Computational Methods in Genomics

Introduction

Sami Khuri
Department of Computer Science
San José State University
San José, California, USA
khuri@cs.sjsu.edu
www.cs.sjsu.edu/faculty/khuri

Outline

- Top 25 Questions for the next 25 years [Science]
- Importance of Algorithms: RECOMB 2010
- Genetics in Medicine – Sixth and Seventh Ed.
- Finnish Disease Heritage

Pathway to Genomic Medicine

- Part One: Human Genome Project (HGP)
- Part Two: ENCODE Project
- Part Three: International HapMap Project
- Part Four: Genomic Medicine

Science: Top 25 Questions (I)

* What Is the Universe Made Of?
* What is the Biological Basis of Consciousness?
- **Why Do Humans Have So Few Genes?**
- **To What Extent Are Genetic Variation and Personal Health Linked?**

* Can the Laws of Physics Be Unified?
* How Much Can Human Life Span Be Extended?
- **What Controls Organ Regeneration?**
- **How Can a Skin Cell Become a Nerve Cell?**
- **How Does a Single Somatic Cell Become a Whole Plant?**

* How Does Earth's Interior Work?
* Are We Alone in the Universe?
* How and Where Did Life on Earth Arise?

Science: Top 25 Questions (II)

- **What Determines Species Diversity?**
- **What Genetic Changes Made Us Uniquely Human?**

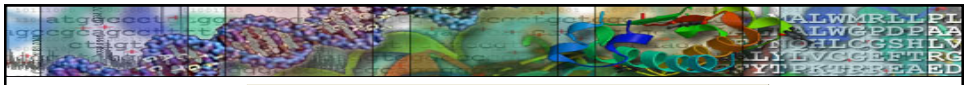
* How Are Memories Stored and Retrieved?
- **How Did Cooperative Behavior Evolve?**
- **How Will Big Pictures Emerge from a Sea of Biological Data?**

* How Far Can We Push Chemical Self-Assembly?
* What Are the Limits of Conventional Computing?
- **Can We Selectively Shut Off Immune Responses?**

* Do Deeper Principles Underlie Quantum Uncertainty and Nonlocality?
* Is an Effective HIV Vaccine Feasible?
* How Hot Will the Greenhouse World Be?
* What Can Replace Cheap Oil -- and When?

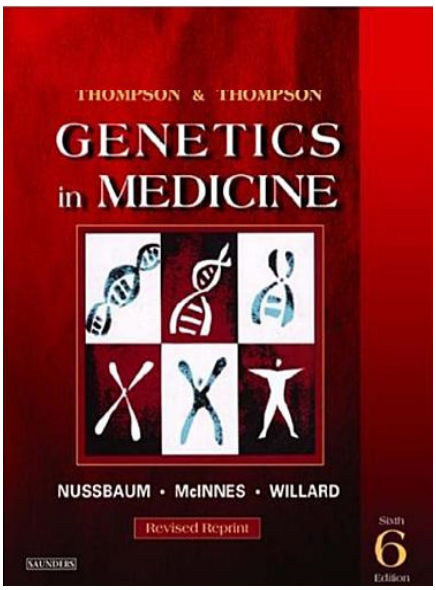
RECOMB 2010

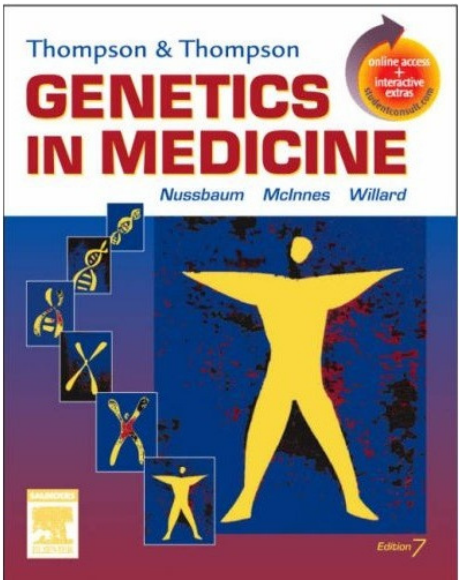
#	Authors, Title
1	Nir Atias and Roded Sharan. An Algorithmic Framework for Predicting Side-Effects of Drugs
2	Ferhat Ay and Tamer Kahveci. SubMAP: Aligning metabolic pathways with subnetwork mappings
3	Sivan Bercovici, Itai Sharon, Ron Pinter and Tomer Shlomi. Pathway-Based Functional Analysis of Metagenomes
4	Sivan Bercovici and Dan Geiger. Admixture Aberration Analysis: Application to mapping in admixed population using pooled DNA
5	José Caldas and Samuel Kaski. Hierarchical Generative Biclustering for MicroRNA Expression Analysis
6	Salim A. Chowdhury, Rod K. Nibbe, Mark R. Chance and Mehmet Koyuturk. Subnetwork State Functions Define Dysregulated Subnetworks in Cancer
7	Manfred Claassen, Ruedi Aebersold and Joachim M. Buhmann. Proteome Coverage Prediction for Integrated Proteomics Datasets
8	Timothy Danford, Robin Dowell and David Gifford. Discovering Regulatory Overlapping RNA Transcripts
9	Constantinos Daskalakis and Sebastien Roch. Alignment-Free Phylogenetic Reconstruction
10	Jianxing Feng, Wei Li and Tao Jiang. Inference of Isoforms from Short Sequence Reads
11	Bjarni Halldorsson, Derek Aguiar, Ryan Tarpine and Sorin Istrail. The Clark Phase-able Sample Size Problem: Long-range Phasing and Loss of Heterozygosity in GWAS
12	Michael Hirsch, Bernhard Schoelkopf and Michael Habeck. A new algorithm for improving the resolution of cryo-EM density maps
13	Richard Jang, Xin Gao and Ming Li. Towards Automated Structure-based NMR Assignment
14	Kyowon Jeong, Sangtae Kim, Nuno Bandeira and Pavel Pevzner. Gapped Spectral Dictionaries and Their Applications for Databases Searches of Tandem Mass Spectra



15	Wei-Chun Kao and Yun S. Song. naiveBayesCall: An efficient model-based base-calling algorithm for high-throughput sequencing
16	David Kelley and Carl Kingsford. Extracting between-pathway models from E-MAP interactions using expected graph compression
17	Yoo-Ah Kim, Stefan Wuchty and Teresa Przytycka. Simultaneous Identification of Causal Genes and Dys-regulated Pathways in Complex Diseases
18	Geoffrey Koh, David Hsu and P. S. Thiagarajan. Incremental Signaling Pathway Modeling by Data Integration
19	Adam Kowalczyk, Justin Bedo, Thomas Conway and Bryan Beresford-Smith. Poisson Margin Test for Normalisation Free Significance Analysis of NGS Data
20	Christos Kozanitis, Chris Saunders, Semyon Kruglyak, Vineet Bafna and George Varghese. Compressing genomic sequence fragments using SlimGene
21	Fumei Lam, Charles H. Langley and Yun S. Song. On the Genealogy of Asexual Diploids
22	Jonathan Laserson, Vladimir Jojic and Daphne Koller. Genovo: De Novo Assembly For Metagenomics
23	Seunghak Lee, Eric Xing and Michael Brudno. MoGUL: Detecting Common Insertions and Deletions in a Population
24	Navodit Misra, Guy Blelloch, R Ravi and Russell Schwartz. Generalized Buneman pruning for inferring the most parsimonious multi-state phylogeny
25	Laurent Noé, Marta Gîrdea and Gregory Kucherov. Seed design framework for mapping AB SOLiD reads
26	Bogdan Pasaniuc, Noah Zaitlen and Eran Halperin. Accurate estimation of expression levels of homologous genes in RNA-seq experiments
27	Benedict Paten, Mark Diekhans, Jian Ma, Bernard Suh and David Haussler. Cactus Graphs for Genome Comparisons
28	Yu Peng, Henry C.M. Leung, SM Yiu and Francis Chin. IDBA - A Practical Iterative de Bruijn Graph De Novo Assembler

29	Sheila Reynolds, Zhiping Weng, Jeff Bilmes and William Stafford Noble. Predicting nucleosome positioning using multiple evidence tracks
30	Barna Saha, Allison Hoch, Samir Khuller, Louiqa Raschid and Xiao-Ning Zhang. Dense Subgraphs with Restrictions and Applications to Gene Annotation Graphs
31	Raheleh Salari, Mathias Mohl, Sebastian Will, S. Cenk Sahinalp and Rolf Backofen. Time and space efficient RNA-RNA interaction prediction via sparse folding
32	Manu Setty, Alexander Gusev and Itsik Pe'er. HLA type inference via haplotypes identical by descent
33	Fabio Vandin, Eli Upfal and Benjamin Raphael. Algorithms for Detecting Significantly Mutated Pathways in Cancer
34	Christian Widmer, Jose Miguel Leiva-Murillo, Yasemin Altun and Gunnar Rätsch. Leveraging Sequence Classification by Taxonomy-based Multitask Learning
35	Yu-Wei Wu and Yuzhen Ye. A novel abundance-based algorithm for binning metagenomic sequences using l-tuples
36	Jianyang Zeng, Pei Zhou and Bruce Donald. A Markov Random Field Framework for Protein Side-Chain Resonance Assignment





Preface of the Seventh Edition

Much has changed, however, since the last edition of this book. Completion of the HGP provides us with a catalogue of all human genes, their sequence, and an extensive, and still growing, database of human variation. Genomic information has stimulated the creation of powerful new tools that are changing human genetics research and medical genetics practice. We therefore have expanded the scope of the book to incorporate the concepts of "**Personalized Medicine**" into ***Genetics in Medicine*** by providing more examples of how genomics is being used to identify the contributions made by genetic variation to disease susceptibility and treatment outcomes.

Finnish Disease Heritage

Hum Genet (2003) 112 : 441–456
DOI 10.1007/s00439-002-0875-3

REVIEW ARTICLE

Reijo Norio

Finnish Disease Heritage I: characteristics, causes, background

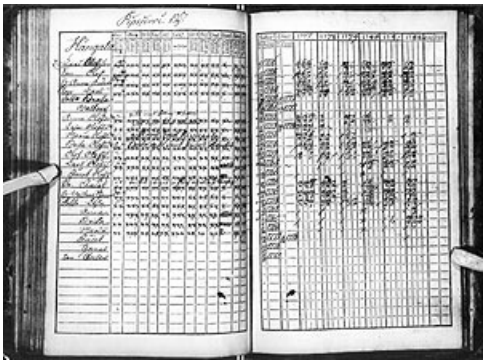
Abstract This review of the Finnish Disease Heritage (FDH), a group of rare hereditary diseases that are over-represented in Finland, includes the following topics: FDH characteristics, causes and background, primary theory, revis(it)ed theory, consanguineous marriages in Finland, internal migration of the 1500s, family series for further FDH studies, geography and population structure as a basis for FDH, geography of individual diseases, the structure of FDH families, family structure in individual diseases, Finnish gene mutations, linkage disequilibrium and haplotypes, age of gene mutations, frequencies of disease genes and carriers, and a short description of the possible future of FDH.

Registers of the Lutheran Church

- Registers from the Lutheran Church are a trove of information for scientists hunting for clues to the inheritance of Finland's distinctive diseases.
- Voluminous congregational records document baptisms, marriages, moves, and deaths throughout the country between the 1700s and the 1960s.
- Geneticists use the registers to trace ancestry back 6 to 10 generations.

"Finland's Fascinating Genes" by J. Wheelwright

Church Records



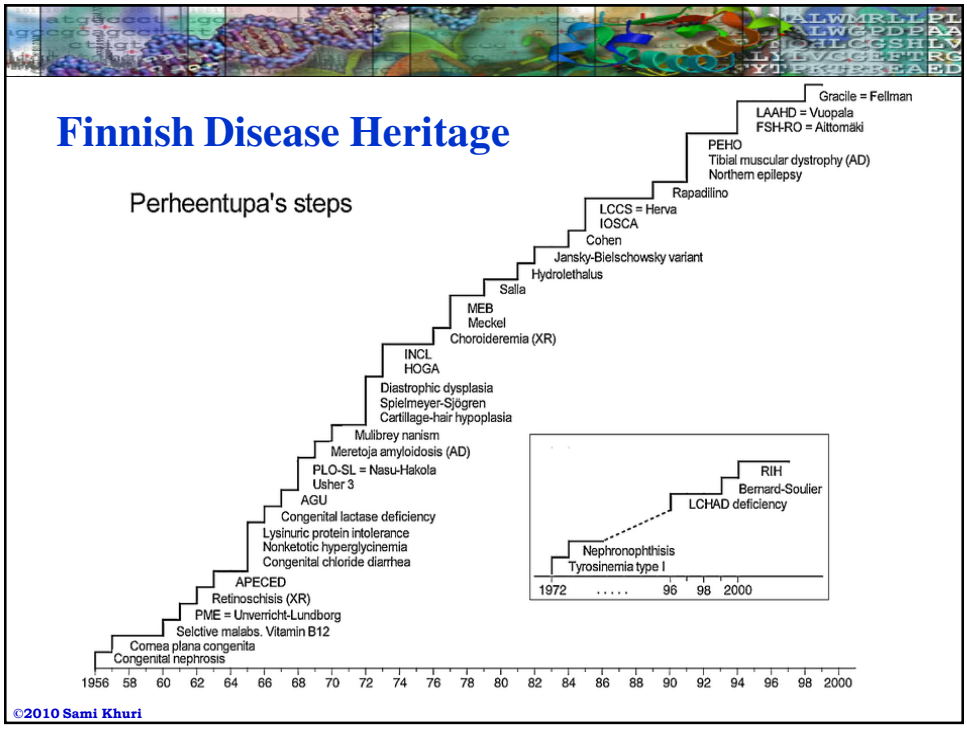
Church Records, Northern Finland, 1777-1782

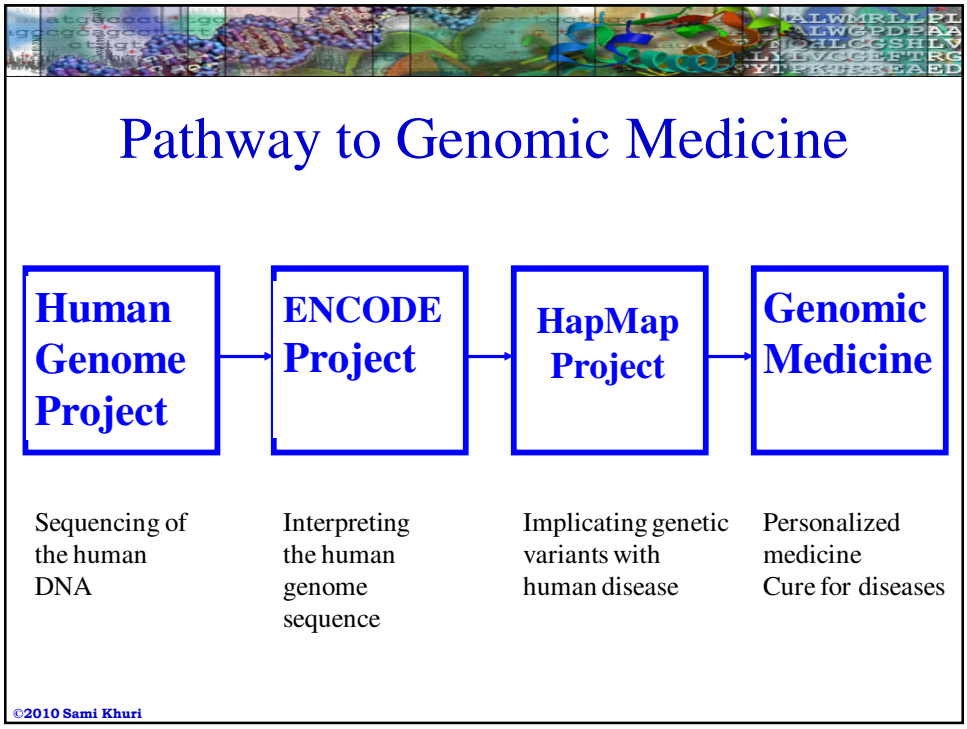
"Finland's Fascinating Genes" by J. Wheelwright

Church Records

- The people in this land of lakes and forests are so alike that scientists can filter out the genes that contribute to heart disease, diabetes, and asthma
- Finland's population has grown tenfold since 1750, with scarcely any of this growth due to immigration. A study of a small Finnish village in the 19th century found that although few weddings occurred between cousins, half of the marriages were between village residents.

"Finland's Fascinating Genes" by J. Wheelwright


Finnish Disease Heritage
Perheentupa's steps
Gracile = Fellman
LAAHD = Vuopala
FSH-RO = Aittomäki
PEHO
Tibial muscular dystrophy (AD)
Northern epilepsy
Rapadilino
LCCS = Herva
IOSCA
Cohen
Jansky-Bielschowsky variant
Hydrolethalus
Salla
MEB
Meckel
Choroideremia (XR)
INCL
HOGA
Diastrophic dysplasia
Spielmeyer-Sjögren
Cartillage-hair hypoplasia
Mulibrey nanism
Meretoja amyloidosis (AD)
PLO-SL = Nasu-Hakola
Usher 3
AGU
Congenital lactase deficiency
Lysinuric protein intolerance
Nonketotic hyperglycinemia
Congenital chloride diarrhea
APECED
Retinoschisis (XR)
PME = Unverricht-Lundborg
Selctive malabs. Vitamin B12
Cornea plana congenita
Congenital nephrosis
1956 58 60 62 64 66 68 70 72 74 76 78 80 82 84 86 88 90 92 94 96 98 2000
RIH
Bernard-Soulier
LCHAD deficiency
Nephronophthisis
Tyrosinemia type I
1972 ..... 96 98 2000
©2010 Sami Khuri


Pathway to Genomic Medicine
Human Genome Project
ENCODE Project
HapMap Project
Genomic Medicine
Sequencing of the human DNA
Interpreting the human genome sequence
Implicating genetic variants with human disease
Personalized medicine
Cure for diseases
©2010 Sami Khuri