表 6.1　MDR の入力ファイルのデータ構成

SNP1	SNP2		SNP9	SNP10	P_C
0	1	…	1	2	1
1	1		1	1	1
⋮		⋱		⋮	⋮
0	0		0	1	0
2	2	…	1	1	0

行が被験者、列が SNP および表現型であり、表中の数値は、各被験者の当該 SNP における遺伝子型（0 がメジャーアレルのホモ、1 がヘテロ、2 がマイナーアレルのホモ）を示す。ただし表現型の場合のみ、0 が健常者、1 が罹患者を表す。列と列の間はタブで区切る。また、列名（SNP および表現型）も必須であることに注意されたい。

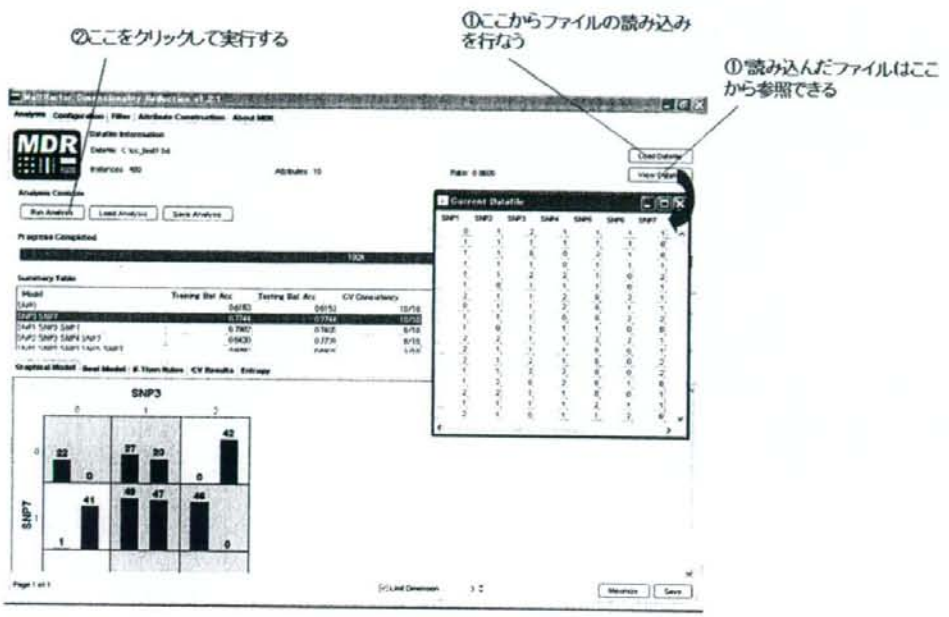

図 6.3　ソフトウェア MDR を用いた遺伝子間相互作用解析

図の結果は、10 セットの SNP のうち、主に SNP3 と 7 が発症リスクに関与している可能性が高いことを示している。

6.3.2 再帰分割法

CART（classification and regression trees）は、データを構成する変数を基準に、段階的なデータのクラス分けを行い、ある現象を説明する最適なモデルを決定するための手法である。基準となる変数の決定にはジニ係数、およびその加重平均が用いられ、より小さな値を与える変数を優先的に用いてデータを分割していく。図 6.4 に示したデータの場合、SNP1 と 2 をそれぞれ基準としてデータを分割すると、前者におけるジニ係数の加重平均は 0.167、後者の場合は 0.476 となり、発症しているかどうかをよく説明するのは SNP1 であることが分かる。したがって、ここではまず SNP1 を、次に SNP2 を基準としてデータの分割が行なわれる。

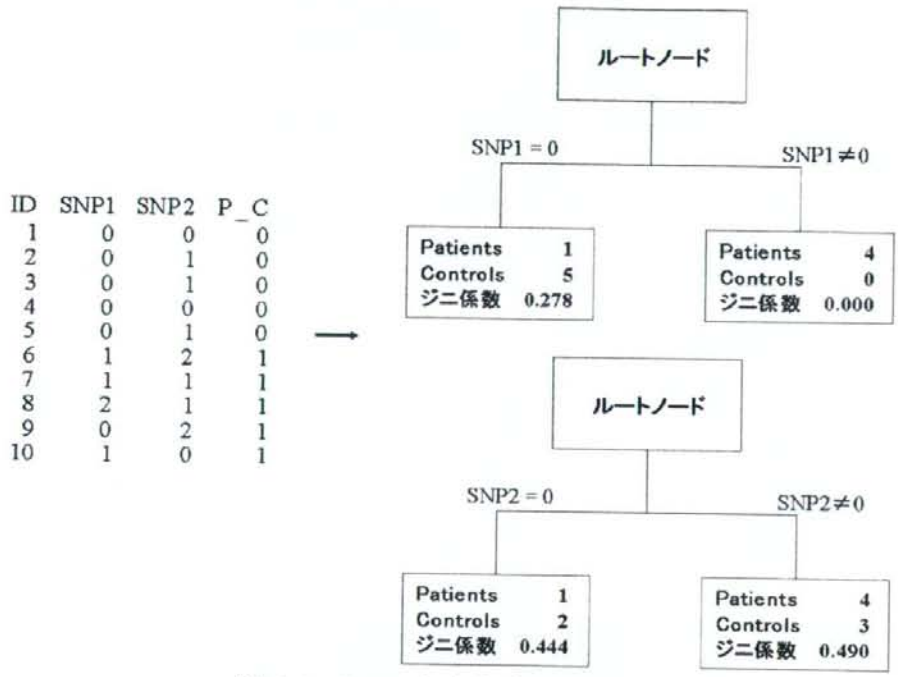

図 6.4　CART によるデータの分割

SNP1 と 2 それぞれについてデータを分割すると、前者におけるジニ係数の加重平均は

$\frac{0.278 \times (1+5)}{10} + \frac{0.000 \times (4+0)}{10} = 0.167$、後者の場合は$\frac{0.444 \times (1+2)}{10} + \frac{0.490 \times (4+3)}{10} = 0.476$

となり、発症しているかどうかをよく説明するのは SNP1 であることが分かる。したがって、

ここではまず SNP1 についてのデータの分割が行なわれる。

例として挙げたデータでは SNP が 2 セットのみであるため、データの分割はここで終了するが、これより大規模なデータの場合は、サンプルが 1 となる、またはすべてのサンプルがまったく同じクラスに属する状態まで、データの分割を繰り返す。ただしその後は、交差検定により、有意でないものについては再度データを統合する、すなわちプルーニング（剪定）を行ない、最終的なモデルとする。

MARS（multivariate adaptive regression splines; 多変量適応的回帰スプライン）は、CART に重回帰分析の理論を組み合わせたものであり、CART で扱われる変数が不連続なカテゴリ変数のみであるのに対し、連続変数が含まれるデータにも適用が可能である。Cook ら（2004）は、CART および MARS を用いて、虚血性脳

梗塞に関する遺伝子間相互作用の同定を試みている。

またランダム森（Breiman 2001）は、大規模な数の変数を扱えるように改良されたアルゴリズムであり、ゲノムワイドなデータの解析に向け、有効な手法の一つとなり得る。具体的なアルゴリズムは以下の通りである。

1）学習用データとして、N名の被験者の記録から、同じ数のブートストラップサンプルを復元抽出する。

2）KセットのSNPから、k（$<K$）セットをランダムに抽出する。

3）抽出したkセットのSNPから、疾患に関する表現型を最もよく説明するSNPの遺伝子型をもとに、1）で得たブートストラップサンプルを分割する。

4）2）～3）を繰り返してデータを分割し、決定木を作成する。なお、プルーニングは行なわない。

5）oob（out-of-bag）データ（ブートストラップサンプルとして抽出されなかったデータ、理論上では約36%のサンプルが該当する）を検証用データとした交差検定を行なう。

6）1）～5）をM回繰り返し、作成された決定木すべてについての交差検定の結果を集計して、各変数の重要度を求める。なお、重要度の計算には、ジニ係数を用いることも可能である。

ランダム森の最大の特徴は、データの分割に用いる変数を、すべての変数からではなく、あらかじめランダムに抽出されたサブグループの中からで、これによって、性急な結論を避けると同時に、大規模なデータからの情報を最大限に活用することが可能となる。

再帰分割法は、遺伝学のみならず、あらゆる分野で広く用いられており、市販のソフトウェアも多く存在するが、Rでも実行が可能である。ここではランダム森の実行方法について説明する。

R でランダム森を実行するには、まずコンソールウィンドウで

```
> library(randomForest)
```

と入力し、パッケージ randomForest を呼び出す。次に、

```
> dat1 <- read.table("C:/cc_test1.dat", header = T)
```

とすれば、C ドライブのテキストファイル cc_test1.dat からデータが変数 dat1 に読み込まれる。なお、cc_test1.dat の構造は以下に示すように、行が被験者、列が SNP および表現型であり、表現型が文字列であること、空白がスペースで区切られていることを除けば、MDR の場合と同じである。

表 6.2　cc_test1.dat ファイルのデータ構成

SNP1	SNP2	…	SNP9	SNP10	P_C
0	1	…	1	2	Patient
1	1		1	1	Patient
⋮		⋱		⋮	⋮
0	0		0	1	Control
2	2	…	1	1	Control

データを読み込んだ後、続いて

```
> result <- randomForest(P_C ~ ., data = dat1, importance = TRUE)
```

と入力すればランダム森が実行され、結果が変数 result に書き出される。この結果を参照したい場合には、

```
> varImpPlot(result)
```

と入力すれば、図 6.5 のような結果が出力される。この結果から、10 セットの SNP のうち、発症リスクに関与しているのは、主に 3 と 7 であることが推察される。

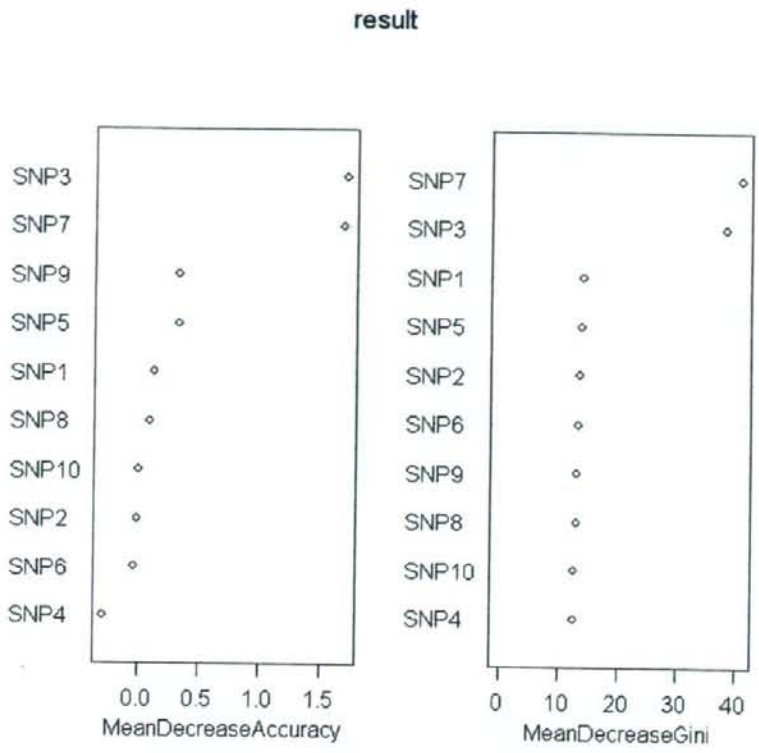

図 6.5　R でのランダム森の実行

左の図は oob データを用いた交差検定の結果から、右はジニ係数から計算された変数の重要度を示す。いずれも、SNP3 と 7 が発症リスクに関与している可能性が高いことを示している。

ランダム森については、提唱者である Dr. Breiman のサイト（http://www.stat.berkeley.edu/users/breiman/RandomForests/）でもフリーのソフトウェアが公開されているが、利用にはやや専門的な知識と技術を要するため、説明は省略する。

6.3.3 セット関連解析法

セット関連解析法（Hoh et al. 2001）は、「次元の呪い」からの回避と、検出力の向上を目的として開発、提唱された手法であり、そのアルゴリズムは以下の通りである。

1）K セットの SNP それぞれについて、疾患との関連に関する検定統計量 $t_i (i = 1, 2, \ldots, K)$ と、症例集団における Hardy-Weinberg 平衡からのずれの検定統計量 u_i を計算する。ただし、Hardy-Weinberg 平衡からのずれがタイピングエラーに起因している可能性を考慮し、対照集団において、Hardy-Weinberg 平衡からのずれが一定の水準（1%、5%など、ユーザーが設定する）以上である d の SNP については、$u_i = 0$ とする（トリミング）。

2）K セットの SNP それぞれについて $s_i = t_i \times u_i$ を計算し（重み付け）、降順に並べ替える。

3）2）で並べ替えた順に s_i（$i = 1, 2, \ldots, k$、ただし $k < K$）を一つずつ加え（グルーピング）、新たな検定統計量 $S_i(d)$ を求める（$S_i(d) = s_1 + s_2 + \ldots + s_i$）。また、そのつどパーミュテーションテストによって $S_i(d)$ の p 値を計算する。なお、d と同様、k はユーザーが各自で設定する。

4）得られた k の p 値のうち、最小のもの（$min_k p_k$）を与える SNP のリストが、疾患に関与する SNP の組み合わせとして最も「それらしい」と推定された結果である。

5）パーミュテーションテストにより、ランダムに並べ替えたデータセットに対して1）～4）を繰り返し、得られた $min_k p_k$ の分布から、オリジナルなデータにおける検定統計量の最終的な p 値（p_{min}）を求める。

セット関連解析法は、提唱者の意図を非常によく反映しており、「次元の呪い」に対して頑健であると同時に、高い検出力を誇る手法である反面、どの SNP 間に有意な相互作用が存在するかなどを特定できない、単独で有意な関連を示さない SNP が除外されてしまう、また連鎖不平衡の関係にある SNP をどう取り扱うかといった問題を有しており、他の手法と比較して認知度はやや低い。

この手法についても、Dr. Ott のサイト（http://www.genemapping.cn/）でフリーのソフトウェアが公開されている。実行ファイルは Windows 用のみであるが、ソー

スファイルをコンパイルすれば Linux でも利用が可能である。実行ファイルは sumstat.exe、および statpval.exe の二つに分けられており、前者は上記の 1)〜4)を、後者が 5)を実行する。利用者は、行を SNP および表現型、列を被験者とするテキストファイルを用意すればよい（行と列、および被験者と SNP/表現型の関係が、MDR などの場合と逆であることに注意されたい。また、SNP 名などは不要である）。そのほか、d の設定や、パーミュテーションテストにおける繰り返しの回数など、様々なオプションの選択が可能である。詳細は上記のサイトを参照されたい。

6.4 ニューラルネットワーク

ニューラルネットワークは、脳内のニューロン（神経細胞）が、他のニューロンから樹状突起を介して信号を受け取り、シナプスを通じて、その信号をまた別のニューロンへ伝えるしくみをコンピューター上で再現し、それによって最適解を得るための手法である。

ニューラルネットワークと言えば多くの場合、図 6.6 に示したような多層パーセプトロンを想定し、バックプロパゲーション（出力と実際の値との二乗誤差を、その偏導関数を用いて最小化し、パラメーターである重み付け値を推定する手法。日本語では誤差逆伝播法と呼ばれる。）を用いて最適解を得るアプローチを指すが、Ritchie ら（2003）は、遺伝的プログラミングを用いた GPNN（genetic programming neural network）を提唱している。シミュレーション実験の結果、バックプロパゲーションを用いた場合と比較して、GPNN は正しい解に到達する割合が高い（Ritchie et al. 2003）一方、高次元相互作用の検出力が低いことが指摘されている（Motsinger et al. 2006）。

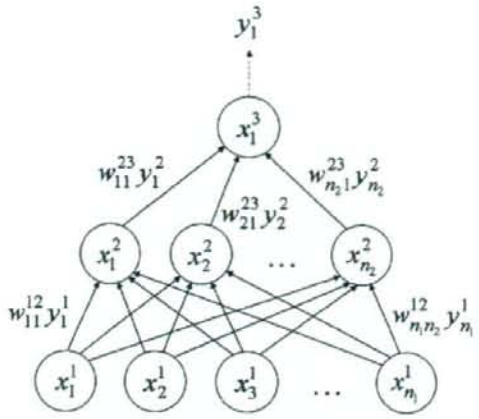

図 6.6　ニューラルネットワーク

N 層のネットワークを仮定し、i 層 j 番目のユニットへの総入力を x_j^i、このユニットからの出力を y_j^i、$i-1$ 層 j_1 番目のユニットから i 層 j_2 番目のユニットへの重み付け値を $w_{j_1 j_2}^{i-1,i}$、また $y_j^i = \frac{1}{1+e^{-x_j^i}}$、$x_k^i = \sum_j w_{jk}^{i-1,i} y_j^{i-1}$ とする。ここで、SNP の遺伝子型 x_j^1 と、疾患に関する表現型 t_k が与えられたとき、最適な重み付け値は、最上層のユニットからの出力 y_k^N と t_k との二乗誤差 $E = \sum_k \left(y_k^N - t_k\right)^2$ が最小となるように求められる。一般的には、各重み付け値を E の偏導関数（$\Delta w_{jk}^{i-1,i} = -\eta \frac{\partial E}{\partial w_{jk}^{i-1,i}}$）で繰り返し修正しながら、解を収束させるアルゴリズム（バックプロパゲーション）が用いられる。

また Tomita ら（2004）は、最適なモデルの推定を目的として、PDM（parameter decreasing method）なるアルゴリズムを取り入れた手法を提唱している。そこでは、すべての SNP を含めたモデルを起点として、ニューラルネットワークの構築と、交差検定によるモデルの評価を、SNP を一つずつモデルから除去しながら行ない、最も高い精度を与える SNP の組み合わせを、最適なモデルと見なす。

ニューラルネットワークもまたあらゆる分野で広く用いられており、無論 R でも実行が可能であるが、遺伝子間相互作用の解析を目的としたフリーのソフトウェアは、上述の PDM を提唱したグループのサイト（http://www.nubio.nagoya-u.ac.jp/proc/english/indexe.htm）で公開されている。

GPNN については現在のところ、ソフトウェアなどは公開されていない。

6.5 グラフィカルモデリング

グラフィカルモデリングは、条件付き独立性を基本的概念とし、多変量データの関連構造を表す統計モデルを、ネットワークグラフによって表す手法である。遺伝子間相互作用解析の場合、その一種であるグラフィカル対数線形モデリングが用いられる。

三つの変数 A、B、および C で構成される以下のデータにおいて、変数 A が与えられた場合の、B と C の独立性に関する検定を行なうものとする。

表 6.3　変数 A, B, C で構成されるデータ

	A_1		A_2	
	B_1	B_2	B_1	B_2
C_1	n_{111}	n_{121}	n_{211}	n_{221}
C_2	n_{112}	n_{122}	n_{212}	n_{222}

ここから、B、C、および B と C についての分割表を作成する。

表 6.4　A と C で集計した分割表

	A_1	A_2
C_1	$n_{1.1}$	$n_{2.1}$
C_2	$n_{1.2}$	$n_{2.2}$

表 6.5　A と B で集計した分割表

A_1		A_2	
B_1	B_2	B_1	B_2
$n_{11.}$	$n_{12.}$	$n_{21.}$	$n_{22.}$

表 6.6　A で集計した分割表

A_1	A_2
$n_{1..}$	$n_{2..}$

ここから、B と C が互いに独立である、すなわち関連はないとした場合の各クラスにおける度数の期待値は、以下の式（1）によって得られる。

$$m_{ijk} = \frac{n_{ij\cdot} \cdot n_{i\cdot k}}{n_{i\cdot\cdot}} \qquad (1)$$

これらの期待値と、実際の度数との差に関する統計量は逸脱度（式（2））と呼ばれ、近似的に χ^2 分布にしたがう。これに基づいた検定の結果、逸脱度が有意であれば、B と C は互いに独立でなく、両者をつなぐ辺を除くことは適切でないと判断される。

$$G^2 = 2\sum_i \sum_j \sum_k \ln \frac{n_{ijk}}{m_{ijk}} \qquad (2)$$

具体例として、以下の分割表に示すようなデータにおいて、SNP2 が与えられた場合の、変数 SNP1 と疾患との関連解析を行なうものとする。

表 6.7　SNP1,2 と疾患の分割表の具体例

SNP2	BB		Bb or bb	
SNP1	AA	Aa or aa	AA	Aa or aa
Patients	24	10	25	41
Controls	27	37	25	11

次に、SNP1、疾患、および両者についての分割表を作成する。

表 6.8　SNP2 と疾患で集計した分割表

SNP2	BB	Bb or bb
Patients	34	66
Controls	64	36

表 6.9　SNP1 と SNP2 で集計した分割表

SNP2	BB		Bb or bb	
SNP1	AA	Aa or aa	AA	Aa or aa
	51	47	50	52

表 6.10　SNP2 で集計した分割表

SNP2	BB	Bb or bb
	98	102

ここから、各クラスに属する被験者の数の期待値は以下の通りとなる。

表 6.11　各クラスの被験者数の期待値

SNP2	BB		Bb or bb	
SNP1	AA	Aa or aa	AA	Aa or aa
Patients	17.694	16.306	32.353	33.647
Controls	33.306	30.694	17.647	18.353

したがって逸脱度は、$G^2 = 2\times\left[24\times\ln\frac{24}{17.694}+27\times\ln\frac{27}{33.306}+\cdots+11\times\ln\frac{11}{18.353}\right]=16.814$

となる。自由度は(2 - 1) × (2 - 1) × (2 - 1) = 1 であり、p 値は 4.0×10^{-5} 以下、すなわち、SNP2 が与えられた場合、SNP1 と疾患との間には、非常に強い関連を示唆している。

グラフィカルモデリングは現在までに、複数の座位の連鎖不平衡解析（Thomas and Camp 2004）や、子宮頸ガンのデータの解析（Horng et al. 2004）に応用されているが、遺伝子間相互作用解析に応用した例は、他の手法と比較してまだ少ない。しかし、「次元の呪い」とともに、研究者を苦しめてきた多重共線性の問題が、このアプローチによって解決されるという点は非常に意義深いと言える。ゲノムワイドデータへの応用を視野に入れた検討（Verzilli et al. 2006）もすでに始まっており、新たな可能性を秘めたアプローチとして、今後さらなる検討が期待される。

6.6 遺伝子間相互作用解析法の課題と今後の展望

以上述べてきたように、ここ数年で様々な遺伝子間相互作用解析法が提唱されると同時に、このテーマの重要性も広く認知されてきており、研究者の努力がようやく実を結びつつある。しかしながら、ゲノムワイド関連解析など、非常に規模の大きなデータを対象とした研究が現実のものとなった現在、「次元の呪い」は再び研究者の前に大きな壁として立ちはだかる問題である。ランダム森やセット関連解析法は、大規模なデータの解析に適しており、「呪い」に対しても頑健ではあるが、その場合、最適なモデルが推定されても、モデルを構成する変数の関連構造を明らかにすることは断念せざるを得ない。

また、今後は遺伝子間相互作用解析にとどまらず、図 6.7 に示すように、非遺伝的因子なども同時に考慮した、生物学的なメカニズムにより近いモデルによるアプローチが求められるであろう。そのためには、臨床検査データなどを含む、質的・量的双方の変数が混在したデータをどう取り扱うかなど、さらなる難問が待ち構えている。

7. 遺伝子コピー数変異（CNV）と疾患

7.1 CNV とは

最近まで、「遺伝子多型といえば SNP」といえるほど、SNP は様々な表現型をつかさどる主要な遺伝子多型として考えられてきた。実際のところ、本書籍のタイトルからして SNP である。たった一塩基の違いが遺伝子産物に影響を与えることで遺伝病を患ったり、また、生活習慣病のような様々な多因子疾患の罹りやすさに影響を与えたりする。大量の SNP を検出できる DNA チップやマイクロアレイは日進月歩で開発され、ヒトの疾患に関連する遺伝子の探索にきわめて強力なツールとなっている。また、テーラーメード医療で目指しているような、個々人の SNP 情報を読み取ることで薬の効きやすさを知り、処方箋に反映させようとする試みも行われている。

ところが、近年、個体間でコピー数が異なる遺伝子領域が、ゲノム全般にわたって存在することが明らかにされ、それが多様な表現型を担っているらしいことがわかってきた(Iafrate et al. 2004; Sebat et al. 2004)。このゲノム構造変異は「Copy Number Variation（CNV)」と呼ばれ、世界中のゲノム研究者たちの注目を集めている (Feuk et al. 2006)。これまでにも、ゲノム配列が 1 kb 以上のまとまった分節単位で挿入、増幅、そして欠失する現象が知られていた。このような変異の生じた配列上に遺伝子が乗っていた場合、その遺伝子コピー数には変化が生じる。つまり、「CNV」という単語には、遺伝子コピー数に影響を与えるような変異、これまでに「分節重複（segmental duplication)」、「欠失変異（deletion variants)」という名前で知られた現象をそっくり包含した概念を有している(表 7.1)。ただし、トランスポゾンによる挿入・欠失は CNV の定義に含まないことが取り決められた。また、ある集団に 1%以上の頻度で存在する CNV を特に copy number polymorphism (CNP：コピー数多型)と呼んでいる。現在までに多型といえるほど頻度が確定しているアレルはまだ数えるほどしかない。詳細は後の項で説明したい。

表 7.1　ゲノム構造変異の用語説明

語句	Term	定義
構造変異	Structural variant	1 kb 以上のまとまった配列で観察されるゲノムの変化
コピー数変異	Copy number variant (CNV)	1 kb以上のDNAで生じた増幅または欠失の総称
デュプリコン	Duplicon	1 kb 以上のまとまった配列が増幅し、その配列中の遺伝子間で９０%以上の相同性があること
欠失	Deletion	1 kb以下の塩基配列が欠失する変異
挿入	Insertion	1 kb以下の塩基配列が挿入する変異
中規模構造変異	Intermediate-sized structural variant (ISV)	8 ～ 40 kb 程度のサイズで生じるゲノム変異
低反復コピー	Low copy repeat (LCR)	分節重複と同義
多所変異	Multisite variant (MSV)	PSVやSNPとも異なるタイプの複雑な多型・変異
パラログ変異	Paralogous sequence variant (PSV)	増幅した遺伝子コピー（パラログ）の間でみられる塩基配列の変異
分節重複	Segmental duplication	1 kb 以上のまとまった配列が増幅し、その配列間で９０%以上の相同性があること。非相同染色体間で起こるもの（interchromosomal）と、染色体内で起こるもの（intrachromosomal）がある。
一塩基多型	Single nucleotide polymorphism (SNP)	一塩基の置換・挿入・欠失。ヒト集団中に１%以上の頻度で存在するSNPは１０００万以上あるとみられる。無作為に選んだ２人の間では平均して約1250塩基あたり1箇所でSNPの違いがある。
反復配列多型	variable number of tandem repeat (VNTR)	数塩基～数十塩基の反復配列からなる多型。ゲノム中に数百～数千箇所で確認されている。
（えすてぃー あーるぴー）	short tandem repeat polymorphism (STRP)	2～5塩基の繰り返し配列からなる多型。マイクロサテライトとも呼ばれる。

CNV が存在する領域に遺伝子が存在すると、文字通り遺伝子コピー数に個体差がでるので、コピー数にしたがって発現量に違いがでたり、発現調節が常に乱された状態になったり、コード配列の異常のため機能不全のタンパク質が産生されたりする (Kleinjan and van Heyningen, 2005)。遺伝性の遺伝子発現変化の 8.75-17.7%が CNV により説明される (Stranger et al. 2007)。

こういった遺伝子発現量の違いは、生活習慣病などの多因子疾患に関連する可能性がある。現在、疾患関連解析では SNP の遺伝子型データを用いるのが主流であるが、CNV による関連解析も今後一層進むことが予想される。それでは、SNP と CNV とのあいだに、従属関係はみられるのであろうか？ SNP と連鎖不平衡にある CNV は、現在明らかになっている CNV のうち 20%のみである。したがって CNV と SNP の遺伝子発現量への寄与は独立といえよう。このことからも、SNP、CNV どちらか一方だけでなく、両方を検証する必要があるといえる。幸い、後述するゲノム全域 SNP タイピングにより、SNP の検出のみならず、CNV の検出も可能である。実際に検出された CNV のゲノム上の位置からみると、タンパク質をコードしている領域よりもむしろ遺伝子構造や発現制御に関与する領域のコピー数が影響して発現量が変化しているようだ。そして、後述するように、CNV が原因となる疾患も続々と報告されている。

7.2 CNV の成り立ち

現在までに検出された CNV は、相同性の高い配列が何度も繰り返された領域（例えばサブテロメア領域、セントロメア近傍、など）か、分節重複が起きている領域が大半を占めていて、欠失変異は少数である。特に、100 万塩基にわたる大きいサイズの CNV はそのほとんどが分節重複によるといっていい。何故にこういった偏りが生じるのであろうか？ 例えば、広範な領域で欠失変異が生じてホモ接合になった場合、その領域にあった遺伝子がごっそりなくなってしまうわけだから、そこに生存に不可欠な遺伝子がふくまれていようものなら、重篤な遺伝病のようなバイアスがかかり、その変異は淘汰される結果になるだろうことが容易に推察できる。対して、そこで重複が生じてその領域にある複数の遺伝子のコピー数が増幅した場合、遺伝子発現量を本来よりも抑える方向に働きかけることで、通常のコピー数と同レベルの発現量にうまく調節して対処できるかもしれない。無論、発現量補正がきかない遺伝子が重複してしまったり、また、重要な遺伝子が欠失した領域がヘテロ接合であったりしたためにバイアスを逃れて維持される CNV は実際に存在し、それゆえ CNV は維持され、表現型に影響を及ぼしているのである。一方、小さいサイズの CNV は、相同性の低い領域で生じている傾向にある。変異の生じる際の分子メカニズムの違いによって何かしらのバイアスがかかっているためであると考えられるがその詳細は不明である。

それでは、これら CNV を生み出したイベントについて説明していきたい。

分節重複（Segmental duplication）は、1〜400 kb のまとまった配列が増幅し、鋳型になった配列と増幅した配列の間で 90 %以上の相同性があるゲノム領域のことを指す (Eichler 2001)。*In situ* hybridization やデータベース解析から、ヒトゲノムの 5%以上が分節重複で構成され、ゲノム上のいたるところに存在している (Bailey et al. 2002; Cheung et al. 2003; Cheung et al. 2001; She et al. 2004)。染色体の内部 (intrachromosomal) で増幅するタイプと、非相同染色体間（interchromosomal）で増幅するタイプがある。特筆すべきは、セントロメア近傍やサブテロメア領域でよく観察されるということである (Linardopoulou et al. 2005; She et al. 2004)。重複配列がぴったり隣り合う「タンデム重複」とは異なり、ゲノム全体にわたりラ

ンダムに配置されている傾向がある。実際、分節重複は非アレル相同組み換え（離れた位置にある相同配列が組み換わること；図 7.1）による構造的配置転換を起こすことから、疾患、進化、構造変異に大きな影響を及ぼす (Armengol et al. 2003; Bailey et al. 2004; Iafrate et al. 2004; Ji et al. 2000; McCarroll et al. 2006; Samonte and Eichler 2002; Sebat et al. 2004; Sharp et al. 2005; Tuzun et al. 2005)。これまでに報告されている、分節重複や欠失が生じている遺伝子座を表 7.2 にまとめた。また、分節重複を検出できるようデザインされたマイクロアレイを用いた研究グループは、複数の集団間で共通した CNV が多型として存在することを明らかにしている (Sharp et al. 2005; Tuzun et al. 2005)。

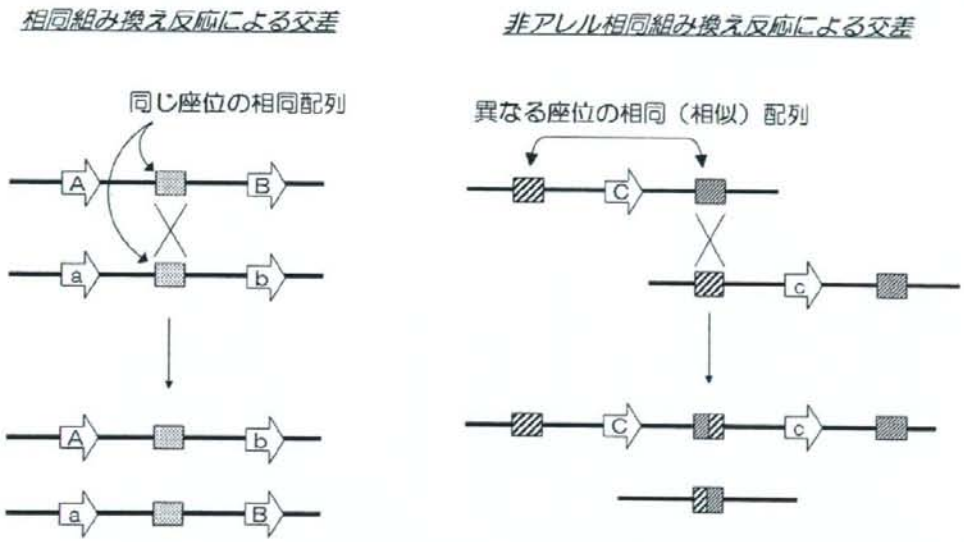

図 7.1 「相同組み換え」と「非アレル相同組み換え」