Gold Standard Validation of GRID, a Gene Panel for Population Level PID Screening

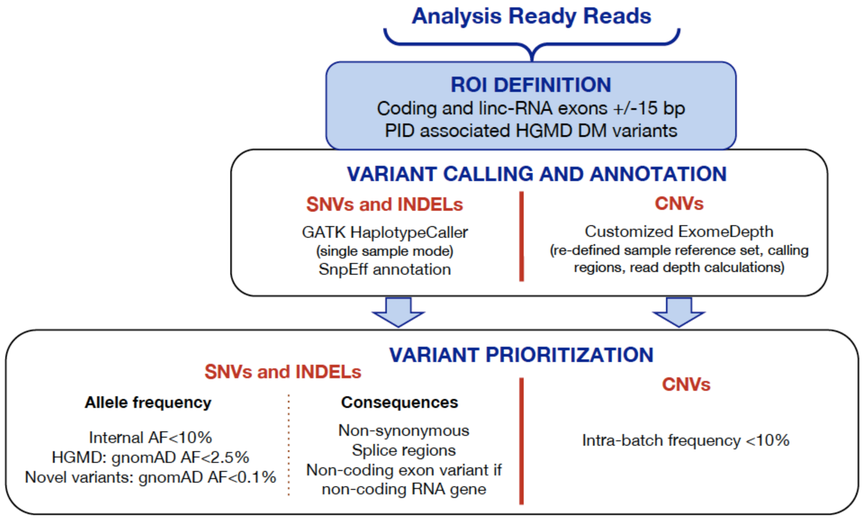
Hana Lango Allen[1,2], Olga Shamardina[1,2], Sri V.V. Deevi[1,2], Jonathan C. Stephens[1,2], Karyn Megy[1,2], Daniel Duarte[1,2], Deborah Whitehorn[1,2], Rutendo Mapeta[1,2], NIHR BioResource[2], Kenneth G.C. Smith[3], James E. Thaventhiran[4], Ilenia Simeoni[1,2]

1. Department of Haematology, University of Cambridge, UK
2. NIHR BioResource, Cambridge University Hospitals NHS Foundation Trust, UK
3. Department of Medicine, University of Cambridge, Cambridge, UK
4. MRC Toxicology Unit, School of Biological Sciences, University of Cambridge, UK

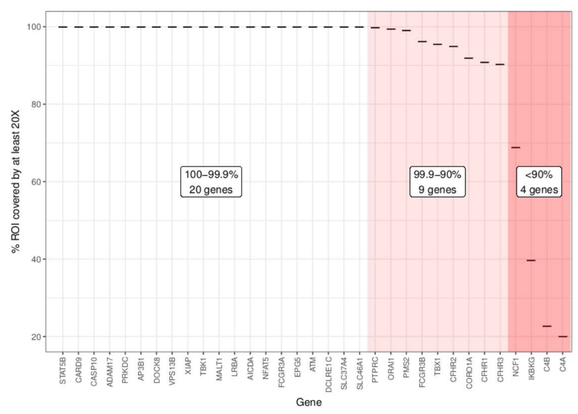
SUMMARY

- A conclusive molecular diagnosis in primary immune disorders (PID) is challenging, owing to the heterogeneity of clinical presentations
- GRID is comprehensive high-throughput gene panel that enables screening of all known genetic causes of PID in a single test
- We validated GRID to a diagnostic level using a gold standard set of >58,000 variants obtained by whole genome and Sanger sequencing of 176 patient samples
- We demonstrate high sensitivity (99.8%) and specificity (99.5%) for single nucleotide variants (SNVs) and insertion/deletion variants (INDELs)
- A bespoke automated copy number variant (CNV) calling pipeline was validated by comparison of 32 GRID CNV calls with WGS data
- We show that GRID is a reliable, first-line, unbiased genetic test for PID patients, with a potential to lead to: timely diagnosis -> early initiation of appropriate treatment -> better clinical outcomes -> and bring PID to the forefront of precision medicine

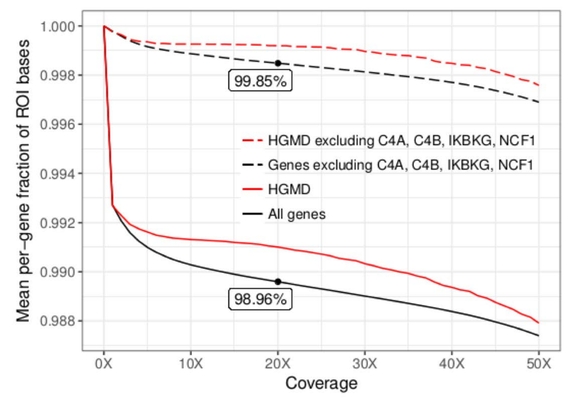
DATA PROCESSING



GENE AND VARIANT COVERAGE



Of the 279 genes on GRID, only 13 have <99.9% coverage at 20X for all ROI positions



Mean per-gene coverage of all ROI bases and all PID-causing HGMD variant positions

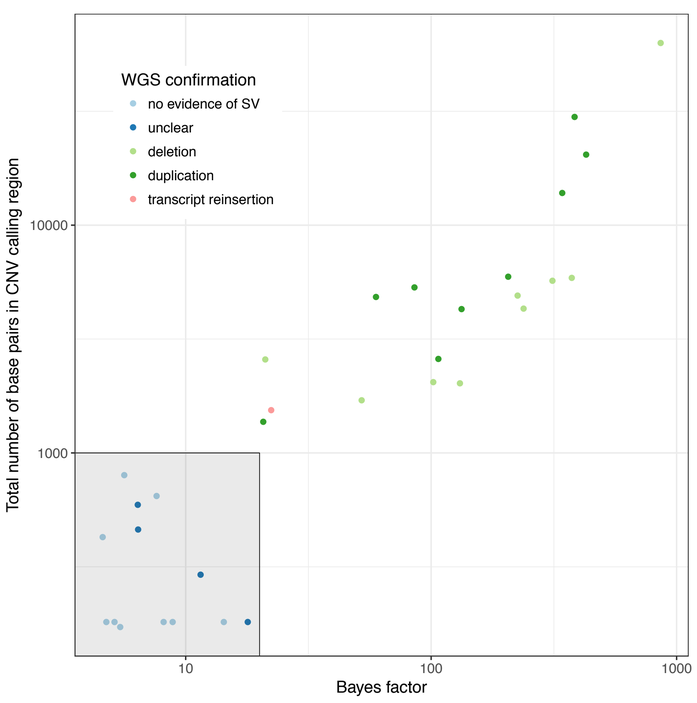
SNV and INDEL SENSITIVITY AND SPECIFICITY

- 176 PID patient samples were sequenced by both GRID and WGS, 53 with pathogenic variant
- We compared SNV and INDEL calls across ROI for 279 PID genes
- Discrepant calls were resolved by Sanger sequencing to create gold standard set of variants
- GRID specificity calculation was restricted to positions with WGS 20X coverage and pass call

SENSITIVITY	SNV (unique)	INDEL (unique)	TOTAL (unique)
Number of gold standard variants	52,731 (2,549)	1,178 (102)	53,909 (2,651)
Maximum GRID false negative calls	69 (7)	2 (2)	71 (9)
GRID Sensitivity [95%CI]	**98.87 [98.87-99.89]**	**99.83 [99.47-99.97]**	**99.87 [98.87–99.89]**

SPECIFICITY	SNV (unique)	INDEL (unique)	TOTAL (unique)
Number of GRID variants	56,819 (2,658)	1,315 (130)	58,134 (2,788)
Maximum GRID false positive calls	68 (22)	7 (5)	75 (27)
GRID Specificity [95%CI]	**99.88 [98.88-99.90]**	**99.47 [99.01-99.75]**	**99.87 [98.87-99.89]**

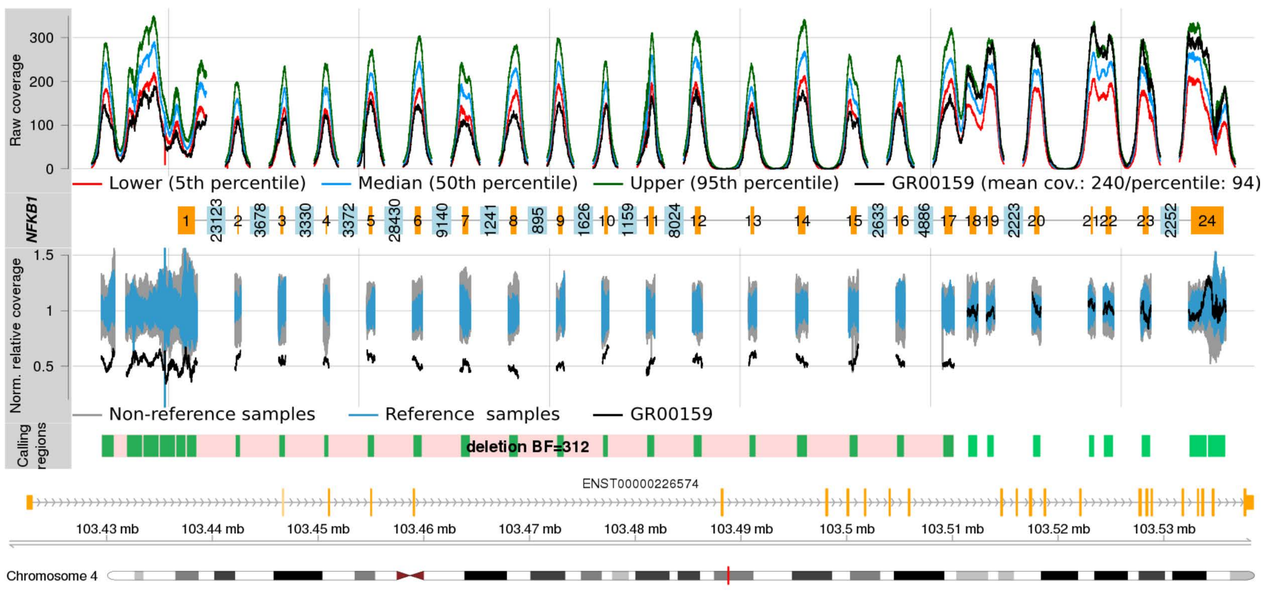
CNV VALIDATION



- GRID pipeline called 11 deletions and 21 duplications
- WGS data confirmed 9 deletions and 10 duplications
- All GRID false calls had low BF statistic, allowing for setting a threshold for high specificity
- 9 deletion calls included *ARPC1B, CFHR4, DOCK8, IGKC, LRBA, NFKB1* and *TBX1* single gene deletions, and one multi-gene call
- 10 duplication calls spanned *CD8A, DOCK8, IKBKB, LYST, MASP2, MCM4, PRKDC, STAT2,* and *UNG* genes

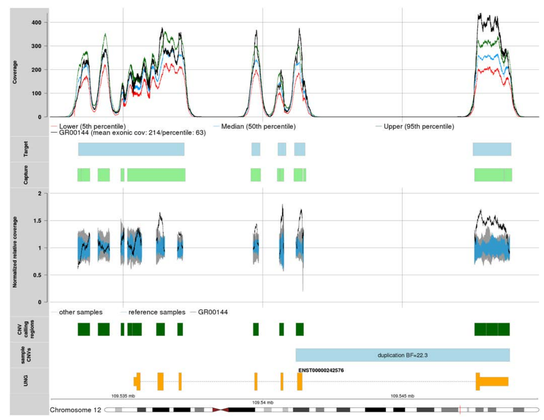
AUTOMATED CNV CALLING

- Automated CNV call and plot example of *NFKB1* heterozygous deletion call (pink block)
- Single gene call is made by combining coverage-based individual exon calls (green bars)
- Normalised relative coverage visually confirms ~50% drop in coverage in this sample (black lines) compared to reference samples from the same batch (blue lines)

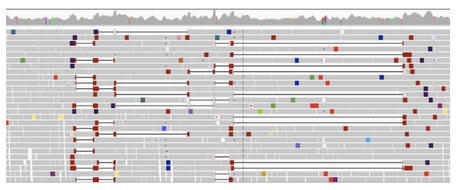


Transcript re-insertion

- Based on coverage, one of the GRID CNV calls was *UNG* gene duplication
- Based on read mapping, WGS data suggested deletions of individual introns, and increased coverage over *UNG* exons
- This was resolved as UNG transcript re-insertion into an unknown genomic location
- This event is not known to be pathogenic



Automated GRID partial gene duplication call



WGS reads suggest multiple intron deletions