

ゲノム研究における Copy Number Variants (CNV) の重要性

CNV

Single Nucleotide Polymorphisms (SNP) と Copy Number Variants (CNV)

2003年のヒトゲノム計画の完了により、ゲノムに一塩基多型(Single Nucleotide Polymorphisms: SNP)とよばれる配列上の多様性が存在することが明らかになり、SNPがヒト表現型に多様性を与える要因の一つとして注目を集めました。癌や糖尿病などの疾患に対するかかりやすさや、薬剤に対する効果や副作用といった、いわゆる個人の”体質”がSNPにより影響されることが予想され、テラーメイド医療やゲノム創薬の実現を目的に、国際HapMapプロジェクトが進められ、SNPの位置やそのハプロタイプ(各遺伝子座位にある対立遺伝子のいずれか一方の組合せ)、同一染色体上で遺伝的に連鎖しているSNPの組合せなどが大規模に調べられました。また、疾患の原因遺伝子のSNP情報と疾患表現型を関連付ける研究、Genome Wide Association Study (GWAS)も進められ、糖尿病、ぜんそく、高血圧や、癌、B型慢性肝炎、摂食障害など、様々な疾患のリスクに関与する多型が報告されています。

さらに最近、SNPとは異なるゲノム多様性として、DNAの量的な変化を生じるコピー数多型(Copy Number Variants: CNV)が、当初考えられていたよりはるかに高頻度にヒトゲノム中に存在することが明らかになりました^{1~10)}、生活習慣病や自己免疫疾患を含むヒトの形質差に影響

を与える可能性が注目されています。これまでに検出されたヒトCNVの公的データベースも構築が進められており、

Database of Genomic Variants (DGV)

(<http://projects.tcag.ca/variation/>)

Human Structural Variation Database

(<http://humanparalogy.gs.washington.edu/structuralvariation/>)

などが参照できます。(さらに新しいデータベース構築の動きもあります。<http://www.emory.edu/home/research/stimulus/go-challenge-grants.html>)。CNVの定義としては、ゲノム上のコピー数が個人間で異なる領域のうちで通常サイズが1kb以上のものとされ、1kbより小さいサイズのコピー数変化はInDelsと呼ばれています¹¹⁾。CNVは、その領域と直接overlapする遺伝子の数を変化させたり、発現制御因子に変化を与えたりすることにより、遺伝子発現量を変化させる可能性が重要視されています^{12,13)}。さらに、CNVがその領域のクロマチン構造を変化させることにより、近傍の遺伝子発現にも影響する機構(neighborhood effect)も報告されており¹⁴⁾、CNVが遺伝子発現へ影響を与える機構の研究が進んでいます。このように、SNPだけではなく、CNVをも考慮に入れた研究の重要性はますます高まっています。以下に最新のCNV研究の一端を紹介します。

■ Copy number variantsと疾患の関連性

William-Beuren症候群やCharcot-Marie Tooth neuropathy Type 1A¹⁵⁾などのいわゆるメンデル型疾患のように、単一遺伝子の量的变化が、疾患の発生やその表現型に影響することは以前から知られていましたが、本来個人の多様性に関わっていると考えられていたCNVの中にも、癌^{16~18)}、ぜんそく¹⁹⁾、HIV感染²⁰⁾、リューマチ性関節炎²¹⁾、全身性自己免疫疾患^{22,23)}、クローン病²⁴⁾など多くの一般的かつ複雑な疾患へのかかりやすさに関わるものがあることがわかつきました。その中にはリスクが高いものも多く含まれ、遺伝的な病因を考える上でCNVの重要性が示唆されています。これらの生活習慣病を含

む一般的な疾患については、Wellcome Trust Case Control Consortiumによる大規模なCNV association studyが行われています。

(<http://www.agilent.co.jp/newsjp/presrel/fy2008/ca06wtccc.shtml>)

また、CNVの多くがいわゆる”環境センサー”遺伝子（嗅覚受容体、免疫、炎症反応遺伝子、細胞シグナリング、細胞間接着分子、構造タンパク質、イオンチャネルなど）に存在し^{2,3)}、自然淘汰圧により影響されるものもある²⁵⁾ことも、CNVとヒト形質の多様性の関わりを示唆しており、注目を集めています。

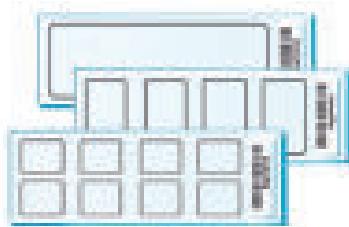
SNP + CNV = !

■ GWASの次に求められるもの

GWASによるSNPに基づく疾患原因遺伝子型と疾患表現型を関連付ける研究においては、その成果とともに問題点も指摘されています。common SNPに着目した場合に遺伝性を示さないものも多数あり、また一般的かつ複雑な疾患の原因遺伝子の特定や、重要な原因遺伝子に存在するrare SNPの検出が困難である点などが、特に診断への応用にあたっての問題点として考えられています。そこで現在多くの研究者よりGWASの次の段階として、今後フォーカスすべき2つのポイントが挙げられています。1つはシーケンサなどの併用により、集団全体においては影響が小さくても複雑に関与することで疾患リスクに影響を与えるrare SNPを探すこと、そしてもう1つはCNV検出です。

rare SNPの探索としては、個々の臨床試験の結果をさらに集めて統計的に解析するメタ解析による方法²⁶⁾や候補遺伝子領域とその近傍のシーケンシングを行う²⁷⁾などの方法があります。また標的領域を選択的に濃縮する方法との組み合わせによるシーケンス効率化²⁸⁾が原

因遺伝子の絞り込みやrare SNP探索のための効率的で強力なツールとなると期待されています。一方CNVについては、最近、特に小児発達障害や神経性疾患などの領域において研究が進んでいます。例えば統合失調症は遺伝的要因が高い²⁹⁾にも関わらず、その遺伝様式は複雑であり、これまでのGWASにより同定された関連遺伝子の影響は小さく、その原因の解明が困難でした。そこで、メタ解析^{30~33)}以外に、ゲノムワイドなCNV解析^{34~38)}を行うことにより、rare CNV、de novo CNVの例を含めた関連性のあるCNVが多く検出され、CNV解析の重要性が示されました。精神発達遅延^{39~44)}、てんかん^{45~46)}、自閉症^{47~50)}、多動性障害などにおいても、ゲノム構造異常との関連について多くの報告があり、診断ツールとしてのマイクロアレイの有用性についても注目されています^{51~53)}。



て注目されていますが、現時点ではduplicationの正確な検出が困難であることに加えて、マッピングするリファレンスゲノム配列にゲノム構造情報がなく、ギャップが多く存在し、2倍体の情報でない、などの問題点が挙げられています⁵⁴⁾。これらの理由から、現時点ではDNAオリゴマイクロアレイがCNV探索に多く用いられており、この先も活用されることが期待されます。

■ CNV探索のスタンダードテクノロジー

CNVを含めゲノムの量的な構造変化の検出手法としては、感度、分解能、分析コストなどの観点からDNAオリゴマイクロアレイが広く利用されており、精神疾患以外にも、癌や白血病やその他の疾患研究に幅広く応用されています。特に次世代シーケンサと比較すると、DNAオリゴマイクロアレイはコピー数変化を検出するのに”もっと費用効率が高い方法”であると言われています（Pollack博士、スタンフォード大学）。シーケンスもCNV検出の手法とし

■効率的なCNV探索に重要なポイント

CNVと表現型との関連性を探索していく上で重要なポイントは、高感度かつ高精度にCNVを検出することです。プラットフォームの解像度^{55,56)}(プローブの位置や密度)のみならずデータのシグナル/ノイズ比、コピー数検出のLog₂Ratio値なども、コピー数の定量性、正確なCNV breakpoint検出、データの再現性などに大きく関わります^{9,11)}。

マイクロアレイを大別すると、SNPマイクロアレイと、CGH(Comparative Genomic Hybridization)マイクロアレイがあり、それぞれ複数のメーカーからアレイプラットフォームが販売されています。SNPマイクロアレイでは解像度の高いものが数多くあり、CNV解析への応用例は数多く報告されていますが、シグナル/ノイズ比が低い問題などにより既知・潜在的CNVを検出しないことがあります^{8,57)}、また

rare CNVの検出が困難であることや、解析に時間や労力を要することも問題点としてあげられています^{58~60)}。各メーカー間のマイクロアレイの性能については多くの論文において比較されていますが^{61~63)}、ノイズレベルが低く、コピー数検出の際のLog₂Ratioが正確で、小さなサイズの構造変化も再現性よく検出するマイクロアレイを選択することが重要です。生データを加工するノイズ低減アルゴリズムの使用をできるだけ抑えるために、マイクロアレイ上に搭載するプローブがあらかじめコピー数検出目的に最適化されていることも、大事なポイントです。さらに、Log₂Ratioが圧縮されることなく、コピー数変化の大きさをできるだけ正確に捉えられることは、モザイク性のある変化領域の検出においても非常に重要です^{64,65)}。

■コスト、スピード、カスタム化も重要

rare CNVを検出するには大規模スケールの実験が必要となるため、低コストとハイスループット性が強く求められます。また、ある程度対象領域が絞られてきた場合には、ゲノムワイドな解析ではなく、目的とする複数の標的loci検出をより低コストで検出するカスタム化も望まれています。定量PCRの手法は標的lociを絞ることは可能ですが、複数の領域を一度に検出することが困難なため、マイクロアレイのカスタム化、スループット性もCNV研究には

求められています。

今、遺伝的要因の高い疾患において、遺伝的に影響のある(遺伝率のある)ゲノムの変化を探索する目的に、大きな注目を集めているCNV研究。今後、マイクロアレイを用いたCNV研究はさらに加速していくことでしょう。そして、その効果的な研究アプローチの鍵となるものは、用いるマイクロアレイの特性にあるかもしれません。

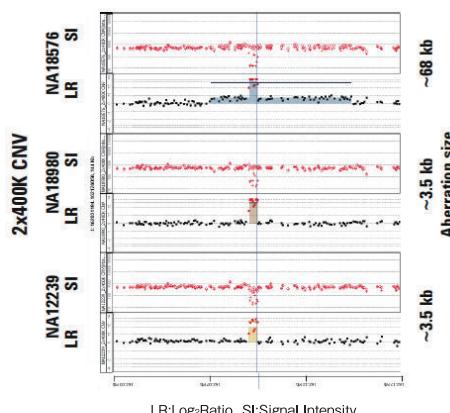
■アジレントCNVマイクロアレイ

CGH/CNVマイクロアレイは、そもそも健常人で基本的に2コピー存在するDNAのわずかなコピー数変化を検出するため、遺伝子発現用アレイよりもさらに高い精度が求められます。またゲノムDNAは複雑性が高く、特にCNVはSegmental Duplicationなどゲノム上で相同性が高い領域に高頻度に分布するため、プローブ設計には細心の注意が求められます。

アジレント・テクノロジーは、Tmやホモロジーを考慮したプローブ設計、高い効率のハイブリダイゼーション、クロスハイブリダイゼーションを最小限に抑えるための厳しい洗浄条件、スキャナーの高感度化や数値化ソフトウェアの改良など、アレイプローブとプロトコール全体を最適化することで、これまでにない高い精度と信頼性でコピー数変化を検出できるマイクロアレイを開発しました。アジレントCNVマイクロアレイは、リファレンスサンプルを用いた2色法により、LogRatioがほとんど圧縮されることなく、コピー数の比を正確に検出することができます。またCNVはその多型性のため、1コピー以下の微細なコピー数変化をも検出できる解像度が必要ですが、アジレントCNVマイクロアレイはこの点でも、高い評価をうけています。

右図は、DGVに登録されているCNVの領域にフォーカスした40万のプローブを搭載しているアジレント2x400K CNVマイクロアレイのデータの一例です。このデータは生データをそのままプロットしたもので、ノイズ低減アルゴリズムを使用していません。生データからでも、研究者が視覚的に、コピー数の増幅、欠失およびそのBreak Pointを明確に捉える事が可能です。

アジレントのマイクロアレイは1枚のスライドグラスに複数のアレイを載せてコストを低減させることができます。CNVカタログアレイとしては2x400K、1x1MとWTCCCで用いられている2X105K⁶⁶⁾のラインアップがあります。また絞り込まれた領域に、ヒト2800万のプローブデータベースから選択したプローブを極めて高密度に配置するカスタムアレイも1枚から製造、発注できます。アジレントCNVマイクロアレイは、ますます重要性の高まるCNV研究を加速していきます。



文献

- 1) Lafrate, AJ, et al., *Nat. Genet.* (2004) 36(9)
- 2) Sebat, J, et al., *Science*, (2004) 305, 525-528
- 3) Tuzun, E, et al., *Nature Genetics*, (2005) 37(7), 727-732
- 4) Redon, R, et al., *Nature*, (2006) 444, 444-453
- 5) Sharp, AJ, et al., *Nature Genetics*, (2006) 38(9), 1038-1042
- 6) Lin, C-H, et al., *BMC Genet.*, (2008) 9:92
- 7) Levy, S, et al., *PLoS Biol.* 2007 October; 5(10): e254
- 8) Kidd, JM, et al., *Nature*, (2008) 453(7191): 56-64.
- 10) Carter, NP, *Nat. Genet. Suppl.* (2007) 39, S16-21
- Smith, AJ, et al., *Human Molecular Genetics*, (2007) 16 (23), 2783-2794
- 11) Scherer, SW, et al., *Nat. Genet. Suppl.* (2007) 39, S7-S15
- 12) Feuk, L, et al., *Hum Mol Genet.* (2006) 15 Review Issue 1, R57-66
- 13) McCarrol, SA and Altshuler, DM, *Nature Genetics Supplement*, (2007) 39, S37-42
- 14) Kleinjan, DA and van Heyningen, V, *Am. J. Hum. Genet.* (2005) 76, 8-32
- 15) Lupski, JR, et al., *Cell*, (1991) 66(2), 219-232
- 16) Shlien, A, et al., *PNAS*, (2008) 105 (32) 11264-11269
- 17) Park, J, et al., *Cancer Epidemiol, Biomarkers Prev.*, 15, 1473-1478 (2006)
- 18) Jakobson, J, et al., *Endocrinol. Metab.* (2006) 91(2), 687-693
- 19) Brasch-Andersen, C, et al., *Hum. Mutat.* (2004) 24, 208-214
- 20) Gonzalez, E, et al., *Science*, (2005) 307, 1434-1440
- 21) McKinney, C, et al., *Ann. Rheum. Dis.* (2007)
- 22) Aitman, TJ, et al. *Nature*, (2006) 439, 851-855
- 23) Fanciulli, M, et al., *Nat. Genet.* (2007) 39, 721-723
- 24) Fellermann, K, et al., *Am. J. Hum. Genet.* (2006) 79, 439-448
- 25) Perry, GH, et al., *Nat. Genet.* (2007) 39(10), 1256-1260
- 26) Scott, LJ, *Natl Acad Sci U S A.* (2009) 106(18), 7501-7506
- 27) Romeo, S, et al., *Nat. Genet.* (2007) 39, 513-516
- 28) Turner, EH, et al., *Annu. Rev. Genomics Hum. Genet.* (2009) Jul 16
- 29) Sullivan, PF, *PLoS Med.* (2005) 2(7) e212, 614-618
- 30) Shi, J, et al., *Nature*, Published online 1 July 2009
- 31) The International Schizophrenia Consortium, Common polygenic variation contributes to risk of schizophrenia and bipolar disorder, *Nature*, Published online 1 July 2009
- 32) Stefansson, H, et al., *Nature*, Published online 1 July 2009
- 33) O'Donovan, MC, et al., *Nat. Genet.* (2008) 40(9), 1053-1055
- 34) Need, AC, et al., *PLoS Genetics*, (2009) 5(2)
- 35) The international Schizophrenia Consortium, *Nature* (2008) 455, 237-241
- 37) Stefansson, H, et al., *Nature*, (2008) 455 232-237
- 38) Xu, B, et al., *Nat. Genet.* (2008) 40, 880-885
- 39) Walsh, T, et al., *Science*, (2008) 320, 539-543
- Toruner, GA, et al., *American Journal of Medical Genetics*
- 40) Part A (2007) 143A, 824-829
- Fan, YS, et al., *Human Mutation*, (2007) 0,1-9
- 41) Aradhya S, et al., *American Journal of Medical Genetics Part A* (2007) 143A, 1431-1441
- 42) Rooryck, C, et al., *Eur J Med Genet.* (2008) 51(1), 74-80
- 43) Sharp, AJ, et al., *Nat. Genet.* (2008) 40(3), 322-8
- 44) Bijlsma, EK, et al., *European Journal of Medical Genetics*, (2009) 52(2-3), 77-87
- 45) Helbig, I, et al., *Nat. Genet.* (2009) 41(2), 160-162
- 46) Dibbens, LM, et al., *Hum. Mol. Genet.* (2009)
- 47) Glessner, JT, et al., *Nature*, (2009) 459, 569-573
- 48) Sebat, J, et al., *Science*, (2007) 316, 445-449, *Nat. Genet.* (2007) 39(3), 319-328
- 49) Autism Genome Project Consortium, *Nat. Genet.* (2007) 39(3), 319-328
- 50) Marshall, CR, et al., *Am. J. Hum. Genet.* (2008) 82(2), 477-88
- 51) Stankiewicz, P and Beaudet, AL, *Current Opinion in Genetics & Development*, (2007) 17(3), 182-192
- 52) Edelmann, L and Hirschhorn, K, *Ann. N.Y. Acad. Sci.* (2009) 1151, 157-166
- 53) Aston, E, et al., *J. Med. Genet.*, (2008) 45, 268-274
- 54) Maher, Nature, (2008) 456, 18-21
- 55) Estivill, X and Armengol, L, *PLoS Genetics*, (2007) 3(10), 1787-1799
- 56) Perry, GH, et al., *Am. J. Hum. Genet.* (2008) 82, 685-695
- 57) Cooper, GM, et al., *Nat. Genet.* (2008) 40, 1199-1203
- 58) Carlson, CS, et al., *Hum. Mol. Genet.* (2006) 15, 1931-1937
- 59) McCarroll, SA, et al., *Nat. Genet.* (2006) 38, 86-92
- 60) Newman, TL, et al., *Hum. Mol. Genet.* (2006) 15, 1159-1167
- 61) Greshock, J, et al., *Cancer Res.*, (2007)
- 62) Gunnarsson, R, et al., *Genes, Chromosomes & Cancer* (2008) 47, 697-711
- 63) Zhang et al., *European Journal of Human Genetics* (2008)
- 64) Shinawi, M, et al., *Blood*, 112(4), 1042-1047 (2008)
- 65) Kamath, A, et al., *Cancer Genetics and Cytogenetics*, (2008) 183, 117-120
- 66) Conrad, DF, et al., *Nature*, Published online 7 October 2009



Agilent Technologies

アジレント・テクノロジー株式会社

〒192-8510 東京都八王子市高倉町9-1

カスタマコンタクトセンター

phone: 0120-477-111 fax: 042-660-8676