

博士論文

一細胞 RNA-seq による肺腺がん細胞株薬剤反応多様性の解明

鹿島幸恵

目次

要旨

緒言

本文中の略語一覧

第一章：『一細胞 RNA-seq プラットフォームの技術評価』

1.1 実験方法

1.2 結果

1.3 考察

第二章：『データセットの相互補完と生物学的意義の解明』

2.1 実験方法

2.2 結果

2.3 考察

総括

データ公開

論文目録

参考文献

謝辞

要旨

scRNA-seq は、不均一性 heterogeneity を明らかにする際に非常に有力なツールである。しかしながら、現在までに使用されている一細胞解析のためのプラットフォームは、それぞれメリットとデメリットが存在する。

第一章では、五種類の肺腺がん細胞株から合計 536 細胞(micro-chamber system)と 5,4631 細胞(micro-droplet system)の scRNA-seq ライブラリーを作成し、各プラットフォームの特徴を明らかにした。はじめに、各プラットフォームから得られたデータセットを元に、データの再現性や、細胞ごとのシーケンス深度の検証、データセットの特徴を記載し、技術的評価を行った。バルクレベルの比較から、どちらのプラットフォームにおいてもデータは高い再現性を有することが示された。一方、一細胞レベルの比較から、解析対象細胞数の違いがシーケンスの深度に反映され、検出遺伝子数等のデータの相違に影響することが示唆された。続いて個別の遺伝子に着目して、両者のデータを比較した。高発現遺伝子はどちらのプラットフォームでも発現変化を計測することができるが、低発現遺伝子は、micro-droplet system においては、タグ数の不足から定量的解析が困難である、あるいは対応するシーケンスタグが検出されないことからその発現自体が検知できない細胞が相当数存在することが示唆された。MAPK/ERK パスウェイ全体で発現遺伝子数・タグ数を比較した場合に、両プラットフォームの差異は特に顕著であった。比較的発現レベルが高いと考えられる細胞周期依存的に発現が変化する遺伝子を用いた細胞周期の推定でも、micro-droplet system ではより多くの遺伝子のデータを必要とした。

二つのプラットフォームのデータセットには多くの相違点が見出された。しかし、細胞数やシーケンス深度を同じレベルに情報的に補正した解析からは、これらのデータセットは本質的に同一のものであることが示唆された。そこで、micro-chamber system のデータセットを元に LASSO を利用し、micro-droplet system のデータセットの欠損値の推定を試みた。いくつかの遺伝子については統計学的推定が可能であったが、妥当な推定が困難である遺伝子も相当数存在した。データセットの補完には、個別遺伝子に関する情報に加えて、複数の遺伝子の情報の統合が必要となる可能性が示唆された。

第二章では、得られた肺腺がんのデータセットから、ゲフィチニブに対する反応不均一性の解明を試みた。第一章の結果から、個々の一細胞 RNA-seq プラットフォームでは、単一のプラットフォームを用いても解析できる細胞数や遺伝子数が限定されることが示された。がん細胞の多様性の詳細を明らかにするためには、単独プラットフォームからのデータセットだけでは不十分であり、各データセットを相互補完する必要があると考えた。

データセット補完の手段として、複数の遺伝子をクラスター化することで情報量の集約を図る手法を検討した。micro-chamber system のデータセットをもとに WGCNA(weight gene co-relation network analysis)を介して、遺伝子共発現ネットワークを構築した。得られたモジュール(共発現遺伝子群)を利用した解析から、がん治療抵抗性との関与が示唆された DUSP1 と AURKA を中心とす

るモジュールに関して特異的な反応を示す細胞が同定された。さらに臨床がん検体のデータセットに照らして、その妥当性を検証した。TCGA(The Cancer Genome Atlas)を用いて DUSP1/AURKA 遺伝子発現と患者生存曲線（ Kaplan-Meier 曲線）を比較解析した。同遺伝子が高発現である検体においては、その臨床予後が不良であることが示唆された。これは、異なる一細胞 RNA-seq プラットフォームの異なるデータセットを組み合わせることにより、単一プラットフォームでは困難である解析を行うことができる可能性を示唆するものであった。

緒言

一細胞 RNA-seq の背景

近年、一細胞シーケンス技術の進歩が目覚ましい。細胞の変化を平均ではなく、個々の細胞レベルで解析することが可能となっている。現在までに、一細胞解析を用いた細胞集団の多様性に関する *in vivo*、*in vitro* の研究が数多く報告されている。

一細胞 RNA-seq(scRNA-seq)においては、細胞集団ははじめにマイクロピペットや、レーザーキャプチャー、FACS、微小流路、微小液滴などの様々な方法により一つずつの細胞へと分離される¹。その後、個々の細胞についてシーケンスライブラリーが調整され、次世代シーケンサーによるシーケンスが行われる。現在主流となっている細胞単離方法は技術的な面から、**micro-chamber system** と **micro-droplet system** の二つへ大別される。前者は、細胞を微小流路、FACS などのプラットフォームによって物理的に分離、その後逆転写、cDNA 増幅、アダプター付加といった処理を行う²⁻⁵。後者は、細胞を微小液滴技術により一つ一つ液滴に閉じ込め、細胞ごとに固有のバーコード配列をつけることで一細胞化を行い、以降、バーコードがつけられた細胞はまとめられ処理がされる。これらの形式の代表的なプラットフォームとして、前者ではフリーダイト社の C1 システム、後者では 10xGenomics 社の Chromium が挙げられる。どちらのプラットフォームも scRNA-seq に広く応用されているが、下記の第一章で述べるようにそれぞれに固有の利点と欠点がある。

micro-droplet system による scRNA-seq は、2015 年に Macosko らと Klein らの二つのグループから発表された。従来の scRNA-seq においては、多くの労力と高いコストが大きな問題となっていたが、**micro-droplet system** では、コストが低下し、同時に処理可能である細胞数も急激に増加した。使用する機器も一般的な研究室で設置可能なものであり、より多くの研究者が scRNA-seq を利用できるようになる可能性を提示した^{4,5}。この報告以降、**micro-droplet system** に基づくプラットフォームとその他のプラットフォームを比較検討する論文がいくつか発表された。Svenson らの論文では 19 のプロトコルが比較されている。ここでは複数の研究室にまたがるデータが比較され、**spike-in** を基準にプロトコル間での比較が行われている。Ziegenhain らの論文では、6 つの方法から得られた一細胞 RNA-seq のデータを解析法ごとに効率、正確性などに焦点を当て解析している。しかし、これらのデータセットはすべてのデータセットを総計してもわずか 583 細胞が扱われているのみであり、細胞集団に関するデータの検証は不十分である。また、いずれの論文も方法論の技術評価と検証が中心であり、異なるデータセットの相互補完は行われていない^{6,7}。

がんと一細胞解析

本研究では、解析対象に肺腺がん細胞を選択した。一細胞解析の対象の中でも、がんは特に重要なターゲットである⁸⁻¹¹。腫瘍組織は、がん細胞だけでなく、線維細胞、がん浸潤白血球、血管内皮細胞といった異なる様々な細胞から構成される複雑な細胞エコシステムであり、がんの病態の理解にはその全体像の理解が不可欠である¹²。がん微小環境の中では、常に変化が起き、個々の細胞が相互作用している。同じクローンのがん細胞であっても、がん進化の中で異なる性質を獲得し、全

く同じ細胞は存在しないため、がん不均一性(cancer heterogeneity)が生じる¹³⁻¹⁵。がんの治療抵抗性や悪性化が生じるメカニズムを理解する上で、これらのがん組織の細胞不均一性を解析することは非常に重要である^{16,17}。例えば、多くのがんはがん治療薬に対して抵抗性を獲得する。肺がんにおけるゲフィチニブは、抗がん剤の中でもそれに対する抵抗性に関して非常によく研究が進んでいる薬剤の一つである^{18,19}。抗がん剤耐性変異は、はじめはわずかな数の細胞で始まり、徐々に集団の中で拡大していくと考えられている。また、ゲノム解析に限らず、薬剤耐性変異はゲノム変異として固定される前に、トランスクリプトーム多様性として多くの細胞に現れることという報告もある²⁰。

従来の RNA シークエンスを用いた方法では、細胞は集団状態（バルク）のまま解析されるため、これらの細胞不均一性を明らかにすることは困難であった。実際に TCGA, ICGC,あるいは COSMIC といったがんのシークエンス解析における代表的なデータベースにおいても、がん不均一性 (cancer heterogeneity) に関するデータは限られている²¹⁻²⁴。

肺腺がんとうがん治療抵抗性

本研究では、scRNA-seq の対象として、肺腺がん由来細胞株である PC9, II-18, H1650, H1975, H2228 の5つの細胞株を用いた。モデルがん治療薬として、ゲフィチニブを選択した。ゲフィチニブは肺腺がんにおける EGFR をターゲットとした分子標的薬である。PC9 細胞は Exon19 での欠損、II-18 細胞は L858R 変異という形で EGFR 遺伝子のドライバー変異をそれぞれ有しており、これらの細胞株はゲフィチニブ感受性である。また、ゲフィチニブには T790M をはじめとするいくつかの薬剤抵抗性変異が報告されている²⁵。H1975 は EGFR の L858R 変異をドライバー変異にもつが、同時に T790M に二次変異を持ち、第一世代 EGFR チロシンキナーゼ阻害薬には完全な耐性を持つ。H1650 は EGFR の Exon19 欠損と PTEN 遺伝子の欠損を有し、部分的な耐性を示す^{26,27}。H2228 はドライバー変異として ALK-fusion を持つため、ゲフィチニブには非感受性である^{28,29}。これらの耐性獲得の初期過程については、いくつかの先行報告もあるが、その一細胞レベルでの詳細は未だ明らかではない³⁰。

本研究の目的

本研究では複数のプラットフォーム由来のデータセットの体系的に相互比較、さらにそれらを組み合わせ、それぞれの欠点を相互に相補する枠組みの構築を試みた。共発現ネットワーク解析手法を用いて、異なるデータセットを相互補完することで、十分なレベルで遺伝子発現と細胞の多様性を解析するための有力な手段となりうることを示した。本提案手法により、これまでに解析することが困難であった低発現量遺伝子の動態および希少細胞集団の詳細を明らかにすることが可能となったと考えている。また本研究では肺腺がんを対象としたが、他のがんにおける治療抵抗性の解明にも貢献することができると考えている。

*なお、本論文の主要な内容は、論文目録に記載した Kashima et al (*Scientific Reports, in press*) にて別途発表予定である。

本論文中の略語一覧

AURKA	aurora kinase A
AURKB	aurora kinase B
COSMIC	Catalogue of Somatic Mutation in Cancer
DUSP1	Dual specify phosphatase1
GO	Gene Ontology
ME	Module Eigengene
PPM	parts per million reads
RNA-seq	RNA sequencing
RPKM	reads per kilobase per million reads
scRNA-seq	Single-cell RNA sequencing
TASCTD2	Tumor Associated Calcium Signal Transducer 2
TCGA	The Cancer Genome Atlas

第一章：『一細胞 RNA-seq プラットフォームの技術評価』

第一章：『プラットフォームの比較検討・技術評価』

1.1 実験方法

細胞培養

PC9、II-18 細胞株は理研バイオセンターから取得した(catalogue number RCB4455 ,RCB2093)。H1650, H1975 と H2228 細胞株は American Type Culture Collection から取得した (catalogue numbers CRL5883, CRL5908, CRL5953)。細胞は RPMI1640 培地 (Wako, 189-02145) に 10%ウシ胎児血清(FBS)と MEM 非必須アミノ酸(catalogue number M7145, Sigma-Aldrich, St. Louis, MO)、ペニシリン/ストレプトマイシンを添加した培地で 37°C、5% CO₂条件下で培養した。

micro-chamber system による一細胞 RNA-seq

先行論文 Matsunaga et al に基づき micro-chamber system の一つである bead-seq による一細胞 RNA-seq を行なった²。bead-seq では、各反応器において細胞が溶解され、逆転写反応が行われる。その後、polyA 付加、第二鎖合成、PCR による増幅反応を経て、ライブラリーが作製される。本方法においては、細胞ごとに独立したライブラリーが調整されるという特徴がある(Sup Fig.1a)。作製したライブラリーは Hiseq2500(illumina)を用いて 50bp のシングルエンドでのシークエンスを行った。PC9 の再現実験に関しては、35bp シングルエンドでのシークエンスを行った。リボソーマル RNA 配列を除去するために、rmRNA に対して mapping を行い、マップされなかったリードを回収し、UCSC hg19 にマップした。マッピングには TopHat/Bowtie を用いた。自作した Perl スクリプトにより、RNA-seq のタグ数をカウントし、RPKM 値を算出した³¹。micro-chamber system において RPKM 値を使用したのは、本方法では全長のシークエンスを得られるためである。

micro-droplet system による一細胞 RNA-seq

micro-droplet system として Chromium Single cell 3' kit を使用し、キットの手順書に従ってライブラリーの作成を行った。Chromium では、流路のそれぞれに細胞希釈液、オリゴつきの Gel Bead、液滴精製のオイルを流し込み、プラットフォーム内で微小液滴の精製を行う。version1 での対応細胞数は、1,200-6,000 細胞程度であるが、本研究においてはターゲットを 5,000 細胞と設定した。本方法では、複数の細胞から一つのライブラリーが調整される(Sup Fig.1b)。作製されたライブラリーは、Hiseq2500 により 50bp のペアエンドでシークエンスを行った³²。得られた RNA-seq のタグは付属する Cellranger を用いてヒトゲノムにマップした。自作の Perl スクリプトを用いて、品質の低いリードと PCR sister を除去した。得られたリードを、セルバーコードに基づいて仕分けし、5,000 以上のリードを持つ細胞だけを選択、ppm 値を算出した。micro-chamber system において PPM 値を使用したのは、本方法では得られるシークエンスが 3'側に偏るためである。

scRNA-seq 解析を用いた肺腺がん細胞株の解析

5種類のヒト肺腺がん細胞株を用いて、ライブラリーの作製を行った(表1)。ゲフィチニブ(CAS 184475-35-2, Santa Cruz Biotechnology)は、最終濃度 $1\ \mu\text{M}$ で培養培地に添加し、薬剤刺激を行ない、24時間後に細胞を回収した。コントロール群には DMSO を投与した。これらのライブラリー作製は、同じ細胞株を出発材料に用いて、micro-droplet system と micro-chamber system のそれぞれの方法論で構築した(表2)。

遺伝子ごとの発現レベルの分散の比較

遺伝子ごとの発現レベルの分散の比較には、micro-chamber system より得られた PC9 細胞、コントロール群 44 細胞のデータセットを使用した。対象遺伝子は synthetic bulk における発現レベルの順位(上位 250 位、500 位、1000 位、2500 位、5000 位)を元に出した。図 3C, インセット表 1 における各遺伝子の発現レベルは synthetic bulk として算出した。図 3C 上においては、マップされたリード数を 5,000, 10,000, 50,000, 100,000, 500,000 タグと仮定した場合での、各遺伝子での発現レベルの分散を算出した。相対的な分散を示すために、図 3C, インセット表 2 においては、“相対分散”を、SD と平均 tag 数から算出した。図 3D においては、マップされたリード数が 5,000 タグだと仮定した場合におけるタグ数の頻度を算出した。

細胞株のクラスタリング

図 3E では、5つの細胞株(PC9,II-18, H1650, H1975, H2228)のデータセットを、プラットフォームごとにまとめ、クラスタリングを行なった。各細胞株の各プラットフォームでのコントロール群の細胞、micro-chamber system では 232 細胞、micro-droplet system では 22,912 細胞を使用した (Table1)。遺伝子リストとして、127SMG(significant muted genes)を使用した(Sup Table 2)³³。PCA の描画には、R パッケージの ggplot2 を使用した³⁴。

細胞周期の同定

図 3F では、細胞周期を図示した。micro-chamber system の PC9 コントロール群 44 細胞のデータセットに対して、先行論文にて細胞周期依存的に発現が変化することが報告されている 20 遺伝子の発現情報を用いて細胞がそれぞれ細胞周期のどのステージにあるかを推定した³⁵。ヒートマップでは、スケール処理を行い、各周期の遺伝子発現の平均値に基づき細胞を順番に並べた。この解析により推定された周期に基づいて、細胞を色分けし、PCA により図示した。

micro-droplet system の PC9 コントロール群の 5,166 細胞の発現量(ppm)を用い、micro-chamber system のデータセットと同様の解析を試みた。先行報告に基づく方法で使用する遺伝子を選び出し、細胞のヒートマップにおける並び替えを行なった⁴。micro-chamber system のデータセットと同様に細胞を色分けし、PCA により図示した。それぞれ、ヒートマップの描画には R パッケージの gplots を使用した。二次元に投影するための PCA の描画には R パッケージの ggplot2 を使用した^{34,36}。

PC9細胞の MAPK パスウェイ解析

図 3 G において、MAPK/ERK パスウェイに含まれる遺伝子の発現レベルを可視化するために、各プラットフォームのタグ数が第一位の細胞を選び出した。これらの細胞における各遺伝子発現レベルとしてタグ数を算出し、カラーバーにより図示した。図示には、DBTSS を使用した^{37,38}。

欠損値推定

欠損値推定をするために、micro-chamber system のコントロール群とゲフィチニブ群の細胞を合わせた。平均発現レベルが 10RPKM 以上の遺伝子を micro-chamber system から選び出した。平均 10RPKM 以上の発現レベルを示す遺伝子として、コントロール群では 4,901 遺伝子、ゲフィチニブ群では 4,845 遺伝子を用いた。発現推定はこれらの遺伝子を用いて行った。発現レベルは、対数によるスケールに変換した。対数化を行う前に 0.01 を追加した。スケールされた値を用いて LASSO による回帰分析を行なった³⁹。

欠損値推定の結果を確認するために、micro-droplet system データセットのうち欠損していない発現情報を推定結果と比較した。全体の相関をピアソンの相関を用いて算出した。

R パッケージ version 3.3.1 の “glmnet” を用いて Lasso regression を行なった³⁹。ラムダのパラメータは、“glmnet” のリストの 10 番目の値を、その他はデフォルトの値を使用した⁴⁰。

1.2 結果

各プラットフォームの技術評価

表 1 に示したヒト肺腺がん細胞株 5 種を用いて micro-chamber system である bead-seq、micro-droplet system である Chromium Single-cell 3' を用いて、それぞれライブラリーを調整し、次世代シーケンサー Illumina HiSeq2500 によりシーケンス解析した^{2,41}(表 1)。得られたシーケンスデータから、各遺伝子の発現量を算出し、それぞれのプラットフォームの技術評価を行なった。はじめに、バーチャルバルク（データセットごとの平均発現量）を計算し、バルクレベルでの再現性の評価を行った。独立した実験間で、二つのプラットフォームはそれぞれピアソン係数 $r=0.96$, $r=0.98$ の高い相関を示した(図 1 A-B)。この結果が、PC9 細胞のデータセットだけに限定したものではないことを、他の細胞のデータセットについて同様の解析を行って示すことができた(図 S2A -S2E)。ゲフィチニブ刺激有無での比較も同様にバルクレベル、バーチャルバルクレベルで行ったところ、 $r=0.91$, $r=0.94$ と $r=0.94$ 同様に高い相関を示した(図 1 C-E)。

シングルセルレベルで同様の評価を行った。それぞれのプラットフォームから、最もマッピングされたリード数が多い細胞を top1、次に多い細胞を top2 として選び出した。top1, top2 のリード数は、micro-chamber システムが top1 細胞 2,944,039 タグ、top2 細胞 2,773,525 タグであったのに対して、micro-droplet システムでは、top1 細胞 35,753 タグ、top2 細胞 32,472 タグであった。それぞれのプラットフォームにおける top1 と top2 を比較すると、その相関係数は $r=0.83$ と $r=0.67$ であった(図 1F, 1G)。これらの結果から、それぞれのプラットフォームは、それぞれに高い再現性を有することが示唆された。

表 1 : 本研究に用いた細胞株

Cell line	Response to gefitinib	Mutation
PC9	sensitive	EGFR Exon19 del
II-18	sensitive	EGFR L858R
H1650	resistance	EGFR Exon 19del, PTEN loss
H1975	resistance	EGFR L858R, T790M
H2228	resistance	EML4-ALK fusion gene

表 2: 本研究に用いた二つのデータセット

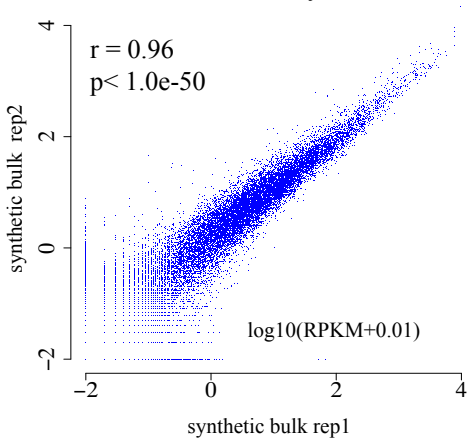
(A) micro-chamber system データセット

Cell line	Treatment	Average raw reads	Average used reads	Number of cells
PC9	DMSO	1,938,496	1,683,528	44
	gefitinib	1,728,426	1,502,134	22
II-18	DMSO	1,071,312	904,648	47
	gefitinib	1,169,322	985,941	47
H1650	DMSO	1,158,863	1,018,527	47
	gefitinib	1,283,908	1,126,651	47
H1975	DMSO	846,215	725,523	47
	gefitinib	866,011	744,728	47
H2228	DMSO	1,362,097	1,183,384	47
	gefitinib	1,269,704	1,092,511	47
PC9 replicate	DMSO	1,362,097	1,183,384	47

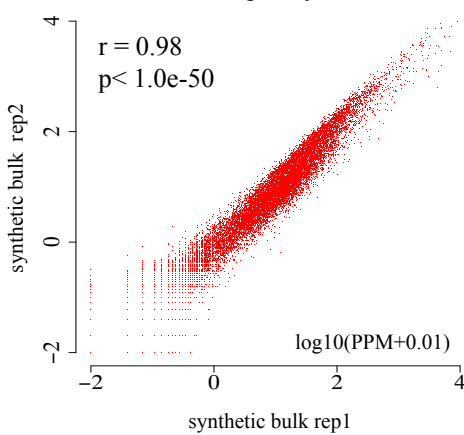
(B) micro-droplet system データセット

Cell line	Treatment	Total raw reads	Total used reads	Number of Cells with >5 k tags
PC9	DMSO	127,631,437	64,433,148	5,166
	gefitinib	170,913,768	55,473,258	4,378
II-18	DMSO	173,605,651	57,839,961	4,965
	gefitinib	164,734,523	65,419,730	5,287
H1650	DMSO	100,094,271	49,223,558	4,348
	gefitinib	111,008,829	54,577,170	5,140
H1975	DMSO	161,777,133	33,499,006	3,079
	gefitinib	173,096,952	56,825,197	4,940
H2228	DMSO	176,666,104	61,861,660	5,354
	gefitinib	149,011,834	43,740,708	5,008
H1975 replicate	DMSO	139,741,411	59,100,428	5,152

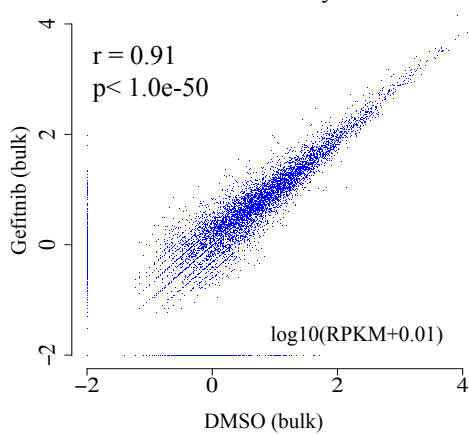
A micro-chamber system



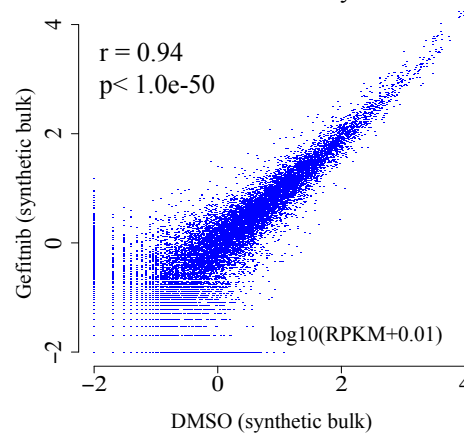
B micro-droplet system



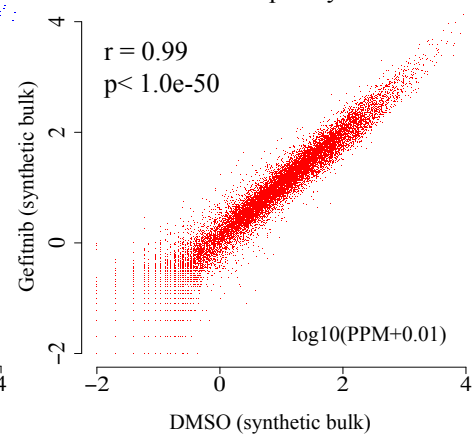
C micro-chamber system



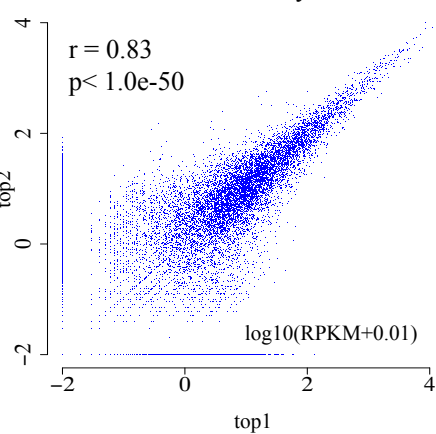
D micro-chamber system



E micro-droplet system



F micro-chamber system



G micro-droplet system

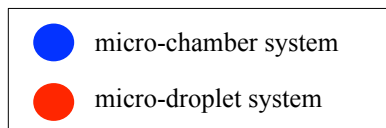
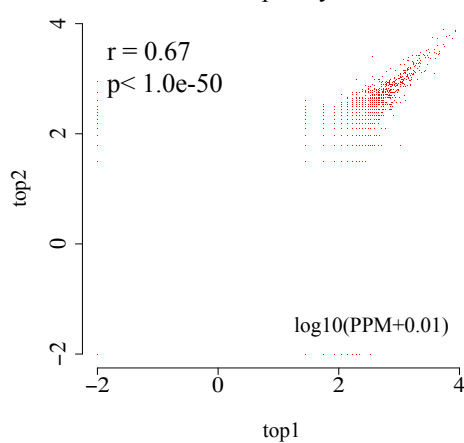


図 1: micro-chamber システムと micro-droplet システムを用いた RNA シークエンスデータセットの比較

(A, B) バーチャルバルクレベルでの実験間比較, micro-chamber system (A) と micro-droplet system (B)。 (C) micro-chamber system を用いたバルクレベルでの gefitinib に対する反応。 (D, E) バーチャルバルクレベルでの gefitinib に対する反応、micro-chamber system (D) と micro-droplet system (E)。 (F, G) top1 細胞と top2 細胞を用いた比較 micro-chamber system (F) と micro-droplet system (G)。ピアソン係数(r)は図中に記載。

各プラットフォームの特徴

二つのプラットフォーム間でデータセットの比較を行った。バーチャルバルクレベルの比較では、 $r=0.90$ と比較的高い相関が認められた($r=0.90$, 図 2A)。micro-droplet system のデータセットからランダムに 50 細胞を選び作製したバーチャルバルクのデータセットと、micro-chamber system のバーチャルバルクのデータセットの比較も併せて行った。細胞数の減少に伴う総タグ数の変化によると思われる相関の減少($r=0.83$)が認められた。しかし依然として、二つのプラットフォームには高い相関があることが示された(図 2B)。一方で、各プラットフォームから得た top1 細胞を用いた一細胞レベルのプラットフォーム間比較では相関係数は $r=0.62$ にとどまった(図 2C)。top1 細胞だけではなく、データセットにおいて二番目に多いタグ数を持つ top2 細胞、三番目に多いタグ数を持つ top3 細胞も用い、同様の解析を行った。それぞれに対して相関係数を算出したところ、top1 細胞と同様の結果が認められた($r=0.55-0.63$, 図 S3)。

観察された相関の低下の要因を明らかにするために、それぞれのデータセットの内在的な要因を調べた。各データセットの一細胞あたりの検出遺伝子数を調べた。シーケンス深度が増すにつれて、検出遺伝子数も増加することが確認された(図 2D, 左)。今回のデータセットのシーケンス深度では、micro-chamber では平均 10,661 遺伝子、micro-droplet では平均 3,096 遺伝子がそれぞれ検出された(図 2E)。また、マップされたタグ数と検出遺伝子数の関係から、micro-chamber system と micro-droplet system のデータセットは類似していることが示唆された(図 2D, 左)。

micro-chamber system の top1 細胞と top2 細胞のデータを用いて、これらの細胞が micro-droplet system の top1,2 細胞のタグ数に相当する 30,000 タグを持つと仮定した場合のスキッタープロットを描画した。この仮想データセットにおいては、二つの細胞の相関係数は $r=0.63$ となった(図 2F)。micro-chamber のデータセットの情報学的疑似的希釈により、micro-droplet のデータセットを用いて同様の解析を行った結果と類似する結果が得られることが示された(図 1D, $r=0.67$, 図 2E, $r=0.61$)。遺伝子に対応するタグ数の順番に並べかえ、累積タグ数を検証した。二つのプラットフォームでは同様の曲線が観察された(図 2G)。検出遺伝子数は、シーケンスの深さに依存し、micro-chamber system ではほぼプラトーに達しているのに対して、(傾き 0.0007; $r=0.47$, $p=0.0013$, 図 2D, 右)、micro-droplet システムのデータセットでは飽和に達していなかった。これらの結果は、二つのプラットフォームから得られるデータセットは本質的には同一であることを示唆していると考えた。当初観測されたデータ上の差異は、シーケンスの深さが異なるために得られたものであると考えている。すなわち、バーチャルバルクレベルでの比較においては、シーケンスの深さが十分であるために高いレベル類似性を示すが、一細胞レベルではその類似性は低下すると考えられた。

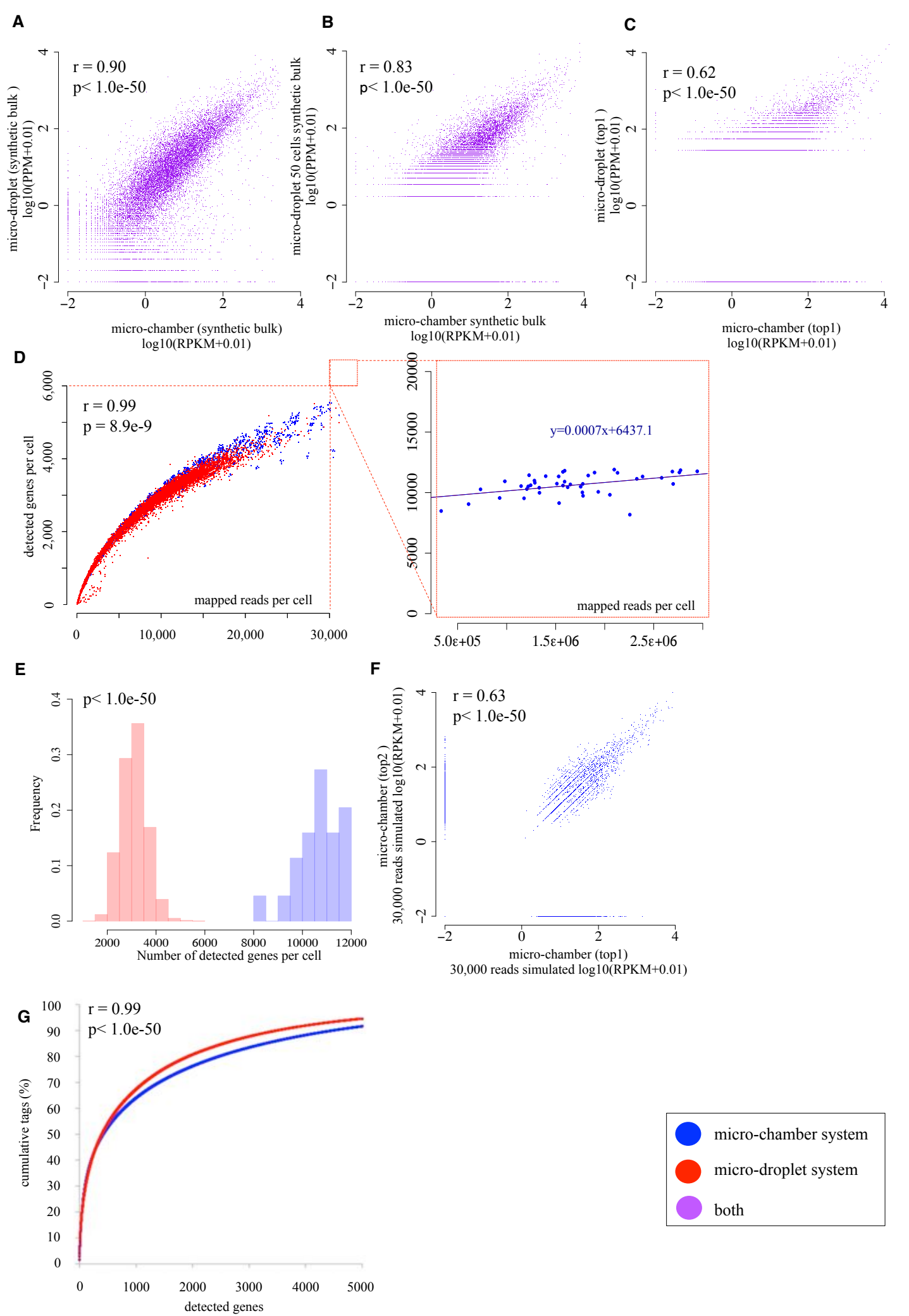


図2：細胞と遺伝子ごとの発現レベルの多様性

(A-C)micro-chamber system と micro-droplet の発現レベルの相関。バルクレベルでの発現の相関(A)と micro-droplet の細胞数を 50 細胞とした場合の相関(B)、両プラットフォームで一細胞レベルでの相関(C)。*ピアソン係数と p-value は図中に記載。(D) マッピングされたリード数と検出された遺伝子数の相関。(E)検出された遺伝子数(x 軸)とその相対頻度(y 軸)。*ピアソン係数とその p-value は図中に記載。(F)30,000tag/cell での二つのプラットフォームでの比較。*ピアソン係数とその p-value は図中に記載。(G)遺伝子数(x-軸)と積算頻度(y-軸)の相関。

各遺伝子でのプラットフォームの比較

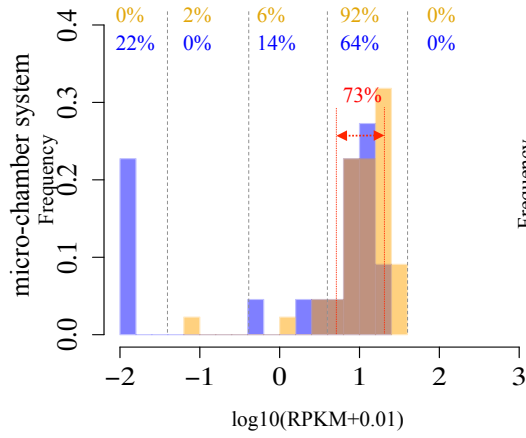
個別の遺伝子に着目した場合、異なるプラットフォーム間で遺伝子発現量とゲフィチニブ刺激に対する発現量変化がどのように観測されるかを調べた(図 3A、図 3B)。micro-chamber system では、EGFR 遺伝子と MYC 遺伝子の発現はコントロール群では、全ての細胞で発現が観測された。MYC の発現については、ゲフィチニブ刺激により遺伝子発現が観測されなくなる細胞が一定の割合で存在した。これらの細胞では EGFR シグナル経路が阻害された結果、MYC の発現が完全に抑制されたのかもしれない。一方、micro-droplet system のプラットフォームにおいては、非刺激群においても、EGFR, MYC 遺伝子の発現が確認されない細胞が認められた。前述の結果から、scRNA-seq のタグが、MYC, EGFR の発現を捉えるには不十分であった可能性が考えられた。また、発現が観察された細胞でも、検知されたタグ数は少数にとどまった(図 3A)。micro-droplet system において、これら機能的に最も重要な遺伝子の細胞多様性を発現がタグ数の制約により観測できないとすれば、大きな問題であると考えた。

発現レベルが EGFR 遺伝子、MYC 遺伝子とは異なる遺伝子についても同様の検証を行った。先行研究において、がんとの関連が示唆される TACSTD2 遺伝子と AURKB 遺伝子についても同様に一細胞レベルでの発現を確認した⁴²⁻⁴⁴。TACSTD2 遺伝子はゲフィチニブ刺激の有無で、micro-droplet system においても発現レベルが 439ppm 対 1,984ppm と大きく変化した(図 3B, 左下)。TACSTD2 遺伝子程度の発現量を有する遺伝子については、micro-droplet system においても発現の情報や、細胞ごとの差、薬剤刺激による差が観測できることが示唆された。AURKB 遺伝子は、micro-chamber system では、細胞により多様な発現パターンが観測された(図 3B, 右上)。これは該当する細胞が属する細胞周期によるものであると考えられ、既知の知見と一致している。一方で、micro-droplet system では、このような特徴的な発現パターンを捉えることはできなかった(図 3B, 右下)。これは、AURKB 遺伝子の発現レベルに対して、micro-droplet system での細胞あたりのタグ数が不十分であったことが一つの要因であると考えられた。ゲフィチニブに対する反応として起きるバルクでの遺伝子発現のレベルの変化に対する測定精度は、その遺伝子の発現レベルに依存する可能性があることが示唆された(表 3)。

A**EGFR**

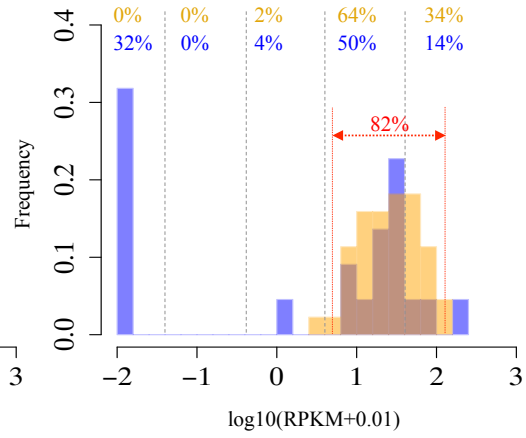
bulk (RPKM)	12.6	10.2
synthetic bulk (RPKM)	13.9	9.7

p = 0.0014

**MYC**

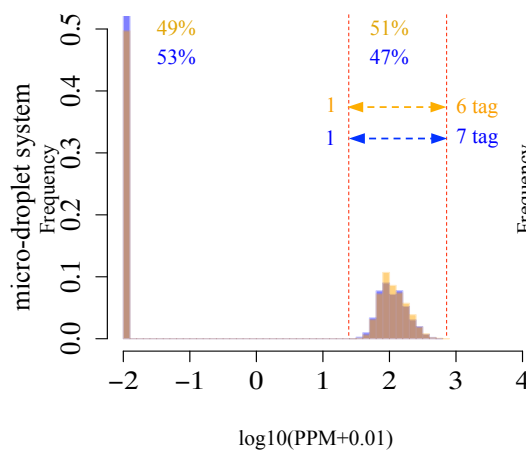
bulk (RPKM)	32.9	23.8
synthetic bulk (RPKM)	43.7	28.3

p = 0.015

**EGFR**

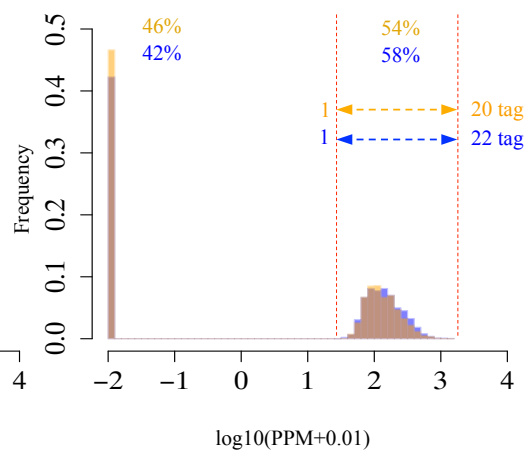
synthetic bulk (PPM)	64.5	60
----------------------	------	----

p = 0.00053

**MYC**

synthetic bulk (PPM)	83.6	97.4
----------------------	------	------

p = 1.4e-07

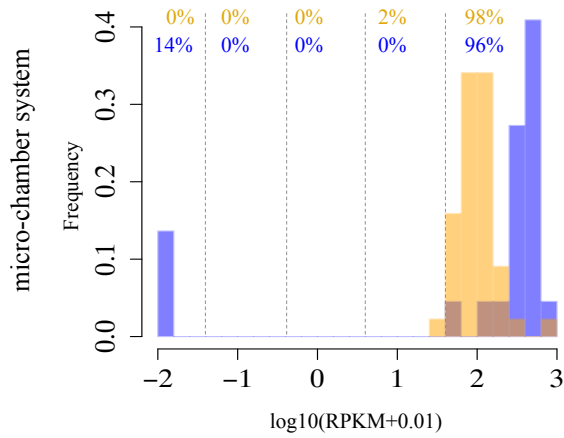


B

TACSTD2

bulk (RPKM)	96.1	469
synthetic bulk (RPKM)	118	425

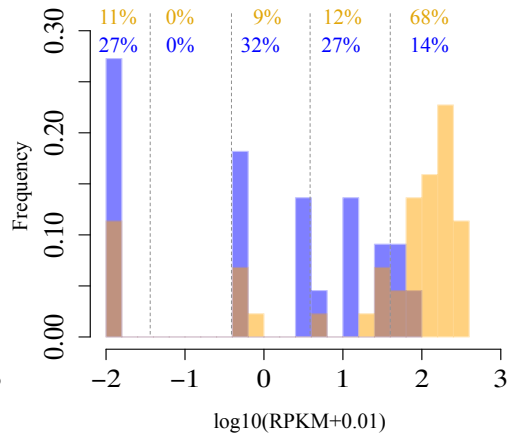
p = 0.00028



AURKB

bulk (RPKM)	89.4	30.4
synthetic bulk (RPKM)	126	17

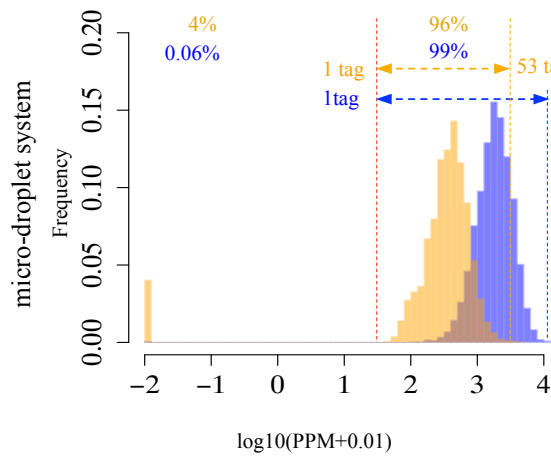
p = 1.5e-05



TACSTD2

synthetic bulk (PPM)	439	1984
----------------------	-----	------

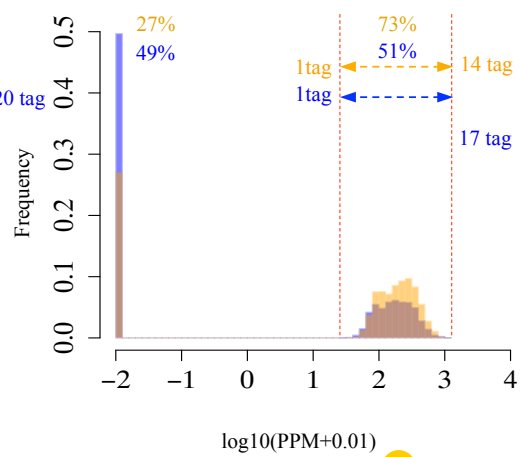
p < 1.0e-50



AURKB

synthetic bulk (PPM)	182.9	124
----------------------	-------	-----

p < 1.0e-50



● DMSO

● Gefitinib

図3：個別の遺伝子に関する各プラットフォームでの発現情報

(A,B) がん関連遺伝子の EGFR、MYC、TACSTD2 と AURKB 遺伝子の個々の細胞での発現レベルとその分布。発現レベル(*x-axis*)と発生頻度(*y-axis*)。 micro-chamber system:上と micro-droplet system:下。

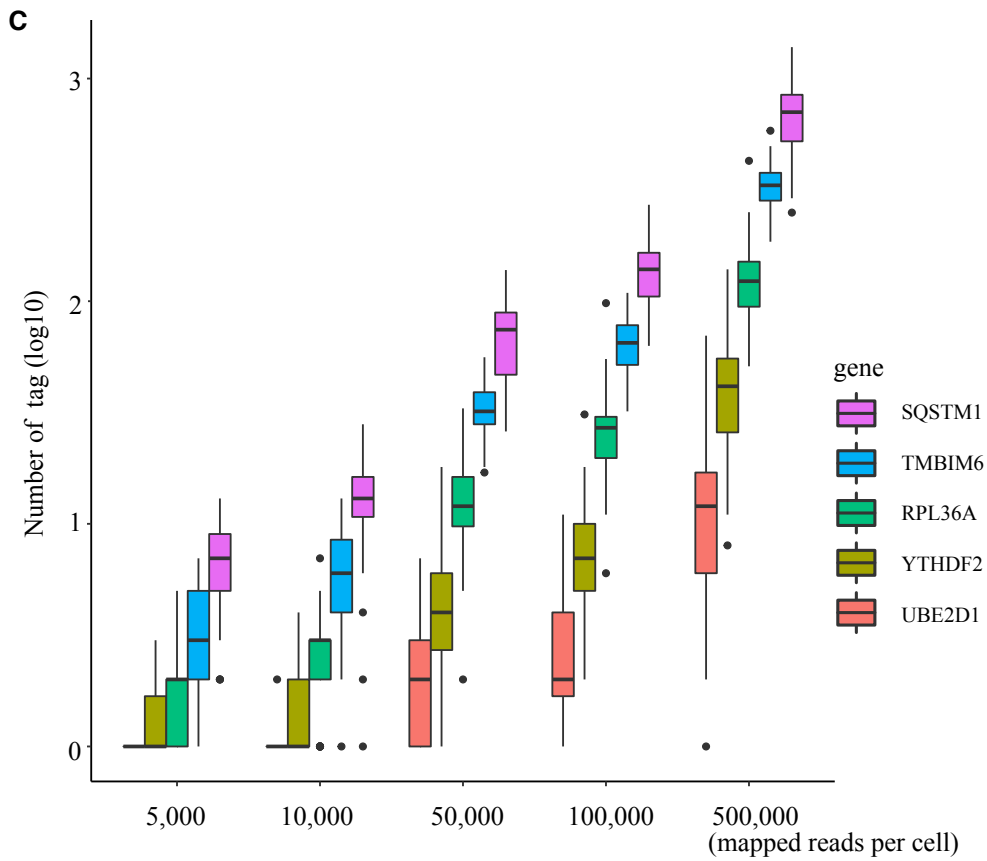
表 3 : 各遺伝子の発現レベルの比較

		EGFR	MYC	TACSTD2	*AURKB	
micro-chamber	バルク (rpkm)	treated	10.18	12.6	468.68	30.35
		untreated	12.6	32.86	96.1	89.43
		fold	1.23	2.61	0.21	2.95
		(untreated/treated)				
micro-chamber	バーチャルバルク (rpkm)	treated	9.69	28.31	424.69	17.2
		untreated	13.87	34.68	118.34	126.39
		fold	1.43	1.23	0.28	7.35
		(untreated/treated)				
micro-droplet	バーチャルバルク (ppm)	treated	59.9	97.4	1985	124.25
		untreated	64.45	83.6	439.47	182.9
		fold	1.08	0.86	0.22	1.47
		(untreated/treated)				

* 最もプラットフォーム間の差が大きい遺伝子。

シーケンス深度と発現レベルの関係

次に、各プラットフォームにおいて、検出可能である遺伝子発現変化と細胞間多様性の詳細に着目した。シーケンスの深度が、各遺伝子の発現レベルとその分散にどのような影響を与えるかを検証した。micro-chamber system の PC9 コントロール群の 44 の細胞のデータを使用し、細胞あたりのマップされたリードが 5,000, 10,000, 50,000, 100,000, そして 500,000 である場合を想定した。各遺伝子のタグ数を算出し、どの程度の分散を検出できているかを確認した(図 3C)。分散の評価のために、各遺伝子の平均タグ数で、標準偏差を除いた相対分散の値もそれぞれ算出した(図 3C, インセット表)。発現レベルが低い遺伝子においては、リード数が少なくなるにつれてタグ数が 0 となる細胞数が増加した。そのため、相対分散は疑似的に大きくなる傾向が認められた。また、micro-droplet system で細胞の cut-off に利用した 5,000 タグレベルでは、各遺伝子のタグ数はどのようになるのかを図示した(図 3D)。上位 5 0 0 0 位である UBE2D1 遺伝子においては、このレベルでは全体の 91%に当たる 40 細胞が 0 タグを示すことが示された。これに対し、SQSTM1 や TMBIM6 といった高発現レベルの遺伝子においては、低いリード数においても相対分散は維持されていた(図 3C, 組み込み表 1,2)。



Gene	Rank	synthetic bulk (rpkm)	number of cells with 0 tag per cell									
			5,000 reads		10,000 reads		50,000 reads		100,000 reads		500,000 reads	
SQSTM1	250	417.7	1	(2)	0	(0)	0	(0)	0	(0)	0	(0)
TMBIM6	500	193.2	3	(7)	1	(2)	0	(0)	0	(0)	0	(0)
RPL36A	1000	91.8	15	(34)	3	(7)	0	(0)	0	(0)	0	(0)
YTHDF2	2500	27.1	30	(68)	16	(36)	0	(0)	1	(2)	0	(0)
UBE2D1	5000	8.63	40	(91)	37	(84)	20	(45)	12	(27)	5	(11)

	relative divergence				
	5,000 reads	10,000 reads	50,000 reads	100,000 reads	500,000 reads
SQSTM1	0.45	0.44	0.37	0.34	0.34
TMBIM6	0.61	0.51	0.27	0.28	0.25
RPL36A	0.93	0.58	0.45	0.52	0.47
YTHDF2	1.74	1.02	0.67	0.68	0.59
UBE2D1	3.20	2.45	1.30	1.07	1.00

D

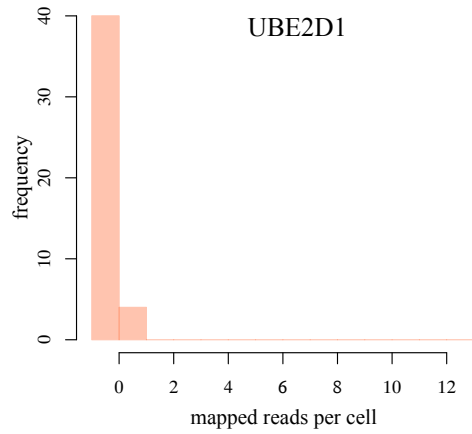
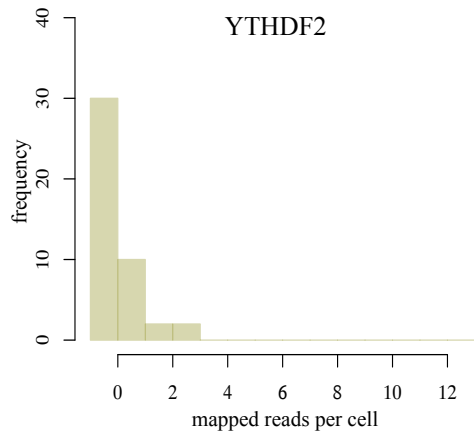
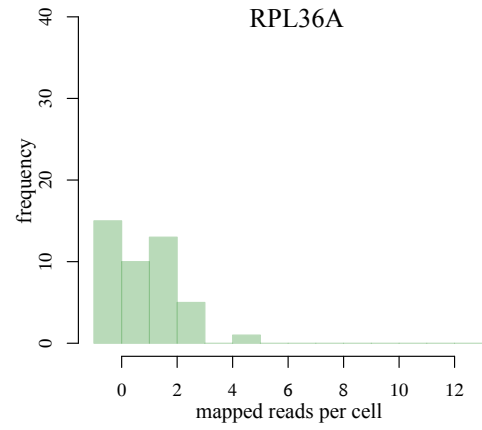
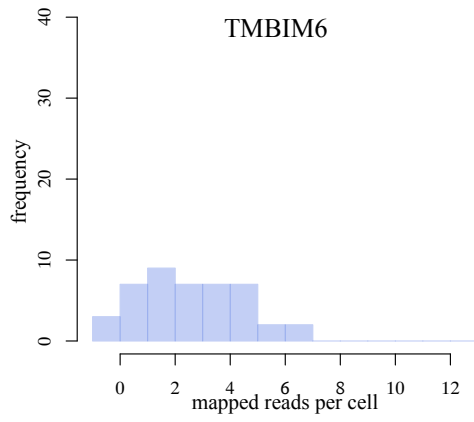
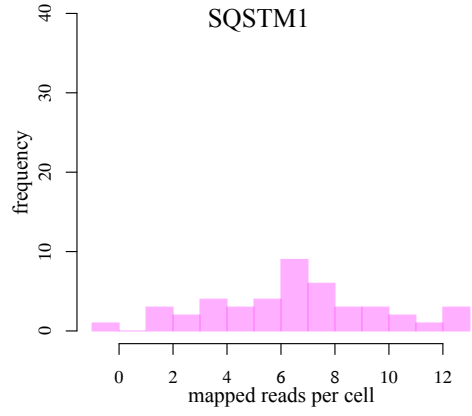
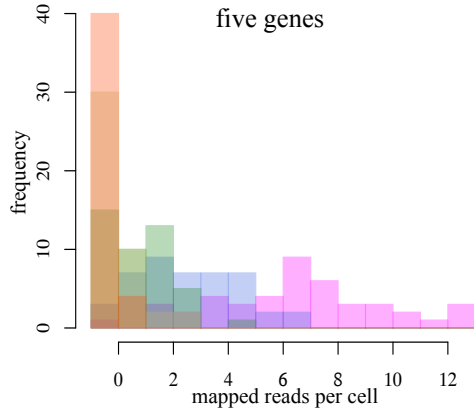


図3：個別の遺伝子に関する各プラットフォームでの発現情報

(C)各遺伝子のマップされたリードにおける分散の変遷。細胞ごとに 5000, 10000, 50000, 100000, 500000tag(x -axis)をそれぞれ得られたと仮定した場合に得られる各遺伝子の絶対タグ数(y -axis)。各遺伝子のバルクレベルでの発現順位と 0tag となる細胞数、()内の数字は割合(%)を示す(インセットテーブル、上)。各遺伝子の relative divergence(インセットテーブル、下)。(D)各遺伝子の tag 数(x -axis)と 頻度(y -axis)(右)。

異なる scRNA-seq のデータセットにより示唆される細胞の状態

これらのデータセットには、どのように生物学的情報が反映されているのかを検証した。はじめに、各プラットフォームの5種類の細胞株のデータセットを、遺伝子発現プロファイリングに基づいてクラスタリングした際に、どのような差が現れるのかを検証した。がんにおいて重要だとされている遺伝子リストを利用し、それぞれのプラットフォームごとにクラスタリングを行なった(Sup Table 2)。PCA では micro-chamber のデータセットと micro-droplet system のデータセットのどちらにおいても、細胞株ごとにまとまってクラスタリングがされることが明らかにされた。しかし、micro-droplet system のデータセットにおいては、5つの細胞株は非常に近接した状態で示された。これは細胞株を特徴付けている遺伝子の発現を micro-droplet system では検出しきれていないことが原因であると考えられる (図 3E)。

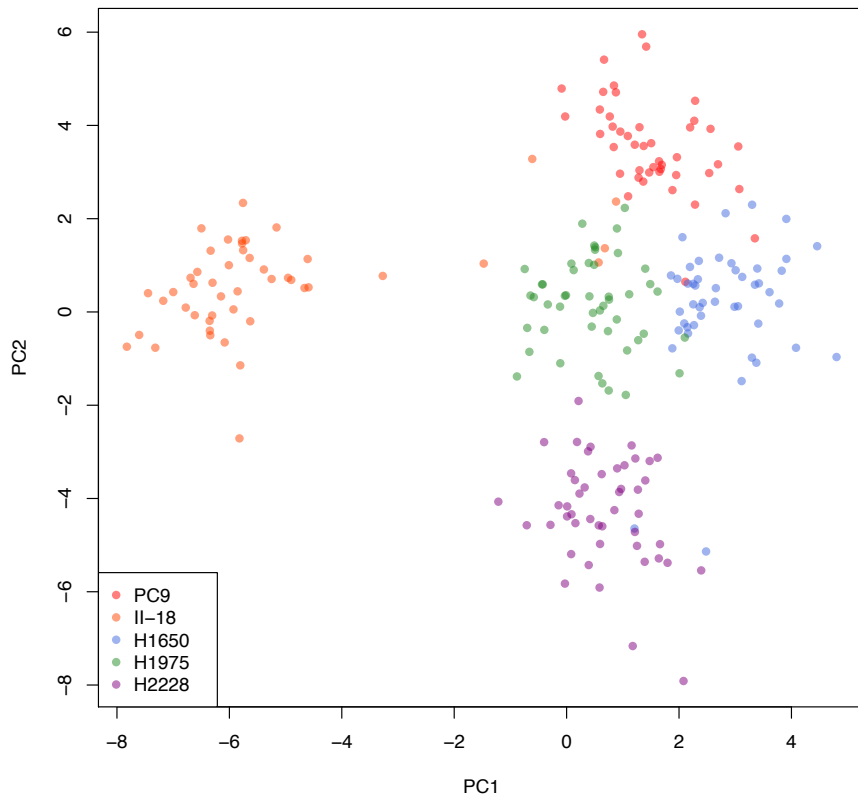
次に、細胞周期依存性が報告されている遺伝子に注目して、その発現量を解析し、細胞周期を推定した(Sup Table S2A)。一般的に細胞周期依存的である遺伝子は、高発現であることが知られている。図 3F に示されるように、micro-chamber system のデータセットからは、先行論文で細胞周期依存的に変化することが知られている 20 遺伝子の発現情報から、ほとんどの細胞について、その細胞周期状態を推定することが可能であった⁴⁵(図 3F, 左上)。一方で、micro-droplet system データセットからでは、同じ条件で同様の結果を得ることは困難であった(図 S3)。micro-droplet system データセットについて、先行論文に従い、使用する遺伝子数を増やし、先行研究で提案される計算機手法を実践したところ、各細胞の細胞周期を推定するは可能であった⁴¹。しかし、micro-droplet データセットにおいては細胞周期の同定は、micro-chamber system データセットに比して困難であった(図 3F, 下)。

最後に、特定のパスウェイでの遺伝子発現がどのように示されるのかを一細胞レベルで検証した。各プラットフォームの top1 細胞の遺伝子発現(タグ数)のデータを MAPK/ERK パスウェイの図上に反映させた(図 3G)。micro-chamber system では top1 の細胞において、それぞれに妥当と思われる発現レベルが観測されていた(図 3G, 上)。一方、micro-droplet system においては、top1 細胞においても該当遺伝子に対応するシーケンスタグが0タグとなるものが多く、正確な遺伝子発現情報を反映してない可能性が示唆された (図 3G, 下)。

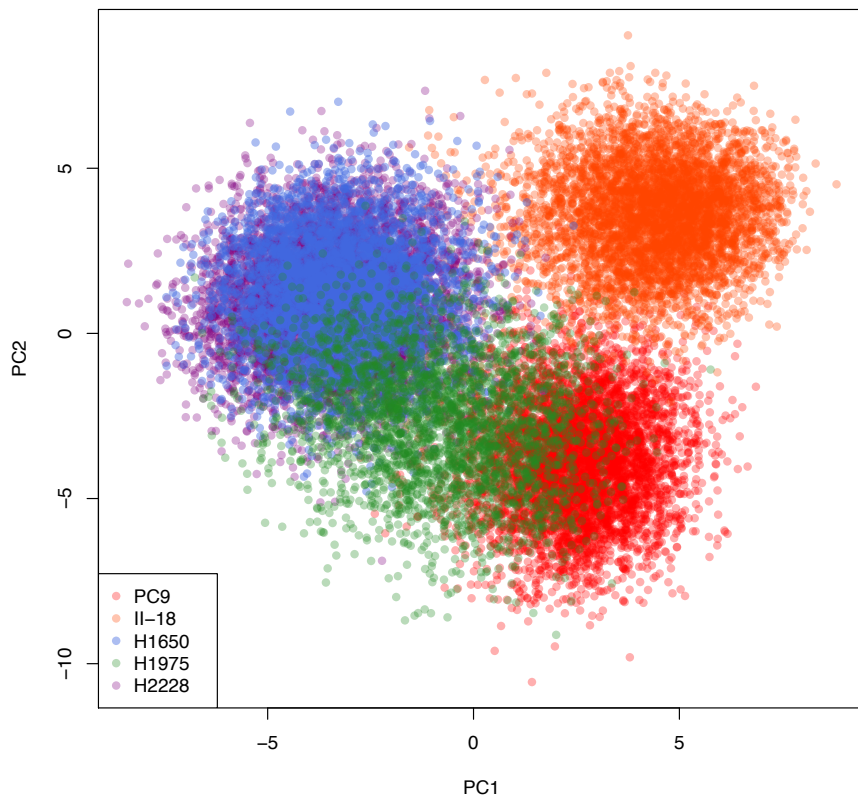
以上の結果から、micro-droplet system では micro-chamber system と比較して多くの細胞の情報が得られる一方で、一細胞レベルでは、必ずしも十分な発現情報が得られない可能性が示唆された。

E

micro-chamber system

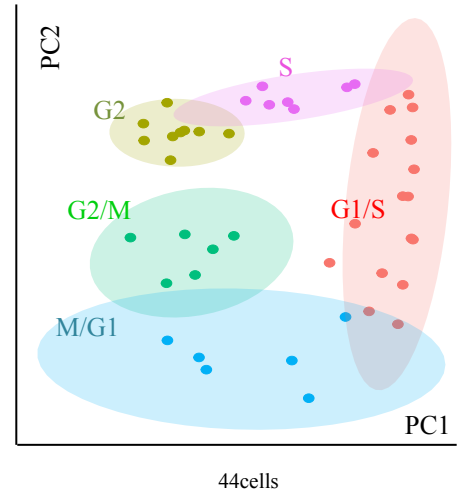
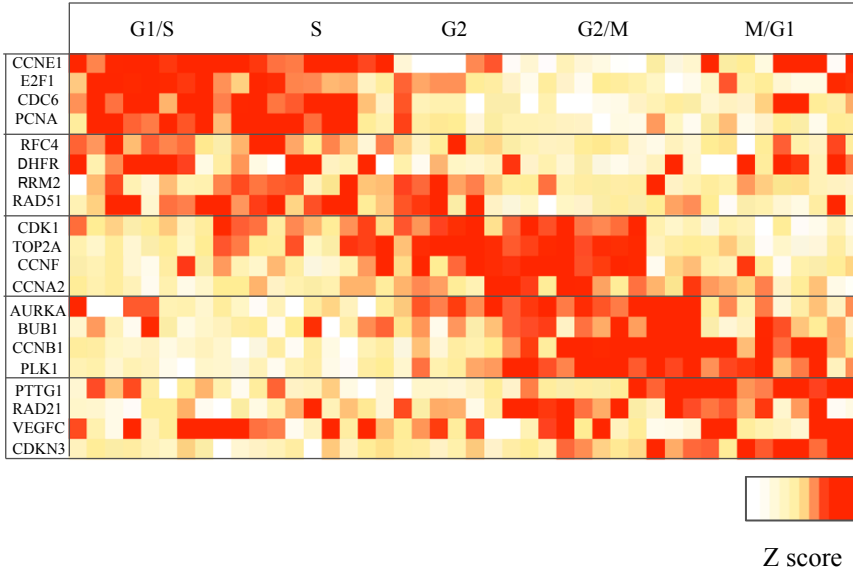


micro-droplet system

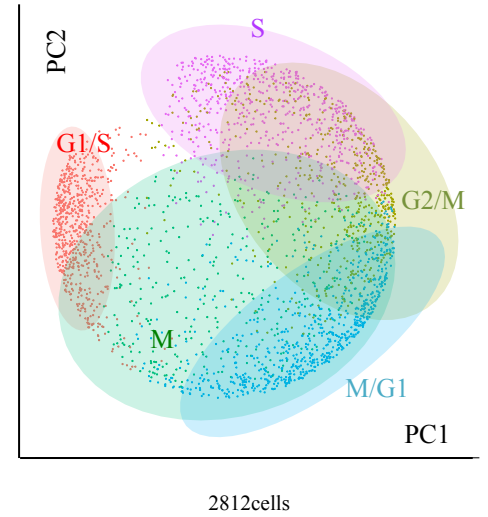
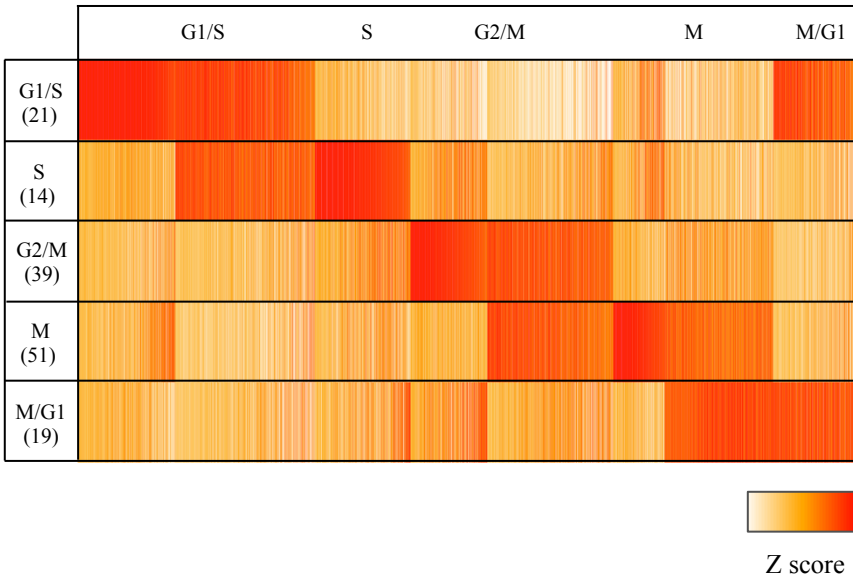


F

micro-chamber system

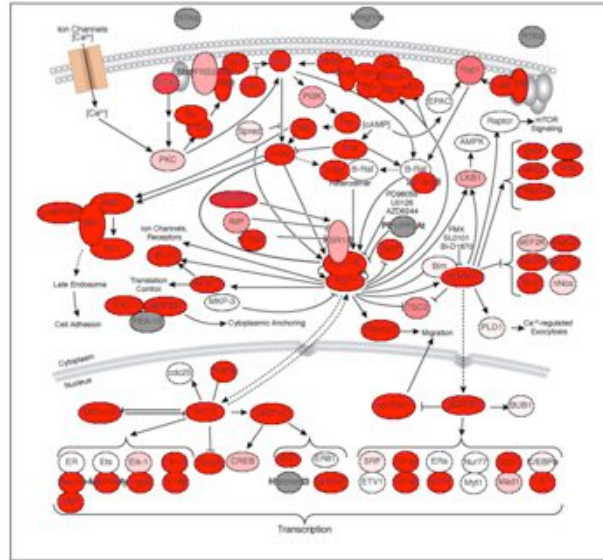



micro-droplet system



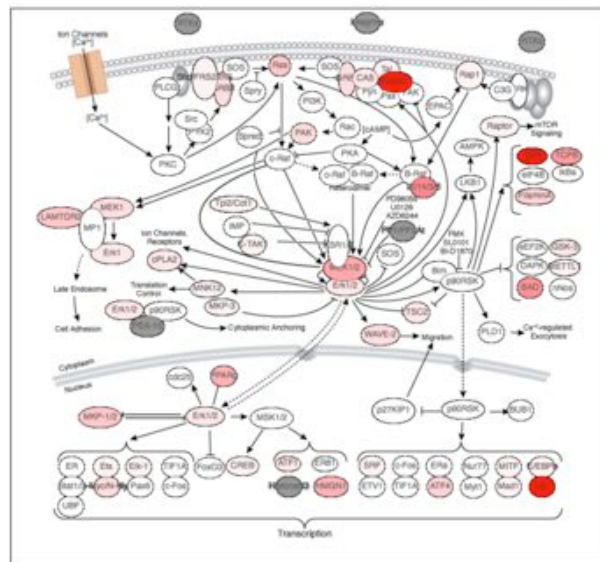
G


micro-chamber system



Coloring density: 0  25

micro-droplet system



Coloring density: 0  25

*p-value = 9.4e-23

図3：個別の遺伝子に関する各プラットフォームでの発現情報

(E) 肺腺がん細胞株のPCA。127遺伝子の発現情報に基づき、micro-chamber system(上)と micro-droplet system(下)からの細胞をそれぞれPCAによりクラスタリングした。

(F) PC9細胞の細胞周期の推定。ヒートマップとPCA。micro-chamber system:上と micro-droplet system:下。

(G) 第一位細胞を用いたMAPK/ERKパスウェイの可視化。micro-chamber system:上と micro-droplet system:下。

micro-chamber system のデータセットによる micro-droplet system の欠損値推定

micro-droplet system において不正確である可能性がある遺伝子発現情報について、その補完を試みた。補完のための方法として、micro-chamber system のデータセットを用いて、micro-droplet system での遺伝子レベル欠損値推定を行った。表 1A に示される micro-chamber system によるデータを全ての細胞について結合した。平均遺伝子発現レベルが 10rpkm 以上である遺伝子を抽出した。選ばれた遺伝子の発現情報を micro-droplet のデータセットに対する予測因子として使用した。統計学的推定には、遺伝子の linear combination を使用した。予測因子として使用する遺伝子は、LASSO により変数選択することで overfitting を回避した(詳細は材料・方法を参照)。パスウェイに含まれるすべての遺伝子について、他の遺伝子発現との Pearson の相関係数を考慮することで推定した。いくつかの遺伝子に関しては、ある程度の推定が可能であり、421 遺伝子に関して、推定値と実測値の相関が $r > 0.5$ 以上となった(図 4A)。また、393 遺伝子に関しては、遺伝子発現の推定値の 93.3% が正解に比して二倍以内に収まった(図 4B)。データを精査したところ、予測の精確さは、予測したい遺伝子と予測因子の発現レベルに依存することが示唆された (図 4C)。比較的発現レベルの高い遺伝子を推定した場合、例えば micro-chamber でのバルクでの発現量が 7.61 rpkm である PTGDR2 の推定精度はほとんどの細胞において高くなった (図 4D 左)。これに対して、PTGDR2 よりも低発現である遺伝子、すなわち micro-chamber system のバルクでの発現量が 5.7 rpkm である ATF5 のような遺伝子、あるいは micro-chamber system のバルクで 1.47 rpkm である FGF2 のような遺伝子においては、精確に推定できる細胞数は減少した(図 4D 中央、右)。

最後に、MAPK/ERK パスウェイ遺伝子全体について、micro-droplet system データセットでの遺伝子発現所法の推定による補完を試みた。予測因子が得られた 7 つの遺伝子については $r \geq 0.39$ の精度で推定をすることができた(表 4)。一方でその他の多くの遺伝子においては、予測因子の欠損値が多いために、その推定は困難であった(図 S4)。

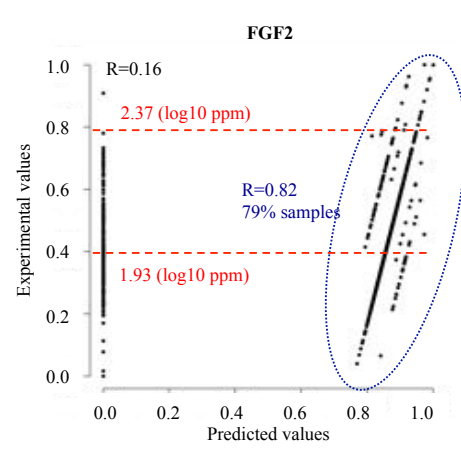
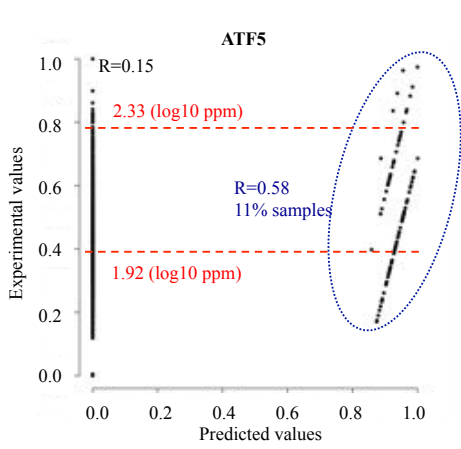
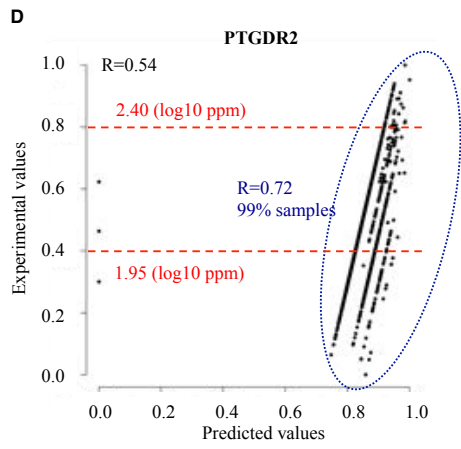
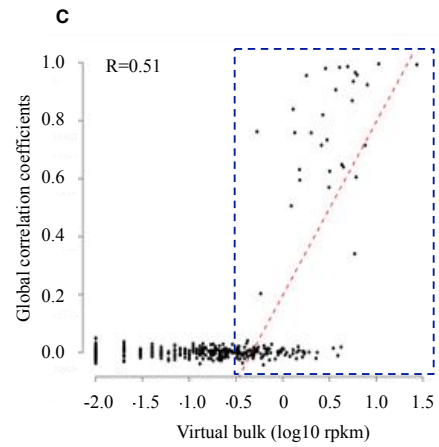
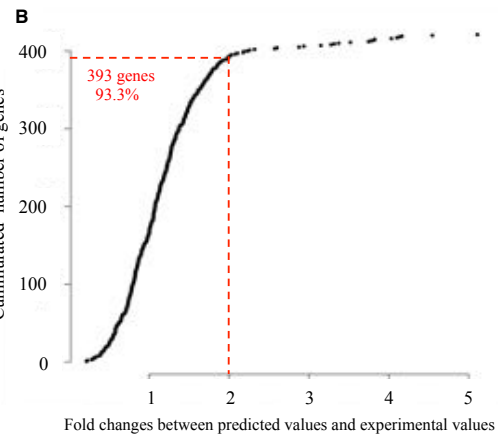
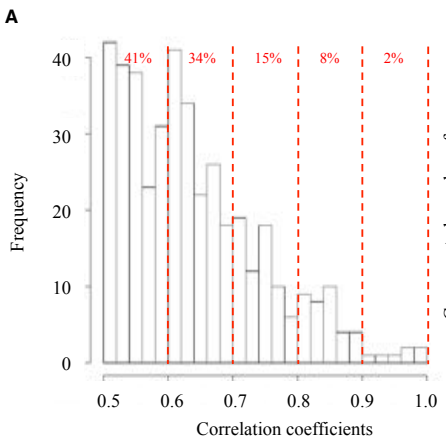


図 4: micro-chamber system データセットによる micro-droplet system データセットの推定

(A) 予測精度の分布。相関係数(x -axis)と遺伝子数(y -axis)。各相関係数における割合(赤)を図中に記載。
(B) 推定値と実測値間の(x -axis) と積算遺伝子数(y -axis)。 (C) 欠損値推定の検証。遺伝子発現(x -axis)と相関係数(y -axis)の関係。 $R > 0.5$ の遺伝子 and validated samples > 5 are shown (421 genes)。バルク発現レベル > -0.5 (\log_{10} rpkm)である遺伝子に対しての近似直線では、 $r = 0.74$ の相関が得られた (D) 3 遺伝子における標準化された推定値(x -axis)と標準化された実測値(y -axis) の相関。比較のため、異なる値を 0-1 に標準化。全ての細胞に対するピアソン相関係数(黒)と図中の囲みの中の細胞のみに対してのピアソン相関係数(青) を図中に示した。図中に記載された実測値はコントロール群のデータセット由来である。標準化されたスケールにおいて 0.8 と 0.4 に相当する実測値 (\log_{10} PPM) は図中に記載(赤)。

表4：MAPK パスウェイの遺伝子の発現レベルの推定

MAPK gene	Predictive precisions	Partially predictive precisions	Proportion of samples
PAK6	-0.12	0.7	30%
SPRED2	-0.16	0.69	3%
MKNK2	-0.03	0.59	16%
MAP2K1	0.03	0.49	3%
EEF2K	-0.22	0.4	53%
ETS1	-0.19	0.39	42%
IRS1	0.06	0.47	16%

1.3 考察

本研究の前半では、各プラットフォームから 536 細胞 (micro-chamber system) と 54,631 細胞 (micro-droplet system) について一細胞 RNA-seq データセットを獲得し、その評価を行なった。micro-chamber system では、個々の細胞において十分な精度での遺伝子発現計測が可能、すなわちトランスクリプトームに対するカバー率は優れていた。同 system では一度に解析できる細胞数が限られるため細胞集団に対する細胞カバー率では劣っていた。一方で micro-droplet system においては、トランスクリプトームカバー率が劣り、細胞カバー率に優れていた。特に、micro-droplet system においては、個々の遺伝子発現計測精度に疑問が残った。そこで、micro-droplet system データについて個別の遺伝子の発現データの統計推定が可能であるかの検討を行った。LASSO による解析の結果、高発現遺伝子について一定の推定精度が得られたものの、低発現遺伝子については依然として、その推定は困難であった。

以上の結果から、個々のデータセットはそれぞれに固有の利点が存在するものの、単独では十分なトランスクリプトームおよび細胞カバー率の両方の観点で、十分に精度での高い scRNA-Seq 解析を行うことは困難であった。現状の scRNA-seq プラットフォームを用いて、より詳細な解析を行うには、サンプルにより方法を選択する必要性や、情報学的あるいは実験学的な手法により、不足している情報を補う必要があると考えられる。

サンプルによる方法の選択としては、ターゲットとする細胞の発生頻度が一つの指標になると考えられる。先行報告などから発生頻度が高いことが保証されている細胞の解析を行う場合には、トランスクリプトームのカバー率が高い micro-chamber system のみによる解析も有効な手段となりうると考えられる。

実験側からのアプローチとしては、細胞数の調整が一つ的手段として考えられる。micro-droplet system では、設定可能な細胞数が 1,000 細胞から 5,000 細胞である。micro-droplet system では、ライブラリーの調整がまとめて行われるため、リードを複数の細胞が奪い合う状態になる。解析の結果から、二つのプラットフォームは、本質的には同等であると示されていることから、ターゲットとする細胞数を調整することで、一細胞あたりのタグ数を増やし、より多くの遺伝子発現プロファイルを得られる可能性もあると考えられる。これに対して、micro-chamber system では、一細胞化の過程の人的コストや機器の制約から、細胞数を増加させることは困難であると考えられる。

第二章 『データセットの相互補完と生物学的意義の検証』

第二章：『データセットの相互補完と生物学的意義の検証』

2.1 実験方法

モジュールに基づく一細胞解析

R パッケージの“WGCNA(Weight co-related gene network analysis)”を使用し、共発現ネットワークモジュールの構築を行なった。micro-chamber データセットの PC9 細胞、コントロール群(44 細胞)と DMSO 群(22 細胞)、合計 66 細胞から、5RPKM 以上の発現量を持つ遺伝子が 5000 以下である細胞を除去した。得られた 61 細胞の遺伝子発現データを用いて、71 のモジュールを作成した。それぞれのモジュールについて ME 値(module eigengene)を算出した。モジュールごとの特徴を評価するために、モジュールに含まれる遺伝子について GO 解析を行なった。同様に、他の四種の細胞株においても WGCNA による共発現ネットワークモジュール解析を行った。

生存曲線による解析

TCGA のデータセットを解析するために、TCGA 肺腺がん(TCGA lung adenocarcinoma, TCGA-LUAD)の RNA-seq v2 データと臨床情報を NCI Genomic Data Common から TCGA-Assembler v2.0.1(2017/03/09 にダウンロード)入手した⁴⁶。これらのデータのうち 506 症例に関しては RNA-seq のデータセットと臨床情報を同時に得ることができた。

RNA-seq のデータセットは、assayPlatform が “gene.normalized_RNAseq”であり、cancerType が “LUAD”であるものを選択した。発現レベルは $\log_2(\text{expression} + 1)$ へと変換した。それぞれの遺伝子について、その平均発現量+ 0.5 s.d.以上にあるものを高発現、 average - 0.5 s.d 以下にあるものを低発現と定義した。

臨床情報に関しては、cancerType が “LUAD”であるものについて、clinical patient file と follow-up file を使用した。カプランマイヤー解析は R の “survival package”を用いて行なった。

2.2 結果

モジュール発現レベルでの一細胞解析

以降の解析では、ゲフィチニブに対して感受性であるとされている PC9 と II-18 の 2 つの細胞株に特に着目した (表 1)。緒言にも述べたように、耐性化に関与する変化はトランスクリプトームから始まるという先行報告もあることから、ゲフィチニブに対して耐性を獲得している過程にある細胞を、ほかの細胞とは異なる挙動を示す”外れ細胞”として、検出しようと試みた。”外れ細胞”の検出には、単独の遺伝子だけの評価では困難であると考えたため、トランスクリプトームモジュールを利用した。

はじめに、相互に発現相関する遺伝子から構成されているトランスクリプトームモジュールを作成した。個々の遺伝子発現の情報が正確ではない場合にも、特定のトランスクリプトームモジュールのレベルにおいて、その発現情報を解析することができるかどうかを検討した。本解析では、トランスクリプトームモジュールは共発現ネットワークモジュールとして作成した。PC9 細胞の場合、micro-chamber system での 66 細胞の scRNA-seq データからモジュールを構築した。まず、少なくとも一つ以上の細胞で 5rpkm 以上の発現を示す 13,619 遺伝子を選択した。その発現データをインプットとして、weighted gene co-expression network analysis (WGCNA; 詳細は材料・方法を参照)を用いてモジュール構築を行った。図 S5A に示すような 71 のモジュールが構築された。それぞれのモジュールに関して、固有値 ME を算出した。先行研究にしたがって、ME をそのモジュールの活性を示す指標とした⁴⁷。次に、モジュールの特徴を明らかにするために GO 解析を行った。特定の生物学的機能を反映するモジュールが同定された。

これらのモジュールの中から”lightsteelblue1”と呼ばれる 38 遺伝子からなるモジュールに着目した(表 S4A)。このモジュールに含まれる遺伝子は細胞周期の G2/M あるいは細胞分裂に関与するものがほとんどであった(表 S4B)。このモジュールに含まれる遺伝子を利用して、61 細胞に対するクラスター解析を行った(図 5A)。作成されたヒートマップから、s_062 という特定の細胞に着目した。この細胞は他のゲフィチニブ処理細胞から、このモジュールについての ME 値が外れていた(図 5A、赤枠)。このモジュールに関しては、同細胞はむしろコントロール群に類似していた。実際、他のゲフィチニブ刺激群の細胞において、同モジュールの ME 値は負の値を示すが、この細胞ではコントロール群と同様に ME 値は正の値を示していた。特に同細胞では”lightsteelblue1”モジュールのメンバーの遺伝子である AURKA の遺伝子発現が亢進していることが確認された(図 5B)。AURKA 遺伝子は、セリン/スレオニンキナーゼの一つで、p53 シグナリングパスウェイの阻害により腫瘍産生や化学療法耐性に大きな役割を果たしていることが知られている^{48,49}。興味深いことにいくつかの非小細胞がんにおいて、AURKA の発現亢進はがん治療抵抗性に関与することが先行論文で示唆されている。ゲフィチニブは細胞周期の遅延を誘導し、G0/G1 停止と G2/M ブロックを引き起こすことが知られている⁵⁰。これらの先行報告に照らして、s_062 細胞においては、ゲフィチニブの効

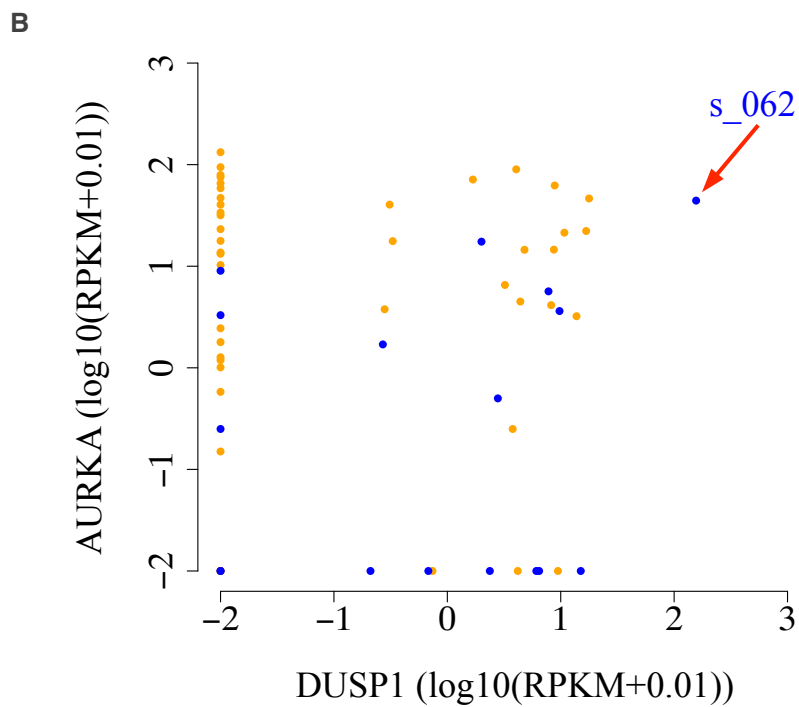
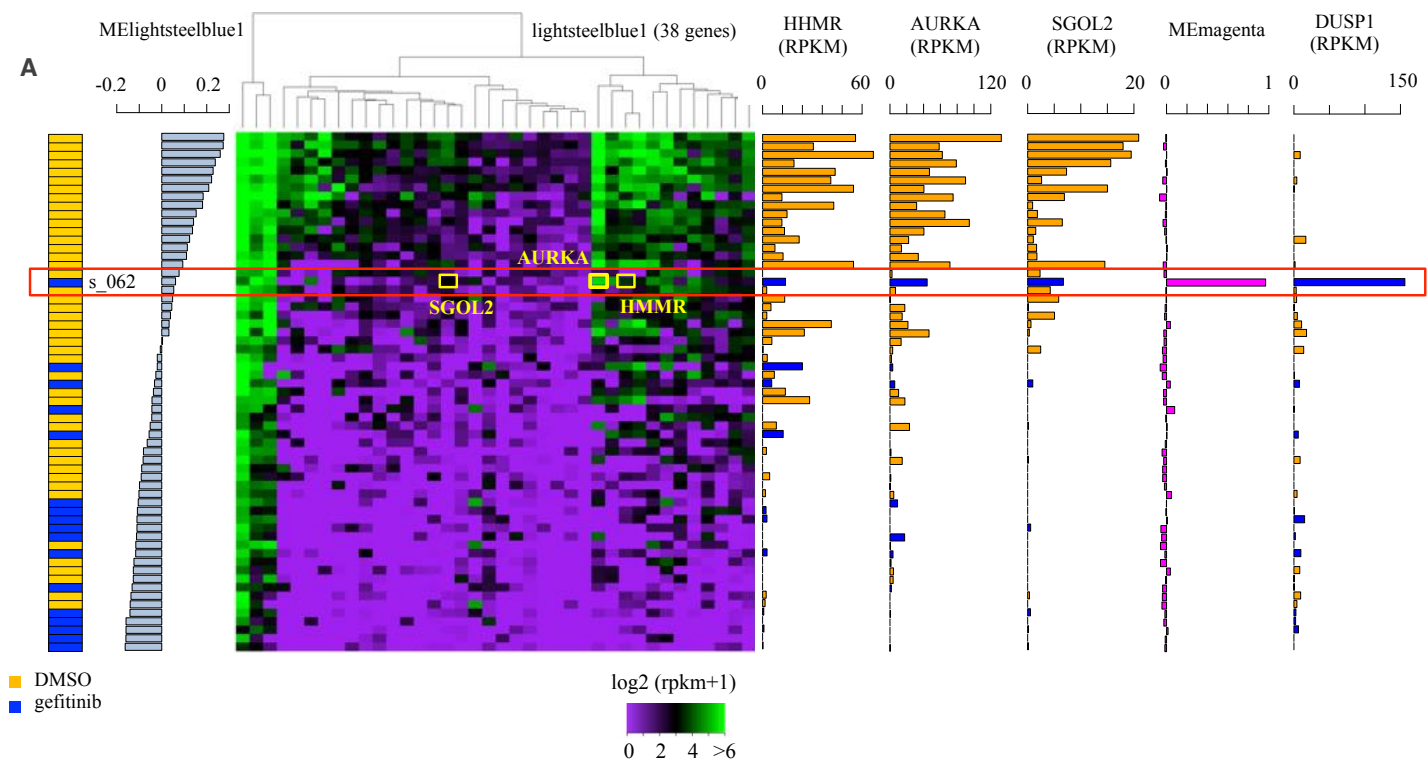
果を AURKA あるいはこのモジュールに含まれる他の遺伝子を高発現することで中和している可能性が示唆された。

外れ細胞の生物学的な可能性

他のモジュールに関してもこの s_062 細胞を精査した。同細胞では、“magenta”と呼ばれるモジュールにおいてその ME 値が高値を示した(図 5A 右から 2 段目、表 S5A: magenta 遺伝子リスト)。GO 解析から、このモジュールには、免疫反応に関与する遺伝子が多く含まれていた(表 S4B)。特に、DUSP1 の発現が高くなっていった(図 5A 右、図 5B)。DUSP1 遺伝子は核タンパクをエンコードし、心臓、肺、肝臓などで高発現している遺伝子である。先行研究から MAPK パスウェイの制御によって、DUSP1 は血管新生、がん浸潤、転移を促進することが知られている⁵¹⁻⁵³。DUSP1 の亢進した発現は、アポトーシスによる細胞死を制御し、ゲフィチニブ治療抵抗性の獲得につながると思われているが、詳細なメカニズムは未だ明らかではない⁵⁴。最近報告された論文では、キナーゼ異常による白血病において DUSP1 の発現レベルがチロシンキナーゼ阻害剤の治療効果に関連するという報告もある⁵⁵。s_062 では、AURKA-DUSP1 のモジュールが同時に活性化することが細胞における静止(dormancy)を誘導、これによりゲフィチニブの効果を低減している可能性があると考えた。

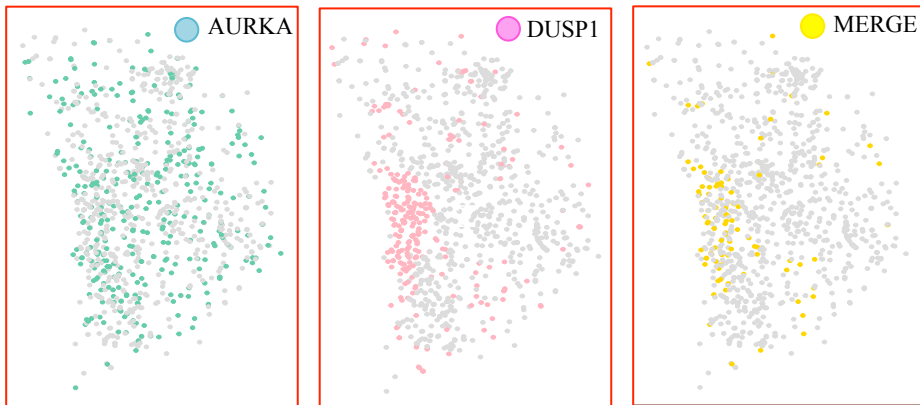
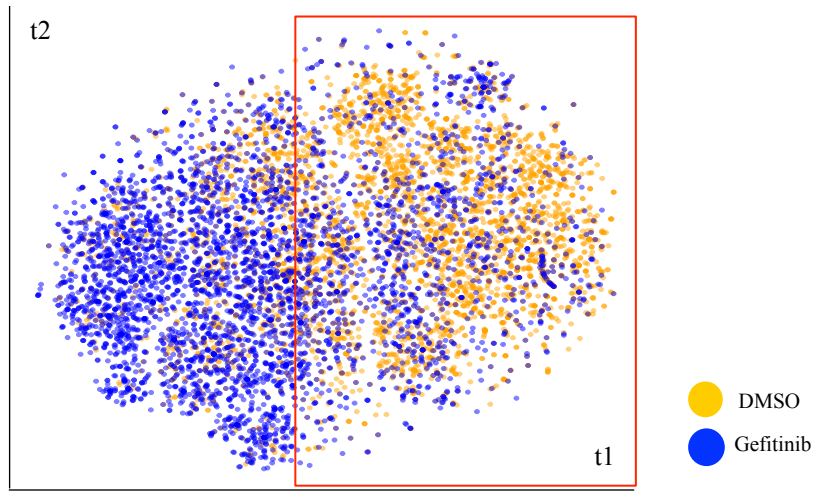
s_062 細胞と同様に AURKA と DUSP1 について発現量高値を示す細胞について、さらに解明を進めた。まず、PC9 細胞における s_062 様細胞の頻度を micro-droplet system のデータセットで解析した。さらに、異なる細胞種においても、同様にそれぞれのモジュールについてどのような多様性があるのかを解析した。この目的のために micro-chamber system で行なったものと同様の解析を micro-droplet system を用いて試みた。

9,544 細胞の PC9 細胞(5,166 細胞のコントロール群と 4,378 細胞のゲフィチニブ群)の“magenta”モジュールの発現プロファイルを tSNE により描画した。“magenta”は前述の通り DUSP1 が含まれるモジュールである(表 S5A)。tSNE によるプロットでは、細胞は大まかにコントロール群(cluster2)とゲフィチニブ刺激群(cluster1)にそれぞれ分かれた(図 S6)。micro-chamber system のデータセットから得られた結果と同様にゲフィチニブ処理された細胞は、コントロール群に属していた(図 5C, 左)。これらの細胞が、s_062 に類似した特徴を持っているのかを検証するために、AURKA と DUSP1 の発現レベルを調べた。AURKA の高発現の細胞は、ゲフィチニブ処理細胞群中に存在していた(図 5C, 右、“merged”パネルの黄色枠)。92 個の細胞(cluster2 に存在するゲフィチニブ処理細胞のうちの 8.8%)については、DUSP1 と AURKA 両方の遺伝子で高い発現を示していた(図 5C 中の表)。また、統計学的な評価により、この二つの遺伝子の発現上昇は、独立した事象であると考えられた(図 5C 組み込み表, Fisher の関連性検定, $p = 0.39$)。



C

tSNE (module magenta)



Gefitinib		DUSP1		total
		+	-	
AURKA	+	92 (9%)	264 (25%)	356 (34%)
	-	183 (17%)	498 (48%)	681 (66%)
total		275 (27%)	762 (73%)	1,037 (100%)

*p=0.39

図5： micro-chamber system データセットの micro-droplet system への写像

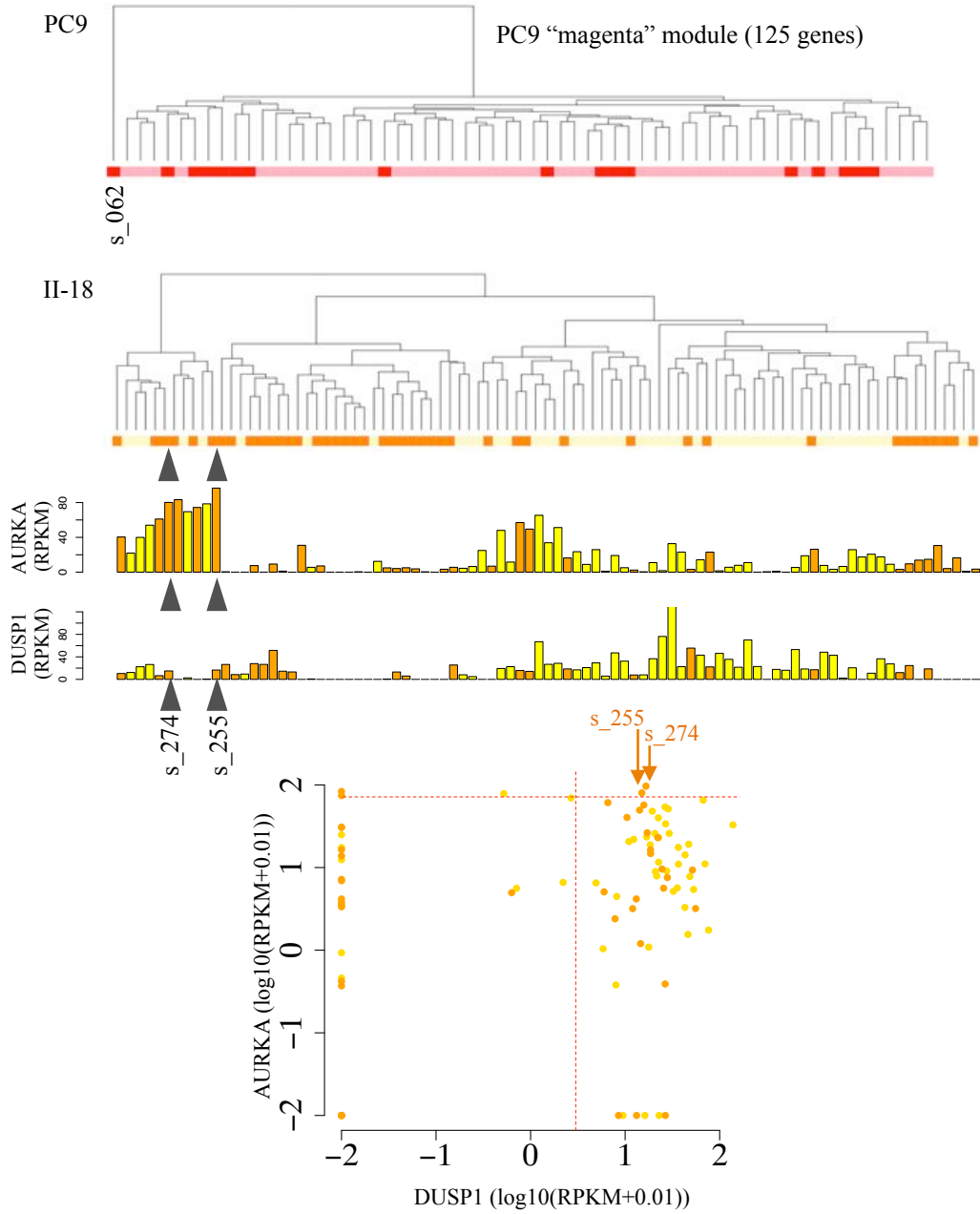
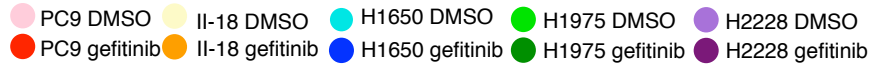
(A) “lightsteelblue1”モジュールによる階層クラスタリング。MElightsteelblue1 と各細胞に対する処理はヒートマップ左側の余白に示した。モジュールメンバーシップの上位3位の遺伝子の発現レベルと MEmagenta、DUSP1 の発現レベルはヒートマップの右側の余白に示した。(B) DUSP1 (x-axis) と AURKA (y-axis) の micro-chamber system データセットでの発現レベル。(C) 9,544 PC9 一細胞発現プロファイルによるクラスタリング。PC9-magenta のモジュールに含まれる遺伝子の発現レベルに基づいて tSNE による二次元プロットを行なった。オレンジ色はコントロール群、青色はゲフィチニブ群の細胞を示している(左)。PC9-magenta によりクラスタリングした際に cluster1 に存在する PC9 ゲフィチニブ群の 1,037 細胞のうち、AURKA を発現 (パネル：緑) と DUSP1 を発現 (パネル：ピンク)、どちらも発現(パネル：黄色)。AURKA と DUSP1 を発現している細胞数は組み込まれた表に示した。フィッシャーの有意差検定による p 値は下に示されている(*p=0.39)。

他の細胞でのモジュールの活性

DUSP1-AURKA の遺伝子の高い活性がゲフィチニブ処理された PC9 細胞に特有なものであるのかを検証した。他の細胞における micro-chamber system のデータセットを使用し、同様のモジュール解析を行った(図 S7A)。H1975 細胞と H2228 細胞においては、それぞれの細胞はコントロール、ゲフィチニブによってクラスターが分離されなかった。これらの細胞においては、ゲフィチニブによる大きなトランスクリプトームの変化が起きていないことが示唆された(図 S7B)。これは、それぞれの細胞株がゲフィチニブ耐性ゲノム変異を有していることに照らして、妥当であると考えられた(表 1)。一方で、PC9 細胞と II-18 細胞においてはゲフィチニブ処理された細胞群とコントロール群に特徴的な発現パターンが観測された。H1650 細胞においては、ゲフィチニブ処理された細胞群とコントロール群が混在していた。これは、H1650 細胞がゲフィチニブ感受性であることを反映したものであると考えられた。II-18 細胞において、PC9 細胞の”magenta”のモジュール(DUSP1 の含まれるモジュール)を使用してクラスター解析を行うと、s_062 のような明確な外れ細胞は検出されなかった。いくつかの細胞は AURKA と DUSP1 の発現が高かったものの(図 5D 矢印、s_255 と s_274)、PC9 細胞での s_062 細胞ほど顕著なものはいだされなかった。

II-18 細胞のデータセットを用いて、新規にモジュール”red”を同定した。このモジュールは、SOX4 をコアとする 132 遺伝子から構成されていた(表 S6A)。同モジュールを利用してクラスタリング解析を行うと二つの細胞(s_252 細胞と s_247 細胞)が外れ値細胞として検出され、これらの外れ細胞では、”turquoise”モジュールの ME 値が高く、CD44 が高発現であった(図 5E)。前章と同様の micro-droplet system データセットによる解析は、このような細胞が、全体の 0.06%存在することを示していた(図 S8)。SOX4 は腫瘍抑制因子として振る舞うことが知られているが、細胞周期停止やアポトーシスを促進することも報告されている⁵⁶⁻⁵⁸。CD44 は、がん幹細胞マーカー、あるいはがん前駆細胞マーカーである⁵⁹。CD44 を発現する細胞は、幹細胞様の特徴を有するとする報告もある^{60,61}。これらのモジュールの亢進は、PC9 細胞では観察されなかった。以上の結果から、がん細胞は、いくつかの共通するモジュールを持つ一方で、多様なモジュールを固有に持つ可能性が示唆された。

D



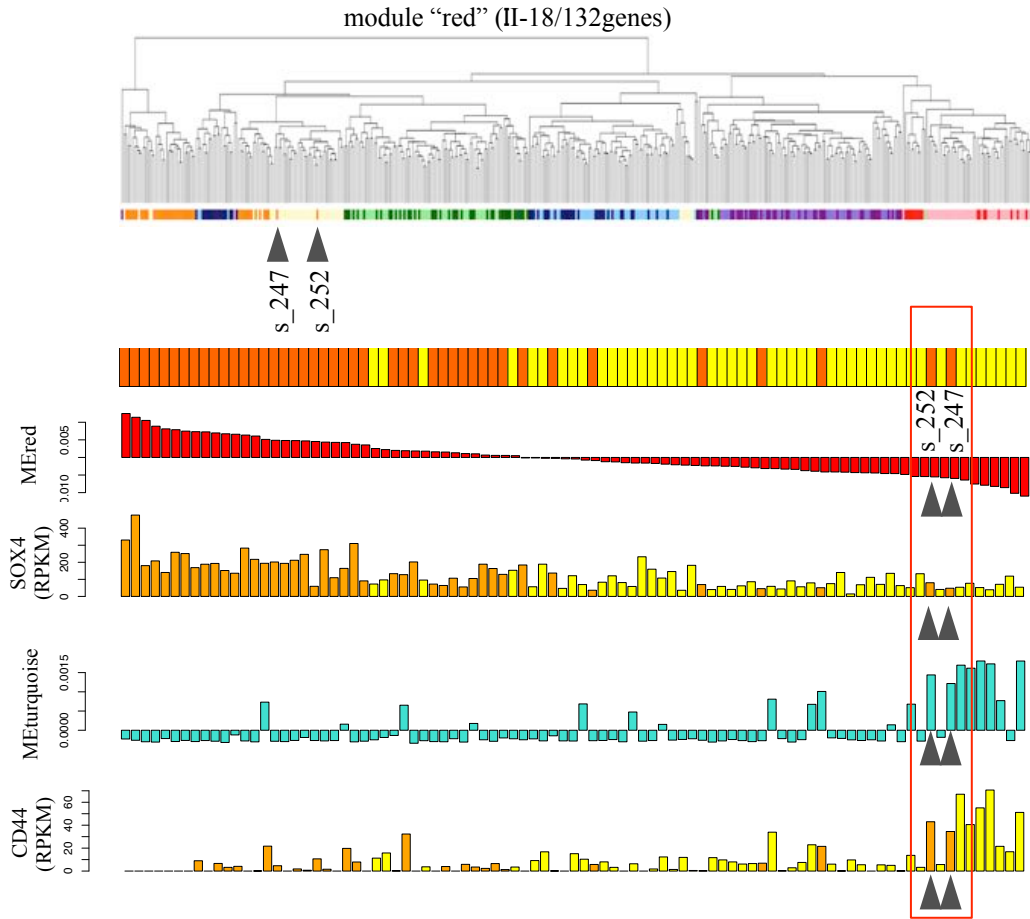
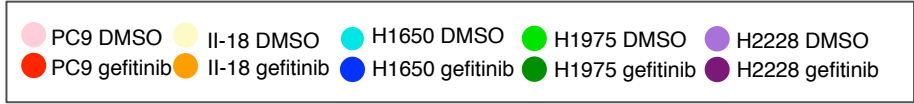
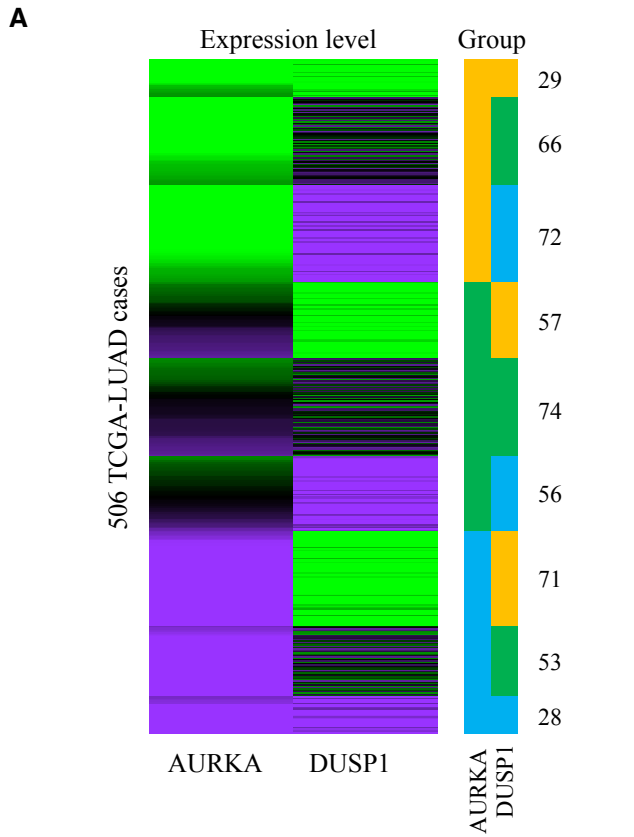
E

図5： micro-chamber system データセットの micro-droplet system への写像

(D) PC9 細胞 "magenta"モジュールによる階層クラスタリング。上段：PC9 細胞 micro-chamber のクラスタリング、下段：II-18 細胞 micro-chamber system のクラスタリング、AURKA の発現量 (RPKM)と DUSP1 の発現量(RPKM)。(E)II-18 細胞"red"モジュールによる階層クラスタリング。上段：全ての細胞種を用いたクラスタリング。下段：II-18 細胞のみのクラスタリング、II-18 細胞モジュール"red"の ME 値、Sox4 発現量(RPKM)、II-18 細胞モジュール"turquoise" ME 値、CD44 発現量(RPKM)。

外れ細胞の臨床的な意義

最後に臨床がんゲノムデータセットとして TCGA のデータセットを利用し、見出されたモジュール活性亢進細胞の臨床的意義を検証した。DUSP1 が高発現かつ AURKA が高発現である外れ細胞に該当するがん検体を調べた。この観点から TCGA でのトランスクリプトーム情報に、臨床情報を対応させた。はじめに、TCGA の肺腺がん患者由来データを AURKA と DUSP1 の発現に基づいてグループ分けを行った。506 肺腺がん症例の中で、29 症例で AURKA と DUSP1 が高い発現を示した(図 6A, 組み込み表)。細胞株での実験結果と同様に、二つの遺伝子間に直接的な相関は見られなかった。AURKA と DUSP1 の発現レベルと生存期間の比較を行った。両方の遺伝子の発現が高値を示す症例では、同遺伝子レベルが平均もしくは低発現レベルである症例と比べると予後が不良であることが明らかになった(図 6B)。AURKA と DUSP1 の活性化は、肺腺がん細胞株の生存に好影響を与えることが示唆された。逆に AURKA モジュールと DUSP1 モジュールの外れ細胞はがん細胞の生存に有利にはたらく可能性が示唆された。これらのモジュールは、抗がん剤治療抵抗性を微小細胞集団に惹起し、最終的な抗がん剤耐性獲得株の発生母地として寄与しているのかもしれない。



		DUSP1			
		High	Medium	Low	
AURKA	High	29	66	72	167
	Medium	57	74	56	187
	Low	71	53	28	152
		157	193	156	506

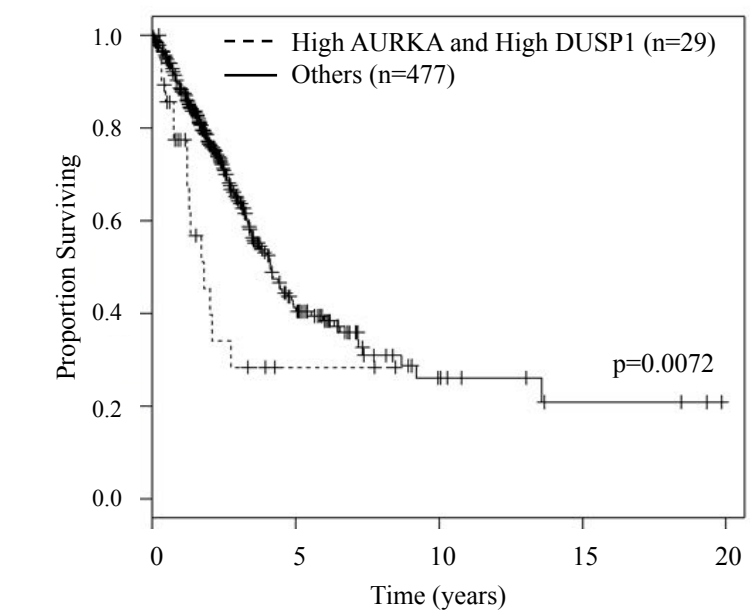
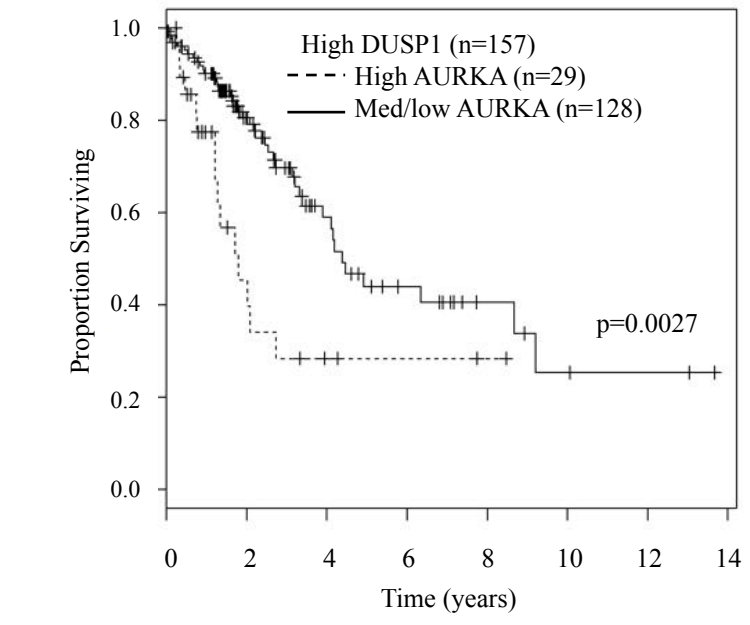
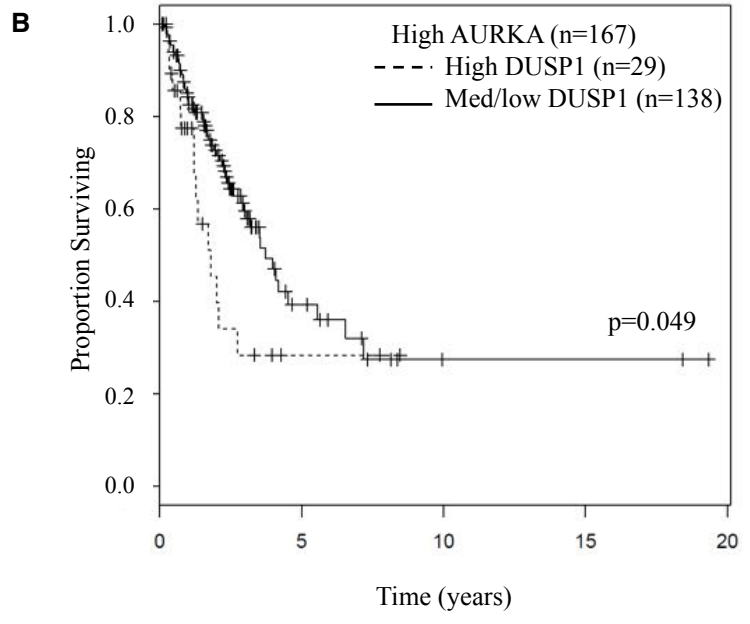
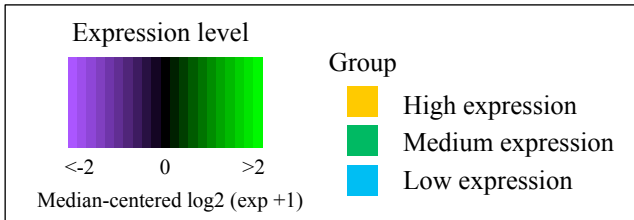


図 6: TCGA-LUAD データセットによる生物学的意義の解明

(A) TCGA データベース 506 患者の AURKA と DUSP1 の発現レベルを示すヒートマップ。患者は AURKA と DUSP1 の発現レベルに基づき 9 つの群に分類された。各群の患者数は、右の余白とインセットテーブルに示した。(B) カプランマイヤー曲線による解析。AURKA と DUSP1 が高発現である群においては、予後不良であることが示されている。二群間の統計学的有意差は、図中に示した。

2.3 考察

本研究の後半では、各プラットフォームから得られた一細胞 RNA-seq データセットを WGCNA により組み合わせることで、がん細胞のゲフィチニブに対する反応多様性の解明を行なった。遺伝子を単独ではなく、グループとして扱うことのできる module を抽出し、この module に基づいた解析により、ゲフィチニブ刺激下で、他の細胞とは異なる挙動を示す”外れ細胞”の検出を試みた。

はじめに、トランスクリプトームに対するカバー率が優れている micro-chamber system のデータセットを元にモジュールの抽出を行なった。PC9 細胞のデータセットからは、71 のモジュールが抽出された。固有値を利用したクラスタリングでは、s_062 細胞が lightsteelblue1 と呼ばれるモジュールで、コントロール群に入り込むことが示された。さらに、この一方で micro-droplet system においては、他の magenta と呼ばれるモジュールの値も特異的に高くなっていることが示された。

次に、この“外れ細胞”に該当する細胞を micro-droplet system のデータセットからも検出することで、トランスクリプトームに対するカバー率の低い micro-chamber system のデータセットを相互補完しようと試みた。micro-droplet system のデータセットからは、モジュールと遺伝子発現に基づき、92 細胞が”外れ細胞”として存在することが示唆された。また、この二つのモジュールを用いて、同じくゲフィチニブ感受性であると報告されている II-18 細胞の解析を行なったところ、s_062 細胞に相当するような細胞は検出されなかった。II-18 細胞からは、異なる遺伝子をコアとするモジュールが抽出され、PC9 細胞での解析と同様の方法で二つの”外れ細胞”が検出された。ここまでの結果から、本解析による”外れ細胞” “検出法は、他の細胞株にも応用できることが明らかにされた。

さらに、TCGA の臨床情報と RNA-seq の情報を利用し、ここまで解析してきた AURKA と DUSP1 が臨床検体において、予後にどのような影響を与えるのかを解析した。二つの遺伝子が高発現である症例では、統計学的に有意に予後が不良であることが示された。以上の結果から、二つのプラットフォーム由来のデータセットの組み合わせによる解析は一細胞解析での不均一性の解明において有効な手段になりうると思われる。

現状の一細胞解析では、本解析で検出された”外れ細胞” “に関して、さらなる解析を行うことは不可能である。今後の課題として、”外れ細胞” “の検出を可能にするサロゲートマーカーの探索が挙げられる。micro-droplet system のデータセットからは、外れ細胞として 92 細胞が検出された。これらの細胞で共通して発現している遺伝子がサロゲートマーカーの候補となりうるのではないかと考えられる。s_062 細胞は全長のシーケンスが得られる micro-chamber system のデータセットから検出されたため、今後の展開として、GATK(The Genome Analysis Toolkit)等を用いたさらなる解析なども考えられる⁶²。

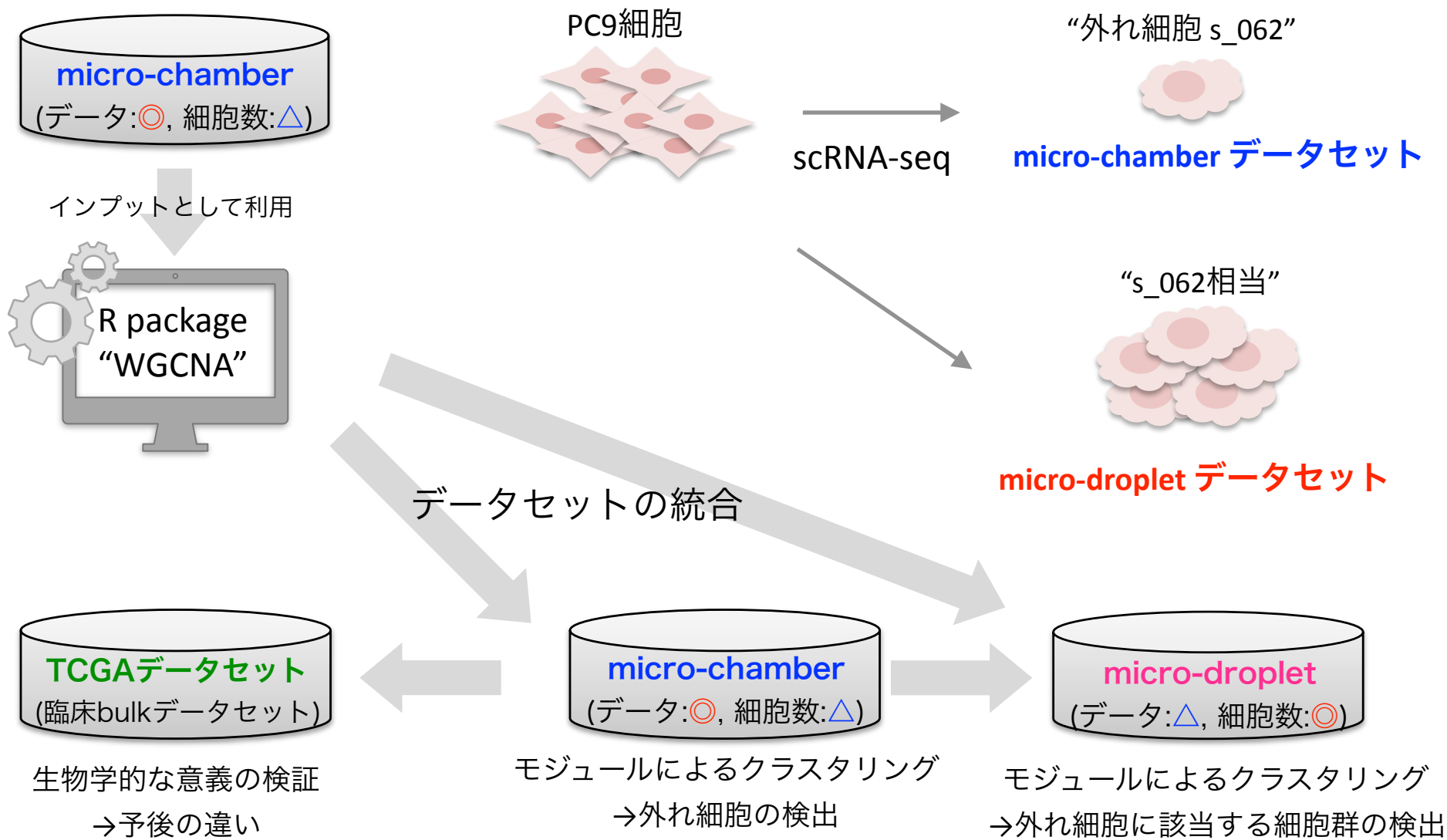
また、本解析の結果からのみでは”外れ細胞” “が、ゲフィチニブ刺激による結果であるのか、刺激前から変異を持っていた特異的な細胞であるのかを同定することは困難である。”外れ細胞” “の薬剤耐性における意義を明らかにするためには、さらなる検証が必要になると考えられる。

総括

本研究では、異なる一細胞 RNA-seq のプラットフォームの検証と異なるデータセットの組み合わせを試みた(図 S-1)。論文の第一章では、**micro-chamber system** と **micro-droplet system** を使用してデータセットを作出した。詳細なデータの評価によりこれらのプラットフォームの比較を行った。それぞれのプラットフォームは、高い再現性を示した。それぞれのプラットフォームでシークエンスの深度は大きく異なっていた。これは、解析された細胞数の違いに起因していた。シークエンス深度を主因として、プラットフォーム間でのデータの差異が見出された。シークエンス深度を情報学的に標準化したデータセットを用いた解析の結果、両プラットフォームのデータには質的差異は認められなかった。それぞれのプラットフォームには固有の利点や欠点が存在する。遺伝子発現情報に関しては **micro-chamber system** は各細胞の特徴の詳細な検証を可能にするが、対象となる細胞数が少ない。そのため希少な細胞集団の同定、および特徴的な細胞の集団中の存在比を解析することか困難である。一方、**micro-droplet system** では、個々の細胞のシークエンスの深度は一般に十分でなく、低発現量遺伝子を中心に、個々の発現データの精度に問題があった。

本研究の後半部では、二つのデータセットを組み合わせることで、より有効な解析を行うことが可能であるか検証した。**micro-chamber** のデータセットを出発材料に、**micro-droplet** の統計推定を試みた。**LASSO** を用いた発現量推定は、一定の有効性を有するものの、遺伝子の発現に依存して全ての遺伝子の発現情報を補完することは困難であった。そこで、共発現ネットワーク解析により遺伝子をモジュールとして解析する手法を検討、実践した。**micro-chamber** のデータセットを用いた **WGCNA** 解析を行うことで、**AURKA** と **DUSP1** をコア遺伝子として含む二つのモジュールを同定した。これらのモジュールに含まれる遺伝子の発現レベルを指標に、他の細胞とは異なる発現パターンを示す外れ細胞を同定した。さらに **micro-droplet** のデータセットにより、この外れ細胞が集団においてどの程度の頻度で出現するかを解析した。興味深いことに、**TCGA** に基づく解析からこれらの遺伝子が同時に活性化されることは、臨床予後不良因子である可能性が示唆された。

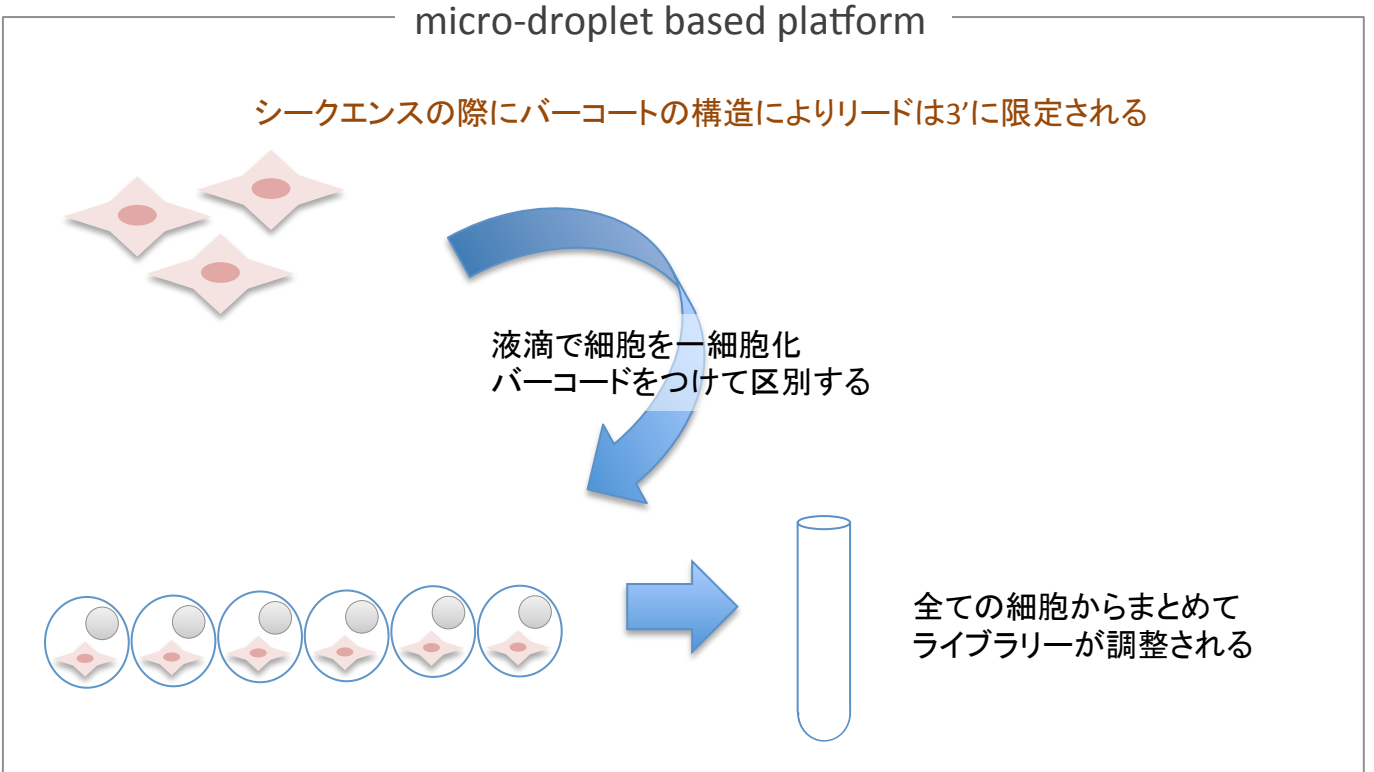
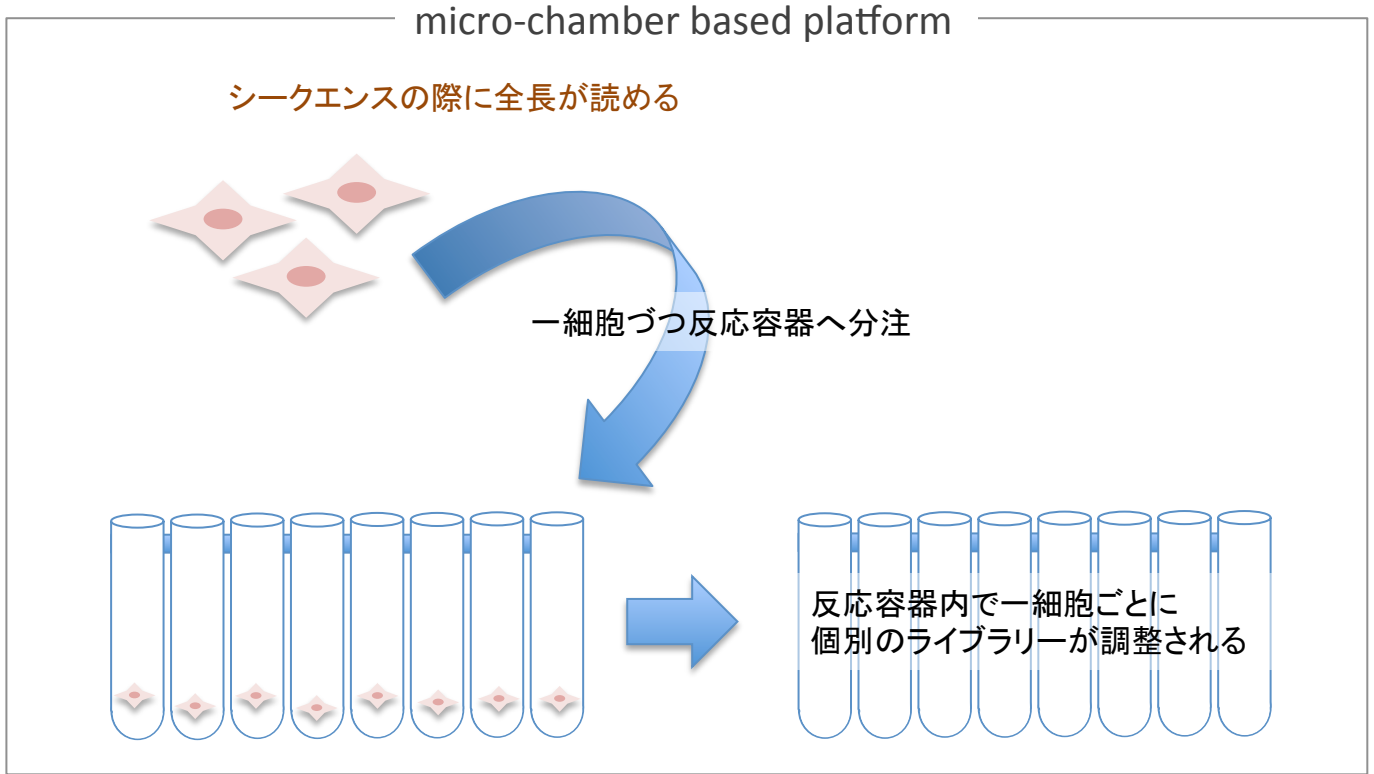
観察されたがん不均一性の臨床的な意義の解明のためには、さらなる検証が必要である。特に本研究では、*in vivo* で重要な要素である多様な微小環境が持つ細胞レベルでの遺伝子発現への影響を評価することはできなかった。近年、*in vivo* で一細胞レベルでのトランスクリプトームを測定する手法が提案されているが、その多くはまだ開発途上にある。正確性や空間分解能についてさらなる改良が必要であると思われる。様々な局面でより発展的な改良を要するものの、一細胞解析はがん細胞の多様性の分子基盤の解明に向けて、不可欠な技術になっている。本研究ががんの多様性に関する研究における一助になることを期待する。



図S-1：本研究の総括

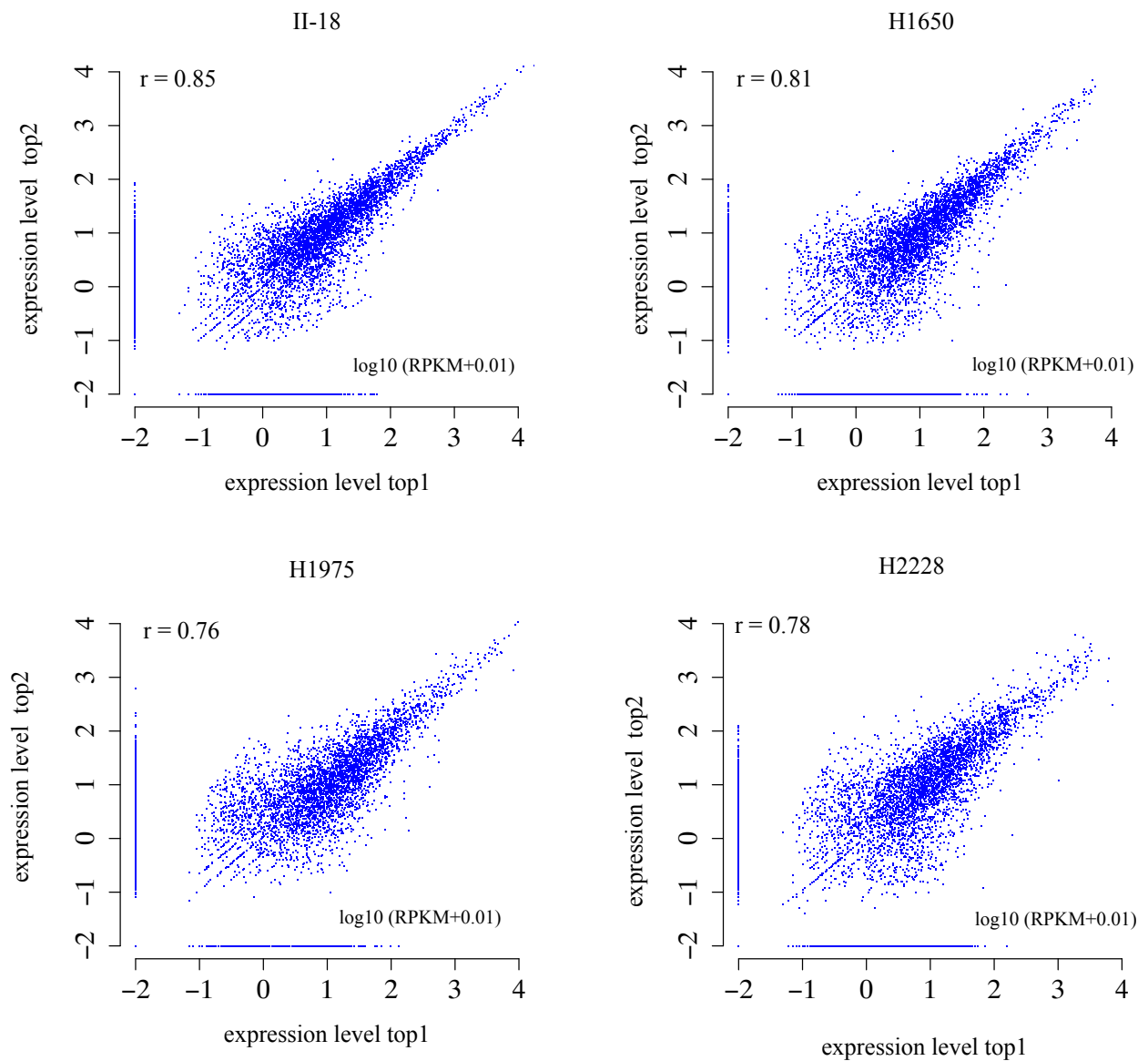
一細胞RNA-seqのプラットフォームであるmicro-chamber systemとmicro-droplet systemはそれぞれデータセットに特徴があるが、一方のデータセットだけでは、希少な細胞の詳細な解析や集団における頻度の推定をすることは困難である。しかし、互いのデータセットを組み合わせ、共発現ネットワーク解析を行うことで、がん抵抗性に関与する可能性が示唆される外れ細胞を検出することを可能にした。

補足データ

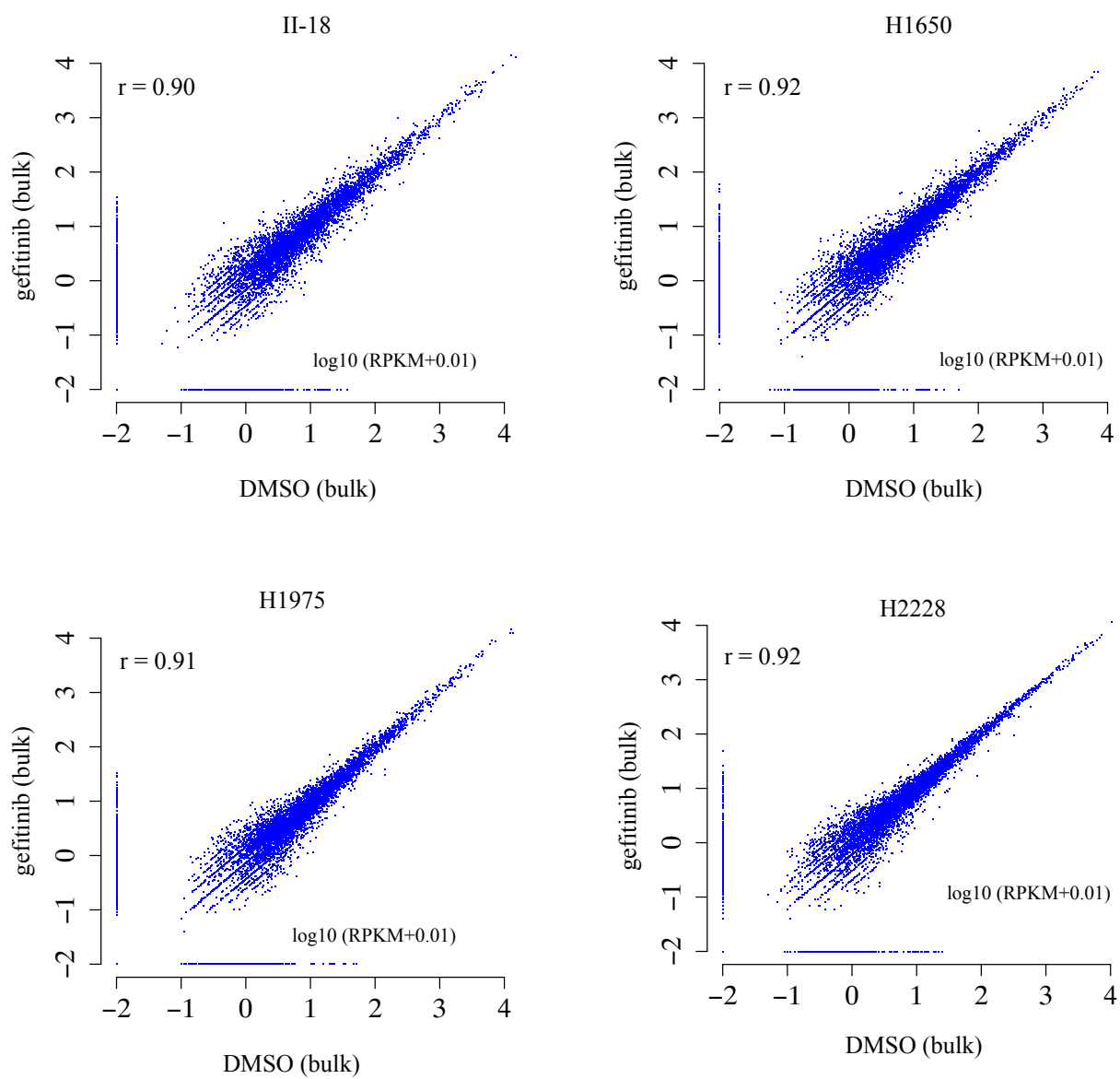


FigS1

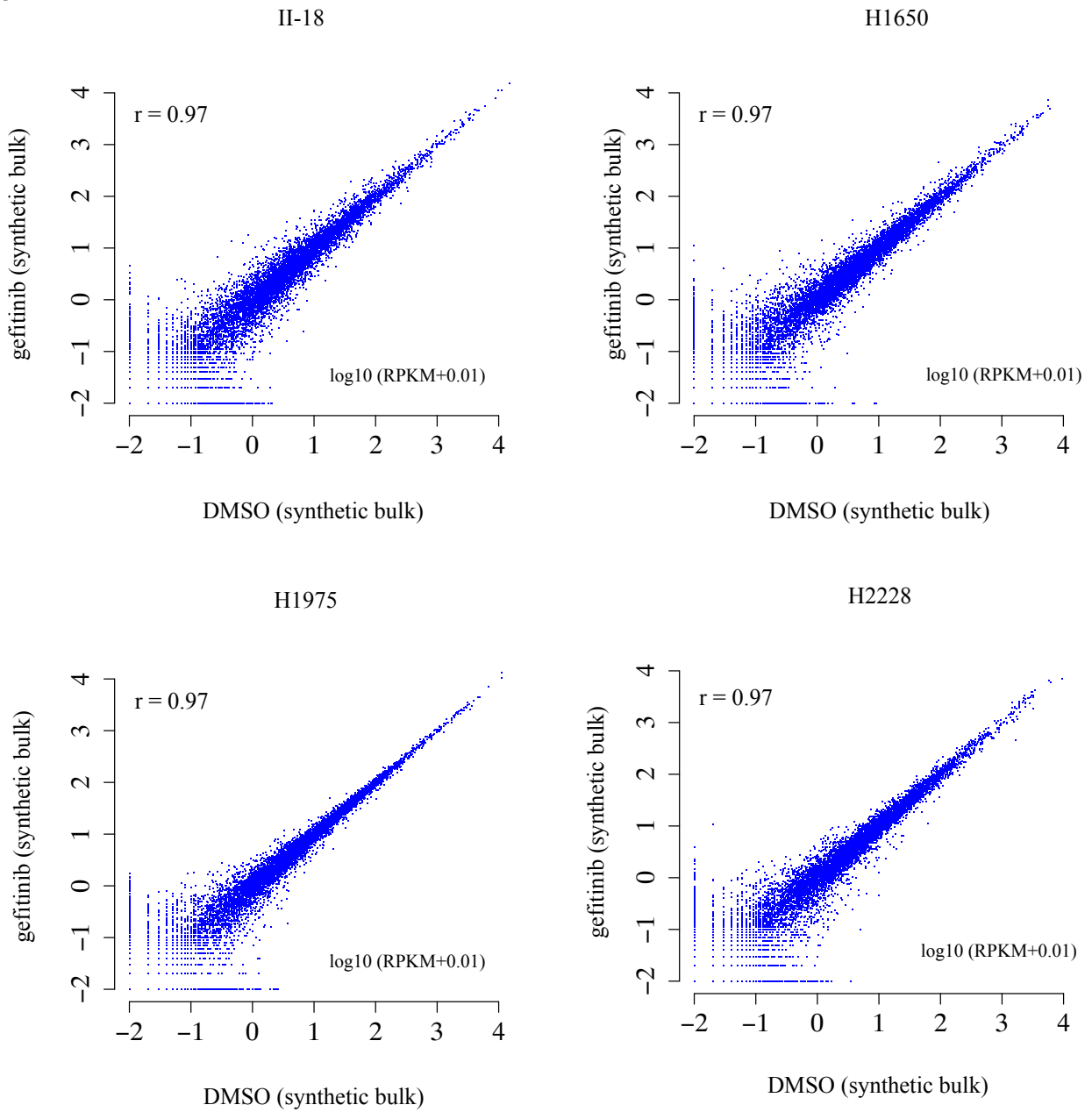
- (A) micro-chamber systemでの一細胞分離。
- (B) micro-droplet systemでの一細胞分離。

A

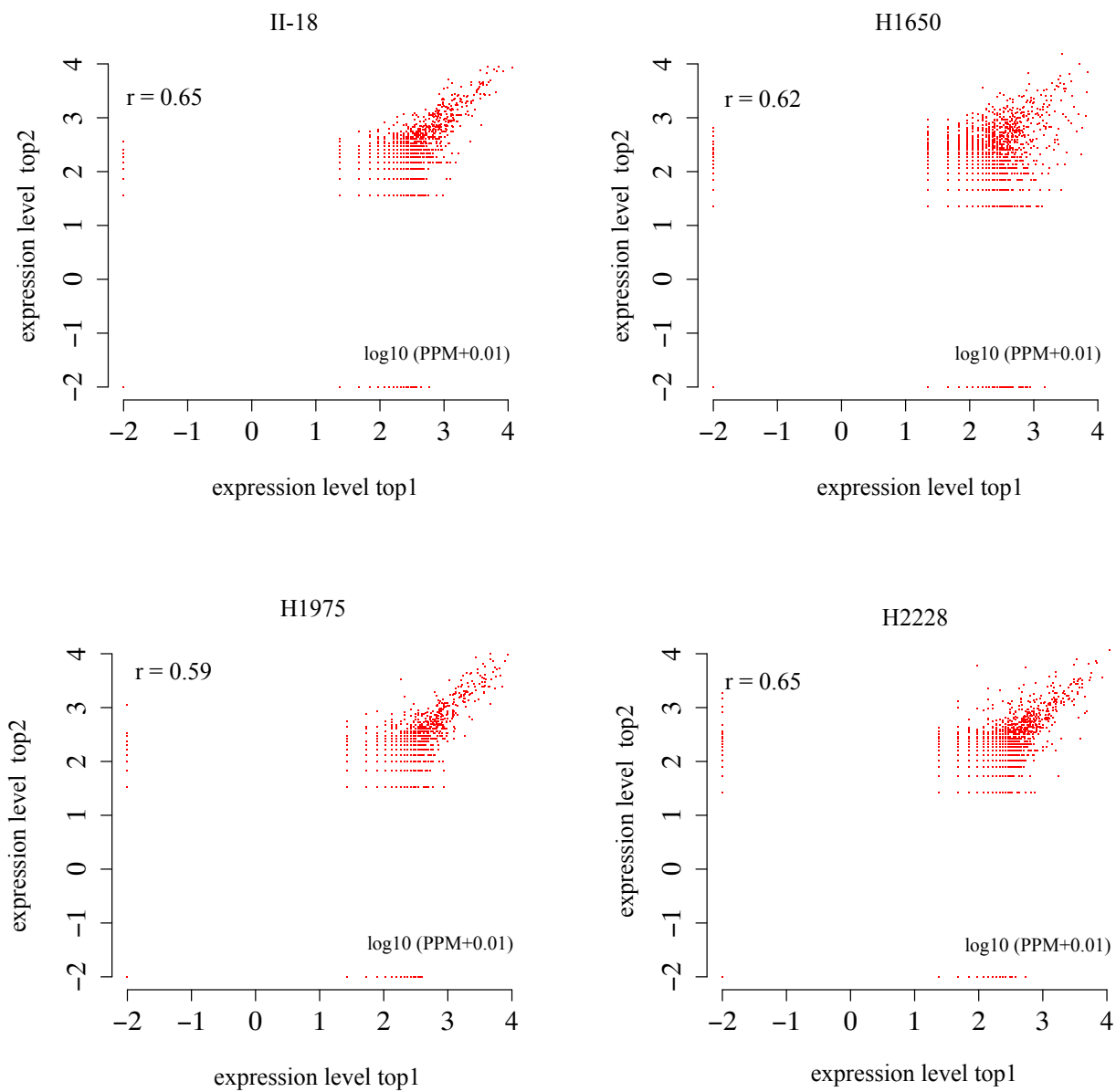
FigS2 micro-chamberシステムとmicro-dropletシステムを用いたRNAシーケンスデータセットの比較
(A)micro-chamber systemでのtop1細胞とtop2細胞を用いた比較。II-18細胞(左上)、H1650細胞(右上)、H1975細胞(左下)、H2228細胞(右下)。ピアソン係数(r)は図中に記載

B

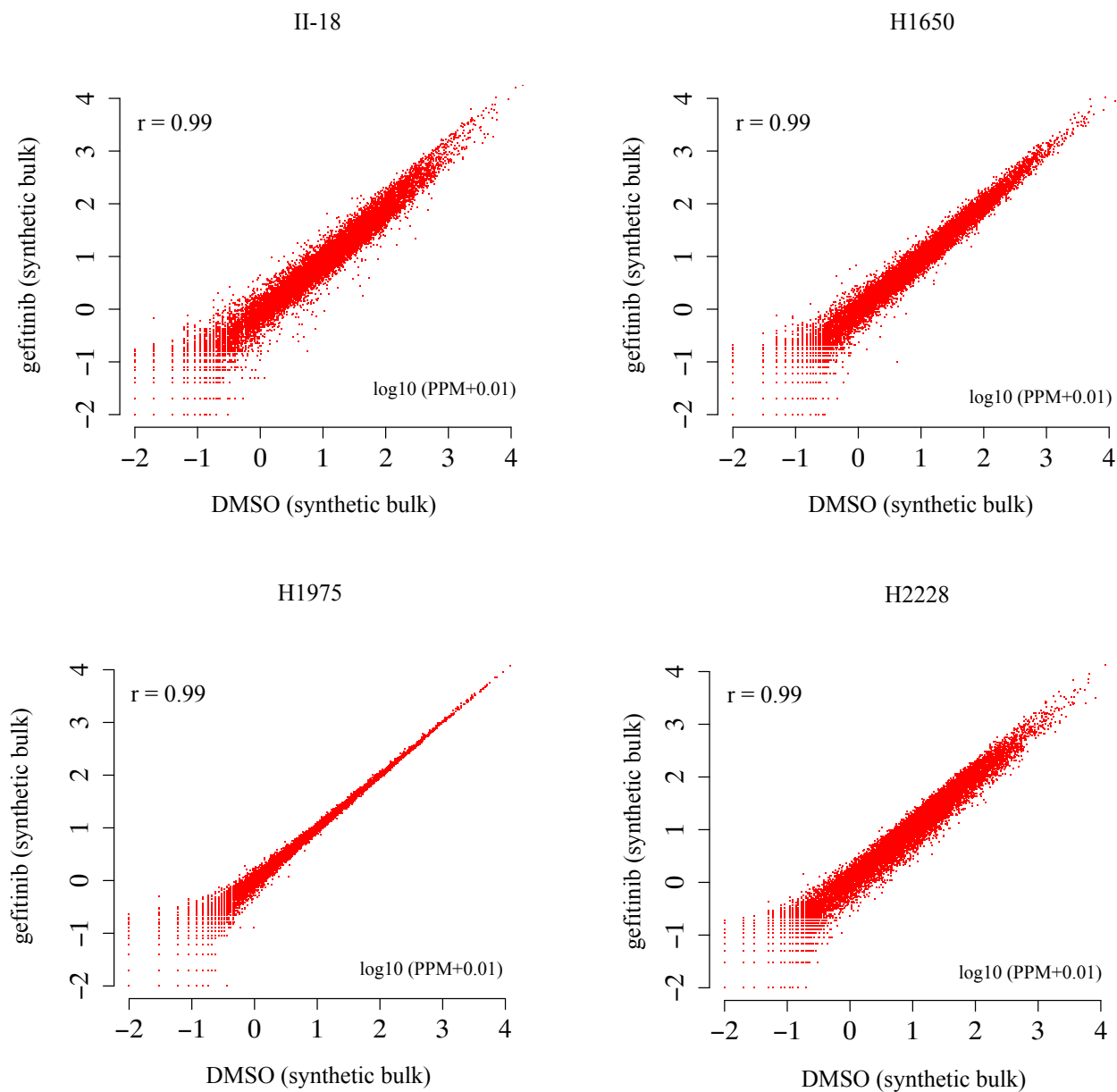
FigS2 micro-chamberシステムとmicro-dropletシステムを用いたRNAシーケンスデータセットの比較
(B)micro-chamber systemでのコントロール群とGefitinib刺激群でのbulkレベルでの比較。H118細胞(左上)、H1650細胞(右上)、H1975細胞(左下)、H2228細胞(右下)。ピアソン係数(r)は図中に記載

C

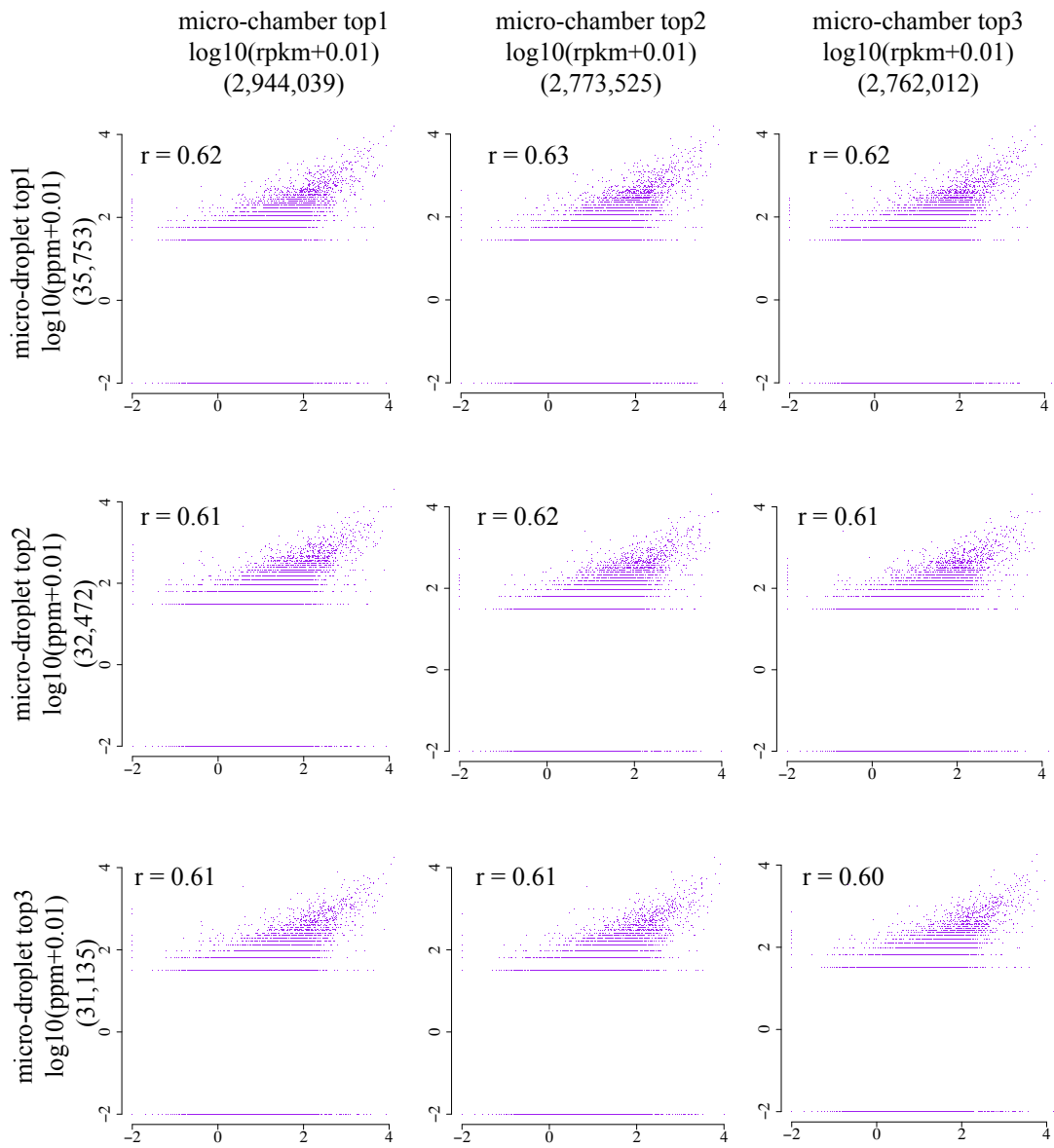
FigS2 micro-chamberシステムとmicro-dropletシステムを用いたRNAシーケンスデータセットの比較
(C)micro-chamber systemでのコントロール群とGefitinib刺激群でのsynthetic bulkレベルでの比較。II-18細胞(左上)、
H1650細胞(右上)、H1975細胞(左下)、H2228細胞(右下)。ピアソン係数(r)は図中に記載

D

FigS2 micro-chamberシステムとmicro-dropletシステムを用いたRNAシーケンスデータセットの比較
(D)micro-droplet systemでのtop1細胞とtop2細胞で一細胞レベルでの比較。II-18細胞(左上)、H1650細胞(右上)、
H1975細胞(左下)、H2228細胞(右下)。ピアソン係数(r)は図中に記載

F

FigS2 micro-chamberシステムとmicro-dropletシステムを用いたRNAシーケンスデータセットの比較
(D)micro-droplet systemでのコントロール群とGefitinib刺激群でのsynthetic bulkレベルでの比較。H1-18細胞(左上)、
H1650細胞(右上)、H1975細胞(左下)、H2228細胞(右下)。ピアソン係数(r)は図中に記載



FigS3 一細胞レベルでのmicro-chamberシステムとmicro-dropletシステムデータセットの比較
各システムでのtop1細胞、top2細胞、top3細胞での一細胞レベルでの比較。ピアソン係数(r)は図中に記載

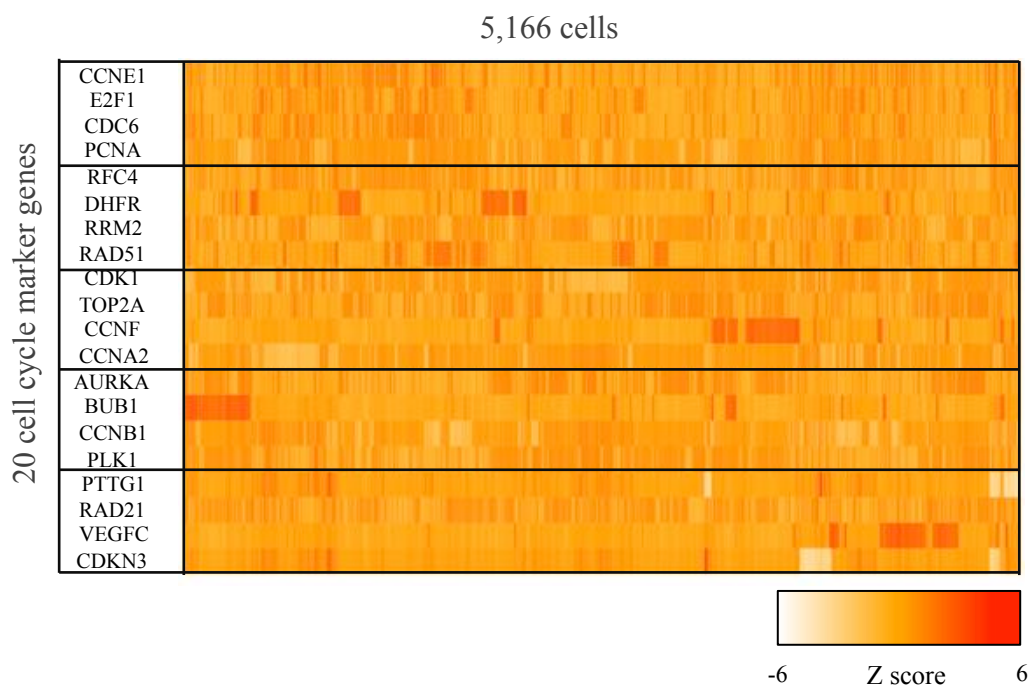
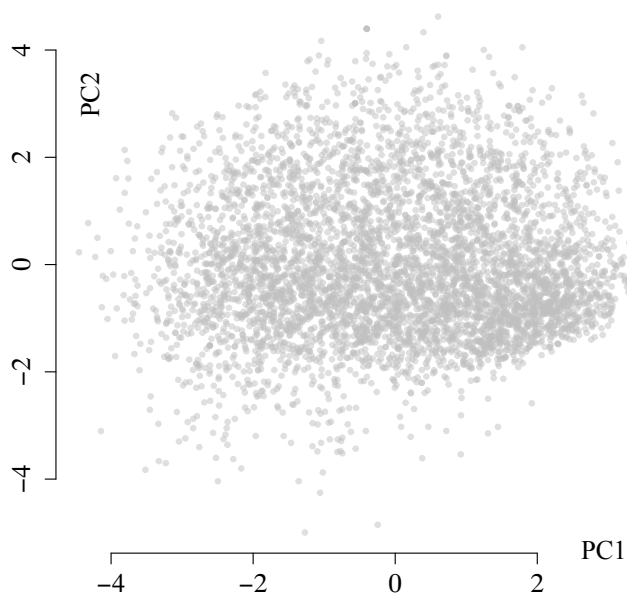
A**B**

Fig.S4 micro-droplet システムでの20遺伝子による細胞周期の推定
micro-chamber systemでの細胞周期の推定と同じ遺伝子により細胞周期を推定。(A)ヒートマップによる図示,(B)PCA
による図示。

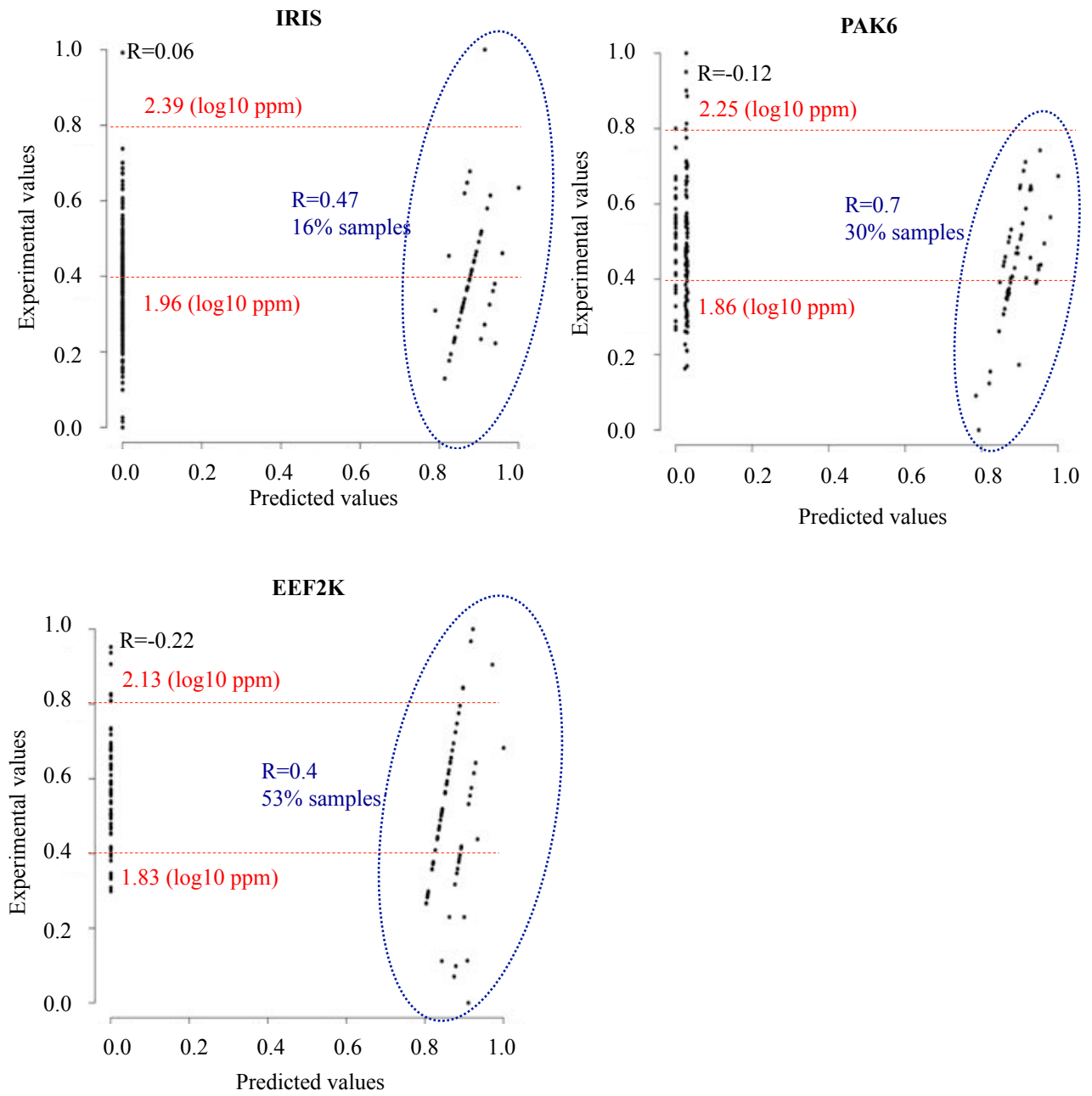
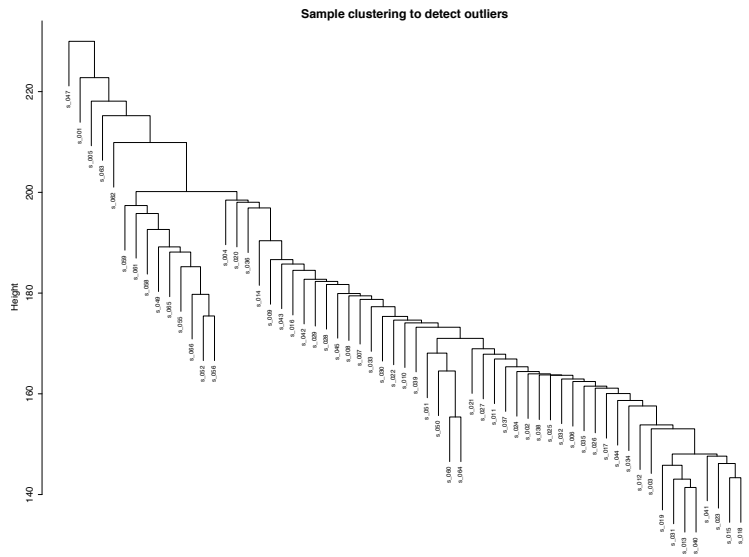


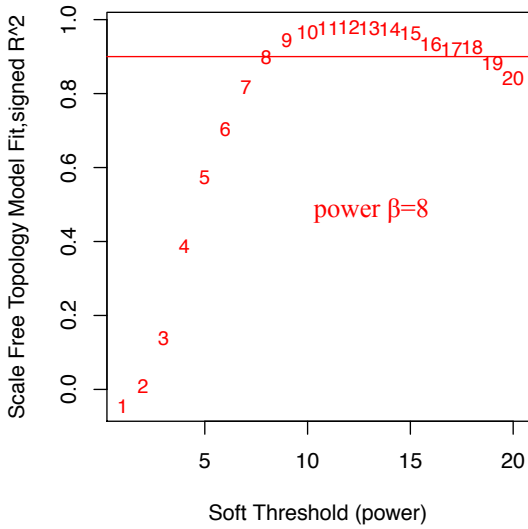
Fig.S4 欠損値推定の評価

LASSOによる欠損値推定の結果(x-axis)と実験での実測値(y-axis)の相関。各軸の値は標準化されており、0.4と0.2に相当する発現量が示されている(赤字)。全体の相関係数 r は黒字、青枠内の細胞の相関係数は青字で図中に記載。

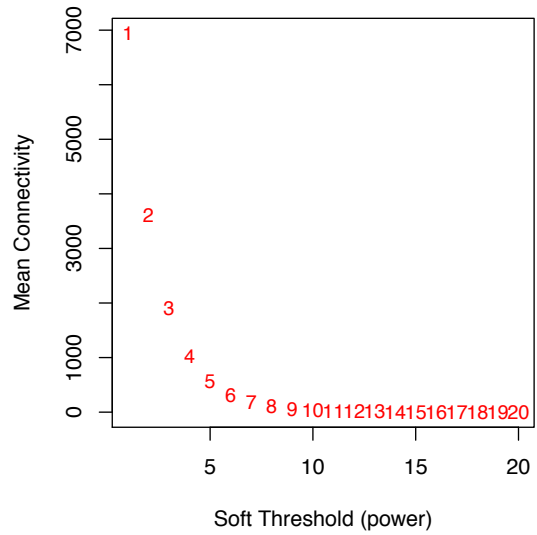
micro-chamber dataset 22913 genes/ 66 cells
 --> remove outlier cells
 (s_046, s_048, s_053, s_054, s_057)
 --> remove low expression genes
 (5rpkm > at least 1 cell)
 --> 13,619 genes / 61 cells
 (DMSO 44 cells, gefitinib 17 cells)



Scale independence



Mean connectivity



71 modules : identified

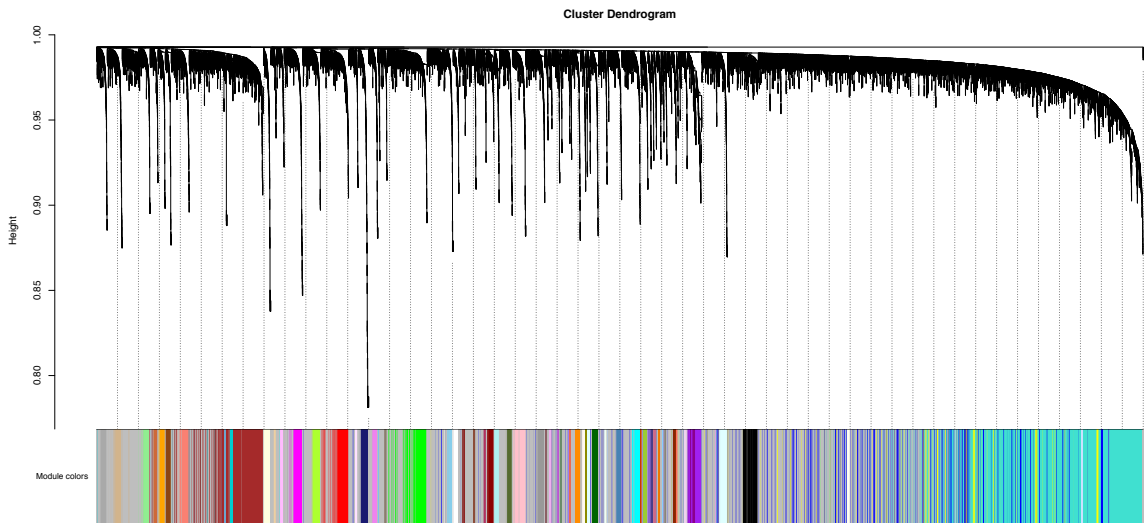
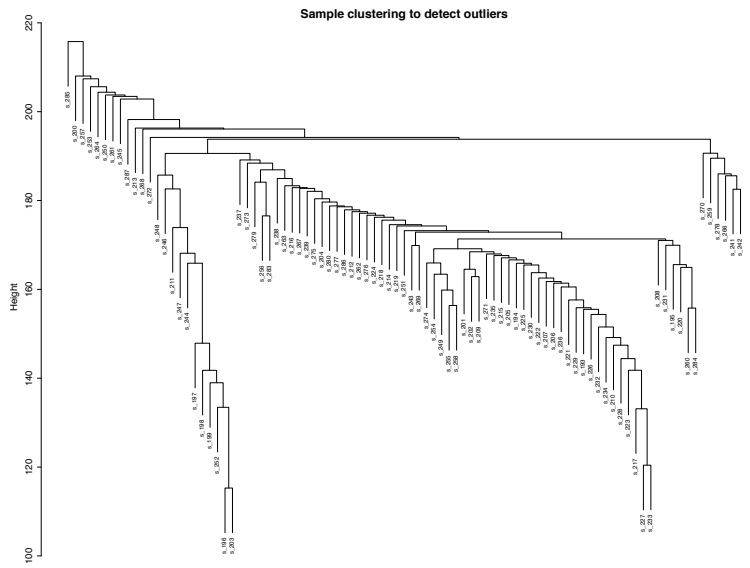
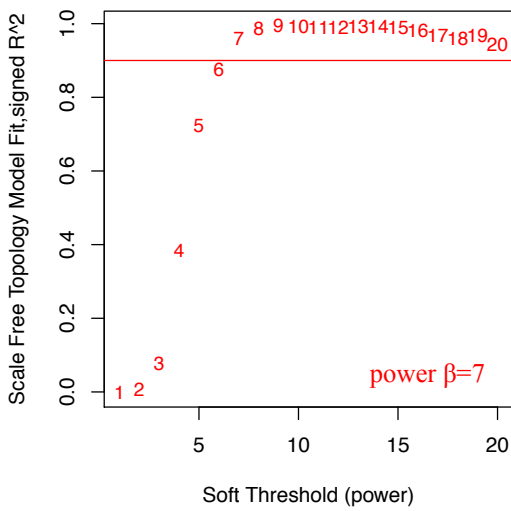


Fig.S6 WCGGAによる各細胞株でのモジュールの抽出
 (A) PC9細胞でのモジュール抽出。細胞のクラスタリング(上)、powerの設定(中央)、モジュールの抽出(下)。

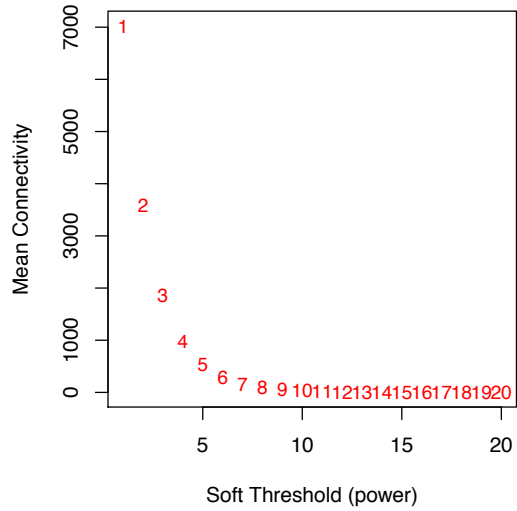
micro-chamber dataset 22913 genes/ 94 cells
 --> remove outlier cells
 (s_265, s_281, s_282)
 --> remove low expression genes
 (5rpkm> at least 1 cell)
 --> 13594 genes / 91 cells
 (DMSO 47 cells, gefitinib 44 cells)



Scale independence



Mean connectivity



58 modules : identified

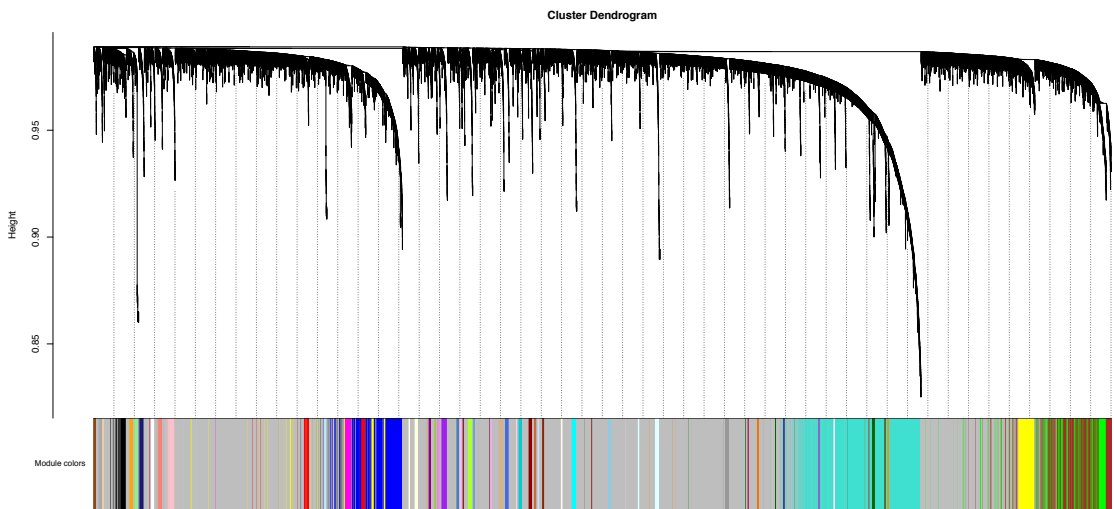
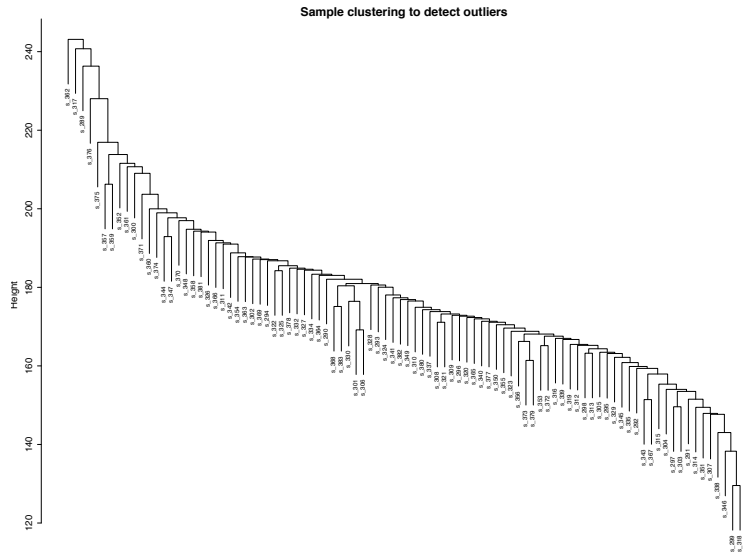
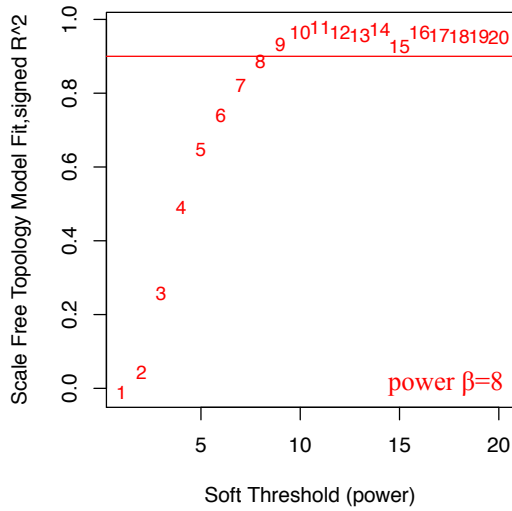


Fig.S6 WCGGAによる各細胞株でのモジュールの抽出
 (B) II-18細胞でのモジュール抽出。細胞のクラスタリング(上)、powerの設定(中央)、モジュールの抽出(下)。

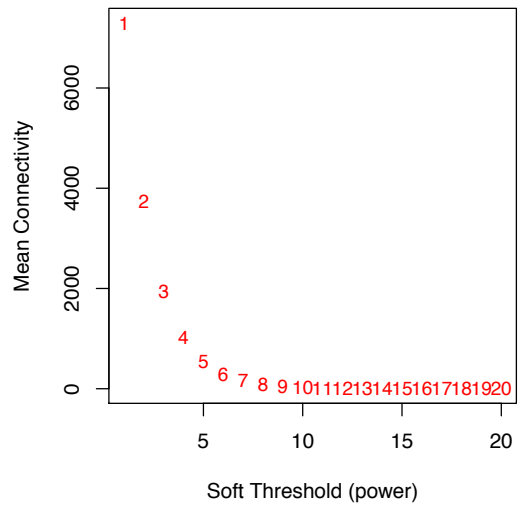
micro-chamber dataset 22913 genes/ 94 cells
 --> remove outlier cells
 (s_331, s_333)
 --> remove low expression genes
 (5rpkm> at least 1 cell)
 --> 14429 genes / 92 cells
 (DMSO 45 cells, gefitinib 47 cells)



Scale independence



Mean connectivity



80 modules : identified

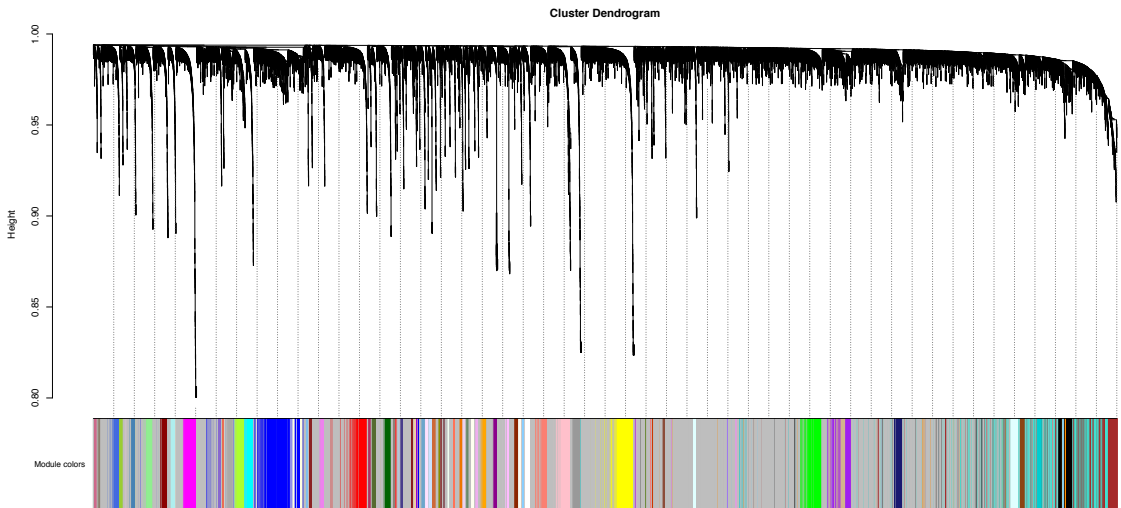
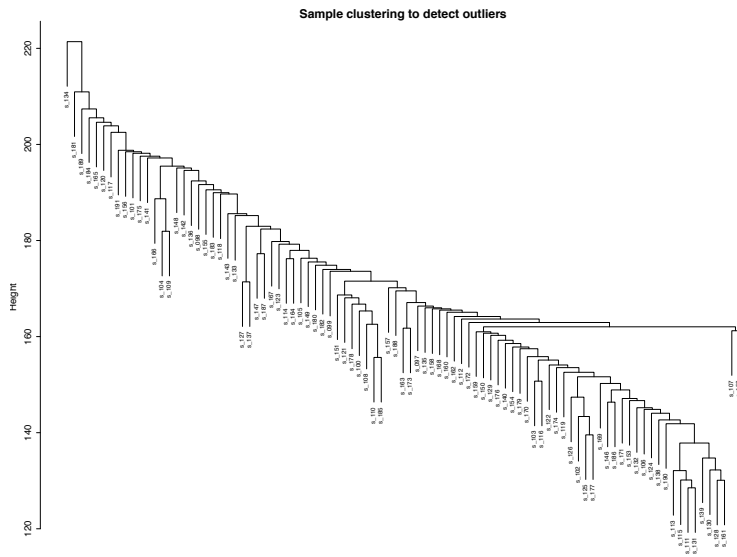
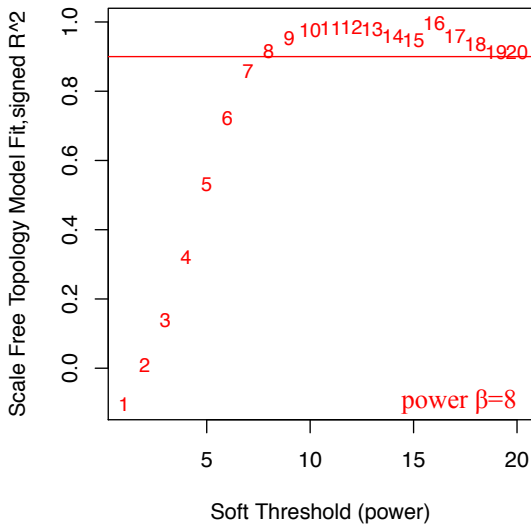


Fig.S6 WGCNAによる各細胞株でのモジュールの抽出
 (C) H1650細胞でのモジュール抽出。細胞のクラスタリング(上)、powerの設定(中央)、モジュールの抽出(下)。

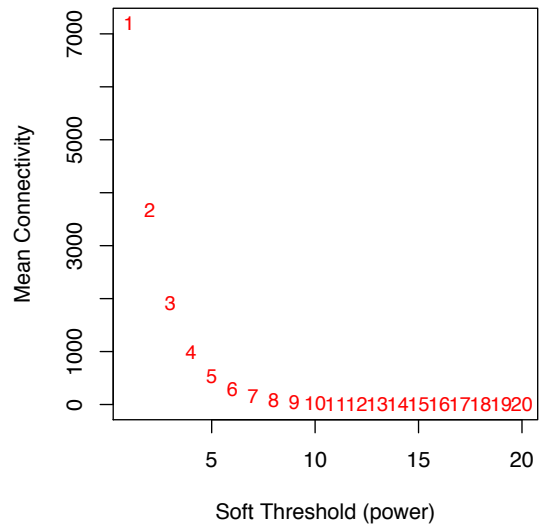
micro-chamber dataset 22913 genes/ 94 cells
 --> remove outlier cells (s_145)
 --> remove low expression genes (5rpkm> at least 1 cell)
 --> 14272 genes / 93 cells (DMSO 47 cells, gefitinib 46 cells)



Scale independence



Mean connectivity



83 modules : identified

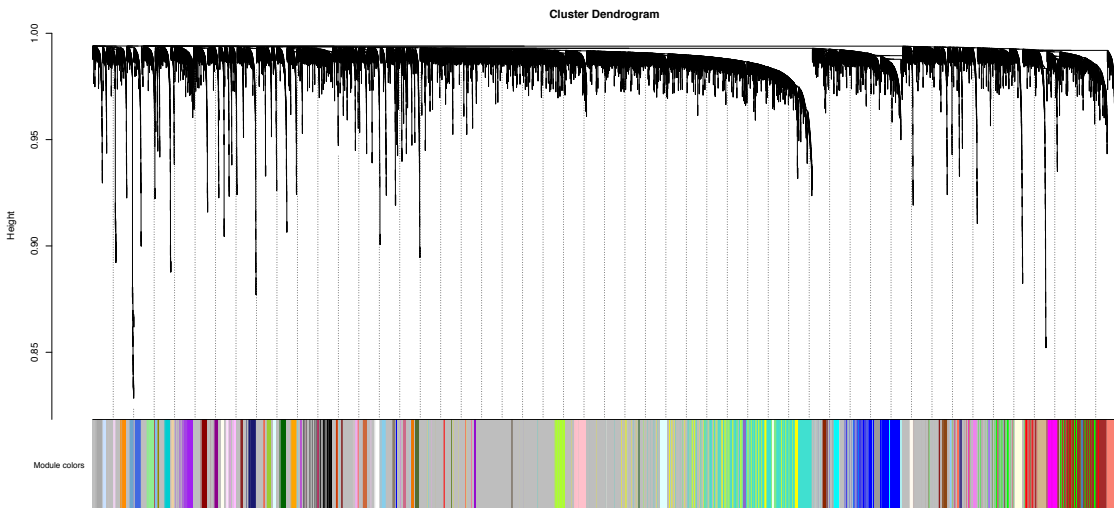
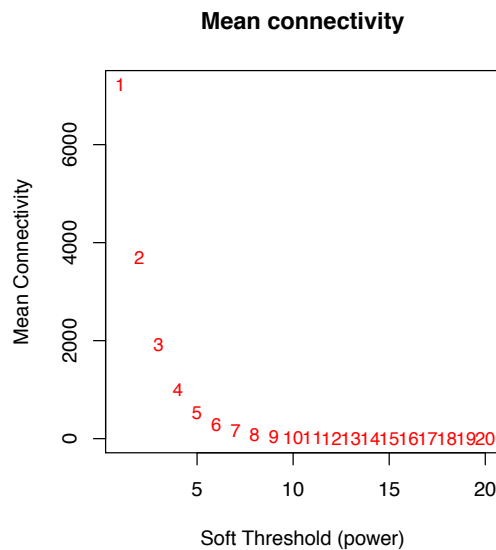
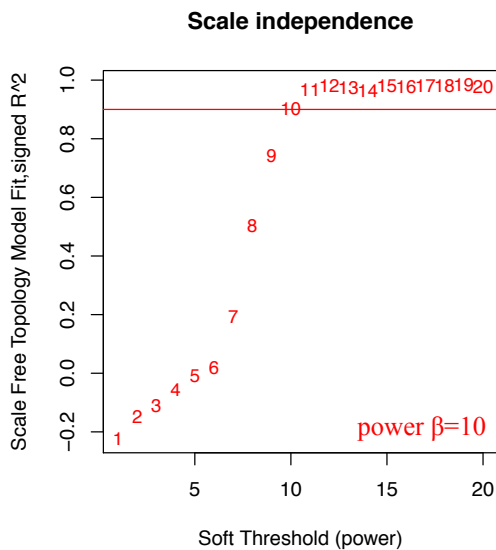
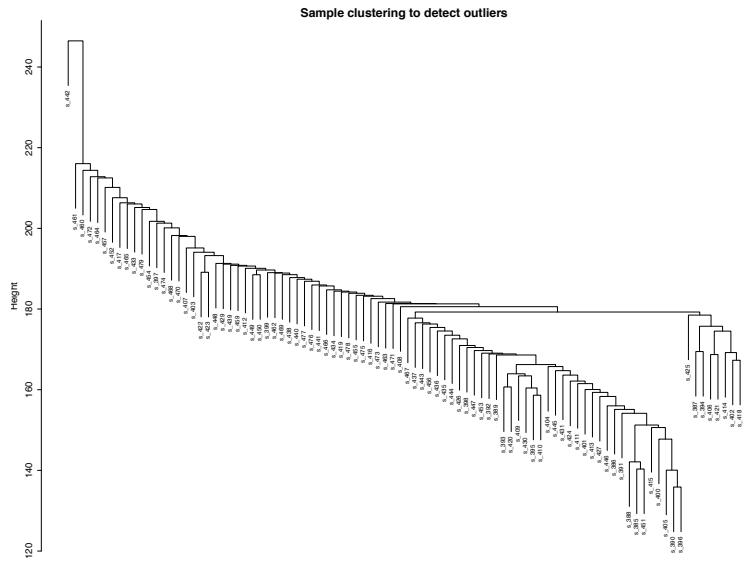


Fig.S6 WGCNAによる各細胞株でのモジュールの抽出 (D) H1975細胞でのモジュール抽出。細胞のクラスタリング(上)、powerの設定(中央)、モジュールの抽出(下)。

micro-chamber dataset 22913 genes/ 94 cells
 --> remove outlier cells
 (s_428, s_458)
 --> remove low expression genes
 (5rpkm> at least 1 cell)
 --> 14293 genes / 92 cells
 (DMSO 47 cells, gefitinib 45 cells)



75 modules identified

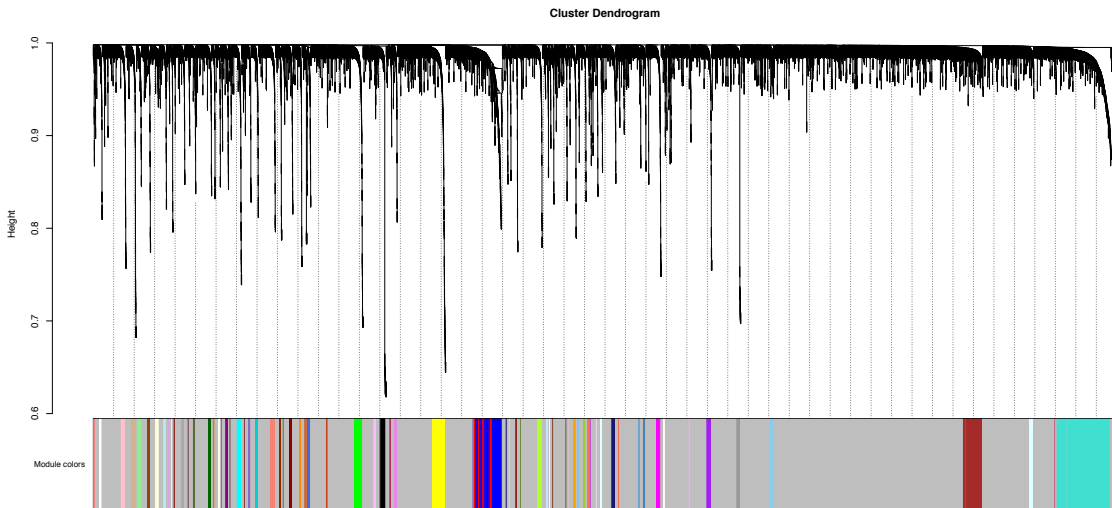


Fig.S6 WGCNAによる各細胞株でのモジュールの抽出
 (E) H2228細胞でのモジュール抽出。細胞のクラスタリング(上)、powerの設定(中央)、モジュールの抽出(下)。

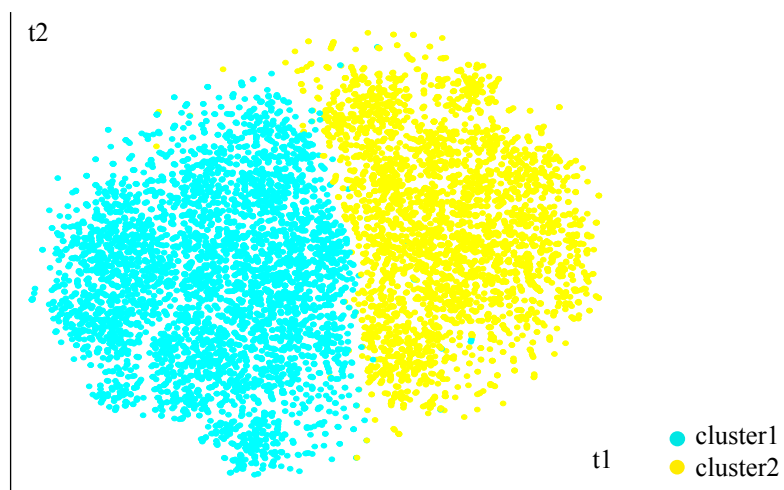


Fig.S6 micro-droplet system PC9細胞のクラスタリング
micro-droplet systemでmodule “magenta”によるクラスタリングの結果、細胞はcluster1(水色)とcluster2(黄色)の2つに分けられた。

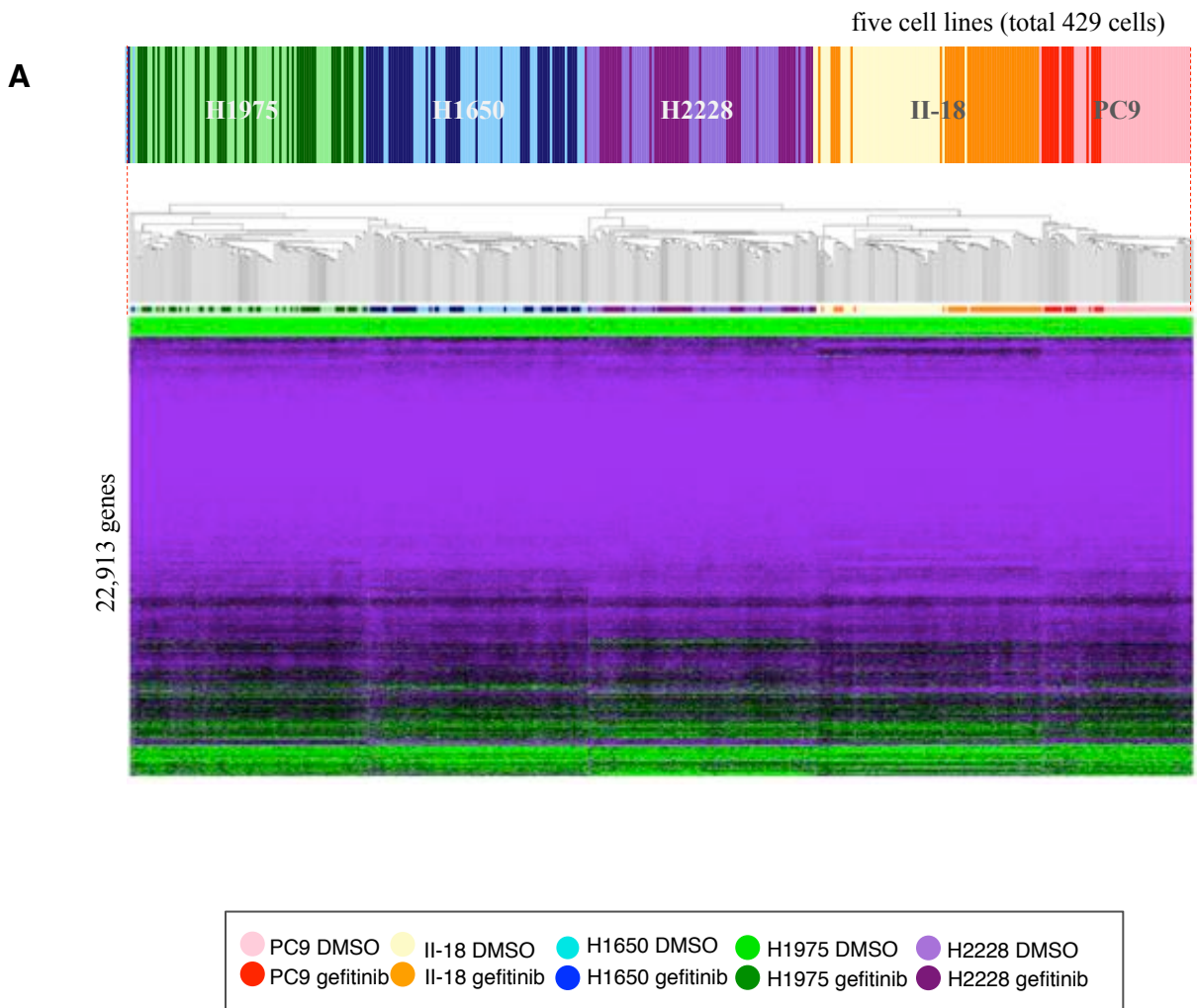


Fig.S7 micro-chamber system 5細胞株の階層クラスタリング

(A)micro-chamber systemで全遺伝子の発現情報を用い、5種類の細胞株のコントロール群とゲフィチニブ群を全てまとめてクラスタリング。

B

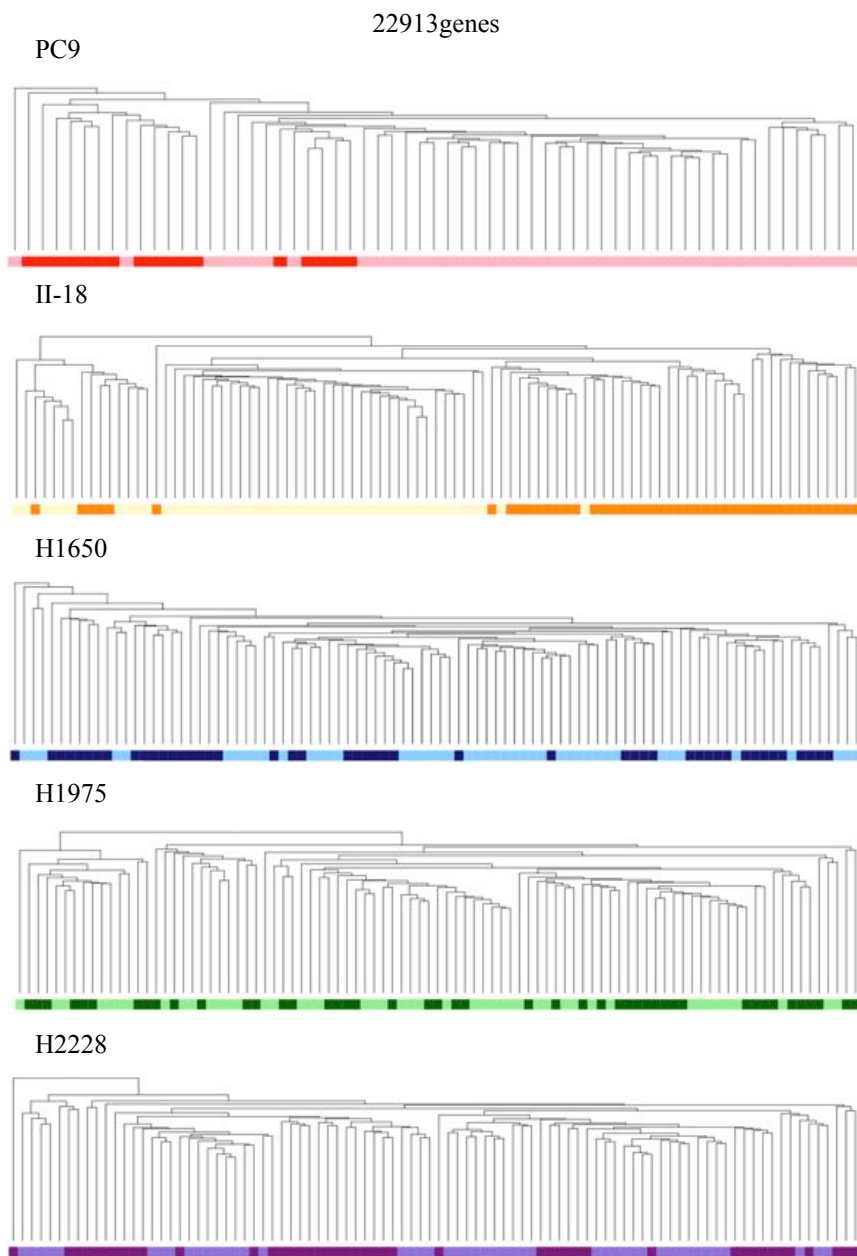


Fig.S7 micro-chamber system 5細胞株の階層クラスタリング

(B)micro-chamber systemで全遺伝子の発現情報を用い、5種類の細胞株のコントロール群とゲフィチニブ群を細胞株ごとにクラスタリング。

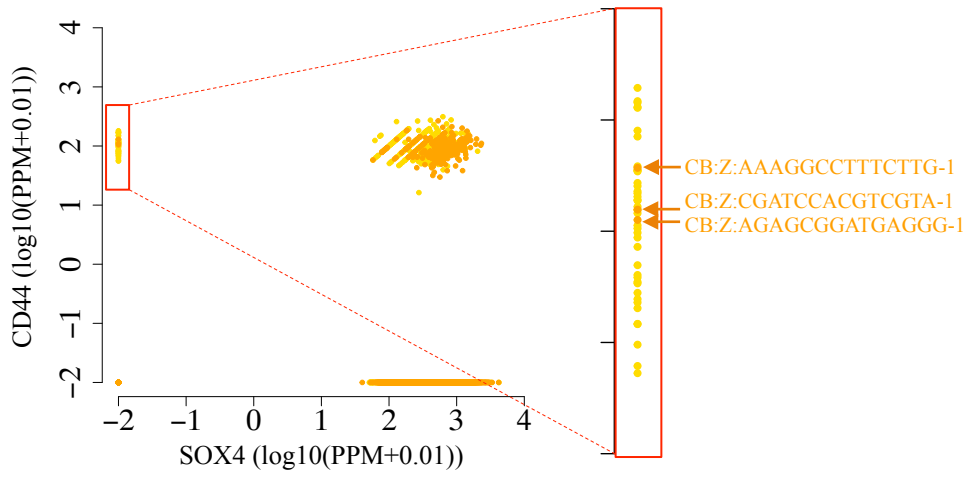


Fig.S8 micro-droplet system II-18細胞でのSOX4とCD44の発現
 micro-droplet systemでのII-18細胞コントロール群とゲフィチニブ群の各細胞でのSOX4(x-axis)とCD44(y-axis)の発現
 レベル。黄色がコントロール群、オレンジ色がゲフィチニブ群を示す。矢印は、外れ細胞を示す。

Supplemental Tables

Supplemental Table S1, PCR sister除去後のリード数とその割合

		Total raw reads	Reads without PCR sisters	%
PC9	DMSO	127631437	86320712	68
	gefitinib	170913768	73809216	43
II-18	DMSO	173605651	82652598	48
	gefitinib	164734523	98278555	60
H1650	DMSO	100094271	68329141	68
	gefitinib	111008829	76768723	69
H1975	DMSO	161777133	70871586	44
	gefitinib	173096952	115455843	67
H2228	DMSO	176666104	96259797	54
	gefitinib	149011834	67743954	45
		average	83649013	57

Supplemental Table S2, 5細胞株のPCAによるクラスタリングに用いた127SMGs遺伝子リスト

group_id	symbol	cellular processes	group_id	symbol	cellular processes
15780	VHL	Transcription factor/regulator	6698	KRAS	MAPK signaling
11582	GATA3	Transcription factor/regulator	3380	NF1	MAPK signaling
750	TSHZ3	Transcription factor/regulator	3091	MAP3K1	MAPK signaling
2425	EP300	Transcription factor/regulator	2763	BRAF	MAPK signaling
14934	CTCF	Transcription factor/regulator	21	NRAS	MAPK signaling
10569	TAF1	Transcription factor/regulator	962	MAP2K4	MAPK signaling
2057	TSHZ2	Transcription factor/regulator	9160	MAPK8IP1	MAPK signaling
5925	RUNX1	Transcription factor/regulator	6196	PIK3CA	PI(3)K signaling
15078	MECOM	Transcription factor/regulator	2176	PTEN	PI(3)K signaling
12139	TBX3	Transcription factor/regulator	7229	PIK3R1	PI(3)K signaling
10387	SIN3A	Transcription factor/regulator	4321	TLR4	PI(3)K signaling
3514	WT1	Transcription factor/regulator	5418	PIK3CG	PI(3)K signaling
5407	EIF4A2	Transcription factor/regulator	263	AKT1	PI(3)K signaling
4302	FOXA1	Transcription factor/regulator	1435	SMAD4	TGF- β signaling
2703	PHF6	Transcription factor/regulator	5665	TGFBR2	TGF- β signaling
8619	CBFB	Transcription factor/regulator	14533	ACVR1B	TGF- β signaling
17119	SOX9	Transcription factor/regulator	9481	SMAD2	TGF- β signaling
11323	ELF3	Transcription factor/regulator	10005	ACVR2A	TGF- β signaling
9263	VEZF1	Transcription factor/regulator	8959	APC	Wnt/ β -catenin signaling
3535	CEBPA	Transcription factor/regulator	15479	CTNNB1	Wnt/ β -catenin signaling
14411	FOXA2	Transcription factor/regulator	14569	AXIN2	Wnt/ β -catenin signaling
14629	MLL3	Histone modifier	11268	TBL1XR1	Wnt/ β -catenin signaling
2450	MLL2	Histone modifier	5016	SOX17	Wnt/ β -catenin signaling
6259	ARID1A	Histone modifier	7041	HIST1H1C	Histone
4921	PBRM1	Histone modifier	17525	H3F3C	Histone
13212	SETD2	Histone modifier	2288	HIST1H2BD	Histone
2495	NSD1	Histone modifier	17014	FBXW7	Proteolysis
13785	SETBP1	Histone modifier	6619	KEAP1	Proteolysis
3547	KDM5C	Histone modifier	7437	SPOP	Proteolysis
6927	KDM6A	Histone modifier	15136	U2AF1	Splicing
8873	MLL4	Histone modifier	10460	SF3B1	Splicing
15388	ARID5B	Histone modifier	12086	PCBP1	Splicing
12304	ASXL1	Histone modifier	8391	CDH1	HIPPO signaling
11980	EZH2	Histone modifier	4334	AJUBA	HIPPO signaling
3772	TP53	Genome integrity	966	DNMT3A	DNA methylation
14586	ATM	Genome integrity	6061	TET2	DNA methylation
2811	ATRX	Genome integrity	18428	IDH1	Metabolism
14757	BRCA2	Genome integrity	3927	IDH2	Metabolism
7304	ATR	Genome integrity	6922	NFE2L2	NFE2L
552	STAG2	Genome integrity	12129	NFE2L3	NFE2L
11636	BAP1	Genome integrity	12950	PPP2R1A	Protein phosphatase
1909	BRCA1	Genome integrity	8706	PTPN11	Protein phosphatase
8980	SMC1A	Genome integrity	3212	RPL22	Ribosome
14209	SMC3	Genome integrity	2557	RPL5	Ribosome
8773	CHEK2	Genome integrity	11217	MTOR	TOR signaling
1402	RAD21	Genome integrity	4039	STK11	TOR signaling
17511	ERCC2	Genome integrity	12262	NAV3	Other
17604	EGFR	RTK signaling	14014	NOTCH1	Other
15676	FLT3	RTK signaling	13036	LRRK2	Other
4528	EPHA3	RTK signaling	25173	MALAT1	Other
18334	ERBB4	RTK signaling	7657	ARHGAP35	Other
6865	PDGFRA	RTK signaling	10485	POLQ	Other
3398	EPHB6	RTK signaling	2713	NCOR1	Other
7547	FGFR2	RTK signaling	1777	USP9X	Other
11079	KIT	RTK signaling	1200	NPM1	Other
15888	FGFR3	RTK signaling	7369	HGF	Other
12210	CDKN2A	Cell cycle	12547	EPPK1	Other
11469	RB1	Cell cycle	7760	AR	Other
2739	CDK12	Cell cycle	2300	LIFR	Other
3326	CDKN1B	Cell cycle	7322	PRX	Other
5431	CCND1	Cell cycle	15580	CRIPAK	Other
10831	CDKN1A	Cell cycle	12800	EGR3	Other
13024	CDKN2C	Cell cycle	5724	B4GALT3	Other
			25281	MIR142	Other

Supplemental Table S3,細胞周期の解析に使用された遺伝子リスト
(A) micro-chamber systemの細胞周期の推定に利用された遺伝子

Symbol	gene name
CCNE1	cyclin E1
E2F1	E2F transcription factor 1
CDC6	cell division cycle 6
PCNA	proliferating cell nuclear antigen
RFC4	replication factor C (activator 1) 4, 37kDa
DHFR	dihydrofolate reductase
RRM2	ribonucleotide reductase M2
RAD51	RAD51 recombinase
CDK1	cyclin-dependent kinase 1
TOP2A	topoisomerase (DNA) II alpha 170kDa
CCNF	cyclin F
CCNA2	cyclin A2
AURKA	aurora kinase A
BUB1	BUB1 mitotic checkpoint serine/threonine kinase
CCNB1	cyclin B1
PLK1	polo-like kinase 1
PTTG1	pituitary tumor-transforming 1
RAD21	RAD21 homolog (S. pombe)
VEGFC	vascular endothelial growth factor C
CDKN3	cyclin-dependent kinase inhibitor 3

**Supplemental Table S3.細胞周期の解析に使用された遺伝子リスト
(B) micro-droplet systemの細胞周期の推定に利用された遺伝子**

phase	Symbol	R	phase	Symbol	R	phase	Symbol	R
G1/S	ACD	0.2084462	G2/M	ANLN	0.2723649	M	AKIRIN2	0.2031315
G1/S	APEX2	0.2148424	G2/M	AURKB	0.5321099	M	ANLN	0.2958071
G1/S	CCNE1	0.2663608	G2/M	BORA	0.2714558	M	ANP32B	0.2541205
G1/S	CCNE2	0.3252323	G2/M	BUB3	0.2748968	M	ANP32E	0.255035
G1/S	CDC6	0.3118351	G2/M	CCNA2	0.4965628	M	ARL6IP1	0.628574
G1/S	CDCA7L	0.2902344	G2/M	CCNF	0.2104578	M	AURKA	0.6092834
G1/S	CHAF1A	0.2369656	G2/M	CDCA2	0.2311073	M	BIRC5	0.3754414
G1/S	CLSPN	0.3776165	G2/M	CDCA3	0.400363	M	BUB1	0.2367238
G1/S	E2F1	0.2836067	G2/M	CDCA8	0.2783313	M	CCNA2	0.5522542
G1/S	GINS2	0.4020135	G2/M	CDK1	0.5821498	M	CCNB2	0.5141196
G1/S	GMNN	0.4269952	G2/M	CDKN2C	0.2262832	M	CDC20	0.6619391
G1/S	INTS8	0.2561322	G2/M	CKAP2	0.3141243	M	CDC25B	0.2580322
G1/S	LUC7L3	0.2050314	G2/M	CKAP2L	0.2641564	M	CDCA3	0.4926722
G1/S	MCM2	0.2240236	G2/M	FAM83D	0.5195265	M	CENPA	0.476688
G1/S	MCM5	0.2868178	G2/M	GAS2L3	0.2245058	M	CENPE	0.4081029
G1/S	MCM6	0.2503445	G2/M	H2AFX	0.2285597	M	CENPF	0.5463907
G1/S	NASP	0.3089949	G2/M	HJURP	0.4282288	M	CKAP2	0.3407493
G1/S	PCNA	0.5877051	G2/M	HMGB2	0.6292196	M	CKS1B	0.2007793
G1/S	POLD3	0.2613784	G2/M	KIF5B	0.3209594	M	CKS2	0.49405
G1/S	SLBP	0.5079985	G2/M	KIF20B	0.2873333	M	DEPDC1	0.475011
G1/S	UNG	0.2762654	G2/M	KIF22	0.2844896	M	DEPDC1B	0.2824315
S	ASF1B	0.241012	G2/M	KIF23	0.332756	M	DLGAP5	0.2684237
S	CDC45	0.2327182	G2/M	KIFC1	0.2672676	M	DNAJA1	0.2247345
S	CENPM	0.2157942	G2/M	KPNA2	0.6223326	M	FAM64A	0.4000655
S	ENOSF1	0.4931348	G2/M	MAD2L1	0.2339825	M	GAS2L3	0.229844
S	FEN1	0.2346011	G2/M	MALAT1	0.409941	M	GTSE1	0.3815512
S	GCLM	0.2522648	G2/M	MALAT1	0.409941	M	HMGB3	0.3296062
S	H1FO	0.2029316	G2/M	NDC80	0.3773631	M	HMMR	0.525004
S	HIST1H4C	0.36078	G2/M	NUCKS1	0.2874382	M	HN1	0.3157104
S	RRM2	0.3137284	G2/M	NUSAP1	0.4998711	M	HSPA8	0.4904216
S	RSRC2	0.2007431	G2/M	PIF1	0.2399757	M	KIF2C	0.310216
S	TOP2A	0.2312718	G2/M	PSMD11	0.2376324	M	KIF5B	0.3977087
S	TYMS	0.4957512	G2/M	PSRC1	0.3390074	M	KIF20B	0.3202733
S	UBE2T	0.3587391	G2/M	RCCD1	0.2086352	M	LBR	0.2530543
S	ZWINT	0.4421044	G2/M	SAP30	0.2066004	M	MK167	0.3967232
			G2/M	SMC4	0.449287	M	MZT1	0.2896028
			G2/M	TOP2A	0.4822954	M	NEK2	0.4019897
			G2/M	TUBB	0.4278037	M	NUF2	0.407485
			G2/M	TUBB4B	0.4368572	M	NUSAP1	0.5026313
			G2/M	UBE2C	0.6418814	M	PBK	0.2675716
						M	PLK1	0.6684096
						M	PRR11	0.2214031
						M	RBM8A	0.2013137
						M	SFPQ	0.2199022
						M	SGOL2	0.4225035
						M	SRSF3	0.2975546
						M	TACC3	0.2837015
						M	THRAP3	0.2032257
						M	TPX2	0.5248069
						M	TTK	0.2213745
						M	TUBB4B	0.5222375
						M-G1	ANP32E	0.259491
						M-G1	CDKN3	0.47832
						M-G1	DKC1	0.2441299
						M-G1	DYNLL1	0.3258742
						M-G1	GRPEL1	0.2068026
						M-G1	HSPA8	0.4251107
						M-G1	ILF2	0.3342597
						M-G1	KIF5B	0.2619027
						M-G1	LYAR	0.2317684
						M-G1	MORF4L2	0.3570747
						M-G1	MRPS18B	0.2187978
						M-G1	NUCKS1	0.3798808
						M-G1	NUP37	0.2350708
						M-G1	PBK	0.2099027
						M-G1	PRC1	0.2334793
						M-G1	PTTG1	0.6069561
						M-G1	RAD21	0.258954
						M-G1	RAN	0.4816124
						M-G1	SRSF3	0.3415955

Supplemental Table S4, PC9のモジュール“lightsteelblue1”

(A) PC9のモジュール“lightsteelblue1”のリスト

gene symbol	NCBI gene id
CENPA	NM_001042426,NM_001809
SGOL2	NM_001160033,NM_001160046,NM_152524
CKAP2L	NM_001304361,NM_152515
RAD21	NM_006265
CNTRL	NM_007018
STIL	NM_001048166,NM_001282936,NM_001282937,NM_001282938,NM_001282939,NM_003035
POLH	NM_001291969,NM_001291970,NM_006502
GPSM2	NM_013296
TTK	NM_001166691,NM_003318
NUF2	NM_031423,NM_145697
LARGE	NM_004737,NM_133642
HMMR	NM_001142556,NM_001142557,NM_012484,NM_012485
KIFC1	NM_002263
AURKA	NM_003600,NM_198433,NM_198434,NM_198435,NM_198436,NM_198437
BUB1	NM_001278616,NM_001278617,NM_004336
PIF1	NM_001286496,NM_001286497,NM_001286499,NM_025049
TROAP	NM_001100620,NM_001278324,NM_005480
TRAIP	NM_005879
IFNAR1	NM_000629
DEPDC1	NM_001114120,NM_017779
SPDL1	NM_017785
NCAPG	NM_022346
ERN2	NM_033266
KNSTRN	NM_001142761,NM_001142762,NM_033286
CKAP2	NM_001098525,NM_001286686,NM_001286687,NM_018204
FAM72B	NM_001100910
CENPE	NM_001286734,NM_001813
HACD4	NM_001010915
KIF14	NM_014875
CDC25C	NM_001287582,NM_001287583,NM_001790,NM_022809
CDCA2	NM_152562
RNF26	NM_032015
PSRC1	NM_001005290,NM_001032291,NM_032636
GAS2L3	NM_001303130,NM_001303131,NM_174942
KIF15	NM_020242
ESPL1	NM_012291
AURKAPS1	NR_001587
HMMR-AS1	NR_109892

(B) モジュール“lightsteelblue1”におけるGO解析の結果

rank	enrichmentP	BonferoniP	termID	termOntology	termName
1	6.42E-16	9.45E-12	GO:0022402	BP	cell cycle process
2	4.05E-15	5.95E-11	GO:0000278	BP	mitotic cell cycle
3	4.81E-15	7.08E-11	GO:0007067	BP	mitotic nuclear division
4	5.98E-15	8.79E-11	GO:0007059	BP	chromosome segregation
5	6.21E-15	9.13E-11	GO:0051301	BP	cell division
6	2.80E-13	4.12E-09	GO:0007049	BP	cell cycle
7	4.95E-12	7.28E-08	GO:0007017	BP	microtubule-based process
8	7.64E-12	1.12E-07	GO:0000819	BP	sister chromatid segregation
9	4.02E-11	5.91E-07	GO:0005819	CC	spindle
10	1.68E-10	2.47E-06	GO:0015630	CC	microtubule cytoskeleton

**Supplemental Table S5, PC9のモジュール “magenta”
(A) PC9 のモジュール“magenta”のリスト**

Gene Symbol	NCBI gene id	Gene Symbol	NCBI gene id
COG3	NM_031431	HIST1H2AJ	NM_021066
ZNF845	NM_138374	TULP2	NM_003323
PLD1	NM_001130081,NM_002662	SBNO2	NM_001100122,NM_014963
DUSP1	NM_004417	ID4	NM_001546
RGCC	NM_014059	VAMP4	NM_001185127,NM_003762
TRIM52	NM_032765	TMEM65	NM_194291
HMOX1	NM_002133		
	NM_000325,NM_001204397,NM_001204398,NM_001204399,NM_153426,NM_153427	TRPV4	NM_001177428,NM_001177431,NM_001177433,NM_021625,NM_147204
PITX2			
GDPD1	NM_001165993,NM_001165994,NM_182569	DCUN1D4	NM_001040402,NM_001287755,NM_001287757,NM_015115
RNF2	NM_007212	WDSUB1	NM_001128212,NM_001128213,NM_152528
WDR78	NM_024763,NM_207014	PUS10	NM_144709
JMY	NM_152405	AOC2	NM_001158,NM_009590
ELF1	NM_001145353,NM_172373	PKP1	NM_000299,NM_001005337
	NM_001111018,NM_001111019,NM_001244963,NM_145117,NM_182964	IL6	NM_000600
NAV2		LRRC69	NM_001129890
ASAP3	NM_001143778,NM_017707	SLC2A10	NM_030777
ASB7	NM_024708,NM_198243	COL17A1	NM_000494
ATXN7	NM_000333,NM_001128149,NM_001177387	AMIGO3	NM_198722
PSG8	NM_001130167,NM_001130168,NM_182707	IFNE	NM_176891
PARP16	NM_017851	ATP6AP1L	NM_001017971
ACTG1	NM_001199954,NM_001614		
PRR34	NM_018280	RARA	NM_000964,NM_001024809,NM_001145301,NM_001145302
SNAI2	NM_003068	MYLIP	NM_013262
C19orf73	NM_018111		
CCDC153	NM_001145018	RNF38	NM_022781,NM_194328,NM_194329,NM_194330,NM_194332
	NM_001167827,NM_001271943,NM_001271944,NM_001271945,NM_007162	ANKRD65	NM_001145210,NM_001243535,NM_001243536
TFEB		SCRN3	NM_001193528,NM_024583
	NM_001042471,NM_001165958,NM_001165959,NM_018530	GLP2R	NM_004246
GSDMB		SOCS6	NM_004232
	NM_001006636,NM_001164629,NM_001284233,NM_001284234,NM_001284235,NM_001284238,NM_024659	AQP11	NM_173039
GTDC1		FBXO32	NM_001242463,NM_058229,NM_148177
TMEM14C	NM_001165258,NM_016462	IRGQ	NM_001007561
RAB42	NM_001193532,NM_152304		
THBD	NM_000361	MAST4	NM_001164664,NM_001290226,NM_001290227,NM_001290228,NM_001297651,NM_015183,NM_198828
KPNA5	NM_002269	IGIP	NM_001007189
PLCE1	NM_001165979,NM_001288989,NM_016341	ACAD10	NM_001136538,NM_025247
CFHR1	NM_002113		
KBTBD6	NM_152903	ZNF248	NM_001267597,NM_001267605,NM_001267606,NM_001267607,NM_021045
KRTAP2-1	NM_001123387	GPR25	NM_005298
OSTM1	NM_014028	ZFP36	NM_003407
CSRP3	NM_003476	TRAPP8	NM_014939
HDAC10	NM_001159286,NM_032019	UPRT	NM_145052
	NM_001009991,NM_001242384,NM_001242394,NM_001242395	CELSR2	NM_001408
SYTL3		FOXJ2	NM_018416
	NM_001030287,NM_001040619,NM_001206484,NM_001206486,NM_001206488,NM_001674	TMEM81	NM_203376
ATF3		LOC100130331	NR_027247
ARHGEF37	NM_001001669	ZNF197-AS1	NR_046658
ANO8	NM_020959	LOC155060	NR_036573
HIST1H2AI	NM_003509	LINC00511	NR_033876
CFP	NM_001145252,NM_002621	LOC100126784	NR_015384
EPM2AIP1	NM_014805	LOC101927571	NR_110782
SAMD4A	NM_001161576,NM_001161577,NM_015589	MIR6792	NR_106850
KRTAP2-4	NM_033184	DKFZP434I0714	NR_033797
KRTAP2-2	NM_033032	LOC283440	NR_033958
TSSK2	NM_053006	PPP1R26-AS1	NR_038969
PTGER4	NM_000958	LOC100379224	NR_033341
CCNT1	NM_001240,NM_001277842	PAN3-AS1	NR_029383
FAM209A	NM_001012971	LOC101927267	NR_120449
	NM_001143676,NM_001143677,NM_001143678,NM_001291995,NM_005627	PIK3IP1-AS1	NR_110542
SGK1		LINC01006	NR_103858
CCDC18	NM_206886	MIR6880	NR_106940
GAS1	NM_002048	MIR1231	NR_031599
TDRP	NM_001256113,NM_175075	LOC102724000	NR_121615
RND3	NM_001254738,NM_005168	ACTG1P20	NR_033926
RHOB	NM_004040	AGBL5-AS1	NR_046730
KRTAP2-3	NM_001165252		
CLEC2B	NM_005127		
SLC22A20	NM_001004326		
SPTBN2	NM_006946		
TNFRSF10D	NM_003840		
CDK18	NM_002596,NM_212502,NM_212503		
NFIL3	NM_001289999,NM_001290000,NM_005384		

(B) モジュール“magenta”のGO解析の結果

rank	enrichmentP	BonferoniP	termID	termOntology	termName
1	9.01E-06	0.132589083	GO:0006959	BP	humoral immune response
2	0.00013173	1	GO:0002366	BP	leukocyte activation involved in immune response
3	0.000143901	1	GO:0002286	BP	T cell activation involved in immune response
4	0.000225331	1	GO:0050710	BP	negative regulation of cytokine secretion
5	0.000393841	1	GO:0045630	BP	positive regulation of T-helper 2 cell differentiation RNA polymerase II core promoter proximal region sequence-specific DNA binding transcription factor activity involved in negative regulation of transcription
6	0.000445651	1	GO:0001078	MF	RNA polymerase II core promoter proximal region sequence-specific DNA binding transcription factor activity
7	0.000463626	1	GO:0000982	MF	activity
8	0.000529063	1	GO:0008081	MF	phosphoric diester hydrolase activity
9	0.000548049	1	GO:0006956	BP	complement activation
10	0.000548049	1	GO:0042093	BP	T-helper cell differentiation

Supplemental Table S6, II-18細胞 モジュール“red”のリスト

Gene symbol	NCBI gene id	Gene symbol	NCBI gene id
CCNO	NM_021147	PLEKHB1	NM_001130033,NM_001130034,NM_001130035,NM_00130036,NM_021200
AKR1A1	NM_001202413,NM_001202414,NM_006066,NM_153326	HLA-F	NM_001098478,NM_001098479,NM_018950
ATP6V0E1	NM_003945	SRI	NM_001256891,NM_001256892,NM_003130,NM_198901
CDKN2B	NM_004936,NM_078487	TMED9	NM_017510
ASAH1	NM_001127505,NM_004315,NM_177924	PDCD4	NM_001199492,NM_014456,NM_145341
CSF1	NM_000757,NM_172210,NM_172211,NM_172212	SC5D	NM_001024956,NM_006918
SOX4	NM_003107	BET1	NM_005868
HIST2H2BE	NM_003528		NM_001256486,NM_001256487,NM_001256488,NM_004487
ALG1L	NM_001015050,NM_001195223	GOLGB1	
DSTN	NM_001011546,NM_006870	TTC3	NM_001001894,NM_003316
HLA-A	NM_001242758,NM_002116	OAZ2	NM_001301302,NM_002537
BMF	NM_001003940,NM_001003942,NM_001003943,NM_033503	ELF3	NM_001114309,NM_004433
HIST1H2BD	NM_021063,NM_138720	SCARB2	NM_001204255,NM_005506
	NM_001256658,NM_001256659,NM_001256660,NM_001256661,NM_001256662,NM_003598	IFT57	NM_018010
TEAD2	NM_001256662,NM_003598	ST13	NM_001278589,NM_003932
ANXA4	NM_001153	LRIG1	NM_015541
GOLGA2	NM_004486	ATP6AP1	NM_001183
MXD4	NM_006454		NM_000694,NM_001030010,NM_001161473,NM_001290058,NM_001290059
	NM_001258320,NM_001258321,NM_001258322,NM_001258323,NM_001258324,NM_006034	ALDH3B1	
TP53I11	NM_001113182,NM_001199455,NM_001199456,NM_001291986,NM_005104	PEL1	NM_020651
BRD2	NM_024552	CREB3	NM_006368
CERS4	NM_001243042,NM_002117	SQLE	NM_003129
HLA-C	NM_002253	COPA	NM_001098398,NM_004371
KDR	NM_004048	MIR205HG	NM_001104548
B2M	NM_001018070,NM_020441	NICN1	NM_032316
CORO1B	NM_004344	CBLB	NM_170662
CETN2	NM_001017369,NM_006745	UBE2R2	NM_017811
MSMO1	NM_004428,NM_182685	MAGED1	NM_001005332,NM_001005333,NM_006986
EFNA1	NM_001277163,NM_001815	STEAP4	NM_001205315,NM_001205316,NM_024636
CEACAM3	NM_001302777,NM_004332	KDELRL1	NM_006801
BPHL	NM_004888		NM_001033518,NM_001033519,NM_001033520,NM_001278299,NM_015610,NM_016003
ATP6V1G1	NM_001134335,NM_014413	WIPI2	NM_001098272,NM_002130
EIF2AK1	NM_006315	HMGCS1	NM_005542,NM_198336,NM_198337
PCGF3	NM_174921	INSIG1	NM_002127
SMIM14	NM_032102	HLA-G	NM_001142339,NM_001261443,NM_001261444,NM_182978
SRSF8	NM_001031738	GNAL	
TMEM150A	NM_001083613,NM_194280	NARF	NM_001038618,NM_001083608,NM_012336,NM_031968
TMEM219	NM_001136020,NM_001276478,NM_004968,NM_022307	SCD	NM_005063
ICA1	NM_014718		NM_001040097,NM_001040098,NM_001040099,NM_023948
CLSTN3	NM_015966,NM_198398	MOSPD3	
ERGIC3	NM_021999	KLHL24	NM_017644
ITM2B	NM_001009923,NM_001009924,NM_001009925,NM_014145	SMPDL3B	NM_001009568,NM_001304579,NM_014474
TMEM230	NM_152261	TKT	NM_001064,NM_001135055,NM_001258028
TMEM263	NM_015388	HBP1	NM_001244262,NM_012257
YIPF3	NM_004872	CAPN1	NM_001198868,NM_001198869,NM_005186
TMEM59	NM_004892	UBXN1	NM_001286077,NM_001286078,NM_015853
SEC22B	NM_001305	PBXIP1	NM_020524
CLDN4	NM_012477	DPP3	NM_001256670,NM_005700,NM_130443
WBP1	NM_001660	DNAJC4	NM_005528
ARF4	NM_138389		NM_001127399,NM_001127400,NM_001127401,NM_016061
FAM114A1	NM_001291484,NM_004363	YPEL5	
CEACAM5	NM_001001438,NM_001145436,NM_001145437,NM_002340	TMPRSS2	NM_001135099,NM_005656
LSS	NM_020412	SEC24C	NM_004922,NM_198597
CHMP1B	NM_001287742,NM_001287743,NM_001287744,NM_001287745,NM_001287747,NM_001287748,NM_001287749,NM_001287750,NM_001287751,NM_001287756,NM_004462	PARP14	NM_017554
FDFT1	NM_004657	PPAPDC1B	NM_001102559,NM_001102560,NM_032483
SDPR	NM_001512	LITAF	NM_001136472,NM_001136473,NM_004862
GSTA4	NM_022121	RPL4	NM_000968
PERP	NM_003041	ST8SIA4	NM_005668,NM_175052
SLC5A2	NM_025048,NM_153840	C16orf58	NM_022744
GPR110	NM_005514	NIPSNAP1	NM_001202502,NM_003634
HLA-B	NM_001007253	PON2	NM_000305,NM_001018161
ERV3-1	NM_014390	LOC102723885	NR_120517
SND1	NM_005125	LOC101927043	NR_110207,NR_110208
CCS	NM_182757	CIRBP-AS1	NR_027271
RNF144B	NM_004354	ZNF528-AS1	NR_125345
CCNG2		MIR4800	NR_039964
		LOC440173	NR_027471
		HLA-J	NR_024240
		HLA-H	NR_001434
		ST13P4	NR_002183
		UGDH-AS1	NR_047679
		ZNRD1-AS1	NR_026751
		INO80B-WBP1	NR_037849

データ公開

本研究に用いられたシーケンスデータは、DNA Data Bank of Japan(DDBJ)にアクセッションナンバーDRA005922- DRA005929 で登録済みである。

論文目録

Combinatory use of distinct single-cell RNA-seq analytical platforms reveals the heterogeneous transcriptome response

Yukie Kashima, Ayako Suzuki, Ying Liu, Masahito Hosokawa, Hiroko Matsunaga, Masataka Shirai, Koji Arikawa, Sumio Sugano, Takashi Kohno, Haruko Takeyama, Katsuya Tsuchihara, and Yutaka Suzuki

Scientific Reports (in press)

参考文献

1. Wang, Y. & Navin, N. E. Advances and Applications of Single-Cell Sequencing Technologies. *Mol. Cell* **58**, 598–609 (2015).
2. Matsunaga, H. *et al.* A highly sensitive and accurate gene expression analysis by sequencing ('bead-seq') for a single cell. *Anal. Biochem.* **471**, 9–16 (2015).
3. Wu, A. R. *et al.* Quantitative assessment of single-cell RNA-sequencing methods. *Nat. Methods* **11**, 41–46 (2013).
4. Macosko, E. Z. *et al.* Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
5. Klein, A. M. *et al.* Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201 (2015).
6. Svensson, V. *et al.* Power analysis of single-cell RNA-sequencing experiments. *Nat. Methods* **14**, 381–387 (2017).
7. Ziegenhain, C. *et al.* Comparative Analysis of Single-Cell RNA Sequencing Methods. *Mol. Cell* **65**, 631–643.e4 (2017).
8. Tirosh, I. *et al.* Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. *Nature* **539**, 309–313 (2016).
9. Ramsköld, D. *et al.* Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol.* **30**, 777–782 (2012).
10. Venteicher, A. S. *et al.* Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-seq. *Science (80-.).* **355**, eaai8478 (2017).
11. Ting, D. T. *et al.* Single-cell RNA sequencing identifies extracellular matrix gene expression by pancreatic circulating tumor cells. *Cell Rep.* **8**, 1905–1918 (2014).
12. Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science (80-.).* **352**, 189–196 (2016).
13. Elenbaas, B. & Weinberg, R. A. Heterotypic Signaling between Epithelial Tumor Cells and Fibroblasts in Carcinoma Formation. *Exp. Cell Res.* **264**, 169–184 (2001).
14. Hu, M. & Polyak, K. Microenvironmental regulation of cancer development. *Curr. Opin. Genet. Dev.* **18**, 27–34 (2008).
15. Quail, D. & Joyce, J. Microenvironmental regulation of tumor progression and metastasis. *Nat. Med.* **19**, 1423–1437 (2013).
16. McGranahan, N. & Swanton, C. Biological and therapeutic impact of intratumor heterogeneity in cancer evolution. *Cancer Cell* **27**, 15–26 (2015).
17. Gawad, C., Koh, W. & Quake, S. R. Single-cell genome sequencing: current state of the

- science. *Nat. Rev. Genet.* **17**, 175–188 (2016).
18. Sharma, S. V., Bell, D. W., Settleman, J. & Haber, D. A. Epidermal growth factor receptor mutations in lung cancer. *Nat. Rev. Cancer* **7**, 169–181 (2007).
 19. Collisson, E. A. *et al.* Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**, 543–550 (2014).
 20. Takahashi, T. *et al.* Genomic and transcriptomic analysis of imatinib resistance in gastrointestinal stromal tumors. *Genes, Chromosom. Cancer* **56**, 303–313 (2017).
 21. Zhang, J. *et al.* International cancer genome consortium data portal—a one-stop shop for cancer genomics data. *Database* **2011**, 1–10 (2011).
 22. Forbes, S. a *et al.* *The Catalogue of Somatic Mutations in Cancer (COSMIC)*. (2009). doi:10.1002/0471142905.hg1011s57.The
 23. Forbes, S. A. *et al.* COSMIC: Exploring the world’s knowledge of somatic mutations in human cancer. *Nucleic Acids Res.* **43**, D805–D811 (2015).
 24. Hudson (Chairperson), T. J. *et al.* International network of cancer genome projects. *Nature* **464**, 993–998 (2010).
 25. Kobayashi, S. *et al.* EGFR Mutation and Resistance of Non–Small-Cell Lung Cancer to Gefitinib. *N. Engl. J. Med.* **352**, 786–792 (2005).
 26. Sos, M. L. *et al.* PTEN Loss Contributes to Erlotinib Resistance in EGFR-Mutant Lung Cancer by Activation of Akt and EGFR. *Cancer Res.* **69**, 3256–3261 (2009).
 27. Pao, W. *et al.* Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain. *PLoS Med.* **2**, 0225–0235 (2005).
 28. Koivunen, J. P. *et al.* EML4-ALK Fusion Gene and Efficacy of an ALK Kinase Inhibitor in Lung Cancer. *Clin. Cancer Res.* **14**, 4275–4283 (2008).
 29. Soda, M. *et al.* Identification of the transforming EML4–ALK fusion gene in non-small-cell lung cancer. *Nature* **448**, 561–566 (2007).
 30. Hata, A. N. *et al.* Tumor cells can follow distinct evolutionary paths to become resistant to epidermal growth factor receptor inhibition. *Nat. Med.* **22**, 262–269 (2016).
 31. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**, 621–8 (2008).
 32. Achim, K. *et al.* High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. *Nat. Biotechnol.* **33**, 503–509 (2015).
 33. Kandoth, C. *et al.* Mutational landscape and significance across 12 major cancer types. *Nature* **502**, 333–339 (2013).
 34. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York,

2009).

35. Whiteside, T. L. The tumor microenvironment and its role in promoting tumor growth. *Oncogene* **27**, 5904–5912 (2008).
36. Warnes, G. R. *et al.* *gplots: Various R Programming Tools for Plotting Data*. (2016).
37. Suzuki, A. *et al.* DBTSS as an integrative platform for transcriptome, epigenome and genome sequence variation data. *Nucleic Acids Res.* **43**, D87–D91 (2015).
38. Cell signaling Technology. MAPK/Erk in Growth and Differentiation Signaling Pathway. Available at:
[https://www.cellsignal.com/common/content/content.jsp?id=pathways-mapk-erk&pathway=MAPK/Erk in Growth and Differentiation](https://www.cellsignal.com/common/content/content.jsp?id=pathways-mapk-erk&pathway=MAPK/Erk%20in%20Growth%20and%20Differentiation).
39. Tibshirani, R. Regression selection and shrinkage via the lasso. *Journal of the Royal Statistical Society B* **58**, 267–288 (1996).
40. Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **33**, (2010).
41. Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
42. Pak, M., Shin, D., Lee, C. & Lee, M. Significance of EpCAM and TROP2 expression in non-small cell lung cancer. *World J. Surg. Oncol.* **10**, 53 (2012).
43. Shvartsur, A. & Bonavida, B. Trop2 and its overexpression in cancers: regulation and clinical/therapeutic implications. *Genes Cancer* **6**, 84–105 (2015).
44. Smith, S. L. *et al.* Overexpression of aurora B kinase (AURKB) in primary non-small cell lung carcinoma is frequent, generally driven from one allele, and correlates with the level of genetic instability. *Br. J. Cancer* **93**, 719–729 (2005).
45. Whitfield, M. L. *et al.* Identification of Genes Periodically Expressed in the Human Cell Cycle and Their Expression in Tumors. *Mol. Biol. Cell* **13**, 1977–2000 (2002).
46. Zhu, Y., Qiu, P. & Ji, Y. TCGA-Assembler: open-source software for retrieving and processing TCGA data. *Nat. Methods* **11**, 599–600 (2014).
47. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
48. Chen, J. *et al.* AURKA upregulation plays a role in fibroblast-reduced gefitinib sensitivity in the NSCLC cell line HCC827. *Oncol. Rep.* **33**, 1860–1866 (2015).
49. Zhong, N. *et al.* Silencing Aurora-A with siRNA inhibits cell proliferation in human lung adenocarcinoma cells. *Int. J. Oncol.* (2016). doi:10.3892/ijo.2016.3605
50. Zhou, X. *et al.* Gefitinib Inhibits the Proliferation of Pancreatic Cancer Cells via Cell Cycle Arrest. *Anat. Rec. Adv. Integr. Anat. Evol. Biol.* **292**, 1122–1127 (2009).

51. Lawan, A., Shi, H., Gatzke, F. & Bennett, A. M. Diversity and specificity of the mitogen-activated protein kinase phosphatase-1 functions. *Cell. Mol. Life Sci.* **70**, 223–237 (2013).
52. Jeffrey, K. L., Camps, M., Rommel, C. & Mackay, C. R. Targeting dual-specificity phosphatases: manipulating MAP kinase signalling and immune responses. *Nat. Rev. Drug Discov.* **6**, 391–403 (2007).
53. Moncho-Amor, V. *et al.* DUSP1/MKP1 promotes angiogenesis, invasion and metastasis in non-small-cell lung cancer. *Oncogene* **30**, 668–678 (2011).
54. Lin, Y. C. *et al.* DUSP1 expression induced by HDAC1 inhibition mediates gefitinib sensitivity in non-small cell lung cancers. *Clin. Cancer Res.* **21**, 428–438 (2015).
55. Kesarwani, M. *et al.* Targeting c-FOS and DUSP1 abrogates intrinsic resistance to tyrosine-kinase inhibitor therapy in BCR-ABL-induced leukemia. *Nat. Med.* **23**, 472–482 (2017).
56. Hur, W. *et al.* SOX4 overexpression regulates the p53-mediated apoptosis in hepatocellular carcinoma: Clinical implication and functional analysis in vitro. *Carcinogenesis* **31**, 1298–1307 (2010).
57. Pan, X. *et al.* Induction of SOX4 by DNA damage is critical for p53 stabilization and function. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 3788–93 (2009).
58. Vervoort, S. J., van Boxtel, R. & Coffey, P. J. The role of SRY-related HMG box transcription factor 4 (SOX4) in tumorigenesis and metastasis: friend or foe? *Oncogene* **32**, 3397–3409 (2013).
59. McFarlane, S. *et al.* CD44 increases the efficiency of distant metastasis of breast cancer. *Oncotarget* **6**, 11465–76 (2015).
60. Du, L. *et al.* CD44 is of functional importance for colorectal cancer stem cells. *Clin. Cancer Res.* **14**, 6751–6760 (2008).
61. Leung, E. L.-H. *et al.* Non-Small Cell Lung Cancer Cells Expressing CD44 Are Enriched for Stem Cell-Like Properties. *PLoS One* **5**, e14062 (2010).
62. McKenna, A. *et al.* The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

謝辞

本研究に際し指導教官としてご指導いただきました鈴木穰先生、菅野純夫先生に深く感謝いたします。

本研究の内容に関し、貴重なコメントとアドバイスをくださった国立がん研究センター鈴木絢子博士、土原先生に深く感謝いたします。

本研究の一細胞解析に関してサポートしてくださいました早稲田大学 竹山春子教授、細川先生、日立研究所 白井博士、松永博士、有川博士、国立がん研究センター 河野先生に感謝いたします。

本研究のシーケンスデータの解析は、東京大学医科学研究所ヒトゲノム解析センタースーパーコンピュータを利用いたしました。

本研究は、創薬等ライフサイエンス研究支援基盤事業、CREST・JST「統一細胞解析のための革新的技術基盤」、新学術領域研究「先進ゲノム支援」、文部科学省プログラム「社会構想マネジメントを先導するグローバルリーダー養成プログラム」による支援を受けて行われました。

研究生生活を助力していただいた菅野研究室、鈴木研究室の皆様、ありがとうございました。