Ranking Dataset Mutations

Using Combined Annotation Scoring Matrices and Gene Network Approach

Mutation Zero

DNAnexus

Rationale

Massive and diverse knowledge of genomics

Ranking mutations to discover treatment target

Human known variation
1000 genomes
dbSNPs

Conservation metrics
GERP
PhastCons
PhyloP

Gene interaction network
BioGrid

Functional genomic
DNase hypersensitivity
Transcription factor binding
Distance to exon-intron boundaries
Expression levels in common cell lines
Synonymous non-synonymous mutation

Protein-level scores
Grantham
SIFT
PolyPhen

Scoring variants using Combined Annotation Dependent Depletion (CADD) (Kircher M et al, 2014)

- Integrate annotation tracks from Ensembl Variant Effect Predictor (VEP), ENCODE project, and UCSC genome browser to one matrix
- Score SNPs and indels for all bases

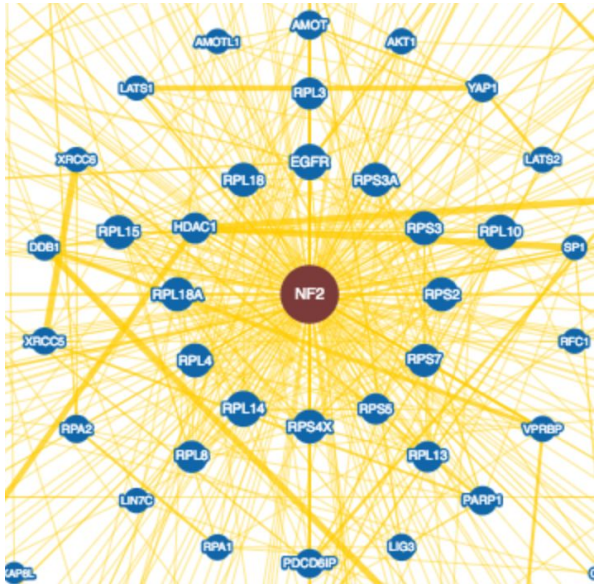
Somatic mutation detection and filtering

Mutect

144,284 somatic mutations

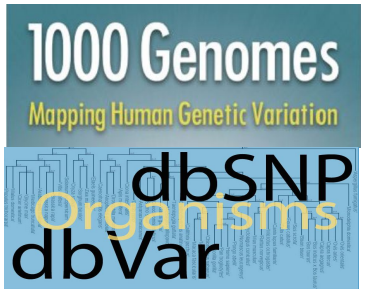
Filter for NF2 network gene

NF2 gene network



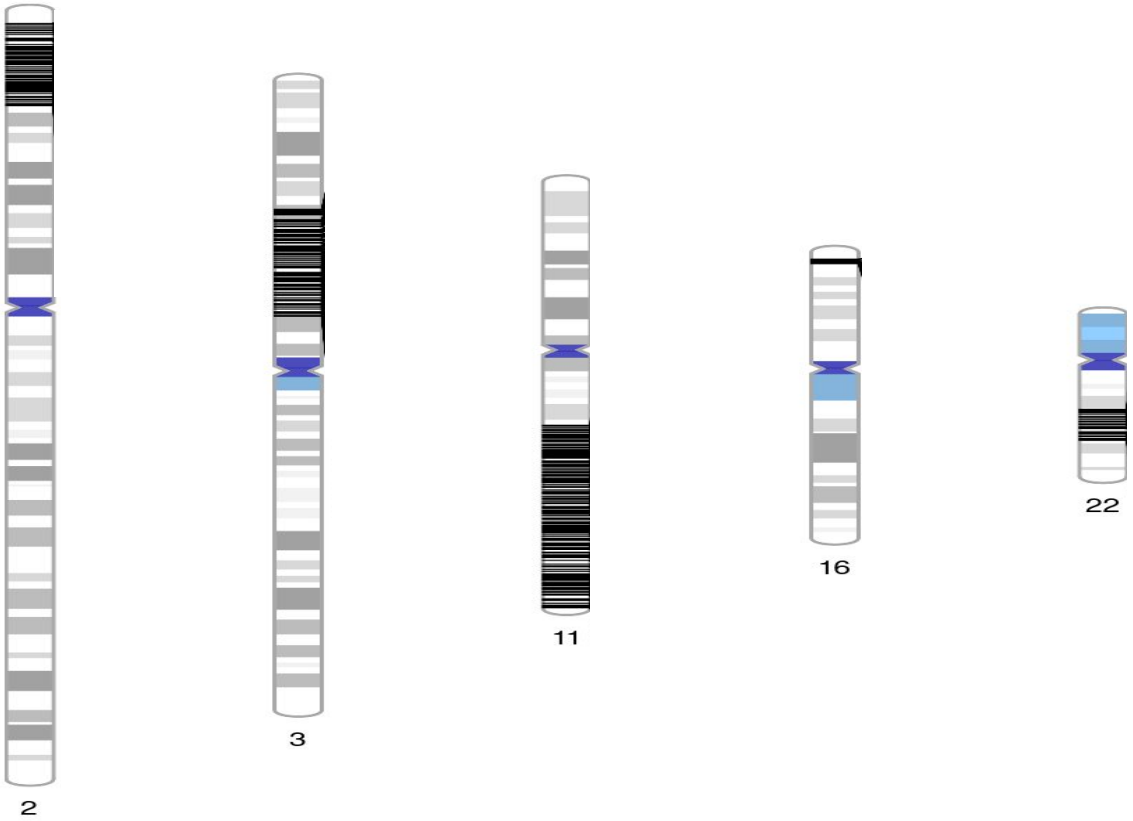
BioGrid: curated interaction repository traversing publications for protein and genetic interactions

Remove known variants

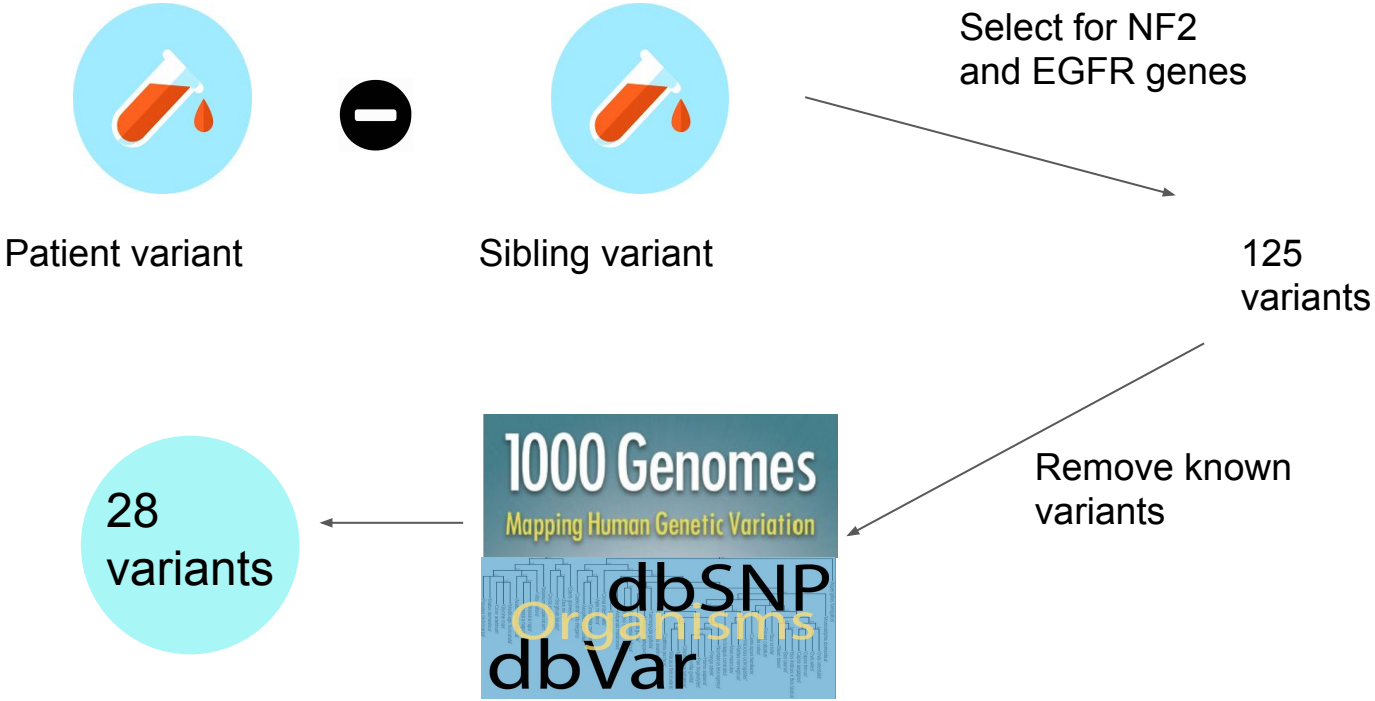


1,628 variants

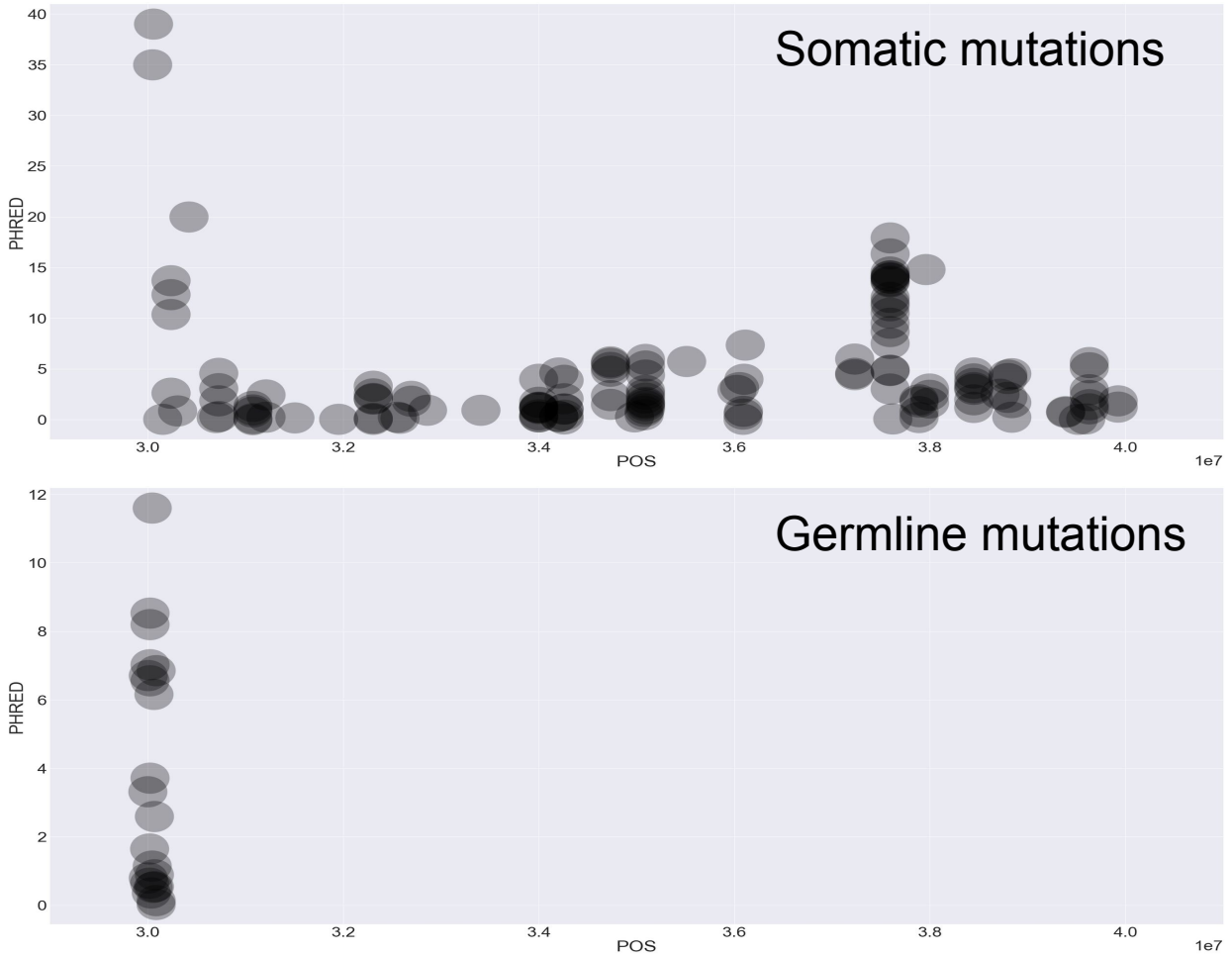
Distribution of somatic mutations in NF2 gene network



Germline mutation detection and filtering



Landscape of variants in the NF2 gene

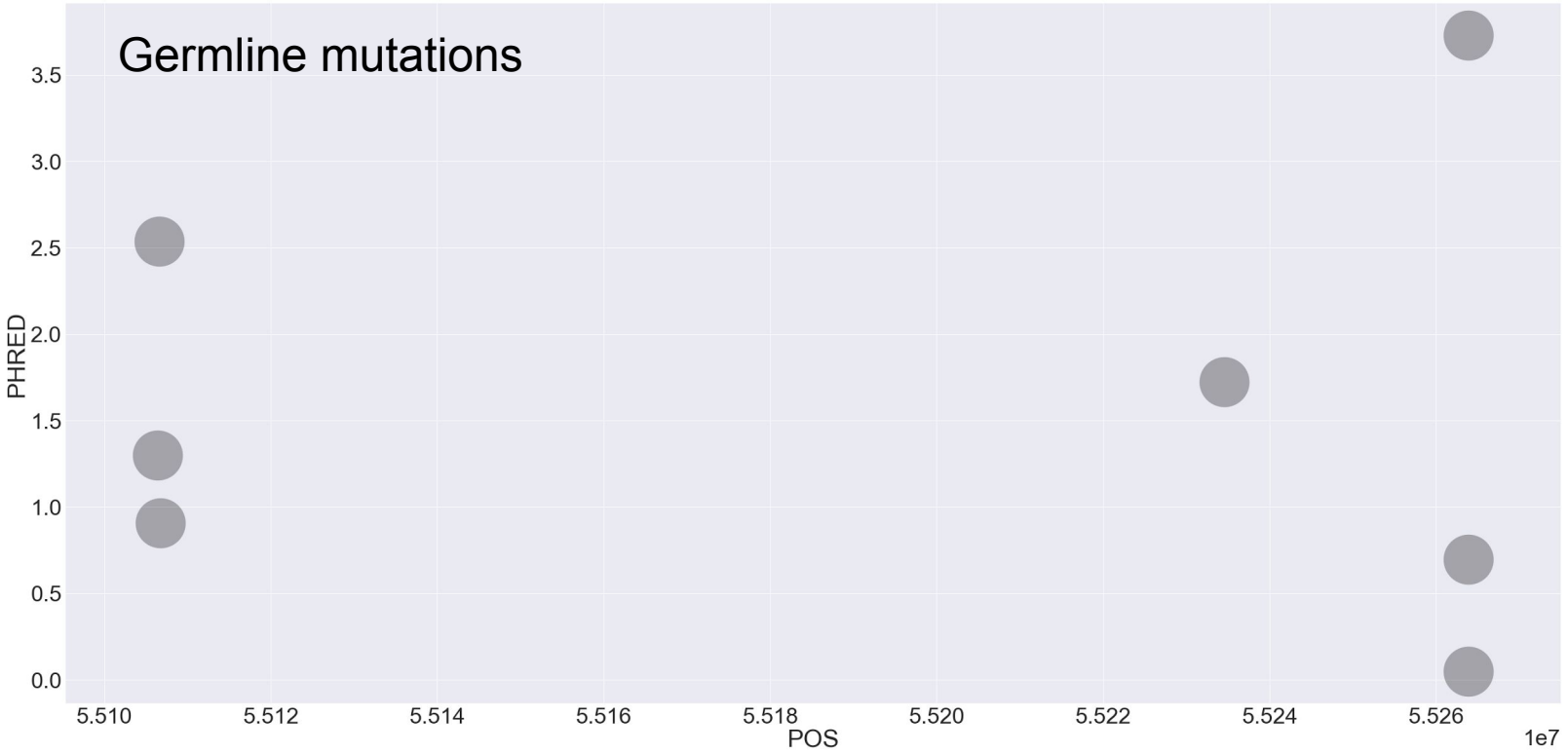


Conclusions and future directions

- Combine annotation scoring matrix + genome sequencing of tumor/normal can be used to prioritize variants that could be implicated in NF2 disease

- Collect the CADD scores of all somatic mutations.
- Genotyping parents or other NF2 patients to reduce false-positives in germline mutation detection.

Landscape of variants in the EGFR gene



SOMATIC VARIANTS:

Original number of the somatic variants (CADD output): **1813 (BEFORE)**

Removed: 181 SNPs and 16 indels from 1KGP

Final number of variants: **1626 (AFTER)** (no overlap with ClinVar but <1% of ClinVar variants are somatic so the probability of finding of a overlap is low)

GERMLINE VARIANTS:

Original number of the germline variants (CADD output): **125 (BEFORE)**

Removed: 96 SNPs and 1 indel

Final number of variants: **28 (AFTER)**