

表 7.2 既知の Duplication/Deletion 領域

遺伝子名	遺伝子産物	遺伝子座	遺伝子産物の機能	参考文献
<i>OR7E24</i>	olfactory receptor	19p13.2	臭い分子の受容体。	Newman et al. 2003
<i>C4A</i>	complement component 4A	6p21.3	補体群の一つ。一般にマクロファージや肝臓で作られ、細胞外に存在し、病原体を攻撃する血中タンパク質で、病原体やこれと結合した抗体によって、活性化される。	Chariem et al. 1988
<i>DEFB</i>	β -defensin	8p23.1	テイクエンジン群。一般に、粘膜組織や皮膚の微生物感染に反応して産生される、自然免疫システムを担うペプチドである。	Hollox et al. 2003
<i>AMY1A</i>	amylase	1p21.1	唾液や汗液に含まれる消化酵素。テングソウ中のグリコシド結合を加水分解して、グルコースなどの糖に分解する。	Groot et al. 1991
<i>GOLGA6</i>	Golgin subfamily A member	15q24.1	精細胞の分化に関与すると考えられている。	Gilles et al. 2000
<i>CHRNA7</i>	$\alpha 7$ -nicotinic acetylcholine receptor	15q13-q14	シナプスでの信号伝達を仲介する、イオンチャンネル。	Pilay et al. 2002

コピー数が増幅する「分節重複」に対して、コピー数が減少するものは「欠失変異」(deletion variant)といわれる。McCarrollらの研究グループが行った、HapMap ProjectのSNPデータを用いた解析では、全ゲノム中541ヶ所の領域で1~745 kbに及び欠失がみられ、そのうち120ヶ所は両アレルで欠失していた。さらにそのうちの10ヶ所はステロイド代謝や嗅覚、薬物代謝に関連した遺伝子の少なくとも一つのエクソンが欠失しており、その領域は個体間で比較的共通していた (McCarroll et al. 2006)。また、Conradら (2006) は0.3~1200 kbの新たな欠失変異を発見した。これらはメンデルの法則に従わない遺伝領域としてみいだされ、遺伝子はほとんど存在していない (Conrad et al. 2006)。おそらくそのために、ゲノム領域が欠失しても、淘汰されず保たれている可能性が高い。このことは、分節重複がCNVの大半を構成していることと無関係ではないだろう。それでも92遺伝子は完全に、201遺伝子は部分的に、コード領域が欠失していた。これらのうち23遺伝子は疾患に関連するものとしてOMIM(Online Mendelian Inheritance in Man: 遺伝性疾患のオンライン版データベース)にも掲載されている。さらに、Hindsらがオリゴヌクレオチドアレイを使って解析したところ、24人のゲノムから、215か所の領域で70~1万塩基対の欠失変異が見いだされた (Hinds et al. 2006)。そのうち41領域の欠失が、10%以上の頻度で存在していた。

多くの研究から、CNVと染色体再編成との間には強い関連があると指摘されている (Aitman et al. 2006; Sharp et al. 2006)。十分長くてよく似た配列が近い位置にあると、非アレル相同組み換えによる染色体再編成が起こり、その結果、欠失、増幅、逆位をもたらす (図 7.2)。実際、重複が生じた分節の両端がAlu配列であるものは分節重複全体の3割にのぼることもわかっている (Babcock et al. 2003; Bailey et al. 2003; Sharp et al. 2006)。このように、非常によく似た配列同士での組み換え反応がCNVを引き起こす引き金になるのである。一方、まったく相同性のない領域が引き金になることがある。たとえば、A型やZ型DNA構造をとっている領域にもCNVが存在していることが見いだされている。これらの構造は、安定なB型と異なり、染色体の不安定性を引き起こしやすいことが知られているので、やはり、ここでも染色体の再編成が生じていると推察される。このように、一次構造の相同性で生じる重複や欠失、そして二次構造の性質によって引き起こされる変異が、進化の過程で、幾度となく生じたことでCNVが成立していった

と考えられている。

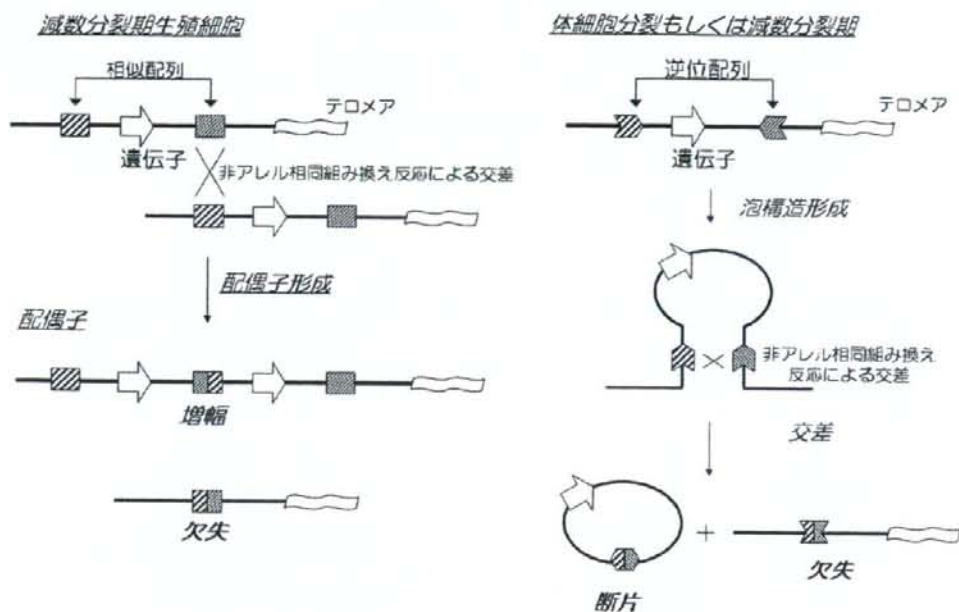


図 7.2 配列の増幅と欠失

ここまでみてきたように、CNV は組換えがベースになって生じた変異である。したがって、CNV と疾患の関連をみようとしたときにもっとも気になりになるのが、CNV とその周辺にある SNP マーカーとの連鎖が保たれているかどうか、ということである (図 7.3)。もし進化の途上で頻発した変異のもとで CNV が成立してきたのであるならば、その周囲の SNP マーカーとはほとんど連鎖不均衡は保たれていないだろう。逆に、進化上たった一回の変異で成立した CNV であれば、きっとその隣にある SNP と連鎖していることが予想される。これらの回答の一つが、上記した McCarroll らや Hinds らの欠失変異の研究で提示されている (Hinds et al. 2006; McCarroll et al. 2006)。多くの場合、欠失変異の周辺にあるタグ SNP をタイピングすることで、欠失領域を同定することができる。このことは、すなわち、連鎖不均衡をもとにした一般的な関連解析を用いることで、欠失変異と疾患との関連を導くことができることを意味する。かといって、それが全ての欠失変異に適用されるわけではない。というのも、異なる家系で比較したとき、欠失領域のつなぎ目にあたる配列が異なる例がある。つまり、これらの変異領域は同じ遺伝

子座にありながら異なるイベントが生じたことを意味し、頻りに組み換えが生じたことが予想されるのである。ここまでみたように実際のところ、SNP と CNV の間の連鎖不平衡度の値は文献によってまちまちである。現時点では、「連鎖不平衡が保たれているかどうかはゲノム領域によって異なる」と説明するのが妥当であろう。これらの確かな値を得ることは、CNV 研究のこれからの課題である。

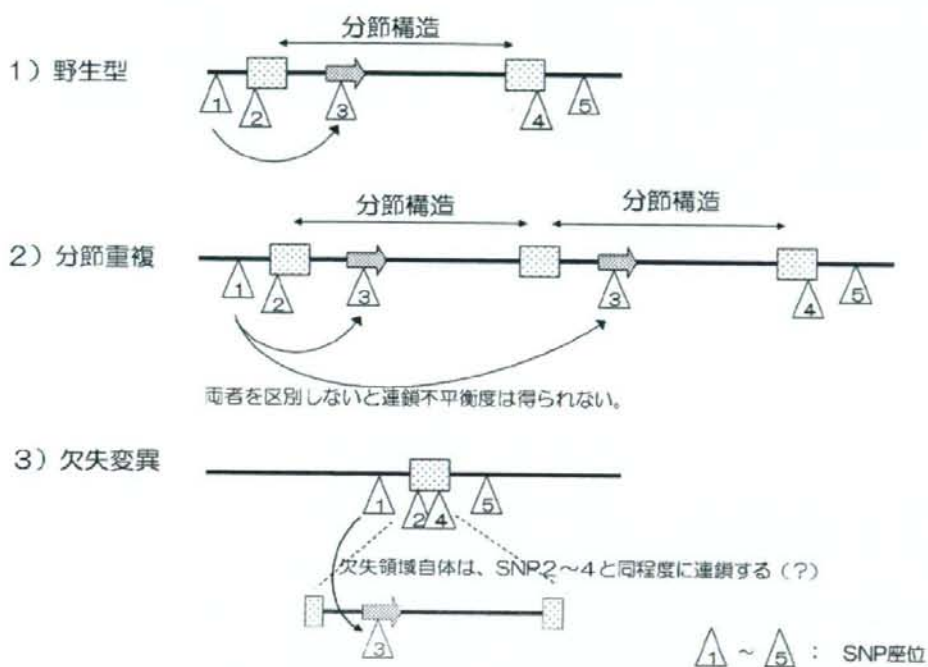


図 7.3 CNV と SNP と連鎖不平衡

7.3 CNV と疾患

2 つの相同染色体間で、ある遺伝子のコピー数が異なる状態を、ハプロ不全 (haploinsufficiency) と呼ぶ。もしたった 1 コピー増えたり減ったりしただけで発現量や機能が乱されてしまうような遺伝子がハプロ不全であれば、成長遅滞や発生異常を引き起こす。遺伝性疾患の多くは、ある遺伝子座が完全に、もしくは一部で欠失したり (例 ファンコニ貧血、ウェルナー症候群)、染色体が丸ごと一本多かったり (例 ダウン症、クラインフェルター症候群) することによって、その遺伝子座にある遺伝子産物の機能を全く喪失したり、生体活動に必須な反応系を阻害したりして、重篤な症状を引き起こす。このような大きなゲノム構造変異だけでなく、CNV による遺伝子発現量のかく乱が疾患に関連している例が、近年続々と報告されてきた。

まずは、1990 年代半ばにみつかった分節重複で少々古典的ではあるが、10 年以上経った今からみると実は CNV と疾患の関係を理解するうえで非常によい例としてあげられるのが、「シャルコ・マリー・トゥース病」(Charcot-Marie-Tooth disease) である。末梢神経症のひとつで、下腿・足をはじめとした四肢遠位筋の萎縮や筋力の低下を主な特徴とする疾患である。病型にいくつかあり、CMT1A タイプは、節性脱髄 (神経線維をくるむ絶縁体としてはたらく髄鞘が脱落すること) が生じるために末梢神経伝導速度が低下することで発症する。欧米では多い病気で、2500 人に一人の割合で罹患し、平均 12 歳で発症する。1995 年に、この疾患に関連したゲノム領域の報告がなされ、CMT1A 患者の「CMT1A-REP 領域」で分節重複が生じていることがわかった (Kiyosawa et al. 1995)。図 7.4 にその模式図を示す。ここには複数の遺伝子が存在していたが、その中で注目されたのが *PMP22* で、その遺伝子産物 PMP22 タンパク質は、髄鞘を構成する分子の一つである。つまり、分節重複でこの遺伝子のコピー数が増加することにより、PMP22 タンパク質の発現が攪乱され、脱髄が生じる。また、別の神経性疾患である、遺伝性圧迫神経麻痺 (HNPP) は、CMT1A-REP 領域とともに *PMP22* が欠失することが原因で発症し、CMT1A よりも軽度の症状がでる。前項で CNV の成り立ちには組み換えが関わっていることを示したが、非アレル相同組み換えによって生じた 2 種類の組み換え体から、それぞれ別の疾患が生じた例でもある。

シャルコーマリートゥース病 (タイプ1A)

健康者

遺伝性圧迫神経麻痺

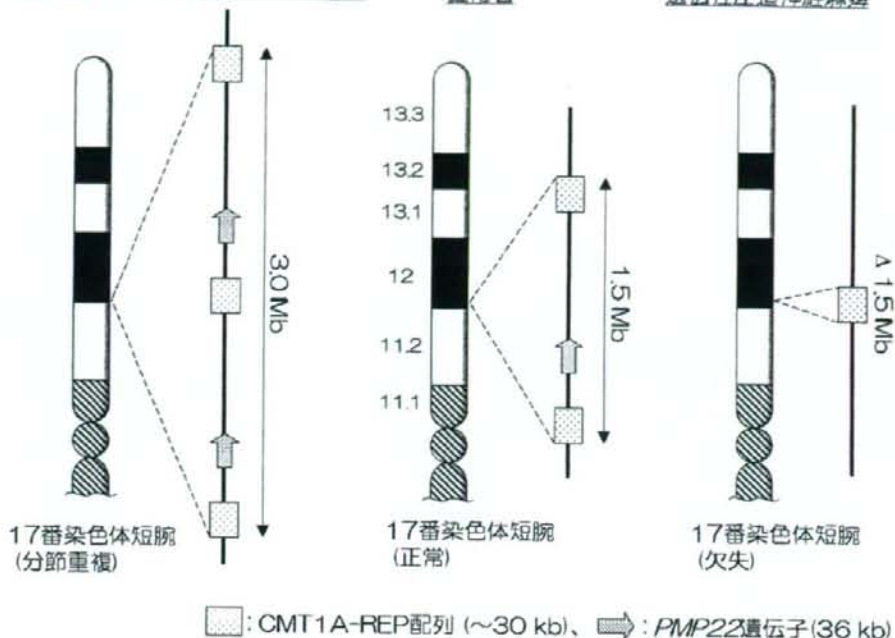
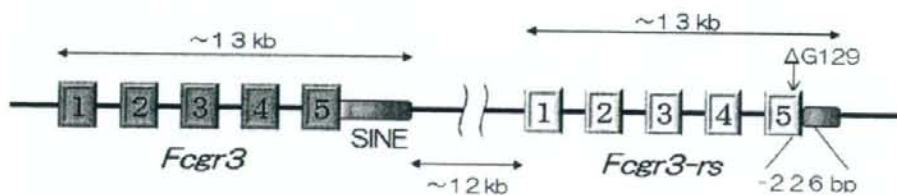


図 7.4 CMT1A 領域の模式図

では、別の疾患の例をみてみよう。糸球体腎炎 (glomerulonephritis) は糸球体の炎症によってタンパク尿や血尿が出る病気で、IgA 腎症、膜性腎症、膜性増殖性腎炎、急速進行型糸球体腎炎など様々な腎炎の総称である。この疾患モデル動物であるラット系統 Whister Kyoto (WKY) は、毒性血清を投与すると 10 日ほどでヒトの糸球体腎炎とよく似た症状を呈する。同じ処理を施してもその他のラット系統 (Lewis など) では異常は見られない。Aitman らはポジショナルクローニングを用いることで、WKY では *Fcgr3* のパラログ遺伝子 *Fcgr3-rs* (*Fcgr3*-related sequence) が欠失していることをみいだした (図 7.5)。そして、このことがマクロファージの過剰活性化とヒトの糸球体腎炎に良く似た症状を引き起こすのである (Aitman et al. 2006)。系統間で共通して存在する *Fcgr3* は、IgG の Fc 断片に対する膜貫通型受容体 (Fc γ receptor III) をコードしていて、活性化したマクロファージの細胞膜表面に局在する分子であるが、WKY 以外のラットがもつ *Fcgr3-rs* は *Fcgr3* の機能を抑制する働きがあることがわかった。それゆえ、WKY では *Fcgr3-rs* による抑制がはたらかない分、免疫反応が促進され、結果、自己免疫反応もあら

われる。ヒトでも、*Fcgr3* ホモログである *FCGR3B* のコピー数は個人により 0~4 コピーの違いがあり、関連解析の結果では、*FCGR3B* のコピー数が少ないほど、全身性エリテマトーデスに絡んだ系球体腎炎の発症率が高いことが明らかになった。

系球体腎炎 抵抗性ラット系統 (Lewis など)



系球体腎炎 感受性ラット系統 (WKY など)



図 7.5 ラットの 13 番染色体上にある、*Fcgr3* - *Fcgr3-rs* 遺伝子座の模式図

最後も欠失変異の例である。補体因子 H 関連遺伝子 (*CFHR*) 群でも、パラログ遺伝子数が欠失することによって、非定型溶血性尿毒症症候群の発症リスクが高まることが報告された(Zipfel et al. 2007)。この遺伝子座には *CFH*、*CFHR3*、*CFHR1*、*CFHR4* 各遺伝子が同じ向きに直列に並んでいる (図 7.6)。*CFH* の 3'配列と *CFHR1* の 5'配列には 98%以上の相同な 28 kb の配列があり、これらは分節重複によって生じたものと考えられている。患者の遺伝子座では 28%でこの相同配列間での組み換えによる欠失変異が起こり、*CFHR1* 遺伝子のコピー数が減少していた。健常者では 6%しかこの欠失変異をもっていない。こういったことから、重複した分節間での欠失によって *CFHR1* 遺伝子が完全に欠失し、*CFHR4* 遺伝子の 5'配列が部分的に削れることが発症率に関連していることが明らかにされた。

遺伝子のコピー数変異が疾患の危険因子になるという知見は、ヒトのありふれた病気に対する感受性をはじめとする遺伝的に複雑な表現型の進化において、ゲ

ノムの可塑性が重要な役割を果たすことを示す直接の証拠となる。このように、CNV は、SNP と並んで、多様な表現型を生み出すゲノム構造変異という地位を得、今後疾患との関連がさらに明らかになるであろう。

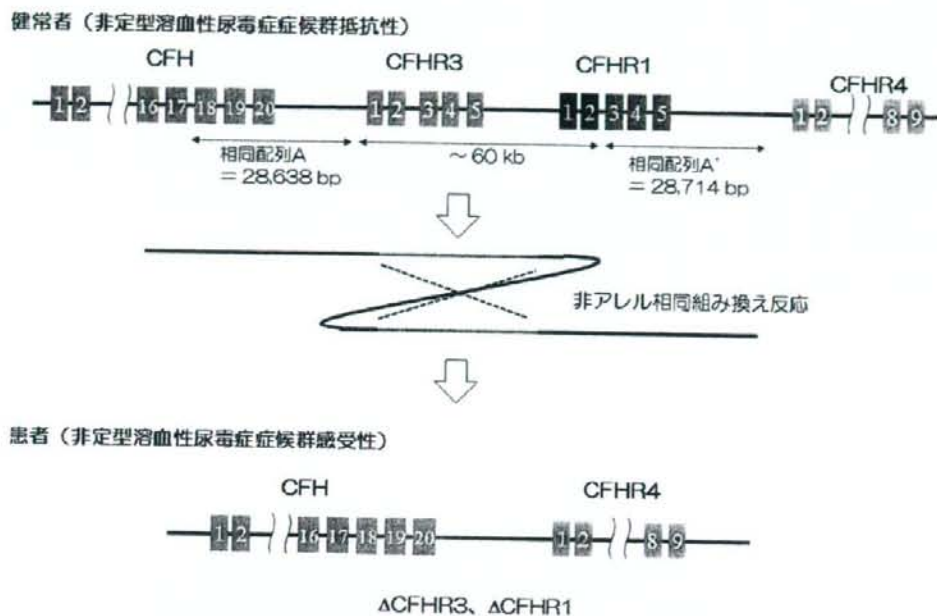


図 7.6 非定型溶血性尿毒症症候群患者でみられる、CFHR 遺伝子座の CNV

7.4 CNV の検出方法

1960 年ごろから細胞核の観察でゲノムの大きな重複と欠失が存在することが報告されはじめた。しかし、解像度などの技術的な問題もあって、特別な遺伝病にみられる大規模な重複・欠失や腫瘍細胞でみられるカリオタイプ異常といった、劇的な構造変化を伴う特別なケースが観察されていた。分裂中期細胞の染色体標本を染色し、光学顕微鏡で観察するカリオタイプ解析では、3~5 Mb という低い解像度しか得られなかったためである。その後、1992 年には、Kallioniemi らが CGH (Comparative Genome Hybridization) 法を発表した (Kallioniemi et al. 1992)。これは Fluorescence in situ hybridization (FISH) 法を応用した手法で、全染色体を対象として DNA が増幅、欠失した領域を調べることができる。感度が不十分なため、1 コピーの違いをみるには適していないが、固形腫瘍で生じたゲノムの染色体異常を解析するときには十分効果的な手法であり、臨床の場においては現在でも利用されている。さらに、ファイバー-FISH 法は、折りたたまれている染色体を、まるでバネを引っ張って 1 本の針金にするように引き伸ばしてスライドガラス上に貼り付けた状態にし、ターゲット配列のプローブをハイブリダイズすると、ターゲット配列の配置を高い解像度で検出できる (Parra and Windle 1993)。

もっとも進歩したプラットフォームとしてはマイクロアレイが挙げられる (図 7.7)。マイクロアレイ上に、レファレンス (基準) DNA と検体 DNA を競合ハイブリさせることにより、両者間のコピー数の違いを検出できる。原理的には CGH と同様の効果を狙ったものであるため、特に アレイ CGH と呼ばれる。2004 年に、BAC (大腸菌人工染色体)ライブラリで構成したアレイをつかって CGH が行われた (Iafrate et al. 2004)。いわゆる BAC アレイ CGH である。この方法は、150 kb 程度のプローブを用いて、ヒトゲノムを 1Mb 間隔で網羅しているので 50 kb 程度の解像度をもっていた。このアレイを使って、5 つの集団で 39 人のゲノムを調べたところ 255 遺伝子座でコピー数の増減がみられた。代表的なものとして 1p13.3 のアミラーゼ遺伝子座があげられる。それに続いてファイバー-FISH 法を用いた解析で、おおよそ 150~425 kb の幅で個人による差異が認められた。

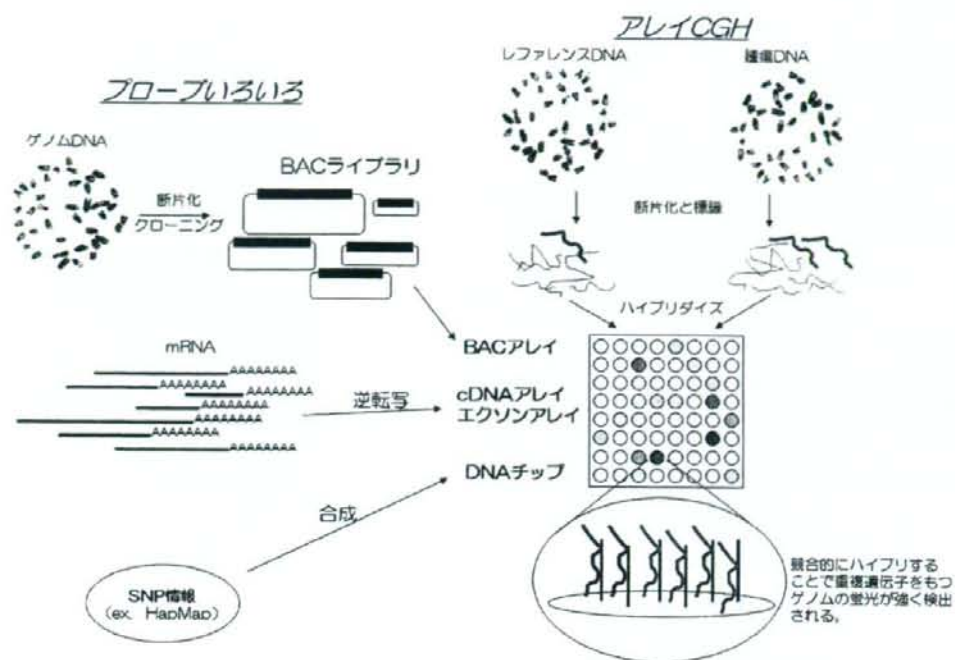


図 7.7 CGH で適用されるマイクロアレイの例

2000 年には Lucito らのグループが、30~70 塩基長の合成オリゴヌクレオチドで、全ゲノムを 35 kb 間隔でカバーしたアレイを開発した(Lucito et al. 2000)。これは Representational Oligonucleotide Microarray Analysis (ROMA) と呼ばれている。彼らは、20 人のゲノムを ROMA で解析し、76 領域、平均 465 kb 長の CNV を検出した。このように、BAC アレイ CGH と ROMA は大量のサンプルをスクリーニングすることができ、手軽でかつ効率的な手法として多くの研究者たちに用いられるようになってきている。

アレイより高解像度のものとして、クローニングした cDNA と一本鎖の PCR 産物をプローブに使ったアレイがある(Dhami et al. 2005; Pollack et al. 1999)。これはエクソン 1 個分の違いまで検出でき、かつ品質が安定していて、高密度に配置でき、カスタマイズも簡単なので、より効果的なツールといえる。ただし、全ゲノムを対象にした解析では、個々人の遺伝子配列の違いにあわせて、プローブを複数設定する必要があるし、全ゲノムをカバーしようとする大きなコストがかかるのが現実である。それに対して、オリゴヌクレオチドを使ったアレイの中でも

SNP タイピングのためにデザインされたマイクロアレイは、CNVを検出するには若干解像度は劣るものの、もっとも手軽といえるだろう (Lucito et al. 2000; Rauch et al. 2004; Sebat et al. 2004)。

このように、ハイスループットでしかもコストパフォーマンスもよいアレイ CGH だが、もっとも注意しなければならないのは、正確にコピー数を測定することはできない、ということである。アレイ CGH では、レファレンスとなる DNA を基準にして増幅があるか欠失しているかをみているので、レファレンスゲノムのコピー数がわからなければ、サンプルゲノムのコピー数もわからない、ということになる。しかも、増幅しているのを検出したと思いきや、じつはレファレンスゲノムで欠失変異が起きているのかもしれないので注意を要する。

もう一つ紹介したいのが、ゲノム全域を網羅した SNP タイピングによる CNV の検出である。先述した SNP タイピング用マイクロアレイによるアレイ CGH と異なるのは、競合ハイブリを行わない点である。そもそも DNA マイクロアレイは核酸定量のために開発されたもので、SNP タイピングのみならず、プローブにハイブリダイズした DNA 量も測定できる。つまり、遺伝子コピー数を測定することができるのである。ゲノム全域の SNP タイピングと同時に CNV 解析を行うこともできるとあって、多くの研究グループで採用されている。製品化されているものでは、Affymetrix 社の GeneChip®、Illumina 社の BeadChip®などが代表的なものである。これらについては、8、9 章で詳細な解説がなされているのでそちらを参照されたい。

これまで様々な CNV 検出方法を紹介してきたが、どれも長短あわせもっている。これが絶対という方法はない。いずれにしろ、多くの検体で複数の手法を組み合わせて検定し、そこで得られた結果については、さらに個々の検体を FISH、Q-PCR、塩基配列決定などで確認することが望ましい。

7.5 利用可能な CNV データベース

現在までに解析された CNV はデータベース化され、インターネットで公開されているものがある。「Database of Genomic Variants」は The Centre for Applied Genomics (TCAG; カナダ)が管理しているデータベースであり、健常者もつ代表的なゲノム構造変異を統括している。2004年に Iafrate ら、2005年に Sharp らがアレイ CGH で、2004年に Sebat らが ROMA で、2005年に Tuzun らがフォスミド末端配列解析で、それぞれ得られたデータが取りまとめられたのを皮切りに、最近では DNA チップによるデータも続々と追加されている (Iafrate et al. 2004; Redon et al. 2006; Sebat et al. 2004; Sharp et al. 2005; Tuzun et al. 2005)。ウェブサイトの表紙には、ヒト染色体のイラストが掲載されていて、その染色体上にある CNV や分節重複がリストになって表示されている (図 7.8)。検索ボックスに遺伝子名またはゲノム塩基配列を入力することで、その領域もしくは周辺に CNV が存在するかどうかが表示される。このデータベースは「国際ハップマップ計画 (International HapMap Project)」のウェブサイトにもリンクしているので、同サイトの膨大な SNP 情報とあわせて閲覧することができるようになっている。

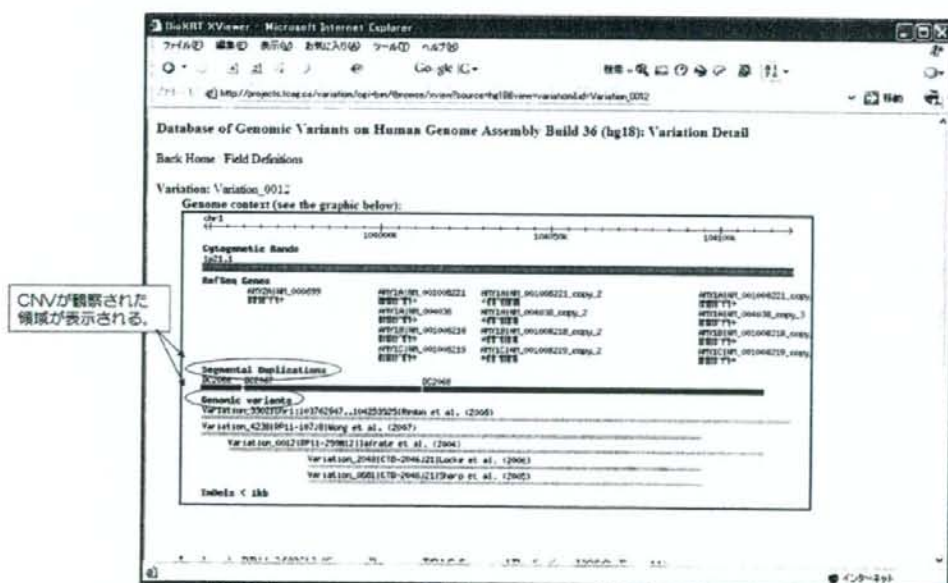


図 7.8 「Database of Genomic Variants」ウェブサイトであミラーゼ遺伝子座を検索した結果

ヒト集団の CNV の網羅的な探索を目的として、Wellcome Trust Sanger Institute (イギリス) が中心となり、「Copy Number Variation Project」という研究プロジェクトが結成された。CNV 解析のプラットフォームとして、WGTP (Whole Genome Tiling Path) array と Affymetrix 500K アレイが採用されている。HapMap に登録されている 270 人のゲノムについて CNV のデータがカタログ化され、Wellcome Trust Sanger Institute のウェブサイトで閲覧・ダウンロードできる。また、同じく Wellcome Trust Sanger Institute から公開されている「DECIPHER」は、CNV の情報と臨床情報を対応させているところが特徴である。将来的に、遺伝性疾患などで アレイ CGH を用いた医療診断を行なう際の手助けになることを期待されている。

「Human Structural Variation Database」も、CNV を含めた中～大規模の構造変異のデータを掲載したサイトであるが、塩基配列レベルまで調査しているのが特徴である。2005 年に Tuzun らが報告したフォスミド末端配列解析をベースにしてデータを収集している。カリフォルニア大学サンタクルズ校で管理されている「UCSC Human Genome Browser」の染色体マップ上でも閲覧することができる。

CNV のデータベースは、様々なプラットフォームを用いたデータを統合したものであるため、それぞれの検出法の解像度によって、CNV の長さには大きな幅がある。データベース間相互の照合や、将来的な技術向上により精度の高い CNV データベースが構築されるだろう。

7.6 最後に

これからは、疾患関連遺伝子を探索することを念頭において、ゲノム構造変異の正確なマップを作成し、全領域にわたってデータベース化することが将来の目標となる。これから公表される研究成果によってデータベースの更新が随時行われ、ますます多くの情報が蓄積されていくであろう。しかしながら、さまざまな研究グループから報告されるデータは、偽陽性があるいは信頼性の低いものも含まれる可能性がでてくる。そのような問題に対処するためにも、複数の研究室で相互にデータの評価をするなどデータを吟味しつつデータベースを構築する必要がある。

構造変異のある領域、特に挿入や欠失のジャンクション部位の塩基配列を決定することも重要である。構造変異が生じた部位がどのような組み換え反応によるものかを明らかにできるからである。また、構造変異の進化的な成り立ちもより詳細に知ることができるはずである。さらには、多型として存在しているのか、ということの証拠となるかもしれない。だからこそ、変異領域を同定するだけでなく、その塩基配列決定が求められている。

もしかすると、一見、非コード領域にある構造変異が、実は microRNA やその他非コード RNA などのコピー数に変化を及ぼしていたり、プロモーターやエンハンサー配列を改変したりしていた場合、見逃してはならない影響が生じることが予想される。

CNV は、SNP と肩を並べる、複雑な表現型をつかさどるゲノム情報として、そして、ヒトのみならず多生物種のゲノムを形づくってきた進化の推進力として、医学のみならず生命科学において重要な役目を担うこととなるであろう。

8. 高密度 SNP タイピングアレイを用いたゲノム全域解析による疾患遺伝子解析

8.1 はじめに

最近のゲノム科学分野の技術革新は目覚しく、特に 2003 年 4 月のヒトゲノム解読完了以降、シーケンス技術の革新、DNA マイクロアレイの高密度化などその動きは更に加速している。それに伴い、特にヒトゲノム配列情報の蓄積が進み、一千万以上の SNP（一塩基多型）の存在が明らかとなった。これらの大量の遺伝マーカーの利用が、DNA マイクロアレイの高密度化により技術的には比較的容易になり、アレイの低価格化とあわせて、現実的に利用が可能となってきている。また、SNP タイピング用に開発された高密度 DNA マイクロアレイが、ヒトゲノムの構造解析にも使用可能なことが示され、ゲノムコピーナンバーの解析にも応用され始めている。本章では、SNP タイピングアレイ用の高密度 DNA マイクロアレイを用いた実験とデータ解析について、主に現在使用している Affymetrix 社製の GeneChip system を中心として述べる。

8.2 高密度 SNP タイピングアレイ実験

GeneChip Mapping Array を用いた SNP タイピングアレイ実験操作については、製品の User Guide に詳しい記述があり、Affymetrix 社のホームページから入手可能である (<http://www.affymetrix.com/>)。導入に当たっては、Affymetrix 社のトレーニングを受けることとなっており、導入後のサポート体制も整っている。以下簡単に実験工程を解説する (図 8.1)。

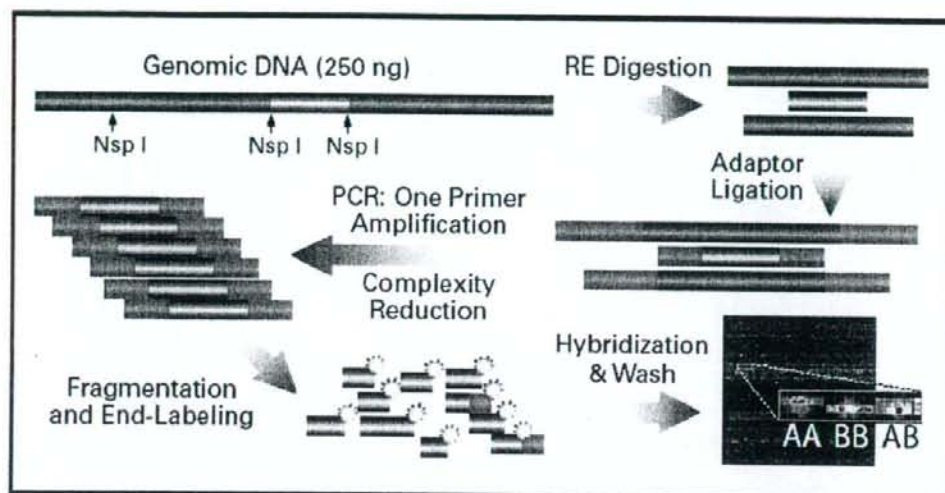


図 8.1 GeneChip Mapping Array を用いた SNP タイピングアレイ実験工程

タイピングアレイ実験は、大まかに 3 つの段階に分けることができる。1) ターゲット配列の増幅、末端ラベル。2) アレイとターゲット配列のハイブリダイゼーション。3) アレイの洗浄とイメージスキャン。このうち 3) の工程は専用の洗浄機とスキャナーが行うので、実験者が主に行う作業としては、1) と 2) である。500K Array Set 以降の GeneChip Mapping Array では、High Throughput Protocol と称して 96 well plate で多検体を同時処理するプロトコールが標準となっている。SNPs タイピング、コピーナンバー解析のデータ解析時に多検体からのデータを標準化して処理する手法が主流となりつつある。均一なデータ取得が望ましいことから、出来るだけ同一な条件で反応を進める必要があり、High Throughput Protocol に従うべきだろう。

8.2.1 ターゲット配列の増幅、末端ラベル

GeneChip Mapping Array では、whole-genome sampling analysis (WGSA) という手法を用いて、ゲノムの complexity を下げる工夫を行っている。

すなわち、ゲノム DNA を制限酵素 (500K Array Set 以降、最新の SNP6.0 までは、*SlyI*, *NspI*) で消化後、断片化されたゲノム DNA 両端にアダプターを付加する。続いてアダプター配列と共通配列を持つプライマーを用いて PCR を行う。PCR は、1.5kb 程度までの DNA 断片を増幅するように条件設定が行われているので、この範囲の DNA 断片のみが選択的に増幅される。

増幅領域外の DNA 断片は実質的に解析対象外となるため、30 億塩基対のヒトゲノムが 5 億塩基対程度までに単純化されるこの手法では、解析が不可能なゲノム領域が存在することは、注意する必要がある (Kennedy et al. 2003)。25bp のアレイ上プローブへのターゲット配列の効率的なハイブリダイゼーションのために、PCR 産物は DNaseI の消化を受け、更に断片化され、terminal deoxynucleotidyl transferase により、末端にピオチンが導入される。以上の工程は、ピペッティング操作に慣れた実験者であれば、特に難しいところは無い。User Guide には各工程でのチェックポイントが詳細に記載されているので、これに従って、確認作業を行えばよい。特に注意を払うとすれば、DNA の濃度、DNA 溶解バッファー、DNA のクオリティーであるが、これらについても User Guide に従えばよい。

8.2.2 アレイとターゲット配列のハイブリダイゼーション

この工程も特に難しいところは無い。遺伝子発現アレイ等で GeneChip を使用した経験があれば、ほぼ同様の操作である。ただし、Mapping Array では Expression Array と異なり、pre-hybridization のステップがない。更に搭載コンテンツの増大 (最新の SNPArray6.0 では、約 90 万 SNPs タイピングプローブと約 95 万のコピー数プローブが搭載されている) により、プローブ径の微小化とターゲット DNA の高濃度化を余儀なくされている (表 8.1)。詳細については 9-3. データ解析 (SNP タイピング) で述べるが、これにより実験条件としてはより厳密さが求められることとなった。核酸のハイブリダイゼーションは、溶液中の核酸濃度、塩濃度、

温度などの影響を受ける。これらのうち核酸濃度、塩濃度は、User Guide に規定されている。温度についても同様であるが、User Guide ではチップへの核酸溶液注入時、チップは室温への平衡化が求められているのみで、ハイブリダイゼーション温度への平衡化については記載がない。我々は、核酸溶液注入前に恒温槽、ホットプレートを使用してあらかじめハイブリダイゼーション温度への平衡化を行っている。また、ハイブリダイゼーション終了時のチップからの核酸溶液拔出時にもチップの温度を低下させないように注意を払っている。これらは実験結果を安定させるためのコツと考えている。

8.2.3 アレイの洗浄とイメージスキャン

この工程は、主に専用の洗浄機とスキャナーが行う。洗浄機のメンテナンスを定期的に行い、機械の不具合やスキャン画像の不整（図 8.2）があった場合には、Affymetrix 社のサポートを受ける。

A



B



図 8.2 機械の不具合やスキャン画像の不整の例

8.3 データ解析 (SNP タイピング)

スキャン画像の各プローブのシグナル強度から SNP の遺伝子型を決定する。Mapping 500K array set では、発現アレイの設計方法を踏襲したためか 10K、100K と同じく、図 8.3 の様にタイリングプローブを用いている。

500K Probe Array Tiling

1	2	3	4	5	6	7
PM A	PMA	PMA	PMA	PMA	PMA	PMA
MM A	MM A	MM A	MM A	MM A	MM A	MM A
PM B	PMB	PMB	PMB	PMB	PMB	PMB
MM B	MM B	MM B	MM B	MM B	MM B	MM B

Quartet

- 6 quartets interrogate each SNP either the Forward and/or on the Reverse strand. Some probe sets have all probes on one strand.
- 24 probes total are used to interrogate a single SNP

図 8.3 500K プローブアレイの構成

ただし、発現アレイと違い、マッピングアレイでは、25bp のプローブ長のほぼ中央に SNP を配する必要があるため、タイリングプローブ間の設定場所の違いはせいぜい+/-4b であり、ほぼ同様の配列を持つ。ほぼ同様であるが、実際には配列が全て異なる点が、データ解析上悩ましい。Mapping 500K array set 後の SNP5.0、SNP6.0 では、タイリングプローブを止め 1SNP、1probe set に変更している。Mapping 500K array set のタイピングに話を戻すと、1SNP をタイピングするのに 6カルテット、24 プローブ(一部 40 プローブ)を用いる(図 8.3)。6カルテットを用いるという冗長性のため、スキャン画像の不整(図 8.2)にもある程度頑強な実験系となっている。

著者は現在 SNP5.0 を使用しているが、SNP5.0 では、冗長性が 4、SNP6.0 では、冗長性が 3 と少なくなっているため、スキャン画像の不整に対しては、脆弱になっ

ているという感触は否めない。Mapping 500K array set では、当初シグナル強度から遺伝子型を決定する際に Mapping 100K array set から導入された Dynamic Model Mapping (DM) algorithm を用いていた。詳細については User Guide に譲るが、このアルゴリズムは、シグナル検出用の perfect match probe signal とバックグラウンド検出用の mismatch probe signal の強弱によって判定する手法である。解析精度が実験の S/N 比に直接的な影響を受けるので、実験のクオリティーコントロール (QC) には適したアルゴリズムであると考ええる。

ただし、ホモ接合体の場合に比べてヘテロ接合体のシグナルが弱くなるため、SNPs タイピング用のアルゴリズムとしては、ヘテロコールが難しいという弱点を抱えている。この弱点を克服できなかったためか、Mapping 500K array set では、途中から BRLMM と呼ばれるアルゴリズムを採用し、以降 SNP5.0、SNP6.0 では、DM は QC 目的のみに一部採用されるのみで、BRLMM が採用されている。

BRLMM は、多検体データを標準化、分布モデルでのスコア化後、遺伝子型によるシグナル強度のクラスターでタイピングを行う手法であり、perfect match probe signal のみを用いている。BRLMM を用いる場合、多検体データの標準化という操作がはじめに入ることには十分注意を払う必要がある。同時に解析するのは、可能な限り均質なデータとする。また、DM でのコールが基準以下のもの (Mapping 500K array set では、DM アルゴリズム閾値 0.33 でコール率 93%未滿)、すなわち実験の S/N 比が基準以下のデータは除外すること。実験の S/N 比が基準以下のデータを含めると、全体のデータがゆがみ、タイピング結果の信頼性が損なわれることが容易に考えられる。