EMIF - European Medical Information Framework

Worldwide-ADNI Update Meeting
Friday, July 11, 2014 Copenhagen
EMIF Vision

To be the trusted European hub for health care data intelligence enabling new insights into diseases and treatments
EMIF consortium

- 56 partners
- 14 European countries represented
- 56 MM € worth of resources (in-kind / in-cash)
- “3 projects in one”
- 5 year project (2013 – 2017)

To be the trusted European hub for health care data intelligence enabling new insights into diseases and treatments
The EMIF project
European Medical Information Framework

❖ Leadership team
  – Bart Vannieuwenhuyse (Janssen R&D, Beerse, Belgium)
  – Simon Lovestone (King's College, London, United Kingdom)
  – Johan van der Lei (Erasmus Universitair Medisch Centrum Rotterdam, Rotterdam, The Netherlands)

❖ AD topic leads
  – Pieter-Jelle Visser (VU Medical Center Amsterdam)
  – Johannes Streffer (Janssen R&D, Beerse, Belgium)
Secondary use of health data to improve clinical research
- Opportunities

The “burning platform” for Life Sciences

The value of healthcare data for secondary uses in clinical research and
development - Gary K. Mallow, Merck, HIMSS 2012
EMIF roadmap

**Mission**

The EMIF project aims to improve access to human health data for life sciences research - this will be achieved via a 3 phased approach:

**Knowing**
- Realize cohort integration and data source profiling to allow meta data browsing

**Accessing**
- Allow searches on aggregated data across different sources and countries within a single point of access

**Using**
- Allow advanced data analysis for specific research questions into the identification and validation of novel biomarkers
EMIF – platform for modular extension

EMIF governance

Research Topics

EMIF - Metabolic

Metabolic

EMIF - Metabolic

Patient generated data

Call 5

Risk stratification

EMIF - AD

Risk factor analysis

Call 5

Prevention algorithms

TBD

Predictive screening

IMI Structure and Network

Data Privacy

Analytical tools

Semantic Integration

Information standards

Data access / mgmt

Data Privacy

Analytical tools

Semantic Integration

Information standards

Data access / mgmt

EMIF Platform
EMIF-AD – Overall approach

- Use existing datasets through data platform
- Use extreme phenotypes as an outcome
  - Amyloid pathology
  - APOE4 negative/positive
  - Hippocampal atrophy
  - Cognitive decline

*Figure 1. Stages of AD*
Cohorts linked

- Preclinical AD studies
  - Barcelona study (Spain)
  - SIGNAL study (Spain)
  - GAP study (Spain)
  - Leuven study (Belgium)
  - Geneva study (Switzerland)
  - EMIF-AD 60-80 cohort (Netherlands/UK)
  - EMIF-AD 90+ cohort (Netherlands/UK)

- Population studies
  - Heinz-Nixdorf recall study (Germany)
  - LOLIPOP (UK)
  - CFAS-1 (UK)
  - Manchester and Newcastle longitudinal study of cognitive aging (UK)
  - Rotterdam study (Netherlands)
  - Maastricht Aging study (Netherlands)
  - SNAC-K (Sweden)
  - CAIDE study (Finland)

- Single centre memory-clinic based studies
  - Finger study (Finland)
  - Athens study (Greece)

- European memory-clinic based multi-centre studies
  - Pharma-Cog
  - Addneuromed
  - DESCRIPA
  - EDAR
  - DiMi
  - NEST-DD
  - PredictAD
  - EADC FDG PET study
  - EADC prodromal AD study

- International multi-centre studies
  - ADNI

- Prodromal AD trials
  - Lipididiet
  - MCI Donepezil study

- National memory-clinic based multi-centre studies
  - Brainpower (Sweden)
  - String of Pearls (Netherlands)
  - PreAL (France)
  - Memento (France)

- Diverse data ownership and availability
Clinical and imaging Work package: Simon Lovestone (London), Hilkka Soininen (Kuopio), Patrizia Mecocci (Perugia), Bruno Vellas (Toulouse), Magda Tsolaki (Thessaloniki), Iwona Kłoszewska (Lodz), Christian Spenger (Karolinska)

45 publications to 2006-2014
Fingerprinting

- A detailed questionnaire to characterize AD cohorts was sent to cohort owners of 52 European AD cohorts
- Information on 27 AD cohorts made available in the EMIF browser

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>SCI</th>
<th>MCI</th>
<th>AD</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>4147</td>
<td>1469</td>
<td>5245</td>
<td>2725</td>
</tr>
<tr>
<td>Cognitive data</td>
<td>3261</td>
<td>1394</td>
<td>4969</td>
<td>1940</td>
</tr>
<tr>
<td>Plasma</td>
<td>1402</td>
<td>511</td>
<td>1504</td>
<td>962</td>
</tr>
<tr>
<td>Serum</td>
<td>3704</td>
<td>227</td>
<td>766</td>
<td>1272</td>
</tr>
<tr>
<td>DNA</td>
<td>1853</td>
<td>511</td>
<td>1607</td>
<td>1594</td>
</tr>
<tr>
<td>RNA (blood)</td>
<td>963</td>
<td>341</td>
<td>1077</td>
<td>185</td>
</tr>
<tr>
<td>CSF</td>
<td>195</td>
<td>342</td>
<td>1513</td>
<td>1048</td>
</tr>
<tr>
<td>Urine</td>
<td>306</td>
<td>0</td>
<td>85</td>
<td>0</td>
</tr>
<tr>
<td>MRI</td>
<td>1872</td>
<td>950</td>
<td>3643</td>
<td>1345</td>
</tr>
<tr>
<td>PET-FDG</td>
<td>45</td>
<td>103</td>
<td>801</td>
<td>71</td>
</tr>
<tr>
<td>PET-amyloid</td>
<td>115</td>
<td>0</td>
<td>72</td>
<td>0</td>
</tr>
<tr>
<td>SPECT</td>
<td>0</td>
<td>24</td>
<td>143</td>
<td>21</td>
</tr>
<tr>
<td>EEG</td>
<td>61</td>
<td>757</td>
<td>648</td>
<td>1015</td>
</tr>
<tr>
<td>MEG</td>
<td>0</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>
Selection of data based on Research questions

- Combination of EHR and Cohort studies from the same region
- Collection of samples for biomarker discovery/confirmation
- Pooling of epidemiologic cohorts and/or EHR for risk factors
### Scientific Advisory Board

<table>
<thead>
<tr>
<th>Name</th>
<th>Institution</th>
<th>Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peter St. George Hyslop</td>
<td>University of Cambridge</td>
<td>Co-Chair</td>
</tr>
<tr>
<td>Robert Green</td>
<td>Harvard</td>
<td>Co-Chair</td>
</tr>
<tr>
<td>David Bennett</td>
<td>Rush, ROS/MAP PI</td>
<td>Member, Data Contributor</td>
</tr>
<tr>
<td>Michael Weiner</td>
<td>UCSF, ADNI PI</td>
<td>Ex Officio, Data Contributor</td>
</tr>
<tr>
<td>Simon Lovestone</td>
<td>University of Oxford, AddNeuroMed PI</td>
<td>Member, Data Contributor</td>
</tr>
<tr>
<td>John Kauwe</td>
<td>Brigham Young University</td>
<td>Member</td>
</tr>
<tr>
<td>Alan Evans</td>
<td>McGill University</td>
<td>Member</td>
</tr>
<tr>
<td>George Vradenburg</td>
<td>USAgainstAlzheimer's</td>
<td>Member</td>
</tr>
<tr>
<td>Gil Rabinovici</td>
<td>UCSF</td>
<td>Member</td>
</tr>
<tr>
<td>Kai Blenow</td>
<td>Göteborg University</td>
<td>Member</td>
</tr>
<tr>
<td>Kristine Yaffe</td>
<td>UCSF</td>
<td>Member</td>
</tr>
<tr>
<td>Maria Issac</td>
<td>EMA</td>
<td>Member</td>
</tr>
<tr>
<td>Nolan Nichols</td>
<td>University of Washington</td>
<td>Member</td>
</tr>
<tr>
<td>Paul Thompson</td>
<td>UCLA</td>
<td>Member</td>
</tr>
<tr>
<td>Reisa Sperling</td>
<td>Harvard</td>
<td>Member</td>
</tr>
<tr>
<td>Scott Small</td>
<td>Columbia</td>
<td>Member</td>
</tr>
<tr>
<td>Maria Corlito</td>
<td>Alzheimer's Foundation</td>
<td>Ex Officio</td>
</tr>
<tr>
<td>Neil Buckholz</td>
<td>NIA</td>
<td>Ex Officio</td>
</tr>
</tbody>
</table>

### Challenge Organizers

<table>
<thead>
<tr>
<th>Name</th>
<th>Institution</th>
<th>Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>Andy Simmons</td>
<td>King’s College London</td>
<td>Data Scientist</td>
</tr>
<tr>
<td>Arno Klein</td>
<td>Sage Bionetworks</td>
<td>Neuroimaging Lead</td>
</tr>
<tr>
<td>Benjamin Logsdon</td>
<td>Sage Bionetworks</td>
<td>Data Scientist</td>
</tr>
<tr>
<td>David Fardo</td>
<td>University of Kentucky</td>
<td>Data Scientist</td>
</tr>
<tr>
<td>Christine Suver</td>
<td>Sage Bionetworks</td>
<td>Data Governance</td>
</tr>
<tr>
<td>Christopher Bare</td>
<td>Sage bionetworks</td>
<td>Synapse Software Engineer</td>
</tr>
<tr>
<td>Gustavo Stolovitzky</td>
<td>IBM</td>
<td>Sage/DREAM Executive Committee</td>
</tr>
<tr>
<td>John &quot;Keoni&quot; Kauwe</td>
<td>BYU</td>
<td>Scientific Lead</td>
</tr>
<tr>
<td>Mette Peters</td>
<td>Sage Bionetworks</td>
<td>Challenge Co-Lead</td>
</tr>
<tr>
<td>Nicholas Tustison</td>
<td>UVA</td>
<td>Data Scientist</td>
</tr>
<tr>
<td>Richard Dobson</td>
<td>King’s College London</td>
<td>Data Scientist</td>
</tr>
<tr>
<td>Satrajit Ghosh</td>
<td>MIT</td>
<td>Data Scientist</td>
</tr>
<tr>
<td>Stephen Friend</td>
<td>Sage Bionetworks</td>
<td>Sage/DREAM Executive Committee</td>
</tr>
<tr>
<td>Stephen Newhouse</td>
<td>King's College London</td>
<td>Data Scientist</td>
</tr>
<tr>
<td>Taylor Maxwell</td>
<td>GWU</td>
<td>Data Scientist</td>
</tr>
<tr>
<td>Thea Norman</td>
<td>Sage Bionetworks</td>
<td>Challenge Strategy and Logistics</td>
</tr>
</tbody>
</table>
Subchallenge 1:
Predict the change in cognitive scores 24 months after initial assessment.

Scientific Rationale: Answers to this question will help predict cognitive trajectory and potentially provide new approaches for early diagnosis of AD. This earlier identification would allow for more efficient selection of samples for clinical trials and possibilities for earlier disease treatment.

Training Set
Ancillary Data

Test Set
ROS/MAP
Subchallenge 3: Classify individuals into diagnostic groups using MR Imaging.

Scientific Rationale: If a single MR image could be used to differentiate AD patients from people with mild cognitive impairment or from healthy individuals, research can focus on the specific anatomical structures that are different between the groups. Currently, MRI data are acquired routinely in hospitals: thus a winning algorithm could potentially be retrospectively applied to existing archives of clinical data as well as to future scans without requiring additional resources or expertise.
Data aggregation and access
– US and Europe

- **Phase 1**
  - Exchange and mapping of meta-data (‘fingerprints’)
  - 24 attributes in GAAIN; >200 in EMIF
  - Cross programme mapping underway

- **Phase 2**
  - “search on GAAIN = search on EMIF” and vice versa
  - Access to cohorts across Europe and USA

- **Phase 3**
  - Joint analysis
  - Data-standardisation