病気に関係する遺伝子の探索

長浜バイオ大学 バイオサイエンス学部 生命情報科学コース バイオ分野の産業や研究分野では医学や医療に係わる課題が益々重要となってき た。本学が滋賀医科大学と「バイオ医療学」という課題で文科省の戦略的教育プロジ ェクトに採択されていることもこの反映である。世界的な状況としてはゲノム塩基配 列を高速で安価に解読する、「次世代シーケンサー」と呼ばれる塩基配列の解読装置 が開発されたことで、バイオや医療分野に大きな変化が起きつつある。日米欧の公的 研究機関を中心に、2000年に最初のヒトゲノム配列が解読された際には、約10年の 年月と世界全体で数千億円の経費がかかったと言われている。ところが、最近登場し た次世代シーケンサーを用いれば数百万円の価格と一週間程度の期間で、一個人のゲ ノム配列が解読可能となっている。さらに、数年以内には、数十万円で一人の個人ゲ ノムが一日以内で解読可能と予想されている。人間ドックと似た発想で、予防医学の 観点から個人のゲノム配列が解読される時期が近いと考えられている。

ヒトゲノム配列の解読、特に特定の個人のゲノム配列の解読は、従来はプライバシ ーの問題を中心に論じられる傾向にあり、何か怖いことをする研究と考えられがちで あった。しかしながら、現在の医学やゲノム科学の発展に伴い、ほとんどの人のゲノ ム配列内には、なんらかの重篤な病気と係わる遺伝因子を持っていることが明らかと なった。ヒトのゲノム配列が、進化の産物として出来あがったことを考えると、人類 の進化の過去の過程において、80歳や100歳まで生きるように淘汰されて来たわけで はない。ほとんどの人が100歳を超える程には長生きはしておらず、高齢になればな んらかの病気を発症する。ウイルスや微生物の感染や生活環境が引き起こす病気も多 いが、ガンを代表例として、各個人のゲノム配列に起因して発症をする病気も多い。 これらの病気について、それを遺伝子やゲノム配列上の多様性の一環として、理解す るようになってきている。「ゲノム配列の多様性」が人類集団にとって重要との認識 も深まっている。

ゲノム科学や医学の現状を考えると、これからの高齢化社会における予防治療・予防医学においては、希望する個人が、各自のゲノム配列を解読することが有意義と考えられる。さらには、ガンを発症した患者を例にとっても、その患者に適した抗がん剤を処方するための「テーラーメード・オーダーメード医療」を行うには、その患者の遺伝子やゲノム配列の特徴を知っておくことが重要となる。プライバシーの問題に十分に配慮すれば、個人ゲノムの解読は、医学・医療に大きく貢献できる。

日米欧の公的研究機関とは別個に、ベンチャー企業としてヒトゲノム解析を独自に 行ったベンター博士が、彼のゲノム配列を全世界に公開している。DNAのダブルヘリ ックス構造をクリック博士とともに発見した、ワトソン博士も彼の配列を公開してい るが、彼らの行動は上述の世界の流れを先導しようとの試みと思える。彼らのゲノム 配列中にも当然のこととして、疾患の原因となる遺伝子配列の特徴が既に見つかって いる。それにも係わらず、あえて公開したことは、予防治療・予防医学の発展の視点 からは意義深いことである。しかしながら、ゲノム配列は彼らの親族とも共通性が高 いので、全世界に公開することに親族からの異議が提起されているとも聞く。ベンタ ー博士やワトソン博士のように、個人名入りでゲノム配列を世界に公開することの是 非には問題があろうが、この問題を十分に承知の上で、彼らは敢えて個人ゲノム情報 を公開している。その意味は、ゲノム配列に起因して発症をする病気は、遺伝子やゲ ノム配列上の多様性の一環として理解すべきであることを伝えるための、先導的な行 動とも考えられる。全地球レベルでの生物の多様性が重要なように、ヒトのゲノム配 列の多様性も重要と考えられるようになっている。

今回の本実習では、これらの世界的な動向ならびに、更には1~2年内には次世代 シーケンサーを用いて1000人レベルの個人のゲノムが解読される状況を考え、また その解読自体が医療に関係した新しいバイオ産業を創出することを想定して、ヒトゲ ノム中の病気と関係した遺伝子配列を収集する。次世代シーケンサーは、我が国でも 百台近くが導入済みか導入予定であるが、その次世代シーケンサーが産出する大量な ゲノム配列の情報解析をする人材が極端に不足している。この我が国の状況を考え、 生命情報の3回生の後期の実習では、次世代シーケンサーが産出する大量配列の情報 解析を行っている。

世界の状況には遅れるとは思うが、我が国でも数年以内には個人のゲノム配列の解 読が普及を始めると思える。個人情報の管理に十分な配慮を行った上で、個人ゲノム 配列の情報解析を行い、「各個人がどのような病気に罹る可能性が高いのか、それを できるだけ避けるにはどのような注意が必要なのか」を、個人別に明らかにすること が重要になる。本実習は、その種のゲノム情報解析の準備と言える。可能であれば、 ベンター博士やワトソン博士がどのような病気になる可能性が高いのかを、本実習中 に探索してみたい。この種のテーマに驚きを持つかもれしないが、これからの医療を 考える上では世界的に重要となる課題である。我が国においては、研究と教育の両面 において、世界に遅れをとっている傾向にあるので、敢えて挑戦を試みよう。

病気に関係する遺伝子と言った場合、いかにも「病気を引き起こすための遺伝子」 がヒトゲノム中に存在するように聞こえるが、それは誤解である。例えばガンを例に とると、ガンを引き起こす(あるいはガンに関係する)遺伝子と表現した場合、その 着目する遺伝子自体が「ガンを引き起こすことを目的として存在している遺伝子」で はない。着目遺伝子自体は、通常はヒトにとって重要な働きをしている遺伝子である。 その遺伝子配列の一部に突然変異が起こることで、その遺伝子の生産するタンパク質 の機能が変化や欠損を起こし、その結果としてガンを生じさせる場合が考えられる。 一方で、タンパク質の配列には一切変化を与えないが、その遺伝子の発現を制御して いる領域に変化(突然変異)を起こすことで、遺伝子の発現状態が変化し、その結果 としてガンを生じさせる例も多い。本来は、発生の特定の時期に特定の臓器でのみ発 現をすべき遺伝子が、発生の異なった時期、ないしは異なった臓器で発現してしまう ことがガンにつながる例も多い。このような予備的な知識を頭に入れた上で疾患(病 気)に係わる遺伝子の探索を行おう。

I. まず、各自で自分の関心のある病気を考えてほしい。ここでは個々人のゲノムの配列に関係する疾患を考えているので、ウイルスや細菌が引き起こす感染症、(例えばインフルエンザや食中毒等)は対象から外す必要がある。各自で関心のある病気について、正確な知識を得ておくためにメルクマニュアル医学百科(Google で検索せよ。http://mmh.banyu.co.jp/mmhe2j/index.html)を参照しておこう。勿論、メルクマニュアル医学百科を見てから、関心のある病気を決めて良い。ガンのように広い範囲の疾患を選択すると以降の操作が難しくなるので、肺ガンや白血病のように、より限定した疾患名を選択する。

Google で "メルクマニュアル医学百科"をキーワードにして検索する。 次の画面で、メルクマニュアル医学百科[家庭版]をクリックする。





メルクマニュアル医学百科 トップ画面(<u>http://mmh.banyu.co.jp/mmhe2j/index.html</u>) 興味のある病気、例えば"白血病"とキーワードを入力して、検索ボタンをクリックすると、 次の検索結果画面が表示される。

	<mark>検索結果</mark> "白血病"
メルクマニュアルとは	表題 例えば"白血病"をクリックする からの検索
医子情報提供の歴史 >>	白血病 白血病は、白血球または成熟して白血球になる細胞のがんです。 <u>急性リンパ球性白血病</u>
を検索 検索	急性リンパ球性白血病は、正常ならリンパ球になる細胞ががん化して、急速に骨髄の正常細 胞に取って代わる、命にかかわる病気です。 <u>急性骨髄性白血病</u>
索] "卢贞"之及世,方	急性骨髄性(骨髄芽球性、骨髄単球性)白血病は、正常な状態では好中球、好塩基球、好酸 は、逆球によりまえないない、ない、ない、ない、ない、ない、ない、ない、ない、ない、ない、ない、ない

"<u>白血病</u>"をクリックすると、次の解説画面が表示される。



課題1)関心のある病気の選択を終えたら、その病気について、メルクマニュアル 医学百科の記載事項で、特に重要と考えられる部分をExcelファイルに記載をし、自 分が設定した病気を十分に理解しておく。この際、自分の設定した疾患がウイルスや 微生物よる感染症でないことも再度確認をしておく。着目した疾患について、メルク マニュアル医学百科なかで、遺伝子が関係することを記載してある際は、その記載事 項をExcelファイルに記載せよ。以降の解析では、着目した疾患の英語名を知ること も重要である。メルクマニュアル医学百科の記載事項中に英語名が出ていれば、それ をExcelファイルに記載せよ。併せて、ライフサイエンス辞書(Googleで検索せよ) のオンライン辞書サービスを用いて、疾患の英語名を調べてExcelファイルに記載せ よ。複数の英語名がある場合には、それらも、/ で区切って、Excelファイルの同じ セル内に記載せよ。メルクマニュアル医学百科中で記載されている英語名と一致する のかも調べよ。もし一致しない場合には、もう一度メルクマニュアル医学百科の記載 事項を確認せよ。

ライフサイエンス辞書は、Google で"**ライフサイエンス辞書**"あるいは"LSD"というキー ワードを入力し、検索ボタンをクリックする。

次の画面が表示される。ライフサイエンス辞書プロジェクトをクリックする。

Google	ライフサイエンス辞書	検索
0	約 524,000 件 (0.11 秒)	検索オプション
3] すべて ▼ もっと見る	他のキーワード: <u>Isd ライフサイエンス辞書</u> ライフサイエンス辞書 fire	ライフサイエンス辞書 オンライン fox ライフサイエンス辞書 iphone
ウェブを検索 日本語のページを検索	<u>ライフサイエンス辞書 pdi</u>	<u>。</u> クリックする
▼その他のツール	<u>ライフサイエンス辞書プロジェクト</u>	
	生命科学の学問領域で使われる専門用語 及びソフトダウンロードを無償で提供。	などのオンライン 辞書 、変換サービス

次の画面で、"WebLSD オンライン辞書"をクリックする。

LIFE SCIENCE DICTIONARY	LIFE SCIE ライフサイコ	ENCE DICTIONARY ニンス辞書プロジェクト	フォントサイズ:1 业 1 虫 PROJECT	太 掊太 ENGLISH	
			ミラー	サイト: <u>京都1</u> <u>京都2</u> <u>東京</u>	
ホーム	サービス	資料	About Us	ダウンロード	クリックする
オンライン辞書	++7				
英和·和英検索/2010 張! 共起検索/8千5	0年1月1日改訂。英語10 万語の生命科学コーパス	1万語,日本語11万語に拡 から検索。	WebLSD オン	ライン辞書	
オンデマント英	語教材				
マスコミ科学記事を注 適。	主釈つきで学習できます。	大学院入試の勉強にも最	オンデマンド英	英語教材 う	
最新記事					
- 貝知らめ人への恐	怖いから生まれる人種的	5(原員 (20100505)			

次の画面で調べたい単語(英語、日本語いづれも可)を入力し、探索する。

	ONLINE	LIFE SCII	ENCE DIC サイエンス辞書፣	TIONARY tンラインサービ	z
ブロジェクト	WebLSD	0/D英語教材	EtoJ Vocab	EtoJ	WebSpell
		ミラーサイト[京	(都1 京都2 東京]	フォントサイズ[小]中	大 持大] ENGLISH
英和·和英	シソーラ	ラス 英語	新規対 共起表現	訳の受け付け 英和・和 オプ・	^{英検姆} クリックす アン表示/非表示
英和·和英検索:				searc	h
▶お知らせ 調	べたい単語を	:入力し、" se	arch"をクリ	リックする	
• iPhone用 WebL	SD mini はじめまし	た。WebLSDの機能!	おまぼそのままに、i	PhoneやXperiaなどの)スマートフォ 📍

ライフサイエンス辞書 http://lsd.pharm.kyoto-u.ac.jp/ja/service/weblsd/index.html

課題2) 次にその日本語の疾患名を Jabion 遺伝子百科(Google で検索せよ)の検索 の項目に入力し、その疾患に関する遺伝子を探索する。Jabion 遺伝子百科においては、 その疾患ならびにその疾患に関する遺伝子の日本語での解説部分が存在するので、そ の部分を Excel ファイルに記載をしておく。まずその疾患と関係をしている遺伝子が 本来果たしている正常な働きを記載し、突然変異を起こして病気を引き起こすことに ついての記載があれば、その突然変異に関する事項も Excel ファイルに記載する。な お Jabion 遺伝子百科の検索結果を見ると、大半の病気について複数の遺伝子が関係 していることが分かる。各々の遺伝子について、Excel ファイルの別の行に記載する こと。

課題3)次に、正常な機能をはたしている際の着目遺伝子の配列(言い換えれば、病気を引き起こす可能性のある突然変異を起こしていない配列)を Jabion 遺伝子百科に備わった機能で収集を行い、Excel ファイルに記載する。この際、以下の4種類の配列を収集する。

- (1) cDNA 配列
- (2) CDS (アミノ酸翻訳部分)の塩基配列
- (3) 遺伝子産物のタンパク質配列
- (4) ゲノム **DNA** 配列

なお、大半の病気について複数の遺伝子が関係している。各々の遺伝子について、上 記の4配列を収集して、遺伝子ごとにExcelファイルの別の行に記載する。 Jabion 遺伝子百科は、Google で"遺伝子百科"をキーワードに入力し検索する。次の様な 結果画面が表示されるので、"Jabion-日本語バイオポータルサイト-遺伝子百科"をクリックす る。



次の画面が表示される。



自分の関心のある病気名を、(ここでは例えば"白血病")キーワードに入力し検索ボタンを クリックすると、その病気に関連する遺伝子の一覧が表示される。(病気によっては Jabion 遺 伝子百科では表示されない場合もある。)

Jabion	国初 ビギナーとプロの	ための日本語	╭ <u>_</u> _ [_] // ⁰⁰ /233	h	ab	io	n
用時辞書 コラム一番	3 遺伝子百利	文計論家	研究プロジェクト	新日報	日本語バー	イオボータノ ンク トップ	レサイ ページ
遺伝子百科		~ 10/12 7	WINDYHAVEVI	274	001 <u>x</u> ±57	00 000	
	_						
シャーワート検知	来	(土・ワードの		« / hamaalahini	`		
	候補小發現		1977 [19]: 肖/	かん hemoglobin.	,		
	大阪街が トーー	an Ot a	ふた、連切・	+~ ~~	でけ説り	1 П Г Ъ	
検索結果:	土物性/ num	ロロクセク	いり、選択	りる。 ーー ン#ひょう	ては記り	л с	
		を選択した	こ例で説明を	進める。	H- Hom		
道伝ナンノハレン		遺伝	子名		11初 種	Link	
ABCC5	ATP-binding cassett	e, sub-family C	(CFTR/MRP), me	mber 5	human	NCBI	
<u>Abl1</u>	v-abl Abelson murine	leukem	ia oncogene 1		mouse	NCBI	
ABL1	v-abl Abelson murine	leukem	a viral oncogene	homolog 1	human	NCBI	
Abl1_mapped	v-abl Abelson murine (mapped)	leukem	ia viral oncogene	homolog 1	rat	<u>NCBI</u>	
Abl2	v-abl Abelson murine Abelson-related gene	leukem	ia viral oncogene	2 (arg.	mouse	NCBI	
ABL2	v-abl Abelson murine (arg, Abelson-related	e <mark>leukem</mark> I gene)	ia viral oncogene	homolog 2	human	NCBI	
Abl2 mapped	Abelson murine et (Mapped)	ukemia _{vi}	al (v-abl) oncoger	e homolog 2	rat	NCBI	
ANPEP	alanyl (membrane) a aminopeptidase M, m	minopeptidase (icrosomal amin	aminopeptidase N, opeptidase, CD13,	p150)	human	NCBI	
ANXA8	annexin A8				human	NOBI	

"ABL1"をクリックすると次の画面が表示される。

Jabion ABL1	
Homologeneの相同遺伝子(HomoloGene: <u>3783</u>)	
H.sapiens ABL1 c-abl oncogene 1. receptor tyrosine kinase.	Ptroglodytes ABL1 c-abl oncogene 1, receptor tyrosine kinase.
NP 009297_2 1149 aa	XP 001166213.1 1149 aa (XM 001166213.1)
C <u>familiaris</u> ABL1 c-abl oncogene 1, receptor tyrosine kinase. XP 5484132 1131 aa XM 5484132 D	B taurus ABL1 c-abl oncogene 1, receptor tyrosine kinase. XP 613548.2 XM 613548.4)
M.musculus Abi1 c-abi oncogene 1, receptor tyrosine kinase.	Rnorvegicus Abl1 c-abl oncogene 1, receptor tyrosine kinase.
G.gallus ABL1	D.rerio LOC100000720

画面を下にスクロールしていくと、次の画面が現れる。要約として、着目している遺伝子の機能 や特徴等が日本語で記載されているので、この部分を読むこと。

H.sapiens H.sapiens(HA)	1) P.troglodytes M.musculus R.norvegicu	s Cfamiliaris G.gallus Q.latipes	
Basic Information			
遺伝子シンボル:ABL1(A	BL, JTK7, bcr/abl, c-ABL, p150, v-ab)	
遺伝子名: c—ablor	ncogene 1		
要約		ゲノム配列上位置	
ABL1《原癌遺伝子》は細胞質お 化》 細胞公列 細胞培美 たけ	よび核の《チロシンリン酸化酵素》を指令し、細胞《分	chromosome 9 Chr. 9 Chr. 9	
の活性はSH3領域によって抑制	い人下レ人に合い通信に関係する。	133763062	
広于川に変える。t(9:22)転座は1 (MIM: 151410)とABL1 遺伝子の	愛性的な骨髄性&日囲病》の多くの症例におけるBCR 直列融合をもたらす。広く発現されるABL1チロシンリ	extract sequence	
ン酸化酵素のDNA結合活性はC 《細胞周期》機能を示している。	;DC2の仲介するリン酸化によって調節され、ABL1の ABL1 遺伝子は6か7キロ塩基のmRNA転写物として発	エクソン・イントロン構造(クリックするとゲノム	ビューアヘジャンプします)
現され、選択的スプライシングに ろ。	こより一般的なエキソン2-11に第一エキソンが継がれ	NCBI	ABLI
		ABLI	
オリジナルDBにおけるID・	ACC クリックするとオリジナル DB ヘジャンプします)	403110	ENS F00000400412
no. transc	ript ACC peptide ACC	407465	
1 <u>NM 00</u> 2 <u>NM 00</u>	05157.3 NP 005148.2 07313.2 NP 009297.2	136 1426	BNS F00000 32 3315
Ensembl ENSG00000	<u>097007</u>	Necan	chr9.pasa.3333.a chr9.pas
no. transcrip 1 ENST00	ot ACC peptide ACC 0000372348 ENSP00000361423	8M 133.60M 133.62M 133.64M 133	.son 133.sem 133.70m 133.72m 133.74m 133.76m
2 ENST00	1000393293 ENSP00000376971	Ensembl	
4 ENST00	000318560 ENSP0000323315 000426772 ENSP00000406308	HISGUUUU224747	
5 ENST00 6 ENST00	000438426 ENSP00000407756 0000444970 ENSP00000400412		
			
オリジナルDBI	こおけるID・ACC(クリックす-	るとオリ <u>ENSG00000097</u>	<u>007</u> をクリックすると、
NCBL	25	全てのスプラ	イシングバリアントの
ПОВГ			、手できる画面が表示さ
Alignened RefSeq	1 NM 007212.2	や れる。	
	1 <u>NW 007313.2</u>		
Ensembl	ENSG00000097007		
	no. transcript ACC	peptide ACC	
	1 ENST0000372348	ENSP00000361423	1
	2 ENST0000393293	ENSP00000376971	1
	3 ENST0000318560	ENSP00000323315	1
	4 ENST00000426772	ENSP00000406308	1
	5 ENST0000438426	ENSP00000407756	1
	6 ENST00000444970	ENSP00000400412	1
Massa	11110 H0C0 KKKK 0		_

|--|

C.Ensembl							2] -	
Home > Human [GRCh37]						Login / R	egister BLAST	/BLAT BioMa
Location: 9:133,589,333-133,76	3,062 Ge	ne: ABL1						
Gene-based displays		🖰 You'ı	/e been redirected	to your nea	rest mirror - uswes	t.ensembl.c	org 🗵	
Gene summary Splice variants (6) Supporting evidence		• Tal	ke me back to <u>www.e</u>	ensembl.org				
 Sequence External references (3) 	Gene:	ABL1 (EN	ISG0000009700	7)				
- Regulation - Comparative Genomics - Genomic alignments (52)	c-abl onc	ogene 1, re	ceptor tyrosine kinas Chromosome 9: 1	e [Source:HG] 33.589.333-13	NC Symbol;Acc:76] 3 763 062 forward str	and.		
Gene Tree (image)	Transcri	ipts	There are 6 transc	ripts in this ge	ne: hide transcripts			
Gene Tree (text) Gene Tree (alignment)		Name	Transcript ID	Length (bp)	Protein ID	Length (aa)	Biotype	CCDS
- Orthologues (47)		ABL1-001	ENST0000318560	5766	ENSP00000323315	1130	Protein coding	CCDS35166
Paralogues (17) Protein families (4)		ABL1-002	ENST0000372348	3824	ENSP00000361423	1149	Protein coding	CCDS35165
Genetic Variation		ABL1-003	ENST0000393293	532	ENSP00000376971	64	Protein coding	-
Variation Table		ABL1-201	ENST00000426772	144	ENSP00000406308	48	Protein coding	-
		ABL1-202	ENST0000438426	3543	ENSP00000407756	265	Protein coding	-
Personal annotation		ABL1-203	ENST00000444970	345	ENSP00000400412	115	Protein coding	-
Gene history	🛈 TI	ranscript a	and Gene level dis	plays				
✤ Configure this page	l In E	Ехро	rt data を	クリック	する。 ^{vide di}	splays at two	levels:	
🕞 Manage your date		 Transcrip domain a 	t views which provide nnotation.	e information s	pecific to an individua	l transcript su	ich as the cDNA	and CDS sequ
🔒 Export data		Gene view	ws which provide disp	plays for data a	associated at the gene	e level such a	s orthologues ar	nd paralogues, r

次の画面が表示され、送ってもらうデータ(Export data)の内容や形式を指定できる。

Export Configuration - Feature List	
Location to export:	chromosome:GRCh37:9:133589333:133763062:1
Gene to export:	ENSG0000097007 (ABL1)
Output:	FASTA sequence
Strand:	Feature strand 💌
5' Flanking sequence (upstream):	0 * (Maximum of 1000000)
3' Flanking sequence (downstream):	0 * (Maximum of 1000000)
	Next >
Fields marked with * are required	
Options for FASTA sequence	
Genomic:	Unmasked 💙
Select/deselect all	
cDNA	
Coding sequence	
Peptide sequence	
5' UTR	
3' UTR	
Exons	
Introns	

Export Configuration - Feature List	
Location to export:	chromosome:GRCh37:9:133589333:133763062:1
Gene to export:	ENSG0000097007 (ABL1)
Output:	FASTA sequence
Strand:	Feature strand
5' Flanking sequence (upstream):	0 * (Maximum of 1000000)
3' Flanking sequence (downstream):	0 * (Maximum of 100000)
	Next >
Fields marked with * are required	
Options for FASTA sequence	ピクリックする
Genomic:	Unmasked 💌
Select/deselect all	
cDNA	V
Coding sequence	V
Peptide sequence	
5' UTR	
3' UTR	Exons, Introns
Exons	□ 「 のチェックを外し INext>1 をクリックする
Introns	
この画面で、Options for FASTA sequence の設定で、	
'UTR, 3'UTR, Exons, Introns の 🛛 チェックを外	して [Next>]をクリックする。
Export Configuration - Output Format	

💻 Text をクリックする

Please choose the output format for your export

• <u>HTML</u>

Text

<u>Compressed text (.gz)</u>

< Back

が表示され、最後に 全ゲノム DNA 配列 が表示される。

同じパターンで、					
ENST00000393293	の	cdna,	cds,	peptide	の配列
ENST00000426772	の	cdna,	cds,	peptide	の配列
ENST00000438426	の	cdna,	cds,	peptide	の配列
ENST00000444970	の	cdna,	cds,	peptide	の配列
	同じパターンで、 ENST00000393293 ENST00000426772 ENST00000438426 ENST00000444970	同じパターンで、 ENST0000393293の ENST0000426772の ENST0000438426の ENST00000444970の	同じパターンで、 <u>ENST00000393293</u> の cdna, <u>ENST00000426772</u> の cdna, <u>ENST00000438426</u> の cdna, <u>ENST00000444970</u> の cdna,	同じパターンで、 <u>ENST00000393293</u> の cdna, cds, <u>ENST00000426772</u> の cdna, cds, <u>ENST00000438426</u> の cdna, cds, <u>ENST00000444970</u> の cdna, cds,	同じパターンで、 <u>ENST00000393293</u> の cdna, cds, peptide <u>ENST00000426772</u> の cdna, cds, peptide <u>ENST00000438426</u> の cdna, cds, peptide <u>ENST00000444970</u> の cdna, cds, peptide

>ENSG0000097007:ENST00000372348 peptide: ENSP00000361423 pep:KNOWN_protein_coding MGQQPGKVLGDQRRPSLPALHFIKGAGKKESSRHGGPHCNVFVEHEALQRPVASDFEPQG LSEAARWNSKENLLAGPSENDPNLFVALYDFVASGDNTLSITKGEKLRVLGYNHNGEWCE AQTKNGQGWVPSNYITPVNSLEKHSWYHGPVSRNAAEYLLSSGINGSFLVRESESSPGQR

>ENSG0000097007:ENST00000318560 peptide: ENSP0000323315 pep:KNOWN_protein_coding MLEICLKLVGCKSKKGLSSSSSCYLEEALQRPVASD NDPNLFVALYDFVASGDNTLSITKGEKLRVLGYNHNGEW SLEKHSWYHGPVSRNAAEYLLSSGINGSFLVRESESSPG SLEKHSWYHGPVSRNAAEYLLSSGINGSFLVRESESSPG (中略)

cDNA 配列をコピーして excel に貼り付ける。

>ENSG0000097007:ENST00000318560 cdna:KNOWN_protein_coding GGCCTTCCCCCTGCGAGGATCGCCATTGGCCCGG

TTGGGCCGGGCTCGGCCTCGGGAACGCCAGGGGCCC CDNA 配列

次の画面が表示される。

>9 dna:chromosome chromosome:GRCh37:9:133589333:133763062:1 TGTGACATTGAAATGACTAGCATTATTGACCCTTTTCAGCATCCCCTGTGAATATTTCT GTTTAGGT TTTCTTCTTGAAAAGAAATTGTTATTCAGCCCGTTTAAAACAAATCAAGAA ACTTTTGGG ゲノム DNA 配列 ABL1 の例 CTCCT ">"で始まる先頭行のみコピーして excel に貼り付ける GCTTGCAAG CTGTG ABL1 の例では 17 万文字程度の長い配列の為、excel の1マスに貼り TGCCA ATTGACTTC 付けることが出来ないので、コメント行(">"で始まる行)のみを excel GGGTA TTTCCCTCT に貼り付ける。 CCTATTATT GCCAA GCTTGCCTGCCCTGCATTTTATCAAAGGAGCAGGGAAGAAGGAATCATCGAGGCATGGGG GTCCACACTGCAATGTTTTTGTGGAACATGGTGAGTGCTTTTCAAAATTTCTGCTCATGG (以下略; ABL1の例では約17万文字の長い配列)

このように、遺伝子 ABL1 の場合では、

	· ABL1-001 Transcript ID	ENST00000318560	の	cdna,	cds,	peptide	の各配列
	ABL1-002 Transcript ID	ENST00000372348	の	cdna,	cds,	peptide	の各配列
	ABL1-003 Transcript ID	ENST0000393293	の	cdna,	cds,	peptide	の各配列
)	ABL1-201 Transcript ID	ENST00000426772	の	cdna,	cds,	peptide	の各配列
7	ABL1-202 Transcript ID	ENST00000438426	の	cdna,	cds,	peptide	の各配列
	ABL1-203 Transcript ID	ENST00000444970	の	cdna,	cds,	peptide	の各配列
	そして最後に、						
	・ 全ゲノム DNA 配列						

計19の配列情報が表示される。

ABL1の例では、合計 19の各配列情報(FASTA形式)を、レポートの excel 様式の所定の欄に貼り付けることになる。

- 注) テキストの "ABL1" の例では 19 の配列情報であるが、調べる遺伝子によって、その数 (スプライシングバリアントの数) は異なる。
- 注) FASTA 形式の配列では、行頭に ">" で始まるコメントの行があり、その次の行以降 に配列が記述されている。各配列をコピーする時は、">" で始まるコメントの行も 含めてコピーすること。

ここで、自分が選択した疾患の日本語を Jabion 遺伝子百科で検索しても、遺伝子の情報が得られない場合を考える。Jabion を使わずに、選択した疾患に関する遺伝子の情報を得る方法を例にして、バイオ分野・生命科学分野で、信頼性の高い情報を日本語文献類から検索する方法を紹介する。具体的には、統合データベースセンター (DBCLS)の作成している統合ホームページ(http://lifesciencedb.jp/)にある、横断検索の機能を使って、選択した疾患に関係する遺伝子に関する情報を探し出そう。

Google で"LSDB"をキーワードにして検索する。

Google	LSDB		検索
0	約 342,000 件 (0.09 秒)	食索才	ブション
3]すべて ▼ もっと見る	他のキーワード: <u>Isdbとは ospf</u>	le la construcción de la	
ウェブ全体から検索 日本語のページを検索	<u>ライフサイエンス統合データベース</u> 文部科学省「統合データベースプロジェ	<u>スプロジェクト-統合ホームページ-LSDB</u> クト」のボータルサイト。生命科学分野のデータベー	ースカ
■その他のツール	タログや横断検索、総説や特許、研究報告書の文献検索、ツール、バイオNLP、教材、事業 説明など、		
	lifesciencedb.jp/ - キャッシュ - 類似ペー	-9	
	生命科学データベース横断検索	生命科学系データベースアーカイブ	
	データベース	サイトボリシー	
	「統合データベースプロジェクト」とは	蛋白質核酸酵素全文検索	
	生物アイコン	DNAデータベース総覧と検索	

ライフサイエンス統合データベースプロジェクト-統合ホームページ-LSDBクリックすると次の画面が表示される。



まず「統合ホームページへようこそ」の空白欄へ、「選択した病気名 原因遺伝子」と記入 して、横断検索の文字部分をクリックすると、キーワードが日本語なので、我が国で公開して いる多数の日本語文献やデータベースを横断的に検索してくれる。



文献の左の + をクリックすると、次の画面が表示される。



左上のデータベースの欄の全てのデータベース(数字)の括弧内の数字は、該当したキ

ーワードの持つ(キーワードを2つ入れた場合はその両方を持つ)文献数を示している。括 弧内の数字がゼロでないからと言って、この検索操作で着目する病気の原因遺伝子が探し 出せた訳ではないことに注意して欲しい。その文献内に、「選択した病気名」と「原因遺伝 子」の両方が存在する文献の全体を探し出しただけである。その文献中で、2つの用語が離 れた場所に登場したのでは、意味がないことが多い。この実習では、「データベースの欄」の 「全てのデータベース(数字)」の下の「統合 DB プロジェクト」の更に下ある、「文献(数字)」の 項目の数字がゼロでない場合にはそこをクリックし、最初に位置している蛋白質核酸酵素 (数字)をクリックしてみる。この操作で、「蛋白質核酸酵素」と呼ばれている雑誌の掲載記事 の中で、「選択した病気名」と「原因遺伝子」が登場する部分が右の欄に表示される。長い文 献中で、検索のキーワードが登場する文章の一部分を切り出して、羅列しているだけなので、 画面上の表示で2つの用語が近くにあっても、文献中で近くに位置していたとは限らない。 しかし、文章を注意深く読めば、「選択した病気名」の「原因遺伝子」が正確に記載されてい る例も存在する。間違った情報を収集しないためには、複数の文献を比較して、共通的に 登場する遺伝子名を探し出す事が重要になる。並行して、蛋白質核酸酵素の検索結果とし て得られた「遺伝子名(遺伝子シンボルでも良い)」の方に着目して、その用語と注目してき た「病気名」の両方を「統合ホームページの横断検索」のキーワードとして検索して見よう。 今まで登場した以外の文献を注意深く読むことで、得られた情報の確実性を検証できる。

雑誌「蛋白質核酸酵素」の各文献の PDF ファイルが用意されているので、下線の付いている論文名をクリックすると、その全文を読むことが出来る。但し、出版社側の方針で、印刷やダウンロードが出来ない。蛋白質核酸酵素は専門家を想定して書かれているので、やや難しく感じるかもしれない。文献の項目内の蛋白質核酸酵素以外の雑誌等に掲載記事があれば、そちらの方を参照することも考えられる。あるいは、文献以外の項目、例えば学会要旨や特許関連文章等を参照することも考えられる。そのように複数の情報源を調べることで、自分の選択した病気の原因となる遺伝子を正確に探し出せる。

ここで再度指摘しておくべき点は、統合データベースセンター(DBCLS)が提供している横 断検索は、生命科学分野の広範囲な用語を用いた検索を可能にしていることである。更に、 文献・用語解説・学会要旨・特許関係等を指定して内容を参照出来る点も便利であり、更に 検索できた文献を参照する際に、雑誌名や新聞名を指定出来るので、信頼のおける情報が 能率的に取得できる。一般の検索サイトを用いた検索では、信頼度の低い情報が混在する 傾向が避けられないので、統合データベースセンターの提供する横断検索のように、信頼 度の高い情報源からの情報収集が可能な横断検索の意義は大きい。学会要旨のような先 端的な情報を含む例については、作業仮説のような事項の記載もあるので、確実な事柄に ついての情報収集を目的とする場合には、学会要旨や科学研究費の報告書等は使用しな い方が安全である。このような情報源の絞り込みが容易に行えることも、この横断検索の特 徴と言える。

統合データベースセンターの横断検索で遺伝子を検索することで、Jabion に記載されて いない病気の遺伝子を探し出せる可能性が高い。この検索の過程で、該当する遺伝子の英 語名や遺伝子シンボルを探し出せる可能性も高い。遺伝子の日本語名しか見つからない場 合には、ライフサイエンス辞書等を利用して英語名を探し出す。遺伝子の英語名が見つか れば、Ensemb1 のような外国で作成しているゲノム配列情報に関するデータベースを 直接に参照して遺伝子の配列情報を取得できる。Ensemb1 の一般的な使用方法は、統 合 TV (Google で検索せよ) と呼ばれる動画サイトに、使用法が日本語で解説されて いる。その動画を参考にして Ensembl から上記の4種(cDNA, CDS, peptide, 全ゲノム DNA)の配列を収集して、遺伝子ごとに Excel ファイルの別の行に記載する。但し、 Ensembl を用いた配列の取得は、研究者が行う操作であり、Ensembl の機能が余りに も多いので、使用することは簡単ではない。15~17頁にその操作を説明しているが、 困難さを感じたら、Jabion で遺伝子が検索できるように設定した病気を変更しても良 い。この際、以前に設定をして検索を行った結果は、この部分もレポートの評価の対 象に含めるので、Excel ファイルからは除去しないこと。新たに設定した病名を、Excel ファイルの別の行に記載し、上記の操作を再開始する。







CENSEMBI West Home > Human [GRCh37] Genome		Login / R
Search Ensembl	Results Summary	
└ New Search		クリックする
A Configure this page	By Feature type	
	Total	5
💼 Manage your data	Family	1
Export data	▼ Gene	4
Bookmark this page	Homo sapiens (4)	

<i>C</i> Ensembl [®]	en e) -
Home > Human [GRCh37]	Login / Register BLA	ST/BLAT BioMart Docs & F/
Search Ensembl	Result in Detail	クリックする
└ New Search	ABL1 corporate/tree:"Top/Species/Homo sapiens" c Search	
🥜 Configure this page	Your query matched 4 entries in the search database	al: APL1) [Pagion in detail]
📭 Manage your data	Ensembliprotein_coding Gene. ENSG0000001007 (HGNC Symb	
🕞 Export data	Description: c-abl oncogene 1, receptor tyrosine kinase [Source:HGf	NC Symbol;Acc:76]
🙀 Bookmark this page	Source: e57; Feature type: Gene; Homo sapiens; Species	: Homo sapiens; Gene;
	Havana protein_coding Gene: OTTHUMG00000020813 (ABL1) [R	tegion in detail]
	Description: c-abl oncogene 1, receptor tyrosine kinase	
[Region in de	tail]をクリックすると、グラフィカルな画面が表示	される。配列データ
を取得する前	前に、こちらの画面も確認しておくとよい。	

ClEnsembl [®]								
Home > Human [GRCh37]								Login / Register
Location: 9:133,589,333-133,763	3,062 Ge	ne: ABL1						
Gene-based displays	Gene:	ABL1 (EN	ISG00000097007	7)				
 Gene summary 	c-abl onc	ogene 1, re	ceptor tyrosine kinas	e [Source:HGN	IC Symbol;Acc:76]			
 Splice variants (6) 	Location	1	Chromosome 9: 13	33,589,333-13	3,763,062 forward str	and.		
- Sequence	Transcri	ipts	There are 6 transc	ripts in this ge	ne: hide transcripts	i		
- Regulation		Name	Transcript ID	Length (bp)	Protein ID	Length (aa)	Biotype	CCDS
Comparative Genomics		ABL1-001	ENST00000318560	5766	ENSP00000323315	1130	Protein coding	CCDS35166
Genomic alignments (52)		ABL1-002	ENST00000372348	3824	ENSP00000361423	1149	Protein coding	CCDS35165
Gene Tree (Image)		ABL1-003	ENST00000393293	532	ENSP00000376971	64	Protein coding	-
Gene Tree (alignment)		ABL1-201	ENST00000426772	144	ENSP00000406308	48	Protein coding	-
Orthologues (47)		ABL1-202	ENST00000438426	3543	ENSP00000407756	265	Protein coding	-
Paralogues (17)		ABL1-203	ENST00000444970	345	ENSP00000400412	115	Protein coding	-
Protein families (4) Genetic Variation Variation Table Variation Image External Data Personal annotation ID History Gene history Configure this page Manage your data Export data	クリ	ックす	 Transcript and In Ensembl a get Transcript sequences Gene view regulatory This view is a ge navigate to the ir information click 	nd Gene leve ne is made up views which p s and protein o s which provic regions and s re level view. formation you on the Gene t	el displays of one or more trans- provide information sp lomain annotation. Je displays for data as plice variants. To access the transc want using the menu ba in the menu bar at	cripts. We pro ecific to an inc ssociated at th ript level displ i at the left har the top of the	vide displays at dividual transcrip ne gene level sur ays select a Tra nd side of the pa page	two levels: ot such as the c ch as orthologu inscript ID in the ge. To return to
<u>この</u> 面面以降は	_ 先程 P	9からと	「同じ」各配な	別をレポー	- FØ excel	様式に貼	り付ける。	

(参考) 統合TV (動画による Ensembl の操作説明)

Google で "統合 TV 配列を取得する"をキーワードにして検索する。 Google 統合TV 配列を取得する 検索 約 31,600 件 (0.25 秒) 検索オブション 🛃 すべて 統合TV (togotv)[塩基配列] 🎬 動画 2007-09-27#p01 遺伝子発現バターンが似ている遺伝子群のゲノム上流配列を一気に取得 💌 もっと見る する; 2007-09-14#p01 ウイルスの持ち出した宿主の遺伝子配列が クリックする ミノ酸配列レベルでゲノム中から探し当てる; 2007-09-06#p01 st rectosirina ... ウェブを検索 日本語のページを検索 togotv.dbcls.jp/?category... - キャッシュ - 類似ページ その他のツール 統合TV (togotv) - Ensembl tips ~ 記列を取得する~ 2008年2月21日 ... 塩基配列やcDNA配列、アミノ酸配列は研究の多くの場面で必要となり ますが、今回紹介するちょっとした使い方を知るだけで自分の欲しい配列や領域をこれまで 以上に簡単に取得することができるかもしれません。画像をクリックするとFlash版 ..



統合 TV 画面 http://togotv.dbcls.jp/movie/090206en_f.html

Ⅱ. 次に病気を引き起こす突然変異を知る操作に移る。

この場合には、OMIM (Online Mendelian Inheritance in Man; ヒトの疾患デー タベース)を使用する。遺伝子が関係する病気に関する情報や、各病気を引き起こす 遺伝子の突然変異を収録している、世界最大規模で最も信頼性の高いデータベースで ある。

課題4)各自が選択した病気に関係する遺伝子について、病気を引き起こすことに 関係する遺伝子の突然変異のデータを収集する。タンパク質遺伝子について、病気と 関係しているアミノ酸配列上の変化の位置とその変化の様子(どのアミノ酸がどのア ミノ酸へと変化したか等)を Excel に記載する。そして、余力のある学生は その記 載した突然変異体が、人種によって特徴があるかどうか、あるいはベンター博士やワ トソン博士のゲノム塩基配列上に存在するのか等を調べてみる。

Google で "OMIM"をキーワードにして検索する。 次の画面が表示されるので、<u>OMIM Home</u>をクリックする。





ここで、白血病に関連する "ABL1"の遺伝子を OMIM で調べてみる。for の右の入

カボックスに ABL1 と入力し Go のボタンをクリックすると、次の画面が表示される。



検索結果の各番号の次に表示されている記号について、*, + は遺伝子に関するデータであることを示し、#, % は疾患に関するデータであることを示している。本自習では、遺伝子に関する データに着目する。この場合、*印のついた ABL1 に関する遺伝子データが一番目にヒットして いるので、これをクリックする。次の画面が表示される。

S NCBI Online Mendelian Inheritance in Man	hns <u>My NCBI</u> opkins <u>[Sign In]</u>
All Databases PubMed Nucleotide Protein Genome Structure	PMC OMIM
Search OMIM Y for Go	Clear
Limits Preview/Index History Clipboard Details	
Display Detailed Show 20 Send to V	
All: 1 OMIM UniSTS: 0 OMIM dbSNP: 0	
MIM *189980 ABELSON MURINE LEUKEMIA VIRAL ONCOGENE HOMOLOG 1; ABL1	Table of Contents MIM *189980 Description
Alternative titles; symbols	Cloning Gene Structure
TRANSFORMATION GENE: ONCOGENE ABL ABELSON STRAIN OF MURINE LEUKEMIA VIRUS; ABL ABL1/BCR FUSION GENE, INCLUDED ABL1/NUP214 FUSION GENE, INCLUDED	Gene Function Cytogenetics Molecular Genetics Animal Model
Gene map locus: <u>9q34.1</u>	Allelic Variants See Also
Description	References

この画面を下にスクロールしていくか、もしくは 右の Table of Contents の中の Allelic Variants

をクリックすると次の様な Allelic Variants の情報が現れる。 次の画面で <u>See allelic variants in tabular display</u> をクリックする。				
translocated from chromosome 9 to chromosome 22.				
Allelic Variants クリックする Back to				
(Selected Examples) Notes				
See allelic variants in tabular display				
.0001 LEUKEMIA, PHILADELPHIA CHROMOSOME-POSITIVE, RESISTANT TO IMATINIB [ABL1, THR315ILE]				
<u>Gorre et al. (2001)</u> reported a 944C-T transition in the region coding for the ATP- binding site of the ABL kinase domain of the BCR-ABL fusion product in 6 of 9 people with advanced-stage leukemia who were Ph-positive and imatinib-refractory. This mutation resulted in a thr315-to-ile (T315I) substitution. Thr315 forms an important hydrogen bond with the secondary amino group of imatinib (<u>Schindler et al., 2000</u>), suggesting that this mutation could be a major cause of resistance to imatinib.				

次の画面が表示される。1つの遺伝子の中に複数の変異の箇所があることがわかる。

Display Allelic Variants Show 20 Send to Send to							
All: 1	All: 1 OMIM UniSTS: 0 OMIM dbSNP: 0 🙊						
MIM *18	9980						
ABEL SO	MURINE LEUKEMIA VIRAL ONCOGENE HOMOLOG 1: ABI	1					
ABEEGG							
Allelic Var	iants (Selected Examples) <u>Notes</u>						
Number	Phenotype	Mutation	dbSNP				
<u>.0001</u>	LEUKEMIA, PHILADELPHIA CHROMOSOME-POSITIVE, RESISTANT TO IMATINIB	ABL1, THR315ILE	-				
<u>.0002</u>	LEUKEMIA, PHILADELPHIA CHROMOSOME-POSITIVE, RESISTANT TO IMATINIB	ABL1, GLU255VAL	-				
<u>.0003</u>	LEUKEMIA, PHILADELPHIA CHROMOSOME-POSITIVE, RESISTANT TO IMATINIB	ABL1, GLU255LYS	-				
<u>.0004</u>	LEUKEMIA, PHILADELPHIA CHROMOSOME-PO9ITIVE, RESISTANT TO IMATINIB	ABL1, TYR253HIS	-				
<u>.0005</u>	CHRONIC MYELOID LEUKEMIA, RESISTANT TO IMATINIB	ABL1, PHE311LEU	-				
<u>.0006</u>	CHRONIC MYELOID LEUKEMIA, RESISTANT TO IMATINIB	ABL1, MET351THR	-				

••• "THR315ILE"の表記は、遺伝子産物のタンパク質配列の 315 番目の THR (スレオニン)が ILE (イソロイシン)に変異していると、発病の可能性が高いことを示している。この突然変異を起 こしたアミノ酸配列情報を含むこの画面を Excel に記載しておく。

Number	Phenotype	Mutation	dbSNP
.0001	LEUKEMIA, PHILADELPHIA CHROMOSOME-POSITIVE, RESISTANT TO IMATINIB	ABL1, THR315ILE	•
.0002	LEUKEMIA, PHILADELPHIA CHROMOSOME-POSITIVE, RESISTANT TO IMATINIB	ABL1, GLU255VAL	8
.0003	LEUKEMIA, PHILADELPHIA CHROMOSOME-POSITIVE, RESISTANT TO IMATINIB	ABL1, GLU255LYS	
.0004	LEUKEMIA, PHILADELPHIA CHROMOSOME-POSITIVE, RESISTANT TO IMATINIB	ABL1, TYR253HIS	
.0005	CHRONIC MYELOID LEUKEMIA, RESISTANT TO IMATINIB	ABL1, PHE311LEU	
.0006	CHRONIC MYELOID LEUKEMIA, RESISTANT TO IMATINIB	ABL1, MET351THR	

↓ この枠内だけを、コピーし EXCEL に貼り付ける(課題 4)。

これまでの操作で、自分が着目した病気(この場合は白血病、LEUKEMIA)に関係す るタンパク質遺伝子(ABL1)について、そのタンパク質のどの位置(複数の位置の場合 も多い:ここでは、315 や 255、253 番目等のアミノ酸の位置)でどのようなアミノ酸 変化が起こると、その病気になりやすいのかが調べられた。個人のゲノム配列が解読 された際に、この Excel の表を参照すると、その人が罹り易い病気が推定できる。

今回の 遺伝子 ABL1 では、dbSNP(SNP に関するデータベース)へのリンク情報 が無かったが、調べる遺伝子によっては、dbSNP(SNP に関するデータベース)へ のリンク情報が記載されている場合がある。

例えば、白皮症に係る遺伝子(OCA2)について、このテキストの手順通りに調べると、次のような allelic variants as a separate report の画面が表示される。

SNP は "Single Nucleotide Polymorphism"の略で、日本語では 一塩基多型 と表現される。SNP について、このテキストでは説明を行わないが、各自 Google 等でキーワード検索して内容を調べ理解すること。

Number	Phenotype	Mutation	dbSNP
<u>.0001</u>	ALBINISM, OCULOCUTANEOUS, TYPE II	OCA2, 2.7-KB DEL, EX7DEL	- カリックす
<u>.0002</u>	ALBINISM, OCULOCUTANEOUS, TYPE II	OCA2, IVS17DS, G-T, +1	-
<u>.0003</u>	ALBINISM, OCULOCUTANEOUS, TYPE II	OCA2, ALA481THR	<u>rs74653330</u>
<u>.0004</u>	ALBINISM, OCULOCUTANEOUS, TYPE II	OCA2, VAL443ILE	<u>rs28934272</u>
<u>.0005</u>	ALBINISM, OCULOCUTANEOUS, TYPE II	OCA2, PRO743LEU	-
<u>.0006</u>	ALBINISM, OCULOCUTANEOUS, TYPE II	OCA2, 1-BP DEL	-
<u>.0007</u>	ALBINISM, OCULOCUTANEOUS, TYPE II	OCA2, ALA334VAL	-
<u>.0008</u>	ALBINISM, OCULOCUTANEOUS, TYPE II	OCA2, 122.5-KB DEL	
<u>.0009</u>	ALBINISM, OCULOCUTANEOUS, TYPE II	OCA2, TRP679CYS	主ての SNF 情報 在 R
<u>.0010</u>	ALBINISM, OCULOCUTANEOUS, TYPE II	OCA2, ASN489ASP	-
<u>.0011</u>	SKIN/HAIR/EYE PIGMENTATION 1, BLUE/NONBLUE EYES	OCA2, ARG305TRP	<u>rs1800401</u>
<u>.0012</u>	SKIN/HAIR/EYE PIGMENTATION 1, BLUE/NONBLUE EYES	OCA2, ARG419GLN	<u>rs1800407</u>
<u>.0013</u>	SKIN/HAIR/EYE PIGMENTATION 1, BLUE/NONBLUE EYES	OCA2, IVS1, HAPLOTYPE 1	rs4778138,4778241,7495174
<u>.0014</u>	ALBINISM, OCULOCUTANEOUS, TYPE II	OCA2, MET394ILE	-

課題5)病気と関係する SNP データの取得を以下の操作に従って行う。

dbSNP の各番号をクリックすると、NCBI ENTREZ の SNP データの画面が表示される。重要 なデータであるので、dbSNP のデータがある場合は、必ず SNP の塩基配列情報も収集し ておく。

SNC 2/ Jy/jto	
All Dat ibases PubMed Nucleotide Protein Genome Structure OMIN	
Search SNP v for (参考): ここをクリックするとゲノムプロジェクトで得られた	2
Limits Preview/Index History SNP 配列情報が表示される	
Display raphic Summary Show 20 Sort By Send to V	
All: 1 1000 Genomes: 1 Cited in PubMed: 0 Clinical/LSDB Submission: 1 Human 1 ついて、	
□1: rs74653330 [Homo sapiens] excel に貼り付ける	
ATTTGGAGGGTCCCCGATGGCAGTGG[C/T] AGCTCCTCCAATGTTTGTGAAGATC	
15 MapView VarView No PubMed GeneView SeqView No 30 OMIM	

<u>rs74653330</u>をクリックすると次の画面が表示される。この中の Fasta sequence の配列を コピーして Excel に貼り付ける。





ここで、公開されている、ワトソン博士、ベンダー博士のゲノムブラウザを紹介す る。(URL: <u>http://jimwatsonsequence.cshl.edu/cgi-perl/gbrowse/jwsequence/</u>)

Google で"James watson Genotype Viewer"をキーワードにして検索する。

James watson Genot	/pe Viewer	
約 1,830 件 (0.21 秒)	クリックする	
ヒント: <u>日本語のページだけを</u> 植	<u>:索 (孝 <u>赤設定</u>で検索対象言語を指定できます)</u>	
About the Genotype Vie	wer - [このページを訳す]	
24 Apr 2008 James (Jim)	Natson's personal genome sequence, showing his genot	ypes i
jimwatsonsequence.cshl.edu	/about.html - キャッシュ - 類似ページ	
Watson Genotype View	ver Now On Line - Cold Spring Harbor	

<u>About the Genotype Viewer</u> をクリックすると次の画面が表示される。

About the James Watson Genotype Viewer

On May 31, 2007, Nobel Laureate James Watson received his personal genome sequence in a ceremony at the Baylor College of Medicine. This genome sequence describes the six billion base pairs of DNA that James Watson received from his two parents, the unique combination of which are responsible for James Watson's genetic individuality. Dr. Watson is making his genome sequence available to the public in the hope that it will encourage the development of an era of "personalized medicine" when the information contained in our genomes is used to identify and prevent diseases to which we are genetically prone before they appear, and to create personalized medical therapies that have the maximum benefit and the minimum risk. This simple browser allows you to view the places where Watson's sequence is different from the "reference" human genome sequence, as well as to view the genes and some of the common diseases associated with them.

文章中の simple browser をクリックすると、次の画面が表示される。

James Watson's Personal Genome Sequence

README: How do I use the James Watson Genome Browser? Downloads: Download bulk JW polymorphisms. For the complete data set, please go to the NCBI Trace Archive and search for CENTER_NAME = 'CSHL' and CENTER_PROJECT = 'Project Jim'.



34.46 kbp の範囲を chr7 から表示、塩基番号 75,221,807 から 75,256,264

Ξ	説明
あフ	同小マ

□ ##### 超列の名前、遺伝子名、ゲノム上の位置、その他のランドマークなどを使って検索します。ワイルドカード文字として * を使うことができます。 ルーラー上でクリックした 位置が 中心になります。 スクロールとズームボタンを使って拡大率と位置を変更します。

ff: HTR2A, macular degeneration, rs726455, DAOA, chr22:20230140..20330139, PARK3, SNP:rs131693, SPTB, NM_001008496, 3q21.2, ENm010.

[パナーを厚す] [この表示をブックマーク] [この画像へのリンク] [高品質SVG画像] [ヘルブ] [リセット] □ 検索

ランドマークまたは領域:	出力や解析などの操作
chr7:75221807752562ť 検索	強調 SNP Properties
データソース	フクロ、サノマ、ノ、🔨 🛑 東子 34.46 kbn 👽 🕂 💙 🗆 日本年
James Watson genotypes, on NCBIB36 assembly, dbSNP b126 💌	
ロ <u>オーバービュー</u>	
chr7	
OM 10M 20M 30M 40M 50M 60M 70M	80M 90M 100M 110M 120M 130M 140M 150M
🗆 🖸 Ideogram	
□ III NT contigs	
OMIM disease associations	
EITI GWA studies (NHCEL Catalog)	
Citility a studies (mon Catalog)	
•• • •• • • •	•••• • •



公開されているワトソン博士のゲノム情報が閲覧できる。



例えば、James Watoson genotypes の rs1146532 のあたりにカーソルをあわせると、ワトソン博士 のこの SNP の genotype に関する情報が表示される。



OMIM disease associationsのrs2302009 あたりにカーソルをあわせると、esophagitis(食道炎) に関係する遺伝子のワトソン博士の genotype に関する情報が表示される。



HapMap Genotyped SNPs の rs1146532 のあたりにカーソルをあわせると、HapMap 計画(ヒトゲノム 計画の成果であるヒトゲノム全配列を基準として、各地域のヒト集団を対象として一塩基多型(SNP) の頻度からハプロタイプを決定し、比較を行うための標準データの作成を目指した国際計画)で得ら れた、各地域の集団ごとの SNP の genotype が表示される。この画面では、ヨーロッパ、北京(中国)、 東京、イバダン(ナイジェリア)の各地域のヒトの genotype の情報が表示されている。



検索の中で、

- データソースで、Grag Vendeter genotypes on・・・を選択すると、ベンダー博士のゲノム 情報の画面が表示される。
- ② ランドマークまたは領域に、表示させたい染色体番号、位置番号を入力すると、その部分が表示される。ここに遺伝子シンボル名を入力すると、その遺伝子周辺の画面が表示される。ただし表示されない遺伝子シンボル名もある。
- ③ スクロール/ズームで、表示サイズの変更や、表示部分を先に進める、戻すことが出来る。

課題 6) 一つの遺伝子の変異が色々な病気を引き起こす可能性のあることを調べる。

病気に関係をする遺伝子と言った場合、「病気を引き起こすための遺伝子」がヒト ゲノム中に存在するのではないことを、テキストの最初の部分で説明した。通常はヒト にとって必要な働きをしている遺伝子であるが、その遺伝子配列上に変異が起こるこ とで、その遺伝子の生産するタンパク質の機能が変化や欠損を起こし、病気を引き起 こす例を述べた。課題4で利用した OMIM で <u>See allelic variants in tabular display</u> をクリックして得られた表の Phenotype (表現型)の欄には、Mutation (変異)の 欄に記載されている変異が、着目した遺伝子に存在する場合に引き起こされる可能性 のある病気の英語名が記載されている。

Number	Phenotype	Mutation	dbSNP
<u>.0001</u>	LEUKEMIA, PHILADELPHIA CHROMOSOME-POSITIVE, RESISTANT TO IMATINIB	ABL1, THR315ILE	-
<u>.0002</u>	LEUKEMIA, PHILADELPHIA CHROMOSOME-POSITIVE, RESISTANT TO IMATINIB	ABL1, GLU255VAL	-
<u>.0003</u>	LEUKEMIA, PHILADELPHIA CHROMOSOME-POSITIVE, RESISTANT TO IMATINIB	ABL1, GLU255LYS	-
<u>.0004</u>	LEUKEMIA, PHILADELPHIA CHROMOSOME-POSITIVE, RESISTANT TO IMATINIB	ABL1, TYR253HIS	-
<u>.0005</u>	CHRONIC MYELOID LEUKEMIA, RESISTANT TO IMATINIB	ABL1, PHE311LEU	-
<u>.0006</u>	CHRONIC MYELOID LEUKEMIA, RESISTANT TO IMATINIB	ABL1, MET351THR	-

病気名以外の用語も存在するが、遺伝子機能や病気に係わる用語であり、それらも病気 を理解する上では有用な情報である。Phenotypeの欄の用語について、ライフサインス 辞書を利用して和訳しながら、病気名に対応する英語名を探し出し、最初に設定した病 気名と同じであるかを確認すること。最初に設定した病気名と異なっている例も多いはずで ある。理由は、着目している遺伝子にMutation(変異)が起きた場合、Mutationの入る場所 によって、あるいはそのMutationの内容により、異なった病気を引き起こす例が多いからで ある。ライフサインス辞書を利用して見つかった病気の日本語名を、課題4で作成した EXCELの各行について、dbSNPの後ろのライフサインス辞書の欄に記入せよ。

次に、Phenotype の欄で見つかった病気の英語名を利用して、新たに見つかった病気の内容を調査する。ここでは再度、メルクマニュアル医学百科を用いるが、今回は医師等の 医療従事者用のメルクマニュアル日本語版(http://merckmanual.jp/mmpej/index.html)を利用する。

Google で "メルクマニュアル医学百科"をキーワードにして検索する。 次の画面で、<u>メルクマニュアル18版日本語版</u>をクリックする。

Google	メルクマニュアル 医師向け 検索
U	約 49,100件(0.24秒) 検索オブション
<mark>∛]</mark> すべて ▼ もっと見る	メルクマニュアル18版 日本語版 メルクマニュアルは、医薬品の研究開発力で高い評価を得てMerck & Co., Inc., Whitehouse
ウェブ全体から検索 日本語のページを検索 ▼その他のツール	Station, N.J., U.S.A.とメルクマニュアル「Merck & Co., Inc., Whitehouse Station, N.J., U.S.A. が非営利事業として提供しているもので、世界で最も信頼されて 皮膚科疾患 - 神経疾患 - 消化器疾患 - 心血管疾患 merckmanual.jp/ - キャッシュ - 類似ページ



医療従事者用 MERCK MANUALS <u>http://merckmanual.jp/mmpej/index.html</u>

最初に利用したメルクマニュアル医学百科の家庭版とは異なり、この医療従事者用の版を用 いれば、英語の病名を検索欄に入れても検索が可能であり、検索結果にはその病気の正 確な和名が記載されている。その正確な病気の和名を、エクセルファイルの医療従事者用 版の欄に記載をせよ。その病気の和名をクリックし、その病気についての内容を読み、遺伝 子が関係することが記載してある際は、その記載事項をエクセルファイルの医療従事者用版 遺伝子の欄に記載せよ。なお、医療従事者用版の内容を理解することが難しいと感じた場 合には、病気の和名を用いて、メルクマニュアル医学百科家庭版を用いて検索をして、遺伝 子が関係することが記載してある際は、その記載事項をエクセルファイルの家庭版遺伝子の 欄に記載せよ。なお、メルクマニュアル医学百科家庭版を用いて検索をして、遺伝 子が関係することが記載してある際は、その記載事項をエクセルファイルの家庭版遺伝子の 欄に記載せよ。なお、メルクマニュアル医学百科で見つかった病気の日本語名が、ライフ サインス辞書を利用して見つかった病気の日本語名と異なる場合も考えられる。同じ病気 に異なった日本語名が使用される場合もあるが、メルクマニュアル医学百科の方が正式名と 考えられる。

このように OMIM の Phenotype の欄を参照することで、一つの遺伝子が多くの病気に関わっていることが理解できる。

課題 7) 今回の実習で行った操作は、これからの医学において重要になる技術である。これまでの課題を行うことで操作の概略が理解できたと思う。比較的早く 課題 6) までを終えた学生は、課題 3) で見つかる別の遺伝子について 課題 3) 以降のデ ータを収集し excel の別の行に追加すること。このことで今回の実習の全貌が理解で きる。 (参考)

	核酸塩基略号表
略号	塩基(略称の由来)
Α	アデニン (Adenine)
T	チミン (Tymine)
G	グアニン (Guanine)
С	シトシン (Cytosine)
U	ウラシル (Uracil)
R	プリン (puRine) A あるいは G
Y	ピリミジン (pYrimidine) T あるいは C
Μ	A あるいは C (aMino)
K	G あるいは T (Keto)
S	G あるいは C (G と C の結合は強い (Strong))
W	A あるいは T (A と T の結合は弱い (Weak))
В	G あるいは T あるいは C (A の次は B)
Н	A あるいは T あるいは C (G の次は H)
V	A あるいは G あるいは C (TU の次は V)
D	A あるいは G あるいは T (C の次は D)
Ν	AGTC のどれか (aNy)

※引用:Wikipedia 核酸より

アミノ酸の名称と略記法

和名	英名	3文字	1 文字	和名	英名	3文字	1文字
メチオニン	methionine	Met	М	セリン	serine	Ser	S
アラニン	alanine	Ara	Α	トレオニン	threonine	Thr	Т
バリン	valine	Val	V	チロシン	tyrosine	Tyr	Y
ロイシン	leucine	Leu	L	アスバラギン	asparagine	Asn	Ν
インロイシン	isoleucine	lle	I	グルタミン	glutamine	Gln	Q
プロリン	proline	Pro	Р	アスバラギン酸	aspartic acid	Asp	D
フェニルアラニン	phenylalanine	Phe	F	グルタミン酸	glutamic acid	Glu	E
トリプトファン	tryptophan	Trp	W	リシン	lysine	Lys	K
システイン	cysteine	Cys	С	アルギニン	arginine	Arg	R
グリシン	glycine	Gly	G	ヒスチジン	histidine	His	Н

(参考)

ヒト統合ボディーマップは、国立遺伝学研究所大久保研究室が、文部科学省 統合 データベースプロジェクトの成果の一つとして公開している。着目した遺伝子の発現 パターンが確認出来る。統合 TV に動画での操作説明があるので参考にするとよい。

Google で、"ヒト統合ボディーマップ"をキーワードにして探索をかける。



次のトップ画面が表示される。(http://okubolab.genes.nig.ac.jp/bodymap_i/)

画面右上のボックスに、例えば白血病に係わる遺伝子 "ABL1"を入力して、検索ボタンをクリックする。



次の画面が表示される。

100		部科学省データベース統合こ 、統合ボディーマップ	プロジェクト ^{ABI}	L1		検索		
\mathbb{P}_{2}	ップ 発現バ	ターンから探す 染色体領域で眺める 遺伝子言	ファミリーで眺める 機能で眺める	る データ取得 データ処	理の説明]		
3	遺伝子 存在	します。						
		クリックす	3		絶対発 (ビーク)	現量 ウウント数	/100万)	相対発現量 ■脳 ▲ ■結
#	遺伝子ID ▲ ▼	道伝子名	 	染色体上の位置 ▲ ▼	EST	NCBI 解釈	大久保研 解釈 ▲ ▼	
1	Hs.431048	V-abl Abelson murine leukemia viral oncogene homolog 1	_	9q34.1 (+)	167.1	207.3	133.5	未設計
2	Hs.11108	Cdk5 and AbI enzyme substrate 1	Cdk5 and c-AbI linker prote cables	ein 18q11.2(+)	90.5	207.3	133.5	
3	Hs.301040	Cdk5 and AbI enzyme substrate 2	Cdk5 and c-AbI linker prote cables	^{;in} 20q13.33(-)	55.3	49.8	46.8	

トップ 発現バタ	ーンから:	探す 染色	白体領域で眺	跡る 遺(法子ファミリ	ノーで眺める	機能で	眺める う	データ取得	データ処理の	兑明
■ 遺伝子情報											
	〒子 ID カス ID 伝子名 伝子名	Hs.4310 25 V-abl A ABL, <mark>A1</mark> c-ABL	48 belson murir <mark>3L1</mark> , Abelsor p150 v-abl	ne leukemi n murine le	ia viral ono eukemia vi	cogene homol iral oncogene	og 1 homolo	g 1, JTK7	, Proto-onc	ogene tyrosine	-protein kinase <mark>ABL1</mark> ,
זעע	一構成	総数 3' EST 5' EST EST(Oth HTC mRNA BetSen	: 597 : 128 : 291 her): 151 : 2 : 22 : 22								
iAFLPクラスタ	刻ング	101004	• -								
染色	体位置	<u>9q34.1</u> 1	32,579,089~	132,752,8	83(+)						
		chr9 🛛									
■ 相対発現量 遺伝子の相対	, 発現量 [.]	です。									
 クラフ・ グラフ・ エース、 2002 エース、 2002 <	の高さ の色は 測定が 張色 は 願のグラ	は、相対 、臓器・綿 失敗してい 、臓器・綿 フの高さ	発現量を表しま 狙織を表しま 11名ことをま 狙織が測定 は便宜上44	します。 をす (内容 むします (されてい) 0分類の4	容は組織: GAFLPの。 ないことで グラフの高	分類を参照) み)。 を表します。 話さをカテゴ!	。 ノー(色)ごとに刊	² 均した高る	きです。	
		組	織10分類	組	目織40分業 224日(₫ 879012	2156	7100	10275	1212241	112122456
	iAF	LP 未	設計	*	20400 設計	070812	040C	120	12040	121204	112 <mark>123430</mark>
	GeneC	hip									
	E	sт 📕									_
SAGE N	ICBI 駕	深 🖕									
SAGE 大久	保研餐	RR .									
■ 組織分類											
解剖学による	蔵器・組	織の分類	表です。								
│ 組織分類 │ 10~40 1		2	3	4	5	6	7	8 9	9		
1 5 7	大脳	小脳	脳幹	脳梁	松果体	末梢神経	脊柱	網膜	T and the second	ドの	あり (1) (神) ア
	助脈	靜脈	リンバ節	末梢血	脾臓	胸腺	骨髄				
結	脂肪	骨	皮膚	50.224	1 = 24					相对日	りに発現重か
● 22 月 ● 23 月 ● 25 月	倍盛 心時	十呂 丹杦族	則立脉	丱果	栢朱					多いズ	いが わかる。
	L'mex 食道	月11日期	賜	結腸						ä	
肝見	-~~= 肝臓	~	499	400099							
肺	市										
腎腸	膀胱 てチ/ナ	腎臓 田44mg	百川田久	時些	회 66	0.新公本8 白					
<u> </u>	1"	甲氨脲	副背	胖瓶	孔脉	"要/仪服					

(参考)

余力のある人は、慶應大学(清水信義ブループが作成)が公開している Mutation View(ミューテーションビュー)を参照することも、病気に関するデータベースへの 理解を深める。なお、清水信義先生は長浜バイオ大の特別招聘教授です。

Google で "mutation view" をキーワードにして検索する。

Google	mutation view 検索
	約 9,320,000 件 (0.19秒) クリック ヒント: <u>日本語のページだけを検索</u> (表示設定で検索対象言語を指定できます)
ウェブを検索 日本語のページを検索 ▼その他のツール	KMDB/MutationView Top - [このページを訳す] KMDB/MutationView (Keio Mutation Databases) Version 1.2. An Integrated Database for Human Disease-associated Using this, it is possible to construct a Locus-Specific DataBase with similar functions with MutationView mutview.dmb.med.keio.ac.jp/ - キャッシュ - 類似ページ
	Japanese Society for Bioinformatics - Mutation View : A Network Mutation View : A Network-Transparent Database System for Mutations in the



MutationView						
This system has been developed by the Department of Molecular Biology Keio University School of Medicine in collaboration with Chi Co.,Ltd.						
Condit	tion to use <i>MutationView</i> and <i>MV4LSDB</i>					
2. It is prohibited to 3. It is not allowed to	arofit by re-dy Java Version をクリック					
Java Version	Java Applet (JDK1.1) Version Netscape Communicator Ver40 or later (except Macintosh) Internet Explorer Ver40 or later					
HTML Version	HTML Version Netscape Communicator Ver3 Internet Explorer Ver3					
Curators	Curating Members					
Online Manual	Online Manual (Java Version)					



OMIM に比べると、登録されている疾患に係わる遺伝子数は少ないかもしれないが、優れた CG 技術を用いて身体の臓器別や病名別に病気に関係する遺伝子が表示されるので有用である。mutation view から疾患に係わる遺伝子を探索して、再び OMIM でその遺伝子の変異情報を取得してもよい。

(その他の参考サイト)

● Ensembl チュートリアル http://www-bird.jst.go.jp/minicourses/Ensembl1.html

2010年6月1日初版 2010年9月3日改訂 長浜バイオ大学 バイオサイエンス学部 コンピュータバイオサイエンス学科 生体分子情報学研究室(池村研究室)編集 URL: <u>http://www.nagahama-i-bio.ac.jp/</u>

※本テキストの内容等についてのお問合わせは、

生体分子情報学研究室 上原 h_uehara@nagahama-i-bio.ac.jp 宛に願います。