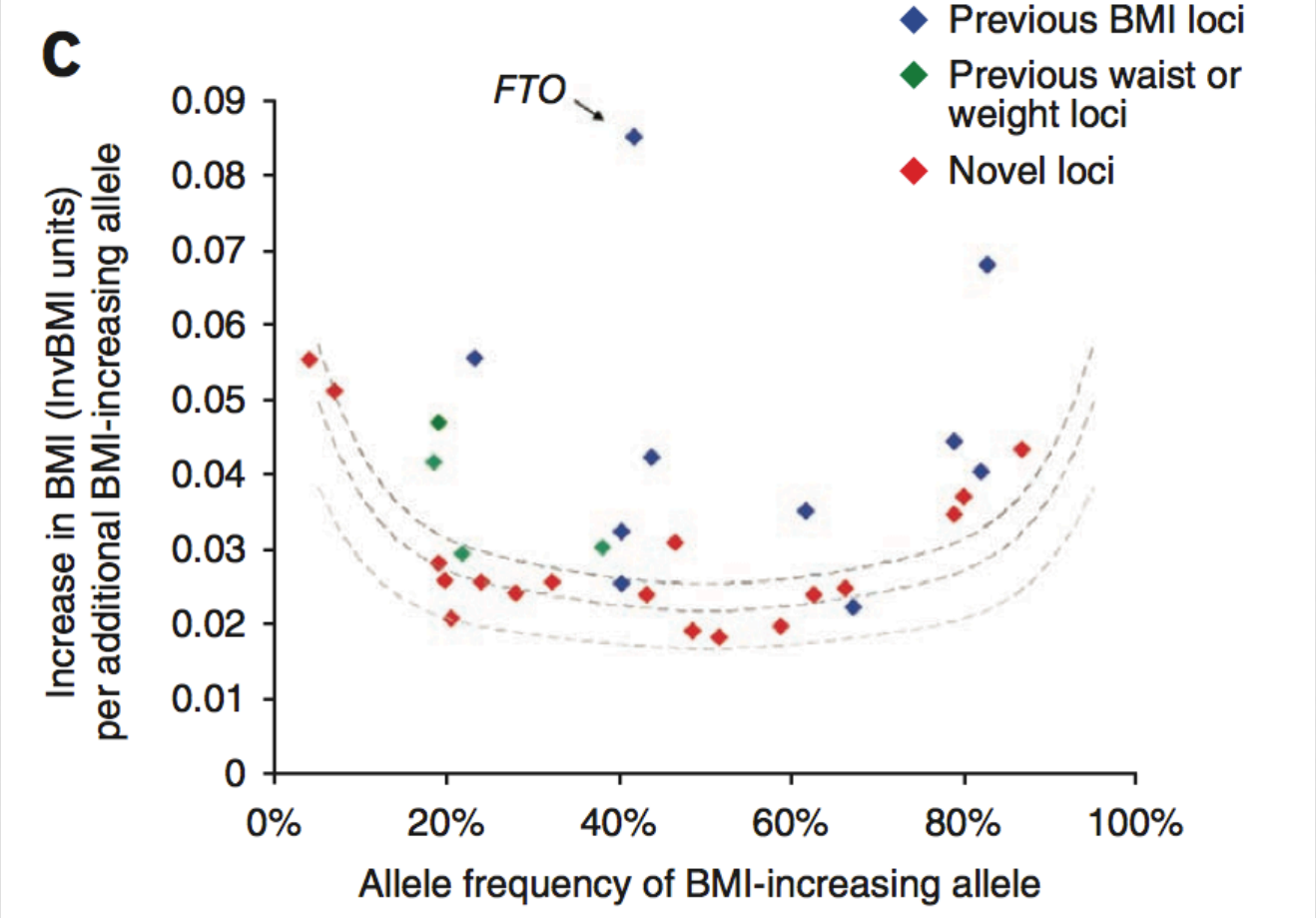
RARE VARIANT ASSOCIATION ANALYSIS METHODS

BIOINFORMAATIKA JOURNAL CLUB

EFFECT SIZE VS. ALLELE FREQUENCY

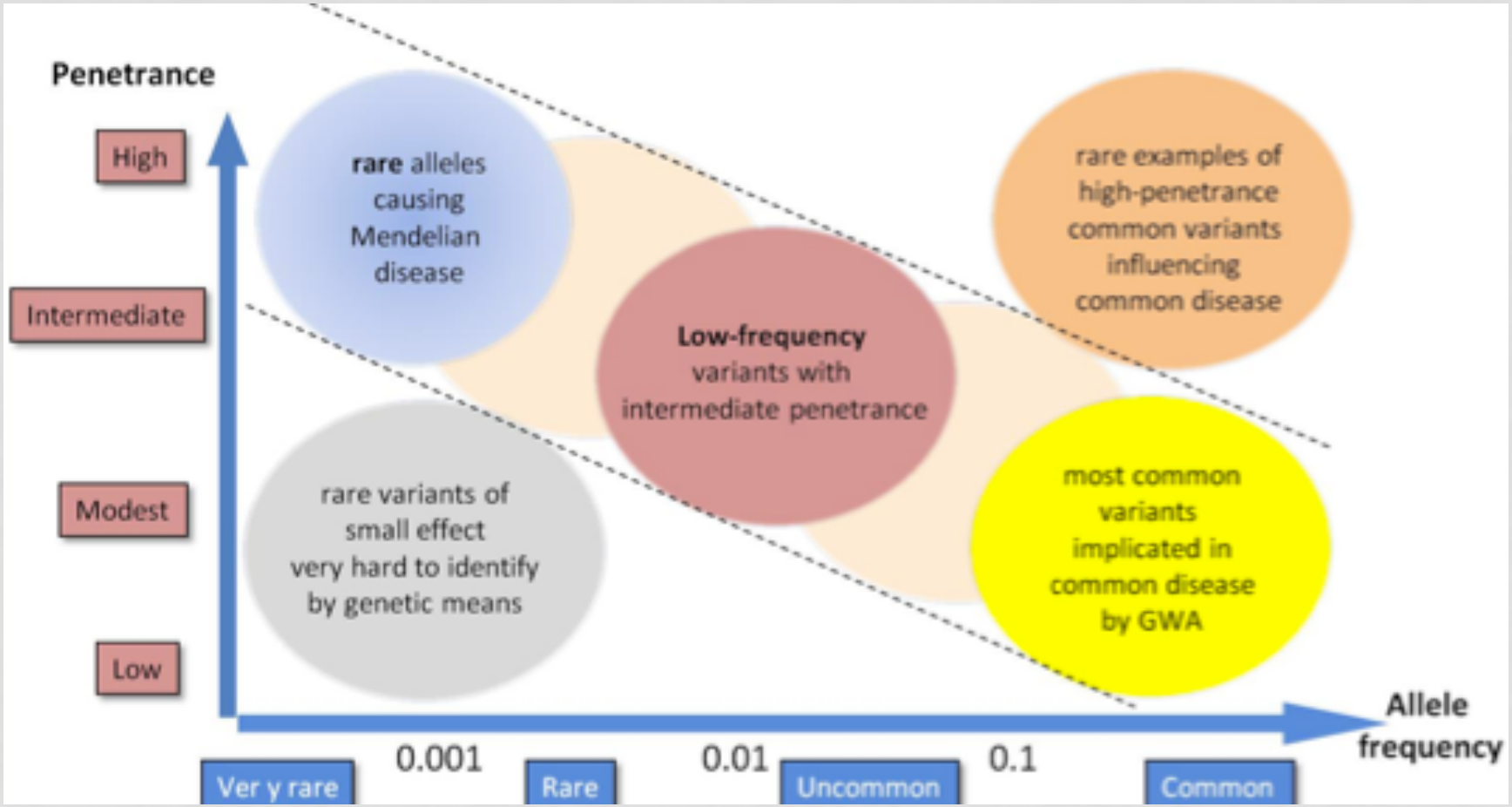


Speliotes et al. 2010

WHERE TO LOOK?

- CDCV – common disease / common variant hypothesis
- MRV – multiple rare variant hypothesis

HIDDEN HERITABILITY



Manolio et al. 2009

DEFINITION OF RARE VARIANTS

- MAF:
  - 5%-50% — common variant
  - 1%-5% — uncommon variant
  - <1% — rare variant

RARE VARIANTS IN CURRENT GENOTYPING PLATFORMS

Table 1 Approximate-low frequency/rare variant GWAS platform content

Platform	Affymetrix 500k	Affymetrix 6.0M	Illumina 370k	Illumina 550k	Illumina 610k	Illumina 1.2M
MAF < 0.05	55k	106k	9k	32k	35k	62k
MAF < 0.01	17k	35k	1k	7k	8k	22k

Asimit & Zeggini 2010

IMPUTATION PLATFORMS

- HapMap II
  - All markers ~2.5M
  - <5% ~380k
  - <1% ~90k
- 1000G data
  - All: 11.5M
  - <5% ~6.3M
  - <1% ~4.1M

KNOWN RARE VARIANT ASSOCIATIONS

- T1D - Nejentsev et al. 2009
  - Markers with MAF ~1% in *IFIH1* locus
  - Resequencing exons
- BP – Ji et al. 2008 (FHS)
  - Markers selected from *SLC12A1*, *SLC12A3*, and *KCNJ1* loci
  - Using phylogenetic conservation nad MAF < 0.001 as criteria
  - Possibly damaging markers were selected using bioinformatic tools
- Lipids – Cohen et al. 2004 (Dallas Heart Study)
  - 256 samples sequenced for candidate genes (*ABCA1*, *APOA1*, and *LCAT*)

HOW TO ANALYSE?

- Single marker analysis
- Multiple marker tests
- Collapsing and aggregation of multiple markers

SINGLE MARKER ANALYSIS

- χ^2 test for contingency table
- Fisher's exact test
- Cochran-Armitage test for trends
- Linear and logistic regression

MULTIPLE MARKERS ANALYSIS

- Fisher's method
- Hotelling's T^2 test
- Multiple regression
  - Require multiple degrees of freedom☹
- multivariate distance matrix regression (MDMR)
- kernel-based association test (KBAT)

COLLAPSING AND AGGREGATION METHODS

- Multiple variants
  - Collapsing methods
  - Aggregation methods
    - Weighted sum
    - Li & Leal partition according to MAF
  - Bi-directional Effect Methods: for example C-alpha

	callrate	mutations	mut. load
Ind. 1	9	5	5/9
Ind. 2	10	5	5/10
Ind. 3	10	1	1/10
Ind. 4	8	2	2/8

ISSUES

- not attain a power as high as when complete resequencing data are used
- appropriate genetically similar reference panel

CONCLUSION

- stringent quality control procedures is necessary
- rare variants are difficult to genotype and challenging to impute, so meticulous quality checks are essential before declaring association