

博士論文

レトロ遺伝子 *PIPSL* の転写機構に関する研究

2018年9月

長浜バイオ大学大学院 バイオサイエンス研究科

バイオサイエンス専攻

バイオ科学技術研究領域

氏名 松村 研哉

## 要旨

新規遺伝子は、既存の遺伝子配列や転写活性またタンパク機能を変化させる突然変異により誕生し、表現型の進化に寄与していると考えられている。新しく遺伝子が創られる過程には、祖先遺伝子の重複、新規創生または遺伝子水平伝播等複数の報告がある。

新規遺伝子誕生の原動力となる現象として遺伝子重複が挙げられる。また、新規遺伝子は DNA を介するだけでなく、レトロ転移を経て誕生することもある。哺乳類において、レトロ転移はレトロトランスポゾンの LINE1 転移機構により行われ、イントロンとプロモーターを欠いた retrocopies を生成する。さらに、retrocopies の中でも CDS を保持している retrogenes が見出され、さらに哺乳類とショウジョウバエはその多くを有しているということが明らかになっている。祖先遺伝子 mRNA を鋳型とした cDNA がゲノムに挿入されて形成される retrocopies は、挿入された直近領域に存在している遺伝子プロモーターを利用するかもしくは、周辺ゲノム配列が変化したプロモーターやエクソンを利用していると考えられる。これは retrocopies が配列により周辺配列を変化させる可能性がある。

本研究では、類人猿 5 系統（ヒト、チンパンジー、ゴリラ、オランウータンとギボン）のみが有するレトロ遺伝子 *PIPSL* を対象として、レトロ転移により失われたプロモーターの獲得とその配列が系統間でどのように異なっているのかを明らかにすることを目的とした。*PIPSL* は類人猿の共通祖先において、レトロ転移により生じたと考えられている。その配列は 1 番染色体上に隣接して存在する 2 つの遺伝子（*PIP5K1A*: リン脂質キナーゼ, *S5a*: 26S プロテアソームサブユニット）が転写終結点の読み過ごしにより、タンデムに連結し mRNA に転写され、LINE1 の転移機構によるレトロ転移により 10 番染色体に挿入されて誕生したと考えられている。タンパクコーディング領域 (CDS) はチンパンジーとオランウータンで完全に保存されており、ヒトにおいては第 1 開始コドンと第 2 開始コドンの間に 1 塩基欠損が生じているため、第 2 開始コドンを採用すると考えられている。また、ゴリラは 2 つの祖先遺伝子に由来する領域に多数の有害変異を蓄積しており、ギボンは上流の領域に相当する *PIP5K1A* 由来領域に有害変異が蓄積している。

*PIPSL* に関して既知のことは、ヒトとチンパンジーの精巣で強い RNA 発現を示すことと *PIPSL* タンパク質の一部が発見されていることである。そこで、類人猿で最も早くに分岐したギボン（シロテテナガザル）の精巣組織を用いて RT-PCR を行い、内在性 *PIPSL* RNA を検出した。ポジティブコントロールにはチンパンジーの精巣を用いた。これにより、ヒト、チンパンジーとギボンの 3 系統で内在性 *PIPSL* RNA が発現していることを確認した。また、*PIPSL* 上流領域配列のプロモーター活性を調べるために、ヒトゲノム DNA を用いてレポーターアッセイを行った。*PIPSL* が精巣で特に発現が高いことを踏まえて、ヒト精巣がん由来細胞株 NEC8 と、プロモーターとして他の組織で転写活性があることを確認するために、ヒト肝がん由来細胞株 HepG2 とヒト子宮頸がん由来細胞株 HeLa を使って実験を行った。*PIPSL* の *PIP5K1A* 由来配列と *PIPSL* が挿入される以前から存在する複数の因子を含む *PIPSL* 上流約 600 bp の配列の転写活性を調べた。*PIPSL* 上流領域には、ヒトとチンパンジーとゴリラの共通祖先で誕生したと考えられる TATA 様配列 (TATAAA) が存在する。4 領域の転写活性を測定した(1: 実験対象領域全長、2: TATA 様配列、3: 起源遺伝子 *PIP5K1A* と同じ配列 (TGTAAG) に置換した配列、4: 起源遺伝子 *PIP5K1A* の上流領域の部分配列)。置換した配列の活性は全長領域の転写活性より有意に減少した。この結果から、*PIPSL* 転写に重要なのはこの TATA 様配列である可能性が高まった。さらに、公開されている RNA-seq データ (ヒトとオランウータンの精巣) とギボン精巣 (シロテテナガザル) を用いた RNA-seq データの解析を行なった。ギボンのデータは公開データベースでもほとんどなく、非常に珍しい。RNA-seq 解析データをもとに *PIPSL* の転写開始点の推定を行った。ヒト、オランウータンとギボンの転写開始点候補を見出し、ヒトとギボンにおいて、その位置が非常に近いことを明らかにした。この発見は、*PIPSL* がギボンの分岐前の類人猿共通祖先の早い段階でプロモーターを獲得し、すべての系統で同様の転写開始点を保持していた可能性を示唆している。その上、*PIPSL* の上流配列に相当する親遺伝子の *PIP5K1A* 5'UTR の部分配列とその近傍の反復配列はプロモーター活性を保持している可能性がある。最後に、TATA-box 様配列がヒトに近い系統において塩基置換を生じ、これによりヒト *PIPSL* 転写開始点の広範な分布に寄与している可能性を見出した。

# 目次

1	序論	6
1-1	研究背景	6
1-2	ヒト上科霊長類 5 系統における <i>PIPSL</i> 進化様式	9
1-3	<i>PIPSL</i> の翻訳	11
1-4	ヒト組織における <i>PIPSL</i> RNA の発現量	13
1-5	起源遺伝子と異なる発現パターン	16
1-6	<i>PIPSL</i> 遺伝子周辺のゲノム構造	17
2	材料および研究方法	20
2-1	ヒト DNA	20
2-2	レポータープラスミドの構築	20
2-2-1	PCR	20
2-2-2	制限酵素処理	21
2-2-3	Ligation	23
2-2-4	Transformation	23
2-2-5	プラスミド単離	23
2-3	<i>PIPSL</i> の発現解析	24
2-3-1	細胞培養	24
2-3-2	Transfection	24
2-4	チンパンジー RNA およびギボン RNA	27
2-5	Total RNA 抽出	27
2-6	RT-PCR	27
2-7	RNA-seq 解析 (転写開始点候補推定)	30
3	結果	31
3-1	NEC8 内在性 <i>PIPSL</i> RNA 検出 (NEC8: 睪丸由来奇形腫細胞)	31
3-2	シロテテナガザル精巣における <i>PIPSL</i> RNA 検出	32
3-3	RNA-seq 解析による転写開始点推定	33
3-4	<i>PIPSL</i> 上流領域配列を用いたプロモーターアッセイ	34
3-4-1	<i>PIPSL</i> 上流全域	34

3-4-2	<i>PIPSL</i> 転写開始点近傍領域のプロモーターアッセイ .....	35
3-4-3	霊長類間で高度に保存された配列を含む領域のプロモーターアッセイ .....	37
4	考察 .....	39
4-1	レトロ遺伝子 <i>PIPSL</i> が初期に獲得したプロモーター活性.....	39
4-2	霊長類 <i>PIPSL</i> の機能に関する考察 .....	40
5	参考文献 .....	42
6	補足資料 .....	48
7	謝辞 .....	57

# 1 序論

## 1-1 研究背景

レトロ遺伝子は様々な遺伝子を起源とし、将来的に新規遺伝子となり得る可能性を秘めた配列であり、プロセス型偽遺伝子に類似したメカニズムで誕生すると考えられている (Brosius 1999, Wang 2004, Casola and Betran 2017, Kubiak and Makalowska 2017)。しかし、その詳細なメカニズムや生物学的な機能は未だ明らかになっていない。

プロセス型偽遺伝子は mRNA を鋳型とした cDNA コピーがゲノムに挿入されて誕生する (Vanin 1985, Weiner, Deininger et al. 1986)。特に哺乳類ゲノムに大量に存在することが報告されている (Ohshima, Hattori *et al.* 2003, Zhang, Harrison *et al.* 2003, Abyzov, Iskow *et al.* 2013, Ewing, Ballinger *et al.* 2013, Schrider, Navarro *et al.* 2013, Zhang 2013, Kabza, Ciomborowska *et al.* 2014, Navarro and Galante 2015, Wang 2017)。その配列は転写に必要なプロモーターが欠落しているため基本的に、転写されない。そのため、生物学的な機能はないとみなされてきた。

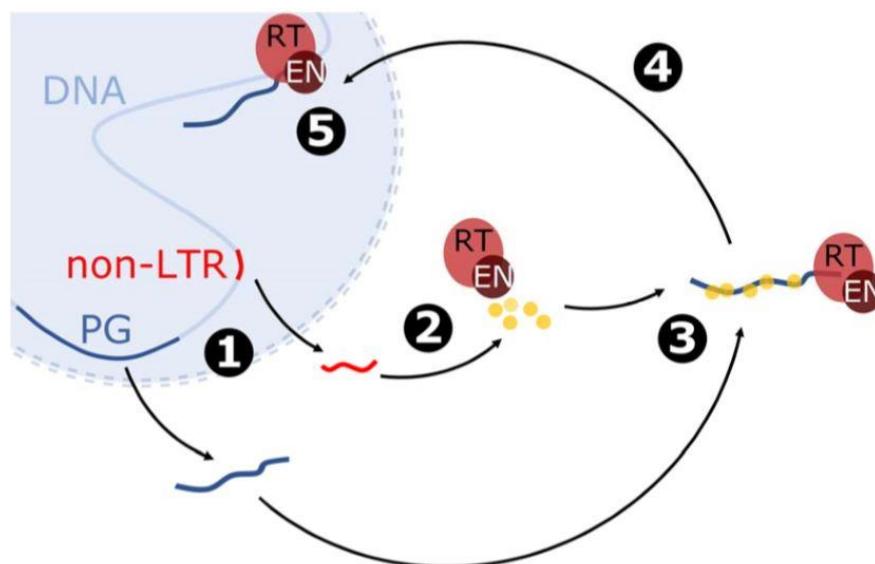


図 1. プロセス型偽遺伝子 (Processed Pseudogene) の生成機構

PG : Parental Gene, non-LTR : LINE1, RT : Reverse Transcriptase, EN : Endonuclease

(Kubiak and Makalowska, 2017)

プロセス型偽遺伝子の他に、2種類の偽遺伝子が存在する。1つは単一の遺伝子が偽遺伝子化した配列である。1例として、哺乳類のゲノムにおいて、嗅覚に関与する遺伝子ファミリーはゲノム中で大きな遺伝子ファミリーを構成している。しかし、霊長類は視覚に依存した進化を経たため、新世界サルが分岐した後、つまり類人猿と旧世界サルの共通祖先において、多くの嗅覚関連遺伝子機能を失ったことが報告されている。しかしながら、配列が消失したのではなく単一遺伝子が偽遺伝子化したことにより機能喪失を生じたことが明らかとなっている(Matsui, Go et al. 2010, Zhang, Frankish et al. 2010)。

2つめはDNA重複(非プロセス型)偽遺伝子である。(Torrents, Suyama et al. 2003)。ゲノム配列の重複により生じる一部の遺伝子コピーが、突然変異の蓄積によりその機能を失うことにより生じる偽遺伝子である。

偽遺伝子の多くは進化過程で有害な変異を多く蓄積しているため、それらの多くが転写活性やタンパク機能を失っていると考えられている(Mighell, Smith et al. 2000, Douglas, Wilson et al. 2016)。しかし、近年増加している研究によると、プロセス型偽遺伝子が獲得した制御因子によって転写されているという報告があり(Harrison, Zheng et al. 2005, Sakai, Koyanagi et al. 2007, Sorourian, Kunte et al. 2014)、一部は新規のタンパク質をコードしている可能性がある(Shashidharan, Michaelidis et al. 1994, Betran, Wang et al. 2002, Betran and Long 2003, Rosso, Marques et al. 2008, Parker, VonHoldt et al. 2009, Young, Menetrey et al. 2010, Ciomborowska, Rosikiewicz et al. 2013, Abdelsamad and Pecinka 2014)。他の偽遺伝子の一部は親遺伝子の発現制御に関与している(Poliseno, Salmena et al. 2010, Cheetham, Gruhl et al. 2013, Ha, Song et al. 2014, Hirano, Iwasaki et al. 2014)。また、レトロ遺伝子が獲得したプロモーターについて、いくつかのパターンが存在することを明らかにした報告もある(Carelli, Hayakawa et al. 2016)。

*PIPSL*は特異なメカニズムを経て誕生したと考えられている。*PIPSL*は異なる遺伝子(*PIP5K1A*: Gene ID: 8394 と *PSMD4*: Gene ID: 5710)の機能ドメインがRNAレベルで融合し、生じたキメラ mRNA(Akiva, Toporik et al. 2006)がLINE1転移機構(Babushok, Ohshima et al. 2007, Doucet-O'Hare, Rodic et al. 2015)により逆転写とゲノムへの挿入を経て誕生したと考えられている。

親遺伝子である PIP5K1A は phosphatidylinositol 4-phosphate (PtdIns4P) をリン酸化し、phosphatidylinositol 4,5-bisphosphate (PtdIns(4,5)P2) を生成するリン脂質キナーゼである。あらゆる組織で発現しており、神経系と生殖系の臓器において発現が特に高い (Hasegawa H, et al. 2012)。PSMD4 はユビキチン化タンパク質を分解する 26S プロテアソームである。*PSMD4* は HGNC (HUGO Gene Nomenclature Committee) における名称であり、統一名称は *Rpn10*、ヒト別称は *S5a* である。

親遺伝子と異なる細胞局在であり、非常に弱いキナーゼ活性である一方で、細胞のユビキチン化タンパクに有意な親和性を示すキメラタンパクをコードしている (Babushok, Ohshima et al. 2007)。*PIPSL* 配列は現存のヒト上科霊長類の共通祖先ゲノムで誕生し (Ohshima and Igarashi 2010)、その遺伝子座はヒト集団で保存されておりアミノ酸変化を伴う新規ハプロタイプを有している (Ohshima and Igarashi 2010)。

先行研究において、ヒトとチンパンジーの精巣で、*PIPSL* RNA の高い発現を確認している (Babushok, Ohshima et al. 2007, Zhang, Lu et al. 2009)。しかしながら、*PIPSL* の部分的なアミノ酸配列がプロテオームデータベースで入手可能にもかかわらず (Ohshima and Igarashi 2010)、完全な *PIPSL* タンパクは検出されなかった (Babushok, Ohshima et al. 2007)。*PIPSL* は隣接する遺伝子から離れた遺伝子間領域に挿入されたので、進化の早い段階で自身の転写制御メカニズムを獲得した可能性がある。本研究では、そのプロセスを明らかにするために、ヒトとチンパンジー以外の種における *PIPSL* の発現を調べプロモーターとして可能性のある配列を探した。

*PIPSL* について既知の事柄を詳細に説明する。



## 1-2 ヒト上科霊長類5系統における *PIPSL* 進化様式

*PIPSL* はヒト上科霊長類の共通祖先で生成された後、各系統で独自の配列進化を経て今に至る。ヒトは *PIPSL* 配列前半に当たる *PIP5K1A* 由来領域の第1と第2開始コドンの間にフレームシフトが生じている。後半の *S5a* (*PSDM4*) 由来領域に変異は見られない。チンパンジーとオランウータンは融合構造を完全に保持しており、有害変異はない。ゴリラは両領域 (*PIP5K1A* と *S5a* (*PSMD4*)) に有害変異を蓄積しており、タンパク機能は失活していると推測される。ギボンは *PIP5K1A* 由来領域に有害変異を蓄積しているが *S5a* (*PSMD4*) 由来領域は完全に保存されている (Ohshima and Igarashi 2010) (図2)。

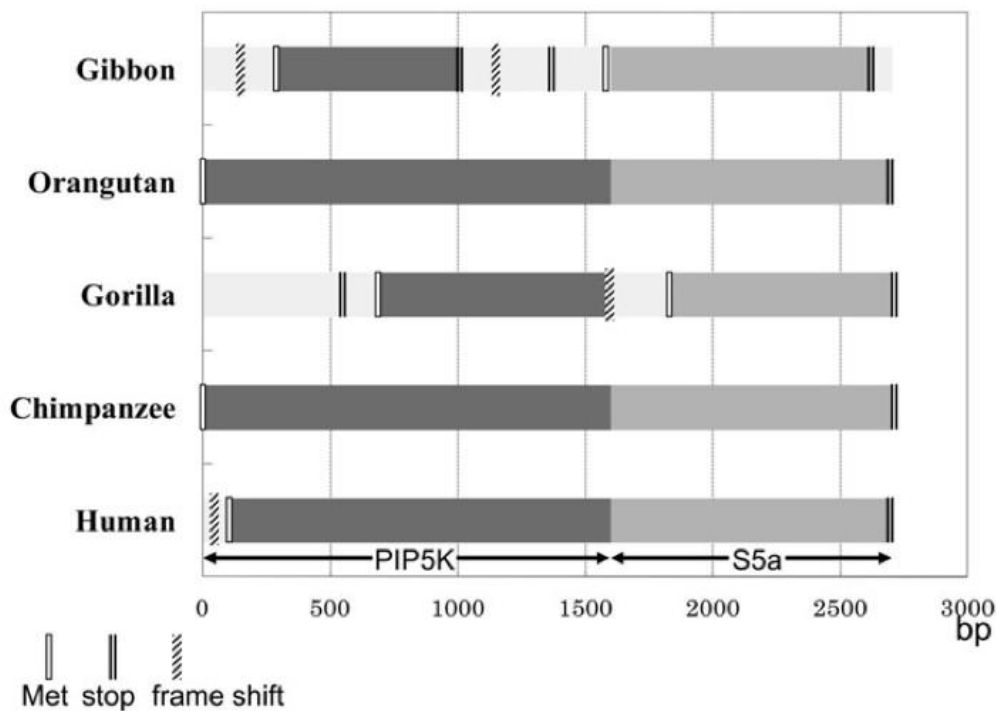


図2. *PIPSL* 配列の保存性

PIP5K : *PIP5K1A* (リン脂質キナーゼ) の exon1~13 に由来する配列

S5a : *PSMD4/S5a* (26S プロテアソームサブユニット) の exon2~10 に由来する配列

ヒト : 他の生物種より約 100 塩基下流から翻訳開始

チンパンジー、オランウータン : 融合構造完全保存

ゴリラ : 2つの起源遺伝子に由来する領域にナンセンス変異とフレームシフト有

ギボン : *S5a* (*PSMD4*) 由来領域のみ保存

(Ohshima and Igarashi, 2010)

PIPSL は生成されてからヒトとチンパンジーの共通祖先までの期間及びギボン系統において、*S5a* 由来領域の非同義置換が同義置換を有意に超過していることが報告されている(図 3)。

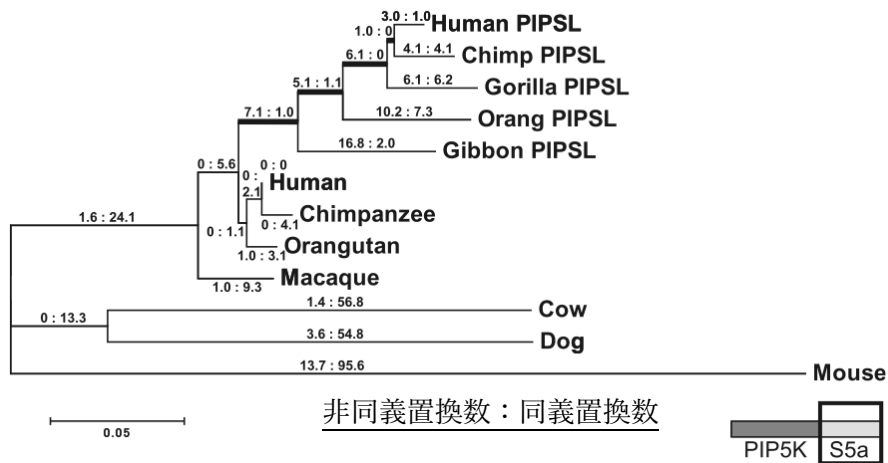


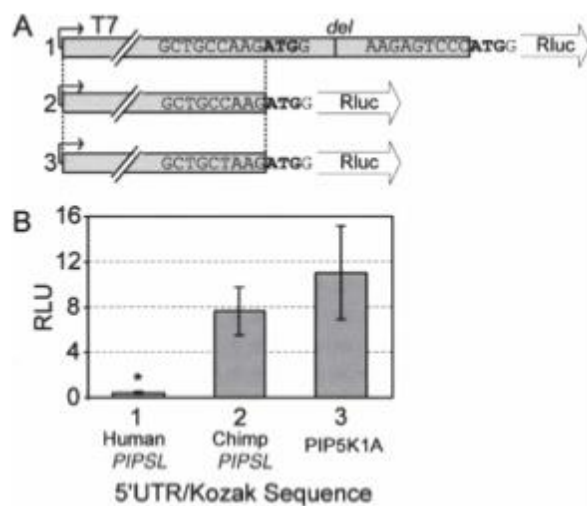
図 3. *S5a* 由来領域の分子系統樹と塩基置換様式

各種 PIPSL は *S5a* 由来領域における非同義置換と同義置換の比を示し、PIPSL の記載がない種名の枝は親遺伝子である *PSMD4/S5a* の非同義置換と同義置換の比を示す。

(Ohshima and Igarashi, (2010)の図を改変)

### 1-3 PIPSLの翻訳

先行研究において、*Renilla* ルシフェラーゼレポーターコンストラクトの *Renilla* Kozak 配列をヒト PIPSL 5'-UTR/Kozak 領域に置き換えて、転写及び翻訳効率を測定した報告がある (Babushok *et al.*, 2007)。PIP5K1A (起源遺伝子) の翻訳効率を 100%としたとき、チンパンジーPIPSL の翻訳効率は 69.5%、



A:

1. ヒト PIPSL
2. チンパンジーPIPSL
3. PIP5K1A

それぞれ5'-UTR/Kozak 領域に定量のためルシフェラーゼ (Rluc) を連結している。

B:

縦軸はルシフェラーゼの相対発光量を示している。

図 4. PIPSL 翻訳効率

(Babushok DV, *et al.*, (2007)の図を改変)

ヒト PIPSL の翻訳効率は 3.6%であることが明らかとなっている(図 4)。これはヒトの PIPSL では、起源遺伝子に由来する ORF 開始直後にフレームシフトが生じているため、起源遺伝子と同じ開始コドンが使用できず第 2 開始コドンが使用される必要があるためと考えられている。また、PIPSL タンパク質のアミノ酸部分配列がヒト生体内で確認されている (Ohshima and Igarashi, 2010)(図 5)。

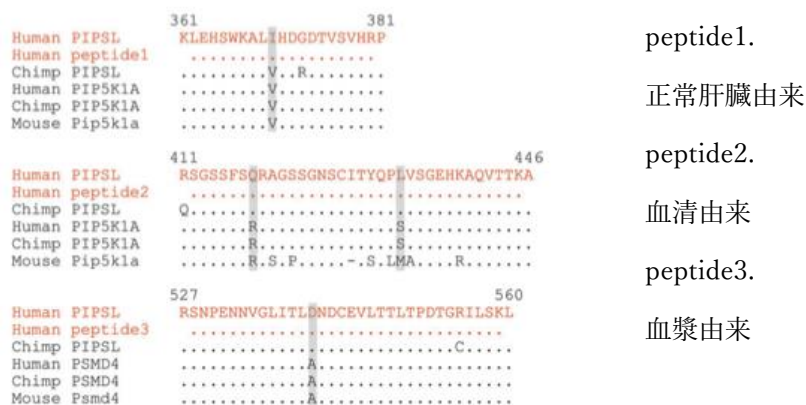


図 5. 同定された PIPSL タンパク質の部分配列と起源遺伝子配列比較 (Ohshima and Igarashi, 2010 の図を改変)

#### 1-4 ヒト組織における *PIPSL* RNA の発現量

多様な生物種のレトロ遺伝子およびその親遺伝子の RNA-seq から得られた発現プロファイルを格納したデータベースとして RetrogeneDB が知られている (Kabza *et al.*, 2014)。このデータベースからヒト *PIPSL* RNA の発現を確認した。*PIPSL* は精巣で高い発現を示すことが報告されている (Bubushok *et al.*, 2007)。精巣以外の他の組織でも *PIPSL* RNA が少量ながら発現していることが明らかになった (図 6)。ヒト組織における *PIPSL* と起源遺伝子 (*PIP5K1A* および *S5a*) の発現量を比較した (表 1)。また、ヒト、チンパンジー、ゴリラの *PIPSL* RNA の発現量を比較した (表 2)。*PIPSL* の RNA 量は起源遺伝子の発現量を大きく下回っていた (表 1)。ヒト、チンパンジー、およびゴリラの間で *PIPSL* RNA の発現量を比較すると、先行研究の報告通りチンパンジー *PIPSL* の発現量は、ヒト *PIPSL* の発現量を上回っていた (表 2)。各系統における *PIPSL* の起源遺伝子である *PIP5K1A* と *S5a* の発現量を比較したところ、起源遺伝子は各系統で *PIPSL* よりも高い発現量を示した (表 3, 4) (補足資料図 1-10)。

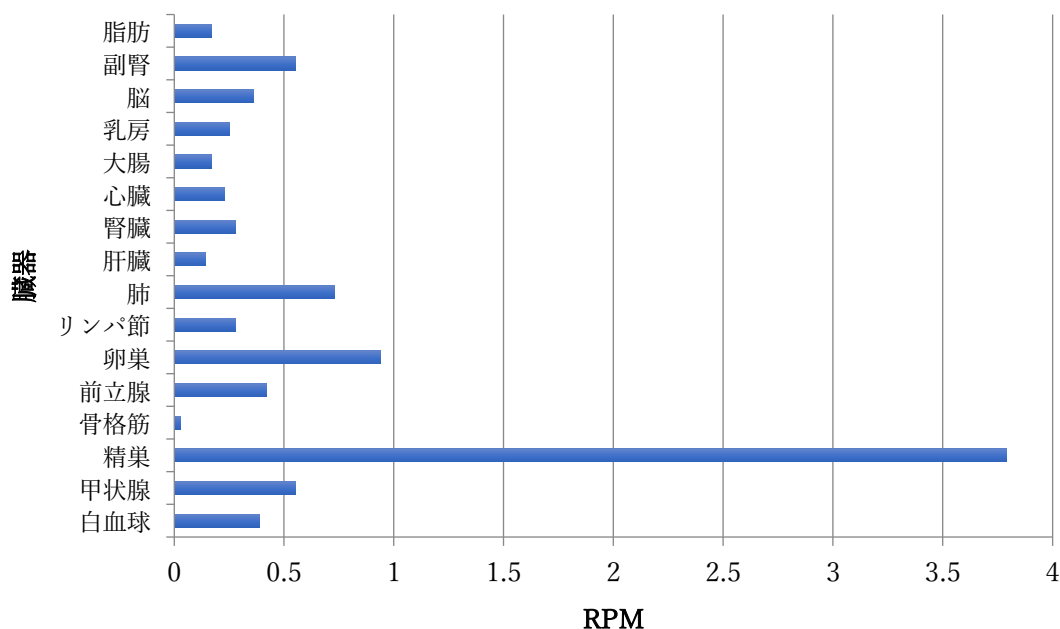


図 6. ヒト組織における *PIPSL* RNA 発現量

RPM : Read Per Million (100 万本あたりの各遺伝子にマップされたリード数)

(Kabza M *et al.*, 2014)

表 1. ヒト精巣における *PIPSL* と親遺伝子との発現量比較

遺伝子名	RPM[比率]
<i>PIPSL</i>	3.79 [1.0]
<i>PIP5K1A</i>	31.21 [8.2]
<i>S5a</i>	47.31 [12.5]

RPM[Human を基準にした比率]

表 2. 各系統における *PIPSL* RNA 発現量比較

生物種名	RPM[比率]
Human	3.79 [1.0]
Chimpanzee	23.66 [6.2]
Gorilla	4.38 [1.2]
Orangutan	データ無し

RPM[Human を基準にした比率]

表 3. 各系統の *PIP5K1A* (親遺伝子) の発現量比較

臓器	Human	Chimpanzee	Gorilla	Orangutan
心臓	14.32 [1.00]	5.32 [0.37]	0.33 [0.02]	6.17 [0.43]
腎臓	17.8 [1.00]	13.8 [0.78]	0.85 [0.05]	9.0 [0.51]
肝臓	8.4 [1.00]	7.31 [0.87]	0.4 [0.05]	7.6 [0.90]
精巣	31.21 [1.00]	65.89 [2.1]	7.84 [0.25]	データ無し

RPM[Human を基準にした比率]

RPM : Read Per Million

次世代シーケンサーから得られる総リード数の 100 万本あたり各遺伝子にマップされたリード数

表 4. 各系統の *S5a* (親遺伝子) の発現量比較

臓器	Human	Chimpanzee	Gorilla	Orangutan
心臓	19.4 [1.00]	34.62 [1.68]	データ無し	19.64 [1.01]
腎臓	22.53 [1.00]	42.79 [1.90]	データ無し	32.99 [1.46]
肝臓	24.45 [1.00]	42.79 [1.75]	データ無し	34.13 [1.40]
精巣	47.31 [1.00]	218.98 [4.63]	データ無し	データ無し

RPM[Human を基準にした比率]

RPM : Read Per Million

次世代シーケンサーから得られる総リード数の 100 万本あたり各遺伝子にマップされたリード数

## 1-5 起源遺伝子と異なる発現パターン

精巣は他の組織に比べて RNA 発現に許容的であり、他の組織で転写される遺伝子が精巣でも転写される現象（転写産物の漏出）が起りやすいと報告されている（Schmidt, 1996）。しかしながら、精巣は偽遺伝子の発現に必ずしも許容的ではないことが明らかとなっている（Zhang, 2009）。*PIPSL* はプロセス型偽遺伝子と同じ機構でゲノムに挿入されて誕生しており、Zhang 等の結果は *PIPSL* の転写も転写産物の漏出ではないことを示す報告である。また、同報告で、ヒトゲノム全体のメチル化データ解析（Weber, 2007）から、*PIPSL* 領域の CpG は始原肺線維芽細胞では高度にメチル化されている一方で、精子では低メチル化状態であり、生殖細胞系列に特異的な遺伝子メチル化パターンと同じ傾向を示すことにも言及している。

表 5. Exon-array による偽遺伝子発現プロファイル

Tissue	Expression intensity of Pseudogenes <sup>A</sup>	Number of pseudogenes with highest transcription <sup>B</sup>	Number of transcribe-able ( $\geq 0.2$ ) pseudogenes <sup>C</sup>
	A	B	C
Breast	0.00	42	39
Cerebellum	0.00	58	63
Heart	0.02	74	54
Kidney	0.00	58	64
Liver	0.01	53	43
Muscle	0.01	58	42
Pancreas	0.00	57	56
Prostate	-0.02	24	43
Spleen	0.00	37	31
Testis	0.00	51	38
Thyroid	0.00	39	39

A: 各組織での全偽遺伝子の発現強度の中央値（発現の中央値より上を正の値、下を負の値で示している。Log 比率）

B: 対象組織で最も RNA 発現が多い偽遺伝子数

C: EST との発現の相関関係を示す指標が 0.2 以上の偽遺伝子を用いた場合 (Zhang *et al.*, 2009)



## 1-6 PIPSL 遺伝子周辺のゲノム構造

PIPSL の上流には *PCLE1* 遺伝子、下流には *SLC35G1* 遺伝子が存在している (図.7)。PLCE1 はホスファチジルイノシトール-4, 5-二リン酸の加水分解を触媒するホスホリパーゼ C である。加水分解の結果、細胞内 2 次メッセンジャーとして機能するイノシトール 1,4,5-三リン酸 (IP3) とジアシルグリセロール (DAG) を生成する。SLC35G1 は基質トランスポーターをコードしている。PIPSL の起源遺伝子 PIP5K1A はホスファチジルイノシトール-4-リン酸をリン酸化し、ホスファチジルイノシトール-4, 5-二リン酸を生じるキナーゼである。*PLCE1* と *SLC35G1* は様々な組織で発現している。また、*PIPSL* 近傍には精巣特異的発現を示す遺伝子は存在しない。

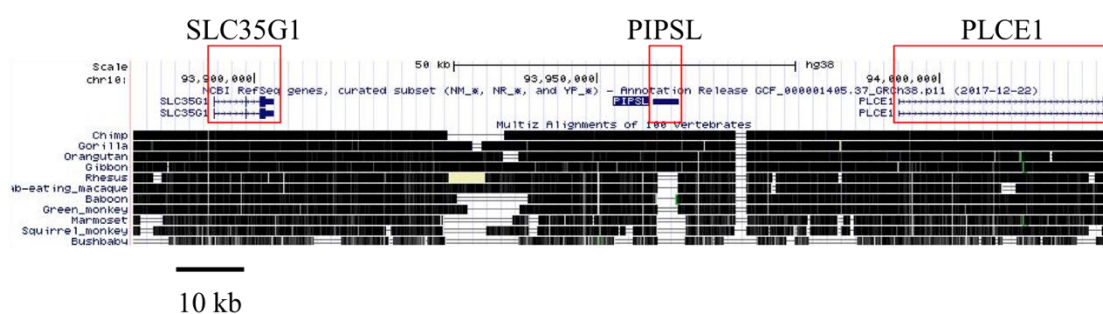


図 7. *PIPSL* 周辺のゲノム構造

(Rosenbloom KR, *et al.*, 2015)

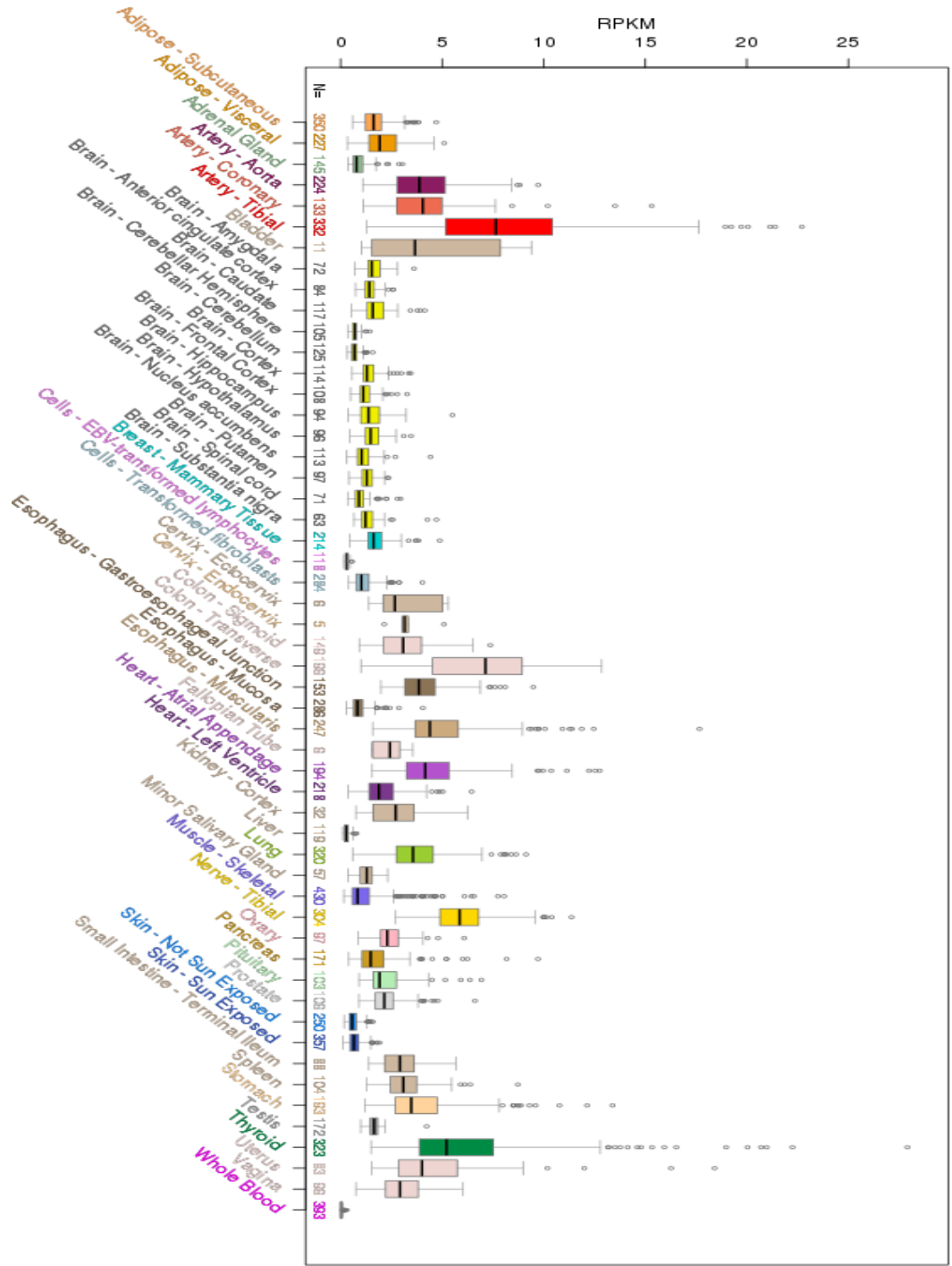


図 8. Human PLCE1 発現プロファイル (RNA-seq data)

RPKM : Reads Per Kilobase of exon per Million mapped reads

取得したリードカウントデータに対して、総リード数による補正後、転写産物長による補正を行って得られるデータ (single-end read は RPKM、paired-end read は FPKM) (Rosenbloom KR, *et al.*, 2015)

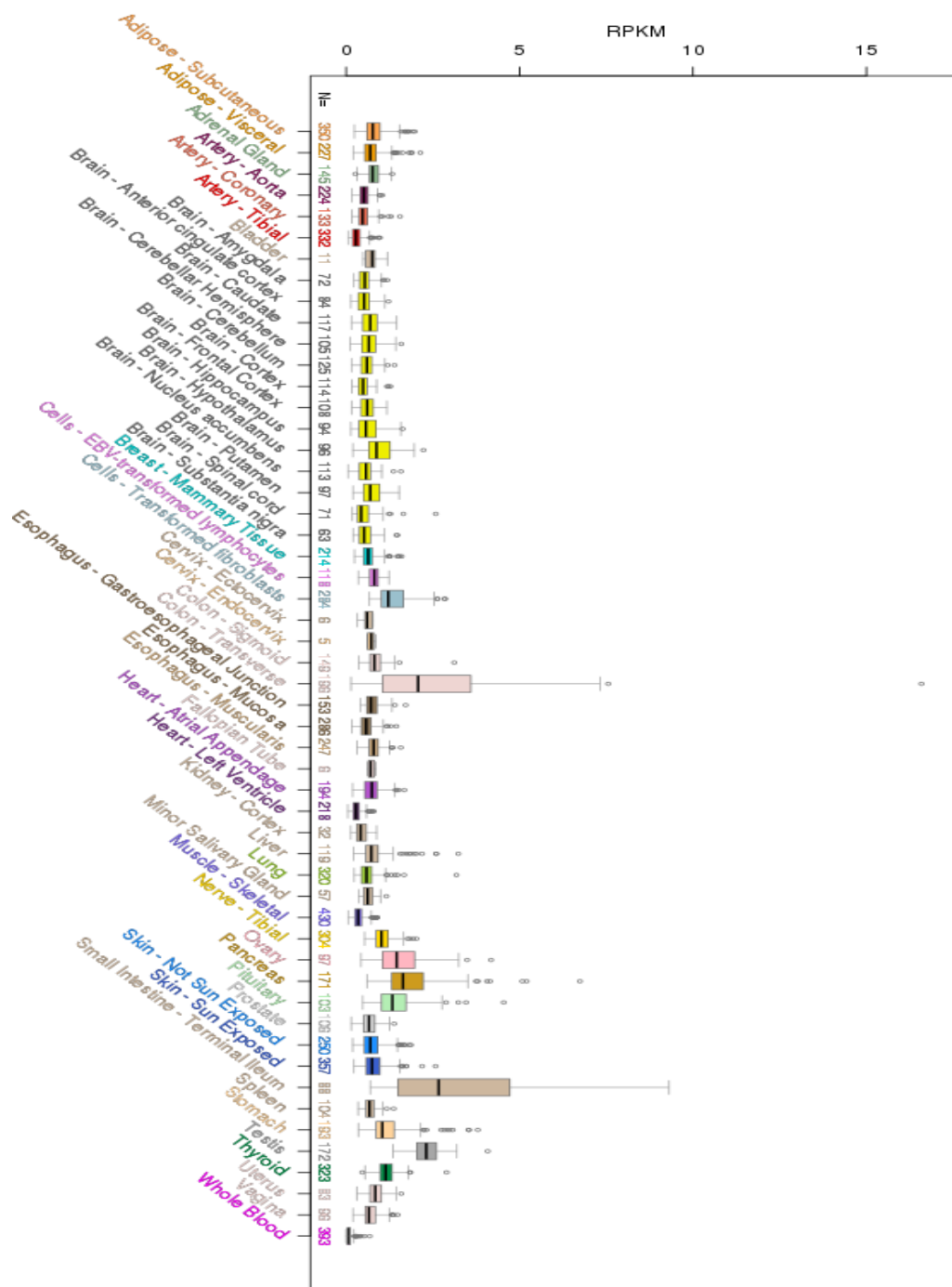


図 9. Human SLC35G1 発現プロファイル (RNA-seq data)

RPKM : Reads Per Kilobase of exon per Million mapped reads

取得したリードカウントデータに対して、総リード数による補正後、転写産物長による補正を行って得られるデータ (single-end read は RPKM、paired-end read は FPKM)

(Rosenbloom KR, *et al.*, 2015)

## 2 材料および研究方法

### 2-1 ヒト DNA

財団法人ヒューマンサイエンス振興財団ヒューマンサイエンス研究資源バンクより分譲を受けた日本人のゲノム DNA (PSCDA0303 33F) を PCR 反応の鋳型に用いた。

### 2-2 レポータープラスミドの構築

#### 2-2-1 PCR

PIPSL の上流領域をクローニングするために、以下に示す組成、プログラムとプライマーを用いて PCR 反応を行った。PCR 反応には TaKaRa 社 Tks Gflex DNA polymerase 及び付随する緩衝液を使用した。サーマルサイクラーは MJ Research 社 PTC-100 を用いた。反応終了後、Min Elute PCR Purification Kit (QIAGEN) を使用して PCR 産物の精製を行った。操作は付属のプロトコールに従った。

滅菌水		プログラム
2 × Gflex PCR Buffer (Mg <sup>2+</sup> , dNTP Plus)	25 μl	1. 94°C, 1min
dNTP mixture (2.5 mM each)	2 μl	2. 98°C, 10sec
Primer F (50 pmol / μl)	0.5 μl	3. 50°C, 15sec
Primer R (50 pmol / μl)	0.5 μl	4. 68°C, 30sec
Human Genomic DNA (200 ng)		5. go to 2, for 35 cycles
Tks Gflex DNA polymerase (1.25 units / μl)	1 μl	6. 4°C, ∞
Total	50 μl	

各領域増幅に用いたプライマー

増幅領域	プライマー名	配列
95,721,300	5'_DNATrans_F-2	5'- <u>GACTAGTATCTCCTATTCGTTCTG</u> -3'
95,722,124	3'_DNATrans_R1	5'- <u>TAAGAATTCTCCGGCTCTTCCTCTTC</u> -3'
95,721,300	5'_DNATrans_F-1	5'- <u>GACTAGTACAGCAGGCATTTTCATC</u> -3'
95,722,082	3'_DNATrans_R1	5'- <u>TAAGAATTCTCCGGCTCTTCCTTCTTC</u> -3'
95,721,876	5'_DNATrans_F-2	5'- <u>GACTAGTATCTCCTATTCGTTCTG</u> -3'
95,722,124	3'_DNATrans_R2	5'- <u>TAAGAATTCCGTGATGGTTGCCAGTCT</u> -3'

増幅領域はヒトゲノム GRCh37/hg19 に対応する。また、下線部は付加した制限酵素認識配列である。

ヒトゲノム (GRCh37/hg19) の 95,721,300~95,722,124 の配列をレポーターアッセイに用いた。プライマーのペアは、5'\_DNATrans\_F-2 primer と 3'\_DNATrans\_R1 primer、5'\_DNATrans\_F-1 と 3'\_DNATrans\_R1、5'\_DNATrans\_F-2 と 3'\_DNATrans\_R2 である。

## 2-2-2 制限酵素処理

最初に、制限酵素 Spe I (TaKaRa 社) を使用し、以下に示す組成で制限酵素処理を 37°C で 1 時間行った。同時にクローニングに用いるプラスミドベクターとして使用した pMCS-*Gaussia* – Dura Luc Vector (Thermo Fisher) も制限酵素処理を行った。

### 反応液組成 (PCR 産物)

滅菌水

10×M Buffer                      2 μl

PCR 産物 (< 1μg)

Spe I (10 units / μl)              1 μl

---

Total                                      20 μl

#### 反応液組成 (プラスミドベクター)

滅菌水

10×M Buffer 2  $\mu$ l

pMCS-*Gaussia* – Dura Luc Vector (Thermo Fisher)

Spe I (10 units /  $\mu$ l) 1  $\mu$ l

---

Total 20  $\mu$ l

制限酵素反応液の入ったチューブを軽く遠心し、125 mM EDTA 2  $\mu$ l、100% EtOH 50  $\mu$ l、3M CH<sub>3</sub>COONa 2  $\mu$ l を加えてタッピングし、室温で 30 分置いた。その後、13,000 rpm、4°Cで 20 分間遠心し、上清を除いて 70% EtOH 100  $\mu$ l を加え、13,000 rpm、4°Cで 10 分間遠心し上清を除いた。13,000 rpm、4°Cで 1 分間遠心後、上清を完全に除き室温で 15 分間乾燥させ、TE Buffer 6  $\mu$ l を加えた。続いて、制限酵素 EcoR I (TaKaRa) を使用し、以下に示す組成で制限酵素処理を 37°Cで 1 時間行った。

#### 反応液組成 (PCR 産物)

滅菌水

10×H Buffer 1  $\mu$ l

PCR 産物 (< 1 $\mu$ g)

EcoR I (15 units /  $\mu$ l) 1  $\mu$ l

---

Total 10  $\mu$ l

#### 反応液組成 (プラスミドベクター)

滅菌水

10×H Buffer 1  $\mu$ l

pMCS-*Gaussia* – Dura Luc Vector (Thermo Fisher)

EcoR I (15 units /  $\mu$ l) 1  $\mu$ l

---

Total 10  $\mu$ l

制限酵素反応液の入ったチューブを軽く遠心し、125 mM EDTA 1  $\mu$ l、100% EtOH 25  $\mu$ l、3M CH<sub>3</sub>COONa 1  $\mu$ l を加えてタッピングした。室温で 30 分間放置後、13,000 rpm、4°Cで 20 分間遠心した。上清を除き 70% EtOH 100  $\mu$ l を加え、13,000 rpm、4°Cで 10 分間遠心し上清を除いた。13,000 rpm、4°Cで 1 分間遠心後、上清を完全に除き室温で 15 乾燥させ、TE Buffer 4.5  $\mu$ l を加えた。

### 2-2-3 Ligation

Ligation Kit <Mighty Mix> (TaKaRa 社) の Ligation Mix を使用し、以下の組成で Ligation 反応を 16°Cで 30 分間行った。vector には Thermo Fisher 社の pMCS-*Gaussia* – Dura Luc Vector を使用した。

pMCS – <i>Gaussia</i> – Dura Luc Vector	0.5 $\mu$ l
Insert DNA	4.5 $\mu$ l
Ligation Mix	5.0 $\mu$ l
<hr/>	
Total	10 $\mu$ l

### 2-2-4 Transformation

TaKaRa 社 *E. coli* DH5  $\alpha$  Competent Cells を氷上で融解し、穏やかに混和した。Ligation 反応液を全量加え氷上で 30 分間放置した後、42°Cで 45 秒ヒートショックを行った。氷上で 2 分間放置し、SOC 培地を 900  $\mu$ l 加え 37°Cで 1 時間培養した。13,000 rpm、4°Cで 2 分間遠心し上清を 900  $\mu$ l 除いた後、残りを懸濁し LB プレート (Amp : 100  $\mu$ g / ml) にプレーティングし、37°Cで一晩培養した。

### 2-2-5 プラスミド単離

コロニーをつつき、L 液体培地 (Amp : 100  $\mu$ g / ml) を 5ml 入れた培養管に入れ、37°Cで 8 時間振盪培養した。QIAprep Spin Miniprep Kit (QIAGEN) を使用し、プラスミド DNA の単離を行った。操作は付属のプロトコールに従った。

## 2-3 PIPSL の発現解析

### 2-3-1 細胞培養

本研究に用いたヒト精巣癌由来細胞株 NEC8 細胞は、ヒューマンサイエンス研究資源バンクより入手した (JCRB0250)。培地には GIBCO 社 RPMI 1640 に牛胎児血清 (FBS) を 10% 添加したものをを用いた。HeLa 細胞 P128 は伊藤正恵教授 (長浜バイオ大学) より譲り受けた。培地には SIGMA 社 DMEM に牛胎児血清 (FBS) 10% を添加したものをを用いた。ヒト肝癌由来細胞株 HepG2 細胞は RIKEN Bio Resource Center, Cell Bank より入手した。培地には SIGMA 社 DMEM に牛胎児血清 (FBS) 10% 添加したものをを用いた。いずれの細胞も 37°C、CO<sub>2</sub> 5% で培養し、継代時にはトリプシン (Nakarai 社 0.05% - Trypsin / 0.53 Mm – EDTA Solution) を用いて細胞を剥離した。

### 2-3-2 Transfection

Transfection 前日に 24 穴プレートに 80% コンフルエントになるように細胞を播種した。測定用プラスミドを 20 ng 分と内部補正用のコントロールプラスミドとして pCMV-*Cypridina* Luc Vector (Thermo Fisher 社) を 200 ng 分を培養細胞に co-transfection した。トランスフェクションから 24 時間経過後、96 穴プレート中の 2 穴に 24 穴プレートの 1 穴に由来する培養上清 20  $\mu$ l を各々分取した。*Gussia* Luciferase Glow Assay Working Solution あるいは *Cypridina* Luciferase Glow Assay Working Solution を分取した 2 穴に 50  $\mu$ l ずつ加え、反応させた。測定には Perkin Elmer 社の Fusion  $\alpha$ -FP を用い、波長は 485 nm を用いた。*Cypridina* Luciferase と *Gussia*-Dura Luciferase の波長は非常に近いが反応基質をそれぞれ異なるウェルで反応させて測定した。測定した発光強度を、以下に示した計算式を用いて解析した。いずれの実験も 3 ウェルの平均値を結果とした。



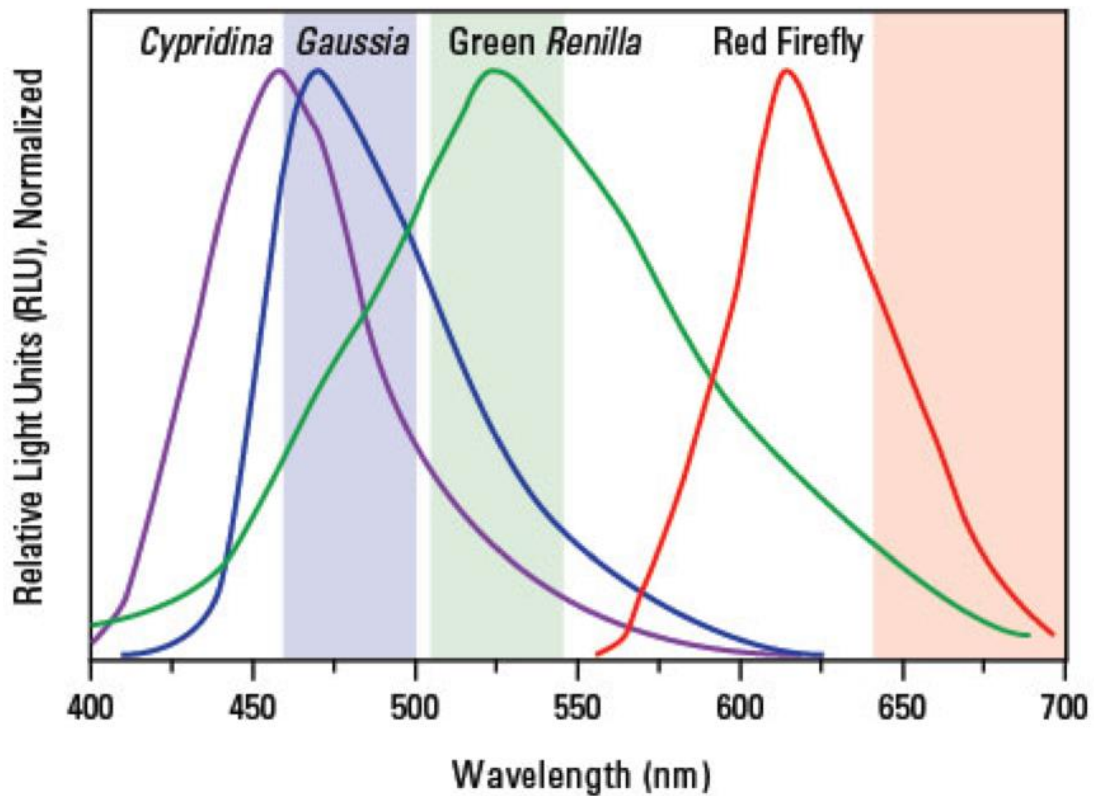


図 10. *Cypridina* Luc および *Gaussia* Luc の基質スペクトル

表 6. コントロールプラスミド

プラスミドの種類	挿入配列の有無	測定値 (発光強度)
<i>Cypridina</i>	—	C <sub>0</sub>
<i>Gaussia</i> -Dura	—	G <sub>0</sub>

表 7. 測定用プラスミド

プラスミドの種類	挿入配列の有無	測定値 (発光強度)
<i>Cypridina</i>	—	C <sub>1</sub>
<i>Gaussia</i> -Dura	+	G <sub>1</sub>

表 8. 計算式および発光強度を用いて算出した計算値

計算値	記号	計算式
コントロールプラスミドの相対発光強度	$X$	$G_0 / C_0$
コントロールプラスミドの相対発光強度の平均値	$\bar{X}$	$\frac{1}{3} \sum^3 X$
測定プラスミドの相対発光強度	$Y$	$G_1 / C_1$
測定プラスミドの相対発光強度の平均値	$\bar{Y}$	$\frac{1}{3} \sum^3 Y$
発光強度の比	$Z$	$Y / \bar{X}$
発光強度の比の平均値	$\bar{Z}$	$\frac{1}{3} \sum^3 Z$
コントロールプラスミドの標準偏差	$A$	$\sqrt{\frac{\sum^3 (X - \bar{X})^2}{3}}$
発光強度比の標準偏差	$B$	$\sqrt{\frac{\sum^3 (Z - \bar{Z})^2}{3}}$

## 2-4 チンパンジーRNA およびギボン RNA

RNA 抽出に用いたサンプル（精巢）は京都大学霊長類研究所および日本モンキーパークより提供を受けた。抽出した Total RNA は RT-PCR と RNA-seq 解析に用いた。

## 2-5 Total RNA 抽出

Total RNA 抽出には RNeasy Mini Kit (QIAGEN) を用いた。最初に、 $-80^{\circ}\text{C}$  で凍結保存した組織を 20 mg 切り出し、Buffer RLT を 350  $\mu\text{l}$  添加して氷上で破碎した。次に溶解液を 13,000 rpm で 3 分間遠心し、上清を新しいマイクロチューブに移した。移した上清に等量の 70% EtOH を添加してピペッティングにより混和した。RNeasy スピンカラムを 2ml コレクションチューブにセットし、混和液をアプライして 10,000 rpm で 15 秒遠心後、ろ液を破棄した。次に RNeasy スピンカラムに 500  $\mu\text{l}$  の Buffer RPE (100% EtOH 添加済) を添加し、10,000 rpm で 15 秒遠心後、ろ液を破棄した。さらに RNeasy スピンカラムに 500  $\mu\text{l}$  の Buffer RPE を添加し、10,000 rpm で 2 分遠心した。最後に、RNeasy スピンカラムを新しい 1.5ml コレクションチューブにセットし、RNase Free  $\text{dH}_2\text{O}$  を 50  $\mu\text{l}$  添加し、10,000 rpm で 1 分遠心して Total RNA を溶出した。

## 2-6 RT-PCR

NEC8 を使った RT-PCR には、ヒトとチンパンジーを対象とした *PIPSL* RNA の検出を目的としたプライマーを使用した (Babushok *et al.*, 2009)。理論長はヒトにおいて 575bp である。

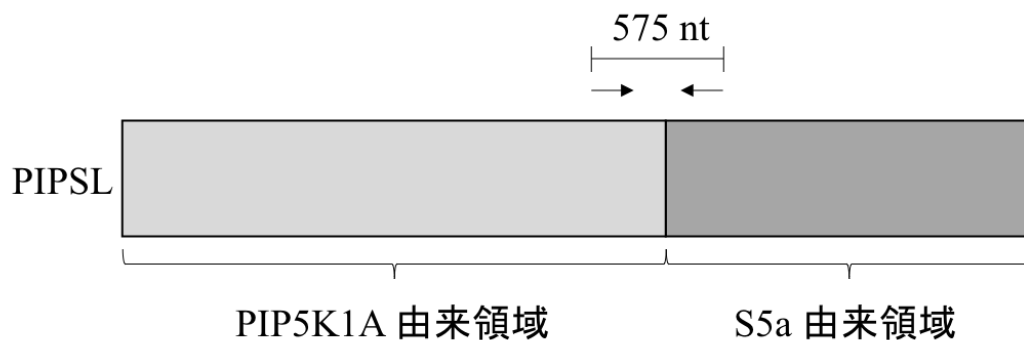


図 11. RT-PCR で増幅する領域

ギボンの *PIPSL* RNA の検出には 5 系統全てに共通して使用できる RT-PCR プライマーを使用した。ヒトにおける増幅領域は chr10:93,959,365–93,960,164 (GRCh38) である。理論長はヒトにおいて 800bp である。この 2 種の RT-PCR プライマーはともに両起源遺伝子の融合境界点を挟むように設計されている。

#### RT-PCR プライマー

プライマー名	配列
RT-PCR-F-primer	5' - GCCTTCTCCTTCCAAAAAAC - 3'
RT-PCR-R-primer	5' - TGATGCGGATCTTGTGATTG - 3'
RT-PCR-F-primer(5 種)	5' - GTCTGGCTCATCTTTCTCTCA - 3'
RT-PCR-R-primer(5 種)	5' - CTTACCAGCCAAAATCGGAA - 3'
GAPDH - F	5' - TTCACCACCATGGAGAAGG - 3'
GAPDH - R	5' - AGGGATGATGTTCTGGAGAG - 3'
PIP5K1A - F	5' - GCCTTCTCCTTCCAAAAAGT - 3'
PIP5K1A - R	5' - AAAATCCATACCACCCAACTCT - 3'
S5a - F	5' - GAGGAGTTGTTGTTAGGCCGTC - 3'
S5a - R	5' - TGATGCGCATCTTGTGATTC - 3'

Prime Script High Fidelity RT-PCR kit (TaKaRa 社) を用いて、以下に示す組成とプログラムで RT-PCR を行った。逆転写反応で Prime Script RTase を加えたサンプルを RT+ とし、代わりに RNase Free dH<sub>2</sub>O を加えたサンプルを RT- とした。

RNA 編成とアニーリング		プログラム
dNTP Mixture (10 mM each)	1 $\mu$ l	1. 65°C, 5min
Specific Primer (2 $\mu$ M)	1 $\mu$ l	2. 4°C
Template RNA (300 ng 分)	8 $\mu$ l	
Total	10 $\mu$ l	

逆転写反応		プログラム
反応液	10 $\mu$ l	1. 50°C, 30 min
5×Prime Script Buffer	4 $\mu$ l	2. 95°C, 5min
RNase Inhibitor (40 U/ $\mu$ l)	0.5 $\mu$ l	3. 4°C
Prime Script RTase	0.5 $\mu$ l	
RNase Free dH <sub>2</sub> O	5 $\mu$ l	
<hr/>		
Total	20 $\mu$ l	
PCR		プログラム
PrimeSTAR Max Premix (2×)	25 $\mu$ l	1. 98°C, 10sec
Primer-F (20 $\mu$ M)	0.5 $\mu$ l	2. 55°C, 5sec
Primer-R (20 $\mu$ M)	0.5 $\mu$ l	3. 72°C, 1min
逆転写反応液	5 $\mu$ l	4. go to 2, 30 cycles
滅菌水		5. 4°C, $\infty$
<hr/>		
Total	50 $\mu$ l	

## 2-7 RNA-seq 解析 (転写開始点候補推定)

公共データベース (NCBI, SRA) からヒト精巢の RNA sequencing データとオランウータン精巢の RNA sequencing データを取得した。また、RT-PCR に用いたギボン Total RNA (精巢) を使って RNA sequencing を行った。これらの RNA sequencing read を使って RNA-seq 解析を行った。リファレンスゲノムには GRCh38 (ヒト)、WUGSC 2.0.2/ponAbe2 (オランウータン) と Nleu\_3.0/nomLeu3 (ギボン) を用いた。ギボンのみ、公開されているリファレンスゲノムの種と本研究で扱った種は異なる。公開されているリファレンスゲノムはクロテナガザルであるのに対し、本研究で扱ったのはシロテナガザルである。シロテナガザルの complete genome は公開されていないため、同系統のギボンゲノムをリファレンスゲノムとして使用した。ギボンは9種存在しており、種間のゲノム相違率は平均 1.1% である。

リファレンスゲノムへのインデックス付与には bowtie2 を用いた。次に、取得した RNA sequencing read からアダプター配列を除去し、read quality 30 (Q30) 以上のリードのみを抽出した。これらのリードのみを使用し、Tophat2 を実行してそれぞれのリファレンスゲノムへのマッピングを行った。マッピングは 1read につき 2 塩基のミスマッチを許容した。得られた alignment データを IGV (Integrative Genomics Viewer) を使って可視化し、sequencing read がどの座位から始まっているかを種間で比較した。

### 3 結果

#### 3-1 NEC8 内在性 *PIPSL* RNA 検出 (NEC8: 辜丸由来奇形腫細胞)

NEC8 において、内在性 *PIPSL* RNA の有無を明らかにするために、RT-PCR を行った。親遺伝子 *PIP5K1A* と *S5a* 由来配列の融合境界点を挟む様に設計したプライマーを用いて、575 nt と想定した理論長通りの増幅産物を得た (図 12)。ネガティブコントロールである *PIPSL* (RT-) にバンドが検出されないことから、この増幅産物が RNA 由来であると断定した。一方で、想定 of 2 倍の長さ (約 1000 bp) になる増幅産物が同時に検出された (図 15, 補足資料図 11)。そのため、この増幅産物をクローニングし、塩基配列を決定した。この配列は、半分が *PIPSL* (10 番染色体) 由来配列でもう半分が 5 番染色体に相同性のある配列が連結していた (補足資料図 12)。

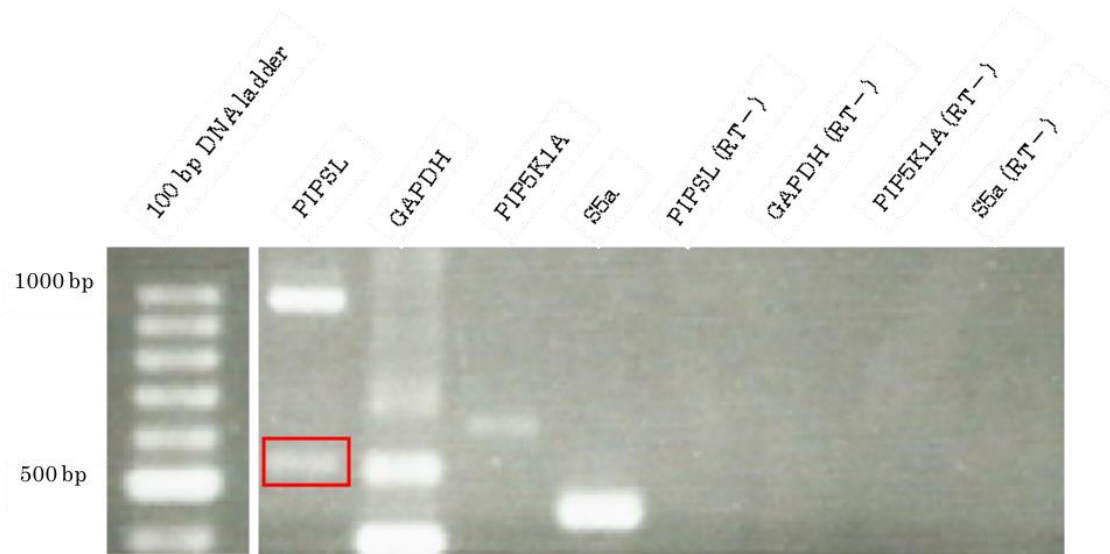


図 12. NEC8 RT-PCR

赤枠で囲われた増幅産物が *PIPSL* に由来

ポジティブコントロール (GAPDH) や親遺伝子の当該領域の鎖長は次の通り

*GAPDH*: 321 nt, *PIP5K1A*: 584 nt, *S5a*: 385 nt

### 3-2 シロテテナガザル精巣における *PIPSL RNA* 検出

類人猿は、ヒト、チンパンジー、ゴリラ、オランウータン、ギボンの5系統が存在する。この5系統における内在性 *PIPSL RNA* は、ヒトとチンパンジーの精巣で検出されている (Babushok, Ohshima et al. 2007)。本研究では、類人猿で最も遠位のギボン (シロテテナガザル) の精巣を用いて、RT-PCR により内在性 *PIPSL RNA* の検出を行った。

コントロールとして用いたチンパンジーとほぼ同じ長さのバンドを検出した (図 13)。チンパンジーは *PIPSL RNA* の発現量が多いために、バンドが太く見える。実験には total RNA を等量用いており、ギボン精巣において、内在性 *PIPSL RNA* はチンパンジーよりもその発現量が少ないことが明らかになった。

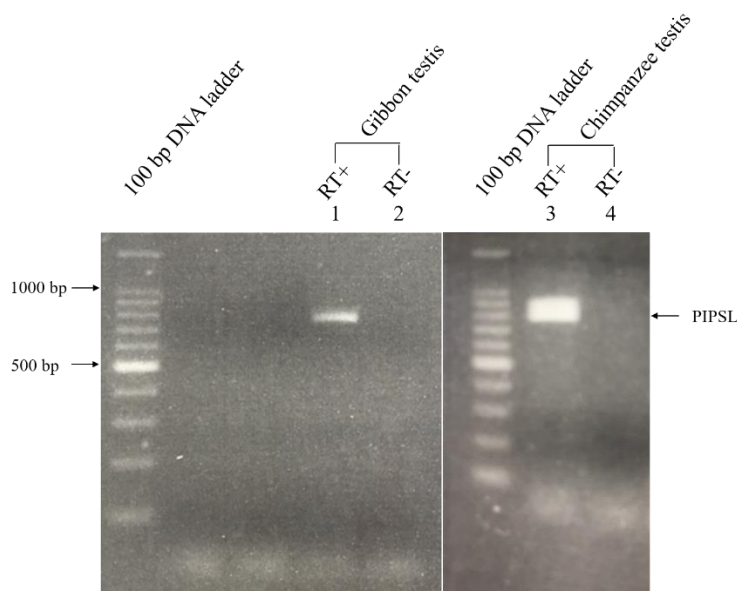


図 13. チンパンジーとギボンの精巣における *PIPSL RNA* 検出



### 3-3 RNA-seq 解析による転写開始点推定

ギボン（シロテテナガザル）の精巣から抽出した total RNA を次世代シーケンサーで配列決定し、リファレンスゲノム（クロテナガザル）にマッピングした。ヒトとオランウータンについて、公共データベースから取得した配列データをリファレンスゲノムにマッピングし、3 系統間で転写開始点候補を比較した。ヒトとギボン間で非常に近い座位に転写開始点候補を見出した(図 14)。また *PIPSL* の配列は親遺伝子である *PIP5K1A* と相同性が高いため、マップされたリードが *PIPSL* 配列であることは確認している(補足資料図 12-14)。

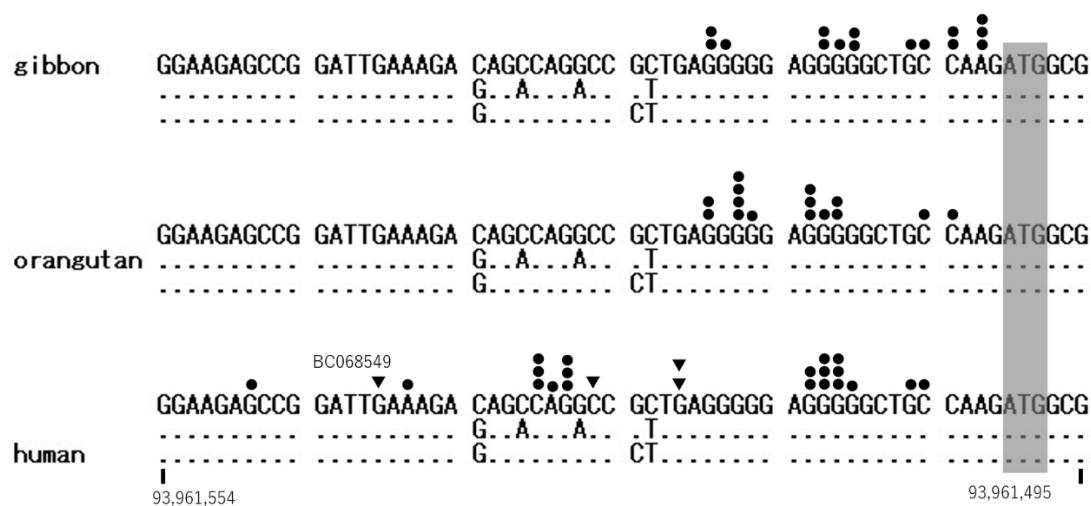


図 14. 転写開始点候補の 3 系統比較

配列の上段がギボン・中段がオランウータン・下段がヒトを示す。●は RNA sequencing read の 5'末端を示す。▼は公共 DB に登録されているヒト *PIPSL* 転写開始点を示す。灰色の網掛けは *PIPSL* の第 1 開始コドンを示す。

### 3-4 PIPSL 上流領域配列を用いたプロモーターアッセイ

#### 3-4-1 PIPSL 上流全域

最初にプロモーターアッセイに用いた配列のゲノムにおける全体図を示す。

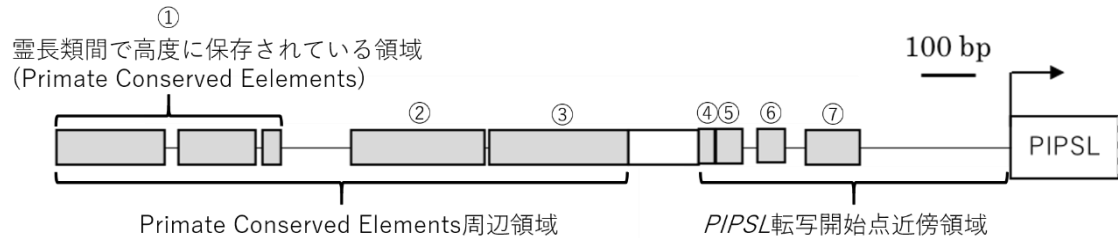


図 15. プロモーターアッセイに用いた *PIPSL* 上流全域

図の左から Primate Conserved Elements (3ヶ所) (①)、LINE2 (②)、LTR (MLT2A1) (③)、(TA)<sub>n</sub>、LTR (MLT2A1) (④)、LINE (LINE2) (⑤)、LTR (MSTB) (⑥)、DNA transposon (MER5A) (⑦)。

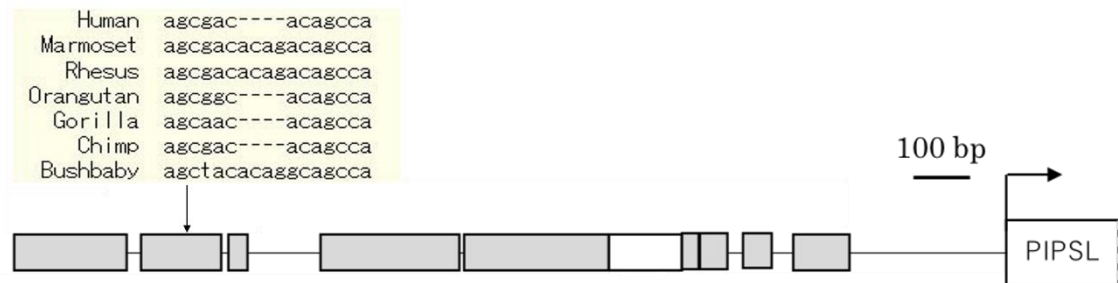


図 16. Primate Conserved Elements における 4 塩基欠損

Primate Conserved Elements において、*PIPSL* を有する系統のみで 4 塩基欠損している。

### 3-4-2 PIPSL 転写開始点近傍領域のプロモーターアッセイ

PIPSL 上流領域のプロモーター活性を調べるために、睾丸由来奇形腫細胞株 (NEC8) を含む 3 種の細胞株を用いて、レポーターアッセイを行った。調査した配列は PIPSL 転写開始点近傍領域に相当する配列 (DNA transposon, LINE, LTR, PIP5K1A 由来配列を含む) である。PIPSL の親遺伝子である PIP5K1A 由来配列は、本来の PIP5K1A 5'UTR 配列の部分配列である。PIP5K1A はあらゆる組織で恒常的に発現しており、そのプロモーター活性は高い。3 種全てにおいて、non-promoter construct と比較して、この転写開始点近傍の領域がプロモーター活性を有していることが明らかとなった (図 17)。

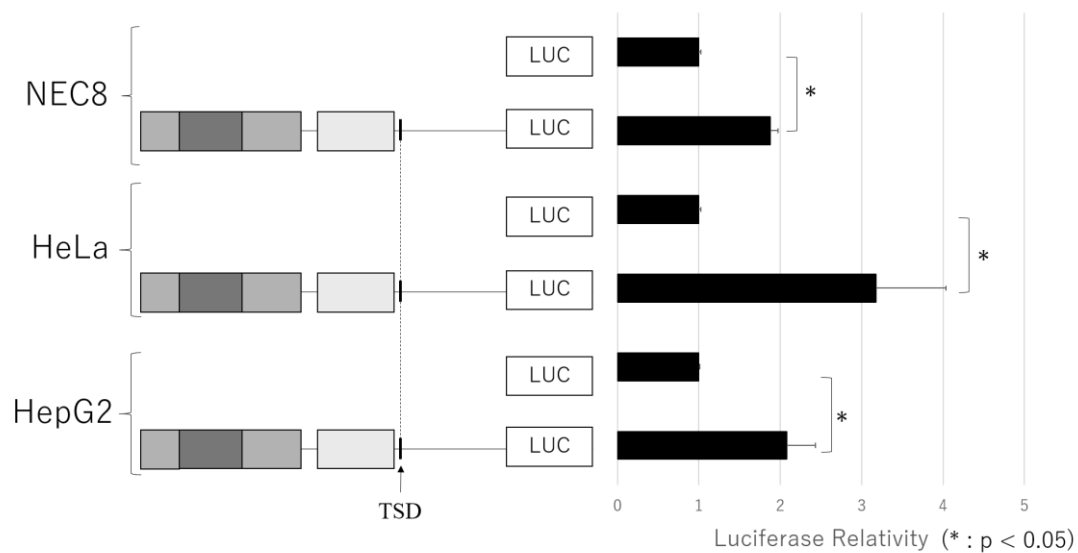


図 17. PIPSL 上流領域配列のプロモーター活性

コンストラクトの構造は左から LTR-LINE2-LTR-DNAtransposon-TSD-PIP5K1A 由来配列-LUC である。TSD は Target Site Duplication の略称であり、プロセス型偽遺伝子生成の過程で、プロセス型偽遺伝子の前後に生じる短い同方向の繰り返し配列である。

また、この領域の中で *PIPSL* から最も遠い LTR の有無で転写活性に違いが見られた。HepG2 のみ LTR を含んだ全領域の場合、転写活性が上昇した。

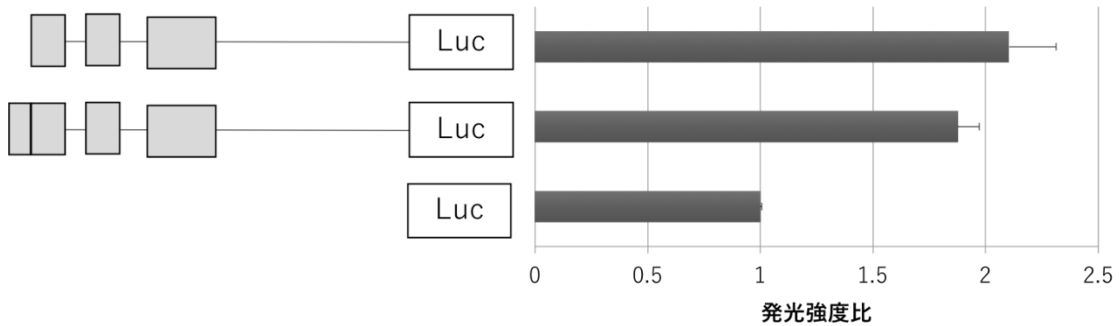


図 18. NEC8 を用いた *PIPSL* 転写開始点近傍における LTR 配列の影響

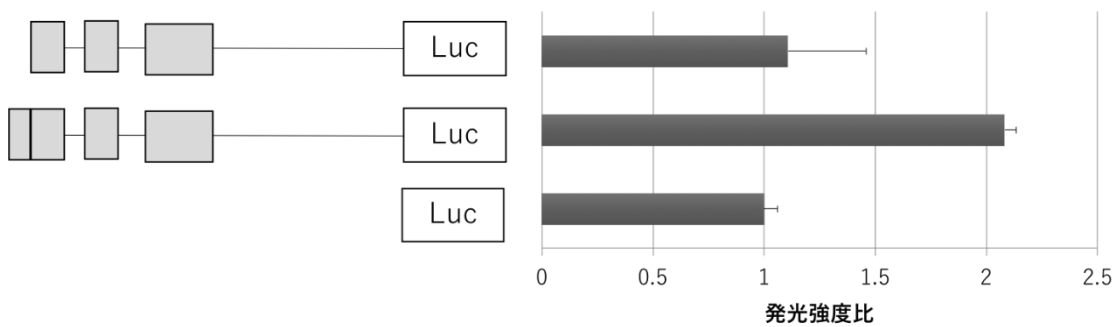


図 19. HepG2 を用いた *PIPSL* 転写開始点近傍における LTR 配列の影響

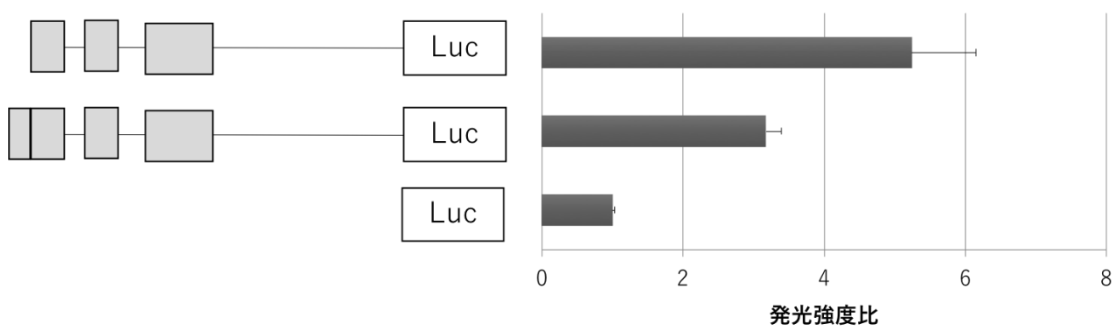


図 20. HeLa を用いた *PIPSL* 転写開始点近傍における LTR 配列の影響

### 3-4-3 霊長類間で高度に保存された配列を含む領域のプロモーターアッセイ

*PIPSL* 上流領域に霊長類間で高度に保存された領域 (Primate Conserved Elements : PCEs) が存在していることを見出し、この領域を含む配列のプロモーター活性を調べた。この PCE 中に、*PIPSL* を有している類人猿 (ヒト、チンパンジー、ゴリラ、オランウータン、ギボン) のみ塩基が一部欠失していることが判明した。相対的に NEC8 の転写活性が他の細胞種を用いた場合の転写活性よりも高かった。しかしながら、どの細胞種を用いても転写活性は低く、配列挿入の無いコンストラクトと同水準である。

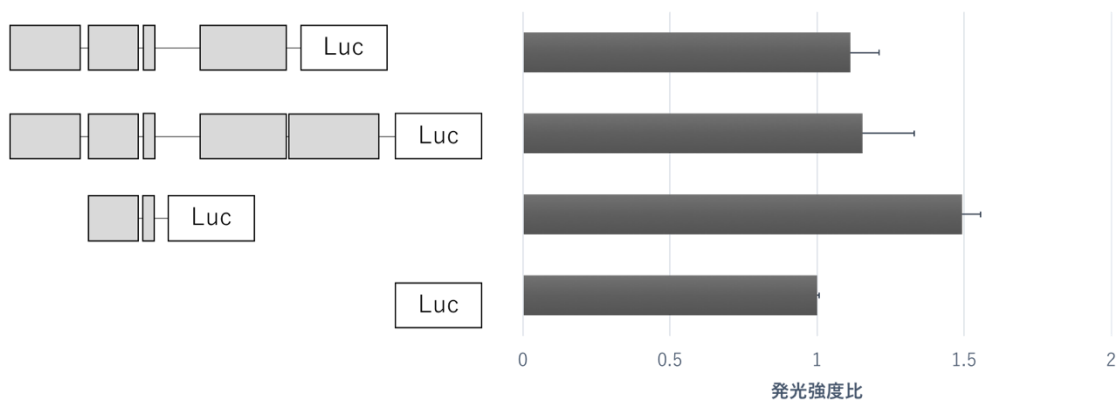


図 21. NEC8 を用いた PCE 周辺配列のプロモーターアッセイ

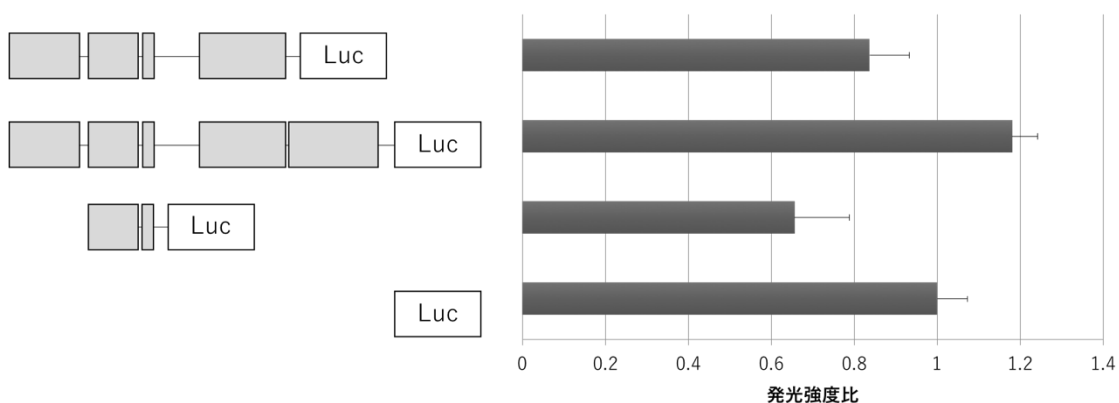


図 22. HepG2 を用いた PCE 周辺配列のプロモーターアッセイ

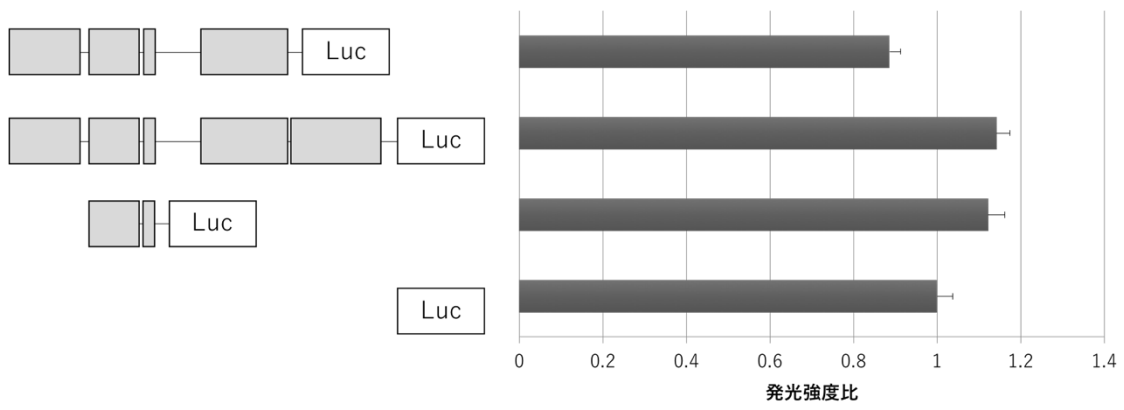


図 23. HeLa を用いた PCE 周辺配列のプロモーターアッセイ

## 4 考察

### 4-1 レトロ遺伝子 *PIPSL* が初期に獲得したプロモーター活性

近年、レトロ遺伝子に関して多様な生物種を対象に、数多くのレトロ遺伝子に共通にみられる現象やメカニズムを報告する研究がある (Carelli FN, *et al.*, 2016)。次世代シーケンサーの登場により、数多くの生物種やサンプルを同時に扱った大規模解析が可能になったためである。一方で、1つのレトロ遺伝子に関する研究報告は少ない (Parker H, *et al.*, 2009)。

*PIPSL* RNA 発現を調査したギボンは、類人猿の中で早くに分岐したヒトから最も遠い系統である。RNA-seq 解析によりギボン *PIPSL* RNA 配列から明らかとなったことは、*PIPSL* TSS がギボンとヒト間で近接していることを強く示唆するものである。本研究で得られた結果に基づき、初期の *PIPSL* における転写活性化に関与する分子メカニズムを説明するプロモーター早期獲得仮説を提案する (図 24C)。

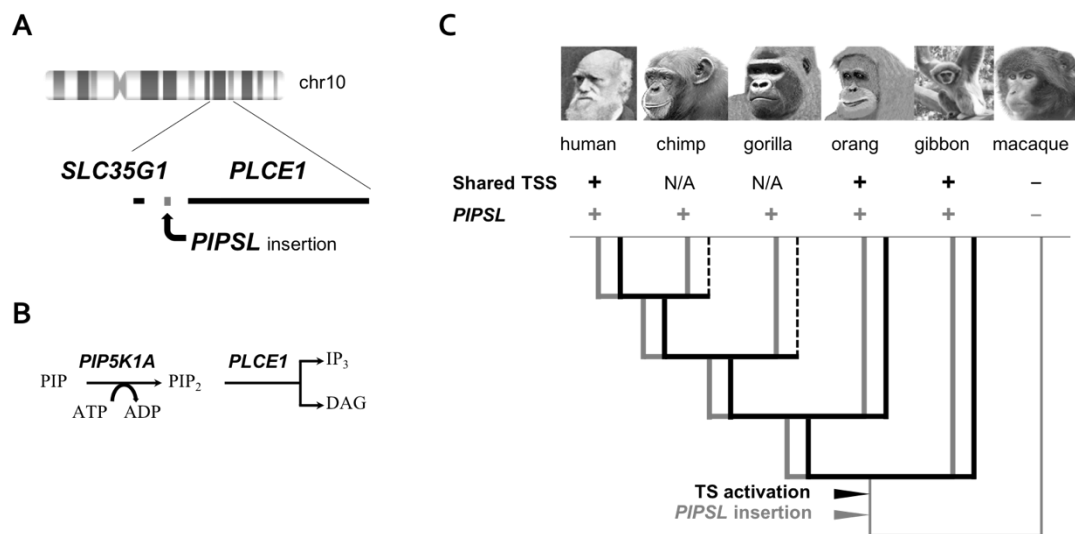


図 24. *PIPSL* のゲノム上の位置およびプロモーター活性早期獲得

*PIPSL* 祖先配列は旧世界サル (macaque) からの分岐後、ヒト上科霊長類の共通祖先のゲノムに出現し、その後ヒト上科霊長類 5 系統全てに保持されている (図 24C)。*PIPSL* は既存の遺伝子から離れた遺伝子間領域に挿入されているため、*PIPSL* は自身の転写制御機構を獲得した可能性が高い。

本研究で、*PIPSL* の親遺伝子である *PIP5K1A* 5'UTR 由来の一部、およびいくつかの転移因子を含む領域について、ヒト培養細胞を用いたヒト *PIPSL* 上流配列のプロモーター活性を同定した。*PIPSL* 5'UTR 配列は *PIP5K1A* の 5'UTR 配列の一部に由来し、*PIPSL* の生成と同時に生じたものである。そのため、*PIP5K1A* 5'UTR の 5'末端と *PIP5K1A/PIPSL* の開始コドンの間に、*PIPSL* の TSS が局在している。このことから、*PIP5K1A* の 5'UTR が *PIPSL* 特異的プロモーター活性を示すという可能性がある。レポーターアッセイの結果から、*PIPSL* 5'UTR 配列より上流に位置する霊長類間で保存性の高い領域にプロモーター活性がないことが明らかとなっている。つまり、この領域より上流にはプロモーター活性を有する領域は存在しないと考えられる。

ヒト培養細胞を用いた実験から、異なる転写活性を示す配列を見出した。この配列は LTR-レトロトランスポゾン (MLT2A1) LTR 配列の一部であり、HepG2 のみで転写活性を高める効果を示した。したがって、この領域における反復配列は、細胞種で *PIPSL* 転写活性に異なる影響を及ぼし得る。類人猿進化の過程で、ヒト *PIPSL* TSS 上流配列の一部が塩基置換 (TGT から TAT) したことにより、TATA-box 様配列 (TATAAA) が出現したことを見出した。この配列を有するのはヒト、チンパンジーとゴリラのみである。この TATA-box 様配列は、ヒト *PIPSL* TSSs の幅広い分布の原因となった可能性がある。

#### 4-2 霊長類 *PIPSL* の機能に関する考察

本研究では、得られた結果をもとに、*PIPSL* 転写の早期獲得仮説を提案した。この仮説が正しければ、*PIPSL* 遺伝子は約 2000 万年間のヒト上科霊長類の進化を通じて、その構造のみならず転写調節に関して保存されていることになる。それは、*PIPSL* の保存を促す機能的制約が存在することを強く示唆する。*PIPSL* の RNA 発現は証明されているが、全長の *PIPSL* タンパク質はこれまでのところ検出されていない (Babushok *et al.*, 2007)。PIPSL 部分アミノ酸配列がプロテオミクスデータベースに見られるのみである (Ohshima and Igarashi, 2010)。このため、*PIPSL* 産物に関する生物学的機能はいまだ明らかではない。しかし *PIPSL* の周辺領域の情報が、*PIPSL* 機能に有用な手がかり



になる可能性がある (図 24A)。 *PIPSL* は、ヒト第 10 番染色体において、ホスホリパーゼ C  $\epsilon$  1 (PLCE1) に近接して位置する。 PLCE1 は、ホスファチジルイノシトール-4,5-ビスホスフェート (PtdIns (4, 5) P2) の加水分解を触媒し、2つのセカンドメッセンジャー(イノシトール 1,4,5-トリホスフェートおよびジアリルグリセロール)を生成する (図 24B) (Hachem *et al.*, 2017)。

PIPSL の親遺伝子である PIP5K1A は、ホスファチジルイノシトール 4-リン酸のリン酸化を触媒して PtdIns (4, 5) P2 を形成するキナーゼをコードしている (図 24B)。 PIPSL に機能的制約がなければ、このように強く関連した遺伝子の隣に、無作為に挿入された遺伝子コピーが長期間にわたって保存されるのはありそうにない。少なくとも、進化の初期段階においては、PIPSL は PIP5K1A 様タンパク質として作用していた可能性が高い。しかしながら、*in vitro* では、ヒト PIPSL タンパク質の脂質キナーゼ活性が非常に弱く、細胞性ユビキチン化タンパク質に対して顕著な親和性を示している (Babushok *et al.*, 2007)。 PIPSL の機能的役割は進化過程で変化した可能性がある。

## 5 参考文献

Abdelsamad, A. and A. Pecinka (2014). "Pollen-specific activation of Arabidopsis retrogenes is associated with global transcriptional reprogramming." *Plant Cell* **26**(8): 3299-3313.

Abyzov, A., R. Iskow, O. Gokcumen, D. W. Radke, S. Balasubramanian, B. Pei, L. Habegger, C. Lee and M. Gerstein (2013). "Analysis of variable retroduplications in human populations suggests coupling of retrotransposition to cell division." *Genome Res* **23**(12): 2042-2052.

Akiva, P., A. Toporik, S. Edelheit, Y. Peretz, A. Diber, R. Shemesh, A. Novik and R. Sorek (2006). "Transcription-mediated gene fusion in the human genome." *Genome Res* **16**(1): 30-36.

Babushok, D. V., K. Ohshima, E. M. Ostertag, X. Chen, Y. Wang, P. K. Mandal, N. Okada, C. S. Abrams and H. H. Kazazian, Jr. (2007). "A novel testis ubiquitin-binding protein gene arose by exon shuffling in hominoids." *Genome Res* **17**(8): 1129-1138.

Betran, E. and M. Long (2003). "Dntf-2r, a young Drosophila retroposed gene with specific male expression under positive Darwinian selection." *Genetics* **164**(3): 977-988.

Betran, E., W. Wang, L. Jin and M. Long (2002). "Evolution of the phosphoglycerate mutase processed gene in human and chimpanzee revealing the origin of a new primate gene." *Mol Biol Evol* **19**(5): 654-663.

Brosius, J. (1999). "RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements." *Gene* **238**(1): 115-134.

Carelli, F. N., T. Hayakawa, Y. Go, H. Imai, M. Warnefors and H. Kaessmann (2016). "The life history of retrocopies illuminates the evolution of new mammalian genes." *Genome Res* **26**(3): 301-314.

Casola, C. and E. Betran (2017). "The Genomic Impact of Gene Retrocopies: What Have We Learned from Comparative Genomics, Population Genomics, and Transcriptomic Analyses?" *Genome Biol Evol* **9**(6): 1351-1373.

Cheetham, S. W., F. Gruhl, J. S. Mattick and M. E. Dinger (2013). "Long noncoding RNAs and the genetics of cancer." *Br J Cancer* **108**(12): 2419-2425.

Ciomborowska, J., W. Rosikiewicz, D. Szklarczyk, W. Makalowski and I. Makalowska (2013). "'Orphan' retrogenes in the human genome." *Mol Biol Evol* **30**(2): 384-396.

Doucet-O'Hare, T. T., N. Rodic, R. Sharma, I. Darbari, G. Abril, J. A. Choi, J. Young Ahn, Y. Cheng, R. A. Anders, K. H. Burns, S. J. Meltzer and H. H. Kazazian, Jr. (2015). "LINE-1 expression and retrotransposition in Barrett's esophagus and esophageal carcinoma." *Proc Natl Acad Sci U S A* **112**(35): E4894-4900.

Douglas, G. M., M. D. Wilson and A. M. Moses (2016). "Decreased Transcription Factor Binding Levels Nearby Primate Pseudogenes Suggest Regulatory Degeneration." *Mol Biol Evol* **33**(6): 1478-1485.

Ewing, A. D., T. J. Ballinger, D. Earl, C. C. Harris, L. Ding, R. K. Wilson and D. Haussler (2013). "Retrotransposition of gene transcripts leads to structural variation in mammalian genomes." *Genome Biol* **14**(3): R22.

Ha, H., J. Song, S. Wang, A. Kapusta, C. Feschotte, K. C. Chen and J. Xing (2014). "A comprehensive analysis of piRNAs from adult human testis and their relationship with genes and mobile elements." *BMC Genomics* **15**: 545.

Harrison, P. M., D. Zheng, Z. Zhang, N. Carriero and M. Gerstein (2005). "Transcribed processed pseudogenes in the human genome: an intermediate form of expressed retrosequence lacking protein-coding ability." *Nucleic Acids Res* **33**(8): 2374-2383.

Hirano, T., Y. W. Iwasaki, Z. Y. Lin, M. Imamura, N. M. Seki, E. Sasaki, K. Saito, H. Okano, M. C. Siomi and H. Siomi (2014). "Small RNA profiling and characterization of piRNA clusters in the adult testes of the common marmoset, a model primate." *Rna* **20**(8): 1223-1237.

Hasegawa H, Noguchi J, Yamashita M, Okada R, Sugimoto R, Furuya M, Unoki T, funakoshi Y, Baba T and Kanaho Y (2012). "Phosphatidylinositol 4-phosphate 5-kinase is indispensable for mouse spermatogenesis." *Biol Reprod* **86**(5) : 136, 131-112

Kabza, M., J. Ciomborowska and I. Makalowska (2014). "RetrogeneDB--a database of animal retrogenes." *Mol Biol Evol* **31**(7): 1646-1648.

Kubiak, M. R. and I. Makalowska (2017). "Protein-Coding Genes' Retrocopies and Their Functions." *Viruses* **9**(4).

Matsui, A., Y. Go and Y. Niimura (2010). "Degeneration of olfactory receptor gene repertoires in primates: no direct link to full trichromatic vision." *Mol Biol Evol* **27**(5): 1192-1200.

Mighell, A. J., N. R. Smith, P. A. Robinson and A. F. Markham (2000). "Vertebrate pseudogenes." *FEBS Lett* **468**(2-3): 109-114.

Navarro, F. C. and P. A. Galante (2015). "A Genome-Wide Landscape of Retrocopies in Primate Genomes." *Genome Biol Evol* **7**(8): 2265-2275.

Ohshima, K., M. Hattori, T. Yada, T. Gojobori, Y. Sakaki and N. Okada (2003). "Whole-genome screening indicates a possible burst of formation of processed pseudogenes and Alu repeats by particular L1 subfamilies in ancestral primates." *Genome Biol* **4**(11): R74.

Ohshima, K. and K. Igarashi (2010). "Inference for the initial stage of domain shuffling: tracing the evolutionary fate of the PIPSL retrogene in hominoids." *Mol Biol Evol* **27**(11): 2522-2533.

Parker, H. G., B. M. VonHoldt, P. Quignon, E. H. Margulies, S. Shao, D. S. Mosher, T. C. Spady, A. Elkahlon, M. Cargill, P. G. Jones, C. L. Maslen, G. M. Acland, N. B. Sutter, K. Kuroki, C. D. Bustamante, R. K. Wayne and E. A. Ostrander (2009). "An expressed *fgf4* retrogene is associated with breed-defining chondrodysplasia in domestic dogs." *Science* **325**(5943): 995-998.

Poliseno, L., L. Salmena, J. Zhang, B. Carver, W. J. Haveman and P. P. Pandolfi (2010). "A coding-independent function of gene and pseudogene mRNAs regulates tumour biology." *Nature* **465**(7301): 1033-1038.

Rosenbloom, K. R., J. Armstrong, G. P. Barber, J. Casper, H. Clawson, M. Diekhans, T. R. Dreszer, P. A. Fujita, L. Guruvadoo, M. Haeussler, R. A. Harte, S. Heitner, G. Hickey, A. S. Hinrichs, R. Hubley, D. Karolchik, K. Learned, B. T. Lee, C. H. Li, K. H. Miga, N. Nguyen, B. Paten, B. J. Raney, A. F. Smit, M. L. Speir, A. S. Zweig, D. Haussler, R. M. Kuhn and W. J. Kent (2015). "The UCSC Genome Browser database: 2015 update." *Nucleic Acids Res* **43**(Database issue): D670-681.

Rosso, L., A. C. Marques, M. Weier, N. Lambert, M. A. Lambot, P. Vanderhaeghen and H. Kaessmann (2008). "Birth and rapid subcellular adaptation of a hominoid-specific *CDC14* protein." *PLoS Biol* **6**(6): e140.

Sakai, H., K. O. Koyanagi, T. Imanishi, T. Itoh and T. Gojobori (2007). "Frequent emergence and functional resurrection of processed pseudogenes in the human and mouse genomes." *Gene* **389**(2): 196-203.

Schrider, D. R., F. C. Navarro, P. A. Galante, R. B. Parmigiani, A. A. Camargo, M. W. Hahn and S. J. de Souza (2013). "Gene copy-number polymorphism caused by retrotransposition in humans." *PLoS Genet* **9**(1): e1003242.

- Shashidharan, P., T. M. Michaelidis, N. K. Robakis, A. Kresoali, J. Papamatheakis and A. Plaitakis (1994). "Novel human glutamate dehydrogenase expressed in neural and testicular tissues and encoded by an X-linked intronless gene." *J Biol Chem* **269**(24): 16971-16976.
- Sorourian, M., M. M. Kunte, S. Domingues, M. Gallach, F. Ozdil, J. Rio and E. Betran (2014). "Relocation facilitates the acquisition of short cis-regulatory regions that drive the expression of retrogenes during spermatogenesis in *Drosophila*." *Mol Biol Evol* **31**(8): 2170-2180.
- Torrents, D., M. Suyama, E. Zdobnov and P. Bork (2003). "A genome-wide survey of human pseudogenes." *Genome Res* **13**(12): 2559-2567.
- Vanin, E. F. (1985). "Processed pseudogenes: characteristics and evolution." *Annu Rev Genet* **19**: 253-272.
- Wang, P. J. (2004). "X chromosomes, retrogenes and their role in male reproduction." *Trends Endocrinol Metab* **15**(2): 79-83.
- Wang, Y. (2017). "PlantRGDB: A Database of Plant Retrocopied Genes." *Plant Cell Physiol* **58**(1): e2.
- Weiner, A. M., P. L. Deininger and A. Efstratiadis (1986). "Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information." *Annu Rev Biochem* **55**: 631-661.
- Young, J., J. Menetrey and B. Goud (2010). "RAB6C is a retrogene that encodes a centrosomal protein involved in cell cycle progression." *J Mol Biol* **397**(1): 69-88.
- Zhang, Q. (2013). "The role of mRNA-based duplication in the evolution of the primate genome." *FEBS Lett* **587**(21): 3500-3507.

Zhang, Y., S. Lu, S. Zhao, X. Zheng, M. Long and L. Wei (2009). "Positive selection for the male functionality of a co-retroposed gene in the hominoids." *BMC Evol Biol*9: 252.

Zhang, Z., P. M. Harrison, Y. Liu and M. Gerstein (2003). "Millions of years of evolution preserved: a comprehensive catalog of the processed pseudogenes in the human genome." *Genome Res* 13(12): 2541-2558.

Zhang, Z. D., A. Frankish, T. Hunt, J. Harrow and M. Gerstein (2010). "Identification and analysis of unitary pseudogenes: historic and contemporary gene losses in humans and other primates." *Genome Biol*11(3): R26.

## 6 補足資料

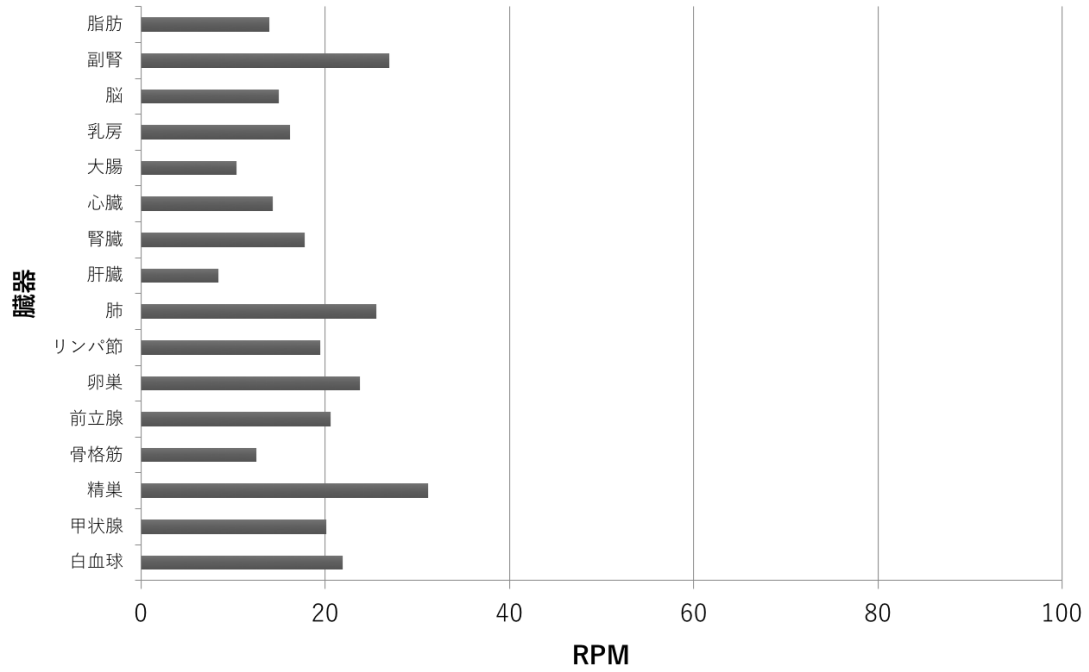


図 1. ヒト各組織における *PIP5K1A* RNA 発現量

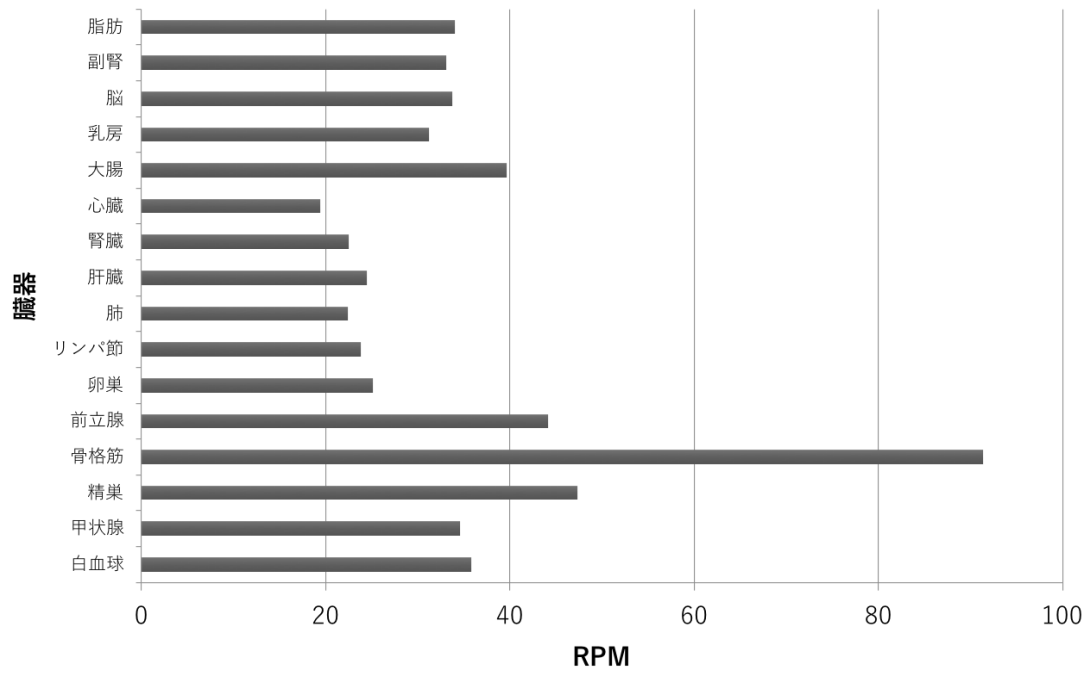




図 2. ヒト各組織における *S5a* RNA 発現量

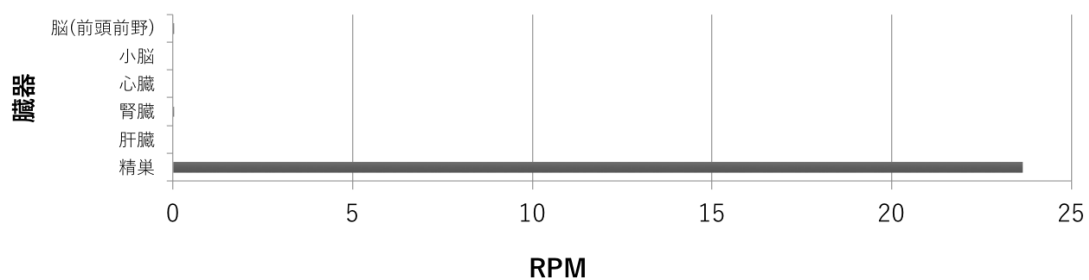


図 3. チンパンジー各組織における *PIPSL* RNA 発現量

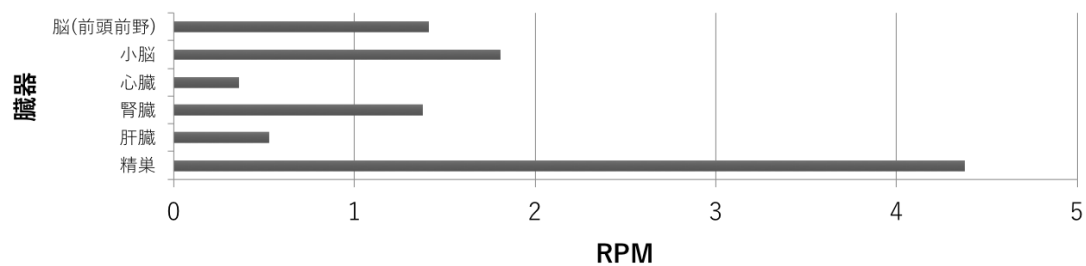


図 4. ゴリラ各組織における *PIPSL* RNA 発現量

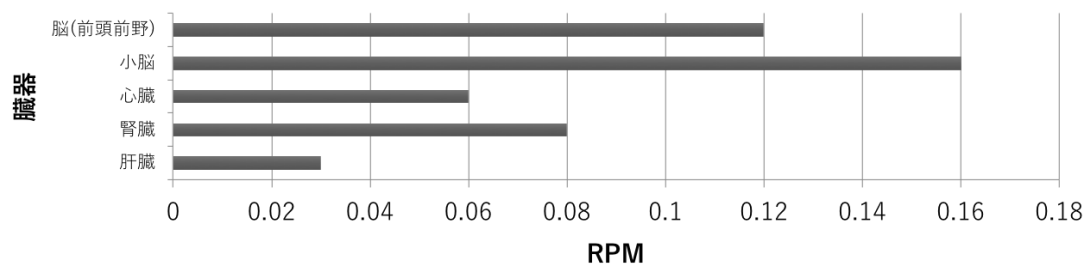


図 5. オランウータン各組織における *PIPSL* RNA 発現量

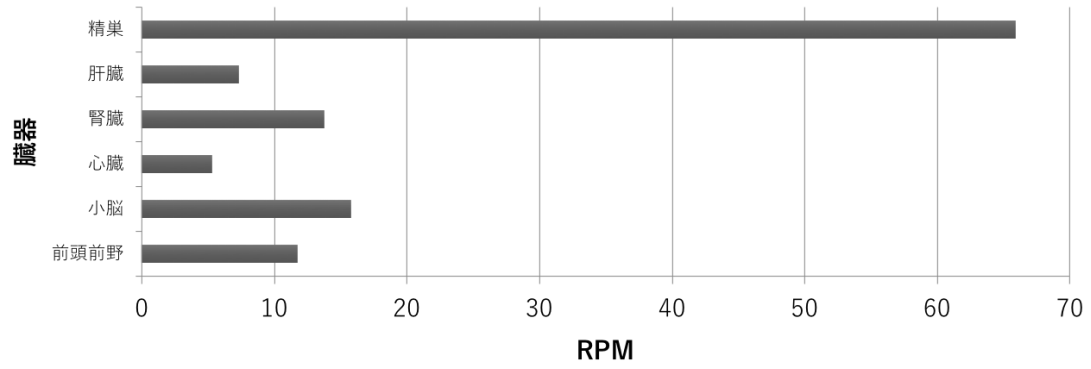


図 6. チンパンジー各組織における *PIP5k1A* RNA 発現量

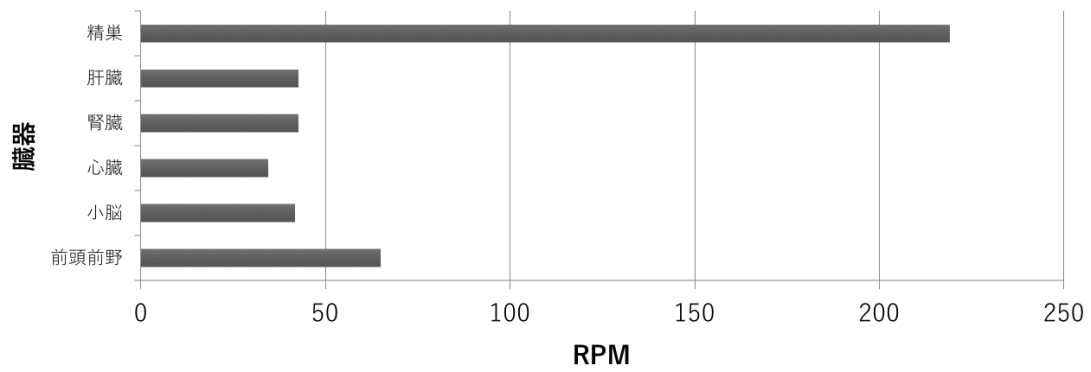


図 7. チンパンジー各組織における *S5a* RNA 発現量

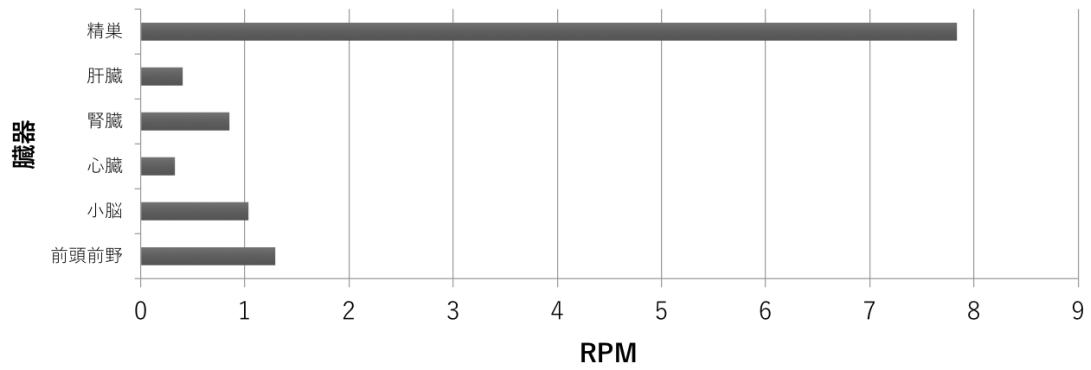


図 8. ゴリラ各組織における *PIP5K1A* RNA 発現量

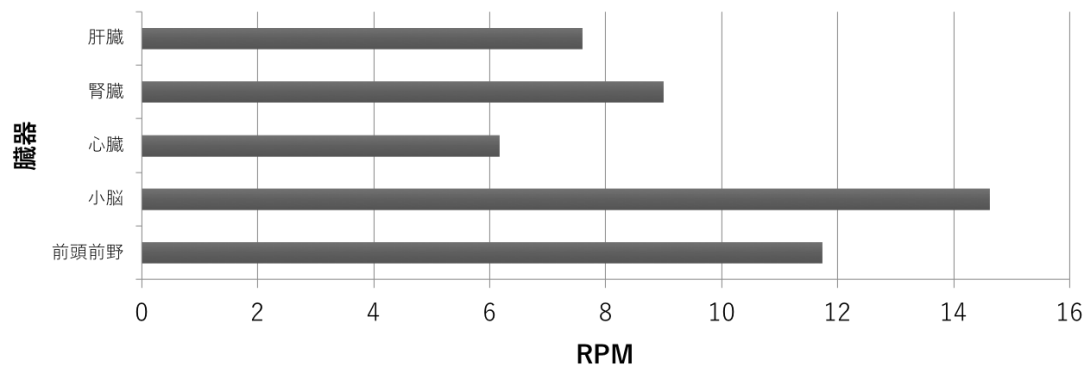


図 9. オランウータン各組織における *PIP5K1A* RNA 発現量

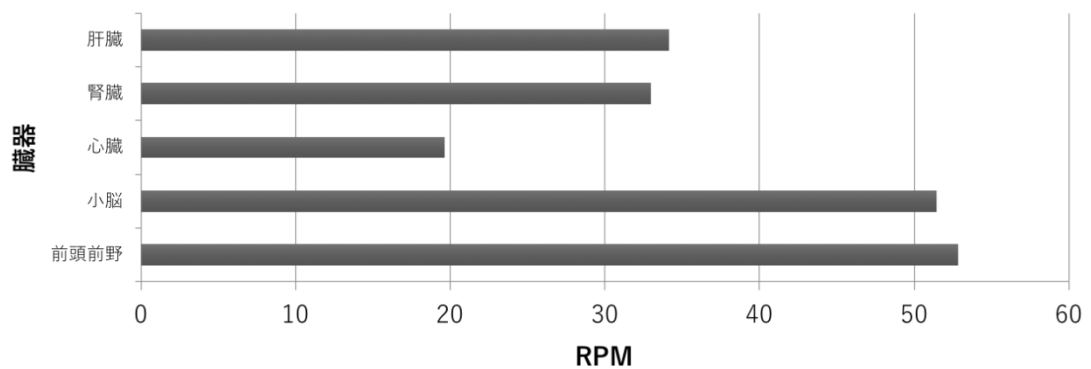


図 10. オランウータン各組織における *S5a* RNA 発現量

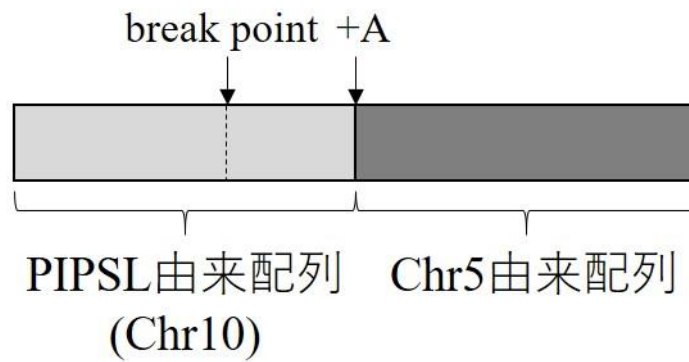


図 11. NEC8 細胞を用いた RT-PCR で検出した約 1000 bp の増幅産物  
*PIPSL* 由来配列と 5 番染色体に由来する配列が融合した配列構造。それぞれの由来配列は 1 塩基の A を挟んで融合していた。break point は *PIP5K1A* 由来配列と *S5a* 由来配列の融合境界点を指す。



```

PIP5K1A_human          GGAAGAACCGGATTGAAAGAGAGCCAGGCCGCTGAGGGGGAGGGGGCTGC 50
PIP5K1A_gibbon        GGAAGAGCCGGATTGAAA-----GCCAGGCCGCTGAGGGGGAGGGGGCTGC 46
PIPSL_human           GGAAGAGCCGGATTGAAAGAGAGCCAGGCCCTTGAAGGGGAGGGGGCTGC 50
PIPSL_orangutan       GGAAGAGCCGGATTGAAAGAGAGACAGCCGCTGAGGGGGAGGGGGCTGC 50
PIPSL_gibbon          GGAAGAGCCGGATTGAAAGACAGCCAGGCCGCTGAGGGGGAGGGGGCTGC 50
SRR2040581.1807470_93961450  -----GCCGGATTGAAAGAGAGCCAGGCCCTTGAAGGGGAGGGGGCTGC 44
SRR2040581.15837181_93961467  -----AAGAGAGCCAGGCCCTTGAAGGGGAGGGGGCTGC 34
SRR2040581.3829339_93961432  -----CAGGCCCTTGAAGGGGAGGGGGCTGC 26
SRR2040581.4038465_93961432  -----CAGGCCCTTGAAGGGGAGGGGGCTGC 26
SRR2040581.9743660_93961431  -----CAGGCCCTTGAAGGGGAGGGGGCTGC 26
SRR2040581.14726955_93961431  -----AGGCCCTTGAAGGGGAGGGGGCTGC 25
SRR2040581.10599783_93961430  -----GGCCCTTGAAGGGGAGGGGGCTGC 24
SRR2040581.15553157_93961430  -----GGCCCTTGAAGGGGAGGGGGCTGC 24
SRR2040581.13166075_93961444  -----GGCCCTTGAAGGGGAGGGGGCTGC 24
SRR2040581.4199401_93961414  -----GGGGGCTGC 9
SRR2040581.7903437_93961415  -----GGGGGCTGC 9
SRR2040581.107858_93961415  -----GGGGCTGC 8
SRR2040581.2296059_93961414  -----GGGGCTGC 8
SRR2040581.4798408_93961413  -----GGGGCTGC 8
SRR2040581.6456886_93961412  -----GGGCTGC 7
SRR2040581.7320176_93961412  -----GGGCTGC 7
SRR2040581.7574988_93961412  -----GGGCTGC 7
SRR2040581.13870115_93961411  -----GGCTGC 6
SRR2040581.13350890_93961408  -----GC 2
SRR2040581.3060534_93961406  -----C 1
*

PIP5K1A_human          TAAGATGGCGTCGGCTCCTCCGGGCGCTGCTTTCGGTCGGTTTTTCAT 100
PIP5K1A_gibbon        CAAGATGGCGTCGGCTCCTCCGGGCGCTGCTTTCGGTCGGTTTTTCAT 96
PIPSL_human           CAAGATGGCGTCGGCTCCTCCGGGCGCTGCTTTCGGTCGGTTTTTC-T 99
PIPSL_orangutan       CAAGATGGCGTCGGCTCCTCCGGGCGCTGCTTTCGGTCGGTTTTTCAT 100
PIPSL_gibbon          CAAGATGGCGTCGGCTCCTCCGGGCGCTGCTTTCGGTCGGTTTTTCAT 100
SRR2040581.1807470_93961450  CAAGATGGCGTCGGCTCCTCCGGGCGCTGCTTTCGGTCGGTTTTTC-T 93
SRR2040581.15837181_93961467  CAAGATGGCGTCGGCTCCTCCGGGCGCTGCTTTCGG----- 72
SRR2040581.3829339_93961432  CAAGATGGCGTCGGCTCCTCCGGGCGCTGCTTTCGGTCGGTTTTTC-T 75
SRR2040581.4038465_93961432  CAAGATGGCGTCGGCTCCTCCGGGCGCTGCTTTCGGTCGGTTTTTC-T 75
SRR2040581.9743660_93961431  CAAGATGGCGTCGGCTCCTCCGGGCGCTGCTTTCGGTCGGTTTTTC-T 75
SRR2040581.14726955_93961431  CAAGATGGCGTCAGCCTCCTCCGGGCGCTGCTTTCGGTCGGTTTTTC-T 74
SRR2040581.10599783_93961430  CAAGATGGCGTCGGCTCCTCCGGGCGCTGCTTTCGGTCGGTTTTTC-T 73
SRR2040581.15553157_93961430  CAAGATGGCGTCGGCTCCTCCGGGCGCTGCTTTCGGTCGGTTTTTC-T 73
SRR2040581.13166075_93961444  CAAGATGGCGTCGGCTCCTCCGGGCGCTGCTTTCGGTCGGTTTTTC-T 73
SRR2040581.4199401_93961414  CAAGATGGCGTCGGCTCCTCCGGGCGCTGCTTTCGGTCGGTTTTTC-T 58
SRR2040581.7903437_93961415  CAAGATGGCGTCGGCTCCTCCGGGCGCTGCTTTCGGTCGGTTTTTC-T 58
SRR2040581.107858_93961415  CAAGATGGCGTCGGCTCCTCCGGGCGCTGCTTTCGGTCGGTTTTTC-T 57
SRR2040581.2296059_93961414  CAAGATGGCGTCGGCTCCTCCGGGCGCTGCTTTCGGTCGGTTTTTC-T 57
SRR2040581.4798408_93961413  CAAGATGGCGTCGGCTCCTCCGGGCGCTGCTTTCGGTCGGTTTTTC-T 57
SRR2040581.6456886_93961412  CAAGATGGCGTCGGCTCCTCCGGGCGCTGCTTTCGGTCGGTTTTTC-T 56
SRR2040581.7320176_93961412  CAAGATGGCGTCGGCTCCTCCGGGCGCTGCTTTCGGTCGGTTTTTC-T 56
SRR2040581.7574988_93961412  CAAGATGGCGTCGGCTCCTCCGGGCGCTGCTTTCGGTCGGTTTTTC-T 56
SRR2040581.13870115_93961411  CAAGATGGCGTCGGCTCCTCCGGGCGCTGCTTTCGGTCGGTTTTTC-T 55
SRR2040581.13350890_93961408  CAAGATGGCGTCGGCTCCTCCGGGCGCTGCTTTCGGTCGGTTTTTC-T 51
SRR2040581.3060534_93961406  CAAGATGGCGTCGGCTCCTCCGGGCGCTGCTTTCGGTCGGTTTTTC-T 50
*****

PIP5K1A_human          CCTTTGATCCCGCGGTCCCTTCTGTACCTCGTCTCAGCATCTGGAATC 150
PIP5K1A_gibbon        CCTTTGATCCCGCGGTCCCTTCTGTACCTCGTCTCAGCATCTGGAATC 146
PIPSL_human           CCGTTGATCCCGCGGTCCCTTCTGTACCTCGTCTCAGCATCTGGAATC 149
PIPSL_orangutan       CCTTTGATCCCGCGGTCCCTTCTGTACCTCGTCTCAGCATCTGGAATC 150
PIPSL_gibbon          CCTTTGATCCCGCGGTCCCTTCTGTACCTCGTCTCAGCATCTGGAATC 150
SRR2040581.1807470_93961450  CCGTTG----- 99
SRR2040581.15837181_93961467  ----- 99
SRR2040581.3829339_93961432  CCGTTGATCCCGCGGTCCCTTCTCT----- 99
SRR2040581.4038465_93961432  CCGTTGATCCCGCGGTCCCTTCTCT----- 99
SRR2040581.9743660_93961431  CCGTTGATCCCGCGGTCCCTTCTCTG----- 100
SRR2040581.14726955_93961431  CCGTTGATCCCGCGGTCCCTTCTCTG----- 99
SRR2040581.10599783_93961430  CCGTTGATCCCGCGGTCCCTTCTCTG----- 99
SRR2040581.15553157_93961430  CCGTTGATCCCGCGGTCCCTTCTCTG----- 99
SRR2040581.13166075_93961444  CCGTTGATCCCG----- 85
SRR2040581.4199401_93961414  CCGTTGATCCCGCGGTCCCTTCTGTACCTCGTCTCAGCAT----- 100
SRR2040581.7903437_93961415  CCGTTGATCCCGCGGTCCCTTCTGTACCTCGTCTCAGCA----- 99
SRR2040581.107858_93961415  CCGTTGATCCCGCGGTCCCTTCTGTACCTCGTCTCAGCA----- 98
SRR2040581.2296059_93961414  CCGTTGATCCCGCGGTCCCTTCTGTACCTCGTCTCAGCAT----- 99
SRR2040581.4798408_93961413  CCGTTGATCCCGCGGTCCCTTCTGTACCTCGTCTCAGCATC----- 100
SRR2040581.6456886_93961412  CCGTTGATCCCGCGGTCCCTTCTGTACCTCGTCTCAGCATCA----- 100
SRR2040581.7320176_93961412  CCGTTGATCCCGCGGTCCCTTCTGTACCTCGTCTCAGCATCT----- 100
SRR2040581.7574988_93961412  CCGTTGATCCCGCGGTCCCTTCTGTACCTCGTCTCAGCATCT----- 100
SRR2040581.13870115_93961411  CCGTTGATCCCGCGGTCCCTTCTGTACCTCGTCTCAGCATCTG----- 100
SRR2040581.13350890_93961408  CCGTTGATCCCGCGGTCCCTTCTGTACCTCGTCTCAGCATCTGGA----- 99
SRR2040581.3060534_93961406  CCGTTGATCCCGCGGTCCCTTCTGTACCTCGTCTCAGCATCTGGAATC 100

```

図 13. ヒト RNA-seq *PIPSL* 5'末端配列

青枠の配列はゲノムにマップされたシーケンズ配列

赤枠は *PIP5K1A*(親遺伝子)はマップされた read が親遺伝子ではなく、*PIPSL* 配列である

ことを示す箇所

```

PIP5K1A_human          GGAAGAACC GGATTGAAAGAGAGCCAGGCCGCTGAGGGGGAGGGGGCTGC 50
PIP5K1A_gibbon         GGAAGAGCCGGATTGAAA-----GCCAGGCCGCTGAGGGGGAGGGGGCTGC 46
PIPSL_human           GGAAGAGCCGGATTGAAAGAGAGCCAGGCCCTTGAGGGGGAGGGGGCTGC 50
PIPSL_orangutan       GGAAGAGCCGGATTGAAAGAGAGACAGCCGTTGAGGGGGAGGGGGCTGC 50
PIPSL_gibbon          GGAAGAGCCGGATTGAAAGACAGCCAGGCCGCTGAGGGGGAGGGGGCTGC 50
SRR2176206.21252451_92737557 -----GGGGGAGGGGGCTGC 15
SRR2176206.15382647_92737557 -----GGGGGAGGGGGCTGC 15
SRR2176206.8312962_92737581 -----GGGAGGGGGCTGC 13
SRR2176206.23186180_92737557 -----GGGAGGGGGCTGC 13
SRR2176206.20047362_92737555 -----GGGAGGGGGCTGC 13
SRR2176206.22547589_92737555 -----GGGAGGGGGCTGC 13
SRR2176206.13341916_92737554 -----GGGAGGGGGCTGC 12
SRR2176206.8194596_92737552 -----GGGGGCTGC 9
SRR2176206.10169540_92737551 -----GGGGGCTGC 9
SRR2176206.2703842_92737551 -----GGGGGCTGC 9
SRR2176206.14442079_92737550 -----GGGGCTGC 8
SRR2176206.3409892_92737549 -----GGGCTGC 7
SRR2176206.5336830_92737549 -----GGGCTGC 7
SRR2176206.9374928_92737543 -----C 1
SRR2176206.1150352_92737542 -----

PIP5K1A_human          TAAGATGGCGTCGGCCTCCTCCGGGCCGTCGCTTCGGTCGGTTTTTTCAT 100
PIP5K1A_gibbon         CAAGATGGCGTCGGCCTCCTCCGGGCCGTCGCTTCGGTCGGTTTTTTCAT 96
PIPSL_human           CAAGATGGCGTCGGCCTCCTCCGGGCCGTTCTCTTCGGTCGGTTTTTTC-T 99
PIPSL_orangutan       CAAGATGGCGTCGGCCTCCTCCGGGCCGTTCTCTTCGGTCGGTTTTTTCAT 100
PIPSL_gibbon          CAAGATGGCGTCGCTCCTCCTCCGGGCCGTTCTCTTCGGTCGGTTTTTTCAT 100
SRR2176206.21252451_92737557 CAAGATGGCGTCGGCCTCCTCCGGGCCGTTCTCTTCGGTCGGTTTTTTCAT 65
SRR2176206.15382647_92737557 CAAGATGGCGTCGGCCTCCTCCGGGCCGTTCTCTTCGGTCGGTTTTTTCAT 65
SRR2176206.8312962_92737581 CAAGATGGCGTCGGCCTCCTCCGGGCCGTTCTCTTCGGTCGGTTTTTTCAT 63
SRR2176206.23186180_92737557 CAAGATGGCGTCGGCCTCCTCCGGGCCGTTCTCTTCGGTCGGTTTTTTCAT 63
SRR2176206.20047362_92737555 CAAGATGGCGTCGGCCTCCTCCGGGCCGTTCTCTTCGGTCGGTTTTTTCAT 63
SRR2176206.22547589_92737555 CAAGATGGCGTCGGCCTCCTCCGGGCCGTTCTCTTCGGTCGGTTTTTTCAT 63
SRR2176206.13341916_92737554 CAAGATGGCGTCGGCCTCCTCCGGGCCGTTCTCTTCGGTCGGTTTTTTCAT 62
SRR2176206.8194596_92737552 CAAGATGGCGTCGGCCTCCTCCGGGCCGTTCTCTTCGGTCGGTTTTTTCAT 59
SRR2176206.10169540_92737551 CAAGATGGCGTCGGCCTCCTCCGGGCCGTTCTCTTCGGTCGGTTTTTTCAT 59
SRR2176206.2703842_92737551 CAAGATGGCGTCGGCCTCCTCCGGGCCGTTCTCTTCGGTCGGTTTTTTCAT 59
SRR2176206.14442079_92737550 CAAGATGGCGTCGGCCTCCTCCGGGCCGTTCTCTTCGGTCGGTTTTTTCAT 58
SRR2176206.3409892_92737549 CAAGATGGCGTCGGCCTCCTCCGGGCCGTTCTCTTCGGTCGGTTTTTTCAT 57
SRR2176206.5336830_92737549 CAAGATGGCGTCGGCCTCCTCCGGGCCGTTCTCTTCGGTCGGTTTTTTCAT 57
SRR2176206.9374928_92737543 CAAGATGGCGTCGGCCTCCTCCGGGCCGTTCTCTTCGGTCGGTTTTTTCAT 51
SRR2176206.1150352_92737542 CAAGATGGCGTCGGCCTCCTCCGGGCCGTTCTCTTCGGTCGGTTTTTTCAT 50
*****

PIP5K1A_human          CCTTTGATCCCGCGGTCCCTTCCTGTACCTTGCTCCTCAGCATCTGGAATC 150
PIP5K1A_gibbon         CCTTTGATCCCGCGGTCCCTTCCTGTACCTCGTCTCCTCAGCATCTGGAGTC 146
PIPSL_human           CCGTTGATCCCGCGGTCCCTTCCTGTACCTCGTCTCCTCAGCATCTGGAATC 149
PIPSL_orangutan       CCTTTGATCCCGCGGTCCCTTCCTGTACCTCGTCTCCTCAGCATCTGGAATC 150
PIPSL_gibbon          CCTTTGATCCCGCGGTCCCTTCCTGTACCTCGTCTCCTCAGCATTTGGAATC 150
SRR2176206.21252451_92737557 CCTTTGATCCCGCGGTCCCTTCCTGTACCTCGTCTCCTCAGCATCTGGAATC 115
SRR2176206.15382647_92737557 CCTTTGATCCCGCGGTCCCTTCCTGTACCTCGTCTCCTCAGCATCTGGAATC 115
SRR2176206.8312962_92737581 CCTTTGATCCCGCGGTCCCTTCCTGTACCTCGTCTCCTCAGCATCTGGAATC 113
SRR2176206.23186180_92737557 CCTTTGATCCCGCGGTCCCTTCCTGTACCTCGTCTCCTCAGCATCTGGAATC 113
SRR2176206.20047362_92737555 CCTTTGATCCCGCGGTCCCTTCCTGTACCTCGTCTCCTCAGCATCTGGAATC 113
SRR2176206.22547589_92737555 CCTTTGATCCCGCGGTCCCTTCCTGTACCTCGTCTCCTCAGCATCTGGAATC 113
SRR2176206.13341916_92737554 CCTTTGATCCCGCAGTCCCTTCCTGTACCTCGTCTCCTCAGCATCTGGAATC 112
SRR2176206.8194596_92737552 CCTTTGATCCCGCGGTCCCTTCCTGTACCTCGTCTCCTCAGCATCTGGAATC 109
SRR2176206.10169540_92737551 CCTTTGATCCCGCGGTCCCTTCCTGTACCTCGTCTCCTCAGCATCTGGAATC 109
SRR2176206.2703842_92737551 CCTTTGATCCCGCGGTCCCTTCCTGTACCTCGTCTCCTCAGCATCTGGAATC 109
SRR2176206.14442079_92737550 CCTTTGATCCCGCGGTCCCTTCCTGTACCTCGTCTCCTCAGCATCTGGAATC 108
SRR2176206.3409892_92737549 CCTTTGATCCCGCGGTCCCTTCCTGTACCTCGTCTCCTCAGCATCTGGAATC 107
SRR2176206.5336830_92737549 CCTTTGATCCCGCGGTCCCTTCCTGTACCTCGTCTCCTCAGCATCTGGAATC 107
SRR2176206.9374928_92737543 CCTTTGATCCCGCGGTCCCTTCCTGTACCTCGTCTCCTCAGCATCTGGAATC 101
SRR2176206.1150352_92737542 CCTTTGATCCCGCGGTCCCTTCCTGTACCTCGTCTCCTCAGCATCTGGAATC 100
**


```

図 14. オランウータン RNA-seq *PIPSL* 5'末端配列

青枠の配列はゲノムにマップされたシーケンス配列

赤枠は *PIP5K1A*(親遺伝子)はマップされた read が親遺伝子ではなく、*PIPSL* 配列である

ことを示す箇所

```

PIP5K1A_human      GGAAGAACCGGATTGAAAGAGAGCCAGGCCGCTGAGGGGGAGGGGGCTGCTAAGATGGCG 60
PIP5K1A_gibbon    GGAAGAGCCGGATTGAAA----GCCAGGCCGCTGAGGGGGAGGGGGCTGCCAAGATGGCG 56
PIP5L_human       GGAAGAGCCGGATTGAAAGAGAGCCAGGCCCTTGAGGGGGAGGGGGCTGCCAAGATGGCG 60
PIP5L_orangutan   GGAAGAGCCGGATTGAAAGAGAGACAGACCCTTGAGGGGGAGGGGGCTGCCAAGATGGCG 60
PIP5L_gibbon      GGAAGAGCCGGATTGAAAGACAGCCAGGCCGCTGAGGGGGAGGGGGCTGCCAAGATGGCG 60
1108_14882_34820  -----GGGGGAGGGGGCTGCCAAGATGGCG 25
1310_17759_33207  -----GGGGGAGGGGGCTGCCAAGATGGCG 25
2316_14547_40082  -----GGGGGAGGGGGCTGCCAAGATGGCG 24
1214_20098_47921  -----GGGGCTGCCAAGATGGCG 18
2105_10759_12471  -----GGGGCTGCCAAGATGGCG 18
2305_16606_59878  -----GGGCTGCCAAGATGGCG 17
1311_10481_48667  -----GGCTGCCAAGATGGCG 16
2214_18526_88710  -----GGCTGCCAAGATGGCG 16
1104_16750_46558  -----GCCAAGATGGCG 12
1308_3374_67232   -----CCAAGATGGCG 11
1112_9666_72350   -----CAAGATGGCG 10
2108_15938_51648  -----CAAGATGGCG 10
*****

PIP5K1A_human      TCGGCCTCCTCCGGGCCGTCGTCCTTCGGTCGGTTTTTCATCCTTTGATCCCGCGGTCCT 120
PIP5K1A_gibbon    TCGGCCTCCTCCGGGCCGTCGTCCTTCGGTCGGTTTTTCATCCTTTGATCCCGCGGTCCT 116
PIP5L_human       TCGGCCTCCTCCGGGCCGTTTCCTTCGGTCGGTTTTTC-TCCGTTGATCCCGCGGTCCT 119
PIP5L_orangutan   TCGGCCTCCTCCGGGCCGTTTCCTTCGGTCGGTTTTTCATCCTTTGATCCCGCGGTCCT 120
PIP5L_gibbon      TCGGCCTCCTCCGGGCCGTTTCCTTCGGTCGGTTTTTCATCCTTTGATCCCGCGGTCCT 120
1108_14882_34820  TCGGCCTCCTCCGGGCCGTTTCCTTCGGTCGGTTTTTCATCCTTTGATCCCGCGGTCCT 85
1310_17759_33207  TCGGCCTCCTCCGGGCCGTTTCCTTCGGTCGGTTTTTCATCCTTTGATCCCGCGGTCCT 85
2316_14547_40082  TCGGCCTCCTCCGGGCCGTTTCCTTCGGTCGGTTTTTCATCCTTTGATCCCGCGGTCCT 84
1214_20098_47921  TCGGCCTCCTCCGGGCCGTTTCCTTCGGTCGGTTTTTCATCCTTTGATCCCGCGGTCCT 78
2105_10759_12471  TCGGCCTCCTCCGGGCCGTTTCCTTCGGTCGGTTTTTCATCCTTTGATCCCGCGGTCCT 78
2305_16606_59878  TCGGCCTCCTCCGGGCCGTTTCCTTCGGTCGGTTTTTCATCCTTTGATCCCGCGGTCCT 77
1311_10481_48667  TCGGCCTCCTCCGGGCCGTTTCCTTCGGTCGGTTTTTCATCCTTTGATCCCGCGGTCCT 76
2214_18526_88710  TCGGCCTCCTCCGGGCCGTTTCCTTCGGTCGGTTTTTCATCCTTTGATCCCGCGGTCCT 76
1104_16750_46558  TCGGCCTCCTCCGGGCCGTTTCCTTCGGTCGGTTTTTCATCCTTTGATCCCGCGGTCCT 72
1308_3374_67232   TCGGCCTCCTCCGGGCCGTTTCCTTCGGTCGGTTTTTCATCCTTTGATCCCGCGGTCCT 71
1112_9666_72350   TCGGCCTCCTCCGGGCCGTTTCCTTCGGTCGGTTTTTCATCCTTTGATCCCGCGGTCCT 70
2108_15938_51648  TCGGCCTCCTCCGGGCCGTTTCCTTCGGTCGGTTTTTCATCCTTTGATCCCGCGGTCCT 70
** *****

PIP5K1A_human      TCCTGTACCTTGTCTCAGCATCTGGAATCAAGAGACCCATGGCATCTGAGGTGCCTTAT 180
PIP5K1A_gibbon    TCCTGTACCTCGTCCTCAGCATCTGGAGTCAAGAGACCCATGGCATCTGAGGTGCCTTAT 176
PIP5L_human       TCCTGTACCTCGTCCTCAGCATCTGGAATCAAGAGTCCCATGGCATCTGAGGTGCCTTAT 179
PIP5L_orangutan   TCCTGTACCTCGTCCTCAGCATCTGGAATCAAGAGACCCATGGCATCTGAGGTGCCTTAT 180
PIP5L_gibbon      TCCTGTACCTCGTCCTCAGCATTTGGAATCAAGAGACCCATGGCATCTGAGGTGCCTTAT 180
1108_14882_34820  TCCTGTACCTCGTCCT----- 101
1310_17759_33207  TCCTGTACCTCGTCCT----- 101
2316_14547_40082  TCCTGTACCTCGTCCT----- 101
1214_20098_47921  TCCTGTACCTCGTCCTCAGCATC----- 101
2105_10759_12471  TCCTGTACCTCGTCCTCAGCATC----- 101
2305_16606_59878  TCCTGTACCTCGTCCTCAGCATCT----- 101
1311_10481_48667  TCCTGTACCTCGTCCTCAGCATCTG----- 101
2214_18526_88710  TCCTGTACCTCGTCCTCAGCATCTG----- 101
1104_16750_46558  TCCTGTACCTCGTCCTCAGCATCTGGAAT----- 101
1308_3374_67232   TCCTGTACCTCGTCCTCAGCATCTGGAATC----- 101
1112_9666_72350   TCCTGTACCTCGTCCTCAGCATCTGGAATCA----- 101
2108_15938_51648  TCCTGTACCTCGTCCTCAGCATCTGGAATCA----- 101
*****

```

図 15. ギボン RNA-seq *PIPSL* 5'末端配列

青枠の配列はゲノムにマップされたシーケンス配列

赤枠は *PIP5K1A*(親遺伝子)はマップされた read が親遺伝子ではなく、*PIPSL* 配列であることを示す箇所



## 7 謝辞

指導教官である大島一彦准教授には、研究室に入った学部4年次から現在まで、未熟な筆者を辛抱強く指導して頂くと共に、大学院での研究活動をする機会を与えてくださったことをこの場を借りて深くお礼申し上げます。

長浜バイオ大学の白井剛先生、斎藤修先生、山本博章先生には、研究に関する助言をいただき自分の研究に大変助かりました。

京都大学霊長類研究所ゲノム細胞研究部門の今井啓雄先生、生理学研究所の郷康広先生には貴重なサンプルを提供していただき、深く感謝しております。

最後に、大学院での研究生活を支えてくださった両親に深く感謝申し上げます。