Hasso Plattner Institut
IT Systems Engineering | Universität Potsdam

How Real-time Analysis turns Big Medical Data into Precision Medicine?

Dr. Matthieu-P. Schapranow

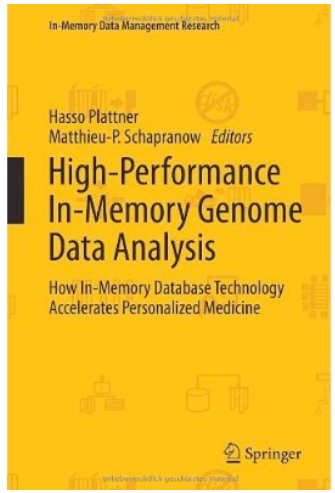
GLOBAL HEALTH, Rome, Italy

August 27, 2014

Important things first:
Where to find additional information?

Hasso Plattner Institut

- Online: Visit http://we.analyzegenomes.com for latest research results, tools, and news
- Offline: Read more about it, e.g. High-Performance In-Memory Genome Data Analysis: How In-Memory Database Technology Accelerates Personalized Medicine, In-Memory Data Management Research, Springer, ISBN: 978-3-319-03034-0, 2014
- In Person: Join us for the **International Workshop on Big Data in Bioinformatics and Healthcare Informatics (BBH14) in Washington D.C. on Oct 27, 2014 (http://bbh14.analyzegenomes.com)**



How Real-time Analysis turns Big Medical Data into Precision Medicine?

Dr. Schapranow, GLOBAL HEALTH, Aug 27, 2014

Who you are dealing with?

158,952 km

This is the number of kilometres that you have already clocked up on your travels.

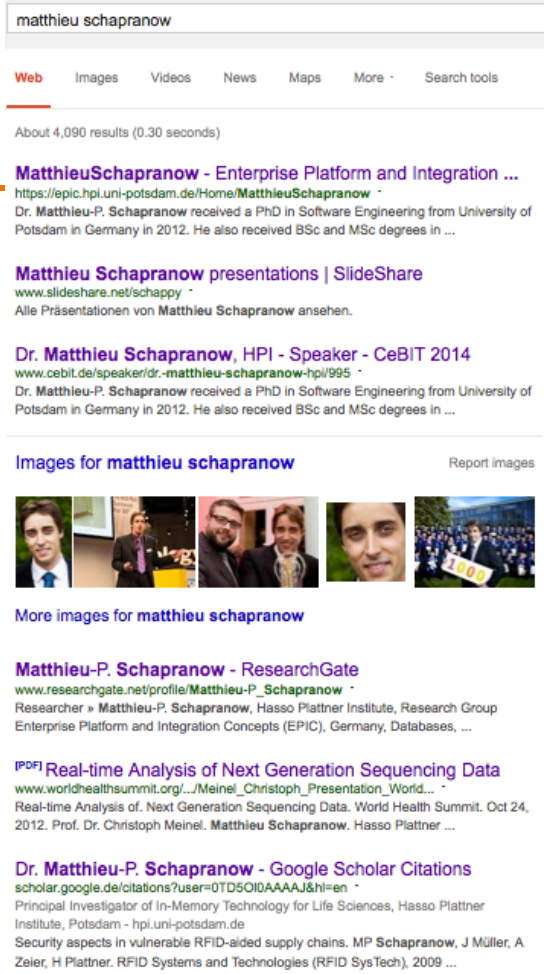
Flights around earth	3.966
Flights to moon	0.414
Shortest flight	FRA-AMS (367 km)
Longest flight	MUC-SFO (9,437 km)
Northernmost airport	Dublin (DUB)
Southernmost airport	Miami International, FL (MIA)

Dr.-Ing. Matthieu Schapranow
Program Manager E-Health at Hasso Plattner Institute
Berlin Area, Germany | Research

Previous: Hasso Plattner Institute, SAP, Travel IQ
Education: Universität Potsdam

Complete your profile | Edit Profile | 441 connections

- Software Engineer by training (B.Sc., M.Sc., PhD)
- 2006-2014 with SAP
- Since 2007 at Chair of Prof. Hasso Plattner, HPI
- Since 2009 focusing on Life Sciences / E-Health



Hasso Plattner Institut

EUROPEAN LIFE SCIENCE AWARDS
Barcelona 13.05.2014

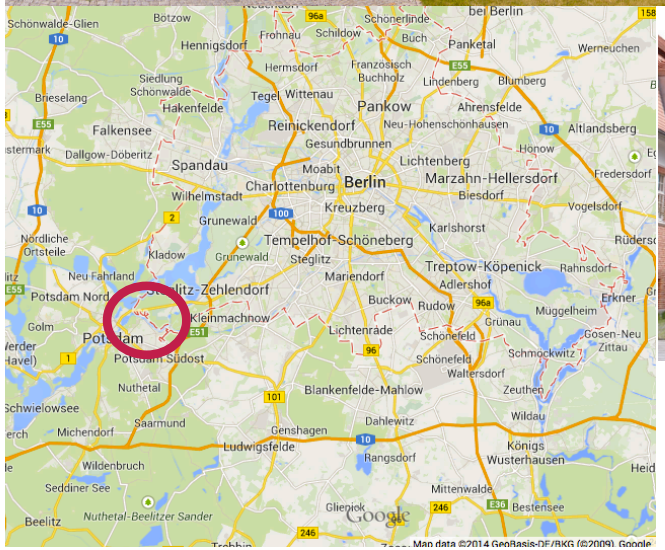
INNOVATIONSPREISTRÄGER BERLIN BRANDENBURG 2012

How Real-time Analysis turns Big Medical Data into Precision Medicine?

Dr. Schapranow, GLOBAL HEALTH, Aug 27, 2014

Where do I work?
Hasso Plattner Institute, Potsdam, Germany

Hasso Plattner Institut



How Real-time Analysis turns Big Medical Data into Precision Medicine?

Dr. Schapranow, GLOBAL HEALTH, Aug 27, 2014

Hasso Plattner Institute
Key Facts

- Founded as a public-private partnership in 1998 in Potsdam near Berlin, Germany
- Institute belongs to the University of Potsdam
- Ranked 1st in CHE since 2009
- 500 B.Sc. and M.Sc. students
- 10 professors, 150 PhD students

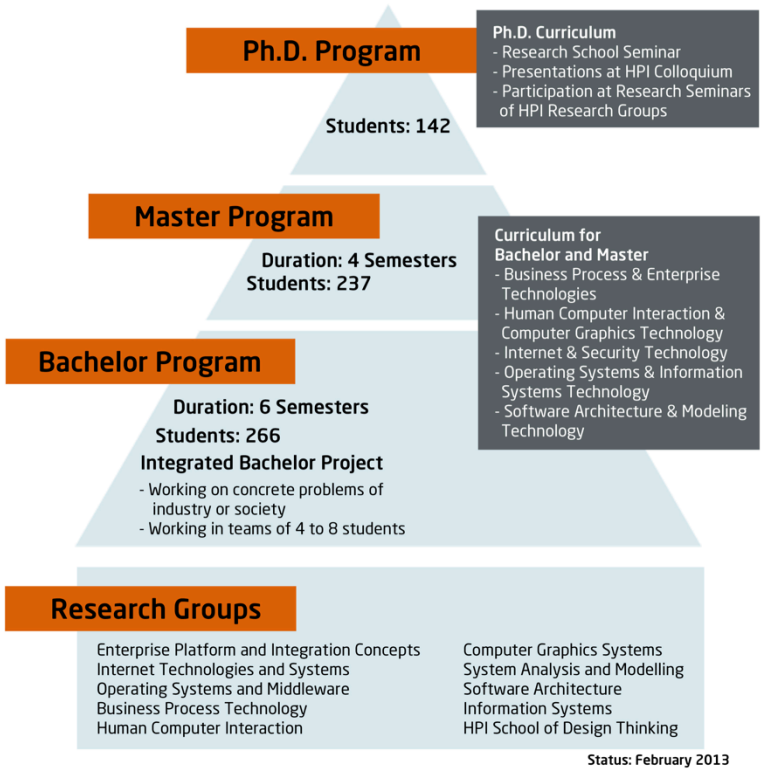
- Course of study: IT Systems Engineering

How Real-time Analysis turns Big Medical Data into Precision Medicine?

Dr. Schapranow, GLOBAL HEALTH, Aug 27, 2014

Hasso Plattner Institute Programs

- Full university curriculum
- Bachelor (6 semesters)
- Master (4 semesters)
- Orthogonal Activities:
  - E-Health Consortium
  - School of Design Thinking
  - Research School



How Real-time Analysis turns Big Medical Data into Precision Medicine?

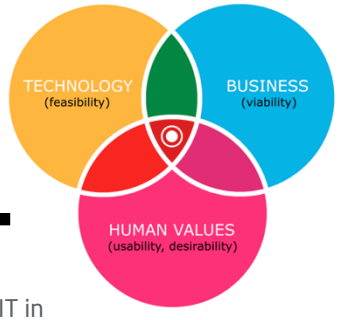
Dr. Schapranow, GLOBAL HEALTH, Aug 27, 2014

Hasso Plattner Institute
Enterprise Platform and Integration Concepts Group

Hasso Plattner Institut

Chair of Prof. Dr. h.c. Hasso Plattner

- Research focuses on the technical aspects of enterprise software and design of complex applications
  - In-Memory Data Management for Enterprise Applications
  - Enterprise Application Programming Model
  - **Scientific Data Management**
  - Human-Centered Software Design and Engineering
- Industry cooperations, e.g. SAP, Siemens, Audi, and EADS
- Research cooperations, e.g. Stanford, MIT, and Berkeley



Partner of Stanford Center for Design Research

Partner of MIT in Supply Chain Innovation and CSAIL

Partner at UC Berkeley RAD / AMP Lab

Partner of SAP AG

How Real-time Analysis turns Big Medical Data into Precision Medicine?

Dr. Schapranow, GLOBAL HEALTH, Aug 27, 2014

The Setting
Actors in Oncology

Hasso Plattner Institut

- **Patients**

  - Individual anamnesis, family history, and background
  - Require fast access to individualized therapy

- **Clinicians**

  - Identify root and extent of disease using laboratory tests
  - Evaluate therapy alternatives, adapt existing therapy

- **Researchers**

  - Conduct laboratory work, e.g. analyze patient samples
  - Create new research findings and come-up with treatment alternatives

How Real-time Analysis turns Big Medical Data into Precision Medicine?

Dr. Schapranow, GLOBAL HEALTH, Aug 27, 2014

Our Motivation
Make Precision Medicine Come Routine in Real Life

- Motivation: Can we enable **patients** to:
  - Understand and monitor their diseases to document the impact on their lives,
  - Receive latest information about their (chronic) diseases,
  - Cooperatively exchange with physicians and patients to improve quality of living

How Real-time Analysis turns Big Medical Data into Precision Medicine?

Dr. Schapranow, GLOBAL HEALTH, Aug 27, 2014

Our Motivation
Make Precision Medicine Come Routine in Real Life (cont'd)

- Motivation: Can we enable **clinicians** to take their therapy decisions:
  - Incorporating all available specifics about each individual patient,
  - Referencing latest lab results and worldwide medical knowledge, and
  - Interactively during their ward round?

How Real-time Analysis turns Big Medical Data into Precision Medicine?

Dr. Schapranow, GLOBAL HEALTH, Aug 27, 2014

IT Challenge: How to Integrate Distributed and Heterogeneous Sources of Big Medical Data

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

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

Cancer patient records
>160k records at NCT

How Real-time Analysis turns Big Medical Data into Precision Medicine?

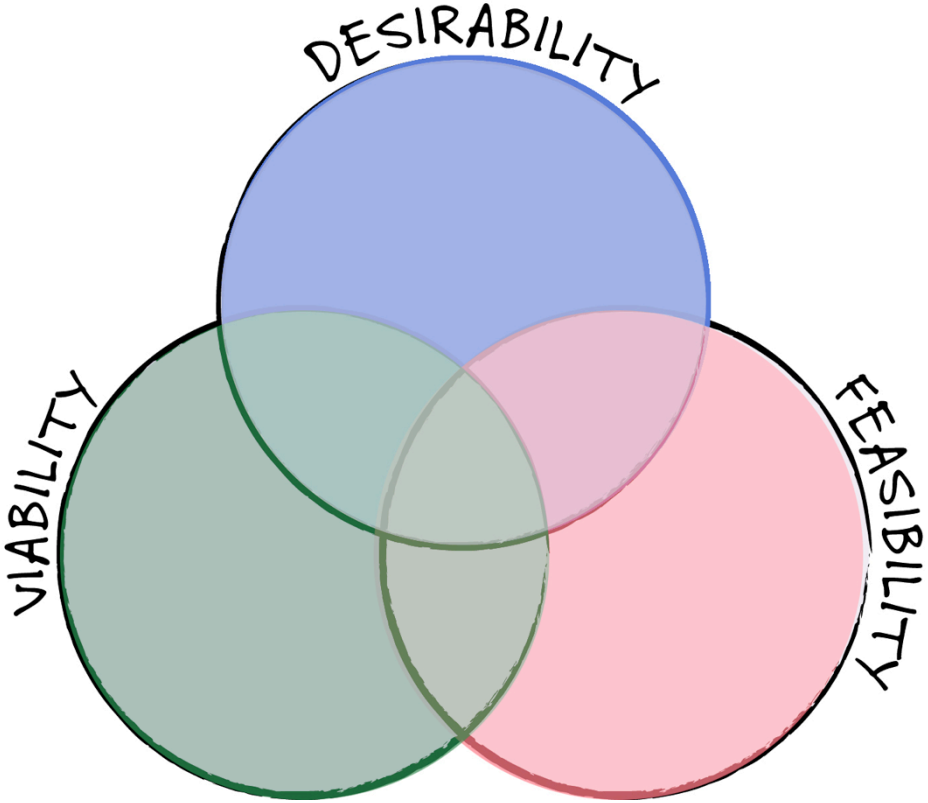
Dr. Schapranow, GLOBAL HEALTH, Aug 27, 2014

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

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

Our Methodology
Design Thinking Methodology

Hasso Plattner Institut



How Real-time Analysis turns Big Medical Data into Precision Medicine?

Dr. Schapranow, GLOBAL HEALTH, Aug 27, 2014

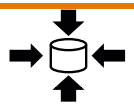
Our Technology: In-Memory Database Technology Enabling Real-time Data Analysis

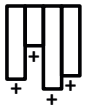
Hasso Plattner Institut

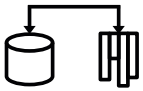

Combined column and row store


Map/Reduce

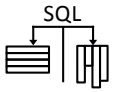

Single and multi-tenancy


Lightweight compression

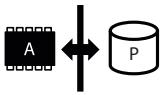

Insert only for time travel


Real-time replication

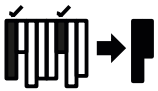

Working on integers

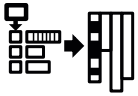


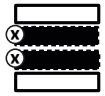
SQL interface on columns and rows



Active/passive data store


Minimal projections


Group key

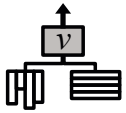

Reduction of software layers


Dynamic multi-threading

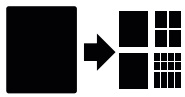

Bulk load of data

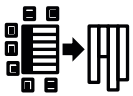

Object-relational mapping

Text retrieval and extraction engine

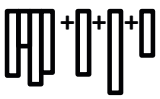


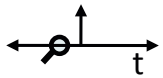
No aggregate tables


Data partitioning


Any attribute as index

No disk


On-the-fly extensibility



Analytics on historical data

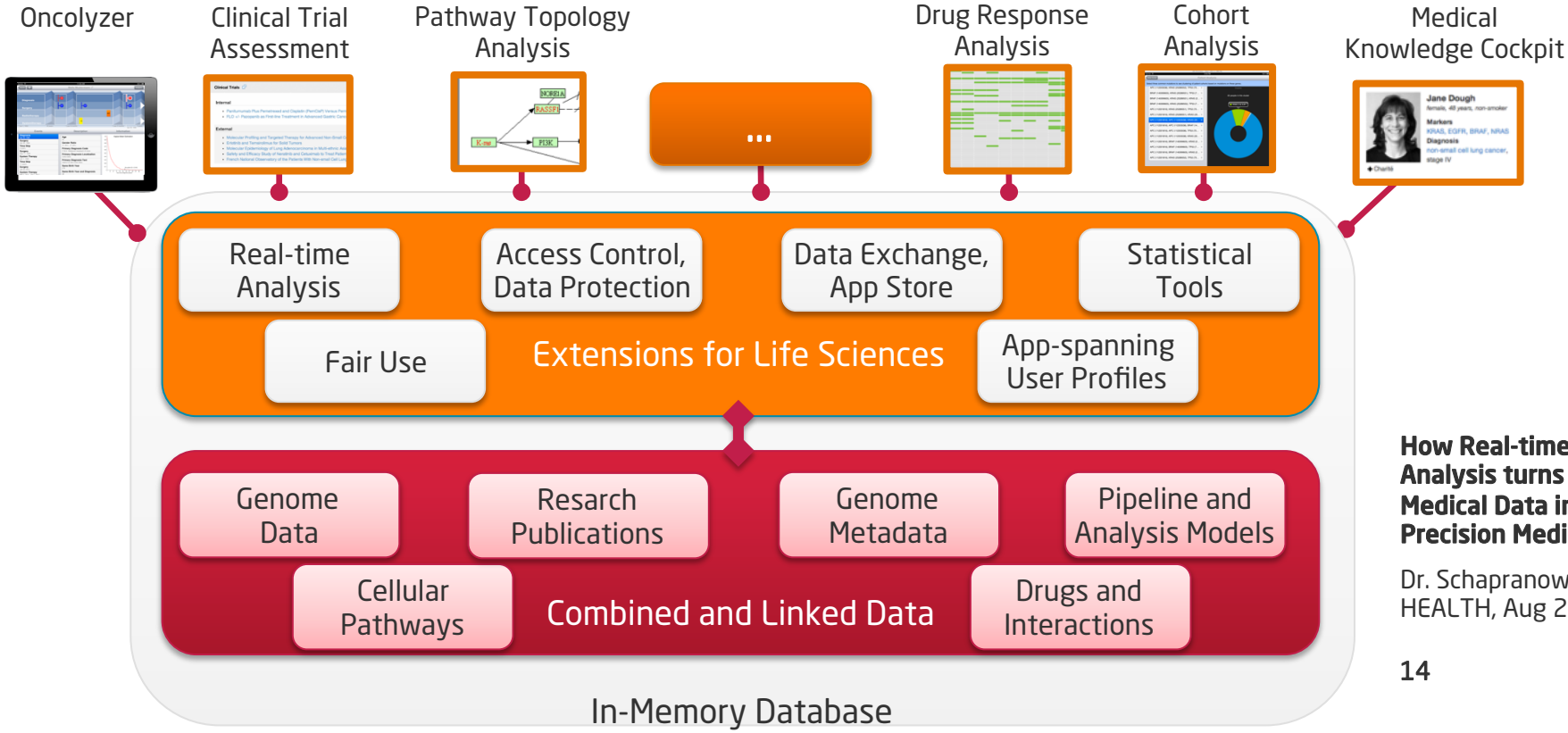

Multi-core/ parallelization

How Real-time Analysis turns Big Medical Data into Precision Medicine?

Dr. Schapranow, GLOBAL HEALTH, Aug 27, 2014

Our Approach: Analyze Genomes – A Cloud Platform Enabling Real-time Analysis of Big Medical Data

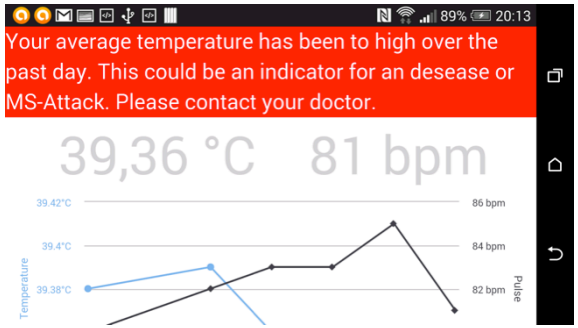
Hasso Plattner Institut

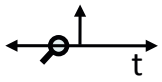


How Real-time Analysis turns Big Medical Data into Precision Medicine?

Dr. Schapranow, GLOBAL HEALTH, Aug 27, 2014

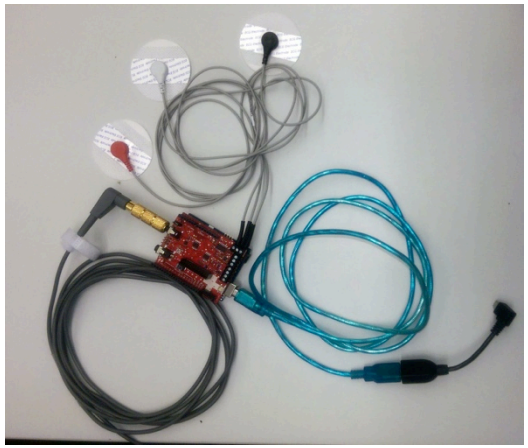
Real-time Processing of Event Data from Medical Sensors





Comparison of waveform data with history of similar patients

- Processing of sensor data, e.g. from Intensive Care Units (ICUs) or wearable sensor devices (quantify self)
- Multi-modal real-time analysis to detect indicators for severe events, such as heart attacks or strokes
- Incorporates machine-learning algorithms to detect severe events and to inform clinical personnel in time
- Successfully tested with 100 Hz event rate, i.e. sufficient for ICU use



Hasso Plattner Institut

CSAIL

SAP

HST

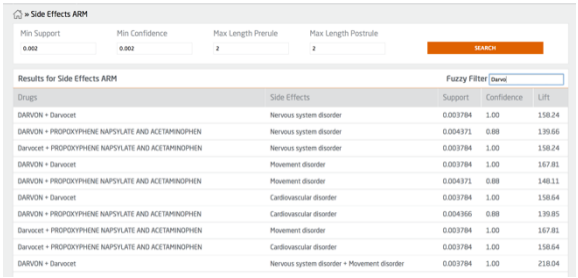
Harvard–MIT Health Sciences & Technology

Future SOC Lab

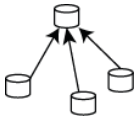
How Real-time Analysis turns Big Medical Data into Precision Medicine?

Dr. Schapranow, GLOBAL HEALTH, Aug 27, 2014

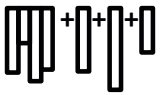
Drug Safety
Statistical Analysis of Drug Side Effects Data



- Combines confirmed side effect data from different data sources
- Interactive statistical analysis, e.g. apriori rules, to discover still unknown interactions
- Integrates personal prescription data and directly report side effects
- Work together with your doctor to prevent interaction with already prescribed drugs



Unified access to international side effect data

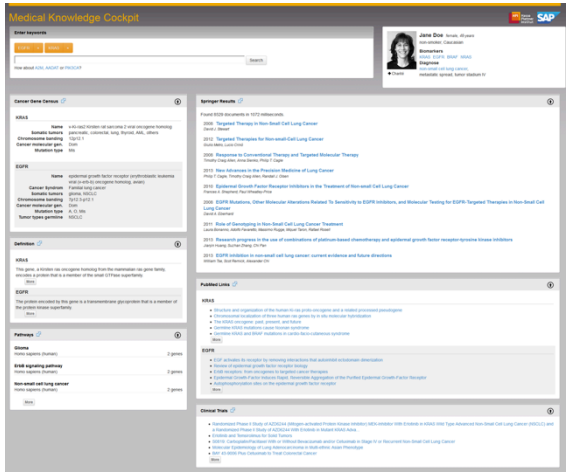


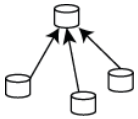
On-the-fly extension of database schema to add side effect databases

How Real-time Analysis turns Big Medical Data into Precision Medicine?

Dr. Schapranow, GLOBAL HEALTH, Aug 27, 2014

Medical Knowledge Cockpit





Unified access to structured and un-structured data sources

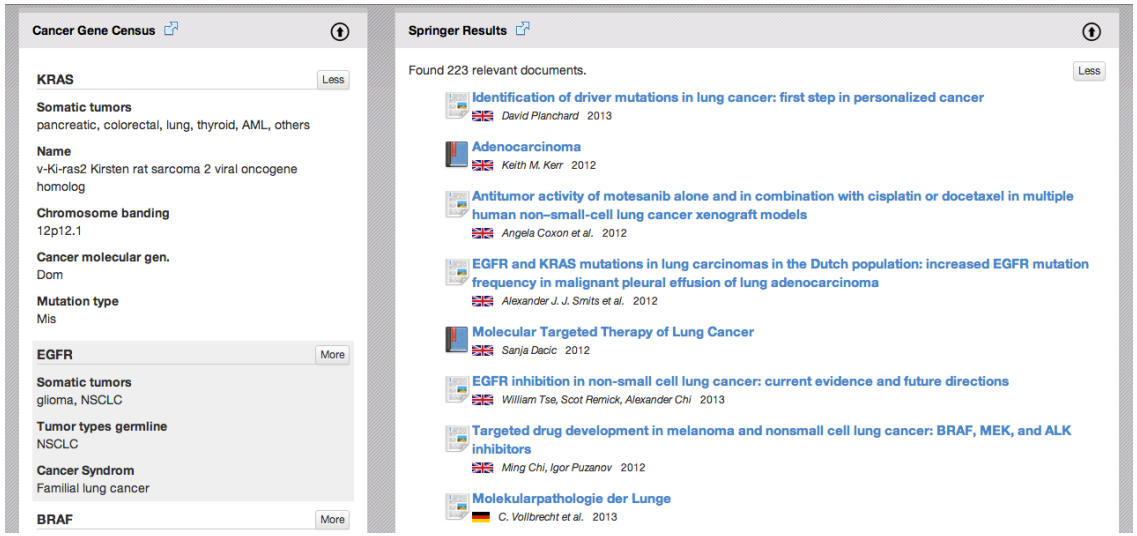
Automatic clinical trial matching build on text analysis features

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

How Real-time Analysis turns Big Medical Data into Precision Medicine?

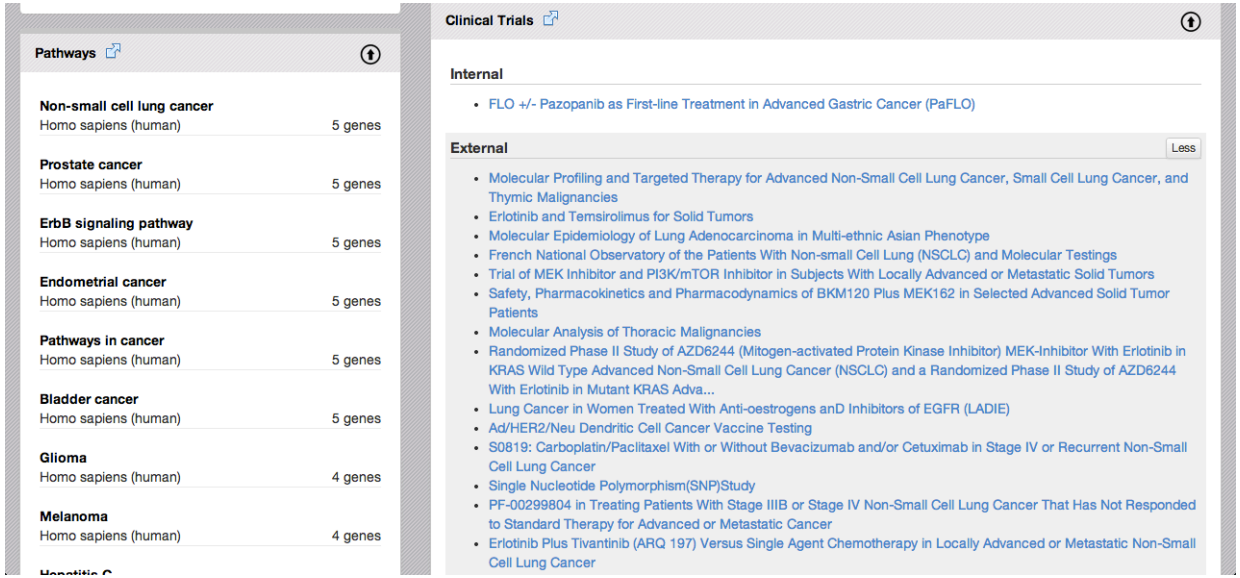
Dr. Schapranow, GLOBAL HEALTH, Aug 27, 2014

Medical Knowledge Cockpit
Publications



- In-place preview of relevant data, such as publications and publication meta data
- Incorporating individual filter settings, e.g. additional search terms

Medical Knowledge Cockpit
Latest Clinical Trials

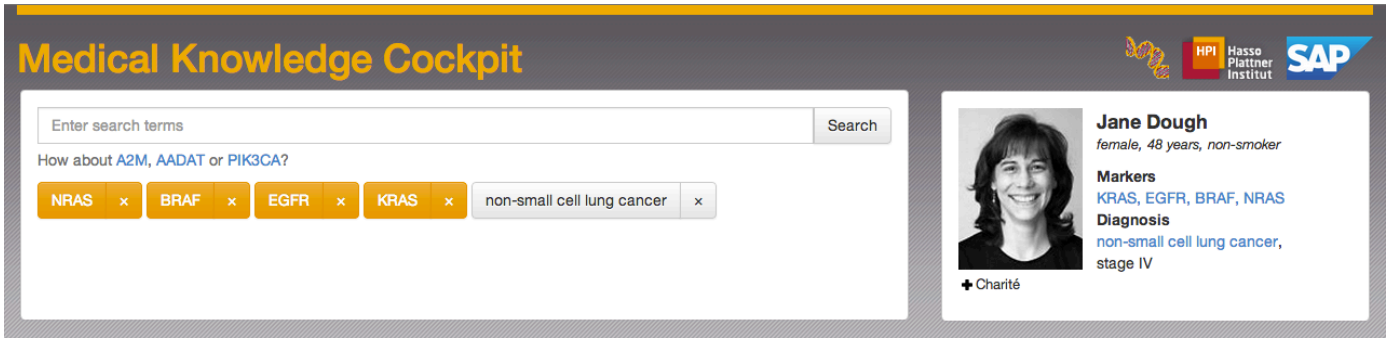


- Personalized clinical trials, e.g. by incorporating patient specifics
- Classification of internal/external trials based on treating institute

How Real-time Analysis turns Big Medical Data into Precision Medicine?

Dr. Schapranow, GLOBAL HEALTH, Aug 27, 2014

Medical Knowledge
Seamless Integration of Patient Specifics

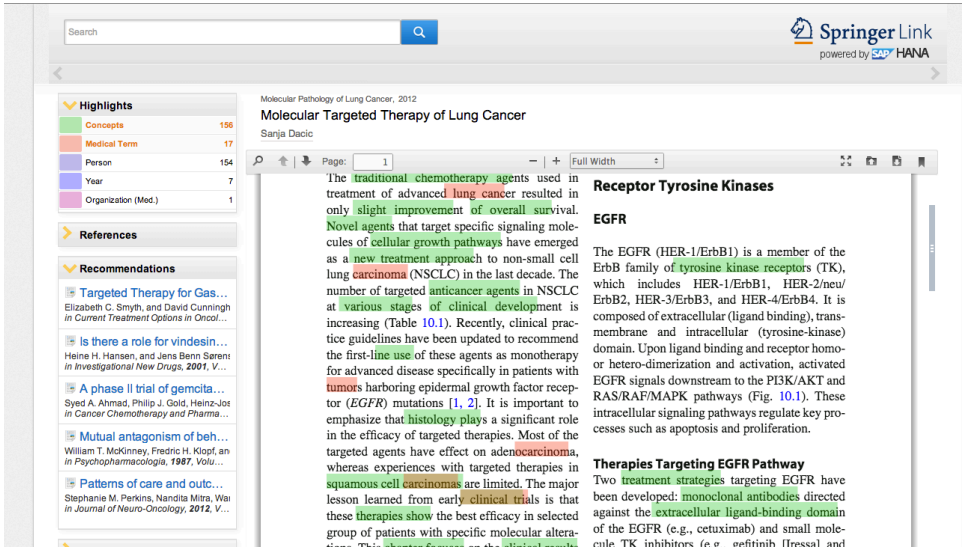


- Google-like user interface for searching data
- Seamless integration of individual EMR data
- Search various sources for biomarkers, literature, and diseases

How Real-time Analysis turns Big Medical Data into Precision Medicine?

Dr. Schapranow, GLOBAL HEALTH, Aug 27, 2014

Medical Knowledge Cockpit
Publications

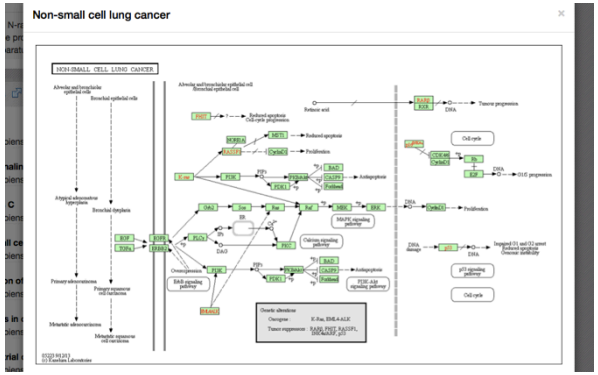


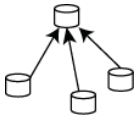
- Interactively explore relevant publications, e.g. PDFs
- Improved ease of exploration, e.g. by highlighted medical terms and relevant concepts

How Real-time Analysis turns Big Medical Data into Precision Medicine?

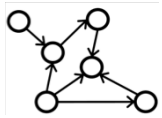
Dr. Schapranow, GLOBAL HEALTH, Aug 27, 2014

Medical Knowledge Cockpit
Pathway Topology Analysis





Unified access to multiple formerly disjoint data sources



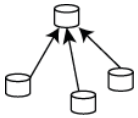
Pathway analysis of genetic variants with graph engine

- Search in pathways is limited to “is a certain element contained” today
- Integrated >1,5k pathways from international sources, e.g. KEGG, HumanCyc, and WikiPathways, into HANA
- Implemented graph-based topology exploration and ranking based on patient specifics
- Enables interactive identification of possible dysfunctions affecting the course of a therapy before its start

How Real-time Analysis turns Big Medical Data into Precision Medicine?

Dr. Schapranow, GLOBAL HEALTH, Aug 27, 2014

Medical Knowledge Cockpit
Search in Structured and Unstructured Medical Data



Unified access to structured and unstructured data sources

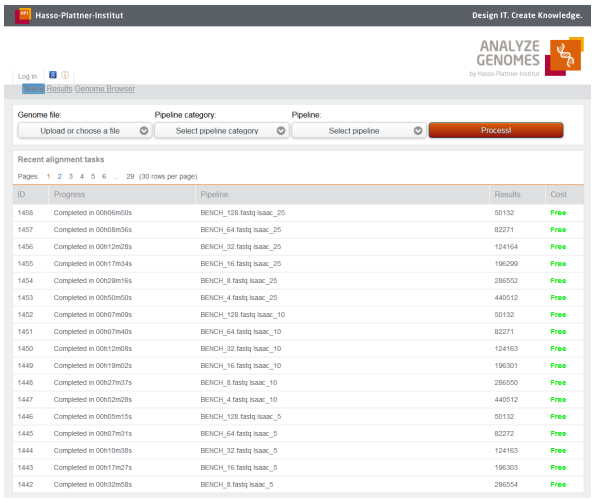
Clinical trial matching using text analysis features

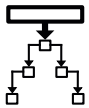
- Extended text analysis feature by medical terminology
  - Genes (122,975 + 186,771 synonyms)
  - Medical terms and categories (98,886 diseases, 47 categories)
  - Pharmaceutical ingredients (7,099)
- Indexed clinicaltrials.gov database (145k trials/ 30,138 recruiting)
- Extracted, e.g., 320k genes, 161k ingredients, 30k periods
- Select studies based on multiple filters in less than 500ms

How Real-time Analysis turns Big Medical Data into Precision Medicine?

Dr. Schapranow, GLOBAL HEALTH, Aug 27, 2014

Cloud-based Services for Processing of DNA Data





Standardized Modeling and runtime environment for analysis pipelines

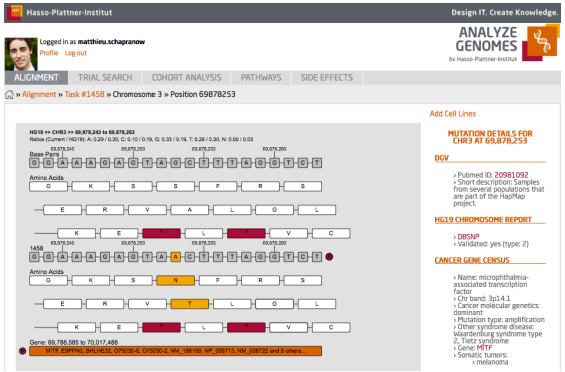
- Control center for processing of raw DNA data, such as FASTQ, SAM, and VCF
- Personal user profile guarantees privacy of uploaded and processed data
- Supports reproducible research process by storing all relevant process parameters
- Implements prioritized data processing and fair use, e.g. per department or per institute
- Supports additional service, such as data annotations, billing, and sharing for all Analyze Genomes services
- **Honored by the 2014 European Life Science Award**

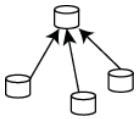
EUROPEAN LIFE SCIENCE AWARDS
Barcelona 13.05.2014

How Real-time Analysis turns Big Medical Data into Precision Medicine?

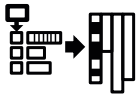
Dr. Schapranow, GLOBAL HEALTH, Aug 27, 2014

Interactive Genome Browser





Unified access to multiple formerly disjoint data sources



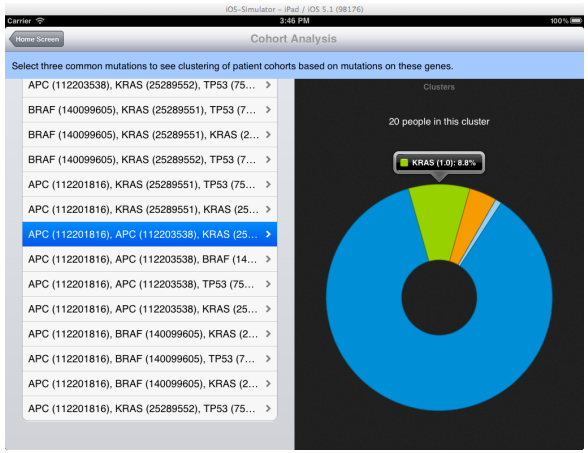
Matching of genetic variants and relevant annotations

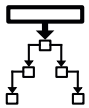
- Genome Browser enables interactive comparison of multiple genomes
- Combined knowledge by integrating latest international annotations and literature, e.g. from NCBI, dbSNP, and UCSC
- Detailed exploration of genome locations and existing associations
- Ranked variants, e.g. accordingly to known diseases
- Links always back to primary data sources to guarantee validity of discovered findings

How Real-time Analysis turns Big Medical Data into Precision Medicine?

Dr. Schapranow, GLOBAL HEALTH, Aug 27, 2014

Analysis of Patient Cohorts





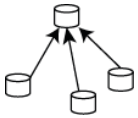
Fast clustering directly performed within the in-memory database

- In a patient cohort, a subset does not respond to therapy – why?
- Clustering using various statistical algorithms, such as k-means or hierarchical clustering
- Calculation of all locus combinations in which at least 5% of all TCGA participants have mutations: 200ms for top 20 combinations
- Individual clusters are calculated in parallel directly within the database
- K-means algorithm: 50ms (PAL) vs. 500ms (R)

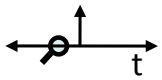
How Real-time Analysis turns Big Medical Data into Precision Medicine?

Dr. Schapranow, GLOBAL HEALTH, Aug 27, 2014

Oncolyzer



Unified access to formerly disjoint oncological data sources



Flexible analysis on patient's longitudinal data

- Research initiative for exchanging relevant tumor data to improve personalized treatment
- Real-time analysis of tumor data in seconds instead of hours
- Information available at your fingertips: In-memory technology on mobile devices, e.g. iPad
- Interdisciplinary cooperation between clinicians, clinical researchers, and software engineers
- Honored with the 2012 Innovation Award of the German Capitol Region

Hasso Plattner Institut

CHARITÉ

SAP

INNOVATIONSPREISTRÄGER 2012 BERLIN BRANDENBURG

How Real-time Analysis turns Big Medical Data into Precision Medicine?

Dr. Schapranow, GLOBAL HEALTH, Aug 27, 2014

Oncolyzer
Patient Details Screen

- Combines patient's longitudinal time series data with individual analysis results
- Real-time analysis across hospital-wide data using always latest data when details screen is accessed
- http://epic.hpi.uni-potsdam.de/Home/HanaOncolyzer

How Real-time Analysis turns Big Medical Data into Precision Medicine?

Dr. Schapranow, GLOBAL HEALTH, Aug 27, 2014

Oncolyzer
Patient Analysis Screen

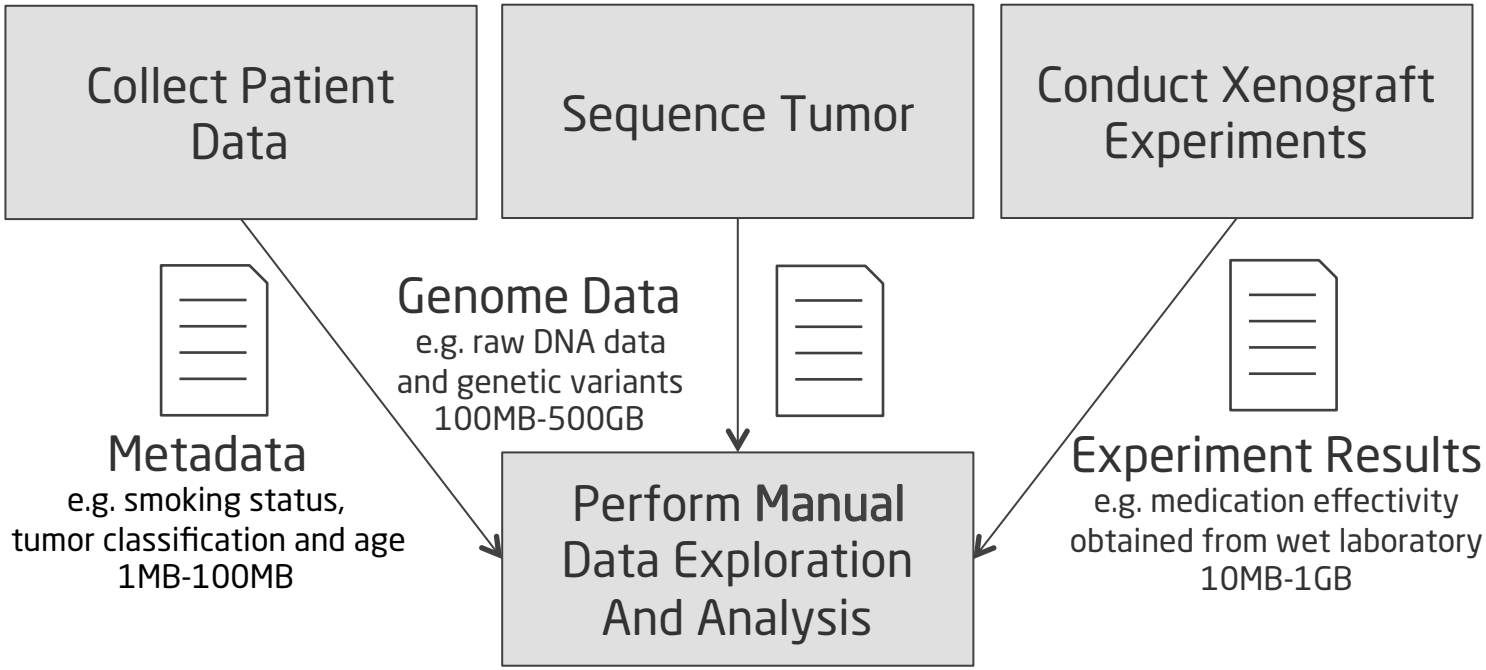
- Allows real-time analysis on complete patient cohort
- Supports identification of clinical trial participants based on their individual anamnesis
- Flexible filters and various chart types allow graphical exploration of data on mobile devices

How Real-time Analysis turns Big Medical Data into Precision Medicine?

Dr. Schapranow, GLOBAL HEALTH, Aug 27, 2014

Drug Response Analysis
Data Sources and Matching

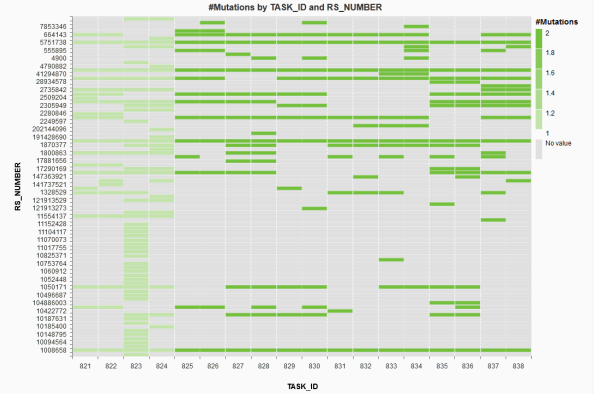
Hasso Plattner Institut

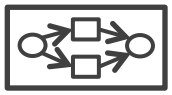


How Real-time Analysis turns Big Medical Data into Precision Medicine?

Dr. Schapranow, GLOBAL HEALTH, Aug 27, 2014

Drug Response Analysis
Interactive Data Exploration





Interactive analysis of correlations between drugs and genetic variants

- Drug response depends on individual genetic variants of tumors
- Challenge: Identification of relevant genetic variants and their impact on drug response is a ongoing research activity, e.g. Xenograft models
- Exploration of experiment results is time-consuming and Excel-driven
- In-memory technology enables interactive exploration of experiment data to leverage new scientific insights

How Real-time Analysis turns Big Medical Data into Precision Medicine?

Dr. Schapranow, GLOBAL HEALTH, Aug 27, 2014

Interactive Clinical Trial Recruitment

Hasso Plattner Institut

cytolon

- Switch from trial-centric to patient-centric clinical trials
- Real-time matching and clustering of patients and clinical trial inclusion/exclusion criteria
- No manual pre-screening of patients for months: In-memory technology enables interactive pre-screening process
- Reassessment of already screened or already participating patient reduces recruitment costs

Assessment of patients preconditions for clinical trials

How Real-time Analysis turns Big Medical Data into Precision Medicine?

Dr. Schapranow, GLOBAL HEALTH, Aug 27, 2014

Join us for upcoming projects!

Hasso Plattner Institut

- Analysis of waveform data from intensive care stations (MIT)
- Design and discovery of clinical trials (Cytolon)
- Discovery of new drugs (Bayer)
- Detect cardiovascular diseases and evaluate treatment options (BMBF)
- Use health insurance data to improve health care (BMWi)
- Processing of big medical data (Bachelor's project)

How Real-time Analysis turns Big Medical Data into Precision Medicine?

Dr. Schapranow, GLOBAL HEALTH, Aug 27, 2014

What to take home?
Test-drive it yourself: http://we.AnalyzeGenomes.com

Hasso Plattner Institut

- For patients

  - Identify relevant clinical trials and medical experts
  - Start most appropriate therapy as early as possible

- For clinicians

  - Preventive diagnostics to identify risk patients early
  - Indicate pharmacokinetic correlations
  - Scan for similar patient cases, e.g. to evaluate therapy

- For researchers

  - Enable real-time analysis of medical data and its assessment, e.g. assess pathways to identify impact of detected variants
  - Combined free-text search in publications, diagnosis, and EMR data, i.e. structured and unstructured data

Keep in contact with us after returning from Rome

Dr. Matthieu-P. Schapranow
schapranow@hpi.de
http://we.analyzegenomes.com/

Hasso Plattner Institute
Enterprise Platform & Integration Concepts (EPIC)
Program Manager E-Health
Dr. Matthieu-P. Schapranow
August-Bebel-Str. 88
14482 Potsdam, Germany