



# ExCAPE: Exascale Compound Activity Prediction Engines

EXDCI Event  
Prague, May 2016

- Exascale machine learning (ML)
  - Machine learning is already a very large consumer of cycles in data centres
  - **New algorithms** need to be developed to turn extra available computation into **more accurate, more useful** models
  - Applying analysis and scaling techniques from HPC will help **exploit maximum potential** of these algorithms
    - Enable move to exascale
  - We aim to open up **new areas** for application of HPC expertise with important societal benefits
- **Our proposal**
  - *Apply Exascale machine learning to problems in the pharmaceutical industry*
    - (But techniques applicable to other uses of ML)

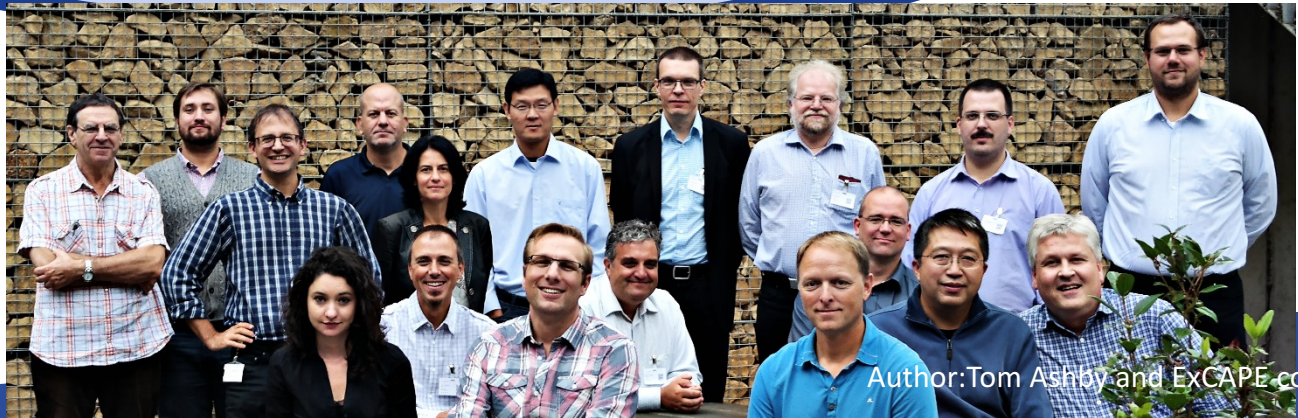


# ExCAPE: Overview

- The big picture: *the constituent threads...*
  - Exascale supercomputers
    - Large clusters
    - Accelerators
  - Machine learning
    - Supervised and unsupervised learning
    - Confidence estimation, dyadic data etc.
    - Learning performance (accuracy of model)
  - Computation in the Pharma Industry
    - Chemogenomics
    - ADMET



# The Partners



06/10/2016

Author: Tom Ashby and ExCAPE consortium

- **Offering to Ecosystem**
  - An *application* and *algorithmic techniques* that are on the edge of the areas covered by traditional HPC
    - Machine learning for life sciences
    - Software and techniques to use
  - A view on how such applications will interact with current and future HPC systems
    - Both hardware and software

- **International collaboration**

- Progress

- Initial contact (through Intel) with the *K supercomputer-based drug discovery (KBDD)* consortium (Biogrid Pharma Consortium), Japan
  - Presentation and discussion on deep learning infrastructure used by U.Kyoto

- Help from EXDCI?

- Possible projects to connect with:
  - RaPyDLI (US: Geoffrey Fox, Jack Dongarra etc)
- News about other *Machine Learning on HPC* projects...

- **cPPP (ETP4HPC + CoEs)**
  - Current
    - Discussions with *Bioexcel CoE*: looking for application synergy
  - Plans
    - Set up something more concrete with *Bioexcel*
    - Look for hardware and systems partners interested in results on e.g. large scale collaborative filtering

- **Extreme Scale Demonstrators**
  - We are a possible *application/software* partner
  - Less typical application: **big data**, machine learning, **pharma industry** etc.
  - **Porting** of application to a demonstrator platform, and **benchmarking**
  - Feedback from potential industrial end users