

疾患関連コピー数多型解析の基盤整備

●石川 俊平

東京大学大学院医学系研究科・人体病理学・病理診断学分野

<研究の目的と進め方>

申請者石川の参画した国際コンソーシアムによる第一世代ヒトゲノムコピー数多型地図の作成によりこれまで未知であったコピー数多型 (CNV) が多数見つかリ、ヒトゲノムの少なくとも1割以上の領域を占めることが明らかになった。

今後 CNV を用いた疾患関連解析が加速するに従い大きな問題になってくるのは

①これまでの CNV の discovery と、実際の関連解析に用いるのとは解析アルゴリズムに要求される精度が全くことなるが、現行のいずれの手法もそれを満たしているとは言い難い。

②それに従い正常人における CNV の頻度の情報が実際に関連解析できる精度に達していない。

本研究ではこれらの問題を解決し、我が国で行われる CNV 疾患関連解析の根本的基盤となるべき技術、データベースの作成を目的とする。

<2008 年度の研究の当初計画>

平成 20 年度には、全ゲノムで 180 万箇所の情報が取れる高密度オリゴヌクレオチドマイクロアレイを用いて、アルゴリズム開発におけるシグナルのトレーニングを目的として、主に計 50 検体程度の DNA サンプルについて情報を取る。ここでは異なる施設、設備、実験条件などから生まれる様々な特有のシグナルパターンを学習させる為、様々な抽出法、温度、時間などの実験条件でデータを取得する。数 Kbp 程度の小さな CNV、非常にわずかなシグナルの差の CNV も安定して検出するためのアルゴリズムの改良を行う。また検出された CNV について個々に確認実験を行い、いくつかの家系サンプルを用いても検証する。現在研究室にある質量分析器 (MassArray) を用いれば定量 PCR では難しい 1 コピー差の検出も迅速かつ正確に測定することが可能であるため検証実験に用いる予定である。この検証実験の結果、および国際共同機関における同一検体からのシグナルファイルや他のアッセイ法による結果と照合し、アレイデータからの検出アルゴリズムの再確認と検証を行う。また現行のアレイにおいては複雑な CNV 領域のアレル情報を取り出すことができないという現状を踏まえアレル別のコピー数を解析可能な次世代マイクロアレイ、もしくはフォーカスアッセイ法の開発を行う。

<2008 年度の成果>

1) Real-time PCR は特定の領域において CNV を検出するために通常使われるが、複雑なゲノム領域において 1 コピーの差を正確に測定することは難しい。この問題を解決するために Nanofluidic chip によるデジタル PCR を試みた。ゲノムのサンプルを段階的に希釈して、TaqMan の試薬とともにそれぞれ 6 nL の容量をもつ約 10000 個の微小区画に流し込む。テンプレートが

入った区画のみ PCR 後シグナルが得られるため、シグナルの得られる微小区画の数によってテンプレートの濃度がデジタルカウント可能である。重要な薬剤の代謝遺伝子であり、コピー数多型の見られる CYP2D6 について Nanofluidic chip を使ったデジタル PCR によりコピー数を測定し、高密度オリゴヌクレオチドマイクロアレイとの比較を行った。デジタル PCR は、高密度オリゴヌクレオチドマイクロアレイに比して定量性にすぐれ配列の類似した CYP2D7 とも明瞭に区別して測定できた。

2) ゲノムワイドにおける正確な絶対コピー数探索を可能な技術を確立するために、Illumina Genome Analyzer による配列 Tag の出現頻度を計測することによりコピー数のデジタルカウントを試みた。HCT 116 細胞株のゲノムを断片化し、138M 個の配列 Tag (5Gbp 相当) の配列を 3Kbp 間隔で集計すると、従来の SNP6.0 array に比して高解像度のコピー数情報が得られ、マイクロアレイでは検出できなかった微小な増幅、欠失が観測された。

<国内外での成果の位置づけ>

2006 年に第一世代の CNV ドラフト地図が作成されて以来、以来多くの CNV の報告がなされ、様々な性質や関連解析の報告などもされた。

最近の解析によると SNP やそこから推定される haplotype の系統樹と、CNV のそれとは様子が異なり、CNV の系統樹ではおそらく実際の伝播とは異なると思われるパターンが見られる。コピー数がダイナミックに増減を繰り返している、同祖でない CNV が発生しているなどの様々な理由が考えられるがいずれにせよ CNV を周辺の SNP でタグする (マーキングする) という発想は単純には成功しないことを示唆している。疾患関連解析においては CNV 領域内の情報を直接測定するプラットフォームが必要とされると考えられるようになってきた。

近年 CNV を直接測定するために、多数の高密度アレイが市販されているにも関わらず、アレイのデータは実験条件やサンプル状態の影響を受けやすく、CNV を推定することは単純ではない。大規模関連解析を含めた様々な解析に対してより高感度で安定した新しい手法、アルゴリズムの開発が求められる。

<達成できなかったこと、予想外の困難、その理由>

当初はマイクロアレイによる CNV の解析基盤整備を予定していたが、すでに共同研究を行っている Affymetrix 社を通じて比較的解析環境が整備されてきたこと、高速シーケンサーの技術革新が予想以上に進んだことなどから、解析の中心をシーケンサー、Nanofluidics chip を用いたコピー数のデジタルカウントに移行させた。ゲノム特定内部でも高速シーケンサーを使用する班員が予想され、これらを想定して解析可能な基盤を作っていく。

<今後の課題>

Illumina Genome Analyzer による配列 Tag を用いたコピー数のデジタルカウントのデータは、実験条件、解読エラー、マッピングアルゴリズムによって大きく変わることが判明した。これらの問題を克服し、関連解析、臨床検査に用いる精度のアルゴリズム、実験手法の改良を図る。また配列を利用して、アレル情報、パラログ情報を取得しながら、

複雑なゲノム領域についてアレル情報とコピー数の情報を統合した解析技術を開発する。

<成果公表リスト>

1)論文／プロシーディング（査読付きのものに限る）

1. 0701132011

Richard Redon, Shumpei Ishikawa, Karen R. Fitch, Lars Feuk, George H. Perry (First authors with equal contribution) et al.: Global variation in copy number in the human genome, *Nature*, vol 444, p444-454(2006).

2. 0701132015

Daisuke Komura, Fan Shen, Shumpei Ishikawa (First authors with equal contribution) et al.: Genome-wide detection of human copy number variations using high-density DNA oligonucleotide arrays, *Genome Research*, vol 16, p1575-1584(2006)

3. 0710291623

Fujii K, Ishikawa S, Uchikawa H, Komura D, Shapero MH, Shen F, Hung J, Arai H, Tanaka Y, Sasaki K, Kohno Y, Yamada M, Jones KW, Aburatani H, Miyashita.: High-density oligonucleotide array with sub-kilobase resolution reveals breakpoint information of submicroscopic deletions in nevoad basal cell carcinoma syndrome. *Hum Genet.* 2007 Dec;122(5):459-66

4. 0901112155

Nakamura Y, Matsubara D, Goto A, Ota S, Sachiko O, Ishikawa S, Aburatani H, Miyazawa K, Fukayama M, Niki T.: Constitutive activation of c-Met is correlated with c-Met overexpression and dependent on cell-matrix adhesion in lung adenocarcinoma cell lines. *Cancer Sci.* 2008 Jan;99(1):14-22.

5. 0901112202

Midorikawa Y, Yamamoto S, Tsuji S, Kamimura N, Ishikawa S, Igarashi H, Makuuchi M, Kokudo N, Sugimura H, Aburatani H.: Allelic imbalances and homozygous deletion on 8p23.2 for stepwise progression of hepatocarcinogenesis. *Hepatology.* 2008 Oct 24. [Epub ahead of print]

2)データベース／ソフトウェア

1. 0701132037

Genotyping Microarray based CNV Analysis (GEMCA) http://www.genome.rcast.u-tokyo.ac.jp/CNV/gemca_details.html

2. 0701132015

Database of Genomic Variants
<http://projects.tcag.ca/variation/>