Неуловимый эпистаз

ИЦиГ
Новосибирск
7 декабря 2016 года

План доклада

- Эпистаз в классической генетике
  - Определение эпистаза и примеры
  - Классические модели эпистаза
- Эпистаз в статистической генетике
  - Наследуемость
  - Полигенная аддитивная модель
  - Полигенная модель со взаимодействиями
  - Кто виноват и что делать?

Эпистаз в классической генетике

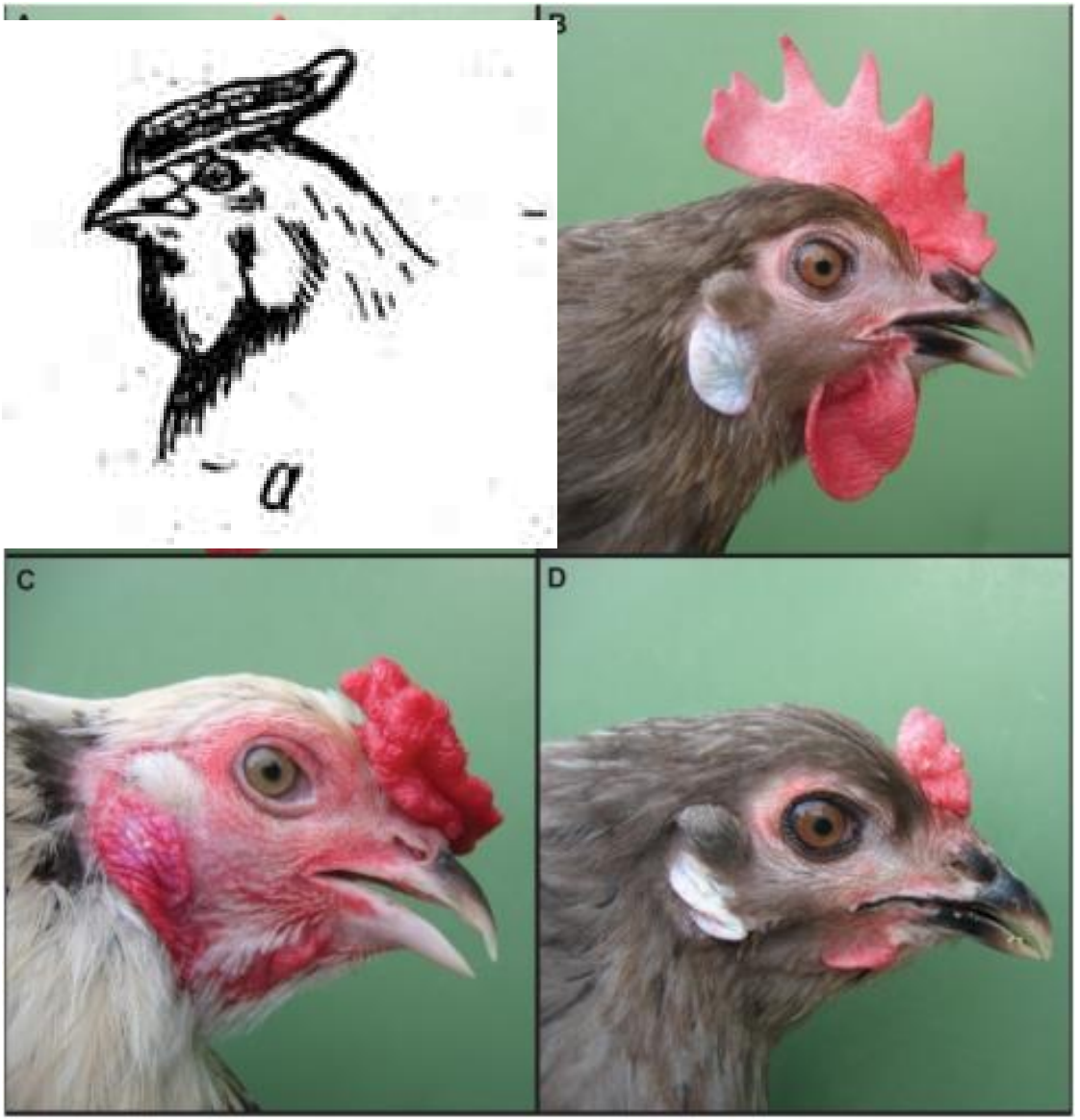
Эпистаз – феномен, при котором эффект гена на признак подавляется вариантом другого гена

Под эпистазом понимается отклонений от независимости эффектов генетических локусов на признак.

Эпистаз в классической генетике

Типы гребня у кур

A – розовидный (A_bb)
B – простой (aabb)
C – гороховидный (aaB_)
D – ореховидный (A_B_)



Эпистаз в классической генетике

Окрас шерсти у лабрадоров

B- E-

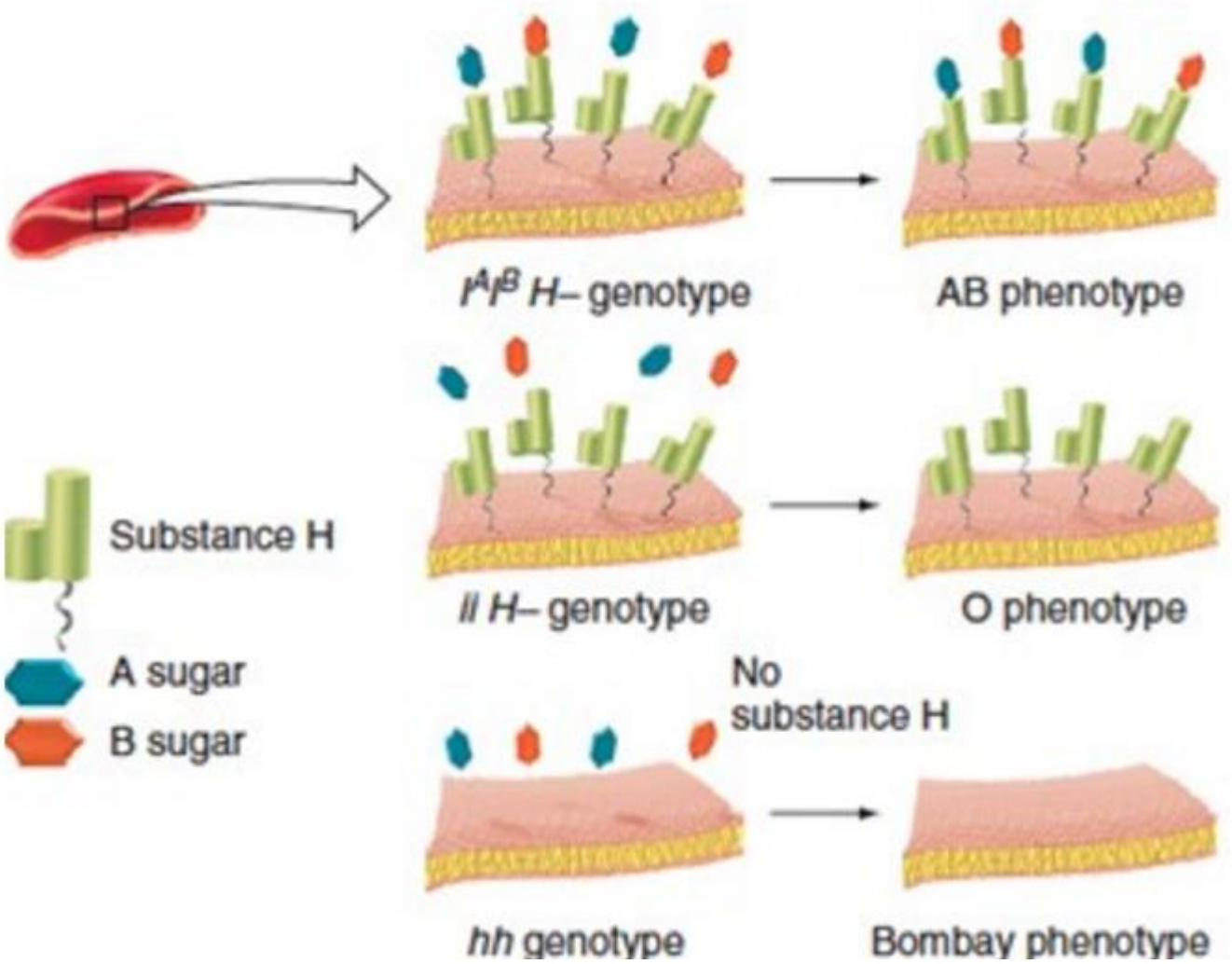
bb E-

B- ee

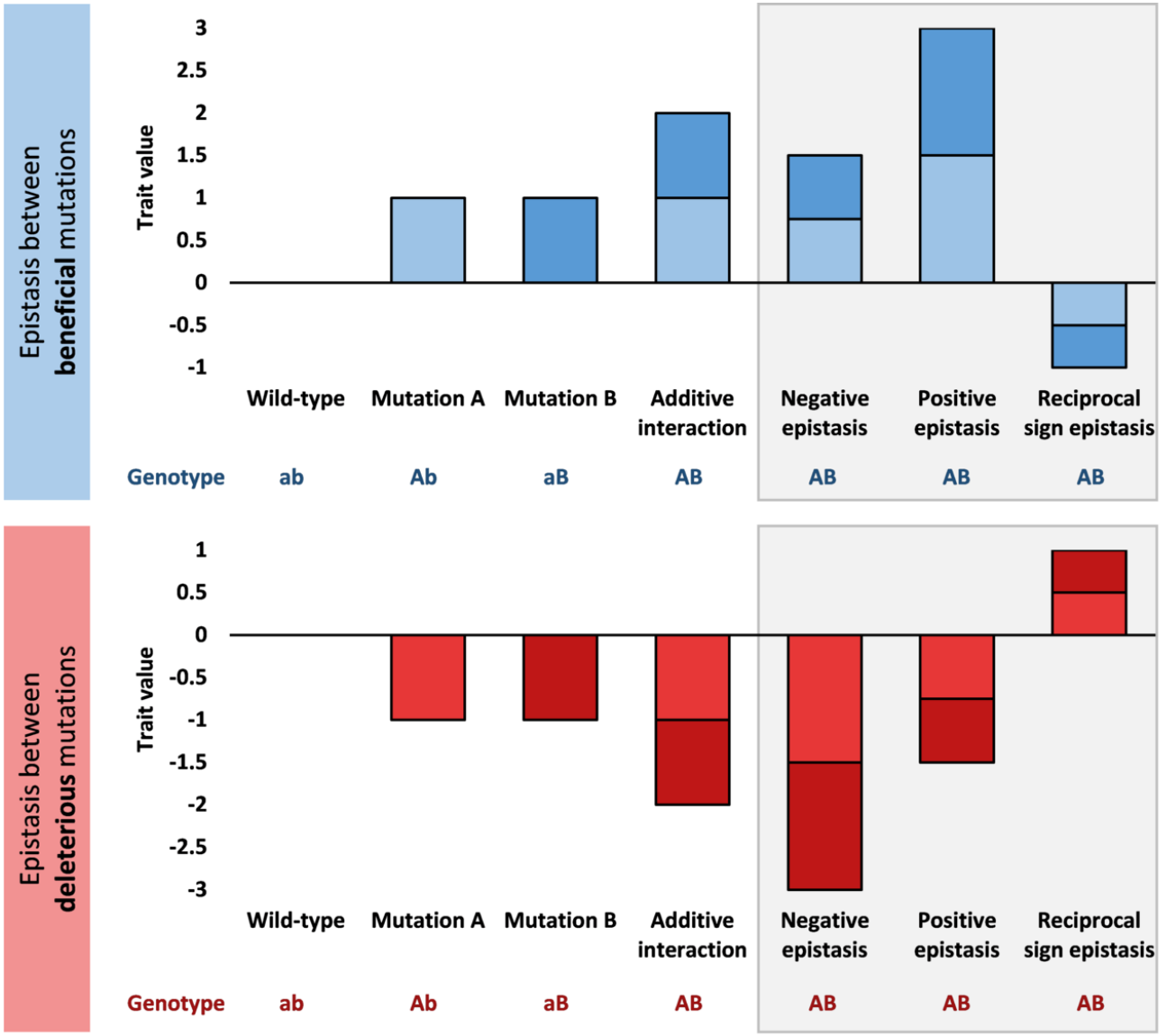
bb ee

Эпистаз в классической генетике

Система АВО группы крови у человека



Эпистаз в классической генетике



Эпистаз в статистической генетике

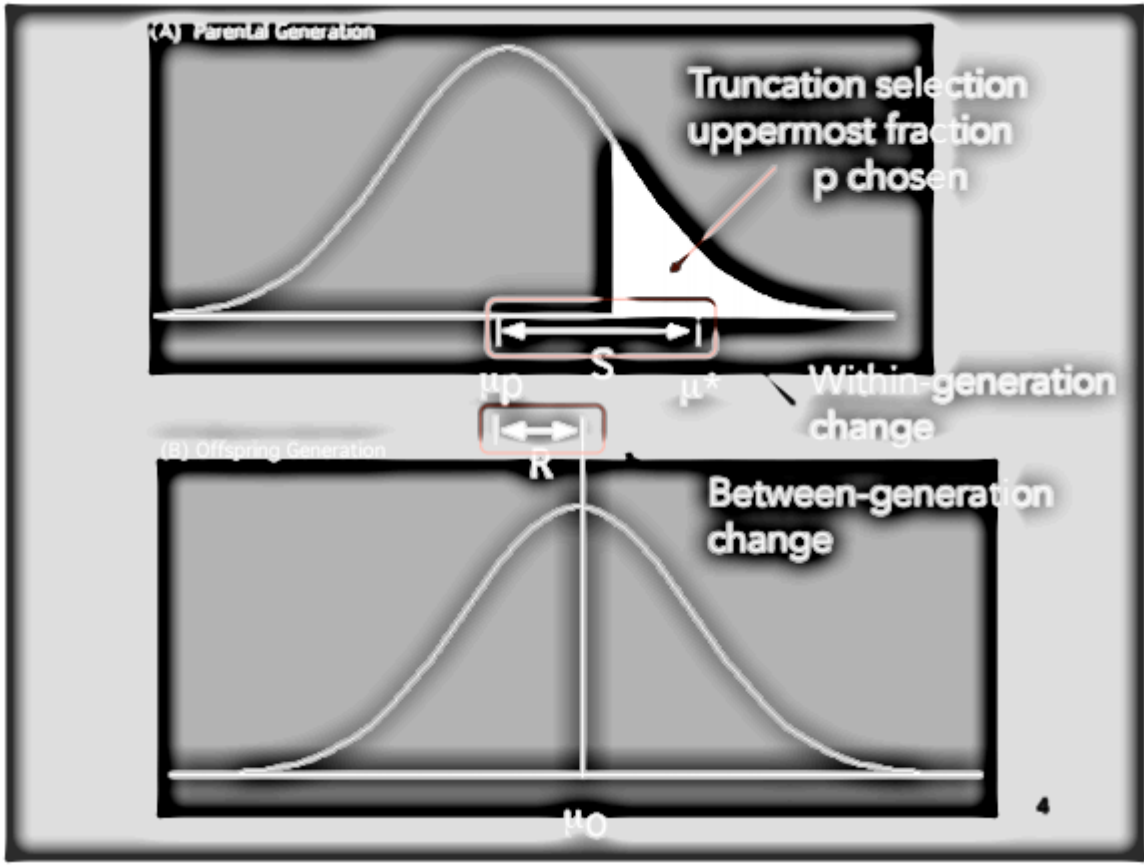
$P = G + E$

$V_P = V_G + V_E$

$V_P = V_A + V_D + V_I + V_E$

$h^2 = V_A / V_P$

$H^2 = (V_A + V_D + V_I) / V_P$



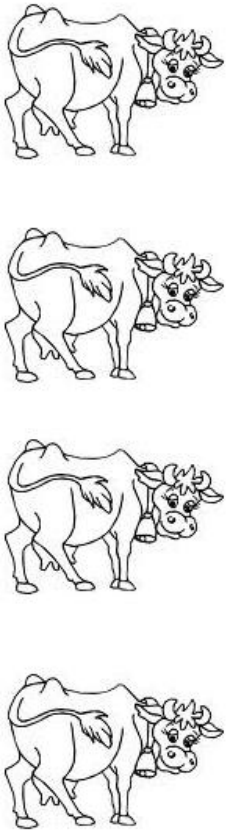
$R = h^2 \times S$

Анализ ассоциаций

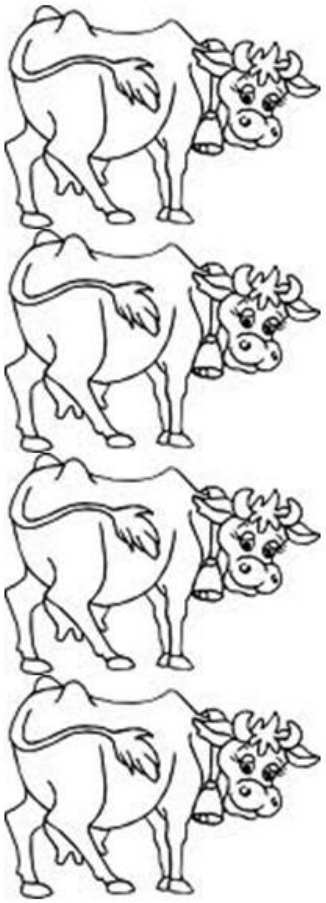
A-G-A-G-C-**G**-T-A-T-C-G

A-G-A-G-C-**A**-T-A-T-C-G

Genotype AA



Genotype AG,GG

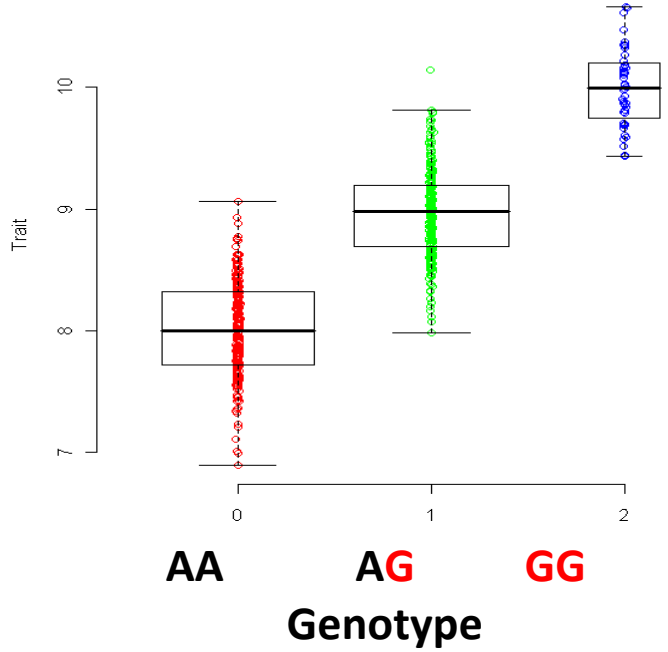


(Lars Rönnegård)

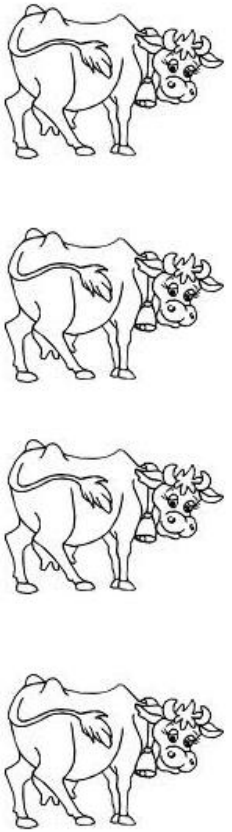
Анализ ассоциаций

A-G-A-G-C-**G**-T-A-T-C-G

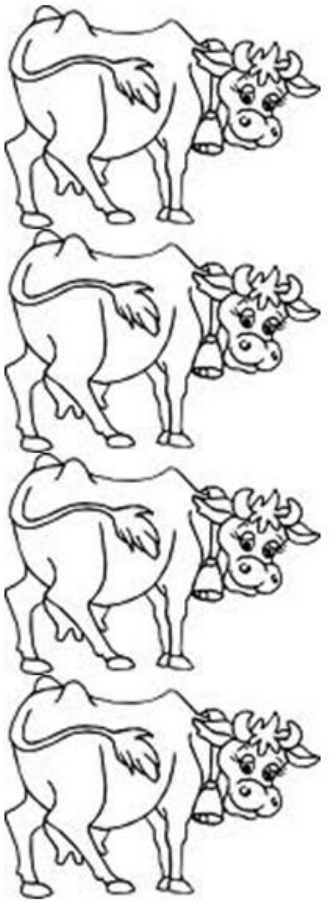
A-G-A-G-C-**A**-T-A-T-C-G



Genotype AA



Genotype AG,GG

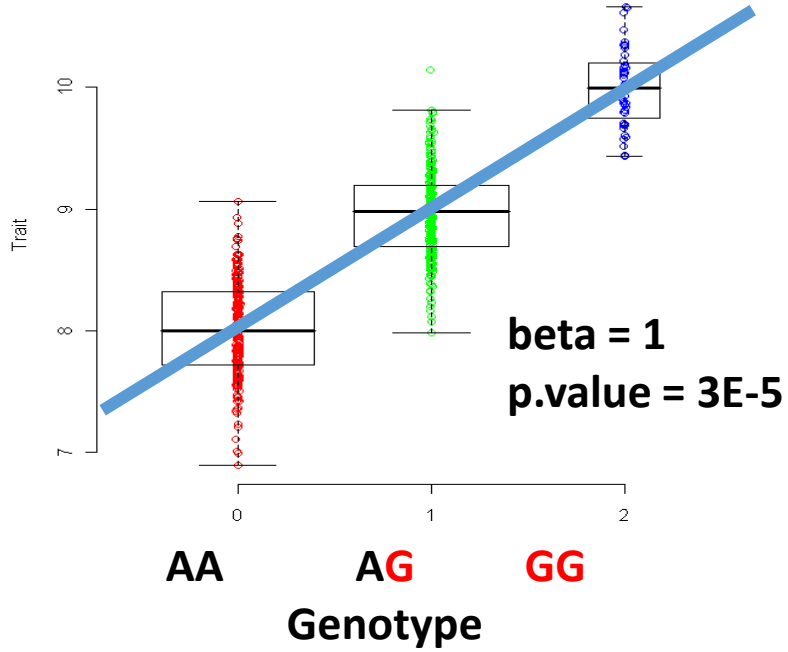


(Lars Rönnegård)

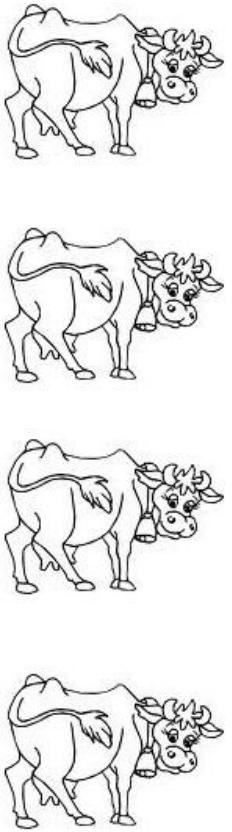
Анализ ассоциаций

A-G-A-G-C-**G**-T-A-T-C-G

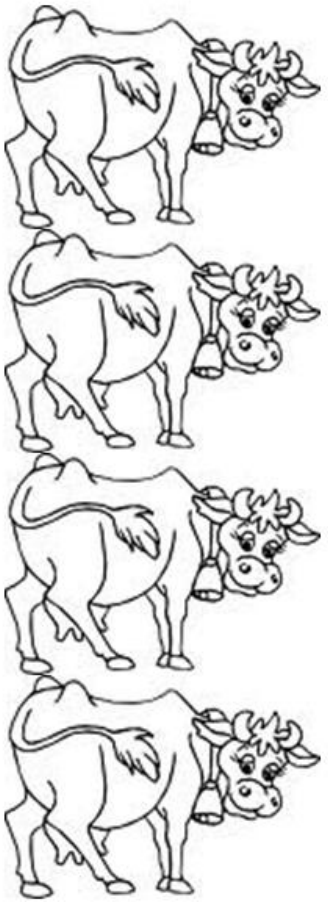
A-G-A-G-C-**A**-T-A-T-C-G



Genotype AA

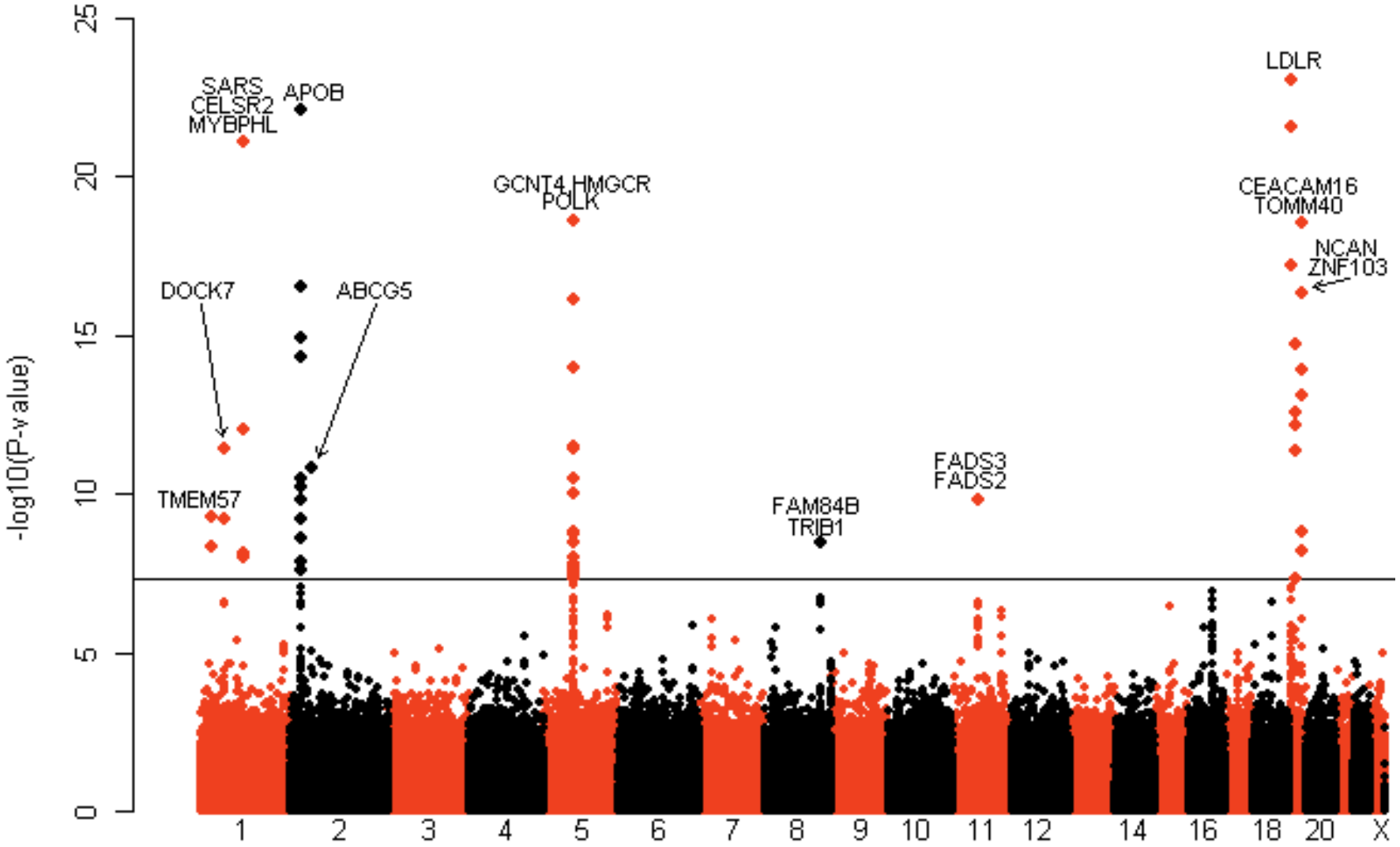


Genotype AG,GG



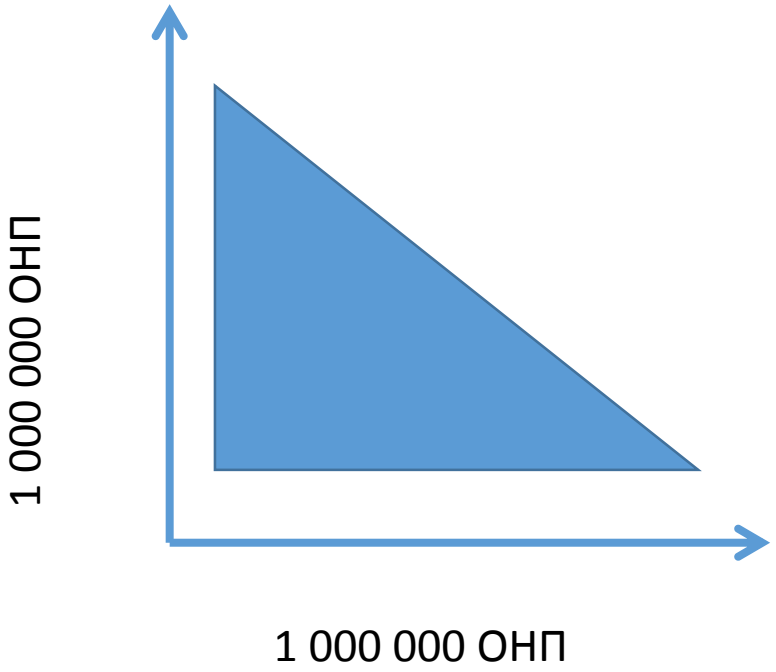
(Lars Rönnegård)

Полногеномный анализ ассоциаций



Поиск взаимодействий генов

$$Y \approx \mu + \beta_g g + \beta_F F + \beta_{gF} * g*F$$



Число тестов для 1 000 000 ОНП		Runtime
GxG	$5 \bullet 10^{11}$	57 дней
GxGxG	$1{,}66 \bullet 10^{17}$	5 300 лет
GxGxGxG	$4{,}16 \bullet 10^{27}$	13 миллиардов лет

Поиск взаимодействий генов

PLOS | ONE

Publish About Browse

OPEN ACCESS PEER-REVIEWED

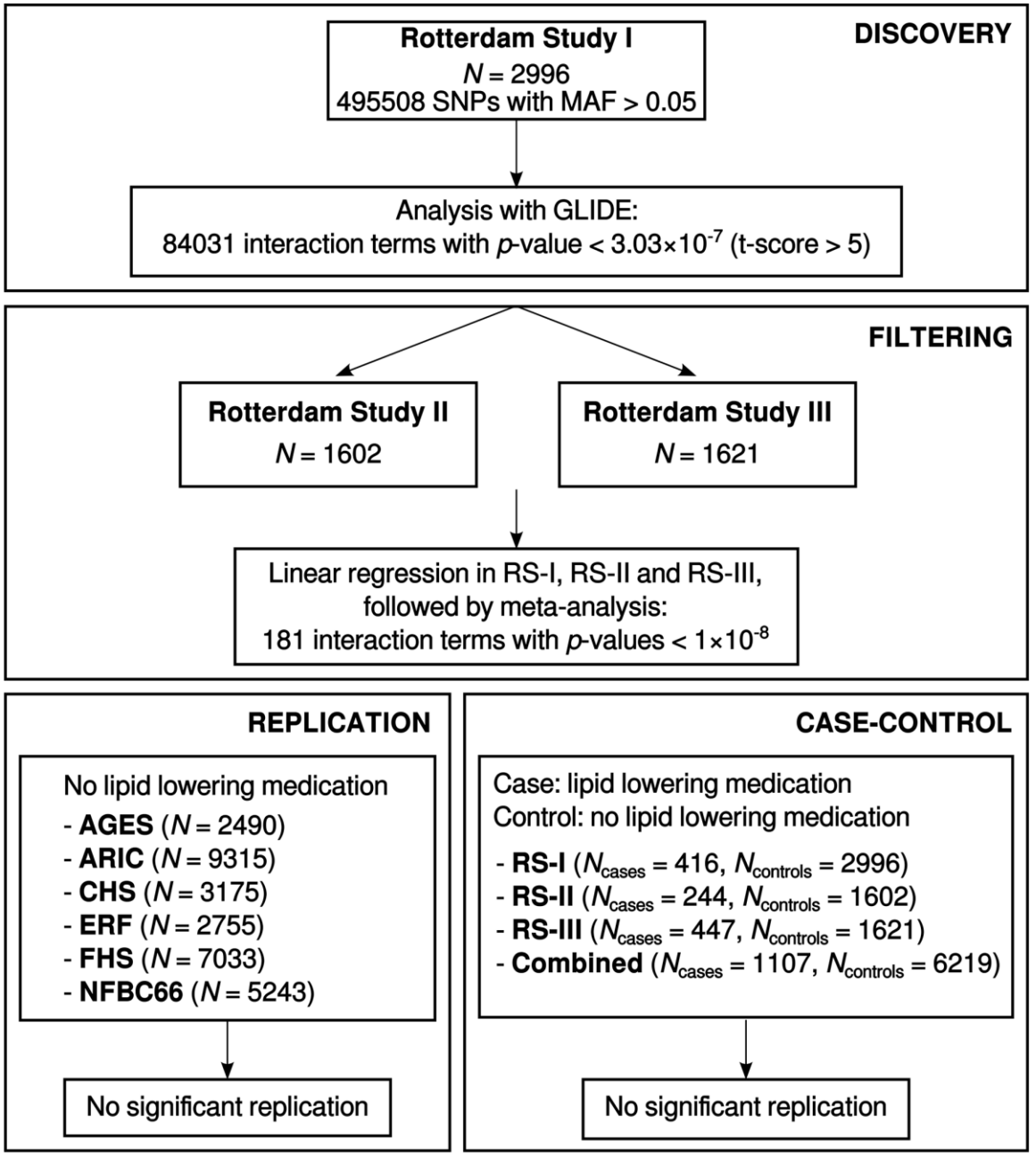
RESEARCH ARTICLE

The Challenges of Genome-Wide Interaction Studies: Lessons to Learn from the Analysis of HDL Blood Levels

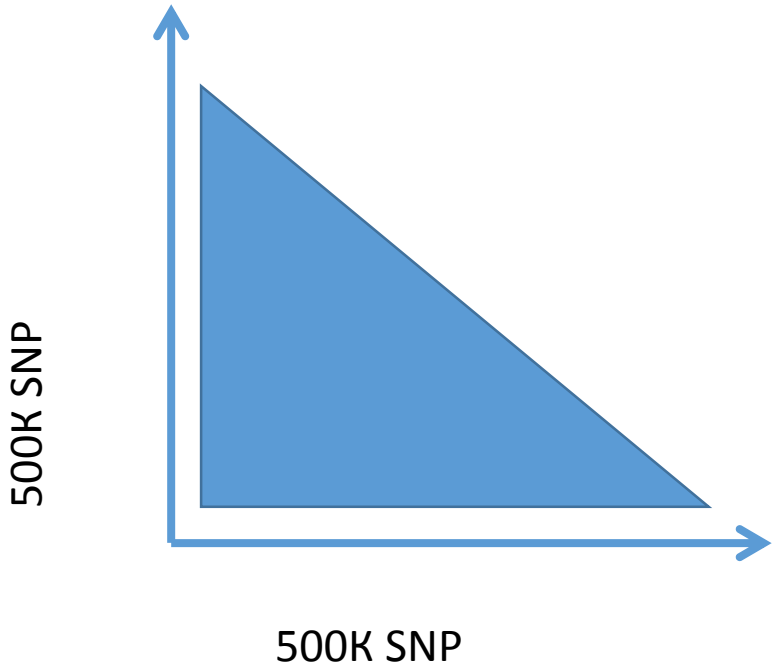
Elisabeth M. van Leeuwen, Françoise A. S. Smouter, Tony Kam-Thong, Nazanin Karbalai, Albert V. Smith, Tamara B. Harris, Lenore J. Launer, Colleen M. Sitlani, Guo Li, Jennifer A. Brody, Joshua C. Bis, Charles C. White, Alok Jaiswal, Ben A. Oostra, Albert Hofman, Fernando Rivadeneira, Andre G. Uitterlinden, Eric Boerwinkle, Christie M. Ballantyne, Vilmundur Gudnason, Bruce M. Psaty, L. Adrienne Cupples, Marjo-Riitta Järvelin, Samuli Ripatti, Aaron Isaacs, Bertram Müller-Myhsok, Lennart C. Karssen, Cornelia M. van Duijn [view less]

Published: October 20, 2014 • DOI: 10.1371/journal.pone.0109290

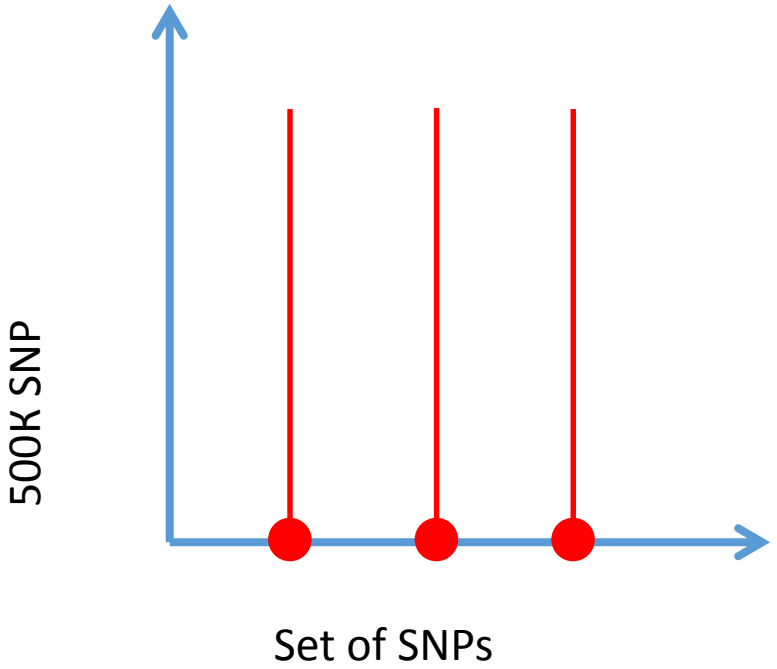
Поиск взаимодействий генов



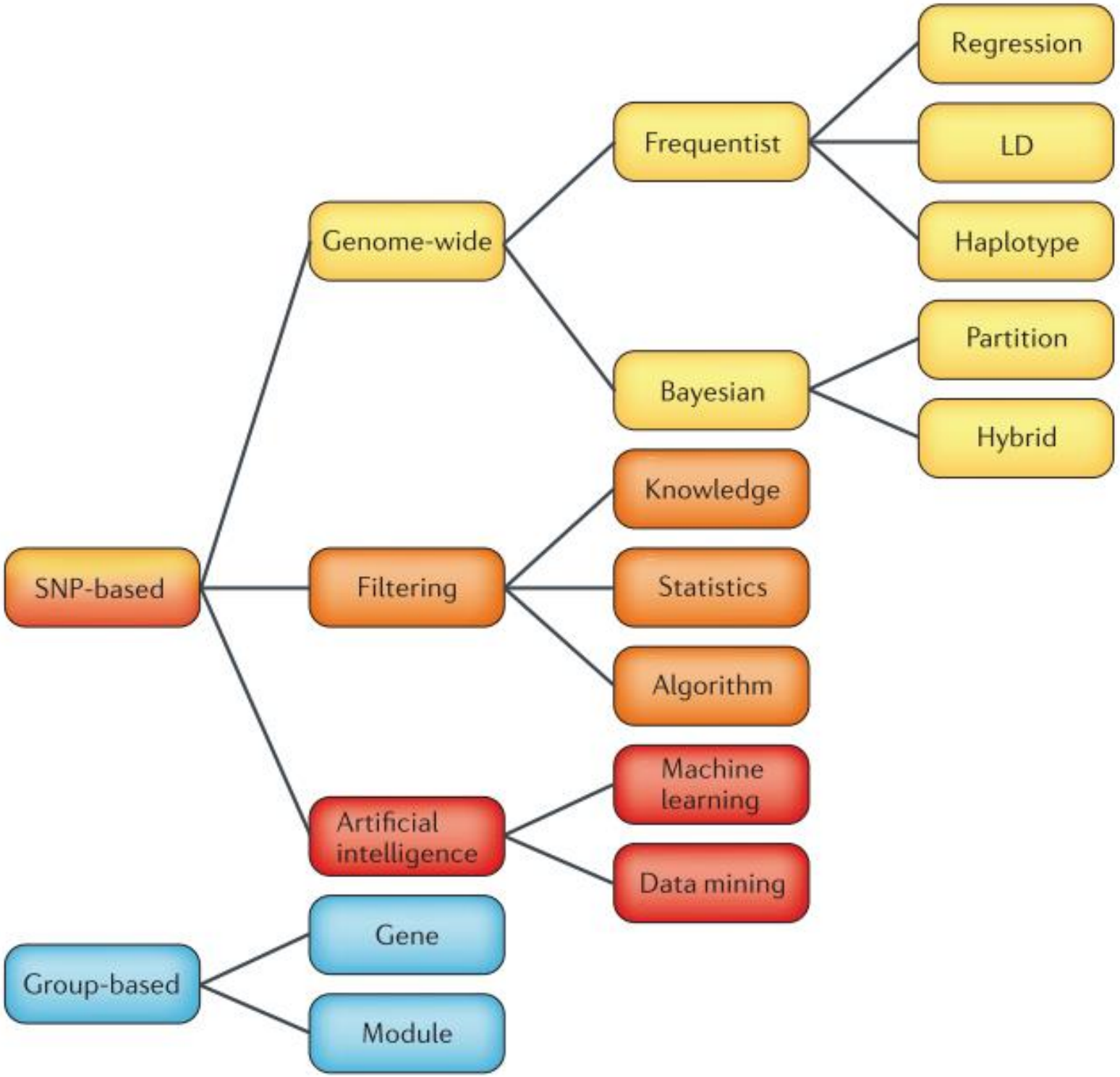
Методы префильтрации данных



Полный перебор



Сужение области перебора


SNP-based
Genome-wide
Frequentist
Regression
LD
Haplotype
Bayesian
Partition
Hybrid
Filtering
Knowledge
Statistics
Algorithm
Artificial intelligence
Machine learning
Data mining
Group-based
Gene
Module

В результате:

- Для модельных организмов:
  - Взаимодействия генов находятся и воспроизводятся.
  - Признаки без давления отбора

- Для человека:

  - Недостаточная мощность методов поиска взаимодействий генов
  - Нулевая воспроизводимостью результатов

American Journal of Epidemiology

Vol. 155, No.
Printed in U.S..

Sample Size Requirements for Association Studies of Gene-Gene Interaction

W. James Gauderman

In the study of complex diseases, it may be important to test hypotheses related to gene-gene ($G \times G$) interaction. The success of such studies depends critically on obtaining adequate sample sizes. In this paper, the author investigates sample size requirements for studies of $G \times G$ interaction, focusing on four study designs: the matched-case-control design, the case-sibling design, the case-parent design, and the case-only design. All four designs provide an estimate of interaction on a multiplicative scale, which is used as a unifying theme in the comparison of sample size requirements. Across a variety of genetic models, the case-only and case-parent designs require fewer sampling units (cases and case-parent trios, respectively) than the case-control (pairs) or case-sibling (pairs) design. For example, the author describes an asthma study of two common recessive genes for which 270 matched case-control pairs would be required to detect a $G \times G$ interaction of moderate magnitude with 80% power. By comparison, the same study would require 319 case-sibling pairs but only 146 trios in the case-parent design or 116 cases in the case-only design. A software program that computes sample size for studies of $G \times G$ interaction and for studies of gene-environment ($G \times E$) interaction is freely available (http://hydra.usc.edu/gxe). *Am J Epidemiol* 2002;155:478–84.

association; case-control studies; genetics; interaction; research design; sample size

nature genetics

Meta-analysis of the heritability of human traits based on fifty years of twin studies

Tinca J C Polderman[1,10], Beben Benyamin[2,10], Christiaan A de Leeuw[1,3], Patrick F Sullivan[4–6], Arjen van Bochoven[7], Peter M Visscher[2,8,11] & Danielle Posthuma[1,9,11]

Despite a century of research on complex traits in humans, the relative importance and specific nature of the influences of genes and environment on human traits remain controversial. We report a meta-analysis of twin correlations and reported variance components for 17,804 traits from 2,748 publications including 14,558,903 partly dependent twin pairs, virtually all published twin studies of complex traits. Estimates of heritability cluster strongly within functional domains, and across all traits the reported heritability is 49%. For a majority (69%) of traits, the observed twin correlations are consistent with a simple and parsimonious model where twin resemblance is solely due to additive genetic variation. The data are inconsistent with substantial influences from shared environment or non-additive genetic variation. This study provides the most comprehensive analysis of the causes of individual differences in human traits thus far and will guide future gene-mapping efforts. All the results can be visualized using the MaTCH webtool.

Specifically, the partitioning of observed variability into underlying genetic and environmental sources and the relative importance of additive and non-additive genetic variation are continually debated[1–5]. Recent results from large-scale genome-wide association studies (GWAS) show that many genetic variants contribute to the variation in complex traits and that effect sizes are typically small[6,7]. However, the sum of the variance explained by the detected variants is much smaller than the reported heritability of the trait[4,6–10]. This 'missing heritability' has led some investigators to conclude that non-additive variation must be important[4,11]. Although the presence of gene-gene interaction has been demonstrated empirically[5,12–17], little is known about its relative contribution to observed variation[18].

In this study, our aim is twofold. First, we analyze empirical estimates of the relative contributions of genes and environment for virtually all human traits investigated in the past 50 years. Second, we assess empirical evidence for the presence and relative importance of non-additive genetic influences on all human traits studied. We rely on classical twin studies, as the twin design has been used widely