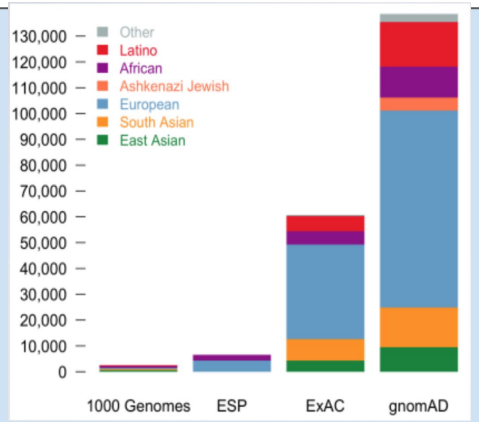
# Zimmerone Projects summary slides

2017-2023

# 2017 group 1

No genes prioritized

Summary Figure:



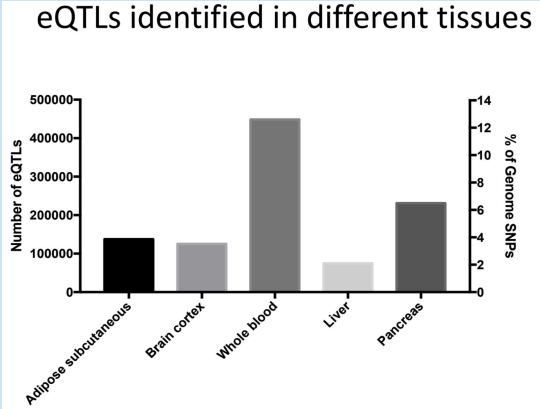

Summary:

1. Project purpose: comparative analysis of personal genomes
   a. Compare the variant in Carl's genome with those in genomAD and 1000 genome databases
2. genomic databases are important for understanding the "normal" range of genetic variation
3. vcfR retrieves online references
4. ANNOVAR retrieves genome frequencies and annotations

# 2017 group 2

No genes prioritized

Summary Figure:



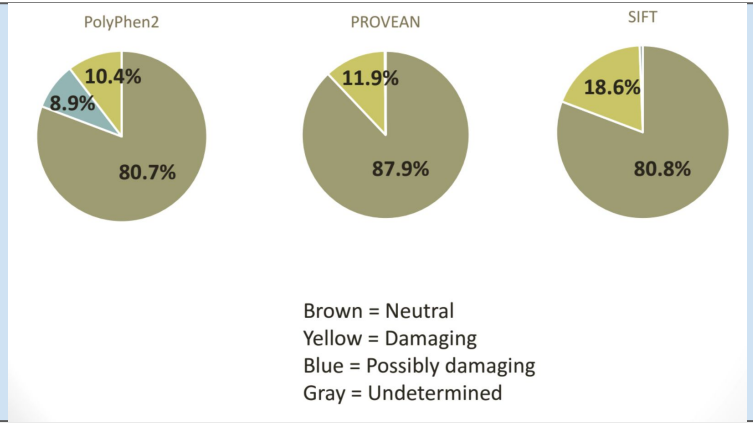

Summary:

1. Project purpose: comparative analysis of personal genomes
   a. Compare the variants in Carl's genome with those in GTEx databases
2. Carl's SNPs are interestingly enriched in immune-related chromosome regions
3. A small proportion of SNPs serve as eQTLs and regulate the expression of genes (mainly in the immune system)

# 2017 group 3

No genes prioritized

Summary Figure:



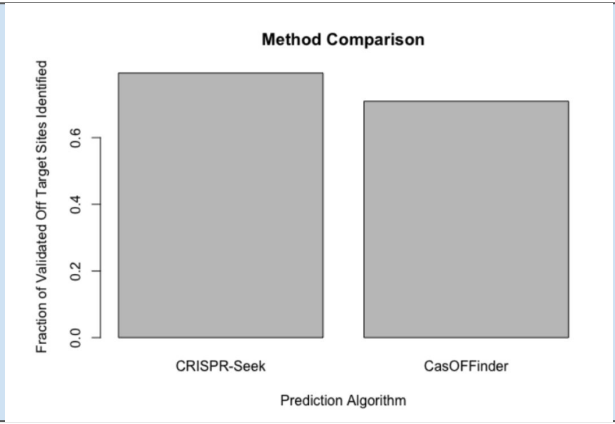

Summary:

1. Project purpose: comparative analysis of personal genomes
   a. Assessing deleteriousness of SNV's in Carl's genome
2. Used PolyPhen2, PROVEAN and SIFT for SNPs classification
3. Estimated the deleteriousness of a mutation based on properties of amino acids and evolutionary constraint

# 2017 group 4

No genes prioritized

Summary Figure:



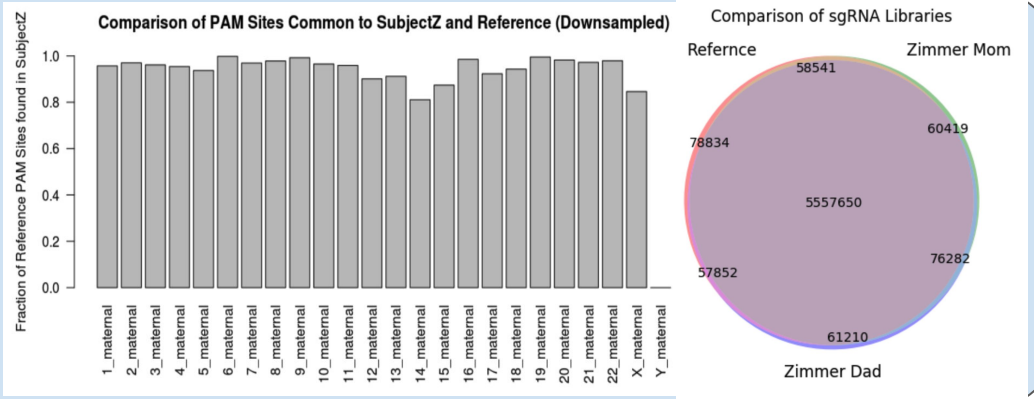

Summary:

1. Project purpose: personal genomes and personalized medicine (CRISPR)
   a. Identifying off-target CRISPR site
2. Used algorithms Cas-OFF and CRISPR-SEEK to identify off-target CRISPR sites
3. Results on hg38 showed that CRISPR-Seek predicted 516/650 and CasOFFinder 461/650

# 2017 group 5

No genes prioritized

Summary Figure:



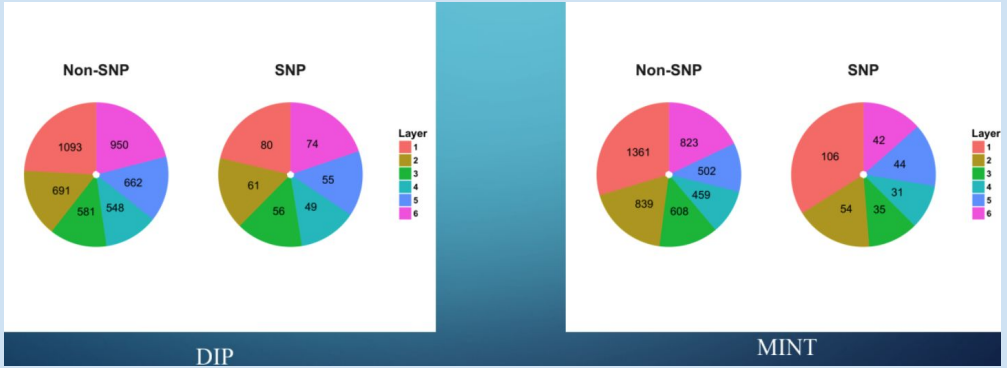

## Summary:

1. Project purpose: personal genomes and personalized medicine (CRISPR)
   a. The impact of SNPs on sgRNA sets and off target mutations
2. Found PAM sites in the human reference genome as well as Carl's genome and compares the similarity of the two sets:
   a. Many PAM sites with matched upstream sequences are present in Carl's sets
3. Generated sgRNA libraries from Carl's genome and compared it to the reference genome
   a. Zimmer mom and dad have slightly more in common than with either does with the reference

# 2017 group 6

No genes prioritized

Summary Figure:



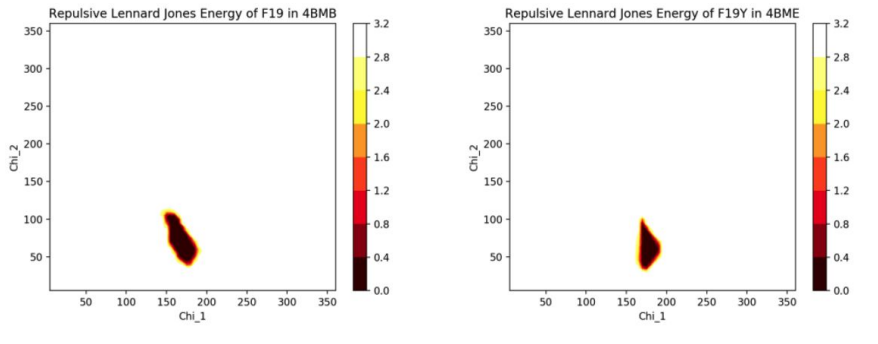

Summary:

1. Project purpose: network analysis of personal genomes
   a. Propose a tool that calculates the degree centrality and betweenness centrality of proteins containing and not containing SNPs in Carl's genome using a PPI file
2. Calculated the degree centrality and the betweenness centrality for all nodes in the PPI network
3. Found no statistically significant difference in distributions of proteins containing SNPs and not containing SNPs
4. Observed no statistically significant enrichment of proteins/protein types in any hierarchical layers of the PPI network

# 2017 group 7

No genes prioritized

## Summary Figure:



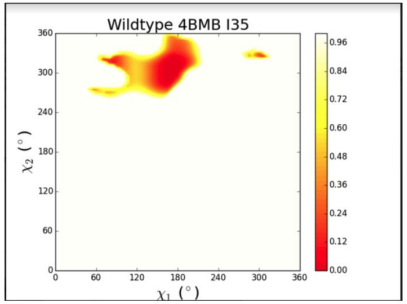

## Summary:

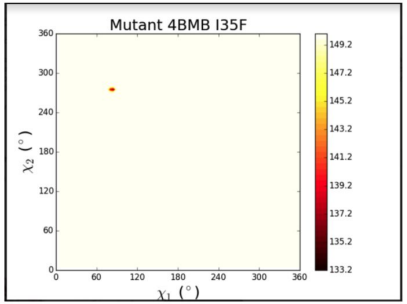
1. Project purpose: structural analysis
   a. Analyze the structure of the mutation F19Y found in Carl's genome for galectin-8
2. Amino acids mutations can be favorable or unfavorable
3. Analyzed Carl's genome for protein galectin-8 mutations
   a. Compared WT vs Mutant's repulsive energy and RMSD

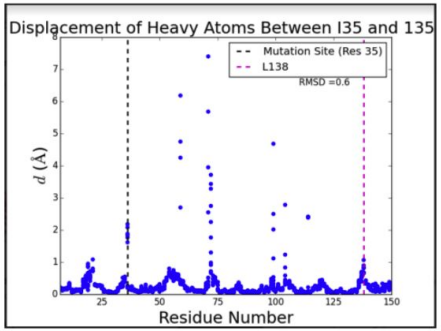
# 2017 group 8

No genes prioritized

## Summary Figure:



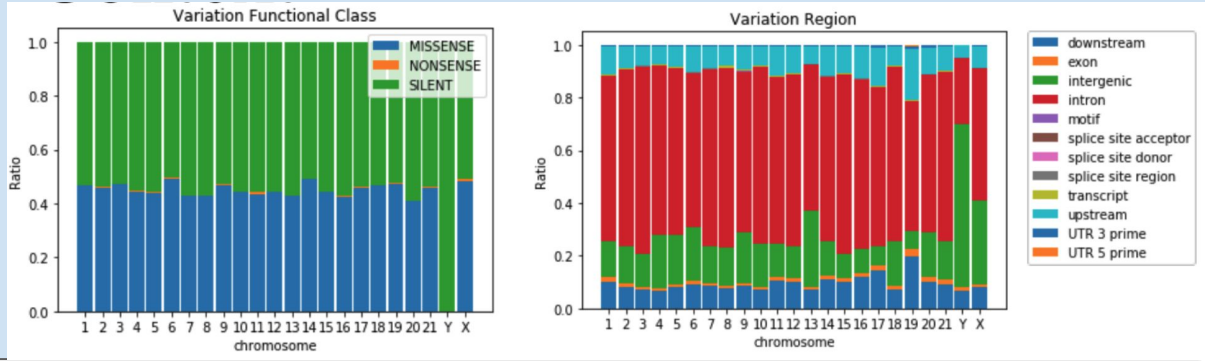

## Summary:

1. Project purpose: structural analysis
    a. Analyze the structure of the mutation I35F found in Carl's genome for galectin-8 at location 35
2. Amino acid mutations include changes in size, charge, hydrophobic effect, and change in potential energy
    a. Size and charge are important in predicting mutation effects
3. Side chain angle potential energies increased by I35F mutations
4. RMSD increased with I35F mutation
5. Unexpected mutation results (D/E, F/Y)

# 2018 group 1

No genes prioritized

Summary Figure:



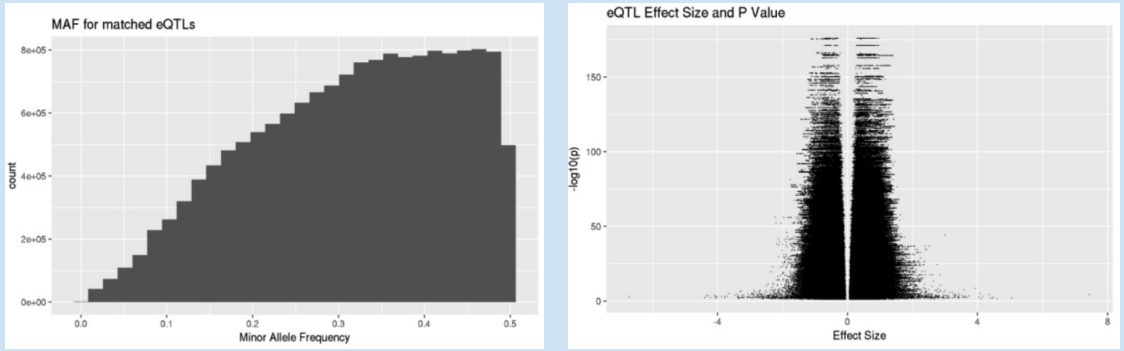

Summary:

1. Project purpose: comparative analysis of personal genomes
   a. Compare Carl's genome with the Neanderthal data
2. Neanderthal variants have been observed in humans in a frequency higher than expected by chance; some of these variants are associated with disease risk.
3. Retrieved Neanderthal data came from The Max Planck Institute and used bedtools intersect to analyze basic statistics of %N SNP composition of Carl's SNP set

# 2018 group 2

No genes prioritized

Summary Figure:



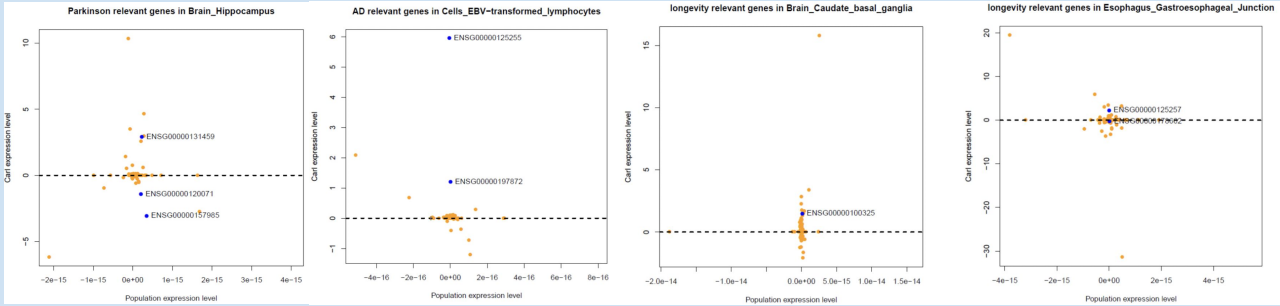

Summary:

1. Project purpose: comparative analysis of personal genomes
    a. Compare the variants in Carl's genome with those in GTEx databases to assess whether Carl should take up sprinting
2. Carl's genome indicates that he:
    a. may have abnormal regulation of axon guidance (related to Parkinson's)
    b. may have more (or less!) explosive fast-twitch muscles than average
    c. may be at risk for gastrointestinal tract cancer

# 2018 group 3

No genes prioritized

## Summary Figure:



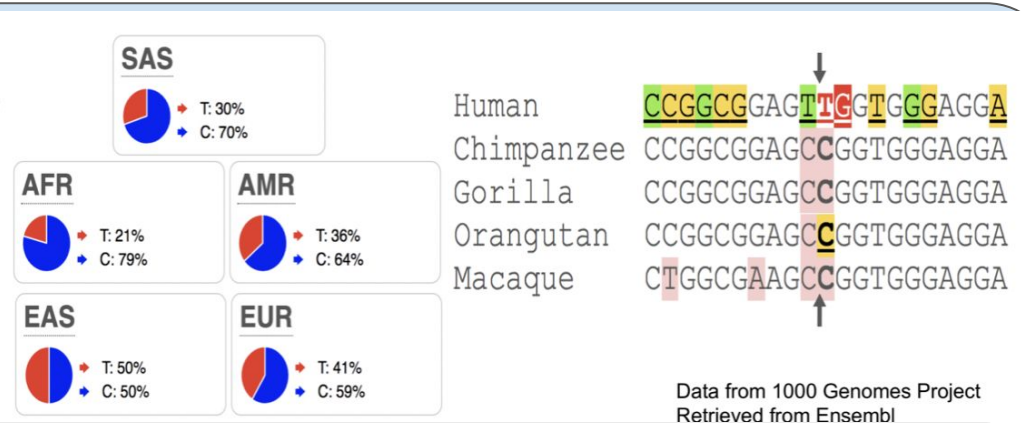

## Summary:

1. Project purpose: predict gene expression values from carl's snp information
2. Potentially higher risk of Parkinson's disease
   a. Lowered expression level for a protective gene AGAP1
   b. Lowered expression level for a important gene KANSL1 (in mental development)
3. Potentially higher risk for Alzheimer's disease
   a. Increased expression level for a risk gene SLC10A2
4. More likely to have a longer life span
5. Increased expression for two protective genes (ASCC2 and MRP4)

# 2018 group 4

No genes prioritized

Summary Figure:




Summary:

1. Project purpose: analyzing a common variant associated with inflammatory response
2. Investigated in cannabinoid receptor type 2
3. The CC variant is the common and ancestral variant
4. Cannabinoid receptor type 2 is involved in immune signaling
5. Disease significance: the cannabinoid receptor type 2 Q63R variant increases the risk of celiac disease

# 2018 group 5

No genes prioritized

Summary Figure:

| Gene (SNP) | Beta coefficient (log odds ratio) |
|---|---|
| APOE(rs429358 & rs7412) | 0.566 |
| CR1 (rs6656401) | 0.165514 |
| PICALM (rs10792832) | -0.13926 |
| MS4A6A (rs983392) | -0.10536 |
| CD2AP (rs10948363) | 0.09531 |
| EPHA1 (rs11771145) | -0.10536 |
| INPP5D (rs35349669) | 0.076961 |
| NME8 (rs2718058) | -0.07257 |
| ZCWPW1 (rs1476679) | -0.09431 |
| CELF1 (rs10838725) | 0.076961 |

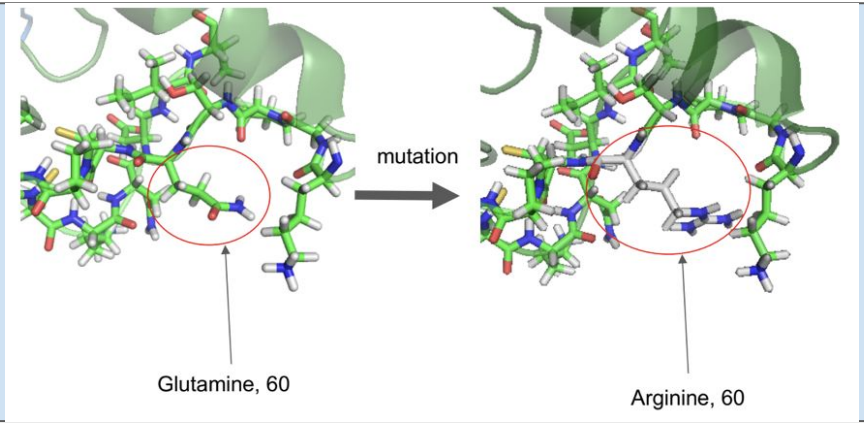
Summary:

1. Project purpose: calculating Carl's risk for Alzheimer's disease
2. Alzheimer's disease phenotype is indicated by the amyloid cascade hypothesis
3. Carl has 10 out of 22 associated SNPs
   a. Probability of Alzheimer's is 13.72% vs probability of Alzheimer's in general population is 10%

## 2018 group 6

No genes prioritized

Summary Figure:



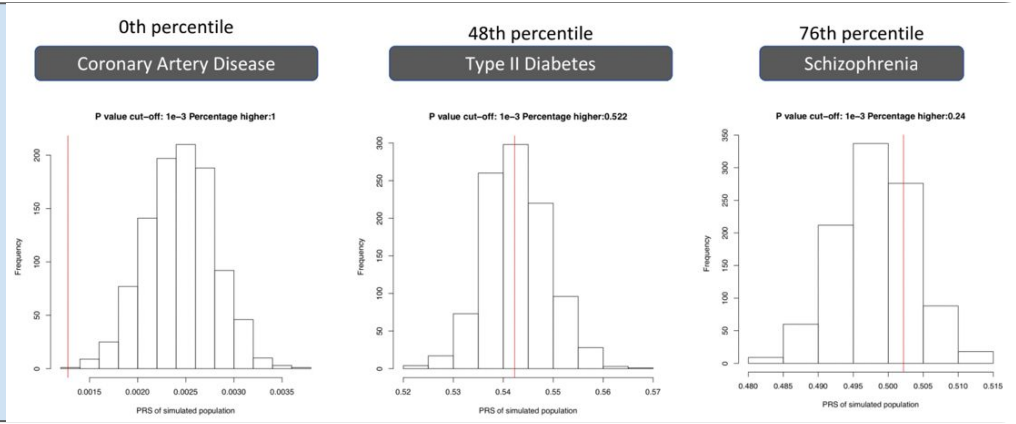

Summary:

1. Project purpose: identifying significant protein-coding mutations in Carl's genome
2. Hypothesizes that Carl's mutation will not drastically decrease protein functionality and thus not increase his risk of developing cancer
3. Conclusions:
    a. Mutations appear to be tolerated in coding regions
    a. Olfactory receptors
    b. Polarity is preserved among most mutation

# 2018 group 7

No genes prioritized

Summary Figure:




Summary:

1. Project purpose: PRS prediction in coronary artery disease, type II diabetes, and schizophrenia
1. Pre-processed Carl's genotype file then calculated his polygenic risk scores for coronary artery disease, type II diabetes, and schizophrenia
2. Limitations:
   a. PRS is only interpretable when compared to that of cases/controls
   b. Simulation result is not robust and depends heavily on simulation hypothesis

# 2019 group 1 (chr 1)

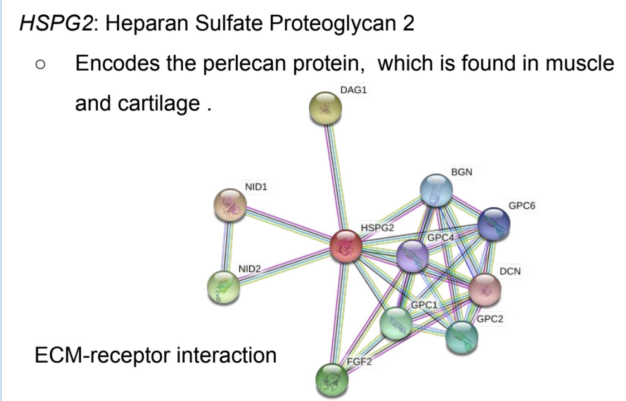
Top 10 Prioritized Genes

1. NBPF19
2. LMX1A
3. OBSCN
4. KIAA1324
5. HSPG2
6. TESK2
7. UBXN11
8. SPATA6
9. ADORA1
10. DCAF6

Additional Gene(s)

HRNR

Summary Figure:




Summary:

1. Prioritization approach: exonic mutational burden
2. Downstream analysis: STRING network analysis of protein interactions
3. Findings:
   a. STRING analysis of the HSPG2 gene shows that Carl has six SNPs associated with Schwartz Jampel syndrome
   b. STRING analysis of the HRNR gene (from the SIFT sorting list) shows that Carl has one SNP that's associated with eczema in caucasian Europeans

# 2019 group 2 (chr 2)

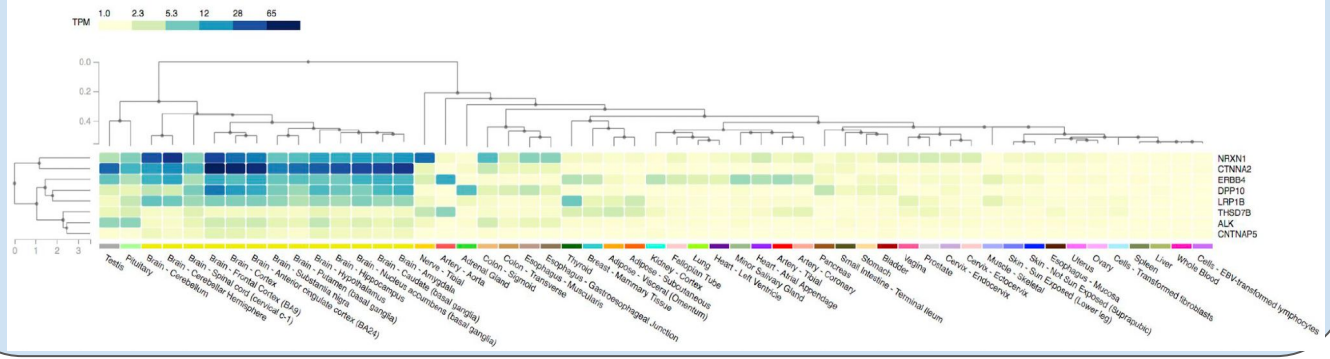
Top 10 Prioritized Genes

1. LRP1B
2. AC007682
3. CTNNA2
4. NRXN1
5. ERBB4
6. AC009499
7. ALK
8. THSD7B
9. DPP10
10. CNTNAP5

Additional Gene(s)

C2orf16

Summary Figure:




Summary:

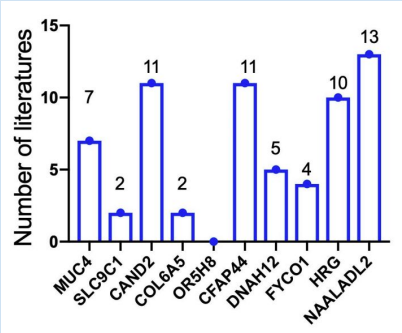
1. Prioritization approach: mutational burden
2. Downstream analysis: gene expression analysis
3. Findings:
   a. NRXN1,CTNNA2, ERBB4, DPP10, LRP1B have high expression levels in brain
   b. SNP in C2orf16 is associated with the MEtabolic Syndrome

# 2019 group 3 (chr 3)

Top 10 Prioritized Genes

1. MUC4
2. SLC9C1
3. CAND2
4. C0L6A5
5. OR5H8
6. CFAP44
7. DNAH12
8. FYC01
9. HRG
10. NAALADL2

## Summary Figure:




| Gene | Most Frequent Term |
|---|---|
| MUC4 | Inflammatory Bowel Disease |
| SLC9C1 | Sodium-Hydrogen Exchanger |
| CAND2 | Atrial Fibrillation |
| COL6A5 | Asthma |
| OR5H8 | None |
| CFAP44 | Multiple morphological abnormalities of the sperm flagella |
| DNAH12 | Dynein |
| FYCO1 | Cataract |
| HRG | Thrombosis |
| | Histidine-rich glycoprotein |
| | Activated partial thromboplastin time |
| NAALADL2 | N-acetylated alpha-linked acidic dipeptidase-like 2 |
| | Prostate cancer |
| | Intellectual disability |
| | Cornelia de Lange syndrome |

## Summary:

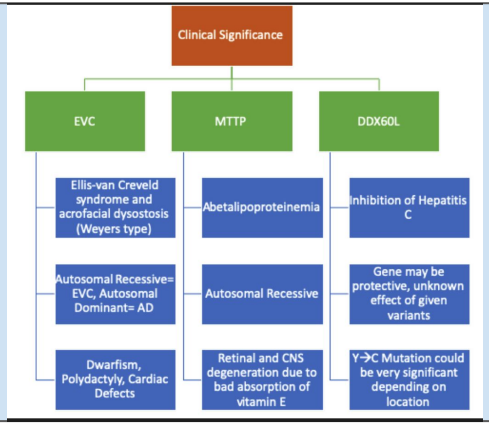
1. Prioritization approach: number of nonsynonymous SNVs
2. Downstream analysis: text mining analysis through APIs and PubTator
3. Findings:
   a. Different SNPs in the same gene may have distinct consequences: one SNP in CAND2 is associated with Atrial fibrillation and Ataxia Telangiectasia while another SNP is related to obesity
   b. Identified a gene that has high rare nonsynonymous SNVs burden: ALDH1L, which has a SNP that's associated with Cleft Palate

# 2019 group 4 (chr 4)

Top 10 Prioritized Genes

1. FAT1
2. DCHS2
3. DUX4L4
4. FRAS1
5. KIAA1211
6. MTTP
7. NEIL3
8. PIGG
9. SHROOM3
10. ALKPK1

Summary Figure:




Summary:

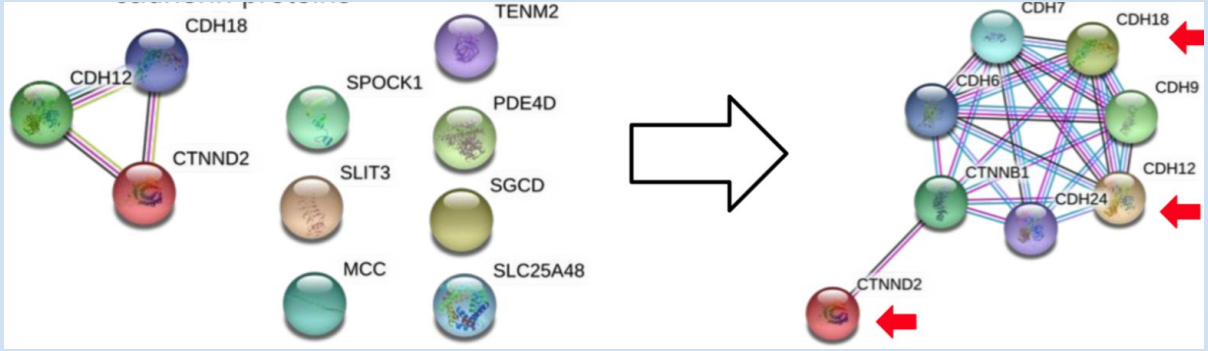
1. Prioritization approach: mutational burden of nonsynonymous variants
2. Downstream analysis: PDB structural analysis
3. Findings:
   a. Found three variants DDX60L has that could potentially reduce the ability of this protein's ability to perform its protective function
      i. However, without structural information this is highly speculative
   b. Found variation in MTTP, a gene associated with abetalipoproteinemia when functionality is reduced

# 2019 group 5 (chr 5)

Top 10 Prioritized Genes

1. CDH12
2. CDH18
3. PDE4D
4. CTNND2
5. SGCD
6. SPOCK1
7. MCC
8. SLC25A48
9. TENM2
10. SLIT3

Summary Figure:




Summary:

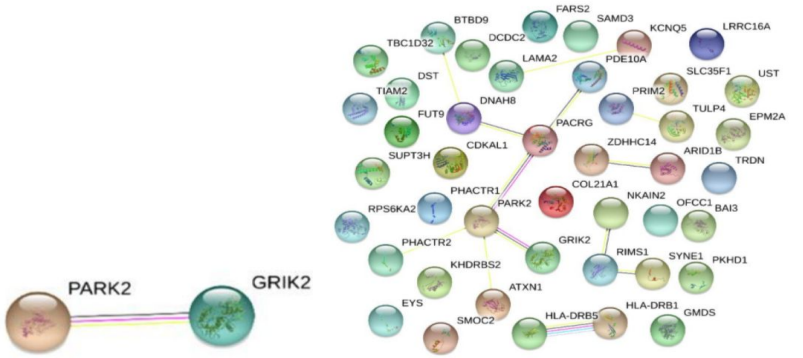
1. Prioritization approach: mutational burden
2. Downstream analysis: STRING network analysis
3. Findings:
   a. Genes from original mutational burden analysis shows connection with cadherin proteins
   b. Carl has high mutational burden for multiple genes linked to CDH12
   c. Chromosome 5p deletion including CDH12 causes cri-du-chat syndrome

# 2020 group 1 (chr 6)

Top 10 Prioritized Genes

1. HLA-DRB1
2. HLA-DRB5
3. HLA-DRB6
4. HLA-DRA
5. HLA-DQA1
6. HLAB
7. GRIK2
8. PACRG
9. DDR1
10. CCHCR1

Summary Figure:




Summary:

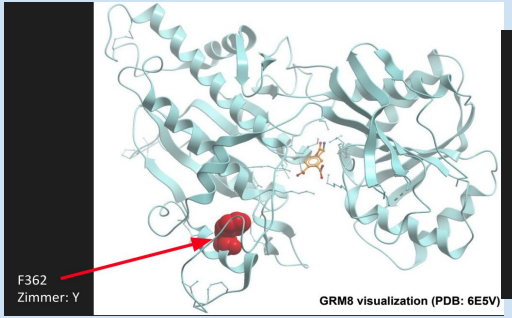
1. Prioritization approach: Putative variant impact and total count of variants
2. Downstream analysis: STRING network analysis
3. Findings:
   a. Biggest Aggregation of rare SNVs are located near a gene complex (HLA) coding for major histocompatibility complex on Chromosome 6
   b. PACRG has high degrees in the network analysis and are associated with Parkinson Disease

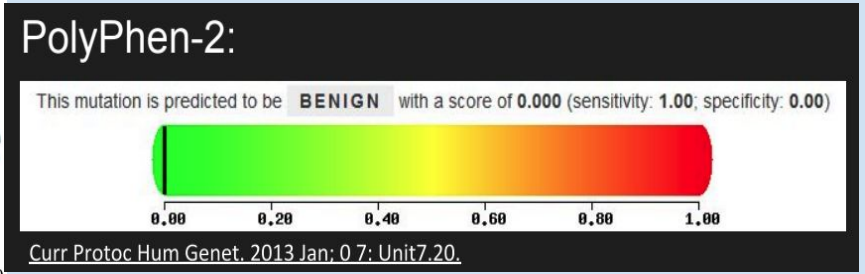
# 2020 group 2 (chr 7)

Top 10 Prioritized Genes

1. CNTNAP2
2. MAGI2
3. PTPRN2
4. DPP6
5. SDK1
6. DGKB
7. AUTS2
8. HDAC9
9. GRM8
10. PDE1C

Summary Figure:




Summary:

1. Prioritization approach: mutational burden
2. Downstream analysis: PDB structural analysis
3. Findings:
   a. Among the top 10 most mutated genes on chromosome 7, there are 6 missense variants within 4 genes
   b. Only one variant, conferring a protein, is characterized: GRM8
   c. PolyPhen analysis shows that substitution at pos 362 from F to Y is predicted to be tolerated

# 2020 group 3 (chr 8)

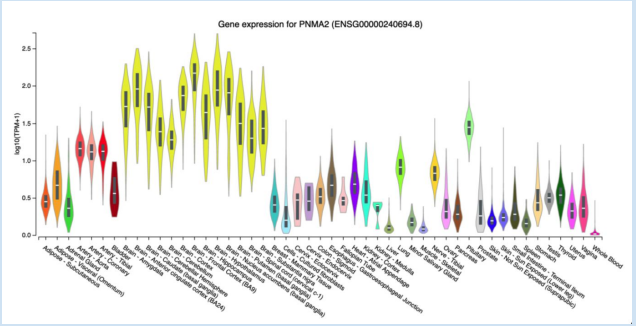
Top 10 Prioritized Genes

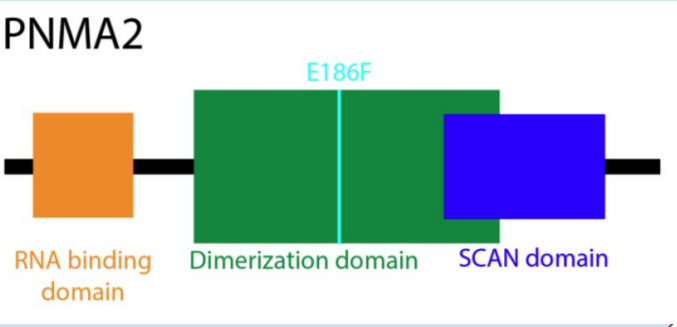
1. CSMD1
2. SGCZ
3. LINC02055
4. NRG1
5. DLGAP2
6. DLC1
7. CSMD3
8. CCDC26
9. PSD3
10. NKAIN3

Additional Gene(s)

PNMA2

Summary Figure:




Summary:

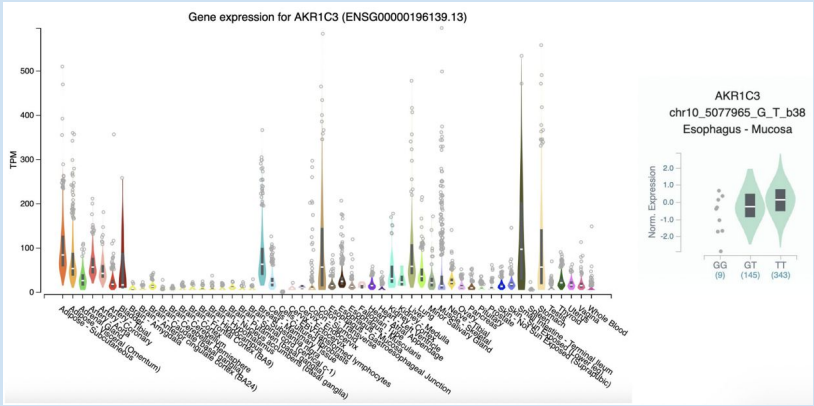
1. Prioritization approach: mutational burden
2. Downstream analysis: gene expression analysis
3. Findings:
   a. PNMA2 is highly expressed in the brain and is implicated in cancer and neurological disorders
   b. PNMA2 C>T causes nonsynonymous variant which lies in the dimerization domain

# 2020 group 4 (chr 9)

Top 10 Prioritized Genes

1. PTPRD
2. ASTN2
3. TRPM3
4. LINGO2
5. KDM4C
6. GLIS3
7. ADAMTSL1
8. CCDC171
9. BNC2
10. PCSK5

Summary Figure:




Summary:

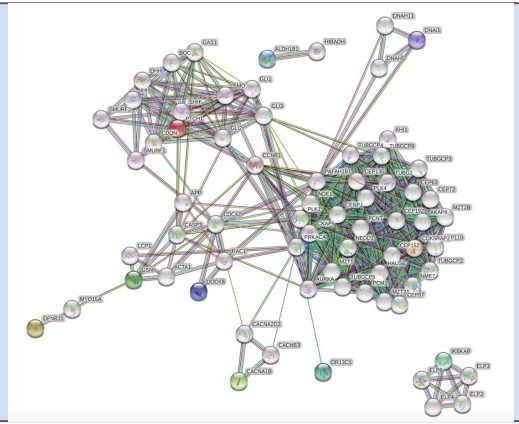
1. Prioritization approach: mutational burden
2. Downstream analysis: gene expression analysis
3. Findings:
   a. ADAMTSL1 is an interesting genes that's involved in extracellular matrix organization
   b. BNC2 is found to be strongly associated with congenital lower urinary-tract obstruction
   c. Gene PTPRD has been implicated as a locus for restless legs syndrome and genetic variants of PTPRD are associated with bronchial asthma

# 2020 group 5 (chr 10)

Top 10 Prioritized Genes

1. AKR1E2
2. AKR1C2
3. AKR1C3
4. TMEM254-AS1
5. ASAH2B
6. WDR37
7. SEPHS1
8. PTCHD3
9. SNCG
10. PITRM1

Summary Figure:



Summary:

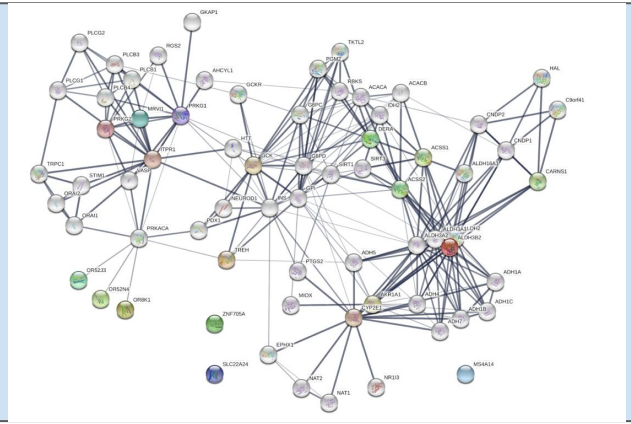
1. Prioritization approach: mutational burden
2. Downstream analysis: STRING network analysis
3. Findings:
   a. Rare, deleterious variant found within ASAH2B
   b. PTCHD3 variant found to be associated with decreased risk for cataracts
   c. AKR1E2 stop gained mutation is associated with lower height and fat-free body mass; AKR1C2 splice-donor variant is associated with lower weight, fat-free body mass; AKR1C3 stop gained mutation is associated with lower fat-free body mass and basal metabolic rate

# 2020 group 6 (chr 11)

Top 10 Prioritized Genes

1. SLC22A24
2. CASP12
3. OR52J3
4. ZNF705E
5. ALDH3B2
6. MRVI1
7. OR52N4
8. TREH
9. MS4A14
10. OR8K1

Summary Figure:



Summary:

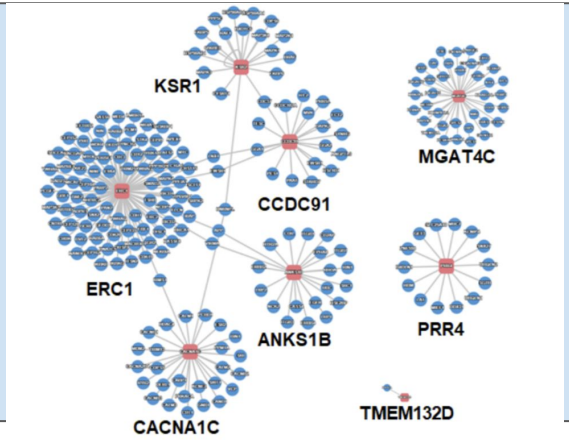
1. Prioritization approach: impact score using SnpEff
2. Downstream analysis: STRING network analysis
3. Findings:
   a. Several olfactory receptors were found in this analysis
   b. Some genes were involved in metabolic processes

# 2021 group 1 (chr 12)

Top 10 Prioritized Genes

1. TMEM132D
2. MGAT4C
3. PRR4
4. ANKS1B
5. FAM19A
6. ERC1
7. CACNA1C
8. KSR2
9. TAS2R14
10. CCDC91

Summary Figure:




Summary:

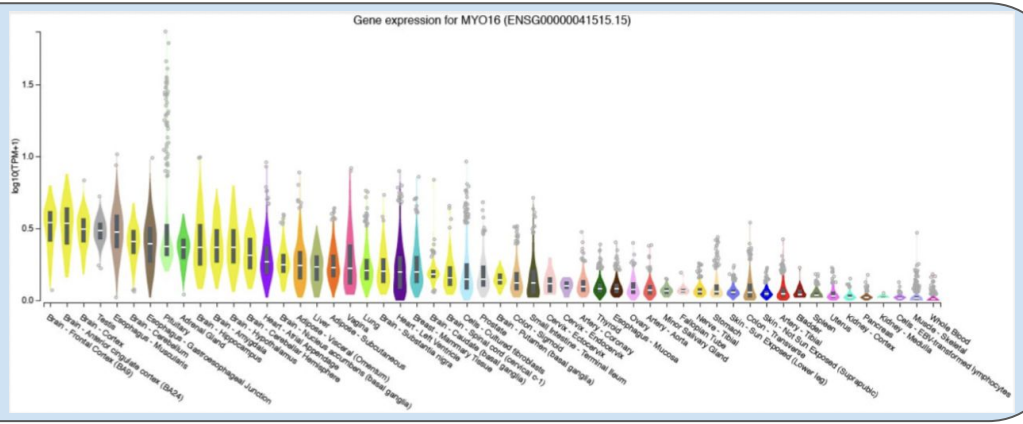
1. Prioritization approach: mutational burden
2. Downstream analysis: PINA and STRING network analysis
3. Findings:
   a. mutations may be relevant to taste transduction pathway
   b. ERC1 expression is associated with a number of genes correlated with cancer prognosis; KSR2 interacts with MAPK (signaling-pathway involved in oncogenesis)
   c. alternative approach to gene ranking finds interesting gene:
      i. KSR2 variant related to facial morphology

# 2021 group 2 (chr 13)

Top 10 Prioritized Genes

1. MYO16
2. ATP8A2
3. ATP11A
4. MCF2L
5. FARP1
6. PCDH9
7. DLEU1
8. RNF219-AS1
9. CCDC169
10. DIAPH3

Summary Figure:




Summary:

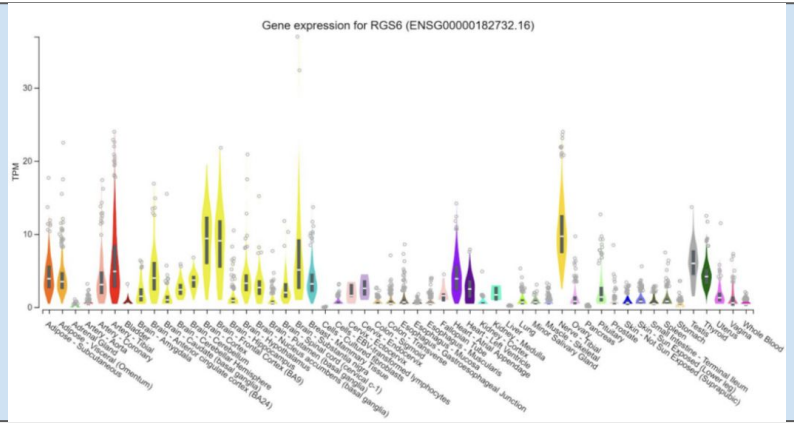
1. Prioritization approach: mutational burden
2. Downstream analysis: gene expression analysis
3. Findings:
   a. MYO16 has genetic association with Autism Spectrum Disorder
   b. Deletions of gene DLEU1 are found at high rates in patients with lymphocytic leukemia

# 2021 group 3 (chr 14)

Top 10 Prioritized Genes

1. RGS6
2. LINC00871
3. FOXN3
4. MDGA2
5. NPAS3
6. NRXN3
7. RAD51B
8. NUBPL
9. SLC25A21
10. AKAP6

Summary Figure:




Summary:

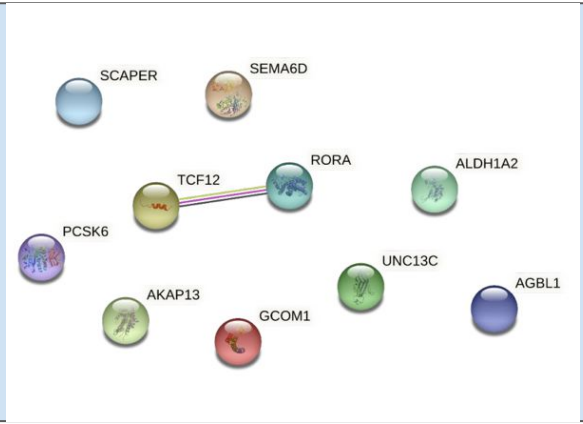
1. Prioritization approach: mutational burden
2. Downstream analysis: gene expression analysis
3. Findings:
   a. RGS6 modulates neuronal and cardiovascular activities, through regulating G protein signaling
   b. MDGA2 is overall highly expressed in brain tissues but also highly expressed in testis

# 2021 group 4 (chr 15)

Top 10 Prioritized Genes

1. ALDH1A2
2. RORA
3. SEMA6D
4. AKAP13
5. PCSK6
6. TCF12
7. SCAPER
8. GCOM1
9. UNC13C
10. AGBL1

Summary Figure:




Summary:

1. Prioritization approach: mutational burden
2. Downstream analysis: STRING network analysis
3. Findings:
   a. Two genes [RORA/TCF12] predicted to interact
   b. RORA and TCF12 are implicated in TH17 cell differentiation pathway
   c. Potential implications of RORA/TCF12 mutation
      i. RORA: neurodevelopmental delay; epilepsy; cerebral ataxia
      ii. TCF12: coronal craniosyntosis; anaplastic oligodendroma

# 2021 group 5 (chr 16)

Top 10 Prioritized Genes

1. PKD1L2
2. METTL22
3. ZNF469
4. UBE2I
5. SCNN1G
6. ADCY9
7. ACSF3
8. BCAR1
9. ABAT
10. CEP20

Summary Figure:

**Table 1: Two Most Common Phrases for Each Gene***

| *Gene* | *Most Common Phrase(s)* | *Second Most Common Phrase(s)* |
|---|---|---|
| PKD1L2 | Polycystem (or polystin) | Cerebrospinal-fluid-contacting neurons |
| UBE2I | E2-conjugating/differentially expressed genes | oogenesis |
| METTL22 | Hematopoietic stem cells | Human proteins |
| ACSF3 | Next-generation sequencing/RNA-sequencing | Breast cancer |
| SCNN1G | Liddle sundrome (ls) | Nephrotic syndrome |
| BCAR1 | osteoclasts | cancer/protein tyrosine kinase 6 |
| ABAT | Gamma-aminobutyric acid | Cabamazepine |
| ADCY9 | microRNA | asthma/molecular mechanisms |
| ZNF469 | Ehlers-Danlos Syndrome | Brittle cornea syndrome |
| CEP20 | Centriolar satellites | Pericentriolar satellites/centrosomal proteins |

**multiple terms in a cell indicate that they were detected with same frequency*

Summary:

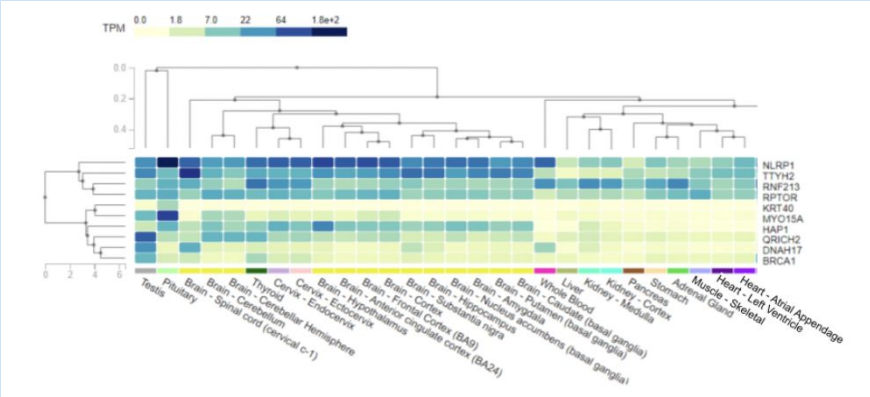
1. Prioritization approach: mutational burden
2. Downstream analysis: text mining analysis
3. Findings:
   a. SCNN1G is associated with liddle syndrome
   b. ADCY9 if part of the G-protein signaling cascade that begins with beta-adrenergic receptor activation
   c. ACSF3 is associated with combined malonic and methylmalonic aciduria

# 2021 group 6 (chr 17)

Top 10 Prioritized Genes

1. MYO15A
2. RNF213
3. KRT40
4. HAP1
5. BRCA1
6. RPTOR
7. TTYH2
8. DNAH17
9. NLRP1
10. QRICH2

Summary Figure:




Summary:

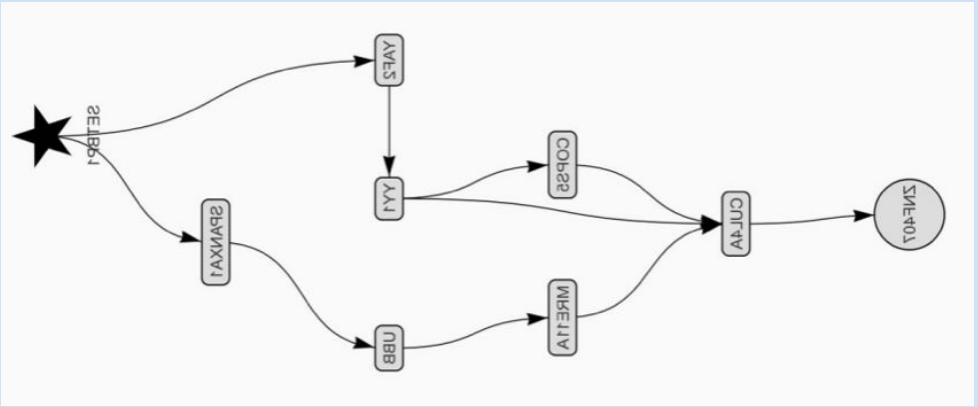
1. Prioritization approach: mutational burden
2. Downstream analysis: gene expression analysis
3. Findings:
   a. DNAH17 has highest (24) SNVs
      i. Missense mutations are associated with flagellar mutations and asthenozoospermia
      ii. Asthenozoospermia is only one of the causes of male infertility
      iii. Investigated if Carl's DNAH17 has a variant that increases risk of asthenozoospermia

# 2022 group 1 (chr 18)

Top 10 Prioritized Genes

1. ZNF407
2. SETBP1
3. PIEZO2
4. MYL12B
5. DCC
6. ZNF532
7. RAB27B
8. PTPRM
9. ANKRD12
10. DOK6

Summary Figure:




Summary:

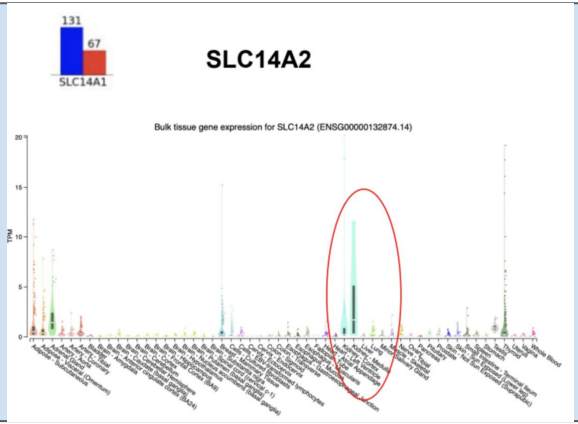
1. Prioritization approach: impact score using SnpEff
2. Downstream analysis: network analysis using KNIT
3. Findings:
   a. The network identified seems to be related to DNA damage repair/DNA replication
   b. ZNF407 is affected by WDR5 that could lead to colon cancer progress
   c. Mutation in SETBP1 is associated with Schinzel-Giedion midface retraction syndrome

# 2022 group 2 (chr 18)

Top 10 Prioritized Genes

1. SLC14A2
2. L3MBTL4
3. DOK6
4. PIGN
5. DCC
6. PIEZO2
7. DLGAP1
8. PTPRM
9. LINC00907
10. AQP4-AS1

Summary Figure:




Summary:

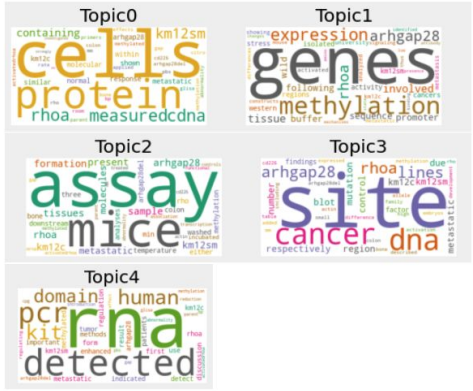
1. Prioritization approach: mutational burden
2. Downstream analysis: gene expression analysis and structural analysis
3. Findings:
   a. Comparison of gene expression levels of target genes indicates mutations affect tissue-specific genes and ubiquitous genes
   b. Protein structure illuminate the possible effects of the mutational burden, since several amino acid were changed in the structures. However, fortunately, none affected active sites or protein protein interfaces

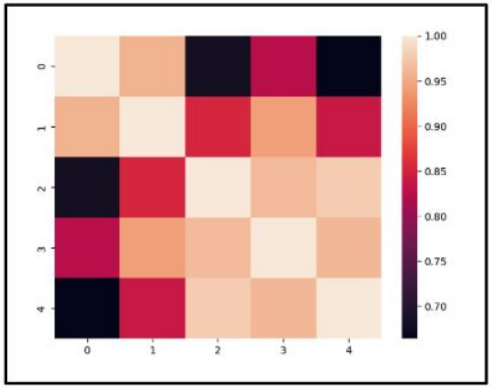
# 2022 group 3 (chr 18)

Top 10 Prioritized Genes

1. MBP
2. LDLRAD4
3. SMAD2
4. SMCHD1
5. PIGN
6. PIEZO2
7. ALPK2
8. NEDD4L
9. CD226
10. ARHGAP28

Summary Figure:




Summary:

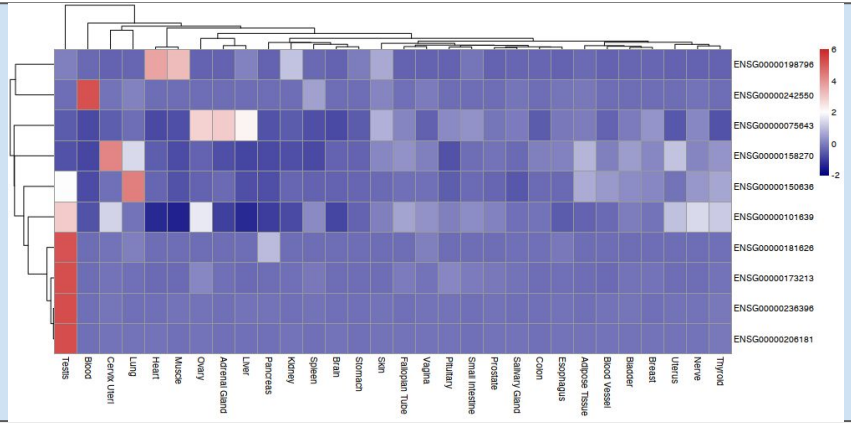
1. Prioritization approach: mutational burden
2. Downstream analysis: text mining analysis
3. Findings:
   a. NEDD4L has the highest correlation with the word hypertension
      i. It is involved in blood pressure maintenance
   b. CD226 has a correlation of 0.79 with the word autoimmune
      i. It is indeed associated with many autoimmune diseases including refractory celiac disease

# 2022 group 4 (chr 18)

Top 10 Prioritized Genes

1. ALPK2
2. CEP192
3. SERPINB10
4. ANKRD62
5. MOCOS
6. SLC35G4
7. COLEC12
8. ELOA2
9. TUBB8B
10. CCDC102B

Summary Figure:




Summary:

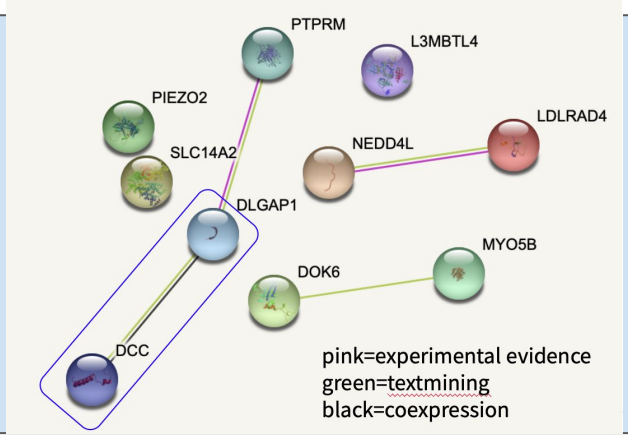
1. Prioritization approach: mutational burden of non-synonymous mutations
2. Downstream analysis: gene expression analysis
3. Findings:
   a. Most prioritized genes are expressed in heart and testes tissues
   b. The most saturated association in list is cancer/cell cycle regulation
   c. ALPK2 is associated with developmental processes primarily in heart
   d. Almost all mutations observed are in unstructured regions and not disruptive to the protein structure

# 2022 group 5 (chr 18)

Top 10 Prioritized Genes

1. DCC
2. DLGAP1
3. PIEZO2
4. PTPRM
5. AQP4-AS1
6. DOK6
7. MYO5B
8. LINC00907
9. L3MBTL4
10. DSEL-AS1

Summary Figure:




Summary:

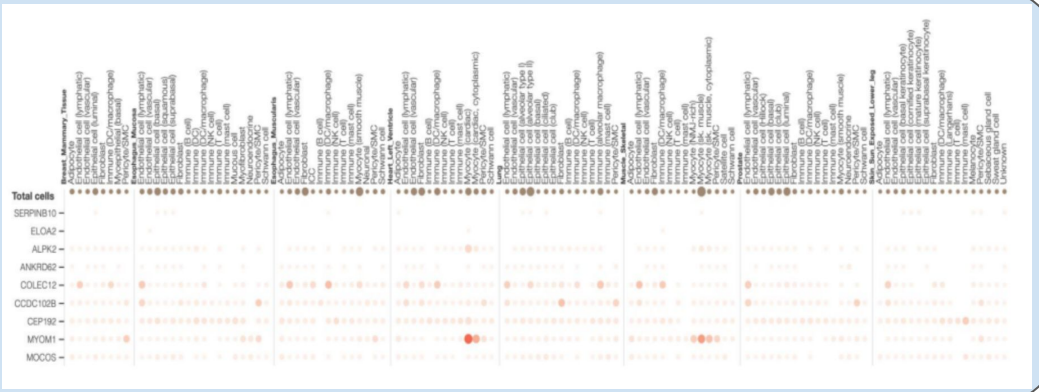
1. Prioritization approach: mutational burden
2. Downstream analysis: gene expression analysis and STRING network analysis
3. Findings:
   a. Some of these genes expressed in few tissues consistent with known function
   b. Differential expression may point to non-canonical functions: DCC in testis (tumor suppression) : PIEZO2 in lungs (stretch sensation)
   c. Mutated genes are pieces in larger pathways: disruption → downstream effects

# 2022 group 6 (chr 18)

Top 10 Prioritized Genes

1. ALPK2
2. CEP192
3. SERPINB10
4. ANKRD62
5. MOCOS
6. SLC35G4
7. COLEC12
8. ELOA2
9. TUBB8B
10. CCDC102B

Summary Figure:




Summary:

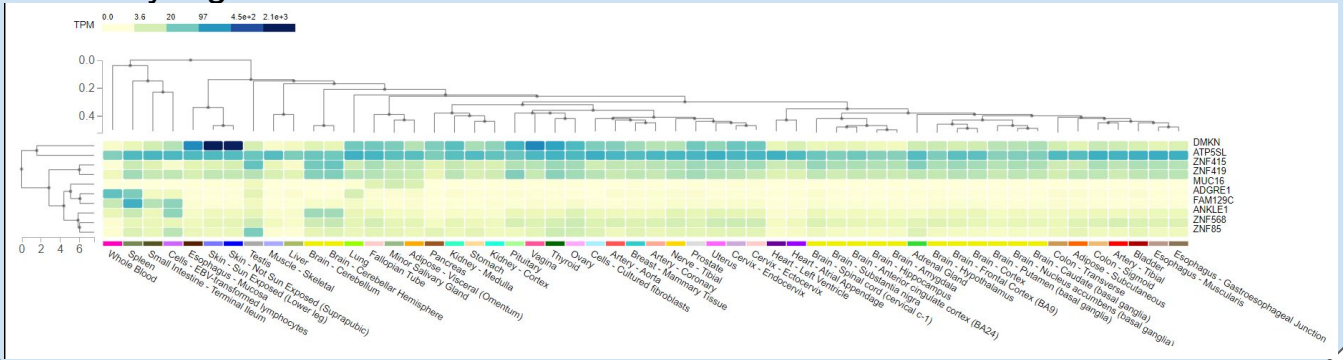
1. Prioritization approach: nonsynonymous exonic mutational burden
2. Downstream analysis: gene expression analysis (single cell analysis)
3. Findings:
   a. Gene expression profiles suggest that differentially expressed genes may associate with tissue-specific functions
   b. Most genes are associated with polymorphisms and are part of disease-proliferation pathways and mechanisms.

# 2023 group 1 (chr 19)

Top 10 Prioritized Genes

1. MUC16
2. ZNF85
3. ANF568
4. DMKN
5. ZNF419
6. ANKLE1
7. ZNF415
8. FAM129C
9. EMR1
10. ATP5SL

## Summary Figure:




## Summary:

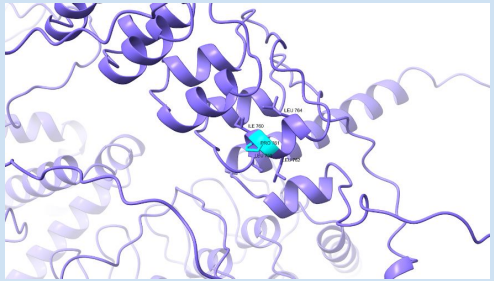
1. Prioritization approach: mutational burden, count number of high/moderate impact mutations on each gene
2. Downstream analysis: gene expression analysis
3. Findings:
   a. MUC16 has 71 moderate impact mutations
   b. MUC16 high expression in the Visceral adipose, cervix, salivary gland
   c. ZNF419 has 7 high impact mutations, most of any other genes
   d. ZNF419, ZNF569, and ZNF415 predict enabling DNA-Binding transcription factor activity

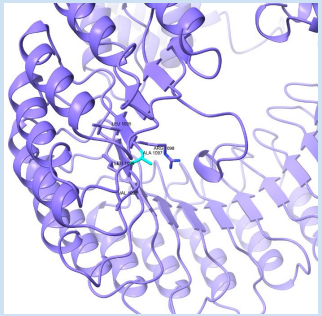
# 2023 group 4 (chr 19)

Top 10 Prioritized Genes

1. Perilipin-4
2. ZNF225
3. CYP2A7
4. ANKRD62
5. ZNF568
6. CYP2A7
7. MUC16
8. ZNF568
9. J3KP41
10. P06213

Summary Figure:





Summary:

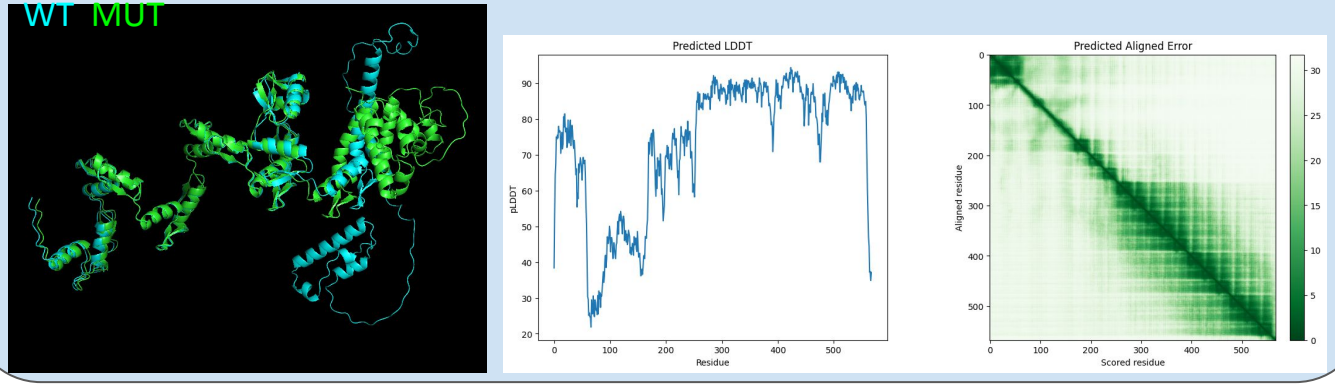
1. Prioritization approach: mutational burden
2. Downstream analysis: protein structure analysis
3. Findings: none

# 2023 group 2 (chr 19)

Top 10 Prioritized Genes

1. CACNA1A
2. VSTM2B-DT
3. ZNF675
4. INSR
5. ZNF536
6. LINC02987
7. MUC16
8. GNG7
9. ENSG00000269110
10. KDM4B

Summary Figure:




Summary:

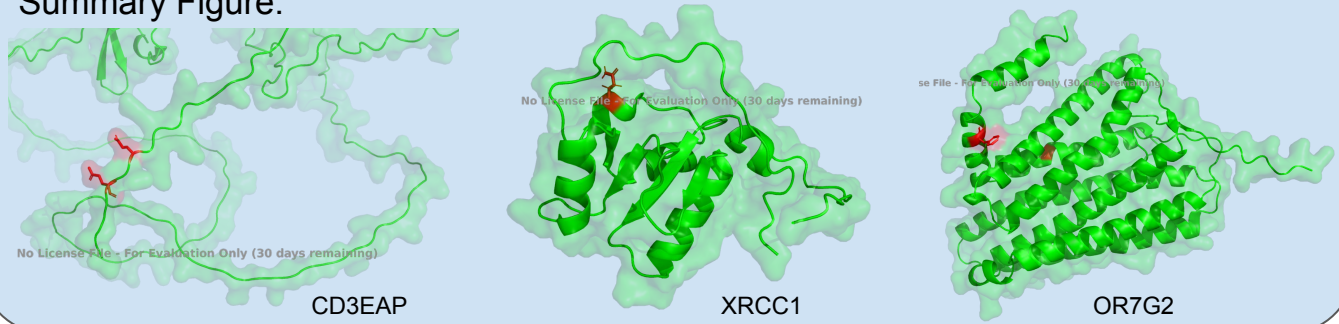
1. Prioritization approach: mutational burden
2. Downstream analysis: structure prediction of wildtype and mutant ZNF675 protein
3. Findings:
   a. Valine to alanine substitution (V197A) on exon 4 of ZNF675 causes a notable change in the protein structure
   b. ZNF675 is co-expressed with other zinc finger proteins that maybe involved in transcriptional regulation via the NFkB pathway.
   c. Low specificity of ZNF675 protein in tissue makes it difficult to characterize its physiologic role.

# 2023 group 3 (chr 19)

Top 10 Prioritized Genes

1. CD3EAP
2. XRCC1
3. OR7G2
4. OR10H5
5. LILRB4
6. KIR2DL3
7. KIR2DL1
8. KIR3DL2
9. ZFP28
10. GP6

Summary Figure:




Summary:

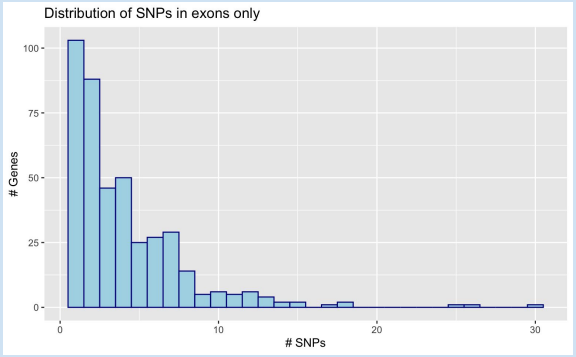
1. Prioritization approach: exonic mutational burden
2. Downstream analysis: protein structure analysis
3. Findings:
    a. Resistance to Cisplatin-based cancer treatment
    b. Dysregulation of certain immune cell subtypes ability to distinguish host and tumor cells
    c. Dysregulation of collagen-based platelet adhesion

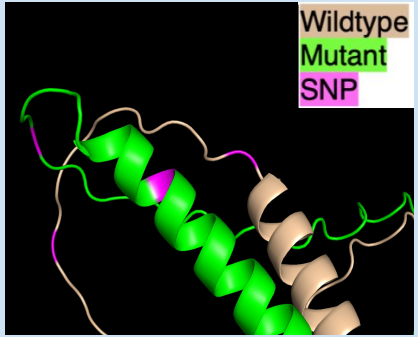
# 2023 group 5 (chr20)

Top 10 Prioritized Genes

1. CTSZ
2. MTG2
3. SLC52A3
4. PROKR2
5. HELZ2
6. LAMA5
7. TMC2
8. JPH2
9. KIAA1755
10. PTPRT

## Summary Figure:




Gene: MTG2

Mutation: P267L

## Summary:

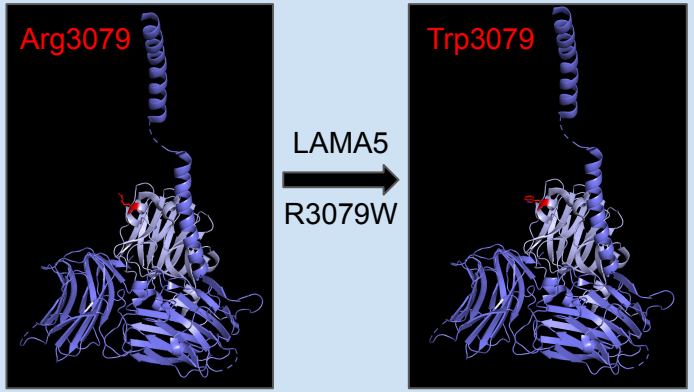
1. Prioritization approach: exonic SNP number
1. Downstream analysis: (1) Pymol visualization and global lDDT; (2) Local lDDT and AlphaFold2 sequence coverage analysis
2. Findings:
   a. Benign mutation P267L on MTG2 alters the secondary structure from loop to α-helix
   b. All 9 SNPs we studied are benign
   c. Loop regions tend to have low sequence coverage and local lDDT

# 2023 group 6 (chr 20)

Top 10 Prioritized Genes

1. LAMA5
2. HELZ2
3. JPH2
4. PTPRT
5. CDH4
6. DNAJC5
7. RAB22A
8. EU95_1860
9. 5LC52A3
10. SIGLEC2

Summary Figure:




Summary:

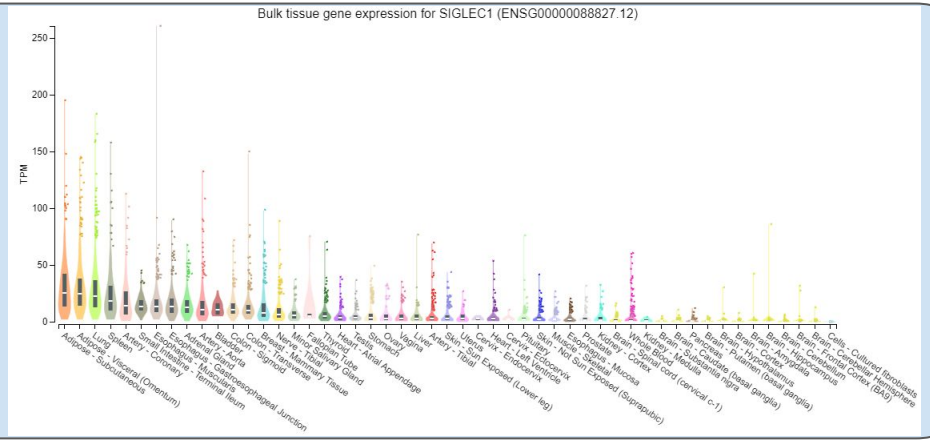
1. Prioritization approach: exonic mutational burden
2. Downstream analysis: protein structure analysis
3. Findings:
   a. No disease-associated variants identified
   b. No variants caused significant changes in protein structures
   c. Most genes with the highest mutational burden regulate cell structure and integrity
   d. Mutations in proteins were non-deleterious since they occurred in loops or did not disrupt important interfaces.

# 2023 group 7 (chr 20)

Top 10 Prioritized Genes

1. TAF4
2. FERMT1
3. MROH8
4. BPIFB4
5. KIAA1755
6. MYH7B
7. SIRPG
8. HELZ2
9. SIGLEC1
10. LAMA5

Summary Figure:




Summary:

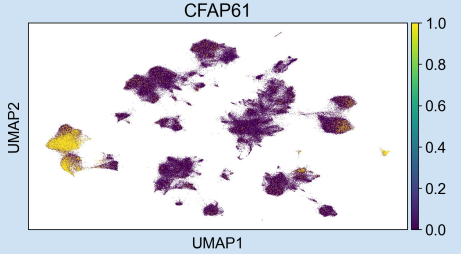
1. Prioritization approach: exonic mutational burden
2. Downstream analysis: tissue gene expression analysis
3. Findings:
   a. BPIFB4 is associated with longevity and decreased heart disease
   b. SIGLEC1 is an immune marker with high lung expression. Low expression is correlated with severe COVID-19 disease

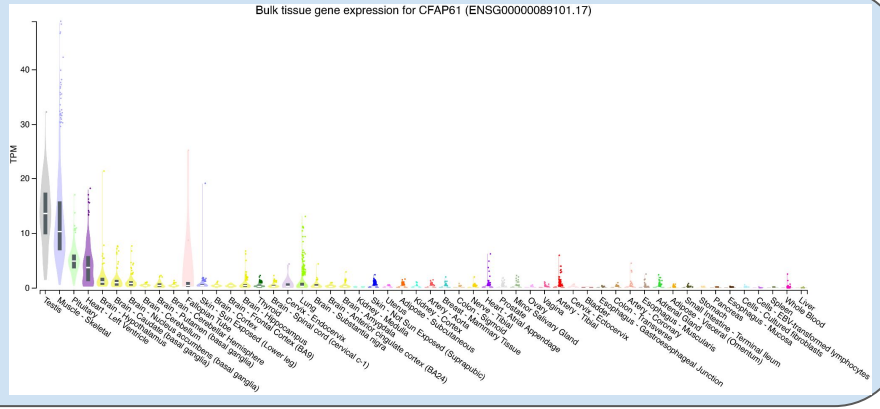
# 2023 group 8 (chr 20)

Top 10 Prioritized Genes

1. MACROD2
2. CDH4
3. PTPRT
4. PLCB1
5. SLC24A3
6. PAK5
7. CFAP61
8. EYA2
9. TSHZ2
10. SLX4IP

## Summary figure:




## Summary:

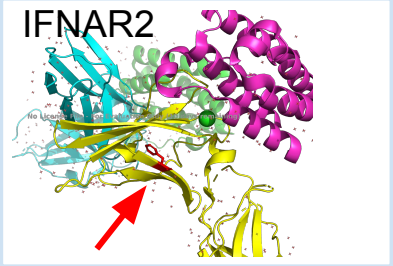
1. Prioritization approach: exonic and intronic mutational burden
2. Downstream analysis: gene expression at bulk and single-cell levels
3. Findings:
   a. We observed little correlation in bulk or single-cell data between mutational burden and specificity of expression by tissue
   b. scRNA-seq gives more detailed picture of how cells in a given tissue express mutationally burdened genes
   c. Further work is required to uncover functional significance of mutations we observed and how it relates to tissue expression

# 2023 group 9 (chr 21)

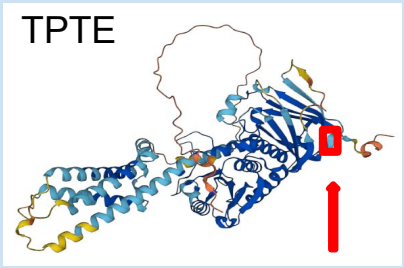
Top 10 Prioritized Genes

1. UMODL1
2. PCNT
3. TPTE
4. SON
5. LSS
6. LIPI
7. IFNAR2
8. WDR4
9. N6AMT1
10. COL6A2

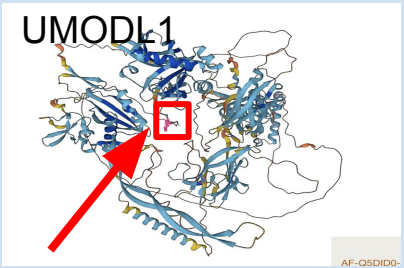
## Summary Figure:



Phenylalanine → Serine



Leucine → Proline



Threonine → Methionine

## Summary:

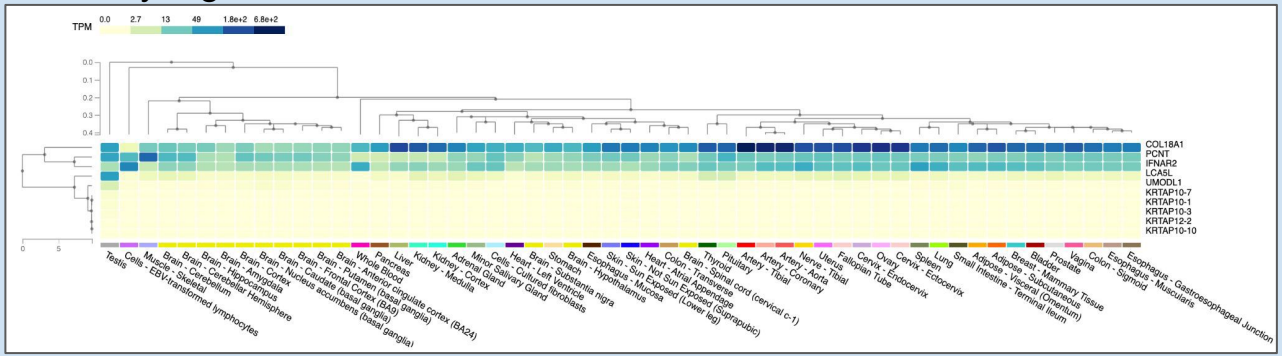
1. Prioritization approach: mutational burden
2. Downstream analysis: Compare the variant in Carl's genome with those in genomAD and 1000 genome databases
3. Findings:
   a. Structure of mutation found using AlphaFold
   b. Studies on mutations found using UNIPROT (most mutations did not have significant effect)

# 2023 group 10 (chr 21)

Top 10 Prioritized Genes

1. UMODL1
2. KRTAP12-2
3. KRTAP10-10
4. PCNT
5. KRTAP10-7
6. LCA5L
7. COL18A1
8. IFNAR2
9. KRTAP10-1
10. KRTAP10-3

## Summary Figure:




## Summary:

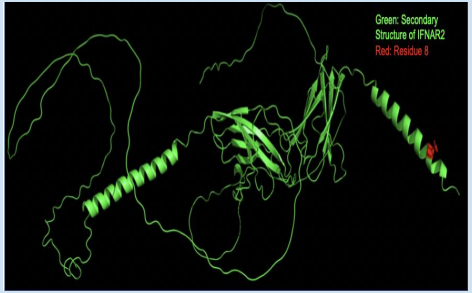
1. Prioritization approach: Non-synonymous mutations
2. Downstream analysis: Gene expression analysis
3. Findings:
   a. Gene expression levels in tissues seem to be highly correlated with function
   b. Protein-coding mutations in UMODL1 have been associated with various conditions like myopia or effects on hormones
   c. PCNT has shown to be a positive association betwee non-synonymous mutations in mood disorders.

# 2023 group 11 (chr 21)

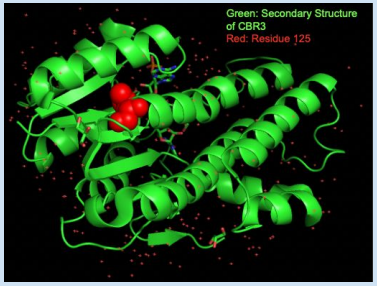
Top 10 Prioritized Genes

1. IFNAR2
2. IL10RB
3. CBR3
4. BACH1
5. KRTAP 10-1
6. KRTAP 10-11
7. TPTE
8. ADAMTS
9. KRTAP 13-2
10. KRTAP 10-6

## Summary Figure:








## Summary:

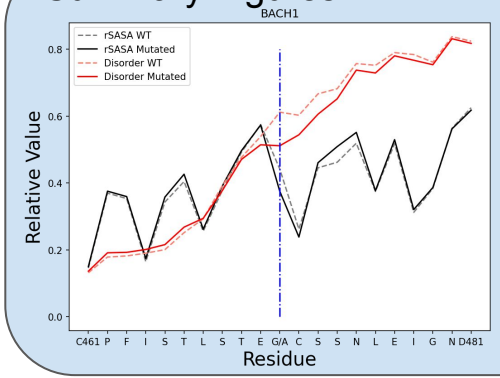
1. Prioritization Approach: Identifying protein-encoding genes with highest mutational burden on chromosome 21 of Carl Zimmer
2. Downstream Analysis: Protein Structural Analysis on Genes with High Mutational Burden and show clinical significance with PyMol
3. Findings
   a. Three clinically significant mutations were observed on chromosome 21 of Carl Zimmer
   b. Mutation(s) of IFNAR2 and IL10RB is expected to increase susceptibility to viruses
   c. Mutation of CBR3 is expected to pose no functional consequence

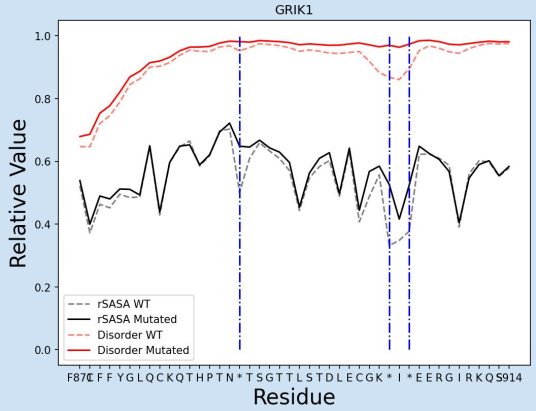
# 2023 group 12 (chr 21)

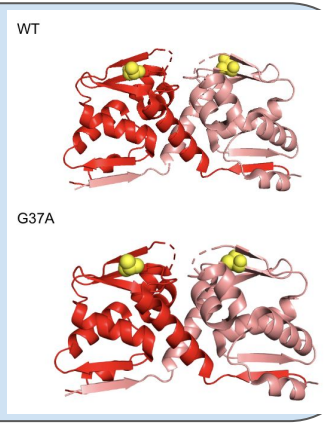
Top 10 Prioritized Genes

1. DSCAM
2. RUNX1
3. NCAM2
4. KCNJ6
5. TSPEAR
6. GRIK1
7. ERG
8. APP
9. CHODL
10. BACH1

## Summary Figures:




## Summary:

1. Prioritization approach: protein coding genes with highest mutational burden
2. Downstream analysis: protein structure (rSASA) analysis with NetSurfP
3. Findings:
   a. APP causes protein aggregates that are well-linked to Alzheimer's: found no protein coding variants GRIK1 L902S has a correlation with ADHD
      - it is a cationic channel in the Cerebellum and hypothalamus; binds to excitatory neurotransmitter L-glutamate