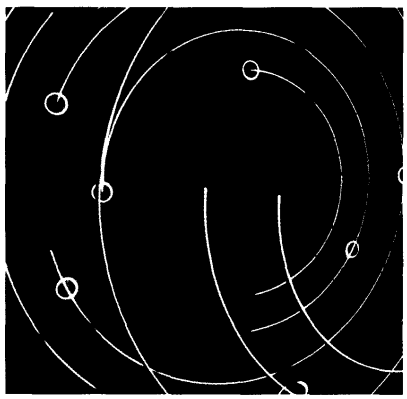
［特集］ポストゲノム時代に高まるバイオ自然言語処理への期待：バイオ自然言語処理最新事情



2 バイオ研究者からのバイオNLPへの期待

a）遺伝子変異データベース構築のための情報収集と抽出の現状

大坪 正史
ohtsubo@dmb.med.keio.ac.jp / 慶應義塾大学医学部分子生物学教室

清水 信義
shimizu@dmb.med.keio.ac.jp / 慶應義塾大学医学部分子生物学教室

蓑島 伸生
mino@hama-med.ac.jp / 浜松医科大学光量子医学研究センター・慶應義塾大学医学部分子生物学教室

遺伝子の変動性（疾患原因変異と多型）と表現型についてのデータベース*MutationView*を構築してきた．現在，ゲノム塩基配列読み取りが完了して，さまざまな公的データベースが構築されて資料を提供していることもあり，コンピュータによる遺伝子構造データの自動作成は可能である．一方，疾患と変異のかかわりをデータベース化するためには，変異の詳細以外にも，症状などの患者情報等多岐にわたる情報をデータとして構築することが必要となる．現状では，これらの情報は，主に発表済み論文から収集，抽出しているが，使用専門用語の冗長性や階層性などの問題に対処する必要性から，もっぱら研究者の手作業によって進めている．本稿では情報処理，計算機科学の専門家による支援の可能性を少しでも開くために，収集しているデータの種類，性質，情報元などを整理し，「手作業」の内容も分析してみたい．

ゲノムの変動性—表現型データベース構築の必要性

2003年4月14日「ヒトゲノム塩基配列読み取り」の完了が宣言され，ゲノム研究の重点は塩基配列の意味を読み解くことにシフトした．その中で重要なポイントの1つは，ゲノムの塩基配列の個人差（変動性：variation）とそれによる表現型（形に現れる性質）の差異を追究することである．ゲノムの変動性と表現型の極値の1つとしての「疾患」について論ずるには，遺伝子とその機能の基本概念の基盤に基づく必要があるので，以下に用語の説明を兼ねて若干の解説を加える（**図-1**）．

遺伝子はゲノムDNA上の特定の座位（Locus：染色体上の位置）に存在する塩基（A，T，G，C）の特異的な配列である．各遺伝子は，各々特異的なアミノ酸配列からなるタンパク質を作り出すための情報を，アミノ酸1個に対して塩基3個がセット（トリプレットコードあるいは単にコード）で暗号になり，その暗号が必要な数だけ連続した塩基配列として保持している．ただし，そのアミノ酸配列のための塩基配列情報は，通常連続的にはなっておらず，アミノ酸配列情報を持たないイントロンという部分でいくつかの部分に遮られている．遺伝子の機能の具現化（発現）のためには機能分子であるタンパク質にその情報が受け渡される必要がある．そのために，遺伝子DNAが持つ塩基配列情報は，転写（Transcription）と呼ばれるステップで，まずメッ

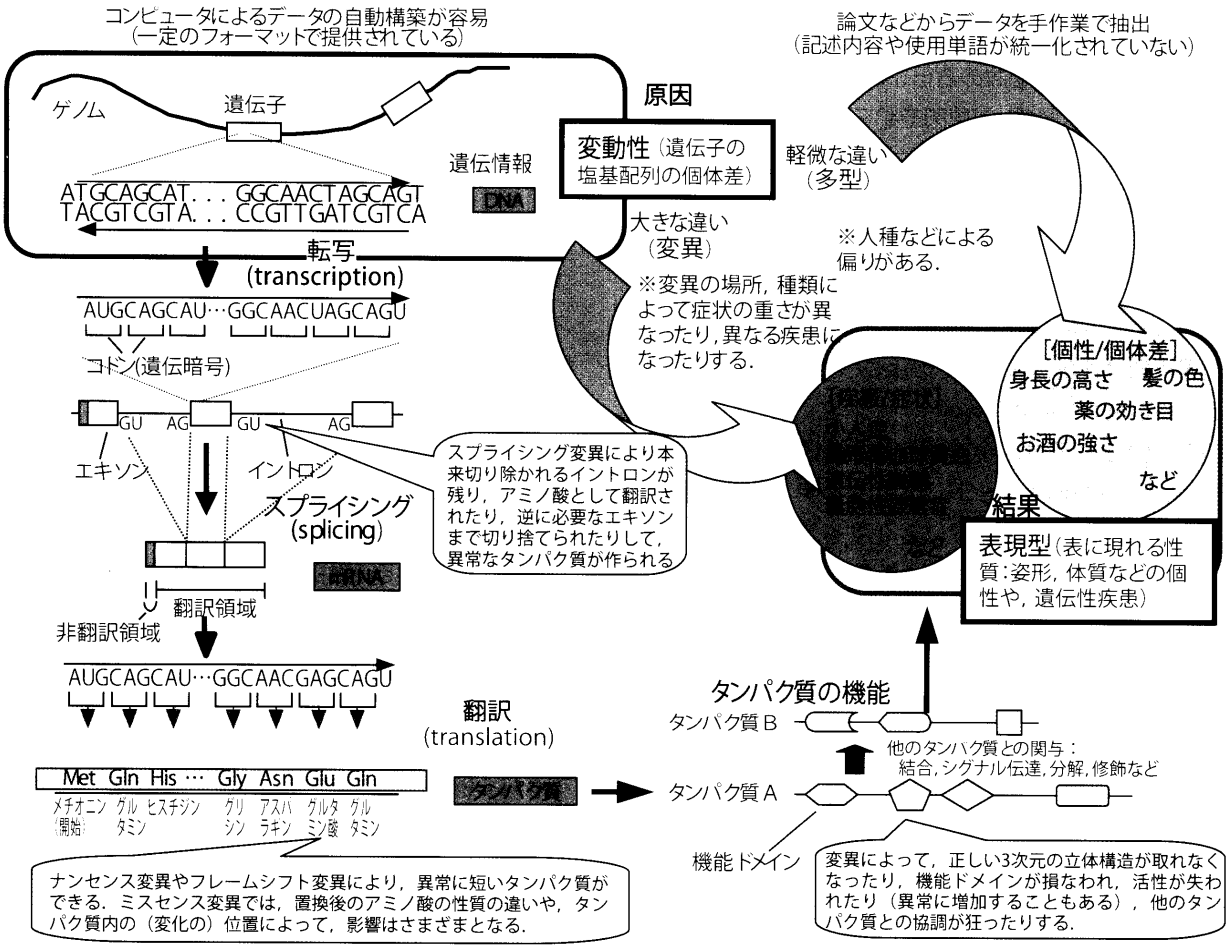


図-1　ゲノムの変動性と表現型：遺伝子の構造，発現，タンパク機能の体現と変異の影響

＜変異の種類＞

- ナンセンス変異：アミノ酸が停止コドン（Stop）に変わる変異．タンパク質は異常な短いものとなる．
- ミスセンス変異：アミノ酸が他のアミノ酸に変わる変異．変化後のアミノ酸の性質の違いや，タンパク質内の（変異アミノ酸の）位置によって，影響はさまざまとなる．
- フレームシフト変異：塩基の（3の倍数以外の）挿入あるいは欠失．遺伝暗号がズレてしまうことで，変異場所より以降は，異常なアミノ酸配列となる．早めにStopコドンが現れて，短いタンパク質となることも多い．3の倍数の塩基の挿入あるいは欠失の場合は，余分なアミノ酸が追加あるいは欠失する．この場合は，タンパク質内の変異の場所によって影響は異なる．
- スプライシング変異：エキソンの端，またはイントロンにあるスプライシングの信号となる配列の変化．イントロンが正常に切り出されないことで，余分なアミノ酸が追加されたり，異常に短いタンパク質が作られることがある．しばしば，正常なスプライシングも同時に起こることもあり，正常タンパク質の量の減少として影響する場合もある．

センジャーRNA（mRNA）に写し取られる．転写においてmRNAは前駆体として形成され，それは通常，エキソン（完成型mRNAに最終的に残る部分）に加え，上述のイントロン（エキソンとエキソンの間にあって転写後切って除かれる部分）に対応する配列を含んでいる．イントロンに対応する部分はスプライシングと呼ぶ反応により取り除かれて，完成型のmRNAとなる．mRNA上で「ひと続き」になったタンパク質のアミノ酸配列情報はリボソーム（RNAとタンパク質からなる巨大複合体でタンパク質生合成機能を担う）で翻訳（Translation）されて，各遺伝子に特異的なタンパク質が作られる．タンパク質コーディング領域（アミノ酸配列の情報を持つ部分．翻訳領域とも呼ぶ）の前後には翻訳の開始や終了のコードもあり，それ以外にも，エキソンとイントロンの切れ目を示す信号や，その遺伝子の発現のタイミングや発現されるべき生体組織，器官などを規定する信号等，遺伝子発現の調節を担うさまざまな塩基配列もイントロン内や翻訳領域上流などの近傍に存在し，それらも合わせて遺伝子の機能が実現される．

ゲノムの変動性は，遺伝子DNAの塩基配列およびコピー数に関するあらゆる変化と定義される．ゲノムの変動は，表現型に与える影響によって「多型」と「疾患原因変異（あるいは単に変異）」に分類される．すなわち，結果として個体の表現型がいわゆる疾患の状態に陥る場合には「疾患原因変異」と呼ばれ，表現型に変化があっても疾患ではない場合は「多型」と呼ばれる．遺伝病は

ほとんどの場合親から遺伝した先天的に持つ変異で，多くの癌は後天的に起きた複数の変異で起こる．一方，生活習慣病は，おそらく複数の遺伝子の変異や多型の組合せと環境要因（生活習慣）が原因で起こると考えられる．一方，薬の効き目や副作用の個人差のように，特殊な条件でのみ表現型に差が現れる多型もある．本稿では，主に「疾患原因変異」に関して述べる．

ゲノムの変動性と表現型の関連を記述したデータベースは，他のゲノム情報のデータベースに比較して未発達で，世界標準が存在しない．そのようなデータベースで扱われるべき情報を考えてみる．

変異の内訳としては，小規模な塩基配列の変化（アミノ酸のコードが停止コードに変化するナンセンス変異，異なるアミノ酸のコードに変化するミスセンス変異，塩基が1個または数個失われるデリーション欠失等）のほか，エキソンや遺伝子全体が失われたり，異なる染色体が切れてつなぎ換わったり，染色体が完全に1本余分になったりするような大規模な変化も含まれる．これらのさまざまな原因により遺伝子（ひいてはタンパク質の機能）が異常をきたした結果として，疾患に至るので，変異をゲノムレベルで正確に記述する必要がある．また，変異と表現型の間をつなぐ情報として，作られるタンパク質の機能を担う共通アミノ酸配列（ドメイン）と変異によるその変化も重要である．さらに，各変異を持っていた症例の詳細な症状が変異の「結果」として貴重な情報となる．一方，遺伝子疾患と原因遺伝子は1対1で対応することは少ない．たとえば，これまでは（臨床診断上の所見，症状の違いなどから）別の疾患として分類されていた複数の疾患（たとえば，致死性骨異形成症や軟骨無形成症および軟骨低形成症）が，同一の遺伝子（FGFR3）の異なる部位（機能ドメイン）の変異に起因する，つまり，変異による機能の損なわれ方によって表現型（疾患の症状）が異なるという現象だったというようなことも明らかになってきている．また，一連の機能を分担する複数のタンパク質の遺伝子（たとえばCOMP, COL9A1, COL9A2, COL9A3, MATN3）は，その塩基変化により同系統の疾患（これらの遺伝子の場合は多発性骨端異形成症）になる場合があることが分かっている．また，疾患の原因となる遺伝子変異および体質などの基となる多型の分布は人種によって異なることも知られている．さらには，1つの遺伝子だけを見ても，身体の部位（臓器／器官）によって発現状態は異なり，成長の段階のある時期にだけ発現する遺伝子も少なくない．このように，遺伝子変異と疾患をデータベースとして記述し，そのデータベースを有効なものとしていくには，必要となるデータは多岐にわたる．そのような現状で，我々は世界標準を目指した「自前の」ゲノム変動性—表現型データベースを構築している．本稿ではその現状と，データ収集に関する問題点，困難さに関して紹介する．

遺伝子疾患変異データベース *MutationView* の概要

筆者らは，疾患遺伝子情報に関する統合データベース化の重要性を強く感じ，数年前からその第1世代のシステムとして，主に単一遺伝子疾患の原因遺伝子と変異データに関するデータベースシステム*MutationView*（http://mutview.dmb.med.keio.ac.jp）を開発してきた[1]．

*MutationView*は，原因遺伝子に関する疾患の変異データをグラフィカルな環境で検索，表示，解析できるシステムである（**図-2**）．個々の遺伝子がコードするタンパク質の機能ドメインの表示や，変異に伴う諸情報（患者の民族，疾患症状など）を用いた変異分類機能など，さまざまな機能を有している．さまざまな情報が疾患の変異情報とともに有機的に結びついて，ビジュアルに表示されるので，データを種々の観点から容易に再評価できる．これまでに眼科疾患，神経疾患，筋肉疾患，家族性腫瘍，聴覚疾患，心臓疾患，骨系統疾患，自己免疫疾患を中心に，259遺伝子，407疾患に関する情報を収集している．

本稿では，*MutationView*データ構築のための情報収集と抽出の現状について述べる．

データ構成と抽出用資料

*MutationView*では，データを以下の5種類に分け，遺伝子記号などで相互に関連付けている．データ抽出に用いている資料を｛｝内に示す．

(A) 染色体情報：染色体バンド情報，遺伝子および疾患リスト（遺伝子名あるいは病名，座位，遺伝子記号，発症組織器官名）{OMIM genemap, morbidmapファイル[2]}

(B) 画像データ（染色体バンド，人体図，臓器図など）{ISCN95[3]}

(C) 遺伝子構造データ：領域種別（エキソン，イントロンなど），塩基配列情報，領域別塩基数 {DDBJ/EMBL/Genbank[4]}

(D) 変異基本情報：染色体変異情報（転座など染色体異常），変異種別（ナンセンス，ミスセンス，デリーションなど），変異位置，塩基変異情報，変異名，結果 {HGMD[5] および論文}

(E) 拡張情報：頻度／症例数，疾患名，発症年齢，人種，遺伝形質／症状，文献等 {論文およびOMIMの

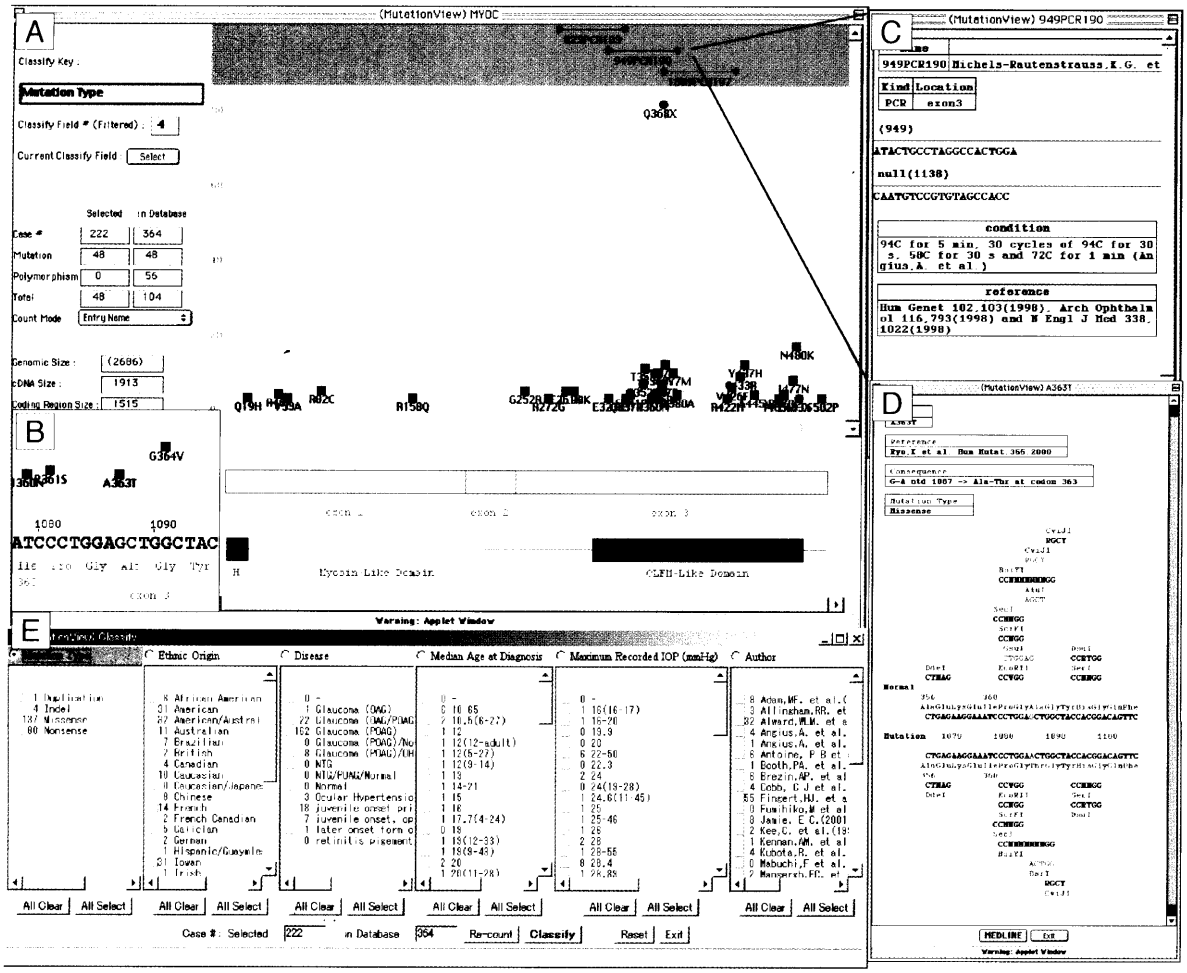


図 -2 *MutationView* の代表的表示画面

A. ミオシリン（MYOC）遺伝子の遺伝子構造ウィンドウ：開放隅角緑内障（POAG）の原因遺伝子 MYOC．現在，40 以上の遺伝子変異が同定されており，変異の発生頻度に人種差があると考えられている．画面では，下部に（横軸として）遺伝子のエキソン構造とタンパク質の機能ドメインが表示されている．縦軸は変異ごとの症例数を表すヒストグラムである．

B. 最大拡大による塩基表示：遺伝子構造ウィンドウの横軸は拡大できる．最大拡大では塩基配列まで表示される．

C. プライマー情報ウィンドウ：遺伝子構造ウィンドウの上部に表示されるプライマーをクリックすると表示される．エキソンあるいは変異を含む領域を PCR で増幅させるためのプライマー配列情報および反応条件などが表示される．

D. A363T の変異詳細ウィンドウ：ミスセンス変異 A363T の変異詳細ウィンドウ．Normal における C が変異により A に変化し，その結果アミノ酸配列が Ala（アラニン）から Thr（トレオニン）に置換する．正常配列では存在した *AluI* および *CviJI* 制限酵素切断部位が突然変異により消失し，新たに *BarI* 制限酵素認識部位が生じたことが分かる．

E. クラシファイウィンドウ：変異の種類，人種，疾患名，診断時年齢，最大 IOP（眼圧）値，文献（著者）などの項目について，統計処理されて表示されている．機能の詳細は本文を参照．

Clinical Synopsis[2]}

以上の項目のうち，拡張情報はクラシファイ機能（後述）に用いられる．

各項目に関するデータベース構築時のデータ収集方法については，以下に詳細を述べる．

■塩基配列情報およびプローブ情報（前項Cに対応）

遺伝子構造（エキソン／イントロン構造，翻訳領域）は，主として塩基配列に関する公的データベース（GenBank など）から取得した mRNA 配列およびゲノム配列に対して，配列比較（ペアワイズアラインメント）アプリケーション SIM4[6] を用いて抽出している．SIM4 はイントロンとエキソンの境界に注目したアラインメントを行うことができる．該当するゲノム配列情報がない場合には，相同性検索プログラム BLAST[7] を用いて，GenBank などに収集されている DNA 配列に対して検索を行い，該当するゲノム配列を得ている．

PCR は，微量の DNA の特定部分と同じ塩基配列を持つ DNA を多量に増幅することができる方法で，増幅したい部分の両端の配列（約 20 塩基ずつ）と同じ配列を持つ DNA（プライマーと呼ぶ）があれば容易に実行できる．PCR は目的遺伝子上の変異を含む部分などを容易に増

幅することができ，検査などに有用であるので，資材となるPCRプライマーなどの情報も，文献（PDFやhtmlファイル）から抽出する．多くの場合，遺伝子の塩基配列上の位置が明示されておらず，またゲノム配列上の塩基番号での記載しかないものもあるため，前述のようにして抽出したエキソン／イントロンとの相対位置情報を，検索して得ている．

■アミノ酸配列情報およびタンパク質ドメイン情報（前項Cに対応）

画面表示上のアミノ酸配列への変換は，遺伝子情報（前項の塩基配列情報およびエキソン／イントロン情報）からアプレットプログラムが自動で行う．

タンパク質機能ドメインは変異とタンパク機能，さらには疾患症状との関連を考察する上で重要な情報である．*MutationView*では，文献にて記載されているドメイン情報を取り込むことを基本とするが，その情報が得られない場合には，アミノ酸配列情報を基に，ドメイン検索データベース（SMART[8]やPfam[9]など）を利用して，既知ドメインを検索し，推定している．これらのデータベースは，これまで研究され解析が進められてきたドメインに関する情報（配列アラインメントやその特徴，ドメインが担う機能を記述したドキュメントなど）を保有している．

■遺伝子ー疾患対応情報（前項Aに対応）

OMIM（Online MIM）は，ジョンズ・ホプキンス大学のVictor A. McKusick教授が，1966年以来執筆を続けている，遺伝病を含めたヒトの遺伝形質に関する医学的，分子遺伝学的記述を集めた著述 Mendelian Inheritance in Man（MIM）のオンライン版である．*MutationView*では，疾患と原因遺伝子の対応情報に関して，OMIM Gene MapやMorbid Mapファイルから得ている．また，OMIMは変異に関する全症例報告を含んではいないため，頻度／症例数情報は得られないが，疾患情報に関しては非常に詳細な記述が多く，疾患研究の歴史，個々の症例の説明など多岐にわたる情報を含んでいる．特に，個々の疾患に関する病態の詳細は，“Clinical Synopsis”という症状の項目に別項として列記されているので，変異情報を作成する場合にも参照している．

■変異基本情報（前項Dに対応）

*MutationView*では，遺伝子の変異の塩基配列変化の部分は，変異基本情報として疾患あるいは症状情報と区別している．変異の種類に関しては主として，英国Wales大学で作成しているヒト遺伝子変異データベースHGMD（Human Gene Mutation Database）から変換して自動生成し，参考データとして用いている．HGMDのデータはすべて，基本的には文字情報に徹している．変異については，変異位置や塩基／アミノ酸置換情報，関連疾患名および報告文献など最低限の情報ではあるが，入力遺伝子数1657，変異データ数42521について，表形式でカタログ化されており，充実している．しかし，前述のOMIMと同じく，変異の頻度／症例数，疾患症状などは収集されていない．

■変異拡張情報（前項Eに対応）

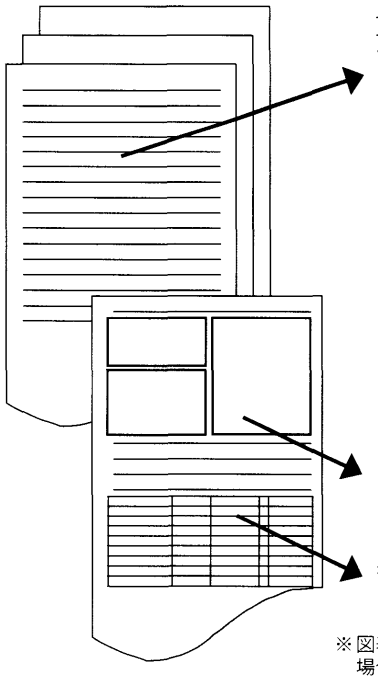
前項までの4項目に関しては，比較的データベース化も，*MutationView*への取り込みも容易である．疾患情報は，前述のOMIMでも，“Allelic Variants”項目として，遺伝子の変異の記載が，論文抜粋として書いてあり，ここで，家族性／孤発例の別，発症年齢，症状，人種などのデータを得ることができる場合もあるが，従来のシステムでは知識ベース化が困難なデータである．*MutationView*データは頻度情報を持たせているため，PubMedを使って当該遺伝子の変異情報に関する全文献を検索し，人種・症状・発症時期などの症例情報（変異拡張情報）を研究者が読んで手作業で収集している（**図-3**）．

臨床症状の抽出に関しては，文献ごとに症状記載に用いる用語が必ずしも一定しておらず，また，症状も，ほとんど記述がないものから，詳細に記述してあるものまで千差万別であり，情報抽出の自動化は困難を極める．しかしながら，*MutationView*の特徴の1つである，データ分類機能（クラシファイ機能）は，この変異拡張情報を用いることで，症状や人種による変異の傾向や偏りを明示することができる機能であり，各種データ上でも最も重要で特徴的なものである．現在は，この部分のデータ構築は手作業に依存しており，（自動化が困難なことから）データ作成の律速段階となっている．

■疾患情報の自動抽出の試み

現在，自然言語処理によるテキストからのデータ抽出や，臨床情報の標準化の必要性が問われている．MeSHや統一医学用語システムUMLS（Unified Medical Language System），ICD10（国際疾病分類第10版）あるいは医療情報交換のための標準規約HL7など整備が始められている．

我々も独自にこれらの疾患情報を扱うための支援機能の充実を行っている．前述のOMIMの，その多岐にわたる疾患情報に関する詳細な記述や，個々の症例の説明およびClinical Synopsis項目を解析し，検索に役立てるツールを開発した．OMIMの同一パラグラフ内での



文献からの情報抽出
・項目：
遺伝形質（遺伝子変異）の情報に加えて，人種，発症年齢，症状，飲酒や喫煙などの環境情報　……などの患者の情報
・問題点：
症状などは規格化された用語ではなく，さまざまな言い方が使われる．たとえば，小頭症：MicrocephalyとOccipitofrontal Head Circumstance，聴覚障害：DeafnessとHearing LossあるいはHearing Impairmentなどである．
これらの言い換えによる問題は，遺伝子名やタンパク質名についても存在する．
また，それ以外にも，発育遅延（Developmental Delay）という一語で説明している論文と，Age at SittingやAge at First Wordsなど詳細な項目で述べている論文など．症状を記述するのに用いられる専門用語間の階層的理解が必要となる．
同様の問題は，人種などでも見られる．たとえば，イタリア人と白人（コーカシアン）などの階層（内包）関係など．

図からの情報抽出
・診療情報
・タンパク質の細胞内局在
・家系図（父母兄弟など親類の発症歴）……など

表からの情報抽出
・臨床情報　……など

※図表で示されたデータの中には，他の文献からの引用されたものが含まれる場合もあるので，データのだぶり（冗長性の解消）が問題となる．

図-3　文献からの情報抽出の困難な点と，文章以外のデータソース

任意単語対の共起情報を用いて重要な単語と単語間関係情報を抜き出し，それらの重要単語間の相互関連を図示することで内容の理解を強力に支援するとともに，関連語まで含めた高度な検索を可能にするツールである（**図-4** A～C）．また，用語が標準化されていないテキストに対応するために，重要単語の抽出を，辞書に依存せずに，OMIMの記述の中の各単語の“分散”を利用して実行できる新規アルゴリズムを開発している．

MutationViewの主な機能

▪ システム概要

分散データベースとしてWebサイト上の同一形式データにアクセス可能であり，染色体模式図（イディオグラム）上に遺伝子疾患を，また人体図の各臓器に対して疾患／遺伝子名を一覧でき，それらを検索の入口にできる．さらに，遺伝子／疾患／組織名を項目とした検索機能を持つ．

■ 表示機能

指定した遺伝子のDNA構造／タンパク質のドメイン構造に対する変異の種類と位置，および変異部位近傍の制限酵素切断部位の変化，PCRプライマー情報を表示できる（図-2 A～D）．

■ クラシファイ機能

遺伝子構造ウィンドウ（図-2 E）の縦軸には，クラシファイ機能を用いてさまざまな情報を分類表示することができるので，たとえば，人種による変異の偏り，疾患症状と変異の相関などを検討するのに有用である．

クラシファイウィンドウでは，個々の突然変異に付随して収集した拡張情報は分類項目ごと（変異のタイプ，報告症例数，優性／劣性，疾患の種類，発症年齢，症状など）に区分され，それぞれ具体的な記事（種別）ごとの症例数が表示される．*MutationView*は，これらの1つあるいは2つ以上を用いたデータ選択・統計分類表示機能（クラシファイ機能）を有しており，拡張情報を利用した分類表示のほかに，データのグループ化（複数のデータ種別を統合する機能），フィルタリング（特定の群の変異を表示から除外する機能）で構成され，ユーザの必要に応じて統計的に表示を加工することができる．

情報処理分野の専門家に望むこと

症例や変異の種類の多い疾患は，主要なものだけでも100編以上の英語論文を読み，そこから症状，変異情報，論文間の患者の重複や異同などを読み取る必要がある．常に最新の情報を追尾するためにはさらに大きな努力を要する．重要な情報は，本文だけでなく，図表にまとめられていることも多いので，本文に加えて図表からの自動情報取得も強く望まれる（図-3）．我々は，手作

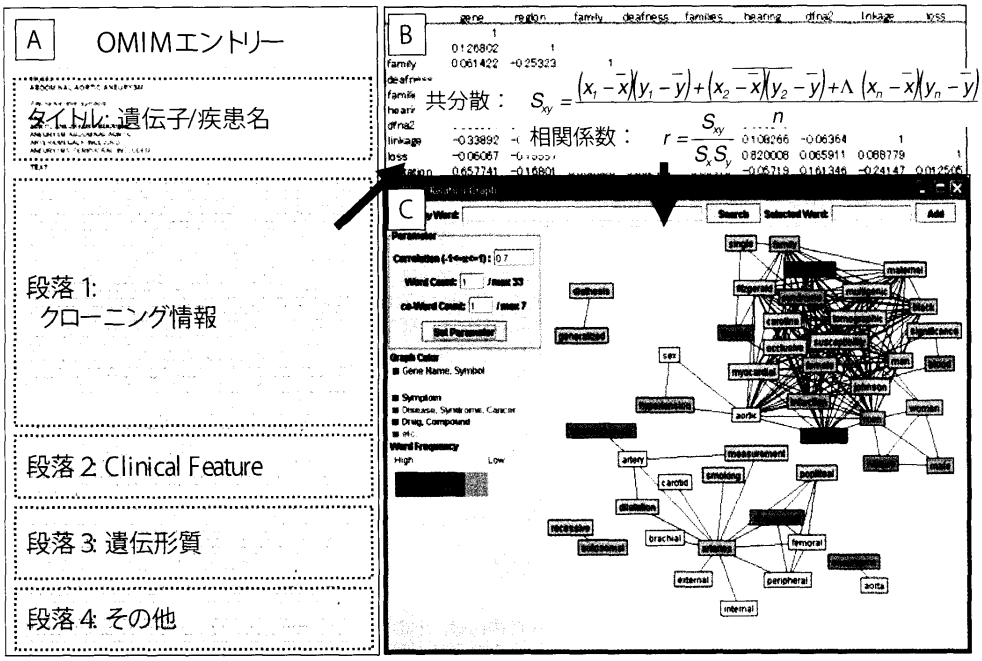


図 -4　疾患情報の自動抽出の概念図

A.　OMIM エントリーからの共起による単語抽出：OMIM の各エントリーは，複数の項目（段落）から構成される．原則として 1 つの段落は 1 つのテーマに関して記述している．よって，同じ段落内で使用されている単語同士は，1 つのテーマを説明するために用いられており，「関係がある」ことが予測される．段落ごとの共起によって，重要な単語間の関係を抽出する．

B.　相関係数による統計処理：段落の長短などで標準化して，単語同士の相関係数を得る．

C.　単語間の関係情報の図示：キーワードにより検索された OMIM エントリーに含まれる重要単語（と単語間の関係）は，グラフィカルに表示され，ユーザに（1 単語の検索という手段を通じて）関連ある単語群を提示できる．ユーザは表示された単語群から，より検索に用いたい（有効である）単語を選択して，再帰的に検索を続けることにより，スムーズに望むエントリー（＝知識）に至ることができる．

業による努力を継続していくが，情報処理の専門家には自動化への期待をお伝えして結語に代えたい．

謝辞　*MutationView* は，筆者らの教室の満山進氏，河村隆氏ほか，多くの教室員の協力を得て作成された．また *MutationView* のコンピュータプログラムは（株）カイとの共同開発である．上記すべての関係各位に感謝します．

また，本稿で紹介した研究の一部は，文部科学省科学研究費補助金特定領域研究「ゲノム情報科学」および「ゲノム医科学」，さらに研究成果公開促進費「データベース」，日本学術振興会未来開拓学術研究推進事業のサポートにより行った．

参考文献

1) Minoshima, S., Mitsuyama, S., Ohtsubo, M., Kawamura, T., Ito, S., Shibamoto, S., Ito, F. and Shimizu, N.: The *KMDB/MutationView*: A Mutation Database for Human Disease Genes, Nucleic Acids Res. 29, pp.327-328 (2001).
2) Hamosh, A., Scott, AF., Amberger, J., Bocchini, C., Valle, D. and McKusick, VA.: Online Mendelian Inheritance in Man (OMIM), A Knowledgebase of Human Genes and Genetic Disorders, Nucleic Acids Res. 30, pp.52-55 (2002).
3) ISCN 1995, An International System for Human Cytogenetic Nomenclature (1995), Ed. F. Mitelman, Karger Publishers, Inc. (1995).
4) Benson, DA., Karsch-Mizrachi, I., Lipman, DJ., Ostell, J. and Wheeler, DL.: GenBank: Update, Nucleic Acids Res. 32, D23-26 (2004).
5) Stenson, PD., Ball, EV., Mort, M., Phillips, AD., Shiel, JA., Thomas, NS., Abeysinghe, S., Krawczak, M. and Cooper, DN.: Human Gene Mutation Database (HGMD): 2003 Update, Hum Mutat. 21(6), pp.577-581 (2003).
6) Florea, L., Hartzell, G., Zhang, Z., Rubin, GM. and Miller, W.: A Computer Program for Aligning a cDNA Sequence with a Genomic DNA Sequence, Genome Res. 8(9), pp.967-974 (1998).
7) McGinnis, S. and Madden, TL.: BLAST: at the Core of a Powerful and Diverse Set of Sequence Analysis Tools, Nucleic Acids Res. 32, w20-25 (2004).
8) Letunic, I., Copley, RR., Schmidt, S., Ciccarelli, FD., Doerks, T., Schultz, J., Ponting, CP. and Bork, P.: SMART 4.0: Towards Genomic Data Integration, Nucleic Acids Res. 32(1), D142-144 (2004).
9) Bateman, A., Coin, L., Durbin, R., Finn, RD., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, ELL., Studholme, DJ., Yeats, C. and Eddy, SR.: The Pfam Protein Families Database, Nucleic Acids Res. 32(1), D138-141 (2004).

（平成17年1月11日受付）