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

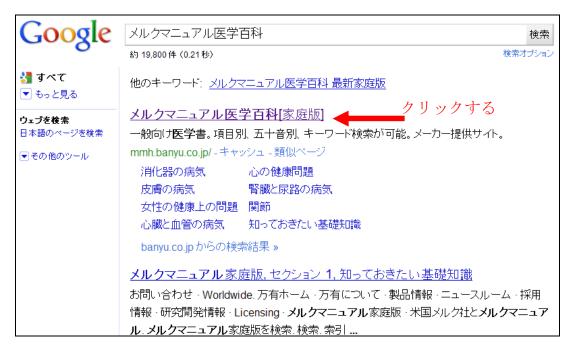
ノム中の病気と関係した遺伝子配列を収集する。次世代シーケンサーは、我が国でも百台近くが導入済みか導入予定であるが、その次世代シーケンサーが産出する大量なゲノム配列の情報解析をする人材が極端に不足している。この我が国の状況を考え、生命情報の3回生の後期の実習では、次世代シーケンサーが産出する大量配列の情報解析を行っている。

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。

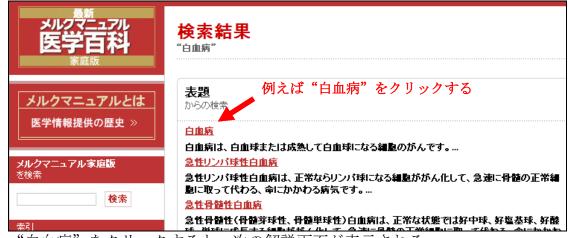
I. まず、各自で自分の関心のある病気を考えてほしい。ここでは個々人のゲノムの配列に関係する疾患を考えているので、ウイルスや細菌が引き起こす感染症、(例えばインフルエンザや食中毒等)は対象から外す必要がある。各自で関心のある病気について、正確な知識を得ておくためにメルクマニュアル医学百科(Google で検索せよ。http://mmh.banyu.co.jp/mmhe2j/index.html)を参照しておこう。勿論、メルクマニュアル医学百科を見てから、関心のある病気を決めて良い。ガンのように広い範囲の疾患を選択すると以降の操作が難しくなるので、肺ガンや白血病のように、より限定した疾患名を選択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

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。

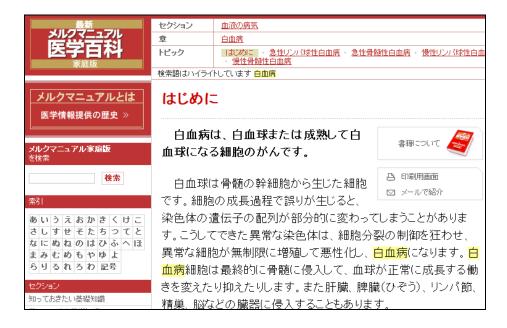




メルクマニュアル医学百科 トップ画面(http://mmh.banyu.co.jp/mmhe2j/index.html) 興味のある病気、例えば"白血病"とキーワードを入力して、検索ボタンをクリックすると、 次の検索結果画面が表示される。



"白血病"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。



次の画面で、"WebLSD オンライン辞書"をクリックする。



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。



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。

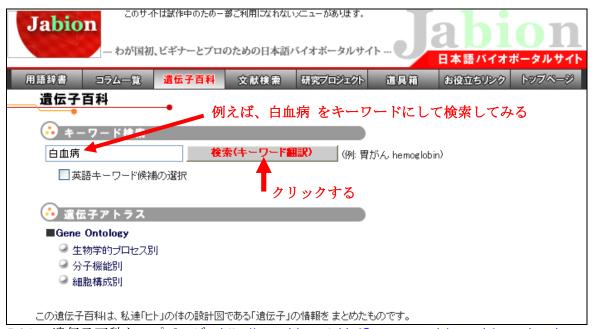
- (1) cDNA 配列
- (2) CDS (アミノ酸翻訳部分) の塩基配列
- (3) 遺伝子産物のタンパク質配列
- (4) ゲノム **DNA** 配列

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。



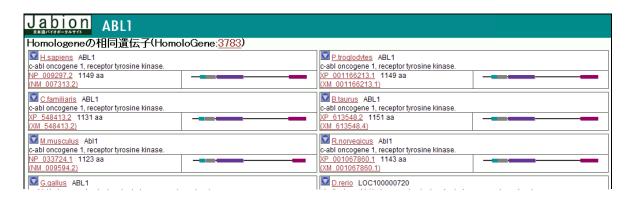
次の画面が表示される。



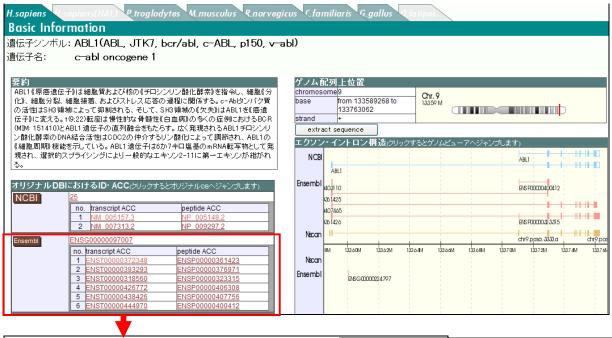
Jabion 遺伝子百科トップページ <a href="http://www.bioportal.jp/Gene_search/search

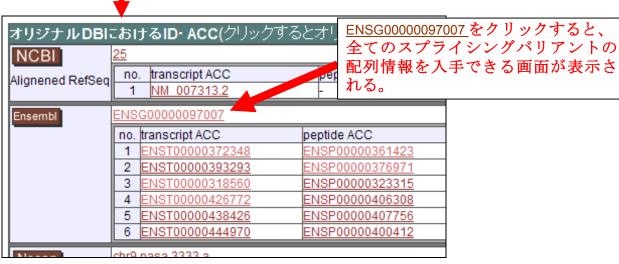


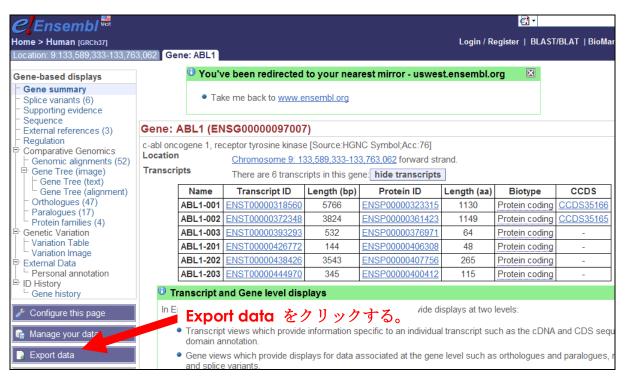
"ABL1"をクリックすると次の画面が表示される。



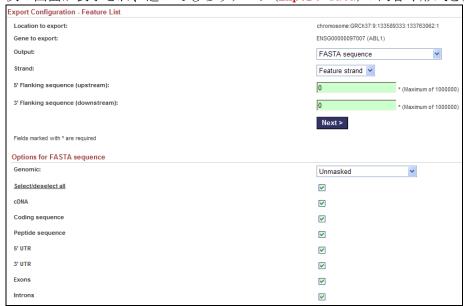
画面を下にスクロールしていくと、次の画面が現れる。要約として、着目している遺伝子の機能 や特徴等が日本語で記載されているので、この部分を読むこと。

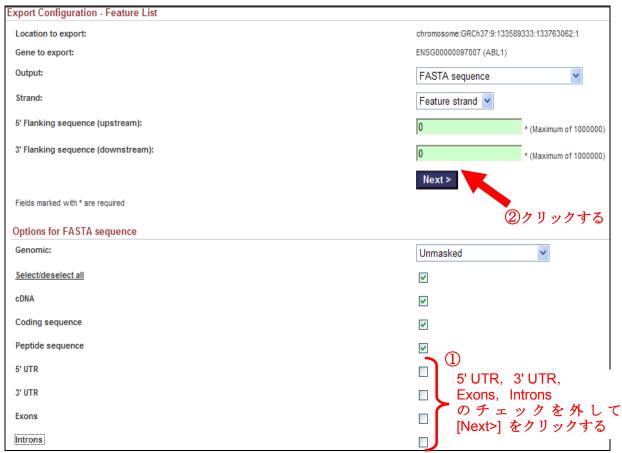






次の画面が表示され、送ってもらうデータ(Export data)の内容や形式を指定できる。





この画面で、Options for FASTA sequence の設定で、

5' UTR, 3' UTR, Exons, Introns の ☑ チェックを外して [Next>]をクリックする。



次の画面が表示される。

```
>ENSG00000097007:ENST00000318560 cdna:KNOWN protein coding
GGCCTTCCCCCTGCGAGGATCGCCATTGGCCCGG T™GCC™™™CGAAAcccCCCCCCCCCCCC
TTGGGCCGGGCTCGGGAACGCCAGGGGC CDNA 配列
                                          cDNA 配列をコピーして excel に貼り付ける。
コメント行(">"で始まる行)も含めてコピー
   ( 中
>ENSG00000097007:ENST00000318560 cds:KNOWN protein coding
ATGTTGGAGATCTGCCTGAAGCTGGTGGGCTGCAAATCCAAGAAGGGGCTGTCCTCGTCC
TCCAGCTGTTATCTGGAAGAAGCCCTTCAGCGGCC A CDS 配列
GGTCTGAGTGAAGCCGCTCGTTGGAACTCCAAGGAA
                                          コピーして excel に貼り付ける
                                          コメント行(">"で始まる行)も含めてコピー
AATGACCCCAACCTTTTCGTTGCACTGTATGATTTT
\Delta CCD \Delta \Delta CCT \Delta \Delta \Delta CCT CCCCCT CTT \Delta \Delta \Delta CCT CCCCCCCT CTT \Delta \Delta
   (中略)
>ENSG00000097007:ENST00000318560 peptide: ENSP00000323315 pep:KNOWN protein coding
MLEICLKLVGCKSKKGLSSSSSCYLEEALQRPVASD POCT SEAADWNSKENT LACDSE
NDPNLFVALYDFVASGDNTLSITKGEKLRVLGYNHNOW peptide 遺伝子産物のタンパク質配列
SLEKHSWYHGPVSRNAAEYLLSSGINGSFLVRESESSPG コメント行 (">" で始まる行) も含めてコピー
```

>ENSG00000097007:ENST00000372348 peptide: ENSP00000361423 pep:KNOWN_protein_coding MGQQPGKVLGDQRRPSLPALHFIKGAGKKESSRHGGPHCNVFVEHEALQRPVASDFEPQG LSEAARWNSKENLLAGPSENDPNLFVALYDFVASGDNTLSITKGEKLRVLGYNHNGEWCE AQTKNGQGWVPSNYITPVNSLEKHSWYHGPVSRNAAEYLLSSGINGSFLVRESESSPGQR

```
(以下 ABL1の場合には、同じパターンで、
```

DOMESTIC OF THE PROPERTY OF TH

(中略)

```
ABL1-003 Transcript ID <u>ENST00000393293</u> の cdna, cds, peptide の配列 ABL1-201 Transcript ID <u>ENST00000426772</u> の cdna, cds, peptide の配列 ABL1-202 Transcript ID <u>ENST00000438426</u> の cdna, cds, peptide の配列 ABL1-203 Transcript ID ENST00000444970 の cdna, cds, peptide の配列
```

が表示され、最後に 全ゲノム DNA 配列 が表示される。

```
>9 dna:chromosome chromosome:GRCh37:9:133589333:133763062:1
TGTGAGATTGAAATGACTAGCATTATTGACCCTTTTCAGCATCCCCTGTGAATATTTCT
GTTTAGGT TTTCTTCTTGAAAAGAAATTGTTATTCAGCCCGTTTAAAACAAATCAAGAA
ACTTTTGGG ゲノム DNA 配列 ABL1 の例
           ">"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
GCTTGCCTGCCTGCATTTTATCAAAGGAGCAGGGAAGAAGGAATCATCGAGGCATGGGG
GTCCACACTGCAATGTTTTTTGTGGAACATGGTGAGTGCTTTTCAAAATTTCTGCTCATGG
   ( 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 ENST00000426772 の cdna, cds, peptide の各配列・ABL1-201 Transcript ID ENST00000438426 の cdna, cds, peptide の各配列・ABL1-203 Transcript ID ENST00000444970 の cdna, cds, peptide の各配列・ABL1-203 Transcript ID ENST000000444970 の cdna, cds, peptide の各配列・ABL1-203 Transcript
```

· 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。

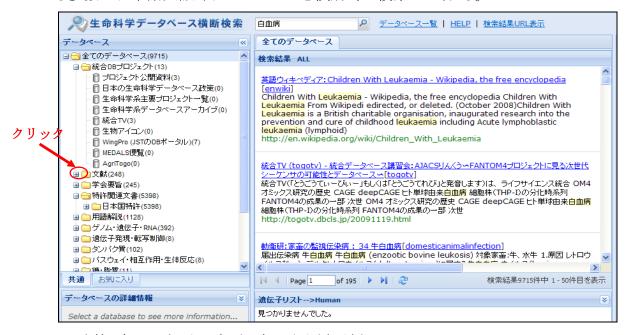
Google で "LSDB" をキーワードにして検索する。



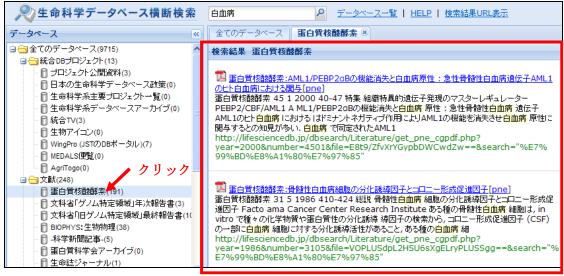
ライフサイエンス統合データベースプロジェクト-統合ホームページ-LSDBクリックすると次の画面が表示される。



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、Ensembl 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。Ensembl 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。

(参考) Ensembl を使った配列データの取得

Google で "ensembl" をキーワードにして検索する。

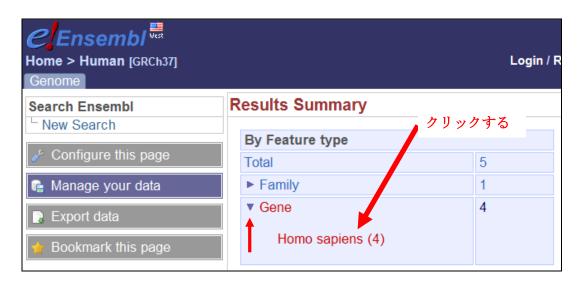


Ensembl をクリックすると、次のトップ画面が表示される。

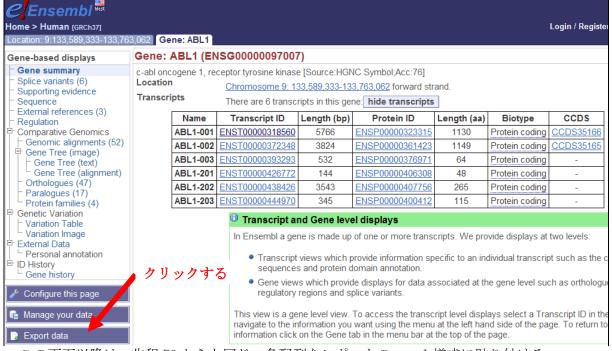




Ensembl 検索フォーム





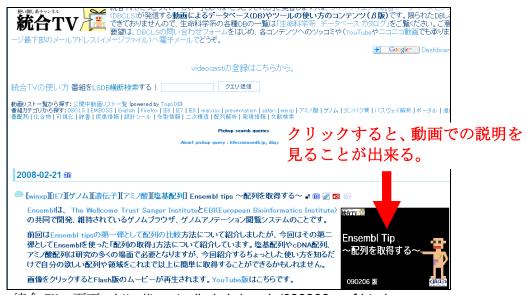


この画面以降は、先程 P9 からと同じ。各配列をレポートの excel 様式に貼り付ける。

(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。





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。

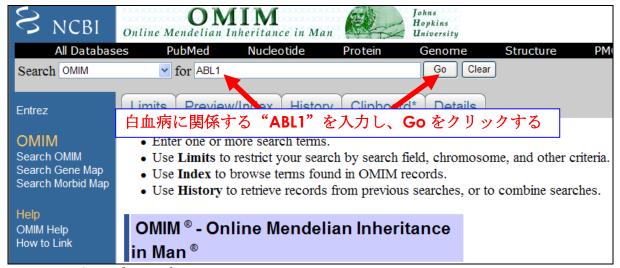
この場合には、OMIM (Online Mendelian Inheritance in Man; ヒトの疾患データベース)を使用する。遺伝子が関係する病気に関する情報や、各病気を引き起こす遺伝子の突然変異を収録している、世界最大規模で最も信頼性の高いデータベースである。

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、OMIM Home をクリックする。

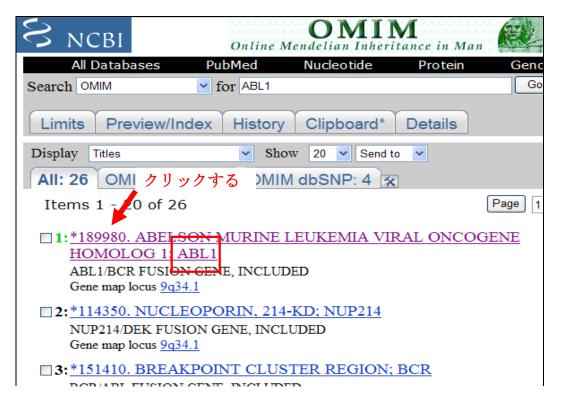




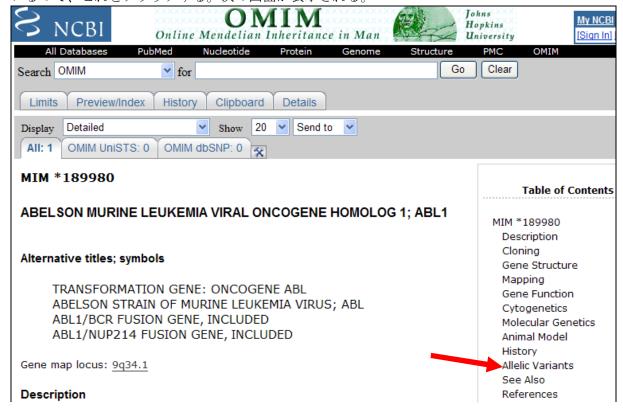
OMIM トップページ http://www.ncbi.nlm.nih.gov/omim/

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。

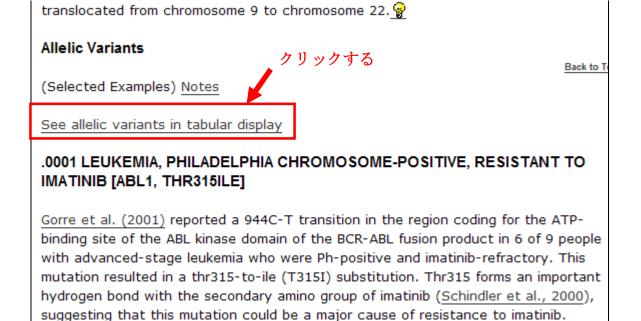


検索結果の各番号の次に表示されている記号について、*, + は遺伝子に関するデータであることを示し、#, % は疾患に関するデータであることを示している。本自習では、遺伝子に関するデータに着目する。この場合、*印のついた ABL1 に関する遺伝子データが一番目にヒットしているので、これをクリックする。次の画面が表示される。



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。**



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						
All: 1 OMIM UniSTS: 0 OMIM dbSNP: 0						
MIM *189980						
ABELSON MURINE LEUKEMIA VIRAL ONCOGENE HOMOLOG 1; ABL1						
Allelic Variants (Selected Examples) Notes						
Number	Phenotype	Mutation	dbSNP			
.0001	LEUKEMIA, PHILADELPHIA CHROMOSOME-POSITIVE, RESISTANT TO IMATINIB	ABL1, THR315ILE	-			
<u>.0002</u>	LEUKEMIA, PHILADELPHIA CHROMOSOME-POSITIVE, RESISTANT TO IMATINIB	ABL1, GLU255VAL	-			
.0003	LEUKEMIA, PHILADELPHIA CHROMOSOME-POSITIVE, RESIGNANT TO IMATINIB	ABL1, GLU255LYS	-			
<u>.0004</u>	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	-			

"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	dbsnf
.0001	LEUKEMIA, PHILADELPHIA CHROMOSOME-POSITIVE, RESISTANT TO IMATINIB	ABL1, THR315ILE	•
.0002	LEUKEMIA, PHILADELPHIA CHROMOSOME-POSITIVE, RESISTANT TO IMATINIB	ABL1, GLU255VAL	=
.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	E

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。

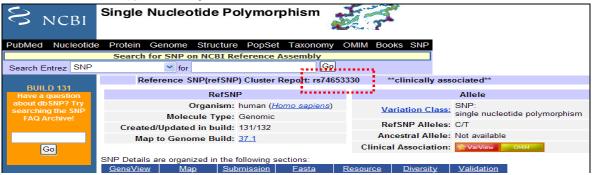
課題5)病気と関係するSNPデータの取得を以下の操作に従って行う。

Number	Phenotype	Mutation	dbsnp
.0001	ALBINISM, OCULOCUTANEOUS, TYPE II	OCA2, 2.7-KB DEL, EX7DEL	- クリックす
.0002	ALBINISM, OCULOCUTANEOUS, TYPE II	OCA2, IVS17DS, G-T, +1	- /
.0003	ALBINISM, OCULOCUTANEOUS, TYPE II	OCA2, ALA481THR	rs74653330
.0004	ALBINISM, OCULOCUTANEOUS, TYPE II	OCA2, VAL443ILE	rs28934272
.0005	ALBINISM, OCULOCUTANEOUS, TYPE II	OCA2, PRO743LEU	-
.0006	ALBINISM, OCULOCUTANEOUS, TYPE II	OCA2, 1-BP DEL	-
.0007	ALBINISM, OCULOCUTANEOUS, TYPE II	OCA2, ALA334VAL	-
.0008	ALBINISM, OCULOCUTANEOUS, TYPE II	OCA2, 122.5-KB DEL	- 全での SNP 情報を収
.0009	ALBINISM, OCULOCUTANEOUS, TYPE II	OCA2, TRP679CYS	すること すること
.0010	ALBINISM, OCULOCUTANEOUS, TYPE II	OCA2, ASN489ASP	-
.0011	SKIN/HAIR/EYE PIGMENTATION 1, BLUE/NONBLUE EYES	OCA2, ARG305TRP	rs1800401
.0012	SKIN/HAIR/EYE PIGMENTATION 1, BLUE/NONBLUE EYES	OCA2, ARG419GLN	rs1800407
.0013	SKIN/HAIR/EYE PIGMENTATION 1, BLUE/NONBLUE EYES	OCA2, IVS1, HAPLOTYPE 1	rs4778138,4778241,7495174
.0014	ALBINISM, OCULOCUTANEOUS, TYPE II	OCA2, MET394ILE	-

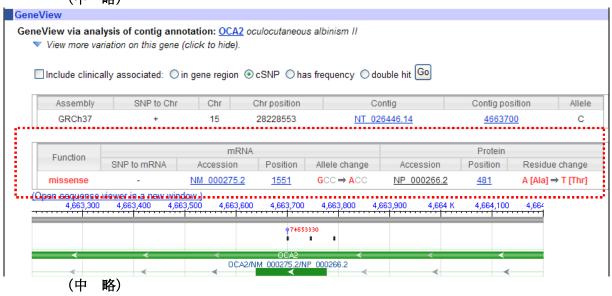
dbSNP の各番号をクリックすると、NCBI ENTREZ の SNP データの画面が表示される。重要なデータであるので、dbSNP のデータがある場合は、必ず SNP の塩基配列情報も収集しておく。

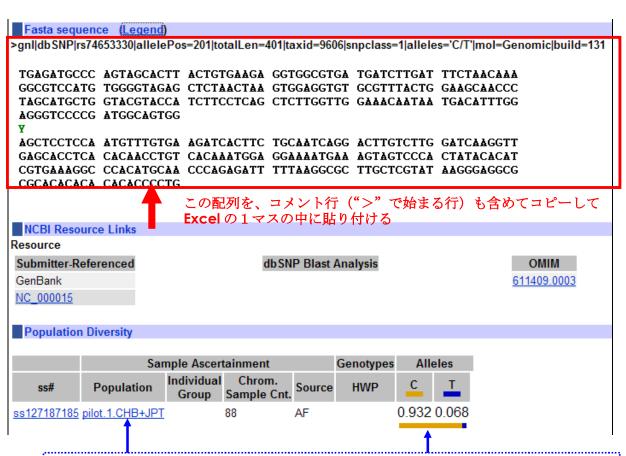


<u>rs74653330</u> をクリックすると次の画面が表示される。この中の Fasta sequence の配列をコピーして Excel に貼り付ける。



(中略)





統計をとった北東アジア人集団 (CHB+JPT) では、93.2%が C, 6.8%が T であったことを示している。 CHB (Chinese in Beijing 北京), JPT (Japanese in Tokyo)

数は少ないが、選んだ SNP によっては、Population の列にワトソン博士 (ENSEMBL Watson) やベンダー博士 (ENSEMBL Venter) の SNP 塩基配 列情報が記載されている場合がある。調べた病気に係わる遺伝子の突然変異体がワトソン博士やベンダー博士の塩基配列上に存在するのかを確認することも出来る。

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: http://jimwatsonsequence.cshl.edu/cgi-perl/gbrowse/jwsequence/)

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

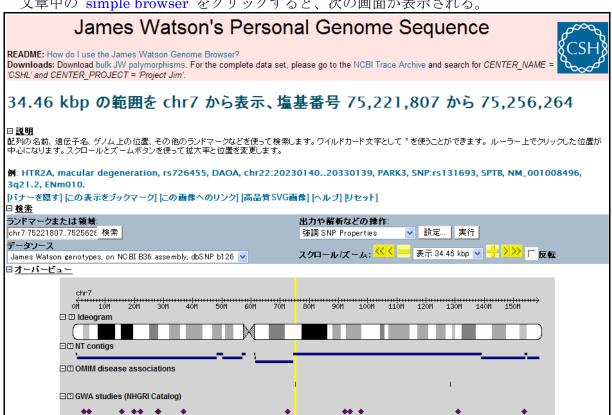


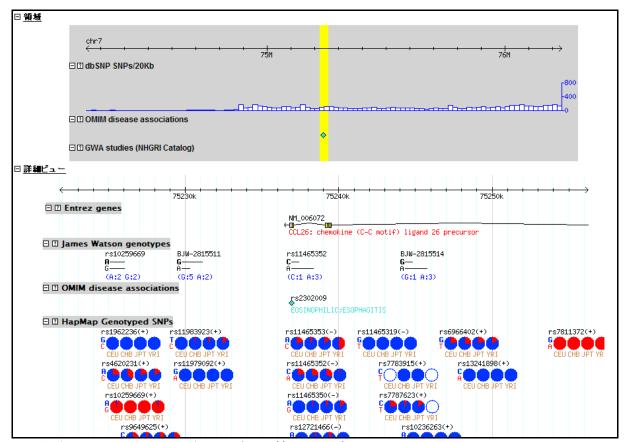
<u>About the **Genotype Viewer**</u> をクリックすると次の画面が表示される。

About the James Watson Genotype Viewer

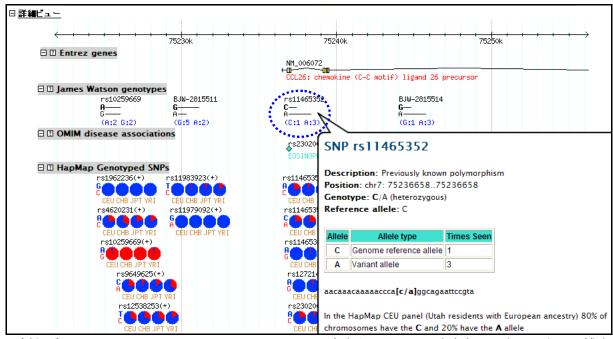
On May 31, 2007, Nobel Laureate James Watson received his personal genome sequence in a ceremony at the Baylor College of Medicine. This genome sequence describes the six billion base pairs of DNA that James Watson received from his two parents, the unique combination of which are responsible for James 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 personalizes 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。

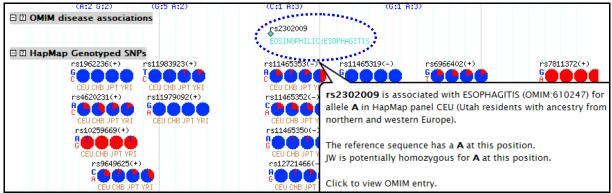




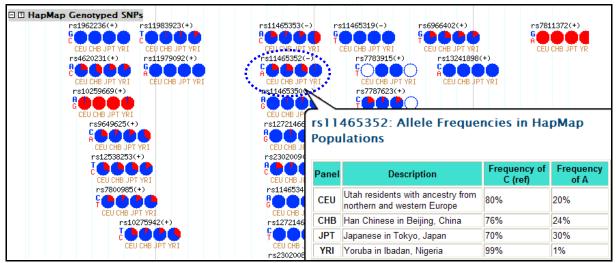
公開されているワトソン博士のゲノム情報が閲覧できる。



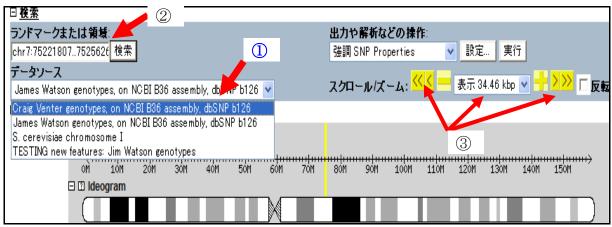
例えば、James Watoson genotypes の rs1146532 のあたりにカーソルをあわせると、ワトソン博士のこの SNP の genotype に関する情報が表示される。



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、

- ① 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 See allelic variants in tabular display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
.0001	LEUKEMIA, PHILADELPHIA CHROMOSOME-POSITIVE, RESISTANT TO IMATINIB	ABL1, THR315ILE	-
.0002	LEUKEMIA, PHILADELPHIA CHROMOSOME-POSITIVE, RESISTANT TO IMATINIB	ABL1, GLU255VAL	-
.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	-

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、メルクマニュアル 18 版 日本語版をクリックする。





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

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
M	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)
N	AGTC のどれか (aNy)

※引用: Wikipedia 核酸より

アミノ酸の名称と略記法

The state of the s							
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	Υ
ロイシン	leucine	Leu	L	アスパラギン	asparagine	Asn	N
イソロイシン	isoleucine	He	- 1	グルタミン	glutamine	Gln	Q
ブロリン	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	Е
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)

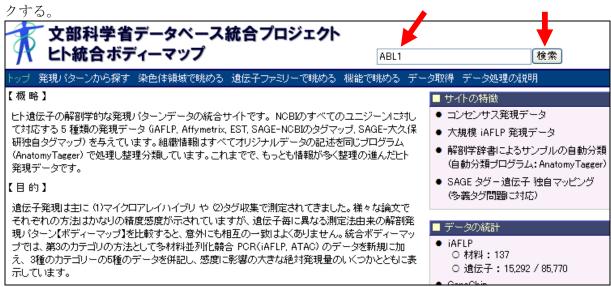
ヒト統合ボディーマップは、国立遺伝学研究所大久保研究室が、文部科学省 統合 データベースプロジェクトの成果の一つとして公開している。着目した遺伝子の発現 パターンが確認出来る。統合 TV に動画での操作説明があるので参考にするとよい。

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。



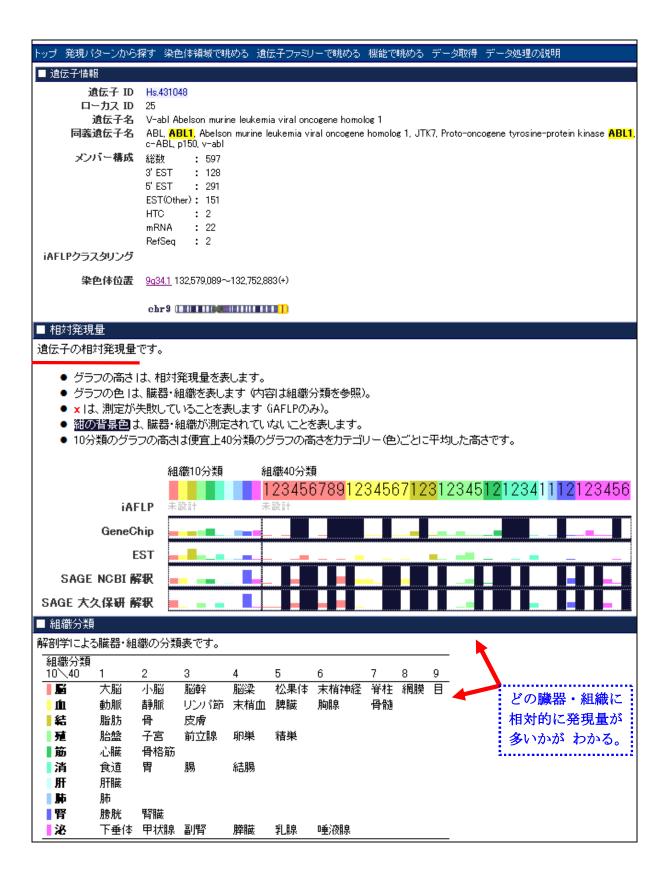
次のトップ画面が表示される。(http://okubolab.genes.nig.ac.jp/bodymap_i/)

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。



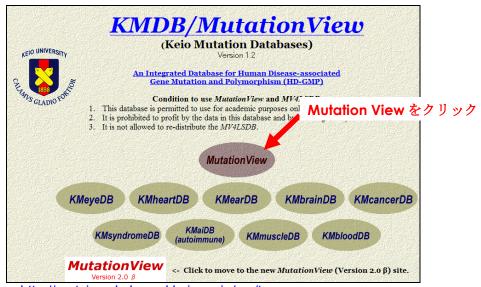


(参考)

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。

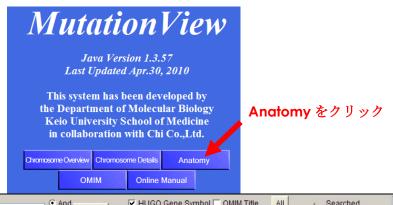
Google で "mutation view" をキーワードにして検索する。

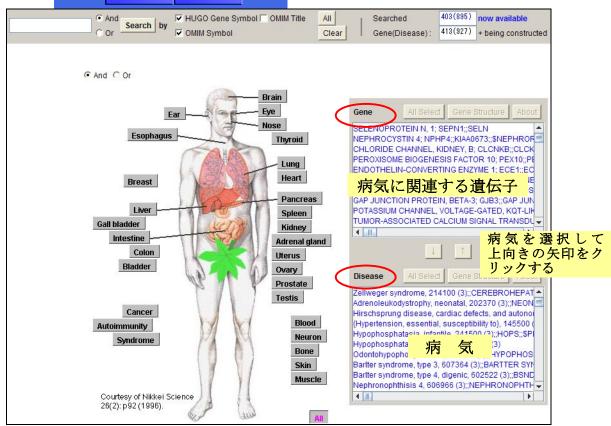




http://mutview.dmb.med.keio.ac.jp/mv/top







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: http://www.nagahama-i-bio.ac.jp/

※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。