

論文の内容の要旨

論文題目 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

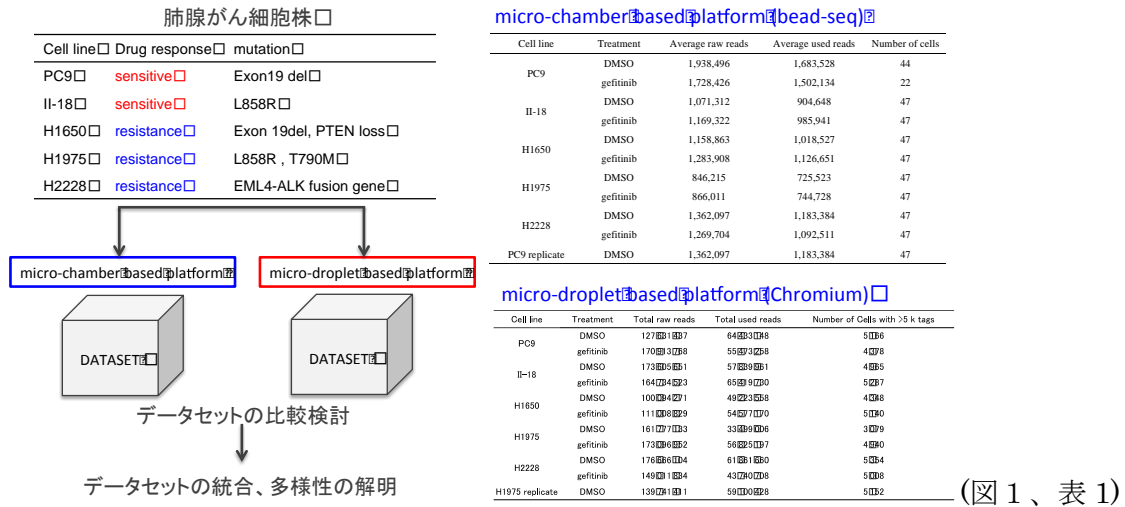
氏名： 鹿島幸恵

背景・目的

発生研究、細菌叢の解析、あるいはがんをはじめとした多くの疾患研究において、一細胞解析が開始されている。発生や分化の過程で生じる細胞ごとの不均一性(heterogeneity)あるいはがん細胞の細胞多様性等、細胞集団（バルク）において一つとして同一な細胞は存在しない。それにも関わらず、その多様性が全体としての集団の振る舞いを決定するのに重要な役割を占めることが、近年の一細胞解析から次第に明らかになってきている。特に知見の集積が著しいのが、in vivo および in vitro でのがん研究の分野である。がんはがん細胞だけでなく、線維芽細胞、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エコシステムを持ち、一細胞 RNA-seq 解析において重要なターゲットである。常に変化が起きているがん微小環境の中に存在し、互いに相互作用している。がん組織の中に存在する細胞の中には、完全に同一なものはない。たとえば、同じクローンのがん細胞といえども、がん進化の中で異なる性質を獲得する。そのため、がんの治療抵抗性や悪性化が起こるメカニズムを理解する上で、がん細胞の多様性を知ることは非常に重要である。しかし、TCGA, ICGC, そして COSMIC といったがんにおける有力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。

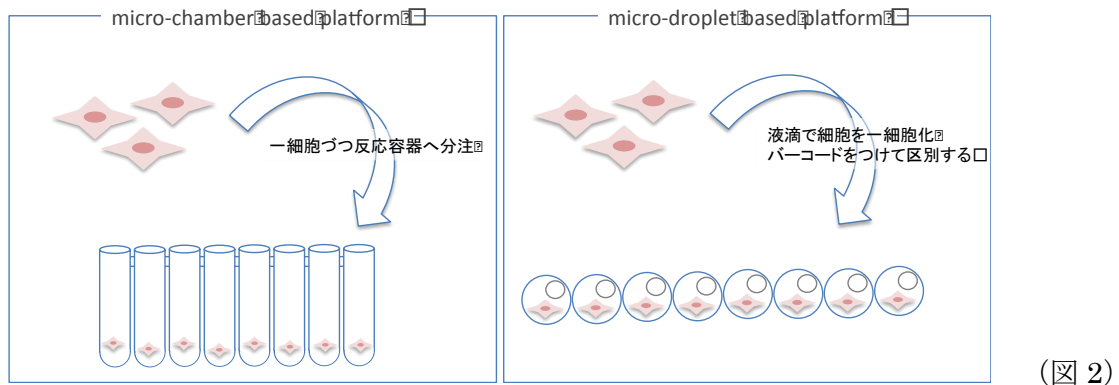
本研究では、肺腺がん細胞株を一細胞解析の対象とし、一細胞 RNA-seq を用いて抗ガン剤の一つであるゲフィチニブに対する反応の多様性の解析を試みた。異なる二つの一細胞 RNA-seq プラットフォームの検証を行った上で、肺腺がん細胞薬剤反応性の一細胞レベルでの解明に取り組んだ。現段階の一細胞 RNA-seq プラットフォームでは、単一のプラットフォームを用いても解析できる細胞数や遺伝子数が限定される。本研究により、異なるプラットフォームを組み合わせることが有力な解析手段となりうることが示唆された。

材料と方法



本研究のサンプルとして、ゲフィチニブに感受性を示す PC9, II-18 と、耐性を示す H1650, H1975, H2228、計 5 種類の肺腺がん細胞を用いた。ゲフィチニブ、あるいはコントロールである DMSO により 24 時間刺激を与えたのちに細胞を回収した。回収された細胞を用い、異なる原理に基づく 2 種類のプラットフォームで、それぞれの定法に従いライブラリを作製した。作製されたライブラリは、illumina 社 HiSeq2500 を用いて解析した(図 1, 表 1)。なお、本実験により得られた一細胞 RNA-seq のデータセットは DDBJ に登録済みである。これらのデータセットを用いて、一細胞 RNA-seq プラットフォームの比較評価、データセットの相互補完と TCGA による臨床データとの整合性を検証した。

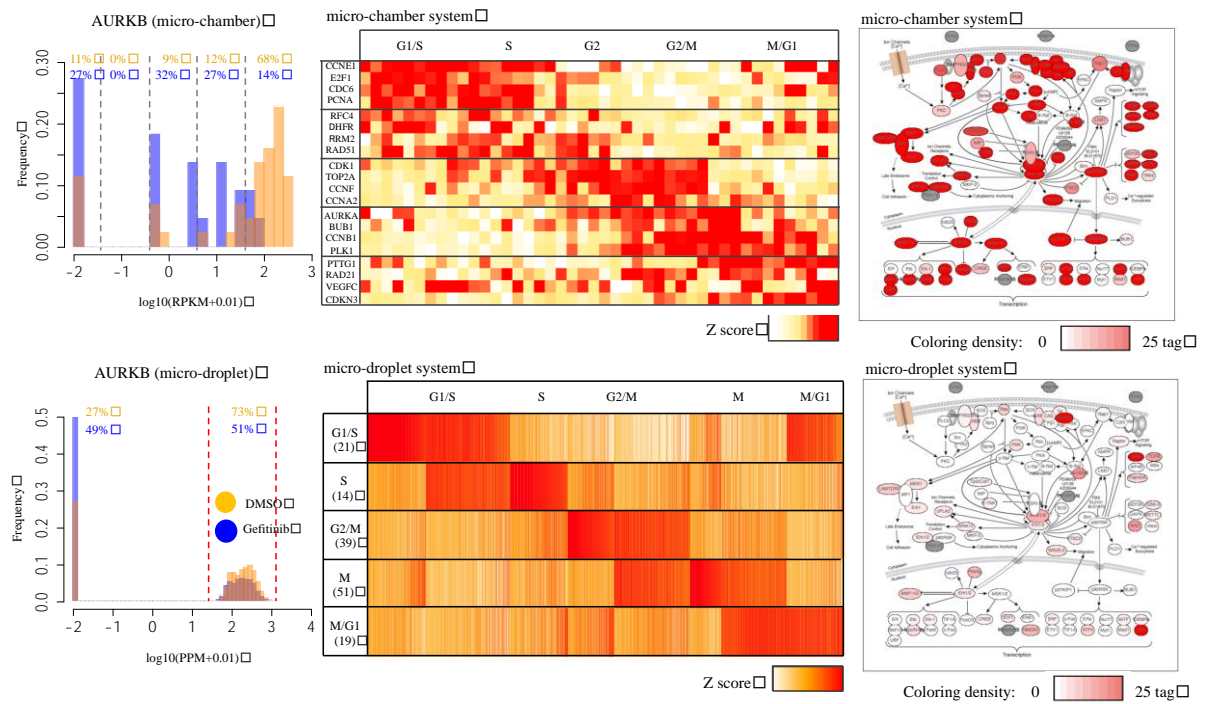
一細胞 RNA-seq プラットフォームの比較評価



一細胞 RNA-seq 解析プラットフォームに焦点を当て、一細胞 RNA-seq プラットフォームの技術評価を行った。一細胞 RNA-seq では、細胞の分離に用いる技術に基づきプラットフォームを 2 つのカテゴリーへと分けることができる。前者は、C1 を含む微小流路、FACS あるいは他のプラットフォームによって細胞を物理的に分離する方法(micro-chamber based system、左図)である。一方、後者は、Chromium に代表される液滴を生成し、個々の細胞にバーコードを付加することで一細胞化を図る方

法(micro-droplet based system、右図)である。ここでは、異なる原理をもつ二つのプラットフォームを比較し、各プラットフォームの利点と弱点を明らかにしようと試みた。

<結果・考察>



(図 3)

	cell number	seq depth	遺伝子数	細胞周期
micro-chamber	~50	~1,000,000 reads/cell	10,661/cell (ave)	20遺伝子で推定
micro-droplet	~5,000	5,000 reads/cell	3,096/cell (ave)	144遺伝子で推定

(表 2)

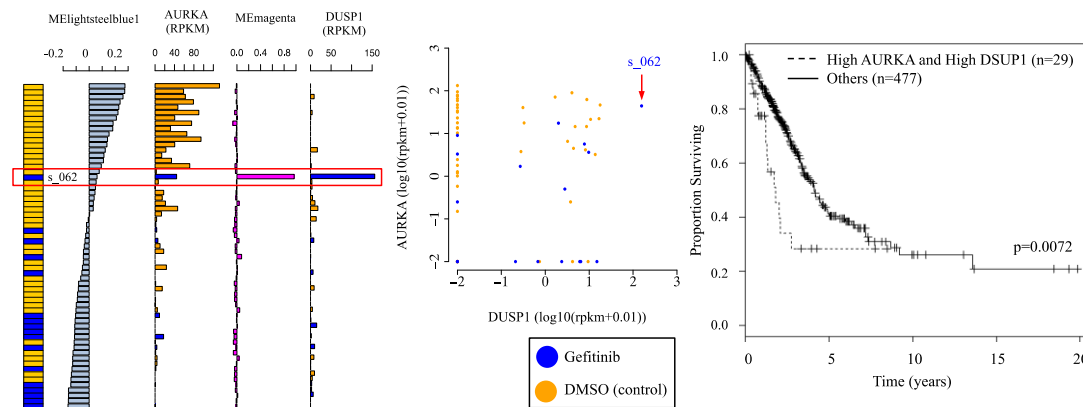
各プラットフォームから 536 細胞 (micro-chamber based system)と 54,631 細胞 (micro-droplet based system)の一細胞 RNA-seq データセットを獲得した。micro-chamber based system は十分なトランスクリプトームカバー率を示すが、一度に解析できる細胞数が限られるため、大きな細胞集団を捉えるのは困難であった。一方、micro-droplet based system にはその逆が当てはまり、トランスクリプトームカバー率は低いが、大規模な集団を処理することが可能であった。また遺伝子によっては、全ての分布を捉えることが困難である可能性が示唆された(図 3、表 2)。また、個別の遺伝子に着目し、欠損するデータの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、欠損値の推定は遺伝子によっては困難であることが明らかにされた。

データセットの相互補完と臨床データとの整合性の検証

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。そこで、以下に共発現ネットワークを用いた肺腺がん細胞株細胞不均一性の解析を試みた。さらに、この解析でがん治療抵抗

性との関与が示唆された DUSP1 と AURKA について、臨床データを用いた検証を行なった。臨床データセットとして、TCGA(The Cancer Genome Atlas)を用いた。

<結果・考察>



(図 4)

micro-chamber based system 由来のデータセットをインプットとして、WGCNA と呼ばれる R パッケージを用いた共発現ネットワークの構築を行った。得られたモジュールの中から、細胞分裂や G2/M 期に関与する遺伝子が多く含まれる”lightsteelblue1”と名付けられたモジュールに着目した。”lightsteelblue1”には、がんに関与することが報告されている AURKA が含まれていた。このモジュールの各細胞の固有値から、ゲフィチニブ刺激をした細胞群では負の値を示す細胞が大部分を占めた。s_062 細胞のみが正の値を示した。この細胞では”magenta”と名付けられたモジュールの固有値が高値を示した。”magenta”にはチロシンキナーゼ阻害への関与が報告されている DUSP1 が含まれていた(図 4:左、中央)。これらのモジュールを micro-droplet based system 由来のデータセットへと写像することで、単独の遺伝子では発現を予測することが困難である場合にも、モジュールを利用して情報を補うことが可能であると考えられた。

TCGA データベースの肺腺がん患者のデータセットを DUSP1 と AURKA の発現に着目し群に分け、それぞれの群における予後を比較した。生存予後に関する Kaplan-Meier 解析から、DUSP1 の発現が高い群、AURKA の発現が高い群はそれぞれの発現が低い群と比べ、有意に予後が不良であることが示された。また、DUSP1 が高発現かつ AURKA が高発現である患者群では、その他の群と比べて予後が悪いことが示された (図 4:右)。共発現ネットワークを介したデータセット補完による解析で得られた結果は、最終的に臨床症例におけるがん細胞多様性を理解するうえで、有用な解析手法であることが示唆された。