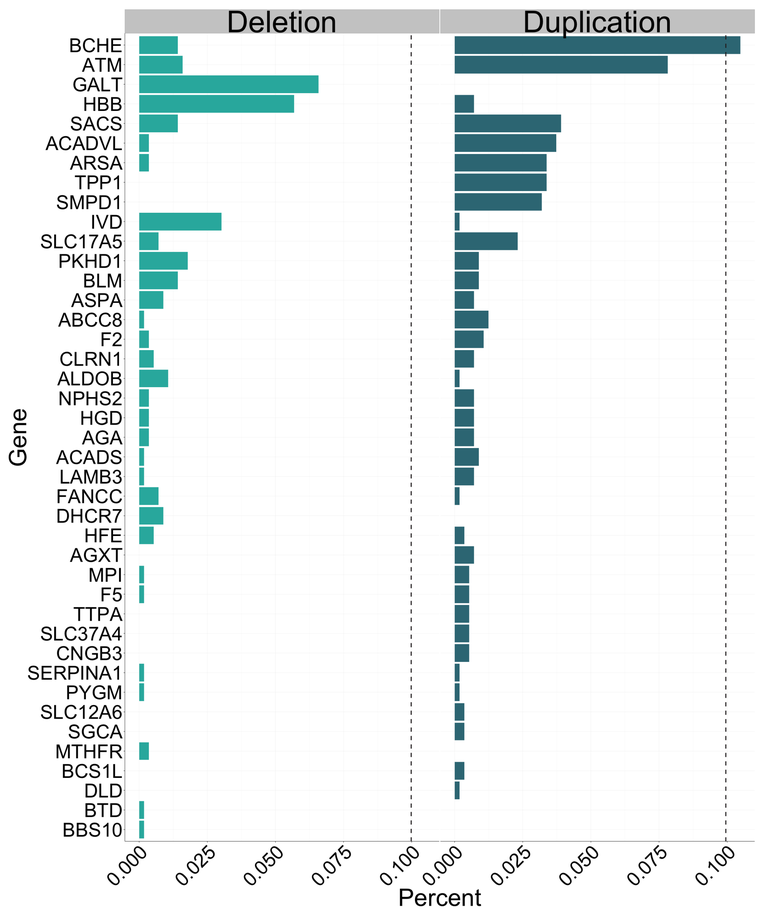
Using simulation for improved design of clinical next-generation sequencing panels

Genevieve Gould PhD, Xin Wang PhD, Peter Grauman, Greg Hogan PhD, Alex Robertson PhD, Jared Maguire, H. Peter Kang MD, Imran S. Haque PhD, Eric A. Evans PhD

South San Francisco, California

Introduction

Historically, variant genotyping tests have been validated by concordance with known reference materials. However, for NGS-panel tests detecting any variant in the genes of interest, this strategy is not effective as positive controls for variants that are difficult to detect may not be readily available. While large numbers of SNPs in genes of interest are available from reference materials, indels (especially large indels) and copy number variants (CNVs) are quite sparse or unavailable for most clinically relevant targets. Thus to comprehensively assess the performance of an entire region of interest, in silico methods must be used to supplement reference material comparisons. Here we present in silico simulations to assess the ability to detect copy number variants (CNVs).



Percent of samples containing a deletion or duplication in each gene based on ~56,000 Counsyl *Family Prep Screen* samples.

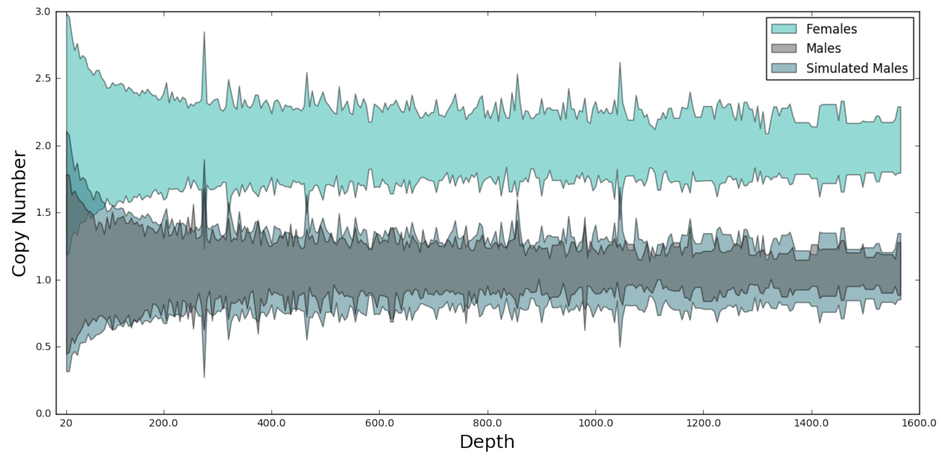
Fraction of nucleotides in the *Family Prep Screen* with at least one CNV observed in ~56,000 samples.

CNV rates estimated from simple depth-based caller.[1]

Methods and results

In-silico simulated CNVs recapitulate naturally occurring CNVs

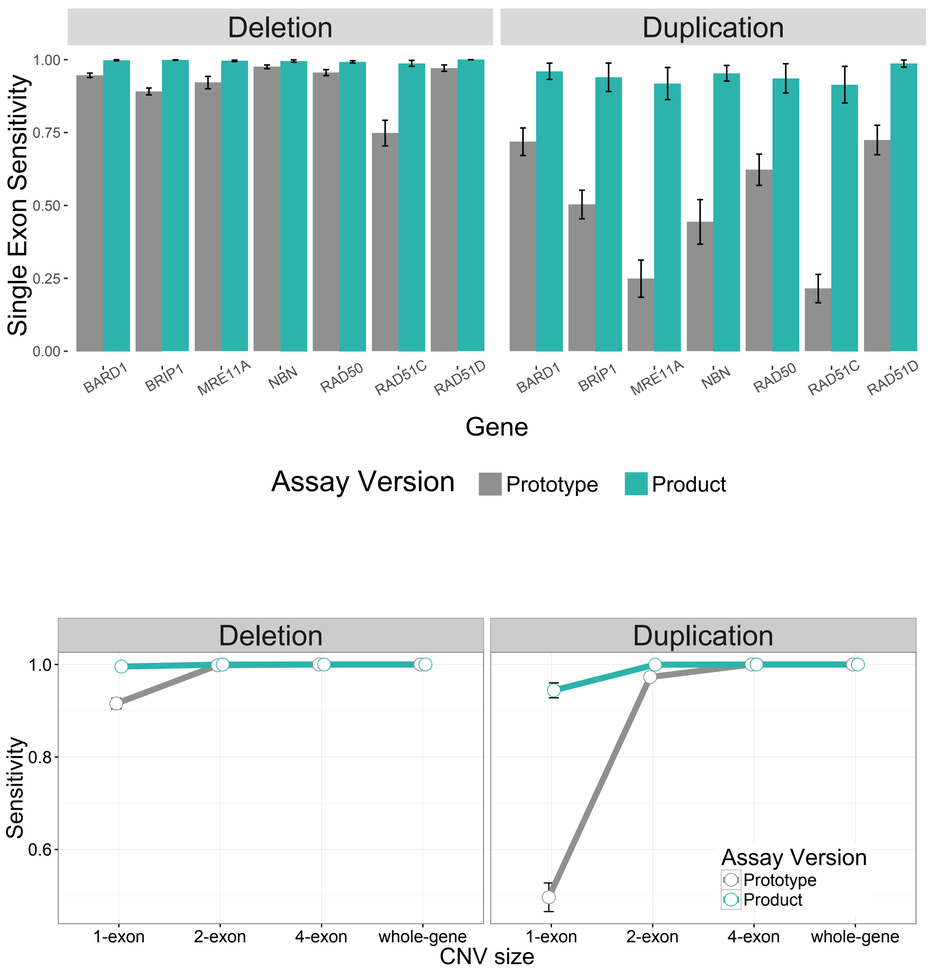
To more comprehensively assess our ability to call deletions and duplications compared to using reference samples alone, we generated synthetic CNVs ranging in length from single-exon to whole gene. A sample was randomly selected and the read depth was adjusted in the specified region to mimic a deletion or duplication while preserving the original variance in the measurement.



Read depth on the X chromosome can be used to compare simulated deletions to true deletions because XY individuals have half as many X chromosomes as XX individuals.

In-silico simulated CNVs inform assay development of *Inherited Cancer Screen*

The simulated CNV data were analyzed by the Counsyl bioinformatic pipeline to assess assay performance. CNV simulation led to a refinement of the initial assay design for the Counsyl *Inherited Cancer Screen* from the baseline design suggested by sequencing reference biological samples. This allowed us to maintain ≥99% CNV sensitivity while reducing our sample retest rate and improving our sensitivity in identified noisy regions.



CNV simulations were applied to the 36 genes in the Counsyl *Inherited Cancer Screen*. Summing the total number of TPs and FNs for the single exon, 5-exon, 10-exon, and whole gene CNV simulations from the ICS Validation study yields the following sensitivity and specificity:

Deletion sensitivity (by simulation):
99.60% [99.50%, 99.72%]

Duplication sensitivity (by simulation):
99.00% [98.82%, 99.19%]

Conclusion

In cases where informative biological samples are not always available, CNV simulations are valuable for validating and improving the quality of a clinical NGS pipeline in order to provide accurate results to patients and better inform medical management.

REFERENCES: 1. Beauchamp KA et al. bioRxiv 080713 (2016).