Team H-No-H

Marc Cuevas - Machine and Deep Learning

Rod Wilhelmy - Data Science

Luis Esquivies - Translational Science / Structural Biology

Sofia Medina - Developmental Biology and Genomics

Rene Lopez - Bioinformatics Cancer and Gut Microbiome

Standard Tumor/Normal Analysis

Case - control (1000 genomes, Tumor/Normal, Sibling)

Somatic variants (Tumor / Normal Analysis)

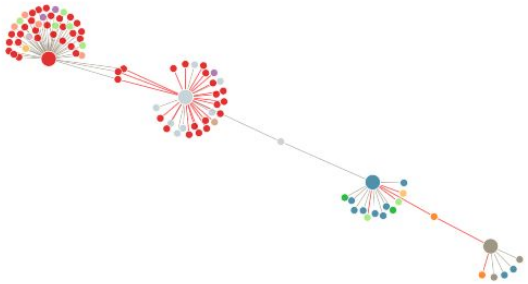
Regions of Interest (Protein Effect and Coexpression analysis)

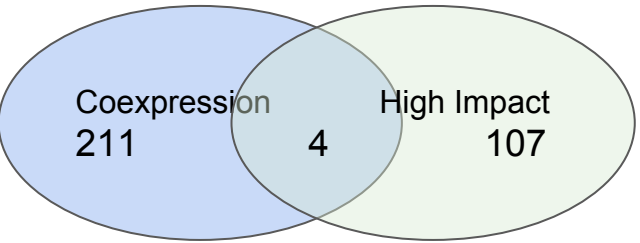
Variant filtering (Protein effect using regions of interest)

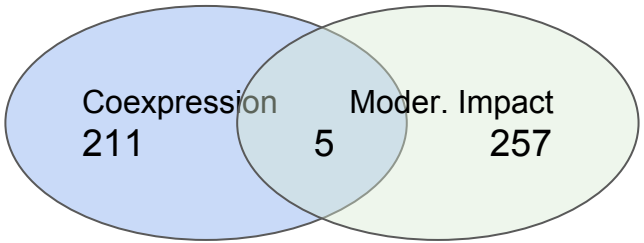
Translational Science (Interpretation)

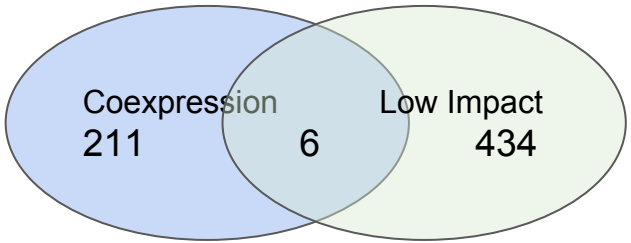
Regions of interest (Genes)

- NF2, FAM208B, SLC23A2, SRSF4, ELP3, C3, EGFR
- Spatio-temporal co-expression Analysis
  - List of 211 genes
  - Relevant genes from expression analysis
- Genes from High Impact / Deleterious variants
  - List of 107 genes
  - High impact effect / Deleterious variants



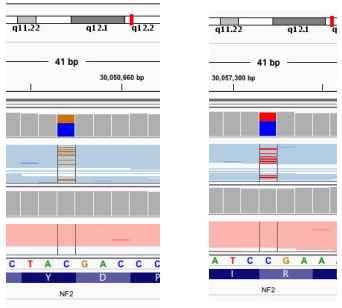






Translational Science - Structural biology

NF2 - Loss of function on heterozygous variants

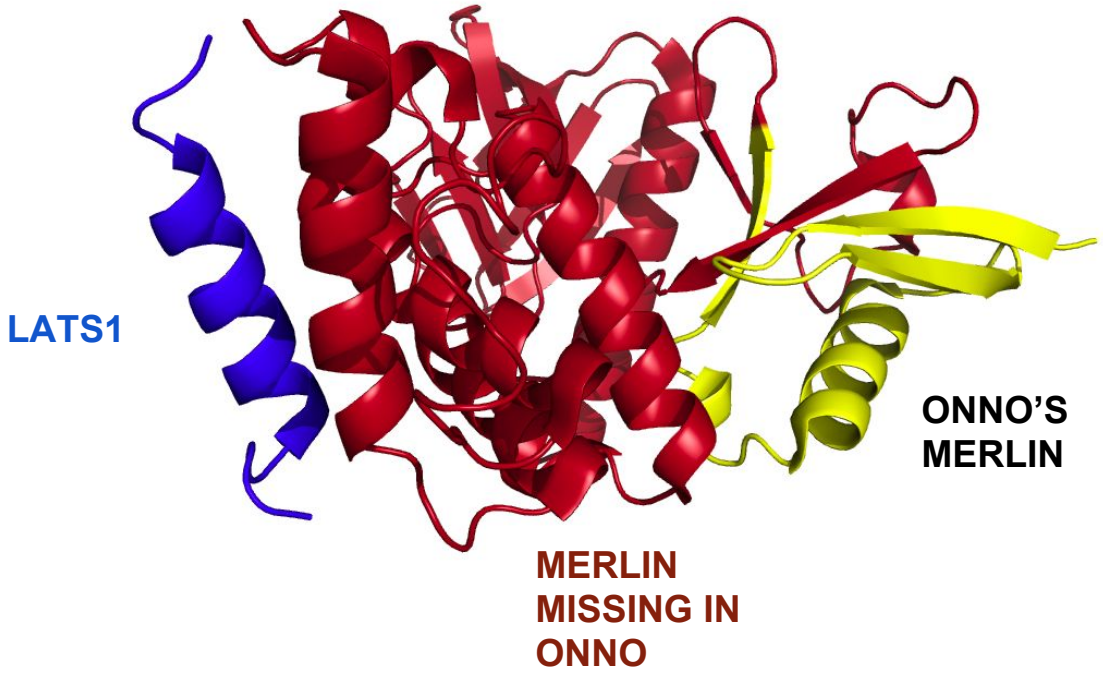


- NF2, 22:30050657, Premature Stop Codon Variant (LOF)
- NF2, 22:30057302, Stop Codon Variant (LOF), rs74315496

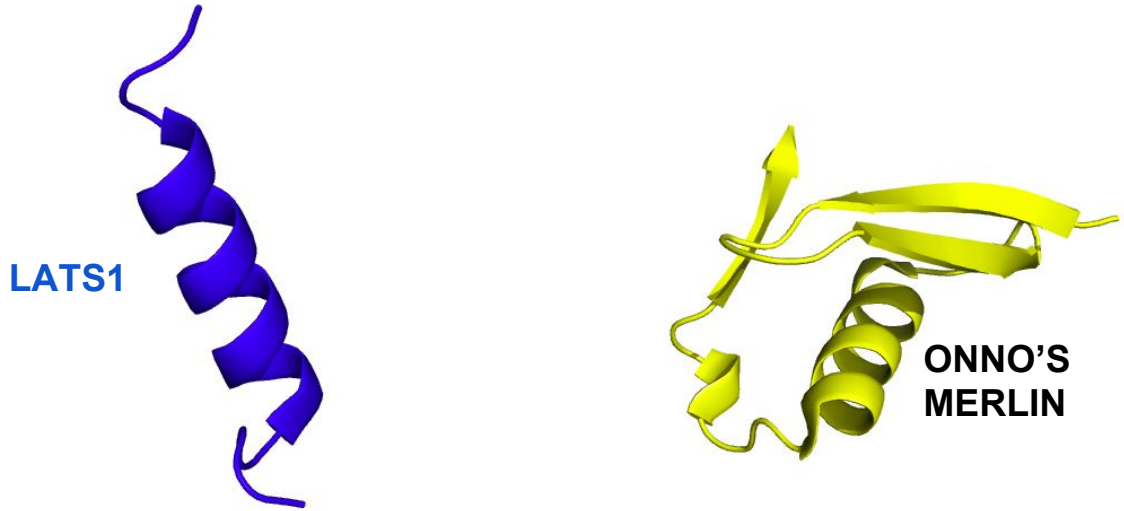
EGFR - Structural Interaction Variants

- 7:55249063, G>A, Structural Interaction Variant, rs1050171
- 7:55268949, A>G, Structural Interaction Variant, rs55737335
  - Structural Interaction Variant: A variant that impacts the internal interactions of the resulting polypeptide structure
  - It affects insertion TRNA probability because of rare codons
  - Synonymous mutations might also affect integration of the protein into the membrane

Truncated NF2/Merlin Protein



Truncated NF2/Merlin Protein



Onno's protein is not able to interact with tumor supressor protein

Variant Scoring (M-CAP - Machine Learning)

M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity, Nature Genetics, 2016

- Gradient boosting tree classifier
  - Random forest with gradient boosting
- Training and testing sets constructed from HGMD and ExAC databases
  - Pathogenic single-nucleotide variants obtained from HGMD
  - Predominantly benign variants obtained from ExAC
- Combines well established measures pathogenicity and conservation:
  - SIFT13, PolyPhen-2 (ref. 14), CADD15, MutationTaster20, MutationAssessor21, FATHMM22, LRT23, MetaLR16, and MetaSVM16.
  - RVIS24, PhyloP25, PhastCons26, PAM250, BLOSUM62, SIPHY28, and GERP29.
- Adds 298 additional features
  - Derived from 99 alignments of genomes from human, mammalian and vertebrate
  - Derives a sophisticated measure of evolutionary constraint via learning model