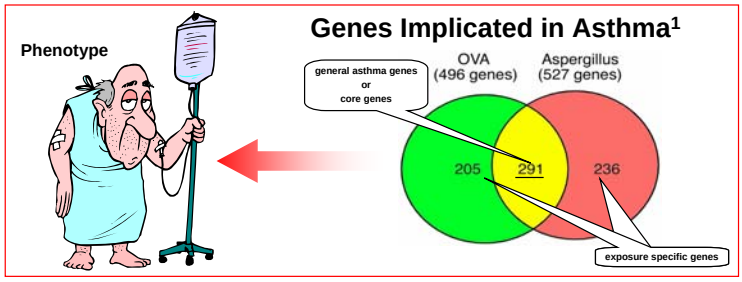
Polygenic Model for Complex Diseases: Genetic Susceptibility and Risk Factor

Eugene Demchuk,† Michael I. Luster,‡ Berran Yucesoy,‡ Victor J. Johnson,‡ Christopher T. De Rosa†
† Agency for Toxic Substances and Disease Registry, Atlanta, GA
‡ National Institute for Occupational Safety and Health, Morgantown, WV
E-mail:

INTRODUCTION

Common diseases, such as asthma, Alzheimer's and cardiovascular diseases, are complex in nature, being variably influenced by physiological, life-style, environmental and genetic factors. As such, variations in individual genes that have the potential to affect a disease generally possess low or incomplete penetrance and, consequently, in epidemiological studies show low risk associations with typical odd ratios around 1.5 - 2.0. The disease phenotype is, in part, a result of joint co-expression of multiple genes. In particular:

- **hundreds of genes** simultaneously shape the susceptible phenotype
- a degree of susceptibility to environmental exposures is determined by the joint effect of unique **combination of genes** specific for a given individual
- how to establish appropriate **risk factors** taking into account subpopulations with genetically susceptible combinations of genes?



METHODS

Population-based genetic association studies deal with relatively small effects against a complex background. Therefore, often they are statistically underpowered and poorly standardized. In the present work[2] source data were extracted from PubMed using the following criteria:

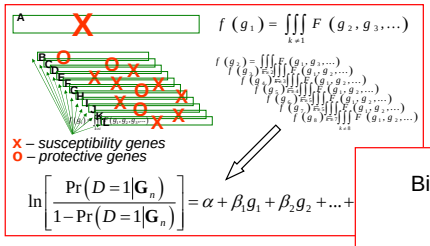
- physician diagnosed asthma as the diagnosis
- case-control study design
- reported associations with $p < 0.05$

For each of 16 genes the following information was collected, e.g.:

Gene	SNP	Reference	Frequency	Odds Ratio
TGF-β	- 509	[3]	0.117	2.456

THEORY

PROBLEM: to estimate gross genetic susceptibility of individual from known single-gene association studies. Mathematically, it narrows down to reconstruction of unknown joint multivariate distribution from known univariate marginals of this distribution.



$$f(g_1) = \iiint_{k \neq 1} F(g_2, g_3, \ldots)$$

$$\ln\left[\frac{\Pr(D=1|\mathbf{G}_n)}{1-\Pr(D=1|\mathbf{G}_n)}\right] = \alpha + \beta_1 g_1 + \beta_2 g_2 + \ldots +$$

- frequencies (p) of genetic markers (minor types) and major types ($q = 1 - p$) enter the model with binomial coefficients

$$f_{genotype} = p_1 p_2 p_3 \ldots \times q_1 q_2 q_3 \ldots$$

- association between a marker and disease is expressed in terms of the standard logistic regression model:

$$\ln\left[\frac{\Pr(Disease=1|g)}{1-\Pr(Disease=1|g)}\right] = \alpha + \beta g$$

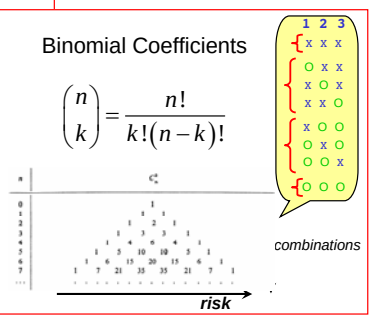
single gene, = 0 or = 1

Odds Ratio (OR)

$$\ln\left[\frac{\Pr(g_1, g_2 \ldots g_n)}{1-\Pr(g_1, g_2 \ldots g_n)}\right] = \alpha + \sum_{i=1}^{n} \beta_i g_i$$

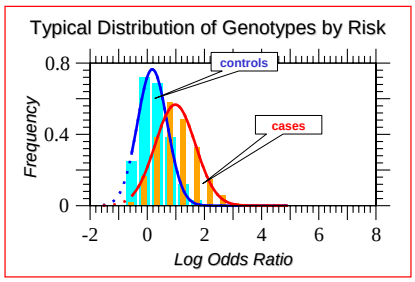
Binomial Coefficients

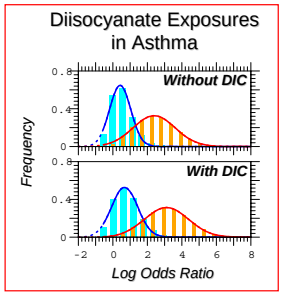
$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$



RESULTS

- Modelling joint impact of the multiple disease variants provides a pseudo-continuous log-normal relative disease risk distribution in the population.

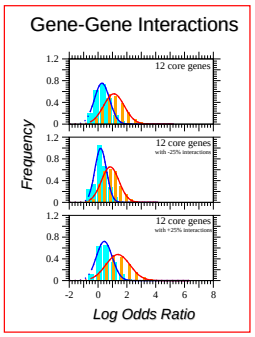


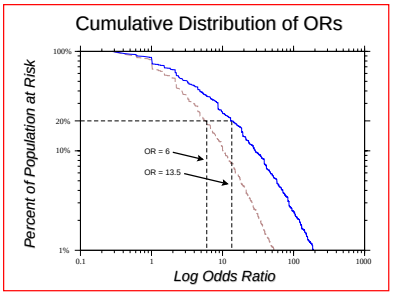


As the number of jointly considered genes increases,

- the distribution of risk further shifts towards the higher risk;
- the standard deviation of disease risk in the population increases;
- the separation between disease population and controls increases.

- **Gene-environment** interactions enter in the model by means of exposure-sensitive genes, *NAT1* in the shown example, which is sensitive to diisocyanates.





- Cumulative distribution quantifies population at risk (" with risk higher than …")
- Mock simulations do not suggest overwhelming effect of **gene-gene** interactions in the model

CONCLUSIONS

A computational approach has been presented, which allows estimation of the **joint contribution of variations** in individual genes to the risk of developing a disease. As an example, variants of 16 asthma susceptibility genes, including those associated with asthma mediators, atopy and chemical metabolism, were analyzed. A 6-fold increased risk of developing asthma for 20% in the general population occurred when only gene variants of asthma mediators were considered. The disease risk was more than doubled (OR=13.5) when the atopy variants were added. Inclusion of all variants resulted in the increased odds ratio of 24. The model helps establish the relative changes in risk associated with **genetic-risk profiles** in the population and provides a framework for comprehensive genetic risk assessment. The software is available on the CDC Intranet, http://158.111.214.63

REFERENCES

[1] N. Zimmermann, N.E. King, J. Laporte, M. Yang, A. Mishra, S.M. Pope, E.E. Muntel, D.P. Witte, A.A. Pegg, P.S. Foster, Q. Hamid & M.E. Rothenberg (2003) Dissection of experimental asthma with DNA microarray analysis identifies arginase in asthma pathogenesis. *J. Clin. Invest.* **111**:1863–1874.
[2] E. Demchuk, B. Yucesoy, V.J. Johnson, A. Weston, D. Germolec, C.T. De Rosa & M.I. Luster (2007) A statistical model to assess genetic susceptibility as a risk factor in multifactorial diseases: Lessons from occupational asthma. *Environ. Health Persp.* **115**:231–234.
[3] E.S. Silverman, L.J. Palmer, V. Subramaniam, A. Hallock, S. Mathew, J. Vallone, D.S. Faffe, T. Shikanai, B.A. Raby, S.T. Weiss & S.A. Shore (2004) Transforming growth factor-beta1 promoter polymorphism C-509T is associated with asthma. *Am. J. Respir. Crit. Care Med.* **169**:214-219.

Disclaimer: The findings and conclusions in this presentation have not been formally disseminated by the Agency for Toxic Substances and Disease Registry or National Institute for Occupational Safety and Health and should not be considered to represent any agency determination or policy.