Genetic Insights from a Genetically Isolated Community

Alyssa Wheeler

About Myself

- Graduate student
- Bioinformatics
- Healthcare research

Introduction

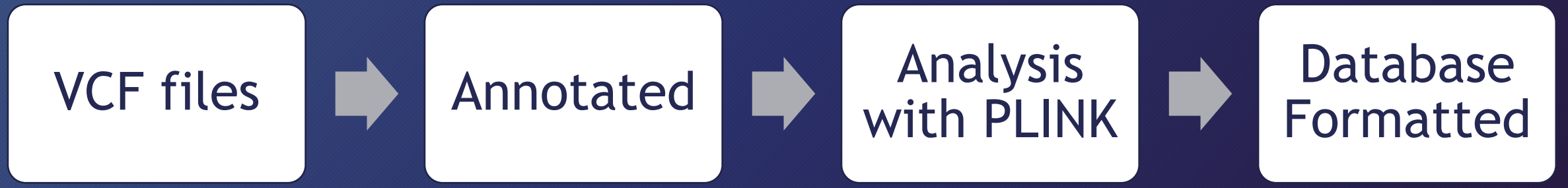
- Plain people
  - Genetically isolated
  - Founder effect
- Broader society
  - Leaving Plain community
- Genetic databases
  - Healthcare providers

Project Goal

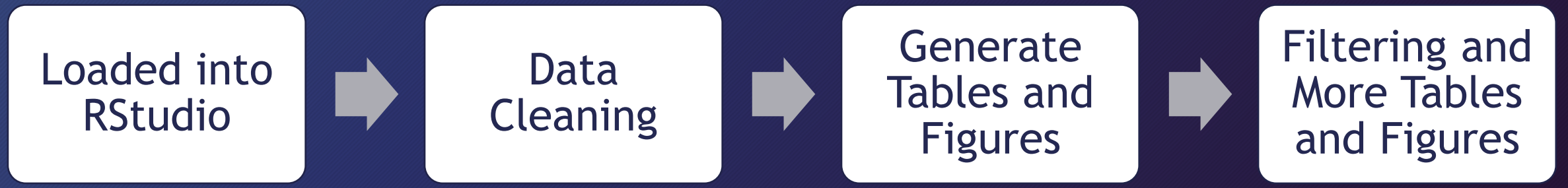
This study created a database from 210 Kish Valley Amish and Mennonite patient exomes. Then, exploratory data analysis was performed to identify meaningful gene variants with a focus on genes of interest identified by Dr. Morton.

Methods

Part 1: Database Creation



Part 2: Exploratory Data Analysis



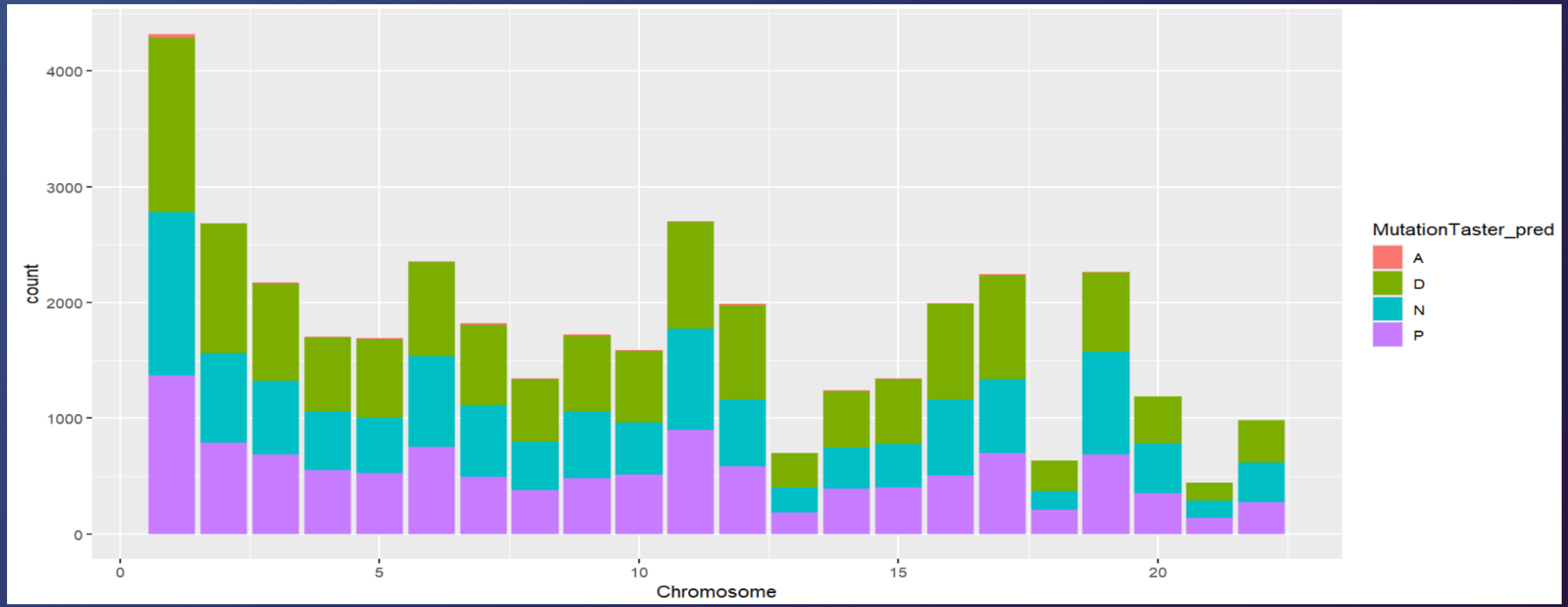
Results

Database

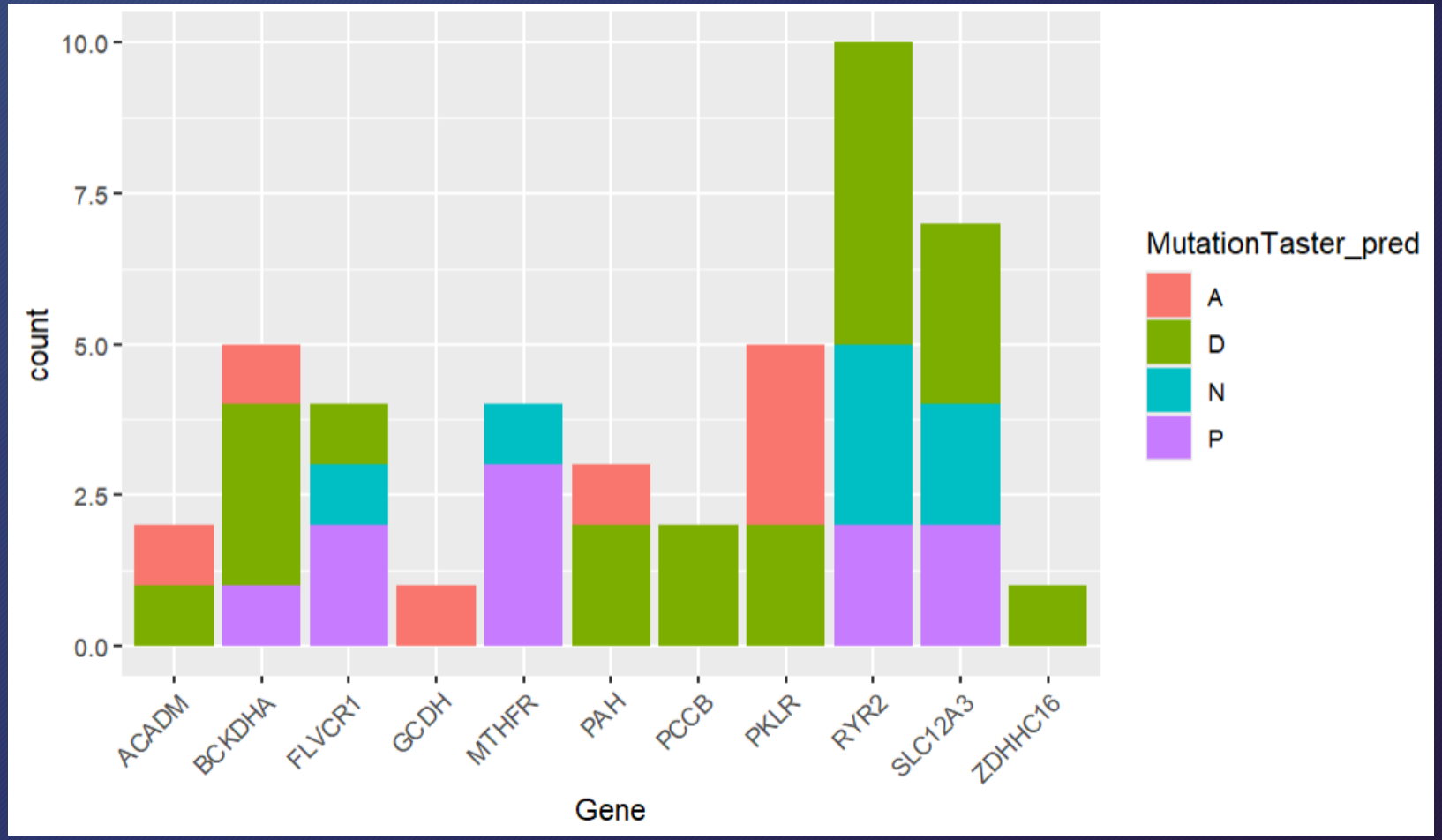
gnomad	SNPid	Chr	Position	Ref	Alt	Func	Gene	GeneDetail	ExonicFunc	AAChange	C(HOM A1)	C(HET)	C(HOM A2
gnomad	1:69063:T:C	1	69063	T	C	upstream	OR4F5	dist=28	.	.	0	1	0
gnomad	1:69081:G:C	1	69081	G	C	upstream	OR4F5	dist=10	.	.	1	1	0
gnomad	1:69270:A:G	1	69270	A	G	exonic	OR4F5	.	synonymous SNV	OR4F5:NM_00100	112	5	0
gnomad	1:69428:T:G	1	69428	T	G	exonic	OR4F5	.	nonsynonymous S	OR4F5:NM_00100	1	0	0
gnomad	1:69511:A:G	1	69511	A	G	exonic	OR4F5	.	nonsynonymous S	OR4F5:NM_00100	140	0	0
gnomad	1:69552:G:C	1	69552	G	C	exonic	OR4F5	.	synonymous SNV	OR4F5:NM_00100	1	0	0
gnomad	1:69594:T:C	1	69594	T	C	exonic	OR4F5	.	synonymous SNV	OR4F5:NM_00100	0	1	0
gnomad	1:69648:CAGGCT:C	1	69648	AGGCT	-	exonic	OR4F5	.	frameshift deleti	OR4F5:NM_00100	0	1	0
gnomad	1:69657:T:C	1	69657	T	C	exonic	OR4F5	.	synonymous SNV	OR4F5:NM_00100	0	1	0
gnomad	1:69682:GT:G	1	69682	T	-	exonic	OR4F5	.	frameshift deleti	OR4F5:NM_00100	0	1	0

MAF_PLIN	MAF	MAF_%	NCHROBS	AF	AF_nfe	SIFT_pre	LRT_pre	MutationTaster_p	MutationAssessor_p	FATHMM_pred	REVEL_score	CADD_raw	GERP++_NR	GERP++_RS	rsNum
0.75	0.002381	0.238095	4	.	.	.	.	.	.	.	.	.	.	.	.
0.9786	0.002381	0.238095	234	0.0103	0.0262	.	.	.	.	.	.	.	.	.	.
1	0.011905	1.190476	2	0.8380	0.9146	.	.	.	.	.	.	.	.	.	.
1	0	0	280	0.0245	0.0417	D	N	N	M	T	0.055	3.839511	2.31	0.891	.
1	0	0	2	0.9497	0.9726	T	N	P	N	T	0.053	0.804646	2.31	1.15	.
0.5	0	0	2	1.234e-05	0	.	.	.	.	.	.	.	.	.	.
0.5	0.002381	0.238095	2	0.0012	4.169e-05	.	.	.	.	.	.	.	.	.	.
0.5	0.002381	0.238095	2	.	.	.	.	.	.	.	.	.	.	.	.
0.5	0.002381	0.238095	2	.	.	.	.	.	.	.	.	.	.	.	.
0.5	0.002381	0.238095	2	.	.	.	.	.	.	.	.	.	.	.	.

MutationTaster Predictor Results



MutationTaster Predictions with Genes of Interest



PCCB had highest MAF

SNPid	Gene	Chr	MAF_percent	REVEL_score	CADD_raw	GERP++_NR
3:136048854:A:G	PCCB	3	3.571429	0.919	5.072618	5.3
19:41930487:T:A	BCKDHA	19	1.904762	0.953	5.097043	5.45
16:56919275:C:G	SLC12A3	16	1.428571	0.511	6.313977	5.4
12:103246653:C:T	PAH	12	0.714286	0.985	6.13065	5.72
1:155261636:C:T	PKLR	1	0.47619	0.947	5.750361	4.85
12:103234252:T:C	PAH	12	0.238095	0.982	4.988886	5.63
12:103249091:C:T	PAH	12	0.238095	0.89	5.048905	5.73

Discussion

- Database
  - Key genetic metrics
  - Resource for healthcare and broader society
- Pathogenicity
  - Most likely disease causing: ACADM, BCKDHA, GCDH, PAH, PCCB, PKLR, RYR2, SLC12A3, and ZDHHC16
  - Neutral or harmless: FLVCR1 and MTHFR
- Filtering for Harmful Predictions
  - Highest minor allele frequency: PCCB
  - CADD scores

Conclusion

- Prioritize harmful gene variants
- Broader impacts
- Improved healthcare

Limitations

- Smalls sample size (210 individuals)
- Predictor tool errors
- Future research

Acknowledgements

Thank you to Dr. Lamendella, Dr. Morton, and Brittney McMullen for their help and guidance throughout this project