Advances in Genetics: From Base Pairs to Bedside

Pediatric Grand Rounds

Regina Margarita Zambrano, MD
Department of Pediatrics
Division of Genetics
LSUHSC and CHNOLA

January 18, 2012

Objectives

- To review the recent advances in Medical Genetics
- To discuss the benefits and limitations of recent technologic advances in genetic testing
- To recognize the value of family history and clinical examination in the evaluation of patients as well as ethical, legal and social implications
- To be aware of the complexity of current practice of Genetics requiring referral to specialized centers
- To recognize individualized medicine as our ultimate goal

Disclosure

Nothing to disclose

"We live in a revolutionary age. Our science has caught the spirit of the times, and more improvements have been made in all its branches in the last 20 years than have been made in a century before"

Benjamin Rush, 1791

"All the News That's Fit to Print"

The New York Times

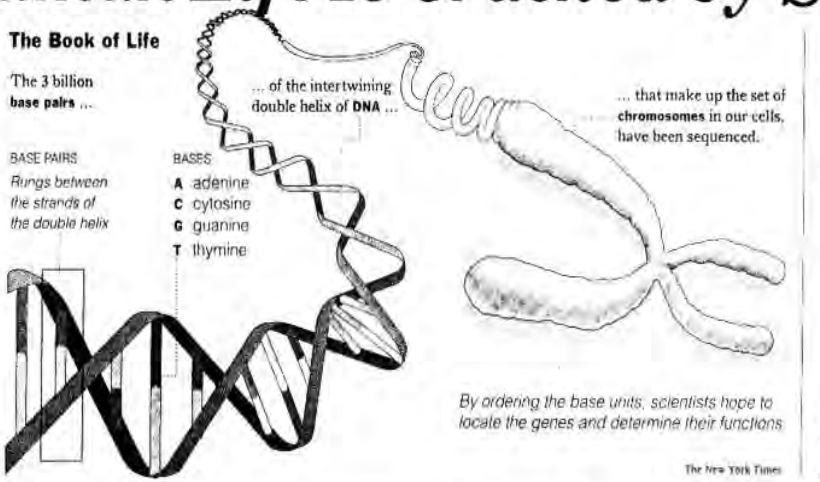
Washington Edition
Washington and Baltimore: Th
storms by afternoon, high ne
Tonight, showers end, lows in th
60's. Tomorrow, showers. Highs
low 80's. Weather map is on Pag

VOL. CXLIX . . . . No. 51,432 Copyright © 2000 The New York Times NU 21205 TUESDAY, JUNE 27, 2000 ONE DOLI

Genetic Code of Human Life Is Cracked by Scientists

Human Genome Project

- Conceived in the mid 1980's
- Debated and argued
- Oct 1, 1990 start date
- Initial efforts focused on technology, maps & sequencing model organisms
- Ended in 2003



Science Times
A special issue

- Putting the genome to work.
- Some information has already paid research dividends.
- Two research methods, two results
- More articles, charts and photos of the genome effort.
- From Mendel to helix to genome.

Section D

Francis S. Collins, head of the Human Genome Project, right, with J. Craig Venter, head of Celera Genomics, after the announcement yesterday that they had finished the first survey of the human genome.

Paul Hosefros/The New York Times

A Pearl and a Hodgepodge: Human DNA

Collins, director of the National Hu-

Though scientists underscore the

A SHARED SUCCI

2 Rivals' Announcem Marks New Medic Era, Risks and All

By NICHOLAS WADE

WASHINGTON, June 26 — achievement that represents nacle of human self-knowledge rival groups of scientists said that they had deciphered the h tary script, the set of instru that defines the human organi

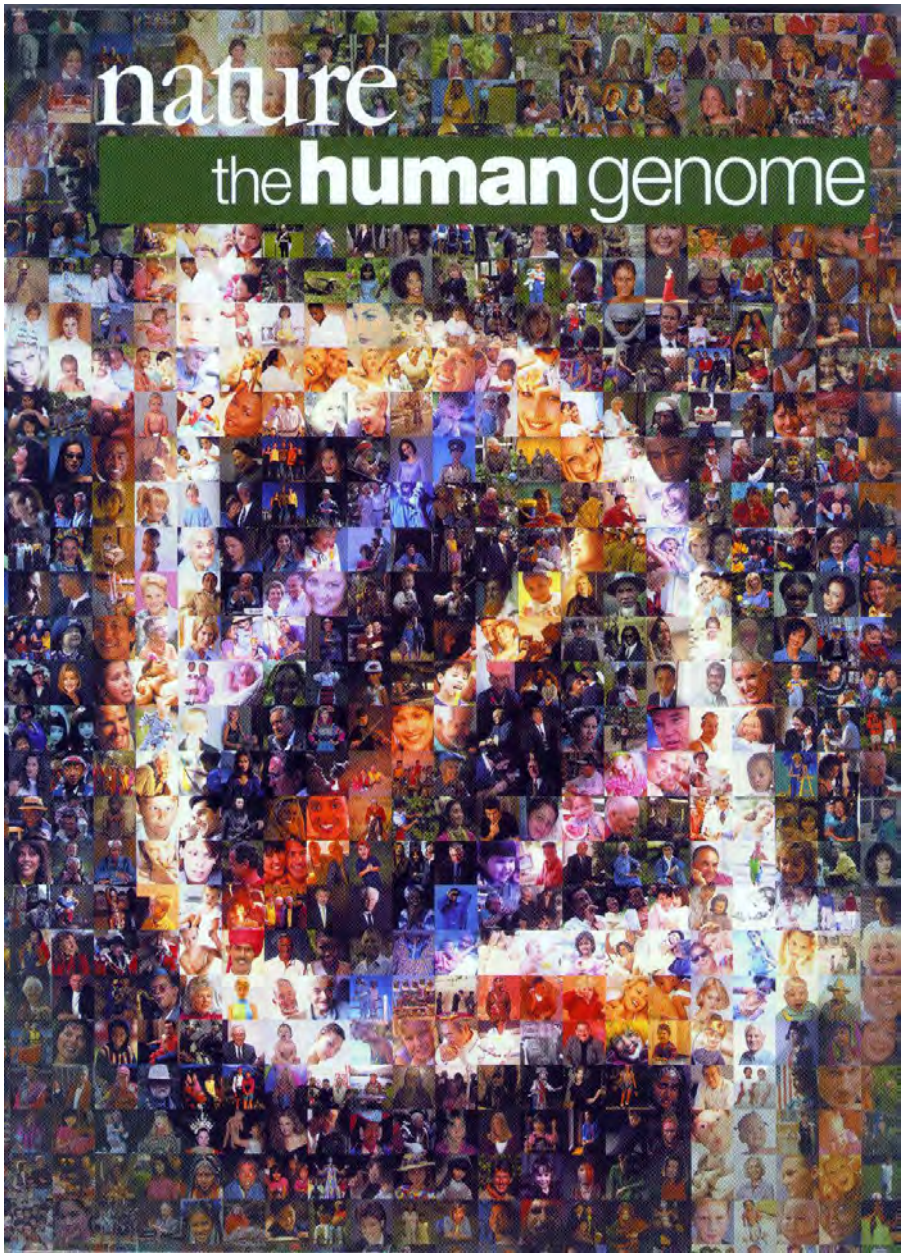
"Today we are learning th guage in which God created President Clinton said at a House ceremony attended by bers of the two teams and, via lite, Prime Minister Tony Bl England. [Excerpts, Page D8.]

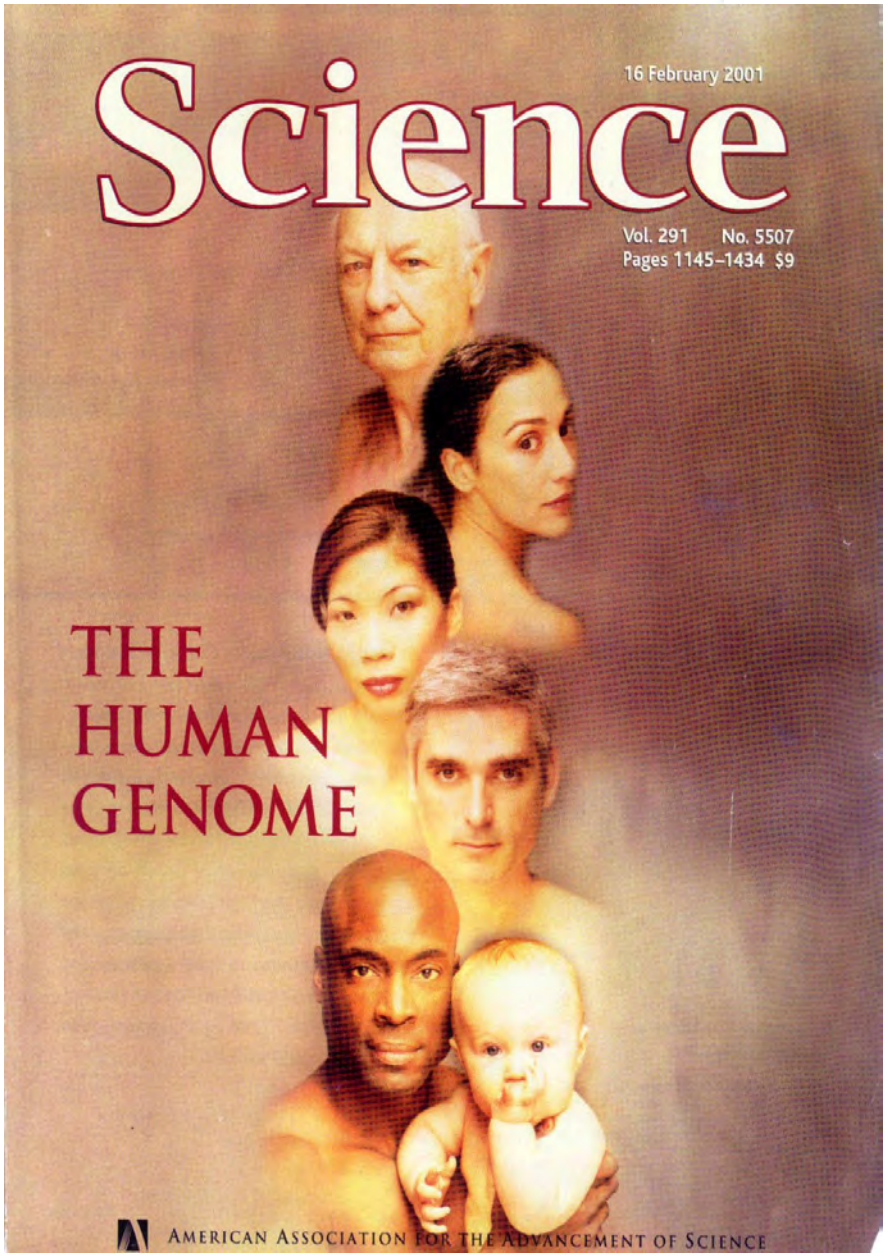
The teams' leaders, Dr. J. Venter, president of Celera G mics, and Dr. Francis S. Collin rector of the National Human nome Research Institute, pr each other's contributions and naled a spirit of cooperation now on, even though the two eff will remain firmly independent.

The human genome, the anc script that has now been deciphe consists of two sets of 23 giant molecules, or chromosomes, each set — one inherited from parent — containing more than billion chemical units.

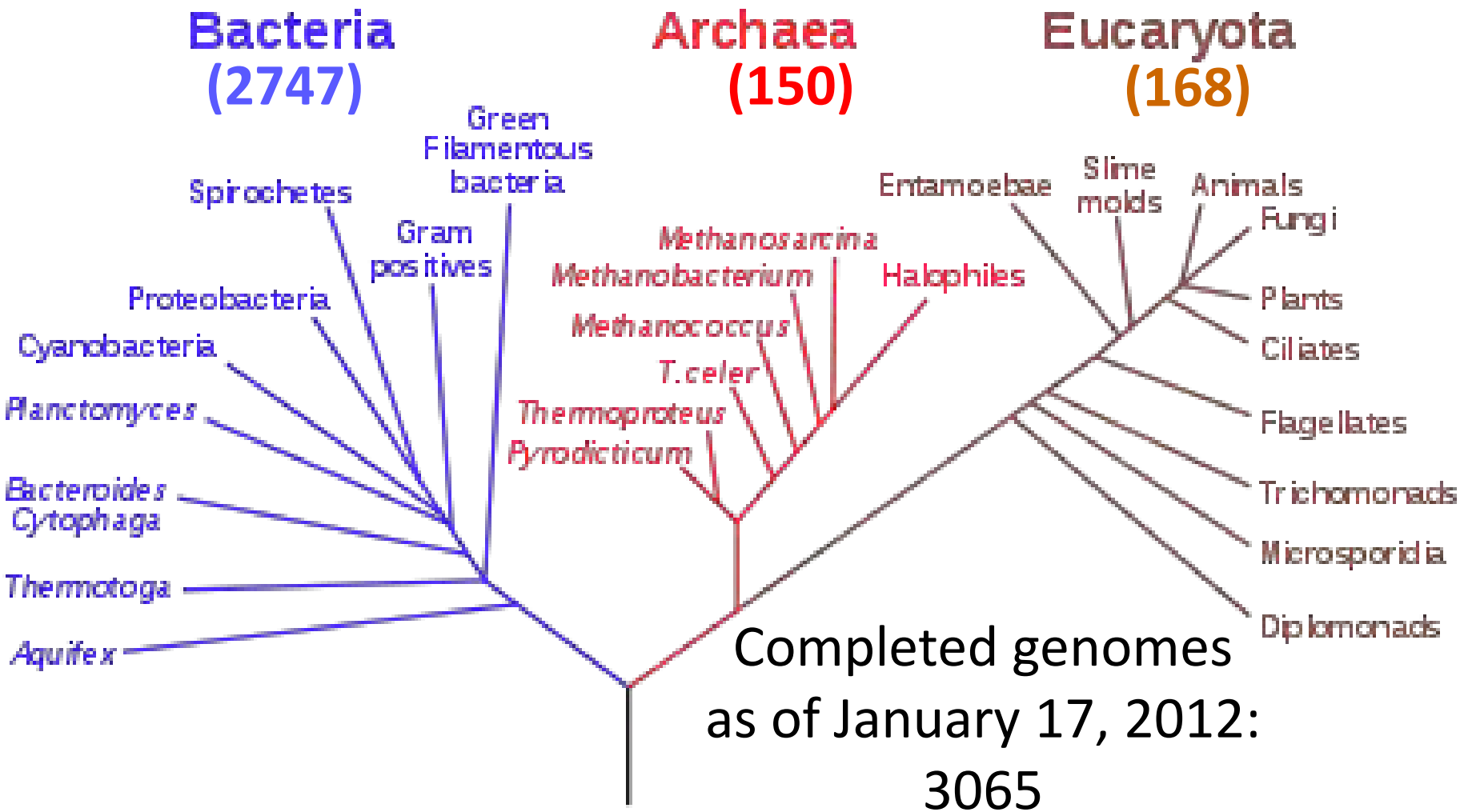
The successful deciphering of vast genetic archive attests to extraordinary pace of biology's

The Human Genome in Print February 2001





Whole Genome Sequences : Progress Report

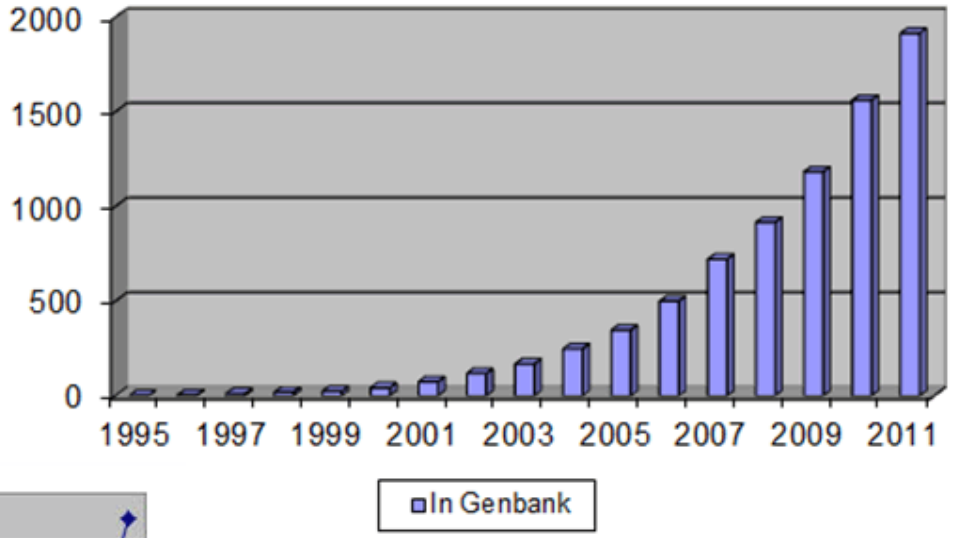


Completed genomes as of January 17, 2012: 3065

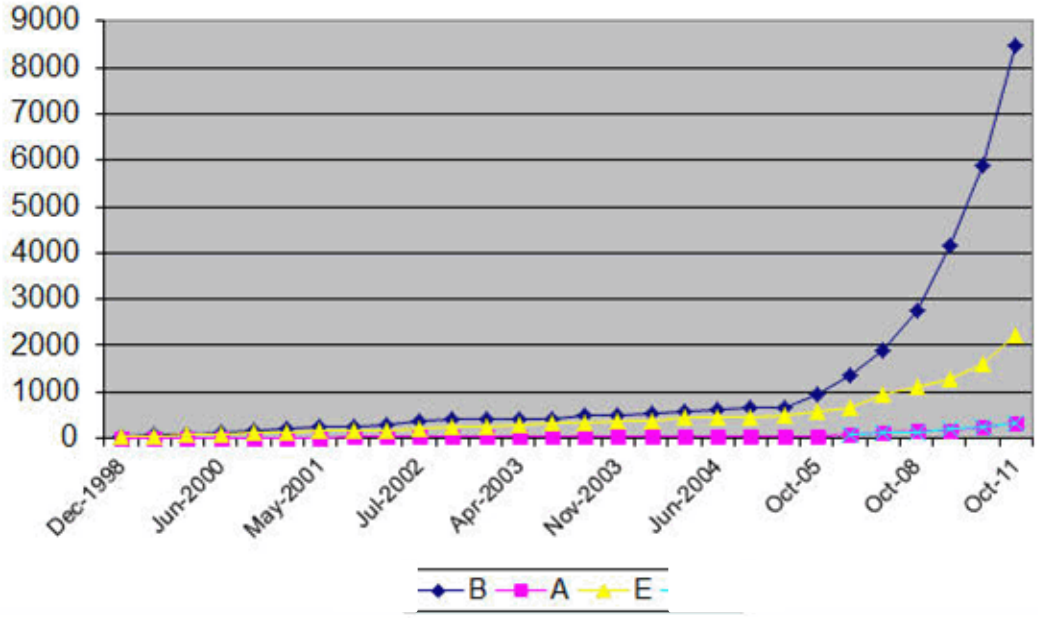
www.genomesonline.org

Genomes Online Statistics

www.genomesonline.org



Completely Sequenced Genomes



Genome Projects according to Phylogenetic Groups

Human Genome Project

What have we learned about the human genome itself over the past decade?

How has the human sequence propelled our understanding of human biology?

What is the road ahead?

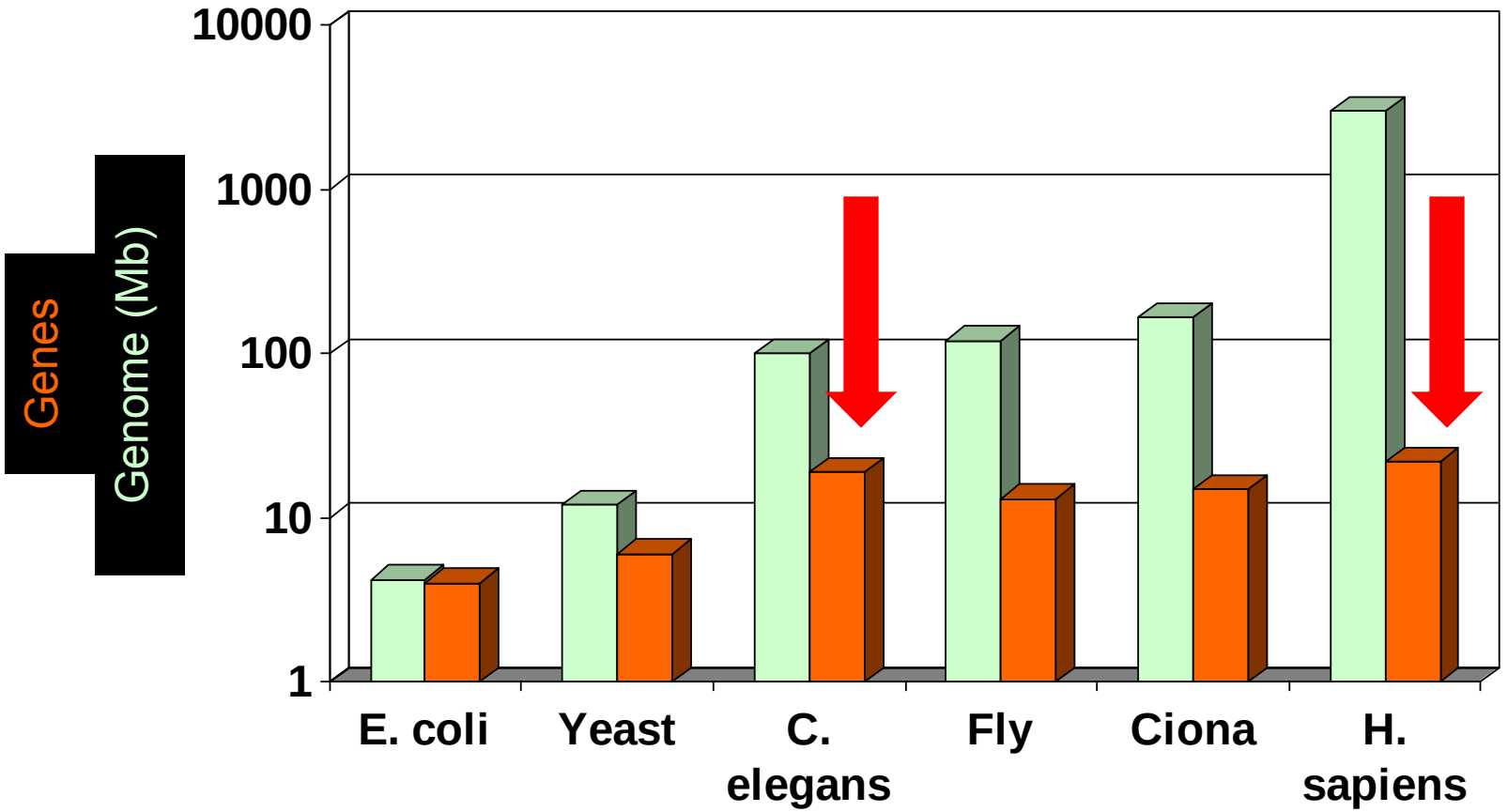
Understanding the structure of genomes

- In the early 2000 our knowledge of the contents of the human genome was surprisingly limited:

- Today the human genome is known to contain between 21,000 and 23,000 protein coding genes

- The human genome is more complex than imagined, regulatory elements outweigh coding regions

Gene Number Independent of Genome Size



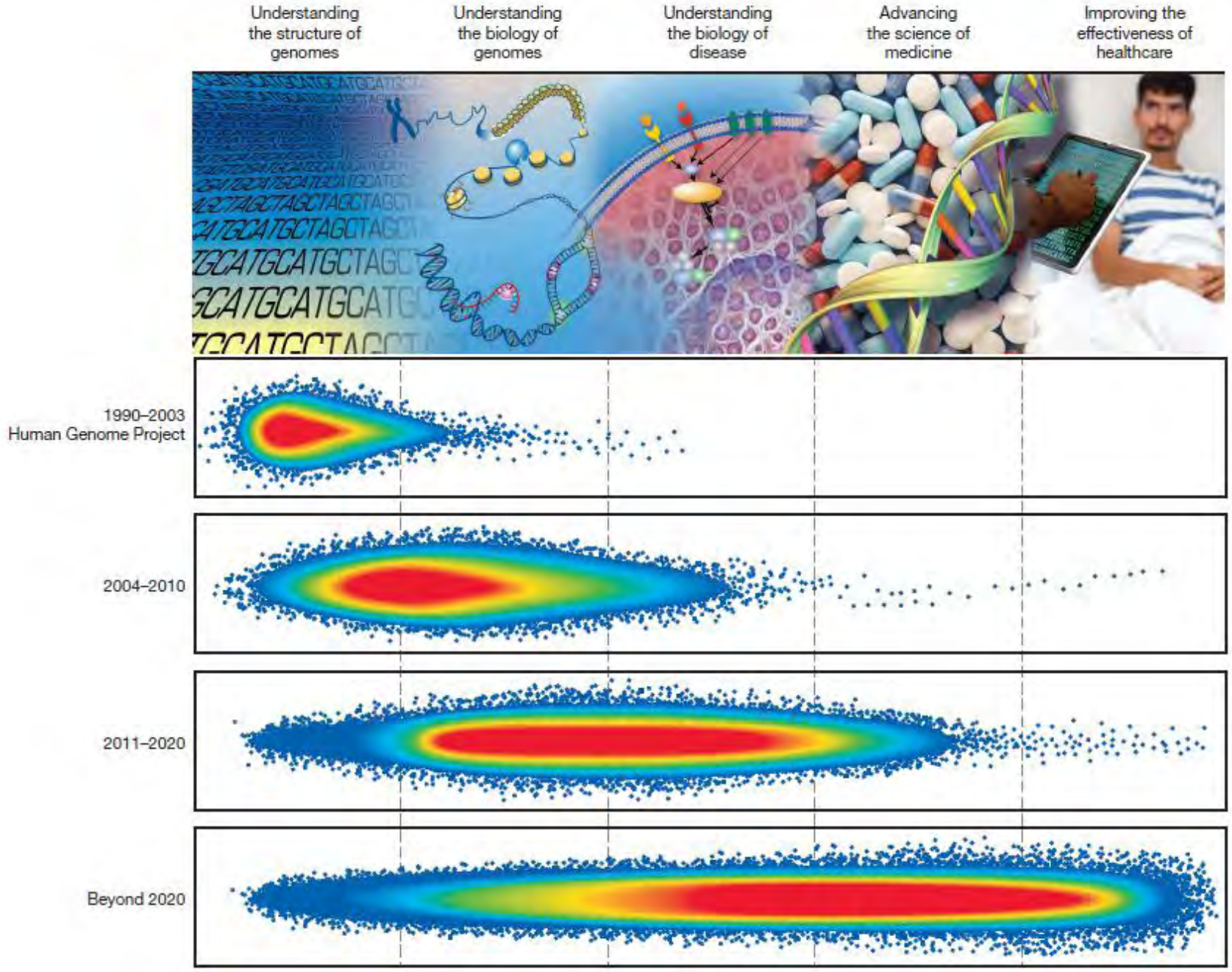
Worm vs. Human

Caenorhabditis elegans
- 20,470 protein-coding genes
~ 972 cells

Homo Sapiens (Ella Sapiens)
~ 22,000 protein-coding genes
~ 100 trillion cells

Generation of Complexity

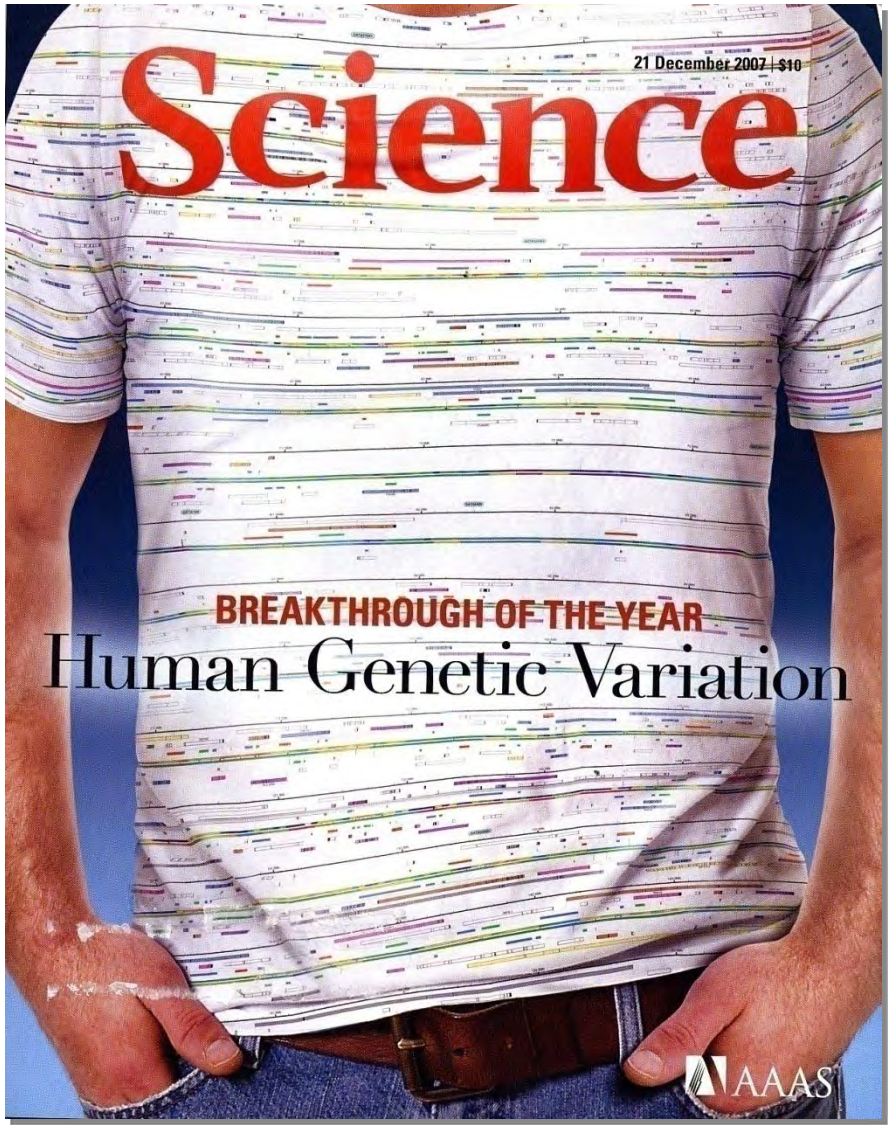
- Only modest increase in gene number
- More sophisticated regulation of gene expression
- More sophisticated post- transcriptional regulation


Understanding the structure of genomes
Understanding the biology of genomes
Understanding the biology of disease
Advancing the science of medicine
Improving the effectiveness of healthcare
1990–2003 Human Genome Project
2004–2010
2011–2020
Beyond 2020

“Variation is the Spice of Life”

- Since the early 1900’s Sir Archibald E. Garrod recognized the importance of human variation

- In 2007 Science Magazine recognized Human Genetic Variation as the Breakthrough of the year



Scales of DNA

- 1 base pair (bp): SNP

- 1,000 bp: size of a typical protein-coding region

- 100,000 bp (100 kb): size of a typical gene

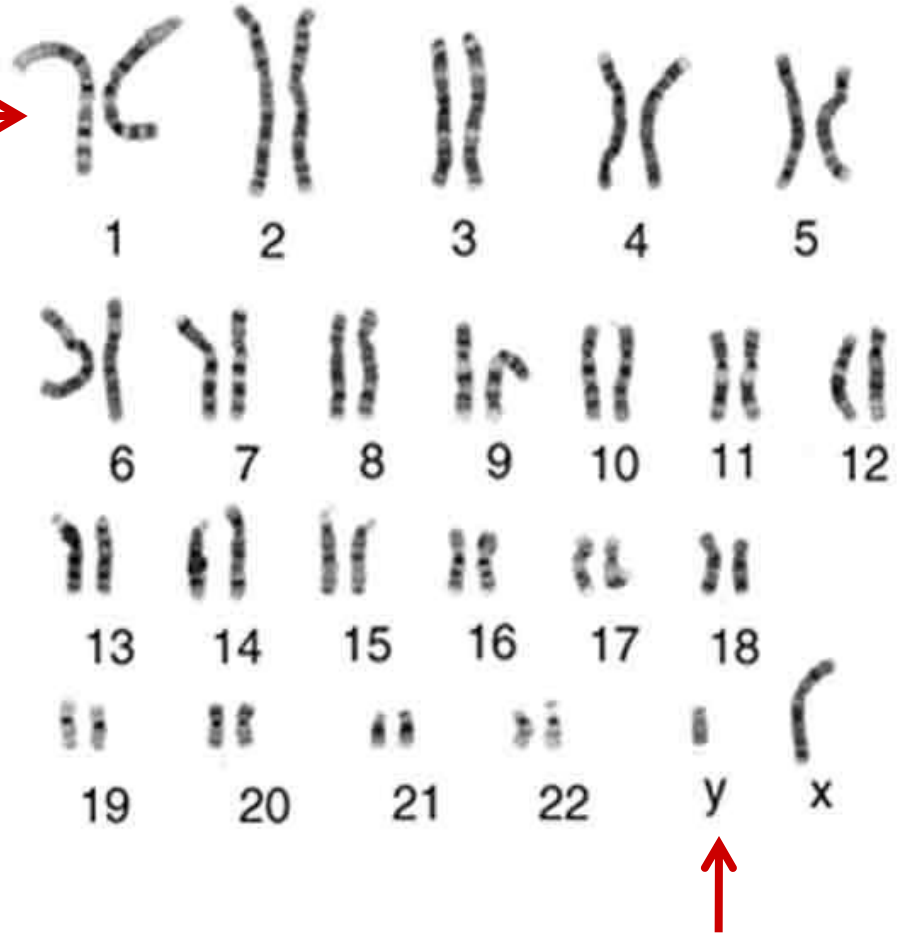
- 3 million bp (3 Mb): minimum size visible with conventional cytogenetics

- 150 Mb: size of average chromosome

- 3,000 Mb (3 Gb): Size of the human genome

Scales of DNA

Chromosome 1:
246 Mb
(2968 genes)



Human genome:
3 billion base
pairs

Chromosome Y: 50 Mb (231 genes)

Genetic Variation

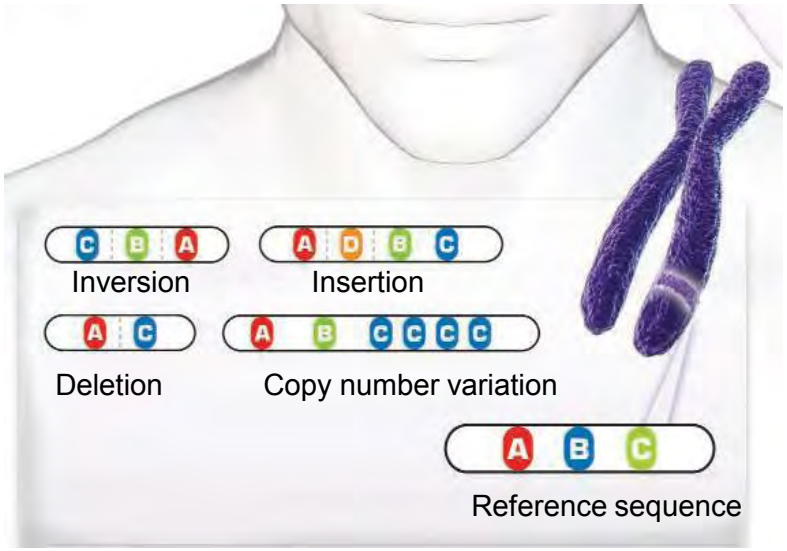
• Single nucleotide polymorphisms (SNPs)	1 bp
• Insertion/deletions (Indels)	few bp
• Short tandem repeats (STRs)	few bp
• Copy number variation (CNVs)	1 to 100s kb
• Cytogenetic deletions/insertions	>3 Mb
• Aneuploidy	>100 Mb

Copy number variants (CNVs)

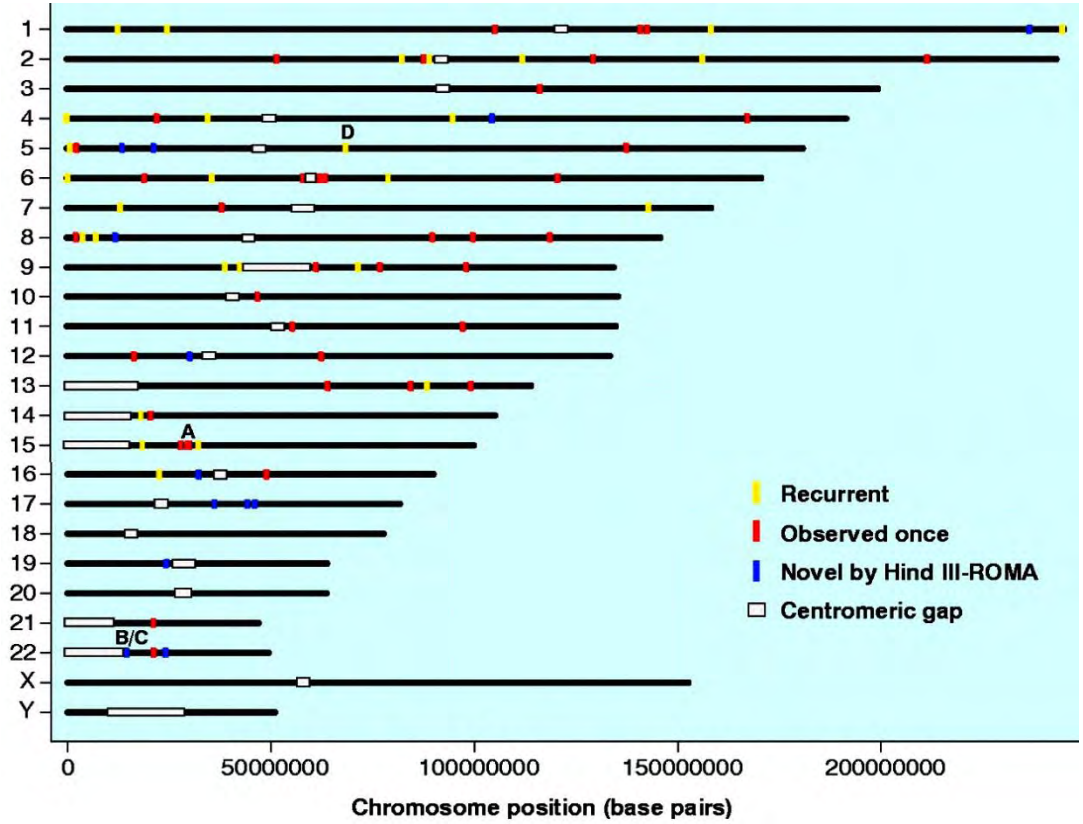
- DNA segment >1kb with a variable copy number compared to the reference genome
- Microdeletions and microduplications
- Currently limited information regarding clinical significance:

  Pathological?

  Benign?



What is the extent of CNV between the genomes of normal individuals?



Large-Scale Copy Number Polymorphism in the Human Genome Sebat et al.

Science Vol 305, July 23, 2004

Analyzed 20 normal individuals

~ 11 CNP differences between individuals

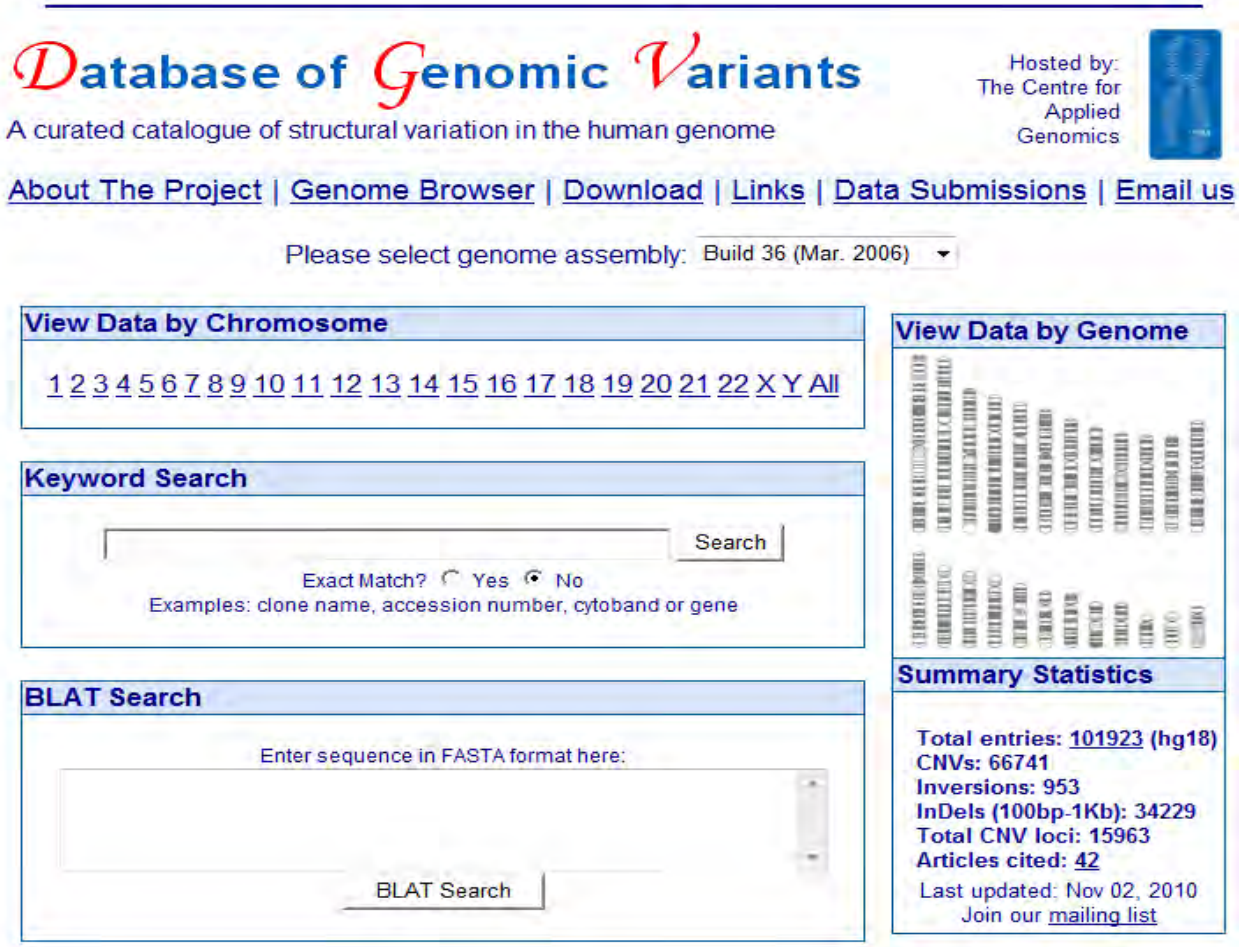
Average length of CNP was 465 kb

About 50% of identified CNP were recurrent in multiple individuals

CNP frequently located near regions responsible for neurodevelopmental disorders

Copy number variation

The ultimate goal is to catalog all the CNV's that can be examined for associations with phenotypes and interpreted in the clinical setting



http://projects.tcag.ca/variation/

The road ahead

- Affordable sequencing
- Understanding all the functional elements of the human genome
- Improve diagnosis of unexplained congenital disorders and identify therapeutic targets for genomic disorders
- Identification of susceptibility loci and functional validation studies of common diseases (polygenic disorders)

Role of Genes in Disease

Online Mendelian Inheritance in Man OMIM ® lists 3,364 phenotype descriptions with known molecular bases (January 17, 2012)

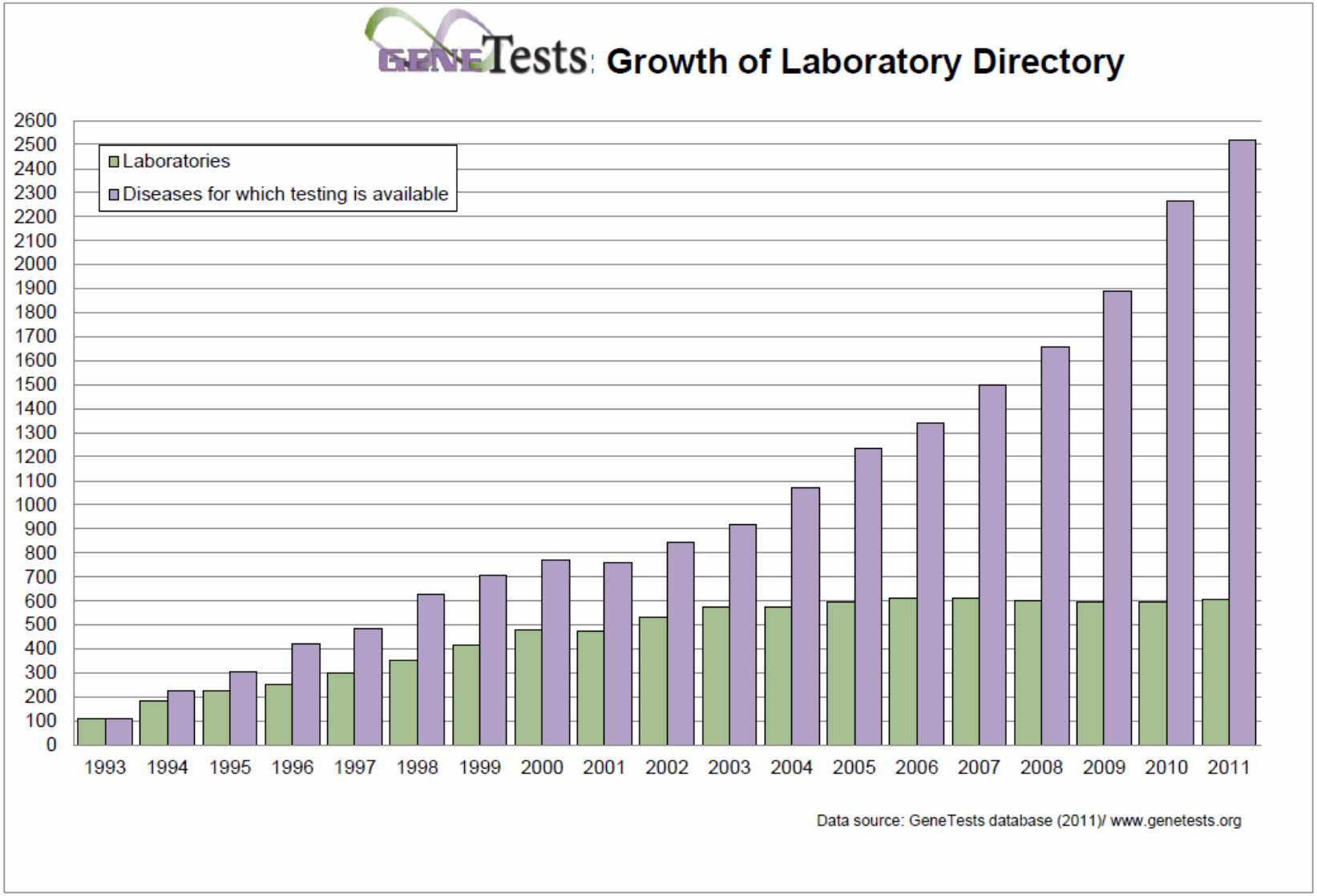
GeneTests

Gene Tests lists 2,528 diseases with molecular tests

2,273 clinically available

(January 17, 2012)

Role of Genes in Disease



Benefits of understanding Mendelian diseases?

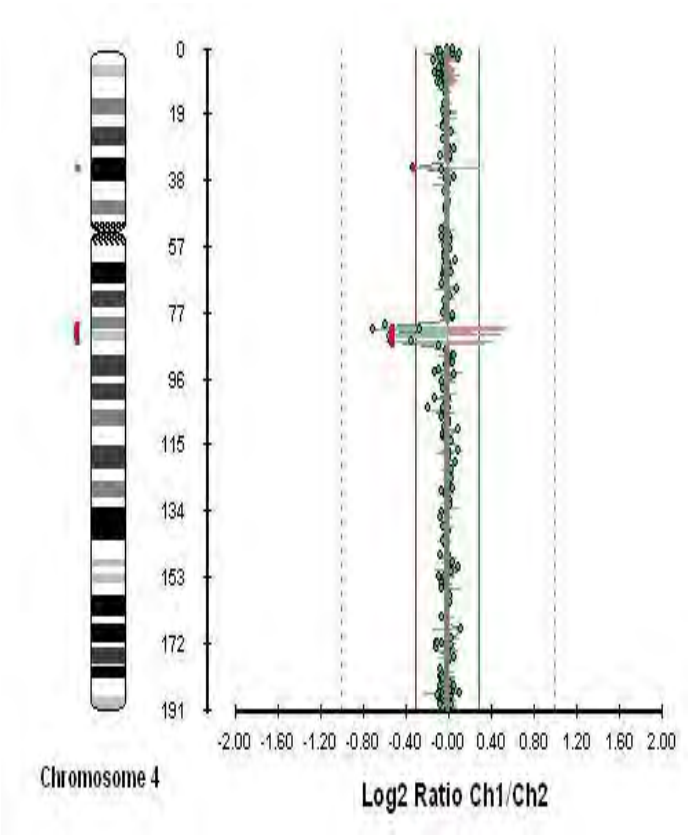
- Elucidating gene function and regulation
- Recognizing normal and pathological pathways
- Development of effective diagnostic tools
- Development of therapeutic targets
- Contribute to understanding of the molecular genetic basis of common complex diseases

Traditional methods of finding Mendelian disease causative genes

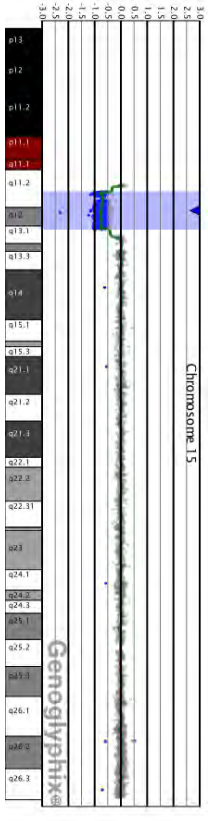
- Linkage analysis
- Cytogenetics
- Compelling biology - candidate genes
- Animal models with similar phenotypes

New strategies of finding Mendelian disease causative genes

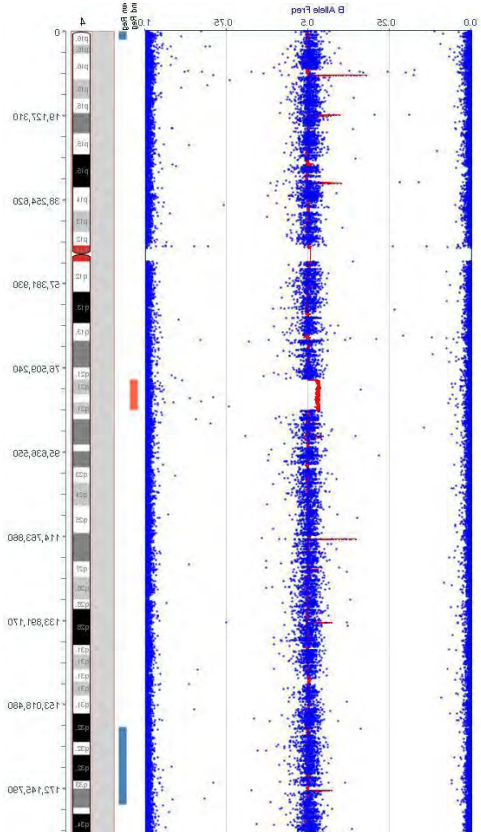
Molecular Karyotyping



1. BAC array
4,200 BAC Clones



2. Oligonucleotide array
135,000 Oligonucleotiodes

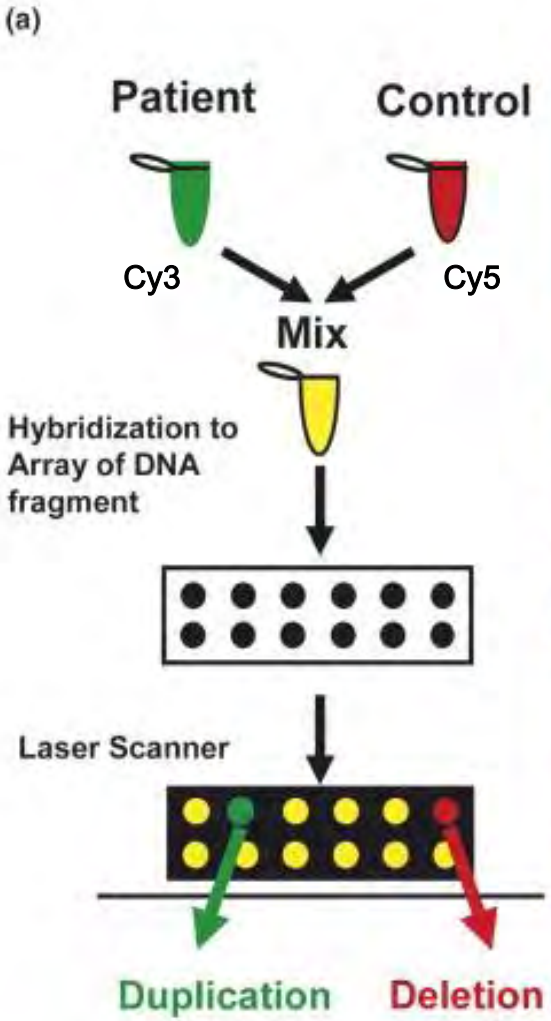


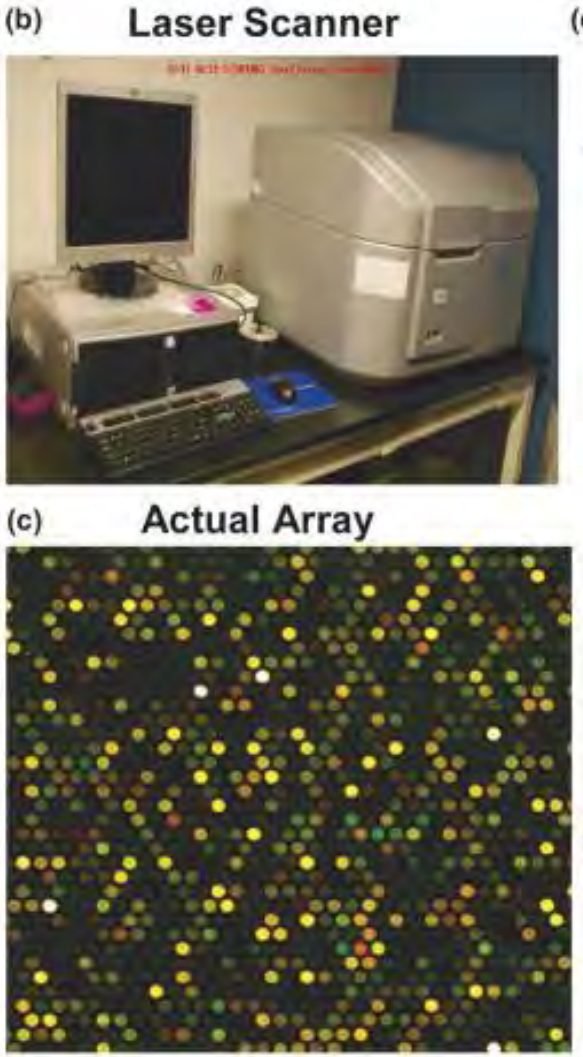
3. SNP array
610,000 SNPs

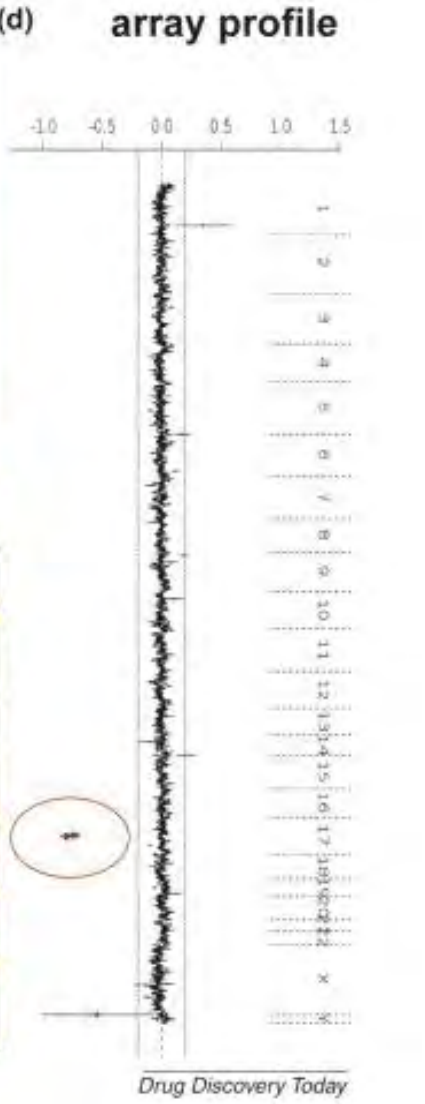
Array Comparative Genomic Hybridization (aCGH)

- AKA chromosome microarray
- Developed as a research tool in cancer cytogenetics
- Revolutionized the diagnostic work-up of patients and facilitated identification of the molecular bases of many genetic disorders

aCGH


(a)
Patient
Control
Cy3
Cy5
Mix
Hybridization to Array of DNA fragment
Laser Scanner
Duplication
Deletion


(b)
Laser Scanner
(c)
Actual Array

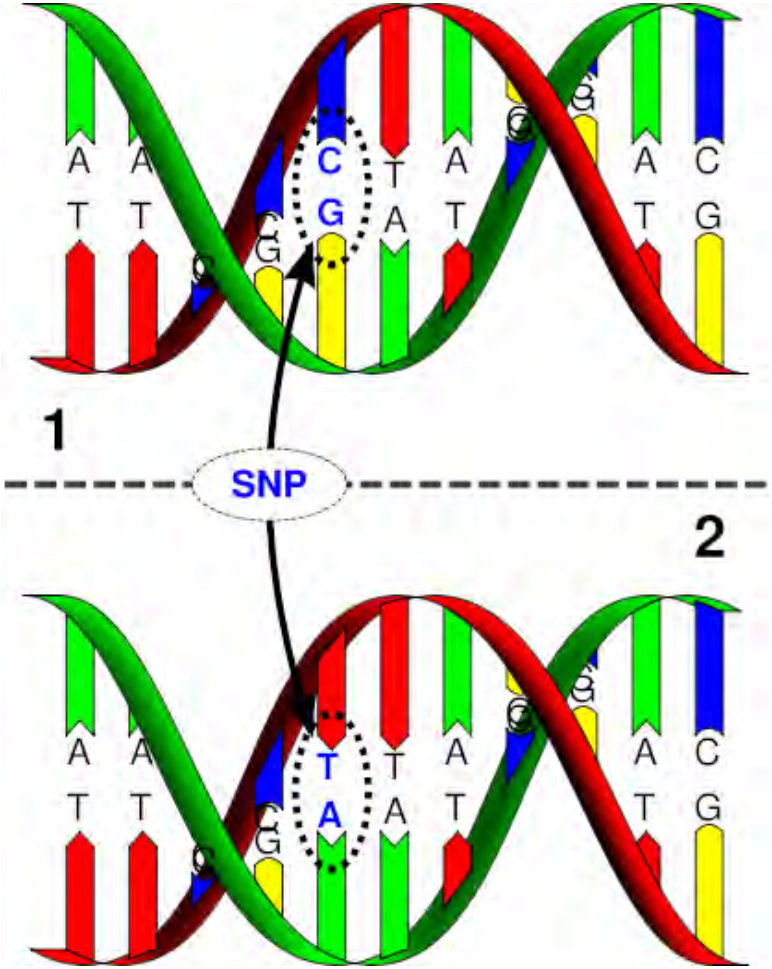

(d)
array profile
-1.0
-0.5
0.0
0.5
1.0
1.5
Drug Discovery Today

Pros and Cons of Arrays

- Pros
  - High detection rate 5-20%
  - Better mapping of aberrations
  - No need for cell culture
  - Shorter result time
  - Reveals unsuspected genomic imbalances
  - Detects genomic duplications
- Cons
  - Not able to identify balanced chromosomal rearrangements or inversions
  - Detection of imbalances of unclear significance
  - Costly

SNP Array

- What is a SNP?
  - A SNP is defined as a single base change in a DNA sequence that occurs in a significant proportion (more than 1 percent) of a large population



SNP Array

- SNP arrays are comprised of oligonucleotides that correspond to SNPs along the human genome
- Used to: -detect polymorphisms within a population; -find disease susceptibility alleles; -find important variants in pharmacogenomics; -to detect LOH
- Allows the use of DNA sequence variation to identify copy number changes
- Identification of long stretches of homozygosity (consanguinity) and copy neutral genetic abnormalities like uniparental disomy

New strategies of finding Mendelian disease causative genes

- Next-generation sequencing:
  - Whole genome sequencing
  - Exome sequencing
- Considerations:
  - Exons (180,000) are about 1% of human genome or 30 Mb
  - About 85% of the disease causing mutations are found in coding regions or canonical splice sites
  - Difficult to interpret the functional consequences of variations in non-coding regions

WGS vs ES

Whole Genome Sequencing

Determining the sequence of the entire human genome

Exome Sequencing

Determining the sequence of the coding regions (exome) of our genome

Whole Genome and Exome Sequencing

- By sequencing the entire genome or exome of affected individuals different groups have identified the molecular bases of Mendelian disorders

Whole Genome and Exome Sequencing

- WGS and ES will identify a large number of sequence variants (Raw data is HUGE!)
- Using additional methods like linkage and bioinformatics, variants are prioritized and the clinically significant (causal) variant is identified
- Recent AR and AD Mendelian disorder causal variants identified using ES and WGS data include: Miller syndrome, Metachondromatosis, Kabuki Syndrome, Schinzel-Gideon Syndrome and the list is growing!

Exome/Genome Sequencing

Exome sequencing identifies *MLL2* mutations as a cause of Kabuki syndrome

Sarah B. Ng[1,*], Abigail W. Bigham[2,*], Kati J. Buckingham[2], Mark C. Hannibal[2,3], Margaret McMillin[2], Heidi Gildersleeve[2], Anita E. Beck[2,3], Holly K. Tabor[2,3], Greg M. Cooper[1], Heather C. Mefford[2], Choli Lee[1], Emily H. Turner[1], Josh D. Smith[1], Mark J. Rieder[1], Koh-ichiro Yoshiura[4], Naomichi Matsumoto[5], Tohru Ohta[6], Norio Niikawa[6], Deborah A. Nickerson[1], Michael J. Bamshad[1,2,3,†], and Jay Shendure[1,†]

[1]Department of Genome Sciences, University of Washington, Seattle, Washington, USA

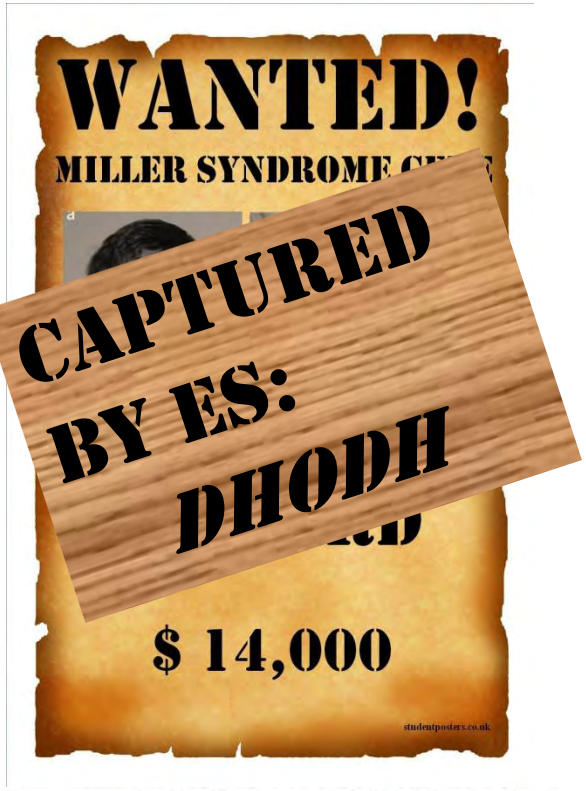
[2]Department of Pediatrics, University of Washington, Seattle, Washington, USA

[3]Seattle Children's Hospital, Seattle, Washington, USA

[4]Department of Human Genetics, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan

[5]Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Japan

[6]Research Institute of Personalized Health Sciences, Health Sciences University of Hokkaido, Hokkaido, Japan



Exome sequencing identifies the cause of a Mendelian disorder

Sarah B. Ng[1,*], Kati J. Buckingham[2,*], Choli Lee[1], Abigail W. Bigham[2], Holly K. Tabor[2], Karin M. Dent[3], Chad D. Huff[4], Paul T. Shannon[5], Ethylin Wang Jabs[6,7], Deborah A. Nickerson[1], Jay Shendure[1,†], and Michael J. Bamshad[1,2,8,†]

[1]Department of Genome Sciences, University of Washington, Seattle, Washington, USA [2]Department of Pediatrics, University of Washington, Seattle, Washington, USA [3]Department of Pediatrics, University of Utah, Salt Lake City, Utah, USA [4]Department of Human Genetics, University of Utah, Salt Lake City, Utah, USA [5]Institute of Systems Biology, Seattle WA, USA [6]Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York, New York, USA [7]Department of Pediatrics, Johns Hopkins University, Baltimore, Maryland [8]Seattle Children's Hospital, Seattle, Washington, USA

Whole-Genome Sequencing of a Single Proband Together with Linkage Analysis Identifies a Mendelian Disease Gene

Nara L. M. Sobreira[1,2,9], Elizabeth T. Cirulli[2,9], Dimitrios Avramopoulos[1,4,9], Elizabeth Wohler[3], Gretchen L. Oswald[1], Eric L. Stevens[1,2], Dongliang Ge[3], Kevin V. Shianna[3], Jason P. Smith[3], Jessica M. Maia[2], Curtis E. Gumbs[3], Jonathan Pevsner[6,7], George Thomas[1,5], David Valle[1,8*], Julie E. Hoover-Fong[1,8,9*], David B. Goldstein[9*]

1 McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, United States of America, 2 Predoctoral Training Program in Human Genetics, Johns Hopkins University School of Medicine, Baltimore, Maryland, United States of America, 3 Center for Human Genome Variation, Duke University School of Medicine, Durham, North Carolina, United States of America, 4 Department of Psychiatry, Johns Hopkins University School of Medicine, Baltimore, Maryland, United States of America, 5 Department of Cytogenetics, Kennedy Krieger Institute, Baltimore, Maryland, United States of America, 6 Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, Maryland, United States of America, 7 Department of Neurology, Kennedy Krieger Institute, Baltimore, Maryland, United States of America, 8 Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, Maryland, United States of America, 9 Greenberg Center for Skeletal Dysplasias, Johns Hopkins University School of Medicine, Baltimore, Maryland, United States of America

Tools used with ES/WGS

- At present, the sequencing of each new individual identifies ~402 novel variants introducing changes in the encoded protein
- Prioritize mutations that introduce truncations of the encoded protein
- Prioritize mutations that occur at highly conserved positions
- Consult SNP databases and use mutation prediction software

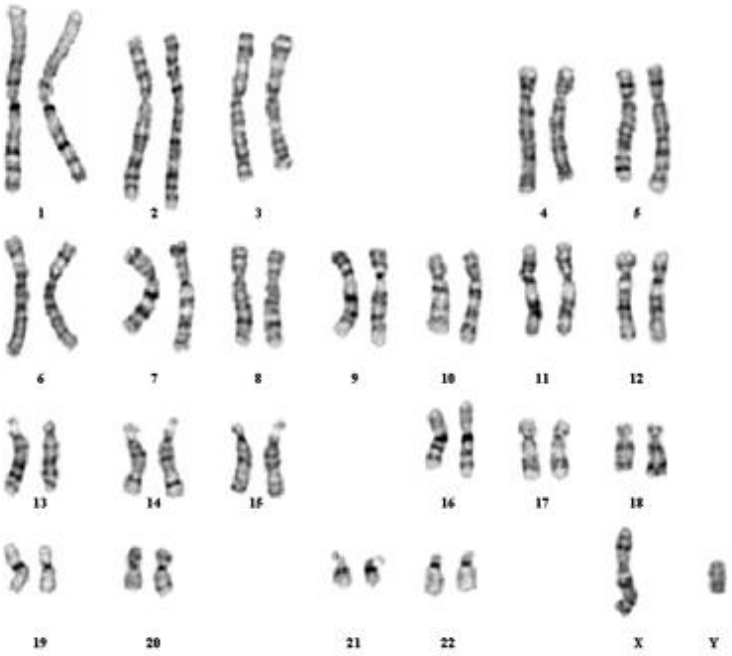
Tools used with ES/WGS

Table 1 Example of tools to analyze variants

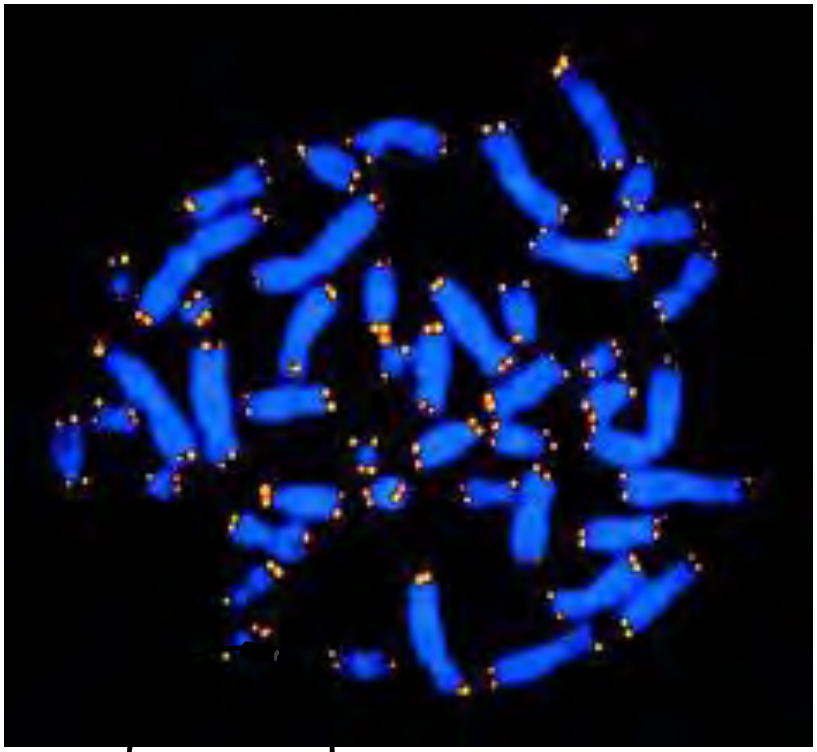
Has the variant been seen in a patient with a particular disease?	
OMIM	http://www.ncbi.nlm.nih.gov/omim
HGMD	http://www.hgmd.cf.ac.uk/ac/index.php
PubMed	http://www.ncbi.nlm.nih.gov/pubmed
Is the variant evolutionarily conserved?	
phastCons	http://compgen.bscb.cornell.edu/phast/index.php
If it is a coding region variant will it result in a premature stop codon, a read-through of a stop codon, affect protein structure or function?	
SIFT	http://sift.jcvi.org
PolyPhen-2	http://genetics.bwh.harvard.edu/pph2
Has the variant been seen in humans before?	
dbSNP	http://www.ncbi.nlm.nih.gov/projects/SNP/
Does the variant affect RNA splicing?	
GeneSplicer	http://www.cbcb.umd.edu/software/GeneSplicer/
In genes with several transcripts, does the variant affect a well documented transcript?	
RefSeq	http://www.ncbi.nlm.nih.gov/RefSeq/
CCDS	http://www.ncbi.nlm.nih.gov/projects/CCDS/CcdsBrowse.cgi
Is there a region of homozygosity in consanguineous families to help find the location of a causal recessive disease?	
BEAGLE	http://faculty.washington.edu/browning/beagle/beagle.html

Whole exome and whole genome sequencing. Bick, D; Dimmock, D. Curr Opin Pediatr 23:594-600.

Costs of genetic testing



Karyotype

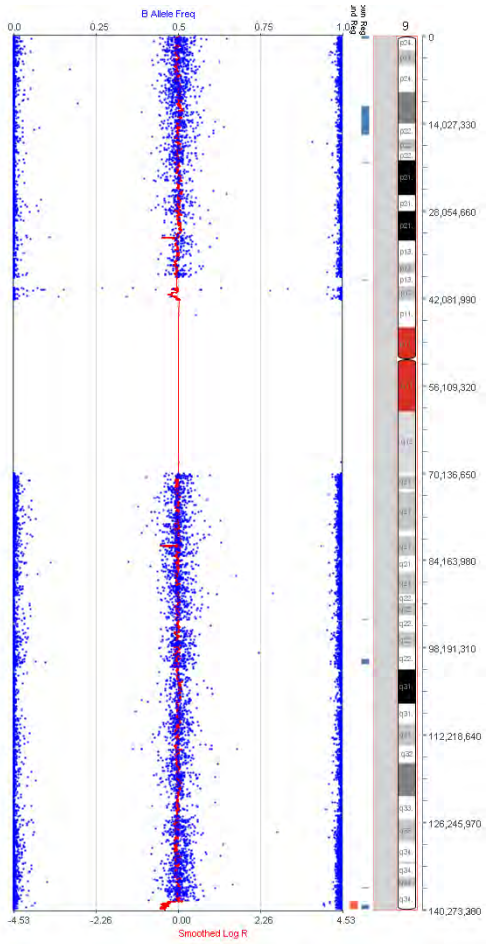


FISH

$ 500-600

$ 260

Costs of genetic testing



$1,000-2,000

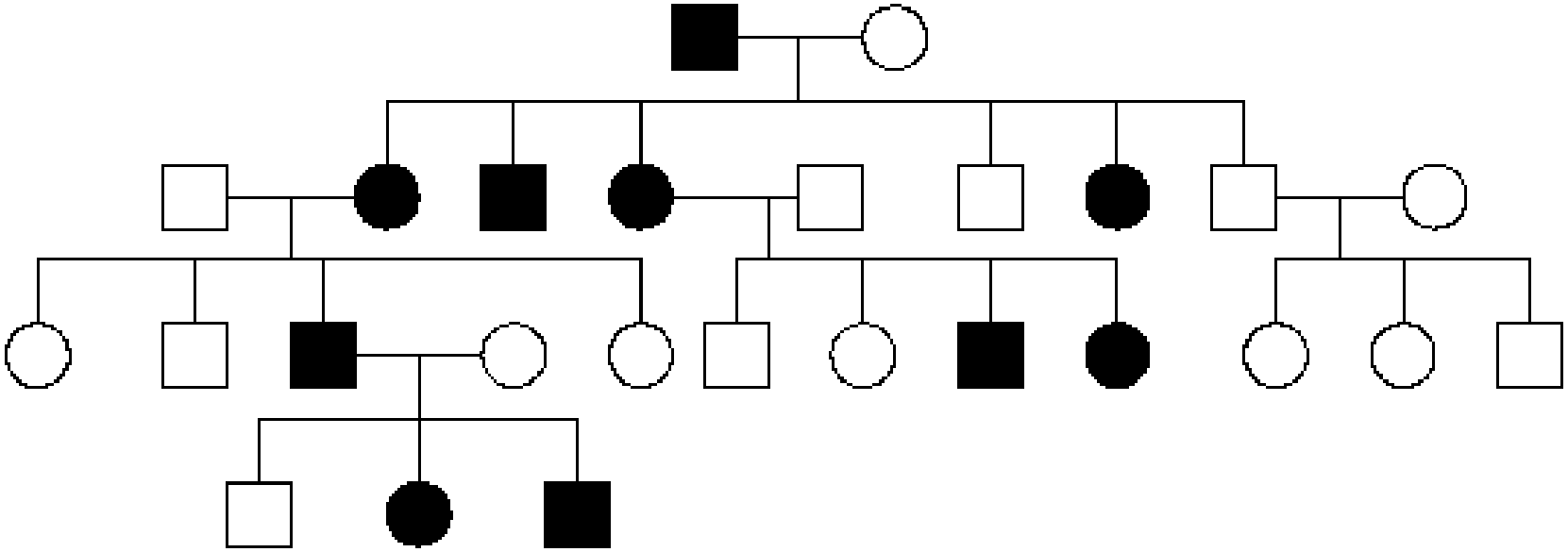
SNP Array

$14,000

Exome Sequencing

Family History

The pedigree is one of the most useful tools in genetic diagnosis



Genotype first diagnosis controversy

•New technologies have increased the number of patients with an established diagnosis

•Recent publica ’ approach to the evaluation o opment by primary care pr

Technology is a powerful tool when used right

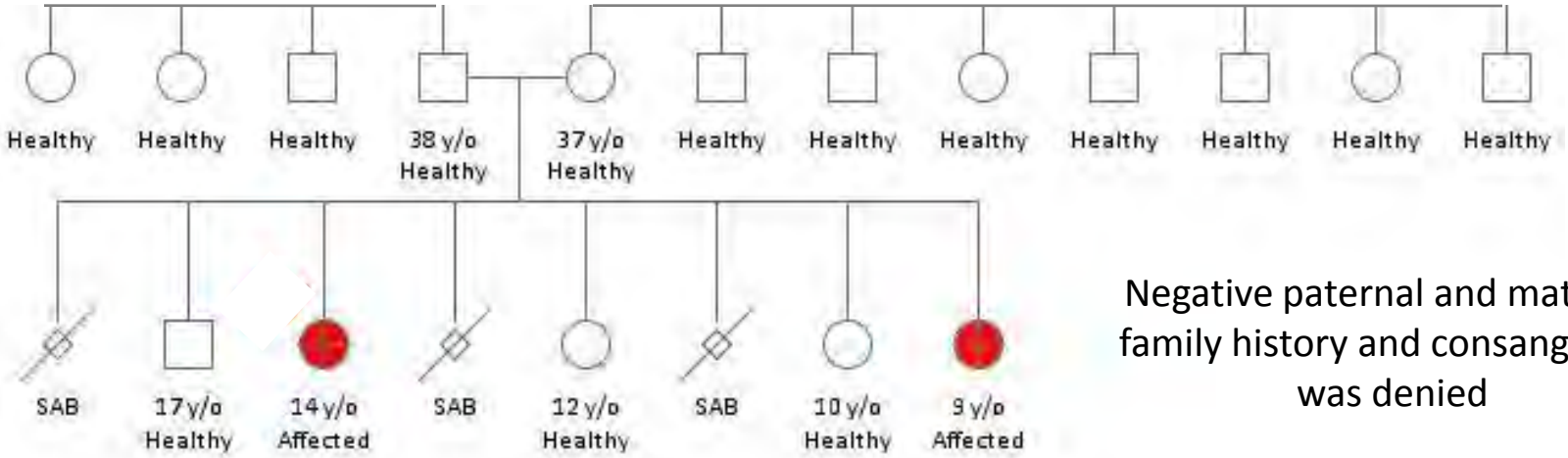
•The yield of ab her after the patient has undergone a thorough PE and FH (YL)

•I feel the family history and PE by an expert clinician should always precede testing in patients with disorders of development

Lacassie, Y. AJMG and Genet Med. 2009 May;11(5):371; author reply 371-2.

Case Report

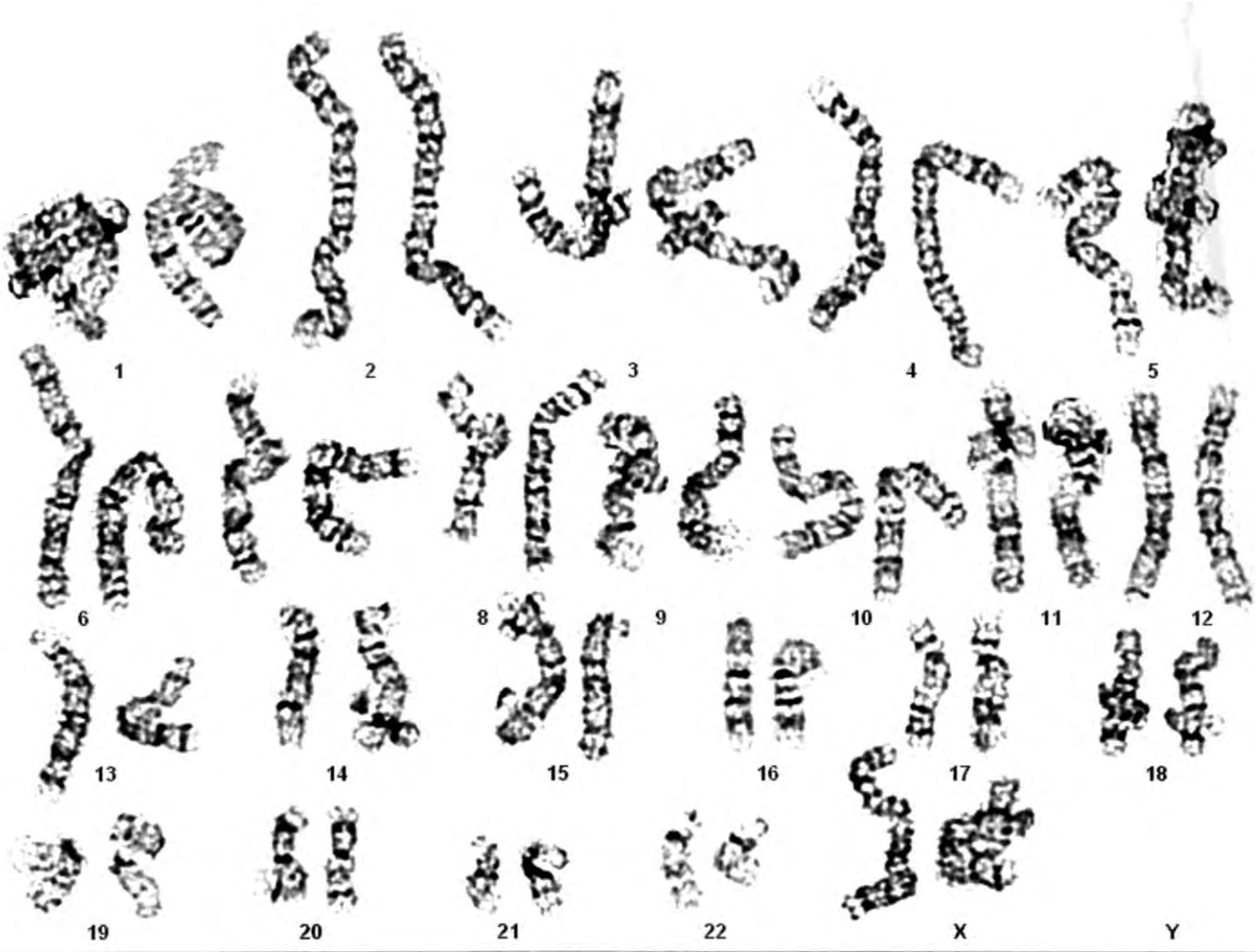
- 14yo and 9yo Hispanic females
- DDM, microcephaly and dysmorphic features
- Seen in 2005



Negative paternal and maternal family history and consanguinity was denied

Case Report

Suspected chromosomal anomaly. Karyotype was normal (2005)



Case Report

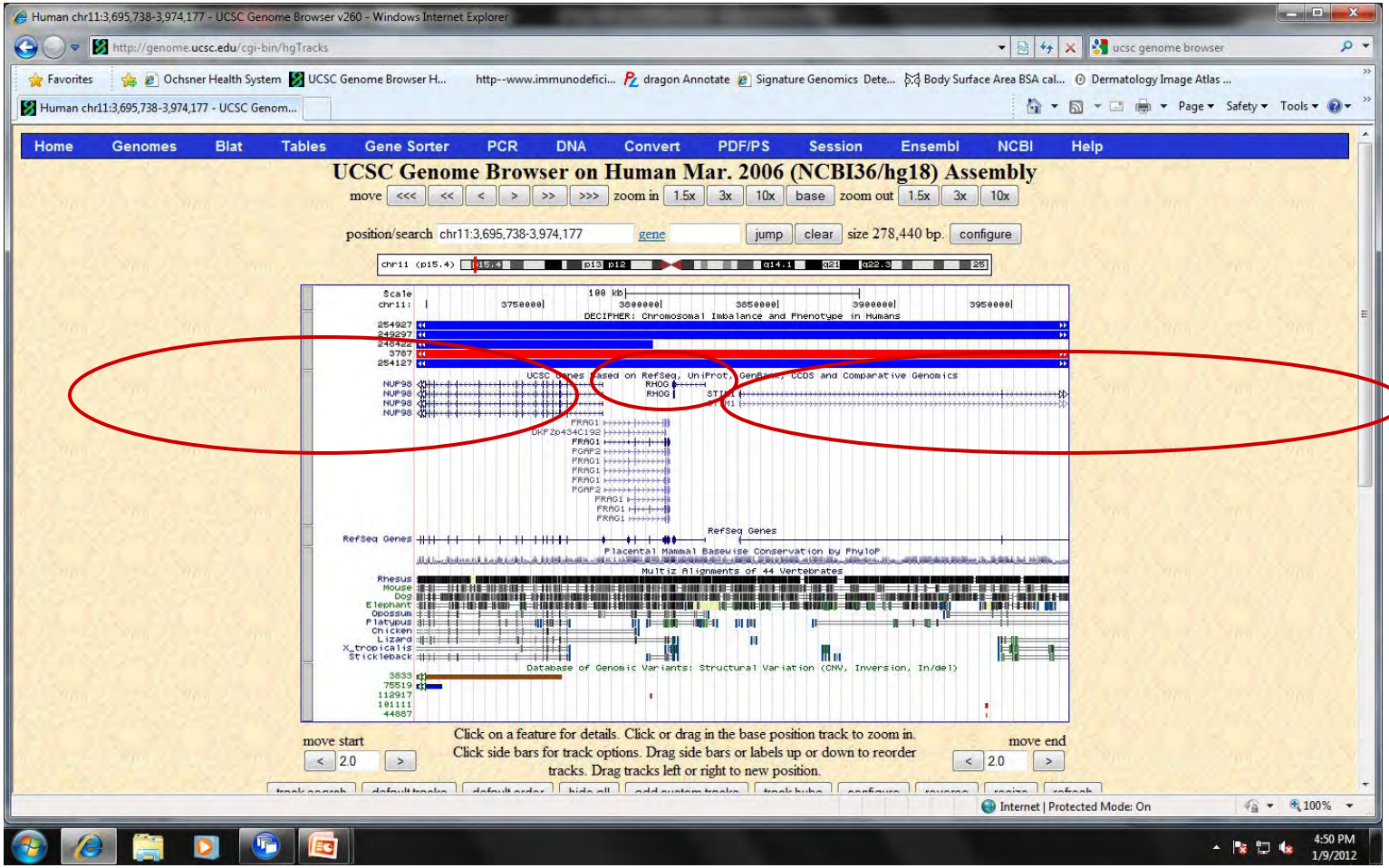
- 5 years later in 2010 returned to clinic and CMA was requested:

dup 11p15.4 (278.4kb)

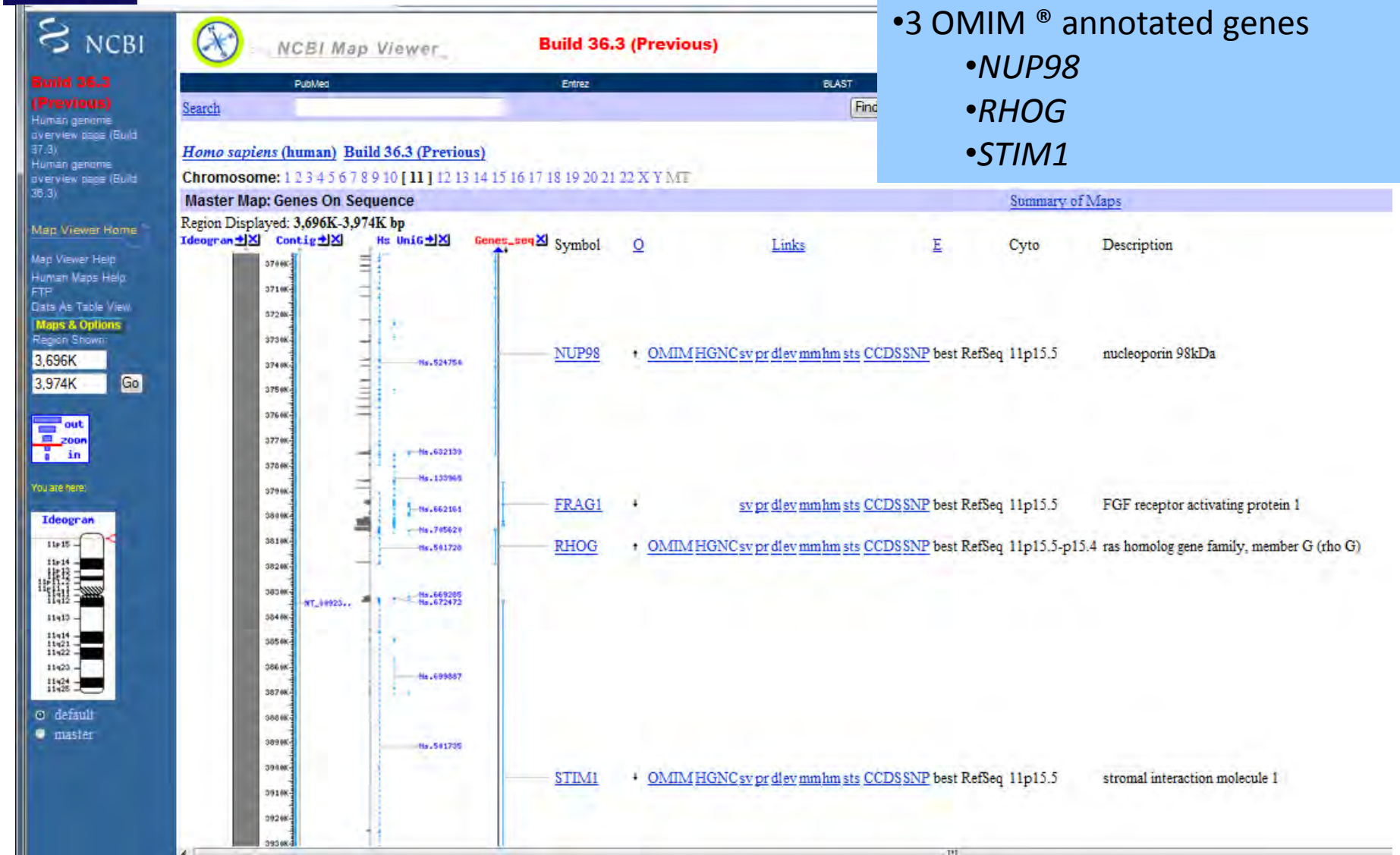
- Sister same abnormality

- Parental testing pending

UCSC Genome Browser



NCBI Map Viewer

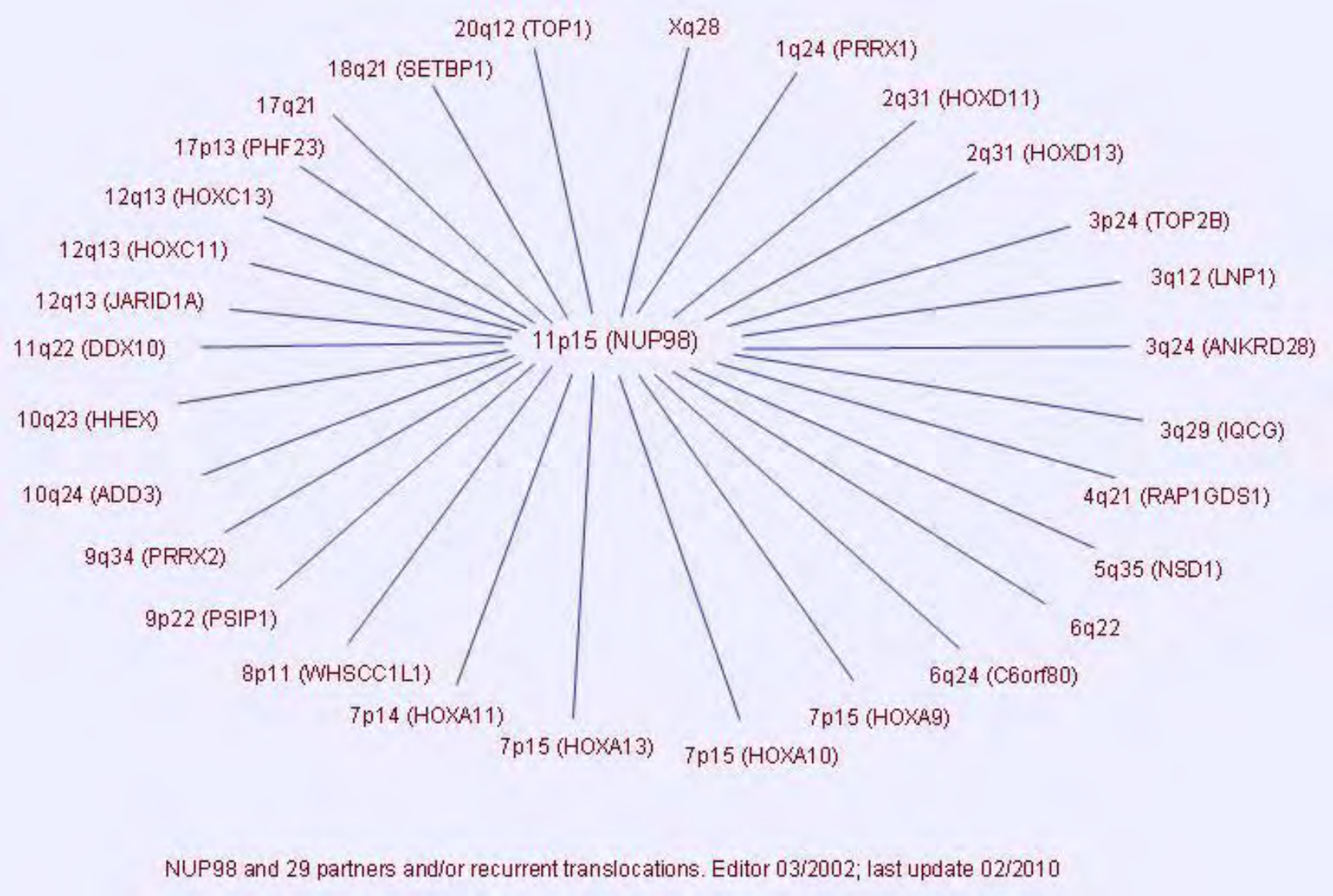


- 3 OMIM ® annotated genes
  - *NUP98*
  - *RHOG*
  - *STIM1*

NUP98

- 920 amino acids; 97 kDa; contains repeated motifs (GLFG and FG) in N-term and a RNA binding motif in C-term
- Mediates nucleo-cytoplasmic transport of protein and RNA
- Chromosomal translocations involving *NUP98* have been identified in patients with MDS, T-ALL, CML and AML

NUP98


20q12 (TOP1)
Xq28
1q24 (PRRX1)
18q21 (SETBP1)
2q31 (HOXD11)
17q21
2q31 (HOXD13)
17p13 (PHF23)
3p24 (TOP2B)
12q13 (HOXC13)
3q12 (LNP1)
12q13 (HOXC11)
12q13 (JARID1A)
11p15 (NUP98)
3q24 (ANKRD28)
11q22 (DDX10)
3q29 (IQCG)
10q23 (HHEX)
4q21 (RAP1GDS1)
10q24 (ADD3)
5q35 (NSD1)
9q34 (PRRX2)
6q22
9p22 (PSIP1)
6q24 (C6orf80)
8p11 (WHSCC1L1)
7p15 (HOXA9)
7p14 (HOXA11)
7p15 (HOXA10)
7p15 (HOXA13)
NUP98 and 29 partners and/or recurrent translocations. Editor 03/2002; last update 02/2010

NUP98

- Unclear what are the implications of duplications of *NUP98*

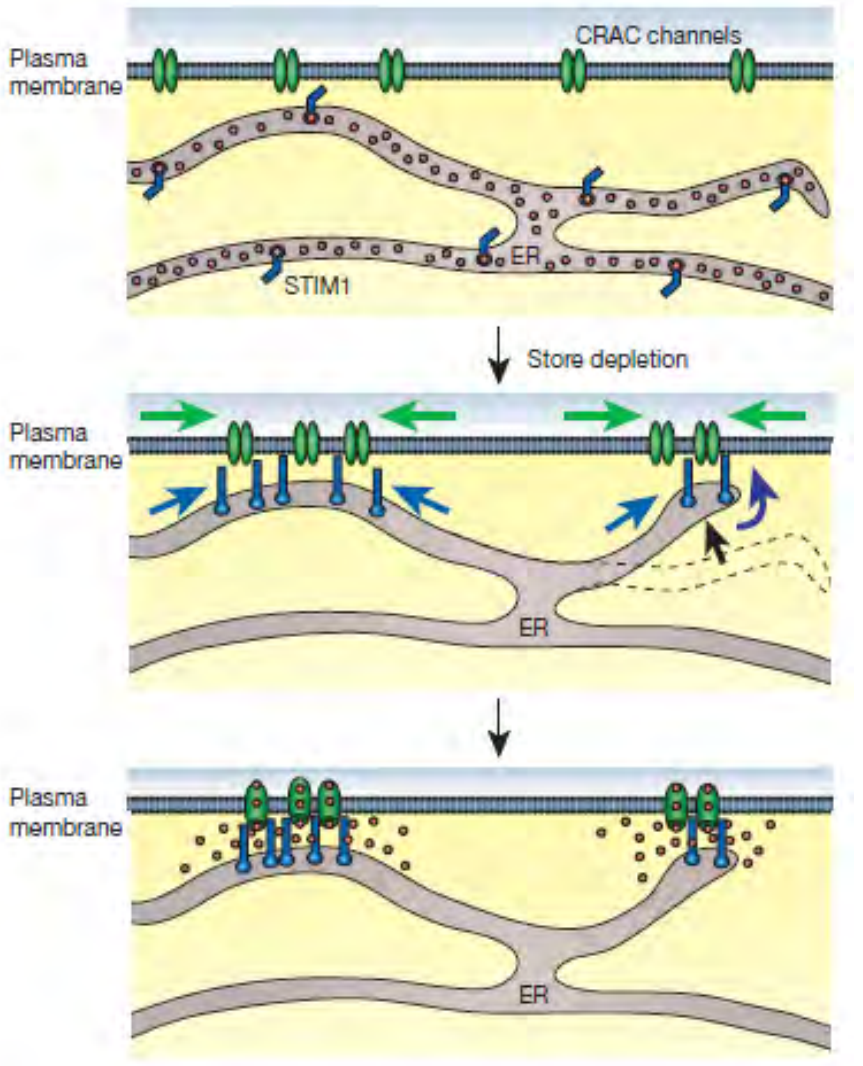
- Dual haploinsufficiency of *Nup98* and *Rae1* has been shown to result in premature separation of sister chromatids, leading to severe aneuploidy

RHOG

- Member of the RAS family of supergenes
- Encodes a GTP binding protein that acts in the pathway of signal transduction and plays a role in the regulation of cellular functions
- Unclear what are the implications of duplications of *RHOG*

STIM1

- 746 amino acids, 90 % sequence identity to mouse Stim1, conserved from Drosophila to human
- Encodes a calcium sensor that conveys calcium load of the ER to store operated calcium channels at the plasma membrane

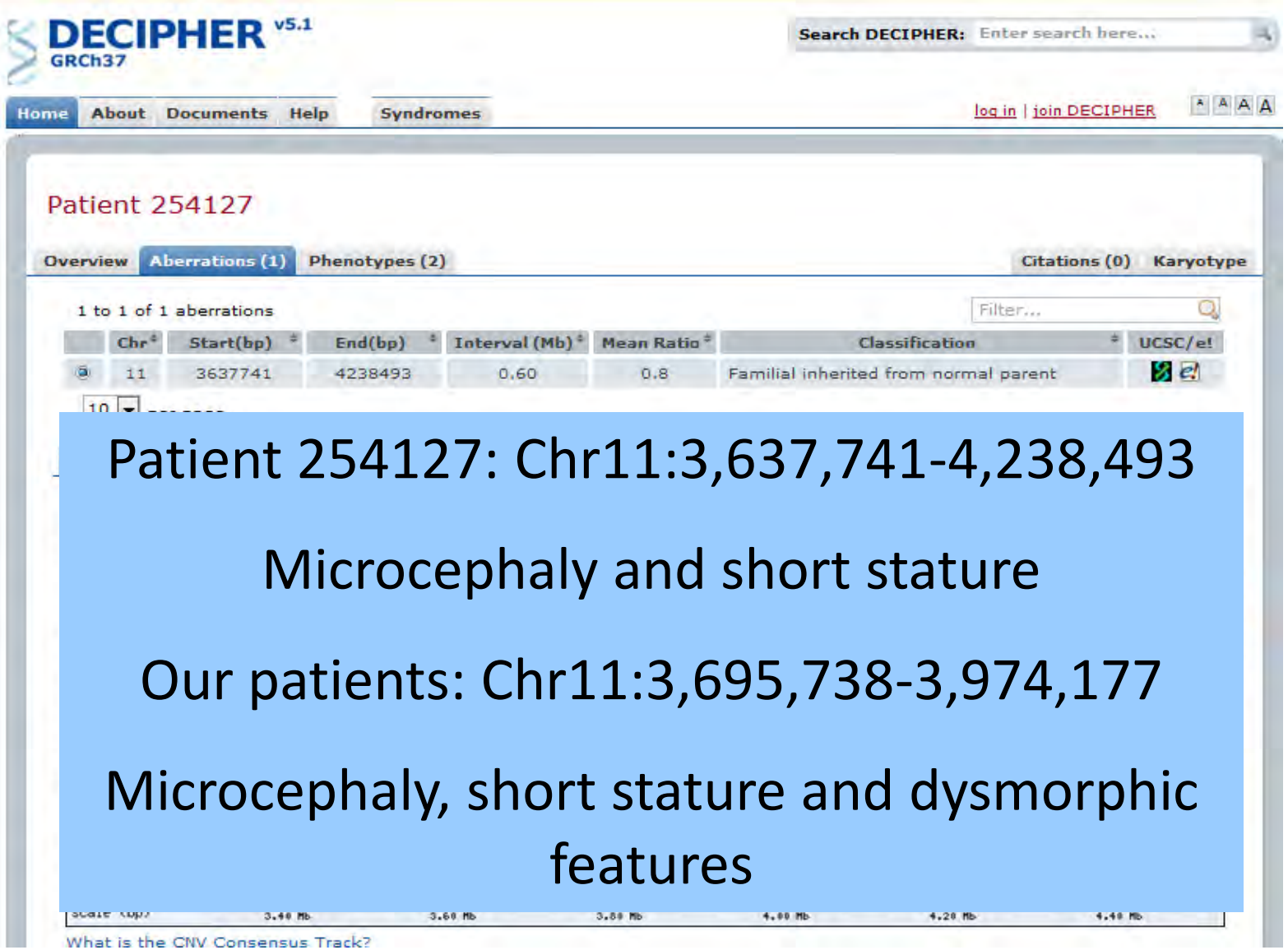


STIM1

- Mutations in STIM1 cause immune dysfunction with T-cell inactivation due to Ca entry defect

- It is not clear what is the effect of duplications of *STIM1*, but overexpression in HEK293 cells modestly enhanced calcium entry

Decipher



Patient 254127: Chr11:3,637,741-4,238,493

Microcephaly and short stature

Our patients: Chr11:3,695,738-3,974,177

Microcephaly, short stature and dysmorphic features

Summary

- Duplications at 11p15.4 have not been reported in the literature and the information regarding clinic significance is still unclear
- Both sisters have the duplication, it must have been inherited from a normal parent
- Region is located near a known imprinted region (BWS region at 11p15.5)

Evolution of Medicine

19th Century	20th Century	21st Century
Treat symptoms	Treat diseases	Predict and Preempt symptoms & disease

Annual Meeting of the Roadmap Multidisciplinary Clinical Research Career Development Program

Elias A. Zerhouni, M.D.
Director

Average Medicine vs. Individualized Medicine

- When it comes to the prac[illegible]edicine, we have been educa[illegible]the "classic case mentality"
- With the technologic advances we have learned of genetic variation giving each one of us our own "flavor of health or disease"
- With our increasing ability to identify and interpret genetic variations, a paradigm shift in medicine is occurring

One-size-fits-all !

One size fits all medicine or "Average Medicine"

Individualized medicine

Individualized medicine

- Individualized medicine is the use of information from a patient's genotype to:
  - Initiate a preventative measure against the development of a disease or condition, or
  - Select the most appropriate therapy for a disease or condition that is particularly suited to that patient

Pharmacogenomics

- Analyzes patient genotypes for cytochrome P450 (CYP) genes CYP2D6 and CYP2C19
- AmpliChip CYP450 Test is based on microarray technology
  - 2 CYP2C19 phenotypes
  - 4 CYP2D6 phenotypes
- It is a major step toward introducing personalized prescribing into the clinical environment

Obstacles to overcome

- Millions of genetic variant exist and identification of them all will take years
- Medication adverse reactions may depend on not only one variant but interacting variants
- Determination of such interactions is also going to take sometime
- Expense of testing

Breaking News

- Mayo Clinic announced in December 2011 that it will start a pilot project to sequence the genomes of volunteers
- "The project will help managers at the clinic decide whether it makes sense to read and store a patient's whole genome early on, instead of ordering single genetic tests as and when the need arises"

But…

- Who is going to store the information?
- How is it stored securely?
- Who has access?
- What are you going to do with information that you or the patient might not necessarily want to find out?
- There are some significant ethical and privacy issues and they are probably more difficult to solve than storing the information

Ethics and Genetics

- In 2008 the Genetic Information Nondiscrimination act (GINA) was passed in the US
  - Prohibits insurers from using personal genetic information to determine eligibility or premiums
  - Prohibits an insurer from requiring a person to have a genetic test
  - Prohibits employers from using a person's genetic information in making hiring, firing, job assignments decisions
  - Prohibits employers from requesting, purchasing, requiring personal or familial genetic information

Direct To Consumer Genetic testing

- When GINA was conceived the availability of genetic testing was limited
- By 2008 when the law was passed genetic testing was available for > 1,300 conditions
- Today we have available tests for > 2,500 disorders
- Most of these tests are offered in the clinical setting
- Some can be offered via DTC test kits

DTC Genetic Testing

Questions ?

- A bright future awaits
- Be mindful of Ethical, legal and Social implications

Bio-Informatic Resources

- OMIM
  - http://www.omim.org
- Gene tests
  - http://www.ncbi.nlm.nih.gov/sites/GeneTests/review?db=GeneTests
- dbSNP
  - http://www.ncbi.nlm.nih.gov/projects/SNP/
- UCSC Genome Browser
  - http://genome.ucsc.edu/
- Ensembl Genome Browser
  - http://www.ensembl.org/index.html
- DECIPHER
  - https://decipher.sanger.ac.uk/application/
- DGV
  - http://projects.tcag.ca/variation/