ARTICLE

Comprehensive epigenetic landscape of rheumatoid arthritis fibroblast-like synoviocytes

Rizi Ai[1], Teresina Laragione[2], Deepa Hammaker[3], David L. Boyle[3], Andre Wildberg[1], Keisuke Maeshima[3], Emanuele Palescandolo[4], Vinod Krishna[4], David Pocalyko[4], John W. Whitaker[4], Yuchen Bai[4], Sunil Nagpal[4], Kurtis E. Bachman[4], Richard I. Ainsworth[1], Mengchi Wang[1], Bo Ding[1], Percio S. Gulko[2], Wei Wang[1,5] & Gary S. Firestein[3]

Epigenetics contributes to the pathogenesis of immune-mediated diseases like rheumatoid arthritis (RA). Here we show the first comprehensive epigenomic characterization of RA fibroblast-like synoviocytes (FLS), including histone modifications (H3K27ac, H3K4me1, H3K4me3, H3K36me3, H3K27me3, and H3K9me3), open chromatin, RNA expression and whole-genome DNA methylation. To address complex multidimensional relationship and reveal epigenetic regulation of RA, we perform integrative analyses using a novel unbiased method to identify genomic regions with similar profiles. Epigenomically similar regions exist in RA cells and are associated with active enhancers and promoters and specific transcription factor binding motifs. Differentially marked genes are enriched for immunological and unexpected pathways, with "Huntington's Disease Signaling" identified as particularly prominent. We validate the relevance of this pathway to RA by showing that Huntingtin-interacting protein-1 regulates FLS invasion into matrix. This work establishes a high-resolution epigenomic landscape of RA and demonstrates the potential for integrative analyses to identify unanticipated therapeutic targets.

[1] Department of Chemistry and Biochemistry, 9500 Gilman Drive, UC San Diego, La Jolla, CA 92093, USA. [2] Division of Rheumatology, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York, NY 10029, USA. [3] Division of Rheumatology, Allergy and Immunology, 9500 Gilman Drive, UCSD School of Medicine, La Jolla, CA 92093, USA. [4] Janssen Pharmaceuticals, 1400 McKean Road, Spring House, PA 19477, USA. [5] Department of Cellular and Molecular Medicine, 9500 Gilman Drive, UCSD School of Medicine, La Jolla, CA 92093, USA. Correspondence and requests for materials should be addressed to P.S.G. (email: percio.gulko@mssm.edu) or to W.W. (email: wei-wang@ucsd.edu) or to G.S.F. (email: gfirestein@ucsd.edu)

Rheumatoid arthritis (RA) is an aggressive immune-mediated joint disease with synovial inflammation and joint destruction[1]. Despite improvements in therapy, many patients have persistent inflammation and progressive disability. Epigenetic alterations such as DNA methylation and histone modification might contribute to RA pathogenesis and provide clues to identify novel therapeutic targets[2–7]. Much of the effort to define the RA epigenome has focused on fibroblast-like synoviocytes (FLS) of the synovial intimal lining, which invade the cartilage and assume a unique aggressive phenotype in patients with RA[8].

Individual epigenetic marks, such as DNA methylation measured by arrays, have been investigated in RA, sometimes together with gene expression or DNA polymorphisms[9, 10]. The previous integrative analyses of these data were relatively simple, using overlapping marks to identify multi-evidence genes, for example. Despite insights gained from these analyses, a comprehensive and global characterization of the epigenomic landscape is needed to define its contribution to pathogenesis of RA. Furthermore, although, multiple omics technologies could provide a unique opportunity to define the global epigenomic landscape of RA, they also pose a great challenge to analyze in one integrative model. Segmentation methods, such as ChromHMM[11] and Segway[12] can identify functional elements but focus on histone modifications and have not incorporated other data such as DNA methylation. Furthermore, the epigenomic signals in RA and OA FLS mostly differ in magnitude rather than lead to distinct epigenetic states, which also limits the applicability of segmentation methods in which signals are discretized and the subtle amplitude difference diminishes. Therefore, a systematic approach is needed to integrate large-scale autoimmune disease epigenomes to define the epigenomic landscape.

In this study, we apply multiple omics technologies on primary cells from the site of disease in RA and control (osteoarthritis [OA]) FLS, including (1) chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) to map six histone modifications, (2) assay for transposase accessible chromatin with high-throughput sequencing (ATAC-seq) for mapping chromatin accessibility genome-widely, (3) RNA sequencing (RNA-seq) to measure whole genome RNA expression, and (4) whole-genome bisulfite sequencing (WGBS) for DNA methylation level quantification. To map the epigenomic landscape in RA, we integrate diverse epigenomic data into a single analysis using our unbiased method called EpiSig (http://github.com/Wang-lab-UCSD/EpiSig, see Supplementary Methods). This platform clusters regions with similar epigenomic profiles across all the RA and OA FLS samples. The epigenomic co-modifications identify clustered regions that share common functionality. As a result, this is the first time that histone modifications, WGBS, ATAC-seq and RNA-seq data have been combined into a single analysis to capture the deep complexity of the epigenomic landscape in any immune-mediated disease. Our approach incorporates signal intensities across all samples and can capture subtle epigenetic differences that distinguish RA. EpiSig clusters enriched with differentially modified epigenetic regions (DMER) were then extracted to identify RA-specific pathways and transcription factor motifs that could be mined for novel therapeutic targets. In addition to pathways known to be relevant to RA, unexpected ones also emerged. Of particular interest, the "Huntington's Disease Signaling" pathway was discovered as highly significant and biologically validated. The new methodology and dataset provide a new way to identify RA-specific targets that can be used to develop novel therapeutic agents.

Results

Epigenomic landscape of RA and OA FLS. A total of 191 genome-wide datasets were generated across 11 RA and 11 OA

samples, including 130 histone modification datasets, 22 open chromatin datasets, 20 RNA-seq datasets and 19 DNA methylation datasets. Six histone modifications marks were analyzed, including histone H3 lysine 4 trimethylation (H3K4me3), associated with promoter regions; H3 lysine 4 monomethylation (H3K4me1), associated with enhancer regions; H3 lysine 27 acetylation (H3K27ac), associated with increased activation of promoter and enhancer regions; H3 lysine 36 trimethylation (H3K36me3), associated with transcribed regions; H3 lysine 27 trimethylation (H3K27me3), associated with Polycomb repression; and H3 lysine 9 trimethylation (H3K9me3), associated with heterochromatin regions. Additional epigenomic marks include: open chromatin regions, profiled by ATAC-seq, denoting regions of accessible chromatin, and typically associated with regulator binding; DNA methylation, commonly associated with repressed regulatory regions, profiled with WGBS; and RNA-seq used for measuring gene expression levels (Fig. 1a). We integrated nine datasets and provided global views of the first high-resolution functional epigenomic landscape of RA and OA FLS (Fig. 1b).

Identification of epigenomically co-modified regions. For integrative analysis of these diverse epigenomic data, EpiSig was used to find regions sharing similar epigenomic profiles across all the samples (Fig. 1c). The method first divided the whole genome into 5 kb regions and identified regions with any enriched epigenomic signals. We then iteratively identified clusters: in each iteration, regions that shared the most similar signal patterns were determined as a seed and the remaining regions in this cluster were assigned or rejected using a Gibbs sampling alike method. In each step, the 5 kb regions were shifted and oriented to align the patterns correctly. This is the first time WGBS, ATAC-seq and RNA-seq data have been integrated with histone marks to capture the greater complexity. The software is available at http://wanglab.ucsd.edu/star/EpiSig and http://github.com/Wang-lab-UCSD/EpiSig, and is further described in Supplementary Methods).

EpiSig identified 169,502 signal-enriched 5 kb regions that were grouped into 125 epigenomic clusters. Each cluster contains an average of 1356 regions that share similar epigenomic profiles across all FLS lines (Fig. 2a and Supplementary Fig. 1). We then grouped similar epigenomic patterns (Fig. 2a). All the samples were sequenced together, so there was no batch effect that could contribute to these observations. To facilitate interpretation of the results, individual clusters were further combined into nine groups, which we call "sections", using a self-organizing map algorithm (SOM).

Overall, the 9 sections show distinct epigenomically co-modified functional characteristics. Examples of these signal profiles in the epigenomic landscape are shown in Fig. 2a. Of the epigenomically defined sections, section I is highly enriched in H3K27ac, H3K4me3 and open chromatin regions with relatively weaker H3K4me1, which has high coverage in the TSS, 1 kb upstream of the TSS and 5′-UTR regions. CpG islands are also highly covered in this section, but most are de-methylated, indicating active promoter regions. Section II is enriched in H3K27ac, H3K4me1 signals and open chromatin regions, and low in DNA methylation, which is likely associated with active enhancer regions. This section is highly enriched with intergenic regions. Section III is largely composed of bivalent promoters, characterized by intermediate H3K4me3 signals together with the repressive mark H3K27me3, and enriched with regions around annotated TSSs. Notably, the poised promoters are highly enriched with chromosome X loci (~21% of regions in section III are located on chromosome X).
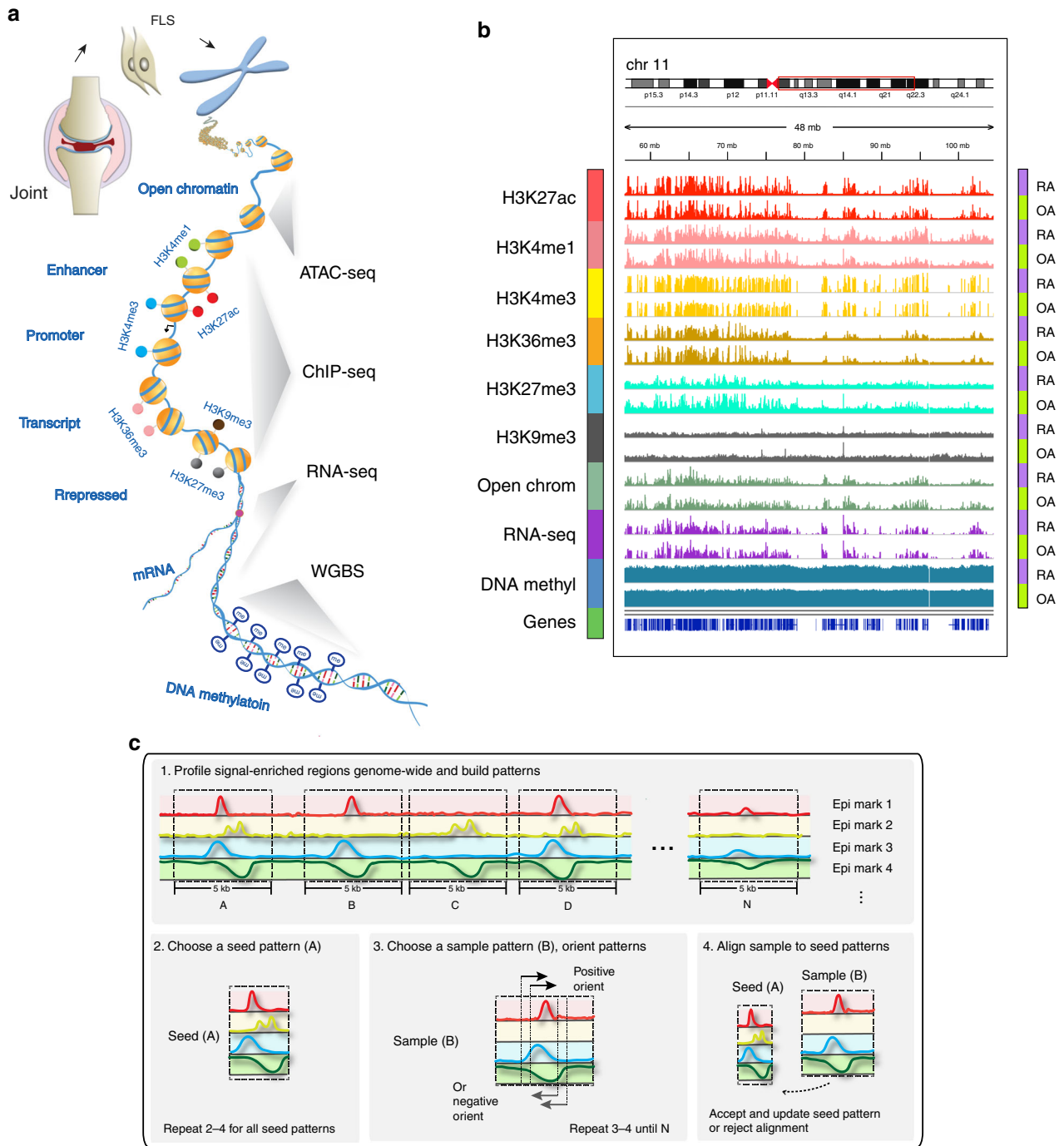
Fig. 1 Overview of RA integrative analysis pipeline. **a** Schematic representation of major omics methods used in RA FLS. **b** Epigenomic landscape of RA FLS with six histone modifications, open chromatin, RNA-seq and DNA methylation. The figure shows an example of the relative signal intensity across a selected region of the genome for each mark in RA and OA FLS. In addition, the locations across chromosomes for the selected regions are indicated. **c** Schematic overview of how EpiSig integrates different epigenetic marks. The algorithm detects significant signals from sequencing data in 5 kb genome regions, clusters regions based on similar epigenomic patterns, and identify epigenomically similar regions

Section IV includes signals of H3K27ac and H3K4me1, together with strong signals of H3K36me3 and RNA-seq. In addition, this section covers low-intergenic regions and is likely associated with intragenic enhancers. Both section V and VII represent actively transcribed regions as indicated by strong H3K36me3 and RNA-seq signals, albeit relatively weaker in section V compared to section VII and with locations highly biased toward the gene body. Section VI and IX represent repressive regions marked by strong signal of H3K27me3. Compared with section VI, section IX shows a strong signal for heterochromatin mark H3K9me3 but much lower signals of open chromatin or DNA methylation. The coverage of intergenic region is high in these two sections. Lastly, H3K9me3 is enriched in section VIII together with high H3K36me3 signal, suggesting that these regions are associated with heterochromatin or zinc finger protein genes[13]. Interestingly, the genomic location of this
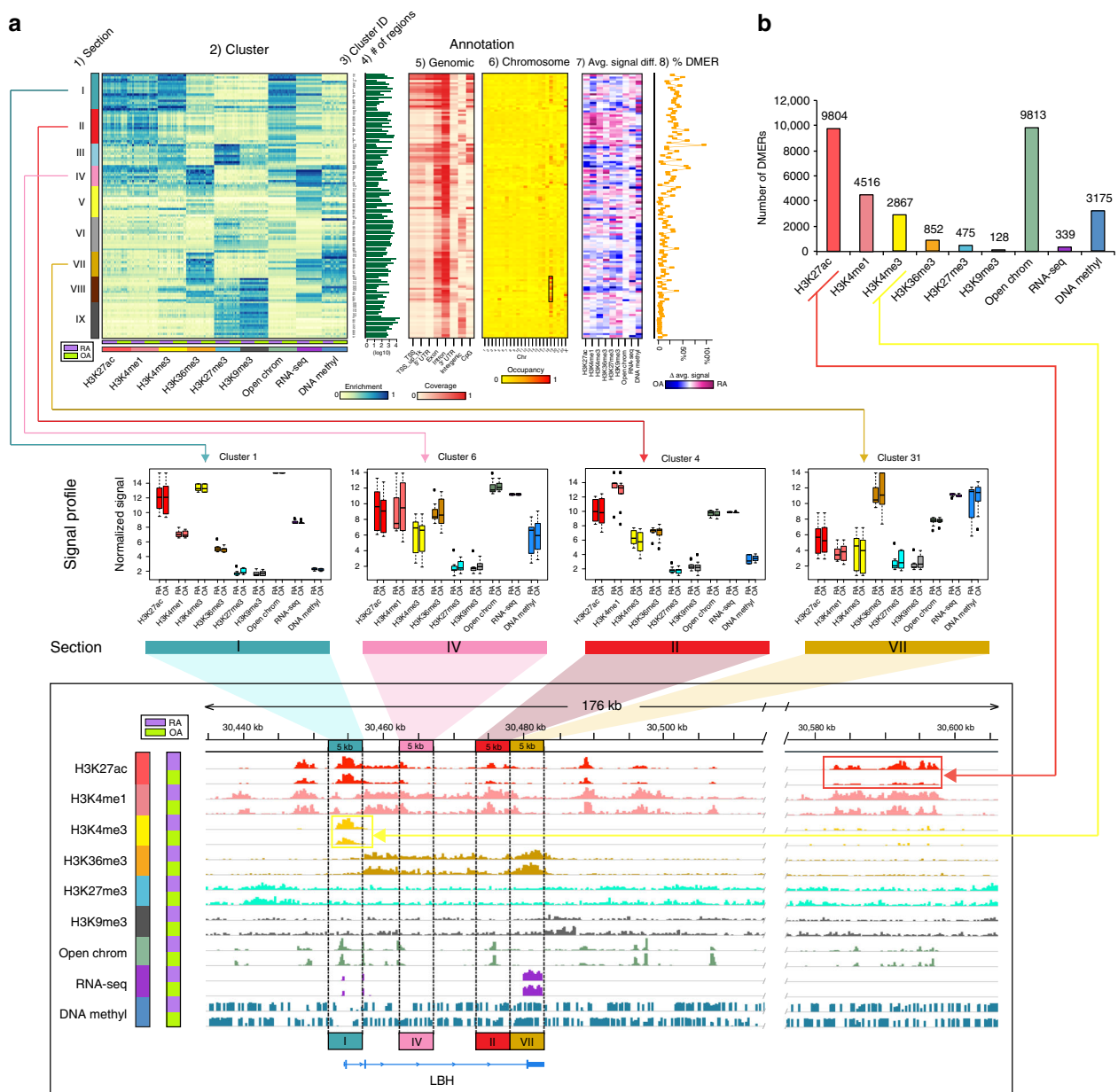
Fig. 2 Genome-wide multidimensional clusters in RA and non-RA genomes. **a** A total of 125 genome-wide 5 kb multidimensional clusters were grouped into nine sections based epigenetic marks. From left to right of the upper panel shows the aligned 125 clusters, each of which includes 11 RA and 11 OA FLS lines: (1) the organization of clusters into epigenetically defined sections (see text); (2) genome annotation heatmap with coverage of each cluster, with 22 separate FLS lines shown within each cluster; (3) 125 cluster IDs; (4) the number of 5 kb regions in each epigenetic cluster (log10 scale); (5) the characteristics of each region within the clusters, such as the relative abundance of transcription start sites (TSS), gene bodies, or introns; (6) chromosome location of each region within the clusters (box shows high occupancy at chromosome 19); (7) differences in signals of epigenetic marks between RA and OA; and (8) percentage of DMERs each cluster. In the middle panel, boxplot examples of clusters in different sections corresponding to *LBH* regions in the genome browser (bottom panel) demonstrating differences between RA and OA and how the FLS marks can be categorized as sections I, II, IV, and VII[14]. **b** DMERs with enhancer and promoter histone marks account for the most common differences between RA and OA. The red and yellow lines connect H3K27ac and H3K4me3 bars to the genome browser, which shows an example of how those marks are differentially modified between RA and OA

section is largely biased toward chromosome 19 (35%) and X (35%) (Supplementary Fig. 2).

DMERs between RA and non-RA. After defining the overall epigenomic landscape for FLS, we compared the relative intensity of each epigenetic mark between RA and our non-RA control. A total of 31,969 DMERs were identified across the 125 clusters (Fig. 2b and Supplementary Data 1). Open chromatin regions were found in most regions, with 9813 DMERs between RA and

OA. A total of 339 differentially expressed genes (DEGs, >2-fold change and Benjamini-Hochberg [B-H] adjusted *p*-value <0.05) were also identified between RA and OA (Supplementary Table 1). Of these, 124 and 215 were over-expressed or under-expressed in RA relative to OA, respectively. For ChIP-seq DMERs, H3K27ac predominated in 9804 DMERs, followed by H3K4me1 and H3K4me3 with 4516 and 2867 DMERs, respectively (Supplementary Fig. 3). Compared to narrow peaks (H3K4me3, H3K4me1, and H3K27ac), histone marks with broad

peaks (H3K27me3, H3K9me3, and H3K36me3) tended to have less peak abundance. In addition, 3175 DNA methylation DMERs were identified. Examples of H3K27ac and H3K4me3 DMERs are shown in the genome browser screenshot in Fig. 2a. The pairwise correlations between DMERs (histone modifications, open chromatin and DNA methylation) and gene expression are shown in Supplementary Fig. 4. Low correlations were observed for pairwise correlations, while the three-dimensional plots showed higher correlation among H3K4me1/3, RNA-seq and H3K27ac. To find the co-localization of DMERs and RA-associated genetic variations, 378 unique SNPs in 524 RA associations were downloaded from GWAS Catalog, 10 unique SNPs were overlapped with DMERs identified in our study (Supplementary Table 2). Interestingly, the co-localization was found in *LBH*[14, 15], which was identified by our previous DNA methylation studies and plays a critical role in cell proliferation. It also provides further evidence that our unbiased methods have biologic relevance. The SNP associated with gene *FADS1, FADS2,* and *FADS3* overlapped with three DMERs (H3K27ac, H3K27me3, and H3K4me3).

Integrative analysis reveals novel biological pathways. To reveal the top enriched differentially modified biological pathways in RA, we used a stringent cutoff for DMERs (*q*-value <0.01) and selected 13 clusters with significant DMER enrichment for functional pathway analysis (Fig. 3a and Supplementary Table 3). Of these, nine clusters that associate with active enhancer regions had predominant differences in the number of H3K27ac DMERs (44.3% of the DMERs) that distinguish RA from OA (see Fig. 3a for the specific clusters). Moderate differences between RA and OA were also observed in open chromatin (21.1% of the DMERs). 8.4% of the differences between RA and OA in these clusters were due to the H3K4me1. The other four clusters had characteristics of active promoter regions, with large differences between RA and OA observed in mark H3K4me3 (47.4%) and intermediate differences in H3K27ac (25.1%). H3K4me1 DMERs account for 2.7% of all DMERs between RA and OA in these clusters. Thus, the key differences that distinguish the RA epigenomic landscape related to transcriptional regulatory regions like enhancers and promoters. The size, percentage and significance of enriched DMERs are shown in Fig. 3a.

Statistically significant differentially modified pathways (B-H adjusted *p*-value from Fisher's exact test <0.05) involving the genes in the clusters with DMERs were found in 8 of the 13 clusters (Supplementary Table 4). Multiple enriched pathways in each individual cluster involved inflammation, immune response, matrix regulation, and cell migration, which supports the different known disease mechanisms of RA. Equally striking, however, was the number of biological pathways that could not be predicted based on the literature or current understanding of RA (see below). Cluster 8, with 579 DMERs, showed the highest DMER enrichment (Fig. 3a). Cluster 8 is associated with active enhancer regions, and 52% of the DMER differences between RA and OA are due to the H3K27ac mark. DMERs in open chromatin regions and DEGs are 16% and 10% of the DMERs, respectively. The cluster with the next most significantly enriched DMERs is cluster 1, which is associated with active promoter regions. About 74% of difference between RA and OA in cluster 1 is due to promoter mark H3K4me3. The activation mark H3K27ac accounted for ~11% of DMERs. Cluster 7 is the third most DMER enriched cluster and with a 59% difference between RA and OA for H3K27ac.

Interestingly, RA-specific pathways "Role of Osteoblasts, Osteoclasts and Chondrocytes in Rheumatoid Arthritis" and "Role of Macrophages, Fibroblasts and Endothelial Cells in Rheumatoid Arthritis" were significant and only enriched in clusters associated with active enhancer regions. Pathways enriched in clusters associated with active promoter regions are more generally associated with regulatory function. For instance, clusters 7 and 8, associated with active enhancers, are especially important as "RA-specific" clusters and have the two enriched RA-associated pathways (Fig. 3b and Supplementary Table 4). Therefore, these active enhancer associated regions are a characteristic chromatin state that distinguishes RA and OA.

Other notable pathways that have been observed in our previous integrative analysis[5, 16] using less robust methodology and different technology (e.g., methylation chips) include "Integrin Signaling", "Leukocyte Extravasation Signaling", and "p53 Signaling". The current dataset confirm those findings, validating the current methodology. As noted, many additional non-obvious pathways also emerge from the epigenomic landscape and could be used to identify potential pathogenic genes.

To select biological pathways for subsequent target identification, pathways were also prioritized based on their frequency in the 8 clusters with significantly enriched pathways, and top 28 most frequently enriched ones that appeared in two or more clusters are shown in Fig. 4a. Among the most significant pathways, we found that "Phospholipase C Signaling", "p53 Signaling", "Integrin Signaling", and "Protein Kinase A signaling" were particularly notable. Of interest, the one that appeared the most often was totally unexpected, namely "Huntington's Disease Signaling". This pathway is enriched in four different clusters with 45 differentially modified genes (DMGs) associated (Fig. 4a) (see below for biologic validation).

Cluster-specific transcription factor motif discovery. Motif analysis in DMERs regions was then performed on the 13 clusters with significantly enriched DMERs in order to understand the drivers of pathogenic pathways. Known motifs and de novo motifs that appeared in two or more clusters are listed in Fig. 4b. The top abundant known motifs are from bZIP and Runt families, binding to promoter regions to control gene expression. For example, TF AP-1 in bZIP family, containing c-Jun and c-Fos, regulates gene expressions of inflammatory cytokines, growth factors, and matrix-degrading matrix metalloproteinases (MMPs), and is crucial to joint destruction in RA[17]. The Runt family potentially plays a role in development of autoimmunity in RA. Interestingly, multiple transcription factors from Stat family are enriched in cluster 7 (Supplementary Table 5). One of these, STAT3, is a key pathway for IL-6 signal transduction, and IL-6 blockade is a highly effective therapy for RA and acts by blocking synovial STAT3 phosphorylation[18].

Biological validation of differentially modified pathways. To determine the functional relevance of non-obvious pathways, we focused on the "Huntington's Disease Signaling" pathway because it was the most commonly enriched pathway and was not known to be associated with immune-mediated disease (Fig. 4a and Supplementary Fig. 5). There are a multitude of genes in this pathway that could serve as potential targets in this pathway, and we noted that on Huntingtin-interacting protein-1 (HIP1) appears in many key locations. The protein participates in clathrin-mediated endocytosis and cell signaling. It also regulates cell invasion in oncology[19, 20] but its potential role has not been explored in inflammatory diseases.

Initial studies showed that HIP1 protein is constitutively expressed in FLS and can be depleted in cells with siRNA (Fig. 5a and Supplementary Fig. 6). We then examined the effect of control and HIP1-deficient FLS in an in vitro invasion model, a model that correlates with in vivo cartilage and joint damage[21, 22].
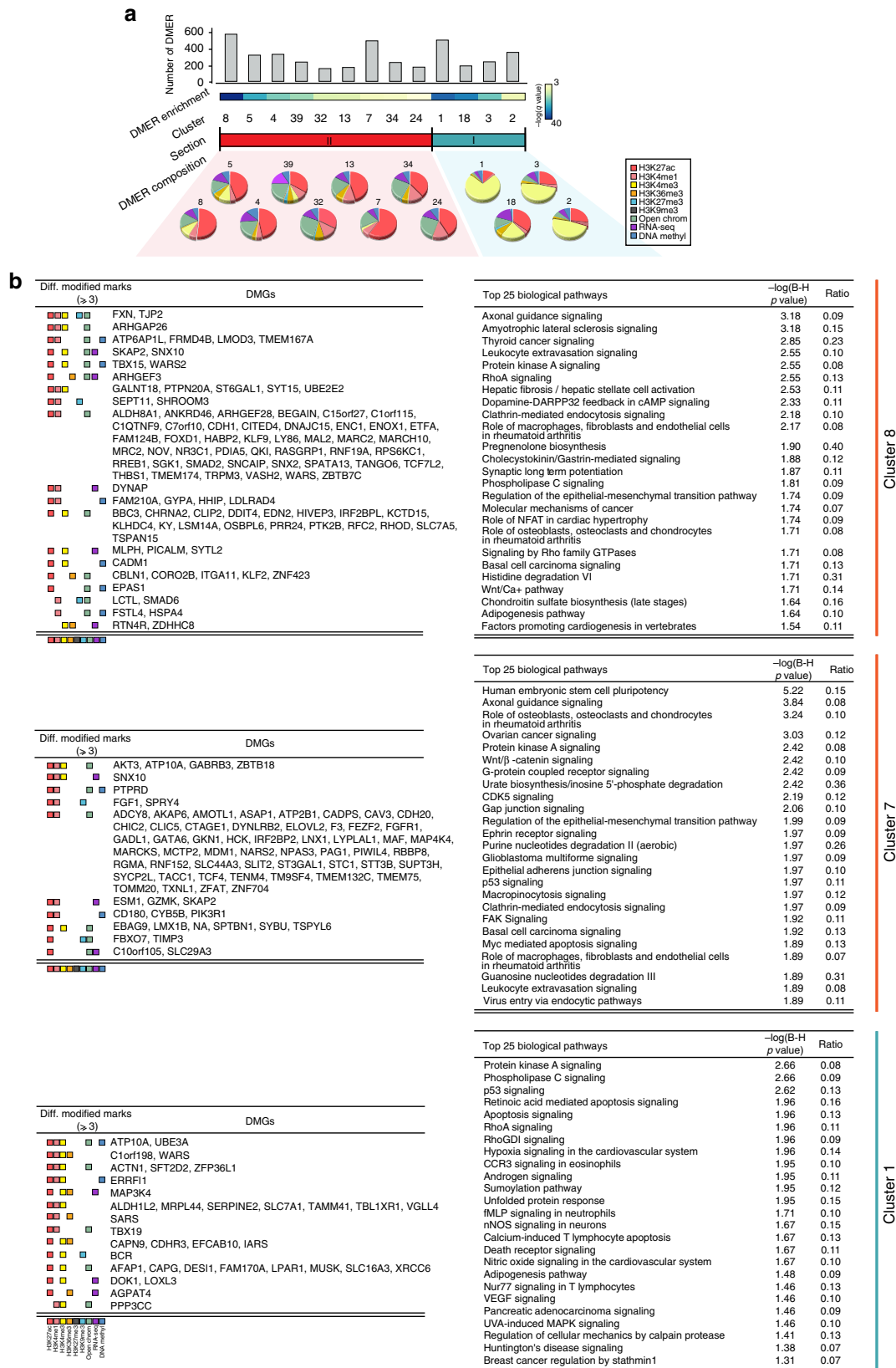
Fig. 3 Integrative analysis identified 13 clusters with significantly enriched DMERs. **a** The 13 enriched clusters are shown with the number of DMERs in each cluster. Below, the relative abundance of various types of differential marks is shown and demonstrates the high abundance of active enhancer and active promoter marks (section I and II). Nine clusters are associated with active enhancers and four with active promoters. **b** Characterization of three clusters with the most DMERs (1, 7, and 8). Top DMGs (left) and top enriched biological pathways associated with these genes (right) are shown
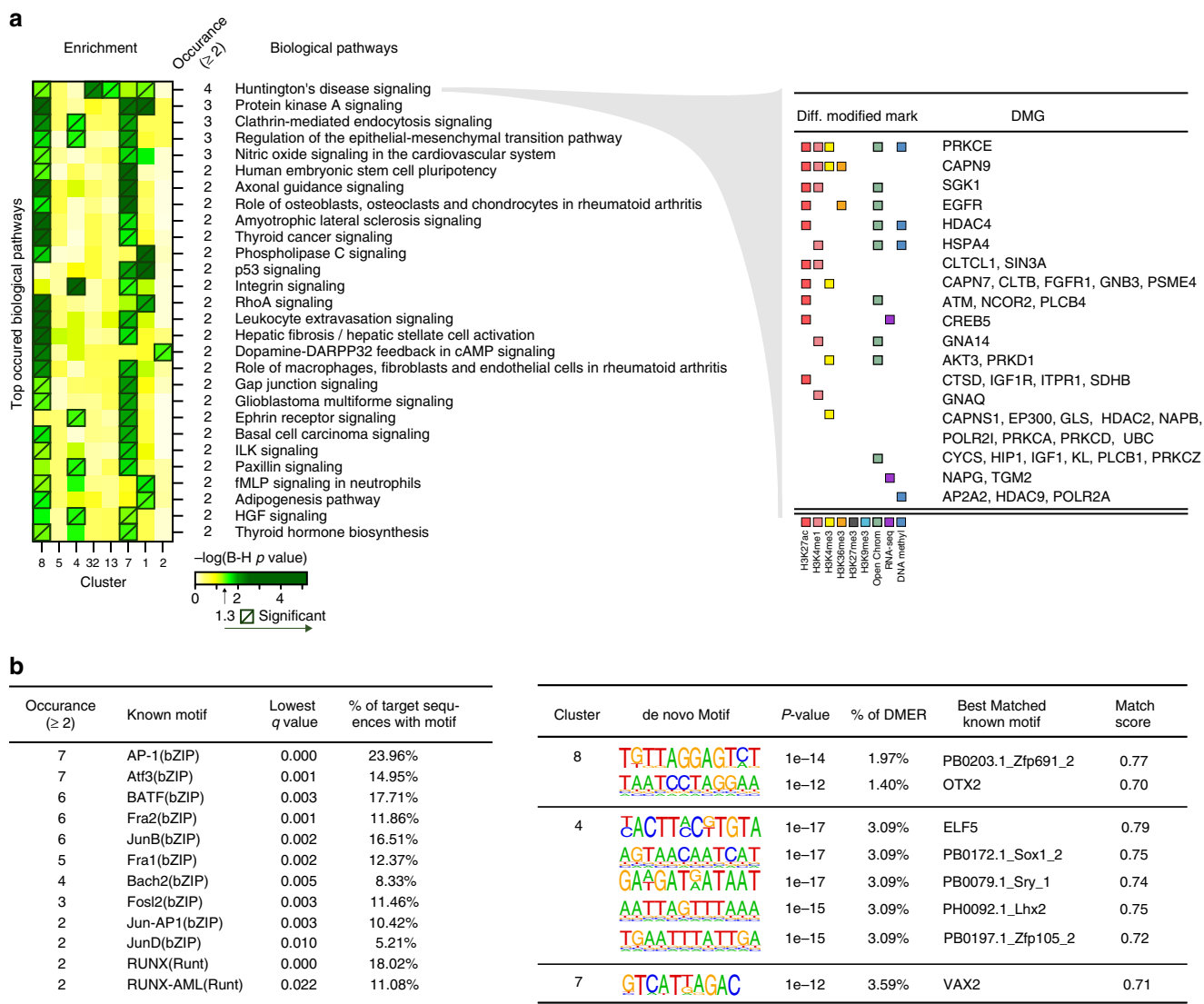
Cluster 8 — DMGs

FXN, TJP2
ARHGAP26
ATP6AP1L, FRMD4B, LMOD3, TMEM167A
SKAP2, SNX10
TBX15, WARS2
ARHGEF3
GALNT18, PTPN20A, ST6GAL1, SYT15, UBE2E2
SEPT11, SHROOM3
ALDH8A1, ANKRD46, ARHGEF28, BEGAIN, C15orf27, C1orf115, C1QTNF9, C7orf10, CDH1, CITED4, DNAJC15, ENC1, ENOX1, ETFA, FAM124B, FOXD1, HABP2, KLF9, LY86, MAL2, MARC2, MARCH10, MRC2, NOV, NR3C1, PDIA5, QKI, RASGRP1, RNF19A, RPS6KC1, RREB1, SGK1, SMAD2, SNCAIP, SNX2, SPATA13, TANGO6, TCF7L2, THBS1, TMEM174, TRPM3, VASH2, WARS, ZBTB7C
DYNAP
FAM210A, GYPA, HHIP, LDLRAD4
BBC3, CHRNA2, CLIP2, DDIT4, EDN2, HIVEP3, IRF2BPL, KCTD15, KLHDC4, KY, LSM14A, OSBPL6, PRR24, PTK2B, RFC2, RHOD, SLC7A5, TSPAN15
MLPH, PICALM, SYTL2
CADM1
CBLN1, CORO2B, ITGA11, KLF2, ZNF423
EPAS1
LCTL, SMAD6
FSTL4, HSPA4
RTN4R, ZDHHC8

Cluster 8 — Top 25 biological pathways

Top 25 biological pathways	−log(B-H p value)	Ratio
Axonal guidance signaling	3.18	0.09
Amyotrophic lateral sclerosis signaling	3.18	0.15
Thyroid cancer signaling	2.85	0.23
Leukocyte extravasation signaling	2.55	0.10
Protein kinase A signaling	2.55	0.08
RhoA signaling	2.55	0.13
Hepatic fibrosis / hepatic stellate cell activation	2.53	0.11
Dopamine-DARPP32 feedback in cAMP signaling	2.33	0.11
Clathrin-mediated endocytosis signaling	2.18	0.10
Role of macrophages, fibroblasts and endothelial cells in rheumatoid arthritis	2.17	0.08
Pregnenolone biosynthesis	1.90	0.40
Cholecystokinin/Gastrin-mediated signaling	1.88	0.12
Synaptic long term potentiation	1.87	0.11
Phospholipase C signaling	1.81	0.09
Regulation of the epithelial-mesenchymal transition pathway	1.74	0.09
Molecular mechanisms of cancer	1.74	0.07
Role of NFAT in cardiac hypertrophy	1.74	0.09
Role of osteoblasts, osteoclasts and chondrocytes in rheumatoid arthritis	1.71	0.08
Signaling by Rho family GTPases	1.71	0.08
Basal cell carcinoma signaling	1.71	0.13
Histidine degradation VI	1.71	0.31
Wnt/Ca+ pathway	1.71	0.14
Chondroitin sulfate biosynthesis (late stages)	1.64	0.16
Adipogenesis pathway	1.64	0.10
Factors promoting cardiogenesis in vertebrates	1.54	0.11

Cluster 7 — DMGs

AKT3, ATP10A, GABRB3, ZBTB18
SNX10
PTPRD
FGF1, SPRY4
ADCY8, AKAP6, AMOTL1, ASAP1, ATP2B1, CADPS, CAV3, CDH20, CHIC2, CLIC5, CTAGE1, DYNLRB2, ELOVL2, F3, FEZF2, FGFR1, GADL1, GATA6, GKN1, HCK, IRF2BP2, LNX1, LYPLAL1, MAF, MAP4K4, MARCKS, MCTP2, MDM1, NARS3, NPAS3, PAG1, PIWIL4, RBBP8, RGMA, RNF152, SLC44A3, SLIT2, ST3GAL1, STC1, STT3B, SUPT3H, SYCP2L, TACC1, TCF4, TENM4, TM9SF4, TMEM132C, TMEM75, TOMM20, TXNL1, ZFAT, ZNF704
ESM1, GZMK, SKAP2
CD180, CYB5B, PIK3R1
EBAG9, LMX1B, NA, SPTBN1, SYBU, TSPYL6
FBXO7, TIMP3
C10orf105, SLC29A3

Cluster 7 — Top 25 biological pathways

Top 25 biological pathways	−log(B-H p value)	Ratio
Human embryonic stem cell pluripotency	5.22	0.15
Axonal guidance signaling	3.84	0.08
Role of osteoblasts, osteoclasts and chondrocytes in rheumatoid arthritis	3.24	0.12
Ovarian cancer signaling	3.03	0.12
Protein kinase A signaling	2.42	0.08
Wnt/β -catenin signaling	2.42	0.10
G-protein coupled receptor signaling	2.42	0.09
Urate biosynthesis/inosine 5'-phosphate degradation	2.42	0.36
CDK5 signaling	2.19	0.12
Gap junction signaling	2.06	0.10
Regulation of the epithelial-mesenchymal transition pathway	1.99	0.09
Ephrin receptor signaling	1.97	0.09
Purine nucleotides degradation II (aerobic)	1.97	0.26
Glioblastoma multiforme signaling	1.97	0.09
Epithelial adherens junction signaling	1.97	0.10
p53 signaling	1.97	0.11
Macropinocytosis signaling	1.97	0.12
Clathrin-mediated endocytosis signaling	1.97	0.09
FAK Signaling	1.92	0.11
Basal cell carcinoma signaling	1.92	0.13
Myc mediated apoptosis signaling	1.89	0.13
Role of macrophages, fibroblasts and endothelial cells in rheumatoid arthritis	1.89	0.07
Guanosine nucleotides degradation III	1.89	0.31
Leukocyte extravasation signaling	1.89	0.08
Virus entry via endocytic pathways	1.89	0.11

Cluster 1 — DMGs

ATP10A, UBE3A
C1orf198, WARS
ACTN1, SFT2D2, ZFP36L1
ERRFI1
MAP3K4
ALDH1L2, MRPL44, SERPINE2, SLC7A1, TAMM41, TBL1XR1, VGLL4
SARS
TBX19
CAPN9, CDHR3, EFCAB10, IARS
BCR
AFAP1, CAPG, DESI1, FAM170A, LPAR1, MUSK, SLC16A3, XRCC6
DOK1, LOXL3
AGPAT4
PPP3CC

Cluster 1 — Top 25 biological pathways

Top 25 biological pathways	−log(B-H p value)	Ratio
Protein kinase A signaling	2.66	0.08
Phospholipase C signaling	2.66	0.09
p53 signaling	2.62	0.13
Retinoic acid mediated apoptosis signaling	1.96	0.16
Apoptosis signaling	1.96	0.13
RhoA signaling	1.96	0.11
RhoGDI signaling	1.96	0.09
Hypoxia signaling in the cardiovascular system	1.96	0.14
CCR3 signaling in eosinophils	1.96	0.10
Androgen signaling	1.95	0.11
Sumoylation pathway	1.95	0.12
Unfolded protein response	1.95	0.15
fMLP signaling in neutrophils	1.71	0.10
nNOS signaling in neurons	1.71	0.15
Calcium-induced T lymphocyte apoptosis	1.67	0.13
Death receptor signaling	1.67	0.11
Nitric oxide signaling in the cardiovascular system	1.67	0.10
Adipogenesis pathway	1.48	0.09
Nur77 signaling in T lymphocytes	1.46	0.13
VEGF signaling	1.46	0.10
Pancreatic adenocarcinoma signaling	1.46	0.09
UVA-induced MAPK signaling	1.46	0.10
Regulation of cellular mechanics by calpain protease	1.41	0.13
Huntington's disease signaling	1.38	0.07
Breast cancer regulation by stathmin1	1.31	0.07

Fig. 4 Prioritized enriched biological pathways and transcription factor (TF) motifs analysis. **a** Prioritized biological pathways ranked by the number of occurrences in the eight clusters with significantly enriched pathways (left) and DMGs with associated DMERs in "Huntington's Disease Signaling" pathway (right). **b** Prioritized TF motifs ranked by the frequency in 13 clusters (left). The significantly enriched de novo TF binding motifs are shown on the right

HIP1 deficiency unexpectedly decreased RA FLS invasion by nearly 50% ($n = 7$ separate FLS lines, p-value <0.001 (paired t-test), Fig. 5b). Morphology of HIP1-deficient RA FLS was then evaluated by confocal microscopy (Fig. 5c, d). HIP1 was depleted using siRNA, which decreased expression of the 116 kDa band on western blot but not other control proteins. HIP1-deficient cells lost their fusiform morphology and assumed a 'stellate-like' shape. Actin fibers became less organized and lamellipodia-like structures did not form or were present in a non-polarized manner without phospho-FAK co-localization. This phenotype is commonly observed in cells that cannot migrate or invade and suggests that HIP1 plays a role in FLS-induced invasion into cartilage in RA[23].

Discussion

Although treatment of RA has improved dramatically in recent years, current targeted therapies usually fail to induce remission and many patients do not respond to any available agent. Most efforts to develop new therapies rely on candidate gene approaches that have emerged from traditional biological research that dissect pathogenesis. This approach has limited potential to discover entirely new and unanticipated pathogenic processes that contribute to disease. To address this need, we developed genome-wide unbiased methodology and datasets that can identify surprising pathways and genes involved in immune dysfunction. By integrating multiple epigenetic technologies we hoped to define pathogenesis of disease and discover non-obvious targets.

The present studies focused on RA FLS which, unlike OA or normal synoviocytes, display an imprinted aggressive phenotype and invade cartilage explants when implanted in immunodeficient mice[24]. RA FLS can grow under anchorage-independent conditions, are less susceptible to contact inhibition, and are resistant to apoptosis[25, 26]. They are more invasive in vitro and produce cytokines and MMPs that serve as amplifying disease mechanisms[27]. Imprinted RA FLS can also migrate from joint to joint, potentially serving as a vector to disseminate the disease[28]. While the adaptive immune system clearly contributes to RA[17],
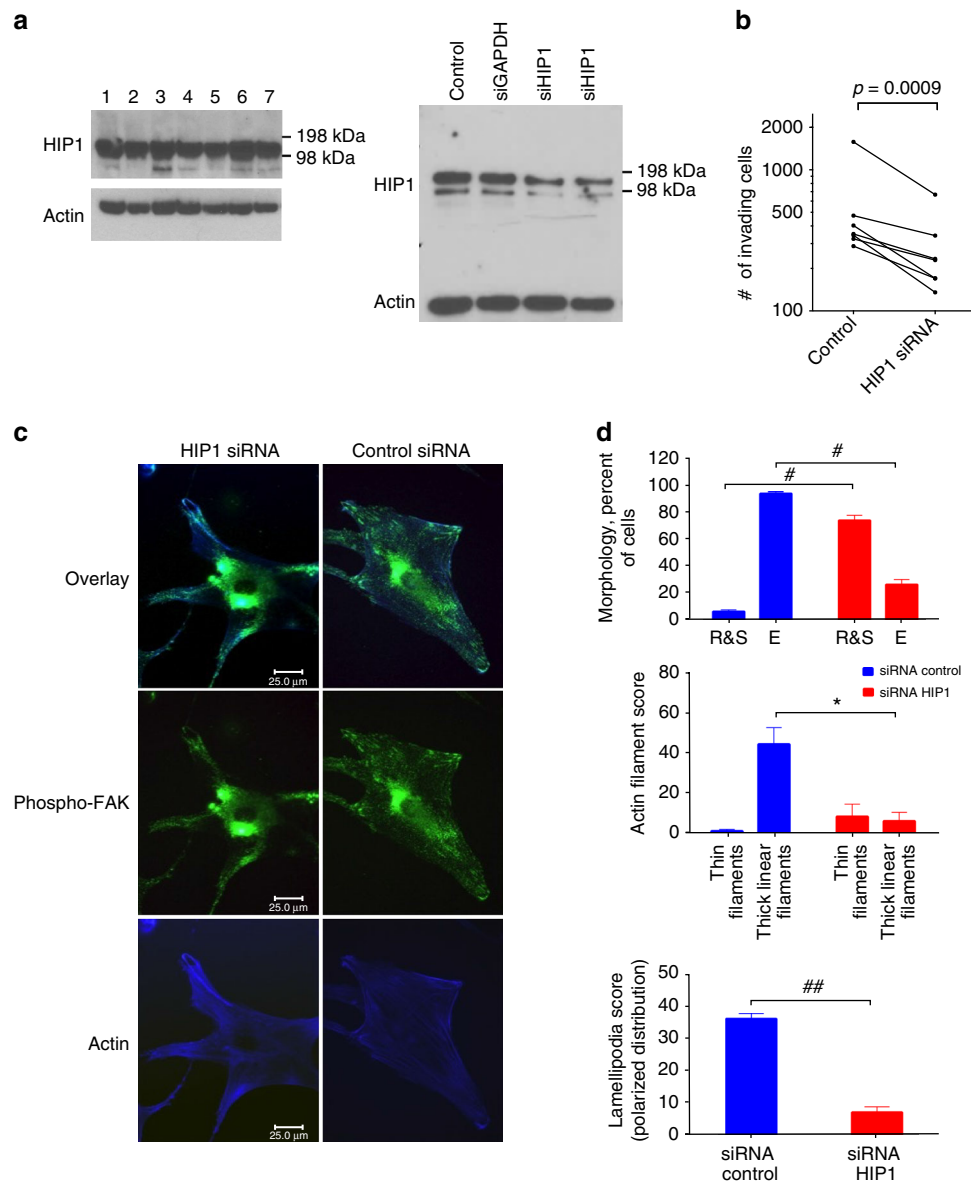
Fig. 5 Biologic validation of HIP1 as a key pathway in RA FLS. **a** Left: HIP1 protein expression in FLS; right: decreased HIP1 expression induced by siRNA knockdown. **b** HIP1 deficiency significantly reduced RA FLS invasion through Matrigel. **c** HIP1 and cytoskeletal rearrangement. Left three panels: morphologic differences in HIP1 deficient cells demonstrate a stellar morphology with reduced number and size of lamellipodia and reduced co-localization of phospho-FAK. Right three panels: RA FLS transfected with siRNA control (green = phospho-FAK; blue = phalloidin/actin filaments; ×600 magnification). **d** Altered phenotype of the HIP1-deficient FLS. Top panel: siRNA HIP1 had reduced cell elongation (E elongated; R&S round and stellar morphology; #p-value = 0.0006). Middle panel: siRNA HIP1 decreased thin central filaments and thick actin filaments. *p-value = 0.003). Bottom panel: siRNA HIP1 knockdown also reduced the number of lamellipodia and their polarized distribution. (##p-value = 0.0008). All p-values are calculated by paired t-test, and error bars indicate standard deviations

innate immune functions regulated by FLS also participate[29] but have not been successfully targeted in humans.

We generated the largest collection of epigenomes for RA FLS by profiling six histone modification patterns, open chromatin regions, RNA expression and whole-genome DNA methylation and established the first high-resolution global epigenomic landscape for RA. To address the complex multidimensional relationships of those epigenomes and identify co-modified regions, integrative analysis was performed using a novel unbiased method, EpiSig. The analysis grouped the RA genome into 125 clusters based on epigenetic marks in multiple regulatory/functional elements. This is the first time whole-genome DNA methylation, chromatin accessibility, and the transcriptome

have been incorporated into the integration method with histone modifications which allows a detailed understanding of the epigenomic landscape.

Our systematic integrative analysis identified 13 of the 125 clusters with statistically significant DMER enrichment for pathway analysis, which allowed us to prioritize biological pathways featured in promoter and enhancer regulatory modules. A variety of pathways were found, some of which were expected because they are related to immune responses or matrix regulation or were previously identified by our previous limited analyses. However other pathways were totally unexpected, such as "Huntington's Disease Signaling" and a variety of others that still need to be explored. In the "Huntington's Disease Signaling"

pathway, many genes that could serve as potential targets. Previous studies found HIP1 is associated with cell movement in cancer[19, 20]. However, HIP1 is particularly prominent in this pathway and its potential involvement in oncology led us to reason that this might be especially interesting and worth exploring further. The fact that we were able to validate the role of a key protein in the "Huntington's Disease Signaling" pathway indicates that our unbiased method can identify novel potential therapeutic targets. Other genes that are differentially marked in this pathway include phosphatidylinositide 3-kinases, activator protein-2, histone deacetylases, and heat shock protein-70. We are particularly intrigued by the potential use of histone deacetylase inhibitors to remodel the epigenome and alter enhancer activity, thereby modifying aggressive RA FLS behavior. In addition, transcription factor motifs were identified that distinguish RA from OA and the bZip motifs were particularly noted as enriched in DMERs. These motifs play a critical role in the regulation of metalloproteinases and cytokines that are pathogenic in RA.

Because, we have previously evaluated joint location-specific changes in RA[16], we also assessed the differences between hip and knee FLS lines ($n = 5$ and 3, respectively for WGBS data). However, the present study was underpowered due to limited sample size and identified <200 DMERs using the same criteria. The previous studies showed that the magnitude of the differences between joints within RA is much less than the difference between OA and RA. Thus it is not surprising that many more samples are required to evaluate more subtle joint location-specific marks, and it is unlikely that these modest differences affect the overall RA vs. OA analysis.

The use of FLS permitted evaluation of a homogeneous cell population of primary cells from the site of disease and avoided the confounding problem of deconvoluting lineage-specific marks in mixed cell populations. The comparators for this analysis were OA FLS, which are derived at the time of joint replacement and are a commonly used comparator for RA because matched normal synovial cells are not readily available. We have previously shown that analysis of OA FLS DNA methylation patterns do not identify any systematic differences compared to normal that would affect the present study[4]. The use of cultured cells rather than freshly isolated cells could potentially influence the results, but we have previously shown that the DNA methylation signature is remarkably stable in RA FLS for many months in culture and they retain the aggressive phenotype[4]. Despite these potential limitations, the present studies provide a full dataset required for detailed analysis of primary cells isolated directly from the site of this destructive disease.

Our methodology focused on RA, but it is disease agnostic and can be applied to any immune-mediated disease if the datasets are available. The use of FLS simplified our studies because they are a homogenous cell lineage and play a critical role in rheumatoid joint damage. Studies on peripheral blood cells or other cells in disease tissue would require separation into individual lineages because single cell assays are not feasible with current technology for most of the marks studied. Future studies are required to validate the large number of unanticipated targets and pathways as well as explore diversity between patients and individualize treatment.

Methods

Human FLS. This study was approved by the Institutional Review Board of University of California, San Diego, and informed consent was obtained from all participants. Synovial tissue was obtained from patients with RA and OA at the time of total joint replacement or synovectomy. The diagnosis of RA conformed to American College of Rheumatology 1987 or 2010 criteria[30, 31], and the diagnosis of OA conformed to American College of Rheumatology 1986 or 1991 criteria[32, 33]. The mean ages of RA and OA patients were 55 ± 9 and 66 ± 12, respectively. Joint

locations were available for 10 RA patients (6 knee, 3 hip, 1 hand) and 11 OA (8 knee, 3 hip). Of the 7 RA patients with clinical information available, treatment included NSAIDs ($n = 5$), low dose prednisone (<5 mg/d) ($n = 4$), methotrexate ($n = 4$), methotrexate plus leflunomide ($n = 1$), TNF blocker ($n = 3$), IL-6 blocker ($n = 1$). Biologics, methotrexate and leflunomide were generally discontinued 4 weeks before surgery. Serology was available for 3 RA patients. One patient was RF positive and 2 patients were CCP and RF negative. Therapy information was available for 5 OA patients, who were treated with NSAIDs ($n = 2$) or analgesics. Enzymatic disaggregation of synovium was performed as previously described[14]. Cells were allowed to adhere overnight and non-adherent cells were removed. Adherent FLS were split at 1:3 when 70–80% confluent and used from passages 4 through 7, when they are a homogeneous population of fibroblasts. Primary FLS were cultured (at 5% CO_2, 37 °C) in Dulbecco's modified Eagle's medium (DMEM) supplemented with L-glutamine, gentamicin, penicillin/streptomycin, and 10% heat-inactivated fetal bovine serum (FBS) as previously described[34]. For synchronization experiments, cells were serum starved for 2 days and transfected with siRNA in DMEM containing 1% FBS and supplements, followed by addition of 10% FBS.

RNA-seq data processing. Total RNA was extracted and the quality of all samples was evaluated using an Agilent Bioanalyzer. The samples had an average RNA Integrity Number (RIN) of 9.4 with a minimum of 7.5. Sequencing libraries were prepared using TruSeq Stranded Total RNA RiboZero protocol from Illumina. Libraries were pooled and sequenced with an Illumina HiSeq 2000. Raw read quality was evaluated using FastQC.

Adapter and low quality bases below a quality score of 15 were trimmed from raw RNA-seq reads. After trimming, reads with less than 30 bp were further discarded. The remaining reads were aligned to human reference genome hg19 using STAR (2.3.0) and assembled and quantified by HTSeq (0.5.4p5). DEGs were identified using DESeq2 package in R. To be considered a DEG, twofold change of gene expression levels between RA and OA should be achieved and the B-H adjusted p-value is <0.05. For the following analysis, transcription levels were then converted to log2 of the normalized counts.

ChIP-seq data processing. The sample preparation for ChIP-seq was performed using the Zymo-spin ChIP kit (Zymo Research, Irvine, CA) according to the manufacturer's instructions. Raw ChIP-seq reads of 11 RA and 11 OA samples, each with six histone modifications (H3K4me1, H3K4me3, H3K9me3, H3K27ac, H3K27me3, and H3K36me3) and input data, were mapped to hg19 genome using bwa (0.7.7-r441). Peaks were called using MACS2 (2.1.0.20150420.1) using a p-value cutoff of 0.1 for broad marks (9me3, 27me3, 36me3) and 0.01 for narrow marks (4me1, 4me3, 27ac). To assess the library complexity and the enrichment of the data, quality control was done using several metrics, including non-redundant fraction of reads, fraction of reads that fall into peak regions, PCR bottleneck coefficient, normalized strand cross-correlation coefficient and the relative strand cross-correlation coefficient. ChIP-seq DMERs were determined by DiffBind package in R with q-value <0.05 for six marks.

ATAC-seq and WGBS data processing. ATAC-seq sample preparation was followed by previous protocol[35] and was sequenced by Epinomics (San Francisco, CA). ATAC-seq data were processed using Bowtie and read depth was normalized. ATAC-seq DMERs were determined by DiffBind package in R with q-value <0.05. The bisulfite conversion and sequencing were performed by GATC Biotech AG (Germany). WGBS data were processed with DSS package in R. Differentially methylated loci (DMLs) were determined and Differentially methylated regions were defined by callDMR function with delta beta of DMLs >0.05 and q-value <0.0015.

Genome-wide multidimensional clustering by EpiSig. EpiSig is derived from the previously published method ChromaSig[36] with greatly improved performance and capability in capturing sequencing profile patterns among large number of datasets. Compared to previous ChromaSig developed in Perl script, EpiSig was rewritten with C++ in a single standalone program and provides two running modes, which significantly improved speed and performance. Regarding to searching accuracy, EpiSig is an unsupervised learning approach to simultaneously clusters, aligns sequencing patterns without prior knowledge. Before the pattern alignment, one important step is to find the initial seeds among the remained loci. EpiSig provides another information entropy based method to find the candidate seed loci. First, the maximum intra-entropy loci (enriched signal window) are selected, then the minimum inter-entropy loci are added to the seed set. Actually they measure the similarities among them. To handle the different sequencing types such as histone modification data, RNA-Seq data and other data types, EpiSig utilizes a bi-clustering thinking to concatenate different data types into one larger data window when aligning them. This provides a more meaningful and accurate pattern discovering approach.

To assess the co-modified regions in RA and non-RA, all sequencing datasets were processed, normalized and filtered ensuring a uniform description of the information based on read depth. For final data processing, inputs were subtracted from the histone mark data. The comprehensive dataset is then preprocessed by

EpiSig by dividing the data into binned binary files. After this, EpiSig performs data normalization using a sigmoid function, followed by scanning the data for enriched regions. To facilitate interpretation of the result, these individual EpiSig clusters were further combined into sections using a SOM algorithm.

Genome and chromosome annotation of targeted regions. The human genome hg19 was used as reference genome with functional annotation in GENCODE v19. CpG islands and TSSs were downloaded from UCSC table browser. The genome annotation was performed by overlapping targeted regions with gene exon, intron, intergenic region, 3′- and 5′-UTR, TSS, 1 kb upstream of TSS, and CpG island. The coverage is defined as the percentage of overlapped regions. The chromosome annotation was calculated as the percentage of 5 kb regions in each chromosome and followed by chromosome length and cluster size normalization.

Pairwise correlations of DMERs and gene expressions. To plot correlations between DMERs and gene expressions, DMERs (from six histone modifications, open chromatin regions, and DNA methylation) were assigned to nearby genes using GREAT analysis with default settings[37]. If multiple DMERs were assigned to the same gene, the average of fold changes was calculated. Log2 fold change was used for the plot except for WGBS.

Pathway prioritizing. To prioritize enriched pathways, stringent cutoff of q-value <0.01 was applied to DMERs from six histone modification marks and open chromatin region. Because different analysis methods were used for identifying DEG and DNA methylation DMERs, no additional filter was applied. Enriched DMERs (from six histone modifications, open chromatin regions and DNA methylation) were assigned to nearby genes using GREAT analysis with default settings. The DMER assigned genes together with DEGs were collected for pathway analysis in each cluster. However, we did not assign different weights to DEGs/DMERs. To test for significant enrichment of DMER in each cluster, hypergeometric test was performed using R package. For prioritizing pathways, clusters with hypergeometric test q-value <0.05 and number of associated genes >200 were selected. Ingenuity pathway analysis (IPA) (QIAGEN Redwood City) was used for pathway analysis. Finally, pathways were prioritized based on their frequency in the enriched clusters.

Motif discovery. Motif analysis was performed in 13 clusters using HOMER[38]. Due to small input regions, q-value of 0.1 was used as the cutoff of known motif enrichment. de novo motif was also discovered and scores were calculated indicating matches to known motifs. de novo motifs with p-value <1e−12 and score >0.7 were listed.

siRNA knockdown of HIP1. Dharmacon SMARTpool siRNA targeting HIP1, GAPDH, and a non-coding control were purchased from Thermo Scientific (Lafayette, CO) and transfected into RA FLS according to the manufacturer's instructions. Briefly, FLS derived from RA FLS were transfected siRNA in the presence of DharmaFECT reagent 1 in DMEM media with NEAA. Two days later, cells were used for invasion experiments or immunofluorescence. Aliquots were used for cell lysis to confirm gene knockdown with western blot and qPCR.

FLS invasion assay. Invasion was assayed in a transwell system using Matrigel-coated inserts (BD Biosciences, Franklin Lakes, NJ), as previously described[39, 40]. FLS in serum free medium were placed in the upper compartment of the Matrigel-coated inserts. The lower compartment was filled with 10% FBS and incubated for 24 h. The insert was stained with crystal violet and the total number of cells that invaded through Matrigel was counted using ImageJ cell counter software.

Western blot analysis. A quantity of 5 μg of FLS protein/lane were loaded into a NuPAGE 10% Bis-Tris gel (Invitrogen) in the presence of MES buffer (Invitrogen) and electrophoresed under reducing conditions. After transfer, the membranes were incubated with anti-HIP1 rabbit monoclonal antibody (Abcam, Cambridge, MA). HRP-conjugated anti-rabbit IgG antibody (GE Healthcare, Pittsburgh, PA) was used as secondary antibody. Proteins were visualized with Clarity Western ECL substrate (Bio-Rad, Hercules, CA). Rat anti-actin antibody (Cell Signaling technology, Danvers, MA) was used as loading control.

HIP1 gene expression. qPCR was performed as previously described[41]. Ct (threshold cycle) values were adjusted for Ubiquitin in each sample (ΔCt), and fold differences were calculated with the 2−ΔΔCt method.

Confocal immunofluorescence microscopy. RA FLS were cultured on coverslips to 20–30% confluence, transfected for 6 h with HIP1 or control siRNA, serum starved overnight and then treated for 5, 15, 30 min with complete media with 10% FBS. Cells were then fixed with 4% formaldehyde for 15 min and permeabilized with PBS/Triton X-100, 0.1% for 5 min. Non-specific binding was blocked with 5% nonfat milk. A rabbit anti-phospho-FAK (Y397) (Abcam) was used as primary antibody, and a Alexa Fluor 488 (green) donkey anti-rabbit IgG (Invitrogen) used

as secondary antibody. Alexa Fluor 350 (blue) Phalloidin (Invitrogen) was used to stain the actin filament. A Leica DMi8 fluorescent microscope system was used for visualization with the appropriate filters, with Leica Lax X software. Cell cytoskeleton was scored as previously described[39] using a ×600 magnification.

References

1. Asif Amin, M., Fox, D. A. & Ruth, J. H. Synovial cellular and molecular markers in rheumatoid arthritis. *Semin. Immunopathol.* **39**, 385–393 (2017).
2. Richardson, B. C. & Patel, D. R. Epigenetics in 2013. DNA methylation and miRNA--key roles in systemic autoimmunity. *Nat. Rev. Rheumatol.* **10**, 72–74 (2014).
3. Nielsen, H. M. & Tost, J. Epigenetic changes in inflammatory and autoimmune diseases. *Subcell. Biochem.* **61**, 455–478 (2013).
4. Whitaker, J. W. et al. An imprinted rheumatoid arthritis methylome signature reflects pathogenic phenotype. *Genome Med.* **5**, 40 (2013).
5. Ai, R. et al. DNA methylome signature in early rheumatoid arthritis synoviocytes compared with longstanding rheumatoid arthritis synoviocytes. *Arthritis Rheumatol.* **67**, 1978–1980 (2015).
6. Araki, Y. et al. Histone methylation and STAT-3 differentially regulate interleukin-6-induced matrix metalloproteinase gene activation in rheumatoid arthritis synovial fibroblasts. *Arthritis Rheumatol.* **68**, 1111–1123 (2016).
7. Nakano, K., Whitaker, J. W., Boyle, D. L., Wang, W. & Firestein, G. S. DNA methylome signature in rheumatoid arthritis. *Ann. Rheum. Dis.* **72**, 110–117 (2013).
8. Bartok, B. & Firestein, G. S. Fibroblast-like synoviocytes: key effector cells in rheumatoid arthritis. *Immunol. Rev.* **233**, 233–255 (2010).
9. Song, X. & Lin, Q. Genomics, transcriptomics and proteomics to elucidate the pathogenesis of rheumatoid arthritis. *Rheumatol. Int.* **37**, 1257–1265 (2017).
10. Dozmorov, M. G., Wren, J. D. & Alarcón-Riquelme, M. E. Epigenomic elements enriched in the promoters of autoimmunity susceptibility genes. *Epigenetics* **9**, 276–285 (2014).
11. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods.* **9**, 215–216 (2012).
12. Hoffman, M. M. et al. Unsupervised pattern discovery in human chromatin structure through genomic segmentation. *Nat. Methods* **9**, 473–476 (2012).
13. Ernst, J. et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43–49 (2011).
14. Hammaker, D. et al. LBH gene transcription regulation by the interplay of an enhancer risk allele and DNA methylation in rheumatoid arthritis. *Arthritis Rheumatol.* **68**, 2637–2645 (2016).
15. Matsuda, S. et al. Regulation of the cell cycle and inflammatory arthritis by the transcription cofactor LBH gene. *J. Immunol.* **199**, 2316–2322 (2017).
16. Ai, R. et al. Joint-specific DNA methylation and transcriptome signatures in rheumatoid arthritis identify distinct pathogenic processes. *Nat. Commun.* **7**, 11849 (2016).
17. Firestein, G. S. & McInnes, I. B. Immunopathogenesis of rheumatoid arthritis. *Immunity* **46**, 183–196 (2017).
18. Boyle, D. L. et al. The JAK inhibitor tofacitinib suppresses synovial JAK1-STAT signalling in rheumatoid arthritis. *Ann. Rheum. Dis.* **74**, 1311–1316 (2015).
19. Li, D. et al. Knockdown of HIP1 expression promotes ligand-induced endocytosis of EGFR in HeLa cells. *Oncol. Rep.* **38**, 3387–3391 (2017).
20. Wang, J. et al. Prognostic significance of huntingtin interacting protein 1 expression on patients with acute myeloid leukemia. *Sci. Rep.* **7**, 45960 (2017).
21. Tolboom, T. C. et al. Invasiveness of fibroblast-like synoviocytes is an individual patient characteristic associated with the rate of joint destruction in patients with rheumatoid arthritis. *Arthritis Rheum.* **52**, 1999–2002 (2005).
22. Laragione, T., Brenner, M., Mello, A., Symons, M. & Gulko, P. S. The arthritis severity locus Cia5d is a novel genetic regulator of the invasive properties of synovial fibroblasts. *Arthritis Rheum.* **58**, 2296–2306 (2008).
23. Laragione, T. & Gulko, P. S. Liver X receptor regulates rheumatoid arthritis fibroblast-like synoviocyte invasiveness, matrix metalloproteinase 2 activation, interleukin-6 and CXCL10. *Mol. Med.* **18**, 1009–1017 (2012).
24. Müller-Ladner, U. et al. Synovial fibroblasts of patients with rheumatoid arthritis attach to and invade normal human cartilage when engrafted into SCID mice. *Am. J. Pathol.* **149**, 1607–1615 (1996).

25. Lafyatis, R. et al. Anchorage-independent growth of synoviocytes from arthritic and normal joints. Stimulation by exogenous platelet-derived growth factor and inhibition by transforming growth factor-beta and retinoids. *J. Clin. Invest* **83**, 1267–1276 (1989).

26. Baier, A., Meineckel, I., Gay, S. & Pap, T. Apoptosis in rheumatoid arthritis. *Curr. Opin. Rheumatol.* **15**, 274–279 (2003).

27. Yamanishi, Y. et al. Regulation of joint destruction and inflammation by p53 in collagen-induced arthritis. *Am. J. Pathol.* **160**, 123–130 (2002).

28. Lefèvre, S. et al. Synovial fibroblasts spread rheumatoid arthritis to unaffected joints. *Nat. Med.* **15**, 1414–1420 (2009).

29. Kiener, H. P. et al. Cadherin 11 promotes invasive behavior of fibroblast-like synoviocytes. *Arthritis Rheum.* **60**, 1305–1310 (2009).

30. Arnett, F. C. et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum.* **31**, 315–324 (1988).

31. Aletaha, D. et al. 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Arthritis Rheum.* **62**, 2569–2581 (2010).

32. Altman, R. et al. Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. Diagnostic and Therapeutic Criteria Committee of the American Rheumatism Association. *Arthritis Rheum.* **29**, 1039–1049 (1986).

33. Altman, R. et al. The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. *Arthritis Rheum.* **34**, 505–514 (1991).

34. Rosengren, S., Boyle, D. L. & Firestein, G. S. Acquisition, culture, and phenotyping of synovial fibroblasts. *Methods Mol. Med.* **135**, 365–375 (2007).

35. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).

36. Hon, G., Ren, B. & Wang, W. ChromaSig: a probabilistic approach to finding common chromatin signatures in the human genome. *PLoS Comput. Biol.* **4**, e1000201 (2008).

37. McLean, C. Y. et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).

38. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).

39. Laragione, T., Brenner, M., Sherry, B. & Gulko, P. S. CXCL10 and its receptor CXCR3 regulate synovial fibroblast invasion in rheumatoid arthritis. *Arthritis Rheum.* **63**, 3274–3283 (2011).

40. Laragione, T. et al. The cation channel Trpv2 is a new suppressor of arthritis severity, joint damage, and synovial fibroblast invasion. *Clin. Immunol.* **158**, 183–192 (2015).

41. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods* **25**, 402–408 (2001).

Author contributions

G.S.F., W.W., R.A., T.L., P.S.G., S.N., and K.E.B. provided substantial contributions to study conception and design; D.H., K.M., D.P., and T.L. contributed to data acquisition; R.A., T.L., D.H., D.L.B., A.W., E.P., V.K., J.W.W., Y.B., S.N., K.E.B., R.I.A., M.W., B.D., P.S.G., W.W., and G.S.F. participated in data analysis and interpretation; The article was drafted and revised critically for important intellectual content by G.S.F., W.W., R.A., and P.S.G.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-04310-9.

Competing interests: E.P., V.K., D.P., J.W.W., Y.B., S.N., and K.E.B. are employees of Janssen Pharmaceuticals. All the remaining authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.