Identification of rare-disease genes from RNA-seq of diverse undiagnosed cases using large control cohorts

Laure Frésard[1], C. Smail[2], K.S. Smith[1], N.M. Ferraro[2], N.A. Teran[3], K. Kernohan[4], D. Bonner[5], X. Li[1], S. Marwaha[5], Z. Zappala[1,3], B. Balliu[1], J.R. Davis[1,2], B. Liu[6], C.J. Prybol[3], J.N. Kohler[5], D.B. Zastrow[5], D.G. Fisk[7], M.E. Grove[7], J.M. Davidson[5], T. Hartley[4], R. Joshi[7], B.J. Strober[8], S. Utiramerur[6], Care4Rare Canada Consortium[4], Undiagnosed Disease Network[9], L. Lind[10], E. Ingelsson[11,12], A. Battle[8,13], G. Bejerano[14], J.A. Bernstein[11,15], E.A. Ashley[3,11], K. Boycott[4], J.D. Merker[1,7,16], M.T. Wheeler[5,11], S.B. Montgomery[1,3]

1 Department of Pathology, School of Medicine, Stanford, CA, USA; 2 Biomedical Informatics Program, Stanford, CA, USA; 3 Department of Genetics, School of Medicine, Stanford, CA, USA; 4 Children's Hospital of Eastern Ontario, Ontario, Canada; 5 Stanford Center for Undiagnosed Diseases, Stanford, CA, USA; 6 Department of Biology, Stanford, CA, USA; 7 Stanford Medicine Clinical Genomics Program, Stanford, CA, USA; 8 Department of Biomedical Engineering, John Hopkins University, MD, USA; 9 NIH Undiagnosed Diseases Network, Bethesda, MD, USA; 10 Department of Medical Sciences, Uppsala University , Sweden; 11 Department of Medicine, School of Medicine, Stanford, CA, USA; 12 Stanford Cardiovascular Institute Stanford, CA, USA; 13 Department of Computer Science, John Hopkins University, Baltimore, MD, USA; 14 Department of Computer Science, Stanford, CA, USA; 15 Department of Pediatrics, School of Medicine, Stanford, CA, USA; 16 Department of Pathology, Lineberger Comprehensive Cancer Canter. University of North Carolina, Chapel Hill, NC, USA

Background

Investigating rare diseases is of great importance

For science
- Reveals biochemical pathways
- Helps genome annotation and interpretation
- Better understanding of Mendelian diseases

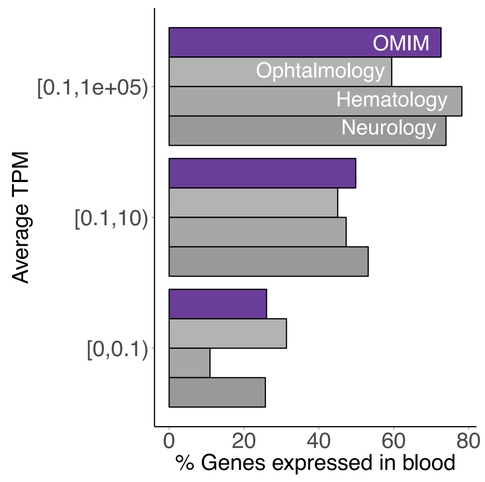
For families
- Identification of the causal gene
- Family planning
- Drug development
- Treatment/disease management

RNA-based applications show considerable promise
- Exome sequencing alone is insufficient in 68% of rare disease diagnosis[1]
- RNA-seq has shown diagnostic utility in specific tissues and diseases[2,3]
- We can observe the consequences of both coding and non-coding variants in gene expression and splicing events

The generalizability of this approach to diverse Mendelian diseases has yet to be evaluated

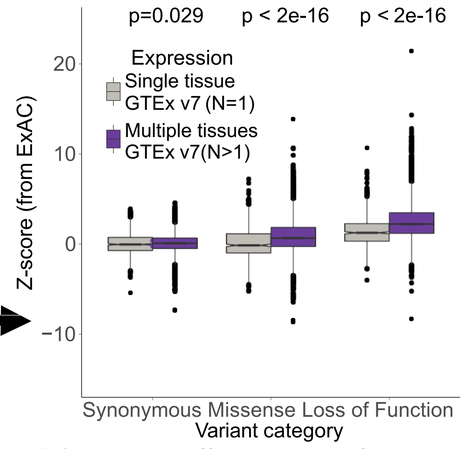
➔ To assess RNA-seq from blood as general diagnostic tool, we sought to evaluate it for rare diseases of different pathophysiologies.

Blood transcriptome as a surrogate for pathological tissue



➔ The disease tissue is the best tissue to analyze
➔ ~70% of OMIM protein-coding genes are expressed in blood

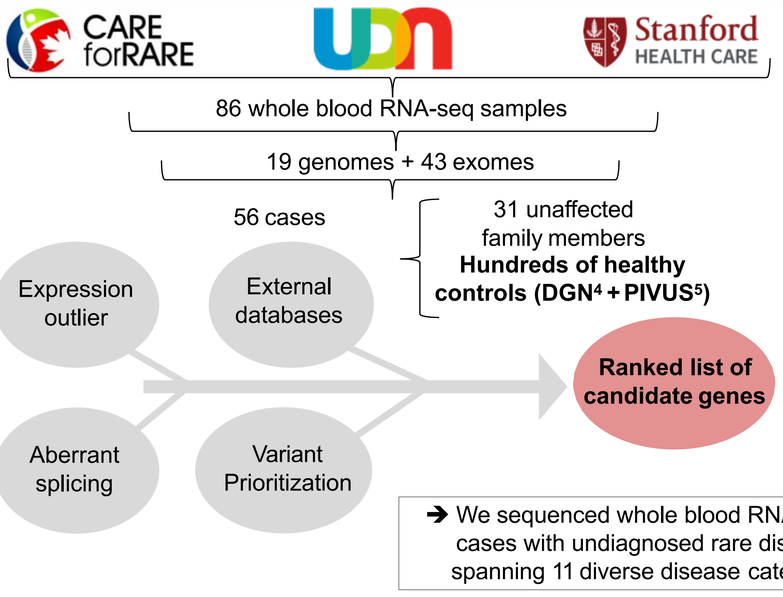
➔ Genes that are expressed in multiple tissues tend to be more sensitive to missense and loss-of-function mutations



Expression status for genes of interest in blood

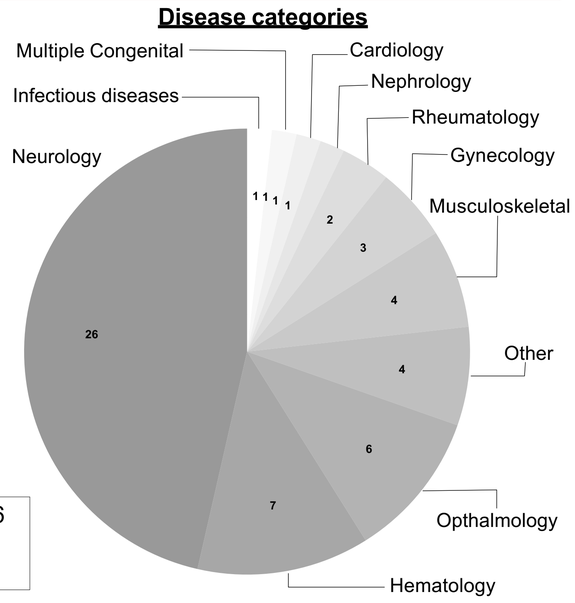
Tolerance to different types of mutations (ExaC) in function of the expression status in a single tissue vs multiple tissues.

Material and Methods



➔ We sequenced whole blood RNA from 56 cases with undiagnosed rare diseases spanning 11 diverse disease categories

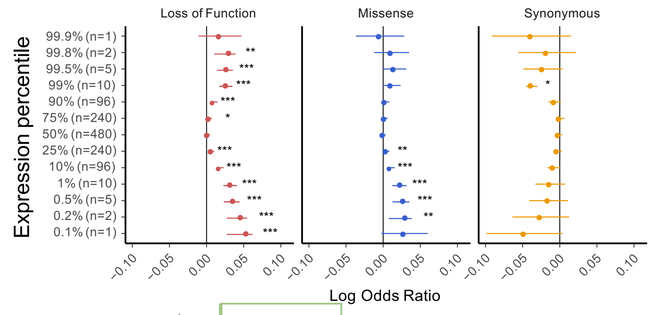
Disease categories



Outlier patterns of cases

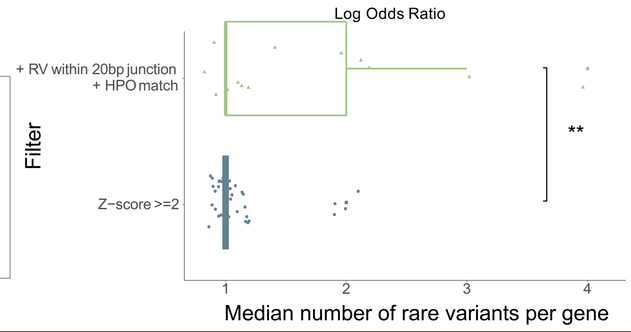
Expression outlier

➔ Case outliers are enriched for genes that are sensitive to Loss of Function variation



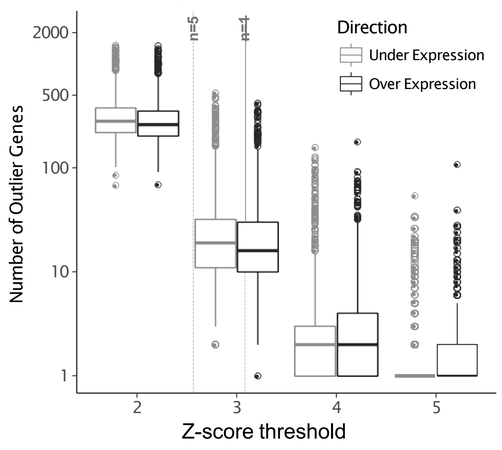
Splicing outlier

➔ Significant increase in number deleterious rare variants in the gene when filtering for outliers with a rare variant in a gene relevant to the disease phenotype

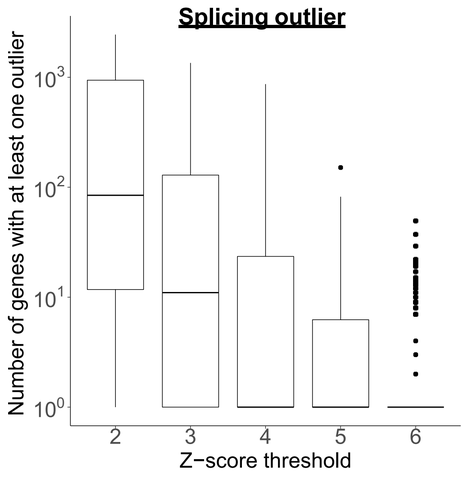


Aberrant transcripts can be detected as outliers

Expression outlier



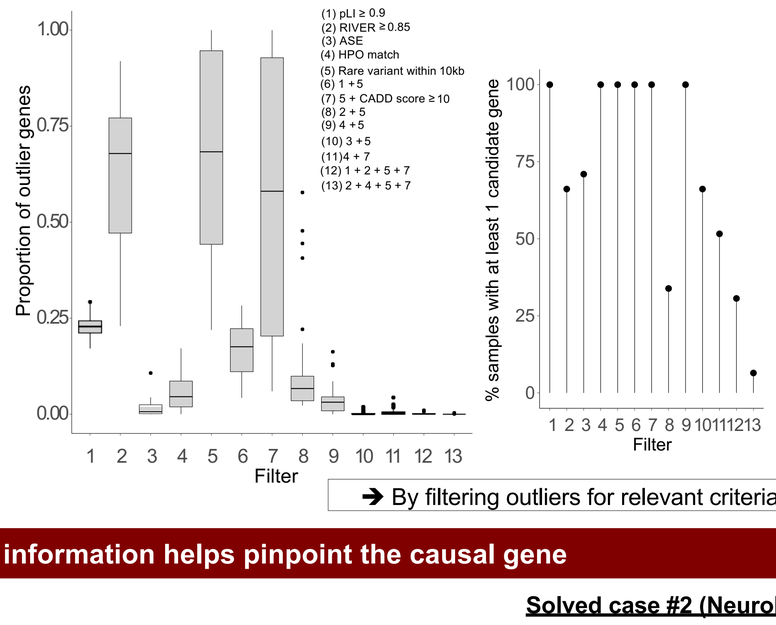
Splicing outlier



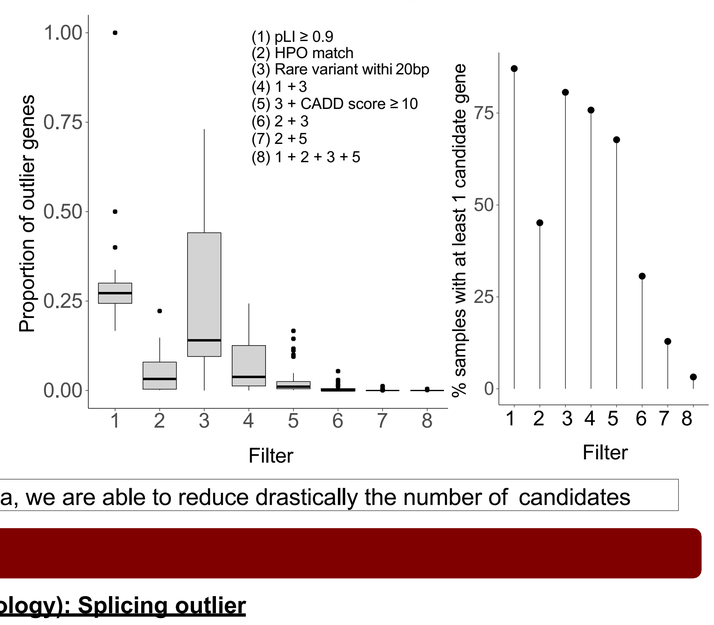
➔ On average we find a few hundreds candidate genes to a handful depending on stringency of the Z-score threshold

Filtering outliers reduces the number of candidate genes

Expression outlier



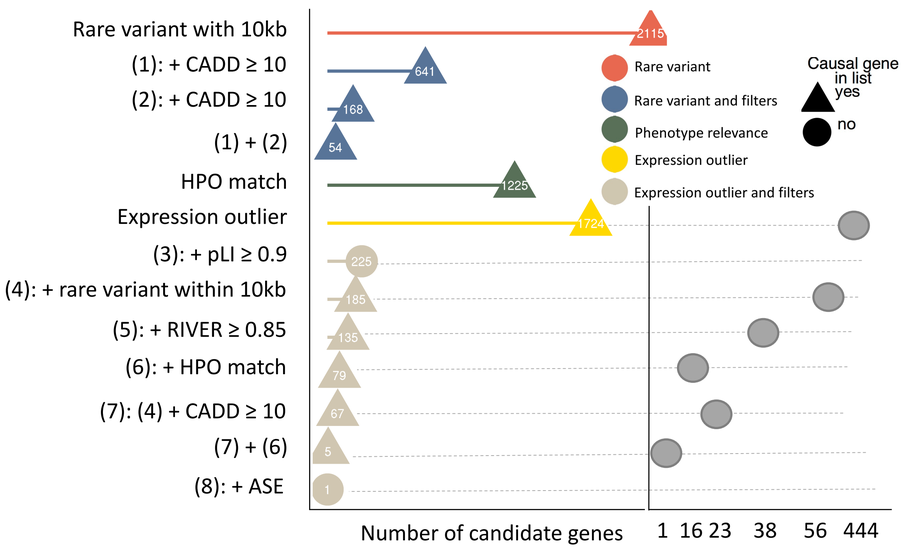
Splicing outlier



➔ By filtering outliers for relevant criteria, we are able to reduce drastically the number of candidates

Combining information helps pinpoint the causal gene

Solved case #1 (Neurology): Expression outlier

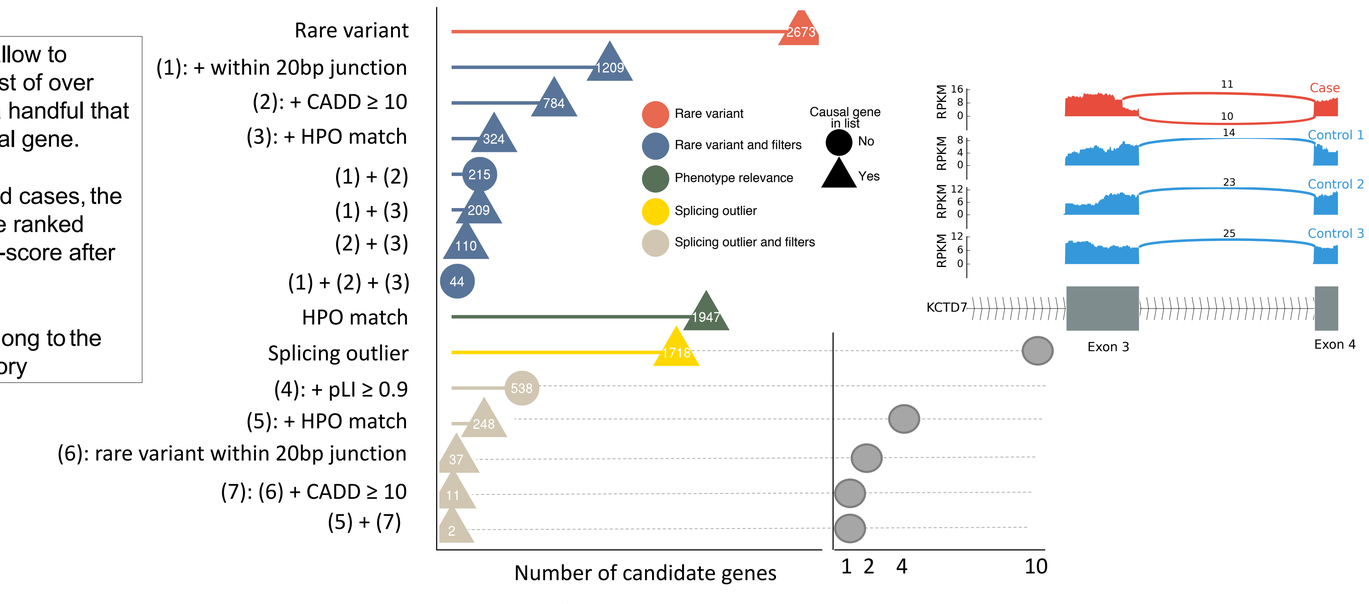


➔ Relevant filters allow to narrow down a list of over 2,000 genes to a handful that contain the causal gene.

➔ In both presented cases, the causal genes are ranked **first** in term of Z-score after filters

➔ Those cases belong to the neurology category

Solved case #2 (Neurology): Splicing outlier



References

1 Zappala and Montgomery, *Hum. Hered.* 2016
2 Cummings et al., *Sci. Transl. Med.* 2017
3 Kremer et al., *Nat. Commun. 2017*
4 Battle et al, *Genome Res.* 2014
5 Lind et al, *Arterioscler. Thromb. Vasc.* Biol. 2005
6 Kernohan et al., *Hum. Mutat* 2017
7 Merker et al., *Genet. Med* 2017

Acknowledgments

- Care4Rare
- the Undiagnosed Disease Network
- the Stanford Clinical Genomics Service
- Montgomery lab

Conclusions

- **Blood** can be an effective surrogate for the investigation of rare diseases by RNA-seq
- Across our cohort, we observed that RNA-seq yields **a 8.5% diagnostic rate**
- RNA-seq can be used to **confirm the impacts** of known mutations **and discover new mutations** affecting expression[6,7]