

審査の結果の要旨

氏名 河村 大輔

本論文は「Genome-wide detection of human copy number variations using high density DNA oligonucleotide arrays」と題し、5章から構成されている。近年、正常個人間におけるゲノム配列の大規模な重複や欠失の多型であるコピー数多型(Copy Number Variation, CNV)がヒトの多様性の形成や疾患に大きく関係していることが明らかになりつつあるが、その全貌は明らかではなかった。DNA マイクロアレイは、このコピー数多型をゲノムワイドに検出することができる技術であり、またシーケンシングなどの技術に比べると廉価で短時間に結果が得られることが利点であるが、シグナルに含まれるノイズにより検出力が低下すること、リファレンスとなる理想的なゲノムが存在しないことにより検出結果に生じるバイアスと検出力の低下が、大きな課題であった。本論文は、これらの問題を解決するために、コピー数多型を高精度に検出する新たなアルゴリズムの開発を行っている。また、提案手法を多様な人種から成る実際のサンプルへ適用をすることで、ヒトゲノムのコピー数多型の全貌を明らかにしている。

第1章「Introduction」では、本論文の背景と目的を述べ、本論文の構成を述べている。まず、本研究では、DNA マイクロアレイの中でも一塩基多型を検出するために開発された SNP アレイを用いているが、SNP アレイをコピー数解析に用いることの利点として、一塩基多型とコピー数多型が一度に検出可能であることなどを挙げている。また、本論文で提案する手法が、シグナルに含まれるノイズの低減を行う前処理と、コピー数多型検出の2つのステップから構成されていることを述べている。

第2章「Pre-processing for Microarray Data」では、シグナルに含まれるノイズの低減を行う部分を論じており、SNP アレイのシグナルに含まれるノイズをモデル化し、それらのノイズを除去する統計的手法の提案とその有効性の評価を行っている。まず、SNP アレイにおけるノイズを (1)クロスハイブリダイゼーションによって生じるノイズ、(2)制限酵素認識配列上に SNP が存在することで生じるノイズ、(3)異なる遺伝子型を持つ2つのサンプルを比較する際に、プローブと DNA の結合力の違いが原因で生じるノイズ、(4) 2つのサンプルを比較する時の実験条件の違いに起因するノイズ、の4種類に分類している。そして、各ノイズをモデル化し、除去する方法について述べている。(1),(2)は原理的に除去不可能なため、そのようなノイズを引き起こす可能性のあるプローブを除くことで対処している。(3)は Gaussian

Mixture Clustering を用いることにより、多数の訓練データから結合力を推定し補正を行う手法を提案している。(4)は PCR product の長さや GC 含有率、プローブの結合自由エネルギーなどの値がノイズの強さと相関していることを示し、ロバスト回帰分析を用いることでノイズ成分を正確に推定し除去する手法を提案している。また、実際のデータを用いて、すでに提案されているアルゴリズムとのノイズ低減効果の比較を行い、提案手法が優れていることを定量的に示している。

第3章「Identification of Copy Number Variations」では、コピー数多型の検出を行う部分を論じており、従来提案されていたアプローチが、検出力の低下や、結果のバイアスを招くことを指摘し、それらの問題を解決する新たなアルゴリズムの提案、及び評価を行っている。解決している問題は2点ある。1点目は、homozygous deletion の検出に関する問題である。従来の遺伝子型に依存する手法では、解像度が低下することを示し、homozygous deletion を検出するための新たな指標と検出手法を提案している。2点目は、リファレンスサンプルの選択に関する問題である。従来はリファレンスサンプルを固定していたために、検出力の低下、及び結果にバイアスが生じることを指摘している。そのため、多数のサンプルの全組み合わせで比較し、その情報をもとにして、コピー数多型領域ごとに適切なりファレンスサンプルをダイナミックに選択する、新たな手法を提案している。また、実データに対し提案手法を適用し、PCR や質量分析計などの確認実験を通してこれらの提案手法の有効性を示している。

第4章「Copy Number Variations in Hapmap Samples」では、提案する手法を用いて、アジア人、アフリカ人、ヨーロッパ人から成る270人のゲノムを解析し、その結果について考察を行っている。提案手法により解析した結果、1203カ所のコピー数多型領域が見つかり、またそれは全ゲノム領域の約5%をカバーしていた。これは従来考えられていたよりもはるかに広く、コピー数多型が一塩基多型以上に個人間の塩基配列の違いを生み出している可能性が示されたと述べている。また、様々な疾患関連遺伝子のコピー数多型や、特定の人種のみに見られるコピー数多型も発見された。これらの結果により、コピー数多型が、医学的にも進化学的にも重要であることが示されたと述べている。

第5章「Conclusions」では、以上の成果を要約し、今後の課題を展望している。

以上を要するに、本論文はゲノム配列の大規模な重複や欠失の多型であるコピー数多型解析を目的とし、マイクロアレイを用いて効率よく高精度に検出するための解析手法を確立し、提案手法を多数の実データに適用することでヒトゲノムのコピー数多型の全貌を初めて明らかにしている点で非常に意義がある研究であり、その成果は工学的に貢献するところが大きい。

よって本論文は博士(工学)の学位請求論文として合格と認められる。