Primary ESP Article

Primary ESP Article

- The Primary ESP article will be modeled on the highly cited WTCCC GWAS article

  (Wellcome Trust Case Control Consortium Nature 2007)

- Five years after publication it is highly cited on QC and analysis of GWAS data

- Analysis of many loci across many traits allows for general lessons that can’t be drawn from one trait

Primary ESP Article

- Goal of Primary ESP article is to paint big picture

- Strategies and lessons learned for future association studies using exome sequence data
  - Calling of SNVs & Indels
  - Exome data QC
  - Design of exome sequence association studies
  - Association analysis of exome sequence data

Benefit of Primary ESP Article

- High impact and quintessential article on the analysis of exome sequence data

- Will document to NHLBI and the world scientific community the success of the NHLBI-ESP program
  - Demonstrate that future funding is warranted and important

Benefit Primary ESP Article

- First publication on a large well phenotyped exome sequence data set (~6900 exomes)
- Proof of principle setting standards for analysis and interpretation of exome data
  - High impact journal
    - Nature, Science or Cell
  - Expectation
    - Highly cited article

Benefit of Primary ESP Article

- All ESP collaborators benefit by being co-authors on a high impact article

- Individual working groups can still publish detailed analysis of their traits
  - Primary manuscript will not preclude publication of individual analyses and follow-up efforts

ESP needs to Publish Main Findings ASAP or Potentially Not at All

- ESP exome data is being made publically available through dbGAP
- Investigators outside of ESP can analyze your trait of interest and publish it
- An analysis of many traits could also be published by investigators who are not members of ESP

BUT

- When outside investigators publish their findings no co-authorship for ESP investigators
- No ESP authorship banner

Logistics of Primary ESP Article

Study Design, QC and Analysis of
6900 Exomes

ESP Study Design

- Description of initial study design
- Benefits and limitations of study design
- Modifications of the study design
- Description of available ESP phenotypes
  - Including covariates
    - e.g. smoking status, age

SNV & Indel Calling

- Description of
  - Alignment of reads
  - Quality of exome data
  - Method used to call SNVs
  - Method used to call indels

QC of Exome Data

- Evaluate various QC approaches
  - More or less aggressive filtering?
  - More or less missing data?
  - What are the major pitfalls?

- Draw general lessons and recommendations that reflect our collective best view and advice for future studies

Analysis of Traits

- Evaluate analysis methods systematically and draw general lessons
  - Choice of tests and type of variants to be analyzed
    - Single variant tests
    - Aggregate rare variant association methods
      - CMC, MZ, VT, SKAT, many others …
- In analysis of single trait, hard to draw objective lessons. With many traits, easier to draw conclusions
  - Which set of tests captures most of our findings?
  - Which set of tests would we recommend for the future?

Analysis of Phenotypes

- Analysis will be trait specific
  - Need guidance from working groups
  - Which covariates (e.g. sex, smoking, BMI)
  - Which exclusions for which trait

- Expert information obtained from conveners of the phenotype specific groups

Phenotypes to be Included in Primary ESP Article

- Relevant to Heart, Lung and Blood
- Inclusion of traits in article not based upon positive association findings
  - Mixture of traits with positive & negative findings
- WTCCC paper on CNVs did not have major new signals but is still a landmark Nature paper

Timeline for Publication of Primary ESP Article

	April	May	June	July	August
QC of Exomes					
Call & QC Indels					
Analysis of Phenotypes					
Prepare Manuscript					
Submit Manuscript					

Impact on ESP Articles in Preparation

- The Primary ESP article will not be published immediately
- Manuscripts currently under preparation can still be submitted for publication

Comments and Questions?

We welcome your input

Publication of Imputation Results

- The ESP exome sequence data has been imputed into very large data sets of well phenotyped African-American and European American
  - European American
    - ~15,000 individuals
  - African American
    - ~30,000 individuals
- This strategy has been extremely successful
  - Blood related traits article in African Americans is currently under review at Nature Genetics

Publication of Imputation Results

- Additional success stories being obtained from analysis of imputed data
  - Some results may overlap with findings from the exome sequence data
- It may be ideal to try to get back to back publication of the primary ESP article and imputation article
  - They complement each other nicely
  - Imputation article demonstrates additional utility of the ESP exome data for a wide variety phenotypes
    - Including phenotypes not related to heart, lung and blood

Analysis of Phenotypes

- Analysis of African-Americans and European-Americans will be performed separately
  - Race will be determined via PCA analysis
- Race specific association analysis results will be combined using meta-analysis