Primate-Specific Evolution of an LDLR Enhancer

Qian-fei Wang[1,3*], Shyam Prabhakar[1,3*], Qianben Wang[2], Alan M. Moses[1], Sumita Chanan[1], Myles Brown[2], Michael B. Eisen[1], Jan-Fang Cheng[1,3], Edward M. Rubin [1,3**] and Dario Boffelli [1,3]

1 Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720 USA.

2 Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA 02115 USA

3 U.S. Department of Energy Joint Genome Institute, Walnut Creek, California 94598 USA.

*: These authors contributed equally to this work.

**: Correspondence and requests for materials should be addressed to E.M.R. (emrubin@lbl.gov).

Running Title: Primate-Specific Enhancer Evolution

Abstract

Sequence changes in regulatory regions have often been invoked to explain phenotypic divergence among species, but molecular examples of this have been difficult to obtain. In this study, we identified an anthropoid primate specific sequence element that contributed to the regulatory evolution of the LDL receptor. Using a combination of close and distant species genomic sequence comparisons coupled with *in vivo* and *in vitro* studies, we show that a functional cholesterol-sensing sequence motif arose and was fixed within a pre-existing enhancer in the common ancestor of anthropoid primates. Our study demonstrates one molecular mechanism by which ancestral mammalian regulatory elements can evolve to perform new functions in the primate lineage leading to human.

Introduction

Since King and Wilson's provocative paper in 1975 [1], differences in gene regulatory sequences have been predicted to be one of the major sources of phenotypic evolution and divergence among animals. Consistent with this hypothesis, cis-regulatory changes have been found to play an important role in the evolution of morphological features in model organisms [2]. In contrast, evolution of physiology has been linked to changes in protein coding sequences, when studied in animal vision, digestive metabolism and host defense [3-7]. The contribution of regulatory sequence changes to the evolution of physiological differences, however, has been largely unexplored [8, 9].

To examine the role of cis-regulatory changes in the emergence of novel physiological traits in primates, we investigated the evolution of regulatory elements of the LDL receptor gene (*LDLR*), a key player in maintaining lipid homeostasis. Cholesterol metabolism in humans has diverged in a variety of ways from that of many distant mammals such as rodents and dogs, with humans in general being more susceptible to diet-induced hypercholesterolemia [10]. The pivotal role of *LDLR* in cholesterol metabolism coupled with its known expression differences among mammals[11] makes it a prime candidate for investigating primate-specific evolution of regulatory sequences. Here we present molecular data supporting the gain of a cholesterol sensing DNA motif in an ancestral mammalian *LDLR* regulatory element at a specific stage in primate evolution.

Results and Discussion

Identification of primate-specific noncoding elements in the *LDLR* locus

To identify putative primate-specific *LDLR* regulatory sequences, we examined orthologous regions from a panel of mammals closely and distantly related to human for the presence of evolutionarily conserved noncoding sequences using Gumby, an algorithm that detects sequence blocks evolving significantly more slowly than the local neutral rate (see

Methods)[12,13, 14]. Since humans and non-human primates share many features of cholesterol metabolism, we specifically scanned for elements that are preferentially conserved in primates under the hypothesis that primate specific regulatory sequences contribute to the distinctive biology of those species. Pairwise sequence comparisons of the 83-kb genomic region containing *LDLR* and its entire 5' and 3' intergenic regions between human and each of a panel of distantly related species consisting of the prosimian lemur, mouse and dog, identified either the known promoter sequence alone (Fig. 1, Panel A and data not shown), or a limited number of noncoding elements (Fig. S1 and data not shown). The promoter region was the only noncoding region consistently identified as conserved in the three pairwise comparisons. In contrast, multiple sequence comparisons between human and a set of five anthropoid primate species, chosen on the basis of their evolutionary relationship using the "phylogenetic shadowing" strategy [15], identified 2 human non-coding DNA elements, named PS (primate specific) 1 and 2, that are very significantly conserved ($P=\sim10^{-5}$) in primates (Fig. 1, Panel B) but undetected in comparisons involving human and each of the distant species (Fig. 1 and Fig. S1).

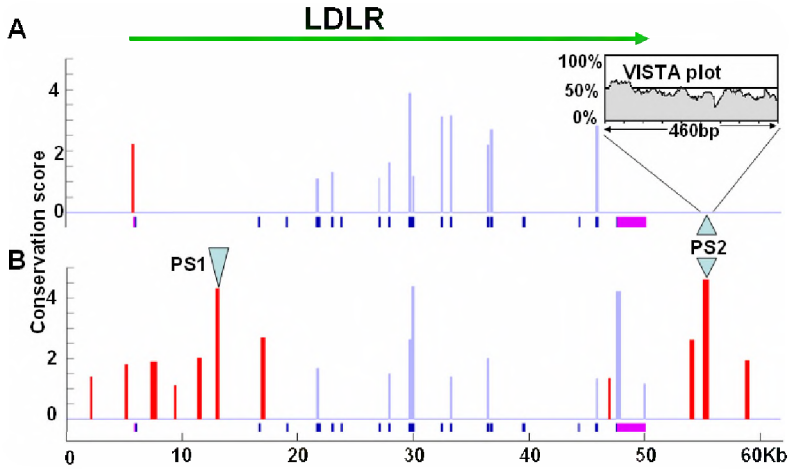

Fig.1 Conservation profiles of the *LDLR* locus using close (primate) and distant (human-mouse) species comparisons.

Human-mouse (A) and multiple-primate (human, baboon, colobus, dusky titi, marmoset, owl monkey) (B) conservation profiles were calculated using Gumby and visualized using RankVISTA (see Methods) and displayed with the human sequence as reference. Only ~6Kb of the 5' intergenic region is shown due to incomplete primate sequence availability. The entire 3' intergenic region was included in the analysis. Vertical bars depict conserved exonic (light blue) and non-exonic (red) sequences, with height indicating statistical significance of sequence conservation (see Methods). LDLR coding exons (dark blue) and UTRs (magenta) are marked below the conservation plots. Arrows denote the two highest-scoring primate-specific elements (PS1 and PS2). The inset shows the human-mouse VISTA plot for element PS2, with the vertical axis representing sequence identity calculated over a 100bp window.

To independently confirm the lack of significant conservation of the PS1 and PS2 elements between human and distant mammals, we also analyzed the human-mouse alignment using a sliding-window percent identity conservation criterion. We found that the human-mouse percent identities across PS1 and PS2 were lower than 50%(Fig. 1 and data not shown), which is close to the background percent identity in aligned intergenic DNA, and well below the threshold of 70% identity normally applied to the detection of conserved functional sequences [16]. We

further verified that the phastCons program [17] detects no conserved sequences overlapping PS1 and PS2 (data not shown). While the phastCons predictions, obtained from the UCSC Genome Browser, are in general based on alignment of 17 mammalian and non-mammalian species, conservation scores in the *LDLR* locus reflect only mammalian conservation, since more distant genomes show very limited non-exonic alignment in this locus.

To quantitatively assess the conservation level of PS1 and PS2 between human and distant mammals, we identified the orthologous aligned counterparts of the human PS1 and PS2 elements in lemur, mouse and dog. Gumby analysis of conservation scores indicated that each of these non-anthropoid primate sequences showed a level of similarity to the human sequence consistent with unconstrained evolution at the neutral rate (conservation p-value, Table I). Together, these analyses strongly suggest the lack of significant sequence constraint between the anthropoid primate and mammalian PS1 and PS2 sequences.

Table I. PS2 Enhancer Functional Divergence Correlates with Sequence Constraint. Conservation p-values are calculated using Gumby [12] under the null hypothesis of evolution at the neutral (background) rate. Low p-values indicate that the null model of neutrality should be rejected, with the lowest p-values identify the most significantly conserved sequences. The sequences analyzed for human-mammal conservation or enhancer activity correspond to the Gumby predicted conserved sequence and approximately 200bp of flanking sequence on either side (See Methods). Enhancer strength is shown as fold increase over promoter alone in luciferase assays in 293T cells.

Sequence Analysis			Functional Test		
Species compared	**Conservation p-value**		**Species assayed**	**Relative enhancer strength**	
	PS1	**PS2**		**PS1**	**PS2**
Human/5primates	4.8×10^{-5}	10^{-5}	Human	0.9	5.1
Human/Lemur	~1	0.76	Lemur	ND	2.6
Human/Mouse	~1	~0.99	Mouse	ND	1.5
Human/Dog	0.28	0.45	Dog	ND	2.6

ND: Not done.

The human *LDLR* PS2 element shows significantly higher enhancer activity than its mammalian orthologs

To explore the potential regulatory function of these two primate-specific conserved elements, we examined their ability to drive reporter gene expression in both a transient transfection assay in human 293T cells and an *in vivo* mouse liver gene transfer assay [18]. Each human element plus approximately 200bp of flanking sequence on either side was cloned upstream of the human *LDLR* promoter fused to a luciferase reporter gene. Human element PS2, but not PS1, consistently increased luciferase expression approximately 5-fold relative to the human promoter alone in both the *in vitro* and *in vivo* assays (Fig. 2). Enhancer activity of this

element was further confirmed by the finding that genomic region corresponding to PS2, but not PS1, is a DNase I hypersensitive site in human liver cells (Fig S2 and data not shown).

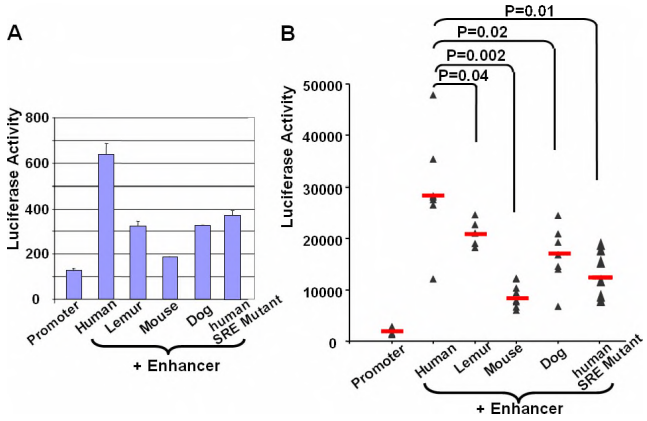

Fig. 2 The human *LDLR* PS2 enhancer shows significantly higher activity than the orthologous lemur, mouse and dog enhancers. Luciferase assay analysis of (A) transient transfections into human 293T cells and (B) plasmid DNA transfer into mouse liver. The luciferase reporter constructs tested are either the LDLR promoter alone (promoter), or the promoter in combination with the LDLR PS2 enhancer from one of the indicated species. Error bars indicate standard deviation. "SRE mutant" refers to the mutagenized human SRE with 4 point substitutions relative to the wild-type (WT) SRE (Fig. 4A). Luciferase activity is reported in arbitrary units. Each triangle in (B) represents luciferase activity in an individual mouse. Red bars denote the median activity of each construct.

To explore the regulatory function, if any, of mammalian sequences orthologous to human PS2, we cloned the PS2-aligned sequences from lemur, mouse and dog into the luciferase reporter vector described above and compared their activity to that of the human sequence. Despite the lack of statistically significant sequence constraint between the human enhancer and its lemur, mouse and dog orthologs, the latter three sequences exhibited enhancer activity both *in vitro* and *in vivo* (Fig. 2). The human regulatory element, however, consistently exhibited stronger enhancer activity in both assays, driving two-fold greater expression than lemur or dog PS2, and four-fold greater expression than mouse (Fig. 2A). This observation, coupled with the evidence of negative selection acting on the primate enhancer and the lack of significant sequence constraint between the anthropoid primate and mammalian PS2 sequences (conservation p-value, Table I), suggests that the stronger enhancer activity in human is a gain of function in the anthropoid primate lineage with a potentially important adaptive role in these species.

An anthropoid-primate specific sterol regulatory element (SRE) contributes to distinct human PS2 enhancer activity

To identify the molecular basis of the primate-specific activity of PS2, we computationally dissected the 860-bp human PS2 enhancer (see Methods) and found a sterol regulatory element (SRE), a binding site specifically recognized by the cholesterol sensing proteins SREBP (sterol regulatory element binding protein) which are known to play a key role

in the regulation of *LDLR* [19, 20]. Phylogenetic analysis of the orthologous PS2 sequences from 3 distant mammals (mouse, rat and dog), 3 prosimians (lemur, mouse lemur and galago), and 9 anthropoid primates covering all major lineages including hominoids, old-world and new-world monkeys revealed the presence of the SRE exclusively in anthropoid primates (Fig 3). This phylogenetic distribution of the SRE in mammals can be most parsimoniously explained by the appearance of the SRE in the ancestor of anthropoid primates after its divergence from prosimians (Fig 3).

The functional role of the binding motif identified by computational analysis was explored by site-specific mutagenesis. A 4-bp substitution was introduced into the SRE, which was expected to completely inactivate the site based on a previously reported mutagenesis study [21]. The 4-bp substitution in the SRE decreased human enhancer activity in the human cell culture assay and the *in vivo* mouse liver DNA transfer assay to a level comparable to that of the lemur, mouse and dog enhancers, species which lack a computationally predicted SRE (Fig. 2). The functionality of the SRE, found exclusively in anthropoid primates, suggests that this element is likely to contribute to the stronger activity found in these species. We also identified within the 860-bp enhancer a 21-bp subregion that shows strong conservation across mammalian species including lemur, mouse lemur, galago, mouse and dog and contains predicted binding sites for transcription factors AP-4 and AP-1. Deletion of the conserved 21-bp sequence from either human or dog PS2 resulted in a significant reduction in enhancer activity (data not shown), suggesting that the evolutionarily conserved AP-4 and AP-1 sites are important for the core enhancer activity shared among mammals. It is worth noting that such short blocks of genuinely constrained sequence are not easily distinguishable from the numerous "coincidentally conserved" sequence fragments that are likely to occur in large genomic regions as a consequence of stochastic variation in the incidence of neutral mutations. Incorporation of additional information, namely the binding specificities of transcription factors, was required to classify this 21-bp fragment as a functional candidate. Thus, conservation of this short subsequence in multiple mammals does not detract from the fact that the enhancer sequence is significantly conserved only in anthropoid primates, as described above.

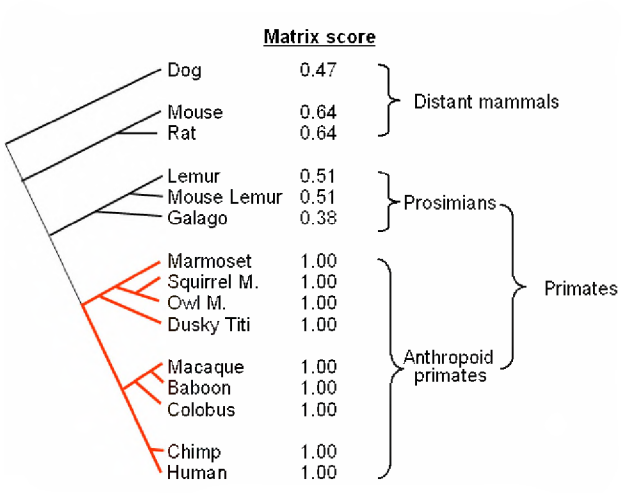

Fig. 3 Phylogenetic analysis of the sterol regulatory element (SRE). The human SRE motif and its orthologs were scored for transcription-factor binding affinity, with low motif scores

indicating low predicted affinity to SREBP (see Methods). Since the SRE is present in all the analyzed anthropoid primates (indicated by the red branches in the tree) and absent from the prosimians, rodents and dog, emergence in the lineage leading to anthropoid primates is the most parsimonious explanation.

Since SREBP-2, the major regulator of LDLR [19, 20] specifically binds to the SRE [11], we examined the responsiveness of the human, lemur, mouse and dog orthologous PS2 enhancers to this transcription factor. Co-transfection of the reporter gene driven by PS2 and the human *LDLR* promoter with a construct expressing the mature form of SREBP-2 indicated that the human enhancer was strongly activated by the exogenous SREBP-2, to a level 5-fold higher than that of the human *LDLR* promoter alone, which is known to be SREBP-responsive as well [22]. The lemur, mouse and dog enhancers were activated to a significantly lesser extent, consistent with their much lower SRE prediction score and with their lack of additional consensus SRE motifs within the PS2 element (Fig. 3, Fig. 4A and data not shown). To determine whether the observed differential SREBP-2 response among tested mammalian PS2 enhancers was directly mediated by the predicted SRE, we inactivated or restored the consensus SRE by site-specific mutagenesis at the orthologous positions of the human and dog PS2 element, respectively. Substituting 4 bases in the human SRE motif, which reduced the motif matrix score from 1 to 0.35 (see Methods), resulted in a reduction in SREBP-2 enhancer response to a level comparable to that of the lemur, mouse and dog enhancers. These results indicate that the anthropoid-specific SRE mediates the activation of the PS2 enhancer by SREBP-2 and contributes to the strong enhancer activity characterizing human and other anthropoid primates. Furthermore, substituting 3 bases in the dog SRE, so as to increase the SRE motif score from 0.47 to 1 (representing a perfect SRE), led to a significant increase in the dog enhancer response to SREBP-2, though only to half the level of the human PS2 enhancer (Fig. 4B). This suggests that the anthropoid primate-specific SRE is part of a combinatorial mechanism [23], including possible additional substitutions in the core enhancer element, that contribute to the stronger human PS2 enhancer activity.

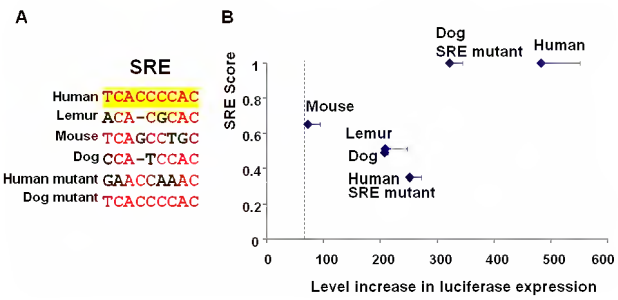

Fig. 4 Relation between SRE motif score and response of PS2 enhancer to SREBP-2. (A) Alignment of sequences from indicated species at position orthologous to human PS2 sterol regulatory element (SRE). "Human mutant" refers to the mutated human SRE with 4 point substitutions relative to the human wild-type. "Dog mutant" refers to the consensus SRE introduced into dog PS2 by means of 3 point substitutions. (B) Luciferase assay analysis of reporter construct and the SREBP-2 expression vector cotransfection into 293T cells. The LDLR PS2 element from each of the indicated species was tested in combination with the LDLR promoter. The Y-axis denotes SRE motif score (likeness to known SRE motifs, see Methods). Response of the PS2 element to SREBP-2 is shown as the increase in luciferase expression

level (arbitrary units) upon activation by 3 ng of SREBP-2 expression vector. Expression level increase for LDLR promoter alone is indicated by the dotted line. Error bars indicate standard deviation.

The role of SREBP-2 in regulating the human PS2 enhancer was further explored in its native chromosomal context in HepG2 cells, which actively express SREBP-2 and are a well-defined system for studying *LDLR* regulation [24-26]. Our analysis showed that the PS2 sequence is a DNaseI hypersensitive site in HepG2 cells (see Supplementary Information Fig. S2), suggesting that the corresponding DNA element is involved in transcriptional regulation of the endogenous gene. Using the ChIP (chromatin immunoprecipitation) assay, we were able to show that fractionation of chromatin with an anti-SREBP-2 antibody specifically enriched for endogenous PS2 and *LDLR* promoter DNA relative to control region (Fig. 5); the latter has previously been shown to be bound by SREBP-2 [27].Together, the DNAse I hypersensitivity and ChIP assays provide strong evidence that SREBP-2 binds in the vicinity of the human PS2 enhancer in its native genomic locus. Regulation of the enhancer by SREBP-2 also suggests that the PS2 element plays a role in the activation of its upstream gene *LDLR* rather than the downstream gene *Spbc24*, which encodes a component of the kinetochore Ndc80 protein complex [28]. It has recently been noted based on genome -wide analysis of gene expression that SREBP targets are largely restricted to lipid-metabolism genes, including *LDLR* [19]. No connection was found between SREBP and kinetochore structural genes such as *Spbc24*.

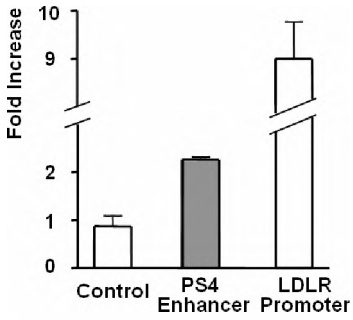

Fig. 5 Anti-SREBP-2 antibody specifically enriches the human PS2 DNA sequences in HepG2 cells in a ChIP assay. DNA precipitates were measured by real-time PCR using primers spanning the indicated regions. Control region (control) corresponds to the first coding exon of the neighboring gene SPBC24, which is approximately 7.5 kb away from the PS2 sequence. The results are presented as fold increase in the enrichment of precipitated DNA by anti-SREBP-2 antibody over non-specific IgG. Graphical representations of the mean ± SE from three independent experiments are shown.

Conclusion

We have shown phylogenetic and molecular data supporting the evolution of differential gene expression of *LDLR* in mammals. Transcriptional control of *LDLR* is mainly effected through the intracellular cholesterol sensor SREBP-2, which mediates the increased transcription of *LDLR* and other cholesterol synthesis genes in response to low cholesterol levels [22]. The differential response among mammals of the PS2 enhancer to SREBP-2 suggests that species-

specific regulation of *LDLR* is expected in conditions that result in decreased intracellular cholesterol levels, such as reduced availability of dietary cholesterol, and has implications for the study of *LDLR* response to cholesterol lowering drugs in animal models.

While the human *LDLR* coding sequence and promoter are well conserved in all sequenced mammals (Fig.1 and Fig. S1), our data support the modification of the expression characteristics of this gene through the primate-specific evolution of a distal regulatory element. We have shown the emergence and fixation of a sterol-responsive element (SRE) in the common ancestor of anthropoid primates, which modifies the expression driven by a pre-existing mammalian enhancer shared by all tested mammals. This demonstrates one mechanism by which mammalian regulatory elements can evolve to perform new functions. Given the vital importance of LDLR in energy storage, the appearance of a new cholesterol sensing element in the LDLR enhancer may have played a role in the evolution of new physiological features, as the ancestor of anthropoid primates adapted to different metabolic requirements and diets.

Materials and Methods

Plasmid constructs.
The human LDLR promoter was cloned in the proper orientation upstream of the luciferase cDNA in the pGL3Basic construct (Promega). The human PS1 element and the PS2 elements from the human, lemur, mouse, and dog LDLR loci were PCR cloned into polylinker sites in the (-) orientation upstream of the promoter. The cloned human PS1 element corresponds to the Gumby predicted conserved sequence and approximately 200bp of flanking sequence on either side (hg18, chr19:11067913-11068639). To clone the human PS2 element, the region containing human PS2 was PCR cloned into pGL3Basic (see Supplementary Information Table 1 for primer sequences), and digested with Spe I and Nhe I to only include the Gumby predicted conserved sequence and approximately 200bp of flanking sequence on either side (hg18, chr19:11110333-11111194). Site-specific point mutations and deletions were introduced into human and dog PS2 elements using QuikChangeII site-directed mutagenesis kit (Stratagene) according to manufacturer's protocol and were confirmed by sequencing. The expression construct for human mature form of SREBP-2 (pcDNA.2FLAG SREBP-2) was kindly provided by Dr. Timothy F. Osborne (UC Irvine).

Transient-transfection reporter assay.
293T cells (ATCC CRL-11268), a human embryonic kidney cell line, were grown at 37°C and 5% CO2 in Dulbecco's modified Eagle's medium (ATCC), supplemented with 10% FBS (Hyclone), L-glutamine, penicillin-streptomycin. Cells with a passage number below 15 were used. The cells were grown in 12 well plates ($4x10^4$ cells/well) and transfected using Fugene (Roche Molecular Biochemicals) following the manufacturer's protocol. Briefly, 100 ng of each assayed plasmid and 10 ng pCMVβ (BD Biosciences) were mixed with 1.5 μl Fugene and added to each well. Following 42-48 hours of incubation cells were harvested and lysed. Activity of luciferase and β-Galactosidase was measured using the Luciferase Assay System (Promega) and the Galacto-Light Plus (Applied Biosystems) respectively. Luciferase activity for each sample was normalized to the β-galactosidase assay control. For cotransfection experiments, 100 ng of the lulciferase reporter gene construct, 3 ng of SREBP-2 expression vector, and 10 ng of pCMVβ were used. Transfections were carried out in duplicates. All experiments are representative of at least three independent transfections.

Tail vein plasmid DNA transfer assays
Tail vein injection was performed as described by Herweijer and Wolff [18] following the TransIT® *In Vivo* Gene Delivery System Protocol (Mirus Corporation). Six to nine FVB male mice (Charles River Laboratory) at age 7-8 weeks were used for each reporter gene construct. Ten μg of each reporter construct, along with 2 μg of pCMVβ(BD Biosciences) to correct for delivery efficiency, were injected into each mouse. The entire content of the syringe was delivered in 3-5 seconds. Animals were sacrificed 24 hours later, livers extracted, measured to correct for size, homogenized, and centrifuged for 15 minutes at 4°C, 14,000 rpm. Activity of luciferase and β-Galactosidase was measured as described above. All p-values are from the two-sample Wilcoxon rank-sum (Mann-Whitney) test using STATA (STATA Corporation). All experimental results are representative of two independent plasmid DNA transfer assays.

Chromatin immunoprecipitation
HepG2 cells (ATCC HB-8065) were cultured in DMSF (Defined medium serum free) medium for 24 hours for induction of endogenous SREBPs[29]. Chromatin immunoprecipitaion assays were performed as described[30]. Crosslinked chromatin was immunoprecipitated with a specific SREBP-2 antibody (Santa Cruz sc-8151)[27] or IgG control.

Sources of sequence data
Draft sequences of baboon, colobus, dusky titi, marmoset, owl monkey, lemur and galago BACs were determined by sequencing ends of 3 Kb subclones to 8-10-fold coverage using BigDye terminators (Applied Biosystems) and assembling reads into contigs with the Phred-Phrap-Consed suite as described previously[31]. All BAC sequences were submitted to GenBank with the following species and accession numbers: baboon (*Papio hamadryas*), AC140974; colobus (*Colobus guereza*), AC150433; marmoset (*Callithrix jacchus*), AC145530; dusky titi (*Callicebus moloch*), AC144655; owl monkey (*Aotus hybrid*), AC171393; squirrel monkey (*Saimiri boliviensis boliviensis*), AC146467;lemur (*lemur catta*), AC118569; mouse lemur (*microcebus murinus*), AC175656; Galago (*otolemur garnetti*), AC175655). Human, chimpanzee, rhesus, mouse, rat and dog sequences were downloaded from the UCSC Genome Bioinformatics website (http://genome.ucsc.edu).

Analysis of sequence conservation
We aligned the human *LDLR* locus (chr19:11,055,219-11,117,169; NCBI Build 35) to its orthologs in baboon, colobus, dusky titi, marmoset, owl monkey, lemur, mouse and dog using MLAGAN[32]. Due to incomplete primate sequence availability, we included only ~6 kb of the 5' intergenic region in the analysis. Aligned sequences were scanned for statistically significant evolutionarily conserved regions using Gumby [12] [13] [14]. Gumby goes through the following three-step process to identify statistically significant conservation in the global alignment input. (i) Non-coding regions in the alignment are used to estimate the local neutral mismatch rates among all pairs of aligned sequences. The rates are used to derive a log-likelihood scoring scheme for slow versus neutral evolution, where the slow rate is set to some fraction (in this case, half) of the neutral rate. (ii) Each alignment position is then assigned a conservation score using a phylogenetically weighted sum-of-pairs scheme. (iii) Conserved regions of any length are identified as alignment blocks with a high cumulative conservation score and assigned p-values using Karlin-Altschul statistics [33]. We set a threshold p-value of 0.1 in a baseline human sequence length of 10 kb. Conserved regions identified by Gumby were visualized using RankVISTA. In addition, human-mouse sequence conservation was analysed using the VISTA web server [34], with the standard criteria of 70% sequence identity in window of size 100 bp.

Binding-site prediction
We scanned the aligned enhancer sequences for predicted transcription-factor binding sites using DiAlign TF (http://www.genomatix.de, Genomatix Software GmbH). Anthropoid primate (human, baboon, colobus, dusky titi, marmoset and owl monkey) sequences were assessed for the presence of sites conserved across all six species that were predicted to bind one of the following liver-expressed transcription factors: C/EBP, LXR, FXR, COUP-TF, PPAR, HNF1, HNF3, HNF4 and SREBP. Binding sites common to primates and mammals were predicted on the basis of conservation of any vertebrate transcription factor motif in at least 8 of the 10

analysed species (6 anthropoid primates, lemur, mouse, rat and dog). Motif scores of the SREs or SRE orthologs of individual species were calculated using rVISTA [35], and normalized so that the maximum achievable score is 1.0, and the expected score of a random nucleotide sequence with the local GC content is zero. The score distribution of functional SREs was calculated from the binding profile of SREBP (http://www.gene-regulation.com/pub/databases.html), assuming that nucleotide frequencies at each position in the motif are independent. We retrospectively augmented the species set with SRE orthologs from chimpanzee, rhesus, squirrel monkey, mouse lemur and galago, based on pairwise alignments of those species to human.

Acknowledgements: We thank J. Noonan, L. Pennacchio, A. Visel, S. Tringe, N. Ahituv, and other Rubin laboratory members for suggestions and criticisms on the manuscript, and Tim Osborne for providing the SREBP-2 expression vector. B. Kullgren and S. Phouanenavong provided technical assistance for tail vein plasmid DNA transfer assay.

Research was conducted at the E.O. Lawrence Berkeley National Laboratory and at the Joint Genome Institute. This work was supported by the Director, Office of Science, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 and NIH-NHLBI grant numbers THL007279F and U1HL66681B

References:

1. King MC, Wilson AC: **Evolution at two levels in humans and chimpanzees**. *Science* 1975, **188**:107-116.
2. Carroll SB: **Evolution at two levels: on genes and form**. *PLoS Biol* 2005, **3**:e245.
3. Hughes AL: **Natural selection and the diversification of vertebrate immune effectors**. *Immunol Rev* 2002, **190**:161-168.
4. Dorus S, Vallender EJ, Evans PD, Anderson JR, Gilbert SL, Mahowald M, Wyckoff GJ, Malcom CM, Lahn BT: **Accelerated evolution of nervous system genes in the origin of Homo sapiens**. *Cell* 2004, **119**:1027-1040.
5. Zhang J, Zhang YP, Rosenberg HF: **Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey**. *Nat Genet* 2002, **30**:411-415.
6. Jessen TH, Weber RE, Fermi G, Tame J, Braunitzer G: **Adaptation of bird hemoglobins to high altitudes: demonstration of molecular mechanism by protein engineering**. *Proc Natl Acad Sci U S A* 1991, **88**:6519-6522.
7. Yokoyama S: **Molecular evolution of color vision in vertebrates**. *Gene* 2002, **300**:69-78.

8. Rockman MV, Hahn MW, Soranzo N, Zimprich F, Goldstein DB, Wray GA: **Ancient and Recent Positive Selection Transformed Opioid cis-Regulation in Humans**. *PLoS Biol* 2005, **3**:e387.
9. Hahn MW, Rockman MV, Soranzo N, Goldstein DB, Wray GA: **Population genetic and phylogenetic evidence for positive selection on regulatory mutations at the factor VII locus in humans**. *Genetics* 2004, **167**:867-877.
10. Chiang JY: **Bile acid regulation of gene expression: roles of nuclear hormone receptors**. *Endocr Rev* 2002, **23**:443-463.
11. Horton JD, Goldstein JL, Brown MS: **SREBPs: activators of the complete program of cholesterol and fatty acid synthesis in the liver**. *J Clin Invest* 2002, **109**:1125-1131.
12. Prabhakar S, Poulin F, Shoukry M, Afzal V, Rubin EM, Couronne O, Pennacchio LA: **Close sequence comparisons are sufficient to identify human cis-regulatory elements**. *Genome Research* 2006, **in press**.
13. Ahituv N, Prabhakar S, Poulin F, Rubin EM, Couronne O: **Mapping cis-regulatory domains in the human genome using multi-species conservation of synteny**. *Hum Mol Genet* 2005, **14**:3057-3063.
14. Hughes JR, Cheng JF, Ventress N, Prabhakar S, Clark K, Anguita E, De Gobbi M, de Jong P, Rubin E, Higgs DR: **Annotation of cis-regulatory elements by identification, subclassification, and functional assessment of multispecies conserved sequences**. *Proc Natl Acad Sci U S A* 2005, **102**:9830-9835.
15. Boffelli D, McAuliffe J, Ovcharenko D, Lewis KD, Ovcharenko I, Pachter L, Rubin EM: **Phylogenetic shadowing of primate sequences to find functional regions of the human genome**. *Science* 2003, **299**:1391-1394.
16. Boffelli D, Nobrega MA, Rubin EM: **Comparative genomics at the vertebrate extremes**. *Nat Rev Genet* 2004, **5**:456-465.
17. Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S, Weinstock GM, Wilson RK, Gibbs RA, Kent WJ, Miller W, Haussler D: **Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes**. *Genome Res* 2005, **15**:1034-1050.
18. Herweijer H, Wolff JA: **Progress and prospects: naked DNA gene transfer and therapy**. *Gene Ther* 2003, **10**:453-458.
19. Horton JD, Shah NA, Warrington JA, Anderson NN, Park SW, Brown MS, Goldstein JL: **Combined analysis of oligonucleotide microarray data from transgenic and knockout mice identifies direct SREBP target genes**. *Proc Natl Acad Sci U S A* 2003, **100**:12027-12032.
20. Horton JD, Shimomura I, Brown MS, Hammer RE, Goldstein JL, Shimano H: **Activation of cholesterol synthesis in preference to fatty acid synthesis in liver and adipose tissue of transgenic mice overproducing sterol regulatory element-binding protein-2**. *J Clin Invest* 1998, **101**:2331-2339.
21. Smith JR, Osborne TF, Goldstein JL, Brown MS: **Identification of nucleotides responsible for enhancer activity of sterol regulatory element in low density lipoprotein receptor gene**. *J Biol Chem* 1990, **265**:2306-2310.
22. Hussain MM, Strickland DK, Bakillah A: **The mammalian low-density lipoprotein receptor family**. *Annu Rev Nutr* 1999, **19**:141-172.

23. Sanchez HB, Yieh L, Osborne TF: **Cooperation by sterol regulatory element-binding protein and Sp1 in sterol regulation of low density lipoprotein receptor gene**. *J Biol Chem* 1995, **270**:1161-1169.
24. Kong W, Wei J, Abidi P, Lin M, Inaba S, Li C, Wang Y, Wang Z, Si S, Pan H, Wang S, Wu J, Li Z, Liu J, Jiang JD: **Berberine is a novel cholesterol-lowering drug working through a unique mechanism distinct from statins**. *Nat Med* 2004, **10**:1344-1351.
25. Maxwell KN, Fisher EA, Breslow JL: **Overexpression of PCSK9 accelerates the degradation of the LDLR in a post-endoplasmic reticulum compartment**. *Proc Natl Acad Sci U S A* 2005, **102**:2069-2074.
26. Izem L, Rassart E, Kamate L, Falstrault L, Rhainds D, Brissette L: **Effect of reduced low-density lipoprotein receptor level on HepG2 cell cholesterol metabolism**. *Biochem J* 1998, **329 (Pt 1)**:81-89.
27. Zeng L, Lu M, Mori K, Luo S, Lee AS, Zhu Y, Shyy JY: **ATF6 modulates SREBP2-mediated lipogenesis**. *Embo J* 2004, **23**:950-958.
28. Wei RR, Sorger PK, Harrison SC: **Molecular organization of the Ndc80 complex, an essential kinetochore component**. *Proc Natl Acad Sci U S A* 2005, **102**:5363-5367.
29. Toth JI, Datta S, Athanikar JN, Freedman LP, Osborne TF: **Selective coactivator interactions in gene activation by SREBP-1a and -1c**. *Mol Cell Biol* 2004, **24**:8288-8300.
30. Wang Q, Carroll JS, Brown M: **Spatial and temporal recruitment of androgen receptor and its coactivators involves chromosomal looping and polymerase tracking**. *Mol Cell* 2005, **19**:631-642.
31. Wang QF, Liu X, O'Connell J, Peng Z, Krauss RM, Rainwater DL, VandeBerg JL, Rubin EM, Cheng JF, Pennacchio LA: **Haplotypes in the APOA1-C3-A4-A5 gene cluster affect plasma lipids in both humans and baboons**. *Hum Mol Genet* 2004, **13**:1049-1056.
32. Brudno M, Do CB, Cooper GM, Kim MF, Davydov E, Green ED, Sidow A, Batzoglou S: **LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA**. *Genome Res* 2003, **13**:721-731.
33. Karlin S, Altschul SF: **Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes**. *Proc Natl Acad Sci U S A* 1990, **87**:2264-2268.
34. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I: **VISTA: computational tools for comparative genomics**. *Nucleic Acids Res* 2004, **32**:W273-279.
35. Loots GG, Ovcharenko I, Pachter L, Dubchak I, Rubin EM: **rVista for comparative sequence-based discovery of functional transcription factor binding sites**. *Genome Res* 2002, **12**:832-839.

Supplemental Data:

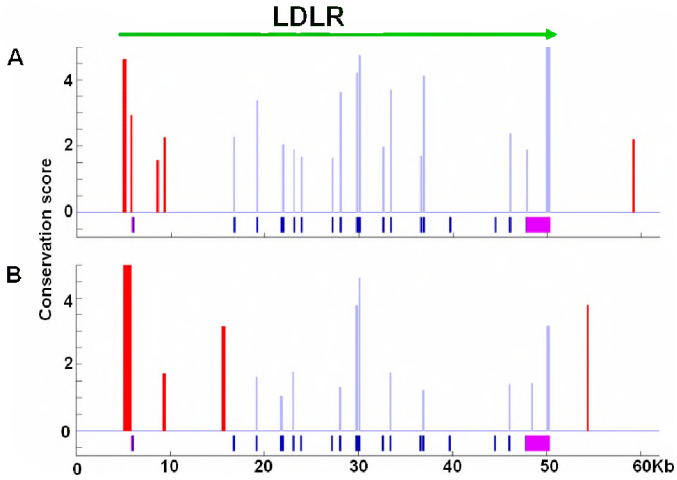

Fig. S1 Conservation profiles of the *LDLR* locus using human-dog (A) and human-lemur (B) sequence comparisons. Conservation profiles were obtained using RankVISTA and displayed with the human sequence as reference as described in Fig. 1.

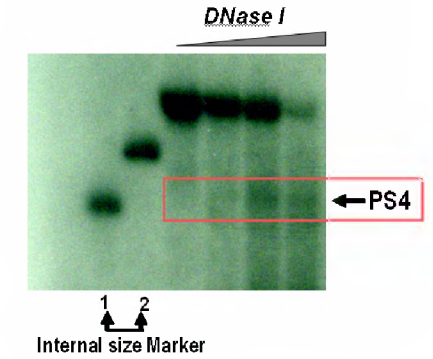

Fig. S2 DNaseI hypersensitive site mapping around LDLR PS2 region in human liver cell line HepG2. Vertical arrows indicate lanes with internal size markers that were generated by enzyme digestion of the LDLR_PS2 sequence (lane 1) and a position approximately 90bp downstream of the element (lane 2). Horizontal arrow points to the hypersensitive site (HS). Co-migration of the internal size marker 1 with the HS localizes the HS to LDLR PS2 sequence.

A total of two tables listing all relevant primer sequences used.

Table 1: Primers used in the cloning of human *LDLR* promoter and PS2 elements from indicated species.

Element ID	Primers
Human LDLR promoter	F: ATGCGTTTCCAATTTTGAGG R: TCTAGCAGGGGGAGGAGTTT
Human PS1	F: AGCCTCAGTCATGCCACTG R: GGCCTAGGCAACATACCAAG
Human PS2	F: GGAGGCCACTGTGTCAGTTT R: ACTCCAGCCTGGGAAACTCT
Lemur PS2	F: ATCCCCAGCTGGTATCCTCT R: ACAGCTGGTTTCACAGCATT
Mouse PS2	F: GCAGCAGCTGATTTCTGACA R: AGGCATGCTTGTGAGAGGTA
Dog PS2	F: TTGGTACCCCCAACTCTGTC R: CAGAGCCAGGATTTGACCTG
HumanPS2 21-bp deletion:	CATCCCAGACCACCTGGAGCCTCTG
DogPS2 21-bp deletion:	GGCGGCCCAAACGCTGGAGCCTGC
HumanPS2 SRE mutant:	GTCCCTGGTCCCCAGAACCAAACTGAGCATGGCCGC
DogPS2 SRE mutant:	GCCCTGGGCACATCACCCCACCCCGAGCGGGGC

Table 2: primers used in ChIP assay

Region	Primers
Human PS2	F: TCTGAGTGGGAGTCCCTGGT R: GAAAGTTGCCAGGAAACCCC
Human LDLR promoter	F: AATGACGTGGGCCCCG R: ACCTGCTGTGTCCTAGCTGGA
Control	F: TGATGAATTTCCTGGAGAGCTG R: TCCATCATGGCCCCAGTG