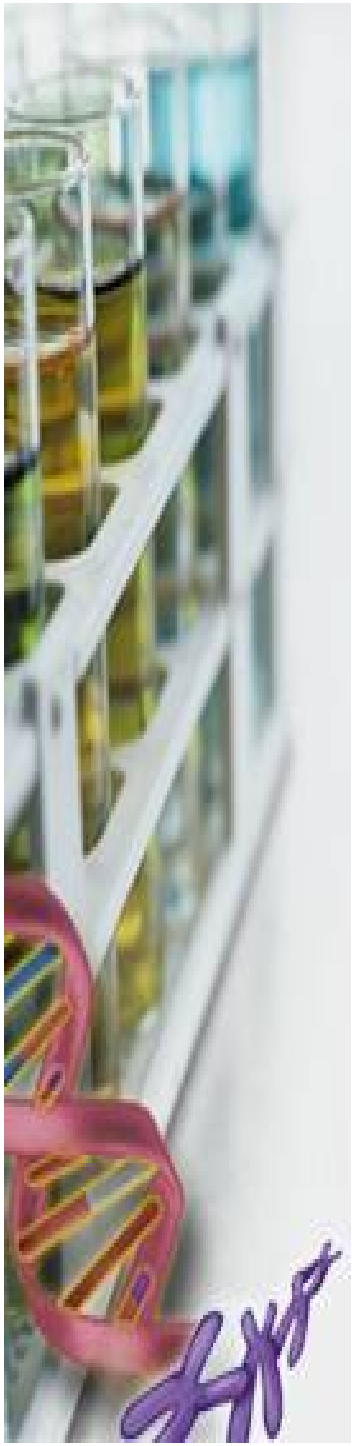


MOLECULAR PROFILING

GAIN Kick-Off and Analysis Workshop

"Genome-Wide Association Studies: A Pharmaceutical Research Perspective"

Patrice M. Milos, Ph.D.
Executive Director, Molecular Profiling
Pfizer Global Research and Development

Herndon, VA
November 29th /30th, 2006

Pfizer

Challenges In Delivering New Medicines

- Cost of Drug Discovery and Development:
  - R&D productivity challenges business model
  - *Need to deliver target portfolio with human relevance to disease*
  - *Need to identify best indications for targets*
  - *Need to enhance our clinical trial designs to make better/faster decisions*
- Market Forces:
  - We aim to deliver the medicines which are best in class
  - Delivering the best medicine for the right patient
- Increased expectations:
  - Regulators
  - Shareholders
  - Public perception

Productivity Improvement Becomes Key Focus on Quality and Quantity

Ideas

45 Preclinical Starts

70%

31 First in Humans

66%

24 First in Patients

28%

6 Phase III Starts

80%

5 NDAs

90%

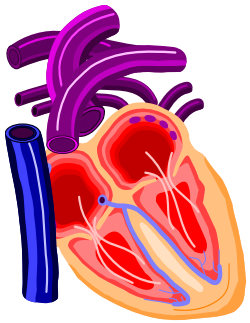
4 Products

Need to Reduce Attrition

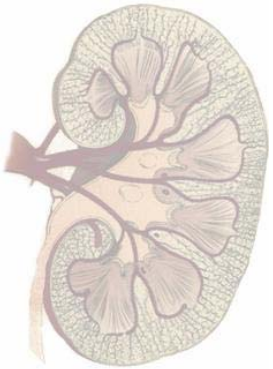
Pharmacogenomics at Pfizer

Vision…

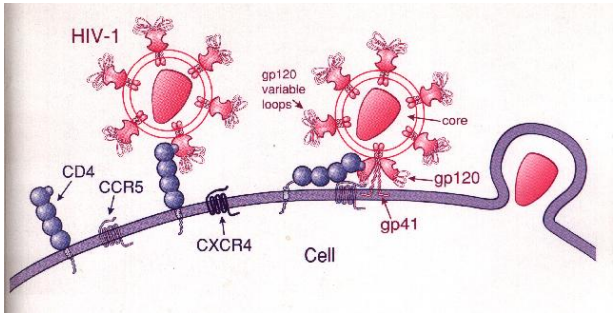
Genetics and genomics will revolutionize the diagnosis and treatment of disease and will be crucial for the successful discovery, development and delivery of effective new medicines



CETP Inhibitor

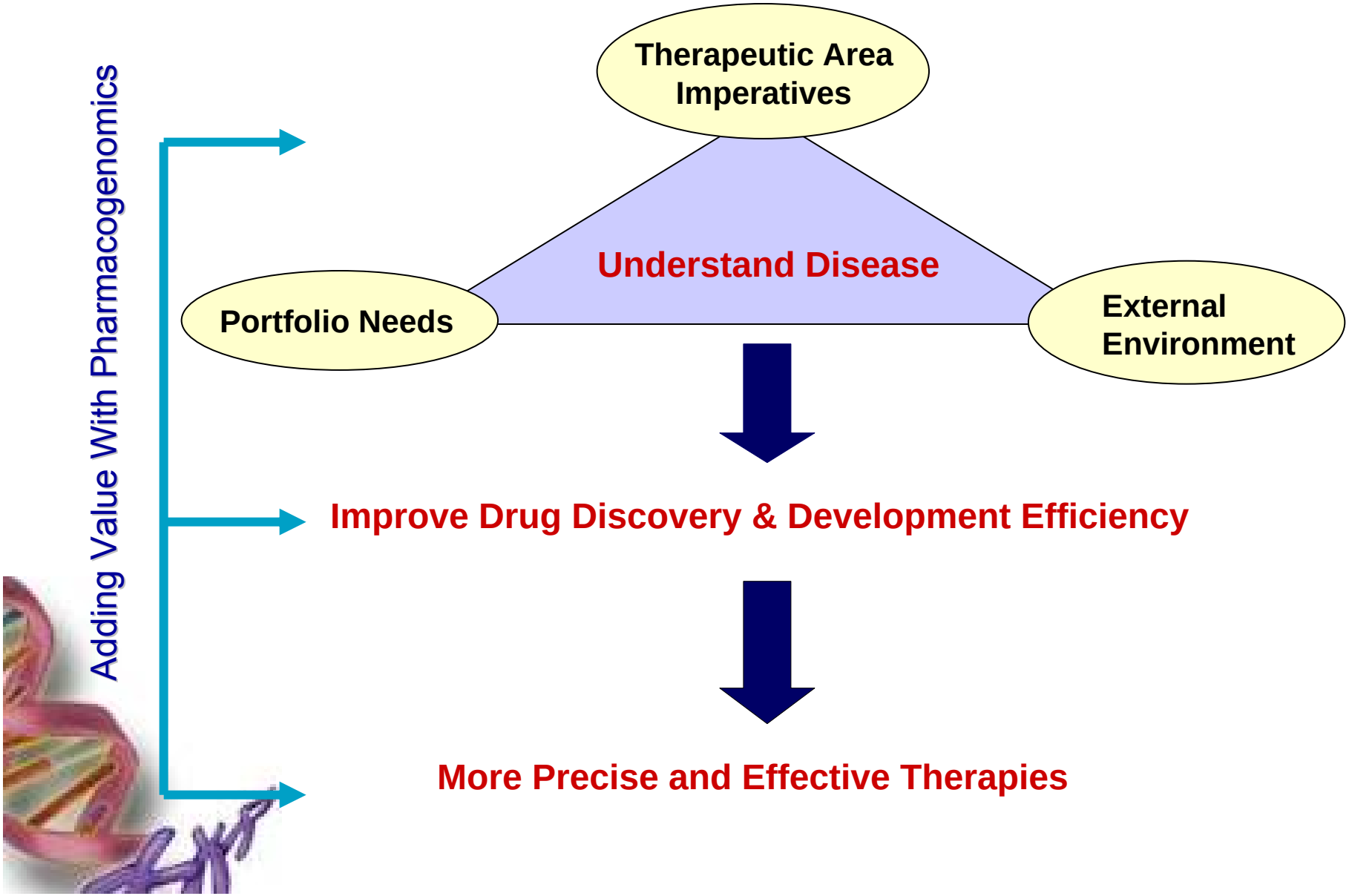


JAK3 Inhibitor

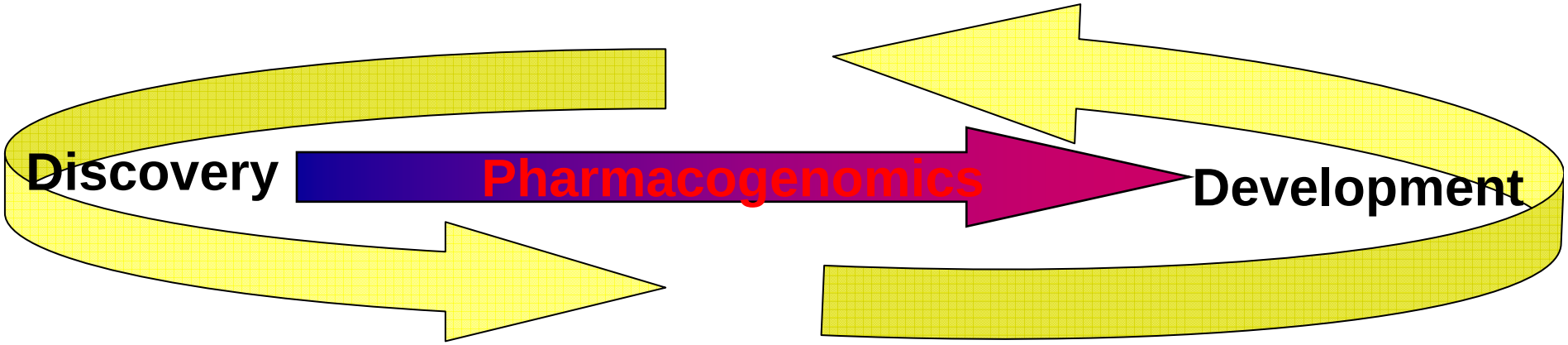


CCR5 Antagonist

Strategic Imperatives



Pharmacogenomics Across the Pipeline



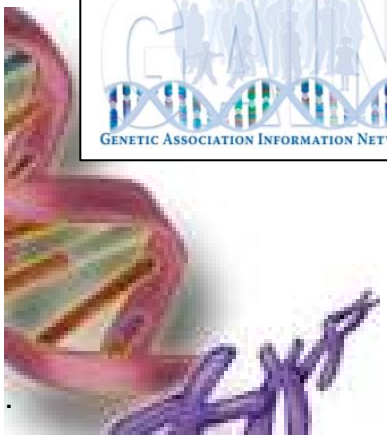
Choosing the Best Targets

Better Understanding of Our Targets

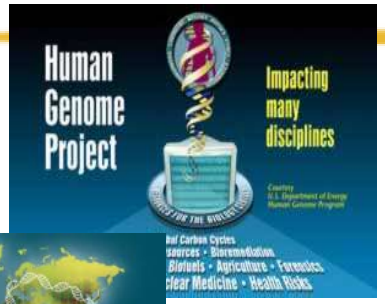
Genetic-based Selection of Optimal Population

Predicting Efficacy and Safety

Differentiating and Defending our Brands



The Pharmacogenomics Opportunity



The SNP Consortium Ltd.

PERLEGEN SCIENCES

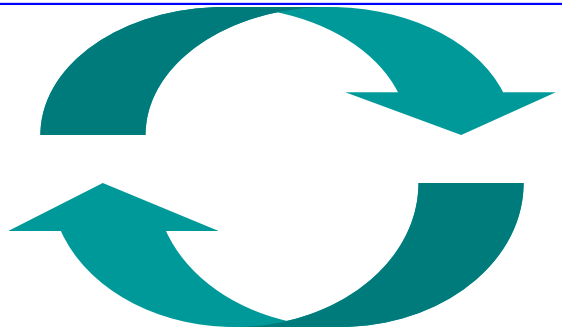
GENETIC ASSOCIATION INFORMATION NETWORK

AFFYMETRIX

IBM

Clinical Samples: DNA Linked with Phenotype

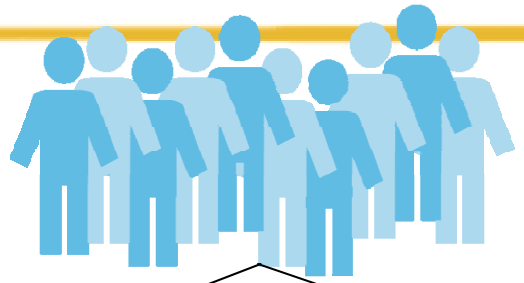
Exploration of Discovery Targets to Add Human Relevance



Building the Infrastructure to Support Pharmacogenomics

Application to Clinical Development

Human Genetics and Disease Definition



Phenotypically the same disease

Genetically
Different Diseases
Different responses

Disease risk driven by Genes x, y, and z

Disease risk driven by Genes a, b, and c

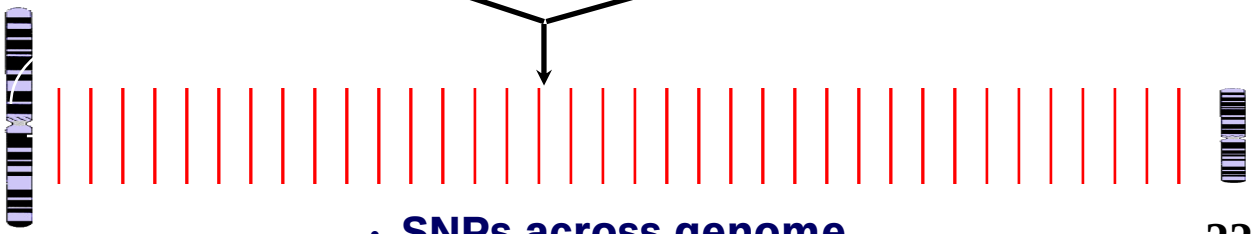
Human genetics can decipher the genetic differences in clinical phenotypes

- ***Disease risk- align therapeutic with genetic risk***
- ***Disease outcome risk- predict subjects likely to express rapid disease progression***
- ***Biomarker variability- identify genetic causes to intersubject variation in biomarkers***
- ***Safety risk – identify subjects at increased risk for safety event***

Using Whole Genome Scans for Therapeutic Targets

Disease Population

Matched Control Population

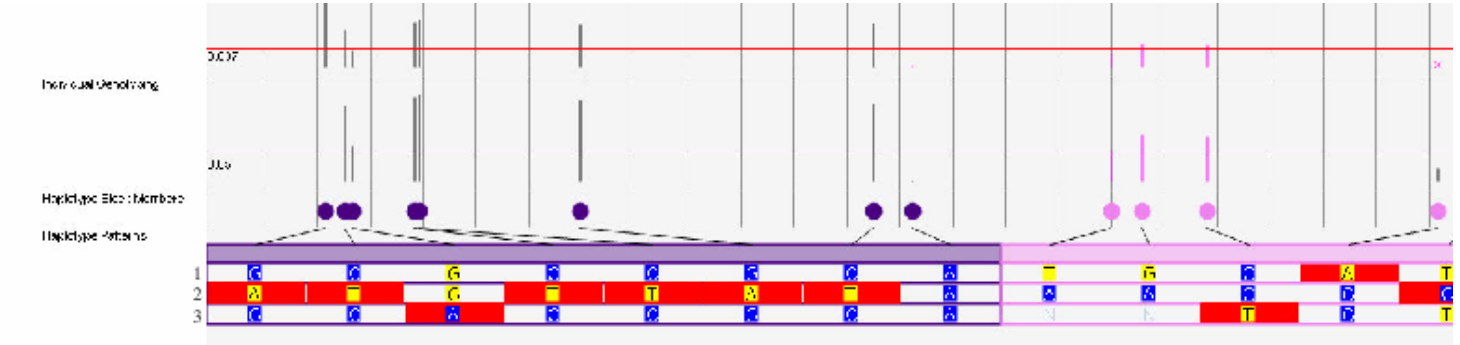


1

- **SNPs across genome**

22

Evaluate every gene in the genome in one experiment



Regions of association

Identify new targets for common diseases

Identify molecular signatures predictive of response

Metabolic Syndrome
ATP III* Guidelines set by the NCEP
National Heart, Lung, and Blood Institute of NIH

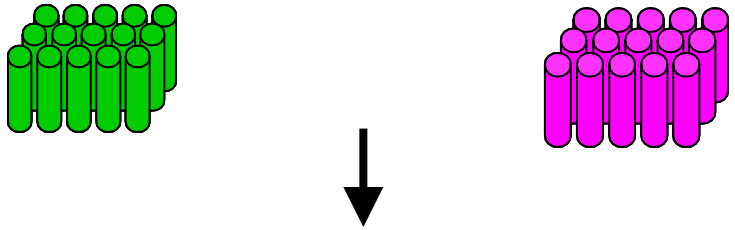
❑ Abdominal obesity	
❑ Men	>40 in
❑ Women	>35 in
❑ Trigs	>150 mg/dL
❑ HDL cholesterol	
❑ Men	<40 mg/dL
❑ Women	<50 mg/dL
❑ Blood pressure	>130/85 mm Hg
❑ Fasting glucose	>110 mg/dL

***Metabolic Syndrome defined as having three or more of the five component phenotypes**

Study Design

Metabolic syndrome and its component phenotypes

Case vs. Control



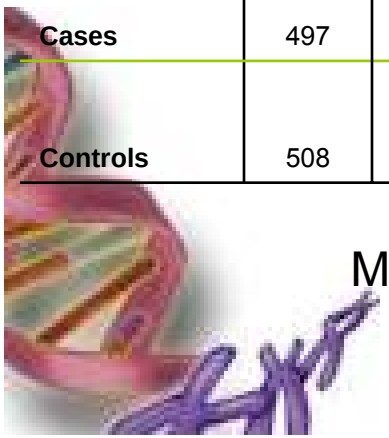
Two genome-wide screens: > 200,000 SNPs
(Two populations: Indian Asian males and Caucasian males)

Replication screen: 5,800 SNPs
(Four populations: Indian Asian females, Caucasian females, Mexican females, and Mexican males)

Genome-wide Scans – Populations
Indian Asian and Caucasian Males

* ATPIII criteria

	No.	Age	Waist > 40 inches*	TG ≥150 mg/d*	HDL <40 mg/dl*	SBP >130 mmHg*	DBP >85 mmHg*	Glucose ≥ 110 mg/dl*
Indian Asian								
Cases	500	52.60 ±8.09	41.42 ±4.20	236.74 ±151.30	42.51 ±8.97	142.70 ±20.19	87.80 ±11.83	132.84 ±54.54
Controls	499	52.50 ± 8.23	36.69 ± 3.15	116.59 ± 61.41	51.09 ± 12.48	131.80 ± 17.05	80.60 ± 9.68	94.86 ± 16.20
Caucasian								
Cases	497	55.60 ± 8.39	43.19 ± 4.44	234.07 ± 157.53	42.51 ± 8.97	142.80 ± 18.71	86.10 ± 10.77	121.50 ± 48.96
Controls	508	55.50 ± 9.52	36.22 ± 3.44	96.12 ± 40.05	55.38 ± 13.26	131.30 ± 17.54	78.00 ± 8.96	93.06 ± 19.08



Measurements for component phenotypes as well as other biometrics and environmental confounders

Individuals chosen to perform case-control study of Metabolic Syndrome based on ATP III criteria

	Number of risk factors					
	0	**1**	**2**	**3**	**4**	**5**
Indian Asian						
Cases	0	0	0	354	124	22
Controls	224	236	40	0	0	0
Caucasian						
Cases	0	0	0	375	103	22
Controls	328	172	0	0	0	0

Our Approach to Genome Scans

First Genome Scan

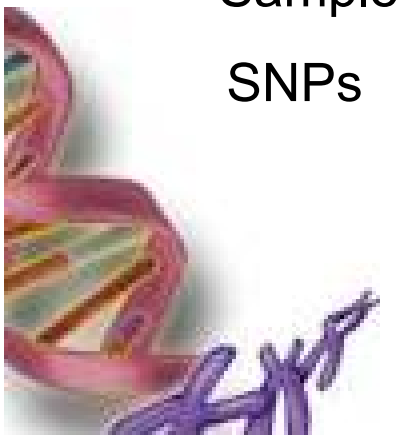
Samples: 500 cases, 500 controls males of Indian Asian ancestry

SNPs

- 248,000 SNPs attempted
- Common haplotype tagging SNPs: blocks defined by >80% coverage of patterns with freq>0.1
- Haplotype map based on re-sequencing 24 individuals of mixed ancestry, ~1 million common SNPs
- Additional SNPs selected from dbSNP to obtain uniform spacing (every 13.5 Mb) across the genome

Second Genome Scan

Samples: 500 cases, 500 controls of European ancestry



SNPs

- 267,000 SNPs attempted
- SNPs selected to tag European LD bins:

  $r^2 > 0.8$, MAF > 0.10
- LD map based on genotyping 1.7 million SNPs in 24 samples of European descent, contained ~1 million SNPs with MAF > 0.10

Genome-Wide Scans – Analyses

Univariate analyses - Models included terms for age and key environmental confounders

- Logistic regression for discrete outcomes

  Metabolic syndrome status ~ age + genotype

- Linear regression for quantitative outcomes

  log (waist) ~ age + genotype

  1/sqrt (trig) ~ age + alcohol + genotype

  sqrt (hdl) ~ age + alcohol + genotype

  log(sbp) ~ age + (blood pressure meds)*genotype

  log(sbp) ~ age + (blood pressure meds)*genotype

  log(HOMA) ~ age + (diabetes mellitus)*genotype

5800 SNPs Chosen for Replication Screen

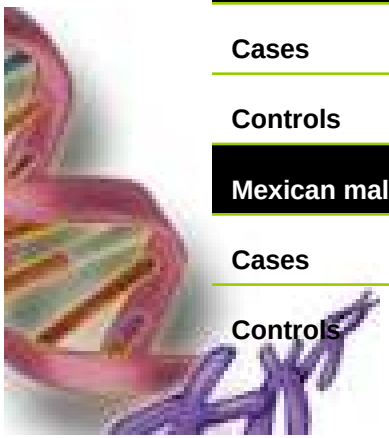
Came from two categories:

1. SNPs significantly associated with metabolic syndrome or component phenotypes in genome-wide scans
    - $p<0.0001$ in either scan
    - $p<0.001$ in both scans

2. SNPs chosen
    - To improve coverage of intervals containing SNPs associated in genome-wide scans
    - In genes of interest due to previously identified associations with metabolic syndrome or component phenotypes

SNP Replications: Metabolic Syndrome Populations defined by ATP III criteria

	Number of risk factors					
	0	1	2	3	4	5
Indian Asian females						
Cases	0	0	0	261	119	27
Controls	140	307	327	0	0	0
Caucasian females						
Cases	0	0	0	84	53	16
Controls	288	247	172	0	0	0
Mexican females						
Cases	0	0	0	333	315	126
Controls	218	286	281	0	0	0
Mexican males						
Cases	0	0	0	236	136	30
Controls	128	129	309	0	0	0



Replication Screen – Analyses

Univariate analyses - models allowed for heterogeneity of genotype effects across populations

- Logistic regression for discrete outcomes

  Metabolic syndrome status ~ age + pop*genotype

- Linear regression for quantitative outcomes

  log (waist) ~ age + pop*genotype

False discovery rates - estimated across subset of SNPs selected for each component phenotype

Subphenotype: HDL Cholesterol Replicated Associations

dbSNP ID	Chr	Position	Genes
rs7205804	16	56780286	[CETP]
rs711752	16	56771608	[CETP]
rs5880	16	56790488	[CETP]
rs1800777	16	56792716	[CETP]
rs9282541	9	103000673	[ABCA1]
rs5882	16	56791489	[CETP]
rs10509681	10	96463336	[CYP2C8]
rs326	8	19829712	[LPL]
rs328	8	19829997	[LPL]
rs325	8	19829601	[LPL]

Results

- Significant – FDR ≤ .007

  Thirteen SNPs all located within genes previously shown to be associated with component phenotypes

  - HDL levels – *CETP, ABCA1,* and *LPL*
  - Triglyceride levels - the *apoAI/apoAV/apoCIII* gene cluster and *LPL*

- FDR ≤ 0.35

  Eleven additional SNPs

  - HDL levels – three in *LPL,* one novel
  - Triglyceride levels – two in *LPL*, one novel
  - Diastolic blood pressure – one novel
  - HOMA / Insulin – three novel

Coverage of the Genome-wide Scans

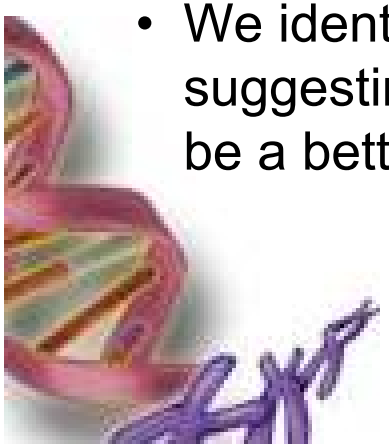
- Used HapMap Phase II data to estimate power
  - Determine direct power based on CEU allele frequency and disease model parameters
  - Determine maximum r^2 with a genotyped SNP and adjusted power
  - Conservative $\alpha = 10^{-4}$ level
- Disease model parameters
  - 500 cases & 500 controls
  - Multiplicative risk
  - Discrete trait with prevalence 0.2

Summary

- Several genes previously shown to be associated at with component phenotypes, such as *CETP* with HDL levels and *LPL* and the chromosome 11 *Apo* gene cluster with triglyceride levels had genome-wide significance in our study

- Some new loci with smaller effect sizes were associated with component phenotypes but examination in additional populations is required to distinguish from false positives

- We identified no genetic risk factors for metabolic syndrome, suggesting that analyzing the individual component phenotypes may be a better means of studying this disease



The Need to Accelerate Our Pace and Our Learnings

GAIN: The Genetic Information Association Network

A Unique Public/Private Partnership

Pfizer/Perlegen's Goals in Funding and Supporting GAIN

- Accelerate our understanding of the genetic basis of human diseases – many of which remain major medical needs for patients
- Generate and quickly release genotype data (pre-competitive) for several these important human diseases
- Encourage analysts around the world to participate in the analysis of these important disease data sets and thereby advance the science of whole genome analyses for the entire scientific community
- Finally, build the foundation for more precise therapies where the potential exists to better diagnosis and treat human disease

Why Are Whole Genome Methods So Important?

The Challenges in Drug Development: Applying Pharmacogenomics – Why We Need to Examine the Whole Genome

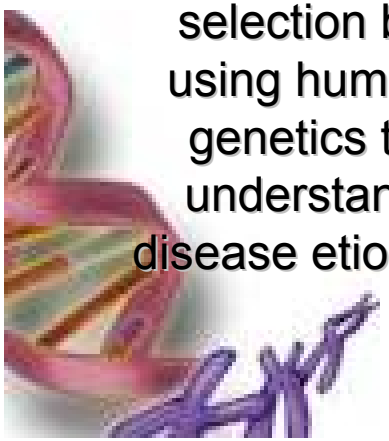
Discovery → **Development**

Increase target selection by using human genetics to understand disease etiology

Evaluate how common variants in the target will influence the efficacy of the compound

Select subjects or stratify populations in early efficacy studies to improve quality of decision making

Prediction of efficacy and adverse events based on a subject's genotype



Torsade de Pointes

Torsade de Pointes (TdP)

Business Drivers for Genetic Studies

- Low incidence of drug-induced *torsade de pointes* during development of dofetilide resulted in alteration of dosing scheme for patients
- Phenotype mimics the phenotype of inherited long QT syndrome
- Does the genetic basis of Long QT Syndrome provide any insight into drug-induced TdP?

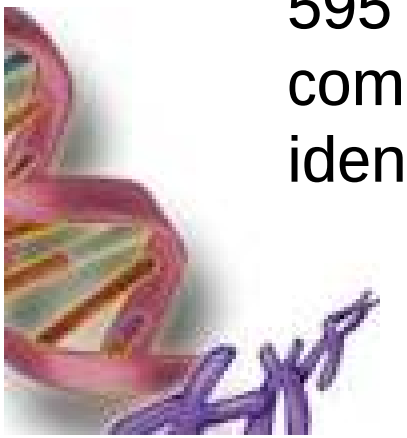
Can Familial Genes Teach Us About Drug-Induced TdP?

Locus	Chromosomal Location	Gene	Product	Function
LQT1	11p15.5	KCNQ1	K+ channel	Structural alpha subunit Iks
LQT2	7q35-36	KCNH2 (HERG)	K+ channel	Structural alpha subunit Ikr
LQT3	3p21-24	SCN5A	Na+ channel	Functional sodium channel
LQT4	4q25-27	ANK2	Structural Protein	Structural
LQT5	21q22.1	KCNE1 (Mink)	KCNE1	Regulatory beta subunit Iks
LQT6	21q22.1	KCNE2 (MiRP1)	KCNE2	Regulatory beta subunit Ikr
LQT7	17q23-24	KCNJ2	Kir2.1	Inward rectifying K channel
LQT8	12p13.3	CACNA1C		K+ channel subunit

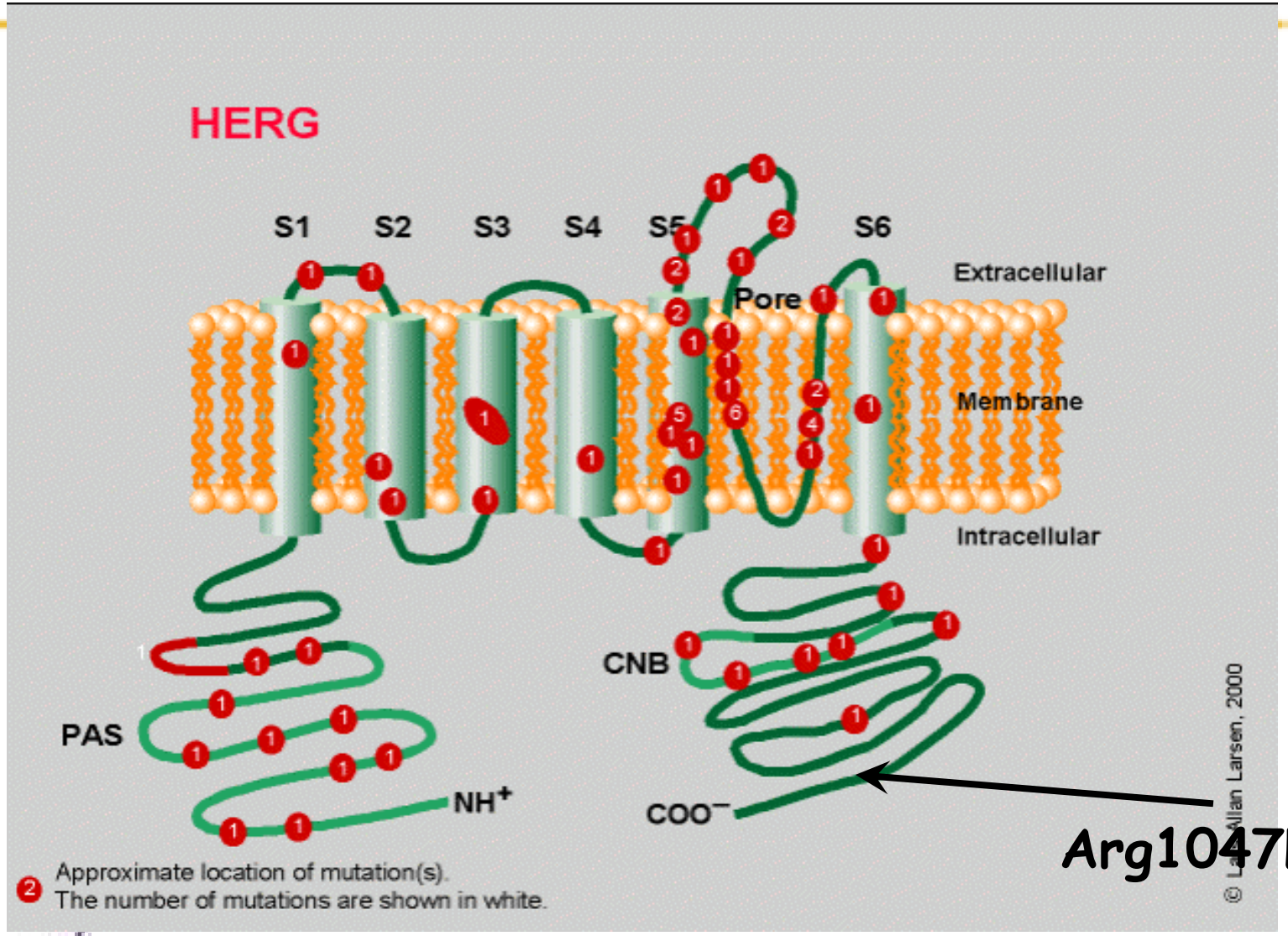
Can Genetics Predict TdP Prior to Therapy?

- DNA samples not collected during dofetilide clinical trials
- Heroic efforts made to collect DNA post-trial
- 40 Patients/families consented to DNA analysis
- For some, only plasma was available (Whole genome amplified for residual DNA where possible)
- All exons of 7 LQT genes scanned for mutations in 34 individuals who developed drug-induced TdP

- All TdP patients, 95 controls from the same study, and 595 controls from another study genotyped for all common amino acid changing SNPs and rare SNPs identified via SNP scanning



HERG Mutation



Highlighted in red are familial variants linked to long QT syndrome.

			Pop:	Mayo		KORA	DIAMOND		Pfizer Other	
			Ancestry:	Euro	Af Am	Euro	Euro	Euro	Euro	Af Am
			Healthy?:	Healthy		Healthy	CHD/MI		Healthy	
			TdP?:	TdP-		TdP-	TdP+	TdP-	TdP-	
			#	187/295	305/319	3966	34	95	555	40
Gene	**SNP**	**AA Change**	**AA Conserved?**	MAF %	MAF %	MAF %	MAF %	MAF %	MAF %	MAF %
ANK2	rs36210415	G475R	Yes				1.5	0	0	0
	rs29372	N687S	Yes				0	0	0	0
	rs36210416	V708M	Yes				1.5	0	0	0
	rs28377576	V2369A	T (mouse)				10.3	9.3	11.5	12.5
	rs3733616	A2423T	P (mouse)				0	0	0	0
	rs3733617	P2835S	T (mouse)				7.4	2.6	3.5	28.8
	rs36210417	I3285T	Yes				1.5	2.1	1.1	0
	rs36210418	S3300R	Yes				1.5	2.6	2.3	1.3
KCNE1	rs1805127	G38S	D (many)	34.1	23.6	35.5	33.8	37.4	36.8	23.8
	rs1805128	D85N	N (rabbit)	0.6	0.4		0	2.1	1	2.6
KCNE2	rs2234916	T8A	Yes	0.6	0	0.7	0	0.5	0.3	0
KCNH2	rs36210422	R176W	Yes	0.3	0		1.5	0.5	0	0
	Ref 13	R887H	Yes				0	0	0	0
	rs1805123	K897T	R (dog)	21	4.3	24.1	15.2	26.8	23.9	5
	Ref 32	P917L		0.8	0		0	0	0	0
	rs36210421	R1047L	Yes	1.9	0.2	2.4	7.6	2	1.8	1.3
KCNJ2	None found									
KCNQ1	rs36210419	K218E	Yes				1.5	0	0	0
	rs12720457	K393N	Yes	0.6	0		0	0	0.1	0
SCN5A	rs6791924	R34C		0	4.8		0	0	0.1	7.5
	rs36210420	N291H	Yes				1.5	0	0	0
	rs1805124	H558R	R (many)	11.5	17.6		37.5	14.4	24.1	23.1
	rs36210423	A572D	T (mouse, cow)				1.5	0	0.4	0

TdP Studies

A single genetic variant does not account for drug-induced TdP.

LQT genotypes alone could not be used to completely predict susceptibility to TdP, even when used in conjunction with phenotype.

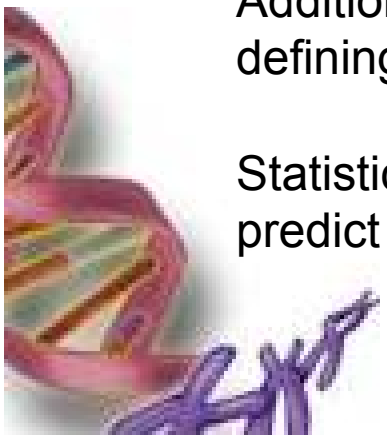
Statistical modeling using genotypic and phenotypic variables was unable to predict all adverse events.

Current research suggests genetic variation can be identified in one of the LQT candidate genes, approximately only 20% of the time.

In other subjects the effect is mediated by other undetermined factors.

Additional research through whole genome approaches may offer opportunity fo defining other genes involved in TdP.

Statistical modeling using genotypic and phenotypic variables was unable to predict all adverse events.



Candidate gene analyses, even though strongly based in selection, provides limited opportunity to define genetic basis of outcome.

Acknowledgements

Pfizer

Groton:

John Thompson
Greg Warnes
Craig Hyde
Poulabi Banerjee
Patrice Milos
Albert Seymour
Hakan Sakul
Doug Lee
Kelly Longo

St. Louis:

Kay Broschat
Nick Staten
Rich Mazzarella

Ann Arbor:

Alu Chatterjee
John Wetterau

Perlegen

David Hinds
David Cox

Imperial College of London

John Chambers
Jaspal Kooner
James Scott

National Institute of Medical Sciences, Mexico

Carlos Aguilar