SUMMARY FOR PUBLICATION

RHAPSODY
for precision therapy and prevention of diabetes

1. SUMMARY OF THE CONTEXT AND OVERALL OBJECTIVES OF THE ACTION

Diabetes is subdivided into two main forms, Type 1 Diabetes (T1D) and Type 2 Diabetes mellitus (T2D). Of them, T1D is more uniform and accounts for about 10% of all patients, whereas T2D, which accounts for about 80-90% of all patients, is much more heterogeneous, but the extent of this diversity is not known. RHAPSODY will explore whether diverse sub-forms of T2D are characterised by different rates of progression from pre-diabetes to T2D and by differences in disease progression, e.g. time to insulin. Establishing a better patient stratification already at diagnosis of diabetes will support the design of novel strategies for precision therapy and prevention of diabetes and of more efficient clinical trials.

We have now obtained lipidomic analysis on a total of 2,775 samples belonging to three T2D progression cohorts. Quantitative plasma peptide and protein data have been obtained from 1,200 plasma samples. Polar metabolites data have been obtained from > 5,000 plasma samples from the same cohorts as well as from 268 plasma samples from pancreas surgery patients (partial pancreatectomy patients) for whom we also have diabetes-related clinical data and islet transcriptomic profiles. We have also obtained miRNAs quantification from 500 blood RNA samples from one diabetes progression cohort. Through various bioinformatic ranking analyses, including literature search we have now established a list of plasma biomarker candidates for the prediction of T2D progression. This list includes 4 metabolites, 6 proteins, and 5 lipids. The expression of the 6 protein biomarkers has been replicated using validated, high throughput techniques in three additional cohorts.

2. WORK PERFORMED DURING YEAR 4 AND MAIN RESULTS ACHIEVED SO FAR

Data federation and systems biology

We have completed the full harmonisation of data annotation from 10 cohorts from 5 different European countries, representing more than 50,000 individuals. Two additional cohorts have been identified and will be integrated in the federated dataset. The newly obtained omics data have been uploaded, they include c-peptide measurements, lipidomics, peptidomics, metabolomics, microRNA and genetics data from 400 single-nucleotide polymorphisms (SNPs) for these cohorts. Analytical tools have been added to the database, including KNN (k nearest neighbour) imputation, Random forest ensemble-learning method, Similarity Network Fusion (SNF). We also established a biomarker prioritisation matrix and a web-based prioritisation tool allowing (i) to systematically evaluate biomarker candidates with relation to high quality and relevant external data sources and (ii) to place candidate biomarkers in the context of all RHAPSODY data generated and visualise results across multiple experiments. A new open access software, dsSwissKnife, has been written to improve server-side and client-side functions of the dsCDISC. The federated database has been actively exploited by all Work Packages to identify new molecular clusters, for integrating human islet transcript profiling with plasma metabolomics and clinical phenotypes, for modelling pre-diabetes progression in preclinical models using phenotypic, transcriptomic and lipidomic data, and to prepare upcoming meeting with the European Medicines Agency (EMA) to discuss biomarker validation.

Predictive biomarkers of glycaemic deterioration

Statistical tools have been implemented through the federated database to perform T2D progression modelling using genetic, metabolomic, lipidomic and proteomic data in combination with the clinical measurements previously used to stratify T2D patients in five subgroups. This integrated analysis allowed refining the previously characterised five patient subgroups. It showed that distinct underlying molecular mechanisms related to pancreatic islets, liver and adipose tissue metabolism can be identified, which provide novel biological insights into the diverse aetiological processes.

Pre-diabetes cohorts have been selected for analysis of six plasma protein biomarkers to help generate diabetes development models. Pre-diabetes modelling also includes genome-wide association study

(GWAS) data from the ADDITION, DESIR, BOTNIA, MDC and GLACIER cohorts, totalling greater than 10,000 individuals.

Predictive biomarkers of beta cell dysfunction

Preclinical models have been studied to identify the link between plasma lipids, beta cell dysfunction and gene expression modules, and the liver metabolic pathways that contribute to biosynthesis of these plasma lipids. A specific link between plasma triglycerides, liver lipid metabolic pathways and beta cell insulin secretion has been identified and is being validated by functional studies. Comparison of correlations between plasma triglycerides and islet gene expression in mice and in humans has led to the identification of common genes associated with the control of insulin secretion. These studies show that circulating triglycerides can be plasma biomarkers of beta cell function.

Partial pancreatectomy patients (121) were stratified in five subgroups as previously described. Their pancreatic islets have been characterised by RNAseq and complete plasma lipidomic and partial proteomic data have been obtained. Integrated clinical and multi-omics analyses have revealed islet gene co-expression modules and their correlation with plasma lipids, HbA1c and other diabetic parameters that lead to a predictive mechanistic model of beta cell failure.

Human islet transcriptomic, proteomic and lipidomic data are being generated to explore the islet cell mechanisms leading to amyloid plaque deposition, a marker of T2D islets, and to identify mechanisms by which gluco-lipotoxic conditions induce irreversible beta cell dysfunction.

Predictive biomarkers of insulin target tissue dysfunction

To identify circulating biomarkers of pre-diabetes progression or of rapid deterioration of T2D, we searched for liver, fat, and muscle metabolic pathways, which produce plasma lipids that have been identified as biomarkers of glycemic deregulations as these pathways may become new therapeutic targets. This search is now aided by a newly developed unsupervised multiblock analysis bioinformatics tool that helps to visualise and analyse, in an integrated manner, the physiological and multi-omics analy-

RHAPSODY – Assessing risk and progression of pre-diabetes and type 2 diabetes to enable disease modification.

Academia

Université de Lausanne Switzerland
Lunds Universitet Sweden
Technische Universität Dresden Germany
Università di Pisa Italy
Université Paris Diderot – Paris 7 France
INSERM France
Université Libre de Bruxelles Belgium
Institut Suisse de Bioinformatique Switzerland
University of Oxford United Kingdom
CNRS France
University of Eastern Finland Finland
University of Dundee United Kingdom
Imperial College London United Kingdom
Eberhard Karls Universität Tübingen Germany
Kobenhavns Universitet Denmark

University hospitals

A.O.U. Città della Salute e della Scienza di Torino Italy
Rijksuniversiteit Groningen The Netherlands
CHRU Lille France
LUMC – Leiden The Netherlands
VUmc – Amsterdam The Netherlands

Pharmaceutical Industries (EFPIA)

Institut de Recherches Servier France
Janssen Pharmaceutica NV Belgium
Eli Lilly and Company Limited United Kingdom
Novo Nordisk A/S Denmark
Sanofi-Aventis Deutschland GMBH Germany

Small and Medium-sized Entreprises

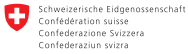
SCIPROM Sàrl Switzerland
Lipotype GmbH Germany

IMI-RHAPSODY.ORG – INFO@IMI-RHAPSODY.ORG

This project receives funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 115881. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA.

This work is supported by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number 16.0097-2.

The opinions expressed and arguments employed herein do not necessarily reflect the official views of these funding bodies.



ses of mouse adaptation to metabolic stress. Our analysis led to the initial establishment of an islet-liver-fat-muscle gene modules interaction network. This provides a framework for a new way of exploring inter-organ communication in T2D development. One of the specific genes studied was Spt2. Investigation of the impact of its inactivation in hepatocytes revealed important new information about a link between ceramide production, bile acid production, Fgf15 action, and control of glucose homeostasis.

3. PROGRESS BEYOND THE STATE OF THE ART, EXPECTED RESULTS UNTIL THE END OF THE PROJECT AND POTENTIAL IMPACTS

RHAPSODY's ambition is to fully characterise novel biomarkers for pre-diabetes and the prediction of T2D development and deterioration as well as to validate them for clinical use and pharmacological development. Biomarker identification, replication, and characterisation towards the use in clinical and pharmacological applications will be guided by the requirement for their validation by the EMA and for coherence with cost-benefits assessment.

Regulatory consensus for diabetes disease modification

RHAPSODY has engaged in a dialogue with the EMA to develop various operational definitions related to the use of biomarkers and to review existing regulatory guidelines for the use of such biomarkers. A second meeting with EMA, focussing on biomarker qualification, based on the analysis of the data accumulated in the federated database as described above, is being prepared.

Modelling economic and public health impact of disease modification

We aim to develop and validate economic models to quantify health outcomes and costs of pre-diabetes and T2D populations that are adaptable across EU jurisdictions and to evaluate the economic usefulness of biomarkers identified by RHAPSODY. To inform on the need to develop an economic model of pre-diabetes populations a systematic review of existing pre-diabetes models was performed. In addition, the PIs of pre-diabetes cohorts have been approached to gain access to the patient level data required to build the pre-diabetes model. A review of the published economic models to quantify health outcomes and costs of pre-diabetes was performed and published in year 4. It showed the limited validity of these models. A new model is being generated based on the analysis of data from 4 different cohorts totalling greater than 38,000 individuals.