

癌マイクロアレイデータ解析に向けた
マイニング技術の開発

水野 英明

目次

要旨	4
第一章 遺伝子シグネチャーを用いたマイクロアレイデータからの細胞周期分布予測法	
1. 1 背景	6
1. 1. 1 癌の本質	6
1. 1. 2 細胞周期の概要	7
1. 1. 3 癌組織の細胞周期を測定する既存の手法	8
1. 1. 4 既存の細胞周期測定法の課題	9
1. 1. 5 遺伝子シグネチャー	9
1. 2 CCS法の概要	10
1. 2. 1 細胞周期に制御されて発現する遺伝子群の大規模同定	10
1. 2. 2 サイクリング細胞特異的に発現する遺伝子の同定	11
1. 2. 3 CCSマスターセットの作成	12
1. 2. 4 CCSサブセットの作成	12
1. 2. 5 CCS法での細胞周期測定	13
1. 3 結果	15
1. 3. 1 細胞周期データセットを使ったコンセプト証明	15
1. 3. 2 非サイクリング細胞を含むデータセットを使ったコンセプト証明	17
1. 3. 3 マウスデータセットでの動作確認	19
1. 3. 4 マウス腫瘍モデルデータセットの解析	20
1. 3. 5 ヒト乳癌データセットの解析	23
1. 4 考察	25

1. 5	材料および方法	28
1. 5. 1	細胞培養および細胞同調	28
1. 5. 2	マイクロアレイ実験	29
1. 5. 3	シグネチャースコアとデータ可視化	29
1. 5. 4	生存時間解析	30
第二章	PrognoScan : 遺伝子発現と癌患者予後の関連の横断解析データベース	
2. 1	背景	31
2. 2	結果	32
2. 2. 1	癌マイクロアレイデータの収集	32
2. 2. 2	解析アルゴリズム	34
2. 2. 3	横断解析の実例	35
2. 3	考察	41
	謝辞	44
	本研究の業績	45
	参考文献	46
	補足資料	58

要旨

マイクロアレイは、遺伝子の転写産物である mRNA と特異的に結合する核酸配列を配置したスライドに対し、サンプルから調製したラベル化 mRNA をハイブリダイズさせることで、サンプル中でどの遺伝子がどれだけ転写されているかを測定する技術である。高密度化が進んでおり、数万の遺伝子の発現量を一度の実験で測定することも可能なため、近年の癌研究において重要ツールとして位置付けられている。一方で、生み出す情報量が多いことから、研究者がデータを解釈しきれない現象も起き始めている。この状況を鑑み、本研究では癌サンプルのマイクロアレイデータから有用な知見を導くためのマイニング技術の開発に取り組んだ。

第一章では、遺伝子シグネチャーを用いたマイクロアレイデータからの細胞周期分布予測について報告する。癌の本質は細胞周期異常であり、癌の特徴は最終的には細胞周期機構へ反映される。これまで、顕微鏡やフローサイトメトリー、あるいは免疫組織化学染色などを使った細胞周期解析手法が癌診断のためのツールを提供してきたが、いずれの手法も一つまたは少数の測定値に依存しているため、得られる情報が限られてきた。マイクロアレイのような網羅的な技術を使って細胞周期分布を俯瞰する解析手法の開発が望まれていた。今回、筆者は細胞周期の各フェーズを代表するマーカー遺伝子セット「細胞周期シグネチャー」(Cell Cycle Signature; CCS) を作成し、その発現を調べることでマイクロアレイのデータからサンプルの細胞周期分布を推定する新規手法を開発した。多くのパラメータの上に成り立つ CCS 法はこれまでの方法と異なり、増殖細胞(サイクリング細胞)と静止細胞(非サイクリング細胞)を同時に考慮することが可能で、静止細胞に「埋もれた」細胞周期分布を調べることができる。CCS 法を用いてマウス腫瘍モデルデータセットを解析したところ、非サイクリング細胞の影響を除いた場合に、癌化イベントに特異的な細胞周期分布パターンが明瞭に浮かび上がり、この手法が癌の特徴づけに利用できることが示唆された。さらに、CCS 法によるヒト乳癌データセットの解析結果は患者の予後とより強い相関を

示し、診断における有用性も示唆された。CCS 法は癌の特徴分類・診断に役立つことが期待される。

第二章では、遺伝子発現と癌患者予後の関連の横断解析データベース「Prognoscan」について報告する。ある遺伝子の発現が患者予後と関連することが分かった場合、その遺伝子と癌進行プロセスとの間に何らかの因果関係を疑えるため、研究を推進する動機となる。実際、これまでこの前提の下に多くの癌遺伝子候補が提案されてきた。近年、臨床情報の付随した癌マイクロアレイデータが大量に公開され、新規の実験を行わずとも、こうしたデータを解析することで遺伝子発現と患者予後とを結びつけることが可能となった。しかし、データ解析のための効率的なプラットフォームの開発が遅れているため、活用が進んでいない。筆者はここに着目し、1) 臨床情報の付随した癌マイクロアレイデータの網羅的なコレクション、2) **minimum *P*-value approach**を用いた遺伝子発現に基づく生存解析ツール、の二つの特徴を持つデータベース「Prognoscan」を開発した。Prognoscanを使うことで、遺伝子発現と癌患者予後の関連性のデータセット横断的な解析が容易に実現できる。本研究でも、例として癌遺伝子候補 **SIX1**の発現と乳癌、**MCTS1**の発現と脳腫瘍・血液癌・乳癌・肺癌の予後が関連することを新規に示し、さらなる研究の足がかりを提供した。筆者はPrognoscanを、誰もが自由に利用することができるようインターネット上で公開した (<http://gibk21.bse.kyutech.ac.jp/Prognoscan/index.html>)。Prognoscanは今後、潜在的腫瘍マーカーや創薬標的を評価するための強力なプラットフォームとして、癌研究を加速させることが期待される。

第一章

遺伝子シグネチャーを用いたマイクロアレイデータからの細胞周期分布予測法

1. 1 背景

あらゆる癌の本質的な特徴は細胞周期異常である (Whitfield *et al.* の総説, 2006)。点変異、遺伝子増幅、癌遺伝子活性化、癌抑制遺伝子不活性化など、さまざまな因子が癌の進展に関与することが明らかにされているが、それらは結局のところ、直接的・間接的に細胞周期機構を攪乱することによって達成されている。細胞周期機構は上流に存在する様々な癌化シグナルが最終的に統合・反映される地点となるため、それぞれの癌において細胞周期がどのように影響を受けているかを調べることは、その性質を知るための有力な手がかりとなる。これまでに、顕微鏡を用いた有糸分裂期細胞数の測定、フローサイトメトリーでの DNA 合成期細胞数の測定、あるいは細胞周期マーカー遺伝子に対する免疫組織化学染色など、様々な方法が開発され、癌診断のツールを提供してきた (Landberg and Roos, 1997; Gonzalez *et al.*, 2004; Colozza *et al.*, 2005; Beresford *et al.*, 2006)。しかし、これらの手法は一つもしくは少数の測定値に依存するため、得られる情報が限られていた。より網羅的な技術を用いた細胞周期解析法が開発が望まれている現状がある (Colozza *et al.*, 2005; Beresford *et al.*, 2006)。本研究は、癌研究の主要ツールとなったマイクロアレイのデータから、遺伝子シグネチャー法を用いて、細胞周期分布を推定する手法の開発に取り組んだ。

1. 1. 1 癌の本質

ヒトの体は1つの受精卵が細胞分裂を繰り返し、細胞数が増えることによって形作られる。やがて成体になると、細胞分裂機構に制御がかかり、古い細胞を置き換える細胞のみが供給され、体細胞数は一定に維持されるようになる。しかし、遺伝子の変異・増幅・欠損、あるいはシグナル伝達経路の異常な活性化・不活性化などで制御に狂いが生じると、細胞は無制限に増殖を続けるようになる。

これが癌である。癌の本質はこの異常な細胞分裂にあり、それを司る細胞周期に何らかの異常が生じている。

1. 1. 2 細胞周期の概要

細胞周期のモデルを図1に示す。受精もしくは細胞分裂によって新たに誕生した細胞は、最初、**G1期** (**1st gap phase**) と呼ばれるフェーズにある。この**G1期**に細胞は、**DNA** 損傷の修復やさまざまな生合成に必要な酵素の活性化など、細胞分裂への準備を行う。**G1期**の末期に **G1** チェックポイントと呼ばれる、細胞分裂を開始するかどうかを決定するポイントがあり、**DNA** に損傷がない・十分な栄養がある・十分なスペースがあるなどの一定の条件が整うと、細胞はこのポイントを突破する。続く **S期** (**Synthesis phase**) では、一つの母細胞が持つ遺伝情報を二つの娘細胞細胞に伝えるための染色体の複製が進行する。染色体コピーが完了すると、**G2期** (**2nd gap phase**) と呼ばれる短いフェーズを経る。そして **M期** (**Mitotic phase**) に入ると、それぞれの染色体セットが細胞の二極に引っ張られる形で分配される。最後に細胞の分割が起こり、細胞分裂が完了する。生み出された二つの娘細胞はこの時点からそれぞれ **G1期**に入り、次の分裂に向けて一連のプロセスを繰り返す。このサイクルが細胞周期である。

発生の早い段階では、組織形成に必要な細胞数を確保するために細胞分裂が活発に行われるが、成体では様々な制御機構が細胞周期を抑制するようになり、組織中の **S~M期**の細胞存在比が減る。成体の完成した組織では、ほとんどの細胞は **G1期**のどこかのタイミングで細胞周期自体を抜け、**G0期**と呼ばれる静止期に入っていると考えられている。

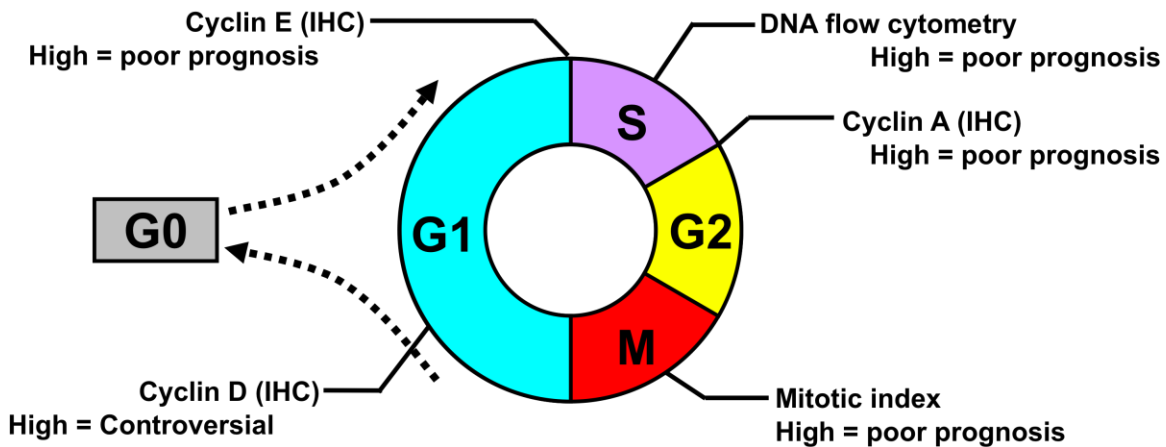


図1 細胞周期のモデル図

1. 1. 3 癌組織中の細胞周期分布を測定する既存の手法

癌は上述のような細胞周期機構に異常が生じた状態にある。正常組織に比べ活発な細胞増殖が行われているため、細胞集団内に S～M 期の細胞が多く、G1/G0 期の細胞が少なくなっている。この存在比は増殖速度が早く、より悪性度の高い癌ほど高くなる傾向があることも知られている。このような観察から、癌組織の細胞周期測定が癌を特徴付ける有力な手段となることが認識され、さまざまな方法が開発されてきた。主要なものを以下に示す。

イ) S phase fraction : ヒト細胞は通常状態で、父方・母方それぞれから受け継いだ $2n$ の染色体数を有する。S 期にある細胞はこの複製を行っているため、 $2n \sim 4n$ 分の染色体を含有する。よって、染色体の構成成分である DNA を蛍光染色し、フローサイトメトリーで各細胞の蛍光強度を測定することで、S 期にある細胞数を測定することができる。

ロ) Mitotic index : 染色体の分配は M 期に起こる。複製されたそれぞれの染色体セットが、細胞の両端から伸びてきた紡錘糸に引っ張られ分離していく様子は特徴的で、顕微鏡で容易に確認できる。ここから、顕微鏡視野内の有糸分裂体数を数え M 期の指標とする Mitotic index が開発された。

ハ) 免疫組織化学染色法 (Immunohistochemistry; IHC) : 細胞周期は非常に精密に制御されている機構である。プロセスが逆方向へ進行しないようにするため、細胞周期の特定の時期になると発現・反応を仲介し、役割を終えると姿

を消す遺伝子が多数存在する。例えば、細胞が G1 期から S 期へ移行する時期に発現する Cyclin E、S 期から G2 期へかけて発現する Cyclin A、S-G2-M 期に発現する Geminin などである。このような細胞周期特異的な遺伝子 (Cycling gene) を抗体染色することで、対応するフェーズの細胞数を推定することができる。

1. 1. 4 既存の細胞周期測定法の課題

上記のいずれの手法も一つまたは少数の測定値に依存しており、得られる情報が限られている。例えば、Mitotic index で識別できるのは M 期のみであり、DNA フローサイトメトリーは G0 と G1 期、G2 と M 期の区別ができない。マイクロアレイのような、より網羅的な技術を用い、細胞周期分布を俯瞰する解析方法が必要とされている (Colozza *et al.*, 2005; Beresford *et al.*, 2006)。

1. 1. 5 遺伝子シグネチャー

遺伝子シグネチャーは膨大なマイクロアレイデータに特定の表現型が見られるかどうかを調べるために考案された技術である。基本的なコンセプトは、モデル実験のマイクロアレイデータからマーカー遺伝子群をシグネチャーとして同定し、それを未知サンプルのデータに当てはめ、モデルと類似しているか否かを見るというものである。癌研究への応用がめざましく、様々な視点から多様なシグネチャーが開発されている。例えば、乳癌転移・非転移患者群の比較から作成された「70 遺伝子シグネチャー」は、独立の癌サンプルに対しても転移リスクの予測能力があることを示し (van't Veer *et al.*, 2002)、特定シグナル伝達経路を活性化したモデル細胞から作成された「パスウェイシグネチャー」は、乳癌・肺癌・卵巣癌のシグナル伝達経路の活性化状態を予測した上で癌を特徴付ける能力を示した (Bild *et al.*, 2006)。薬剤高感受性細胞株と低感受性細胞株の比較から作成された「薬剤応答シグネチャー」は、乳癌・卵巣癌において化学療法の効果を高精度で予測できることを示した (Potti *et al.*, 2006)。

以上のような背景から、細胞周期を表現する遺伝子シグネチャー (Cell Cycle Signature; CCS) を作製・癌マイクロアレイデータへ適用し、細胞周期分布解析を行うアイデアが生まれた。

1. 2 CCS法の概要

細胞周期の各フェーズを代表する一連の CCS を以下の手順で作成した。

1. 2. 1 細胞周期に制御されて発現する遺伝子群の大規模同定

まず、Whitfield *et al.* (2002) が公開した細胞周期データセットの解析を行った。このデータセットは、Hela S3 細胞株を DNA 合成阻害剤である Thymidine で処理・強制的に S 期で同調させた後、Thymidine を取り除き細胞周期を再開させ、そこから 4 6 時間分の遺伝子発現を 1 時間おきに cDNA マイクロアレイで計測したものである。この間に細胞は 3 回分裂を行っているため、細胞周期によって転写が制御されている遺伝子 (Cycling gene) は、周期的な発現のピークを 3 ないし 4 持つ波動のような発現パターンによって同定することが可能である。生データを Stanford Microarray Database (<http://smd.stanford.edu/>) からダウンロードし、各サンプル毎に Cy5・Cy3 の二つのチャンネルからの信号強度を Quantile normalization (Bolstad *et al.*, 2003) で処理した後、 $\log(\text{Cy5}/\text{Cy3})$ を求めた (チャンネルレベルの正規化)。アレイ毎のシグナル強度のバラつきを一定に揃えるため、データセット全体に対して再度 Quantile normalization を実施した (アレイレベルの正規化)。ノイズ除去の目的で、各プローブについてタイムコースに沿って 3 時間分のウィンドウサイズで移動平均をとり、発現量を平滑化、最後に Z 変換を行って変動幅を揃えた (プローブレベルの正規化)。各プローブのタイムコース内での発現変動の周期性と位相を特定するため、1 ~ 40 時間の 15 分おきの各時間長についてフーリエ変換を適用し、その周期性の強さを調べた。Whitfield *et al.* (2002) の原著論文に記載されている既知の 51 の Cycling gene 群は全体として 14.75 時

間の時間長にピークを示したため（図3）、以下の基準を用いて全プローブから Cycling gene 候補を選抜した。

$$Z\text{-score}(P_i) > 1.96$$

ここで P_i はフーリエ変換で求めたプローブ ($i = 1, \dots, 44,160$) の 14.75 時間の時間長における周期性の強さである。この結果、細胞周期で定期的な発現を示す 1,633 プローブ、全 976 遺伝子のリストを得た。

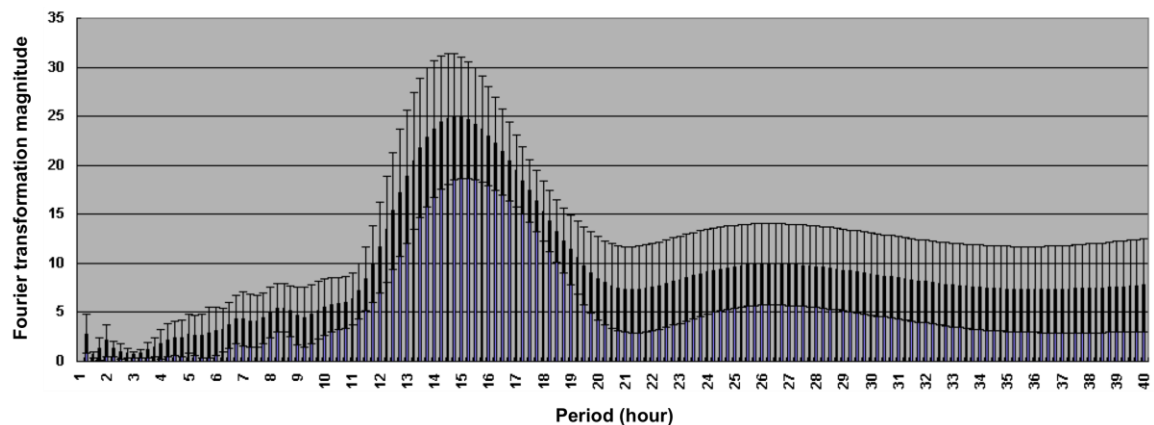


図2 Whitfield *et al.* データセットにおける既知の細胞周期遺伝子群の周期性。横軸は時間長、縦軸はフーリエ変換の強度（平均値）を示す。

1. 2. 2 サイクリング細胞特異的に発現する遺伝子の同定

細胞周期は $G1 \cdot S \cdot G2 \cdot M$ 期に、静止期である $G0$ 期を加えたものとしてモデル化されている（図1）。Whitfield *et al.* (2002) のデータセットは計測間隔の短さ・観測回数の多さから Cycling gene を同定するための最も良いデータセットであるが、残念ながら $G0$ 期細胞のデータを含まない。1. 2. 1 で求めた Cycling gene が $G0$ 期も代表する可能性を排除するため、Bar-Joseph *et al.* (2008) データセットを追加的に解析した。このデータセットは包皮線維芽細胞 (Foreskin fibroblast; FF) を 1) 血清飢餓状態に置いた、もしくは 2) Thymidine ブロックによって細胞周期同調した (0~32 時間, 2 時間毎)、の遺伝子発現データを含む。通常、ヒト細胞の増殖には血清成分が必要で、これが

不足すると細胞は細胞周期を出て G0 期へ入り、非サイクリング細胞となる (Prather *et al.*, 1999)。従って、各 **Cycling gene** について、血清飢餓細胞と各細胞周期フェーズでの発現量を比較し、G0 期を代表する可能性のある遺伝子を除外した。具体的には、以下の評価基準によって、細胞周期のどの期間においても常に非サイクリング細胞より発現が高く保たれている遺伝子を選抜した。

$$\max(e_{ij}) < \min(e_{ik})$$

ここで e_{ij} は血清飢餓 FF ($j = 1, 2$) のプローブ i のシグナル値、 e_{ik} は細胞周期同調 FF ($k = 1, \dots, 17$) のプローブ i のシグナル値である。この結果、全 22, 277 のうち 2,304 プローブ、1,779 遺伝子のリストを得た。

1. 2. 3 CCS マスターセットの作成

1. 2. 1 および 1. 2. 2 で得たリストを比較し、両者の交わりから最終的に 252 遺伝子のリストを得た (別表 1)。これら、1) 遺伝子発現が細胞周期によって調節され、且つ、2) サイクリング細胞で特異的に発現する、遺伝子群を $\text{CCS}_{\text{cycling}}$ と呼ぶことにする (図 3, $\text{CCS}_{\text{cycling}}$)。 $\text{CCS}_{\text{cycling}}$ は G0 期以外のすべての細胞周期を代表するため、全体的な細胞周期活性の指標となる。 $\text{CCS}_{\text{cycling}}$ には Ki67、geminin、TOP2A、Aurora A、PCNA などの既知の細胞周期マーカー (Landberg *et al.*, 1997; Whitfield *et al.*, 2002; Gonzalez *et al.*, 2004; Colozza *et al.*, 2005; Beresford *et al.*, 2006; Williams and Stoeber, 2007) が含まれていた。一方、細胞周期によって発現が調節されているものの、静止期でも発現が上昇することの知られている p21 や Cyclin G1 といった遺伝子 (Ezoe *et al.*, 2004; Zhou *et al.*, 2006) は含まれなかった。

1. 2. 4 CCS サブセットの作成

$\text{CCS}_{\text{cycling}}$ として同定された遺伝子を、それぞれが代表するフェーズに応じてサブセットに分割した (図 3, **CCS subset**)。具体的には細胞周期を 36

0°の円とみなし、フーリエ変換で得られた位相に従って20°毎に均等に18分割した。各 CCS サブセットはそれぞれが三つ以上の遺伝子を含むようになっている（別表1）。尚、マイクロアレイでは同じ遺伝子に複数プローブがデザインされていることがあり、これによって同じ遺伝子が隣接する複数の CCS サブセットに現れることがあることに留意されたい。

それぞれの CCS サブセットは細胞周期の特定のフェーズを代表する。この論文では以降、各 CCS サブセットを、CCS という単語の後に代表する細胞周期フェーズを添えた CCS_{phase} の規則を使って示す。例えば、G1 期の CCS サブセットは CCS_{G1} 、G2 から M 期にかけてのサブセット群は CCS_{G2-M} という具合である。

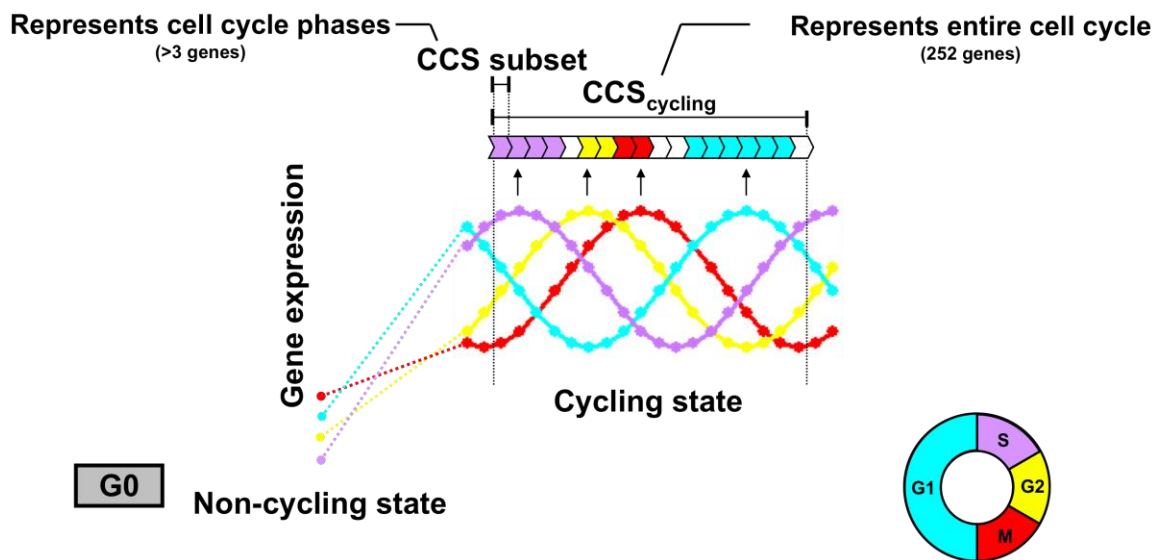


図3 CCS の概念図。 CCS_{cycling} は細胞周期に調節され、かつサイクリング細胞で優先的に発現する遺伝子で構成される。それぞれの CCS サブセットは細胞周期の特定のフェーズで発現がピークに達する遺伝子から成る。

1. 2. 5 CCS 法での細胞周期測定

癌組織にはサイクリング細胞と非サイクリング細胞が様々な割合で混ざって存在している (Baker *et al.*, 1995)。細胞周期分布を考える場合、総細胞数に対する分布を求め、癌組織としての特徴を捉える考え方と、サイクリング細胞に対する分布を求め、癌細胞としての特徴を捉える考え方の二通りの見方ができる。

マイクロアレイはサンプル中に含まれるすべての細胞の mRNA 量をまとめて検出するので、通常のデータはサイクリング細胞と非サイクリング細胞の遺伝子発現量の総計となり、これに対しての解析結果は総細胞数あたりのものとなる (図 4, **Total gene dataset**)。マイクロアレイデータセットからサイクリング細胞あたりのデータを得るため、本研究では **Total gene dataset** から **CCS_{cycling}** を構成する遺伝子の発現値を抽出したサブデータセットを作成する工夫を用いる (図 4, **Cycling gene dataset**)。そして、**Total gene dataset**、**Cycling gene dataset** の両方について **Quantile normalization** (Bolstad *et al.*, 2003) を実施する。この結果、**Total gene dataset** では全遺伝子の発現量を元に正規化が行われ、**Cycling gene dataset** では **CCS_{cycling}** 構成遺伝子の発現量だけに従って正規化が行われる。**CCS_{cycling}** はサイクリング細胞で優先的に発現する遺伝子のみで構成されているため、非サイクリング細胞由来の遺伝子発現の影響は **Cycling gene dataset** では限定的となっているはずである。

この操作の後、各 **CCS** スコアをそれぞれのデータセットについて計算する (1. 5. 3 シグネチャースコアとデータ可視化の項参照)。**Total gene dataset** に対しての **CCS_{cycling}** スコアと **CCS_{phase}** スコアは、サンプル中に含まれる総細胞あたりのサイクリング細胞数比率と各細胞周期の細胞数比率をそれぞれ意味することになる。一方、**Cycling gene dataset** に対しての **CCS_{phase}** スコアは上で述べた原理によって、サイクリング細胞あたりの各細胞周期の細胞数比率を意味することになる。尚、**Cycling gene dataset** における **CCS_{cycling}** スコアは、サイクリング細胞あたりのサイクリング細胞の割合を意味するので、常に一定の値となるはずである。

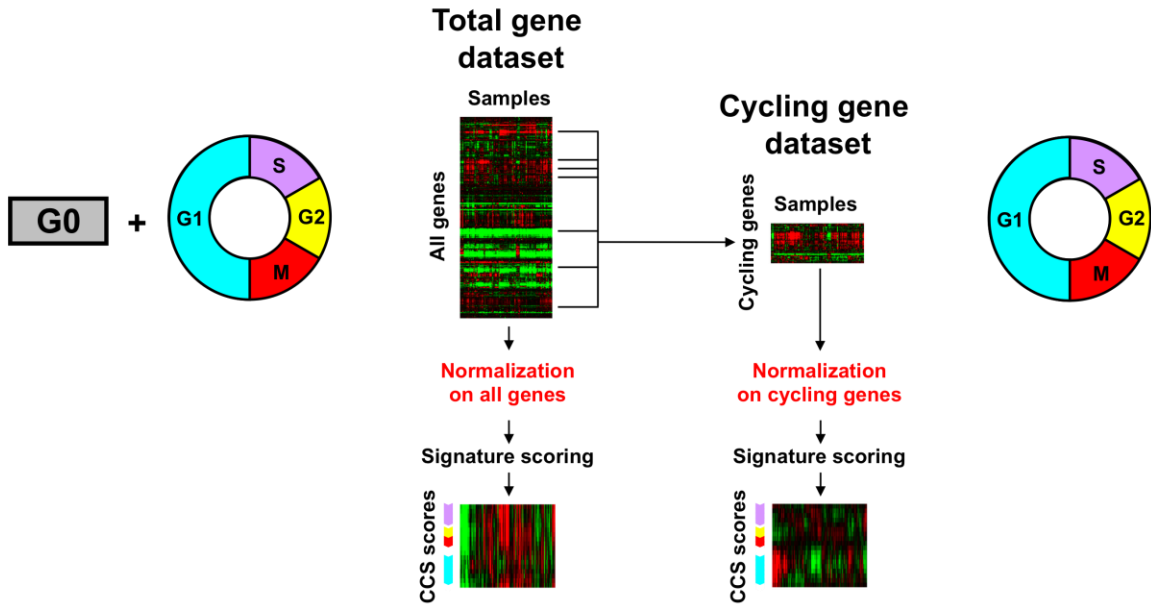


図4 CCS法におけるスコア計算までの流れ。与えられた Total gene dataset から、 $CCS_{cycling}$ 遺伝子の発現値のみを抽出することによって Cycling gene dataset を作成する。それぞれのデータセットが個別に正規化され、CCS スコアが計算される。

1. 3 結果

1. 3. 1 細胞周期データセットを使ったコンセプト証明

はじめに、CCS 作成の元データである Whitfield *et al.* (2002) の細胞周期データセットを解析したところ、予想通り細胞周期分布を推定することができた (図5)。

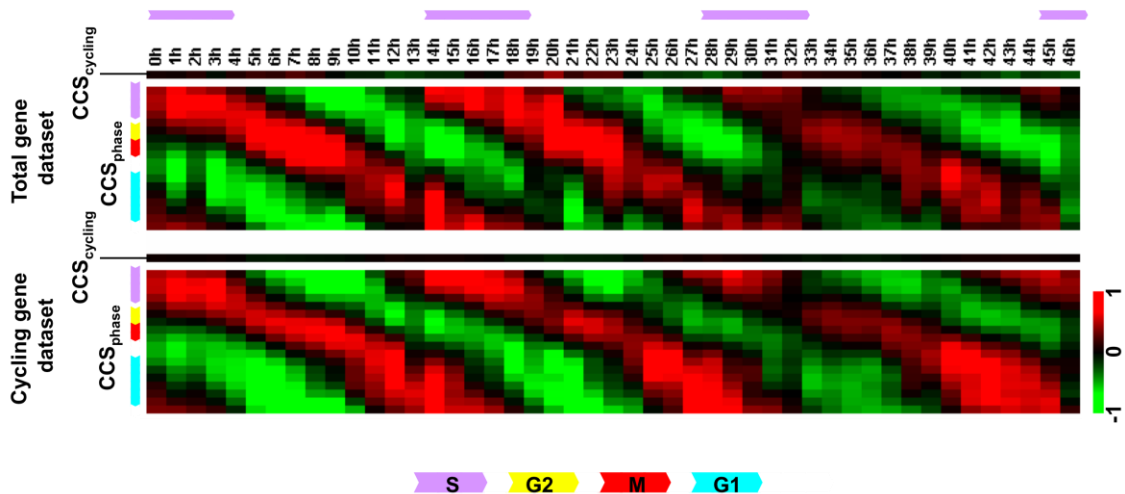


図5 Whitfield *et al.* データセットでの細胞周期分布予測。Total gene setとCycling gene datasetの両方についてCCSスコアを計算した。各列は実験サンプルを表し、各行がそれぞれのCCSに対応する。赤は対応するフェーズの細胞が相対的に多いことを示し、緑は少ないことを示す。細胞周期フェーズは色で対応づけてある (S; 紫、G2; 黄色、M; 赤、G1; 水色)。最上部にある紫のバーは、原著論文で推定されたS期を示す。

CCS 法が他の独立データセットに対しても有効であることを確認するため、HCT116 大腸癌細胞株を用いて細胞同調実験を行い (1. 5. 1 細胞培養および細胞同調の項参照)、Affymetrix 社のマイクロアレイで発現量を測定し、得られたデータを解析した (1. 5. 2 マイクロアレイ実験の項参照)。この実験ではほとんどの細胞がサイクリング細胞であると考えられるが、予想通り Total gene dataset と Cycling gene dataset の両方に類似したヒートマップパターンが観測された (図6)。

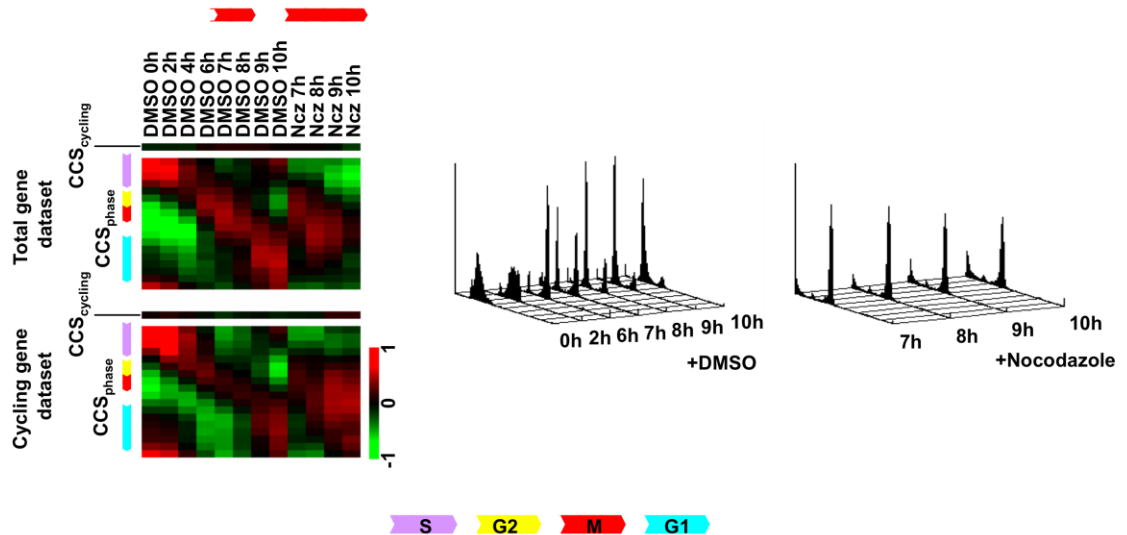


図6 細胞周期を同調させた HCT116 細胞の解析。Thymidine ブロックからのリリース後 0、2、4、6、7、8、9、10 時間の各タイムコースの細胞 (DMSO)、および Nocodazole 処理後 7、8、8、10 時間の細胞 (Ncz) を CCS 法および DNA フローサイトメトリーで解析した。最上部の赤いバーは推定される M 期

DMSO 添加のコントロール群では CCS_{phase} スコアのピークが細胞周期進行に従って移行していく様子が確認できた (図6, DMSO 0–10h)。一方、有糸分裂阻害剤 Nocodazole で処理した群ではピークが M 期周辺で停止する様子が観察できた (図6, Ncz 7-10h)。これらのパターンは DNA フローサイトメトリーによる測定値 (図6) とも良く一致していた。こうした結果は、CCS 法が異

なる細胞株・プラットフォームから得られたデータセットに対しても細胞周期分布測定能を持つことを示した。

1. 3. 2 非サイクリング細胞を含むデータセットを使ったコンセプト証明

癌の生検サンプルには、サイクリング細胞だけではなく、様々な数の非サイクリング細胞が含まれる (Baker *et al.*, 1995)。理論的には、サンプル中のサイクリング細胞の割合が変化すれば、静止期以外のすべての細胞周期フェーズにある細胞数が同じ割合だけ変化すると考えることができる。サンプル中のサイクリング細胞の割合変化がどのように CCS スコアに表れるか調べるため、ヒト乳腺上皮細胞 (Human mammary epithelial cell; HMEC) を Leucine-rich extra cellular 培地で培養した Fournier *et al.* (2006) データセットを解析した。

この細胞培養系では、HMEC はまず指数関数的に細胞数を増やし (day 3)、立体的な組織構造を形成した後、静止状態に入る (day 7) (Petersen *et al.*, 1992; Fournier *et al.*, 2006)。原著論文に記載されている DNA フローサイトメトリーでの測定値を確認したところ、タイムコースに沿って S 期分画は $15\% \pm 5.1$ (day 3) から $5.5\% \pm 0.5$ (day 7)、G2+M 期分画は $12\% \pm 1.1$ (day 5) から $7\% \pm 2.5$ (day 7) (day 3 のデータは掲載なし) まで減少していた。一方、G0+G1 期分画は $73\% \pm 6.3$ (day 5) から $86\% \pm 4.6$ (day 7) へと増えていた。DNA フローサイトメトリーは G1 期細胞と G0 期細胞を区別できないため、断定はできないものの、HMEC は徐々に増殖停止していることと、次に示す Total gene dataset における CCS_{G1} スコアの減少から、この増加は主に G0 期細胞数の増加によるものと考えられる。

CCS 法を使った解析では、HMEC がサイクリング状態 (day 3) から非サイクリング状態 (day 7) へ移行するのに伴って、Total gene dataset の CCS_{cycling} スコアと CCS_{phase} スコアは一様に減少した (図 7, 上部パネル)。この結果は、サンプル中のサイクリング細胞の割合変化は Total gene dataset における CCS_{cycling} および CCS_{phase} スコアの一様な変化として現れることを示した。意外なことに、Cycling gene dataset に対するヒートマップでは、 CCS_{G1} スコア

は day 7 (図7, 下部パネル) にかけて上昇していた。発生の完了や栄養飢餓など、G0 期が誘導される条件のもとで G1 期が延びることが知られているが (Prather *et al.*, 1999; Nygren *et al.* 2006)、この CCS_{G1} スコア上昇はそうした G1 延長によるものかもしれない。

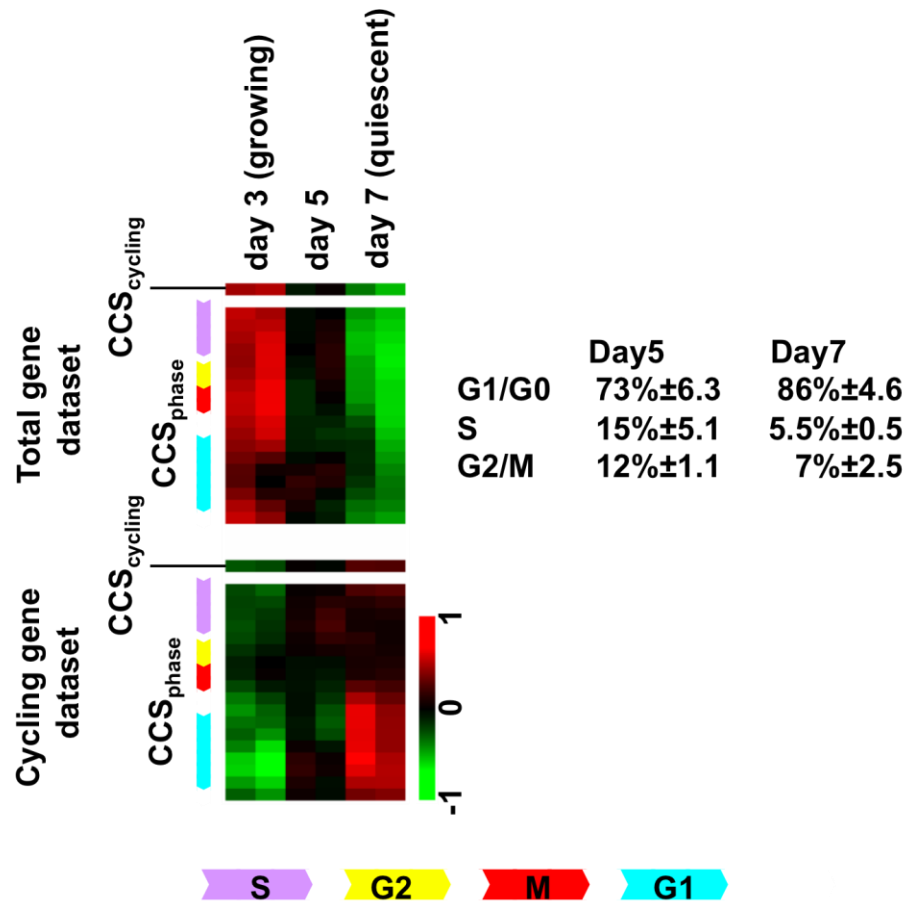


図7 Fournier *et al.* データセットの解析。HMEC を 3D 培養するこのシステムでは、細胞は急速に増加した後 (day 3)、静止状態に入る (day 7)。

観察を強固にするため、Cam *et al.* (2004) データセットを追加的に解析した。これは増殖中の T98 乳癌細胞株を血清飢餓状態に移し、G0 期を誘導した条件の発現プロファイルである。結果は HMEC のものとほぼ同様、サイクリング細胞 (Growing) が栄養飢餓 (Starved) に入った後、Total gene dataset で $CCS_{cycling}$ および CCS_{phase} スコアの一様の減少が観測された (図8, 上部パネ

ル)。さらに、G1 期の長期化を示唆する **Cycling gene dataset** における CCS_{G1} スコア上昇も観測された (図 8, 下部パネル)。

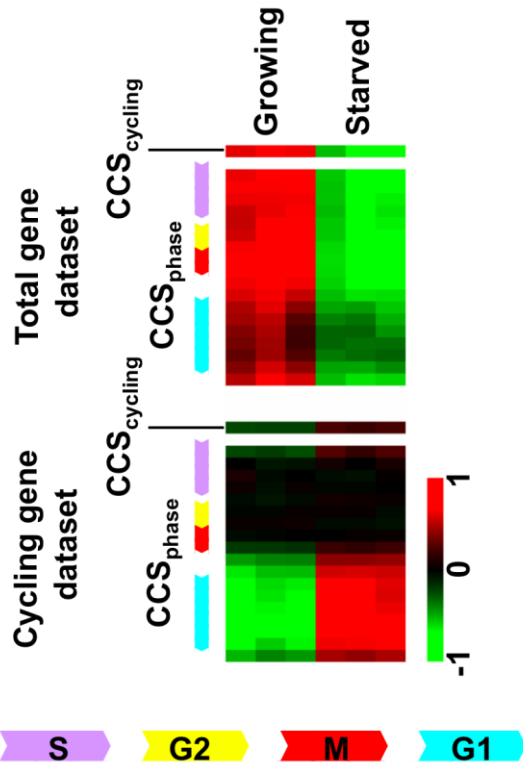


図 8 Cam *et al.* データセットの解析。T98 乳癌細胞の増殖時、および血清飢餓時のプロファイルを CCS 法で調べた。

1. 3. 3 マウスデータセットでの動作確認

細胞周期機構は非常に精密にできており、哺乳類間では高度に保存されていると考えられている (Harper and Brooks の総説, 2005)。ヒト細胞周期データセットから作成した CCS が近縁種に適用できるかどうか確認するため、Yamamoto *et al.* (2006) データセットを解析した。このデータセットは栄養飢餓状態に置いた NIH3T3 マウス線維芽細胞を、増殖誘導因子である Fibroblast Growth Factor (FGF) で刺激・細胞周期を強制的に再開させ、S 期まで進行する過程の遺伝子発現を追ったものである。ヒト CCS で解析したところ、FGF 刺激後、非サイクリング細胞の割合が減っていく一方 (図 9, 上部パネル)、細胞集団で優勢なフェーズが G1 期から S 期へ移っていく様子 (図 9, 下部パネル)

が明確に観察できた。これらの結果から、本研究で作成したヒト CCS がマウスデータセットの解析にも適用できることが示された。

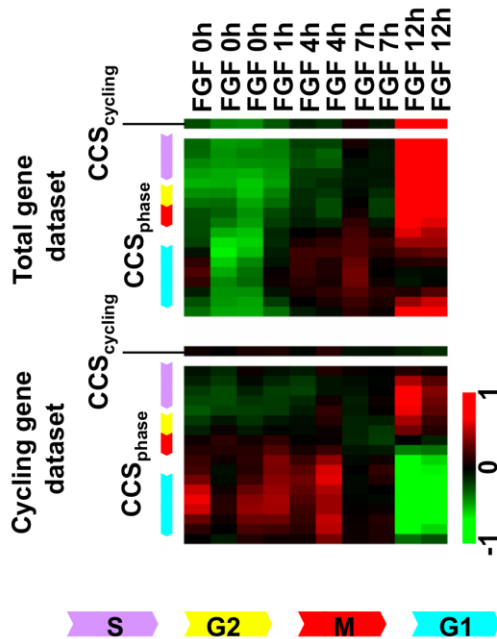


図9 Yamamoto *et al.*データセットの解析。このシステムでは血清飢餓状態のマウス NIH3T3 細胞を FGF で刺激し、細胞周期に再入させている。

1. 3. 4 マウス腫瘍モデルデータセットの解析

癌は様々な原因から生じ、その増殖能力や進行速度は多様である。癌の発生イベントの違いが細胞周期分布に異なった影響を与えるかどうかを調べるため、CCS 法を Herschkowitz *et al.* (2007) データセットに適用した。このデータセットは13の異なる発癌イベントから生じたマウス乳癌モデル (n=122)、および正常乳腺 (n=19) のプロファイルから成るものである。この実験には、1) 同じモデル内で同様の癌化プロセスにより発生すると考えられる **Homogeneous model** と、2) 同じモデル内であっても二次的に異なる癌化イベントを引き起こして発生すると見られる **Heterogeneous model**、の両方が含まれている。

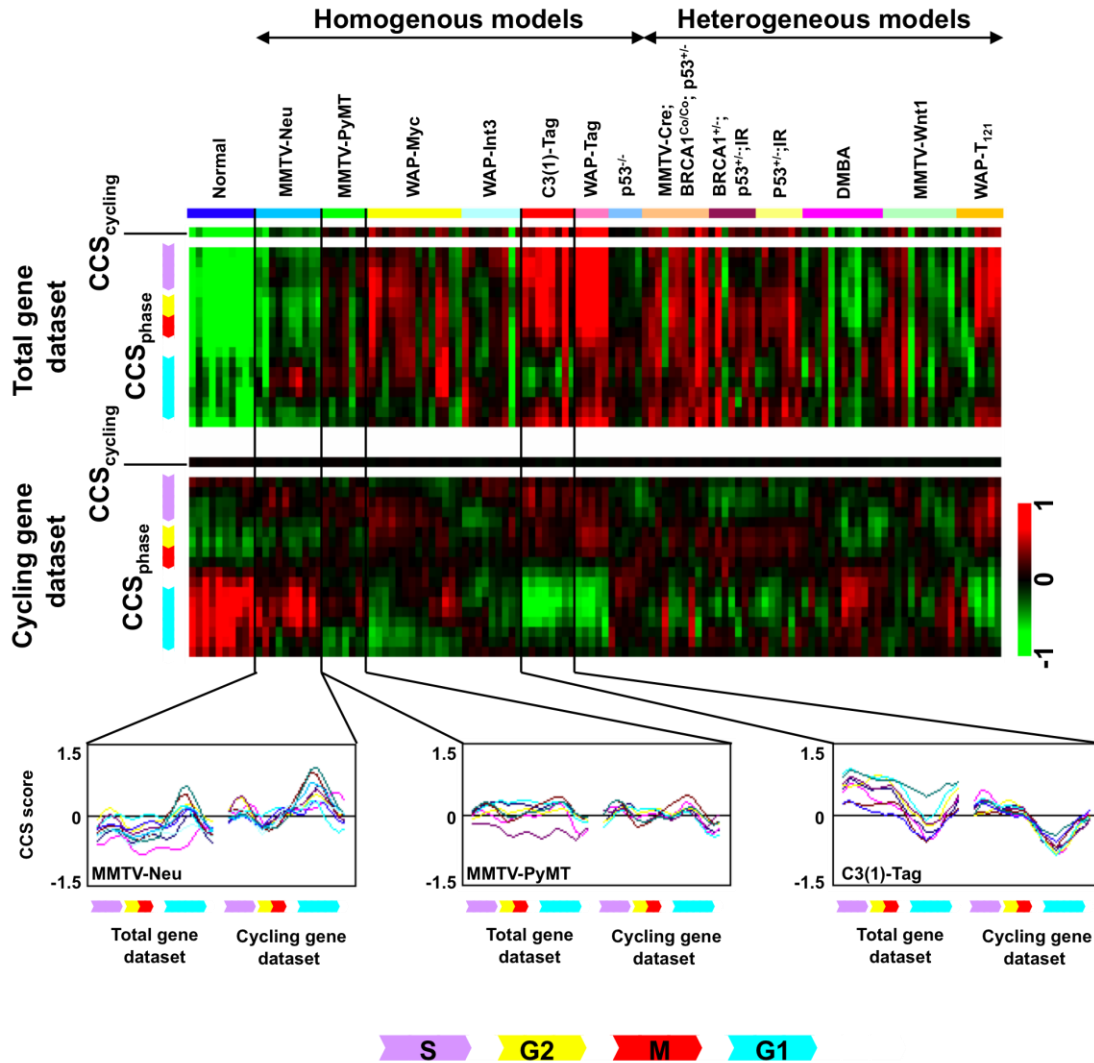


図10 Herschkowitz *et al.*データセットの解析。正常乳腺および13種のマウス腫瘍モデルから成る122のプロファイルを調べた。Herschkowitz *et al.*が定義した Homogeneous・Heterogeneous クラスによってモデルを並べてある。下部のプロットは MMTV-Neu、MMTV-PyMT、および C3(1)-Tag モデルについて CCS_{phase} スコア示したもの。X軸は細胞周期フェーズを、Y軸は各 CCS スコアの大きさを表す。

Total gene dataset において、サイクリング細胞の割合を示す CCS_{cycling} スコアは正常乳腺サンプルで一貫して低く、他方、癌モデルでは様々な度合いで正常より高い値となっていた (図10)。実質的な癌化イベントが異なると考えられる Heterogeneous model がバラつきのある CCS_{cycling} スコアと CCS_{phase} スコアを示したのは予想通りであった。しかし、バラつきは、Simian virus 40 由来の発癌性抗原 Tag を用いたモデルが高スコア、癌遺伝子 ERBB2 を用いた Neu

モデルが低スコアを示すなどの傾向はあったものの、各 Homogeneous model においても見られた。

観察を **Cycling gene dataset** へ移してみると、面白いことに、**Myc** 遺伝子を用いた系を除くそれぞれの Homogeneous model 間に、類似の CCS_{phase} スコアパターンが認められた (図 1 0)。これを詳細に見るため、いくつかのモデルについて、**Total gene dataset** と **Cycling gene dataset** それぞれにおける CCS_{phase} スコアをプロットした (図 1 0, 最下部)。図は各モデルに特定の細胞周期分布があることをはっきりと示している。例えば、**Neu** モデルは高い CCS_{G1} と低い CCS_{S-G2-M} スコアで特徴付けることができ、**Tag** モデルはその反対のパターンで識別できる。例外的に、**Myc** モデルは二つの異なった細胞周期分布パターンを示した (図 1 1)。この理由は明確でないが、**Myc** 遺伝子はゲノムを不安定化し、追加的な癌化イベントを誘発する (**Dominant mutator effect**) との報告があるため (Felsner and Bishop, 1999)、一部のサンプルでそうした効果が発揮されたのかもしれない。

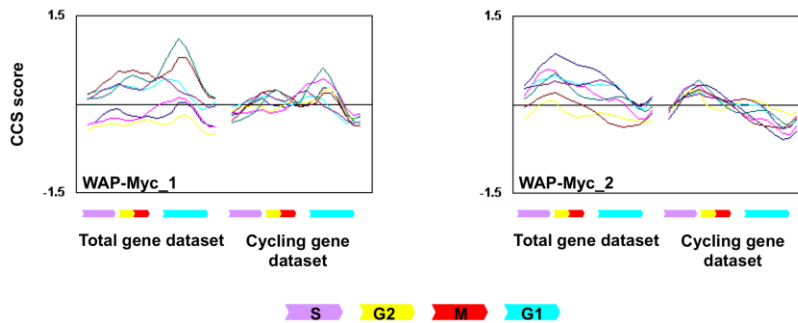


図 1 1 WAP-Myc モデルの CCS スコアのプロット。

すべてのモデルにおいて、**Total gene dataset** のプロットは垂直移動をすする形のバラつきを生じていた (図 1 0, 1 1)。これは、**HMEC** と **T98** 細胞のデータで見てきたように、サイクリング細胞の存在比が影響しているためと考えられる。一方、非サイクリング細胞の影響を限定した **Cycling gene dataset** では、垂直方向へのバラつきは最小限に抑えられ、パターン比較が容易となっていた。

これらの結果は次の二点を示す。(イ) 癌化イベントの違いは細胞周期分布へと反映される、(ロ) 非サイクリング細胞の影響を考慮に入れた細胞周期分布はこの違いをより明確に示す。既存の細胞周期測定法は、少数の測定値に依存する、あるいは、非サイクリング細胞を区別することができないため、こうした特徴を見分けるのが難しい。一方、CCS法はこれを可能にする。

1. 3. 5 ヒト乳癌データセットの解析

本研究の総仕上げとして、CCS法をヒトの乳癌パネル (n=249) である Ivshina *et al.* (2006) データセットに適用した。

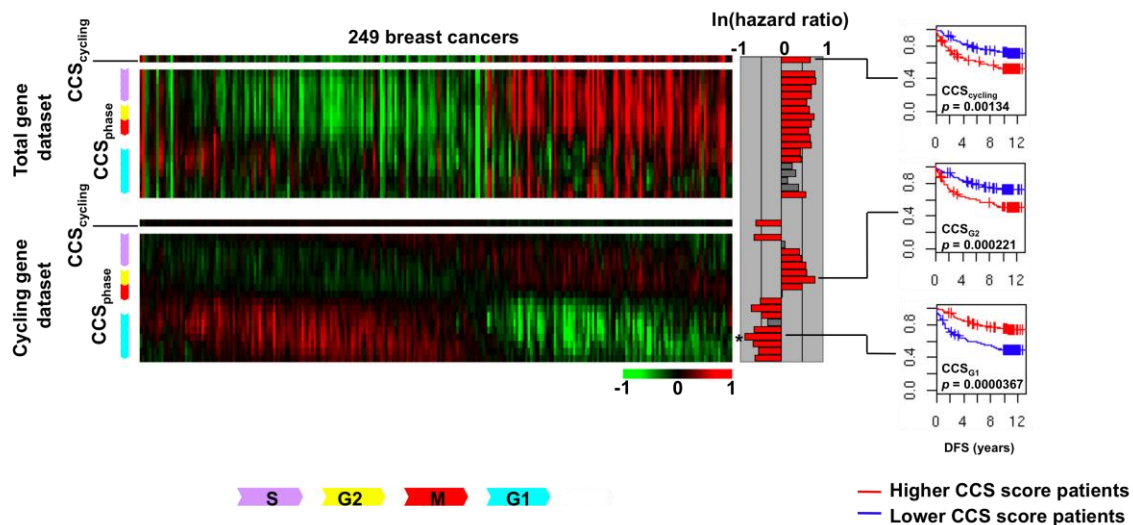


図 12 Ivshina *et al.* データセットの解析。249 の乳癌患者データは Cycling gene dataset における CCS_{phase} スコアのピーク位置によって整列してある。患者をそれぞれの CCS スコアの中央値によって二分割し、それぞれのグループの DFS に対するリスクを Log-rank テストと Cox model で評価した。Log-rank テストで 5% 水準の有為差のついたグループ間でのハザード比は赤色で示してある。最も高い P 値にはアスタリスクを配置してある。

Total gene dataset では様々な CCS_{cycling} スコアが認められた (図 12, 上部パネル)。これまでの観察から、これはサンプル中のサイクリング細胞存在比の違いを反映するものと推測することができる。Cycling gene dataset のヒートマップでは "rolling wave" パターンが観察できた (図 12, 下部パネル)。Total gene dataset で高い CCS_{cycling} スコアを示す患者は、Cycling gene dataset で高

CCS_{S-G2-M} スコアと低 CCS_{G1} スコアを持つ傾向があったが、いくつかの例外が存在していた。これはマウス腫瘍モデルの解析で見た CCS_{phase} スコアの垂直移動を連想させる。

S phase fraction や Mitotic index など、既存の細胞周期測定値が癌の悪性度と相関することが様々な研究で示されている。Ivshina *et al.* データセットで患者の臨床情報が利用可能であったため、CCS スコアと患者予後の間に関連性があるかを調べた (1. 5. 4 生存解析の項参照)。患者をそれぞれの CCS スコアの中央値によって二つのグループに分け、次にグループ間の無病生存率 (Disease free survival; DFS) に関するリスク差を、Kaplan-Meier 曲線、Log-rank テストおよび Cox model によって評価した。Total gene dataset の CCS_{cycling} スコアは予後不良に対し、強い予測能を示した (図 1 2, 右パネル, Hazard ratio; HR = 1.98, $P = 0.00134$)。これはサイクリング細胞が多いほど、臨床結果がより悪くなるという一般的な認識と一致するものである。一方、Total gene dataset における CCS_{S-G2-M} といくつかの CCS_{G1} スコアも予後不良の予測能があった。興味深いことに、Cycling gene dataset における CCS_{G1} スコアは、予後良好に対する予測能を示したうえ、実施したすべての生存解析の中で最も高い P 値を示した (図 1 3, 右パネル, HR = 0.41, $P = 0.0000367$)。

これらの結果がデータセット特異的である可能性を排除するため、CCS 法を Langerød *et al.* (2007) 乳癌パネル (n=80) へも適用した (図 1 3)。

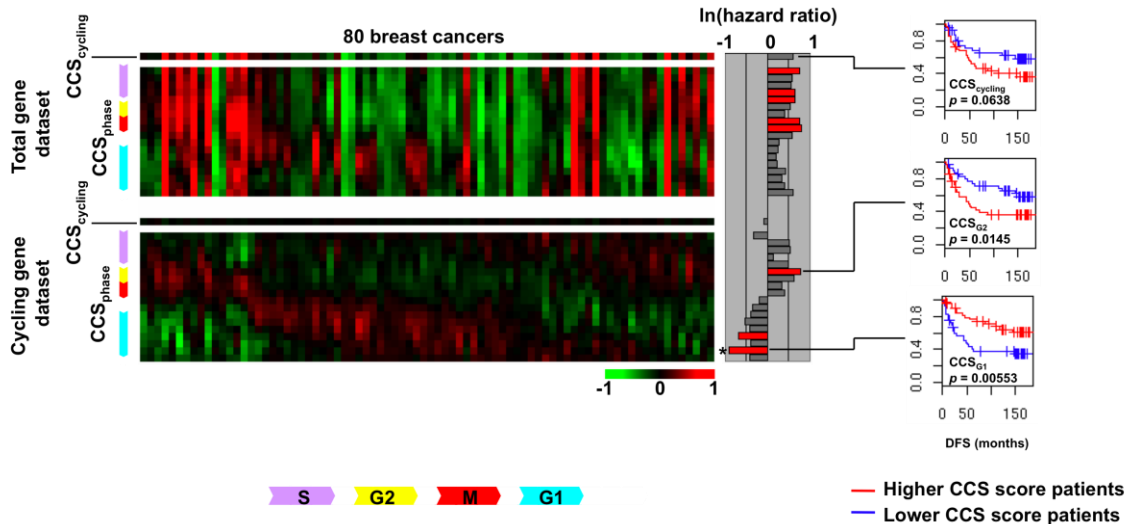


図 1 3 Langerød *et al.* データセットの解析

その結果は Ivshina *et al.* データセットのものと同様であった。Total gene dataset では $CCS_{cycling}$ スコアのバラつきが観察された。Total gene dataset で高い $CCS_{cycling}$ スコアを示す患者が Cycling gene dataset で高い CCS_{S-G2-M} と低い CCS_{G1} スコアを持つ傾向、そして、若干の例外が示された。さらに Cycling gene dataset での CCS_{G1} スコアが、良好な DFS に対して予測能があり、全生存解析中で最も高い有意性を示すことも一致していた (図 1 3, $HR = 0.41$, $P = 0.00553$)。これらの結果から、次のことが示された。(イ) 腫瘍中のサイクリング細胞の割合はバラついている、(ロ) 腫瘍中のサイクリング細胞の割合はサイクリング細胞あたりの細胞周期分布と関連するが例外もある、(ハ) サイクリング細胞あたりの細胞周期分布は腫瘍中のサイクリング細胞の割合よりも患者予後に対する高い予測能を持つ。

1. 4 考察

本研究で筆者は、サイクリング細胞と非サイクリング細胞の両方を考慮して、マイクロアレイデータから細胞周期分布を推定する遺伝子シグネチャーに基づいた方法を開発した。この方法は癌に関する二つの貴重な情報を提供する。

情報の一つはサンプル中の全細胞に対するサイクリング細胞の存在比である。Mitotic index、S phase fraction、細胞周期マーカーに対する IHC など、現在の細胞周期測定法は「不良な予後につながる高い増殖性腫瘍は多くのサイクリング細胞を含む」という前提のもとに成り立っている。確かに、ヒト乳癌データセットの解析では、Total gene dataset の高い CCS_{cycling} スコア（サンプル中に多くのサイクリング細胞が含まれていることを示唆する）は予後不良と強く関連していた（図 12, 13 CCS_{cycling}）。しかし、Whitfield *et al.* (2002) は、いくつかの細胞周期遺伝子の発現が乳癌のグレードと相関しないことを観察している。また、G1 期は細胞周期の一部であるにも関わらず、このフェーズのマーカーである Cyclin D1 の発現は乳癌の予後良好と関連することが度々報告されている（Landberg *et al.*, 1997; Barnes and Gillett, 1998; Colozza *et al.*, 2005; Beresford *et al.*, 2006）。サイクリング細胞数の増加は、すべての細胞周期における細胞数を一律に増加させるはずなので、サイクリング細胞の割合だけを考えるモデルでは、こうした観察を十分に説明できない。

もう一つの情報はサイクリング細胞あたりの細胞周期分布である。多くの癌化イベントが細胞周期の各フェーズでの所要時間を攪乱することが知られている。例えば、v-H-Ras、v-Src、v-Raf、cyclin D1、cyclin E、c-myc といった癌遺伝子の活性化、Pten など癌抑制遺伝子の不活性化は G1 期を短縮する（Karn *et al.*, 1989; Wimmel *et al.*, 1994; Liu *et al.*, 1995; Sun *et al.*, 1999）。SV40-Tag や HTLV-1 Tax などの癌化ウイルスの抗原が発現した場合についても、同様の報告がある（Sladek and Jacobberger, 1992; Lemoine and Marriott, 2001）。他方、Lzts1 と Lats2 の欠失は M 期を短くすることが報告されている（Vecchione *et al.*, 2007; Yabuta *et al.*, 2007）。このように、それぞれの癌化イベントが細胞周期の各フェーズに異なった影響を与えると、それはサイクリング細胞あたりの細胞周期分布に直接反映される。実際、マウス腫瘍モデルの解析では、異なる癌化イベントが特定の細胞周期分布パターンを持つことが確認できた（図 10）。これは、非サイクリング細胞の影響を除いた細胞周期分布が、癌の特徴づけに利用できることを示唆する。

各癌でサイクリング細胞の絶対数と細胞周期の時間配分が同時に異なっていることを考慮すると、Whitfield *et al.* (2002) の観察や、度々報告される Cyclin D1 発現と予後良好との関連 (Landberg *et al.*, 1997; Barnes and Gillett, 1998; Colozza *et al.*, 2005; Beresford *et al.*, 2006) は上手く説明できるようになる。そのモデルを図 1 4 に示す。

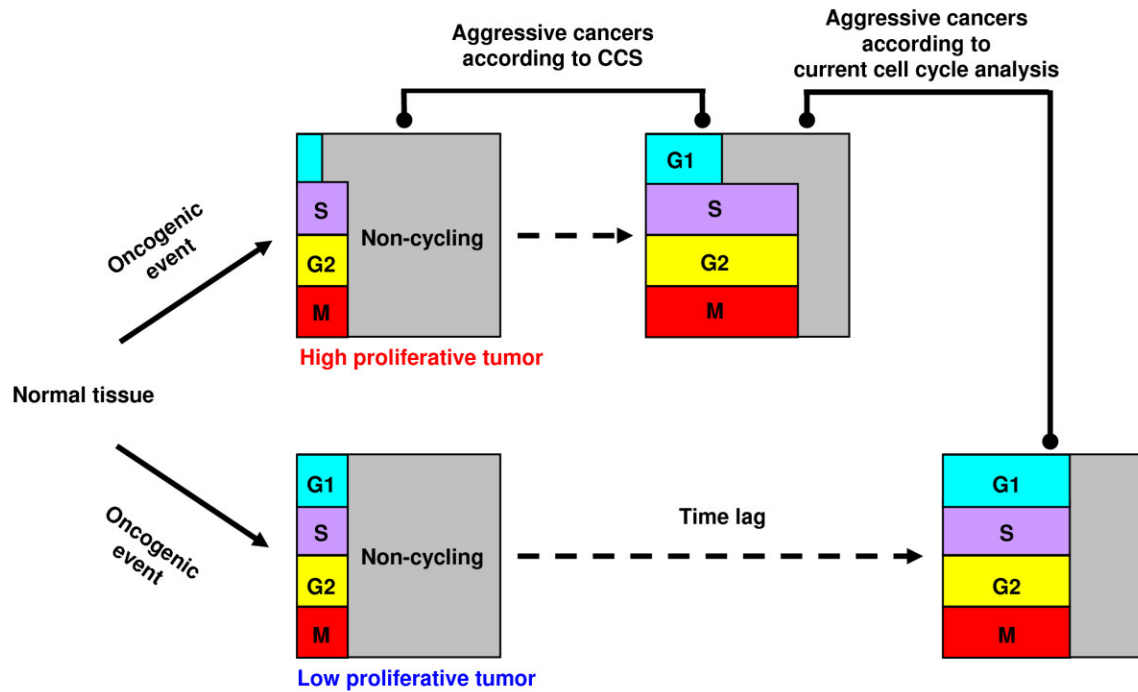


図 1 4 癌が異なる細胞周期分布を持つモデル。癌化イベントは独自の方法で細胞周期を攪乱し、異なった増殖能の癌を生み出す。サイクリング細胞の数は増殖能力に大きく影響を受けるが、例外が低い確率で存在する。CCS 法は非サイクリング細胞の存在を考慮して細胞周期分布を解析することができるため、こうした癌を識別することが可能である。

それぞれの癌化イベントは独自の方法で細胞分裂機構を攪乱し、細胞周期分布と増殖スピードを変化させる。高増殖性癌では細胞分裂が急速に起こるため、短時間で多くのサイクリング細胞が生まれる (図 1 4 上段)。一方、低増殖性癌では、サイクリング細胞の絶対数が増えるまでに時間がかかる (図 1 4 下段)。このメカニズムで、癌の増殖速度とサイクリング細胞数は関連することとなる。しかし、癌の発生から発見までの時間は一定ではない。診断に至るまでに長い時間を経たため、多くのサイクリング細胞を得た低増殖性の癌や、早期発見

されたため、サイクリング細胞が増えきっていない高増殖性の癌が低い確率で存在すると考えられる。

このような癌を検出するのに、現在の細胞周期分布測定法は不十分である。**Mitotic index** と **S phase fraction** はサイクリング細胞と非サイクリング細胞を分けて考えることができない。IHC に関しては、**CCS** 法と同様の理論に基づき、同一サンプル中で複数のマーカーを調べ、細胞周期分布を見積もる **Combinatorial IHC** が提唱され始めているものの (Williams and Stoeber, 2007)、まだ発展途上である。尚、マイクロアレイデータから細胞周期を解析する技術として、Lu *et al.* (2003) が開発した **Expression deconvolution** という手法がある。この手法は、細胞周期の各フェーズにおける細胞存在比を説明する変数を用意し、複数の細胞周期遺伝子の発現を連立方程式として表した上で、それらを最も満たす最適解を探索するというものである。酵母の細胞周期分布を調べるために開発されたこの方法は、癌の細胞周期分布を解析する目的でも、**CCS** 法に匹敵、あるいはより有力なツールとなる可能性がある。しかし、局所最適化問題を避けながら多数の方程式の最適解を見つけるためには、膨大な計算量が必要となる。特に今回の研究のように高い解像度で細胞周期分布解析を行おうとする場合は (Liu *et al.*の細胞周期分割数は5、本研究では18+1)、可能な組み合わせが乗数的に増加していくため、実用化にはコンピューターリソースおよびアルゴリズム面での技術革新が必要である。以上より、実際の癌データを使って有用性が検証された **CCS** 法は、現時点で最も先行した手法と言える。

1. 5 材料および方法

1. 5. 1 細胞培養及び細胞同調

HCT116 大腸癌細胞株 (ATCC 社) を、10% FBS (JBS) を添加した McCOY'S 5A MEDIUM MODIFIED 培地 (Sigma Aldrich 社)、37°C、5% CO₂ の条件下で培養した。これを 2 mM 濃度の Thymidine で19時間処理し、9時間通常培地に戻した後、再度16時間処理して細胞周期を同調させた。得られた同調細胞を通常培地で洗浄した後、0.1 mg/ml の Nocodazole (Sigma Aldrich 社)、

もしくは DMSO で処理し、それぞれから 7、8、9 および 0、2、4、6、7、8、9 時間後に採取した。採取細胞をマイクロアレイ実験及び DNA フローサイトメトリーで分析した。

1. 5. 2 マイクロアレイ実験

Total RNA を逆転写し、蛍光ラベル化した後、メーカー指定の方法に従って Human Genome U133 Plus 2.0 アレイ (Affymetrix 社) にハイブリダイゼーションさせた。スキャン後、得られた生データを GC-RMA アルゴリズムを用いて処理し、mRNA 発現量を示すシグナル値を計算した。

得られたマイクロアレイデータは、全て NCBI が運営する Gene Expression Omnibus (GEO) データベース (Barret *et al.*, 2007) に登録した (<http://www.ncbi.nlm.nih.gov/projects/geo/query/acc.cgi?acc=GSE14103>)。

1. 5. 3 シグネチャースコアとデータ可視化

Total gene dataset として所与のマイクロアレイデータセットをそのまま使用した (図 4)。Cycling gene dataset は、Total gene dataset から $CCS_{cycling}$ の構成遺伝子の発現値を抽出することによって作成した (図 4)。Total gene dataset、Cycling gene dataset それぞれについて、以下のステップで CCS スコアを計算した。発現値を Log 変換後、Quantile normalization (Bolstad *et al.*, 2003) を適用し、プローブごとに Z 変換を行って標準化した。各 CCS についてそれぞれを構成する遺伝子のシグナル値を平均し、CCS スコアとした。頑強なスコアを得るため、それぞれの CCS_{phase} スコアを隣接する CCS スコアで 2 度平均し平滑化した。ヒートマップは Java Treeview (Saldanha, 2004) によって作成した。ヒト CCS を用いたマウス腫瘍データセットの解析は、HomoloGene データベース (Wheeler *et al.*, 2008) のヒト-マウスのオーソログ情報を用い、遺伝子 ID をマッピングすることで実現した。ヒト乳癌データセットの解析結果では、Cycling gene dataset の CCS_{phase} スコアのピーク位置によってサンプルの並び替えを行った。

1. 5. 4 生存時間解析

患者を各 **CCS** スコアの中央値によって二分し、**DFS** に対するグループ間のリスク差を評価した。**Kaplan-Meier** 法を使って生存曲線を描き、**Log-rank** テストで二群間の生存率の差を検定、**Cox** 単変量解析でハザード比を求めた。これらの計算には **R** (<http://www.r-project.org/>) の **survival** パッケージを使用した。

第二章

Prognoscan : 遺伝子発現と癌患者予後の関連の横断解析データベース

2. 1 背景

目下、様々な遺伝子について、癌との潜在的な関連が報告されている。こうした遺伝子を評価する主要な方法の一つは、その発現と患者予後との関連を調べることである。近年、臨床情報の付随した癌マイクロアレイデータが大量に公開され、遺伝子発現と患者予後とを結びつける機会が提供されている。しかし、データ解析のための効率的なプラットフォームの開発が遅れているため、その活用は進んでいない。

一般的に生存解析は1) 患者を群分けする、2) 群間のリスクを比較する、の2ステップから成る。遺伝子発現のような連続的な測定値を元に生存解析を行う場合、発現の高い順に二等分、三等分といった便宜的な分け方がよく用いられる。しかし、これは必ずしも生物学的意味を反映するものではない。従って、仮説をあらかじめ持たない場合、適切な閾値を決めることが解析上の困難の一つとなる。ITTACA (Elfilali *et al.*, 2006) やREMBRANDT (<http://caintegrator-info.nci.nih.gov/rembrandt>) といった癌マイクロアレイデータの先駆的データベースでの解析が効率的に行えないのはこのためである。

Minimum *P*-value approachは連続的な測定値の中から、リスク分離を行う最適な閾値を見つけるための網羅的な手法であり、癌ではこれまで腫瘍サイズ、細胞周期指標、遺伝子コピー数などと予後との関連性の解析で有効性が示されている (Abel *et al.*, 1984; Kronqvist *et al.*, 2000; Jensen *et al.*, 2008)。最近、BUB1、HOXB4、MYCといった遺伝子で、癌へ寄与する発現量の閾値があることが報告され (Will *et al.*, 2006; Jeganathan *et al.*, 2007; Shachaf *et al.*, 2008)、この手法を癌マイクロアレイデータへ適用する妥当性が示された。こうした状況を踏まえて、筆者は、1) 臨床情報の付随した癌マイクロアレイデータの網羅的なコレクション、2) minimum *P*-value approachを用いた遺伝子発現に基づく

生存解析ツール、の二つの特徴を持つデータベース「Prognoscan」の開発に取り組み、成果物をインターネット上で公開した：

(<http://gibk21.bse.kyutech.ac.jp/Prognoscan/index.html>)。

Prognoscanは遺伝子発現と癌患者予後の関連の横断解析を可能にし、癌研究を加速するプラットフォームとなることが期待される。

2. 2 結果

2. 2. 1 癌マイクロアレイデータの収集

患者予後の臨床情報の付随した癌のマイクロアレイデータセットを Gene Expression Omnibus (GEO) (Barret *et al.*, 2007)、ArrayExpress (Parkinson *et al.*, 2007)、および各研究室のウェブサイトなどの公開データベースから収集した。データ収集に当たっては次の評価基準を用いた。1) 各患者に生存イベントの有無と生存期間が注釈されていること、2) 生存解析が可能な十分大きいサンプルサイズであること、3) ゲノム規模のプラットフォームから得られていること、4) プローブに一般的なID (例: Gene symbol、GenBank accession number、UniGene ID) が注釈されており遺伝子名と関連付けが可能なこと、5) 欠損値を含まずQuantile normalizationが適切に機能すること。

2009年2月時点でのコレクションは、膀胱癌・血液癌・乳癌・脳腫瘍・食道癌・頸椎癌・腎臓癌・肺癌・卵巣癌などの様々な癌種由来の40以上のデータセットを含む(表1)。これは膀胱癌・乳癌・ブドウ膜黒色腫に特化しているITTACAや、脳腫瘍に特化しているREMBRANDTよりも遥かに網羅的な内容である。

生存解析を適切に行うため、集めたマイクロアレイデータセットのチェックを行った。データセット中にサンプル重複が見つかった場合、任意の代表を一つ残すことで対処した。データセット間での重複は、研究毎に設定された実験デザインには価値があると考え、解消を行わずそのまま使用した。次に、各実験について、イ) コホート・癌種・サブタイプ・エンドポイント・治療歴・病的所見などの研究デザイン、ロ) 試料保管/調製法・プラットフォーム種類・シグナ

ル計算法などの実験条件、の精査を行って情報を整理した。予後データに関しては、全生存（Overall survival; OS）・無再発生存（Relapse free survival; RFS）・無イベント生存（Event free survival; EFS）・無遠隔転移生存（Distant metastasis free survival; DMFS）などの利用可能なエンドポイントを可能な限り採用した。遺伝子発現データはQuantile normalizationを用いて正規化した。さらに、各マイクロアレイのプローブ情報をGEOとArrayExpressから入手し、UniGeneデータベースのID対応表を用いてプローブIDと遺伝子ID（Entrez Gene ID）の対応づけを行った。すべてのテーブルを関連づけた上でMySQLサーバに格納した。

表1 PrognScanのデータコレクション

Dataset	Cancer type	Subtype	Cohort	Author/Contributor	Array type	n	Data source
GSE13507	Bladder cancer	Transitional cell carcinoma	Cheongju	Kim	Human-6 v2	n = 165	GEO
GSE5287	Bladder cancer		Aarhus (1995–2004)	Als <i>et al.</i> , 2007	HG-U133A	n = 30	GEO
GSE12417-GPL570	Blood cancer	AML	AMLCG (2004)	Metzeler <i>et al.</i> , 2008	HG-U133_Plus_2	n = 79	GEO
GSE12417-GPL96	Blood cancer	AML	AMLCG (1999–2003)	Metzeler <i>et al.</i> , 2008	HG-U133A	n = 163	GEO
GSE12417-GPL97	Blood cancer	AML	AMLCG (1999–2003)	Metzeler <i>et al.</i> , 2008	HG-U133B	n = 163	GEO
GSE8970	Blood cancer	AML	San Diego	Raponi <i>et al.</i> , 2008	HG-U133A	n = 34	GEO
GSE4475	Blood cancer	B-cell lymphoma	Berlin (2003–2005)	Hummel <i>et al.</i> , 2006	HG-U133A	n = 158	GEO
E-TABM-346	Blood cancer	DLBCL	GELA (1998–2000)	Jais <i>et al.</i> , 2008	HG-U133A	n = 53	ArrayExpress
GSE2658	Blood cancer	Multiple myeloma	Arkansas	Zhan <i>et al.</i> , 2006	HG-U133_Plus_2	n = 559	GEO
E-TABM-158	Breast cancer		UCSF, CPMC (1989–1997)	Chin <i>et al.</i> , 2006	HG-U133A	n = 129	ArrayExpress
GSE11121	Breast cancer		Mainz (1988–1998)	Schmidt <i>et al.</i> , 2008	HG-U133A	n = 200	GEO
GSE1378	Breast cancer		MGH (1987–2000)	Ma <i>et al.</i> , 2004	Arcturus 22 k	n = 60	GEO
GSE1379	Breast cancer		MGH (1987–2000)	Ma <i>et al.</i> , 2004	Arcturus 22 k	n = 60	GEO
GSE1456-GPL96	Breast cancer		Stockholm (1994–1996)	Pawitan <i>et al.</i> , 2005	HG-U133A		
			Stockholm (1994–1996)	Pawitan <i>et al.</i> , 2005	HG-U133B	n = 159	GEO

GSE2034	Breast cancer		Rotterdam (1980–1995)	Wang <i>et al.</i> , 2005	HG-U133A	n = 286	GEO
GSE2990	Breast cancer		Uppsala, Oxford	Sotiriou <i>et al.</i> , 2006	HG-U133A	n = 187	GEO
GSE3143	Breast cancer		Duke	Bild <i>et al.</i> , 2006	HG-U95A	n = 158	GEO
GSE3494-GPL96	Breast cancer		Uppsala (1987–1989)	Miller <i>et al.</i> , 2005	HG-U133A	n = 236	GEO
GSE3494-GPL97	Breast cancer		Uppsala (1987–1989)	Miller <i>et al.</i> , 2005	HG-U133B	n = 236	GEO
GSE4922-GPL96	Breast cancer		Uppsala (1987–1989)	Ivshina <i>et al.</i> , 2006	HG-U133A	n = 249	GEO
GSE4922-GPL97	Breast cancer		Uppsala (1987–1989)	Ivshina <i>et al.</i> , 2006	HG-U133B	n = 249	GEO
GSE6532-GPL570	Breast cancer		GUYT	Loi <i>et al.</i> , 2008	HG-U133_Plus_2	n = 87	GEO
GSE7378	Breast cancer		UCSF	Zhou <i>et al.</i> , 2007	U133AAofAv2	n = 54	GEO
GSE7390	Breast cancer		Uppsala, Oxford, Stockholm, IGR, GUYT, CRH (1980–1998)	Desmedt <i>et al.</i> , 2007	HG-U133A	n = 198	GEO
GSE7849	Breast cancer		Duke (1990–2001)	Anders <i>et al.</i> , 2008	HG-U95A	n = 76	GEO
GSE9195	Breast cancer		GUYT2	Loi <i>et al.</i> , 2008	HG-U133_Plus_2	n = 77	GEO
GSE9893	Breast cancer		Montpellier, Bordeaux, Turin (1989–2001)	Chanrion <i>et al.</i> , 2008	MLRG Human 21 K V12.0	n = 155	GEO
GSE11595	Esophagus cancer	Adenocarcinoma	Sutton	Giddings	CRUKDMF_22 K_v1.0.0	n = 34	GEO
GSE7696	Glioma	Glioblastoma	Lausanne	Murat <i>et al.</i> , 2008	HG-U133_Plus_2	n = 70	GEO
GSE4271-GPL96	Glioma		MDA	Phillips <i>et al.</i> , 2006	HG-U133A	n = 77	GEO
GSE4271-GPL97	Glioma		MDA	Phillips <i>et al.</i> , 2006	HG-U133B	n = 77	GEO
GSE2837	Head and neck cancer	Squamous cell carcinoma	VUMC, VAMC, UTMDACC (1992–2005)	Chung <i>et al.</i> , 2006	U133_X3P	n = 28	GEO
HARVARD-LC	Lung cancer	Adenocarcinoma	Harvard	Beer <i>et al.</i> , 2002	HG-U95A	n = 84	Author's web site
MICHIGAN-LC	Lung cancer	Adenocarcinoma	Michigan (1994–2000)	Beer <i>et al.</i> , 2002	HuGeneFL	n = 86	Author's web site
GSE11117	Lung cancer	NSCLC	Basel	Baty	Novachip human 34.5 k	n = 41	GEO
GSE3141	Lung cancer	NSCLC	Duke	Bild <i>et al.</i> , 2006	HG-U133_Plus_2	n = 111	GEO
GSE4716-GPL3694	Lung cancer	NSCLC	Nagoya (1995–1996)	Tomida <i>et al.</i> , 2004	GF200	n = 50	GEO
GSE4716-GPL3696	Lung cancer	NSCLC	Nagoya (1995–1996)	Tomida <i>et al.</i> , 2004	GF201	n = 50	GEO
GSE8894	Lung cancer	NSCLC	Seoul	Son	HG-U133_Plus_2	n = 138	GEO
GSE4573	Lung cancer	Squamous cell carcinoma	Michigan (1991–2002)	Raponi <i>et al.</i> , 2006	HG-U133A	n = 129	GEO
DUKE-OC	Ovarian cancer		Duke	Bild <i>et al.</i> , 2006	HG-U133A	n = 134	Author's web site
GSE8841	Ovarian cancer		Milan	Mariani	G4100A	n = 83	GEO
E-DKFZ-1	Renal cell carcinoma		RZPD	Sueltmann	A-RZPD-20	n = 74	ArrayExpress

2. 2. 2 解析アルゴリズム

PrognScanの生存解析では、連続的な遺伝子発現値の中から患者を分ける最適な分割点を見つけるため、minimum *P*-value approach (Abel *et al.*, 1984) を採用した。これはまず、患者を各遺伝子の発現値によって並び替え、次にすべての可能な分割点で二つのグループに分割、そしてグループ間のリスク差をLog-rankテストで計算し、最も顕著な*P*-value (P_{\min}) を与える最適分割点を選択するというものである。

このアプローチについては複数の検定を行うため、偽陽性を増やしてしまう欠点が知られている (Altman *et al.*, 1994; Mazumdar and Glassman, 2000; Holländer and Schumacher, 2001)。従って、Miller and Siegmund (1982) の公式を用い、 P -valueの補正を行った。

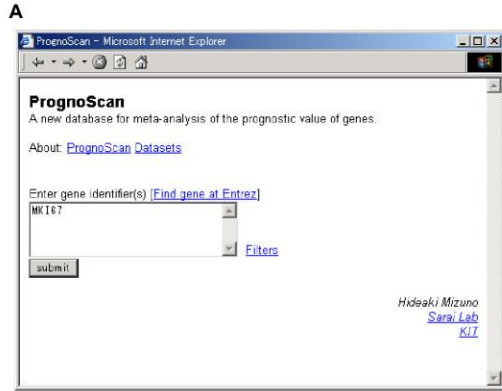
$$P_{\text{cor}} = 4\varphi(z) / z + \varphi(z)\{z - (1/z)\}\log\{(1 - \varepsilon)^2 / \varepsilon^2\}$$

ここで z は正規分布の $(1 - P_{\text{min}} / 2)$ 分位点、 φ は正規分布の確率密度関数、 $[\varepsilon, 1 - \varepsilon]$ は分割点として考慮する分位点範囲を示す。小さすぎるグループ分けを避けるため、PrognoScanでは $\varepsilon = 0.1$ を設定した。所与の遺伝子について、データセット・エンドポイント・プローブの可能な組み合わせのすべてで、この分割点決定と生存解析を行った。この論文では便宜的に、それぞれの組み合わせを「テスト」と呼ぶことにする。尚、一つの遺伝子に複数のプローブが設計されていることがあるため、遺伝子ごとに可能なテスト数が異なることに留意されたい。統計解析およびデータ可視化には、Rパッケージ (<http://www.r-project.org>) を使用した。

2. 2. 3 横断解析の実例

PrognoScanの機能を示すため、三つの横断解析の例を紹介する。最初の例として、良く知られた腫瘍マーカーであるMKI67 (Antigen identified by monoclonal antibody Ki-67) を取り上げる。MKI67は、その発現が非ホジキンリンパ腫などの一部の例外を除き、脳腫瘍・乳癌・肺癌など多くの癌で悪性度と関連することが報告されている遺伝子である (Brown and Gatter, 2002)。

PrognoScanの利用は非常に簡便であり、利用者は遺伝子IDを入力するだけで良い (図 1 5 A)。遺伝子IDを与えると、PrognoScanはデータセット・癌種・サブタイプ・エンドポイント・コホート・データ取得者・プラットフォーム・プローブID・最適分割点・ P_{min} ・ P_{cor} の列を持つ表形式で各テストの結果を表示する (図 1 5 B)。



PROBE ID	CANCER TYPE	SUBTYPE	ENDPOINT	COHORT	DATA OBTAINER	PLATFORM	PROBE ID	N	CORP	P VALUE	CORRECTED P VALUE
S182001	Breast Cancer	Luminal	Overall Survival	TCGA (2012)	TCGA	HT-155A	S182001_1_#	1130	0.01	0.00021	0.00010
S182002	Breast Cancer	Luminal	Overall Survival	TCGA (2012)	TCGA	HT-155A	S182002_1_#	1130	0.77	0.00114	0.00050
S182003	Breast Cancer	Luminal	Overall Survival	TCGA (2012)	TCGA	HT-155A	S182003_1_#	1130	0.83	0.00044	0.00040
S182004	Breast Cancer	Luminal	Overall Survival	TCGA (2012)	TCGA	HT-155A	S182004_1_#	1130	0.57	0.00041	0.00050
S1812007	Breast Cancer	Transcriptome at single nucleotide	Overall Survival	TCGA (2012)	TCGA	HT-155A	S1812007_1_#	1130	0.01	0.00021	0.00010
S1812012	AML	AML	Overall Survival	AMLCG (1999-2005)	Mizuno	HT-155A	S1812012_1_#	1130	0.14	0.01855	0.00500
S1812013	AML	AML	Overall Survival	AMLCG (1999-2005)	Mizuno	HT-155A	S1812013_1_#	1130	0.59	0.00124	0.00500
S1812017	AML	AML	Overall Survival	AMLCG (1999-2005)	Mizuno	HT-155A	S1812017_1_#	1130	0.10	0.00704	0.12400
S1812018	AML	AML	Overall Survival	AMLCG (1999-2005)	Mizuno	HT-155A	S1812018_1_#	1130	0.21	0.01304	0.20500
S1812019	AML	AML	Overall Survival	AMLCG (2006)	Mizuno	HT-155A	S1812019_1_#	1130	0.77	0.01993	0.30400
S1812017	AML	AML	Overall Survival	AMLCG (2006)	Mizuno	HT-155A	S1812017_1_#	1130	0.80	0.00359	0.30400
S1812018	AML	AML	Overall Survival	AMLCG (2006)	Mizuno	HT-155A	S1812018_1_#	1130	0.77	0.18883	0.30400
S1812019	AML	AML	Overall Survival	AMLCG (2006)	Mizuno	HT-155A	S1812019_1_#	1130	0.53	0.24289	0.30400
S182000	AML	AML	Overall Survival	AMLCG (2006)	Mizuno	HT-155A	S182000_1_#	1130	0.47	0.11265	0.30400
S182001	AML	AML	Overall Survival	AMLCG (2006)	Mizuno	HT-155A	S182001_1_#	1130	0.01	0.00021	0.00010
S182002	AML	AML	Overall Survival	AMLCG (2006)	Mizuno	HT-155A	S182002_1_#	1130	0.22	0.01095	0.15600
S182003	AML	AML	Overall Survival	AMLCG (2006)	Mizuno	HT-155A	S182003_1_#	1130	0.44	0.04287	0.15600
S182004	AML	AML	Overall Survival	AMLCG (2006)	Mizuno	HT-155A	S182004_1_#	1130	0.88	0.01118	0.17900
S182005	AML	AML	Overall Survival	AMLCG (2006)	Mizuno	HT-155A	S182005_1_#	1130	0.72	0.00118	0.17900
S182006	AML	AML	Overall Survival	AMLCG (2006)	Mizuno	HT-155A	S182006_1_#	1130	0.04	0.00185	0.05000

図 1 5 PrognScan画面その 1。(A) トップページは簡素で、遺伝子ID を入力するだけで使用できる。(B) サマリテーブル。列はデータセット・癌種・サブタイプ・エンドポイント・コホート・データ取得者・プラットフォーム・プローブID・患者数・最適分割点・ P_{min} ・ P_{cor} から成る。5%水準で有意な P_{cor} は赤く強調される。各データセットには元データがアーカイブされてあるパブリック・ドメインへのリンクがついている。サマリテーブル内のプローブIDをクリックすることで各テストの詳細なレポートを見ることができる。下部にあるボタンを使ってテーブルをタブ区切りファイルでダウンロードすることができる。

MKI67を評価したところ、152のテストのうち52が5%水準の有意性を示した（膀胱癌3/5、血液癌6/28、乳癌39/83、脳腫瘍3/8、食道癌0/1、頸椎癌0/4、腎臓癌0/1、肺癌1/16、卵巣癌0/6）（別表2）。このうちBリンパ腫における全生存以外のすべてのテストがMKI67発現上昇と予後不良の関連を示していた。この結果はこれまでの報告と一致するものである（Brown and Gatter, 2002）。

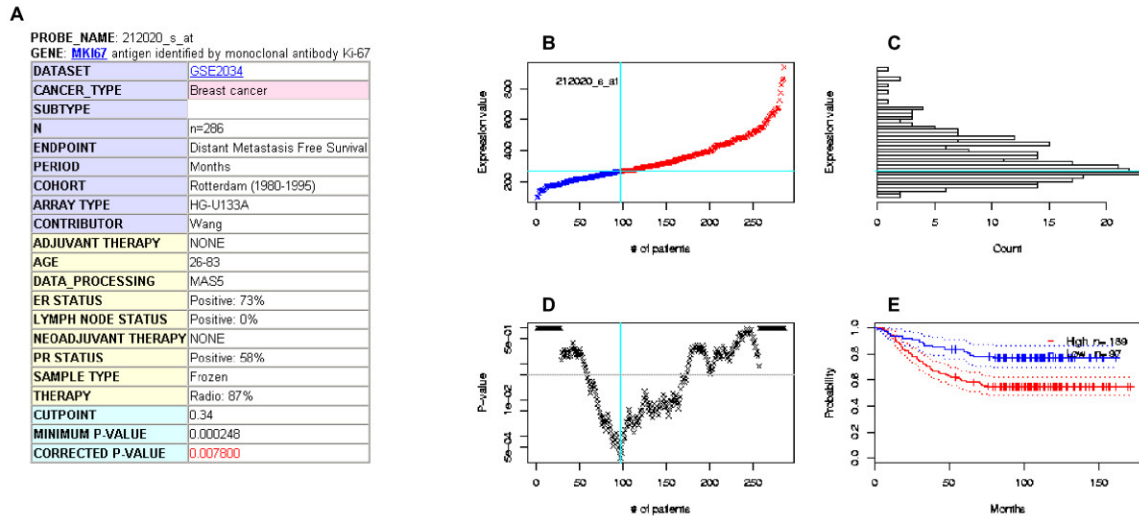


図 1 6 PrognScan画面その 2。(A) アノテーションテーブル。治療歴・サンプル種類・臨床パラメータなどの詳細を表示する。(B) 遺伝子発現プロット。遺伝子発現量順に患者が並びかえられている。X軸は累積の患者数、Y軸は遺伝子発現量を表す。水色の直線は患者を高発現群(赤)および低発現群(青)に二分するのに最適な分割点を示す。(C) 遺伝子発現ヒストグラム。X軸に患者数、Y軸に遺伝子発現プロットと同じスケールで分布を表示する。最適分割点が水色の直線で示される。(D) *P*-valueプロット。各分割点において患者が二分され、高発現群と低発現群の生存率の違いがLog-rankテストによって検定される。X軸は発現プロットと同様に患者の累積数を表し、Y軸は補正前の*P*-valueを対数スケールで表す。*P*値を最小にする点は水色の直線で示される。灰色の直線は5%有意水準を示す。(E) Kaplan-meier生存曲線。最適分割点で二分された高発現群(赤)および低発現群(青)の生存曲線が示される。X軸は時間、Y軸は生存率を表す。各群の95%信頼区間が点線で示される。

こうして得られる各項目の詳細を知りたい場合、リスト中のプローブIDをクリックすると、データセットアノテーションと四つの画像パネルから成るページへと移動する。図 1 6はRotterdam乳癌コホートでMKI67発現の無遠隔転移生存との関連を調べた例であるが、34%地点で P_{\min} が得られ、高発現患者群の予後が悪いことが見てとれる($P_{\text{cor}} = 0.0078$)。

MKI67と並んで代表的なTOP2A、PCNA、Aurora Aなどの増殖マーカーについても、様々なテストにおいて予後不良との関連が示された(表2)。これらはPrognScanの解析データの妥当性を保証する結果である。

	Bladder	Blood	Breast	Brain	Esophagus	Head and Neck	Kidney	Lung	Ovary
MKI67	3/5	6(1)/28	39/83	3/8	0/1	0/4	0/1	1/16	0/6
TOP2A	1/3	2/17	21/53	3/6	0/1	0/3	0/1	2/16	0/3
PCNA	0/2	1/7	6/26	1/2	0/2	0/1	0/1	2/9	0/4
Aurora A	0/4	2/21	26/65	2/6	0/1	0/2	ND	0/14	0/6

() represents inverse association

表2 腫瘍増殖マーカーのTOP2A、PCNA、Aurora A、MKI67をPrognoScanで評価し、予後不良との関連性を示すテスト数を数えた。

二つ目の例は、新規癌遺伝子として研究が活発になってきている **SIX1** (**SIX homeobox 1**) である。このホメオボックス遺伝子については、これまでに、細胞周期遺伝子 **Cyclin A1** を直接活性化して癌化プロセスを促進すること (**Coletta et al., 2004; Coletta et al., 2008**)、発現が肝癌と後期卵巣癌の予後と関連すること (**Ng et al., 2006; Behbakht et al., 2007**)、が報告されている。また、乳癌において遺伝子増幅や発現亢進が見られることも報告されている (**Ford et al., 1998; Reichenberger et al., 2005**)。しかし、この遺伝子の発現と患者予後との関連性はまだ示されていない。そこで、**SIX1** を **PrognoScan** で評価することにした。

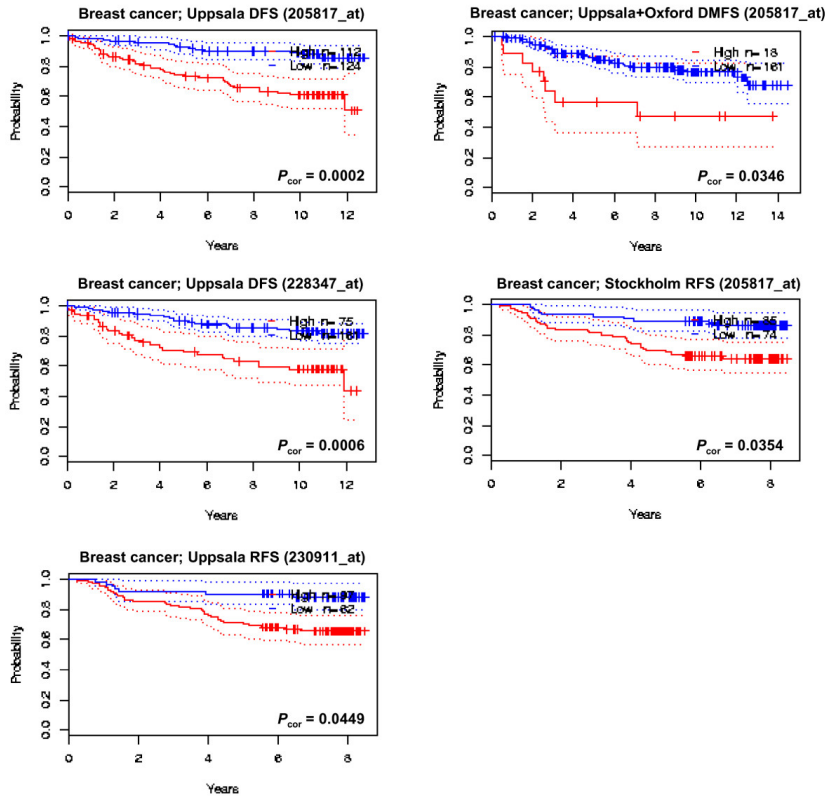


図 17 各コホートにおける SIX1 高発現群 (赤) と低発現群 (青) の生存曲線

卵巣癌に関しては、参照可能な三つのテストのいずれにも明確な関連は観察されなかった。SIX1 発現と卵巣癌予後の関連は後期ステージでのみ確認されているので (Behbakht *et al.*, 2007)、データセット中の癌サンプルをステージ毎にさらに細分して解析する必要があるのかもしれない。他方、乳癌については、参照可能な 28 のテストのうち 5 つで SIX1 発現と予後不良との有意な関連が示された (図 17, Uppsala コホート; $P_{cor} = 0.0002, 0.0006, 0.0449$, Uppsala+Oxford コホート; $P_{cor} = 0.0346$, Stockholm コホート; $P_{cor} = 0.0354$)。また、5%水準に達しなかったものの、SIX1 発現は GUYT2・MGH コホートについても予後不良に関連する傾向を示した ($P_{cor} = 0.0601, 0.0729$)。筆者の知る限り、SIX1 発現と乳癌予後との関連がデータで示されるのはこれが初めてである。これまで論文で報告されている知見と合わせると、SIX1 発現が乳癌の悪性化に関与していることが強く示唆される。PrognScan を利用することで、こうした高次の考察を行うことが容易になる。

三つ目の例は、MCTS1 (Malignant T cell amplified sequence 1) である。MCTS1 は、NIH3T3 マウス線維芽細胞では形質転換を引き起こし (Prosniak *et al.*, 1998)、ゼノグラフトモデルでは血管新生の促進と細胞死の抑制によって癌化を誘導する (Levenson *et al.*, 2005)。調べた限り、この遺伝子の発現と癌患者予後との関連はいずれの癌においてもまだ報告されていなかったため、PrognScan を用いて評価することにした。

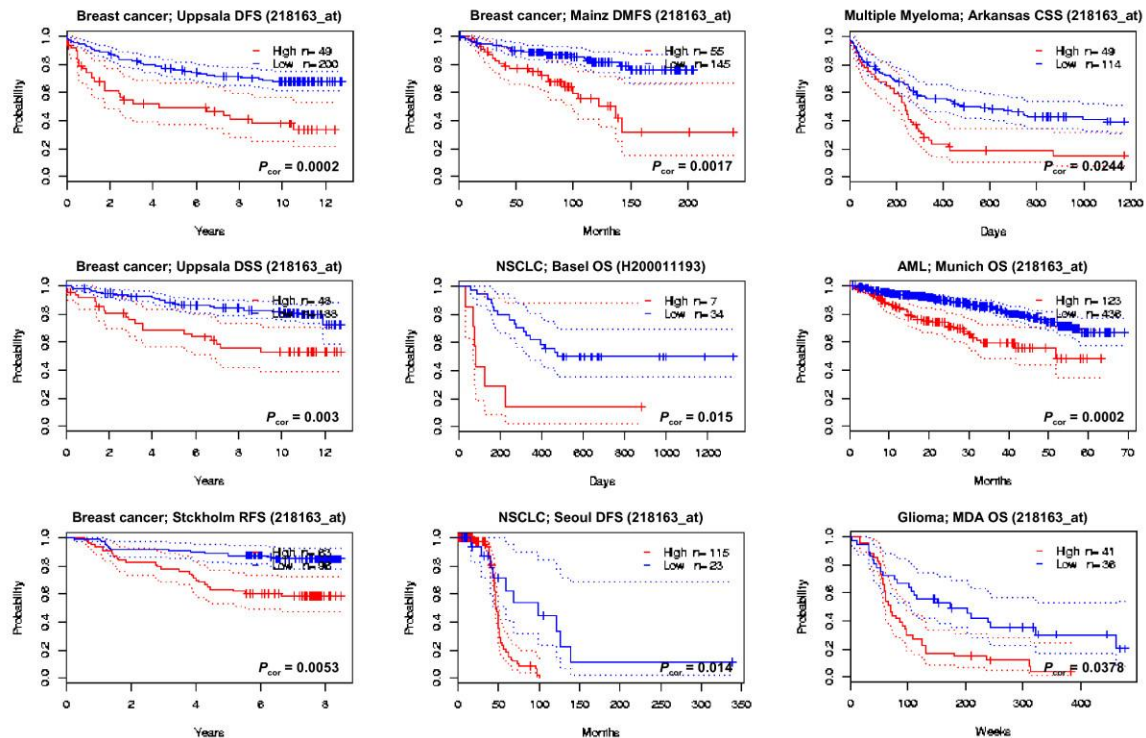


図 18 各データセットにおける MCTS1 高発現群 (赤) と低発現群 (青) の生存曲線

結果、いくつかのテストにおいて MCTS1 発現と患者予後との間に統計的に有意な関係が見つかった (図 18, 血液癌 2/7、乳癌 4/21、脳腫瘍 1/2、肺癌 2/5)。これらのテストすべてで MCTS1 の高発現は不良な予後と関連していた。これまでの知見と併せて、MCTS1 遺伝子の癌化プロセスに対する積極的な関与が示唆された。SIX1 のときと同様、MCTS1 の発現と予後との関連がデータで示されるのはこれが最初である。

2. 3 考察

PrognScanは個々の遺伝子発現と患者予後との関連に焦点を合わせて設計されたデータベースで、必ずしも予後予測への貢献を想定したものではない。予後予測の精度を高める観点では、1章で紹介したように、複数の遺伝子発現を同時に考慮する手法のほうが発展の余地が大きいだろう。このデータベースの意義は、パブリックドメインに蓄積された癌マイクロアレイデータを使って遺伝子発現と予後との関連を示し、各遺伝子の研究対象としての潜在的価値を素早く評価する点にある。本報告で、癌遺伝子候補SIX1の発現と乳癌、MCTS1の発現と脳腫瘍・血液癌・乳癌・肺癌の予後が関連することを示し、さらなる研究の足がかりを提供したが、こうした利用を主に想定している。ちなみに、このアイデアは新しいものではない。実際、Mehra *et al.* (2005)、Paulson *et al.* (2007)、Kim *et al.* (2007) は、各々が着目していたGATA3、HBP1、CUL7といった遺伝子の発現が予後と関連することを公共マイクロアレイデータで示し、そこから研究を展開している。PrognScanの開発は、以前からあったこのような潜在的ニーズに応えたものという位置づけになる。

さて、公共マイクロアレイデータを使用して生存解析を行う場合、次のようないくつかの制約を考慮する必要がある。

イ) コホート： データセットは世界中の様々な研究機関で取得されたものであり、患者の背景はそれぞれ異なっている。さらに、コレクションには、最初から特定の集団に焦点をあてて解析されたデータセットも含まれる。例えば、Rotterdam乳癌コホートはリンパ節転移のない癌を対象を絞ってあるし、Aarhus膀胱癌コホートはシスプラチン系の化学療法を受けた患者のデータである。コホートに異なったバイアスが掛かっている以上、同じ癌種であっても、特定のデータセットにのみ遺伝子発現と予後との関連が見つかる可能性がある。例えばDai *et al.* (2005) は、乳癌において、年齢の割にエストロゲン受容体発現の高い患者群では細胞周期遺伝子の発現が予後と関連するのに対し、それ以外の患者群では関連が弱くなる・無くなることを報告している。

ロ) 治療の質： 医療機関自体が臨床結果に影響を与える要素となることがしばしば報告されている (Schrag *et al.*, 2000; Bilimoria *et al.*, 2007; Vernooij *et al.*, 2008)。従って、マイクロアレイデータ取得時にコホートが同等であったとしても、その後の治療の質が患者の臨床経過に影響することはあり得る。

ハ) 実験的要因： マイクロアレイによる発現測定は、実験レベルの様々な要素の影響を受ける。例えば、生検サンプルに対するMicrodissection処理 (例：乳癌MGHコホート) は非癌細胞からのmRNA夾雑を抑え、測定の質を高める (Mizuarai *et al.*, 2005)。一方、サンプルのホルマリン固定 (例：頸椎癌VUMCコホート) はmRNAの質を劣化させる (Mazumder and Wang, 2006)。プラットフォーム (Affymetrix、cDNAマイクロアレイ) やデータ処理法 (MAS、RMA) の違いも測定値に影響を及ぼすことが知られている (Irizarry *et al.*, 2003)。さらに、マイクロアレイで使用されているプローブの中には、不正なデザインのもものが混ざっていることも知られている (Draghici *et al.*, 2006)。

ニ) 偽陽性： 遺伝子発現と予後との関連が全くない場合でも、偶然、偽の関連性が作られることがある。

元になっているデータセット間で、こうした影響がそれぞれ違うので、結果の解釈には一定の注意が必要である。現在、Prognoscanは、コホート、治療歴、病理所見、プラットフォームなどの情報を整理した形で提供し、結果の適切な解釈をサポートしている。

今後のPrognoscanの拡張であるが、まずデータセット収集を続け、六カ月毎にデータをアップデートする予定である。データの増加はより頑強な予測をするのに役立つだろう。その他に、患者を複数の分割点で分けるためのアルゴリズム開発にも取り組んでいる。現在は計算リソースの制約があるため、発現値を上下二分分割する点のみの探索を行っているが、治療目的には高・中・低のような三分分割も有用である。具体的には、「X-tile」 (Camp *et al.*, 2004) で示されたようなグリッド検索アルゴリズムの実装を検討中である。さらに長期的には、複数の解析結果を統合して解釈するアルゴリズムの開発も視野に入れている。

今後の機能追加・拡張を含め、この新しいデータベースは、潜在的腫瘍マーカーや創薬標的を評価するための強力なプラットフォームを提供し、癌研究を加速させていくと期待される。

謝辞

本研究をまとめるにあたり、終始ご指導ご鞭撻をいただいた九州工業大学情報工学研究科・皿井明倫教授に感謝する。

本稿を上梓するにあたって、実験面での協力・原稿へのアドバイスをいただいた、東京大学ヒトゲノムセンターの中井謙太教授、中外製薬（株）の **Detlef Schmitt**、**Frances Ford**、青木雅弘、芦原基起、大森寛、北田邦雄、高橋恒太、中西義人、原村昌幸の各氏に感謝する。

本研究の業績

本研究の一章・二章の研究成果は、それぞれ以下の学術誌に掲載された。

Mizuno H, Nakanishi Y, Ishii N, Sarai A, Kitada K: A signature-based method for indexing cell cycle phase distribution from microarray profiles. *BMC Genomics*. 2009, 10:137. (<http://www.biomedcentral.com/1471-2164/10/137>)

Mizuno H, Kitada K, Nakai K, Sarai A: PrognoScan: A new database for meta-analysis of the prognostic value of genes. *BMC Med Genomics*. 2009, 2:18. (<http://www.biomedcentral.com/1755-8794/2/18>)

参考文献

- Abel U, Berger J, Wiebelt H: CRITLEVEL: an exploratory procedure for the evaluation of quantitative prognostic factors. *Methods Inf Med* 1984, 23:154-6.
- Als AB, Dyrskjøt L, Maase H, Koed K, Mansilla F, Toldbod HE, Jensen JL, Ulhøi BP, Sengeløv L, Jensen KM, Orntoft TF: Emmprin and survivin predict response and survival following cisplatin-containing chemotherapy in patients with advanced bladder cancer. *Clin Cancer Res* 2007, 13:4407-14.
- Altman DG, Lausen B, Sauerbrei W, Schumacher M: Dangers of using "optimal" cutpoints in the evaluation of prognostic factors. *J Natl Cancer Inst* 1994, 86:829-35.
- Anders CK, Acharya CR, Hsu DS, Broadwater G, Garman K, Foekens JA, Zhang Y, Wang Y, Marcom K, Marks JR, Mukherjee S, Nevins JR, Blackwell KL, Potti A: Age-specific differences in oncogenic pathway deregulation seen in human breast tumors. *PLoS ONE* 2008, 3:e1373.
- Baker FL, Sanger LJ, Rodgers RW, Jabboury K, Mangini OR: Cell proliferation kinetics of normal and tumour tissue in vitro: quiescent reproductive cells and the cycling reproductive fraction. *Cell Prolif* 1995, 28:1-15.
- Bar-Joseph Z, Siegfried Z, Brandeis M, Brors B, Lu Y, Eils R, Dynlacht BD, Simon I: Genome-wide transcriptional analysis of the human cell cycle identifies genes differentially regulated in normal and cancer cells. *Proc Natl Acad Sci USA* 2008, 105:955-60.
- Barnes DM, Gillett CE: Cyclin D1 in breast cancer. *Breast Cancer Res Treat* 1998, 52:1-15.
- Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, Evangelista C, Kim IF, Soboleva A, Tomashevsky M, Edgar R: NCBI GEO: mining tens of millions of expression profiles – database and tools update. *Nucleic Acids Res* 2007, 35:D760-5.
- Beer DG, Kardia SL, Huang CC, Giordano TJ, Levin AM, Misek DE, Lin L, Chen G, Gharib TG, Thomas DG, Lizyness ML, Kuick R, Hayasaka S, Taylor JM,

- Iannettoni MD, Orringer MB, Hanash S: Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat Med* 2002, 8:816-24.
- Behbakht K, Qamar L, Aldridge CS, Coletta RD, Davidson SA, Thorburn A, Ford HL: Six1 overexpression in ovarian carcinoma causes resistance to TRAIL-mediated apoptosis and is associated with poor survival. *Cancer Res* 2007, 67:3036-42.
- Beresford MJ, Wilson GD, Makris A: Measuring proliferation in breast cancer: practicalities and applications. *Breast Cancer Res* 2006, 8:216.
- Bild AH, Yao G, Chang JT, Wang Q, Potti A, Chasse D, Joshi MB, Harpole D, Lancaster JM, Berchuck A, Olson JA Jr, Marks JR, Dressman HK, West M, Nevins JR: Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* 2006, 439:353-7.
- Bilimoria KY, Bentrem DJ, Ko CY, Tomlinson JS, Stewart AK, Winchester DP, Talamonti MS: Multimodality therapy for pancreatic cancer in the U.S. *Cancer* 2007, 110:1227-34.
- Bolstad BM, Irizarry RA, Astrand M, Speed TP: A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 2003, 19:185-93.
- Brown DC, Gatter KC: Ki67 protein: the immaculate deception? *Histopathology* 2002, 40:2-11.
- Cam H, Balciunaite E, Blais A, Spektor A, Scarpulla RC, Young R, Kluger Y, Dynlacht BD: A common set of gene regulatory networks links metabolism and growth inhibition. *Mol Cell* 2004, 16:399-411.
- Camp RL, Dolled-Filhart M, Rimm DL: X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. *Clin Cancer Res* 2004, 10:7252-9.
- Chanrion M, Negre V, Fontaine H, Salvetat N, Bibeau F, Mac Grogan G, Mauriac L, Katsaros D, Molina F, Theillet C, Darbon JM: A gene expression signature that can predict the recurrence of tamoxifen-treated primary breast cancer. *Clin Cancer Res* 2008, 14:1744-52.

- Chin K, DeVries S, Fridlyand J, Spellman PT, Roydasgupta R, Kuo WL, Lapuk A, Neve RM, Qian Z, Ryder T, Chen F, Feiler H, Tokuyasu T, Kingsley C, Dairkee S, Meng Z, Chew K, Pinkel D, Jain A, Ljung BM, Esserman L, Albertson DG, Waldman FM, Gray JW: Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. *Cancer Cell* 2006, 10:529-41.
- Chung CH, Parker JS, Ely K, Carter J, Yi Y, Murphy BA, Ang KK, El-Naggar AK, Zanation AM, Cmelak AJ, Levy S, Slebos RJ, Yarbrough WG: Gene expression profiles identify epithelial-to-mesenchymal transition and activation of nuclear factor-kappaB signaling as characteristics of a high-risk head and neck squamous cell carcinoma. *Cancer Res* 2006, 66:8210-8.
- Coletta RD, Christensen K, Reichenberger KJ, Lamb J, Micomonaco D, Huang L, Wolf DM, Muller-Tidow C, Golub TR, Kawakami K, Ford HL: The Six1 homeoprotein stimulates tumorigenesis by reactivation of cyclin A1. *Proc Natl Acad Sci USA* 2004, 101:6478-83.
- Coletta RD, Christensen KL, Micalizzi DS, Jedlicka P, Varella-Garcia M, Ford HL: Six1 overexpression in mammary cells induces genomic instability and is sufficient for malignant transformation. *Cancer Res* 2008, 68:2204-13.
- Colozza M, Azambuja E, Cardoso F, Sotiriou C, Larsimont D, Piccart MJ: Proliferative markers as prognostic and predictive tools in early breast cancer: where are we now? *Ann Oncol* 2005, 16:1723-39.
- Dai H, van't Veer L, Lamb J, He YD, Mao M, Fine BM, Bernardis R, Vijver M, Deutsch P, Sachs A, Stoughton R, Friend S: A cell proliferation signature is a marker of extremely poor outcome in a subpopulation of breast cancer patients. *Cancer Res* 2005, 65:4059-66.
- Desmedt C, Piette F, Loi S, Wang Y, Lallemand F, Haibe-Kains B, Viale G, Delorenzi M, Zhang Y, d'Assignies MS, Bergh J, Lidereau R, Ellis P, Harris AL, Klijn JG, Foekens JA, Cardoso F, Piccart MJ, Buyse M, Sotiriou C, TRANSBIG Consortium: Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. *Clin Cancer Res* 2007, 13:3207-14.

- Elfilali A, Lair S, Verbeke C, La Rosa P, Radvanyi F, Barillot E: ITTACA: a new database for integrated tumor transcriptome array and clinical data analysis. *Nucleic Acids Res* 2006, 34:D613-6.
- Ezoe S, Matsumura I, Satoh Y, Tanaka H, Kanakura Y: Cell cycle regulation in hematopoietic stem/progenitor cells. *Cell Cycle* 2004, 3:314-8. Zhou T, Chou JW, Simpson DA, Zhou Y, Mullen TE, Medeiros M, Bushel PR, Paules RS, Yang X, Hurban P, Lobenhofer EK, Kaufmann WK: Profiles of global gene expression in ionizing-radiation-damaged human diploid fibroblasts reveal synchronization behind the G1 checkpoint in a G0-like state of quiescence. *Environ Health Perspect* 2006, 114:553-9.
- Felsher DW, Bishop JM: Transient excess of MYC activity can elicit genomic instability and tumorigenesis. *Proc Natl Acad Sci USA* 1999, 96:3940-4.
- Ford HL, Kabingu EN, Bump EA, Mutter GL, Pardee AB: Abrogation of the G2 cell cycle checkpoint associated with overexpression of HSIX1: a possible mechanism of breast carcinogenesis. *Proc Natl Acad Sci USA* 1998, 95:12608-13.
- Fournier MV, Martin KJ, Kenny PA, Xhaja K, Bosch I, Yaswen P, Bissell MJ: Gene expression signature in organized and growth-arrested mammary acini predicts good outcome in breast cancer. *Cancer Res* 2006, 66:7095-102.
- Gonzalez MA, Tachibana KE, Chin SF, Callagy G, Madine MA, Vowler SL, Pinder SE, Laskey RA, Coleman N: Geminin predicts adverse clinical outcome in breast cancer by reflecting cell-cycle progression. *J Pathol* 2004, 204:121-30.
- Harper JV, Brooks G: The mammalian cell cycle: an overview. *Methods Mol Biol* 2005, 296:113-53.
- Herschkowitz JI, Simin K, Weigman VJ, Mikaelian I, Usary J, Hu Z, Rasmussen KE, Jones LP, Assefnia S, Chandrasekharan S, Backlund MG, Yin Y, Khramtsov AI, Bastein R, Quackenbush J, Glazer RI, Brown PH, Green JE, Kopelovich L, Furth PA, Palazzo JP, Olopade OI, Bernard PS, Churchill GA, Van Dyke T, Perou CM: Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biol* 2007, 8:R76.

- Holländer N, Schumacher M: On the problem of using 'optimal' cutpoints in the assessment of quantitative prognostic factors. *Onkologie* 2001, 24:194-9.
- Hummel M, Bentink S, Berger H, Klapper W, Wessendorf S, Barth TF, Bernd HW, Cogliatti SB, Dierlamm J, Feller AC, Hansmann ML, Haralambieva E, Harder L, Hasenclever D, Kühn M, Lenze D, Lichter P, Martin-Subero JI, Möller P, Müller-Hermelink HK, Ott G, Parwaresch RM, Pott C, Rosenwald A, Rosolowski M, Schwaenen C, Stürzenhofecker B, Szczepanowski M, Trautmann H, Wacker HH, Spang R, Loeffler M, Trümper L, Stein H, Siebert R, Molecular Mechanisms in Malignant Lymphomas Network Project of the Deutsche Krebshilfe: A biologic definition of Burkitt's lymphoma from transcriptional and genomic profiling. *N Engl J Med* 2006, 354:2419-30.
- Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP: Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res* 2003, 31:e15.
- Draghici S, Khatri P, Eklund AC, Szallasi Z: Reliability and reproducibility issues in DNA microarray measurements. *Trends Genet* 2006, 22:101-9.
- Ivshina AV, George J, Senko O, Mow B, Putti TC, Smeds J, Lindahl T, Pawitan Y, Hall P, Nordgren H, Wong JE, Liu ET, Bergh J, Kuznetsov VA, Miller LD: Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer. *Cancer Res* 2006, 66:10292-301.
- Jais JP, Haioun C, Molina TJ, Rickman DS, de Reynies A, Berger F, Gisselbrecht C, Brière J, Reyes F, Gaulard P, Feugier P, Labouyrie E, Tilly H, Bastard C, Coiffier B, Salles G, Leroy K, Groupe d'Etude des Lymphomes de l'Adulte: The expression of 16 genes related to the cell of origin and immune response predicts survival in elderly patients with diffuse large B-cell lymphoma treated with CHOP and rituximab. *Leukemia* 2008, 22:1917-24.
- Jeganathan K, Malureanu L, Baker DJ, Abraham SC, van Deursen JM: Bub1 mediates cell death in response to chromosome missegregation and acts to suppress spontaneous tumorigenesis. *J Cell Biol* 2007, 179:255-67.
- Jensen KC, Turbin DA, Leung S, Miller MA, Johnson K, Norris B, Hastie T, McKinney S, Nielsen TO, Huntsman DG, Gilks CB, West RB: New cutpoints to identify increased HER2 copy number: analysis of a large, population-

- based cohort with long-term follow-up. *Breast Cancer Res Treat* 2008, 112:453-9.
- Karn J, Watson JV, Lowe AD, Green SM, Vedeckis W: Regulation of cell cycle duration by c-myc levels. *Oncogene* 1989, 4:773-87.
- Kim SS, Shago M, Kaustov L, Boutros PC, Clendening JW, Sheng Y, Trentin GA, Barsyte-Lovejoy D, Mao DY, Kay R, Jurisica I, Arrowsmith CH, Penn LZ: CUL7 is a novel antiapoptotic oncogene. *Cancer Res* 2007, 67:9616-22.
- Kronqvist P, Kuopio T, Collan Y: Quantitative thresholds for mitotic counts in histologic grading: confirmation in nonfrozen samples of invasive ductal breast cancer. *Ann Diagn Pathol* 2000, 4:65-70.
- Landberg G, Roos G: The cell cycle in breast cancer. *APMIS* 1997, 105:575-89.
- Langerød A, Zhao H, Borgan O, Nesland JM, Bukholm IR, Ikdahl T, Karesen R, Borresen-Dale AL, Jeffrey SS: TP53 mutation status and gene expression profiles are powerful prognostic markers of breast cancer. *Breast Cancer Res* 2007, 9:R30.
- Lemoine FJ, Marriott SJ: Accelerated G(1) phase progression induced by the human T cell leukemia virus type I (HTLV-I) Tax oncoprotein. *J Biol Chem* 2001, 276:31851-7.
- Levenson AS, Thurn KE, Simons LA, Veliceasa D, Jarrett J, Osipo C, Jordan VC, Volpert OV, Satcher RL Jr, Gartenhaus RB: MCT-1 oncogene contributes to increased in vivo tumorigenicity of MCF7 cells by promotion of angiogenesis and inhibition of apoptosis. *Cancer Res* 2005, 65:10651-6.
- Liu JJ, Chao JR, Jiang MC, Ng SY, Yen JJ, Yang-Yen HF: Ras transformation results in an elevated level of cyclin D1 and acceleration of G1 progression in NIH 3T3 cells. *Mol Cell Biol* 1995, 15:3654-63.
- Loi S, Haibe-Kains B, Desmedt C, Wirapati P, Lallemant F, Tutt AM, Gillet C, Ellis P, Ryder K, Reid JF, Daidone MG, Pierotti MA, Berns EM, Jansen MP, Foekens JA, Delorenzi M, Bontempi G, Piccart MJ, Sotiriou C: Predicting prognosis using molecular profiling in estrogen receptor-positive breast cancer treated with tamoxifen. *BMC Genomics* 2008, 9:239.

- Lu P, Nakorchevskiy A, Marcotte EM: Expression deconvolution: a reinterpretation of DNA microarray data reveals dynamic changes in cell populations. *Proc Natl Acad Sci USA* 2003, 100:10370-5.
- Ma XJ, Wang Z, Ryan PD, Isakoff SJ, Barmettler A, Fuller A, Muir B, Mohapatra G, Salunga R, Tuggle JT, Tran Y, Tran D, Tassin A, Amon P, Wang W, Wang W, Enright E, Stecker K, Estepa-Sabal E, Smith B, Younger J, Balis U, Michaelson J, Bhan A, Habin K, Baer TM, Brugge J, Haber DA, Erlander MG, Sgroi DC: A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen. *Cancer Cell* 2004, 5:607-16.
- Mazumdar M, Glassman JR: Categorizing a prognostic variable: review of methods, code for easy implementation and applications to decision-making about cancer treatments. *Stat Med* 2000, 19:113-32.
- Mazumder A, Wang Y: Gene-expression signatures in oncology diagnostics. *Pharmacogenomics* 2006, 7:1167-73.
- Mehra R, Varambally S, Ding L, Shen R, Sabel MS, Ghosh D, Chinnaiyan AM, Kleer CG: Identification of GATA3 as a breast cancer prognostic marker by global gene expression meta-analysis. *Cancer Res* 2005, 65:11259-64.
- Metzeler KH, Hummel M, Bloomfield CD, Spiekermann K, Braess J, Sauerland MC, Heinecke A, Radmacher M, Marcucci G, Whitman SP, Maharry K, Paschka P, Larson RA, Berdel WE, Büchner T, Wörmann B, Mansmann U, Hiddemann W, Bohlander SK, Buske C, Cancer and Leukemia Group B; German AML Cooperative Group: An 86-probe-set gene-expression signature predicts survival in cytogenetically normal acute myeloid leukemia. *Blood* 2008, 112:4193-201.
- Miller LD, Smeds J, George J, Vega VB, Vergara L, Ploner A, Pawitan Y, Hall P, Klaar S, Liu ET, Bergh J: An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. *Proc Natl Acad Sci USA* 2005, 102:13550-5.
- Miller R, Siegmund D: Maximally selected chi square statistics. *Biometrics* 1982, 38:1011-16.

- Mizuarai S, Takahashi K, Kobayashi T, Kotani H: Advances in isolation and characterization of homogeneous cell populations using laser microdissection. *Histol Histopathol* 2005, 20:139-46.
- Murat A, Migliavacca E, Gorlia T, Lambiv WL, Shay T, Hamou MF, de Tribolet N, Regli L, Wick W, Kouwenhoven MC, Hainfellner JA, Heppner FL, Dietrich PY, Zimmer Y, Cairncross JG, Janzer RC, Domany E, Delorenzi M, Stupp R, Hegi ME: Stem cell-related "self-renewal" signature and high epidermal growth factor receptor expression associated with resistance to concomitant chemoradiotherapy in glioblastoma. *J Clin Oncol* 2008, 26:3015-24.
- Ng KT, Man K, Sun CK, Lee TK, Poon RT, Lo CM, Fan ST: Clinicopathological significance of homeoprotein Six1 in hepatocellular carcinoma. *Br J Cancer* 2006, 95:1050-5.
- Nygren JM, Bryder D, Jacobsen SE: Prolonged cell cycle transit is a defining and developmentally conserved hemopoietic stem cell property. *J Immunol* 2006, 177:201-8.
- Parkinson H, Kapushesky M, Shojatalab M, Abeygunawardena N, Coulson R, Farne A, Holloway E, Kolesnykov N, Lilja P, Lukk M, Mani R, Rayner T, Sharma A, William E, Sarkans U, Brazma A: ArrayExpress – a public database of microarray experiments and gene expression profiles. *Nucleic Acids Res* 2007, 35:D747-50.
- Paulson KE, Rieger-Christ K, McDevitt MA, Kuperwasser C, Kim J, Unanue VE, Zhang X, Hu M, Ruthazer R, Berasi SP, Huang CY, Giri D, Kaufman S, Dugan JM, Blum J, Netto G, Wazer DE, Summerhayes IC, Yee AS: Alterations of the HBP1 transcriptional repressor are associated with invasive breast cancer. *Cancer Res* 2007, 67:6136-45.
- Pawitan Y, Bjöhle J, Amler L, Borg AL, Egyhazi S, Hall P, Han X, Holmberg L, Huang F, Klaar S, Liu ET, Miller L, Nordgren H, Ploner A, Sandelin K, Shaw PM, Smeds J, Skoog L, Wedrén S, Bergh J: Gene expression profiling spares early breast cancer patients from adjuvant therapy: derived and validated in two population-based cohorts. *Breast Cancer Res* 2005, 7:R953-64.

- Petersen OW, Ronnov-Jessen L, Howlett AR, Bissell MJ: Interaction with basement membrane serves to rapidly distinguish growth and differentiation pattern of normal and malignant human breast epithelial cells. *Proc Natl Acad Sci USA* 1992, 89:9064-8.
- Phillips HS, Kharbanda S, Chen R, Forrest WF, Soriano RH, Wu TD, Misra A, Nigro JM, Colman H, Soroceanu L, Williams PM, Modrusan Z, Feuerstein BG, Aldape K: Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. *Cancer Cell* 2006, 9:157-73.
- Potti A, Dressman HK, Bild A, Riedel RF, Chan G, Sayer R, Cragun J, Cottrill H, Kelley MJ, Petersen R, Harpole D, Marks J, Berchuck A, Ginsburg GS, Febbo P, Lancaster J, Nevins JR: Genomic signatures to guide the use of chemotherapeutics. *Nat Med* 2006, 12:1294-300.
- Prather RS, Boquest AC, Day BN: Cell cycle analysis of cultured porcine mammary cells. *Cloning* 1999, 1:17-24.
- Prosniak M, Dierov J, Okami K, Tilton B, Jameson B, Sawaya BE, Gartenhaus RB: A novel candidate oncogene, MCT-1, is involved in cell cycle progression. *Cancer Res* 1998, 58:4233-7.
- Raponi M, Lancet JE, Fan H, Dossey L, Lee G, Gojo I, Feldman EJ, Gotlib J, Morris LE, Greenberg PL, Wright JJ, Harousseau JL, Löwenberg B, Stone RM, De Porre P, Wang Y, Karp JE: A 2-gene classifier for predicting response to the farnesyltransferase inhibitor tipifarnib in acute myeloid leukemia. *Blood* 2008, 111:2589-96.
- Raponi M, Zhang Y, Yu J, Chen G, Lee G, Taylor JM, Macdonald J, Thomas D, Moskaluk C, Wang Y, Beer DG: Gene expression signatures for predicting prognosis of squamous cell and adenocarcinomas of the lung. *Cancer Res* 2006, 66:7466-72.
- Reichenberger KJ, Coletta RD, Schulte AP, Varella-Garcia M, Ford HL: Gene amplification is a mechanism of Six1 overexpression in breast cancer. *Cancer Res* 2005, 65:2668-75.

- Saldanha AJ: Java Treeview – extensible visualization of microarray data. *Bioinformatics* 2004, 20:3246-8.
- Schmidt M, Böhm D, von Törne C, Steiner E, Puhl A, Pilch H, Lehr HA, Hengstler JG, Kölbl H, Gehrman M: The humoral immune system has a key prognostic impact in node-negative breast cancer. *Cancer Res* 2008, 68:5405-13.
- Schrag D, Cramer LD, Bach PB, Cohen AM, Warren JL, Begg CB: Influence of hospital procedure volume on outcomes following surgery for colon cancer. *JAMA* 2000, 284:3028-35.
- Shachaf CM, Gentles AJ, Elchuri S, Sahoo D, Soen Y, Sharpe O, Perez OD, Chang M, Mitchel D, Robinson WH, Dill D, Nolan GP, Plevritis SK, Felsher DW: Genomic and proteomic analysis reveals a threshold level of MYC required for tumor maintenance. *Cancer Res* 2008, 68:5132-42.
- Shetty A, Loddo M, Fanshawe T, Prevost AT, Sainsbury R, Williams GH, Stoeber K: DNA replication licensing and cell cycle kinetics of normal and neoplastic breast. *Br J Cancer* 2005, 93:1295-300.
- Sladek TL, Jacobberger JW: Simian virus 40 large T-antigen expression decreases the G1 and increases the G2 + M cell cycle phase durations in exponentially growing cells. *J Virol* 1992, 66:1059-65.
- Sotiriou C, Wirapati P, Loi S, Harris A, Fox S, Smeds J, Nordgren H, Farmer P, Praz V, Haibe-Kains B, Desmedt C, Larsimont D, Cardoso F, Peterse H, Nuyten D, Buyse M, Vijver MJ, Bergh J, Piccart M, Delorenzi M: Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J Natl Cancer Inst* 2006, 98:262-72.
- Sun H, Lesche R, Li DM, Liliental J, Zhang H, Gao J, Gavrilova N, Mueller B, Liu X, Wu H: PTEN modulates cell cycle progression and cell survival by regulating phosphatidylinositol 3,4,5,-trisphosphate and Akt/protein kinase B signaling pathway. *Proc Natl Acad Sci USA* 1999, 96:6199-204.
- Tomida S, Koshikawa K, Yatabe Y, Harano T, Ogura N, Mitsudomi T, Some M, Yanagisawa K, Takahashi T, Osada H, Takahashi T: Gene expression-based, individualized outcome prediction for surgically treated lung cancer patients. *Oncogene* 2004, 23:5360-70.

- van 't Veer LJ, Dai H, Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH: Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002, 415:530-6.
- Vecchione A, Croce CM, Baldassarre G: Fez1/Lzts1 a new mitotic regulator implicated in cancer development. *Cell Div* 2007, 2:24.
- Vernooij F, Heintz AP, Witteveen PO, Heiden-van der Loo M, Coebergh JW, Graaf Y: Specialized care and survival of ovarian cancer patients in The Netherlands: nationwide cohort study. *J Natl Cancer Inst* 2008, 100:399-406.
- Wang Y, Klijn JG, Zhang Y, Sieuwerts AM, Look MP, Yang F, Talantov D, Timmermans M, Meijer-van Gelder ME, Yu J, Jatkoe T, Berns EM, Atkins D, Foekens JA: Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* 2005, 365:671-9.
- Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, Dicuccio M, Edgar R, Federhen S, Feolo M, Geer LY, Helmberg W, Kapustin Y, Khovayko O, Landsman D, Lipman DJ, Madden TL, Maglott DR, Miller V, Ostell J, Pruitt KD, Schuler GD, Shumway M, Sequeira E, Sherry ST, Sirotkin K, Souvorov A, Starchenko G, Tatusov RL, Tatusova TA, Wagner L, Yaschenko E: Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 2008, 36:D13-21.
- Whitfield ML, George LK, Grant GD, Perou CM: Common markers of proliferation. *Nat Rev Cancer* 2006, 6:99-106.
- Whitfield ML, Sherlock G, Saldanha AJ, Murray JI, Ball CA, Alexander KE, Matese JC, Perou CM, Hurt MM, Brown PO, Botstein D: Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol Biol Cell* 2002, 13:1977-2000.
- Will E, Speidel D, Wang Z, Ghiaur G, Rimek A, Schiedlmeier B, Williams DA, Baum C, Ostertag W, Klump H: HOXB4 inhibits cell growth in a dose-dependent manner and sensitizes cells towards extrinsic cues. *Cell Cycle* 2006, 5:14-22.

- Williams GH, Stoeber K: Cell cycle markers in clinical oncology. *Curr Opin Cell Biol* 2007, 19:672-9.
- Wimmel A, Lucibello FC, Sewing A, Adolph S, Muller R: Inducible acceleration of G1 progression through tetracycline-regulated expression of human cyclin E. *Oncogene* 1994, 9:995-7.
- Yabuta N, Okada N, Ito A, Hosomi T, Nishihara S, Sasayama Y, Fujimori A, Okuzaki D, Zhao H, Ikawa M, Okabe M, Nojima H: Lats2 is an essential mitotic regulator required for the coordination of cell division. *J Biol Chem* 2007, 282:19259-71.
- Yamamoto T, Ebisuya M, Ashida F, Okamoto K, Yonehara S, Nishida E: Continuous ERK activation downregulates antiproliferative genes throughout G1 phase to allow cell-cycle progression. *Curr Biol* 2006, 16:1171-82.
- Zhan F, Huang Y, Colla S, Stewart JP, Hanamura I, Gupta S, Epstein J, Yaccoby S, Sawyer J, Burington B, Anaissie E, Hollmig K, Pineda-Roman M, Tricot G, van Rhee F, Walker R, Zangari M, Crowley J, Barlogie B, Shaughnessy JD Jr: The molecular classification of multiple myeloma. *Blood* 2006, 108:2020-8.
- Zhou Y, Yau C, Gray JW, Chew K, Dairkee SH, Moore DH, Eppenberger U, Eppenberger-Castori S, Benz CC: Enhanced NF kappa B and AP-1 transcriptional activity associated with antiestrogen resistant breast cancer. *BMC Cancer* 2007, 7:59.

補足資料

別表1 CCS 構成遺伝子一覧

CCS SUBSET ID	ASSIGNED PHASE	GENE SYMBOL	DESCRIPTION	Whitfield <i>et al.</i> assignment
1	M	ANP32E	acidic (leucine-rich) nuclear phosphoprotein 32 family, member E	
1	M	ARL6IP1	ADP-ribosylation factor-like 6 interacting protein 1	
1	M	BBS7	Bardet-Biedl syndrome 7	
1	M	BIRC5	baculoviral IAP repeat-containing 5 (survivin)	G2/M
1	M	BUB1	BUB1 budding uninhibited by benzimidazoles 1 homolog (yeast)	G2/M
1	M	C12orf24	chromosome 12 open reading frame 24	
1	M	CBX3	chromobox homolog 3 (HP1 gamma homolog, Drosophila)	
1	M	CCNB2	cyclin B2	G2/M
1	M	CDC27	cell division cycle 27 homolog (<i>S. cerevisiae</i>)	
1	M	CENPF	centromere protein F, 350/400ka (mitosin)	G2/M
1	M	CKAP5	cytoskeleton associated protein 5	
1	M	DHX57	DEAH (Asp-Glu-Ala-Asp/His) box polypeptide 57	
1	M	DLG7	discs, large homolog 7 (Drosophila)	
1	M	ECT2	epithelial cell transforming sequence 2 oncogene	
1	M	FAM64A	family with sequence similarity 64, member A	
1	M	HMGB3	high-mobility group box 3	
1	M	HMMR	hyaluronan-mediated motility receptor (RHAMM)	
1	M	HRB	HIV-1 Rev binding protein	
1	M	KIF14	kinesin family member 14	
1	M	MAPK13	mitogen-activated protein kinase 13	
1	M	NEK2	NIMA (never in mitosis gene a)-related kinase 2	
1	M	PLK1	polo-like kinase 1 (Drosophila)	G2/M
1	M	PRR11	proline rich 11	
1	M	SFPQ	splicing factor proline/glutamine-rich (polypyrimidine tract binding protein associated)	
1	M	SMTN	smoothelin	
1	M	SPAG5	sperm associated antigen 5	
1	M	TPX2	TPX2, microtubule-associated, homolog (<i>Xenopus laevis</i>)	
1	M	TROAP	trophinin associated protein (tastin)	
1	M	TSN	translin	
1	M	VANGL1	vang-like 1 (van gogh, Drosophila)	
1	M	WAPAL	wings apart-like homolog (Drosophila)	
1	M	WDR77	WD repeat domain 77	
2	M/G1	BTBD3	BTB (POZ) domain containing 3	
2	M/G1	C1orf174	chromosome 1 open reading frame 174	
2	M/G1	CACNA2D3	calcium channel, voltage-dependent, alpha 2/delta 3 subunit	
2	M/G1	CDKN3	cyclin-dependent kinase inhibitor 3 (CDK2-associated dual specificity phosphatase)	M/G1
2	M/G1	CENPI	centromere protein I	
2	M/G1	DHX16	DEAH (Asp-Glu-Ala-His) box polypeptide 16	
2	M/G1	HSPA8	heat shock 70kDa protein 8	
2	M/G1	NUP98	nucleoporin 98kDa	
2	M/G1	PPP2R2A	protein phosphatase 2 (formerly 2A), regulatory subunit B, alpha isoform	
2	M/G1	PTTG1	pituitary tumor-transforming 1	
2	M/G1	RBM8A	RNA binding motif protein 8A	

2	M/G1	SEPHS1	selenophosphate synthetase 1	
2	M/G1	SYNCRIP	synaptotagmin binding, cytoplasmic RNA interacting protein	
2	M/G1	VANGL1	vang-like 1 (van gogh, Drosophila)	
3	M/G1	DKC1	dyskeratosis congenita 1, dyskerin	
3	M/G1	DYNLL1	dynein, light chain, LC8-type 1	
3	M/G1	FUBP1	far upstream element (FUSE) binding protein 1	
3	M/G1	GSPT1	G1 to S phase transition 1	
3	M/G1	HERC4	hect domain and RLD 4	
3	M/G1	HSPC111	hypothetical protein HSPC111	
3	M/G1	LSM5	LSM5 homolog, U6 small nuclear RNA associated (S. cerevisiae)	
3	M/G1	SEPHS1	selenophosphate synthetase 1	
3	M/G1	TAF9	TAF9 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 32kDa	
4	G1	CSK	c-src tyrosine kinase	
4	G1	GRPEL1	GrpE-like 1, mitochondrial (E. coli)	
4	G1	GSPT1	G1 to S phase transition 1	
4	G1	KHDRBS1	KH domain containing, RNA binding, signal transduction associated 1	
4	G1	MAPK1	mitogen-activated protein kinase 1	
4	G1	PTP4A2	protein tyrosine phosphatase type IVA, member 2	
4	G1	SYNCRIP	synaptotagmin binding, cytoplasmic RNA interacting protein	
5	G1	CCT4	chaperonin containing TCP1, subunit 4 (delta)	
5	G1	HSPA4	heat shock 70kDa protein 4	
5	G1	PLEKHA2	pleckstrin homology domain containing, family A (phosphoinositide binding specific) member 2	
5	G1	RER1	RER1 retention in endoplasmic reticulum 1 homolog (S. cerevisiae)	
5	G1	SRP19	signal recognition particle 19kDa	
6	G1	HSPA4	heat shock 70kDa protein 4	
6	G1	SR140	U2-associated SR140 protein	
6	G1	TSPYL4	TSPY-like 4	
6	G1	YES1	v-yes-1 Yamaguchi sarcoma viral oncogene homolog 1	
7	G1	CSTF2T	cleavage stimulation factor, 3' pre-RNA, subunit 2, 64kDa, tau variant	
7	G1	HIG2	hypoxia-inducible protein 2	
7	G1	IMPAD1	inositol monophosphatase domain containing 1	
7	G1	RNF138	ring finger protein 138	
7	G1	SNRPA	small nuclear ribonucleoprotein polypeptide A	
7	G1	SRP72	signal recognition particle 72kDa	
8	G1	GPR107	G protein-coupled receptor 107	
8	G1	SLBP	stem-loop (histone) binding protein	G1/S
8	G1	WWTR1	WW domain containing transcription regulator 1	
9	G1	ARGLU1	arginine and glutamate rich 1	
9	G1	C6orf211	chromosome 6 open reading frame 211	
9	G1	CASP8AP2	CASP8 associated protein 2	
9	G1	DHX29	DEAH (Asp-Glu-Ala-His) box polypeptide 29	
9	G1	MTCH1	mitochondrial carrier homolog 1 (C. elegans)	
9	G1	PANK2	pantothenate kinase 2 (Hallervorden-Spatz syndrome)	
9	G1	PLCXD1	phosphatidylinositol-specific phospholipase C, X domain containing 1	
9	G1	SYNCRIP	synaptotagmin binding, cytoplasmic RNA interacting protein	
10	G1/S	ARGLU1	arginine and glutamate rich 1	

10	G1/S	BARD1	BRCA1 associated RING domain 1	
10	G1/S	C1GALT1	core 1 synthase, glycoprotein-N-acetylgalactosamine 3-beta-galactosyltransferase, 1	
10	G1/S	CDC25A	cell division cycle 25 homolog A (S. pombe)	G1/S
10	G1/S	DTL	denticleless homolog (Drosophila)	
10	G1/S	E2F1	E2F transcription factor 1	G1/S
10	G1/S	EIF2S1	eukaryotic translation initiation factor 2, subunit 1 alpha, 35kDa	
10	G1/S	FN3KRP	fructosamine-3-kinase-related protein	
10	G1/S	GINS2	GINS complex subunit 2 (Psf2 homolog)	
10	G1/S	HNRPD	heterogeneous nuclear ribonucleoprotein D (AU-rich element RNA binding protein 1, 37kDa)	
10	G1/S	IVNS1ABP	influenza virus NS1A binding protein	
10	G1/S	KIF1A	kinesin family member 1A	
10	G1/S	MBD4	methyl-CpG binding domain protein 4	
10	G1/S	MCM6	minichromosome maintenance complex component 6	G1/S
10	G1/S	NASP	nuclear autoantigenic sperm protein (histone-binding)	G1/S
10	G1/S	PANK2	pantothenate kinase 2 (Hallervorden-Spatz syndrome)	
10	G1/S	PNN	pinin, desmosome associated protein	
10	G1/S	SENP5	SUMO1/sentrin specific peptidase 5	
10	G1/S	SIVA1	SIVA1, apoptosis-inducing factor	
10	G1/S	STAM	signal transducing adaptor molecule (SH3 domain and ITAM motif) 1	
11	S	APEX2	APEX nuclease (apurinic/aprimidinic endonuclease) 2	
11	S	C14orf130	chromosome 14 open reading frame 130	
11	S	CASP2	caspase 2, apoptosis-related cysteine peptidase (neural precursor cell expressed, developmentally down-regulated 2)	
11	S	CCT4	chaperonin containing TCP1, subunit 4 (delta)	
11	S	CDC6	cell division cycle 6 homolog (S. cerevisiae)	G1/S
11	S	CDK2	cyclin-dependent kinase 2	
11	S	CHAF1B	chromatin assembly factor 1, subunit B (p60)	
11	S	DCC1	defective in sister chromatid cohesion homolog 1 (S. cerevisiae)	
11	S	DONSON	downstream neighbor of SON	
11	S	GMNN	geminin, DNA replication inhibitor	
11	S	MCM4	minichromosome maintenance complex component 4	
11	S	MIS12	MIS12, MIND kinetochore complex component, homolog (yeast)	
11	S	MSH2	mutS homolog 2, colon cancer, nonpolyposis type 1 (E. coli)	G1/S
11	S	MYCBP	c-myc binding protein	
11	S	NUP43	nucleoporin 43kDa	
11	S	ORC1L	origin recognition complex, subunit 1-like (yeast)	
11	S	PCNA	proliferating cell nuclear antigen	G1/S
11	S	PKMYT1	protein kinase, membrane associated tyrosine/threonine 1	
11	S	POLD3	polymerase (DNA-directed), delta 3, accessory subunit	
11	S	POLE	polymerase (DNA directed), epsilon	
11	S	RBBP8	retinoblastoma binding protein 8	
11	S	RFC2	replication factor C (activator 1) 2, 40kDa	
11	S	RFC4	replication factor C (activator 1) 4, 37kDa	
11	S	RSRC2	arginine/serine-rich coiled-coil 2	
11	S	RUSC1	RUN and SH3 domain containing 1	

11	S	SEC23B	Sec23 homolog B (<i>S. cerevisiae</i>)	
11	S	SETD3	SET domain containing 3	
11	S	SH3GLB1	SH3-domain GRB2-like endophilin B1	
11	S	TOPBP1	topoisomerase (DNA) II binding protein 1	
11	S	USP1	ubiquitin specific peptidase 1	
12	S	ABHD5	abhydrolase domain containing 5	
12	S	ADAMTS1	ADAM metallopeptidase with thrombospondin type 1 motif, 1	
12	S	ATAD2	ATPase family, AAA domain containing 2	
12	S	BLM	Bloom syndrome	
12	S	BRCA1	breast cancer 1, early onset	S
12	S	C1orf41	chromosome 1 open reading frame 41	
12	S	CCND3	cyclin D3	
12	S	CENPQ	centromere protein Q	
12	S	DCK	deoxycytidine kinase	
12	S	DHFR	dihydrofolate reductase	S
12	S	EED	embryonic ectoderm development	
12	S	ENOSF1	enolase superfamily member 1	
12	S	EXO1	exonuclease 1	
12	S	EZH2	enhancer of zeste homolog 2 (<i>Drosophila</i>)	
12	S	FANCA	Fanconi anemia, complementation group A	
12	S	FANCG	Fanconi anemia, complementation group G	
12	S	FEN1	flap structure-specific endonuclease 1	
12	S	GINS3	GINS complex subunit 3 (<i>Psf3</i> homolog)	
12	S	GPD2	glycerol-3-phosphate dehydrogenase 2 (mitochondrial)	
12	S	HAT1	histone acetyltransferase 1	
12	S	HELLS	helicase, lymphoid-specific	
12	S	ICMT	isoprenylcysteine carboxyl methyltransferase	
12	S	KIAA0101	KIAA0101	
12	S	KIAA0460	KIAA0460	
12	S	KNTC1	kinetochore associated 1	
12	S	MLF1IP	MLF1 interacting protein	
12	S	ORC6L	origin recognition complex, subunit 6 like (<i>yeast</i>)	
12	S	RMI1	RMI1, RecQ mediated genome instability 1, homolog (<i>S. cerevisiae</i>)	
12	S	RPA2	replication protein A2, 32kDa	
12	S	USP1	ubiquitin specific peptidase 1	
13	S	ASF1B	ASF1 anti-silencing function 1 homolog B (<i>S. cerevisiae</i>)	
13	S	HELLS	helicase, lymphoid-specific	
13	S	KIAA0101	KIAA0101	
13	S	MLF1IP	MLF1 interacting protein	
13	S	PRIM2	primase, DNA, polypeptide 2 (58kDa)	
13	S	RFC3	replication factor C (activator 1) 3, 38kDa	
13	S	RRM2	ribonucleotide reductase M2 polypeptide	S
13	S	STOML1	stomatol (EPB72)-like 1	
13	S	TMEM106C	transmembrane protein 106C	
14	S	C18orf24	chromosome 18 open reading frame 24	
14	S	DUSP6	dual specificity phosphatase 6	
14	S	FANCI	Fanconi anemia, complementation group I	
14	S	GPD2	glycerol-3-phosphate dehydrogenase 2 (mitochondrial)	
14	S	NRD1	nardilysin (N-arginine dibasic convertase)	
14	S	RRM1	ribonucleotide reductase M1 polypeptide	S
14	S	RRM2	ribonucleotide reductase M2 polypeptide	S
14	S	SAR1B	SAR1 gene homolog B (<i>S. cerevisiae</i>)	
14	S	YWHAH	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, eta polypeptide	
15	S/G2	AURKB	aurora kinase B	
15	S/G2	CDC7	cell division cycle 7 homolog (<i>S. cerevisiae</i>)	

15	S/G2	ERCC6L	excision repair cross-complementing rodent repair deficiency, complementation group 6-like	
15	S/G2	HIRIP3	HIRA interacting protein 3	
15	S/G2	NUP50	nucleoporin 50kDa	
15	S/G2	STIL	SCL/TAL1 interrupting locus	
15	S/G2	UMPS	uridine monophosphate synthetase (orotate phosphoribosyl transferase and orotidine-5'-decarboxylase)	
16	G2	AURKB	aurora kinase B	
16	G2	CDCA3	cell division cycle associated 3	
16	G2	CDKN2C	cyclin-dependent kinase inhibitor 2C (p18, inhibits CDK4)	G2
16	G2	CEP76	centrosomal protein 76kDa	
16	G2	H2AFX	H2A histone family, member X	
16	G2	HJURP	Holliday junction recognition protein	
16	G2	HMGB2	high-mobility group box 2	
16	G2	KIF11	kinesin family member 11	
16	G2	KIF23	kinesin family member 23	
16	G2	KIFC1	kinesin family member C1	
16	G2	MAD2L1	MAD2 mitotic arrest deficient-like 1 (yeast)	
16	G2	NCAPG2	non-SMC condensin II complex, subunit G2	
16	G2	NEIL3	nei endonuclease VIII-like 3 (E. coli)	
16	G2	OPCML	opioid binding protein/cell adhesion molecule-like	
16	G2	PPP1R2	protein phosphatase 1, regulatory (inhibitor) subunit 2	
16	G2	PSMD11	proteasome (prosome, macropain) 26S subunit, non-ATPase, 11	
17	G2	ARHGDIB	Rho GDP dissociation inhibitor (GDI) beta	
17	G2	C13orf34	chromosome 13 open reading frame 34	
17	G2	C14orf106	chromosome 14 open reading frame 106	
17	G2	CASP3	caspase 3, apoptosis-related cysteine peptidase	
17	G2	CCNA2	cyclin A2	G2
17	G2	CCNF	cyclin F	G2
17	G2	CDC25C	cell division cycle 25 homolog C (S. pombe)	G2
17	G2	CDCA8	cell division cycle associated 8	
17	G2	CENPA	centromere protein A	G2/M
17	G2	CKAP2	cytoskeleton associated protein 2	
17	G2	CKS1B	CDC28 protein kinase regulatory subunit 1B	
17	G2	DTYMK	deoxythymidylate kinase (thymidylate kinase)	
17	G2	FZR1	fizzy/cell division cycle 20 related 1 (Drosophila)	
17	G2	GABPB2	GA binding protein transcription factor, beta subunit 2	
17	G2	GLI2	GLI-Kruppel family member GLI2	
17	G2	KIF23	kinesin family member 23	
17	G2	KPNA2	karyopherin alpha 2 (RAG cohort 1, importin alpha 1)	
17	G2	LMNB1	lamin B1	
17	G2	LOC727761	similar to deoxythymidylate kinase (thymidylate kinase)	
17	G2	LOC728860	karyopherin alpha-2 subunit like	
17	G2	MELK	maternal embryonic leucine zipper kinase	
17	G2	MKI67	antigen identified by monoclonal antibody Ki-67	
17	G2	NCAPH	non-SMC condensin I complex, subunit H	
17	G2	NUSAP1	nucleolar and spindle associated protein 1	
17	G2	OXS1	oxidative-stress responsive 1	
17	G2	PPP1R12A	protein phosphatase 1, regulatory (inhibitor) subunit 12A	
17	G2	PSRC1	proline/serine-rich coiled-coil 1	

17	G2	RANGAP1	Ran GTPase activating protein 1	
17	G2	SHCBP1	SHC SH2-domain binding protein 1	
17	G2	SMC4	structural maintenance of chromosomes 4	
17	G2	TIMP1	TIMP metalloproteinase inhibitor 1	
17	G2	TOP2A	topoisomerase (DNA) II alpha 170kDa	G2
17	G2	TTK	TTK protein kinase	
17	G2	TUBA4A	tubulin, alpha 4a	
17	G2	TUBB2C	tubulin, beta 2C	
17	G2	TUBD1	tubulin, delta 1	
17	G2	UBE2C	ubiquitin-conjugating enzyme E2C	
18	M	ANP32E	acidic (leucine-rich) nuclear phosphoprotein 32 family, member E	
18	M	ARHGAP19	Rho GTPase activating protein 19	
18	M	AURKA	aurora kinase A	G2/M
18	M	BUB1B	BUB1 budding uninhibited by benzimidazoles 1 homolog beta (yeast)	G2/M
18	M	C12orf24	chromosome 12 open reading frame 24	
18	M	CCDC99	coiled-coil domain containing 99	
18	M	CDCA3	cell division cycle associated 3	
18	M	CENPE	centromere protein E, 312kDa	
18	M	CENPF	centromere protein F, 350/400ka (mitosin)	G2/M
18	M	CEP55	centrosomal protein 55kDa	
18	M	CKAP2	cytoskeleton associated protein 2	
18	M	CKS2	CDC28 protein kinase regulatory subunit 2	G2/M
18	M	EIF2C2	eukaryotic translation initiation factor 2C, 2	
18	M	FZR1	fizzy/cell division cycle 20 related 1 (Drosophila)	
18	M	GTSE1	G-2 and S-phase expressed 1	
18	M	HN1	hematological and neurological expressed 1	
18	M	KIF2C	kinesin family member 2C	
18	M	KIF5B	kinesin family member 5B	
18	M	KPNA2	karyopherin alpha 2 (RAG cohort 1, importin alpha 1)	
18	M	LOC728860	karyopherin alpha-2 subunit like	
18	M	MCM4	minichromosome maintenance complex component 4	
18	M	MPHOSP1	M-phase phosphoprotein 1	
18	M	NUSAP1	nucleolar and spindle associated protein 1	
18	M	PBK	PDZ binding kinase	
18	M	PPIF	peptidylprolyl isomerase F (cyclophilin F)	
18	M	PPP2CA	protein phosphatase 2 (formerly 2A), catalytic subunit, alpha isoform	
18	M	PRC1	protein regulator of cytokinesis 1	
18	M	QSER1	glutamine and serine rich 1	
18	M	RAD51C	RAD51 homolog C (S. cerevisiae)	
18	M	RAN	RAN, member RAS oncogene family	
18	M	RANGAP1	Ran GTPase activating protein 1	
18	M	RCAN1	regulator of calcineurin 1	
18	M	RNF126	ring finger protein 126	
18	M	SMC4	structural maintenance of chromosomes 4	
18	M	SPAG5	sperm associated antigen 5	
18	M	TPX2	TPX2, microtubule-associated, homolog (Xenopus laevis)	
18	M	TRAIP	TRAF interacting protein	
18	M	YWHAH	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, eta polypeptide	

別表2 PrognoScanでのMKI67の評価結果

DATASET	CANCER TYPE	SUBTYPE	ENDPOINT	COHORT	CONTRIBUTOR	ARRAY TYPE	PROBE ID	N	CUTPOINT	MINIMUM P-VALUE	CORRECTED P-VALUE
GSE5287	Bladder cancer		Overall Survival	Aarhus (1995-2004)	Als	HG-U133A	212021_s_at	n=30	0.77	0.000194	0.0063
GSE5287	Bladder cancer		Overall Survival	Aarhus (1995-2004)	Als	HG-U133A	212020_s_at	n=30	0.67	0.004441	0.0862
GSE5287	Bladder cancer		Overall Survival	Aarhus (1995-2004)	Als	HG-U133A	212023_s_at	n=30	0.6	0.002021	0.0457
GSE5287	Bladder cancer		Overall Survival	Aarhus (1995-2004)	Als	HG-U133A	212022_s_at	n=30	0.83	0.001044	0.0264
GSE13507	Bladder cancer	Transitional cell carcinoma	Overall Survival	Cheongju	Kim	Human-6v2	ILLMN_1734827	n=165	0.88	0.01568	0.2251
GSE12417-GPL96	Blood cancer	AML	Overall Survival	AMLCCG (1999-2003)	Metzeler	HG-U133A	212023_s_at	n=163	0.53	0.003224	0.0667
GSE12417-GPL96	Blood cancer	AML	Overall Survival	AMLCCG (1999-2003)	Metzeler	HG-U133A	212022_s_at	n=163	0.21	0.013904	0.2063
GSE12417-GPL96	Blood cancer	AML	Overall Survival	AMLCCG (1999-2003)	Metzeler	HG-U133A	212021_s_at	n=163	0.1	0.007084	0.1242
GSE12417-GPL96	Blood cancer	AML	Overall Survival	AMLCCG (1999-2003)	Metzeler	HG-U133A	212020_s_at	n=163	0.14	0.018785	0.2563
GSE12417-GPL570	Blood cancer	AML	Overall Survival	AMLCCG (2004)	Metzeler	HG-U133_P1us_2	212020_s_at	n=79	0.77	0.071563	-
GSE12417-GPL570	Blood cancer	AML	Overall Survival	AMLCCG (2004)	Metzeler	HG-U133_P1us_2	212021_s_at	n=79	0.77	0.138803	-
GSE12417-GPL570	Blood cancer	AML	Overall Survival	AMLCCG (2004)	Metzeler	HG-U133_P1us_2	212022_s_at	n=79	0.68	0.033558	0.3816
GSE12417-GPL570	Blood cancer	AML	Overall Survival	AMLCCG (2004)	Metzeler	HG-U133_P1us_2	212023_s_at	n=79	0.53	0.242668	-
GSE8970	Blood cancer	AML	Overall Survival	San Diego	Raponi	HG-U133A	212021_s_at	n=34	0.44	0.104767	-
GSE8970	Blood cancer	AML	Overall Survival	San Diego	Raponi	HG-U133A	212020_s_at	n=34	0.5	0.043767	0.4528
GSE8970	Blood cancer	AML	Overall Survival	San Diego	Raponi	HG-U133A	212023_s_at	n=34	0.32	0.010696	0.1698
GSE8970	Blood cancer	AML	Overall Survival	San Diego	Raponi	HG-U133A	212022_s_at	n=34	0.47	0.112995	-
GSE4475	Blood cancer	B-cell lymphoma	Overall Survival	Berlin (2003-2005)	Hummel	HG-U133A	212023_s_at	n=158	0.64	0.001166	0.029
GSE4475	Blood cancer	B-cell lymphoma	Overall Survival	Berlin (2003-2005)	Hummel	HG-U133A	212022_s_at	n=158	0.18	0.011198	0.1758
GSE4475	Blood cancer	B-cell lymphoma	Overall Survival	Berlin (2003-2005)	Hummel	HG-U133A	212021_s_at	n=158	0.14	0.146069	-
GSE4475	Blood cancer	B-cell lymphoma	Overall Survival	Berlin (2003-2005)	Hummel	HG-U133A	212020_s_at	n=158	0.72	0.062168	-
E-TABM-346	Blood cancer	DLBCL	Event Free Survival	GELA (1998-2000)	Jais	HG-U133A	212023_s_at	n=53	0.74	0.023231	0.2976
E-TABM-346	Blood cancer	DLBCL	Overall Survival	GELA (1998-2000)	Jais	HG-U133A	212021_s_at	n=53	0.87	0.000448	0.0129
E-TABM-346	Blood cancer	DLBCL	Event Free Survival	GELA (1998-2000)	Jais	HG-U133A	212022_s_at	n=53	0.81	0.133429	-
E-TABM-346	Blood cancer	DLBCL	Overall Survival	GELA (1998-2000)	Jais	HG-U133A	212020_s_at	n=53	0.43	0.074494	-
E-TABM-346	Blood cancer	DLBCL	Event Free Survival	GELA (1998-2000)	Jais	HG-U133A	212021_s_at	n=53	0.87	0.000237	0.0074

E-TABM-346	Blood cancer	DLBCL	Overall Survival	GELA (1998-2000)	Jais	HG-U133A	212023_s_at	n=53	0.74	0.042956	0.4475
E-TABM-346	Blood cancer	DLBCL	Event Free Survival	GELA (1998-2000)	Jais	HG-U133A	212020_s_at	n=53	0.36	0.241147	-
E-TABM-346	Blood cancer	DLBCL	Overall Survival	GELA (1998-2000)	Jais	HG-U133A	212022_s_at	n=53	0.51	0.19056	-
GSE2658	Blood cancer	Multiple myeloma	Cause Specific Survival	Arkansas	Zhan	HG-U133_P1 us_2	212021_s_at	n=559	0.84	0.000003	0.0002
GSE2658	Blood cancer	Multiple myeloma	Cause Specific Survival	Arkansas	Zhan	HG-U133_P1 us_2	212022_s_at	n=559	0.78	0.000042	0.0016
GSE2658	Blood cancer	Multiple myeloma	Cause Specific Survival	Arkansas	Zhan	HG-U133_P1 us_2	212023_s_at	n=559	0.83	0.000058	0.0022
GSE2658	Blood cancer	Multiple myeloma	Cause Specific Survival	Arkansas	Zhan	HG-U133_P1 us_2	212020_s_at	n=559	0.83	0.005044	0.0953
GSE3143	Breast cancer		Overall Survival	Duke	Bild	HG-U95A	419_at	n=158	0.67	0.000591	0.0163
GSE3143	Breast cancer		Overall Survival	Duke	Bild	HG-U95A	418_at	n=158	0.62	0.00106	0.0268
GSE7849	Breast cancer		Disease Free Survival	Duke (1990-2001)	Anders	HG-U95A	419_at	n=76	0.28	0.133407	-
GSE7849	Breast cancer		Disease Free Survival	Duke (1990-2001)	Anders	HG-U95A	418_at	n=76	0.49	0.320109	-
GSE6532-GPL570	Breast cancer		Recurrence Free Survival	GUYT	Loi	HG-U133_P1 us_2	212022_s_at	n=87	0.75	0.000246	0.0077
GSE6532-GPL570	Breast cancer		Distant Metastasis Free Survival	GUYT	Loi	HG-U133_P1 us_2	212023_s_at	n=87	0.63	0.000898	0.0233
GSE6532-GPL570	Breast cancer		Distant Metastasis Free Survival	GUYT	Loi	HG-U133_P1 us_2	212020_s_at	n=87	0.36	0.052057	-
GSE6532-GPL570	Breast cancer		Recurrence Free Survival	GUYT	Loi	HG-U133_P1 us_2	212023_s_at	n=87	0.63	0.000898	0.0233
GSE6532-GPL570	Breast cancer		Recurrence Free Survival	GUYT	Loi	HG-U133_P1 us_2	212020_s_at	n=87	0.36	0.052057	-
GSE6532-GPL570	Breast cancer		Distant Metastasis Free Survival	GUYT	Loi	HG-U133_P1 us_2	212021_s_at	n=87	0.78	0.000073	0.0027
GSE6532-GPL570	Breast cancer		Recurrence Free Survival	GUYT	Loi	HG-U133_P1 us_2	212021_s_at	n=87	0.78	0.000073	0.0027
GSE6532-GPL570	Breast cancer		Distant Metastasis Free Survival	GUYT	Loi	HG-U133_P1 us_2	212022_s_at	n=87	0.75	0.000246	0.0077
GSE9195	Breast cancer		Distant Metastasis Free Survival	GUYT2	Loi	HG-U133_P1 us_2	212020_s_at	n=77	0.88	0.001684	0.0393
GSE9195	Breast cancer		Recurrence Free Survival	GUYT2	Loi	HG-U133_P1 us_2	212022_s_at	n=77	0.79	0.005568	0.103
GSE9195	Breast cancer		Distant Metastasis Free Survival	GUYT2	Loi	HG-U133_P1 us_2	212021_s_at	n=77	0.87	0.002832	0.0601
GSE9195	Breast cancer		Recurrence Free Survival	GUYT2	Loi	HG-U133_P1 us_2	212023_s_at	n=77	0.51	0.00005	0.0019
GSE9195	Breast cancer		Recurrence Free Survival	GUYT2	Loi	HG-U133_P1 us_2	212020_s_at	n=77	0.69	0.000491	0.0139

GSE9195	Breast cancer	Distant Metastasis Free Survival	GUYT2	Loi	HG-U133_Plus_2	212022_s_at	n=77	0.74	0.004292	0.0839
GSE9195	Breast cancer	Recurrence Free Survival	GUYT2	Loi	HG-U133_Plus_2	212021_s_at	n=77	0.3	0.011734	0.182
GSE9195	Breast cancer	Distant Metastasis Free Survival	GUYT2	Loi	HG-U133_Plus_2	212023_s_at	n=77	0.51	0.000447	0.0129
GSE11121	Breast cancer	Distant Metastasis Free Survival	Mainz (1988-1998)	Schmidt	HG-U133A	212022_s_at	n=200	0.89	0.000001	0.0001
GSE11121	Breast cancer	Distant Metastasis Free Survival	Mainz (1988-1998)	Schmidt	HG-U133A	212021_s_at	n=200	0.89	0	0
GSE11121	Breast cancer	Distant Metastasis Free Survival	Mainz (1988-1998)	Schmidt	HG-U133A	212020_s_at	n=200	0.89	0.025152	0.3143
GSE11121	Breast cancer	Distant Metastasis Free Survival	Mainz (1988-1998)	Schmidt	HG-U133A	212023_s_at	n=200	0.9	0.000007	0.0003
GSE1378	Breast cancer	Recurrence Free Survival	MGH (1987-2000)	Ma	Arcturus 22k	3150	n=60	0.88	0.007407	0.1285
GSE1379	Breast cancer	Recurrence Free Survival	MGH (1987-2000)	Ma	Arcturus 22k	3150	n=60	0.83	0.009133	0.1508
GSE9893	Breast cancer	Overall Survival	Montpellier, Bordeaux, Turin (1989-2001)	Chanrion	MLRG Human 21K V12.0	6175	n=155	0.17	0.015742	0.2258
GSE2034	Breast cancer	Distant Metastasis Free Survival	Rotterdam (1980-1995)	Wang	HG-U133A	212022_s_at	n=286	0.47	0.000549	0.0153
GSE2034	Breast cancer	Distant Metastasis Free Survival	Rotterdam (1980-1995)	Wang	HG-U133A	212021_s_at	n=286	0.57	0.02854	0.3426
GSE2034	Breast cancer	Distant Metastasis Free Survival	Rotterdam (1980-1995)	Wang	HG-U133A	212020_s_at	n=286	0.34	0.000248	0.0078
GSE2034	Breast cancer	Distant Metastasis Free Survival	Rotterdam (1980-1995)	Wang	HG-U133A	212023_s_at	n=286	0.22	0.075951	-
GSE1456-GPL96	Breast cancer	Overall Survival	Stockholm (1994-1996)	Pawitan	HG-U133A	212020_s_at	n=159	0.7	0.002856	0.0605
GSE1456-GPL96	Breast cancer	Recurrence Free Survival	Stockholm (1994-1996)	Pawitan	HG-U133A	212021_s_at	n=159	0.55	0.000361	0.0107
GSE1456-GPL96	Breast cancer	Recurrence Free Survival	Stockholm (1994-1996)	Pawitan	HG-U133A	212020_s_at	n=159	0.14	0.020654	0.2741
GSE1456-GPL96	Breast cancer	Overall Survival	Stockholm (1994-1996)	Pawitan	HG-U133A	212023_s_at	n=159	0.5	0.004338	0.0846
GSE1456-GPL96	Breast cancer	Overall Survival	Stockholm (1994-1996)	Pawitan	HG-U133A	212022_s_at	n=159	0.31	0.000011	0.0005
GSE1456-GPL96	Breast cancer	Recurrence Free Survival	Stockholm (1994-1996)	Pawitan	HG-U133A	212023_s_at	n=159	0.43	0.00264	0.0568
GSE1456-GPL96	Breast cancer	Recurrence Free Survival	Stockholm (1994-1996)	Pawitan	HG-U133A	212022_s_at	n=159	0.56	0.000023	0.001

GSE1456-GPL96	Breast cancer	Overall Survival	Stockholm (1994-1996)	Pawitan	HG-U133A	212021_s_at	n=159	0.54	0.001405	0.0338
GSE7378	Breast cancer	Disease Free Survival	UCSF	Zhou	U133AA ofAv2	212023_s_at	n=54	0.85	0.003121	0.065
GSE7378	Breast cancer	Disease Free Survival	UCSF	Zhou	U133AA ofAv2	212020_s_at	n=54	0.8	0.001466	0.0351
GSE7378	Breast cancer	Disease Free Survival	UCSF	Zhou	U133AA ofAv2	212021_s_at	n=54	0.89	0.020197	0.2698
GSE7378	Breast cancer	Disease Free Survival	UCSF	Zhou	U133AA ofAv2	212022_s_at	n=54	0.72	0.019667	0.2648
E-TABM-158	Breast cancer	Overall Survival	UCSF, CPMC (1989-1997)	Chin	HG-U133A	212020_s_at	n=129	0.68	0.058697	-
E-TABM-158	Breast cancer	Distant Metastasis Free Survival	UCSF, CPMC (1989-1997)	Chin	HG-U133A	212020_s_at	n=129	0.68	0.027221	0.3318
E-TABM-158	Breast cancer	Recurrence Free Survival	UCSF, CPMC (1989-1997)	Chin	HG-U133A	212020_s_at	n=129	0.68	0.072504	-
E-TABM-158	Breast cancer	Overall Survival	UCSF, CPMC (1989-1997)	Chin	HG-U133A	212023_s_at	n=129	0.19	0.025889	0.3206
E-TABM-158	Breast cancer	Distant Metastasis Free Survival	UCSF, CPMC (1989-1997)	Chin	HG-U133A	212023_s_at	n=129	0.11	0.012299	0.1884
E-TABM-158	Breast cancer	Recurrence Free Survival	UCSF, CPMC (1989-1997)	Chin	HG-U133A	212023_s_at	n=129	0.26	0.042315	0.4432
E-TABM-158	Breast cancer	Distant Metastasis Free Survival	UCSF, CPMC (1989-1997)	Chin	HG-U133A	212022_s_at	n=129	0.15	0.197926	-
E-TABM-158	Breast cancer	Recurrence Free Survival	UCSF, CPMC (1989-1997)	Chin	HG-U133A	212022_s_at	n=129	0.7	0.056973	-
E-TABM-158	Breast cancer	Overall Survival	UCSF, CPMC (1989-1997)	Chin	HG-U133A	212022_s_at	n=129	0.74	0.03218	0.3712
E-TABM-158	Breast cancer	Overall Survival	UCSF, CPMC (1989-1997)	Chin	HG-U133A	212021_s_at	n=129	0.86	0.049598	0.4892
E-TABM-158	Breast cancer	Distant Metastasis Free Survival	UCSF, CPMC (1989-1997)	Chin	HG-U133A	212021_s_at	n=129	0.74	0.057109	-
E-TABM-158	Breast cancer	Recurrence Free Survival	UCSF, CPMC (1989-1997)	Chin	HG-U133A	212021_s_at	n=129	0.87	0.032434	0.3732
GSE3494-GPL96	Breast cancer	Disease Specific Survival	Uppsala (1987-1989)	Miller	HG-U133A	212022_s_at	n=236	0.52	0.000109	0.0038
GSE3494-GPL96	Breast cancer	Disease Specific Survival	Uppsala (1987-1989)	Miller	HG-U133A	212021_s_at	n=236	0.88	0.000167	0.0055
GSE3494-GPL96	Breast cancer	Disease Specific Survival	Uppsala (1987-1989)	Miller	HG-U133A	212020_s_at	n=236	0.75	0.017337	0.242
GSE3494-GPL96	Breast cancer	Disease Specific Survival	Uppsala (1987-1989)	Miller	HG-U133A	212023_s_at	n=236	0.36	0.000166	0.0055
GSE4922-GPL96	Breast cancer	Disease Free Survival	Uppsala (1987-1989)	Ivshina	HG-U133A	212022_s_at	n=249	0.52	0.000009	0.0004

GSE4922-GPL96	Breast cancer	Disease Free Survival	Uppsala (1987-1989)	Ivshina	HG-U133A	212021_s_at	n=249	0.88	0.000549	0.0154
GSE4922-GPL96	Breast cancer	Disease Free Survival	Uppsala (1987-1989)	Ivshina	HG-U133A	212020_s_at	n=249	0.31	0.005634	0.1039
GSE4922-GPL96	Breast cancer	Disease Free Survival	Uppsala (1987-1989)	Ivshina	HG-U133A	212023_s_at	n=249	0.43	0.00001	0.0004
GSE2990	Breast cancer	Distant Metastasis Free Survival	Uppsala, Oxford	Sotiriou	HG-U133A	212023_s_at	n=179	0.78	0.026246	0.3236
GSE2990	Breast cancer	Relapse Free Survival	Uppsala, Oxford	Sotiriou	HG-U133A	212021_s_at	n=187	0.89	0.003521	0.0716
GSE2990	Breast cancer	Distant Metastasis Free Survival	Uppsala, Oxford	Sotiriou	HG-U133A	212022_s_at	n=179	0.6	0.000073	0.0027
GSE2990	Breast cancer	Relapse Free Survival	Uppsala, Oxford	Sotiriou	HG-U133A	212020_s_at	n=187	0.56	0.003686	0.0743
GSE2990	Breast cancer	Relapse Free Survival	Uppsala, Oxford	Sotiriou	HG-U133A	212023_s_at	n=187	0.74	0.030127	0.3553
GSE2990	Breast cancer	Distant Metastasis Free Survival	Uppsala, Oxford	Sotiriou	HG-U133A	212021_s_at	n=179	0.61	0.001607	0.0378
GSE2990	Breast cancer	Relapse Free Survival	Uppsala, Oxford	Sotiriou	HG-U133A	212022_s_at	n=187	0.28	0.001236	0.0304
GSE2990	Breast cancer	Distant Metastasis Free Survival	Uppsala, Oxford	Sotiriou	HG-U133A	212020_s_at	n=179	0.48	0.004083	0.0806
GSE7390	Breast cancer	Recurrence Free Survival	Uppsala, Oxford, Stockholm, IGR, GUYT, CRH (1980-1998)	Desmedt	HG-U133A	212022_s_at	n=198	0.4	0.00003	0.0012
GSE7390	Breast cancer	Overall Survival	Uppsala, Oxford, Stockholm, IGR, GUYT, CRH (1980-1998)	Desmedt	HG-U133A	212022_s_at	n=198	0.45	0.000015	0.0007
GSE7390	Breast cancer	Distant Metastasis Free Survival	Uppsala, Oxford, Stockholm, IGR, GUYT, CRH (1980-1998)	Desmedt	HG-U133A	212020_s_at	n=198	0.35	0.000525	0.0148
GSE7390	Breast cancer	Distant Metastasis Free Survival	Uppsala, Oxford, Stockholm, IGR, GUYT, CRH (1980-1998)	Desmedt	HG-U133A	212023_s_at	n=198	0.22	0.018245	0.251
GSE7390	Breast cancer	Recurrence Free Survival	Uppsala, Oxford, Stockholm, IGR, GUYT, CRH (1980-1998)	Desmedt	HG-U133A	212021_s_at	n=198	0.29	0.003341	0.0687

GSE7390	Breast cancer		Overall Survival	Uppsala, Oxford, Stockholm, IGR, GUYT, CRH (1980-1998)	Desmedt	HG-U133A	212021_s_at	n=198	0.29	0.000255	0.0079
GSE7390	Breast cancer		Distant Metastasis Free Survival	Uppsala, Oxford, Stockholm, IGR, GUYT, CRH (1980-1998)	Desmedt	HG-U133A	212022_s_at	n=198	0.45	0.00005	0.0019
GSE7390	Breast cancer		Recurrence Free Survival	Uppsala, Oxford, Stockholm, IGR, GUYT, CRH (1980-1998)	Desmedt	HG-U133A	212020_s_at	n=198	0.35	0.000283	0.0087
GSE7390	Breast cancer		Overall Survival	Uppsala, Oxford, Stockholm, IGR, GUYT, CRH (1980-1998)	Desmedt	HG-U133A	212020_s_at	n=198	0.46	0.000269	0.0083
GSE7390	Breast cancer		Overall Survival	Uppsala, Oxford, Stockholm, IGR, GUYT, CRH (1980-1998)	Desmedt	HG-U133A	212023_s_at	n=198	0.29	0.005634	0.1039
GSE7390	Breast cancer		Recurrence Free Survival	Uppsala, Oxford, Stockholm, IGR, GUYT, CRH (1980-1998)	Desmedt	HG-U133A	212023_s_at	n=198	0.89	0.040036	0.4279
GSE7390	Breast cancer		Distant Metastasis Free Survival	Uppsala, Oxford, Stockholm, IGR, GUYT, CRH (1980-1998)	Desmedt	HG-U133A	212021_s_at	n=198	0.29	0.000147	0.0049
GSE11595	Esophagus cancer	Adenocarcinoma	Overall Survival	Sutton	Giddings	CRUKD MF_22K_v1.0.0	108718	n=34	0.85	0.017642	0.2451
GSE4271-GPL96	Glioma		Overall Survival	MDA	Phillips	HG-U133A	212021_s_at	n=77	0.3	0.00177	0.041
GSE4271-GPL96	Glioma		Overall Survival	MDA	Phillips	HG-U133A	212020_s_at	n=77	0.82	0.134886	-
GSE4271-GPL96	Glioma		Overall Survival	MDA	Phillips	HG-U133A	212023_s_at	n=77	0.87	0.000361	0.0107
GSE4271-GPL96	Glioma		Overall Survival	MDA	Phillips	HG-U133A	212022_s_at	n=77	0.39	0.000919	0.0237
GSE7696	Glioma	Glioblastoma	Overall Survival	Lausanne	Murat	HG-U133_P1us_2	212021_s_at	n=70	0.33	0.103863	-
GSE7696	Glioma	Glioblastoma	Overall Survival	Lausanne	Murat	HG-U133_P1us_2	212022_s_at	n=70	0.53	0.247063	-
GSE7696	Glioma	Glioblastoma	Overall Survival	Lausanne	Murat	HG-U133_P1us_2	212023_s_at	n=70	0.13	0.047799	0.4782

GSE7696	Glioma	Glioblastoma	Overall Survival	Lausanne	Murat	HG-U133_Plus_2	212020_s_at	n=70	0.46	0.042958	0.4475
GSE2837	Head and neck cancer	Squamous cell carcinoma	Recurrence Free Survival	VUMC, VAMC, UTMDA CC (1992-2005)	Chung	U133_X3P	Hs_80976.1_S4_3p_a_at	n=28	0.86	0.019863	0.2666
GSE2837	Head and neck cancer	Squamous cell carcinoma	Recurrence Free Survival	VUMC, VAMC, UTMDA CC (1992-2005)	Chung	U133_X3P	Hs_80976.1_S1_3p_a_at	n=28	0.46	0.101345	-
GSE2837	Head and neck cancer	Squamous cell carcinoma	Recurrence Free Survival	VUMC, VAMC, UTMDA CC (1992-2005)	Chung	U133_X3P	Hs_80976.1_S2_3p_a_at	n=28	0.75	0.034818	0.391
GSE2837	Head and neck cancer	Squamous cell carcinoma	Recurrence Free Survival	VUMC, VAMC, UTMDA CC (1992-2005)	Chung	U133_X3P	Hs_80976.1_S3_3p_a_at	n=28	0.14	0.019381	0.2621
HARVARD-LC	Lung cancer	Adenocarcinoma	Overall Survival	Harvard	Beer	HG-U95A	418_at	n=84	0.54	0.264372	-
HARVARD-LC	Lung cancer	Adenocarcinoma	Overall Survival	Harvard	Beer	HG-U95A	419_at	n=84	0.71	0.000966	0.0247
MICHIGAN-LC	Lung cancer	Adenocarcinoma	Overall Survival	Michigan (1994-2000)	Beer	HuGene FL	X65550_at	n=86	0.45	0.062645	-
GSE3141	Lung cancer	NSCLC	Overall Survival	Duke	Bild	HG-U133_Plus_2	212023_s_at	n=111	0.32	0.34174	-
GSE3141	Lung cancer	NSCLC	Overall Survival	Duke	Bild	HG-U133_Plus_2	212020_s_at	n=111	0.12	0.007437	0.1289
GSE3141	Lung cancer	NSCLC	Overall Survival	Duke	Bild	HG-U133_Plus_2	212021_s_at	n=111	0.14	0.016956	0.2382
GSE3141	Lung cancer	NSCLC	Overall Survival	Duke	Bild	HG-U133_Plus_2	212022_s_at	n=111	0.18	0.00981	0.1591
GSE4716-GPL3694	Lung cancer	NSCLC	Overall Survival	Nagoya (1995-1996)	Tomida	GF200	1583	n=50	0.7	0.184956	-
GSE8894	Lung cancer	NSCLC	Recurrence Free Survival	Seoul	Son	HG-U133_Plus_2	212021_s_at	n=138	0.41	0.027116	0.3309
GSE8894	Lung cancer	NSCLC	Recurrence Free Survival	Seoul	Son	HG-U133_Plus_2	212022_s_at	n=138	0.36	0.160049	-
GSE8894	Lung cancer	NSCLC	Recurrence Free Survival	Seoul	Son	HG-U133_Plus_2	212023_s_at	n=138	0.19	0.162646	-
GSE8894	Lung cancer	NSCLC	Recurrence Free Survival	Seoul	Son	HG-U133_Plus_2	212020_s_at	n=138	0.25	0.037713	0.4118
GSE4573	Lung cancer	Squamous cell carcinoma	Overall Survival	Michigan (1991-2002)	Raponi	HG-U133A	212020_s_at	n=129	0.11	0.23612	-
GSE4573	Lung cancer	Squamous cell carcinoma	Overall Survival	Michigan (1991-2002)	Raponi	HG-U133A	212023_s_at	n=129	0.18	0.073251	-
GSE4573	Lung cancer	Squamous cell carcinoma	Overall Survival	Michigan (1991-2002)	Raponi	HG-U133A	212022_s_at	n=129	0.38	0.03014	0.3554
GSE4573	Lung cancer	Squamous cell carcinoma	Overall Survival	Michigan (1991-2002)	Raponi	HG-U133A	212021_s_at	n=129	0.32	0.053071	-
DUKE-OC	Ovarian cancer		Overall Survival	Duke	Bild	HG-U133A	212023_s_at	n=134	0.22	0.123857	-
DUKE-OC	Ovarian cancer		Overall Survival	Duke	Bild	HG-U133A	212022_s_at	n=134	0.75	0.257483	-
DUKE-OC	Ovarian cancer		Overall Survival	Duke	Bild	HG-U133A	212021_s_at	n=134	0.59	0.037868	0.4129
DUKE-OC	Ovarian cancer		Overall Survival	Duke	Bild	HG-U133A	212020_s_at	n=134	0.32	0.029157	0.3475

<u>GSE8841</u>	Ovarian cancer	Overall Survival	Milan	Mariani	G4100A	<u>11832</u>	n=83	0.61	0.003968	0.0788
<u>GSE8841</u>	Ovarian cancer	Overall Survival	Milan	Mariani	G4100A	<u>17302</u>	n=83	0.82	0.105103	-
<u>E-DKFZ-1</u>	Renal cell carcinoma	Overall Survival	RZPD	Sueltmann	A-RZPD-20	<u>rzpd.de:huber1:Reporte:IMAGE:376442</u>	n=74	0.41	0.032751	0.3756