

Data-driven Temporal Phenotyping of the GCAT Cohort from Electronic Health Records

Xavier Duran-Albareda
GCAT Genomes for Life
Institut de Recerca Germans Trias i Pujol (IGTP)

FI-2017 AGAUR
October 28th, 2016

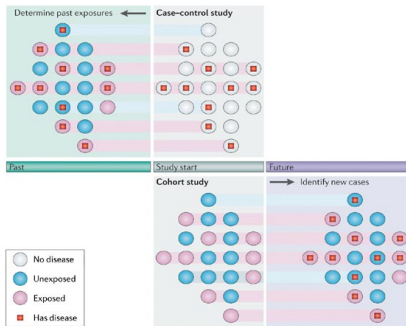
A Cohort Study of the Genomes of Catalonia



- ▶ Genomic population-based cohort
- ▶ 25.000 participants by 2018
- ▶ General, asymptomatic population
- ▶ Analyze genomics and health interplay in the next 20 years



Cohort studies linked to Electronic Health Records



Copyright © 2006 Nature Publishing Group
Nature Reviews | Genetics

- ▶ Identify the **genomic basis** of **common disorders**
- ▶ **Electronic Health Records** are the major source of phenotype data available
- ▶ From hypothesis-driven to **data-driven** phenotyping

Temporal disease trajectories and disease clusters



ARTICLE

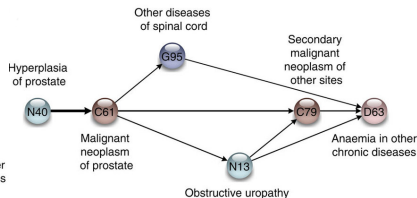
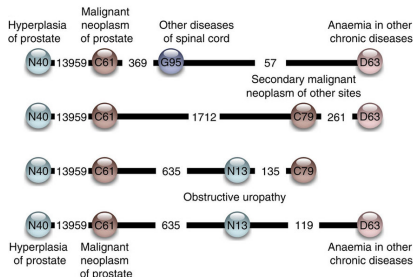
Received 22 Jan 2014 | Accepted 1 May 2014 | Published 24 Jun 2014

DOI: 10.1038/ncomms5022

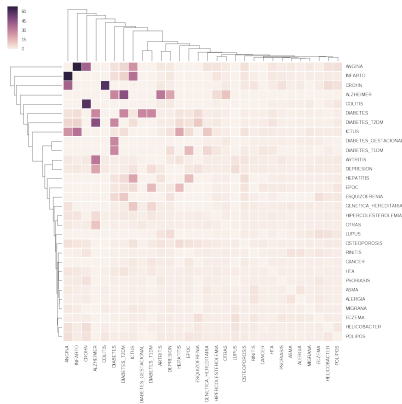
OPEN

Temporal disease trajectories condensed from population-wide registry data covering 6.2 million patients

Anders Boeck Jensen^{1,2}, Pope L. Moseley^{2,3}, Tudor I. Oprea^{1,3,4}, Sabrina Gade Ellesee², Robert Eriksson^{1,2}, Henriette Schmock⁵, Peter Bjødstrup Jensen², Lars Juhl Jensen² & Søren Brunak^{1,2}



GCAT Snapshot and the Diseasome



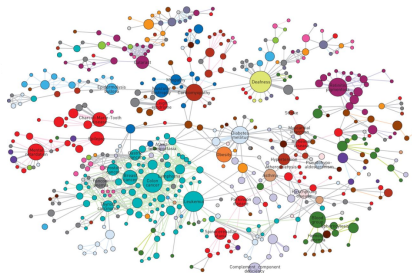
OPEN ACCESS Freely available online

PLOS COMPUTATIONAL BIOLOGY

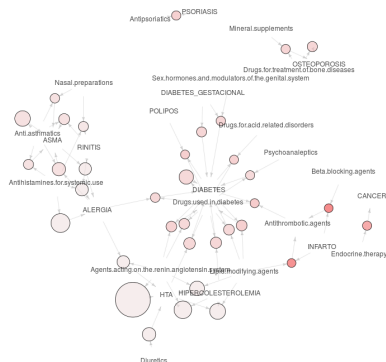
A Dynamic Network Approach for the Study of Human Phenotypes

César A. Hidalgo^{1,*}, Nicholas Blumm^{2,3}, Albert-László Barabási^{2,4,5}, Nicholas A. Christakis⁶

¹ Center for International Development and Harvard Kennedy School, Harvard University, Cambridge, Massachusetts, United States of America, ² Center for Network Science, Department of Physics, Biology and Computer Science, Northeastern University, Boston, Massachusetts, United States of America, ³ College of Computer and Information Science, Northeastern University, Boston, Massachusetts, United States of America, ⁴ Center for Complex Network Research and Department of Physics, University of Notre Dame, Notre Dame, Indiana, United States of America, ⁵ Department of Medicine, Harvard Medical School, Boston, Massachusetts, United States of America, ⁶ Department of Health Care Policy and Department of Medicine, Harvard Medical School, Boston, Massachusetts, United States of America



Association rule mining



rules	support	confidence	lift	chiSquare
{Drugs used in diabetes} => {DIABETES}	0.039209193	0.9157088	16.936870	3940.3809
{Agents acting on the renin angiotensin system} => {HTA}	0.156305245	0.9406028	3.968346	3690.2555
{Anti asthmatics} => {ASMA}	0.042133176	0.8194842	10.917174	2953.0285
{HTA,Drugs used in diabetes} => {DIABETES}	0.018709487	0.9338235	17.271919	2100.2424
{DIABETES, Lipid modifying agents} => {Drugs used in diabetes}	0.014142605	0.8648649	22.493114	2084.5341
{ALERGIA,Ans asthmatics} => {ASMA}	0.028874484	0.8376068	11.214350	2041.0416
{HIPERCOLESTEROLEMIA,Drugs used in diabetes} => {DIABETES}	0.016352387	0.9652174	17.852577	1898.9545
{DIABETES,Agents acting on the renin angiotensin system} => {Drugs used in diabetes}	0.013111373	0.8018018	20.852991	1778.5342
{Drugs used in diabetes, Lipid modifying agents} => {DIABETES}	0.014142605	0.9142857	16.910549	1543.0976
{Agents acting on the renin angiotensin system,Drugs used in diabetes} => {DIABETES}	0.013111373	0.8900000	16.461362	1386.7701
{Antihistamines for systemic use} => {ALERGIA}	0.061431939	0.8424242	3.998884	1281.5520
{HIPERCOLESTEROLEMIA,Agents acting on the renin angiotensin system} => {HTA}	0.054065999	0.9683377	4.075063	1183.1037
{Drugs for treatment of bone diseases} => {OSTEOPOROSIS}	0.010312316	0.8433735	17.723899	1174.1679
{DIABETES, GASTACIONAL} => {DIABETES}	0.009675722	1.0000000	18.496913	1148.2295
{RENITIS, Anti asthmatics} => {ASMA}	0.014731880	0.8849558	11.848283	1091.6168
{Endocrine therapy} => {CANCER}	0.004419564	0.9677419	36.698504	1074.8972
{HTA, Lipid modifying agents} => {HIPERCOLESTEROLEMIA}	0.043753683	0.8272981	3.525235	741.1850
{ALERGIA,Drugs used in diabetes} => {DIABETES}	0.006629346	0.9375000	17.339918	737.7111
{Agents acting on the renin angiotensin system, Lipid modifying agents} => {HTA}	0.036535562	0.9117647	3.836986	710.8339

Diseases and medications rule network from GCAT cohort

Summary

- ▶ Data-driven temporal analysis of the cohort
- ▶ Identify subgroups of patients with distinct mechanisms of disease
- ▶ Genome associations with disease subgroups
- ▶ Risk models for common disorders
- ▶ Stratified medicine

Acknowledgements



Rafael de Cid, PhD



LARCA. Laboratori d'Algorísmia Relacional, Complexitat i Aprenentatge
UNIVERSITAT POLITÈCNICA DE CATALUNYA

Ricard Gavaldà, PhD