Multi-Analyte Data Analysis Reduces False Positives in Cystic Fibrosis

T Henry, J Marcy, D Shirazi, and S Berberich

Iowa Newborn Screening Program

March 3, 2016
Multi-analyte data analysis

• Cancer genomics
  – MammaPrint breast cancer recurrence assay
  – Predicts risk for breast cancer recurrence based on gene expression data from 70 target genes
  – Developed using unbiased gene selection based on patient outcomes
Multi-analyte analysis: Cystic Fibrosis

- Why can’t NBS steal a page from the cancer book?
- Many NBS disorders are multi-system disorders
  - CF: respiratory, pancreatic, hepatobiliary, reproductive, etc.

- Perhaps additional biomarkers can be found that improve discrimination of disease states
Multi-analyte analysis: Cystic Fibrosis

• Number of specimens: 27,961 (approx. 9 mos)
• Confirmed CF Dx: 9
• CRMS: 7
• Carriers: 29
• Total clinical cases: 45
• Normals: 27, 916
• Data points per specimen: 76 (demographic, facility, analyte, etc)
• Total dataset: 2,125,036
Multi-analyte analysis: Cystic Fibrosis

- Analysis performed using Partek Genomic Suite
- Data median normalized
- ANOVA performed across all analytes using CF diagnosis as the comparator groups (confirmed CF versus CRMS/carriers)
- Analytes with a greater than 2-fold change and false discovery rate of $p<0.05$ used for unsupervised hierarchical clustering
Multi-analyte analysis: Cystic Fibrosis
Multi-analyte analysis: Cystic Fibrosis

- Normals
- CRMS/carriers
- Confirmed CF

Hierarchical Clustering

[C8]  
Cit/arg  
IRT

confirmed CF  
CRMS  
carriers  
normals
Multi-analyte analysis: Cystic Fibrosis

• IRT + cit/arg + [C8] (dataset n=27,961):
  – Accurately stratified 8 of 9 confirmed CF cases (1 false negative) from CRMS/carriers
  – Accurately stratified 7 of 7 CRMS from confirmed CF cases
  – Accurately stratified 28 of 29 CF carriers from confirmed CF cases
  – Misclassified 4 normals into CRMS/carrier group

• IRT/DNA: potentially less reflexed for DNA mutation, less for sweat testing (1 or 2 mutations forwarded)
Multi-analyte analysis: Cystic Fibrosis

• What’s next:
  – Additional data being added to refine/improve CF model training set
  – Pilot project with multi-analyte model run in parallel with current CF screening algorithm

• Multi-analyte analysis is challenging but provides opportunities for discovery
Multi-analyte: special populations

- Preterm, LBW, NICU

GA, weight, and gender variance scatterplot
Multi-analyte: special populations

- Preterm, LBW, NICU

Sub-population at GA $\leq 33$ weeks
Multi-analyte: special populations

- Preterm, LBW, NICU
Multi-analyte: special populations

- TPN signature (in progress)
Conclusions

• Screening is for risk stratification
• Multi-analyte data analysis improves risk stratification for true CF cases and CRMS/carriers
• Algorithms can be embedded in the laboratory to reduce the number of cases forwarded to follow up
• Existing analyte data may be leveraged to provide improved screening risk stratification
Acknowledgements

• Iowa Department of Public Health
• University of Iowa Hospitals and Clinics
• State Hygienic Laboratory at the University of Iowa