

REVIEW

Open Access

Multi-omics approach to precision medicine for immune-mediated diseases



Mineto Ota^{1,2*}  and Keishi Fujio¹

Abstract

Recent innovation in high-throughput sequencing technologies has drastically empowered the scientific research. Consequently, now, it is possible to capture comprehensive profiles of samples at multiple levels including genome, epigenome, and transcriptome at a time. Applying these kinds of rich information to clinical settings is of great social significance. For some traits such as cardiovascular diseases, attempts to apply omics datasets in clinical practice for the prediction of the disease risk have already shown promising results, although still under way for immune-mediated diseases. Multiple studies have tried to predict treatment response in immune-mediated diseases using genomic, transcriptomic, or clinical information, showing various possible indicators. For better prediction of treatment response or disease outcome in immune-mediated diseases, combining multi-layer information together may increase the power. In addition, in order to efficiently pick up meaningful information from the massive data, high-quality annotation of genomic functions is also crucial. In this review, we discuss the achievement so far and the future direction of multi-omics approach to immune-mediated diseases.

Keywords: Multi-omics analysis, Immune-mediated disease, Genome, Transcriptome, Expression quantitative trait loci, Polygenic risk score

Background

Immune-mediated diseases (IMDs) consist of a wide range of etiologies from autoimmune to autoinflammatory conditions [1]. Although a variety of therapeutic agents and regimens have been developed for each IMD, the treatment response varies from patient to patient. The heterogeneous pathogenesis of an IMD may be associated with different outcomes [2].

Recent advances in high-throughput sequencing technologies have enabled capturing comprehensive profiles of samples at multiple levels, referred to as omics analysis, for example, genomic, epigenomic, transcriptomic, and proteomic analyses. Multi-omics analysis refers to the collective analysis of omics data at multiple levels [3] and enables deep phenotyping of patients.

Personalized medicine, more recently referred to as precision medicine, aims to develop drugs and to optimize prescription of the appropriate drugs at the optimal dose and time [4, 5]. To achieve this aim, a deep understanding of the disease etiology, accurate and early diagnosis, appropriate stratification of patients based on disease phenotype, and prediction of the treatment response are required.

Applying multi-omics information to precision medicine is a clinically attractive challenge. Although still in its infancy, there have been numerous efforts to apply omics data in clinical practice using various approaches (Fig. 1). Here, we review some of these approaches and discuss their potential future directions.

Prediction of treatment response in rheumatoid arthritis

A growing number of drugs, from small molecule compounds to biologics, have been approved for IMDs so far [6]. In particular, many biological and non-biological disease-modifying anti-rheumatic drugs are widely used to treat rheumatoid arthritis (RA) [7, 8]; however, the

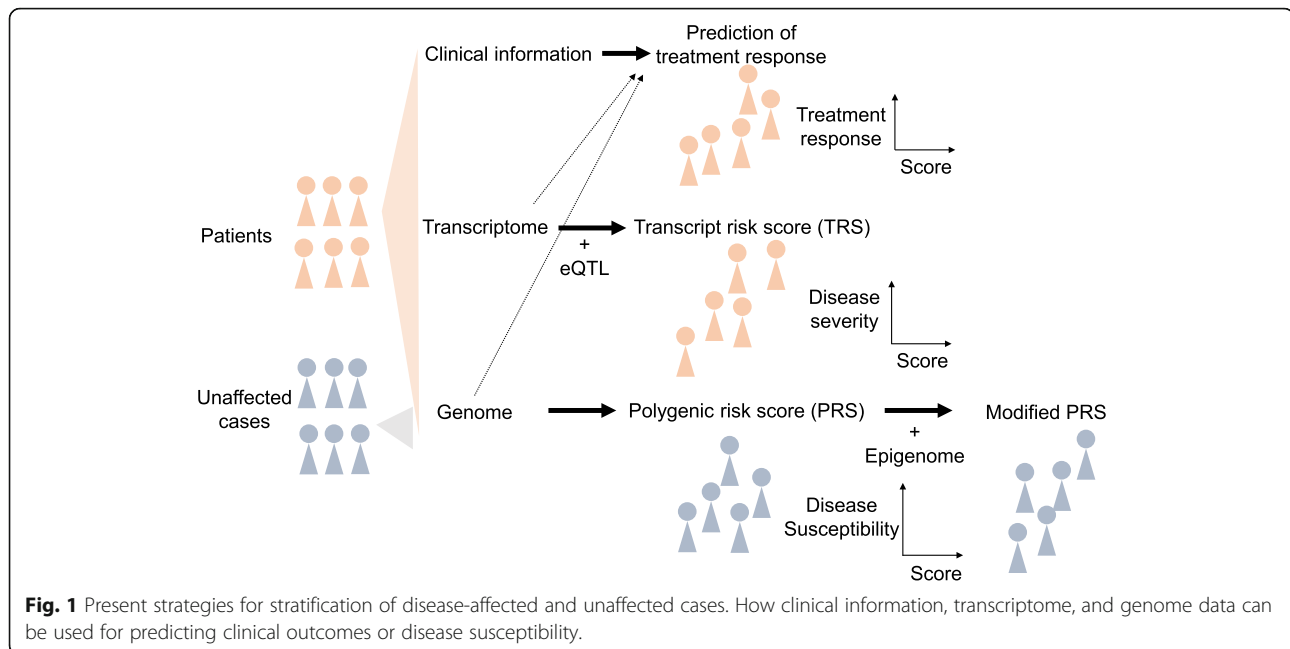
* Correspondence: otam-int@h.u-tokyo.ac.jp

¹Department of Allergy and Rheumatology, Graduate School of Medicine, The University of Tokyo, Tokyo 113-0033, Japan

²Department of Functional Genomics and Immunological Diseases, Graduate School of Medicine, The University of Tokyo, Tokyo 113-0033, Japan



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.



treatment response varies from patient to patient, with only 30–40% of patients showing an adequate response to first-line biological disease-modifying anti-rheumatic drugs [9, 10]. Clinically, prediction of the treatment response before administering therapy to reduce the risk of side effects and the economic burden has gained great interest [10].

For this purpose, many studies have been conducted over the last 10 years (reviewed in detail elsewhere [11]). Briefly, patients serologically positive for rheumatoid factor or anti-citrullinated protein antibodies seem to show good responses to rituximab (RTX), an anti-CD20 monoclonal antibody [12, 13], to abatacept (CTLA4-Ig fusion protein) [14] and to tocilizumab (TCZ) [15], although serological positivity did not seem to be predictive of the response to TNF inhibitors [16]. In addition, a strong interferon (IFN) signature in serum was reported to predict a good response to TCZ [17] but a poor response to RTX [18].

Based on these findings, more specific and accurate markers for predicting treatment response have been investigated using omics data. Potential genetic variants associated with treatment response based on genomic data have been reported. For example, variants in the *PDE3A-SLCO1C1* [19], *CD84* [20], and *PTPRC* [21] loci and variants affecting the expression of *CD40* and *CD39* [22] were found to be associated with the response to TNF inhibitors; however, a limited number of those findings could be reproduced by a subsequent study [23], and thus further validation is required.

For fair evaluation of the usefulness of genomic data, one community-based assessment aimed to develop

models predicting the response to TNF inhibitors in RA patients in 2016 [24]. In this open challenge, genomic data and clinical information from over 2000 patients were provided to 73 research groups to generate a prediction model using various approaches, including machine learning methods. Contrary to expectations, all of the groups failed to show a significant contribution of genetics to prediction accuracy. The highest performing model in this challenge greatly relied on clinical parameters, especially the 28-joint disease activity score at baseline [25]. Although this study reinforced the importance of clinical parameters for predicting treatment responses, clinical application of genetic data may still have a role in the prediction of drug responses, especially in larger sample sizes [26], and in combination with other omics levels such as transcriptomic data.

In contrast to genomic data, transcriptomic data reflect the variance acquired from environmental factors as well as from genetics. For instance, exposure to inflammation, cellular activation, or cellular composition can be reflected in transcriptomic data [27]. In some reports, transcriptomic data proved useful for patient stratification. In RA, synovial fibroblasts (SFs) play important roles in joint inflammation [28] and show a dynamic response to inflammation at the epigenomic and transcriptomic levels [29]. Lewis et al. classified RA patients into three distinct pathotypes (fibroblastic pauci-immune, macrophage-rich diffuse-myeloid, and lymphomyeloid pathotypes) based on gene expression in the synovium and assessed the treatment response associated with each pathotype using transcriptomic data [30]. They found that the pauci-immune pathotype was

predictive of an inadequate response to TNF inhibitors [31]. In addition, Humby et al. recently conducted a biopsy-driven randomized controlled study (RTX vs. TCZ) and reported that “B-cell poor” patients (defined by RNA sequencing of SFs) showed a better response to TCZ than to RTX [32]. Although further validation is warranted, the transcriptomic data from diseased tissue might enable better stratification of IMD patients. Taken together, combining multi-level information, such as clinical information, serological information, and genomic/transcriptomic data, will improve the prediction of treatment responses in RA patients.

Prediction of disease susceptibility using genomic data

Application of genomic data for precision medicine, especially for disease prevention, has gained wide interest in this era [33]. In particular, the polygenic risk score (PRS) has been applied to a variety of diseases, and some of the results have been promising. PRS is calculated by summing the effects of all common variants on disease onset to estimate the overall risk of developing a particular disease. It is especially useful for polygenic traits, in which small effects of numerous common variants contribute to disease onset. For instance, coronary artery disease (CAD) is a prevalent polygenic trait. In one study, individuals with a high PRS (8% of the population) had at least a threefold increased risk of developing CAD [34]. A clinical trial using the PRS for treatment intervention in CAD patients also had a promising outcome [35]. These results highlight the possibility of using genetic information to stratify high-risk patients, although predicting the risk of disease onset using only genomic data is still a challenge.

Generally, a larger sample size in genome-wide association studies (GWAS) is necessary for better prediction of the PRS, although the explained genetic variance of each trait also has an influence [36]. Among IMDs, relatively large-scaled GWAS for inflammatory bowel disease (IBD) have been performed [37, 38], and those datasets enabled promising disease risk prediction in the subsequent studies [34, 39]. In one of those studies, subjects with a PRS in the top 10% of distribution had a 2.43-fold increased disease susceptibility compared to the remaining 90% [34]. Although pre-disease interventions in IBD have not been established, some dietary habits have been associated with IBD onset [40, 41]. Thus, dietary intervention in high-risk cases could reduce the risk of disease onset.

In clinical practice in IMDs, correct diagnosis at the first outpatient visit is important. One interesting study evaluated the value of genetic data combined with clinical examination for accurate diagnosis of inflammatory arthritis [42]. In that study, based on retrospective data,

the authors indicated that calculating disease risk using genetic information significantly improved the accuracy of the initial diagnosis of arthritis. As another example, Zhao et al. reported that the DNA methylation level at the *IFI44L* locus distinguished SLE patients from healthy controls, RA patients, and Sjögren’s syndrome patients, suggesting its potential as a diagnostic marker for SLE [43]. Those studies support the utility of genomic and epigenomic data for improving clinical practice in rheumatology, although further validation by prospective studies in real clinical settings is required.

Attempts to improve the performance of the PRS

Although clinical application of the PRS to classify increased risks of certain traits has become realistic, PRSs developed by GWAS for a specific population tended to underperform when tested in a different population [44, 45]. As most large-scale GWAS have been performed in European populations, this might lead to health disparities among populations [45].

Some attempts to overcome this issue have already been reported. Amariuta et al. reported that the trans-ethnic portability of the PRS was significantly improved by prioritizing variants with regulatory annotation which was constructed based on epigenomic data [46]. Another group reported that calculating the PRS based on variants discovered among diverse populations improved the trans-ethnic portability of the PRS [47]. Although conducting large-scale GWAS in diverse populations is also an important direction [48], combination with other omics data, such as epigenomic data, would improve the performance of the PRS in the clinical setting.

Application of epigenomic data for multi-omics analysis

As exemplified above, epigenomic data is a quite valuable resource for multi-omics analysis. Recently, Encyclopedia of DNA elements (ENCODE) Project has released their phase III data, which consists of wide variety of epigenomic annotations from 5992 new experimental datasets [49]. Together with Roadmap Epigenomics data [50], these large-scaled epigenome datasets inform cis-regulatory elements of variety of cell types. As epigenomic data is not influenced by linkage disequilibrium or allele frequencies, these data can be utilized for prioritization of disease-associated genetic variants in the context of precision medicine.

Application of expression quantitative trait loci (eQTL) data for precision medicine

eQTL analysis is used to identify the association between genetic variants and gene expression. During the past 10 years, large consortia including Geuvadis [51], GTEx [52], DICE [53], and eQTLGen [54] performed eQTL

analyses in various tissues and cell types, using large sample sizes, and reported the effects of genetic variants on gene expression. The resulting datasets are quite useful for estimating the effects of disease-associated variants. Generally, majority of non-coding disease-associated genetic variants are assumed to modulate the expression of genes that play roles in disease pathogenesis. Thus, it is reasonable to integrate GWAS results with eQTL data and estimate gene-level associations with diseases to reduce the multiple testing burden and facilitate biological interpretation. This approach, referred to as transcriptome-wide association studies (TWAS), has garnered great interest, and many such studies have been performed [55–57].

Integration of eQTL and GWAS datasets can be applied for patient-level estimation of disease risk. Mari-gorta et al. developed a transcriptional risk score (TRS) based on the gene expression data in ileal mucosal samples from Crohn's disease patients, risk variants of Crohn's disease determined by GWAS, and the eQTL effects of these variants [58]. The TRS outperformed genetic risk scores in terms of not only distinguishing Crohn's disease from healthy samples but also identifying patients who will progress to complicated disease [58]. Although further validation in other centers is warranted, that study raises the possibility of treating gene expression data derived from biopsy specimens as an index for patient classification, in combination with GWAS and eQTL data.

To enhance understanding of the functions of IMD-associated genetic variants, their functions in immune cells should be evaluated. Recently, we constructed an eQTL atlas based on 28 types of immune cells (ImmuNexUT; Immune Cell Gene Expression Atlas from the University of Tokyo) [27]. Our atlas showed enrichment of IMD-associated genetic variants in immune cell eQTLs and identified a number of IMD-associated genes and cell types. This information could be used to prioritize disease-relevant cell types and genes and subsequently stratify IMD patients in the future. In addition, information on genetic variants associated with heterogeneity within a disease is limited so far. In a GWAS of IBD, few genetic variants were associated with disease prognosis [59]. The difficulty of clinically defining heterogeneity within a disease and the small number of patients with a rare disease entity make it a challenge to identify disease heterogeneity-associated variants. Using ImmuNexUT, we identified immune cell eQTL variants that show heterogeneity in effect sizes in an inflammatory context-dependent manner. These context-dependent eGenes (genes possessing context-dependent eQTL variants) are enriched in genes induced by inflammation or vaccination, indicating their roles in diversifying an individual's response to inflammation. We

surmise that these context-dependent eQTL variants are candidates for IMD heterogeneity and will aid prioritization of variants and stratification of IMD patients.

Conclusion

For predicting disease susceptibility, disease severity, and treatment response, multi-omics data may play an important role in clinical practice in the near future. So far, some attempts at patient stratification have been performed using genomic, epigenomic, transcriptomic, and clinical data; however, most of those studies were based on information from a single level. Combination of multi-level information will improve the prediction of these outcomes. Construction of large-scale patient cohorts with high-quality clinical data (e.g., treatment response, clinical prognosis, and genomic, transcriptomic, and epigenomic data, preferably from diseased tissues) and refined analytic approaches to handle these data would contribute to a better understanding of IMD biology and accelerate precision medicine in IMD patients.

Abbreviations

IMD: Immune-mediated disease; RA: Rheumatoid arthritis; RTX: Rituximab; TCZ: Tocilizumab; IFN: Interferon; PRS: Polygenic risk score; SF: Synovial fibroblast; CAD: Coronary artery disease; GWAS: Genome-wide association study; IBD: Inflammatory bowel disease; eQTL: Expression quantitative trait loci; TWAS: Transcriptome-wide association study; TRS: Transcriptional risk score; ImmuNexUT: Immune Cell Gene Expression Atlas from the University of Tokyo

Acknowledgements

Not applicable.

Authors' contributions

M.O. and K.F. designed and wrote the manuscript. The authors read and approved the final manuscript.

Funding

Not applicable.

Availability of data and materials

Not applicable.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

M.O. belongs to the Social Cooperation Program, Department of Functional Genomics and Immunological Diseases, supported by Chugai Pharmaceutical.

Received: 11 May 2021 Accepted: 26 June 2021

Published online: 01 August 2021

References

- van Kempen TS, Wenink MH, Leijten EF, Radstake TR, Boes M. Perception of self: distinguishing autoimmunity from autoinflammation. *Nat Rev Rheumatol*. 2015;11(8):483–92. <https://doi.org/10.1038/nrrheum.2015.60>.

2. Cho JH, Feldman M. Heterogeneity of autoimmune diseases: pathophysiologic insights from genetics and implications for new therapies. *Nature Med.* 2015;21(7):730–8. <https://doi.org/10.1038/nm.3897>.
3. Subramanian I, Verma S, Kumar S, Jere A, Anamika K. Multi-omics data integration, interpretation, and its application. *Bioinform Biol Insights.* 2020; 14:1177932219899051.
4. Hamburg MA, Collins FS. The path to personalized medicine. *N Eng J Med.* 2010;363(4):301–4. <https://doi.org/10.1056/NEJMp1006304>.
5. Marson FAL, Bertuzzo CS, Ribeiro JD. Personalized or precision medicine? The example of cystic fibrosis. *Front Pharmacol.* 2017;8:390.
6. Li P, Zheng Y, Chen X. Drugs for autoimmune inflammatory diseases: from small molecule compounds to anti-TNF biologics. *Front Pharmacol.* 2017;8: 460. <https://doi.org/10.3389/fphar.2017.00460>.
7. Sepriano A, Kerschbaumer A, Smolen JS, van der Heijde D, Dougados M, van Vollenhoven R, et al. Safety of synthetic and biological DMARDs: a systematic literature review informing the 2019 update of the EULAR recommendations for the management of rheumatoid arthritis. *Ann Rheum Dis.* 2020;79(6):760–70. <https://doi.org/10.1136/annrheumdis-2019-216653>.
8. Smolen JS, Landewé RBM, Bijlsma JWJ, Burmester GR, Dougados M, Kerschbaumer A, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2019 update. *Ann Rheum Dis.* 2020;79(6):685–99. <https://doi.org/10.1136/annrheumdis-2019-216655>.
9. Rein P, Mueller RB. Treatment with biologicals in rheumatoid arthritis: an overview. *Rheumatol Ther.* 2017;4(2):247–61. <https://doi.org/10.1007/s40744-017-0073-3>.
10. Shi Q, Li KJ, Treuer T, Wang BCM, Gaich CL, Lee CH, et al. Estimating the response and economic burden of rheumatoid arthritis patients treated with biologic disease-modifying antirheumatic drugs in Taiwan using the National Health Insurance Research Database (NHIRD). *PLoS One.* 2018;13(4): e0193489. <https://doi.org/10.1371/journal.pone.0193489>.
11. Romão VC, Vital EM, Fonseca JE, Buch MH. Right drug, right patient, right time: aspiration or future promise for biologics in rheumatoid arthritis? *Arthritis Res Ther.* 2017;19(1):239. <https://doi.org/10.1186/s13075-017-1445-3>.
12. Isaacs JD, Cohen SB, Emery P, Tak PP, Wang J, Lei G, et al. Effect of baseline rheumatoid factor and anticitrullinated peptide antibody serotype on rituximab clinical response: a meta-analysis. *Ann Rheum Dis.* 2013;72(3):329–36. <https://doi.org/10.1136/annrheumdis-2011-201117>.
13. Chatzidionysiou K, Lie E, Nasonov E, Lukina G, Hetland ML, Tarp U, et al. Highest clinical effectiveness of rituximab in autoantibody-positive patients with rheumatoid arthritis and in those for whom no more than one previous TNF antagonist has failed: pooled data from 10 European registries. *Ann Rheum Dis.* 2011;70(9):1575–80. <https://doi.org/10.1136/ard.2010.148759>.
14. Gottenberg JE, Courvoisier DS, Hernandez MV, Iannone F, Lie E, Canhão H, et al. Brief report: association of rheumatoid factor and anti-citrullinated protein antibody positivity with better effectiveness of abatacept: results from the pan-European registry analysis. *Arthritis Rheumatol.* 2016;68(6): 1346–52. <https://doi.org/10.1002/art.39595>.
15. Maneiro RJ, Salgado E, Carmona L, Gomez-Reino JJ. Rheumatoid factor as predictor of response to abatacept, rituximab and tocilizumab in rheumatoid arthritis: systematic review and meta-analysis. *Semin Arthritis Rheum.* 2013;43(1):9–17. <https://doi.org/10.1016/j.semarthrit.2012.11.007>.
16. Lv Q, Yin Y, Li X, Shan G, Wu X, Liang D, et al. The status of rheumatoid factor and anti-cyclic citrullinated peptide antibody are not associated with the effect of anti-TNF α agent treatment in patients with rheumatoid arthritis: a meta-analysis. *PLoS One.* 2014;9(2):e89442. <https://doi.org/10.1371/journal.pone.0089442>.
17. Sanayama Y, Ikeda K, Saito Y, Kagami S, Yamagata M, Furuta S, et al. Prediction of therapeutic responses to tocilizumab in patients with rheumatoid arthritis: biomarkers identified by analysis of gene expression in peripheral blood mononuclear cells using genome-wide DNA microarray. *Arthritis Rheumatol.* 2014;66(6):1421–31. <https://doi.org/10.1002/art.38400>.
18. Thurlings RM, Boumans M, Tekstra J, van Roon JA, Vos K, van Westing DM, et al. Relationship between the type I interferon signature and the response to rituximab in rheumatoid arthritis patients. *Arthritis Rheum.* 2010;62(12): 3607–14. <https://doi.org/10.1002/art.27702>.
19. Acosta-Colman I, Palau N, Tornero J, Fernández-Nebro A, Blanco F, González-Alvaro I, et al. GWAS replication study confirms the association of PDE3A-SLCO1C1 with anti-TNF therapy response in rheumatoid arthritis. *Pharmacogenomics.* 2013;14(7):727–34. <https://doi.org/10.2217/pgs.13.60>.
20. Cui J, Stahl EA, Saevarsdottir S, Miceli C, Diogo D, Trynka G, et al. Genome-wide association study and gene expression analysis identifies CD84 as a predictor of response to etanercept therapy in rheumatoid arthritis. *PLoS Genet.* 2013;9(3):e1003394. <https://doi.org/10.1371/journal.pgen.1003394>.
21. Cui J, Saevarsdottir S, Thomson B, Padyukov L, van der Helm-van Mil AH, Niittham J, et al. Rheumatoid arthritis risk allele PTPRC is also associated with response to anti-tumor necrosis factor alpha therapy. *Arthritis Rheum.* 2010;62(7):1849–61. <https://doi.org/10.1002/art.27457>.
22. Spiliopoulou A, Colombo M, Plant D, Nair N, Cui J, Coenen MJ, et al. Association of response to TNF inhibitors in rheumatoid arthritis with quantitative trait loci for CD40 and CD39. *Ann Rheum Dis.* 2019;78(8):1055–61. <https://doi.org/10.1136/annrheumdis-2018-214877>.
23. Ferreira-Iglesias A, Montes A, Perez-Pampin E, Cañete JD, Raya E, Magro-Checa C, et al. Evaluation of 12 GWAS-drawn SNPs as biomarkers of rheumatoid arthritis response to TNF inhibitors. A potential SNP association with response to etanercept. *PLoS One.* 2019;14(2):e0213073.
24. Sieberts SK, Zhu F, García-García J, Stahl E, Pratap A, Pandey G, et al. Crowdsourced assessment of common genetic contribution to predicting anti-TNF treatment response in rheumatoid arthritis. *Nat Commun.* 2016; 7(1):12460. <https://doi.org/10.1038/ncomms12460>.
25. Guan Y, Zhang H, Quang D, Wang Z, Parker SCJ, Pappas DA, et al. Machine learning to predict anti-tumor necrosis factor drug responses of rheumatoid arthritis patients by integrating clinical and genetic markers. *Arthritis Rheumatol.* 2019;71(12):1987–96. <https://doi.org/10.1002/art.41056>.
26. Nair N, Wilson AG. Can machine learning predict responses to TNF inhibitors? *Nat Rev Rheumatol.* 2019;15(12):702–4. <https://doi.org/10.1038/s41584-019-0320-9>.
27. Ota M, Nagafuchi Y, Hatano H, Ishigaki K, Terao C, Takeshima Y, et al. Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases. *Cell.* 2021;184(11):3006–3021.e17. <https://doi.org/10.1016/j.cell.2021.03.056>.
28. Nygaard G, Firestein GS. Restoring synovial homeostasis in rheumatoid arthritis by targeting fibroblast-like synoviocytes. *Nat Rev Rheumatol.* 2020; 16(6):316–33. <https://doi.org/10.1038/s41584-020-0413-5>.
29. Tsuchiya H, Ota M, Sumitomo S, Ishigaki K, Suzuki A, Sakata T, et al. Parsing multiomics landscape of activated synovial fibroblasts highlights drug targets linked to genetic risk of rheumatoid arthritis. *Ann Rheum Dis.* 2020; 80(4):440–50. <https://doi.org/10.1136/annrheumdis-2020-218189>. Epub ahead of print.
30. Lewis MJ, Barnes MR, Blighe K, Goldmann K, Rana S, Hackney JA, et al. Molecular portraits of early rheumatoid arthritis identify clinical and treatment response phenotypes. *Cell Rep.* 2019;28(9):2455–2470.e2455.
31. Nerviani A, Di Cicco M, Mahto A, Lliso-Ribera G, Rivellese F, Thorborn G, et al. A pauci-immune synovial pathotype predicts inadequate response to TNF α -blockade in rheumatoid arthritis patients. *Front Immunol.* 2020;11:845. <https://doi.org/10.3389/fimmu.2020.00845>.
32. Humby F, Durez P, Buch MH, Lewis MJ, Rizvi H, Rivellese F, et al. Rituximab versus tocilizumab in anti-TNF inadequate responder patients with rheumatoid arthritis (R4RA): 16-week outcomes of a stratified, biopsy-driven, multicentre, open-label, phase 4 randomised controlled trial. *Lancet.* 2021; 397(10271):305–17. [https://doi.org/10.1016/S0140-6736\(20\)32341-2](https://doi.org/10.1016/S0140-6736(20)32341-2).
33. Chatterjee N, Shi J, García-Closas M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nat Rev Genet.* 2016; 17(7):392–406. <https://doi.org/10.1038/nrg.2016.27>.
34. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet.* 2018;50(9):1219–24. <https://doi.org/10.1038/s41588-018-0183-z>.
35. Damask A, Steg PG, Schwartz GG, Szarek M, Hagström E, Badimon L, et al. Patients with high genome-wide polygenic risk scores for coronary artery disease may receive greater clinical benefit from alirocumab treatment in the ODYSSEY OUTCOMES trial. *Circulation.* 2020;141(8):624–36. <https://doi.org/10.1161/CIRCULATIONAHA.119.044434>.
36. Dudbridge F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet.* 2013;9(3):e1003348. <https://doi.org/10.1371/journal.pgen.1003348>.
37. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature.* 2012;491(7422):119–24. <https://doi.org/10.1038/nature11582>.
38. Liu JZ, van Sommeren S, Huang H, Ng SC, Alberts R, Takahashi A, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel

- disease and highlight shared genetic risk across populations. *Nat Genet.* 2015;47(9):979–86. <https://doi.org/10.1038/ng.3359>.
39. Gattler K, Levantovsky R, Moscati A, Giri M, Wu Y, Hsu NY, et al. Common and rare variant prediction and penetrance of IBD in a large, multi-ethnic, health system-based biobank cohort. *Gastroenterol.* 2021;160(5):1546–57. <https://doi.org/10.1053/j.gastro.2020.12.034>.
 40. Tragnone A, Valpiani D, Miglio F, Elmi G, Bazzocchi G, Pipitone E, et al. Dietary habits as risk factors for inflammatory bowel disease. *Eur J Gastroenterol Hepatol.* 1995;7(1):47–51.
 41. Principi M, Losurdo G, Iannone A, Contaldo A, Deflorio V, Ranaldo N, et al. Differences in dietary habits between patients with inflammatory bowel disease in clinical remission and a healthy population. *Ann Gastroenterol.* 2018;31(4):469–73. <https://doi.org/10.20524/aog.2018.0273>.
 42. Knevel R, le Cessie S, Terao CC, Slowikowski K, Cui J, Huizinga TWJ, et al. Using genetics to prioritize diagnoses for rheumatology outpatients with inflammatory arthritis. *Sci Transl Med.* 2020;12(545):eaay1548. <https://doi.org/10.1126/scitranslmed.aay1548>.
 43. Zhao M, Zhou Y, Zhu B, Wan M, Jiang T, Tan Q, et al. IFI44L promoter methylation as a blood biomarker for systemic lupus erythematosus. *Ann Rheum Dis.* 2016;75(11):1998–2006. <https://doi.org/10.1136/annrheumdis-2015-208410>.
 44. Márquez-Luna C, Loh PR, Price AL. Multiethnic polygenic risk scores improve risk prediction in diverse populations. *Genet Epidemiol.* 2017;41(8):811–23. <https://doi.org/10.1002/gepi.22083>.
 45. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet.* 2019;51(4):584–91. <https://doi.org/10.1038/s41588-019-0379-x>.
 46. Amariuta T, Ishigaki K, Sugishita H, Ohta T, Koido M, Dey KK, et al. Improving the trans-ancestry portability of polygenic risk scores by prioritizing variants in predicted cell-type-specific regulatory elements. *Nat Genet.* 2020;52(12):1346–54. <https://doi.org/10.1038/s41588-020-00740-8>.
 47. Cavazos TB, Witte JS. Inclusion of variants discovered from diverse populations improves polygenic risk score transferability. *HGG Adv.* 2021; 2(1):100017. <https://doi.org/10.1016/j.xhgg.2020.100017>.
 48. Duncan L, Shen H, Gelaye B, Meijssen J, Ressler K, Feldman M, et al. Analysis of polygenic risk score usage and performance in diverse human populations. *Nat Commun.* 2019;10(1):3328. <https://doi.org/10.1038/s41467-019-11112-0>.
 49. ENCODE Project Consortium, Moore JE, Purcaro MJ, Pratt HE, Epstein CB, Shores N, et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature.* 2020;583(7818):699–710.
 50. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilienky M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. *Nature.* 2015;518(7539):317–30.
 51. Lappalainen T, Sammeth M, Friedländer MR, t Hoen PA, Monlong J, Rivas MA, et al. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature.* 2013;501(7468):506–11. <https://doi.org/10.1038/nature12531>.
 52. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science.* 2020;369(6509):1318–30. <https://doi.org/10.1126/science.aaz1776>.
 53. Schmiedel BJ, Singh D, Madrigal A, Valdovino-Gonzalez AG, White BM, Zapardiel-Gonzalo J, et al. Impact of genetic polymorphisms on human immune cell gene expression. *Cell.* 2018;175(6):1701–1715.e1716.
 54. Vösa U, Claringbould A, Westra H-J, Bonder MJ, Deelen P, Zeng B, et al. Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. *bioRxiv.* 2018:447367. <https://doi.org/10.1101/447367>.
 55. Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BW, et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet.* 2016;48(3):245–52. <https://doi.org/10.1038/ng.3506>.
 56. Mancuso N, Shi H, Goddard P, Kichaev G, Gusev A, Pasaniuc B. Integrating gene expression with summary association statistics to identify genes associated with 30 complex traits. *Am J Hum Genet.* 2017;100(3):473–87. <https://doi.org/10.1016/j.ajhg.2017.01.031>.
 57. Ishigaki K, Kochi Y, Suzuki A, Tsuchida Y, Tsuchiya H, Sumitomo S, et al. Polygenic burdens on cell-specific pathways underlie the risk of rheumatoid arthritis. *Nat Genet.* 2017;49(7):1120–5. <https://doi.org/10.1038/ng.3885>.
 58. Marigorta UM, Denson LA, Hyams JS, Mondal K, Prince J, Walters TD, et al. Transcriptional risk scores link GWAS to eQTLs and predict complications in Crohn's disease. *Nat Genet.* 2017;49(10):1517–21. <https://doi.org/10.1038/ng.3936>.
 59. Lee JC, Basci D, Roberts R, Geary RB, Mansfield JC, Ahmad T, et al. Genome-wide association study identifies distinct genetic contributions to prognosis and susceptibility in Crohn's disease. *Nat Genet.* 2017;49(2):262–8. <https://doi.org/10.1038/ng.3755>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

