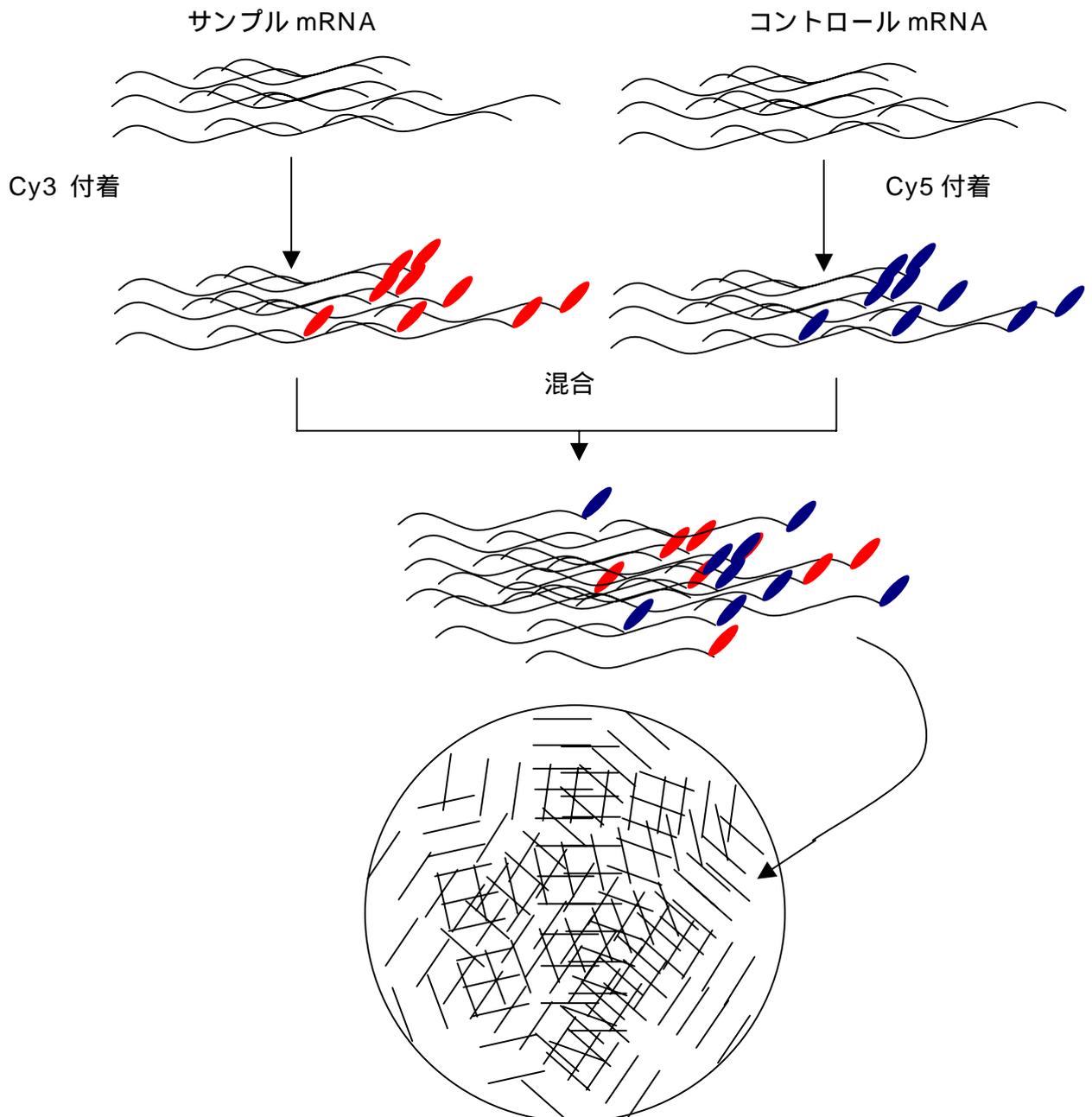


Microarray の原理

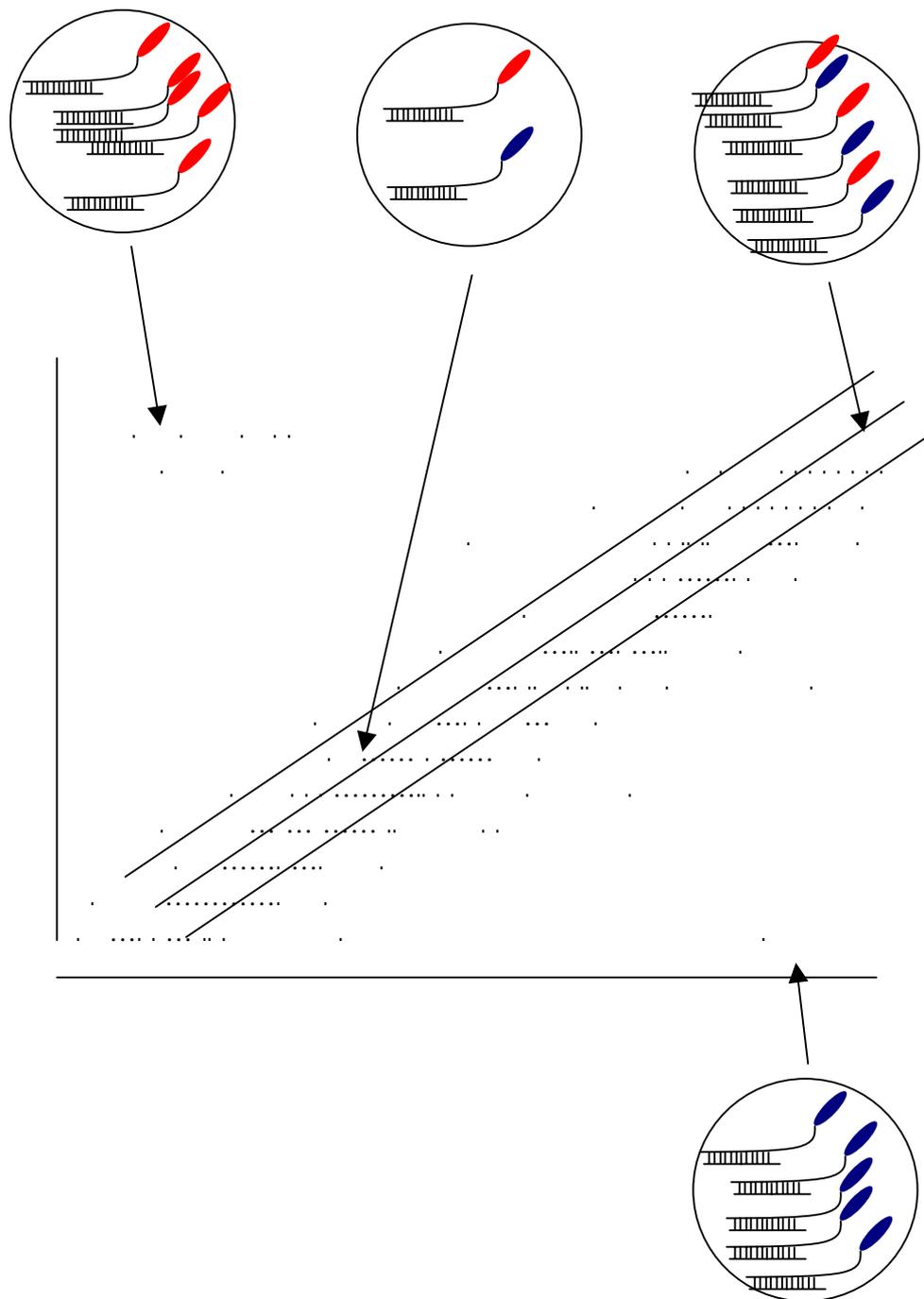
Microarray は主に SNP などの point mutation の検出と mRNA 発現量の同時定量に用いられます。ここでは、後者について解説します。

Microarray の測定には大きく言って 2 種類の方法があります。1 つは Affymetrix 社の光を用いる方法、もう 1 つは遺伝子をプリンタ - で印字するのと類似の原理を利用する方法があります。



ある遺伝子(1)の oligonucleotide が、円形にスポットされています。ここに全ての mRNA を流し込むと、遺伝子(1)の mRNA がスポットに特異的に付着します。そして

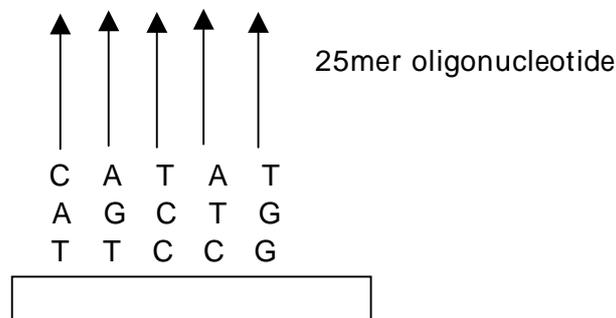
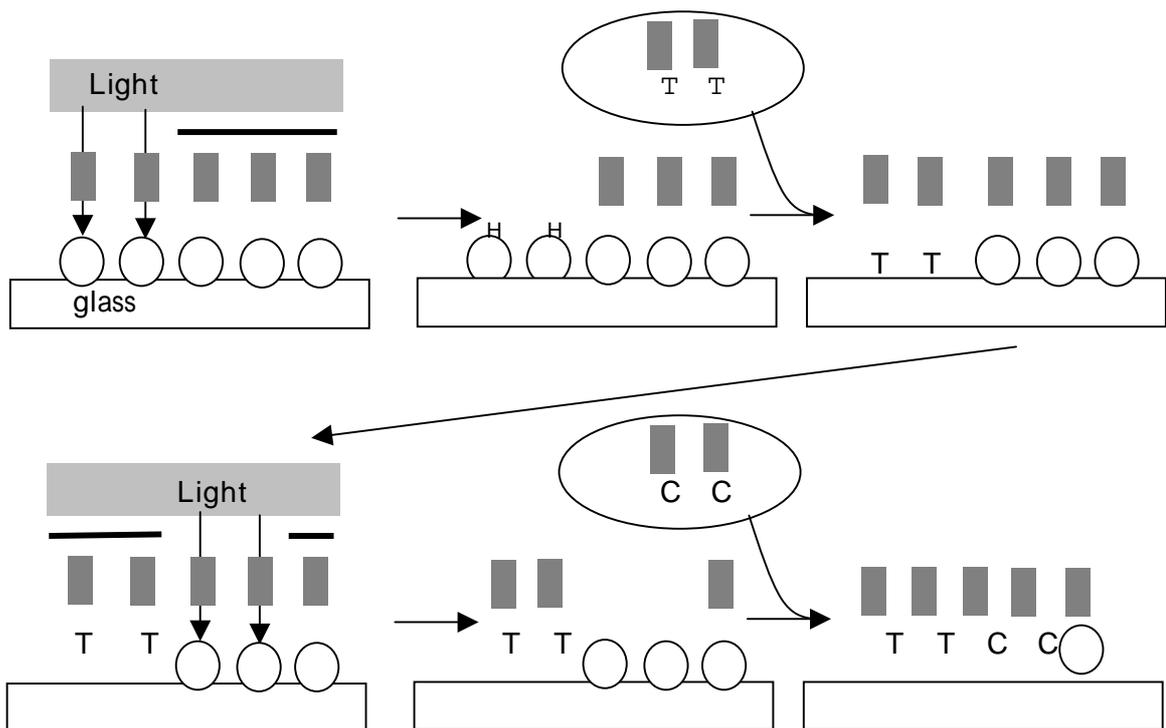
サンプルとコントロールの間で遺伝子 1 に対する mRNA の量が等しければ、赤と青は同じ程度発現されます。一方、サンプルで強ければ赤くなり、コントロールが強ければ青くなります。発現量が多ければ、それぞれの蛍光の強さが増し、逆に発現量が少なければ蛍光強度が弱くなります。図に表わすと以下のようになります。



この場合サンプルを赤、コントロールを青にしているわけですから、もしも両者で差が無ければ $x = y$ の線上にのって来るはずですが。実際には多少の誤差がでるため、バッチ毎に quality control しなくてはなりません。大概是 1.2 倍から 1.6 倍の範囲にくるようです。

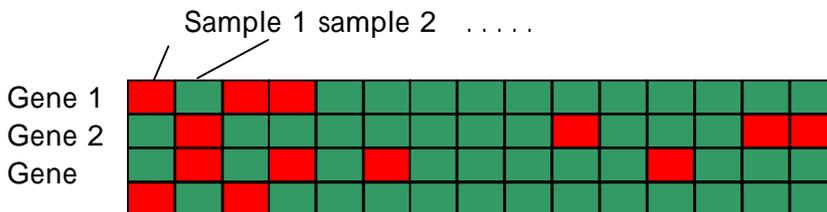
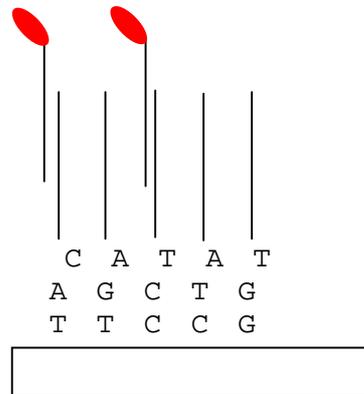
Affymetrix High-Density Oligonucleotide Arrays

Photolithography



ガラス版の上には多数の oligomer が上の方に伸びていきます。まさに海底からワカメがたくさん生えているところを想像してください。上の光を当てた際の穴が四角いものですから、oligonucleotide も四角い形でガラスの上についていきます。この方法により多数の、100 万レベルの、oligonucleotide を 1.28 cm² のガラス板に付着させるこ

とが可能となりました。これに遺伝子を振りかけると、相補的に結びつくものを検出できます。



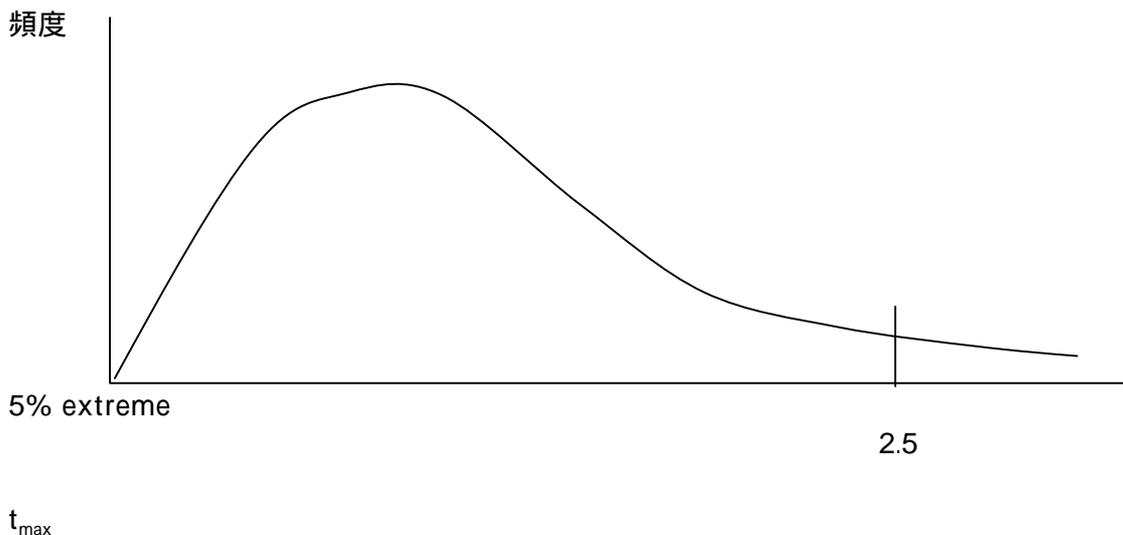
数千の遺伝子

先に挙げた spotted array では clone, PCR 産物、cDNA など種々のサイズの DNA をスライドガラスにプリントあるいは吹き付けて行われます。よって、この技術はプリンターから採用されているのです。これに対して oligonucleotide array の場合は 25 程度の塩基配列を合成しガラスに付けてある点が大きく異なります。ですから、塩基配列を 1 つ変えて合成することも実に簡単にできる点が特徴です。遺伝子の世界では 25 の塩基対で十分な特異性と感受性を保つことができるのです。もちろん、25 の配列はそれぞれの遺伝子に特異的な部分を選択しなくてはなりません。あるいは、rRNA, tRNA, alu-like sequences と類似の塩基配列も避けるようにします。さらに、1 つの種類の mRNA に対していくつかの oligonucleotide を作り、異なった場所に配置するようにする方法も採用されています。このことにより signal-to-noise ratio を減じ、極端な値を探している遺伝子なのか、false positive なのか判定するためにも有用です。ある mRNA に対応する oligonucleotide 4 つが全て高ければ noise でなく signal であり、どれか 1 つが高ければ signal でなく noise と考える方が合理的です。特に 25 程度の oligonucleotide ですと、基本的に特異的とはいえ、時に異なった mRNA と結合する心配がいつもつきまといます。

Permutation 2-sample t tests on large arrays

この microarray の特徴は sample 数に比較して遺伝子の数がとてつもなく多いことです。Basic で行ったように2つの sample を比較するとき t test を行うとします。P=0.05 ということは、100回繰り返すと本当は同じなのに5回は間違ってしまうということです。もしも1万の遺伝子を解析した場合、500の遺伝子解析は本当は2群間で差がないのに差があると言ってしまうわけです。よって $p < 0.05$ では cut off が甘すぎます。いくつを設定基準にしたらいでしょうか？

例えばダウン症の遺伝子発現を正常のものと繊維芽細胞を用いて比較しようと思います。そして比較する遺伝子を96個とします。検体は4人のダウン症患者由来のもの、4人の正常人由来のもの、合計8種類あります。まず、この中からランダムに4つを選ぶことを考えます。そして、残り4つと比較します。それぞれの遺伝子でt値をだすことができますよね。そしてt値の最も大きい値を t_{max} とします。8個から4個選ぶわけですから70個の組み合わせがあるわけですが、仮に500回行ったときの t_{max} を横軸に、その頻度を縦軸にとったら下図のような分布を得たとします。



仮に t_{max} が 2.5 以上のとき、 t_{max} 分布の極端な 5% に相当したとします。このことは t は 1.96 でなく、2.5 をもってめったに起こり得ない事象であり、有意差ありと考えるべきであることを示唆します。ですから、上の例で考えると遺伝子 1 について 4 人のダウン症患者の発現量と 4 人の正常人のそれを比較して $t = 2.7$ を得れば、ダウン症患者の遺伝子 1 の発現量は違うということになります。一方、遺伝子 2 の t 値が 2.0 であったとすると、通常のカット・オフ 1.96 より大きいのですが、2.5 より小さいので、有意とはいえません。

癌の分子分類

今まで癌の病理分類は予後を予測したり癌治療法を選択する上で無くてはならない存在でした。更に、免疫組織学的手法も癌分類に大きく貢献してきました。しかしながら、病理分類で同じ疾患であり、同じ予後因子に対して同じ治療を施しても、しばしば全く異なった結果をみます。近年、癌抑制遺伝子等の変異などを検討することにより、そのギャップを埋めようとする研究が行われてきました。そして、microarray technology に行き着いたのです。この1-2年、microarray を駆使して数千の遺伝子発現状態を同時に解析することにより癌の分子分類が成されつつあります。このことにより、癌を現在の病理分類より更に細分化し、木目細かい治療を行うことができるようになるでしょう。そして、究極的には患者個人に合わせたオーダーメイドの治療を励行できる日も来るかもしれません。

クラス解析

白血病分子分類

Bolub らは初診の白血病患者の 38(ALL27, AML11)の骨髄より検体を採取し、Affymetrix の high-density oligonucleotide microarray にて 6817 の RNA 発現パターンを比較しています。そして、全く新しい白血病亜分画をみつけられるか?(class discovery)と、病理免疫分類を分子発現パターンから予測できるか?(class prediction)について検討しています。

Neighborhood analysis. 検討した遺伝子のうち 1100 は AML あるいは ALL のどちらかに有意に高く、あるいは低く発現されていました。

Class predictor. 既知の検体を使って未知の検体を ALL か AML かに分類します。まず ALL か AML の分類とよく関連する遺伝子を informative genes として固定します。そして、これらの遺伝子の発現レベルに応じて未知の検体が ALL になるのか AML になるのか検討します。その informative gene がどちらかの白血病と関連が強い程、あるいは未知の白血病 informative gene の発現が偏っている程、予測しやすいことになります。この偏りを weighted vote と呼びます。この vote を合計して prediction strength とします。この prediction strength は 0 - 1 で表される数値となり、ALL である可能性などを示唆します。

この論文では ALL か AML か判っている 38 の検体のうち 36 を明確に診断 (predictor strength 0.3 未満を不明とする) できるように 50 個の informative gene によりを算出し、36 の未知の白血病検体を ALL か AML かに分け、どれくらい正解するかを調べたところ全例診断(ALL あるいは AML)と一致していました。これを用いて未知の白血病検体 34 について検討したところ、29 は診断可能であり(prediction strength \geq 0.3)、全例についてこの molecular classification を用いて正しく ALL か AML かを診断できました。未知の検体は、骨髄、末梢血、小児のものなど多岐にわたっていました。そして prediction strength は 0.77 と非常に高いものでありました。今回の 50 という数値は適当なのですが、10 - 200 の間で検討したところ、どれでも診断率 100%であり、仮に 10 でも十分でした。

もちろん CD33 といった細胞表面抗原によっても ALL か AML かに判断でき

ますが、cyclin D3, RbAp48, SNF2, TFIIIB, c-Myb, E2A HOXA9, topoisomerase II などの多くの遺伝子が informative gene として使用でき、単に未知の白血病細胞を分類するだけでなく、癌細胞内に異常や薬物治療などとも関連しそうな因子がつかまりました。

この解析を ALL か AML かだけでなく、予後とも関係しないかどうか検討したところ、15人の成人 AML 中 8人が寛解不全、残る 7人が 46 - 84 ヶ月寛解を維持しました。Neighborhood analysis では明らかな ALL-AML 型分類と違って、明らかな相関を見出せる遺伝子を同定することができませんでした。しかし、検討した数が少ないため有意差を検出できなかった可能性があります。有意差はないながらも、HOXA9 の過剰発現が白血病の予後と最も相関していました。この HOXA9 は t(7;11)(p15;p15)染色体転座により異常を持ち、この染色体異常を持つものは予後が悪いと言われていますので、この HOXA9 遺伝子過剰発現が予後不良と相関することは十分あり得ますが、新しい事実でもあります。

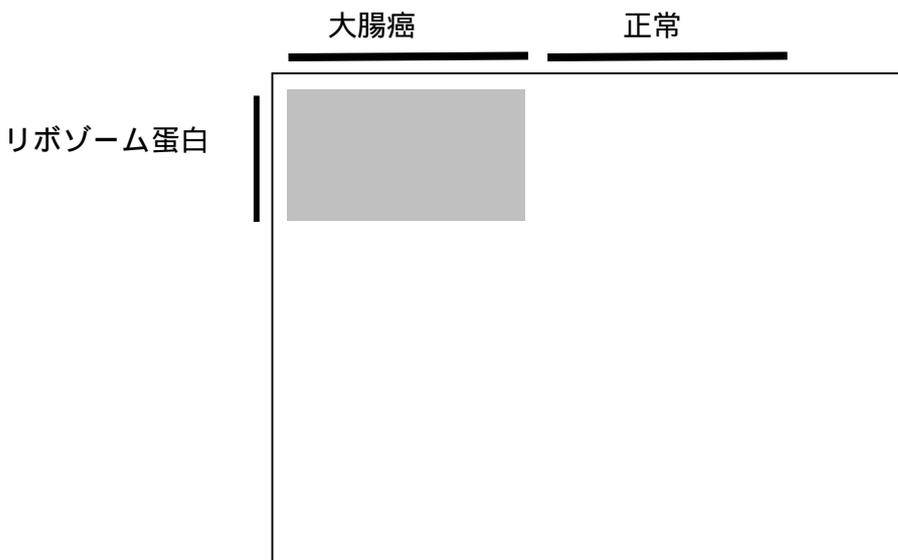
Golub TR et al. Molecular Classification of Cancer: Class discovery and Class Prediction by Gene Expression Monitoring. Science 286:531-537,1999.)

クラスター解析

大腸癌分子分類

44 の大腸癌組織と 22 の正常大腸組織の遺伝子 (6500) 発現を Affymetrix oligonucleotide array を用いて調べ、two-way clustering algorithm にて解析しました。2000 の遺伝子の組み合わせで、極一部は $P < 0.001$ で高い相関を示していました。つまり、1つの遺伝子が強く発現されているとき、他の遺伝子は強くあるいは弱く発現されています。このことは遺伝子 - 遺伝子相互作用を示しているかもしれませんが、同じ転写調節因子の影響を受けているかもしれません。つまり遺伝子が同じような動きをする (clustering) ということは、共通の生理機能を有していると考えられます。このように clustering するものを gene tree あるいは tissue tree と呼び、ある公式を使って遺伝子群、あるいは組織群を2つに分けます。さらにこれを2つに分け、また分けといった形で続けると、トーナメント戦で使うような網だになります (つまり2つ一組のところまで続ける)。通常、このような cluster で整理した tissue tree を横軸、gene tree を縦軸にとります。そうすることにより、水平軸において似たような組織はかたまり、垂直軸において似たような動きをする遺伝子はかたまります。例えば水平軸は、最初の10個列が diffuse large cell lymphoma, 次の10列が lymphoblastic lymphoma, 次の10列が Burkitt's lymphoma, 次の10列が正常 B 細胞といった具合です。一方垂直軸には細胞周期関連の遺伝子群、B 細胞分化抗原遺伝子群、免疫グロブリン関連遺伝子群、T 細胞系遺伝子群、などと clustering していきます。この2つの座標で示された遺伝子発現で、赤い領域があれば遺伝子と病理組織が一致するわけで、同じ病理組織内でもある遺伝子群に対応する領域に赤い領域と緑の領域が存在すれば、病理組織を遺伝子発現パターンに従って2つに細分化することができることを示唆します。

従来の報告通り、大腸癌では ribosomal protein gene-set が多く発現されており、一方正常大腸組織では僅かしか発現されていませんでした。



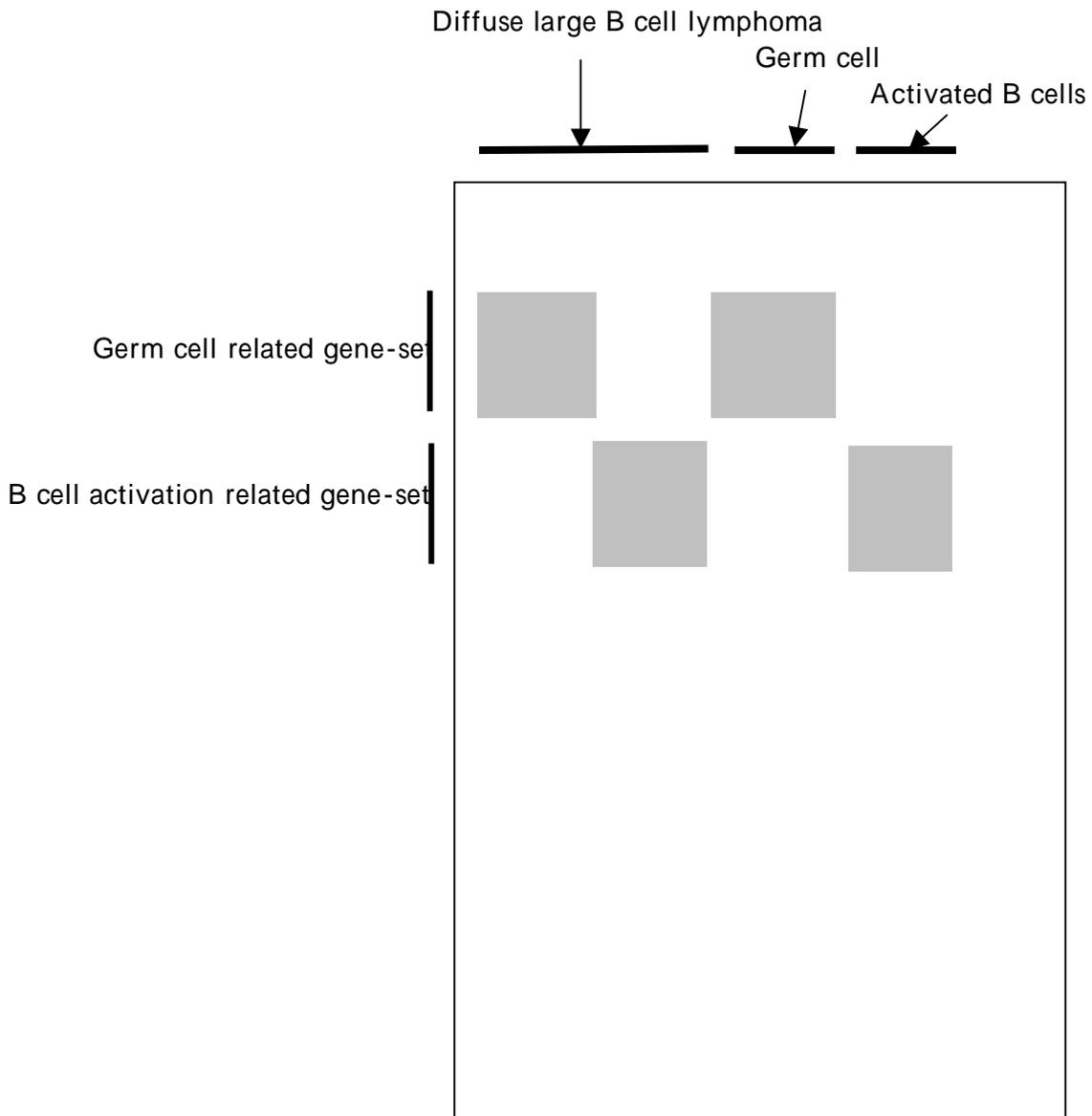
大腸癌でリボゾーム蛋白が高いことを言うためには、大腸癌と正常大腸での数値を単純に t test で比較します。数値とは cy5/cy3 ratio です。

(Alon U, et al. Broad patterns of gene expression revealed by clustering analysis of tumor and

normal colon tissues probed by oligonucleotide arrays. Proc Natl Acad Sci USA. 96:6745-6750,1999)

悪性リンパ腫分子分類

diffuse large B-cell lymphoma (DLBCL)は non-Hodgkin's lymphoma の中でも最も多い型ですが、臨床的には heterogeneous であり、およそ 40%は治療によく反応し生存率も良いのですが、他は予後不良です。我々は、従来の方法では検出できない分子生物学的相違がこの予後の良い DLBCL と予後の悪いそれとの間に存在するのではないかと考えました。そこで DNA microarray を用いて B 系悪性腫瘍の遺伝子発現について検討してみました。



上図をみて判る通り、DLBCL は germinal center B-like DLBCL (正常の germinal center B-cell と遺伝子発現パターンが類似)と activated B-like DLBCL(正常の activated B-cell と遺伝子発現パターンが類似)に分けることができました。前者では予後が良く (19 人中 6 人死亡)、後者では予後が悪く (21 人中 16 人死亡)、従来の病理

分類では識別できなかった DLBCL を更に分子分類することにより予後の差を明らかにすることができました。

Germinal center B-like DLBCL では BCL-6 が、一方 activated B-like DLBCL では BCL-2 が高い傾向にありましたが、必ずしも同部を含む染色体転座とは関連していませんでした。更にいくつかの遺伝子発現がどちらかに偏って発現されていました。上のような関係より diffuse large cell B cell lymphoma には germ cell 由来と peripheral B cell 由来があることが理解できます。

(Alizadeh AA, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 403:503-511,2000.)

乳癌分子分類

乳癌組織は癌細胞だけでなく、正常上皮細胞、正常間質細胞、死亡細胞、内皮細胞、リンパ球など様々な組織が混在しています。また乳癌細胞自体も形態的に遺伝子的に雑多な性格を持ち合わせています。Microarray においても、組織に B リンパ球や間質細胞の混入が多いと microarray でもその影響を受ける傾向にありました。他、細胞増殖に関連のある遺伝子、インターフェロンに関連のある遺伝子が cluster として乳癌と関係することが確認されました。

(Perou CM, et al. Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. Proc Natl Acad Sci USA. 96:9212-9217,1999)

マイクロアレイの将来性

マイクロアレイの問題点は、遺伝子をつけたスライドガラスの値段が高い点にありました。しかし、最近各社が競っているため値段がどんどん下がっています。また、1つの目的を達成するのに必ずしも全ての遺伝子について検討する必要はありません。例えば悪性リンパ腫を分子分類するのに 96 の遺伝子を調査すれば、治療決定のための十分な情報が得られることでしょうか。そうなると、病院中央検査室で保険点数 1,000 点程度で検査ができる日も近いのではないのでしょうか？