Clinical Bioinformatics: a research agenda to support health care transformation

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

Hrvoja Bosnjak, Kresimir Pavelic & Sandra Kraljevic Pavelic

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

<table>
<thead>
<tr>
<th>Focus of article</th>
<th>Reference</th>
</tr>
</thead>
</table>
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
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.
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

Translational bioinformatics in the cloud: an affordable alternative

Joel T Dudley1,2,3, Yannick Poulion2,3, Rong Chen2,3, Alexander A Morgan1,2,3, Atul J Butte2,3*


Current methodologies for translational bioinformatics.

Lussier YA, Butte AJ, Hunter L.


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
HETEROGENEOUS DATA

Paul Lewis, Cancer Informatics Group
Clinical by Informatics & Translational Bioinformatics

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

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 BIO MEDICAL ONTOLOGY
http://www.bioontology.org/

iDASH
http://idash.ucsd.edu/
Infrastructures
Informatics for Integrating Biology and the Bedside (i2b2)
Isaac Kohane, PI

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

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

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

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

Center for Computational Biology (CCB)
Arthur Toga, PI

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

National Center for Integrative Biomedical Informatics (NCIBI)
Brian D. Athey, PI
Ten thousand views of bioinformatics: a bibliome perspective.
Kohane I.

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

Knowledge discovery

research data

Tissue
Microarray
Genotype
Labs
Form Data

Information Warehouse

clinical data

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

analysis

SAS
R
BioConductor
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

Global Warehouse

All Data sources

i2b2

Workflows

Clinical Research Chart

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

Knowledge repositories

Intelligent query / data mining

Reasoning systems

Biobanks

EMR

Discharge letters

Research data-bases

DW / clinical research chart

HIV
BIOINFORMATICS METHODOLOGY AND TECHNOLOGY TO INTEGRATE CLINICAL AND BIOLOGICAL KNOWLEDGE SUPPORTING ONCOLOGY TRANSATIONAL RESEARCH (ONCO-I2B2)
NLP – extracting information from clinical narratives

Diagram:
- Document
- Section splitter
- Text tokenizer
- POS tagger
- NP chunker
- Diagnosis extractor
- Database
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.

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

terapia consigliata: "almotriptan cpr: una cpr al inizio della crisi; in alternativa/dopo due ore: indometacina supp. 50 mg: una supp."
[...] 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**. [..]
Concept Finder

Diagnosis:

1. Emicrania senza aura

2. Cefalea tensiva episodica sporadica

ICHID – 1.1 - Emicrania senza aura

ICHID – 2 - Cefalea di tipo tensivo
Towards knowledge-discovery support systems

Knowledge repositories

Intelligent query / data mining

Reasoning systems

The Phenotype Miner

Genephony

SNP-2-Net

The ST-Model

Nuzzo et al, BMC Bioinformatics, 2008
Malovini et al, BMC Bioinformatics, 2008
Nuzzo et al, BMC Bioinformatics, 2009
Riva et al, JBI, 2009
The phenotype miner

Phenotype Editor

Dynamic inspection table (provided by Mondrian)

Genotypic data retrieving utilities

Phenotype Editor

BMI Lab's Phenotype Miner

Phenotype Definition

OLAP Engine

Query Warehouse

Pedigree Inspection

Genotype searching
The phenotype miner

Phenotype Editor

Dynamic inspection table (provided by Mondrian)

Genotypic data retrieving utilities
Risk stratification
Genome-wide association studies

Controls

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

Nicholas Eriksson¹, J. Michael Macpherson¹, Joyce Y. Tung¹, Lawrence S. Hon¹, Brian Naughton¹, Serge Saxonov¹, Linda Avey¹, Anne Wojcicki¹, Itzik Pe'er², Joanna Mountain¹,³*

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 genotypic and phenotypic 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, TYRIP1, 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; rs734932, near WNT16A; 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.


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

Received June 22, 2009; Accepted April 12, 2010; Published June 24, 2010
Naive Bayes classifier
Diagnosis and therapy planning
**SYSTEMIC ADJUVANT TREATMENT - HORMONE RECEPTOR POSITIVE - HER2 NEGATIVE DISEASE**

- **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
  - pN0 → No adjuvant therapy
  - pN1mi → Consider adjuvant endocrine therapy

- **Consider 21-gene RT-PCR assay (category 2B)**
  - Low recurrence score (< 18) → Adjuvant endocrine therapy (category 1)
  - Intermediate recurrence score (18-30) → Adjuvant endocrine therapy ± adjuvant chemotherapy (category 2B)
  - High recurrence score (≥ 31) → Adjuvant endocrine therapy + adjuvant chemotherapy (category 2B)

**Histology:**
- Ductal
- Lobular
- Mixed
- Metaplastic

**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)**

**See Principles of HER2 Testing (BINV-A).**

**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.**

**Unfavorable features:** Angiolympathic invasion, high nuclear grade, or high histologic grade.

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

**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.**

**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.**

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

Centre for Inherited Cardiovascular Diseases - IRCCS Policlinico San Matteo - Pavia
Clinically oriented genetic investigation

From DCM to...

“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

Symptoms Duration

Physical evaluation

ECG Rest, effort, holter

LAB

Imaging: echo, MRI

RV Cath

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

Cardiac, Extra
Cardiac, Recent
Onset, Long term

Muscle, Skin
Eyes, Kidney,
Liver, Lung

AVB, PR, WPW, etc,

CPK, Leukocytes,
Enzymes, Metab.
Etc

LVNC, DE

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

Harvard Medical School

University of Ljubljana

IRCCS Fondazione C. Mondino

i2b2
Informatics for Integrating Biology & the Bedside

ItaBiNet

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

IRCCS Fondazione S. Maugeri

Project Funded by the European Community’s 7th Framework Programme
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.

Perspective for medical informatics. Reusing the electronic medical record for clinical research.

Biomedical informatics and translational medicine.

Translational informatics: enabling high-throughput research paradigms.