



Title	シングルセルRNA - seqデータを用いた細胞系譜解析による癌治療に対する抵抗性因子の探索手法に関する研究
Author(s)	村山, 皇翔
Citation	大阪大学, 2022, 博士論文
Version Type	VoR
URL	https://doi.org/10.18910/88157
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

シングルセル RNA-seq データを用いた
細胞系譜解析による癌治療に対する
抵抗性因子の探索手法に関する研究

提出先 大阪大学大学院情報科学研究科

提出年月 2022年1月

村山 皇翔

学位取得に関わる発表論文

国際学術論文

[1] Kosho Murayama, Hideo Matsuda. "Detecting lineage-specific marker genes for tumorevolution based on single cell transcriptome", International Journal of Bioscience, Biochemistry and Bioinformatics, Vol. 11, No. 3, pp. 50–57, 2021

(博士論文第 2 章)

国際会議 (採択済み)

[2] Kosho Murayama, Hideo Matsuda. "A Method for Detection of Markers for Epithelial-Mesenchymal Transition based on Single Cell Transcriptomic Data", Proceedings of ICBBB '22: 2022 12th International Conference on Bioscience, Biochemistry and Bioinformatics(in press)

(博士論文第 3 章)

内容梗概

近年、脳心血管障害に関連する高血圧、高脂血症、糖尿病等の生活習慣病に対する薬物治療の進展により余命は延長する一方、癌患者数の増加に伴う生活の質の低下や癌に関連する死亡は社会的な問題となっている。この状況を打破するために、癌の基礎的な特性を考慮した薬物治療の研究開発は目覚ましく進展している。しかしながら、癌は巧妙であり、やがて薬物への抵抗性を示すことから、根治に導くことは至極困難である。そのため、今後の更なる予後の改善を目指し、新たな治療薬開発の標的の検出、並びに検出された標的への薬剤開発が求められ続けている。

既存の手法では検出されていない新規の治療標的を見つけていくためには、多面的な手法が考えられる。その中でも、現在標準治療として使用されている薬物治療に対する抵抗性因子の探索は、薬効の消失が特定の遺伝子異常に依存している可能性があり、次の治療開発につながりやすいという観点で、未治療の状態から新規薬物治療の候補を一から検出するよりも効果的な手法であり、多くの研究者が取り組んでいる領域である。これまでに、癌患者から樹立された細胞株やマウス等の非臨床モデルや、実際の癌患者の組織検体や血液検体より得られた癌細胞、DNA、RNA、そしてタンパク質を対象として、広範な解析が行われてきた。

解析手法の中でも、遺伝子転写産物である RNA の発現を網羅的に評価するトランスクリプトーム解析は有力なツールであり、現在も癌研究の本流にあたる技術である。トランスクリプトーム解析としては、これまでに、細胞集団を平均的に評価する手法として、ガラス基板上に数万種類の DNA（プローブ）を高密度に固定したマイクロアレイや、次世代シーケンサーを用いた RNA シーケンス (RNA-seq) 解析が行われてきた。そして近年、細胞を単離する技術と次世代シーケンサーの解析能力の向上に伴い、単一細胞ごとの RNA の発現量を評価するシングルセル RNA-seq の導入が癌研究でも進んでいる。

シングルセル RNA-seq の解析は、単一細胞ごとの RNA の発現量を評価できるため、表現型が不均一な細胞集団である癌微小環境を、従来の RNA-seq よりも、薬剤抵抗性に関与するマーカーを明瞭に評価できる可能性がある。一方で、シングルセル RNA-seq の技術進展にあわせて、より膨大な量のデータを取り扱う必要があり、現在も更なる解析手法の開発が重要な研究領域である。

これまでに実施されてきたシングルセル RNA-seq データを対象とした先行研究では、細胞集団を細胞タイプごとに分類するクラスタリング手法、細胞の分化や刺激による反応

等に対して遺伝子発現が経時的にどのように変化するかを明らかにする細胞系譜解析等、各々の手法、またその最適な組み合わせを改善する開発が行われてきた。特に、シングルセル RNA-seq のデータを用いてクラスタリングを行い、クラスター間での遺伝子発現変動解析によるマーカーの検出が一般的であるが、この手法は離散化したクラスター間のマーカーの発現量の差の大きさを評価しているため、動的な過程で変化するマーカーの検出には、更なる手法の開発が求められている。

本学位論文は、以下の第 1 章から第 4 章で構成される。第 1 章では、研究の背景に関して述べた。第 2 章では、頭頸部癌に該当する口腔扁平上皮癌 (OSCC) 細胞に対して細胞障害性抗癌剤の一つであるシスプラチンを投薬し感受性が得られた株、抵抗性が導かれた株、およびシスプラチン抵抗性後薬剤を除去し一定の休薬期間を設けて培養された細胞 (薬剤抵抗性後、薬剤を除去して培養した細胞：今後薬剤休止と称す) より得られた単一細胞ごとの RNA-seq を使用した。得られたシングルセル RNA-seq データを対象に、バッチ効果除去のための複数サンプルデータの統合を行い、クラスタリングによって細胞タイプを分類した上で、薬剤抵抗性に関連する動的な過程を決定する細胞系譜を推定し、各系譜の連続的な遺伝子発現の推定、各系譜内および系譜間の発現変動マーカーの検出を行う手法の開発を試みた。シスプラチン抵抗性因子のマーカー検出を行うために上記の手法を試みた結果、先行研究のクラスター間での遺伝子発現変動解析によるマーカーの手法では検出されなかった潜在的なマーカー *UCA1*、*AXL*、*UBE2C*、*PTGR1*、*FTL*、*MT2A* を検出した。*UCA1* は既に OSCC におけるシスプラチンの抵抗性因子であると報告されている。また、*AXL* は、食道癌・卵巣癌・悪性胸膜中皮腫におけるシスプラチンの抵抗性因子としての報告や、OSCC の標準治療の一つである EGFR 抗体セツキシマブへの抵抗性因子としての報告がある。また、*UBE2C* は非小細胞肺癌や卵巣癌、*MT2A* は悪性胸膜中皮腫に対するシスプラチンの抵抗性因子であると報告されている。以上のことから、細胞系譜を用いた抵抗性因子検出の手法が有効であることが示唆された。第 3 章では、肺癌・前立腺癌・乳癌・卵巣癌へ上皮間葉転換 (EMT) を誘引する TGF β 1、EGF、TNF による刺激後の細胞より得られたシングルセル RNA-seq のデータを対象に、EMT に関与するマーカーを見出す手法を開発した。本研究では、EMT の進行時と EMT 反転時の両方向に対して最小全域木により系譜を推定し、系譜ごとの発現変動遺伝子を検出する手法の開発を試みた。本手法により、前立腺癌細胞株への EMT 誘導因子である TNF 刺激による EMT の進行と反転の系譜に沿って変動する *IL32* と *PLPP4* 遺伝子を検出した。更に、本手法を用いて得られた前立腺癌細胞株の EMT に関与するマーカーを対象

に、標準的な EMT 関連遺伝子を陽性リファレンスとして用い、定量的に精度を評価した結果、ROC 曲線での AUC は 0.5 を越えており、系譜推定による EMT に関与するマーカー検出の手法が有効であることが示唆された。第 4 章では本研究の結果より薬物治療や EMT による変化が、細胞の動的な遺伝子発現の変化であることを踏まえ、本手法により考えられる細胞の特性変化を捉えるアプローチの優位性に関して論じ、今後のシングルセル RNA-seq データを用いた薬剤抵抗性機序の解明に対する展望について述べた。

目次

第 1 章 緒論	1
1.1 癌の現状と治療の現状	1
1.2 癌研究における薬物治療への抵抗性	3
1.3 Epithelial Mesenchymal Transition : EMT に関して	4
1.4 癌研究におけるシングルセル RNA-seq 解析	5
1.5 先行研究の目的および研究の必要性	6
1.6 本論文の目的と構成	7
第 2 章 シングルセル RNA-seq データを用いた薬剤抵抗性に関与する系譜特異的 マーカーの検出	10
2.1 緒言	10
2.2 先行研究	10
2.2.1 先行研究の目的	11
2.2.2 先行研究のシングルセル RNA-seq データ	11
2.2.3 先行研究の解析手法	13
2.2.4 先行研究の解析結果	13
2.3 研究の目的	17
2.4 実験材料と解析手法	17
2.4.1 使用細胞株とシングルセル RNA-seq データ	17
2.5 結果	19
2.5.1 シングルセル RNA-seq データの統合とクラスタリング	19
2.5.2 細胞系譜推定	23
2.5.3 系譜に沿ったマーカーの発現の推定	24
2.5.4 系譜内および系譜間での発現差解析によるマーカーの検出	26
2.5.5 クラスターのマーカー解析との比較	29
2.6 考察	29
2.7 結言	34
第 3 章 シングルセル RNA-seq データを用いた上皮間葉転換に関与する系譜特異	

的マーカーの検出	36
3.1 緒言	36
3.2 先行研究	36
3.2.1 先行研究の目的	37
3.2.2 先行研究のシングルセル RNA-seq データ	37
3.2.3 先行研究の解析手法	38
3.2.4 先行研究の解析結果	39
3.3 研究の目的	40
3.4 実験材料と解析手法	40
3.4.1 使用細胞株とシングルセル RNA-seq データ	40
3.5 結果	41
3.5.1 シングルセル RNA-seq データの統合とクラスタリング	41
3.5.2 細胞系譜推定	42
3.5.3 EMT 進行の系譜と EMT 反転の系譜に沿ったマーカーの検出	44
3.5.4 先行研究との比較	47
3.6 考察	50
3.7 結言	51
第 4 章 結論	52
4.1 本研究のまとめ	52
4.2 癌研究における本研究の寄与とシングルセル RNA-seq の可能性	54
参考文献	57
謝辞	66
附録	67

第 1 章 緒論

1.1 癌の現状と治療の現状

現在、日本人の 2 人に 1 人は一生のうちに、何らかの臓器を起源とした癌にかかるといわれており、脳心血管関連の病気と同様に、癌はすべての人にとって大変身近な病気である。最新の癌統計のまとめ（引用：癌情報サービス [1]）では、2018 年に新たに診断された癌は 980,856 例（男性 558,874 例、女性 421,964 例）、2019 年に癌で死亡した人は 376,425 人（男性 220,339 人、女性 156,086 人）と報告されている。癌罹患数としては総数で順に大腸、胃、肺、乳房、前立腺、癌死亡数としては順に肺、大腸、胃、膵臓、肝臓との報告であり、全身の様々な臓器より癌は引き起こされることを踏まえた対策を講じる必要がある（引用：がん情報サービス [1]）。

異常な細胞が増殖し、体内に細胞の塊をつくることがあり、これが腫瘍と呼ばれる（引用：がん情報サービス [1]）。腫瘍は、周囲に広がる浸潤や、他の組織に移行する転移する悪性腫瘍と、浸潤や転移しない良性腫瘍に分けられる。この悪性腫瘍が癌である。癌の治療は、大きく局所療法と全身療法に分けることができる（表 1.1）。局所療法とは、目でみえる癌に対して放射線をあてることで癌細胞を殺す治療や、手術で癌を切除する選択肢である。全身療法とは、薬物を用いた治療選択肢である。癌に対する薬物治療開発の進展はこの数十年の歴史の中で目覚ましく、特に近年では旧来の歴史の中で用いられてきた細胞障害性抗癌剤の有効性、安全性、簡便性の限界をこえるために、癌に対する様々な研究開発の取り組みが行われている。現在日本で承認され病院で受けることができる癌の治療薬の主なものとしては、上記の細胞障害性抗癌薬以外に、下記の表 1.2、表 1.3 に示すように癌自体の特性、癌が存在する微小環境を考慮した薬剤開発が行われ、多大なる研究成果によって、近年各癌種ごとに使用できる選択肢が増えてきている。薬物療法の目的には、患者の病態・病気に応じて、「治癒」と「延命・症状緩和」があり、目的に応じて、薬物療法単独もしくは上記放射線療法や外科的な切除との組み合わせによって、予後の延長、根治率の向上、QOL の改善を目指す取り組みが行われている。

表 1.1 癌に対する治療選択肢

治療方針	具体的な治療選択肢
局所療法	放射線療法： 放射線（エックス線、電子線、ガンマ線等）を用いて、 癌細胞内の遺伝子 (DNA) にダメージを与えて癌細胞を殺す 治療 手術：手術で癌の塊や転移が疑われるリンパ節を切除する治療
全身療法	薬物治療：細胞障害性抗癌剤、内分泌療法、分子標的薬、免疫療法等

表 1.2 癌に対する薬物治療 (1)

治療の種類	薬物特性	例
細胞障害性抗癌剤	細胞が増殖する仕組みを利用して 癌細胞を攻撃する薬。正常細胞 よりも早く増殖する癌細胞の特性 を踏まえたアプローチである。 癌細胞以外にも増殖速度の早い 正常細胞を攻撃するため、脱毛、 胃腸、血液等の強い副作用が一般 的に伴う。	プラチナ製剤（シスプラ チン、カルボプラチン）、 代謝拮抗薬、微小管阻害 薬、トポイソメラーゼ阻 害薬等
内分泌療法薬 (ホルモン療法薬)	ホルモンを利用して増殖するタイプ の癌に対して、ホルモンの分泌や 働きを阻害して攻撃する薬である。 乳癌や前立腺癌などの特定 のタイプの癌でのみ使用される。	抗エストロゲン剤（タモ キシフェン）、抗アンド ロゲン剤（ビカルタミ ド）、アロマターゼ阻害 剤（アナストロゾール） 等

表 1.3 癌に対する薬物治療 (2)

治療の種類	薬物特性	例
分子標的薬	分子標的薬は、癌細胞に特有の標的分子をねらい撃ちすることで、効果を示す薬剤である。癌細胞の増殖を引き起こす遺伝子異常や特異的に発現するタンパク質を選択的にとらえて薬効を発揮する。標的とする遺伝子変異やタンパク質がある細胞へ選択的に働きかけるため、細胞障害性抗癌剤より副作用は、比較的軽いと言われている。	EGFR 抗体（セツキシマブ）、HER-2 抗体（トラスツズマブ）、EGFR チロシンキナーゼ阻害薬（オシメルチニブ）等
免疫療法	免疫療法は、癌微小環境における癌に対する免疫の力を活用した治療である。代表的な免疫療法である免疫チェックポイント阻害剤は、免疫細胞に対するブレーキを解除する治療であり、患者自身の免疫の力を活性化させることで薬効を発揮する。効果が得られた場合は、長期的なコントロールが示されている。	免疫チェックポイント阻害薬（CTLA-4 抗体：イピリムマブ、PD-1 抗体：ニボルマブ、ペムブロリズマブ、PD-L1 抗体：デュルバルマブ）等

1.2 癌研究における薬物治療への抵抗性

薬物療法の新しい選択肢を作り出していくために、現在までに癌細胞やマウス等を用いた非臨床のモデルや、ヒトから得られた腫瘍検体や血液検体より単離された癌細胞、DNA、RNA、タンパク質を用いた解析が行われてきた。これらの検体に対する解析手法の研究開発も、日進月歩で進んでいるが、それでもなお、癌を根治に導くことは困難であり、現在も世界中の大学、企業の研究者の尽力によって様々な知見が得られている。これまでに得られた知見を踏まえても、癌細胞は生存するために巧妙であり、薬物治療による殺傷から逃れるために、様々な機序を利用している。例えば歴史的に長く使用されてお

り、現在も広く様々な癌において標準治療として使用されている細胞障害性抗癌剤は、細胞からの薬剤排出に関わる ABC トランスポーターである P-糖タンパク質 *ABCB1* (別称 *MDR-1*)、*BCRP* (別称 *ABCG2*)、*ABCC1* (別称 *MRP-1*) 等の発現を通して、細胞外に排出する機序が知られている [2]。

また、薬剤への抵抗性機序には複雑性があり、1つの因子で生じることもあれば、複数の因子が組み合わさり導かれることも一般的である [2]。癌細胞自体の内因的な変化だけでなく、癌細胞局所の周囲をとりまく癌微小環境における時間的・空間的な変化、癌細胞の不均一性 (Tumor heterogeneity) も含めて、癌細胞が薬剤に抵抗を示す科学的な理解を追求する必要がある、包括的な解析が求められている。こういった多面性を持つ癌細胞に対する薬剤抵抗性の機序を見出すために、多面的な手法が行われており、その中でも治療薬への抵抗性に関与する癌の進化に伴う特性変化の検出は、次の治療戦略を立てるために非常に重要な研究領域の一つである。

1.3 Epithelial Mesenchymal Transition : EMT に関して

次の治療戦略を考えていくために、これまでに注目されている癌細胞の特性変化の一つとして、Epithelial Mesenchymal Transition : 略称 EMT (別称 : 上皮間葉転換または上皮間葉移行) がある。EMT は、正常な状態では生理学的に重要な胚発生、創傷治癒に関与する。一方で癌細胞にとっては、上皮細胞から特性を変化させることで、血流やリンパを通じた遠隔の別の組織への転移促進や、既存の薬物治療への抵抗性を導く動的な特性変化にあたる [3][4]。具体的には、上皮系の細胞が間葉系細胞へ特性が移行する中で、細胞の極性や周囲細胞との接着能を失うことを背景に、遊走、浸潤能を得る。間葉系へ移行する前の上皮系の細胞は、密着結合 (タイトジャンクション)、接着結合 (アドヘレンスジャンクション)、接着斑/半接着斑 (デスモソーム/ヘミデスモソーム) を介して、細胞の空間的な方向性を持ち、半接着斑を介して基底膜につながれており、状態を維持するために、各分子を発現して、局所に留まることで恒常性を維持している。一方で、間葉系細胞への移行に伴い、癌にとって転移しやすい細胞特性の土台形成が行われており、上皮状態を維持する遺伝子の発現を阻害する EMT 関連転写因子である *ZEB*、*SNAIL*、*TWIST* の発現を誘導し、N-カドヘリン、ビメンチン等の間葉状態に関連する遺伝子の発現が活性化している。特に、間葉系の細胞では、上皮細胞間の接着で役割を果たす接着結合が減退していることを受け、原発巣から癌細胞が血流にのりやすく、遠隔部位に拡散しやすい。またその後の重要なプロセスとして、Mesenchymal Epithelial Transition : 略称 MET (別

称：間葉上皮転換または間葉上皮移行）が引き起こされることで、遠隔の転移先で癌を定着させることも報告されている。

EMT は、上記の転移促進に関連する遊走能の亢進以外にも、細胞障害性抗癌剤等の抗癌剤への抵抗性自体を導くことも報告されている。例として、EMT を誘導する転写因子である *ZEB*、*SNAIL*、*SLUG* が、乳癌、卵巣癌、大腸癌、膵臓癌において、シスプラチンやオキサリプラチン等のプラチナ製剤を基軸とした抗癌剤への抵抗性を導いたことも示されている [4]。

EMT が転移能の亢進や薬剤抵抗性に及ぼす影響を考慮すると [3][4]、次の治療薬開発に有望な因子を探索する対象として有用である。上述してきた通り、EMT は、薬剤抵抗性と同じく細胞ごとの特性が動的に変化しているため、シングルセル RNAseq データを対象に、動的な特性変化の過程を捉える手法が、有効であることが考えられる。

1.4 癌研究におけるシングルセル RNA-seq 解析

特性変化をとらえる手法として、DNA の遺伝子異常や RNA の発現、タンパク質の発現等、様々な検体を対象に解析技術の開発が行われてきた。その中で、RNA の発現を評価する手法は、マイクロアレイ、RNA Sequencing（略称：RNA-seq）と、この 20 年程の間で大きく技術革新が行われてきた。これまでに、得られた細胞より RNA の発現を平均して評価する手法が一般的に行われてきた中で、近年特に単一細胞ごとに遺伝子発現を評価するシングルセル RNA-seq 解析が大変注目されている [5]。この解析手法は、個々の細胞や分子の個性を解明することが可能であり、より包括的な理解を進める強力なツールとして、細胞の分化に関する研究にとどまらず、様々な癌を対象として近年活用されている。シングルセル RNA-seq 解析は、単一細胞ごとの遺伝子発現情報が得られることを受け、今までよりも膨大なデータを取り扱う必要があり、更なる解析手法に関する研究が現在までに取り組みられてきた。

これまでに実施されてきたシングルセル RNA-seq 解析の研究では、細胞集団を細胞タイプごとに分類するクラスティング手法、細胞の分化や刺激による反応等に対して遺伝子発現が経時的にどのように変化するかを明らかにする細胞系譜解析等が行われてきた [5]。一方で、各解析手法自体の改善（バッチ効果除去の改良や、単一の系譜に加えて複数系譜の推定等）や、解析手法の適切な組み合わせ等、開発余地は未だ多く残されている。特に、癌細胞の動的な変化を考慮したマーカーの検出手法の開発は、シングルセル RNA-seq 解析の研究手法の開発で先行している細胞の分化を対象とした研究と比較して、限られた形

での実行に留まっている。具体的には、次の先行研究での目的および研究の必要性にて述べる。

1.5 先行研究の目的および研究の必要性

第2章の研究では、薬剤抵抗性の因子を検出する手法を開発するために、薬剤開発の必要性が高い口腔扁平上皮癌 (oral squamous cell carcinoma: これより OSCC と称す) に、標準治療である細胞障害性抗癌剤のシスプラチンを投与し、抵抗性を誘導した細胞から得たシングルセル RNA-seq のデータを用いた [6]。

先行研究では、表現型として homogeneous (均一) な癌細胞と heterogeneous (不均一) な癌細胞に対して薬剤による選択圧や転移による癌細胞の獲得性の生存戦略 (薬剤抵抗性因子の検出) を明らかにすることが目的とされた。特に、均一な表現型を持つ細胞集団においては、抵抗性が生じる機序として、今までになかった新規の表現型を獲得するという仮説が考えられた。

上記の仮説を検証するために、主に、表現型として均一な細胞株である HN120、不均一な細胞株である HN137 の2種類が用いられた。更に、これらの細胞株にシスプラチンを投与した後の抵抗性株と、抵抗性後薬剤を休薬した薬剤休止株のシングルセル RNA-seq データを用いて解析が実施された。

実施された手法としては、希少なサブ集団を検出するためにデザインされた RaceID [7] を用いて、非階層型クラスタリングの手法の一つである k-means クラスタリングを行い、t 分布型確率的近傍埋め込み法/t-distributed stochastic neighbor embedding (t-SNE) [8] を用いて次元削減が行われた。

次に、薬物投薬による変化が細胞起源 (cell-of origin) [9] に関係しているかを確認するために、癌幹細胞仮説をもとに、幹細胞性関連遺伝子を使用して、シングルセル RNA-seq データを対象に、最短距離法による階層的クラスタリングが行われ、重み付き遺伝子共発現ネットワーク解析 (WGCNA) の dynamicTreeCut [10] によって樹形図よりクラスターの検出が行われた。以上の RaceID により得られたクラスターと幹細胞クラスターのデータを統合することで、幹細胞性関連遺伝子の中から、抵抗性因子に関与する因子の検出が行われた。

更に、HN120 と HN137 細胞間の細胞起源を推定する目的で、RaceID2/StemID [11] が使用された。また、細胞分化の推定による疑似時間解析によって、細胞起源を確認するために、Monocle2.0 [12] が用いられた。

以上の解析手法によって、遺伝子発現が不均一な細胞では幹細胞遺伝子の一つである *SOX2* の発現を維持した状態での薬剤による抵抗性細胞の選択、均一な細胞では細胞自体の特性を位置づける遺伝子として *SOX2* から *SOX9* へのスイッチが示唆された。これらの手法では、細胞起源を軸に、癌幹細胞仮説に基づく幹細胞関連遺伝子を用いて解析された。それ以外の遺伝子の関与を評価するには、細胞起源の仮説を用いない別の手法を考える必要がある。また先行研究においては、クラスタリングを用いて、抵抗性遺伝子の評価が行われたため、細胞特性が変化する過程の中で移行する遺伝子発現を踏まえたマーカーの検出は実施されていなかった。

第3章の研究で使用したシングルセル RNA-seq データの参照先である先行研究 [13] においては、複数の癌腫の細胞株を対象に EMT 誘導因子を用いて実験的に EMT を誘導した単一細胞ごとの RNA-seq データを用いて、EMT 応答マーカーの比較を包括的に行うことでこの条件固有性の仮説を証明することを目的とされた。解析のために、Seurat[14] を用いて分散安定化変換を計算した上で、次元削減時の特徴量として用い、上位の変動 3000 遺伝子が使用された。また、EMT のダイナミクスを評価するために、細胞に付与された時刻ラベルを基に、ロジスティック回帰によって、各細胞を時刻ラベルにあてはめることで、疑似時間の推定も行われた。一方で、これらの手法では、遺伝子発現の刺激因子投与に対する連続的な移行を捉えることはできない。刺激因子投与からの細胞内の遺伝子発現変化を捉えることが困難であることを踏まえると、細胞内の遺伝子発現の連続的な遷移過程で特徴的な発現変動を示す遺伝子を検出するには、経時的变化に着目した別の手法を開発する必要がある。

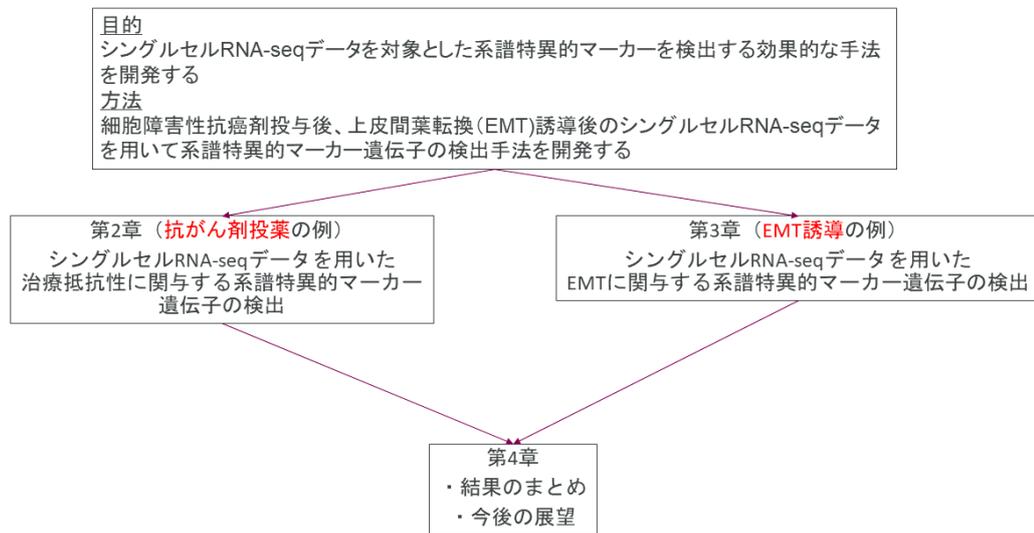
1.6 本論文の目的と構成

これまでに述べてきた通り、現在までに続けられてきた癌治療薬の次の展開を考える研究領域において、新しい技術潮流としてのシングルセル RNA-seq データを対象とした解析手法に関して、更なる手法の改良と提案は現在も求められている。

特に、本研究の対象とした薬剤抵抗性や EMT は動的に遺伝子発現が変化することによって生じる事象である。現在最も一般的に実施されているクラスタリングの手法では、クラスター間の遺伝子発現差の比較に留まるため、動的な変化を捉えるには限界がある。そこで本研究では、細胞の分化に関して適用されてきた細胞系譜特異的マーカーの解析手法に着目し、細胞内の遺伝子発現の連続的な状態遷移を捉える手法の開発を試みた。

本学位論文は、第1章から第4章で構成される(図 1.1)。第1章では、研究の背景に関

して述べた。第2章では、提案手法を展開するために、シスプラチンを投与した際に、口腔扁平上皮癌（OSCC）細胞感受性株、抵抗性株、薬剤休止株のシングルセル RNA-seq データにおいて、抵抗性に関わる特徴的な遺伝子発現の遷移があると仮定した。第3章では、提案手法を展開するために、上皮間葉転換（EMT）を誘引する TGFβ1、EGF、TNF による刺激後、肺癌・前立腺癌・乳癌・卵巣癌の細胞より得られたシングルセル RNA-seq のデータにおいて、EMT に応答する特徴的な遺伝子の発現の推移があると仮定した。第2章および第3章で取り扱う細胞集団の状態の遷移は、細胞系譜に沿って連続的に進行すると仮定して、各系譜内においての疑似時間に沿った遺伝子発現変動、また遺伝子発現の差異を相互に比較することで、対象としている状態遷移で特徴的な発現変動を示すマーカーを検出する手法を開発した。第4章では、薬物治療や EMT による変化が、細胞の動的な遺伝子発現の変動で特徴づけられることを踏まえ、本手法により考えられる細胞の特徴的な遺伝子発現の変動を捉える手法の優位性に関してまとめ、今後のシングルセル RNA-seq データを用いた薬剤抵抗性機序の解明に対する展望について述べた。



18

図 1.1 本論文の構成

第 2 章 シングルセル RNA-seq データを用いた薬剤抵抗性 に関する系譜特異的マーカーの検出

2.1 緒言

第 1 章でも述べたように、シングルセル RNA-seq は、単一細胞分解能でのトランスクリプトーム解析において、現在広く活用され始めている領域である [5]。

本章では、薬剤投薬後の癌の薬剤抵抗性への移行過程を解析するための新しい手法を提案する。本研究の目的としては、薬剤感受性から薬剤抵抗性への移行中において、癌細胞で発現するマーカーを検出する手法を見出すことである。この目的に対する現在の手法として、例えば、先行研究として参照した Sharma ら [6] によって報告されたものは、癌細胞のクラスタリングを行い、各々のクラスターに含まれる細胞上の遺伝子発現を用いて表現型のグループを検出し、各薬剤感受性、薬剤抵抗性、薬剤休止の細胞が含まれるグループを活用して、均一な表現型を持つ細胞と不均一な表現型を持つ細胞が異なる抵抗性因子を持つ仮説を検証するために、薬剤による表現型の変化を解析した。薬剤感受性から薬剤抵抗性への移行過程の解析では、薬剤感受性と薬剤抵抗性のそれぞれの細胞において遺伝子発現を比較したが、先行研究の手法では、薬剤感受性から薬剤抵抗性への移行プロセスの途中での遺伝子発現の変化は考慮されていなかった。

本章で述べる手法は、次の 4 つのステップで構成されている：(1) 薬剤感受性、薬剤抵抗性、薬剤休止における癌細胞のシングルセル RNA-seq を対象に、異なるサンプル間のバッチ効果を除去し、細胞タイプを分類するために、クラスタリングを行う。(2) クラスター上に、細胞集団の状態遷移を表現する細胞系譜を推定する。(3) 各系譜において疑似時間による連続的な遺伝子発現の変化を推定する。(4) 系譜内および系譜間で発現が変動したマーカーを検出する。この一連の手法によって、特定の系譜が薬剤抵抗性に向かって進行するにつれて発現が変動するマーカーを見出すことを可能とする。

2.2 先行研究

ここでは、本研究で取り組んだ系譜推定を用いた抵抗性因子の検出手法との違いを説明するために、先行研究 [6] に関して、目的、使用されたシングルセル RNA-seq データ、解析手法、得られた結果に関して端的に述べる。

2.2.1 先行研究の目的

先行研究 [6] においては、細胞障害性抗癌剤の代表的な薬剤であるプラチナ製剤シスプラチンを用いて、表現型として homogeneous (均一) な癌細胞と heterogeneous (不均一) な癌細胞に対して薬剤による選択圧や転移による癌細胞の獲得性の生存戦略 (薬剤抵抗性因子の検出) を明らかにすることが目的とされた。特に、均一な表現型を持つ細胞集団においては、抵抗性が生じる機序として、今までになかった新規の表現型を獲得するという仮説がおかれた。

2.2.2 先行研究のシングルセル RNA-seq データ

先行研究 [6] では、今までに広く使用されてきた従来の RNA-seq での評価が困難であった癌細胞ごとのクローン進化や腫瘍内不均一性をより正確に評価する手法として、単一細胞単位で解析する手法であるシングルセル RNA-seq データ解析が用いられた。

これより、薬物治療による癌進化を観測可能な細胞株を目指して樹立された一連の薬剤感受性細胞株、そして薬剤投与を経て得られた細胞株に関して説明する。研究では、頭頸部癌の一つである口腔扁平上皮癌 (oral squamous cell carcinoma: これより OSCC と称す) の患者より樹立された細胞株 (patient derived cell lines: PDPCs) が使用された。OSCC は、5 年生存率が 4~5 割と予後不良であり、現在の標準治療の限界を踏まえて、次世代の薬剤開発が求められている癌腫の一つである [15]。OSCC では、治癒目的の切除後微小残存癌細胞を原因とした再発予防や、再発・転移時の標準治療として、細胞障害性抗癌剤の一つであるシスプラチンが使用されている [16]。先行研究においては、このシスプラチンによる薬剤治療の影響を受けた細胞株を樹立し、細胞の特性変化を見出すことが目的とされたので順に説明する。

研究で使用された細胞株を下記表 (表 2.1) にまとめた。

使用された患者由来の細胞株は HN120 と HN137 である。これらの細胞株に関して、原発巣から樹立された細胞株を各々 HN120P (原発巣を意味する Primary) と HN137P と表記する。HN120P は、蛍光免疫染色による評価において、上皮系細胞のマーカである E-カドヘリン (これより *ECAD* と称す) と基底細胞のマーカである *KRT5* を均一に発現する”上皮/基底パターン”の細胞株であった。一方で、HN137P は上記”上皮/基底パターン”の *ECAD/KRT5* だけでなく、管腔細胞のマーカである *KRT18* と間葉系細胞のマーカであるビメンチン (これより *VIM* と称す) の発現 (*KRT18/VIM*) を示

表 2.1 使用細胞株 (細胞数)

細胞株	薬剤抵抗性株	薬剤休止株	由来:原発/転移
HN120P(90)	HN120PCR(95)	HN120PCRDH(93)	原発
HN120M(91)	HN120MCR(92)	HN120MCRDH(87)	転移
HN137P(253)	HN137PCR(78)	HN137PCRDH(75)	原発
HN137M(75)	HN137MCR(87)	該当なし	転移
HN148(186)	該当なし	該当なし	原発/転移

す”管腔/間葉パターン”とが、不均一にみとめられる細胞株であった。これらの細胞はシスプラチンとの培養において反応が認められ、薬剤への感受性を示した。

また、転移性細胞の表現型が薬剤抵抗性細胞と同様の表現型を獲得することが過去の知見より知られていることを考慮し、上記の細胞株を樹立した同じ症例の転移先にあたるリンパ節より、HN120M(転移を意味する Metastasis より) と HN137M が樹立された。

これらの細胞株に関して、 1×10^7 細胞をシスプラチン投与下で 4 か月間培養して樹立した薬剤抵抗性の細胞株を各々 HN120PCR、HN120MCR、HN137PCR、HN137MCR と表記する。

各々の細胞の表現型を確認するために蛍光免疫染色による評価が行われており、HN120PCR と HN120M では、*KRT18/VIM* を発現しており、薬剤抵抗性と転移性の細胞株で同様の”管腔/間葉”パターンの表現型を持つことが示された。また、HN137P では、*ECAD/KRT5* と *KRT18/VIM* が不均一に発現していた一方で、薬剤抵抗性の HN137PCR では *ECAD/KRT5* のみ発現(”上皮/基底パターン”)、転移性の HN137M では *KRT18/VIM* のみ発現(”管腔/間葉パターン”)が認められ、異なる表現型であることが示された。

更にこれらの抵抗性を持った細胞株をシスプラチン非存在下の状態で再培養した細胞株を各々 HN120PCRDH(薬剤休止:休薬を意味する Drug holiday/DH)、HN120MCRDH、HN137PCRDH、HN137MCRDH でと表記する。

これらの細胞株に加えて、解析では薬剤未治療であるが内因性の理由で抵抗性化している HN148 がコントロールとして用いられた。

シングルセル RNA-seq には、フリューダイン社の C1 Single-Cell Auto Prep 集積流体回路 (IFC: Integrated Fluidic Circuit) システムが用いられた。フリューダイン社の C1 system によって調製後、イルミナ社の HiSeq-Hi-output-2500 sequencer によって配

列が決定された。

以上の細胞株、手法より得られた計 1302 の単一細胞ごとの RNA-seq データが解析に使用された。

2.2.3 先行研究の解析手法

シングルセル RNA-seq データの解析には、RaceID[7] が用いられた。RaceID[7] は、教師なし学習によるクラスタリングを行った上で、検出された外れ値で調整を行い、特に希少なサブ集団を検出するためにデザインされた方法である。本データの正規化は、各細胞の遺伝子の転写物のカウント数を、各細胞の転写物の総数で割った後、細胞全体の転写物の総数の中央値を掛けて行われた。RaceID の [7] のシングルセルのライブラリは、遺伝子発現行列を作成した後に、k-means クラスタリングが実施された。高次元データを可視化するための次元削減には t-distributed stochastic neighbor embedding (t-SNE) [8] が使用された。

また、薬物投薬による変化が細胞起源 (cell-of origin) [9] に関係しているかを確認するために、幹細胞関連遺伝子を使用して、シングルセル RNA-seq データを対象に、最短距離法による階層的クラスタリングが行われ、WGCNA の dynamicTreeCut[10] によって樹形図よりクラスターの検出が行われた。

また、各細胞間の細胞起源を確認するために、RaceID2/StemID[11] が使用された。RaceID2/StemID[11] は、RaceID で識別されたクラスタリングの結果を使用して細胞系譜木を推定する方法である。更に、本結果は Monocle2.0[12] を用いて疑似時間解析を行い、結果の確認が行われた。

2.2.4 先行研究の解析結果

RaceID[7] を用いたシングルセル RNA-seq 解析において、遺伝子発現行列の作成とそれに続く k-means クラスタリングが実施された結果、5 つの主要なクラスターが作られた (RaceID クラスター)。これらを可視化するための次元削減には、t-distributed stochastic neighbor embedding (t-SNE) [8] が使用された。先行研究の手法とその結果の全体像を下記に示す (図 2.1)。

これらの 5 つの主要なクラスターは更に、特定の遺伝子の発現パターンによって、2 つの表現型の状態に分類された。1 つは、上皮系細胞のマーカー (*EPCAM/ECAD*) と基底細胞のマーカー (*KRT5/KRT14*) を発現する”上皮/基底細胞パターン”、もう 1 つは

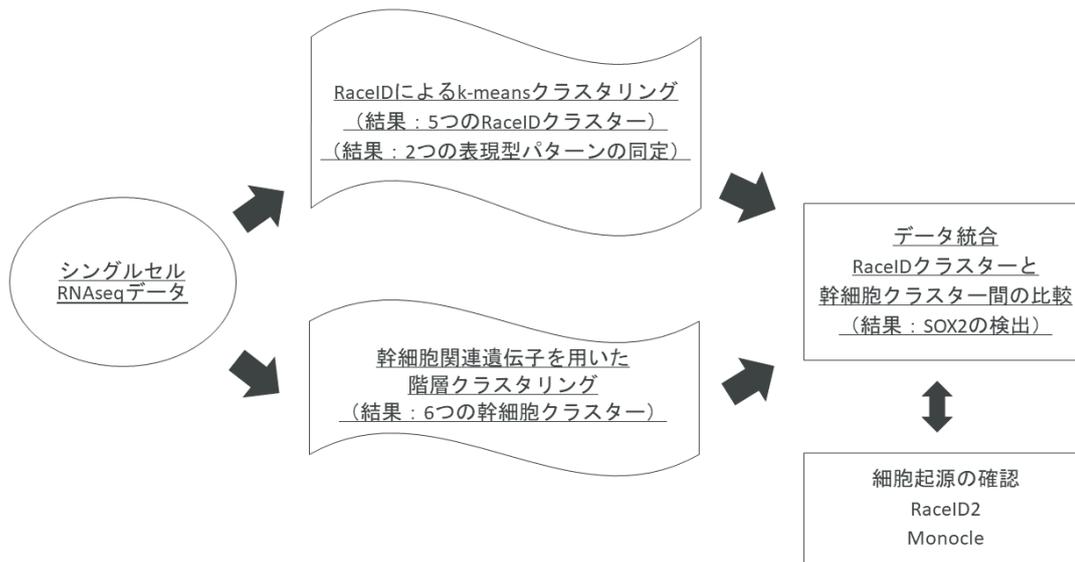


図 2.1 先行研究の手法とその結果の全体像

管腔細胞のマーカー (*KRT8/KRT18*) と間葉系細胞のマーカー (*VIM/SPARC*) を発現する”管腔/間葉パターン”である。

HN120 と HN137 にシスプラチン投薬により与えられた薬剤抵抗性に関与する変化が、細胞起源 (cell-of origin) に関連するか否かを評価するために、幹細胞関連遺伝子を基にシングルセル RNA-seq データを用いて階層的クラスタリングを行った結果、6 つの主要なクラスター (幹細胞クラスター) が見いだされた。上記の RaceID の中で、定義された”上皮/基底細胞パターン”のクラスターにおいて、この階層的クラスタリングによって見いだされた幹細胞関連遺伝子のクラスターが不均一に認められた。

特に、HN120 と HN137 の上皮系細胞から 2 つの異なる幹細胞系クラスターが形成され、いずれのクラスターにおいても、多能性状態にリプログラミングするために重要な転写因子として知られている幹細胞関連遺伝子 *SOX2* が発現していることを見出された。*SOX2* は、癌遺伝子としても知られており、腫瘍内に浸潤している細胞の特性維持に必須と考えられている [17, 18]。彼らはこの結果を、細胞株を対象に *SOX2* の免疫染色で評価し、”上皮/基底細胞パターン”の HN137PCR で高発現していることを確認した。一方で、間葉系様の HN120PCR、HN120M、HN137M では *SOX2* の発現を認めなかった。

更に検証を深めるため実施された細胞株を移植したマウスモデルを対象にシスプラチンを投薬して評価した結果では、HN120P および HN137P いずれにおいてもシスプラチン

抵抗性時に、細胞株の結果と同様に HN120 では *SOX2* の発現が喪失した上で *SOX9* の発現が認められ、HN137 では *SOX2* の発現が認められた。

更にクローン選択の仮説を検証する目的で、細胞起源の共通性を評価するために、RaceID2 が用いられた [11]。結果として、HN120P と HN120PCR の間で、細胞起源の共通性は見いだされなかった。一方で、HN137P と HN137PCR では共通の細胞起源が見いだされたことを受け、HN137 に薬剤を投薬したモデルでは既存の癌幹細胞のクローン選択が生じていることが示唆された。

この HN120P と HN120PCR 間の細胞起源の不一致は、別の手法である Monocle2.0[12] による疑似時間解析でも確認された。HN120 の疑似時間時系列の順序づけによる原発由来の細胞株と薬剤抵抗性/転移性細胞株から推定された細胞系譜は疑似時間に沿って不連続であり、薬剤によるストレス誘導性の細胞状態の切り替えが生じていることが示唆された。一方で、HN137P より、薬剤抵抗性の HN137PCR と転移性の HN137M への連続的な異なる 2 つの分岐が導かれた。これは、クローン選択の仮説を裏付けるものであった。

先行研究の結果をまとめると、シスプラチンへの抵抗性は、癌進化に関連する異なる 2 つの方向性があることが見いだされた。一つは、上皮細胞集団と間葉細胞集団の両方を含む表現型が不均一な状態 (HN137 細胞) へのシスプラチン投薬は、細胞周期が遅く静止状態の上皮細胞の選択をもたらす点であった (図 2.2)。もう一つは、表現型として均質な上皮系の患者由来の原発細胞 (HN120) は、薬剤誘発性の間葉系特性の獲得によって細胞自体の特性の切り替えをもたらす細胞の可塑性を示した点であった (図 2.3)。この可塑性には、薬剤ストレス誘発性の幹細胞へのスイッチが関与することが示唆された。癌幹細胞の特性は、これまでも薬剤抵抗性への関与が示唆されている。特に先行研究のシングルセル RNA-seq のクラスターから、不均一な細胞群では幹細胞関連遺伝子の状態が評価され、その結果として幹細胞関連遺伝子である *SOX2* の関与が示唆された。一方で、均一な細胞では、細胞自体の特性が変化することによって、薬剤への抵抗性が生じていることが示唆された。

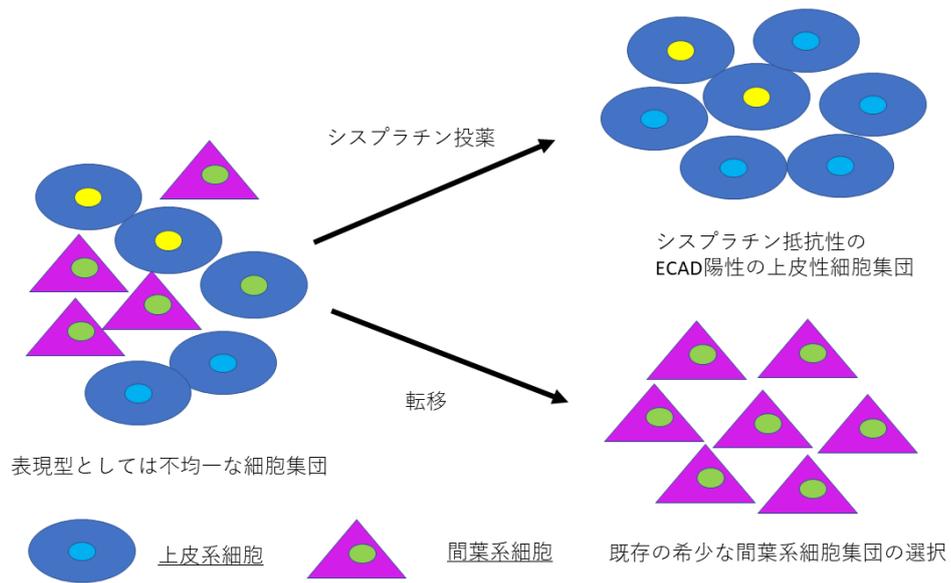


図 2.2 既存細胞の選択

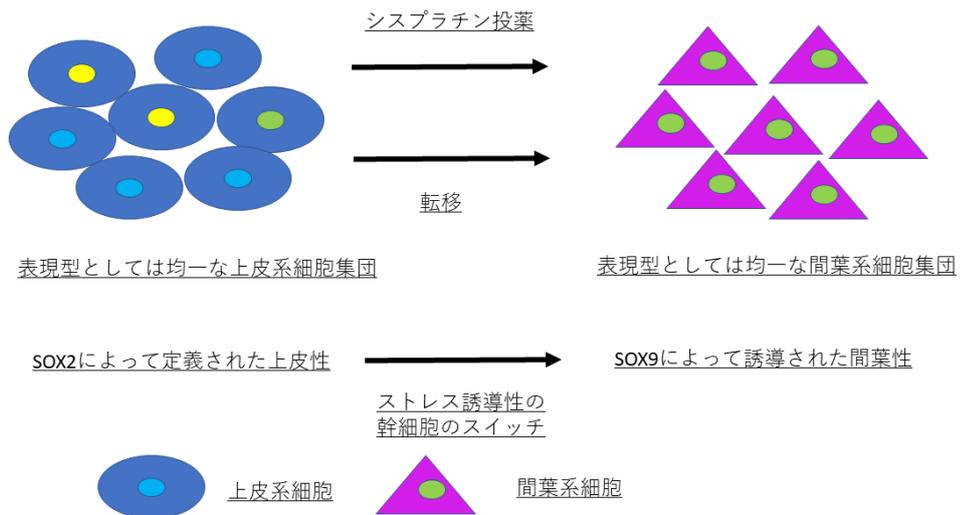


図 2.3 細胞特性の切り替え

2.3 研究の目的

先行研究 [6] では、仮説として腫瘍内細胞が不均一である場合は、癌幹細胞仮説を評価することを目指し、腫瘍内細胞が均一である場合は、細胞の特性自体が変化することを評価することを目指して実施された。先行研究では、クラスタリング解析を活用した上で、既存の知見を踏まえた癌幹細胞仮説を基にして、幹細胞関連因子を対象とした抵抗性因子の検出が行われた。

本研究においては、薬剤感受性から薬剤抵抗性に向かう連続的な遷移過程で特徴的な発現変動を示す遺伝子を検出するシングルセル RNA-seq データの解析手法の開発を目指す。本研究を行うにあたり、癌の薬剤抵抗性の獲得を、薬剤感受性と薬剤抵抗性という離散的な状態間の遷移ではなく、薬剤感受性から薬剤抵抗性に向かう連続的な遷移過程ととらえることにより、細胞分化過程の解析で用いられている細胞系譜の推定手法を応用することを考えた。細胞系譜の推定手法は、遺伝子発現の推移を評価することを可能とする手法であり、薬剤感受性から薬剤抵抗性に向かう連続的な遷移過程で特徴的な発現変動を示す遺伝子を検出し、従来の離散的な状態間での発現変動解析では得られなかった、新たな薬剤抵抗性のマーカー遺伝子が得られるのではないかと考えるに至った。

2.4 実験材料と解析手法

2.4.1 使用細胞株とシングルセル RNA-seq データ

本研究では、先行研究 [6] で使用された細胞株より得られたシングルセル RNA-seq データを格納先の Gene Expression Omnibus リポジトリ (GSE117872) から入手し使用した。本研究では、先行研究の解析では含まれていた内因的に抵抗性のポジティブコントロールである HN148 は、薬剤抵抗性細胞と薬剤休止の細胞を含まないため、細胞系譜解析ができないことを考慮し解析には含めなかった。

解析の手法としては、大きく 4 つに分かれている。順に説明すると、(1) 細胞株の条件の違いによって生じるバッチ効果の影響を除去するために実施する正準相関分析を基にしたサンプルデータの統合とクラスタリング、(2) 薬剤抵抗性に関連する細胞系譜の推定、(3) 各系譜の連続的な遺伝子発現の推定、(4) 各系譜内および系譜間の発現変動マーカーの検出を実施した。本手法の実装は R で行い、ライブラリとして (1) では Seurat[14]、(2) では Slingshot[19]、(3)(4) では tradeSeq[20] を使用した。各手法に関しては下記解析手法に示す。

本解析では、シングルセル RNA-seq のカウント行列データやクラスタリングの結果を格納するために、Seurat オブジェクトを作成した。この Seurat オブジェクトを対象に、低品質な細胞の情報を除くために、少なくとも 3 細胞以上で発現している遺伝子、少なくとも 200 遺伝子以上発現している細胞の結果を用いるようにフィルタリングを行った。更に前処理として、細胞ごとの検出されたリード数と遺伝子数の比を調べ、2 個以上の細胞が入ったデータを除くために、遺伝子数が 10,000 未満であったものを使用した。上記のフィルタリングで使用データの品質を確保した上で、細胞ごとの発現量のばらつきを補正するために正規化を行った。細胞ごとの発現量を、各細胞の総発現量で割り、スケールファクターをかけた後に対数値に変換した。その次に、発現量の対数値を、各遺伝子について全ての細胞での平均値と標準偏差を求め、各遺伝子の発現量の対数値から平均値を引いて標準偏差で割った値である Z スコアに変換した。以上の正規化したデータを対象に、高発現している 2000 遺伝子を選択した。

本研究では、複数の異なる条件下のデータを統合するために、まず 2 つのデータ間で共通する特徴量を正準相関分析 [21] で検出する。具体的には、データ間で発現量の相関が高い遺伝子の検出を行う。次に、検出された遺伝子をもとに、2 個のデータ間でクラスター同士の相互近傍 (mutual nearest neighbor/MNN[22]) を求め、相互近傍が近いクラスターを対として重ね合わさるように配置した。

次元削減には、t-SNE[8] を用いた。背景としては、Sharma ら [6] の先行研究でクラスタリングの結果を t-SNE で可視化していたことを受け、本研究の結果を比較するためである。t-SNE[8] は、高次元データを可視化するために 2 次元もしくは 3 次元の低次元空間へ次元削減する手法の一つである。PCA や自己組織化マップ等の既存の線形データを対象とした手法では難しい高次元における非線形の類似データを低次元上でも近くに維持することを可能とする。t-SNE では、高次元におけるデータポイント間の距離 (ユークリッド距離) を条件付き確率に変換し、次に、低次元でのデータポイント間の距離を条件付き確立に変換した上で、カルバック・ライブラー・ダイバージェンス情報量を用いて、二つの確率分布がどの程度似ているかを比較し、損失関数が最小となるように各配置を行う。また、正規分布に比べて裾が長い自由度 1 の t 分布を使用することによって、近いデータはより近くに、遠いデータはより遠くに表現することを可能としている。以上の特性を活かし、非線形の高次元データであるシングルセル RNA-seq データの解析において、次元削減のために活用されている。

上記の次元削減とクラスターラベルのベクトルの細胞の行列データを対象に、細胞系譜

解析のために、Slingshot[19] を使用した。本研究では、クラスタリングの結果より得られた各クラスターのラベルを入力に使用し、クラスター上に最小全域木を構築した上で、これらのクラスターを系譜に順序付けする。出発点のクラスターのラベルを定義し、全ての系譜が出発点に指定したクラスターを共有し、最小全域木を構築している。各木内の経路に対する主曲線をフィッティングすることによって系譜を平滑化し、この主曲線に沿って、疑似時間を定義し計算を行う。

各系譜の連続的な遺伝子発現の推定、各系譜内および系譜間の発現変動マーカの検出には、tradeSeq[20] を使用した。本研究において使用した一般化加法モデルは、遺伝子発現測定値を疑似時間の非線形関数としてモデル化する手法であり、一般化線形モデルの線形和の限界を改善し、説明変数の重み付き和から、各説明変数の一般的な変換の総和にしたものである。疑似時間に沿った遺伝子発現の推移を結ぶ曲線の作成を行うために、各時点の間を補完した曲線を3次関数で表し、平滑化された曲線として3次スプライン曲線を用いている。

本研究では、系譜推定後の系譜に沿った疑似時間を用いているため、使用する疑似時間は系譜推定の手法と連動する。そのため、示される疑似時間は、前述の手法 [19] で推定した系譜に沿って定義され、計算されたものである。

2.5 結果

2.5.1 シングルセル RNA-seq データの統合とクラスタリング

2.2.2 節で述べたシングルセル RNA-seq データに対して、正準相関分析でデータの統合を行った [14][21]。データの統合で使用した正準相関分析の寄与度を示すため、正準相関分析の寄与率を計算した (図 2.4)。HN120 と HN137 で共通に発現する遺伝子を求め、それを特徴量として正準相関分析を行った。正準相関分析では、主成分分析の主成分に相当する正準変量がどの程度の情報を要約しているかを表す指標として、正準相関係数を2乗した値 (主成分分析の固有値に相当する) を用いている。HN120 と HN137 のデータ統合では、30次元までの正準変量を用いたため、30次元までの値をプロットした。

次に、Seurat[14] を用いてクラスタリングを行い、t-SNE で次元削減を実施した (解像度 1.0) (図 2.5、図 2.6)。別の次元削減の手法である主成分解析 (principal component analysis : PCA) で次元削減を行った解析 (図 2.7) ではクラスターがうまく分かれなかったが、本手法では、図 2.5 に示すように、3つのグループ A、B、および C に分かれた。それぞれ、グループ A は HN120P、HN137P (HN137P 細胞の一部)、および HN137PCR、

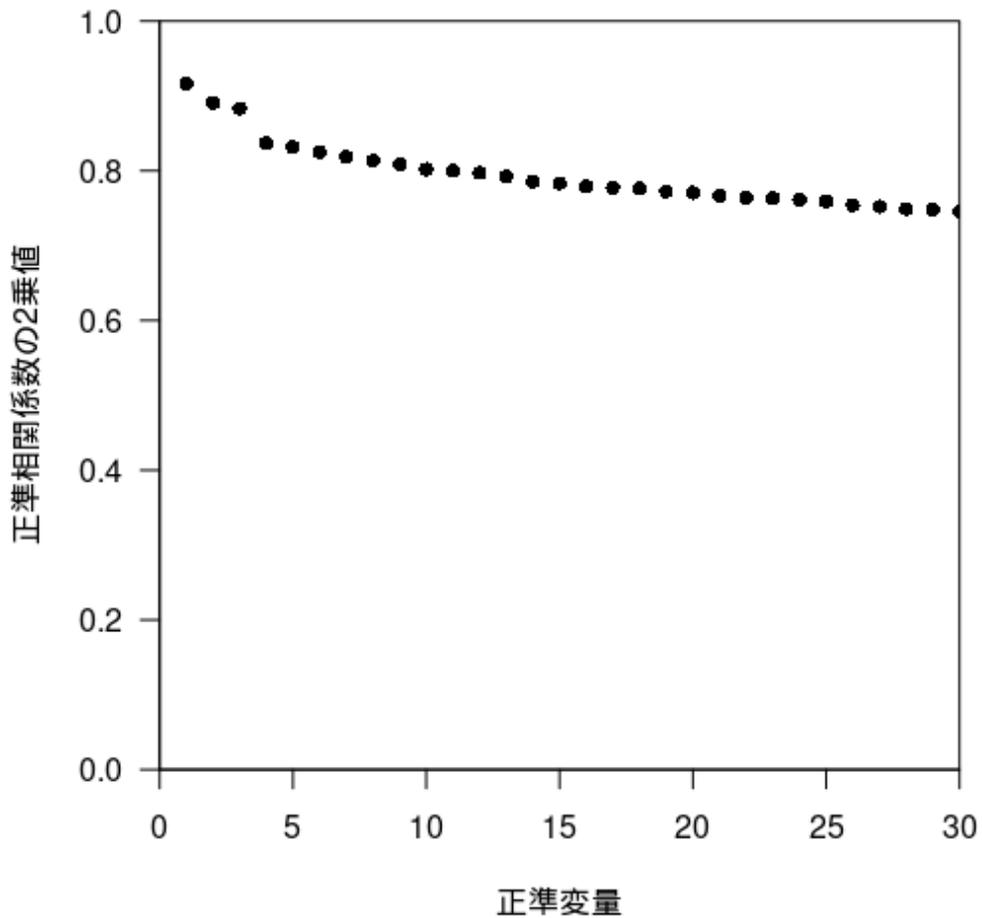


図 2.4 正準相関分析の寄与率

グループ B は HN120PCRDH、HN120MCRDH、および HN137PCRDH、グループ C は HN120PCR、HN120M、HN120MCR、および HN137P (HN137P 細胞の一部) で構成されていた。HN137P (HN137P 細胞の一部、左下のクラスター)、HN137M、および HN137MCR は、これら 3 つのグループに含まれていなかった。グループに含まれる細胞タイプに基づいて、グループ A、B、および C を、それぞれ薬剤感受性、薬剤休止、および薬剤抵抗性と見なした。先行研究 [6] では、データは 5 つのクラスターに分類された ([6] の図 2 (d))。先行研究 [6] では、薬剤休止の (HN120PCRDH、HN120MCRDH、および HN137PCRDH) のクラスターは 2 つのクラスターに分離されていたが、今回使用した手法では、薬剤休止の細胞は、同じグループ B に分類された。また、クラスタリングの結果

(図 2.5) では、グループ A：クラスター 1 および 5、グループ B：クラスター 4、6、8、グループ C：クラスター 0、2、3 に対応することが示された。HN137P 細胞に関しては 3 つのクラスター (2、5、10) に分離された。先述した先行研究 [6] では、HN137P は上記”上皮/基底パターン”の *ECAD/KRT5* だけでなく、管腔細胞のマーカである *KRT18* と間葉系細胞のマーカであるビメンチン (これより *VIM* と称す) の発現 (*KRT18/VIM*) を示す”管腔/間葉パターン”とが、不均一にみとめられる細胞株であった。間葉系の特性を持つ集団は、抗癌剤に抵抗性を持つことが示唆されるため、薬剤抵抗性の HN120PCR や薬剤抵抗性細胞と類似した表現型を持つことが考えられる転移性の HN120M、その薬剤抵抗性株 HN120MCR が含まれるグループにも、クラスターにも分かれたことが考えられた。本研究では、更なる解析を行うために、クラスター 10 を解析から除外した。

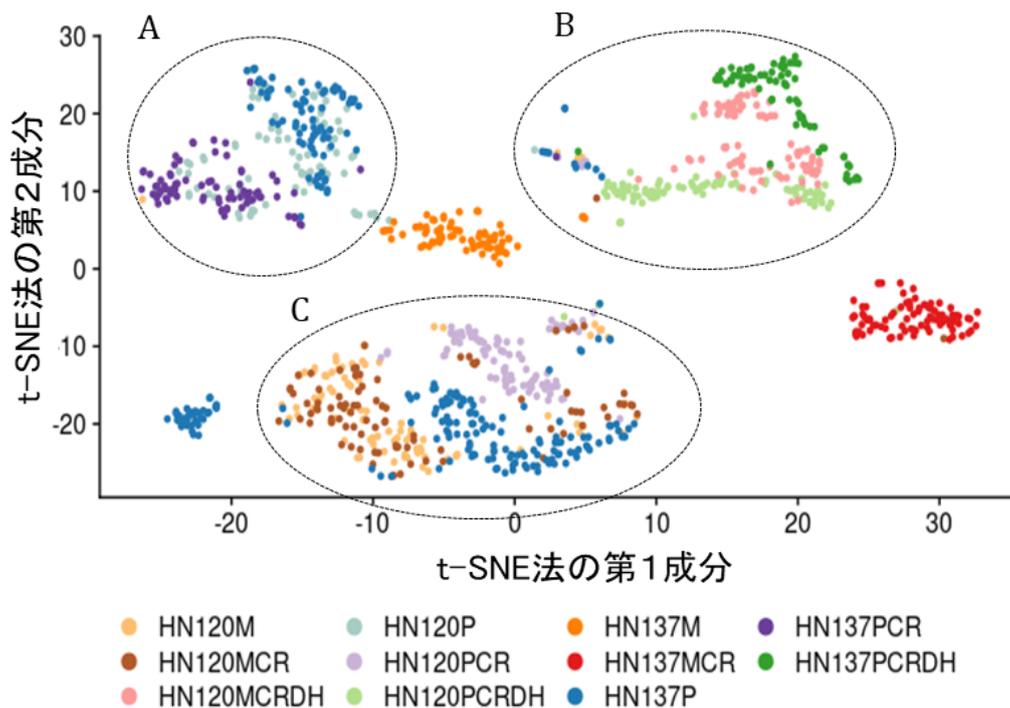


図 2.5 細胞タイプごとの t-SNE プロット

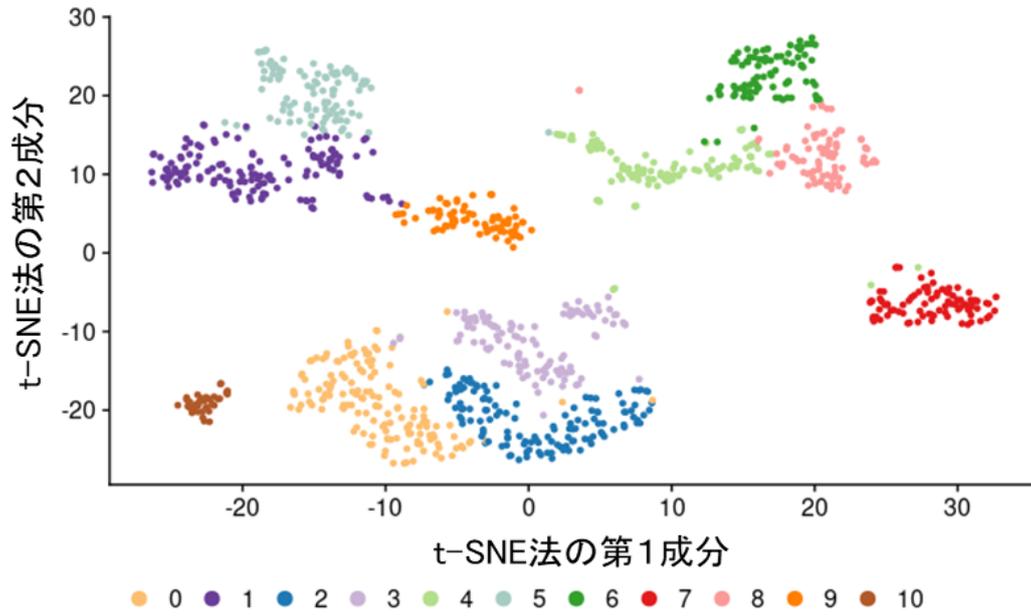


図 2.6 クラスターごとの t-SNE プロット

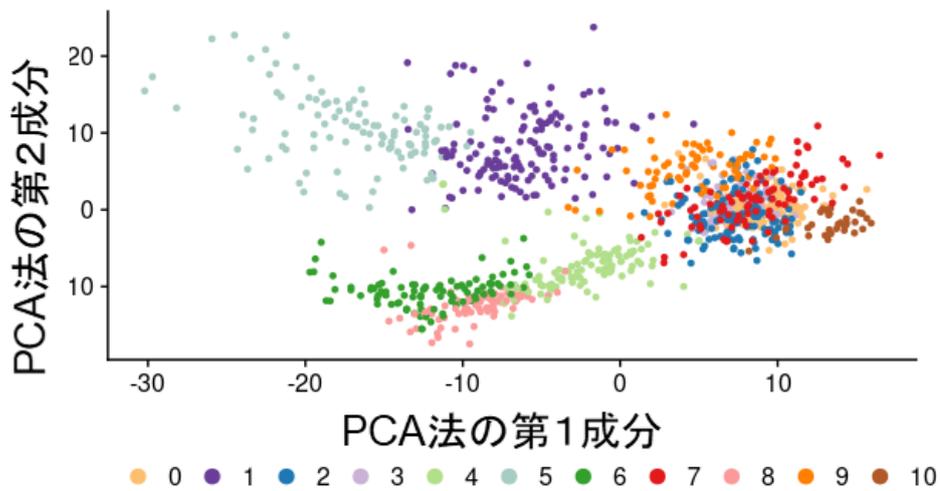


図 2.7 クラスターごとの PCA プロット

2.5.2 細胞系譜推定

本研究の手法により、感受性細胞株が含まれるクラスター 5 (HN120P および HN137P (部分的)) を開始クラスターとして設定することで、3つの系譜が推定された (図 2.8)。これらの系譜は、薬剤感受性から薬剤抵抗性 (系譜 1)、薬剤休止 (系譜 2)、および HN137MCR へ向かう転移性及び薬剤抵抗性 (系譜 3) の 3つの系譜が示された。興味深いことに、以前の先行研究 [6] ではそのような系譜は得られていなかったが、本手法により薬剤感受性から薬剤休止に向かう系譜 (系譜 2) が推定された。本手法では、上述した通り、出発点を HN120P/137P の 2つの薬剤感受性株が含まれるクラスター 5 に設定して各系譜を推定することにより、各薬剤の抵抗性株や、抵抗性後薬剤を休止して培養した細胞株のクラスターと連結したことで、薬剤の抵抗性に向けて、移行する系譜が推定できたと判断した。系譜は、感受性株より、薬剤抵抗性株である HN137PCR、転移性株である HN137M を経て、各々の系譜に分岐していた。系譜 1 では、その後、薬剤抵抗性株 HN120PCR、HN137P (間葉系の細胞が含まれることで薬剤抵抗性に関与している可能性が考えられる)、最終的に薬剤抵抗性細胞と類似した表現型を持つことが考えられる転移性の HN120M、その薬剤抵抗性株 HN120MCR が含まれるクラスターに接続した。系譜 2 に関しては、HN137M からの分岐後、薬剤休止株群が含まれるクラスターに接続した。これは、抵抗性後に、薬剤を除去して培養した場合に、抵抗性時とは異なる遺伝子発現の状態に変化していることが示唆された。系譜 3 に関しては、HN120PCRDH を経て、HN137MCR に接続していた。HN137MCR の細胞は、クラスタリングの際に、HN137M と離れた場所にあることを踏まえると、遺伝子発現のタイプが全く別の表現型に変化していることが示唆された。

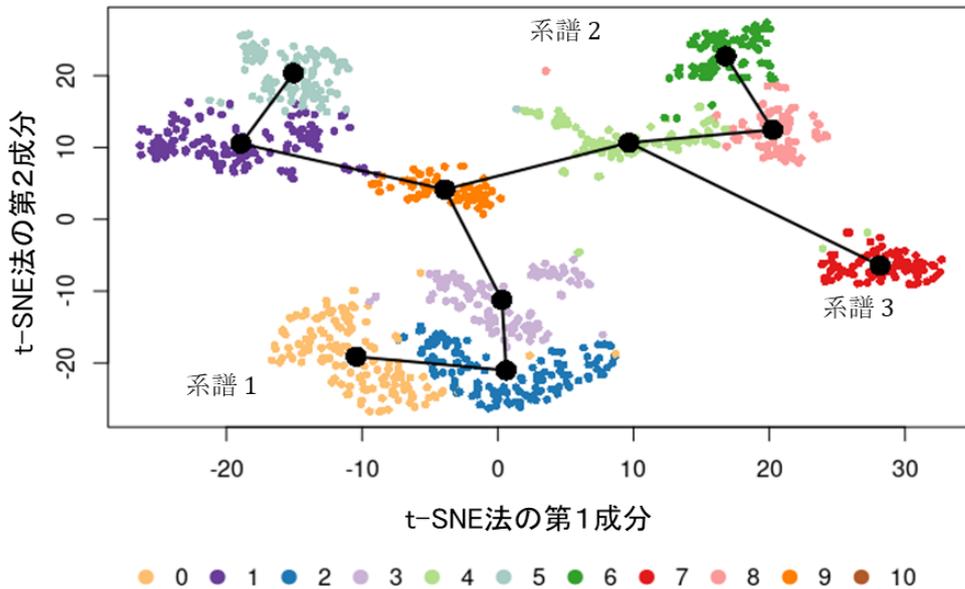


図 2.8 細胞系譜推定による系譜

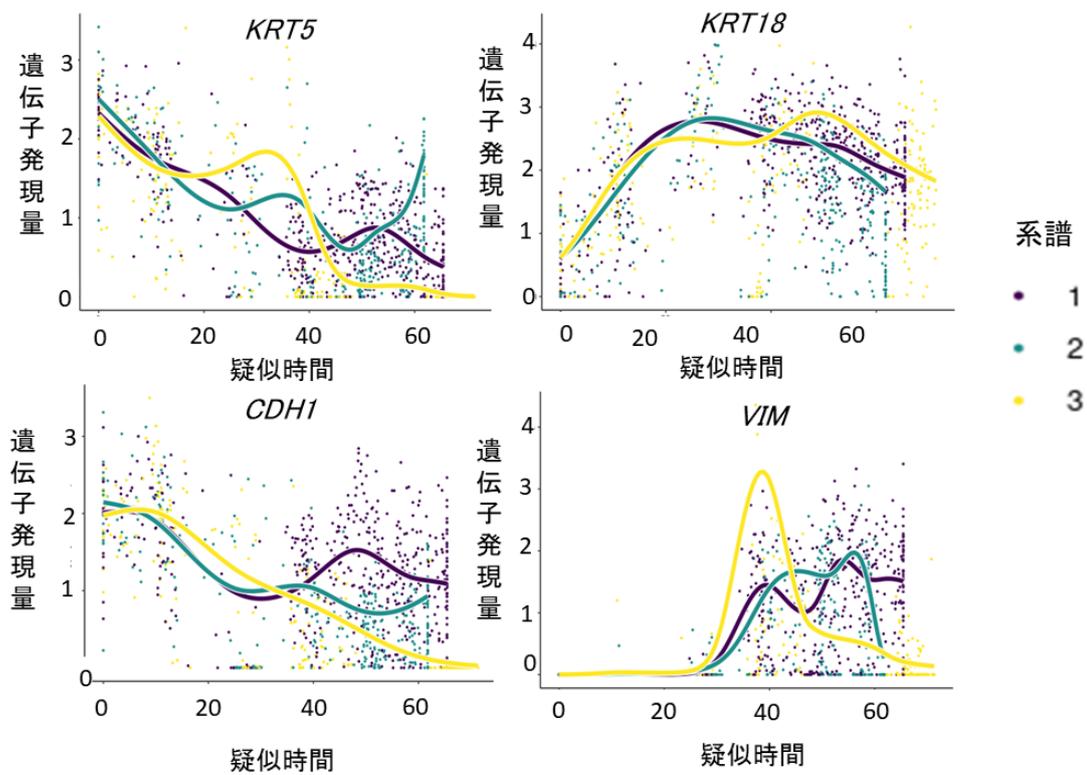
2.5.3 系譜に沿ったマーカーの発現の推定

本研究の手法では、2.5.2 節で推定した系譜を使用し、系譜に沿ったマーカーの発現を推定した。系譜に沿った連続的な遺伝子発現を計算するために、tradeSeq[20] を用い、一般化加法モデルを使用して、各系譜の疑似時間に沿った遺伝子発現を推測した。薬剤感受性、薬剤抵抗性、および薬剤休止間の系譜特異的变化を解析するために、基底細胞のマーカー (*KRT5*)、管腔細胞のマーカー (*KRT18*)、上皮系細胞のマーカー (*CDH1*)、および間葉系細胞のマーカー (*VIM*) の遺伝子の発現を調べた。

図 2.9 に示されるように、疑似時間は、クラスタリングの結果をもとにして、各クラスター上において最小全域木を用いて系譜を推定した上で、系譜に沿って主曲線を作成し、この主曲線上に沿って時間を定義し計算した。そのため、疑似時間は推定された系譜の結果に依存している。本研究では、疑似時間上での各点をもとにした遺伝子発現に対して、平滑化スプライン曲線を作成し、遺伝子発現の推移を評価した。推定された EMT 関連マーカーの推移を評価すると、疑似時間が進むにつれて上皮系細胞のマーカー (*CDH1*) と基底細胞のマーカー (*KRT5*) の発現は低下し、管腔細胞のマーカー (*KRT18*) と間葉系細胞のマーカー (*VIM*) の発現は上昇している傾向が認められ、上皮/基底から管腔/間葉への系譜に沿った発現の遷移が示され、細胞状態の遷移が明確に検出された。

興味深いことに、基底細胞のマーカー (*KRT5*) の発現は、ほとんどが疑似時間に沿っ

て減少するが、系譜 2（薬剤休止系譜）の終わりにのみ増加した。これは、細胞がシスプラチンの非存在下で薬剤感受性状態に戻ることを示唆している。更に、間葉系マーカー（*VIM*）は、系譜 3（転移性及び薬剤抵抗性への系譜）の疑似時間中点で高度に発現していた。これは、間葉状態への移行が転移と薬剤抵抗性の両面への影響を反映している可能性を示唆していると考えられた。本研究では、クラスター上の系譜推定に基づいた結果として、転移かつ薬剤抵抗性に関係すると考えられる系譜 3 における間葉系マーカー（*VIM*）の疑似時間上の中間地点での高度の発現が認められたため、シングルセル RNA-seq データを用いて、系譜推定を行ったことにより得られた知見である。



系譜 1：薬剤抵抗性、系譜 2：薬剤休止、系譜 3：薬剤抵抗性、転移性

図 2.9 EMT 関連マーカーの発現推移

2.5.4 系譜内および系譜間での発現差解析によるマーカーの検出

2.5.2 節で推定した系譜内および系譜間の発現について統計的検定を実施した。癌進行の細胞状態遷移のマーカーを検出するために、系譜に沿った始点と終点での発現量を比較した。遺伝子発現量の平均が、細胞系譜の始点（薬剤感受性集団）と終点（薬剤抵抗性、薬剤休止、転移性及び薬剤抵抗性）で差があることを検証するために startVsEndTest 関数を用いて Wald 検定を行った。

表は、テストによって検出された系譜 1（薬剤抵抗性：表 2.2）、系譜 2（薬剤休止：表 2.3）、および系譜 3（転移性及び薬剤抵抗性：表 2.4）の統計量が高い上位 5 つのマーカーを示している。参考として、系譜 1、系譜 2、系譜 3 の統計量の上位 50 位のマーカーを附録に添付した（表 4.1, 表 4.2, 表 4.3）。興味深いことに、系譜 1 と 3 で検出された統計量が最も高かった遺伝子はともに、*AXL* であった。また、*UBE2C* は、系譜 1 と 3 いずれにおいても 4 番目に統計量が高い遺伝子として検出された。更に、*UCA1* は、系譜 2 で 3 番目に統計量が高く、系譜 1 でも 6 番目に高い統計量であった。系譜 3 で 2 番目に高い統計量であった *MT2A* は系譜 2 では 8 番目に高い統計量であった。

表 2.2 系譜 1(薬剤抵抗性)

	waldStat 系譜 1	logFC 系譜 1
<i>AXL</i>	143.45	4.53
<i>SPC24</i>	96.59	3.94
<i>AAED1</i>	83.95	3.30
<i>UBE2C</i>	74.66	4.45
<i>NUSAP1</i>	69.82	3.43

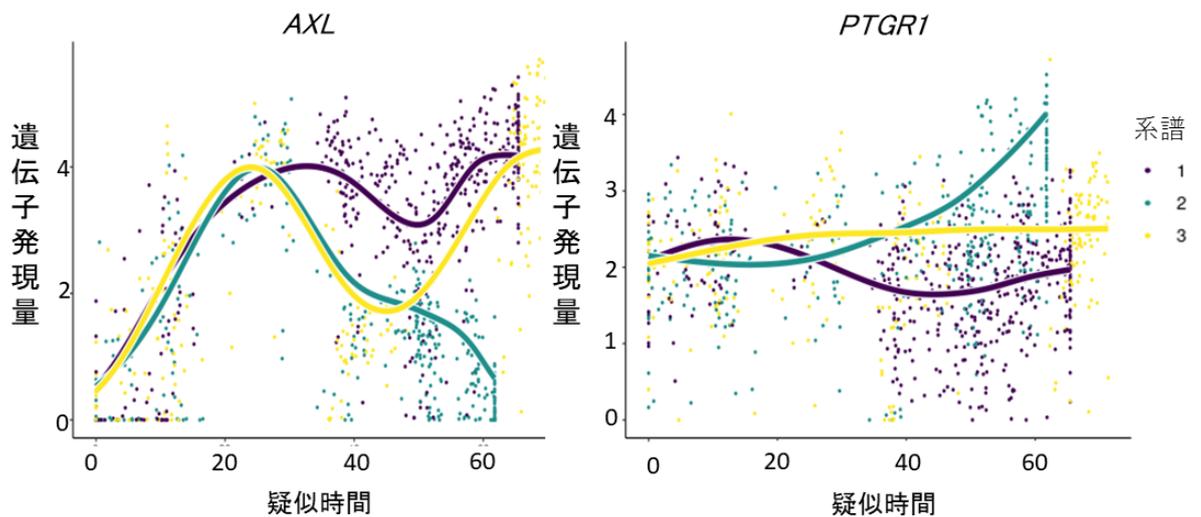
表 2.3 系譜 2(薬剤休止)

	waldStat 系譜 2	logFC 系譜 2
<i>PTGR1</i>	60.14	2.29
<i>ADA</i>	54.45	3.12
<i>UCA1</i>	38.91	3.84
<i>CCDC80</i>	28.49	5.18
<i>OLR1</i>	24.46	6.70

表 2.4 系譜 3(薬剤抵抗性、転移性)

	waldStat 系譜 3	logFC 系譜 3
<i>AXL</i>	166.88	4.64
<i>MT2A</i>	83.50	2.46
<i>HLA-B</i>	78.05	2.50
<i>UBE2C</i>	72.25	3.63
<i>HES1</i>	67.30	2.96

図 2.10 に統計量が最も高かった遺伝子である *AXL*、*PTGR1* の疑似時間に伴う発現量を示す。



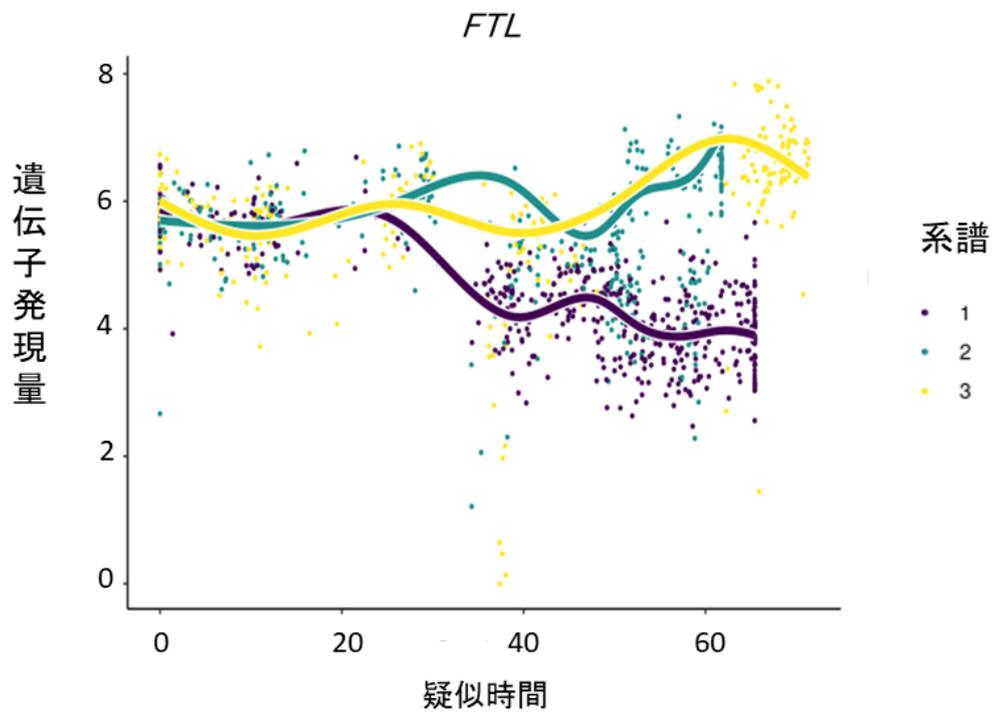
系譜 1：薬剤抵抗性、系譜 2：薬剤休止、系譜 3：薬剤抵抗性、転移性

図 2.10 系譜内解析によるマーカーの発現

また、系譜間での発現量の比較（系譜 1 と 3）も実施した（表 2.5）。本研究では、平滑化された遺伝子発現が疑似時間に沿って 2 つの系譜間で差があることを検証するために Wald 検定を行った。疑似時間に沿って均等に分散された 100 時点を、2 つの系譜間で比較した。結果を表に、統計量が最も高かった遺伝子である *FTL* の発現を図（図 2.11）に示す。

表 2.5 系譜間のマーカー

	waldStat 系譜 1 vs 3	fc Median
<i>FTL</i>	520.41	0.37
<i>HES1</i>	277.62	0.77
<i>F3</i>	251.58	0.41
<i>LIMCH1</i>	231.61	0.90
<i>CHP1</i>	213.41	0.46



系譜 1：薬剤抵抗性、系譜 2：薬剤休止、系譜 3：薬剤抵抗性、転移性

図 2.11 系譜間解析によるマーカーの発現

2.5.5 クラスターのマーカー解析との比較

本解析結果とクラスターの手法による解析結果を比較するために、Seurat[14] の FindMarkers による解析を用いて、クラスターの遺伝子発現を評価した。事前フィルタリングのオプションとして、発現差があるものかつ、2つの集団のいずれかで最低でも25%以上の細胞で発現しているものを使用した。図2.5のグループA（薬剤感受性）とグループC（薬剤抵抗性）におけるクラスター間での遺伝子発現変動解析によるマーカーを検出し、「2.5.4節」で検出したマーカーと比較した。

興味深いことに、*FTL*はクラスター間での遺伝子発現変動解析によるマーカーの検出でも検出されたが、一方で *AXL*はクラスター間での遺伝子発現変動解析によるマーカーの検出による結果リストに含まれていなかった。

更に、過去の研究結果より得られているシスプラチン抵抗性の関連遺伝子を文献よりまとめ [2][16][23]、提案手法とクラスター間での遺伝子発現変動解析で得られたマーカーとの比較を行った。今回の提案手法、またクラスター間比較で検出されたマーカーには、文献レベルで報告されている薬剤排出や解毒作用に関わる遺伝子は含まれていなかった。

これは、細胞系譜に沿ったマーカーの検出が、癌の薬剤抵抗性メカニズムを理解するための異なる手法を提供できることを示唆している。

2.6 考察

現在の標準治療に対する癌の抵抗性メカニズムを解明することは、癌研究において最も重要な領域の一つである。これまでにも多くの研究者が、従来のも得られた細胞のRNAの平均を評価するRNA-seq、そして単一の細胞ごとのRNAを評価するシングルセルRNA-seqを対象にマーカーを検出する手法の開発に取り組んできた。

まず、シングルセルRNA-seqデータが持つ遺伝子発現量のばらつきを踏まえた提案手法の妥当性に関しての考察を示す。シングルセルRNA-seqデータが持つ遺伝子発現量のばらつきには、ランダムな計測誤差と細胞の持つ遺伝子発現量のばらつきがある。ランダムな計測誤差の影響については、図4.1、図4.2を用いて、各細胞のRNA-seq遺伝子発現データの分布を示して考察した。これらの図の横軸は各細胞のRNA分子のカウント数、縦軸は少なくとも一つのカウントが確認された遺伝子の数である。縦軸の検出された遺伝子数と横軸のカウント数の間に、明瞭な相関関係があるため、大きな計測誤差を持つ細胞は、論文内で示した前処理で取り除かれたと考えられる。また、個別の遺伝子の発現量に

関しては、ランダムな計測誤差によると思われるばらつきはあるが、図 2.6 において、細胞系譜推定に利用したクラスタリングの結果では、各々別々のクラスターに分かれていることを踏まえ、生物学的な不均一性よりは小さいと考えられる。

本研究の手法では、正準相関分析を用いてバッチ効果の除去を行った上で、特定の系譜が薬剤抵抗性に向かって進行するにつれて発現が変動するマーカーを検出するために、系譜の遺伝子発現変動解析を実施した。先行研究の手法では、クラスタリングによって、各細胞の特性は評価しているものの、抵抗性因子に関与する候補因子の検出には、過去の知見より注目されている癌幹細胞仮説に基づく幹細胞関連因子より、候補が絞り込まれていた。疑似時間による細胞系譜は推定されていた一方で、活用は細胞起源の推定の活用にとどまっており、疑似時間によって推移する遺伝子発現の影響を考慮したマーカーの解析はなされていなかった。

今回用いた正準相関分析を用いた解析手法 [14] は、他のクラスタリング手法とバッチ効果の除去に関して比較した評価が行われており、高い精度を備えた上で、大きなサイズのデータを取り扱うことによる計算に伴う負荷が少ない点で有益性が示されている [24]。本研究においても、正準相関分析を用いることで、クラスターを細胞タイプとあわせて適切に分類することができたことを踏まえると、異なる条件の細胞の解析を行う場合、特に、本研究で取り組んだ抗癌剤への抵抗性機序研究には、正準相関分析が最適なクラスタリングに寄与することが考えられる。

また、本研究では細胞系譜推定の手法として、他の細胞系譜推定手法より高い精度で疑似時間を推定したことが示された最小全域木 (MST: Minimum Spanning Tree) を用いた [19]。本手法の前提は細胞タイプをクラスタリングにより適切な細胞集団に分けた上で、各細胞の遺伝子発現の連続的な変化を評価することにより、細胞系譜が存在することを仮定している。次元削減されたクラスターの結果を用いるため、まず薬剤抵抗性と薬剤休止の細胞株のクラスタリングでの細胞タイプの分類分けの成否が前提となる。本研究では正準相関分析を用いたことで、遺伝子の発現が生物学的に類似の細胞が適切にクラスタリングされたことで、最小全域木によるその後の各系譜の推定が、生物学的に意義のあるクラスターに準拠してなされたと考えられる。また、今回用いた最小全域木を構成する各枝における細胞の遺伝子発現量の変化は連続的であり、疑似時間に沿って、大きな変動はしておらず、単調に増加もしくは低下していることを仮定している。本仮定のもとでは、各クラスターの重心間での遺伝子発現量の差分が各枝の距離となり、最小の距離でクラスターを接続していくことで細胞系譜が正しく推定できると考えられる。更に、本手法にお

いては、細胞系譜の推定のために、系譜の始点をおくことによって、遺伝子発現の移行が生物学的に正しく推定できると考えられる。

また、系譜内及び系譜間の発現変動を求めるために使用した負の二項分布に基づく一般化加法モデルには、非線形の疑似時間につれて変動する遺伝子発現に連動する平滑な曲線を作成することによって、系譜内および系譜間で変動する遺伝子を検出できる手法であり、薬剤抵抗性因子の検出に向いていると考えられる。また別の疑似時間解析 [25] で必要とされるシーケンシャルラベルが不要であり、汎用性があると考えられる。

本手法では、これらの手法を組み合わせることによって、先行研究では示されていなかった薬剤抵抗性因子の候補マーカーが検出されたと考えられる。

今回の手法によって検出された各抵抗性因子に関して、個別に目を向けていくと、まず系譜内解析によって、受容体型チロシンキナーゼ *AXL* が、OSCC の標準治療の 1 つであるシスプラチンの選択圧下において薬剤抵抗性に関連する潜在的な候補であることを示した (表 2.2、表 2.4、図 2.10)。薬剤休止にあたる系譜 2 では、疑似時間経過とともに、発現の低下が認められていることを踏まえると、薬剤投薬によって生じる *AXL* の変化は、可塑性ではない点が考えられた (図 2.10)。そのため、*AXL* はシスプラチンの恩恵を長期に受けるために、シスプラチン投薬下にこそ、考慮すべき標的であると考えられる。

AXL は、TYRO3、*AXL*、MERTK (TAM) ファミリー受容体型チロシンキナーゼの 1 つであり、様々な癌腫における癌細胞の成長、生存、および転移を促進する上で重要な役割を果たしていることが示されている [26]。以前の研究に基づいて、*AXL* は、上皮成長因子受容体を標的とする抗体であり、頭頸部癌の治療基準の 1 つであるセツキシマブの抵抗性因子として知られている [27][28][29]。したがって、本結果に基づいて、*AXL* はシスプラチンおよびセツキシマブに対する潜在的な抵抗性メカニズムに関与している可能性がある。頭頸部癌では、シスプラチンとセツキシマブを含む治療薬剤投与計画 (通称: EXTREME) が最も一般的な標準治療であることを考慮すると、薬剤抵抗性を防ぎながら、より長くこれらの薬剤を投与していくために *AXL* を阻害する分子標的薬の組み合わせが有望である可能性がある。*AXL* を標的とする薬剤は近年開発が加速しており、他癌腫にて *AXL*/MET/VEGFR を阻害するマルチキナーゼ阻害剤の承認や、*AXL* 単独を阻害する薬剤開発も進んでおり、近い将来の承認が見込まれている。以上を踏まえると、他の癌腫で進んでいる *AXL* 阻害剤の開発が頭頸部癌に進んでいくことによって、治療効果の改善が期待できる可能性がある。

AXL は、他癌種である、食道癌 [30]、卵巣癌 [31]、悪性胸膜中皮腫 [32] においても活

性化されていることも報告されており、シスプラチンへの抵抗性を導くことが示されている。そして重要なこととして、これらの抵抗性時に AXL を阻害する薬剤を投与することでシスプラチンへの感受性が回復することが報告されている [30][31][32]。そのため、これらの知見を考慮すると、シスプラチンに AXL を阻害する戦略は、現実的に取りえる選択肢であることが考えられ、本研究の細胞系譜を用いた抵抗性因子検出手法が有効であることが示唆された。

また、AXL と同じく系譜 1 と系譜 3 で統計量が高かった *UBE2C* は、非小細胞肺癌 [33] や卵巣癌 [34] においてシスプラチンへの抵抗性を導き、*UBE2C* の阻害は、シスプラチンへの感受性を回復させることが報告されている [33][34]。AXL と同様に、シスプラチンへの抵抗性克服に寄与する可能性のある因子を見出したことで、本手法の頑強性が増したと考えられた。

更に、系譜内解析で検出されたマーカーの結果として、*PTGR1* の発現レベルが休薬期間中に増加することが示された (表 2.3、図 2.10)。*PTGR1* は、アラキドン酸代謝経路に関与し、主にプロスタグランジンやロイコトリエンを含むいくつかのエンコサノイド (生体機能の調整で役割を果たす生理活性物質) の不活性化に関与する律速酵素である [35]。*PTGR1* は、肝細胞癌、肺癌、前立腺癌、および膀胱癌などの細胞株で高発現しており、癌細胞の増殖を促進することが示唆されている [35, 36]。また、*PTGR1* の高発現は、頭頸部癌、乳癌、肺癌、胃癌、膵臓癌、肝臓癌における予後不良因子であることが示されている [35]。今回のシスプラチンを中止した後に生じるこの *PTGR1* の発現亢進は、薬剤投薬を中止した後の次の薬剤開発の標的になる可能性が示唆された。

また、系譜 1 と 2 の両方で統計量が高かった *UCA1* は、今回対象とした OSCC において、シスプラチンの抵抗性機序として報告されており [37]、OSCC における薬剤開発の標的として有望である。*UCA1* は、他癌種の非小細胞肺癌 [38] でも同様の報告があり、広くシスプラチンへの抵抗性機序である可能性が高い。また、系譜 2 と 3 の両方で統計量が高かった *MT2A* は悪性胸膜中皮腫に対するシスプラチンの抵抗性機序として報告されている [39]。

更に、系譜 1 (薬剤抵抗性) と系譜 3 (薬剤抵抗性と転移) を比較した系譜間の比較で検出されたマーカーは、*FTL* が転移と関連している可能性があることを示された (表 2.5、図 2.11)。*FTL* は長い間、鉄代謝調節因子の 1 つとして知られていたが、近年の研究成果により、*FTL* の高発現が、神経膠腫、結腸直腸癌、卵巣癌、胃癌、および急性骨髄性白血病を含むいくつかの癌の予後因子として報告されている [40, 41, 42, 43, 44]。更に、*FTL*

の高発現は、5-フルオロウラシル (5-FU) を含む化学療法抵抗性メカニズムに関与しており、細胞障害性抗癌剤の薬剤抵抗性で重要であると考えられる [41]。また、*FTL* は低酸素下で誘導され、EMT の調節因子であることも報告されている [40]。EMT は遊走性を向上させることで転移を促進しているため、本結果とあわせて考慮すると、*FTL* による転移促進の可能性が示唆された。

更に、EMT 自体がシスプラチンへの抵抗性を誘導することが報告 [45][46] されており、今回の解析において、上皮系から間葉系細胞への移行を確認できた点は、抵抗性の重要な要素となる EMT を細胞系譜推定により適切に評価した裏付けと考えられる。

本研究において、系譜内解析は、クラスター上で推定した系譜に沿った疑似時間を用いて、始点での遺伝子発現の平均が終点での平均と等しいという帰無仮説を評価するために実施している。一方、系譜間解析は、平滑化された遺伝子発現が疑似時間に沿って2つの系譜間で等しいかどうかを決定する統計的方法を使用している。以上より、系譜内解析と系譜間解析では、使用している目的と解析手法が異なるため、系譜1と3で見いだされた *AXL* が系譜間解析では、見いだされなかったことが考えられる。以上をまとめると、系譜内解析では、系譜に沿って、発現が上昇もしくは低下するマーカーを見出すことに強みがある。一方で、系譜間解析では、推定した系譜間で発現差があるマーカーを見出すことに強みがある。

また、過去の知見より得られているシスプラチン抵抗性関連遺伝子を用いたポジティブコントロールが、今回の提案手法、またクラスター間解析で得られなかった点に関して考察を行った。まず、最も代表的な抵抗性因子である薬剤を細胞外に排出するトランスポーター *ABCB1* (別称 *MDR-1*)、*ABCC1* (別称 *MRP-1*)、*ABCC2* (別称 *MRP-2*) に関して、どのような手法で見いだされたかに関して調べた。Nakatani ら [47] の解析では、シスプラチン感受性の口腔扁平上皮癌細胞株 (Sa-3) と薬剤抵抗性細胞株 (Sa-3r) 間において、抽出された mRNA を用いて、口腔内の病気に関係がある 2201 遺伝子をもとに作成された cDNA マイクロアレイを用いて発現差のある遺伝子を検出した。この解析により、抵抗性株で発現が高いと評価された遺伝子の一部に *ABCB1*、*ABCC1*、*ABCC2* が含まれていた。この結果を検証するために、リアルタイム PCR (RT-PCR) でこれらの遺伝子の mRNA の発現が、抵抗性細胞の方で高いという結果が確認された。また、Nakamura ら [48] の解析では、シスプラチン感受性の口腔扁平上皮癌細胞株 (H-1) と抵抗性細胞株 (H1-R) 間において、[47] の解析を同様の 2201 遺伝子をもとに作成された cDNA マイクロアレイを用いて、mRNA の発現量を評価し、*ABCB1*、*ABCC1*、*ABCC2* の発現が

抵抗性株で高いことが見いだされた。[47]と同じく、RT-PCRでこの結果は確認された。次のポジティブコントロールとして、抗アポトーシスタンパク質である *Bcl-2*、*Bcl-xl*、*Mcl-1* に関して、どのような手法で見いだされたかに関して評価した。抗アポトーシスタンパク質は、癌細胞死を抑制することで、薬剤抵抗性を導く。Majiら [49]の解析では、シスプラチン感受性の口腔扁平上皮癌細胞株 (H357 CisS, SCC4 CisS) と抵抗性細胞株 (H357 CisR, SCC4 CisR) 間において、抽出された mRNA を対象に RT-PCR で、タンパク質を対象にウェスタンブロット法で、発現を評価した所、いずれの評価系においても *Bcl-2*、*Bcl-xl*、*Mcl-1* の発現亢進が抵抗性株で認められた。以上、今回ポジティブコントロールとして使用した薬剤排出 (*ABCB1*、*ABCC1*、*ABCC2*)、抗アポトーシス (*Bcl-2*、*Bcl-xl*、*Mcl-1*) 関連因子は、それぞれ別々の薬剤感受性と薬剤抵抗性の細胞株を用いており、[47]、[48]、および [49] において、それぞれ別の遺伝子を抵抗性関連遺伝子として報告している。一方で、本研究で用いた遺伝子発現データは、[50]で独自に開発された癌患者由来の初代培養株から得られたものであり、[50]ではマイクロアレイでの遺伝子発現解析を行っているが、上記の遺伝子は抵抗性因子として報告されていない。このため、同じ口腔扁平上皮癌由来であっても薬剤抵抗性を示す細胞株は多様であり、ある細胞株での抵抗性因子が別の細胞株ではあまり発現していないということが起こり得ると考えられる。以上のことから、定性的な考察として、上記のポジティブコントロールとされた遺伝子が本研究の手法で検出されなかったのは対象として用いた細胞株の違いによるものではないかと考えられる。

2.7 結言

本章では、系譜が薬剤感受性状態と薬剤抵抗性状態の間で移行する際のマーカーの検出のための新しい手法を提案した。本手法は、単一細胞ごとのトランスクリプトームデータに基づいて連続的な状態遷移を解析でき、従来のクラスター間での遺伝子発現変動解析では検出されなかった潜在的なマーカー *UCA1*、*AXL*、*UBE2C*、*PTGR1*、*FTL*、*MT2A* を見出した。

UCA1 は既に OSCC におけるシスプラチンの抵抗性因子であると報告されており、更に、*UBE2C* は NSCLC や卵巣癌、*MT2A* は悪性胸膜中皮腫に対するシスプラチンの抵抗性因子であると報告されていることから、本手法の科学的妥当性が裏付けられたと考えられる。

また、今回新規で検出した *AXL* は今回対象とした頭頸部癌における標準治療の一つで

ある上皮成長因子受容体を標的とする抗体セツキシマブの抵抗性因子としても報告があり、頭頸部癌における *AXL* を標的とした手法が重要であることが示唆された。また *AXL* は、他癌種において、シスプラチンへの抵抗性を導き、抵抗性時に *AXL* を阻害する薬剤を投与することでシスプラチンへの感受性が回復することが報告されており、OSCC においても重要な薬剤標的になりえると考えられた。

また、*PTGRI* は、薬剤休止後に発現が亢進しており、薬剤投薬を中止した後の次の薬剤開発の標的になる可能性が示唆された。更に、系譜間解析により、*FTL* による転移促進の可能性が示唆され、その背景に EMT を介した機序が考えられた。

本章で実施した系譜特異的なマーカー解析の手法は、癌の進行中の細胞状態の変化のプロセスを理解するために有効な手段を提供することが考えられる。

癌細胞は巧妙に自身の特性と癌微小環境を変えながら、薬物治療から逃れていくため、経時的に変化する特性を捉えることによる抵抗性因子の検出は、今後の治療標的を考えていく上でも、製薬企業やアカデミアによる創薬につながりやすい有効な手法であると考えられる。

以上の点を踏まえると、経時的に特性を評価する手法は、他の癌腫、他の薬物治療による影響を評価するためにも、展開できる可能性がある。

第3章 シングルセル RNA-seq データを用いた上皮間葉転換に関与する系譜特異的マーカーの検出

3.1 緒言

癌研究においては、癌の進化に伴う特性の変化を検出することが、次の治療戦略を考える上で非常に重要である。上皮間葉転換 (EMT) は、癌の進行に関連する特性変化の移行において、よく知られた重要な要因であり、薬剤抵抗性の原因の 1 つでもあると考えられている [3][4]。近年、細胞の発達や刺激反応に応答する解析にシングルセル RNA-seq 解析が活用され始めており、経時的に変化する過程をたどると考えられている複雑な EMT のメカニズムを解明するためにも、適用されている。

たとえば、EMT 関連のマーカーを検出するために、クラスタリングと疑似時間解析を組み合わせた半教師ありの手法である Monocle も適用された報告がある [51]。また、教師なしの学習法 (QuanTC) は、scRNA-seq データの個々の細胞の遷移特性を推測および定量化するために提案された方法の一つである [52]。これらの方法を用いることで、EMT の単一細胞データセットからの細胞系譜推定の解析に適用することにより、細胞の中間状態によって媒介される細胞系譜推定を踏まえた EMT のダイナミクスが解析されている。

第2章で述べた研究手法では、各細胞のクラスタリング、細胞系譜推定、および遺伝子発現解析を組み合わせることにより、特定の系譜上において、薬剤抵抗性に向かって進行するにつれて発現が変動するマーカーを検出した [53]。また、細胞のタンパク質の発現を介したマスマイトメトリーの経時的解析を使用する別の単一細胞の手法は、TGF β 処理後の肺癌の EMT 状態を解析し、TGF β の投与中止によって明確な EMT が戻る MET 状態を検出した [54]。本章では、EMT が生じてからの経時的な変化・過程に着目して解析するための新しい方法を提案する。本研究の目的は、EMT を介した移行プロセスに沿って発現するマーカーを検出する手法の開発である。

3.2 先行研究

ここでは、本研究で取り組んだ系譜推定を用いた EMT 応答因子の検出手法との違いを説明するために、先行研究 [13] に関して、目的、使用されたシングルセル RNA-seq データ、解析手法、得られた結果に関して端的に述べる。

3.2.1 先行研究の目的

EMT はこれまでに広く研究されてきた領域であり、癌の転移に関わる細胞間結合に関する遺伝子を介した上皮性の低下等、様々な EMT に関する応答遺伝子が報告されてきた [13]。一方で、EMT を完全に定義する遺伝子発現プログラムが同定されていない状況にある。その背景に、EMT に関して古くは上皮状態と間葉状態の間の単純な変換として考えられていたが、近年は上皮系関連遺伝子と間葉系関連遺伝子が共発現する状態を考慮した部分的な EMT の概念も提唱されていることにある。このハイブリッドな状態には、癌幹細胞としての特性や、癌細胞の遊走を通して、転移との関連も示唆されている [55][56][57]。これまでに実施されてきた解析手法としては、マイクロアレイや、得られた細胞の RNA の平均で評価する RNA-seq 等による解析が用いられてきた。これらの評価の中で、EMT 誘導因子と細胞株を組み合わせた各条件において応答するマーカーの重複性が低い点が示唆されてきた [57][58][59]。すなわち、EMT は誘導する因子の種類や使用する細胞株によって応答遺伝子が異なる条件固有性の反応であることが考えられている。先行研究では、複数の癌腫の細胞株を対象に EMT 誘導因子を用いて実験的に EMT を誘導した単一細胞ごとの RNA-seq データを用いて、EMT 応答マーカーの比較を包括的にを行うことにより、この条件固有性の仮説を証明することが目的とされた。

3.2.2 先行研究のシングルセル RNA-seq データ

先行研究 [13] では、これまでに得られた非臨床研究の知見より、EMT を誘導するタンパク質 TGFβ1、EGF、および TNF の 3 つを用いて、EMT が引き起こされる 4 つの異なる癌細胞株 (A549/肺癌、DU145/前立腺癌、MCF7 乳癌、および OVCA420/卵巣癌) を用いた計 12 条件からなるシングルセル RNA-seq データを取得して解析が行われた。サンプルは、EMT 誘導因子による処理後 8 時間から 1 週間 (1: 「0d」、2: 「8h」、3: 「1d」、4: 「3d」、5: 「7d」) の 5 つの異なる時点で収集された (EMT が進む時間軸: 今後 EMT 進行 と称する)。更に、これまでの予備的な知見より、EMT 誘導因子除去後少なくとも 3 日の時点で発現する遺伝子のプロファイルが処理前の状態へ戻ることを考慮して、EMT 誘導因子除去後 8 時間から 3 日間 (6: 「7d + 8h オフ」、7: 「7d + 1d オフ」、8: 「7d + 3d オフ」) の 3 つの異なる時点でサンプルが収集された (EMT が戻る時間軸: 今後 EMT 反転 と称する)。まとめると、EMT 進行 5 つの時点、EMT 反転 3 つの時点の計 8 つの時点で総 576 サンプルより 58,808 のシングルセル RNA-seq のデータが収集

され、解析に利用された。

3.2.3 先行研究の解析手法

先行研究 [13] においては、Seurat[14] を用いて、上記記載した計 12 条件、8 つの時点のシングルセル RNA-seq のデータを対象に、各実際の時間的経過の進行に伴う発現遺伝子が評価された。

シングルセル RNA-seq には、10XGenomics 社のシステムが用いられた。10XGenomics Single Cell 3'RNA-seq キットで調製後、MULTI-seq バーコードライブラリが作られた [60]。その後、バーコードライブラリーが、10XGenomics ライブラリー調製プロトコルを用いて cDNA ライブラリーから分離され、PCR 増幅された [60]。最終的に、イルミナ社の NextSeq500 によって配列が決定された。

Seurat[14] で解析を行うにあたり、解析の質を担保するため、200 以上の遺伝子が検出されたシングルセル RNA-seq データを用い、ミトコンドリア遺伝子の発現が割合が高い細胞を取り除き、1 %未満の細胞でのみしか発現していなかった遺伝子も除去された。その後、遺伝子の発現量を標準のライブラリで正規化し、対数変換された。その後、各遺伝子の発現と平均と分散を用いて分散安定化変換を計算した上で、次元削減時の特徴量として用い、上位の変動 3000 遺伝子が使用された。次元削減には、Uniform Manifold Approximation and Projection(UMAP)[61] が用いられた。UMAP は、先述した t-SNE と同じくシングルセル RNA-seq データの次元削減を行うために、近年注目されている手法の一つである。非線形の類似データを低次元上でも近くに維持することを可能とする。t-SNE では、近さを確率と見なしたが、UMAP では、ファジー集合によって与えられるデータ間の距離から次元削減を行う。

各条件下での応答遺伝子の固有性の仮説を証明するために上記手法で変動した各時間経過時点の上位変動の 1000 遺伝子が Upset plot で評価された。

また、各条件下における応答遺伝子群が、過去の知見より得られた EMT hallmark 遺伝子セットと関連があるかを評価する目的で、Gene Set Enrichment Analysis (GSEA)[62] が実施された。

更に、EMT の時間的ダイナミクスを評価するために、psupertime[25] を用いて、各条件下における疑似時間による順位付けが行われた。psupertime は、シングルセル RNA-seq データを順次ラベルで取得し、指定されたラベル順序で細胞を配置する遺伝子の線形結合を識別する手法である。細胞に付与された物理的な時刻ラベルを基に、ロジスティック回

帰により各細胞を時刻ラベルにあてはめることにより、疑似時刻の推定を行う。疑似時間上の遺伝子発現に対して、一般化加法モデルを用い、平滑化した曲線が作成された。

3.2.4 先行研究の解析結果

先行研究の Seurat[14] を用いた解析結果では、各条件同様に、EMT 誘導因子での処理後、時間依存性に、細胞の遺伝子発現プロファイルは移行し (EMT 進行)、処理中止後、EMT 未処理の状態へ戻っていくことが示された (EMT 反転)。例として、TNF 刺激後の DU145 細胞のプロットを示す (図 3.1)。

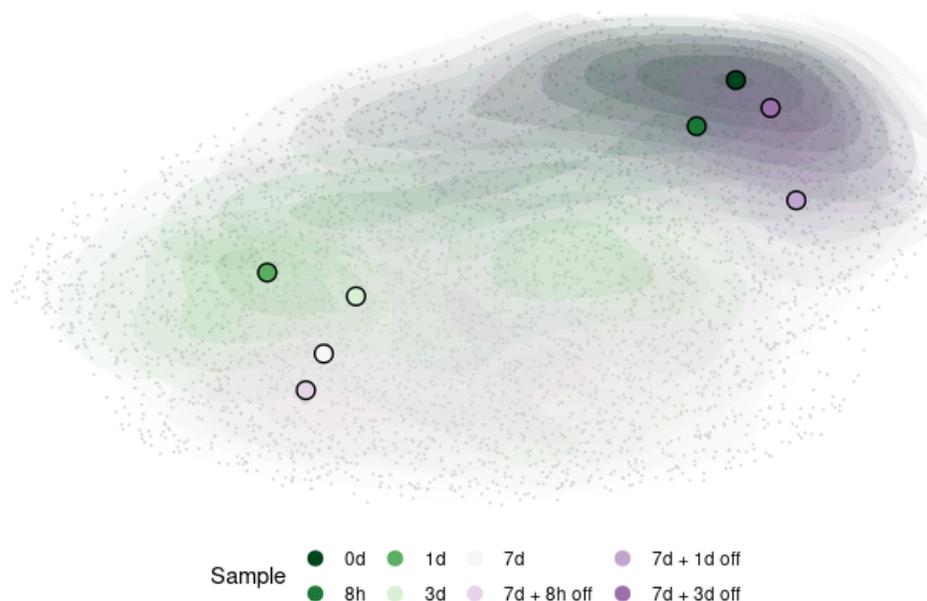


図 3.1 TNF 刺激後の DU145 細胞における時間経過における UMAP プロット

各時間経過時点の上位 1000 の変動遺伝子が評価では、細胞株を通して共通の発現パターンを示すものも認められたが、条件固有の遺伝子が優勢であることが示された。この変化を GSEA[62] を用いて、各時間経過時点の発現差がある遺伝子が、Molecular Signatures Database (MSigDB) の EMT の Hallmark 遺伝子セットと関係するかを評価した所、上記すべての条件で EMT 関連遺伝子セットの発現が高いことが示された。

以上の結果を踏まえ、EMT に関連する遺伝子セットは細胞の形態的变化にあわせて変化しているが、各条件間で変化する上位の遺伝子の重なるの低さを考慮すると、応答する関連マーカーは条件に固有であることが示された。3.2.2 節で述べた 12 の条件下の評価より、異なる 2 つの条件間で共有されているマーカーは平均して 2 割程度であった。また半

分以上のマーカーは、1~2つの実験系でのみ限定的に応答が認められることが示された。

また、EMTの時間的ダイナミクスを評価するために実施された psupertime[25] を用いた各条件下における疑似時間による順位付けを実施、その後の疑似時間系譜と実際の処理後の実時間との関係性を評価した結果により、EMT 誘導因子処理後 7 日間を通して前向きに発現プロファイルは移行し (EMT 進行)、処理中止後 3 日間で EMT 未処理の状態へ戻ることが示された (EMT 反転)。

3.3 研究の目的

先行研究では、肺癌・前立腺癌・乳癌・卵巣癌の細胞に上皮間葉転換 (EMT) を誘導する TGFB1、EGF、TNF による刺激後、EMT 進行 (8h, 1d, 3d, 7d)、EMT 反転 (7d+8h off, +1d off, 3d off) として経時的に評価したシングルセル RNA-seq のデータを対象として、Seurat[14] を用いて特徴量による遺伝子の絞り込みが行われた [13]。また、EMT の時間的なダイナミクスを評価するために、実際の時間経過をラベルにすることで、疑似時間の推移は評価されたものの、細胞系譜上に沿ったマーカーの検出は実施されていなかった [13]。

本研究においては、EMT 進行、EMT 反転に向かう連続的な遷移過程で特徴的な発現変動を示す遺伝子を検出するシングルセル RNA-seq データの解析手法の開発を目指す。本研究を行うにあたり、EMT 応答を、上皮系細胞と間葉系細胞の各々の特性という離散的な状態間の遷移ではなく、上皮系から間葉系、また間葉系から上皮系に向かう連続的な遷移過程ととらえることにより、細胞分化過程の解析で用いられている細胞系譜の推定手法を応用することを考えた。細胞系譜の推定手法は、遺伝子発現の推移を評価することを可能とする手法であり、上皮系から間葉系、間葉系から上皮系に向かう連続的な遷移過程で特徴的な発現変動を示す遺伝子を検出し、従来の離散的な状態間での発現変動解析では得られなかった、新たな EMT/MET 応答のマーカー遺伝子が得られるのではないかと考えるに至った。

3.4 実験材料と解析手法

3.4.1 使用細胞株とシングルセル RNA-seq データ

本研究では、先行研究 [13] のシングルセル RNA-seq データで Seurat[14] によってクラスタリングまで実施された結果を用いて、細胞系譜推定、細胞系譜に沿った遺伝子発現推定を組み合わせることで、EMT に関与する発現の移行に関連する因子を見出す手法の

開発を目指した。

シングルセル RNA-seq データは、格納先の Gene Expression Omnibus リポジトリ (GSE147405) から入手し使用した。

解析の手法としては、大きく 3 つに分かれている。(1) 推定されたクラスター間の距離を UMAP 空間上での距離として求め、最小全域木により細胞系譜を推定、(2) 系譜推定には EMT 進行と反転の両方に対して系譜を推定、(3) 系譜ごとの発現変動遺伝子を検出という流れで実施した。本手法の実装は R で行い、ライブラリとして (1)(2) では Slingshot[19]、(3) では tradeSeq[20] を使用した。

本研究では、Seurat[14] を用いたデータ統合とクラスタリング、次元削減が行われた先行研究のシングルセル RNA-seq のデータの結果を利用したので、細胞系譜解析以降の手法を実施した。

系譜推定には、Slingshot[19] を使用した。本研究で用いた手法では、クラスタリングの結果より得られた各クラスターのラベルを入力として使用し、クラスター上に最小全域木を構築した上で、これらのクラスターを系譜に順序付けする。出発点のクラスターのラベルを定義し、全ての系譜が出発点に指定したクラスターを共有し、最小全域木を構築している。各木内の経路に対する主曲線をフィッティングすることによって系譜を平滑化し、この主曲線に沿って、疑似時間を定義し計算を行う。

各系譜の連続的な遺伝子発現の推定、各系譜内の発現変動マーカーの検出には、tradeSeq[20] を使用した。一般化加法モデルは、遺伝子発現測定値を疑似時間の非線形関数としてモデル化する手法であり、一般化線形モデルの線形和の限界を改善し、説明変数の重み付き和から、各説明変数の一般的な変換の総和にしたものである。疑似時間に沿った遺伝子発現の推移を結ぶ曲線の作成を行うために、各時点の間を補完した曲線を 3 次関数で表し、平滑化された曲線として 3 次スプライン曲線を用いる。

tradeSeq では、系譜推定後の系譜に沿った疑似時間を用いているため、使用する疑似時間は系譜推定の手法と連動する。そのため、示される疑似時間は、Slingshot で推定された系譜に沿って定義され、計算されたものである。

3.5 結果

3.5.1 シングルセル RNA-seq データの統合とクラスタリング

本研究では、先行研究で実施された Seurat[14] を用いて解析、クラスタリングされたシングルセル RNA-seq のデータの結果を利用した。

本解析では、前立腺癌の細胞株 DU145 へ EMT を誘引する TNF による刺激後、EMT 進行 (8h, 1d, 3d, 7d)、EMT 反転 (7d+8h off, +1d off, 3d off) の系譜を推定した。本手法は 3 つのステップで構成されている：(1) 最小全域木 (MST) によって EMT とともに移行プロセスを検出するための細胞系譜推定を実施する。(2) EMT 進行に加えて EMT からの反転の系譜を推定する。(3) 発現変動解析により、EMT 進行系譜と EMT 反転系譜の両方のマーカーを検出する。この方法を利用することにより、EMT の特定の系譜に沿って発現が変動するマーカーを検出する手法を開発した。

3.5.2 細胞系譜推定

本研究では、Slingshot で細胞系譜推定を行った [19]。Slingshot による推論では、EMT 進行時の開始クラスター (緑の円) として「0d」を設定し、終了クラスター (赤の円) として「7d」を設定した (図 3.2)。更に、EMT 反転時では、「7d + 8h」を開始クラスター (緑の円) として設定し、「7d + 3d」を終了クラスター (赤の円) として設定した (図 3.3)。図 3.2 および図 3.3 に示すように、TNF によって刺激された DU145 前立腺細胞に関してそれぞれの系譜を推定した。これらの系譜は、EMT 進行の系譜では上皮から間葉へ、EMT 反転では間葉から上皮への時間的変化をそれぞれ示した。更に、EMT の時間的ダイナミクスの比較を通じて、各条件のセルを疑似時間的に順序付けした。他の条件でも各系譜を推定した (data not shown)。

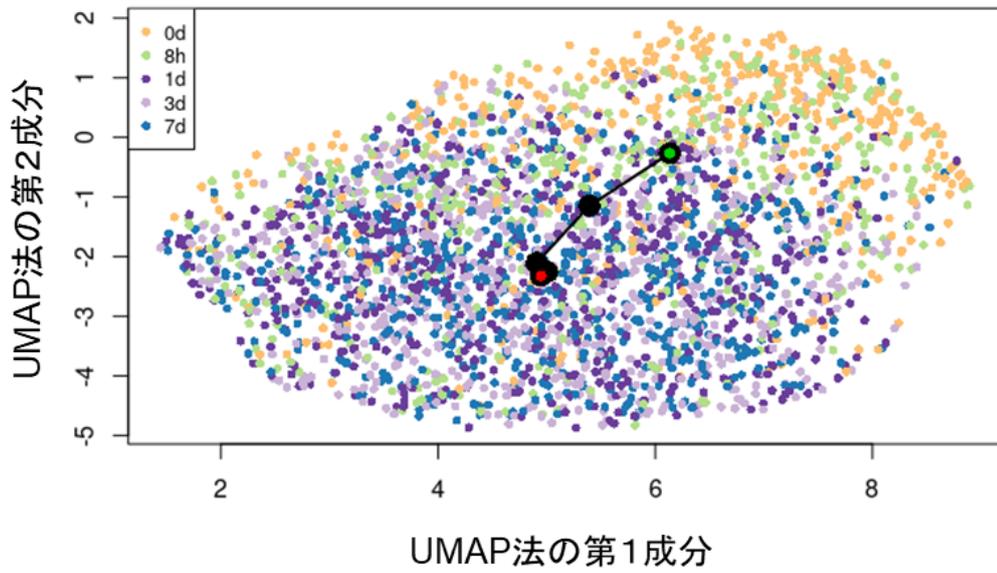


図 3.2 TNF 刺激後の DU145 細胞における EMT 進行の系譜

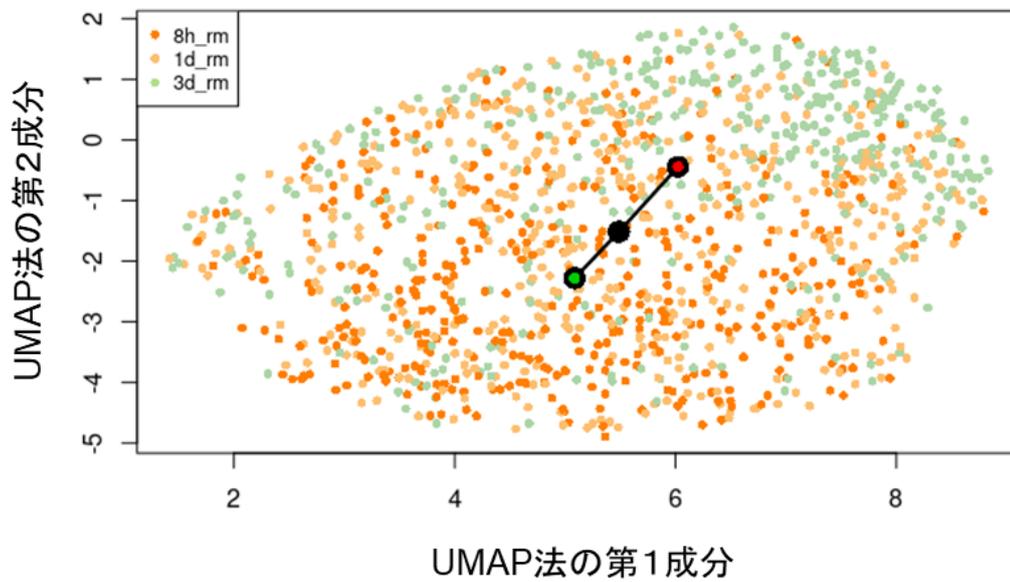


図 3.3 TNF 刺激後の DU145 細胞における EMT 反転の系譜

3.5.3 EMT 進行の系譜と EMT 反転の系譜に沿ったマーカーの検出

本研究では、EMT 進行系譜に沿って発現量が上昇し、EMT 反転系譜に沿って発現量が低下する遺伝子をマーカーと考えて検出する手法を開発した。3.5.2 節で推定した系譜を対象に、tradeSeq[20] を使用して、系譜に沿った遺伝子発現に関する統計的検定を実施した。遺伝子発現量の平均が、細胞系譜の始点と終点で差があることを検証するために startVsEndTest 関数を用いて Wald 検定を行った。検定後、始点と比較して終点で発現量が有意に変動 ($p < 0.05$) した遺伝子のうち、EMT 進行系譜では発現量が上昇した遺伝子、EMT 反転系譜では発現量が低下した遺伝子をそれぞれ検出した。表 3.1 と表 3.2 は、TNF 刺激 DU145 細胞の EMT 進行系譜と EMT 反転系譜において、本手法で検出されたマーカーのうち、Wald 統計量の上位 5 位までをそれぞれ示している。興味深いことに、インターロイキン (IL) -32 とリン脂質ホスファターゼ 4(*PLPP4*) が、2つの系譜の両方で統計的に重要なマーカーとして検出された。図 3.4 と図 3.5 は、それぞれ *IL32* と *PLPP4* の連続的な遺伝子発現を示している。以上の結果から、*IL32* と *PLPP4* が、上皮系から間葉系へ移行する状態 (EMT)、また間葉系から上皮系へ反転する状態 (MET) のいずれにおいても、重要な役割を果たしている可能性がある。参考までに、DU145 細胞への TNF 刺激後の EMT 進行、EMT 反転時の上位 50 の遺伝子を示す (表 4.4、表 4.5)。

表 3.1 TNF 刺激後の DU145 細胞における EMT 進行のマーカ-

	waldStat	logFC change
<i>IL32</i>	84.93	1.84
<i>PLPP4</i>	48.60	2.53
<i>FHL2</i>	44.62	1.42
<i>LTB</i>	40.01	2.29
<i>SFTA1P1</i>	38.78	1.28

表 3.2 TNF 刺激後の DU145 細胞における EMT 反転のマーカ-

	waldStat	logFC change
<i>ISG15</i>	76.76	-2.14
<i>IL32</i>	46.32	-2.02
<i>UBE20</i>	39.13	-1.47
<i>PLPP4</i>	24.66	-2.72
<i>WFDC2</i>	22.79	-2.10

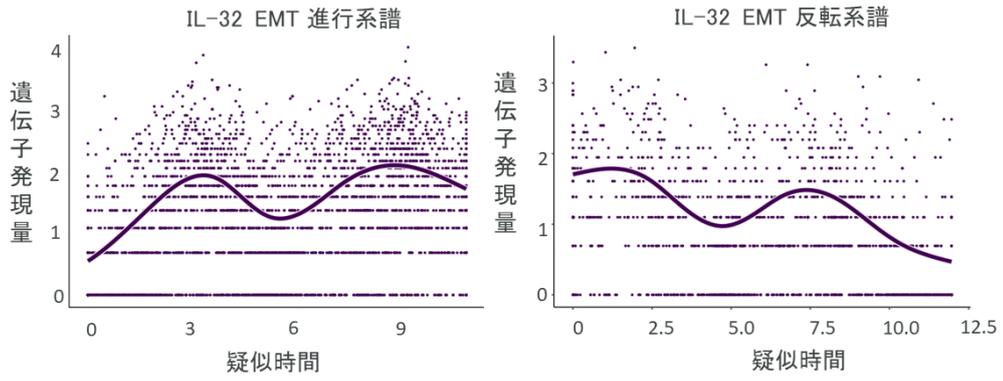


図 3.4 TNF 刺激後の DU145 細胞における各系譜に沿った *IL32* の変化

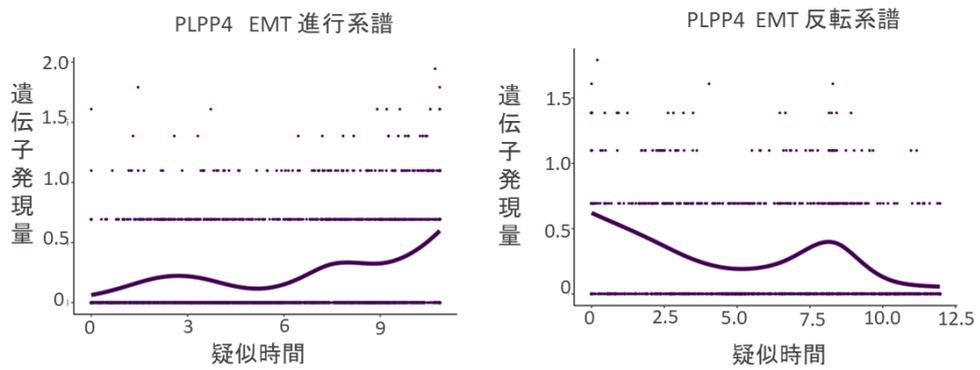


図 3.5 TNF 刺激後の DU145 細胞における各系譜に沿った *PLPP4* の変化

3.5.4 先行研究との比較

3.5.3 節で説明したマーカーを、Seurat[14] を用いたクラスター間で変動したマーカーと比較した。本研究では、TNF 刺激 DU145 細胞の EMT 進行と反転の間で検出されたマーカーを用いて、MSigDB の hallmarks の EMT 遺伝子セットを、標準的な EMT 関連遺伝子の陽性リファレンスとして適用し、評価した [63]。本解析には、縦軸に真陽性、横軸に偽陽性の割合をとる ROC (Receiver Operating Characteristic) 曲線を使用した。精度を評価するために、ROC 曲線下の領域の面積の大きさにあたる AUC (Area Under the Curve) もあわせて計算した。本解析を行うために、ROCR[64] を用いた。結果としては、本研究手法の TNF 刺激 DU145 細胞の EMT 進行 (図 3.6) は AUC が 0.58、Seurat のみを用いた評価 (図 3.7) では AUC が 0.57 であった。また、TNF 刺激 DU145 細胞の EMT 反転 (図 3.8) は AUC が 0.62、Seurat のみを用いた評価 (図 3.9) では AUC が 0.02 であった。

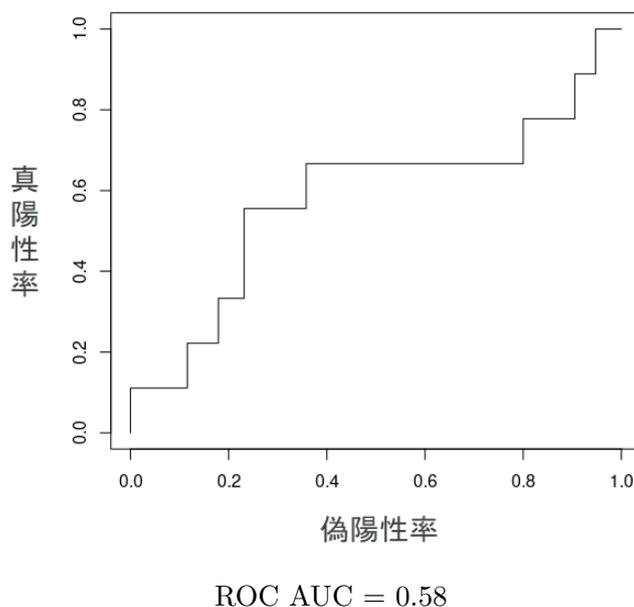
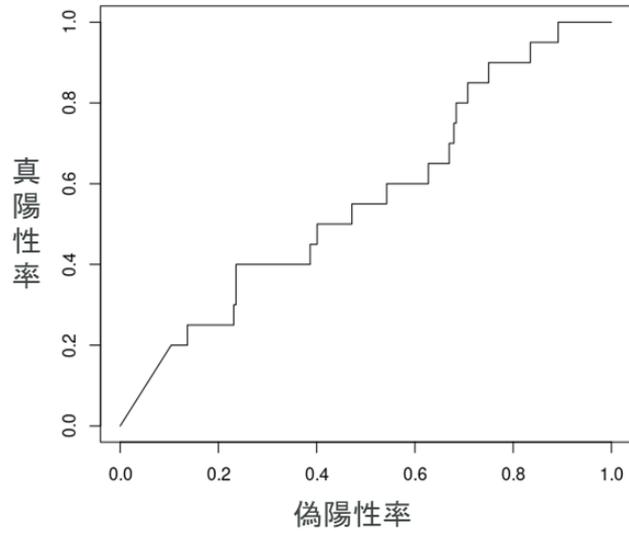
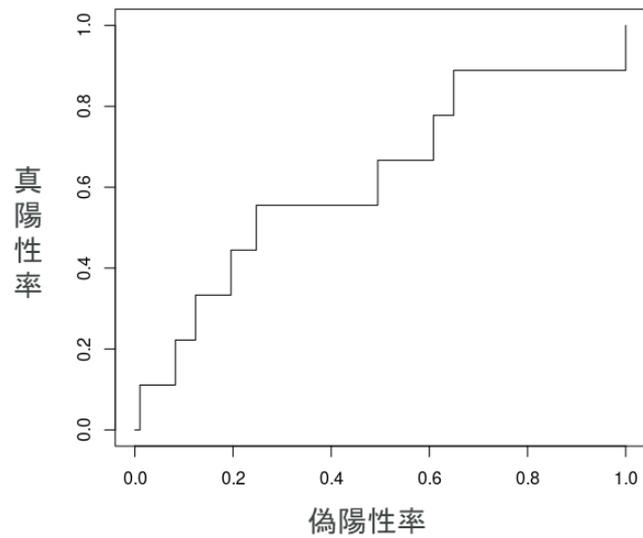


図 3.6 本研究手法で評価した TNF 刺激後の DU145 細胞の EMT 進行時のマーカーと EMT hallmark 遺伝子との関係性を評価した ROC 解析



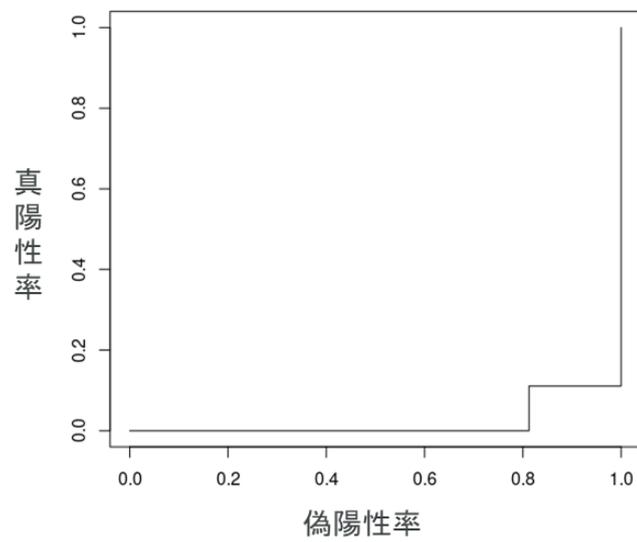
ROC AUC = 0.57

図 3.7 先行研究の Seurat で解析された TNF 刺激後の DU145 細胞の EMT 進行時のマーカーと EMT hallmark 遺伝子との関係性を評価した ROC 解析



ROC AUC = 0.62

図 3.8 本研究手法で評価した TNF 刺激後の DU145 細胞の EMT 反転時のマーカーと EMT hallmark 遺伝子との関係性を評価した ROC 解析



ROC AUC = 0.02

図 3.9 先行研究の Seurat で解析された TNF 刺激後の DU145 細胞の EMT 反転時のマーカーと EMT hallmark 遺伝子との関係性を評価した ROC 解析

3.6 考察

EMT の生物学を解明することは、次の癌治療薬を開発するための重要な研究分野の 1 つである。現在、多くの研究者が、EMT に関与する要因を検出するために、従来の RNA-seq およびシングルセル RNA-seq からの膨大な量のデータからマーカーを検出するための手法を開発しようとしている。

先行研究では、複数の癌腫の細胞株を対象に EMT 誘導因子を用いて実験的に EMT を誘導した単一細胞ごとの RNA-seq データを用いて、EMT 応答マーカーの比較を包括的に行うことでこの条件固有性の仮説を証明することを目的とされた。先行研究で用いられた、Seurat[14] の解析では、各遺伝子の発現と平均と分散を用いて分散安定化変換を計算し、次元削減時の特徴量として用いられた。その後、取得したセルサンプルの物理的な時点に沿って配置され、各条件から細胞を疑似時間的に順序付されたが、細胞系譜の推定が実施されておらず、系譜に沿った遺伝子発現変動によるマーカーの検出は行われていなかった。

先行研究において、EMT 誘導因子を導入してからの間葉系細胞への形態学的変化、並びに EMT 誘導因子除去後の上皮系細胞へ移行する形態学的変化が認められた。そのため、本手法は、EMT 誘導因子による刺激からの時間軸に応じて移行する細胞系譜が存在すると判断して評価を行った。解析では先行研究で実施された Seurat[14] によってクラスタリングされた結果を対象に、EMT 進行と反転両方に関して系譜を推定し、系譜内の疑似時間に沿った特異的な発現変動マーカーを検出する手法の開発を行った。

本手法の精度を定量的に評価するために、検出されたマーカーを、MSigDB の hallmarks の EMT 遺伝子セット [63] を標準的な EMT 関連遺伝子の陽性リファレンスとして適用し、ROC 曲線での AUC を計算した結果、TNF 刺激 DU145 細胞の EMT 進行は AUC が 0.58、TNF 刺激 DU145 細胞の EMT 反転は AUC が 0.62 といずれも 0.5 を越えており、本手法の精度の高さが裏付けられたと考えられる。

また、本手法により得られた具体的なマーカーとしては、*IL32* が、TNF で刺激された DU145 細胞において、EMT 進行時に発現が上昇し、反転時に発現が低下したことで検出された (図 3.4、表 3.1、表 3.2)。EMT 反転時に、統計量が高く発現が低下することから、EMT の逆の生理学的な変化である MET にも関与していることが考えられる。

IL32 は、炎症、自己免疫疾患、ウイルス感染、および癌に関連するサイトカインの一つである [65]。特に、*IL32* を分泌する癌関連線維芽細胞 (CAF) は、EMT マーカーの

発現を増強して癌の浸潤を促進することにより、乳癌細胞の浸潤と転移を促進することが報告されている [66]。

更に、本評価により、*PLPP4* は、TNF 刺激 DU145 細胞の条件における EMT の進行と反転の両方において検出された (図 3.5、表 3.1、表 3.2)。*PLPP4* は、膜内在性タンパク質スーパーファミリーに属し、細胞周期の進行を促進することにより、肺癌の増殖と腫瘍形成に関わる分子であると報告されている [67]。*PLPP4* の高発現は、進行した臨床病理学的特徴と正の関連があり、肺癌患者の生存率の低下と相関している [67]。更に、*PLPP4* は卵巣癌患者の腸転移を促進する重要な遺伝子の 1 つとして同定されている [68]。*PLPP4* はこれまで EMT 関連遺伝子の 1 つとして報告されていないが、これらの報告は、*PLPP4* が癌の進行に加えて EMT に寄与することを示唆している。

3.7 結言

本論文では、EMT に関連して系譜特異的に発現が変化するマーカーを検出するための手法を提案した。先行研究 [13] では、連続する時間経過の実験データと疑似時間を計算されていたが、細胞系譜推定は実施していなかった。本手法では、シングルセル RNA-seq データに基づいて EMT 誘導因子の投与後の連続的な状態遷移を評価することが可能である。本手法の精度を、標準的な EMT 関連遺伝子の陽性リファレンスとして適用して定量的に評価した結果、ROC 曲線での AUC は 0.5 を越えており、本手法の精度の高さが裏付けられた。また、本手法では以前の研究では検出されなかった潜在的なマーカー *PLPP4* を検出した。本研究の系譜特異的マーカーを検出する手法は、EMT の進行、そして EMT の反転いわゆる MET の応答マーカーを見出すために、遷移する EMT の遺伝子発現を捉えることで、既存の手法で見いだされないマーカーを検出することを可能とする。

第4章 結論

4.1 本研究のまとめ

本研究では、現代社会における死因の第一位である癌に対する薬物治療の開発を研究機関・製薬企業が行っていくために、薬剤抵抗性因子の検出を行う手法に関して、研究を実施してきた。

研究対象としては、近年癌微小環境の不均一性をより明瞭に捉える手法として期待されているシングルセル RNA-seq データを用いた。本研究では、特に、癌細胞の特性変化を捉える手法を開発すべく研究を行ってきた。癌細胞の特性変化に着目した背景としては、薬物治療への抵抗性因子の検出や、転移の促進並びに薬物治療の抵抗性に関与する EMT における応答因子の検出を行うことが、次の治療標的を決める一手となるからである。癌細胞は、遺伝子の発現変化を通して個々の特性を変え、薬剤抵抗性や EMT を生じさせている。

シングルセル RNA-seq 解析では、単一細胞ごとの遺伝子発現を評価することで、癌のような不均一な特性を持つ細胞を遺伝子発現の状態により分類することが可能である。これまでのシングルセル RNA-seq 解析では、これらの分類分けされたクラスター間で発現差がある遺伝子をマーカーとして検出する方法が一般的である。

また、細胞の分化や癌の進行において、分類された細胞を対象に細胞系譜を推定し、系譜に沿った疑似的な時間で移行する遺伝子発現の変化を解析することで、特性変化に関連する因子を検出する手法も現在までに開発されてきた。

薬物治療や EMT による癌細胞への影響は、各遺伝子発現の状態遷移を伴う動的な移行に伴った変化である。そのため、系譜推定を用いた手法を、薬物治療の抵抗性因子や EMT 応答因子の検出へ活用することを考えるに至った。

系譜推定を用いた薬物抵抗性因子や EMT 応答因子を検出する手法は、既報では限られているため、遺伝子の発現変化を捉える手法や新規の手法の開発は現在も強く求められている。

第2章では、口腔扁平上皮癌細胞における標準治療の一つであるシスプラチンを投薬して抵抗性が導かれた株、一定の休薬期間を設けて培養された細胞株より得られたシングルセル RNA-seq のデータを用いた薬剤抵抗性因子に着目し、細胞内の RNA の発現変化を系譜により推定する手法を開発した。本研究では、薬物治療による遺伝子発現の変化を

捉える系譜推定を行うために、まず細胞を分類するクラスタリングを実施した。クラスタリングを行うにあたり、本研究対象が複数の細胞株・異なる条件から得られたデータを対象にすることを考慮し、バッチ効果除去を見据えた、正準相関分析による複数サンプルデータの統合を行った。その後、細胞のタイプを分類するクラスタリングを実施し、クラスター上で細胞系譜を推定した上で、各系譜上の連続的な遺伝子発現をもとに疑似時間解析を行い、各系譜内および系譜間の発現変動マーカーを検出する手法を実施した。以上の一連の解析手法の組み合わせにより、薬物治療による影響を受けた癌細胞の特性の変化を捉えた。本細胞系譜推定を用いた手法により、従来のクラスター間での遺伝子発現変動解析では検出されなかった潜在的なマーカー *UCA1*、*AXL*、*UBE2C*、*PTGR1*、*FTL*、*MT2A* が見いだされた。先行研究においては、クラスタリングの結果のみを用いて、抵抗性遺伝子の評価が行われた。そのため、細胞特性が変化する過程の中で移行する遺伝子発現を踏まえたマーカーの検出は実施されていなかった。本手法では、系譜推定による疑似時間解析によって、遺伝子発現の推移を基に抵抗性因子を見出したことにより、既存手法では、見出されていなかったマーカーが検出されたと考えられる。

UCA1 は既に OSCC におけるシスプラチンの抵抗性因子であると報告されており、更に、*UBE2C* は NSCLC や卵巣癌、*MT2A* は悪性胸膜中皮腫に対するシスプラチンの抵抗性因子であると報告されていることから、本手法の科学的妥当性が裏付けられたと考えられる。

また、今回シスプラチンの抵抗性因子として検出された *AXL* は、対象とした口腔扁平上皮癌における標準治療の一つである上皮成長因子受容体を標的とする抗体セツキシマブの抵抗性因子としても報告がある [27][28][29]。以上より、*AXL* はシスプラチンおよびセツキシマブに対する潜在的な抵抗性メカニズムに関与している可能性がある。頭頸部癌では、シスプラチンとセツキシマブを含む治療薬剤投与計画（通称：EXTREME）が最も一般的な標準治療であるが、いずれ抵抗性を示し、効果が得られなくなる。薬剤抵抗性を防ぎながら、より長くこれらの治療を使用していくために、*AXL* を阻害する分子標的薬の組み合わせを行う薬剤開発を製薬企業が実施することが有望であると考えられる。

更に、*PTGR1* は、薬剤休止の系譜で疑似時間に沿って発現が亢進しており、薬剤投薬中止後も発現が上昇し続ける可能性がある。そのため、製薬企業が薬剤開発を行う際に、*PTGR1* が開発標的になる可能性が示唆された。更に、系譜間解析により、*FTL* の転移促進の可能性が示唆され、その背景に、細胞接着性が喪失した EMT を介した機序が考えられた。そのため、*FTL* を阻害する薬剤開発によって、EMT を介した転移の低減につな

がる可能性があると考えられた。

第3章では、EMT 誘導因子による刺激後の癌細胞より得られたシングルセル RNA-seq のデータを対象として EMT 応答マーカーを検出する手法に着目し、EMT の動的な遺伝子発現の変化を捉える手法を開発した。本研究では、遺伝子発現の変化を捉えるために、クラスター間の距離を基にして最小全域木による系譜を推定した。上皮系から間葉系に移行する EMT の進行と、間葉系から上皮系へ状態が移行する EMT 反転の両方向に対して系譜を推定し、系譜ごとの発現変動遺伝子の検出を試みた。本手法の精度を、標準的な EMT 関連遺伝子の陽性リファレンスとして適用して定量的に評価した結果、ROC 曲線での AUC は 0.5 を越えており、本手法の精度の高さが裏付けられた。

また、前立腺癌細胞株への TNF 刺激による EMT の進行と反転の系譜に沿って変動する *IL32* と *PLPP4* 遺伝子の変化を明らかにした。近年の研究成果により、上皮系から間葉系への細胞特性の変化いわゆる EMT は、細胞接着性の喪失による転移の促進を促す一方で、間葉系から上皮系へ細胞特性が変化する MET は細胞接着性が回復することによって転移先に生着することを促していることが考えられている。そのため、*IL32* と *PLPP4* 遺伝子は、これらの EMT と MET の変化にとって、重要な役割を果たしている可能性がある。

IL32 は EMT マーカーの発現を促すことで、乳癌の癌細胞を上皮系から間葉系へ細胞の特性を変化させることにより、細胞接着性が喪失することを通して、血流を介した別組織への転移を促すことが示されている。本研究の知見により、前立腺癌においても、*IL32* が EMT を介した転移に関して、重要な役割を示す可能性が示唆された *IL32*。以上の結果を踏まえると、EMT の進行と反転に対して、系譜特異的なマーカー解析の手法を行うことで、転移に関係する癌細胞状態の特性変化を評価できる可能性が示された。

これまでに述べてきた第2章、第3章で開発してきた系譜推定を用いた手法は、薬物治療の抵抗性因子や EMT 応答因子の検出において、推定した系譜上に沿った疑似時間を用いることで遺伝子発現の状態遷移からマーカーを見出すことにより、既存の手法では見いだされていないマーカーを見出すことを可能とする。

4.2 癌研究における本研究の寄与とシングルセル RNA-seq の可能性

本研究で取り組んだクラスター上に系譜を推定し、系譜に沿った疑似時間をもとにして遺伝子発現の変化を評価する手法は、薬物治療の抵抗性因子検出を目的としたシングルセル RNA-seq データの活用において、標準的に行われているクラスター間での遺伝子発現

変動解析によるマーカーの検出では得られない、経時的に変動する遺伝子発現に着目したマーカーを見出すことが期待される。

RNA の発現量を用いた疑似時間解析による連続的な細胞状態の遷移過程での遺伝子発現変動解析は、細胞の分化や癌の発生を対象とした解析で現在までに使用されるケースが増えてきたが、今回対象とした薬物治療への抵抗性因子、EMT 誘導時に関与するマーカーの検出等、遺伝子発現の変化に伴う特性の移行を捉えることを目的とした研究において強みを発揮すると考えられる。

特に、癌細胞は、巧妙に自身の特性と、癌が存在する微小環境を変化させながら薬物治療から逃れていくため、本研究で行った経時的に変化する特性を捉える抵抗性因子の検出は、今後の治療標的を考えていく上でも、遺伝子発現の変化の継時的な移行に着目したという観点で、既存の手法では見いだされないマーカーを検出する手法であると考えられる。本研究では、頭頸部癌の一つである OSCC に対するシスプラチンへの抵抗性因子の検出に用いたが、各癌種に対して、様々な薬物治療があることは 1 章の表 1.2、表 1.3 でも示した通りである。今後対象の癌種、治療の種類を細胞性抗癌剤から、乳癌・前立腺癌で使用されるホルモン療法、肺癌・乳癌・大腸癌等の特定の遺伝子異常を対象として開発された分子標的薬等の癌細胞に直接的に薬効を発揮するタイプの治療薬への新規の薬物抵抗性因子の検出に貢献できることが期待できる。見いだされた薬物抵抗性因子は、次の治療選択肢を作り出すための新しい創薬標的の開拓につながり、製薬企業やアカデミアの治療薬の研究開発が進展することで、予後改善や、根治に繋がる治療選択肢の開発に展開できる可能性を踏まえ、更に本領域の研究が進んでいくことが期待される。一方で、本手法の limitation としては、細胞のクラスタリングを行うことによって細胞集団を分けた上で、細胞系譜上の疑似時間による連続的な遺伝子発現の移行を評価しているため、解析対象が生命現象として、細胞系譜が存在している必要がある。そのため、細胞系譜が存在しないものを対象とした解析に本手法を用いることはできない。更に、本手法では、細胞系譜の推定に細胞系譜の始点を必要とするため、各細胞においてラベル、具体的には物理的な時刻、幹細胞、受精卵等の発生系、そして本研究の対象とした薬剤感受性/抵抗性、EMT が移行する時間の始点・終点といったものが必要になる。これらのラベリングに必要な情報がないデータを対象として、本手法を使用することはできない。一方でクラスタ間解析のみであれば、これらのラベリングとして必要とされる追加の情報を必要としないため、解析を行うことが可能である。

また、近年、癌研究領域における技術活用の潮流の中で、バイオインフォマティクスの

手法はますます重要な位置を占めている。したがって、日々進展する癌研究の中で、確実に存在感が増している。その中でも RNA の発現量を評価する手法は、マイクロアレイの時代から数千以上の遺伝子の発現を同時に評価することを踏まえ、多くの研究者の尽力によって、新規手法の提案や解析手法の改善がなされてきた。シングルセル RNA-seq はこういった本領域の発展を考えていくために、正確に単一細胞を単離して RNA の発現量を評価することに関しての技術進展だけでなく、得られる RNA 発現データに応じた解析手法に関しての開発・改善も進めていく必要がある [69]。1.1 節で述べた様々な種類の薬剤は、全く効かないケース、効果があっても短期で効かなくなるケース、また長期的に効いた後に最終的に抵抗性化するケースなど多様である。今回の研究により見出した、単一細胞レベルの遺伝子発現の変化をもとにした抵抗性因子の検出手法は、これらの標準治療に対する次の選択肢を考えていく上で、貢献できる可能性がある。

加えて、現在、シングルセル RNA-seq のデータ単独にとどまらず、解釈の質を高めるために、他の手法と組み合わせた動きも加速している。例えば、各細胞単位で T 細胞受容体/B 細胞受容体の発現を確認するシングルセルレパトア解析は、シングルセル RNA-seq 解析と組み合わせることで、癌微小環境における免疫細胞の多様性を検出することを可能とする手法である [70]。また、重金属ラベルされた抗体を用いて、単独の細胞単位でタンパク質の発現を評価するマスサイトメーターとの組み合わせも有望であると考えられている [71]。細胞によっては (NK 細胞等)、細胞タイプを決定するマーカーの発現レベルが、翻訳を制御する microRNA の影響や、ユビキチン修飾によるタンパク質の分解促進等の理由で、異なる場合がある [71]。生体に直接機能するタンパク質の評価は重要である一方で、現在単一細胞単位で同時に評価できるタンパク質の数は数十と限られている。そのため、数千から数万の遺伝子を同時に評価できるシングルセル RNA-seq の解析結果を組み合わせることで相互補完し、より結果を正しく解釈することが期待される。

これらの研究を通して見いだされた抵抗性因子への薬剤開発が行われることによって、現在予後の改善が満たされていない癌種への治療選択肢の増加や生存期間の延長、最終的には根治へ向けて、確実に医療が前に進んでいくことが見込まれる。

参考文献

- [1] 国立研究開発法人国立がん研究センター:がん情報サービス
<https://ganjoho.jp/public/index.html>.
- [2] T. ATiantian, R. L. Jie, Li and, W. Rongrong, Y. Liyan, L. Mengmeng, Z. Yiyiing, Z. Zhaowu, and X. Tian. Understanding the mechanisms of elemene in reversing drug resistance in tumor cells: A review. *Molecules*, Vol. 26, , 2021.
- [3] M. Nieto, R. Y.-J. Huang, R. A. Jackson, and J. P. Thiery. EMT: 2016. *Cell*, Vol. 166, No. 1, pp. 21–45, 2016.
- [4] D. Anushka and A. W. Robert. New insights into the mechanisms of epithelial-mesenchymal transition and implications for cancer. *Nature Reviews Molecular Cell Biology*, Vol. 20, pp. 69–84, 2018.
- [5] L. Zappia, B. Phipson, and A. Oshlack. Exploring the single-cell RNA-seq analysis landscape with the scRNA-tools database. *PLOS Computational Biology*, Vol. 14, No. 6, 2018.
- [6] A. Sharma, E. Y. Cao, V. Kumara, X. Zhang, H. S. Leong, A. M. L. Wong, N. Ramakrishnan, M. Hakimullah, H. M. V. Teo, F. T. Chond, S. Chia, M. T. Thangavelu, X. L. Kwang, R. Gupta, J. R. Clark, G. Periyasamy, N. G. Iyer, and R. DasGupta. Longitudinal single-cell RNA sequencing of patient-derived primary cells reveals drug-induced infidelity in stem cell hierarchy. *Nature Communications*, No. 9, 2018.
- [7] D. Grün, A. Lyubimova, L. Kester, K. Wiebrands, O. Basak, N. Sasaki, H. Clevers, and A. v. Oudenaarden. Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature*, Vol. 525, p. 251–255, 2015.
- [8] L. v. d. Maaten and G. Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, Vol. 9, pp. 2579–2605, 2008.
- [9] Y. Ge, N. C. Gomez, R. C. Adam, M. Nikolova, H. Yang, A. Verma, C. P.-J. Lu, L. Polak, S. Yuan, O. Elemento, and E. Fuchs. Stem cell lineage infidelity drives wound repair and cancer. *Cell*, Vol. 4, pp. 636–650, 2017.
- [10] P. Langfelder and S. Horvath. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, Vol. 9, p. 559, 2008.

- [11] D. Grün, M. J. Muraro, J.-C. Boisset, K. Wiebrands, A. Lyubimova, G. Dharmadhikari, M. v. d. Born, J. v. Es, E. Jansen, H. Clevers, E. J. P. d. Koning, and A. v. Oudenaarden. De novo prediction of stem cell identity using single-cell transcriptome data. *Cell Stem Cell*, Vol. 19, pp. 266–277, 2016.
- [12] X. Qiu, Q. Mao, Y. Tang, L. Wang, R. Chawla, H. A. Pliner, and C. Trapnell. Reversed graph embedding resolves complex single-cell trajectories. *Nature Methods*, No. 14, p. 979–982, 2017.
- [13] D. P. Cook and B. C. Vanderhyden. Context specificity of the EMT transcriptional response. *Nature Communications*, Vol. 11, p. 2142, 2020.
- [14] T. Stuart, A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W. M. Mauck, M. Stoeckius, P. Smibert, and R. Satija. Comprehensive integration of single cell data. *Cell*, Vol. 177, pp. 1888–1902.e21, 2019.
- [15] C. R. Leemans, B. J. M. Braakhuis, and R. H. Brakenhoff. The molecular biology of head and neck cancer. *Nature Reviews Cancer*, Vol. 11, pp. 9–22, 2011.
- [16] L. Kelland. The resurgence of platinum-based cancer chemotherapy. *Nature Reviews Cancer*, Vol. 7, pp. 573–84, 2007.
- [17] J. M. Siegle, A. Basin, A. Sastre-Perona, Y. Yonekubo, J. Brown, R. Sennett, M. Rendl, A. Tsirigos, J. A. Carucci, and M. Schober. Sox2 is a cancer-specific regulator of tumour initiating potential in cutaneous squamous cell carcinoma. *Nature Communications*, Vol. 31, p. 4511, 31.
- [18] S. Boumahdi, G. Driessens, G. Lapouge, S. Rorive, D. Nassar, M. L. Mercier, B. Delatte, A. Caauwe, S. Lenglez, E. Nkusi, S. Brohée, I. Salmon, C. Dubois, V. d. Marmol, F. Fuks, B. Beck, and C. Blanpain. Sox2 controls tumour initiation and cancer stem-cell functions in squamous-cell carcinoma. *Nature*, Vol. 10, pp. 246–50., 2014.
- [19] K. Street, D. Risso, R. B. Fletcher, D. Das, J. Ngai, N. Yosef, E. Purdom, and S. Dudoit. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics*, Vol. 19, No. 1, p. 477, 2018.
- [20] K. V. d. Berge, H. R. d. Bézieux, K. Street, W. Saelens, R. Cannoodt, Y. Saeys, S. Dudoit, and L. Clement. Trajectory-based differential expression analysis for single-cell sequencing data. *Nature Communications*, Vol. 11, No. 1, p. 1201,

2020.

- [21] A. Butler, P. Hoffman, P. Smibert, E. Papalexi, and R. Satija. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature Biotechnology*, Vol. 36, p. 411–420, 2018.
- [22] L. Haghverdi, A. T. L. Lun, M. D. Morgan, and J. C. Marioni. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nature Biotechnology*, Vol. 36, pp. 421–427, 2018.
- [23] J. Zhou, Y. Kang, L. Chen, H. Wang, J. Liu, S. Zeng, and L. Yu. The drug-resistance mechanisms of five platinum-based antitumor agents. *Frontiers in Pharmacology*, Vol. 20, p. 343, 2020.
- [24] H. T. N. Tran, K. S. Ang, M. Chevrier, X. Zhang, N. Y. S. Lee, M. Goh, and J. Chen. A benchmark of batch-effect correction methods for single-cell RNA sequencing data. *Genome Biology*, Vol. 16, p. 12, 2020.
- [25] M. Will and C. Manfred. Psupertime: Supervised pseudotime inference for single cell RNA-seq data with sequential labels. *Preprint at BioRxiv*, 2019.
- [26] R. M. A. Linger, A. K. Keating, H. S. Earp, and D. K. Graham. Tam receptor tyrosine kinases: biologic functions, signaling, and potential therapeutic targeting in human cancer. *Advances in Cancer Research*, Vol. 100, pp. 35–83, 2008.
- [27] T. M. Brand, M. Iida, A. P. Stein, K. L. Corrigan, C. M. Braverman, N. Luthar, M. Toulany, P. S. Gill, R. Salgia, R. J. Kimple, and D. L. Wheeler. AXL mediates resistance to cetuximab therapy. *Cancer Research*, Vol. 74, No. 18, pp. 5152–64, 2014.
- [28] N. K. McDaniel, M. Iida, K. P. Nickel, C. A. Longhurst, S. R. Fischbach, T. S. Rodems, C. A. Kranjac, A. Y. Bo, Q. Luo, M. M. Gallagher, N. B. Welke, K. R. Mitchell, A. E. Schulz, J. C. Eckers, R. Hu, R. Salgia, S. Hong, J. Y. Bruce, R. J. Kimple, and D. L. Wheeler. AXL mediates cetuximab and radiation resistance through tyrosine 821 and the c-ABL kinase pathway in head and neck cancer. *Clinical Cancer Research*, pp. 4349–4359, 2020.
- [29] H. Picon and A. K. Guddati. Mechanisms of resistance in head and neck cancer. *American Journal of Cancer Research*, Vol. 9, p. 2742–2751, 2020.
- [30] J. Hong, D. Peng, Z. Chen, V. Sehdev, and A. Belkhiri. ABL regulation by

- AXL promotes cisplatin resistance in esophageal cancer. *Cancer Res*, Vol. 73, pp. 331–340, 2013.
- [31] M. Tian, X.-S. Chen, L.-Y. Li, H.-Z. Wu, D. Zeng, X.-L. Wang, Y. Zhang, S.-S. Xiao, and Y. Cheng. Inhibition of AXL enhances chemosensitivity of human ovarian cancer cells to cisplatin via decreasing glycolysis. *Acta Pharmacol Sin*, Vol. 42, pp. 1180–1189, 2021.
- [32] D. B. Oien, T. Garay, S. Eckstein, and J. Chien. Cisplatin and pemetrexed activate AXL and AXL inhibitor BGB324 enhances mesothelioma cell death from chemotherapy. *Front Pharmacol*, Vol. 11, p. 970, 2018.
- [33] Y. Wu, D. Jin, X. Wang, J. Du, W. Di, J. An, C. Shao, and J. Guo. UBE2C induces cisplatin resistance via ZEB1/2-dependent upregulation of ABCG2 and ERCC1 in nslc cells. *J Oncol*, Vol. 1, p. 8607859, 2019.
- [34] J. Li, X. Zhi, X. Shen, C. Chen, L. Yuan, X. Dong, C. Zhu, and L. Y. a. Chen. Depletion of UBE2C reduces ovarian cancer malignancy and reverses cisplatin resistance via downregulating CDK1. *Biochem Biophys Res Commun*, Vol. 523, pp. 434–440, 2020.
- [35] X. Wang, G. Yin, W. Zhang, K. Song, L. Zhang, and Z. Guo. Prostaglandin reductase 1 as a potential therapeutic target for cancer therapy. *Frontiers in Pharmacology*, Vol. 6, p. 717730, 2021.
- [36] L. Xue, Z. Zhu, Z. Wang, H. Li, P. Zhang, Z. Wang, Q. Chen, H. Chen, and T. Chong. Knockdown of prostaglandin reductase 1 (PTGR1) suppresses prostate cancer cell proliferation by inducing cell cycle arrest and apoptosis. *BioScience Trends*, Vol. 10, pp. 133–139, 2016.
- [37] Z. Fang, J. Zhao, W. Xie, Q. Sun, H. Wang, and B. Qiao. LncRNA UCA1 promotes proliferation and cisplatin resistance of oral squamous cell carcinoma by sunpressing miR-184 expression. *Cancer Med*, Vol. 12, pp. 2897–2908, 2017.
- [38] C. Li, K. Fan, Y. Qu, W. Zhai, A. Huang, X. Sun, and S. Xing. Deregulation of uca1 expression may be involved in the development of chemoresistance to cisplatin in the treatment of non-small-cell lung cancer via regulating the signaling pathway of microRNA-495/nrf2. *J Cell Physiol*, Vol. 235, pp. 3721–3730, 2020.
- [39] S. Borchert, P.-M. Suckrau, R. F. H. Walter, M. Wessolly, E. Mairinger, J. Stein-

- born, B. Hegedus, T. Hager, T. Herold, W. E. E. Eberhardt, J. Wohlschlaeger, C. Aigner, A. Bankfalvi, K. W. Schmid, and F. D. Mairinger. Impact of metallothionein-knockdown on cisplatin resistance in malignant pleural mesothelioma. *Sci Rep*, Vol. 29, p. 18677, 2020.
- [40] J. Liu, L. Gao, N. Zhan, P. Xu, Ji' anYang, Fan' enYuan, Y. Xu, Q. Cai, R. Geng, Q. Chen. Hypoxia induced ferritin light chain(FTL)promoted epithelia mesenchymal transition and chemoresistance of glioma. *Journal of Experimental and Clinical Cancer Research*, Vol. 39, p. 137, 2020.
- [41] Z. Li, J. Liu, H. Chen, Y. Zhang, H. Shi, L. Huang, J. Tao, R. Shen, and T. Wang. Ferritin light chain (FTL) competes with long noncoding RNA Linc00467 for miR-133b binding site to regulate chemoresistance and metastasis of colorectal cancer. *Carcinogenesis.*, Vol. 41, pp. 467–477, 2020.
- [42] L. Wang, X. Li, Y. Mu, C. Lu, S. Tang, K. Lu, X. Qiu, A. Wei, Y. Cheng, and W. Wei. The iron chelator desferrioxamine synergizes with chemotherapy for cancer treatment. *Journal of Trace Elements in Medicine and Biology*, Vol. 56, pp. 131–138, 2019.
- [43] L. Zhang, Z. Chen, and A. Xu. FTL: a novel predictor in gastric cancer. *International Journal of Clinical and Experimental Pathology*, Vol. 10, p. 7865–7872, 2017.
- [44] S. Bertoli, E. Paubelle, E. Bérard, E. Saland, X. Thomas, S. Tavitian, M.-V. Larcher, F. Vergez, E. Delabesse, A. Sarry, F. Huguet, C. Larrue, C. Bosc, T. Farge, J. E. Sarry, M. Michallet, and C. Récher. Ferritin heavy/light chain (FTH1 /FTL) expression, serum ferritin levels, and their functional as well as prognostic roles in acute myeloid leukemia. *European Journal of Haematology*, Vol. 102, pp. 131–142, 2019.
- [45] M. Ashrafzadeh, A. Zarrabi, K. Hushmandi, M. Kalantari, R. Mohammadinejad, T. Javaheri, and G. Sethi. Association of the epithelial-mesenchymal transition (EMT) with cisplatin resistance. *International Journal of Molecular Sciences*, Vol. 21, p. 4002, 2020.
- [46] H. S. Choi, Y.-K. Kim, and P.-Y. Yun. Molecules in cisplatin-resistant human oral squamous cell carcinoma cell lines. *International Journal of Molecular Sci-*

- ences, Vol. 20, p. 3034, 2019.
- [47] K. Nakatani, M. Nakamura, K. Uzawa, T. Wada, N. Seki, H. Tanzawa, and S. Fujita. Establishment and gene analysis of a cisplatin-resistant cell line, Sa-3R, derived from oral squamous cell carcinoma. *Oncology Reports*, Vol. 13, pp. 709–714, 2005.
- [48] M. Nakamura, K. Nakatani, K. Uzawa, K. Ono, H. Uesugi, K. Ogawara, M. Shiiba, H. Bukawa, H. Yokoe, T. Wada, S. Fujita, and H. Tanzawa. Establishment and characterization of a cisplatin-resistant oral squamous cell carcinoma cell line, H-1R. *Oncology Reports*, Vol. 14, pp. 1281–1286, 2005.
- [49] S. Maji, O. Shriwas, S. K. Samal, M. Priyadarshini, R. Rath, S. Panda, S. K. D. Majumdar, D. K. Muduly, and R. Dash. STAT3- and GSK3 β -mediated Mcl-1 regulation modulates TPF resistance in oral squamous cell carcinoma. *Carcinogenesis*, Vol. 40, pp. 173–183, 2019.
- [50] S. Chia, J.-L. Low, X. Zhang, X.-L. Kwang, F.-T. Chong, A. Sharma, D. Bertrand, S. Y. Toh, H.-S. Leong, M. T. Thangavelu, J. S. G. Hwang, K.-H. Lim, T. Skanthakumar, H.-K. Tan, Y. Su, S. H. Choo, H. Hentze, I. B. H. Tan, A. Lezhava, P. Tan, D. S. W. Tan, G. Periyasamy, J. L. Y. Koh, N. G. Iyer, and R. DasGupta. Phenotype-driven precision oncology as a guide for clinical decisions one patient at a time. *Nature Communications*, Vol. 8, p. 435, 2017.
- [51] I. Pastushenko, A. Brisebarre, A. Sifrim, M. Fioramonti, T. Revenco, S. Boumahdi, A. V. Keymeulen, D. Brown, V. Moers, S. Lemaire, S. D. Clercq, E. Minguijón, C. Balsat, Y. Sokolow, C. Dubois, F. D. Cock, S. Scozzaro, F. Sopena, A. Lanas, NickyD’ Haene, I. Salmon, J.-C. Marine, T. Voet, P. A. Sotiropoulou, C. Blanpain. Identification of the tumour transition states occurring during EMT. *Nature*, Vol. 556, pp. 463–468, 2018.
- [52] Y. Sha, S. Wang, P. Zhou, and Q. Nie. Inference and multiscale model of epithelial-to-mesenchymal transition via single-cell transcriptomic data. *Nucleic Acids Research*, Vol. 48, p. 9505–9520, 2020.
- [53] K. Murayama and H. Matsuda. Detecting lineage-specific marker genes for tumor evolution based on single cell transcriptome. *International Journal of Bioscience, Biochemistry and Bioinformatics*, Vol. 11, No. 3, pp. 50–57, 2021.

- [54] K. Loukia G, B. Anchang, N. Ignatiadis, K. Samuel C, B. Jalen A, S. Joseph B, R. Tibshirani, B. Sean C, and P. Sylvia K. Mapping lung cancer epithelial-mesenchymal transition states and trajectories with single-cell resolution. *Nature Communications*, Vol. 10, p. 5587, 2019.
- [55] C. Kröger, A. Afeyan, J. Mraz, E. N. Eaton, F. Reinhardt, K. Yevgenia L, P. Thiru, B. Bierie, X. Ye, B. Christopher B, and R. A. Weinberg. Acquisition of a hybrid E/M state is essential for tumorigenicity of basal breast cancer cells. *Proceedings of the National Academy of Sciences Natl Acad Sci U S A*, Vol. 116, pp. 7353–7362, 2019.
- [56] F. Bocci, M. K. Jolly¹, J. T. George, H. Levine, and J. N. Onuchic. A mechanism-based computational model to capture the interconnections among epithelial-mesenchymal transition, cancer stem cells and notch-jagged signaling. *Oncotarget*, Vol. 9, pp. 29906–29920, 2018.
- [57] S. V. Puram, I. Tirosh, A. S. Parikh, A. P. Patel, K. Yizhak, S. Gillespie, C. Rodman, C. L. Luo, E. A. Mroz, K. S. Emerick, D. G. Deschler, M. A. Varvares, R. Mylvaganam, O. Rozenblatt-Rosen, J. W. Rocco, W. C. Faquin, D. T. Lin, A. Regev, and B. E. Bernstein. Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell*, Vol. 171, pp. 1611–1624, 2017.
- [58] P. Peixoto, A. Etcheverry, M. Aubry, A. Missey, C. Lachat, J. Perrard, E. Hendrick, R. Delage-Mourroux, J. Mosser, C. Borg, J.-P. Feugeas, M. Herfs, M. Boyer-Guittaut, and E. Hervouet. EMT is associated with an epigenetic signature of ecm remodeling genes. *Cell Death and Disease*, Vol. 10, p. 205, 2019.
- [59] H. José L, McFaline-Figueroa and Andrew J, X. Qiu, D. Jackson, J. Shendure, and C. Trapnell. A pooled single-cell genetic screen identifies regulatory checkpoints in the continuum of the epithelial-to-mesenchymal transition. *Nature Genetics*, Vol. 51, p. 1389–1398, 2019.
- [60] C. S. McGinnis, D. M. Patterson, J. Winkler, D. N. Conrad, M. Y. Hein, V. Srivastava, J. L. Hu, L. M. Murrow, J. S. Weissman, Z. Werb, E. D. Chow, and Z. J. Gartner. Multi-seq: sample multiplexing for single-cell rna sequencing using

- lipid-tagged indices. *Nature Methods*, Vol. 16, pp. 619–626, 2019.
- [61] L. McInnes, J. Healy, and J. Melville. UMAP: Uniform manifold approximation and projection for dimension reduction, 2018.
- [62] G. Korotkevich, V. Sukhov, N. Budin, B. Shpak, A. Maxim N, and A. Sergushichev. Fast gene set enrichment analysis. *Preprint at BioRxiv*, 2019.
- [63] A. Liberzon, C. Birger, H. Thorvaldsdóttir, M. Ghandi, J. P. Mesirov, and P. Tamayo. The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst*, Vol. 1, pp. 417–425, 2015.
- [64] T. Sing, O. Sander, N. Beerenwinkel, and T. Lengauer. ROCr: visualizing classifier performance in R. *Bioinformatics*, Vol. 21, p. 3940–3941, 2005.
- [65] A. Gautam and B. Pandit. IL32: The multifaceted and unconventional cytokine. *Human Immunology*, Vol. 82, pp. 659–667, 2021.
- [66] S. Wen, Y. Hou, L. Fu, L. Xi, D. Yang, M. Zhao, Y. Qin, K. Sun, Y. Teng, and M. Liu. Cancer-associated fibroblast (CAF)-derived IL32 promotes breast cancer cell invasion and metastasis via integrin β 3-p38 MAPK signalling. *Cancer Letters*, Vol. 442, pp. 320–332, 2019.
- [67] X. Zhang, L. Zhang, B. Lin, X. Chai, R. Li, Y. Liao, X. Deng, Q. Liu, W. Yang, Y. Cai, W. Zhou, Z. Lin, W. Huang, M. Zhong, F. Lei, J. Wu, S. Yu, X. Li, S. Li, Y. Li, J. Zeng, W. Long, D. Ren, and Y. Huang. Phospholipid phosphatase 4 promotes proliferation and tumorigenesis, and activates Ca²⁺-permeable cationic channel in lung carcinoma cells. *Molecular Cancer*, Vol. 16, p. 147, 2017.
- [68] A. Mariani, C. Wang, A. L. Oberg, S. M. Riska, M. Torres, J. Kumka, F. Multinu, G. Sagar, D. Roy, D.-B. Jung, Q. Zhang, T. Grassi, D. W. Visscher, V. P. Patel, L. Jin, J. K. Staub, W. A. Cliby, S. J. Weroha, K. R. Kalli, L. C. Hartmann, S. H. Kaufmann, E. L. Goode, and V. Shridhar. Genes associated with bowel metastases in ovarian cancer. *Gynecologic Oncology*, Vol. 154, pp. 495–504, 2019.
- [69] G. Sun, Z. Li, D. Rong, H. Zhang, X. Shi, W. Yang, W. Zheng, G. Sun, F. Wu, H. Cao, W. Tang, and Y. Sun. Single-cell RNA sequencing in cancer: Applications, advances, and emerging challenges. *Molecular Therapy - Oncolytics*, Vol. 8, pp. 183–206, 2021.
- [70] E. Azizi, A. J. Carr, G. Plitas, A. E. Cornish, C. Konopacki, S. Prabhakaran,

- J. Nainys, K. Wu, V. Kiseliovas, M. Setty, K. Choi, R. M. Fromme, P. Dao, P. T. McKenney, R. C. Wasti, K. Kadaveru, L. Mazutis, A. Y. Rudensky, and D. Pe'er. Single-cell map of diverse immune phenotypes in the breast tumor microenvironment. *Cell*, Vol. 23, pp. 1293–1308, 2018.
- [71] S. Nagasawa, Y. Kashima, A. Suzuki, and Y. Suzuki. Single-cell and spatial analyses of cancer cells: toward elucidating the molecular mechanisms of clonal evolution and drug resistance acquisition. *Inflammation and Regeneration*, Vol. 41, p. 22, 2021.

謝辞

本研究を遂行し、学位論文を作成するにあたり、多大なるご指導とご高配を賜りました大阪大学大学院情報科学研究科松田 秀雄教授に謹んで深謝の意を表します。

本論文の審査において、貴重なご指導とご助言を賜りました大阪大学大学院情報科学研究科 清水 浩教授、岡橋 伸幸准教授、古川 正紘准教授に謹んで感謝の意を表します。

本研究の進捗などについて、ご指導およびご助言を賜りました大阪大学大学院情報科学研究科瀬尾 茂人准教授に謹んで深謝の意を表します。

本研究を進めるにあたり、研究室内や研究科内での報告会にて貴重なご意見をいただきました大阪大学大学院情報科学研究科 繁田 浩功助教に心より感謝申し上げます。

本研究を行うにあたってお世話になりました大阪大学大学院情報科学研究科ゲノム情報工学講座の諸先輩方、学生諸氏、事務の方々の皆様に心から感謝の意を表します。

附録

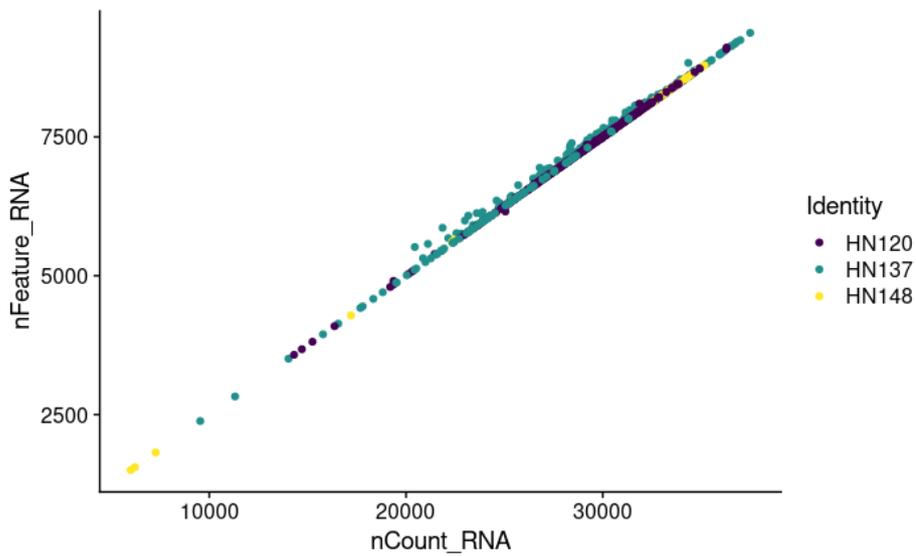


図 4.1 HN120、HN137、HN148 の遺伝子数とカウント数

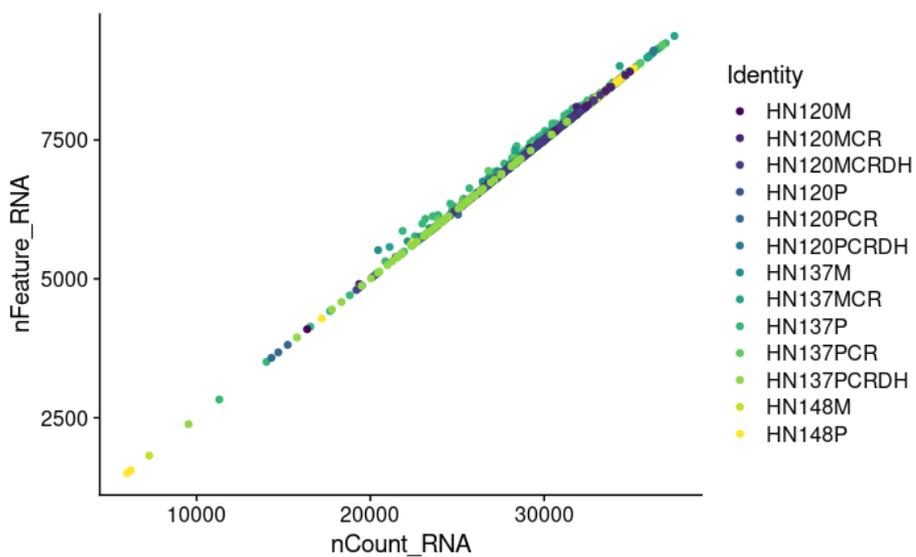


図 4.2 HN120、HN137、HN148 の各感受性、薬剤抵抗性、薬剤休止株の遺伝子数とカウント数

表 4.1 系譜 1(薬剤抵抗性) 上位 50 遺伝子

	waldStat 系譜 1	logFC 系譜 1
AXL	143.45	4.53
SPC24	96.59	3.94
AAED1	83.95	3.30
UBE2C	74.66	4.45
NUSAP1	69.82	3.44
UCA1	68.75	5.04
SAA1	60.87	4.86
HNRNPH1	56.76	2.04
FANCD2	55.92	3.00
KPNA2	53.99	2.64
SLC25A37	49.23	2.31
TPX2	45.05	2.82
CCNB1	41.35	3.17
CTNNB1	40.69	1.96
ADA	38.85	2.60
AURKA	37.87	3.15
CXCL1	36.84	5.56
CDK1	36.25	2.95
CCNB2	35.72	2.80
TINAGL1	35.64	2.24
EXOSC6	35.09	1.97
THBS1	34.59	2.83
KNSTRN	33.75	2.70
AFMID	33.25	2.32
AURKB	32.88	3.15
KIF23	32.10	2.78
TOP2A	31.37	2.77
ARHGDIB	30.92	4.12
UPP1	30.67	2.70
NDC80	30.48	3.36
INHBA	30.14	2.23
TAPBP	28.51	1.69
RRM2	25.73	2.63
TOR1AIP2	25.12	1.43
CDC20	24.79	2.93
KRT8	24.14	2.64
MT2A	23.00	1.70
KRT18	22.68	2.00
ZNF165	22.65	2.84
MOB3A	21.23	2.13
YWHAE	19.34	1.23
CDCA3	19.01	3.17
NEK2	18.97	3.32
HERPUD2	18.57	1.82
CDKN3	18.48	2.16
MROH1	18.38	1.93
NADK2	18.23	1.72
PARD6B	18.12	1.90
PSMC4	15.48	1.71
WTAP	15.37	1.46

表 4.2 系譜 2(薬剤休止) 上位 50 遺伝子

	waldStat 系譜 2	logFC 系譜 2
PTGR1	60.14	2.29
ADA	54.45	3.13
UCA1	38.91	3.84
CCDC80	28.49	5.18
OLR1	24.46	6.70
IL1B	24.09	3.04
HIST1H2BH	23.90	2.76
MT2A	21.79	1.95
SAA1	18.16	9.43
TMEM52B	18.14	3.15
NFKBIA	16.02	1.76
LOXL4	15.92	3.54
ADPRM	15.63	2.17
AK9	15.14	2.70
DEPDC4	15.11	3.82
PLAU	14.36	2.16
CCL20	13.80	4.42
TSPAN1	13.46	2.49
BTN2A3P	12.20	2.97
FTL	11.56	2.10
TM4SF1	11.25	1.70
PROSER2	10.66	3.07
SCARNA9	9.99	3.09
GAS6-AS2	9.85	3.60
TXNIP	9.36	3.32
CRYGS	9.32	3.85
EXOSC5	8.96	2.74
TCN2	8.90	5.17
CIDECP	8.81	2.98
CTGF	8.79	8.82
WFDC2	8.54	8.99
CSF2	7.77	2.75
PLAC8	7.60	2.52
HMSD	7.44	2.54
TMEM156	7.35	3.47
PUS3	7.27	2.61
MIRLET7D	7.19	2.85
HYKK	7.06	1.91
IL1R2	7.06	5.24
PFN4	6.88	5.36
BEX5	6.70	2.21
SCARNA8	6.70	4.04
LINC00525	6.63	4.47
TSC22D3	6.56	2.18
SNX12	6.49	1.46
OOEP	6.38	3.12
GBP2	5.89	2.67
ZNF565	5.82	3.47
DUSP5P1	5.80	3.49

表 4.3 系譜 3(薬剤抵抗性、転移) 上位 50 遺伝子

	waldStat 系譜 3	logFC 系譜 3
AXL	166.88	4.64
MT2A	83.50	2.46
HLA-B	78.06	2.50
UBE2C	72.25	3.63
HES1	67.30	2.96
MAP1B	50.29	3.37
PTTG1	45.02	2.36
CKS2	42.83	2.09
IL7R	42.19	4.23
SDCBP	42.19	1.62
AURKB	41.18	3.07
THBS1	40.89	2.77
UCA1	40.83	3.01
MT1L	38.09	3.33
KRT8	38.00	2.77
FN1	37.67	4.34
HIST1H1C	36.26	2.68
CSF2	35.86	4.28
SCARNA22	34.70	4.85
LIMCH1	33.77	3.09
SNAPC1	32.95	3.94
NDC80	31.63	3.00
KRT18	29.85	2.04
BASP1	29.59	2.39
DDIT3	29.40	2.49
MT1E	29.23	1.66
ANKRD10	28.49	1.92
KNSTRN	27.33	2.34
IL6	26.60	3.90
HIST1H2BG	23.88	5.81
ANTXR1	23.69	4.75
LINC00920	23.45	3.24
TPX2	22.86	2.07
CARD16	22.45	2.63
NUSAP1	21.74	1.98
ARHGEF28	21.08	3.13
CD320	20.15	2.64
HIST1H4E	19.49	3.71
HIST1H2AG	19.46	3.63
TIMP2	19.17	3.92
BIRC3	18.87	3.58
HIST1H2BN	18.86	2.86
CCNB1	18.77	2.15
SPC24	18.71	2.05
TOP2A	18.23	2.02
KRT81	17.98	3.65
VTRNA2-1	17.88	4.09
CPA4	17.76	3.21
UCHL1	17.58	1.74
HIST1H4H	17.31	4.91

表 4.4 TNF 刺激後の DU145 細胞における EMT 進行のマーカー上位 50 遺伝子

	waldStat	logFC change
IL32	84.93	1.84
PLPP4	48.60	2.53
FHL2	44.62	1.42
LTB	40.01	2.29
SFTA1P	38.78	1.28
SERPINA1	29.49	2.26
BCL2A1	28.32	2.28
CDA	25.84	1.16
SERTAD4-AS1	25.24	1.46
CDC42EP3	25.05	1.46
RP11-148B18.1	23.71	2.62
CST6	18.76	1.29
LAMC2	18.04	1.66
LGALS1	17.93	0.95
IL24	17.75	2.49
SERINC2	17.63	1.48
ZNF185	17.44	1.77
LETM2	16.56	1.77
RP5-1159O4.1	16.15	1.86
TNFAIP3	15.74	1.72
ANKRD13A	15.69	1.27
GPR87	15.65	2.99
TJP2	15.17	1.30
BSPRY	14.70	1.99
DSG2	13.80	1.15
TAGLN	13.76	1.56
VEGFC	13.03	2.08
NIPAL3	12.95	1.65
IGFBP6	12.64	1.17
IL11	12.23	1.28
C16orf74	12.08	1.16
SERTAD4	11.84	1.72
OCIAD2	11.66	0.92
TINAGL1	11.37	1.19
KLF6	10.58	0.99
LINC00704	9.94	1.92
TBC1D2	9.76	1.45
ZBED2	9.66	2.54
KRTAP2-3	9.20	2.90
TNC	9.16	1.50
CD24	9.13	1.04
EFEMP1	8.63	1.18
MFI2	8.54	1.16
MT1E	8.53	0.93
HMGA2	8.28	1.74
IFI27L2	8.23	1.06
WWC1	8.12	1.17
FLNC	8.00	1.10
MAL2	7.91	1.23
C1orf116	7.66	1.88

表 4.5 TNF 刺激後の DU145 細胞における EMT 反転のマーカー上位 50 遺伝子

	waldStat	logFC change
ISG15	76.76	-2.14
IL32	46.32	-2.02
ISG20	39.13	-1.47
PLPP4	24.66	-2.72
WFDC2	22.79	-2.10
MAL2	16.55	-1.71
TINAGL1	16.18	-1.49
SERTAD4-AS1	14.63	-1.53
LINC00704	14.12	-2.16
TNFAIP3	14.03	-2.20
RARRES3	13.70	-2.13
CD82	13.62	-1.97
ANKRD13A	13.37	-1.38
KRTAP2-3	13.11	-2.66
CAPG	12.74	-1.16
CSF2	12.67	-2.34
CDC42EP3	12.28	-1.47
RP5-1159O4.1	12.07	-1.85
OASL	11.46	-4.65
EFR3B	10.55	-1.53
HMGA2	9.88	-1.61
SFTA1P	9.86	-1.09
TAGLN	9.84	-1.71
ZNF185	9.64	-1.80
OPTN	9.49	-1.12
IGFBP6	9.17	-1.24
RNF144B	8.92	-2.67
IFI6	8.90	-1.45
LOX	8.85	-2.53
EVA1A	8.85	-2.79
EFNB2	8.81	-2.83
LETM2	8.80	-1.75
PTGER4	8.48	-2.65
SPOCD1	8.46	-1.91
PSMB8	8.43	-1.18
IFI44	8.24	-6.03
TJP2	7.80	-1.20
ATP8B1	7.55	-1.64
IL11	7.06	-1.54
IFITM3	6.94	-1.40
CCDC15	6.92	-1.52
PRSS3	6.90	-1.59
LGALS3BP	6.88	-1.87
GJB2	6.78	-3.46
HOXB8	6.71	-1.61
AC007405.6	6.66	-1.97
NIPAL3	6.63	-1.50
RP11-148B18.1	6.62	-1.86
KRT19	6.49	-1.00
KCNK1	6.29	-1.50