Weight and see: A systematic overview of category rankings within the semi-quantitative metric of the NC NEXUS project

Girnary Z, Milko L, Strande N, O'Daniel J, Foreman K, Booker J, Boshe L, Couser N, Gucsavas-Calikoglu M, Aylsworth AS, Frazier D, Roche M, Vora N, Powell C, Berg J, Powell B
University of North Carolina at Chapel Hill, School of Medicine, Chapel Hill, NC

Introduction

Newborn screening (NBS) enables early detection and presymptomatic intervention for selected conditions based on population impact and availability of efficacious treatments. Next generation sequencing (NGS) would enable the inclusion of vastly more conditions for which early treatment or surveillance is crucial, but also presents significant ethical complexity for conditions where there is no proven medical intervention, or where avoiding the "diagnostic odyssey" might be the only benefit. NGS can also reveal findings where the initial benefit may not be to the child, such as adult-onset conditions or carrier screening.

Through the utilization of a framework that characterizes categories of conditions based on age at onset/intervention and "medical actionability," NC NEXUS (North Carolina Newborn Exome Sequencing for Universal Screening) seeks to develop a core panel of medically actionable, childhood-onset disorders (including standard NBS conditions). This process has been applied to over 750 gene/condition pairs. All parents will receive results involving pathogenic variants that have a high threshold for such a classification in this core panel.

Determining which gene/condition pairs should be included in this group is difficult when phenotypes receive similar scores but the Binning Committee thinks some should be included and others excluded. We present our methods to establish a threshold for medical actionalibity using manual classification through consensus group discussion and linear discriminant analysis.

A semi-quantitative metric is used to score "Medical Actionability" on a 0-3 Point Scale for 5 Criteria:

Category	Description	0	1	2	3
Severity of Disease	"What is the effect on morbidity / mortality to an individual carrying a pathogenic variant in this gene?"	Modest Morbidity	Serious / Chronic Morbidity	Possible Death	Sudden Death or Unavoidable Death in Childhood (<10yo)
Likelihood of Outcome	"What is the chance that a threat will materialize?"	<1%	1-5%	5-49%	>50%
Efficacy of Intervention	"How effective are the interventions for preventing harm in a presymptomatic individual?"	No Effective Intervention	Minimally Effective	Modestly Effective	Highly Effective
Acceptability of Intervention	"How acceptable are the interventions in terms of the burdens or risks placed on the individual?"	No Effective Intervention	Minimally Acceptable	Modestly Acceptable	Highly Acceptable
Knowledge Base	"What is the evidence base for decisions about the natural history of the disease, and interventions used for preventing serious outcomes?"	Poor	Minimal	Modest	Substantial Evidence and/or Practice Guidelines

Results – Group Consensus / Manual Review

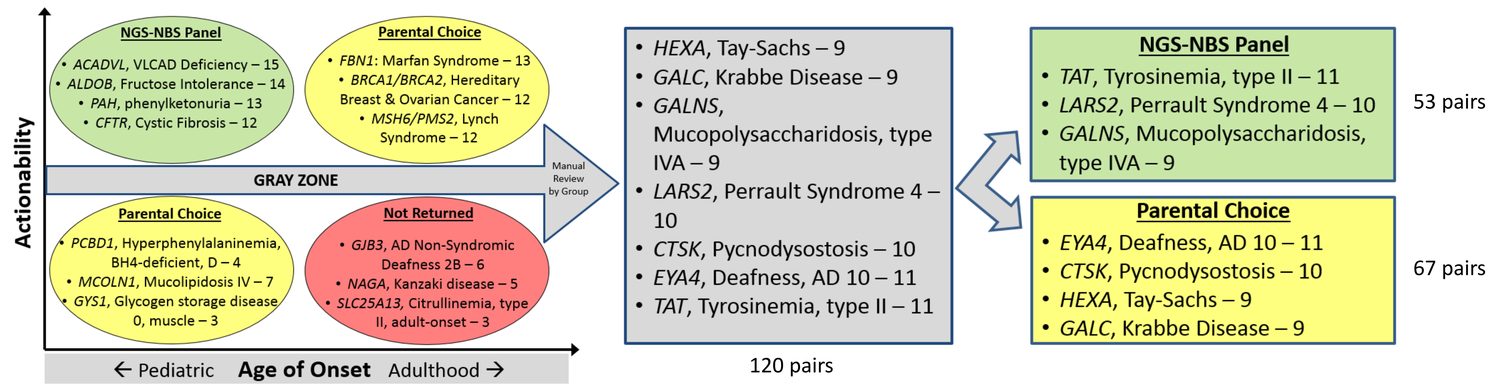

Fig 1. While the SQM provides an overall ranking for actionability, the delineation of which genes are to be included in the core panel still requires making a binary decision about individual gene/condition pairs. Through ongoing group discussion, it was recognized that some conditions with the same overall score were discordant with regard to whether the group felt inclusion on the core panel would be appropriate, suggesting that the different components of the metric might have different weights in determining actionability. 120 gene/phenotype pairs in the gray zone were classified as either "In" or "Out" of the core panel. 53 were classified to be "In" (Green) and 67 were classified to be "Out" (Yellow).

Results – Linear Discriminant Analysis

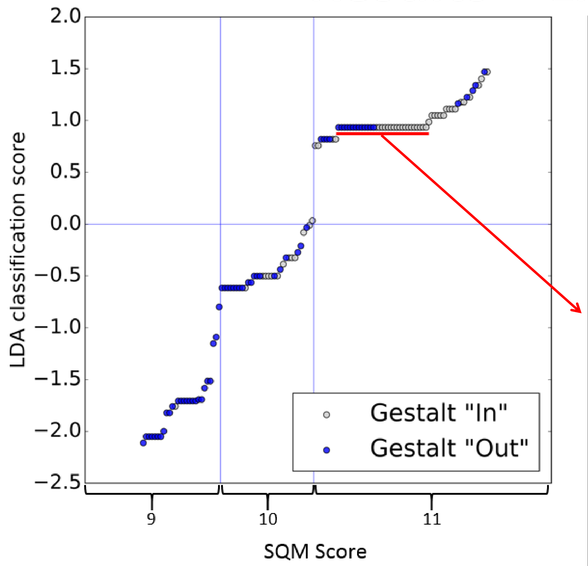

Fig. 2: **Linear Discriminant Analysis**
- "Open" circles = "In" by gestalt
- "Filled-in" circles = "Out" by gestalt
- Data with >0 scaled score = "In" by LDA
- Data with <0 scaled score = "Out" by LDA

←**Fig. 2 Continued**: The same 120 gene/condition pairs were analyzed via Linear Discriminant Analysis. This method seeks a weighted combination of input variables (the individual component scores of the SQM) that maximizes the variance among classified groups (the consensus gestalt classifications of the Binning Committee).

Many gene/condition pairs received the same score across the five components of the metric, but some were classified by the group to be "in" and some were said to be "out".

The LDA algorithm that best fit our "Gestalt" classifications assigned the following relative weights:

A	>	S	>	K	>	E	>	L
1.14		1.10		1.02		0.89		0.85

Although "Acceptability" received the highest weight and "Likelihood" received the lowest, the weights are similar among the components, and the LDA classifier would change the categorization of few gene-disease pairs compared to a simple SQM score cutoff of ≥ 11

	Classifier "In"	Classifier "Out"
Gestalt "In"	38	14
Gestalt "Out"	22	45

Fig. 3: **Remaining misclassification after LDA**

- There were 38 pairs that the Binning Committee thought should be "In" the core panel, and 45 that were decided should be "Out" and therefore be Parental Choice. These classifications were echoed by the classifer (~70% agreement).
- There was a discrepancy between the group and the classifier for ~35 gene/condition pairs (~30%). In other words, about ~35 pairs were misclassified.

Methods

The NC NEXUS Binning Committee:
- Composed of MDs, PhDs, Genetic Counselors, Biocurators, and Experts

Binning by Semi-Quantitative Metric (SQM): How do we define "actionability?"
- 2 reviewers per gene/condition pair → Extensive literature review
- Present to group for consensus review → Document group discussion → Utilize SQM to establish a final score
- There was no consensus for single numeric threshold for inclusion onto core panel
- Are some components of the overall score more important for actionability? (Could we weight factors to better decide?)

Weighting by Linear Discriminant Analysis (LDA): Does weighting components of the score help refine the threshold?
- Analyzed about ~120 gene/condition pairs with intermediate total SQM scores, and found that there existed differences in opinion within the Binning Committee
- 2 reviewers re-assigned gene/condition pairs → Re-reviewed → Documented discussion as a group → Assign "Gestalt" characterization as "training" set
- Apply LDA (using sklearn version 0.17, http://scikit-learn.org) to see if an algorithm can establish categorical weights to classify "Gestalt in" as "In" and "Gestalt Out" as "Out"

Discussion / Conclusions

Semi-Quantitative Metric

- Regarding the actionability of gene-disease associations, there exists a discrepancy between the gestalt impression of the group and the classifier that was developed using LDA. This has the potential to call into question the consistency within the group when using the SQM to score the various phenotypes.
- The discrepancy in the 11s lies within our hearing loss phenotypes, in which age of onset was taken into consideration when determining gestalt classifications → Shows that SQM is not capturing all of the factors guiding the Binning Committee

Linear Discriminant Analysis

- The high misclassification rate (even when applying the classifier to the training set itself) shows that a weighted combination of component scores of the SQM is not sufficient to categorize what should be on the core panel

General Conclusions / Future Directions

- Due to the heterogeneity of gene/condition pairs analyzed, these data make apparent the need for a Binning Committee, as well as a panel of experts on whose knowledge to rely
- Having a Binning Committee composed of diverse experts ensures that there is extensive discussion and documentation of the inclusion status of each gene/condition pair
- Since there does not seem to be a clear-cut actionability threshold, the group will continue to individually classify all gene/phenotype pairs that score 9s, 10s, and 11s on whether inclusion onto the core panel is appropriate
- There is consensus re-evaluation unconstrained by the SQM of what is and is not appropriate to put on a core panel, and that is what the study aims to learn → Do not know what effect it will have on families → Potentially changing professional norms in the genetics community

Acknowledgments

We would like to acknowledge funding from the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD) and The National Human Genome Research Institute (NHGRI) grant U19 HD077632.