H E L I C A L

H2020-MSCA-ITN-2018-813545

HELICAL

Health Data Linkage for Clinical Benefit

Deliverable D2.5

WP2 Report Summary

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 813545

Table of Contents

Work package 2 Identification of key pathogenic pathways suitable for therapeutic targeting in Giant Cell Arteritis (GCA)

Lead: Ann Morgan (UNIVLEEDS)

Participants: CSIC, IBMZ, IDIBAPS, AX, IDIVAL, UNIVLEEDS

Long-term collaborations: University of Newcastle-Upon Tyne, University of Edinburgh, Imperial College London

The pathogenesis of giant cell arteritis (GCA) is incompletely understood, and therapy, largely dependent upon glucocorticoids, is blunt and associated with a heavy burden of adverse events. WP members performed the first large-scale genetic study of this disease identifying a tight association with single nucleotide polymorphisms (SNPs) in *HLA-DRB1*, *HLA-DQA1* and *HLA-B* within the major histocompatibility (MHC) locus. Several GCA risk loci that encode proteins involved in helper T-cell subset function (Th1, Th17, and regulatory T-cells (Treg)), such as *IL6*, *TNFA*, *IL17A* and *PTPN22*, were also identified at a sub-GWAS significance level. Meta-Immunochip analysis across the vasculitides highlighted the importance of *IL12B*. Furthermore, a GCA GWAS identified *PLG* and *P4HA2*, involved in vascular remodelling and angiogenesis, at genome-wide significance, highlighting the importance of studying vascular and matrix biology in addition to immunological pathways.

AIM:

The primary aim of this WP was to assemble and further mine international GCA molecular and clinical databases, in conjunction with public genetic, genomic and epigenomic databases, to investigate the top loci from the recent Immunochip and GWA studies: to identify genotype-phenotype correlations, new genetic associations and pathogenic mechanisms through polygenic risk score and pathway analysis, and to examine the functional effects of these genomic variants on immunological and pathogenic processes by performing both *in silico* analyses and functional studies on disease tissue. We achieved this by combining 5 ESR projects in a programme of genetic, genomic, epigenomic and

functional work linking existing molecular data with GCA phenotype, taking advantage of Bayesian statistical methods and network-based machine learning approaches, as well as validating key findings in *ex-vivo* model systems.

EXPERIMENTS PLANNED AT MID-TERM REPORT: ESR2 (IBMZ) had moved from WP1 to align with genomic workflow and computational expertise. They leveraged the world-leading expertise in machine learning approaches at IBMZ to study the T-cell receptor (TCR) repertoire to derive insights into the role of antigen-specific T-cells in the pathogenesis of GCA. This will provide support for involvement of T-cells in GCA pathobiology and further study of relevant T-cell phenotypes, ultimately leading to clinical trials of therapeutics targeting T-cells. **ESR7** was based in CSIC (Granada), where most of the recent European genetic effort in GCA was coordinated. Using an extensive bank of existing genotyped and clinically phenotyped samples from large international GCA and systemic vasculitis cohorts and ongoing sample collections from IDIBAPS and UNIVLEEDS, CSIC performed a meta-analysis of GWAS data to define variants related to discrete vasculitis phenotypes, fine mapping of GWAS hits to identify causal variants and integration of genetic, epigenetic and gene expression data to further explore functional significance of GWAS results. **ESR8** (UNIVLEEDS) initially prioritised genomic (RNASeq) analysis of GCA tissue to unravel the molecular pathways contributing to immune and vascular remodelling phenotypes. Pathway analyses of GCA molecular traits and immune cell infiltration were performed using an agnostic approach. Publicly accessible transcriptomic and metabolomic data will be used to generate polygenic risk scores (PRS) of vascular traits of interest. Traits analysed will include expression of immune and vascular genes (e.g. from the Genotype-Tissue Expression (GTEx) Project) and markers of inflammation and relevant vascular and tissue (e.g. from UK Biobank). **ESR9** (AX) exploited and extended systems biology approaches developed by this SME to address the following tasks: (1) interpret Real World Evidence derived from public datasets (in particular NHANES) using protein functional network analysis, to define potential molecular mechanisms that could explain such relationships. From these mechanisms one can derive potential biomarkers that could help establish the relevance of

the identified relationship. (2) applied protein functional network analysis for functional understanding of genetic results from ESR10/8 and time permitting, (3) identify (auto) antigens in GCA recognised by disease-associated HLA alleles. The analysis of molecular mechanisms in the context of a human interaction network can unveil functional properties and pathogenic insights that are otherwise hidden, and that can be further validated *in vitro*, *in vivo* or by other *in silico* approaches, including re-purposing of existing pharmacological agents or targeted therapies. The IDIBAPS team trained **ESR10** in models developed to functionally explore GCA pathogenic pathways identified from the systems biology approaches above. She initially explored pathways triggered by the IL-6 cytokine family, based on previous immunopathology studies, further reinforced by recent genetic studies and clinical trials of therapeutics targeting these pathways. Later studies will include functional analysis of vascular smooth muscle cells from freshly-isolated TAB, co-culture systems including the main cell types present in inflamed arteries and culture of TAB in 3D matrix, as appropriate for the gene/pathway of interest. In these model systems changes in transcripts, proteins and activation/inactivation of signalling pathways will be explored, including manipulation of pathogenic pathways with targeted therapies.

Thus, WP2 will drive a workflow linking ESRs in the three leading GCA investigator laboratories in Europe with an SME and a major industry leader at the forefront of commercial genomic methodology, exploiting the overarching consortium strength in ethical linkage of experimental to clinical data, and utilising readily available samples and genetic data from prior studies conducted by these research groups.

EXPERIMENTS CONDUCTED DURING THE SECOND HALF OF THE ACTION:

Identification of Novel Genetic Loci associated with GCA

In light of delays generating new genotypic data in GCA due to the COVID-19 pandemic, CSIC in collaboration with IDIBAP and UNIVLEEDS undertook a cross-vasculitis association study in 8,467 patients with vasculitis compared with 29,795 healthy individuals. Comprehensive assessment of the genetic overlap between the major vasculitis phenotypes including GCA

(n=2,134), Takayasu arteritis (TAK; n=1,091), Kawasaki disease (KD; n=405), IgA vasculitis (IgAV; n=215), Behçet's disease (BD; n=3,197), and ANCA-associated vasculitis (AAV: combined as a single phenotype; n=914) was undertaken. The MHC was excluded from this analysis since it had been explored in depth previously. In total 85 genetic variants at 12 genomic regions were associated with two or more of the vasculitides at genome-wide significance.

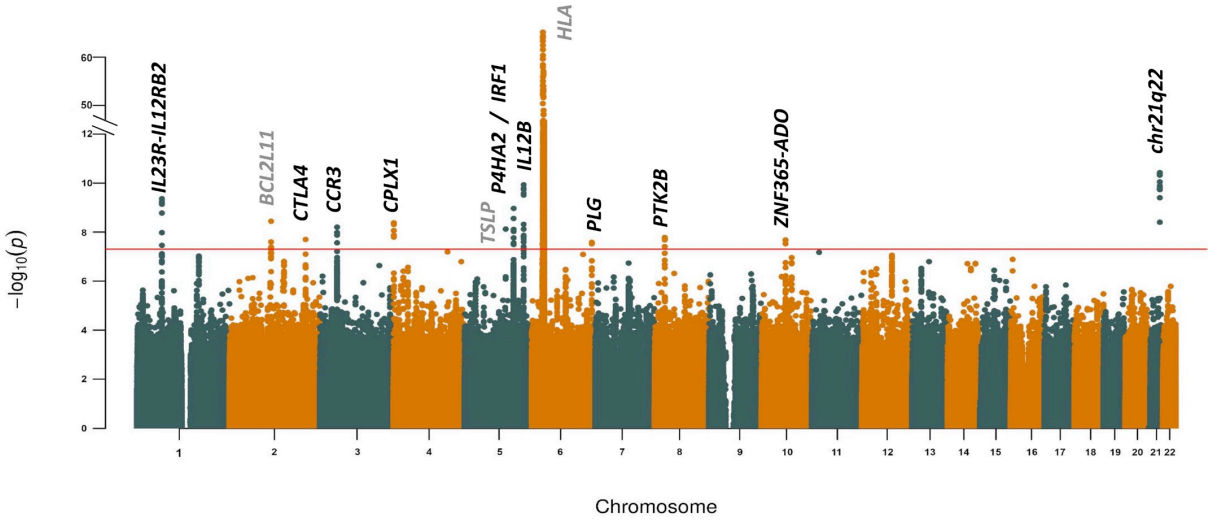

Figure 1. Manhattan plot showing the results of the cross-phenotype meta-analysis. The −log10 of the p values are plotted against their physical chromosomal position. The red line represents the genome-wide level of significance ($p < 5\times10^{-8}$). Loci reaching the significant threshold are annotated in the plot. Loci representing new shared risk loci in vasculitis are highlighted in bold.

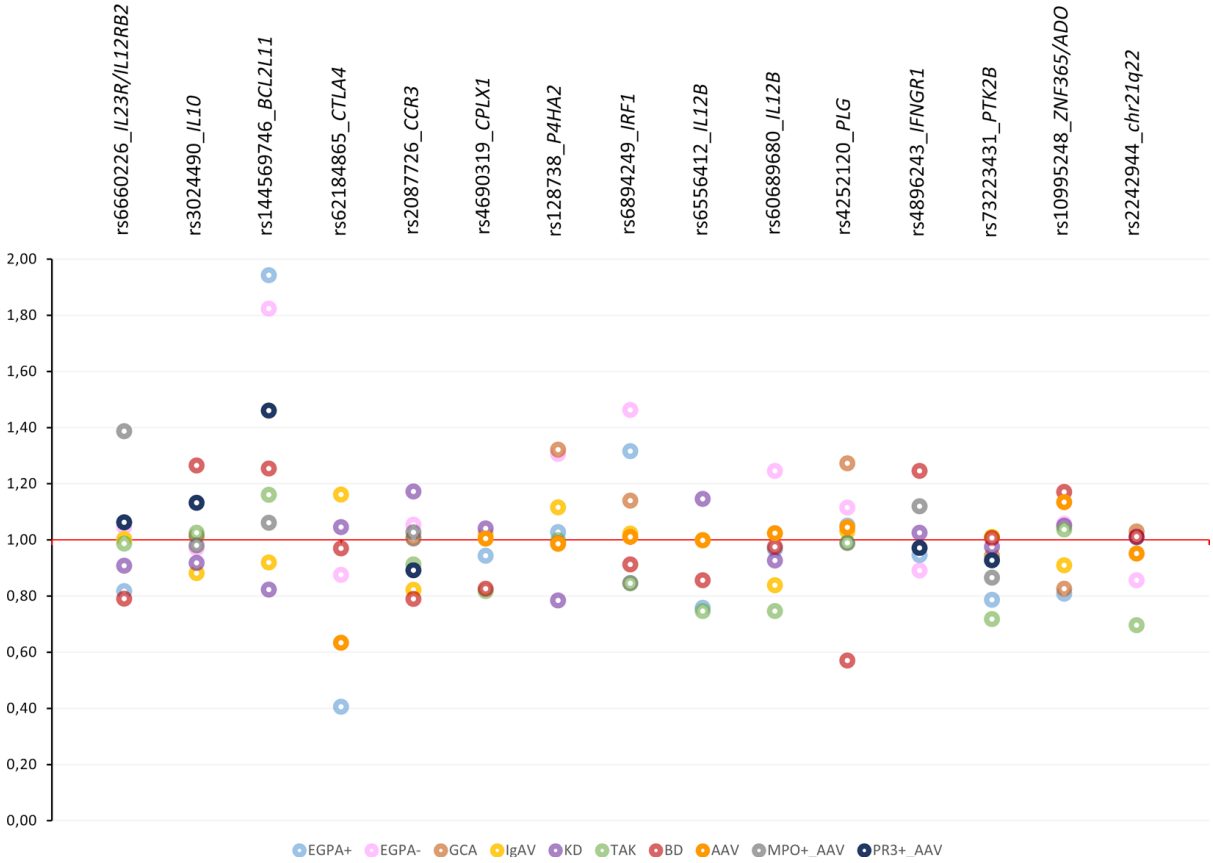

Figure 2. Novel risk loci shared across vasculitides. Effect of the independent genetic variants reaching genome-wide significant level in the subset-based meta-analysis is shown. Circles represent the analyzed phenotypes of vasculitides.

This study confirmed the association of *P4HA2* and *PLG* with GCA and identified *ADO* as a new genetic locus for GCA. This gene encodes a dioxygenase involved in amino acid metabolism. Functional analyses performed *in silico* predicted putative enhancers and chromatin interactions for each of these 3 loci and a possible eQTL for both *P4HA2* and *ADO*.

ESR7 subsequently explored epigenetic signatures in monocytes and ESR8 undertook a transcriptome-wide association study and explored transcriptional changes in TAB (see below). A drug repurposing analysis with the aim of identifying novel therapies for the analysed vasculitides was performed. In total, 103 drugs that could potentially be repurposed in vasculitis were identified, 13 of which are currently indicated for immune-mediated disorders. However, disappointingly, no hits were observed for the genes

specifically associated with GCA, highlighting the importance of gaining further molecular insights from the additional studies outlined in this report.

Prompt diagnosis of GCA is important to avert visual loss. False-negative TAB can occur due to skip lesions and historically many UK patients with typical symptoms did not undergo a TAB. If only TAB-positive cases are included in GWA studies, as has been performed to date, this markedly reduces the sample size. On the other hand, without vascular imaging, GCA may be over-diagnosed in TAB-negative cases, but it was unclear how often this occurred. An unbiased test is one way to address an imperfect reference standard. The genome-wide Manhattan plot for GCA susceptibility is dominated by the signal from the HLA region on chromosome 6 (chr6: 29–34 Mb on build 36/hg18) and we therefore chose to focus on the HLA region to estimate the extent of overdiagnosis in the UK before widespread adoption of temporal artery ultrasound as a first-line diagnostic test.

After QC, HLA alleles were imputed from genome-wide genotypic data from 663 patients diagnosed with GCA between 1991 and 2014 were compared with data from 2619 population controls. TAB-negative GCA (n=147) and GCA without a TAB performed (n=160) had variant frequencies intermediate between those of TAB-positive GCA and population controls. Associations with the three SNPs (rs9268969, rs9275184, rs477515) previously reported as being associated with TAB-confirmed GCA were also examined. Making several strong assumptions, we estimated that around two-thirds of TAB-negative cases and around one-third of cases without TAB result may have been over-diagnosed. From these data, TAB sensitivity was estimated at around 88%. Consequently, all cases with a clinical diagnosis of GCA were included in downstream GWAS to enhance the power of subsequent studies.

We also considered recently reported somatic mutations in *UBA1*, which are associated with VEXUS syndrome, however, there was no evidence of misdiagnosis in our UK GCA cohort.

CSIC in collaboration with IDIBAP and UNIVLEEDS then sought to expand the GCA GWAS cohort and collated a total of 3,498 patients with a clinical diagnosis of GCA and 15,550 healthy donors from ten populations of European ancestry. Through the largest genetic study performed in GCA, to date, we identified three new genetic loci. Two loci were located in genes related to the vascular endothelial growth factor (VEGF) pathway, namely *MFGE8*, encoding lactadherin, and *VTN*, encoding vitronectin; and the third locus was located in the gene *CCDC25*, which codifies a receptor of neutrophil extracellular traps (NETs). Functional annotation showed the GCA-associated loci acted as regulatory variants influencing gene expression in vascular tissue and immune cells. Furthermore, we also found a significant enrichment in histone marks in several immune cell types, particularly in natural killer cells, a cell type that has not been extensively studied in GCA, to date. Interestingly, the results of the drug repurposing analysis evidenced abciximab, an antagonist of the vitronectin protein and approved for the treatment of diabetes mellitus and acute coronary syndrome, as a potential therapeutic for GCA. Finally, a PRS comprising 28 genetic variants identified a fraction of individuals with more than three times the risk of developing GCA.

GCA is a complex disease mediated by multiple genetic factors. To increase statistical power, polygenic risk scores (PRS) were built as genotypic scores for intermediate and related phenotypes and then used to test for associations with GCA. The scores were generated using the GENOSCORES platform (https://pm2.phs.ed.ac.uk/genoscores/) developed by collaborators at the University of Edinburgh, which contains a database of publicly available well-powered GWA Studies with SNP to trait associations along with a software suite that was used to compute the scores and run downstream regression analysis. Our polygenic risk score (PRS) analyses identified several preliminary expression, protein or metabolic quantitative trait loci that were associated with GCA. The results are undergoing further verification and the most significant traits will be validated in the Spanish GCA cohorts and causality explored further by undertaking Mendelian Randomisation analyses.

Utilising genetic strategies to explore T-cell repertoires in GCA

Custom genetic assays allow exploration of somatic mutations, such as TCR recombination events that occur during T-cell development. The activity of the adaptive immune system is governed by T-cells and their specific TCR, which selectively recognise foreign antigens. Recent advances in experimental techniques have enabled sequencing of TCRs and their antigenic targets (epitopes), facilitating research into the missing link between TCR sequence and epitope binding specificity. Scarcity of data and a large sequence space make this task challenging, and to date only models limited to a small set of epitopes have achieved good performance.

IBMZ initially established a generic model for TCR-epitope binding prediction, called TITAN (Tcr epITope bimodal Attention Networks), a bimodal neural network that explicitly encoded both TCR sequences and epitopes to enable the independent study of generalisation capabilities to unseen TCRs and/or epitopes. The best performances were seen when the full TCR sequence was included (rather than restricting to the CDR3 region) along with atom-level representation of the antigenic peptide (SMILES sequences). This representation enabled pretraining the model with drug-protein interactions, resulting in a significantly improved performance (ROC-AUC 0.87 in 10-fold cross-validation) beating the current state-of-the art models (ImRex). The pipeline is generic and easily adaptable to all text-based prediction models and extracts a set of binding rules for clusters of TCRs.

Many different solutions to predicting the cognate epitope target of a TCR have been proposed. However, several questions on the advantages and disadvantages of these different approaches remain unresolved, as most methods have only been evaluated within the context of their initial publications and data sets. We reported the findings of the first public TCR-epitope prediction benchmark performed on 23 prediction models in the context of the ImmRep 2022 TCR- epitope specificity workshop. This benchmark revealed that the use of paired-chain alpha-beta, as well as CDR1/2 or V/J information, when available,

improved classification obtained with CDR3 data, independent of the underlying approach. In addition, we found that straight-forward distance-based approaches achieved a respectable performance when compared to more complex machine-learning models. Finally, we highlighted the need for a truly independent follow-up benchmark and provided recommendations for the design of such a benchmark.

We then sought to explore the repertoire of TCRs found in GCA patients to determine if this may provide insights into the pathogenesis of GCA. High-throughput TCR sequencing techniques were performed at UNIVLEEDS to enable IBMZ to examine TCR repertoires of 73 GCA patients compared with 70 controls with no autoimmune or inflammatory disease, 39 paroxysmal nocturnal haemoglobinuria (PNH) patients and 22 patients with aplastic anaemia (AA). The latter 2 haematological diseases are recognised to have clonal expansions and served as positive control disease states. An in-depth analysis of the peripheral blood TCR repertoires was performed using established TCR clustering methods and simple clustering algorithms. GCA patients were demonstrated to have reduced TCR diversity (species richness, Shannon diversity and Simpson diversity) compared with selected age-matched controls. This indicated that GCA patients may have larger T-cell clonal expansions and consequently fewer clones with low expansion rates compared to age-matched population controls. Several V and J genes were observed at a higher frequency compared to the background distribution of TCRs seen in controls. The most prominent association was present in over 44% of GCA-associated TCRs, while only being present in 1:7% of control TCRs. Further studies are ongoing to determine if these expanded TCRs are also present in GCA tissue and whether they are associated with specific clinical features, including relapse. This work lends additional support for the central role played by T-cells in GCA pathogenesis.

Exploration of Genome-wide Transcriptomic Signatures within GCA Tissue

An alternative agnostic approach was explored by ESR8 at the UNIVLEEDS, whereby the transcriptomic profile of GCA tissue was interrogated with correlations sought between the

transcriptome and different molecular subtypes of GCA. Multiple DAG were identified for 7 histological phenotypes of interest, based on their importance for GCA pathogenesis, after correction for multiple testing. The strongest associations were found for the degree of tissue damage in the media (5,159 genes; 3,801 upregulated), severity of adventitial and medial inflammation (3,503 and 3,333 genes, respectively), with approximately equal proportions of up- and down-regulated genes and degree of arterial occlusion (1,275 genes; 725 upregulated).

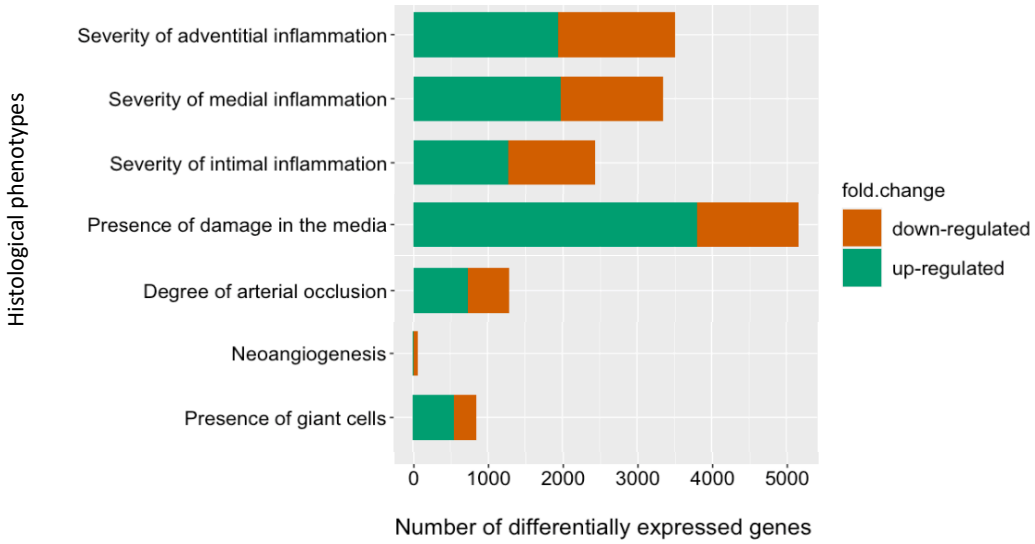

Figure 3. Number of statistically significant DEGs detected with FDR-corrected p-value < 0.01 in group comparisons for each histological phenotype. FDR – False Discovery Rate; DEGs – differentially expressed genes

Two histological phenotypes 'degree of arterial occlusion' and 'severity of inflammation in the media' were selected for further analysis with the former allowing exploration of vascular remodelling, whilst the latter focussed on transcripts associated with extension of inflammatory cells through the arterial wall.

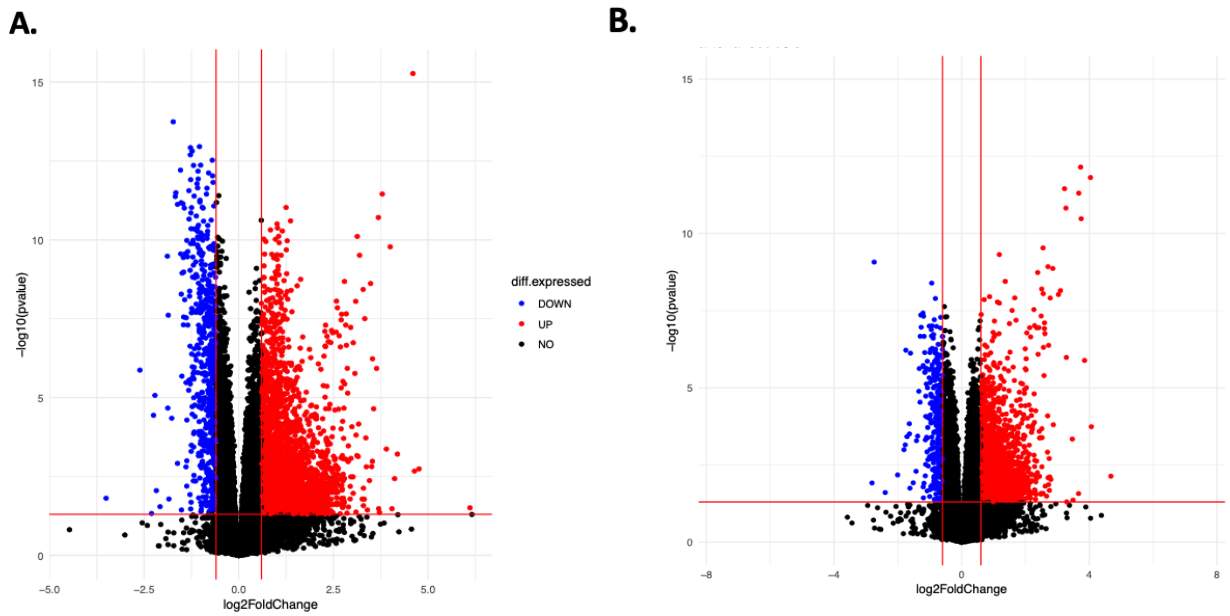

Figure 4. Volcano plots showing differential expression results for 'severity of inflammation in media' (A) and 'degree of arterial occlusion' (B). Statistically significant genes based on thresholds of p-value < $1x10^{-6}$ and log2FoldChange > 1 (up-regulated) / log2FoldChange < -1 (down-regulated) are coloured in red and blue, respectively. Top 10 genes with greatest/smallest log2FoldChange values are labelled with gene symbols.

The most significant and most differentially expressed gene for both features of interest was H3 clustered histone 14 (*H3C14*), with log2FoldChange=23.466; *P*-value=1.215e-23; FDR=2.686e-19 for the degree of occlusion grade and log2FoldChange=23.819; *P*-value=2.622e-26; FDR=7.023e-22 for the severity of medial inflammation. Histone isoforms are transcribed during the S phase of the cell cycle and are differentially expressed under different pathophysiological states. Upregulation of *H3C14* has previously been described in various malignancies, but not inflammatory diseases, but is consistent with the enriched histone marks observed in our latest GWAS data. *IL6* was notably down-regulated in both severe phenotypes, these data are confounded by glucocorticoids exposure prior to TAB, but collectively suggest the IL-6 pathway is rapidly suppressed by high-dose glucocorticoids at presentation. Analysis of other highly differentially expressed transcripts did not reveal obvious biological pathways for therapeutic targeting. Transcripts encoded proteins

associated with stromal proliferation, muscle differentiation and cell signalling pathways that are not typically associated with immunological disorders.

Functional enrichment analysis showed that the statistically significant DEGs detected for the degree of arterial occlusion were mostly involved in *vascular smooth muscle contraction* (hsa04270)*, hypertrophic cardiomyopathy* (hsa05410) and *dilated cardiomyopathy* (hsa05414), *arrhythmogenic right ventricular cardiomyopathy* (hsa05412) and *focal adhesion* (hsa04510) KEGG pathways with the enrichment fold > 3. These data highlight the importance of stromal pathways and the need to study these further to limit the vascular damage in GCA. The most functionally enriched pathway for the severity of medial inflammation was *primary immunodeficiency* (hsa05340), reaching an enrichment fold > 3.

A collaboration between UNIVLEEDS and Dr Gary Reynolds at the University of Newcastle-upon-Tyne enabled deconvolution of bulk RNA-seq dataset to be performed by ESR8. A scRNA-seq dataset generated from GCA tissue was shared by our collaborator and provided a disease-specific single cell reference panel to allow deconvolution of gene expression signatures. The cell type composition in temporal artery tissue from GCA patients was inferred from the bulk RNA-seq dataset using the MuSiC software package. The three most abundant cell types present in TAB were myofibroblasts, M1 macrophages and endothelial cells with myofibroblasts and M1 macrophages varying markedly across samples. A moderate number of transitional and vascular smooth muscle cells (VSMC) were observed. Further immune cell subsets were imputed at low levels e.g. regulatory T-cells (Treg), plasma cells, M2 macrophages, dendritic cells, CD4+ and CD8+ T-cell subsets. Higher proportions of VSMC, M1 macrophages, CD4+ and CD8+ T-cells were estimated in biopsies with more severe cases of inflammation in the media (multifocal or diffuse pattern) than in those with less inflammation (normal and focal patterns), while a reverse pattern was observed in the proportion of myofibroblasts. Similar patterns were observed for VSMC, macrophages, CD8+ T-cells and myofibroblasts with increasing arterial occlusion. The

reciprocal proportions of myofibroblasts to vascular smooth muscle cells in the GCA lesions most likely explains the high levels of DEGs from stromal pathways.

In collaboration with Dr Jason Tarkin at Imperial College London, ESR8 from UNIVLEEDS analysed available molecular data sets to provide supporting molecular evidence for the use of a novel imaging ligand directed at somatostatin receptor 2 (SST2) in GCA. Subsequent PET-CT studies confirmed the potential clinical utility of SST2 imaging and ability to distinguish active vascular inflammation from atherosclerosis.

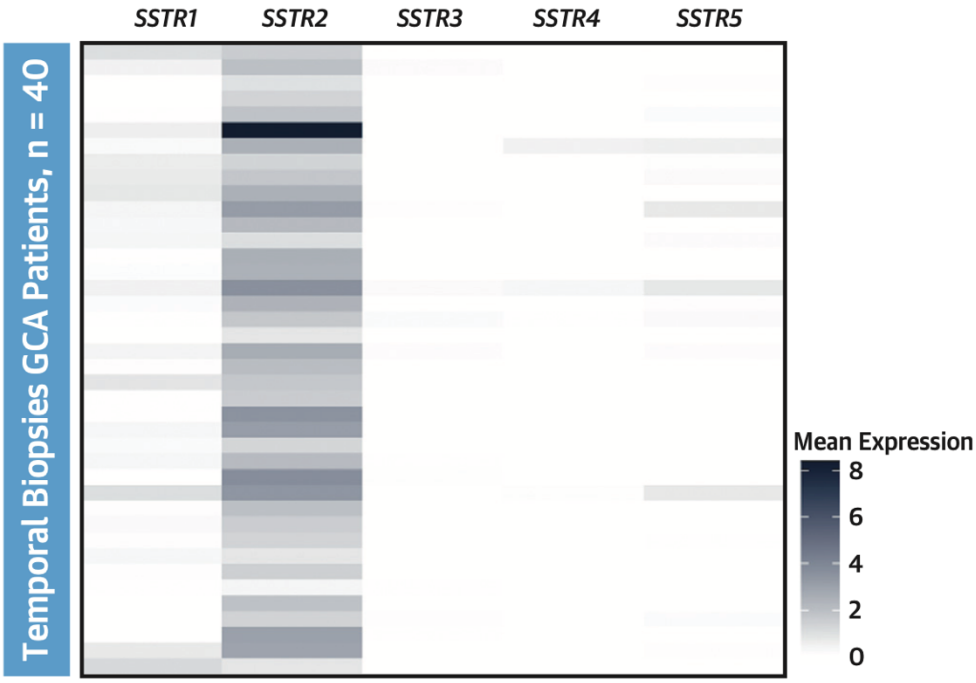

Figure 5. Heatmap of mean expression of somatostatin receptors across all patients. Figure adapted from (Ćorović et al., 2023)

Identification of Transcriptomic Signatures in Circulating CD14+ monocytes in GCA

Building on the observation that M1 macrophages were the most abundant immune cell in GCA tissue, CSIC in collaboration with IDIBAP extended these studies to explore changes in gene expression between CD14+ monocytes (macrophage precursor) from GCA patients compared with controls revealing 54 differentially expressed genes (DEGs), of which 41 were upregulated in GCA. *ADAMTS2*, *CD163*, *AMPH*, *FLT3* and *IL1R2* were observed to be

among the most significantly upregulated DEGs. The majority (72%) of DEGs presented higher levels of expression in the subgroup of patients with active disease. These results were consistent with the previous knowledge of GCA pathogenesis. For example, *IL6* and *MMP9*, as well as other members of the MMP family (MMP2, MMP24, MMP14, MMP19 and MMP25), were upregulated in active disease. We also detected overexpression of several genes of the integrin family, such as *ITGA2B*, *ITGA5*, *ITGA6*, *ITGA7*, *ITGAX*, *ITGAV*, *ITGB1*, *ITGB3*, *ITGB5*, *ITGB7* and *ITGB8*, as well as other remarkable genes that are important in monocyte cell biology like *CCR2*, *CCL2*, *CCL7*, *CXCL5*, *CXCL2* and *CXCL3*. Of note, important biological processes and pathways involved in GCA pathogenesis, such as angiogenesis, TNF signalling pathway, VEGF receptor pathway, chemokine signalling, MAPK cascade, Toll-like receptor signalling pathway and cellular response to IL-6, were enriched among the set of upregulated genes in GCA patients with active disease.

Identification of Epigenetic Signatures in Circulating CD14+ monocytes in GCA

In parallel with the above study, CSIC in collaboration with IDIBAP undertook the first methylome and transcriptome profiling of CD14+ monocytes in GCA. They demonstrated widespread alterations of the DNA methylation landscape compared with controls and revealed 1,371 differentially methylated positions (DMPs), annotated to 1,190 unique genes, across the whole genome. Of note, hypermethylated DMPs located within or close to genes previously associated with immune-mediated diseases, including *P4HA2* identified in our genetic studies. Through gene ontology analysis, we observed enrichment in functional pathways of the immune response, such as regulation of interferon-gamma (IFN-γ) production, leukocyte chemotaxis and integrin biosynthesis processes. In addition, we detected a significant enrichment in monocyte cell biology pathways, such as the Colony Stimulating Factor 1 (CSF1)-colony stimulating factor 1 receptor (CSF1R) complex, differentiation and proliferation of macrophages, and cytokine production like macrophage colony-stimulating factor (MCSF). As expected, CD14+ monocytes from patients with active disease showed large differences in the methylation landscape compared to those from patients in remission with treatment. These hypomethylated DMPs were enriched in pathways implicated in the immunopathogenic processes of GCA, including the cellular

response to IL-6 as well as response to other members of the IL-6 family, specifically IL-11, which warrants further investigation.

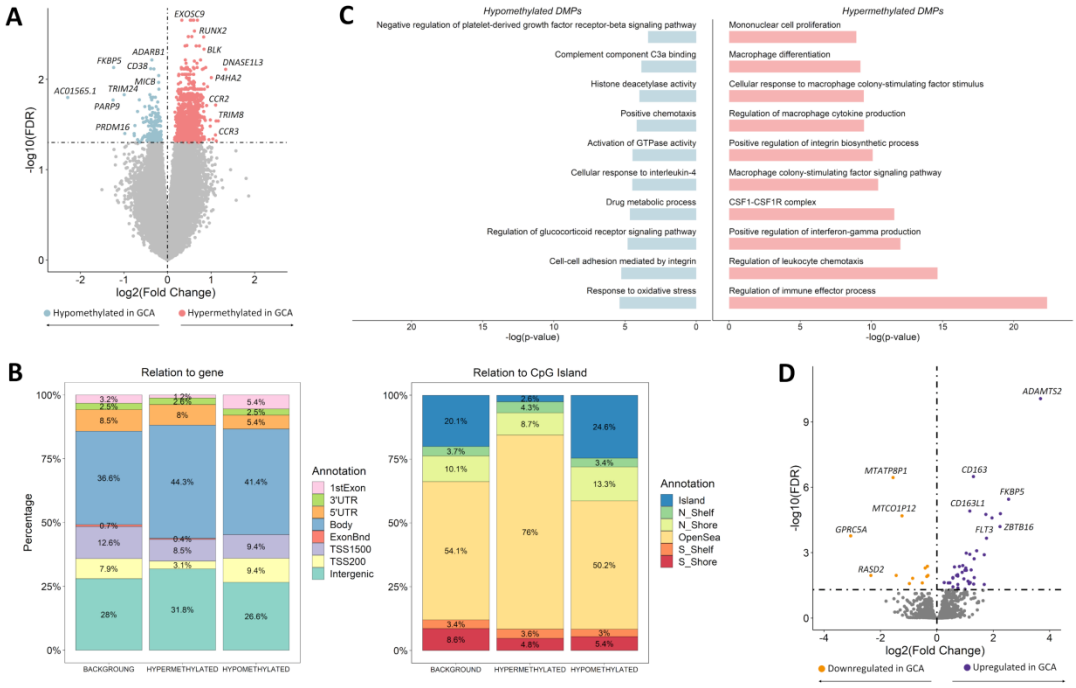

Figure 6. Results from the comparison of both DNA methylation and gene expression patterns of CD14+ monocytes between patients with giant cell arteritis and controls. A) Volcano plot of the epigenome-wide association study results. False discovery rate (FDR) values are represented on the –log10 scale in the y-axis. Significant threshold (FDR < 0.05) is marked by a dashed line. The effect size and direction obtained for each CpG site is depicted in the x-axis. Pink and blue dots represent hypermethylated and hypomethylated differentially methylated positions (DMPs), respectively. B) Bar plots representing the annotation of the significant hypermethylated and hypomethylated DMPs in relation to CpG island (left panel) and gene location (right panel). C) Representation of selected gene ontology categories obtained from the DMPs enrichment analysis using the GREAT online tool. D) Volcano plot of the transcriptome-wide association study results. FDR values are represented on the –log10 scale in the y-axis. Significant threshold (FDR < 0.05) is marked by a dashed line. The effect size and direction obtained for each gene is depicted in the x-axis. Red and green dots represent upregulated and downregulated differentially expressed genes, respectively.

It was noted that similar results were found when active patients were compared with both healthy controls and patients in remission, with and without treatment, which suggest that the pro-inflammatory methylation and expression profiles observed in the active disease are lost during glucocorticoid-induced remission. In fact, no differences were found when DNA

methylation and gene expression levels were compared between patients in remission without treatment and healthy controls.

Functional Characterisation of targeted therapies for GCA

Building on the aforementioned studies, IDIBAPS selected two targeted therapies for functional characterisation, namely tocilizumab (anti-IL-6 receptor) and mavrilimumab (anti-GM-CSFRα), which block different signalling pathways (STAT3 and STAT5A, respectively). Tocilizumab is the first biologic licensed for use in GCA, targeting the IL-6 pathway, whilst the mavrilimumab is an emerging therapeutic targeting granulocyte-monocyte colony stimulating factor (GM-CSF) that is currently being evaluated in phase II clinical trials. Both therapies target non-overlapping immunological pathways that were highlighted in our transcriptomic studies.

IDIBAPS initially validated mavrilimumab as a therapeutic target for GCA, confirming increased GM-CSF and GM-CSFRα transcription and translation in TAB from patients with GCA compared with control arteries. Immunostaining confirmed macrophages were the main immune cells expressing GM-CSF and GM-CSFRα and expression was also observed on luminal endothelial cells and, to a lesser extent, in T-cells, intimal myofibroblasts, and endothelial cells from vasa vasorum and neovessels in GCA arteries. Consistent with other inflammatory diseases, low levels of serum GM-CSF were detected in the serum of GCA patients at diagnosis, but this was not significantly elevated compared with controls, supporting a paracrine function in the tissue microenvironment. They also observed activation GM-CSFR-driven signalling pathways in the tissue with detection of JAK2 and STAT5A phosphorylation in GCA, along with increased expression of STAT5- regulated molecules, such as CD83 and transcription factor Spi1/ PU.1.

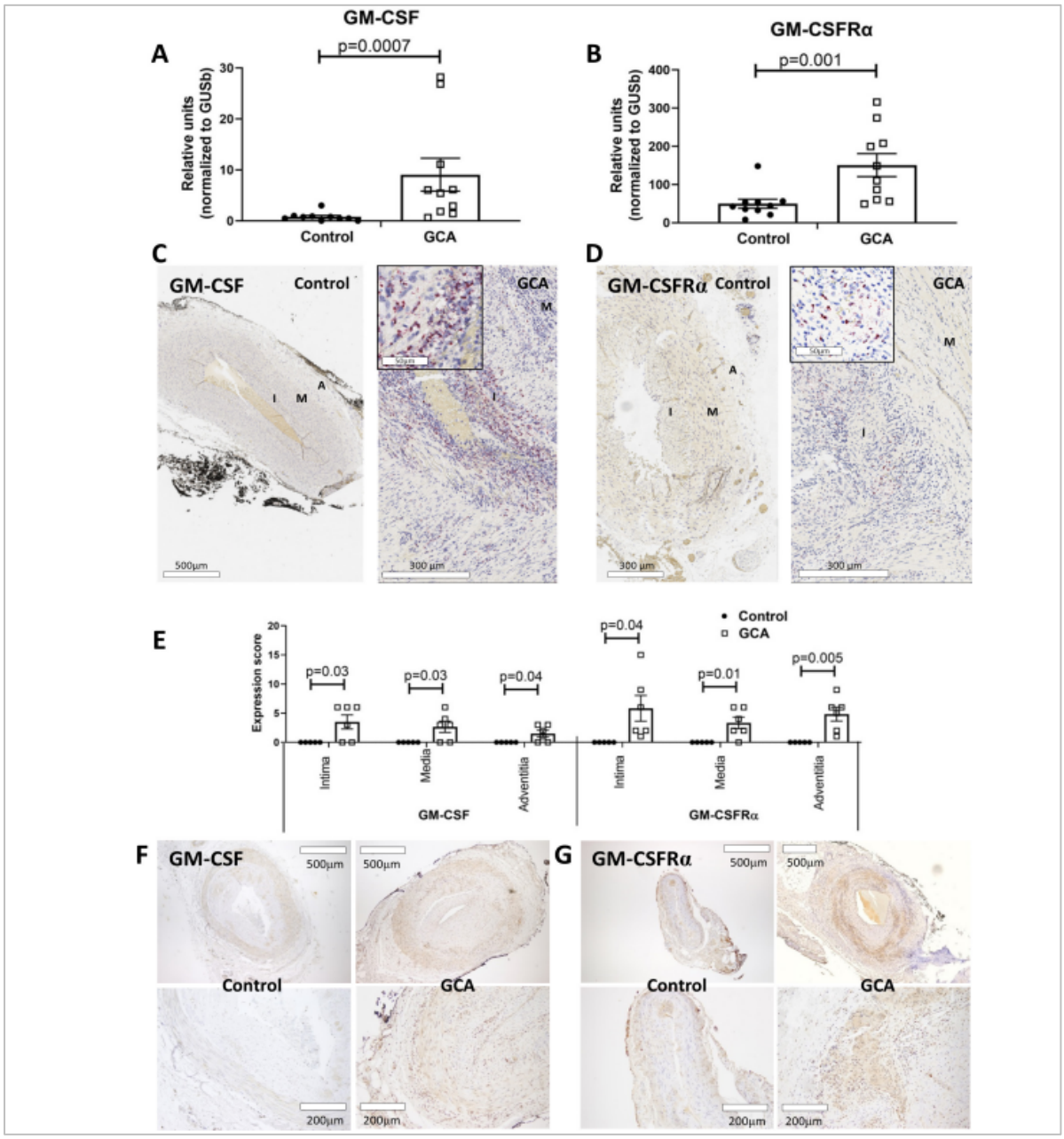

Figure 7. Granulocyte-macrophage colony stimulating factor (GM-CSF) and GM-CSFRα expression in GCA lesions. Concentrations of GM-CSF (A) and GM-CSFRα mRNA (B) measured by qRT-PCR in fresh-frozen histologically negative arteries (controls) (n=10) vs GCA-positive arteries (n=10). Results are expressed in relative units normalised to the housekeeping transcript GUSB. GM-CSF (C) and GM-CSFRα (D) RNA hybridisation signals (red dots) on control temporal arteries and GCA-involved arteries. (E) Quantitation of RS signal (expression score) in different arterial layers in 6 GCA-involved and 5 control arteries. Immunostaining with anti-GM-CSF (F) and anti-GM-CSFRα (G) antibodies (brown colour) of FFPE normal or GCA-involved arteries (representative of 5 controls and 12 GCA arteries). A, adventitia layer; FFPE, formalin-fixed paraffin-embedded; GCA, giant cell arteritis; GM-CSFRα, GM-CSF receptor alpha chain; I, intima layer; M, media layer; qRT-PCR, quantitative real-time PCR; RS, RNAScope.

To confirm GM-CSFR-mediated signalling contributed the increased expression of inflammatory cell markers observed in GCA, cultured temporal arteries from patients with histopathologically proven GCA were exposed to mavrilimumab for 5 days. Treatment with mavrilimumab resulted in significantly decreased transcripts of lymphoid, monocyte and myeloid cell markers and markers of antigen presenting function/T-cell activation/ differentiation, such as CD83, HLADR and interferon-γ. Mavrilimumab also reduced expression of pro-inflammatory cytokines: IL-6, TNF and IL-1β, as well as molecules related to vascular injury (MMP9, lipid peroxidation products and iNOS).

Aligning with our genetic data, our results suggested that GM-CSF regulated VEGFA production. Since neoangiogenesis is prominent in GCA lesions, and newly formed capillaries express adhesion molecules and recruit inflammatory leucocytes into arteries, mavrilimumab may indirectly reduce leucocyte recruitment into the vessel wall by decreasing neoangiogenesis in addition to its direct effects on myeloid and other cells. The results from a recent phase 2 trial in which mavrilimumab was superior to placebo (both with 26-week prednisone taper) in reducing the risk of GCA flare and maintaining sustained remission further validated the role of GM-CSF in GCA pathogenesis.

IDIBAPS then demonstrated mavrilimumab and tocilizumab showed different transcriptomic effects on cultured arteries from patients with GCA, with some overlapping effects, most notably CXCL-1, but no significant changes remained after correction for multiple comparisons, perhaps due to the small sample size. Differential effects may also have been attenuated by prior glucocorticoid use, which is required whenever GCA is strongly suspected clinically to reduce the risk of vision loss.

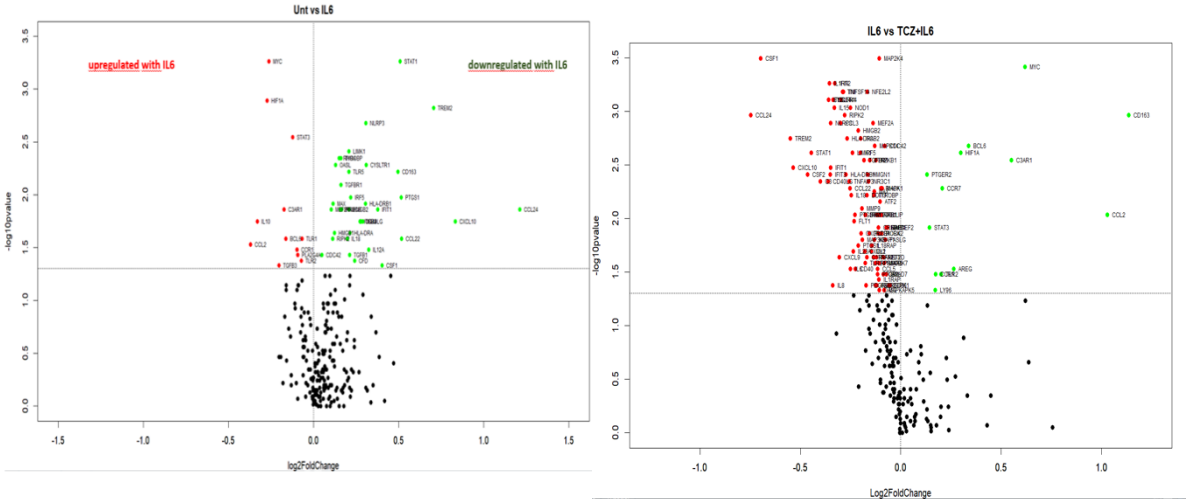

Figure 8. Differentially regulated transcripts after exposure to mavrilimumab (A) or tocilizumab (B) compared to placebo.

Systems biology approaches can further contribute to the holistic understanding of the biology of complex diseases. AX applied their Therapeutic Mapping System (TPMS) to the aforementioned datasets. This approach assesses potential functional relationships between proteins associated with medical conditions (e.g. GCA) to determine whether major non-canonical pathways (e.g. IL-6 pathway) can explain a drug's mechanism of action (e.g. tocilizumab) and evaluates if proteins of interest, like the ones identified experimentally, have any role in the proposed mechanism. TPMS relies on a human protein network (HPN) that incorporates available relationships (edges) between proteins (nodes) from individual literature-supported relationships and from public sources (e.g. KEGG, REACTOME). The TPMS algorithm takes as input the values of activation and inhibition of the proteins from certain stimulus and response sets, and it generates possible mechanisms of actions that connect the stimulus and the response through signalling pathways. These network analyses can unveil functional properties and pathogenic insights that are otherwise hidden, and that can be further validated *in vitro*, *in vivo* or by other *in silico* approaches.

This model was applied to differentially expressed proteins (transcripts) streaming from a high throughput analysis of the effects of tocilizumab in GCA. A HPN mathematical model

was initially used to assess the relationship of the differentially expressed proteins identified in the *in vitro* studies with a gene set representing GCA. The network analysis identified among the proteins repressed by tocilizumab: BCL6, STAT3, HIF1A, MYC and IL10 as the most relevant for GCA pathobiology. A second type of *in-silico* model simulated the action of tocilizumab on GCA and was shown to reproduce the effects observed in the aforementioned *in vitro* studies, and also with CCR1 and CCL2.

IDIBAPS then explored the effects of tocilizumab exposure on *ex-vivo* peripheral blood mononuclear cells (PBMCs) from patients with GCA in remission in comparison with control PBMCs. At a transcriptomic level, 64 transcripts were differentially expressed between GCA patients and controls at baseline, despite the GCA patients being in clinical remission. PBMCs were then exposed to IL-6, in the presence or absence of tocilizumab and a total of 73 transcripts were differentially expressed between these groups, after correction for multiple comparisons. Only 7 transcripts were increased in PBMCs following IL-6 stimulation, and subsequently decreased after exposure to tocilizumab: *STAT3*, *HIF1A*, *CCL2*, *CCR1*, *BCL6*, *MYC* and *C3AR1* (all of them target genes of STAT3). Working with AX, 2 clusters of patients could be classified that differed in their transcriptional response to tocilizumab. In group 1 ("responders") decreased expression of these 7 transcripts was observed, whereas in group 2 ("non-responders") no changes in transcripts were observed.

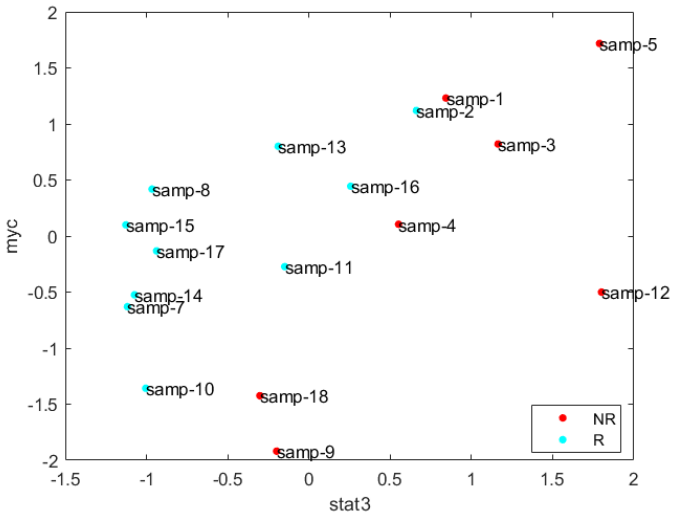

Figure 9. STAT3 vs MYC transcript levels at baseline.

A classifier analysis based on linear regression of transcriptomic baseline data identified the combined downregulation of *MYC* and *STAT3* transcripts as the key for good classification between the *in vitro* responder and non-responder groups. When considering clinical utility as predictors of response to tocilizumab, CCL2 is of particular interest since it can be measured in patients' serum or plasma. IDIBAPS have contacted the principal investigators of the GUSTO trial that used tocilizumab as monotherapy in active patients with GCA and will be able to validate using their samples, which will allow the effects of tocilizumab to be examined without being confounded by concomitant glucocorticoid use.

AX then focussed on glucocorticoid (prednisolone) use to evaluate whether systems biology approaches applied to real world data could identify previously unidentified molecular mechanisms and key molecules that may be used as biomarkers. AX initially screened the NHANES (National Health and Nutrition Examination Survey) database for correlations between prednisone use and pathological conditions. Some of those correlations could point to potential unaccounted adverse drug reactions to prednisone use. In a rheumatoid arthritis population, association of dyspnoea with prednisone use was significant, and in asthma and COPD populations association of hypernatremia with prednisone use was significant, with attention deficit disorder also being observed with COPD. At a molecular level, hypernatremia was associated with the reduced bone resorption pathological process and the protein MCF-1 was highly positively correlated in the COPD model. This protein is known to be involved in both inflammatory processes, including GCA, and in bone remodelling demonstrating the utility of this approach.

Summary

Collectively we have identified numerous new genetic loci in GCA and demonstrated transcriptional changes within the genome of circulating immune cells and the arterial wall. These changes arise from the influx of inflammatory cells and modulation of the stromal compartment within the tissue, alongside inflammation and disease-specific epigenetic changes that are likely to extend beyond the monocyte population, which were investigated in further detail. We demonstrated macrophages were the most abundant immune cell type

in GCA tissue and our transcriptional and epigenetic studies of circulating monocytes highlighted upregulation of CSFs, chemokines, MMPs and integrins amongst other factors. The suppression of IL-6 in GCA biopsies of patients treated with high-dose glucocorticoids highlighted that there may be opportunities for targeting other inflammatory pathways. We validated GM-CSF as a therapeutic target in GCA and compared the functional effects of mavrilimumab (anti-GMCSFα) to tocilizumab (anti-IL-6R) highlighting these therapies targeted different biological pathways.

We have many new therapeutic leads to follow up on such as the role of IL-11, stromal and matrix modelling pathways and our TCR data lend additional support for the ongoing evaluation of T-cell targeting therapeutics, such as abatacept, JAK inhibitors and emerging check point modulators.

Intriguingly our genetic data have highlighted vitronectin as a possible therapeutic target. This glycoprotein has a number of binding partners including plasmin activator inhibitor-1, integrin αV β3, and various protease inhibitors and plays a crucial role in limiting cell damage from the terminal complement pathway. Plasminogen, the protein encoded by PLG, which plays a role in tissue homeostasis may also feed into the same biological pathway. Abciximab which targets the integrin αV β3 (vitronectin receptor) is already approved for use in diabetes and acute coronary syndrome, due to its effect limiting platelet aggregation, making vitronectin an emerging therapeutic target worthy of further validation for GCA. In addition, we have shown the potential for our molecular datasets to be used for validating novel PET/CT ligands and have identified some novel molecular biomarkers requiring validation in external datasets.

KEY SCIENTIFIC OUTPUTS:

* denotes equal contribution between highlighted authors

Conference Presentations

- Gonzalo Borrego-Yaniz, Lourdes Ortiz-Fernández, Martin Kerick, Adela Madrid-Paredes, Augusto Vaglio, José Hernández-Rodríguez, Paul A. Monach, Santos Castañeda, Roser Solans, Bhaskar Dasgupta, Richard Watts, Nader Khalidi, Carol A. Langford, Steven Ytterberg, Lorenzo Beretta, Marcello Govoni, Giacomo Emmi, Marco A. Cimmino, Torsten Witte, Thomas Neumann, Julia Holle, Verena Schönau, Laurent Sailler, Thomas Papo, Julien Haroche, Alfred Mahr, Luc Mouthon, Øyvind Molberg, Andreas P. Diamantopoulos, Alexandre Voskuyl, Thomas Daikeler, Christoph T. Berger, Eamonn S. Molloy, Daniel Blockmans, Benedicte A. Lie, Paul McLaren, International GCA Consortium, Norberto Ortego-Centeno, Elisabeth Brouwer, Antje Mueller, Carlo Salvarani, Peter A. Merkel, **María C. Cid**, Miguel A. González-Gay, **Ann W. Morgan**, **Javier Martín** and **Ana Márquez**. Identification of new risk loci and pathways involved in GCA pathogenesis by a genome-wide study. EULAR Abstract 2023
- Marc Corbera Bellalta; **Farah Kamberovic**; Ferran Araujo; Roser Alba-Rovira; Georgina Espigol-Frigolé; Marco A Alba; Sergio Prieto-Gonzalez; José Hernández-Rodríguez; Patricia Pérez-Galán; Patricia Pérez-Galán; Kent Bondensgaard; John F Paolini; **Maria C Cid**. P175. International Vasculitis and ANCA Workshop, Dublin, April 2022. 10.5281/zenodo.6403187
- **Farah Kamberovic**, Marc Corbera-Bellalta, Roser Alba-Rovira, Marco A Alba, Georgina Espigol-Frigolé, Sergio Prieto-Gonzalez, Javi Marco Hernández, **MC Cid**. P69. 20th International Vasculitis and ANCA Workshop, Dublin, April 2022. 10.5281/zenodo.6401411
- **Farah Kamberovic**, Marco A Alba, Roser Alba, Georgina Espigol Frigolé, Javier Marco Hernández, Sergio Prieto, Marc Corbera-Bellalta and **Maria C Cid**. Transcriptomic Changes Induced by Tocilizumab in Ex-Vivo PBMCs from Patients with GCA in Remission. Predictors of Response. **Arthritis Rheumatol**. 2022; 74 (suppl 9).
- **Zulcinski M,** Reynolds G, Shafi L, Chakrabarty A, Westhead DR, **Iles MM** and **Morgan AW.** Deconvolution of bulk transcriptome from temporal artery biopsies reveals immune cell landscape of inflammatory infiltrates in giant cell arteritis. Poster presentation at: British Society of Immunology Dec 2022. *Awarded Best Poster Prize.*
- **Zulcinski M**, Reynolds G, Shafi L, Chakrabarty A, **Iles MM** and **Morgan AW**. Computational deconvolution of transcriptomics data reveals immune cell landscape of inflammatory infiltrates in giant cell arteritis. Oral presentation at: Proceedings of the 20th International Vasculitis and ANCA workshop 3-6th April 2022. 20th International Vasculitis and ANCA conference (Vasculitis2022), Dublin, Ireland. https://doi.org/10.5281/zenodo.6381941

- **Zulcinski M**, Reynolds G, Shafi L, Chakrabarty A, **Iles MM** and **Morgan AW**. Computational deconvolution of transcriptomics data reveals immune cell landscape of inflammatory infiltrates in giant cell arteritis. Oral presentation at: Proceedings of the Journées Ouvertes en Biologie, Informatique et Mathématiques (JOBIM), 5-8th July 2022, Rennes, France. https://jobim2022.sciencesconf.org/data/pages/JOBIM2022_proceedings_posters_demos.pdf

Preprint Medical Archives for submitted manuscripts

- Chatzigeorgiou C, Barrett JH, Martin J, UK GCA Consortium, Morgan AW*, Mackie SL*. Estimating overdiagnosis in giant cell arteritis diagnostic pathways using genetic data. MedRxiv: 2023.04.17.23288682.

Published Manuscripts

- **Anna Weber**, Jannis Born and **María Rodríguez Martínez.** TITAN: T-cell receptor specificity prediction with bimodal attention networks. **Bioinformatics** 2021; 37 (Suppl 1): i237-i244. The code as well as the dataset used in this study is publicly available at https://github.com/PaccMann/TITAN.
- Marc Corbera-Bellalta, Roser Alba-Rovira, Sujatha Muralidharan, Georgina Espígol-Frigolé,Roberto Ríos-Garcés ,Javier Marco-Hernández, Amanda Denuc, **Farah Kamberovic**, Patricia Pérez-Galán, Alexandra Joseph, Annalisa D'Andrea, Kent Bondensgaard, **Maria C Cid**, John F Paolini. **Ann Rheum Dis** 2022; 81:524–536.
- Iliana Papadopoulou, An-Phi Nguyen, **Anna Weber**, **María Rodríguez Martínez.** DECODE: a computational pipeline to discover T cell receptor binding rules. **Bioinformatics** 2022; 38 (Suppl 1): i246-i254. The code is available publicly at https://github.com/phineasng/DECODE.
- Ćorović A, Wall C, Nus M, Gopalan D, Huang Y, Imaz M, **Zulcinski M**, Peverelli M, Lambert J, Uryga A, Bressan D, Maughan RT, Pericleous C, Dubash S, Jordan N, Rusk RA, Jayne DR, Dean AF, Rassl D, Barwick T, Frontini M, Hannon G, Manavaki R, Fryer TD, Aloj L, Graves M, Dweck MR, Newby D, Fayad ZA, Reynolds G, **Morgan AW**, Aboagye EO, Davenport A, Jorgensen H, Mallat Z, Bennett MR, Peters JE, Rudd JHF, Mason JC, Tarkin JM. Somatostatin receptor 2 PET/MR imaging of inflammatory disease activity in large vessel vasculitis and atherosclerosis. **J Am Coll Cardiol**. 2023; 81: 336-356.
- Pieter Meysmana, Justin Bartonc, Barbara Bravid, Liel Cohen-Lavie, Vadim Karnaukhovh, Elias Lilleskovj, Alessandro Montemurrok, Morten Nielsenk, Thierry Morai, Paul Pereirai,

Anna Postovskayaa, **María Rodríguez Martínez**, Jorge Fernandez-de-Cossio-Diazi, Alexandra Vujkovica, Aleksandra M. Walczaki, **Anna Weber,** Rose Yino, Anne Eugsterg, Virag Sharmap. Benchmarking solutions to the T-cell receptor epitope prediction problem: IMMREP22 workshop report. **ImmunoInformatics** 2023; 9: 100024.

- Ortiz-Fernández L, Carmona EG, Kerick M, Lyons PA, Carmona FD, López-Mejías R, Khor CC, Grayson PC, Tombetti E, Jiang L, Mason JC, Direskeneli H, Saruhan-Direskeneli G, Callejas-Rubio JL, Vaglio A, Salvarani C, **Cid MC, Morgan AW**, Merkel PA, Burgner D, Smith KGC, González-Gay MA, Sawalha AH, **Martín J**, **Márquez A**. Identification of new risk loci shared across systemic vasculitides points towards potential target genes for drug repurposing. **Annals Rheum Dis** 2023 Online first doi: 10.1136/ard-2022-223697.