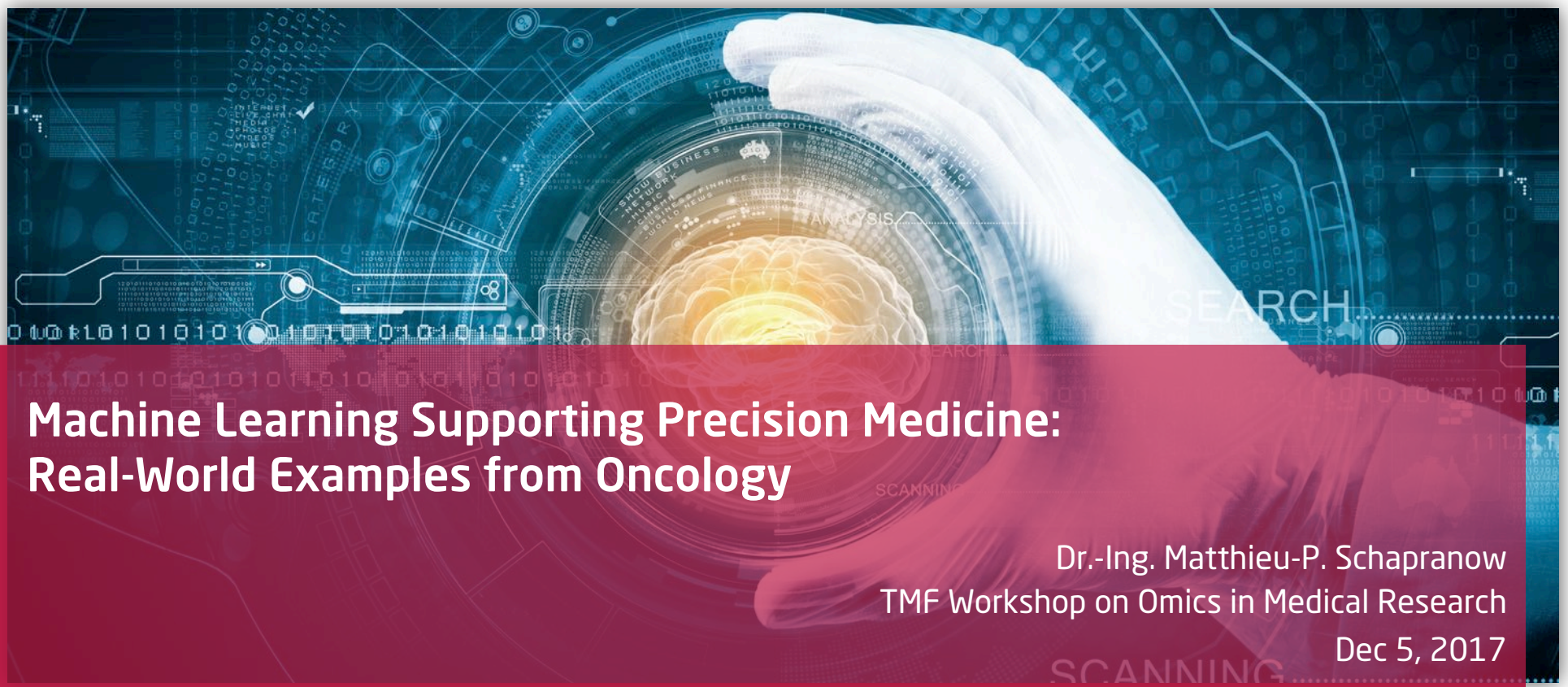
Hasso Plattner Institut
Digital Engineering · Universität Potsdam



Machine Learning Supporting Precision Medicine: Real-World Examples from Oncology

Dr.-Ing. Matthieu-P. Schapranow

TMF Workshop on Omics in Medical Research

Dec 5, 2017

The Challenge
Distributed Heterogeneous Data Sources

Hasso Plattner Institut

Human genome/biological data
600GB per full genome
15PB+ in databases of leading institutes

Human proteome
160M data points (2.4GB) per sample
>3TB raw proteome data in ProteomicsDB

Hospital information systems
Often more than 50GB

PubMed database
>23M articles

Cancer patient records
>160k records at NCT

Medical sensor data
Scan of a single organ in 1s creates 10GB of raw data

Prescription data
1.5B records from 10,000 doctors and 10M Patients (100 GB)

Clinical trials
Currently more than 30k recruiting on ClinicalTrials.gov

Routine Data
Medical and treatment costs

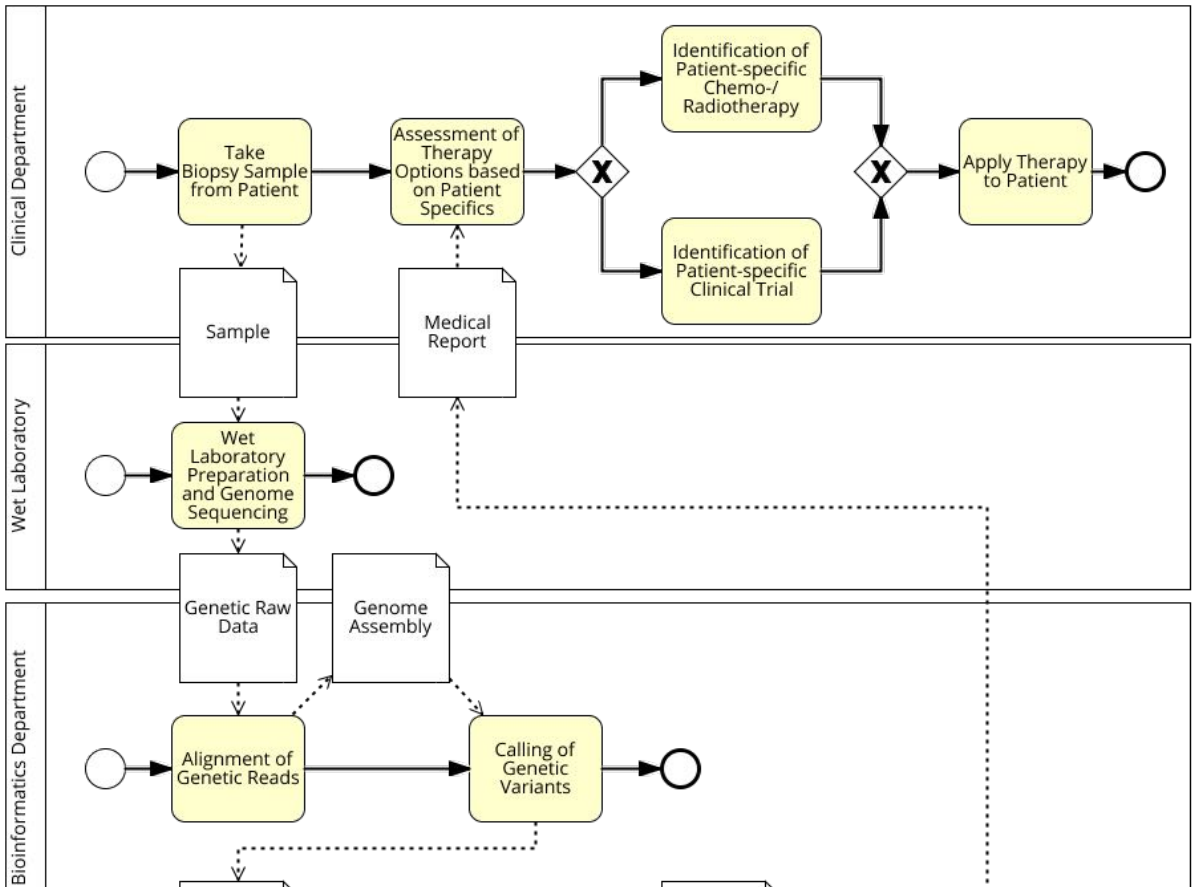
Machine Learning Supporting Precision Medicine

Schapranow, Omics in Medical Research, Dec 5, 2017

Our Motivation
Turn Precision Medicine Into Clinical Routine

- Can we enable clinicians to:
  - Incorporate all available patient specifics, e.g. *-omics data, lab results,
  - Access latest worldwide medical knowledge, and
  - Get questions answered interactively, e.g. during tumor board discussions.

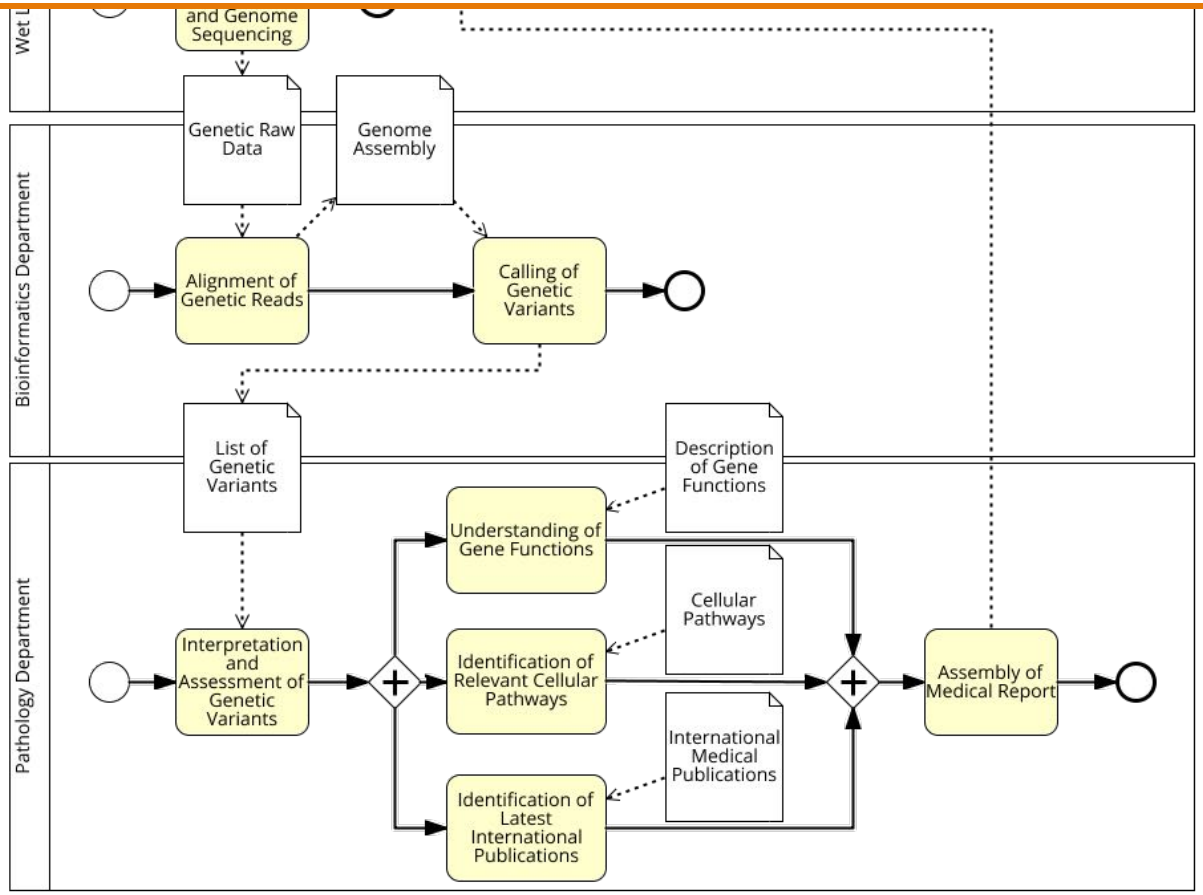
Simplified Clinical Oncology Process (1/2)



Machine Learning Supporting Precision Medicine

Schapranow, Omics in Medical Research, Dec 5, 2017

Simplified Clinical Oncology Process (2/2)

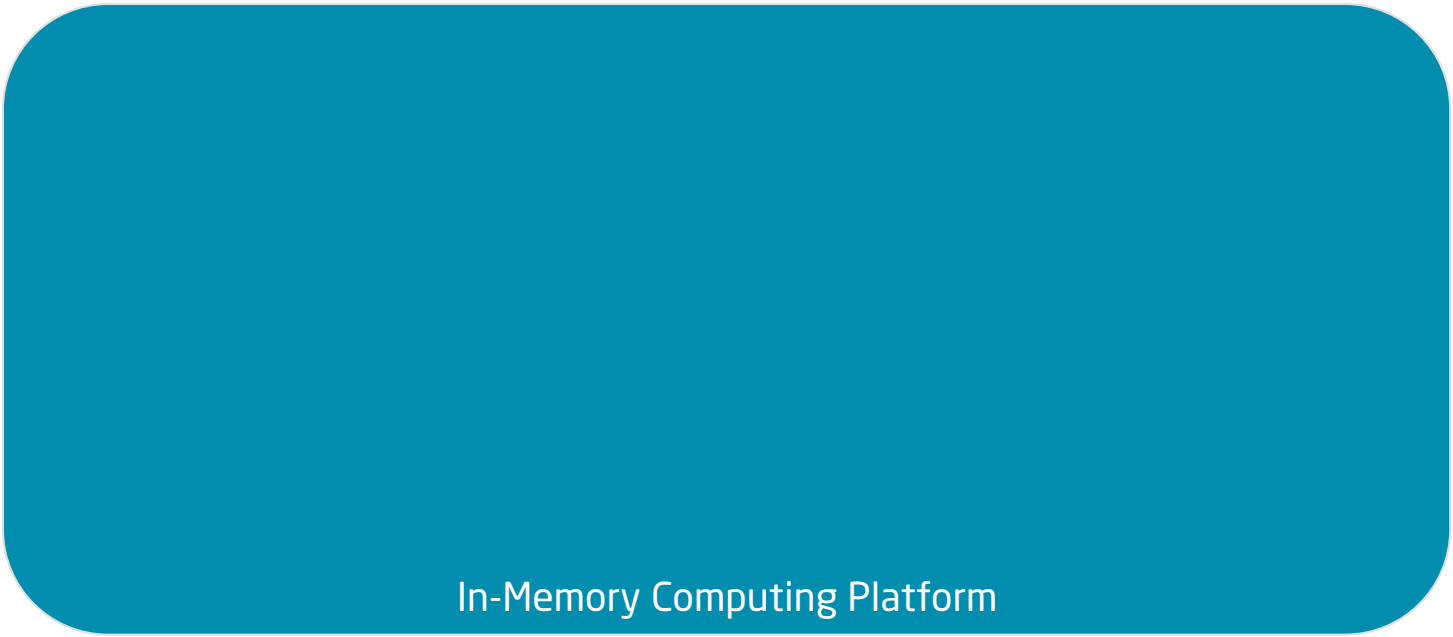


Machine Learning Supporting Precision Medicine

Schapranow, Omics in Medical Research, Dec 5, 2017

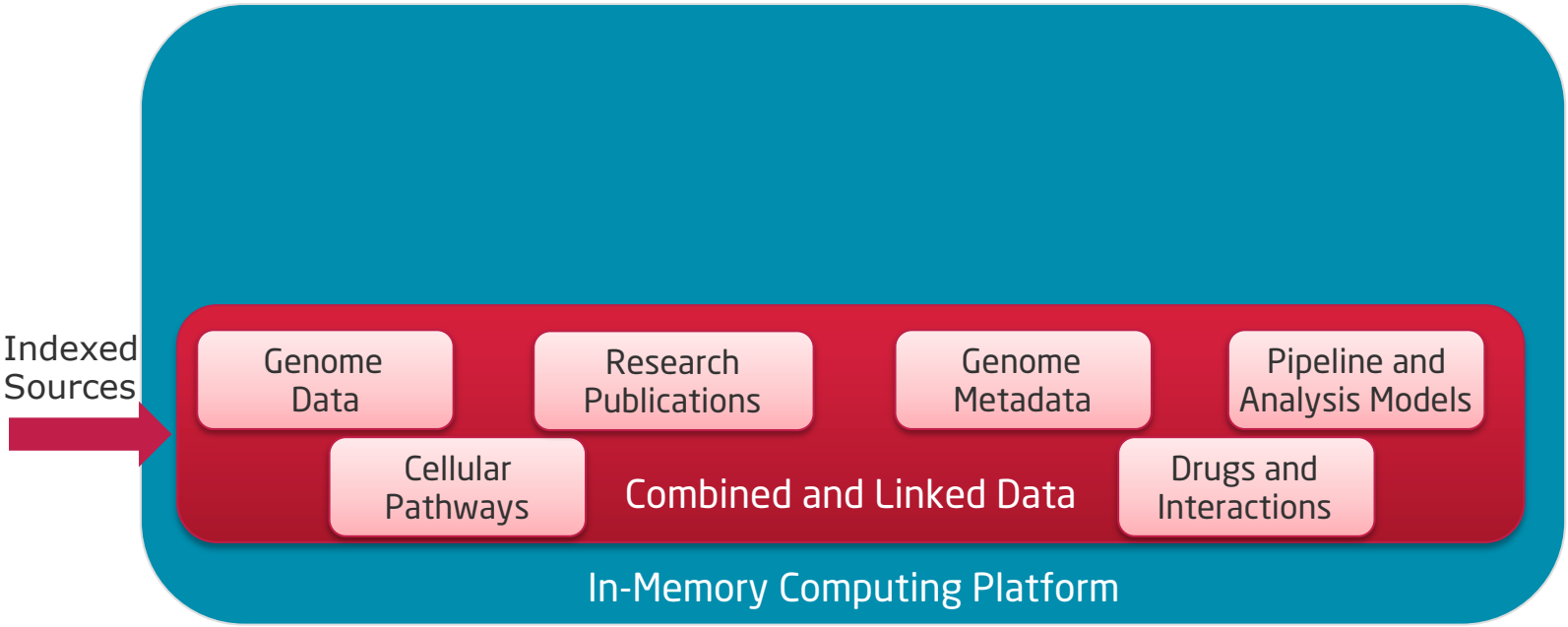
Our Approach:
In-Memory Computing Platform for Big Medical Data

Hasso Plattner Institut



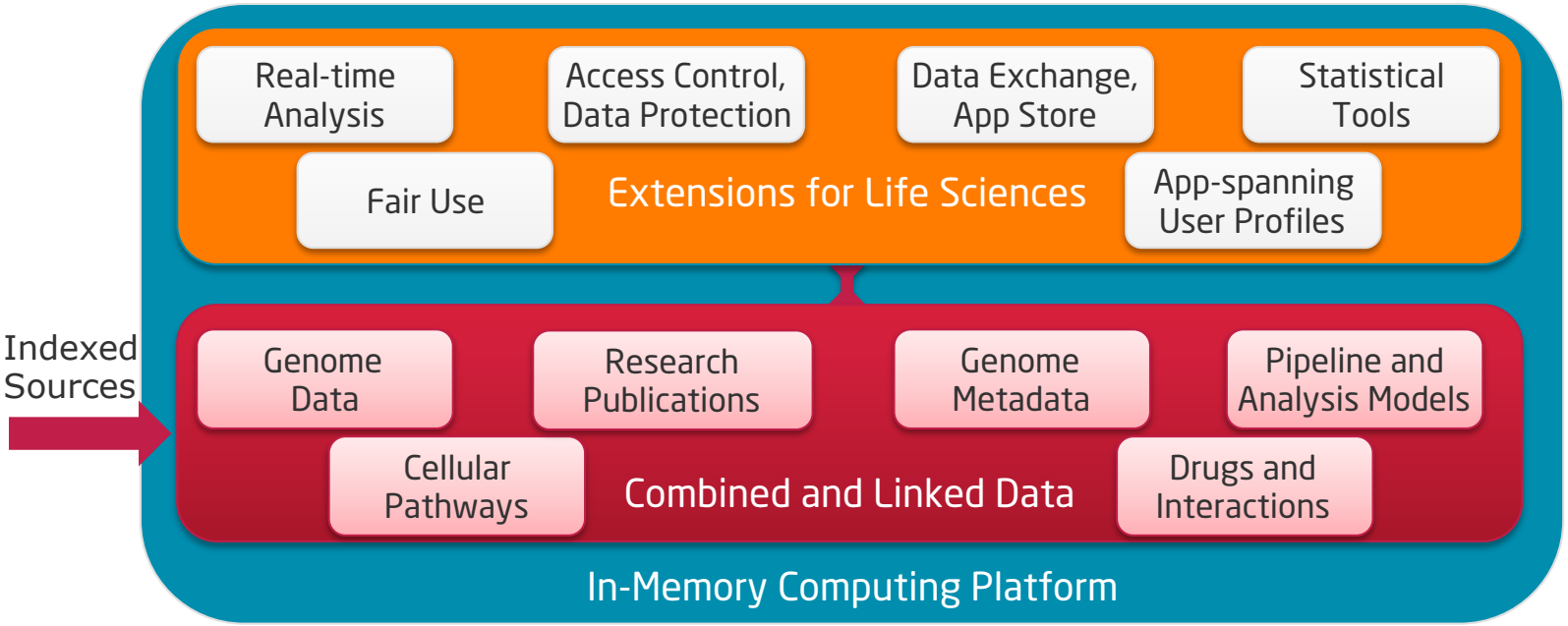
Our Approach:
In-Memory Computing Platform for Big Medical Data

Hasso Plattner Institut



Our Approach:
In-Memory Computing Platform for Big Medical Data

Hasso Plattner Institut

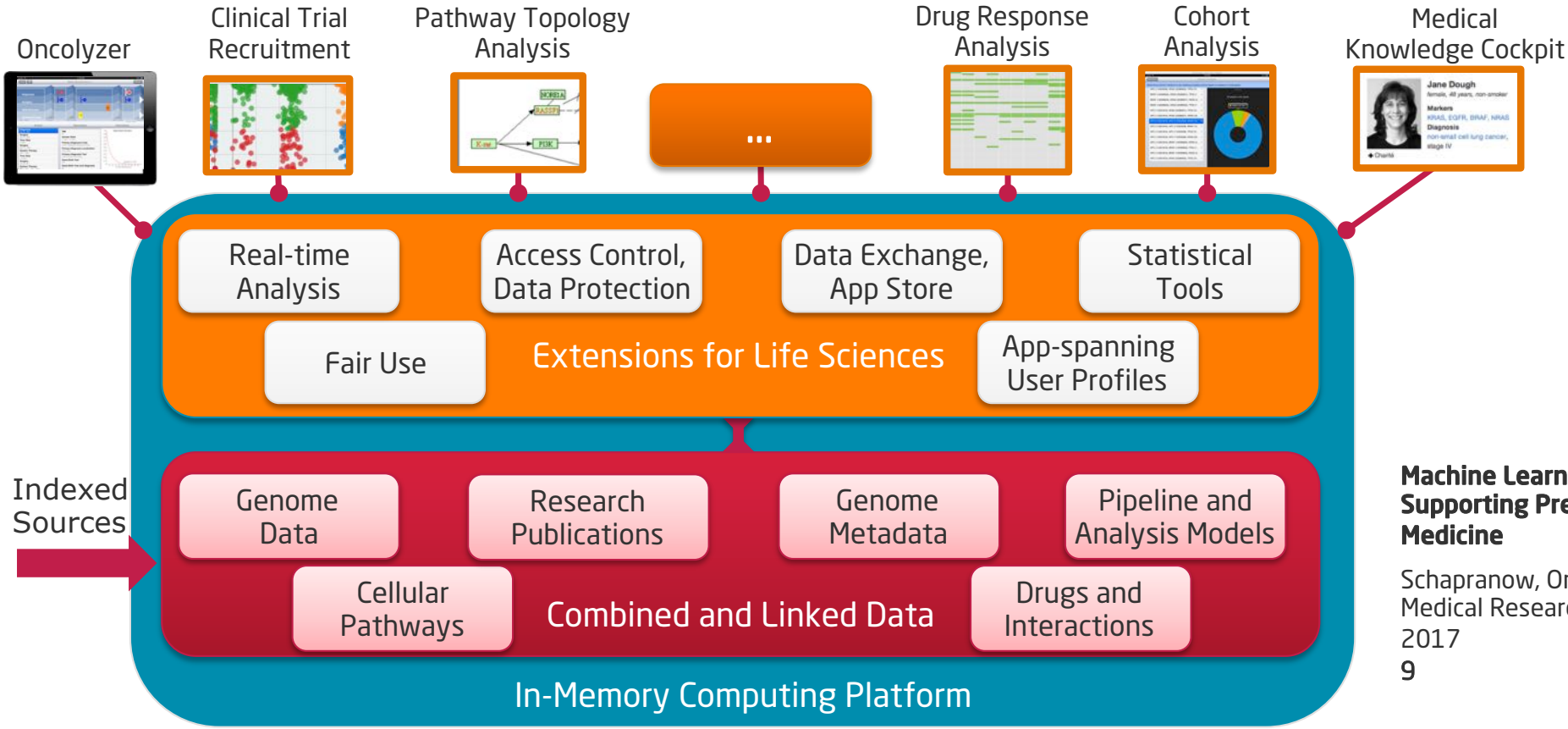


Machine Learning Supporting Precision Medicine

Schapranow, Omics in Medical Research, Dec 5, 2017

Our Approach:
In-Memory Computing Platform for Big Medical Data

Hasso Plattner Institut



Machine Learning Supporting Precision Medicine

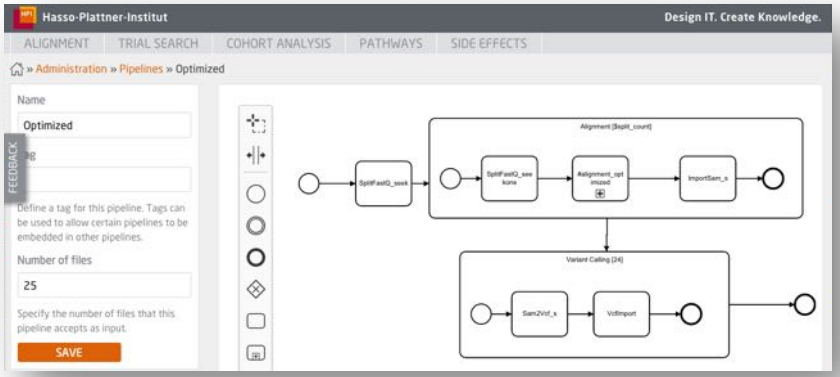
Schapranow, Omics in Medical Research, Dec 5, 2017

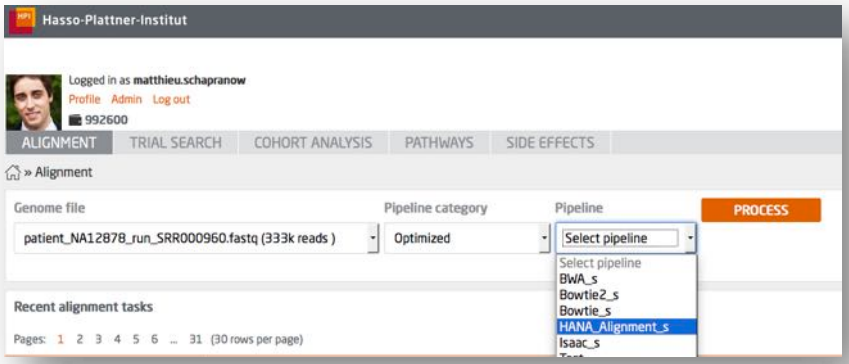
Reproducibility Data Processing
Modeling of Analysis Pipelines

1. Design time (researcher, process expert)
   - Definition of parameterized process model
   - Uses graphical editor and jobs from repository

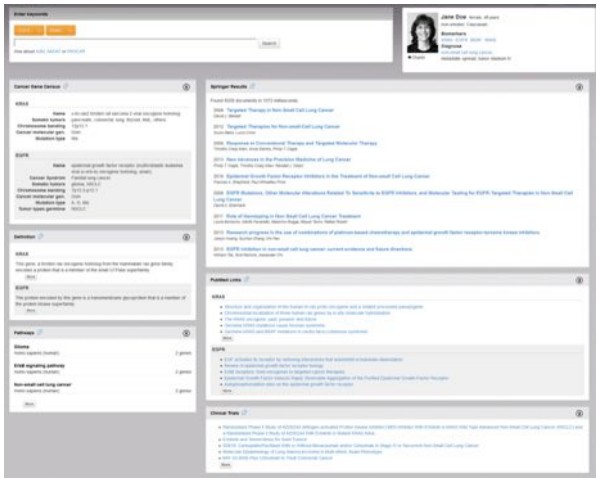
2. Configuration time (researcher, lab assistant)
   - Select model and specify parameters, e.g. aln opts
   - Results in model instance stored in repository

3. Execution time (researcher)
   - Select model instance
   - Specify execution parameters, e.g. input files

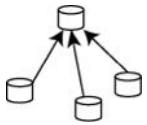




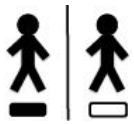
App Example:
Medical Knowledge Cockpit for Patients and Clinicians



- Search for affected genes in distributed and heterogeneous data sources
- Immediate exploration of relevant information, such as
  - Gene descriptions,
  - Molecular impact and related pathways,
  - Scientific publications, and
  - Suitable clinical trials.



Unified access to structured and un-structured data sources



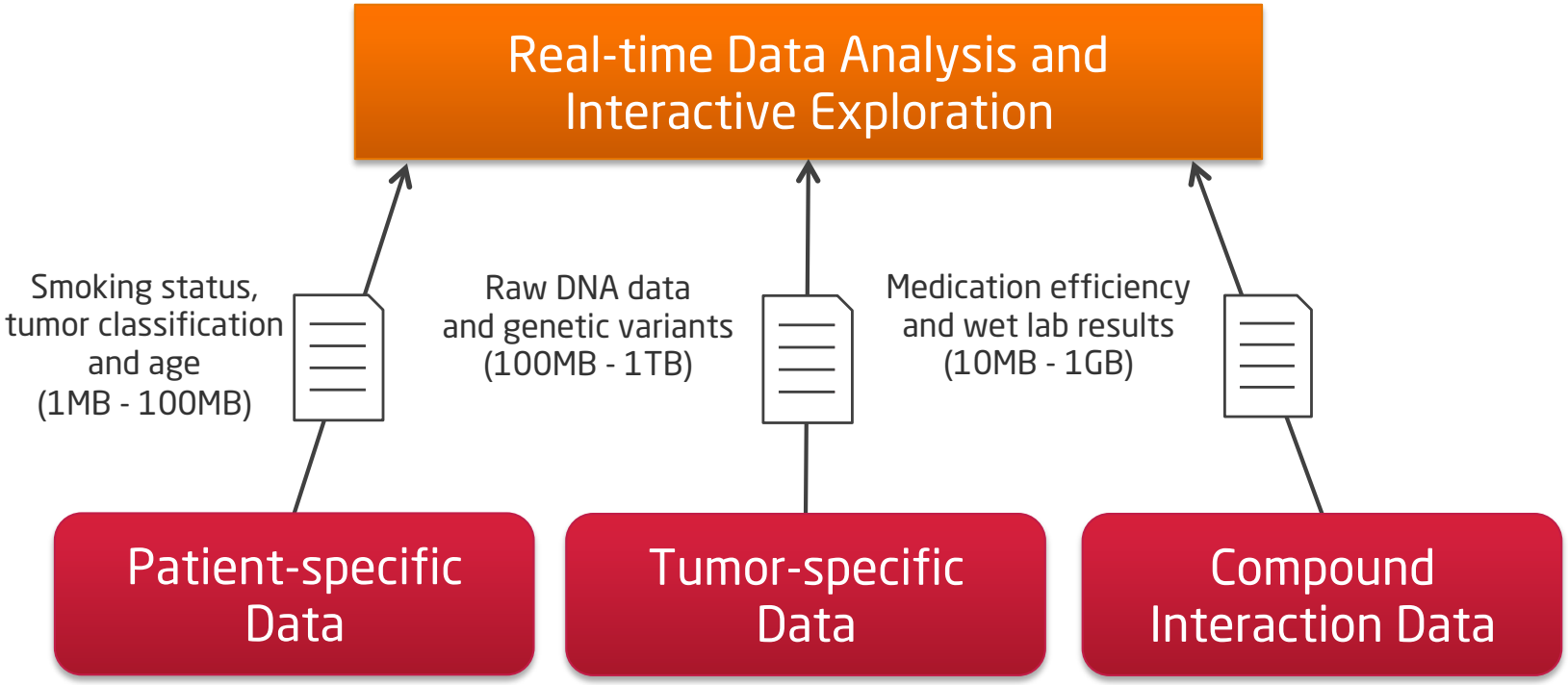
Automatic clinical trial matching build on text analysis features

- No manual searching for hours or days:
  In-memory technology translates searching into interactive finding!

Machine Learning Supporting Precision Medicine

Schapranow, Omics in Medical Research, Dec 5, 2017

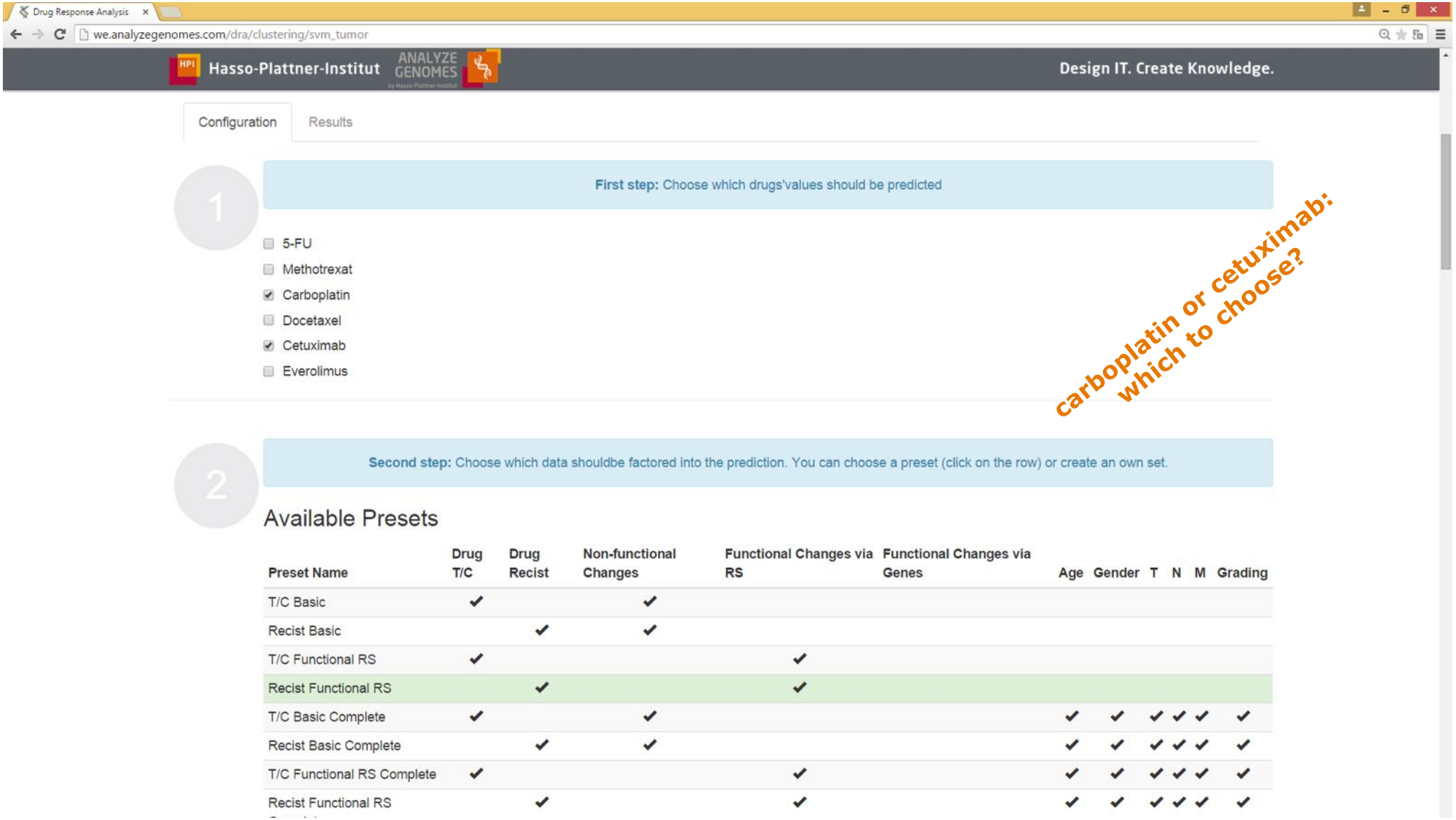
App Example:
Identification of Optimal Chemotherapy

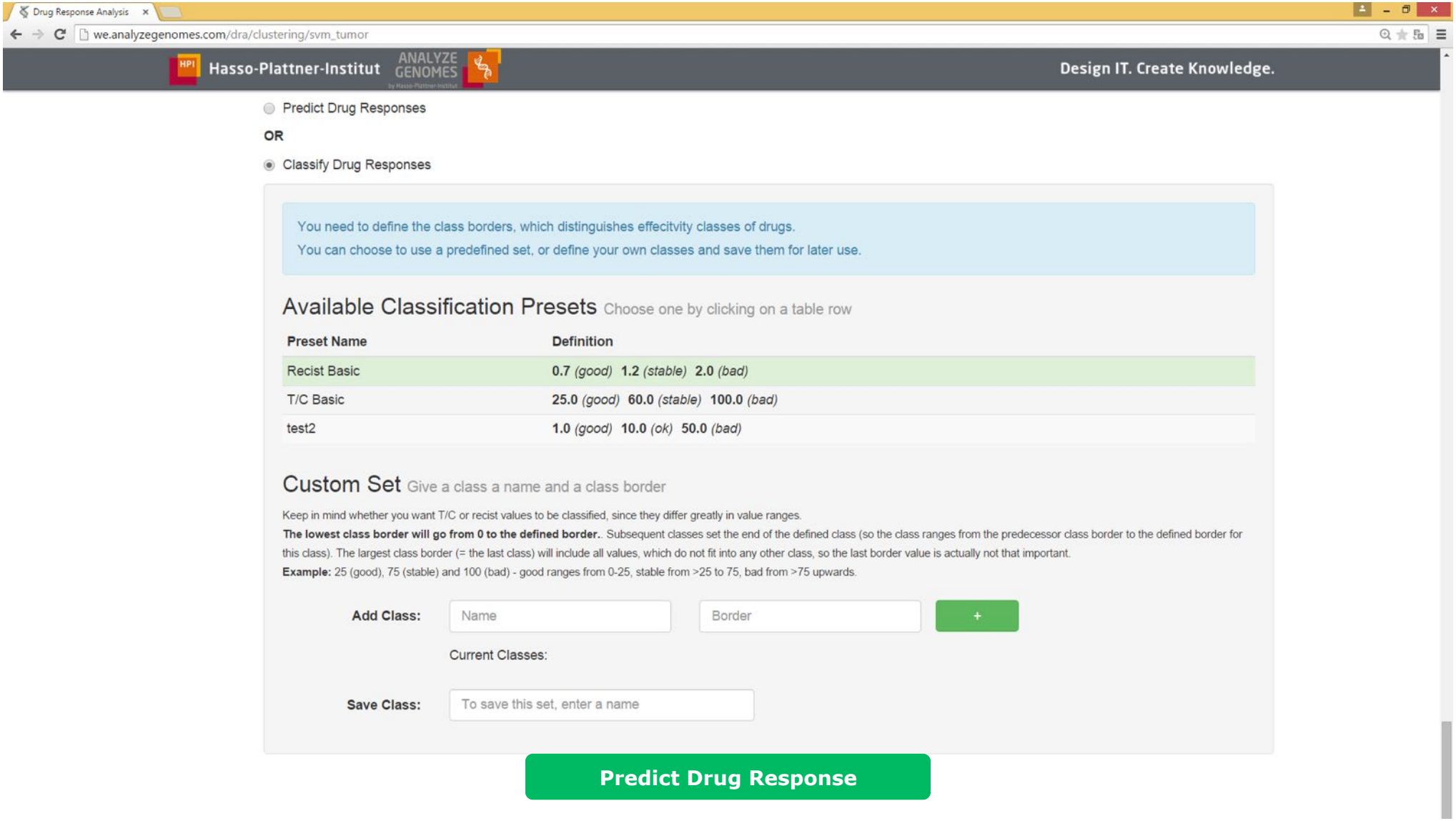


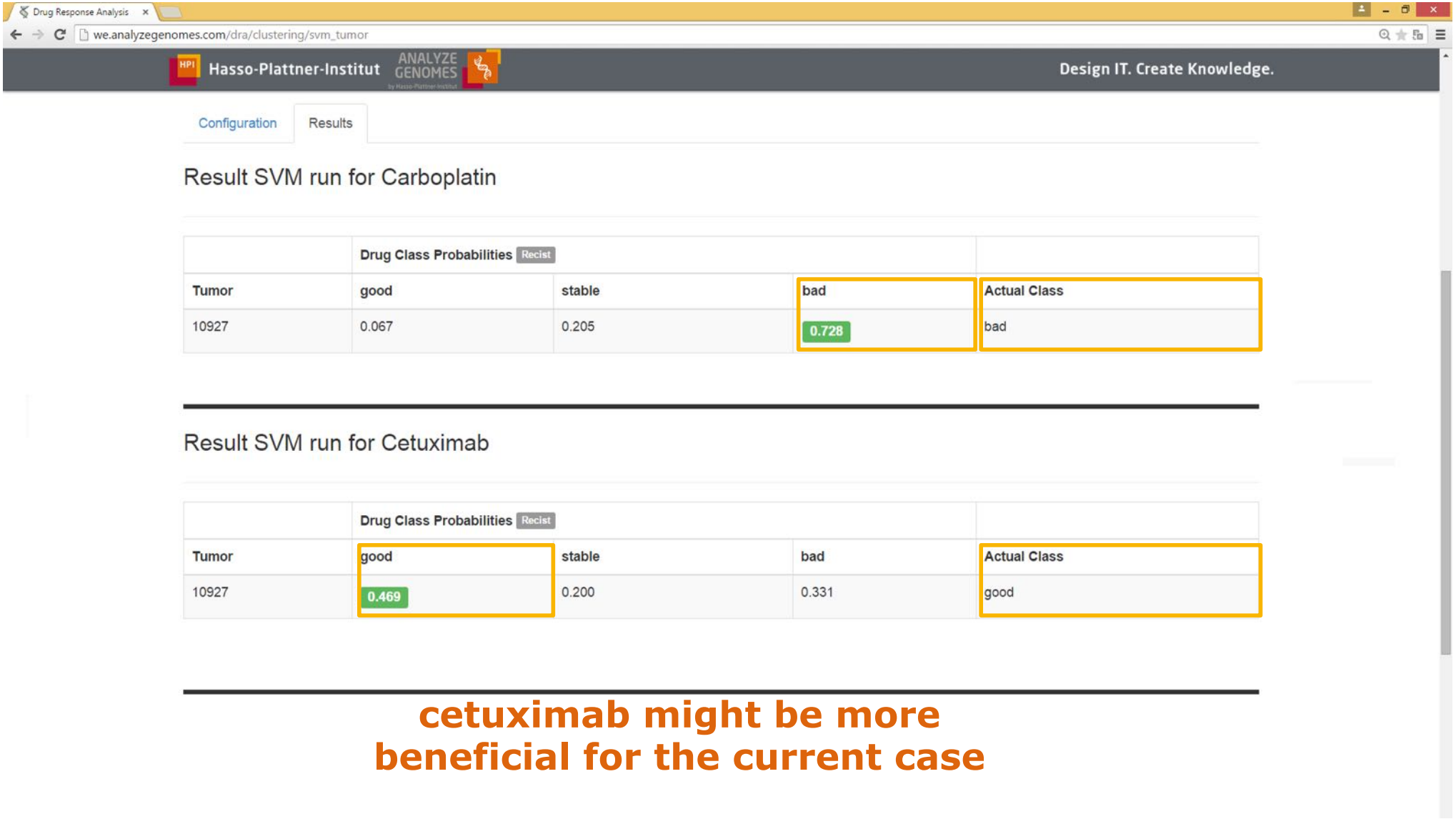
- Honored by the 2015 PerMediCon Award

Machine Learning Supporting Precision Medicine

Schapranow, Omics in Medical Research, Dec 5, 2017


Drug Response Analysis
we.analyzegenomes.com/dra/clustering/svm_tumor
Hasso-Plattner-Institut
ANALYZE GENOMES
Design IT. Create Knowledge.
Configuration
Results
1
First step: Choose which drugs'values should be predicted
5-FU
Methotrexat
Carboplatin
Docetaxel
Cetuximab
Everolimus
carboplatin or cetuximab: which to choose?
2
Second step: Choose which data shouldbe factored into the prediction. You can choose a preset (click on the row) or create an own set.
Available Presets
Preset Name
Drug T/C
Drug Recist
Non-functional Changes
Functional Changes via RS
Functional Changes via Genes
Age
Gender
T
N
M
Grading
T/C Basic
Recist Basic
T/C Functional RS
Recist Functional RS
T/C Basic Complete
Recist Basic Complete
T/C Functional RS Complete
Recist Functional RS


Drug Response Analysis
we.analyzegenomes.com/dra/clustering/svm_tumor
Hasso-Plattner-Institut
ANALYZE GENOMES
Design IT. Create Knowledge.
Predict Drug Responses
OR
Classify Drug Responses
You need to define the class borders, which distinguishes effecitvity classes of drugs.
You can choose to use a predefined set, or define your own classes and save them for later use.
Available Classification Presets Choose one by clicking on a table row
Preset Name
Definition
Recist Basic
0.7 (good) 1.2 (stable) 2.0 (bad)
T/C Basic
25.0 (good) 60.0 (stable) 100.0 (bad)
test2
1.0 (good) 10.0 (ok) 50.0 (bad)
Custom Set Give a class a name and a class border
Keep in mind whether you want T/C or recist values to be classified, since they differ greatly in value ranges.
The lowest class border will go from 0 to the defined border.. Subsequent classes set the end of the defined class (so the class ranges from the predecessor class border to the defined border for this class). The largest class border (= the last class) will include all values, which do not fit into any other class, so the last border value is actually not that important.
Example: 25 (good), 75 (stable) and 100 (bad) - good ranges from 0-25, stable from >25 to 75, bad from >75 upwards.
Add Class:
Name
Border
+
Current Classes:
Save Class:
To save this set, enter a name
Predict Drug Response



cetuximab might be more beneficial for the current case

transkript

Brauchen wir eine Debatte über künstliche Intelligenz?

30.11.2017

Prof. Dr. Roland Benedikter, Center for Advanced Studies, Eurac Research Bozen (Foto: Univ. Breslau); Dr.-Ing. Matthieu Schapranow, Program Manager E-Health & Life Sciences, HPI, Potsdam

Während der Politikanalytiker Roland Benedikter angesichts großen Transformationspotentials eine breite Debatte über künstliche Intelligenz fordert, wollen Forscher sie endlich nutzen dürfen.

Hasso Plattner Institut

Machine Learning Supporting Precision Medicine

Schapranow, Omics in Medical Research, Dec 5, 2017

Stay in Contact!

Dr.-Ing. Matthieu-P. Schapranow
Program Manager E-Health & Life Sciences

Hasso Plattner Institute
August-Bebel-Str. 88
14482 Potsdam, Germany

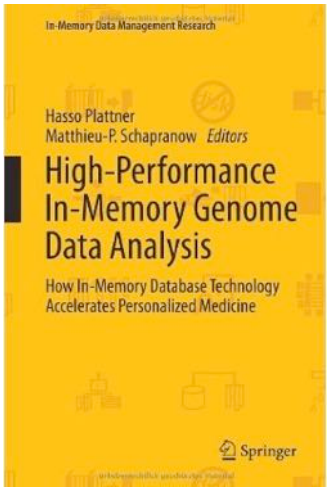
we.analyzegenomes.com
@AnalyzeGenomes

Supported by:

Federal Ministry of Education and Research

Federal Ministry for Economic Affairs and Energy

on the basis of a decision by the German Bundestag



Machine Learning Supporting Precision Medicine

Schapranow, Omics in Medical Research, Dec 5, 2017
17