

博士論文（要約）

トランスクリプトームによる自己免疫疾患病態の解析

波多野 裕明

全身性自己免疫疾患の多くは、いまだ原因が不明であり、病態の理解も十分進んでいない。疾患特異的な治療薬はほとんどなく、ステロイドと免疫抑制剤による治療が中心である。

本研究では、全身性自己免疫疾患の患者末梢血の多種の免疫担当細胞サブセットをフローサイトメトリーで分取し、RNA シークエンスを行い、機械学習手法を用いてトランスクリプトームを解析し、サブセット特異的な発現遺伝子の同定、疾患分類器の作成、疾患特異的な遺伝子発現やスプライシングイベントの同定を行った。

対象とした患者は、全身性強皮症 (SSc) 67 例、全身性エリテマトーデス (SLE) 62 例、炎症性筋炎 (Myo) 59 例、ANCA 関連血管炎 (AAV) 25 例、関節リウマチ (RA) 24 例、ベーチェット病 (BD) 23 例、混合性結合組織病 (MCTD) 19 例、シェーグレン症候群 (SjS) 18 例、成人スティル病 (AOSD) 18 例、高安動脈炎 (16 例)、Overlap 症候群 8 例 (SLE-SSc, SLE-SSc-Myo, Myo-SSc 3 例, Myo-SLE, Myo-RA 2 例)およびボランティアの健常者 (HC) 79 例の計 418 例であった。

これらの患者の以下の 27 種の細胞[naïve CD4⁺ T cells (NCD4), memory CD4⁺ T cells (MCD4), type 1 helper T cells (Th1), type 2 helper T cells (Th2), type 17 helper T cells (Th17), follicular helper T cells (Tfh), activated regulatory T cells (aTreg), fraction I Treg cells (Fra1), fraction III Treg cells (Fra3), naïve CD8⁺ T cells (NCD8), memory CD8 (MCD8), effector CD8⁺ T cells (EffectorCD8), effector memory CD8⁺ T cells (EmCD8), central memory CD8⁺ T cells (CmCD8), naïve B cells (NaiB), unswitched memory B cells (UnswMB), switched memory B cells (SwiMB), plasmablast (PB), double negative B cells (DNB), NK cells (NK), CD16⁻ monocytes (CD16nMo), CD16⁺ monocytes (CD16pMo), nonclassical monocytes (NonClassical), intermediate_monocytes (Intermediate), myeloid dendritic cells (mDC), plasmacytoid dendritic cells (pDC), low density granulocyte (LDG), 好中球 (Neu)]の RNA シークエンスを行った。

まず、解析対象とした 9,897 サンプルについて、遺伝子発現による階層型クラスタリングを行ったところ、各サンプルはそれぞれの親サブセット (CD4⁺ T 細胞, CD8⁺ T 細胞, B 細胞, NK 細胞, 好中球, 単球, 樹状細胞)と対応した群にクラスタリングされたことから、各免疫細胞の遺伝子発現の特徴がデータセットに反映されていることが確認された。

さらに、他サブセットを自身の部分集合として含まない 24 サブセット 8,923 サンプルについて階層型クラスタリングを行ったところ、各サブセットがクラスターを形成し、ほぼ全サンプルがサブセットをまたぐことなくクラスタリングされた。これらの結果から、24 の異なるサブセットがそれぞれ特徴的な遺伝子発現を有すること、ほぼ全てのサンプルが正確に各サブセットの遺伝子発現の特徴を捉えていることが確認された。

次に、64 例の健常者のサンプルを用いて、サブセット特異的な発現パターンをとる遺伝子を検出した。機械学習の 1 つであるランダムフォレスト (RF)により、あるサブセットを他のサブセットと弁別する試行を 1,000 回を行い、弁別重要度によって遺伝子をランキングすることにより、各サブセットを特徴付けるのに重要な上位 100 遺伝子を同定した。同定

した遺伝子リストには、既知のマスターレギュレーター遺伝子が上位に含まれており、結果の妥当性が確認された。また、long non-coding RNA (lncRNA) も含まれており、サブセット特異的な発現パターンをとる lncRNA が存在することが確認された。

次いで疾患の違いが遺伝子発現に与える影響について検討した。サンプルが 300 以上存在する 19 サブセットを対象に、各サブセットで主成分分析を行い、上位 20 個の主成分を算出した。計 380 個の主成分について、疾患を説明変数として主成分得点を線形回帰した結果、140 個の主成分がいずれかの疾患の有無と有意に ($FDR < 0.01$) 関連した。この結果から、疾患による遺伝子発現の変動が、本データセットの遺伝子発現の分散に大きく寄与していることが分かった。

同定した 140 個の主成分を用いて、疾患および個人の階層型クラスタリングを行うことで、疾患間の類似性や疾患内の多様性について検討した。疾患のクラスタリングでは、SLE と MCTD の類似性が高く、ついで Myo, SSc, SjS, RA といった自己免疫疾患が隣接し、一方でこれらと離れて AOSD と BD, AAV と TAK という自己炎症性疾患および血管炎がクラスタリングされた。これは、臨床的な疾患分類と合致する結果であった。

さらに、各主成分がどのような遺伝子群の発現を反映しているかについて概観するため、因子負荷量に応じたエンリッチメント解析を行ったところ、SLE と MCTD に共に関連する主成分がインターフェロン関連の遺伝子群の発現を反映していることや、BD, AAV, AOSD に共に関連する主成分の一部が、Cell Cycle や Translation activity 関連の遺伝子群の発現を反映していることが示唆された。

個人のクラスタリングでは、同じ疾患では比較的隣接した位置にクラスタリングされる傾向があったが、疾患ごとに 1 つにはまともならず、個人間の分散が大きいことが推察され、患者間の病態の多様性が反映された結果と考えられた。

これまでの検討で、各疾患群は特徴的な遺伝子発現のパターンを有する一方、各患者個人単位で遺伝子発現を観察すると同一疾患内でも多様性に富むことが確認された。このことは自己免疫疾患が多彩な臨床症状を呈することに合致するが、その中で各疾患に特徴的な遺伝子発現マーカーが同定されれば、臨床診断および病態理解に役立つと考えられる。そこで、各疾患を特徴づける要素を抽出するため、50 症例以上のデータがある 3 疾患 (SLE, SSc, Myo) を対象として、遺伝子発現及びスプライシングイベントを入力に用いた RF により、i) 健常者 (HC) および ii) その他の疾患 (others) との間の弁別を行った。対象疾患群をトレーニングデータ用、テストデータ用に 6 : 4 にランダムに分割し、対象疾患群とサンプルサイズ・男女比が一致するように比較対象群もランダム抽出した。このようなランダム抽出を 100 回施行し、各回でトレーニングデータによる予測モデルの作成と重要度算出を行い、テストデータによる精度評価を行った。結果として、SLE と HC では 95%、Myo と HC では 90%、SSc と HC では 85% の予測精度が得られた。これらの疾患は治療中であったとしても、免疫細胞の遺伝子発現が健常者とは大きく異なっていることが示唆された。また、SLE と others では 90%、Myo と others, SSc と others でも 70% 前後の予測精度が得られ、各疾患に

特異的な特徴量が存在することが示された。

さらに、疾患の弁別に用いた(サブセット)_(遺伝子)または(サブセット)_(スプライシングイベント)の組を 100 回の試行での重要度の平均でランキングすることにより、これらの疾患に特徴的な遺伝子発現、スプライシングイベントを網羅的に同定し、いくつかの遺伝子において、実際に各疾患での発現に有意差があることを示した。また、疾患弁別重要度が上位であったスプライシングイベントについて、ゲノムにマップされた全リードを疾患別に重ねて可視化することにより、SLE, Myo にそれぞれ特徴的なスプライシングイベントが存在することを示した。

本研究の結果は自己免疫疾患の病態の理解、診断マーカーの作成、疾患特異的な治療薬の開発に寄与することが期待できる。今後、本研究のようなアプローチと、ゲノムデータやメタゲノムデータなどのマルチオミックス解析をすすめることにより、さらに自己免疫疾患の病態に迫ることができると考える。