Pharmacogenomics Research Network
PGRN

P-STAR
PGRN Statistical Analysis Resource

Pharmacogenetics: Past, Present, and Future

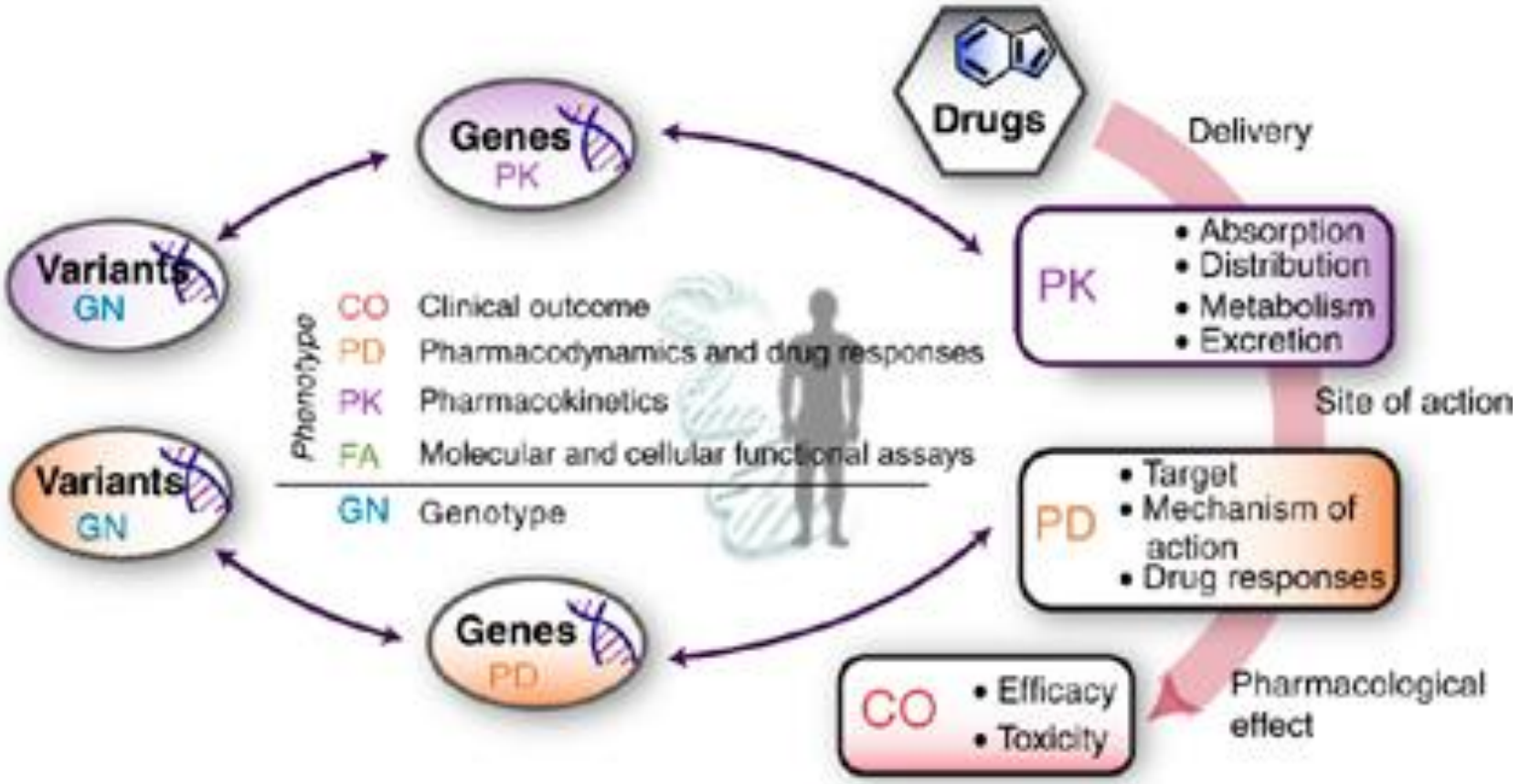
Brooke L. Fridley, PhD

Associate Professor of Biostatistics
University of Kansas Medical Center
Site Director, K-INBRE Bioinformatics Core
Director, Biostatistics and Informatics Shared Resource,
University of Kansas Cancer Center

KU MEDICAL CENTER
The University of Kansas

KU

What is Pharmacogenomics?



The Pharmacogenetics Research Network: From SNP Discovery to Clinical Drug Response

K M Giacomini, C M Brett, R B Altman, N L Benowitz, M E Dolan, D A Flockhart, J A Johnson, D F Hayes, T Klein, R M Krauss, D L Kroetz, H L McLeod, A T Nguyen, M J Ratain, M V Relling, V Reus, D M Roden, C A Schaefer, A R Shuldiner, T Skaar, K Tantisira, R F Tyndale, L Wang, R M Weinshilboum, S T Weiss and I Zineh for the Pharmacogenetics Research Network

Pharmacogenomics & Precision Medicine

Aims to deliver the **correct drug** or treatment:

- At the right **time**
- To the right **patient**
- At the right **dose**

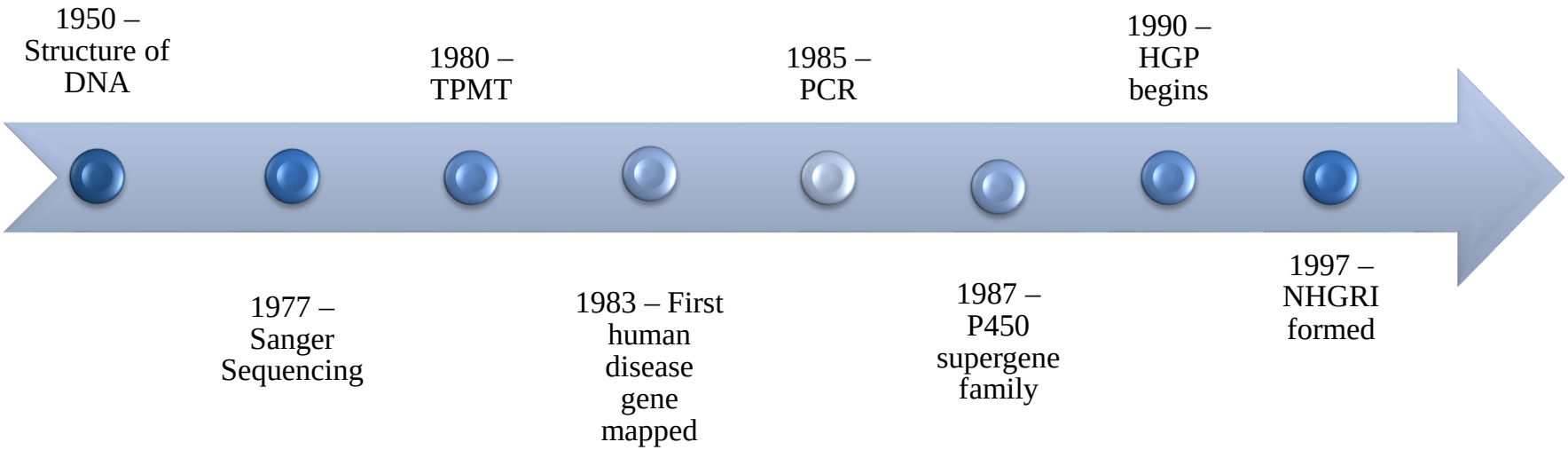
Can only be achieved when we have accurate clinical tests (**BIOMARKER**) and companion drugs at our disposal

What is a Biomarker?

- **Biomarker:** "A characteristic that is objectively measured and evaluated as an indicator of
  - normal biological processes,
  - pathogenic processes, or
  - pharmacologic responses to a therapeutic intervention."

Biomarkers Definitions Working Group. Biomarkers and surrogate end points: preferred definitions and conceptual framework. *Clin Pharmacol Ther* 2001;**69**:89–95

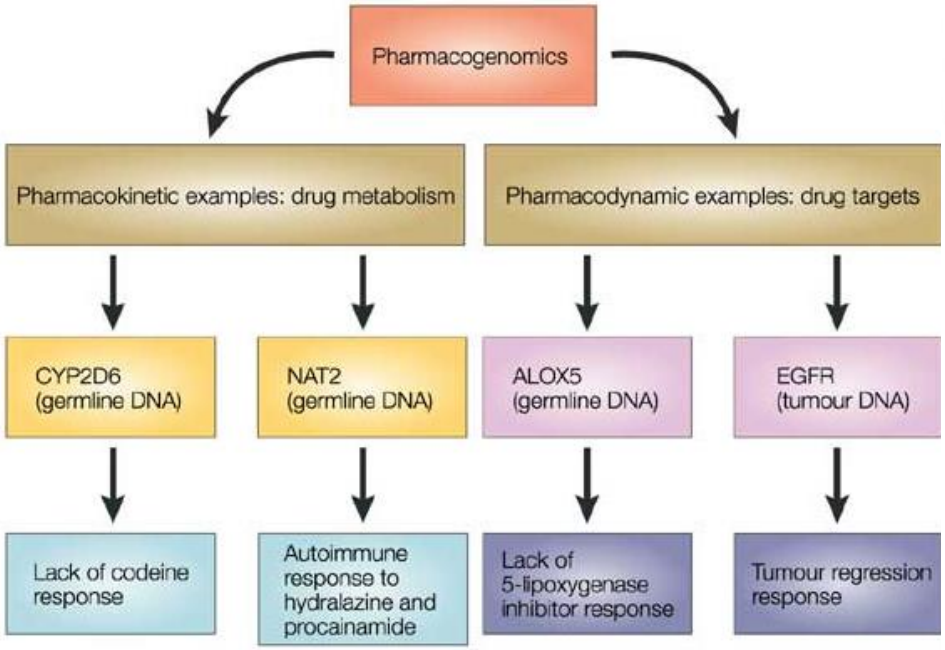
Candidate Gene Era

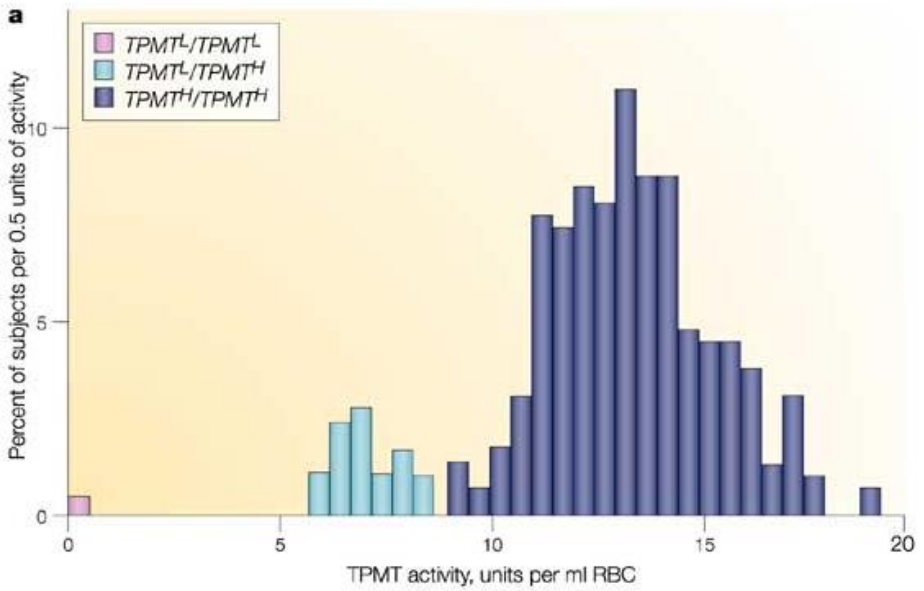


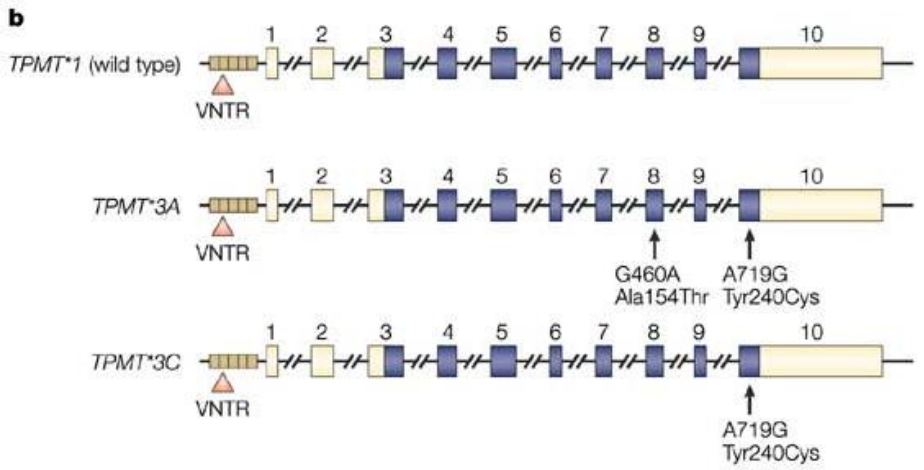
Candidate Variant & Gene Studies

- Genotyping
- RT-PCR
- Re-sequencing genes (exons) with Sanger Sequencing

PK and PD and * Alleles (haplotypes)







FROM THE FOLLOWING ARTICLE:

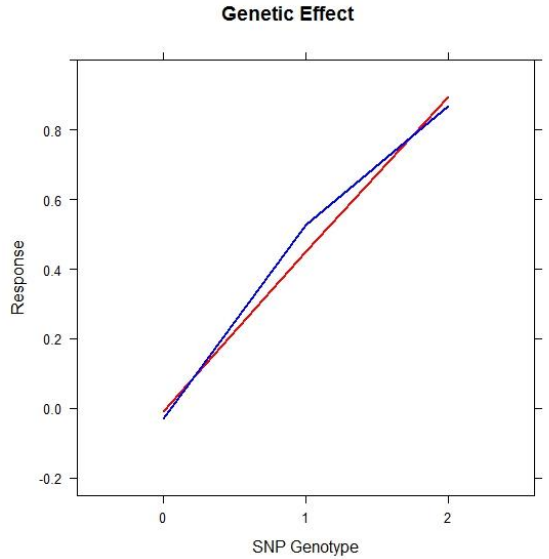
Pharmacogenomics: bench to bedside

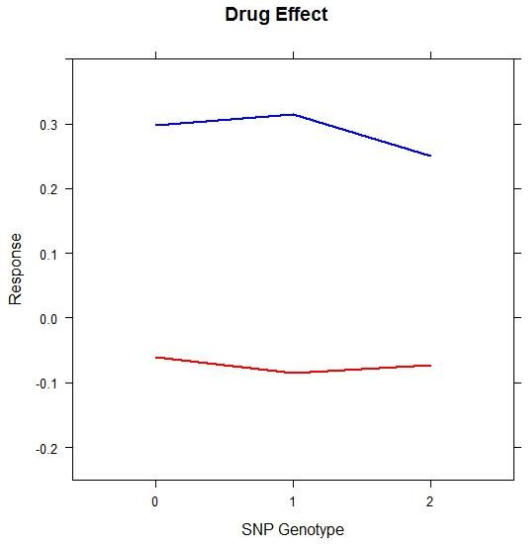
Richard Weinshilboum & Liewei Wang

Nature Reviews Drug Discovery **3**, 739-748 (September 2004)

doi:10.1038/nrd1497

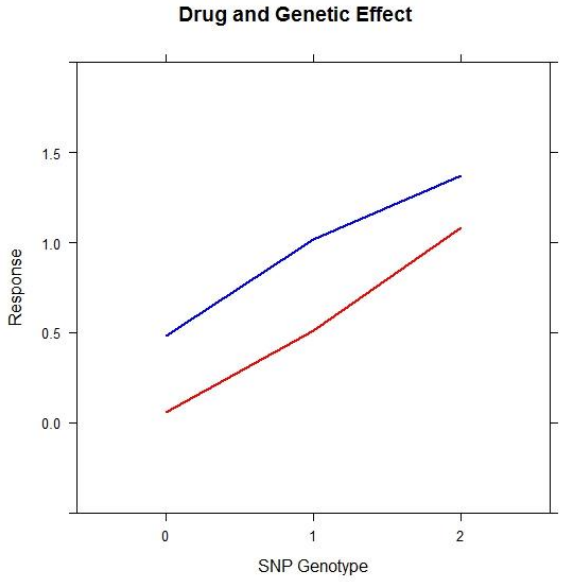
PGx Effect = Gene*Drug Interaction

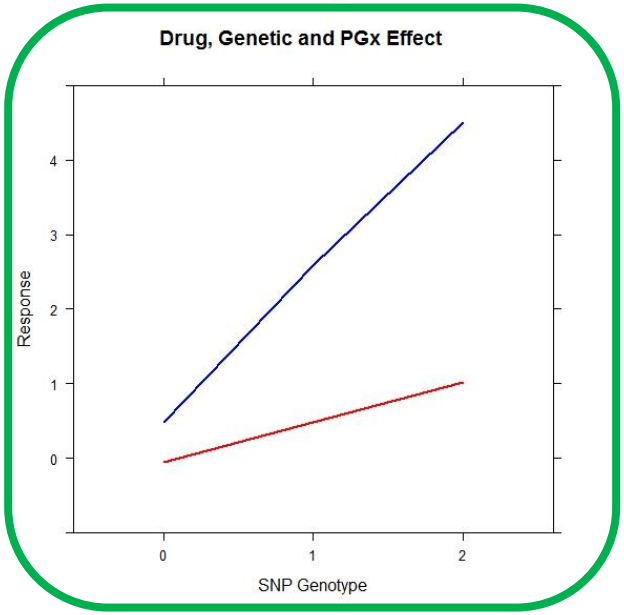




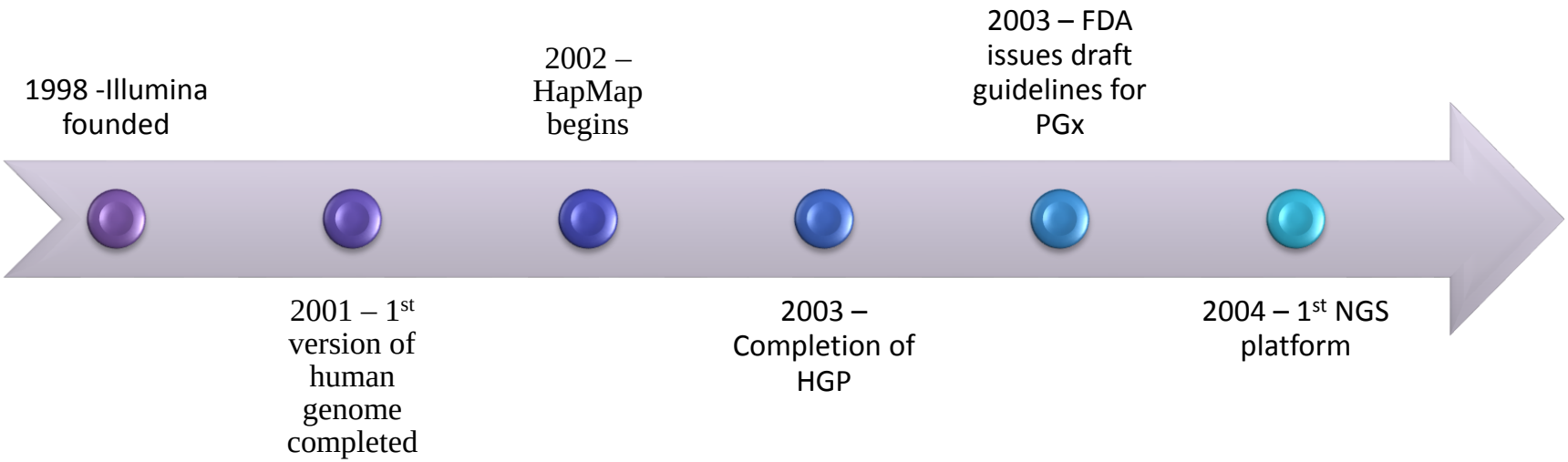
Drug 1

Drug 2





Genome-wide Array Era

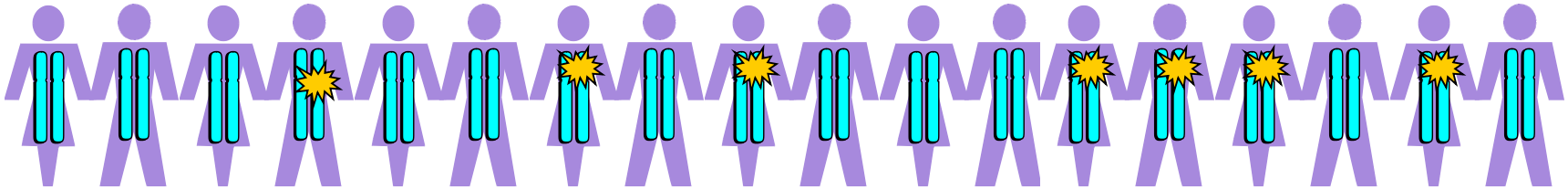


Genome-wide Association Studies

- SNP arrays (~mid 2000s)
- mRNA arrays (~mid 1990s)
- Methylation arrays (~2008-2010)

Study of Genetic Association with Drug Response Phenotypes

Cases: Non-responders



Cases: Responders

Genetic association studies look at the frequency of genetic changes to try to determine whether specific changes are associated with a phenotype.

High Throughput Methods for Measuring DNA

- Many approaches for genotyping
  - Hybridization Methods (Affymetrix, TaqMan)
  - Primer extension (Pyrosequencing)
  - Ligation (Illumina)
- Custom Content / Design
  - GoldenGate, Infinium at Illumina
  - Disease Specific panels (PGx, Cancer, Carbo-Metabo)
- **Standard large arrays**
  - Genome-wide arrays (> 1 million SNPs)
  - Exome Arrays (rare variants)
- **Next-Generation Sequencing**

Model Systems used in PGx

- **Lymphoblastoid Cell Lines (LCLs)**
- **Liver Banks** (drugs metabolized by liver)
- **Cancer Cell Lines**
- Yeast Models
- Drosophila Models
- **Animal Models**
  - Mouse, Zebrafish
- Cultured neuronal cells derived from olfactory neuroepithelium (CNON)
- Patient-derived Induced pluripotent stem **(iPS)** cells

PGx Study Designs

- **Clinical Trials**
  - Use of control group to determine PGx effect
  - **Extensive clinical outcomes** (efficacy and toxicity)
  - **Very expensive**
  - **Limited sample size**
- **Observational Studies**
  - Often **retrospective** studies
  - **Confounding** of other co-medications
  - Lack of **detailed** clinical and drug data
  - Less expensive
  - **Larger sample size**

Challenges of PGx Studies as compared to Genetic Epidemiology Studies

- Often don't have **"control" group**
- Often have limited sample size and no replication study
  - **Functional studies** for validation

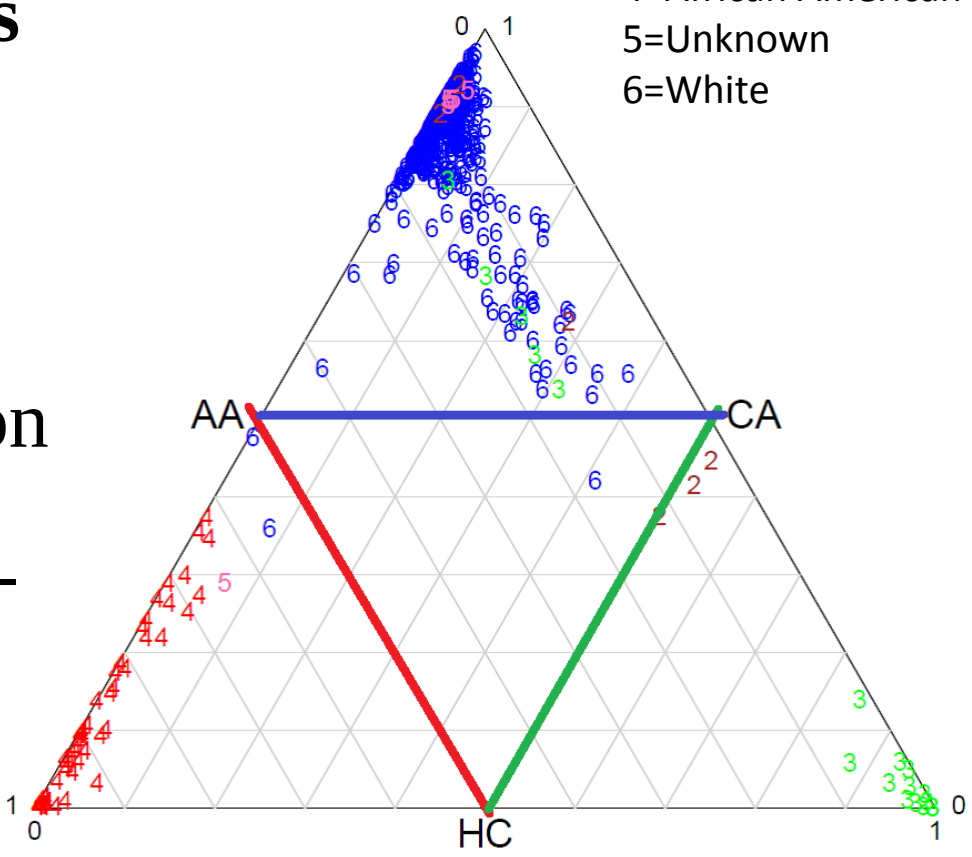
    Caveat: FUNCTIONAL $\not\rightarrow$ PATHOGENIC
- No **pedigree** information
- Toxicity and response **information limited** (except in a clinical trial setting)
- **Confounding issues** with other co-medications and comorbidities

Population Stratification and PGx

- Many known **PGx markers allele frequencies vary across racial populations**.
  - *TPMT, NAT2, GSTs, SULT1A1*

- Assessment of population substructure can be completed with genome-wide SNP data

Race coding:
1=NA (missing race)
2=Native American
3=Asian
4=African American
5=Unknown
6=White



EXAMPLE I:

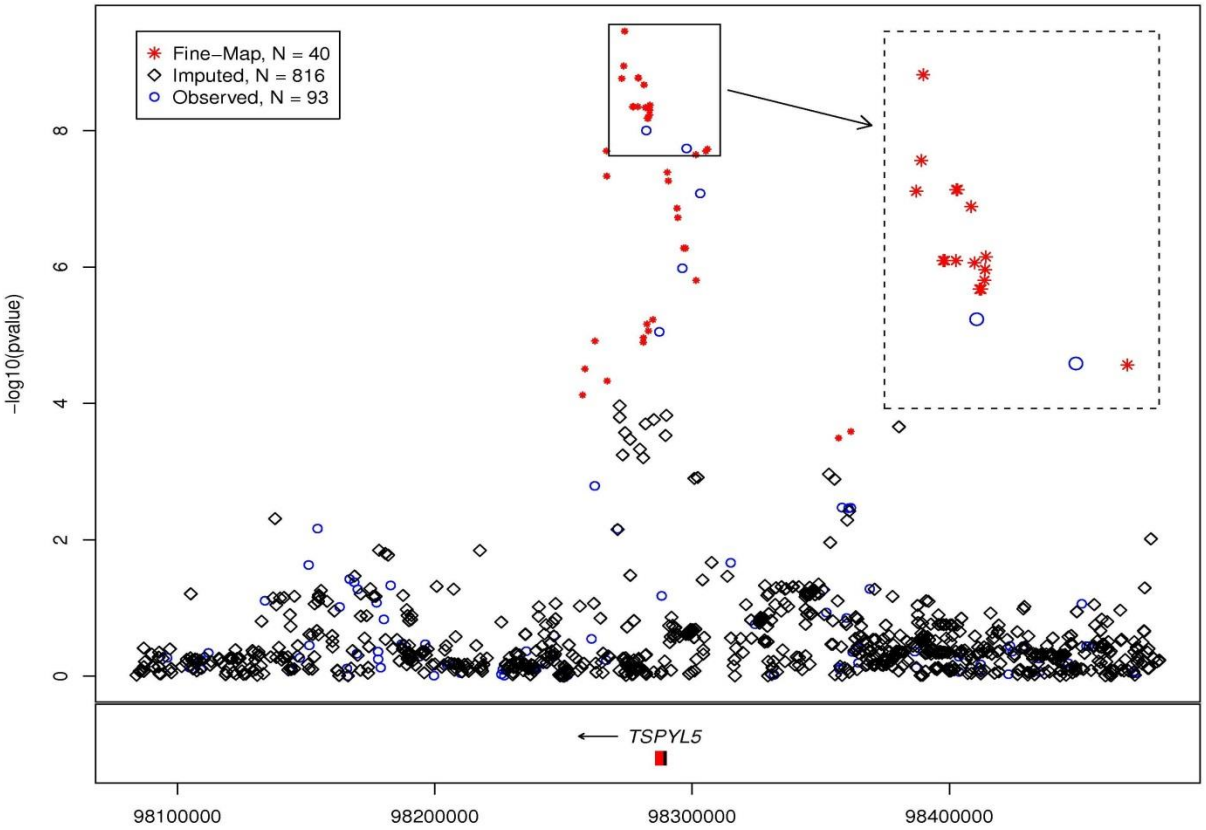
GENOME WIDE ASSOCIATION STUDY

Breast cancer treatment

- **Aromatase Inhibitors (AI)** are commonly used in treatment of **breast cancer**. However, response varies between women.
- Goal of study is to determine **genetic predictors of response** to **AI** treatment.
- N = 835 women genotyped on **Illumina 610** Array
- To determine genetic markers associated with:
  - **Baseline hormone levels**
  - Δ hormone levels following AI treatment
  - Blood drug levels following AI treatment
  - Δ in breast density, Δ BMD

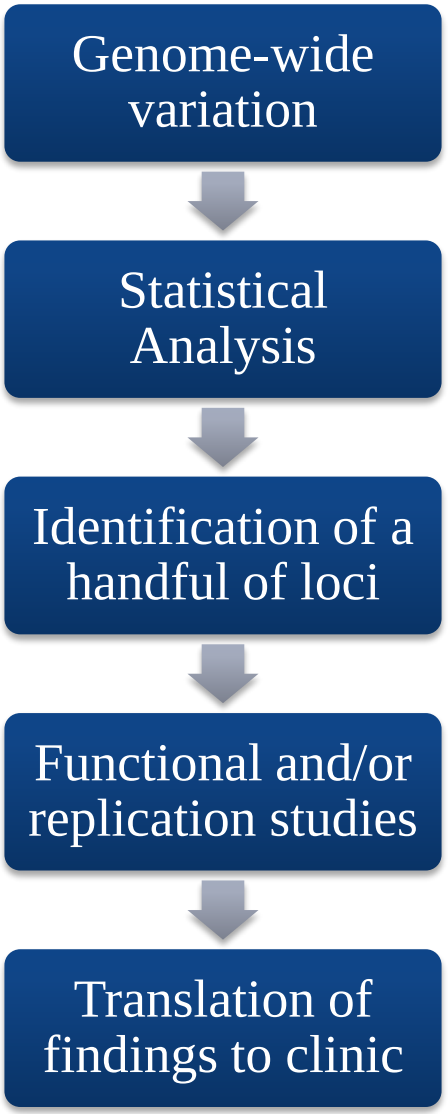
Refining Region with Imputation

- Genotype Imputation with 1KGP to refine region/signal
- Confirm imputation results with genotyping



TSPYL5 and Aromatase

- No previous evidence relating TSYPL5 to estrogen levels
- Functional studies confirm this observed statistical association of TSYPL5 with E2 (estradiol).
- These results represent a new mechanism for the control of aromatase and, thus E2, in postmenopausal women

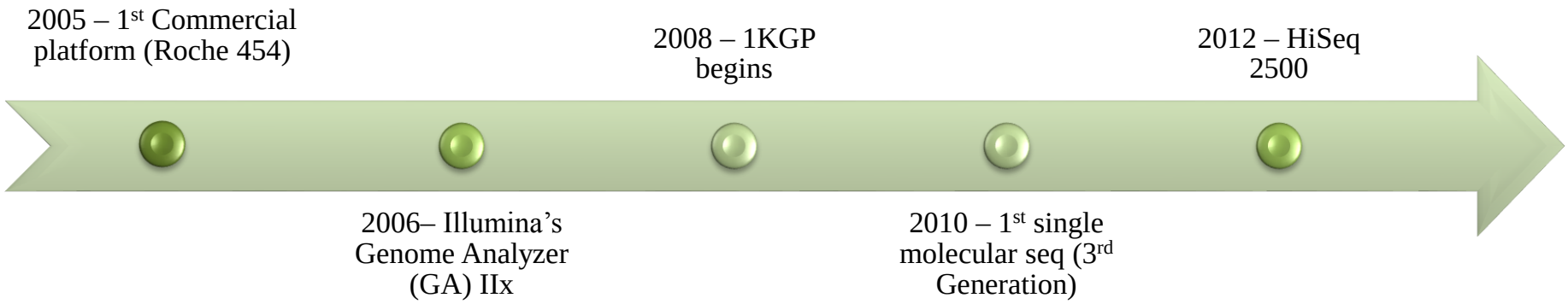


Mol Endocrinol. 2013 Apr;27(4):657-70. doi: 10.1210/me.2012-1397. Epub 2013 Mar 21.

TSPYL5 SNPs: association with plasma estradiol concentrations and aromatase expression.

Liu M[1], Ingle JN, Fridley BL, Buzdar AU, Robson ME, Kubo M, Wang L, Batzler A, Jenkins GD, Pietrzak TL, Carlson EE, Goetz MP, Northfelt DW, Perez EA, Williard CV, Schaid DJ, Nakamura Y, Weinshilboum RM.

Next-Generation Sequencing Era

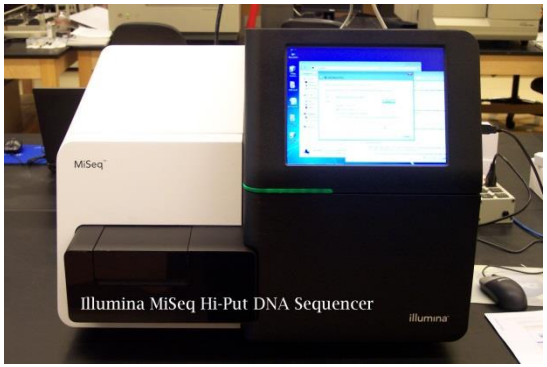


Next-Gen Sequencing (NGS)

- DNA (Exome & WGS)
- RNA-seq
- Bisulfite or RRBS (methylation)
- Custom panels (MiSeq)
- Exome arrays

NGS Technologies

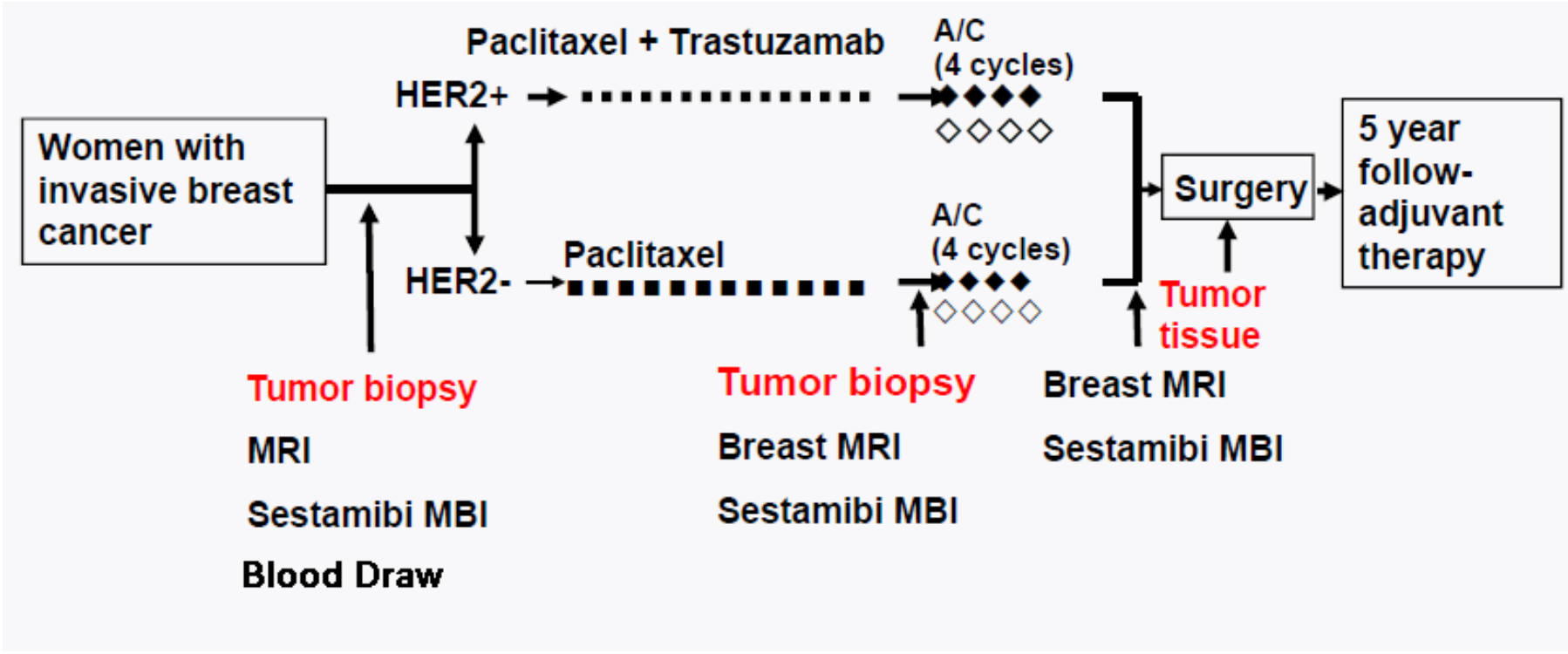
- Illumina (Solexa) HiSeq 2000 (2500) & MiSeq, Life Technologies SOLiD, PacBio, Ion Torrent PGM, Roche 454, ... , and many more to come
  - **No one-size-fits-all solution**
  - Each has pros and cons



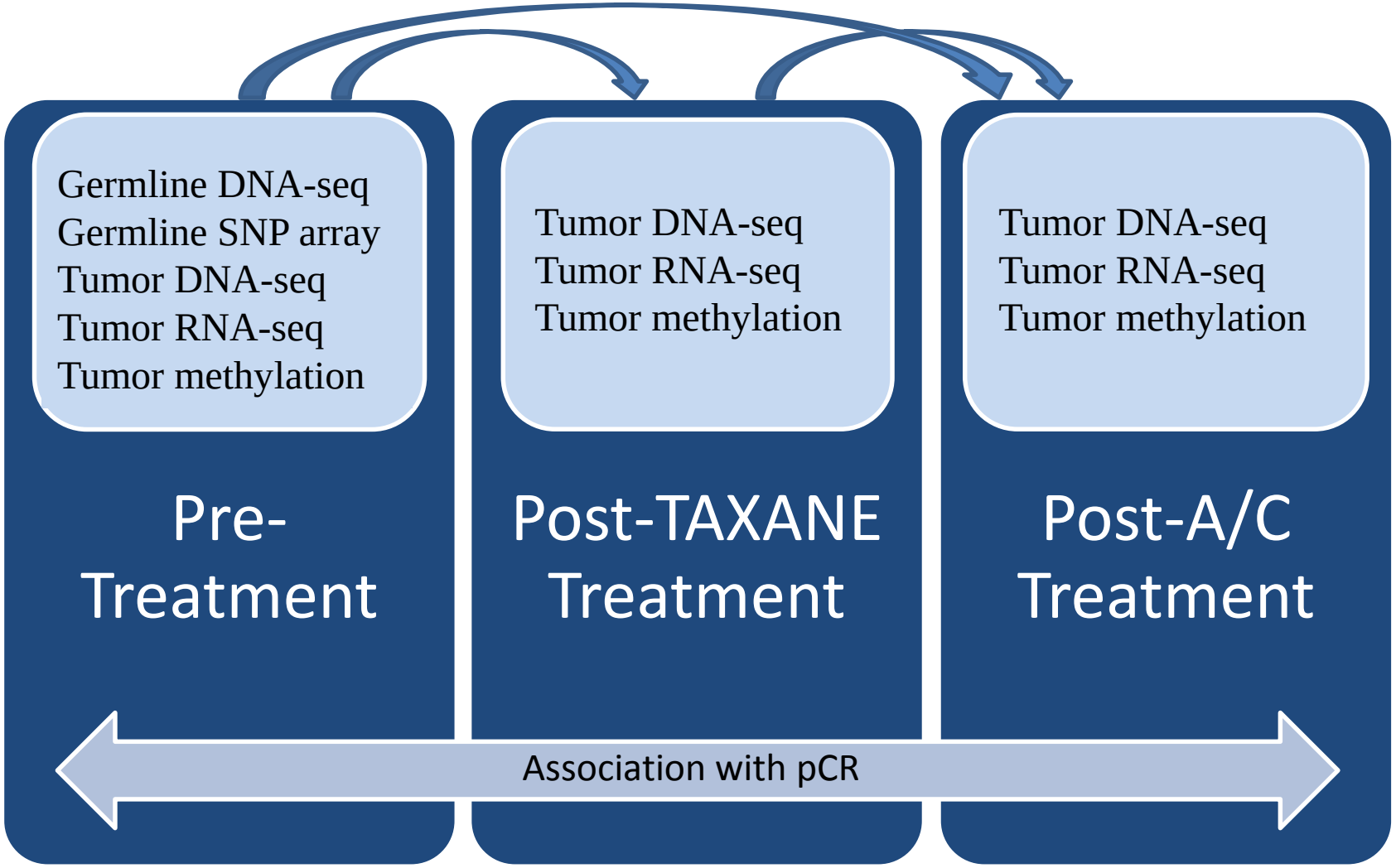
EXAMPLE II:

NGS PHARMACOGENOMIC STUDY

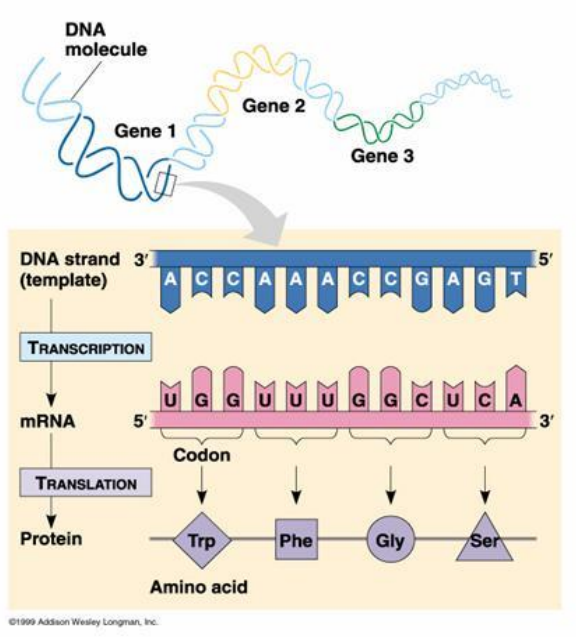
Mayo Clinic's BEAUTY Study Design

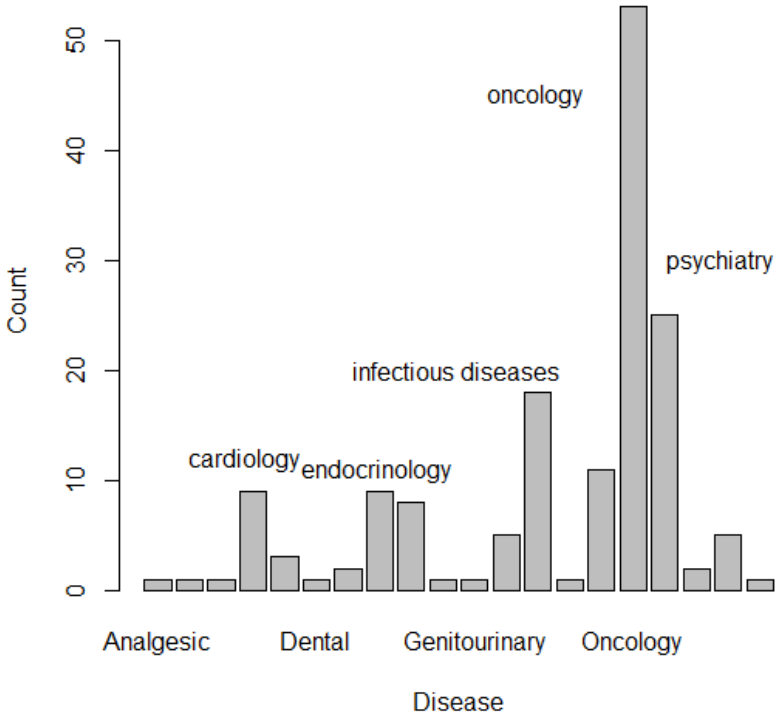


Sequencing for BEAUTY



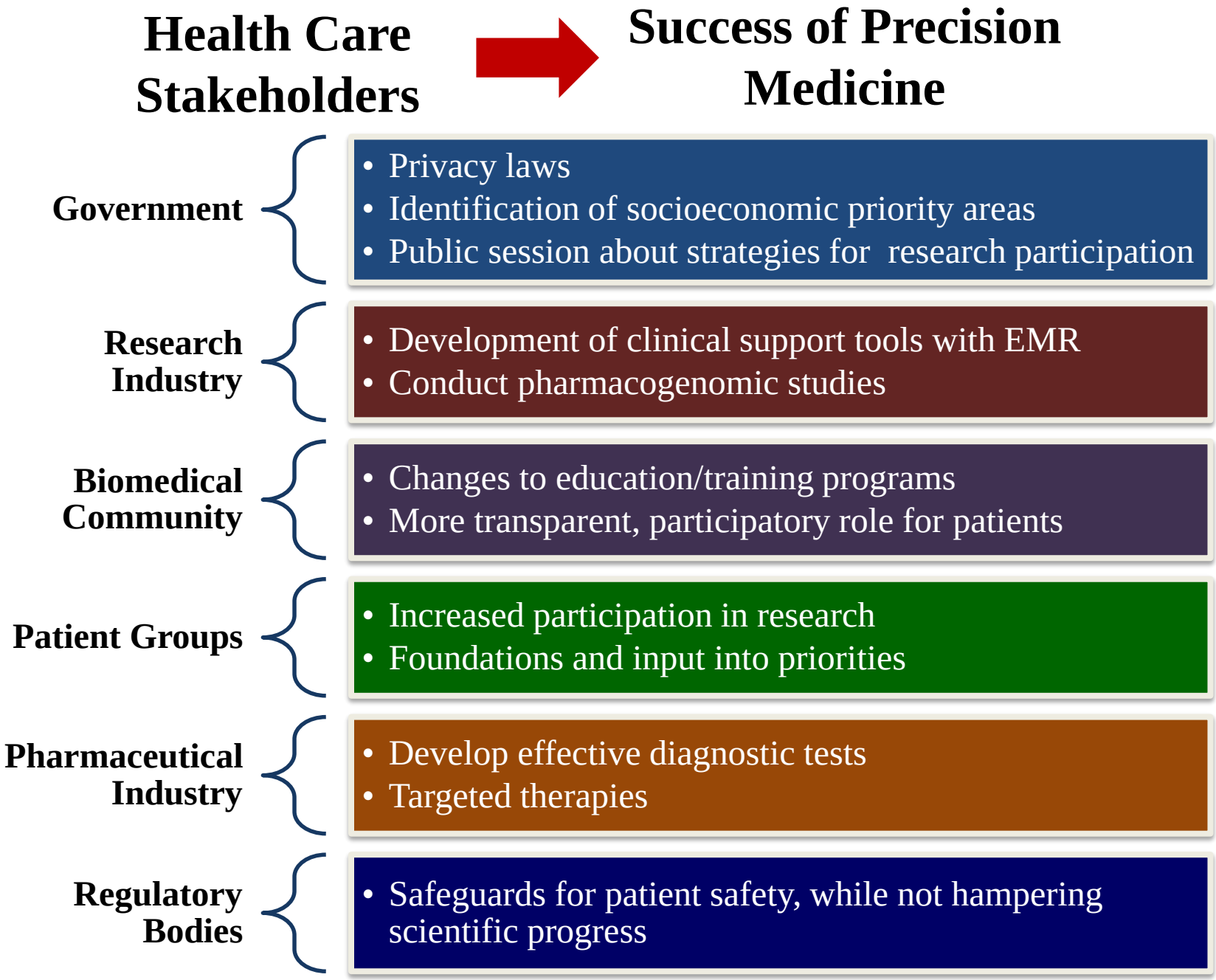
Future & Precision Medicine Era





Future of Precision Medicine

- Single Molecule Sequencing
- **Translation to clinical practice**
- **Data / Information Integration**
- Additional FDA recommendations (currently 158 FDA approved PGx labeling)
- Genomic based clinical trials
- **Using and incorporation results into EMRs (later talk)**



Mirnezami R et al. N Engl J Med 2012;366:489-491.

Examples of from Bench to Bedside

- **FDA warning labels** regarding PGx biomarkers:
  - **CV: Warfarin** *(CYP2C9, VKORC1)*
  - **Breast Cancer**: Tamoxifen (CYP2D6)
  - **Childhood ALL**: Mercaptopurine & Thioguanine (*TPMT)*

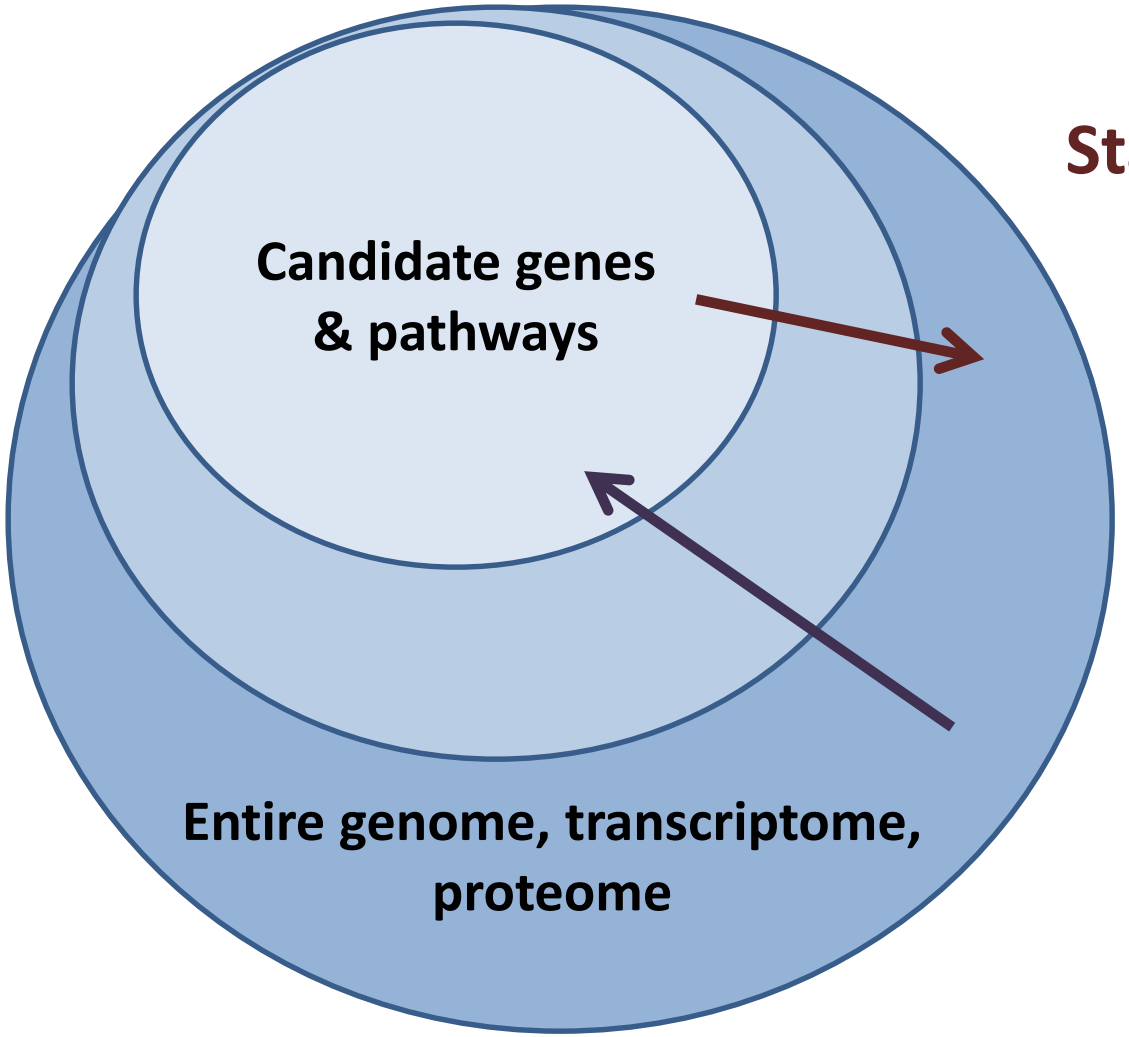
- **Targeted Cancer Treatments**:
  - *Imatinib* was designed to inhibit an altered enzyme produced by a **gene fusion** in **chronic myelogenous leukemia.**
  - **Breast cancer** drug *trastuzumab* works only for women who have **HER-2 positive** tumors.
  - **Lung cancer** patients with mutations in ***EGFR*** respond to the drugs *gefitinib* and *erlotinib*.
  - **Colon cancer** patients with a mutation in ***KRAS*** have little benefit from drugs *cetuximab* and *panitumumab*.

Finding THE remaining needle in a haystack with multiple types of hay

Systems Biology

- Biological systems are **complex** and therefore data has been collected on **multiple scales**
  - genome, transciptome, epigenome, proteome, metabolome and phenome.
- However, is often the case that each data set is analyzed in solitude
- **Multi-scale integration** of data types to answer fundamental and practical **questions in complex disease and traits** is a **significant challenge**

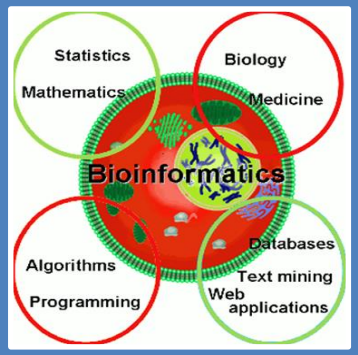
Computation Approaches



Bottom Up:
Start with a candidate and build up

Top Down:
Start with genome and filter down

Bioinformatics-Statistics “continuum”



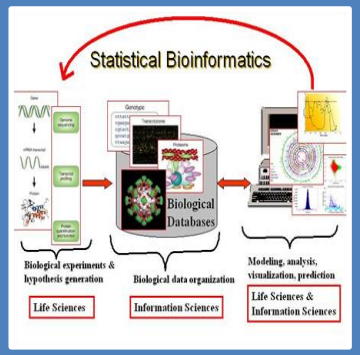
Processing of data via computers

Biological knowledge/annotation

Algorithms to determine function, structure

Informatics

New algorithms for processing next-generation sequence data

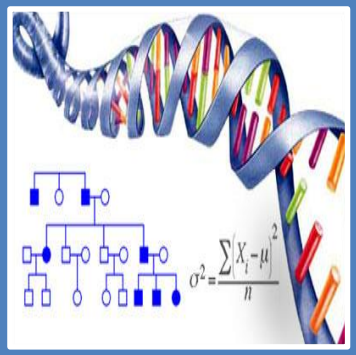


Data mining

Clustering/Profile

Network and Interactions

Gene set and pathway analysis



Experimental Design

Association Analysis

Differential Analysis

GWAS & Haplotype

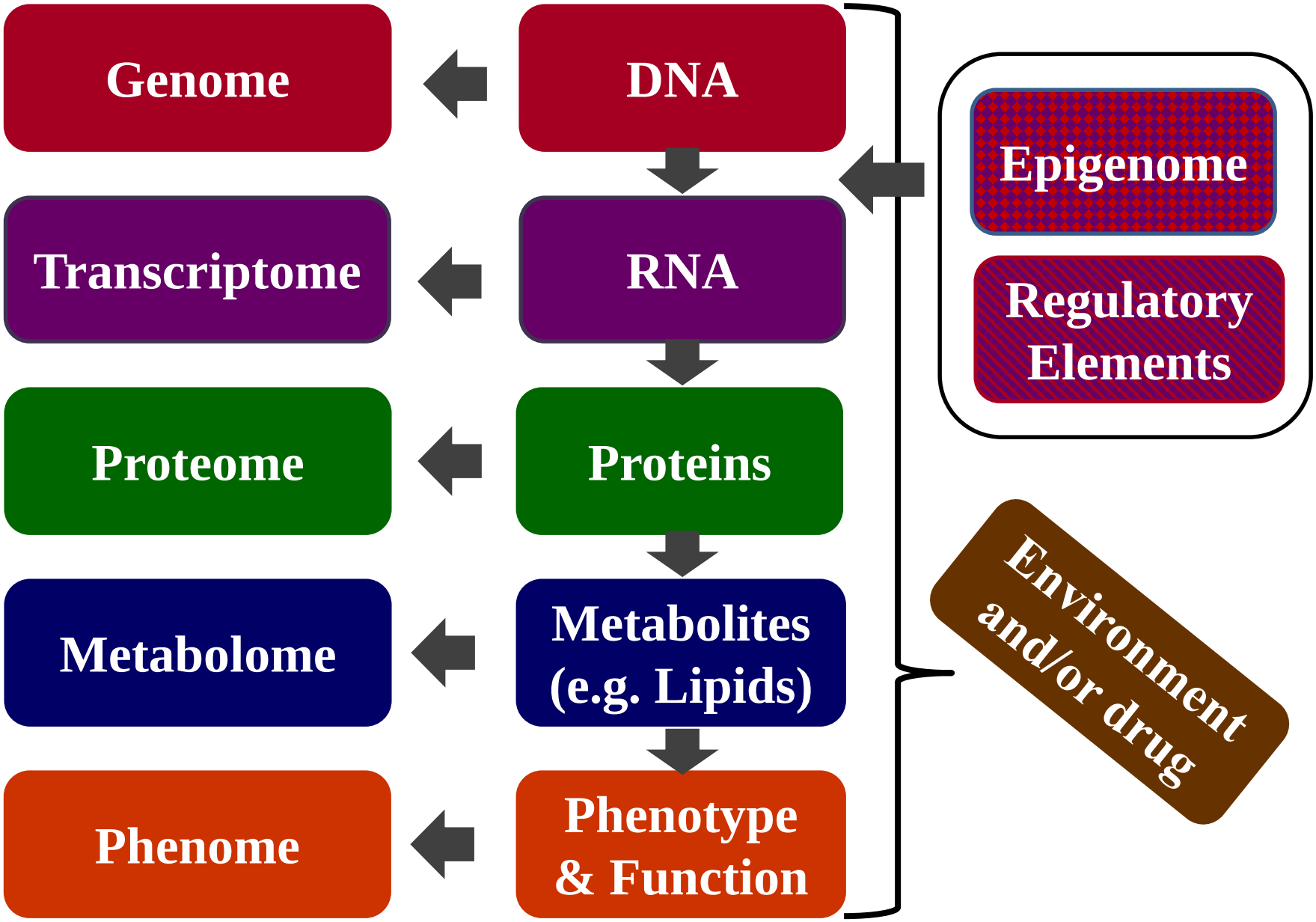
Modeling & Prediction

Pedigree Studies (Linkage)

New statistical methods

Bioinformatics → **Statistical Genomics**

Integrative 'Omics



Public Data and Information

NCBI Resources How To

NCBI
National Center for Biotechnology Information

All Databa

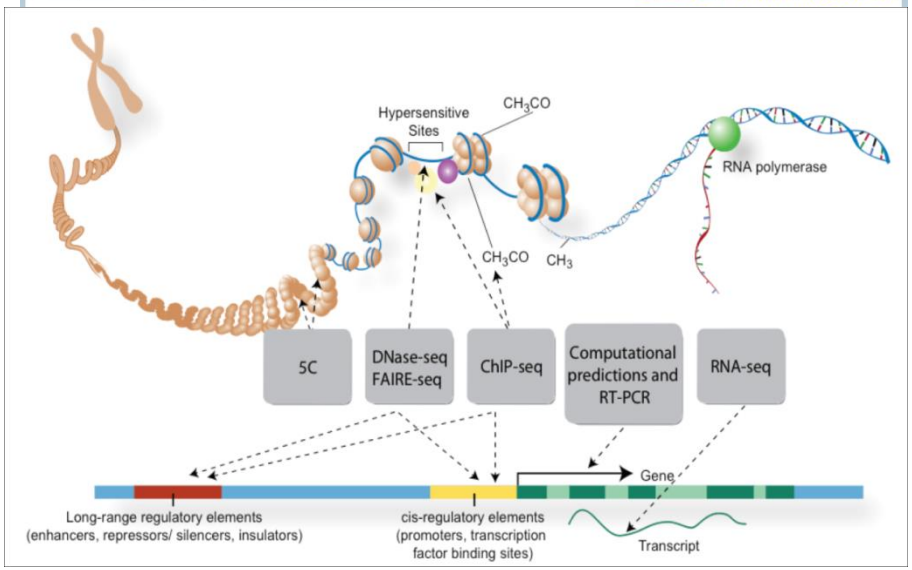
- NCBI Home
- Resource List (A-Z)
- All Resources
- Chemicals & Bioassays
- Data & Software
- DNA & RNA
- Domains & Structures
- Genes & Expression
- Genetics & Medicine
- Genomes & Maps
- Homology
- Literature
- Proteins
- Sequence Analysis
- Taxonomy
- Training & Tutorials
- Variation

GEO
Gene Expression Omnibus

nature | ENCODE

Search ENCODE Go

Home | Research | Threads | Additional Research | News and Comment | About | Sponsor



1000 Genomes

A Deep Catalog of Human Genetic Variation

The Cancer Genome Atlas

Understanding genomics to improve cancer care

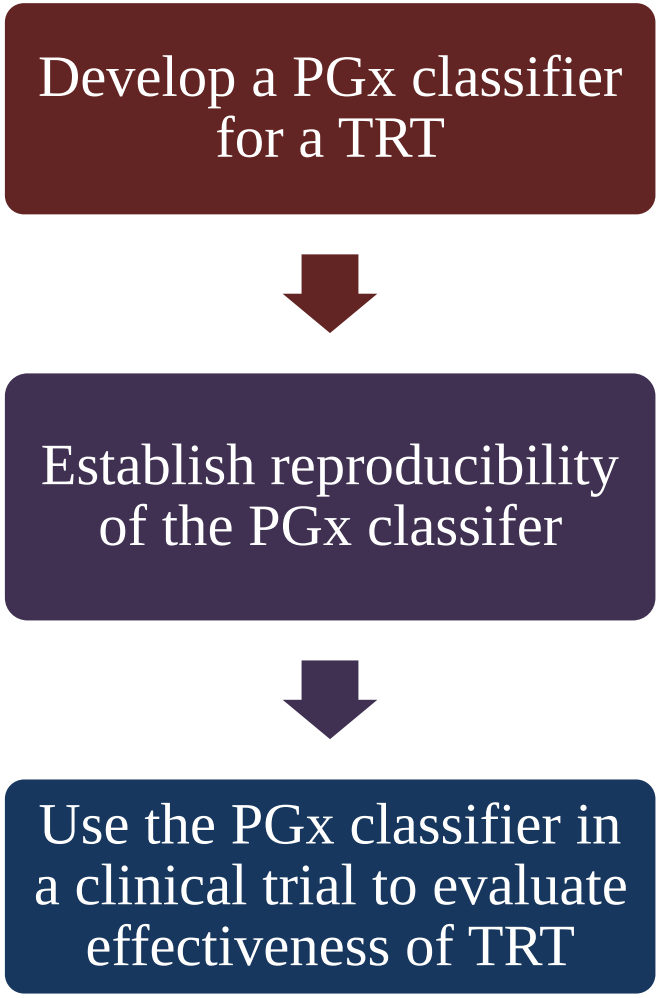
dbGaP

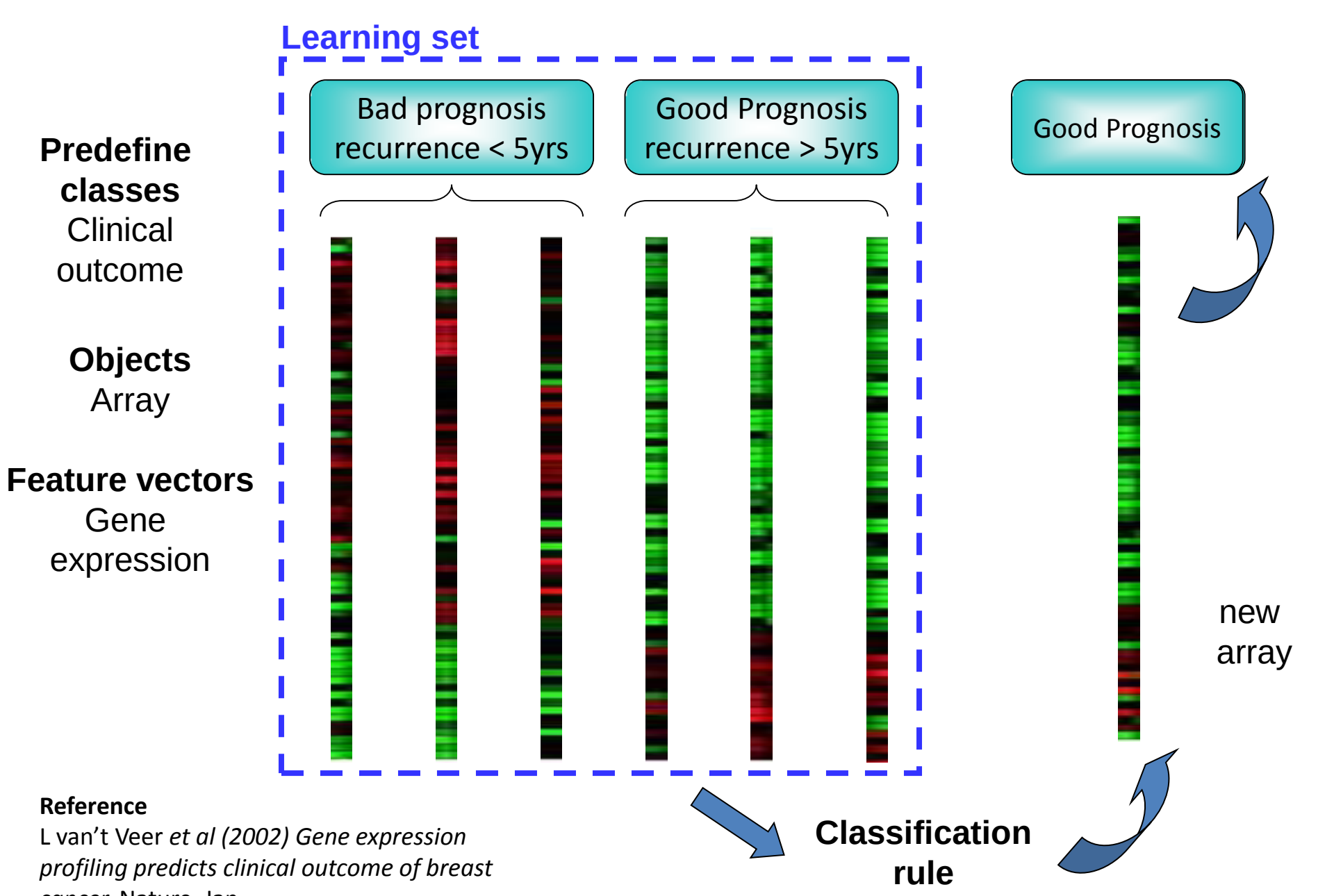
GENOTYPES and PHENOTYPES

Pharmacogenomic (PGx) Classifiers

Benefits:

- Enables patients to be treated with drugs that actually work for them
- **Avoids false negative trials** for heterogeneous populations
- Avoids erroneous generalizations of conclusions from positive trials





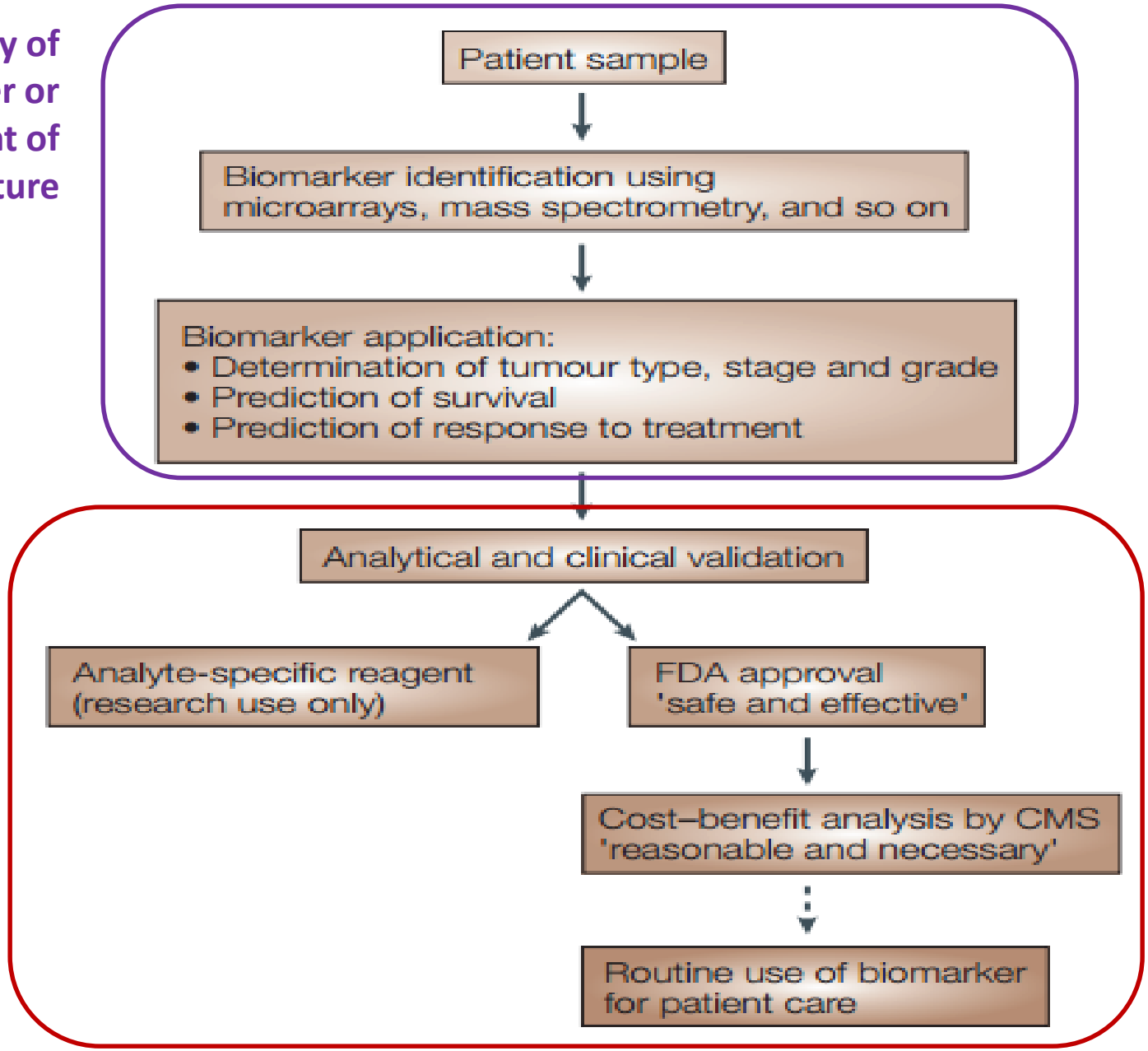
Reference

L van't Veer *et al (2002) Gene expression profiling predicts clinical outcome of breast cancer.* Nature, Jan.

Process for development of a biomarker

Discovery of biomarker or development of signature

Validation (assay and signature) and FDA approval

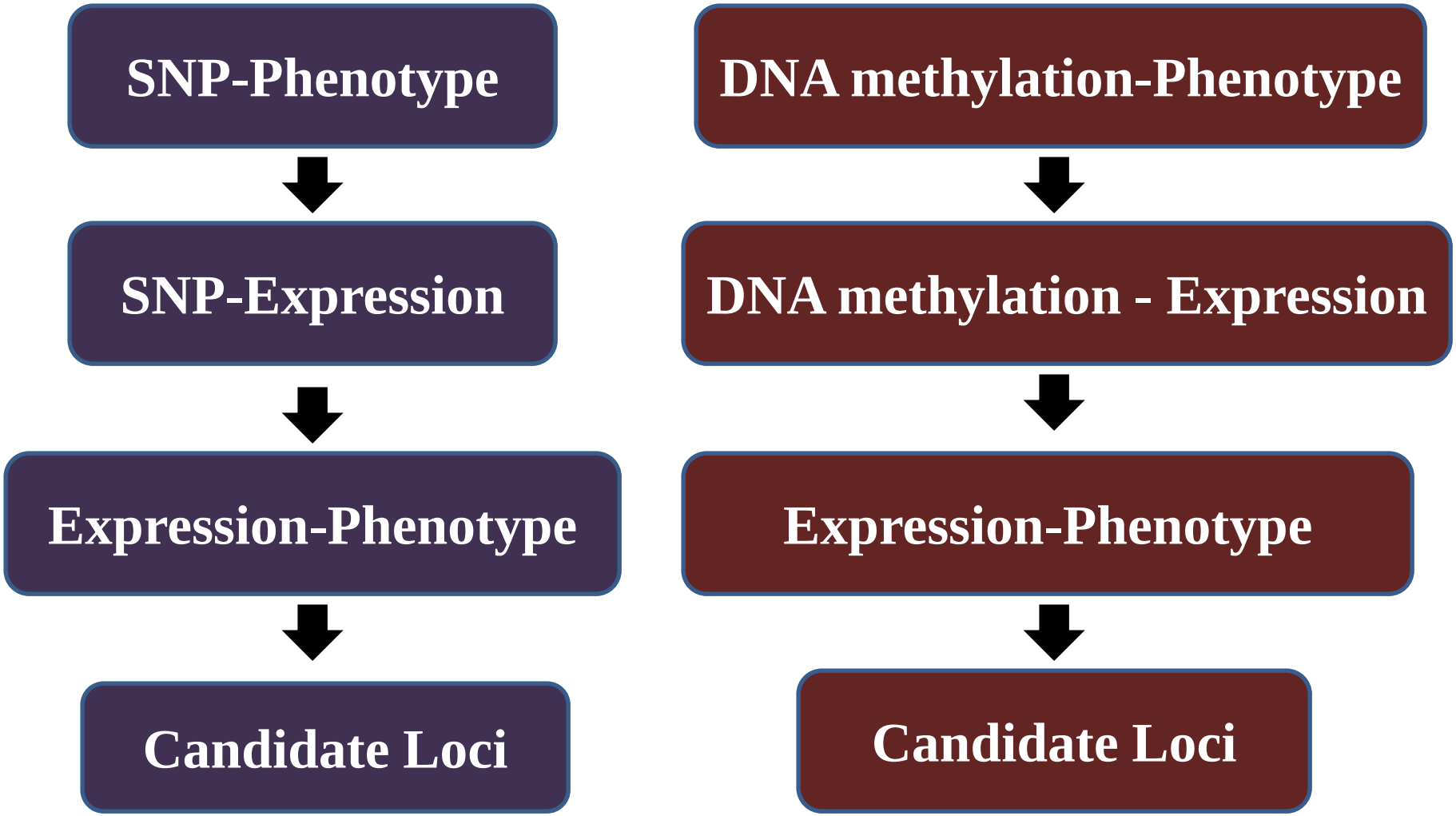


Ludwig & Weinstein. (2004) *Nature Reviews Cancer.*

EXAMPLE III:

DATA INTEGRATION IN PHARMACOGENOMICS STUDIES

Step-Wise Data Integration



Radiation PGx

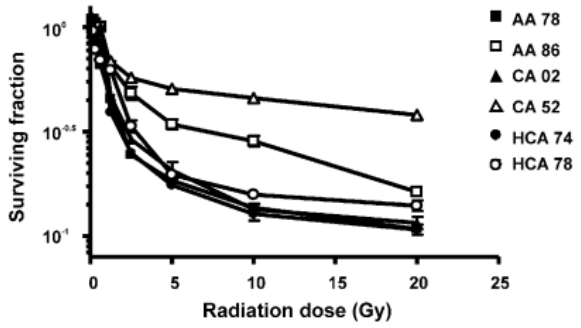
GENOME RESEARCH
CSHL Press | Journal Home | Subscriptions | eTOC Alerts | BioSupplyNet

Genome Res. 2010 November; 20(11): 1482–1492. PMCID: PMC2963812
doi: 10.1101/gr.107672.110

Radiation pharmacogenomics: A genome-wide association approach to identify radiation response biomarkers using human lymphoblastoid cell lines

Nifang Niu,[1,3] Yuxin Qin,[1,3] Brooke L. Fridley,[2] Junmei Hou,[1] Krishna R. Kalari,[1,2] Minjia Zhu,[1,4] Tse-Yu Wu,[1] Gregory D. Jenkins,[2] Anthony Batzler,[2] and Liewei Wang[1,5]

- Integrated analysis for response to radiation in **277 LCLs**
  - Area under radiation dose response curve (AUC)
  - Illumina 550K, 510S & Affymetrix 6.0 arrays
  - Affymetrix U133plus2.0 mRNA array
- **Functional validation** using siRNA knockdown in multiple tumor cell lines
  - *C13orf34*,*MAD2L1*, *PLK4*, *TPD52*, and *DEPDC1B* each significantly altered radiation sensitivity



175 SNPs associated with AUC ($p < 10^{-3}$)

2432 SNP-expression associations ($p < 10^{-4}$)

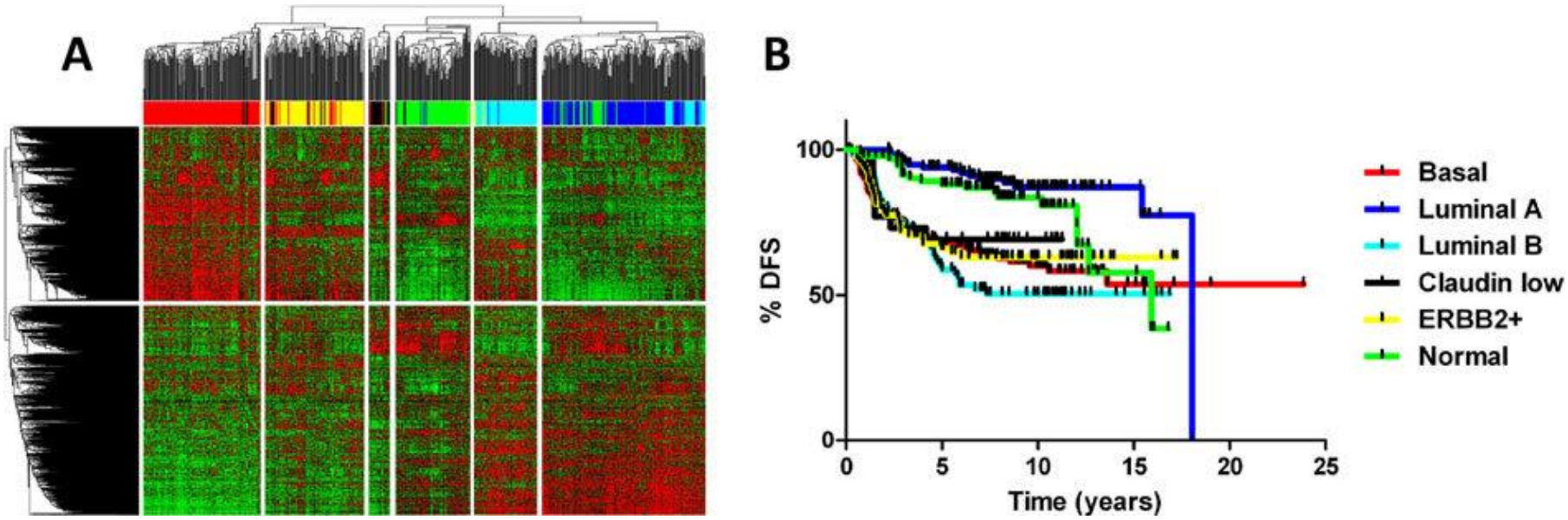
47 expression probe sets (39 genes) associated with AUC ($p < 10^{-3}$)

EXAMPLE IV:

MOLECULAR PHENOTYPING AND INTEGRATIVE CLUSTERING ANALYSIS

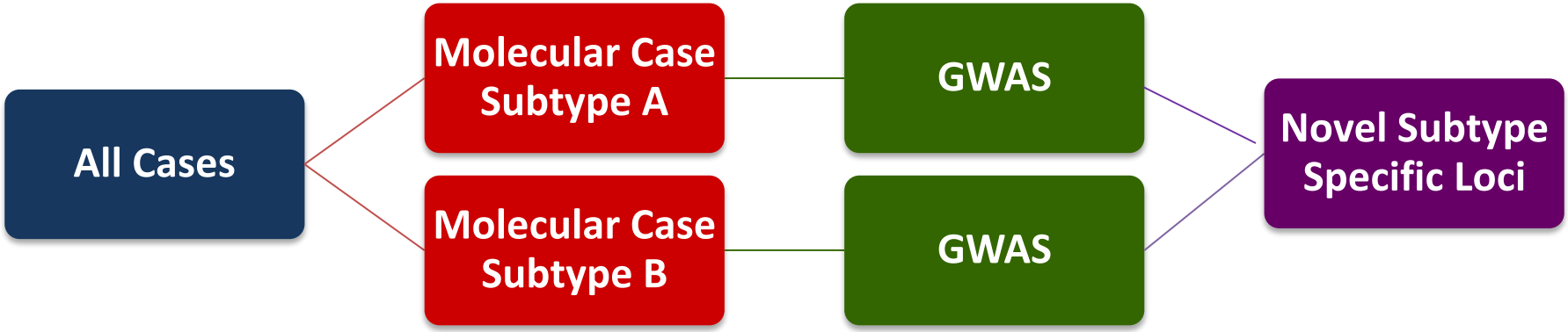
Molecular Based Phenotype Definition

- **Disease Heterogeneity**: Determine disease subtypes or case definition
- **Clinical Heterogeneity**: Determine profiles that classify into subtypes with different prognosis or treatment response



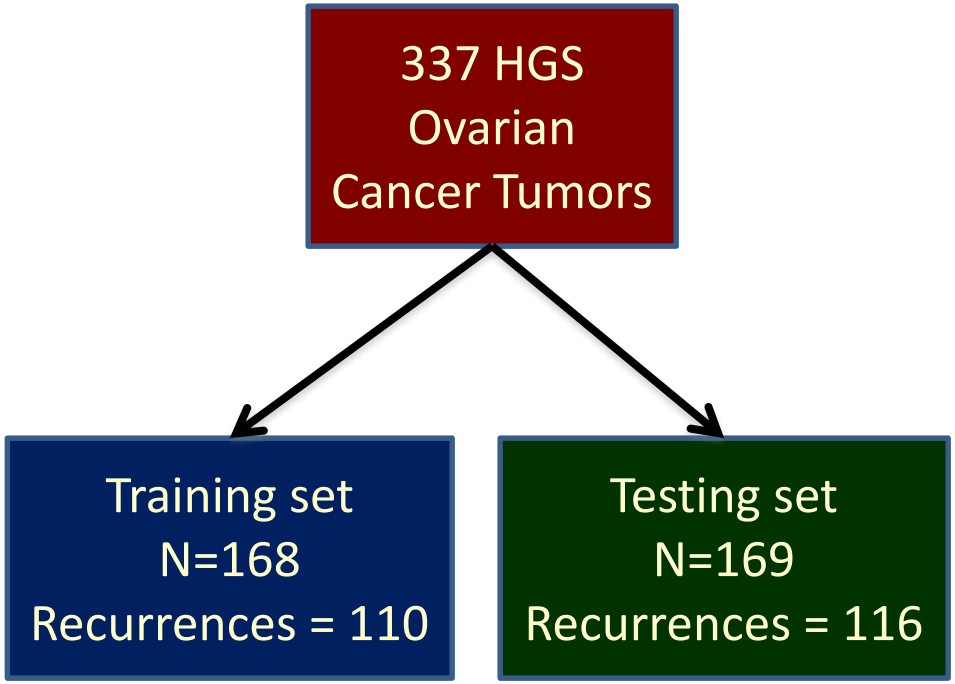
Hallett, et al. (2012) A gene signature for predicting outcome in patients with basal-like breast cancer. *Scientific Reports*

Molecular Phenotype Based GWAS



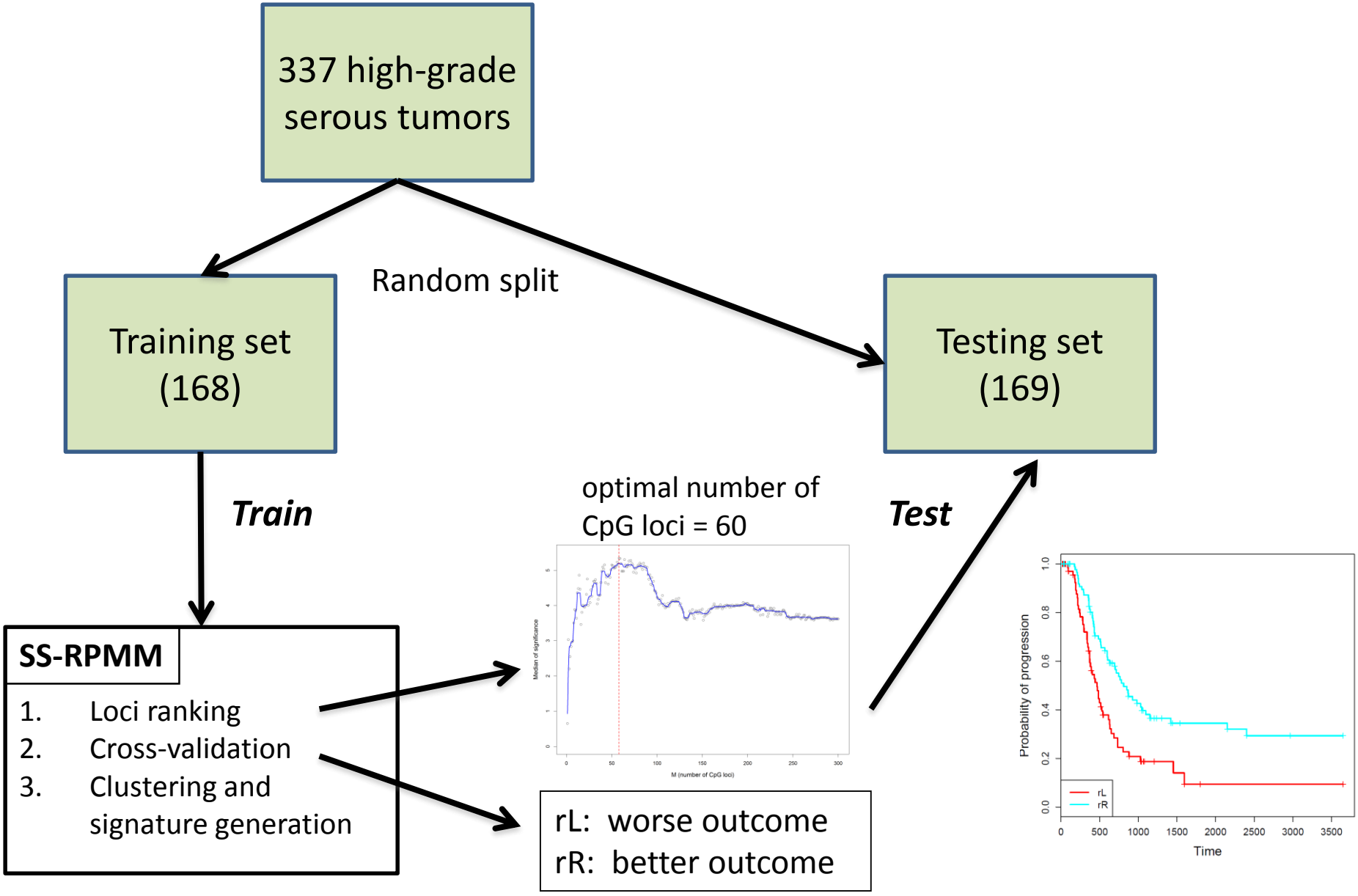
- **Molecular Subtype GWAS:**
  - For risk with existing controls
  - For clinical outcome
  - For quantitative trait

Subtypes of Ovarian Cancer



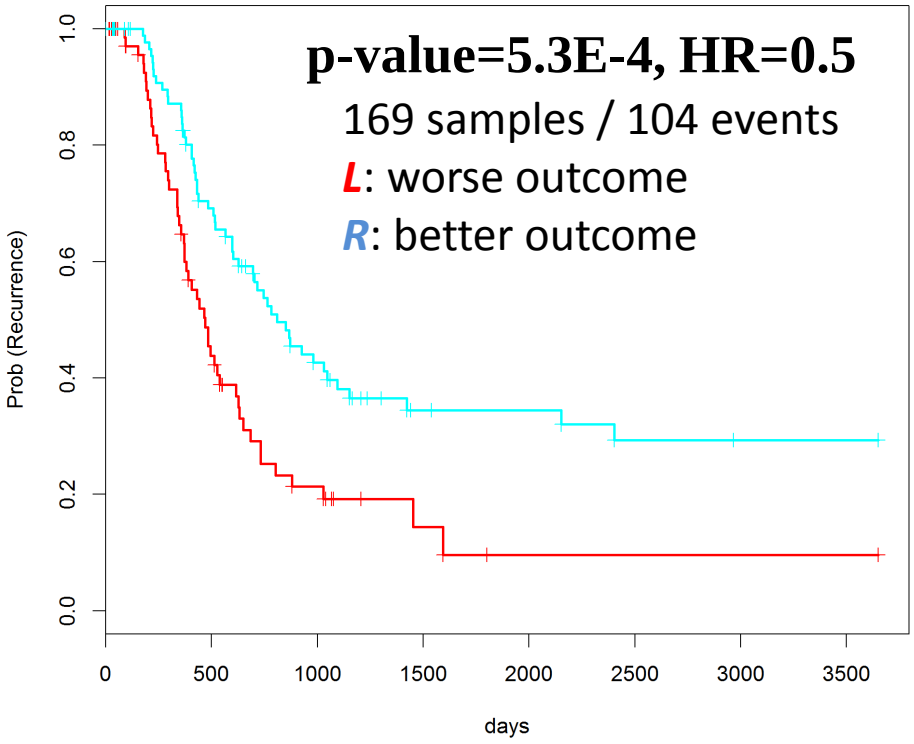
- Restricted to High Grade Serous (HGS) histology
- Pre-chemo tumor sample
- 450K Illumina Methylation Array
- Similar stage and recurrence status between testing and training data sets

Wang, et al (2014). Tumor hypomethylation at 6p21.3 associates with longer time to recurrence of high-grade serous epithelial ovarian cancer. *Cancer Research*

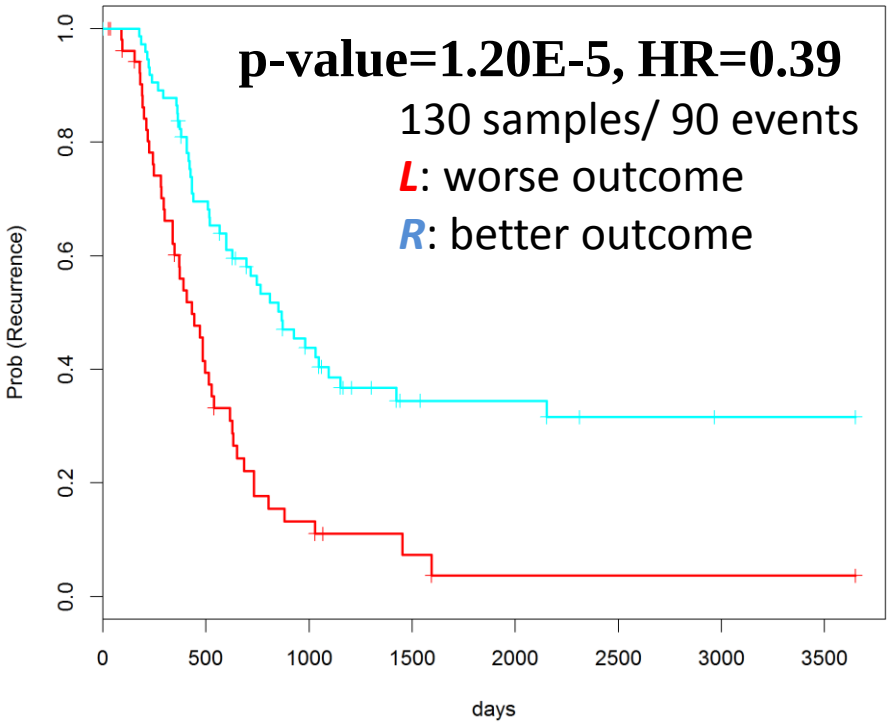


Analysis workflow of semi-supervised clustering used in this study.

All the HGS (testing) samples



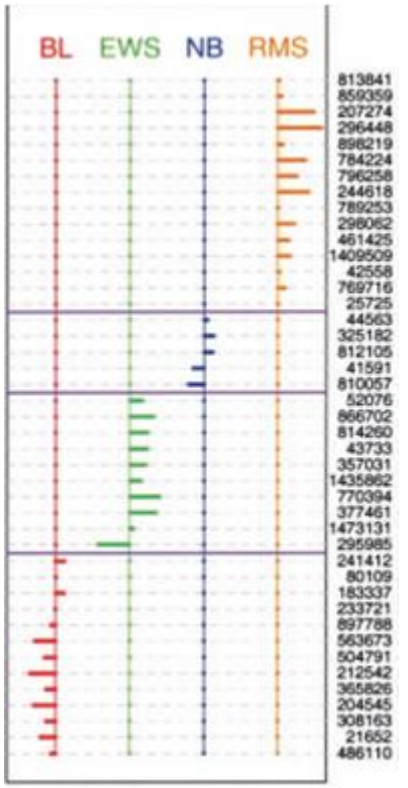
HGS testing samples with platinum and taxane trt



Kaplan Meier plot of association between groups (R or L) and recurrence time

Integration of Gene Expression

- **Goal:** In the two HGS ovarian cancer methylation subtypes, what genes are differentially expressed (i.e., what genes can separate of these DNAm sub-classes)?



One Solution:

- **PAM:** Shrinks each class centroid towards the overall centroid. The shrinkage factor is determined by CV.
- The shrinkage denoises large effects while setting small ones to zero (i.e., selection of key genes)

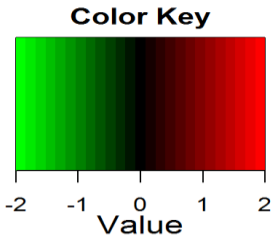
Diagnosis of multiple cancer types by shrunken centroids of gene expression

Robert Tibshirani[†‡], Trevor Hastie[§], Balasubramanian Narasimhan[§], and Gilbert Chu[¶]

Departments of [†]Health, Research and Policy, and Statistics, [§]Statistics and Health, Research and Policy, and [¶]Medicine and Biochemistry, Stanford University, Stanford, CA 94305

Communicated by Bradley Efron, Stanford University, Stanford, CA, February 19, 2002 (received for review October 10, 2001)

Gene expression differences (from PAM)

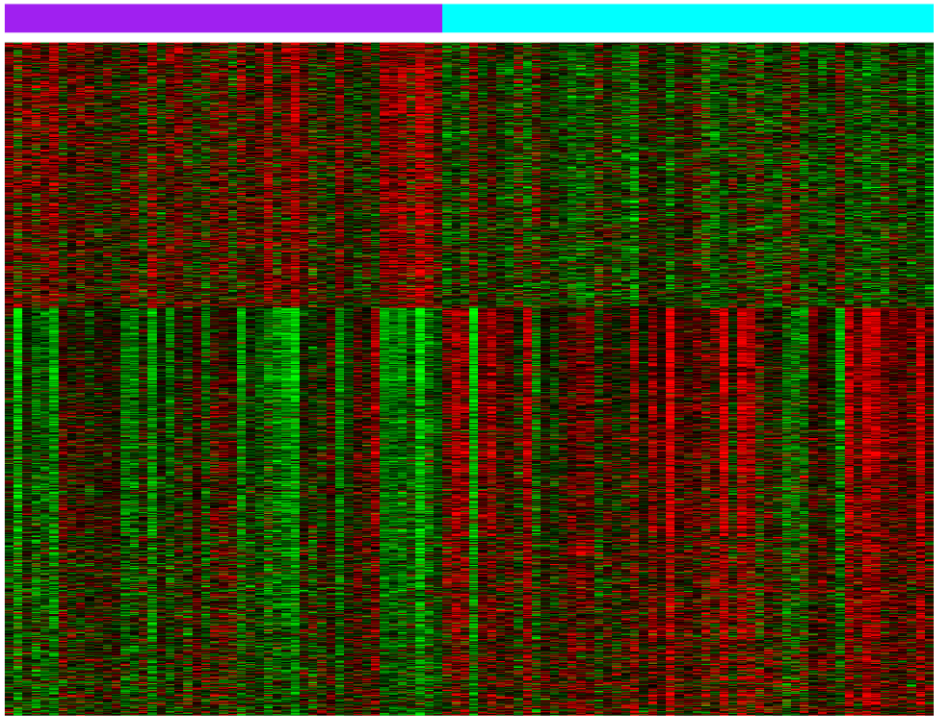


L: worse outcome (n = 48)
R: better outcome (n = 56)

- 104 of the HGS cases have Agilent gene expression data

921 probes with higher expression in L group

1413 probes with higher expression in R group



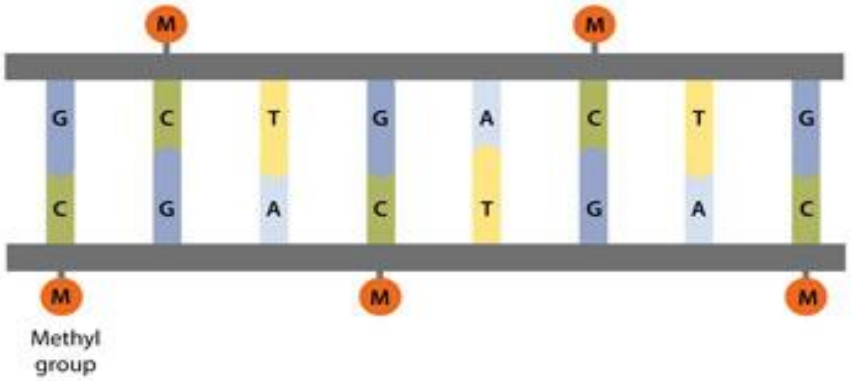
Expression heatmaps of signature genes selected by PAM analysis using shrinkage factor 1.5, which was selected based on minimum cross-validation error.

What are the genes that distinguish between the molecular subtypes?

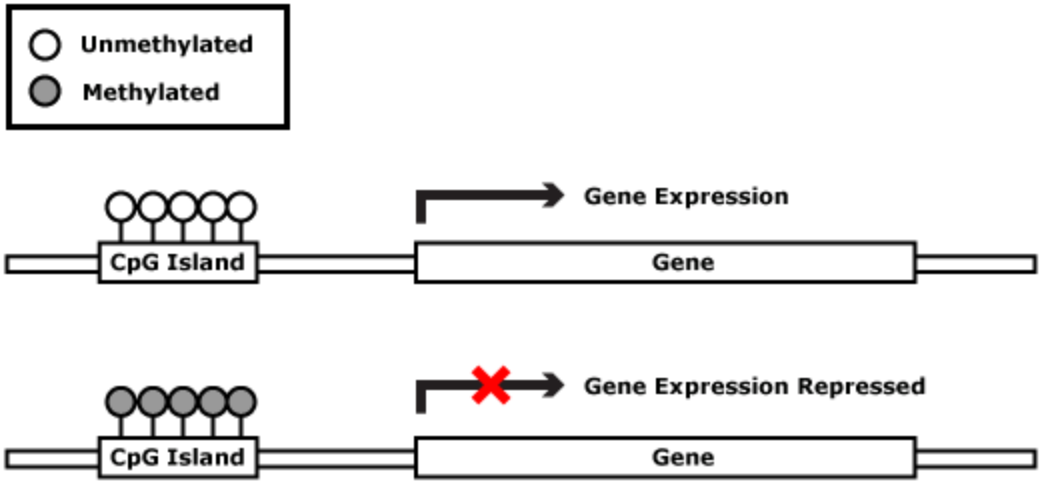
- 958 genes are over expressed in patients with better outcome
  - extremely enriched in immune related pathways, such as **Antigen Presentation Pathway** (p-value=1.6E-32), **Crosstalk between Dendritic Cells and Natural Killer Cells** (p-value=2E-24), and **Communication between Innate and Adaptive Immune Cells** (p-value=5E-24).
  - Might explain why this group is associated with better outcome (blessed by protection of boosted immune mechanism)

Epigenetics and Drug Development

- Study of heritable changes in gene expression that are not due to changes in DNA sequence.
- A methyl group may be added to cytosine to form 5-methylcytosine.
- This process is known as DNA methylation (DNAm) and only occurs in cytosines that are followed by a guanine (5' CG 3').

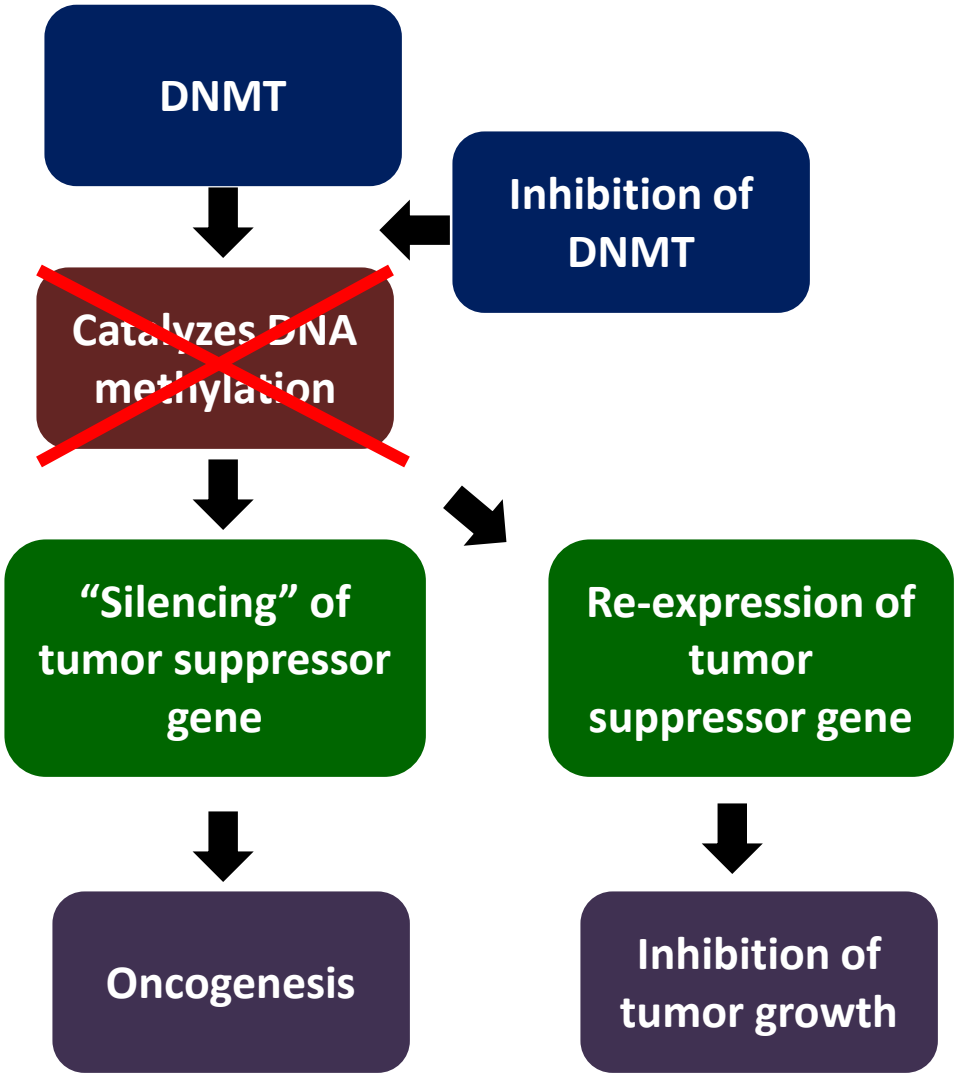


- Many genes have upstream CG-rich regions called **CpG islands**.
- DNAm of a gene's CpG island represses gene expression ("**gene silencing**").



- Different cell types have different DNAm patterns

DNA Methylation & Cancer Therapies



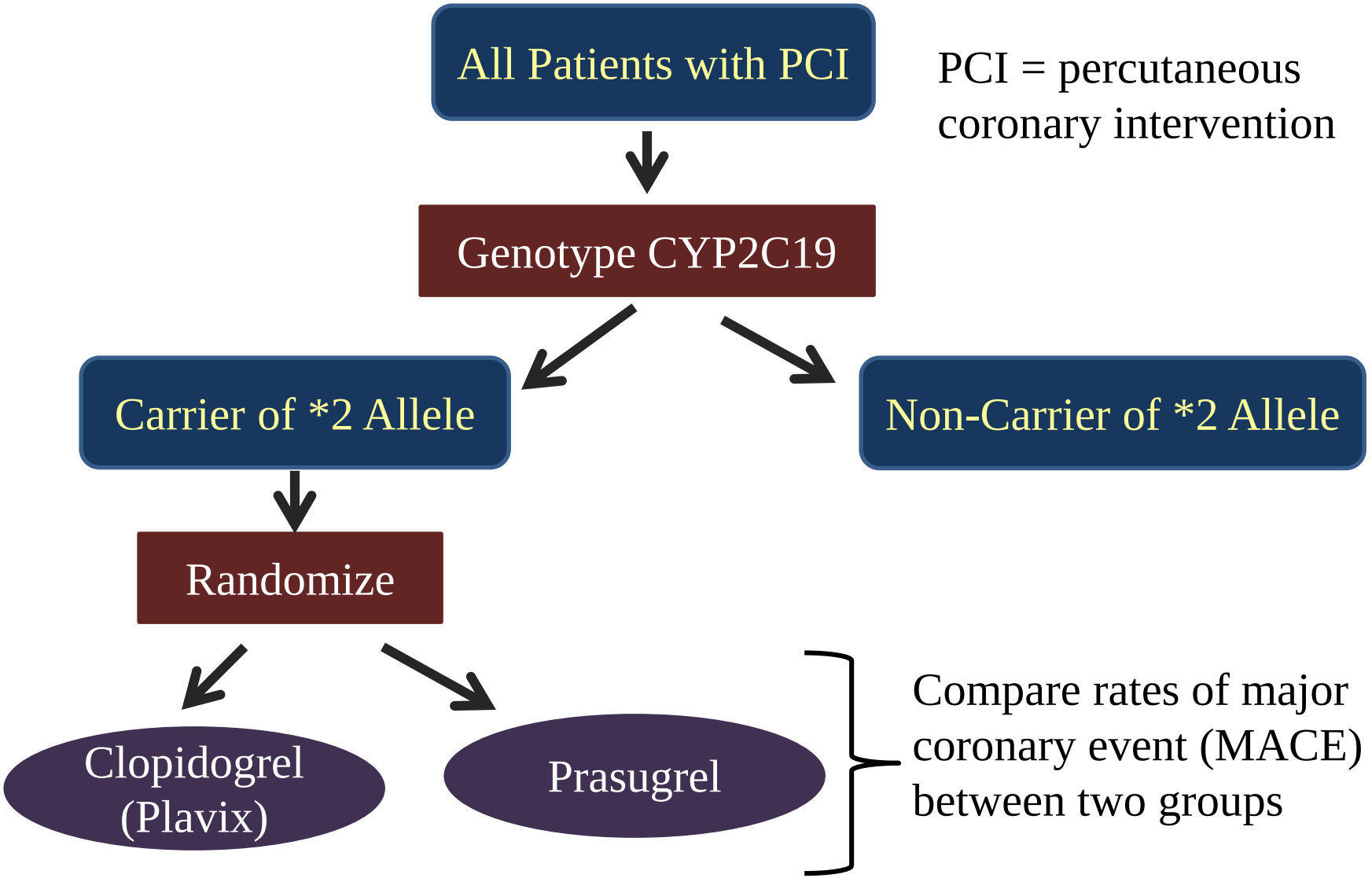
- **DNAm changes are reversible.**
- The potential to reverse DNAm and re-express critical genes **presents a therapeutic option**
  - DNMT inhibitors (5-Azacytidine)
  - Histone acetylation, Histone methylation, miRNAs
- **Pharmacoepigenomics**

Kelly, et al (2010) Epigenetic modifications as therapeutic targets. *Nature Biotechnology.* 28:1069-1078

EXAMPLE V:

CLINICAL TRIAL USING PHARMACOGENOMICS

Study Design using Pharmacogenomics



Carriers of *2 allele “poor metabolizers” of clopidogrel

Conclusion & Discussion

- **PGx studies in this era require a team science approach to go from bench to bedside.**
  - clinicians, geneticist, pharmacologist, basic scientists, pathologist, statisticians, bioinformaticians, government / FDA, hospital administrators, insurance companies, and most importantly the patients.
- Pharmacogenomic studies are continually evolving.
  - Novel **Statistical & Bioinformatics** Methods
  - **Integration** of multiple types of 'omic data and annotation information.
- **Functional (mechanistic) studies** are needed to follow-up the findings to determine the "causative" loci.

Acknowledgements

Pharmacogenomic Studies:

- Liewei Wang
- Richard Weinshilboum
- Nifang Nui
- Dan Schaid
- Greg Jenkins
- Anthony Batzler
- Matt Goetz
- Jim Ingle
- Mohan Liu
- RIKEN
- PGRN

Ovarian Study:

- Ellen Goode
- Julie Cunningham
- Mine Cicek
- Chen Wang
- Kimberly Kalli
- Melissa Larson
- Robert Vierkant
- Sebastian Armasu
- Zach Fogarty

Thank you for your attention.