

Gene

by environment interaction in linkage and association

David Duffy 20040921

*Queensland Institute of Medical Research
Brisbane, Australia*



Introduction

All gene action is with respect to a genetic and environmental background. The questions are rather:

Can detailed modelling of $G \times E$ (or epistasis) improve power for gene discovery?

Does $G \times E$ for a given gene underlie surprising biology?

Is a given trait **labile** or **nonlabile** (can we observe response of a genotyped individual in the different environments)?

What sample sizes are needed?

ERCC2 genotype, smoking, and lung cancer

ERCC2 (19q13.2) codes for a DNA repair protein (Type D xeroderma pigmentosum).

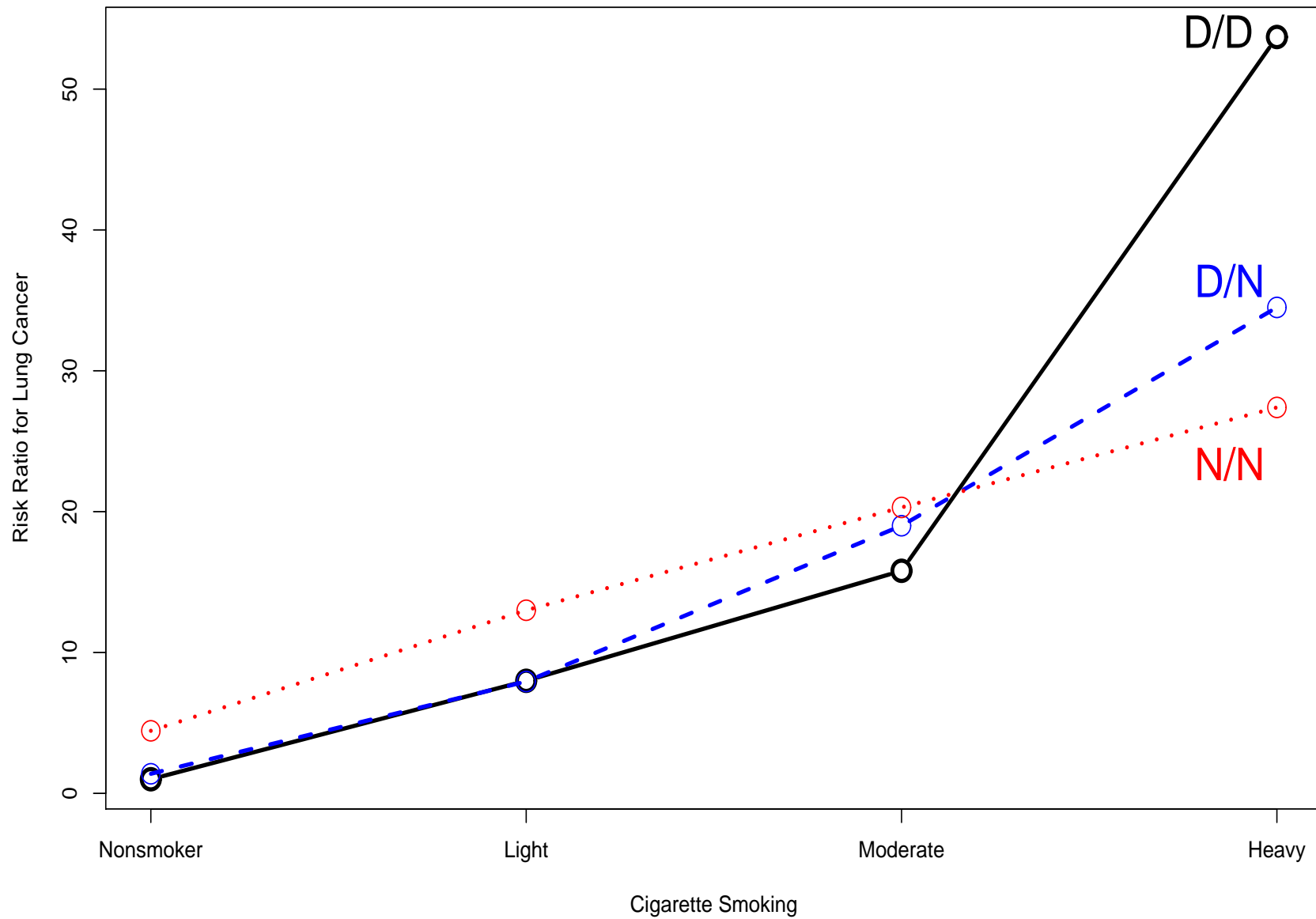
ERCC2*312Asp reported to increase risk in light but not heavy smokers.

ERCC2*751Gln reported to increase risk in female smokers but not male smokers.

ERCC2*312Asp suggested to increase rate of p53 somatic mutations in smokers.

One **large** case-control study (Zhou et al 2002) found:

ERCC2 genotype, smoking, and lung cancer



Family v. individual level

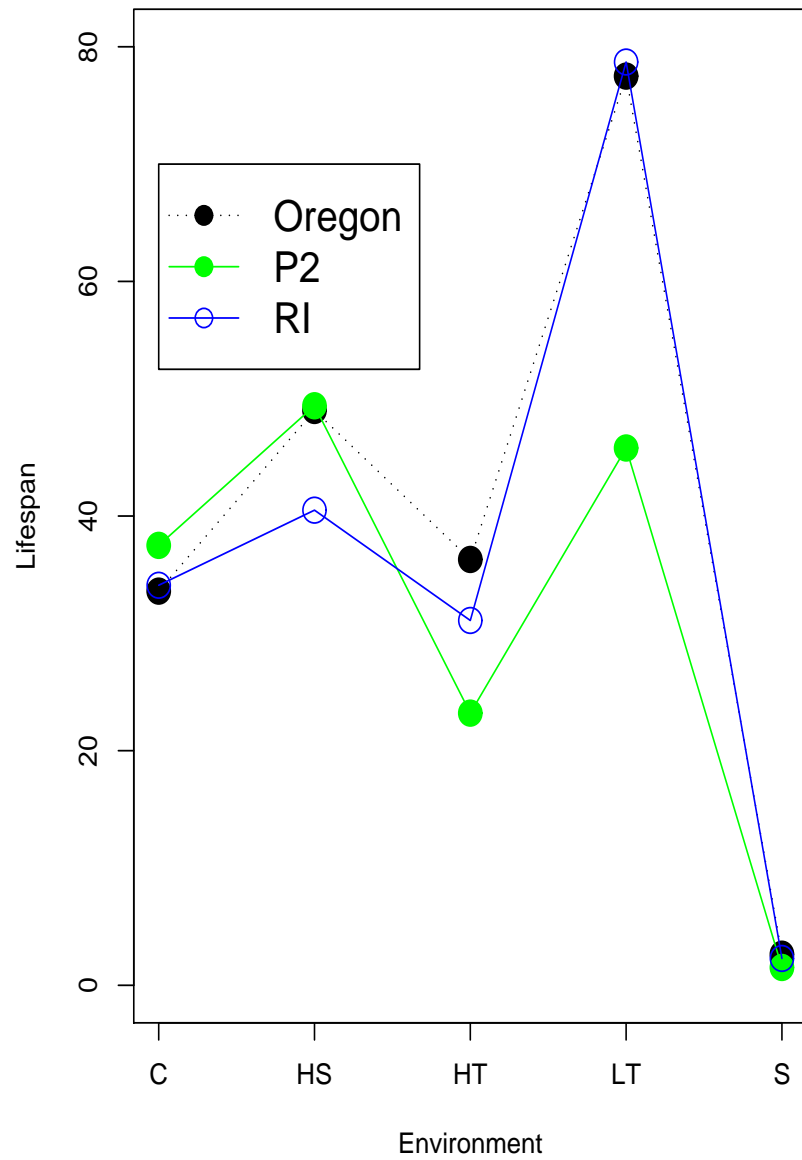
Studies of $G \times E$ in other species are often in experimental setups, where a particular line cross is examined in different environments. The **moderator variables** usually vary only at the family level.

Alternatively, the moderator variable might vary from individual to individual within a family. With respect to the autosomes, the sex of the individual is a highly variable random environmental event.

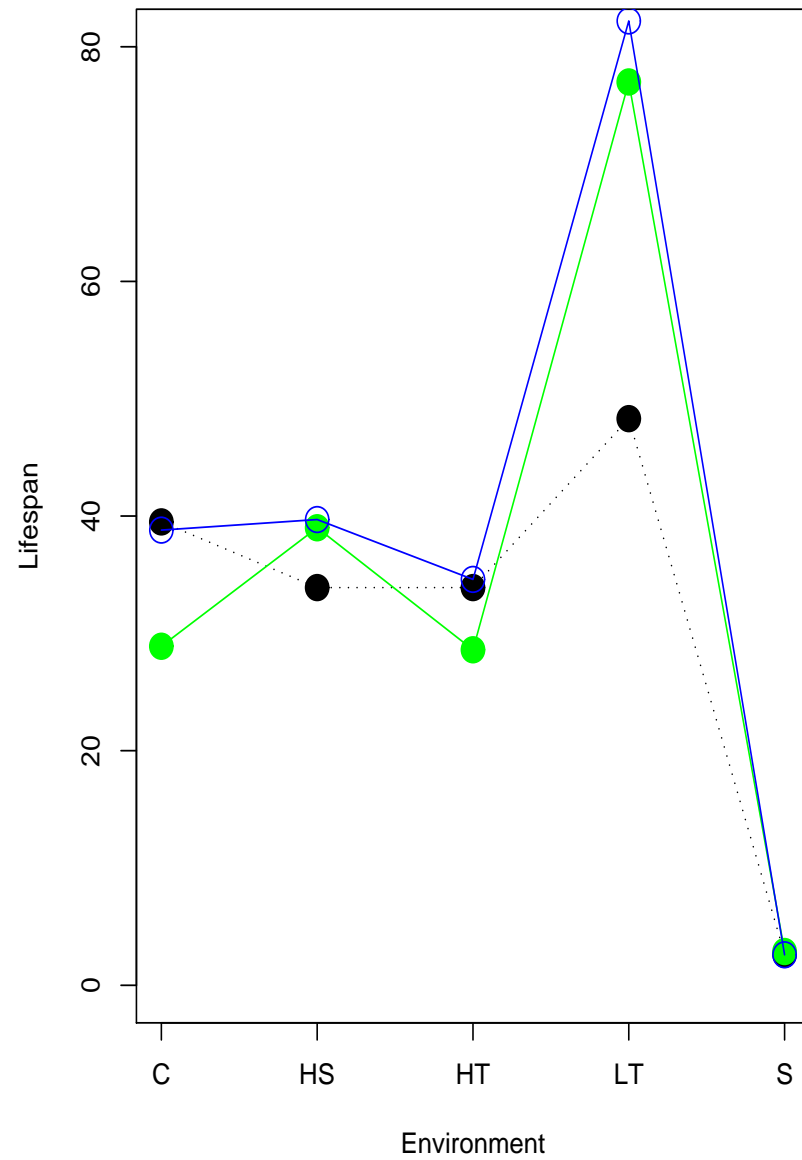
In studies on humans, both classes of moderator variable are commonly measured, and are commonly continuous rather than discrete in nature.

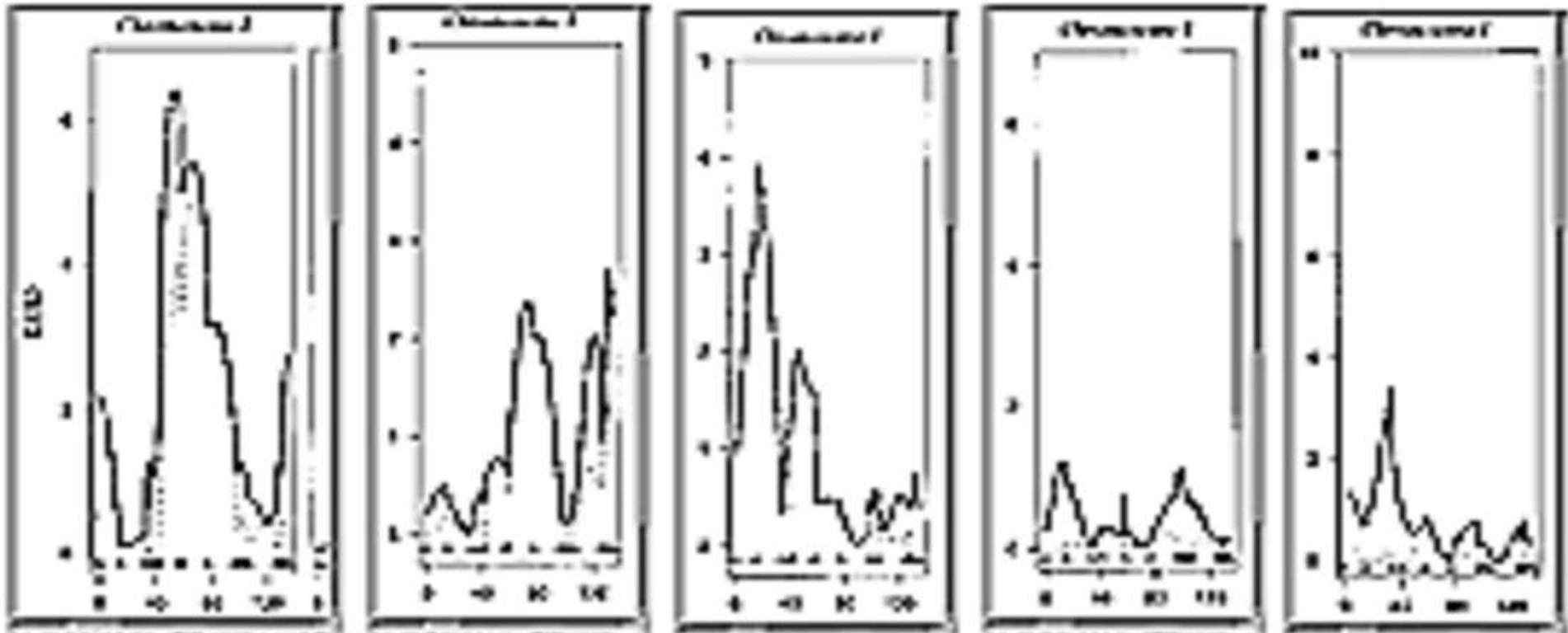
The moderator may be unobserved. One classic example is amount of contact between twins or relatives (moderating the effects of C).

Males



Females





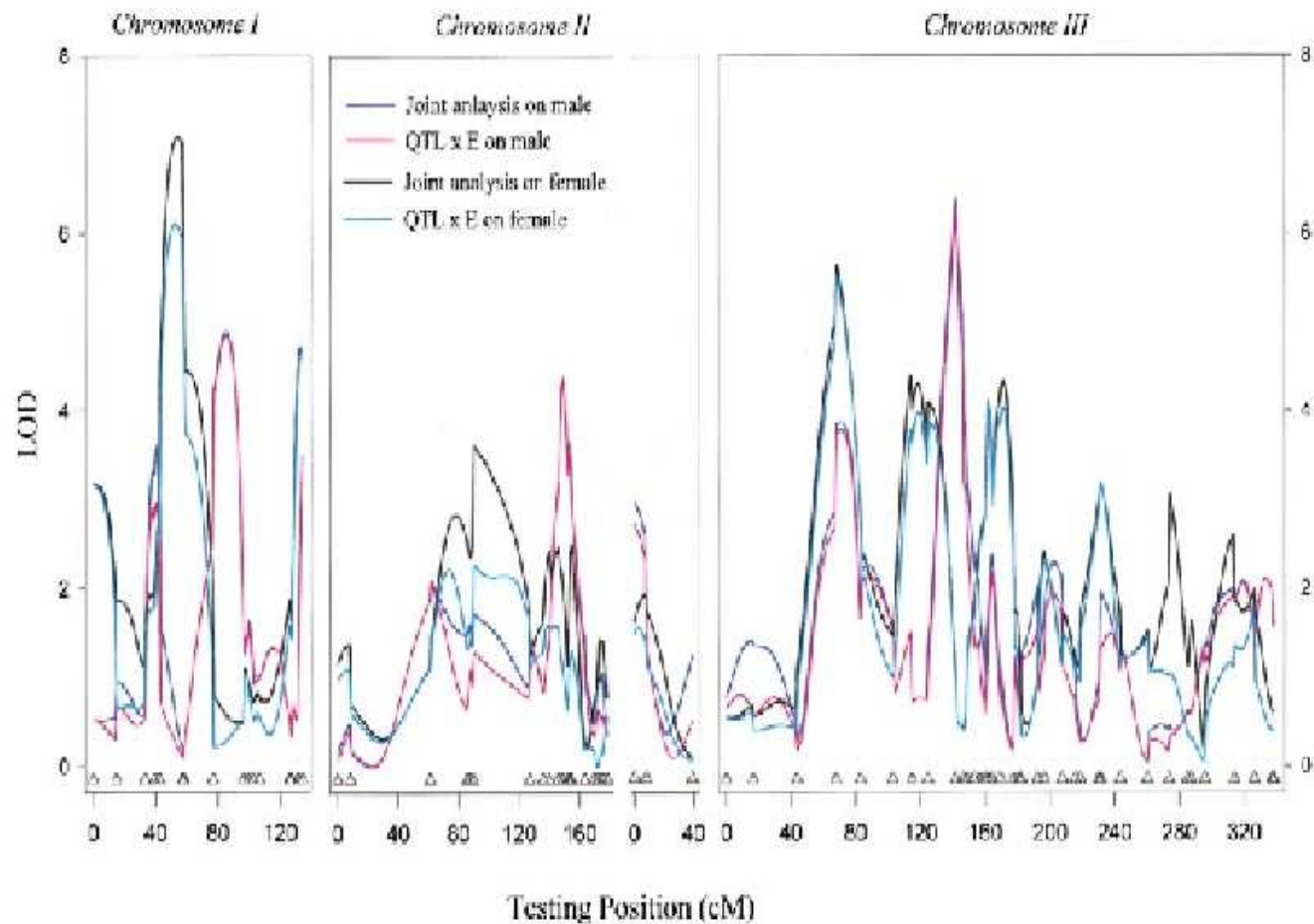
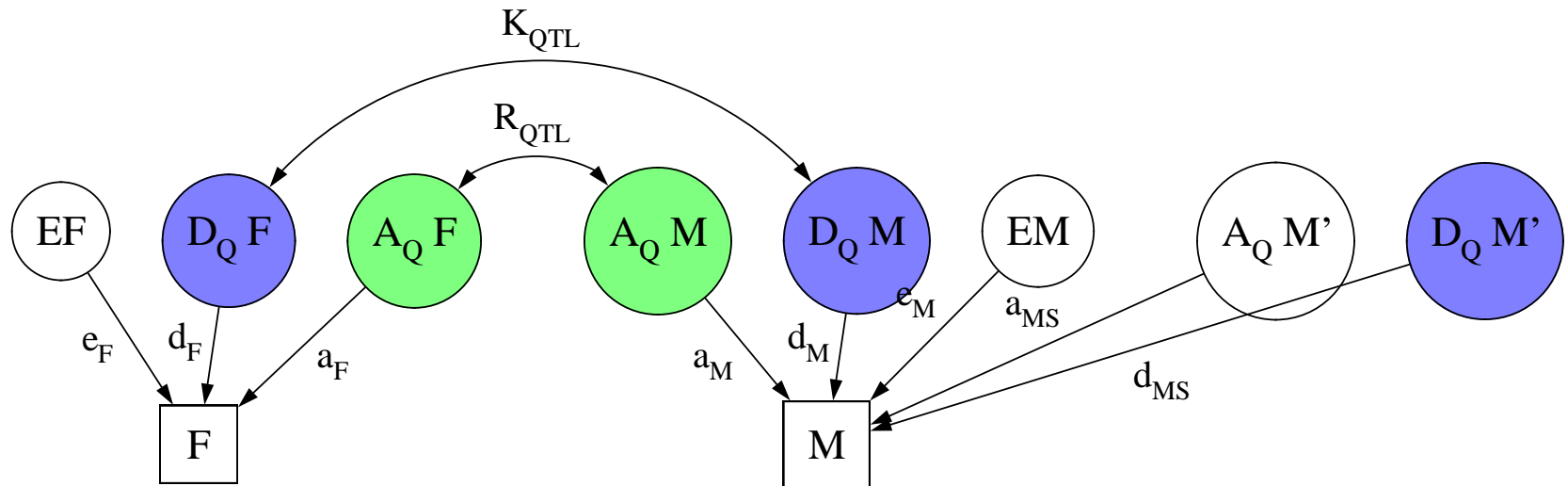


FIGURE 6.—Plot of LOD scores against estimated map position from multiple-trait composite interval mapping of male and female life span. The joint analysis considers the control, heat shock, high temperature, and low temperature environments. Total QTL and QTL \times environment interaction LOD scores are shown. (Vieira et al. Genetics 2000; 154:213-227)

Sex-limitation is a common example of $G \times E$



In the polygenic sex-limitation model, **different** sets of genes can act on the trait in the two sexes eg $h_M^2 = h_F^2$, but $r_g = 0$.

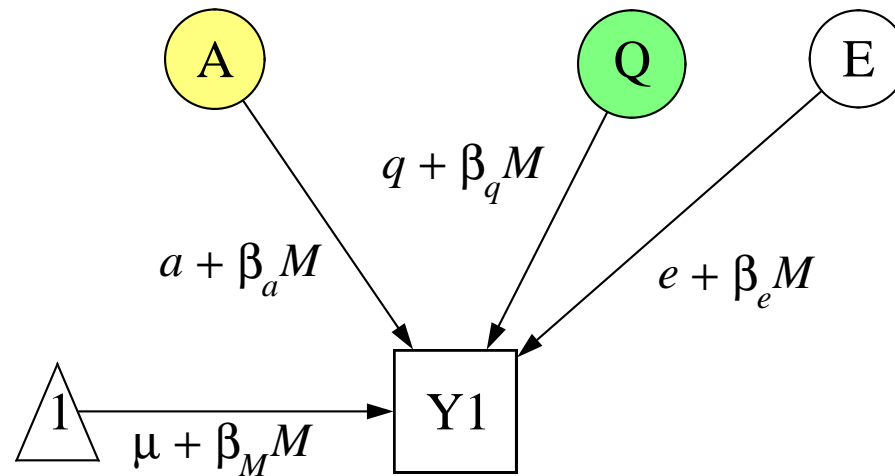
Sex-specific genotypic means or penetrances at a single QTL can mimic this type of pattern, but the solutions require overdominance in one sex.

Sex-limitation is a common example of $G \times E$

The GAW10 simulated data included a sex-limited QTL. Towne et al (1997) found the ELOD for one QTL increased from 1.5 to 1.9 with inclusion of a $Q \times \text{Sex}$ term.

Continuous moderator models ($G \times E$)

Nice papers by Purcell (2002) and Sham & Purcell (2002) cover this material:



Effects of the measured modifier variable on both means and variances are being modelled.

Hypertension in migrant Greeks

Hopper and coworkers (1992) describe blood pressure in 628 individuals on the island of Levkadis and 846 of their relatives who had migrated to Melbourne, Australia (total of 204 pedigrees). Migrants eat and work quite differently in Australia: drinking one-third the amount of alcohol, weighing more, and having higher blood pressures as they age.

Group	V_G	V_C	V_E
Male Migrant	25.8 (4.9)	0.0 (2.8)	74.5 (7.5)
Female or Non-migrant	25.8	0.0	48.3 (4.7)

Regression of total variance of dBP on age in males	0.77 (0.25) per year
Regression of total variance of dBP on age in females	0.18 (0.24) per year
Regression of V_G on age	0.23 (0.31) per year
Regression of V_E on age	0.16 (0.31) per year

Simulated QTL data based on Hopper et al (1992)

I've simulated 300 nuclear families (3 children per family) based on the above data: one QTL explains half the genetic variance.

1. Simulate markers along a map for pedigrees of the required size.
2. Simulate the main QTL as linked to markers on the map
3. Simulate five unlinked polygenes
4. Simulate the phenotypes based on the genotype.
5. Add in the effects of the two moderators to E for the phenotypes.
(5b. Use rejection sampling to mimic ascertainment.)
6. Analyse the resulting data.

Simulated QTL data II

Script for the *twinsim* program:

```
set map 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 100
set qtl 51 0.25
set va 13
set vq 13
set ve 0
set mzfamilies 0
set halvesib_families 50
set nfamilies 300
set nsibs 3
set ple 1
write fam300.ped
```

Simulated QTL data III

Script for the *sib-pair* program:

```
set locus mztwin qua
set locus dBP qua
set loc marker001 marker      10.00
set loc marker002 marker      15.00
...
set loc marker009 marker      50.00
set locus qtl marker          51.00
set loc marker010 marker      55.00
set loc marker011 marker      60.00
...
set loc marker018 marker     100.00
read pedigree fam300.ped
set ple -1
run
```

Simulated QTL data IV

Script for the *sib-pair* program:

```
set loc age qua
set loc env qua
set loc mal aff
mal=mal
env=0
if (mal and rand<0.43) then env=1
if (isnon) then age=30+30*rand
if (isfou) then age=60+10*rand
dBP = 80 + env*6 + dBP + \
      sqrt(age*(0.2+mal*0.5)+48)*rnorm + env*5*rnorm
drop qtl
write loc merlin merlin.dat
write loc merlin merlin.map
write merlin.ped
```


Simulated QTL data V

60% of replicates pass the Shapiro-Wilk's test for normality

Effect of migration on male total variance easily detected

ELOD small: 0.5

ELOD for VC linkage analysis including migration smaller

Association designs for $G \times E$

In the absence of complicating population stratification, a conventional case-control design with unrelated controls is the most efficient design for detecting main genetic effects, but various family based designs are superior for picking up $G \times E$ (Witte et al 1999; Siegmund & Langholz 2001).

Case-only and case-combined-control studies are two designs being explored.

A simple cotwin-control example

A dichotomous trait is determined by a single major locus and a dichotomous environmental exposure (or measured genotype) under the following model:

Risk Factor	Gene Penetrance			Mean Risk	
	AA	AB	BB		
Exposure Positive	f_2	f_1	f_0	$p^2 f_2 + 2p(1-p)f_1 + (1-p)^2 f_0$	R_1
Exposure Negative	f_0	f_0	f_0	f_0	R_0

The probability of exposure is E (exposure is uncorrelated in families).

A conventional study using unrelated cases and controls will measure the mean population risk ratio R_1/R_0 .

A simple cotwin-control example

Prospectively:

Exposure Level of pair	Marginal Risk		Recurrence Risk
	Twin 1	Twin 2	
Both Unexposed	R_0	R_0	R_0
Twin 1 exposed	R_1	R_0	R_0
Both twins exposed	R_1	R_1	$R_1 + \text{Cov}/R_1$

A simple cotwin-case-control example

The equivalent matched cotwin-case-control study:

	Unaffected Member	Affected Member
	Exp Pos	Exp Neg
Exposure Positive	$E^2[R_1(1 - R_1) - Cov]/T$	$E(1 - E)[R_0(1 - R_1)]/T$
Exposure Negative	$E(1 - E)[R_1(1 - R_0)]/T$	$(1 - E)^2[R_0(1 - R_0)]/T$

where the total probability $T = R_0(1 - R_0) + E(R_1 - R_0)^2 + (1 - E)R_0(1 - R_0)^2$.

Testing differences in exposure concordances between two strata of pairs (MZ v DZ) will allow one to test $Cov=0$,