

論文の内容の要旨

論文題目 「ゲノム情報解析のためのインタラクティブ可視化環境に関する研究」

氏名 西村 邦裕

コンピュータが身の回りにあふれるようになり，インターネットなどを通じて，誰もが多くの情報に接することが可能な情報化社会となった．大量の情報へのアクセスが可能となったものの，その中から必要な情報や知見を見出すことが課題となってきている．そのため，大量の情報をいかに把握し，理解していくか，などを支援する手法が必要となっている．

ゲノムサイエンス分野において，whole genome shotgun法などによる塩基配列の解読や，DNAチップを利用したmicroarray法による発現量情報の取得など，生命情報の高速取得技術が発展してきている．この結果，ヒトゲノムの解析が完了し，さまざまな生命情報がNCBIやUCSCといった公共データベースに格納されるようになり，データベースのエントリ数も急激に増えている．

蓄積されている生命情報の解析には，情報科学による支援が不可欠となっており，近年，ゲノム情報科学，情報生物学，bioinformaticsと呼ばれる分野が生まれた．そのため，ゲノムサイエンスの分野において，生物学的な実験をすることも必要であることは変わらないものの，より多くの生命情報を利用した解析の重要性が高まってきている．手に入るゲノム情報の全体像をつかみ，俯瞰しながら注目すべき点を見つけることや，データ全体の傾向や特徴を捉え，新規機能や新規メカニズムを発見あるいは類推していく作業が必要となってきている．ゲノムサイエンス分野において，大量のゲノム情報の解析作業支援が要請されている，と言える．

一方，情報科学分野において，VR(Virtual Reality)技術が発展し，大画面を利用した情報提示や，さまざまな情報とのインタラクションを行うことが可能となった．また，コンピュータのグラフィックス性能の進化などにより，情報の可視化分野において，より多くの情報を可視化の対象として扱えるようになった．

以上の背景を踏まえ，本論文では，VR技術，特にインタラクションと情報の可視化によるアプローチによって，ゲノム情報の解析作業支援が行えるのではないかと，いう発想を得た．ゲノム情報をいかに可視化し，知的作業支援ともいえる解析支援を実現するか，というのが，本論文の課題である．本論文では，ゲノム情報の解析のために，インタラクティブ可視化環境を提案し，ゲノム情報解析における可視化について論じることを目的とする．具体的には，ゲノム情報を見ながら直観的に操作し，判断し，解析できる作業支援環境を，ゲノム研究者に提供できるように，ゲノム情報解析作業を整理し，VR技術を利用し

た可視化手法を提案・評価し、メタレベルからの考察を加えることによって、ゲノム情報解析における効果的な可視化手法の体系化を行う。

まず、解析に必要な情報の可視化について議論し、そのための可視化概念を提案する。解析は、情報の中から必要な情報にアクセスし、特徴を把握し、比較・判断しながら知見を見出す作業であると言える。適切に情報を可視化することは解析支援につながる。つまり、解析手法と可視化手法は切り離せないと考える。既存の可視化手法は、解析結果をいかに絵にするか、についてであり、かつ、ひとつの絵の中に全てを盛り込む手法ばかりである。しかし、試行錯誤などを要する解析の場合、解析結果だけではなく、その結果に至る過程も重要である。そのため、解析結果をひとつの絵に盛り込むことが重要ではなく、一つ一つの試行に対して、適切に可視化をし、それらを総体として見たときに、解析が進んでいることが望ましいと考える。

そのため、試行的解析を支援するための可視化、という意味で、「試行の可視化」、および、試行の流れを把握できるように空間上へマッピングすることや、作業履歴を利用することなどをする「履歴の可視化」を提案する。「試行」を繰り返し、かつ、「履歴」を利用することは総じて「プロセス」であり、本論文では、提案する可視化を「プロセス可視化」と呼び、試行錯誤が必要な解析作業におけるインタラクティブな可視化として位置づける。「プロセスの可視化」を満たしたインタラクティブ可視化環境の実現により、総合的にインタラクションをしながら解析支援を行うことが可能となる。

次に、対象とするゲノム情報解析について議論を行う。ゲノム情報の特徴としては、多層の情報であること、多異種な相互関係情報であること、があげられる。そのため、これらの層に応じ、かつ、参照や比較可能な「試行」の可視化が必要である。ゲノム情報解析の際の要求事項としては、全体像の提示、興味ある部分の抽出、他のデータとの比較、パラメータ変更などのデータとのインタラクション、繰り返し判断可能な仕組みがあげられる。また、ゲノム研究者の思考を観察し、染色体に沿った可視化が求められていることも導き出した。これらを踏まえて、ゲノム情報解析への要求事項を満たした「試行の可視化」および「履歴の可視化」の実装を行った。

対象となるゲノム情報解析として、現在、注目が集まっているゲノムコピー数解析に焦点を絞った。ゲノムコピー数は、疾患になると一部あるいは全体が増幅・欠損することがある。これらのコピー数変化領域に疾患に関係する遺伝子が存在することが推測され、ゲノムコピー数解析は重要と考えられている。また、正常の人同士でもゲノムコピー数の変化が発見され、集団間の差を見るために注目が集まっている。さらにDNAチップ技術の発展により、ゲノムコピー数情報を全ゲノム網羅的に取得することが可能となった。そのため、網羅的なコピー数情報からコピー数変化部位を抽出し、他のサンプルや既存のゲノム情報と比較・参照することにより、解析が行われている。

提案する「試行の可視化」として、上記のゲノムコピー数解析を対象にして行った。可視化手法として、大きく分け2つ、本数の可視化手法とはずれ値の可視化手法を提案した。

本数の可視化手法としては、ゲノムコピー数情報を、1サンプル・複数サンプルに対して可視化し、2次元・3次元、形、色、振動などを利用し、特徴的な部位(Loss of Heterozygosity: LOH, Homozygous Deletionなど)を抽出・比較・参照できる仕組みを提案した。また、コピー数情報同士の関係(対立遺伝子ごとのコピー数変化: UniParental Disomy, UPD)の可視化も行った。はずれ値の可視化手法としては、正常検体におけるコピー数変化(Copy Number Polymorphism: CNP)を対象として、異常値の染色体上における偏りを可視化する方法を提案した。

提案する「履歴の可視化」として、作業履歴を利用した手法と、空間を利用した手法を提案した。作業履歴を利用した可視化手法は、ユーザが意識的に履歴をつける主観的作業履歴と、自動的に履歴をつける客観的作業履歴とを用いて、解析支援をする方法である。この結果、作業履歴があることにより、情報探索などにおいては効率が上がるという示唆を得た。また、空間を利用した可視化手法としては、解析作業の一つ一つを空間に配置し、それを連続的に接続することによって、履歴を可視化する方法である。これにより、解析作業への理解度が上がることが主観的評価からわかった。

以上を踏まえて、提案する可視化手法を実際のゲノム情報解析に応用した。具体的には、肺がんのコピー数情報(LOH, Homozygous Deletionなど)に関する解析、および、正常の人のコピー数情報(CNP)についての解析、の2例を提示した。この結果、集団としてLOHが起きている部位、Homozygous Deletion部位、肺がんの分類と一致したコピー数変化部位などが抽出でき、CNPの候補領域についても発見することができ、有効性を示せた。

以上をまとめると、本論文では、ゲノム情報の解析支援を行うための解析手法と可視化手法をあわせた意味での「可視化」について論じた。現在、ゲノム情報は多量高速に取得できるDNAチップ技術などの発展により、急速に蓄積されつつあり、可視化を利用した解析に対する要求が非常に大きくなっている。しかしながら、情報の解析だけでも手一杯な状況であり、ゲノム情報の可視化に対する議論はほとんどなされていないのが現状である。本論文では、ゲノム情報解析と密接に絡んだ可視化手法を検討し、一つ一つが簡単な可視化手法であっても解析の過程それぞれで情報をインタラクティブに可視化することが重要であることを示し、「試行の可視化」として提案した。また、それらを空間に位置づけることや、履歴を利用することとして「履歴の可視化」を提案した。さらに、ゲノム情報解析支援のための「試行の可視化」および「履歴の可視化」を合わせた総合的な「プロセスの可視化」を提案し、インタラクティブ可視化環境の実現を提唱した。そして、実際のゲノムコピー数解析を対象として、提案する可視化を行い、ゲノム情報解析を整理し、解析支援のレベルに応じた新規的な可視化手法を開発した。実際のゲノムコピー数解析に応用し、評価・検証を行い、提案する可視化が理解度や効率の面および実際の解析の面で有効であることを示した。また、提案する「プロセスの可視化」が、研究者の試行錯誤を可能とし、解析における思考支援につながっていると考える。

今後、本論文で提案する「プロセスの可視化」を多くの事例に適応することで、ゲノム

情報解析がより進展すると考える。また、「プロセスの可視化」をゲノム以外の分野に適応することで、大量情報の解析や複雑な処理およびユーザに思考支援が必要な際に、有効であると考えている。