§ 74 features used to build our prediction model §

Features about evolution (2)

Number of human paralogues, average dS within human paralogues.

Features about codon usage (10)

Codon usage: CTA (L)	Codon usage: ATT (I)	Codon usage: TAT (Y)
Codon usage: GCG (A)	Codon usage: CAC (H)	Codon usage: TGC (C)
Codon usage: CGT (R)	Codon usage: CGA (R)	Codon usage: CGG (R)
Codon usage: AGA (R)		

Features about genetic composition (40)

DNA AC content	Dinucleotide CpT composition	DNA 4-mer CGCG composition
DNA 4-mer AATC composition	DNA 4-mer TCGT composition	DNA 4-mer GATG composition
DNA 4-mer AACA composition	DNA 4-mer TGAG composition	DNA 4-mer GACC composition
DNA 4-mer ATAT composition	DNA 4-mer TGTA composition	DNA 4-mer GACG composition
DNA 4-mer ATGT composition	DNA 4-mer CACG composition	DNA 4-mer GAGT composition
DNA 4-mer ACAC composition	DNA 4-mer CTCC composition	DNA 4-mer GTAC composition
DNA 4-mer ACTA composition	DNA 4-mer CCAC composition	DNA 4-mer GTGT composition
DNA 4-mer ACTC composition	DNA 4-mer CCTA composition	DNA 4-mer GTGC composition
DNA 4-mer ACCG composition	DNA 4-mer CCTC composition	DNA 4-mer GTGG composition
DNA 4-mer TATG composition	DNA 4-mer CCGT composition	DNA 4-mer GCAA composition
DNA 4-mer TTCT composition	DNA 4-mer CGAG composition	DNA 4-mer GCTC composition
DNA 4-mer TTCG composition	DNA 4-mer CGTG composition	DNA 4-mer GCCT composition
DNA 4-mer TTGA composition	DNA 4-mer CGCA composition	DNA 4-mer GGGG composition
DNA 4-mer TCAT composition		

Features about proteomic composition (9)

Arginine composition, cysteine composition, methionine composition;

Basic amino acid composition (R/H/K)	Sulfur amino acid composition (C&M)
Hydroxyl amino acid composition (S&T)	Small amino acid composition (N/D/C/P/T)

Large amino acid composition (R/I/L/K/M)

Uncharged amino acid composition (A/N/C/Q/G/I/L/M/F/P/S/T/W/Y/V)

Features about human interactome network (3)

Shortest paths, betweenness, neighborhood connectivity.

Features about motifs (8)

SLim_DNA ATA[AG][TG]	SLim_DNA TAT[AT]T	SLim_DNA T[AT]AAA
SLim_DNA [ATG]TGTA	SLim_AA S[A-Z]N[A-Z]E	SLim_AA ENE
SLim_AA SVI	Co-occurence of SLim_AAs	

Abbreviations: dS, synonymous substitutions per synonymous site; SLim_DNAs, short linear nucleotide motifs; SLim_AAs, short linear amino acid motifs.