IMI1 Final Project Report
Public Summary

Project Acronym: AETIONOMY

Project Title: Organising Knowledge about Neurodegenerative Disease Mechanisms for the Improvement of Drug Development and Therapy

Grant Agreement: 115568
Project Duration: 01/01/2014 – 31/12/2018

1. Executive summary

1.1. Project rationale and overall objectives of the Project

Today, diseases are still defined largely based on the presentation of signs and symptoms, yet while two patients may share the same diagnosis, the underlying causes of their symptoms may be very different. Naturally, this means that a treatment that works in one patient may prove ineffective in another. There is now broad acceptance that a new direction for disease classification is needed; built not on symptomatology but derived from the inherent pathogenic mechanisms that drive the disease.

AETIONOMY has paved the way towards an innovative approach to the classification of neurodegenerative diseases focussing on Alzheimer's (AD) and Parkinson's (PD) diseases. The project has successfully delivered an initial prototype of a new mechanism-based taxonomy for both conditions. This taxonomy prototype will change the view on pathophysiology mechanisms underlying Alzheimer´s disease and Parkinsonism in the future; we do hope that our taxonomy will play a role in the paradigm shift in drug discovery and development in both syndromes. Whether the insights generated during AETIONOMY will ultimately contribute to the development of new and effective treatments remains to be shown in the future.

1.2. Overall deliverables of the project

The AETIONOMY team, through its 5 work packages, tackled the problem of how to obtain, dynamically organise, structure, integrate and interpret a broad range of distinct types of data (ranging from molecular data, to information on symptoms) available across the neurodegeneration community at the start of the project. During the project, we brought new structure to the classification of disease by dissecting the underlying mechanistic/molecular causes of disease, and by linking these to clinical evidence to attempt to partially validate a selection of these mechanistic drivers.

The ambitious goal of AETIONOMY was, and is, far beyond the scope of any single company or university; the key to the success of AETIONOMY was the broad nature of the project consortium and the intense collaboration between essential project partners of a very different background and expertise, which normally do not have such an opportunity to work together in a research project. The project brought together 18 partners including pharmaceutical companies, universities, and patient groups, and combined substantial expertise in neurodegenerative diseases, molecular biology, clinical research, research ethics and law, neuroimaging, data modelling and simulation, data standards, and patient engagement in research among others.

Collaborations, beyond the original scope of the project but initiated through the identification of common goals, were very fruitful and key for success; in particular, with the University of Oxford (Prof. Simon Lovestone) and with the Human Brain Project (through Prof. Viktor Jirsa). The collaboration initiated with the Human Brain Project finally resulted in the inclusion of a leading team of the Human Brain Project into the AETIONOMY project; effectively building bridges between IMI projects and a Horizon 2020 Future and Emerging Technologies (H2020 FET) flagship project on brain research for the first time. The group of Prof. Viktor Jirsa (University of Aix Marseille) has not only brought one of the leading brain simulation platforms to AETIONOMY, but also paved the way for the successful uptake and continuation of the AETIONOMY approach in the context of the H2020 project

“VirtualBrainCloud”. For Alzheimer Europe, as an NGO promoting the rights and interests of people with dementia across Europe, the AETIONOMY project was a valuable opportunity to network with leading European researchers in the field and to help ensure that the perspectives and wellbeing of people with dementia were taken into consideration.

AETIONOMY has greatly influenced other IMI-projects and played a key role in bringing together the community of IMI projects in the neurodegeneration arena during two important networking events. The AD/PD conference in 2015 in Nice, which saw the inauguration of the IMI-AD platform, and the FINAL SYMPOSIUM of the AETIONOMY project, which hosted a review of talks representing most IMI projects in the field of neurodegeneration and quantitative neuropsychiatry.

AETIONOMY's innovative approach has delivered on the promise. Curated data, curated and computable knowledge, tools and mechanistic hypotheses have been structured in a way that we hope now enables the biomedical community and regulators to direct the development, approval and use of new diagnostic tests and treatments for Alzheimer's and Parkinson's diseases.

After 5 years of work, AETIONOMY has generated the following key deliverables:

1) A publicly accessible knowledge-base with inventories of mechanistic hypotheses that form the basis for the prototypic mechanism-based taxonomies for AD and PD, and the resultant data from the recruited clinical cohort from the novel cross-sectional study and additional integrated cohorts. This knowledge-base combines curated clinical and relevant OMICS-data, disease models for AD and PD and dedicated analysis and visualisation services.
2) An initial validation of the prototypic, mechanism-based taxonomies and the demonstration that the mechanism-based taxonomy can be used for patient subgroup identification in PD.
3) A comprehensive molecular profiling of clinical subjects aiming to characterise the mechanistic hypotheses identified via in-silico methods.
4) The concept of Virtual Dementia Cohorts (VDCs), for effective data sharing enabling mechanism based drug discovery for neurodegenerative disorders.

Strategic implications

AETIONOMY has achieved most of its objectives; in particular, a comprehensive selection of candidate mechanisms (the “inventory” of candidate mechanisms in the Multimodal Mechanistic Signatures Database for Neurodegenerative Diseases (NeuroMMSigDB)) has been generated and all relevant data and knowledge have been curated and re-annotated. The initial implementation and validation of a prototype mechanism-based taxonomy for AD and PD, and its partial validation in a clinical study, in addition has made it possible to share freely with the research community not only the inventory of computable models representing disease mechanisms hypotheses, but also a dedicated set of algorithms that facilitate the interrogation of these mechanistic hypotheses against patient-level data.

AETIONOMY was able to perform a partial validation of a selection of candidate mechanisms for AD and PD; this validation was done in silico and by means of dedicated wet lab experiments performed by our clinical research, biotech and pharmaceutical partners:

1) **In-silico validation** was performed on the candidate mechanisms in NeuroMMSigDB, with a special focus on the seven shortlisted candidate mechanisms, using independent data sets. Demonstrating the potential of these candidate mechanisms to identify strata of patients in patient-level data was a major challenge and turned out to be non-trivial. So far, this crucial validation was only able to be demonstrated for one mechanism and only in PD patient data.
2) **Wet lab validation** was attempted in all seven shortlisted candidate mechanisms based on exploring biomarkers representing these mechanisms. A challenge here was the link between the readouts in the candidate mechanisms and readouts (variables) in patient-level data. Dedicated in-silico enrichment procedures were used to bridge between wet lab biomarker measurements and computable mechanism hypotheses. Based on clustering of genetic variation information, mechanism-based strata could be identified in two independent PD cohorts.

Bridging between the world of "pathways" and multimodal mechanism graphs on one side and the readouts (variables) that can be found in patient-level (cohort) data sets on the other side has turned out to be one of the major challenges in our work. Deep biomarker profiling of the samples accessible to AETIONOMY was one way to increase the overlap between entities in our inventory of computable mechanism graphs (NeuroMMSigDB) and the entities (biomarkers) measured in patients.

Furthermore, establishing the relationship between the internally derived biomarker measurements and those external clinical biomarker data sets measured in publicly accessible cohorts (i.e. ADNI, EMIF-1000, AddNeuroMed, AIBL for AD; PPMI, MAP2PD) was a non-trivial undertaking and was not always successful (some of our mechanism-graphs are simply too devoid of biomarker information).

In an impressive effort, clinical partners in AETIONOMY coordinated their resources and managed the recruitment of a significant cohort of patients and controls, the logistics of sample transfers and the execution of a range of biomarker assays. In the end, we had our patient cohorts profiled and characterized to a degree that allows us now to perform "mechanism-enrichment" procedures on our patient-level biomarker data.

Of course, we had to be creative on the algorithmic side as well. The work package 3 team had to update the various analysis plans again and again to cope with the progress and the modalities of biomarker assays and measurements. The ability of the AETIONOMY consortium to dynamically adapt their strategy to link algorithmic approaches to available biomarkers was both, impressive and absolutely required. Consequently, AETIONOMY has demonstrated how to handle extremely diverse, heterogeneous data sets in integrative modelling and mining approaches and has developed a strategy to do so, which should enable future mechanism-taxonomy projects.

To overcome hurdles to access patient data for the in-silico validation of disease mechanisms and ultimately a first validation of a mechanism-based taxonomy of neurodegenerative diseases, we developed the concept of VDCs, which are synthetic (artificial) data sets that share features and

characteristics of real-world study cohorts in the area of neurodegenerative diseases. VDCs bear the potential to overcome some of the substantial challenges we face in translational neurodegeneration research, namely the:

- sharing of patient-level data without compromising patient data privacy,
- blending and merging of heterogeneous, complex clinical data sets,
- increasing the number of virtual patients to match statistical requirements,
- post-hoc enrichment of limited clinical data with additional features,
- integration of data and knowledge,
- ability to address counterfactual questions,
- computing of “what-if” scenario,
- ability to simulate trials in silic,
- ability to “play with data” in the sense that modelers and miners can test new methods.

We hope, through the conceptualisation and implementation of the VDCs, the AETIONOMY project may pave the way for future data sharing and data integration of patient-level data without compromising patient data privacy. The Virtual Patient topic was not a novel concept invented by the AETIONOMY consortium but is in line with globally emergent trends. Synthetic data sets and their use in pre-competitive collaboration have made their way into the finance sector (see https://mostly.ai) and are the subject of a future FET flagship project initiative (see https://www.health-eu.eu). With the Virtual Cohort concept being closely developed with the EFPIA industry partners (in particular UCB), AETIONOMY has laid the foundations that may enable a future pharmaceutical industry initiative for sharing and pre-competitive usage of patient-level data.Initiated by Martin Hofmann-Apitius, the academic lead of AETIONOMY, a group of academic and industry researchers have started a new series of annual workshops on longitudinal modelling of disease progression. The first workshop took place in Bonn in April 2019; the next one is planned for 2020 in Paris. This workshop brought together researchers from three different IMI-projects (AETIONOMY, EPAD, and RADAR-AD) and one H2020 project (VirtualBrainCloud). As a research topic, longitudinal disease progression modelling is of great interest for both, academic researchers and industry scientists. Given the hughe interest by both communities, we will push this topic further and aim at organising funding through a public-private partnership model.

An important further learning from AETIONOMY is that organizing data and knowledge from ongoing projects such as EPAD (European Prevention of Alzheimer's Dementia), or even getting permission to jointly work on legacy data such as ADNI (Alzheimer´s Disease Neuroimaging Initiative), turned out to be a greater challenge than expected at the start of the project. With the VDC approach, we hopefully will overcome these hurdles in future.

1.3. Summary of progress versus plan since last period

WP1	The project office successfully coordinated the last deliverables of the project and the final results were showcased in a symposium on Neurodegeneration together with other IMI projects in Bonn at the end of November. The final workshops to discuss the results obtained by the work package 3 & 5 partners were organized in London (UCB) and September (Barcelona) and these were fundamental to finalizing the project.
WP2	Final curation, mapping and incorporation of data sets into the AETIONOMY Knowledge base (AKB). Update and redesign of the AKB frontend according to recommendations of the interim reviews offering information on models, methods, web services, mechanistic disease hypotheses and data sets. Preparation (including the updated business plan) of the maintenance of the AKB for further 5 years as part of the ELIXIR-LU node. Generation of the first virtual dementia cohort for AD (ADNI avatar).
WP3	Data-driven (including preclinical lab work in pre-existing cohorts as well as data analyses on pre-existing datasets) and knowledge-driven (based on literature-mining and expert knowledge gathering) approaches were combined to identify candidate mechanisms, which were tested by WP3 on patient-level data (in silico validation and validation in the AETIONOMY study cohort). WP3 successfully established the relationship between our internally derived biomarker measurements and external clinical biomarker data sets. Finally, a number of web-recorded webinars on algorithmic approaches for mechanism-based patient-stratification and a number of publications were made available.
WP4	In T4.5, LUH assessed and implemented for the project the complex legal changes to EU data protection law resulting from the commencement of the General Data Protection Regulation (GDPR) in May 2018. This required LUH to draft and negotiate with partners the completion of a new agreement updating the earlier project data protection framework (which had implemented the older 1995 Directive). LUH participated, together with the legal team of UL-LSCB, in framing the terms of a further agreement to cover post-project data use in the project's Elixir legacy repository. The analysis of these matters formed the subject of a new additional Deliverable D4.5.2, submitted in M60.
WP5	As planned, WP5 provided a large amount of data and samples which were analysed and tested to support the validation of the mechanism-based taxonomy of neurodegenerative diseases . Both the AETIONOMY cross-sectional study dataset, and the additional external datasets from the consortium partners served as data sources for this effort. Thanks to the extensive clinical and laboratory efforts of partners, these resources provide a comprehensive clinical and multi-omic description of the studied populations An initial genomic-feature-based classification of PD patients has been achieved and validated in collaboration with WP3. Subsequently, we expect, via the integration of the large range of available meta-data and the deep exploration and further definition of patient subsets, the emergence of links between phenotype of those subsets to biologically testable parameters, which we see as a critical step towards our goals of supporting precision medicine.

1.4. Significant achievements since last report

Work - Package Number	Major Achievement
1	• Submission of all final deliverables (except 1) to IMI before the end of the project. • Organization of the final symposium on neurodegeneration where the project results were highlighted together with other IMI projects on NDD. • Management of all pending final tasks and communication of results and publications on a monthly basis to all partners, advisory boards, etc.
2	• Overall, the AETIONOMY Knowledge base (AKB) with its frontend intends to serve as a one point stop to offer access to all curated and harmonised data, databases and services along with tutorials and webinars to help users gain a deeper understanding and utilise the services developed during the AETIONOMY project. Following reviews from users and previous reviews, the AKB was improved and adapted to user feedback. Available at https://data.aetionomy.scai.fraunhofer.de/. • Using Bayesian modelling, the **Virtual Dementia Cohort (VDC)** was generated as a first synthetic (artificial) data set that shares features and characteristics of real-world study cohorts, like ADNI or Parkinson's Progression Markers Initiative (PPMI)). Our implementation of the VDC has a time dimension and is able to reflect disease progression.
3	• Strategies for in-silico validation of candidate mechanisms and its publication (D3.9.3.1); • Validation of o in-silico candidate mechanisms (high-throughput with NeuroMMSigDB server, D3.9.1.3), o surrogate markers in clinical data (e.g. observational data from published studies) related to syndecan-dependent uptake and signalling mechanisms (D3.9.2.3), and o imaging-related candidate mechanisms using advanced strategies like regression analysis on genetic and imaging biomarkers (HASE), Voxel-Based Morphometry, and Event-Based Modelling (D3.9.3.2); • Target validation and evaluation of druggability of syndecan-dependent ligands related to hypotheses of the clinical study (D3.9.2.4); • Mining on Bayesian representations of major cohort studies including ADNI and PPMI and their disease specific clinical features, disease risk models and stratification of patient subgroups (D3.9.1.4).
4	• Analysing and responding to changes in EU data protection regime, and implementing these in data-sharing agreements to ensure legally compatible data use by Project and its legacy resource.

5	• All the tasks related to the AETIONOMY Cross-sectional Study (AETIONOMY-CS) were achieved this year with the AETIONOMY-CS dataset management and lock, and the delivery of the final Clinical Study Report (CSR). • WP5 partners have additionally generated 8 datasets on biomarkers that were chosen for the mechanism-based taxonomy of neurodegenerative diseases. • A first genomic-based classification of PD patients was proposed.

1.5. Scientific and technical results/foregrounds of the project

WP2 – Sustainability of the AKB

The AETIONOMY Knowledge Base (AKB) and associated tools and services will be maintained for an additional 5 years by partners Fraunhofer and UL. A data access and management plan has been conceived and all the data and samples generated will be accessible, subject to approval by the data access committee (discussed in detail in the sustainability plan).

Moreover, all the disease models (BEL models AD and PD, and PD Map) and services (BEL commons, NeuroMMSigDB) will continue to be updated and improved by the respective partners. All the services and pipelines developed during AETIONOMY are described by tutorials and webinars to introduce the methods and use cases to interested users. These materials and webinars are available the AKB website (https://data.aetionomy.scai.fraunhofer.de/)

Besides these disease modelling aspects, the approach to generate a Virtual Dementia Cohort was realized. Here, two different, complementary research activities have been used to generate synthetic data:

a) Generation of structural and functional connectivity data with Mechanistic Modeling for virtual brain simulation – AMU
b) A Machine Learning based framework for virtual cohort simulation – Fraunhofer and UCB

Virtual Brain Simulations using Mechanistic Modeling

AMU proposed a whole-brain computational approach to model the whole-brain structural and functional connectivity of each subject by project The Virtual Brain (TVB). All analyses and simulations were carried out using connectivity matrices based on a 96-brain parcellation. In the figure 1 below the TVB pipeline for completing missing data in ADNI has been illustrated.

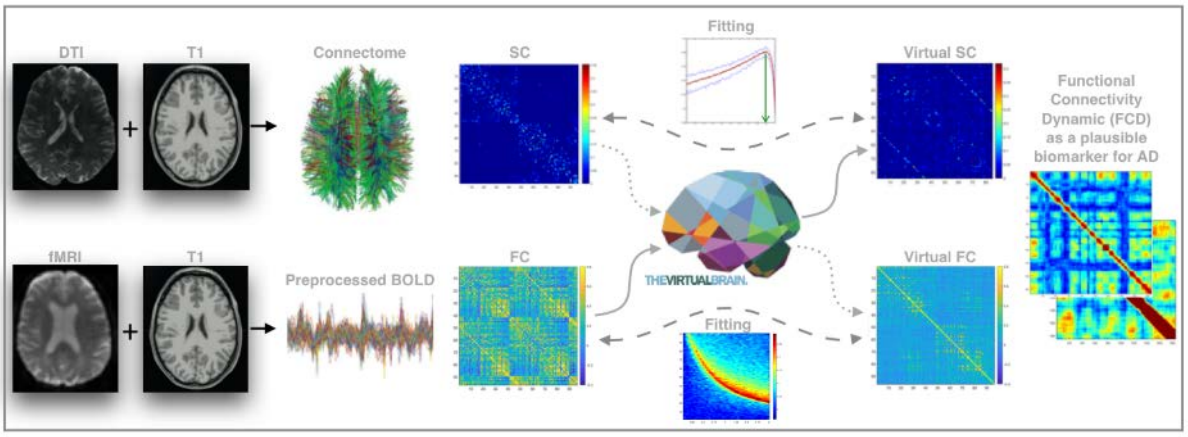


Figure 1 – The TVB pipeline for completing ADNI data.

As a first step, tractography was performed using data from Diffusion Tensor Imaging (DTI) and by using T1 images and implementing the 96-brain parcellation, the structural connectivity (SC) matrices for each subject were built. For constructing the functional connectivity (FC) matrices, after preprocessing the fMRI images, the parcellation were implemented in order to reach BOLD signals. Afterwards, by calculating the Pearson correlation coefficient of signals, the FC matrices were built.

At the second step, the simulation and parameter fitting were performed on 12 subjects for Ornstein-Uhlenbeck process and Wong-Wang model. By linear Ornstein-Uhlenbeck process we could simulate SC from FC and when SC were missing and vice versa. As it is illustrated in table 1, we could virtualize FC of 76 patients from their SC and SC of 156 patients based on their FC.

- Patient data base completed for 76 patients with SC only
- and 156 patients with FC only
- Validated for 12 patients with SC and FC

Type of data	SC/FC	SC	FC
Empirical patient data	12	76	156
Virtual patient data	12	76	156

Table 1- The number of data completed for the group of SC only and FC only

The Wong-Wang model was used for the simulations of the group of patients, for which their functional data were missing in order to simulate their BOLD signals besides having more sophisticated FC and FCD matrices.

We developed methods to compute the missing data on basis of the available data, and for the computation of the connectivity matrices, either data for functional or structural connectivity is necessary. Metrics applied in this study are Functional Connectivity Dynamics (FCD) of simulated and empirical time-series. We demonstrated data completion is feasible when one of the structural or functional data is missing and we have solved the problem by filling all the gaps. The ADNI database has been now extended from 12 complete data sets to 156 comprising DTI, Magnetic Resonance Imaging (MRI) and functional Magnetic Resonance Imaging (fMRI) data.

A Machine Learning based framework for virtual cohort simulation

The underlying concept is straightforward, translational research of many disease areas requires a longitudinal understanding of disease development and progression across all biologically relevant scales. Several corresponding studies are now available. However, to compile a comprehensive picture of a specific disease, multiple studies need to be analyzed and compared. Many clinical studies are nowadays conducted in the context of drug development in pharmaceutical research. However, legal and ethical constraints typically do not enable sharing of sensitive patient data in a form which supports the type of integration, data processing and assimilation that is needed to bring multiple independent studies together for these analyses. Consequently, there exist data ``silos'', which can hinder progress in translational research.

In WP2, we suggested the idea of a Virtual Cohort (VC) to address this limitation. To realize this goal, UCB and Fraunhofer together developed an innovative data analytical approach combining Bayesian Networks and deep learning methods. The underlying model is trained based on pre-existing data (e.g. ADNI, PPMI) and subsequently can be used to simulate virtual subjects. We showed that, with the help of our method, we can simulate AD and PD subjects that are largely indistinguishable from real ones. Our approach (see Figures below) allows for the incorporation of arbitrary multi-scale, multi-modal data without making specific distribution assumptions. Moreover, we demonstrated the possibility to simulate interventions (e.g. via a treatment) in the VC (see Figure). Overall, our proposed approach offers the possibility to build sufficiently realistic VCs for multiple disease areas in the future.

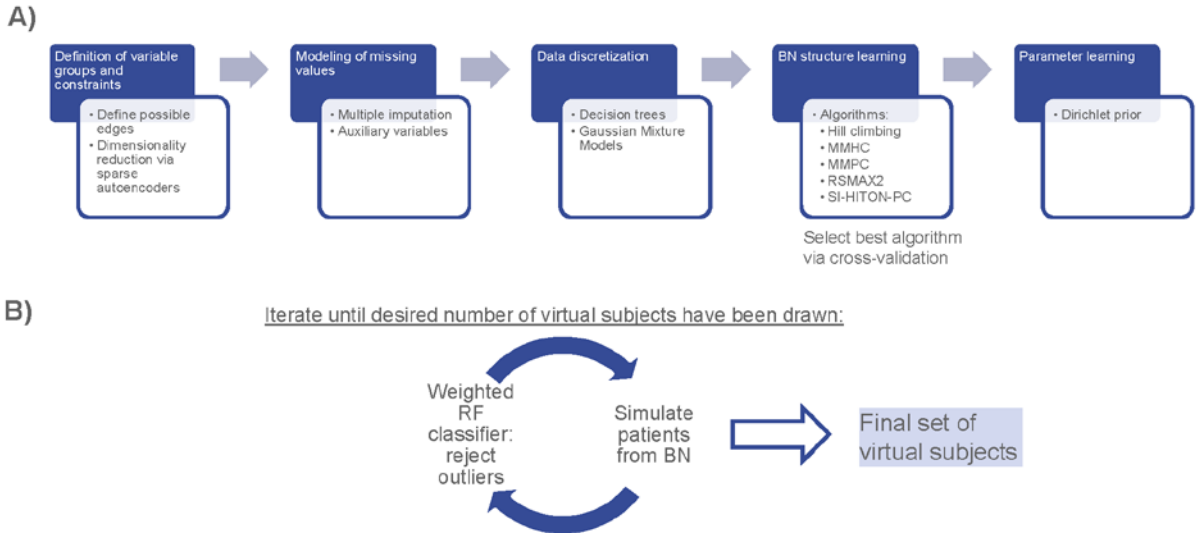


Figure 2: Overview about our developed approach for virtual patient simulation

As shown in Figure 2 (above), complete workflows for the generation of longitudinal synthetic patient cohorts have been implemented and tested in preliminary test scenarios.

By combining the generative Bayesian approach for virtual patient generation with the simulation of brain connectivity and function, we implemented a strategy to generate a Virtual Dementia Cohort (VDC) that displays the typical features and distributions of a real-world cohort (e.g. ADNI) but comprises only simulated patients.

Below in figure 3, we show an example for the use of Virtual Dementia Cohorts: the simulation of counterfactual interventions. The term “counterfactual” describes scenarios that cannot be easily found in the real world. A typical example would be the extrapolation of a study cohort in time towards earlier time points (e.g. as if recruitment would have taken place already in childhood). We can also make assumptions about the presence or absence of certain factors (e.g. synuclein expression levels) and “play” with “in-silico-knockdowns” of key pathways.

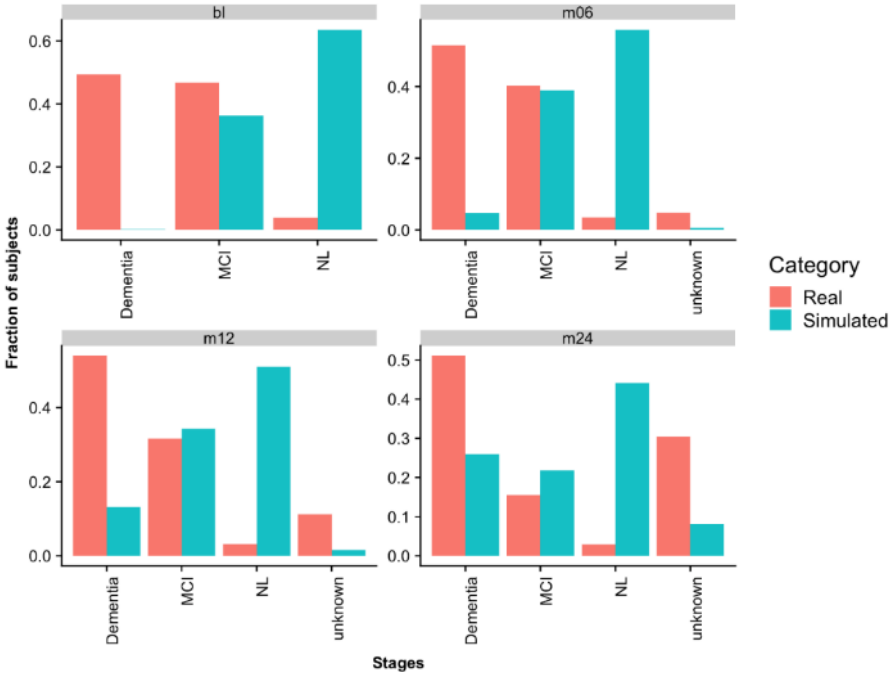


Figure 3: Counterfactual simulation of an intervention into a virtual AD cohort: real = original data; simulated = simulated intervention

The preliminary application examples worked out during AETIONOMY demonstrate impressively, how Virtual Cohorts may help answering complex scientific questions.

These complementary research activities by AMU and Fraunhofer/UCB have resulted in the successful generation of synthetic data for patients provided as a Virtual Dementia Cohort prototype.

WP3 - Identification of disease candidate mechanisms and their validation

An essential activity of AETIONOMY was to generate hypotheses about multiscale mechanisms of neurodegenerative pathophysiology and to explore new ways of stratifying patient subgroups utilizing OpenBEL graphs which are linked by mechanism (causal chains) and include entities that can serve as “biomarkers”. In the final months of AETIONOMY, WP3 has done impressive work on mechanism-based patient-stratification; on algorithms suited for the identification of candidate mechanisms for stratification and for the mechanistic interpretation of patterns found in clustering experiments.

The following sheds light on some of the work done in the very last months of the project:

WP3 generated hypotheses and in-silico candidate mechanisms

WP3 proposed on basis of their disease modeling and analytics the following candidate mechanisms and biomarkers for further clinical validation:

1. ***Role of insulin pathway in AD and PD progression*** (biomarkers STK11, INSR): Comorbidity association of Alzheimer's disease (AD) and type 2 diabetes mellitus (T2DM) by genetic variants of clusterin (CLU) and serine/threonine kinase 11 (STK11) genes. In figure 4, the NeuroMMSigDB AD T2DM SNPs subgraph is shown:

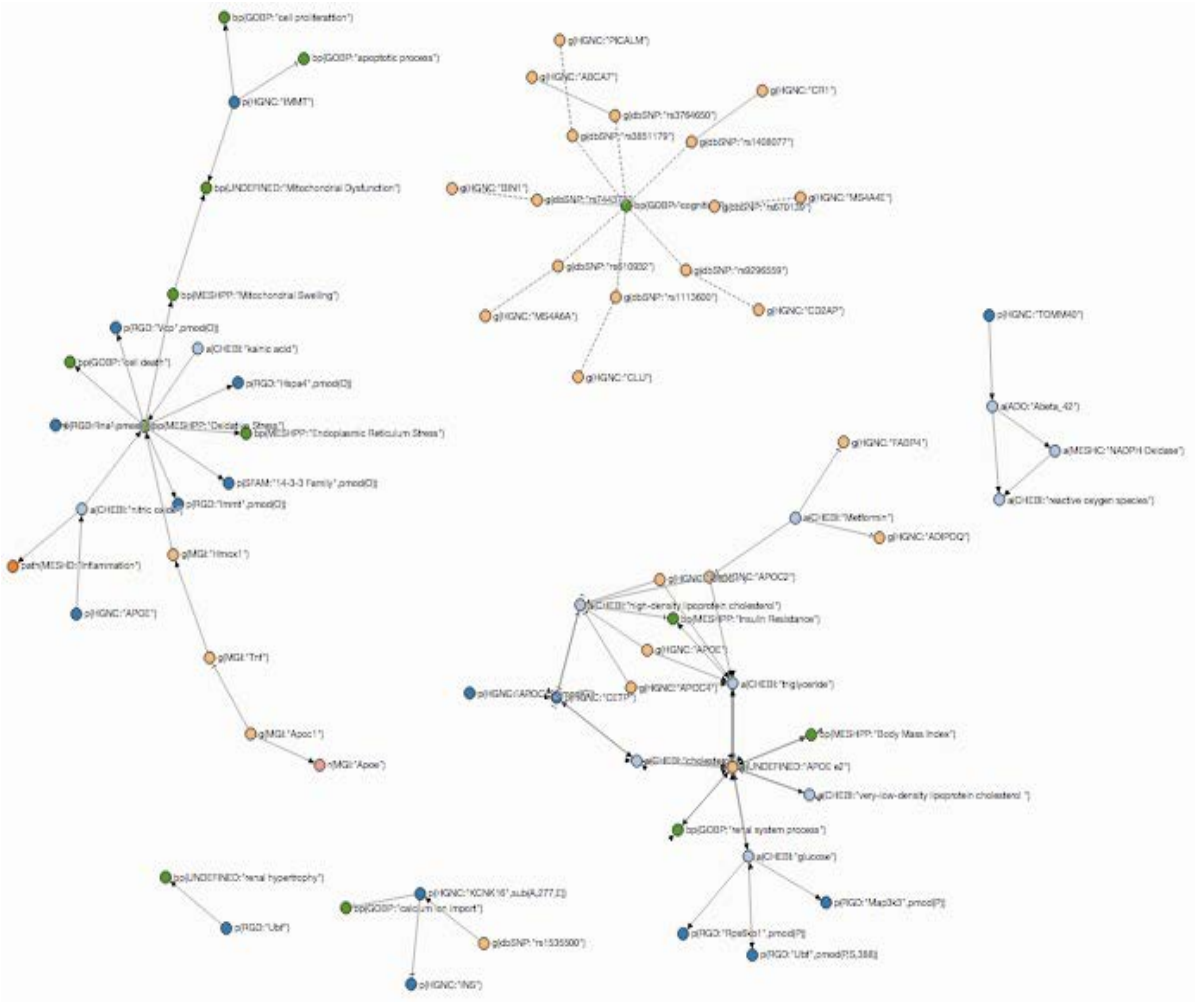


Figure 4: NeuroMMSigDB AD T2DM SNPs subgraph

Comorbidity association of Alzheimer's disease (AD) and type 2 diabetes mellitus (T2DM) by genetic variants of clusterin (CLU) and serine/threonine kinase 11 (STK11) genes: In the normal state (green colour edges), insulin protein binds to its receptor insulin receptor and this binding event activates INSR through phosphorylation. The activated INSR binds to insulin-like growth factor 1 (IGF1) and activates insulin receptor substrate 1 (IRS1). Activated IRS1 activates the phosphoinositol signalling system which activates protein kinase B (AKT) signalling and controls glycogenesis. Activated INSR binding to IGF1 also activates Src homology 2 domain containing protein (SHC) and thereby activates the MAPK signalling pathway. In the disease state (red colour edges), CLU promotes neuron apoptosis. Amyloid beta peptides bind to INSR, effectively preventing activation of INSR by insulin. As a

consequence, through inactivation of the phosphoinositol signalling system, AKT signalling and mitogen-activated protein kinase (MAPK) signalling pathways, binding of APP peptides suppresses the insulin signalling pathway. CLU single nucleotide polymorphisms (SNPs) are associated with an increased production of amyloid beta peptides and the CLU variants increase the risk of T2DM by primarily inducing the insulin resistance and secondly by decreasing the production of insulin. In the case of insulin resistance, the level of INS is increased due to its accumulation in the blood. Normally, under the condition of energy stress, STK11 activates adenosine monophosphate-activated protein kinase (AMPK) by phosphorylation and AMPK activation decreases mTOR, the mechanistic target of rapamycin serine/threonine kinase signalling activity, thereby helping degradation of β-amyloid. In T2DM, the SNP rs8111699, which maps to the enhancer region of the STK11 gene, is influencing insulin sensitivity. The other SNP (rs741765) is located in the insulator region, which may block the interaction between the enhancer and promoter of the gene, resulting in downregulation of the STK11 gene. Deficiency and dysfunction of STK11 inhibits the AMPK phosphorylation, thereby reducing the activity of AMPK, which hyper-activates mTOR signalling in AD. Moreover, in T2DM, hyperactivation of mTOR signalling inhibits IRS1 via activation of S6K1 and the IRS1 inhibition leads to insulin resistance (linking the STK11 causal graph to the CLU graph), which leads to increase in INS and glucose in blood. The black coloured arrows (up and down) indicate over- or under-expression of the nodes in diseased state; while dotted arrows are inferring the possible effect of genetic variants.

2. ***Stress-induced comorbidity of AD and PD*** (biomarkers CRH, CRHR1, MAPT, KANSL1). The following figures 5 and 6 present NeuroMMSigDB CRH and KANSL1 subgraphs:

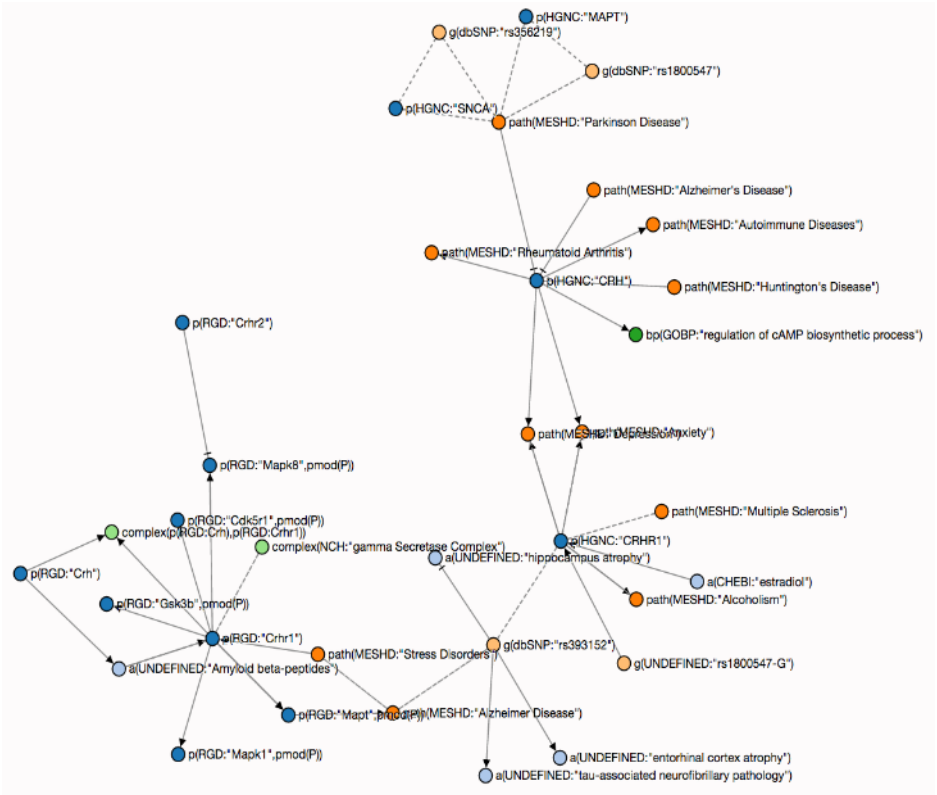


Figure 5: NeuroMMSigDB CRH subgraph

Stress up-regulates CRH gene expression, which interacts with its receptor, the CRHR1 protein; the CRHR1 gene is highly expressed in hippocampus and the complex between the hormone and its receptor (CRH+CRHR1) can be detected in that brain region. In addition, the CRHR1 protein also interacts with γ-secretase, which is associated with Aβ accumulation, one of the hallmarks of AD pathophysiology. The hormone receptor protein complex (CRH+CRHR1) is further linked to the up-regulation of GSK3β and the phosphorylation of essential elements of the ERK1/2/MAPK pathway. Up-regulation of GSK3βis associated with MAPT hyper-phosphorylation; in addition, phosphorylated MAPT and ERK1/2/MAPK pathway up-regulate Neurofilament phosphorylation, which has been associated with AD. The complex physiology is even increased through the interaction of the 'CRH+CRHR1' protein complex with the BDNF protein; this interaction has already been associated with AD pathology. The complex also enhances neuronal activity by interacting with adenylate cyclase, cAMP, act(PAK), Ca2 signalling pathways. The resulting enhanced neuronal activity has been shown to further accumulate interstitial fluid amyloid beta (ISF Aβ), while this accumulation of ISF Aβ is also linked with up-regulation of CRH gene expression, effectively establishing a feedback loop that can enhance negative dysregulation events. MAPT hyper-phosphorylation also increases its dissociation from microtubules, a process that has been linked to Lewy-bodies and Parkinsonism, in the PD context. Finally, the CRHR1 antagonist 'Antalarmin', which is used in response of chronic stress, has been shown to reduce Aβ Accumulation in brain, adding further meaningful, supportive evidence in context.

In the following figure 6 we show the NeuroMMSigDB subgraph for the second biomarker related to this mechanisms KANSL1:

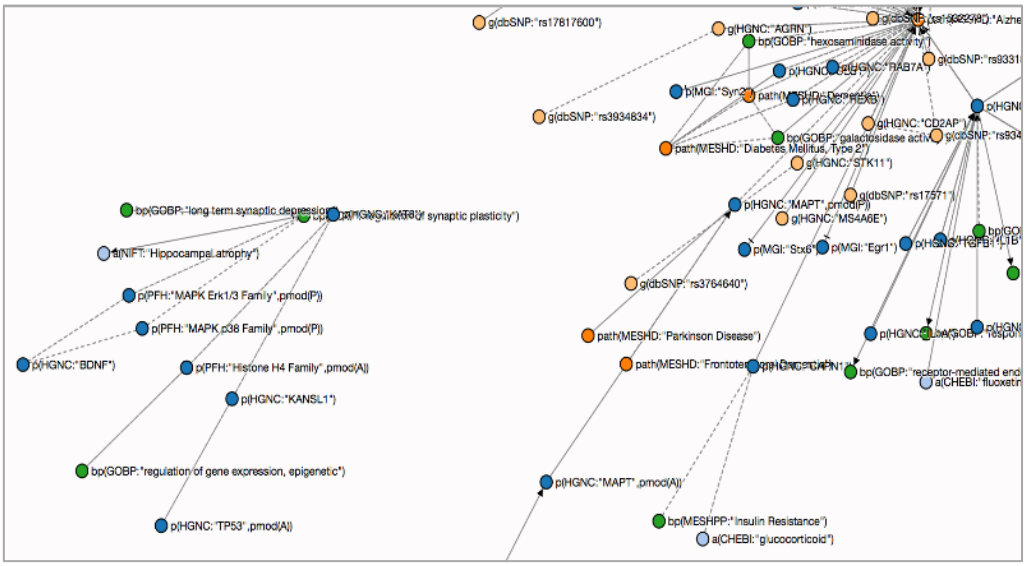


Figure 6: NeuroMMSigDB KANSL1 subgraph

KANSL1 is a subunit of two protein complex NSL1 and MLL1 that plays a role in histone acetylation. This gene encodes a nuclear protein that is involved in chromatin modification and belongs to histone acetyltransferase (HAT) complexes. In the AD context, KANSL1 participates in histone acetylation inhibition, resulting in disruption of gene expression. KANSL1 also influences MAPT hyperphosphorylation through the MAPK pathway, as MAPT is a well-known biomarker for PD. The Amyloid-beta and oxidative stress inhibited CREB activity and further inhibits BDNF expression resulting in Long Term Potentiation which ultimately results in Hippocampal atrophy. The activated

acetylated Tau inhibited the KIBRA normal function in actin polymerization which ultimately results in long Term Potentiation dysfunction and hippocampal atrophy. The CLU mutation which prevents the phagocytosis of microglial cells, leads to white matter degeneration through the dysregulation of oligodendrocytes.

In the 'Webinar on In-depth validation of imaging-related candidate mechanisms using advanced strategies, like regression analysis on genetic and imaging biomarkers (HASE), Voxel-Based Morphometry (VBM), and Event-Based Modeling (EBM)' (D3.9.3.1), EMC demonstrated the results from the imaging-related analyses. For eight selected (strong-signalling) SNPs, VBM analysis was carried out to explore hypothesis free association with gray matter volume. Results (top20-list) were presented in a report from EMC's HASE-experiments on ADNI using the 2.4 million of SNPs and about 370 imaging biomarkers computed by FreeSurfer v6.0, including volumes of 12 subparts from the hippocampus. The p-values of the top 4 are significant also after Bonferroni-correction.

Furthermore, EMC selected 33 KANSL1-proxy-SNPs to redo the HASE (linear regression) association studies and VBM approaches. For all SNPs mentioned, Gene Expression was measured using the Allen Human Brain Atlas. Figure 7 below shows these results:

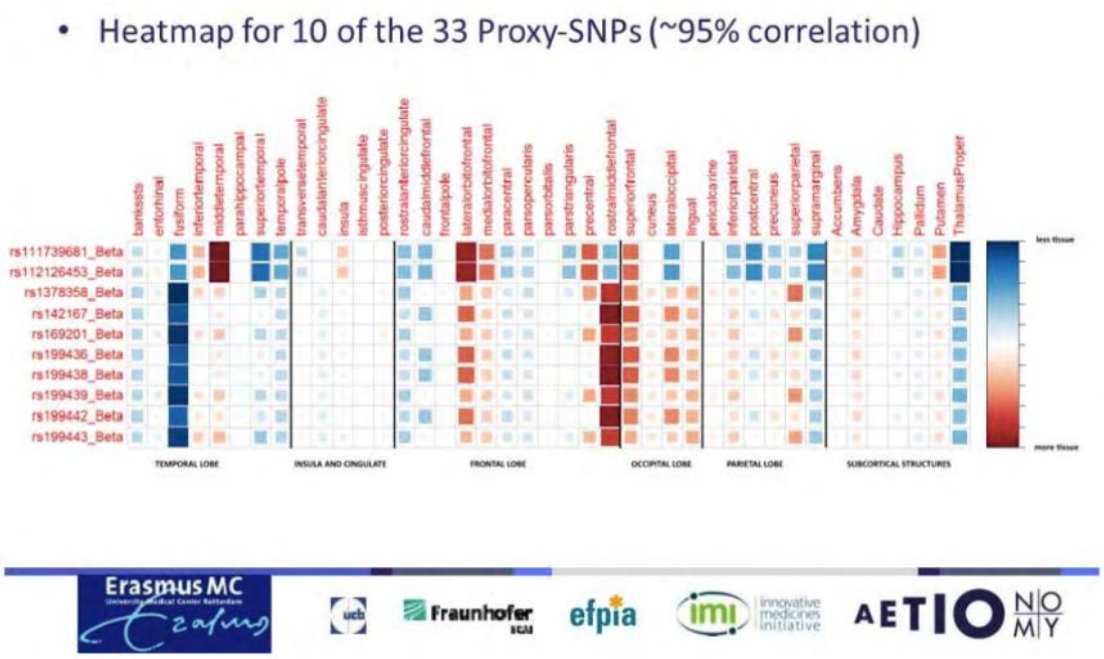


Figure7: KANSL1-proxy-SNPs analysis (HASE association studies and VBM approaches)

More information about the Webinar and EMC's results can be found under the URL: *https://www.aetionomy.eu/en/media/webinars.html* .

3. ***Syndecans uptake of aggregate proteins in neurodegenerative diseases*** (biomarkers: AD: GSK3, APOE, DPYSL2, MAP1B, PTP1B, RTN1. STMN1; PD: hnRNPK, PARK7; and ALS: MAP1B). In figure 8 below, the NeuroMMSigDB syndecan subgraph is shown:

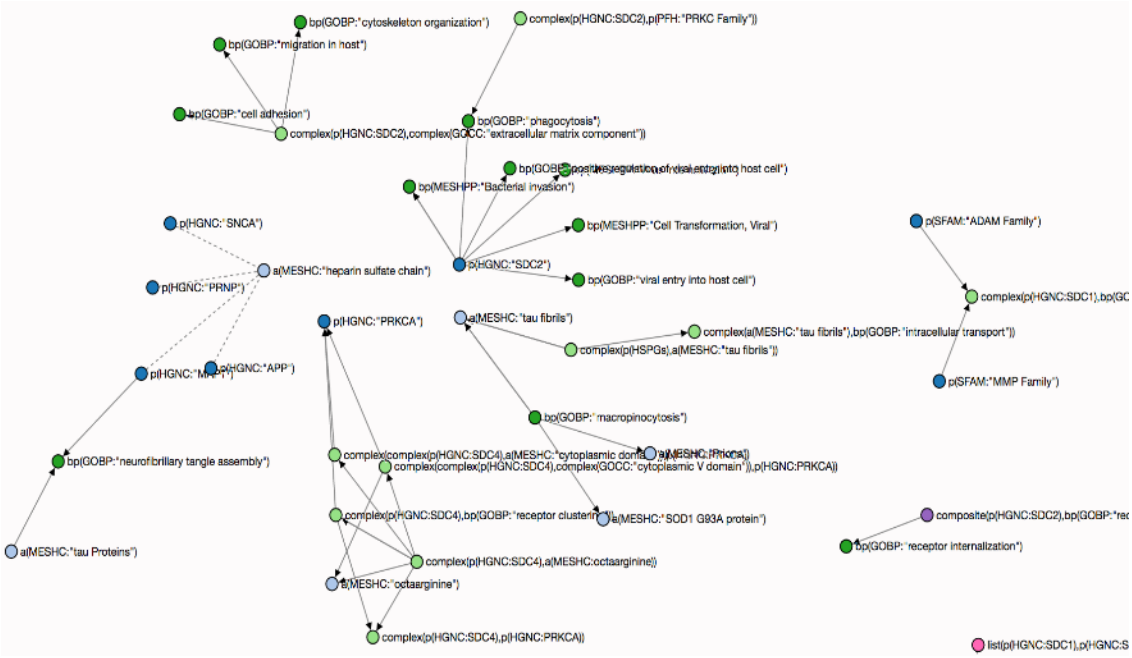


Figure 8: NeuroMMSigDB syndecan subgraph

Fundamental mechanisms underlying the biological processes governing the cellular spreading of pathologically misfolded proteins in AD and PD. Letoha et al. (PHARMACOIDEA) pursued these objectives with established cellular platforms overexpressing distinct isoforms of syndecans (SDCs), a heparan sulfate proteoglycan (HSPG) family with an established role in mediating intracellular delivery of various macromolecular entities including growth factors, viruses and bacteria.

PHARMACOIDEA confirmed in their analysis (report D3.9.2.4) previous findings on the role of SDCs, especially SDC3 in the interaction with neurodegeneration-related misfolded proteins, namely Aβ1-42. They gathered further evidence on the correlation of SDC3 expression and key steps of neurodegeneration in cellular binding & uptake and fibrillation of Aβ1-42. Druggability assessment established the major role of polyianionic HS chains in SDCs' interactions with Aβ1-42. Thus SDCs' contribution to Aβ1-42 attachment or fibrillation could be only achieved by blocking HS chains of SDCs, interventions targeting other parts of the SDC core protein is less effective. Studies on cells lines grown in high glucose media also showed the high glucose concentrations increase SDC3 and HS concentration, thus suggesting a mode in which glucose might affect Alzheimer's disease patients. The following figure 9 of report D3.9.2.4 shows this aspect of high glucose concentrations on SDC3 and HS expression:

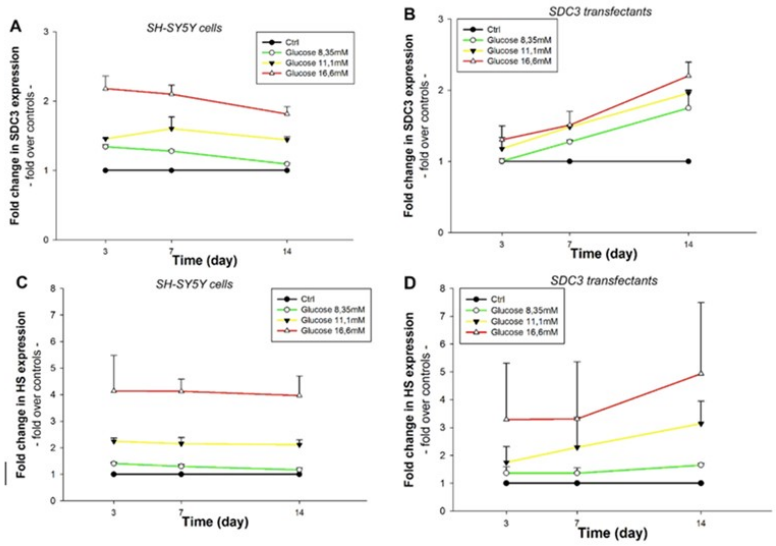


Fig. 9. Effect of high glucose concentrations on SDC3 and HS expression. SDC3 transfectants (created in K562 cells), along with wild type SH-SY5Y cells were grown with or without high glucose concentrations.(8,35mM; 11,1mM; 16,6mM). The medium was replaced in every 3 days and SDC3 and HS expression was measured on Day 3, 7 and 14 with flow cytometry. Detected SDC3 and HS expression levels were normalized to cells grown in medium with standard glucose concentration (5,5 mM). The lines represent mean ± SEM of two independent experiments.

The observed influence of glucose levels on SDC-mediated uptake suggests the investigation of different patterns of "spreading" in diabetic patients (diabetes as co-morbidity to AD). Some of the findings by partner PHARMACOIDEA have been published.

Validation of candidate mechanisms and approaches to stratify patient subgroups.

Regarding the in-silico validation of candidate mechanisms, we followed the agreed strategy to analyse candidate mechanisms against real patient data, e.g. AddNeuroMed data (ADNI- and PPMI-independent data sets) to look mainly for genetic overlaps to our candidate mechanisms. We proceeded to map data-derived information (pathway level summaries of SNPs) to literature-derived knowledge (NeuroMMSigDB) resulting in an AD risk model and explored in greater extent for a mixed AD/PD cohort from ADNI and PPMI. Clustering experiments were done, and the individual resulting clusters were tested for their association with mechanism hypotheses encoded in BEL graphs (coming from NeuroMMSigDB). Amongst the features used in the clustering experiments were:

- Various clinical features, including neuropsychological assessment scores,
- Genomic features,
- Neuro-imaging features and PET scan results.

Based on these data, Fraunhofer and UCB have developed a model predicting the AD risk for each individual patient. Our results indicate a high prediction performance with a concordance index of ~86%. The model can be used to stratify patients based on their disease risk. Stable dependencies between relevant features were estimated via a Bayesian Network approach and could (at least partially) be linked to causal biological mechanisms. Altogether, the utility of our work can be thus seen two-fold: First, our AD risk model could be used to identify patients for early therapeutic

intervention. Second, with the help of our Bayesian Network approach combined with the OpenBEL encoded cause-effect relationships, we were able to achieve mechanistic insights into the key mechanisms that relate to the conversion from normal / MCI to AD. These insights might support the development of novel therapies. Finally, we demonstrated the possibility to model entire clinical cohorts in a fully longitudinal manner while appropriately considering missing data. We showed that with the help of a Bayesian Networks, we can build models that are predictive as well as generative. They thus allow for predicting the effect of counter-factual situations (e.g. therapeutic alpha-synuclein reduction; see the section on WP2, page 13) as well as simulating virtual patient trajectories that are largely indistinguishable from real patient trajectories.

Our framework could thus become a major step towards the simulation of clinical trials as well as building virtualized versions of proprietary data, hence addressing a key bottleneck in data sharing. More details are described in the Webinar on 'Mining on Bayesian representations of major cohort studies including ADNI and PPMI' describing disease specific clinical features, disease risk models and stratification of patient subgroups (D3.9.1.4). More information can be found under the URL: *https://www.aetionomy.eu/en/media/webinars.html* .

Efforts were invested into improving the interoperability between mechanism-representations and conventional pathway databases (described in two publications). We already planned for additional mining efforts beyond the end of the funded period of AETIONOMY.

Candidate Mechanism Perturbation Amplitude (CMPA)

With the CMPA approach our goal was to quantify mechanisms using gene expression data sets. Fraunhofer described, how the approach – a heat diffusion algorithm, can be applied to NeuroMMSigDB mechanistic subgraphs in order to investigate which mechanisms are perturbed in different patient subgroups or brain regions. This algorithm supports the identification of mechanisms regulated at different intensities at different stages of a disease and brain regions.
The CMPA algorithm was implemented for 3 different clinical candidate mechanisms:

- Mitochondrial dysfunction in PD,
- NFT aggregation in AD, and
- Neuroinflammation in AD.

The approach is illustrated in the figure 10 below:

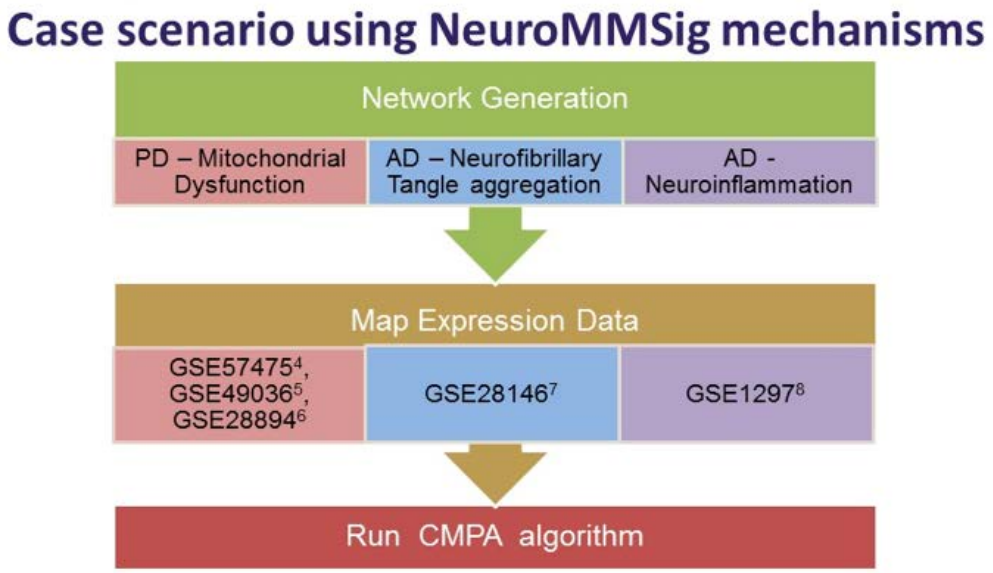


Figure10: CMPA heat diffusion algorithm applied for 3 clinical candidate mechanisms for AD/PD and furthermore for all NeuroMMSigDB signatures to quantify the intensity of perturbed mechanisms associated with AD/PD (as the consequence of differentially expressed genes).

The CMPA algorithm quantifies the intensity of perturbed mechanisms associated with diseases as the consequence of differentially expressed genes as seen below in figure 11. Therefore, the CMPA algorithm reduces the existing caveat of mere mechanism-disease associations by showing that mechanisms regulate with different intensities across spatial and temporal dimensions.

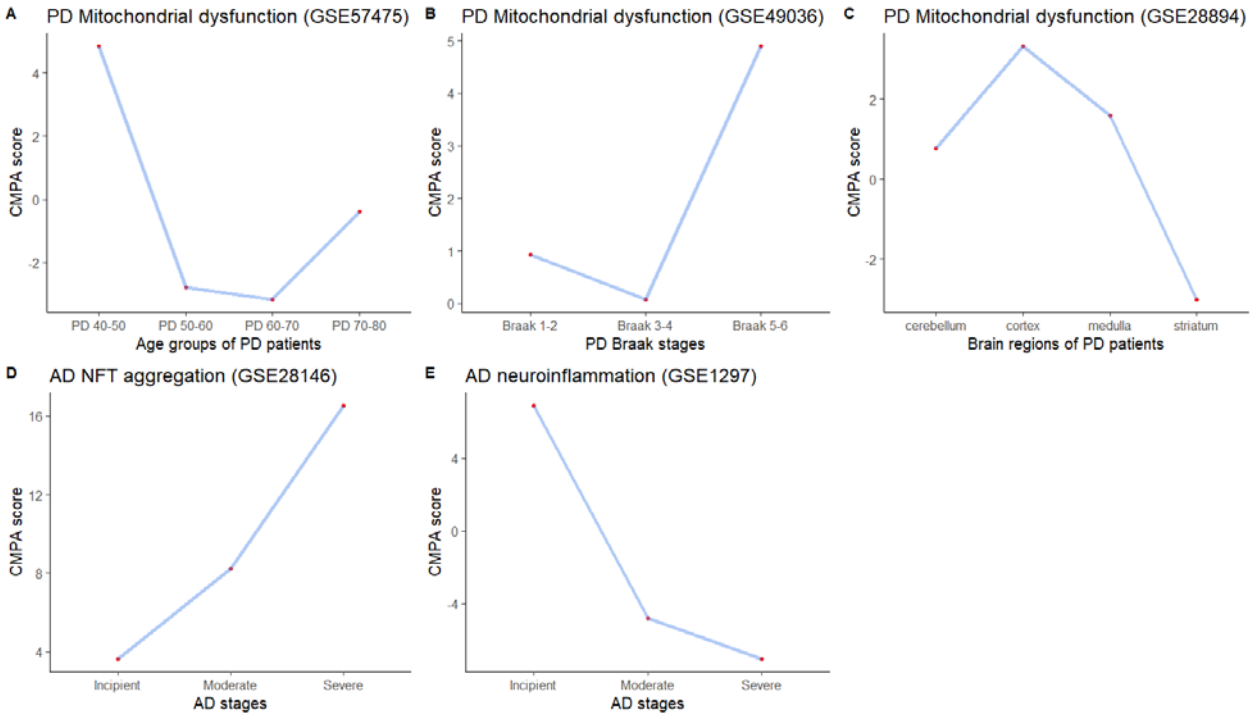


Figure 11: CMPA scores for mitochondrial dysfunction in PD (A, B and C), NFT aggregation (D) and neuroinflammation (E) in AD

Results of the CMPA analysis

- CMPA analysis on GSE57475 shows that the mitochondrial dysfunction in PD is perturbed the most in the age group 40-50 of PD patients with a CMPA score of 4.84.
- In dataset GSE49036, mitochondrial dysfunction is perturbed the most in Braak stage 5-6 of PD patients with a CMPA score of 3.33.
- Among brain regions, the cortex is the region with the highest perturbation of mitochondrial dysfunction with a CMPA score of 4.9.
- Similarly, with a CMPA score of 16.53, patients with severe AD had the highest perturbation of aggregation of NFTs as compared to incipient and moderate stages of AD.
- The highest perturbation of mechanism representing neuroinflammation in AD is observed in incipient AD patients with a CMPA score of 6.9.

To conclude, WP3 established the relationship between our internally derived biomarker measurements and external clinical/biomarker data sets – ADNI, EMIF-1000, AddNeuroMed, partly also Australian Imaging, Biomarker & Lifestyle Flagship Study of Ageing (AIBL) for AD; PPMI, "Mapping Proteomics to Parkinson's disease" (MAP2PD-UK) for PD. Mining on Bayesian representations of major cohort studies including ADNI and PPMI and their disease specific clinical features, disease risk models and stratification of patient subgroups were reported in the Webinar D3.9.1.4 mentioned above (more information can be found under the URL: *https://www.aetionomy.eu/en/media/webinars.html*).

Based on the above-mentioned data Fraunhofer and UCB have developed a model predicting the AD risk for each individual patient, which can be used to stratify patients based on their disease risk.

Fraunhofer described how the CMPA has been applied to NeuroMMSigDB mechanistic subgraphs in order to investigate which mechanisms are perturbed in different patient subgroups or brain regions. Additionally, clustering experiments were done and the individual resulting clusters were tested for their association with mechanism hypotheses encoded in BEL graphs (coming from NeuroMMSigDB).

Finally, in dedicated wet-lab experiments, partner PHARMACOIDEA demonstrated the potential of Syndecans to mediate uptake and spreading of misfolded proteins in the brain. The experimental validation of the Syndecan mechanism was worked out so well that the results could be published in Nature Scientific Reports.

Several web-recorded webinars and several publications had been generated to raise awareness for our new approaches and results.

WP4 – Ethical and Legal Governance

WP4 was led by Leibniz University Hannover (LUH), which could draw on its long experience in dealing with the ethical and legal aspects of health informatics, and particularly the European data protection framework. LUH set up and liaised with the Project Legal and Ethics Advisory Board (LEAB), organizing annual meetings in which the LEAB offered added expertise on areas affecting the Project (and also the PRECISESADS Project, see below). Valuable practical input was provided also by partner UCB, in terms of the practical and industrial aspects of medical research. A further key participant in the Work package was Alzheimer Europe (AE), with its patient-focused perspective, which provided valuable

insights into issues of significance for the relevant research study participants and the optimizing of mechanisms for safeguarding their autonomy and dignity.

Following the development of the Project Data Protection Framework in the first year, LUH assisted in implementing the framework's contractual approach (consisting of a series of data sharing agreements covering the diverse Project data flows in project). It liaised with the signatory partners throughout the signing process, and subsequently monitored and approved relevant data transfers under the framework through UL 'ticketing system'. LUH updated this framework in the last year of the Project to ensure compliance with the new EU data protection regime, following the entry into force of the new General Data Protection Regulation (GDPR) 2016/629. For this purpose, it devised and circulated a supplemental data protection agreement for the partners to sign that updated their obligations under the original Project data-sharing framework in light of the new legal position.

A further key task saw WP4 cooperate on legal and ethical matters with AETIONOMY's IMI sister-project, PRECISESADS (which did not have an ethics/legal work package of its own). This collaboration proved informative and useful in terms of comparing approaches to similar issues and advising on challenges (such as the implications of the GDPR rules), and involved frequent communication, as well as joint legal and ethical workshops and other face-to-face meetings, including PRECISESADS participation at two of the AETIONOMY LEAB meetings. In addition, LUH together with AE carried out a detailed legal and ethical assessment of the implications of the prospective clinical study in WP5 with its aim of stratifying patients into distinct AD and PD patient subgroups. This considered the terms of the patient information sheets and ICFs, various risks and benefits to subject interests from study participation, as well as wider societal implications of such studies and the trend to personalized medicine. It also included best practice guidance on research involving vulnerable participant groups, based on AE's extensive experience, as well as considering issues arising from collection and analysis of genetic materials, as regards information feedback and ongoing data/sample stewardship obligations

A key benefit for the WP4 partners from participating in the Project consists in the further experience acquired in managing data protection law and ethical challenges in the context of a scientifically valuable and innovative data-driven medical research project. This involved addressing practical challenges of balancing and safeguarding various interests, including through the data protection framework developed for the Project and, as noted, the need to update the framework due to changes in the external EU data protection law. The analysis made of the EU legal changes (as set out in the two T4.5 Deliverables towards the close of the Project may hopefully also prove an enduring asset that may be of use in similar further projects, as a guide for researchers in achieving legally and ethically compliant data processing.

WP5 – Clinical Validation

Preamble: WP5 aimed at validating the mechanism-based stratification of patients, by testing in real clinical samples the hypotheses underlying pathogenic mechanisms identified with WP3.

To this end, a multicentre European cohort of patients was implemented, with the help of the European Clinical Research Infrastructure Network (ECRIN). The unique AETIONOMY cross-sectional

clinical study (AETIONOMY-CS) built during the 5-year IMI project gave the opportunity to collect clinical data, biological samples (blood, DNA, CSF and skin biopsies) and brain imaging data necessary for the validation of the taxonomy provided by WP3 partners, from more than 400 PD patients (idiopathic or genetic), subjects at risk for PD and matched healthy controls (Figure 12 and 13 below).

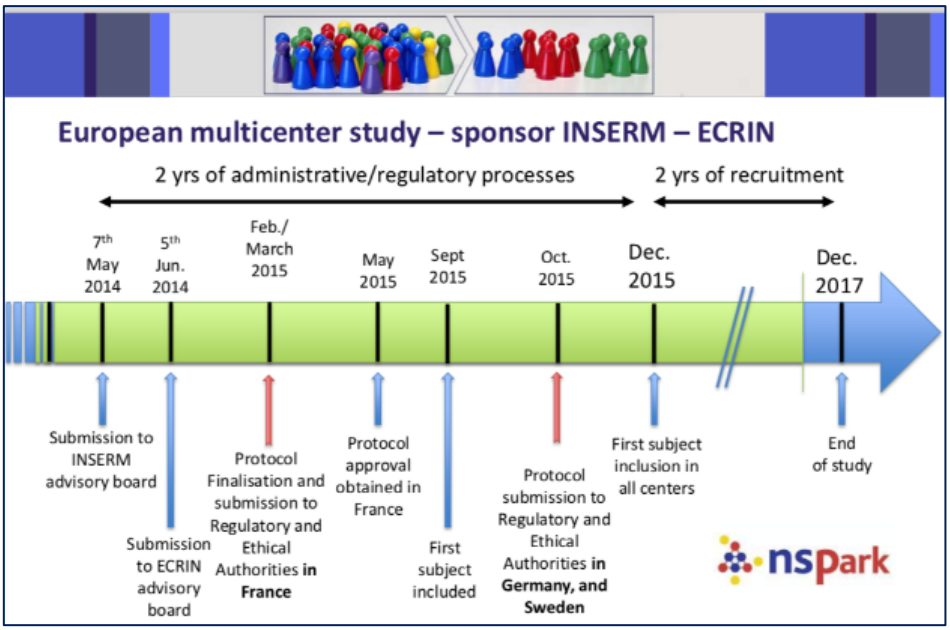


Figure 12: AETIONOMY Clinical Study regulatory flow chart

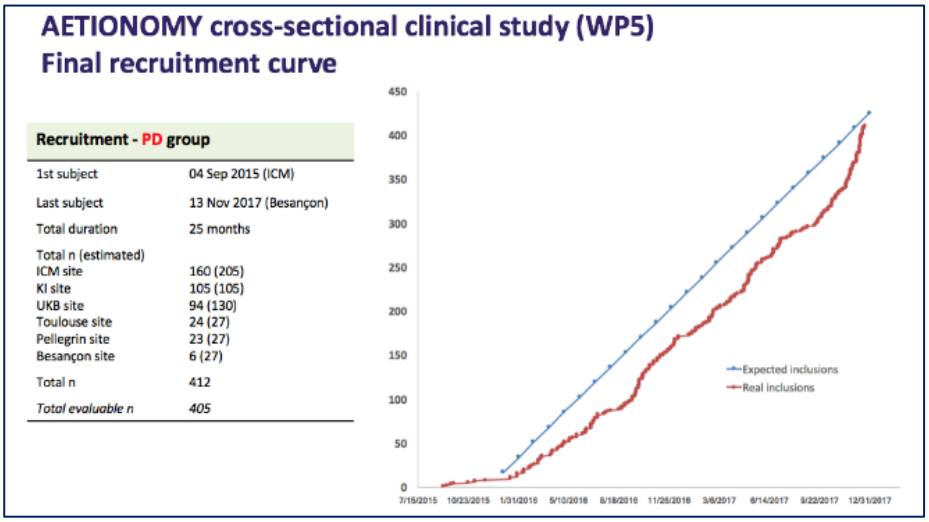


Figure 13: AETIONOMY Clinical Study recruitment. Left: PD (Parkinson's disease) group global recruitment information and actual versus estimated recruitment per clinical site. Right: Final recruitment curve.

Data management of the AETIONOMY-CS dataset was achieved with the collaboration of UCB, demonstrating the engagement of EFPIA partners to our project. UCB also provided a statistical description of the cohort which is planned to be published early 2019.

Additionally, thanks to the partners contributions (WP5, UCB), several external cohorts fitting with our criteria were also identified and accessed leading to an important number of data, samples and brain images, supporting testing and replication the mechanism-based taxonomy, both in PD and AD groups, through a panel of selected biomarkers of mechanisms (see Table 2 below).

		PD						AD		
Origin	**Total n**	**DNA**	**CSF**	**Plasma**	**Serum**	**Fibroblasts**	**MRI**	**DNA**	**CSF**	**MRI**
AETIONOMY-CS (PD group)	405	396	99	391	391	160	30			
External cohorts	1556	645	84	230			14	436	380	23
TOTAL samples		**1041** [0,1,4]	**183** [0,2,3,6]	**621** [0,3,5]	**391** [0]	**160** [0]	**44** [0,7]	**436** [0,1]	**380** [0,2,6]	**23** [0,7]

Available associated data:
[0]Clinical Data
[1]Genomics (ICM)
[2]Inflammatory (IDIBAPS/UKB)
[3]Proteomics (KI)
[4]Methylation (UKB)
[5]Cholesterol (UKB)
[6]IRS (SARD)
[7]Brain imaging (BBRC)

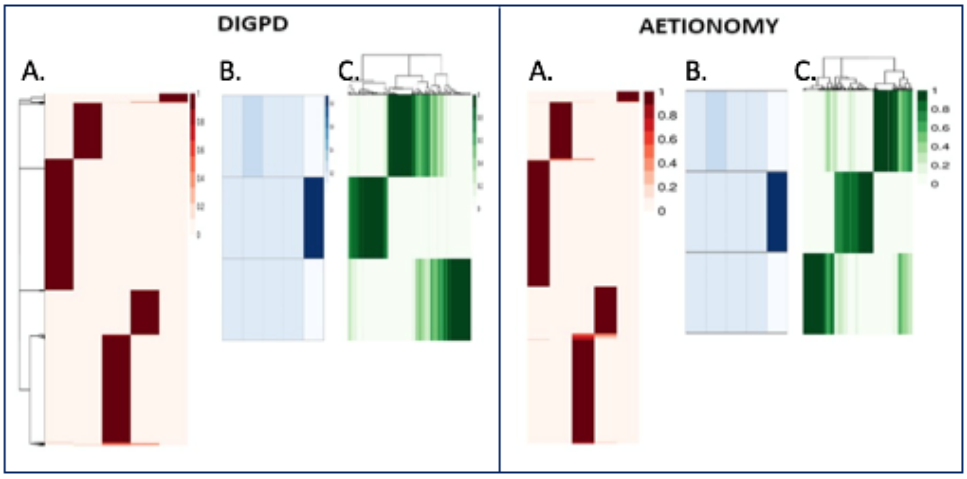
Table 2: WP5 Overview of AETIONOMY available samples and data. The table details the actual numbers of samples collected during the AETIONOMY cross-sectional clinical study (AETIONOMY-CS), as well as the samples from external cohorts (Consortium's contribution). Availability of clinical data and results of WP5 investigations are references in red (0 and 1 to 7, respectively).

Genomics (ICM, Jean-Christophe Corvol)

Pursuing the goal of validating the mechanism-based taxonomy of neurodegenerative diseases, the ICM proposed an evaluation of a genomic-based classification of the PD patients.
Five mechanisms were selected, among which 3 were known hypotheses (synuclein methylation, neuroinflammation, and mitochondrial dysfunction) and two were novel mechanisms found by WP3 partners with unbiased methods (insulin pathway, and stress-induced comorbidity). A list of genes was then selected from these pathways, and functional variants (based on their functional Cadd score) were filtered for inclusion into the model.

We then genotyped all the samples of interest (AETIONOMY-CS and additional WP5 cohorts, approximate total N individuals=1,400) and secondly applied an unsupervised method of clustering (NMTF for Non-Negative Tri- Factorization) on the available genetic data, to cluster the individuals in regard to the five WP5-selected mechanisms.

Two independent PD cohorts were analyzed: DIGPD as a discovery cohort and AETIONOMY-CS as a replication cohort. The set of variants selected in DIGPD cohort was thus applied to AETIONOMY-CS cohort.
Our model uses three matrices, presented in the Figure 14 below, for both tested datasets: the matrix of variants (A), the matrix of patients (C), and the connection matrix between both types of clusters (B).



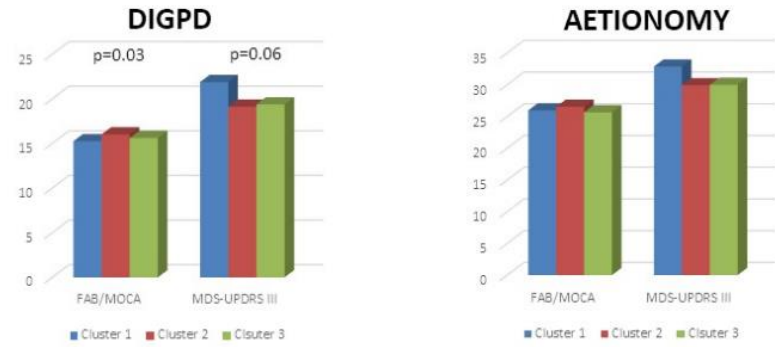
Genomics – Figure 14: NMTF clustering application in AETIONOMY project

Prior knowledge on the association of the variants to the WP5-selected mechanisms was used as a constraint for the variants clustering. The final number of clusters was defined using exploratory parameters, leading to a consensus of three clusters of patients.

Based on the genetic data decomposition, we observed one cluster linked to the "Stress Induced Comorbidity" mechanism, while the two others were randomly distributed. We could also note that *PARKIN* mutated patients of the AETIONOMY-CS were found in the same cluster.

Looking at the Movement Disorder Society Unified Parkinson Disease Rating Scale (MDS-UPDRS) and the Montreal Cognitive Asessment (MoCA) , a proxy for the Frontal Assessment Battery (FAB) which was not assessed in AETIONOMY-CS, we observed the same profile in the cohort 'Drug Interaction With Genes in Parkinson's (DIGPD)' and AETIONOMY-CS, even if those variables were not significantly associated with the clusters in AETIONOMY-CS (Figure 15 below).

	DIGPD (n=407)				AETIONOMY (n=224)			
	Cluster 1	Cluster 2	Cluster 3	P-value	Cluster 1	Cluster 2	Cluster 3	P-value
N	130 (32)	145 (36)	132 (32)		81 (36)	63 (28)	80 (35)	
Age	62.59 ± 10.64	61.8 ± 10.17	62.62 ± 8.63	0.79	63.52 ± 7.88	65 ± 7.57	63.89 ± 10.02	0.45
Female (%)	52 (40%)	55 (37.93%)	58 (43.94%)	0.59	25 (30.86%)	17 (26.98%)	26 (32.5%)	0.77
FAB/MOCA	**15.25 ± 2.66**	**16.02 ± 2.15**	**15.62 ± 2.19**	**0.03**	**25.95 ± 3.09**	**26.53 ± 2.58**	**25.65 ± 3.04**	**0.29**
MMSE	28.2 ± 1.82	28.3 ± 1.84	28.08 ± 2.01	0.68	28.28 ± 1.73	28.72 ± 1.44	28.25 ± 1.86	0.25
Non motor	7.41 ± 4.66	7.66 ± 4.8	7.24 ± 4.35	0.87	10.39 ± 4.47	8.4 ± 4.77	8.49 ± 4.55	0.01
MDS-UPDRS I	8.27 ± 5.07	7.87 ± 4.94	7.82 ± 4.31	0.81	9.04 ± 5.12	8.35 ± 4.67	8.24 ± 4.9	0.54
MDS-UPDRS II	8 ± 4.75	7.08 ± 4.47	7.68 ± 4.67	0.3	9.53 ± 5.68	8.5 ± 5.03	9.21 ± 5.45	0.52
MDS-UPDRS III	**21.88 ± 10.85**	**19.12 ± 9.2**	**19.38 ± 10.76**	**0.06**	**32.85 ± 14.81**	**29.91 ± 16.07**	**29.97 ± 12.71**	**0.29**
MDS-UPDRS IV	0.7 ± 1.99	0.49 ± 1.33	0.51 ± 1.62	0.79	1.54 ± 2.82	1.18 ± 2.33	1.13 ± 2.3	0.81
Anxiety/depression	4.77 ± 3.11	4.63 ± 3.35	4.51 ± 3.03	0.77	4.96 ± 3.97	3.79 ± 3.05	3.42 ± 3.3	0.03



Genomics – Figure 15: Clinical overview of DIGPD and AETIONOMY-CS clusters. Preliminary observations in clusters: Cluster 1 (blue) seems to present a higher motor severity, whereas the cluster 2 (red) has a higher cognitive impairment.

Epigenetics (UKB, Ullrich Wuellner)

DNA-methylation was chosen among the mechanisms to be tested in order to create a mechanism-based taxonomy of neurodegenerative diseases, as several lines of evidence pointed to the involvement of epigenetic mechanisms in neurodegenerative disorders. In the initial investigations in existing cohorts, we first identified differential methylation of the alpha – synuclein (SNCA) gene and second also differential methylation of other genes in Parkinson's disease (PD) patients, i.e. phosphodiesterase (PDE) 4 and Meteorin, Glial Cell Differentiation Regulator-Like (METRNL). SNCA is prominently involved in PD, as the expressed protein is the main component of intraneuronal protein aggregates (Lewy bodies) and was initially identified as the non-amyloid component (NAC) of plaques in Alzheimer's disease (AD) . PDEs have 3',5'-cyclic-AMP phosphodiesterase activity and degrade cAMP, which acts as a signal transduction molecule. PDE regulates the cellular levels, localization and duration of action of these second messengers (cAMP). METRNL is expressed in undifferentiated neural progenitors and in the astrocyte lineage, including radial glia and plays important roles in both glial cell differentiation and axonal network formation during neurogenesis. Upon completion of the AETIONOMY cross sectional study, the changes of DNA methylation in these genes, i.e. SNCA, PDE4D and METRNL have been investigated in greater detail with regard to available meta data, in particular tobacco smoking and coffee consumption. Apparently, the additional information of exposure and gender improved the accuracy of the observed changes which might suffice as a biomarker for a subgroup of PD patients. While a meta-classifier of the investigated genes alone achieved a specificity

of ~ 0.65 for men with a sensitivity of ~ 0.55; with the additional information the specificity rose to ~ 0.9 and a sensitivity of ~ 0.9 for current smokers among the male PD patients. We will continue to develop this potential biomarker further in the future.

Neuroinflammation (UKB, Michael Heneka)

The group of Michael Heneka at UKB focused on analysis of inflammatory / immune biomarkers in cerebrospinal fluid (CSF) as readouts of neuroinflammatory mechanisms in AD and PD. The lab analyzed 399 CSF samples from a local UKB biobank using a panel of 15 reliably detectable markers and in the 227 samples from the AETIONOMY cross-sectional study and external WP5 cohorts (AD from IDIBAPS and PD from ICM) with a larger panel of 20 markers. The studies underlined the importance of confirming detectability of inflammatory biomarkers in CSF. They furthermore reproducibly showed stronger associations of immune markers with molecular pathology hallmarks, such as amyloid and tau, than with clinical diagnosis alone. At first glance, the discriminative power of the immune markers appeared to be limited. Yet, further analysis revealed age-dependent trajectories of immune markers that add up with pathological effects. Other covariates, such as sex or genotype, furthermore, had significant influence on some markers such as complement factors. In conclusion, these covariates need to be stratified to unlock the potential of immune biomarkers for diagnostic or treatment readout purposes.

Astroglial inflammation (IDIBAPS, Raquel Sanchez-Valle)

CSF YKL-40, a marker of astroglial inflammation, was chosen among the mechanisms to be tested in order to create a mechanism-based taxonomy of neurodegenerative diseases. In addition, AD biomarkers characterization was considered crucial to perform different analysis related to the WP5 objectives. IDIBAPS analyzed YKL-40, Aβ1-42, total tau, phospho tau levels using commercial ELISA kits in 502 samples from 6 centers (WP5 partners). These samples included samples from Alzheimer's disease patients, mild cognitive impairment, Parkinson disease patients and healthy controls. Raw data have been uploaded to the AETIONOMY knowledge database and are available for further analysis. The data analysis of the YKL40 results was performed with other inflammatory/ immune biomarkers at UKB. Within this study, YKL-40 was not influenced by pre-analytical protocol, APOE genotype or sex, while age was a significant modifier of YKL-40 levels. YKL-40 was associated with clinical diagnosis, pathological biomarkers, and in Idiopathic Parkinson's disease patients with the Hoehn & Yahr PD progression staging. YKL-40 levels were associated to tau isoforms levels.

Noteworthy, when plotted against age, YKL-40 was on lower trajectory in amyloid positive in both the AETIONOMY/ICEBERG/IDIBAPS as well as the UKB dataset, but on (more pronounced) higher trajectory in tau positive. This could be in line with previous reports of potential bi-directional regulation of YKL-40 throughout disease, though more research will be necessary for validation. When we used YKL-40 as a marker of neurodegeneration in the AD continuum subgroup (A+T+) 53% subjects were classified as N + using a pre-established cut-off. AD and YKL-40 levels were also used to investigate the biomarker-neuroimaging correlations and influence of CSF biomarkers in neuroimaging changes in the IDIBAPS retrospective cohort in collaboration with BBRC (main findings are described below in BBRC section).

Proteomics (KI, Per Svennigsson)

The Stockholm site has performed antibody suspension bead arrays on wet biomarker samples obtained from the patient cohorts (AETIONOMY cross-sectional study and external cohorts from the consortium). The protein array used is a single binder assay based on direct biotinylation labeling of proteins and detection with antibodies on color coded magnetic beads. It is set up for multiplexing in two dimensions (384x384). It provides measures of relative protein levels and is a very powerful screening and discovery platform.

Quality check has been performed from the data of the protein array which covered 824 CSF and plasma samples from patients with Parkinson´s disease (PD), Alzheimer´s disease (AD) and controls. Data have been generated from 216 proteins, which were covered by 307 antibodies. In particular, the 38 proteins, covered by 66 antibodies, that specifically examine AETIONOMY-based mechanisms have been investigated. In addition, some proteins were also different between PD patients and controls (Figure 16 below).

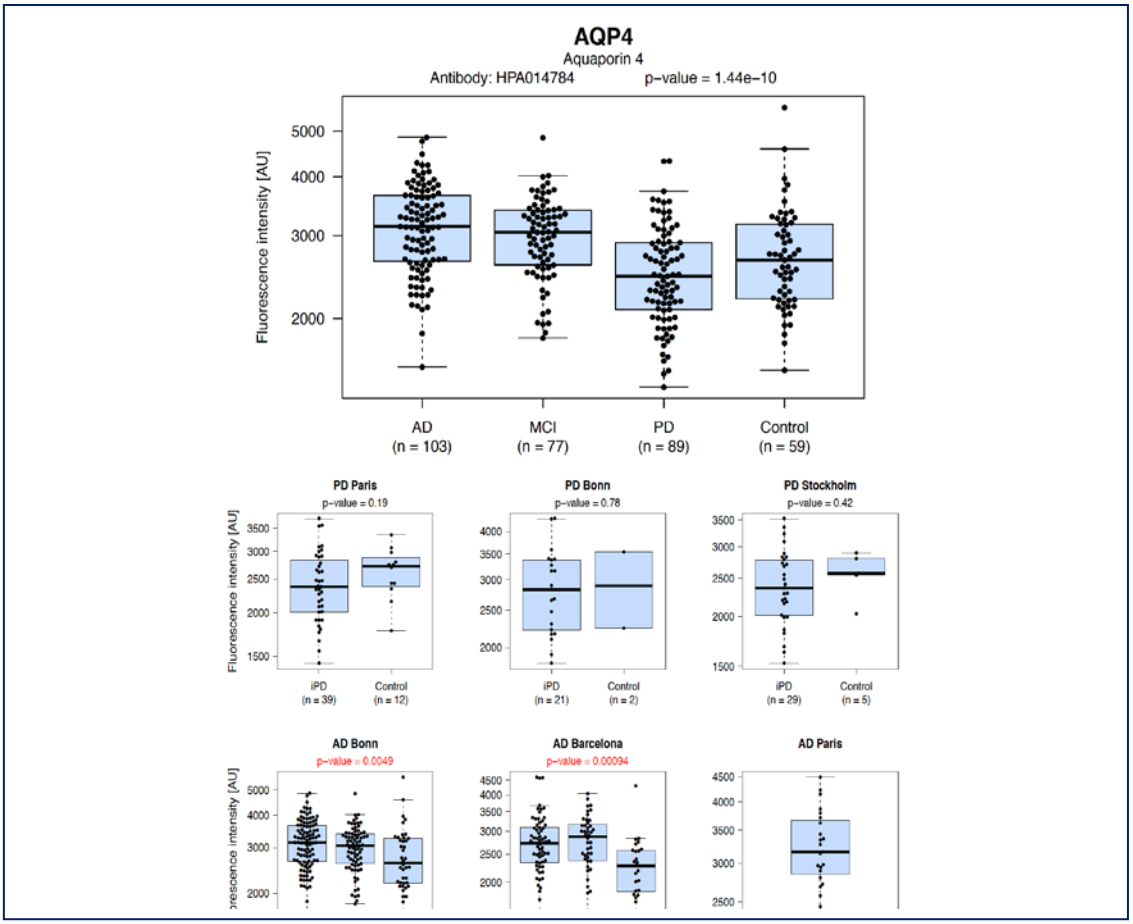


Figure 16: changes of aquaporin 4 in the different cohorts. Note the highly significant upregulation in MCI and AD patients

One postdoc, Dr Wojciech Paslawski and one PhD student, Sofia Bergström, have been heavily engaged in the project from the KI side. Dr Paslawski has spent time selecting antibodies to relevant targets for the large antibody array. as well as validating and analyzing findings from the antibody array experiments. Ms Sofia Bergström has successfully executed the antibody array and analyzed data generated from it. Both of them are now writing scientific papers from the experiments.

Briefly, to conclude, CSF levelS of several different proteins from AD cohorts were found different from controls (Figure 16). In addition, some proteins also showed level differences between PD patients and controls: Vgf, scg2, aplp1, clstn1, cntnap4, basp1, sparcl1, cck, hsp90b1, megf10. Among the AETIONOMY-based mechanisms, we have found correlations between changes in the insulin/diabetes-related pathway markers, DPP-4 and GIP, and cognitive status in the PD patients.

Brain Imaging (BBRC, José Luis Molinuevo)

BBRC has demonstrated that age-corrected YKL-40 CSF levels were increased in prodromal AD versus preclinical and dementia due to AD, and showed an inverse u-shaped association with p-tau values. A similar nonlinear relationship was found between grey matter volume and YKL-40 in inferior and lateral temporal regions spreading to the supramarginal gyrus, insula, inferior frontal cortex and cerebellum in MCI and AD. These findings for YKL-40 were unchanged after adjusting for p-tau, which was found to be associated with grey matter volumes in distinct anatomic areas. Taking together, these

observations reveal that CSF YKL-40, a biomarker of glial activation, is associated with a cerebral structural signature distinct from that related to p-tau neurodegeneration at the earliest stages of cognitive decline due to AD.

The subsequent goal was the stratification of our available cohorts as a function of YKL-40 measures (cluster high vs low YKL-40) and the identification of specific MRI signatures or cognitive differences (cross-sectional, dichotomic analysis), with the initial hypothesis that YKL-40 levels stratify the AD sample as measured by cognition.

BBRC finally performed a mechanism-focused analysis of brain imaging features using data from two populations with distinct pathologies, Alzheimer's Disease (AD) and Parkinson's Disease (PD), from external cohorts provided by WP5 (INSIGHT for the AD group and ICEBERG for the PD group, ICM). These data included Magnetic Resonance Imaging (MRI, diffusion-weighted (DW) and T1-weighted sequence) data, inflammation (YKL40) and core pathological markers (ABeta42 and t-Tau) levels in CSF, from N=65 subjects. The partners analysed parameters from diffusion-weighted images and looked for possible common associations between diffusion metrics and CSF markers across the two subject groups, as a potential reflection of shared neurobiological mechanisms independent from the pathology. Voxel-wise associations were assessed using a voxel by voxel permutation nonparametric test (1000 permutations) with threshold-free cluster enhancement. BBRC performed voxel-wise multiple linear regression analysis using various models: 1) a first model including level of Abeta42 in CSF, plus covariates such as age, sex and the study group; 2) a second model including level of YKL40 in CSF, plus covariates such as age, sex and the study group; 3) a third model including level of t-Tau in CSF, plus covariates such as age, sex and the study group; 4) a fourth version was tested modelling the study group as an interaction factor with the level of YKL40.

After performing inter-subject registration on parametric maps, mean values were extracted from a set of regions of interest taken from the John Hopkins University tract-based atlas of the white matter (WM). Diffusion Weighted Imaging (DWI) metrics were compared over tracts to assess possible differences between study groups (AD and PD) and a General Linear Model (GLM) analysis on these ROI values to look for associations between levels of YKL40 and diffusion metrics at the tract level was performed. Levels of YKL40 were dichotomized into high and low levels taking the group median as cut off value. We also looked-for associations between CSF levels of t Tau and diffusion metrics at the tract level. Levels of YKL40 were dichotomized and the cut off threshold was set at 242 pg/mL.

Results showed significant differences ($p<0.05$) between the two study groups. Subjects from the AD group displayed reduced fractional anisotropy (FA), and increased mean diffusivity (MD), axial diffusivity (AxD) and radial diffusivity (RD) in extended regions of the white matter, as compared to the PD group.

A main effect of age was observed on the 4 DW parameters in the expected direction (Model #1). As age increases, MD, AxD and RD also increase in extended regions of the WM. Accordingly, FA decreases, and a small cluster was found significant.

A main effect of ABeta42 was also found in all the 4-diffusion metrics. Significant positive association was observed between FA and levels of Abeta42 in a small posterior part of the left superior

longitudinal fasciculus, with the two study groups taken together. A negative association between MD, AxD and RD in extended bilateral regions of the white matter.

With the second model including levels of YKL40, main effect of age was also found significant in the exact same direction as with model #1. Association with FA was mostly found in bilateral anterior corona radiata and left cingulate gyrus. Association with MD, AxD and RD were found in extended bilateral regions of the white matter.

Main effect of levels of YKL40 were not found significant in all the four parametric maps.

Voxel-wise tract-based analysis of the four parametric maps with Model #3 returned no significant effect of t-Tau on any of the two study groups. A positive association was found between CSF levels of t-Tau and axial diffusivity in two tracts.

Amyloid deposition confers an additional burden to observed age-related changes in WM microstructure, regardless of the underlying neuropathological process.
Neuroinflammation as measured by high levels of YKL40 is captured by diffusion parameters in a set of WM tracts across the two patient groups, independently of the pathology, implying that it may be capturing a neurobiological process potentially common to both diseases.
Finally, high levels of t Tau were found locally associated with increased axial diffusivity, both reflecting potential axonal loss.

Insulin resistance pathway (SARD, Luc Canard)

The initial planned work was to deliver data on plasma neuroexosomes and CSF but the currently existing technique for neuroexosome preparation from plasma did not perform well enough a) to deliver robust data, b) to ensure the central nature of the plasma neuroexosomes that were collected. Instead of pursuing on the path of this challenging biological matrix, we only measured biomarkers in CSF samples, which were more representative of the brain functions.

Sanofi has contributed to the wet-lab biomarker analysis in WP5 by investigating the insulin resistance pathway, one of the two mechanisms selected with unbiased methods by WP3.

The work occurred in two phases, the preliminary and then confirmatory studies – described below.

1) A preliminary study

The preliminary study was performed on 25 patients (see characteristics in Table 3) and consisted in the measurement of the Biomarkers listed in Table 4.

List of samples

iPD : idiopathic PD n=10
FPD : familial PD
AR-PD : at risk PD
HC-PD : healthy controls matchés pour PD n=10
PAD : prodromal AD n=1
AR-AD: at risk AD n=4
HC-AD: healthy control matchés pour AD

Box name: LCR-AETIONOMY
Type of samples: CSF (stored at -80°C)
Recipient: Dr Nathalie SCHUSSLER
Date: 04/06/2018
Date de réception Sanofi: 07/06/2018

Position in the box	Individual Original Code				AETIONOMY ID (for ICEBERG patient)	Group (1=iPD, 2=FPD, 3=AR-PD, 4=HC-PD, 5=PAD, 6=AR-AD, 7=HC-AD)	Type of sample	Sampling date	Sampling time	Processing-Freezing date	Freezing time	Volume (µl)	Visit Number	Comments	N° Sanofi	N° Sanofi issus de la randomisation
1	ICEBERG	G	C	018	01-099-HC-PD	4	LCR	03/04/2015	09:10	03/04/2015	09:40	1000	V0		01-099-HC-PD	18
2	ICEBERG	G	C	018	01-099-HC-PD	4	LCR	03/04/2015	09:10	03/04/2015	09:40	1000	V0			
3	ICEBERG	W	R	030	01-002-IPD	1	LCR	12/06/2015	09:10	12/06/2015	09:40	1000	V0		01-002-IPD	7
4	ICEBERG	W	R	030	01-002-IPD	1	LCR	12/06/2015	09:10	12/06/2015	09:40	1000	V0			
5	ICEBERG	B	J	045	01-008-IPD	1	LCR	02/10/2015	10:08	02/10/2015	10:36	1500	V0		01-008-IPD	2
6	ICEBERG	A	J	046	01-006-IPD	1	LCR	07/10/2015	09:35	07/10/2015	10:25	1500	V0		01-006-IPD	16
7	ICEBERG	V	P	058	01-014-IPD	1	LCR	25/11/2015	09:28	25/11/2015	09:57	1500	V0		01-014-IPD	5
8	ICEBERG	M	D	071	01-017-IPD	1	LCR	22/01/2016	10:00	22/01/2016	10h36	1500	V0		01-017-IPD	1
9	ICEBERG	L	V	072	01-018-IPD	1	LCR	28/01/2016	12:44	28/01/2016	13:18	1500	V0		01-018-IPD	21
10	ICEBERG	C	D	083	01-023-IPD	1	LCR	04/03/2016	09:23	04/03/2016	09:51	1500	V0		01-023-IPD	15
11	ICEBERG	P	C	086	01-034-IPD	1	LCR	16/03/2016	09:08	16/03/2016	10:10	1500	V0		01-034-IPD	6
12	ICEBERG	F	J	095	01-032-IPD	1	LCR	13/04/2016	09:44	13/04/2016	10:14	1500	V0		01-032-IPD	10
13	ICEBERG	T	A	101	01-031-IPD	1	LCR	11/05/2016	09:37	11/05/2016	10:07	1500	V0		01-031-IPD	20
14	ICEBERG	H	C	106	01-042-HC-PD	4	LCR	03/06/2016	09:30	03/06/2016	09:52	1500	V0		01-042-HC-PD	19
15	ICEBERG	P	J	110	01-040-HC-PD	4	LCR	17/06/2016	10:12	17/06/2016	10:33	1500	V0		01-040-HC-PD	8
16	ICEBERG	C	G	150	01-090-HC-PD	4	LCR	17/02/2017	09:12	17/02/2017	09:22	1500	V0		01-090-HC-PD	23
17	ICEBERG	S	M	158	01-104-HC-PD	4	LCR	12/04/2017	11:05	12/04/2017	11:36	1500	V0		01-104-HC-PD	25
18	AETIONOMY	01	110	HC-PD	/	4	LCR	05/05/2017	12:38	05/05/2017	13:11	1500	/		01-110-HC-PD	11
19	ICEBERG	A	A	161	01-114-HC-PD	4	LCR	10/05/2017	10:33	10/05/2017	10:52	1500	V0		01-114-HC-PD	13
20	ICEBERG	L	M	166	01-122-HC-PD	4	LCR	02/06/2017	12:07	02/06/017	12:30	1500	V0		01-122-HC-PD	24
21	ICEBERG	S	I	172	01-137-HC-PD	4	LCR	21/07/2017	10:50	21/07/2017	11:28	1500	V0		01-137-HC-PD	12
22	ICEBERG	G	L	178	/	4	LCR	27/10/2017	09:42	27/10/2017	10:15	1500	V0		GL-178-HC-PD	3
23	AETIONOMY	02	001	PAD	/	5	LCR	06/06/2017	13:37	06/06/2017	15:22	500	/		02-001-PAD	17
24	AETIONOMY	02	001	PAD	/	5	LCR	06/06/2017	13:37	06/06/2017	15:22	500	/			
25	AETIONOMY	02	001	PAD	/	5	LCR	06/06/2017	13:37	06/06/2017	15:22	500	/			
26	AETIONOMY	02	007	AR-AD	/	6	LCR	12/01/2018	12:06	12/01/2018	12:45	500	/		02-007-AR-AD	22
27	AETIONOMY	02	007	AR-AD	/	6	LCR	12/01/2018	12:06	12/01/2018	12:45	500	/			
28	AETIONOMY	02	007	AR-AD	/	6	LCR	12/01/2018	12:06	12/01/2018	12:45	500	/			
29	AETIONOMY	02	006	AR-AD	/	6	LCR	26/01/2018	11:53	26/01/2018	12:45	500	/		02-006-AR-AD	4
30	AETIONOMY	02	006	AR-AD	/	6	LCR	26/01/2018	11:53	26/01/2018	12:45	500	/			
31	AETIONOMY	02	006	AR-AD	/	6	LCR	26/01/2018	11:53	26/01/2018	12:45	500	/			
32	AETIONOMY	02	005	AR-AD	/	6	LCR	29/01/2018	12:27	29/01/2018	13:09	500	/		02-005-AR-AD	9
33	AETIONOMY	02	005	AR-AD	/	6	LCR	29/01/2018	12:27	29/01/2018	13:09	500	/			
34	AETIONOMY	02	005	AR-AD	/	6	LCR	29/01/2018	12:27	29/01/2018	13:09	500	/			
35	AETIONOMY	02	003	AR-AD	/	6	LCR	02/02/2018	12:18	02/02/2018	12:54	500	/		02-003-AR-AD	14
36	AETIONOMY	02	003	AR-AD	/	6	LCR	02/02/2018	12:18	02/02/2018	12:54	500	/			
37	AETIONOMY	02	003	AR-AD	/	6	LCR	02/02/2018	12:18	02/02/2018	12:54	500	/			

1) Pool des aliquots lorsque plusieurs aliquots pour 1 échantillon
2) Aliquots de 110µl
3) dernier aliquot (si volume <110µl) avec un point au dessus du couvercle

Table 3: Data associated with the 25 samples received for the preliminary study

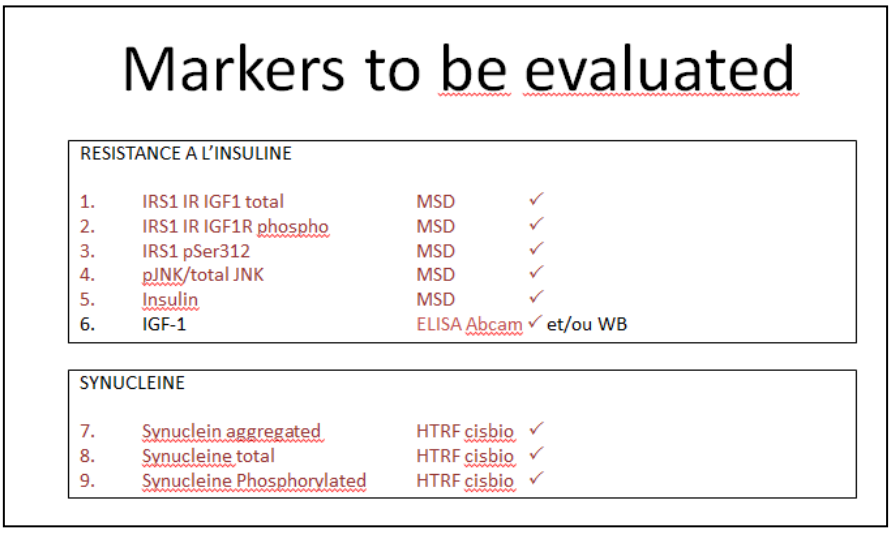


Table 4: List of the evaluated biomarkers

Results:

- On this small sample set, the only biomarker that showed consistency is the Insulin Receptor (see figure 17) for which results were confirmed in a technical replicate.
- In this small set we had HC-PD (10), iPD (10) prodromal AD (4) and AR-AD (1) and showed that we were able to measure IR total (figure 17), Insulin, IGF1R in the provided samples. IRS1 was only detectable in 2 out of the 25 samples provided. For Synuclein, both total, phopho Ser129 and aggregated synuclein were undetectable. The full table of results can be found below (Table 5)

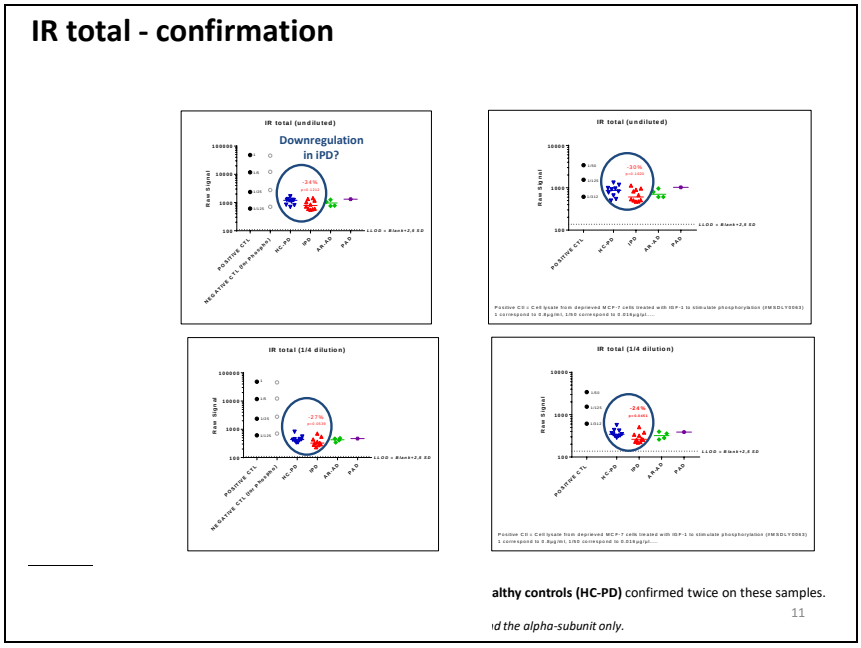


Figure 17: Insulin receptor measurement in the preliminary cohort

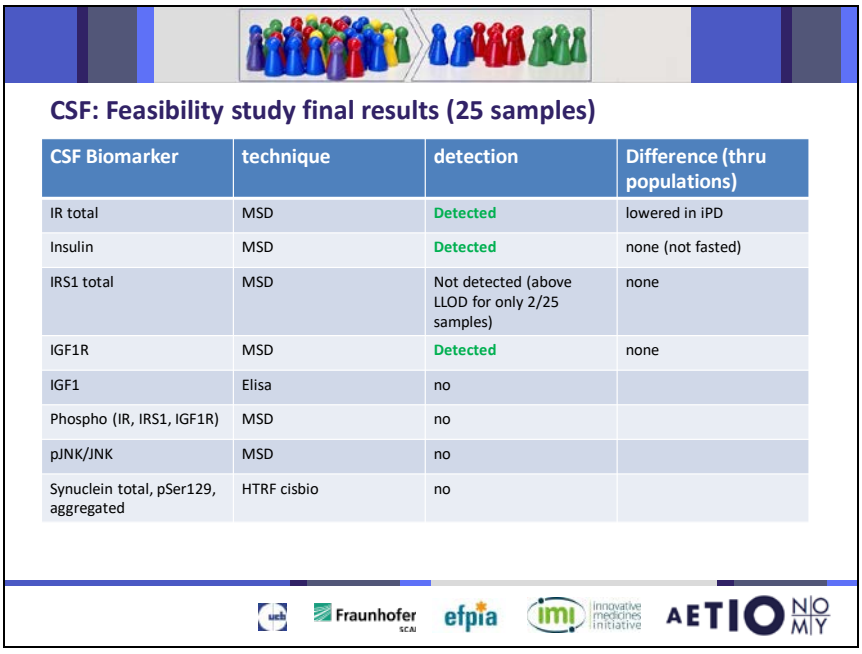


Table 5: Recapitulative table of measurements

2) Confirmation in the AETIONOMY cohort

In this second phase, the work focused on the determination of Insulin Receptor (IR), Insulin Like Growth Factor-1 Receptor (IGF-1R) and Insulin Receptor Substate-1 (IRS-1) levels (together with Insulin) by a multiplex immunoassay in Human Cerebrospinal Fluid samples from Alzheimer's and Parkinson's disease patients collected in the AETIONOMY clinical study. We received CSF from a total of 197 patients from three clinical centers: ICM (54), Karolinska Institute (34) and IDIBAPS (109).

Partner SANOFI performed an evaluation of the all biomarkers in the various populations. The figures that are provided below are showing data for the Insulin Receptor. In the following views, our colleague pooled HC-PD and HC-AD, as well as fPD and iPD and also pooled the sites (see figure 18) as indicated in the right part for the pooled measurements.

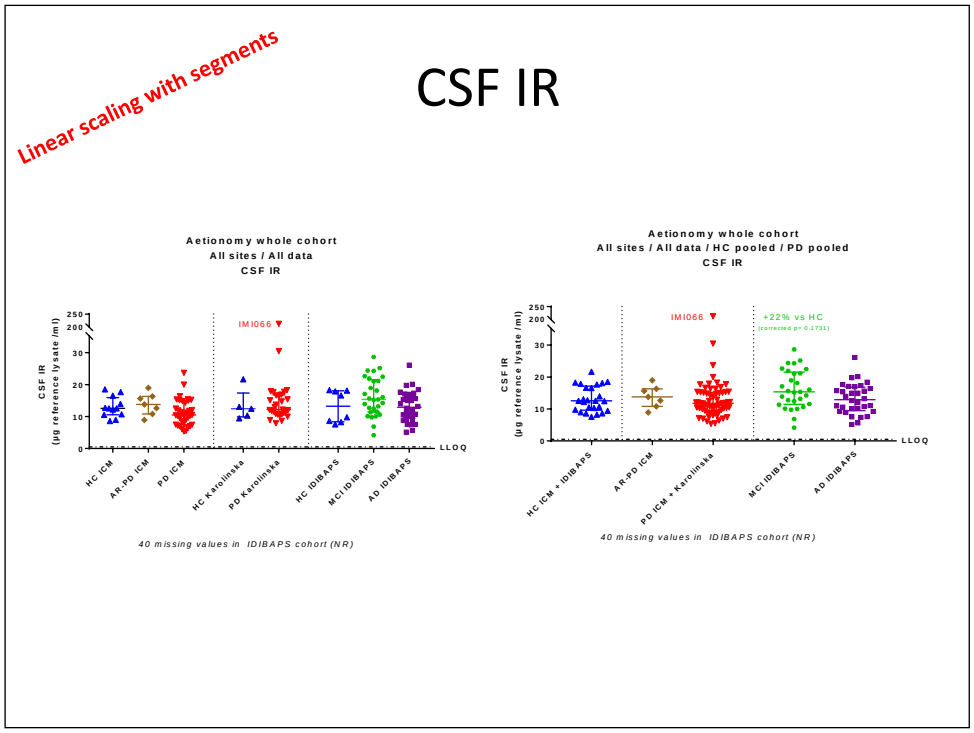


Figure 18: Measurements of IR

The associated descriptive statistics are shown in Figure 19.

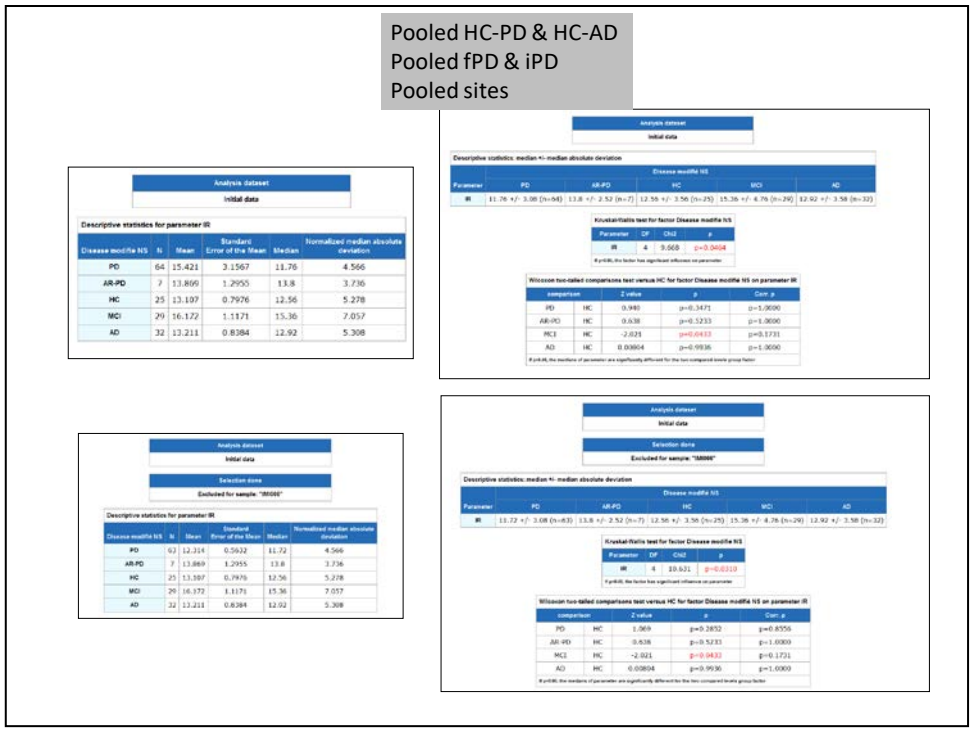


Figure 19: Descriptive statistics for Insulin Receptor (AETIONOMY cohort)

Results: In the AETIONOMY cohort, the IR was measurable and seemed to tend to be associated with the MCI disease condition (though corrected P value not adequate, +22%, p=0.1731).

Conclusions:

- It is very important to note that the results shown for the AETIONOMY cohort are partial and to be considered as purely indicative as, for the time being rigorous statistical review has not been applied.
- IR is a biomarker that can be measured and that is showing some level of association with a disease status.
- The resulting quality-controlled data were made available to the AETIONOMY consortium. This data could be integrated as an additional modality to the mechanism-based stratifications of patients.
- Additional Biomarkers that are linked to autophagy will be measured and made available to the consortium during 2019.

Conclusions from WP5

From a clinical perspective, AETIONOMY has been a pioneer in proposing a new classification of neurodegenerative diseases. So far, these diseases are defined according to their clinical representation, i.e. by their symptoms reflecting localization of the pathological process in the brain (e.g. hippocampus involvement leading to memory impairment, dopaminergic system leading to motor symptoms).

Stratifying diseases and patients according to the underlying molecular (and anatomical) mechanisms represents a revolution in the way we address therapeutic innovation in the field. AETIONOMY has thrust together the disparate worlds of clinicians, basic scientists and computer scientists to revisit the taxonomy of these diseases. To make this interaction efficient was a challenge because of conceptual and language barriers, but altogether AETIONOMY finally achieved this goal and we believe that our preliminary results will pave the way for future development towards more personalized and comprehensive treatment for neurodegenerative disease.

Collaboration with EFPIA partners was an important asset for the project, not only to drive our scientific goals towards applicable solution towards drug development, but also for their support in terms of scientific input and samples (statistical analysis particularly). Despite this commitment, there are still important issues for data sharing between academic and industrial partners that remains to be resolved.

One main achievement of our project was to show that neurodegenerative processes like neuroinflammation affect both PD and AD in a similar manner both at the molecular (neuroinflammatory markers in CSF) and at the anatomical (brain atrophy) levels. Another major result is the development of new tools to stratify populations of patients according to their mechanisms based on genomic information (the NMTF method). Of course, these preliminary results remain to be replicated in independent cohorts by others and validated biologically by identifying relevant biological markers for each mechanism.

Considering the high complexity of neurodegenerative processes, interactions with individual and environmental factors that affect both susceptibility of the disease and modifications of their progression, there is still substantial work to be done to transfer the AETIONOMY findings into clinical practice and drug development. Particularly, it has not been possible to address the longitudinal aspects of disease progression in the anatomical and the time spaces due to the constraints of the AETIONOMY program. Such efforts will be pursued in other projects in which consortium members are involved, supported either by IMI (EPAD) or other European funding mechanisms (VirtualBrainCloud).

1.6. Potential impact and main dissemination activities and exploitation of results

AETIONOMY was initiated to explore the idea that conventional disease definition is an increasingly outdated concept in the current medical environment. Opportunities for so-called ‘precision medicine’, in which molecular features both identify and direct treatment for disease, have been observed in a range of disorders, most prominently in the oncology field. The AETIONOMY consortium chose to seek molecular characteristics of Alzheimer’s disease (AD) and Parkinson’s disease (PD) that might contribute to a ‘taxonomy’ of these conditions, and help our community move towards a precision-medicine approach.

The project has developed innovative computational tools to manage and interpret the complex healthcare and research data environment. It was a Herculean task for the teams to clean, associate

and relate data that had been collated to answer diverse research questions over the years, and to make them available for our cause. Among the legacies of AETIONOMY are the resultant 'cleaned' data, managed in a knowledge base that will support future research.

The availability and sustainability of the AETIONOMY Knowledge Base (AKB) including the datasets, services and models will ensure the dissemination of the important results of the project to other researchers beyond the consortium. The related webinars and documentation will serve as a self-starter kit using which these approaches can be applied to other research domains or used for subsequent development in the Neurodegenerative Disease domain. The AKB will continued to be maintained for another 5 years and the services and disease models will be continued to be updated and hosted by relevant partners.

Another outcome - as a clinical contribution - is represented by the AETIONOMY cross-sectional study itself (AETIONOMY-CS): the delivery of this European study represents a unique prospective cross-sectional dataset, from individuals of multicenter origin, in the field of PD. The full dataset, which combines not only clinical data routinely collected during clinical trials but also a complete panel of OMICS data (genomic, epigenomic, proteomic, metabolomic, transcriptomic), has served as a validation dataset for the patients' stratification. This dataset has been implemented in the AKB and and provided to the Consortium and open to the scientific community in a secure environment for further projects. The generated AETIONOMY datasets are fully accessible to the community through the AKB (subject to a data access committee agreement). More information about methods, models, services and data can be found here: *https://data.aetionomy.scai.fraunhofer.de/*

The outcomes are still being explored and our findings will support new research directions for many years to come. Excitingly, the computational models appear to have identified groups of patients that differ significantly from each other. The differences can either be traced to molecular characteristics or describe different risk profiles for the development of neurodegenerative disorders. Both findings offer insights into future opportunities for precision-medicine approaches and increased hope in the search for treatments.

The proof of concept of the genomic-based stratification of PD patients is a step forward to future precision medicine clinical trials. However, the results of the AETIONOMY project published so far in the scientific literature reflect only some selected aspects and have not yet displayed the full, broad combination of deep clinical phenotyping on the one hand and genetic, epigenetic and biological/biochemical data on the other hand. The ongoing analyses of these datasets will provide a rather unique dataset which facilitates the connection of additional parameters of relevance, e.g. environmental factors, such as caffeine, nicotine and alcohol intake with PD and PD related symptoms, for epigenetic investigations.

A collaborative analysis of serum biochemistry together with CSF protein arrays and genetic and epigenetic markers will improve our understanding of the metabolic changes occurring in PD and the respective genetic basis.

Additionally, neuroinflammation is currently not considered in routine clinical diagnostics of AD or PD and its therapeutic potential is not fully explored. This project has identified several biomarkers of neuro-immunity / neuroinflammation that could serve as readouts for immune-based therapeutic

trials and help understanding the time-course of neuroinflammation in these disorders. This information may help to improve mechanistic understanding and diagnosis of neurodegenerative disorders and provide treatment options in the long term.

We have also generated robust protein profiling data, from samples from different countries and cohorts. There are many CSF changes in AD, but also potential candidates in PD are revealed by this initial analysis. Early explorations of the data provide further evidence for increased neuroinflammation and altered diabetes-related signaling in neurodegeneration along with some novel biomarker candidates. All these results were presented to the IMI/HBP community during the AETIONOMY Final Symposium. Presentations of the project outcomes and of related dementia research initiatives can be found under: *https://www.aetionomy.eu/en/events/aetionomy-final-symposium.html.*

Additionally, the consortium generated a long list of presentations and publications, many of which are listed on the project website. Finalized clinical analyses are planned to be published in 2019:

- A comprehensive description of the AETIONOMY cross-sectional study is currently being finalized to be published (ICM, UCB, WP5).
- Of the two generated datasets by UKB, one is already published and a second one is in preparation for publication.
- IDIBAPS is preparing a publication on longitudinal cerebral volumetric changes associated to CSF YKL-40 and sTREM2 levels in preclinical Alzheimer's disease.
- A manuscript from KI on the first results from the protein array is being compiled and will be published in 2019. A series of manuscripts, more closely related to different symptomatology's, are also being planned.
- A manuscript focusing on the methodology used to classify the patients with the genomic-based model will be submitted for publication in 2019 (ICM).

Finally, the AETIONOMY project had a transformative effect on our corner of the scientific community by bringing together researchers from different fields. This fruitful collaboration between physicians, basic scientists, bioinformaticians specialized in modeling and system biology, from academic and industrial partners helped reinforce the multidisciplinary demands of the era of "big data" and "artificial intelligence" and has implications for drug development and in medicine in general. We are confident that AETIONOMY will be considered as a pioneer in this emerging field which will have an important impact on the society in a near future.

As an important aspect of outreach of an IMI-project to another "flagship" project, the uptake of a large portion of the AETIONOMY semantic framework formalizing knowledge about neurodegenerative diseases in the Human Brain Project (HBP) in the course of the ROHAN project marks a significant crosstalk between two major scientific funding schemes in Europe. This is ongoing work, but the following achievements of this collaboration can already be reported:

- Construction of a dedicated epilepsy ontology that harmonizes the currently accessible, partial epilepsy ontologies is under way

- A referential semantic hub that can be used for terminology and ontology management (based on the OLS technology developed at the European Bioinformatics Institute (EBI)) has been developed. It will be deployed shortly as part of the medical informatics platform of HBP
- A dedicated literature mining services ("SCAIView Neuro") has been developed; after updating of terminologies and indices, this system will be rolled-out as a service to the entire HBP.

Moreover, a large portion of the resources and the lessons learned in AETIONOMY will be re-used in the VirtualBrainCloud (VBC) project; a H2020 project that aims at linking the world´s leading platform for brain simulation (The Virtual Brain, http://www.thevirtualbrain.org))) with the mechanistic approaches developed in the course of AETIONOMY. Efforts on the side of the AETIONOMY leadership to bridge between IMI projects and the FET flagship Human Brain Project and, namely the uptake of Viktor Jirsa and his team as a partner in AETIONOMY have resulted in this future activity. VBC will demonstrate the power of the AETIONOMY mechanism modeling and mining in the context of personalized brain simulation; with a special focus on neurodegeneration. Remarkably, VBC will deal with both, Alzheimer´s and Parkinson´s Disease and will continue the approach taken in AETIONOMY to tackle two of the most prominent neurodegenerative diseases in a simultaneous fashion.

1.7. Lessons learned and further opportunities for research

Although some of the project deliverables may have been achieved without the collaboration between partners, the joint effort increased the speed, quality and diversity of the delivery in most areas of the project – below we summarise a selection of contributions which emphasise the value of a PPP.

From the beginning, the coordination was a strong partnership between UCB and Fraunhofer. The leadership of Prof. McHale, who has a clinical background, greatly helped shape the research targets, and the close collaboration of the joint Project Office (PO) has efficiently driven the project throughout the duration.

The contribution of EFPIA partners can be considered as important for several aspects of the ICM led Clinical Research efforts. As the project evolved, the staffing and expertise of clinical data management, statistics and clinical research departments from UCB supported the project objectives. EFPIA partners participated in the design of the clinical study and provided important support for data management and analysis, and Novartis brought the work package leadership and coordination.

UCB was also able to support with the delivery of 419 participant samples that helped to supplement the human bio samples available to validate the hypotheses of the selected mechanisms. SARD was able to bring scientific input for the data generation and analysis of the biomarker measurement of central insulin resistance network and will deliver subsequent autophagy related biomarkers during 2019.

Collaborative projects like AETIONOMY require a standardized sampling effort to be provided by all the involved partners. It is well-known that sample handling is a critical point in wet biomarker studies, being a source of variation in multicentre studies. The AETIONOMY clinical study protocol and its instructions on sample handling were carefully reviewed by all centres before submission and the

external cohorts were selected fitting as much as possible with those instructions, before sharing. This is reflected, for instance in Proteomics assessments, by the similarity of the overall signal intensity in the different cohorts (Figure 20 below). We have, thus, established conditions which could be used in future collaborative efforts on wet biomarkers between the partners.

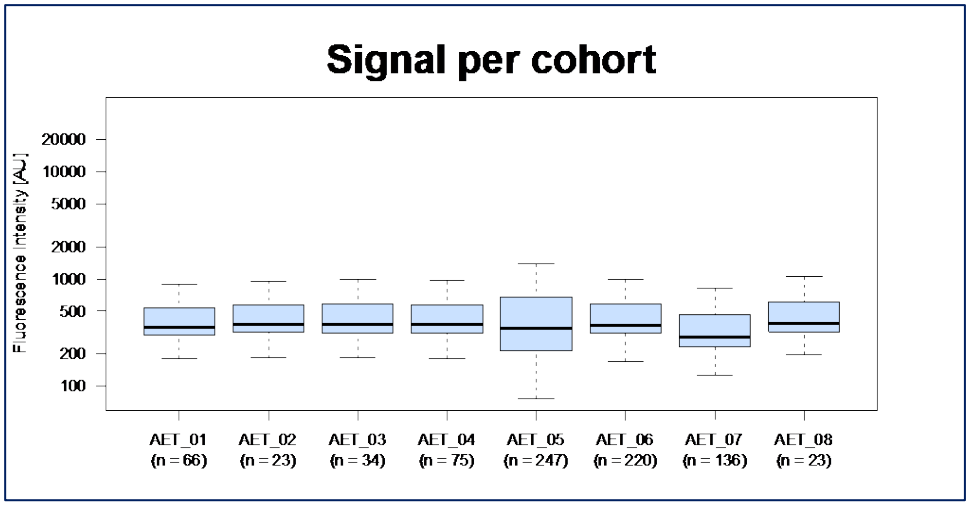


Figure 20: Total signal for all antibodies from each of the studied cohorts

Following Partner BI's portfolio changes and switch in scientific direction, the main role of BI in the project, in addition to the data provision mentioned above, was its partnership with LU and its provision of support to the data management activities related to setting up and improving the knowledge base, the decision to use tranSMART as the platform of the knowledge base and the methods for data incorporation, as well as to assessing the quality of data curation and assignments of relevant brain regions.

Under the co-leadership of Partner SARD and during the early phase of the project, there was significant engagement between UCB and KI in the development of the initial exploratory computational models to assess the potential for data stratification and in managing access to EFPIA datasets (including a collaboration between Data Transparency teams from all EFPIA partners). In the later phases, significant expertise/resource from UCB was leveraged to impact the analytical aspirations of the project and ultimately deliver against the key goal of the project.

These activities were all developed and managed in the true spirit of collaboration, with significant debate, co-development and resourcing from each of the partners engaged.

We have mentioned above some of the value that EFPIA partners brought to the project, but it would be remiss to ignore the value gained by these partners. In addition to the scientific value that a diverse population like this community brings during its myriad interactions and discussions, there will be transversal applications in each of the partner organisations from some of the intermediate deliveries of the project.

The AKB, a highly-curated knowledge base, is going to demonstrate its value as a unique resource in this domain and will support further hypothesis generation for the development of novel therapeutic concepts. The experience of building towards the AKB, and an appreciation of the effort involved, will

certainly deliver a lasting influence on EFPIA partner processes for information retrieval from literature, data management, data mining and disease modelling.

The technologies implemented in the knowledge base and analytical tools will be adapted for internal use within some of the EFPIA teams and spin off collaborations are already planned to develop these methods further.

The development of computational methodologies for the integration and interpretation of diverse data types will naturally have impact as we apply these in many other disease settings, in particular, the innovative perspectives on data use and reuse will have long reaching implications. The generation of models such as the VPC and the demonstration of the use of method like the NNMF approaches applied in AETIONOMY will translate into new strategies for the development of drugs in the era of precision medicine.

A range of markers associated with mechanisms identified in this project may also serve as candidates for such approaches, but further replication will naturally be needed in independent cohorts with samples obtained under strictly standardized conditions for full validation of these findings.

From your experience, please propose any recommendations/ solutions which could be useful for a PPP.

Data

The contribution of data by EFPIA partners was a non-trivial subject. During the early years of the project, significant investments of time from both industry and academic partners was consumed in addressing the critical issue of attempting to meet partner commitments described in the consortium agreement. This involved consideration of data sensitivity, security provision as well as legal and ethical challenges and standardisation of processes. Ultimately the result, despite it being an accepted standard practice for data transparency, provisioning of data in the SAS-sponsored environment made it almost impossible to utilize patient-level data for mechanism-validation as intended.

It is appreciated that in the second wave of IMI projects, better processes have been put in place to facilitate data sharing, but it is still not clear that this effort would be enough to support all forms of data modelling and joint analysis. We recommend subjecting data sets to rigorous pre-processing following best practices established in this project and to provide well-documented “merged” data sets, or indeed utilise the methods developed in WP2 (Virtual Dementia Cohort).

Major efforts in WP2 were spent on data harmonization and integration. While the availability of data can be dependent on several factors such as delays in the analysis, quality control and legal work from participating partners, a standard format for sharing of data could increase ease in adaptability and integration of platform. There are several formats available and researchers are gradually being aware of its importance and adapting them. However, it would be helpful if there are suggested or enforced formats and templates from the start of the project. Also, important along with proper formats for sharing are the accompanying data dictionary and metadata associated with generated data. During curation it is important to have knowledge of data generation or collection in order curate data accurately. This will reduce delays which arise from several back and forth communication between the data generator and processor.

Another important aspect that was developed in course of the project are the guidelines on documenting sharing of data among partners. In order to document the stages of data life cycle from initial generation of raw data until integration into the platform and availability of the harmonized data to the research community we adopted an open source ticketing system in LU to handle requests of inclusion and access to data after proper approval in terms of data sharing agreements. It would be helpful and easy if there is a recommended platform or tool available to projects from their initiation especially with data sharing among multiple partners.

In addition, we did face challenges with the intended collaboration with other IMI Projects when it came to the sharing of patient-level data. The strategy to leverage AD patient-level data collected in EPAD did not work out; neither did the “fall-back-solution” of using EMIF-1000 data and samples.

The lesson learned here is that clinical data being earmarked as being “open” or being “accessible” does not mean necessarily that you can use them automatically in a consortium context. We strongly suggest that a clear due diligence on the availability and accessibility (including informed consent) of patient-level data is performed ahead of the acceptance of partners in a consortium that are supposed to bring data into the project.

Regulatory Landscape

The changing regulatory landscape was a challenge for us with the implementation of the GDPR and some guidance from IMI to the projects would have been appreciated. Fortunately, in this project we had a Legal and Ethical Working Group that spent a lot of time reviewing the changes that were needed from a project perspective and set up joint workshops with the sister taxonomy project PRECISESADS to discuss these regulations and implement action plans, especially around the sustainability aspect.

As a recommendation, it is critical to support future PPP initiatives through building and centrally disseminating clear guidelines on the challenges, and mitigation mechanisms needed to ensure that partners have freedom to exchange data within projects like these. The central infrastructure and expertise that has been build up by institutions like Partner UL are critical to the success of any project where simple bilateral exchanges of information are insufficient. In projects like AETIONOMY where central, accessible and computationally available data are needed, the complexity of building these capabilities cannot be overestimated. These challenges are most evident in the need to make data available for complex computational tasks such as model building that might require integration of multiple sources of data.

We learned painfully, despite having some insight from the beginning, that data provision does not equal data accessibility and does not guarantee usability – within the project different stakeholders met these challenges resulting in time allocated to the provision of data management infrastructure, data cleansing and harmonisation and the development of tools where tool provision was lacking. Each of these steps delayed the abilities of the teams to generate insights and ultimately resulted in a significant proportion of the findings occurring in the last phase of the project or awaiting post project efforts.

In terms of regulatory compliance issues, particularly the accountability requirements in new EU General Data Protection Regulation for parties that process sensitive biomedical data, there is scope for IMI, and similar frameworks to better guide data management practices or indeed provide

mechanisms to support or simplify these. During the course of AETIONOMY, Work package 4 produced several documents that analysed these issues (especially in the T4.5 deliverables), and flagged project organisational implications. Now the new legislative regime is in force, these matters could be embodied in future Project or Grant Agreements (PA/GA), such as by fixing one or two partners (e.g. the project coordinator(s) and/or a party with an especially significant role in organizing and rationalizing project data processing flows), with the status of lead 'data controller'. By contrast, other partners, playing a more peripheral or limited role in project data processing operations, may have subordinate status as 'joint controllers' of a lesser degree, or indeed be seen as data processors under the GDPR, rather than (joint) controllers. The PA/GA could thereby serve to concretize the parties' own understanding of their roles and set out their expected contribution to the GDPR compliance framework (in terms of such matters as participating in Data Protection Impact Assessments, prior consultation with supervisory authorities, and establishing mechanisms for cooperation between partner Data Protection Officers).

EFPIA partner commitments

The change in strategy of BI and Novartis from the project objectives affected the leadership of WPs 2 & 5 despite efforts from both organisations to bring additional value. In both WPs, we hoped for more strategic input through the EFPIA partners.

In the case of Novartis, the expected sample contributions ultimately could not be made due to changes in data privacy laws, and unfortunately their involvement in other IMI-projects (e.g. EPAD) which offered an excellent opportunity to help build bridges between AETIONOMY and related IMI-projects, was not able to be leveraged.

The consortium could certainly have gained from both partners experience in clinical trial planning, statistics and trial data modelling, and as has been mentioned the contribution of EFPIA data sets, could have helped us substantially in the field of hypothesis validation, yet the form in which this was possible did not facilitate the ultimate goal of the project.

Public private partnership in a consortium requires long-term involvement over several years both from the academic and the EFPIA partners as well. The main lesson learned is that strategic changes of focus areas unexpectedly led to a reduced interest of senior management in this partnership. Staying involved in AETIONOMY and delivering the agreed upon share requires major internal adaptations to provide the continued support for the development of tools and biomarkers.

Management

Any project delivery is a result of leveraging successes and managing or mitigating challenges, for AETIONOMY the goal and approach was ambitious from the outset and the methods remained flexible until the closing days, and as such this balance was a central theme of the management efforts throughout.

There were many successes, such as the key intellectual collaboration between partners AMU, Fraunhofer and UCB, which lead to multiple breakthroughs in computational approaches; the collaboration between ICM, UCB and Fraunhofer to demonstrate clinical stratification and pointers to

the taxonomy; the clinical efforts of the entire clinical team to recruit, gain access to samples, and generate an extensive biomarker dataset; the considerable effort of the legal teams of all partners in navigating the complexities of data sharing (not once but twice to accommodate the advent of the GDPR era), and the technical developments and sustainability that Fraunhofer and LU built into the AKB. However, these did not come easily, and were the consequence of significant debate, influencing and reinforcing of multiple stakeholder's positions through the management of all the work packages as well as the Project Office.

Mitigating and managing challenges was an equally significant effort for work package leadership and the Project Office, whether it be the loss of partners such as NEURORAD or UCL, or the gain of partners to address shortfalls such as AMU and BBRC. A significant challenge for the project was the fact that, written over 6 years ago, the Description of Work sometimes felt like a millstone around the neck of the project teams. Timelines and plans that rapidly became untenable, despite amendments, needed to be adhered to and deliverables generated -while insights generated in the project sought different directions and outcomes. Again, the efforts of the management of the project were employed to good effect to ensure reporting compliance was balanced against the need for innovation in the project.

A key challenge, that was ultimately only partly resolved over the duration of the project, was the different world views of the participants, with clinical and computational worlds rarely fully coming into clear alignment. It remains a problem to address in the community at large as some commentators point out, despite the clinical domain becoming more data rich and the computational teams more clinically literate.

As a consortium, the decision-making process seemed overly complicated, especially when disciplinary actions or structural changes were needed. The time needed to recognize the causes, explore options, and importantly come to a consensus on the actions can be excessive, and this is time that a project even of a five-year duration can ill afford. This is further compounded by the diversity of understanding of detailed project objectives held by the consortium leadership, rendering a consensus much harder to achieve.

One interesting observation was the fact that a difference in the mechanism by which success is measured generated an interesting dynamic in the project during the final year and the race to the end product. Academic and clinical partners drove towards individual quality and publication-based outputs, while the EFPIA partners were able to leverage project management skillsets to support the deliveries which required coordinated actions and a clear understanding of collaborative dependencies. Again, the insight gained is the benefit that comes from the collaboration of diverse partners.

Large Consortium involving several academic and industrial partners necessitate a close monitoring of the activities, which is sometimes difficult to align with the necessary flexibility and plasticity of a research project like AETIONOMY, at the frontiers of knowledge and methodology. In a field as dynamic as neurodegeneration it should be considered in the planning phase of the projects that new findings are to be expected during the project and that these findings might need to be integrated fast into the already running project to keep up to date with current research.

Face to face meetings helped to promote engagement and motivation of the partners and are a necessary part of any project and as such should be built into the IMI structure of all projects, i.e. a minimum of two face-to-face steering committee meetings per year. The rotation of the meetings hosted by a different partner each time gave the participants a chance to learn more about the research and resources at each site and allowed for other participants to join in on some innovative scientific sessions.

As the project was in the initial wave of projects under IMI1, we were not able to benefit from some innovations such as standard templates available to use for the Project Agreement, Project Manual, Communication Plan, etc. When a project is starting up, there is lots to think about and these types of standard templates are very helpful and ensures critical points are not forgotten.

When it came to the question of sustainability outside of the period of the project agreement, some guidance documents or templates on how to approach this part would have been appreciated instead of having to reinvent the wheel. This is especially important when there is a large legacy of bio samples that need to be available at the end of the study for other EU researchers. There is a cost associated, but funds are not available to sustain this.

Summary Recommendations

IMI is a powerful platform to facilitate industrial and academic partners working together. Over our collective experience of IMI projects there are certainly significant benefits to be observed from this partnership. Above we have highlighted some of the recommendations that we see as valuable insights.

AETIONOMY was affected by data and the question of its availability from start to end – we have learned a lot about the challenges associated with making data available and that there is no one-size fits all solution to this. It goes without saying that future projects should pay critical attention to the quality, accessibility, availability of existing datasets that are planned to be analysed, and the feasibility of data sharing in its different aspects (quality, reproducibility, administrative, legal, IP, ethics, consents, logistics), and as a footnote it is heartening to see that in more recent times projects have been initiated with mechanisms in place to mitigate some of these challenges.

AETIONOMY has clearly launched an important concept of mechanism-based redefinition of neurodegenerative diseases, but also raised the difficulties of obtaining large datasets of sufficient quality to answer this question. This may be achieved by increasing the collaborations and partnerships between the academic and industrial worlds working together to build state of the art platforms for generating such datasets that would fit both interests, not only for specific projects but also in real life patients care.

In view of your project achievements, please provide your views on potential new research to further advance the field.

Mechanism-based stratification of patients is an emerging field and we are far from having established a best practice for the association between mechanisms (e.g. represented in graph models) and patient-level data. Furthermore, all mechanism-associations we have evaluated so far have been largely ignorant of progression over time. In fact, a quantitative representation of mechanisms that

could explain the changes of biomarkers (e.g. as seen in ADNI and PPMI progression models) is still missing. First attempts at linking mechanisms to longitudinal data and progression patterns (defined by biomarkers) are under way, but these first approaches are far from becoming routine. We expect future endeavours to focus here.

We are aware that the mechanistic analysis of neurodegenerative diseases, and their progression, represents one of the most challenging areas in translational research. Indeed, it is often with envy that we consider the position of colleagues who work in oncology, where access to sample materials; numbers of patients and observation times required to link diagnosis to clinical outcome are considerably more forthcoming. Neurodegeneration research, as we tried to establish it in AETIONOMY, cannot be adapted to other areas without substantial modification. First encouraging results were obtained in psychiatry, and we expect an uptake of AETIONOMY modelling and mining strategies in the area of quantitative neuropsychiatry before we consider adaptation to other indication areas. However, the fundamental approaches at the core of the project will be adapted and applied broadly.

One aspect of AETIONOMY is currently being ported to the oncology domain: the functional interpretation of genetic variation information using OpenBEL was pioneered in the course of AETIONOMY and there is now a strong uptake of these principles in the precision medicine / personalized tumour treatment optimization domain. For instance, a collaboration of Fraunhofer SCAI with Harvard Medical School (Peter Sorger and his team) aims at creating a “mutation interpretation machine” that links data and computable knowledge in a way that personalized assessment of the functional consequences of mutation patterns in the tumours of patients is enabled.

Our clinical study had some limitations due to the observational and cross-sectional nature of the cohort. It was unfortunately not possible in the timeframe of the project, and considering the relatively limited funding, to build a longitudinal cohort of patients. Longitudinal datasets are needed to better understand the complexity of the progression of neurodegenerative diseases. Other limitations are related to the relatively small amount of data analysed in regard to the complexity of the diseases. For instance, one of the main WP5 findings of AETIONOMY was that age and sex of patients have significant impact on immune biomarker levels. If not stratified for these covariates, the power of neuroinflammation biomarkers was very limited. However, when combining for example aging and pathology traits, higher potential is indicated. Again, we expect future projects to include numbers of patients large enough to cover wide ranges of ages in patients and controls, and to support stratification by sex, genotypes and other confounders.

AETIONOMY had substantial influence on the way problem-solving was addressed at computational level in the IMI2-project PHAGO (https://www.phago.eu) and has contributed to programmatic discussions in IMI-project RADAR-AD (https://www.imi.europa.eu/projects-results/project-factsheets/radar-ad). Beyond the discourse within IMI, which was and is always strong and which was boosted by events like the FINAL SYMPOSIUM of the AETIONOMY project in November 2018, the crosstalk to the Human Brain Project (mediated by including Prof. Viktor Jirsa, AMU into AETIONOMY) has proven to bring new and exciting perspectives to EU-funded research.

We would go so far to state that IMI-funded projects have a lot to bring to H2020 and related projects; it shows that in the collaboration between EFPIA partners and academia, a wide spectrum of robust

and reliable solutions have been developed that can support other EU-funded research. Our recommendation is therefore clear: IMI needs to encourage IMI-funded projects to actively seek options to increase the dissemination to and partnering with other EU-funded projects. In the area of neurodegenerative disease research, stronger emphasis should be put on intensified collaboration with EraNET NEURON and JPND; and – of course – the flagship project of neurology in Europe, the Human Brain Project.

We once again recommend that EU bodies, including IMI and H2020 officers, need to team up to work together and organise workshops and small conferences that bring together projects in similar indication areas. EU funding can be leveraged much better through this crosstalk and it is a core duty of EU officers and programs, to enable more of this crosstalk in the future.

Naturally, as we have discussed, the full interrogation and exploitation of the data that now resides in the AKB is far from complete, and we expect not only the partners but researchers from across the community to engage in mining these data to explore the taxonomy further over the coming years.

AETIONOMY's existing partners will of course in the short term continue to explore the nascent clusters presented by the approaches described above, both to further explore robustness and to elaborate the biological associations that drive those signals, as it is these that hold the key to unlocking the potential for the development of new diagnostic or treatment modalities. Ultimately, the goal of the project was to bring the development of the taxonomy into clinical practice, and despite this still being somewhat distant, future research in this direction will leverage the deliveries of this project in order to achieve these goals.