



Title	シングルセルRNA-seq を用いた細胞系譜推定と細胞間情報伝達推定による免疫疾患関連遺伝子の探索手法に関する研究
Author(s)	野村, 明弘
Citation	大阪大学, 2023, 博士論文
Version Type	VoR
URL	https://doi.org/10.18910/91999
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 を用いた細胞系譜
推定と細胞間情報伝達推定による免疫疾患
関連遺伝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 2023 年 1 月

野村 明弘

学位取得に関わる発表論文

国際学術論文

Akihiro Nomura, Hideo Matsuda. Identification of Lineage Markers for T Cell Immune Dysregulation in Sarcoidosis using Single-Cell RNA-seq, *International Journal of Bioscience, Biochemistry and Bioinformatics*, Volume 12, Number 4, pp.63-70 (Oct.2022) (博士論文第 2 章)

国際会議

Akihiro Nomura, Hideo Matsuda. Inference of Cell-Cell Communication for Detecting Immune Disease-Associated Signaling Pathways using Single-Cell RNA-seq, *International Conference on Biomedical and Bioinformatics Engineering (ICBBE 2022)*. 2023 (in press) (博士論文第 3 章)

内容梗概

アレルギー、リウマチといった疾患は、人間の持つ免疫システムが正常に働かなくなった状態と言い換える事が出来る。これらは一般的には免疫疾患と言われている。これらの疾患は QOL (Quality of life) に大きくかかわるため、今後、さらに QOL を改善するために、治療薬開発のための新たなターゲットの提案と提案されたターゲットに対する薬剤の研究開発が必要である。

薬剤開発の研究においては、何らかの手法で疾患の原因と考えられる蛋白質に関して、低分子薬剤や抗体により、制御することで治療とすることが一般的に行われている。したがって、何らかの手法により薬剤開発のための創薬候補を検出することが必要である。

数ある蛋白質の解析手法の内、発現の途中段階である転写産物の発現解析も必要であり、そのためには RNA シーケンス (RNA-seq) に基づくトランスクリプトーム解析が行われている。この技術は、生体内の遺伝子発現量を測定し、その翻訳産物である蛋白質の数を推定するために普及し、様々な研究に活用されている。近年、細胞分離技術やシーケンサーの性能向上により、1 細胞レベルで遺伝子発現を計測するシングルセル RNA-seq が利用できるようになった。この技術を活用し、健常人と患者あるいは患者に薬剤を投与したような試料をそれぞれ比較することで、発現量が変動する RNA を検出し、創薬のターゲットを探索することができる。また、シングルセル RNA-seq は疾患の根底にある遺伝子発現および遺伝的多様性に基づいて患者を層別化し、標的療法の開発に新たな知見をもたらすことができる可能性があると期待されている。

シングルセル RNA-seq 解析では、遺伝子発現量を個々の細胞レベルで評価できる。これまでの RNA-seq では細胞集団の平均値を算出することができたが、特定の細胞集団の遺伝子発現を個別に解析することは困難であった。このため免疫疾患などの異なる免疫細胞が関与するような複雑な疾患を、単一の細胞レベルで評価することが可能である。このようにして、免疫疾患に関わる遺伝子を、従来の細胞種をすべて平均化してしまうような RNA シーケンスよりもより詳細に評価することができるようになってきている。一方、シングルセル RNA-seq 配列決定の技術的進歩は目覚ましく大量の細胞からデータを取得できるようになってきているため、これまでよりも大量のデータを扱う必要が出てきている。そのため、さらなる効果的な解析手法の開発が求められている。

疾患と健常など遺伝子発現の異なる複数の細胞集団での状態を比較して解析する手法について、細胞種としてクラスリングを行った上で、それらのクラスター間のマーカーの発

現量の差の大きさを評価することが一般的に行われている。しかしながら、健常と疾患との比較を特定の細胞種間で行う方法の他に、免疫疾患のような複数の免疫細胞が関与する複雑な現象に関しては、免疫細胞の各々のサブタイプへの分化を考慮する必要がある。また、免疫疾患においては、発症から増悪、寛解、重症化、など様々な病態の段階があるため、なんらかの時系列を表現するような手法が必要である。このためには細胞の分化を系譜という細胞種間の移行過程を疑似的な時間軸に沿って表現する方法で比較し、遺伝子情報を検出するような手法が有効である可能性がある。

このような手法は、より病態を反映したような遺伝子を検出できる可能性があり、がんの抵抗性因子の獲得に関する細胞系譜の研究以外でも、適用できるような更なる手法の開発や改良が求められている。また、各々の細胞は何らかの方法で細胞間の情報伝達を行うことにより、機能を発揮すると考えられる。このため複数の細胞種のリガンドとそのターゲットの受容体の発現を測定し、疾患と健常などの状態の異なる細胞集団におけるそれらの変化を説明する細胞間の情報伝達を詳細に理解することが疾患の理解に必要であると考えられる。例えば、細胞系譜解析によるがんの抵抗性因子を検出する研究や、細胞間の情報伝達解析により皮膚炎の病変部、非病変部の比較による原因遺伝子を検出した報告はある。しかしながら、免疫疾患と健常など遺伝子発現パターン状態の異なる複数の細胞集団を対象とした細胞系譜解析や、細胞間の情報伝達解析でリガンドと受容体遺伝子ペアを解析し、疾患に関わる遺伝子を検出するような手法に関してはこれまで報告されていない。細胞間の情報伝達解析のための手法としては、免疫細胞と T 細胞の細胞間相互作用のあるシグナルパスウェイをデータベースとパスウェイモデルを構築して解析することが有効であると考えられる。また、アゴニスト、アンタゴニストの分子相互作用も考慮することにより、より生体の状態を反映した情報伝達の活性が理解できると考えられる。したがってこの手法の開発についても検討が必要であると考えられる。

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、免疫疾患におけるシングルセル RNA-seq 研究における背景及び本研究の目的と構成に関して述べる。第 2 章では、免疫疾患のシングルセル RNA-seq を使用した解析により疾患と健常を比べたときに発現差のある遺伝子を検出する手法を提案する。具体的には、疾患と健常など遺伝子発現パターン状態の異なる複数の細胞集団を対象としてシングルセル RNA-seq データを使用してクラスタリングによって細胞種を分類した上で、いくつかある免疫系細胞のうち免疫疾患に関わるとされ、自己免疫疾患や後天的な免疫の中心的な役割を果たすとされている T 細胞を抽出する。その上で T 細胞からさらなる分化が進んだ状態である、T 細

胞を由来とするそれぞれのサブタイプへの動的な分化過程を推定する。次に各系譜の遺伝子発現の推定、各系譜内および系譜間において疾患と健常の状態の異なる複数の細胞集団を比較することにより発現変動マーカーを検出する手法を提案する。

上記の手法を試みた結果、先行研究の細胞種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疾患関連遺伝子を検出した。これらの遺伝子はいずれも免疫疾患に関連することが知られている遺伝子であった。一方で、細胞系譜を使用しない先行研究の細胞種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たが、本研究において検出されなかった遺伝子も検出された。これらの検出されなかった遺伝子について、T 細胞に共通して発現している遺伝子であり、自己免疫疾患やがん免疫に関与しているものとして知られていた。今回の手法においては細胞種ごとにクラスタリングを行い、系譜間の発現差遺伝子を解析したことにより、T 細胞に共通して発現している遺伝子に関しては、検出されなかったと考えられる。また、T 細胞が分化して機能を果たすことができなくなった老化した T 細胞に発現する遺伝子が検出された。解析では分化の始めと終わりを考慮し、高発現する遺伝子を検出している。このため T 細胞種に関わらず終わりの状態のときに特に発現している遺伝子に関しても分化の進んだ系譜では発現していたと考えられる。

以上のことから、疾患に直接あるいは間接的に関わる遺伝子が検出されたことから、細胞系譜を用いた T 細胞のサブタイプへの分化を系譜として比較する解析手法が有効であることが示唆される。

第 3 章では、同様なシングルセル RNA-seq のデータを対象に、細胞間情報伝達解析による疾患において健常と発現差のある遺伝子を検出する手法を提案した。免疫においては、細胞種ごとにリガンド-受容体のやり取りを通じて、その機能を発揮している。その一方、疾患においてはそのやり取りが健常と比べて変化が生じていると考えられる。この結果、免疫の不具合が起きていると考えられる。リガンド-受容体はその上流や下流へ情報を受けたり与えたりすることにより機能している。したがってそのリガンド-受容体を含む上流下流のシグナルパスウェイの状態を検出することが疾患の理解には重要である。本研究では、T 細胞をサブタイプに分類したのち、T 細胞の細胞間情報伝達を表す確率を推定した。すなわち免疫疾患と健常な細胞を比べて T 細胞と T 細胞以外のその他の各種の免疫細胞間において特に発現変動する遺伝子を、シグナルパスウェイごとに疾患において発現差のある遺伝子を検出した。これら検出したシグナルパスウェイや発現差遺伝子については、新規、既知の両者が含まれていた。

また先行研究でも $TGF\beta 1$ については T 細胞との関連が報告がなされていた。新規に検出したシグナルパスウェイに含まれる遺伝子については、いずれも免疫関連遺伝子であることが報告されていた。今回の研究で新規に検出されたシグナルパスウェイが多かったことに関しては、細胞間の情報伝達を考慮した場合と、クラスタリングした細胞種間の疾患と健常との比較を解析した場合とでは、通常のクラスタリングによる比較の解析では、リガンドー受容体といった関係性を考慮していないために結果が大きく異なったと考えられる。このように解析の結果が先行研究と一部重複があることから、解析手法としては異なるにも関わらず、共通に検出される遺伝子というのは疾患との関連性がある可能性を示唆していると考えられる。また新規に検出した遺伝子については、いずれも疾患に直接あるいは間接的に関わる遺伝子として知られているものが検出されたことから、疾患に関わる新規の遺伝子候補の検出にも有効であることが示唆される。

第 4 章では、免疫疾患の病態変化が細胞内の遺伝子発現の動的な変化により引き起こされることを踏まえ、シングルセル RNA-seq に基づく解析や細胞間情報伝達解析による疾患メカニズム解明の今後の展望を述べている。

目次

第 1 章 緒論	2
1.1 免疫疾患と自己免疫疾患	2
1.2 免疫における免疫細胞	3
1.2.1 免疫細胞の種類と役割	3
1.2.2 T 細胞の種類と役割	4
1.2.3 免疫疾患における T 細胞の関わり	7
1.2.4 自己免疫疾患における T 細胞の関わり	7
1.2.5 免疫細胞の細胞系譜	8
1.2.6 免疫疾患における細胞系譜解析の重要性	8
1.2.7 免疫疾患における細胞間情報伝達の重要性	10
1.3 シングルセル RNA-seq の免疫疾患研究への応用	10
1.4 本論文の目的と構成	12
第 2 章 シングルセル RNA-seq を用いた細胞系譜推定による発現差遺伝子の探索	
手法の研究	14
2.1 緒言	14
2.2 先行研究	15
2.2.1 先行研究の目的	15
2.2.2 先行研究のデータセット	16
2.2.3 先行研究の解析手法	16
2.2.4 先行研究の解析結果	18
2.2.5 先行研究における課題	23
2.3 本研究	24
2.3.1 本研究の目的	24
2.3.2 本研究の解析手法	24
2.3.3 本研究の解析結果	26
2.4 本研究の考察	31
2.5 結言	36

第 3 章 シングルセル RNA-seq データを用いた細胞間情報伝達の推定による免疫	
制御異常マーカー遺伝子の探索手法の研究	37
3.1 緒言	37
3.2 先行研究	38
3.2.1 先行研究の目的	38
3.2.2 先行研究のデータセット	39
3.2.3 先行研究の解析手法	39
3.2.4 先行研究の解析結果	42
3.2.5 先行研究の課題	42
3.3 本研究の目的	43
3.4 本研究の解析手法	44
3.4.1 細胞間情報伝達解析の概要	44
3.4.2 細胞間情報伝達の解析手法	44
3.4.3 シグナルパスウェイにおけるリガンド受容体の寄与の解析手法	46
3.5 細胞間情報伝達解析の結果	47
3.5.1 シグナルパスウェイにおけるリガンド受容体の寄与の解析結果	47
3.6 考察	49
3.7 結言	52
第 4 章 結論	55
4.1 本研究のまとめ	55
4.2 免疫疾患研究における本研究の寄与と今後の展望	58
参考文献	60
謝辞	71
附録	72

第 1 章 緒論

1.1 免疫疾患と自己免疫疾患

花粉症、食物アレルギー、喘息、アトピー性皮膚炎などの免疫疾患や、関節リウマチなどの自己免疫疾患の多くは、人生の長期にわたり、体の様々な部位にまたがる症状がある。また、年齢や体質、生活習慣、環境などによって改善と悪化を繰り返すため、日常生活に多大な影響を与えている [1]。アナフィラキシーや薬物アレルギーといった一部の疾患では、命を落とす危険もある。

免疫は本来人間が持つ異物に対する排除システムであり複雑な生命現象である。これにより病原微生物に対する防御反応、がん化した細胞の排除などが起こるが、このシステムが破綻すると自分自身の細胞や組織に対して過剰に反応して攻撃してしまう状態である自己免疫疾患になる。

特に多くみられる自己免疫疾患には、バセドウ氏病、関節リウマチ、橋本甲状腺炎、1 型糖尿病、全身性エリテマトーデス、潰瘍性大腸炎、クローン病などがある。自己免疫疾患の原因はこれまで不明なところが多く、細菌の感染や物理的な人体への損傷きっかけになったり、なんらかの刺激により自己抗原が生成してしまうことによるとも言われている。治療法は自己免疫疾患の種類によって異なるが、免疫機能を抑制する薬（免疫抑制剤、ステロイド、抗炎症剤など）がしばしば使用される。免疫機能の抑制は、ウイルスや細菌の疾患に対する抵抗力を低下させ、様々な抗原に対する免疫反応を劇的に抑制するため、免疫系のバランスが崩れ、感染症などの病気にかかりやすくなる可能性がある。そのため、病気を引き起こす免疫反応だけをピンポイントで抑えることができる新しい治療法の開発が望まれている。

これまで複雑な免疫疾患を詳細に特徴付けることが困難で、免疫疾患か自己免疫疾患かをはっきり区別できないようなサルコイドーシスのような疾患も知られている [2]。

このように免疫疾患は多種多様な原因が考えられ、喘息、アトピー性皮膚炎、花粉症、食物アレルギーなどの免疫疾患や、リウマチ、膠原病、潰瘍性大腸炎、クローン病などの自己免疫疾患は、症状が長期的に持続するため、健康な生活を送るうえでの QOL の著しい低下が懸念されていることから、予防や治療に対する社会的なニーズが大きい。

したがって、これらの疾患のメカニズムを詳細に解明することは、疾患の予防、診断、治療の方法を提案する上での重要である。細胞レベルでその遺伝子発現情報を取得できるシ

表 1.1 免疫疾患の特徴

種類	役割	例
免疫疾患（アレルギー疾患）	外部からの異物（非自己）に対する 免疫応答	花粉症 アトピー性皮膚炎 喘息
自己免疫疾患	自分の身体（自己）に対する 免疫応答	バセドウ病 関節リウマチ 1 型糖尿病 全身性エリテマトーデス 潰瘍性大腸炎、クローン病
どちらか不明な疾患	未解明な部分が多く、何に対する 免疫応答か不明	サルコイドーシス

シングルセル RNA-seq はその疾患のメカニズムの解明に対して寄与できると期待される。

まとめとして代表的な免疫疾患の種類と特徴及びその疾患例を以下示した（表 1.1）。

1.2 免疫における免疫細胞

1.2.1 免疫細胞の種類と役割

免疫細胞の中心的な役割を担っている細胞として、白血球が知られている。白血球は外敵の侵入から体を守るために異物を検知して防御を行う。リンパ球は白血球の一部で、さらに B 細胞、T 細胞、NK 細胞（ナチュラルキラー細胞）などに分類される。NK 細胞は生得的に存在する免疫細胞で、細菌やウイルスなどの病原体に感染した細胞を細胞表面に発現する蛋白質などの情報を基に攻撃する。B 細胞は細菌やウイルスなどの病原体が侵入してきたときに抗体を作る。また、T 細胞は病原体が侵入してきたときに抗体を作ることをサポートする役割を持つ。T 細胞は、体を守るために独自に働き、侵入してきた病原体を記憶することができるため、様々な T 細胞のサブタイプに分化をして細かな役割分担を行いながら、素早く反応し、細菌やウイルスの異物を破壊する。T 細胞はリンパ球の一種で、血液中のリンパ球の過半数以上を占めていることから免疫の中心的な存在である。T 細胞は、マクロファージや樹状細胞などの抗原提示細胞から、細胞表面に発現する T 細胞の抗原受容体を介して抗原情報を受け取ることにより、さまざまな機能を発揮する。

免疫細胞は、慣例的に細胞表面に発現する蛋白質によって、細胞の種類ごとに分類される。これらの蛋白質は細胞マーカーと呼ばれ、CD(Cluster of Differentiation) に続く番号表記でマーカー蛋白質の種類ごとに表現されている。

その分類によればヘルパー T 細胞は CD4、キラー T 細胞は CD8 を持っている。自己と非自己をどのように識別するための指標となるのが、細胞表面にある MHC (Major Histocompatibility Complex、主要組織適合遺伝子複合体) と呼ばれる蛋白質で、ヒトでは HLA (Human Leukocyte Antigen、ヒト白血病抗原) と呼ばれている。HLA 抗原は自己のすべての細胞に存在している。それらを認識し、自己と非自己を区別するのは T 細胞である。これゆえ T 細胞は免疫における中心的な役割を担っている。T 細胞は自己として提示された抗原を T 細胞受容体で認識する。その後、サイトカイン産生や抗体が病原体に結合後、抗体を認識する受容体を持つマクロファージや NK 細胞などの免疫細胞が引き寄せられることで細胞や病原体を死滅させる方法を通じて、免疫としての役割を果たす。キラー T 細胞は、ウイルスに感染した細胞やがん細胞を直接的に死滅させる細胞性免疫に関与している。一方、ヘルパー T 細胞は、抗原刺激に反応し、他の免疫細胞の作用を制御する司令塔的な役割を担っている。

1.2.2 T 細胞の種類と役割

T 細胞は様々な細胞へ分化していくことにより特化した役割を取得し免疫システムを担っている。したがって、T 細胞の分化についての一般的な背景は本研究において分化の起点や終点を指定する上で参考となるために以下に記載する。T 細胞の分化は次のように行われる。骨髓内の造血幹細胞からプロ T 細胞という細胞が生成する。このプロ T 細胞は細胞マーカーとしては、CD4、CD8 を共に発現していない。その後、プロ T 細胞は胸腺へと移動する。その後、胸腺においてプレ T 細胞と呼ばれる CD4、CD8 を共に発現している細胞に変化する。プレ T 細胞はさらに、CD4 あるいは CD8 のどちらかを発現した細胞に分化する。この時点においては特定の抗原刺激を受けていないためナイーブ T 細胞と呼ばれる。その後胸腺を離れて、ナイーブ CD4+T 細胞または CD8+T 細胞として全身循環に入る。これらのナイーブ T 細胞は血液やリンパ系を循環し、リンパ節や脾臓あるいは粘膜関連リンパ組織で樹状細胞が提示する特異的抗原に出会うと活性化してエフェクター T 細胞へ分化する。それぞれの T 細胞のサブタイプの種類と役割について以下に示した (表 1.2)。

また、CD4+ ヘルパー T 細胞の種類と役割について以下に示した (表 1.3)。

表 1.2 T 細胞の種類と役割

種類	役割	疾患例
CD4+ ナイーブ T 細胞	抗原刺激を受けていない成熟 T 細胞	免疫疾患
初期エフェクター T 細胞	ヘルパー T 細胞への分化の途中の細胞 (発現するサイトカインの種類 によって Th1,Th2,Th17 に分類される)	免疫疾患 自己免疫疾患
CD4+ メモリー T 細胞	寿命が長い 同一の抗原が再侵入してきた場合に 速やかな免疫応答を発動することが可能	免疫疾患 自己免疫疾患
制御性 T 細胞	免疫応答を抑制することで自己寛容や過剰 な免疫応答を抑制する	がん、免疫疾患 自己免疫疾患
ガンマデルタ T 細胞	多様な機能を持ち、 がん細胞を認識して攻撃する	がん 自己免疫疾患
CD8+ ナイーブ T 細胞	抗原刺激を受けていない成熟 T 細胞	免疫疾患
CD8+ エフェクター T 細胞 (細胞障害性 T 細胞)	MHC クラス I 分子に代表される ウイルス抗原を認識・活性化し、 感染細胞を破壊する	免疫疾患

表 1.3 CD4+ ヘルパー T 細胞の種類と役割

種類	役割
Th1	感染部位に移動した後に IFN- γ を放出し、マクロファージによる 病原体の取り込みと破壊を促進する
Th2	リンパ節、脾臓や粘膜リンパ組織といった二次リンパ組織に留まり、 IL4 を分泌し、ナイーブ B 細胞の抗原特異的活性を促進する
Th17	IL17 を産生するは細胞が抗菌ペプチドやの産生を促進し、 好中球を活性化して粘膜表面を感染から守る

表 1.2 にも示したが、CD4+T 細胞のサブセットの中には、FOXP3 を発現し、抗原特異的に免疫応答を抑制する制御性 T 細胞 (Treg、Regulatory T cell) が存在する。FOXP3 に変異を持つ患者は、制御性 T 細胞の機能疾患により IPEX (immune dysregulation polyendocrinopathy enteropathy X-linked) 症候群と呼ばれる甲状腺異常や I 型糖尿病、炎症性腸疾患などの自己免疫疾患を発症することが知られている。制御性 T 細胞は、末梢組織に発現する自己抗原を認識する自己反応性 T 細胞の活性化と増殖を抑制し、過剰な免疫反応を抑制する役割を担っていて、自己免疫疾患の発症を予防していることが明らかになりつつある。このように制御性 T 細胞による免疫制御は、重要な役割を担っている。また、Th17 細胞の分化は、制御性 T 細胞の分化と相互に対立する形で成立していることが明らかにされつつある。例えば、Th17 細胞が過剰に作られると、自己の組織が攻撃され、関節リウマチや全身性エリテマトーデス (SLE、systemic lupus erythematosus) などの自己免疫疾患につながる。制御性 T 細胞がそれらの過剰な働きを抑制して疾患の発症を制御する。

免疫細胞は多種多様なヘテロ細胞集団でありこれらが原因となり発症する疾患も極めて複雑である。まとめると、近年になって、CD4+ ヘルパー T 細胞は、Th1, Th2 細胞の他にも感染防御に重要な Th17 細胞、過剰な炎症を抑える制御性 T 細胞、リンパ組織の形成を促す濾胞ヘルパー T 細胞の他にも、自然免疫 T 細胞を起源とする腸の粘膜に多数存在するガンマデルタ T 細胞種の T 細胞抗原受容体を持つ T 細胞 (通常の T 細胞は $\alpha\beta$ 型)、NK 細胞と T 細胞の両方の特徴をもった NKT 細胞など、新たな T 細胞種が次々に発見されてきている [3]。これらの細胞と疾患との関係を明らかにする研究も行われてきている。

このように T 細胞には様々な種類があり、多様な役割をそれぞれが担っている。近年、腸管粘膜に多数存在するガンマデルタ T 細胞抗原受容体を持つ T 細胞があることがわかってきている。ガンマデルタ T 細胞は通常はアルファベータ型の抗原受容体を持つが、自然免疫系 T 細胞に由来する細胞抗原受容体を持つ T 細胞である。本研究においても解析の対象として扱っているガンマデルタ T 細胞については、比較的新しく発見されたことから、未解明な部分も残されていると考えられている。したがって、これらの免疫細胞と疾患との関係を調べる研究は現在も進行中である。

1.2.3 免疫疾患における T 細胞の関わり

前節で説明したように、免疫細胞には数多くの種類があり、それぞれの細胞がその役割を担っていることで全体として免疫システムはバランスを取りながら機能している。そのため、T 細胞に関する障害は生体の免疫系にとっては致命的になりうる。例えば、がん細胞を移植しても T 細胞に問題があると移植された組織を非自己として拒絶することができない。その結果、がん細胞が移植された生体内で増殖してしまい本来の健常な状態の生理機能が害されてしまう。AIDS（後天性免疫不全症候群、acquired immunodeficiency syndrome）は、HIV ウイルスによって引き起こされる疾患として知られている。この HIV ウイルスは CD4+ ヘルパー T 細胞に感染する。このため、HIV が感染すると CD4+ ヘルパー T 細胞は本来持っている B 細胞の抗体を促す役割ができなくなってしまう。このため、B 細胞を含む他の免疫系が体内においてウイルスや細菌を排除することができなくなる。その結果として、これらの病原体が増殖し、通常であれば免疫に抑え込まれていたウイルスや細菌が日和見感染症として発症してしまい、肺炎などの病気を引き起こしてしまう。

このように、免疫系の中心的な役割を担う T 細胞が失われると、免疫機能全体が弱まるため、さまざまな病原体の侵入を許してしまう。まとめると、免疫疾患と T 細胞は密接に関連していることが分かっている。

1.2.4 自己免疫疾患における T 細胞の関わり

自己免疫疾患やアレルギー疾患の患者のリンパ球においては、CD4+ ヘルパー T 細胞に何らかの異常があることが多いことが知られている [4, 5, 6, 7]。CD4+ ヘルパー T 細胞は、B 細胞を中心とした他の免疫細胞に指令を出し、抗体を産生させたり、キラー T 細胞や NK 細胞（Natural Killer）の持つ殺菌作用を活性化する。無数の種類が考えられる抗原は、それに対応する遺伝情報としてヘルパー T 細胞ごとに、それぞれの抗原の情報が記憶される。これにより、膨大な種類の抗原に対して免疫反応を行うための備えを行うことができる。CD4+ ヘルパー T 細胞は、骨髄で前駆細胞として生まれ、胸腺に移動して、役割が決まっていない状態のナイーブ T 細胞（Th0 細胞）の状態になる。胸腺を出たナイーブ T 細胞は、抗原の情報を受け取ると抗原の種類に応じてエフェクター T 細胞（Th1 細胞、Th2 細胞、Th17 細胞、制御性 T 細胞）や制御性 T 細胞、ガンマデルタ T 細胞へと分化していく。分化した T 細胞はその機能が固定化されているわけではなく、

ある程度の可塑性を持っていることも特徴である。したがって疾患も状態により T 細胞のサブタイプの細胞の変化を伴うことがある。

一方、免疫反応は適切なタイミングでその役割を減弱させたり終了させる必要がある。また、自己の蛋白質に反応しないことも必要である。免疫系が異常に活性化したままであったり、自己の蛋白質や花粉などの本来は無害あるいはそれほど害のない物質に反応してしまうと、関節リウマチや炎症性腸疾患などの自己免疫疾患になってしまう。また花粉症などのアレルギーについては免疫が過剰に反応してしまった結果、引き起こされる。

このように、自己免疫疾患と T 細胞は密接に関連していることが知られている。

1.2.5 免疫細胞の細胞系譜

これまでに説明した免疫細胞の種類以外にも造血幹細胞から免疫細胞は様々なサブタイプへと分化していく。

この様子を細胞系譜の図 1.1 として示した。これまで述べたように、CD4+T 細胞は様々な機能をもつサブタイプに分化することによって多様な役割を担っている。またガンマデルタ T 細胞や制御性 T 細胞など比較的最近発見された細胞種を含み未解明な部分が多い。このため本研究においては、疾患との関連をこの細胞種群で解析することにより、新たな知見が得られる可能性があると期待した。本研究において着目した CD4+T 細胞のサブタイプについては黄色で示した。このように免疫細胞は多種多様に分化しそれぞれの役割を担うことで免疫システムは正常に機能することができる。

1.2.6 免疫疾患における細胞系譜解析の重要性

これまで説明してきたように免疫疾患と免疫細胞とは密接に関わっている。がんの免疫回避は、抗腫瘍免疫療法に対する障壁となっているが、腫瘍の成長、進行、または転移に影響を与える様々な細胞種またはその状態が関与していることが知られている [8]。しかしながら時間変化を伴うような現象である免疫耐性の取得といった問題においては、細胞種ごとの状態とがんとの関係性の解析だけでなく、時間的な要素を含んだ細胞の分化の解析が有効であると考えられる。近年普及しつつある細胞系譜解析はシングルセル RNA-seq を使用して細胞同士の類似性を評価する手法の一つである。しかしながらクラスタリングのように離散的ではなく、類似性のある細胞を疑似時間とよばれるパラメータでつなぐことによりその移行過程を軌跡として解析する手法である。細胞系譜解析のような手法は、免疫疾患のような複雑で細胞種の分化のような時間変化の影響を受ける現象に対して有効

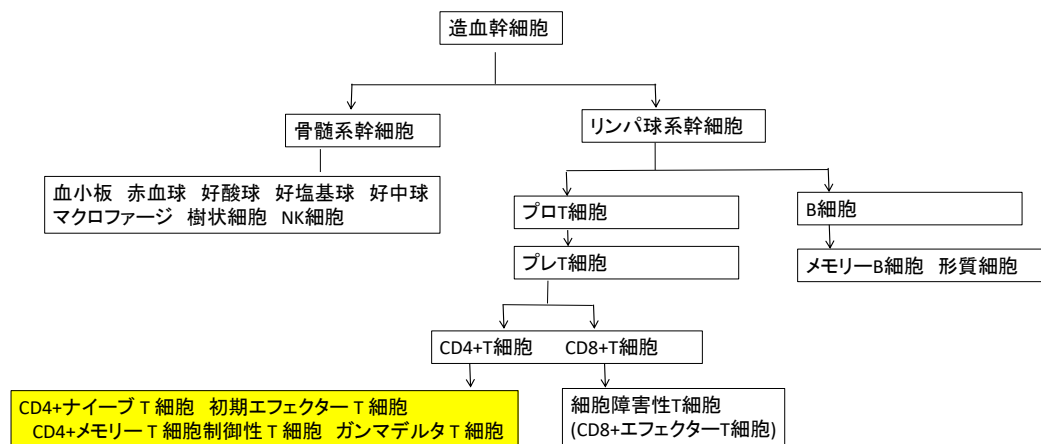


図 1.1 免疫細胞の細胞系譜

であると考えられる。細胞系譜解析が有効であった例としては以下のような解析例が知られている。

掌蹠膿疱症 (PPP;palmoplantar pustulosis) は、手のひらと足の裏に好中球で満たされた痛みを伴う膿疱の噴出を特徴とする重度の炎症性皮膚疾患である。PPP は生活の質に大きな影響を与えるが、理解が不十分であり、治療が難しいことで有名である。PPP シングルセル RNA-Seq と細胞系譜解析により、PPP 患者の CD4 メモリー T 細胞は、Th2 と Th17 細胞との分化の偏りと関連することが分かった。このように PPP は複雑な T 細胞活性化パターンと関連していることが明らかになった [9]。

また、原発性硬化性胆管炎 (PSC;Primary sclerosing cholangitis) の病因を調べた研究では、炎症を起こしたヒト肝臓の T 細胞の詳細な特性評価をシングルセル RNA-seq と細胞系譜解析により行った。その結果、PSC 患者の肝臓における組織常駐ナイーブ様 T 細胞の存在と、これらの細胞が Th17 細胞への分極状態を獲得する発達傾向を持っていることを示唆した [10]。

このように時間軸を考慮して病因を理解するために細胞系譜解析は有効であることが示されている [11]。細胞系譜解析には、まずクラスター解析を行い細胞種を定義する。次に細胞種間の系譜を推定する。その後、系譜に沿った遺伝子発現情報を取得することが必要である。しかしながら、この手法を利用した、疾患と健常のそれぞれの系譜を比較する手法については、確立されていない。

1.2.7 免疫疾患における細胞間情報伝達の重要性

生物は、分化などの恒常性を維持するために、細胞間でさまざまな情報をやり取りしている。細胞間の情報伝達の種類には、1) 細胞同士が直接的に表面の蛋白質を介して行う細胞接着、2) 細胞マトリクスと受容体を介した方法、3) 細胞から分泌されるリガンドと細胞膜表面に存在する受容体と呼ばれる蛋白質を介して行う方法がある。3) の細胞間情報伝達に關与する因子として、特に免疫に關与するものはサイトカインと呼ばれている。これらのサイトカインは細胞の分化や機能の発現のために影響を及ぼしている。疾患においても健常と比較した場合、細胞間の情報伝達の機能が本来の役割を果たせない分子メカニズムが存在している可能性もある。さらに分子メカニズムが分かれば創薬ターゲットとなる標的蛋白質を検出することができるかもしれない。したがって疾患の病態をより詳細に理解するためには、免疫の中心的な役割を担う T 細胞と、T 細胞を含む免疫細胞がどのように相互作用するかを理解する必要がある。実際 Jak キナーゼシグナルパスウェイにおける細胞間の情報伝達が皮膚炎などの自己免疫疾患に關与しているという報告 [12] や T 細胞間の情報伝達が自己免疫疾患に關与して T 細胞の機能に深く関わっているという報告もある [13] シングル RNA-seq データでは細胞種ごとに遺伝子の発現量を定量的に解析できることから、細胞間情報伝達についても定量的に解析できると考えられる。細胞間情報伝達には、まずクラスター解析を行い細胞種を定義する。次に細胞種において発現する遺伝子を検出する。その後、細胞種間で注目した遺伝子の発現量とそれに関連するデータベースにある遺伝子の発現量に基づき細胞間の情報伝達確率を計算する。しかしながら、この手法を利用した、疾患と健常のそれぞれの病態を比較した手法については、確立されていない。

1.3 シングルセル RNA-seq の免疫疾患研究への応用

これまで説明したように、T 細胞は細胞表面に CD4 や CD8 などの細胞表面マーカー蛋白質を発現する。この情報をマーカーとして細胞の分化の過程や状態を判断することが行われている。しかしながら必ずしも細胞表面マーカーだけですべての細胞種を判断できるとは限らないケースも知られている [14]。

近年、普及してきているシングルセル RNA-seq は 1 細胞ごとに mRNA の配列と量を網羅的に検出する技術である。この技術は、従来のまとまった細胞種の母集団平均を測定するような RNA-seq 解析では不可能であった、細胞ごとの遺伝子情報を検出することが

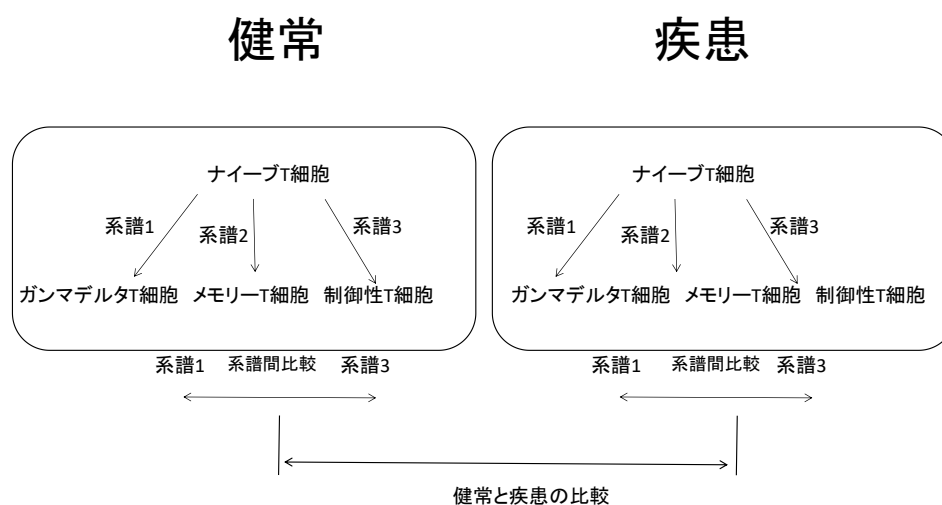


図 1.2 細胞系譜解析による健康と疾患の比較の概念図

できるため、幅広い研究に利用がなされてきている。

これまでのシングルセル RNA-seq の研究では、細胞集団を特徴的に発現している遺伝子ごとに細胞種に分類する方法、細胞種を解析して細胞の分化や刺激に応じた遺伝子発現の経時変化を調べる方法などが開発されている [15]。

しかしながら、免疫疾患のような複数の免疫細胞が関与する複雑な現象に関して生体の状況をより反映させるためには、特定の細胞集団を細胞種へ分類することにより細胞種間での比較を行う方法の他に、健康と疾患のような複数の細胞集団での免疫細胞の各々のサブタイプへの分化といった状態遷移を考慮し、その比較により遺伝子情報をの検出するような、更なる手法の開発が必要であると考えられる。すなわち従来のバルクの RNA-seq による解析では埋もれてしまうような細胞集団を規定し、さらに疾患と健康など遺伝子発現パターンの状態の異なる複数の細胞集団において、免疫細胞の細胞系譜による系譜間の遺伝子の発現変動を比較することで、状態遷移過程の時間軸を含めた理解ができると考えられる。そのような解析により疾患特異的に起きている変化を検出するような、免疫疾患をより詳細に特徴付ける新しい手法の開発が疾患メカニズムを詳細に解明するためには必要であると考えられる。

特に、免疫疾患に関わる中心的な役割を担っている T 細胞のサブタイプへの分化に注目して、健康と疾患との遺伝子発現変化を解明することが期待される。以上の内容を行うために本研究における解析手法の概念図を示した (図 1.2)。

さらに T 細胞は他の細胞集団との間では情報伝達を行うことにより、お互い関わり合いながら、情報を受け取ったり、発信したりしている。したがって、T 細胞の集団と、他の細胞集団の間の情報伝達を解析することにより、免疫疾患に関する T 細胞を含む細胞間の情報伝達をより詳細に明らかにすることができれば、疾患のメカニズムを理解できると期待される。

これら 2 つの解析手法を開発することで、免疫疾患と健常など遺伝子発現パターン状態の異なる複数の細胞集団で T 細胞の分化状態や細胞間情報伝達が免疫疾患において、どのように関与しているのかを明らかにできることが期待される。

1.4 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免疫疾患と健常の異なる状態における T 細胞の分化や細胞間情報伝達を明らかにする研究領域において、新たな技術的なアプローチとして、シングル RNA-seq データを利用した解析手法には、さらなる改善と提案が必要である。

特に、本研究で対象とした T 細胞のサブタイプは、免疫の多様な特化した作用を発揮するために、多様性に富んでいることが想定される。現在、最もよく使われているクラスタリング手法は、細胞種として定義したクラスター間における遺伝子発現の違いを比較する。本研究では、免疫細胞の一つである T 細胞に着目してそのサブタイプへの分化過程において、疾患と健常など遺伝子発現パターン状態の異なる複数の細胞集団での状態遷移過程を比較することにより、異なる状態での細胞内の遺伝子発現の差異を検出す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。

第 1 章では、研究の背景、既往の研究、課題、本論文で取り扱う課題に関して述べた。

第 2 章では、健常と免疫疾患における状態の異なる細胞種において、ナイーブ CD4⁺T 細胞から分化の進んだ T 細胞のサブタイプへの移行過程を細胞系譜として解析し、その状態変化を検出する手法を提案した。提案手法を検討するために、免疫疾患あるいは自己免疫疾患としての定義すらきちんと決められることができていない不明な点が多いサルコイドーシスと健常人のシングルセル RNA-seq データを使用して、T 細胞の分化の過程における細胞系譜解析を行った。この際には、免疫疾患の患者には健常者と比較して特徴的な遺伝子発現があると仮定した上で健常と疾患各々で解析した。その後、健常と疾患とで検出した系譜間における差分解析を行い、特徴的な遺伝子を検出した。

第3章では、同様なシングルセル RNA-seq のデータを対象に、細胞間情報伝達解析による疾患において健常と発現差のある遺伝子を見出す手法を提案した。免疫においては、細胞種ごとにリガンドと呼ばれる何らかのシグナルが出力され、受け手である受容体への入力によるやり取りを通じて、その機能を発揮しているが、疾患においてはそのやり取りが健常と比べて変化することにより免疫の不具合が起きていると考えられる。リガンドー受容体はその上流や下流へ情報を受けたり与えたりすることにより情報を伝達して働いている。したがってそのリガンドー受容体を含む上流下流への情報伝達経路であるシグナルパスウェイの状態を検出することが疾患の理解には重要である。本研究では、提案手法を検討するために、T 細胞とその他のすべての免疫細胞の間に情報伝達があると仮定したときに、どのシグナルパスウェイが高い確率で情報伝達しているか、さらにそのシグナルパスウェイの中でどのリガンドと受容体が主要な因子であるかを推定した。免疫疾患と健常な細胞とを比べて T 細胞と各種の免疫細胞間において特に発現変動する遺伝子を、シグナルパスウェイごとに検出した。

第4章では、健常と疾患で状態の異なる免疫細胞が、分化した細胞種の遺伝子発現の変動で特徴づけられることについて、異なる状態間での系譜間における細胞の遺伝子発現の状態を検出する手法を評価し、解析法による違いを考察した。また、T 細胞とその他のすべての免疫細胞の間における情報伝達を、従来の解析と比較して本解析における違いを考察した。今後のシングルセル RNA-seq データを用いた異なる状態における免疫疾患での T 細胞の分化の状態の解明に対する展望について述べた。

第 2 章 シングルセル RNA-seq を用いた細胞系譜推定による発現差遺伝子の探索手法の研究

2.1 緒言

シングルセル RNA-seq は単一細胞ごとでの遺伝子発現情報を取得することができ、従来の細胞を平均集団として解析する RNA-seq にはない解像度をもつため、幅広く活用されてきている [16]。これまで免疫に関する疾患の研究に関しては、様々な研究が行われてきた [17, 18, 19, 20, 21, 22, 23, 24, 25]。しかしながらこれらの研究は、疾患患者の組織や血液中に多く発現している蛋白質や遺伝子の変異を調べた報告であり、細胞種との関係については調べられていない。

先行研究として参照した Garman ら [26] によって報告されている免疫疾患の一つであるサルコイドーシスに関する研究は、シングルセル RNA-seq を使用して、免疫細胞を既知のマーカー遺伝子を基に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グループを検出し、細胞種ごとの頻度が患者と健常人の間にあるか、また、それぞれの細胞種ごとに異なる発現状態がある遺伝子が存在するかどうかという研究である。この解析では、細胞種ごとに疾患に多く発現する遺伝子が明らかになった。しかしながら、Garman らの先行研究の手法では、ナイーブ T 細胞からその分化した T 細胞である、メモリー T 細胞、ガンマデルタ T 細胞、制御性 T 細胞への分化過程を考慮した疾患と健常との間での違いを見出すことは行われていない。細胞の分化過程を系譜として解析することにより、クラスタリングのみの解析では得られないクラスター間の状態遷移を考慮することができるため、疾患と健常との違いや疾患における病態の変化を疑似的な時間軸を介して解釈することができる長所があると考えられる。一方で、複数のクラスター間の移行に関して始点と終点を生物学的な意味合いなどから決定する点において、疑似的な時間軸が真の病態の変化とリンクしているかどうかについての恣意性がぬぐい切れないという短所がある。そこで本章では、健常と免疫疾患における状態の異なる細胞種において、ナイーブ CD4+T 細胞から分化の進んだ T 細胞のサブタイプへの移行過程を細胞系譜として解析し、その状態変化を検出する手法を提案する。本研究の目的としては、シングルセル RNA-seq という 1 細胞の遺伝子発現情報を取得できる長所を利用することで、免疫疾患における T 細胞からサブタイプへの分化の系譜において、同じ細胞種の中でも時間的な情報を加味して特徴的に発現する遺伝子を検出する手法を考案することで

ある。

2.2 先行研究

Garman らの先行研究 [26] 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述べる。この先行研究は免疫疾患の一つであるサルコイドーシスに関して、シングルセル RNA-seq を使用して健常と患者での遺伝子発現差解析を細胞種のクラスタリングにより行った研究である。また、細胞系譜解析を用いたがんの薬剤抵抗性因子の検出に関する村山らの先行研究 [15] においては、発現変動を示す遺伝子を検出する手法が提案されている。Garman らの先行研究におけるデータセットに、細胞系譜解析を応用し、疾患と健常という異なる細胞状態において系譜間発現差解析を行うことで、従来のクラスタリングによる手法との比較ができるという点においてこの先行研究を選んだ。

2.2.1 先行研究の目的

免疫疾患（自己免疫疾患を含む）は多様な免疫細胞が関与する複雑な生体内現象であり、その詳細を明らかにすることが極めて困難であり、疾患患者を長い間悩ませてきた。免疫疾患は数多くあり、また、同じ疾患でも個人により分子レベルの病態は異なると考えられるため、その正確な診断が、治療法の適用や薬剤開発に求められている。従来の細胞のトランスクリプトームを平均化して解析するバルク RNA-seq ではこれらの多様な免疫細胞ごとに情報を取得することができなかったために、細胞種特異的に発現する蛋白質を標的として薬剤を設計することが困難であり、免疫抑制剤のように多くの細胞種が関わる免疫機能の全体的な抑制を行うことで、過剰に働いている免疫関連の遺伝子発現を網羅的に抑制し正常な状態に近づけるために、適用されてきた。近年、細胞ごとに RNA-seq を行い個々の情報のトランスクリプトームを解析するシングル RNA-seq が発明されて普及してきたために、免疫疾患に関しても、この技術を利用することにより、今まで取得できなかった細胞ごとの情報から、新たに疾患に関する知見を検出できる可能性がでてきた。そこで全身性の免疫疾患（中でも肺で発症することが多い）に関して、Garman らの先行研究 [26] においては、その免疫疾患に特異的なマーカー遺伝子を、健常と疾患との免疫細胞種ごとの遺伝子発現の比較により検出することを目的に研究が行われた。一方、村山らの先行研究 [15] では、細胞系譜解析を用いたがんの薬剤抵抗性因子の検出を行い、時間変化を考慮した発現変動を示す遺伝子を検出する手法が提案されている。

2.2.2 先行研究のデータセット

Garman らの先行研究では、98,741 個の細胞からシングルセルのトランスクリプトームを取得、血小板と赤血球をフィルタリングした。シングルセル RNA-seq には、10x Genomics Chromium 集積流体回路 (IFC: Integrated Fluidic Circuit) システムが用いられた。ハイスループット マイクロフルイディクスを利用して個々の細胞を油滴内に自動的にカプセル化して調製後、イルミナ HiSeq3000sequencer によって配列が決定された。各液滴に由来する単一細胞 3' トランスクリプトームを回収した。シーケンシングが完了すると、10x Genomics Cell Ranger (v3.0.2) を使用して処理された。以上の手法より得られた計 53,756 の抹消血液細胞の単一細胞ごとの RNA-seq データが解析に使用された。この内、疾患と健常の細胞数の内訳は、33 人の患者 (47,276 細胞) と 9 人の健常者 (6,480 細胞) となっている。

免疫疾患の患者及び健常人より得られたシングルセル RNA-seq データが Gene Expression Omnibus リポジトリに GSE132338 として登録されており、本研究でもこのデータを使用する。

2.2.3 先行研究の解析手法

Garman らの先行研究では、細胞をクラスタリングした後、健常と疾患の細胞の遺伝子の発現量を比較して、疾患の細胞で高発現している遺伝子を、疾患のマーカー遺伝子の候補として検出している。

そのための解析手法は主に以下の 5 つのステップからなる。

ステップ 1 細胞のクラスタリングのためのデータセットの前処理

ステップ 2 データセットのクラスタリング

ステップ 3 各クラスターの発現遺伝子の検出

ステップ 4 クラスターの細胞種の同定

ステップ 5 健常人と患者での細胞種ごとの発現差遺伝子の同定

以下でそれぞれのステップについて詳細な解析手法を説明する。

2.2.3.1 先行研究におけるデータセットのクラスタリングのためのデータセットの前処理

シングル RNA-seq データを解析する上で、前処理として正規化を行い、細胞ごとの発現量の差を補正する。細胞あたりの発現量を各細胞の総発現量で割り、スケーリングファクターを掛け、対数値に変換する。次に、各遺伝子の発現量の対数値から平均値を引き、標準偏差で割ることにより、各遺伝子の全細胞における平均値と標準偏差を表す Z スコアに変換する。上記の正規化されたデータを用いて高発現遺伝子を選択する。まず、以下の条件のいずれかに該当する細胞を個別に削除する。1) 陽性発現遺伝子の総数が 200 以下または 2000 以上、2) ミトコンドリアリードの割合が 20 %以上、3) ランダムに選択された細胞ペアの転写プロファイルを平均化することによって作成された遺伝子発現に対しての細胞の類似度に従って、一つの細胞に複数の細胞が混在してしまうダブレットという状態を取り除くために予測を行う。これにより推定した細胞ダブレットである確率が 7.5 % 以上のいずれかの条件を満たした細胞をまず除去する。

2.2.3.2 先行研究におけるデータセットのクラスタリング

細胞のクラスタリングと細胞種の決定に関して、始めに健常または疾患の細胞を対数正規化する。スケーリングしたデータを用いて、graph-based clustering approach[27]を用いてクラスタリングを行う。このクラスタリングは K 近傍法 (KNN、K Nearest Neighbor) を用いて近傍に存在する細胞同士をつないでグラフ構造にし、相互に高度に連結した部分細胞種をクラスターとして抽出する方法である。まず上位 70 主成分の PCA(principal component analysis) 空間でユークリッド距離に基づいて K 最近傍グラフを作成し、任意の 2 細胞について、2 細胞間の全エッジの重なり (Jaccard 類似度) に基づいて 2 細胞間のエッジの重みを調整する。次に Louvain アルゴリズム [28] で細胞のクラスターを特定する。Louvain アルゴリズムは (1) モジュラリティというグラフを分割する際の質の最適化の計算を局所的に行う (2) 密に結合しているノードの集合であるコミュニティとして集約する。この手法は、従来の手法に比べ高速にコミュニティ抽出できる手法とされている。

2.2.3.3 各クラスターの発現差遺伝子の検出

遺伝子の発現量は、正規化、スケール化されたデータを使用する。各クラスターについて、他のクラスターと比べて発現差の大きい遺伝子を検出する。

2.2.3.4 クラスターの細胞種の同定

各クラスタについて、各クラスターを特定する上位のアップレギュレーションした遺伝子を同定する。正規化、スケーリングされたデータを用いて、Wilcoxon 検定を用い、10 %以上の細胞で発現している遺伝子に対して発現差の対数比の閾値が 0.25 の基準で、それぞれの細胞種において計算する。細胞種は、これらのクラスター特異的に発現している標準的な細胞種マーカーとして知られている特異的な遺伝子を指標として決定する（表 2.1）。

2.2.3.5 健常と疾患の細胞種ごとの発現差遺伝子の同定

発現差遺伝子の同定は、細胞種および健常か疾患かによって選択した細胞について、正規化、スケーリングされたデータを用いて、Wilcoxon 検定を用い、発現差の対数比の閾値が 0.01 の基準で、それぞれの細胞種において細胞間の遺伝子の発現差を計算する。

2.2.4 先行研究の解析結果

2.2.4.1 細胞のクラスタリングのためのデータセットの前処理結果

まず、解析に適さない低品質の細胞を除去するため、以下の条件のいずれかに該当する細胞を個別に削除する。その結果、1) 陽性発現遺伝子の総数が 200 以下の細胞数は 35,207 個または遺伝子として 2000 個以上ある細胞の個数は 413 個ある。2) ミトコンドリアリードの割合が 20 %以上の細胞数は 1,217 個あった。3) 細胞分離が一部不十分であったため 2 個以上の細胞の遺伝子発現データが得られてしまう現象である細胞ダブレットの指標が 7.5 以上の細胞を除去した結果、残った合計 53,756 個の細胞が細胞種に割り当てられ、さらなる解析に使用する。この内、疾患と健常の細胞数の内訳は、33 人の患者（47,276 細胞）と 9 人の健常者（6,480 細胞）となっている。

2.2.4.2 データセットのクラスタリング結果

Louvain アルゴリズムにより細胞のクラスターが特定された。その結果、18 の異なるクラスターが同定された。

2.2.4.3 各クラスターの発現差遺伝子の検出結果

クラスターごとに遺伝子発現を計算して、検出した遺伝子を細胞種ごとに健常と疾患との対数発現差が大きい順にランキングした (表 2.2, 表 2.3, 表 2.4)。

2.2.4.4 クラスターの細胞種の同定結果

クラスターごとに遺伝子発現を計算して、細胞マーカーを基に各クラスターの細胞種を決定した。これらの結果は、t-SNE(t-Distributed Stochastic Neighbor Embedding) アルゴリズムにより、細胞種の発現パターンを可視化した (図 2.1)。t-SNE アルゴリズムは、複数の特徴量を持つ発現データのような高次元のデータを低次元に変換して可視化する

表 2.1 標準的な細胞種特異的細胞表面マーカー

Cell Type	Expression Pattern	References
CD4 ナイーブ T 細胞	CD3D+, CCR7+, SELL+, S100A4-,CD8A-,CD8B-	[29] [30]
CD4 メモリー T 細胞	CD3D+, S100A4+, CCR7-, CD8A-, CD8B-	[31]
初期エフェクター T 細胞	CD3D+, GZMK+, GZMB-	[32]
制御性 T 細胞	CD3D+, CD27+, FOXP3+	[33]
ガンマデルタ T 細胞	CD3D+, TRGC1+	[34]
CD8 ナイーブ T 細胞	CD3D+, CD8A+, CD8B+, S100A4+	
細胞障害性 T 細胞	CD3D+, CD8A+, CD8B+, GZMB-, GZMK-, GNLY+	[35]
ナチュラルキラー細胞	CD3D-, NKG7+, GZMB+	[36]
CD14 単球	LYZ+, FCGR3A-	[37]
CD16 単球	LYZ+, FCGR3A+	[38]
ミエロイド樹状細胞	FCER1A+, CLEC10A+	[39], [40]
形質細胞様樹状細胞	GZMB+, IL3RA+	[41], [42]
ナイーブ B 細胞	CD79A+, TCL1A+	[43]
メモリー B 細胞	CD79A+,TCL1A-	[44]
巨核球	PF4+, PPBP+, CD79A+	[45]
血小板	PPBP+, CD79A-	
造血幹細胞	SOX4+	[46]
赤血球	HBA1+	

表 2.2 ガンマデルタ T 細胞で検出された遺伝子

遺伝子	健常と疾患の logFC
<i>ECH1</i>	0.36
<i>DYNLL1</i>	0.29
<i>IMP3</i>	0.28
<i>RNF167</i>	0.27
<i>PLGRKT</i>	0.25
<i>SRSF1</i>	0.23
<i>STAG2</i>	0.23
<i>MARCH7</i>	0.17
<i>CERK</i>	0.15
<i>DNAJB9</i>	0.12

表 2.3 CD4 メモリー T 細胞で検出された遺伝子

遺伝子	健常と疾患の logFC
<i>IRF2BP2</i>	0.21
<i>AP2B1</i>	0.21
<i>CD82</i>	0.15
<i>PPM1B</i>	0.14
<i>SMIM27</i>	0.14
<i>SH3YL1</i>	0.13
<i>RAB11A</i>	0.12
<i>BTF3L4</i>	0.11
<i>PASK</i>	0.11
<i>RAB6A</i>	0.10

る次元削減アルゴリズムの一つである [47]。

細胞の各サブタイプごとに色付けしている。CD4+ T (CD3D、SELL)、CD8+ T (CD3D、CD8A/B)、NK 細胞 (GNLY、NKG7)、B 細胞 (CD79A)、ナイーブ B 細胞 (TCL1A)、CD14+ 単球 (LYZ、FCGR3A)、CD16+ 単球 (LYZ、FCGR3A)、および制御性 T 細胞 (FOXP3) のマーカーが使用された [26]。

2.2.4.5 健常と疾患の細胞種ごとの発現差遺伝子の同定結果

健常と疾患の細胞細と比較して、異なる発現を示す遺伝子を検出した。

健常と比較して疾患では、CD4 ナイーブ T 細胞は、JAK/STAT、PI3K/AKT、および ERK/MAPK シグナルパスウェイ遺伝子の発現上昇を示した。また、抗原による T 細胞刺激を示す証拠は見出すことはできなかった。CD4 ナイーブ T 細胞は、アポトーシス、分化調節疾患のマーカーが示された。また、ナイーブ T 細胞の生存を維持することが知られている ATM シグナルパスウェイ遺伝子が発現上昇していた。T 細胞の生存と分化の両方を調節する可能性がある PPAR シグナルパスウェイ遺伝子も発現低下していた。制御性 T 細胞を調節することが知られている PDGF および IGF-1 シグナルパスウェイ遺伝子を含む、自己免疫疾患関連遺伝子が発現上昇していた。Th17 と制御性 T 細胞分化に影響を与えることが知られている、HIPPO シグナルパスウェイ遺伝子と TGF β シグナルパスウェイ遺伝子の両方が発現上昇していた。以上をまとめると、これらの結果は、

表 2.4 制御性 T 細胞で検出された遺伝子

遺伝子	健常と疾患の logFC
<i>ATP6AP2</i>	0.38
<i>NT5C</i>	0.17
<i>YWHAE</i>	0.11
<i>PLEKHJ1</i>	0.08
<i>MRPL23</i>	0.08
<i>VPS26B</i>	0.08
<i>FAM45A</i>	0.07
<i>ELAVL1</i>	0.07
<i>C1D</i>	0.06
<i>ZBTB7A</i>	0.06

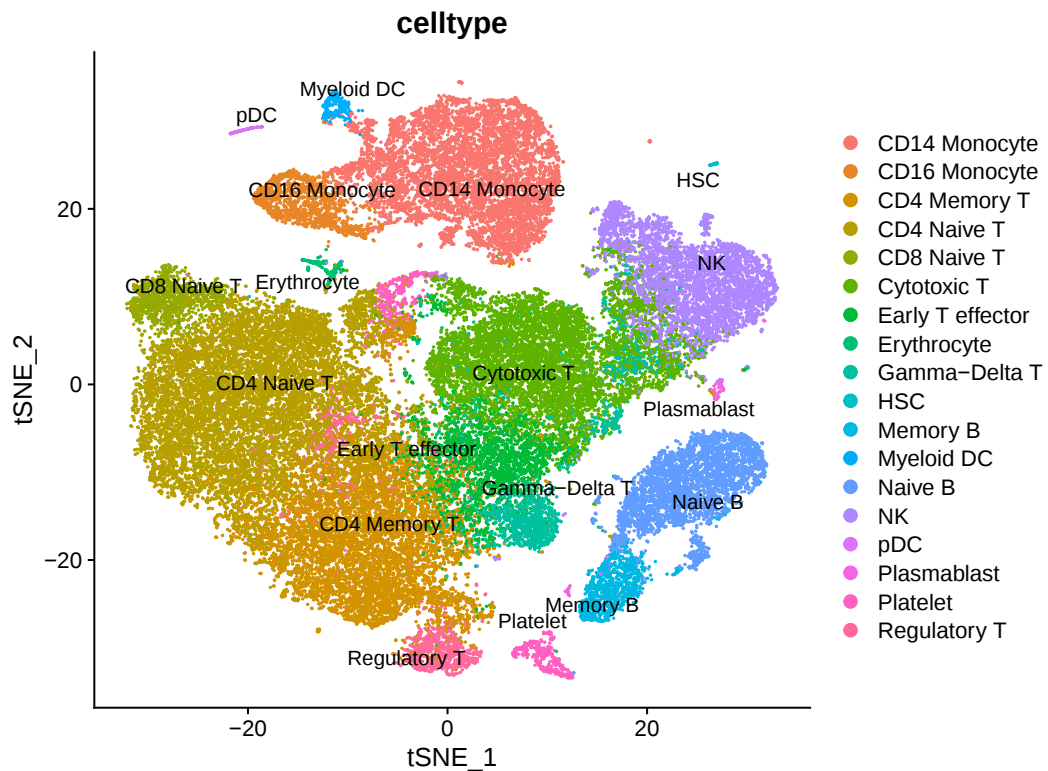


図 2.1 すべての細胞種の t-SNE 図

CD14 Monocyte:CD14 単球, CD16 Monocyte:CD16 単球, CD4 Memory T:CD4 メモリ T 細胞, CD4 Naive T:CD4 ナイーブ T 細胞, CD8 Naive T:CD8 ナイーブ T 細胞, Cytotoxic T: 細胞障害性 T 細胞, Early T effector: 初期エフェクター T 細胞, Erythrocyte: 赤血球, Gamma - Delta T: ガンマデルタ T 細胞, HSC: 造血幹細胞, Memory B: メモリー B 細胞, Myeloid DC: ミエロイド樹状細胞, Naive B: ナイーブ B 細胞, NK: ナチュラルキラー細胞, pDC: 形質細胞様樹状細胞, Plasmablast: 形質芽球, Platelet: 血小板, Regulatory T: 制御性 T 細胞

自己免疫疾患の特徴と一致していた。また、アポトーシスおよび分化メカニズムの両方の調節疾患が示唆された。

エフェクター T 細胞はアネルギー関連遺伝子を発現上昇させていた。すなわち、T 細胞受容体シグナルパスウェイを抑制していた。T 細胞のアネルギーは、CD28 結合などの共刺激の非存在下で抗原が感知されたときに T 細胞で確立される低応答状態として古典的に定義される末梢寛容のメカニズムである。アネルギーは、T 細胞受容体 (TCR、T cell receptor) シグナル伝達の遮断を介して、その後の抗原との遭遇に応答して細胞増殖とサイトカイン産生を防ぎ、自己免疫にならないように作用していると考えられてい

る。制御性 T 細胞またはエフェクター T 細胞の機能が働かないことが、この疾患では示唆されている。エフェクター T 細胞はアネルギー関連遺伝子をしていた。疾患における初期のエフェクター T 細胞 (ETE、early T effector) は、TCR シグナルパスウェイおよび ICOS-ICOSL シグナルパスウェイ、さらには PI3K/AKT、ERK/MAPK、NFAT、ERBB2、サーチュイン、および mTOR シグナルパスウェイが発現上昇していた。

先行研究 [26] では、免疫疾患が生じる機序として、古典的な単球と CD4 ナイーブ T 細胞を介した自然免疫と適応免疫の持続的な過剰活性化、制御性 T 細胞の機能疾患、およびエフェクター T 細胞のアネルギーが、今までになかった免疫疾患の原因であるという仮説を提唱している。

以上、まとめると既知の細胞種マーカーに基づくクラスタリングに続いて、18 種類の細胞について発現差のある遺伝子が特定された。さらにパスウェイ解析により古典的な単球の持続的な活性化を明らかにした。また、CD4+ ナイーブ T 細胞では、アポトーシスおよび Th17 のマーカーが多く発現していた。エフェクター T 細胞はアネルギー関連シグナルパスウェイの高発現を示していた。制御性 T 細胞で差次的に発現する遺伝子として、*TNFR2* をはじめとしていくつかを検出した。

2.2.5 先行研究における課題

先行研究では、クラスタリングを行った後、既知のマーカー遺伝子を基に細胞種を決定した。さらに健常と疾患との細胞種ごとに発現差のある遺伝子を検出することにより、疾患の特徴を解析した。この研究においては、健常と疾患における異なる状態における T 細胞の分化と遺伝子発現との関係を考慮していなかった。CD4+T 細胞とそのサブタイプの免疫疾患における重要性については、これまで報告がいくつかある [5, 48, 49, 50]。ただこれらの研究においても、どのように CD4+T 細胞がそのサブタイプへと分化していくのか、またそれがどのように疾患に結びついていくのかについてははっきりしたことはわかっていない。

一方村山らの先行研究の解析手法では、(1) がん細胞株のクラスタリング、(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といった解析を行っており、健常と疾患との差分析といったことは行われていない。

このため、村山らの解析手法を、T 細胞のサブタイプへの分化の細胞系譜推定に適用することで、健常と疾患とで発現差のある遺伝子を検出することを考えた。

2.3 本研究

2.3.1 本研究の目的

本研究においては、先行研究では解析されていなかった、疾患における T 細胞の細胞分化の過程を考慮した発現差遺伝子の検出を行うことを目的とする。すなわち、疾患に関わると報告されている T 細胞が疾患の進行に伴いどのように変化するかを明らかにし、通常のクラスタリングによる方法では検出できない、疾患の機序に関わる潜在的なマーカーを検出する手法開発の開発を目指した。Garman らの先行研究においても行われたように、よく行われている研究としては、クラスタリングを行い細胞種を決定した後、細胞種間の遺伝子発現の違いを比較することである。しかしながら、分化のような変化を伴う状態を正確に捉えるには限界がある。そこで本研究では、健常と疾患との異なる状態での細胞系譜解析を独立に行い、さらに特定のサブタイプへの分化系譜を比較することで、これまで検出できなかった遺伝子を検出する手法の開発を行い、従来のクラスタリングによる発現変化の解析では得られない、免疫細胞の分化と疾患に関する新しい情報を得ることを目的としている。この解析により Garman らの先行研究では得られないような、免疫細胞の分化の状態を考慮した新たな遺伝子が検出できることを期待した。提案する手法により疾患の理解が深まり、将来的には病気の診断や治療に応用されることが期待される。先行研究による方法の概略図を図 2.2 に、1.3 節で図 1.2 に本研究の方法の概略図を示している。

2.3.2 本研究の解析手法

解析では、疾患における T 細胞の細胞分化の過程を明らかにするために、まず分化過程を反映した系譜を同定し、さらに系譜ごとに健常状態と比較して疾患において発現差のある遺伝子を検出する。

本研究の解析手法は主に以下の 4 つのステップからなる。

ステップ 1 解析のためのデータセットの前処理

ステップ 2 細胞系譜の推定

ステップ 3 系譜間の差分発現解析により遺伝子を検出

ステップ 4 健常と疾患との差分発現解析により遺伝子を検出

といった手法で解析を行う。

それぞれのステップについての詳細について説明する。

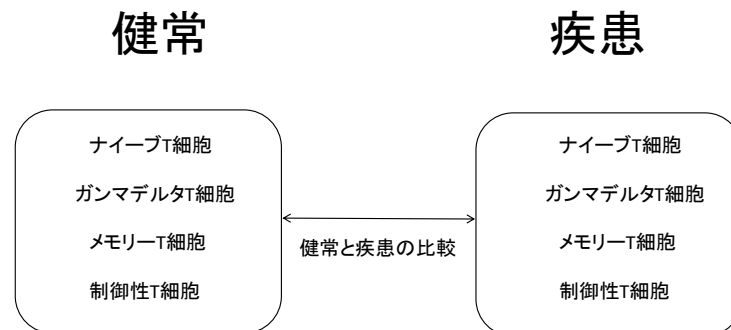


図 2.2 クラスタリング解析による健康と疾患の比較の概念図

2.3.2.1 解析のためのデータセットの前処理

シングル RNA-seq データを解析する上で、前処理として正規化を行い、細胞ごとの発現量の差を補正する。細胞あたりの発現量を各細胞の総発現量で割り、スケーリングファクターを掛け、対数値に変換する。次に、各遺伝子の発現量の対数値から平均値を引き、標準偏差で割ることにより、各遺伝子の全細胞における平均値と標準偏差を表す Z スコアに変換する。

2.3.2.2 細胞系譜の推定

細胞系譜の推定では、村山らの先行研究 [15] の手法に従って行う。具体的にいうと、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する。出発点のクラスターのラベルを定義する全ての系譜が出発点に指定したクラスターを共有するような、最小全域木を作成する。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する。この主曲線に沿って、疑似時間を定義する。細胞系譜を推定した後、各遺伝子を系譜の始点と終点の発現差を基に Wald 検定により評価する。有意差が得られた遺伝子をスコア順にランク付けする。系譜推定には、Slingshot[51] を使用する。

2.3.2.3 健常と疾患での系譜間の差分発現解析

同定された系譜を基に、系譜特異的な遺伝子発現を推定する。系譜のクラスターにまたがる遺伝子発現を計算するために、各系譜の疑似時間に沿った遺伝子発現を推定する。具体的には健常と疾患のそれぞれの系譜の分化終了点の間の発現の差を評価した [52]。健常と疾患のそれぞれについてガンマデルタ T 細胞（系譜 1）および制御性 T 細胞（系譜 3）の間に Wald 検定を実施した。この際発現量の差が 2 以上の遺伝子を対象にした。さらに疾患から健常の系譜間の比較による Wald 統計量を差し引く。本手法では健常と疾患との比較という適用の新規性に加えて、手法として、系譜間の差分解析を始めに行い、さらに健常と疾患の差分という 2 段階の発現差解析を行った。

これによりこれまで検出できなかった遺伝子が検出できると考えられる。ステップ 1,2,3 では村山らの細胞系譜推定手法を応用し、本研究においてステップ 4 の部分を拡張した。

以上、本解析手法は、以下の言語と計算環境で開発し実行した。言語 R 4.2.1 でパッケージとして Seurat 4.2.0, slingshot 2.4.0, tradeSeq 1.10.0, 計算環境 CPU Intel Xeon Gold 6138, メモリ 1TB, OS Ubuntu 20.04 を使用した。

2.3.3 本研究の解析結果

2.3.3.1 解析のためのデータセットの前処理結果

解析には Gene Expression Omnibus リポジトリに登録されている GSE132338 シングルセル RNA-seq データを使用した。このデータは、35 名の患者解析には者と 13 名の健常対照者の患者由来 PBMC(Peripheral Blood Mononuclear Cells) の細胞のシングルセル RNA-seq で得られたものである。インフォームドコンセントを行った後、被験者は血液サンプルを提供し、人口統計、薬物使用、臨床歴のアンケートに回答した。まず、以下の条件のいずれかに該当する細胞を個別に削除した。1 細胞あたりに発現する遺伝子の総数が 300 以上、ミトコンドリアの遺伝子の RNA リード数が 10 %未満の細胞を死細胞 RNA の遺伝子数が 8000 以上の細胞を細胞ダブレットであるとしてと条件を満たした細胞をのみを選択した。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。

2.3.3.2 細胞系譜の推定結果

細胞のクラスタリングと細胞種の決定に関しては、先行研究 [26] に従った。その後、細胞系譜の推定では、村山らの先行研究 [15] の手法に従った。

この解析では、先行研究 [26] にある細胞種のアノテーション情報を使用した。すべての細胞種の UMAP 図を示した (図 2.3)。T 細胞の分化を解析するために、CD4 メモリー T 細胞、CD4 ナイーブ T 細胞、初期エフェクター T 細胞、ガンマデルタ T 細胞、制御性 T 細胞の抽出を行った。PCA (principal component analysis) 解析により 20 次元までのデータで UMAP による次元削減を行った。この結果、5 つの大きなクラスター (CD4 メモリー T 細胞、CD4 ナイーブ T 細胞、初期エフェクター T 細胞、ガンマデルタ T 細胞、制御性 T 細胞) が観察された (図 2.4)。

細胞系譜の推定では、分化の起点として考えられる CD4 ナイーブ T 細胞を出発クラスターとし、終止クラスターをガンマデルタ T 細胞 (系譜 1)、CD4 メモリー T 細胞 (系譜 2)、制御性 T 細胞 (系譜 3) として設定する。

その 3 つの系譜を健常な細胞と疾患のそれぞれの結果を示した。(図 2.5, 図 2.6)。これらの系譜は、CD4 ナイーブ T 細胞からガンマデルタ T 細胞 (系譜 1)、CD4 メモリー T 細胞 (系譜 2)、制御性 T 細胞 (系譜 3) へと進行する系譜を表している。CD4 ナイーブ T 細胞から各 T 細胞サブタイプへの分化過程は、解析によって可視化された。

2.3.3.3 健常と疾患での発現差遺伝子の解析結果

統計的に有意性の高い上位遺伝子を示した (表 2.5)。

出発点 (CD4 ナイーブ T 細胞) の平均発現量は、終点 (ガンマデルタ T 細胞、CD4 メモリー T 細胞、制御性 T 細胞) の平均発現量と等しいという帰無仮説を、Wald 検定で評価した。次に、ガンマデルタ T 細胞 (系譜 1) の Wald 統計量から制御性 T 細胞 (系譜 3) の Wald 統計量を差し引いた。さらに疾患の系譜 1 と系譜 3 の Wald 統計量から健常の系譜 1 と系譜 3 の Wald 統計量を差し引いた。検出された遺伝子が真に T 細胞の系譜を特徴づけているという生物学的な証拠としては系譜 1 に関してガンマデルタ T 細胞に特徴的な *GNLY*, *NKG7* が発現 [34] しているため系譜 1 がガンマデルタ T 細胞への分化の特徴をとらえていると考えられた [53]。遺伝子 *NKG7* と *GNLY* は Garman らの先行研究 [26] で報告されており、*CCL*、*CST7*、*GZMA*、*TRDC* は 2.2.4 節で述べたマーカー解析により見出されている。興味深いことに、本手法により、*CTSW*、*PLAC8*、*ID2*、

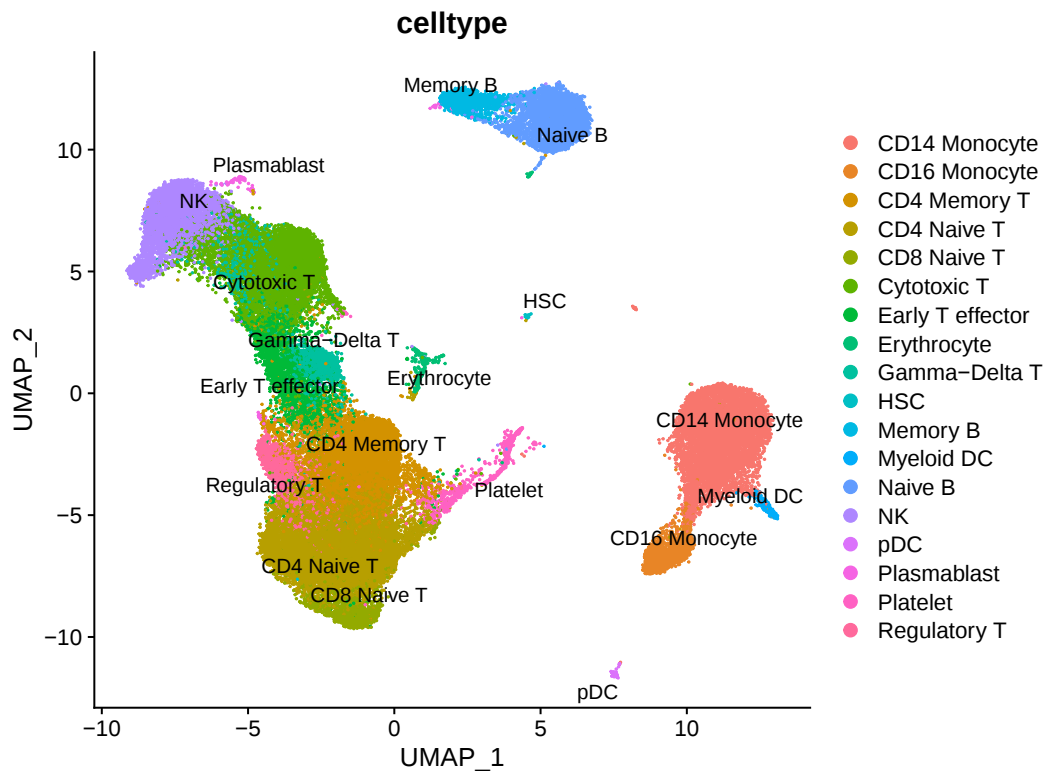


図 2.3 すべての細胞種の UMAP 図

CD14 Monocyte:CD14 単球, CD16 Monocyte:CD16 単球, CD4 Memory T:CD4 メモリ T 細胞, CD4 Naive T:CD4 ナイーブ T 細胞, CD8 Naive T:CD8 ナイーブ T 細胞, Cytotoxic T: 細胞障害性 T 細胞, Early T effector: 初期エフェクター T 細胞, Erythrocyte: 赤血球, Gamma - Delta T: ガンマデルタ T 細胞, HSC: 造血幹細胞, Memory B: メモリー B 細胞, Myeloid DC: ミエロイド樹状細胞, Naive B: ナイーブ B 細胞, NK: ナチュラルキラー細胞, pDC: 形質細胞様樹状細胞, Plasmablast: 形質芽球, Platelet: 血小板, Regulatory T: 制御性 T 細胞

ANXA1 という新しい遺伝子が検出された。上記の各遺伝子について、系譜の進行に伴うマーカーの発現の変化を示す (図 2.7)。

この 4 つの遺伝子は、健常と疾患の系譜 1 と系譜 3 の Wald 統計量に大きな差があることがわかる (表 2.7)。それぞれの遺伝子の系譜と系譜時間の解析と検出された遺伝子の系譜時間との対応は以下のとおりである。*ANXA1* については系譜 1 において疾患においては系譜の最終段階で健常ではやや下がってきている。系譜 3 ではそれほど大きな違いは見られない。共に最終段階では減少傾向にある。また、系譜 1 の疑似時間の中間的な位置で急激に健常と疾患の発現量が上昇している。これはナイーブ T 細胞から初期エフェク

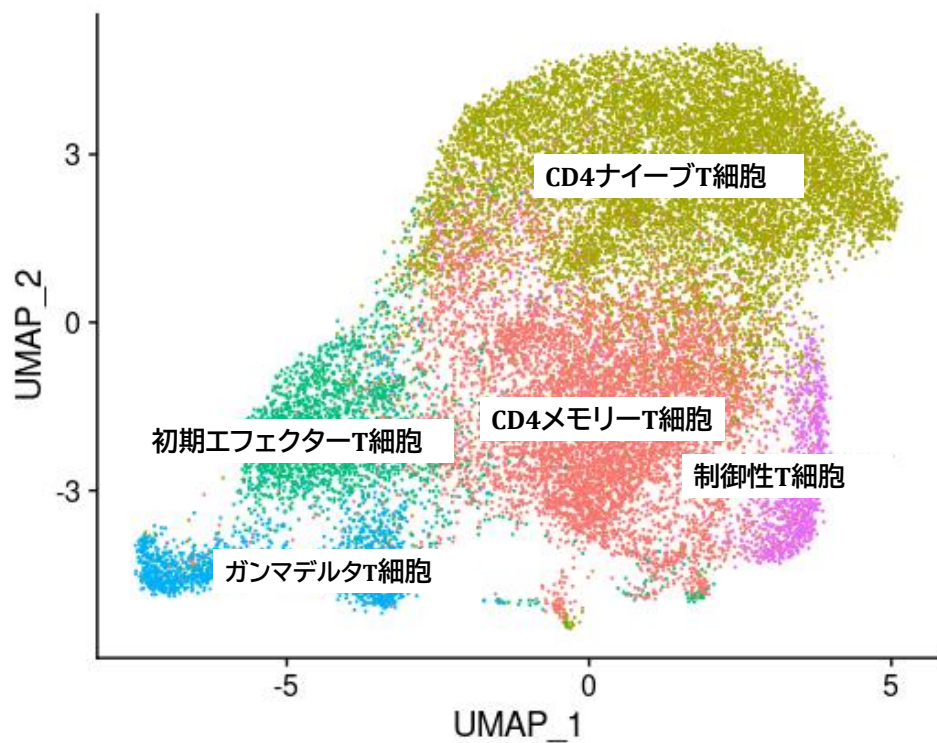


図 2.4 T 細胞の UMAP 図

表 2.5 系譜 1 と系譜 3 の差で検出された遺伝子

遺伝子	系譜 1 と系譜 3 の Wald 統計量の差	系譜 1 と系譜 3 の logFC
<i>CCL5</i>	203.60	4.88
<i>NKG7</i>	115.73	6.08
<i>GNLY</i>	113.99	6.29
<i>CST7</i>	111.71	2.57
<i>CTSW</i>	99.62	3.77
<i>PLAC8</i>	91.74	3.42
<i>GZMA</i>	84.04	2.73
<i>ID2</i>	83.04	3.29
<i>TRDC</i>	80.39	5.26
<i>ANXA1</i>	78.82	2.11

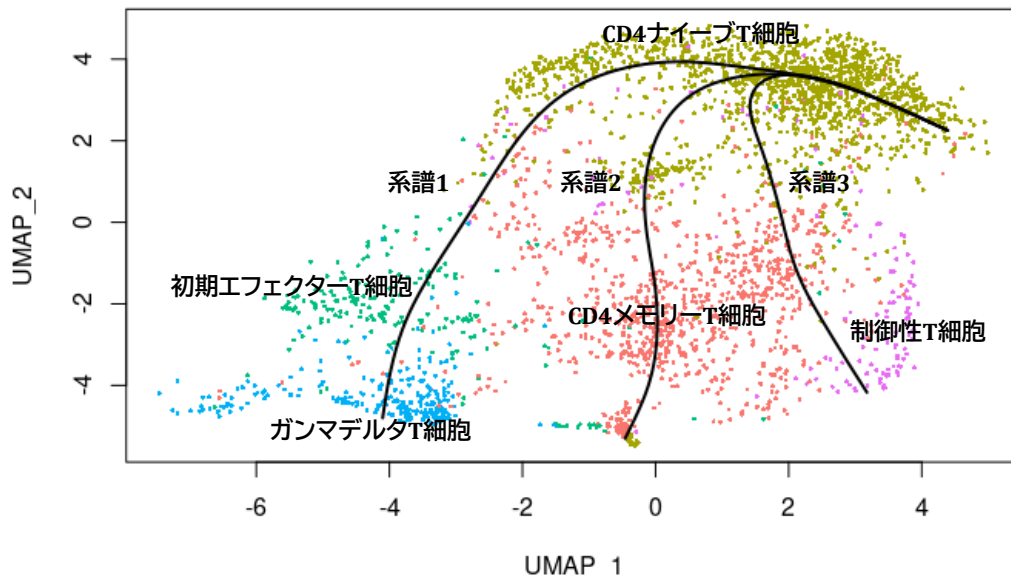


図 2.5 健康細胞の細胞系譜

ター T 細胞に分化している位置と重なり、分化の特徴をとらえていると考えられる。さらに分化が進むといったん下がるが疾患だけ再上昇しているというのは疾患との関連性が示唆される。*CTSW* については系譜 1 では疾患、健康共に上昇傾向であるが、疾患は最終的な発現が大きい。系譜 3 については共に低発現の状態が続いている。また、系譜 1 の疑似時間の中間的な位置で急激に健康と疾患の発現量が上昇していし、疾患の方が発現量が常に健康よりも多い。これはナイーブ T 細胞から初期エフェクター T 細胞に分化している位置と重なり、分化の特徴をとらえていると考えられる。さらに分化が進んでも同じように上昇し続ける特徴がある。これは *ANXA1* とは分化に伴う発現が異なっている。*PLAC8* については、系譜 1 では疾患、健康共に上昇傾向であるが、分化の最終段階で疾患ではさらに上昇している。系譜 3 ではそれほど大きな違いは見られない。共に最終段階では減少傾向にあり疾患ではより大きく減少している。また、系譜 1 の疑似時間の中盤から後半にかけて健康と疾患の発現量が上昇していし、疾患の方が発現量が常に健康よりも多い。系譜 3 はナイーブ T 細胞から制御性 T 細胞への系譜であるが、疾患では最終分化段階のやや手前で、健康ではよりその前で低下傾向にあり、現象までが長くどまり続け

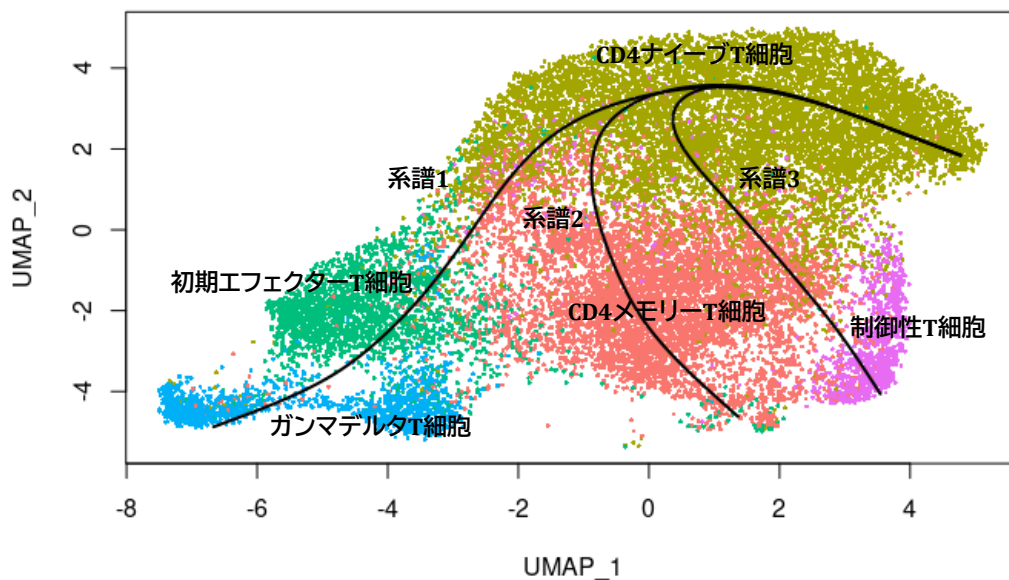


図 2.6 疾患細胞の細胞系譜

ることが疾患との関連性を示す可能性が考えられる。*ID2* については、系譜 1 では疾患、健常共に上昇傾向であるが、健常において分化の最終段階ではさらに上昇している。系譜 3 については共に低発現の状態が続いているが、開始時は健常の方が発現量も多い。

遺伝子 *ID2* に関しては、疾患の系譜 1 と系譜 3 の発現量の差は健常のそれよりも小さいように見えるが、統計学的に判断すると疾患と健常の系譜間の Wald 統計量の差は大きいことがわかる (表 2.7)。

2.4 本研究の考察

提案手法では細胞系譜解析による T 細胞の分化を考慮した健常と疾患の発現差遺伝子の検出手法について提案した。サルコイドーシスは、遺伝的素因と免疫制御異常が関与すると報告されている確立されているが [18], [19]、それ以外の病因がない謎めいた疾患である。このため、サルコイドーシスに固有の免疫調節異常を理解することが重要である。従来の複数の細胞を平均化する RNA-seq を超えるシングル RNA-seq による遺伝子発現マーカー同定に関する有効な手法の開発が現在も進められている。本研究では、T 細胞の

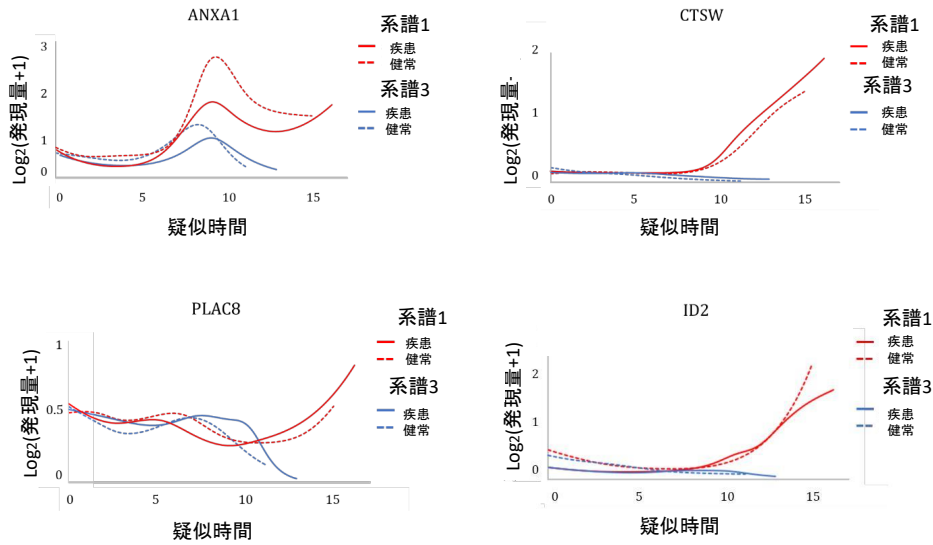


図 2.7 疑似時間に沿った発現遺伝子の変動

1 つであるガンマデルタ T 細胞の系譜と制御性 T 細胞の系譜を比較し発現差のあった遺伝子について健常と疾患とで比較した。ガンマデルタ T 細胞は、主に可溶性メディエーターの産生を介して、病原体の除去、創傷治癒、自己免疫、がんにおいて無数の役割を果たすことが知られている未解明な部分の多い細胞種である [54]。

ここで本研究において検出された遺伝子を評価するため、クラスターベースの発現差遺伝子解析法との比較を行う。疾患のガンマデルタ T 細胞と分化の進んでいない細胞である CD4⁺ ナイーブ T 細胞と比較しガンマデルタ T 細胞の発現差遺伝子を検出する。疾患の正規化したデータを用いて、Wilcoxon 検定を用い、発現差の対数の変化が 0.1 の閾値以下のパラメータで、それぞれの細胞のサブタイプにおいて 10% 以上は発現する遺伝子に関して細胞間の遺伝子の発現差を計算する。検出された遺伝子について表 2.6 に示した。

検出された遺伝子と比較すると、トップ 10 のうち 6 つの遺伝子 (*GZMA*, *TRDC*, *CST7*) はクラスターベースと細胞系譜解析とで共通であった。このうち、*KLRB1* は先行研究で報告されている遺伝子であり、*TRGC1* と *DUSP2* (dual specificity phosphatase 2) は細胞系譜間差解析では検出されなかった。*PLAC8*、*ID2*、*CTSW*、*ANXA1* はクラスターベースの解析では見つからなかった。

我々の提案する手法である健常と疾患における差分系譜解析と従来のクラスターベースの手法により共通して検出される上位 10 遺伝子の内の 6 つの遺伝子のうち、*NKG7* と

GNLY の 2 つは既に先行研究においても報告されていた [1]。先行研究では検出されていない他の 4 つの遺伝子 (*CCL5*, *CST7*, *GZMA*, *TRDC*) は、以下のように知られている。

CCL5 は肺サルコイドーシスと関連がある [20]。*CST7* は、サルコイドーシスで最も高発現している遺伝子の一つである [21]。*GZMA* は、highly effector memory T cell (TEMRA, Terminally differentiated effector memory T cells re-expressing CD45RA) で高発現している [55]。TEMRA メモリー T 細胞は、エフェクターメモリー T 細胞のサブセットで、抗原刺激後に CD45RA を細胞表面に表す未知の役割や機能を持つ細胞である。CD4+TEMRA 細胞は、デングウイルスなどの病原体に対する防御免疫に関与していることが示唆されている [56]。*TRDC* は、ガンマデルタ T 細胞受容体の構成要素の一つで、ガンマデルタ T 細胞の発生に不可欠な T 細胞受容体 δ の一定領域をコードしている [57]。クラスターベースの発現差遺伝子解析でだけ検出された遺伝子の中に、*TRGC1* と *DUSP2* がある。*TRGC1* (T 細胞受容体 γ 鎖の定数領域) は、抗原認識に関与している。ガンマデルタ TR は、様々な自己および外来の非ペプチド抗原を認識する [58]。抗原を認識すると、病原体の除去や組織の修復に関与する、迅速で自然な免疫応答が誘導される [59]。*DUSP2* は、phosphatase of activated cells 1、*PAC1* として知られ、T 細胞の抗腫瘍免疫における免疫チェックポイントとして機能している。*PAC1* は、疲弊した腫瘍浸潤リンパ球で選択的に発現が上昇し、がん患者の予後不良と関連している [60]。本手法では、特定の系譜間の発現差解析を行い、ガンマデルタ T 細胞で差次的に発

表 2.6 クラスターベースの解析で検出された遺伝子

遺伝子	avg.logFC
<i>GNLY</i>	3.349
<i>NKG7</i>	3.33
<i>CCL5</i>	2.77
<i>KLRB1</i>	2.74
<i>GZMA</i>	2.27
<i>TRDC</i>	2.22
<i>CST7</i>	2.09
<i>TRGC1</i>	2.08
<i>DUSP2</i>	1.87

現する新たな遺伝子を同定した。すなわち、系譜 1 (ガンマデルタ T 細胞) と系譜 3 (制御性 T 細胞) を比較した。各系譜間の差分解析により、*CTSW*、*PLAC8*、*ID2*、*ANXA1* が検出された。*CTSW* は *GZMA* と同様に TEMRA で高発現している [55]。*PLAC8* は、肺結核と肺サルコイドーシスで転写された血液と比較して、肺サルコイドーシスでよく発現している遺伝子であった [21]。*ID2* は肺 ILC 細胞 (Innate lymphoid cell) の発生に関与する転写因子であり、*ILC* は肺機能、バリア性、肺損傷後の呼吸器組織のリモデリングに関与している [61]。*ANXA1* リポコルチン-1 (アネキシン-1 としても知られる) は、副腎皮質ホルモンによって発現が上昇し、その抗炎症作用の多くに顕著な役割を果たし [22]、内因性抗炎症のメディエーターの 1 つとして、多くの文脈において白血球の輸送および活性化の調節を与えてる [62]。*ANXA1* ペプチド模倣品を治療的に投与すると、炎症と線維化の両方の徴候が改善されたという報告もある [62]。

さらに、偽陽性の遺伝子を除くことを目的として、先行解析以外の別の既知情報を基に本解析において検出された遺伝子について評価した。本解析において上位に検出された *CCL5*、*NKG7*、*GNLY* はいずれも最終分化エフェクター T 細胞 (TEMRA) に多く発現しているという報告がある [63]。最終分化エフェクター T 細胞とは、*NKG7* は細胞障害性のマーカー遺伝子としても知られていて最終分化エフェクター T 細胞に特に発現している遺伝子である。これらの知見を踏まえて我々の解析結果を考えると、CD4+T 細胞の分化後のサブクラスであるガンマデルタ T 細胞にもこれらの遺伝子が発現していることは興味深い。また、また細胞傷害性 T 細胞にも上記の遺伝子や *GZMA* は高発現することが知られている [64]。また、CD8 + エフェクター T 細胞 (TEM) は *CCL5* を高レベルで発現することが知られている。レジデントメモリー T 細胞 (TRM) シグネチャー遺伝子は、血液と比較して組織、特に肺でアップレギュレートされている。これらの知見は、組織における T 細胞の局在化には、組織マトリックスとの相互作用を促進する細胞の構造変化が関与している可能性が高いことを示唆している。その一つである *ANXA1* はその中でリンパ液と血液を比較して血液で高い発現があった [65]。このことは我々の研究結果でも血液中で高い発現をしていたことと一致している。

また、上位の 10 遺伝子を含むそれ以下の遺伝子について、Garman らの先行研究、提案手法、クラスターベースの手法で検出された遺伝子遺伝子のベン図を示す (図 2.8)。この結果では、先行研究により検出された遺伝子の内、本研究で提案した細胞系譜推定を用いた手法の方がクラスターベースの検出よりも重複した遺伝子が少なかった。結果の重なり的大小については、Garman の先行研究とクラスターベースの検出が共通の手法を用い

ていることが寄与したと考えられる。

参考として上位 50 までの Garman らの先行研究 (表 4.3)、クラスター解析 (表 4.2)、細胞系譜解析 (表 4.1) による検出遺伝子を付録に記載した。本結果においても、クラスター解析 (表 4.2)、細胞系譜解析 (表 4.1) による共通する遺伝子については、Garman らの先行研究と比較して多い。このため Garman らの先行研究について重複が少ないのは、クラスター解析における発現差遺伝子の検出方法などが影響した可能性が考えられる。クラスター解析で用いた従来知られている遺伝子マーカーと新規に検出された系譜解析による遺伝子とで、どのような系譜の発現パターンの差があるかについては、クラスター解析で用いた従来知られている遺伝子マーカーと新規に検出された系譜解析による遺伝子とで、どのような系譜の発現パターンの差があるかについては、前者ではガンマデルタ T 細胞と制御性 T 細胞の細胞種間の発現量の平均値の差が大きい遺伝子が検出される。後者では系譜の開始と終了において発現差が大きい遺伝子を検出しやすいと考えられる。

また Garman らの先行研究では、各クラスターで他のクラスターと比べて発現差の大きい遺伝子を疾患と健常のデータそれぞれについて検出しており、本論文のクラスターベースの解析ではガンマデルタ T 細胞のクラスターとナイーブ T 細胞のクラスター間で疾患と健常で発現差がある遺伝子を検出している。すなわち、発現差を比較するクラスターのペアが一部異なるだけで、発現差を検出する手法は同じであるため、両者の検出結果の重なりが大きいことが考えられる。

このことから、疾患の進行機構を理解するための細胞系譜解析による検出は、T 細胞のサブタイプへの分化を考慮することにより、従来のクラスタリング法とは異なる、別の手段を提供することができることが示唆された。

これらの結果から、クラスター解析では検出できない免疫疾患に関連する新規の遺伝子の検出に、健常と疾患との細胞系譜解析において、系譜間の差異解析法が有効であること

表 2.7 健常と疾患の系譜 1 と系譜 3 の差の Wald 統計量

遺伝子	疾患の系譜間の Wald 統計量の差	健常の系譜間の Wald 統計量の差
<i>PLAC8</i>	92.85	0.00
<i>CTSW</i>	125.51	0.43
<i>ID2</i>	114.57	0.008
<i>ANXA1</i>	90.14	2.24

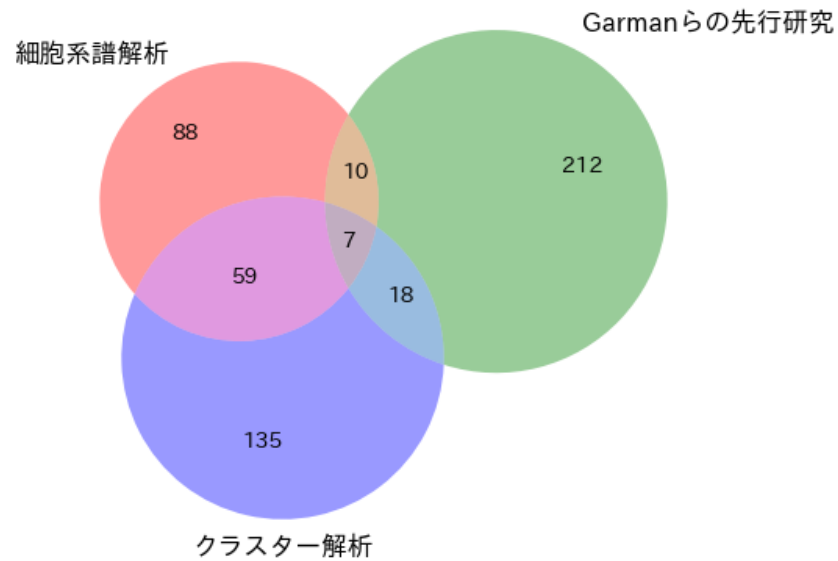


図 2.8 解析手法による Garman らの先行研究と検出遺伝子の比較

が示唆された。

2.5 結言

本論文では、シングルセル RNA-seq データを使用して、免疫疾患の一つであるサルコイドーシスと健常対照者の T 細胞のサブタイプで異なる発現の遺伝子を検出する新しい方法を提案した。すなわち、ナイーブ T 細胞からガンマデルタ T 細胞への系譜と、ナイーブ T 細胞から制御性 T 細胞への系譜から、その分化の最終段階の系譜間の発現差遺伝子を検出した。その後、さらに健常と疾患との発現差遺伝子を検出した。本解析法を行うことにより、*CTSW*、*PLAC8*、*ID2*、*ANXA1* といった遺伝子を検出することに成功した。これは、従来のクラスターを用いた変動遺伝子解析ツールや通常の解析では検出できなかったものである。健常と疾患との状態の異なる系譜間での差分解析は、免疫疾患の進行に伴う T 細胞の状態変化の過程を理解するための有効なツールとなりうる。一方で、系譜の比較において健常と疾患ともに同じような発現量で発現しているような遺伝子については検出しづらいという短所もある。今後は、より多くの潜在的なバイオマーカーを標的として同定できるツールの開発を目指している。

第 3 章 シングルセル RNA-seq データを用いた細胞間情報伝達の推定による免疫制御異常マーカー遺伝子の探索手法の研究

3.1 緒言

シングルセル RNA-seq は、個々の細胞における遺伝子発現を定量的に解析することができる。細胞間の情報伝達の種類には、細胞同士が直接的に表面の蛋白質を介して行う細胞接着や細胞マトリクスと受容体を介した相互作用や細胞から分泌されるリガンドと細胞膜表面に存在する受容体と呼ばれる蛋白質を介して行う方法があるが、細胞間の情報伝達を解析することで、疾患と関連するシグナル伝達経路に関する知見を得ることができると考えられる。

シングル RNA-seq データでは細胞種ごとに発現遺伝子を定量的に解析できることから、細胞間情報伝達についても定量的に推測および分析できると考えられる。さらに健常と疾患との違いを分析することで、疾患に特異的に関連するシグナル伝達経路の洞察を得られることが期待される。

近年、シングルセル RNA-seq によるサルコイドーシスの末梢血液の免疫細胞に関する先行研究が報告された [26]。その結果では、古典的単球の持続的な活性化と、それに続くトラフィック分子というリンパ球の組織への移動を促進する遺伝子の発現上昇が明らかになった。さらに、疾患特有の T 細胞サブセットにはアネルギー関連の遺伝子の異常のパターンが見られた。このようにシングルセル RNA-seq による細胞種の異常に関する知見も明らかになりつつある。

第 2 章では、この先行研究の結果に基づいて疾患および健常の T 細胞に関する細胞系譜解析を行った [66]。その結果、疾患と健常の状態の異なる T 細胞の細胞系譜の解析は、これまで知られていなかったいくつかの遺伝子が疾患に関与していることを示唆するものであった。

一方で、サルコイドーシスをはじめ、様々な疾患の病態をより詳細に理解するためには、免疫の中心的な役割を担っている T 細胞と、T 細胞を含むその他のすべての細胞種がどのように情報伝達しているかを明らかにすることが有効であると考えられる。また、細胞間情報伝達が全身性エリテマトーデス (SLE、systemic lupus erythematosus) や関節リウマチなどの様々な自己免疫疾患に関与しているという報告もある。全身性エリテマ

トーデスと関節リウマチの細胞間情報伝達解析の結果から、この2つの疾患は、免疫的に異なる病態であることが示唆された [67]。

このような背景の中、T細胞のような免疫細胞に関する疾患に特徴的な細胞間情報伝達を明らかにすることにより、疾患の理解を深め治療につながる可能性が考えられる [17]。

具体的には、健常と疾患との細胞間の情報伝達の違いから、より詳細な疾患の状態を理解すること必要であると考えられる。このためには、例えば健常と疾患との、シグナルパスウェイの情報伝達の活性の違いを示すことにより、その詳細な理解に対する足がかりを得られると考えられる。つまり、数多くあるシグナルパスウェイからどのシグナルパスウェイが疾患の発症や進行に関わるのかといったことを理解することが疾患の理解やさらには創薬のターゲットを検出することには必要である。なぜなら、生体内では様々なシグナルパスウェイとして機能を持つことにより機能を生み出しているために、単独の遺伝子の発現情報のみの情報よりも、シグナルパスウェイとしての単位で疾患における健常との状態変化を検出することが疾患の理解には有効であると考えられるためである。T細胞が関与する自己免疫疾患の発症や進行の現象に関与するようなシグナルパスウェイを分子レベルで明らかにすることができれば、疾患の理解や治療法の手がかりを得やすいだろうと考えられた。この目的を検討するために、自己免疫疾患おいてのT細胞間の情報伝達が関与することが報告されている [17] 中、シングルセル RNA-seq を利用することにより、T細胞が関与する細胞種間の情報伝達の活性を解析できる可能性が浮上していた。

以上、疾患の病態をより詳細に理解することにより、診断や治療への応用が期待されるため、我々は、健常と疾患との異なる細胞における T細胞が関与する細胞間情報伝達解析手法の開発を行い疾患においての細胞の情報伝達活性の変化をシグナルパスウェイとして検出することを目指した。

3.2 先行研究

3.2.1 先行研究の目的

Jin らの先行研究 [68] 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述べる。細胞間の相互作用を理解するためには、細胞間の情報伝達を表現しシステムレベルで効果的に解析することが必要である。Jin らは、リガンド、受容体、およびそれらの補因子間の相互作用のデータベースを構築した。シングルセル RNA-seq データから細胞間情報伝達ネットワークを定量的に推測・解析するツールを開発した。これにより細胞間の情報伝達におけるシグナルパスウェイを検出できる。ヒ

トの皮膚の病変と非病変部の比較により、シグナルパスウェイの活性化状態を検出した。

3.2.2 先行研究のデータセット

Jin らの先行研究では、ヒト疾患皮膚データセット アトピー性皮膚炎患者の病変部および非病変部のヒト皮膚サンプルの 17,349 細胞のトランスクリプトームデータを利用している。得られたシングルセル RNA-seq データが Gene Expression Omnibus リポジトリに GSE14742470 として登録されている。

3.2.3 先行研究の解析手法

現在までに提案されているシングル RNA-seq を使用した細胞間情報伝達解析手法を以下に説明する。

3.2.3.1 リガンド-受容体相互作用のデータベースの構築

まず、対象とするリガンド-受容体相互作用のデータベースを構築するために、一般に公開されている他のシグナル伝達パスウェイデータベースや査読付き文献を手作業でレビューする。このデータベースのリガンド-受容体相互作用は、KEGG シグナリングパスウェイデータベースに基づいて手動でキュレーションされたものである。リガンド-受容体相互作用の組成だけでなく、可溶性アゴニストとアンタゴニスト、そしてリガンド-受容体を介した情報伝達をに関わる共刺激性と共抑制性の膜結合型受容体を含む補因子分子も考慮に入れる。さらに生物学的に意味のある方法で細胞間情報伝達解析をするために、229 種類のシグナル伝達経路ファミリーにグループ化する。

3.2.3.2 細胞間情報伝達の解析

細胞間情報伝達の推定についての概要を説明する。それぞれの細胞種で他の細胞種との発現差遺伝子を検出する。次に、それぞれの細胞種において遺伝子の平均発現量を計算する。

細胞間情報伝達の推定については以下のように行う。

ステップ a 細胞種により発現量の異なる遺伝子を同定する。

細胞種特異的な遺伝子を推測するために、与えられた scRNA-seq データセット内のすべての細胞種にわたって、有意水準 0.05 で Wilcoxon 順位和検定を使用して、差次的に発現した遺伝子を同定する。

ステップ b アンサンブル平均発現の計算をする。

ステップ a で同定された遺伝子に関して、ノイズの影響を考慮し、統計的にロバストな平均法を用いて、与えられた細胞種におけるシグナルパスウェイ遺伝子の平均発現を算出した。1 つの細胞種における遺伝子が発現している細胞の割合が 10 % 未満の場合、その遺伝子の平均遺伝子発現はゼロとする。

$$EM = \frac{1}{2}Q_1 + \frac{1}{4}(Q_2 + Q_3) \quad (3.1)$$

ここで、 Q_1 、 Q_2 、 Q_3 は、それぞれ細胞種におけるシグナル伝達遺伝子の発現量の第 1、第 2、第 3 四分位を表す。

ステップ c 細胞間情報伝達確率の計算。ステップ a で同定、ステップ b) の発現量を基に、リガンドと受容体の結合の物理的過程は蛋白質-蛋白質相互作用を含むので、遺伝子発現プロファイルを実験的に検証された信頼度の高い蛋白質-蛋白質ネットワークに含まれる遺伝子について伝達確率をモデル化した数式により計算する。投影されたリガンドと受容体のプロファイルに基づいて、特定のリガンドと受容体のペア k に対する細胞種 i から j への伝達確率 $P_{i,j}$ をモデル化する。

$$P_{i,j}^k = \frac{L_i R_j}{K_h + L_i R_j} \times \left(1 + \frac{AG_i}{K_h + AG_i}\right) \cdot \left(1 + \frac{AG_j}{K_h + AG_j}\right) \times \frac{K_h}{K_h + AN_i} \cdot \frac{K_h}{K_h + AN_j} \quad (3.2)$$

$$L_i = \sqrt[m_1]{L_{i,1} \cdots L_{i,m_1}}, R_i = \sqrt[m_2]{R_{i,1} \cdots R_{i,m_2}} \cdot \frac{1 + RA_i}{1 + RI_j}$$

ここで、 L_i と R_j は細胞種 i と j におけるリガンド L と受容体 R の発現量を表している。サブユニットが m_1 個のリガンド L の発現量は幾何平均 (i.e. $L_{i,1} \cdots L_{i,m_1}$) により推定されており、どのサブユニットの発現量がゼロでもリガンドの活性はなくなることが示唆される。同様に、 m_2 サブユニットを用いて、受容体 R の発現量を算出する。さらに、膜結合型共刺激性受容体と共抑制性受容体は、受容体活性を調節することによってシグナル伝達を修正する [69]。多数の共刺激性受容体 (RA)、共抑制性受容体 (RI) を有するリガンド-受容体ペアの共刺激性受容体の平均発現量を計算し、線形関数を利用して受容体発現量をモデル化している。正規化した入力データの範囲を 0 から 1 であるようにするために、パラメータ K_h をデフォルト値の 0.5 に設定して、 L と R の間の相互作用を表現する。 K_h については

Jin らの先行研究において値を変動させて、検出した細胞間情報伝達の種類について検討した結果、 K_h を細かく振った結果との細胞間情報伝達の種類との類似性が高いことから、検出性能が高い値であるという事前の検討をした上で選択される。また、検出されるシグナルパスウェイの評価については、オリジナルデータからサブサンプルデータを抽出し、サブサンプルデータから推定された相互作用の頑健性を、サブサンプルデータとオリジナルデータの比較により、真陽性率（TPR）、偽陽性率（FPR）、精度（ACC）の3つの指標で評価している。

$$TPR = \frac{TP}{TP + FN} \quad (3.3)$$

$$FPR = \frac{FP}{TN + FP} \quad (3.4)$$

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.5)$$

TP（真陽性）は、サブサンプルデータセットから推定された相互作用が、オリジナルデータセットから推定された相互作用と一致した数であり、FP（偽陽性）は、サブサンプルデータセットから推定された相互作用が、オリジナルデータセットから推定された相互作用と一致した数である。FP (false positive) はサブサンプルデータセットの相互作用の総数から TP を引いた値である。TN (true negative) はサブサンプルデータセットとオリジナルデータセットの両方に含まれない相互作用の数である。FN (false negative) はサブサンプルデータセットでマッチしないオリジナルデータセットの相互作用の数である。

リガンド-受容体相互作用は、送り手と受け手の両方の細胞からの細胞外アゴニストとアンタゴニストによって直接的または間接的に修正できる。リガンド-受容体ペアに対する多数の可溶性アゴニストの平均発現量を計算する（AG で示す）。多数の可溶性アンタゴニスト、アンタゴニストを持つリガンド-受容体ペアを同じ方法でモデル化する。3次元配列 $P (K \times K \times N)$ を用いて、リガンド-受容体のすべてのペアにまたがる細胞種間の伝達確率を表現する。K は細胞種の数、N はリガンド-受容体のペアリングまたはシグナル伝達経路の数である。関連するリガンド・レセプターのペアリングの確率を加算することで、シグナル伝達経路の伝達確率が計算される。

ステップ d 統計的に有意な細胞間情報伝達の同定

生物学的に有意な細胞間情報伝達を推定するため、並べ替え検定を行い、各相互作用に確率値を割り当てる。細胞種ラベルをランダムに並べ替えることで、並べ替え検定を行い、2つの細胞種間の有意な接続を見つけることができる。そして、細胞種 i と細胞種 j の間の情報伝達確率 $P_{i,j}$ は、リガンド L と受容体 R のペアを用いて再計算される。各 $P_{i,j}$ の p 値は、次のように計算される

$$p = \frac{\#\{mP_{i,j}^{(m)} \leq P_{i,j}, m = 1, 2, \dots, M\}}{M} \quad (3.6)$$

ここで、 M は並べ換えの総数であり、デフォルト値は 100 である。 $p < 0.05$ の相互作用は有意であるとみなす。

3.2.4 先行研究の解析結果

全細胞のグラフに基づく各細胞の k -最近傍の計算を介して構築されるグラフに Louvain アルゴリズムを適用して、細胞種にクラスタリングした。その結果、線維芽細胞 (FIB)、樹状細胞 (DC)、T 細胞 (TC) を含む 10 の細胞型に分別された細胞集団が同定された。細胞間情報伝達の解析には、病変および非病変皮膚からの 7563 細胞を使用した。病変皮膚では、既知の炎症性シグナルである CSF、IL4、CCL、TNF を含む 4 つの経路が特異的に活性化しており、これらの経路が疾患の進行に決定的に寄与している可能性が示唆された。CCL シグナルにおいては、リガンド-受容体ペア CCL19-CCR7 を重要なシグナルと同定した。リガンド MIF とそのマルチサブユニットレセプター CD74/CD44 は、病変皮膚では非病変皮膚と比較して、シグナルとして活性化することが分かった。

3.2.5 先行研究の課題

Jin らの研究では、皮膚疾患の病変と非病変部の細胞間の情報伝達解析を行い、病変部において活性化しているシグナルパスウェイを情報伝達確率として解析している。一方、健常と疾患の比較による細胞間の情報伝達解析やそのシグナルパスウェイの解析については行われていない。したがって、免疫疾患のような多様な病態について細胞間の情報伝達解析を行うことにより、疾患で特に活性化しているシグナルパスウェイを検出するような方法が必要であった。

3.3 本研究の目的

細胞間情報伝達が全身性エリテマトーデス（SLE、systemic lupus erythematosus）や関節リウマチなどの様々な自己免疫疾患に関与しているという報告 [67] や最近の研究例として、全身性エリテマトーデスと関節リウマチの細胞間情報伝達解析の結果から、この2つの疾患は、免疫的に異なる病態であることが示唆されている [67]。また自己免疫疾患には、T 細胞が関わるということが報告されている [70]。先行研究においての課題は疾患の病変部と非病変部という特定の状態での細胞間の情報伝達を解析するにとどまっていることである。すなわち、アトピー性皮膚炎患者の病変部と非病変部の比較を通じて、細胞間の情報伝達の解析を行っているにとどまっており、健常と疾患との比較を行っていない。特に全身に影響している免疫疾患や自己免疫疾患に関しては、免疫の中心的な役割を担っているとされる T 細胞に注目した健常と疾患との比較が詳細な病態の理解に必要であると考えられる。また、第2章で行われた研究では、疾患のある細胞種に特に発現している遺伝子を検出することを目的としていた。そのため、その遺伝子が細胞間においてどのように相互作用し、情報伝達を行っているのかについては、解析できていなかった。このため、健常と疾患との異なる細胞がどのように細胞間の情報伝達を行い相互作用し、活動を調整するのかを解析することが必要であると考えられた。さらに、疾患における病変部の比較だけでは得られない詳細な疾患の状態の理解の手がかりになる可能性もある。このためには、例えば健常と疾患との比較により、細胞間情報伝達の疾患におけるパスウェイの情報伝達が健常と違うことを示せるとその詳細な理解に必要である。さらに、ある遺伝子の単独の発現情報だけだとそれが疾患と本当に関与しているのかがわかりにくいいため、複数の遺伝子の塊であるシグナルパスウェイとして疾患における健常との状態変化を検出することにより、T細胞が関与する自己免疫疾患の発症や進行の現象に関与するようなシグナルパスウェイを分子レベルで明らかにしT細胞間の情報伝達をシングルセル RNA-seq を利用することにより、細胞種間の情報伝達の活性を解析する必要がある。以上、病態をより詳細に理解することにより、診断や治療への応用が期待されるため、我々は、健常と疾患との異なる細胞における細胞間情報伝達解析手法の開発を目指した。

本章では疾患における T 細胞を含むすべての免疫細胞の間の情報伝達が健常と疾患で異なるという仮説に基づいて、違いについて解析することにより、細胞間情報伝達に関する免疫疾患との関連を解析することを目的とした。

3.4 本研究の解析手法

3.4.1 細胞間情報伝達解析の概要

本研究では解析のデータセットとして第2章と同様に Garman らの先行研究 [26] で使用された GSE132338 を用いる。細胞間情報伝達解析では細胞種ごとに特に発現している遺伝子を検出し、あらかじめデータベースに登録されている情報から、その遺伝子と相互作用する遺伝子を抽出する。細胞種の決定に関しては、クラスタリングについては Garman らの先行研究の方法と同様に行いマーカーベースで行う。次に関連する遺伝子に関しても発現解析を行う。これらの情報と確率モデルからそれぞれの細胞種間の情報伝達確率を見積もる方法である。

3.4.2 細胞間情報伝達の解析手法

本研究では、細胞間情報伝達の内、リガンドと受容体間の情報伝達に限定して解析する。また免疫疾患における T 細胞と各種の免疫細胞との細胞間シグナル伝達を理解するために、(1) T 細胞とそれ以外の細胞種の細胞の間の情報伝達において情報の送り手と受け手に T 細胞を含む場合を解析対象とする、(2) 健常状態と疾患でのシグナル伝達の違いを検出するという手法を提案する。一般的に免疫細胞には T 細胞以外にも単球、樹状細胞、NK 細胞、血小板といった多くの種類のサブタイプが含まれる。T 細胞とそれらの細胞は情報伝達を行い、何らかの機能を発揮していると考えられている。疾患における細胞種の役割というのも明らかにされつつあり、それらの細胞間の情報伝達の重要さがより増してきている [71]。以下にその概略図を示す (図 3.1)

細胞間の情報伝達の解析では以下の 4 ステップから構成される。

この解析では差分解析のため、健常と患者を各々独立に解析する。

ステップ 1 細胞種のクラスタリング

ステップ 2 健常と疾患との異なる細胞状態における細胞種ごとの発現差解析

ステップ 3 情報の送り手と受け手に T 細胞を含む場合に関して、それらすべての免疫細胞サブタイプ群ごとに対して細胞間情報伝達確率の差分解析を行う

ステップ 4 情報伝達確率解析で見出されたシグナルパスウェイに着目し、発現差解析により有意に情報伝達が増えた遺伝子を特定する

以下についてそれぞれのステップの詳細について説明する。

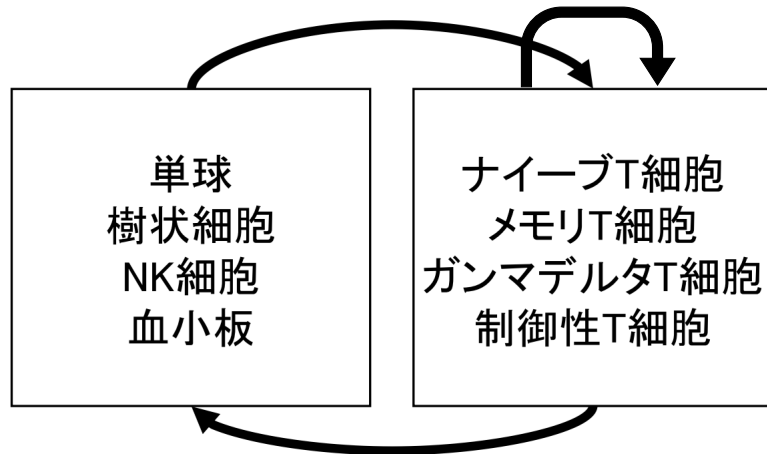


図 3.1 免疫細胞と T 細胞の細胞間のシグナル伝達の概要図

(ステップ 1) 細胞のクラスタリングに関しては、Garman らの先行研究 [26] と同様に行う。すなわち、細胞のクラスタリングと細胞種の決定に関して、始めに健常または疾患の細胞を対数正規化する。スケーリングしたデータを用いて、graph-based clustering approach を用いてクラスタリングを行う。まず上位 70 主成分の PCA 空間でユークリッド距離に基づいて K 最近傍グラフを作成し、任意の 2 細胞について、2 細胞間の全エッジの重なり（Jaccard 類似度）に基づいて 2 細胞間のエッジの重みを調整する。次に Louvain アルゴリズムで細胞のクラスターを特定する。各クラスターについて、各クラスターを特定する上位のアップレギュレーション遺伝子を同定する。細胞種は、これらのクラスター特異的に発現している標準的な細胞種マーカーとして知られている特異的な遺伝子を指標として決定する（表 2.1）。

(ステップ 2) の細胞間情報伝達の推定については以下のように行う。まず、発現量の異なる情報伝達遺伝子を同定する。遺伝子の差分発現は、正規化、スケール化されたデータを使用する。細胞種のサブタイプごとに遺伝子発現を計算して、サブタイプ間の差分を計算する。この際に Wilcoxon 検定を行い、 p 値として 0.1 以上、発現差の対数比の閾値が 0.1 の基準で行う。この基準で差次的に発現したシグナル伝達遺伝子を同定する。（ここまでの先行研究の a に相当する）(ステップ 3) 情報の送り手と受け手に T 細胞を含む場合に関して、それらすべての免疫細胞サブタイプ群ごとに対して細胞間情報伝達確率の差分解析については以下のように行う。T 細胞に着目したのは免疫疾患では、T 細胞の関与

を示唆する報告があるためである。このため、細胞間情報伝達の解析対象は、5種類の T 細胞が細胞間の情報伝達に関わる場合に限定する。CD4 メモリー T 細胞、CD4 ナイーブ T 細胞、初期エフェクター T 細胞、ガンマデルタ T 細胞、制御性 T 細胞の 5 種類の T 細胞に限定し、すべての細胞種ではなく、この 5 種類の T 細胞が関与する細胞間情報伝達を解析する。細胞種グループ内の細胞数が 10 以下の場合、その遺伝子は細胞間情報伝達解析の解析から除外する。健常と疾患の T 細胞をそれぞれ送り手と受け手あるいはその逆とし、すべての細胞について細胞間情報伝達解析を行い、送り手と受け手の組み合わせにおいての健常と疾患との情報伝達確率の差分を計算する。解析の結果、細胞間情報伝達の確率が 5% 以下の場合、細胞間相互作用も除外する。情報伝達確率が上昇するシグナルを、 p 値 <0.01 の統計指標で解析する。(ここまでが先行研究の b,c,d に相当する)

(ステップ 4) 情報伝達確率解析で見出されたシグナルパスウェイに着目し、発現差解析により有意に情報伝達が増えた遺伝子の特定は以下のように行った。有意に検出したシグナルパスウェイについて、発現差解析を行う。この解析では、クラスター内で 10% 以上発現している細胞、および、 p 値 <0.01 とする。Jin らの先行研究にないこのステップを行うことで、細胞間の情報伝達確率としてではなく、細胞間の発現差としての解析を行うことで、疾患において有意に発現が増えた遺伝子を特定する。

3.4.3 シグナルパスウェイにおけるリガンド-受容体の寄与の解析手法

各シグナルパスウェイにおいて、どの遺伝子の組み合わせが支配的な要因であるかを解析した。入力としてはシグナルパスウェイ名と、細胞間情報伝達解析の結果を用いる。細胞間情報伝達解析の結果には、各細胞種間の伝達確率が含まれている。4 つのシグナルパスウェイのそれぞれについて、含まれているリガンド-受容体経路を使って、各々の情報伝達確率を計算した。各シグナルパスウェイにおいて含まれているリガンド-受容体の伝達確率の総和から、各リガンド-受容体経路の相対的な情報伝達確率の大きさの割合を計算し寄与とする。これらの結果を基に、各シグナルパスウェイにおいての寄与が最も大きいリガンド-受容体のペアを特定する出力としては、各シグナルパスウェイを構成するリガンド-受容体について相対的な寄与が出力される。

本手法は、以下の言語と計算環境で開発し実行した。言語 R 4.2.1 でパッケージとして Seurat 4.2.0, CellChat 1.5.0, 計算環境 CPU Intel Xeon Gold 6138, メモリ 1TB, OS Ubuntu 20.04 を使用した。

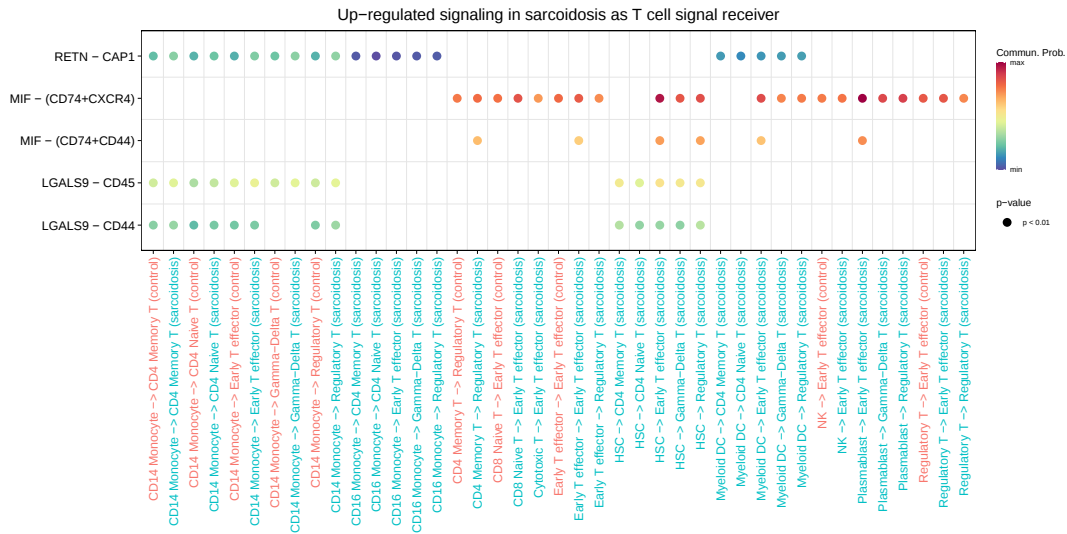


図 3.2 T 細胞と T 細胞を含む免疫細胞との間の細胞間情報伝達、T 細胞を受け手とした細胞間情報伝達確率の差によって見出される遺伝子発現の差のあるシグナルパスウェイ

3.5 細胞間情報伝達解析の結果

図 2.3 は、解析に用いた 18 個の大規模クラスターを UMAP で埋め込んだものである。

図 3.2 は、差分解析により疾患で有意に増加するパスウェイを決定した結果である。

その結果、T 細胞が受け手の場合において 5 つのシグナルパスウェイから 3 つのシグナル伝達経路を検出した (RESISTIN, MIF、GALECTIN)(図 3.2)。

また、T 細胞が送り手の場合において 10 つのシグナルパスウェイから 7 つのシグナル伝達経路 (TGF β 、MIF、GALECTIN、プロテアーゼ活性化受容体 (PARs)、FLT3、CCL、ANNEXIN の 7 つのシグナル伝達経路を検出した (図 3.3)。

バブルプロットの色は、情報伝達確率に比例している。情報伝達確率に比例し、濃い色と赤い色がそれぞれ最小値と最大値に対応している。横軸のラベルは、送信側と受信側を表す細胞種のペアを示している。健常対照者 (赤) と疾患 (水色) に由来するサンプルをラベルの色で示す。

3.5.1 シグナルパスウェイにおけるリガンドー受容体の寄与の解析結果

送り手と受け手のどちらかの T 細胞にのみ存在し、いくつかの遺伝子の組み合わせが各シグナルパスウェイの支配因子として、いくつかの遺伝子の組み合わせが存在する。

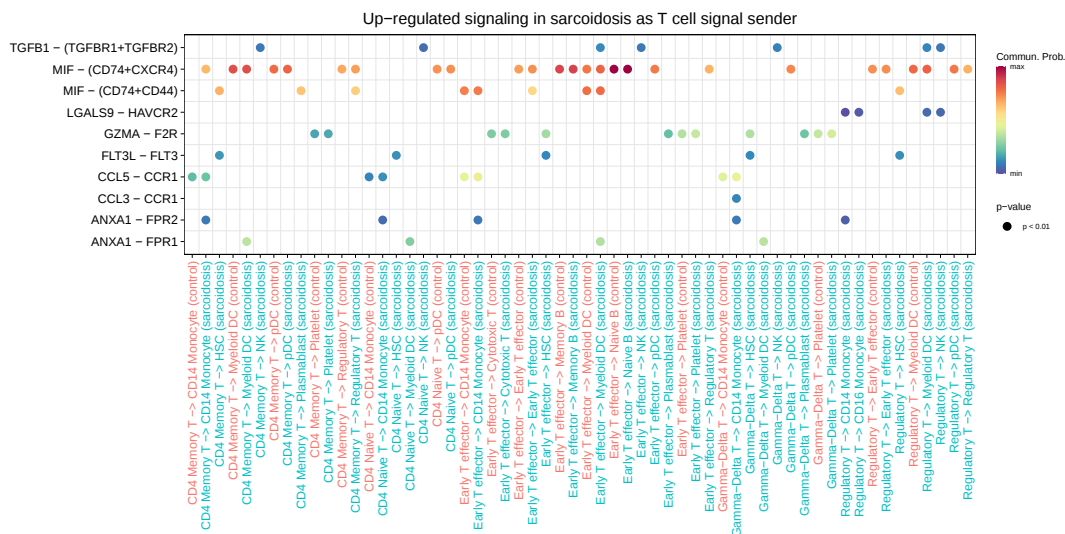


図 3.3 T 細胞と T 細胞を含む免疫細胞との間の細胞間情報伝達、T 細胞を送り手とした細胞間情報伝達確率の差によって見出される遺伝子発現の差のあるシグナルパスウェイ

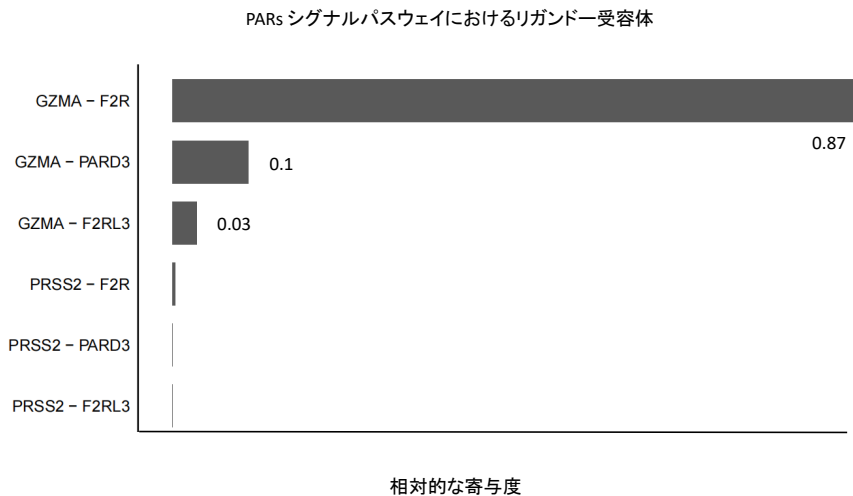


図 3.4 PARs シグナルパスウェイの各リガンド・レセプター対の相対的な寄与度

PARs シグナルパスウェイには 6 つのリガンド-レセプター経路が存在する。これらのうち、GZMA-F2R リガンド-レセプターのペアが最も高い寄与を示した（図 3.4 参照）。RESISTIN シグナルパスウェイは 2 つのリガンド-受容体ペアを含んでいた。これらのうち、RETN-CAP1(cyclase-associated protein 1) のペアが最も貢献度が高かった（図 3.5 上参照）。ANNEXIN シグナルパスウェイは 2 つのリガンド-レセプターのペアを含んで

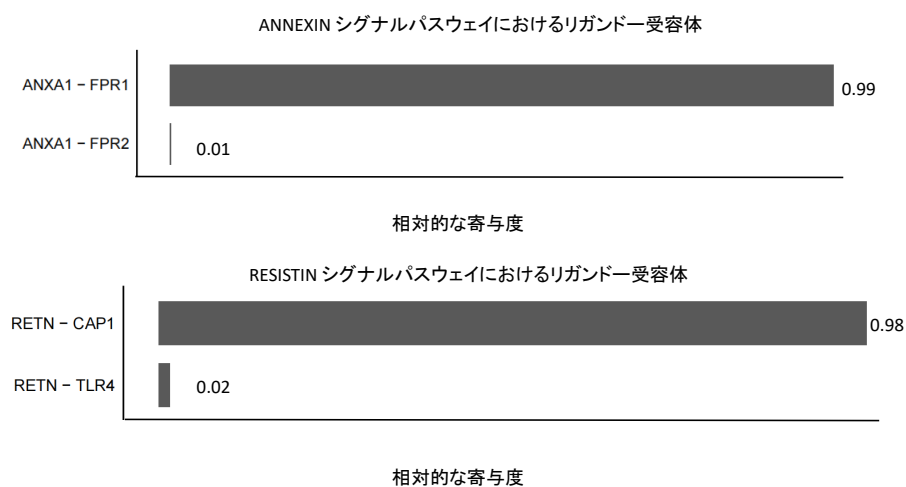


図 3.5 ANNEXIN、RESISTIN シグナルパスウェイの各リガンド・レセプター対の相対的な寄与度

いた。ANNEXIN シグナルパスウェイにおける支配的なものは、以下の通りであった。(図 3.5 下参照)。FLT3 シグナルパスウェイは FLT3L-FLT3 リガンド-レセプターのみをペアを含んでいた。

3.6 考察

今回の解析で見つかった 8 つのシグナルパスウェイのうち、2 つ (MIF と GALECTIN) が T 細胞が情報伝達の受け手あるいは送り手の場合によらず両方で検出された。残りの 6 つのシグナルパスウェイ (RESISTIN、 $TGF\beta$ 、PARs、FLT3、CCL、ANNEXIN) は、T 細胞が情報伝達の受け手、あるいは送り手のどちらか一方のみで検出された。

CCL5、GZMA、ANXA1 リガンドは、第 2 章の研究において、T 細胞からそのサブタイプである制御性 T 細胞への系譜解析においても見出された免疫に関連する遺伝子であった [66]。T 細胞が送り手あるいは受け手となる場合の解析で見出されたシグナルパスウェイについては、多くの報告があり、MIF、 $TGF\beta$ 、CCL がサルコイドーシスに関連していることが示されている [24, 23, 20, 72, 73]。例えば、その一つである MIF シグナルパスウェイは、サルコイドーシス患者の血清 MIF 活性を調べたところ、サルコイドーシス患者 10 例中 9 例で血清 MIF 活性が陽性であったことが知られている。そして MIF 発現の増加が、いくつかの肉芽腫の発症や重症化と関連していることが報告されている [73]。このためサルコイドーシスの遺伝子マーカーとしての利用が期待された。しかしな

がら、最近の報告によると、MIF の増加は、いくつかの肉芽腫性自己免疫疾患の発症とサルコイドーシスではこの事実は当てはまらないとする報告もある [23]。

TGF β に関しては、サルコイドーシス患者の血清において、そのレベルが上昇していることが報告されている [24]。

CC ケモカインの一つである CCL5 は、肺サルコイドーシスと関連している。CCL5 は、免疫反応における肺サルコイドーシスと関連しており、CCR1 を発現する単核球を動員する重要なメディエーターである。CCL5 のレセプターである CCR5 は、肺サルコイドーシスのすべてのステージにおいて、CCR1、CCR2、CCR3 を発現する単核球と マスト細胞を動員する重要なメディエーターである。[20]。また、サルコイドーシス患者では、CD4+ 細胞に比べ、CD8+ 細胞で CCL3 の発現が有意に上昇していることが報告されている [72]。第 2 章の研究においてもサルコイドーシスのガンマデルタ T 細胞種では、健常な細胞の対照と比較して CCL5 の発現量が増加していた。この結果は、これまでの報告と一致しており、本法の解析法による妥当性が示されたと考えられる。

GZMA はサルコイドーシスとの関連は知られていないが、第 2 章の研究における解析でも見出した遺伝子の一つであり、興味深い [66]。GZMA の受容体である F2R (Factor 2 receptor) は第 2 因子受容体のことであり、そのプロモーターに一塩基多型 (SNP, single nucleotide polymorphism) が存在することが示されている。この遺伝子プロモーターの一塩基多型は、高いプロモーター活性を与えることが示されておる。この高い発現レベルは慢性閉塞性肺疾患の増悪に対して保護的である [25]。

RETN-CAP1 は、脂肪組織が CAP1 への作用を通じて炎症促進作用を持つレジスチンおよびアデニル酸をさらに分泌する肥満と関連していることが示されており、閉経後女性乳がん患者の血清レジスチンレベルは、閉経後対照群と比較して高かった。組織における CAP1 の発現は、転移の有無によって違いがみられた [74]。この遺伝子に関してサルコイドーシスとの関連を示す報告はない。

ANXA1 は、内因性の抗炎症メディエーターとして知られており、線維性疾患におけるアネキシン A1 を標的とした治療の可能性を持っている [75]。アネキシン A1 は様々な病理学的プロセスにおいて重要な役割を担っている [75]。また、自然免疫系と適応免疫系の極めて重要なレギュレーターとしても知られている [76]。一方、FPR1 (Formyl Peptide Receptor 1、ホルミルペプチド受容体) リガンドは、肺上皮細胞において創傷閉鎖を促進する。すなわち、肺上皮のホルミルペプチド受容体が肺上皮の修復に関与している可能性が報告されている [77]。しかし、これらのシグナルパスウェイがサルコイドーシスと関連

しているとの報告はない。

このように細胞間情報伝達解析による発現差異解析は従来には知られていなかった炎症に関与するシグナルパスウェイが免疫疾患の一つであるサルコイドーシスにも関わっている可能性を示唆していると考えられる。

最後に、FLT4L (Vascular endothelial growth factor) についての疾患との関連について考察する。樹状細胞 (Dendritic Cell) がサルコイドーシスの重要なメディエーターである可能性が報告されている [78]。この上で、FLT4L は樹状細胞の増殖に関与していることが示されている [79]。したがって、サルコイドーシス患者では FLT4L の発現が上昇することで、樹状細胞が増殖することで疾患の発症に関与している可能性がある。

また、これまでに疾患に関連することが知られているシグナルパスウェイとしては、NOD2,NLRP3,JAK-STAT,Wnt が報告されている [80, 81, 82, 83, 84, 85, 86]。これらの他、MIF (Microphage Microphage Inhibition Factor) シグナルパスウェイに関しては疾患の制御に関与しているか関与していないかについては今だ確定していない [87, 23]。

また、 $TGF\beta$ と galectin シグナルパスウェイについても、疾患で活性が上昇していることが知られている [88, 24]。また免疫系の疾患関与することが知られているケモカインシグナルパスウェイもこの疾患との関連が報告されている [89]。

Garman らの先行研究におけるパスウェイ解析において CD4 ナイーブ T 細胞、初期エフェクター T 細胞、および制御性 T 細胞について、疾患での活性が上昇している結果として検出されているシグナルパスウェイの内、 $TGF\beta$ シグナルパスウェイは T 細胞で有意に活性が上昇していた。このパスウェイ解析では、既知の生物学的機能を持つ文献やデータベースから手動でキュレートした約 500 万件の実験結果にあるパスウェイを利用している。また、発現遺伝子とパスウェイ内の遺伝子との重複の有意性は、右側フィッシャーの検定によって計算された p 値として表される。パスウェイとオーバーラップする遺伝子の発現差の大きさ、予測したものに基つきパスウェイの活性が示される。Z score が 1 以上または -1 以下、有意な p 値 0.05 以下、発現差遺伝子とパスウェイの遺伝子の重複が 5% 以上の場合にパスウェイが濃縮されていると定義されている [90]。先行研究において報告されているシグナルパスウェイの内、 $TGF\beta$ については、T 細胞に関連する細胞間の情報伝達解析においても検出された。

本研究では、Garman らの先行研究 [26] やその他の先行研究 [80, 81, 82, 83, 84, 85, 86] において知られていない、疾患との関連が示唆されるシグナルパスウェイとして 4 つのシグナルパスウェイ (RESISTIN,ANNEXIN,FLT3L,PARs) が検出された。本研究で解析

した結果が正当かどうかについては、実証を行えない時点ではできないが、第 2 章の細胞系譜と第 3 章の細胞間の情報伝達解析という別の視点から解析した際に共通している遺伝子というのは疾患との関連が高い可能性がある。今後、これらのシグナルパスウェイと疾患との関連が明らかにされることを期待している。

次に、検出されたシグナルパスウェイについて定性的に評価を行うために、第 2 章の細胞系譜解析、クラスター解析、Garman らの先行研究と本章の細胞間情報伝達解析の結果を比較した。その結果を示した (図 3.6)。この結果では、細胞系譜解析、細胞間情報伝達、クラスター解析の手法において共通する遺伝子を 3 つ検出した。これらは *GZMA*, *ANXA1*, *CCL5* であった。このうち、*CCL5* については疾患との関連が報告されている。一方、*GZMA*, *ANXA1* については疾患との関連が報告されていない。また、*CCL3* については、細胞系譜解析、細胞間情報伝達の両者の手法においてのみ検出された。*CCL3* については炎症を引き起こすことが知られている蛋白質であり、疾患と関連することが知られている [91]。

2 章の細胞系譜解析、3 章の細胞間情報伝達解析に共通に検出された新規の *ANXA1*, *CCL3* 遺伝子については、疾患においてこれらの遺伝子が活性化している可能性があることから、遺伝子あるいは蛋白質に対する阻害を行うことにより、疾患が緩和するなどの実験的な検証が必要であると考えられる。

3 章の細胞間情報伝達解析についてどのようなデータセットに適用可能かということに関しては、様々な疾患と健常のあるシングル RNA-seq データにおいて適用可能だと考えられる。例えばいくつもの遺伝子が創薬ターゲット候補として挙がっていて絞り込まなければならない場合、細胞間の情報伝達解析により病変部と非病変部の比較、本解析の疾患と健常あるいは疾患の病態の伴う変化（発症時、増悪期、緩和期、終末期）との比較など複数の条件においての比較を行うことにより、その際に重要なシグナルパスウェイ及びリガンド受容体を検出できると考えられる。また、細胞系譜の解析と組み合わせて系譜の分岐点のような生物学的に重要な曲面においてどのような細胞間の情報伝達が行われているのかといった解析にも有効であると考えられる。

3.7 結言

本研究では、健常と疾患の異なる細胞集団の比較により、細胞間情報伝達解析による疾患に関連する遺伝子マーカーを検出するための新しい方法を提案した。

第 2 章の研究 [66] では、T 細胞の分化に対する細胞系譜解析を行い、細胞系譜間及び疾

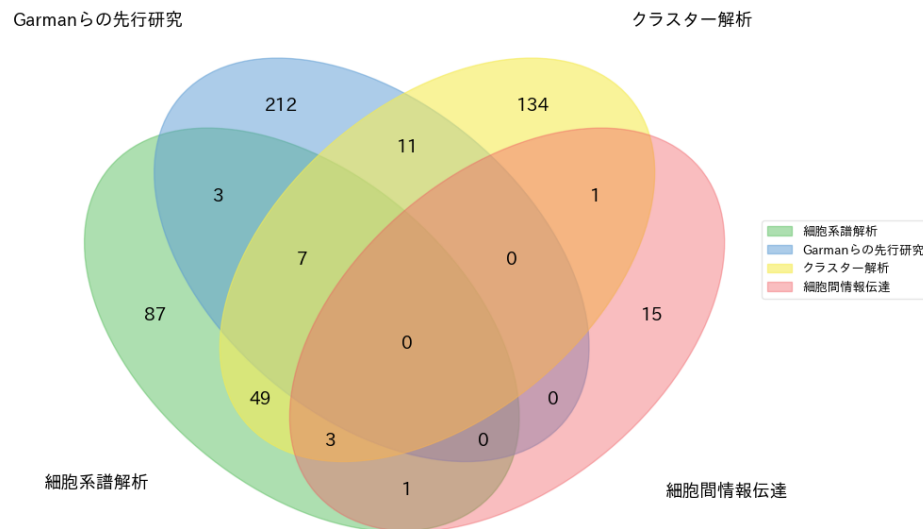


図 3.6 解析手法による Garman らの先行研究と検出遺伝子の比較

患と健常の遺伝子発現の差分解析を行った。しかしながらこの研究においては細胞間の相互作用という観点の解析はできていなかった。

本研究において、細胞間の情報伝達解析では細胞間で情報伝達確率が高いシグナルパスウェイと検出された遺伝子の細胞種間の差分発現解析から、*CCL5*、*GZMA*、*ANXA1* など、疾患の T 細胞において関連する細胞間情報伝達のリガンドとなる遺伝子を見出した。細胞間の情報伝達のリガンドとなる *CCL5*、*GZMA*、*ANXA1* などの遺伝子については、2 章の細胞系譜解析の研究 [66] でも報告されている。この 3 つの遺伝子は、T 細胞の細胞系譜解析において検出されたその他の遺伝子と比較して、T 細胞と他の細胞との細胞間情報伝達の点においてその他の遺伝子よりもより活性化されていると考えられる。したがって、これらのリガンドに対する受容体との特定の蛋白質間相互作用を模倣するような薬剤を創ることにより疾患の制御ができるかもしれない。また、T 細胞が情報の送り手となる場合のシグナルパスウェイを明らかにしたことで、数ある免疫細胞の中で T 細胞で特に作用する可能性のあるターゲット遺伝子を検討する手法としても有用である可能性がある。免疫細胞には T 細胞以外の細胞種もあり、疾患に関連する報告も多数あることから、T 細胞以外の細胞種に対する細胞間情報伝達解析による、病態のより詳細な理解が必要であると考えられる。

本研究で提案した手法が従来の遺伝子を検出していながらも、別のこれまでに知られていない新規の遺伝子を検出したことは、有意な結果であると考えられる。疾患がどのよ

うにこれらの遺伝子と関連があるかについては、実証実験の検討が必要であると考えられる。

第 4 章 結論

4.1 本研究のまとめ

本研究の第 2 章ではシングルセル RNA-seq のデータを使用して、健常と疾患で異なる細胞集団について、免疫細胞の一つである T 細胞に着目して、そのサブタイプへの分化を細胞系譜で表現し、系譜間の差分解析により、疾患に関連する遺伝子を検出する手法を提案した。Garman[26] らの先行研究においての細胞種間の比較による手法では検出されなかった遺伝子が健常と疾患との異なる細胞においての系譜の比較において検出された。上位の 10 を含むそれ以下の遺伝子について、Garman[26] らの先行研究結果、第 2 章の提案手法、クラスターベースの手法により検出された遺伝子の比較を行った。Garman らの先行研究により検出された遺伝子の内、我々の提案した手法の方がクラスターベースの検出よりも重複した遺伝子が少なかった。細胞系譜の解析では分化の開始と終了地点で比較を行うため、クラスターベースの細胞種の平均よりもより分化状態に影響して発現する遺伝子を検出しやすいと考えられる。この結果、手法として共通のクラスターベースの解析の比較に比べて細胞系譜と Garman らの先行研究との重複が少なかったと考えられる。本研究により検出した遺伝子が疾患との関連を示す実験はないのだが、疾患が健常と比較し免疫細胞の異常により起こっているとすると、免疫細胞の異常を検出しやすい細胞系譜の手法は有効であると考えられる。これまでの手法では検出されなかった 4 つの遺伝子が検出された。これら 4 つの遺伝子について健常と疾患との細胞系譜の発現差が統計的に大きかった。また、これらの遺伝子は直接的にサルコイドーシスとの関連性を示す報告はなかったが免疫関連の遺伝子であった。細胞系譜においての疑似時間と分化状態との比較により、ナイーブ T 細胞からそれぞれのサブタイプであるガンマデルタ T 細胞や制御性 T 細胞への分化の過程での第 2 章の細胞系譜で検出された提案手法でしか検出できていない 88 遺伝子のうちの 4 遺伝子の遺伝子発現量を健常と疾患で比較すると免疫疾患においての特徴が分かり、それぞれの遺伝子の発現状態と分化状態との関連性が示唆された。

Garman らの先行研究により検出された遺伝子の内、我々の提案した手法の方がクラスターベースの検出よりも重複した遺伝子が少なかったことに関して、その理由を考察する。細胞系譜の解析では分化の開始と終了地点で比較を行うため、クラスターベースの細胞種の平均よりもより分化状態に影響して発現する遺伝子を検出しやすいと考えられる。この結果、手法として共通のクラスターベースの解析の比較に比べて細胞系譜と Garman

らの先行研究との重複が少なかったと考えられる。新規に検出した遺伝子についてはいずれも疾患との直接的な関係の報告はないものの、免疫関連の遺伝子であった。細胞系譜の解析では、系譜に沿って疾患と健常との発現差が観察されており、系譜における遺伝子発現と疾患との関連が示唆された。疾患と健常を細胞分化の過程を考慮し比較することできる細胞系譜の手法は有効であると考えられる。

また、第3章では健常と疾患で異なる細胞集団について、T細胞とT細胞を含む他の免疫疾患細胞との関わりを細胞間の情報伝達解析の差分解析により、シグナルパスウェイ及びそのリガンドー受容体を検出する新たな手法として提案した。

提案した手法について先行研究やその他の解析法と比較し、その解析法の妥当性を評価した。その結果、先行研究において知られていた疾患に関与する遺伝子やシグナルパスウェイ(CCL,TGF β 1,MIF)との重複があった。また、複数の先行研究において知られていない疾患との関連が示唆される新規のシグナルパスウェイとして4つのシグナルパスウェイ(RESISTIN,ANNEXIN,FLT3L,PARs)を検出することができた。これらの遺伝子が免疫に関与していることから、解析法が妥当であることが示唆された。この中でも特にANNEXINシグナルパスウェイについては第2章の細胞系譜解析による結果でもそのリガンドであるANXA1として検出されている。このことから、複数の解析手法により数ある遺伝子の中から疾患のマーカー遺伝子を絞り込むことが有効である可能性を示唆された。今後、これらのシグナルパスウェイと疾患との関連が明らかにされることを期待している。

第3章の細胞間情報伝達解析の手法で発見された遺伝子と、第2章の手法である細胞系譜間差分解析の手法で発見された遺伝子是一部重なる部分がある。

細胞系譜解析と細胞間の情報伝達解析には以下のような関連性があると考えられる。

検出されやすい遺伝子の特徴に関しては、細胞系譜解析においては、健常と疾患を比較した際の各々の系譜の最終分化状態である細胞種において特に疾患で発現している遺伝子を検出しやすいと考えられる。一方、細胞間情報伝達解析においては、任意の2種類の細胞種において、どちらか一方をT細胞で固定しリガンドー受容体の関係を基に、発現する遺伝子の平均量の積でその大きさを見積もる。このためこの方法は、遺伝子に関与するシグナルパスウェイの遺伝子群としての活性が大きい遺伝子を検出しやすいと考えられる。したがって前者では分化が進んだ細胞群において特に発現している遺伝子を検出しやすく、後者では細胞種間で関係する遺伝子がどれも発現していると検出しやすくなると考えられる。

いいかえると、細胞分化の系譜において健常と比べて疾患で発現差が大きいとして検出された遺伝子と細胞情報伝達において健常と比べて疾患で情報伝達確率が高いとして検出された遺伝子が一致すると、健常と疾患を比較した際の各々の系譜の最終段階と最初の状態である細胞種において特に疾患で発現している遺伝子かつ遺伝子が関与するあるクラスターでのシグナルパスウェイのリガンドとなる遺伝子と、別のクラスターで受容体となる遺伝子の両方について発現が見られないと検出されない遺伝子であると考えられる。

つまり、細胞系譜解析による検出した遺伝子について、細胞間の情報伝達解析を行うことで、シグナルパスウェイとしての活性化を検出することができると考えられる。

また本研究では既知のマーカー遺伝子に基づいて細胞種を定義したが、より細分化して細胞種を定義し、細胞系譜解析と細胞間情報伝達解析を組み合わせる行うことにより、疾患でしか活性化していないシグナルパスウェイ遺伝子ペアについても検出できる可能性があると考えられる。

第2章と第3章で共通して見つかった遺伝子の内、疾患との関連が知られていないが炎症のメディエーターとして知られていた *ANXA1* について、*ANNEXIN* シグナルパスウェイの受容体である *FPR1* を含めた疾患との関連について、最近 *IL-6* と相互作用し、乳がんの微小環境の炎症に関わっているという報告があった [92]。このことから、*ANXA1-FPR1* パスウェイが活性化してしまうと、炎症の持続的な状態を継続させ、がんへ発展させてしまう可能性が考えられている。

直接的にサルコイドーシスとの *ANNEXIN* シグナルパスウェイが関連する報告はいまだないが、複数の解析方法を組み合わせることで、単独の手法のみでは絞り込めない候補遺伝子の中から、より重要な免疫疾患に関連する遺伝子を検出できる可能性がある。

細胞系譜解析のおいての発現遺伝子の変動の結果を観ると健常と疾患との差は疑似時間の最終段階の状態においての差が大きいと考えられる。したがって、疑似時間が本来の時間を反映したものであるかはわからないが、細胞種のパターンが通常の分化よりも極端に分化している状態というのが疾患であるとも考えることもできる。また情報伝達の解析においては、検出した2遺伝子 (*ANXA*, *GZMK*) についてはシグナルパスウェイに疾患特異的な活性がみられたことから、異常に情報伝達の活性化した状態というのが免疫疾患という病態を表していると考えられる。

数多く存在する遺伝子の中で、数個の遺伝子候補にまで標的遺伝子を絞れたことは、細胞系譜解析及び細胞間情報伝達確率による疾患と健常との比較による遺伝子マーカーの検出手法の有用性を示すものであると考えられる。

また、各々の手法にのみ検出されるような遺伝子についても、疾患の個別化あるいは層別化医療に対する新たな治療の開発の手がかりが見つかる可能性もある。

一方で、第2章と第3章で用いたシングルセル RNA-seq データは同じものであるため、データが同じであれば、解析方法が違っていても共通の遺伝子が挙がってくることは当然であるという見方もある。このような偽陽性の遺伝子を除くには、別の疾患に対する解析を行い同様な遺伝子が出てくるかどうかを検討する必要があると考えられる。

今後さらに T 細胞以外の免疫細胞あるいは免疫細胞以外の細胞に関してあるいは違う疾患のデータセットで解析手法を検討し、より検出精度確度の高い解析方法の開発へと発展させることができると考えられる。

4.2 免疫疾患研究における本研究の寄与と今後の展望

第1章でも述べたように免疫疾患は複雑な生命現象であり、その動態は多種多様である。また、免疫には多様な免疫細胞が関わっており、一つの疾患をとってみても様々な病態があることが近年明らかにされつつある。このためシングルセル RNA-seq のような細胞ごとの情報を検出できる技術を利用することにより疾患の複数存在する病態を層別化することにより、より正確な疾患の診断や治療法が提案できると考えられる。このような提案ができると、今後より効果的な薬剤開発やその適用の個別化した医療に寄与できる可能性がある。しかしながら、細胞ごとの遺伝子発現情報を検出できる技術からどのようにに疾患に関連する有意な遺伝子変化を検出するかについてはいまだ決定的な方法はない。本研究においては、先行研究やその機能的な重要性から免疫疾患に特に関連すると考えられる T 細胞に注目してその免疫疾患と健常のシングルセル RNA-seq を利用して、2通りの解析方法を提案し適用することにより、有意な遺伝子を検出できるような方法論を検討した。遺伝子マーカーでは検出する遺伝子が多く検出され、真に疾患に関連する遺伝子を絞り込むことが重要と考えられる。本研究から、細胞系譜解析により検出した遺伝子を細胞間の情報伝達解析におけるシグナルパスウェイの活性化を通して検討することにより、疾患に関連する遺伝子を絞り込む方法としての可能性を提案できた。

すなわち、2章の健常と疾患との T 細胞の分化における細胞系譜解析と3章の T 細胞に関連する細胞間情報伝達解析がそれらに相当する。これらの解析の結果、先行研究で知られている遺伝子やシグナルパスウェイを含んだいくつかの新規の遺伝子やシグナルパスウェイが検出された。両手法において検出される遺伝子が疾患にとって重要であるという考えと、今までに疾患との関連が報告されていない遺伝子といった新規性の観点から

ANXA1 という 1 つの遺伝子が浮かび上がった。それぞれの手法の特徴から ANXA1 は分化が進んだガンマデルタ T 細胞において特に発現していて、その遺伝子を含むシグナルパスウェイが疾患特異的に活性化していると考えられる。サルコイドーシスに対する治療ターゲットとして 2 つの解析法の組み合わせで得られた遺伝子である ANXA1 は細胞間の情報伝達確率差が健常と比較して大きいことからシグナルパスウェイが活性化し、さらに系譜内、間での発現状態が疾患で亢進していることから、またこの遺伝子が多面的な機能をもつという既報告からマーカーとして挙げられると考えられる。このように細胞の分化や細胞間の情報伝達といった、複数の生物プロセスの理解に基づいた手法を組み合わせることで利用することにより、膨大な遺伝子発現情報の中から疾患に関連する遺伝子を絞り込めたことは、今後の免疫疾患の関連遺伝子候補を絞り込む手法において寄与できるものであると考えられる。今後このような手法が様々な疾患に有効であることを期待するとともに、まだ、現状においてはこの遺伝子の疾患との関連が示されていないため、今後の実験的な検証にも期待している。

一方、今後の展望としては、シングルセル RNA-seq に基づく疾患メカニズム解明のためには、より遺伝子情報の詳細な変化を解像度よく検出し、解析していく必要があると考えられる。エピジェネティクス情報を含んだ解析が詳細な遺伝子発現情報に寄与できるものであると考えられる。エピジェネティクスにおいて主な変化として挙げられる DNA のメチル化は遺伝子を活性化したり不活性化すること知られている。エピゲノムの変化は遺伝子発現を始めがんなどの疾患の進行や予後に影響を及ぼしていることが明らかにされつつある。一方、空間的な情報を加味したトランスクリプトーム解析も必要になってくると考えられる。同じ疾患においても、組織などの細胞のサンプリング場所によって異なるトランスクリプトームになることが知られてきている。また同じ細胞でも細胞の核や細胞質などの場所によってもそのトランスクリプトームは異なっている。したがって、疾患とエピゲノム、空間的な情報を組み合わせることによるシングルセル RNA-seq を行うことにより詳細な疾患のメカニズム解明に近づくことができると考えられる。

参考文献

- [1] 足立剛也, 貝沼圭吾, 浅野浩一郎, 天谷雅行, 新井洋由, 石井健, 伊藤浩明, 内尾英一, 海老澤元宏, 岡野光博ほか. 免疫アレルギー疾患研究 10 か年戦略 2030: 「見える化」による安心社会の醸成. *アレルギー*, Vol. 69, No. 1, pp. 23–33, 2020.
- [2] 本多紘二郎. サルコイドーシス. *内科*, Vol. 129, No. 4, pp. 543–546, 2022.
- [3] D. G. Pellicci, H.-F. Koay, and S. P. Berzins. Thymic development of unconventional T cells: how NKT cells, MAIT cells and $\gamma\delta$ T cells emerge. *Nature Reviews Immunology*, Vol. 20, No. 12, pp. 756–770, 2020.
- [4] A. O’Garra, L. Steinman, and K. Gijbels. CD4+ T-cell subsets in autoimmunity. *Current Opinion in Immunology*, Vol. 9, No. 6, pp. 872–883, 1997.
- [5] I. Raphael, R. R. Joern, and T. G. Forsthuber. Memory CD4+ T cells in immunity and autoimmune diseases. *Cells*, Vol. 9, No. 3, p. 531, 2020.
- [6] S. Park, N. L. Anderson, D. A. Canaria, and M. R. Olson. Granzyme-producing CD4 T cells in cancer and autoimmune disease. *ImmunoHorizons*, Vol. 5, No. 12, pp. 909–917, 2021.
- [7] R. Qiu, L. Zhou, Y. Ma, L. Zhou, T. Liang, L. Shi, J. Long, and D. Yuan. Regulatory T cell plasticity and stability and autoimmune diseases. *Clinical Reviews in Allergy & Immunology*, Vol. 58, No. 1, pp. 52–70, 2020.
- [8] J. W. Leavenworth, L. Z. Shi, X. Wang, and H. Wei. Immune cell lineage reprogramming in cancer. *Frontiers in Immunology*, Vol. 12, , 2021.
- [9] D. McCluskey, N. Benzian-Olsson, S. K. Mahil, N. K. Hassi, C. T. Wohnhaas, D. Baudry, V. Cornelius, H. Lachmann, H. McAteer, F. Meynell, et al. Single-cell analysis implicates Th17 to Th2 cell plasticity in the pathogenesis of palmoplantar pustulosis. *Journal of Allergy and Clinical Immunology*, Vol. 150, No. 4, pp. 882–893, 2022.
- [10] T. Poch, J. Krause, C. Casar, T. Liwinski, L. Glau, M. Kaufmann, A. E. Ahrenstorf, L. U. Hess, A. E. Ziegler, G. Martrus, et al. Single-cell atlas of hepatic T cells reveals expansion of liver-resident naive-like CD4+ T cells in primary sclerosing cholangitis. *Journal of Hepatology*, Vol. 75, No. 2, pp. 414–423, 2021.
- [11] Z. Wang, Z. Li, K. Zhou, C. Wang, L. Jiang, L. Zhang, Y. Yang, W. Luo, W. Qiao,

- G. Wang, et al. Deciphering cell lineage specification of human lung adenocarcinoma with single-cell RNA sequencing. *Nature Communications*, Vol. 12, No. 1, pp. 1–15, 2021.
- [12] K. Ghoreschi, A. Laurence, John JO’ Shea. Janus kinases in immune cell signaling. *Immunological Reviews*, Vol. 228, No. 1, pp. 273–287, 2009.
- [13] G. C. Tsokos and N. R. Rose. Immune cell signaling in autoimmune diseases. *Clinical Immunology*, Vol. 181, pp. 1–8, 2017.
- [14] A. Cossarizza, H.-D. Chang, A. Radbruch, A. Acs, D. Adam, S. Adam-Klages, W. W. Agace, N. Aghaeepour, M. Akdis, M. Allez, et al. Guidelines for the use of flow cytometry and cell sorting in immunological studies. *European Journal of Immunology*, Vol. 49, No. 10, pp. 1457–1973, 2019.
- [15] 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.
- [16] G. Chen, B. Ning, and T. Shi. Single-cell RNA-seq technologies and related computational data analysis. *Frontiers in Genetics*, p. 317, 2019.
- [17] R. Lepzien, M. Nie, P. Czarnewski, S. Liu, M. Yu, A. Ravindran, S. Kullberg, A. Eklund, J. Grunewald, and A. Smed-Sörensen. Pulmonary and blood dendritic cells from sarcoidosis patients more potently induce IFN γ -producing Th1 cells compared with monocytes. *Journal of Leukocyte Biology*, Vol. 111, No. 4, pp. 857–866, 2022.
- [18] D. R. Moller, B. A. Rybicki, N. Y. Hamzeh, C. G. Montgomery, E. S. Chen, W. Drake, and A. P. Fontenot. Genetic, immunologic, and environmental basis of sarcoidosis. *Annals of the American Thoracic Society*, Vol. 14, No. Supplement 6, pp. S429–S436, 2017.
- [19] G. A. Bello, I. Adrianto, G. G. Dumancas, A. M. Levin, M. C. Iannuzzi, B. A. Rybicki, and C. Montgomery. Role of NOD2 pathway genes in sarcoidosis cases with clinical characteristics of blau syndrome. *American Journal of Respiratory and Critical Care Medicine*, Vol. 192, No. 9, pp. 1133–1135, 2015.
- [20] V. Palchevskiy, N. Hashemi, S. S. Weigt, Y. Y. Xue, A. Derhovannessian, M. P. Keane, R. M. Strieter, M. C. Fishbein, J. C. Deng, J. P. Lynch, et al. Im-

- mune response CC chemokines CCL2 and CCL5 are associated with pulmonary sarcoidosis. *Fibrogenesis & Tissue Repair*, Vol. 4, No. 1, pp. 1–12, 2011.
- [21] C. I. Bloom, C. M. Graham, M. P. Berry, F. Rozakeas, P. S. Redford, Y. Wang, Z. Xu, K. A. Wilkinson, R. J. Wilkinson, Y. Kendrick, et al. Transcriptional blood signatures distinguish pulmonary tuberculosis, pulmonary sarcoidosis, pneumonias and lung cancers. *PloS One*, Vol. 8, No. 8, p. e70630, 2013.
 - [22] T. Mikuniya, S. Nagai, M. Shigematsu, K. Hamada, Y. Hoshino, T. Mio, and T. Izumi. Significance of increase in lipocortin-1 gene transcripts in the blood monocytes of patients with pulmonary sarcoidosis. *Sarcoidosis, Vasculitis, and Diffuse Lung Diseases: Official Journal of WASOG*, Vol. 18, No. 3, pp. 272–278, 2001.
 - [23] C. D. Odio, E. J. Miller, M. Sauler, L. Leng, M. Piecychna, W. P. Drake, and R. Bucala. Macrophage Migration Inhibitory Factor is not associated with sarcoidosis susceptibility or severity in whites or blacks. *Sarcoidosis, Vasculitis, and Diffuse Lung Diseases*, Vol. 37, No. 3, 2020.
 - [24] S. Kobak, M. Akyildiz, A. Gokduman, T. Atabay, and H. Vural. Serum galectin-3 and TGF-beta levels in patients with sarcoidosis. *Reumatologia Clinica*, Vol. 17, No. 10, pp. 562–565, 2021.
 - [25] M. Platé, P. J. Lawson, M. R. Hill, R. P. Marshall, H. L. Booth, M. Kumari, G. J. Laurent, J. R. Hurst, and R. C. Chambers. Role of a functional polymorphism in the F2R gene promoter in sarcoidosis. *Respirology*, Vol. 20, No. 8, pp. 1285–1287, 2015.
 - [26] L. Garman, R. C. Pelikan, A. Rasmussen, C. A. Lareau, K. A. Savoy, U. S. Deshmukh, H. Bagavant, A. M. Levin, S. Daouk, W. P. Drake, et al. Single cell transcriptomics implicate novel monocyte and T cell immune dysregulation in sarcoidosis. *Frontiers in Immunology*, Vol. 11, p. 567342, 2020.
 - [27] C. Xu and Z. Su. Identification of cell types from single-cell transcriptomes using a novel clustering method. *Bioinformatics*, Vol. 31, No. 12, pp. 1974–1980, 2015.
 - [28] X. Que, F. Checconi, F. Petrini, and J. A. Gunnels. Scalable community detection with the louvain algorithm. In *2015 IEEE International Parallel and Distributed Processing Symposium*, pp. 28–37. IEEE, 2015.

- [29] K. M. Mullen, A. R. Gocke, R. Allie, A. Ntranos, I. V. Grishkan, C. Pardo, and P. A. Calabresi. Expression of CCR7 and CD45RA in CD4+ and CD8+ subsets in cerebrospinal fluid of 134 patients with inflammatory and non-inflammatory neurological diseases. *Journal of Neuroimmunology*, Vol. 249, No. 1-2, pp. 86–92, 2012.
- [30] Y. Elyahu, I. Hekselman, I. Eizenberg-Magar, O. Berner, I. Strominger, M. Schiller, K. Mittal, A. Nemirovsky, E. Eremenko, A. Vital, et al. Aging promotes reorganization of the CD4 T cell landscape toward extreme regulatory and effector phenotypes. *Science Advances*, Vol. 5, No. 8, p. eaaw8330, 2019.
- [31] K. Weatherly, M. Bettonville, D. Torres, A. Kohler, S. Goriely, and M. Y. Braun. Functional profile of S100A4-deficient T cells. *Immunity, Inflammation and Disease*, Vol. 3, No. 4, pp. 431–444, 2015.
- [32] C. M. L. Munier, D. v. Bockel, M. Bailey, S. Ip, Y. Xu, S. Alcantara, S. M. Liu, G. Denyer, W. Kaplan, K. Suzuki, et al. The primary immune response to vaccinia virus vaccination includes cells with a distinct cytotoxic effector cd4 t-cell phenotype. *Vaccine*, Vol. 34, No. 44, pp. 5251–5261, 2016.
- [33] R. C. Duggleby, T. N. Shaw, L. B. Jarvis, G. Kaur, and J. Hill Gaston. CD27 expression discriminates between regulatory and non-regulatory cells after expansion of human peripheral blood CD4+ CD25+ cells. *Immunology*, Vol. 121, No. 1, pp. 129–139, 2007.
- [34] G. Pizzolato, H. Kaminski, M. Tosolini, D.-M. Franchini, F. Pont, F. Martins, C. Valle, D. Labourdette, S. Cadot, A. Quillet-Mary, et al. Single-cell RNA sequencing unveils the shared and the distinct cytotoxic hallmarks of human TCRV δ 1 and TCRV δ 2 $\gamma\delta$ T lymphocytes. *Proceedings of the National Academy of Sciences*, Vol. 116, No. 24, pp. 11906–11915, 2019.
- [35] G.-H. Park, K.-Y. Kim, J. Y. Cheong, S. W. Cho, and K. Kwack. Association of GNLY genetic polymorphisms with chronic liver disease in a korean population. *DNA and Cell Biology*, Vol. 31, No. 9, pp. 1492–1498, 2012.
- [36] M. A. Turman, T. Yabe, C. McSherry, F. H. Bach, and J. P. Houchins. Characterization of a novel gene (NKG7) on human chromosome 19 that is expressed in natural killer cells and T cells. *Human Immunology*, Vol. 36, No. 1, pp. 34–40,

1993.

- [37] P. Sampath, K. Moideen, U. D. Ranganathan, and R. Bethunaickan. Monocyte subsets: phenotypes and function in tuberculosis infection. *Frontiers in Immunology*, Vol. 9, p. 1726, 2018.
- [38] H. L. Ziegler-Heitbrock, B. Passlick, and D. Flieger. The monoclonal antimonocyte antibody my4 stains b lymphocytes and two distinct monocyte subsets in human peripheral blood. *Hybridoma*, Vol. 7, No. 6, pp. 521–527, 1988.
- [39] T. Hruz, O. Laule, G. Szabo, F. Wessendorp, S. Bleuler, L. Oertle, P. Widmayer, W. Gruissem, and P. Zimmermann. Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. *Advances in bioinformatics*, Vol. 2008, , 2008.
- [40] L. Heger, S. Balk, J. J. Lühr, G. F. Heidkamp, C. H. Lehmann, L. Hatscher, A. Purbojo, A. Hartmann, F. Garcia-Martin, S.-I. Nishimura, et al. CLEC10A is a specific marker for human CD1c+ dendritic cells and enhances their toll-like receptor 7/8-induced cytokine secretion. *Frontiers in Immunology*, Vol. 9, p. 744, 2018.
- [41] B. Jahrsdörfer, A. Vollmer, S. E. Blackwell, J. Maier, K. Sontheimer, T. Beyer, B. Mandel, O. Lunov, K. Tron, G. U. Nienhaus, et al. Granzyme B produced by human plasmacytoid dendritic cells suppresses T-cell expansion. *Blood, The Journal of the American Society of Hematology*, Vol. 115, No. 6, pp. 1156–1165, 2010.
- [42] B. J. Masten, G. K. Olson, C. A. Tarleton, C. Rund, M. Schuyler, R. Mehran, T. Archibeque, and M. F. Lipscomb. Characterization of myeloid and plasmacytoid dendritic cells in human lung. *The Journal of Immunology*, Vol. 177, No. 11, pp. 7784–7793, 2006.
- [43] M. Aggarwal, R. Villuendas, G. Gomez, S. M. Rodriguez-Pinilla, M. Sanchez-Beato, D. Alvarez, N. Martinez, A. Rodriguez, M. E. Castillo, F. I. Camacho, et al. TCL1A expression delineates biological and clinical variability in B-cell lymphoma. *Modern Pathology*, Vol. 22, No. 2, pp. 206–215, 2009.
- [44] P. G. Chu and D. A. Arber. CD79: a review. *Applied Immunohistochemistry & Molecular Morphology*, Vol. 9, No. 2, pp. 97–106, 2001.

- [45] M. P. Lambert, R. Meng, D. Harper, L. Xiao, M. S. Marks, and M. Poncz. Megakaryocytes exchange significant levels of their alpha-granular PF4 with their environment. *Blood*, Vol. 124, No. 21, p. 1432, 2014.
- [46] E. Laurenti, S. Doulatov, S. Zandi, I. Plumb, J. Chen, C. April, J.-B. Fan, and J. E. Dick. The transcriptional architecture of early human hematopoiesis identifies multilevel control of lymphoid commitment. *Nature Immunology*, Vol. 14, No. 7, pp. 756–763, 2013.
- [47] L. Van der Maaten and G. Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, Vol. 9, No. 11, 2008.
- [48] Z. Gao, Y. Feng, J. Xu, and J. Liang. T-cell exhaustion in immune-mediated inflammatory diseases: New implications for immunotherapy. *Frontiers in Immunology*, Vol. 13, p. 977394, 2022.
- [49] T. V. Zhao, Y. Sato, J. J. Goronzy, and C. M. Weyand. T-cell aging-associated phenotypes in autoimmune disease. *Frontiers in Aging*, Vol. 3, , 2022.
- [50] W. Yang, T. Yu, and Y. Cong. CD4+ T cell metabolism, gut microbiota, and autoimmune diseases: implication in precision medicine of autoimmune diseases. *Precision Clinical Medicine*, Vol. 5, No. 3, p. pbac018, 2022.
- [51] 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, pp. 1–16, 2018.
- [52] K. Van den 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, pp. 1–13, 2020.
- [53] D. I. Godfrey, A. P. Uldrich, J. McCluskey, J. Rossjohn, and D. Moody. The burgeoning family of unconventional T cells. *Nature Immunology*, Vol. 16, No. 11, pp. 1114–1123, 2015.
- [54] S. R. Carding and P. J. Egan. Gamma delta T cells: functional plasticity and heterogeneity. *Nature Reviews Immunology*, Vol. 2, No. 5, pp. 336–345, 2002.
- [55] E. Cano-Gamez, B. Soskic, T. I. Roumeliotis, E. So, D. J. Smyth, M. Baldrighi, D. Willé, N. Nakic, J. Esparza-Gordillo, C. G. Larminie, et al. Single-cell tran-

- scriptomics identifies an effectorness gradient shaping the response of CD4⁺ T cells to cytokines. *Nature Communications*, Vol. 11, No. 1, pp. 1–15, 2020.
- [56] Y. Tian, M. Babor, J. Lane, V. Schulten, V. S. Patil, G. Seumois, S. L. Rosales, Z. Fu, G. Picarda, J. Burel, et al. Unique phenotypes and clonal expansions of human CD4 effector memory T cells re-expressing CD45RA. *Nature Communications*, Vol. 8, No. 1, pp. 1–13, 2017.
- [57] B. Petersen, R. Kammerer, A. Frenzel, P. Hassel, T. H. Dau, R. Becker, A. Breithaupt, R. G. Ulrich, A. Lucas-Hahn, and G. Meyers. Generation and first characterization of TRDC-knockout pigs lacking $\gamma\delta$ T cells. *Scientific Reports*, Vol. 11, No. 1, pp. 1–13, 2021.
- [58] M.-P. Lefranc. Immunoglobulin and T cell receptor genes: Imgt[®] and the birth and rise of immunoinformatics. *Frontiers in Immunology*, Vol. 5, p. 22, 2014.
- [59] M. M. Nielsen, D. A. Witherden, and W. L. Havran. $\gamma\delta$ T cells in homeostasis and host defence of epithelial barrier tissues. *Nature Reviews Immunology*, Vol. 17, No. 12, pp. 733–745, 2017.
- [60] D. Lu, L. Liu, Y. Sun, J. Song, Q. Yin, G. Zhang, F. Qi, Z. Hu, Z. Yang, Z. Zhou, et al. The phosphatase PAC1 acts as a T cell suppressor and attenuates host antitumor immunity. *Nature Immunology*, Vol. 21, No. 3, pp. 287–297, 2020.
- [61] L. A. Monticelli, G. F. Sonnenberg, M. C. Abt, T. Alenghat, C. G. Ziegler, T. A. Doering, J. M. Angelosanto, B. J. Laidlaw, C. Y. Yang, T. Sathaliyawala, et al. Innate lymphoid cells promote lung-tissue homeostasis after infection with influenza virus. *Nature Immunology*, Vol. 12, No. 11, pp. 1045–1054, 2011.
- [62] A. S. Damazo, A. L. Sampaio, C. M. Nakata, R. J. Flower, M. Perretti, and S. M. Olini. Endogenous annexin A1 counter-regulates bleomycin-induced lung fibrosis. *BMC Immunology*, Vol. 12, No. 1, pp. 1–13, 2011.
- [63] P. A. Szabo, H. M. Levitin, M. Miron, M. E. Snyder, T. Senda, J. Yuan, Y. L. Cheng, E. C. Bush, P. Dogra, P. Thapa, et al. Single-cell transcriptomics of human T cells reveals tissue and activation signatures in health and disease. *Nature Communications*, Vol. 10, No. 1, pp. 1–16, 2019.
- [64] K. Hashimoto, T. Kouno, T. Ikawa, N. Hayatsu, Y. Miyajima, H. Yabukami, T. Teruoate, T. Sasaki, T. Suzuki, M. Valentine, et al. Single-cell transcrip-

- tomics reveals expansion of cytotoxic CD4 T cells in supercentenarians. *Proceedings of the National Academy of Sciences*, Vol. 116, No. 48, pp. 24242–24251, 2019.
- [65] J. M. Brenchley, N. J. Karandikar, M. R. Betts, D. R. Ambrozak, B. J. Hill, L. E. Crotty, J. P. Casazza, J. Kuruppu, S. A. Migueles, M. Connors, et al. Expression of CD57 defines replicative senescence and antigen-induced apoptotic death of CD8+ T cells. *Blood, The Journal of the American Society of Hematology*, Vol. 101, No. 7, pp. 2711–2720, 2003.
- [66] A. Nomura and H. Matsuda. Identification of lineage markers for T cell immune dysregulation in sarcoidosis using single-cell RNA-seq. *International Journal of Bioscience, Biochemistry and Bioinformatics*, 2022 in press.
- [67] Y. Zhang and T.-Y. Lee. Revealing the immune heterogeneity between systemic lupus erythematosus and rheumatoid arthritis based on multi-omics data analysis. *International Journal of Molecular Sciences*, Vol. 23, No. 9, p. 5166, 2022.
- [68] S. Jin, C. F. Guerrero-Juarez, L. Zhang, I. Chang, R. Ramos, C.-H. Kuan, P. Myung, M. V. Plikus, and Q. Nie. Inference and analysis of cell-cell communication using cellchat. *Nature Communications*, Vol. 12, No. 1, pp. 1–20, 2021.
- [69] C. Chang. Agonists and antagonists of TGF- β family ligands. *Cold Spring Harbor Perspectives in Biology*, Vol. 8, No. 8, p. a021923, 2016.
- [70] U. Khan and H. Ghazanfar. T lymphocytes and autoimmunity. *International Review of Cell and Molecular Biology*, Vol. 341, pp. 125–168, 2018.
- [71] H. Zhang, U. Costabel, and H. Dai. The role of diverse immune cells in sarcoidosis. *Frontiers in Immunology*, Vol. 12, , 2021.
- [72] M. J. Trifilo, C. C. Bergmann, W. A. Kuziel, and T. E. Lane. CC chemokine ligand 3 (CCL3) regulates CD8+-T-cell effector function and migration following viral infection. *Journal of Virology*, Vol. 77, No. 7, pp. 4004–4014, 2003.
- [73] P. Umbert, R. W. Belcher, and R. Winkelmann. Lymphokines (MIF) in the serum of patients with sarcoidosis and cutaneous granuloma annulare. *British Journal of Dermatology*, Vol. 95, No. 5, pp. 481–485, 1976.
- [74] A. Muñoz-Palomeque, M. A. Guerrero-Ramirez, L. A. Rubio-Chavez, R. C.

- Rosales-Gomez, M. G. Lopez-Cardona, V. H. Barajas-Avila, A. Delgadillo-Barrera, J. C. Canton-Romero, H. Montoya-Fuentes, T. A. Garcia-Cobian, et al. Association of RETN and CAP1 SNPs, expression and serum resistin levels with breast cancer in Mexican women. *Genetic Testing and Molecular Biomarkers*, Vol. 22, No. 4, pp. 209–217, 2018.
- [75] Z. Yan, X. Cheng, T. Wang, X. Hong, G. Shao, and C. Fu. Therapeutic potential for targeting Annexin A1 in fibrotic diseases. *Genes & Diseases*, 2022.
- [76] F. D’acquisto, M. Perretti, and R. J. Flower. Annexin-A1: a pivotal regulator of the innate and adaptive immune systems. *British Journal of Pharmacology*, Vol. 155, No. 2, pp. 152–169, 2008.
- [77] G. Shao, M. W. Julian, S. Bao, M. K. McCullers, J.-P. Lai, D. L. Knoell, and E. D. Crouser. Formyl peptide receptor ligands promote wound closure in lung epithelial cells. *American Journal of Respiratory Cell and Molecular Biology*, Vol. 44, No. 3, pp. 264–269, 2011.
- [78] L. C. Zaba, G. P. Smith, M. Sanchez, and S. D. Prystowsky. Dendritic cells in the pathogenesis of sarcoidosis. *American Journal of Respiratory Cell and Molecular Biology*, Vol. 42, No. 1, pp. 32–39, 2010.
- [79] E. Maraskovsky, E. Daro, E. Roux, M. Teepe, C. R. Maliszewski, J. Hoek, D. Caron, M. E. Lebsack, and H. J. McKenna. In vivo generation of human dendritic cell subsets by Flt3 ligand. *Blood, The Journal of the American Society of Hematology*, Vol. 96, No. 3, pp. 878–884, 2000.
- [80] M. Bhargava, K. Viken, B. Barkes, T. Griffin, M. Gillespie, P. Jagtap, R. Sajulga, E. Peterson, H. Dincer, L. Li, et al. Novel protein pathways in development and progression of pulmonary sarcoidosis. *Scientific Reports*, Vol. 10, No. 1, pp. 1–24, 2020.
- [81] A. Borzutzky, A. Fried, J. Chou, F. A. Bonilla, S. Kim, and F. Dedeoglu. NOD2-associated diseases: Bridging innate immunity and autoinflammation. *Clinical Immunology*, Vol. 134, No. 3, pp. 251–261, 2010.
- [82] C. Huppertz, B. Jäger, G. Wieczorek, P. Engelhard, S. J. Oliver, F.-G. Bauernfeind, A. Littlewood-Evans, T. Welte, V. Hornung, and A. Prasse. The NLRP3 inflammasome pathway is activated in sarcoidosis and involved in granuloma

- formation. *European Respiratory Journal*, Vol. 55, No. 3, 2020.
- [83] T. Zhou, N. Casanova, N. Pouladi, T. Wang, Y. Lussier, K. S. Knox, and J. G. Garcia. Identification of Jak-STAT signaling involvement in sarcoidosis severity via a novel microrna-regulated peripheral blood mononuclear cell gene signature. *Scientific Reports*, Vol. 7, No. 1, pp. 1–9, 2017.
 - [84] B. Levänen, Å. M. Wheelock, A. Eklund, J. Grunewald, and M. Nord. Increased pulmonary Wnt (wingless/integrated)-signaling in patients with sarcoidosis. *Respiratory Medicine*, Vol. 105, No. 2, pp. 282–291, 2011.
 - [85] T. Zhou, W. Zhang, N. J. Sweiss, E. S. Chen, D. R. Moller, K. S. Knox, S.-F. Ma, M. S. Wade, I. Noth, R. F. Machado, et al. Peripheral blood gene expression as a novel genomic biomarker in complicated sarcoidosis. *PLoS One*, Vol. 7, No. 9, p. e44818, 2012.
 - [86] J. Ding, J. Dai, H. Cai, Q. Gao, and Y. Wen. Extensively disturbance of regulatory T cells-Th17 cells balance in stage II pulmonary sarcoidosis. *International Journal of Medical Sciences*, Vol. 14, No. 11, p. 1136, 2017.
 - [87] M. Elahi, J. Talreja, B. Steinbauer, L. L. Koth, and L. Samavati. Modulatory role of macrophage migration inhibitory factor on cytokines and clinical features of sarcoidosis. *Scientific Reports*, Vol. 12, No. 1, pp. 1–11, 2022.
 - [88] W. Piotrowski, J. Kiszalkiewicz, D. Pastuszek-Lewandoska, A. Antczak, P. Górski, M. Migdalska-Sęk, W. Górski, K. Czarnecka, E. Nawrot, D. Domańska, et al. TGF- β and SMADs mRNA expression in pulmonary sarcoidosis. In *Respiratory Carcinogenesis*, pp. 59–69. Springer, 2014.
 - [89] M. Wu, Y. Zhu, S. Han, Y. Kou, X. Li, X. Gao, H. Liu, and R. He. Bioinformatics analysis of potential biomarkers associated with diagnosis and treatment of sarcoidosis involved with different tissues. *Research Square*, 2022.
 - [90] A. Krämer, J. Green, J. Pollard Jr, and S. Tugendreich. Causal analysis approaches in ingenuity pathway analysis. *Bioinformatics*, Vol. 30, No. 4, pp. 523–530, 2014.
 - [91] I. Bhavsar, C. S. Miller, and M. Al-Sabbagh. Macrophage inflammatory protein-1 alpha (MIP-1 alpha)/CCL3: as a biomarker. *General Methods in Biomarker Research and their Applications*, p. 223, 2015.

- [92] L. Vecchi, S. T. S. Mota, M. A. P. Zóia, I. C. Martins, J. B. d. Souza, T. G. Santos, A. d. O. Beserra, V. P. d. Andrade, L. R. Goulart, and T. G. Araújo. Interleukin-6 signaling in triple negative breast cancer cells elicits the annexin a1/formyl peptide receptor 1 axis and affects the tumor microenvironment. *Cells*, Vol. 11, No. 10, p. 1705, 2022.

謝辞

本論文の作成にあたり、大阪大学大学院情報科学研究科の松田秀雄教授から貴重なご指導と多大なご支援をいただきましたことに、深く感謝申し上げます。

本論文の審査にあたり、大阪大学大学院情報科学研究科の清水浩教授、瀬尾茂人准教授、岡橋伸幸准教授から貴重なご指導、ご助言をいただきましたことに、多大な感謝を申し上げます。

大阪大学大学院情報科学研究科の繁田浩功助教には、研究室や大学院の会合で貴重なご意見をいただきましたことに、誠に感謝を申し上げます。

本研究を行うにあたり、大阪大学大学院情報科学研究科バイオ情報工学専攻ゲノム情報工学講座の教員、学生、事務職員に多大なご協力をいただきました。心より感謝申し上げます。

附録

表 4.1 系譜 1 と系譜 3 の差で検出された遺伝子 (上位 50)

遺伝子	系譜 1 と系譜 3 の Wald 検定の差	系譜 1 と系譜 3 の logFC
<i>CCL5</i>	203.60	4.88
<i>NKG7</i>	115.73	6.08
<i>GNLY</i>	113.99	6.29
<i>CST7</i>	111.71	2.57
<i>CTSW</i>	99.62	3.77
<i>PLAC8</i>	91.74	3.42
<i>GZMA</i>	84.04	2.73
<i>ID2</i>	83.04	3.29
<i>TRDC</i>	80.39	5.26
<i>ANXA1</i>	78.82	2.11
<i>TRGC2</i>	75.96	4.50
<i>PRF1</i>	67.77	2.58
<i>HOPX</i>	65.26	4.94
<i>GZMB</i>	63.20	5.92
<i>EFHD2</i>	60.42	2.06
<i>KLRD1</i>	58.67	7.04
<i>KLRB1</i>	50.74	2.66
<i>TYROBP</i>	50.34	4.24
<i>IFITM2</i>	46.46	1.44
<i>SPON2</i>	45.08	3.16
<i>AOAH</i>	43.32	3.63
<i>LINC01871</i>	42.76	4.53
<i>SYTL2</i>	40.68	3.76
<i>KLRG1</i>	38.98	3.68
<i>FCGR3A</i>	37.39	4.78
<i>ABHD17A</i>	36.55	1.63
<i>ADRB2</i>	36.00	5.12
<i>CD300A</i>	34.69	3.69
<i>GZMH</i>	34.57	6.02
<i>ZEB2</i>	29.74	4.07
<i>GZMM</i>	29.21	1.76
<i>CCL4</i>	29.01	5.27
<i>A2M-AS1</i>	28.94	3.99
<i>XPB1</i>	28.28	1.82
<i>FGFBP2</i>	27.87	7.67
<i>CHST12</i>	27.72	1.57
<i>STOM</i>	26.75	2.49
<i>FGR</i>	26.17	6.34
<i>CLIC3</i>	25.06	3.02
<i>TRG-AS1</i>	24.21	2.03
<i>CMC1</i>	22.70	2.43
<i>CCL3</i>	22.30	6.56
<i>LINC00987</i>	21.58	3.01
<i>KLRF1</i>	21.51	5.10
<i>TBX21</i>	21.21	4.10
<i>DUSP2</i>	20.59	2.10
<i>C1orf21</i>	19.52	6.98
<i>ABI3</i>	19.40	2.37
<i>PATL2</i>	18.35	1.95
<i>NSG1</i>	18.25	3.45

表 4.2 クラスター解析で検出された遺伝子 (上位 50)

遺伝子	avg_logFC
<i>GNLY</i>	3.34
<i>NKG7</i>	3.33
<i>CCL5</i>	2.77
<i>KLRB1</i>	2.74
<i>GZMA</i>	2.27
<i>TRDC</i>	2.22
<i>CST7</i>	2.09
<i>TRGC1</i>	2.08
<i>DUSP2</i>	1.87
<i>KLRD1</i>	1.84
<i>FGFBP2</i>	1.64
<i>CTSW</i>	1.58
<i>GZMK</i>	1.58
<i>GZMH</i>	1.58
<i>TYROBP</i>	1.57
<i>PRF1</i>	1.57
<i>HOPX</i>	1.55
<i>GZMB</i>	1.55
<i>S100A4</i>	1.53
<i>DUSP1</i>	1.52
<i>KLRG1</i>	1.52
<i>ID2</i>	1.49
<i>CCL4</i>	1.46
<i>TRGC2</i>	1.40
<i>LYAR</i>	1.33
<i>CMC1</i>	1.30
<i>NCR3</i>	1.30
<i>MYO1F</i>	1.29
<i>C12orf75</i>	1.16
<i>EFHD2</i>	1.06
<i>CEBPD</i>	1.06
<i>LINC01871</i>	1.01
<i>KLRF1</i>	0.98
<i>MATK</i>	0.87
<i>MT-CO1</i>	0.64
<i>B2M</i>	0.51
<i>PLEK</i>	0.77
<i>KLRC1</i>	0.96
<i>SLC4A10</i>	0.58
<i>SRGN</i>	1.08
<i>IL32</i>	0.78
<i>ZFP36</i>	1.20
<i>SH3BGRL3</i>	0.89
<i>MT-CO2</i>	0.46
<i>HCST</i>	0.96
<i>FCGR3A</i>	0.79
<i>S1PR5</i>	0.58
<i>MYBL1</i>	0.68
<i>CLIC1</i>	0.95

表 4.3 Garman らの先行研究で検出された遺伝子 (上位 50)

遺伝子	avg_logFC
<i>ECH1</i>	0.36
<i>DYNLL1</i>	0.29
<i>IMP3</i>	0.28
<i>RNF167</i>	0.27
<i>PLGRKT</i>	0.25
<i>SRSF1</i>	0.23
<i>STAG2</i>	0.23
<i>MARCH7</i>	0.17
<i>CERK</i>	0.15
<i>DNAJB9</i>	0.12
<i>SUMO1</i>	0.12
<i>SURF1</i>	0.12
<i>BTN3A2</i>	0.10
<i>BIRC6</i>	0.09
<i>GRB2</i>	0.09
<i>GRK6</i>	0.09
<i>MRPS15</i>	0.08
<i>BNIP2</i>	0.08
<i>LSM4</i>	0.08
<i>GCHFR</i>	0.08
<i>RAB29</i>	0.07
<i>MANF</i>	0.07
<i>GSTP1</i>	0.07
<i>PTGDR</i>	0.07
<i>ETHE1</i>	0.07
<i>EFCAB14</i>	0.07
<i>ARMCX3</i>	0.07
<i>DDX21</i>	0.06
<i>ZMYM2</i>	0.06
<i>PTPN1</i>	0.06
<i>NUTF2</i>	0.06
<i>BAX</i>	0.06
<i>CCT6A</i>	0.05
<i>S1PR5</i>	0.05
<i>PPA1</i>	0.05
<i>TMEM50B</i>	0.05
<i>FLI1</i>	0.05
<i>ZEB2</i>	0.05
<i>AC006369.1</i>	0.05
<i>PHF1</i>	0.05
<i>AC016831.5</i>	0.05
<i>CAND1</i>	0.05
<i>FCGR3A</i>	0.04
<i>ACTR10</i>	0.04
<i>OFD1</i>	0.04
<i>DARS</i>	0.04
<i>IDH3G</i>	0.04
<i>ZFAND1</i>	0.04
<i>TFAM</i>	0.04