

Supplementary Information 1 – Supplementary Tables

Supplemental Table S1

>A1_Chimpanzee

GGGGGCGGGGAGGAGGGGAGAGCACAGGCTTTGACCGATAGTAACCTCTGCGCTCGGTGCA
GCCGAATCTATAAAAGGAACTAGTCTCGGCAAAAACCCGTAATTGCGAGCGAGAGTGAG
TGGGGCCGGGACCCGAGAGCCGAGCCGACCCCTTCTCTCCCGGGCTGCGGCAGGGCAGGG
CGGGGAGCTCCGCGCACCAACAGAGCCGGTTCTCAGGACGCTTTGCTCCTTGTTTTTCC
CCGGTTCTGTTTTCTCCCTTCTCCGGAAGCCTTGTCAGGGGTAGGAGAAAGAGACGCA
AACACAAAAGTGAAAAACAGTTAATGACCAGCCACGGCGTCCCTGCTGTGAGCTCTGGCC
GCTGCCTTCCAGGGCTCCCAAGCCACACGCTGTGGGTGCTGGCTGAGGGAAC

>A1_Bonobo

GGAGGAGGGGAGAGCACAGGCTTTGACCGATAGTAACCTCTGCGCTCGGTGCAGCCGAATC
TATAAAAGGAACTAGTCTCGGCAAAAACCCGTAATTGCGAGCGAGAGTGAGTGGGGCGG
GGACCCGCAGAGCCGAGCCGACCCCTTCTCTCCCGGGCTGCGGCAGGGCAGGGCGGGGAGC
TCCGCGCACCAACAGAGCCGGTTCTCAGGACGCTTTGCTCCTTGTTTTTCCCGGTTCT
GTTTTCTCCCTTCTCCGGAAGGCTTGTCAGGGGTAGGAGAAAGAGACGCAAAACAAAA
AGTGGAACACAGTTAATGACCAGCCACGGCGTCCCTGCTGTGAGCTCTGGCCGCTGCCTT
CCAGGGCTCCCAAGCCACACGCTGTGGGTGCTGGCTGAGGGAAC

>A1_Human

AGAGCACAGGCTTTGACCGATAGTAACCTCTGCGCTCGGTGCAGCCGAATCTATAAAAGG
AACTAGTCCCGCAAAAACCCGTAATTGCGAGCGAGAGTGAGTGGGGCCGGGACCCGCA
GAGCCGAGCCGACCCCTTCTCTCCCGGGCTGCGGCAGGGCAGGGCGGGGAGCTCCGCGCAC
CAACAGAGCCGGTTCTCAGGGCGCTTTGCTCCTTGTTTTTCCCGGTTCTGTTTTCTCC
CCTTCTCCGGAAGGCTTGTCAGGGGTAGGAGAAAGAGACGCAAAACAAAAAGTGAAAAA
CAGTTAATGACCAGCCACGGCGTCCCTGCTGTGAGCTCTGGCCGCTGCCTTCCAGGGCTC
CCGAGCCACACGCTGGGGGTGCTGGCTGAGGGAAC

>A1_Gorilla

GGAGGAGGGGAGAGCACAGGCTTTGACCGATAGTAACCTCTGCGCTCGGTGCAGCCGAATC
TATAAAAGGAACTAGTCTCGGCAAAAACCCGTAATTGCGAGCGAGATTGAGTGGGGCGG
GGACCCGCAGAGCCGAGCCGACCCCTTCTCTCCCGGGCTGCGGCAGGGCAGGGCGGGGAGC
TCCGGGCACCAACAGGGCCGGTTCTCAGGGCGCTTTGCTCCTTGTTTTTCCCCGTTCT
ATTTTCTCCCTTCTCCGGAAGGCTTGTCAGGGGTAGGAGAAAGAGACGCAAAACAAAA
AGTGGAACACAGTTAATGACCAGCCACGGCGTCCCTGCTGTGAGCTCTGGCCGCTGCCTT
CCAGGGCTCCCGAGCCACACGCTGTGGGTGCTGGCTGAGGGAAC

>A1_Golden_s_n_monkey

GGAGGAAGGCGAGAACAAGCTTTGACCGATAGTAACCTCTGCGCTCCGGGCAGCCGAATC
TATAAAAGGAACTAGTCTCGGCAAAAACCCATAATTGCGAGCGAGAGTGAGTGGGGCGG
GGACCCGCAGAGCCGAGCCGACCCCTTCTCTCCCGGGCTGCGGCAGGGCAGGGCAGAGCAG
GGCGGGGGGCTCCGCGCACCAACAGGACCAAGTTCTCAGGGCGCTTTGCTCCTTGTTTTT
CCCCTGTTCTGTTTTTCCCCCTTCTCCGGAAGGCTTGTCAGGGGTAGGAGAAAGAGAC
GCAAAACAAAAAGTGAAAAACAGTTAATGACCAGCCACGGCGTCCCTGCTGTGAGCTCTG
GCCGCTGCCTTCCAGGGCTCCCGAGCCACACACTGTGGGTGCTGGCTGAGGGAAC

>A1_Black_s_n_monkey

GACTGAACTACATAAACAGAGGCCGGGAAGGGGGCGGGAGGAAGGCGAGAACAAGCTTT
GACCGATAGTAACCTCTGCGCTCCGGGCAGCCGAATCTATAAAAGGAACTAGTCTCGGCA
AAAACCCATAATTGCGAGCGAGAGTGAGTGGGGCCGGGACCCGCAGAGCCGAGCCGAGCCGACC

CTTCTCTCCCGGGCTGCGGCAGGGCAGGGCAGAGCAGGGCGGGGGGCTCCGCGCACCAAC
AGGACCAGTTCTCAGGGCGCTTTGCTCCTTGTTTTTCCCCTCGTTCTGTTTTCCCCCT
TCTCCGGAAGGCTTGTC AAGGGGTAGGAGAAAGAGACGCAAAACACAAAAGTGAAAAACAG
TTAATGACCAGCCACGGCGTCCCTGCTGTGAGCTCTGGCCGCTGCCTTCCAGGGCTCCCG
AGCCACACACTGTGGGTGCTGGCTGAGGGAAC

>A1_Angola_colobus

GGAGGAAGGCGAGAACAGGCTTTGACCGATAGTAACCTCTGCGCTCCGGGCAGCCGAATC
TATAAAAGGAACTAGTCTAGGCAAAAACCTGTAATTGCGAGCGAGAGTGAGTGGGGCGG
GGACCCGCGAGAGCCGAGCCGACCCTTCTCTCTGGGCTGCGGCAGGGCAGGGCAGAGCAG
GGCGGGGGGCTCTGCGCACCAACAGGACCAGTTCTCAGGGCGCTTTGCTCCTGTTTTTT
CCCCCTCGTTCTGTTTTCTCCCCTTCTCCGGAAGGCTTGTC AAGGGGTAGGAGAAAGAGAC
GCAAAACACAAAAGTGAAAAACAGTTAATGACCAGCCACGGCGTCCCTGCTGTGAGCTCTG
GCCGCTGCCTTCCAGGGCTCCCGAGCCACACGCTGTGGGTGCTGGCTGAGGGAAC

>A1_Crab_e_macaque

GGGGGATTCAAGTATGGAGCTCGCTCTCCCCAATCCCTCCCTCCGGCTGAGGAACTAAC
AAAGGGAAAAAAGATTGCGGAAAGCACGATTTAGGGGAAGCAAATTCACCTGGTGCCCTT
GGCTGCCGGGAACATGGACTAGAGAGTCTGCGGCGCAGCCCGAGCCAGCGCTTCCCGC
GGGTCTTAGGCCGGCGGGCCCCGGCGGGGAGCGGGGGCGCAGACCGCGGACCCTGAGACA
CCTGATACACCTCCACCCCCACCCACCTCCTCCCAACTCCCTAGCAGCTTCGCG
GGCGGCTGAACGTCGCCCGTTTAAGAGGCGGGCCGCGCTCCACGTGCTTTCTGCTGAGT
GACTGAACTACATAAACAGAGCGCGGAAGGGGGCGGGGAGGAAGGCTAGAACAGGCTTT
GACCGATAGTAACCTCTGCGCTCCGGGCAGCCGAATCTATAAAAGGAACTAGTCTCGGCA
AAAACCCCGTAATTGCGAGCGAGAGTGAGTGGGGCCGGGACCAGCAGAGCCGAGCCGACC
CTTCTCTCCCGGTTGCGGCAGGGCAGAGCAGGGCGGGGGCTCTGCGCACCAACAGGAC
TAGTTCTCAGGGCGCTTTGCTCCTTGTTTTTCCCCTCGTTCTGTTTTCTCCCCTTCTCC
GGAAGGCTTGTC AAGGGGTAGGAGAAAGAGACGCAAAACACAAAAGTGAAAAACAGTTAAT
GACCAGCCACGGCATCCCTGCTGTGAGCTCTGGCCGCTGCCTTCCAGGGCTCCTGAGCCA
CACGCTGTGGGTGCTGGCTGAGGGAAC

>A1_Pig_t_macaque

CAGCTTCGCGGGCGGCTGAACGTCGCCCGTTTAAGAGGCGGGCCGCGCTCCACGTGCTT
TCTGCTGAGTGACTGAACTACATAAACAGAGGCCGGAAGGGGGCGGGGAGGAAGGCGAG
AACAGGCTTTGACCGATAGTAACCTCTGCGCTCCGGGCAGCCGAATCTATAAAAGGAACT
AGTCTCGGCAAAAACCCGTAATTGTGAGCGAGAGTGAGTGGGGCCGGGACCCGCAGAGC
CGAGCCGACCCTTCTCTCCCGGTTGCGGCAGGGCAGAGCAGGGCGGGGGGCTCTGCGCA
CCAACAGGACTAGTTCTCAGGGCGCTTTGCTCCTTGTTTTTCCCCTCGTTCTGTTTTCT
CCCCTTCTCCGGAAGGCTTGTC AAGGGGTAGGAGAAAGAGACGCAAAACACAAAAGTGGA
AACAGTTAATGACCAGCCACGGCATCCCTGCTGTGAGCTCTGGCCGCTGCCTTCCAGGGC
TCCTGAGCCACACGCTGTGGGTGCTGGCTGAGGGAAC

>A1_Macaque

GACTGAACTACATAAACAGAGGCCGGAAGGGGGCGGGGAGGAAGGCGAGAACAGGCTTT
GACCGATAGTAACCTCTGCGCTCCGGGCAGCCGAATCTATAAAAGGAACTAGTCTCGGCA
AAAACCCCGTAATTGCGAGCGAGAGTGAGTGGGGCCGGGACCAGCAGAGCCGAGCCGACC
CTTCTCTCCCGGTTGCGGCAGGGCAGGGCAGAGCAGGGCGGGGGGCTCTGCGCACCAAC
AGGACTAGTTCTCAGGGCGCTTTGCTCCTTGTTTTTCCCCTCGTTCTGTTTTCTCCCCT
TCTCCGGAAGGCTTGTC AAGGGGTAGGAGAAAGAGACGCAAAACACAAAAGTGAAAAACAG
TTAATGACCAGCCACGGCATCCCTGCTGTGAGCTCTGGCCGCTGCCTTCCAGGGCTCCTG
AGCCACACGCTGTGGGTGCTGGCTGAGGGAAC

>A1_Drill

GAGGAAGGCGAGAACAGGCTTTGACCGATAGTAACCTCTGCGCTCCGGGCAGCCGAATCT
ATAAAAGGAACTAGTCTCGGCAAAAACCCCGTAATTGCGAGCGACAGTGAGTGGGGCCGG
GACCCGCAGAGCCGAGCGGACCCTTCTCTCCCGGTTGCGGCAGGGCAGAGCAGGGCGGG
GGGCTCCGCGCACCAACAGGACTAGTTCTCAGGGCGCTTTGCTCCTTGTTTTTACCCTC
GTTCTGTTTTTCTCCCTTCTCCGGAAGGCTTGTCAGGGGTAGGAGAAAGAGACGCAAAAC
ACAAAAGTGAAAACAGTTAATGACCAGCCACGGCATCCCTGCTGTGAGCTCTGGCCGCT
GCCTTCCAGGGCTCCCGAGCCACACGCTGTGGGTGCTGGCTGAGGGAAC

>A1_Vervet_AGM

GGAGGAAGGCGAGAACAGGCTTTGACCGATAGTAACCTCTGCGCTCCGGGCAGCCGAATC
TATAAAAGGAACTAGTCTCGGCAAAAACCCCGTAATTGCGAGCGAGAGTGAGTGGGGCCG
GGACCCGCAGAGCCGAGCCGACCCTTCTCTCCCGGCTGCGGCAGAGCAGGGCGGGGGC
TCCGCGCACCAACAGGACTAGTTCTCAGGGCGCTTTGCTCCTTGTTTTTCCCTCCTTC
TGTTTTTCCCTTCTCCGGAAGGCTTGTCAGGGGTAGGAGAAAGAGACGCAAAACAAA
AAGTGAAAACAGTTAATGACCAGCCACGGCTCCCTGCTGTGAGCTCTGGCCGCTGCCT
TCCAGGGCTCCCGAGCCACACGCTGTGGGTGCTGGCTGAGGGAAC

>A1_Bolivian_s_monkey

GGAGGAGGGAGAGAACAGGCTTTGACCGATAGTAACCTCGCGCTCGGTGCAGCCGAATC
TATAAAAGGAACTAGTCTCGGCAAAAACCCCGTAATTCCGAGGGAGAGTGAGTGGGGCCG
GGACCCGCAGAGCCGAGCCGACGCTCTCTCCCGGCTGCGGCAGCGCAGGGCAGGGCAGG
GTGGGGGGCGCCGCGCCTGCAGGGCCAGTTCTCAGGACGCTTGCTGCTTGTTTTTTC
TCTGGTTCAGTTTTTCTCCCGTCTCCGGAAGGCTTGTCAGGGGTGGGAGAAAGAGACGC
CAACACAAAAGTGAAAACAGTTAATGACTAGCCACGGTGTCCCTGCTGTGAGCTCTGGC
CACTGCCTTCCAGGGCTCCTGAGCCACACACTGTGGGTGCTGGCTGAGGTAAC

>A1_Marmoset

GTAATTCCGAGGGAGAGTGAGTGGGGCCGGGACCCGCAGAGCCGAGCCGACGCCTCTCTC
CGGGGTGCGGCAGCGCAGGGCAGGGTGGGGGGCGCCGCGGCCAACAGGGCCAG
TTCTCAGGACGCTTGCTTCTTTTTTCTCCGGTTCTGTTTTCTCCCTTCTCTGGAAGG
CTTGTCAGGGGTGGGAGAAAGAGACGCCAACACAAAAGTGAAAACAGTTAATGACCAG
CCACGGCGTCCCTGCTGTGAGCTCTGGCCGCTGTCTTCCAGGGCTCCTGAGCCACACGCT
GTGGGTGCTGGCTGAGGTAAC

>A1_Tarsier

AGAGAACAGGCTTTGACCGATAGTAACCCCTGCGCTCAGTGCAGCCGAATCTATAAAAGG
AACTAGTCTAGGCAAAAACCCGTAATTCCGTGGGAGAGTGAGTGGGGCCGGGACCCGCG
AGCCGAGCCGACATCTCTCCCTGGGAGCGGGTGGGCAGAGCAGAGGTGCCCCACGCC
ACCAGGGTTAATTCTCGGGCGCTTTGCTCCCTGGTTTTTTTTTTTTTTTTTTTCCC
CTTTCCCTTTCTTTTTCTCCCTTACCAGGAAGGGTTGCCTAGGGGTGGGGGAGAGA
CGCAAAACACAAAAGCAGAAAACAGTTAATGACCAGCCACTGTGTCCCTGCTGCGAGCTCT
GGCTACTGCCTTCCAGGGCTCCCAAGCCACACGCTGTGGGTGCTGGCGAGGGAAC

>A1_Mouse_lemur

GGAGGAGGCAGAGAGCAGGCTCTGACCGATAGTAACCCCTGCGCTCAGTGCAGCCGAATC
TATAAAAGGAACTAGTCTCGGCAAAAACCCGTAATTCCGAGAGAGACTGAGAGGGGCCG
GGACCCGCCGGGCCGGGTGGACACCGCTGTCTCCGGCGCGGGTGGGCAGGGCAGGGGG
CGCCGAGGGCCAGTTCTCGGGCGCTCCGCTCCCTGTTTTTTTCTCTCACTTTTCCC
CCCTCCCTTCTCGGGAAGGGCTTGTCAGAGGTGAGGGGAAGAGACGCAAAACAAAAG
TGGAAGCAGTTAATGACCAGCCACAGTATCCCTGCTGTGAGCTCTGGCCACTGCCTTCC
AGGGCTCCTGAGCCACACGCTGTGGGTGCCGGCTGAGGTGAC

>A1_Bushbaby

GACCAGCCACAGCATCCCTGCTGTGAGCTCTGGCCACTGTCTTCCAGGGCTCCTGAGGAG

CCAGACTCCGTGGGTGCCAGCGGAGGTGAC

>A1_Long_t_chinchilla

TACATAAACAGAGGCCGCGGAGGGGCGGGGGCGGAGGAGAGAACAGGCTTTGACCGCTG
GTAACCCCTGCGCTCGGCGCAGCCGAATCTATAAAAGGAACTAGTCGCGGCAAAAACCAG
TGATTCCGAGGGAGAGTGCGAGGGGTGCGGACCTGCGGAGCCGAGCCGCACATCTGGTCC
GCGAGCCGGGCACGCAGGGCAGGGCAGGGCAGGGCAGGGCAGGGGGCGCCGAGGACGCG
CGCAGCTAGAGCCAGTTCTCCGGGGCGCGCCGCCCTTGCGTCCTCTCCCCGTTTTTGT
TTTCCCCCTCCCCTTCACTGGAAGGGTTTGTCAAGGGGTGGGGGCAAGAAACGCAAAAC
AAAAGTGAAAAACAGTTAATGACCAGCCACGGTGTCCCTGCCGTGAGCTCTGGCCGCCG
CTCCAGGGCTCCCAAGCCACACACAAGCTTCTGCGGTGTTGGGGGTGTAAC

>A1_Degu

TTAATGACCAGCCACAGTGTCCCTGCTGTGAGCTCTGGCCACCGCCTCCAGAGCTCCAA
GCCACACACAAGCTTCTGCGGCATTGGAGGAGTAAC

>A1_Naked_mole_rat_m

ACTAGTCGCGGCAAAAACCAAGTGATTCCGAGGGAGAGTGAGAGGGGCCGGGGACCTGCA
GAGCCGAGCCCACTCTTCTGGTCCCCGAGCTGGGCAGGGGAGGGCAGGGGACGCCGCGAT
ACCCACGCACCTAGAGCCAGTTTCTCGGGGCGCTTCGCTCCCTGCGCCCCACCCCTT
TTTTCTCGCCTTCTGGAAGGGTTTGTCAAGGGGTGGGCAAGAAACGCAAAACGCAAAAG
AGGAAAACAGTTAATGACCAGCCACAGCGTCCCTGTTGTGAGCTCTGGCTGCTGCCTCA
GGGCTCCCAAGCCACACACGGGCTTCTGTGGTGTGGAGGAGGCAAC

>A1_Naked_mole_rat_f

ACTAGTCGCGGCAAAAACCAAGTGATTCCGAGGGAGAGTGAGAGGGGCCGGGGACCTGCA
GAGCCGAGCCCACTCTTCTGGTCCCCGAGCTGGGCAGGGGAGGGCAGGGGACGCCGCGAT
ACCCACGCACCTAGAGCCAGTTTCTCGGGGCGCTTCGCTCCCTGCGCCCCACCCCTT
TTTTCTCGCCTTCTGGAAGGGTTTGTCAAGGGGTGGGCAAGAAACGCAAAACGCAAAAG
AGGAAAACAGTTAATGACCAGCCACAGCGTCCCTGTTGTGAGCTCTGGCTGCTGCCTCA
GGGCTCCCAAGCCACACACGGGCTTCTGTGGTGTGGAGGAGGCAAC

>A1_Squirrel

GCTCGGCGCAGCCGAATCTATAAAAGGAACTAGTCGCGGCAAAACCCAGTAATCCGAGG
GAGAGTGAGTGGGGCCGGGACCTGCCGAGCCGAGCCACACCTCTCTTCCCTACGTGCG
TGGGCAGGACAGGGGGCTCCACAGGACGGCGCAGCCGAAGCCATTTCAGGGGCGCTT
GCTCCCCGTTTTCCCACTTCTTCTCTGCCATTACCCGGAACGGTTGGTCAAGGGGTG
GGGAAAGAGACGCAAAACAAAAAGTGAAAAACAGTTAATGACCAGCCACAGCGTCTGCT
GTGAGCTTCGGCCAGTGCTCCAGGGCTCTGAGCCACACAGGTTGCTTTGGGTGCCG
GCTGAGGTAAC

>A1_Chinese_hamster_CG

TAACCCCGTGCTCGGCACAGCCGAATCTATAAAAGGAACTAGTCGCGGCAAAAACCAGT
AATTCGAGGGCGAGCGAGCGGGCCGGGACCGGCCGAGCCACATCTCTACACAGCGC
AGGGCTGGGCAGGGGCGCCTCGGGACCCGCTACCCACAGCCGGTTTGGGGCGCTGCTCT
GCTCCCTGTTTCCCTTCACTTTTCCTCCCCCTTGCTGGAAGGGTTTGTGACGGGTGGC
TAAACAGACGCAAAACACAAAGTGAAAAACAGTTAATGACCAGCCACAGCGTCGCTGCTG
TGCACTCTGGCTGCTTCTCCAGGGCTCTGGAGCCACACGCACGGGTTGCTGTGGGTTC
AGCGGAGGTAAC

>A1_Golden_Hamster

GCGAGCGGGCCGGCACGGGCCGGGCCCACTTCTCTCCACCTAGCGCAGGGCTGGGAGGG
GCGCCGCGGGGACCCGCTCGCCACAGCCGGTTTGGGGCGCTGCTCTGCTCCCTGTTTCT
TCCTCTCCCCCAACACACACTTTTCCCTCCCCCTTTCTGGAAGGGTTTGTGACGGGTG
GGAGAAACAGACGCCAACACCGAAGTGGAACAGTTAATGACCAGCCACAGTGCCGTGT

GCTCTGGCTGCTTCTCCAGGGCTCTGGGGCCACACACACGGCTCGCCGTGGGTTCAGC
AGACGCAAC

>A1_Prairie_vole

AGCAATTCCGAGGGCGAGCGAGCGGGCCGGGACCGAGCCCACACCTCTCTCCGCAG
CGCAGGGCTGGGCAGGGGCGCCGCGGGACCCGCTCAGCCACAGCCGGTTTGGGGCGCTGC
TCTGCTCCATTTCCCCCACCCCACTTTTCTCTCCCCCTTCTGGAAGGGTTTGTGCAG
GGTGGGAGACACAGACGAAACACCAAAGTGGAAGACAGTTAATGACCAGCCACAGCGTC
ACTGCTGCGCGCGCTGGTTGCTTCTCCAGGGCTCTCGAGCCACATGCACAGGTTGCTCT
GGGTGCCAGCAGAGGTAAC

>A1_N_A_deer_mouse

AGTAATTCGAGGGCGAGCGAGCGGGCCGCGACCGGCCAAGCCACATCTCTCTCCGCAG
CGCAGGGCTGGGCAGGGGCGCCGCGGGACCCGCTCAGTCACAGCCGGTTTGGGGCGCTCC
TCTGCTCCCTGGTCCCCCACTTTTCTCCCCCTCCCCCTTCTGGAAGGGTTTGTGCAAG
GTGGGAGAACACAGACAAAACACAAACTGGAAAACAGTTAATGACCAGCCACAGCGTC
ACTGCTGTGTCTGTGCTGCTTCTCCAGGGCTCGCGAGCCACACACAGGTCGCTGT
GGGTGCCAGCGGAGGTAAC

>A1_Mouse

AGTAATTCGAGGGCGAGCGAGCGGGCCGGGACCGGCAGAGCCCACTTCTCTCCGCGGCG
CAGCGCAAAGCTGGGCAGGGGGCGCCGCGGGACCCGCGCAACCACAGCCGGCTTGGGGAG
CTGCTCTGCTCCCTGTTTCCCCCACTTTTCTTCCCCCTTCTGGAAGGGTTTGTGCAG
GGGTAGGGAAAACAGACTCAAACAGCAAAGTGAAAAACAGTTAATGACCAGCCACAGAGT
CACAGCTCTGTGCTCTGGCTGCTCCCTCCAGGGCTCTCGAGCCGACAGCGCAGGTCGCTG
TGGGTGCCGGCTGTGGTGAC

>A1_Algerian_mouse

AGTAATTCGAGGGCGAGCGAGCGGGCCGGGACCGGCAGAGCCCACTTCTCTCCGCGGCG
CAGCGCAAAGCTGGGCAGGGGGCGCCGCGGGACCCGCGCAGCCACAGCCGGCTTGGGGAG
CTGCTCTTCTCCCTGTTTCCCCCCCCCACTTTTCTCCCCCTTCTGGAAGGGTTTGTG
TGCAGGGGTAGGGAAAACAGTCTCAAACAGCAAAGTGAAAAACAGTTAATGACCAGCCAC
AGAGTCACAACTCTGTGCTCTGGCTGCTCCCTCCAGGGCTCTCGAGCCACAGACGCAGGT
CGCTGTGGGTGCCGGCTGCGGTGAC

>A1_Ryukyu_mouse

AGTAATTCGAGGGCGAGCGAGCGGGCCGGGACCGGCAGAACCCACTTCTCTCCGCGGCG
CAGCGCAAAGCTGGGCAGGGGGCGCCGCGGGACCCGCGCAGCCACAGCCGGCTTGCTCCG
TGTTTCCCCCACTTTTCTCCCCCTTCTGGAAGGGTTTGTGCAGGGGTAGGGAAAAC
AGACTCAGACAGCAAAGTGAAAAACAGTTAATGACCAGCCACAGAGTCACAGCTCTGTGC
TCTGGCTGCTCCCTCAGGGCTCTCGAGCCACAGACGCAGGTCGCTGTGGGTGCCGGCTG
CGGTGAC

>A1_Shrew_mouse

AGTAATTCGAGGGCGAGCGAGCGGGCCGGGACCGGCAGAGCCCACTTCTCTCCGCGGCG
CAGCGCAAAGCTGGGCAGGGGGCGCCGCGGGACCCGCGCAGCCATAGCCGGCTCGGGAAG
CTGCTCTGCTCCCTGTTTCTCCCACTTTTCTCCCCCTTCTGGAAGGGTTTGTGTAG
GGGTGGGGAAAACAAGACTCAAACAGCAAAGTGAAAAACAGTTAATGACCAGCCACGGAG
TCACAGCTCTGTGCTCTGGCTGCTTCTCCAGGGCTCTCAAGCCACACGCGCAGGTTGCT
GTGGGTGCCGGCTGCGGTGAC

>A1_Rat

TAACCCCGGCGCTCGGCACAGCCGAATCTATAAAAGGAACTAGTCGCGGCAAAAACAGT
AATTCGAGAGCGAGCGAGCGGGCCGGGACCGGCAGAGCCCACTTCTGTCTCTGCGGCTC
AGCGCAAAGCTGGGCAGGGGGCGCCGCGGGACCCGCGCAACCACAGCCGGCTTGGGGAGC

TGCTCTGCTCCCTGTCCCCACCCCTACTTTTTCTCCCGGTTTCTGGAAGGGTTTGTGT
AGGGGTTGGGTAAACAGACGCAAACAGCAAAGTGGAACAGTTAATGACCAGCCACAGT
GTCACAGCTCCGTGCTCTGGCTGCTTCCTCCAGGGCTCTCAAGCCACACACAGATCAC
TGTGGGTGCTGGCAGAGGTAAC

>A1_U_m_b_mole_rat

GGCGAGCGACGGGGCCGAAGCTGCGCAGGGCAGGGGCACGCCGACCCGCAACCCACAGGCA
CTCTGGGGGTGCTCCACTCCCCGCCCCACCCCGCTTTCTTTCTCTCCCCTTCTCTAGA
AGGGTTTGTCCAGGGGTGGGGAACAGACGCAAACATCAAAGTGGAACAGTTAATGA
CCAGCCAAAGCGTCCCTGCTGTGAGCTCTGGCCACTGCCTCCAGAGCTCTGAGCCACAC
ACGCAGGTCACCGTGGGTGCCAGCAGAGGTAAC

>A1_Lesser_E_jerboa

CAGTTAATGACCAGCCACAGCGTCCCTGCTGTGAGCTCTGGCCACTGCCTCCAGGCCTGT
GGAGCCACAAGCAGGACAGTGTGAAGGGTCAGCAGAAGAAAC

>A1_Rabbit

AGGAGAGAACAGGCTTTGACCGATAGTAACCCCCGGCGCCCGGAACGCCCAATTCTTTA
AAAGGAACTAGTCGCGGAAAAAACAGTTATTTCCGAACGAAGATGAGTTGCGGTCGGAA
CGGCCAGCGCCGCGCCGACCCCTTCTTCCAAGGCGGGCGGACAGGGTTGGGGGCC
CCAGGACCCCGACAGCCAGGCCTATTGCTTCCTGCGCTTTGCTCCCTAATTTTTTTTT
GCCCTCTTCTTTTCTCCCTTCTCCGGAAGGTTTGTACGGGTTGGGGAAAGAG
GCGCAAACAAAGTGAAAGCAGTTAATGACCAGCCACAGCGCTTGTCTACAAGCTCTG
GCCGCTGCCTCCAAGGGCGAGCCACAAGCAGGTTGCCGTGGGTGCCTGCTGAGGTAAC

>A1_Ferret

GGTTTTTCAAGGGTTGGGGACAAAGACGCACGCACAAAAGTGGAACAGTTAATGACC
AGCCACAGCGGCCCTGCCGTGAGTTGCGGCCACTGCCTCCACGGGCTCCTGAGCCACAG
CATCTGCTGGGGTGCTAGCAGCGGTAAC

>A1_Dog

TTAATGACCAGCCACGGCGGCCCTCCACGGGCCCCGAGCCACACGCACGGTGCTGGC
GCCGCCGCTGAGCCGAC

>A1_Cat

GAAACAGGCTTTGACCGATAGTAACCTGGCGCTCAGCGAGCCGAATCTATAAAAGGAAC
TAGTCGCGGCAAAACCCGGTAATCCGAGGGAGAGTGAGTGGCGCCGGATCCGCCGAG
CCCAGCGGACACCTCTCTCTCGGGCGCAGGTGGGCAGGGCAGGGGGGCGCCGCTGGAC
CGGCGCTGCCAGAACCAATTTTCGGGGCGCTTTGCTCCTCTTTTCCCCCTTTTCCCT
TCTCCGGAAGGGTTTTTCAAGGGGTGGGGAAAAGACGCACACAAAAGTGGAACAG
TTAATGACCAGCCACAGCGTCCCTGCTGTGAGCTGCTGCTGCCTCCAGAAGCTCCTG
AGCCACACACAGGTTGCTGCGGCTGCCGGCTGAGGAAAC

>A1_Horse

GTAATTCCGAGAGAGAGTGAGTGGGGCTGGCACCGGCCGAGCCGAGCTGACCCCTGTCCG
CCCGGGCGCAGGTGGGCAGGGCCGGGGCCCCGCGGACCAGCGCCACCAGAGCCAATTC
GCGGCGCGCTTTGCTCCTTTTCCCCCTTTTACCAGGAAGGATTTCAAAGGGTGGG
GAAAGACAGGCAACACAAAAGTGGAACAGTTAATGACCAGCCACAGCGTCCCTGCCC
TGAGCTCTGGCCACGGCCTTCACAGGCTCCAAGCCACACGCAGGTTGCCGTGGGTGCCA
GCTGAGGTAAC

>A1_Cow

AAGAGGGTGGCCTATTGACTGGCCATGACTGTGAAGTGGGACGCTTTCAACCAAGTAAC
GGAAATGTGCACAGGGGCAGAAATGAGTGTGCTACTAGCTTTGAAGCATTGAGAAATCTA
GTCAATACATAGTACGGCCTCTTTTGGTTTGGTTTGTGTTGCTTCCTTGAAAA
CTGATGTAGAAATGTCCCTTTGGCCCCAGTTACCTGAGAAGAGGATTTGGGCATTGCCAA

AACCTCCTCTTTGGATAAAATTGCTCCCAGCAGTGGTCACCATGTTCTGAGGTGGGAGGA
GGGAATTATGCAAGAGGCCAGGGCTATGCCAAGGTGGGTAGGAAAGCGGGGCAGGTACC
CAGCGAAGCGGAGCGATACCTTCCACCGCTGCATAGGCTACTTTTGCTGACGTCTCTGC
GTGTCTCCTGGAGGGTCTCTGAAGCCAGGATCTGTGTCTGGAAATGCAGAAACCCCGCTT
TAGGAAAACACTCTCCAGAAAGCTTCTCCTGGCCGTTCTCGTCTCGCGCCCATTCTGCA
GAGTTAATGACCAGCCACAGCGTCCCTGCTGTGAGCTCCAGCCACTTCTCTCCCGGGCTC
CCGAGCCTCACGCAGGCTGCTGTGGTGTGGCTGAGGTAAC

>A1_Sheep

CTCTCTACCCCGCGCAGGTGGGCAGAACAGGGGGCGCCGTGCAACCAGCGCCGCACAG
CCAATTTCTCGGGGCGCTTTGCTCCCTTTACCCCTTCTTTTCCGAAGGGTTTTTCAAG
GGGTGGGGAAAGAGCCACAAACACAAAAGTGGAAAACAGTTAATGACCAGCCACAGCGT
CCCTGCTGTGAGCTCCAGCCACTTCTTCCAGGCTCCCGAGCCTCACGCAGGCTGCTGT
GGTGTGGCTGAGGTAAC

>A1_Armadillo

GCCGAGCCCAGGCTTCGACCGATAGTAACCCCGCGCCACCGAATCTATAAAAGGAECTA
GTCGCGGCAAAACCCGGTAATTCCAAGCCAGAGTGAGTGGGGCCCCGATCGGCAGAGGC
CAGCCGACGCGTCGCTGCCCCGGTGCGGGTGGGCTGGGCAGGGGCGTCGCAGGACCCGC
GCCGCCGCCCGGGCCAGTTCCACGAGGCTTTGCCATTGTTTTCTGCTTTTCCCCAGTT
TCCAGAAGGGTTCAGTCGAGGGGTGGGGAAAGAGTCGCAGACACCAAAGTGAAAAACAG
TTAATGACCAGCCACAGCGTCCCTGCCGTGAGTTCCGGCCACTGCCCCACGGGCTCCCG
ACGAGCCACACGCAGGTTGCCTGGGCACGGCTGAGGTAAC

>A1_Tasmanian_devil

TTAATGACCAGCCACAGTGTCCCTGCCAGCGCTACCTCTGCTCCCTCCACTATCTGCCT
GGCTGAAGGGAGCTGTGGGTGTATGCTGAGGAAAG

>A1_Opossum

GAAAGAGAACAGAGTTTGACCGATAATAACCCCGCTGTTGAGCGCAGCCGAATCTATAAA
AGGAACTAGTCCAGGCTAAAAACCTTAATTCTGGGAGAGAGGGACTGAGGACCATATGC
TAGGCTGTGCAGGAGGAGGCCGTTGGTGCCCACAGGAAAGCCTAGGCAGGGGTTGGCAAA
GGCAAGGGAGGGGGTCAACAAAGGGGTATCGGAGGGGCGAGAGCCCGTTTGTGCGCTCTC
TCTCTCTCTCTCTTTTTTTTTTTTGTTCGCAAGTATTTCCAAAGGGGTAGGGGAG
AGAGTCCCAGAAACAAAAGTGAAAGCAGTTAATGACCAGCCACGGTGTCTTGCCAGT
GCTGCCTCTGCTCCCTCCACTATCTGCTGGACTGAGGGGAGCTGTGGGTGCGTGCTGAGG
AGAG

>A1_Platus

TGCTGTAATGACTAATGAGAACTCTCTCTCTCTGTTTTATCTTCCAGTTAATGACCA
GCCACAGTGTCCCTGCAGTGTGCTGTTGCCACTGCCCTGTACACCAGGGGAGCTGTGC
CTGCATGCTGAGGAAAG

>A1_Chicken

CCGGGAGCGCGGCCGCGTCGGGGCCGGATCCGTGAGTGGCAGAGCTGCGCTCGGGAGAA
GTGGAGCGAAGCGAGCCGACTCGGCTCGGTGCGGGACCGAGGGCGGCTGCCAGGGCTCC
CCGGCCCGAGTTAATGCCAGCCATAACGTCCTTGTGTGTACTGCTGCTGCTGCACAA
AGGGAAGAAGGCATTGCTGAAGAACG

>A1_Duck

AGTTAATGCCAGCCACAACCTCCTTGCTGTATACTGCTGCTGCTGCACAAGAGGAGCAA
GGCGTTGCTGAAGAGCG

>A1_Flycatcher

TTAATGCCTGGCCACAACATCTGTACTGTACTGCTGCTGCTACACAAAGAGAGCTTCTAG
TCTTTGCTGAAGGAAG

>A1_Anole_lizard

GGCAGAAAGGAAGCCGCGCTTGACCGATAATAACCTCCGCGCGCAGCGCCGGCGGGATC
TATATAAGGCGCTCCGAGTGCGGGACTGGCCCTGGTTTGGCCCCAGCCTCGTCGTTGGGA
GCCACAAAGGACGACCGGAGAAGCCCTCCATCGACTGGCTCTCCGATCTACTCCGGGAGG
GAGAGAAAGATCGCTCCCCATCAAAGTGGCGAGGAGTTGATGACCAGCCACAGTAGCTC
TGCTGTGTGCTGTTTCCATCCAAGGTGCTGATTGCTGGTACATACAGAAGGAAG

>A1_Coelacanth

GCCTGTACAGCGCTGCTCGGTGTCACCTATAAAAGCAACTAGAGGTTAAAAAAAACAT
TCAGTCTTACATCGAATCAGAGACACCTCAAGACCATATATATCGTTCACAGTAAACAAA
GAAAAGAAAAAATATATATATTTAAGGATTTTAAAGAAAGATACTATATATTTCTAGTA
GGGATTTGTCATTTTTCTGGTTTAACTATATTTTACATATACAGCCAAGTGGGGAAAAAA
GTTAATGTCCGACAACAATATCCCTGCTGCATGAGTGGAGCTGCTACTGGGAAAAG

>A1a_Platyfish

TGACCGTTAGTTACCTTTTGCTGCGCTCCGCGTCGAGCTGCGTATAAAAGCCTGCCTGGG
GAGAAAAAACAGTTTTCCTAGACACTGACAAAGAGCAAGGAGGACCGCGCGTCAGAGGG
ACGCAGTTGTTAACTCGTTTTTCGCTTTTGTGTTGAGAAAAGACGTTCAATTGATGAGA
GGCAGAAGAGCGAGTAAACACATTAAGATCAAGTGAGGAGCTGGCGTTGGCTTGTGAGG
TTAGGAAGACGGCAAC

>A1a_Medaka

GCGAGACTACCGCGAGTCCCGACAAGGGAAGTTTATCCTCGTTTTCTGGGACACCCT
ATGACACGAACAAGAGTGGATAGAAAAAAGTGAGGATCAGGAGTTTGCTTGTGAAGGTGG
GGAGAAGACAAC

>A1a_Tilapia

TGAGGAGCTGGAGCGTGCTCAGTGAGGTTGGGAAGACGGCAAC

>A1a_Tetraodon

TCTAGCCCGTTTACTGAGCGTCTTTTCTCTGTCAACTTTGCTTCATTAGTGAGGAGCTGG
AGCGTGCTCGGTGAGGTTGGGACGACATCAAC

>A1a_Zebrafish

GGAAAAAGGTCCGTCCATAAGAACACAAAGCACACCAGTACAATATCGCTCGGAGAGAG
TTTAAACAGCTGCTTTATCTCTCTTTTCTAAAAGGATTTACAGTTTATATGAAATAAGG
AACAATCATTAAAGGAGCAAGAAATAGCGCTTCTGTAAGGCGGCACGGATAGTGTGTATCC
GTGGTAAG

>A1b_Tilapia

ACCGGTTTGTAAAACTTCCACCAACATTCTTGTAAGTTCGTTTCTCCTCGCGGTGCGGT
TTTATTTGAAGTCTCGCGGAAGTGTGTGTGACTATTGGGGCTTTACTCCGGTGAGT
TTTGTCGGCGGCTGCTGCTCCGTGAGCAGCAG

>A1b_Cave_fish

TTTATTTATTCATTTTACTCCAACATATACCTATAAATAGAAAAATCATAGAAAAATCAGA
GCTGTATAATAACTATATCAAGACAGAAAGAACAAAATGGAGCTGCCAACATTGTATTCT
TTAATTTTACATTTTCACTCATTATCCCTAACTATTCTGGCATAGTCCTCACACTTTC
TTTGATCTCCAAATCAATTCTCTTCTTATTTCTTAATTTTCCCATTTTCTAGTTT
TTCGGCATGTGATTGAATCAGTGTAAGTCAATTGGCCACATTAACTTTGAACAGTGATG
GCTTGACCGTAAGACCACACACAAGCCCTCGGCCCATATGAATAGTTTCATTAGCATAGA
ACAATGGCGACAGATTGACTCTCAGACTGCCATTGAAGAGCTTTGATCAAAAGCTGCCGT
TCCTGGAGCAGTACCTGGAGCTGTGAAATTCCTTACAGAACGATTACCAGACTACAGT
GAAGTCCATTATTACTATTATAAAAGAAGAGAGGGGTGGGTGTTCTGAGTGAGATG
AGTCAGGACACAACAGGACAGGAGAAGACAGGAGGACGGGGGAAACGTGAGCAGAACGGA
TGCAGATAGGCTGGTGCACTTTGTCTGGACCTTTGAGCTGCTCTGTTGGTTGCACTTGG


```

AATCCTTCTTTTACTCAGCCTGTGCTCTCTGTGTCTCACTGTCAGCTGGAGAGCGCAG
CGTGAGGAGGACCTGAGTGAAG

>A1b1_Zebrafish

GACCGACAGTAACCCAGGCGCTCCGAGCGGATGCGCTATAAAACCACTTTAAACCCTTCA
GATGCCCTTCACACGAGAACCACCGCGTCAGACTCCACAAGGACGAATAAACATCTGA
TTCACCTGGATTACAACGAATGTCCTACACAAGTCATCAGCTGTGTCTGGTTGAGAGCTG
AGCGCGAAGAGGGACCCGAGTGAAG

>A1b2_Zebrafish

GAATAGCACCATTAGCATAGAGCAAGATCTCGACTTCCAGCAAACCTGTTTCATTGAAAT
GATCAAACCATTAAGAGACTCTAACGGCCTGCGAAATTCCTTTCTTGATGAATCTCTGAC
ACTACAAGGTCCACCACTTCCAGGAATTCTGGAAAAAAAAAAGACAGGAGATGAAAAGGG
AAGACGTGTGAGAGACGAAGCGAACGCTGGAGAGTAAGATAGGAGCTGCTGCACTTTGTC
CTGGACTTTTGAGCTGCTCTGTTGGTCGCTGGATGCTTTTTTGTCACTCAGCCCGT
CTCCGTTGTGTCATTGCAGTTGAGAGCTGAGCGCGAAGAGGGACCCGAGTGAAG

>A1a_Spotted_gar

CTGGACTCGAAGCCTGCACTTCCCTCACAGACTTCGTTATTGCGTGGCCCTTACCTTAA
TAAATCGATTTTAAATGCTCATAGATCCTCACATATGCCTTGGGAAATTTGTTCCATC
TATAGGAGTTAAATCCTGTTTTCAACGCTCGGATTCTTAATTTTCATTGGGACCTCTTT
GATTGTGAGGAATAAACTCACATTTCTTGAGCGTTCATAGCTTAATTTTCTCTTGCGAA
ACGAATAGTTTTTTCTTTATTTGTCGCGCTGGAAGGCAGAAGGACAAAACACAATGGCAG
GGTAGCCCTTTTAAGAGCTGTTTTGACTCTACACGTGCTTCGCTTGCAATGACTGAGCCA
CACAAACACAGCGGAGACTGGGATCTACTCGACCGAGAGTAACCCCTGCCTCCTCAGCGCA
GTGTACAGAGAGCGCTATAAAAAACAGCTAAAGGTTAAAAACGGCAGTTTCACAGACCGAG
ACCGCACGGTTCACAGCGCAGAGAGAAGATACAACAAAAGGAATCTTTGCTTCGTTTA
CTTTAAAGGATTTAGCATTTATGCATAAAGACACATAAAGACTGCGCGAGATAATAGTCG
AGCGCCGGAGTCTGTGAGCTGGAACCTCTGAGGAAAG

```

Supplemental Table S1: Multi-FASTA format file containing *ABCA1* 5'UTR sequences of 55 species (59 transcripts without intron sequences) downloaded from the Ensembl database and evaluated within comparative analyses.

Supplemental Table S2

Table S2: Summary of 59 transcripts from Ensembl and comparison of 5'UTR sections

Sp. No.	Sp. Name	Transcript			Lenght of 5'UTR sections (bp)				Length of protein (aa)	
		No.	Name	ID	Before intron	After intron	Whole 5'UTR	Intron		
1	Chimpanzee (<i>Pan troglodytes</i>)	1	ABCA1-201	ENSPTRT00000039236.5	320	92	412	31889	2261	
2	Bonobo (<i>Pan paniscus</i>)	2	ABCA1-201	ENSPPAT00000058048.1	312	92	404	24206	2102	
3	Human (<i>Homo sapiens</i>)	3	ABCA1-202	ENST00000374736.7	303	92	395	24163	2261	
4	Gorilla (<i>Gorilla gorilla</i>)	4	ABCA1-201	ENSGGOT00000009477.3	312	92	404	24188	2261	
5	Golden snub-nosed monkey (<i>Rhinopithecus roxellana</i>)	5		ENSRROT00000053936.1	323	92	415	26077	2261	
6	Black snub-nosed monkey (<i>Rhinopithecus bieti</i>)	6	ABCA1-201	ENSRBIT00000047284.1	360	92	452	25732	2261	
7	Angola colobus (<i>Colobus angolensis palliatus</i>)	7	ABCA1-201	ENSCANT00000035942.1	323	92	415	25552	2261	
8	Crab-eating macaque (<i>Macaca fascicularis</i>)	8	ABCA1-201	ENSMFAT00000023951.1	715	92	807	24586	2261	
9	Pig-tailed macaque (<i>Macaca nemestrina</i>)	9	ABCA1-201	ENSMNET00000039779.1	425	92	517	24876	2261	
10	Macaque (<i>Macaca mulatta</i>)	10	ABCA1-201	ENSM MUT00000028999.3	360	92	452	24582	2261	
11	Drill (<i>Mandrillus leucophaeus</i>)	11		ENSMLET00000035934.1	317	92	409	31834	1919	
12	Vervet-AGM (<i>Chlorocebus sabaeus</i>)	12	ABCA1-201	ENSCSAT00000009382.1	313	92	405	24875	2261	
13	Bolivian squirrel monkey (<i>Saimiri boliviensis boliviensis</i>)	13	ABCA1-201	ENSSBOT00000055395.1	321	92	413	26130	2261	
14	Marmoset (<i>Callithrix jacchus</i>)	14	ABCA1-201	ENSCJAT00000032583.2	229	92	321	25833	2261	
15	Tarsier (<i>Carlito syrichta</i>)	15	ABCA1-201	ENSTSYT00000031652.1	324	92	416	22034	2261	
16	Mouse Lemur (<i>Microcebus murinus</i>)	16	ABCA1-201	ENSMICT00000063500.1	310	92	402	20008	2261	
17	Bushbaby (<i>Otolemur garnettii</i>)	17	ABCA1-201	ENSOGAT00000009127.2	0	90	90	0	2264	
18	Long-tailed chinchilla (<i>Chinchilla lanigera</i>)	18	ABCA1-201	ENSCLAT00000022302.1	375	97	472	22789	2260	
19	Degu (<i>Octodon degus</i>)	19	ABCA1-201	ENSODET00000008072.1	0	97	97	0	2260	
20	Naked mole-rat male (<i>Heterocephalus glaber male</i>)	20	ABCA1-201	ENSHGLT00100013811.1	250	97	347	16484	2260	
20	Naked mole-rat female (<i>Heterocephalus glaber female</i>)	21	ABCA1-201	ENSHGLT00000016174.1	250	97	347	16244	2260	
21	Squirrel (<i>Ictidomys tridecemlineatus</i>)	22	ABCA1-201	ENSSTOT00000001269.3	275	96	371	20032	2261	
22	Chinese hamster CriGri (<i>Cricetulus griseus crigri</i>)	23	Abca1-201	ENSCGRT00000004943.1	272	100	372	16750	2261	
23	Golden Hamster (<i>Mesocricetus auratus</i>)	24	Abca1-201	ENSMAUT00000018446.1	215	94	309	17080	2261	
24	Prairie vole (<i>Microtus ochrogaster</i>)	25	Abca1-201	ENSMOCT00000016062.1	219	100	319	17510	2261	
25	Northern American deer mouse (<i>Peromyscus maniculatus bairdii</i>)	26	Abca1-201	ENSPENT00000010703.1	219	100	319	16816	2260	
26	Mouse (<i>Mus musculus</i>)	27	Abca1-201	ENSMUST00000030010.3	220	100	320	15580	2261	
27	Algerian mouse (<i>Mus spretus</i>)	28		MGP_SPRETEiJ_T0064864.1	225	100	325	16109	2261	
28	Ryukyu mouse (<i>Mus caroli</i>)	29	Abca1-201	MGP_CAROLIEiJ_T0062396.1	207	100	307	16091	2277	
29	Shrew mouse (<i>Mus pahari</i>)	30	Abca1-201	MGP_PahariEiJ_T0077533.1	221	100	321	16933	2261	
30	Rat (<i>Rattus norvegicus</i>)	31	Abca1-201	ENSRNOT00000024564.7	282	100	382	15826	2261	
31	Upper Galilee mountains blind mole rat (<i>Nannospalax galili</i>)	32	Abca1-201	ENSNGAT00000004944.1	173	100	273	24815	2261	
32	Lesser Egyptian jerboa (<i>Jaculus jaculus</i>)	33	Abca1-201	ENSJJAT00000027060.1	0	102	102	0	2261	
33	Rabbit (<i>Oryctolagus cuniculus</i>)	34	ABCA1-201	ENSOCUT00000009288.3	323	95	418	24629	2261	
34	Ferret (<i>Mustela putorius furo</i>)	35	ABCA1-201	ENSMPUT00000007173.1	51	98	149	20857	2261	
35	Dog (<i>Canis familiaris</i>)	36	ABCA1-201	ENSCAFT00000004346.3	0	78	78	0	2262	
36	Cat (<i>Felis catus</i>)	37	ABCA1-201	ENSFCAT00000004356.4	300	99	399	20563	2261	
37	Horse (<i>Equus caballus</i>)	38	ABCA1-201	ENSECAT00000011043.1	212	99	311	20272	2261	
38	Cow (<i>Bos taurus</i>)	39	ABCA1-201	ENSBTAT00000027538.3	543	98	641	2481	2261	
39	Sheep(<i>Ovis aries</i>)	40	ABCA1-201	ENSOART00000008003.1	160	98	258	19661	2261	
40	Armadillo (<i>Dasypus novemcinctus</i>)	41	ABCA1-201	ENSNDOT00000009223.3	300	100	400	20170	2264	
41	Tasmanian devil (<i>Sarcophilus harrisii</i>)	42	ABCA1-201	ENSSHAT00000008717.1	0	95	95	0	2264	
42	Opossum (<i>Monodelphis domestica</i>)	43	ABCA1-201	ENSMODT00000020693.2	329	95	424	25789	2260	
43	Platypus (<i>Ornithorhynchus anatinus</i>)	44	ABCA1-201	ENSOANT00000004496.2	0	137	137	0	2266	
44	Chicken (<i>Gallus gallus</i>)	45	ABCA1-201	ENSGALT00000024891.7	131	75	206	15067	2260	
45	Duck (<i>Anas platyrhynchos</i>)	46	ABCA1-201	ENSAPLT00000015088.1	0	77	77	0	2261	
46	Flycatcher (<i>Ficedula albicollis</i>)	47	ABCA1-201	ENSFALT00000005231.1	0	76	76	0	2262	
47	Anole lizard (<i>Anolis carolinensis</i>)	48	ABCA1-201	ENSACAT00000014788.2	217	77	294	35441	2258	
48	Coelacanth (<i>Latimeria chalumnae</i>)	49	ABCA1-201	ENSLACT00000009960.1	241	55	296	33177	1525	
49	Platyfish (<i>Xiphophorus maculatus</i>)	50	abca1a-201	ENSXMAT00000006594.1	213	43	256	3530	2270	
50	Medaka (<i>Oryzias latipes</i>)	51	abca1a-201	ENSORLT00000017693.1	90	42	132	2863	2283	
51	Tilapia (<i>Oreochromis niloticus</i>)	52	abca1a-201	ENSONIT00000020384.1	0	43	43	0	2269	
52	Tetraodon (<i>Tetraodon nigroviridis</i>)	53	abca1a-201	ENSTNIT00000001913.1	49	43	92	119	2269	
53	Zebrafish (<i>Danio rerio</i>)	54	abca1a-203	ENSDDART00000167514.2	148	40	188	2551	2268	
51	Tilapia (<i>Oreochromis niloticus</i>)	55	abca1b-201	ENSONIT00000002004.1	0	153	153	0	2337	
54	Cave fish (<i>Astyanax mexicanus</i>)	56	abca1b-201	ENSAMXT00000016919.1	0	742	742	0	2293	
53	Zebrafish (<i>Danio rerio</i>)	57	abca1b-201	ENSDDART00000051556.5	170	35	205	5575	2282	
53	Zebrafish (<i>Danio rerio</i>)	58	abca1b-202	ENSDDART00000148461.2	0	354	354	0	2271	
55	Spotted gar (<i>Lepisosteus oculatus</i>)	59	abca1a-201	ENSLOCT00000014184.1	597	40	637	3435	2262	
Primates (N=17)					Median:	320	92	412	24875	2261
Rodents (N=16)					Median:	221	100	321	16750	2261
Placental mammals (N=7)					Median:	212	98	311	20170	2261
Marsupials and Platypus (N=3)					Median:	0	95	137	0	2264
Reptiles and Birds and Coelacanth (N=5)					Median:	131	76	206	15067	2260
Ray finned fishes - abca1a (N=6)					Median:	119	43	160	2707	2269
Ray finned fishes abca1b (N=3)					Median:	0	254	280	0	2288

Abbreviations: sp., species;

Supplemental Table S2: Summary of all species and names and IDs of transcripts downloaded from the Ensembl database, lengths of 5'UTR sections evaluated - whole 5'UTR, 5'UTR section located before-Intron-1, section located after-Intron-1, whole Intron 1 lengths, and ABCA1 protein lengths in amino acids.

Supplemental Table S3

Table S3: Summary of alignment analyses (59 transcripts from Ensembl) performed with the help of 9 alignment programs.

Program	No. of most cons. nts	No. of most cons. nts within after Intron 1 sec.	Consensus seq. - start of 5'UTR sec. after Intron 1	Consensus seq. - end of 5'UTR sec. after Intron 1 preceding mORF
<i>Clustal</i>	40	38	AG↓TTA ATG ACCAGCCAC	GCTGAGGTAAC(ATG)
<i>ClustalO</i>	16	16	AG↓TTA ATG ACCAGCCAC	GCGGGGGTAAC(ATG)
<i>GLprobs</i>	37	37	AG↓TTA ATG ACCAGCCAC	GCTGAGGTAAC(ATG)
<i>Mafft</i>	30	28	AG↓TTA ATG ACCAGCCAC	GCTGAGGTAAC(ATG)
<i>MSAprobs</i>	38	38	AG↓TTA ATG ACCAGCCAC	GCTGAGGTAAC(ATG)
<i>Muscle</i>	41	39	AG↓TTA ATG ACCAGCCAC	GCTGAGGTAAC(ATG)
<i>Probcons</i>	39	39	AG↓TTA ATG ACCAGCCAC	GCTGAGGTAAC(ATG)
<i>TCoffee</i>	22	21	AG↓TTA ATG ACCAGCCAC	GCTGAGGTAAC(ATG)
<i>WebPRANK</i>	26	26	AG↓TTA ATG ACCAGCCAC	GCTGAGGGAAC(ATG)
Median	37	37		
Range	16-41	16-39		

Notes: The most conserved nts had >80% identity and >84% occupancy scores. Position of the Intron 1 was specified by an arrow sign. ATG sequence in brackets is the first codon of the mORF.

Abbreviations: cons., conserved; no, number; nts, nucleotides; sec., section; seq., sequence;

Supplemental Table S3: Summary of multiple-sequence (59 transcripts) alignment analyses performed with the help of 9 alignment programs. Highly conserved nucleotides are those with >80% identity and >84% occupancy scores.

Supplemental Table S4

Table S4: Upstream ATG and non-ATG start as well as stop codon comparison

Sp. No.	Sp. Name	Tr. No.	No. of upstream start codons										No. of upstream stop codons					
			uATGs		uGTGs		uCTGs		uTTGs		uACGs		uTAGs		uTAAs		uTGAs	
			Wh. 5'UT R	Af. In. 1 section	Wh. 5'UT R	Af. In. 1 section	Wh. 5'UT R	Af. In. 1 section	Wh. 5'UT R	Af. In. 1 section	Wh. 5'UT R	Af. In. 1 section	Wh. 5'UT R	Af. In. 1 section	Wh. 5'UT R	Af. In. 1 section	Wh. 5'UT R	Af. In. 1 section
1	Chimpanzee (Pan troglodytes)	1	1	1	7	3	10	7	5	0	4	2	3	0	4	1	5	3
2	Bonobo (Pan paniscus)	2	1	1	7	3	10	7	5	0	4	2	3	0	4	1	5	3
3	Human (Homo sapiens)	3	1	1	6	2	10	7	5	0	3	2	3	0	4	1	5	3
4	Gorilla (Gorilla gorilla)	4	1	1	6	3	9	7	6	0	3	2	3	0	4	1	5	3
5	Golden snub-nosed monkey (Rhinopithecus roxellana)	5	1	1	6	3	10	7	5	0	2	1	3	0	4	1	5	3
6	Black snub-nosed monkey (Rhinopithecus bieti)	6	1	1	6	3	11	7	5	0	2	1	3	0	5	1	6	3
7	Angola colobus (Colobus angolensis palliatus)	7	1	1	6	3	13	7	5	0	3	2	4	0	4	1	5	3
8	Crab-eating macaque (Macaca fascicularis)	8	3	1	9	3	21	8	8	0	6	2	9	0	7	1	13	4
9	Pig-tailed macaque (Macaca nemestrina)	9	1	1	9	3	15	8	6	0	5	2	4	0	6	1	11	4
10	Macaque (Macaca mulatta)	10	1	1	6	3	12	8	6	0	3	2	4	0	5	1	7	4
11	Drill (Mandrillus leucophaeus)	11	1	1	6	3	9	7	6	0	3	2	4	0	4	1	5	3
12	Vervet-AGM (Chlorocebus sabaeus)	12	1	1	6	3	10	7	5	0	3	2	4	0	4	1	5	3
13	Bolivian squirrel monkey (Saimiri boliviensis boliviensis)	13	1	1	10	4	12	8	4	0	4	1	3	1	5	2	6	4
14	Marmoset (Callithrix jacchus)	14	1	1	8	3	11	8	2	0	5	2	0	0	3	2	5	4
15	Tarsier (Carlito syrichta)	15	1	1	10	3	10	7	3	0	3	1	4	0	5	1	3	1
16	Mouse Lemur (Microcebus murinus)	16	1	1	9	4	12	7	1	0	2	1	2	0	4	1	8	5
17	Bushbaby (Otolemur garnettii)	17	0	nr.	4	nr.	5	nr.	0	nr.	0	nr.	0	nr.	0	nr.	3	nr.
18	Long-tailed chinchilla (Chinchilla lanigera)	18	1	1	8	4	8	3	5	1	4	1	2	0	5	2	4	2
19	Degu (Octodon degus)	19	1	nr.	2	nr.	4	nr.	1	nr.	0	nr.	0	nr.	2	nr.	2	nr.
20	Naked mole-rat male (Heterocephalus glaber male)	20	1	1	6	3	10	5	3	2	5	1	2	0	1	1	4	2
20	Naked mole-rat female (Heterocephalus glaber female)	21	1	1	6	3	10	5	3	2	5	1	2	0	1	1	4	2
21	Squirrel (Ictidomys tridecemlineatus)	22	1	1	9	3	6	4	4	2	4	0	1	0	4	2	5	4
22	Chinese hamster CriGri (Cricetulus griseus crigri)	23	1	1	6	2	11	6	4	1	3	2	1	0	6	2	1	1
23	Golden Hamster (Mesocricetus auratus)	24	1	1	6	3	8	3	2	0	4	2	1	0	1	1	1	1
24	Prairie vole (Microtus ochrogaster)	25	2	2	4	1	8	4	4	2	1	0	0	0	2	2	1	1
25	Northern American deer	26	1	1	4	2	11	6	2	0	0	0	0	0	3	2	1	1

	mouse (Peromyscus maniculatus bairdii)																	
26	Mouse (Mus musculus)	27	1	1	7	5	10	5	2	0	1	1	1	0	2	1	2	2
27	Algerian mouse (Mus spretus)	28	1	1	6	4	9	5	2	0	1	1	1	0	2	1	2	2
28	Ryukyu mouse (Mus caroli)	29	1	1	7	4	7	5	2	0	1	1	1	0	2	1	2	2
29	Shrew mouse (Mus pahari)	30	1	1	7	4	10	5	2	1	2	2	2	0	2	1	2	2
30	Rat (Rattus norvegicus)	31	1	1	6	4	11	4	3	0	1	0	2	0	6	2	1	1
31	Upper Galilee mountains blind mole rat (Nannospalax galili)	32	1	1	6	3	7	5	1	0	3	1	1	0	2	2	3	3
32	Lesser Egyptian jerboa (Jaculus jaculus)	33	1	nr.	3	nr.	5	nr.	0	nr.	0	nr.	0	nr.	1	nr.	3	nr.
33	Rabbit (Oryctolagus cuniculus)	34	2	1	3	2	5	4	10	2	4	0	2	0	5	2	4	2
34	Ferret (Mustela putorius furo)	35	1	1	3	2	5	5	2	1	4	2	1	1	2	2	3	3
35	Dog (Canis familiaris)	36	1	nr.	1	nr.	2	nr.	0	nr.	4	nr.	0	nr.	1	nr.	2	nr.
36	Cat (Felis catus)	37	1	1	6	1	13	10	3	1	1	0	2	0	4	1	6	4
37	Horse (Equus caballus)	38	1	1	7	2	7	4	2	1	2	2	0	0	3	2	5	3
38	Cow (Bos taurus)	39	10	1	13	3	21	6	15	0	4	1	7	0	4	2	14	3
39	Sheep(Ovis aries)	40	1	1	6	3	6	6	2	0	1	1	0	0	2	2	3	3
40	Armadillo (Dasypus novemcinctus)	41	1	1	7	1	7	4	3	1	6	4	2	0	5	2	4	3
41	Tasmanian devil (Sarcophilus harrisii)	42	2	nr.	3	nr.	7	nr.	0	nr.	0	nr.	0	nr.	1	nr.	3	nr.
42	Opossum (Monodelphis domestica)	43	2	1	8	5	11	7	6	1	1	1	4	0	6	1	5	3
43	Platypus (Ornithorhynchus anatinus)	44	4	nr.	3	nr.	9	nr.	1	nr.	0	nr.	0	nr.	3	nr.	4	nr.
44	Chicken (Gallus gallus)	45	1	1	5	1	7	5	3	3	2	2	0	0	2	2	2	1
45	Duck (Anas platyrhynchos)	46	1	nr.	0	nr.	6	nr.	2	nr.	0	nr.	0	nr.	1	nr.	1	nr.
46	Flycatcher (Ficedula albicollis)	47	1	nr.	0	nr.	7	nr.	1	nr.	0	nr.	1	nr.	1	nr.	1	nr.
47	Anole lizard (Anolis carolinensis)	48	1	1	4	2	9	5	4	2	1	0	1	1	3	0	4	3
48	Coelacanth (Latimeria chalumnae)	49	2	2	3	1	7	4	1	0	0	0	3	0	7	1	1	1
49	Platyfish (Xiphophorus maculatus)	50	1	0	0	0	6	1	8	3	3	1	3	1	4	0	7	2
50	Medaka (Oryzias latipes)	51	1	0	3	2	1	0	2	2	1	0	1	0	0	0	3	2
51	Tilapia (Oreochromis niloticus)	52	0	nr.	2	nr.	1	nr.	1	nr.	1	nr.	0	nr.	0	nr.	2	nr.
52	Tetraodon (Tetraodon nigroviridis)	53	0	0	2	2	3	1	2	1	1	1	2	0	0	0	3	2
53	Zebrafish (Danio rerio)	54	1	0	2	2	2	1	0	0	1	1	2	1	7	2	1	0
51	Tilapia (Oreochromis niloticus)	55	0	nr.	4	nr.	4	nr.	6	nr.	0	nr.	1	nr.	1	nr.	4	nr.
54	Cave fish (Astyanax mexicanus)	56	7	nr.	14	nr.	18	nr.	14	nr.	4	nr.	8	nr.	10	nr.	20	nr.
53	Zebrafish (Danio rerio)	57	3	0	2	1	4	1	2	1	3	0	0	0	4	0	4	3
53	Zebrafish (Danio rerio)	58	4	nr.	3	nr.	13	nr.	8	nr.	3	nr.	4	nr.	3	nr.	10	nr.
55	Spotted gar (Lepisosteus oculatus)	59	5	0	5	1	12	3	12	0	5	0	7	0	16	0	8	2
Primates (N=17)		M. :	1	1	6	3	10	7	5	0	3	2	3	0	4	1	5	3

Rodents (N=16)	M. :	1	1	6	3	8	5	2	1	2	1	1	0	2	2	2	2
Placental mammals (N=7)	M. :	1	1	6	2	7	6	2	1	4	2	1	0	3	2	4	3
Marsupials and Platypus (N=3)	M. :	2	1	3	5	9	7	1	1	0	1	0	0	3	1	4	3
Reptiles and Birds and Coelacanth (N=5)	M. :	1	1	3	1	7	5	2	2	0	0	1	0	2	1	1	1
Ray finned fishes - abca1a (N=6)	M. :	1	0	2	2	3	1	2	1	1	1	2	0	2	0	3	2
Ray finned fishes abca1b (N=3)	M. :	4	0	4	1	9	1	7	1	3	0	3	0	4	0	7	3

Abbreviations: M., median; nr., not relevant; sp., species; tr., transcript; Wh., whole; Af., after; In., intron;

Supplemental Table S4: Upstream ATG and non-ATG start as well as stop codon comparison – number and position.

Supplemental Table S5

Table S5: 5'UTR GC content versus length comparison

Species	GC content (%)				Whole 5'UTR length (nts)
	Before In. 1	After In. 1	Whole r.	Subtr. After_In._1 - Before_In._1	
Human	58.7	67.4	60.8	8.7	395
Macaque	58.1	64.1	59.3	6	452
Mouse Lemur	61.9	62	61.9	0.1	402
Rabbit	57.9	61.1	58.6	3.2	418
Mouse	63.2	64	63.4	0.8	320
Squirrel	60	60.4	60.1	0.4	371
Cat	60	61.6	60.4	1.6	399
Armadillo	63.7	67	64.5	3.3	400
Tasmanian devil	n. r.	n. r.	57.9	n. r.	95
Opossum	51.7	61.1	53.8	9.4	424
Platypus	n. r.	n. r.	52.6	n. r.	137
Chicken	77.9	49.3	67.5	-28.6	206
Flycatcher	n. r.	n. r.	46.1	n. r.	76
Anole lizard	63.6	49.4	59.9	-14.2	294
Coelacanth	32	47.3	34.8	15.3	296
Zebrafish (A1a)	37.8	52.5	41	14.7	188
Spotted gar (A1a)	42.5	60	43.6	17.5	637

Abbreviations: In, Intron; n. r., not relevant; r., region; subtr., subtraction;

Supplemental Table S5: GC content versus length comparison in 5'UTRs of orthologous *ABCA1* genes.

Supplemental Table S6

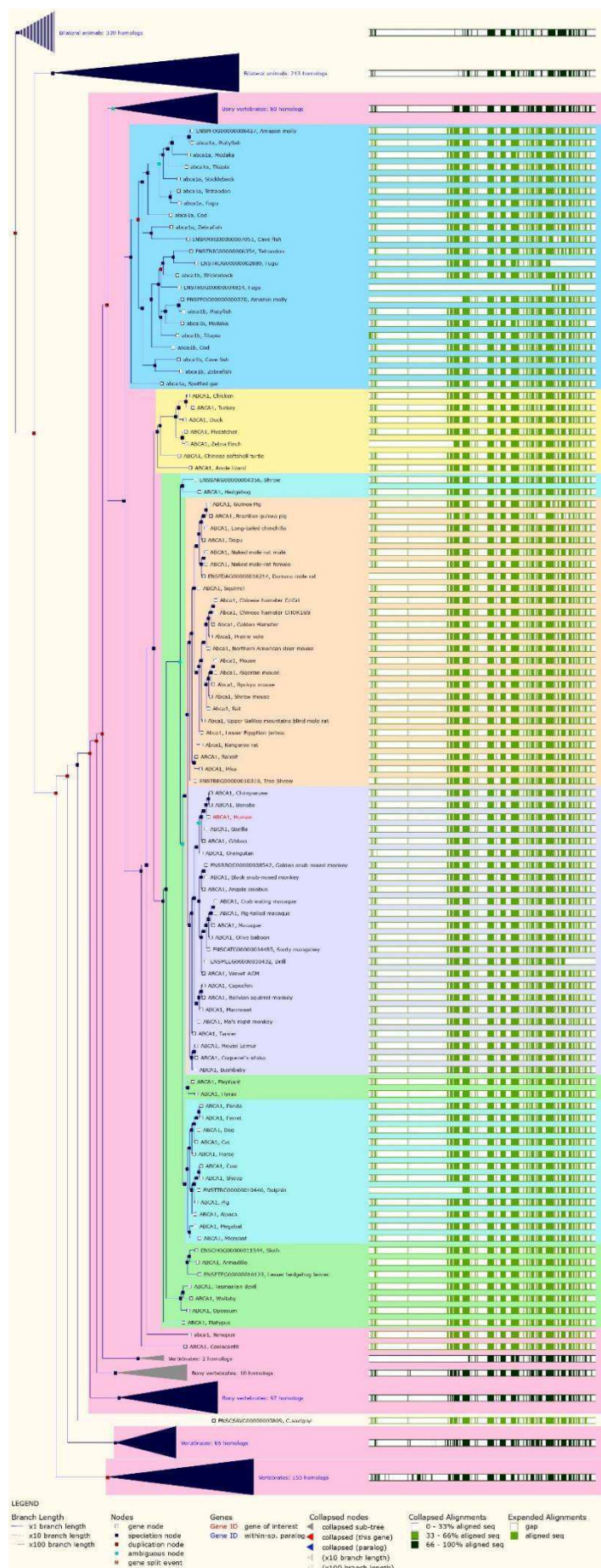
Table S6: Analysis of conserved RNA motifs and sites in human ABCA1 5'UTR - detailed information

Motif_Type	Motif_Name	Position	Human	
			Length	Sequence
exon splicing enhancer (ESE)	ct/cgrp intron downstream from exon 4	217 ~ 225	9	tttttccc
	sc35 - exonic splicing enhancer	113 ~ 120	8	gacccgca
	sc35 - exonic splicing enhancer	356 ~ 363	8	ggctcccg
exon splicing silencer (ESS)	fibronectin eda exon	260 ~ 265	6	caaggg
intron splicing enhancer (ISE)	gh-1 intron 3	384 ~ 393	10	ggctgagggga
	gh-1 intron 3	105 ~ 114	10	ggggccggga
	ighg2 cgamma2 (immunoglobulin heavy chain subclass g2 - cgamma2 gene) - intron 1	99 ~ 104	6	gtgagt
	ighg2 cgamma2 (immunoglobulin heavy chain subclass g2 - cgamma2 gene) - intron 1	331 ~ 336	6	gtgagc
	ctnt, exon 5	253 ~ 259	7	ggcttgt
	ctnt, exon 5	384 ~ 390	7	ggctgag
rho-independent terminator	Rho-independent Terminator	197 ~ 215	19	cagggcgctttgtccttg
Transcriptional regulatory motif (TRM)	Sp1	158 ~ 170	13	gcagggcggggag
	Elk-1	243 ~ 256	14	ttctccggaaggct
	TATA	51 ~ 65	15	ctataaaaggaacta
	Msx-1	81 ~ 89	9	ccgtaattg
	TFIIA	52 ~ 63	12	tataaaaggaac
	Kid3	315 ~ 319	5	ccacg
	NFAT1	295 ~ 300	6	ggaaaa
	NF-AT4	295 ~ 300	6	ggaaaa
	HOXA13	53 ~ 58	6	ataaaa
	ER81	246 ~ 255	10	tccggaaggc
	PEA3	246 ~ 255	10	tccggaaggc
	ETV7	246 ~ 255	10	tccggaaggc
	E2F-3	162 ~ 168	7	ggcgggg
UTR motifs	Musashi binding element (MBE)	20 ~ 24	5	atagt
microRNA target sites	hsa-miR-4435	323 ~ 345	23	tccctgctgtgagctctggccgc
	hsa-miR-5581-5p	236 ~ 256	21	tctccccttctccggaaggct

Supplemental Table S6: Analysis of conserved RNA motifs and sites in human ABCA1 5'UTR – detailed information.

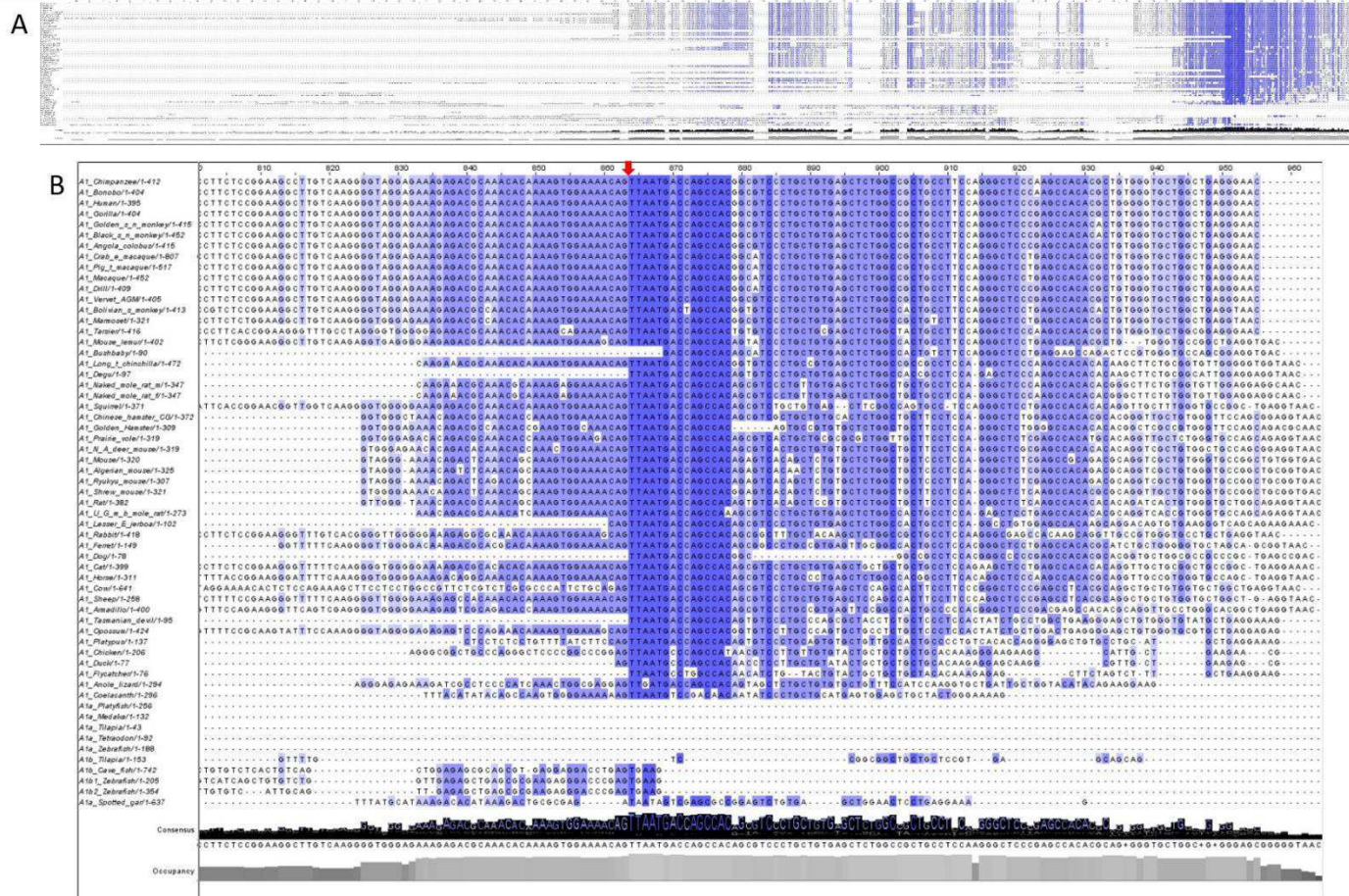
Supplementary Information 2 – Supplementary Figures

Supplemental Figure S1



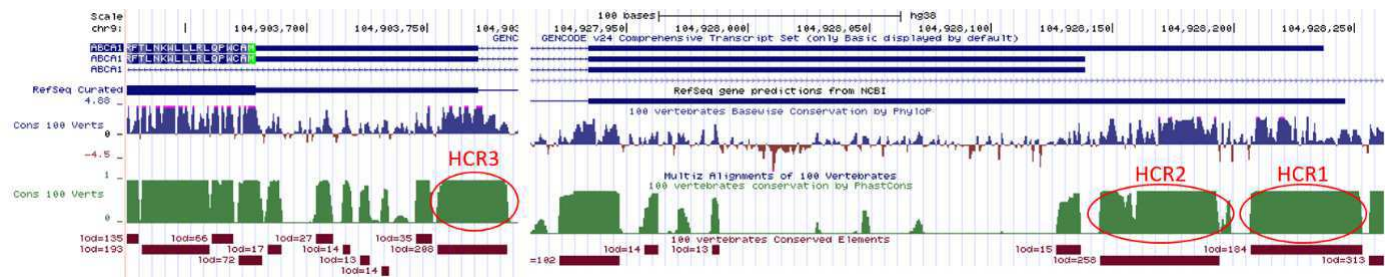
Supplemental Figure S1: The *ABCA1* gene tree created by the Gene Tree tool – Ensembl database.

Supplemental Figure S2



Supplemental Figure S2: Results of the ABCA1 5'UTR multi-sequence alignment analysis in the ClustalO program performed on 59 transcripts downloaded from the Ensembl database. Lines represent transcripts, which were aligned according to the ClustalO algorithm; the conservation of nucleotides among vertebrate species is highlighted with the most conserved ones having the darkest color. A) Alignment of the whole 5'UTRs. 5'UTR segments, which contain conserved nucleotides can be distinguished. B) Detail of the 5'UTR segment containing the most conserved sequences; the position of the spliced Intron 1 is marked by a red arrow.

Supplemental Figure S3



Supplemental Figure S3: Results of the conservation track focused on the 5'UTR region of the human *ABCA1* gene performed in UCSC Genome Browser. Three highly conserved regions (HCR1-HCR3) can be distinguished. HCR3 is overlapping with the most conserved sequence displayed on Figure 4 in the main text.

CCGGAGCGCGCCGCGCTCGGCGCCGATCCGTGAGTTGCGAGATCTCTCGGAGAAATGAGAGCGAGAGCCGACTCGGCTCGGTGCGGACCGAGGCGCGCTCCAGGCGTCCCCGCCCCGAGTTAATGCGCCAGCCATAACGTCCTGTGTGTATGCTGCTGCTGCTGACAAAGGAGAGAGAGGCTATTTGAGAGAGC

AGTTAATGCCAGCCACACCTCTCTTATATATGCTCTCTCCACAAAGGAGCAAGGCGTTTCTAGAGGCC

TTAATGCTGGCCACAACATCTGTATCTGATCTCTGCTCTACACAAGAGAGCTTCTAGTCTTTCTGAAGGAAG

[illegible]

GC TGGTACAGAGCTCTCTGGTGTGACACTATAAAGCACTATGAGTGTAAAAAAACAACTTCAGTCTCTACATCGATGATCAGAGACACCTCAAGACCATATATATATCTGTCACGATAAACAAGAAAAAGAAAAATATATATATTATTAAAGGATTTTAAAGAAAGATACTATATATTCTCTAGTAGGAGATTGTGTCATTTTGGTTTAACTATATTTCATCATACACCAAGTGGGAAAAAATCTAAATGATCGCCACAACTATCTCTCTCATGAGTGGAGCTCTAAGGAAAAAG

TACCGTGAATACCTTTTCTCCCTCCGCTCAAGCTGATATAAAGCTTCTGGGAGAAAAACAGTTTCTTACACACAAAAGACAAAGAGGACCCGCGCTCAAGAGACGCACTGTAACTCTGTTTTCCTTTGCTTTGTTTGAAGAAAGACCTTCAATTGATGAGAGCGAAGAGCCAGTAACACATTAAATCAAGAGAGAGCTTCTGCTTGGAGGTTAGGTTAGGAGACGCAAG

GCGAGACTCACCSCGAGTCCCGACAGGGAGTTTATTCTCTGTTTTTGGACACCCCTATGACACGAACACAGAGTGGATAGAAAAAATGAGGATCAGSAGTTTCTTGTAAGGTGGGAGAGACAC

TSAGGAGCTGGAGCGTCTCNGTGAAGTTGGGAGACGGCAGC

TCTAGCCGCTTTTATGAGCGTCTTTCTCTCAACTTTCCTTCATTAGTGGAGGAGCGCTGCTCGGTGAGGTTGGGACGACATCAAG

GGAAAAGGTCCCTCCATAGAACACAAACACCAACCTACACATATCTCCGAGAGAGTTTAAACAGCTTTTATCTCTCTTTCTAAAAGGATTACAGTTTATATGAATAAAGAACATCTATTAAGGACCAAGAAATAGCGCTTTTAAAGGCGACGAGATAGTGTGTATCCCGGTAAAG

ACCGGTTTGTAAAGCTCCACCAACATTCTGTAGTTCGTTTCCCTCCCTTTCGCGGCGGGTTTATTTTGAAGGTCCTTTCGCGAAGTGTTGTGTGACTATTGGGGCTTTACTCCGGTGAGTTTGTCCGGCGGGTTCCTCTCCGTGASCAGACAT

[illegible][illegible][illegible]

5

CCGGAGCGCGCCGCGTCCGGGCGGAFCCGTGAGTGACAGCTCCGCTCGGAGAAGTGAGSCGAAGCGAGCCGACTCGGCTCGGTGCGGGACCGAGGGCGGCTGCCCAAGGCTCCCGGGCCCGGAGTTAATGCCCAAGCATAAGCTCTGGTGTGTACTGCTGCTGCTGCAAAAGGGAGAAGGCACTGTAAGAAAC

AGTTAATGCCAGCCACACCTCTCTGCTGATACTGCTGCTGCTGCACAGAGSAGCAAGGCCCTGCTGAGAGCCG

TTAATGCTTGCCACACATCTGTACTGTACTGCTGCTACACAAAGAGACTTCTAGTCTTCTGGAAGGAAG

[illegible]

GCCTGTACAGCGCTGCTCGGTGTGCACATATAAAAGCACTGAGGGTAAAAAANATTTCAGTCTTACATCGAATCAGAGACACTCAGAGCCATATATATCGTTCACAGTAAACAAGAAAAAAGAAAAATATATATATTTAGGGAATTTAAAGAAAGTACTATATTTCTAGTAGGGAATTCATTTTCTGGTTTAACTATATTTTACATATACAGCCAGTGGGGAAAAAGTTAATGTCCGACACAAATATCCCTGTGCATGAGTGGAGCTGCTACGGGAAAG

TGAACCTTAGTTACCTTCTGCGCTCCGGTCGAGCTGCGTATAAAGCGCTGCTGCGGGAGAAAAACAGTTTCTAGACACTGACAAAGAGCAAGGAGGACGCGCGCTGCGAGGGGAGCGAGTTTAACTCGTTTTTCGCTTTGAGAAAAGGACGTCATGTTGAGGGGAGAGAGCGAGTAAACACATTAGATCAGTGGAGGCTGGCGGCTGAGTTAGGAGAGCGCAAG

CGCGACTCACCGGAGTCCCGACAAAGGGAAGTTTATTCCTCGTTTTCTGGGACACCCCTATGACACGAACAAGAGTGGATAGAAAAAAGTGGAGGATCAGGAGTTGGAGGTGGGAGAGAGCAAC

TSAGGAGCTGGAGCGTCTCAGTGAAGTGGAGGACGGCAC

TCATGCCCGTTACTGAGCGTCTTTCCTCTGTCAACTTCTTCATTATGTTGAGGAGCTGGAGCGTGCTCTGGTGAAGTTGGAGCGACATCAAC

(G)AAGAGG(TCC)GTCCATAGAACACAAAGCACCACAGTACANTATGCTCGGAGAGAGTTTAAACAGCTGCTTTATCTCTCTTTCTAAGAGATTACAGTATTATATGAATAGGAGACATCATTAAGGAGCAGGAATAGCGCTCTCTGTAGGCGGCACGGATAGTGTGTATCCGTGTGTAA

ACCGGTTCTAAACCTTCACCAACATCTCTGATTCGTTTCCTCCCTGCGGCTCGGATTTTATTTGAGGCTCTGCGGGAAGTGTTCTGACTATGAGGCTTTACTCCGTTGAGTTTCCGACCGGCTGCTGCTCCGTTGAGCAGCA

[illegible]

GACCCACAGTAACCCAGGCGCTCCGAGCGGATGCGCTATAAAACCACTTTAAACCTTTCAGATGCCCTTCACACAGAACCAACCGCGTCAGACTCCACAGGACGGAATAAATCTGTATTCACCTGATTACACGGAATGTCTACACAGGTCATCAAGCTGTGTCTCGAGAGCTGAGCGCGAAGAGGGACCCCGAGTGAAG

[illegible][illegible]

7

NCBI Resources ☒ How To ☒ Sign in to NCBI

ORFfinder

Open Reading Frame Viewer [Help](#)

Sequence

ORFs found: 5 Genetic code: 1 Start codon: 'ATG' only

Find:

(U) ORFfinder_11.26.154554586

1: 1..697 (697bp)

ORF1 (111 nt)

ORF1 CDS translation

```

1 ATGACCAGCCACGGGGTCCCTGCTGTGAGCTCTGGGDDGCTGGCTT
  M T S N G V P A V S S G R C L
46 CCAGGGCTCCCGAGCCACAGCTGGGGTCTGCTGAGGGGAACA
  P G L P S H T L G V L A E G T
91 TGGCTTGTGGGCTCAGCTGA
  M L V G L S *

```

Mark subset... Marked: 1 as

Label	Strand	Frame	Start	Stop	Length (nt aa)
ORF5	-	1	397	>2	396 131
ORF4	+	3	396	>695	300 99
ORF2	+	1	541	>696	156 51
ORF1	+	1	307	417	111 36
ORF3	+	2	461	544	84 27

ORF1

Marked set (1)

[SmartBLAST best hit titles...](#)

BLAST Database:

Upstream ORF discussed within the main text was predicted as ORF1 and known main ORF as ORF4.

Supplemental Figure S7

Translation Initiation Site Prediction

Prediction Results

```
>ABCA1 Human E1 E2 E3 E4
AGAGCACAGGCTTTGACCGATAGTAACCTCTGCGCTCGGTGACGCCGAATCTATAAAAGG
AACTAGTCCCGGCAAAAACCCGTAATTGCGAGCGAGAGTGAGTGGGGCCGGGACCCGCA
GAGCCGAGCCGACCTTTCTCTCCCGGGCTGCGGCAGGGCAGGGCGGGGAGCTCCGCGCAC
CAACAGAGCCGGTTCTCAGGGCGCTTTGCTCCTTGTTTTTCCCGGTTCTGTTTTCTCC
CCTTCTCCGGAAGGCTTGTCAAGGGTAGGAGAAAGAGACGCAACACAAAAGTGGAAA
CAGTTAATGACCAAGCCACGGCGTCCCTGCTGTGAGCTCTGGCCGCTGCCTTCCAGGGCTCCCG
AGCCACACGCTGGGGGTGCTGGCTGAGGGAACATGGCTTGTTGGCTCAGCTGAGGTTGC
TGCTGTGGAAGAACTCAGTTTCAGAAAGAGACAAACATGTCAGCTGCTGCTGGAAGTGGCCT
GGCCTCTATTATCTTCTGATCCTGATCTCTGTTCCGGCTGAGCTACCCACCCTATGAACAACAT
GAATGCCATTTTCCAAATAAAGCCATGCCCTCTGCAGGAACACTTCTTGGGTTGAGGGGATTA
TCTGTAATGCCAACAAACCCCTGTTTCCGTTACCCGACTCCTGGGAGGGCTCCCGAGTTG
TTGGAAACTTTAACAAATCCAT
```

Position	Context	Prediction Score	Predicted
307	AGTTAATGACCAG	1.000484834932	Yes
396	GGAAACATGGCTTG	0.767714554146	---
461	CAAAACATGTCAGC	0.883724617599	---
541	ACCCTATGAACAA	1.000669928285	Yes
550	ACAAACATGAATGC	0.882354680182	---
554	CATGAATGCCATT	1.000033349677	Yes
576	AAGCCATGCCCTC	0.930832456891	---
622	CTGTAATGCCAAC	0.931964887995	---

Designed & Developed By: Suhas Tikole

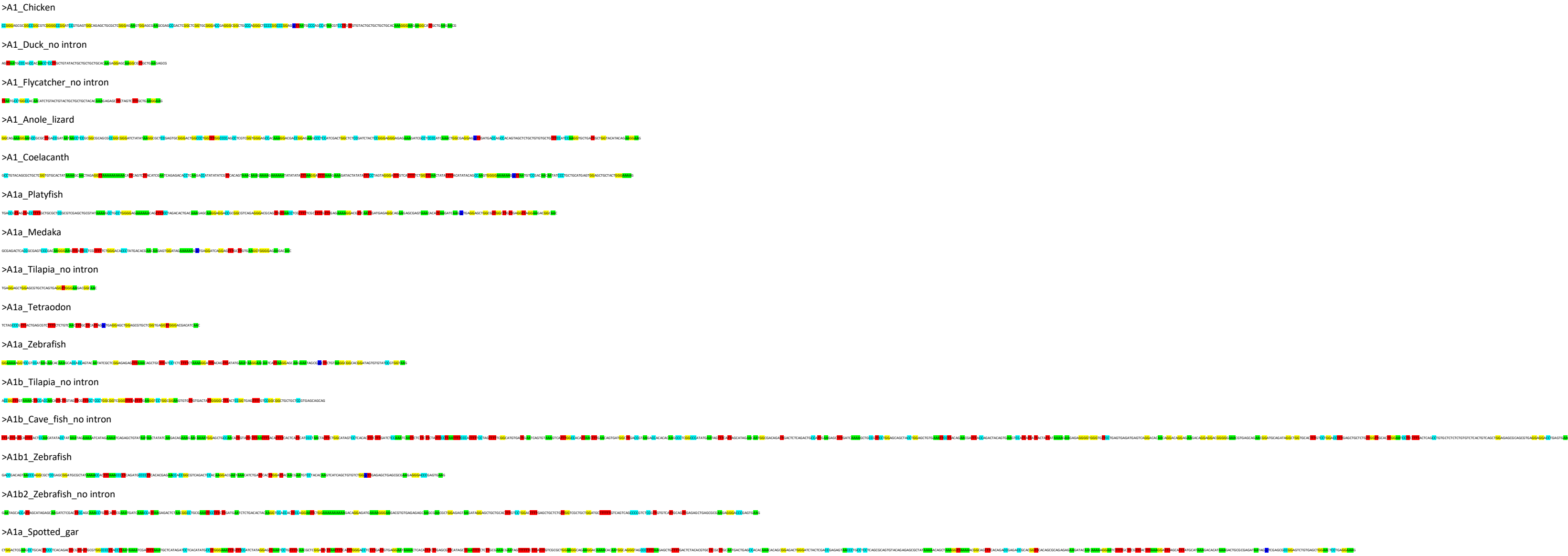
Copyright (c) 2006 All rights reserved, IIT Kanpur.

Laboratory of Computational Biology
Department of Biological Sciences and Bioengineering
Indian Institute of Technology Kanpur -208016
INDIA

Supplemental Figure S7: Results obtained from the WeakAUG server.

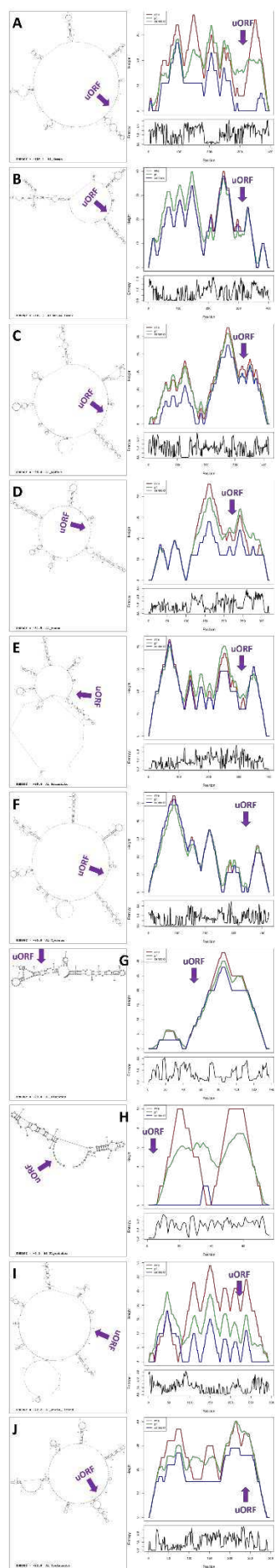
↓ specifies the position of Intron 1

10



Supplemental Figure S8: Analysis of specific base-repetition-rich subregions in 5'UTRs of orthologous *ABCA1* genes, 5'UTR sequences of 55 species were downloaded from the Ensembl database.

Supplemental Figure S9



Supplemental Figure S9: Secondary structure comparison of *ABCA1* 5'UTRs in 10 vertebrate species. Two web servers – RNAstructure (secondary structure diagrams) and RNAfold (mountain plots) were employed independently. Results of the following species are disclosed: A) Human, B) Mouse lemur, C) Rabbit, D) Mouse, E) Armadillo, F) Opossum, G) Platypus, H) Flycatcher, I) Anole lizard, and J) Coelacanth. Start position of the uORF discussed in the main text is marked with an arrow.

Supplemental Figure S10

Gene: ABCA1

Human (GRCh37.p13)

Transcript: ABCA1-002 ENST00000374736.3

5'UTR sequence

TATA TFIIA HOXA13

AGAGCAAGGCTTTGACCGA TAGTAACCTCTGCGCTCGGTGCAGCCGAATCTATAAAAGG

Ighg2 Gh-1 Sc35

AACTAGTCCCGGC~~AAAA~~ACCCGTAATTGCGAGCGAGAGTGAGTGGGGCCGGGACCCGCA

GAGCCGAGCCGACCCTTCTCTCCCGGGCTGCGGCAGGGCAGGGCGGGGAGCTCCGCGCAC

CAACAGAGCCGGTTCTCAGGGCGCTTIGCTCTTGTTTTTTCCCCGGTTCTGTTTTCTC

Ctnt FEE NFAT1/4

CCTTCTCCGGAAGGCTTGTCAAGGGGTAGGAGAAAGAGACGCAAACACAAAAGTGGAAAA

uORF Ighg2 Sc35

CAGTTAATGACCAGCCACGGCGTCCCTGCIGTGAGCTCTGGCCGCTGCCTTCCAGGGCTCCCG

Ctnt Gh-1

AGCCACACGCTGGGGGTGCTGGCTGAGGGAAC

Notes:

5'UTR variants according to Ensembl in blue color

5'UTR hot spots defined by the study in red

Nucleotides matching both criteria in orange

Supplemental Figure S10: Analysis of the human *ABCA1* 5'UTR variants from Ensembl versus hot spots defined by the study.