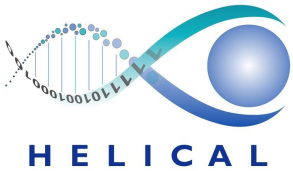


H2020-MSCA-ITN-2018-813545

HELICAL

Health Data Linkage for Clinical Benefit
Deliverable D6.8 Final Report

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

Explanation of the work carried out by the beneficiaries and overview of the progress

Objectives

Project objectives. Work Package 1 – 8

WP 1 Environmental Impacts on Autoimmunity

To investigate the impact of environmental factors on vasculitis onset and relapse.

WP 2 Identification of key pathogenic pathways suitable for therapeutic targeting in GCA

To study the molecular contributions to GCA pathogenesis.

WP 3 Linkage of clinical data to plasma and tissue analyses

To link clinical and experimental readouts to develop novel biomarker and risk stratification techniques.

WP 4 Governance of electronic health record sharing and linkage

To investigate the impact of recent data protection legislation on health research and to oversee data governance across the consortium.

WP 5 Structured training

To train ESRs in a broad range of techniques relevant to vasculitis research, to provide ESRs with a foundation in commercialisation and to equip ESRs with transferable skills relevant to their future careers.

1. To train ESRs in advanced data science, machine learning, systems biology and clinical applications allowing them to become pioneering experts in the fields of health data linkage and autoimmune disease,
2. To ensure that all ESRs exit the PhD programme highly competitive for careers across academic, industry and NGO sectors,
3. To enhance awareness by ESRs of FAIR and GDPR data principles, as well as developing their skills in dissemination and exploitation of research results.

WP 6 Management and Recruitment

To put in place and implement a strong management structure which will monitor progress of project, development of ESRs and will ensure achievement of deliverables and milestones.

WP 7 Communication, dissemination, exploitation & outreach

To effectively communicate and disseminate the project findings to key stakeholders, including the public, patient groups, cancer researchers, pharmaceutical/diagnostics companies and to involve ESRs in dissemination and exploitation activities and to effectively manage intellectual property generated during the project.

WP 8 Ethics

To manage ethical considerations identified by the commission at project inception.

Scientific objectives:

1. Leverage the power of cutting-edge **data science technologies** by ethically linking health records to environmental information to learn about factors leading to onset and relapse of autoimmune disease (*Scientific WP1*).
2. Investigate **genetic**, **genomic** and **tissue** signatures associated with GCA to identify key pathogenic pathways amenable to therapeutic targeting (*Scientific* WP2).
3. Combine linkage of **proteomic** and **immunohistological data** to clinical outcomes, using machine learning to uncover novel mechanisms and develop innovative ***biomarkers*** for autoimmune vasculitis (*Scientific WP3*)
4. Define the most effective and **ethically appropriate model** for linkage of these health and research data (*Scientific WP4*)

European objectives:

1. Establish durable interdisciplinary and intersectoral cooperation in the field of autoimmunity and health data linkage among researchers, and between academic and non-academic partners
2. Enhance attractiveness of research-linked careers for the next generation to address the European data science skill gap
3. Build a strong network and establish a forum open to all those operating in the arena of data linkage for clinical benefit.

Summary of work carried out to date. Work Packages 1- 8

Work Package 1 Environmental impacts on autoimmunity

Develop methodologies to investigate the interaction between vasculitis onset and relapse

Lead: Mark Little (TCD)
Participants: P7D, UG, ISG, UNIABDN, KSG, Univ Padova, DFKI, IHD, RPTU, TCD

Aim
This WP is adopting a novel approach to the question of environmental impact on autoimmunity by studying the whole system in which the patient moves through time, comparing this to formal reductionist case-control study design. To achieve this, it is necessary to develop IT systems that facilitate fusing large quantities of data deriving from multiple sources, some of it sensitive; the first aim is to build on existing systems to create frameworks that support diverse dataset fusion and advanced analytics, linking environmental data streams with registry and app data. Focusing on the pre- and post-diagnosis periods in parallel, we are addressing the following questions: (a) Are there clusters within the population under study and can we identify specific factors associated with vasculitis flare and/or onset? (b) Can we predict for a given patient the probability of suffering a flare within a specified time? (c) Can an approach that combines agnostic machine learning and traditional focused case-control epidemiological study improve predictive fidelity? (d) Is it thus possible to develop a prototype physician/patient interface integrating environmental and patient level data to provide individual flare risk estimates, forming the basis for future clinical tool development?

EXPERIMENTS PERFORMED:

Enabling infrastructure

The twin challenges of studying environmental triggers of autoimmune disease are dynamic spatio-temporal linkage and irregular time series modelling using sparse datasets. These were addressed by ESRs 1 and 3.

ESR1 adopted a 3-phase approach:

1. Identify researcher requirements in linking data for environmental health research.
2. Develop a framework that enables a researcher to link environmental data with particular health events based on user data inputs.
3. Evaluate and refine the developed framework through rare disease case studies. This was undertaken iteratively through a series of systematic user interaction sessions involving 50 researchers.

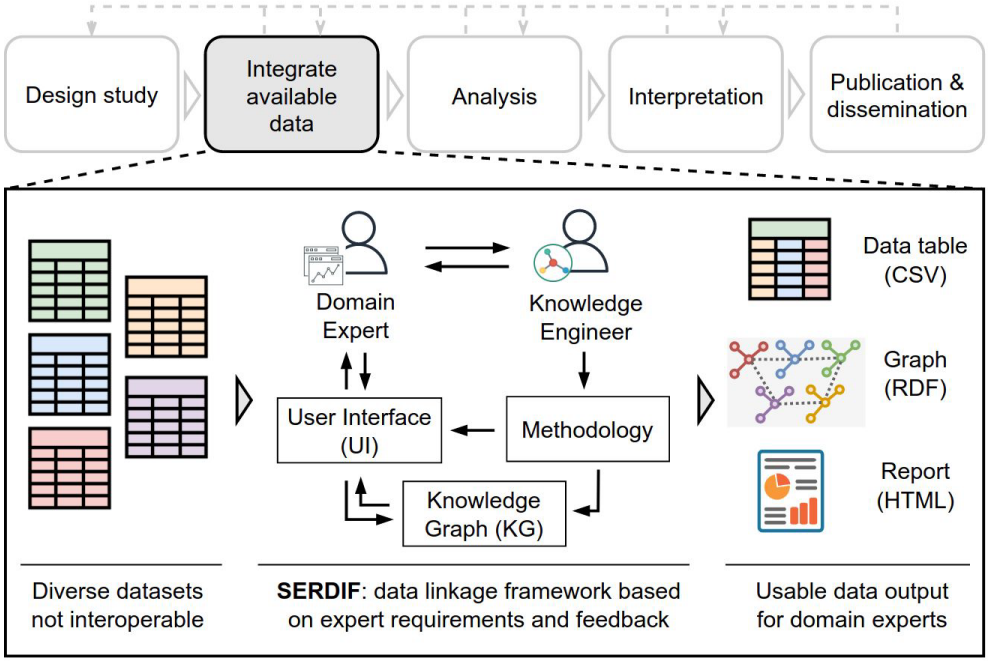


***Figure 1**. Framework for development and evaluation of the SERDIF data integration tool.*

ESR3 addressed the challenge of developing a model that can detect a correlation between ANCA vasculitis flare propensity and environmental exposure. In brief, the approach comprised a specialised regression architecture, which accounts for the presence of a distributed lag period before relapse that can be inferred from the data in the context of an irregular time series. This integrates the Mixed-data sampling (MIDAS) model concept, generalising them for binary response data and progressing to a Bayesian implementation. This model performs variable selection and accounts for binary response imbalance. It was applied to linked environmental data derived from the Irish Rare Kidney Disease registry (using the software developed by ESR1), implemented in R and developed as an R package.

Linking registry data to environmental information

The two objectives of HELICAL WP1 are identification of triggers of vasculitis ***relapse*** and vasculitis ***onset***. By using the Irish RKD and UKIVAS registries, we studied **longitudinal** environmental triggers of relapse in patients *known to have AAV*. Thus, by considering the time series nature of each participant,

it was possible to study spatio-temporal changes in environmental conditions and to link these to relapse at the subject level.

As patients who develop de novo AAV are generally not known before diagnosis, it was necessary for ESR4 to use large publicly available datasets to allow comparison to appropriate control groups: re-use of primary healthcare data from **NHS Scotland** and leveraging the **UK Biobank** to link new onset AAV with preceding environmental conditions.

The approach of ESR6 was focused on **time series analyses** to discover patterns within the temporal dynamics of disease incidences and establishing connections to environmental factors. Deep analysis of the spatiotemporal dynamics of Kawasaki Disease in Japan allowed ESR6 to test the hypothesis that tropospheric winds play a role in the development of the disease. He conducted a comprehensive computational analysis of wind transport phenomena, as well as air sampling to assess the chemical and biological properties of the air at the destination. This has involved both the on-site air-sampling in Japan and the computational analysis of previous sampling campaigns.

Modelling approaches to exploring environmental triggers of autoimmune vasculitis

The relevant methodology is described in Deliverable 1.3. In addition to the MIDAS modelling approach described above, we used the following techniques:

1. The median prodrome period (that interval between onset of symptoms and recording of the clinically evident initial AAV diagnosis, or diagnosis of relapse) was estimated from existing registry data as being approximately 70 days. This derived value was applied in time series and data linkage analyses.
2. Relapse was studied using an **n-of-1 design** where each participant acts as their own control, eliminating confounding by time-independent factors such as gender, occupation and genotype. Case (relapse) and control (remission) windows were defined using an algorithm that considered the effect of residual disease activity (Figure 2). A multi-level model was applied to investigate the association between environmental factors (such as ambient vitD-UVB/CW-D-UVB) and AAV relapse. A random effect was included to account for repeated measures and the varying relapse risk between individuals.

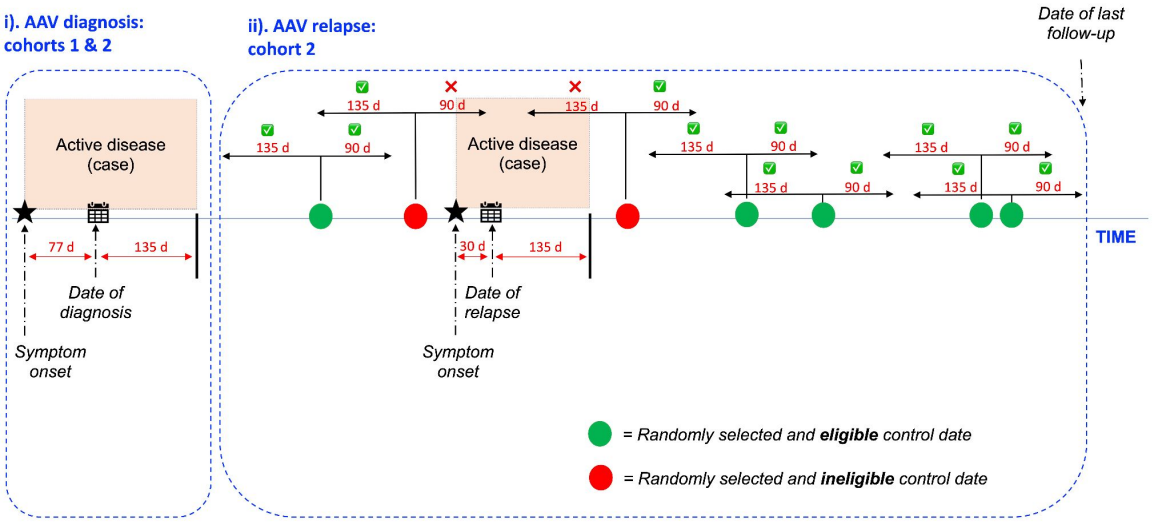


Figure 2. *Strategy for an n-of-1 study design to explore environmental triggers of vasculitis.*

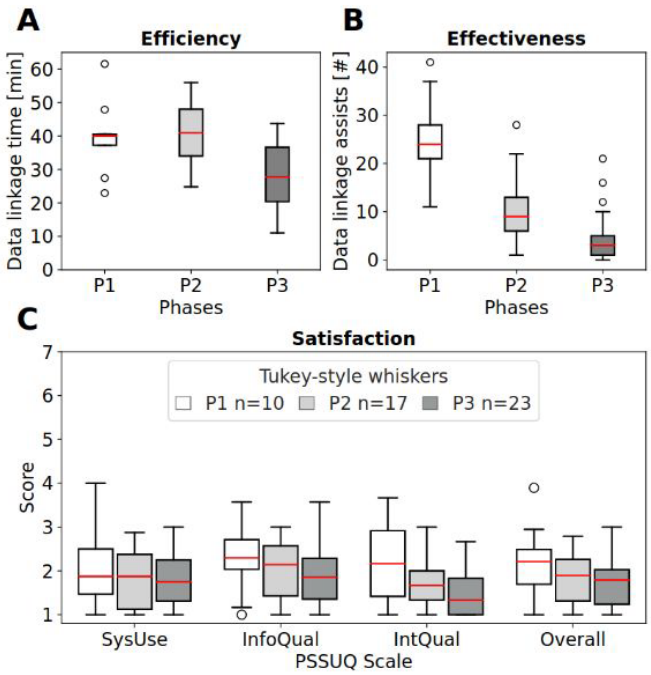
3. To study AAV onset (diagnosis), logistic regression was used to examine the relationships between cumulative dose of pollution exposure and vasculitis risk, and between AAV phenotype and serotype (outcome variables) and measures of ambient vitD-UVB. Each of the 18-25 model's two-tail p-values were adjusted for multiple comparison using Simes-Benjamini-Hochberg false discovery rate. A stratified analysis was undertaken to obtain insight into geographic variation of air pollution by population density.
4. To evaluate the spatiotemporal dynamics in Kawasaki disease, we used a variety of methods.
    a. The implementation of Multiple Seasonal-trend Decomposition using LOESS (MSTL) served to break down the time series into trend, seasonal cycles, and residual components, offering an effective way to understand the underlying patterns in the data. In addition, to detect cycles of an uncertain periodicity that may be present in the series, Singular Spectrum Analysis (SSA) was employed.
    b. Hierarchical clustering, using the Ward method and Spearman correlation as a distance metric, was chosen for grouping regions based on the similarity of their temporal dynamics. For assessing similarities where the correlations between the time series might be transient and non-linear, Scale Dependent Correlation (SDC) analysis was applied.
    c. The task of evaluating the spatial autocorrelation of disease incidence was addressed with the use of Global Moran's I. To further supplement this analysis and determine the presence of spatial clusters, Local Indicators of Spatial Association (LISA) were used.

Case-control study

ESR5 designed a questionnaire-based study to compare lifetime occupational history between patients with AAV and community dwelling controls. The “Canjem” job exposure matrix was used to convert these occupations into a hierarchical list of specific pollutants. Latent class mixed models were used to study nonlinear associations between protracted exposures (e.g. asbestos and tobacco smoke) and vasculitis onset. Given extreme delay to commencement of this study, the same approach was applied to a publicly available Swedish dataset.

RESULTS:

A software tool for spatio-temporal environmental data integration



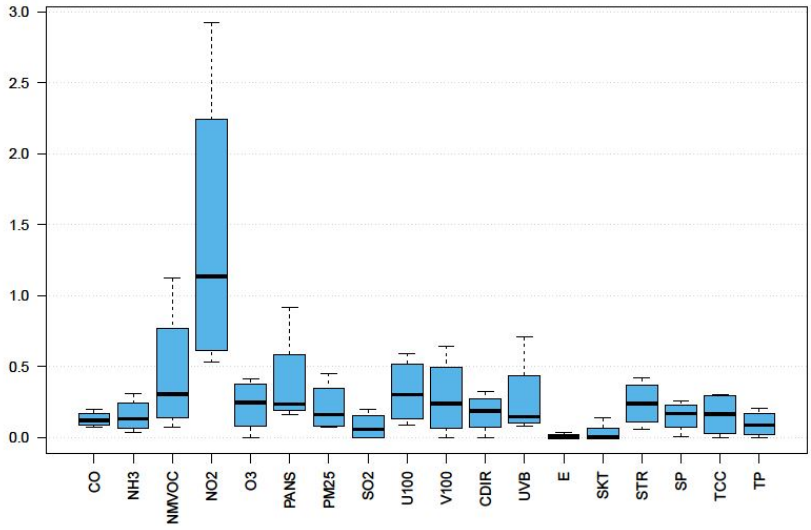
Extensive incremental evaluation of the SERDIF tool demonstrated progressive improvement in efficiency, effectiveness and satisfaction with each iteration (Figure 3). The tool is now applicable not only to the HELICAL use case, but more generically to any biomedical question that requires spatio-temporal linkage of patient location to environmental values.

Figure 3. *Summary of evaluation experiments illustrating progressive improvement in usability and satisfaction of the SERDIF tool.*

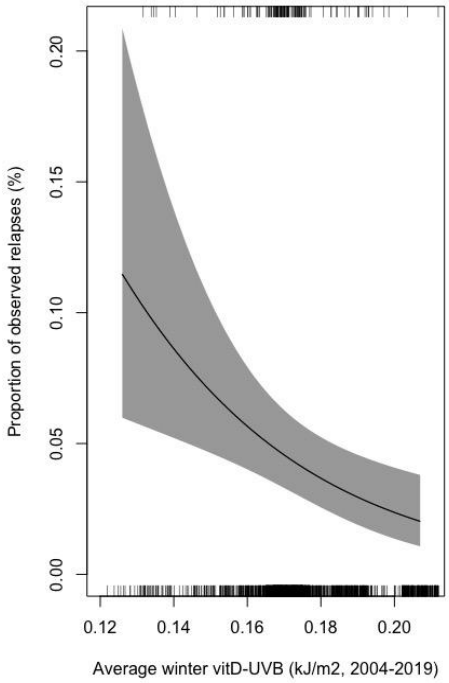
The association between environmental exposures and AAV relapse

The final dataset analysed using the MIDAS technique contained 87 patients from the RKD registry who had suffered a definite relapse, each with 4 years' worth of daily data and 18 location specific environmental explanatory variables. The Bayesian posterior prior values across all runs are given in Figure 4, which shows a low rate of variable selection. The most often selected value across all runs, at just under 3%, was Nitrogen Dioxide ($N0_2$). We see from the boxplot that $N0_2$ was consistently one of the most frequently selected variables, with the next most selected (on average) being Non-Methane Volatile Organic Compounds (NMVOC). No other variable achieved more than 1% inclusion across all 50,000 runs, and $N0_2$ was the only variable that never fell below 0.5%. These results suggest that, using this methodology and dataset, there is no clear association between the tested individual environmental triggers and vasculitis relapse, but $N0_2$ merits further investigation.

Figure 4. *Testing of association between environmental pollutants and relapse in occurrence in the Irish vasculitis population. Variable selection results after 4 runs with different starting points. Note the scale of the y-axis; across all runs, no variable is selected more than 3% of the time*.



The association between UVB exposure and AAV relapse and onset



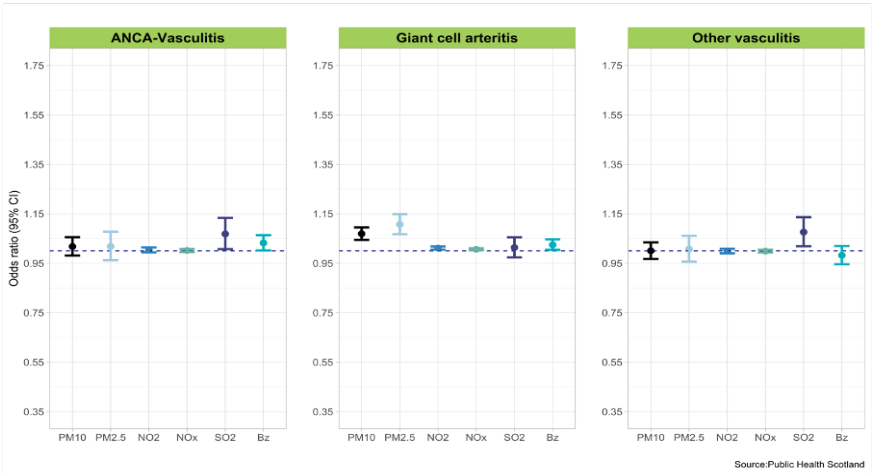
Residential latitude was positively correlated (OR:1.41, 95% CI 1.14-1.74, p=0.002) and average vitD-UVB negatively correlated (0.82, 0.70-0.99, p=0.04) with relapse risk, with a stronger effect when restricting to winter measurements (0.71, 0.57-0.89, p=0.002). However, contrary to our hypothesis, these associations were not restricted to granulomatous phenotypes. We observed no clear relationship between latitude, vitD-UVB or CW-D-UVB and AAV phenotype or serotype.

Figure 5. *Correlation between winter Vitamin D exposure and relapse.*

The association between AAV onset and air-borne pollutants

Analysis of the Scottish NHS dataset, with validation using UKIVAS and UK Biobank dataset illustrated a consistent association between sulphur dioxide, which was primarily observed in rural dwellers (Figure 6).

A



B

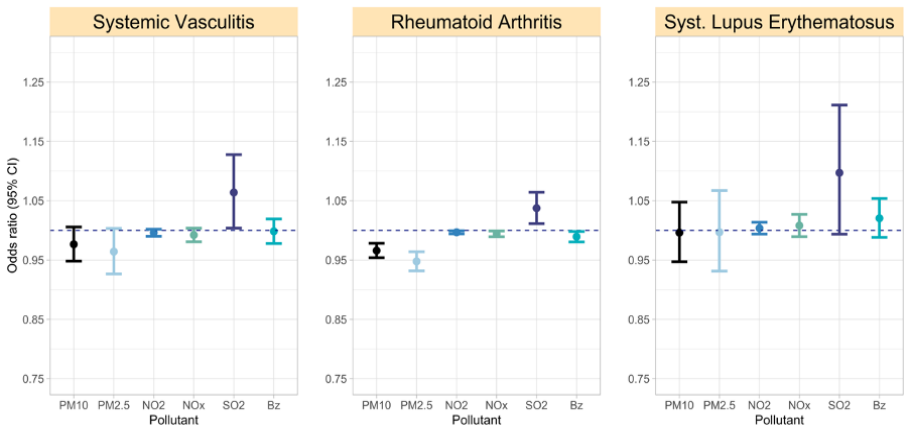


Figure 6. *Association between SO2 and vasculitis onset in Scotland (A), validated using UK Biobank data (B).*

Examination of prolonged occupational exposures on AAV occurrence

By applying a novel job exposure matrix to a publicly available AAV dataset, we observed associations with occupational exposure to 13 or 188 potential agents, including hydrogen sulphide, organic alkanes and aldehydes, and aromatic hydrocarbons, but not silica. These exposures tend to be enriched in agricultural and food processor workers (Figure 7). For example, 92% of substances associated with GPA are present in Agricultural and Animal Husbandry Workers, while only 17% are present in Cabinet Makers and related Woodworkers. These are often the product of acid catalysis of SO_2, providing a potential link between the data linkage and case control results. These findings are consistent with the concept of AAV as an oxidative disease.

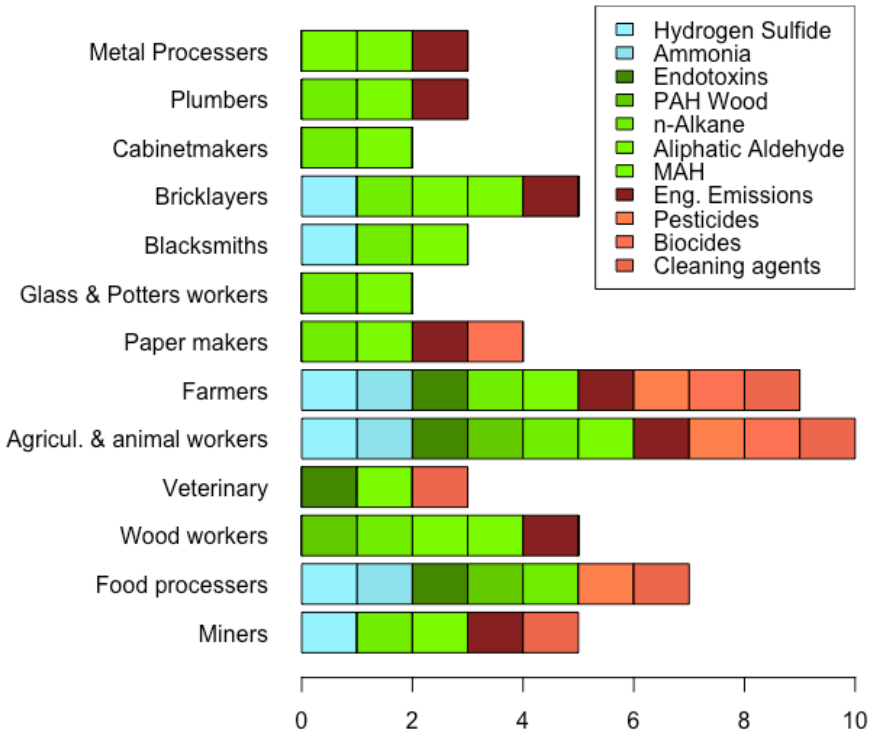


***Figure 7**. Distribution of the 11 substances found to be associated with GPA among the major 13 ISCO-68 groups. MAH = Monocyclic Aromatic Hydrocarbons*

Time series analysis of Kawasaki disease incidence

Daily variability of fine aerosols in a surveillance campaign in south Japan shows a striking co-evolution between their trace elements (metal and metalloid, MM) content and Kawasaki Disease (KD) admissions, suggesting a strong dynamical link. This association may account for >40% of total variability in the disease, being dominated by a clear sub-weekly cycle (SWC1). This SWC1 appears to connect or disconnect Japan to air intrusions from above the planetary boundary layer (PBL), having their source in industrial and agricultural areas in NE Asia, and points to a stronger case for an agricultural source for the exposure as opposed to urban pollution. KD maxima always occur in full synchrony with the arrival of very small (<1 µm; PM1) particles showing that ultrafine aerosols appear to be a necessary cofactor in the occurrence of KD (Figure 8).

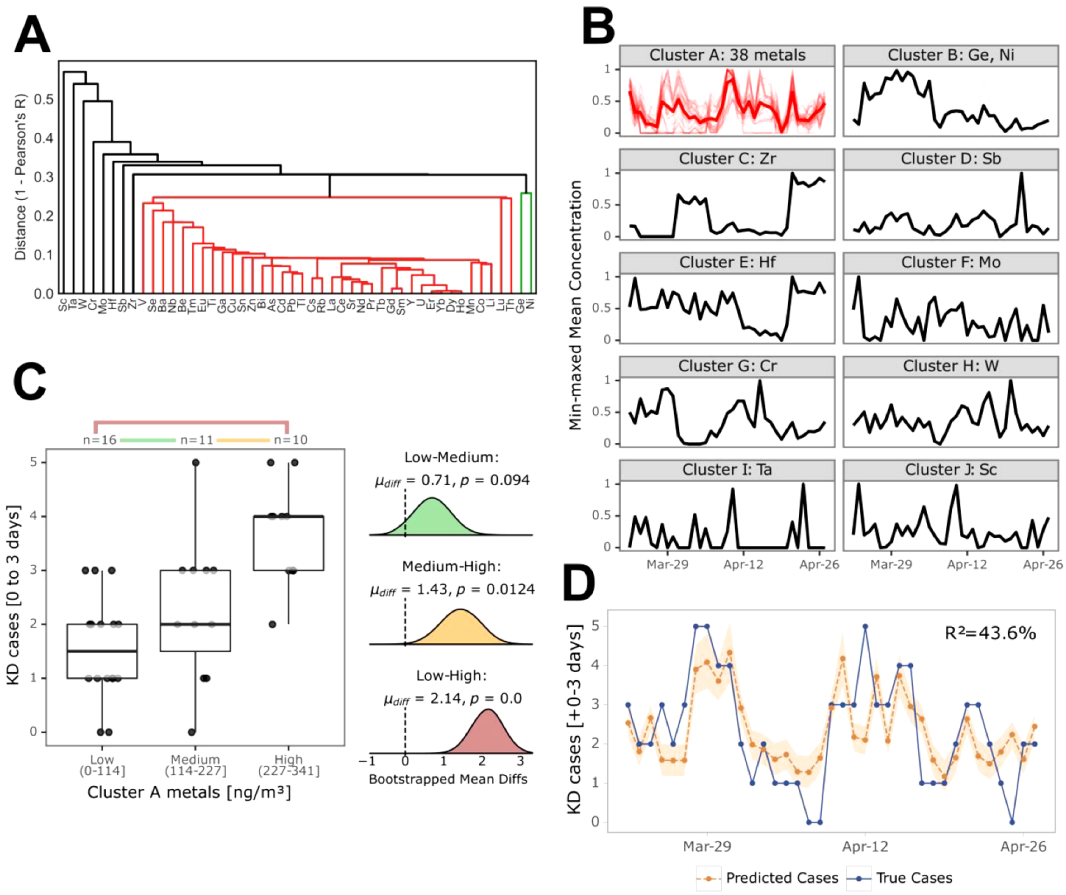


Figure 8. *Cluster analysis of the metal content in aerosols and its synchrony with KD incidence. (A) Dendrogram generated when using Pearson's correlation as distance metric and the Nearest Point Algorithm to compute distances across the newly formed clusters. Colours denote the different clusters assigned when setting a distance threshold of 0.3. (B) Variation during 37 sampled dates of the (min-max normalized) concentration of MM of each cluster. The thick line represents the median value of all cluster members at each time-point, with the individual metal contributions shown in the shaded thin lines. (C) Boxplots for the distribution of daily KD cases as a function of the daily concentration of Cluster A categorized into three groups: Low (0 to 114 ng/m³, n=16), Medium (114 to 227 ng/m³, n=11) and High (227 to 341 ng/m³, n=10). On the side, distribution of the difference of means for the pairwise comparison of groups performing a bootstrap test, showing significant differences in KD cases for days with high concentrations compared to those both in medium and low concentrations. (D) Reported (true) KD cases in the Kumamoto prefecture (blue) and predicted cases with a Poisson model using the concentration of Cluster A MM as predictor. The shaded area represents a bootstrapped 95% CI.*

KEY SCIENTIFIC OUTPUTS:

- SERDIF software tool (https://serdif-ui.adaptcentre.ie/) that uses knowledge graph technology to integrate patient location with environmental factors to allow study of the impact of these factors on disease progression. Critically, this software is open source and FAIR; its use of a flexible interoperable semantic web approach led to an unexpected scalability to other disorders. Indeed, the software is not limited to use in vasculitis. It can be applied in any

condition where there is a need to link patient location over time with a wide range of environmental data streams, suggesting that it will have wider application in exposome research (Milestone 2: https://github.com/navarral/serdif/; described in publications 2, 12 and 14).

- This tool was used to study the association of Ultraviolet-B radiation and the occurrence and relapse of ANCA vasculitis, with the finding that relapse risk was associated with low cumulative winter UVB exposure. This work also demonstrated, for the first time, the use of an n-of-1 study design for investigation of longitudinal outcomes in patients (Publication 10).
- Development of an R-package for distributed lag modelling of unbalanced binary data through use of Bayesian quantile regression. Variable selection is possible and useful for quantifying potentialities of impacts of environmental exposures on binary responses (Deliverable 1.2; https://github.com/jsnwyse/dlvarsel).
- Identification of SO2 as a key environmental potentiator of ANCA vasculitis (Deliverable 1.3, manuscript in preparation).
- Description of strong consistent negative effects of both temperature and absolute humidity at large spatial scales associated with the spread of SARS-CoV-2 around the world. ESR6 classified, for the first time, COVID-19 as a seasonal low-temperature infectious disease (Publication 13).
- Development of a Python package, designed specifically for the execution of Scale Dependent Correlation (SDC) analysis. This tool facilitates the exploration of transient synchronicities within time-series data sets. The source code for this package has been uploaded onto GitHub (https://github.com/AlFontal/sdcpy), and a copy has been archived on the HELICAL community in Zenodo for long-term preservation (https://doi.org/10.5281/zenodo.4949813). Further expanding on accessibility, a web application has been developed to provide an interactive GUI for the package, thereby ensuring ease of use. The application can be found on GitHub (https://github.com/AlFontal/sdcpy-app), and is also readily available for use on the Heroku platform (https://sdcpy-app.herokuapp.com/).
- Association of ultrafine aerosols enriched in metals sourced from areas of intensive farming and urban pollution to Kawasaki Disease in Japan. Accepted Manuscript available at https://doi.org/10.1088/1748-9326/acd798. Accompanying code open sourced at Github (https://github.com/AlFontal/kd-metals-swc) and deposited in Zenodo (https://doi.org/10.5281/zenodo.7948389)
- Assessment of the spatial coherence in the temporal dynamics of Kawasaki Disease at the prefectural level. Spatial clusters were observed not only during the early epidemic events (1979-1986), but also in current seasonal patterns. This provides a critical insight into the geographical and temporal factors influencing the disease's spread, furthering the evidence towards an environmental driver (manuscript in preparation).

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

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) implement public (in particular the NHANES dataset) and new experimental knowledge on regulation of gene expression by transcription factors, and incorporated epigenetic elements into protein functional network analysis, (2) applied high-throughput structure-function analysis for functional understanding of genetic results from ESR7/8 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 drove 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.

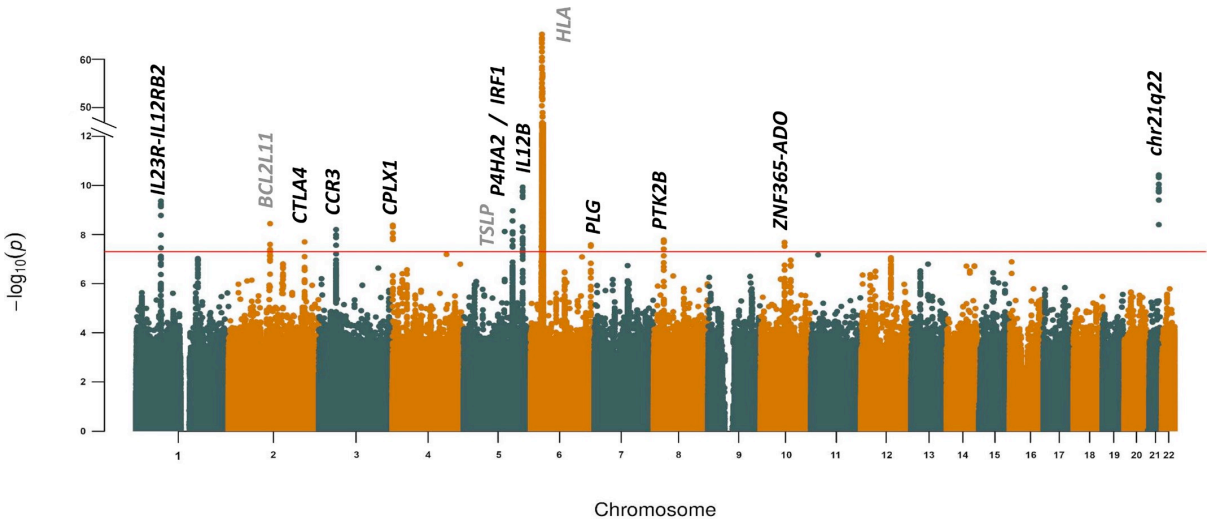


Figure 9. Manhattan plot showing the results of the cross-phenotype meta-analysis. The −log10 of the p values are plotted against their physical chromosomal position. The red line represents the genome-wide level of significance ($p < 5\times10^{-8}$). Loci reaching the significant threshold are annotated in the plot. Loci representing new shared risk loci in vasculitis are highlighted in bold.

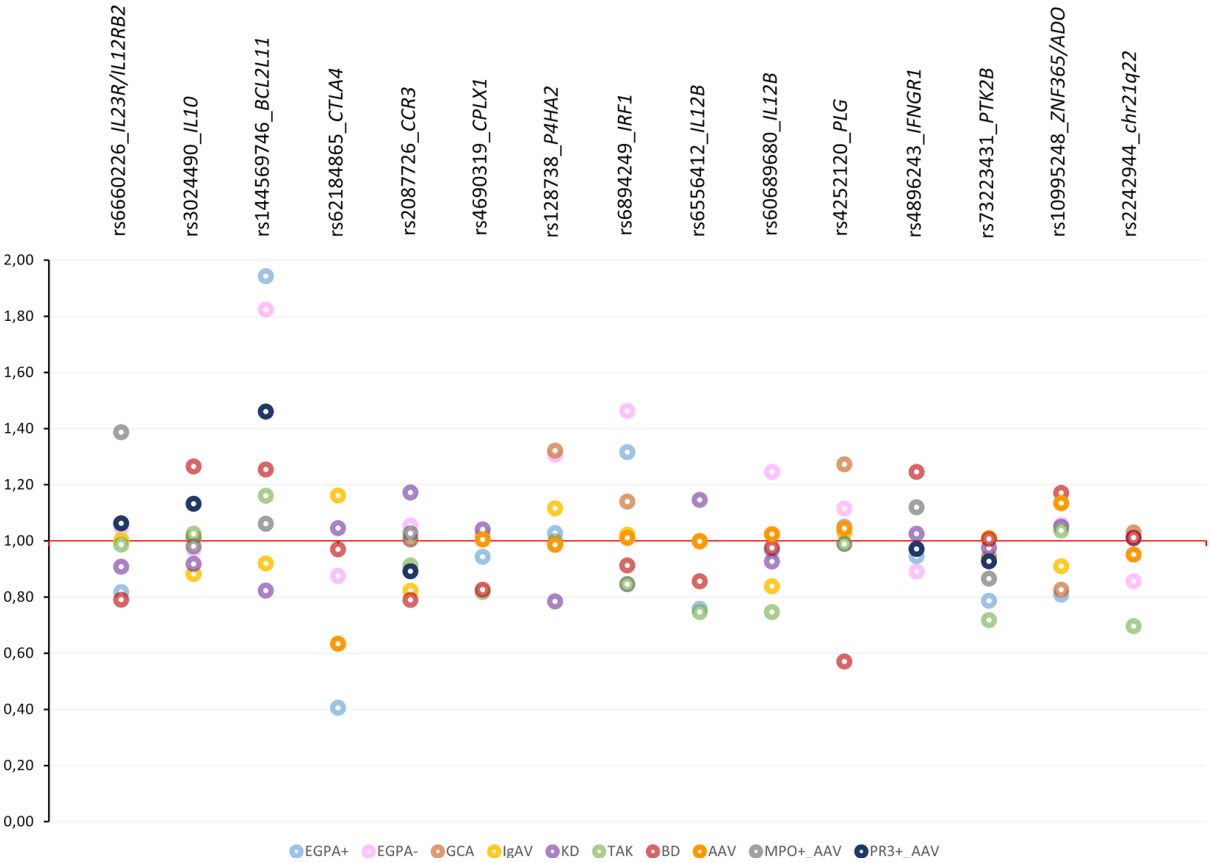


Figure 10. Novel risk loci shared across vasculitides. Effect of the independent genetic variants reaching genome-wide significant level in the subset-based meta-analysis is shown. Circles represent the analyzed phenotypes of vasculitides.

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 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).

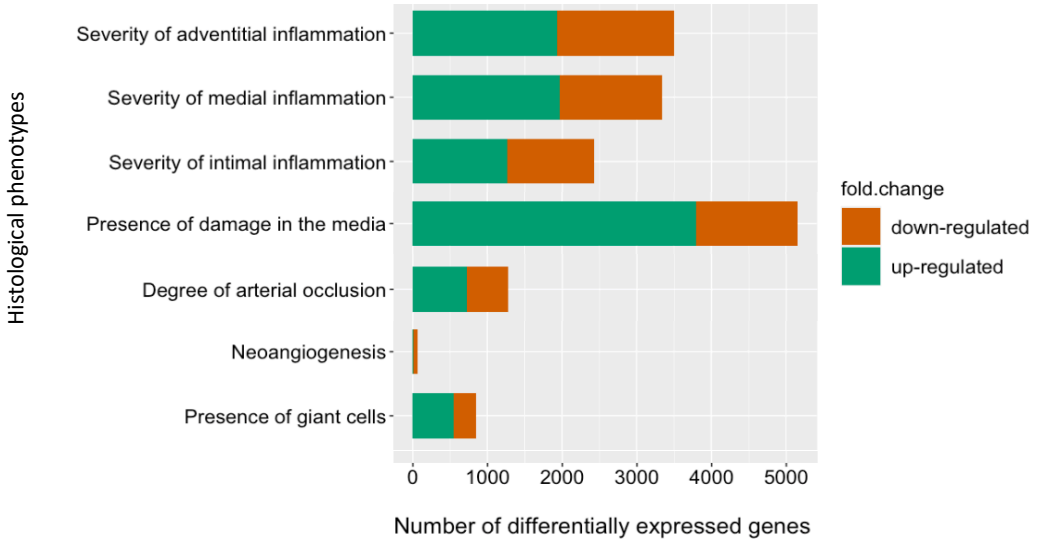


Figure 11. Number of statistically significant DEGs detected with FDR-corrected p-value < 0.01 in group comparisons for each histological phenotype. FDR – False Discovery Rate; DEGs – differentially expressed genes

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.

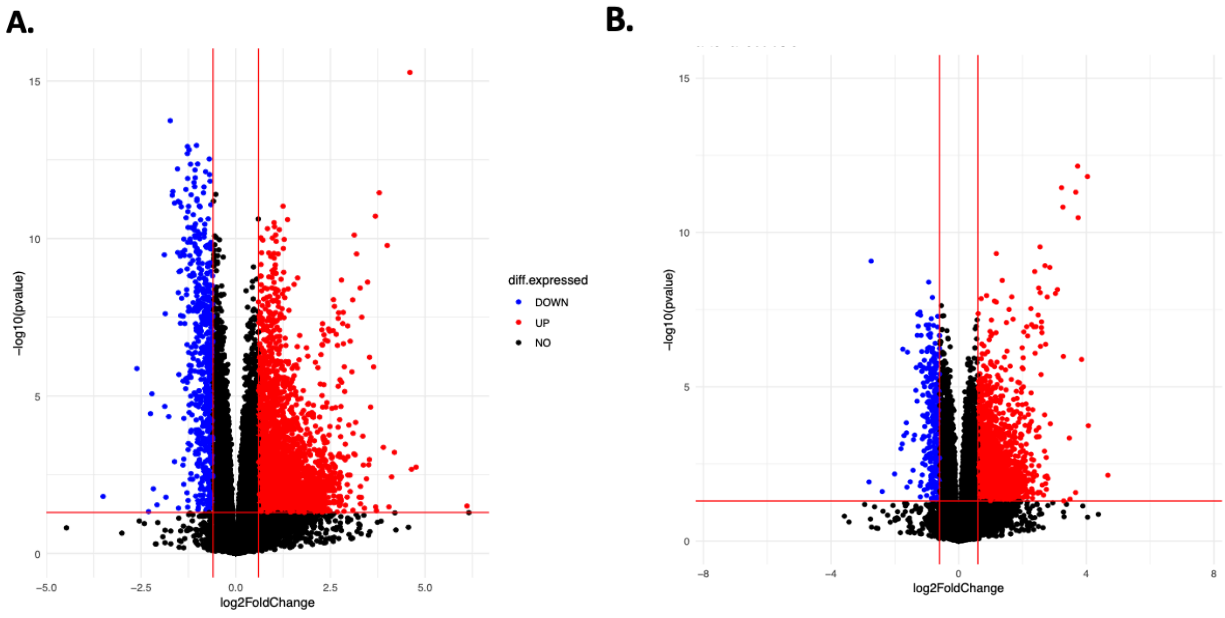


Figure 12. Volcano plots showing differential expression results for 'severity of inflammation in media' (A) and 'degree of arterial occlusion' (B). Statistically significant genes based on thresholds of p-value < $1x10^{-6}$ and log2FoldChange > 1 (up-regulated) / log2FoldChange < -1 (down-regulated) 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.

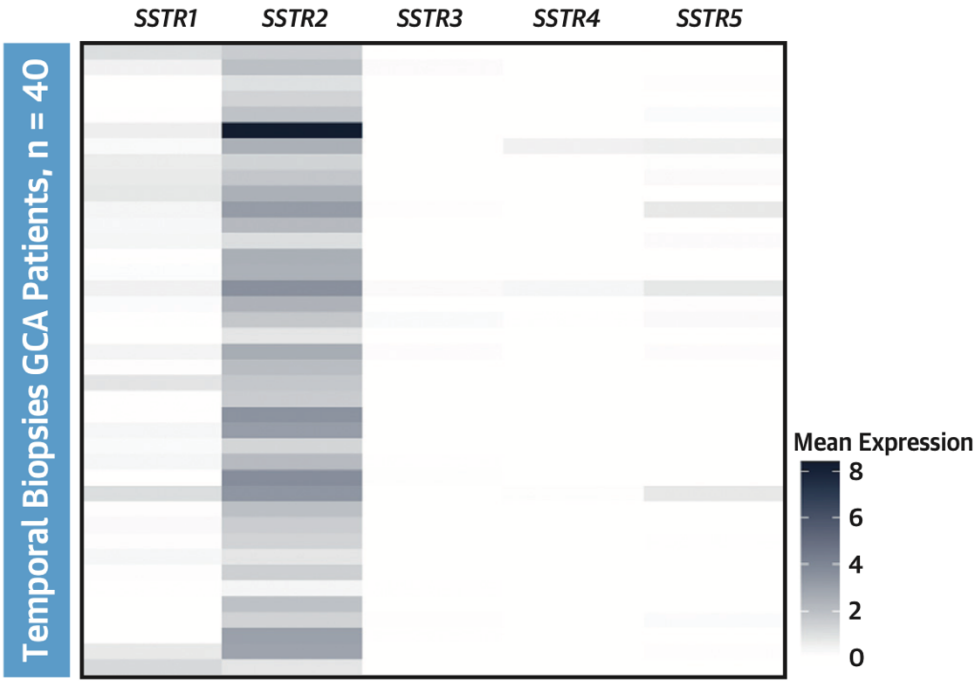


Figure 13. Heatmap of mean expression of somatostatin receptors across all patients. Figure adapted from (Ćorović et al., 2023)

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.

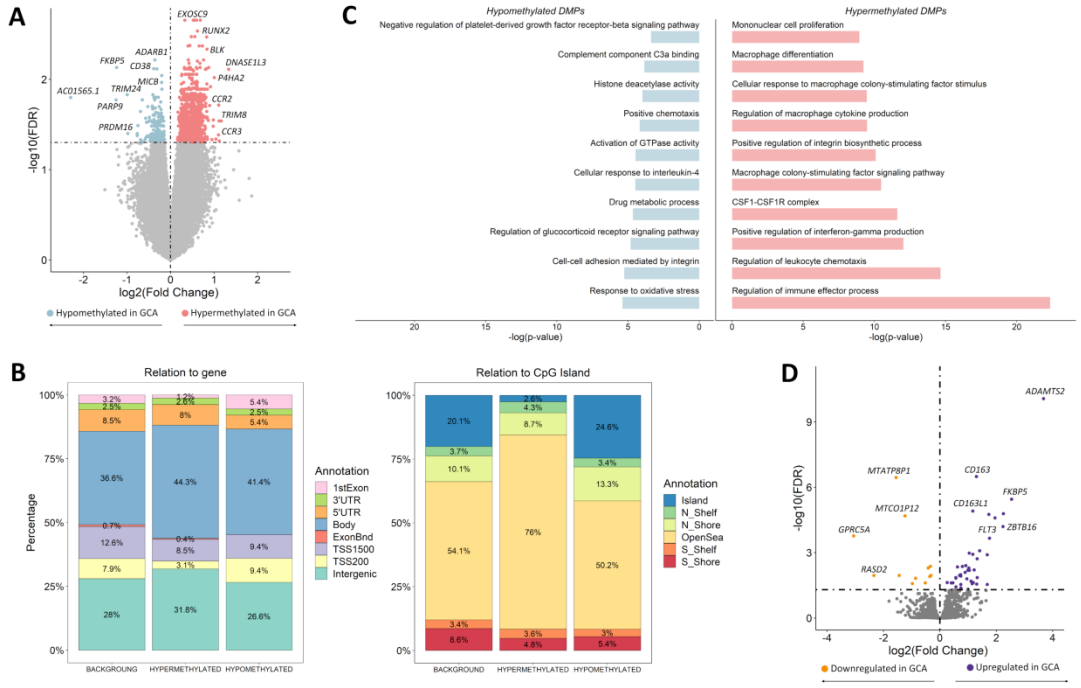


Figure 14. Results from the comparison of both DNA methylation and gene expression patterns of CD14+ monocytes between patients with giant cell arteritis and controls. A) Volcano plot of the epigenome-wide association study results. False discovery rate (FDR) values are represented on the –log10 scale 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 these 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.

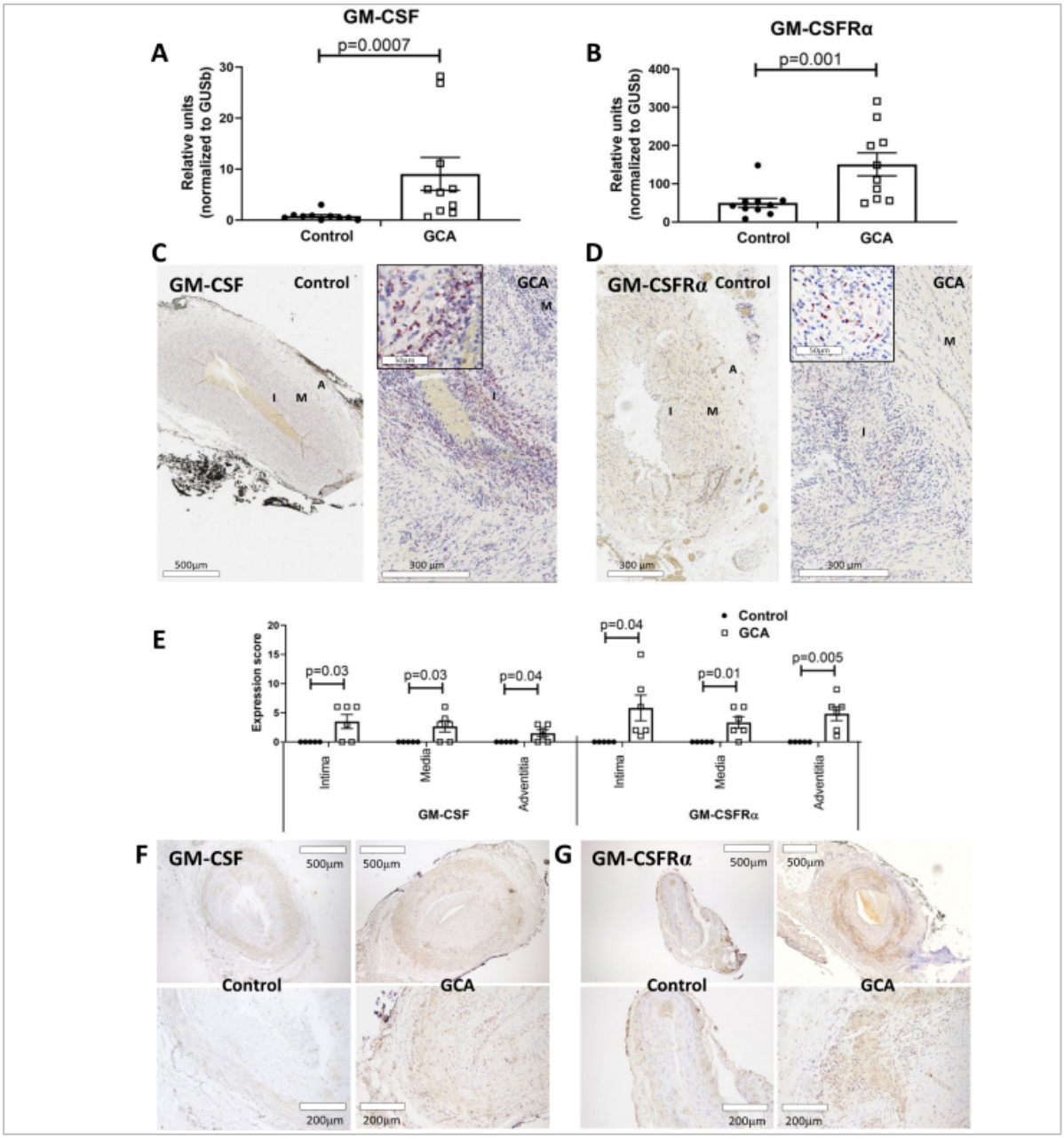


Figure 15. Granulocyte-macrophage colony stimulating factor (GM-CSF) and GM-CSFRα expression in GCA lesions. Concentrations of GM-CSF (A) and GM-CSFRα mRNA (B) measured by qRT-PCR in fresh-frozen histologically negative arteries (controls) (n=10) vs GCA-positive arteries (n=10). Results are expressed 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 prednisolone 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.

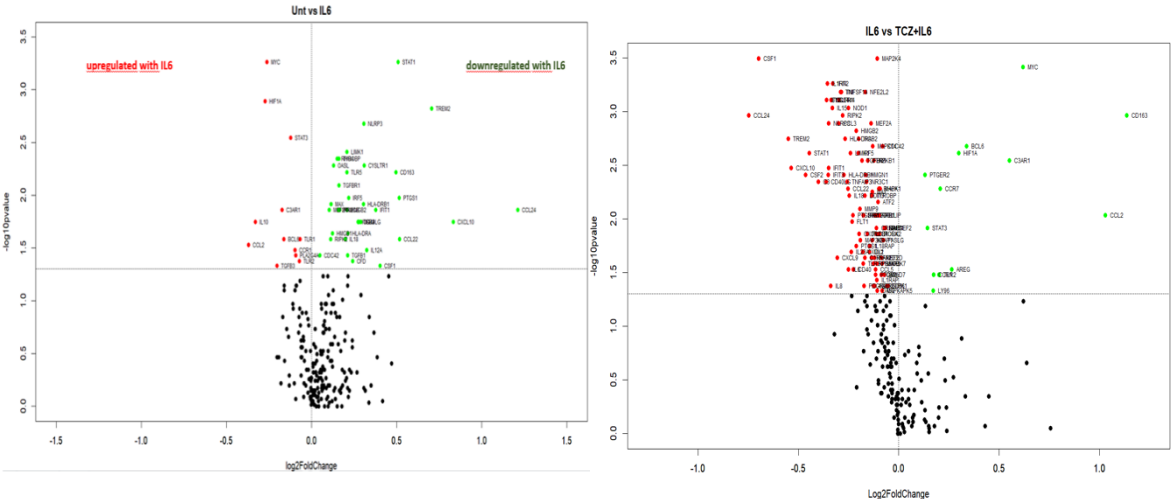


Figure 16. Differentially regulated transcripts after exposure to mavrilimumab (A) or tocilizumab (B) compared to placebo.

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.

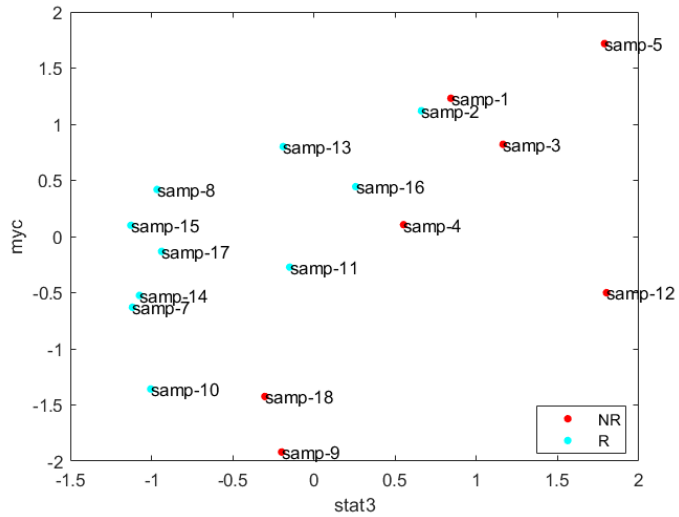


Figure 17. STAT3 vs MYC transcript levels at baseline.

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 prednisolone use and pathological conditions. Some of those correlations could point to potential unaccounted adverse drug reactions to prednisolone use. In a rheumatoid arthritis population, association of dyspnoea with prednisolone use was significant, and in asthma and COPD populations association of hypernatremia with prednisolone 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-GMCSFa) 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 several binding partners including plasmin activator inhibitor-1, integrin aV b3, 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 aV b3 (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.

Work package 3 Linkage of clinical data to plasma and tissue analyses

Lead: Peter Nilsson (KTH)

Participants: MUW, TG, FI, TCD, IBMZ, UNISTRA, KTH

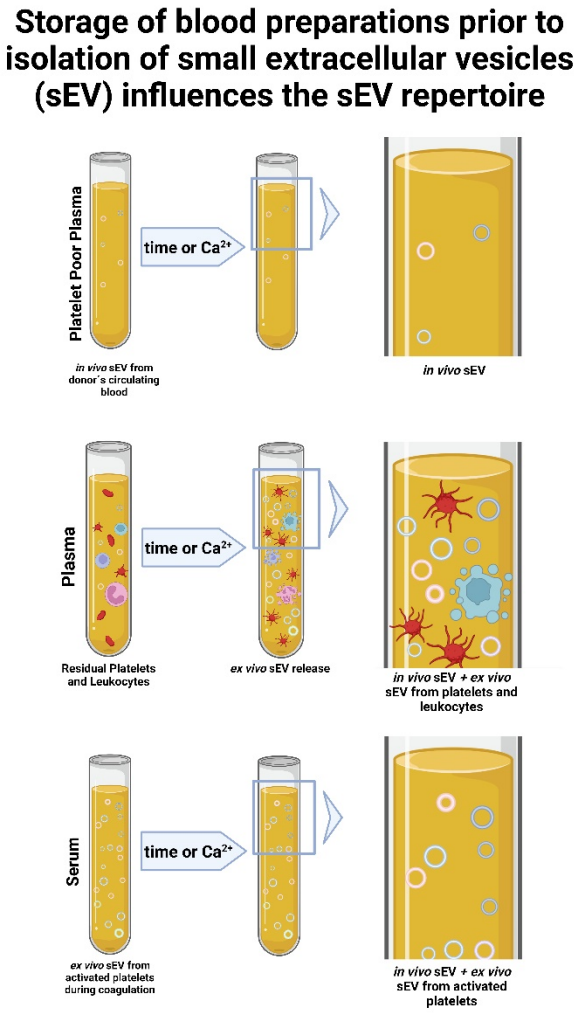
Aim

WP3 uses novel techniques to correlate results from proteomic and morphometric analyses of plasma and tissue with clinical outcome, aiming at stratifying patients with vasculitis according to disease activity, while building on the overall HELICAL theme of ethical linkage of experimental and clinical data to improve patient care. Using existing resources established in past and ongoing collaborations, the common goal of the four ESR projects is to develop clinical tools, supported by machine learning algorithms, that assist clinicians in diagnosing active vasculitis and predicting outcome. Extracellular Vesicles (EV) are derived from activated resident or circulating cells and carry a cargo of (surface) proteins, DNA and RNA that they deliver to other cells. In autoimmune diseases like AAV they are produced by endothelial cells and leukocytes in response to injury, with the specific vascular beds affected by this condition producing a different pattern of EVs. One aim of this WP is to develop and optimise techniques to identify the source of EVs and establish them as a diagnostic and predictive tool in AAV using consortium samples. Adopting an alternative strategy, we will exploit the increasing evidence from large multinational consortia like Neptune and INTEGRATE that have defined tissue morphological changes (descriptors) that can be used as predictors of outcome in renal ANCA vasculitis. However, analysis and validation of these descriptors is time consuming, hindered by inter-observer variation and by difficulties in interpreting them due to clinical outcome data variability. Based on existing descriptors and algorithms, this project aims to define morphological changes in renal biopsies from patients with ANCA vasculitis that are suited to automated morphometric analysis and subsequent validation using existing clinical outcome data. In parallel, we aim to investigate novel autoantibodies to combinations of peptides and EV proteins with a view to diagnosing vasculitis flare and defining flare risk. Part of this aim is linked to a novel test developed by SME Firalis (FI), based on an innovative set of selective soluble biomarkers that will be combined with an algorithm based on the patients' clinical data, and a web-based user interface to facilitate diagnosis of active vasculitis. 30 biomarkers have been distilled to 4 key peptides, which display excellent biomarker characteristics. The aim of this project is to take this to Technology Readiness Level (TRL) 4 in the path of commercial test development.

EXPERIMENTS CONDUCTED DURING THE SECOND HALF OF THE ACTION:

The importance of the pre-analytic pathway when studying sEV

ESR11 (MUW) worked on optimizing methods to purify and measure sEV from human blood and establish them as markers for AAV. Standardised procedures, including complete depletion of residual blood cells, are indispensable for obtaining reliable results from sEV circulating in the blood stream. The isolation of sEVs requires appropriate methods to avoid contamination, and differential ultracentrifugation (DUC) resulted as the most robust method for sEV isolation. Serum stored at -80 was identified as the most appropriate sample type for sEV studies. The results also identified potential sources of variation, highlighting the need for strict pre-analytical protocols. For sEV biomarker discovery, stored platelet poor plasma (PPP) was identified as the best sample type for sEV biomarker discovery studies, facilitating the quantification of sEV from retrospectively characterized and biobanked sample cohorts. Stored plasma and serum are less suitable for sEV studies.



Use of sEV to define organ-specific disease activity in AAV

ESR11 also worked on the development of a custom microarray for high-throughput analysis of sEV from the blood of patients with ANCA-associated vasculitis (AAV). The approach did not generate reliable or consistent results, possibly due to the high degree of glycosylation of the sEV surface proteins making antibody binding sites inaccessible. Additional approaches (i.e. enzymatic removal of glycans, heat and chemical pre-treatment) are being tested to improve the reliability of sEV profiling using antibody bead-arrays.

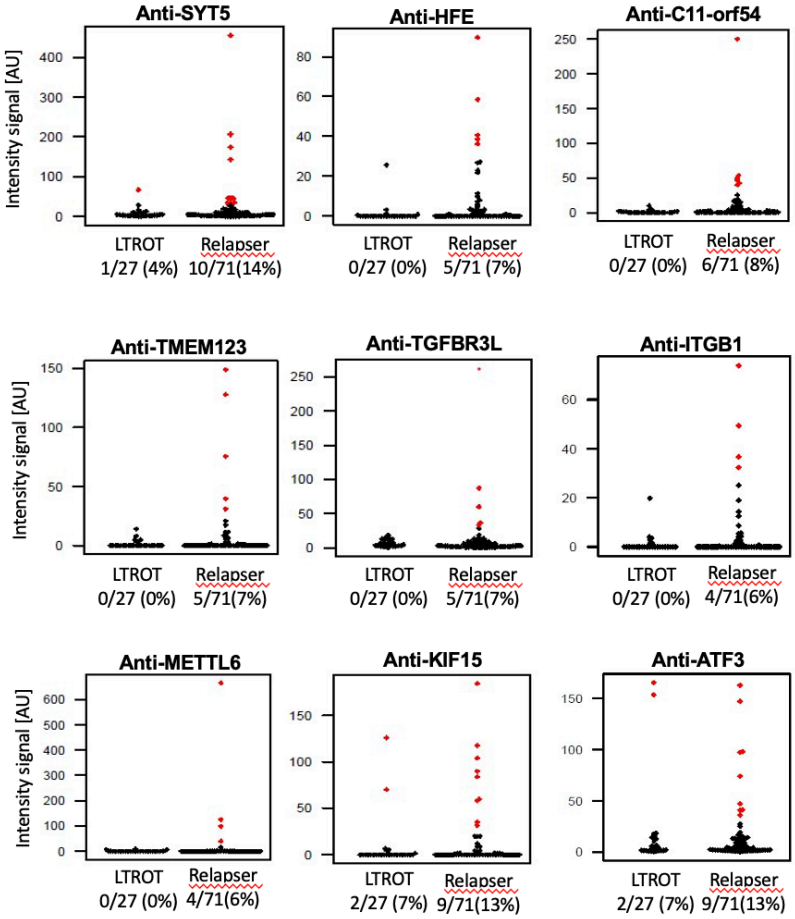
Automated analysis of renal histopathology

ESR 12 (TG) worked on developing and testing machine learning models that can identify features in biopsies of patients with AAV and assessing their suitability in predicting patient outcomes. Initially, a U-net model was used to identify glomeruli in kidney biopsies. However, the model´s performance suffered when dealing with images containing cropped glomeruli. For this reason, the model was switched to the YOLO model, which showed to be faster and better suited for clinical use to identify

PAS-stained whole slide images (WSIs). The YOLO model was trained to classify glomeruli as normal, crescentic or sclerosed. The model has been refined using a large number of WSIs from Trinity College, with linkage to granular longitudinal data to correlate with clinical outcomes. Further development of a multiple instance model (MIL) for clinical data is ongoing to predict outcomes and a framework to develop ore user friendly machine learning models.

Proteomic and transcriptomic profiling of peripheral blood to define AAV disease activity and predict relapse

ESR13 (KTH) and ESR 14 (Firalis) has been working on the serum/plasma cohort from the Irish Rare Kidney Disease (RKD) Registry and Biobank to identify autoantibodies and mRNA markers as predictors of relapse. **ESR13 (KTH)** performed **autoantibody profiling** of 98 samples collected at remission from patients classified as LTROT and relapsers by using **proteome-wide antigen array technology** and optimized workflow developed at the Protein Science Department at KTH. Nine autoantibodies with higher reactivity were observed in subgroups of samples collected from AAV patients experiencing subsequent relapse (see figure below). These bind to activating transcription factor 3 (ATF3), methyltransferase-like protein 6 (METTL6), synaptotagmin 5 (SYT5), homeostatic iron regulator (HFE), chromosome 11 open reading frame 54 (c11orf54), transmembrane protein 123 (TMEM123), transforming growth factor beta receptor 3 like (TGFBR3L), integrin subunit beta 1 (ITGB1), and kinesin 15 (KIF15). These findings are under verification for their potential to predict future relapses as single autoantibodies or autoantibody panel(s).



ESR14 (Firalis) applied the **BIOPRED RNA profiling technology**, developed to identify mRNA targets associated to autoimmune-inflammatory disorders such as rheumatoid arthritis and lupus, on the cohort from the Irish Rare Kidney Disease (RKD) Registry and Biobank to identify new transcriptomic biomarkers able to predict relapse. The technology is based on a targeted gene sequencing panel of 2155 mRNA targets associated with inflammatory and pro-inflammatory pathways, immune response pathways, interleukins, chemokines, growth factors, metalloproteinases, and more. Incorporating the 2155 mRNA from the BIOPRED panel into a logistic regression model significantly improved the accuracy of predicting different AAV clinical stages (AAV active, remission, LTROT, and relapse) and separating them from healthy individuals. The findings therefore suggest that the BIOPRED panel could be used as a precision medicine tool for diagnosing active vasculitis (TRL3). We went on to assess whether the peripheral transcriptomic profile could predict flares, allowing adjustment of medication dosage and developing personalized treatment. Here, the results were more subtle, the study under-powered and the assay not currently ready for taking forward for this diagnostic indication (TRL2). Further validation in additional patient cohorts and prospective studies are needed.

KEY SCIENTIFIC OUTPUTS:

- Optimization of methods for purifying and measuring sEVs from human blood.
- Ongoing development of a custom microarray for high-throughput analysis of sEVs from AAV patients.
- Development of machine learning models to identify features in biopsies of AAV patients and predict patient outcomes.
- Identification of 9 autoantibodies at high reactivity in subgroups of patients that will experience relapses.
- A proof-of-concept study showing the potential of the BIOPRED assay in separating active from remission vasculitis, and in predicting AAV relapse.

Published manuscripts:

- Małys, M. S. S., Aigner, C., Schulz, S. M. M., Schachner, H., Rees, A. J. J., & Kain, R. (2021). Isolation of Small Extracellular Vesicles from Human Sera. *International journal of molecular sciences*, *22*(9), 4653. https://doi.org/10.3390/ijms22094653
- Małys et al. Small extracellular vesicles are released *ex vivo* during coagulation from activated platelets and during storage from residual blood cells. Submitted to Journal of Extracellular Biology on the 29.03.2023.

Manuscripts in preparation:

- Bayati S, Scott J, Little M, Nilsson P, Pin E. Identification of autoantibodies for the prediction of relapses in ANCA-associated vasculitis.
- Mescia F, Bayati S, Brouwer E, Rees A, Kain R, Lyon P, Nilsson P, Pin E. Anti-kinesin antibodies associated to ANCA-vasculitis. To be submitted to the International Journal of Molecular Sciences.
- Matthieu Coq, Yagmur Dogay, Gisella Pattarone, Huseyin Firat, Jennifer Scott, Caitriona McEvoy, Mark A Little. Quantification of ANCA vasculitis disease activity and relapse risk using peripheral blood transcriptomic profiling. To be submitted to Arthritis and Rheumatology.

Work package 4 Governance of electronic health record sharing and linkage

Leads: Dipak Kalra and Nathan Lea (IHD), ESR 15 Maria Christofidou

Aim

Regulatory differences between countries and uncertainty as to what ethics and respect for privacy requires is a major bottleneck when trying to enhance the sharing and utility of personal data for biomedical research. There is ample anecdotal evidence of how overly strict measures for privacy protection are detrimental for bringing research to patient benefit. However, differing Member State and institutional interpretations of Europe's flagship data protection regulation, the GDPR, add to the challenges through inconsistency of decision making and the difficulty of determining what would be acceptable practice in applying data protection safeguards. Rare disease research is especially challenging due to small patient numbers and distinctive clinical patterns. WP4 participants developed a Data Management Plan that is GDPR and FAIR consistent. Through novel research they have worked on an Ethical Framework that facilitates sharing and access to data across borders, while respecting national and European law.

Experiments Conducted:

These are summarised below along with the scientific results.

Key Scientific Results:

Work done and outputs supporting information governance, data protection and data sharing across the project

- Created a tailor-made Data privacy Impact Assessment (DPIA) for the project - used both for the project in its entirety and by the ESRs as living documents for the duration of their projects - which reflects in a comprehensive manner how the FAIR principles are met, the GDPR-relevant grounds used for the obtaining data, the infrastructure in place to keep the data secure and addresses concerns that patients may have.
- Created the HELICAL Open Research Guidelines, a document aimed at being used as a set of recommendations and step-by-step instructions to help ESRs and experts share their research findings (whether aggregated data, software or algorithms) as open as possible in accordance with the relevant data protection rules and legislation.
- Aided in the review and finalisation of the individual DPIA across the partners, aligned with an over-arching project wide DPIA and other educational material on data protection/privacy relating to the project.
- The creation, finalisation and presentation of the Proposed Governance Framework, a deliverable which proposes a governance framework which can act as a "tool kit" that may be utilised by European projects conducting research in rare disease areas.
- Reviewed and updated the Data Management Plan to ensure that the changes that occurred since the first draft were reflected accurately.
- Assistance with convening, running and reporting on the Ethics Advisory Board and addressing of ethical questions raised during the mid-term review.

Research undertaken by ESR 15: Maria Christofidou

- Publication of the first article contributing to the ESR 15 PhD, containing a literature review on the topics of GDPR, COVID-19 and the ethical considerations of data protection in the time of crisis arising as a result of practical learnings through the project. The article was published in the IMIA Yearbook in 2021.
- Research then went into the area of **'consent'** (Arts 6 and 9 GDPR) and how the legal basis has caused practical difficulties when used in the space of healthcare and scientific research in the reuse of data. This research then carried into and encompassed the newly proposed Data Governance Act and the grounds to provide consent based on 'altruism'. This therefore then gave rise to research around this area, which is highly relevant to the project, and to a PROPOSERO paper and manuscript which is being finalised in June 2023.
- Presentation of the research conducted as part of HELICAL and WP4 as well as the papers drafted during this time was presented in a variety of conferences and workshops. Some examples include, the BioData World Congress with a presentation on "Data Protection, GDPR and Rare Diseases" in 2020, the Internet of Things (IoT) Week in 2021, the DigiHealthDay Workshop/Conference in 2021, the European Association of Health Law (EAHL) Annual Conference in 2022, more recently the Computers, Privacy and Data Protection conference (CPDP) in 2023.

Publications:

- Christofidou M, Lea N, Coorevits P. A Literature Review on the GDPR, COVID-19 and the Ethical Considerations of Data Protection During a Time of Crisis. Yearb Med Inform 2021; 30(01): 226-232. DOI: 10.1055/s-0041-1726512

Work package 5 Structured training

The primary aim of HELICAL is to provide outstanding training to 15 ESRs delivered through a multidisciplinary and multisector structure that will enhance their employability and innovative capacity by structured doctoral training, collaboration and exposure to academia and industry. The training programme was structured through **supervised research towards a doctoral award** and **local / network-wide training activities**.

The Doctoral Studies Committee oversaw network-wide training. It was led by Renate Kain, and included Jessica Grene, Nathan Lea, Elisa Pin, Lucy Hederman, Javier Martín, Helen Cameron, and Alfred Mahr. The DSC monitored selection procedures, gender balance and transparency of the national procedures, as well as monitoring training progress. The committee met twice a year.

Core research training, acquired mainly through the IRP; 2) **Advanced/Additional Research Skills** (delivered by the consortium); 3) **Transferable Skills** (delivered by the consortium, particularly those useful in non-academic careers). These were delivered across three frameworks, embedded in Open Science policy 1) Local training; 2) Network-wide training; 3) Secondments.

Core research training

In HELICAL, skills were acquired via network-wide training modules, secondments and individual research projects (IRPs), under supervision of their Supervisory Panel.

Doctoral Awards

The completed or expected dates of PhD submission of thirteen HELICAL Early-Stage Researchers are listed below.

ESR number	Name	Thesis defence date	
1	Albert Navarro Gallinad	Defended 23rd May 2023	Awarded
2	Anna Weber	July 2023	
4	Enock Havyarimana	July 2023	
5	Solange Gonzalez Chiappe	Defended December 2022	Awarded
6	Alejandro Fontal	September 2023	
7	Elkyn Estupiñan Moreno	Defended 19th May 2023	Awarded
8	Michal Zulcinski	July 2023	
9	Filippo Guerri	December 2023	
10	Farah Kamberovic	December 2023	
11	Malgorzata Malys	July 2023	
13	Shaghayegh Bayati	Defended June 2nd 2023	Awarded
15	Maria Christofidou	September 2023	

The exceptions are as follows:

- ESR3 Bahareh Khosravi, at Trinity College Dublin, did not progress to the PhD register after her first year. With the permission of the project officer, she continued her fellowship and research project without being enrolled in a PhD.
- ESR12 Marco Zanet with TissueGnostics joined the project in April 2021, after the resignation of the researcher initially recruited to the role, with a contract duration of 19 months. Marco decided against continuing a PhD after the end of the action.

Two researchers, Matthieu Coq and Yagmar Dogay, were recruited to Trinity College Dublin to work on the ESR14 research project on short-term contracts which did not allow enrolment in a PhD programme. Please see WP6 Management and Recruitment report below, for the details of the ESR14 role.

Figure 18. *ESR7 Elkyn Estupiñan Moreno with his PhD thesis after successfully defending at CSIC 19.05.2023*

Network-wide training

In the second reporting period of the action, **Module 6** Intellectual Property, **Module 7**: The Innovation Pathway, and **Module 8**: Using Your Transferable Skills to Drive Your Career: ESR Innovation and Leadership, were held in accordance with the Description of Action. Additional training was held in response to perceived needs, and requests from ESRs. These were on communications, researcher mental health and well-being, and career building. Scientific lectures on *Feature Selection in Rare Disease research*, and *Aspects of Gender Medicine in Biomedical Research* were included in the HELICAL network-wide meeting in November 2022.

Module 6 Intellectual Property, and Module 7: The Innovation Pathway, IBM Zurich, June – July 2021

Module 6 Intellectual Property, and Module 7: The Innovation Pathway, were combined and hosted by IBM Zurich online over three sessions, on the 23rd and 30th June, and 7th July 2021.

The aim of these modules was to provide the ESRs with a good understanding of how to identify intellectual property for commercial exploitation, and the processes involved in bringing it to market. The ESRs heard from experts about these topics, and from entrepreneurs about their own experiences of innovation. The ESRs then participated in a practical group expertise in teams who planned and executed pitches to a panel of judges.

Schedule of Modules 6 and 7:

HELICAL Modules 6 and 7 Schedule
Day 1 - Wednesday 23.06.21 (13:00- 16:30 CET)
13:00 – 13:45 From idea to market, introduction of basic concepts: Aurelien Pelissier
13:45 – 14:30 Practical cases: Experiences creating your own company: Christoph Aufricht
15:00 – 15:30 Practical cases: Experiences working in a mid-size company: Rupert Ecker
15:30 – 16:00 Practical cases: Experiences working in a mid-size company: Judith Farrés
16:00 – 16:30 Q&A
Day 2 - Wednesday 30.06.21 (13:00- 16:30 CET)
Teamwork: assembling teams, identifying ideas.
Day 3 - Wednesday 07.07.21 (13:00- 16:00)
Pitch presentation to "Investor's panel" (the PIs).

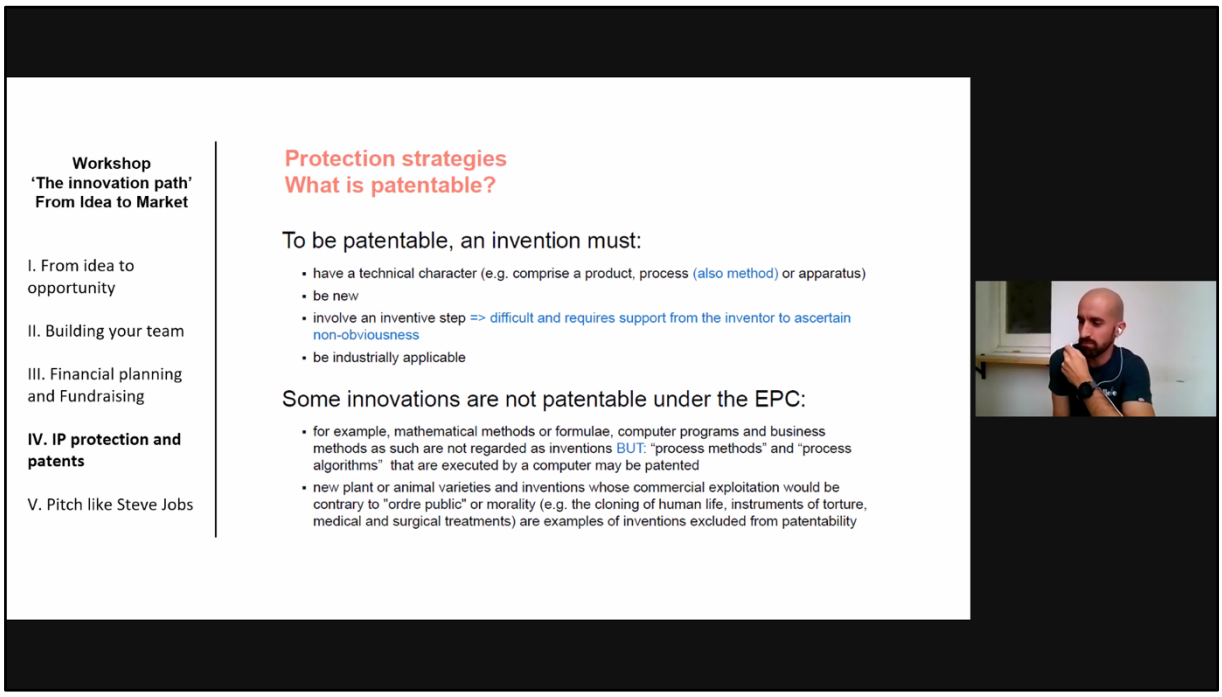


Figure 19. *Module 6 Presentation on Intellectual Property 23.06.2021 Screenshot*

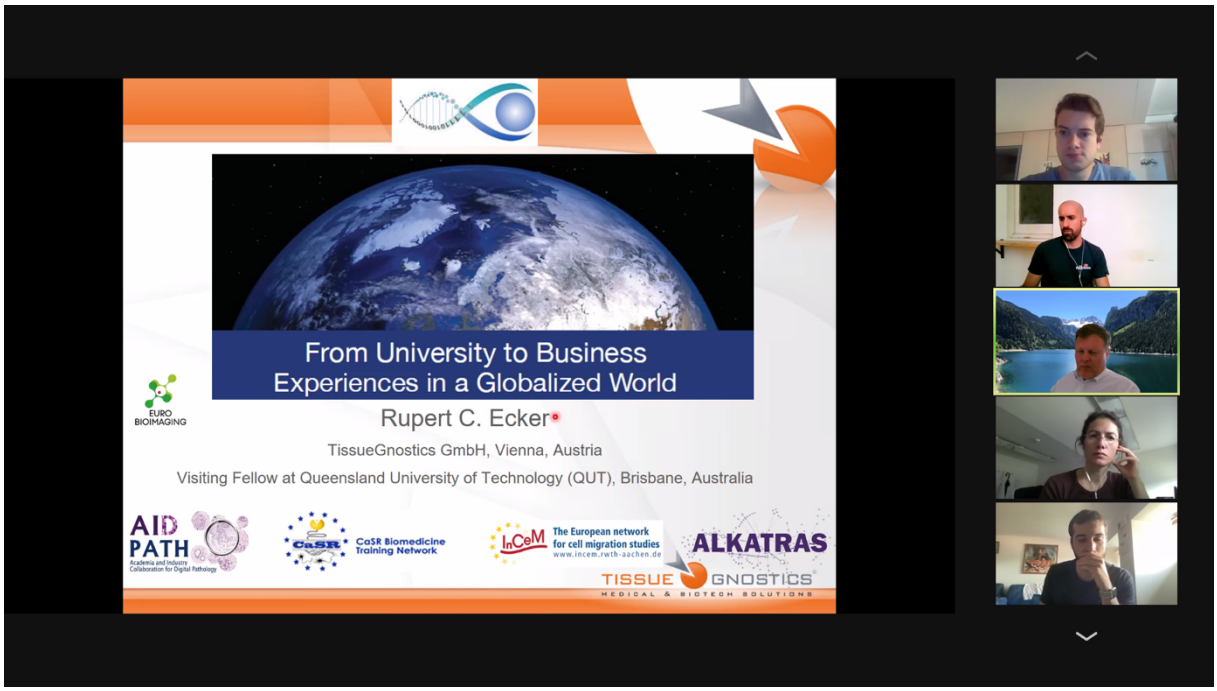


Figure 20. *Module 6 Presentation on practical experience in business; Screenshot 23.06.2021*

The researchers formed four teams, and each developed an idea for a start-up, assigning roles within the companies. Each presented their pitch and received scores and feedback.
The judges for the exercise were Aurelian Pelissier, IBM Zurich, Judith Farrés, Anaxomics, Renate Kain, MUW, and Lucy Hederman, TCD.

List of start-up companies developed and pitched, with presenting teams and judges' scores

Start-up name	Presenting team	Score
AAA Precision Cancer Treatment	Albert Navarro Gallinad, Anna Weber, Alejandro Fontal	14.375
SignUp (translation service for deaf people)	Marco Zanet, Michal Zulcinski, Maria Christofidou	17.75
SeekBIO	Malgorzata Malys, Filippo Guerri, Bahareh Khosravi	10.875
Human Teeth Biobank	Solange Gonzalez Chiappe, Shaghayegh Baygati, Enock Havyarimana	14.5625

Evaluation criteria for start-up presentation pitches.

Evaluation Criteria /20
Presentation quality (/5)
Gesture, story, link between slides, slide simplicity, energy of the speaker
Feasibility (/5)
Is the proposed product realistic?
Market opportunity (/5)
Is there really a market for it, are you solving a real problem?
Business model and financial planning (/5)
Is the cost estimate realistic? Are you profitable?

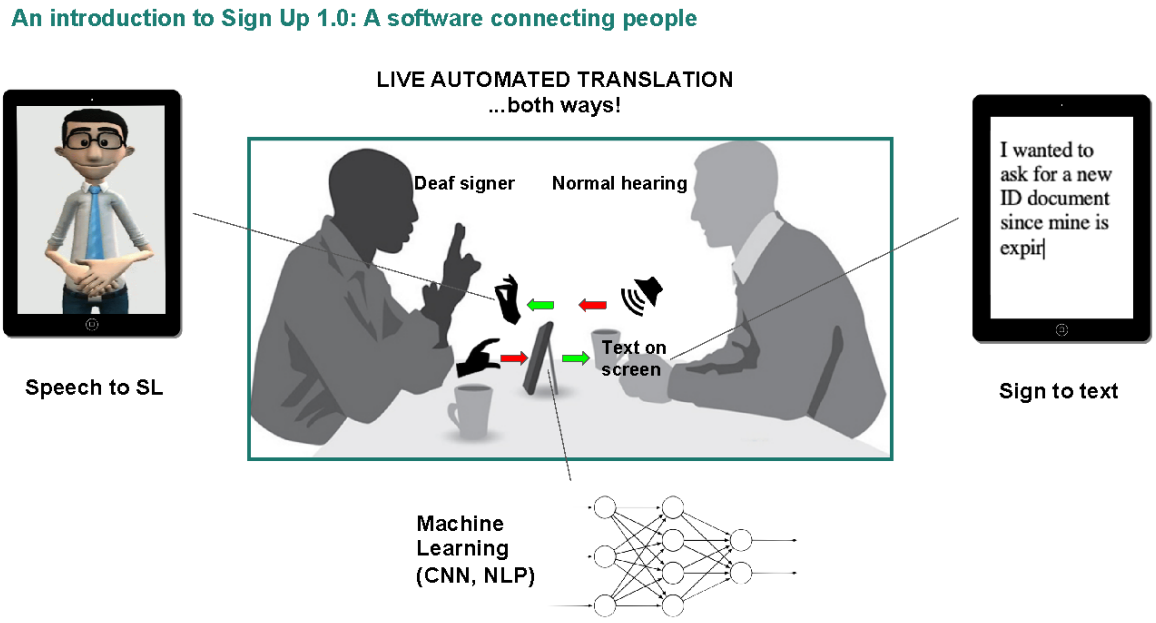


Figure 21. *Module 7: Slide from SignUp, the winning start up pitch presentation.*

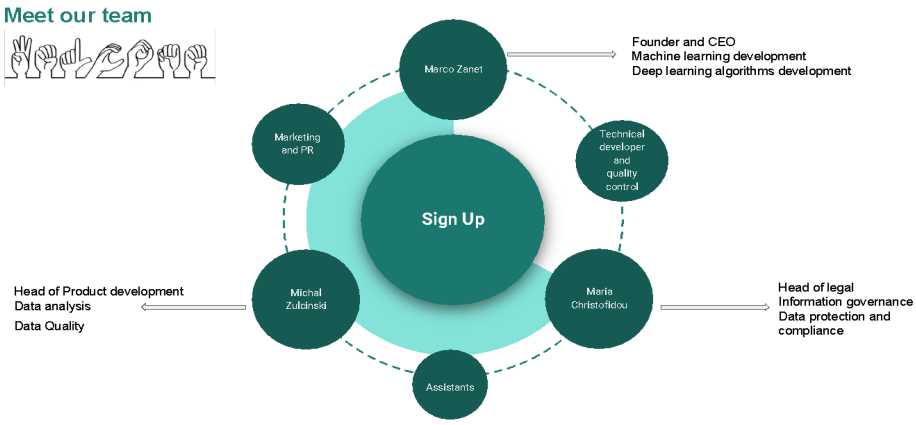


Figure 22. *Module 7. Slide from Sign Up, the winning start-up pitch presentation*

How would you rate your overall experience of the IBM Zurich workshop?	**Please comment on why this is.**
Excellent	it was a new topic for me, and I was interested to know how to make my results impactful in the future after I graduate It highlighted important points on how to set up your own startup. From idea formation to execution and the overview of finance and what investors are looking for. The practical experience was also very helpful and i learned a lot in 3 sessions. Very well-done Aurelien!
Good	I've learned some new interesting things and become aware of many aspects of business.
Very good	Good training The organizer and speakers efficiently transferred their experiences to us and they also showed us another possible career option after our PhD It would need more time, but given the time available it was really nice and interesting Interesting topic, gave both an overview of how different models work and how these are then applied into practice by looking into examples. The fact that we were then able to apply these into practice made the understanding of the material better the presenters were well organised It was good practice for giving a pitch to investors. It was very interesting to listen the stories of experienced researcher in the private field

Figure 14. *ESR Feedback on Module 6 and 7.*

Module 8: Using Your Transferable Skills to Drive Your Career: ESR Innovation and Leadership, Trinity College Dublin, March 2022.

This module was hosted in-person by Trinity College Dublin from 28th -31st March 2022. It took place in Tangent, a space within Trinity offering programmes in in innovation, creative thinking, and entrepreneurship. The workshop aimed to empower researchers with the confidence and agility to lead out on innovation, and to innovatively respond to opportunities and challenges. It moved from abstract concepts to practical methods to exercise innovative ability. The workshop was interactive and helped the researchers to consider their associations with innovation, and their innovative potential, through discussion and practical group exercises.

The workshop topics over the four days were as follows:

March 28th: Developing Innovative Leadership

March 29th: Introducing Creative Thinking and Creative Problem Solving (CPS)

March 30th: Design Thinking for Research

March 31st: Inspiring Creative Collaboration

Tangent Trinity's Ideas Workspace

Helical Research Training Programme, March 28th - 31st

High level programme outline (subject to change)

Day 1: Developing Innovation Leadership: 28th March 2022	
10am-1pm	Basadur Innovation profile for individuals and teams
1pm-2pm	Lunch
2pm-3:30pm	Guest Speaker – Mike Wride *Transformative Pedagogies Lead – University of Limerick* *Founder and Director – Learning with Creativity*
3:30pm-5pm	High Performing Creative Teams

Day 2: Introducing Creative Problem Solving: 29th March 2022	
10am-1pm	The Creative Problem Solving (CPS) Process – An introduction
1pm-2pm	Lunch
2pm-5pm	Creative Thinking for Researchers – Convergent and Divergent thinking

Day 3: Design Thinking for Research: 30th March 2022	
10am-1pm	Design Thinking Process and Methods – An introduction
1pm-2pm	Lunch
2pm-5pm	Human-centred techniques – Customer journeys and empathy mapping

Day 4: Inspiring Creative Collaboration: 31st March 2022	
10am-1pm	Presenting with impact
1pm-2pm	Lunch
2pm-4pm	Action Planning - For 2022 and beyond Helical programme

Hope to see you in-person in the Tangent building but for those who don't make it to Dublin:
Zoom meeting link: Helical Network - Innovation Training Programme
Meeting ID: 975 3740 8926
Password: innovation

Figure 23. *Module 8 TCD Workshop schedule*

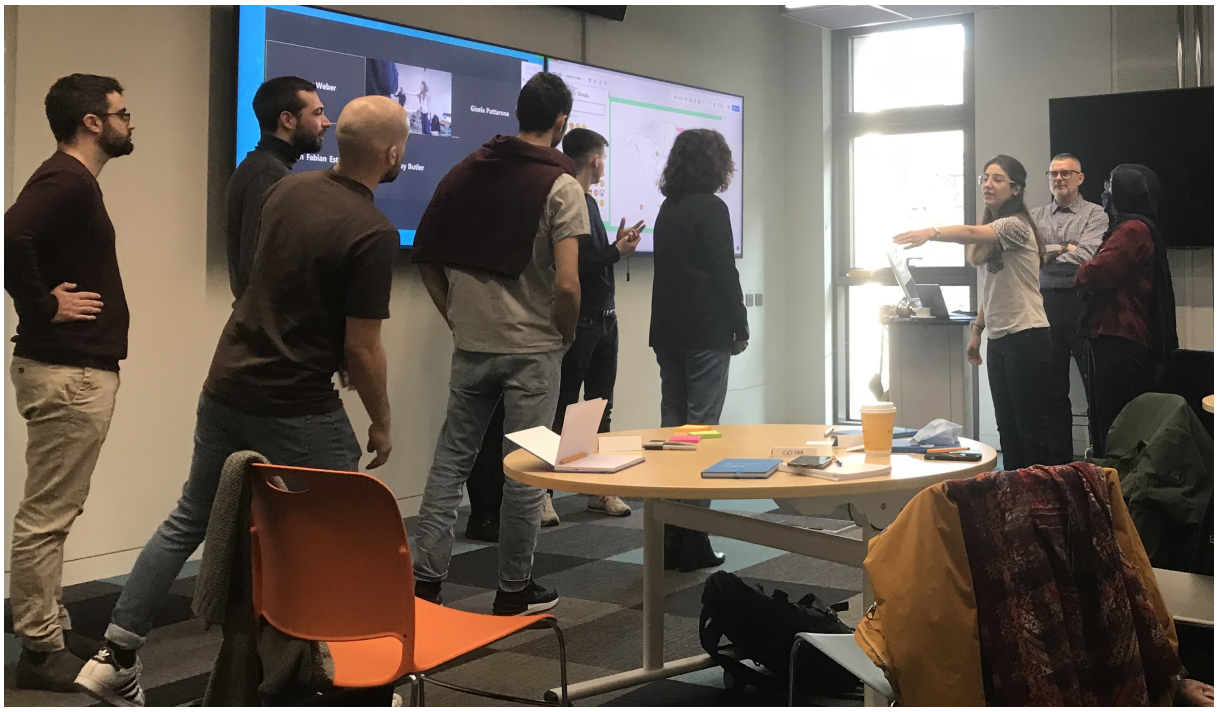


Figure 24. *Researchers, Module 8 TCD 28.03.2022*

Figure 25*. Researchers, Module 8, 29.03.2022*

Figure 26. *Researchers, Module 8 30.03.2022*

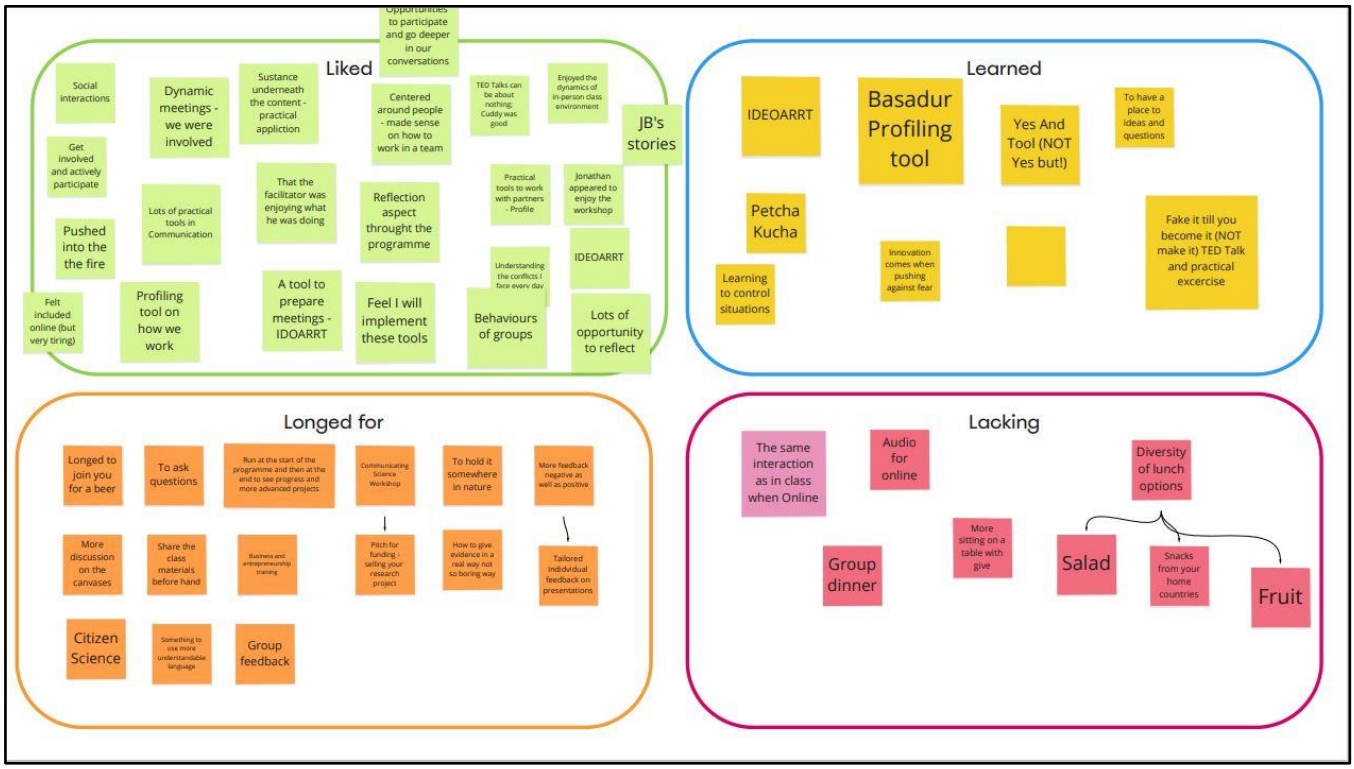


Figure 27. *ESR Feedback on Module 8*

Additional Network-Wide Training

In response to needs and requests from the research fellows, several additional trainings were held over the action. The Covid 19 pandemic and its resultant restrictions on travel and meetings necessitated a shift in dissemination practices, and the ESRs participated in a workshop on facilitation skills in online fora in February 2021, and a workshop on scientific messaging for the public in July 2021. The researchers were at risk of feelings of isolation and anxiety the social restrictions across Europe during the pandemic. To help to address this, Dr Darragh McCashin spoke to the researchers about mental health, and the resources available to MSCA Fellows in April 2021.

Prior to the HELICAL Symposium hosted by KTH Sweden in February 2023, the fellows requested some further training on career development. This took place on 9th February 2023.

Moving Dialogue Online: Facilitation Skills Training, 25 Feb 2021

Moving Dialogue On-line Facilitation Skills

Workshop by Sticky Dot Consulting

10:00am - 12:00pm

Programme:
10:00 - 10:10 Introduction and icebreaker
10:10- 11:00 Know your audience: short presentation followed by practical task on audiences and what can be relevant for them
11:00 - 11:05 Break
11:05 - 11:10 Twitter 101: brief presentation about using Twitter for science communication
11:10 - 11:20 Keep It Simple Stupid a.ka. KISS: participants write a lead to a story according to rule: "What, where, when, who" in less than 280 characters
11:20 - 11:50 Live action exercise: write a Tweet
11:50 - 12:00 Discussion and wrap up

JUL 30, 2021
HELICAL ESR training: Scientific Messaging for the Public
10:00am - 12:00pm

Schedule:
10:00 - 10:15 Introduction and icebreaker
10:15 - 10:35 Why and how to engage people with science on social media
10:35 - 11:00 In a galaxy far, far away… - what makes a great story? Short introduction about storytelling followed by practical exercise on creating imaginary stories about science
11:00 - 11:05 Break
11:05 - 11:15 Presentation of results of previous exercise
11:15 - 11:20 Videos 101: Brief presentation about using videos for science communication
11:20 - 11:50 Live action exercise: create a storyboard
11:50 - 12:00 Discussion and wrap up

Researcher Mental Health: Dr Darragh McCashin Wed, Apr 21

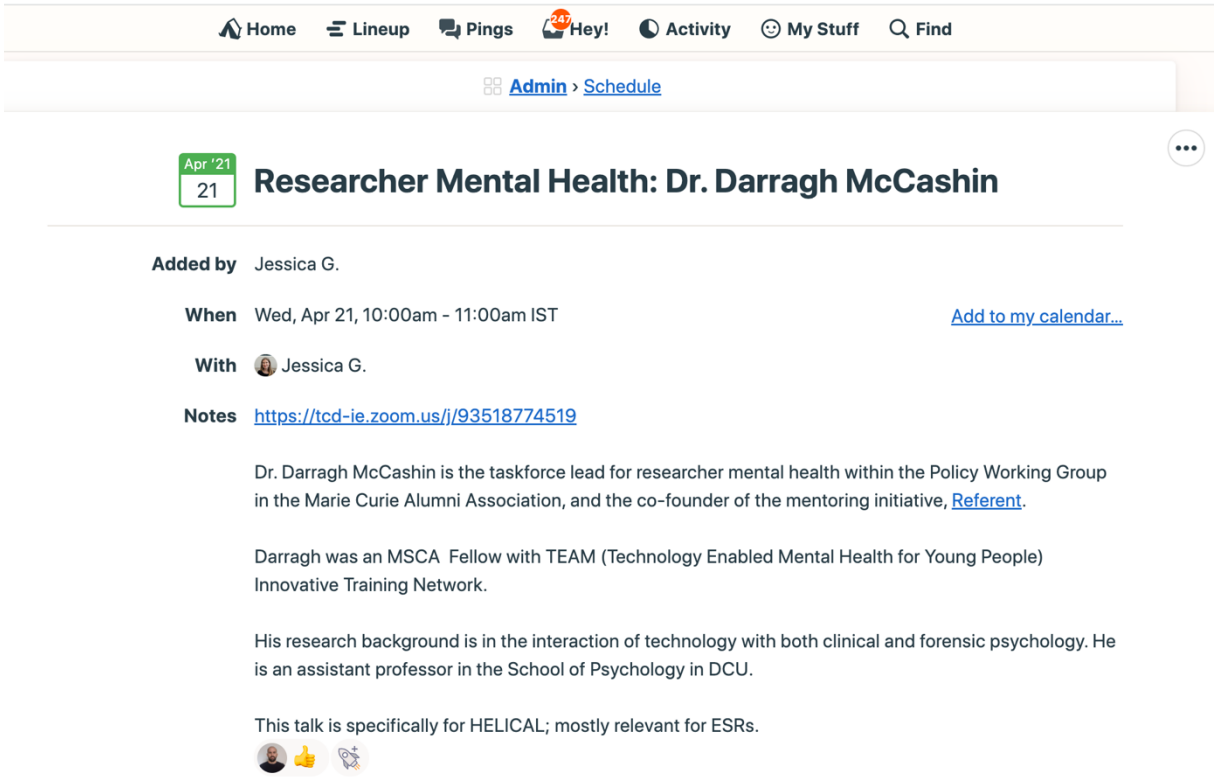


Figure 28. *Researcher Mental Health session announcement, project calendar.*

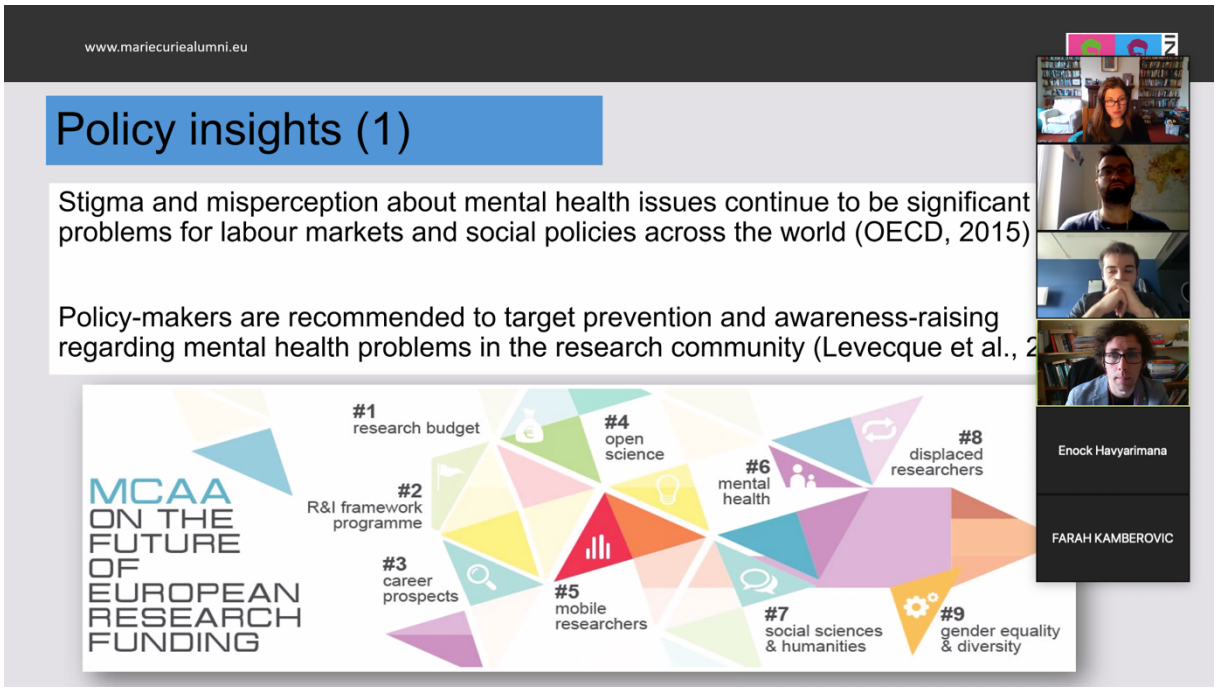


Figure 29. *Researcher Mental Health Session 21.04.2021 Screenshot*

Scientific Messaging for the Public: JUL 27, 2021

Figure 30. *Scientific Messaging for the Public: Schedule*

HELICAL Career Coaching, KTH, Sweden Feb 8 -10 2023

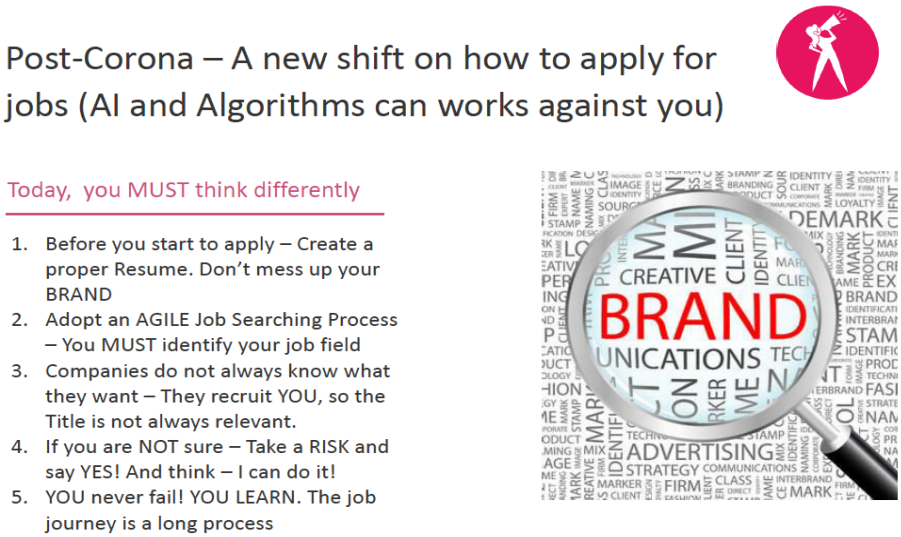


Figure 31. *Resources from Career Coaching Training, Sweden*

Secondments

As outlined in Deliverable D5.2 Evaluation of Secondment Outcomes, there were disruptions to the secondment planned because of the Covid-19 pandemic due to travel restrictions and campus closures.

In the cases of four planned secondments, a continuous period of full-time work at the secondment host organisation was not possible. In these cases, scientific collaboration based on regular interaction and data-sharing over the second and third year of the ESRs' fellowships are in place in lieu of secondments.

Four new partners were added to the HELICAL consortium. The University of Edinburgh, with host supervisor Athina Spiliopoulou, Chancellor's Fellow at the Usher Institute of Population Health Sciences and Informatics, and Newcastle University, whit host supervisor Dr. Gary Reynolds, Clinical Lecturer in Rheumatology at the Translational & Clinical Research Institute, both acted as hosts for collaboration with Michał Zulcinkski, HELICAL ESR8, University of Leeds.
The University of Kaiserslautern-Landu was added as a partner, with Prof. Dr. Sebastian Vollmer, Professor of Machine Learning Applications in the Faculty of Informatics supervising the secondment of Bahareh Khosravi, HELICAL ESR3, TCD.

Secondments Outline

Name	**Home**	**Secondment host**	**Secondment supervisor**	**Dates**
ESR7 Elkyn Estupiñan Moreno	CSIC	IDABAPS	Maria Cid	15-09-2020 to 15-12-2020

The ESR will perform this secondment with the aim of learning different strategies for the analysis of epigenetic data, in particular DNA methylation data. Thus, he will fulfil one of the objectives set out in his thesis project; "To analyse and to compare the methylation patterns of different subgroups of individuals affected with giant cell arteritis (GCA) and healthy individuals". In addition, the ESR will obtain samples and clinical data from patients with GCA, information necessary for the execution of our project "Identification of functionally relevant genetic variants associated with GCA" as part of the consortium HEalth data Linkage for ClinicAL benefit (HELICAL) a Marie Skłodowska-Curie Innovative Training Network. With this secondment, we hope that the ESR will advance and obtain relevant results that allow us to understand gaps in the clinic and in GCA physiopathogenesis.				
ESR13 Shaghayegh Bayati	KTH	TCD	Mark Little	28-09-2020 to 23-10-2020
I had access to the patient list of the RKD biobank and with the guidance of Biobank organizers and managers, I will learn how the biobank organized all the data during past years, patient's history of disease, their status of treatment, terminology of the biobank data and choose samples based on the agreed criteria. For my first stage of experiment, I will choose 12 samples and will need both Long Term Remission Off Therapy (LTROT) and frequent relapsing patients. We agreed to choose, 3 LTROT MPO, 3 Relapser MPO, 3 LTROT PR3, 3 Relapser PR3. Samples should be either serum or plasma (not any mix). Furthermore, I will select and include more samples for the second phase of the study.				
ESR10 Farah Kamberovic	IDIBAPS	CSIC	Javier Martin	15-09-2020 to 15-12-2020
The ESR will perform this secondment with the aim of learning strategies for the obtention and analysis of genetic data (GWAS, exome sequencing) and contribute to the project "Identification of functionally relevant genetic variants associated with giant-cell arteritis (GCA)". The ESR will also explore genetic variants in the IL-6 receptor gene of relevance to her PhD project (predictors of response to IL-6 receptor blockade with tocilizumab) in a large cohort of patients with GCA. With this secondment, the ESR will learn state of the art techniques in large-scale genetic studies as well as bioinformatics analysis.				
ESR1 Albert Navarro Gallinad	TCD	I-HD	Dipak Kalra	06-04-2021 to 30-06-2021

ESR1 is developing a Semantic Environmental and Rare Disease data Integration Framework (SERDIF) that aims to support researchers who require a flexible methodology to integrate environmental data with longitudinal and geospatial diverse clinical data in their hypothesis exploration of environmental factors for rare disease research. However, the patient data protection risk when combining, accessing and exporting environmental data with clinical rare disease data has not yet been assessed.
During the first secondment aligned with the HELICAL funding in i~HD from April to June 2021, the ESR will work towards the achievement of the following objectives:

1. Gain understanding in data protection and GDPR topics within the scope of ESR1 research. This will especially focus on how the GDPR applies to the reuse of data for research, and the challenges of applying data protection safeguards in a rare disease context.
2. Assess whether SERDIF data lineage transparency complies with GDPR for environmental-patient linked records, and also what lineage metadata is required for trusted research use (data provenance).
3. Provide SERDIF with the required data protection steps to be applicable in international context besides GDPR, for example by re-examining the initially conducted DPIA to ensure it is as robust as possible.
4. Explore the steps required to make SERDIF available as an Open research framework, including how best to incorporate FAIR metadata and to align with the FAIR principles.
5. Assess the suitability of SERDIF to become a component of an early warning system for public health management Subject Matter Experts (SME), such as the APIs may need to support and what quality assurance would be ideally needed for this to be a trusted component within a decision support architecture.

ESR11 Małgorzata Malys	MUW	KTH	Elisa Pin	18-10-2021 to 05-11-2021

The ESR will perform this secondment with the aim of optimising the bead-based antibody assay for detection of proteins on small extracellular vesicles purified from human blood and cell culture supernatant. This is an essential step to determine whether the assay can be applied on small extracellular vesicles and further be used for the characterisation of protein repertoire on small extracellular vesicles isolated from patient sera that are suffering from ANCA-associated vasculitis. Additionally, the initial analysis of small extracellular vesicles derived from various cell lines will enable to find the linkage between the specific protein repertoire and the cell origin. With this secondment, we hope that the ESR will optimise the assay and characterise the cell culture derived small extracellular vesicles that play the key role in ANCA-associated vasculitis pathogenesis.

ESR4 Enock Havyarimana	UG	University of Aberdeen	Corri Black	01-03-2021 to 10-05-2021

The ESR will perform this secondment with the aim of learning about Scotland's Public Health data environment. He will have access to a rich dataset containing detailed information about patients with ANCA-Associated Vasculitis. He will take part in the cleaning, quality control and exploration of this data, preparing him for when he gets access to his own dataset. He will learn about the different methods for linkage and matching in clinical research and will get insight into the geographical

patterns of healthcare use in AAV across Scotland. This is in line with his HELICAL research objective and training.				
ESR6 Alejandro Fontal	ISGlobal	*Università degli Studi di Padova*	Lucia Delogu	01-11-2021 to 15-01-2022
The ESR will be responsible of discussing, organizing and performing the computational analysis of an experiment designed to evaluate the activation of immune biomarkers via the stimulation of Peripheral Blood Mononuclear Cells with collected and standardized Particulate Matter (as a proxy of urban pollution). The cell cultures will go through CyTOF and Luminex assays to be able to determine changes in biomarker expression. The objective is to pinpoint what biomarkers and associated immune routes might be activated by exposure to pollution.				
ESR12 Marco Zanet	TG	IBM Zurich	Antonio Foncubierta Rodriguez	01-12-2022 to 01-05-2022
Knowledge and experience acquisition on machine learning cutting edge technologies to be applied to the context histopathology and specifically to renal biopsies that are relevant within the scope of the project, i.e. to be applied to predict an outcome of AAV based on renal biopsy images. Such technologies include: - Machine learning techniques: semantic segmentation (e.g. U-net), GANs for anomaly detection, Graph Neural Networks for topological features extraction on WSIs, other relevant architectures. - Image processing techniques to be applied to whole slide images: tiles creation, tiles stitching, stain colour normalization, virtual staining etc. - Other: parallel computing techniques, object-oriented programming, best practices, containers etc.				
ESR1 Albert Navarro Gallinad	TCD	IS Global	Xavier Rodó	06-11-2021 to 28-02-2022
ESR1 (Albert Navarro) research is about developing a framework to facilitate the linkage between health and environmental data in a meaningful way using graph-based technologies. The framework proposed has the potential to be a solution for researchers who are trying to answer complex questions in environmental health research. The aim of the secondment is to validate the health-environmental linkage methods in terms of (1) adequacy for the health outcomes and (2) usefulness for Kawasaki disease environmental research in Japan.				
ESR3 Bahareh Khosravi	TCD	Rheinland-Pfälzische Technishe Universität	Sebastian Vollmer	25-09-2022 to 20-11-2022
Neutrophilic Cytoplasmic Autoantibody (ANCA) vasculitis is a type of autoimmune disease in which half of suffers have a relapse within five years. In this disease, the condition of illness repeatedly relapses over time, and the evidence demonstrates potential clinical triggers. The raw data provided				

by the previous research in TCD needed to be wrangled so that it could be modelled in R software. Therefore, we decided to implement models that follow patients through time and present a relapse risk measurement that evolves longitudinally according to the patient's clinical therapy regimes. Before modelling data, we will implement an algorithm to describe the vasculitis prodrome according to available clinical factors. Then, we perform a primary model to estimate the risk of the first relapse and consider the impact of explanatory variables on the risk in patients with definite ANCA vasculitis. After that, we will use the survival recurrent event analysis to estimate the risk of relapse and evaluate the impact of clinical variables on the risk over time. Eventually, the efficiency of the methods will be assessed by the goodness of fit models to determine which model works better to analyze the data.

Scientific collaborations taking place over years 2 and 3 of the researcher's PhD:

ESR2 Anna Weber	**IBM Zurich**	**University of Leeds**
The aim of the project is to identify T cell receptor (TCR) clones that are likely involved in giant cell arteritis (GCA). For this, we will work with TCR sequencing data of GCA patients, identify suspicious clones in their T cell repertoire and compare to control groups (healthy subjects as well as PNH patients). Additionally, the use of RNAseq data will be explored. We expect to identify GCA-specific T cell clones and will try to find indications of what target they bind. Ideally, we hope to predict likely autoantigens.		
ESR9 Filippo Guerri	**Anaxomics**	**IDIBAPS**
Filippo will get a better understanding of GCA and associated pathologies. He will select **protein data sets describing the current knowledge of GCA pathophysiology.** This work will be done in collaboration with Maria's lab as this description has to be point of reference from both sites. Selection of molecular data sets representing different pathophysiological pathways of GCA. Filippo will screen Pubmed publications, pathway repositories to collect potential descriptions. The definition of pathophysiological pathways as well as the revision of the proteins that will constitute the different sets will be reviewed by secondment lab. **Systems biology based analysis of differential immunological markers** streaming from ex-vivo stimulation of biopsies treated with tocilizumab and non-treated streaming from the ESRIO experiments. The molecular data sets defining GCA pathology will be used for the purpose. The data is expected by end 2021. The same type of analysis but using PBMCs may also be performed. We will evaluate the possibility that using the same strategy we can analyse other data sets from Maria's lab relating to other drugs. **Analysis of clinical data sets using Anaxomics pattern recognition pipeline.** We will evaluate the possibility of using Anaxomics strategy for pattern recognition in a longitudinal study on aortic dilatation property of IDIBAPs or request access to CDVAS data. We need to stablish if the biological question we want to pursue has not yet been analysed in those data sets or if the analyses Filippo can do bring anything new to the ones already done. In the event of CDVAS data request this will be done in collaboration between ESR 9 and 10, and they will seek advice on how to formulate the request to ESR15 from I-HD.		

ESR8 Michał Zulcinki	**University of Leeds**	**Edinburgh**
The main aims are to: (1) compute genotypic predictors (scores) based on GWAS summary statistics for relevant transcript traits and genotypic data (SNPs) using the GENOSCORES platform, (2) perform case-control association analysis using these computed scores and (3) perform association analysis for different phenotypes in GCA. The expected results will show associations between genetic traits and GCA and also the associations with some phenotypes in GCA (mainly inflammatory patterns of the arterial wall).		
ESR8 Michał Zulcinki	**University of Leeds**	**Newcastle**
The aim of the secondment is to perform a deconvolution analysis to integrate bulk and single-cell transcriptomic data generated from temporal artery biopsies. The main steps of the analysis will include: (1) infer sample-specific cell-type proportions and (2) perform cell-type-specific differential expression analysis to test for associations with distinct histological patterns of arterial inflammation in GCA. Expected results will reveal a landscape of cell population abundance levels in GCA biopsies and their associations with different inflammatory phenotypes. This work will also provide novel insights into cell-type-specific expression profiles of both, transcripts already known to be involved in GCA pathogenesis, as well as novel molecular signatures that might have potential for therapeutic targeting. Host supervisor will share single-cell RNA-seq data generated from temporal artery biopsies of 9 patients		

Researcher Careers post-HELICAL

These HELICAL fellows shared the roles and sectors they are moving on to after their MSCA Fellowships.

- ESR1 Albert Navarro Gallinad is joining the life sciences institute, Human Technopole in Milan, Italy as a health data scientist.
- ESR2 Anna Weber is going into science journalism and has a position at the NZZ (Neue Zürcher Zeitung) Media Group, Zurich, Switzerland.
- ESR3 Bahareh Khosravi is seeking to develop here career as a biostatistician and is seeking posts in Europe.
- ESR4 Enock Havyarimana will be joining the NIHR Centre for Environmental health and Sustainability based at the University of Leicester, England.
- ESR5 Solange Gonzalez Chiappe has a role as an epidemiologist at Max-Delbrück-Centrum für Molekulare Medizin, Germany.
- ESR7 Elkyn Estupinan Moreno will be working with the research group on epigenetics and immune disease at Josep Carreras International Leukaemia Foundation in Barcelona.
- ESR9 Michal Zulcinski will do postdoctoral research at the University of Leeds.
- ESR 10 Farah Kamberovic is interested in pursuing a career focused on clinical trials.
- ESR15 Maria Christofidou has begun a role as a Legal and Policy officer with the European Cybersecurity Competence Centre at the European Commission in Brussels.

Work package 6 Management and Recruitment

Supervisory board and Project Management team

The HELICAL ITN Supervisory Board was established in March 2019**,** to support the Coordinator and Project Manager in overseeing the research, training, and dissemination activities of the project. The Project Manager and Coordinator meet weekly. The management team elected to use Basecamp project management software as a coordination forum for the consortium.

The Supervisory Board is composed of representatives from all beneficiaries**,** one representative from each partner organisation, and one ESR representative. ESR2 Anna Weber, IBM Zurich, was elected by the ESR group to act as their representative from 2021 to 2023.

The committees listed below were also established. All committees were coordinated through Basecamp.

Research Coordination Committee

Comprising the WP leaders:
WP1: Mark Little, WP2: Ann Morgan, WP3: Peter Nilsson, WP4: Dipak Kalra and Nathan Lea, and Jessica Grene, project manager.

- Point of contact for WP leaders
- Reports to Supervisory board
- 6-monthly meeting (before SB meeting to allow feeding into that)
- Feed novel results into the Dissemination committee

Doctoral Studies Committee

Led by WP5 - MUW, Renate Kain
Includes one representative from each scientific WP
Monitors ESR selection procedures, gender balance and transparency of the national procedures, as well as monitoring training progress.

Dissemination, Exploitation and Communication Committee

Led by WP7 – MUW, Renate Kain
Includes one representative from each scientific WP
Input from academic beneficiary tech transfer offices
Responsible for establishing and monitoring the of website and social media, monitors publications and communication activities.

Information Governance Board (IGB)

i~HD coordinates the Information Governance Board, comprising Dipak Kalra, Nathan Lea, ESR15, Maria Christofidou, (iHD), Julie Power, Vasculitis Awareness Ireland.

Deliverables:

D1.1	Report describing vasculitis flare prodrome	11 Apr 2023
D1.2	Statistical Software Package and Code	16 Jan 2023
D1.3	Report on the role of environment in AAV onset	20 Jun 2023

D1.4	WP1 Summary Report	27 Jun 2023
D2.1	Report on genetic/epigenetic markers of GCA & related vasculitides	21 Mar 2023
D2.2	Report on identification of new genetic associations	23 Jun 2023
D2.3	Report on the analysis of the molecular mechanisms in the context of a human interaction network	16 Feb 2023
D2.4	Report summarising key pathways activated in tissue from patients with GCA	03 Mar 2023
D2.5	WP 2 Report Summary	27 June 2023
D3.1	Report on protein composition of EV	09 Dec 2021
D3.2	Report on suitability of EV as disease marker in AAV	16 May 2023
D3.3	Report on clinical evaluation of automated descriptor reading in tissue morphometry	29 Jun 2023
D3.4	Report on the analysis of the vasculitis associated autoantibody repertoire	01 Mar 2023
D3.5	Report on minimal biomarker set for diagnosis of active vasculitis	31 May 2023
D3.6	Report on development of multivariate assay for the diagnosis of different forms of vasculitis	25 June 2023
D3.7	WP 3 Summary report	26 June 2023
D4.1	Data Management Plan (DMP)	24 Jun 2019
D4.2	Helical IG policies	17 Dec 2020
D4.3	Governance framework for research in rare diseases	20 Dec 2021
D4.4	Validated / endorsed ethical and governance framework	03 Jan 2023
D5.1	Personal career development plan (PCDP)	20 Nov 2020
D5.2	Evaluation of secondment outcomes	16 Dec 2021
D6.1	Consortium Agreement	17 Apr 2019
D6.2	Network meetings 1 (1st 24 mths)	28 Dec 2020
D6.3	Network meetings 2 (2nd 24 mths)	24 Nov 2022
D6.4	Inception report	24 Jun 2019
D6.5	Supervisory Board of the Network	14 Mar 2019
D6.6	Progress Report	31 Jan 2020
D6.7	Open Data research Pilot	09 Jan 2023
D6.8	Final Report	29 Jun 2023
D6.9	15 ESRs recruited	16 Dec 2020
D6.10	Research Declarations	20 Oct 2020
D7.1	Plan for Exploitation and Dissemination of Results	23 Dec 2019
D7.2	Communication/Public Engagement 1	29 Mar 2021
D7.3	Communication/Public Engagement 2	28 Dec 2022
D7.4	Dissemination/Exploitation 1	29 Mar 2021
D7.5	Dissemination/Exploitation 2	09 Jan 2023
D8.1	H - Requirement No. 1	14 Apr 2021
D8.2	HCT - Requirement No. 2	23 Dec 2019
D8.3	POPD - Requirement No. 3	23 Dec 2019
D8.4	NEC - Requirement No. 4	23 Dec 2019
D8.5	GEN - Requirement No. 5	28 Dec 2020

Milestones

- All environmental and patient level data uplifted into RDF 09.12.2020
- Prototype software with spatiotemporal algorithms 15.11.2022
- Developed machine learning tools for morphological changes in renal biopsies 31.12.2022
- Developed multiplexed assay with panel of autoantigens 17.05.2022
- Secondments completed 17.10.2022
- Project Management System 25.03.2019
- ESRs recruited & appointed 12.05.2020
- Project check 14.12.2020
- Website 25.03.2019
- Final Conference 15.06.2023

Supervisory Board meetings:

In the second reporting period, Supervisory Board Meetings were held on 21.06.2021, 22.11.2021, 01.04.2022, 18.10.2022, 09.02.2023, and 14.06.2023.

The consortium applied for an amendment, requesting a no-cost extension of the action for six months to compensate for the research time lost because of the pandemic. The amendment was accepted, and the final date of the action was 30.06.2023.

Recruitment:

The recruitment process for Early-Stage Researchers was decentralised as described in Deliverable 6.9; each beneficiary managed their own recruitment process, with input from co-supervisors, and guidance from the HELICAL recruitment steering committee; (Mark Little (TCD), Renate Kain (MUW), Ann Morgan (LEEDS), Huseyin Firat (FIRALIS), Jenny Barrett (LEEDS). The process adhered to the principles set out in The Code of Conduct for the Recruitment of Researchers. The process was tracked centrally; shortlisting, panels, and appointment decisions were made using Basecamp project management system to ensure consistency and transparency in the recruitment process.

All vacancies were advertised on EURAXESS. The vacancies were shared with Scholars at Risk, and by Women in Science Ireland. Interview panels were made up from the consortium as far as possible and maintained a gender balance not exceeding 60/40. 8 female and 7 male ESRs were recruited.

Applications were added to Basecamp, the ITN's project management tool, for shortlisting by the interview panel. Interviews were held by video conference. Progress of the recruitment process for all HELICAL ESRs was tracked using a shared spreadsheet.

All 15 ESRs were recruited by 12.05.2020.

ESR Resignations and re-recruitment

ESR12 TissueGnostics

ESR12 Mihael Galinac with TissueGnostics resigned and left the project on 31.01.2021. The role was re-advertised, and Marco Zanet was recruited, beginning on 15.04.2021 for the remaining 19.03 months, until 15.11.2022.

ESR14 Firalis

Several resignations occurred in the Firalis beneficiary:

1. The first and second researchers recruited to the role of ESR14 with Firalis resigned after a short period; Rayan Mahmoud 05.02.2020 to 25.02.2020 (0.7 months) and
2. ESR14 Elisa Gómez de Lope 15.06.2020 to 17.07.2020 (1.1 months).
3. Samaneh Zareian was appointed as ESR14 on 25.08.2020 and finished on 15.05.2021.
4. Gisela Pattarone was recruited to the role on 01.10.2021 and resigned on 31.07.2022.

In consultation with the Project Officer, the consortium agreed that Trinity College Dublin would recruit the remaining months on the ESR14 role. To maximise the capacity to complete the research project, two ESRs were recruited simultaneously for the final months of the action. Matthieu Coq for seven months, 01.12.2022 to 30.06.2023, and Yağmur Doğay for 5.73 months, from 09.01.2023 to 30.06.2023.

Progress monitoring and evaluation of individual projects

Fellows on the action have a supervisory panel which includes primary supervisor, academic PhD supervisor (where different) and secondment host supervisors. These panels meet at least once a year, to address progress on the individual research project. All Fellows completed an initial Personal Career Development plan, which were submitted in Deliverable 5.1. These plans were updated by the ESRs and discussed with their supervisory panel.

Financial Management Strategy

Details of financial management were established in the Consortium Agreement at the start of the project, and are the responsibility of the coordinator, with project manager support. Funds were distributed to beneficiaries, with some funds retained centrally from management and RTN budgets to support project management and network-wide meetings respectively. An agreement was reached between the beneficiaries to weight RTN budgets in favour of “wet lab” projects, which have much higher consumable budgets.

Risk management at consortium level

The coordinator and supervisory board are responsible for all legislative and contractual issues concerning contract amendments, conflict resolution, non-compliance, and financial auditing. The project plan has been carefully designed to maximize information gained from results and to minimize potential risks, which have been evaluated and alternative approaches incorporated.

Risk	Description of Risk	Encountered?	Mitigation measures
5.1	Delay in recruitment	Yes	Advertised widely and worked with networks to encourage suitable candidates to apply; there is a 12 mth contingency built in
6.1	Conflict with background IP licensing	No	
6.2	Barriers to exploiting IP	No	
6.3	Researcher quits network	Yes	Re-advertised, support given to host institution on Article 32 of the Grant Agreement.
7.1	Ineffective dissemination	No	
1-4.1	Approach inadequate/ unachievable	No	
1-4.2	Failure of partner to perform or they leave the consortium	No	

Work package 7 Communication, dissemination, exploitation & outreach

Dissemination, Exploitation and Communication Committee

The Dissemination, Exploitation and Communication Committee (DECC) has been convened and held regular meetings on 27.5.2020, 24.7.2020 and 9.12.2020 and 04.11.2021. It is led by WP7 leader Renate Kain (MUW – Dissemination and Communication) and María Rodriguez Martinez (WP2, IBMZ - Exploitation), and includes Gertrude Krainz (MUW), Jessica Grene (TCD), Mark Little (WP1, TCD), Rupert Ecker (WP3, TG) and Nathan Lea (WP4). This committee is responsible for establishing, implementing and monitoring the plan for exploitation and dissemination of results, and for overseeing and monitoring the strategy for communications and public engagement activities. The DECC is responsible for preparing and submitting deliverables under WP7.

Dissemination of the research results

HELICAL outputs are designed to be disseminated to:

1. **healthcare professionals** caring for patients with vasculitis
2. **data science** and **information governance** communities,

3. potential **industry** users of HELICAL outputs,
4. **patients** with vasculitis and
5. the **public,** via:

- Publications

  Each ESR is expected to produce at least 1 paper (as first authors) in a top journal. WP1 ESRs will publish in leading machine learning conferences (NIPS, UAI, KDD[1]), WP2/3 ESRs will target biomedical journals such as Nature, J Am Soc Nephrol and Arth Rheum, and the WP4 ESR will publish in the leading bioethics journals (Bioethics, Europ J Epidem, J Med, Health Care and Philosophy). **Open access** principles are implemented by publications being made freely accessible on institutional repositories using budget assigned for open access. Results will also be targeted at a magazine with a wide readership for a more general audience such as *Research EU*.

  Below is a selected list of publications in addition to the publications recorded in the previous deliverable, deliverable *D7.4 Dissemination/Exploitation 1*. These publications were achieved from March 2021 by the ESRs and/or supervisors, all of which formally acknowledge the HELICAL project.

ESR	Title	Journal / conference	Metrics (2/12/22)
Albert Navarro Gallinad	Enhancing Rare Disease Research with Semantic Integration of Environmental and Health Data	https://dl.acm.org/doi/10.1145/3502223.3502226 in the 10th *International Joint Conference on Knowledge Graphs*: https://dl.acm.org/doi/proceedings/10.1145/3502223	3596 downloads 23 citations ACM: 30 downloads 1 self-citation
Albert Navarro Gallinad	Evaluating the usability of a semantic environmental health data framework: Approach and study	Semantic Web Journal https://content.iospress.com/articles/semantic-web/sw223212	
Albert Navarro Gallinad and Alejandro Fontal	Sub-weekly signatures relate ultrafine aerosols enriched in metals from intensive farming and urban pollution to Kawasaki disease	Environmental Research https://iopscience.iop.org/article/10.1088/1748-9326/acd798	244 Downloads
Anna Weber	TITAN: T-cell receptor specificity prediction with bimodal attention	Bioinformatics https://doi.org/10.1093/bioinformatics/btab294	5278 downloads 16 citations

[1] Neural Information Processing Systems, Uncertainty in Artificial Intelligence, Knowledge Discovery and Data Mining

	networks		
Anna Weber	DECODE: a computational pipeline to discover T cell receptor binding rules	Bioinformatics https://academic.oup.com/bioinformatics/article/38/Supplement_1/i246/6617535	1328 downloads 1 citation
Enock Havyarimana	The association between ambient UVB dose and ANCA-associated vasculitis relapse and onset	Arthritis Research & Therapy https://arthritis-research.biomedcentral.com/articles/10.1186/s13075-022-02834-6	1212 downloads 1 citation
Alejandro Fontal	COVID-19 Pandemic Sets New Clues on the Transmission Pathways in Kawasaki Disease	JAMA Network Open https://doi.org/10.1001/jamanetworkopen.2021.4624	3636 downloads 1 citation
Alejandro Fontal	Climatic signatures in the different COVID-19 pandemic waves across both hemispheres	Nature Computational Science https://www.nature.com/articles/s43588-021-00136-6	19000 downloads 26 citations
Elkyn Estupiñan Moreno	Methylome and transcriptome profiling of giant cell arteritis monocytes reveals novel pathways involved in disease pathogenesis and molecular response to glucocorticoids	Annals of the Rheumatic Diseases https://ard.bmj.com/content/81/9/1290	2297 downloads 1 citation
Michal Zulcinski	Somatostatin Receptor PET/MR Imaging of Inflammation in Patients with Large Vessel Vasculitis and Atherosclerosis	Journal of the American College of Cardiology https://www.sciencedirect.com/science/article/pii/S0735109722074411?via%3Dihub	2 Citations
Solange Gonzalez-Chiappe	Incidence of giant cell arteritis in six districts of Paris, France (2015–2017)	Rheumatology International https://link.springer.com/article/10.1007/s00296-022-05167-4	140 downloads 0 citations
Farah Kamberović	Blocking GM-CSF receptor α with	Annals of the Rheumatic Diseases https://ard.bmj.com/content/81/	3819 downloads 11 citations

	mavrilimumab reduces infiltrating cells, pro-inflammatory markers and neoangiogenesis in ex vivo cultured arteries from patients with giant cell arteritis	4/524.long	
Małgorzata S. Małys	Isolation of Small Extracellular Vesicles from Human Sera	International Journal of Molecular Sciences https://www.mdpi.com/1422-0067/22/9/4653	1918 downloads 5 citations
Małgorzata S. Małys	Small extracellular vesicles are released ex vivo from platelets into serum and from residual blood cells into stored plasma	Journal of Extracellular Biology https://onlinelibrary.wiley.com/doi/10.1002/jex2.88	519 Full text views
Shaghayegh Bayati	Persisting Salivary IgG Against SARS-CoV-2 at 9 Months After Mild COVID-19: A Complementary Approach to Population Surveys	Journal of Infectious Diseases https://academic.oup.com/jid/article/224/3/407/6274637	2687 downloads 29 citations
Maria Christofidou	A Literature Review on the GDPR, COVID-19 and the Ethical Considerations of Data Protection During a Time of Crisis	Yearbook of Medical Informatics https://www.thieme-connect.de/products/ejournals/html/10.1055/s-0041-1726512	8 citations

- Scientific Presentations

  ESRs and supervisors are encouraged to present and publish at workshops and local, regional, and international conferences targeting both industry and researchers. Each ESR is expected to present at a minimum of 2 conferences, focusing on the **20th International Vasculitis and ANCA Conference** in April 2022 (https://vasculitis2022.org/patient-conference/), which was hosted by the coordinator and is considered a key outreach vehicle. Several ESRs gave platform presentations at this conference: Albert Navarro Gallinad, ESR 1, and Enock Havyarimana, ESR 4, spoke about environmental triggers in vasculitis, Anna Weber, ESR 2, spoke about Giant Cell Arteritis, Maria Christofidou, ESR 15, spoke about data protection in vasculitis research, and different ESRs spoke about molecular influences in vasculitis (A link to the full programme: https://vasculitis2022.org/wp-content/uploads/2022/03/Vasculitis-2022-International-Patient-

Programme-1.pdf). Posters were also presented by various ESRs. Figure 32 is an example of a poster submitted by Michal Zulcinski, ESR 8.

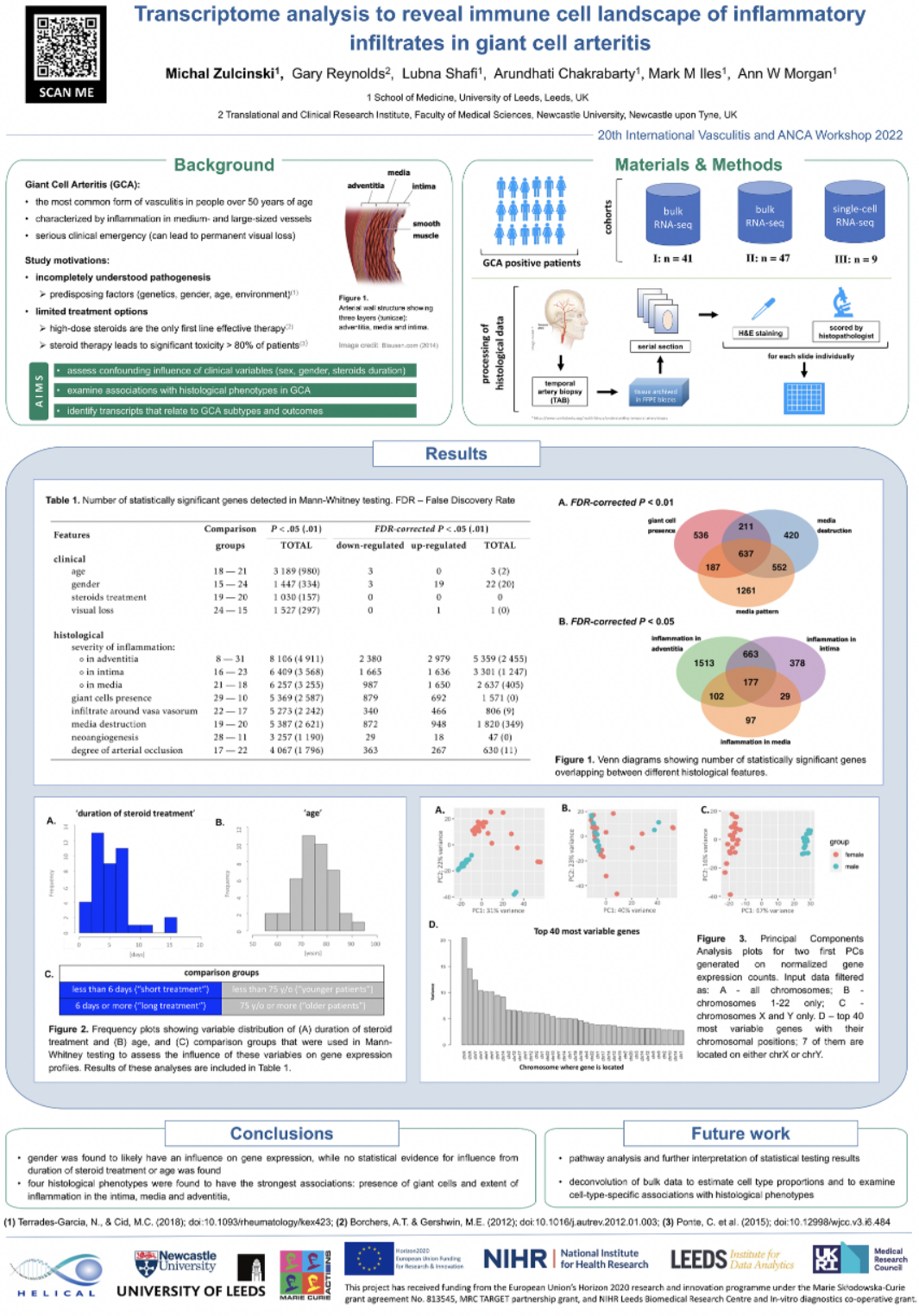


Figure 32. *Example of a poster, submitted by Michal Zulcinski, ESR 8.*

The ESRs presented at the HELICAL scientific symposium in September 2021 (Figure 33 shows a screenshot of this virtual scientific symposium), a symposium was also held in April 2022 where the ESRs presented (Figure 34 shows Maria Christofidou, ESR 15, presenting) and a meeting was held in Stockholm in February 2023 where the ESRs presented their work. There was an end meeting held in Barcelona in June 2023.

Figure 33. *Scientific symposium September 2021.*

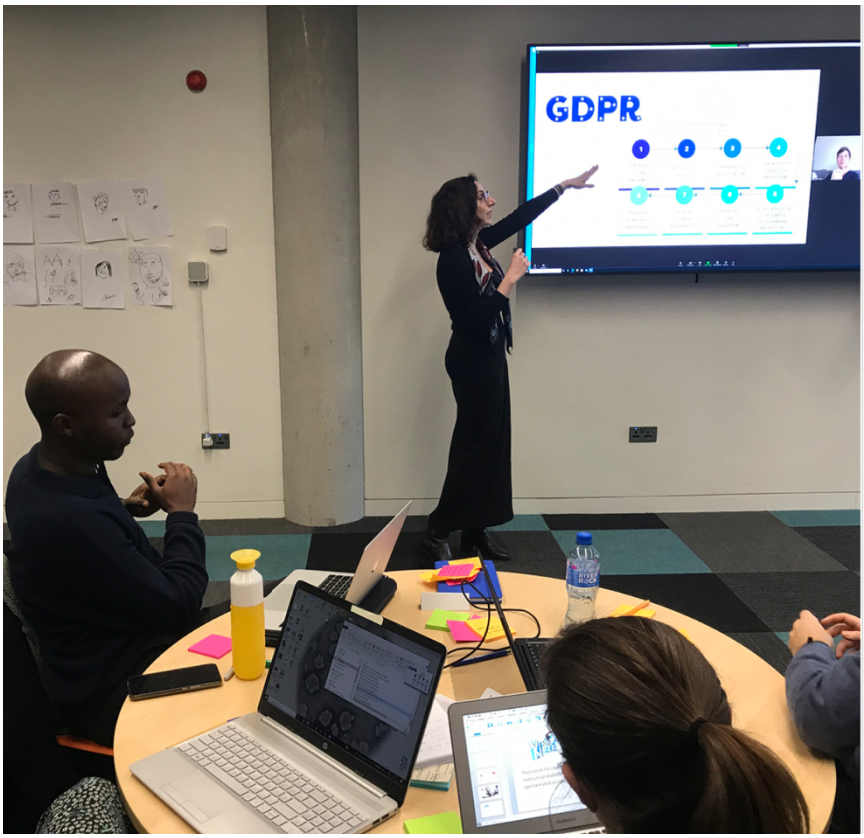


Figure 34. Maria Christofidou, ESR 15, presenting *at the scientific symposium April 2022.*

Numerous **poster presentations** were given by the ESRs:

- Albert Navarro Gallinad, ESR1, presented a poster on *Rare Diseases and the importance of being able to combine several sources to enable research on them* at 2021 Marie Curie Alumni (MCAA) Annual Conference (https://www.mariecuriealumni.eu/2021-conference-home).

- A poster presentation was given by Enock Havyarimana, ESR 4, at the UK Biobank Winter Scientific Conference 2021 on *The association of air pollution exposure and rare disease risk (https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/winter-scientific-conference-2022)* and at the 2021 Marie Curie Alumni (MCAA) Annual Conference on *The association of air pollution exposure and inflammatory rheumatic disease risk (https://www.mariecuriealumni.eu/2021-conference-home).* He also attended the *New England Statistical Society* with a presentation on environmental triggers of vasculitis.

- Solange Gonzalez Chiappe, ESR 5, presented a poster at the Autoimmunity Congress 2022 (https://app.frame.io/presentations/2973c4b0-2a81-44cd-bbf7-fffd3eef23cc).

- Michal Zulcinski, ESR 8, presented a poster at JOBIM 2021 (https://jobim2021.sciencesconf.org/).

- Filippo Guerri, ESR 9, contributed to PhD in Bioinformatics Scientific Workshop with a poster on *Systems Biology and Bioinformatics Approaches to Provide a Holistic Understanding of GCA biology*.

- Farah Kamberovic, ESR 10, did poster presentations in several conferences: American College of Rheumatology 2021 (https://acrabstracts.org/abstract/transcriptomic-changes-induced-by-mavrilimumab-versus-tocilizumab-in-ex-vivo-cultured-arteries-from-patients-with-giant-cell-arteritis/), American College of Rheumatology 2022 (https://acrabstracts.org/abstract/transcriptomic-changes-induced-by-tocilizumab-in-ex-vivo-pbmcs-from-patients-with-gca-in-remission-predictors-of-response/).

- A poster presentation was given by Małgorzata S. Małys, ESR 11, at different conferences: the international conference TRAIN-EV (http://train-ev.eu/train-ev-international-conference-2021-2/), the annual meeting organised by Young Scientist Association in Vienna (https://www.meduniwien.ac.at/web/ueber-uns/events/2021/ysa-symposium-juni-2021/), Small New World (ASEV and GSEV) 2022, and the Urinary EV Symposium 2022 by ISEV (https://www.isev.org/isev2022-annual-meeting).

The following ESRs gave **oral presentations** at conferences:

- Maria Christofidou, ESR 15, participated in several conferences such as: Future Digital Health (https://hopin.com/events/future-digital-health-event#speakers) and IoT week 2021 (https://iotweek.org/program-2021/). She also participated and presented her work at Last-JD-RIOE (https://rioe.oeg.fi.upm.es/conference/).

- Albert Navarro Gallinad, ESR1, did a paper presentation *Enhancing Rare Disease Research with Semantic Integration of Environmental and Health Data* at IJCKG 2021 (https://language-semantic.org/ijckg2021/program/).

- Anna Weber, ESR 2, presented a highlight talk *TITAN: T-cell receptor specificity prediction with bimodal attention networks* at ECCB Conference Barcelona 2022 (https://eccb2022.org/final-programme/). She also gave a talk on this topic at the virtual ISMB/ECCB 2021 (https://www.youtube.com/watch?v=S35isRyoSbg).

- Elkyn Estupiñan Moreno, ESR 7, gave a presentation at II Congreso Investigación PTS Granada on *Los perfiles de metilación y expresión génica de monocitos revelan nuevas vías implicadas en la patogénesis de la arteritis de células* (https://investiga.granadaessalud.es/wp-content/uploads/2022/01/orales-precision.pdf).

- Małgorzata S. Małys, ESR 11, gave an oral presentation in Lyon at the conference of International Society of Extracellular Vesicles, ISEV 2022 (https://www.isev.org/assets/PastISEVAnnualMeetings/ISEV2022%20Final%20Schedule%20at%20a%20Glance%205.26.2022.pdf).

Open data

This has been considered as part of the data management plan (DMP, WP4 deliverable); suitable repositories are identified for the outputs and all generated research data will adhere to FAIR data

principles. Data management costs are eligible for reimbursement during the duration of the project and can be claimed under the conditions defined in the grant agreement. ESRs have received training on OA/Open Science. Deliverable 6.7 describes consortium specific mechanisms for depositing in open repositories.

In addition to the open access repositories mentioned in *D7.4 Dissemination/Exploitation 1*, a **HELICAL Zenodo Community** was created on 24 August 2022 to facilitate access to ESRs´ research outputs from this central repository: https://zenodo.org/communities/itn_helical/.

Website
The HELICAL website (http://helical-itn.eu/) advertised the PhD positions, provides training programme details, and was set up to serve as a key communication tool within the network, as well as to the wider public, scientific, and medical communities. The running of the website is discussed in deliverable *D7.3 Communication/Public engagement 2*

Scientific Reports (Deliverables of WPs)
These will be openly available by way of the website after clearance from the Exploitation team.

Theses
The theses of the ESRs will be disseminated via Open Access Theses and Dissertations (https://oatd.org/).

Exploitation of results and intellectual property (IP)

The procedures outlined in the previous deliverable D7.4 and the monitoring of the potential for exploitable results are managed strategically under the jurisdiction of the DECC. To ensure maximal protection of IP and possible routes to commercialisation, ESRs received IP training provided by IBM Zurich, *The Innovation Path*, in Module 6 (June 2021). Further training, *Innovation in Research,* was provided in Module 8 by Tangent (March 2022).

An approach has been undertaken where everything that is being published is done so as open source. Two specific cases were considered by the DECC:

1. **TITAN** (Tcr epITope bimodal Attention Networks, designed as part of Anna Weber, ESR's 2 research). No exploitable foreground IP artefacts were arising.
2. **SERDIF** (Semantic Environmental and Rare disease Data Integration Framework, designed as part of Albert Navarro Gallinad, ESR's 1 research) an approach has been taken where all related components are based on W3C open standards, and the artefacts have also been made available in an open manner to the community. No exploitable foreground IP artefacts were arising.

Work package 8 Ethics requirements

The following deliverables were submitted as part of work package 8:

D8.1	H - Requirement No. 1	14 Apr 2021
D8.2	HCT - Requirement No. 2	23 Dec 2019

D8.3	POPD - Requirement No. 3	23 Dec 2019
D8.4	NEC - Requirement No. 4	23 Dec 2019
D8.5	GEN - Requirement No. 5	28 Dec 2020

Introduction

Each ESR project has received appropriate ethical approval from competent ethics committees. It was never the intention of the consortium to replace the function of these ethics committees with a new "independent ethics board"; reference to such an entity in the Description of Action was an error and intended to refer to an "Information Governance Board". This board was established in the first six months of the action and has overseen development of a HELICAL Data Management Plan, Information Governance Policies and Information Governance Framework.
The mid-term review mandated the creation, in addition to these actions, of an Independent ethics board, comprising members with expertise in the ethical issues related to this project.

Convening of an independent ethics board (IEB)

This process was led by IHD beneficiary, with support from the coordinator and central management team. ESR15 (Maria Christofidou) project managed the activity. The IEB comprised two experts in the field of medical ethics, specifically in the ethical challenges of re-use of primary healthcare data and sharing of that data across borders:

1. **Petra Wilson**, a senior advisor in Healthcare Information and Management Systems Society (HIMSS) on EU Policy, with a particular focus on data use and re-use in the context of GDPR and other EU level legislation. She is a public health lawyer with over 25 years' experience in EU level digital health policy and practice. Alongside her work at HIMSS she also acts as consultant advisor to many leading pharmaceutical companies through FTI Consulting in Brussels. She is a member of the WHO Digital Health Technical Advisory Group, Board Member of the European Association for the Study of Diabetes and sits on committees for European Patients' Forum and the European Rare Diseases Organization (EURORDIS). She has previously engaged in key EU level digital health projects such as EPSOS, ANTILOPE, CALLIOPE, METHOTELEMED, LEGALLY EHEALTH.
2. **Zoi Kolitsi** has a long record of eHealth policy support, including a decade in the Greek Public Health Administration as a senior policy advisor. Over the last 15 years, she has led work in national and EU eHealth projects on eHealth interoperability, legal and regulatory, eHealth governance, business planning and sustainability domains. Particular contribution has been to cross-border eHealth and cross sectorial aspects of electronic identification and authentication, while she maintains an active involvement in eGovernment activities.

IEB activity

Petra and Zoi were briefed on the HELICAL concept and consortium, and provided with ethics deliverables (D4.2, D4.3, D4.4, D8.1 and mid-term report ethics document), along with the description of action. After reviewing these documents, a meeting was convened in July 2022 to clarify matters and answer specific questions posed by the IEB. The discussion ranged between the impact of the European Health Data Space and Data Governance Act on HELICAL procedures (theoretical, as HELICAL will be effectively completed by the time these are implemented), data altruism and dynamic consent. With respect to the latter, the HELICAL team acknowledged that it would have been ideal to include the HELICAL

research in a dynamic consent framework, but all parties agreed that it would not have been feasible.

The IEB concluded that there were no ethical concerns with the HELICAL project, and that further meetings were not required.