Clinical Bioinformatics: a research agenda to support health care transformation

Riccardo Bellazzi
Biomedical Informatics Labs
University of Pavia
IRCCS Fondazione S. Maugeri
Pavia

Laboratory
for
Biomedical
Informatics

Translational Bioinformatics

Summit on Translational Bioinformatics
March 10-12, 2008
San Francisco, California

"It is the responsibility of those of us involved in today's biomedical research enterprise to **translate the remarkable scientific innovations** we are witnessing **into health gains** for the nation... At no other time has the need for a robust, **bidirectional information flow between basic and translational scientists** been so necessary."

--Dr. Elias Zerhouni, Director of the National Institutes of Health, 2005

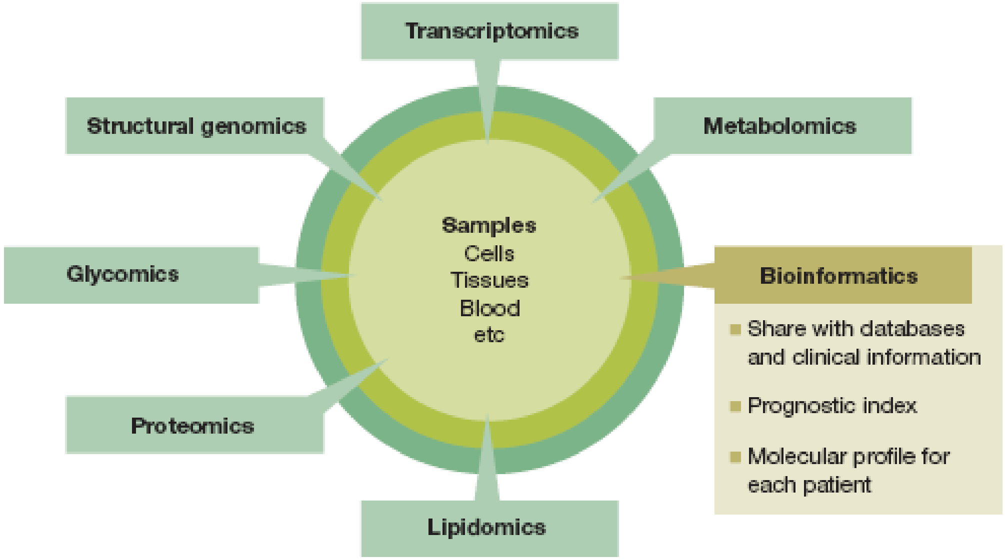
Towards preventive medicine

High-throughput methods from molecular biology are about to change daily clinical practice

Hrvojka Bosnjak, Kresimir Pavelic & Sandra Kraljevic Pavelic

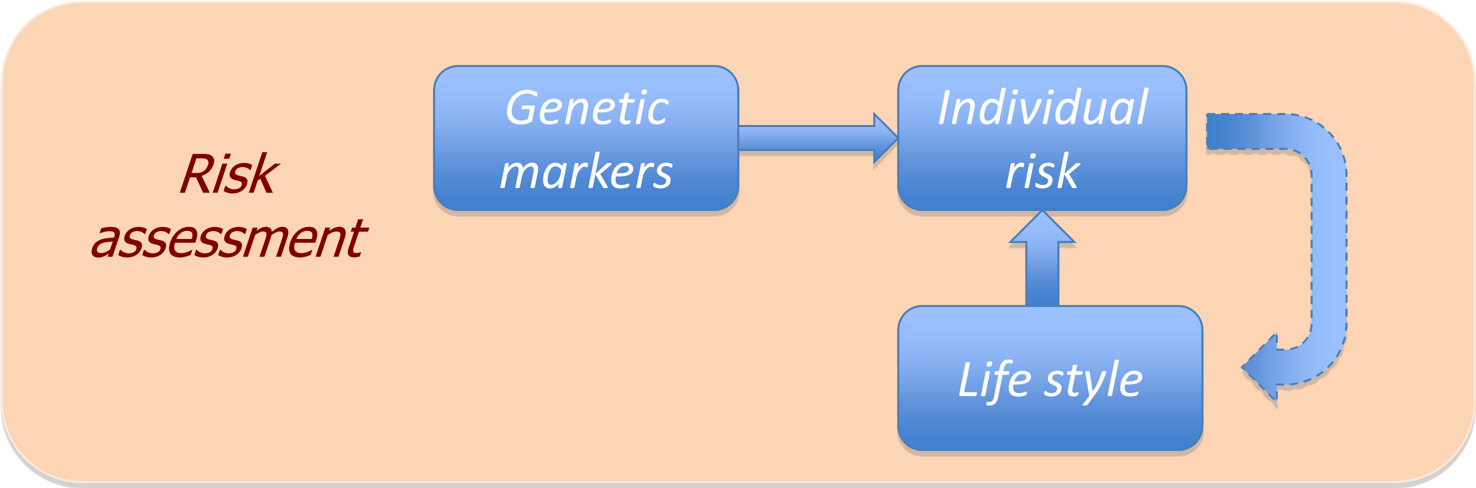
Table 1 | Randomly chosen examples of clinical studies using the -omics methods

Focus of article	Reference
Gene expression data in building prognostics models for early-stage lung cancer.	Director's Challenge Consortium for the Molecular Classification of Lung Adenocarcinoma *et al* (2008) Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat Med* 14: 822–827
Gene expression in cytologically normal large-airway epithelial cells can act as a lung cancer biomarker, potentially owing to a cancer-specific airway-wide response to cigarette smoke.	Spira A *et al* (2007) Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer. *Nat Med* 13: 361–366
Predicting medulloblastoma clinical outcome by gene expression profiling.	Korenberg MJ (2004) On predicting medulloblastoma metastasis by gene expression profiling. *J Proteome Res* 3: 91–96
Discovery of bone metastasis genes. Metastasis requires the concerted action of multiple genes that belong to different functional classes.	Kang Y *et al* (2003) A multigenic program mediating breast cancer metastasis to bone. *Cancer Cell* 3: 537–549
186 genes as the "invasion signature".	Liu R *et al* (2007) The prognostic role of a gene signature from tumorigenic breast cancer cells. *N Engl J Med* 356: 217–226
Prediction: which of the node-negative breast cancer patients would develop distant metastasis during the 5-year follow up period and which would remain disease-free?	Wang Y *et al* (2005) Proteomic profiling of urine for the detection of colon cancer. *Lancet* 365: 671–679
Proteomic profiling of urine for the detection of colon cancer.	Ward DG *et al* (2008) Proteomic profiling of urine for the detection of colon cancer. *Proteome Sci* 6: 19
Gene expression profiling allows reliable prediction of lymph node metastasis in cervical cancer.	Kim T-J *et al* (2008) Gene expression profiling for the prediction of lymph node metastasis in patients with cervical cancer. *Cancer Sci* 99: 31–38
Pathological changes within an organ might be reflected in the proteomic pattern in serum. Proteomics analysis of serum could recognize ovarian cancer cases.	Petricoin E *et al* (2002) Use of proteomic patterns in serum /to identify ovarian cancer. *Lancet* 359: 572–577



AIMS

- Learning on biochemical pathways
- Detection of disease in the very early stages
- Prediction of outcome of therapies

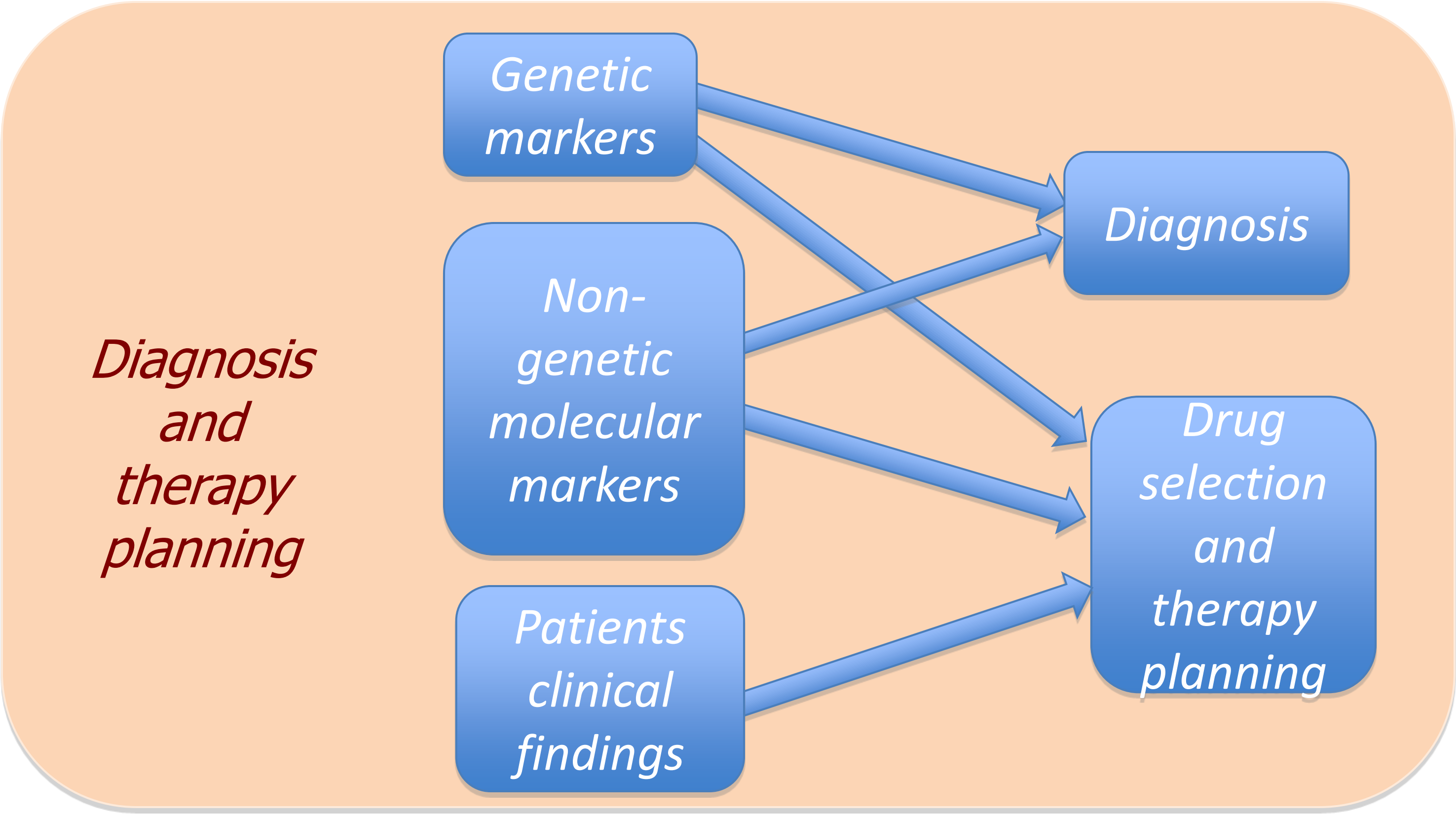


Risk assessment

Genetic markers

Individual risk

Life style



Diagnosis and therapy planning

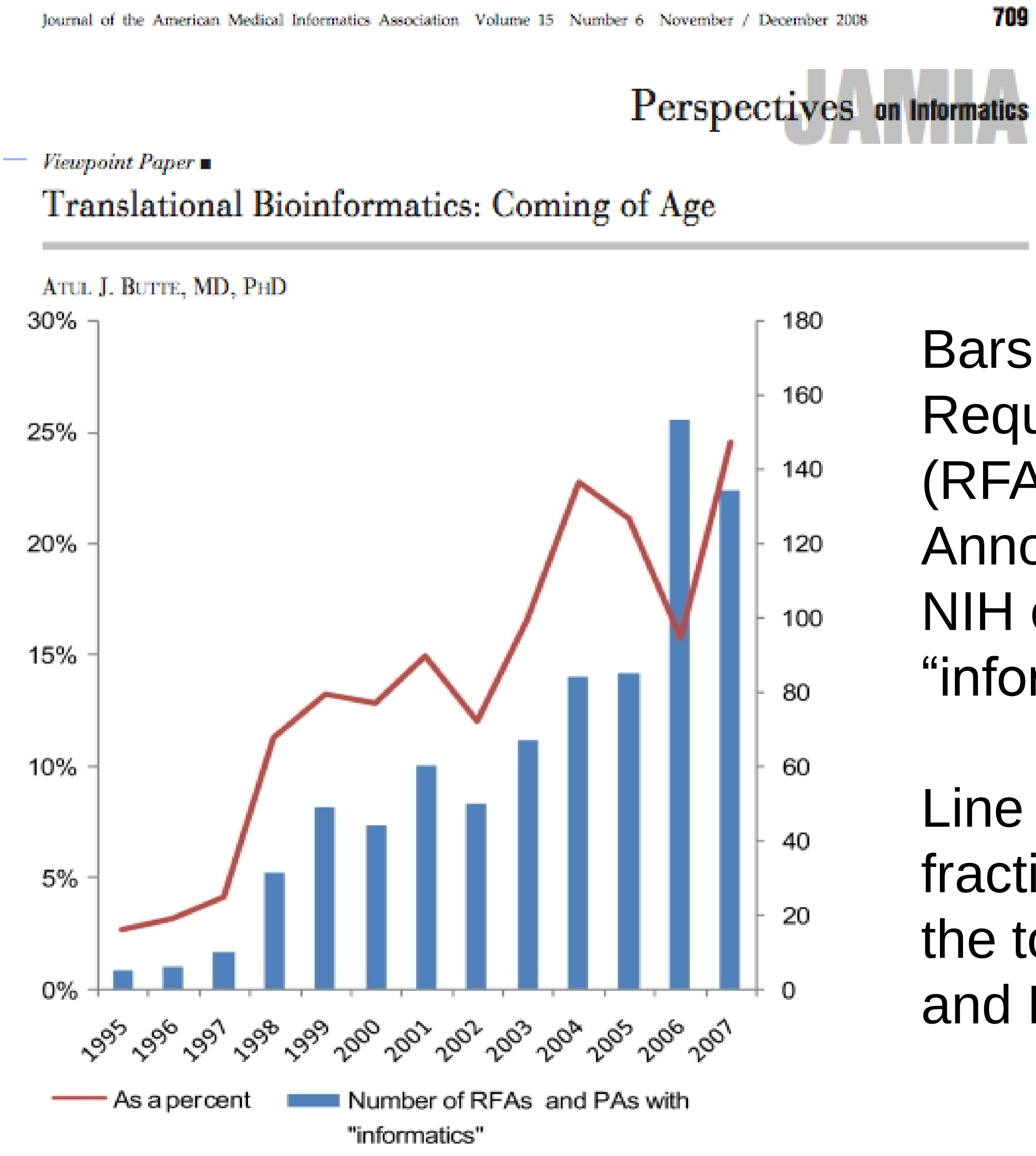
Genetic markers

Non-genetic molecular markers

Patients clinical findings

Diagnosis

Drug selection and therapy planning



Journal of the American Medical Informatics Association Volume 15 Number 6 November / December 2008 709

JAMIA Perspectives on Informatics

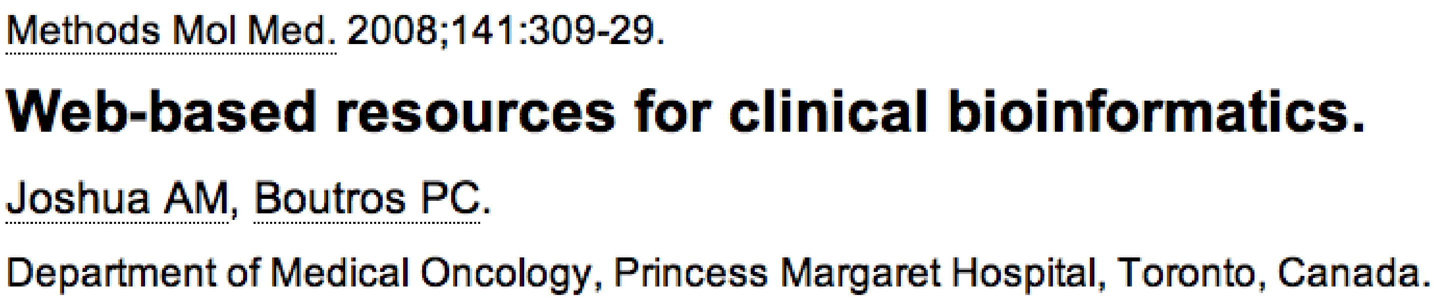
Viewpoint Paper ■

Translational Bioinformatics: Coming of Age

ATUL J. BUTTE, MD, PHD

Bars represent the Request for Applications (RFAs) and Program Announcements (PAs) by NIH containing the term “informatics”.

Line represents the fraction of this count over the total count of RFAs and PAs that year.

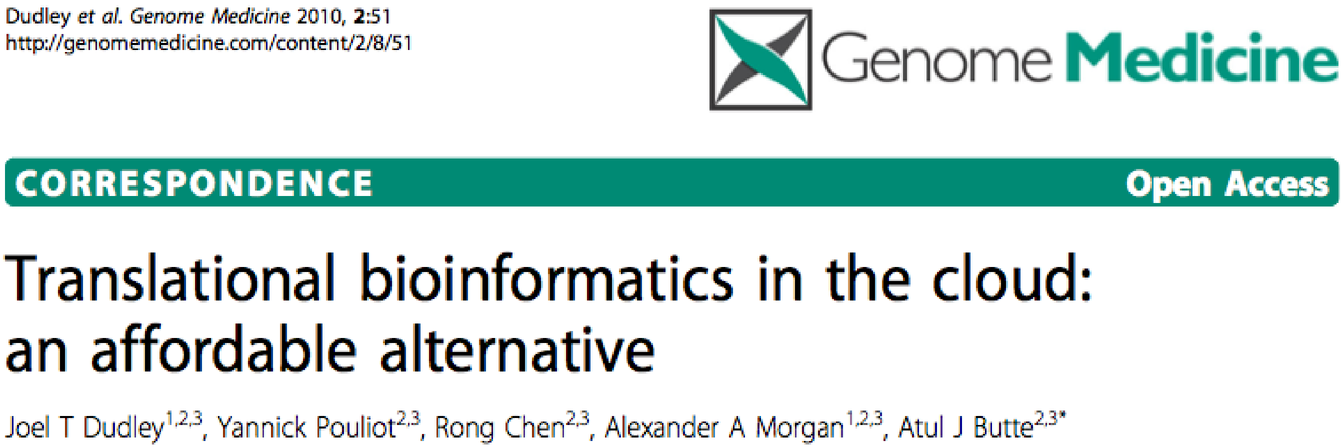


Methods Mol Med. 2008;141:309-29.

Web-based resources for clinical bioinformatics.

Joshua AM, Boutros PC.

Department of Medical Oncology, Princess Margaret Hospital, Toronto, Canada.

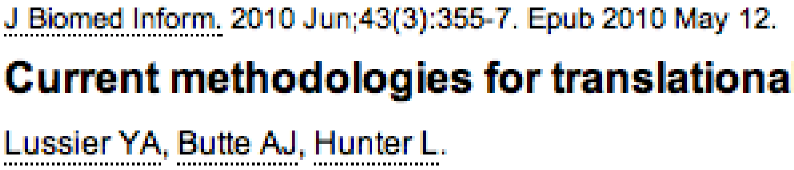


Dudley *et al. Genome Medicine* 2010, **2**:51
http://genomemedicine.com/content/2/8/51

CORRESPONDENCE Open Access

Translational bioinformatics in the cloud: an affordable alternative

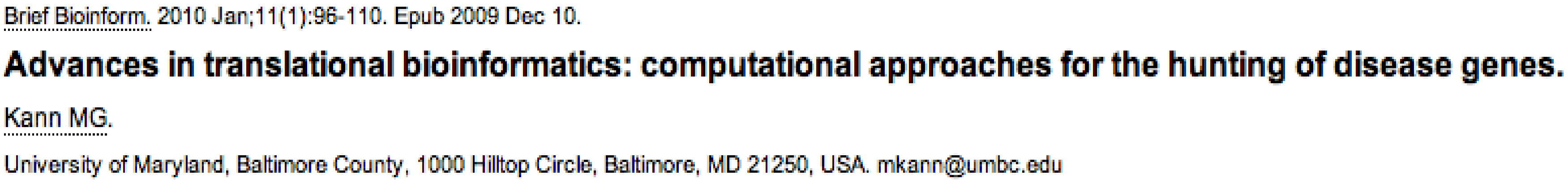
Joel T Dudley[1,2,3], Yannick Pouliot[2,3], Rong Chen[2,3], Alexander A Morgan[1,2,3], Atul J Butte[2,3*]



J Biomed Inform. 2010 Jun;43(3):355-7. Epub 2010 May 12.

Current methodologies for translational bioinformatics.

Lussier YA, Butte AJ, Hunter L.



Brief Bioinform. 2010 Jan;11(1):96-110. Epub 2009 Dec 10.

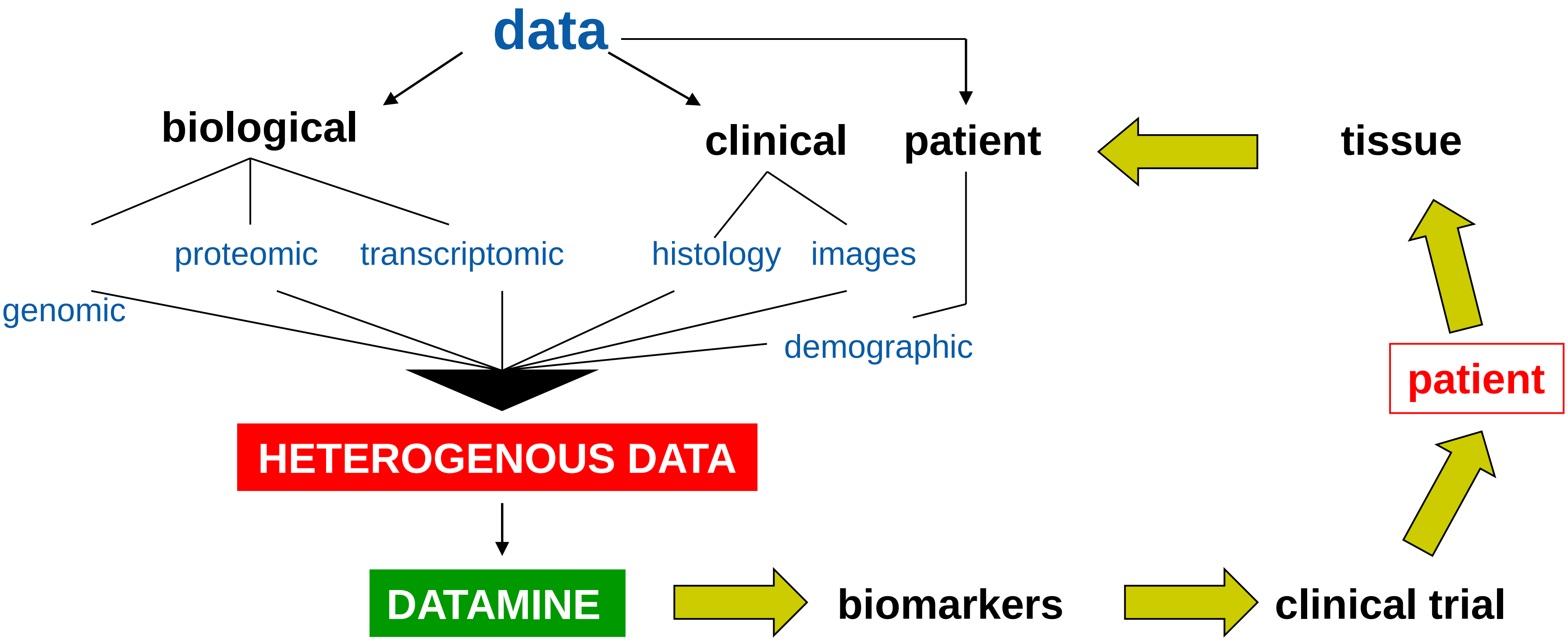
Advances in translational bioinformatics: computational approaches for the hunting of disease genes.

Kann MG.

University of Maryland, Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250, USA. mkann@umbc.edu

Clinical Bioinformatics

Translational Cancer Research & Clinical Bioinformatics



data

biological

clinical

patient

tissue

genomic

proteomic

transcriptomic

histology

images

demographic

patient

HETEROGENOUS DATA

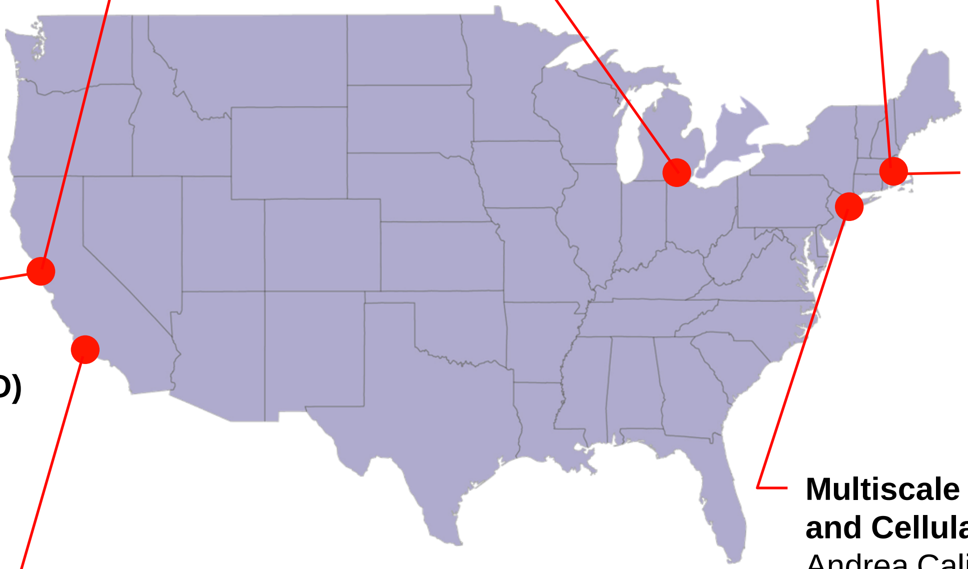
DATAMINE

biomarkers

clinical trial

Paul Lewis, Cancer Informatics Group

NIH Roadmap National Centers for Biomedical Computing (NCBC)



Physics-Based Simulation of Biological Structures (SIMBIOS)
Russ Altman, PI

National Center for Integrative Biomedical Informatics (NCIBI)
Brian D. Athey, PI

Informatics for Integrating Biology and the Bedside (i2b2)
Isaac Kohane, PI

National Alliance for Medical Imaging Computing (NA-MIC)
Ron Kikinis, PI

The National Center For Biomedical Ontology (NCBO)
Mark Musen, PI

Multiscale Analysis of Genomic and Cellular Networks (MAGNet)
Andrea Califano, PI

Center for Computational Biology (CCB)
Arthur Toga, PI

Clinical by Informatics &Translational Bioinformatics

i2b2
Informatics for Integrating Biology & the Bedside
https://www.i2b2.org/

STANFORD SCHOOL OF MEDICINE | Stanford Center for Clinical Informatics
http://clinicalinformatics.stanford.edu/

NCIBI
National Center for Integrative Biomedical Informatics
http://portal.ncibi.org/gateway/

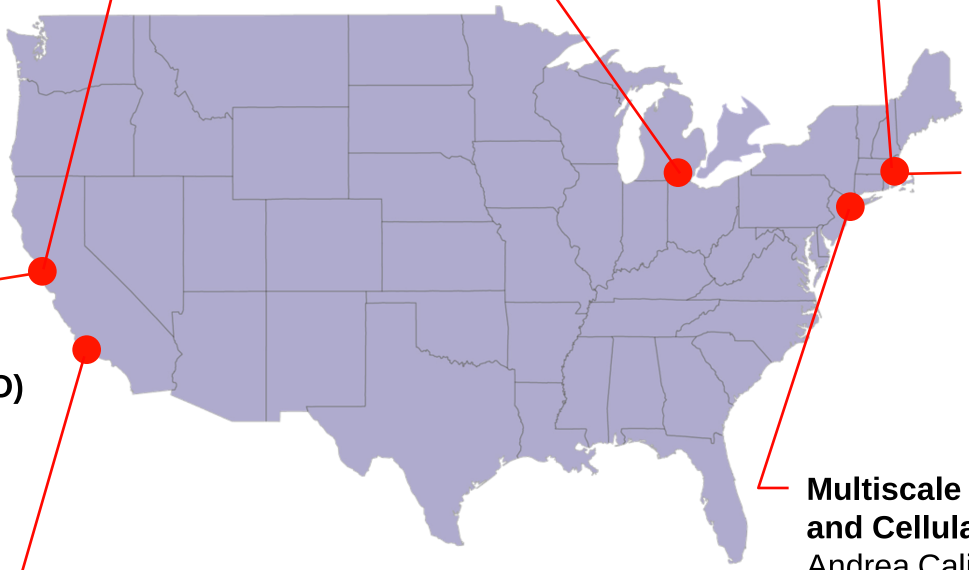
Yale Center for Medical Informatics
http://ycmi.med.yale.edu/index.html

THE NATIONAL CENTER FOR BIOMEDICAL ONTOLOGY
http://www.bioontology.org/

iDASH
http://idash.ucsd.edu/

Infrastructures

NIH Roadmap National Centers for Biomedical Computing (NCBC)



Physics-Based Simulation of Biological Structures (SIMBIOS)
Russ Altman, PI

National Center for Int
Biomedical Informatic
Brian D. Athey, PI

Informatics for Integrating Biology and the Bedside (i2b2)

Isaac Kohane, PI

National Alliance for Medical Imaging Computing (NA-MIC)
Ron Kikinis, PI

The National Center For Biomedical Ontology (NCBO)
Mark Musen, PI

Multiscale Analysis of Genomic and Cellular Networks (MAGNet)
Andrea Califano, PI

Center for Computational Biology (CCB)
Arthur Toga, PI

Ten thousand views of bioinformatics: a bibliome perspective.

Kohane I.

Yearb Med Inform. 2009:113-6.

Instrumenting the health care enterprise for discovery research in the genomic era

Shawn Murphy, Susanne Churchill, Lynn Bry, et al.

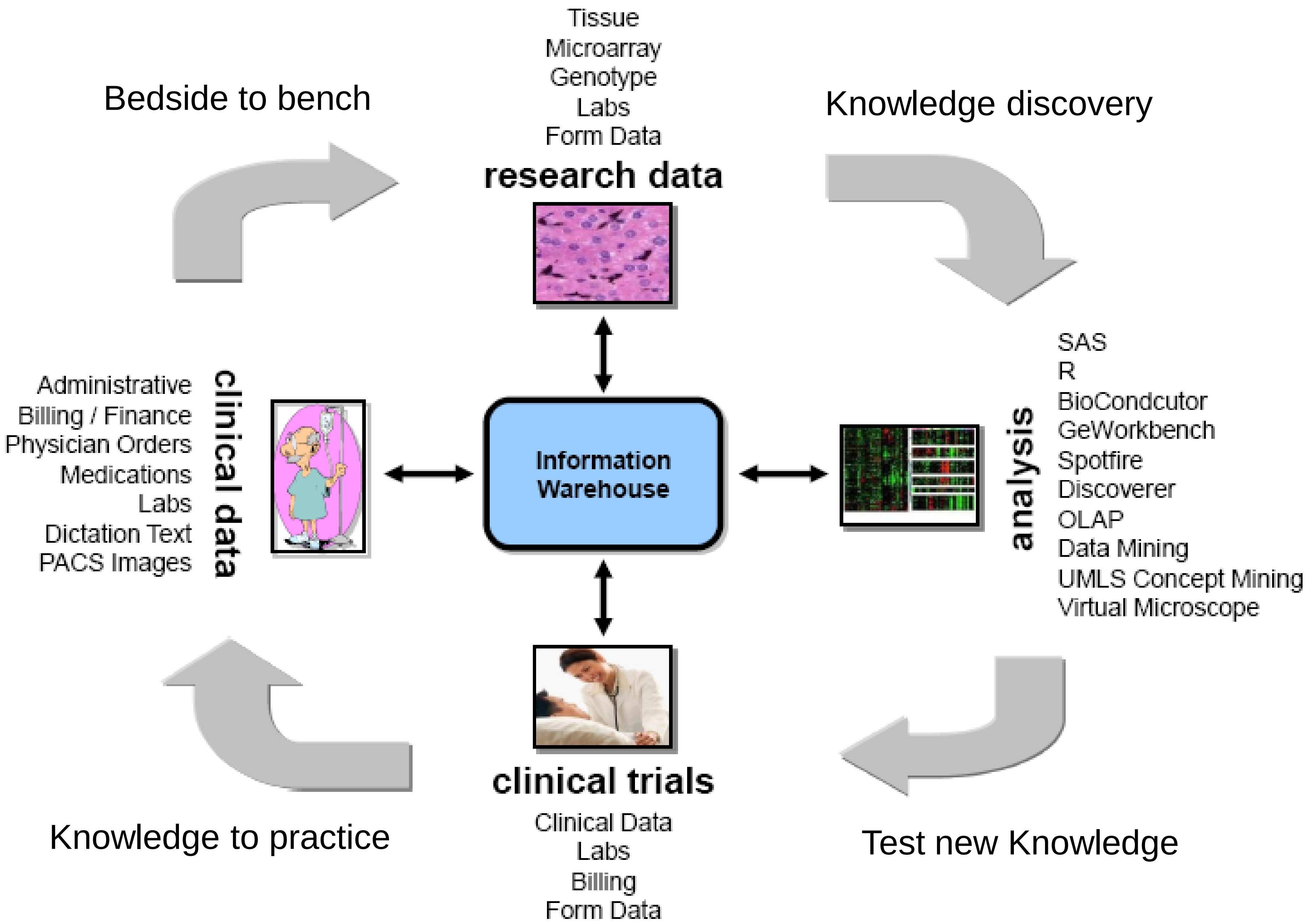
Genome Res. 2009 19: 1675-1681 originally published online July 14, 2009

Foundational biomedical informatics research in the clinical and translational science era: a call to action

Philip R O Payne, Peter J Embi and Joyce Niland

JAMIA 2010 17: 615-616

Clinical Bioinformatics



Bedside to bench

Tissue
Microarray
Genotype
Labs
Form Data

research data

Knowledge discovery

Administrative
Billing / Finance
Physician Orders
Medications
Labs
Dictation Text
PACS Images

clinical data

Information Warehouse

analysis

SAS
R
BioCondcutor
GeWorkbench
Spotfire
Discoverer
OLAP
Data Mining
UMLS Concept Mining
Virtual Microscope

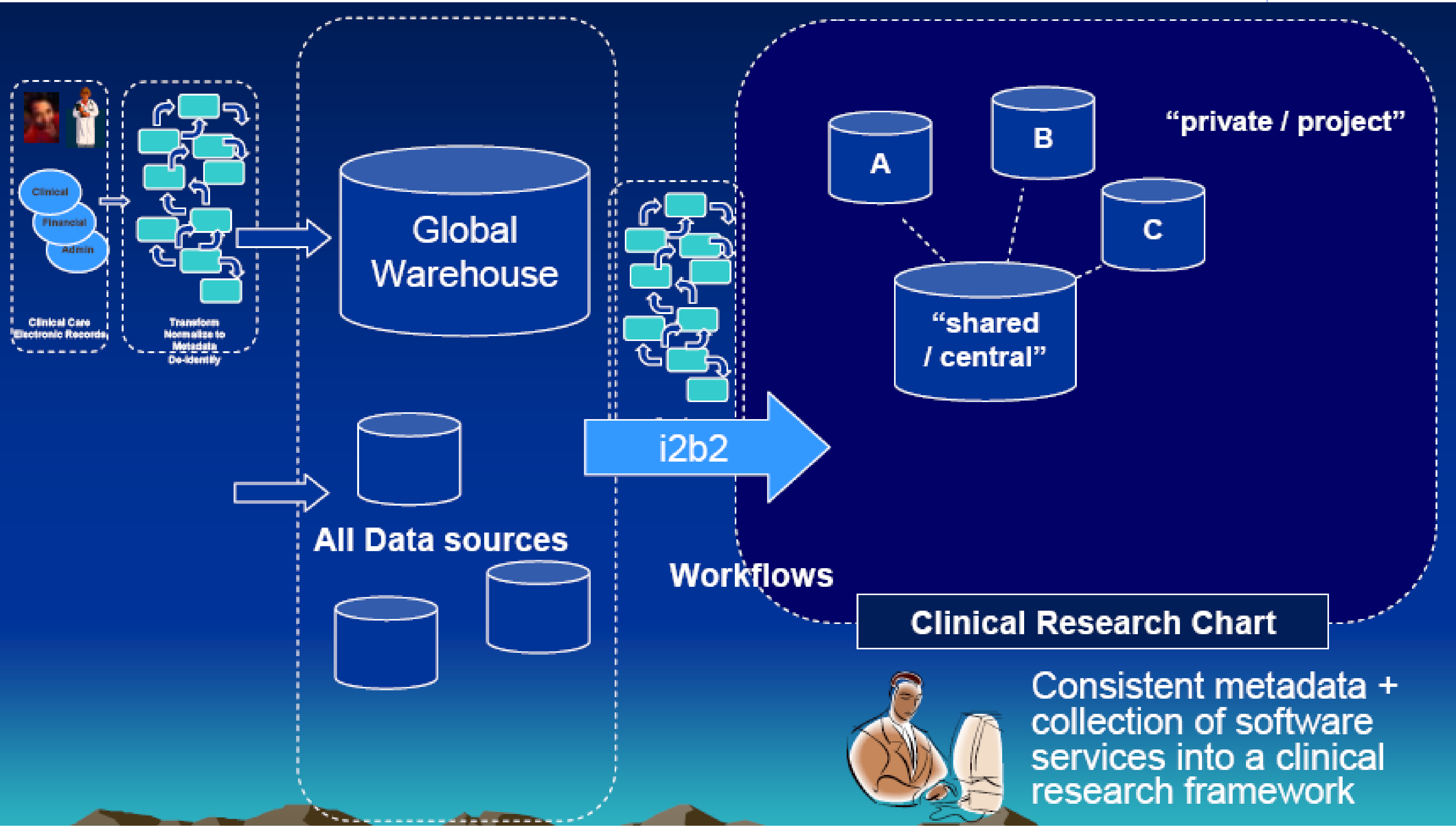
clinical trials

Clinical Data
Labs
Billing
Form Data

Knowledge to practice

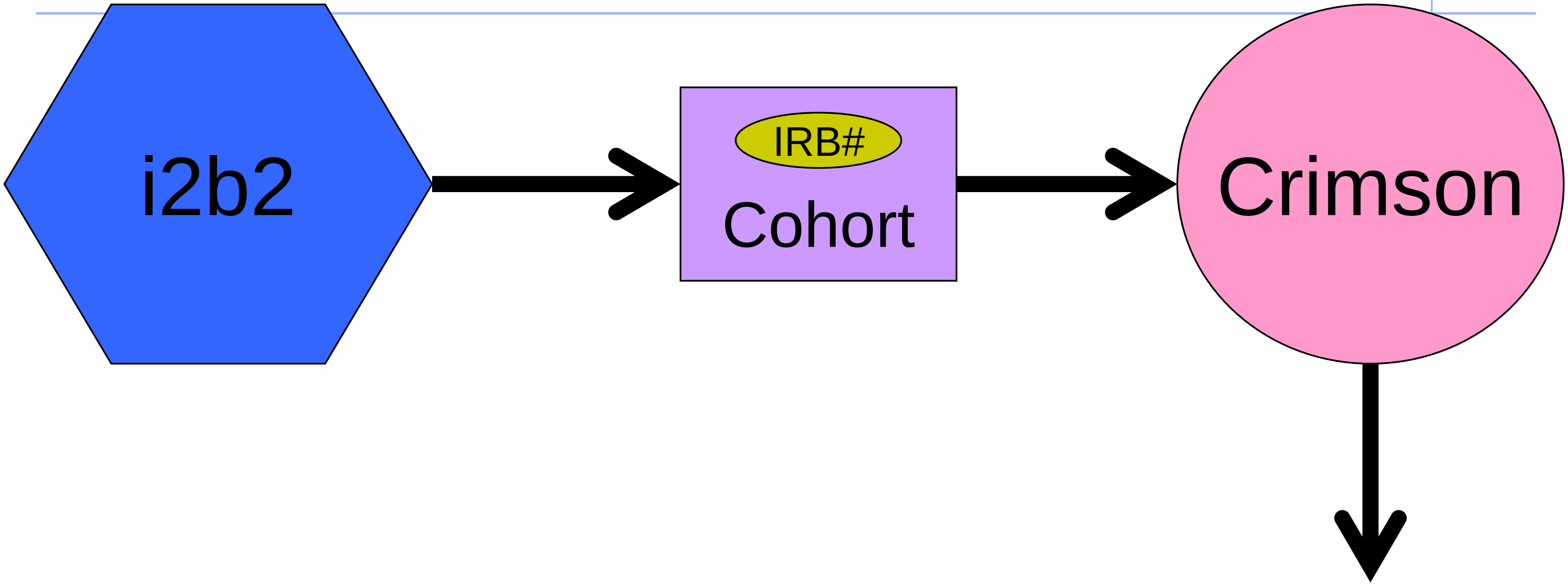
Test new Knowledge

CLINICAL RESEARCH CHART

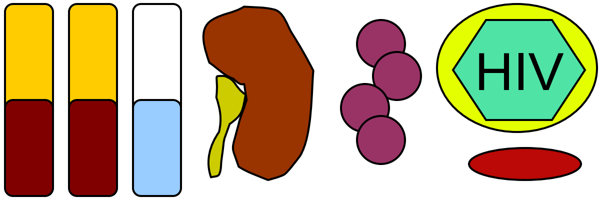


Consistent metadata + collection of software services into a clinical research framework

Crimson



Samples:
Pathology Depts/Clinical Labs
BWH labs discard >5000
Clinical samples/day >2 million/year
Partners hospitals: >20,000 samples/day
BWH AP >200,000 samples/year

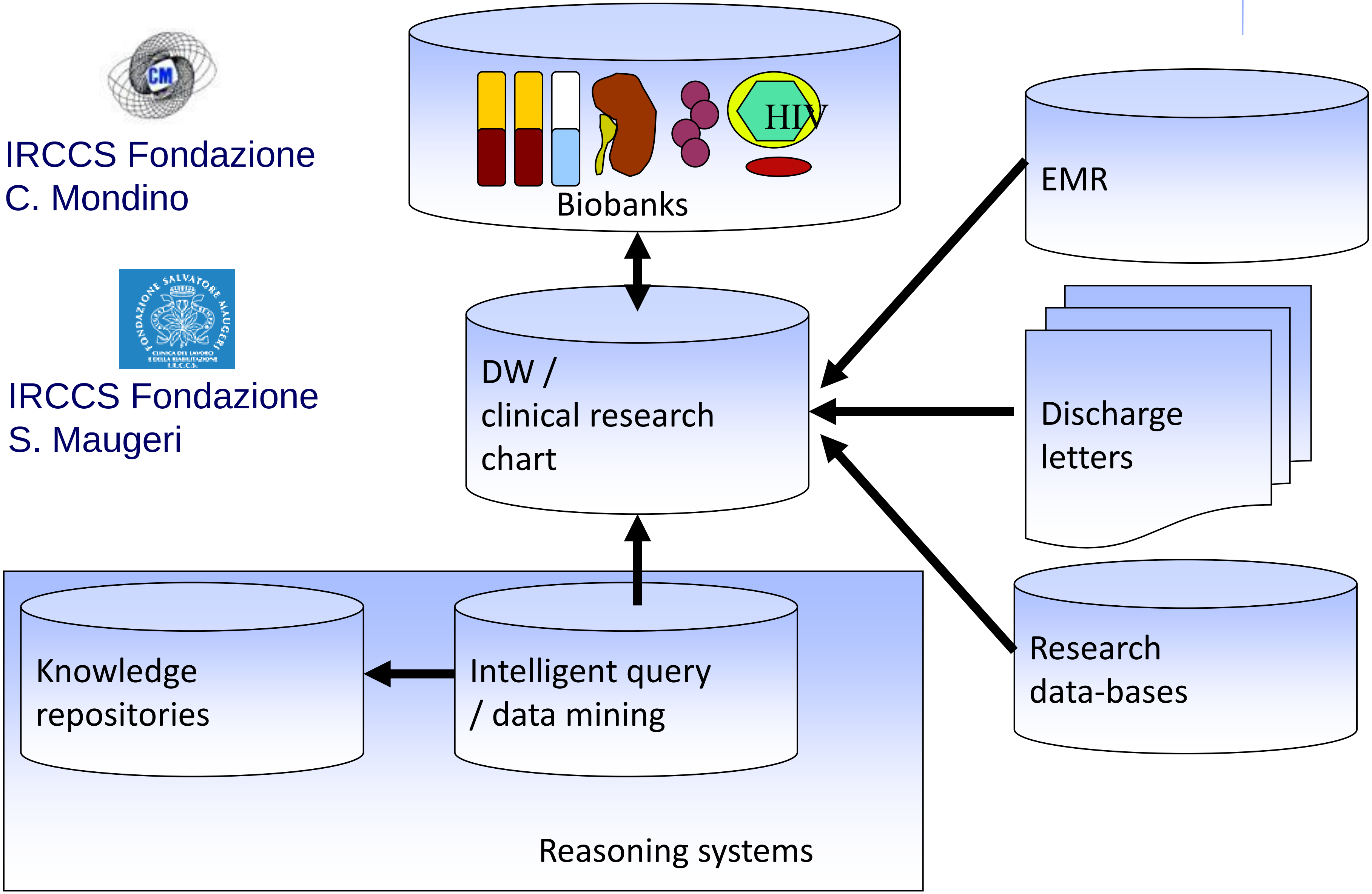


High-throughput sample collection

Clinical Bioinformatics – the i2b2 Pavia project

IRCCS Fondazione
C. Mondino

IRCCS Fondazione
S. Maugeri



Biobanks

EMR

DW /
clinical research
chart

Discharge
letters

Research
data-bases

Knowledge
repositories

Intelligent query
/ data mining

Reasoning systems

Università
degli Studi di Pavia

IUSS
Sapere Aude
Istituto Universitario
di Studi Superiori
Pavia

BIOINFORMATICS METHODOLOGY AND TECHNOLOGY TO INTEGRATE CLINICAL AND BIOLOGICAL KNOWLEDGE SUPPORTING ONCOLOGY TRANSATIONAL RESEARCH

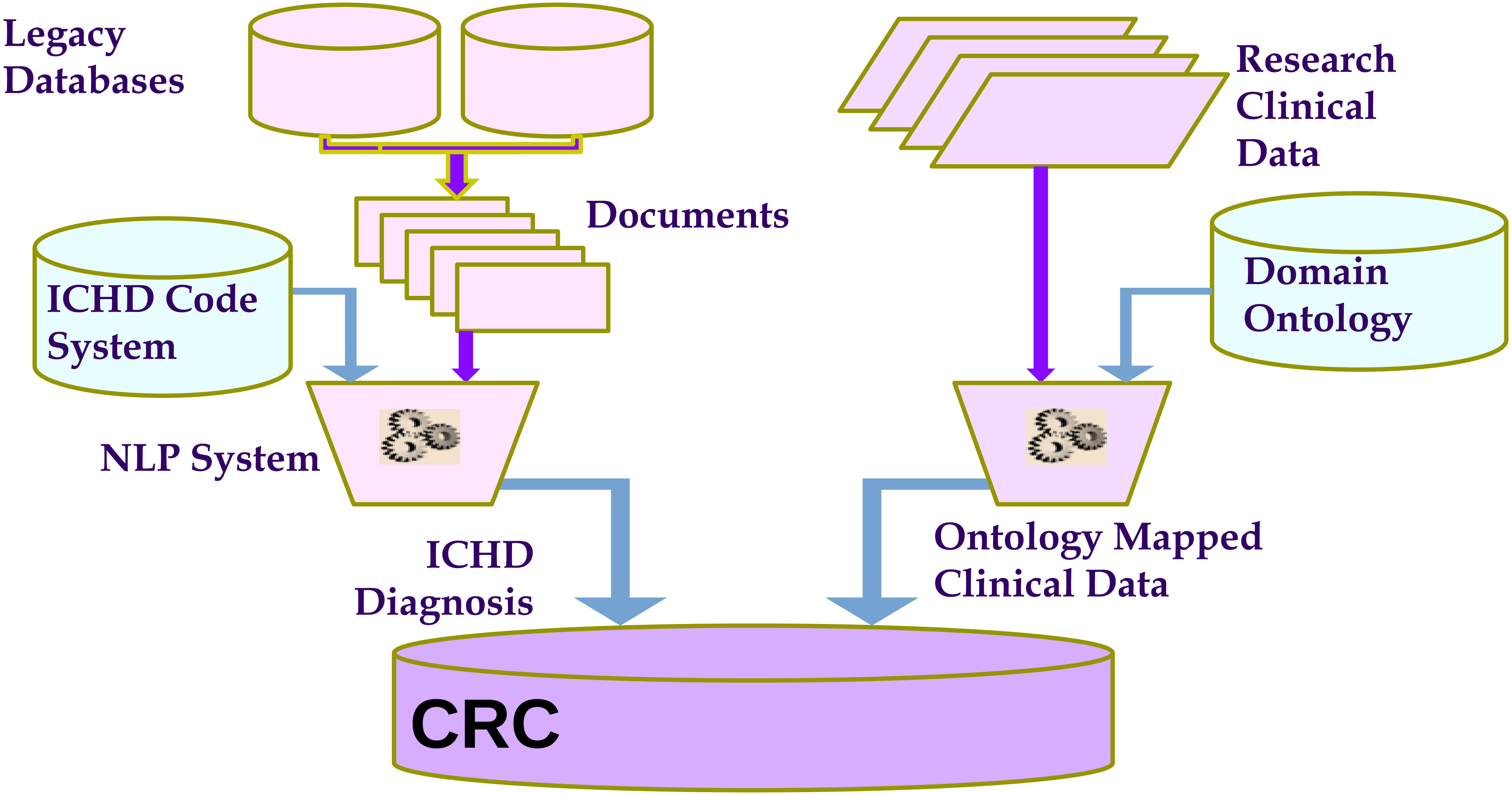
(**ONCO-I2B2**)

FONDAZIONE SALVATORE MAUGERI
CLINICA DEL LAVORO
E DELLA RIABILITAZIONE
I.R.C.C.S.

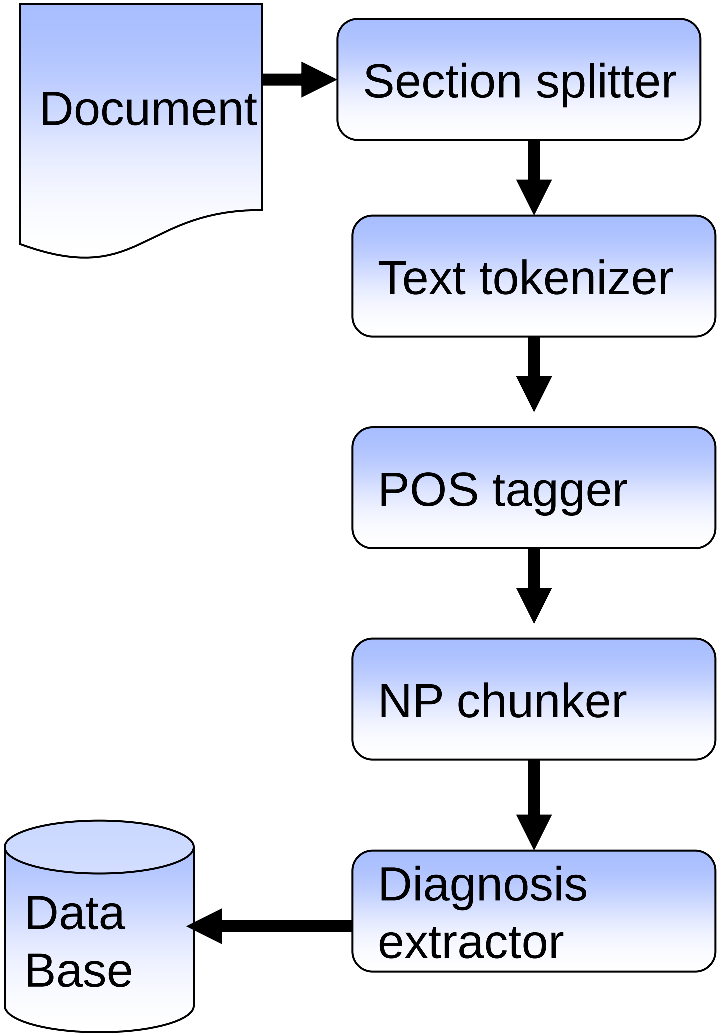
FINLOMBARDA

RegioneLombardia

I2b2-Pavia populating the datawarehouse



NLP – extracting information from clinical narratives



Document

Section splitter

Text tokenizer

POS tagger

NP chunker

Diagnosis extractor

Data Base

Section splitter

Sections

[…]

TREATMENT

trattamento raccomandato: sulla base della storia clinica e della obiettività neurologica ho concordato con il paziente di tenere un diario della cefalea (indicando durata, frequenza, intensità delle crisi, uso di analgesici). si consiglia di mantenere un regolare ritmo sonno-veglia, di riprendere un'attività fisica periodica.

ritengo che vi sia una tensione a livello dei muscoli epicranici e del collo che potrebbe essere migliorata con della fisioterapia.

DIAGNOSIS

diagnosi: 1. emicrania senza aura
2. cefalea tensiva episodica sporadica.

THERAPY

terapia consigliata:
" almotriptan cpr: una cpr all'inizio della crisi; in alternativa/dopo due ore: indometacina supp. 50 mg: una supp.

[…]

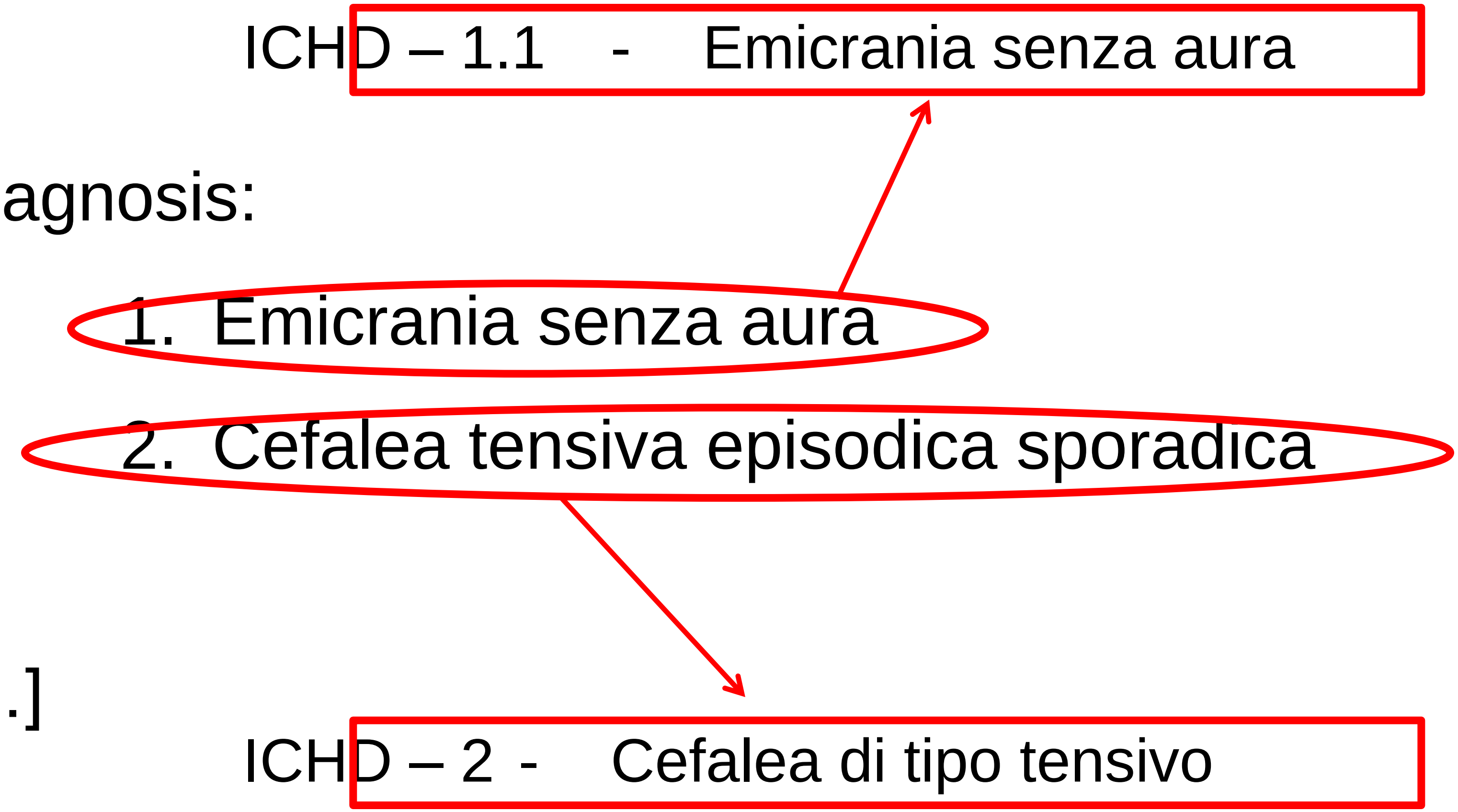
NP - chunker

[…]
Si consiglia di mantenere un regolare ritmo sonno-veglia, di riprendere un'attività fisica periodica.
Ritengo che vi sia una tensione a livello dei muscoli epicranici e del collo che potrebbe essere migliorata con della fisioterapia.
[…]

Noun Phrase

Concept Finder

[…]



ICHD – 1.1 - Emicrania senza aura

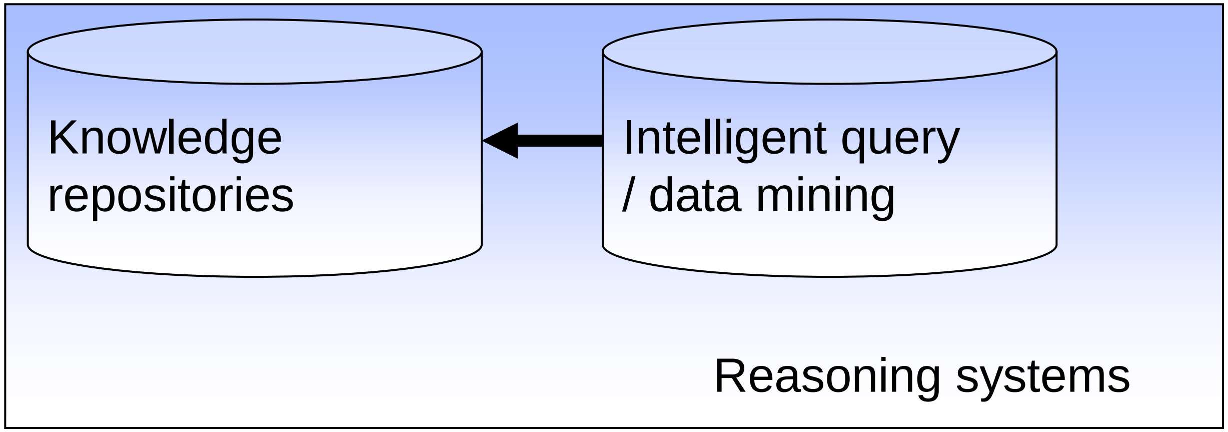
Diagnosis:

1. Emicrania senza aura
2. Cefalea tensiva episodica sporadica

[…]

ICHD – 2 - Cefalea di tipo tensivo

Towards knowledge-discovery support systems



Knowledge repositories

Intelligent query / data mining

Reasoning systems

The Phenotype Miner[1]

Genephony[3]

SNP-2-Net[3]

The ST-Model[4]

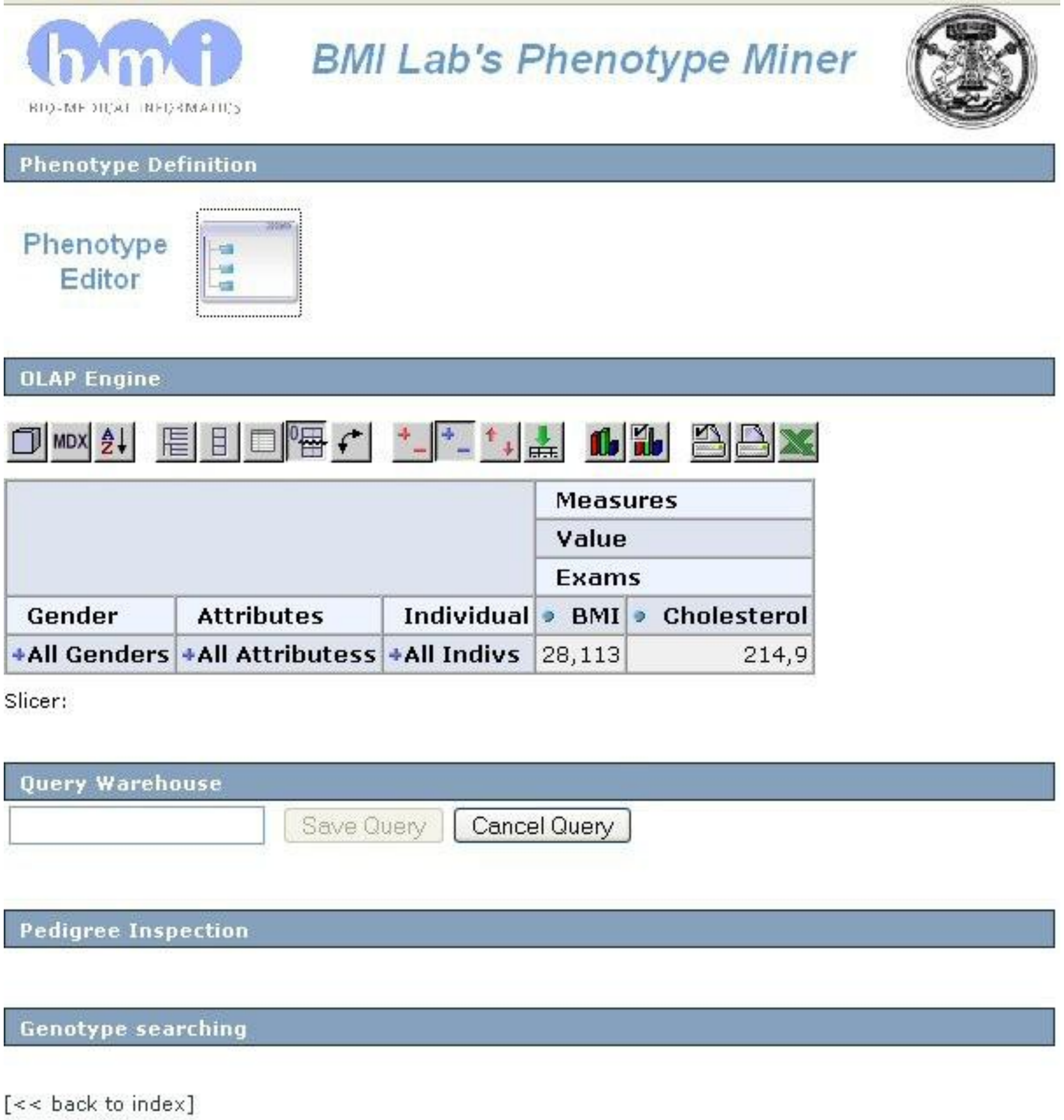
Nuzzo et al, BMC Bioinformatics, 2008[1]
Malovini et al, BMC Bioinformatics, 2008[2]
Nuzzo et al, BMC Bioinformatics, 2009[3]
Riva et al, JBI, 2009[4]

The phenotype miner

Phenotype Editor

Dynamic inspection table (provided by Mondrian)

Genotypic data retrieving utilities



BMI Lab's Phenotype Miner

BIO-MEDICAL INFORMATICS

Phenotype Definition

Phenotype Editor

OLAP Engine

			Measures	
			Value	
			Exams	
Gender	Attributes	Individual	BMI	Cholesterol
+All Genders	+All Attributess	+All Indivs	28,113	214,9

Slicer:

Query Warehouse

Pedigree Inspection

Genotype searching

[<< back to index]

The phenotype miner

Phenotype Editor

Dynamic inspection table (provided by Mondrian)

Genotypic data retrieving utilities

- 1or2.xml
- 2or.xml
- 3and.xml
- colesteroloHDL_magg_80.xml
- colest_gozzo.xml
- col_gozzo_tsh.xml
- or3.xml
- test1.xml
- test2.xml

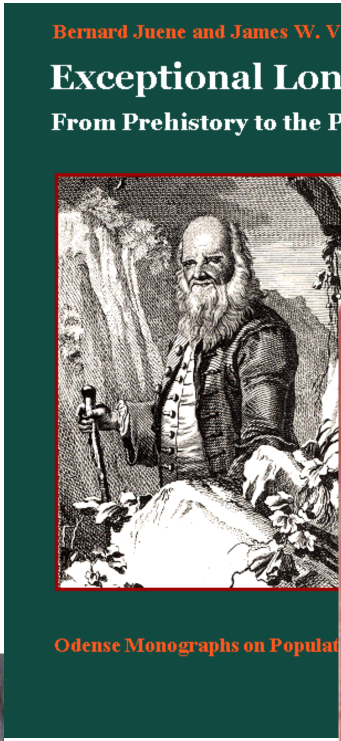
- 2or.xml
  - 3and.xml
    - Tsh - III Generazione > 4.0
      - Ematocrito > 60.0
        - Colesterolo HDL > 80.0
  - tsh_3generaz_magg_4.xml
    - Tsh - III Generazione > 4.0

Condition From 3and.xml
Tsh - III Generazione > 4.0
Ematocrito > 60.0
Colesterolo HDL > 80.0
Condition From tsh_3generaz_magg_4.xml
Tsh - III Generazione > 4.0

Risk stratification

The Longevity Revolution is Here

World-renowned economist Ranga Chand in partnership with Stone & Co. Ltd. introduce the Longevity Fund – the first and only Canadian mutual fund poised to profit from an unprecedented and inevitable growth in global aging.



Bernard Juene and James W. Vaupel(Eds.)

Exceptional Longevity:

From Prehistory to the Present

Odense Monographs on Populat

TIME

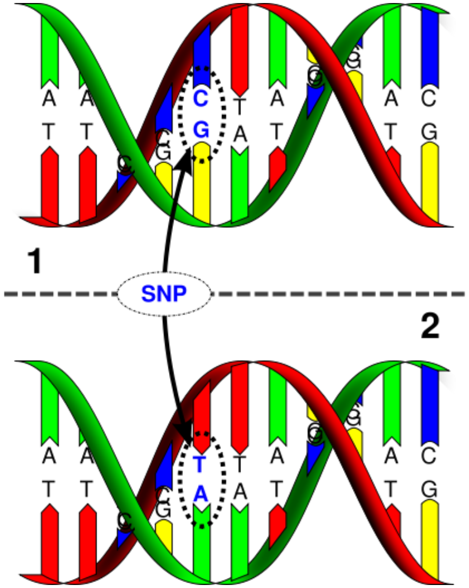
Secrets of Asian Longevity

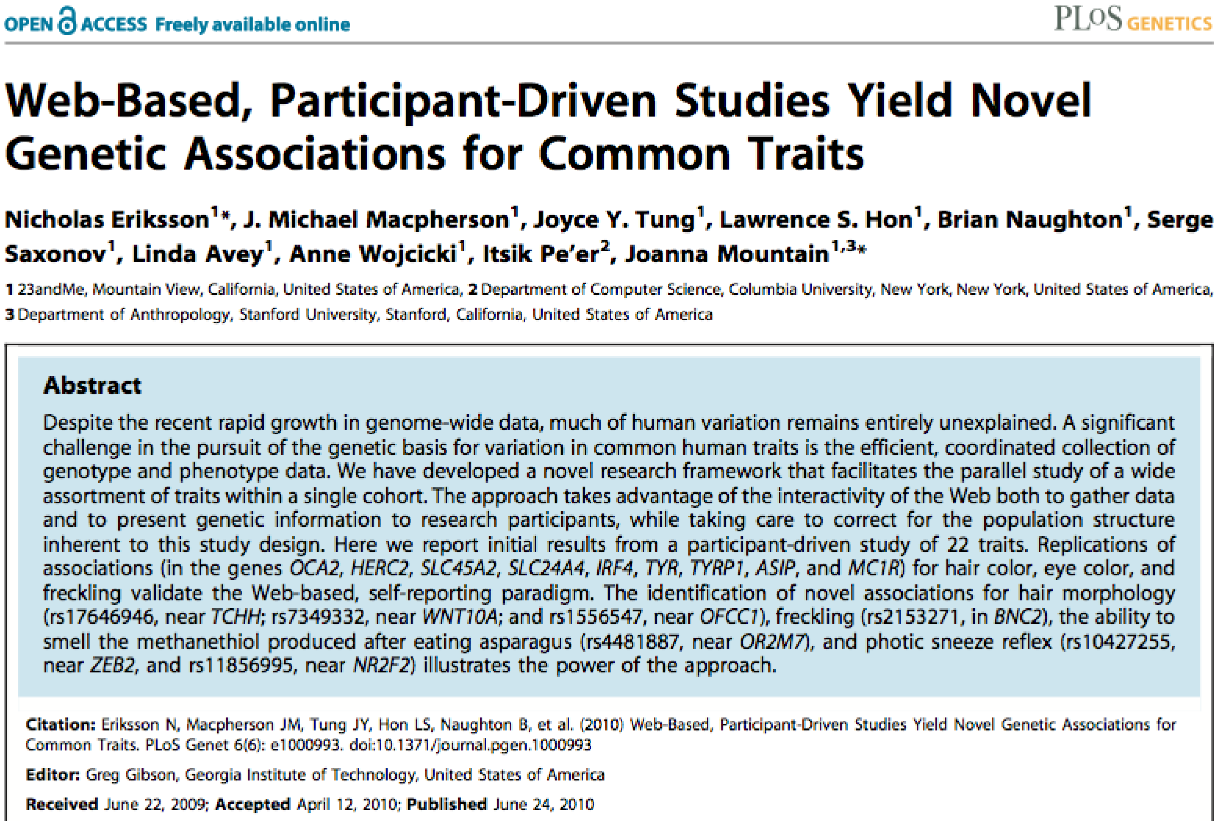
The latest science on how to live longer, healthier and happier

Genome-wide association studies

Controls

Cases





OPEN ACCESS Freely available online PLoS GENETICS

Web-Based, Participant-Driven Studies Yield Novel Genetic Associations for Common Traits

Nicholas Eriksson[1]*, J. Michael Macpherson[1], Joyce Y. Tung[1], Lawrence S. Hon[1], Brian Naughton[1], Serge Saxonov[1], Linda Avey[1], Anne Wojcicki[1], Itsik Pe'er[2], Joanna Mountain[1,3]*

1 23andMe, Mountain View, California, United States of America, **2** Department of Computer Science, Columbia University, New York, New York, United States of America, **3** Department of Anthropology, Stanford University, Stanford, California, United States of America

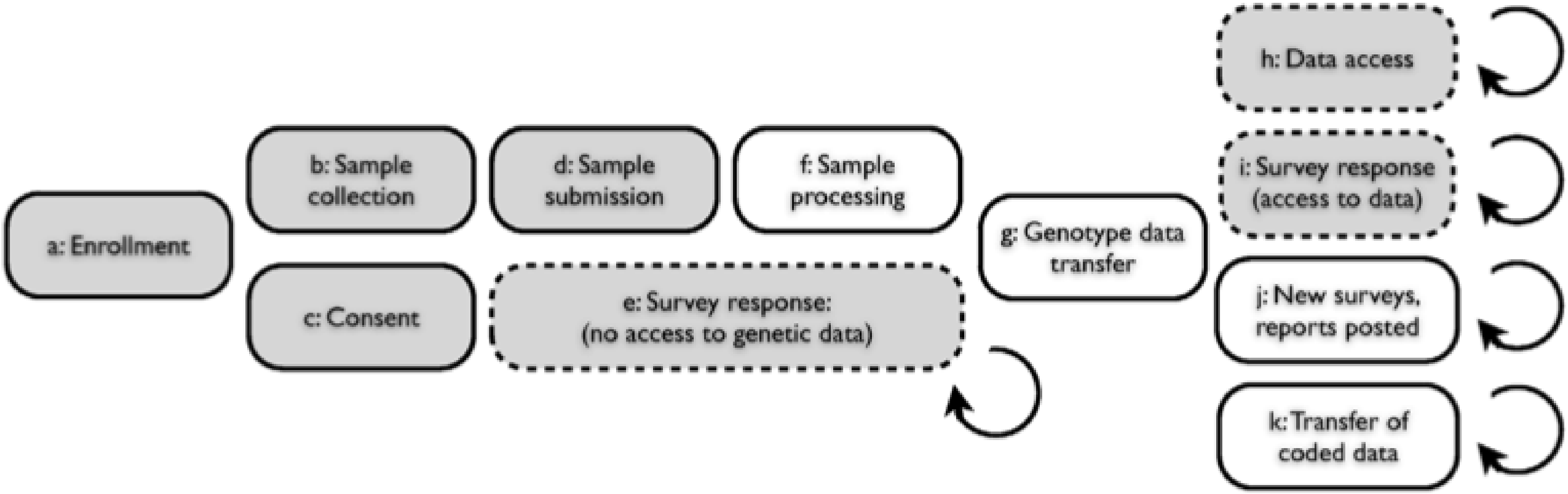
Abstract

Despite the recent rapid growth in genome-wide data, much of human variation remains entirely unexplained. A significant challenge in the pursuit of the genetic basis for variation in common human traits is the efficient, coordinated collection of genotype and phenotype data. We have developed a novel research framework that facilitates the parallel study of a wide assortment of traits within a single cohort. The approach takes advantage of the interactivity of the Web both to gather data and to present genetic information to research participants, while taking care to correct for the population structure inherent to this study design. Here we report initial results from a participant-driven study of 22 traits. Replications of associations (in the genes *OCA2*, *HERC2*, *SLC45A2*, *SLC24A4*, *IRF4*, *TYR*, *TYRP1*, *ASIP*, and *MC1R*) for hair color, eye color, and freckling validate the Web-based, self-reporting paradigm. The identification of novel associations for hair morphology (rs17646946, near *TCHH*; rs7349332, near *WNT10A*; and rs1556547, near *OFCC1*), freckling (rs2153271, in *BNC2*), the ability to smell the methanethiol produced after eating asparagus (rs4481887, near *OR2M7*), and photic sneeze reflex (rs10427255, near *ZEB2*, and rs11856995, near *NR2F2*) illustrates the power of the approach.

Citation: Eriksson N, Macpherson JM, Tung JY, Hon LS, Naughton B, et al. (2010) Web-Based, Participant-Driven Studies Yield Novel Genetic Associations for Common Traits. PLoS Genet 6(6): e1000993. doi:10.1371/journal.pgen.1000993

Editor: Greg Gibson, Georgia Institute of Technology, United States of America

Received June 22, 2009; **Accepted** April 12, 2010; **Published** June 24, 2010



h: Data access

a: Enrollment

b: Sample collection

c: Consent

d: Sample submission

e: Survey response: (no access to genetic data)

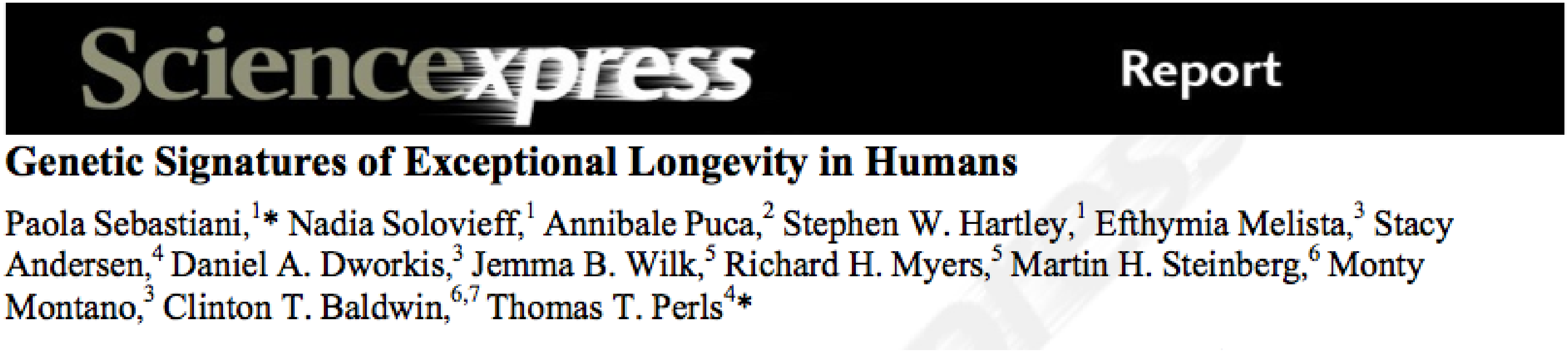
f: Sample processing

g: Genotype data transfer

i: Survey response (access to data)

j: New surveys, reports posted

k: Transfer of coded data

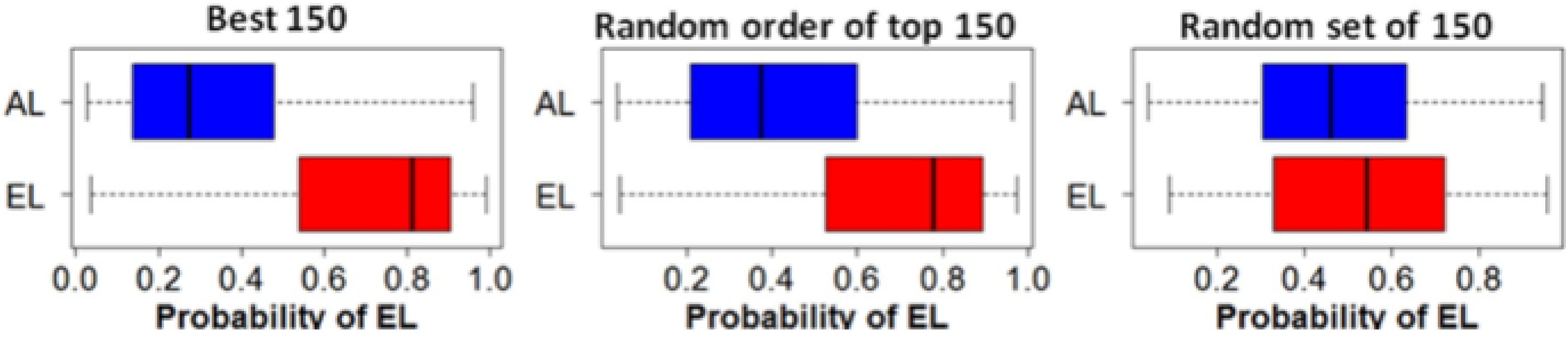


Sciencexpress Report

Genetic Signatures of Exceptional Longevity in Humans

Paola Sebastiani,[1]* Nadia Solovieff,[1] Annibale Puca,[2] Stephen W. Hartley,[1] Efthymia Melista,[3] Stacy Andersen,[4] Daniel A. Dworkis,[3] Jemma B. Wilk,[5] Richard H. Myers,[5] Martin H. Steinberg,[6] Monty Montano,[3] Clinton T. Baldwin,[6,7] Thomas T. Perls[4]*

Naive Bayes classifier



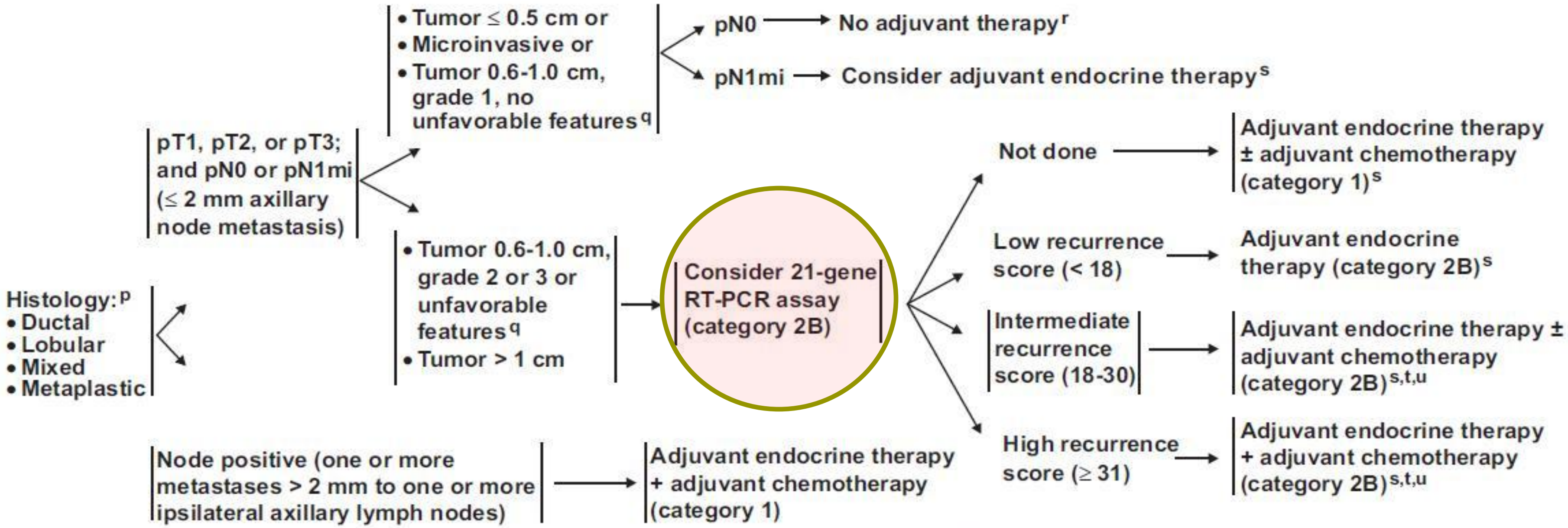
Diagnosis and therapy planning

NCCN® Practice Guidelines in Oncology – V.2.2010

Invasive Breast Cancer

Guidelines Index
Breast Cancer TOC
Staging, Discussion, References

SYSTEMIC ADJUVANT TREATMENT - HORMONE RECEPTOR POSITIVE - HER2 NEGATIVE DISEASE[b]



Histology:[p]
- Ductal
- Lobular
- Mixed
- Metaplastic

pT1, pT2, or pT3; and pN0 or pN1mi (≤ 2 mm axillary node metastasis)

- Tumor ≤ 0.5 cm or
- Microinvasive or
- Tumor 0.6-1.0 cm, grade 1, no unfavorable features[q]

pN0 → No adjuvant therapy[r]

pN1mi → Consider adjuvant endocrine therapy[s]

- Tumor 0.6-1.0 cm, grade 2 or 3 or unfavorable features[q]
- Tumor > 1 cm

→ Consider 21-gene RT-PCR assay (category 2B)

Not done → Adjuvant endocrine therapy ± adjuvant chemotherapy (category 1)[s]

Low recurrence score (< 18) → Adjuvant endocrine therapy (category 2B)[s]

Intermediate recurrence score (18-30) → Adjuvant endocrine therapy ± adjuvant chemotherapy (category 2B)[s,t,u]

High recurrence score (≥ 31) → Adjuvant endocrine therapy + adjuvant chemotherapy (category 2B)[s,t,u]

Node positive (one or more metastases > 2 mm to one or more ipsilateral axillary lymph nodes) → Adjuvant endocrine therapy + adjuvant chemotherapy (category 1)

See Adjuvant Endocrine Therapy (BINV-I) and Adjuvant Chemotherapy (BINV-J)

---

[b] See Principles of HER2 Testing (BINV-A).

[p] Mixed lobular and ductal carcinoma as well as metaplastic carcinoma should be graded based on the ductal component and treated based on this grading. The metaplastic or mixed component does not alter prognosis.

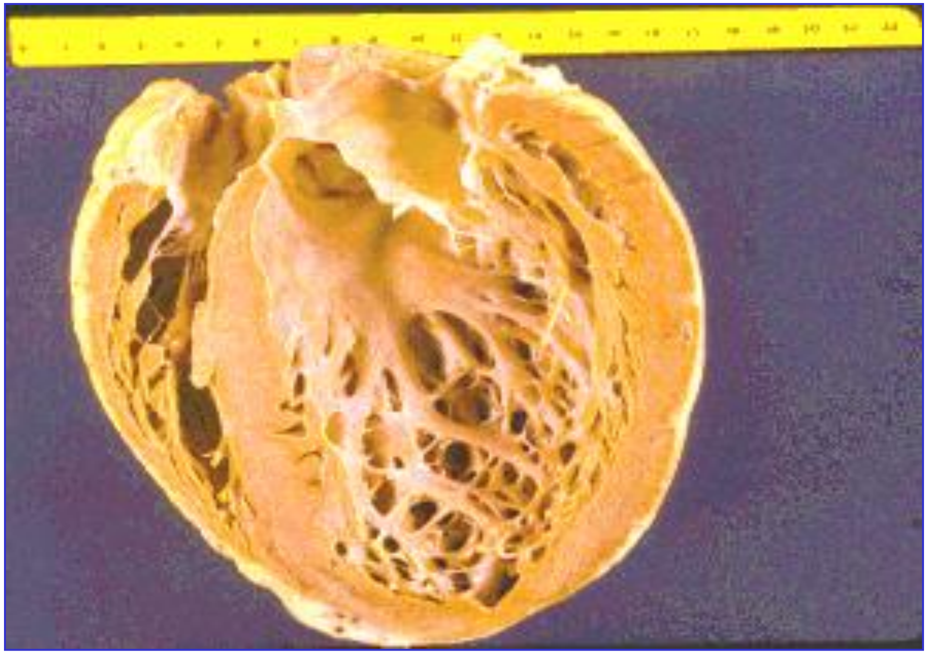
[q] Unfavorable features: angiolymphatic invasion, high nuclear grade, or high histologic grade.

[r] If ER-positive consider endocrine therapy for risk reduction and to diminish the small risk of disease recurrence.

[s] Evidence supports that the magnitude of benefit from surgical or radiation ovarian ablation in premenopausal women with hormone-receptor-positive breast cancer is similar to that achieved with CMF alone. Early evidence suggests similar benefits from ovarian suppression (ie, LHRH agonist) as from ovarian ablation. The combination of ovarian ablation/suppression plus endocrine therapy may be superior to suppression alone. The benefit of ovarian ablation/suppression in premenopausal women who have received adjuvant chemotherapy is uncertain.

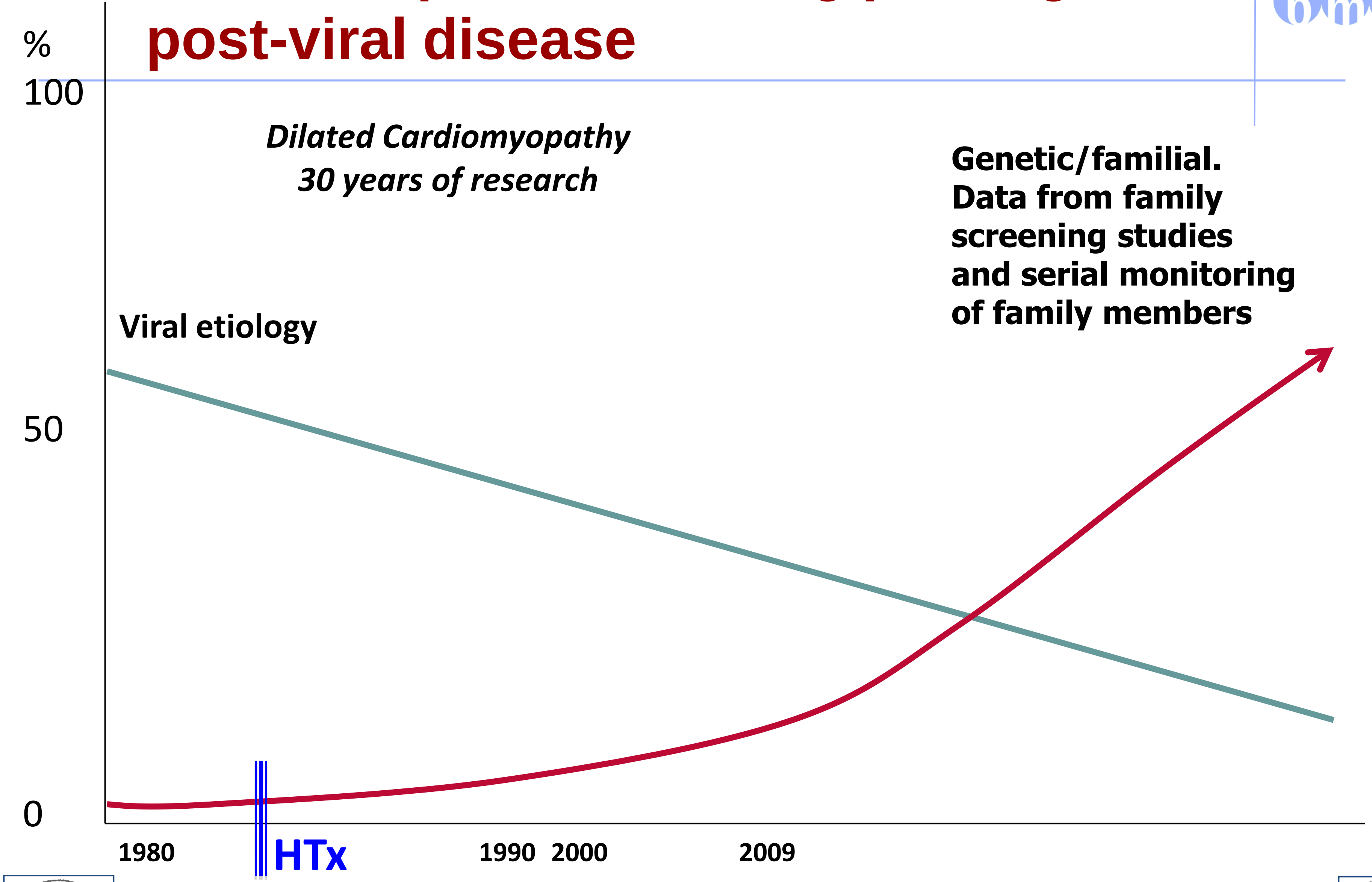
[t] Chemotherapy and endocrine therapy used as adjuvant therapy should be given sequentially with endocrine therapy following chemotherapy. The benefits of chemotherapy and of endocrine therapy are additive. However, the absolute benefit from chemotherapy may be small. The decision to add chemotherapy to endocrine therapy should be individualized, especially in those with a favorable prognosis and in women age ≥ 60 y where the incremental benefit of chemotherapy may be smaller. Available data suggest sequential or concurrent endocrine therapy with radiation therapy is acceptable.

[u] There are insufficient data to make chemotherapy recommendations for those over 70 y old. Treatment should be individualized with consideration of comorbid conditions.



Dilated cardiomiopathy

DCM: the past and wrong paradigm of post-viral disease



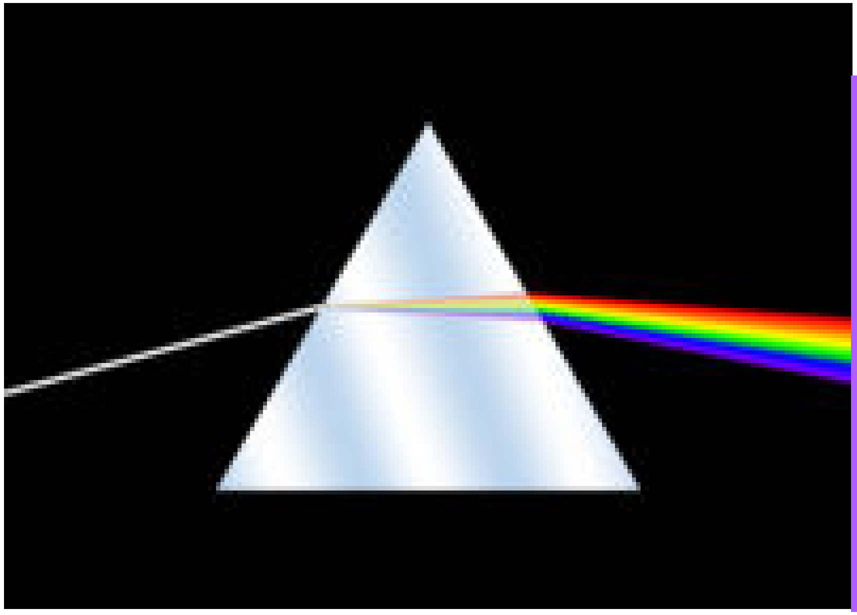
Dilated Cardiomyopathy
30 years of research

Viral etiology

Genetic/familial. Data from family screening studies and serial monitoring of family members

From DCM to…

Clinically oriented genetic investigation

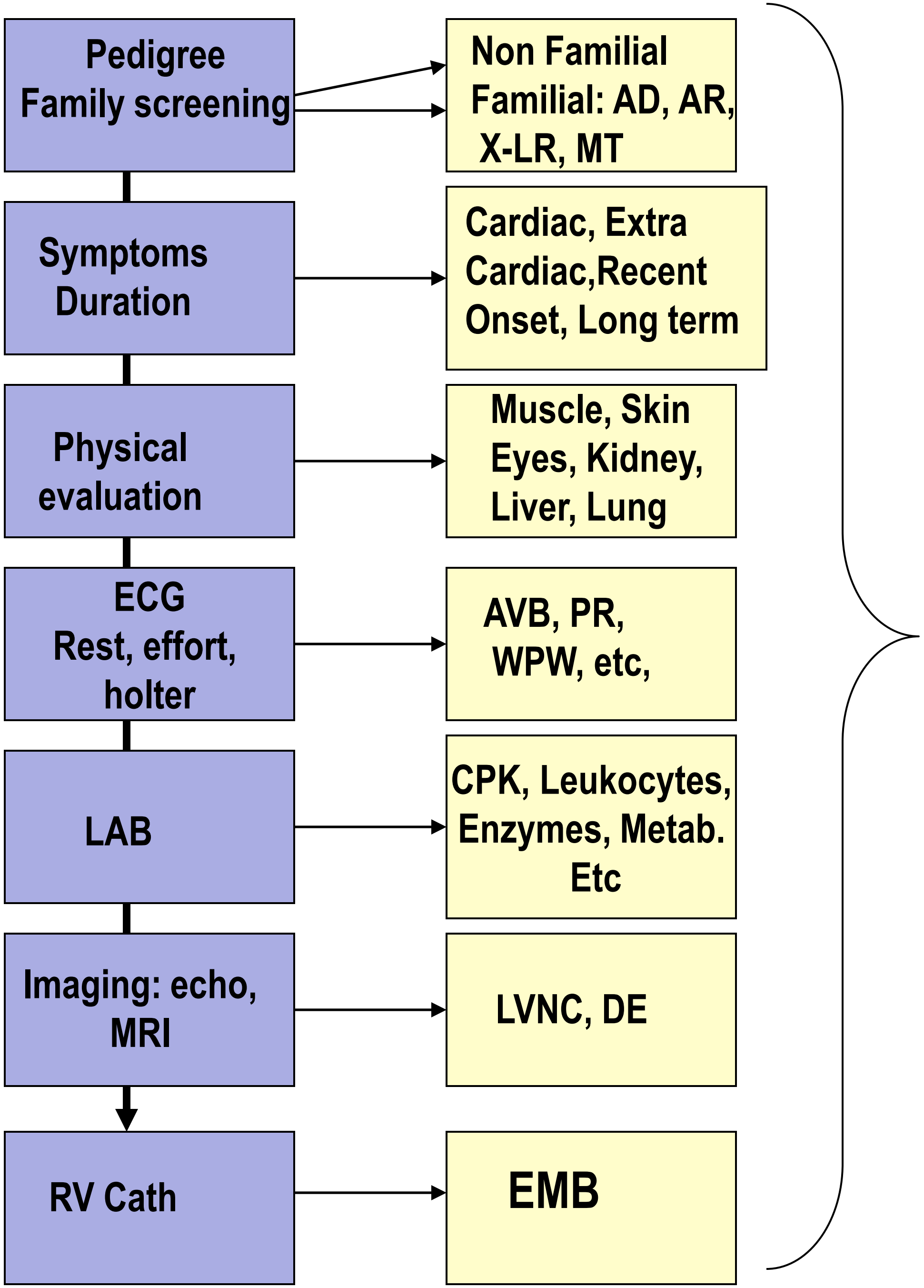


“DCM”

- *Dystrofinopathies*
- *Laminopathies*
- *Desminopathies*
- *Mitocondriopathies*
- *Epicardinopathies*
- *Actinopathies*
- *Zaspopathies*
- *Desmosonopathies*

Concept

- More than 35 genes may cause DCM
- DCM is sometimes accompanied by gene-specific traits → red flags
- Grouping patients according to phenotypes
  - DCM + type of inheritance + cardiac markers + extracardiac markers + any clinical data that may “specify” the subgroups



Pedigree Family screening

Non Familial
Familial: AD, AR, X-LR, MT

Symptoms Duration

Cardiac, Extra Cardiac,Recent Onset, Long term

Physical evaluation

Muscle, Skin Eyes, Kidney, Liver, Lung

ECG Rest, effort, holter

AVB, PR, WPW, etc,

LAB

CPK, Leukocytes, Enzymes, Metab. Etc

Imaging: echo, MRI

LVNC, DE

RV Cath

EMB

Family screening Clinical markers

Diagnostic Hypothesis: Before Genetic Testing

Increasing the number of genotyped CMP
One gene ---> one disease

Case-based-ranking

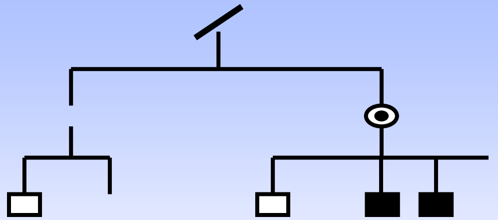
Assign a score to genes based on similarity of the clinical case with previous, and already known, clinical cases

PTPN11 → DCM, HCM

EYA4 → DCM

Noonan syndrome

Current patient



Patient 1 – gene PTPN11

Patient 2 – gene EYA4

…

Case base

Acknowledgements

Bioinformatics and Data Mining group (http://bioinfo.unipv.it)

Harvard Medical School

A.I.LAB
Ljubljana

University of Ljubljana

IRCCS Fondazione C. Mondino

i2b2
Informatics for Integrating Biology & the Bedside

ItalBioNet

SUMMIT
surrogate markers for micro- and macro-vascular hard endpoints for innovative diabetes tools

Home | Site Map | Contact webmaster

INHERITANCE PROJECT

Project Funded by the European Community's 7th Framework Programme

IRCCS Fondazione S. Maugeri

NETTAB 2011: CLINICAL BIOINFORMATICS PAVIA

PAVIA: UNIVERSITY AND HOSPITALS

Organizers

- Paolo Romano
  (IST Genova)

- Riccardo Bellazzi
  (University of Pavia)

- Isaac (Zak) Kohane
  (Harvard Medical School)

Location: Collegio Ghislieri

Aula Magna

Quadriportico

Lardirago Castle

Clinical by Informatics
&
Translational Bioinformatics

Future Developments of Medical Informatics from the Viewpoint of Networked Clinical Research.
http://www.ncbi.nlm.nih.gov/pubmed/19151883

Perspective for medical informatics. Reusing the electronic medical record for clinical research.
http://www.ncbi.nlm.nih.gov/pubmed/19151882?dopt=Abstract

Biomedical informatics and translational medicine.
http://www.ncbi.nlm.nih.gov/pubmed/20187952

Translational informatics: enabling high-throughput research paradigms.
http://www.ncbi.nlm.nih.gov/pubmed/19737991