Categorizing variants after whole genome sequencing:

Implementation of “binning” -- a structured algorithm for the identification of clinically relevant incidental findings

Jonathan S. Berg, MD/PhD

Department of Genetics
Lineberger Comprehensive Cancer Center
Carolina Center for Genome Sciences
The University of North Carolina at Chapel Hill

Implementation of a next-generation clinical diagnostic test

- How do we find the clinically relevant variants?
  - PGx variants / common SNPs are easy, just a (somewhat difficult) question of which ones to use
  - Rare disease causing mutations are a challenge
- Three "sweeps" through the data:
  1. Diagnostic results (max. sensitivity)
  2. Incidental findings (max. specificity)
  3. Research (long-term analyses)

Context matters!

- Symptomatic patient
  - **Diagnostic assessment**
    - Needs to report full range of variants, including VUS (as with standard genetic testing)
- Asymptomatic patient
  - **Incidental assessment**
    - Prior probability of a genetic disorder approaches 0
    - Must maximize specificity so as to provide a clinically relevant posterior probability

Incidental analysis

- **Goal**: identify clinically relevant findings ***unrelated to the patient's presentation***
- **Premise**: the vast majority of genomic variants have no clinical relevance ***and must be ignored in a medical context***
  - Therefore imperative to maximize specificity and avoid reporting VUS
  - Set a "high bar" to ensure that variants reported to physicians/patients can be incorporated into clinical care in an evidence-based fashion

Genetics IN Medicine • Volume 13, Number 6, June 2011

COMMENTARY

Deploying whole genome sequencing in clinical practice and public health: Meeting the challenge one bin at a time

Jonathan S. Berg, MD, PhD[1], Muin J. Khoury, MD, PhD[2], and James P. Evans, MD, PhD[1]

- Current practices in medical genetics may not be suitable for genome-wide analysis
- Need for evidence-based, structured computational analysis for diagnostic and incidental results

Bins for incidental findings

- **<u>Strategy</u>**: Automated annotation of all variants
  1. *A priori* categorization of **genes** according to clinical utility and risk for harm
  2. *A priori* definition of the types of **variants** that should be reported

  – Sort variants into predetermined "bins," then review/report ***only those likely to be deleterious***
  – "Versioning" of analyses to allow one to know precisely what parameters were used

Bins for incidental findings

- Bin 1: Clinically actionable (Lynch, Long QT)
- Bin 2: Clinically valid but not directly actionable
  - 2a: low risk for harm (GWAS risk SNPs, PGx)
  - 2b: medium risk for harm (most Mendelian disorders)
  - 2c: high risk for harm (Huntington's, Presenilin)
- Bin 3: No known clinical significance
- Carrier status category*

Bins for incidental findings

- Informatics screening of OMIM genes
- 2016 genes “binned”
- Final bin decision was a judgment call

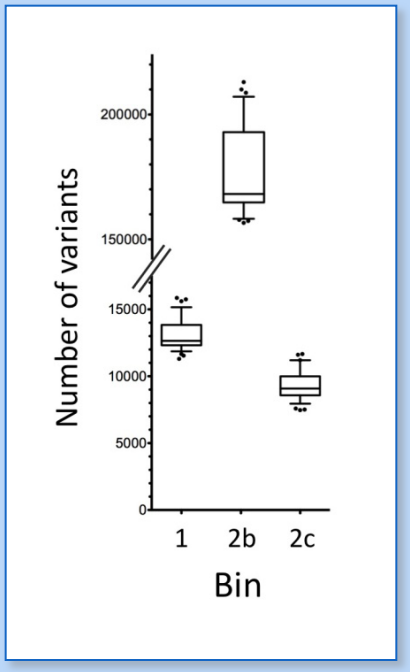
- Bin 1: 161 genes
- Bin 2:
  - 2b: 1798 genes
  - 2c: 57 genes

Theories are good, but data are better…

- 80 genomes sequenced by Complete Genomics
  - 19 patients with likely hereditary cancer susceptibility enrolled in a WGS study at UNC
  - 61 genomes made publically available by Complete Genomics
- 1000 Genomes Project allele frequency data
- Human Gene Mutation Database

Expectations

- The likelihood for any given person to have a disease-causing Mendelian mutation is LOW
  - Expect ***very few*** bin 1 or 2 findings per person
  - Most individuals will be carriers of heterozygous mutations in autosomal recessive genes



All variants in binned genes:

~13,000 variants in bin 1 genes

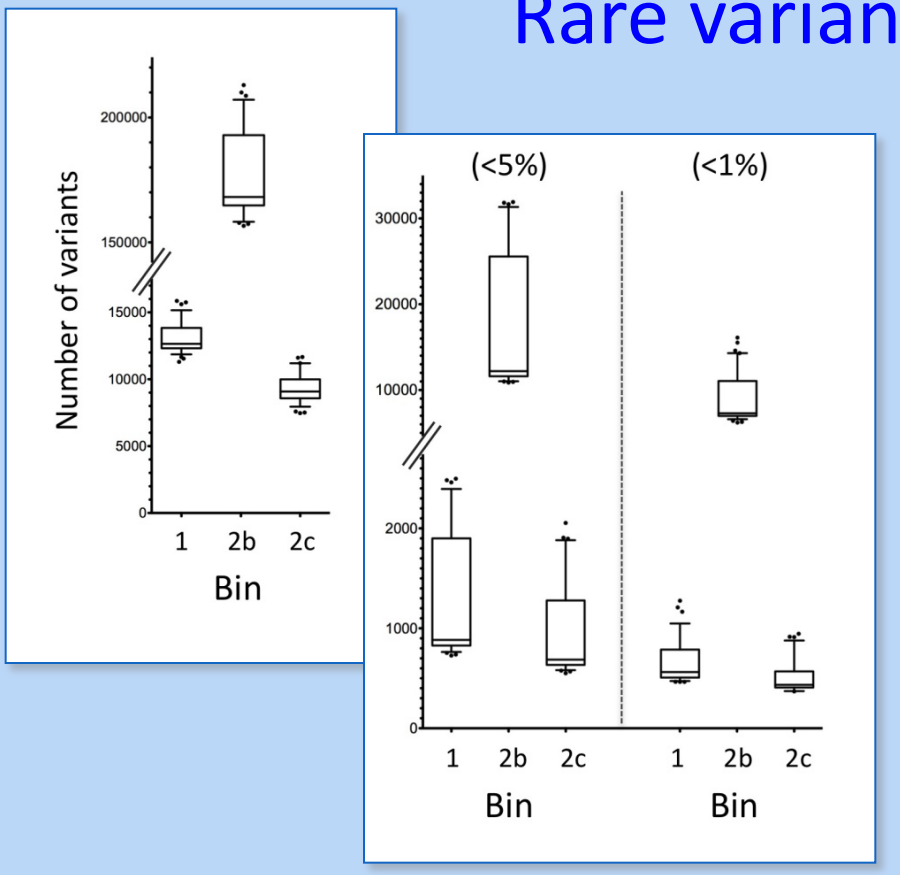
~175,000 variants in bin 2b genes

~9,200 variants in bin 2c genes

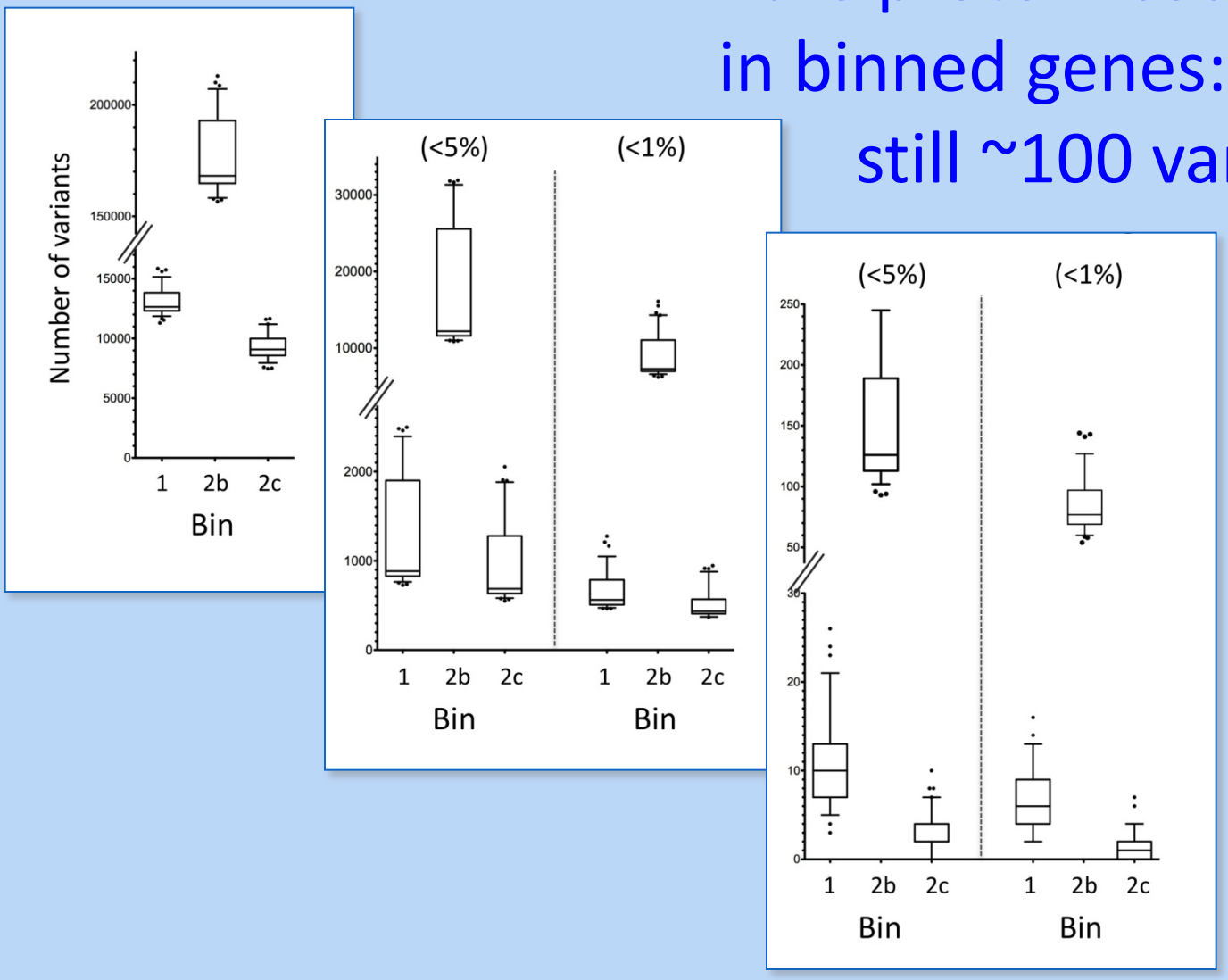
Rare variants in binned genes:

10-fold reduction (<5% AF)

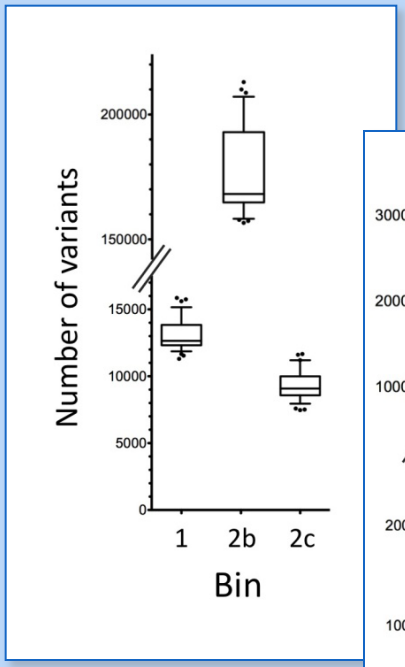
15-fold reduction (<1% AF)

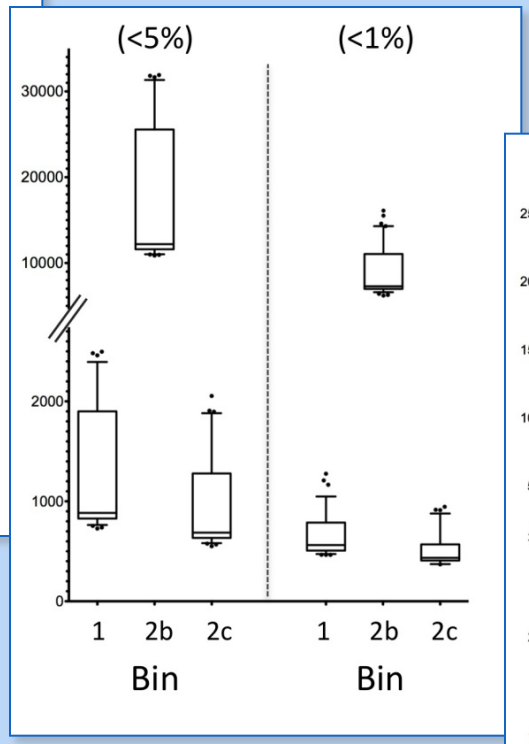


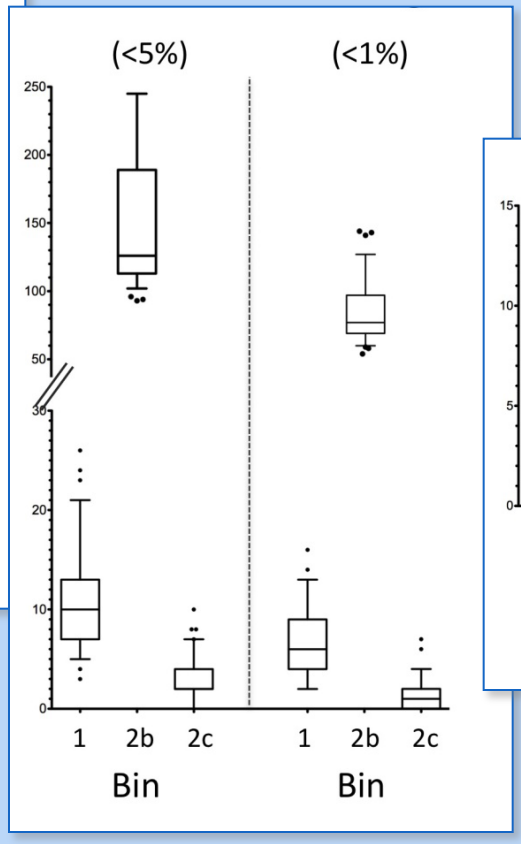
Rare protein-coding variants in binned genes:
still ~100 variants/person

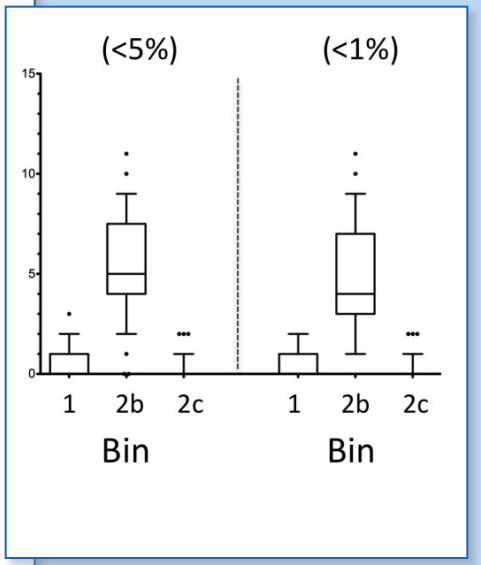


Rare truncating variants in binned genes: <10 variants/person





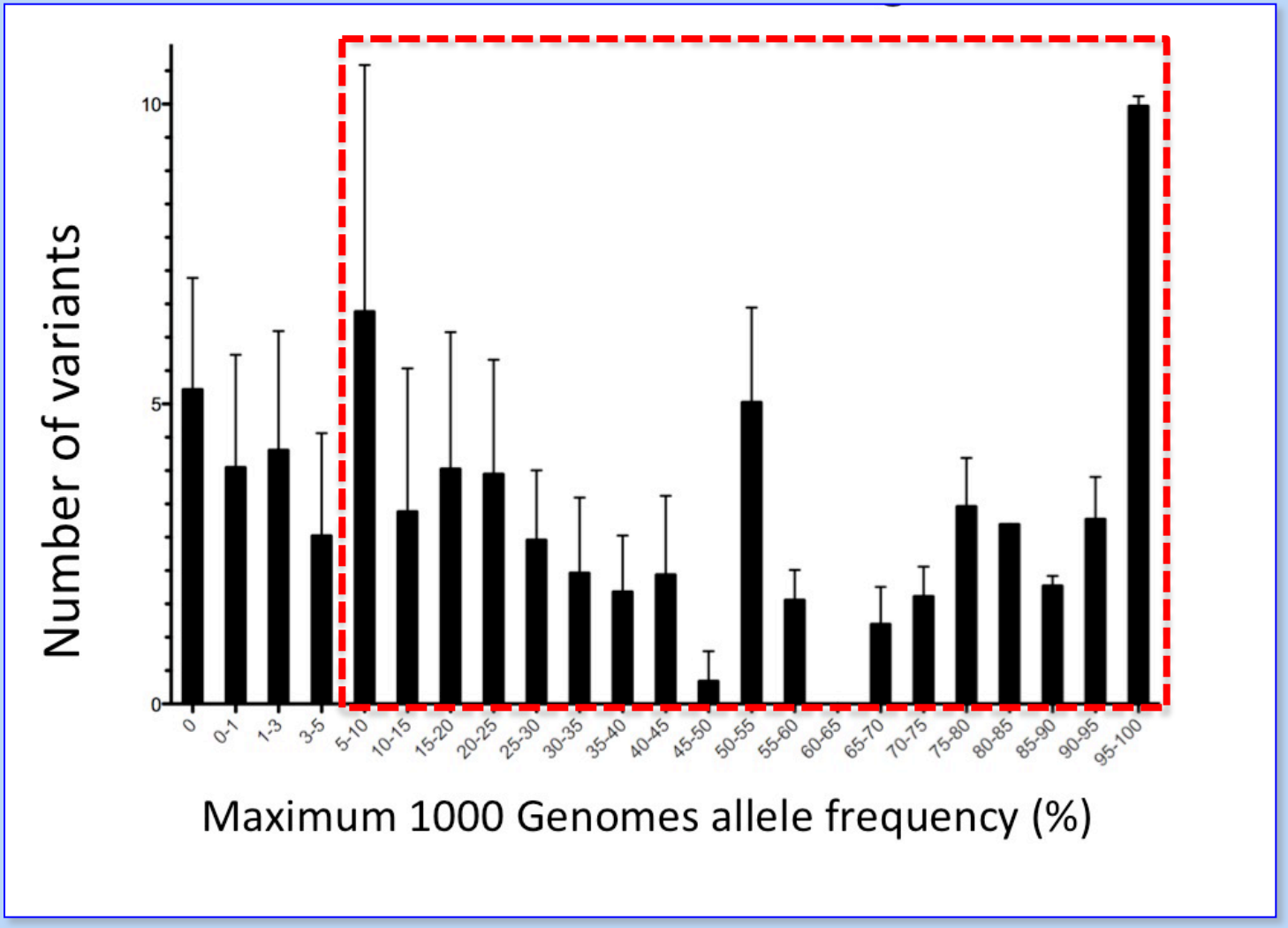




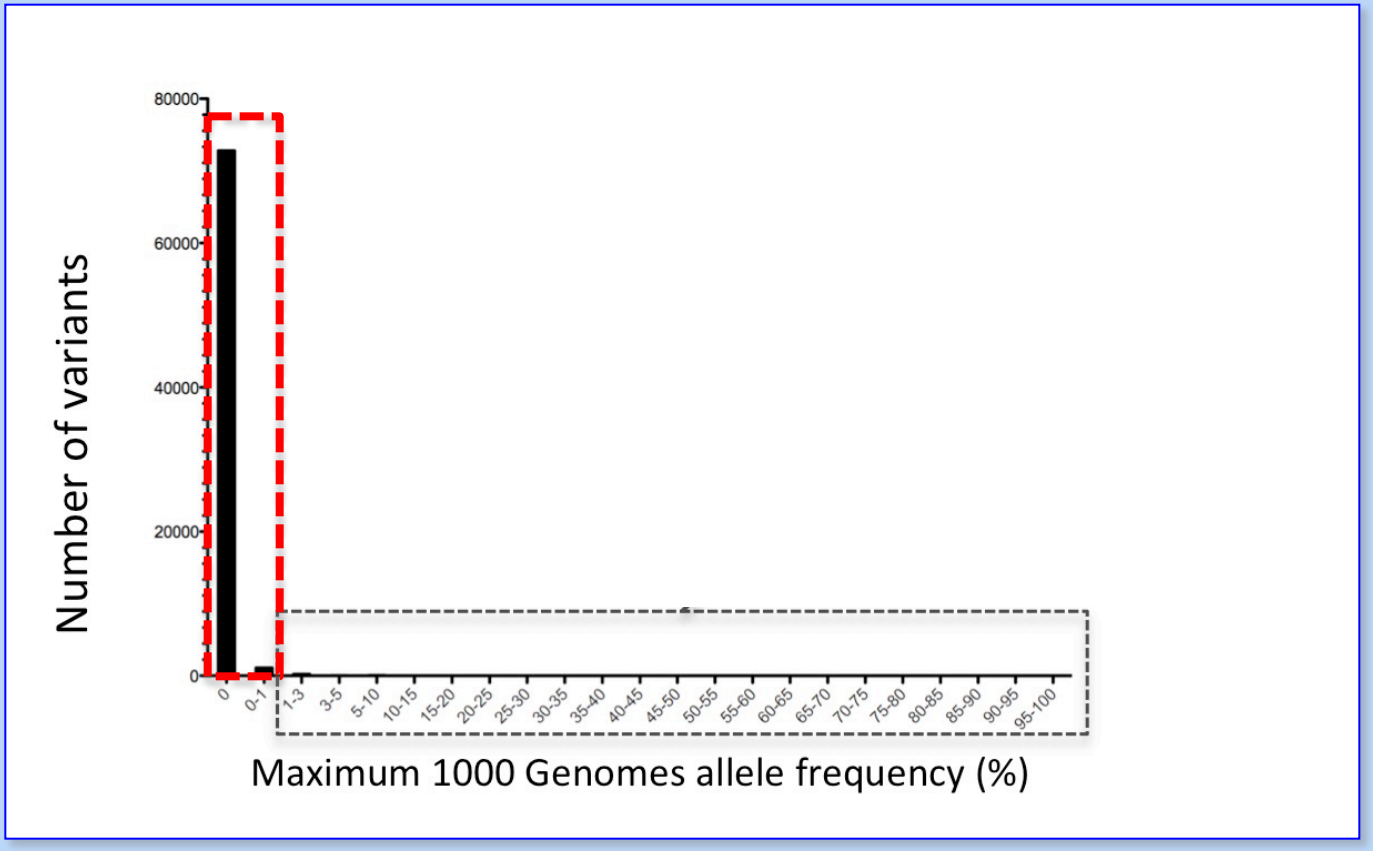
No missense variants?

- Limiting to truncating variants sacrifices sensitivity, excludes known disease causing missense mutations
- Possible solution: query the Human Gene Mutation Database for “DM” variants
  - Identified 871 unique variants, 771 missense
  - Average 74 (61-106) per person
  - Surprisingly little overlap with rare missense variants

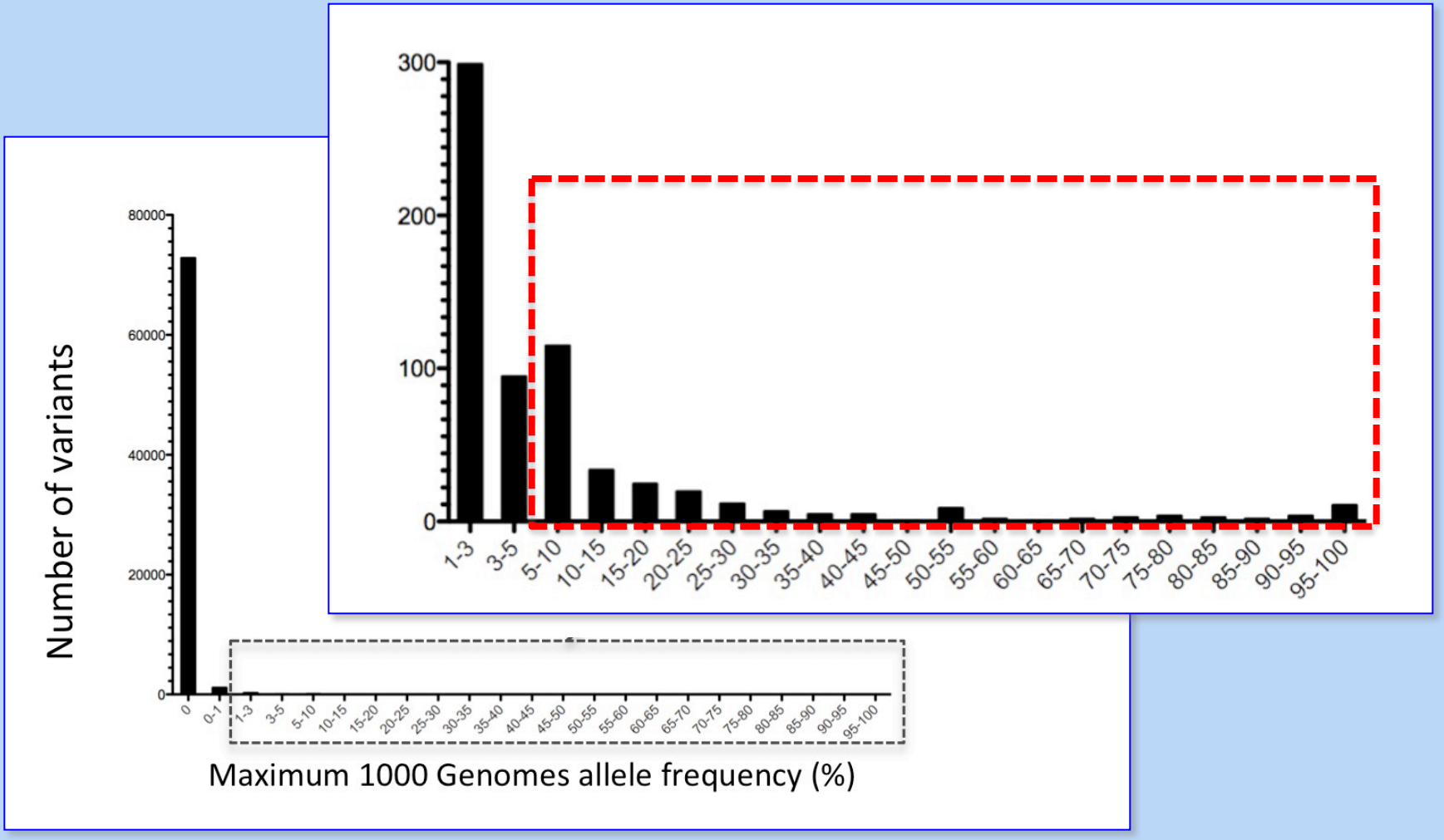
~80% of HGMD “DM” variants identified have >5% allele frequency



Most HGMD “DM” variants are rare…



…but too many are common



Final binning algorithm

- Variant annotated within “binned” gene

and

- <5% allele frequency

and

  - “DM” in HGMD

  OR

  - Protein truncating

Final binning algorithm
Variants per genome

	Bin 1	Bin 2b	Bin 2c	Carrier
Binned variants	1.5 (0-5)	6.4 (2-14)	0.2 (0-2)	9.2 (0-17)

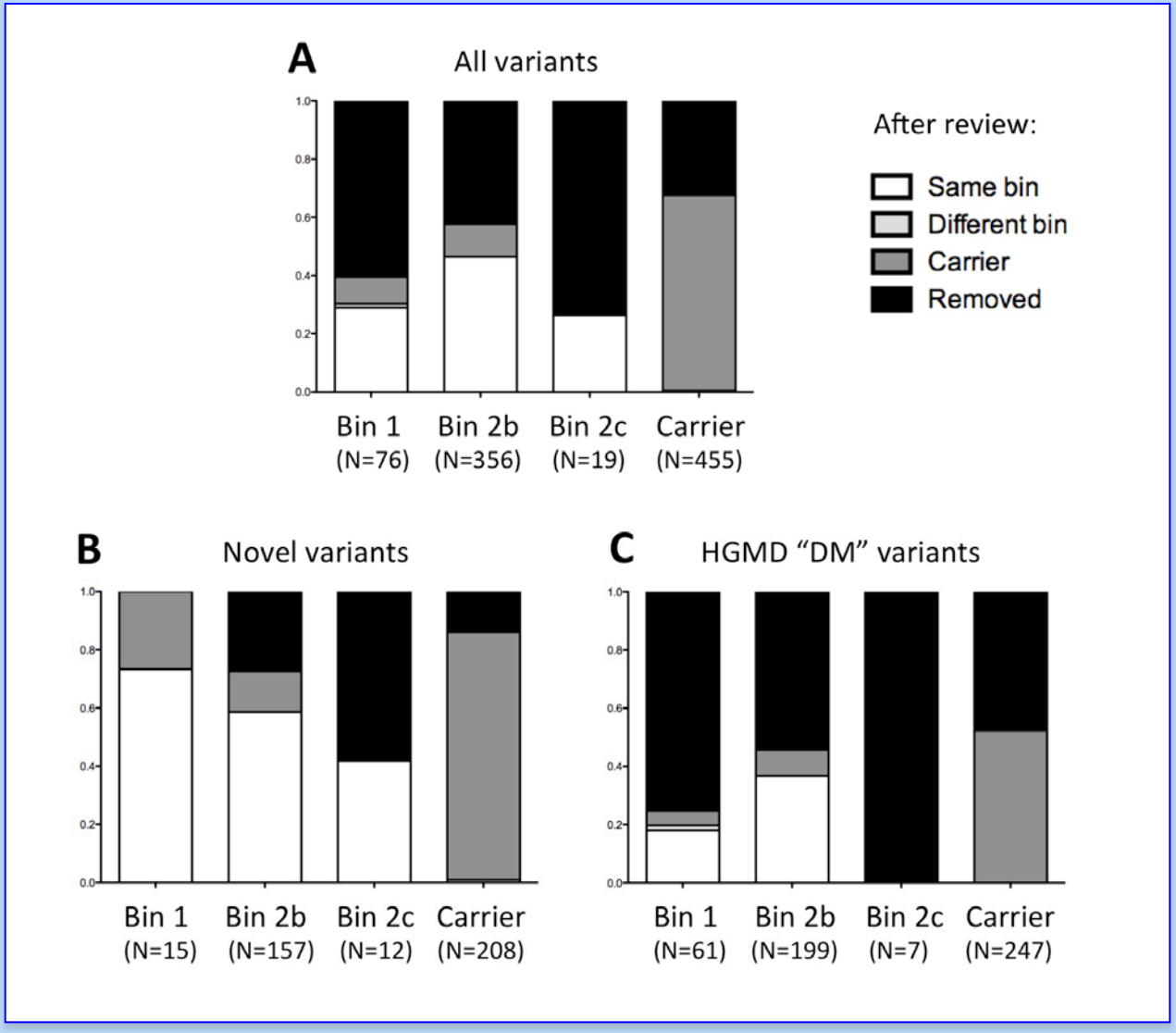
- A very tractable number for a human to review on a per person basis
- Close to expected numbers, but still too many
- Needed to perform manual curation of 1391 variants to remove or reassign

Final binning algorithm
Variants per genome

	Bin 1	Bin 2b	Bin 2c	Carrier
Binned variants	1.5 (0-5)	6.4 (2-14)	0.2 (0-2)	9.2 (0-17)
Upon review, ~50% were removed from consideration, 5% moved to carrier status				

- Used "Goldilocks" approach to reviewing variants (not too harsh, not too lenient)
  - Reviewed literature
  - Assessed type/location of variant
  - Used allele frequency information, especially in dominant disorders

Reclassification of variants after review



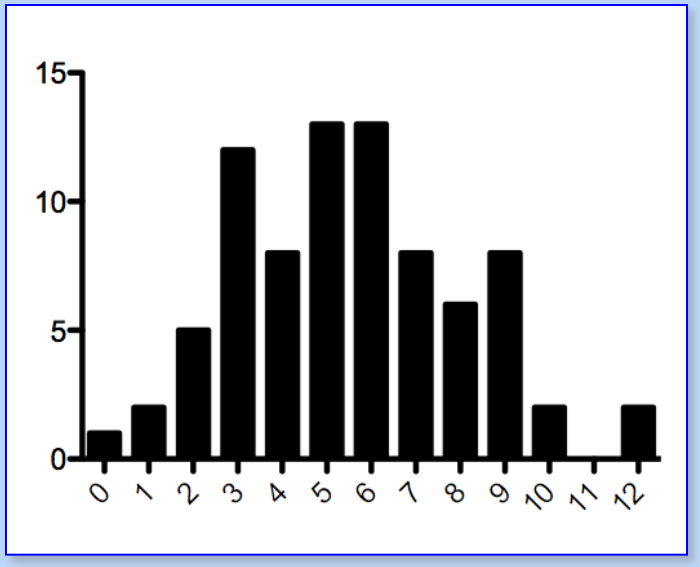
Final binning algorithm
Variants per genome

	Bin 1	Bin 2b	Bin 2c	Carrier
Binned variants	1.5 (0-5)	6.4 (2-14)	0.2 (0-2)	9.2 (0-17)
~50% removed from consideration, 5% moved to carrier status				
After review	0.3 (0-2)	2.6 (0-8)	0.06 (0-1)	5.5 (0-12)
~8.5 variants to confirm/report per sample				

- Still more variants than expected
- Sequencing artifacts? False positive reports in the literature? Incomplete penetrance?

Incidental carrier status findings

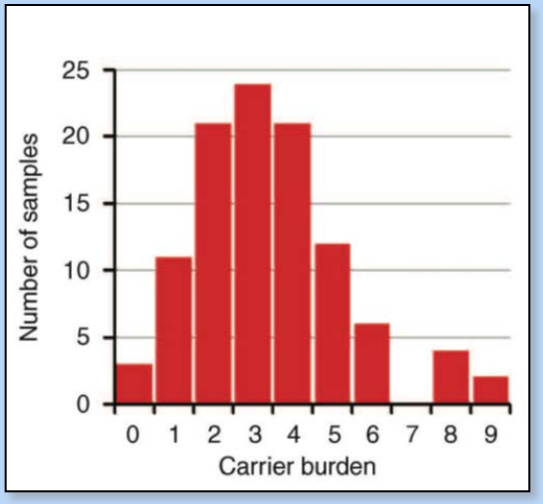
- 79/80 were “carriers” for at least one recessive condition
  - Range 0 – 12



RESEARCH ARTICLE
www.ScienceTranslationalMedicine.org 12 January 2011 Vol 3 Issue 65 65ra4
HUMAN GENOMICS

Carrier Testing for Severe Childhood Recessive Diseases by Next-Generation Sequencing

Callum J. Bell,[1]* Darrell L. Dinwiddie,[1,2]* Neil A. Miller,[1,2] Shannon L. Hateley,[1] Elena E. Ganusova,[1] Joann Mudge,[1] Ray J. Langley,[1] Lu Zhang,[3] Clarence C. Lee,[4] Faye D. Schilkey,[1] Vrunda Sheth,[4] Jimmy E. Woodward,[1] Heather E. Peckham,[4] Gary P. Schroth,[3] Ryan W. Kim,[1] Stephen F. Kingsmore[1,2]†



Summary

- A structured framework permits consistent analysis of WGS data, critical for clinical work
  - Every report will be linked to a “version” of the analytic scheme for reproducibility and future updating
- Clinical WGS analysis for incidental findings **can be** a tractable problem
  - High quality variant database critical
  - Predefined “rules” for automated annotation

Future Directions

- Refinement of bins
  - Expect debate over the genes in each category
  - Anticipate changes with advances in medical genetics
- Refinement of “rules” for reporting variants
  - More nuanced, gene- and disease-specific criteria
  - Development and utilization of clinical-grade databases
- Validation and deployment of risk prediction models with proven clinical utility

Future Directions

- Best practices for informed consent, return of results, integration with medical record
- ELSI issues related to clinical use of NGS
  - The field is in a state of equipoise regarding return of incidental findings
    - Should all results be divulged automatically?
    - How can we best enable patient preferences?
  - Need to study patient decision-making and outcomes from return of incidental findings
  - Incidental findings in infants/children?

UNC Clinical Genetics

Jim Evans, Kristy Lee, Kate Foreman,
Cecile Skrzynia, Catherine Fine, Ofri Leitner

Berg Lab

Michael Adams, Laura Milko
David DeWeese, Jonathan Mathew

RENCI

Kirk Wilhelmsen, Nassib Nassar,
Charles Schmitt, Chris Bizon

UNC Molecular Genetics Lab

Karen Weck, Jessica Booker, Kristy Crooks