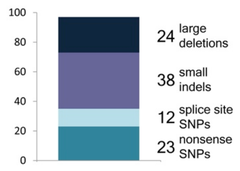
Functional Interpretation of Loss-of-Function Mutations in DNA Sequencing Studies

M.A. Rivas[1], T. Lappalainen[2], M. Sammeth[3,11], D.G. MacArthur[4], M. Lek[4], M. Pirinen[1], T. Wieland[5], A. Kirby[4], N. Kurbatova[6], T. Griebel[3], M. Barann[7], M. Sultan[8], J. Carlsson[9], P.A.C. 't Hoen[10], T. Strom[6], S. Schrieber[7], H. Lehrach[8], S. Antonarakis[2], G. van Ommen[10], R. Sudbrak[8], R. Hasler[7], A. Brazma[5], A. Syvanen[9], P. Rosenstiel[7], T. Meitinger[6], R. Guigo[11], I. Gut[3], X. Estivilli[11], P. Donnelly[1], M.I. McCarthy[1], E.T. Dermitzakis[2], The Geuvadis Consortium

[1]Nuffield Department of Clinical Medicine, University of Oxford; [2]Department of Genetic Medicine and Development, University of Geneva; [3]Centro Nacional de Analisis Genomico, Barcelona; [4]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston; [5]Institute of Human Genetics, Helmholtz Zentrum München, Munich; [6]European Bioinformatics Institute, Hinxton; [7]Institute of Clinical Molecular Biology, University of Kiel; [8]Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Berlin; [9]Department of Medical Sciences, Uppsala University; [10]Center for Human and Clinical Genetics, Leiden University Medical Center; [11]Center for Genomic Regulation and UPF, Barcelona

Loss-of-Function Mutations in DNA Sequencing Studies

- Protein truncating variants (frameshift, stop gain, splice site) are commonly referred to as Loss-of-Function (LoF).
- A typical genome has ~100 annotated germine LoF variants.



- Annotated LoF variants may have a strong impact on disease outcome[3,4,5].

Aims

- Using RNASeq data to develop more precise description of functional effect for loss-of-function variants, e.g with respect to Nonsense Mediated Decay (NMD) and splicing.

mRNA Sequencing of 465 Samples from the 1000 Genomes Project

- RNA Sequencing done in seven institutes with Illumina TruSeq Protocol.
- Lymphoblastoid cell lines (LCL) from 465 samples.

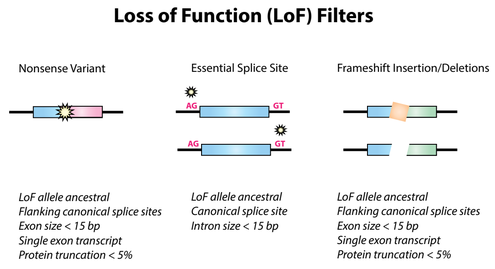
	1000G Phase1 + Phase2	RNAseq
TSI	92 + 1	93
GBR	86 + 10	96
FIN	89 + 6	95
CEU	79 + 13	92
YRI	77 + 12	89
TOTAL	423 + 42	465

- Genotypes from 1000 Genomes 27M total variants. 90% of samples in Phase 1, remaining imputed from Omni 2.5
- 660 informative stop gain variants; ~5000 informative variants near splice junctions.

See Poster 567F for more information about the Geuvadis RNA Sequencing Project.

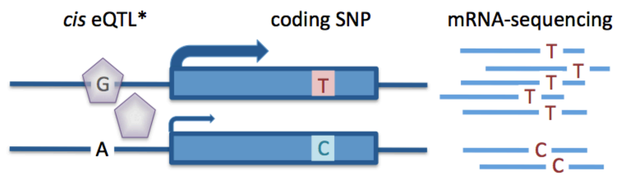
Methods

Annotation



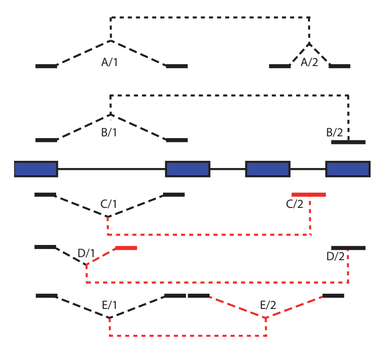
- DNA Sequence variants annotated with ALOFT and PLINK/SEQ[6] using Gencode version 12 transcripts.

Allele Specific Expression (ASE)



- Read counts over all heterozygous sites of an individual partitioned by the allele.
- Binomial test to detect deviation from the expected ~50/50 ratio.

Splice Junction Quantifications

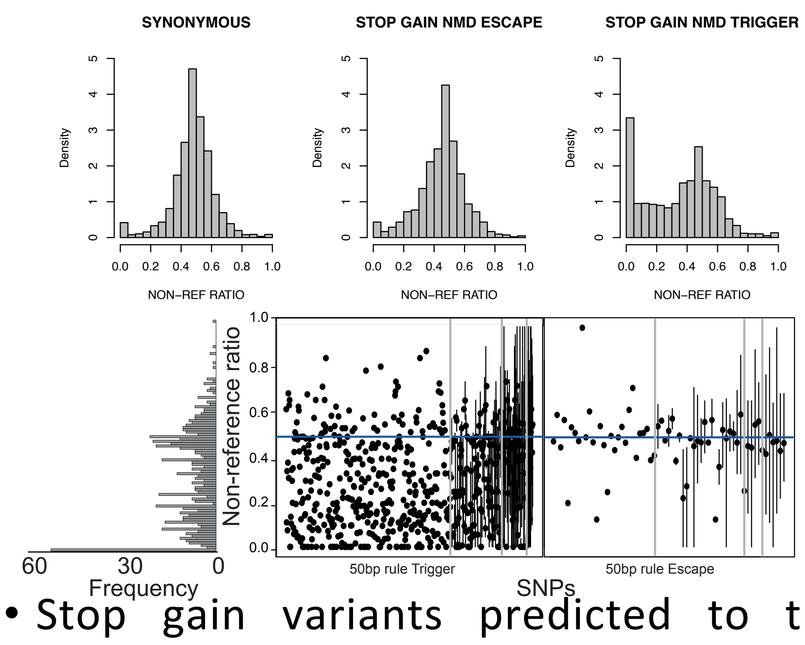


Two examples for valid, annotation-paired split-mappings (A and B) as well as three split-mappings that are incompatible with the annotation (C, D, E),

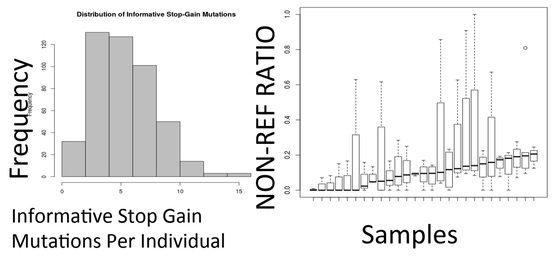
- Splice junction quantifications using annotation-compatible split-mappings.

Results

Stop Gain variants and ASE Signal

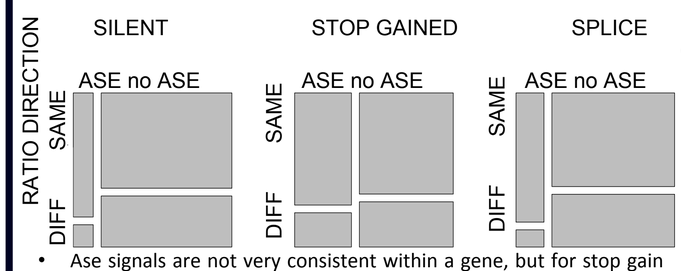


- Stop gain variants predicted to trigger nonsense-mediated decay (NMD) exhibit different ASE distribution from Synonymous variants and STOP gain variants predicted to escape NMD (P < 2.2e-16). Better predictions needed for NMD.



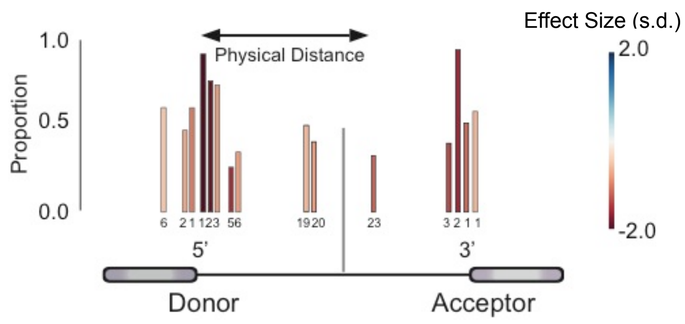
Informative Stop Gain Mutations Per Individual

Substantial variance in ASE signal between individuals.

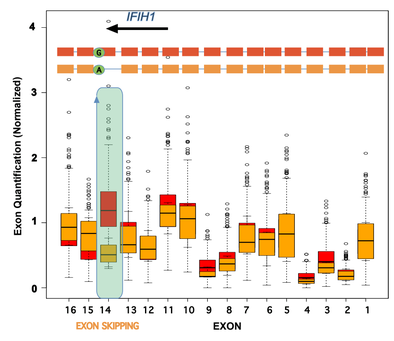


- ASE signal for stop gain variants tend to affect the entire gene.
- Ase signals are not very consistent within a gene, but for stop gain variants the consistency is higher – i.e. they tend to affect the entire gene, more often than regulatory variation that only affect a subsection of the gene.

Variants near splice junctions



MCMC algorithm applied to splice junction quantifications highlight shared impact of rare variants near splice junctions.



- Rare protective splice variant c.IVS14+1G>A *at IFIH1* disrupts splicing of exon 14 and leads to exon skipping while escaping nonsense mediated decay.

Functional interpretation of LoF variants associated to disease.

References

[1]MacArthur, D.G. *et* al. A Systematic Survey of Loss-of-Function Variants in Human Protein-Coding Genes. *Science* 2012.
[2]The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012.
[3]Nejentsev. S. *et* al. Rare variants of *IFIH1,* a Gene Implicated in Antiviral Responses, Protect Against Type I Diabetes. *Science* 2009.
[4]Rivas, M.A. *et* al. Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nature Genetics* 2011.
[5]Neale, B.M. *et* al., Sanders et al., O'Roak et al., 2012. *Nature* 2012.
[6]http://atgu.mgh.harvard.edu/plinkseq/; PLINK/SEQ v0.09. ALOFT MacArthur, D.G. et al.