and blood vessels. Two types of alternating domains are characterized: (1) hydrophobic (rich in glycine, valine, and proline) and (2) crosslinking (rich in alanine and lysine) [1,2]. Both types of domains are encoded in distinct, usually alternating vertebrate exons, which may be subject to alternate splicing [3-7].

The single-copy ELN gene is located in a region of conserved synteny between human 7q11.23 and mouse chromosome 5G. This region is referred to as the Williams region [5,8] and encompasses roughly 30 genes. Genomic sequences of additional species such as baboon, cow, cat, and rat show conserved synteny around ELN (Pam Thomas, personal communication), thus indicating that the gross anatomy of the region does not contribute to any anticipated changes in ELN function between species. Moreover, when alignments are viewed at the nucleotide level, most coding exons in the large region display contiguous columns of nucleotides that are classified as a match or mismatch, but are devoid of gapped positions. Contrary to this common characteristic, two of the WBS genes, ELN [9] and WBSR15 [10], show strikingly less sequence conservation, both at the nucleotide and amino acid levels. Differences in the patterns of their divergence have not been previously analyzed.

A second, non-conventional feature of ELN is found in its structure. The exon/intron ratio is unusually low, reflecting small exons interspersed within large introns [3]. This feature represents a contradiction to the observed correlation between GC-rich genomic sequences and short intron length [11].

Mutations of ELN are implicated in several human disorders, including supravalvular aortic stenosis (SVAS), Williams syndrome [12], and cutis laxa [13,14]. Disease phenotypes are associated with multiple types of sequence changes including large deletions, translocations, nonsense-, frameshift-, and splice-site mutations [15-18]. Previous studies that examined limited alignments in the 3' end of the ELN locus in five species noted particularly dense accumulations of repetitive sequences

immediately downstream of the gene [4,5] that contribute to rearrangements in the primate lineage.

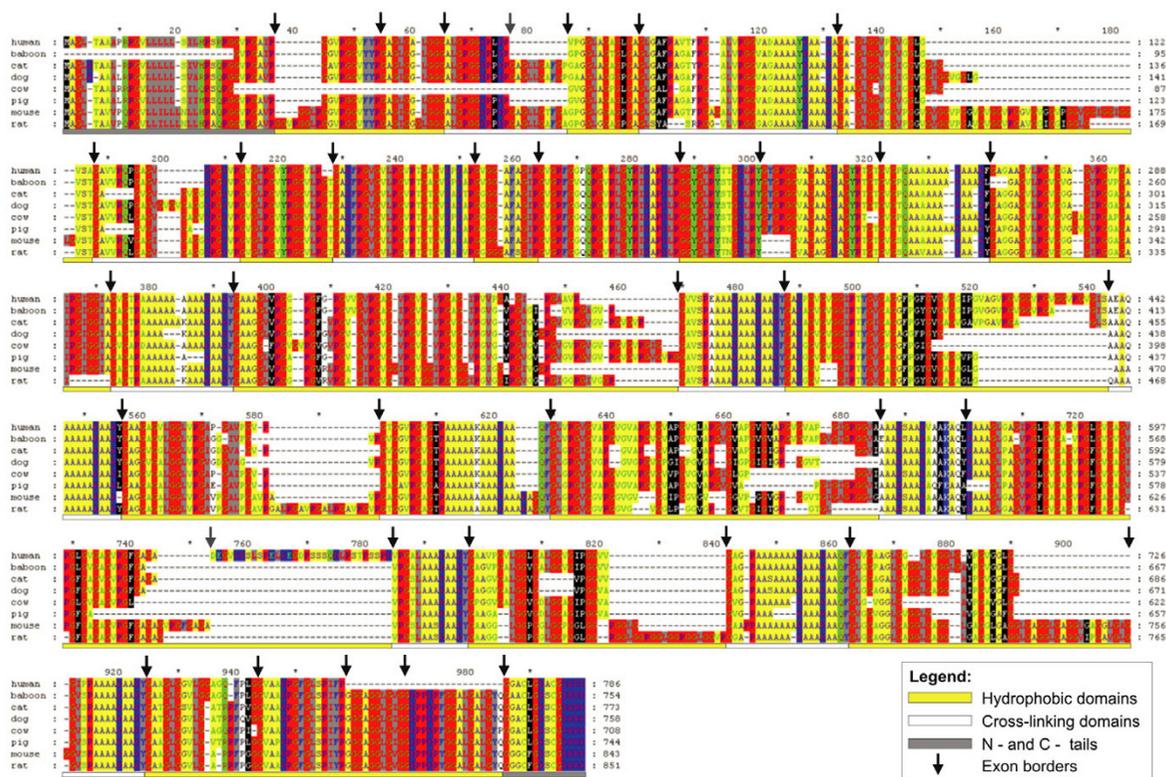
An analysis of multiple sequence alignments of the ELN gene from up to 9 species reveals the nucleotide substitution patterns at synonymous and non-synonymous sites. This approach identifies the presence of numerous in-frame insertions/deletions (in/dels) within coding exons that have lineage-specific characteristics but do not diminish the overall hydrophobic character of the protein. Contrary to the expectation that genes with such a high level of nucleotide divergence would be undergoing positive selection, the remaining bases are highly conserved and are undergoing purifying selection. These results have implications for the use of animal models with ELN mutations as well as for studies of genomic evolution.

## Results

### Gene structure

The expectation that orthologous coding exons in the 1.6-Mb genomic region commonly deleted in Williams Syndrome [9] align as highly conserved (gap-free) sequences among mammalian species is confirmed by the RFC2, LIMK1, and BAZ1B genes. ELN, however, is one example of a gene with a critical role in development that differs among its orthologs. Since the functional mechanism of the ELN protein is not fully elucidated, we sought to discern the nature of the divergence of this unusual gene. Although maintaining an overall interspersed structure of hydrophobic and cross-linking domains in all species, divergence occurs in the form of gaps in aligned coding exons and species-specific exons. The gapped alignment columns reveal a range of exon sizes among species (see Fig. 1 and supplemental Table A). Furthermore cow, pig, cat, and dog ELN genes have 36 exons, whereas mouse and rat ELN genes have 37 exons, due to an additional exon inserted after exon 4 (i.e., 4A). Some of these changes are recent; for instance the loss of two exons in human ELN (giving it 34 total) [19].

Alignment of human and mouse ELN cDNA sequences reveals 64.5% identity at the nucleotide level and 64.1% identity and 72.6% similarity at the amino acid level (with about 20% gaps; Table 2). This is well below the average percent identity of 85% at the nucleotide level and 78.5% at the amino acid level for human and mouse cDNAs [20] and is more similar to the average of 69% identity found in intronic regions. Although such divergence is typical of genes undergoing positive selection, these values are far below that of other genes in the region, with the exception of WBSR15/Wbscr15 [10]. Rat and mouse ELN proteins share the most similarity at 91% consistent with the average percent identity calculated for 11,071 known mouse and rat cDNA orthologs of 92.6%.



Notes: H,K,R - polar /positively charged amino acids; D,E - polar /negatively charged; N,Q - polar /amide; S,T - polar /alcohol; L,I,V - non-polar /aliphatic; F,Y,W - non-polar /aromatic; A,G,P - other non-polar. Domain information is given below the alignment; alternating cross-linking (white boxes) and hydrophobic (yellow boxes) domains are shown. Exon borders are marked with black arrows at the top. Grey arrows mark the beginning of exon 4A (found in rodents) and 26A (human-specific, [4]), respectively.

**Figure 1**

Multi-species alignment of ELN proteins from eight mammalian species. Amino acids with similar chemical characteristics are color-coded (see notes below). Human, cow, mouse, and rat are derived from GenBank sequences; baboon, cat, dog, and pig are predicted from genomic sequences based on the similarity to human and mouse ELN genes. Color legend: H, K, R – polar/ positively charged amino acids; D, E – polar/negatively charged; N, Q – polar/amide; S, T – polar/alcohol; L, I, V – non-polar/ aliphatic; F, Y, W – non-polar/aromatic; A, G, P – other non-polar. Domain information is shown below the alignment; alternating cross-linking (designated as white boxes) and hydrophobic (yellow boxes) domains are shown. Exon borders are marked with black arrows at the top. Grey arrows mark the beginning of exons 4A (found in rodents) and 26A (human-specific, [4]), respectively.

The lower level of sequence conservation among ELN orthologs is also associated with differences in gene structure among species. Although the splice junctions in the human [16] and other mammalian ELN genes in all cases contain consensus GT/AG signals, they vary in their positions in the different orthologs. This leads to differences in

the sizes of orthologous exons. For example, there are six ELN splice junctions that are not shared between mouse and human based on the whole-genome *blastz* alignments ([21] shown on Fig. 2). These splice junctions are either altered by a nucleotide change, fall in a region of sequence expansion/contraction, or are not represented in the

**Table 1: ELN-containing sequences used for comparative analyses.**

Species name	Type of sequence	Source (GenBank accession number)
Human ( <i>Homo sapiens</i> )	genomic cDNA cDNA	<a href="http://www.nisc.nih.gov">http://www.nisc.nih.gov</a> (2003 freeze) Fig. 1 of Tassabehji et al. 1997 [16] XP_004897
Baboon ( <i>Papio hamadryas</i> )	genomic	<a href="http://www.nisc.nih.gov">http://www.nisc.nih.gov</a> (2003 freeze)
Cow ( <i>Bos taurus</i> )	genomic	<a href="http://www.nisc.nih.gov">http://www.nisc.nih.gov</a> (2003 freeze)
Pig ( <i>Sus scrofa</i> )	genomic	<a href="http://www.nisc.nih.gov">http://www.nisc.nih.gov</a> (2003 freeze)
Cat ( <i>Felis catus</i> )	genomic	<a href="http://www.nisc.nih.gov">http://www.nisc.nih.gov</a> (2003 freeze)
Dog ( <i>Canis familiaris</i> )	genomic	<a href="http://www.nisc.nih.gov">http://www.nisc.nih.gov</a> (2003 freeze)
Mouse ( <i>Mus musculus</i> )	genomic cDNA	<a href="http://www.nisc.nih.gov">http://www.nisc.nih.gov</a> (2003 freeze) NM_007925
Rat ( <i>Rattus norvegicus</i> )	genomic cDNA	<a href="http://www.nisc.nih.gov">http://www.nisc.nih.gov</a> (2003 freeze) M60647

**Table 2: Percent similarity (%) in globally aligned amino acid (below diagonal) and nucleotide (above diagonal) sequences of ELN genes from eight mammalian species.**

	Human	Baboon	Cat	Dog	Cow	Pig	Mouse	Rat
Human		86.4	77.5	74.8	69.8	74.8	68.3	67.3
Baboon	87.3		78.5	78	74.8	79.4	68.5	67.1
Cat	81.6	84.4		86.1	75.2	82.4	71.9	71.6
Dog	79.6	83.9	90.6		76.4	80.4	71.9	70.9
Cow	75.4	80.7	80.8	81.9		79.1	63.4	63.5
Pig	80.7	85.8	86.8	86	85.2		68.6	68.2
Mouse	72.6	73.7	78.6	79.4	70.8	75.3		87.9
Rat	70.8	71.6	78.1	77.5	70.9	74.9	91	

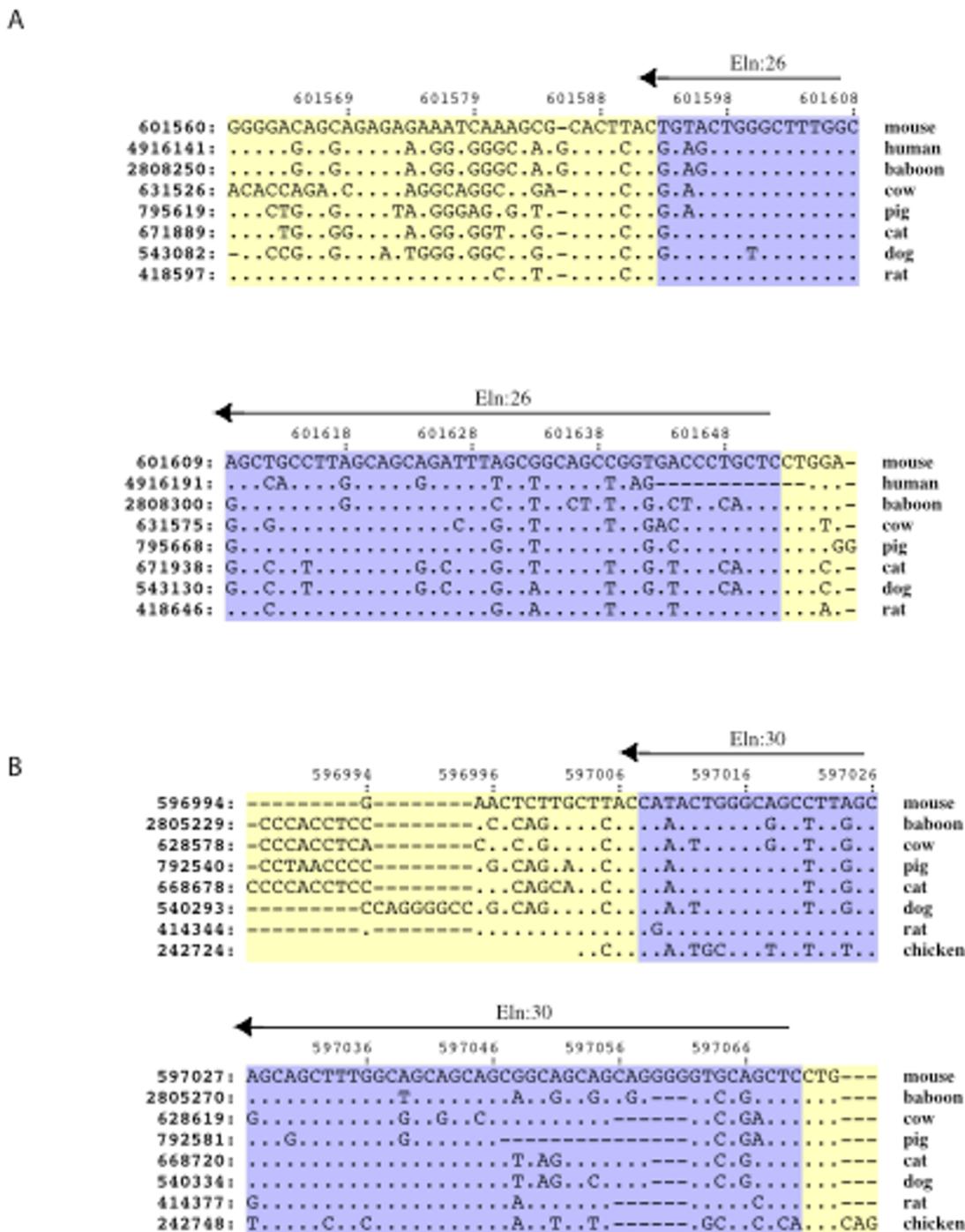
alignment because the exons failed to align. A closer look at these exons identifies a lineage-specific deletion in human ELN (Fig. 2 top panel) that corresponds to mouse exon 26 (at the 5' end of the exon). Internal deletions of varying sizes are illustrated in Fig. 2, bottom panel. Additionally, the use of alternative splice junctions can contribute to variable exon size; for instance, an alternate 5' splice junction in human is also completely conserved in baboon (data not shown).

The splicing pattern in human ELN uses a uniform phase 1 intron selection [16]. Here, the splice acceptor site falls after the first nucleotide of the last codon in an exon and the splice donor site falls immediately prior to the last two nucleotides of the interrupted codon. Interestingly, this pattern is also found in the ELN gene of seven other mammals, while being underrepresented in most other genes. Examination of 35,657 genes from the known gene track of the UCSC Genome Browser [22] shows that introns end in a phase 0, 1, and 2 fashion 43%, 32%, and 24% of the time, respectively; these figures are fairly consistent with previous studies done on fewer genes [23,24]. The

majority of genes, including many within the Williams region, use a mixture of these splice options; however, a minority of genes (2,422 or 6%) contains a uniform splicing pattern throughout. Of these, 42% are phase 0, 46% are phase 1, and 0.8% are phase 2. Such uniform phasing may increase the transcript diversity of a gene by providing options that create maximal flexibility in the alternative use of exons without disrupting the reading frame.

#### Sequence divergence among species

The numbers of synonymous and non-synonymous substitutions per synonymous and non-synonymous sites, respectively, were computed across the ELN gene using the Li-Wu-Luo method [25]. The complete-deletion option was used, thus only those codons shared among all species examined were included in the analyses. While the above findings suggested that ELN is under positive selection, studies of nucleotide divergence suggest otherwise, i.e., that ELN is evolving under strong purifying selection, as indicated by  $K_s$  being much higher than corresponding  $K_a$  values (Tables 3 and 4). Close inspection of the pair-wise and multi-species protein alignments



**Figure 2**

A nucleotide-level alignment of ELN in multiple species displaying variable structure in the coding regions, as aligned by Pip-Maker. The upper panel shows the alteration of a 5' splice junction due to a gap in the human sequence, indicating that some changes in the ELN gene are recent. The lower panel shows gaps in all sequences except one, which would normally indicate an insertion in that one sequence. However, the gaps differ in size (and are thus likely a product of multiple different events), suggesting that the region is indeed undergoing dynamic rearrangements. Panels A and B are from the same alignment, however differences result from exons that do not align in different species.

reveal that the majority of insertions are hydrophobic and targeted more frequently to the hydrophobic domains. Comparison of *Ka/Ks* ratios reveal that hydrophobic regions have accumulated significantly more non-synonymous changes than cross-linking regions (average *Ka/Ks* values are 0.217 and 0.121 for hydrophobic and cross-linking regions, respectively;  $t = 14.8, p < 0.001$ ). Nevertheless, the hydrophobic and crosslinking domains do not show a perceivable disparity in their functional importance, since both have significantly higher levels of synonymous versus non-synonymous substitutions for all species examined.

The majority of repeat elements in the elastin protein contain variations of PGVA or AAAAYKAA amino acid motifs that appear in the hydrophobic and crosslinking domains, respectively. There are many combinations of separated, tandem, and periodic repeats, present in lengths that vary from 5 to 41 amino acids; these data indicate that no particular repeat consensus is maintained. Furthermore, in contrast to deletions in the coding regions of the SCA and PRNP genes that align perfectly between human and mouse [26] the repeat elements in the elastin gene differ among species. Figure 3 shows two examples of the substitution patterns in repeat sites,

where expanded regions of the insertion in exons 6 and 8 are absent in the primate lineage (parts A and B, respectively; see also Figure 1). Many of the inserted amino acids are represented by 4-fold degenerate codons. The occurrence of identical amino acids in the repeat sequences reveals the use of different synonymous codons (e.g., this stretch of alanines: GCT-GCA-GCC-GCT-...GCA-GCT-...;) within the same species. This suggests that these insertion events did not result from the expansion of trinucleotide repeats and/or that mutation and repair mechanisms are not fully balanced. Furthermore, the inserted codons in closely related species show convergent differences at synonymous positions. For example, exon 6 of mouse, rat, cow and pig have an alanine insertion, yet, the mouse and pig gene share a C at the third codon position, while the rat and cow share a T. The same site in dog and primates lacks this insertion. Similarly, in exon 8, the sequence from cow has adjacent GAGV-repeats; yet, at the nucleotide level, this segment exhibits three synonymous differences at the third codon positions. One possible scenario that led to these changes is shown in Figure 3D. Thus none of the segments that contain repetitive variations of only few amino acids (e.g., PGVA or AAAAYKAA motifs) appear to be the result of internal duplications.

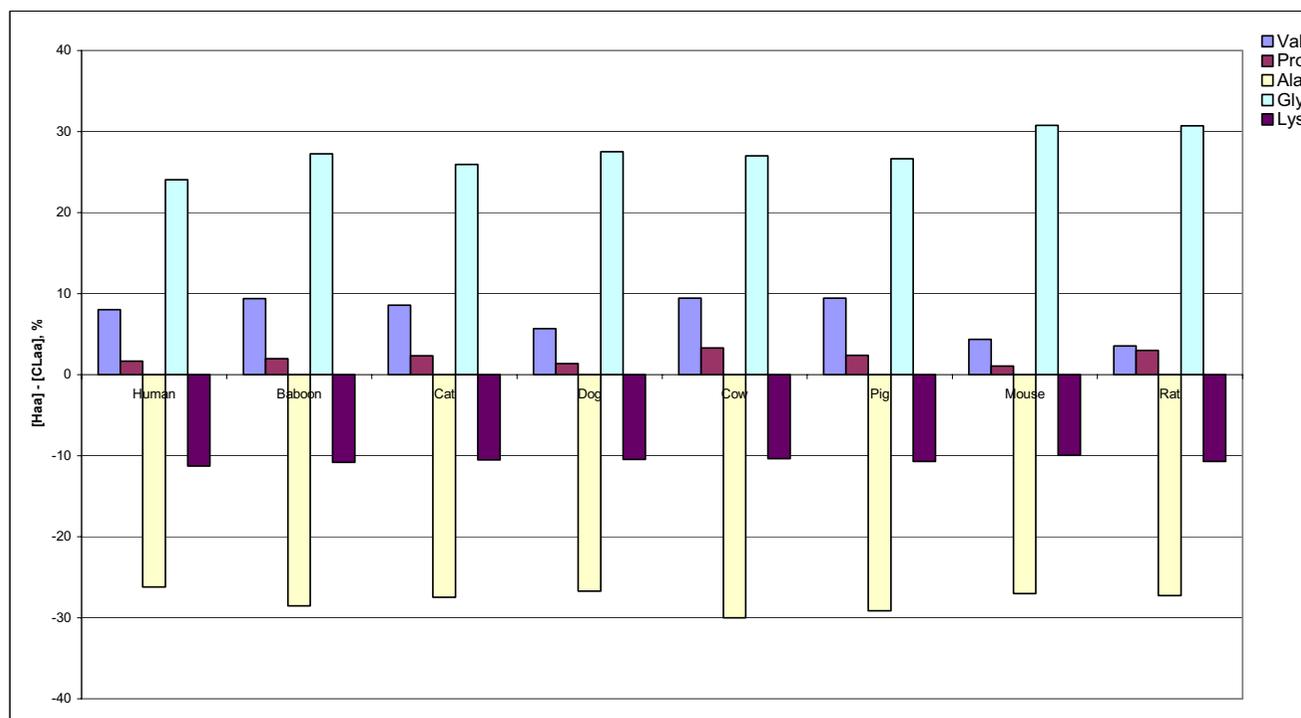
**Table 3: Numbers of synonymous substitutions per synonymous site (*Ks*) at hydrophobic (below diagonal) and cross-linking (above diagonal) regions of ELN gene. The number of substitutions was estimated with Li-Wu-Luo method [25] in pair-wise comparisons.**

<i>K<sub>s</sub></i>	Human	Baboon	Cat	Dog	Cow	Pig	Mouse	Rat
Human		0.13	0.529	0.519	0.546	0.686	0.835	0.785
Baboon	0.138		0.509	0.524	0.614	0.675	0.904	0.827
Cat	0.489	0.556		0.289	0.436	0.553	0.773	0.718
Dog	0.556	0.568	0.388		0.509	0.594	0.854	0.742
Cow	0.664	0.644	0.49	0.544		0.489	1.001	1.032
Pig	0.552	0.579	0.427	0.502	0.529		1.098	0.911
Mouse	0.626	0.673	0.744	0.678	0.847	0.703		0.286
Rat	0.653	0.684	0.811	0.767	0.827	0.736	0.258	

**Table 4: Numbers of nonsynonymous substitutions per nonsynonymous site (*Ka*) at hydrophobic (below diagonal) and cross-linking (above diagonal) regions of ELN gene. The number of substitutions was estimated with Li-Wu-Luo method [25] in pair-wise comparisons.**

<i>K<sub>a</sub></i>	Human	Baboon	Cat	Dog	Cow	Pig	Mouse	Rat
Human		0.048	0.058	0.082	0.072	0.108	0.05	0.025
Baboon	0.123		0.06	0.061	0.057	0.079	0.016	0.039
Cat	0.107	0.077		0.088	0.071	0.114	0.059	0.047
Dog	0.154	0.112	0.14		0.085	0.043	0.055	0.073
Cow	0.095	0.057	0.069	0.123		0.116	0.05	0.058
Pig	0.161	0.114	0.143	0.043	0.129		0.072	0.099
Mouse	0.141	0.061	0.115	0.125	0.102	0.129		0.038
Rat	0.044	0.093	0.087	0.146	0.076	0.151	0.127	





**Figure 4**

Compositional differences between hydrophobic (H) and crosslinking (CL) regions of ELN genes among eight mammalian species ( $H_{aa} - CL_{aa}$ , %).

Both types of domains contain a large amount of hydrophobic amino acids (hydrophobic regions are glycine-rich and crosslinking regions are alanine-rich), and have minor differences in the amino acid content among species (Figure 4). In particular, human and baboon have an excess of valine in their hydrophobic regions, whereas bovine has an excess of both valine and proline.

#### Genomic features

Several features of the ELN locus make it noteworthy. The genomic interval encompassing the gene has a high GC content (56%), which is notably higher than the human genome average of 41%. Similarly high GC content is seen for the orthologs in seven other vertebrates. The lowest GC content (in rodents, 50%), is still significantly higher than the mouse genome average of 42% [20]. Interestingly, the regions immediately flanking ELN show extreme dips in GC-content. For example, in mouse, a region of 42% GC is present 65 kb upstream of ELN, between two uncharacterized genes (matching pig\_EST\_BI340999 and mouse XM\_132419.1). In human, GC content dips to 39.9%, 60-kb upstream of ELN. The regions of low GC content are also found to be rich in SINES, LINES, LTRs, and simple repeats (data not

shown). Not surprisingly, several of the assembled sequences have contig gaps near the repeat insertions.

Ancestral repeats (ancestral repeat families are described in [20]) account for only 7% of the genomic interval containing the human ELN gene, whereas lineage-specific repeats account for 22% of interspersed repeat nucleotides. Typically, SINES are strongly biased toward regions of high GC content, as observed in the ELN locus. However, the relative abundance of SINES in the human (29%) and mouse (25%) ELN genes is higher than the genome averages (13.6% and 8.2%, respectively; [20]). The LTR content is highest in rat and mouse (although slightly lower than the mouse genome average), suggesting that these species have undergone the most insertion events.

Small scale deletion rates calculated from three-way alignments of rat, mouse and human show that rodents have fewer deletions (on the order of 10 bp) in the ELN locus than in the larger 1 Mb neighborhood (3.26 events per nucleotide for rat and 3.59 for mouse; versus 3.97 and 4.40, respectively [27]). Insertion rates, in contrast, are higher in the rat ELN locus (3.34) than in the 1 Mb neigh-

borhood (3.00) and the genome average (2.6 +/- 0.32). Using gap size of equal or less than 20 bp, we computed ratios of gap columns per coding nucleotides of the gene in pairwise human-mouse *blastz* alignments. Compared with the overall average ratio of 0.0218 gaps per coding nucleotide observed for human chromosome 7 genes, the ELN gene exhibited a striking ratio of 0.737 (species-specific exons were excluded). This observation is consistent with numerous small-scale gaps observed at the protein level in multi-species alignments (Figures 1 and 3). A similar analysis in WBSCR5 shows a ratio of .063, in which one insertion of SPGPV hydrophobic residues is found in the human protein, while the majority of the remaining gaps fall within one contiguous area where the mouse WBSCR5 gene is missing sequence orthologous to human exon 10.

To date, 32 SNPs have been mapped within a 40,513-bp region containing the human ELN gene. Only one of these SNPs is located within a coding region, representing a synonymous substitution in exon 20 (REFSNP # 2071307); this gives a frequency of 0.0004 SNPs per base pair in coding regions. The silent and non-coding polymorphisms do not appear to affect the function of the protein, in contrast to missense SNPs found in SVAS patients (with an occurrence of 1/20,000 births) [28], although their role(s) in the regulation of transcription cannot be ruled out. In SVAS patients, 20 different ELN exons show mutations that affect the reading frame [28]). Even at 1 mutation per exon, which is lower than observed, the resulting frequency of 0.009 changes per base pair differs from the frequency of synonymous changes by an order of magnitude. This suggests that although the ELN gene tolerates silent polymorphisms, mutations in the locus frequently lead to events that diminish the function of the protein.

## Discussion

Initial comparisons of the human and mouse genome sequences revealed three classes of genes associated with divergent coding regions, those encoding extracellular, immune system, and reproductive proteins [20]. The findings we report here, showing marked sequence divergence with the ELN gene, are consistent with the elastin protein being a major component of the extracellular matrix in vertebrates. Furthermore, the differential alternative splicing seen with the gene might be used to tailor the structural function of the protein in different tissues. Several alternatively spliced variants are associated with disease phenotypes such as SVAS [28]. On the other hand, one of alternatively spliced variants in human possesses highly hydrophilic exon 26A that may play an important role through its interactions with other matrix macromolecules [4] and therefore is selectively maintained in the genome.

Analyses of the human ELN gene reveal a dynamically evolving region of the genome. For instance, mutant phenotypes associated with various SNPs in SVAS patients suggest that the gene is susceptible to mutation. Additionally, the excision of two exons in the 3' end of the primate gene indicates that the region has undergone recent recombination. Other mechanisms that provide diversity include species-specific alternative splicing and lineage-specific exons. Thus, there are multiple factors driving ELN diversity (e.g., SNPs, recombination between repetitive elements, expansion and contraction of in/del units). These findings are congruent with conclusions drawn in Watanabe et al. (2002) [29], where it was shown that early/late transitions in replication timing correlate with high SNP frequency, increased DNA damage, transitions in GC content, and concentrated occurrence of disease-related genes.

Multi-species nucleotide-level alignments alone do not provide sufficient information to annotate the coding sequence of the ELN genes, primarily due to splice junctions that align out of register among species. This pattern results from the expansion and contraction of in/dels, which makes the alignment output more complex to interpret. Furthermore, amino acid alignments show that the non-conserved amino acids are limited mainly to hydrophobic or neutral sites. Many of these hydrophobic residues are encoded by 4-fold degenerate codons that can expand to "12-fold" degenerate sites. That is, amino acids V, P, A, and G are encoded by 4-fold degenerate codons, which in the case of V, P, and G start with the nucleotide G, followed by any G, C, or T in the second codon position. Finally, 4 nucleotide choices are available in the third codon. This suggests that nearly any nucleotide substitution in these sites is acceptable for the function of the elastin protein, as long as it encodes one of these hydrophobic amino acids. Therefore, our results indicate that nucleotide divergence is generally tolerated, but is localized to specific domains and/or sites (i.e., mainly hydrophobic regions). A potential mechanism for incorporating runs of similar codons is slippage during replication [30]. However, diseases attributed to trinucleotide repeat expansion have runs of a single codon (polyglutamine; reviewed in [31]) and not a varying repeat unit, as seen with ELN.

It appears unlikely that the amino acid insertions occurred in an ancient lineage, since the number of duplicated in/dels varies drastically among the modern sequences. The identity of the variable repeat unit resembles the recognition site of the elastin-binding protein, PGAIPG [32], located in human exon 16 [15]. This repeat is in a conserved position in cow, but not other species. Variations in the hydrophobic repeat sequence could create new recognition sites for this or other interacting molecules, provid-

ing opportunities for modulating the level and type of interaction. For example, some ligands for the bovine elastin receptor are chemotactic for neutrophils and fibroblasts [33,34]. If this were the case for human elastin, then one would expect similar variability in other species. For instance VGVAGP is also recognized as a ligand in human, cow [32], and baboon [19], but is not conserved in other species. Other identified ligands for the elastin receptor in bovine (VGAMPG, VGMAPG, VGSLPG, VGLSPG, and GIAPG) are also chemotactic for neutrophils and fibroblasts [33,34], although they are not represented in the elastin protein. Since the ligand-binding sites are not conserved, it is likely that each species has its own recognition signal and the diversity of the repeats expands the repertoire of interacting molecules.

Beyond the coding region, evolutionary changes of the ELN gene are also occurring through insertion and recombination of repetitive elements. In general, the ELN gene in mammalian species contains a somewhat higher proportion of repeat elements than the rest of the genome, with numerous repeats concentrated ~70 kb upstream of the first coding exon (in a region of extremely low GC content). These repeat elements are related in the rodent lineage, but do not align to other species for which sequence is available (e.g., cat and human), implying that they occurred after the human-rodent speciation event. The cat sequence shows an abundance of LINE1 elements in its region of low GC content (data not shown), whereas rodent sequences show a similar distribution of SINEs, LINEs, and simple sequences. Coupled with evidence of recombination between ALU repeats in primate species [19], these observations suggest that the region may contain a hot spot for insertion of repetitive elements.

Our analyses indicate that although ELN appears to be accumulating mutations (both nucleotide substitutions and in/dels) in all mammalian lineages examined, the locations and identity of these changes vary among species, perhaps with different functional consequences. The fact that the majority of amino acid substitutions and in/dels observed among eight mammalian species involve hydrophobic amino acids may reflect selective pressure to preserve the hydrophobic nature of the elastin protein. Furthermore, most amino acid substitutions do not change the chemical properties of the particular site, but rather interchange similar amino acids (e.g., exchange of glycine with valine or proline). Most of these evolutionary changes are concentrated in the hydrophobic regions of the protein, which represent a more flexible part of the protein (despite the overall hydrophobicity, these regions remain accessible on the protein surface [35]). Our results suggest that changes in the hydrophobic regions do not lead to major alterations to the structure and function of the elastin protein, as long as the overall hydrophobic

properties are maintained. In contrast, crosslinking domains allow fewer substitutions and in/del events than hydrophobic domains. This can be explained by purifying selection that not only maintains the primary structure of the protein, but also affects the spatial arrangement of crosslinking regions (i.e., the main components responsible for elasticity). In short, the combination of purifying selection acting to preserve the hydrophobic nature of the protein and in-frame in/del events have shaped the evolution of the ELN gene in mammals. Similar multi-species comparisons are needed to establish if this pattern is specific for ELN or is more generally the case for genes encoding structural proteins.

## Conclusions

Vertebrate ELN is a divergent gene that plays an important role in human health. Key structural and functional elements of the gene, including sequences of splice junctions and alternation of hydrophobic and crosslinking domains, are retained through purifying selection and remain virtually unchanged across large phylogenetic distances. However, remaining parts of the gene (such as repeated PGVA or AAAAYKAA amino acid motifs, and others) are more flexible and subject to in/del events that are selected against in the majority of all other protein-coding genes in the mammalian genomes. The hydrophobic repeat elements that comprise the insertion elements are likely to play a role in diversifying the repertoire of the interaction domain of the elastin protein in each species.

## Methods

A list of sequences used in this study is provided in Table 1. The multi-species genomic sequences were generated as part of the NISC Comparative Sequencing Program ([36]; <http://www.nisc.nih.gov>; 2003 freeze). The human ELN gene structure was taken from [16]. Alternative splicing in human ELN is documented in GenBank reference XP\_004897.2 and PIR entry EAHU.

Alignment of genomic sequences was performed using the PipMaker and MultiPipMaker servers running the *blastz* alignment program ([21]). The comparisons were computed as local alignments using the default parameters of the program [21]. Pair-wise cDNA and protein comparisons used the global alignment program ALIGN, available at EBI, to calculate a percent identity and similarity. Ratio of gaps per coding region was computed using whole genome human-mouse alignments with *blastz* program [21].

The exon numbers for the ELN gene correspond to those of the human cDNA sequence in [16]. Exons that are absent in humans but present in other species are numbered and lettered consecutively (i.e., the last exon in bovine ELN is 36 and corresponds to human exon 34; see

supplementary Table A). An additional exon is shared between the mouse mRNA sequence (NM\_007925, [37]) and the rat mRNA (M60647, [38]) (see Figure 1). Calculations of mouse-rat similarity in coding sequences used data from reciprocal best matches available from the Homologene database.

Because of the low exon/intron ratio, automated gene prediction fails to find many exons in the ELN gene. In this study, exons were identified by homology at the nucleotide and amino acid levels, and around the putative splice sites [16]. Translated amino acid sequences were aligned using ClustalX [39], and then manually adjusted to preserve the exon correspondence among species. Nucleotide sequences from the corresponding coding regions were then aligned according to the amino acid alignment using the program BioEdit (T. Hall). The shading of the amino acid alignment that reflects chemical properties of amino acids (Figure 1) was performed using the program GeneDoc (Nicholas K.B. and Nicholas H.B. Jr.).

The number of synonymous substitutions per synonymous site ( $K_s$ ) and the number of nonsynonymous substitutions per nonsynonymous site ( $K_a$ ) were estimated using the Li-Wu-Luo method [25], as implemented in the program MEGA2 [40].

### Abbreviations

ELN – Elastin, SVAS – supraaortic stenosis, bp – base pair, SNP – single nucleotide polymorphism, in/del – insertion/deletion

### Authors' contributions

LE contributed alignment data and observations about divergence and oversaw the project. HP carried out gene structure prediction, phylogenetic analysis, alignment refinements, and protein analysis. YZ calculated intron splice site frequencies and homology among human, mouse, and rat orthologs. EG and NISC contributed sequences, access to data, and editorial support. All authors read and approved the final manuscript.

### Web Site Reference

The Human Genome Browser <http://genome.ucsc.edu/cgi-bin/hgGateway>

The SNP Consortium LTD <http://snp.cshl.org/>

ENTREZ SNP <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=snp>

BlastZ Alignment Program <http://bio.cse.psu.edu/>

European Bioinformatics Institute <http://www.ebi.ac.uk>

Homologene Database <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=homologene>

BioEdit Program <http://www.mbio.ncsu.edu/BioEdit/bioedit.html>

GeneDoc Program <http://www.psc.edu/biomed/genedoc>

### Acknowledgements

We thank Webb Miller, Pam Thomas, Anton Nekrutenko, and Ian Korf for discussions about annotations of the region. We also thank all members of the NISC Comparative Sequencing Program for generating the genomic sequence data. LE is supported by NHGRI grant HG-02325 and HP by NIH grant GM20293 to Masatoshi Nei and by NIH grant GM066710 to Austin L. Hughes.

### References

1. Parks WC, Deak SB: **Tropoelastin heterogeneity: implications for protein function and disease.** *Am J Respir Cell Mol Biol* 1990, **2**:399-406.
2. Mecham RP: **Elastin.** *Guidebook to the Extracellular Matrix, Anchor and Adhesion Proteins Second Edition* Edited by: Edited by Thomas Kreis late Professor Département de Biologie Cellulaire Université de r s l l 'eve and Ronald Vale Professor Department of Pharmacology University of California San Francisco. OUP/Sambrook and Tooze Publications (588 pages); 1999:414-417.
3. Indik Z, Yeh H, Ornstein-Goldstein N, Kucich U, Abrams W, Rosenbloom JC, Rosenbloom J: **Structure of the elastin gene and alternative splicing of elastin mRNA: implications for human disease.** *Am J Med Genet* 1989, **34**:81-90.
4. Indik Z, Yeh H, Ornstein-Goldstein N, Sheppard P, Anderson N, Rosenbloom JC, Peltonen L, Rosenbloom J: **Alternative splicing of human elastin mRNA indicated by sequence analysis of cloned genomic and complementary DNA.** *Proc Natl Acad Sci U S A* 1987, **84**:5680-5684.
5. Bashir MM, Indik Z, Yeh H, Ornstein-Goldstein N, Rosenbloom JC, Abrams W, Fazio M, Uitto J, Rosenbloom J: **Characterization of the complete human elastin gene. Delineation of unusual features in the 5'-flanking region.** *J Biol Chem* 1989, **264**:8887-8891.
6. Boyd CD, Christiano AM, Pierce RA, Stolle CA, Deak SB: **Mammalian tropoelastin: multiple domains of the protein define an evolutionarily divergent amino acid sequence.** *Matrix* 1991, **11**:235-241.
7. Mauch JC, Sandberg LB, Roos PJ, Jimenez F, Christiano AM, Deak SB, Boyd CD: **Extensive alternate exon usage at the 5' end of the sheep tropoelastin gene.** *Matrix Biol* 1995, **14**:635-641.
8. Fazio MJ, Mattei MG, Passage E, Chu ML, Black D, Solomon E, Davidson JM, Uitto J: **Human elastin gene: new evidence for localization to the long arm of chromosome 7.** *Am J Hum Genet* 1991, **48**:696-703.
9. DeSilva U, Elnitski L, Idol JR, Doyle JL, Gan W, Thomas JW, Schwartz S, Dietrich NL, Beckstrom-Sternberg SM, McDowell JC, Blakesley RW, Bouffard GG, Thomas PJ, Touchman JW, Miller W, Green ED: **Generation and comparative analysis of approximately 3.3 mb of mouse genomic sequence orthologous to the region of human chromosome 7q11.23 implicated in Williams syndrome.** *Genome Res* 2002, **12**:3-15.
10. Doyle JL, DeSilva U, Miller W, Green ED: **Divergent human and mouse orthologs of a novel gene (WBSR15/Wbscr15) reside within the genomic interval commonly deleted in Williams syndrome.** *Cytogenet Cell Genet* 2000, **90**:285-290.
11. 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, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt

- A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramses J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blocker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kasprzyk S, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ: **Initial sequencing and analysis of the human genome.** *Nature* 2001, **409**:860-921.
12. Schmidt MA, Ensing GJ, Michels VV, Carter GA, Hagler DJ, Feldt RH: **Autosomal dominant supravulvar aortic stenosis: large three-generation family.** *Am J Med Genet* 1989, **32**:384-389.
  13. Zhang MC, Giro M, Quagliano D., Jr., Davidson JM: **Transforming growth factor-beta reverses a posttranscriptional defect in elastin synthesis in a cutis laxa skin fibroblast strain.** *J Clin Invest* 1995, **95**:986-994.
  14. Zhang MC, He L, Giro M, Yong SL, Tiller GE, Davidson JM: **Cutis laxa arising from frameshift mutations in exon 30 of the elastin gene (ELN).** *J Biol Chem* 1999, **274**:981-986.
  15. Li DY, Toland AE, Boak BB, Atkinson DL, Ensing GJ, Morris CA, Keating MT: **Elastin point mutations cause an obstructive vascular disease, supravulvar aortic stenosis.** *Hum Mol Genet* 1997, **6**:1021-1028.
  16. Tassabehji M, Metcalfe K, Donnai D, Hurst J, Reardon W, Burch M, Read AP: **Elastin: genomic structure and point mutations in patients with supravulvar aortic stenosis.** *Hum Mol Genet* 1997, **6**:1029-1036.
  17. Urban Z, Michels VV, Thibodeau SN, Donis-Keller H, Csiszar K, Boyd CD: **Supravulvar aortic stenosis: a splice site mutation within the elastin gene results in reduced expression of two aberrantly spliced transcripts.** *Hum Genet* 1999, **104**:135-142.
  18. Urban Z, Zhang J, Davis EC, Maeda GK, Kumar A, Stalker H, Belmont JW, Boyd CD, Wallace MR: **Supravulvar aortic stenosis: genetic and molecular dissection of a complex mutation in the elastin gene.** *Hum Genet* 2001, **109**:512-520.
  19. Szabo Z, Levi-Minzi SA, Christiano AM, Struminger C, Stoneking M, Batzer MA, Boyd CD: **Sequential loss of two neighboring exons of the tropoelastin gene during primate evolution.** *J Mol Evol* 1999, **49**:664-671.
  20. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, Antonarakis SE, Attwood J, Baertsch R, Bailey J, Barlow K, Beck S, Berry E, Birren B, Bloom T, Bork P, Botcherby M, Bray N, Brent MR, Brown DG, Brown SD, Bult C, Burton J, Butler J, Campbell RD, Carninci P, Cawley S, Chiaromonte F, Chinwalla AT, Church DM, Clamp M, Clee C, Collins FS, Cook LL, Copley RR, Coulson A, Couronne O, Cuff J, Curwen V, Cutts T, Daly M, David R, Davies J, Delehaunty KD, Deri J, Dermitzakis ET, Dewey C, Dickens NJ, Diekhans M, Dodge S, Dubchak I, Dunn DM, Eddy SR, Elnitski L, Emes RD, Eswara P, Eyraes E, Felsenfeld A, Fewell GA, Flicek P, Foley K, Frankel W, Fulton LA, Fulton RS, Furey TS, Gage D, Gibbs RA, Glusman G, Gnerre S, Goldman N, Goodstadt L, Grafham D, Graves TA, Green ED, Gregory S, Guigo R, Guyer M, Hardison RC, Haussler D, Hayashizaki Y, Hillier LW, Hinrichs A, Hlavina W, Holzer T, Hsu F, Hua A, Hubbard T, Hunt A, Jackson I, Jaffe DB, Johnson LS, Jones M, Jones TA, Joy A, Kamal M, Karlsson EK, Karolchik D, Kasprzyk A, Kawai J, Keibler E, Kells C, Kent WJ, Kirby A, Kolbe DL, Korf I, Kucherlapati RS, Kulbokas EJ, Kulp D, Landers T, Leger JP, Leonard S, Letunic I, Levine R, Li J, Li M, Lloyd C, Lucas S, Ma B, Maglott DR, Mardis ER, Matthews L, Mauceli E, Mayer JH, McCarthy M, McCombie WR, McLaren S, McLay K, McPherson JD, Meldrim J, Meredith B, Mesirov JP, Miller W, Miner TL, Mongin E, Montgomery KT, Morgan M, Mott R, Mullikin JC, Muzny DM, Nash WE, Nelson JO, Nhan MN, Nicol R, Ning X, Nusbaum C, O'Connor MJ, Okazaki Y, Oliver K, Overton-Larty E, Pachter L, Parra G, Pepin KH, Peterson J, Pevzner P, Plumb R, Pohl CS, Poliakov A, Ponce TC, Ponting CP, Potter S, Quail M, Reymond A, Roe BA, Roskin KM, Rubin EM, Rust AG, Santos R, Sapojnikov V, Schultz B, Schultz J, Schwartz MS, Schwartz S, Scott C, Seaman S, Searle S, Sharpe T, Sheridan A, Shownkeen R, Sims S, Singer JB, Slater G, Smit A, Smith DR, Spencer B, Stabenau A, Stange-Thomann N, Sugnet C, Suyama M, Tesler G, Thompson J, Torrents D, Trevaskis E, Tromp J, Ucla C, Ureta-Vidal A, Vinson JP, Von Niederhausern AC, Wade CM, Wall M, Weber RJ, Weiss RB, Wendl MC, West AP, Wetterstrand K, Wheeler R, Whelan S, Wierzbowski J, Willey D, Williams S, Wilson RK, Winter E, Worley KC, Wyman D, Yang S, Yang SP, Zdobnov EM, Zody MC, Lander ES: **Initial sequencing and comparative analysis of the mouse genome.** *Nature* 2002, **420**:520-562.
  21. Schwartz S, Kent WJ, Smit A, Zhang Z, Baertsch R, Hardison RC, Haussler D, Miller W: **Human-mouse alignments with BLASTZ.** *Genome Res* 2003, **13**:103-107.
  22. Karolchik D, Baertsch R, Diekhans M, Furey TS, Hinrichs A, Lu YT, Roskin KM, Schwartz M, Sugnet CW, Thomas DJ, Weber RJ, Haussler D, Kent WJ: **The UCSC Genome Browser Database.** *Nucleic Acids Res* 2003, **31**:51-54.
  23. Long M, Deutsch M: **Association of intron phases with conservation at splice site sequences and evolution of spliceosomal introns.** *Mol Biol Evol* 1999, **16**:1528-1534.
  24. Long M, Rosenberg C, Gilbert W: **Intron phase correlations and the evolution of the intron/exon structure of genes.** *Proc Natl Acad Sci U S A* 1995, **92**:12495-12499.
  25. Li WH, Wu CI, Luo CC: **A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes.** *Mol Biol Evol* 1985, **2**:150-174.
  26. Poux C, Van Rheede T, Madsen O, De Jong WW: **Sequence gaps join mice and men: phylogenetic evidence from deletions in two proteins.** *Mol Biol Evol* 2002, **19**:2035-2037.
  27. Yang S, Smit A, Schwartz S, Chiaromonte F, Roskin KM, Haussler D, Miller W, Hardison RC: **Patterns of insertions and their covariation with substitutions in the rat, mouse and human genomes. (in press Genome Res).** 2004.
  28. Metcalfe K, Rucka AK, Smoot L, Hofstadler G, Tuzler G, McKeown P, Siu V, Rauch A, Dean J, Dennis N, Ellis I, Reardon W, Cytrynbaum C, Osborne L, Yates JR, Read AP, Donnai D, Tassabehji M: **Elastin: mutational spectrum in supravulvar aortic stenosis.** *Eur J Hum Genet* 2000, **8**:955-963.
  29. Watanabe Y, Fujiyama A, Ichiba Y, Hattori M, Yada T, Sakaki Y, Ike-mura T: **Chromosome-wide assessment of replication timing for human chromosomes 11q and 21q: disease-related genes in timing-switch regions.** *Hum Mol Genet* 2002, **11**:13-21.
  30. Pearson CE, Wang YH, Griffith JD, Sinden RR: **Structural analysis of slipped-strand DNA (S-DNA) formed in (CTG)<sub>n</sub> (CAG)<sub>n</sub> repeats from the myotonic dystrophy locus.** *Nucleic Acids Res* 1998, **26**:816-823.
  31. Cummings CJ, Zoghbi HY: **Trinucleotide repeats: mechanisms and pathophysiology.** *Annu Rev Genomics Hum Genet* 2000, **1**:281-328.
  32. Grosso LE, Scott M: **Peptide sequences selected by BA4, a tropoelastin-specific monoclonal antibody, are ligands for the 67-kilodalton bovine elastin receptor.** *Biochemistry* 1993, **32**:13369-13374.
  33. Grosso LE, Scott M: **PGAIPG, a repeated hexapeptide of bovine and human tropoelastin, is chemotactic for neutrophils and Lewis lung carcinoma cells.** *Arch Biochem Biophys* 1993, **305**:401-404.

34. Grosso LE, Scott M: **PGAIPG, a repeated hexapeptide of bovine tropoelastin, is a ligand for the 67-kDa bovine elastin receptor.** *Matrix* 1993, **13**:157-164.
35. Wrenn DS, Griffin GL, Senior RM, Mecham RP: **Characterization of biologically active domains on elastin: identification of a monoclonal antibody to a cell recognition site.** *Biochemistry* 1986, **25**:5172-5176.
36. Thomas JW, Touchman JW, Blakesley RW, Bouffard GG, Beckstrom-Sternberg SM, Margulies EH, Blanchette M, Siepel AC, Thomas PJ, McDowell JC, Maskeri B, Hansen NF, Schwartz MS, Weber RJ, Kent WJ, Karolchik D, Bruen TC, Bevan R, Cutler DJ, Schwartz S, Elnitski L, Idol JR, Prasad AB, Lee-Lin SQ, Maduro VV, Summers TJ, Portnoy ME, Dietrich NL, Akhter N, Ayele K, Benjamin B, Cariaga K, Brinkley CP, Brooks SY, Granite S, Guan X, Gupta J, Haghghi P, Ho SL, Huang MC, Karlins E, Laric PL, Legaspi R, Lim MJ, Maduro QL, Masiello CA, Mastrian SD, McCloskey JC, Pearson R, Stantripop S, Tiongson EE, Tran JT, Tsurgeon C, Vogt JL, Walker MA, Wetherby KD, Wiggins LS, Young AC, Zhang LH, Osoegawa K, Zhu B, Zhao B, Shu CL, De Jong PJ, Lawrence CE, Smit AF, Chakravarti A, Haussler D, Green P, Miller W, Green ED: **Comparative analyses of multi-species sequences from targeted genomic regions.** *Nature* 2003, **424**:788-793.
37. Wydner KS, Sechler JL, Boyd CD, Passmore HC: **Use of an intron polymorphism to localize the tropoelastin gene to mouse chromosome 5 in a region of linkage conservation with human chromosome 7.** *Genomics* 1994, **23**:125-131.
38. Pierce RA, Deak SB, Stolle CA, Boyd CD: **Heterogeneity of rat tropoelastin mRNA revealed by cDNA cloning.** *Biochemistry* 1990, **29**:9677-9683.
39. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG: **The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools.** *Nucleic Acids Res* 1997, **25**:4876-4882.
40. Kumar S, Tamura K, Jakobsen IB, Nei M: **MEGA2: molecular evolutionary genetics analysis software.** *Bioinformatics* 2001, **17**:1244-1245.
41. Saitou N, Nei M: **The neighbor-joining method: a new method for reconstructing phylogenetic trees.** *Molecular Biology and Evolution* 1987, **4**:406-425.
42. Jukes TH, Cantor CR: **Evolution of protein molecules.** *Mammalian Protein Metabolism* Edited by: H N Munro. New York, Academic Press; 1969:21-132.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

