



OPEN ACCESS

Novel Insights Into Rheumatoid Arthritis Through Characterization of Concordant Changes in DNA Methylation and Gene Expression in Synovial Biopsies of Patients With Differing Numbers of Swollen Joints

Edited by:

Aridaman Pandit,
University Medical Center Utrecht,
Netherlands

Reviewed by:

W Tao,
University Medical Center Utrecht,
Netherlands

Saikat Chowdhury,
University of Texas MD Anderson
Cancer Center, United States

***Correspondence:**

Rab K. Prinjha
rabinder.prinjha@gsk.com
Paul P. Tak
tak.paulpeter@gmail.com

†Present Address:

Enrico Ferrero,
Autoimmunity Transplantation and
Inflammation Bioinformatics, Novartis
Institutes for BioMedical Research,
Basel, Switzerland;
Paul P. Tak,
Candel Therapeutics, Needham,
MA, United States

†These authors share first authorship

§These authors share last authorship

Specialty section:

This article was submitted to
Molecular Innate Immunity,
a section of the journal
Frontiers in Immunology

Received: 09 January 2021

Accepted: 25 March 2021

Published: 22 April 2021

Citation:

Li Yim AYF, Ferrero E, Maratou K,
Lewis HD, Royal G, Tough DF,
Larminie C, Mannens MMAM,
Henneman P, de Jonge WJ,
van de Sande MGH, Gerlag DM,
Prinjha RK and Tak PP (2021) Novel
Insights Into Rheumatoid Arthritis Through
Characterization of Concordant Changes
in DNA Methylation and Gene Expression
in Synovial Biopsies of Patients With
Differing Numbers of Swollen Joints.
Front. Immunol. 12:651475.
doi: 10.3389/fimmu.2021.651475

Andrew Y. F. Li Yim^{1,2†}, Enrico Ferrero^{1†‡}, Klio Maratou¹, Huw D. Lewis¹, George Royal¹,
David F. Tough¹, Chris Larminie¹, Marcel M. A. M. Mannens², Peter Henneman²,
Wouter J. de Jonge^{3,4}, Marleen G. H. van de Sande^{5,6}, Danielle M. Gerlag¹,
Rab K. Prinjha^{1*§} and Paul P. Tak^{1,5,7,8*†§}

¹ R&D GlaxoSmithKline, Stevenage, United Kingdom, ² Department of Clinical Genetics, Genome Diagnostics Laboratory, Amsterdam Reproduction & Development, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, Netherlands, ³ Tytgat Institute for Liver and Intestinal Research, Amsterdam Gastroenterology & Metabolism, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, Netherlands, ⁴ Department of Surgery, University Clinic of Bonn, Bonn, Germany, ⁵ Department of Rheumatology and Clinical Immunology, Amsterdam Rheumatology and Immunology Center, Amsterdam Institute for Infection & Immunity, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, Netherlands, ⁶ Department of Experimental Immunology, Amsterdam Institute for Infection & Immunity, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, Netherlands, ⁷ Department of Rheumatology, Ghent University, Ghent, Belgium, ⁸ Department of Medicine, University of Cambridge, Cambridge, United Kingdom

In this study, we sought to characterize synovial tissue obtained from individuals with arthralgia and disease-specific auto-antibodies and patients with established rheumatoid arthritis (RA), by applying an integrative multi-omics approach where we investigated differences at the level of DNA methylation and gene expression in relation to disease pathogenesis. We performed concurrent whole-genome bisulphite sequencing and RNA-Sequencing on synovial tissue obtained from the knee and ankle from 4 auto-antibody positive arthralgia patients and thirteen RA patients. Through multi-omics factor analysis we observed that the latent factor explaining the variance in gene expression and DNA methylation was associated with Swollen Joint Count 66 (SJC66), with patients with SJC66 of 9 or more displaying separation from the rest. Interrogating these observed differences revealed activation of the immune response as well as dysregulation of cell adhesion pathways at the level of both DNA methylation and gene expression. We observed differences for 59 genes in particular at the level of both transcript expression and DNA methylation. Our results highlight the utility of genome-wide multi-omics profiling of synovial samples for improved understanding of changes associated with disease spread in arthralgia and RA patients, and point to novel candidate targets for the treatment of the disease.

Keywords: rheumatoid arthritis, arthralgia, DNA methylation, transcriptomics, multi-omics analyses, synovial biopsies, target identification

INTRODUCTION

Rheumatoid arthritis (RA) is a complex, multifactorial, and chronic autoimmune disease that primarily affects the synovial tissue in joints (1). It affects about 1% of the population and manifests with significant unmet medical need (2). Investigating pathogenesis during different stages of disease, or across a spectrum of disease severity, is critical to optimize appropriate therapeutic interventions that affect disease progression (3, 4).

Diagnosis of established RA coincides with the development of painful and swollen joints, although circulating auto-antibodies can be detected up to 10 years before diagnosis (5, 6). Clinically manifested joint swelling reflects synovial tissue inflammation (synovitis), which is characterized by infiltration into the synovium of multiple immune cell types [including T cells, B cells, macrophages, plasma cells, dendritic cells, natural killer (NK) cells, and mast cells (7)], with up to 18 distinct infiltrating cell populations being reported on in a recent single cell transcriptomics analysis (8). As disease progresses, synovial fibroblasts adopt an increasingly aggressive and invasive phenotype, promoting further inflammation and joint damage together with other processes induced by the inflammatory environment, such as the differentiation of bone-resorbing osteoclasts (9, 10). Disease progression in early RA is often associated with the involvement of an increasing number of inflamed joints, but the mechanisms responsible for this spread of disease are poorly understood. Moreover, differences in the rate of disease manifestation and variability of response to therapy indicate that different pathophysiological mechanisms are implicated in disease development and progression compared to disease etiology (11).

An increasing body of evidence indicates that epigenetic modifications play an important role in the regulation of RA pathogenesis (12). Several array-based studies have reported widespread differences in DNA methylation among peripheral blood cells from RA patients and controls, suggesting that epigenetic modifications in circulating cells associate with disease (13, 14). However, wider conclusions may be limited by the unknown correlation of these effects to synovial cells directly at the site of inflammation. Epigenetic modifications have also been implicated in modulating the function of synovial fibroblasts in RA, through comparisons of DNA methylation patterns of cultured cells isolated from RA and osteoarthritis patients (15–20). Such studies have identified DNA methylation patterns that distinguish RA from other forms of arthritis, along with regulatory elements and biomarkers related to the pathological phenotype of RA. However, to identify novel candidate genes for therapeutic interventions that affect disease progression, it is important to study samples from patients across different stages of disease. Two such studies using cultured fibroblast-like synoviocytes found small but statistically significant patterns of hypomethylation in patients with longstanding RA compared to those with early RA, suggesting that the DNA methylome could be associated with the transformation of synovial fibroblasts into invasive cells capable of joint destruction and the resulting disease progression (21, 22).

Here, we gain insights into RA heterogeneity by combining whole-genome DNA methylome and transcriptome analyses from the same synovial biopsies and stratifying patients according to the number of swollen joints, a clinical parameter reflective of disease evolution. We find that swollen joint count based on 66 joints (SJC66), which reflects the amount and spread of inflamed synovial tissue, is associated with major changes in gene transcription and DNA methylation at promoters. By cross-interrogating differentially methylated genomic regions and their associated genes, we reveal novel candidate loci associated with the spread of the disease across joints.

MATERIALS AND METHODS

Sample Description

Synovial tissue was collected *via* a mini-arthroscopic procedure as described previously (23) from patients at the Amsterdam University Medical Center, University of Amsterdam. A total of 17 samples were obtained from three cohorts. The first cohort contained individuals that had either arthralgia and/or a positive family history for RA, but without arthritis (as determined by a clinician), and that were positive for IgM Rheumatoid Factor (IgM-RF) and/or Anti-Citrullinated Protein Antibodies (ACPA) (Pre-synoviomics; $n = 4$) (24). The second cohort consisted of individuals that at inclusion were Disease Modifying Anti-Rheumatic Drug (DMARD)-naïve with early arthritis, as defined by a disease duration of less than 1 year (Synoviomics; $n = 9$) (25, 26). The third cohort contained samples from patients with established RA on active treatment with a disease duration of more than one year who had at least one swollen joint suitable for synovial tissue sampling (Synoviomics II; $n = 4$) (27). For all analyses, samples were prepared simultaneously to mitigate batch effects.

All subjects provided written informed consent and the collection and use of the samples received Institutional Review Board review and approval. Characteristics of patients included in this study are listed in **Table S1**.

RNA-Sequencing and Whole Genome Bisulphite Sequencing

Flash-frozen synovial tissue biopsies were utilized to simultaneously isolate RNA and DNA using an AllPrep DNA/RNA Mini kit (Qiagen), with QIAshredder spin columns (Qiagen) used to disrupt the tissue. RNA samples were quantified and their integrity assessed using Qubit RNA Broad Range Assay Kit (Thermo Fisher Scientific) and an Agilent 2100 Bioanalyzer RNA 6000 Nano Kit (Agilent Technologies), respectively. Depending on sample yield, DNA samples were quantified using Qubit DNA BR or Qubit DNA HS kits (Thermo Fisher Scientific).

RNA-Seq libraries were generated from 150 ng of total RNA. The TruSeq® Stranded Total RNA LT was used with a RiboZero™ Human/Mouse/Rat kit (Illumina), following the 'Low Sample' protocol except for two modifications. Firstly, the time of the 'Elution 2 – Frag – Prime' program was reduced from 8 to

6 min to increase the length of the RNA fragments. Secondly, 11 instead of 15 cycles were used to enrich the DNA fragments. Libraries were quantified with a KAPA Library Quantification Kit (KAPA Biosystems) on a QuantStudio 12 K Flex Real-Time PCR System (Thermo Fisher Scientific). Five samples were multiplexed per lane and libraries were clustered and sequenced using HiSeq[®] PE Cluster Kit v3 – cBot[™] and HiSeq[®] SBS Kit v3 kits (Illumina). Paired-end sequencing (2 x 76 bp) was performed using a HiSeq 1500 (Illumina) to a depth of ~40 M reads per sample.

Whole Genome Bisulphite Sequencing (WGBS) libraries were generated using an EpiGnome Methyl-Seq Library Preparation Kit (Epicentre, now Illumina) from 100 ng of sample DNA. Bisulphite conversion was performed using an EZ DNA Methylation-Lightning Kit (Zymo Research). Each bisulphite conversion reaction contained 500 pg of unmethylated lambda DNA (Promega), which was used as a control to verify that bisulphite conversion efficiencies were at least 98%. Libraries were quantified using a KAPA Library Quantification Kit (KAPA Biosystems) on a 7900HT Real-Time PCR System (Thermo Fisher Scientific). Six samples were multiplexed per lane and libraries were clustered and sequenced using HiSeq[®] PE Cluster Kit v4 – cBot[™] and HiSeq[®] SBS Kit v4 kits (Illumina). Paired-end sequencing (2 x 125 bp) was performed using a HiSeq 1500 (Illumina) to a depth of ~600M reads per sample. To provide sufficient coverage each batch was sequenced over 2 high output runs.

Exploratory Analyses of DNA Methylation and Gene Expression

To concurrently explore the DNA methylation and gene expression data, Multi-Omics Factor Analysis (MOFA) was applied (v1.0.0) (28). In short, MOFA performs unsupervised dimensionality reduction simultaneously across multiple data modalities from the same sample through a small number of inferred latent factors, enabling the detection of co-ordinated changes between the different data modalities (28). Here, we used the 5,000 most variable CpGs and genes as input and with the number of latent factors set to 9, the tolerance to 0.1 and the factor threshold to 0.02.

Gene Expression Data Analysis

Quality assessment of the raw reads was performed using FastQC (v0.11.2) (29). Adapters were removed from reads using Scythe (v0.991) (30) and sequences were quality-trimmed using Sickle (v1.33) (31) using a quality threshold of 20. Alignment to the GRCh38 build of the human genome was performed using Kallisto (v0.44.0) (32).

All subsequent analyses were performed in R (v3.5.0) (33). Gene-level counts were generated from the transcript abundances using tximport (v1.12.0) (34). Allosome-associated genes were removed to mitigate obvious sex effects. Differential Gene Expression (DGE) analyses were conducted using DESeq2 (v1.22.2) (35) where SJC66_{high} was compared with SJC66_{low} whilst correcting for sex, age and DMARD usage using the following design formula:

$$\text{Gene Expression} \sim \text{sex} + \text{age} + \text{DMARD usage} + \text{SJC66}_{\text{dichotomized}}$$

DMARD usage was a binary variable defined by the usage of any medication: conventional DMARDs (cDMARD), including methotrexate) or biological DMARDs (bDMARD). As Kallisto provided abundance levels for individual transcripts, Gene Differential Expression (GDE) analyses were also conducted to identify genes where particular transcripts were differentially expressed (36). In short, differential expression analyses was performed using DESeq2 and the resulting *p*-values were combined using the Lancaster aggregation method found in aggregation (v1.0.1) (36), where observations were weighted by the base expression. DGEs and GDEs were defined as genes with a false discovery rate (FDR)-adjusted *p*-value less than 0.05.

DNA Methylation Data Analysis

Quality assessment of the raw reads was performed using FastQC (v0.11.2) (29). Adapter and quality trimming was performed using Skewer (v0.1.123) (37) and a quality filter of 20. To assess bisulphite conversion rates, Bismark (v0.14.1) (38) was used to align the reads to the genome of the phage lambda, and again for alignment to the GRCh38 build of the human genome. Post-alignment filtering of unmapped reads, reads aligning at multiple locations and reads with a mapping score lower than 10 was carried out using SAMtools (39).

All subsequent analyses were performed in R (v3.5.0). CpG loci located on the allosomes were removed to mitigate the sex effect. The differential methylation analyses were performed using dmrseq (v1.2.5) (40), where we contrasted SJC66_{high} with SJC66_{low} whilst correcting for sex, age and DMARD usage using the following design formula:

$$\text{Methylation} \sim \text{sex} + \text{age} + \text{DMARD usage} + \text{SJC66}_{\text{dichotomized}}$$

Differentially Methylated Regions (DMRs) were annotated using ChipPeakAnno (v3.16.1) (41) to genes if the DMR was located within 5,000 bp upstream or 1,000 bp downstream of the gene as obtained from Gencode (v29) (42).

Integrated DNA Methylation and Gene Expression Analysis

Integrated analyses were based on the DMRs and GDEs found through the separate DNA methylation and gene expression analyses. The overlap between DMRs and GDEs were called Genes displaying both Differential Expression and Methylation (GDEMs). For each GDEM the median percentage methylation was calculated for all constituent CpGs per sample and correlated with the log₂ transformed expression counts to obtain the Pearson correlation coefficient. Confidence intervals (95%) were calculated through 1000 bootstraps for each GDEM. In short, 17 samples were randomly drawn from the original samples with replacement, whereupon the Pearson correlation coefficient was calculated. This process was repeated 1000 times to generate the empirical distribution function, which was then used to estimate the confidence intervals (43). The aforementioned bootstrapping approach was performed using the boot (v1.3) package (44). For inferential purposes, *p*-values

were calculated by means of a permutation approach. In short, per GDEM, 1000 sets of consecutive CpGs equal to the length of the observed DMR were sampled and correlated with the gene expression signal as described above, after which the proportion of correlation coefficients higher than the observed correlation coefficient was calculated yielding the *p*-value.

Functional Enrichment, Cell Type Enrichment and Protein-Protein Interaction Network Analyses

Gene set enrichment analyses were performed using fgsea (v1.8.0) (45) using the Metabase pathways terms (46) as reference. Metabase pathway terms with an FDR-adjusted *p*-value less than 0.05 were considered significant.

Cell proportions were imputed using xCell (v1.1.0) (47), where transcripts per million were used to estimate the proportion of each of the 64 immune and stromal cell types. Subsequent linear regressions were performed to calculate the *p*-values to assess statistical significance. Again, we compared SJC66_{high} with SJC66_{low} also correcting for age, sex and DMARD usage.

Protein-protein interaction (PPI) network analyses were performed using the STRING (v11) database (48), in order to identify whether a set of genes was over-represented for interactions. In short, the PPI analysis returned networks of genes where the encoded proteins interacted, co-expressed or co-evolved with one another, based on text mining, curated databases, and experimental data.

Data Availability

The datasets generated and analyzed for this study can be found in the ArrayExpress repository under accession number E-MTAB-6638 and E-MTAB-6684 for WGBS and RNA-Seq, respectively. All code is hosted on GitHub at <https://github.com/enricoferrero/BTCURE>.

RESULTS

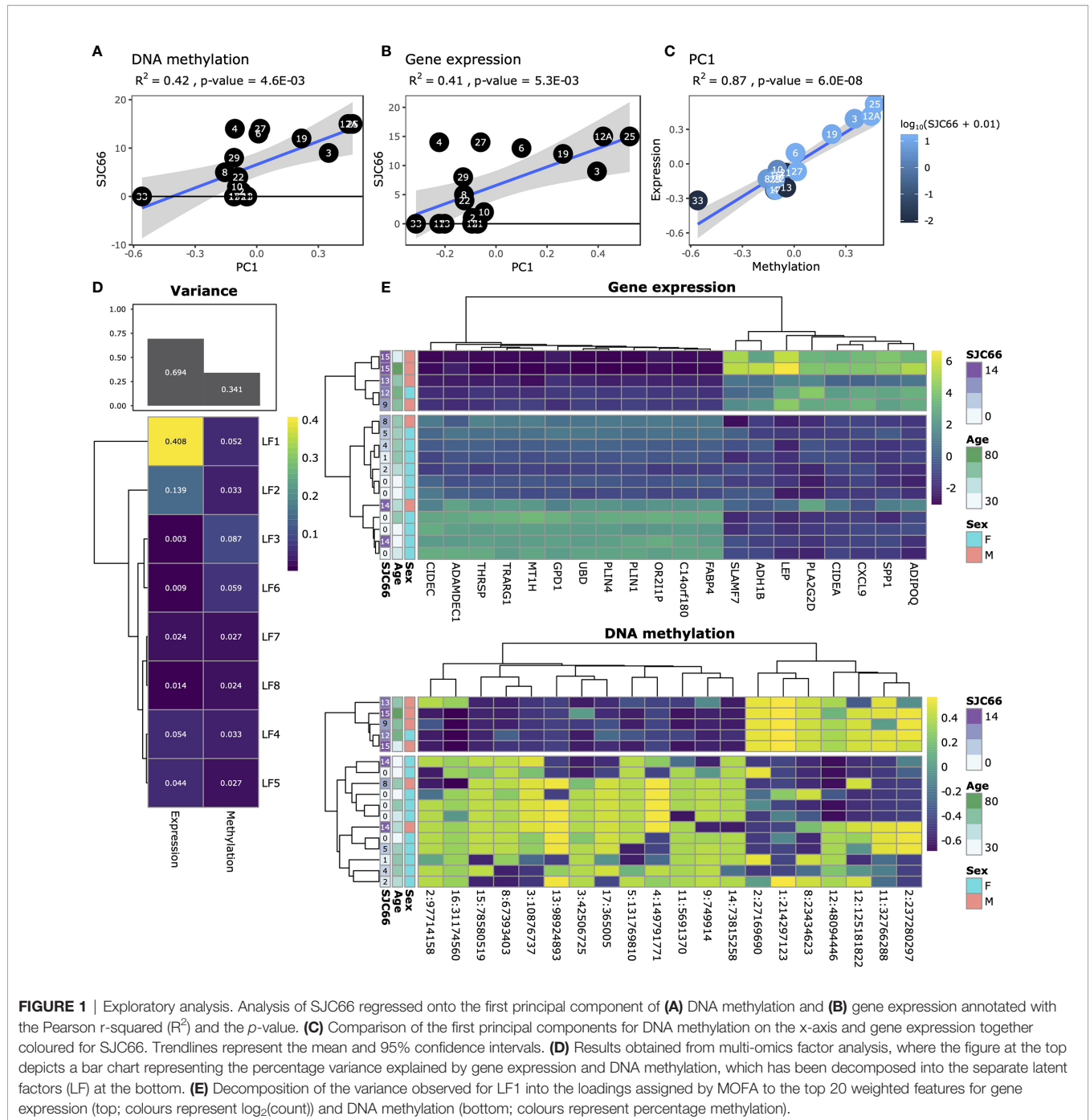
Swollen Joint Count Is Associated With the Latent Factor Explaining Variance in Gene Expression and DNA Methylation

We profiled the DNA methylome and transcriptome of 17 synovial tissue samples (Table S1) using whole genome bisulphite sequencing (WGBS) and RNA-sequencing (RNA-Seq). We initially attempted to link gene expression changes to disease duration as well as to cross-patient variations in the Disease Activity Score of 28 joints (DAS28) variables, but these analyses resulted in weak and non-biologically relevant signals, and were ultimately deemed inconclusive for this set of samples (data not shown). Principal Component Analysis (PCA) indicated that variation in both DNA methylation and gene expression were independently correlated with the swollen joint count in 66 joints (SJC66; Figures 1A, B). Comparison of the first principal component of both the DNA methylation and gene expression data suggested agreement with samples 33, 19, 3,

12A and 25 broadly separating from the other samples for both modalities (Figure 1C). While sample 33 appeared to be an outlier based on the DNA methylation data the removal thereof did not alter the correlation substantially (Figure S1). To further explore DNA methylation and gene expression at a genome-wide level in an integrative fashion, we performed variance decomposition using multi-omics factor analysis (MOFA) (28). MOFA infers a set of latent factors that capture sources of variability across different measured -omic modalities of the same samples. We found that most variance (approx. 70%) was better explained by gene expression as compared to DNA methylation (Figure 1D). Further decomposition of the variance identified 8 latent factors, with LF1 explaining 40% of the variance in gene expression, whereas variance in methylation was more evenly distributed amongst all eight latent factors. Focusing on LF1, we observed a marked separation between samples with SJC66 of 9 and above compared to those with SJC66 of 8 or less (Figures 1E and S2A, B). While most samples were obtained from the knee, two were obtained from the ankle. Interrogation of the first latent factor did not indicate any correlation with the source of the sample (Figure S2C). Importantly, sex was also strongly associated with LF1 (*p*-value = 8.8E-03; Figure S2D) where samples with SJC66 of 9 and above were mostly males and most samples with SJC66 equal to or fewer than 8 were mostly females. Therefore, subsequent comparative analyses accounted for the imbalance in sex by removal of allosome-associated genes as well as including sex as covariate in linear models. females. Therefore, subsequent comparative analyses accounted for the imbalance in sex by removal of allosome-associated genes as well as including sex as covariate in linear models.

Differences in DNA Methylation and Gene Expression in Patients With High and Lower Numbers of Swollen Joints Are Associated With Immune Response and Cell Adhesion Pathways

Having observed genome-wide differences through exploratory analyses in both DNA methylation and gene expression data that associated with SJC66, we next investigated which regions and genes were differentially methylated and expressed. At this point we investigated DNA methylation and gene expression separately. While samples with SJC66 = 0 were included for the aforementioned exploratory analyses, they were excluded from subsequent comparative analyses as they were medically not a homogeneous group consisting of very early arthritis as well as late arthritis with no swelling following medication. The swollen samples were stratified according to the separation observed in the exploratory analyses (Figure 1E) and we compared SJC66_{high} (SJC66 ≥ 9) with SJC66_{low} (SJC66 < 8) correcting for age, sex and DMARD usage. Comparative methylation analysis identified 3,536 DMRs (Figure 2A and Table S2), where 2,140 were hypomethylated and 1,396 were hypermethylated. The majority of DMRs were located within 1 Kb of a TSS or in distal regions (Figure 2B). Notably, the most statistically significant DMRs spanned regions larger than 10 Kb (Table S2).



For example, the top two DMRs were located within the promoter regions of the *MIRLET7B* and *MIR3619* genes (*MIRLET7BHG*; 22:46,072,724-46,090,732; **Figure 2C**) and microRNA 10B (*MIR10B*; 2:176,147,225-176,162,267; **Figure 2D**) and exceeded 15 Kb in length. While the *MIRLET7BHG*-associated DMR was hypermethylated, the *MIR10B*-associated DMR was hypomethylated when comparing SJC66_{high} with SJC66_{low}. Functional analysis of the DMRs identified several pathways with evidence of differential methylation. Among the top 10, we observed that the NGF/TrkA MAPK pathway,

immune response pathways including antigen presentation by MHC class II, macrophage and dendritic cell phenotype shifts, as well as CCR3 signalling in eosinophils, were hypermethylated (**Figure 2E** and **Table S3**).

Comparing SJC66_{high} with SJC66_{low} at the gene expression level identified 142 DGEs, of which 106 up-regulated and 36 down-regulated (**Figure 3A** and **Table S4**). The top 2 DGEs, chemokine ligand 13 (*CCL13*) (**Figure 3B**) and C-Type Lectin Domain Containing 10A (*CLEC10A*) (**Figure 3C**), were both found to be more highly expressed in the SJC66_{high} samples.

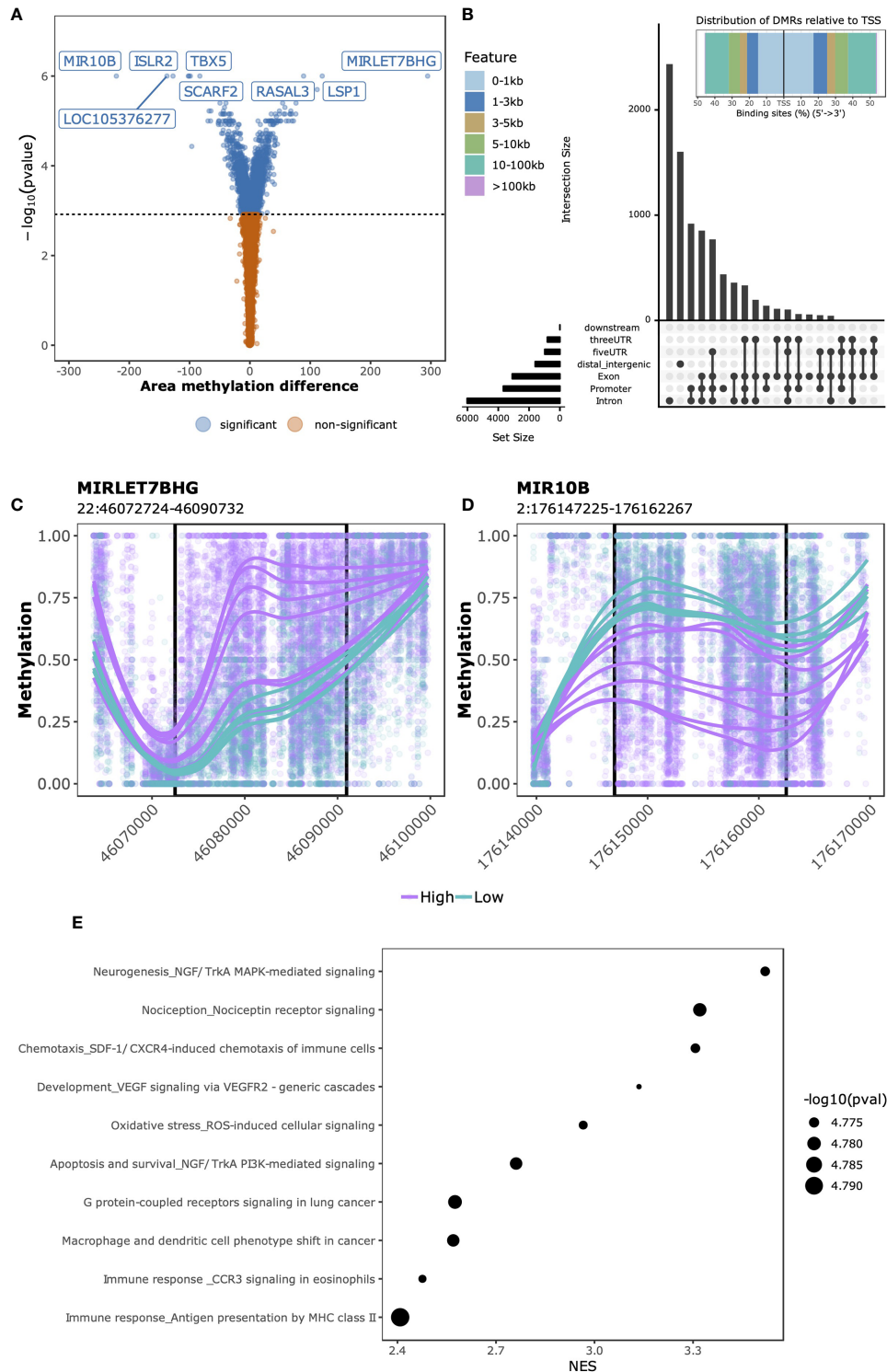
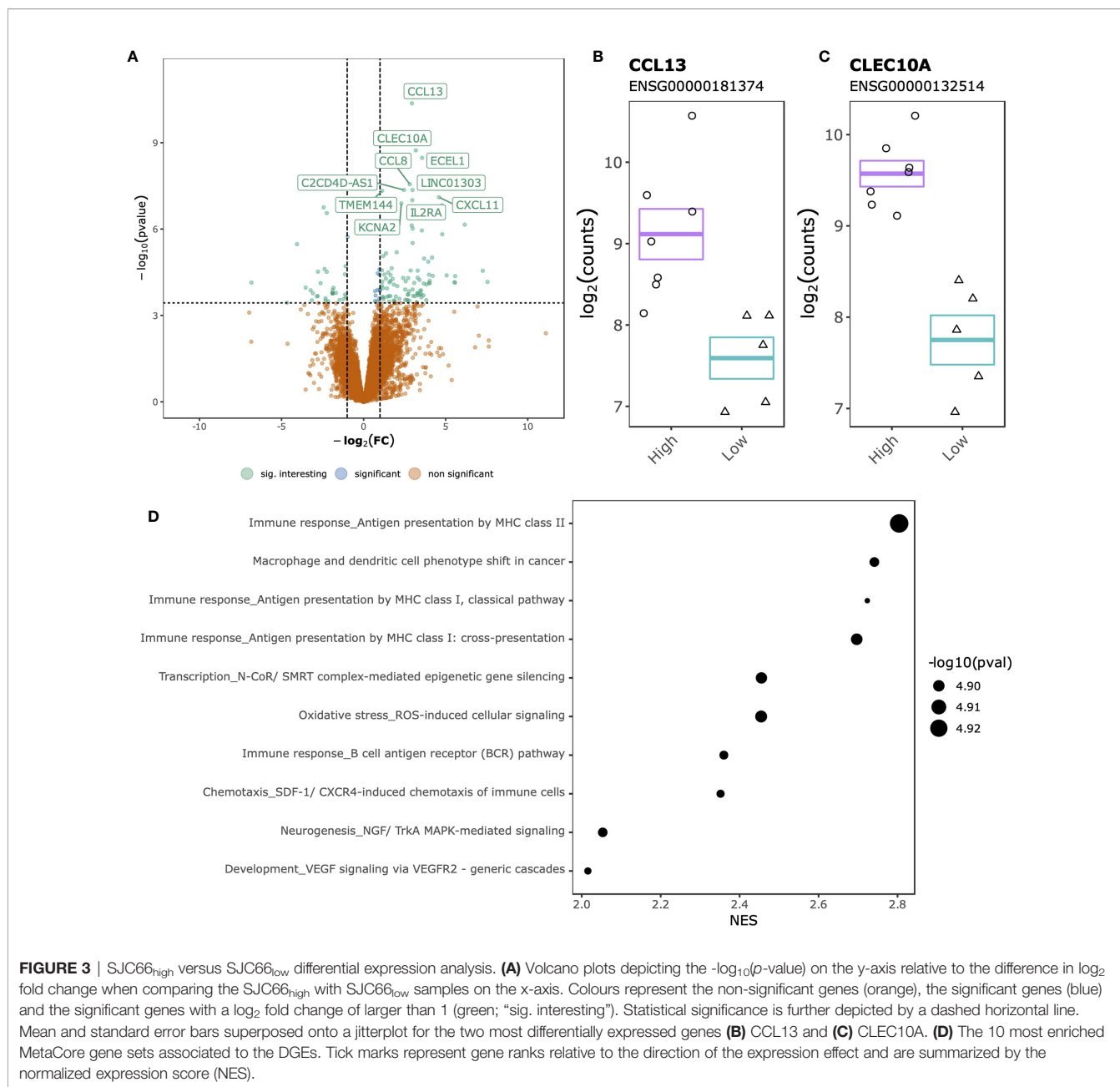


FIGURE 2 | SJC66_{high} versus SJC66_{low} differential methylation analysis. **(A)** Volcano plots depicting the $-\log_{10}(p\text{-value})$ on the y-axis relative to the difference in % methylation when comparing the SJC66_{high} with SJC66_{low} samples on the x-axis. Colours represent the non-significant (orange) and significant regions (blue). Statistical significance is further depicted by a dashed horizontal line. **(B)** At the top: distribution of the DMRs relative to the transcription start site (TSS) and at the bottom: upset plot indicating the distribution of the DMRs across the different genetic features. **(C, D)** DMRs found at genomic co-ordinates **(C)** 22:46,072,724-46,090,732 and **(D)** 2:176,147,225-176,162,267 are located in the promoters of MIRLET7BHG and MIR10B, respectively. DNA methylation is visualized as the percentage methylation (y-axis) with a smoothed trendline per sample. **(E)** The 10 most enriched MetaCore gene sets associated to the promoter-bound DMRs. Tick marks represent gene ranks relative to the direction of the methylation effect and is summarized by the normalized expression score (NES).

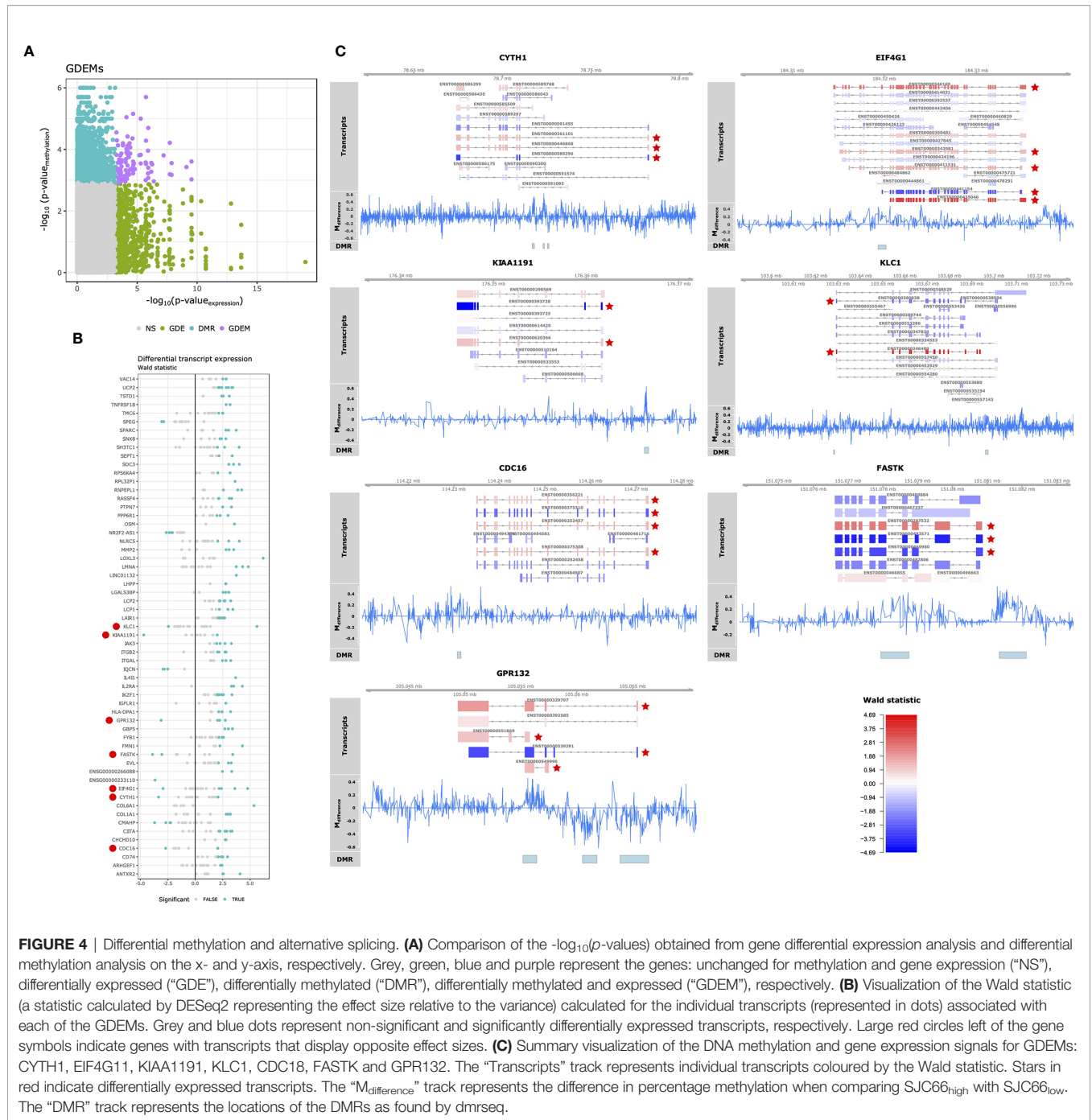


Functional analyses revealed a striking similarity with the differential methylation results, where genes associated with antigen presentation by MHC class II as well as the NGF/TrkA MAPK pathways were overexpressed (**Figure 3D** and **Table S5**).

Transcripts Associated With Differentially Methylated and Expressed Genes Display Concordant Expression and Identify Nodal Points for Key Interactions

Having observed pathway-concordant differences in DNA methylation and gene expression, we were interested in identifying specific genes that were both differentially methylated and expressed. Although DNA methylation has

traditionally been associated with gene-level expression, emerging evidence shows that it also regulates alternative splicing (49–51). Therefore, we complemented our Differential Gene Expression (DGE) analysis, which masks transcript-level dynamics, with Gene Differential Expression (GDE) analysis, which identifies genes that display transcript-level differences, by combining the p -values of individual transcripts associated with a single gene (36). We identified 290 genes that displayed perturbations in the expression of their transcripts (GDEs; **Table S6**). Combining the GDEs with the DMRs yielded 97 unique DMRs associated with 59 unique genes, which we termed Genes Differentially Expressed and Methylated (GDMEs) (**Figures 4A, B**). We observed that most transcripts, those



associated with 52 GDEMs, were concordantly differentially expressed. Transcripts of the remaining 7 GDEMs were found to display opposite directions of expression, namely Cytohesin 1 (*CYTH1*), Eukaryotic Translation Initiation Factor 4 Gamma 1 (*EIF4G1*), a putative monooxygenase (*KIAA1191*), Kinesin Light Chain 1 (*KLC1*), Cell Division Cycle 16 (*CDC16*), Fas-Activated Serine/Threonine Kinase (*FASTK*) and G protein coupled receptor 132 (*GPR132*; **Figure 4B**). Investigation of the methylation status of the gene overlaid onto the transcript location indicated that the observed DMRs for *KIAA1191*,

KLC1, *CDC16*, *FASTK* and *GPR132* were located in the promoter shared by most transcripts (**Figure 4C**). The mechanisms that could cause opposite direction of expression in these genes are currently unclear although studies in Arabidopsis have identified a methylation reader complex that can enhance rather than suppress gene transcription in the presence of methylation (52). Our work therefore supports the rationale for further validation and mechanistic studies.

Functionally, we observed that the 59 GDEMs were primarily over-represented for immune response-associated pathways,

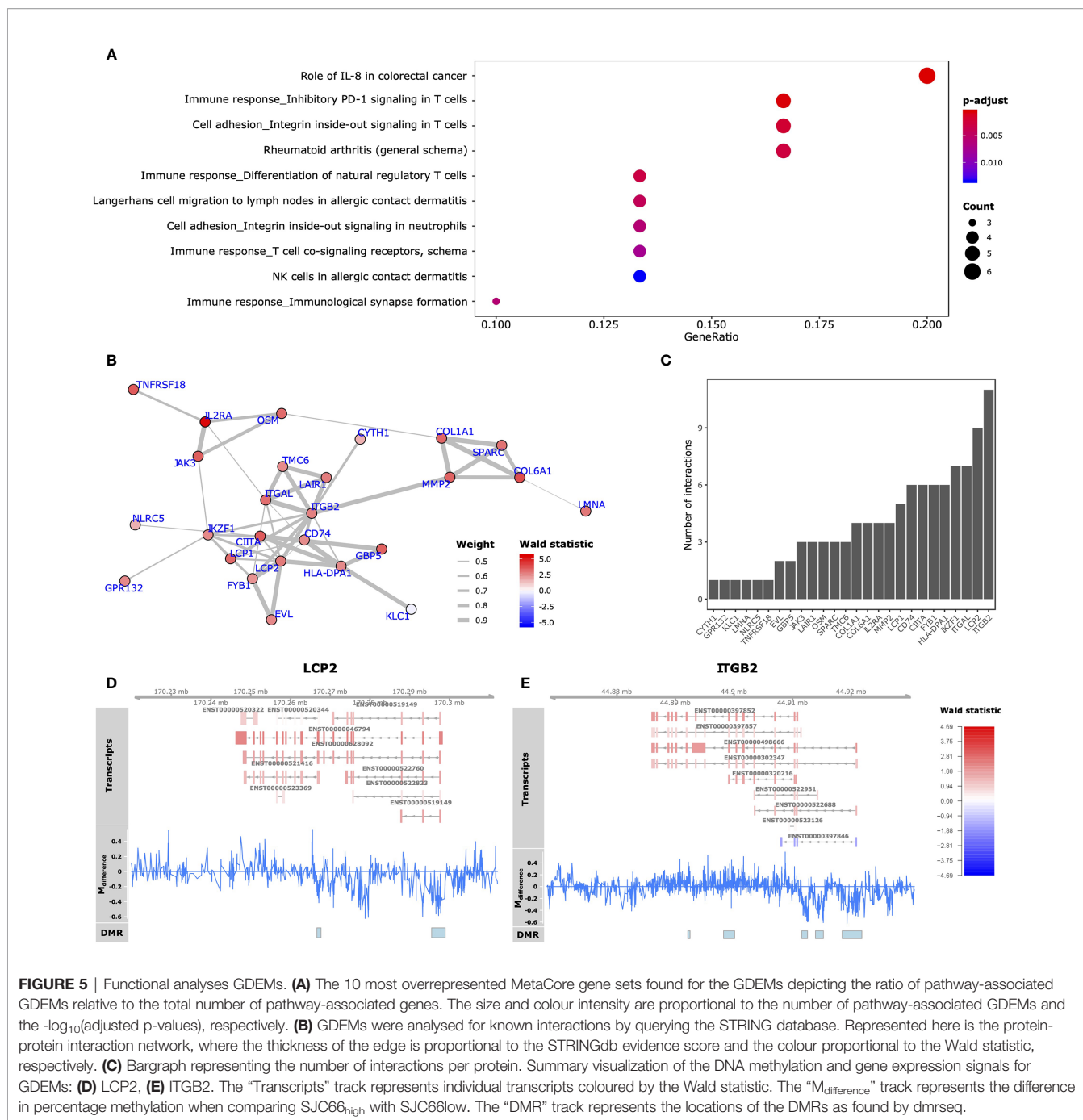


FIGURE 5 | Functional analyses GDEMs. **(A)** The 10 most overrepresented MetaCore gene sets found for the GDEMs depicting the ratio of pathway-associated GDEMs relative to the total number of pathway-associated genes. The size and colour intensity are proportional to the number of pathway-associated GDEMs and the $-\log_{10}(\text{adjusted } p\text{-values})$, respectively. **(B)** GDEMs were analysed for known interactions by querying the STRING database. Represented here is the protein-protein interaction network, where the thickness of the edge is proportional to the STRINGdb evidence score and the colour proportional to the Wald statistic, respectively. **(C)** Bargraph representing the number of interactions per protein. Summary visualization of the DNA methylation and gene expression signals for GDEMs: **(D)** LCP2, **(E)** ITGB2. The “Transcripts” track represents individual transcripts coloured by the Wald statistic. The “ $M_{\text{difference}}$ ” track represents the difference in percentage methylation when comparing SJC66_{high} with SJC66_{low}. The “DMR” track represents the locations of the DMRs as found by dmrseq.

specifically T-lymphocyte associated pathways (**Figure 5A**). We interrogated which of the GDEMs encoded known interaction partners by querying the STRING database for documented interactions (48). Almost half (46%) of the GDEMs encoded for interacting proteins with the most connected GDEMs being *ITGB2* (11 interactions) and *LCP2* (9 interactions) (**Figure 5C**). Both *ITGB2* and *LCP2* were differentially methylated at multiple locations, with the largest visible differences occurring within the TSS and downstream thereof (**Figures 5D, E**). Transcript-wise, *ITGB2*-transcripts ENST00000498666 and ENST00000397852 as

well as *LCP2* transcript ENST0000046794 were most differentially expressed. Expression Quantitative Trait Methylation (eQTM) analyses confirmed strong inverse correlations between the differences in methylation in the promoter regions of *ITGB2* (21:44918461-44921815) and *LCP2* (5:170295513-170298924) with transcripts ENST00000498666 ($r = -0.9$; $p\text{-value} = 1E-04$) and ENST0000046794 ($r = -0.9$; $p\text{-value} = 1E-04$), respectively (**Table 1**). While the association between the *ITGB2* promoter DMR and ENST00000397852 expression was non-significant ($p\text{-value} = 0.2353$), the correlation

TABLE 1 | Differentially methylated and expressed genes.

Gene	GDE _{p-value}	DMR	DMR _{stat}	DMR _{p-value}	EnsT	DTE _{p-value}	eQTM _r	eQTM _{p-value}
ENSG00000233110	2.53E-04	4:185586913-185587794	-9.94	2.78E-04	ENST00000411847.1	2.53E-04	0.88 [0.74;0.95]	2.54E-02
CD74	2.00E-04	5:150410549-150413278	-9.48	3.59E-04	ENST00000009530.11	1.77E-02	-0.88 [-0.96;-0.7]	1.00E-04
CD74	2.00E-04	5:150410549-150413278	-9.48	3.59E-04	ENST00000518797.5	1.23E-02	-0.9 [-0.96;-0.76]	1.00E-04
CIITA	2.93E-10	16:10877103-10878257	-10.23	2.47E-04	ENST00000637439.1	5.06E-03	-0.92 [-0.96;-0.78]	1.00E-04
CIITA	2.93E-10	16:10877103-10878257	-10.23	2.47E-04	ENST00000637439.1	5.06E-03	-0.92 [-0.96;-0.78]	1.00E-04
CIITA	2.93E-10	16:10877103-10878257	-10.23	2.47E-04	ENST00000575513.1	8.19E-04	-0.88 [-0.96;-0.66]	8.10E-03
CIITA	2.93E-10	16:10877103-10878257	-10.23	2.47E-04	ENST00000575513.1	8.19E-04	-0.88 [-0.96;-0.66]	8.10E-03
CIITA	2.93E-10	16:10877103-10878257	-10.23	2.47E-04	ENST00000572665.1	6.67E-03	-0.86 [-0.94;-0.62]	8.80E-03
CIITA	2.93E-10	16:10877103-10878257	-10.23	2.47E-04	ENST00000572665.1	6.67E-03	-0.86 [-0.94;-0.62]	8.80E-03
CIITA	2.93E-10	16:10877103-10878257	-10.23	2.47E-04	ENST00000576601.1	3.12E-02	-0.89 [-0.96;-0.67]	2.51E-02
CIITA	2.93E-10	16:10877103-10878257	-10.23	2.47E-04	ENST00000576601.1	3.12E-02	-0.89 [-0.96;-0.67]	2.51E-02
CIITA	2.93E-10	16:10877103-10878257	-10.23	2.47E-04	ENST00000381835.9	2.59E-02	-0.86 [-0.95;-0.52]	4.06E-02
CIITA	2.93E-10	16:10877103-10878257	-10.23	2.47E-04	ENST00000381835.9	2.59E-02	-0.86 [-0.95;-0.52]	4.06E-02
CMAHP	1.89E-04	6:25232404-25233193	10.46	2.24E-04	ENST00000377993.8	2.12E-04	-0.7 [-0.87;-0.25]	1.00E-04
CMAHP	1.89E-04	6:25232404-25233193	10.46	2.24E-04	ENST00000493257.5	2.08E-02	-0.68 [-0.84;-0.28]	1.00E-04
COL6A1	1.18E-06	21:45978416-45978817	8.14	8.44E-04	ENST00000361866.7	8.83E-08	0.58 [0.07;0.8]	2.20E-03
COL6A1	1.18E-06	21:45983116-45984355	15.46	2.00E-05	ENST00000361866.7	8.83E-08	0.45 [-0.19;0.78]	1.80E-03
CYTH1	1.53E-04	17:78716750-78718002	12.83	7.40E-05	ENST00000589296.5	8.32E-04	-0.42 [-0.78;0.23]	4.12E-02
EIF4G1	4.11E-07	3:184319325-184320192	7.83	1.09E-03	ENST00000342981.8	2.97E-02	0.32 [-0.28;0.72]	1.00E-04
EIF4G1	4.11E-07	3:184319325-184320192	7.83	1.09E-03	ENST00000342981.8	2.97E-02	0.32 [-0.28;0.72]	1.00E-04
EIF4G1	4.11E-07	3:184319325-184320192	7.83	1.09E-03	ENST00000346169.6	3.07E-04	0.68 [0.28;0.9]	1.00E-04
EIF4G1	4.11E-07	3:184319325-184320192	7.83	1.09E-03	ENST00000346169.6	3.07E-04	0.68 [0.28;0.9]	1.00E-04
EIF4G1	4.11E-07	3:184319325-184320192	7.83	1.09E-03	ENST00000411531.5	1.71E-02	0.7 [0.33;0.9]	1.00E-04
EIF4G1	4.11E-07	3:184319325-184320192	7.83	1.09E-03	ENST00000411531.5	1.71E-02	0.7 [0.33;0.9]	1.00E-04
EIF4G1	4.11E-07	3:184319325-184320192	7.83	1.09E-03	ENST00000435046.6	1.78E-06	0.53 [-0.05;0.8]	1.00E-04
EIF4G1	4.11E-07	3:184319325-184320192	7.83	1.09E-03	ENST00000435046.6	1.78E-06	0.53 [-0.05;0.8]	1.00E-04
EVL	2.46E-05	14:100132409-100133139	9.08	4.52E-04	ENST00000556921.1	3.54E-02	0.89 [0.73;0.97]	3.04E-02
FASTK	1.23E-06	7:151081348-151082120	11.88	1.10E-04	ENST00000460980.5	2.32E-03	-0.21 [-0.76;0.52]	1.00E-04
FMN1	1.83E-04	15:33154437-33154950	-7.96	9.88E-04	ENST00000320930.7	2.49E-02	-0.82 [-0.94;-0.59]	1.26E-02
GPR132	3.80E-05	14:105055190-105056414	9.09	4.48E-04	ENST00000329797.7	6.12E-03	0.9 [0.62;0.99]	1.00E-04
GPR132	3.80E-05	14:105063870-105066425	-11.52	1.36E-04	ENST00000549990.1	2.60E-02	-0.38 [-0.72;0.44]	3.57E-02
IKZF1	6.58E-06	7:50305276-50312367	-11.26	1.55E-04	ENST00000471793.1	1.04E-02	-0.8 [-0.92;-0.33]	1.00E-04
IKZF1	6.58E-06	7:50305276-50312367	-11.26	1.55E-04	ENST00000492119.1	3.29E-02	-0.88 [-0.94;-0.7]	1.00E-04
IKZF1	6.58E-06	7:50305276-50312367	-11.26	1.55E-04	ENST00000413698.5	1.30E-02	-0.9 [-0.95;-0.72]	2.55E-02
IL411	2.32E-04	19:49895253-49897628	-7.97	9.82E-04	ENST00000391826.6	2.32E-04	-0.96 [-0.98;-0.74]	1.00E-04
ITGB2	7.75E-05	21:44898213-44900097	8.96	4.89E-04	ENST00000498666.5	5.07E-03	0.91 [0.67;0.98]	1.00E-04
ITGB2	7.75E-05	21:44911555-44912533	-11.65	1.24E-04	ENST00000397852.5	5.15E-03	-0.94 [-0.98;-0.72]	1.00E-04
ITGB2	7.75E-05	21:44911555-44912533	-11.65	1.24E-04	ENST00000498666.5	5.07E-03	-0.93 [-0.98;-0.81]	2.00E-03
ITGB2	7.75E-05	21:44911555-44912533	-11.65	1.24E-04	ENST00000522688.5	3.04E-02	-0.9 [-0.97;-0.63]	3.80E-03
ITGB2	7.75E-05	21:44911555-44912533	-11.65	1.24E-04	ENST00000320216.10	6.78E-03	-0.79 [-0.93;-0.57]	4.08E-02
ITGB2	7.75E-05	21:44911555-44912533	-11.65	1.24E-04	ENST00000302347.9	4.31E-02	-0.9 [-0.97;-0.62]	4.71E-02
ITGB2	7.75E-05	21:44913899-44915246	-12.34	9.60E-05	ENST00000397852.5	5.15E-03	-0.96 [-0.98;-0.88]	1.00E-04
ITGB2	7.75E-05	21:44913899-44915246	-12.34	9.60E-05	ENST00000397852.5	5.15E-03	-0.96 [-0.98;-0.88]	1.00E-04
ITGB2	7.75E-05	21:44913899-44915246	-12.34	9.60E-05	ENST00000498666.5	5.07E-03	-0.92 [-0.98;-0.79]	1.14E-02
ITGB2	7.75E-05	21:44913899-44915246	-12.34	9.60E-05	ENST00000498666.5	5.07E-03	-0.92 [-0.98;-0.79]	1.14E-02
ITGB2	7.75E-05	21:44913899-44915246	-12.34	9.60E-05	ENST00000302347.9	4.31E-02	-0.92 [-0.97;-0.72]	2.00E-02
ITGB2	7.75E-05	21:44913899-44915246	-12.34	9.60E-05	ENST00000302347.9	4.31E-02	-0.92 [-0.97;-0.72]	2.00E-02
ITGB2	7.75E-05	21:44913899-44915246	-12.34	9.60E-05	ENST00000522688.5	3.04E-02	-0.87 [-0.96;-0.61]	3.08E-02
ITGB2	7.75E-05	21:44913899-44915246	-12.34	9.60E-05	ENST00000522688.5	3.04E-02	-0.87 [-0.96;-0.61]	3.08E-02
ITGB2	7.75E-05	21:44913899-44915246	-12.34	9.60E-05	ENST00000320216.10	6.78E-03	-0.79 [-0.95;-0.45]	3.97E-02
ITGB2	7.75E-05	21:44913899-44915246	-12.34	9.60E-05	ENST00000320216.10	6.78E-03	-0.79 [-0.95;-0.45]	3.97E-02
ITGB2	7.75E-05	21:44918461-44921815	-18.22	9.00E-06	ENST00000498666.5	5.07E-03	-0.9 [-0.98;-0.55]	1.00E-04
ITGB2	7.75E-05	21:44918461-44921815	-18.22	9.00E-06	ENST00000498666.5	5.07E-03	-0.9 [-0.98;-0.55]	1.00E-04
ITGB2	7.75E-05	21:44918461-44921815	-18.22	9.00E-06	ENST00000320216.10	6.78E-03	-0.85 [-0.93;-0.69]	3.35E-02
ITGB2	7.75E-05	21:44918461-44921815	-18.22	9.00E-06	ENST00000320216.10	6.78E-03	-0.85 [-0.93;-0.69]	3.35E-02
KLC1	7.08E-05	14:103696092-103697307	8.07	9.05E-04	ENST00000246489.11	1.93E-08	0.62 [0.28;0.84]	2.73E-02
LAIR1	6.55E-07	19:54375122-54375966	-8.84	5.29E-04	ENST00000440716.5	9.38E-03	-0.93 [-0.98;-0.73]	3.80E-03
LCP2	1.90E-08	5:170295513-170298924	-13.19	6.10E-05	ENST00000046794.9	1.23E-03	-0.9 [-0.97;-0.66]	1.00E-04
LCP2	1.90E-08	5:170295513-170298924	-13.19	6.10E-05	ENST00000519149.1	8.96E-03	-0.82 [-0.9;-0.62]	1.00E-04
LCP2	1.90E-08	5:170295513-170298924	-13.19	6.10E-05	ENST00000628092.2	6.87E-03	-0.92 [-0.97;-0.74]	1.00E-04
LCP2	1.90E-08	5:170295513-170298924	-13.19	6.10E-05	ENST00000519594.5	2.89E-02	-0.87 [-0.93;-0.72]	3.70E-03
LCP2	1.90E-08	5:170295513-170298924	-13.19	6.10E-05	ENST00000522760.5	7.55E-03	-0.93 [-0.97;-0.8]	8.10E-03
OSM	4.92E-04	22:30266011-30268071	-13.63	5.10E-05	ENST00000215781.2	3.59E-04	-0.68 [-0.95;-0.13]	1.00E-04

(Continued)

TABLE 1 | Continued

Gene	GDE _{p-value}	DMR	DMR _{stat}	DMR _{p-value}	EnsT	DTE _{p-value}	eQTM _r	eQTM _{p-value}
OSM	4.92E-04	22:30266011-30268071	-13.63	5.10E-05	ENST00000215781.2	3.59E-04	-0.68 [-0.95;-0.13]	1.00E-04
PPP6R1	5.34E-04	19:55251330-55251959	-11.33	1.51E-04	ENST00000592242.1	4.83E-02	-0.86 [-0.94;-0.6]	4.22E-02
PPP6R1	5.34E-04	19:55251330-55251959	-11.33	1.51E-04	ENST00000412770.6	2.83E-03	-0.88 [-0.96;-0.6]	4.54E-02
RASSF4	4.87E-05	10:44977022-44977811	11	1.76E-04	ENST00000489171.5	3.42E-02	0.93 [0.76;0.98]	1.00E-04
RPS6KA4	3.11E-04	11:64359911-64360661	-10.55	2.11E-04	ENST00000334205.8	1.51E-03	-0.7 [-0.86;-0.43]	1.00E-04
RPS6KA4	3.11E-04	11:64359911-64360661	-10.55	2.11E-04	ENST00000334205.8	1.51E-03	-0.7 [-0.86;-0.43]	1.00E-04
SH3TC1	3.47E-04	4:8240533-8241022	-7.96	9.85E-04	ENST00000507891.1	6.46E-05	-0.7 [-0.83;-0.17]	1.64E-02
TMC6	6.31E-05	17:78126612-78127570	-9.6	3.35E-04	ENST00000590602.5	3.31E-02	-0.92 [-0.98;-0.82]	3.02E-02
TMC6	6.31E-05	17:78126612-78127570	-9.6	3.35E-04	ENST00000590602.5	3.31E-02	-0.92 [-0.98;-0.82]	3.02E-02
TMC6	6.31E-05	17:78126612-78127570	-9.6	3.35E-04	ENST00000591436.5	1.21E-02	-0.81 [-0.91;-0.64]	3.78E-02
TMC6	6.31E-05	17:78126612-78127570	-9.6	3.35E-04	ENST00000591436.5	1.21E-02	-0.81 [-0.91;-0.64]	3.78E-02
TMC6	6.31E-05	17:78126612-78127570	-9.6	3.35E-04	ENST00000593044.5	1.28E-02	-0.85 [-0.93;-0.67]	4.06E-02
TMC6	6.31E-05	17:78126612-78127570	-9.6	3.35E-04	ENST00000593044.5	1.28E-02	-0.85 [-0.93;-0.67]	4.06E-02
UCP2	7.66E-10	11:73979728-73982457	-9.4	3.75E-04	ENST00000545562.2	3.59E-02	-0.67 [-0.83;-0.46]	1.33E-02
UCP2	7.66E-10	11:73979728-73982457	-9.4	3.75E-04	ENST00000545562.2	3.59E-02	-0.67 [-0.83;-0.46]	1.33E-02
VAC14	2.54E-04	16:70691109-70693047	11.18	1.56E-04	ENST00000261776.9	4.82E-03	0.88 [0.55;0.96]	2.74E-02
VAC14	2.54E-04	16:70746190-70746702	9.13	4.36E-04	ENST00000261776.9	4.82E-03	0.87 [0.71;0.94]	3.48E-02

Expression quantitative trait methylation (eQTM) analysis of the GDEMs representing the correlation between methylation and transcript expression. Key in table legend: Gene = HGNC gene symbol, DMR = co-ordinates of the DMR (GRCh38), DMR_{p-value} = p-value associated to the differential methylation analysis, EnsT = Ensembl transcript ID, DTE_{p-value} = p-value associated to the differential transcript expression analysis, eQTM_r = Pearson correlation coefficient for the methylation-expression correlation and the 95% confidence intervals, eQTM_{p-value} = p-value associated to the methylation-expression correlation. An extended parsable table including the full eQTM analysis for all GDEMs can be found in **Table S6**.

coefficient remained high ($r = -0.87$) (**Table S7**). Nonetheless, given the centrality of *ITGB2* and *LCP2* among the GDEMs would make them interesting candidates in future targeted studies.

Estimated Cellular Composition Suggests Lower Proportion of Neuronal Cells in SJC66_{high}

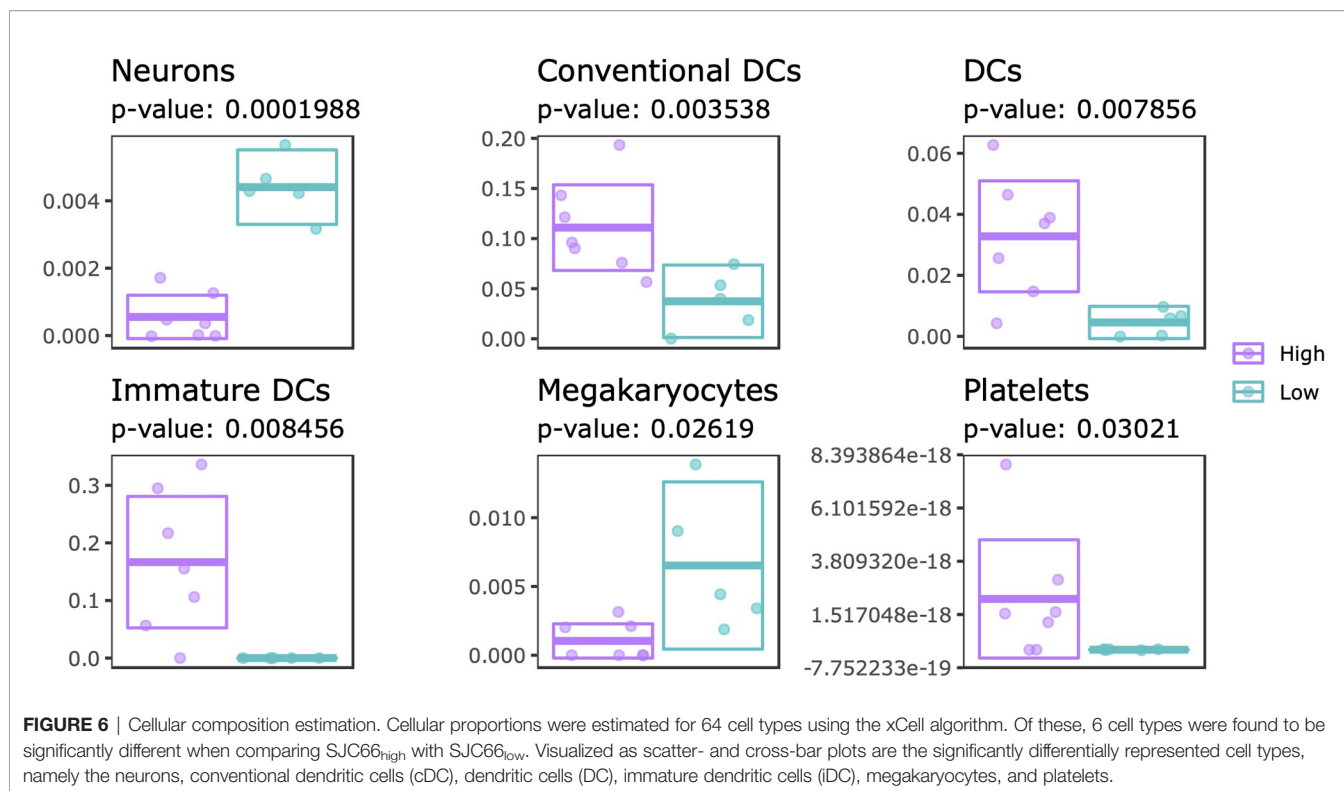
Systematic differences in DNA methylation and gene expression could reflect changes in the cellular composition. To this end, we estimated the cellular composition using the transcriptome data as input for xCell, which is capable of estimating enrichment scores of 64 immune and stromal cell types (47). By comparing the estimated proportions from SJC66_{high} with SJC66_{low} we identified significant differences for 6 cell types: neurons, dendritic cells (DCs: all, conventional and immature), megakaryocytes and platelets (**Figure 6** and **Table S8**). Expectedly, higher enrichment scores for DCs (all subtypes) and platelets were estimated for the SJC66_{high} samples. By contrast, lower proportions of neuron and megakaryocyte signatures were observed for the SJC66_{high} samples. Notably, the difference in neuronal enrichment was found to be the most statistically significant with an almost fourfold difference when comparing SJC66_{high} with SJC66_{low}.

DISCUSSION

In this study, we highlight insights into RA progression by combining the outputs of parallel whole-genome DNA methylome and transcriptome analyses on extracted preparations of synovial biopsies, from auto-antibody positive individuals, early arthritis patients and patients with established RA, stratified by the number of swollen joints. We observed that synovia from patients with a higher number of swollen joints (SJC66 ≥ 9) were different at the level of DNA methylation and

gene expression from synovia from patients with a lower number of swollen joints. Specifically comparing SJC66_{high} with SJC66_{low} revealed 3536 DMRs and 142 DGEs, with both datasets primarily enriched for pathways associated with immune responses. The most significant difference in methylation was found spanning the promoter regions of *MIRLET7B* and *MIR10B*. Interestingly, mouse *miR-let-7b* has been shown to provoke arthritic joint inflammation by remodeling naïve myeloid cells into M1 macrophages *via* TLR-7 ligation (53) and can augment disease severity (54). *MIR10B* has been shown to regulate Th17 cells in patients with ankylosing spondylitis (55) but no studies have specifically associated it with RA. At the level of transcription, *CCL13* and *CLEC10A* were found to be the most differentially expressed. *CCL13* (MCP-4) is an extensively studied chemokine that is thought to be involved with RA pathogenesis and disease progression (56–58). By contrast, not much is known about the role of *CLEC10A* in RA besides it being highly expressed on immature dendritic cells (DCs), monocyte-derived DCs and alternatively activated macrophages (59), as well as having been observed in the inflamed synovium of patients with active RA (60). Chemokines *CCL13*, *CCL8*, *CXCL11*, *CXCL10*, and *CXCL9* regulate the recruitment of leukocytes into tissue and have therefore been implicated in the pathogenesis of RA (61). Differential methylation was observed in the vicinity of the promoter for *CCL13*, *CXCL11*, and *CXCL9*. Such results support a role for epigenetic/transcriptional processes in the spread of pathology to additional joints. While definitive mechanisms of joint spreading remain elusive, possible roles for immune cell migration due to chemokine expression (62, 63) could be further evaluated based on our data.

Altogether, we observed that 3% of DMRs associate with 20% of the differentially expressed GDEs. It is not surprising that not all DMRs could be linked to genes as a large number are found in distal intergenic (18.6%) or intronic (28.3%) regions, making any functional inference challenging. Of the genes that presented



both differentially expression and methylation, protein-protein interaction networks indicated that half encoded for interacting proteins, suggesting that the observed GDEMs function together. The most interconnected GDEMs appeared to be *ITGB2* and *LCP2*, with multiple regions of differential methylation observed surrounding both genes. While we observed transcript differential expression, most transcripts belonging to *ITGB2* and *LCP2* behaved similarly and all displayed reasonably strong inverse correlations with the DMRs located in the promoter area. *ITGB2* encodes an integrin, which would typically be involved in cell-surface mediated signalling. We observed that the gene encoding *ITGB2*'s interaction partner integrin alpha L (*ITGAL*) was also differentially methylated and expressed. *ITGB2* and *ITGAL* together form lymphocyte function-associated antigen (LFA-1), which interacts with intracellular adhesion molecule 1 (ICAM-1 or CD54) resulting in an enhanced immune cell influx into the synovial tissue (64, 65). Inhibiting LFA-1 has been reported to reduce inflammation and joint destruction in murine models of arthritis (66).

Functionally, the 59 GDEMs were primarily over-represented for immune response-associated pathways, specifically T-lymphocyte associated ones. It would be fascinating to understand why the DNA methylome and transcriptome between patients with a SJC66 of 9 and above relative to patients with a SJC66 of 8 present with such a sudden split instead of a gradual difference. It is clear that inflammation is likely an important factor contributing to the observed differences as patients with a high SJC66 also generally present higher levels of inflammation as expressed through using erythrocyte sedimentation rate (ESR) or concentration

C-reactive protein (CRP). Our transcriptomic data indeed suggested an increased proportion of immune cells among SJC66_{high} samples, as would be expected while pathology develops and cells migrate into the affected joints. Previous work has shown that clinically manifest arthritis in established RA is associated with increased infiltration of leukocytes. In synovial tissue samples from clinically involved joints, scores for infiltration by DCs are consistently higher than in clinically uninvolved joints obtained simultaneously from the same RA patients (67). Importantly, when comparing different clinically inflamed joints from the same RA patient simultaneously, leukocyte infiltration in one inflamed joint was shown to be representative of that in other inflamed joints, supporting the notion that leukocytes migrate from one joint to another (68). Indeed, there is continuous influx of leukocytes into the joints in established RA (69). We postulate that if synovial leukocytes exhibit properties that would facilitate cell migration, arthritis might spread from one inflamed joint to another. The results presented here support a disease mechanism in which, after development of clinically established RA, inflammatory and cell adhesion-associated processes play a key role in the progression of RA to greater joint involvement (70–72). Interestingly, we also observed differences of neuronal signatures suggesting a lower relative enrichment of neurons among SJC66_{high} samples. In addition, enrichment analyses on the DMRs suggested hypermethylation of genes encoding nociception receptors, which are typically associated with peripheral sensory neurons (73, 74). A similar decrease in neuronal signature has previously been associated with RA severity, where the authors noted a potential role in the maladaptive response towards damage (75).

This is consistent with a more general loss of anti-inflammatory control by the nervous system in RA (76). Important to note is that our observations are based on estimates made by xCell, which can only calculate enrichment scores based on signatures rather than absolute values of cells. We are therefore unable to discern whether a population increased in size or a different population had decreased. Ideally, a similar estimate would have been generated based on the DNA methylation data, but the currently available reference datasets do not include the cell types measured available in xCell.

There are two limitations of this study, namely the fact that sex confounds the separation between SJC66_{high} and SJC66_{low} and the limited sample size. While we have sought to mitigate the confounding effect of sex by removing genes and CpGs located on the allosomes as well as by including sex as a covariate in our analyses, we acknowledge that we cannot fully eliminate the possibility that a sex effect is present. Accordingly, validation studies would be necessary where the DMEGs are verified using a larger, independent cohort while controlling for an interaction effect between sex and SJC66. The observed differences in transcript expression could be validated using a quantitative PCR approach with primers designed specifically against particular transcripts. Similarly, for validating the DMRs, targeted bisulphite-sequencing using primers for the regions of interest would be a cost-effective approach.

In conclusion, our study constitutes an exploratory analysis of whole genome DNA methylation and gene expression data performed on primary synovial tissue material from auto-antibody positive arthralgia patients without arthritis as well as patients with early and established RA patients. Where previous studies investigated cells from patients with RA versus disease controls and were potentially limited by their use of cultured cells, we focus on an integrative analysis of epigenetic marks and alternative splicing associated with swelling spread, providing novel insights into the mechanisms of disease progression towards more severe phenotypes. Nonetheless, further validation is necessary if the identified target genes are to be used for monitoring or treatment of the swelling and associated inflammatory processes in the joints of RA patients.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://www.ebi.ac.uk/arrayexpress/>, E-MTAB-6638; <https://www.ebi.ac.uk/arrayexpress/>, E-MTAB-6684.

REFERENCES

1. Tak PP, Bresnihan B. The pathogenesis and prevention of joint damage in rheumatoid arthritis: Advances from synovial biopsy and tissue analysis. *Arthritis Rheum* (2000) 43:2619–33. doi: 10.1002/1529-0131(200012)43:12<2619::AID-ANR1>3.0.CO;2-V

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by AMC medisch ethisch toetsingscommissie under MEC 02/152, MEC 05/107, and MEC 07/253 for the Synoviomics, Pre-synoviomics, and Synoviomics II cohorts, respectively. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

EF, HL, DT, CL, MS, DG, PT, and RP conceived the study. KM and GR carried out the laboratory experimental work. EF and CL conceived the analytical design. EF and AL performed the data analysis. HL, DT, CL, MM, PH, and WJ helped supervise the secondary data analyses. EF, AL, KM, HL, and DT led the writing of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

AL was funded by the European Union's Horizon 2020 research and innovation program under Grant Agreement No. ITN-2014-EID-641665. This study was supported by the Dutch Arthritis Foundation (06-1-303 and 11-1-407), IMI BeTheCure (115142), and FP7 Euro-TEAM consortium (305549).

ACKNOWLEDGMENTS

We would like to thank the BTCure research consortium for providing access to the samples, Victor Neduva (Target and Pathway Validation, Target Sciences, GSK) for transferring the raw data files, Erika Cule (Research Statistics, GSK) for input into statistical methodology, and Yee Ying Chang (Clinical Genetics, Genome Diagnostics Laboratory, Amsterdam UMC, University of Amsterdam) for implementing the eQTM function in R. This manuscript has been released as a Pre-Print at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3576744 (77).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2021.651475/full#supplementary-material>

2. Taylor PC, Moore A, Vasilescu R, Alvir J, Tarallo M. A structured literature review of the burden of illness and unmet needs in patients with rheumatoid arthritis: a current perspective. *Rheumatol Int* (2016) 36:685–95. doi: 10.1007/s00296-015-3415-x
3. Okada Y, Wu D, Trynka G, Raj T, Terao C, Ikari K, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* (2014) 506:376–81. doi: 10.1038/nature12873

4. Wijbrandts CA, Tak PP. Prediction of Response to Targeted Treatment in Rheumatoid Arthritis. *Mayo Clin Proc* (2017) 92:1129–43. doi: 10.1016/j.mayocp.2017.05.009
5. Nielen MMJ, van Schaardenburg D, Reesink HW, van de Stadt RJ, van der Horst-Bruinsma IE, de Koning MHMT, et al. Specific autoantibodies precede the symptoms of rheumatoid arthritis: a study of serial measurements in blood donors. *Arthritis Rheum* (2004) 50:380–6. doi: 10.1002/art.20018
6. Kolfenbach JR, Deane KD, Derber LA, O'Donnell CI, Gilliland WR, Edison JD, et al. Autoimmunity to peptidyl arginine deiminase type 4 precedes clinical onset of rheumatoid arthritis. *Arthritis Rheum* (2010) 62:2633–9. doi: 10.1002/art.27570
7. Tak PP, Smeets TJ, Daha MR, Kluin PM, Meijers KA, Brand R, et al. Analysis of the synovial cell infiltrate in early rheumatoid synovial tissue in relation to local disease activity. *Arthritis Rheum* (1997) 40:217–25. doi: 10.1002/art.1780400206
8. Zhang F, Wei K, Slowikowski K, Fonseka CY, Rao DA, Kelly S, et al. Defining inflammatory cell states in rheumatoid arthritis joint synovial tissues by integrating single-cell transcriptomics and mass cytometry. *Nat Immunol* (2019) 20:928–42. doi: 10.1038/s41590-019-0378-1
9. Tak PP, Zvaifler NJ, Green DR, Firestein GS. Rheumatoid arthritis and p53: how oxidative stress might alter the course of inflammatory diseases. *Immunol Today* (2000) 21:78–82. doi: 10.1016/S0167-5699(99)01552-2
10. Ospelt C. Synovial fibroblasts in 2017. *RMD Open* (2017) 3:e000471. doi: 10.1136/rmdopen-2017-000471
11. Pitzalis C, Kelly S, Humby F. New learnings on the pathophysiology of RA from synovial biopsies. *Curr Opin Rheumatol* (2013) 25:334–44. doi: 10.1097/BOR.0b013e32835fd8eb
12. Nemtsova MV, Zaletaev DV, Bure IV, Mikhaylenko DS, Kuznetsova EB, Alekseeva EA, et al. Epigenetic Changes in the Pathogenesis of Rheumatoid Arthritis. *Front Genet* (2019) 10:570. doi: 10.3389/fgene.2019.00570
13. Ospelt C, Gay S, Klein K. Epigenetics in the pathogenesis of RA. *Semin Immunopathol* (2017) 39:409–19. doi: 10.1007/s00281-017-0621-5
14. Nair N, Wilson AG, Barton A. DNA methylation as a marker of response in rheumatoid arthritis. *Pharmacogenomics* (2017) 18:1323–32. doi: 10.2217/pgs-2016-0195
15. Whitaker JW, Shoemaker R, Boyle DL, Hillman J, Anderson D, Wang W, et al. An imprinted rheumatoid arthritis methylome signature reflects pathogenic phenotype. *Genome Med* (2013) 5:40. doi: 10.1186/gm444
16. de la Rica L, Urquiza JM, Gómez-Cabrero D, Islam ABMMK, López-Bigas N, Tegnér J, et al. Identification of novel markers in rheumatoid arthritis through integrated analysis of DNA methylation and microRNA expression. *J Autoimmun* (2013) 41:6–16. doi: 10.1016/j.jaut.2012.12.005
17. Whitaker JW, Boyle DL, Bartok B, Ball ST, Gay S, Wang W, et al. Integrative omics analysis of rheumatoid arthritis identifies non-obvious therapeutic targets. *PLoS One* (2015) 10:1–14. doi: 10.1371/journal.pone.0124254
18. Ekwall A-KH, Whitaker JW, Hammaker D, Bugbee WD, Wang W, Firestein GS. The Rheumatoid Arthritis Risk Gene LBH Regulates Growth in Fibroblast-like Synoviocytes. *Arthritis Rheumatol (Hoboken NJ)* (2015) 67:1193–202. doi: 10.1002/art.39060
19. Hammaker D, Whitaker JW, Maeshima K, Boyle DL, Ekwall A-KH, Wang W, et al. LBH Gene Transcription Regulation by the Interplay of an Enhancer Risk Allele and DNA Methylation in Rheumatoid Arthritis. *Arthritis Rheumatol (Hoboken NJ)* (2016) 68:2637–45. doi: 10.1002/art.39746
20. Ham S, Bae J-B, Lee S, Kim B-J, Han B-G, Kwok S-K, et al. Epigenetic analysis in rheumatoid arthritis synoviocytes. *Exp Mol Med* (2019) 51:22. doi: 10.1038/s12276-019-0215-5
21. Ai R, Whitaker JW, Boyle DL, Tak PP, Gerlag DM, Wang W, et al. DNA Methylome Signature in Synoviocytes From Patients With Early Rheumatoid Arthritis Compared to Synoviocytes From Patients With Longstanding Rheumatoid Arthritis. *Arthritis Rheumatol* (2015) 67:1978–80. doi: 10.1002/art.39123
22. Karouzakis E, Raza K, Kolling C, Buckley CD, Gay S, Filer A, et al. Analysis of early changes in DNA methylation in synovial fibroblasts of RA patients before diagnosis. *Sci Rep* (2018) 8:7370. doi: 10.1038/s41598-018-24240-2
23. Gerlag DM, Tak PP. How to perform and analyse synovial biopsies. *Best Pract Res Clin Rheumatol* (2013) 27:195–207. doi: 10.1016/j.berh.2013.03.006
24. de Hair MJH, van de Sande MGH, Ramwadhoebe TH, Hansson M, Landewé R, Van Der Leij C, et al. Features of the synovium of individuals at risk of developing rheumatoid arthritis : Implications for understanding preclinical rheumatoid arthritis. *Arthritis Rheumatol* (2014) 66:513–22. doi: 10.1002/art.38273
25. de Hair MJH, Harty LC, Gerlag DM, Pitzalis C, Veale DJ, Tak PP. Synovial tissue analysis for the discovery of diagnostic and prognostic biomarkers in patients with early arthritis. *J Rheumatol* (2011) 38:2068–72. doi: 10.3899/jrheum.110426
26. van de Sande MGH, de Hair MJH, Schuller Y, van de Sande GPM, Wijbrandts CA, Dinant HJ, et al. The features of the synovium in early rheumatoid arthritis according to the 2010 ACR/EULAR classification criteria. *PLoS One* (2012) 7:e36668. doi: 10.1371/journal.pone.0036668
27. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO, et al. 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Ann Rheum Dis* (2010) 69:1580–8. doi: 10.1136/ard.2010.138461
28. Argelaguet R, Velten B, Arnol D, Dietrich S, Zenz T, Marioni JC, et al. Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets. *Mol Syst Biol* (2018) 14:e8124. doi: 10.15252/msb.20178124
29. Andrews S. *FastQC: a quality control tool for high throughput sequence data*. Available at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
30. Buffalo V. *Scythe: a Bayesian adapter trimmer*.
31. Joshi N. *Sickle: a windowed adaptive trimming tool for FASTQ files using quality*.
32. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* (2016) 34:525–7. doi: 10.1038/nbt.3519
33. Development Core Team R. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing (2008). Available at: <http://www.r-project.org/>.
34. Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research* (2016) 4:1521. doi: 10.12688/f1000research.7563.2
35. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* (2014) 15:550. doi: 10.1186/s13059-014-0550-8
36. Yi L, Pimentel H, Bray NL, Pachter L. Gene-level differential analysis at transcript-level resolution. *Genome Biol* (2018) 19:52. doi: 10.1186/s13059-018-1419-z
37. Jiang H, Lei R, Ding S-W, Zhu S. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinf* (2014) 15:182. doi: 10.1186/1471-2105-15-182
38. Krueger F, Andrews SR. Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* (2011) 27:1571–2. doi: 10.1093/bioinformatics/btr167
39. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. Genome Project Data Processing S. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* (2009) 25:2078–9. doi: 10.1093/bioinformatics/btp352
40. Korthauer K, Chakraborty S, Benjamini Y, Irizarry RA, User G. Detection and accurate false discovery rate control of differentially methylated regions from whole genome bisulfite sequencing. *Biostatistics* (2018) 00:1–17. doi: 10.1093/biostatistics/kxy007
41. Zhu LJ, Gazin C, Lawson ND, Pagès H, Lin SM, Lapointe DS, et al. ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinf* (2010) 11:237. doi: 10.1186/1471-2105-11-237
42. Frankish A, Diekhans M, Ferreira A-M, Johnson R, Jungreis I, Loveland J, et al. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res* (2019) 47:D766–73. doi: 10.1093/nar/gky955
43. Efron B, Tibshirani RJ. *An introduction to the bootstrap*. New York, N.Y.; London: Chapman & Hall (1993).
44. Canty A, Ripley B. *boot*. (2021).
45. Sergushichev AA. An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. *bioRxiv* (2016), 060012. doi: 10.1101/060012
46. *Clarivate. MetaCore*.
47. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol* (2017) 18:220. doi: 10.1186/s13059-017-1349-1
48. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: Protein-protein association networks with increased coverage,

- supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* (2019) 47:D607–13. doi: 10.1093/nar/gky1131
49. Shayeveitch R, Askayo D, Keydar I, Ast G. The importance of DNA methylation of exons on alternative splicing. *RNA* (2018) 24:1351–62. doi: 10.1261/rna.064865.117
 50. Lev Maor G, Yearim A, Ast G. The alternative role of DNA methylation in splicing regulation. *Trends Genet* (2015) 31:274–80. doi: 10.1016/j.TIG.2015.03.002
 51. Linker SM, Urban L, Clark SJ, Chhatriwala M, Amatya S, McCarthy DJ, et al. Combined single-cell profiling of expression and DNA methylation reveals splicing regulation and heterogeneity. *Genome Biol* (2019) 20:30. doi: 10.1186/s13059-019-1644-0
 52. Harris CJ, Scheibe M, Wongpalee SP, Liu W, Cornett EM, Vaughan RM, Jacobsen SE. A DNA methylation reader complex that enhances gene transcription. *Science* (2018) 362(4619):1182–6. doi: 10.1126/science.aar7854
 53. Kim SJ, Chen Z, Essani AB, Elshabrawy HA, Volin MV, Volkov S, et al. Identification of a Novel Toll-like Receptor 7 Endogenous Ligand in Rheumatoid Arthritis Synovial Fluid That Can Provoke Arthritic Joint Inflammation. *Arthritis Rheumatol* (2016) 68:1099–110. doi: 10.1002/art.39544
 54. Umar S, Palasiewicz K, Van Raemdonck K, Volin MV, Romay B, Amin MA, et al. IRAK4 inhibition: a promising strategy for treating RA joint inflammation and bone erosion. *Cell Mol Immunol* (2020). doi: 10.1038/s41423-020-0433-8
 55. Chen L, Al-Mossawi MH, Ridley A, Sekine T, Hammitzsch A, De Wit J, et al. MiR-10b-5p is a novel Th17 regulator present in Th17 cells from ankylosing spondylitis. *Ann Rheum Dis* (2017) 76:620–4. doi: 10.1136/annrheumdis-2016-210175
 56. Hintzen C, Quaiser S, Pap T, Heinrich PC, Hermanns HM. Induction of CCL13 expression in synovial fibroblasts highlights a significant role of oncostatin M in rheumatoid arthritis. *Arthritis Rheum* (2009) 60:1932–43. doi: 10.1002/art.24602
 57. Yamaguchi A, Nozawa K, Fujishiro M, Kawasaki M, Suzuki F, Takamori K, et al. CC motif chemokine ligand 13 is associated with rheumatoid arthritis pathogenesis. *Mod Rheumatol* (2013) 23:856–63. doi: 10.3109/s10165-012-0752-4
 58. Rump L, Matthey DL, Kehoe O, Middleton J. An initial investigation into endothelial CC chemokine expression in the human rheumatoid synovium. *Cytokine* (2017) 97:133–40. doi: 10.1016/j.cyto.2017.05.023
 59. Hooper JK. ASGR1 and Its Enigmatic Relative, CLEC10A. *Int J Mol Sci* (2020) 21:4818. doi: 10.3390/ijms21144818
 60. McHugh J. Synovial macrophage populations linked to RA remission. *Nat Rev Rheumatol* (2020) 16:471. doi: 10.1038/s41584-020-0481-6
 61. Miyabe Y, Miyabe C, Iwai Y, Luster AD. Targeting the Chemokine System in Rheumatoid Arthritis and Vasculitis. *JMA J* (2020) 3:182–92. doi: 10.31662/jmaj.2020-0019
 62. Elemam NM, Hannawi S, Maghazachi AA. Role of Chemokines and Chemokine Receptors in Rheumatoid Arthritis. *ImmunoTargets Ther* (2020) 9:43–56. doi: 10.2147/itt.s243636
 63. Yu X, Song Z, Rao L, Tu Q, Zhou J, Yin Y, et al. Synergistic induction of CCL5, CXCL9 and CXCL10 by IFN- γ and NLRs ligands on human fibroblast-like synoviocytes—A potential immunopathological mechanism for joint inflammation in rheumatoid arthritis. *Int Immunopharmacol* (2020) 82:1–6. doi: 10.1016/j.intimp.2020.106356
 64. Lowin T, Straub RH. Integrins and their ligands in rheumatoid arthritis. *Arthritis Res Ther* (2011) 13:244. doi: 10.1186/ar3464
 65. Tak PP, Thurkewitz EW, Doha MR, Kluin PM, Smeets TJ, Meinders AE, et al. Expression of adhesion molecules in early rheumatoid synovial tissue. *Clin Immunol Immunopathol* (1995) 77:236–42. doi: 10.1006/clin.1995.1149
 66. Suchard SJ, Stetsko DK, Davis PM, Skala S, Potin D, Launay M, et al. An LFA-1 (alphaLbeta2) small-molecule antagonist reduces inflammation and joint destruction in murine models of arthritis. *J Immunol* (2010) 184:3917–26. doi: 10.4049/jimmunol.0901095
 67. Kraan MC, Versendaal H, Jonker M, Bresnihan B, Post WJ, Hart B, et al. A symptomatic synovitis precedes clinically manifest arthritis. *Arthritis Rheum* (1998) 41:1481–8. doi: 10.1002/1529-0131(199808)41:8<1481::AID-ART19>3.0.CO;2-O
 68. Kraan MC, Reece RJ, Smeets TJM, Veale DJ, Emery P, Tak PP. Comparison of synovial tissues from the knee joints and the small joints of rheumatoid arthritis patients: Implications for pathogenesis and evaluation of treatment. *Arthritis Rheum* (2002) 46:2034–8. doi: 10.1002/art.10556
 69. Thurlings RM, Wijbrandts CA, Bennink RJ, Dohmen SE, Voermans C, Wouters D, et al. Monocyte scintigraphy in rheumatoid arthritis: The dynamics of monocyte migration in immune-mediated inflammatory disease. *PLoS One* (2009) 4:1–6. doi: 10.1371/journal.pone.0007865
 70. Tsubaki T, Arita N, Kawakami T, Shiratsuchi T, Yamamoto H, Takubo N, et al. Characterization of histopathology and gene-expression profiles of synovitis in early rheumatoid arthritis using targeted biopsy specimens. *Arthritis Res Ther* (2005) 7:R25–36. doi: 10.1186/ar1751
 71. Lequerré T, Bansard C, Vittecoq O, Derambure C, Hiron M, Daveau M, et al. Early and long-standing rheumatoid arthritis: distinct molecular signatures identified by gene-expression profiling in synovia. *Arthritis Res Ther* (2006) 11:R105. doi: 10.1186/ar2744
 72. Choy E. Understanding the dynamics: pathways involved in the pathogenesis of rheumatoid arthritis. *Rheumatol (Oxford)* (2012) 51(Suppl 5):v3–11. doi: 10.1093/rheumatology/kes113
 73. Dubin AE, Patapoutian A. Nociceptors: The sensors of the pain pathway. *J Clin Invest* (2010) 120:3760–72. doi: 10.1172/JCI42843
 74. Miller RE, Tran PB, Obeidat AM, Raghu P, Ishihara S, Miller RJ, et al. The Role of Peripheral Nociceptive Neurons in the Pathophysiology of Osteoarthritis Pain. *Curr Osteoporos Rep* (2015) 13:318–26. doi: 10.1007/s11914-015-0280-1
 75. Orange DE, Agius P, DiCarlo EF, Robine N, Geiger H, Szymonifka J, et al. Identification of Three Rheumatoid Arthritis Disease Subtypes by Machine Learning Integration of Synovial Histologic Features and RNA Sequencing Data. *Arthritis Rheumatol* (2018) 70:690–701. doi: 10.1002/art.40428
 76. Koopman FA, van Maanen MA, Vervordeldonck MJ, Tak PP. Balancing the autonomic nervous system to reduce inflammation in rheumatoid arthritis. *J Intern Med* (2017) 282:64–75. doi: 10.1111/joim.12626
 77. Ferrero E, Li Yim A, Maratou K, Lewis HD, Royal G, Tough DF, et al. Novel Insights into Rheumatoid Arthritis Through Characterisation of Concordant Changes in DNA Methylation and Gene Expression in Synovial Biopsies of Patients with Differing Numbers of Swollen Joints. *SSRN Electron J* (2020). doi: 10.2139/ssrn.3576744

Conflict of Interest: AL, KM, GR, HL, DT, CL, DG, and RP were employed by GSK when this study was conducted. EF and PT were employed by Novartis and Candel Therapeutics when this study was conducted, respectively. WJ was financially supported by GSK and Mead Johnson.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Li Yim, Ferrero, Maratou, Lewis, Royal, Tough, Larminie, Mannens, Henneman, de Jonge, van de Sande, Gerlag, Prinjha and Tak. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.