

## ヒト疾患関連コピー数多型解析

●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>

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、もしくは人間間における頻度、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。

### <2007 年度の研究の当初計画>

これまで参画してきた国際コンソーシアムにおけるプロジェクトの過程で得られた CNV の情報について、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 CNV 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 500K 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 500K array 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 カバーのよいより高密度なアレイプラットフォームの開発を行う。

また 500K array user である研究者達が自由にコピー数多型の関連解析を行えるように、解析ツールを充実させ利用できる環境を作る。いくつかの batch process を導入し、大量サンプル処理にも対応できるようにしたい。

### <2007 年度の成果>

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精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 500K 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 を含む最少 165Kbp の欠失が見つかりいずれの症例にも break point(切断点) が同定可能であった。

### <国内外での成果の位置づけ>

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精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。

#### <達成できなかったこと、予想外の困難、その理由>

高頻度な CNV 領域については当初の想像以上に 500K アレイのプローブ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ブが抜け落ちていることがわかった。その結果、高頻度な CNV 領域では数百キロ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 領域内にプローブ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>

高頻度な CNV 領域については当初の想像以上に 500K アレイのプローブ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ブが抜け落ちていることがわかった。こうした 500K array 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 カバーのよいより高密度なアレイプラットフォームの開発を行う。

また 500K array user である研究者達が自由にコピー数多型の関連解析を行えるように、解析ツールを充実させ利用できる環境を作る。いくつかの batch process を導入し、大量サンプル処理にも対応できるようにしたい。

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いえない。これらの問題を解決し、より精度の高いアルゴリズムの完成を目指す。

#### <成果公表リスト>

##### 1) 論文/プロシーディング

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

Katsunori Fujii, Shumpei Ishikawa et al.: High-density oligonucleotide array with sub-kilobase resolution reveals breakpoint information of submicroscopic deletions in nevoid basal cell carcinoma syndrome. *Hum Genet.* 2007 Aug 17

##### 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](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