

1 ヒト疾患関連コピー数多型解析、2 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

<研究の目的と進め方>

1. ヒトの疾患罹患性、薬剤応答性等を規定する遺伝子多型としては SNP が最もよく知られているが、最近になってヒトには 10Kbp～1Mbp の領域に及ぶゲノムの構造的多様性、すなわちコピー数の多型 CNV(copy number variation) が従来の予想よりも高頻度に存在することが明らかになってきた。一般には細胞あたり 2 コピーの遺伝子が存在するが、ある個人には 3 コピーもしくは 1 コピーとなるような多型である。上記のような長い領域にわたるコピー数の変異(多型)が糸球体腎炎の発症リスク、HIV の感染効率等の原因となりうるものが少数であるが知られているが、多様な人種内、もしくは人種間における頻度、分布はほとんどが未知である。CNV の詳細な解析が、従来型 SNP に加え、疾患等のリスクの個性、薬剤応答性の個性の解明という点において新たな可能性を引き出すことが期待される。今回の計画は①期間内に日本人を含む多様な人種における CNV の分布、頻度、遺伝様式を特定し②従来型の多型である SNP との関係、癌、生活習慣病を含めたヒトの表現型への寄与をも明らかにすることを目的とする。

2. 代表者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、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異なるが、現行のいずれの手法もそれを満たしているとは言いがたいということである。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。

具体的には標準的な CNV 解析のプラットフォームになると考えられる高密度オリゴヌクレオチドマイクロアレイによる CNV 解析のアルゴリズムの世界標準を開発する。特に大規模関連解析に対するアプリケーションの強化を重点的に行い国内の研究者が使用しやすい環境をつくる。また現行の市販の Affymetrix や Illumina のマイクロアレイプラットフォームは単に CNV 領域のコピー数を見るのみでありアレルの解像度では見ることはできない。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疾患関連解析を可能にする次世代技術、プラットフォームを開発する。

<研究開始時の研究計画>

1. 全ゲノムで 50 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主に国際 HapMap project のサンプルを中心に日本人の DNA サンプルを中心に情報を取る。数 Kbp 程度の小さな CNV も存在することが予想され、こうした小さな変化を少数のプロープ数でより正確に見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

を行う。従来の定量的 PCR では 1 コピーの増加、減少を正確に見ることは難しく FISH 法ではコストおよび時間がかかる。質量分析器を用いれば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。

上記の過程で得られた CNV 情報について、日本人を中心とする多様な人種における頻度、分布、遺伝様式を始め、様々なゲノム上の特徴との関連について詳細な解析を行う。特に i) 癌、生活習慣病、薬剤感受性の関連遺伝子領域と CNV との関連、ii) 従来の SNP による連鎖不均衡地図、関連解析の結果がどの程度 CNP の存在によって訂正、補完されるかに重点をおいて詳細な解析を行う。

高頻度な CNV 領域については当初の想像以上にオリゴヌクレオチドアレイのプロープが抜け落ちていることが明らかになった。CNV の領域ではアレル数の不均衡による call rate の低下、Mendel Inheritance エラーや H-W エラーによってアレイの開発段階で多くのプロープが抜け落ちていることがわかった。こうした従来のマイクロアレイではカバーできない CNV 領域が多数見られたという事実を踏まえ、CNV カバーのより良い高密度なアレイプラットフォームの開発を行う。

2. 全ゲノムで 100 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。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。

<研究期間の成果>

1. 世界の 7 つの研究機関、企業が国際コンソーシアムを結成し、日本人を含む多様な人種の CNV の全貌を解明するプロジェクトを立ち上げた。我々は合成オリゴヌクレオチドゲノタイプアレイを用いて、癌、遺伝病、または個人間などにおけるゲノムコピー数測定をアレル別に正確に決定する方法を樹立している。この手法は、ノイズを含むデータの中から統計的手法で有意な領域を推定するという一般的な手法ではなく、サンプル調整からデータ

取得までの過程で生じる様々なノイズの原因となりうる実験的パラメーターを数値的に設定し、それを補正することによりノイズそのものを除去するというユニークな手法である。今回は Affymetrix 社と共同で全ゲノムで 50 万箇所以上の情報が取得可能な 500K mapping のシグナル値から上記アルゴリズムを使って CNV を特定した。その結果 HapMap で使われた 3 系統 270 人のサンプルの中に、約 1500 箇所、ヒトゲノムの 12% 以上という、従来考えられていたよりもはるかに広い領域にコピー数多型が見られ、従来知られていた SNP 以上の個人間塩基多様性を生み出している可能性が示された。また約 3000 個の遺伝子が含まれておりこれらの個人差に大きく寄与している可能性が示された

高頻度な CNV 領域については当初の想像以上にオリゴヌクレオチドアレイのプロープが抜け落ちていることが明らかになった。我々が Affymetrix と共同で開発したオリゴヌクレオチドアレイは全ゲノムで 130 万 (1.3M) のデータポイントと従来の 500K array の倍以上の情報が取得可能である。CNV 領域のカバー率も良く、次世代ゲノムアレイ SNP6.0 array のプロトタイプとなった。このアレイに GIM/GEMCA アルゴリズムを実装した。先天性腫瘍性疾患 Gorlin 症候群のゲノム DNA の解析では PTCH を含む最小 165Kbp の欠失が見つかりいずれの症例にも breakpoint (切断点) が同定可能であった。

2. 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集積流体回路によるデジタル 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 について集積流体回路を使ったデジタル 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。また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、並列型シーケンサーによる配列 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。大腸がん細胞株のゲノムを断片化し、Tag 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>

3 系統 270 人のサンプルのなかに約 1500 箇所、ヒトゲノムの 12% 以上という、従来考えられていたよりもはるかに広い領域にコピー数多型が見られ、従来知られていた SNP 以上の個人間塩基多様性を生み出している可能性が示された。また約 3000 個の遺伝子が含まれておりこれらの個人差に大きく寄与している可能性が示された。この事実はサイエンスのコミュニティーに大きなインパクトを与えた (Nature2006 に発表、また 2007 年の Science's breakthrough of the year のトップに選ばれた)。また遺伝子の数の個人間の違いがたくさんあるという事実は、一般の人々からも興味を持たれ世界のマスコミ (日本 5 大紙の他、ABC ニュース、BBC ニュース、ニューズウィーク誌、ワシントンポスト紙、ニューヨークタイムズ紙等) に取り上げられた。また我々の開発し、国際コンソーシアムで使われたオリゴヌクレオチドアレイのコピー数推定アルゴリズム GIM/GEMCA の基本コンセプト

は様々なアルゴリズムに受け継がれコピー数多型のスタンダード手法となっている。

<達成できなかったこと、予想外の困難、その理由>

1. 高頻度な CNV 領域については当初の想像以上にマイクロアレイのプロープが抜け落ちていることが明らかになった。CNV の領域ではアレル数の不均衡による call rate の低下、Mendel Inheritance エラーや H-W エラーによってアレイの開発段階で多くのプロープが抜け落ちていることがわかった。その結果、高頻度な CNV 領域では数百 Kbp の間隔があくこともあり、こうした領域の正確な境界の決定、コピー数の決定は課題を残した。これらの領域を正確に測定するには CNV 領域内にプロープを設計した次世代アレイ、もしくは別の platform 等による targeted approach が必要と考えられた。

<今後の課題、展望>

並列型シーケンサーによる配列 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。

またデジタル PCR についてはコピー数変化の程度が小さい場合、また変化のあるゲノムの割合が少ない場合 (組織中の腫瘍、末梢血中の腫瘍、胎児 DNA) などを想定し実際に検査に用いられるレベルまでの精度を達成することを目的とする

2006 年に第一世代の CNV ドラフト地図が作成されて以来、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。

<研究期間の全成果公表リスト>

1) 論文/プロシーディング

1. 0911121053

Hino R, Uozaki H, Murakami N, Ushiku T, Shinozaki A, Ishikawa S, Morikawa T, Nakaya T, Sakatani T, Takada K, Fukayama M.: Activation of DNA methyltransferase 1 by EBV latent membrane protein 2A leads to promoter hypermethylation of PTEN gene in gastric carcinoma. *Cancer Res.* 2009 Apr 1;69(7):2766-74.

2. 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*. 2009 Feb;49(2):513-22.
3. 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.
4. 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 nevoid basal cell carcinoma syndrome. *Hum Genet*. 2007 Dec;122(5):459-66
5. 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).
6. 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)
- 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/>
- 3) 新聞記事 (抜粋)
1. 0702092015
ヒトゲノム、遺伝子重複の個人差は1447カ所
朝日新聞 11/23/2006
2. 0702092043
International team of scientists unveils map of human genetic variation
NEWS WEEK 誌 11/22/2006
- 3) 著書
1. 0702081159
ヒトゲノムのコピー数多型 : CNV
実験医学増刊 ゲノム情報と生命現象の統合的理解 2007 p69-75
- 4) 学会 / その他 (抜粋)
1. 0608071236
Global Detection of Copy Number Variations in Human Genome by Affymetrix 500K Mapping Arrays.
Cold Spring Harbor Laboratory 2006 Biology of the Genome Meeting ポスター発表
2. 0701132043
Copy Number Variation of the Drug Associated Genes
Cold Spring Harbor Laboratory 2006 Pharmacogenomics Meeting ポスター発表