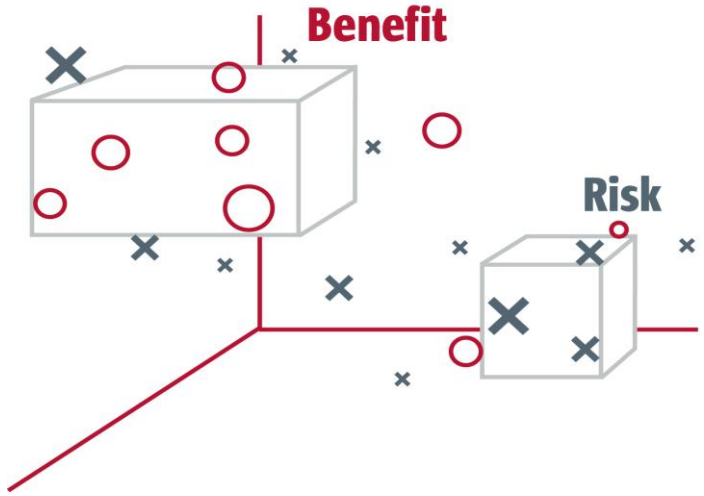
HyperCube
Complex Event Intelligence

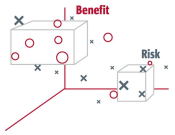
HyperCube®



A breakthrough data mining technology with game changing potential in healthcare and biomedical R&D

April 2013

HyperCube Research



Key issues in biomedical research

High-performance analytics technology is needed to convert “big data” into valuable information

Issues with biological phenomena

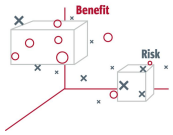
- Heterogeneity of patient populations
- Multi-factorial causality (genetic, epigenetic, environmental…) & Redundancy of pathways
- High number of variables, especially with the emergence of genomics and proteomics
- Data quality and homogeneity issues (missing data, noise, error margins, heterogeneous data formats)

Experts opinion:

- Progress in Analytics is lagging behind the exponential growth in volume of data collected
- Lack of analytic capabilities is becoming the major hurdle for converting this huge mass of data into useful information

→ Huge need for **high-performance** data analytics technologies which can overcome current difficulties, potentiate available data and **accelerate speed of discovery** in the biomedical area

→ **Sub-group discovery** and identification of **interactions between variables** are major stakes to address the wide diversity of patients and situations



HyperCube

Fruit of 15 years of R&D, HyperCube® provides a unique data mining approach

- HyperCube is a unique algorithm developed by the start-up company Effiscience, acquired by BearingPoint in 2011
- The software offers major benefits that differentiate it from competition:

1 … highlights **locally homogenous groups**, each corresponding to **real cases**

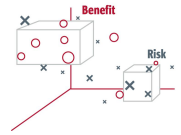
2 … defines for each of these groups the **local and original causes of outcome**

3 … works in an **explorative manner** without any preconceived hypothesis

4 …**robust in cases of missing / uncertain data, whether discrete or continuous**, without heavy pre-processing

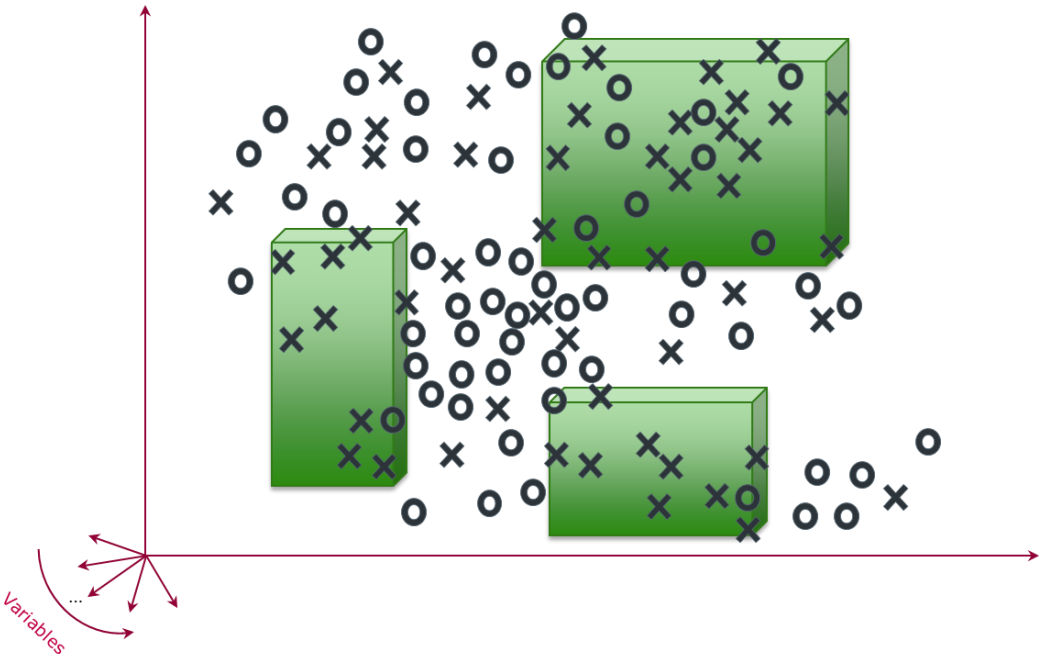
HyperCube is sold:

- Either through data analytics services
- Or as a license, as Software as a Service (access to HyperCube computing center) or installed at your premises



HyperCube (Link to HyperCube Algorithm demonstration)

Hypercube technology is a profoundly original Sub-group Discovery data mining technology



Multi-dimensional space exploration. Key steps:

- **Explore space to identify local areas of over-density**
- **For each identified area (hypercube):**
  1. Select influent variables
  2. Find optimal limits
  3. Test statistical significance
  4. Outputs scientific rule

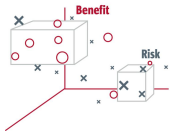
Hypercube key principle:

- Explore the multi-dimensional data space to identify local areas (hypercubes) of over-density of the phenomenon studied

Key features:

- **No pre-established model or assumptions** on the distribution of variables or relationships between input and output
- **Systematic** and **exhaustive** exploration of the space, which provides **several strong local correlations** rather one weak global relationship
- Uncovers the **interactions between variables** which cause the phenomenon
- Uses **non-Euclidean** mathematics i.e., can work with both continuous and discrete variables and is **tolerant to missing data**

HyperCube output: High-interest sub-groups with respect to the outcome

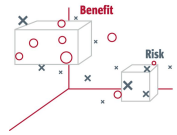


HyperCube

Hypercube characteristics make it a “perfect match” for data-enabled biomedical research

Characteristics of Hypercube technology

- Identifies relevant, and hard-to-find sub-populations
- Finds interactions between variables that are linked to the outcome
- No assumptions of any kind - neither on data distribution, nor on relationships between input and output
- Can handle almost limitless number of variables, including genomic data
- Can handle both continuous (quantitative) and discrete (qualitative) variables
- Can identify interactions between variables of different kinds (clinical, demographic, environmental, genetic)
- Finds patterns even on incomplete data – handles missing data very well



Applications to Life Sciences R&D

Pasteur publication: a rigorous demonstration of the value of HyperCube technology

Objectives: Using Hypercube, detect unknown risk factors of malaria events. Validate findings. Compare Hypercube explanatory and predictive performance with statistical method and regression tree technology

Data set: cohort of 46,837 outcome events from 1,653 individuals with 34 explanatory variables

Results:

- A 4-variable rule generated by Hypercube enabled to identify a high-risk population with relative risk of malaria episode 3.71 higher than general population.
- All rules generated by Hypercube were successfully replicated in a second, independent cohort
- Hypercube enabled to identify a previously unknown risk factor, i.e., number of previous P. malariae episodes
- No other tool used enabled to generate results with as good level of relative risk as the rule generated by Hypercube

OPEN ACCESS Freely available online PLoS one

An Exhaustive, Non-Euclidean, Non-Parametric Data Mining Tool for Unraveling the Complexity of Biological Systems – Novel Insights into Malaria

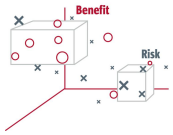
Cheikh Loucoubar[1,2,3], Richard Paul[1], Avner Bar-Hen[2,4], Augustin Huret[5], Adama Tall[3], Cheikh Sokhna[6], Jean-François Trape[6], Alioune Badara Ly[3], Joseph Faye[3], Abdoulaye Badiane[3], Gaoussou Diakhaby[3], Fatoumata Diène Sarr[3], Aliou Diop[7], Anavaj Sakuntabhai[1,8*], Jean-François Bureau[1]

Authors conclusion:

"We describe here a new data mining algorithm that can identify the combinations of variables that **give the optimal prediction** of the outcome of interest. We demonstrate that the model identified by HyperCube has **better predictive value than any other model tested.**"

"Search of local over density in m-dimensional space, explained by easily interpretable rules, is thus seemingly ideal for generating hypotheses for large datasets to unravel the complexity inherent in biological systems."

Loucoubar C, Paul R, Bar-Hen A, Huret A, Tall A, et al. (2011) An Exhaustive, Non-Euclidean, Non-Parametric Data Mining Tool for Unraveling the Complexity of Biological Systems – Novel Insights into Malaria. PLoS ONE 6(9): e24085. doi:10.1371/journal.pone.00240



Applications to Life Sciences R&D

Major applications to biomedical research

Epidemiology : Characterization of high-risk populations and identification of risk factors acting in combination

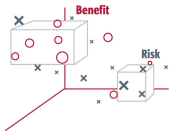
Disease Detection / Diagnosis from patient biological samples

Disease classification, identification of sub-types of diseases with specific phenotypes

Prognosis – both to explain what are the combinations of characteristics associated to good/poor prognosis patients, and to predict outcome in view of patient data

Characterization of good/bad responders profiles in post-hoc clinical trial data analysis, with or without **Pharmacogenomics**

Characterization of patients at risk of specific toxicity, with or without Pharmacogenetics



Applications to Life Sciences R&D

References in Life Sciences R&D

A six year experience in collaborating with Pharma R&D and Medical Affairs

- Clinical trial data analysis
- Epidemiology data analysis
- Safety data analysis
- Identification of patient sub-populations: good responders, patients with specific AEs…
- Identification of risk factors
- Input to clinical trial design

Clients in Life Sciences

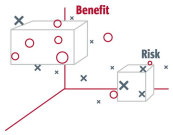
sanofi aventis
JANSSEN-CILAG
ucb Pharma
IPSEN
Innovation for patient care
Pierre Fabre
SERVIER
LFB
L'ENGAGEMENT ÉTHIQUE

Well-established collaborations with leading academic epidemiologists

- Identification of disease risk factors (malaria, asthma)
- Analysis of genetic data (GWAS SNPs data) (tuberculosis, hepatitis C)
- Analysis of gene expression data (cancer)
- Analysis of MRI scan data (AD diagnosis)

Collaborations in academic world

The Rockefeller University
CHARITÉ
Imperial College London
Inserm
Institut national de la santé et de la recherche médicale
Institut de cancérologie GUSTAVE ROUSSY
VILLEJUIF - www.igr.fr
Institut Pasteur
EPFL
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE



In summary: take-home messages

HyperCube: a potential game changer in the days of effectiveness programs, low-cost genomics, and progressive emergence of personalized medicine

Hypercube is a genuinely different and innovative data mining technology

- Key principle = identification of “local” causes of phenomena

Hypercube can extract value from your data

- Finds correlations with unequaled performance
- Can work with all kinds of data
- Is not limited by missing or heterogeneous data
- Can analyze massive genetic data sets

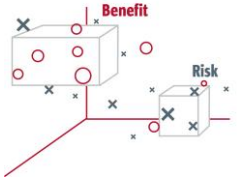
Hypercube is a full match with the needs and issues of biomedical research

- Biological and patient heterogeneity
- Multi-factorial causality
- 00,000s of variables
- Uneven data quality

Hypercube has potential to accelerate analysis of your data and foster new scientific discovery in your research projects

Thank you!

HyperCube
Complex Event Intelligence

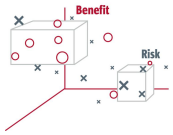


Yann Gaston-Mathé
Director, Life Sciences

Hypercube Research
51, esplanade du G[al] de Gaulle
92907 Paris la Défense Cedex
France

T +33 1 58 86 32 42
M +33 6 23 74 14 14

yann.gastonmathe@hypercube-research.com



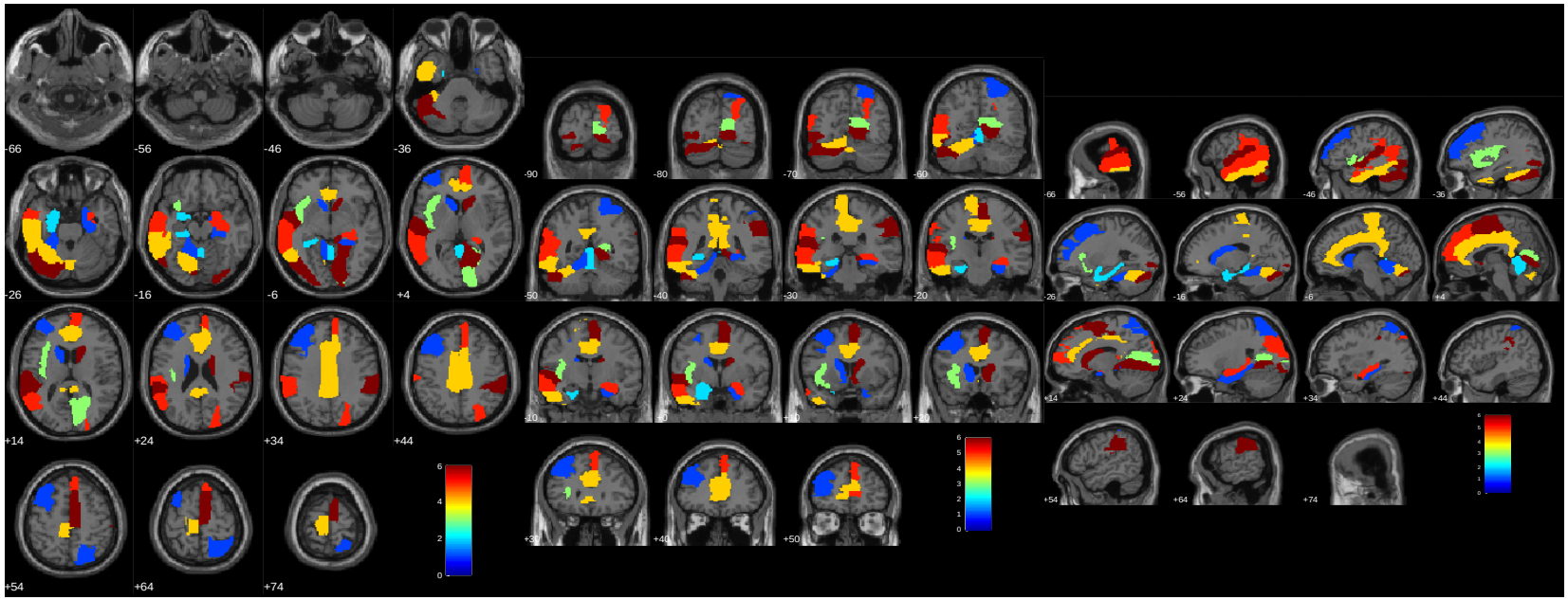
Applications to Life Sciences

HyperCube for Alzheimer patients precocious detection (collaboration with EPFL / CHUV)

Case and preliminary data (confirmation and replication is ongoing)

- ADNI cohorts with Alzheimer and control patients
- Cerebral MRI scan normalized data

HyperCube analysis objective: find rules enabling discriminate patients from controls.



Example of results : 6 rules with 100% purity of Alzheimer patients covering 98% of Alzheimer patients

HyperCube

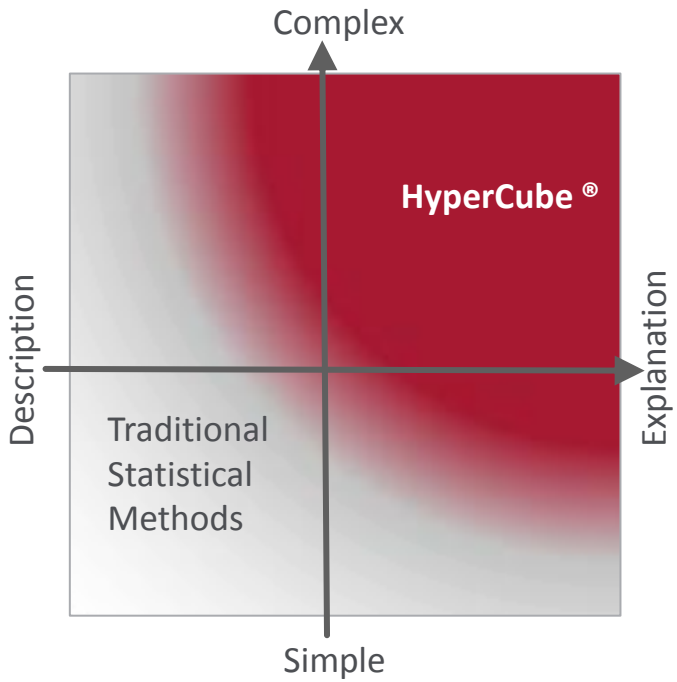
HyperCube can succeed where traditional statistical methods fail to decipher the underlying complexity of phenomena

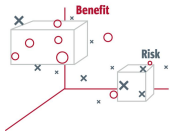
Statistical approach

- **Provides "typical" results that may not describe any real scenario**
- **Do not perform well with partially complete or limited data sets (very common)**
- **Cannot manage non continuous variable**
- **Do not precisely explain "why" something occurs nor quantify impact**
- **Rely on assumptions and hypotheses or pre-ranking of variables**

HyperCube® Technology

- Highlights locally homogenous groups, each corresponding to real cases
- Does not rely on averages/ best-fit line but can identify trends before they become statistically relevant
- Handles natively any type of variable whatever their nature
- Provides root cause analysis which identifies the drivers behind the data
- Works exhaustively with no analytical bias or preliminary assumption





HyperCube

The award-winning HyperCube solution is recognized by the business and scientific community

IDC
Analyze the Future

"HyperCube has the ability to help organisations identify causal links between different factors as represented in a dataset, and present these links as business rules in a simple format to an end user. The fact that it outpaces traditional statistics by firstly requiring no hypothesis and secondly working on a full dataset rather than a sample, is very powerful. It constitutes an excellent reusable asset for BearingPoint's consulting practice."

Mr Alys Woodward, European Business Analytics IDC

MITSloan
CIO SYMPOSIUM

HyperCube **entered the final round of the MIT Sloan School "Innovation Showcase"** in 2010 This award recognises 10 young firms proposing innovative technologies. *"The new state-of-the-art Complex Event Intelligence: proactive/predictive optimisation of problems on a granular level through non-assumptions based analysis of massive scale non-homogenous data sets."*

Mr Jim Champy, MIT board

Institut Pasteur

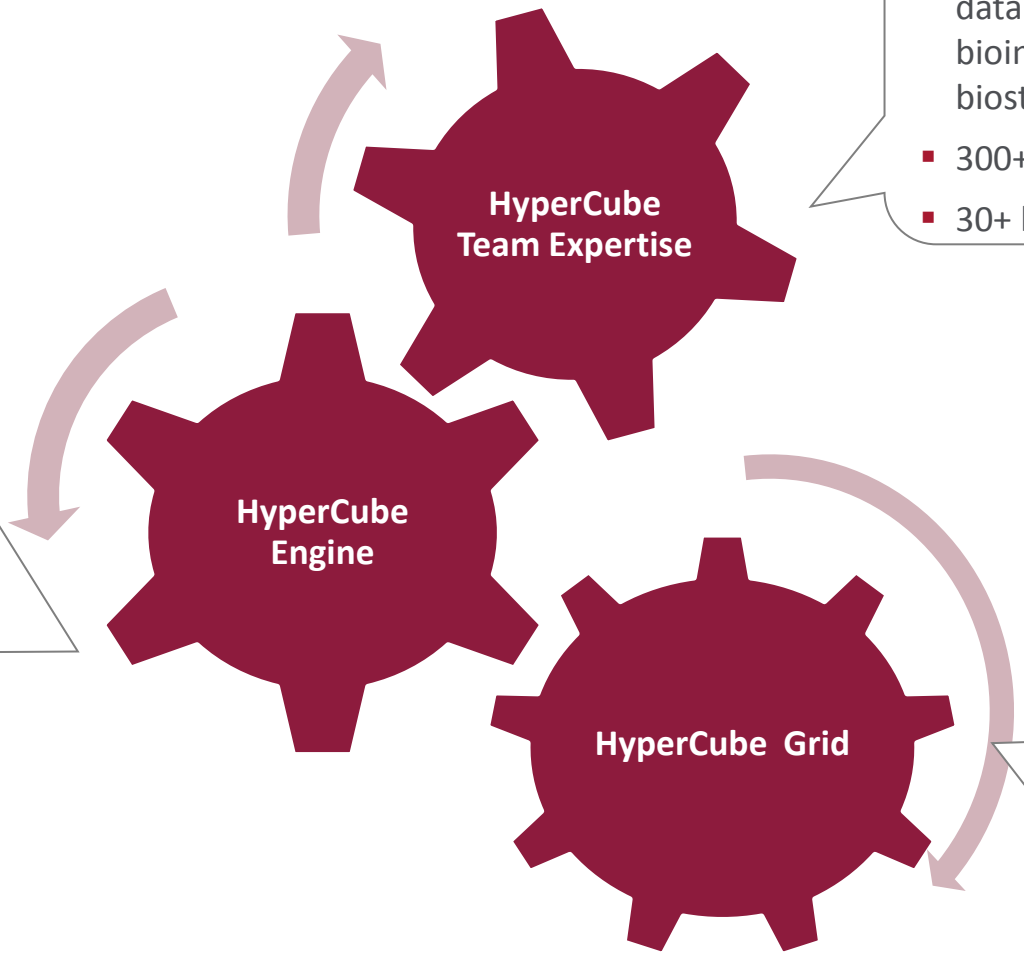
Pasteur Institute recognized HyperCube as the **best performing data analysis tool** as of today: *"We describe here a new data mining algorithm that can identify the combinations of variables that **give the optimal prediction** of the outcome of interest. We demonstrate that the model identified by HyperCube has **better predictive value than any other model tested**. HyperCube was able to identify the **best cut-off value and range** for continuous variables. It classified the population into high and low risk groups and made the results easier to interpret in terms of biology than the probability estimates generated by most statistical methods."*

Source: http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0024085

HyperCube

A solution built on three core assets

- Non-statistical algorithm
- Based on 15 years of research
- Exhaustive and not assumption based
- Simple formal business rules explaining root cause or the condition of appearance of a phenomenon



- Dedicated analysts with data mining, bioinformatics and biostatistics profiles
- 300+ reference projects
- 30+ key business issues

- 40 Teraflops computing center
- Scalable
- Software as a Service