Economics & Genetics Lecture 2

DR. NIELS RIETVELD, NRIETVELD@ESE.EUR.NL

OFFICE MANDEVILLE T18-29

Tutorial 1

How many decimals should I report? (for the paths, “umxACE” and “umxModify” provide 3 decimals, “umxSummary” provides only 2 decimals)

- A consistent number (e.g., 3 decimals: coefficients from “umxACE” and “umxModify”, and summary statistics, e.g., *p*-value, from “umxSummary”)

Today's agenda

Main questions:

- What are the promises of genoeconomics?
- The basics of genetics: What is a "SNP"?
- How can we use Mendel's laws to understand natural selection on socio-economic traits?

Literature:

- Benjamin et al. (2012): "The promises and pitfalls of genoeconomics"
- Beauchamp (2016): "Genetic evidence for natural selection in humans in the contemporary United States"

Economics & Genetics Lecture 2 – Part 1

MAIN QUESTIONS 1 & 2

Last week…

Heritability: The proportion of observed differences in a trait among individuals from a certain population that is due to genetic differences among these individuals

- Twin study: Exploit zygosity differences using ACE models (Falconer's formula)

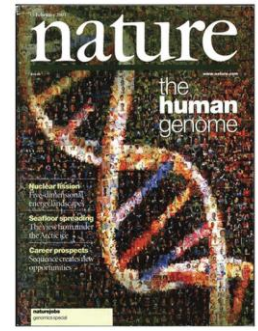
"From twins to genes"

Genes are latent in twin models: We do not observe them, but we infer their effects using MZ and DZ twin differences

- We can't infer the precise effect of a specific gene, and thus we cannot analyze with twin studies how a specific person with a specific genetic profile reacts to environmental (e.g., policy) changes (*Manski*)

Now that the human genome has been sequenced (2001), we can!

In essence, genoeconomics is about using molecular genetic information in economic models



The 4 promises of *genoeconomics* (2012)

Irrespective of their position in the Taubman vs. Goldberger debate, economists started to get excited about the use of genetic information because of four promises (Benjamin et al., 2012, section 4.1):

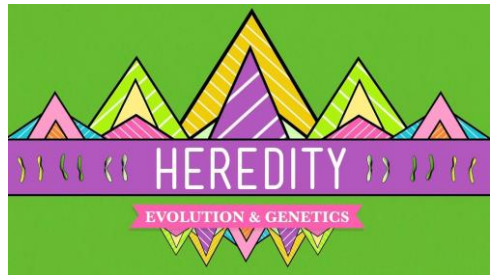
1. *Biological mechanisms:* Better insights into what underlies behavior of interest, e.g. the "nature of preferences", why people make particular (occupational) choices, etc.
2. *Direct measure of construct that is otherwise difficult to measure:* For example measure FTO gene instead of asking about preferences for fatty food
3. *Targeted interventions:* Early detection based on genetic information ("genetic profiling")
4. *Genes as control variables and/or instrumental variables:* Improvement of existing economic models (less omitted variable bias, possibility to infer causal relationships)

Some promises are far-fetched, some impose ethical dilemma's, but some have already been realized

- *Lectures 3-6*

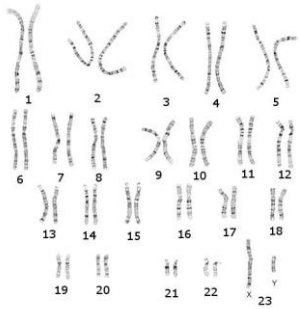
Genetics: The basics

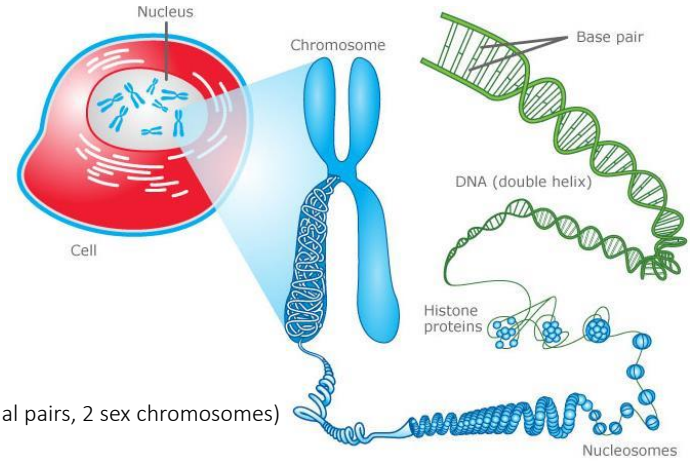
Check out this youtube course



https://www.youtube.com/watch?v=CBezq1fFUEA

DNA (Deoxyribonucleic acid)





Human genome:

- 46 chromosomes (22 autosomal pairs, 2 sex chromosomes)

DNA: Deoxyribonucleic Acid

Deoxyribose = a sugar

Nucleic = stored in cell nucleus (core of cell)

Acid

Sugar binds to **nucleotide base**:

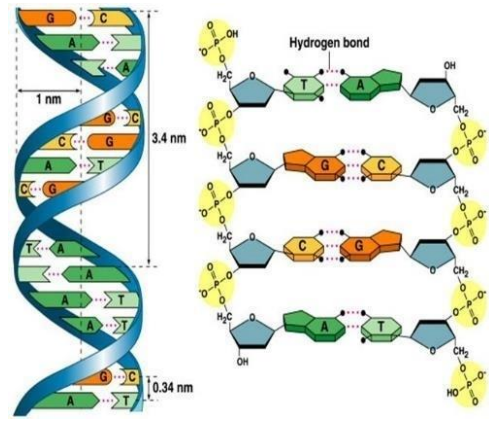
A: Adenine

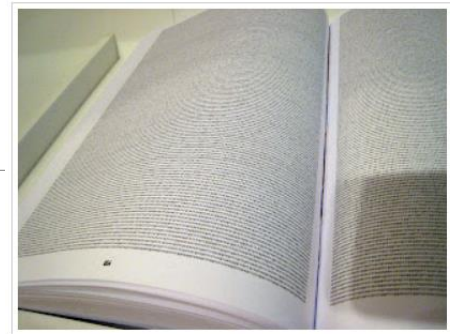
C: Cytosine

G: Guanine

T: Thymine

- The “code of life” is written with 4 letters (~3 billion)





DNA

THE DNA THAT MAKES US HUMAN

99.6% of DNA is equal across humans

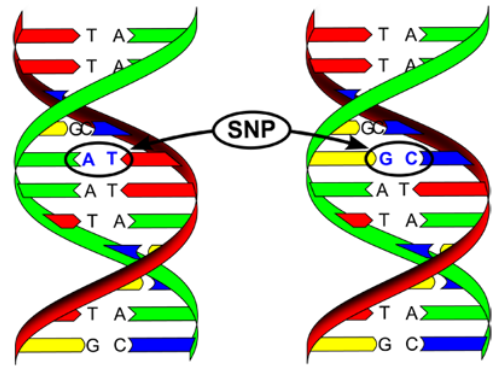
THE DNA THAT MAKES HUMANS DIFFER

The remainder (0.4%) differs across humans

SNPs ("snips")

Single **N**ucleotide **P**olymorphism (**SNP**)

- Most common form of genetic variation
- 1 nucleotide difference between individuals

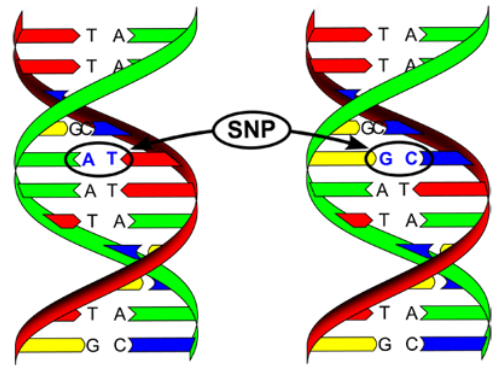


SNPs ("snips")

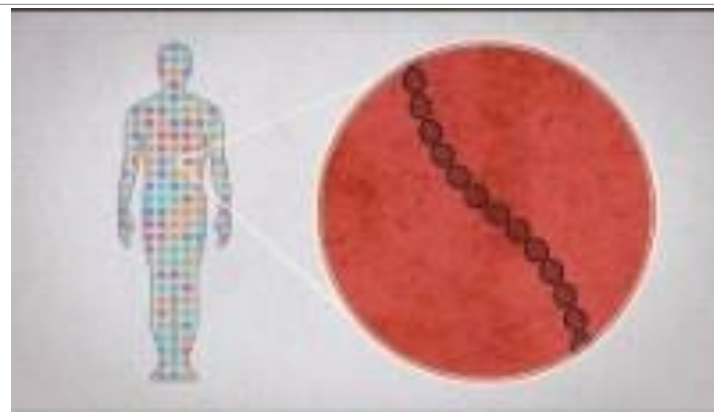
- Started as a "random" mutation in some generation
  - Became more frequent in population
  - Alleles = The possible nucleotides at a SNP location (2 possibilities)
- A pairs with T, G pairs with C → Reading out one "strand" of the double helix suffices
  - A and G are the alleles

Take one of the alleles as "reference allele"

- Count how many you have one each *pair* of chromosomes (0, 1, or 2)



Sequencing the human genome



https://www.youtube.com/watch?v=MvuYATh7Y74

Labelling SNPs (Tutorial 2: Example.bim)

Chromosome: 1-22, X (23), Y (24)

rs-number: "Reference SNP" in dbSNP (SNP database), no logical numbering

(Genetic distance, from previous SNP, 0=Missing)

Basepair: Position of the SNP on a chromosome

Possible alleles

Example.bim

1	1	rs4648808	0	2030796	T	C
2	1	rs2503705	0	2066601	G	A
3	1	rs2172900	0	2212983	T	C
4	1	rs10910111	0	2532513	A	G
5	1	rs12565258	0	2725680	G	C
6	1	rs2485944	0	2873718	T	C
7	1	rs10797373	0	2893916	G	A
8	1	rs2376813	0	3004553	G	A
9	1	rs16823548	0	3130672	A	G
10	1	rs2500289	0	3202806	G	A

Economics & Genetics Lecture 2 – Part 2

MAIN QUESTION 3

Inheritance

- How can we use Mendel's laws to understand socio-economic differences?

- Mendel: The father of genetics
  - Gregor Mendel (1822-1884), Augustinian friar and abbot of St. Thomas' Abbey in Brno with interest in biology
  - His pea plant experiments conducted between 1856 and 1863 established many of the rules of heredity, now referred to as the laws of Mendelian inheritance

Mendel's laws

Without measuring genes, Mendel deducted that genes come in pairs and are inherited as distinct units, one from each parent

1) The Law of Segregation: Each inherited trait is defined by a gene *pair*. Parental genes are randomly separated to the sex cells so that sex cells contain only one gene of the pair. Offspring therefore inherit one genetic allele from each parent when sex cells unite in fertilization

2) The Law of Independent Assortment: Genes for different traits are sorted separately from one another so that the inheritance of one trait is not dependent on the inheritance of another

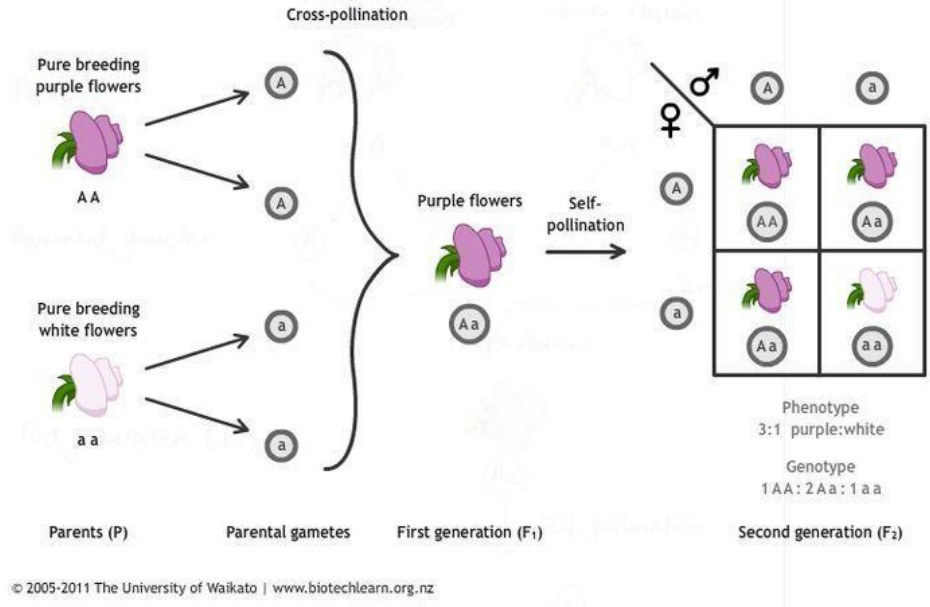
3) The Law of Dominance: An organism with alternate forms of a gene will express the form that is dominant

Mendel's experiments

	Flower Colour	Plant Height	Seed Color	Seed Shape	Pod Colour	Pod Shape	Flower Position
Dominant Trait	 Purple	 Tall	 Yellow	 Round	 Green	 Inflated (full)	 Axial
Recessive Trait	 White	 Short	 Green	 Wrinkled	 Yellow	 Constricted (flat)	 Terminal

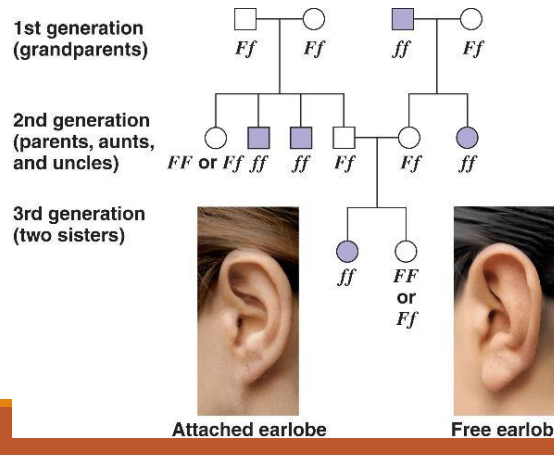
Dominant:
≥1 allele to have trait/disease
Recessive:
2 (alternative) alleles to have trait/disease

A = reference allele
a = other allele



"Mendelian" *monogenic* inheritance

Dominant trait
- You only need 1 allele to have trait/disease



□ = males
O = females

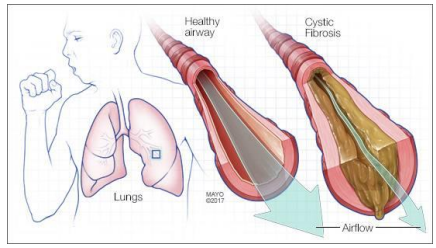
F = reference allele
f = other allele

Purple = "affected" (attached earlobe)

"Mendelian" or "monogenic" traits

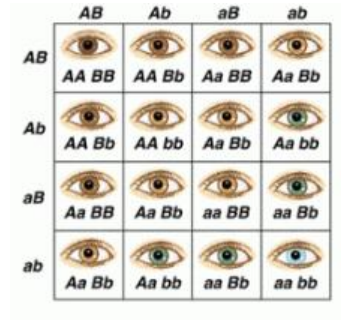
Traits that are determined by <u>one/a few</u> genes

SEVERE DISEASES



Huntington, *BRCA*-linked breast cancer, cystic fibrosis, Rett's disease, sickle cell disease

SIMPLE PHYSICAL TRAITS (≈MENDELIAN)



Ear lobe attachment, eye color, cheek dimples, blood type

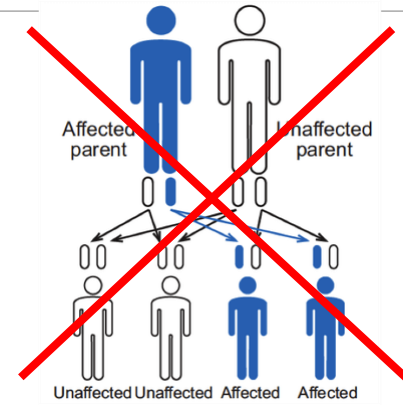
Polygenic traits

Most traits do ***<u>not</u>*** follow <u>simple</u> Mendelian inheritance

<u>*Many*</u> genetic variants affect trait

- Each with *tiny* effect size

"Polygenic" traits



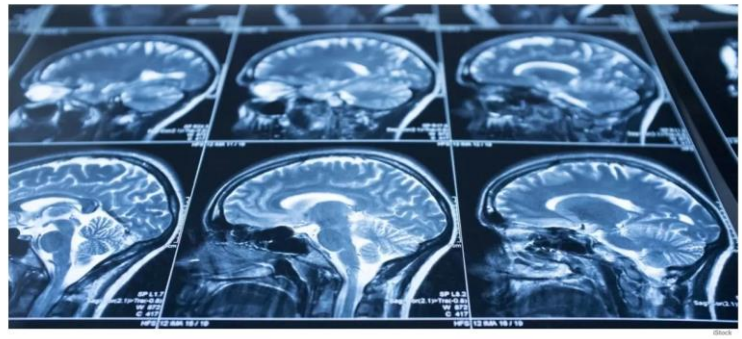
Complex traits (happiness, risk preferences, intelligence, etc.) do not follow simple inheritance patterns!

This gene makes you smarter

FROM

mental_floss

Jessica Hullinger



May 20, 2013

Scientists have long known that our genetic makeup influences our intelligence. But now, thanks to the largest brain study of its kind, [illegible]

NATIONAL

The Case Against the 'Intelligence Gene'

More evidence that DNA and intelligence are related--but let's not get carried away

REBECCA GREENFIELD AUG 9, 2011

This article is from the archive of our partner **The Wire**.

Smart parents breed smart kids. That seems like common sense. And today another study came out reinforcing that claim. By looking at DNA, researchers have linked genes to IQ, according to a study in Molecular Psychology. Scientists found that at least a thousand genes contribute to between 40 and 51 percent of one's intelligence. Now, that doesn't necessarily mean you can predict someone's smarts just from their parents' grades. It just means that

MORE

The I
of 'G
Taxa
ANNIK

The N
Cove
Storn
ALIA W

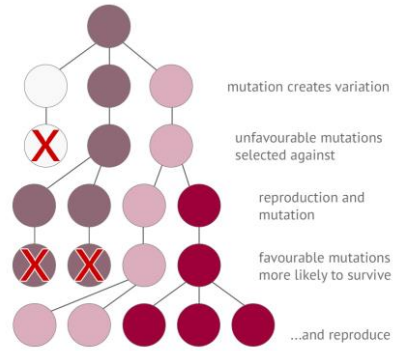
Polygenic traits

Why are traits polygenic?

Mutations with large effects are usually harmful

"Mutation/selection balance"

- Organisms with rare mutations don't tend to (live long enough to) procreate

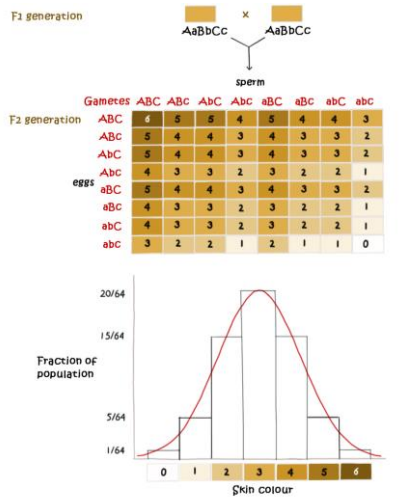


Mendel's laws and polygenic traits

Practice with polygenic inheritance: Handout Lecture 2

Simple setup with 3 genes: Distribution of "trait-increasing" alleles in the offspring generation

Polygenic inheritance



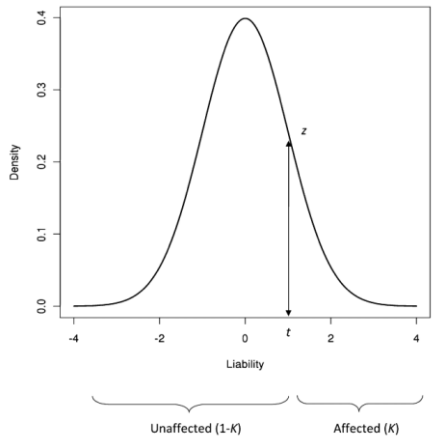
Simple example: 3 genes

Trait following polygenic inheritance will be *normally distributed* in a reasonably large sample

- *"Continuous" traits*

Theoretical example with three SNPs– imagine this figure with >hundreds SNPs...!

Liability threshold model



For binary outcomes (e.g., entrepreneur vs. non-entrepreneur), we assume there is an <u>underlying continuous liability</u>

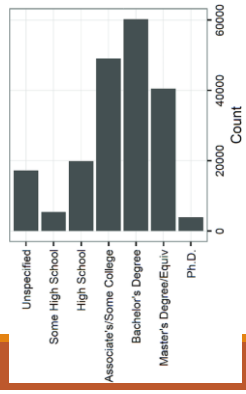
If your liability is sufficiently high, you are "affected"

(See also Lecture 3)

Years of education vs. College degree

Let's assume for now that there exists some SNPs that are related to socio-economic traits (*Lecture 4*), e.g., educational attainment

"Continuous" (quantitative) and heritable trait, so we may expect many underlying genetic variants



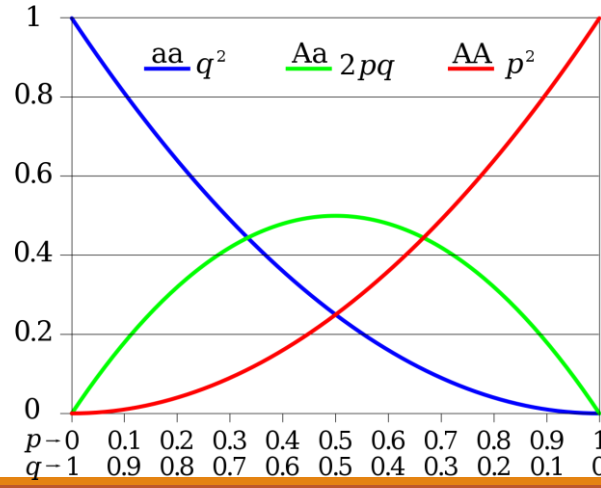
Inheritance: Population genetics

	Alleles (genes)		Genotypes		
	A	a	AA	Aa (=aA)	aa
Frequencies	*p*	*q = (1-p)*	p^2	*pq+pq=2pq*	q^2

A population with constant gene and genotype frequencies is said to be in Hardy-Weinberg equilibrium

If the gene frequencies of two alleles among the parents are *p* and *q*, then the genotype frequencies among the progeny are p^2, $2pq$, and q^2

Hardy-Weinberg equilibrium



Exercise

Exercise follows Falconer & Mackay, 1996 (Example 1.3)

Phenylketonuria (PKU) is a human metabolic disease due to a single recessive gene. Homogyzgotes (carrying "AA") can be detected a few days after birth. Tests of babies in a particular region, detected 1 cases in 10,000 babies. Hence, the frequency of homozygotes in the sample is 1/10,000.

- Assuming Hardy-Weinberg equilibrium, what is the frequency of the recessive allele in the population?
- Are you surprised that for such a rare disease, there are so many individuals carrying this recessive allele?

Exercise (solution)

Exercise follows Falconer & Mackay, 1996 (Example 1.3)

Phenylketonuria (PKU) is a human metabolic disease due to a single recessive gene. Homogyzgotes (carrying "AA") can be detected a few days after birth. Tests of babies in a particular region, detected 1 cases in 10,000 babies. Hence, the frequency of homozygotes in the sample is 1/10,000.

- Assuming Hardy-Weinberg equilibrium, what is the frequency of the recessive allele in the population?
  - Frequency of homozygotes is $q^2 \rightarrow q = \sqrt{(1/10{,}000)} = 0.01$ (1%)
- Are you surprised that for such a rare disease, there are so many individuals carrying this recessive allele?
  - Frequency heterozygotes in the population is $2pq = 2(1-0.01)0.01 = 0.0198$ (≈2%, i.e., 1/50)

Causes of gene frequency changes

Several mechanisms affect gene and genotype frequencies in the process of transmission of genes from one generation to the next

- *Population size: Gene frequencies are subject to sampling variation between successive generation, and the smaller the number of parents the greater is the sampling variation*
- *Migration and mutation: Entrance or leaving of genotypes*
- *Assortative mating: Partners select each other based on heritable characteristics*
- Selection: Fitness differences (relation between genotype and fertility, e.g., through survival)
  - The contribution of offspring to the next generation is called the fitness of the individual
    - If the differences of fitness are in any way associated with the presence or absence of a particular genetic variant, then selection operates on that gene
    - The strength of selection is expressed as the coefficient of selection *s*

Selection and fitness

Selection against a (assuming absence of dominance):

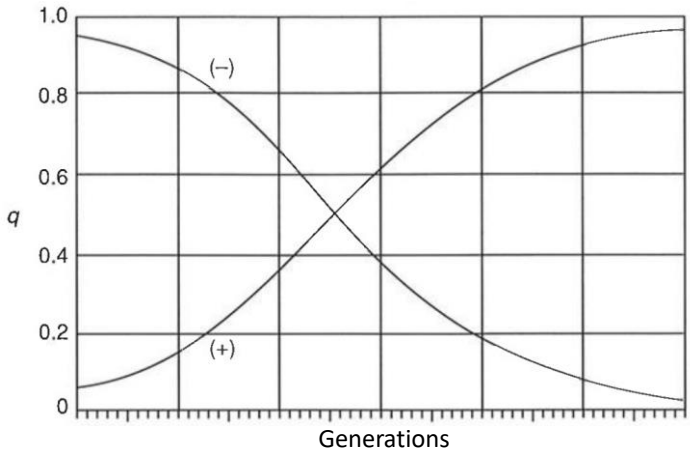
- AA with frequency p^2 is passed on with fitness 1
- Aa with frequency 2pq is passed on with fitness 1 – 0.5s
- aa with frequency q^2 is passed on with fitness 1-s

Let q_1 be the gene frequency in the offspring generation

- If s = 0, then $q_1 = q$
- If s > 0, then $\Delta q = q_1 - q = -(0.5sq(1-q))/(1-sq) < 0$

Let q_0 = 94% and s = 0.2 (20%!), how many generations will it take for allele a to disappear in the population?

Selection and fitness (q_0 = 94%, s = 0.2)



Change of gene frequency during the course of selection from one extreme to the other.

Graph marked (-) refers to selection against the allele whose frequency is q, so that q decreases.

The graph marked (+) refers to selection is favor of the alternative allele

(Figure from Falconer & Mackay, 1996, p. 31)

Selection on socio-economic outcomes

"Genetic evidence for natural selection in humans in the contemporary United States" (Beauchamp, 2011)

Genetic "scores" (combinations of genetic variants, *more about this in Lectures 4 & 5*)

- BMI = Body Mass Index
- EA = Educational attainment
- GLU = Glucose concentration
- HGT = Height
- SCZ = Schizophrenia
- TC = Total cholesterol
- AAM = Age at Menarche

rLRS = (relative) "Lifetime Reproductive Success"

Table 2. Estimates from separate regressions of rLRS on the polygenic score of each phenotype

Score	Females	Males
Score of BMI	0.006 (0.010)	0.016 (0.013)
Score of EA	−0.033*** (0.010)	−0.031** (0.012)
Score of GLU	0.009 (0.010)	−0.013 (0.013)
Score of HGT	−0.011 (0.014)	−0.005 (0.018)
Score of SCZ	−0.001 (0.011)	0.009 (0.013)
Score of TC	−0.012 (0.011)	−0.003 (0.013)
Score of AAM	0.018* (0.011)	—
N	3,416	2,571

This table shows estimates of the coefficients on the polygenic scores and their SEs (in parentheses) from separate regressions of rLRS on the polygenic score of each phenotype and on control variables, for the study sample. All regressions for each sex had the same number of observations. (*$P < 0.10$; **$P < 0.05$; ***$P < 0.01$.)

Natural selection vs. *Artifical* selection

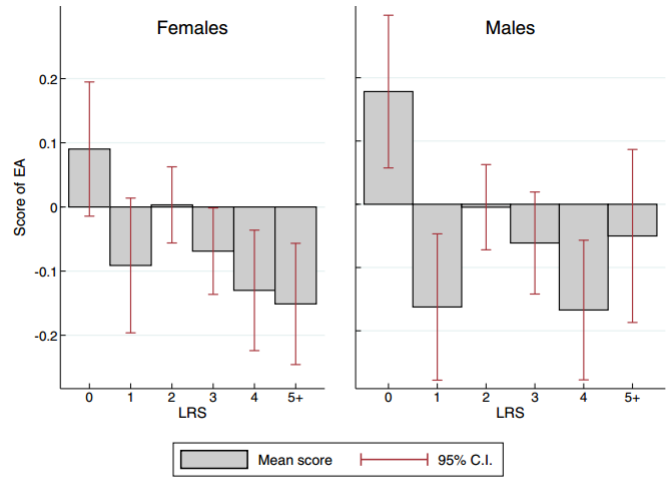


Fig. 2. Mean polygenic score of EA as a function of LRS, for females and males in the study sample.

Natural selection and the environment

"… my estimate of a directional selection differential of EA of about −1.5mo of education per generation pales in comparison with the increase of 6.2 y in the mean level of EA that took place for native-born Americans born between 1876 and 1951 (which is equivalent to about 2 y of education per generation)."

"Thus, although natural selection is still operating, the environment appears to have achieved an "evolutionary override" on the measurable phenotypes I study."

Be careful with using "natural selection" for behavioral traits

Genoeconomics (behavioral genetics) needs to take the environment seriously!

The 4 promises of *genoeconomics* (2012)

The four promises of *genoeconomics* can only be realized when we appreciate that genes operate in particular environments

1. *Biological mechanisms*
2. *Direct measure of construct that is otherwise difficult to measure*
3. *Targeted interventions*
4. *Genes as control variables and/or instrumental variables*

Genes, Environments, Genes × Environment (G × E)

- Lecture 5 ⟶ The policy relevance of genoeconomics

Summary of Lecture 2

What are the promises of genoeconomics?

1. *Biological mechanisms*
2. *Direct measure of construct that is otherwise difficult to measure*
3. *Targeted interventions*
4. *Genes as control variables and/or instrumental variables*

The basics of genetics: What is a "SNP"?

- A SNP is a one nucleotide difference in the genome: A/C/T/G

How can we use Mendel's laws to understand natural selection on socio-economic traits?

- The Law of Segregation, the Law of Independent Assortment, and the Law of Dominance
- Socio-economic traits are polygenic
- Mendel's laws not sufficient: Artificial selection and the environment

What comes next?

- Tutorial 2: Practice with SNP data in PLINK
  - Go through the step-by-step tutorial (Canvas → Pages → Tutorial 2) yourself / together
  - Continue with the Individual assignment (e.g., the Introduction)
- Make teams for Group assignment and register them on Canvas (see Group assignment.pdf on Canvas)
  - The first presentation will *next week*
- Lecture 3:
  - Are unrelated but genetically similar individuals more similar in their behavior?
  - What does it mean that some behaviors are genetically related?
  - Are genetically similar individual attracted to each other? (*Presentation*)

Economics & Genetics Lecture 2

DR. NIELS RIETVELD, NRIETVELD@ESE.EUR.NL

OFFICE MANDEVILLE T18-29