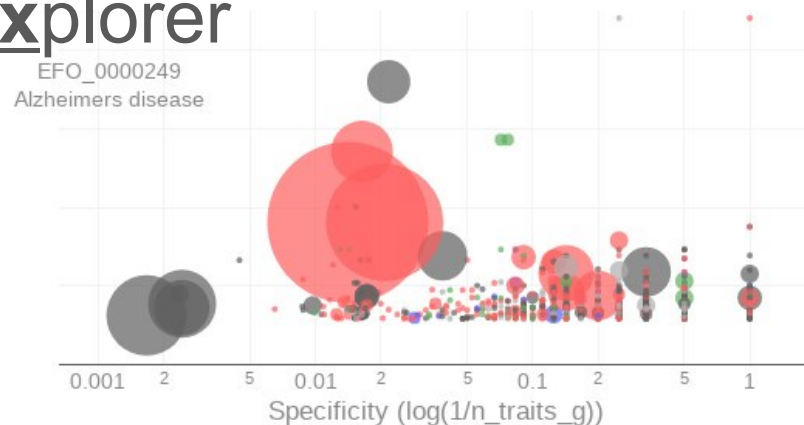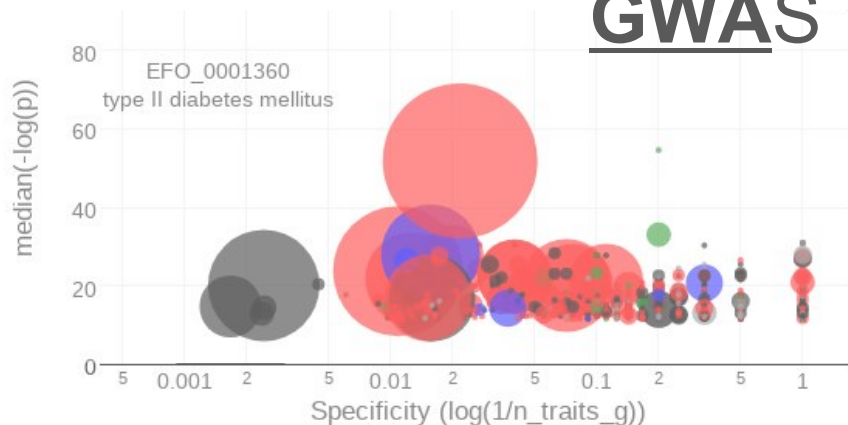# GWA-X Introduction

GWAS Explorer

2017 November 6

Jeremy Yang
UNM Translational Informatics
Division

GWA-X Intro: Concept

Genome wide association studies (GWAS) can reveal important genotype to phenotype associations, however, data **quality** and **interpretability** issues must be addressed. The GWA-X approach enables rational filtering and interpretation of GWAS via metrics, methods, and interactive visualization. Each GWAS association is evaluated, using a continuous numerical scale, for **confidence** and **specificity** linking a protein-coding gene and clinical phenotype. Appropriate thresholds for various use cases are left to the judgement of users.

GWA-X Intro: Background
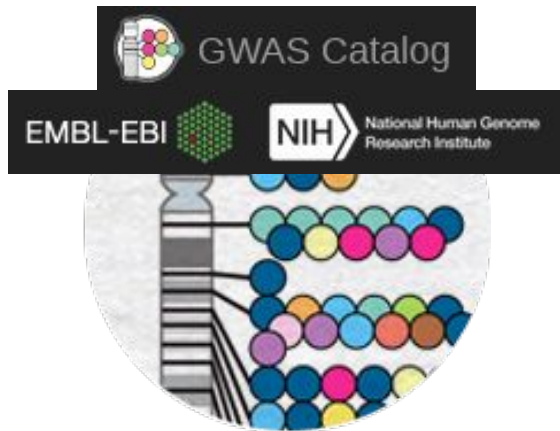
- Dataset: GWAS Catalog, http://www.ebi.ac.uk/gwas/
- GWAS Catalog studies all:
  - Have a "study_accession"
  - Have a publication (PubMedID), but not uniquely.
- Mapped genes:
  - To in-gene SNP, or upstream/downstream genes to intergenic SNP.
- Raw files:
  - gwas_catalog_v1.0.1-studies_r2017-10-10.tsv
  - gwas_catalog_v1.0.1-associations_e90_r2017-10-10.tsv
- Genes specified by symbol (N=25777, TCRD: mapped=12148, not=13629).
- Diseases and traits specified by EFO(5710), GO(274), HP(107) Orphanet(97).

GWA-X Intro: Background

Studies file:

- DATE_ADDED_TO_CATALOG
- PUBMEDID
- FIRST_AUTHOR
- DATE
- JOURNAL
- LINK
- STUDY
- DISEASE/TRAIT
- INITIAL_SAMPLE_SIZE
- REPLICATION_SAMPLE_SIZE
- PLATFORM_[SNPS_PASSING_QC]
- ASSOCIATION_COUNT
- MAPPED_TRAIT
- MAPPED_TRAIT_URI
- STUDY_ACCESSION

Associations file:

- DATE_ADDED_TO_CATALOG
- PUBMEDID
- FIRST_AUTHOR
- DATE
- JOURNAL
- LINK
- STUDY
- DISEASE/TRAIT
- INITIAL_SAMPLE_SIZE
- REPLICATION_SAMPLE_SIZE
- REGION
- CHR_ID
- CHR_POS
- REPORTED_GENE(S)
- MAPPED_GENE
- UPSTREAM_GENE_ID
- DOWNSTREAM_GENE_ID
- SNP_GENE_IDS
- UPSTREAM_GENE_DISTANCE
- DOWNSTREAM_GENE_DISTANCE
- STRONGEST_SNP-RISK_ALLELE
- SNPS
- MERGED
- SNP_ID_CURRENT
- CONTEXT
- INTERGENIC
- RISK_ALLELE_FREQUENCY
- P-VALUE
- PVALUE_MLOG
- P-VALUE_(TEXT)
- OR_or_BETA
- 95%_CI_(TEXT)
- PLATFORM_[SNPS_PASSING_QC]
- CNV
- MAPPED_TRAIT
- MAPPED_TRAIT_URI
- STUDY_ACCESSION

GWA-X Intro: Processing

- Map to gene symbols
- Map to disease/trait terms (EFO, GO, etc.)
- Compute counts for studies:
  - **trait_count**: traits per study
  - **assn_count**: associations per study
  - **snp_count**: SNPs per study
  - **gene_r_count**: genes-reported per study
  - **gene_m_count**: genes-mapped per study
  - **study_count**: studies per publication

Studies vary greatly in their specificity and interpretability. Each publication represents some measure of scientific consensus, but may be dispersed over more or less traits, SNPs and genes.

GWA-X Intro: Processing (2)

- Compute pairwise gene-trait (gt) statistics:
    - **gsymb**: gene symbol
    - **trait_uri**: trait URI
    - **trait**: trait name
    - **n_study**: studies per this gt
    - **n_snp**: SNPs per this gt
    - **n_traits_g**: traits per this gene
    - **n_genes_t**: genes per this trait
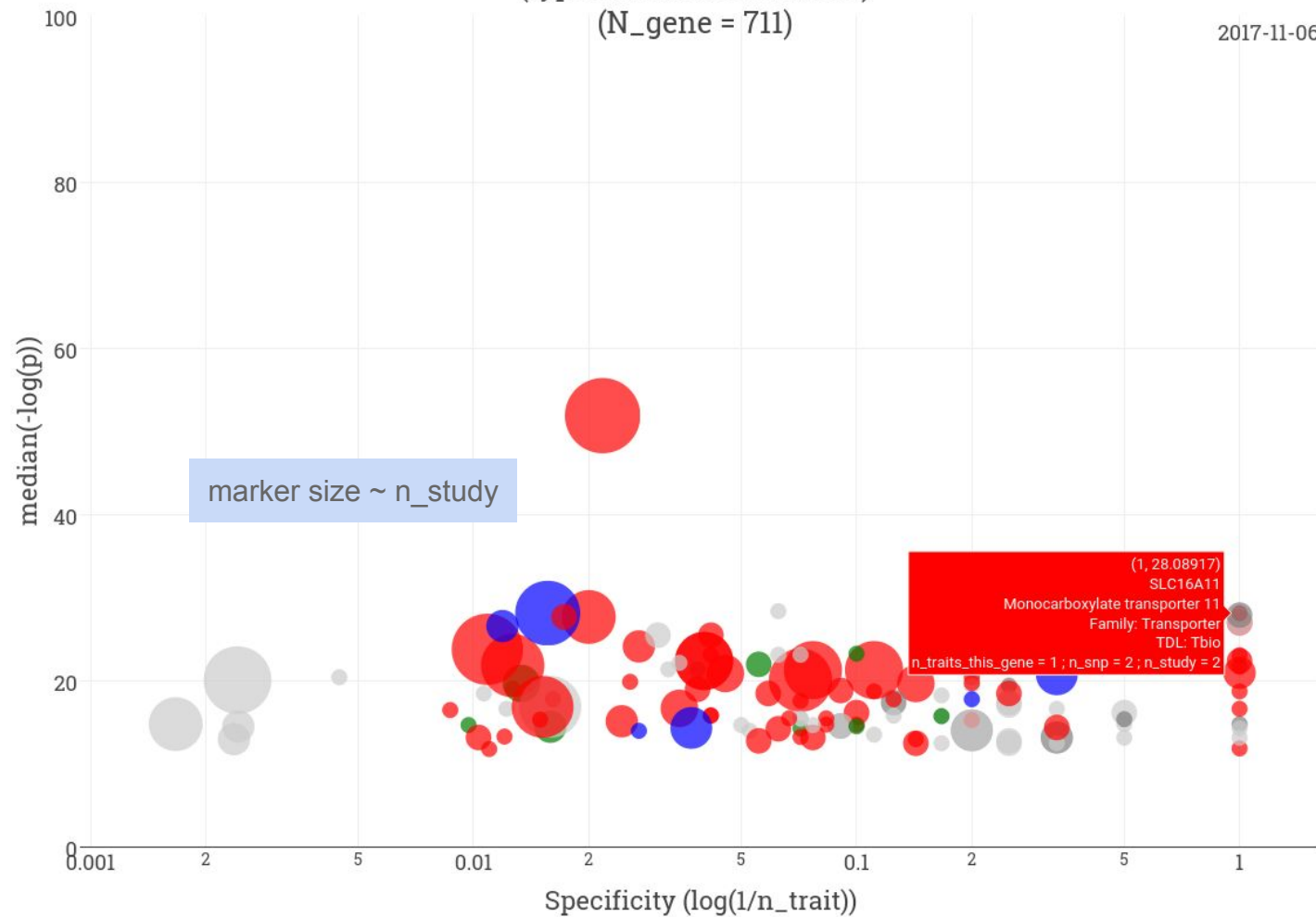   - **p_median_nlog**: median(-log(pvalue))
-

Gene-trait associations vary in their quantity and quality of supporting evidence. Optimally, SNP to gene mapping is strong, p-values are good, and multiple studies replicate the result.

GWAS Catalog TRAIT: EFO_0001360
(type II diabetes mellitus)
(N_gene = 711)

2017-11-06

median(-log(p))

marker size ~ n_study

(1, 28.08917)
SLC16A11
Monocarboxylate transporter 11
Family: Transporter
TDL: Tbio
n_traits_this_gene = 1 ; n_snp = 2 ; n_study = 2

Specificity (log(1/n_trait))

References and links:

- https://www.ebi.ac.uk/gwas/
- https://www.ebi.ac.uk/gwas/docs/file-downloads
- http://rpubs.com/jeremyjyang/gwax_trait_02
- http://rpubs.com/jeremyjyang/gwax_trait_01
- http://rpubs.com/jeremyjyang/gwax_gene_01
- https://github.com/jeremyjyang/idg-gwax-pre (scripts, SQL and R)
- https://www.dropbox.com/sh/xt1niufuv45881k/AABTMqDD1vActskQtMjiij9da? dl=0 (Postgresql dump and gt_stats.csv)