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 01/2015 - 12/2015

Reporting Period 2

Description of work 04 Sept -2015/v3.0

Date for submission 29-02-2015

Version 1.0 _ 2016-02-29

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 while 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 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, mass spectrometry, rheumatology, and immunology, together with five major pharmaceutical companies, with the common aim to give this challenge a practical and clinically relevant answer. 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 pathways to be targeted in individual cases. 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 cohort 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. These will be the MRL-lpr model, a model for spontaneous RA and lupus nephritis and the B6.lpr, thus on two genetic backgrounds; the (NZBxNZW)F1 and B6.Sle123 models, both models of lupus disease on two genetic backgrounds, and the tight skin mouse for SSc on the B6 background (B6.tsk1) complemented by a Bleomycin-induced model.

1.3 Summary of progress versus plan since last period

WP1 - Project Management and Coordination of the Consortium

- First General Assembly took place in Feb 2015, Berlin in presence of both Advisory Boards
- Steering Committee met twice face to face and monthly throughout the year to follow overall consortium progress and implement changes to the Cohorts to ensure recruitment goals are met
- New partner P/29 Charite accepted into Consortium replacing P/24 DRFZ which necessitated amending the Grant Agreement and Project Agreement
- Established connections with AMP RA/Lupus Network Project (https://amp-ralupus.stanford.edu/about/ra-lupus-amp-project) in the US with the idea to sign a MoU.
- Performed site quality visits (2 at the Biobank and 2 recruiting sites)

WP2 - Sample collection and biobanking

- PRECISESADS, recruitment was initiated for the 3 cohorts: Cross sectional Phase 1, Cross sectional Phase 2 and Inception.
- Almost 2000 Biobank recruitment kits were prepared and sent to all clinical partners:
- 25,530 sample tubes from 1123 individuals have been received and registered by the Biobank.

- 1155 samples were sent to analyzing sites. 480 DNA and 463 mRNA samples have been extracted.

WP3 - Clinical data management

- Cross sectional and Inception Cohorts, are recruiting and patient clinical data are entered into the corresponding eCRF databases in an on-going manner.
- Data cleaning on a daily basis and data transfers are performed monthly.
- No technical issues were reported and with the support of the SERVIER data-management group, each site is now fully operational.
- Data quality reports have been established to help in supervising the study progress.

WP4 – Genomics, transcriptomics, epigenomics

- Genomics: 187 samples of SADs and controls were genotyped by the HumanCore Illumina array. The analysis of the GWAS data is ongoing
    - The analysis of meta-GWAS of rheumatoid arthritis versus systemic lupus erythematosus, and rheumatoid arthritis versus systemic sclerosis is final. Common genetic factors between SADs were identified. Two manuscripts were sent for publication.
    - HLA analysis: meta-analysis and imputation of HLA data for rheumatoid arthritis. Manuscript sent for publication.
- Transcriptomics (NGSequencing): Selection of material and patients for RNA-Seq and miRNA-Seq for 4 cell types (B-cells, T-cells, monocytes and neutrophils) and total blood from the same 50 samples for gene expression deconvolution trial. Preparation of the analysis pipeline of the data for RNA-Seq.
    - Tested and optimized a RNA extraction method for separated cell populations and PAXgene tubes, which will be automated in Genyo (FPS).
- Epigenomics: Tested and optimized conditions for bisulphite modification of small amounts of DNA (for DNA methylation analysis), given the large range of material obtained from large cohorts of patients
    - Tested and optimized conditions for chromatin immunoprecipitation assays (for histone modifications) for small amounts of material, given the limited amount of material obtained in patient's cohorts
    - Generation of DNA methylation data for 187 individuals (including SLE, RA, SSc, SjS, PAP, UCTD, MCTD and controls) from cross-sectional 1 and 2 cohorts.
- RNA-SEQ: The new Illumina sequencer (NextSeq 500) at Genyo (FPS) has been set up. Waiting for first Illumina sequencing kits delivery to arrive, in order to start sequencing the first samples from the PRECISESADS project.
    - RNA-Seq and miRNA-Seq pipelines finished and ready to start processing the PRECISESADS samples. Now, a pipeline to discovery and analyse new lncRNAs is being set up

WP5 – Flow cytometry and cellular separation

- An intercalibration verification procedure has been done every 3 months and confirms the mirroring of the 11 cytometers.

- Recruitment of patients is ongoing and no issues have been reported due to flow cytometry problems.
- In January 2016, files from phase I, phase II and from inception cohorts have been uploaded onto the server and their analysis is ongoing.

WP6 – Proteomics, Metabolomics and Serology of SADS

- One deviation due to the delay of D6.3 "Generate a list of biomarker candidates for validation and analysis in CS phase II and in the inception cohort" which is due to the delay in receiving samples from cohorts
- One change was done (and reported) for **19/UNIMI** with a decrease in the number of cytokines to be measured.
- Contract in place with LaboSpace Srl for Luminex analysis. Methodology for the measurement of the selected parameters validated. The order for the reagents submitted.

WP7 – Tissue taxonomy and Imaging Analysis

- RNA extraction from tissue samples has started (where the source of the samples is the biobank at UCL Brussels), and the major challenge was to set up procedures in order to extract enough material of enough quality from very small residual tissue fragments, initially obtained and stored for diagnostic purposes. After the last setting up experiments, RNA and proteins were extracted from SLE and control kidney biopsies in 12 batches, and from skin biopsies in 2 batches.
- Optimization of the protocols led to successful extraction of RNA and proteins from 87 kidney biopsies. (RNA quantity > 50 ng/sample). RNA quality was adequate in 47 samples, used for microarray analysis. Unfortunately, protein yield proved to be too low for proteomic experiments.

WP8 – Data analysis, bioinformatics and biostatistics

- TranSMART v1.2 was installed on the OMIC server. The structure of the OMIC data repository was documented. First data transfers were performed: clinical data (from WP3) and test flow cytometry data (from WP5). At the end of the year, eCRF records of 785 cross-sectional patients were made available to all partners through the PRECISESADS tranSMART. In parallel the data governance plan is being revised following-up on the Ethics and Legal joint workshop with AETIONOMY.
- The clustering analysis roadmap was defined. Two main clustering approaches were identified:
    - *cluster integration*, which refers to methods generating several sets of clusters, independently on each platform, all those sets of clusters then being integrated;
    - *integrated clustering*, which refers to methods generating a single set of clusters by combining the data from each platform.

A list of possible clustering methods, based on the two kinds of approaches described above, was defined and the most promising methods were tested and benchmarked.

- The algorithm to cluster genetic data and deconvolute disease information from ancestry was further developed, tested on previously available genetic data (from WP4), presented to a scientific conference, and an article was submitted to peer review.
- Gene expression data (publicly available) were studied to assess the impact of treatment effects, and also analysed with connectivity-map approaches. As a result, a couple of SADs marker genes were identified to be used for the analysis of PRECISESADS project data. In addition, a very large SLE gene expression dataset was made available by LILLY as an in-kind contribution.
- A new clustering software has been produced, initially prepared for cytometry data but can be used for all clustering, using graphic interphase programme called KNIME and supported on Docker. This software was created by FPS. The program is available as an open source (http://bmuchmore.github.io/CymeR/) and a publication has been prepared.

WP9 – Knowledge dissemination and training

- WP1 & WP9 has organized the Genomics Congress that will take place in Project Year 3 in March 2016 in Granada, Spain with many well-known speakers and CPD credit being offered.
- Publications from the Consortium as well as dissemination activities are detailed in Section 4.

1.4 Significant achievements since last report

WP1 has coordinated all WP activities to continue on schedule within the Consortium and actively sits in on each WP. Steering Committee met regularly and sent letters to sites to ensure recruitment stayed on target. The Annex 1, Description of Work was amended to reflect updates in certain work packages as well as the addition of Partner 29/Charite who replaced P24/DRFZ. This also necessitated amending the Project Agreement. The GA was held in February in presence of the advisory boards and the 2nd GA has been scheduled for March 2016, using the same format with the Advisory Board members. Immediately following will be the PRECISESADS Genomics Conference. Representatives of this WP have also participated in the joint Legal Ethical Workshop with Taxonomy Partner AETIONOMY. The PO has setup means to collaborate with the AMP RA/Lupus Network Project in the US, which is a similar initiative to IMI. (The overall goal of the AMP RA/Lupus Network is to identify relevant targets for diagnostics and treatment of RA, lupus, and related autoimmune diseases by defining shared and disease-specific biological pathways).

The most important achievement of **WP2** is the setting up of the logistics and coordination for the transportation and storage of the samples: sending of kits with all the necessary tubes, instructions, and other necessary documentation, receipt of samples, their registry, storage and processing of the derivatives (DNA, RNA), as well as the optimization of the methods for mRNA isolation (to obtain mRNA as well as short RNAs). The logistics set up are now flowing optimally.

WP3 has established a full set of data quality reports for the cross sectional study to help the sites and the project management office to monitor the quality of the clinical data database. The reports are issued monthly and individual site specific follow-up is organised according to the identified issues. Overall the data quality is very good with few pending queries.
Monthly data transfers to the tranSMART database managed by WP8 started in October 2015.

WP4 - We have produced the first genomic and epigenomic data with PRECISESADS samples. The meta-GWAS between several SADs is finished. Transcriptomics pipelines are set up and fully prepared to start the analysis.

WP5 - An intercalibration verification procedure, using the Duraclone compensation kit with the VersaComp Capture beads kit, is done every 3 months to be sure that the mirroring of the 11 flow cytometers remains optimal (June 22nd, October 19th , 2015 and January 14th , 2016). A bioinformatical procedure for the normalization of the files presenting some discrepancies with the reference has been validated.

WP6 - Methods are in place for: profiling of cytokines (also with contract in place), routine antibody profiling, metabolic profiling and small lipid-antibodies (also including findings on their regulation and role) and are currently run on samples from PRECISESADS.

WP7 - Protocol for RNA and protein extraction from small tissue samples (presently used for systemic sclerosis skin samples, in which protein yield appears to be adequate for proteomic experiments).
Protocol for hybridization of microarrays using RNA from human kidney biopsies has been established.
Proteomic experiments were initiated in kidneys harvested longitudinally in B6.Sle1.Sle2.Sle3 lupus-prone versus control mice. 35 human SLE kidney biopsies and 12 control kidneys were hybridized to microarrays, but the results were not satisfactory (extreme batch effects related to RNA extraction date). We have adapted the protocol to minimize these effects and are ready now to hybridize the samples again as soon as the scanner is functional again.

WP8 - The OMIC server is fully set up and hosts the first PRECISESADS patients' clinical data which is available via tranSMART The data governance is up and running. We are fully prepared to start the clustering analyses on biomarker data as soon as they are available.

WP9 - Several dissemination activities have been organised throughout the year ranging from interviews to national media (TV and newspapers), participation in the European Researcher's night and scientific presentations at a variety of International conferences.