

Simulation

# The Virtual Dementia Cohort

Martin Hofmann-Apitius

AETIONOMY FINAL SYMPOSIUM

November 29

Bonn, Germany



innovative  
medicines  
initiative

**Mission**

*To increase knowledge of the causes of Alzheimer's and Parkinson's Disease by generating a mechanism-based taxonomy; to validate the taxonomy in a prospective clinical study that demonstrates its suitability for identifying patient subgroups (based on discrete disease mechanisms); to support future drug development and lay the foundation for improved identification and treatment of patient subgroups currently classified as having AD or PD.*



Inspired by patients.  
Driven by science.





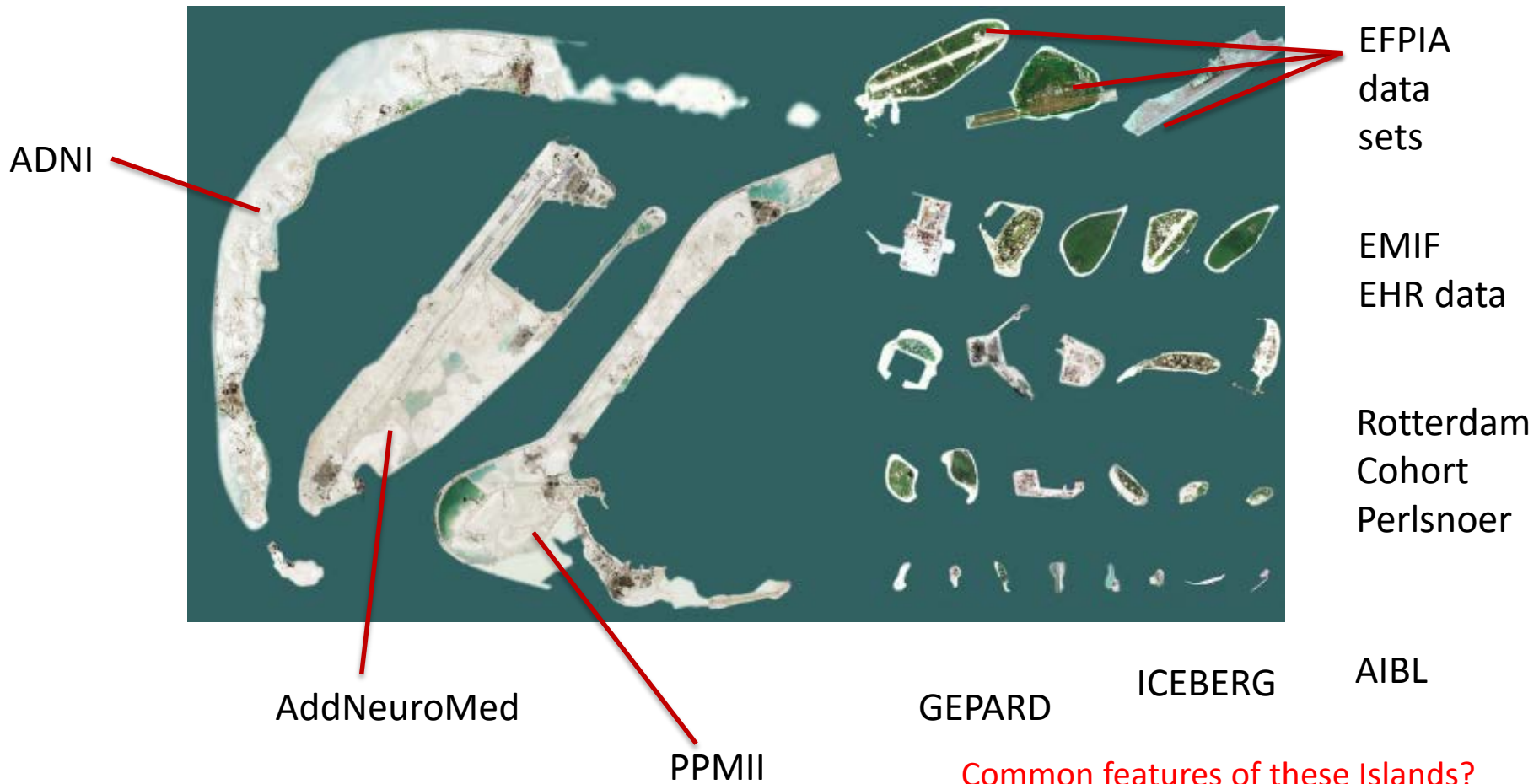
## Patient-Level Data – Essential for AETIONOMY

Patient-level data were an essential cornerstone of the AETIONOMY strategy

- For pattern detection in patient-level data
- For all stratification testing (unbiased, just to detect strata in study data)
- For all mechanism-based stratification (mapping of mechanisms to study data)
- For all hypothesis validation (testing in independent clinical study data, whether we can reproduce identified strata linked to mechanisms)
- For comparison of study data
  - variability and robustness of measured variables
  - understand patterns linked to progression
  - understand the impact of inclusion and exclusion criteria
  - understand study protocols and their influence on biomarker measurements



## Sometimes, Study Data are like Islands





## Some Clinical Datasets are like Ghost Ships



They appear in grant applications again and again, but you will never be able to work with them .....



## Other Data Sets are perfectly siloed



- Variable catalog ?
- Summary statistics ?
- Interoperability ?
- Shared metadata ?
- MERGE dataset ?
- Pre-Processed ?
- Curated ?
- Imputed ?
- ....



## Getting Access to Study Data means nothing ....

Getting access to study data does not mean, that you can use them

- Pre-processing (AddNeuroMED merge took us months)
- Cannot share data within the consortium (everybody pre-processes again and again)
- Sometimes, important variables (visits) are not in the package (AIBL)
- We have to chase special data owners (AddNeuroMed imaging data)
- For comparison of study data
  - Only AddNeuroMed and ADNI could be compared
  - AIBL is almost not comparable to ADNI & AddNeuroMed
  - EFPIA data were made “accessible” in a SAS-sponsored environment that effectively prevents you from being able to compare and mine data



## Data Protection, Data Protection and Data Protection ...





## GDPR, Declaration of Human Rights (1948) et al. ....



....  
ROADBLOCKS  
for  
translational  
biomedical  
research







# The Way Out: The Virtual Dementia Cohort



## You must not play with Patient Data!

- DOCH !

*The German word “Doch!” has a meaning close to “sure!” or “Yes, of course!” .... but it is much stronger and it disproves a previous statement.*

*We love to use that term in disputes when somebody says “you can’t do that” and we simply say: “doch”.*

*A bit like Obamas “yes, we can ...”*





## Actually, we NEED TO PLAY WITH DATA

- **Data science methods** need data to play with
- **Clinical trial simulation** needs data to play with
- **Mechanism-based stratification** needs data to play with
- **Deep learning** needs lots of data to play with
- **Students** need data that can be freely shared to improve their data science skills





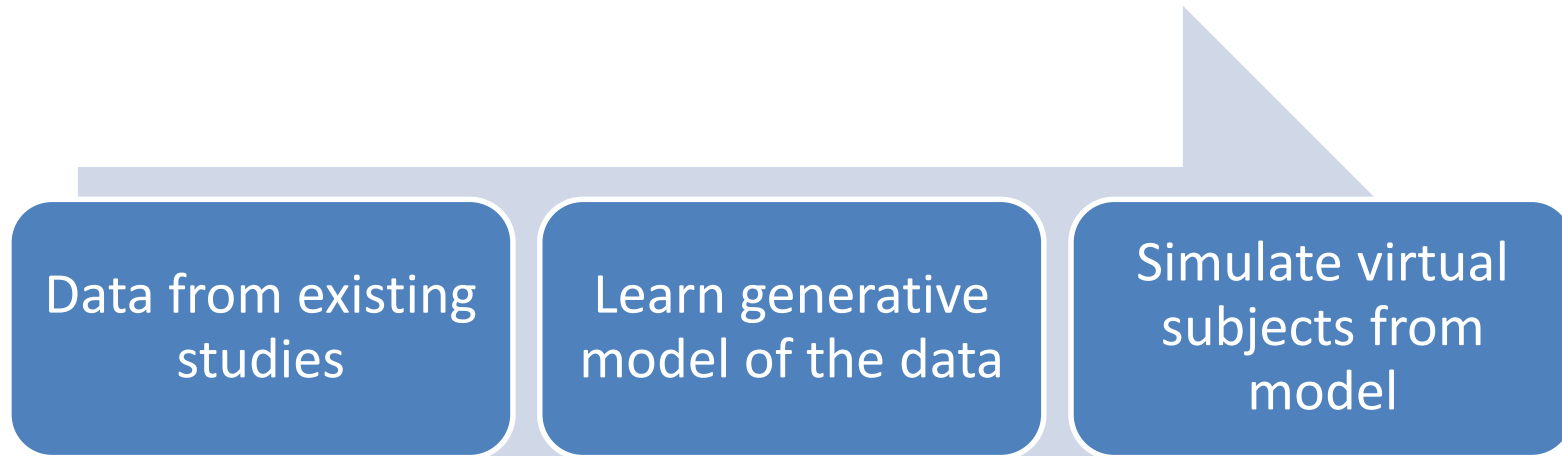
## Virtual Patient Cohorts

- Synthetic data sets
- Instructed by reality
- No patient data privacy rights
- Very close to reality
- Allow to “publish” clinical data
- Allows to share clinical data
- Allow for global meta-cohorts
- Can integrate a priori knowledge
- Can be used to ask “unethical” questions
- Can be used to mix and merge





# Learning Synthetic Data from Real World Studies



Work of:  
**Holger Fröhlich**  
Akrishta Sahai  
Meemansa Sood  
Shashank Khanna  
Asif Emon Khan  
Reagon Karki

**Bayesian Networks**

AutoEncoders

Recurrent Neural Networks

.....

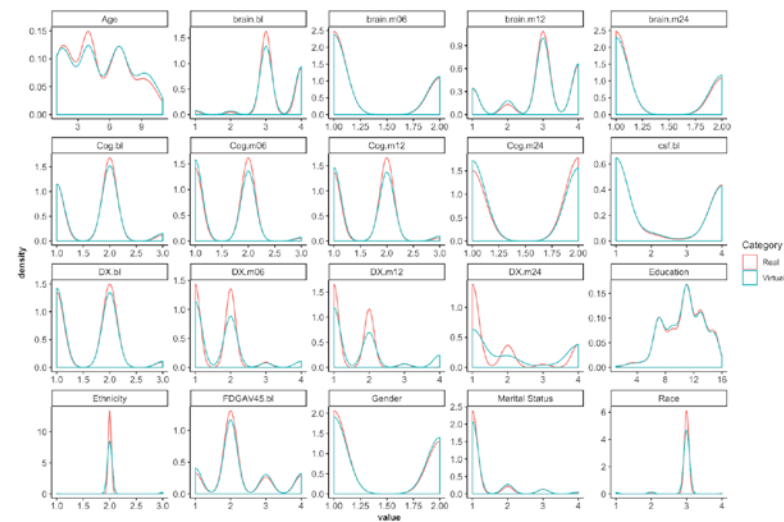
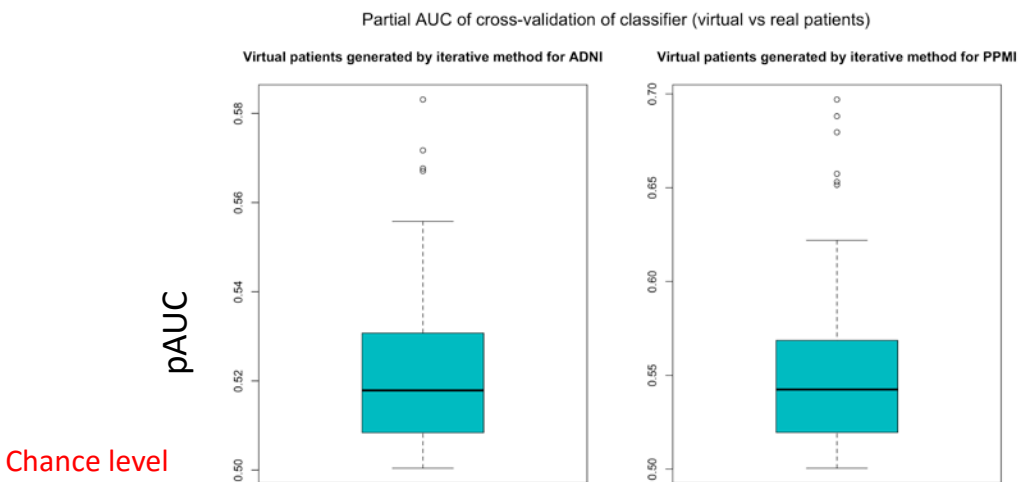
Simulate virtual subjects from model

Never identical to real patients

Khanna, Shashank, et al. "Using Multi-Scale Genetic, Neuroimaging and Clinical Data for Predicting Alzheimer's Disease and Reconstruction of Relevant Biological Mechanisms." *Scientific reports* 8.1 (2018): 11173.



# Virtual Patient Simulations Look Realistic



Classifier cannot detect virtual patients significantly better than chance level

- Partial area under ROC curve (sensitivity  $\geq 90\%$  for detecting real patients)

Marginal distributions of virtual and real patients agree visually

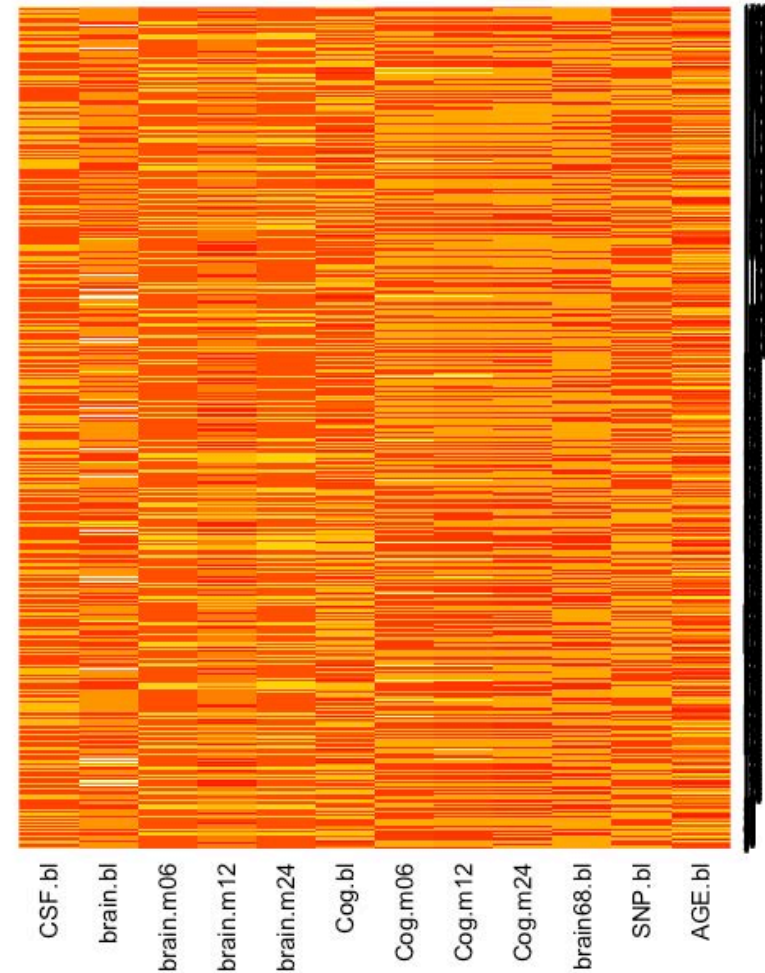


## Thousands of ADNI – like, longitudinal VPs

Virtual Patients generated by a Bayesian Network representing essential variables and their conditional dependency in ADNI are shown.

In the heat map shown, the features of 689 real-world ADNI patients and the features of 1000 virtual patients have been clustered.

Try to tell them from each other ....

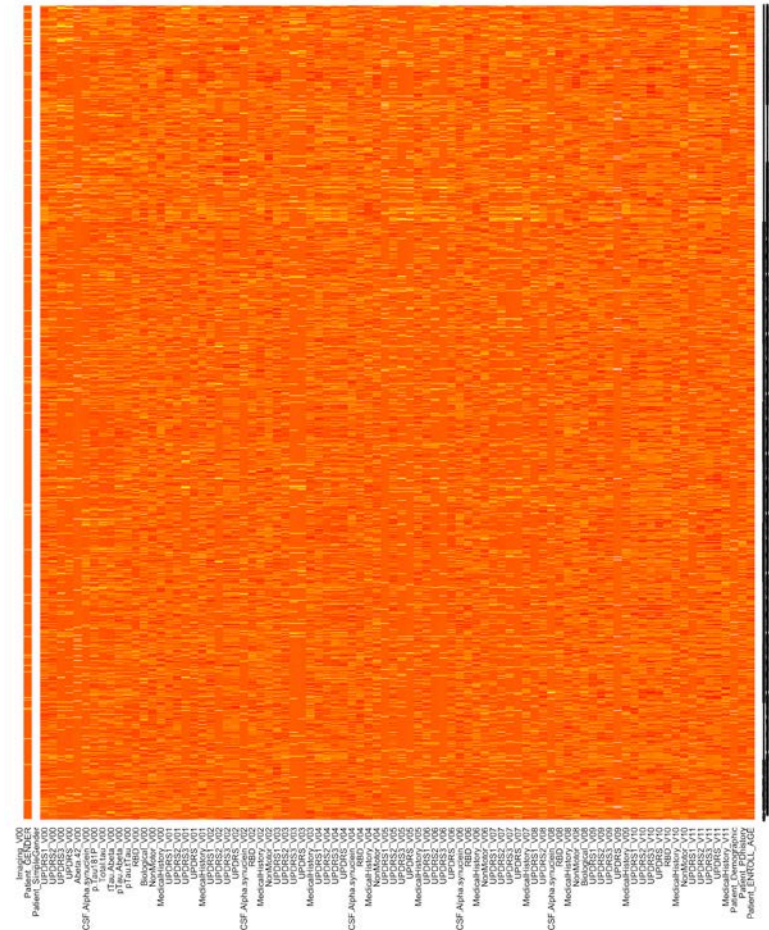




## Millions of PPMI – VPs

Virtual Patients generated by a Bayesian Network representing essential variables and their conditional dependency in PPMI are shown.

The heat map shown here comprises more than 10,000 PPMI like Virtual Patients.

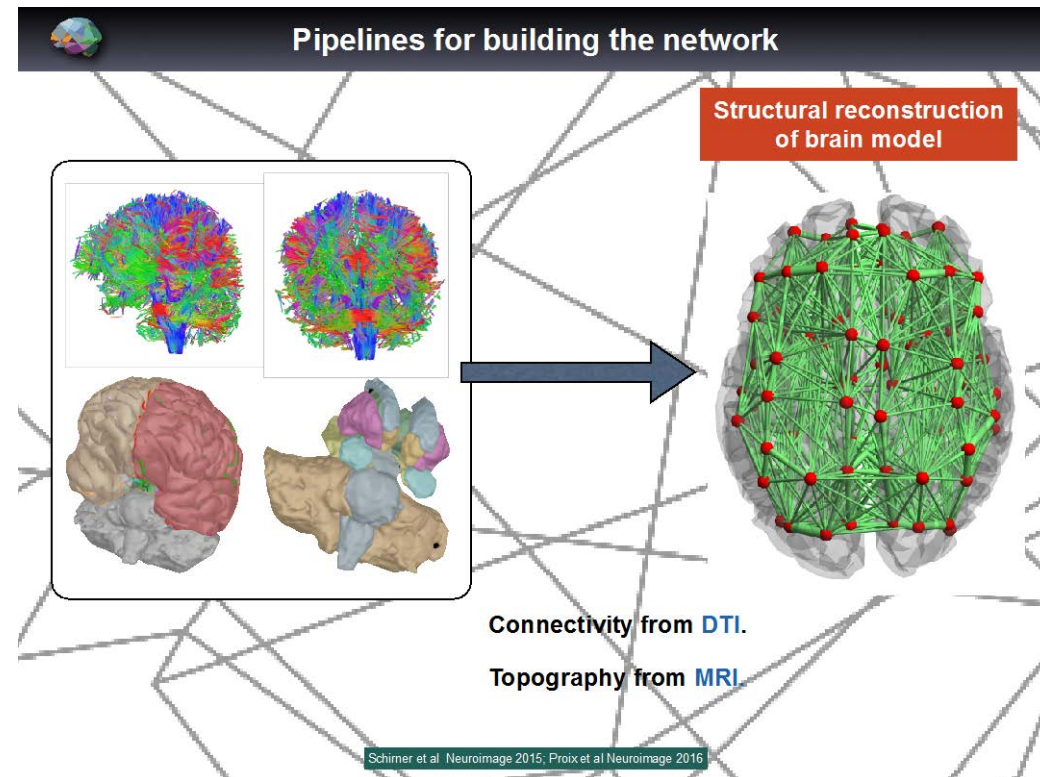






## A Dedicated Task in AETIONOMY

- During the runtime of AETIONOMY, we added a new task: The Virtual Dementia Cohort
- A new partner came on board: AIX Marseille, Prof. Viktor Jirsa
- Viktor can simulate brain states
- Viktor can simulate fMRI data and entire connectomes
- A perfect match for our activities: we can simulate entire cohorts





# Virtual Patient Cohorts – A Focused Research Topic in FRONTIERS



Research Topic

## Virtual Patient Cohorts and Data Sharing: Prison Break for Patient-Level Data

Comment

0



0



0



0



0



New

Submit your abstract

Submit your manuscript

Overview

Articles

Authors

Impact

Comments

VIEWS

118

### About this Research Topic

The Challenge:

Translational computational biomedicine is often hampered by limitations that come with the sharing of patient-level data. In fact, most patient-level data sets cannot be used in collaborative environments and require individual agreements between institutions for research purposes (e.g. the ADNI data set holders in the field of Alzheimer's Disease). The Terms of Use for such 'public' data sets often do not allow researchers to generate subsets, to blend or mix data sets, or to form hybrids and other derivatives.

### Topic Editors



**Martin Hofmann-Apitius**



Follow

Fraunhofer Institute for Algorithms and Scientific Computing (FHG) Sankt Augustin, Germany

73 publications





## Thank You

UCB: Holger Fröhlich

AIX Marseille: Viktor Jirsa

Djouya (ARBABYAZD Mohammad)

Fraunhofer: Shashank Khanna

Akrishta Sahai

Meemansa Sood

Reagon Karki

Asif Emon Khan

