1. Whole phylogeny of genes in genotype to phenotype analysis
   1. Dendrogram ANOVA [https://onlinelibrary.wiley.com/doi/full/10.1111/evo.13492#.Wv23DRFGI\_M.twitter](https://onlinelibrary.wiley.com/doi/full/10.1111/evo.13492" \l ".Wv23DRFGI_M.twitter)
2. eQTL analysis
   1. Matrix eQTL
      1. Shabalin 2012 “Matrix eQTL: ultra fast eQTL analysis via large matrix operations”
      2. See sample R scripts
         1. Format files
         2. Install BLAS on Linux
            1. Or, environment of Revolution R Enterprise 5.0.1
         3. Test runs
      3. Who has used it?
         1. 449 citing papers
         2. <https://scholar.google.com/scholar?cites=11137296657696040463&as_sdt=2005&sciodt=0,5&hl=en>
   2. FastMap
      1. Gatti 2008 “FastMap: fast eQTL mapping in homozygous populations”
      2. <http://comptox.us/fastmap.php>
      3. cited by 41, ~1 study since 2016
   3. MT-eQTL
      1. Li 2017 “An empirical Bayes approach for multiple tissue eQTL analysis”
      2. For multiple tissues.
      3. Cited by 15
   4. FASTASSOC in MERLIN
      1. Chen 2007 “Family-based association tests for genomewide association scans”
         1. 408 citing
         2. <https://scholar.google.com/scholar?cites=9375494418915028580&as_sdt=2005&sciodt=0,5&hl=en>
         3. <http://csg.sph.umich.edu/abecasis/merlin/reference.html> software
      2. Abecasis 2002 “Merlin–rapid analysis of dense genetic maps using sparse gene flow trees”
         1. <https://csg.sph.umich.edu/abecasis/Merlin/> software
   5. QTL Cartographer
      1. Old and probably not fast?
3. GWA meta-analysis
   1. Mega-analysis? One genotype
   2. Binning and manhattan of manhattans
      1. Jeck 2012 Review: a meta-analysis of GWAS and age-associated diseases
   3. Combine p values
      1. Combine p values using Fisher’s method
         1. Chi-squared distribution with 2k (k = number of studies/ phenotypes) df
         2. All studies weighted equally – ok for identical sample sizes
         3. SNP effect direction is ignored
      2. Combine p values using Z-score method
         1. 1:1 p value to z score transformation
         2. Weight can differ by study (pheno)
         3. Begum 2012 ref 8
         4. Use METAL! (see below)
   4. Sum beta scores for each SNP across all phenotypes
      1. Should use linear models (see below)
      2. Still report with manhattan of manhattans?
      3. Stahl 2010 Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci
         1. Inverse variance weighting
         2. Average inverse variance-weighted beta coefficients
         3. Sum inverse variance-weighted z scores after genomic control correction
            1. De Bakker 2008

Alternative z-score: scale to effective sample size

Especially useful if beta/ SE units differ across studies

Software: MANTEL. Perl and R scripts.

Trying in strawberry perl on laptop

<http://debakker.med.harvard.edu/resources.html>

to try in perl on Linux

* + - 1. Cochran’s Q tests to assess heterogeneity across collections
    1. More statistical power than Z scores!
    2. Requires that scale is the same for each study (pheno) – same units and transformations – should work for across transcripts
  1. Fixed effects models
     1. Combine effect sizes
     2. Inverse variance weighting: weight each study (phenotype?) by inverse of squared standard error (Zeggini 2009 Meta-analysis in genomewide association studies)
     3. Cochran-Mantel-Haenszel approach
     4. Assumes no between-study (phenotype) heterogeneity
     5. Assumes genetic effects are the same across studies (phenotypes)
        1. I feel this does not apply in the case of cis effects on the transcriptome? Or for trans hotspots is it fine?
     6. Narrower CI and lower p-val than random fx
     7. See Nakaoka 2009 (ref 15 Begum 2012)
  2. Random effects models
     1. Combine effect sizes
     2. Assumes mean effect of each SNP is different in each study (pheno)
        1. This could be true for transcriptome, ESPECIALLY cis loci… but is it true for trans hotspots?
     3. Can calculate between-study variance in heterogeneity
        1. Could this be an informative measure of cis vs. trans?
     4. Best for: generalizability of trait-SNP associations, estimating average SNP effect size & uncertainty across traits (Pereira 2009; Ioannidis 2007)
     5. Con: low discovery power
     6. Assumes different studies (phenotypes) are estimating different effects
  3. Can try both fixed and random effects (Zeggini 2009)
  4. Bayesian methods
     1. Bayes Factor (Evangelou 2013 ref 38; Wellcome Trust Case Control Consortium 2007)
     2. Posterior probabilities that SNPs are null (Evangelou 2013 ref 39; Samani 2007)
     3. Best for cumulative analysis
  5. CPMA (cross phenotype meta analysis)
     1. Multiple associations at a single marker across different phenotypes (Evangelou 2013 ref 79; Cotsapas 2011)
     2. May not work for traits across same individuals – low power
  6. Bioinformatics tools
     1. PrediXcan
        1. <https://github.com/hakyimlab/PrediXcan>
        2. Can try skipping the GWAS step and just do this
        3. Use on Linux
     2. METAL (Willer 2010) \*\*
        1. Pro: explicitly for meta-analysis of GWA
           1. And flexible input file format!
        2. Con: time to compute is O(n)
           1. 15 studies x 2.5 million SNPs, = 36 mill assoc stats

< 6 min to compute

790 MB memory

* + - 1. FYI formula in paper has typo, but package correctly implements (Begum 2012)
      2. Fixed effect model
         1. I think only??
      3. Weighted Z-score method based on sample size, p-value, effect direction
         1. OR effect-size method weighted by SE within study
      4. Use on Linux
      5. For large studies, try this first
    1. GWAMA (Magi 2010) \*\*
       1. Pro: explicitly for meta-analysis of GWA
       2. Con: may need output format from PLINK
          1. [http://zzz.bwh.harvard.edu/plink/anal.shtml#qt](http://zzz.bwh.harvard.edu/plink/anal.shtml" \l "qt)
          2. <https://link.springer.com/protocol/10.1007%2F978-1-62703-447-0_8>
       3. Nifty features?
       4. Can calculate heterogeneity across studies -- decide between fixed or random effects model
       5. Use on Linux
    2. *MetABEL (Aulchenko 2007)* 
       1. *Within GenABEL in R*
       2. *Fixed effect model*
       3. *Visualization tool*
       4. *Discontinued ~ May 2018* [*http://www.genabel.org/node/309*](http://www.genabel.org/node/309)
    3. PLINK (Purcell 2007) \*\*\*
       1. Pro: Explicitly for meta-analysis of GWA
          1. Con: GWA output needs to be \*.assoc format
       2. Can do fixed effects and random effects meta-analysis
       3. implementation
          1. **HINT** If performing meta-analysis on a large number of large files (e.g. 10+ files of imputed results, each with over 2 million entries), one might need to perform this one chromosome at a time, with the --chr option, as all the result files might not fit in memory in one go otherwise.
          2. Need to input individual .assoc files. May be very inefficient across 1000s of phenotypes. Could bin into sub-chromosomal segments?
       4. <http://zzz.bwh.harvard.edu/plink/metaanal.shtml>
    4. MAGENTA (Broad institute)
       1. Hypothesis testing
       2. Gene set enrichment analysis
    5. Others in R
       1. Metafor
          1. Fixed and random and mixed fx
          2. Can try on windows, cool!
          3. Use forest plot … summarizes each effect across experiments

See Viechtbauer 2010: example forest plot is 13 studies and seems best suited for small numbers.

Also many not make sense for MANY EFFECTS (GWAS)

* + - 1. Rmeta
         1. Can also try on windows
         2. Appears to be only treatment/ control
      2. CATMAP
         1. Can also try on windows
         2. Says only for case-control or family-based (TDT) data

1. Questions
   1. Can we look into LD?
   2. Do I have interaction or environmental effects?
   3. Are my phenotypes correlated?
      1. If so: bivariate meta-analysis (Evangelou 2013 ref 84, 85; Bagos 2008, Bagos 2012)
      2. May not have enough individuals for this
   4. Are my SNPs correlated?
      1. LD correlations of SNPs
      2. Genome-wide stepwise selection: joint effects of all selected SNPs
         1. Do not have enough individuals for this
   5. Evaluate HWE?