National Cancer Institute

Genomics in the (Cancer) GWAS Era: What's Next for Translation?

Personal Health Genomics Interest Group (PHGIG) Seminar Series

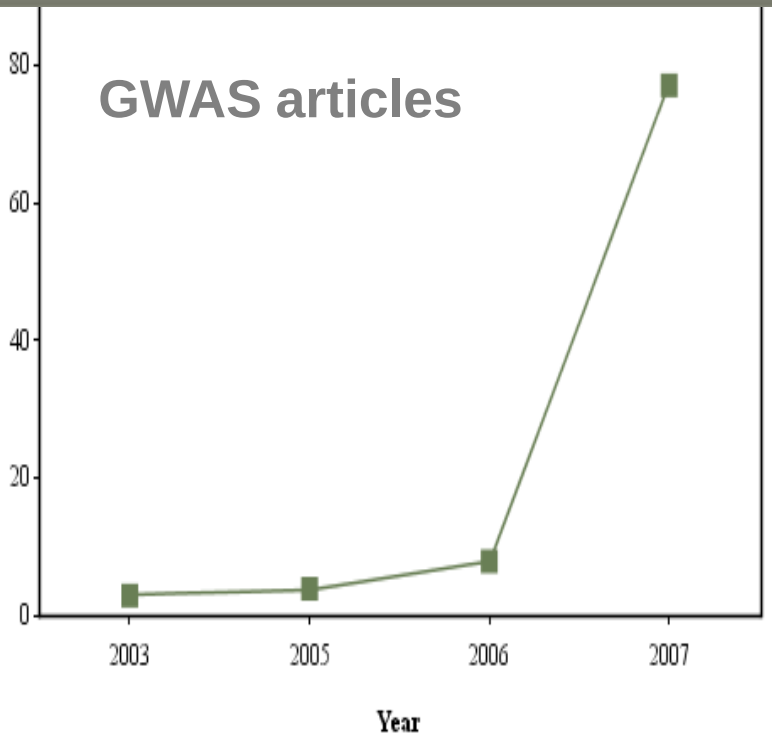
October 22, 2008

Daniela Seminara, Ph.D. M.P.H.
Epidemiology and Genetics Research Program
Division of Cancer Control and Population Sciences

Outline

- Genome-wide association studies (GWAS) of cancer : the story so far
- The future: maximizing the investment in GWAS for translational research
- NCI resources and initiatives

2007: The Year of GWAS



HuGENavigator

http://www.genome.gov/26525384

Catalog of published GWAS

BREAKTHROUGH OF THE YEAR

Human Genetic Variation

Equipped with faster, cheaper technologies for sequencing DNA and assessing variation in genomes on scales ranging from one to millions of bases, researchers are finding out how truly different we are from one another

THE UNVEILING OF THE HUMAN GENOME ALMOST 7 YEARS AGO cast the first faint light on our complete genetic makeup. Since then, each new genome sequenced and each new individual studied has illuminated our genomic landscape in ever more detail. In 2007, researchers came to appreciate the extent to which our genomes differ from person to person and the implications of this variation for deciphering the genetics of complex diseases and personal traits.

Less than a year ago, the big news was triangulating variation between us and our primate cousins to get a better handle on genetic changes along the evolutionary tree that led to humans. Now, we have moved from asking what in our DNA makes us human to striving to know what in my DNA makes me me.

What mak
the number
add variety

Pennisi E, *Science* 2007; 318:1842-43.

Identification of Cancer Susceptibility Genes (in a Nutshell)

- Cancer is a complex, common disease: multiple genetic and environmental factors act in concert.
- Few rare mutations in high penetrance genes have been identified through linkage approaches (*BRCA*, *MMR*, *CDKN2A*)
- Association studies focusing on candidate genes and pathways have yielded very few well-validated associations (*NAT2*)
- Rare coding variants associated with moderate cancer risk have been identified through re-sequencing (*ATM*, *CHEK2*, *MYH*)
- GWAS have recently emerged as a powerful approach to identify lower penetrance common variants associated with cancer susceptibility with an agnostic approach

Cancer GWAS Findings: August 2008

	Studies	# Loci with $p < 10^{-7}$	Promising Regions
Breast	**5**	**9**	**FGFR2, 2q35, CASP8, MAP3K1, TNRC9, 8q24, LSP1, MRPS30, ECHDC1, RNF146**
Prostate	**7**	**16**	**8q24, HNF1B, 10q11, 17q24, Xp11 etc.**
Colon	**4**	**5**	**8q24, SMAD7, 11q23, 10p14, CRAC1, EIF3H**
Lung	**3**	**1**	**15q25.1**
Neuroblastoma	**1**	**1**	**6p22**
Melanoma	**1**	**1**	**20q11.22**

*Ongoing cancer GWAS: pancreas, bladder, testis, kidney, ovary, hematopoietic system, esophageal

Easton and Eales. *Human Molecular Genetics* 2008; 17: R109-R115.

GWAS and Cancer

- GWAS have succeeded in identifying at least 28 new cancer susceptibility loci with variants associated with common cancers (breast, prostate, colon, lung, melanoma)
- Although each common variant confers only a modest increase in risk, it may explain a large % of a given cancer total burden (PAR)
- Most robust associations in cancer GWAS have not been within previously known candidate genes, providing clues for the identification of new pathways
- For most of these studies, results have been independently replicated (mostly in Caucasian, but also African Americans and Asians)

DCCPS-NCI Initial GWAS Investments: March 2008

- EGRP portfolio includes *39 funded and pending projects* that directly or indirectly support GWAS
- Investment has grown:

  $2M in 2005 → *$14.5 M in 2008*

- Center for Inherited Diseases Research (CIDR) : *21 cancer GWAS projects* partially supported
- Division of Cancer Epidemiology and Genetics (examples):
  - CGEMS (Prostate and Breast Cancer)
  - PanScan 1 and 2: Collaborative multi-cohort and case-control Pancreatic Cancer
- More Planned by Cohort Consortium
  http://epi.grants.cancer.gov/Consortia/cohort.html

Results are Intriguing, Sometimes Unexpected

- Some identified variants are associated with increased risk across several cancer types
- Some regions identified contain no known cancer genes
- Preliminary evidence of genetic determinants of precursory traits (nicotine addiction and lung cancer, skin type and melanoma)
- Some GWAS are beginning to address cancer subtypes (estrogen receptors positive breast cancer, aggressive prostate cancer)
- Causality: *FGFR2* (chr. 10) and breast cancer, rs6983267 (8q24)and multiple cancers, rs10993994 in *MSMB* and prostate cancer

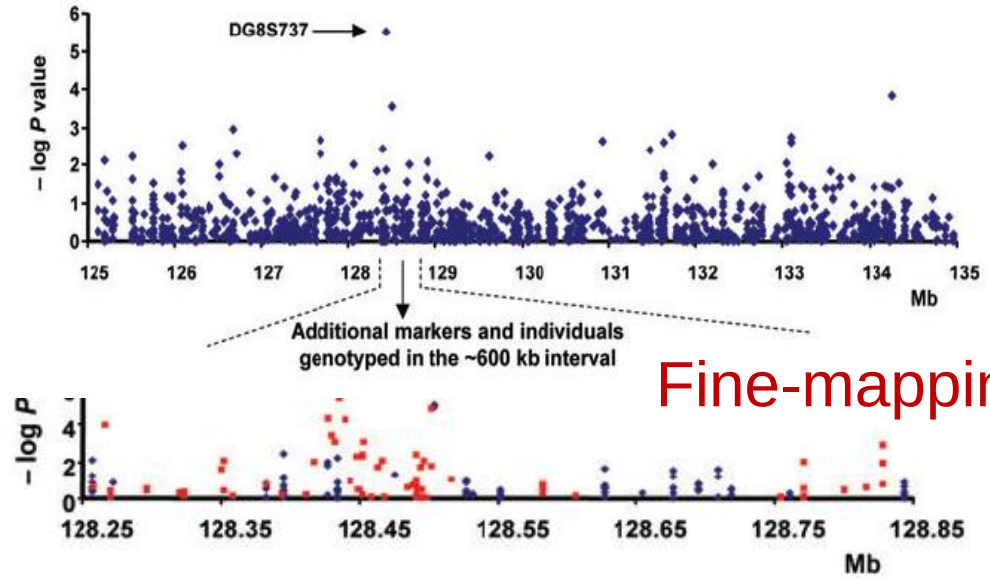
An Example: The Case of 8q24

A common variant associated with prostate cancer in European and African populations

Laufey T Amundadottir[1,12], Patrick Sulem[1,12], Julius Gudmundsson[1,12], Agnar Helgason[1], Adam Baker[1], Bjarni A Agnarsson[2], Asgeir Sigurdsson[1], Kristrun R Benediktsdottir[2], Jean-Baptiste Cazier[1], Jesus Sainz[1], Margret Jakobsdottir[1], Jelena Kostic[1], Droplaug N Magnusdottir[1], Shyamali Ghosh[1], Kari Agnarsson[1], Birgitta Birgisdottir[1], Louise Le Roux[1], Adalheidur Olafsdottir[1], Thorarinn Blondal[1], Margret Andresdottir[1], Olafia Svandis Gretarsdottir[1], Jon T Bergthorsson[1], Daniel Gudbjartsson[1], Arnaldur Gylfason[1], Gudmar Thorleifsson[1], Andrei Manolescu[1], Kristleifur Kristjansson[1], Gudmundur Geirsson[3], Helgi Isaksson[2], Julie Douglas[4], Jan-Erik Johansson[5], Katarina Bälter[6], Fredrik Wiklund[6], James E Montie[7], Xiaoying Yu[8], Brian K Suarez[9], Carole Ober[10], Kathleen A Cooney[7,11], Henrik Gronberg[6], William J Catalona[8], Gudmundur V Einarsson[3], Rosa B Barkardottir[2], Jeffrey R Gulcher[1], Augustine Kong[1], Unnur Thorsteinsdottir[1] & Kari Stefansson[1]

Summary:

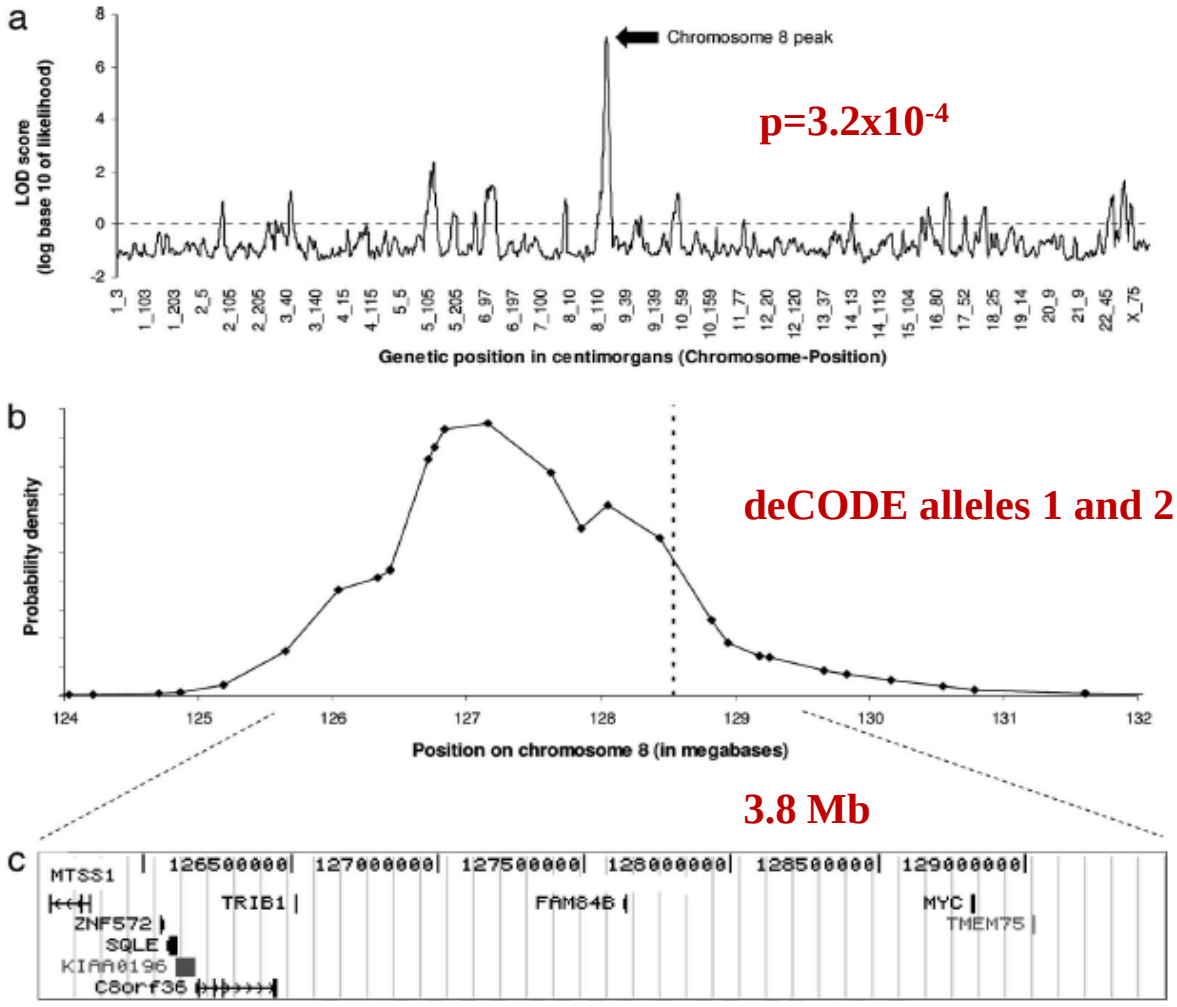
- **1-2 risk alleles at 8q24 in 90kb LD block**
- **Replication in Swedish, European and African Americans**
- **Allele 2:Caucasian PAR 8%; A-A PAR 41%**
- **No known gene (MYC ~200kb)**



Fine-mapping

Risk Allele	OR	P	MAF
1	1.5	$2x10^{-11}$	10%
2	1.6	$3x10^{-11}$	7%

Admixture mapping in African American men identifies 8q24 as prostate cancer risk locus



- A gene mapping strategy that has good power to detect risk variants with large allele frequency differences between populations
- Recent admixture between populations generates large chromosomal segments of discrete ancestry that can be used to track risk alleles
- Best suited for diseases with incidence rates that vary across populations
- Prostate cancer: is the greater risk in African Americans due to alleles that are more common in African vs. European populations?

Freedman et al. *PNAS* 2006; *103:*

Validation of deCODE risk alleles at 8q24

Multiethnic Cohort

Ethnic Group	Cases	Controls	Risk Allele 1, MAF *rs1447295*	OR (95% CI)	P-value
African Americans	674	644	30.7%	1.17 (0.99-1.37)	*0.066*
Native Hawaiians	70	68	16.2%	3.02 (1.66-5.50)	*0.0003*
Japanese	449	465	17.2%	1.48 (1.18-1.86)	*0.0007*
Latinos	640	567	9.5%	1.48 (1.14-1.91)	*0.0028*
Whites	455	447	10.0%	1.35 (1.01-1.80)	*0.044*
All groups	2,288	2,191		1.36 (1.22-1.51)	$2.3x10^{-8}$

Freedman et al. PNAS 103:2006

- Schumaker et Al. NCI Breast and Prostate Cancer Consortium (BPC3) *Can Res* 2007
- Wang et al. *Can Res*, 2007
- Severi et al. *CEBP*, 2007
- Suuriniemi et al. *CEBP*, 2007

Validation of deCODE risk alleles at 8q24 (continued)

- 8q24 is a confirmed risk locus for prostate cancer
- Could the 2 deCODE risk alleles at 8q24 explain the admixture signal?
- Admixture signal still apparent after accounting for these 2 risk alleles (Haiman)
- There must be additional unmapped risk alleles at 8q24 that are highly differentiated in frequency between populations of African and European ancestry…..

Next Piece of the Puzzle: GWAS identify colorectal cancer risk (CRC) locus at 8q24

- July 2007
  - Haiman et al. rs6983267 and closely linked variants on 8q24 contribute to risk for colon cancer
  - Zanke et al., Tomlison et al. Association of rs6983267 and CRC independently replicated
- May 2008 – Tenesa et al. Independent replication
- 5 other prostate cancer-associated variants not associated with CRC
- rs6983267 is common in general population; frequency varies from 85% (African-Americans) to 30% (Japanese)
- CRC-associated variants lay in “gene desert”

  regions with highly conserved DNA
- **What are the underlying mechanisms?**

Genome-wide association study identifies novel breast cancer susceptibility loci

Multi-stage scan: 22,000 cases and 22,000 controls

rs Number	Gene	Position*	m.a.f.†	Per allele OR (95% CI)	HetOR (95% CI)	HomOR (95% CI)	*P*-trend		
							Stages 1 and 2	Stage3	Combined
rs2981582	*FGFR2*	10q 123342307	0.38 (0.30)	1.26 (1.23–1.30)	1.23 (1.18–1.28)	1.63 (1.53–1.72)	4×10^{-16}	5×10^{-62}	2×10^{-76}
rs12443621	*TNRC9/ LOC643714*	16q 51105538	0.46 (0.60)	1.11 (1.08–1.14)	1.14 (1.09–1.20)	1.23 (1.17–1.30)	10^{-7}	9×10^{-14}	2×10^{-19}
rs8051542	*TNRC9/ LOC643714*	16q 51091668	0.44 (0.20)	1.09 (1.06–1.13)	1.10 (1.05–1.16)	1.19 (1.12–1.27)	4×10^{-6}	4×10^{-8}	10^{-12}
rs889312	*MAP3K1*	5q 56067641	0.28 (0.54)	1.13 (1.10–1.16)	1.13 (1.09–1.18)	1.27 (1.19–1.36)	4×10^{-6}	3×10^{-15}	7×10^{-20}
rs3817198	*LSP1*	11p 1865582	0.30 (0.14)	1.07 (1.04–1.11)	1.06 (1.02–1.11)	1.17 (1.08–1.25)	8×10^{-6}	10^{-5}	3×10^{-9}
rs2107425	*H19*	11p 1977651	0.31 (0.44)	0.96 (0.93–0.99)	0.94 (0.90–0.98)	0.95 (0.89–1.01)	7×10^{-6}	0.01	2×10^{-5}
rs13281615		8q 128424800	0.40 (0.56)	1.08 (1.05–1.11)	1.06 (1.01–1.11)	1.18 (1.10–1.25)	2×10^{-7}	6×10^{-7}	5×10^{-12}

Breast cancer risk allele is not linked to the 7 other risk alleles and is not associated with prostate or colorectal cancer risk in the MEC

Setiawan et al. *CEBP* 2007;16:

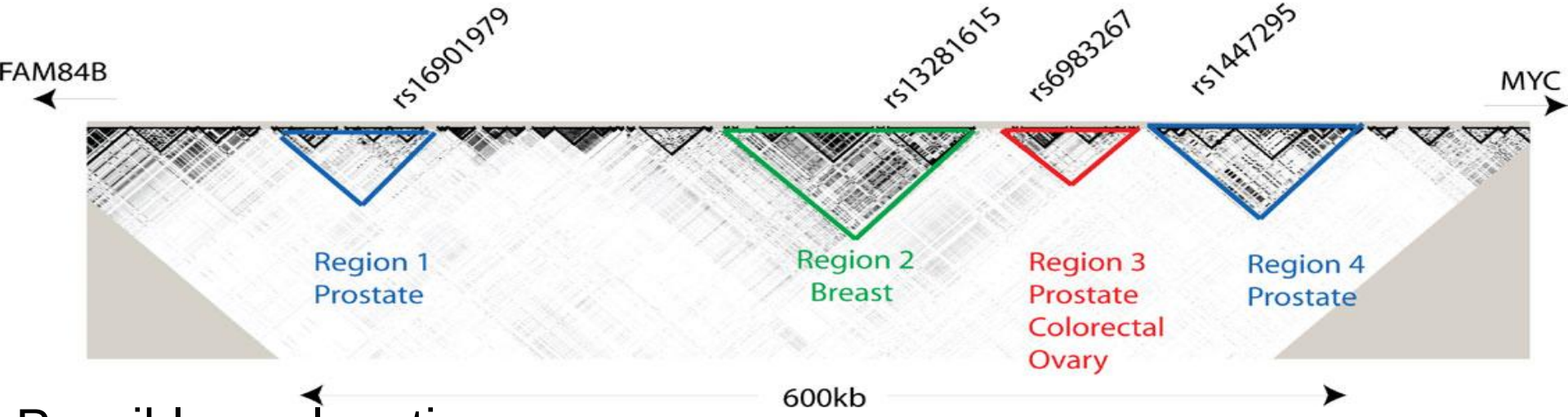
Haiman et al. *Nat Genet* 2007; 39:

Summary

8q24:

- 7 independent risk alleles for prostate cancer in 4 regions of association
- A specific marker of both prostate and colorectal risk
- A common risk allele for breast cancer

Schematic of 8q24 Region (from Easton and Eeles, 2008)



Possible explanations:

- Multiple different mechanisms for cancer risk
- A single mechanism (gene) with tissue-specific effects

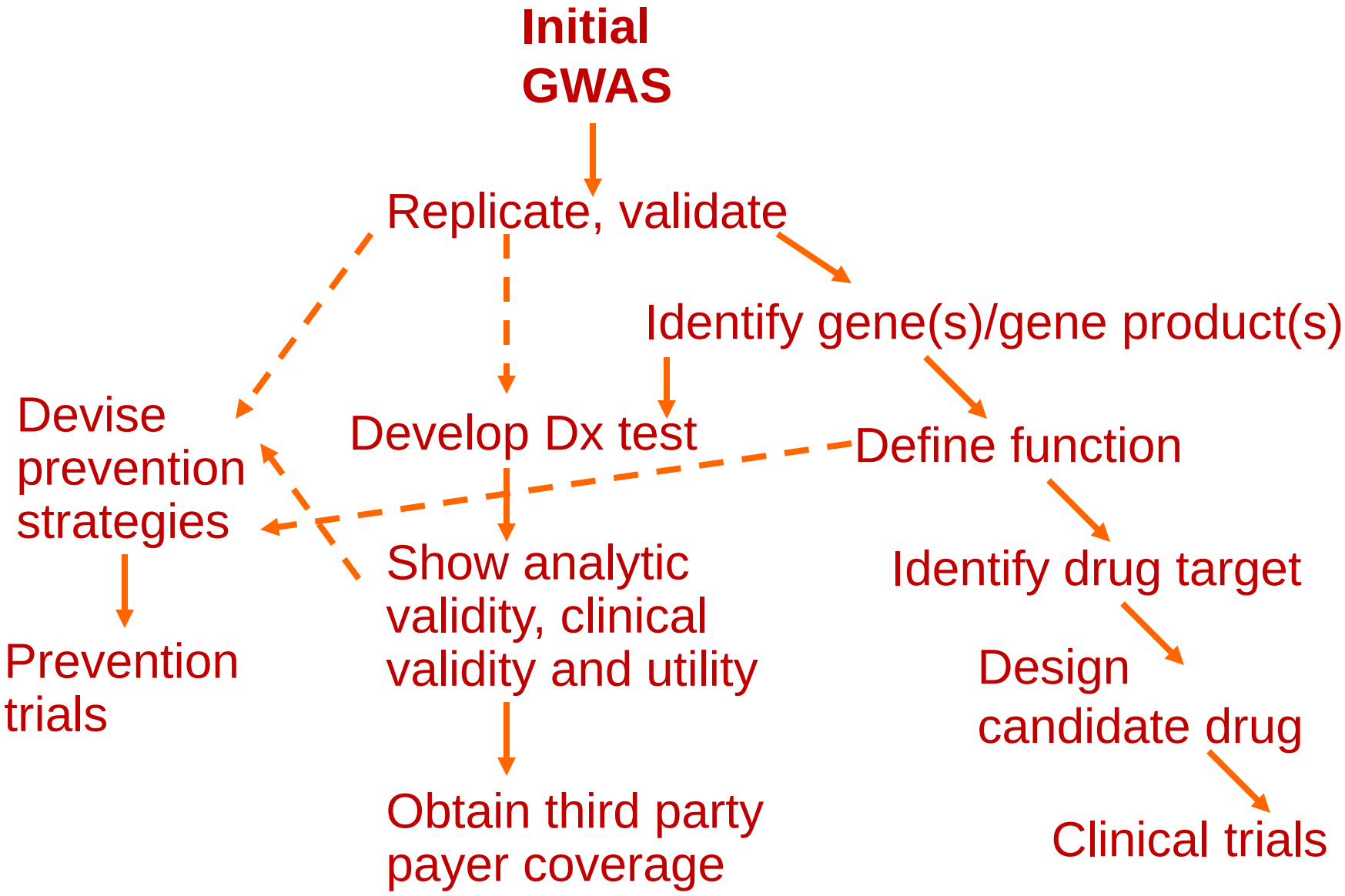
Possible Mechanisms for 8q24 Contribution to Cancer Risk

- Genomic instability
- Un-annotated gene(s) or microRNA(s)
- Long-range regulation of gene expression (Enhancers/Repressors)

Lessons Learned (but more to come)

- First findings of genetic variations strongly associated with risk in multiple cancers: upstream mechanism and tissue-specific regulation?
- Frequencies are different in different populations: may account for different population burden of disease
- Combination of markers (plus environmental factors) may engender composite risk in individuals with exposures comparable to that of high penetrance
- Epidemiologists, clinicians and biologists need to work hand in hand: importance of team science/consortia
- **Theoretical applicability to individualized prevention and early detection**

Capitalizing on GWAS Data for Prevention, Diagnosis and Therapy



Challenge 1 : How to Accelerate Efficient and Effective Translation of GWAs Data

We will be sitting on a “gold mine” of data …
how do we achieve efficient and effective translation into genomic-guided prevention and medicine?

COMMENTARY

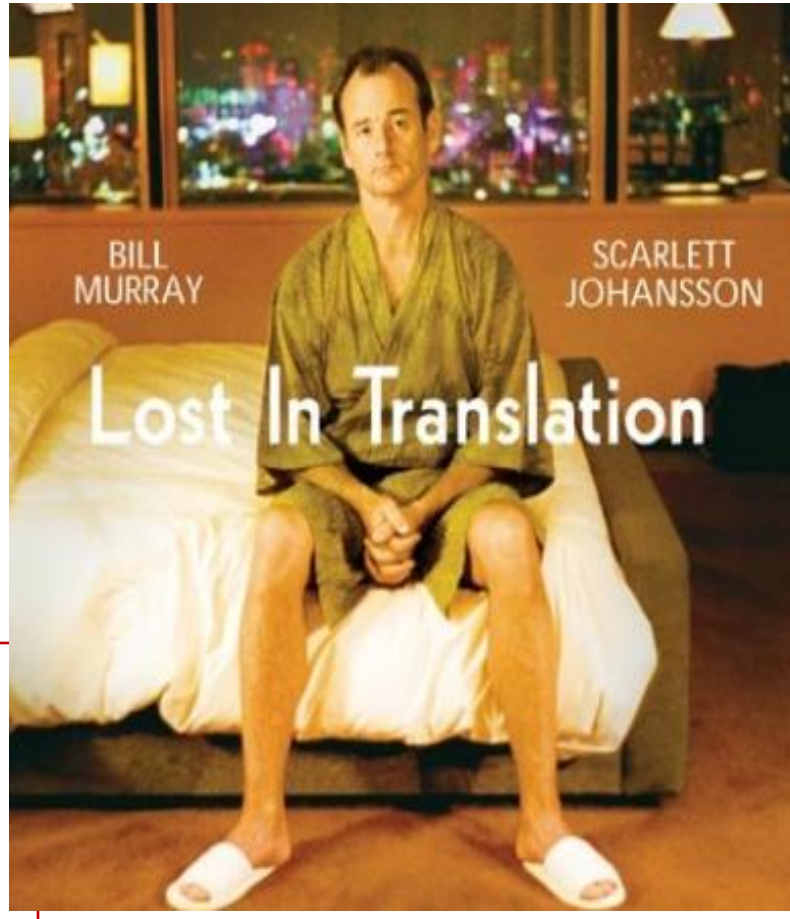
The Genome Gets Personal—Almost

W. Gregory Feero, MD, PhD
Alan E. Guttmacher, MD
Francis S. Collins, MD, PhD

IT'S THE "YEAR OF PERFECT VISION," 2020. AMY, AGE 21 YEARS, visits with her physician and elects to have complete genome sequencing. At a follow-up visit, Amy chooses to learn of her genetic risk factors for heart disease, diabetes, breast cancer, and colon cancer. Amy's physician provides her with risk scores for those disorders, and with suggestions for lifestyle modifications. Specifically, Amy is alerted to her particularly high risk of developing type 2 diabetes, and her physician recommends a rigorous program of diet and exercise

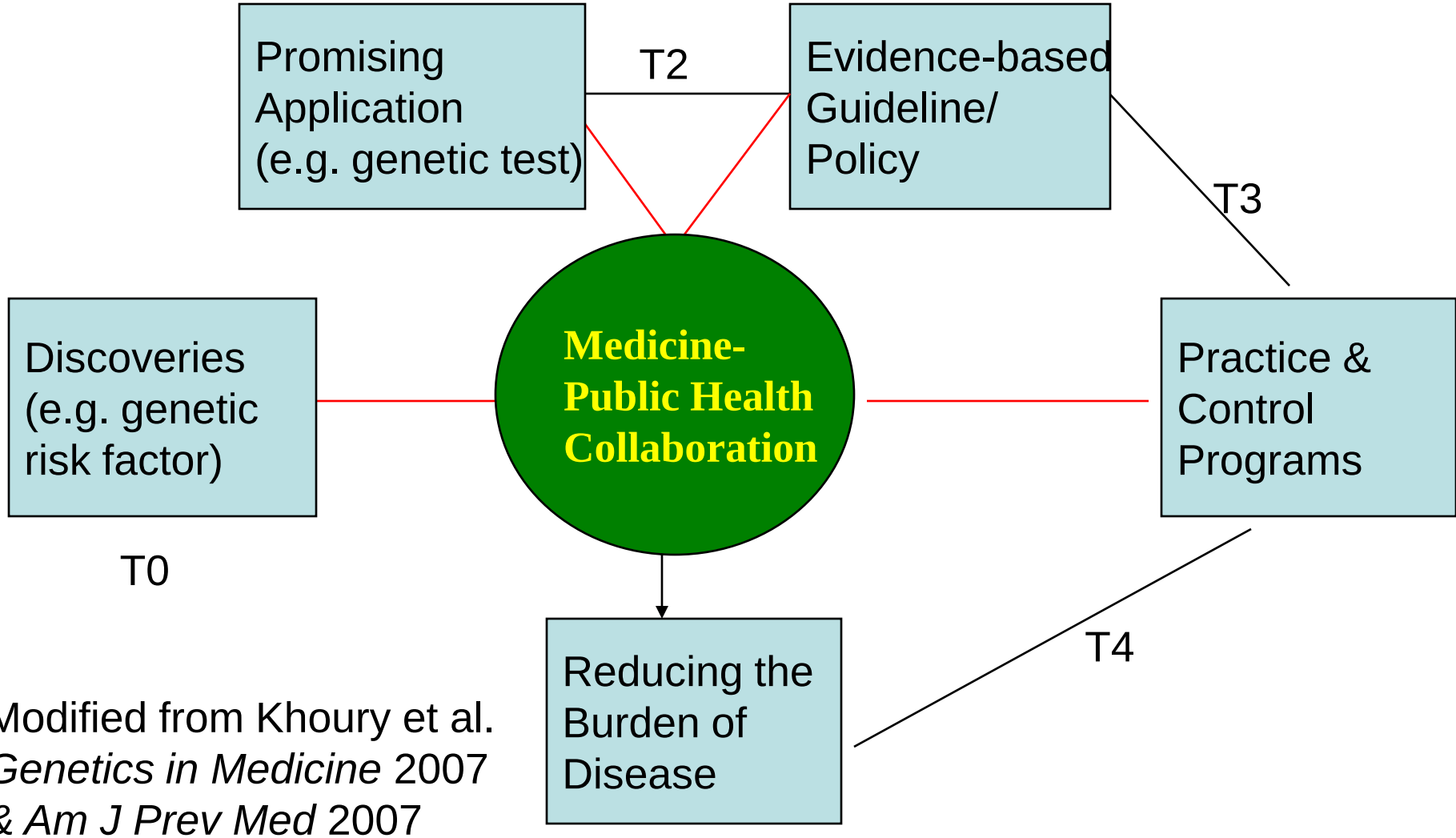
ENCODE project,[5] the "1000 Genomes" project,[6] and initiatives to bring full genome sequencing costs below $10 000[7] promise to accelerate knowledge generation further.

Perhaps the most breathtaking recent advances relevant to personalized medicine come from the current explosion of genome-wide association studies. These studies are based on the ability to search the genomes of large numbers of individuals in an unbiased way for statistical associations between the most common form of genetic variation, single nucleotide polymorphisms (SNPs), and the occurrence of disease. Unthinkably expensive as recently as 2004, genome-wide association studies have been made possible through the availability of HapMap data[3] and the ability to genotype individuals rapidly and accurately at hundreds of thou-



"I predict that comprehensive, genomics-based health care will become the norm with individualized preventive medicine and early detection of illnesses” (Zerhouni, 2006)

Medicine-Public Health Collaboration in Genomics Translation



Modified from Khoury et al. *Genetics in Medicine* 2007 *& Am J Prev Med* 2007

Next Post-GWAS Step for Cancer

Discovery
T0

Promising Applications
T1

Discovery and Replication
- **Finding of new associations through pooled analyses**
- **Independent replication of associations**
- **Fine Mapping of association signals**

Biological studies
- **Identification of risk-enhancing variant**
- **Examination of functional consequence of variant**
- **Determination of biological mechanism of risk-enhancement**

Epidemiologic studies
- **Evaluation of gene-gene interactions**
- **Evaluation of gene-environment interactions**
- **Assessment of penetrance and population attributable risk**
- **Development of complex risk models**
- **Evaluation of clinical validity of risk models in observational studies**

Next step in the translational process

Adapted from M. Khoury

What is Needed to Fully Exploit the Wealth of GWAs Studies (T0 to T1)

I. Finding additional loci containing cancer causal variants

II. Refining the location and phenotypic consequences of causal variants

III. Progressing from known loci and variants to functional mechanisms

Finding additional loci containing cancer causal variants: More GWAS?

- The 28 plus loci discovered by cancer association studies account for relatively low % of the disease: most of the cancer inherited component remains unexplained
- GWAS on less common cancers have not been published/performed yet
- Studies in different ethnic populations: race-specific genetic variations or allele frequencies
- Some of the studies were underpowered for lower effect variants

Finding additional loci containing cancer causal variants: Rare and low frequency intermediate penetrance variants

- Meta analyses of cancer-specific GWAS
- Resequencing
- Recently isolated and self-contained populations

Finding additional loci containing cancer causal variants: Beyond the main effect

- **GXG** detection depends on knowledge of causal variants, larger data sets and more efficient computational approaches
- **GXE** detection depend on identifying and accurately measuring relevant exposures in large populations in a standardized manner

Finding additional loci containing cancer causal variants: Variations not captured by current GWAS

- Heritable epigenetic changes
- Copy number variations (CNV)

Genetic and Phenotypic Characterization of Associated Loci and Causal Variants

- Genetic refinement of GWAS association signal (fine mapping)
  - Resequencing across the region of association
  - Resequencing across the region of association
- Phenotypic refinement of GWAS association signal
  - Proven correlation (diabetes, obesity)
  - Connected phenotypes (cancers, immune diseases)
  - No known connection/correlation: novel mechanisms?
  - Identification of new phenotypes through association with GWAS-identified genotype

Functional Strategies to Support Causal Variant and Gene Identification

- Use of expression quantitative trait locus data (eQTL)
- Use of genome annotation
- Functional experiments in vitro or in animal models
- Deep re-sequencing approach

Need For a Coordinated Multi-Step Approach

- Complete GWAS discovery phase
- Replication phase
- Fine Mapping phase
- Biological Validation phase

This stepwise approach needs to be performed in a timely and coordinated fashion, and usually involves a large number of institutions. Particular attention to quality control and data harmonization are required. Expertise in all pertinent scientific areas needs to be represented in the research team.

NCI's Replication Efforts

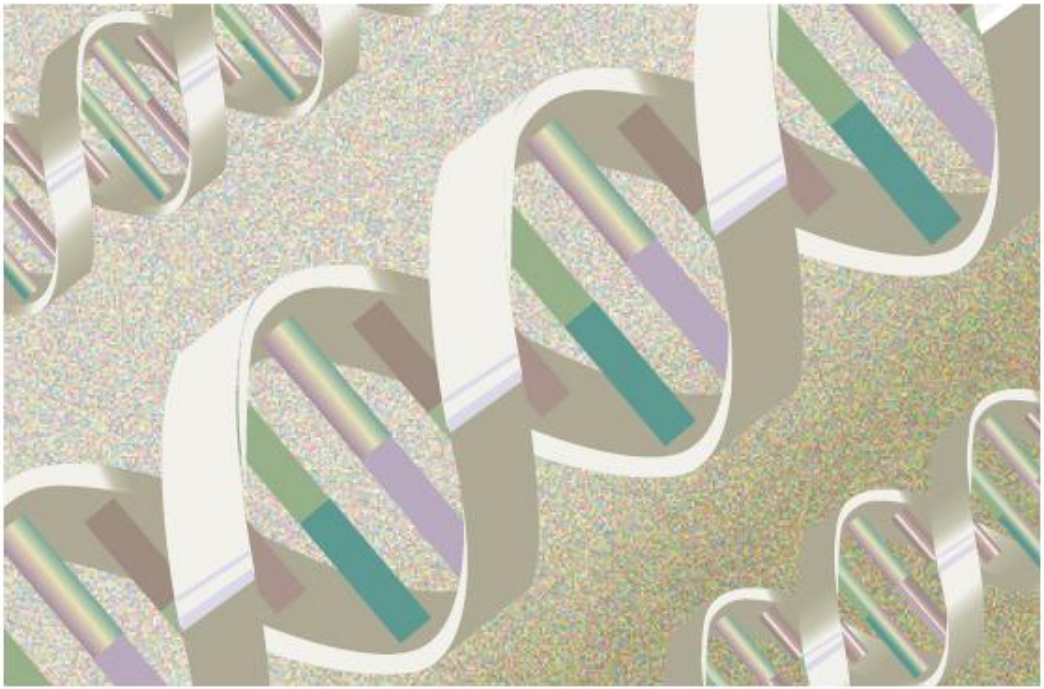
Replicating genotype–phenotype associations

What constitutes replication of a genotype–phenotype association, and how best can it be achieved?

NCI-NHGRI Working Group on Replication in Association Studies

The study of human genetics has recently undergone a dramatic transition with the completion of both the sequencing of the human genome and the mapping of human haplotypes of the most common form of genetic variation, the single nucleotide polymorphism (SNP)[1–3]. In concert with this rapid expansion of detailed genomic information, cost-effective genotyping technologies have been developed that can assay hundreds of thousands of SNPs simultaneously. Together, these advances have allowed a systematic, even 'agnostic', approach to genome-wide interrogation, thereby relaxing the requirement for strong prior hypotheses.

So far, comprehensive reviews of the published literature, most of which reports work based on the candidate-gene approach, have demonstrated a plethora of questionable genotype–phenotype associations, replication of which has often failed in independent studies[4–7]. As the transition to genome-wide association studies occurs, the challenge will be to separate true associations from the blizzard of



studies because of issues in either the initial study or the attempted replication[4–6,32,33]. Small

conclusion from the literature because follow-up studies have not consistently analysed the

Chanock et al, *Nature* 2007; 447:655-660.

NCI's Replication and Fine Mapping Funding Opportunity Announcements

- FY2008 Administrative Supplements for Gene Identification Efforts: Replication and Fine-Mapping Studies for The Genes, Environment, and Health Initiative (GEI)
  - To date, numerous GWAS of complex traits have yielded promising results. However, follow-up studies are needed to eliminate false positives, extend the findings to diverse populations (diverse in terms of ethnicity or environmental exposures), and narrow the association interval.
- FY2009 RFA-CA-09-003

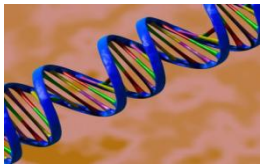
  Replication and Fine-Mapping Studies for the Genes Environment and Health Initiative (GEI)(R01)

The Cancer Post-GWA Initiative (T0 to T1): Coming Soon from NCI

Discovery

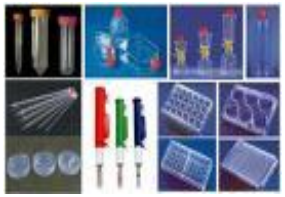
Initial evidence of association between genetic regions and disease (includes pooled analyses)

Replication



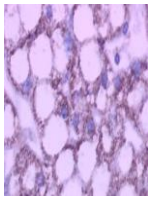
Validation of findings; selection of most promising regions

Fine Mapping



Identification of risk enhancing variants (includes sequencing/genotyping)

Biological Validation



Functional studies in cell lines and animal models

Clinical and public health applications

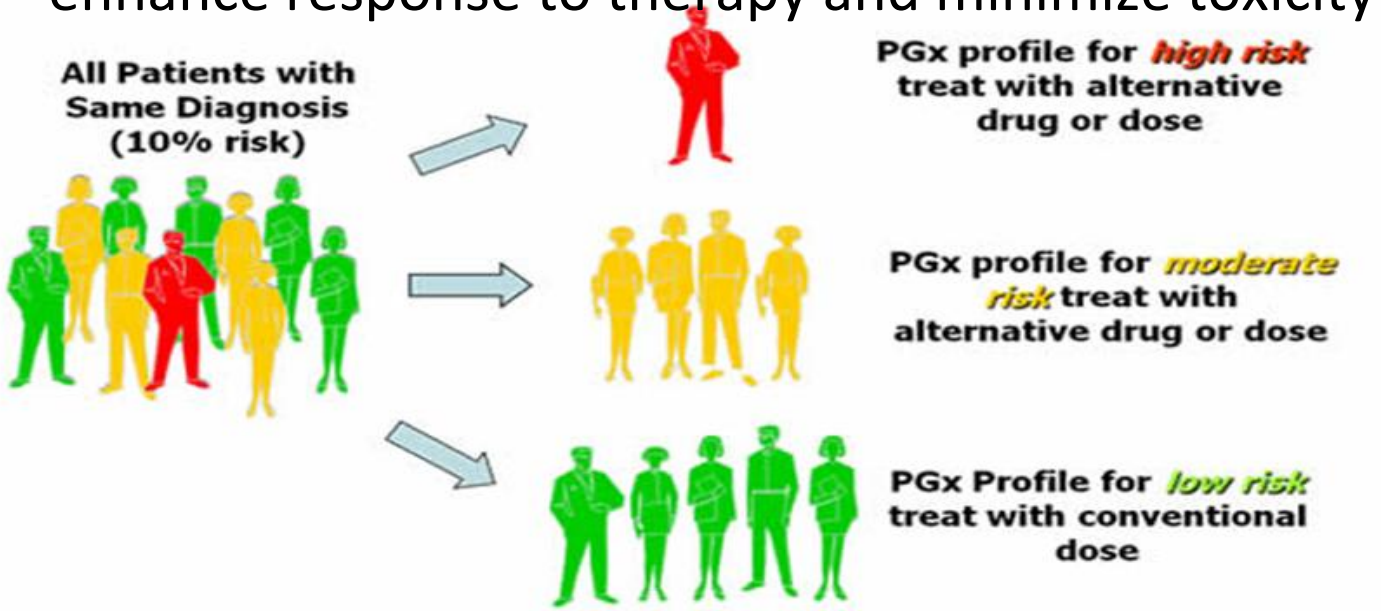
Translational Phases (T1 to T4): Khoury Model, 2008

Beyond T1:

Other NCI GWAS-Related Activities and Initiatives

Pharmacogenomics and Pharmacoepidemiology WG Goals

- Develop recommendations to move NCI's research agenda forward in pharmacogenomics and pharmacoepidemiology
- Identify research opportunities to elucidate specific epidemiologic, clinical, and genomic profiles that could enhance response to therapy and minimize toxicity



National Cancer Institute

U.S. National Institutes of Health | www.cancer.gov

TARGET

Therapeutically Applicable Research to Generate Effective Treatments

Promoting Discoveries
Leading to Cures

What is TARGET? The Therapeutically Applicable Research to Generate Effective Treatments (TARGET) Initiative seeks to harness the power of modern genomics technologies to rapidly identify valid therapeutic targets in childhood cancers so that new, more effective treatments can be developed and ultimately bring new hope to children and their families who face the devastating burden of these diseases.

Learn more

http://target.cancer.gov

Areas of Research

Cancers Selected for Study

Areas for TARGET Initiative Research Focus

- Underlying premise: Genes that are consistently altered by mutation, copy number change, or LOH will highlight cellular pathways for therapeutic exploitation:
  - High-throughput array-based technologies to comprehensively characterize genomic and transcriptomic profiles
  - Gene resequencing to identify genes that are consistently altered in specific childhood cancers
- Functional validation to validate putative therapeutic targets

Childhood Cancer TARGET Initiative

- TARGET program ongoing for high-risk acute lymphoblastic leukemia and for high-risk neuroblastoma
- TARGET programs for childhood sarcomas under development

Resources Needed to Progress from Current Findings of GWA Studies

- Large samples in diverse populations for multiple diseases/traits*
- Complete knowledge of common variation across the genome in multiple populations
- Methods to interrogate efficiently structural variation in large samples
- Improved sequencing technology and/or other methods for interrogating low frequency variation
- Computational methods to interpret sequence data from large samples
- Expression data from densely genotyped human samples and covering diverse tissue types
- Improved genome annotation, especially of noncoding regions
- Relevant and validated functional assays for associated genes*
- Tractable animal models or highly relevant in vitro models in which human causal variants can be assessed*
- Coordinated assessment of environmental exposures and disease outcomes in large cohorts with DNA samples available
- Computational tools for comprehensive assessment of GG and GE joint effects
- Assessment of the role of epigenetics in the inherited risk of disease

McCarthy and Hirschhorn, *Human Molecular Genetics* 2008

Thank You

Acknowledgements ad Slides' Credits

- Elizabeth Gillanders, NCI
- Muin Khoury, CDC
- Deborah Winn, NCI
- Chris Haiman, USC
- Daniela Gerhard, NCI
- Robert Haile, USC
- David Hunter, Harvard
- Malcolm Smith, NCI
- James Jacobson, NCI
- Andy Freedman, NCI
- Leah Sansbury, NCI
- Margaret Tucker, NCI
- And the Investigators from the EGRP-Supported Consortia