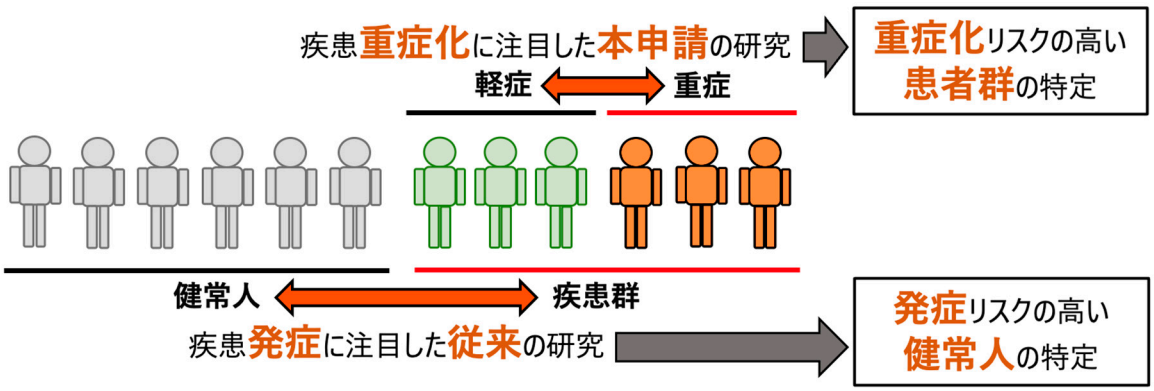
54　ゲノム情報を用いた疾患重症化の予測手法の開発	石垣　和慶

【目的】大多数の疾患の病態形成にはゲノムに内在する遺伝的素因（ゲノム因子）が強く関与する。そのため、個々の患者のゲノム情報を効率的に活用することで医療の質を改善できるという期待が世界的に高まっている。近年の大規模ゲノム研究は患者と健常人のゲノム情報を比較することで疾患発症に関与するゲノム因子を多数同定してきた。そして、これらのゲノム因子にもとづいて発症リスクの高い健常人を特定できることが示されている。しかし、臨床現場では既に発症した患者の診療が一般的であるため、発症リスクの予測の有用性は限定的である。むしろ、“発症”ではなく“重症化”リスクを予測することが医学的・社会的なニーズであり、重要な問いである。発症と重症化には異なる生物学的メカニズムが関与するため、重症化に関わるゲノム因子は重症と軽症の患者の比較によって初めて探索できるが、現状そのような研究は極めて稀である。本研究では、重症化の予測が臨床的に重要課題である全身性エリテマトーデス（SLE）を対象として、重症化に関わるゲノム因子の探索と、その因子にもとづく重症化の予測アルゴリズムを開発することを目指した。

【方法】1. 正準相関分析を応用して、SLE の症例“間”の違いを網羅的に反映した transcriptome と、転写因子による遺伝子発現制御機構（TF-gene regulatory network：TF-GRN）とを有機的に統合する新しい解析手法を開発した。この解析手法を SLE の大規模ゲノム解析結果との統合解析に適応した。2. 最新の単細胞解析である CITE-seq の条件検討を行った。健常人の末梢血免疫細胞を用いて CITE-seq のパイロット実験を行った。3. イルミナ社のマイクロアレイを用いて LUNA コホートの SLE 症例の遺伝子型判定を行った。

【結果】1. SLE の症例間の不均一性と関連するゲノム領域を同定し、その領域に SLE のリスク多型が集積していることを見出した。この結果から SLE の重症化に関わるゲノム因子の推定が可能となった。今後、これらのゲノム因子にもとづく重症化の予測アルゴリズムの開発を計画している。2. CITE-seq の各種実験条件が最適化され、末梢血の免疫細胞を高い解像度で評価できることを確認した。今後、CITE-seq を用いて SLE 患者の末梢血免疫細胞を対象とした高い解像度の transcriptome 情報を回収し、1. で開発した新しい解析手法に応用する計画である。3. アレル頻度情報などが適切であることを確認し、データの QC が完了した。今後、1. で開発する新しい予測アルゴリズムを適応し、臨床情報との関連を評価する計画である。

本研究構想の目標の概念図



54. ゲノム情報を用いた疾患重症化の予測手法の開発

石垣　和慶

理化学研究所 生命医科学研究センター ヒト免疫遺伝研究チーム

Key words：リスク多型，全身性エリテマトーデス，単細胞解析，ゲノムワイド関連解析

緒　言

大多数の疾患の病態形成にはゲノムに内在する遺伝的素因（ゲノム因子）が強く関与する。そのため、個々の患者のゲノム情報を効率的に活用することで医療の質を改善できるという期待が世界的に高まっている。近年の大規模ゲノム研究は患者と健常人のゲノム情報を比較することで疾患発症に関与するゲノム因子を多数同定してきた。そして、これらのゲノム因子にもとづいて発症リスクの高い健常人を特定できることが示されている。

しかし、臨床現場では既に発症した患者の診療が一般的であるため、発症リスクの予測の有用性は限定的である。むしろ、“発症”ではなく“重症化”リスクを予測することが医学的・社会的なニーズであり、重要な問いである。発症と重症化には異なる生物学的メカニズムが関与するため、重症化に関わるゲノム因子は重症と軽症の患者の比較によって初めて探索できるが、現状そのような研究は極めて稀である。本研究では、重症化の予測が臨床的に重要課題である全身性エリテマトーデス（SLE）を対象として、重症化に関わるゲノム因子の探索と、その因子にもとづく重症化の予測アルゴリズムを開発することを目指した。

方法および結果

1. 正準相関分析を用いた転写因子が結合するゲノム領域の推定

SLE の病態は非常に複雑であるが、ゲノムワイド関連解析（genome-wide association study：GWAS）によって多数のリスク多型が同定され、その機能解析によって病態解明が徐々に進んでいる。SLE のリスク多型は非常に数が多いため、リスク多型全体が協同して惹起する機能を評価するアプローチ（polygenic analysis）は病態の全体像の解明に有用であることが知られている［1］。代表的な polygenic analysis が stratified LD score regression（S-LDSC）である。S-LDSC はゲノム全体に散在する無数のリスク多型が平均的にどのようなゲノム領域に集積して分布しているかを頑強に推定できる手法である。S-LDSC によってヒト疾患の病態解明が進んでいる［2］。特に、免疫機能の調整に重要な転写因子の結合領域に免疫疾患のリスク多型の集積が証明されている［3〜4］。

疾患の病態などの様々な生命現象の多様性は transcriptome に反映される。そして、transcriptome の多様性は、本質的には、転写因子による遺伝子発現制御機構（TF-gene regulatory network：TF-GRN）によって形成される［2］。そのため、transcriptome の豊富な情報を GWAS との統合解析（特に polygenic analysis）に用いることは疾患病態解明を進めるために重要である。また、GWAS 結果と transcriptome を有機的に結合するには、TF-GRN を仲介させることが有効な戦略であると期待される。

我々は、SLE の重症度などの症例“間”の違いを反映するゲノム因子の探索を目的としている。その目的において、SLE の症例“間”の違いを網羅的に反映した transcriptome と、SLE の GWAS 結果との統合を有機的に統合する新しい開発手法を開発した。解析の概要を図 1 に示した。この手法では、次元圧縮法の 1 つである正準相関分析（Canonical Correlation Analysis：CCA）を用いて transcriptome と TF-GRN との間における共分散成分を最大限大きく抽出できるように、両データのサンプルに“重み”（loading）を計算する手法である。つま

り、transcriptome 側の loading 値が症例“間”の違いを反映する際に、TF-GRN 側の対応する loading 値は症例“間”の違いを反映した転写因子活性（転写因子が結合するゲノム領域）を示唆すると理解できる。そして、S-LDSC を用いて SLE のリスク多型がそれらのゲノム領域にどの程度集積しているかを評価することで、SLE の GWAS 結果の中に隠れている症例“間”の違いを反映したゲノム因子を推定することができる。

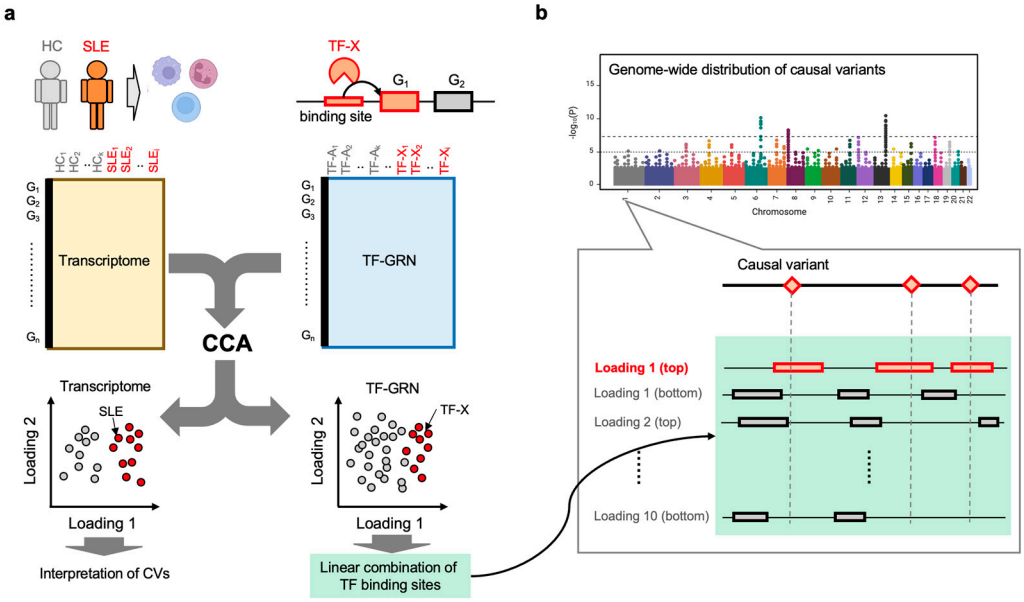


図 1. 本研究で開発した CCA を活用した新規解析手法の概要

本研究では SLE の症例“間”の違いを網羅的に反映した transcriptome と転写因子による遺伝子発現制御機構（TF-gene regulatory network：TF-GRN）とを結合し、SLE の GWAS 結果との統合解析を有機的に実現する解析手法を開発した。その概要を示した。

a）transcriptome と TF-GRN の行列がそれぞれある。行が遺伝子であり、両行列に共通である。CCA によって次元圧縮することで両行列の列の各サンプルに対する重み（loading 値）が順に計算される。図の例では、SLE は transcriptome 側の loading 1 の値が高い。同様に、転写因子 X（TF-X）は TF-GRN 側の loading 1 の値が高い。つまり、SLE の症例間を分別する病態に TF-X が関与することが推定される。また TF-GRN 側の loading を反映したゲノム領域を推定することができる。

b）a）で推定されたゲノム領域に GWAS で同定された SLE のリスク多型がどの程度集積するのかを S-LDSC で推定する。

実データの解析結果を図 2 に示した。Transcriptome には、近年報告されたさまざまな病勢の SLE の症例を含む免疫関連疾患の患者から回収された Memory CD4 陽性 T 細胞のデータを利用した［6〜7］。また、TF-GRN には公共データベースから転写因子の ChIP-seq データをダウンロードし、共通のパイプラインで既報のように作成した［5］。まず、CCA によって SLE 症例を他の疾患群と区別する成分を同定できた（図 2a の loading 7）。そして、これに対応するゲノム領域には SLE のリスク多型が集積していることが分かった。さらに、他の自己免疫疾患（関節リウマチ、クローン病）においても同様の所見が確認された（図 2b）。一方、negative control として 2 型糖尿病のリスク多型も評価したが、予想どおり有意な集積は認めなかった。この解析によって、SLE を非免疫疾患と区別し、かつ、SLE のリスク多型が集積したゲノム領域を同定することができた。今後、この領域にあるリスク多型に重みを与えた polygenic risk score などの新しいリスクスコアの計算方法を検討する予定である。

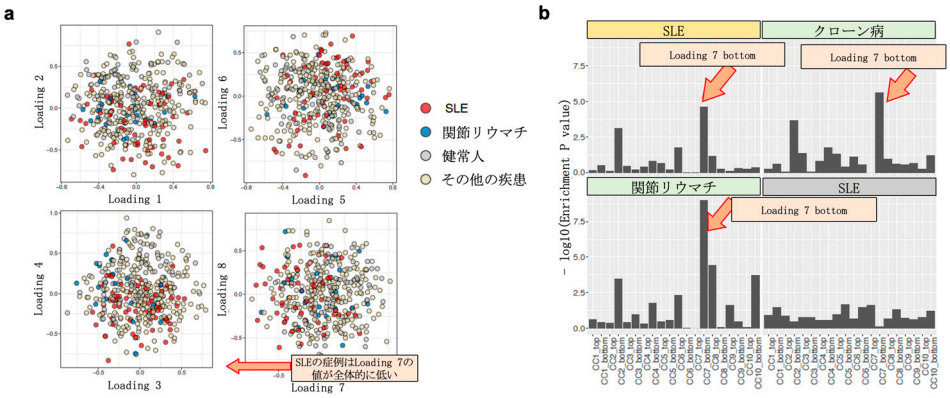


図 2. 本研究で開発した新規解析手法の実データへの応用例

a) CCA の transcriptome 側の次元圧縮結果を示す。7 つ目の loading（loading 7）では SLE が他の症例よりも低い値を有していることが分かる。

b) 各 loading が高いゲノム領域（top）と低いゲノム領域（bottom）に各疾患のリスク多型がどの程度集積しているかを S-LDSC で評価し、集積度の有意性を bootstrap 法で推定した。SLE を始め免疫疾患で loading 7 の低いゲノム領域にリスク多型の集積が確認された。

2. CITE-seq を用いた免疫細胞の単細胞 transcriptome 解析

上述の CCA を活用した解析手法には高い汎用性があり、さまざまな transcriptome のデータに応用することができる。そこで、最新の単細胞 transcriptome 解析技術である CITE-seq を活用して SLE 患者の末梢血免疫細胞の解析を行うことを計画した。CITE-seq は 10X Genomics 社が提供する最新の単細胞解析プラットフォームであり、免疫細胞の表現型の分別に有用な細胞表面マーカーに対する抗体に RNA の標識オリゴが結合している。

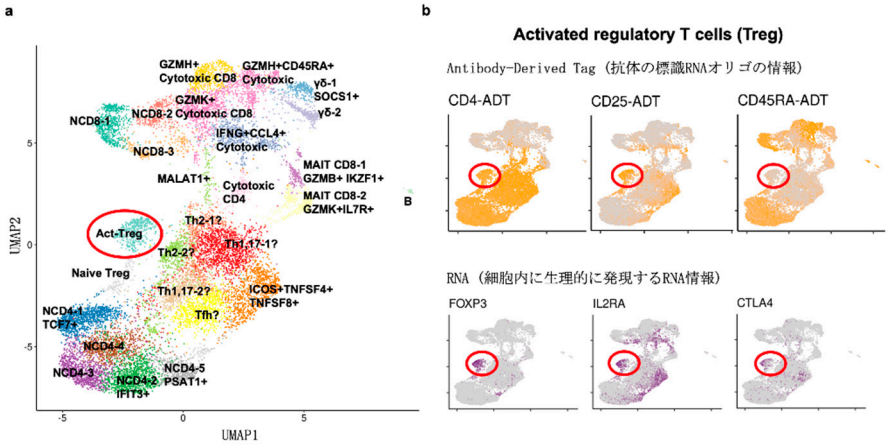


図 3. CITE-seq のテストデータ

健常人から末梢血単核球を採取した。MACS で T 細胞を精製し、CITE-seq で評価した。

a) Antibody derived tag（抗体の標識 RNA オリゴ）の情報と RNA 発現情報を Weighted Nearest Neighbor Analysis を用いて統合解析し、UMAP で次元圧縮を行った結果を図示した。また同一情報からクラスタリングを行い、同定した細胞集団を異なる色で示した。発現プロファイルを参考に各細胞集団のアノテーションを行い、図中に示した。確実ではないアノテーションには“？”を付記した。

b) パネル a の赤で囲んである細胞集団（activated regulatory T cell : act-Treg）の Antibody derived tag と RNA の発現情報を示す。Act-Treg に特徴的な細胞表面マーカーと遺伝子の具体例を 3 つずつ示した。

細胞内の生理的に発現する RNA と共に標識オリゴがシーケンスライブラリーに反映される。そのため、CITE-seq では、各単細胞から細胞表面マーカーと細胞内の RNA とを同時に評価することができる。本研究では、そのテストデータとして、健常人の末梢血免疫細胞を用いた CITE-seq の各種条件の最適化を行い、多くの細胞集団を同定することができた（図 3）。

3. 全身性エリテマトーデス患者の遺伝子型判定実験

国内最大規模の SLE レジストリである LUNA コホートと連携して研究を進めている。本研究では、本コホートと連携し、抽出済みの約 400 人の SLE 患者のゲノムサンプルから、マイクロアレイを用いて遺伝子多型情報を回収した。現在、QC が完了し、適切な情報が回収できていることを確認した（図 4）。今後、上述の新しいリスクスコアと臨床情報との関連を評価し、リスクスコアの臨床的有用性を評価する予定である。

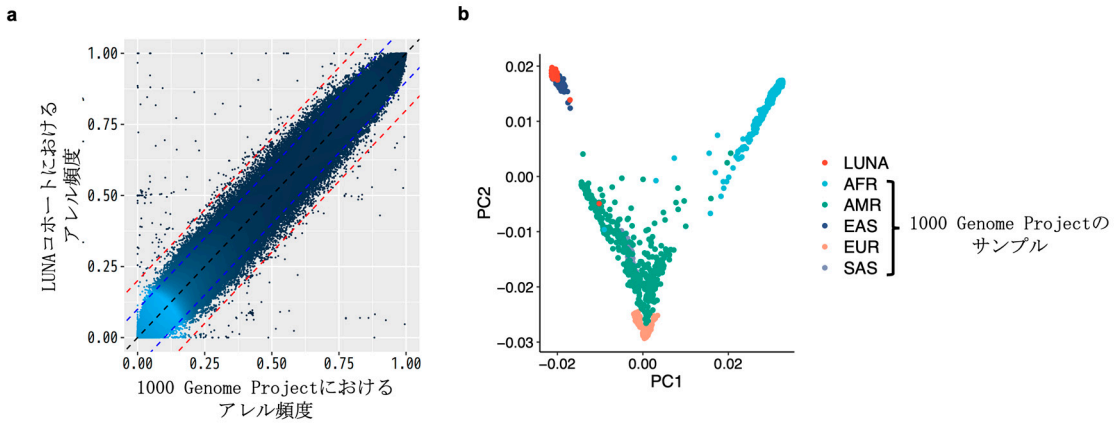


図 4. LUNA コホートの遺伝子型情報の QC

イルミナ社のマイクロアレイ（Infinium Asian Screening Array）を用いて遺伝子型判定を行った。その QC の概要を示す。

a） 各 SNP の LUNA コホートにおけるアレル頻度を 1000 Genome Project の東アジア人と比較した。概ね整合性のとれた結果を確認できた。参考としてアレル頻度の差が 0.2 の部分に赤点線を示している。

b） 遺伝子型情報を用いて主成分解析を実施した結果を示す。

考　察

自己免疫疾患の病態は複雑であり、詳細な分子機序を明らかにすることは容易ではない。しかし、複雑な病態が正確に反映された transcriptome 情報を収集することは可能であり、本研究で使用したものを含め過去に多くのデータが報告されている。本研究では、そのような豊富に存在する transcriptome 情報とリスク多型の分布を有機的に結合するアルゴリズムを開発した。このアルゴリズムを活用することで、詳細な分子機序が未解明の状況であっても、SLE の重症度の違いがどのようなゲノム領域に存在するリスク多型により惹起されているかが推定可能となる。この知見は、個々人の遺伝子型情報から SLE の重症度を予測するゲノムスコアの基盤情報となる。今後、本アルゴリズムを更に高度化・最適化し、LUNA コホートの遺伝子型情報と臨床情報との統合解析を進める計画である。

共同研究者・謝辞

本研究の共同研究者は、国立長寿医療研究センターメディカルゲノムセンターの尾崎浩一センター長、岡山大学学術研究院医歯薬学域腎・免疫・内分泌代謝内科学の勝山隆行助教、長崎大学大学院医歯薬学総合研究科先進予防医学共同専攻リウマチ・膠原病内科学分野の一瀬邦弘准教授（申請当時）である。

文　献

1) Yin X, Kim K, Suetsugu H, Bang SY, Wen L, Koido M, Ha E, Liu L, Sakamoto Y, Jo S, Leng RX, Otomo N, Laurynenka V, Kwon YC, Sheng Y, Sugano N, Hwang MY, Li W, Mukai M, Yoon K, Cai M, Ishigaki K, Chung WT, Huang H, Takahashi D, Lee SS, Wang M, Karino K, Shim SC, Zheng X, Miyamura T, Kang YM, Ye D, Nakamura J, Suh CH, Tang Y, Motomura G, Park YB, Ding H, Kuroda T, Choe JY, Li C, Niiro H, Park Y, Shen C, Miyamoto T, Ahn GY, Fei W, Takeuchi T, Shin JM, Li K, Kawaguchi Y, Lee YK, Wang Y, Amano K, Park DJ, Yang W, Tada Y, Yamaji K, Shimizu M, Atsumi T, Suzuki A, Sumida T, Okada Y, Matsuda K, Matsuo K, Kochi Y; Japanese Research Committee on Idiopathic Osteonecrosis of the Femoral Head; Kottyan LC, Weirauch MT, Parameswaran S, Eswar S, Salim H, Chen X, Yamamoto K, Harley JB, Ohmura K, Kim TH, Yang S, Yamamoto T, Kim BJ, Shen N, Ikegawa S, Lee HS, Zhang X, Terao C, Cui Y, Bae SC. Meta-analysis of 208,370 East Asians identifies 113 susceptibility loci for systemic lupus erythematosus. Ann Rheum Dis. 2021 May;80(5):632-640. doi: 10.1136/annrheumdis-2020-219209. Epub 2020 Dec 3. PMID: 33272962.

2) Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh PR, Anttila V, Xu H, Zang C, Farh K, Ripke S, Day FR; ReproGen Consortium; Schizophrenia Working Group of the Psychiatric Genomics Consortium; RACI Consortium; Purcell S, Stahl E, Lindstrom S, Perry JR, Okada Y, Raychaudhuri S, Daly MJ, Patterson N, Neale BM, Price AL. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat Genet. 2015 Nov;47(11):1228-35. doi: 10.1038/ng.3404. Epub 2015 Sep 28. PMID: 26414678.

3) Amariuta T, Ishigaki K, Sugishita H, Ohta T, Koido M, Dey KK, Matsuda K, Murakami Y, Price AL, Kawakami E, Terao C, Raychaudhuri S. Improving the trans-ancestry portability of polygenic risk scores by prioritizing variants in predicted cell-type-specific regulatory elements. Nat Genet. 2020 Dec;52(12):1346-1354. doi: 10.1038/s41588-020-00740-8. Epub 2020 Nov 30. PMID: 33257898.

4) Ishigaki K, Akiyama M, Kanai M, Takahashi A, Kawakami E, Sugishita H, Sakaue S, Matoba N, Low SK, Okada Y, Terao C, Amariuta T, Gazal S, Kochi Y, Horikoshi M, Suzuki K, Ito K, Koyama S, Ozaki K, Niida S, Sakata Y, Sakata Y, Kohno T, Shiraishi K, Momozawa Y, Hirata M, Matsuda K, Ikeda M, Iwata N, Ikegawa S, Kou I, Tanaka T, Nakagawa H, Suzuki A, Hirota T, Tamari M, Chayama K, Miki D, Mori M, Nagayama S, Daigo Y, Miki Y, Katagiri T, Ogawa O, Obara W, Ito H, Yoshida T, Imoto I, Takahashi T, Tanikawa C, Suzuki T, Sinozaki N, Minami S, Yamaguchi H, Asai S, Takahashi Y, Yamaji K, Takahashi K, Fujioka T, Takata R, Yanai H, Masumoto A, Koretsune Y, Kutsumi H, Higashiyama M, Murayama S, Minegishi N, Suzuki K, Tanno K, Shimizu A, Yamaji T, Iwasaki M, Sawada N, Uemura H, Tanaka K, Naito M, Sasaki M, Wakai K, Tsugane S, Yamamoto M, Yamamoto K, Murakami Y, Nakamura Y, Raychaudhuri S, Inazawa J, Yamauchi T, Kadowaki T, Kubo M, Kamatani Y. Large-scale genome-wide association study in a Japanese population identifies novel susceptibility loci across different diseases. Nat Genet. 2020 Jul;52(7):669-679. doi: 10.1038/s41588-020-0640-3. Epub 2020 Jun 8. PMID: 32514122.

5) Kawakami E, Nakaoka S, Ohta T, Kitano H. Weighted enrichment method for prediction of transcription regulators from transcriptome and global chromatin immunoprecipitation data. Nucleic Acids Res. 2016 Jun 20;44(11):5010-21. doi: 10.1093/nar/gkw355. Epub 2016 Apr 30. PMID: 27131787.
6) Ota M, Nagafuchi Y, Hatano H, Ishigaki K, Terao C, Takeshima Y, Yanaoka H, Kobayashi S, Okubo M, Shirai H, Sugimori Y, Maeda J, Nakano M, Yamada S, Yoshida R, Tsuchiya H, Tsuchida Y, Akizuki S, Yoshifuji H, Ohmura K, Mimori T, Yoshida K, Kurosaka D, Okada M, Setoguchi K, Kaneko H, Ban N, Yabuki N, Matsuki K, Mutoh H, Oyama S, Okazaki M, Tsunoda H, Iwasaki Y, Sumitomo S, Shoda H, Kochi Y, Okada Y, Yamamoto K, Okamura T, Fujio K. Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases. Cell. 2021 May 27;184(11):3006-3021.e17. doi: 10.1016/j.cell.2021.03.056. Epub 2021 May 1. PMID: 33930287.
7) Nakano M, Ota M, Takeshima Y, Iwasaki Y, Hatano H, Nagafuchi Y, Itamiya T, Maeda J, Yoshida R, Yamada S, Nishiwaki A, Takahashi H, Takahashi H, Akutsu Y, Kusuda T, Suetsugu H, Liu L, Kim K, Yin X, Bang SY, Cui Y, Lee HS, Shoda H, Zhang X, Bae SC, Terao C, Yamamoto K, Okamura T, Ishigaki K, Fujio K. Distinct transcriptome architectures underlying lupus establishment and exacerbation. Cell. 2022 Sep 1;185(18):3375-3389.e21. doi: 10.1016/j.cell.2022.07.021. Epub 2022 Aug 22. PMID: 35998627.