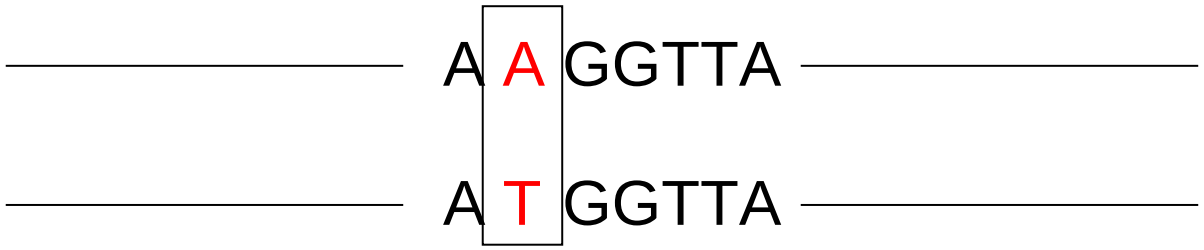
Single Nucleotide Polymorphisms in the Genome

Correlations of Divergence and Diversity and Complicated Things and other Correlations between Strange Matters and Unexpected Turns without Any Solutions and My Attempts to Say Something Really Important and Impressive Although I Dont Like to Speak.

Claudia Fried

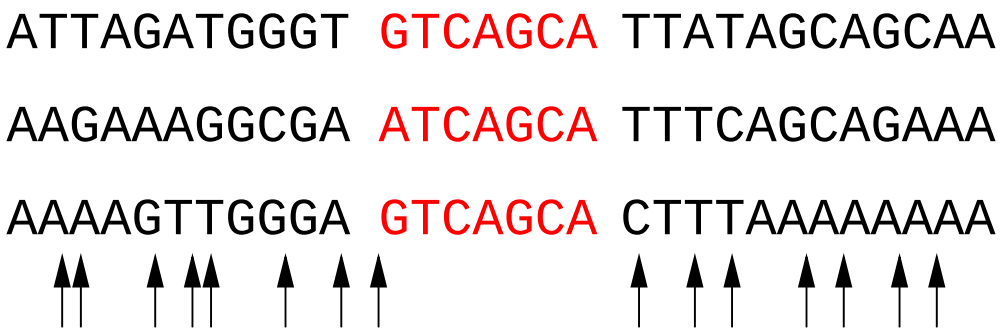
Bled, 21.02.2005

Single Nucleotide Polymorphisms (SNPs)



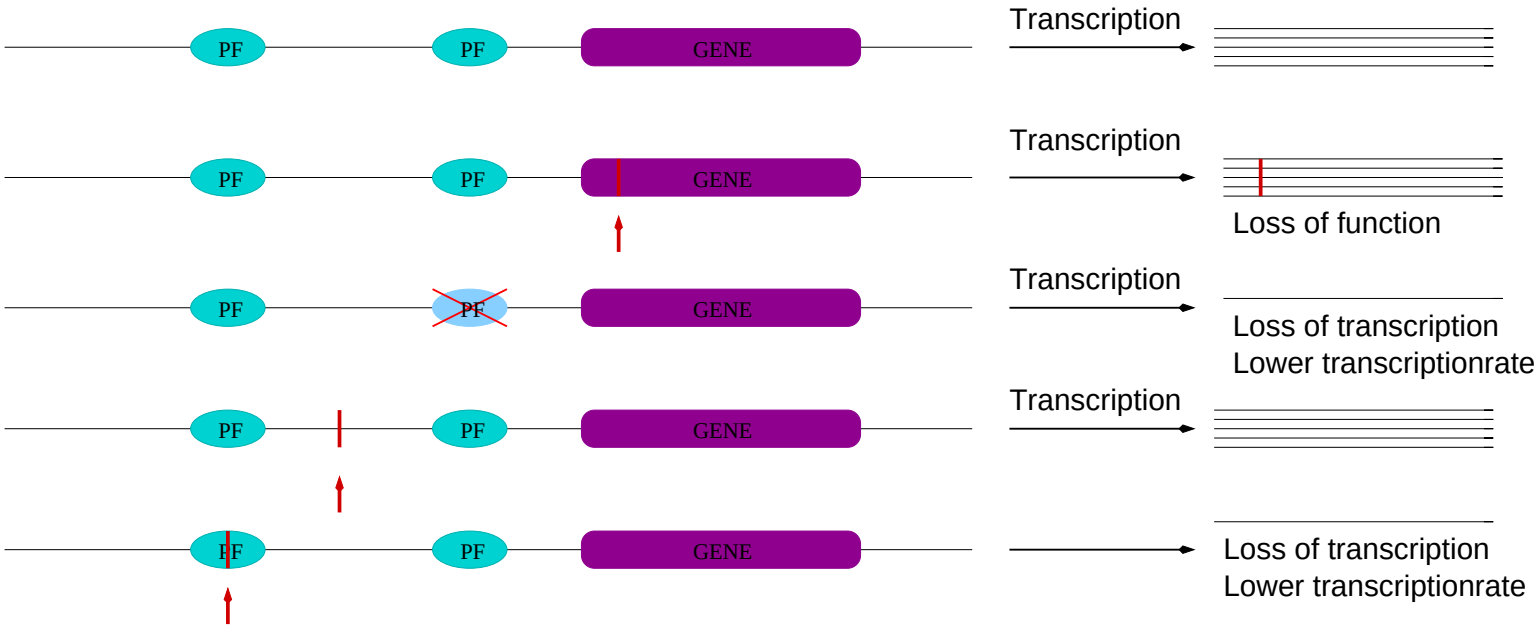
* small genetic variation between different allels
* involving 1 basepair
* occur in at least 1% of the population
* make up about 90% of all human genetic variations
* occur every 100 to 300 bases in the human genome
* ability to alter or destroy function of a gene
* predispose people to diseases (e.g. cancer, diabetes, vascular diseases, rheumatoid arthritis, Alzheimer)
* influence response to drugs, bacteria, viruses, toxins...

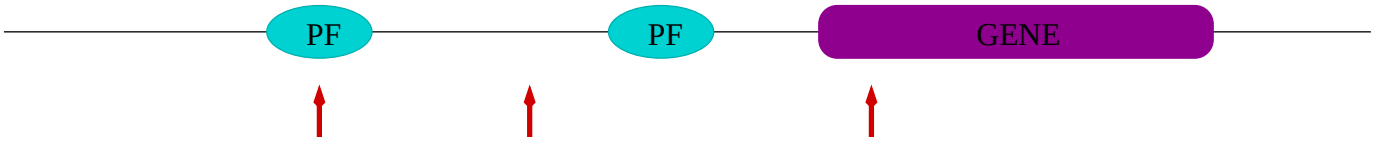
Phylogenetic Footprints



* highly conserved due to high selective pressure
* experimentally verified to be functional
* detected by comparative analysis of homologos sequences
* single functional element (Transcription factor binding site, enhancer elements) (6-20bp)
* Clusters of functional elements (100-300bp)
* ncRNAs

Influence of SNPs





Differs the distribution of SNPs between coding, non-coding but regulatory functional, and non-functional sequences?

Sequence Data

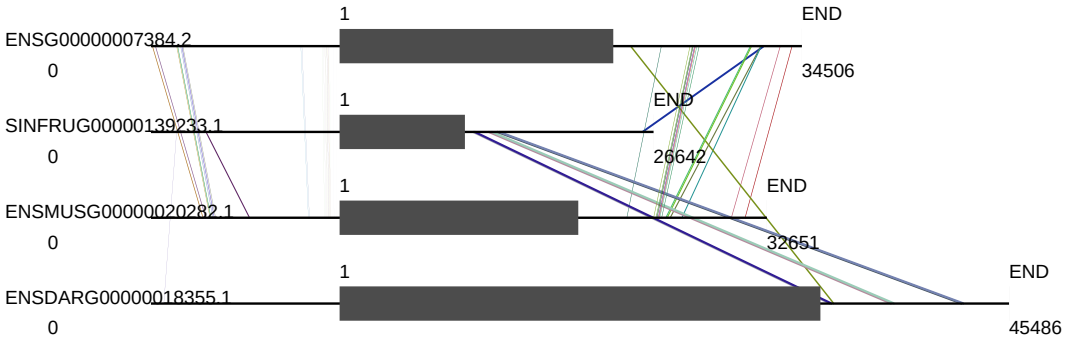
8213 *human (Homo sapiens)* genes with homologous genes in:

* *mouse (Mus musculus)*
* *pufferfish (Fugu rubripes)*
* *zebrafish (Danio rerio)*

The data was retrieved with ENSMART from EBI database and dbSNP

Detection of Phylogenetic Footprints

Detection of phylogenetic footprints in the human intergenic sequences using `tracker`.

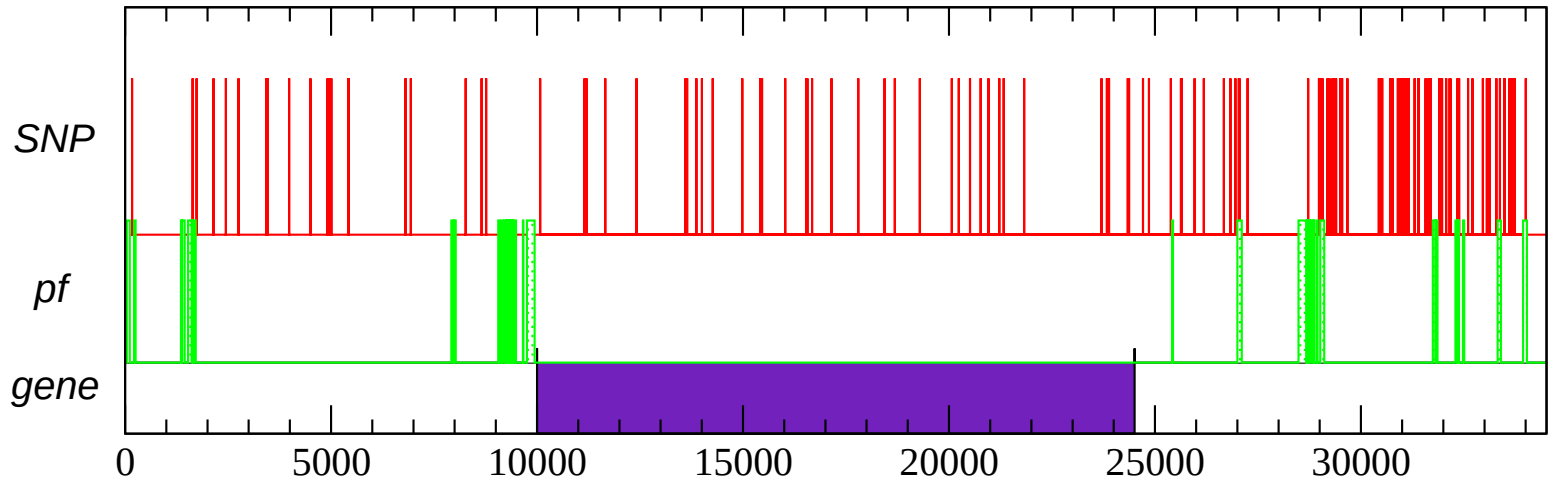


ENSG00000007384.2 41 114
ENSG00000007384.2 219 257
ENSG00000007384.2 1355 1371
ENSG00000007384.2 1384 1447
ENSG00000007384.2 1530 1619
ENSG00000007384.2 1623 1650
ENSG00000007384.2 1666 1714
.
.

Phylogenetic Footprints of ENSG00000007384.2 (rhomboid family 1; epidermal growth factor receptor, related sequence)

Detection of SNP Frequency

Calculation of the SNP frequency in phylogenetic footprints, genes and nonconserved noncoding region



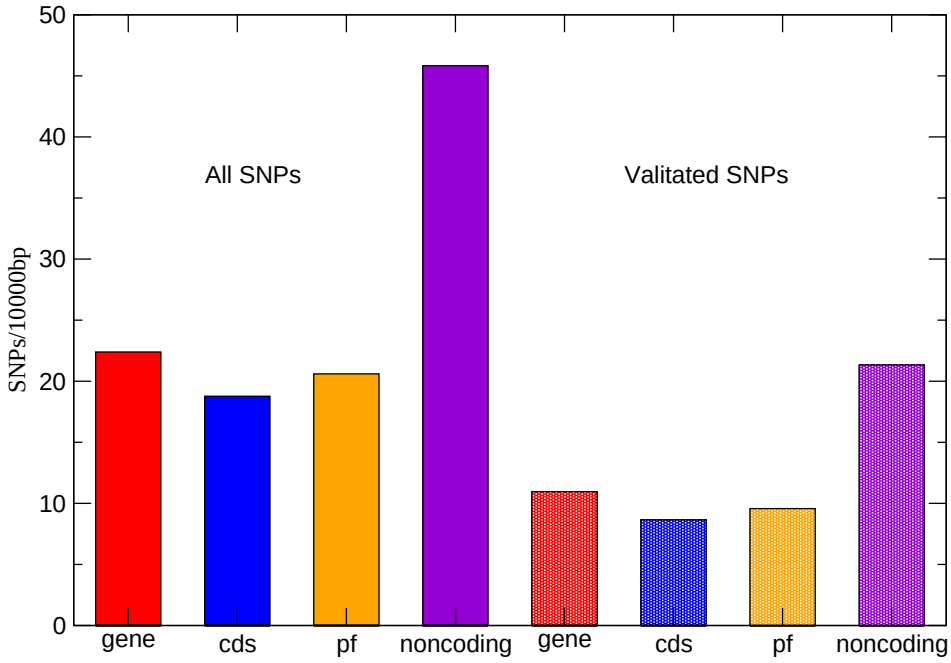
SNPs, Phylogenetic Footprints, and Coding Regions of ENSG00000007384.2

SNPs are Underrepresented in Coding Regions and Phylogenetic Footprints

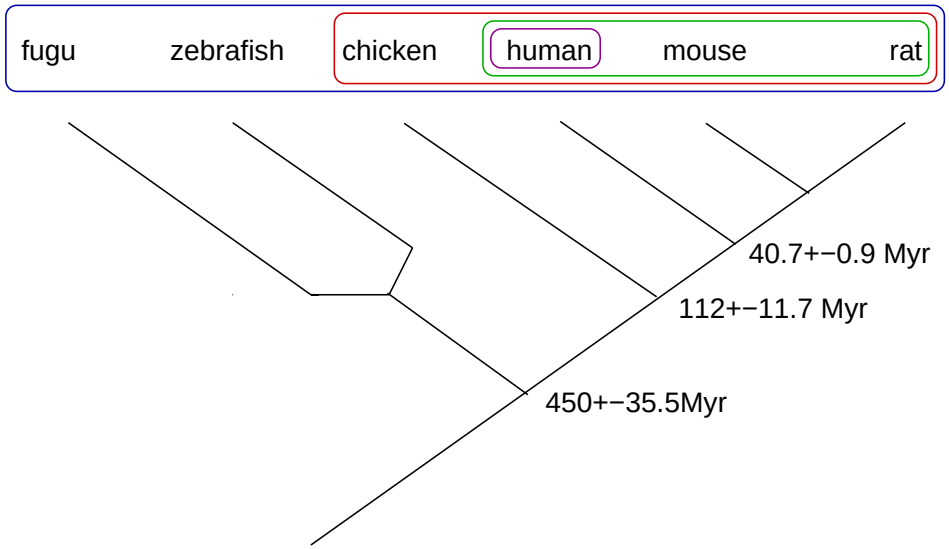
6581 genes	(all SNPs)	(nVal. SNPs)	(Val. SNPs)	SeattleDB
ρ_{Sgene}	22.39	11.43	10.96	40.69
ρ_{Scds}	18.77	10.12	8.649	2.64
FT(SNP $\uparrow$)	1	1	1	0.99
FT(SNP $\downarrow$)	0 $\downarrow$	0 $\downarrow$	0 $\downarrow$	1.17e-11 $\downarrow$
ρ_{Spf}	20.60	11.03	9.57	38.43
FT(SNP $\uparrow$)	1	1	1	0.98
FT(SNP $\downarrow$)	0 $\downarrow$	0 $\downarrow$	0 $\downarrow$	0.031 $\downarrow$
ρ_{Snc}	45.83	24.50	21.33	50.95

SNPs are significantly underrepresented in both coding regions and in conserved noncoding regions in the vicinity of a gene.

SNP Frequency



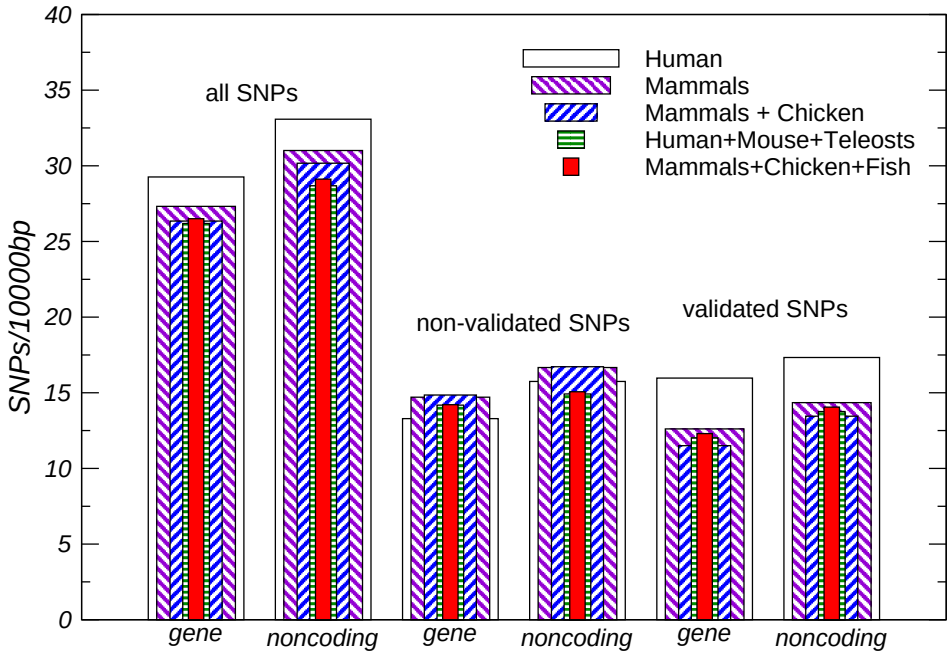
Correlation of Diversity and Divergence



four datasets:

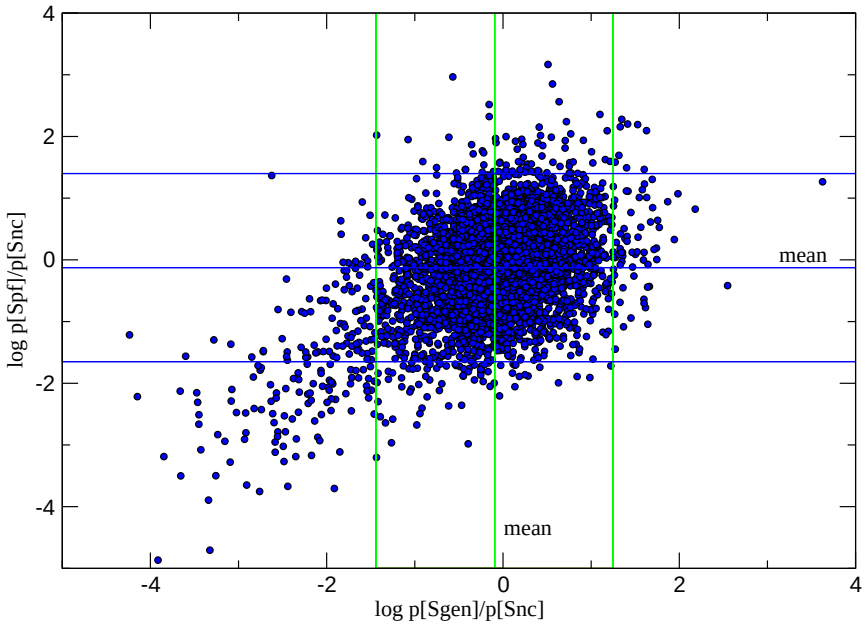
* human genes that have no homologs in either mouse, rat, chicken, fugu, or zebrafish
* human genes with homologs in mouse and rat but not in chicken, fugu, or zebrafish
* human genes with homologs in mouse, rat, and chicken, but not in the teleosts
* human genes with homologs in mouse, rat, chicken, zebrafish, and fugu

SNPs are Underrepresented in Old Genes



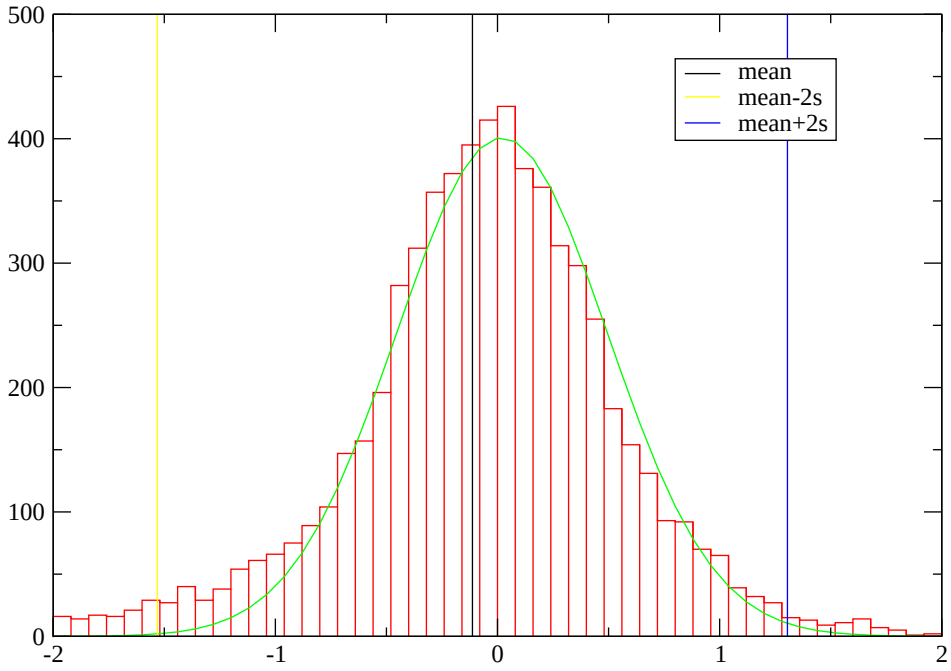
Human genes that have no known homologs in non-hominid species have an increased density of SNPs compared to evolutionarily older genes.

Correlation of the SNP Frequency in the Gene and Phylogenetic Footprints



Genes with low SNP density have a high selective pressure on their regulatory elements

Frequency Distribution



Analysis of genes with a low SNP density (mean-2s)

Analysis of Gene Gntology Groups

Assignment of the genes to their Gene Ontology (GO) term using GeneMapper

	ENSG00000175879			
Taxonomie				
Biological Process	(GO:0006350) transcription	(GO:0006350) segment specification	(GO:0007379) anterior/posterior axis specification	GO:0009948 embryonic ax specificatior
Cellular Component	–			
Molecular Function	(GO:0003700) transcription factor activity			

Functional profiling to search for GO terms that are significantly enriched in the genes with low SNP density

High Selective Sressure on some GO Terms

	Biol. Process	Mol. Function	Cell. Component
$\uparrow$	14 terms	23 terms	9 terms
$\downarrow$	5 terms	3 terms	3 terms

several GO terms show an overrepresentation in the genes with low SNP density whereas only few show an underrepresentation

Assignment to Function

SNPs in Footprints or Genes underrepresented	
Cellular Component underrepresentation	
GO:0000786	nucleosome
GO:0016591	DNA-directed RNA polymerase II, holoenzyme
GO:0030017	sarcomere
Cellular Component overrepresentation	
GO:0000242	pericentriolar material
GO:0005581	collagen
GO:0005624	membrane fraction
GO:0005740	mitochondrial membrane
GO:0005774	vacuolar membrane
GO:0005829	cytosol
GO:0005844	polysome
GO:0008076	voltage-gated potassium channel complex
GO:0019866	inner membrane

Overrepresentation of genes associated with membrane structures, transport and signaling and receptor activity

Conclusion

* SNPs are Underrepresented in Old Genes

* SNPs are Underrepresented in Coding Regions
  and Phylogenetic Footprints

* Enrichment of some GO terms in Genes with low SNP frequency
  transcription factors, transport proteins, membrane proteins
  receptor activity

* detect SNPs in regulatory regions that are suceptible for
  rheumatoid arthritis