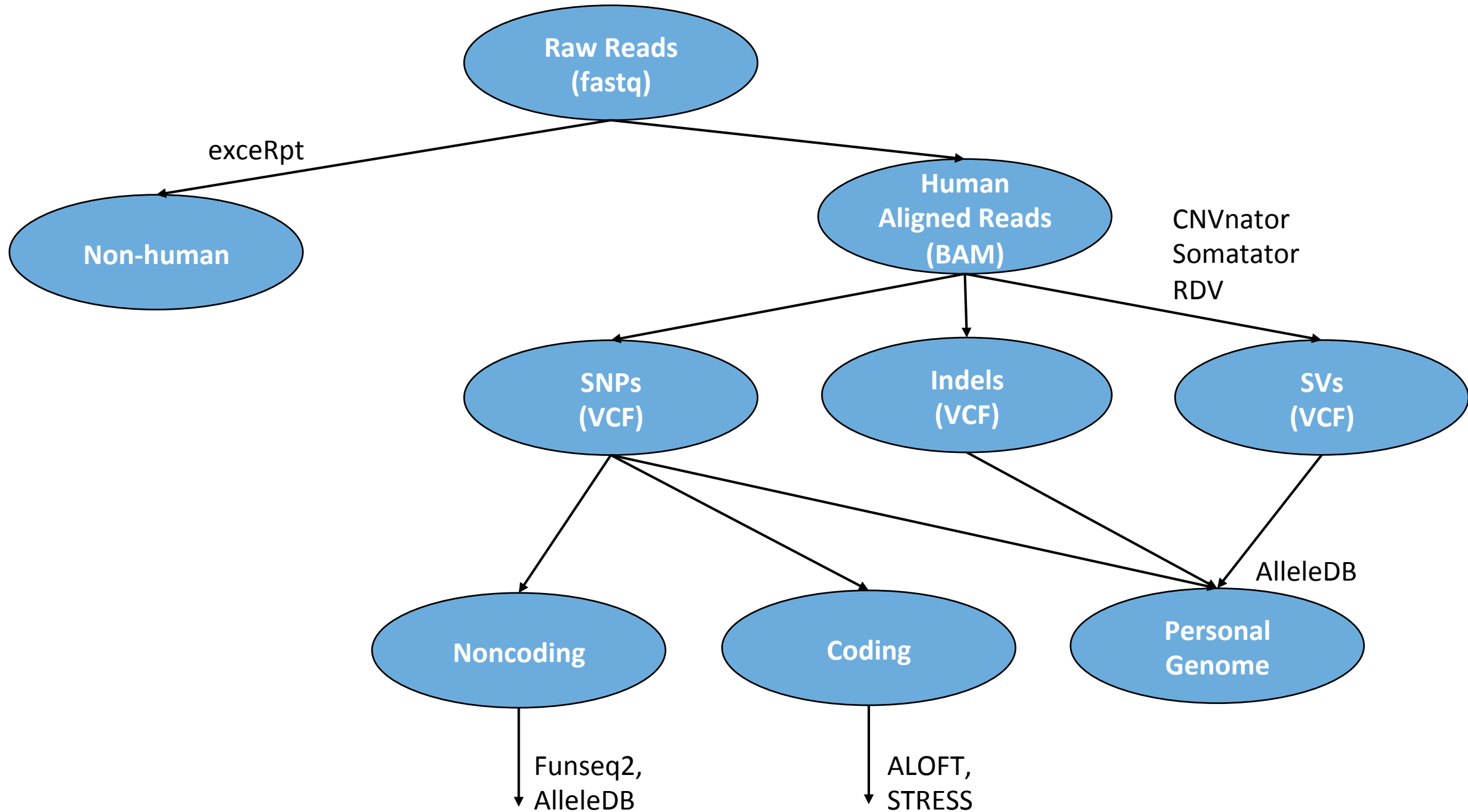


Coding Variation in Subject Z

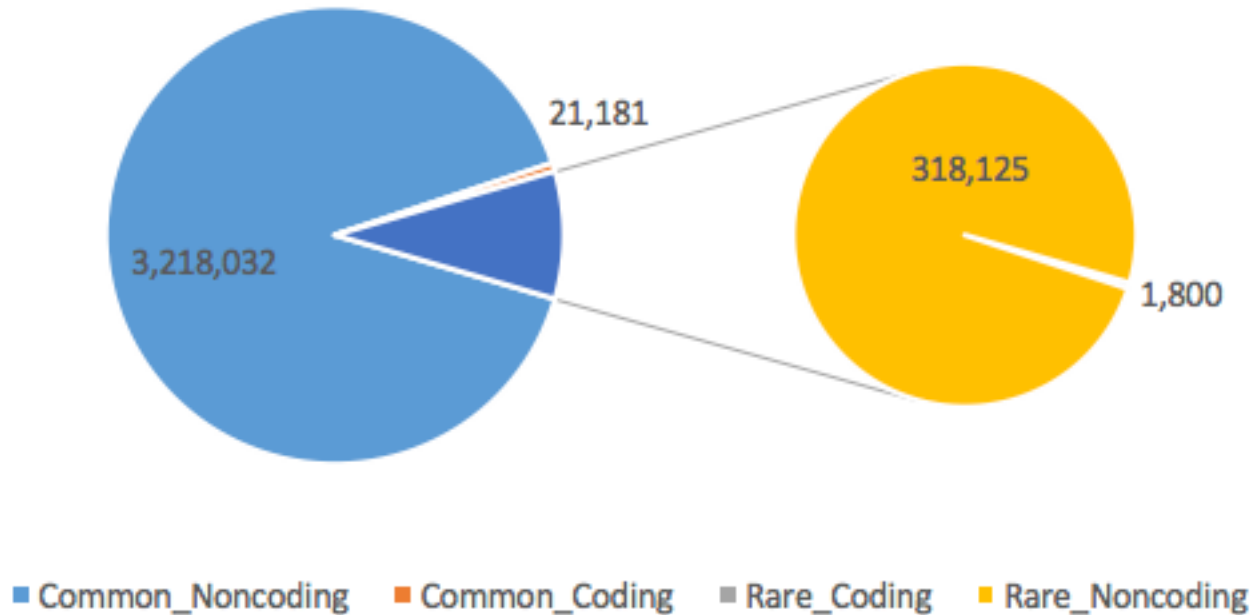
With applications using tools developed in the lab:
STRESS, ALOFT

March 1, 2016

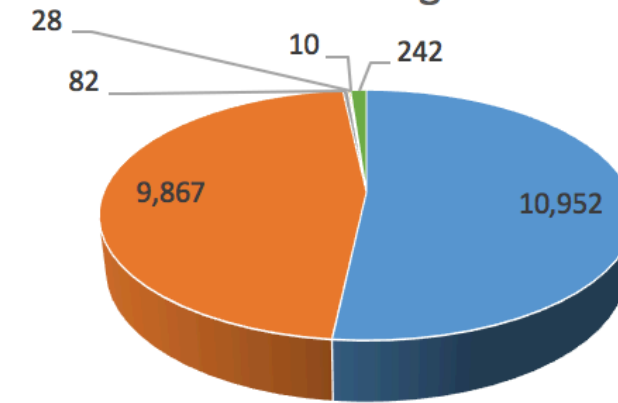


Overview & Coding Variants

SNVs of Individual Z

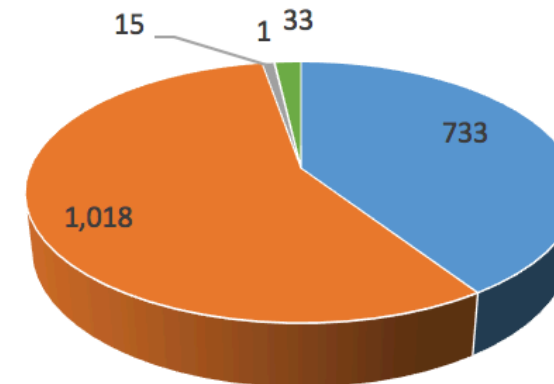


Common Coding Variants



- Synonymous ■ Nonsynonymous ■ PrematureStop
- RemovedStop ■ SpliceOverlap ■ NA

Rare Coding Variants



Rare Non-synonymous Coding Variants

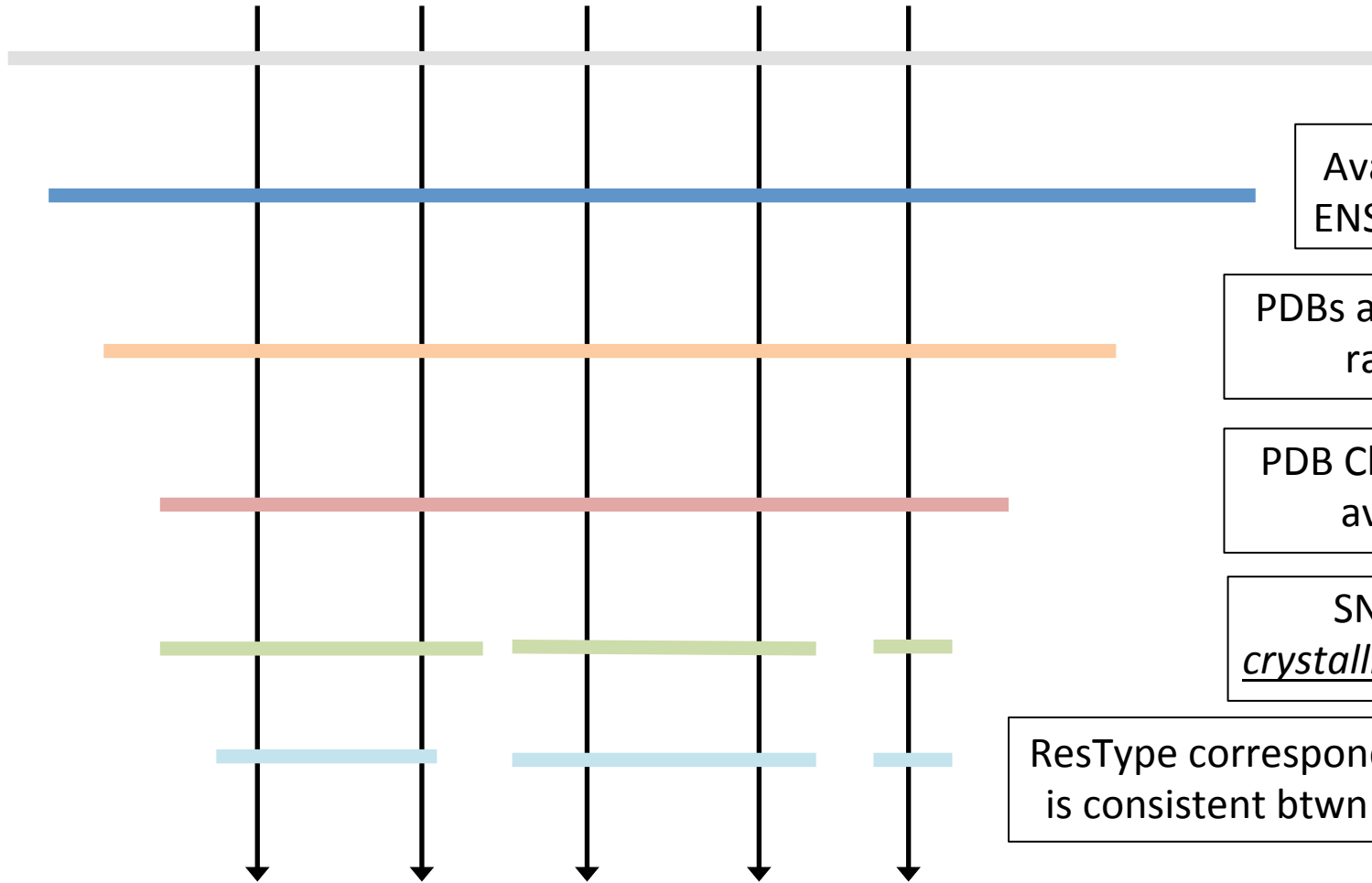
- 1018 SNVs -> **824** target genes

Gene Annotation	Gene Name
Cancer-related	NOTCH2; PDE4DIP; TPR; CRTC3; CDH11; MLLT6; ASXL1; HMGA1; KDM6A
DNA repair	RECQL; RAD51; PPM1D; XRCC1; AP1B1; FANCI; PTPRH; RBBP7; SLX4; POLR2A; DCLRE1C; ANKLE1
Cancer & DNA repair	ATM; PMS2; ERCC5
Actionable Gene	ATM; KDM6A; INSR; FOXP4

- **ATM**: Serine/Threonine Kinase; Regulator of **p53** and **BRCA1**; leukemia; ataxia-telangiectasia; breast cancer
- **PMS2**: Direct **p53** effectors; mismatch repair cancer syndrome; colorectal cancer; hereditary nonpolyposis
- **ERCC5**: Chks in Checkpoint Regulation; DNA Repair; xeroderma pigmentosum
- **KDM6A**: Transcriptional misregulation in cancer
- **INSR**: **Insulin Receptor**; PI3K-Akt signaling pathway; GPCR Pathway; Diabetes mellitus
- **FOXP4**: **Transcriptional repressor** that represses lung-specific expression

At least one SNV from

Subj. Z Subj. S na12878 1KG HGMD



Avail PDB IDs (given ENSTs) from BioMart

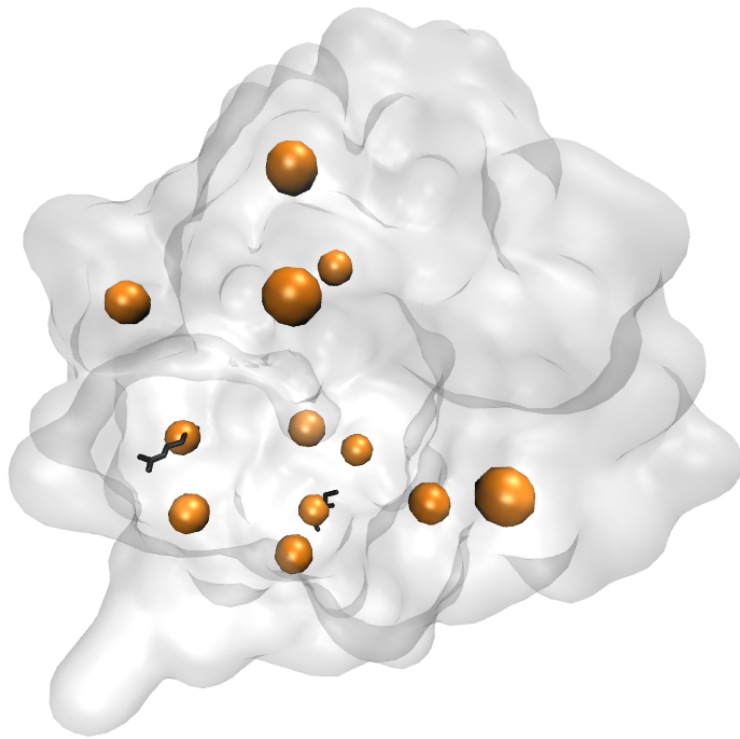
PDBs are high-quality X-ray structures

PDB Chains are known/avail (UniProt)

SNVs hit within crystallized region of PDB

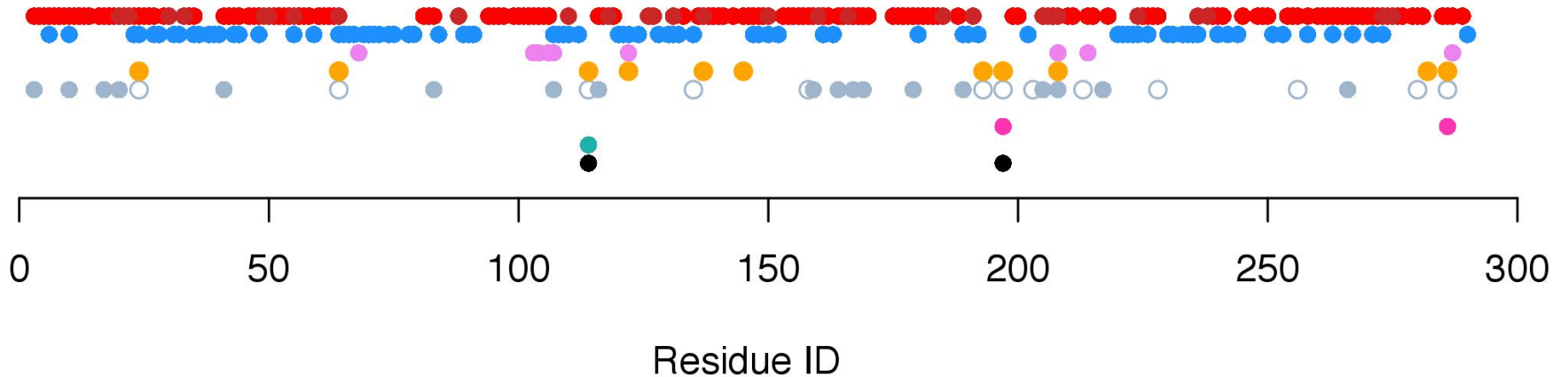
ResType corresponding to wt residue is consistent btwn VAT file and PDB

(remove redundant proteins)

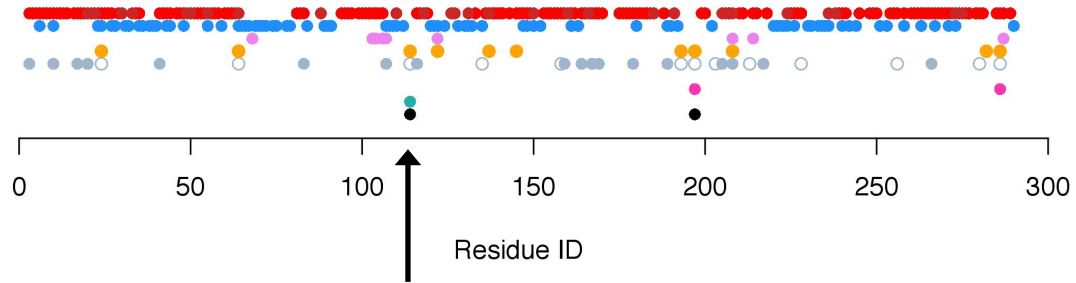


- Predicted allosteric (surface | interior)
- Buried residues
- Protein-protein interaction site
- Post-translational modifications
- ▲ ● HGMD (prem. stop | non-synon)
- ○ 1000 Genomes (rare | common)
- ▲ ● Snyder (prem. stop | non-synon)
- ▲ ● na12878 (prem. stop | non-synon)
- ▲ ● Subj. Z (prem. stop | non-synon)

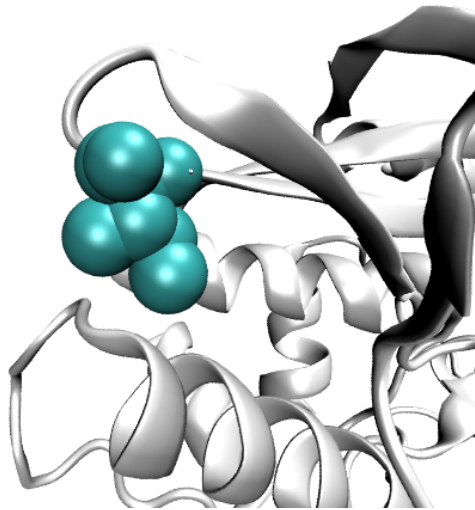
Arylamine N-acetyltransferase 2 (2PFR_A: gene = NAT2)



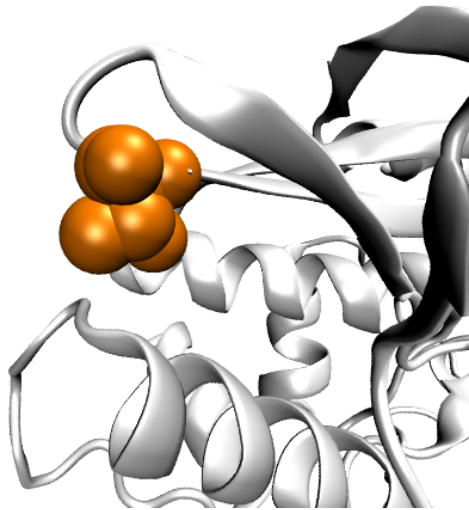
Arylamine N-acetyltransferase 2 (2PFR_A: gene = NAT2)



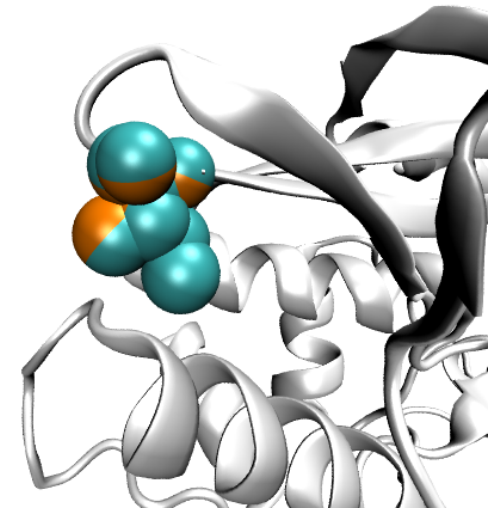
114: I->T



Wild-type

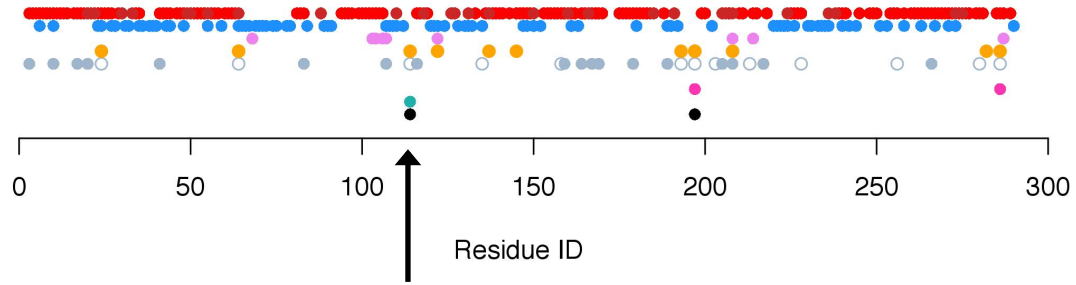


Mutated

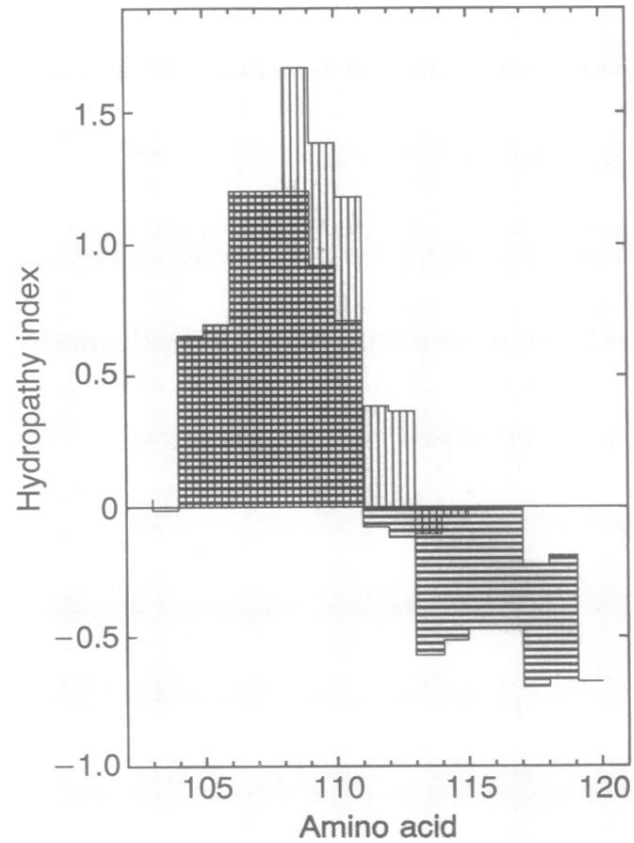


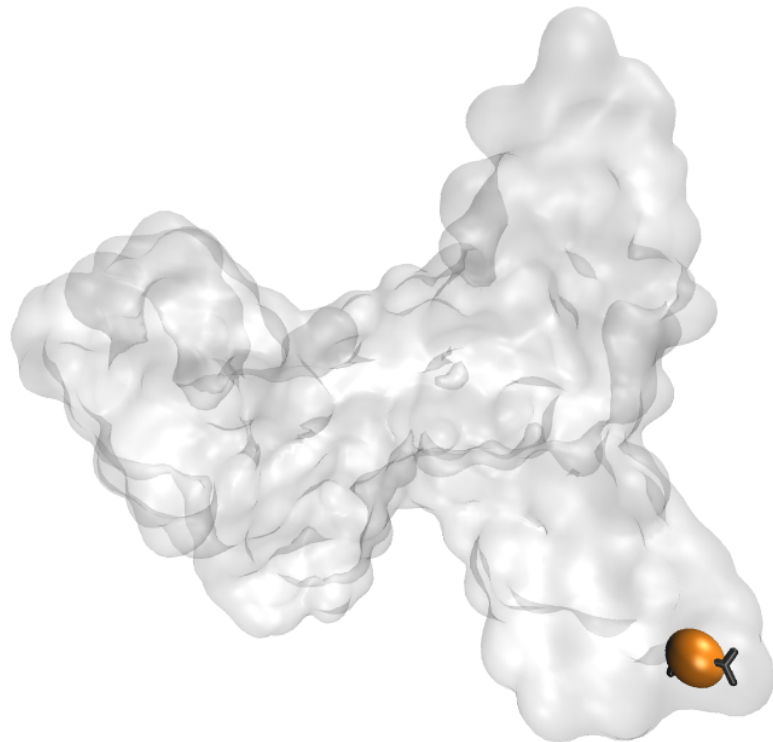
(superimposed)

Arylamine N-acetyltransferase 2 (2PFR_A: gene = NAT2)



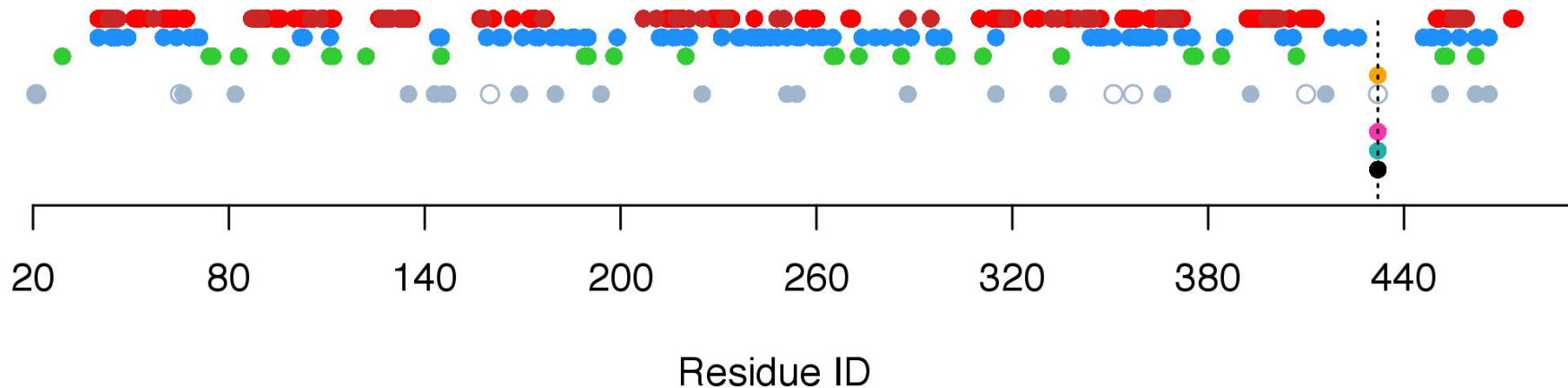
114: I->T



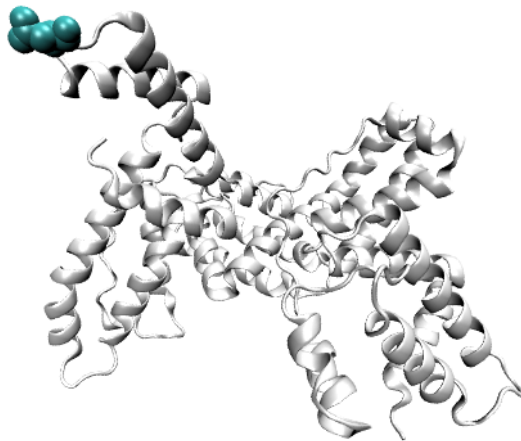
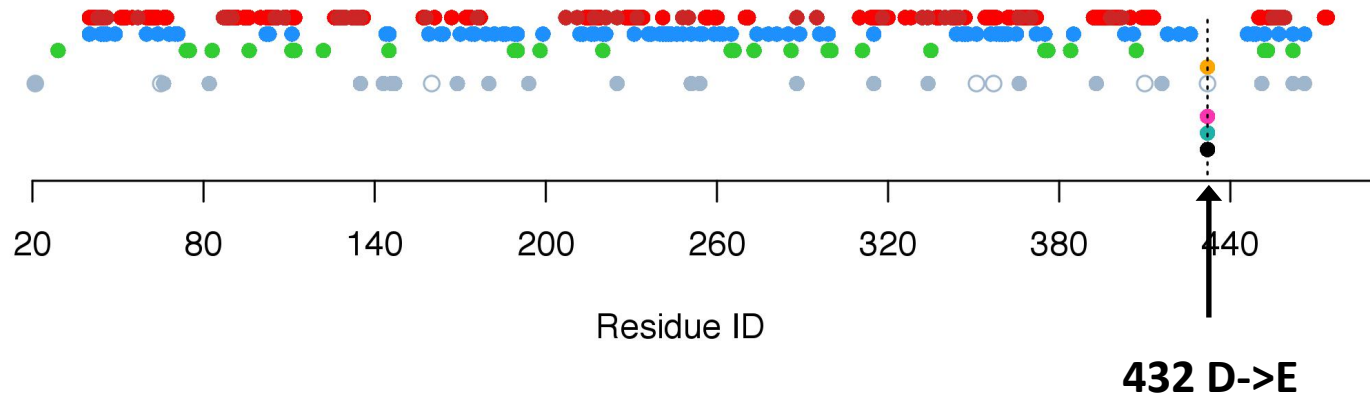


- Predicted allosteric (surface | interior)
- Buried residues
- Protein-protein interaction site
- Post-translational modifications
- ▲ ● HGMD (prem. stop | non-synon)
- ○ 1000 Genomes (rare | common)
- ▲ ● Snyder (prem. stop | non-synon)
- ▲ ● na12878 (prem. stop | non-synon)
- ▲ ● Subj. Z (prem. stop | non-synon)

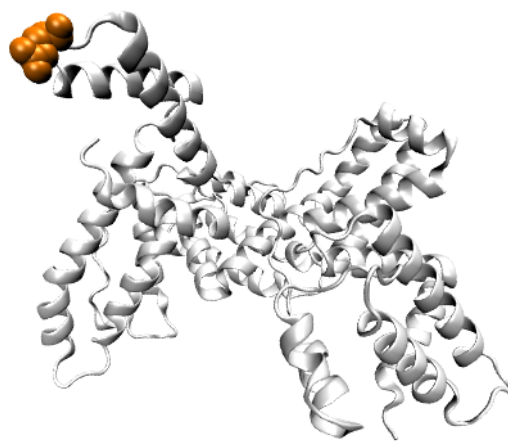
Vitamin D-binding protein (1KW2_A: gene = GC)



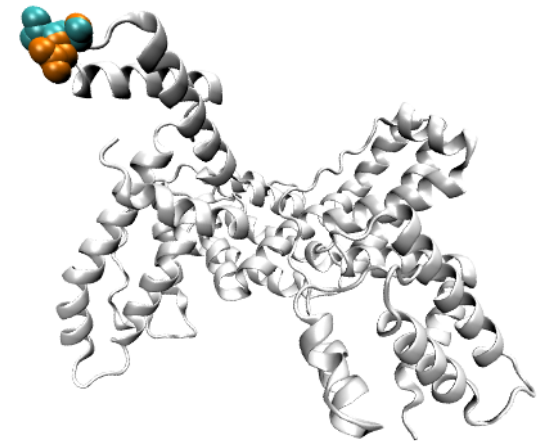
Vitamin D-binding protein (1KW2_A: gene = GC)



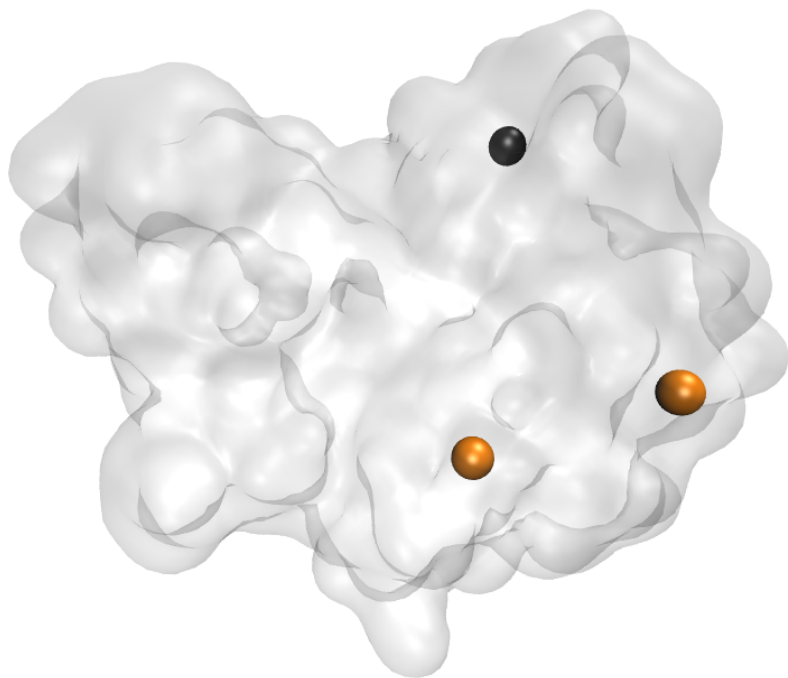
Wild-type



Mutated

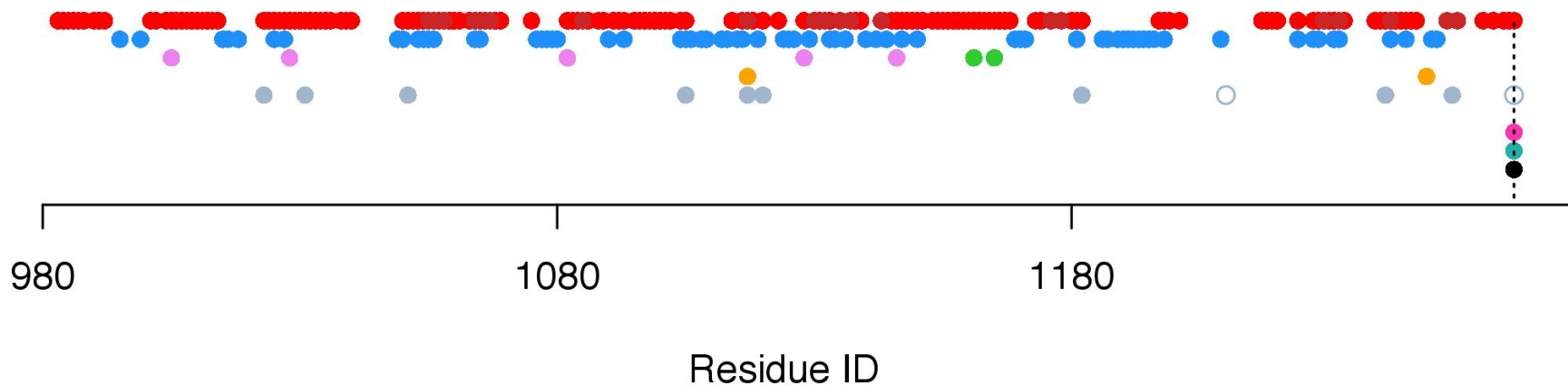


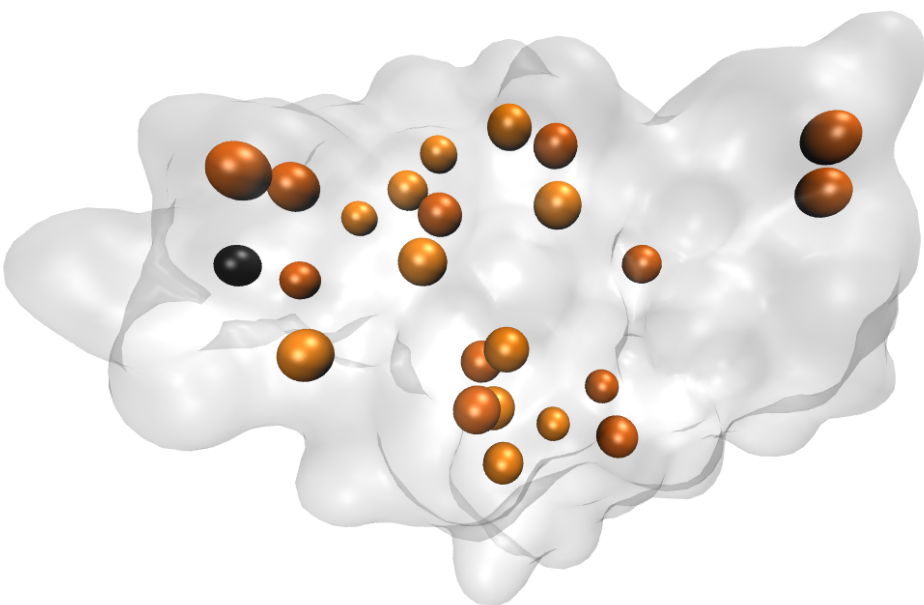
(superimposed)



- Predicted allosteric (surface | interior)
- Buried residues
- Protein-protein interaction site
- Post-translational modifications
- ▲ ● HGMD (prem. stop | non-synon)
- ○ 1000 Genomes (rare | common)
- ▲ ● Snyder (prem. stop | non-synon)
- ▲ ● na12878 (prem. stop | non-synon)
- ▲ ● Subj. Z (prem. stop | non-synon)

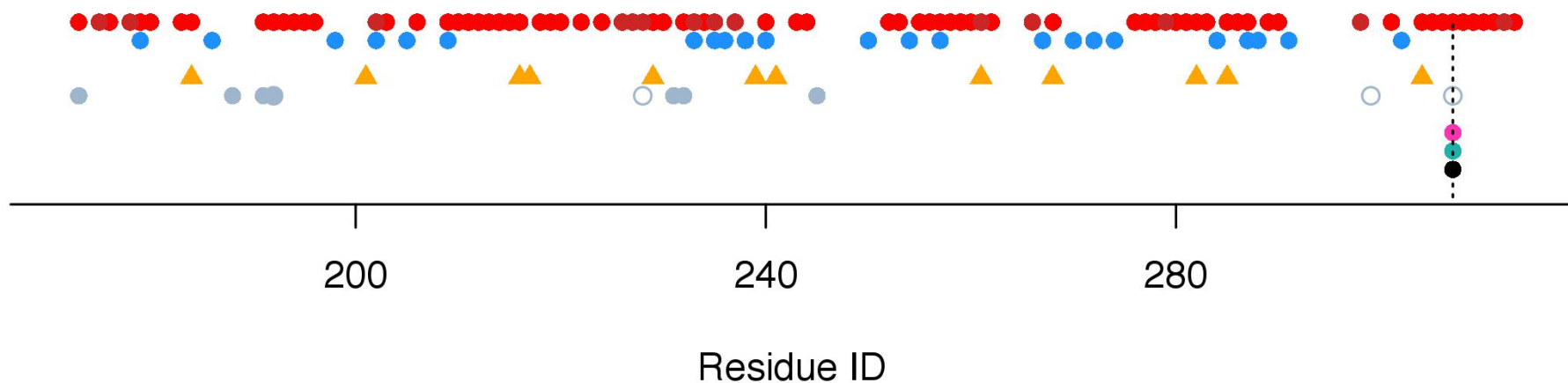
**PAS domain-containing
serine/threonine-protein
kinase (3DLS_A: gene = PASK)**



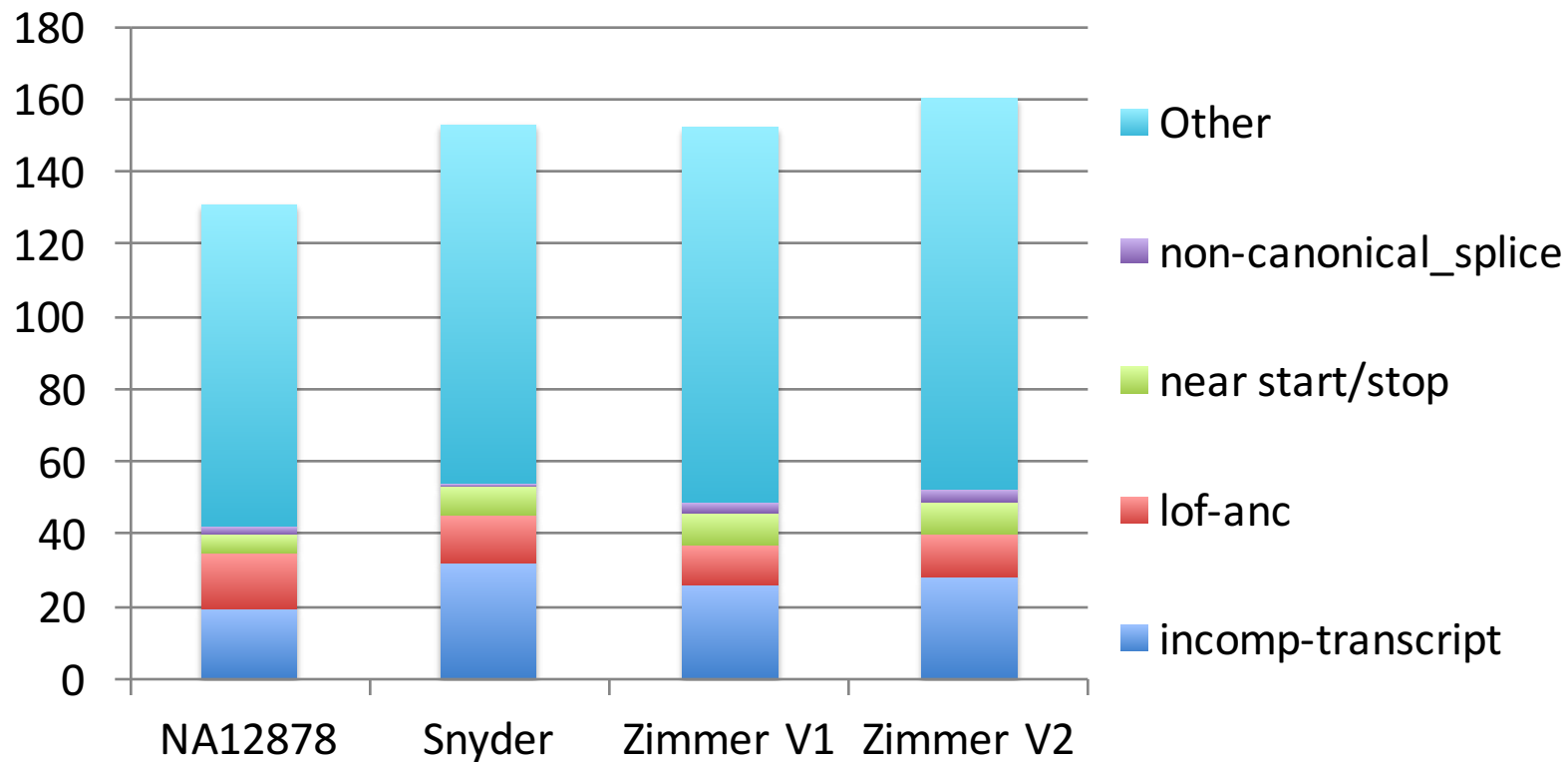


- Predicted allosteric (surface | interior)
- Buried residues
- Protein-protein interaction site
- Post-translational modifications
- ▲ HGMD (prem. stop | non-synon)
- 1000 Genomes (rare | common)
- ▲ Snyder (prem. stop | non-synon)
- ▲ na12878 (prem. stop | non-synon)
- ▲ Subj. Z (prem. stop | non-synon)

Tpr Domain of Aip (4APO_A: gene = AIP)



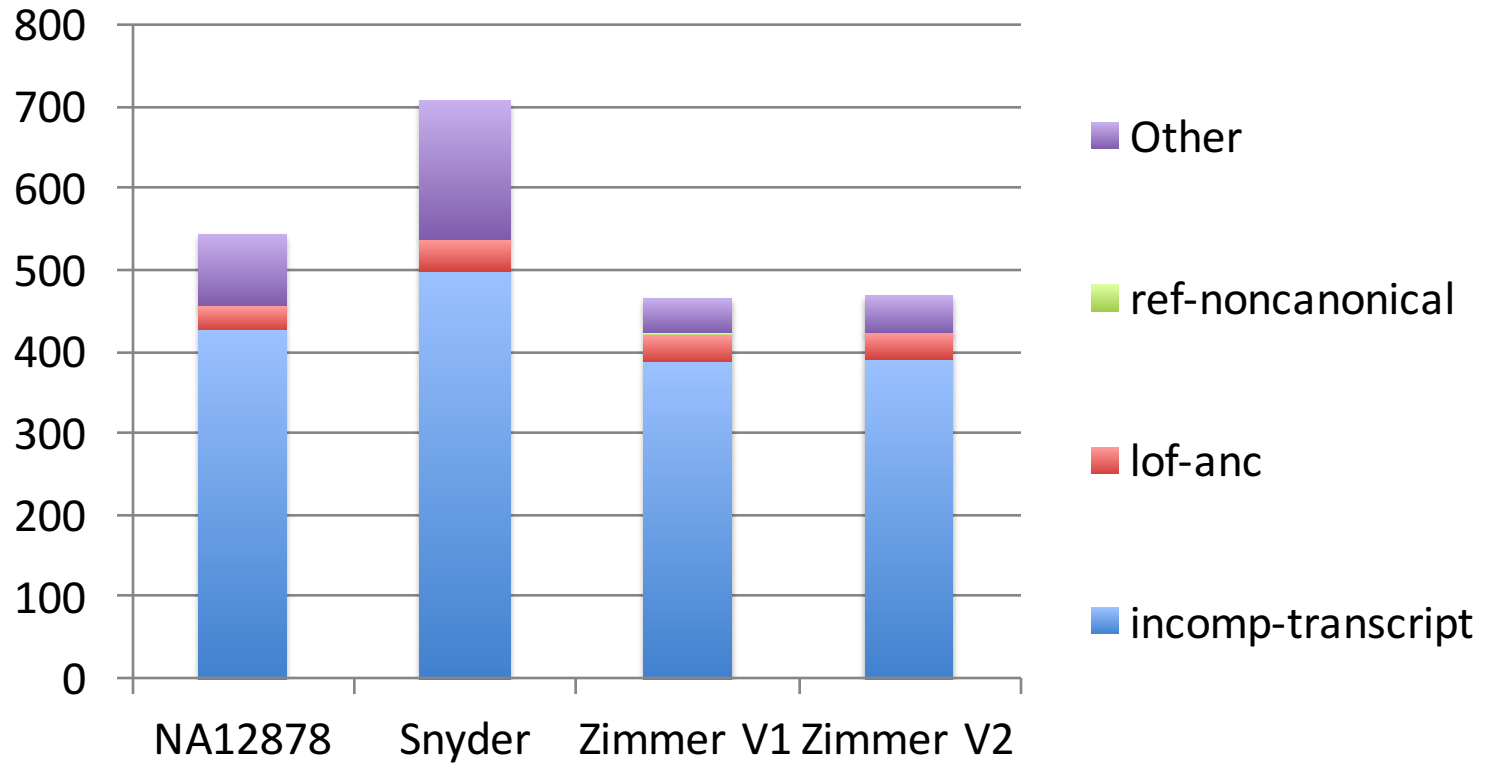
- **Transcript-level Flags**
- **LoF**



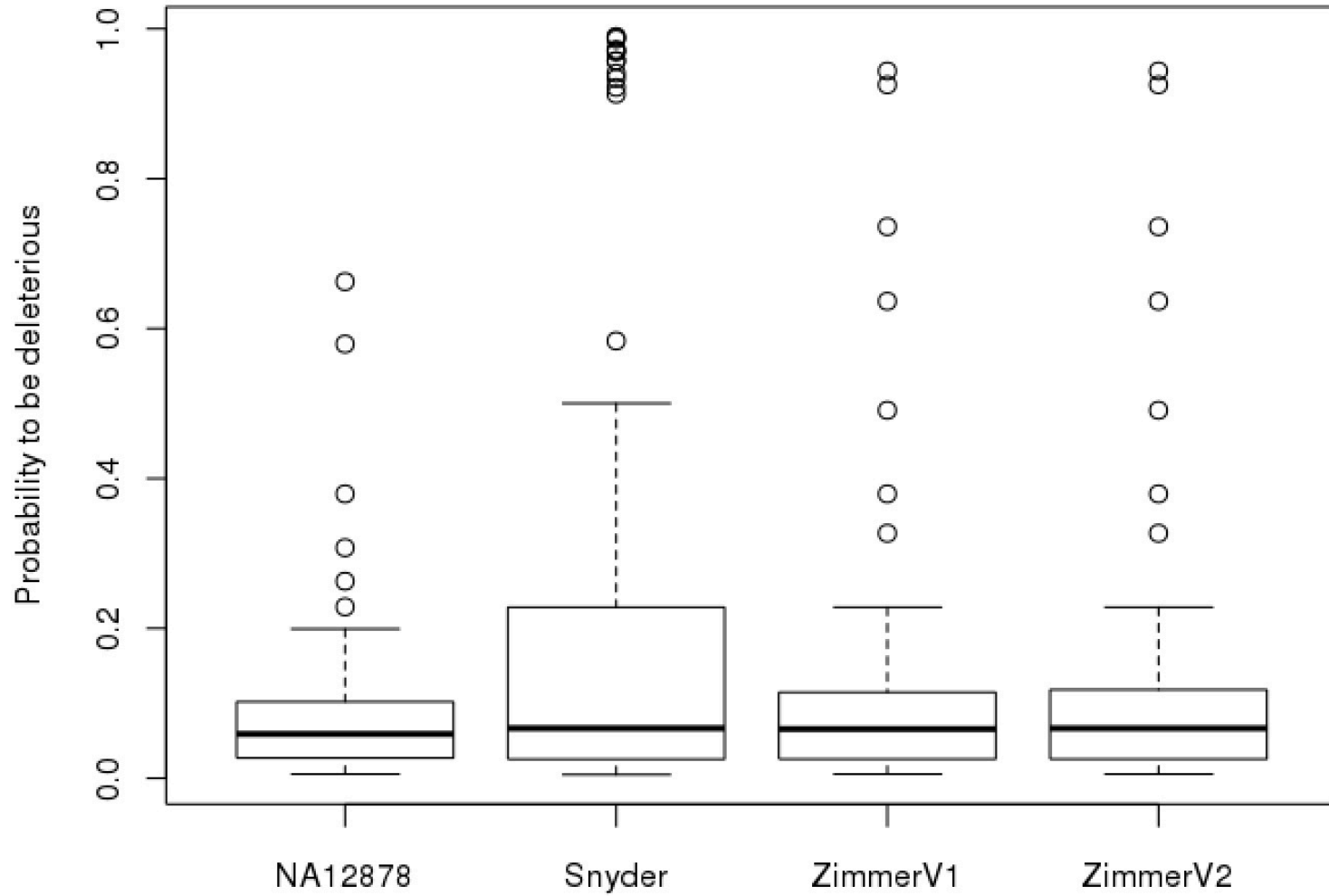
	Other
NA12878	89 (56 NMD)
Snyder	99 (67 NMD)
Zimmer V1	103 (58 NMD)
Zimmer V2	108 (63 NMD)

Transcript-level Flags

Splice



LOF variants



Potential deleterious variant

Zimmer

chr	pos	ref	alt	gene	Score	genotype	Gene function
6	17606162	C	T	FAM8A1	0.94365	0/1	Unknown, Autism related ??? Pubmed: 22495306
6	155577717	T	A	TIAM2	0.63655	0/1	Cell migration
17	61829719	A	C	CCDC47	0.92540	0/1	unknown
19	759925	C	A	MISP	0.73605	0/1	Mitotic spindle positioning

No disease associations in OMIM

CCDC47 associated with Schizophrenia

Potential deleterious variant

Snyder something is wrong ... (metabolism ???)

chr	pos	ref	alt	gene	Score	genotype	OMIM
2	44079970	C	A	ABCG8	0.92190	0/1	Sitosterolemia
2	215854316	T	A	ABCA12	0.97240	0/1	Ichthyosis
2	216240022	G	T	FN1	0.98975	0/1	fibronectin deficiency
9	111718091	G	T	CTNNAL1	0.98845	0/1	
9	130635074	G	T	AK1	0.96915	0/1	Hemolytic anemia
10	29581479	C	A	LYZL1	0.58365	0/1	
11	64056777	C	A	GPR137	0.94075	0/1	
12	18800840	G	T	PIK3C2G	0.95735	0/1	
12	122400030	C	A	WDR66	0.93380	0/1	
14	71570264	C	A	PCNX	0.98635	0/1	
15	68504073	G	T	CLN6	0.97080	0/1	Ceroid lipofuscinosiss
15	93007504	C	A	ST8SIA2	0.91290	0/1	
20	5157344	C	A	CDS2	0.95755	0/1	

- VEP annotation -> ALoFT run
- # of mutations

	# LOF (without flags)	Splice (without flags)
NA12878	80 (49)	308 (28)
Snyder	91 (53)	357 (57)
Zimmer V1	86 (55)	302 (24)
Zimmer V2	90 (57)	310 (26)

- # of genes

	# LOF (without flags)	Splice (without flags)
NA12878	80 (49)	275 (29)
Snyder	89 (51)	319 (56)
Zimmer V1	85 (54)	268 (25)
Zimmer V2	88 (55)	274 (27)

Current Gene List: z.motif.list

Current Background: Homo sapiens

11 DAVID IDs

+ **Options**

Rerun Using Options

Create Sublist

1 chart records

 [Download File](#)

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Benjamini
<input type="checkbox"/>	KEGG_PATHWAY	p53 signaling pathway	RT		2	18.2	6.5E-2	4.5E-1

9 gene(s)

from your list are not in the output

Current Background: Homo sapiens
702 DAVID IDs

Options

Rerun Using Options Create Sublist

[Download File](#)

231 chart records

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Benjamin
<input type="checkbox"/>	SP_PIR_KEYWORDS	polymorphism	RT		556	79.2	5.1E-32	2.5E-29
<input type="checkbox"/>	SP_PIR_KEYWORDS	alternative splicing	RT		338	48.1	3.8E-8	9.3E-6
<input type="checkbox"/>	SP_PIR_KEYWORDS	glycoprotein	RT		215	30.6	6.7E-8	1.1E-5
<input type="checkbox"/>	INTERPRO	EGF-like region, conserved site	RT		32	4.6	9.6E-8	9.4E-5
<input type="checkbox"/>	GOTERM_BP_FAT	cell adhesion	RT		54	7.7	1.1E-7	2.2E-4
<input type="checkbox"/>	GOTERM_BP_FAT	biological adhesion	RT		54	7.7	1.1E-7	1.1E-4
<input type="checkbox"/>	GOTERM_MF_FAT	extracellular matrix structural constituent	RT		16	2.3	4.3E-7	2.7E-4
<input type="checkbox"/>	SP_PIR_KEYWORDS	cell adhesion	RT		38	5.4	5.5E-7	6.6E-5
<input type="checkbox"/>	GOTERM_CC_FAT	plasma membrane	RT		180	25.6	2.1E-6	8.3E-4
<input type="checkbox"/>	PIR_SUPERFAMILY	PIRSF038286:preferentially expressed antigen of melanoma (PRAME) family protein	RT		8	1.1	3.3E-6	7.5E-4
<input type="checkbox"/>	SP_PIR_KEYWORDS	disulfide bond	RT		149	21.2	4.8E-6	4.7E-4
<input type="checkbox"/>	GOTERM_BP_FAT	homophilic cell adhesion	RT		17	2.4	1.5E-5	9.6E-3
<input type="checkbox"/>	GOTERM_BP_FAT	cell-cell adhesion	RT		26	3.7	1.5E-5	7.6E-3
<input type="checkbox"/>	INTERPRO	EGF-like, type 3	RT		21	3.0	2.7E-5	1.3E-2
<input type="checkbox"/>	INTERPRO	MAGE protein	RT		9	1.3	2.8E-5	9.2E-3
<input type="checkbox"/>	KEGG_PATHWAY	Graft-versus-host disease	RT		9	1.3	4.0E-5	5.3E-3
<input type="checkbox"/>	SP_PIR_KEYWORDS	tumor antigen	RT		9	1.3	4.9E-5	4.0E-3
<input type="checkbox"/>	SP_PIR_KEYWORDS	egf-like domain	RT		22	3.1	8.7E-5	6.0E-3
<input type="checkbox"/>	INTERPRO	Neuroblastoma breakpoint family	RT		5	0.7	1.1E-4	2.6E-2
<input type="checkbox"/>	GOTERM_BP_FAT	cellular component morphogenesis	RT		30	4.3	1.7E-4	6.6E-2
<input type="checkbox"/>	PIR_SUPERFAMILY	PIRSF003152:G protein-coupled olfactory receptor, class II	RT		26	3.7	2.8E-4	3.2E-2
<input type="checkbox"/>	PIR_SUPERFAMILY	PIRSF800006:rhodopsin-like G protein-coupled receptors	RT		41	5.8	3.1E-4	2.3E-2
<input type="checkbox"/>	GOTERM_BP_FAT	sensory perception of smell	RT		31	4.4	3.1E-4	9.7E-2
<input type="checkbox"/>	SP_PIR_KEYWORDS	coiled coil	RT		102	14.5	3.2E-4	1.9E-2
<input type="checkbox"/>	GOTERM_BP_FAT	cell morphogenesis	RT		27	3.8	3.7E-4	1.0E-1
<input type="checkbox"/>	GOTERM_BP_FAT	sensory perception of chemical stimulus	RT		33	4.7	4.0E-4	9.3E-2
<input type="checkbox"/>	SP_PIR_KEYWORDS	olfaction	RT		30	4.3	4.0E-4	2.1E-2
<input type="checkbox"/>	KEGG_PATHWAY	Olfactory transduction	RT		27	3.8	4.3E-4	2.8E-2
<input type="checkbox"/>	KEGG_PATHWAY	Antigen processing and presentation	RT		11	1.6	4.7E-4	2.1E-2
<input type="checkbox"/>	PIR_SUPERFAMILY	PIRSF005491:tumor associated protein MAGE	RT		6	0.9	5.2E-4	2.9E-2

Current Gene List: z.premature.list**Current Background: Homo sapiens****86 DAVID IDs****Options**

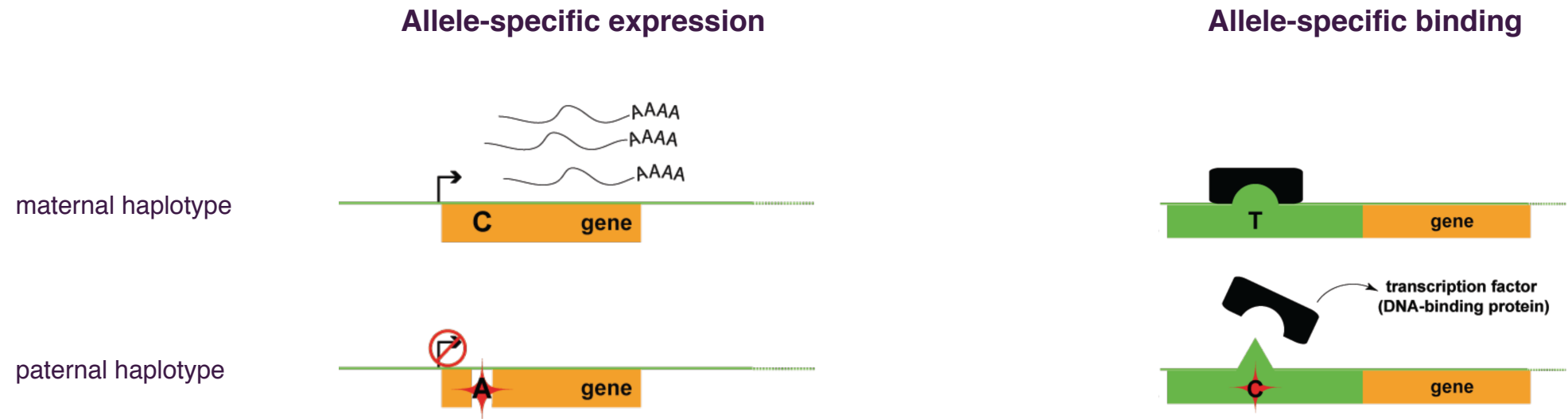
Rerun Using Options

Create Sublist

43 chart records[Download File](#)

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Benjamini
<input type="checkbox"/>	PIR_SUPERFAMILY	PIRSF800006:rhodopsin-like G protein-coupled receptors	RT		12	14.0	1.3E-5	3.2E-4
<input type="checkbox"/>	GOTERM_MF_FAT	olfactory receptor activity	RT		10	11.6	4.0E-5	5.3E-3
<input type="checkbox"/>	GOTERM_BP_FAT	sensory perception of smell	RT		10	11.6	5.4E-5	1.7E-2
<input type="checkbox"/>	SP_PIR_KEYWORDS	olfaction	RT		10	11.6	7.3E-5	9.2E-3
<input type="checkbox"/>	GOTERM_BP_FAT	sensory perception of chemical stimulus	RT		10	11.6	1.2E-4	1.9E-2
<input type="checkbox"/>	INTERPRO	Olfactory receptor	RT		10	11.6	1.4E-4	2.4E-2
<input type="checkbox"/>	PIR_SUPERFAMILY	PIRSF003152:G protein-coupled olfactory receptor, class II	RT		8	9.3	2.1E-4	2.5E-3
<input type="checkbox"/>	KEGG_PATHWAY	Olfactory transduction	RT		9	10.5	4.2E-4	1.7E-2
<input type="checkbox"/>	INTERPRO	GPCR, rhodopsin-like superfamily	RT		12	14.0	4.3E-4	3.7E-2
<input type="checkbox"/>	INTERPRO	7TM GPCR, rhodopsin-like	RT		12	14.0	4.4E-4	2.5E-2
<input type="checkbox"/>	SP_PIR_KEYWORDS	g-protein coupled receptor	RT		12	14.0	8.2E-4	5.1E-2
<input type="checkbox"/>	SP_PIR_KEYWORDS	sensory transduction	RT		10	11.6	1.1E-3	4.7E-2
<input type="checkbox"/>	GOTERM_BP_FAT	sensory perception	RT		11	12.8	1.4E-3	1.4E-1
<input type="checkbox"/>	SP_PIR_KEYWORDS	transducer	RT		12	14.0	1.4E-3	4.4E-2
<input type="checkbox"/>	GOTERM_BP_FAT	G-protein coupled receptor protein signaling pathway	RT		13	15.1	1.6E-3	1.2E-1
<input type="checkbox"/>	SP_PIR_KEYWORDS	disulfide bond	RT		24	27.9	2.9E-3	7.0E-2
<input type="checkbox"/>	GOTERM_BP_FAT	cognition	RT		11	12.8	3.3E-3	1.9E-1
<input type="checkbox"/>	UP_SEQ_FEATURE	glycosylation site:N-linked (GlcNAc...)	RT		30	34.9	3.6E-3	7.1E-1
<input type="checkbox"/>	UP_SEQ_FEATURE	disulfide bond	RT		23	26.7	3.7E-3	4.7E-1
<input type="checkbox"/>	SP_PIR_KEYWORDS	glycoprotein	RT		31	36.0	3.9E-3	7.8E-2
<input type="checkbox"/>	UP_SEQ_FEATURE	transmembrane region	RT		33	38.4	6.7E-3	5.4E-1
<input type="checkbox"/>	GOTERM_BP_FAT	neurological system process	RT		12	14.0	8.5E-3	3.7E-1
<input type="checkbox"/>	SP_PIR_KEYWORDS	transmembrane	RT		33	38.4	9.0E-3	1.5E-1

Subject Z's SNVs associated with allele-specific expression (ASE) and binding (ASB)



- Subject Z has **2,081,803** heterozygous SNVs
- **40,267** and **19,687** of them were found in AlleleDB, assessed for ASE and ASB, respectively
- **8,680** (ASE) and **734** (ASB) SNVs show some evidence for allele-specific behavior: allele-specific in at least one of the 380 AlleleDB individuals
- **173** of SNVs show consistent association with ASE across AlleleDB individuals: present in more than 100 individuals and found to be allele-specific in more than 95% of them



Subject Z's SNVs associated with allele-specific expression (ASE) and binding (ASB)

Expression/binding balanced between alleles

RNA-/ChIP-Seq Reads

ACTTTGATAGCGTCAAC**CG**

CTTTGATAGCGTCAAC**CGC**

CTTTGATAGCGTCAAC**CGC**

TTGACAGCGTCAAT**TGCAC**

TGATAGCGTCAAT**TGCACG**

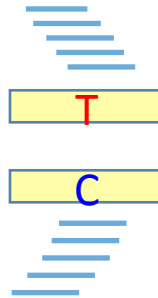
ATAGCGTCAAC**CGC**ACGTC

TAGCGTCAAT**TGCACGTCG**

CGTCAAC**CGC**ACGTCGGGA

GTCAAT**TGCACGTCGAGAG**

CAAT**TGCACGTCGGGAGTT**



Allele-specific expression/binding

RNA-/ChIP-Seq Reads

ACTTTGATAGCGTCAAT**TG**

CTTTGATAGCGTCAAT**TGC**

CTTTGATAGCGTCAAT**TGC**

TTGACAGCGTCAAT**TGCAC**

TGATAGCGTCAAT**TGCACG**

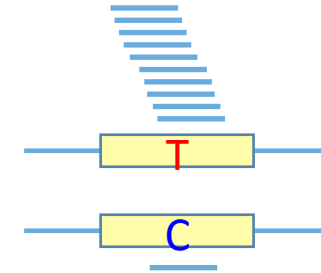
ATAGCGTCAAT**TGCACGTC**

TAGCGTCAAT**TGCACGTCG**

CGTCAAC**CGC**ACGTCGGGA

GTCAAT**TGCACGTCGAGAG**

CAAT**TGCACGTCGGGAGTT**



- Subject Z has **2,081,803** heterozygous SNVs
- **40,267** and **19,687** of them were found in AlleleDB, assessed for ASE and ASB, respectively
- **8,680** (ASE) and **734** (ASB) SNVs show some evidence for allele-specific behavior: allele-specific in at least one of the 380 AlleleDB individuals
- **173** of SNVs show consistent association with ASE across AlleleDB individuals: present in more than 100 individuals and found to be allele-specific in more than 95% of them

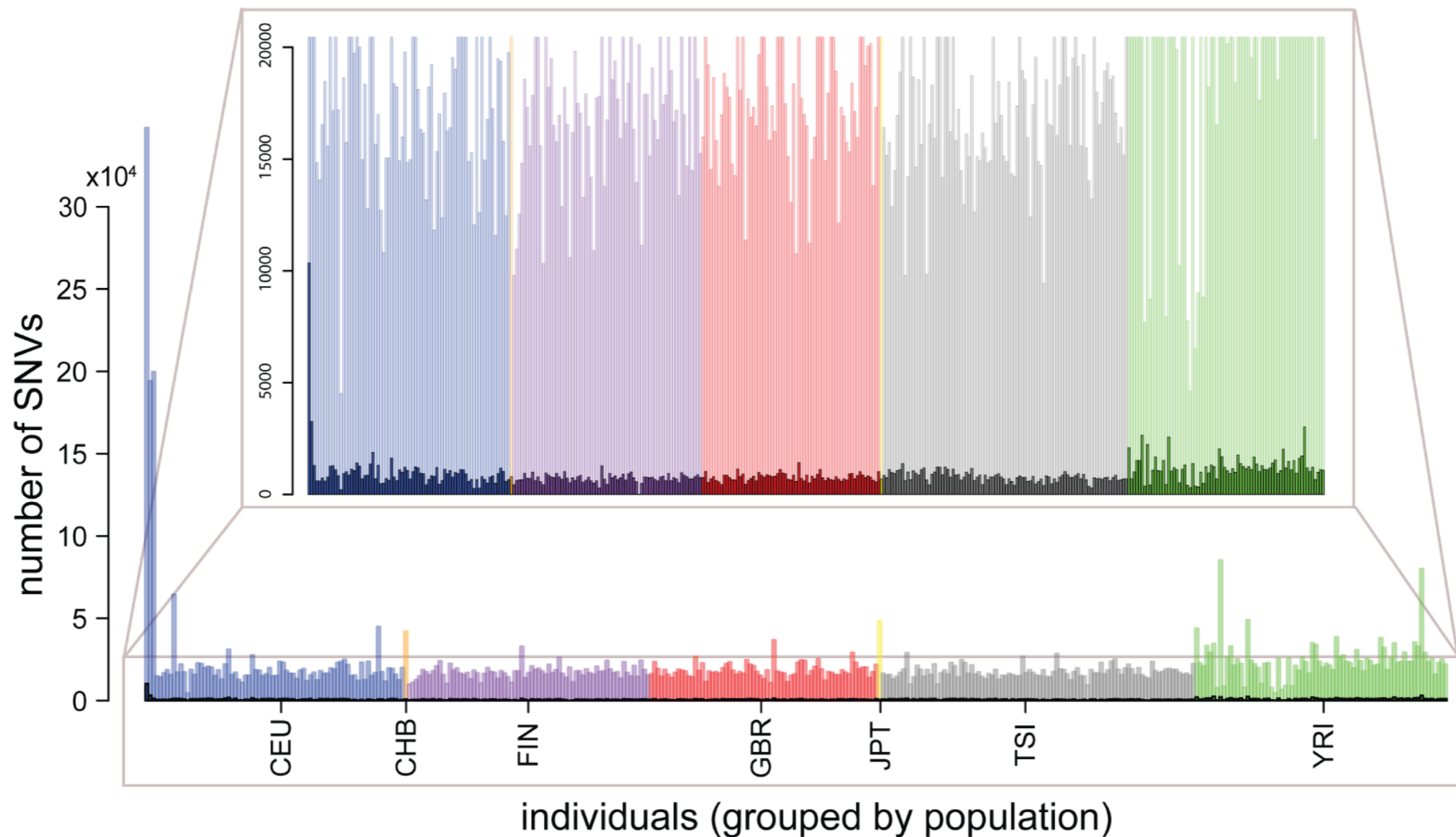


AlleleDB



alleledb.gersteinlab.org

632,126 ASE/nonASE-annotated SNVs across **383** individuals



342,500 ASE/nonASE-annotated SNVs across **380** individuals after removing the NA12878-NA12891-NA12892 trio

40,078 (/3,595,421) *N*, **38,999** (/3,521,063) *S*, and **40,304** (/3,559,138) *Z* HetSNPs in AlleleDB

Number of N/S/Z HetSNPs matching HetSNPs in AlleleDB individuals

columns: total count of individuals with the HetSNP accessible,

rows: fraction of individuals in which the HetSNP is associated with ASE.

	1	2-4	5-19	20-99	100-380
[0.00, 0.05)	9956/9710/10149	7082/6835/6839	5566/5412/5624	7809/7819/8012	4495/4616/4668
[0.05, 0.25)	0/0/0	0/0/0	971/959/992	936/771/863	239/210/201
[0.25, 0.75)	0/0/0	602/554/613	356/324/376	306/259/300	82/68/72
[0.75, 0.95)	0/0/0	23/16/22	61/62/82	77/66/64	24/34/20
[0.95, 1.00]	675/614/689	245/221/250	133/119/126	246/175/169	194/155/173

	1	2-4	5-19	20-99	100-380
[0.00, 0.05)	9956/9710/10149	7082/6835/6839	5566/5412/5624	7809/7819/8012	4495/4616/4668
[0.05, 0.25)	0/0/0	0/0/0	971/959/992	936/771/863	239/210/201
[0.25, 0.75)	0/0/0	602/554/613	356/324/376	306/259/300	82/68/72
[0.75, 0.95)	0/0/0	23/16/22	61/62/82	77/66/64	24/34/20
[0.95, 1.00]	675/614/689	245/221/250	133/119/126	246/175/169	194/155/173

