Estimating risk profiles for common diseases from environmental and genetic factors

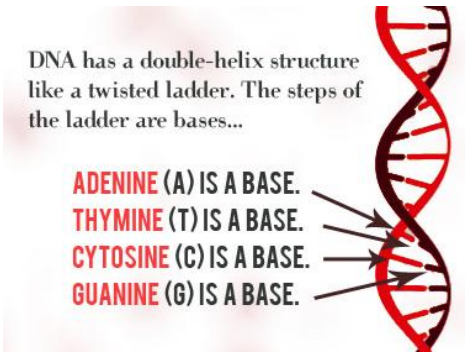
Cathryn Lewis

King's College London

Contents

- Introduction to genetic prediction
- Estimating disease risks
- Implications

Introduction to genetics: 1
DNA structure



www.onlineeducation.net/dna

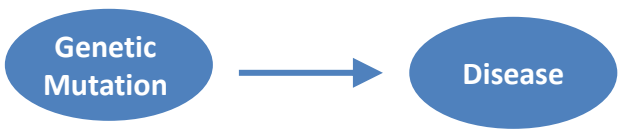

James Watson and Francis Crick with their DNA model at the Cavendish Laboratories in 1953. Photograph copyright A. Barrington

Introduction to genetics: 2
DNA differences

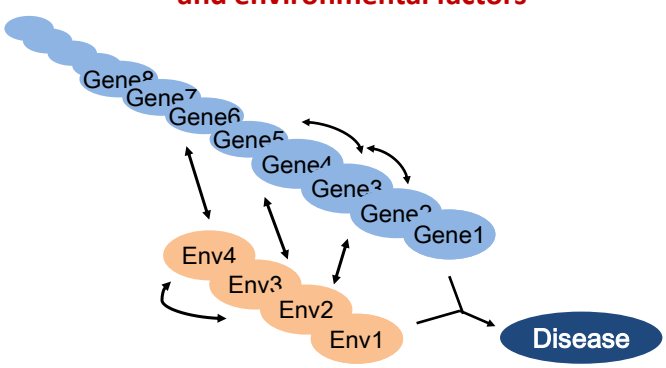
Inherited genetic mutations

Single gene disorders

- Huntington's disease
- Cystic fibrosis
- Breast cancer genes: *BRCA1, BRCA2*



Complex disease: contributions from genetic and environmental factors



Examples: asthma, breast cancer, heart disease, autism, arthritis, migraine, obesity, diabetes, stroke

Most diseases that have a major economic, social and health burden

Genetic variation: Single nucleotide polymorphism (SNP)

....TGGACATGCA....
....TGGACCTGCA....

Alleles A and **C** are present in the population

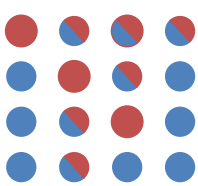
Genotype : carried by an individual, on paternal and maternal inherited chromosomes

....TGGACATGCA....	....TGGACATGCA....	....TGGACCTGCA....
....TGGACATGCA....	....TGGACCTGCA....	....TGGACCTGCA....
Genotype: AA	**AC**	**CC**

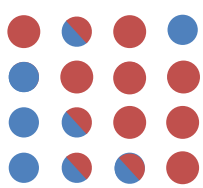
Identifying SNPs that increase risk of disease

Genotype SNP with A, C alleles: AA AC CC

Cases – affected with disease

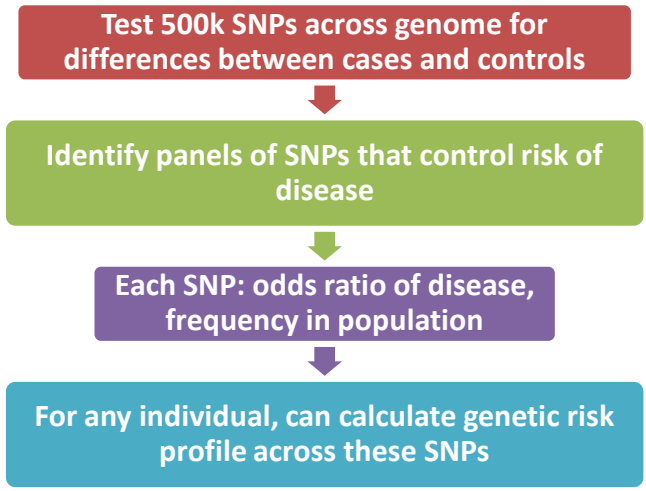


Controls – not affected with disease



More AC and CC genotypes in cases than in controls
Indicates that carrying C allele increases risk of disease

Genetic association studies



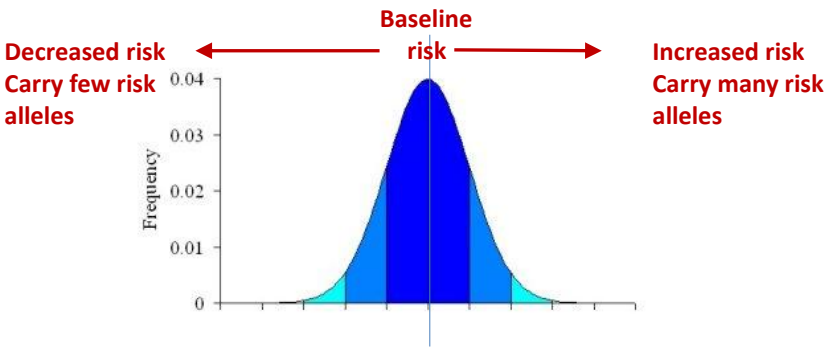
Breast cancer genetics

			Odds Ratio, by number of risk alleles		
Name of SNP	Gene location	Risk allele	0	1	2
1 rs2981579	*FGFR2*	A	1	1.35	1.82

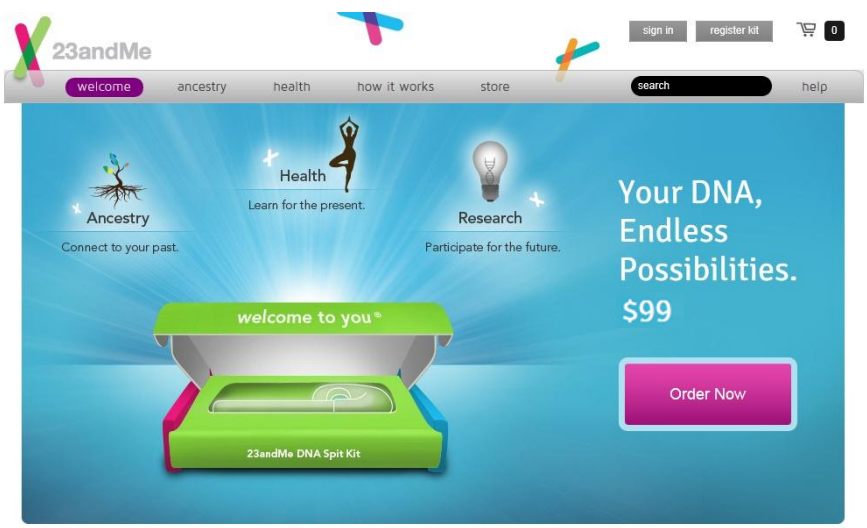
To combine relative risk across SNPs: multiply odds ratio for genotype

Product of odds ratios = 1 x 1.28 x 1.42 x 1.31 x 1 = 2.38
Rescale so OR is compared to an 'average' member of the population

Distribution of genetic risk in the population



How useful is this information?



Research programme: Disease risk estimation for combining genetic and environmental risk factors

- Developed new statistical methodology
  - Combining genetic and environmental risk factors
  - Incorporating confidence intervals
- Issued software program REGENT
- Evaluated utility of risk prediction for common diseases

Risk modelling: genetic factors

- Risk SNP characterised by
  - Minor allele frequency (MAF), *p*
  - Odds ratio for each minor allele (*1, g, g²*)
- Disease prevalence *r*
- Assume risks are multiplicative across SNPs

Frequency
Severity

- *N* SNPs, with genotype k_i = 0, 1, 2, *i=1*, ... *N*

$$P(D \mid k_1, k_2, \ldots, k_N) = r \prod_{k=1}^{N} g_i^{k_i} \Big/ (1 + (g_i - 1) p_i)^2$$

Risk modelling: environmental factors

- Environmental risk factors (*M*), each with
  - OR h_j
  - Confidence interval
  - Exposure prevalence, *e[j]* = 0,1
- Risk component relative to individual with no exposure is: $\prod_{j=1}^{M} h_j^{e[j]}$

Model assume environmental risks are independent

Risk modelling: confidence intervals

Disease risk estimated using multiplicative model between

Genetic risk factors

Environmental risk factors

Calculate empiric confidence intervals for an individual genotype

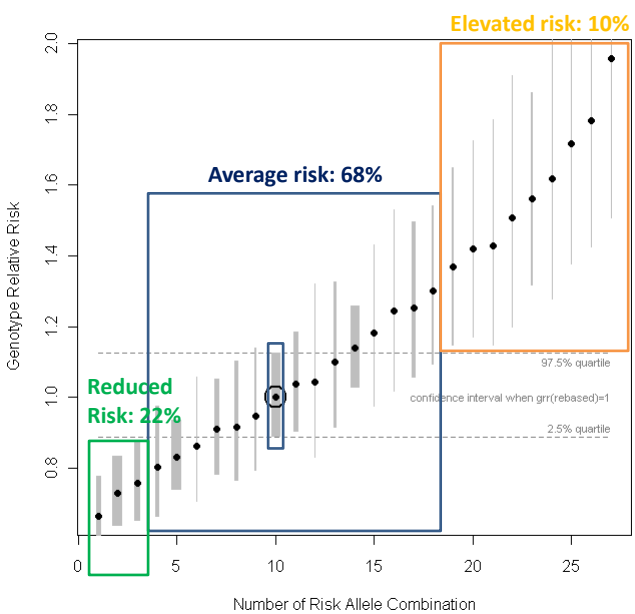
Type 2 diabetes risk SNPs

SNP	Allele frequency	OR
rs5215	0.35	1.14
rs7901695	0.31	1.37
rs4430796	0.47	1.10

Frayling et al., 2007

Three SNPs : Type 2 Diabetes

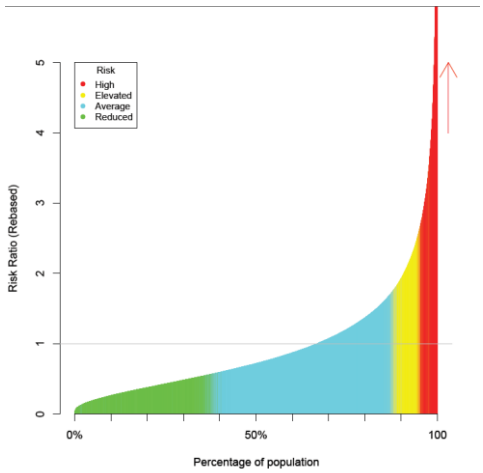
Combination Number	SNP Number 1	2	3	Population Frequency	Rel. risk (Rebased)	Rel. Risk Quartiles 2.50%	97.50%	Risk Category
1	0	0	0	0.0565	0.6636	0.5649	0.7802	Low
2	0	0	1	0.1002	0.7299	0.6395	0.8361	Low
3	1	0	0	0.0609	0.7565	0.6482	0.8807	Low
4	0	0	2	0.0444	0.8029	0.6607	0.9810	Average
5	1	0	1	0.1079	0.8321	0.7385	0.9370	Average
6	2	0	0	0.0164	0.8624	0.7010	1.0566	Average
7	0	1	0	0.0508	0.9091	0.7809	1.0565	Average
8	1	0	2	0.0479	0.9153	0.7579	1.1064	Average
9	2	0	1	0.0291	0.9486	0.7886	1.1414	Average
10	0	1	1	0.0900	1.0000	0.8876	1.1280	Average
11	1	1	0	0.0547	1.0364	0.9026	1.1887	Average
12	2	0	2	0.0129	1.0435	0.8297	1.3171	Average
13	0	1	2	0.0399	1.1000	0.9112	1.3256	Average
14	1	1	1	0.0970	1.1400	1.0288	1.2621	Average
15	2	1	0	0.0147	1.1815	0.9738	1.4377	Average
16	0	2	0	0.0114	1.2455	1.0094	1.5375	Average
17	1	1	2	0.0430	1.2540	1.0509	1.5008	Average
18	2	1	1	0.0261	1.2996	1.0950	1.5469	Average
19	0	2	1	0.0202	1.3700	1.1394	1.6540	Moderate
20	1	2	0	0.0123	1.4198	1.1634	1.7311	Moderate
21	2	1	2	0.0116	1.4296	1.1452	1.7899	Moderate
22	0	2	2	0.0090	1.5070	1.1937	1.9128	Moderate
23	1	2	1	0.0218	1.5618	1.3149	1.8695	Moderate
24	2	2	0	0.0033	1.6186	1.2709	2.0696	Moderate
25	1	2	2	0.0097	1.7180	1.3723	2.1595	Moderate
26	2	2	1	0.0059	1.7805	1.4225	2.2348	Moderate
27	2	2	2	0.0026	1.9585	1.4956	2.5690	Moderate



No high risk genotypes

Different Risk Categories

Crohn's disease risk estimation



REGENT software

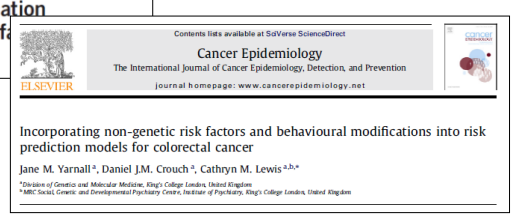
- R package
  - http://cran.r-project.org/web/packages/REGENT/
- Population distribution of disease risk and risk categories
- Individual-level risk assessment
- Genetic risk factors (SNP genotypes) and environmental risk factors (multilevel)


European Journal of Human Genetics (2013) 21, 109–111
© 2013 Macmillan Publishers Limited All rights reserved 1018-4813/13
www.nature.com/ejhg

SHORT REPORT

REGENT: a risk assessment and classification algorithm for genetic and environmental fa

Daniel JM Crouch[1], Graham HM Goddard[1] and Cathryn M Lewis*,[1,2]


Contents lists available at SciVerse ScienceDirect

Cancer Epidemiology

The International Journal of Cancer Epidemiology, Detection, and Prevention

journal homepage: www.cancerepidemiology.net

ELSEVIER

Incorporating non-genetic risk factors and behavioural modifications into risk prediction models for colorectal cancer

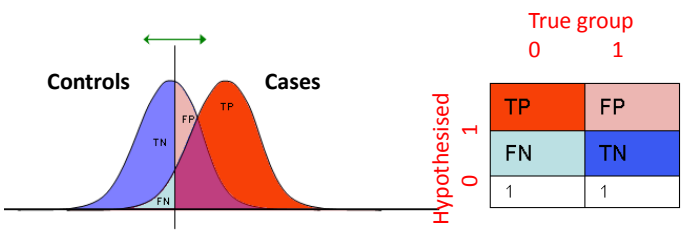
Jane M. Yarnall[a], Daniel J.M. Crouch[a], Cathryn M. Lewis[a,b,*]

[a] Division of Genetics and Molecular Medicine, King's College London, United Kingdom
[b] MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, United Kingdom

Genetic risk profile

- Case studies of three adult-onset disorders :
  - Coronary artery disease
  - Colorectal cancer
  - Type 2 diabetes
- Identified SNPs most strongly associated with disease
- Modelled genetic profiles in the population through simulation
- Assessed ability of model to identify individuals at high risk of disease

Receiver operating characteristic curve



Genetic risk assessment

Disease	No. SNPs modelled	Area under curve	Proportion of population at increased risk		Lifetime risks
			OR > 2	OR > 3	
Coronary artery disease	25	0.60	1.5%	0.0%	6.0%
Colorectal cancer	10	0.59	0.7%	0.0%	6.2%
Type 2 diabetes	19	0.60	1.7%	0.0%	4.0%

Odds ratios: genetic v. conventional risk factors

Disease	Top 5% of genetic risk	Family history (affected sibling)	Epidemiological & risk factors	
Coronary artery disease	1.7	3.2	Total cholesterol Smoking	3.1 1.9
Colorectal cancer	1.6	5.1	Smoking Obesity	1.3 1.5
Type 2 diabetes	1.7	3.5	Obesity	2.5

Summary

- *Scientific* strides in identifying the inherited genetic variants that affect disease risk
- Very limited prediction available from current findings
  - Incomplete knowledge of polygenic component of disease
  - Causal genetic variants are unknown
- Better prediction comes from
  - Family history
  - Environmental risk factors (smoking, body mass index)
  - Pre-clinical factors (blood pressure, cholesterol levels)

Acknowledgements

King's College London

- Graham Goddard
- Daniel Crouch
- Jane Yarnall

Funding

Institute and Faculty of Actuaries

MRC Medical Research Council CENTENARY 1913-2013

"Prediction is very difficult, especially about the future"

Niels Bohr

Institute and Faculty of Actuaries