

表2. ヒト疾患遺伝子対応パンパングー遺伝子配列にみられる病的アミノ酸置換

機能を低下させ症状を悪化させる

Gene symbol	Gene name	Chimp. Seq.	Accession No.	Mutation	Phenotype
ABCA1	ATP-binding cassette A1	UCSC	CM990005	Ile883Met	Coronary artery disease 冠動脈症
AIRE	Autoimmune regulator	UCSC	CM020261	Pro252Leu	Polyendocrinopathy (APECED) 自己免疫疾患
ADRB2	Adrenergic, beta-2, receptor	UCSC	CM950016	Arg166Gly	Asthma, nocturnal, ぜん息
ADRB3	Adrenergic, beta-2, receptor	UCSC	CM970061	Gln276Glu	Obesity, 肥満
ADRB3	Adrenergic, beta-3, receptor	UCSC	CM950017	Trp64Arg	Hyperinsulinaemia 高インスリン血症
APOB	Apolipoprotein B	UCSC	CM980095	Glu3405Gln	Hypocholesterolaemia 低コレステロール血症
APOE	Apolipoprotein E	UCSC	CM900020	Cys130Arg	Coronary artery disease Alzheimer's disease 冠動脈症、アルツハイマー病
APOH	Apolipoprotein H	UCSC	CM993913	Leu247Val	Antihypertensive syndrome 抗リン脂質症
ENPP1	phosphodiesterase 1	UCSC	CM993455	Lys121Gln	Insulin resistance 抗インスリン
GALT	Galactose-1-phosphate uridylyl	UCSC	CM940804	Asn314Asp	Galactosaemia, Duarte ガラクトース血症
GSTP1	Glutathione S-transferase pi	UCSC	CM990713	Ile105Val	Oral cancer, 口腔がん
GYPB	Glycophorin A	UCSC	CM004555	Arg311Trp	Haemolytic disease of the newborn新生児溶血
LAMA2	Laminin alpha 2	UCSC	CM023406	Val2587Ala	Tuberculoïd leprosy, 類結核型らい病
LIPC	Lipase, hepatic	UCSC	CM962679	Leu334Phe	Hepatic lipase deficiency 肝リパーゼ欠損症
MKKS	MKK gene	UCSC	CM010926	Arg518His	McKusick-Kaufman syndrome McKusick-Kaufman 症候群
MCH1	mutL homologue 1 (mismatch repair)	UCSC	CM970965	Ala441Thr	Colorectal cancer, non-polyposis 大腸がん
MTRR	5-methyltetrahydrofolate-homocysteine methyltransferase reductase	UCSC	CM004563	Met221Ile	Coronary artery disease, 冠動脈症
MYOC	Myocilin	UCSC	CM023962	Gln48His	Glaucoma 1, open angle 緑内障
OTC	Ornithine carbamoyltransferase	UCSC	CM961052	Thr125Met	Hyperammonaemia アンモニア血症
PLA2G7	Phospholipase A2, group VII	UCSC	CM001309	Ala379Val	Asthma and atopy ぜんそく、アトピー
PON1	Paraoxonase 1	UCSC	CM031331	Ile102Val	Prostate cancer, increased risk, 前立腺がんリスク上昇
PON1	Paraoxonase 1	UCSC	CM930596	Gln192Arg	Paraoxonase activity パラオキシゲンナーゼ活性
PPARG	Type 2 diabetes	UCSC	CM981614	Pro12Ala	Obesity, 肥満
PRNP	Prion protein	UCSC	CM971239	Asn171Ser	Schizophrenia 統合失調症
TPMP	Thiopurine S-methyltransferase	UCSC	CM991193	arg215His	抗がん剤チオプリン分解
UCP1	Uncoupling protein 1	UCSC	CM031390	Ala64Leu	Waist-to-hip ratio, 肥満
SLC2A2	olute carrier family 2, member 2	UCSC	CM941277	Thr110Ile	Diabetes, NIDDM, 糖尿病
CIBA	Cytochrome b-245, alpha polypeptide	UCSC	CM983302	His72Tyr	Coronary artery disease, lower risk, 冠動脈症 リスク軽減
SCN5A	Sodium channel, voltage gated, type V, alpha polypeptide	UCSC	CM031355	His558Arg	Phenotype modifier, 心臓伝導障害の軽減

赤色：直接病因となる遺伝子、黄色：多因子疾患の一原因遺伝子が間接的に影響する遺伝子

表3. 病的アミノ酸置換のヒト疾患遺伝子対応哺乳動物遺伝子における分布

遺伝子	ADRB2	ADRB2	ADRB3	AIRE	AROE	CYBA	FCGR2A	GALT	GPD2	LAMA2
ヒト正常	Arg	Gln	Trp	Pro	Cys	His	Arg	Asn	His	Val
ヒト患者	Gly16	Glu27	Arg64	Leu252	Arg	Tyr72	His131	Asp314	Arg264	2587Ala
チンパンジー	Gly	Glu	Arg	Leu	Arg	Tyr	His	Asp	Arg	Ala
ゴリラ	Gly	Glu	Arg		Arg					
テナガザル										
マカク	Gly	Glu	Arg	Leu	Arg	Tyr	His		Arg	Ala
ウシ			Arg							
イヌ		Glu			Arg					
マウス	Gly	Glu	Arg	Leu	Arg	Tyr	His	Asp	Arg	Ala
疾患・症状	心不全	肥満	高インスリン血症	自己免疫疾患	冠動脈症、アルツハイマー病	冠動脈症	IgG2 binding上昇	ガラクトース血症	FFA コレステロール上昇	類結核型らい病
遺伝子	LIPC	MKKS	MTRR	MUT	PLA2G7	PON1	PPARG	PRSS1	UCP1	APOB
ヒト正常	Leu	Arg	Met	His	Ala	Gln	Ala	Asn	Ala	Glu
ヒト患者	Phe334	His	Ile22	Arg523	Val379	Arg192	Pro	29Thr	64Thr	Gln3405
チンパンジー	Phe		Ile	Arg	Val	Arg	Pro	Thr	Thr	Gln
ゴリラ			Ile							
テナガザル										
マカク	Phe	His	Ile	Arg	Val	Arg	Pro	Thr	Thr	Gln
ウシ									Thr	
イヌ		His					Pro			
マウス	Phe	His	Ile	Arg	Val	Arg	Pro	Thr	Thre	His
疾患・症状	肝リパーゼ欠損症	McKusick-Kaufman症候群	冠動脈症	メチルマロン尿症	ぜんそくアトピー	cancer risk	肥満	遺伝性膵炎	肥満	低コレステロール血症
遺伝子	AGT	APOH	DIO	ENPP1	SLO2A2	TPMP	PRNP	GYPA	OTO	SCN5A
ヒト正常	Met	Leu	The	Lys	Thr	Arg	Asn	Arg	Thr	His
ヒト患者	Thr235	Val247	Ala92	Gln121	Ile	His215	Ser171	Try31	Met	Arg558
チンパンジー	Thr	Val	Ala	Gln		His	Ser	Try	Met	Arg
ゴリラ	Thr					His	Asn	Gln		
テナガザル							Ser	Try		
マカク	Thr	Val	Ala	Gln	Ile	His	Asn	Ser		His
ウシ		Leu							Thr	
イヌ					Ile				Met	
マウス	Glu	Leu	Pro	His	Val	Gln	Asn	ND	Thr	Arg
疾患・症状	高血圧	血液凝固	insulin resistance	抗インスリン	糖尿病	抗がん剤チオプリン分解	統合失調症	新生児血	溶	アンモニア血症

遺伝子	IL1R	CETP	FCGR2A	ABCA1	GSTP1	MLH1	PON1	MYOC
ヒト正常	Arg	Ile	Gln	Ile	Ile	Ala	Ile	Gln
ヒト患者	His104	Val405	Lys127	Met	Val105	Thr	Val102	His48
チンパンジー	Arg	Ile	Gln	Met	Val	Thr	Val	His
ゴリラ								
テナガザル								
マカク	His	Val	Lys		Leu	Ala	Ile	Gln
ウシ								
イヌ				Val		Ala		
マウス	His	Lys	Val	Val	Gly	Ala	Ile	Gln
疾患・症状	アミロイド沈着抑制	HDL コレステロール上昇	IgG2 binding上昇	冠動脈症	Oral cancer	大腸がん	prostate cancer	緑内障

表 4. 病的アミノ酸置換のヒト正常集団における出現頻度と祖先型

Gene symbol	Gene name	Mutation	Phenotype	ancestral genotype	Frequency
AIRE	Autoimmune regulator	Pro252Leu	Polyendocrinopathy (APECED) 自己免疫疾患	Disease	0.00
GALT	Galactose-1-phosphate uridylyl transferase	Asn314Asp	Galactosaemia, Duarte ガラクトース血症	Disease	0.00
TPMP	Thiopurine S-methyltransferase	arg215His	抗がん剤チオプリン分解	Disease	0.00
CYBA	Cytochrome b-245, alpha polypeptide	His72Tyr	Coronary artery disease, lower risk, 冠動脈症リスク軽減	Disease	0.00
PRSS1	trypsin 1	Asn29Thr	Pancreatitis, hereditary 遺伝性膵炎	Disease	0.00
FCGR2A (CD32A)	Fc fragment of IgG	Arg131His	Lupus nephritis, protection against, 腎炎	Disease	0.00
UCP1	Uncoupling protein 1	Ala64Leu	Waist-to-hip ratio 肥満	Disease	0.12
SLC2A2	Solute carrier family 2, member 2	Thr110Ile	Diabetes, NIDDM, 糖尿病	Disease	0.12
ADRB3	Adrenergic, beta-3, receptor	Trp64Arg	Hyperinsulinaemia 高インスリン血症	Disease	0.13
APOE	Apolipoprotein E	Cys130Arg	Coronary artery disease Alzheimer's disease 冠動脈症、アルツハイマー病	Disease	0.15
APOE	Apolipoprotein E	Cys130Arg	Apolipoprotein E deficiency	Disease	0.15
PLA2G7	Phospholipase A2, group VII	Ala379Val	Asthma and atopy, ぜんそく、アトピー	Disease	0.19
APOB	Apolipoprotein B	Glu3405Gln	Hypocholesterolaemia 低コレステロール血症	Disease	0.20
LIPC	Lipase, hepatic	Leu334Phe	Hepatic lipase deficiency 肝リパーゼ欠損症	Disease	0.20
ADRB2	Adrenergic, beta-2, receptor	Gln27Glu	Obesity 肥満	Disease	0.23
MDJ	methylmalonyl-CoA isomerase	His532Arg	Methylmalonic aciduria メチルマロニル尿症	Disease	0.24
AGT	angiotensinogen	Met235Thr	Hypertension 高血圧	Disease	0.36
GP2	glycerol-3-phosphate dehydrogenase 2	His264Arg	Increased plasma FFA and glycerol levels 血清遊離脂肪酸、グリセロール上昇	Disease	0.38
LAMA2	Laminin alpha 2	Val2587Ala	Tuberculoïd leprosy, 類結核型らい病	Disease	0.41
D102	type 2 iodothyronine deiodinase	Thr92Ala	Insulin resistance 抵抗性インシュリン	Disease	0.49
ADRB2	Adrenergic, beta-2, receptor	Arg166Iy	Asthma, nocturnal, ぜん息	Disease	0.52
PON1	Paraoxonase 1	Gln192Arg	Paraoxonase activity ナーゼ活性	Disease	0.52
MTTR	5-methyltetrahydrofolate-homocysteine methyltransferase reductase	Met22Ile	Coronary artery disease 冠動脈症	Disease	0.57
ENPP1	phosphodiesterase 1	Lys121Gln	Insulin resistance, 抗インスリン	Disease	0.63
PPARG	Type 2 diabetes	Pro12Ala	Obesity, 肥満	Disease	0.85
MYOC	Myocilin	Gln48His	Glaucoma 1, open angle (緑内障)	Wild type	0.00
MLH1	mutL homologue 1 (mismatch repair)	Ala441Thr	Colorectal cancer, non-polypsis 大腸がん	Wild type	0.00
PON1	Paraoxonase 1	Ile102Val	Prostate cancer, increased risk, 前立腺がんリスク上昇	Wild type	0.04
GYPA	Glycophorin A	Arg31Trp	Haemolytic disease of the newborn	Unresolved	0.00

MKKS	MKK gene	Arg518His	McKusick-Kaufman syndrome McKusick-Kaufman 症候群	Unresolved	0.00
OTC	Ornithine carbamoyltransferase	Thr125Met	Hyperammonaemia アンモニア血症	Unresolved	0.00
TTR	transthyretin	Arg104His	Amyloidotic polyneuropathy アミロイド沈着症	Unresolved	0.00
PRNP	Prion protein	Asn171Ser	Schizophrenia 統合失調症	Unresolved	0.01
ABCA1	ATP-binding cassette A1	Ile883Met	Coronary artery disease Alzheimer's disease 冠動脈症、アルツハイマー病	Unresolved	0.14
SCN5A	Sodium channel, voltage gated, type V, alpha polypeptide	His558Arg	Phenotype modifier 遺伝性心臓伝導障害の軽減	Unresolved	0.18
GSTP1	Glutathione S-transferase pi	Ile105Val	Oral cancer. 口腔がん	Unresolved	0.22
CE1P	cholesteryl ester transfer protein	Ile405Val	Higher HDL cholesterol level 高HDLコレステロールレベル	Unresolved	0.28
FCGR2A (CD32A)	Fc fragment of IgG	Gln127Lys	Immunoglobulin binding variant IgG結合変異体	Unresolved	0.36
APOH	Apolipoprotein H	Leu247Val	Antiphospholipid syndrome 抗リン脂質症	Unresolved	0.48

(表5)

病的アミノ酸置換の祖先型

祖先型	置換数 (割合)	病的アミノ酸置換の出現要因
正常型	3 (8%)	ヒト集団では比較的出现頻度が低い。チンパンジーからヒトに進化した後、ヒト集団中に自然発生的に生じたアミノ置換による疾患
疾病型	25 (64%)	疾患の原因となるアミノ酸置換はヒトの祖先である非ヒト霊長類では正常型で、病的症状も非ヒト霊長類にとってはさほど障害とならない。ヒトへ進化した後、症状を抑制するようなヒト型の変異が起きたものと思われる。ヒト自然集団における病的アレルの分布をみると疾病によって様々で、病的アレルが抑制的な正常型アレルで置き換わって行く様子がわかる
確定不能	11 (28%)	病的アミノ酸置換は進化的にはなれた種にも不規則に出現する。種の生活環境や食性に影響を受けた変異

厚生労働科学研究費補助金（ヒトゲノム・再生医療等研究事業）

平成17年度 分担研究報告書

「チンパンジー完全長cDNAの解析」

分担研究者 平井百樹 東京女子医科大学国際統合医科学インスティテュート教授

研究要旨

チンパンジー・ゲノム配列が2005年9月に報告され、ヒトとチンパンジーの遺伝子レベルでの比較が進んでいる。しかしながら、まだチンパンジー・ゲノム配列は22番ならびにY染色体以外は概要版である。したがって、チンパンジーのトランスクリプト配列の情報が少ないことから、遺伝子レベルでの両種間の比較解析にはまだ情報が不十分である。本研究では、このギャップを埋めるための研究の一環として、チンパンジーの完全長cDNAの解析を進めた。脳と皮膚から得た87遺伝子に対応するチンパンジー完全長cDNAを選び、その全長配列と対応するヒトmRNA配列（RefSeq）とをヒトゲノム配列上にマッピングした。そして、従来行っていた置換率算定に加え、exon-intron構造、転写開始点・終了点の位置の比較など、大規模な遺伝子構造における種差の有無について比較解析を行った。この際、公開されたチンパンジー・ゲノム概要配列も参考に用いた。塩基置換を指標にした種間差は、塩基挿入・欠失（indel）を加算すると1.36%になり、従来の置換のみの種間差（0.73%）よりも大きな値となっている。エクソン構造の違い、選択的スプライシング、Alu配列などを含めた大きな種特異的配列の挿入などを考慮に入れると、種間差はさらに大きく見積もられることがわかった。一方、ヒトで疾患に関連する変異型配列がチンパンジーでの野生型として見られる例が今年のチンパンジー・ゲノム解析で報告されている。これらの再検討のためにチンパンジー20個体のゲノム配列の解析を心臓血管疾患に関連する6遺伝子の塩基置換について行った。たとえばADRB3遺伝子においては、第1エクソン部分にヒトとの種差が高頻度にみられること、また疾患と関連するとみられるArg64Trp多型が、チンパンジーの集団研究でも多型としてみられることなど、ヒト疾患関連遺伝子の塩基置換に関する新たな側面がみられた。今後、脳と皮膚以外の組織での発現遺伝子を調べる必要が生ずることを踏まえ、さらに他の組織の凍結試料を保管した。

A. 研究目的

（1）2005年9月にチンパンジー・ゲノム概要版配列の解析結果が報告された。ゲノムの94%をカバーし、ヒトゲノムとの比

較から、約3,500万塩基の置換（1.23%の置換率）、500万ヶ所の欠失および挿入、多数の重複領域などの相違が同定された。また、遺伝子の約70%になんらかのアミノ酸置換が検出されたという。ヒト21番染色体と相同のチンパンジー22番染色体ならびにY染

色体に関しては、日本の研究グループが中心となり、精度の高い配列が報告されている。このようなゲノム情報をもとに、種間差にみられる遺伝子レベルと表現型レベルとの関係についての研究が期待される。しかしながら、チンパンジー・ゲノム情報はまだ不完全であることから、機能単位としての遺伝子に関する情報を得るには、やはりまだ転写産物についての詳細な解析が必要である。本研究では、昨年度にひきつづき、脳と皮膚から得た完全長cDNAライブラリーより得た 5' -cDNA 配列約 1万 5000 から、精度高く配列決定されたものを87選択し、解析を行う。これまで、RefSeqに登録された対応配列による置換率についてはすでに結果を得て来たが、本年度は、チンパンジーcDNA配列と対応するヒトのRefSeq配列をヒトゲノム配列にマップして、正確にortholog間でのエキソン - イントロン構造の比較、種差と考えられる選択的スプライシング (alternative splicing, AS) の検出、転写開始点と転写終了点の位置の比較などの大規模な種差に注目して解析を行うことを目的とした。

(2) また一方、ヒトのゲノムの多様性に目を向ける際に、霊長類、特にチンパンジーの遺伝子・ゲノムデータは重要な情報をもたらす。ヒトの疾患に関連するとおもわれる変異型配列が、チンパンジーでは野生型になっている例が、昨年チンパンジー・ゲノム概要版の論文には16報告されている。これらの疾患遺伝子の進化的側面からの解析についての検討は興味ある。そこで心血管疾患に関連する 6 遺伝子の多型について、チンパンジー20個体についてダ

イレクトシーケンシングにより比較解析をおこなう。特にヒトで肥満に関わるといわれる β 3 アドレナリン受容体遺伝子 (ADRB3) のArg64Trp変異に関して、ヒトではこの変異型が肥満に関連するがチンパンジーでは野生型である、との報告の検証をおこなう。

(3) 今後の研究に備え、さらに他の組織標本を揃える。

B. 研究方法

(1) 解析に用いた 87 クローンのうち、脳由来が49、皮膚由来が38である。解析に用いる配列をシーケンシングの確からしさ (Phrap値) を規準に改めて選択しなおし、また新規のチンパンジー・ゲノムの概要配列情報も参考にしながら解析を進めた。解析項目は、対応できる配列間での塩基置換、ヒトゲノム配列と比較しての挿入・欠失 (indel)、それに選択的スプライシング (AS)、転写開始・終了点の位置の比較による大規模エクソン変化である。具体的には、BLASTプログラムを用い、ヒトmRNA、genomeの各配列データベースに対して相同性検索を行って配列データを分類した。そのうえで、cDNA全長配列データならびに対応するRefSeqをヒトゲノム (NCBI b35, UCSC hg16) 上へマッピングし、exon-intron構造およびexon配列の比較をおこなった。こうして非翻訳領域 (5' -UTR, 3' -UTR) とコード領域を特定して、領域ごとのヒトのorthologとの比較を行なった。

(2) SNP に関する比較ゲノム学的解析のモデルとして、 β 3 アドレナリン受容

体遺伝子(ADRB3)の第1 exon の SNP に注目した。190 位の T が C に置換した allele が日本人で 15%以上存在する。これにより遺伝子産物の N-末端から 64 番目のアミノ酸がトリプトファンからアルギニンへ置換 (Trp64Arg) する。この変異は、はじめ非インシュリン依存性糖尿病 (NIIDM) の罹患と関連して発見された。交感神経から分泌されるアドレナリンは、体脂肪の蓄積を抑え、脂肪を燃焼させる働きがある。Trp64Arg では、 β 3 アドレナリン受容体の機能が低くなる。データベース上、新世界ザル、アカゲザル、チンパンジー、オランウータンを含めた哺乳動物はすべて置換型であった。すなわちヒトの非置換型がむしろ例外的であることになる。そこでチンパンジー20 個体についてゲノム DNA をもちいてダイレクトシーケンシングにより、exon-1 の塩基配列を解析した。

(3) すでに得られている、チンパンジーの死亡後数時間以内に液体窒素に保管された試料について整理した。

C. 研究結果

(1) 全長cDNA配列の種間比較解析：

チンパンジーのトランスクリプト配列の 99.4%がヒトゲノム上にマップされた。22 遺伝子(25%)について、ヒトのorthologと比較してアミノ酸置換がなかった。逆に言うと、65遺伝子(75%)についてはアミノ酸置換が生じていることになる。なお5遺伝子については、塩基配列レベルでCDSが完全に一致した。ヒトゲノム配列を軸にしてチンパン

ジー・トランスクリプト配列にみられる変化の類型を (Figure 1) に示した。a)は塩基配列レベルの置換と挿入・欠失を比較し、b)はエクソンレベルでの構成を比較したものである。置換率は全体、5' -UTR、CDS、3' -UTR、アミノ酸でそれぞれ 0.73%、0.87%、0.59%、0.93%、0.58%であった。これにindelを考慮に入れると、塩基配列レベルでの種差は増加する。置換とindelとを合わせた塩基レベルでの種間差の算定法にはいろいろあるが、ここでは比較対象とするヒトゲノム配列に塩基置換数とindelの塩基数を加算した値を示した (Figure 2)。同義置換率をNei-Gojoboriにより計算すると1.77となる。この値は5' -UTR、3' -UTRのそれぞれより大きいことから、UTR領域が何らかの機能的制約を受けていることが考えられる。本年度特に精度高く解析した挿入・欠失 (indel) に関して領域ごとに比較したところ、チンパンジーにおいては、5' -UTRで挿入あるいは欠失がみられたものがそれぞれ3遺伝子、CDSで挿入がみられたものが7遺伝子、欠失が3 遺伝子、3' -UTRで挿入がみられたものが20遺伝子、欠失が17遺伝子であった。また、ヒトゲノム配列上にマップされない配列が、5' -UTRで 7箇所、3' -UTRで 6箇所存在した。

両種で合計740のエクソンが見いだされた。そのうち9 遺伝子に種特異的エクソンが14個みられ、7 個ずつがそれぞれの種に特異的なエクソンであった。8 エクソンは5-UTRに存在し、3' -UTR、CDS、5' -UTRの平均的長さ (101bp、1,167bp、592bp) の差を考慮すると、5' -UTRにおける種特異的ASが高度に生じていることになる。また、

チンパンジー3' -UTRにAlu配列の挿入がみられるもの、50塩基以上の単位長を持つ反復配列の繰り返し数の違いによる大規模な挿入・欠失が見つかった。転写開始点(5')は転写終了点(3')よりも種間での保存性が低く、開始点はチンパンジーの方がヒトより上流に位置する傾向が見られた。

このように、ヒトゲノム配列を軸に、チンパンジー・トランスクリプト配列の転写開始・終了位置、種特異的とおもわれる選択的スプライシング、大きな反復配列の挿入などの差がもたらす違いを考慮すると、遺伝子の種間差は従来考えられていたより大きくなる。このような大規模な変化を加えた種間差を正確に定量化する必要があると考える。実際に大規模な種差のあった遺伝子例を (Figure 3) に示す。(a)では、5' -UTR に違いがあり、チンパンジーには無いエクソンがヒトにみられる。(b)では、CDS にチンパンジー固有の 126bp の非反復配列の挿入が存在する。(c)では、3' -UTR に 559bp にわたる SINE 配列が付加する。これはチンパンジー・ゲノム概要配列でもマップできている。(d)では、10bp と 55bp の 2 種類の重複配列が 3' -UTR 領域にあり、その一方の 55bp の配列はチンパンジー特異的でしかもチンパンジー・ゲノム配列に当てると個体差とおもわれる数の違いがみられる。

(2) ADRB3 のアルギニンがトリプトファンに変化した型は他の哺乳類にはみられない。exon 1 をチンパンジー20 個体のゲノム DNA について調べたところ、この変異に関してヘテロ接合が3 個体

がみいだされた。したがって、アルギニンのトリプトファンへの置換はヒトではじめて生じたのではなくヒト・チンパンジー系統で生じたものと推定される結果を得た。さらに、エクソン1には塩基置換が高密度にみられ、しかもこのチンパンジーでの置換がヒトのある隔離集団においても新規の SNP として発見された。

D. 考 察

本研究で得られた全長 cDNA のうち、同一遺伝子について得られたクローンの中でもっとも長いトランスクリプト配列で、さらにフラップ値が高い正確なシーケンスデータを選択した。その配列をゲノム配列上にマップすることをとおして、種間差を求めることをおこなった。その結果、遺伝子によっては、大規模な配列の挿入・欠失があること、転写開始点ならびに終了点に種間差が有ること、種特異的とおもわれる選択的スプライシングが約 10%の遺伝子に見られることなどがわかった。この中には、559bp もの配列がヒトゲノムの未決定部位に対応する例 (PccB1795) がある。ヒトゲノムが解読されたといってもまだ残っている未解読領域がある。チンパンジーのトランスクリプト解析により得られる新規とおもわれる遺伝子配列の中には、そのような領域に対応する配列が含まれる可能性がある。またさらに、ゲノム情報からは偽遺伝子と判定される配列が、実は発現している例や、intergenic でなくエクソン領域内

へのトランスポゾンの挿入例も見い出された。このようにゲノム配列からだけでは得られないような遺伝子構造についての情報が、全長 cDNA を解析することにより解明される。本研究で仮に種特異的オルターナティブ・スプライシングと分類したものについても、実際にはまだ種特異的とは断定できない。ヒトの RefSeq 情報に対してチンパンジーの mRNA 配列情報が少ないためである。しかし、indel や大規模構造変化を含めたトランスクリプト全長配列の比較解析は、遺伝子機能解析上も重要な情報をもたらすことを確認した。

昨年のチンパンジー・ゲノム解析の報告によれば、ヒト疾患に関連する対立遺伝子が、チンパンジーにおいては野生型となっている例が 16 見つかった。そのうちの 15 例は 7 個体の解析ですべての個体がホモ接合になっていた。また 1 例ではチンパンジーでも共通の多型であることも示唆された。これらには心血管疾患やインスリン耐性などの疾患に関連するものがあるという。本研究で ADRB3 に見られたチンパンジーでの多型の発見は、ヒト集団で観察される遺伝的多型のうちの変異型がチンパンジーで野生型である、という前提で疾患の遺伝子の進化を推測すると、誤る可能性のあることを示唆する。

E. 結論

系統的に最もヒトに近いチンパンジーとの塩基配列の比較解析は、種間にみら

れる差を解析し、その表現型への効果の生成を知るうえで重要となる。ゲノム情報が完備してきても、*in silico* 解析法では正確な exon-intron 構造の比較や、UTR 領域の役割などの情報がまだ完全には得られない。完全長 cDNA を用いた研究によりこれらの情報が得られることが期待される。また、種特異的 AS や発現偽遺伝子など、大規模でダイナミックな遺伝子の差異を明らかにすることができる。一方で、罹病性に関連するといわれる塩基多型のなかには、霊長類で保存的なものもあるし、ヒトで特異的に生じたものもあるが、そのような生物学的意味を知るうえでも、近縁のチンパンジーのトランスクリプトを比較解析する重要性がある。今後、特に心血管系異常などに関わる遺伝子の解析を進めることが要請されるので、チンパンジー由来の完全長 cDNA クローン資源はさらに重要になると考えられる。また、我々の得ているライブラリーの中に、特にまだヒトゲノム配列にも RefSeq にも当たらない配列のクローンが多数見いだされていることは、上述したようにヒトゲノム配列の未決定領域との対応からも新たな情報が得られるものと考えられる。今後、ヒトの翻訳領域における SNP の意義などを調べるうえでも、チンパンジーの配列情報が価値の高い研究資源となると考えられるので、さらに脂肪組織、肝臓、腎臓についても収集し、RNA 抽出して発現解析に供する準備を行った。

F. 研究発表

1. 論文発表

1) K. Nagao, N. Takenaka, M. Hirai and S. Kawamura: Coupling and decoupling of evolutionary mode between X- and Y-chromosomal red-green opsin genes in owl monkeys. *Gene* (2005) 352-:82-91

2. 学会発表

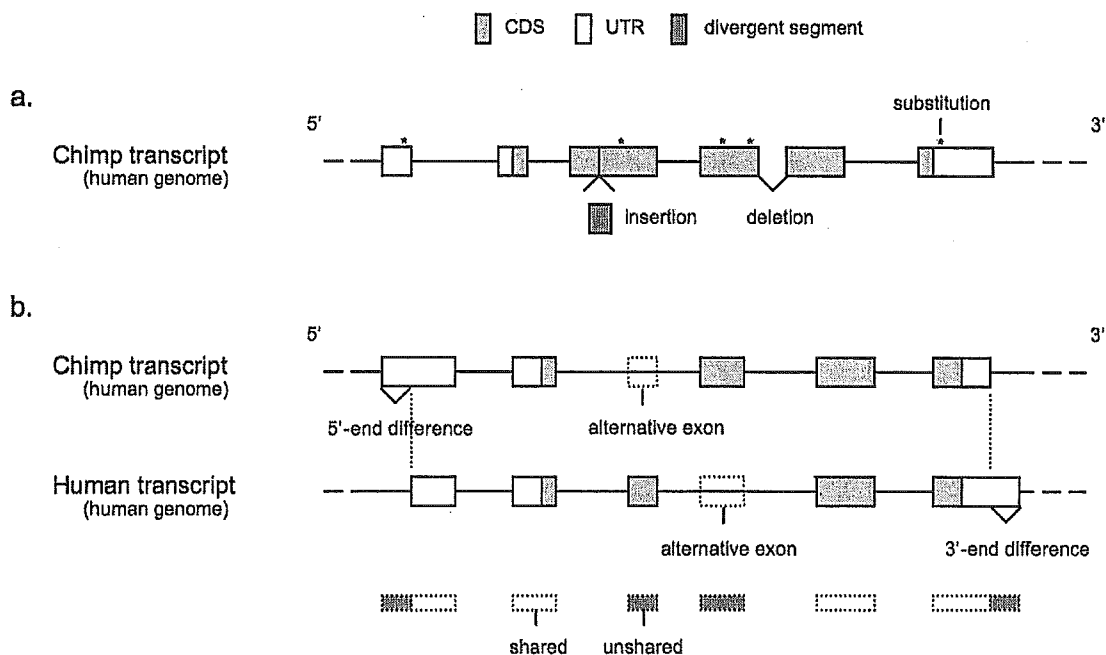
1) Sakate R, Suto Y, Imanishi T,

Hayasaka I, Gojobori T, Hashimoto K, and Hirai M.: Comparative analysis of chimpanzee full-length cDNAs mapped onto human genome unveils potentially large inter-species divergence in gene loci. 55th Annual Meeting of American Society of Human Genetics 2005 (October 25-29, 2005, Salt Lake City, Utah, U. S. A.

G. 知的所有権の取得状況

取得なし

Figure 1



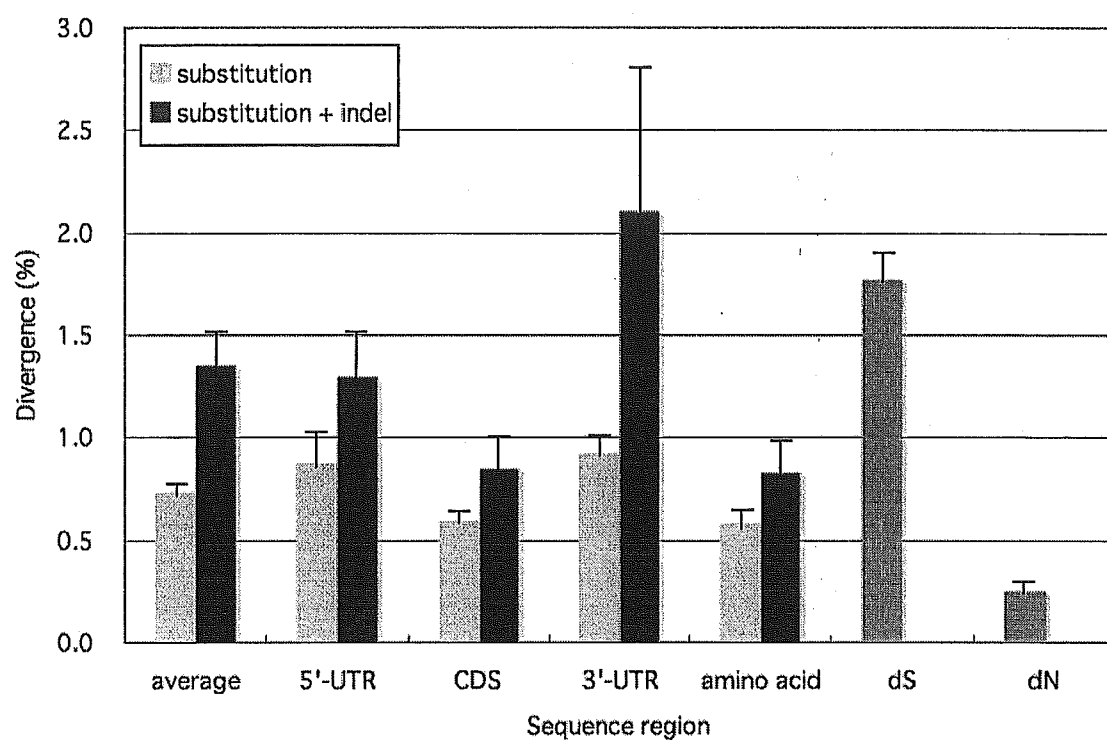
ヒトとチンパンジーの遺伝子の種間差の解析

チンパンジーのトランスクリプト配列と対応するヒトの RefSeq をヒト・ゲノム配列上にマップして比較した。 a. 置換と挿入・欠失 (indel) b. エクソン構成の相違。

5' -end あるいは 3' -end にみられる挿入・欠失と選択的スプライシング

種間の共有 (shared) 配列と、非共有 (unshared) 配列に分類して解析した。

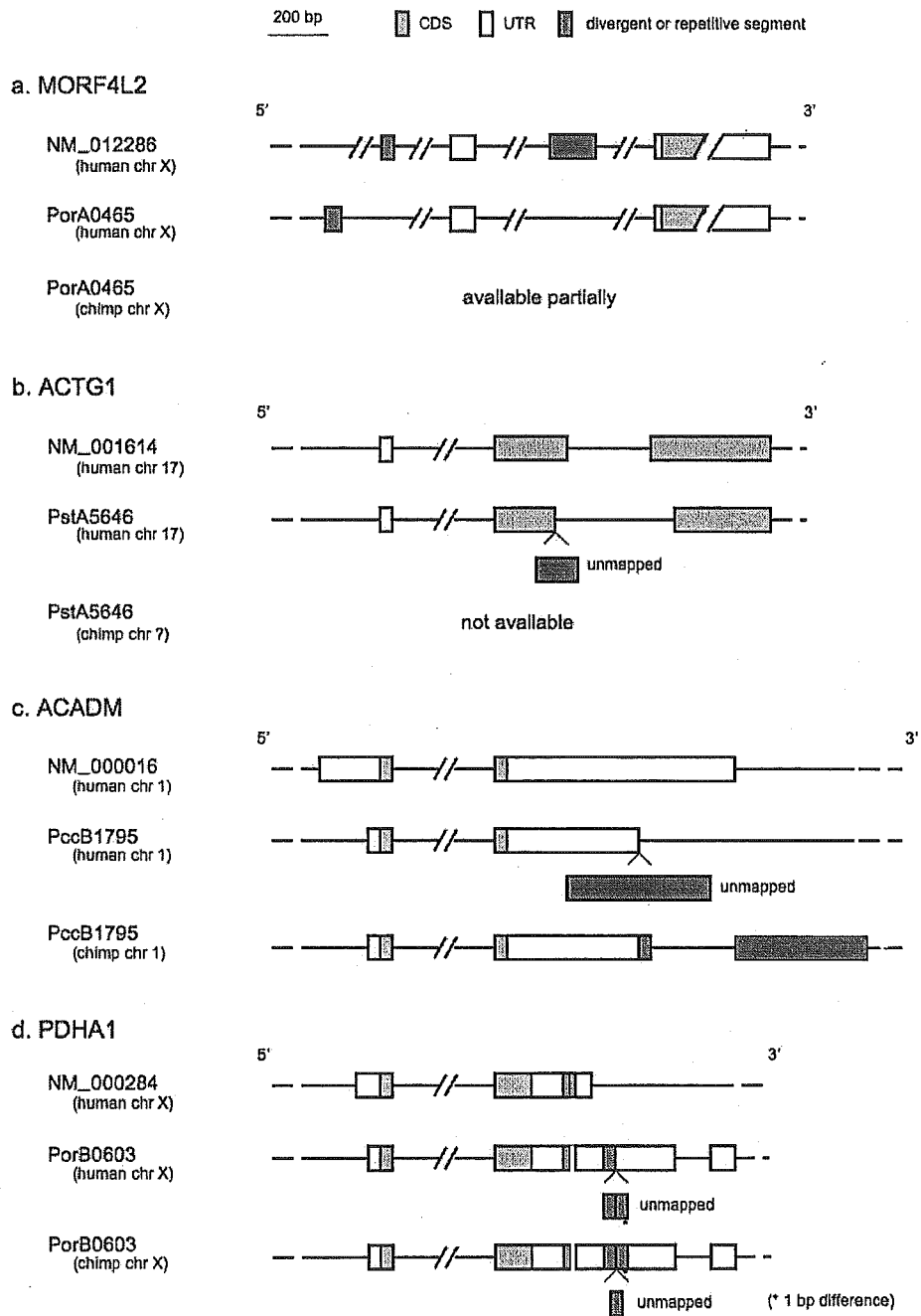
Figure 2



各領域にみられる塩基配列レベルでの種間差

塩基置換のみ (左コラム) と、欠失・挿入 (indel) の塩基数を加算した違い (右コラム)。
同義置換率 (dS) と非同義置換率 (dN) も示した。

Figure 3



大規模な種間差を示す遺伝子の例 (説明は本文参照)

分 担 研 究 報 告 書

カニクイザルcDNAライブラリー作製とヒト疾病関連cDNA分離

分担研究者 菅野純夫 東京大学新領域創成科学研究科メディカルゲノム専攻
ゲノム制御医科学分野 教授

本年度は、カニクイザル肝臓由来完全長cDNAの部分配列情報とヒト遺伝子のホモロジー情報をもとに、ヒトの疾患関連遺伝子を分離する過程で見出されてきた、カニクイザル特異的遺伝子の候補について詳細な検討をおこなった。カニクイザル肝臓cDNA 22324 クローンの部分配列をヒトゲノムと比較したところ約50種類のカニクイザル特異的遺伝子の候補を得ることができた。現在までに、それらのうち22種の全長配列を決定し詳細に解析を行なっている。

A. 研究目的

悪性腫瘍・動脈硬化・糖尿病・精神病等疾病の理解とその解決ためには、システムとしての生体を遺伝子レベルで理解する必要がある。このためにはヒトの全遺伝子が明らかにされ、その発現解析・機能解析が行われることが必要になる。

遺伝子の発現解析、機能解析には、mRNAの完全なコピーである完全長 cDNA が必要であり、遺伝子レベルで疾病を研究していくための欠くことの出来ない基盤となっている。特に、疾病の遺伝子レベルの研究を広範に展開するために、ヒトの全遺伝子の完全長 cDNA クローンを持つことは重要と考えられる。

本研究は、疾病の遺伝子レベルの研究基盤として、完全長 cDNA クローンを多数収集し、その情報をもとに、遺伝子発現の制御や、完全長 cDNA の機能解析に直結するような情報の収集を目指す。

B. 研究方法

部分配列を決定したカニクイザル cDNA ライブラリーのうちの、Q1v (liver) 22,324 件（実際にはこのうち20,895件の3' one-path

read)についてマッピングを行った。その結果、マップ可クローン数は15,898件であり、重複を考慮に入れた総クラスタ数は1,650クラスタであった。そのうちカニクイザルユニーククラスタ数は59クラスタとなった。ここから、クラスタの代表完全 cDNA クローンを選び、その全長を決定し、それをヒトゲノム配列と比較した。現在までに22クラスタについて、全長配列を決定し、それらについて詳細な解析を行なった。

C. 研究結果

現在までにわれわれは、22クラスタについてマッピングを行なっている。そのうち14クラスタが染色体上の単一個所にマップされた。それらは、

Q1v-U012A-D11

Q1v-U022A-G3

Q1v-U035A-A3

Q1v-U038A-D9

Q1v-U042A-F11

Q1v-U050A-D10

Q1v-U095A-H5

Q1v-U166A-D7

Q1v-U275A-H1

Q1v-U361A-G6

Q1v-U365A-D3
Q1v-U385A-B9
Q1v-U389A-C1
Q1v-U426A-F1

である。

このうち、Q1v-U022A-G3、Q1v-U038A-D9、
Q1v-U042A-F11、Q1v-U050A-D10、
Q1v-U389A-C1の5種以外は、結局ヒトの対
応する遺伝子とほぼ一致し、5' 端または
3' 端等で、微妙に配列が異なるものである
事が分かった。

Q1v-U038A-D9とQ1v-U042A-F11は、ヒト遺
伝子と高いホモロジーを有するものの、そ
れぞれCYP2C family、CFHL family遺伝子で
あり、カニクイザル特異的メンバーである
可能性がある。のこる3種のうち、
Q1v-U022A-G3とQ1v-U050A-D10は、スプライ
シングはあるものの、ORFを持たない
non-codingRNAであるように思われる。なお、
図として、Q1v-U022A-G3、Q1v-U050A-D10
および、Q1v-U389A-C1をヒトゲノム配列に
マップした場合をUCSFゲノムbrowserを使
って表示したものを、参考として示した。

D. 考察

本年度は、カニクイザル cDNA 部分配列情
報とヒト遺伝子のホモロジー情報をもとに、
ヒトの疾患関連遺伝子を分離する過程で見
出されてきた、カニクイザル特異的遺伝子
の候補について詳細な検討をおこなった。
その結果、5 種類のカニクイザル特異的
cDNA クローンを見出してきた。このような
カニクイザル特異的遺伝子の存在がヒトと
カニクイザルの差を生みだしている可能性
もある。アルツハイマー病やエイズなどヒ
トで見られる疾患が他の霊長類で見られな
いことは良く知られており、遺伝子の種類
の差の研究も疾患の治療に貢献する可能性
がある。

E. 結論

カニクイザル完全長cDNA部分配列をヒト

ゲノム配列と比較することで種特異的遺伝
子を同定して行くことは、遺伝子機能研究
の基盤を整備する上で有効であるばかりで
なく、疾患の研究に重要な示唆を与えると
考えられる。

F. 研究発表

論文発表

1. Yamashita R, Suzuki Y, Sugano S, Nakai K. Genome-wide analysis reveals strong correlation between CpG islands with nearby transcription start sites of genes and their tissue specificity. *Gene*. 350:129-136, 2005.
2. Osada N, Hirata M, Tanuma R, Kusuda J, Hida M, Suzuki Y, Sugano S, Gojobori T, Shen CK, Wu CI, Hashimoto K. Substitution Rate and Structural Divergence of 5'UTR Evolution: Comparative Analysis Between Human and Cynomolgus Monkey cDNAs. *Mol Biol Evol*. 22: 1976-1982, 2005. 2005 Jun 8; [Epub ahead of print]
3. Tahira T, Baba S, Higasa K, Kukita Y, Suzuki Y, Sugano S, Hayashi K. dbQSNP: a database of SNPs in human promoter regions with allele frequency information determined by single-strand conformation polymorphism-based methods. *Hum Mutat*. 26:69-77, 2005.
4. Yamasaki C, Koyanagi KO, Fujii Y, Itoh T, Barrero R, Tamura T, Yamaguchi-Kabata Y, Tanino M, Takeda JI, Fukuchi S, Miyazaki S, Nomura N, Sugano S, Imanishi T, Gojobori T. Investigation of protein functions through data-mining on integrated human transcriptome database, H-Invitational database (H-InvDB). *Gene*. 364: 99-107 (2005). 2005 Sep 23; [Epub ahead of print]
5. Otsuki T, Ota T, Nishikawa T, Hayashi K, Suzuki Y, Yamamoto J, Wakamatsu A, Kimura K, Sakamoto K, Hatano N, Kawai Y, Ishii S, Saito K, Kojima S, Sugiyama T, Ono T, Okano K, Yoshikawa Y, Aotsuka S, Sasaki N, Hattori A, Okumura K, Nagai K, Sugano S, Isogai T. Signal Sequence and

Keyword Trap in silico for Selection of Full-Length Human cDNAs Encoding Secretion or Membrane Proteins from Oligo-Capped cDNA Libraries. DNA Res. 12: 117-126, 2005.

6. Yamashita R, Suzuki Y, Wakaguri H, Tsuritani K, Nakai K, Sugano S. DBTSS: DataBase of Human Transcription Start Sites, progress report 2006. Nucleic Acids Res. 34: D86-D89, 2006.

7. Kimura K, Wakamatsu A, Suzuki Y, Ota T, Nishikawa T, Yamashita R, Yamamoto J, Sekine M, Tsuritani K, Wakaguri H, Ishii S, Sugiyama T, Saito K, Isono Y, Irie R, Kushida N, Yoneyama T, Otsuka R, Kanda K, Yokoi T, Kondo H, Wagatsuma M, Murakawa K, Ishida S, Ishibashi T, Takahashi-Fujii A, Tanase T, Nagai K, Kikuchi H, Nakai K, Isogai T, Sugano S. Diversification of transcriptional modulation: Large-scale identification and characterization of putative alternative promoters of human genes. Genome Res. 16: 55-65, 2006. Epub 2005 Dec 12.

参考図(別添)

Q1v-U022A-G3

Q1v-U050A-D10

Q1v-U389A-C1

Chromosome Band
STS Markers
UCSC Known Genes (June, 05) Based on Uniprot, RefSeq, and GenBank mRNA
Consensus CDS
RefSeq Genes
Non-Human RefSeq Genes
H
Ensembl Gene Predictions
ACEview Gene Models with ALT-Splicing
GENSCAN Gene Predictions
ExonIntron Human/Mouse/Rat/Dog
ExonIntron ALT-Splicing Transcripts
Human mRNAs from GenBank
CpG Islands (Islands < 300 Bases are Light Green)
Affymetrix Transcription Project Phase 2 (A375 Txn)
Hu/Chimp/Mouse/Rat/Dog/Chicken/Fugu/Zfish Multiz Alignments & Conservation
Simple Nucleotide Polymorphisms (SNPs)
Repeating Elements by RepeatMasker

chr2: 369900000 370000000 370100000 370200000 370300000 370400000 370500000 370600000 370700000 370800000 370900000 371000000

Chromosome Band

STS Markers

Gap

CCDS

Mus Strn

Rattus Strn

Drosoph Cka

Drosoph Cka

Drosoph Cka

ENST00000253918

RiceView Genes

GenScan Genes

ExonInphy

ExonIn1K

AF065351

CpG Islands

No data -

Affy Txn Phase2

No data -

Conservation

chimp

dog

mouse

rat

chicken

fugu

zebrafish

Fugu Blat

SNPs

RepeatMasker

STS Markers on Genetic (blue) and Radiation Hybrid (black) Maps

UCSC Known Genes (June, 05) Based on Uniprot, RefSeq, and GenBank mRNA

Consensus CDS

RefSeq Genes

Non-Human RefSeq Genes

Ensembl Gene Predictions

RiceView Gene Models With ALT-SPLICING

GenScan Gene Predictions

ExonInphy Human/Mouse/Rat/Dog

ExonIn1K ALT-SPLICING Transcripts

Human mRNAs From GenBank

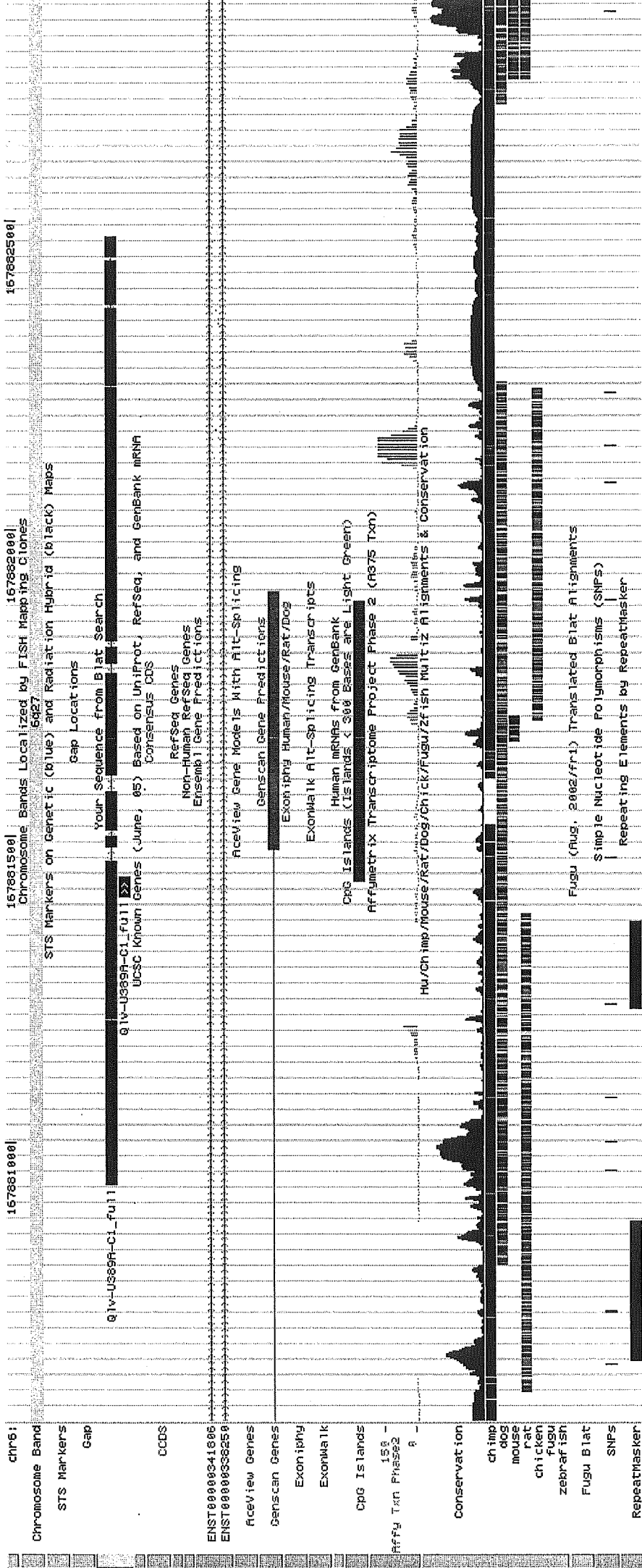
CpG Islands (Islands < 300 Bases are Light Green)

Affymetrix Transcription Project Phase 2 (A375 Txn)

Hu/Chimp/Mouse/Rat/Dog/Chicken/Fugu/Zfish Multiz Alignments & Conservation

Simple Nucleotide Polymorphisms (SNPs)

Repeating Elements by RepeatMasker



研究成果の刊行に関する一覧表

1) K. Nagao, N. Takenaka, M. Hirai and S. Kawamura:

Coupling and decoupling of evolutionary mode between X- and Y-chromosomal red-green opsin genes in owl monkeys.

Gene (2005) 352-:82-91

2) Osada, N., Hirata, M., Tanuma, R., Kusuda, J., Hida, M., Suzuki, Y, Sugano, S., Gojobori, T., Shen, C.-K. J., Wu, C.-I., Hashimoto, K.:

Substitution rate and structural divergence of 5' UTR evolution: Comparative analysis between human and cynomolgus monkey cDNAs.

Mol. Biol. Evol. 22(10):1976-1982(2005)

3) Yamashita R, Suzuki Y, Wakaguri H, Tsuritani K, Nakai K, Sugano S.:

DBTSS: DataBase of Human Transcription Start Sites, progress report 2006.

Nucleic Acids Res. 34: D86-D89, 2006.

4) Kimura K, Wakamatsu A, Suzuki Y, Ota T, Nishikawa T, Yamashita R, Yamamoto J, Sekine M, Tsuritani K, Wakaguri H, Ishii S, Sugiyama T, Saito K, Isono Y, Irie R, Kushida N, Yoneyama T, Otsuka R, Kanda K, Yokoi T, Kondo H, Wagatsuma M, Murakawa K, Ishida S, Ishibashi T, Takahashi-Fujii A, Tanase T, Nagai K, Kikuchi H, Nakai K, Isogai T, Sugano S.:

Diversification of transcriptional modulation: Large-scale identification and characterization of putative alternative promoters of human genes.

Genome Research 16: 55-65, 2006.