PRECISESADS

Molecular Reclassification to Find Clinically Useful Biomarkers for Systemic Autoimmune Diseases

IMI Periodic Report Template

“Molecular reclassification to find clinically useful biomarkers for systemic autoimmune diseases”

PRECISESADS

115565

Prof. Chris Chamberlain

UCB Biopharma SPRL

Allée de la Recherche, 60 1070 Brussels, Belgium

Period 01/2016 - 12/2016

Reporting Period 3

Description of work 4 September 2016/v4.0*

Date for submission 28-02-2017

Revised submission: v2 10 July 2017

efpia

IMI/INT/2013-01039 version 3

PRECISESADS

Molecular Reclassification to Find Clinically Useful Biomarkers for Systemic Autoimmune Diseases

1. Executive summary

1.1 Project rationale and overall objectives of the project

Inflammatory autoimmune diseases such as rheumatoid arthritis and lupus affect 1-3% of the population, and whilst treatments exist, these are costly and have a number of serious side effects. There is growing evidence that many of these conditions may be incorrectly classified. The PRECISESADS project will study over 2000 people with various autoimmune diseases, gathering data on the molecular causes of their disease as well as their clinical symptoms and making comparison to 600 healthy controls. Through the analysis of these patients' data, the project aims to define clusters of individuals who share similar molecular pathways for their disease and so could be treated in a targeted and personalised way. By evaluating the molecular and clinical data using the latest technology, the project will deliver new biomarkers for use in more targeted clinical trials. Clinicians can then tailor therapies according to the specific molecular pathways found in individual cases. In short, treatments will become more personalised.

1.2 Overall deliverables of the project

The PRECISESADS consortium is a translational network of leading clinical, translational and basic researchers and practicing physicians in the fields of genetics, metabolomics, flow and mass spectrometry, rheumatology, and immunology, together with five major pharmaceutical companies, with the common aim to provide a practical and clinically relevant answer to this challenge. Our main hypothesis is that the identification of specific molecular signatures in patients with SADs will enable clinicians to tailor therapies according to the specific aetiological pathways at the root of pathology in each individual case. In short, to implement precision medicine strategies.

We will make use of the available high-throughput 'omics technologies and employing the clinical knowledge that exists for these diseases will perform an integrated bioinformatic analysis of the data to identify biomarkers that can provide clinically relevant information. To accomplish this, we aim to reclassify the individuals affected by SADs into clusters of molecular, instead of clinical entities, going beyond the classical clinical diagnoses. At the same time, we will address several sub aims that intend to answer relevant questions relating to pathology within more defined "tissue" entities, such as kidney, or skin. We believe such a tissue-based taxonomy is relevant because it touches on a second important fact common to these diseases: organ damage, prognosis and response to therapy. These separate taxonomies: a systemic taxonomy based on peripheral blood/serum/urine/plasma/cellular markers, and a parallel tissue-based taxonomy, based on specific tissue markers will complement each other.

The systemic taxonomy will be analysed first in a discovery cross-sectional study of 2000 individuals. These individuals will be studied only once, and their data will include clinical details. In addition, a total of 100 biopsies of individuals with kidney disease and scleroderma will be thoroughly analysed for several molecular parameters.

In parallel, we will initiate the recruitment of individuals with the same diseases by collecting an inception, longitudinal cohort for assessment of the taxonomies discovered in the cross-sectional study. This second study will have three aims: i) to study individuals extensively at baseline with little or no treatment, and to follow the individuals to observe their development towards clustering in the absence of long-term treatments, ii) to observe if these conform to the same clusters we observe in the cross-sectional set, iii) to compare their individual 'omics patterns with the patterns obtained in the initial discovery study. These individuals will be studied three times during the course of the longitudinal study and will be analyzed fully, using the markers, assays and systems we identify as most reliable, clinically adaptable and informative. The analysis of their samples will be done in parallel to the cross-sectional cohort study.

Finally, and with the aim of determining their translational value for the new molecular classification and to obtain a similar re-classification, a more limited study will include the molecular analysis of blood/spleen and tissues obtained at different time points from mouse models of the SADs, those often used in pre-clinical studies of new drugs.
The models that are planned at the moment are:
(NZBxNZW)F1 (Charite/Bayer)
MRL/lpr (Charite/Bayer)
B6.Sle123 and B6 controls (UCL)
B6.TLR7tg (Genyo)
The B6.Sle123 is a model that has disease-causing loci from the (NZBxNZW)F1 model on the B6 background. There is no control model for MRL/lpr mouse planned.

1.3 Summary of progress versus plan since last period

WP1 - Project Management and Coordination of the Consortium

- Second General Assembly took place in March, 2016 in Granada in the presence of both Advisory Boards
- Co-organized with FPS and WP9, the Genomics Conference in Granada in March 2016
- Steering Committee met three times face to face and monthly throughout the year to follow overall consortium progress, to ensure that recruitment goals are met and to discuss data analysis strategies
- Successfully organized and achieved the 24-month Interim Review with minimum comments
- Met with the Taxonomy partner, AETIONOMY, to continue the established collaboration.
- Organized the clinicians meeting at ACR to ensure recruitment goals and coordinate data cleaning activities.
- Performed site quality visits at 5 clinical sites

WP2 - Sample collection and bio banking

- Ongoing patient recruitment of the PRECISESADS Cross Sectional phase 2 study. Recruitment target achieved for Cross Sectional phase 1 participants (302) and Inception (200).
- A total of 3088 Biobank recruitment kits were prepared and sent to all clinical partners
- 67,501 sample tubes from 2126 individuals have been received and registered by the Biobank.
- 8866 samples from 1167 individuals were sent to analysing sites. 1495 DNA and 1315 total RNA samples have been extracted.
- Successful posting of both studies on clinicaltrials.gov website.

WP3 - Clinical data management

- Recruitment in the Cross Sectional study is at 73% of the total objective of 2000 patients and 666 controls. It is foreseen to end by 30 June 2017 in line with expectations.
- The Cross-sectional Phase 1 part has been completed with 302 patients included (247 patients and 55 controls). All corresponding eCRF data are frozen.
- The Inception Cohort has reached the targeted milestone of 200 patients recruited. Specific attention is drawn to ensure patient retention across the 14 months' study duration and recruitment of 10% additional patients is on-going to replace withdrawn patients.
- Data cleaning is controlled on a regular basis, data quality reports are issued monthly and data transfers are performed monthly. All efforts are put to anticipate data cleaning in view of database lock planned dates.
- For both the Cross-Sectional Phase II and the Inception cohort, site-specific monitoring of data entry backlog is organized and an action plan established to maximize the clinical data entry in a timely fashion and have all data available for the analyses.

WP4 – Genomics, transcriptomics, epigenomics

- 764 samples of systemic autoimmune diseases and controls were successfully genotyped by the Illumina Human Core Chip 24 v1 with 306,670 SNPs. Imputation (impute2) leads to the inclusion of data for a total of 2,259,093 high confidence SNPs. The distribution of the analyzed samples is as follows: 164 controls, 178 SSc, 105 RA, 120 SLE, 97 SjS, 47 UCTD, 25 PAPs and 28 MCTD.
- Complete GWAS genotyping of the Phase 1 portion
- We have performed HLA imputation and analysis of the above mentioned samples using SNP2HLA.
- The analysis of meta-GWAS of rheumatoid arthritis versus systemic lupus erytematosus, and rheumatoid arthritis versus systemic sclerosis is done. We identified IRF4 and COG6 as novel common genetic factors between SADs. Two published papers have been reported.

Transcriptomics:

Whole blood gene expression profiling

- Transition to Illumina HiSeq2500 platform, since BeadArray technology discontinued Establishment of new workflow (96 samples / Pool, 4 Pools / Batch, each Batch on 3 HiSeq runs)
- All of phase I + 83 samples from phase II analyzed, data currently undergoing QC check 11 samples excluded from HiSeq runs due to low quality or low RNA yield

Cell types

- RNA-Seq: The four isolated cell populations in WP5 from 20 patients (4 for each main group: SLE, RA, SSc, SjS and Controls) have been selected and sequenced at 20 million paired-end reads on average (expected to sequence up to 40 million on average). Except for 5 samples (which libraries must be repeated), the other 75 have a unique alignment success of around 85%. Libraries for the same 80 samples are being prepared and will be ready to sequence soon.
- miRNA-Seq: Libraries for the same 40 samples are being prepared and will be ready to sequence soon.

Exosomes

- RNA-Seq test has been performed with exosomes from control samples (n=4). Despite the low amount of RNA in the samples (0,33-0,44 ng/ul), after increasing the number of PCR cycles and checking that we did not generate a GC compositional bias, we were able to generate and sequence the libraries.
- The most expressed genes in the samples were enriched in Gene Ontology terms related with exosome function and almost all of them have been described in Vesiclepedia as being expressed in exosomes.

Epigenomics:

- Generation of DNA methylation data (using the 450K platform) for an additional number of 192 individuals (including SLE, RA, SSc, SjS, PAP, UCTD, MCTD and controls) from cross-sectional 1 and 2 cohorts.
- Pre-processing and normalization and QC filtering of the DNA methylation data.
- Cluster analysis of the data using an enthropy based approach to stratify patients according to their molecular profile

- Bisulfite modification and preparation of remaining whole blood samples (a total of 252) for completing the hybridization of the cross-sectional Phase 1 study. These samples are being hybridized in the novel EPIC platform

WP5 – Flow cytometry and cellular separation

- An intercalibration verification procedure has been done every 4 months with the VersaComp capture beads kit and 2 times a year with the same lot of Immuno-Trol cells. Both confirm the mirroring of the 11 cytometers.
- Recruitment of patients is ongoing and no issues have been reported due to flow cytometry problems.
- All the files from phase I cross-sectional study have been normalized and manually analysed by the same operator and downloaded to WP8 for clustering. The analyses from phase II and from inception cohorts are ongoing.
- Flow cytometry analysis for pre-clinical models of SADs is ongoing.

WP6 – Proteomics, Metabolomics and Serology of SADS

- The detection of 100 cytokines was performed using the LUMINEX platform. Of these, 87 mediators were possible to measure in cross sectional phase 1 subjects. Analyses are ongoing within WP8 with the aim to use in the formation of clusters and to select those cytokines that will be measured for phase 2 and inception. Some previously undescribed cytokines were found increased across nearly all diseases including the undifferentiated cases and consideration is been given to their potential publication.
- For metabolomics analysis, both LC-MS and NMR methodology was used to measure plasma and urine samples from the cross-sectional phase 1 study.
- Autoantibodies, all in all 27, established in the field of autoimmunity and also natural antibodies against lipid epitopes were determined in the cross-sectional cohort. Analyses are under way with the data all uploaded in the centralized server and in WP8.

WP7 – Tissue taxonomy and Imaging Analysis

- We finalized our high-throughput transcriptomic studies on SLE and control kidney biopsies. We performed immunohistochemistry and ELISA confirmatory experiments on the same, and on independent cohorts of samples. Our results demonstrate that tubular infiltration by adaptive immune effectors is a major prognostic factor in SLE nephritis but needs to be independently validated. These results have a major impact on the (tissue) taxonomy of SLE nephritis, which is presently based on the description of the glomerular distribution of the inflammatory process, yet needs to consider tubular involvement as well.

- We performed proteomic studies on kidneys from SLE-prone and control mice, harvested at different stages of the disease. Analyses, in particular correlations with histological scores, are ongoing.
- High-throughput transcriptomic studies on systemic sclerosis skin biopsies were delayed for ethical reasons (see below). Preliminary proteomic studies were performed on 2 sets of 5 affected versus normal skin samples from the same patients.

WP8 – Data analysis, bioinformatics and biostatistics

- The Data Governance Office is up-and-running (currently 50 active accounts).
- The Data Governance Plan (DGP) was upgraded following the Legal & Ethics joint workshop with AETIONOMY.
- Clinical eCRF data transfers occurred monthly and patients' data, including first longitudinal data from the Inception cohort, were seamlessly uploaded to tranSMART and made available to all partners.
- OMIC data were transferred from WPs 4-5-6 and integrated in the OMIC server:

WP	Data type	Cross-sectional		Inception			TOTAL
		Phase 1	Phase 2	M000	M006	M014	
WP3	eCRF	302	1684	133	90	27	2236
WP4	Genotyping	62	463	13	*na*	*na*	538
WP4	HLA (imputed)	62	463	13	*na*	*na*	538
WP4	Gene expression**	0	0	0	0	0	0
WP4	Methylation	50	300	11	0	0	375
WP5	Flow cytometry (panels 1 & 2)	298	19	0	0	0	317
WP5	Flow cytometry (panels 3 to 9)	298	1	0	0	0	299
WP6	Auto-antibodies	300	237	5	0	0	542
WP6	Small lipid moiety antibodies	300	237	5	0	0	542
WP6	Metabolomics	296	4	0	0	0	301
WP6	Luminex	266	7	1	0	0	274

**Please see comment on page 17 that addresses this.

- A Clustering Guidelines document was issued and circulated to all WP8 members, who started to work in small teams for each of the specific OMIC data sets. Current major achievements include:
  - Genetics: The previously developed *SNPclust* algorithm was published in *PLoS ONE*. It allowed us to deliver D8.3 Prototype clustering tools and metrics based on preliminary data. Interesting clustering results were obtained with available HLA data (abstract in

preparation), to be confirmed when complete Cross-Sectional Phase 1 DNA data are available;

- Methylation data: A first set of 350 individuals with methylation data using the 450k CpG array was analysed. Using all CpGs did not allow for any clustering. A potential reason for this is that the material comes from whole blood. A different strategy was subsequently adopted. Using two separate batches, with appropriate adjustments for batch and treatment, we obtained 4 preliminary clusters of patients. We are awaiting new methylation array data (EPIC array, as Illumina has discontinued the 450k array) to validate the clusters. The methodology is completely new and we believe it will have great impact.
- Flow cytometry: A novel open-source software (CyMER) was developed to help analyse the flow cytometry data for cluster analysis, which can also be used for analysis of other data types. This software is now being used to analyse all the phase 1 data for cluster analysis (manuscript in preparation, and abstract submitted to the European Workshop in Rheumatology Research and the EULAR). We leveraged the power of machine learning to data-mine a comprehensive list of SAD-pertinent cell populations. Our results verified well-documented relationships between specific cell populations and SADs and also revealed new relationships.
- Antibodies: Latent class analysis classified subjects into five clusters, two of them being enriched in SjS and MCTD, one enriched in Controls and SSc, and another enriched in PAPs and SLE.
- Metabolomics: Exploratory analyses of the metabolomic datasets identified several specific metabolites correlated with individuals taking anti-malarial medications; initial clustering of the metabolomic datasets identified several clusters that are currently being investigated.
- Luminex: Cross-sectional phase 1 data were analysed (association with disease and treatments, and clustering) in order to select cytokines to be measured in phase 2.

In parallel, clustering of longitudinal data was benchmarked to prepare for the Inception data analysis. Several longitudinal data clustering strategies were adopted, including Consensus clustering, NonNegative Matrix Factorization, Similarity Network Fusion and WGCNA. Also, some indices comparing clusters were studied. All these are currently being tested on a publicly available dataset, before implementing them for the PRECISESADS longitudinal data.

The LILLY microarray dataset was normalized and processed in conjunction with the clinical assessment of the patients, to *(i)* identify genes whose expression correlates with disease severity, and following that identify signalling pathways/biological processes involving these genes, uncovering mechanisms that potentially drive the disease phenotype; and *(ii)* segment patient populations to identify patient subgroups with similar gene expression profile.

Finally, secondary bioinformatics analyses were initiated:

- Quantitative Trait (QT) analyses: An initial overview of the phase 1 data (300 individuals) was performed. The data were first analysed for differences between patients and controls followed by correlation analyses between the various data types (*e.g.* Luminex *vs.* autoantibodies, metabolites *vs.* Luminex, metabolites *vs.* autoantibodies). This will provide a framework for selection of the appropriate Luminex analytes and metabolite peaks for the phase 2 analyses. Some of the novel data is currently being prepared for publication (manuscript in preparation).
- Epistasis: An *R* package, *gpatscan*, was developed to detect epistatic interactions between pairs of SNPs using predefined sets of genetic interaction patterns. Performance was benchmarked on previously available genetic data (article in preparation).

WP9 – Knowledge dissemination and training

- Organisation of the PRECISESADS European Conference "Genomics of Complex Diseases: New Challenges" in Granada, Spain.
- The conference gave an overview of the genetic, epigenetic, regulatory and functional aspects of genomics studies as well as the new challenges posed by next generation sequencing and bioinformatics in the understanding of complex diseases, focused on systemic autoimmune diseases.
- Organisation of the Taxonomy Workshop in Barcelona with the participation of 32 researchers from both PRECISESADS and AETIONOMY
- Publications from the Consortium as well as dissemination activities are detailed in Section 3.

1.4 Significant achievements since last report

WP1 - Project Management and Coordination of the Consortium

WP1 has continued to coordinate all WP activities to schedule within the Consortium and actively sits in on each WP. The Steering Committee met regularly. The Annex 1, Description of Work was amended to reflect updates in certain work packages. The GA was held in March in the presence of the advisory boards and was held just before the successful PRECISESADS Genomics Conference.

The 3rd GA has been scheduled for March 2017, using the same format with the Advisory Board members. This WP also coordinated and participated in the joint Workshop with Taxonomy Partner AETIONOMY. The PO successfully choreographed all rehearsals and responses for the Interim Review that was held at month 24.

WP2 - Sample collection and bio banking

The logistics and coordination for the transportation and storage of the patient samples has been effectively carried out. The biobank has also been organizing the sending of 8866 samples from 1167 individuals to the analysis sites. Successful posting of both studies on clinicaltrials.gov website has led to an increase in recruitment due to external contacts:

- PRECISESADS INCP (RB 15.007) - NCT02890134 (Inception)
- PRECISESADS CS (RB 14.106) - NCT02890121 (Cross-sectional)
- PRECISESADS-T - NCT02890147 (controls)

WP3 - Clinical data management

The cleaning of Phase 1 study (302 patients) is complete and all corresponding eCRF data is now frozen.

In addition, as described in the revised Description of Work (Version 4), WP3 has achieved the development of a Study Management Guide which aims at describing in one single document compliance/adherence of both Cross-Sectional study and Inception Cohort to Good Clinical Practices (GCPs) and Regulatory obligations, all study procedures established to ensure standardization of patients and samples management as well as the monitoring plan applicable to both studies and managed by SERVIER (26) and Project Office jointly. This document is a major achievement to ensure the quality of all data collected within the PRECISESADS studies. This Study Management Guide was sent to IMI on December 15th 2016 as Deliverable 3.6.

WP4 – Genomics, transcriptomics, epigenomics

Novel shared genes in SADs were described. Cross-sectional phase 1 study GWAS data completed. Methylation data ongoing, complete phase 1 study shortly. Whole blood gene expression data phase 1 ongoing. The cell types RNA-Seq analysis is ongoing. The pipelines for microRNA-Seq and exosomes are set

WP5 – Flow cytometry and cellular separation

An intercalibration verification procedure has been done every 4 months with the VersaComp capture beads kit (in 2016: January 14th, April 14th, September 6th, and January 9th, 2017) and 2 time a year with the same lot of Immuno-Trol cells (January 14th and December 13th, 2016). Both confirm the mirroring of the 11 cytometers. All the files from phase I cross-sectional cohort have been normalized and manually analysed by the same operator and downloaded to WP8 for clustering analyses.

Regarding pre-clinical animal models of SADs, flow cytometry procedures have been validated for cells obtained from peripheral blood and spleen in NZBxNZWF1 and MRL-lpr mice.

Advances have been made in the establishment of a protocol for complete immunophenotyping of circulating populations using mass cytometry in whole blood samples. An extensive panel comprising more than 30 parameters is being designed and optimized.

WP6 – Proteomics, Metabolomics and Serology of SADS

Measures in all the fields of study have been successful, and some interesting data have begun to emerge. Further analyses will yield information about the role of these measures in SAD taxonomy.

WP7 – Tissue taxonomy and Imaging Analysis

Our findings suggest that lupus nephritis is not just a glomerulonephritis, but a pan-nephritis. Tubular infiltration by adaptive immune effectors plays a determinant role in disease severity and evolution.

WP8 – Data analysis, bioinformatics and biostatistics

A significant volume of eCRF and OMIC data are now shared on the OMIC server, enabling clustering analyses (by WP8) and multiple comparative analyses (by all partners). These activities are now generating a significant volume of new scientific results based on data from the PRECISESADS patients.

WP9 – Knowledge dissemination and training

Several dissemination activities have been organised throughout year 3: the PRECISESADS European Conference of Complex Diseases, interviews to national media (TV and newspapers), presentation of project results either in international congresses or peer-reviewed scientific papers.