Multistage Design Options for Pharmacogenetic Studies

Duncan C. Thomas

with

David Conti

University of Southern California

Scientific Questions in Pharmacogenetics

- Why do some people respond favorably to a particular treatment and others not?
- Who do some experience a particular side effect and others not?
- Why is one treatment better for some, another treatment better for others?
- Can genetics help explain these effects?

Multistage Sampling

- **Designs** that exploit information already collected or readily obtainable on a large sample to improve the cost-efficiency of subsample(s) for other variables
- **Analyses** that combine information from both the main study and subsamples
- Applications in
  - **Epidemiology:** White *Am J Epidemiol* 1982:115:119-28
    Breslow & Chatterjee, *Appl Statist* 1999;48:457-68
  - **Genetics:** Whittemore & Halpern, *Stat Med* 1997; 16:153-67

Design for Studying Rare Exposures and Rare Diseases

- **Stage I:** case-control sample by *Y*, observe surrogate *Z* for exposure
- **Stage II:** subsample 2x2 cells defined by *Y* and *Z* and measure exposure *X* (and other covariates)
- Analysis of stage II data must adjust for differential sampling fractions
- **Better:** analyze stage I and II data jointly

White *Am J Epidemiol* 1982:115:119-28;

Focus on Design Issues in Phamacogenetics

- Primary focus on interactions rather than main effects
- Prior knowledge about pathways targeted by agent under study
- Exposure (treatment) can be randomized
  - Independently of genotype
  - And vice-versa: genes segregate independently of treatment
- Unrelated individuals, not families
- Possibility of case-only designs, particularly where treatment is randomized

Reasons to Consider Multistage Designs

- Cost-efficiency
- Opportunity to use informative sampling
- Joint analysis of data from different samples
- Optimization of design

Chicken or Egg?

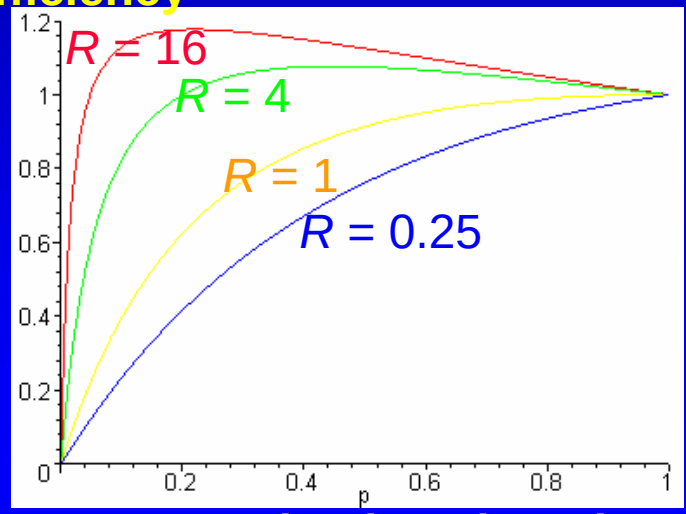
- **Start with clinical trial: add genetic association study to look for modifiers of treatment response**

OR

- **Start with a case-control or cohort study of genes: use genes to target clinical studies of treatment outcomes**

Optimization of Designs

- Compute expected Fisher information for the joint analysis of main and substudy as a function of parameters and sampling probabilities
- Find sampling scheme that maximizes E(information), subject to constraint on total cost
- **Example**: E(info)/Cost as function of overall sampling fraction



Examples

- Candidate gene association study using tag SNPs
- Pathway-based study involving biomarkers
- Genome-wide association study

Candidate Gene Studies

- *A priori* hypotheses about candidate gene(s)
- If functional variants known in a clinical trial: no need for multistage sampling ...
  But not if starting point is a cohort study.
- Negative result could mean gene is not relevant or wrong variant(s) were tested
- Complete characterization would require sequencing of entire gene in full sample
- Focus on common variants ⇒ tagSNP approach

Preservation of Pancreatic β-C
Prevention of Type 2 Diabetes
Treatment of Insulin Resistanc
Women

Thomas A. Buchanan,[1,2,3] Anny H. Xiang,[3,4] Ruth K. Pete
Jose Goico,[1] Cesar Ochoa,[1] Sylvia Tan,[4] Kathleen Berko
and Stanley P. Azen[3,4]

Diabetes 51:2796–2803

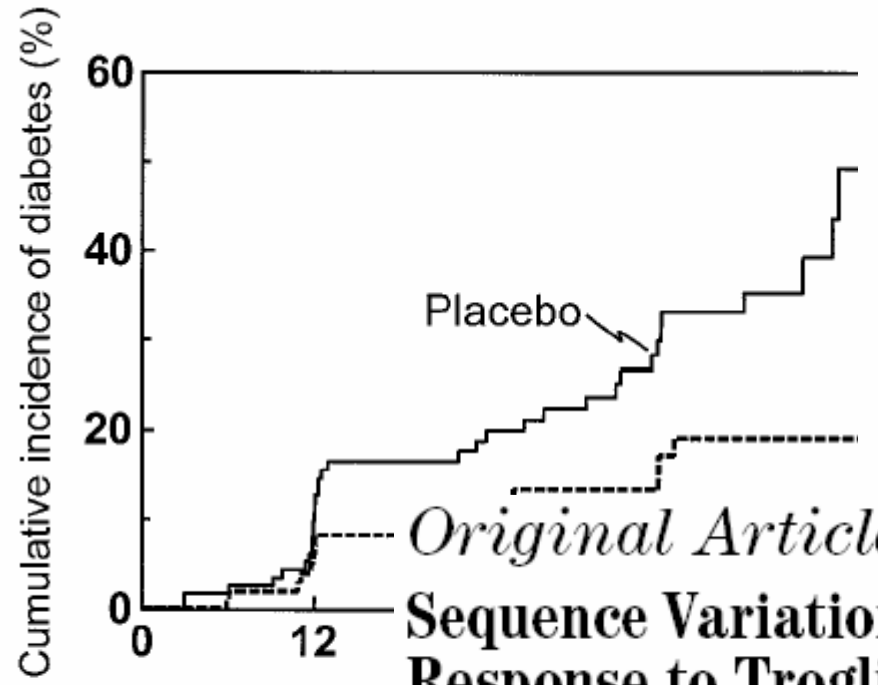


FIG. 1. Cumulative incide
returned for at least one fo
or troglitazone. The rate in the troglitazone group was significantly lower than the rate in the placebo group ($P = 0.009$).

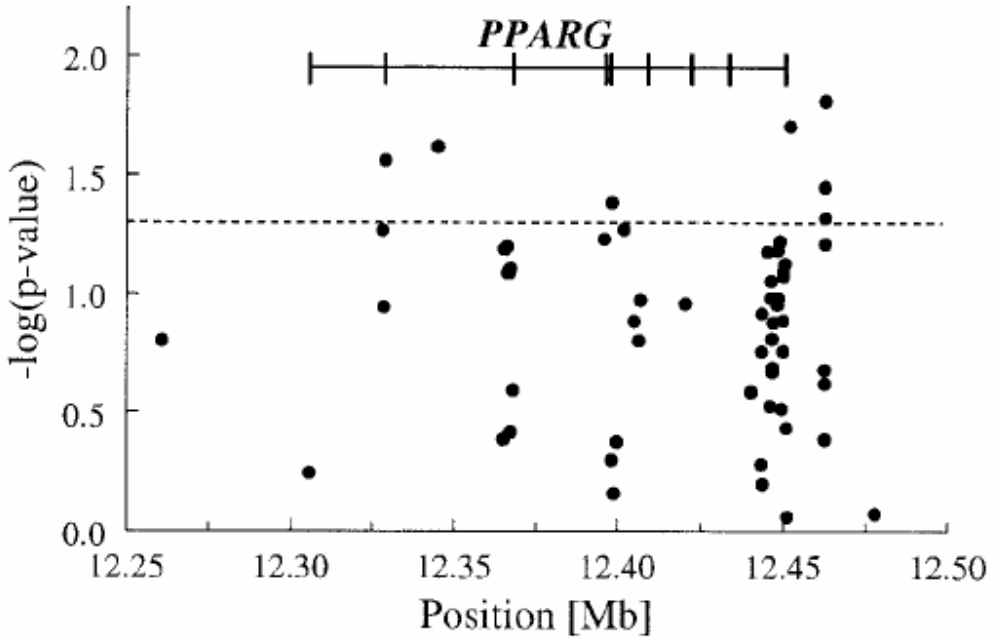


FIG. 1. Single marker association with response to troglitazone. The negative log of the P value for the χ^2 test of association is plotted according to physical distance. Horizontal dashed line denotes P value of 0.05. Two SNPs in close proximity gave identical P values, so only seven of the eight significant results are visible. The gene structure for *PPARG* is shown at the *top* with the A1 promoter on the *left*.

Original Article

Sequence Variation in *PPARG* May Underlie Differential Response to Troglitazone

Johanna K. Wolford,[1] Kimberly A. Yeatts,[1] Sharanjeet K. Dhanjal,[2] Mary Helen Black,[2]
Anny H. Xiang,[2] Thomas A. Buchanan,[3] and Richard M. Watanabe[2]

Diabetes 54:3319–3325, 2005

Multistage Sampling for TagSNP Studies

- Small sample to characterize LD patterns and choose tag SNPs *S*
- Only tag SNPs and treatment *T* are tested in main study
- Joint analysis allows tests of untyped SNPs *G*

$$p_G(Y \mid S,T) = \Sigma_g \, p_\beta(Y \mid G=g,T) \, p_\alpha(G=g \mid S)$$

- Haplotype analysis similar, but requires additional summation over possible haplotype resolutions given unphased genotypes

Thomas et al., Genet Epidemiol 2004;27:401-14

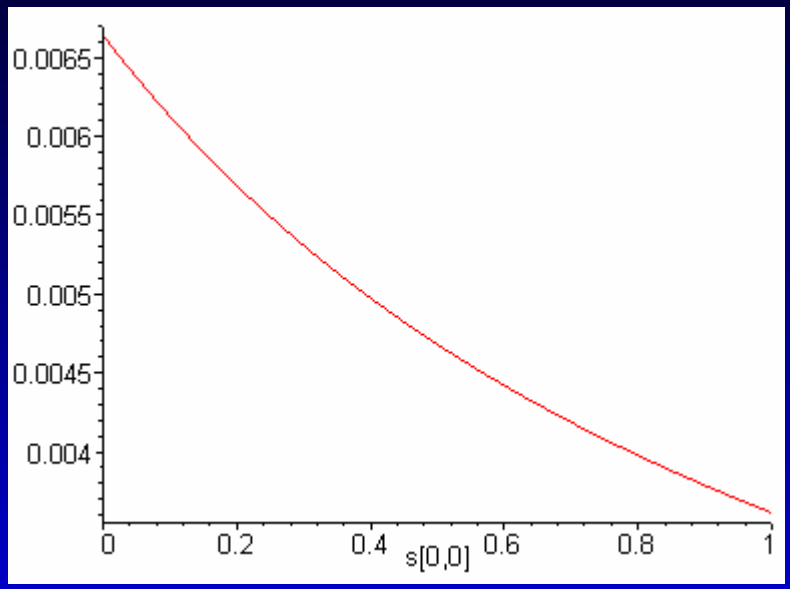
Extensions

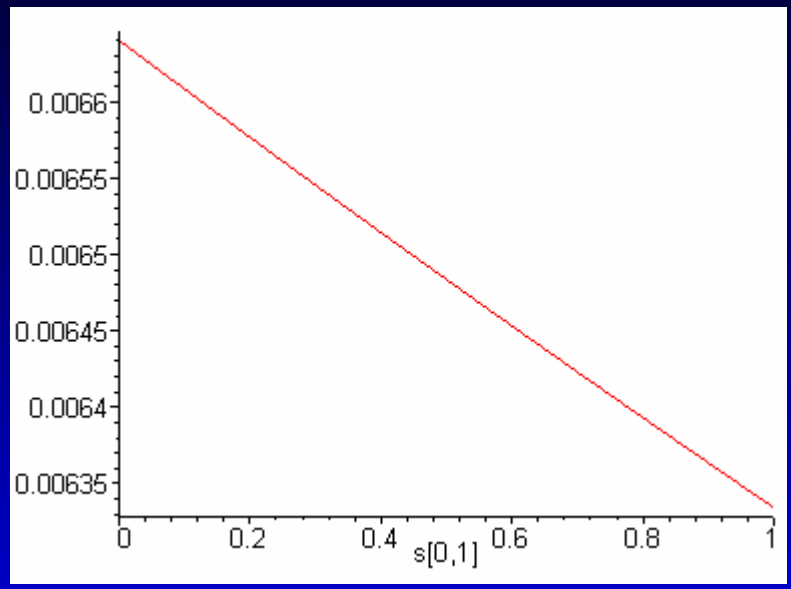
- **Multistage samples incorporating sequencing**
- **Gene-treatment interactions: optimize design by sampling on outcome, treatment, and surrogate for causal variant**

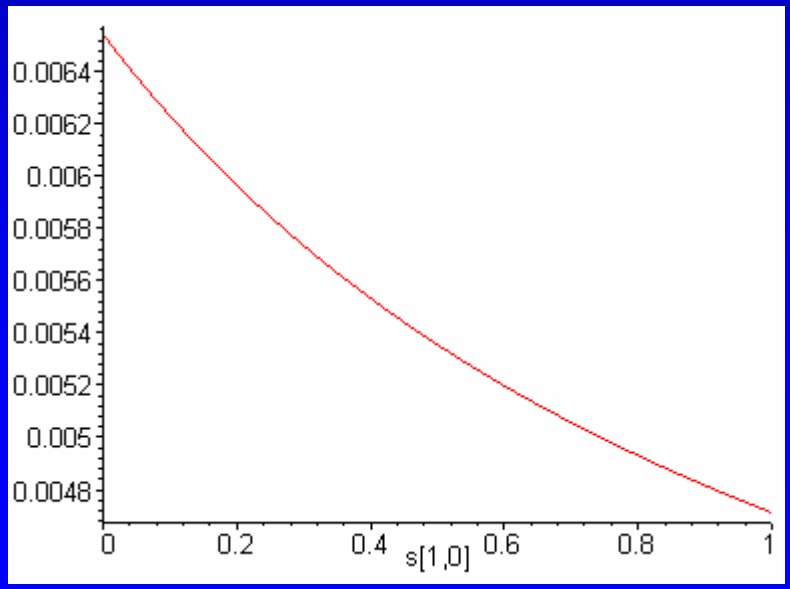
Nested Genetic Study Within a Clinical Trial

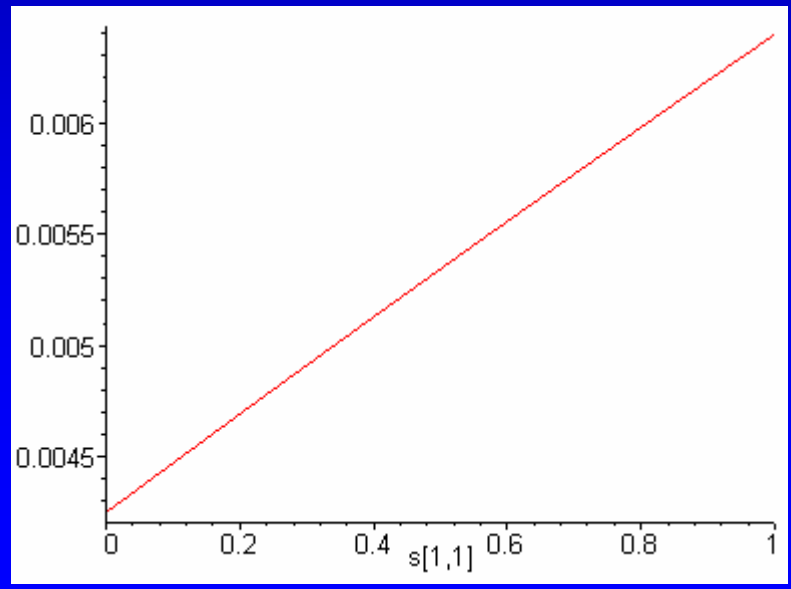
- **Stage I:** observe $Y \mid T$
- **Stage II:** sample conditional on Y, T; observe $G \mid Y, T$
- Likelihood is: $\prod_M p(Y \mid T)^{N_{YT}} \times \prod_S p(G \mid Y,T)^{n_{GYT}}$
- Total information is: $\sum_M N_{YT}\, i_{YT} + \sum_S n_{GYT}(s)\, i_{GYT}$
- Choose s to maximize information per unit cost
- Optimal design might sample only $Y=1$, $T=1$

$Info_{GxT}$/Cost by Sampling Fractions









Is Equal Allocation Optimal?

Sampling plan	Sampling fractions	ARCE (*R*=2)
No subsampling	(1,1,1,1)	.0033
Constant sample	(.086, .086, .086, .086)	.0044
Equal allocation	(.046, .789, .076, 1)	.0064
Case-control	(.051, .051, 1, 1)	.0063
Sample only one cell	(1, 0, 0, 0)	.0024
	(0, 1, 0, 0)	.0036
	(0, 0, 1, 0)	.0044
	(0, 0, 0, 1)	.0071

Clinical Trial Within an Observational Study

- **Stage I:** observe G, Y_1 (disease)
- **Stage II:** sample $Y_1=1$ subjects within strata of G

  assign $T \mid G$ at random

  observe treatment outcomes $Y_2 \mid T, G$
- Optimize sampling fractions given G

Case-Only Designs

- From clinical trial, sample only responders or only nonresponder (whichever is rarer)
- From cohort study of *T*, sample only cases
- In either design, examine *G-T* association
- Assuming *G* and *T* are independent in population, *G-T* association in cases estimates *GxE* interaction

Counter-Matched Design

- In a cohort with *T* and *Y* :
- Match each *T*=1 case with a *T*=0 control and vice-versa from within Cox risk set
- Measure *G* on cases and CM'd controls
- Analysis is by conditional logistic regression with offset term for control sampling fractions

Langholz & Goldstein, *Statist Sci* 1996;11:35-53

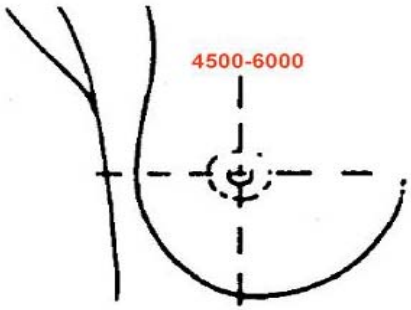
Example: WECARE

- Nested case-control study of second breast cancer in relation to radiotherapy and DNA repair genes (*ATM* etc.)
- 700 cases of bilateral breast cancer
- 1400 controls, counter-matched on radiotherapy (2 treated + 1 untreated per triplet)
- Contralateral radiation doses estimated by phantom dosimetry
- Genotyping of *ATM* and other genes

Bernstein et al. *Breast Ca Res* 2004;6:R199-214

Doses to the Contralateral Breast During RT

Subgroup	RR (95% CI) ≥1.0 Gy vs. no RT
All subjects	1.3 (1.0 – 1.6)
Under age <45 at exposure 5+ y latency	2.0 (1.1 – 3.8)
Under age <40 at exposure 5+ y latency	2.8 (1.1 – 8.8)

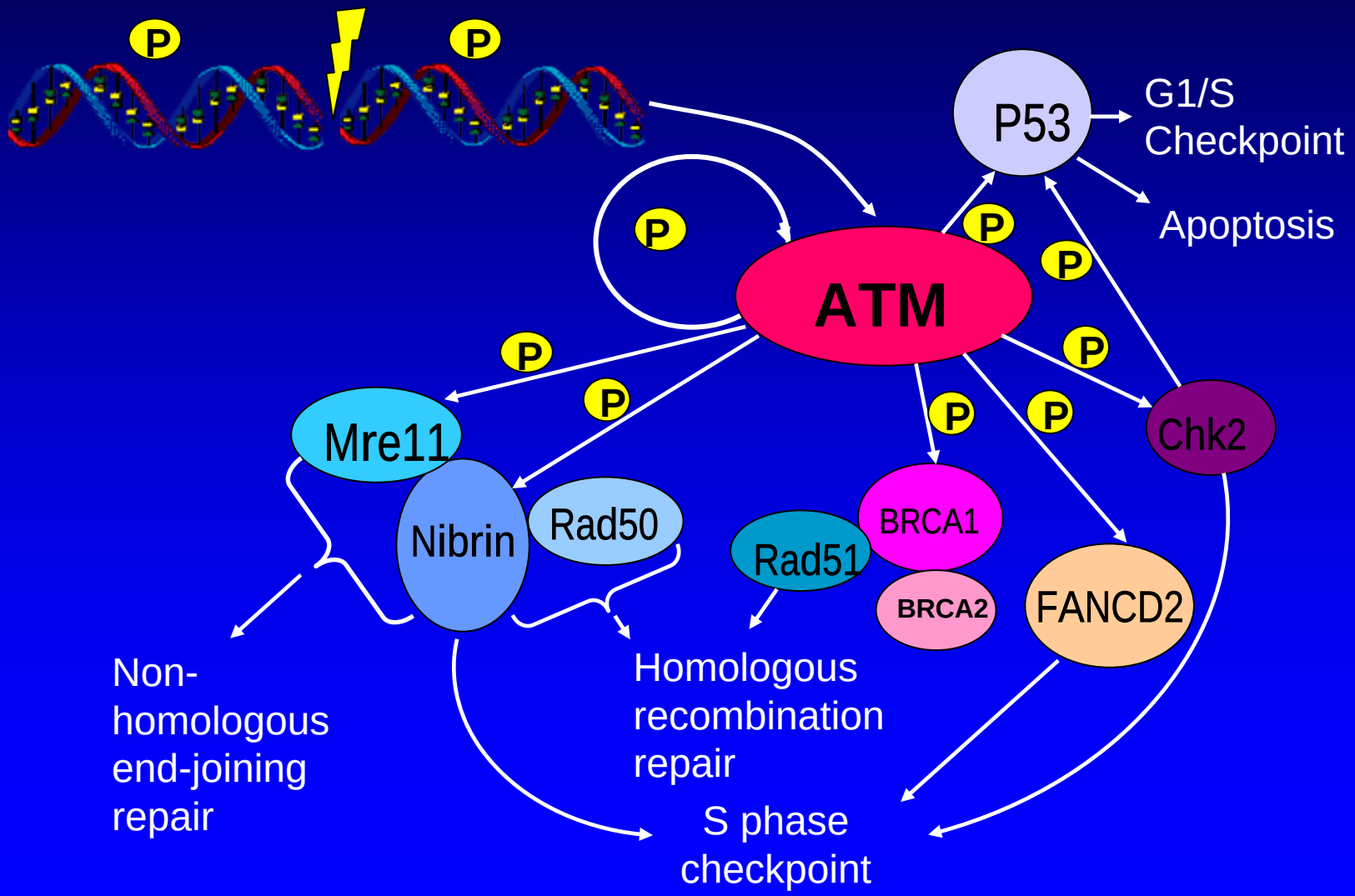


Treated Breast: Tumor Dose

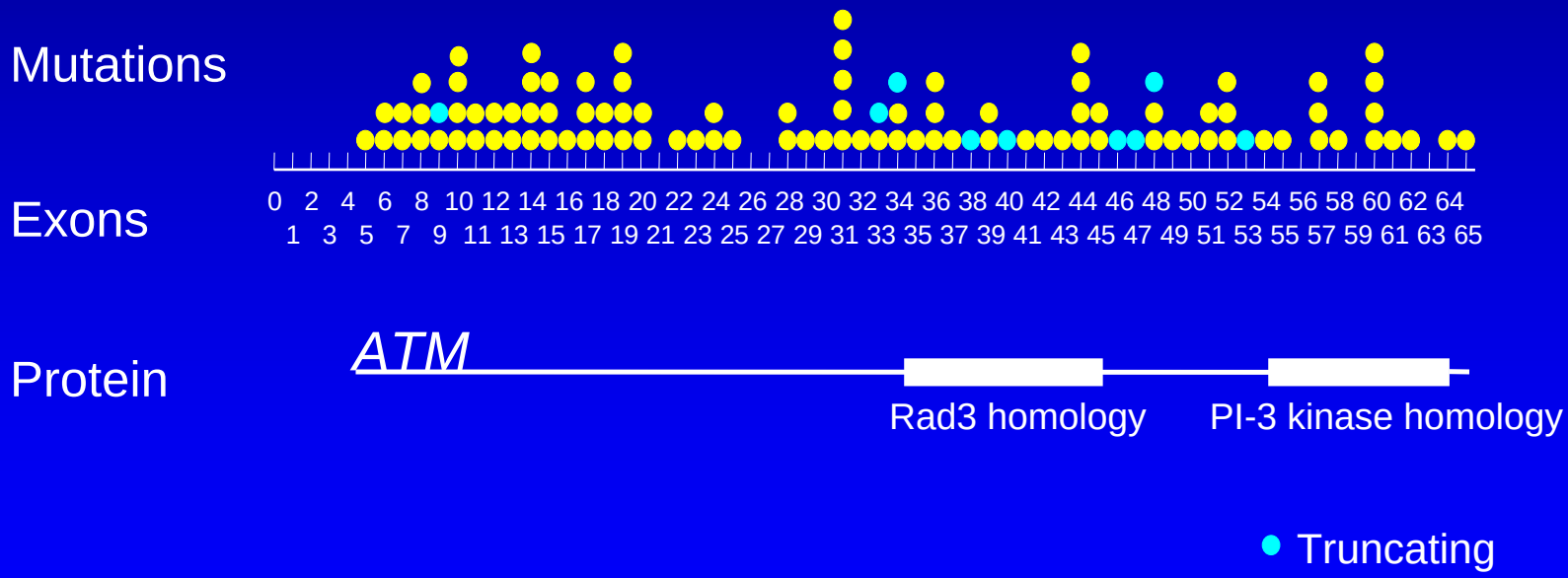
Contralateral Breast: Range of Average Dose per Quadrant Among Patients

Absorbed radiation dose (cGy) to the contralateral breast during RT is estimated using patient-equivalent phantoms and medical/treatment record information. Range limits correspond to techniques used among WECARE participants that resulted in the lowest and highest doses to each quadrant and nipple.

Role of ATM in Cellular DNA Damage Response



Distribution of Unique Non-Silent ATM Variants: WECARE Study Progress



Bernstein et al, *Hum Mut* 2003;21:542-50

Main Effects of *ATM*

Common variants	RR (95% CI)
All 12 w freq > 1%	0.8 (0.6 – 0.97)
IVS 14-55T>G	0.6 (0.4 – 0.96)

Table 2: Risk of developing second primary breast cancer using general categories of ATM variants (compared to wild-type).

	Unadjusted for common variants				Controlling for common variants			
Variant Category	Ca (N)	Co (N)	RR*	95% CI	Ca (N)	Co (N)	RR*	95% CI
Wild-type	271	480	1.0		271	480	1.0	
Silent	256	572	0.9	0.8-1.0	99	190	0.9	0.7-1.3
Missense	296	596	0.9	0.8-1.1	80	140	1.2	0.8-1.7
Splicing	4	16	0.5	0.2-1.7	4	16	0.6	0.2-1.7
Truncation	11	7	2.3	0.8-6.5	11	7	2.1	0.7-6.0
Less likely functional ATM mutation carrier	49	109	0.8	0.5-1.2	43	87	1.1	0.7-1.7
Likely functional ATM mutation carrier	247	487	0.9	0.7-1.1	37	53	1.5	0.9-2.6

* adjusted for confounders.

ATM x Radiation Interaction

Table 3: ATM gene carrier status and radiation on risk of developing second primary breast cancer.

Variable	Cases		Controls		RR	95% CI
	RT+	RT-	RT+	RT-		
Overall*						
Less likely functional ATM mutation carrier	21	22	65	22	1.3	0.6-3.3
Likely functional ATM mutation carrier	24	13	42	11	3.8	1.3-11.0
Age (years)**						
<45 and less likely functional**	9	9	30	4	0.6	0.1-2.6
<45 and likely functional	10	5	15	9	7.6	1.8-32.5
45+ and less likely functional	12	13	35	18	2.2	0.7-6.4
45+ and likely functional	14	8	27	2	1.1	0.2-6.2
Latency (years)**						
<5 and less likley functional	16	13	39	13	1.7	0.6-4.9
<5 and likely functional	13	8	23	6	2.4	0.7-8.9
5+ and less likely functional	5	9	26	9	0.9	0.2-4.0
5+ and likely functional	11	5	19	5	6.0	1.2-29.4
Age and Latency**						
Age <45, latency <5 and likely functional	5	2	8	6	7.8	1.1-55.2
Age <45, latency 5+ and likely functional	5	3	7	3	9.1	0.8-98.5

*RR adjusted for confounders.

**Due to small sample numbers, RRs were adjusted for mutation types, but not confounders.

Examples

- Candidate gene association study using tag SNPs
- Pathway-based study involving biomarkers
- Genome-wide association study

Pathways and Biomarkers

- Studies of many related genes, e.g.,
  - Drug metabolism
  - Repair of DNA damage from therapeutic radiation
- Joint analysis of all relevant genes using hierarchical or pharmacokinetic models
- Wish to incorporate markers of intermediate endpoints, e.g., urine/blood concentrations of metabolites, but expensive or awkward

Effects of a 5-Lipoxygenase–Activating Protein Inhibitor on Biomarkers Associated With Risk of Myocardial Infarction

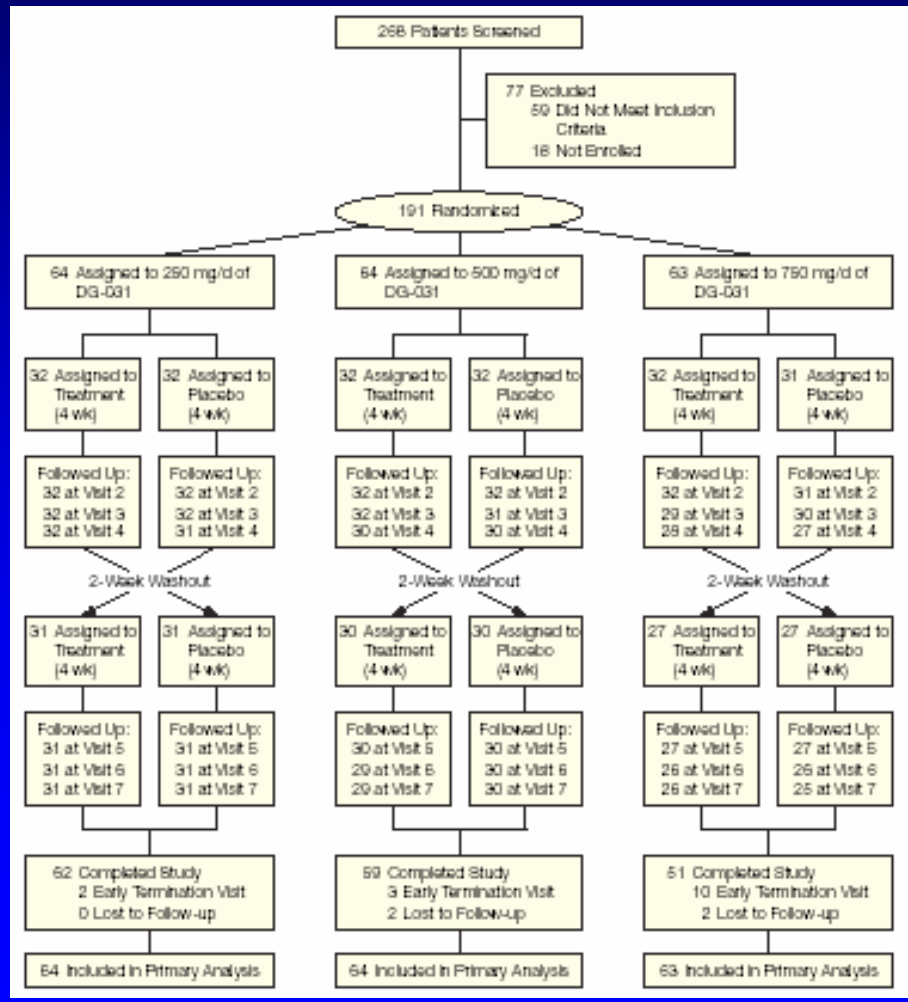
A Randomized Trial

JAMA. 2005;293:2245-2256

Hakon Hakonarson, MD, PhD
Sverrir Thorvaldsson, MSc

Context Myocardial infarction (MI) is the leading cause of death in the world. Variants in the 5-lipoxygenase–activating protein (FLAP) gene are associated with risk of MI.

- **DG-031 is FLAP inhibitor**
- **Aim is to assess treatment effect on biomarkers of MI risk (CRP, leukotrienes, MPO)**
- **Restricted to carriers of risk variants for *ALOX5AP* (87%) or *LTA4H* (13%)**



Effects of a 5-Lipoxygenase–Activating Protein Inhibitor on Biomarkers Associated With Risk of Myocardial Infarction

A Randomized Trial

JAMA. 2005;293:2245-2256

Hakon Hakonarson, MD, PhD
Sverrir Thorvaldsson, MSc

Context Myocardial infarction (MI) is the leading cause of death in the world. Variants in the 5-lipoxygenase–activating protein (FLAP) gene are associated with risk of MI.

Biomarker	Dose / Time	Change (95% CI)	P
Leukotriene B_4	750 mg/d	26% (10 – 39%)	.003
MPO	750 mg/d	12% (2 – 21%)	.02
CRP	500 – 750 2 wk	16% (-2 – 31%)	.07
	4wk post washout	25% (5 – 40%)	.02

Gene-Expression Patterns in Drug-Resistant Acute Lymphoblastic Leukemia Cells and Response to Treatment

Amy Holleman, B.Sc., Meyling H. Cheok, Ph.D., Monique L. den Boer, Ph.D., Wenjian Yang, Ph.D., Anjo J.P. Veerman, M.D., Ph.D., Karin M. Kazemier, Deqing Pei, M.Sc., Cheng Cheng, Ph.D., Ching-Hon Pui, M.D., Mary V. Relling, Pharm.D., Gritta E. Janka-Schaub, M.D., Ph.D., Rob Pieters, M.D., Ph.D., and William E. Evans, Pharm.D.

N Engl J Med 2004;351:533-42.

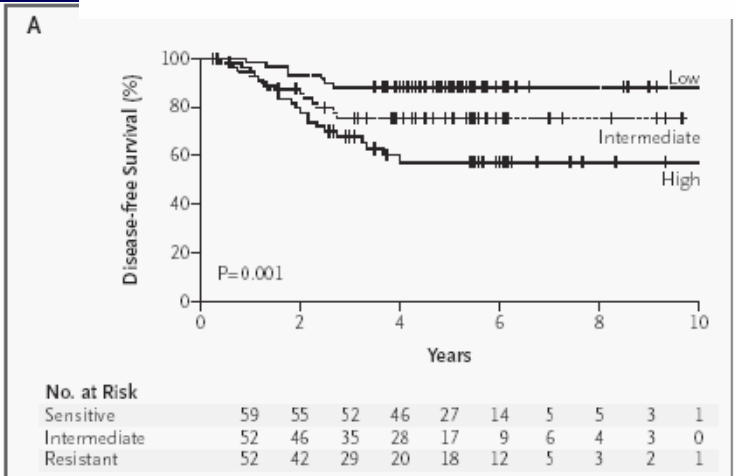


Figure 3. Kaplan–Meier Estimates of Disease-free Survival among 173 Patients in the Original Study Group (Panel A) and 98 Patients in the Validation Cohort (Panel B), According to Whether the Pattern of Gene Expression Indicated Cellular Resistance or Sensitivity to the Four Antileukemic Agents.

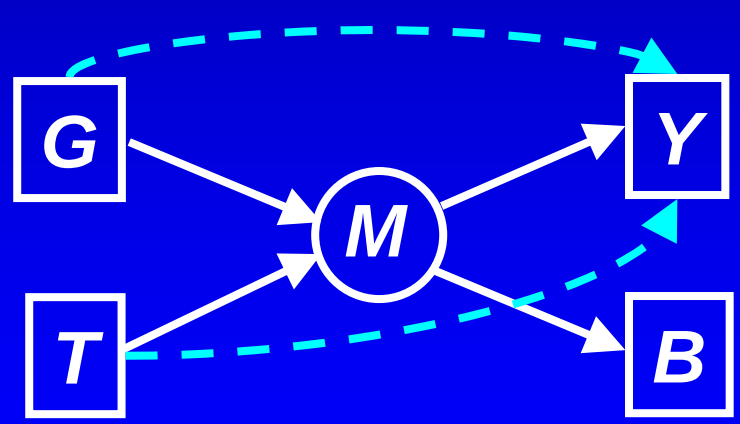
Figure 4. Gene Ontology (GO) Functional Classification of Genes That Discriminated between Drug-Sensitive and Drug-Resistant B-Lineage ALL.

Pathway Analysis with Biomarkers

- **Notation:**

  G **= genes**
  Y **= outcomes**
  T **= treatment**
  M **= intermediate metabolite (unobserved)**
  B **= flawed biomarker for M**

- **Design**
  - **Main study (M): (*G*,*R*,*Y*)**
  - **Substudy (S): (*G*,*R*,*B*)**



- **Model**
  - **Outcomes: $P_\beta(Y|M)$**
  - **Measurement: $P_\sigma(B|M)$**
  - **Metabolic: $P_\alpha(M|R,G)$**

Pathway Analysis

- **Combined analysis of main study and substudy data**
- **Maximum likelihood, integrating over latent variable *M***

$$L(\beta,\alpha,\sigma)=\prod_{i\in S}\int P_{\sigma}(B_i \mid M=m)\,P_{\alpha}(M=m \mid T_i,G_i)\,dm$$

$$\times\prod_{i\in M}\int P_{\beta}(Y_i \mid M=m)\,P_{\alpha}(M=m \mid T_i,G_i)\,dm$$

- **or MCMC, sampling *M***

Conti et al., *Hum Hered* 2003;56:83-93

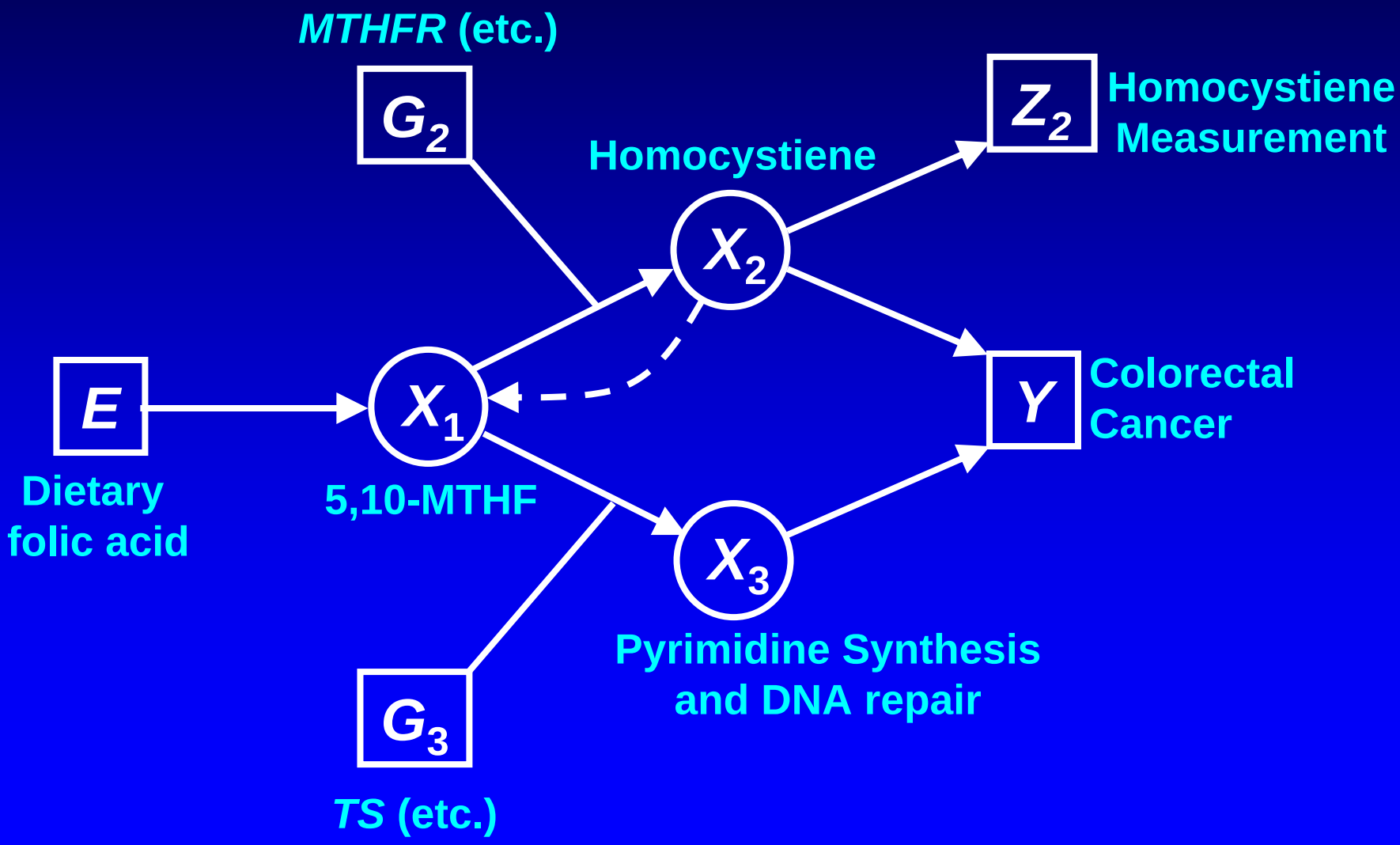
Stratified Sampling for Biomarkers

- **Optimize design by sampling subjects for biomarker measurements by main study data on *T,G,Y***
- **Starting with a clinical trial:**
  - **Observe *Y*|*T***
  - **Sample given *Y,T*; observe *G***
  - **Subsample given *Y,T,G*; measure *M***
- **Starting with an observational study:**
  - **Observe Y,G (and *E*?)**
  - **Sample given *Y,G,E*; apply *T*; measure *M***

Complex Pathways
Example: Folate

Ulrich et al., *Nat Rev Cancer* 2003;3:912-20
Ulrich et al., *Parmacogenet* 2002;3:299-314

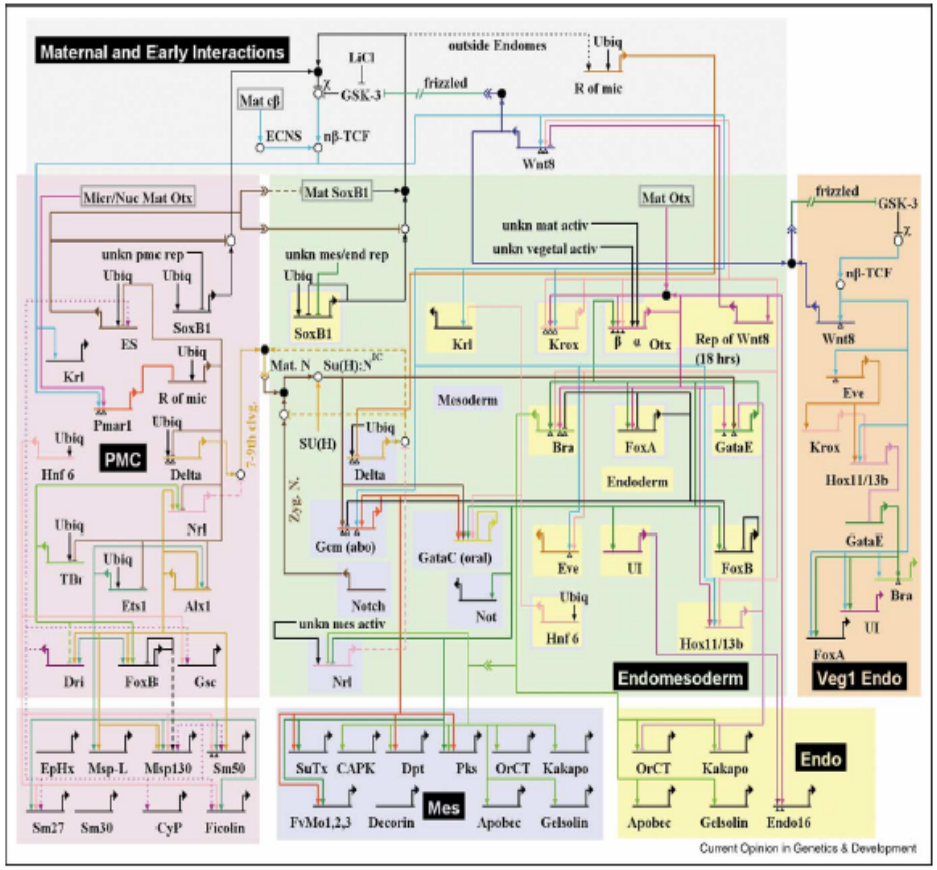
Folate: the Minimalist Version



Topology

- How well do we really understand the structure of a network?
- Incorporate uncertainty in topology into models
  - Bayesian network analysis, e.g., for expression (Friedman et al, J Comp Biol 2000;7:601-20)
  - Basso et al, Nat Genet 2005;38:382-90
- Contribution of systems biology … at the opposite end of detail from molecular epidemiology

Stochastic Boolean Networks



GRN for sea urchin endomesoderm specification: the view from the genome. The architecture of the network is based on perturbation and

Gene regulatory network controlling embryonic specification in the sea urchin

Paola Oliveri and Eric H Davidson

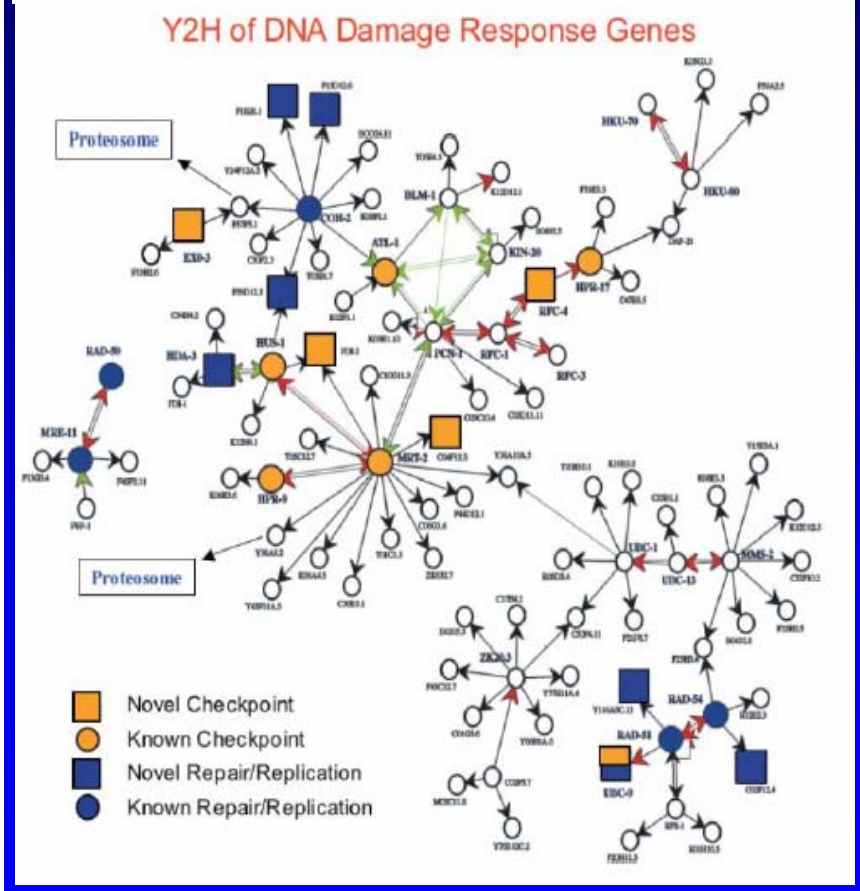
Current Opinion in Genetics & Development 2004, 14:351–360

Human Molecular Genetics, 2005, Vol. 14, Review Issue 2 **R171–R181**
doi:10.1093/hmg/ddi335

Interactome: gateway into systems biology

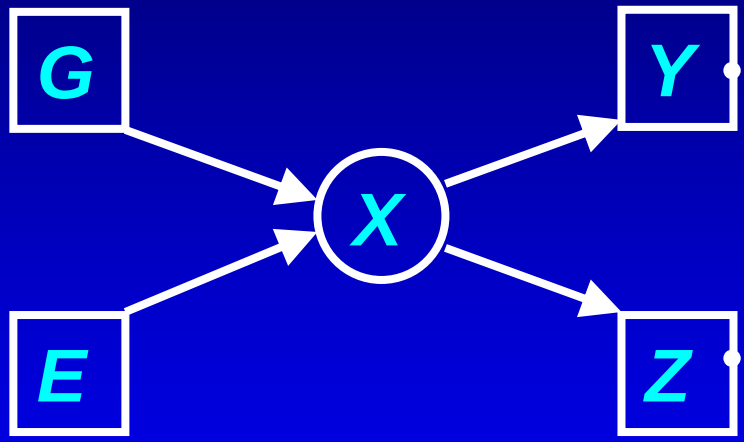
Michael E. Cusick[1,*], Niels Klitgord[1], Marc Vidal[1,2] and David E. Hill[1,2]

[1]Center for Cancer Systems Biology and Department of Cancer Biology, Dana-Farber Cancer Institute, 44 Binney Street, Boston, MA 02115, USA and [2]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA



Causality in Molecular Epidemiology

- We postulate a causal pathway from exposures *E* and genes *G* through a sequence of intermediate steps *X* to a disease.

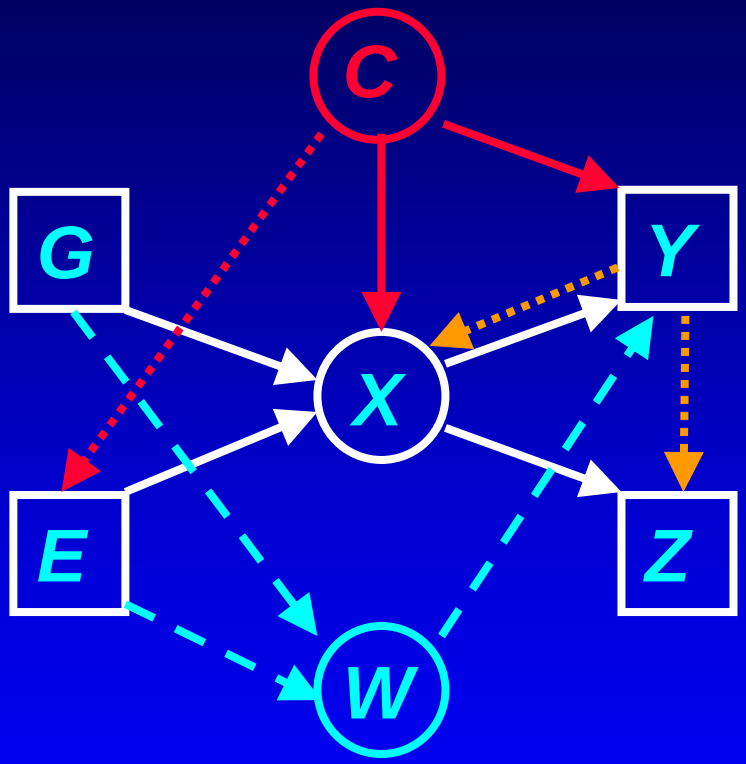


We wish to test the causality of a particular intermediate, as measured by *Z*, on *Y*

By which we mean that holding all other determinants of *Y* fixed, a change in *X* would lead to a change in *Y*

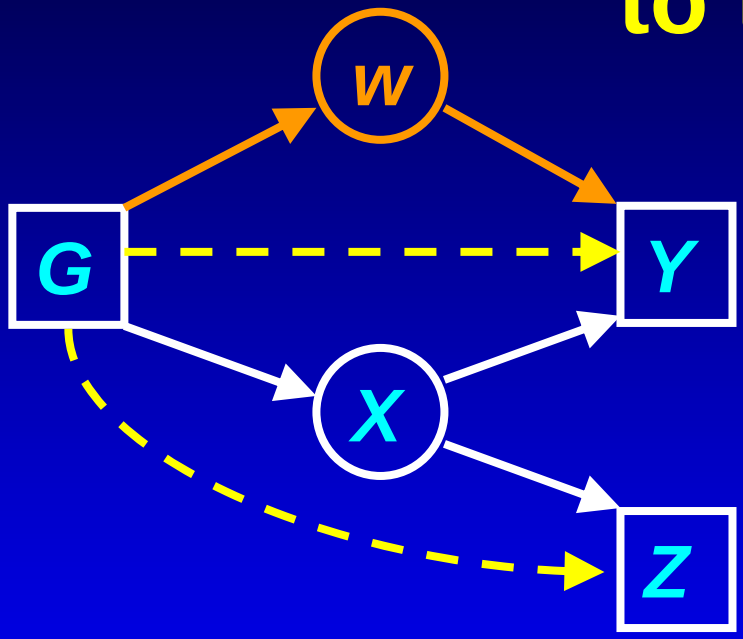
Note: the focus of inference here will be on the causality of *X* (not *G* or *E*) on *Y*, except as *X* is modifiable by *E*

Difficulties in Causal Inference



- Confounding
- Reverse Causation
- Pleiotropy

"Mendelian Randomization" to the Rescue!



- Instead of testing $X \Rightarrow Y$ directly (or more realistically $Z \Rightarrow Y$), test $G \Rightarrow Z$ and $G \Rightarrow Y$ relationships separately
- If both are present, infer a causal connection $X \Rightarrow Y$, because G is not subject to either confounding or reverse causation
- However G could have pleiotropic effects on Y mediated thru W, not X

"Real" Mendelian Randomization

- **Genes are not really assigned randomly across the population, only conditionally on parental mating types**

- **Family-based association studies (e.g., transmission-disequilibrium test (TDT)) exploit this feature:**

$$\Pr(G|Y,G_{par}) = p_{\beta}(Y|G)\, p(G|G_{par}) \,/\, p_{\beta}(Y|G_{par})$$

- **Extension to MR:**

$$\Pr(G,X|Y,G_{par}) = p_{\beta}(Y|X)\, p_{\alpha}(X|G)\, p(G|G_{par}) \,/\, p_{\alpha,\beta}(Y|G_{par})$$

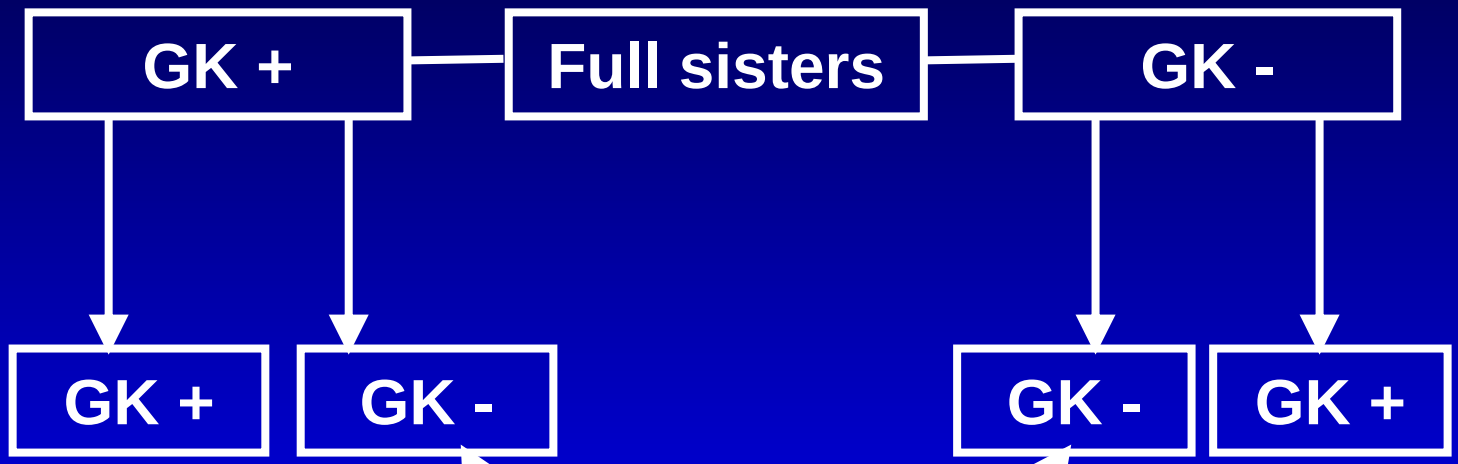
Mendelian Randomization in the Clinical Trials Setting

- Opportunity to randomize both the treatment *T* and the modifiers *G*
- Classical MR assumes *G* are randomly assigned across the population, would treat both *T* and *G* as instrumental variables
  - Models *B* |*T,G* and *Y* |*T,G*
  - Infer causal connection thru *M* if both exist
- Real MR obtains *G* for parents and trial subjects

$$L_i(\beta) = \Pr(G_i \mid Y_i, T_i, G_{P_i}) = \frac{P_\beta(Y_i \mid G_i, T_i)\, P(G_i \mid G_{P_i})}{\sum_g P_\beta(Y_i \mid G_i = g, T_i)\, P(G_i = g \mid G_{P_i})}$$

Double MR to Test a Randomized Environmental Hypothesis

(D.A. Lawlor, *DAE/GDMS* 2005)



Offspring of sisters

Compare BMI:

Equivalent to a randomised trial to slightly higher glucose concentrations *in utero*

Examples

- Candidate gene association study using tag SNPs
- Pathway-based study involving biomarkers
- Genome-wide association study

Genome-wide Association Studies

- Scan of the entire genome to search for genes associated with a trait (or interactions)
- Most scans use multistage design, using commercial chip (~500K SNPs) on first sample to identify promising associations, confirming them on additional samples
- Optimize design with respect to critical value at stage I and allocation of sample size

Wang et al, *Genet Epidemiol* 2006;30:356-68

GWA for Treatment Modifiers

- Select stage I and II samples conditional on treatment and outcome
- Prioritize SNPs for stage II based on test of gene-treatment interactions
  - Based on case-only or case-control comparisons
  - Also based on main effects
  - Incorporate genomic annotation in ranking

Genome-wide discovery of loci influencing chemotherapy cytotoxicity

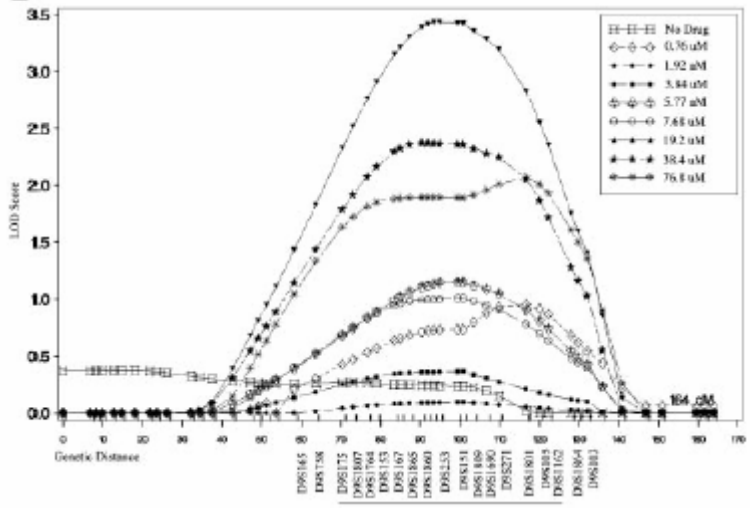
James W. Watters[†], Aldi Kraja[‡], Melissa A. Meucci[†], Michael A. Province[‡], and Howard L. McLeod[†§¶||††]

PNAS | August 10, 2004 | vol. 101 | no. 32 | 11809–11814

Table 2. Regions showing preliminary evidence for linkage using the RCR-derived rate of dose response as the phenotype

Drug	Maximum LOD*	Chromosome[†]
5-Fluorouracil	1.55	9 (99.4 cM)
	1.95	16 (73.98 cM)
Docetaxel	1.24	5 (109.63 cM)
	1.66	6 (100.91 cM)
	1.5	9 (100.74 cM)

*LOD score >1.0 required for preliminary



3. Regions showing supportive evidence for linkage using ual doses of drug as separate phenotypes

	Maximum LOD*	Chromosome[†]	Approximate 1 LOD interval
ouracil	3.44	9 (94.85 cM)	D9S175–D9S1162
xel	2.21	5 (97.21 cM)	D5S502–D5S1965
	2.73	9 (94.85 cM)	D9S175–D9S1162

ore >2.0 required for supportive evidence of linkage.
: map location of maximum LOD score within QTL peak.

Reality Check: Sample Size Needs

- Candidate genes, pathways, genome-wide
- Therapeutic or prevention trials
- Main effects or treatment modifiers
- Power calculations by Quanto
  - Gauderman, Am J Epidemiol 2002;155:478-84 (GxG)
  - Gauderman, Stat Med 2002; 21:35-50 (GxE)

http://hydra.usc.edu/gxe

Candidate Treatment Modifier Gene

- **Sample sizes needed to detect interaction effect at α = .05, 1-β = .90, single stage design**
  - $p(T) = 0.5$, $MAF = 0.2$ (dom), $RR_{T|G=0} = 0.9$, $RR_{G|T=0} = 0.9$

RR_{GxT}	Cases (cohort size) needed	
	Therapeutic $p(Y) = 0.5$	Prevention $p(Y) = 0.01$
0.3	271 (542)	142 (14K)
0.5	789 (1,578)	408 (40K)
0.7	2,900 (6,000)	1,513 (150K)
0.9	33K (67K)	17K (1.7M)

Candidate Treatment Modifier Gene

- **Sample sizes needed to detect interaction effect at α = .05, 1-β = .90, single stage design**
  - $p(T) = 0.5$, $MAF = 0.2$ (dom), $RR_{T|G=0} = 0.9$, $RR_{G|T=0} = 0.9$

RR_{GxT}	Cases (cohort size) needed	
	Therapeutic $p(Y) = 0.5$	Prevention $p(Y) = 0.01$
0.3	271 (542)	142 (14K)
0.5	**789 (1,578)**	**408 (40K)**
0.7	2,900 (6,000)	1,513 (150K)
0.9	33K (67K)	17K (1.7M)

Candidate Treatment Modifier Gene: Two-Stage Design

- Consider prevention trial scenario with $p(Y) = .01$ and $RR_{GxE} = 0.5$
- Nested case-control study with 1:1 matching within treatment arms

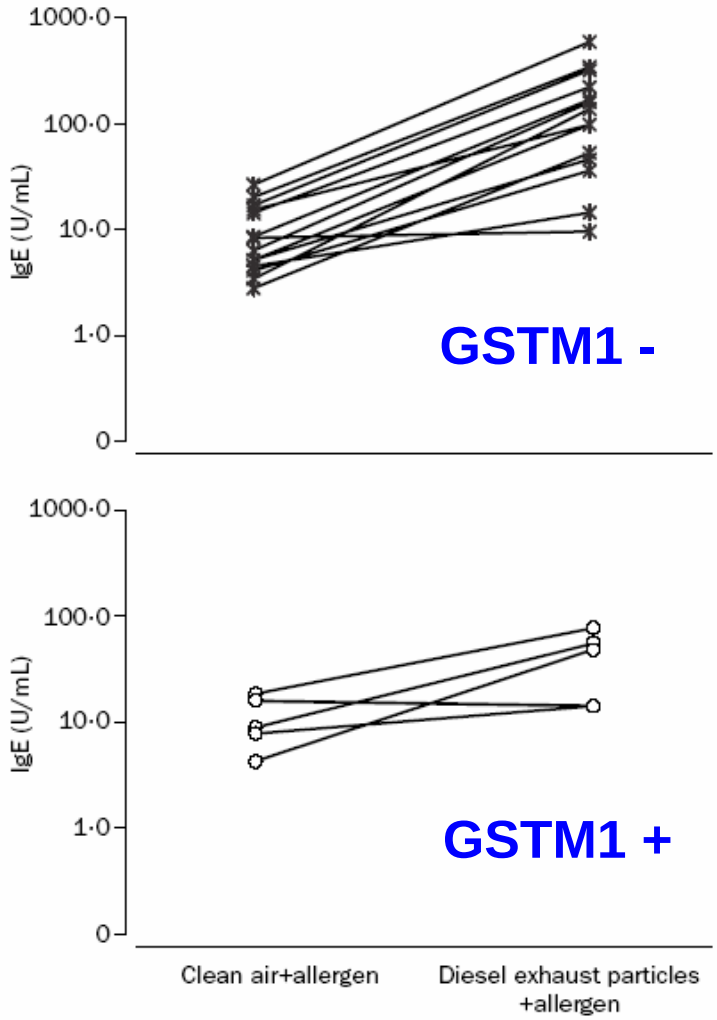
	Cohort	Cases	Controls
1-stage	14K	140	13,860
2-stage	27K	270	270

- Somewhat greater advantage if stages I and II were analyzed jointly

Randomized Trial within an Epidemiologic Cohort

- **Cohort study:** observe *Y*, measure *G*
- **Challenge experiment:**
  sample based on *Y* and *G*,
  assign *T* (randomized or crossover),
  measure acute response *R*
- ***Example:*** Children's Health Study

Challenge Studies:
GSTM1 x *GSTP1* in Allergic Response to DEP



Nasal allergen-specific IgE response to allergens plus clean air and allergen plus diesel exhaust particles for *GSTM1* absent (upper) and present (lower) genotypes

GSTM1	GSTP1	*N*	ΔIGE
+	I/I	2	26 (6.7– 45)
+	I/V	3	49 (-1.5 – 61)
–	I/I	11	137 (29 – 511)
–	I/V	3	9.1 (1.0 – 46)

Gilliland et al, *Lancet* 2004;363:119-25

Biomarker for Pathway

- **Stage I:** Prevention trial, assign T, observe Y
- **Stage II:** Nested case-control study, sample based on Y, T, observe G_1 and G_2
- **Stage III:** Biomarker substudy, sample based on Y, T, G, observe M

	Cases	Controls
Stage I	270	27K
Stage II	135,135	135,135
Stage III	10 x 8 = 80	10 x 8 = 80

Effect	Min det r^2
$Y\|M$	1.7/SD
$M\|T,G$	6.5%

Genome-wide Association Scan for Treatment Modifying Genes

- Same model parameters as for candidate gene study
- Two-stage genotyping strategy with genomewide significance level .05 ($1x10^{-7}$ per SNP) and 90% power

	Cases / controls		
	RR_{GxT} = 0.3	RR_{GxT} = 0.5	Markers
Stage I	600	1700	500K
Stage II	600	1700	5K

Perspectives

- **Well established statistical theory**
- **Increasingly used in epidemiology and genetics, *but underdeveloped in pharmacogenetics***
- **Particularly useful for incorporating pharmacokinetic / pharmacodynamic models**
- **Sample size requirements for detecting interactions are large**

Nature Genetics **38**, 68 - 74 (2006)
Published online: 10 November 2005; |
doi:10.1038/ng1692

A variant of the gene encoding leukotriene A4 hydrolase confers ethnicity-specific risk of myocardial infarction

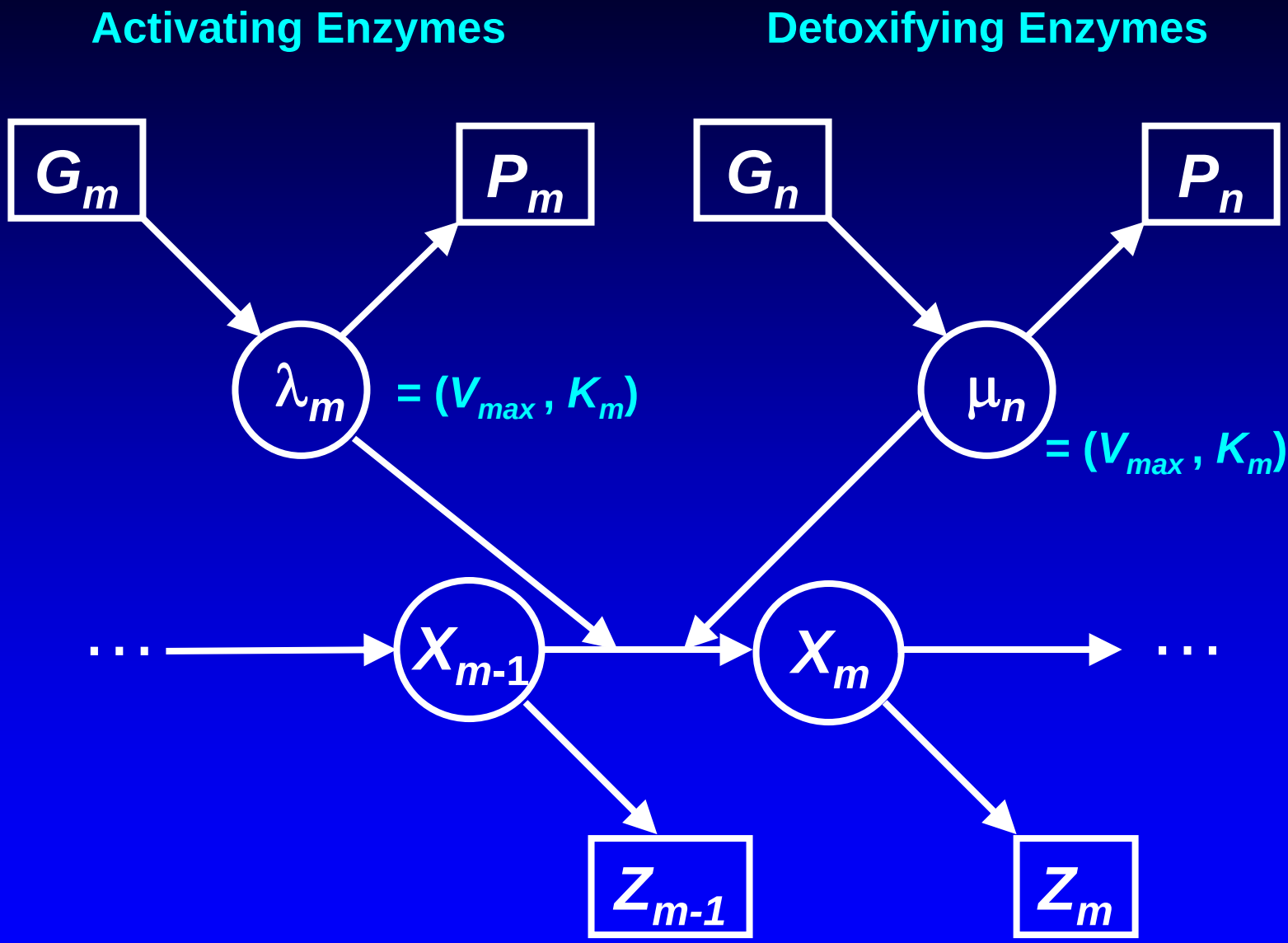
Anna Helgadottir[1], Andrei Manolescu[1], Agnar Helgason[1], Gudmar Thorleifsson[1],

Maternal-Fetal Interactions

- Standard TDT analysis is $\Pr(G_o \mid G_m, G_f, Y_o=1)$
- Suppose:

$$\Pr(Y_o=1 \mid G_o, G_m, G_f) \propto \exp(\beta_1 G_o + \beta_2 G_m + \beta_3 G_m G_o)$$

- Parameters β_1 and β_3 are estimable; β_2 is not
- Instead use $\Pr(G_o, G_m, G_f \mid MT, Y_o=1)$
- Now all three parameters are estimable, assuming $\Pr(G_m, G_f) = \Pr(G_f, G_m)$.
  No controls needed


Activating Enzymes
Detoxifying Enzymes
G_m
P_m
G_n
P_n
λ_m
$= (V_{max}, K_m)$
μ_n
$= (V_{max}, K_m)$
X_{m-1}
X_m
Z_{m-1}
Z_m

Contexts

- Candidate gene known to be functionally relevant to agent
- Biomarkers to inform about pathway
- Genome-wide search for modifier genes

Design alternatives

- Clinical trial within observational study
  - Cohort study: sample based on E, store DNA, observe Y
  - subsample based on Y,E, measure G

    OR
  - Case-control study: sample based on Y, observe E, G
  - Subsample based on Y,E, G, assign T, observe M
- Genetic study within a clinical trial
  - Assign T, observe Y, M
  - Subsample based on Y,M,T
  - Observe G
- Observational study with countermatching (WECARE)
  - Observe T, Y
  - Nested case-control sample based on Y, countermatching on T
  - Observe G

ATM Gene Screening

- **ATM Gene Analyses**
  - Conducted in 4 labs
  - Staged approach: DHPLC followed by Direct Sequencing
  - All conditions, primers standardized across labs
  - Inter- and Intra-lab QC implemented

(Bernstein, … , Concannon, *Hum Mut* 2003)

- **RCT for diabetes prevention in 3234 overweight people with elevated fasting glucose**
- **Randomized to lifestyle intervention, metformin, or placebo**
- **58% reduction in diabetes risk on lifestyle intervention and 31% reduction on metformin over 3 years**
- **Genotyped for two common polymorphisms in *TCF7L2* associated with NIDDM**

The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 JULY 20, 2006 VOL. 355 NO. 3

TCF7L2 Polymorphisms and Progression to Diabetes in the Diabetes Prevention Program

Jose C. Florez, M.D., Ph.D., Kathleen A. Jablonski, Ph.D., Nick Bayley, B.A., Toni I. Pollin, Ph.D., Paul I.W. de Bakker, Ph.D., Alan R. Shuldiner, M.D., William C. Knowler, M.D., Dr.P.H., David M. Nathan, M.D., and David Altshuler, M.D., Ph.D., for the Diabetes Prevention Program Research Group

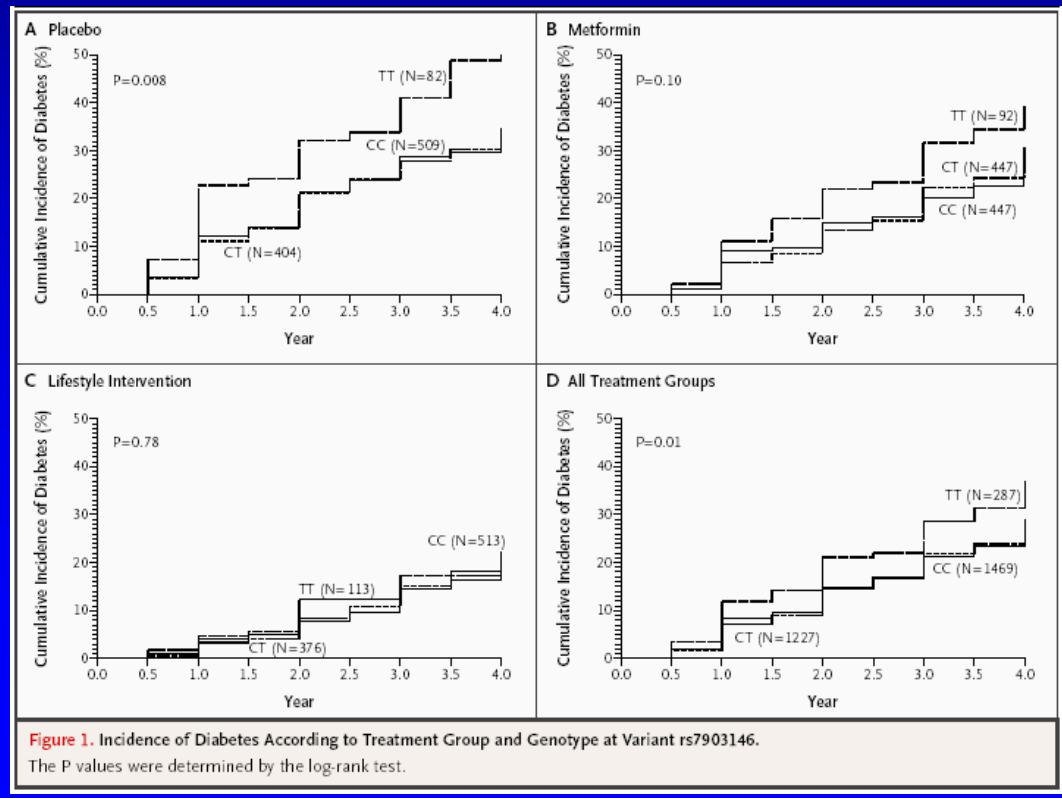


Figure 1. Incidence of Diabetes According to Treatment Group and Genotype at Variant rs7903146.
The P values were determined by the log-rank test.

Genetic Association Studies in the Context of Clinical Research

- Identify and characterize genes that modify response to phamacologic agents or other interventions
  - Preventive or therapeutic (phase I, II, III)
- Approaches:
  - Clinical trials with genetic add-ons
  - Nested challenge or treatment studies within population-based observational studies
  - Integrating separate observational and randomized studies