

Supplementary Tables

Supplementary Table 1 Center contribution

Clones Finished per Center					
Center	Minimum tiling path clones	Additional clones	All clones	Total bases	Accession Numbers
CHGC	39	1	40	6,016,694	BS000165-204
KRIBB	21	0	21	2,810,557	BS000205-225
Germany	47	12	59	4,979,927	AL954200-256, 258-259
NIG	5	2	7	1,050,886	BS000158-164
RIKEN	143	14	157	19,540,921	BS000001-157
YMGC	18	2	20	3,412,759	BS000226-245
Total	273	31	304	42,666,429	AL954200-256, 58-59; BS000001-245

Germany: MPI, IMB and GBF

Center-specific sequencing procedures

CHGC

BAC DNA preparation: QIAGEN Large-Construct Kit Shotgun library: sonication and subcloning into pUC18/SmaI Plasmid DNA preparation: Millipore Multiscreen filter plate Sequencing: Big-Dye terminator on ABI 3700 sequencer, DYEnamic ET terminator on Megabase 1000 and 4000 Assembly and finishing: Phrap/Consed

GBF

DNA fragmentation: HydroShear shotgun vector: pTZ 18R prep: TempliPhi chemistry: dideoxy terminator from Amersham machines: MegaBace 1000 or 4000 (semi automated with Watrex plate feeder, "Caddy"), few finishing runs on Licor slab gel assembly and finishing with pregap/Gap4

IMB

Qiagen BAC prep sonication and subcloning into pUC18/SmaI Qiagen template prep ABI BigDye cycle sequencing and separation on ABI3700 Phred/Phrap assembly manual proofreading & editing using Gap4

KRIBB

BAC prep: Ultra-centrifugation by using CsCl₂ gradient DNA fragmentation: Hydroshear Shotgun vector: pUC118 Plasmid prep: Millipore 96-well prep kit Chemistry: ET (Amersham), BD-Terminator (ABI) Machine: RISA384 or ABI3730 Data management for shotgun data: Phred/Phrap assembling Data management for finishing data: Sequencher V4.14

MPI

BAC DNA was isolated by alkaline lysis and purified on CsCl by standard procedures. For subcloning DNA was sonicated, fragment ends polished with T4 and Klenow polymerase, size selected for 1.5 and 3.5 kb,

ligated in pUC19/SmaI, transformed into E. coli DH10B. Plasmid prep with Millipore kit. Inserts of the libraries were amplified by PCR as templates for sequencing. Sequencing was performed using Big Dye chemistry, M13 primers and ABI 3700 capillary sequencers resulting in more than 10-fold coverage. All raw sequences were processed by PHRED, controlled for vector or E. coli contamination and assembled by Phrap. Analyzed regions were manually edited in GAP4.

NIG

Qiagen BAC prep., DNA fragmentation by HydroShear, Subcloned by pUC118, PCR by M13F&M13R primers, Sequencing by BigDye terminator and ABI3700, PCR and TA cloning for finishing, phredPhrap assembling, manual editing by consed.

RIKEN

BAC DNA was purified through EtBr/CsCl equilibrium centrifugation, then fragmented by shearing (Hattori *et al. Nature* **405**, 311-319, 2000).

YMGC

BAC DNA was isolated by alkaline lysis and followed by phenol/Chloroform extraction. The DNA was then sheared by Hydroshear and ends polished with T4 DNA polymerase and Bal 31nuclease. Fragment sizes ranging from 2.5 to 3.5 kb were selected and ligated to pUC18/SmaI followed by transformation to E. coli DH5a. Plasmid DNA was isolated by alkaline lysis and purified by MultiScreen-HV filter (Millipore). Plasmid DNA was sequenced by Big Dye version 1 and analyzed by ABI 3700. For each BAC clone, at least 10-fold coverage was sequenced and the vector and E. coli contamination were below 15%. The sequencing data were processed and assembled by Phred/Phrap/Consed.

Supplementary Table 2. Summary of the chimpanzee22 sequence quality.

(See SupplementaryTable2 in separate file)

Supplementary Table 3. Comparative gene catalogue.

(See hyperlinked Supplementary Table3 file)

Supplementary Table 4 Genes with internal amino acid insertions or deletions in chimpanzee

Human Gene references	Genomic changes in PTR22q	AA indels in chimp	Mouse ortholog
<i>IFNAR1</i> Interferon-alpha receptor: A26593, A26595, A32389, A32391, BC021825, J03171	ins (CCT)	p.P148_G149insP	NP_034638 no insertion
<i>IFNAR2</i> Interferon alpha/beta receptor: L41942, X89772, X77722, L41944	ins (GAA)	p.L443_E444insE L41942, X89772	NP_034639 divergent structure in this region
<i>C21orf2</i> Nuclear encoded mitochondrial protein, cDNA A2-YF5: BC031300, Y11392, U84569, Z93322	ins (GAG)	p.G323_L324insL BC031300	BAB23134 divergent structure in this region
<i>MCM3AP</i> MCM3 import factor: AJ010089, AB005543, BC013285, BC032750	ins (GAT)	p.V1000_S1001insI AJ010089	NP_062307 Insertion I
<i>ANKRD3</i> ankyrin like, dual-specificity Ser/Thr/Tyr kinase domain: AB047783, AJ278016, AK027424	ins (AGACAC)	p.P315_A316insVS	AAG30871 no insertion
<i>USP16</i> Ubiquitin processing protease, EC3.1.2.15: AF113219, AK023247, AF126736, AK025104	ins (ACTGACTGT)	p.P374_T375insTDC : AF113219, AK023247 p.P553_T554insTDC : AF126736, AK025104	NP_077220 Insertion. EC
<i>ETS2</i> Erythroblastosis virus oncogene homolog 2: AK096841, BC017040, J04102	ins (CCCTCGCCCTCG)	S116_P117insPSPS AK096841	AAA37581 divergent structure in this region
<i>KRTAP10-10</i> Hair keratin-associated protein 10.10: AJ566387	ins (TGCTGCGCCCCCAGC)	P34_A35insSCCAP	no clear ortholog
<i>C21orf45</i> Unknown function: AF231921, AF387845	del (CTC)	p.E72del	AAF72945 No deletion
<i>TRPM2</i> Transient receptor potential-related channel 7, a novel putative Ca ²⁺ channel protein: AB001535	del (GAG)	p.E15del	NP_036165 divergent structure in this region
<i>SLC19A1</i> Reduced folate carrier U19720, AF004354, U15939, U17566	del (TGG)	p.P234del	AAC53287 No deletion

<i>C21orf22</i> Unknown function AY040089	del (TGCAGC)	p.A44_A45del	No ortholog
<i>COL18A1</i> Human type XVIII collagen AF018081, AF018082	ins (GGCCCCCCC) del (GGCCCCCCA)	p.P1176_S1177ins (GPP) p.G1125_P1127del (GPP)	NP_034059 no insertion del (GPPGPR)
<i>TCP10L</i> * T-complex protein 10A-2 AF115967, AK058078, BC022024	del (ACAAAGATCGTCATCTA) corresponds to a duplication of 17 bp in HSA21	p.F156fsS168	No ortholog
<i>PCNT2</i> *; Pericentrin, kendrin U52962	del (195 bp)	p.128Q_E192del (65aa)	AAA17886 Missing exon

Accession numbers in bold refer to the transcript isoforms predicted to be modified in chimpanzee. Coordinates of the genomic changes in the PTR22q sequence, and nucleotide/amino acid sequence alignments are given in supplementary Table 3. Genes with an asterisk are described in the text.

Supplementary Table 5 Genes with altered start or stop codons in chimpanzee

Human Gene references	Genomic Changes in PTR22q	Modified START	Modified STOP	Mouse ortholog
<i>C21orf9</i> Unknown function: AY077697	g.29579632G>C	p.M1_L15del		No ortholog
<i>C21orf122</i> Unknown function: NM_032653	g.44721800T>G	p.M1?		No ortholog
<i>C21orf86</i> Unknown function: AF426264	g.44946998T>C	p.M1_V34del		No ortholog
<i>KRTAP23-1</i> human hair keratin-associated protein 23.1: (KAP ref)	g.30149103C>T g.30149105T>G	p.M1?		No clear ortholog
<i>C21orf97</i> Unknown function: AK024977 BC003651	g.43399866-43399867del (C)	p.M1_D14del		No ortholog
<i>DSCR6</i> Unknown function: AB037158 AB037159	g.36751787C>T		Premature STOP: p.R162X AB037158 p.R78X AB037159	BAB60891 divergent structure in this region
<i>PSMD15</i> Proteasome 26S subunit gene: AF050199	g.36229865C>T		Premature STOP: p.R183X	NP_032977 divergent structure in this region
<i>C21orf118</i> Unknown function: AF304442	g.26044865C>A		Premature STOP: p.S36X	No ortholog
<i>C21orf128</i> Unknown function: NM_152507	g.41815522A>T		Premature STOP: p.Y33X	No ortholog
<i>KRTAP15-1</i> human hair keratin-associated protein 15.1: (KAP ref)	g.30216795C>T		Premature STOP : p.Q72X	No clear ortholog
<i>IGSF5</i> putative gene, immunoglobulin superfamily 5 like: AK092516	g.39483930C>T		Premature STOP: p.Q160X	NP_082354 The mouse and the human proteins have a stop at the same position but have a different methionine.
<i>ABCC13</i> putative gene, multidrug resistance associated protein like: AF418600 AF518320 AY063514 AY063515	g.14356247ins (G) AF418600 AY063514 AY063515 g.14385690T>C g.14385691G>		Frameshift + premature STOP: p.R16fsX20 AF418600, AY063515, AY063514	AAB80938 divergent structure in this region

	AF518320		p.Q180X AF518320	
<i>C21orf124</i> Unknown function: AK056502 BC008008 BC021550	g.43375146_43375147del (A)		Frameshift + premature STOP: p.S72fsX80	No ortholog
<i>C21orf90</i> Unknown function: NM_153204	g.44162879_44162880del (C)		Frameshift + premature STOP: p.G23fsX26	No ortholog
<i>KRTAP19-4</i> human hair keratin-associated protein 19.1: (KAP ref)	g.30265157_30265158ins (GA)		Frameshift + premature STOP: p.Q61fsX64	No clear ortholog
<i>ICOSL</i> Unknown function: AB014553	g.43879282_43879283del (CGCGGAGACCTC GGGG)		Frameshift + premature STOP: p.P35fsX66	AAF34738 divergent structure in this region
<i>C21orf121</i> Unknown function: NM_032653	g.41735691_41735692del (AGTCCACCGTCG TCTTCTAGCCCCA CCATCGTCGTCT)		Frameshift + premature STOP: p.I30fsX102	No ortholog
<i>C21orf74</i> Unknown function: AY077696	g.22190917- 22190918del (1287 bp)		Frameshift + premature STOP: p.T29fsX49	No ortholog
<i>C21orf114</i> Unknown function	g.17810821ins (A)		Frameshift + premature STOP: p.I17fsX37	No ortholog
<i>C21orf79</i> Unknown function	g.26054348ins (T)		Frameshift + premature STOP:p.V29fsX3 9	No ortholog
<i>C21orf71</i> Unknown function: AF086441	g.25411102T>G		STOP deletion : p.X84CinsA84+1_ X+21	No ortholog
<i>C21orf30</i> Unknown function: AL117578	g.44105796T>C g.44105797G>C		STOP deletion: p.X248PinsC248+ 1_X+20	No ortholog
<i>LIP1</i> putative gene, lipase (EC 3.1.1.3) like: BC028732	g.14247293_14247294del (A)		Frameshift and use of another STOP: p.L271fsX294	BB663289 divergent structure in this region
<i>C21orf70</i> Unknown function: AF391113	g.44590485_44590486del		Deletion	AAL34506

AF391114 BC009341	86del (corresponding to a duplication of 40nt in the human)		p.E81_G95del (EAGSSRSVPSI RRG) AF391113 BC009341	The duplication is also absent in the mouse
<i>LSS</i> human lanosterol synthase, EC 5.4.99.7.: AK092334 D63807 S81221 U22526	g.45868808_45868809ins(TA)		Frameshift + premature STOP: p.R20fsX35 AK092334	XP_109587 divergent structure in this region
<i>PDE9A</i> CGMP-specific 3', 5'-cyclic phosphodiesterase type 9, EC 3.1.4.17.: AF048837 AF067223 AF067224 AF067225 AF067226 BC009047	g.42390873C>T		Premature STOP: p.R38X AF067226	AAC24344 divergent structure in this region
<i>ABCG1</i> white protein homolog (ATP-binding cassette transporter 8): AY048757 X91249	g.41964344_41964345del(A)		Frameshift + premature STOP: p.V216fsX247 AY048757	AAB47738 divergent structure in this region
<i>FAM3B</i> Unknown function: AF375989 AJ409094	g.41022241_41022242del(TTTGTTGGTT)		Frameshift + premature STOP: p.C29fsX31 AJ409094	BAB31283 divergent structure in this region
<i>DSCR5</i> human Down syndrome critical region protein C: AB035742 AB035743 AB035744 AB035745 AB037162 AB037163 AB037164 AF216305 BC011007	g.36807416ins (C)		Frameshift + premature STOP: p.T9fsX20 AB035745 AB037163	NP_062416 divergent structure in this region
<i>AIRE</i> autoimmune regulator (APECED protein): AB006682 AB006683 AB006685 Z97990	g.43928300A>G	p.M1_V25del AB006683 AB006685		CAB66141 divergent structure in this region
<i>D21S2056E</i> human NNP-1/Nop52 (NNP-1), novel nuclear protein 1: AY033999 BC014787 U79775	g.43432783G>A	p.M1_I17del AY033999		NP_035055 divergent structure in this region
<i>C21orf66</i> Unknown function: AF231920 AY033903 AY033905	g.32483288ins (C) g.32483291ins (A)		STOP deletion: p.X816Lins?	BAB27645

AY033906 BC030539 HSA279080 HSA279081 AF153208 AY033904			AY033904 p.X248LinsE248+ 1_X+2 AF153208	divergent structure in this region
--	--	--	---	---------------------------------------

Supplementary Table 6 Genes with extreme Ka/Ks values

Group	Group character	Minimum Ka/Ks value	Gene name (Hs)	LocusLink ID (Hs)	Description
Ka/Ks>1		3.37	KRTAP23-1	337963	keratin associated protein 23-1
		2.78	C21orf87	257357	chromosome 21 open reading frame 87
		1.98	C21orf81	114035	chromosome 21 open reading frame 81
		1.79	C21orf128	150147	hypothetical protein
		1.76	C21orf119	84996	chromosome 21 open reading frame 119
		1.73	RPS5L	54022	ribosomal protein S5-like
		1.71	PRED62		
		1.67	KRTAP15-1	254950	keratin associated protein 15-1
		1.57	KRTAP21-1	337977	keratin associated protein 21-1
		1.47	ABCC13	150000	ATP-binding cassette, sub-family C (CFTR/MRP), member 13
		1.45	C21orf94	246705	chromosome 21 open reading frame 94
		1.42	C21orf93	246704	chromosome 21 open reading frame 93
		1.40	C21orf86	257103	chromosome 21 open reading frame 86
		1.37	ANKRD21	317754	Expressed in prostate, ovary, testis, and placenta
		1.34	C21orf111	378823	chromosome 21 open reading frame 111
		1.16	C21orf129	150135	hypothetical protein
		1.15	TMPRSS2	7113	transmembrane protease, serine 2
		1.10	C21orf126	84210	hypothetical protein
		1.08	C21orf22	54089	chromosome 21 open reading frame 22
		1.04	C21orf115	378827	chromosome 21 open reading frame 115
1.01	DSCR6	53820	Down syndrome critical region gene 6		
1.01	PRED61				
(Nd + Sd) >= 10	relatively rapidly	0.56	KRTAP13-3	337960	keratin associated protein 13-3

and $(K_a + K_s) / 2 > 2\%$ ($> 1.44\%$)	evolving genes	0.56	KRTAP6-3	337968	keratin associated protein 6-3
		0.27	KRTAP19-7	337979	keratin associated protein 19-7
		0.32	KCNE1	3753	potassium voltage-gated channel, Isk-related family, member 1
		0.19	ATP5J	498	ATP synthase, H ⁺ transporting, mitochondrial F1 complex, alpha subunit, isoform 1, cardiac muscle
		0.35	UMODL1	89766	uromodulin-like 1
		0.72	TCP10L	140290	t-complex 10 (mouse)-like
		0.34	B3GALT5	10317	UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase, polypeptide 5
		0.54	IGSF5	54046	immunoglobulin superfamily, member 5
p-value < 0.05 , $(K_a + K_s) / 2 > 2\%$	evolving under purifying selection	0.55	KRTAP13-4	284827	keratin associated protein 13-4
		0.00	PFKL	5211	phosphofructokinase, liver
		0.08	C21orf113	378825	chromosome 21 open reading frame 113
		0.27	COL18A1	80781	collagen, type XVIII, alpha 1
		0.19	AIRE	326	autoimmune regulator (autoimmune polyendocrinopathy candidiasis ectodermal dystrophy)
		0.05	ITGB2	3689	integrin, beta 2 (antigen CD18 (p95), lymphocyte function-associated antigen 1; macrophage antigen 1 (mac-1) beta subunit)
		0.06	TMPRSS3	64699	transmembrane protease, serine 3
		0.05	AGPAT3	56894	1-acylglycerol-3-phosphate O-acyltransferase 3

Nd: # of nonsynonymous substitutions; Sd: # of synonymous substitutions

Supplementary Table 7 Genes on human chromosome 21 that show significant expression differences between humans and chimpanzees

Gene ID	Transcript ID	brain	liver	lineage-specific changes	
				brain	liver
C21orf18	NM-017438	higher in chimp	not expressed	chimp	
C21orf33	Y07572	not expressed	higher in human		nonspecific
C21orf5	AJ237839	no difference, but expressed	higher in human		chimp
C21orf97	BC003651, AK024977	higher in human		nonspecific	
CRYAA	U05569, U66584	not expressed	higher in chimp		chimp & human
CRYZL1	AK001293	higher in human		?	
CXADR	Y07593	no difference, but expressed	higher in human		human
DSCR1	U85267	no difference, but expressed	higher in human		human
ETS2	J04102	higher in chimp	no difference, but expressed	chimp & human	
IFNAR2	X77722	higher in chimp	higher in chimp	chimp	chimp & human
IFNGR2	BC003624	higher in chimp		chimp & human	
ITSN1	U61166	no difference, but expressed	higher in chimp		nonspecific
LSS	AK092334	no difference, but expressed	higher in human		nonspecific
PDXK	BC000123	no difference, but expressed	higher in chimp		chimp
PTTG1IP	Z50022	no difference, but expressed	higher in human		chimp & human
TTC3	D84296	higher in human	higher in human	nonspecific	nonspecific
USP16	AK025104	higher in human		nonspecific	
ZNF294	AB018257	higher in human	higher in human	chimp	chimp
		higher in chimp			
		higher in human			
		not expressed			
		no difference, but expressed			
		not on chip			

The assignment of lineage-specific differences was done using the expression values from one orangutan. If the expression measured for the orangutan differed significantly (t-test $p \leq 0.05$) from chimpanzee, but not from human then we assume that the change occurred on the chimpanzee lineage and if the orangutan differed significantly from human but not from chimpanzee then the change occurred on the human lineage.

Supplementary Table 8 Ranking of all genes that were expressed in either brain or liver

	# of genes with expression		mean divergence		p-Value*
	different	not different	different	not different	
non-degenerate sites	18	47	0.0060	0.0043	0.2150
3'UTR	18	47	0.0100	0.0090	0.1240
5'UTR	18	46	0.0115	0.0085	0.023*
CpG-island	12	40	0.0161	0.0135	0.0590
intron	17	46	0.0129	0.0125	0.1290
3' intergenic (10kb)	18	48	0.0129	0.0129	0.9310
5' intergenic (10kb)	18	48	0.0131	0.0127	0.6970
5' intergenic (1kb)	18	48	0.0140	0.0132	0.4080

* Mann-Whitney U -test, two-tailed significance

Calculation is according to the nucleotide divergence at non-degenerate sites, 3'UTR, 5'UTR, CpG-island, introns, 3' and 5' intergenic regions and tested whether the genes with differing mRNA levels between humans and chimpanzees are non-randomly distributed within that list.

Supplementary Table 9 Nucleotide substitution patterns in human and chimpanzee

(A) Proportion of nucleotide substitution in the human genome based on SNP data

		NEW			
		A	T	C	G
OLD	A	-	2.88 (703)	3.60 (880)	14.00 (3422)
	T	2.80 (681)	-	15.05 (3654)	3.45 (838)
	C	4.43 (747)	20.31 (3424)	-	4.51 (760)
	G	19.56 (3292)	4.53 (762)	4.88 (822)	-

Note. An element at the i-th row and j-th column means nucleotide substitution from nucleotide i to j. Figures designate standardized proportion of substitution in %, while figures in parentheses are numbers of mutations observed out of 19,985 SNP loci.

(B) Proportion of nucleotide substitutions in the chimpanzee genome based on BAC-overlap SNP data

		NEW			
		A	T	C	G
OLD	A	-	3.55 (252)	3.68 (261)	12.58 (893)
	T	2.92 (206)	-	13.18 (930)	4.02 (284)
	C	5.08 (250)	19.22 (946)	-	5.24 (258)
	G	20.83 (1024)	4.37 (215)	5.33 (262)	-

Note. An element at the i-th row and j-th column means nucleotide substitution from nucleotide i to j. Figures designate the standardized proportion of substitution in %, while figures in parentheses are numbers of mutations observed out of 5,781 SNP loci.

(C) Nucleotide transition probability matrices for human and chimpanzee.

	Human				Chimpanzee			
	A	T	C	G	A	T	C	G
A	0.950	0.007	0.008	0.034	0.985	0.002	0.003	0.009
T	0.007	0.948	0.036	0.008	0.002	0.985	0.010	0.003
C	0.012	0.050	0.929	0.011	0.004	0.014	0.978	0.004
G	0.049	0.011	0.011	0.930	0.015	0.004	0.004	0.978

(D) Equilibrium frequencies of four nucleotides based on the nucleotide transition probability matrices for human and chimpanzee shown above. Tajima and Nei (1982)'s method was used. Figures in parentheses are observed values.

	Human	Chimpanzee
A	0.290 (0.297)	0.303 (0.296)
T	0.288 (0.295)	0.293 (0.294)
C	0.214 (0.205)	0.205 (0.205)
G	0.208 (0.204)	0.200 (0.205)
GC content	0.422 (0.409)	0.405 (0.410)

<References sited in this table>

Tajima, F. & Nei, M. Biases of the estimates of DNA divergence obtained by the restriction enzyme technique. *J. Mol. Evol.* **18**, 115-120 (1982).

Supplementary Table 10 List of 18 DNA regions that showed statistically significant results based on a Modified H test suggesting positive selection (bin size = 10kb)

Position	$P(H)$	S	Gene name; description (in any)
16299112	0.0430	2(0)	CHODL; A type I membrane protein with a carbohydrate recognition domain characteristic of C-type lectins in its extracellular portion
17880261	0.0320	2 (0)	Noncoding region
18735154	0.0320	1 (0)	Noncoding region
20045321	0.0270	2 (0)	Noncoding region
20724640	0.0070	8 (0)	Noncoding region
20794640	0.0340	5 (0)	Noncoding region
21949854	0.0450	31 (0)	Noncoding region
25259390	0.0480	7 (0)	Noncoding region
27461495	0.0100	4 (0)	Noncoding region
28230953	0.0450	3 (0)	Noncoding region
30396956	0.0370	4 (0)	C21orf108; Chromosome 21 open reading frame 108
32522462	0.0430	4 (0)	KCNE; Potassium voltage-gated channel, Isk-related family, member 1
37213005	0.0210	3 (0)	DSCR2; Down syndrome critical region gene 2
37694053	0.0220	4 (2)	B3GALT5; UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase, polypeptide 5; Exon 1: Synonymous difference (ACG/ACA); Exon 3: Difference in 3' UTR
40131716	0.0250	3 (0)	Noncoding region
40243527	0.0340	7 (0)	Noncoding region
41688508	0.0480	4 (1)	KIAA0179; KIAA0179 protein; Exon 7: Synonymous difference (GCG/GCA)
42533280	0.0380	2 (0)	C21orf90, Chromosome 21 open reading frame 90; C21orf29, Chromosome 21 open reading frame 29

Note. Position: the mid-point of a bin (10 kb for the alignment) along HSA21q. S: the number of segregating sites. The number of segregating sites in exons are in parentheses, and the SNP information is given when it resides in an exon. $P(H)$: the probability of observing a more negative H value under neutrality (significance level = 5%, one sided). If an SNP locus was detected with both maximum and minimum sample sizes, the smaller $P(H)$ was shown.

Supplementary Table 11 Computational analysis of the transcription factor binding sites

Gene	All Transcription factor binding sites											
	Human				Chimp				Mouse			
	name	pos	str	score	name	pos	str	score	name	pos	str	score
C21orf18 not in liver higher in C brain	Oct-1	13	-	0.995	Oct-1	10	-	0.995	Pax	17	+	0.889
	RFX1	27	+	0.827	RFX1	24	+	0.827	GATA-4	100	+	0.904
	TFIIA	102	+	0.961	GATA-4	98	+	0.935	C/EBPdelta	231	+	0.974
	PPAR	111	-	0.893	TFIIA	99	+	0.961	CREB	232	-	0.985
	SRF	247	+	0.892	PPAR	108	-	0.893	GATA-4	278	-	0.938
	AML1	273	-	1	AML1	273	-	1	Myc	328	-	1
	RFX1	400	+	0.887	RFX1	400	+	0.887	Myc	329	+	1
	COMP1	546	+	0.831	c-Rel	596	-	0.99	Pax-6	519	+	0.812
	c-Rel	596	-	0.99	Cdc5	639	+	0.846	Nrf-1	993	-	1
	Cdc5	639	+	0.846	Nrf-1	810	-	1				
	Nrf-1	810	-	1	Myc	980	-	1				
	Elk-1	820	+	0.988	USF	980	+	1				
	Muscle initiator sequence-20	884	+	0.948	Myc	981	+	1				
	Myc	980	-	1								
	USF	980	+	1								
Myc	981	+	1									
C21orf33 not in brain higher in H liver	HNF-3alpha	17	-	0.971	FOXD3	124	-	0.954	Hand1/E47	66	-	0.961
	FOXD3	118	-	0.954	E2F-1	372	-	0.922	Oct-1	70	+	0.948
	E2F-1	366	-	0.922	E2F-1	376	+	0.942	MAZ	100	-	1
	E2F-1	370	+	0.942	E2F-1	377	+	0.982	NF-kappaB	256	+	0.914
	E2F-1	371	+	0.982	NF-kappaB	408	+	0.923	Pax-6	460	-	0.832
	NF-kappaB	402	+	0.923	Elk-1	861	+	0.982	Pax-3	463	-	0.928
	Elk-1	860	+	0.982	STATx	882	+	1	v-Myb	576	-	0.981
	STATx	881	+	1	CP2	941	-	0.976	HNF-1	660	-	0.933
CP2	941	-	0.976					SRF	947	+	0.818	
C21orf5 H,C equal brain higher in H liver	COMP1	5	+	0.874	COMP1	5	+	0.874	AML1	87	+	1
	Xvent-1	128	+	0.962	Xvent-1	128	+	0.962	SREBP-1	515	+	1
	AML1	159	+	1	AML1	159	+	1	SMAD-3	556	+	1
	AML1	262	+	1	AML1	262	+	1	APOLYA	863	+	0.966
	LEF-1	387	-	0.956	LEF-1	387	-	0.956	CDP	873	+	0.925
	GATA-4	522	+	0.923	GATA-4	522	+	0.923				
	Staf	524	-	0.94	Staf	524	-	0.94				
	Hand1/E47	528	+	0.972	Hand1/E47	528	+	0.972				
	Pax	536	+	0.887	Pax	536	+	0.887				
	Nkx2-5	703	-	1	Nkx2-5	703	-	1				

	ER	721	+	0.96	ER	721	+	0.96					
	Pax-6	726	-	0.824	Pax-6	726	-	0.824					
	HNF-4	882	-	0.91	HEB	853	-	1					
	Pax	897	+	0.856	HNF-4	882	-	0.91					
					Pax	897	+	0.856					
CRYAA	HP1site factor	67	+	0.944	HP1site factor	65	+	0.944	SREBP-1	47	-	1	
not in brain	RFX1	209	+	0.821	RFX1	207	+	0.835	c-Ets-168	60	-	1	
higher in C liver	C/EBPdelta	235	+	0.971	ATF4	220	-	0.962	ER	395	+	0.96	
	Myc	332	-	1	CREB	220	+	0.982	AML1	431	-	1	
	Myc	333	+	1	CRE-BP1	220	-	0.972	FOXD3	534	+	0.957	
	SMAD-4	773	-	0.944	HLF	221	+	0.93	HNF-3alpha	537	+	0.985	
	HEN1	809	+	0.975	CREB	222	-	0.983	FOX	561	-	0.939	
	HEN1	809	-	0.982	Myc	330	-	1	FOXD3	562	-	0.966	
	v-Maf	893	+	0.976	Myc	331	+	1	HNF-3alpha	564	-	0.965	
	Nrf2	897	-	0.954	SMAD-4	773	-	0.944	FOX	565	-	0.921	
					HEN1	809	+	0.975	HNF-3alpha	571	-	0.962	
					HEN1	809	-	0.982	HNF-3alpha	575	-	0.965	
					v-Maf	893	+	0.976	FOX	576	-	0.937	
					Nrf2	897	-	0.954	FOXD3	577	-	0.96	
									SRF	725	+	0.834	
									v-Maf	890	+	0.946	
									NF-kappaB	932	+	0.97	
CRYZL1	MAZ	218	-	1	MAZ	94	+	1	E2F	601	+	0.989	
higher in H brain	GATA-4	277	-	0.901	MAZ	215	-	1	E2F	601	+	0.994	
	Evi-1	284	-	0.846	GATA-4	274	-	0.901	E2F	601	+	1	
	SRF	340	+	0.833	Evi-1	281	-	0.846	Rb/E2F-1/DP-1	601	+	1	
	SRF	433	+	0.859	SRF	337	+	0.833	E2F-1	601	+	0.972	
	NF-Y	482	+	0.995	Xvent-1	429	-	0.932	E2F-1	602	+	0.97	
	CCAATbox	483	+	0.994	SRF	430	+	0.909	E2F	602	+	0.977	
	Nkx2-5	657	+	1	SRF	431	+	0.941	HNF-4	612	-	0.892	
	AML1	768	+	1	NF-Y	479	+	0.995	OCT-x	768	+	0.99	
	Oct-1	782	-	0.936	CCAATbox	480	+	0.994	Oct-1	770	-	0.984	
	myogenin/NF-1	832	+	0.772	SREBP-1	532	+	1					
					Nkx2-5	654	+	1					
					AML1	768	+	1					
					Oct-1	782	-	0.936					
					myogenin/NF-1	832	+	0.775					
CXADR	AML1	193	-	1	AML1	193	-	1	IRF-7	64	-	0.942	
higher in H liver	SMAD-4	245	-	0.938	SRF	389	+	0.867	SRF	143	-	0.913	

equal in brain	SRF	389	+	0.867	NF-kappaB	586	-	0.908	SRF	149	+	0.842
	NF-kappaB	586	-	0.908	E2F-1	618	-	0.95	Cdc5	430	-	0.881
	E2F-1	618	-	0.95	E2F	663	+	0.962	Pax-4	584	-	0.924
	E2F	663	+	0.962	Cdc5	678	+	0.86	AML1	599	-	1
	Cdc5	678	+	0.86	RFX1	820	+	0.981				
	RFX1	820	+	0.981	RFX1	820	-	0.988				
	RFX1	820	-	0.988	Elk-1	880	+	0.994				
	Elk-1	880	+	0.994	c-Ets-1(p54)	883	+	0.995				
	c-Ets-1(p54)	883	+	0.995	Pax-4	886	+	0.847				
Pax-4	886	+	0.847									
DSCR1	AP-1	131	+	0.993	Xvent-1	68	-	0.938	Oct-1	151	+	0.924
higher in H liver equal in brain	Lentiviral Poly A	188	-	1	AP-1	128	+	0.993	HNF-1	225	-	0.884
	Pax-6	263	+	0.809	Lentiviral Poly A	185	-	1	XFD-2	248	+	0.992
	SRF	281	+	0.818	Pax-6	260	+	0.809	TTF1	278	+	0.979
	GCM	309	+	0.971	SRF	278	+	0.818	Xvent-1	284	+	0.936
	Osf2	324	+	1	GCM	306	+	0.971	Cdc5	362	+	0.84
	AML1	324	-	1	AML1	321	-	1	Cdc5	365	-	0.868
	SREBP-1	352	+	1	Osf2	321	+	1	HEB	489	+	1
	HEB	395	-	1	SREBP-1	349	+	1	TGIF	492	+	1
	TGIF	406	+	1	TGIF	403	+	1	Xvent-1	532	-	0.934
	HNF-4	431	-	0.926	HNF-4	428	-	0.926	SRF	533	+	0.819
	USF	497	-	0.971	USF	494	-	0.971	Pax-4	552	+	0.833
	SRF	512	-	0.822	SRF	509	-	0.822	USF	582	+	0.997
	Lentiviral Poly A	520	-	1	Lentiviral Poly A	517	-	1	USF	582	-	0.997
E2F-1	651	-	0.941	E2F-1	651	-	0.941	SRF	658	+	0.82	
E2F-1	955	+	0.982	E2F-1	955	+	0.982	E2F-1	952	+	0.982	
ETS2	v-Myb	103	+	0.966	v-Myb	103	+	0.966	SRF	41	-	0.84
higher in C brain equal in liver	core-binding factor	111	-	1	core-binding factor	111	-	1	Pax-3	203	+	0.93
	AML1	113	-	1	AML1	113	-	1	Pax-6	206	+	0.815
	ATF6	193	+	1	ATF6	193	+	1	MAZ	833	-	1
	NF-kappaB	378	-	0.923	NF-kappaB	378	-	0.923	MAZ	854	-	1
	R	404	-	0.985	R	404	-	0.985				
	c-Ets-168	444	+	1	c-Ets-168	444	+	1				
	SRF	446	-	0.876	SRF	446	-	0.876				
	Myc	701	-	1	Myc	701	-	1				
	USF	701	-	1	USF	701	-	1				
	Myc	702	+	1	Myc	702	+	1				
	Elk-1	789	+	0.982	Elk-1	789	+	0.982				
	MAZ	974	-	1	MAZ	974	-	1				

IFNAR2 higher in C brain higher in C liver	MEF-1	524	+	0.921	MuscleTATA box	564	-	0.945	Imperfect Hoagness/Gold Major T- antigen Pax	90	+	1	
	MuscleTATA box	564	-	0.945	HNF-1	570	-	0.871		121	+	1	
	HNF-1	570	-	0.871	XFD-3	704	+	0.976		173	-	1	
	XFD-3	704	+	0.976						AML1	306	-	1
										GATA-4	561	+	0.907
									MAZ	582	+	1	
									GCM	770	+	0.984	
									RFX1	778	+	0.827	
									Pax	814	-	0.863	
									Cdc5	879	-	0.85	
IFNGR2 higher in C brain	C/EBPdelta	147	-	0.973	C/EBPdelta	146	-	0.973	AML1	141	+	1	
	Nkx2-5	189	+	1	Nkx2-5	188	+	1	SREBP-1	153	-	1	
	Pax-4	239	-	0.908	Pax-4	238	-	0.908	AML1	203	-	1	
	C/EBPgamma	403	-	0.931	C/EBPgamma	402	-	0.931	MEF-1	276	-	0.886	
	Evi-1	464	-	0.82	Evi-1	463	-	0.82	NKX3A	322	-	0.991	
	POU1F1	480	+	0.98	POU1F1	479	+	0.98	AML1	345	+	1	
	C/EBPgamma	534	+	0.938	NF-kappaB	848	+	0.899	PPAR	445	+	0.937	
	NF-kappaB	848	+	0.899	Pax	879	+	0.881	Cdx-2	495	+	0.894	
	Pax	879	+	0.881	MAZ	897	-	1	SRF	844	+	0.848	
	MAZ	897	-	1									
ITSN1 higher in C liver equal in brain	GATA-4	13	-	0.932	GATA-4	10	-	0.932	Pax-4	19	+	0.833	
	Evi-1	20	-	0.839	Evi-1	17	-	0.839	HNF-4	41	+	0.893	
	myogenin/NF- 1	411	-	0.772	myogenin/NF- 1	408	-	0.775	Oct-1	371	+	0.984	
	Oct-1	475	+	0.936	Oct-1	472	+	0.936	OCT-x	373	-	0.99	
	AML1	498	-	1	AML1	495	-	1	HNF-4	529	+	0.892	
	Nkx2-5	607	-	1	Nkx2-5	607	-	1	E2F	539	-	0.977	
	NF-Y	777	-	0.995	SREBP-1	730	-	1	E2F-1	539	-	0.97	
	CCAATbox	777	-	0.994	CCAATbox	777	-	0.994	E2F-1	540	-	0.972	
	SRF	821	-	0.859	NF-Y	777	-	0.995	Rb/E2F-1/DP- 1	540	-	1	
	SRF	914	-	0.833	SRF	821	-	0.909	E2F	540	-	1	
	Evi-1	973	+	0.846	SRF	823	-	0.941					
	GATA-4	983	+	0.901	Xvent-1	827	+	0.932					
					SRF	914	-	0.833					
				Evi-1	973	+	0.846						
				GATA-4	983	+	0.901						
LSS higher in H liver equal in brain	AML1	145	+	1	AML1	208	+	1	Myc	33	-	1	
	MAZ	270	-	1	MAZ	334	-	1	USF	35	-	0.976	
	E2F-1	546	+	0.911	E2F-1	610	+	0.911	MuscleTATA box	244	+	0.967	

	AML1	796	+	1	AML1	823	+	1	C/EBPgamma	494	+	0.949
	Pax-3	813	+	0.951	Pax-3	840	+	0.947	TFIIA	585	+	0.961
	Pax-4	816	+	0.839	Pax-4	843	+	0.839	AML1	663	+	1
	NF-Y	837	-	0.986	NF-Y	864	-	0.986	Osf2	663	-	1
	CCAATbox	838	-	0.981	CCAATbox	865	-	0.981	core-binding factor Pax-6	665	+	1
	Egr-2	845	+	1	Egr-2	872	+	1	SF-1	691	-	0.829
	Egr-1	845	+	0.999	Egr-1	872	+	0.999	HNF-3alpha	758	-	1
	E2F-1	903	+	0.946	E2F-1	930	+	0.911	HNF-1	788	+	0.963
	HEB	982	-	1					HNF-1	799	+	0.88
									01-Oct	824	-	0.896
PDXK	TCF-1(P)	69	+	0.979	TCF-1(P)	72	+	0.979	Myc	155	+	1
higher in C liver equal in brain	SREBP-1	114	-	1	SREBP-1	117	-	1	USF	156	+	0.974
	c-Myb	367	+	0.978	c-Myb	370	+	0.978	AR	246	-	0.978
	E2F-1	415	-	0.933	Pax-4	496	+	0.874	c-Myc/Max	375	+	0.989
	E2F-1	419	+	0.992					Hand1/E47	399	-	0.972
	E2F	420	+	1					Myc	449	+	1
	E2F-1/DP-1	420	+	0.987					USF	450	+	0.993
	E2F-4/DP-1	420	+	1					AML1	634	+	1
	Rb/E2F-1/DP-1	420	+	1					Pax-3	646	+	0.89
	Pax-4	493	+	0.874					HNF-4 direct repeat 1 SREBP-1	660	-	0.89
										808	-	1
PTTG1IP	Pax-3	120	+	0.959	Pax-3	61	+	0.959	NF-kappaB	98	+	0.927
higher in H liver equal in brain	GATA-4	161	+	0.953	Elk-1	104	+	0.955	Pax-4	178	-	0.857
	TFIIA	179	+	0.931	TFIIA	119	+	0.931	Barbie	244	+	0.971
	AML1	526	+	1	AML1	466	+	1	Hand1/E47	292	+	0.961
	HNF-3alpha	540	-	0.961	HNF-3alpha	480	-	0.961	Hand1/E47	353	+	0.979
	FOX	541	-	0.95	FOX	481	-	0.95	Cdc5	726	+	0.963
	FOXD3	542	-	0.958	FOXD3	482	-	0.958	HNF-1	780	-	0.884
	ER	728	+	0.974	ER	668	+	0.962	TCF-1(P)	857	-	0.979
	Pax-3	729	-	0.937					MAZ	974	-	1
	ER	732	+	0.975								
TTC3	Ik-1	337	-	0.981	Ik-1	335	-	0.981	MEF-1	281	-	0.99
higher in H brain higher in H liver	Cdx-2	376	+	0.885	Cdx-2	374	+	0.885	AML1	556	+	1
	Osf2	477	-	1	v-Myb	402	-	0.966	AML1	625	+	1
	AML1	479	+	1	Osf2	476	-	1	TFIIA	667	+	0.94
	GATA-4	728	-	0.9	AML1	478	+	1	MRF-2	698	-	0.935
	HNF-1	865	+	0.886	GATA-4	727	-	0.9	NKX3A	815	-	1
	Cdx-2	914	+	0.913	Cdc5	772	-	0.877	Cdx-2	966	+	0.882
					HNF-1	865	+	0.886				

					Cdx-2	914	+	0.913					
USP16	Pax-4	43	+	0.839	Pax-4	27	+	0.839	Evi-1	35	+	0.895	
higher in H brain	Pax-4	165	-	0.878	Pax-4	149	-	0.878	Nkx2-5	360	+	1	
	SREBP-1	182	-	0.995	SREBP-1	166	-	0.995	HNF-1	529	+	0.912	
	PITX2	215	-	1	PITX2	199	-	1	COMP1	533	-	0.82	
	SF-1	422	-	1	SF-1	406	-	1	SF-1	616	-	1	
	HNF-3alpha	589	-	0.964	HNF-3alpha	573	-	0.964	Cdc5	738	+	0.847	
	FOX	590	-	0.937	FOX	574	-	0.937	NF-kappaB	768	-	0.929	
	FOXD3	591	-	0.96	FOXD3	575	-	0.96	Cdc5	817	+	0.879	
	NF-kappaB	761	+	0.926	NF-kappaB	745	+	0.926	Myc	852	+	1	
	SREBP-1	951	-	1	c-Ets-1(p54)	991	+	0.994	SREBP-1	876	+	1	
	c-Ets-1(p54)	991	+	0.994					Elk-1	913	+	0.991	
									c-Ets-1(p54)	916	+	0.994	

Red: TF binding sites found only in human

Blue: TF binding sites found only in chimpanzee

Yellow: TF binding sites common in human, chimpanzee and mouse

Grey: TF binding sites common in human and mouse.

Position 1 locates 1000 bases upstream from the coding sequence of gene

(+) indicate sense strand and (-) anti-sense strand, respectively.

Score 1 indicates perfect match to binding site and <1 indicates not-perfect match.