

***ZCWPWI* loss-of-function variants in Alzheimer's Disease**

Fahri Küçükali^{1,2}, Katrin Nußbaumer³, Jasper Van Dongen^{1,2}, Elisabeth Hens^{4,5,6,7}, Céline Bellenguez⁸, Benjamin Grenier-Boley⁸, Delphine Daian⁹, Anne Boland⁹, Jean-François Deleuze⁹, Jean-Charles Lambert⁸, Christine Van Broeckhoven^{2,4}, Kristel Sleegers^{1,2*}, on behalf of the BELNEU Consortium[§]

¹Complex Genetics of Alzheimer's Disease Group, Center for Molecular Neurology, VIB, Antwerp, Belgium

²Department of Biomedical Sciences, University of Antwerp, Antwerp, Belgium

³University of Applied Sciences Campus Vienna, Vienna, Austria

⁴Neurodegenerative Brain Diseases Group, VIB Center for Molecular Neurology, Antwerp, Belgium

⁵Department of Neurology and Memory Clinic, Hospital Network Antwerp, Middelheim and Hoge Beuken, Antwerp, Belgium

⁶Department of Neurology, University Hospital Antwerp, Edegem, Belgium

⁷Department of Neurology, University Hospital Brussel and Center for Neurosciences, Free University of Brussels, Brussels, Belgium

⁸Univ. Lille, Inserm, CHU Lille, Institut Pasteur de Lille, U1167-RID-AGE facteurs de risque et déterminants moléculaires des maladies liés au vieillissement, Lille, France

⁹Université Paris-Saclay, CEA, Centre National de Recherche en Génomique Humaine, 91057, Evry, France

[§] The following members of the BELNEU Consortium have contributed to the clinical and pathological phenotyping and follow-up of the Belgian patient cohorts: Johan Goeman, Roeland Crols (Hospital Network Antwerp (ZNA) Middelheim and Hoge Beuken, Antwerp, Belgium); Dirk Nuytten (Hospital Network Antwerp (ZNA) Stuivenberg, Antwerp, Belgium); Rudy Mercelis (Antwerp University Hospital, Edegem); Mathieu Vandenbulcke (University of Leuven and University Hospitals Leuven, Leuven, Belgium); Anne Sieben, Jan L. De Bleecker, Patrick Santens (University Hospital Ghent, Ghent, Belgium); Jan Versijpt, Alex Michotte (University Hospital Brussels, Brussels, Belgium); Olivier Deryck, Ludo Vanopdenbosch, Bruno Bergmans (AZ Sint-Jan Brugge, Bruges, Belgium); Christiana Willems, Nina De Klippel (Jessa Hospital, Hasselt, Belgium); Jean Delbeck (General Hospital Sint-Maria, Halle); Adrian Ivanoiu (Saint-Luc University Hospital, Université Catholique de Louvain, Louvain-la-Neuve, Belgium); and Eric Salmon (University of Liege and Memory Clinic, CHU Liege, Liege, Belgium).

*Corresponding author: Prof. Dr. Kristel Sleegers MD PhD

Complex Genetics of Alzheimer's Disease Group; VIB Center for Molecular Neurology

University of Antwerp - CDE

Universiteitsplein 1, B-2610, Antwerp, Belgium

Email: kristel.slegers@uantwerpen.vib.be

Supplementary Figures

Supplementary Figure 1. Genetic ancestry analysis of the case-control cohort with the 1KG samples. The PCA plot was generated on first and second genetic principal components calculated from genotyping array data of $n=1803$ samples that were included in this study, together with 1KG Project samples ($n=2504$). AFR: African population, AMR: Ad mixed American population, EAS: East Asian population, EUR: European population, SAS: South Asian population.

Supplementary Figure 2. Coverage Circos plot for targeted exons and UTRs of *PILRB*, *PILRA* and *ZCWPWI*. *PILRB*, *PILRA*, and *ZCWPWI* exons are colored in blue, red, and green respectively. Each exon is extended by 20 bp to include splice regions. The numbers in the inner rim indicate the exon number of each gene of interest. The outer rim indicates percentage of the study cohort covered $\geq 20X$ on the corresponding position and indicate 10% difference in this scale. The coverage information indicated on the figure is solely based on sequencing on the MiSeq platform, therefore does not include additional Sanger sequencing of *PILRB* exon 5 and *ZCWPWI* exon 14 due to relative drop in $\geq 20X$ coverage in the study cohort.

Supplementary Figure 3. Haplotype sharing analysis of non-singleton *ZCWPWI* predicted LoF variant carriers. **(A)** Haplotype sharing analysis for *ZCWPWI* c.631+1G>T. **(B)** Haplotype sharing analysis for *ZCWPWI* c.-29-1G>A. For each panel, the distances to the respective mutation are provided. Allele frequencies for each STR genotype are given in base lengths. Frequencies of indicated STR lengths are shown in brackets (calculated based on STR lengths in $n=172$ Flanders-Belgian healthy control individuals) and for SNPs gnomAD NFE MAFs are provided. IDs of STRs and SNPs within the targeted resequencing locus are bolded. “?” indicates that the related SNP genotype information is unknown. Risk-increasing and mutation alleles are bolded in red. Main possible shared haplotypes are highlighted in green and blue respectively.

Supplementary Figure 4. IGV snapshot of Illumina short-read sequencing reads to show phasing between *ZCWPWI* p.Glu105Gly and *ZCWPWI* p.Glu95Lys predicted deleterious rare missense variants in LOAD-P7.

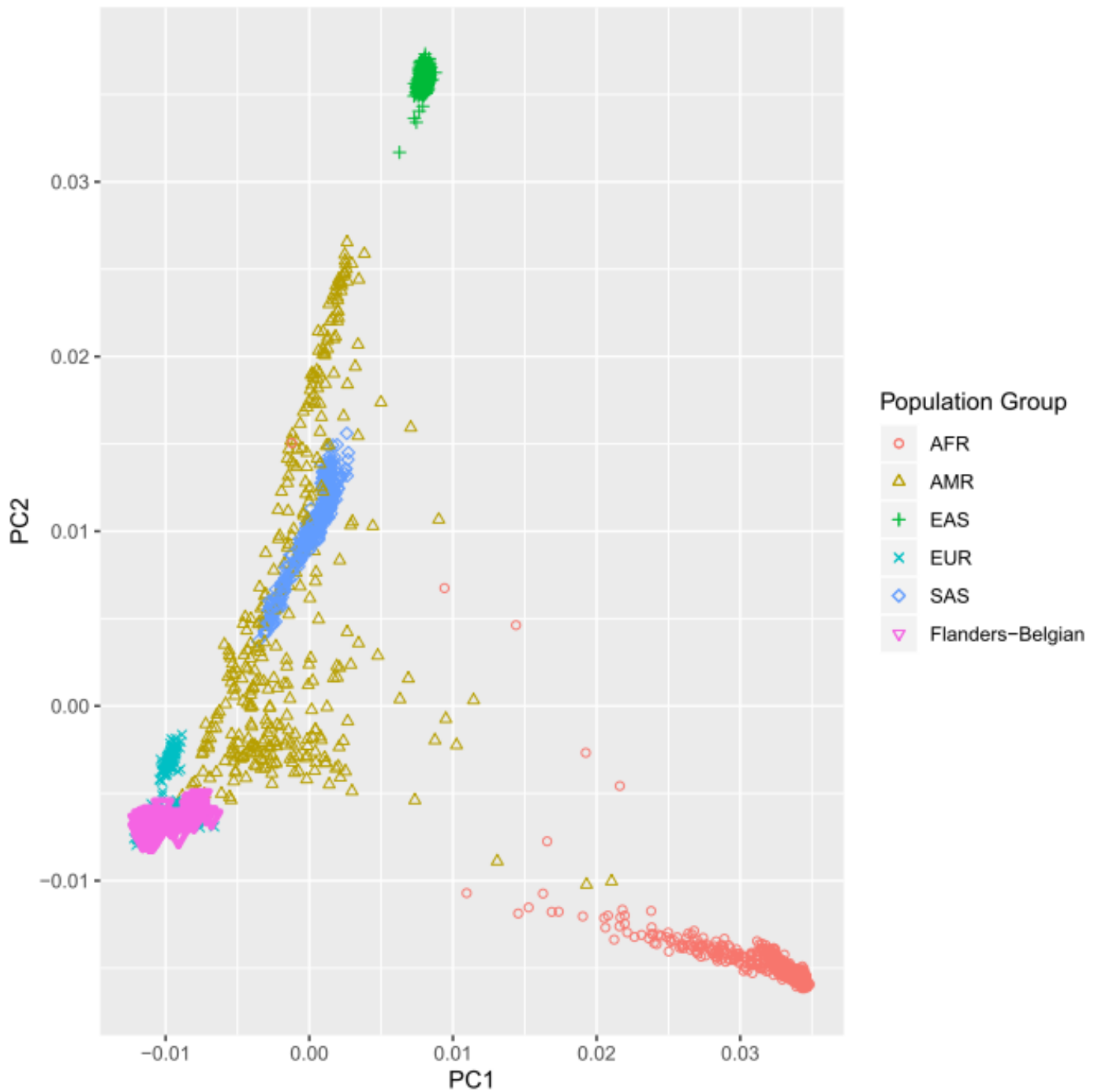
Supplementary Figure 5. IGV snapshot of Nanopore long-read sequencing reads of LOAD-P2 and LOAD-P3 harboring rs34919929 and rs774275324 (*ZCWPWI* c.631+1G>T). Aligned reads are colored by their “HP” tag (haplotype identifier) added by WhatsHap software for

each sample separately. For each sample, haplotypes defined by WhatsHap are shown below its coverage track. The coverage tracks are in \log_{10} scale.

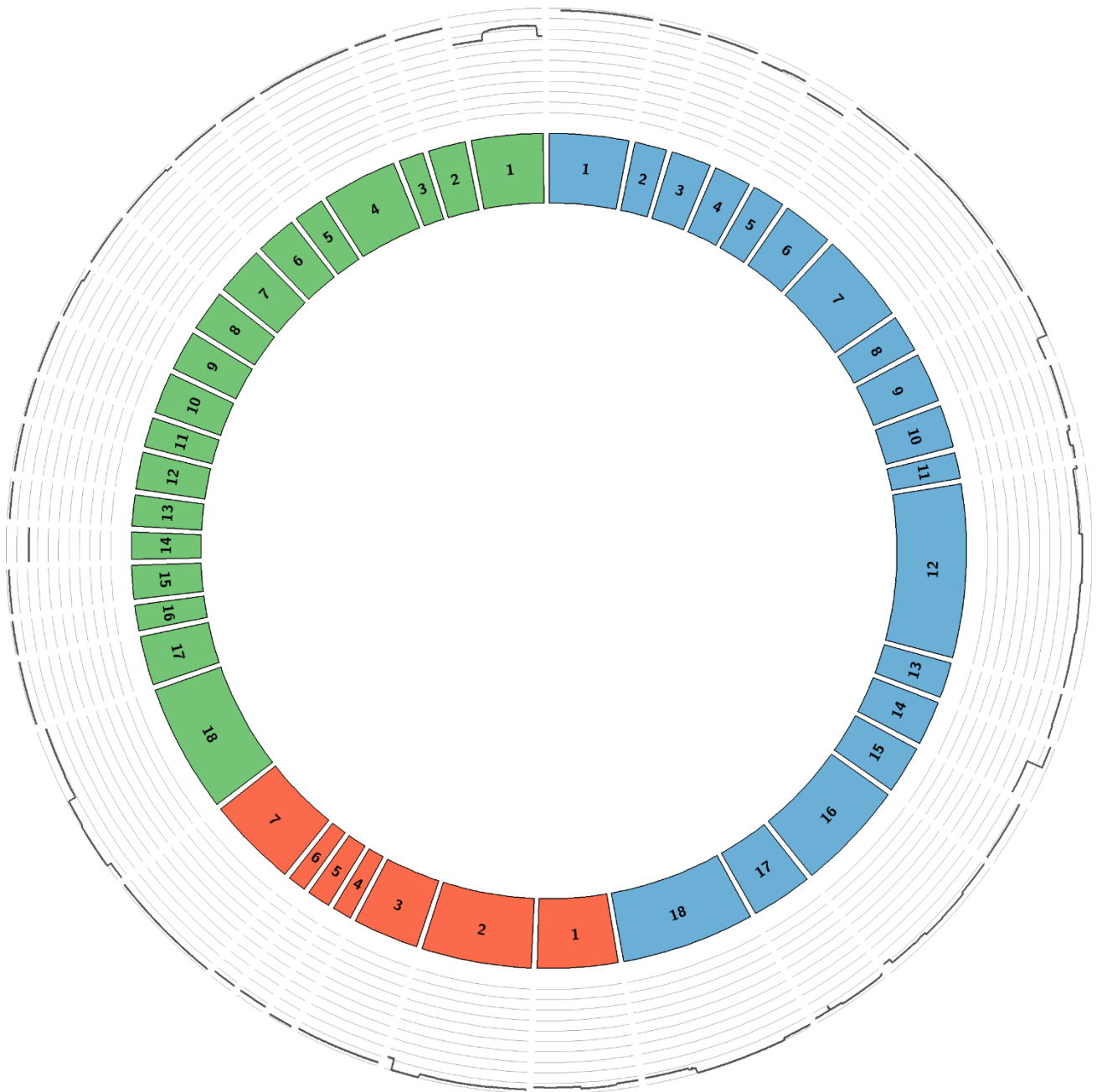
Supplementary Figure 6. IGV snapshot of Nanopore long-read sequencing reads of LOAD-P4, LOAD-P5 and LOAD-P6 harboring rs34919929 and rs1180932049 (*ZCWPWI* c.-29-1G>A). Aligned reads are colored by their “HP” tag (haplotype identifier) added by WhatsHap software for each sample separately. For each sample, haplotypes defined by WhatsHap shown below its coverage track. The coverage tracks are in \log_{10} scale.

Supplementary Figure 7. IGV snapshot of Nanopore long-read sequencing reads of three negative control samples having all possible genotypes for rs34919929 and not harboring any predicted LoF mutations of interest. The negative control samples were named according to their rs34919929 genotype. Aligned reads are colored by their “HP” tag (haplotype identifier) added by WhatsHap software for each sample separately. For each sample, haplotypes defined by WhatsHap shown below its coverage track. The coverage tracks are in \log_{10} scale.

Supplementary Figure 1. Genetic ancestry analysis of the case-control cohort with the 1KG samples.



Supplementary Figure 2. Coverage Circos plot for targeted exons and UTRs of *PILRB*, *PILRA* and *ZCWPW1*.



Supplementary Figure 3. Haplotype sharing analysis of non-singleton *ZCWPWI* predicted LoF variant carriers. **(A)** Haplotype sharing analysis for *ZCWPWI* c.631+1G>T. **(B)** Haplotype sharing analysis for *ZCWPWI* c.-29-1G>A.**A.**

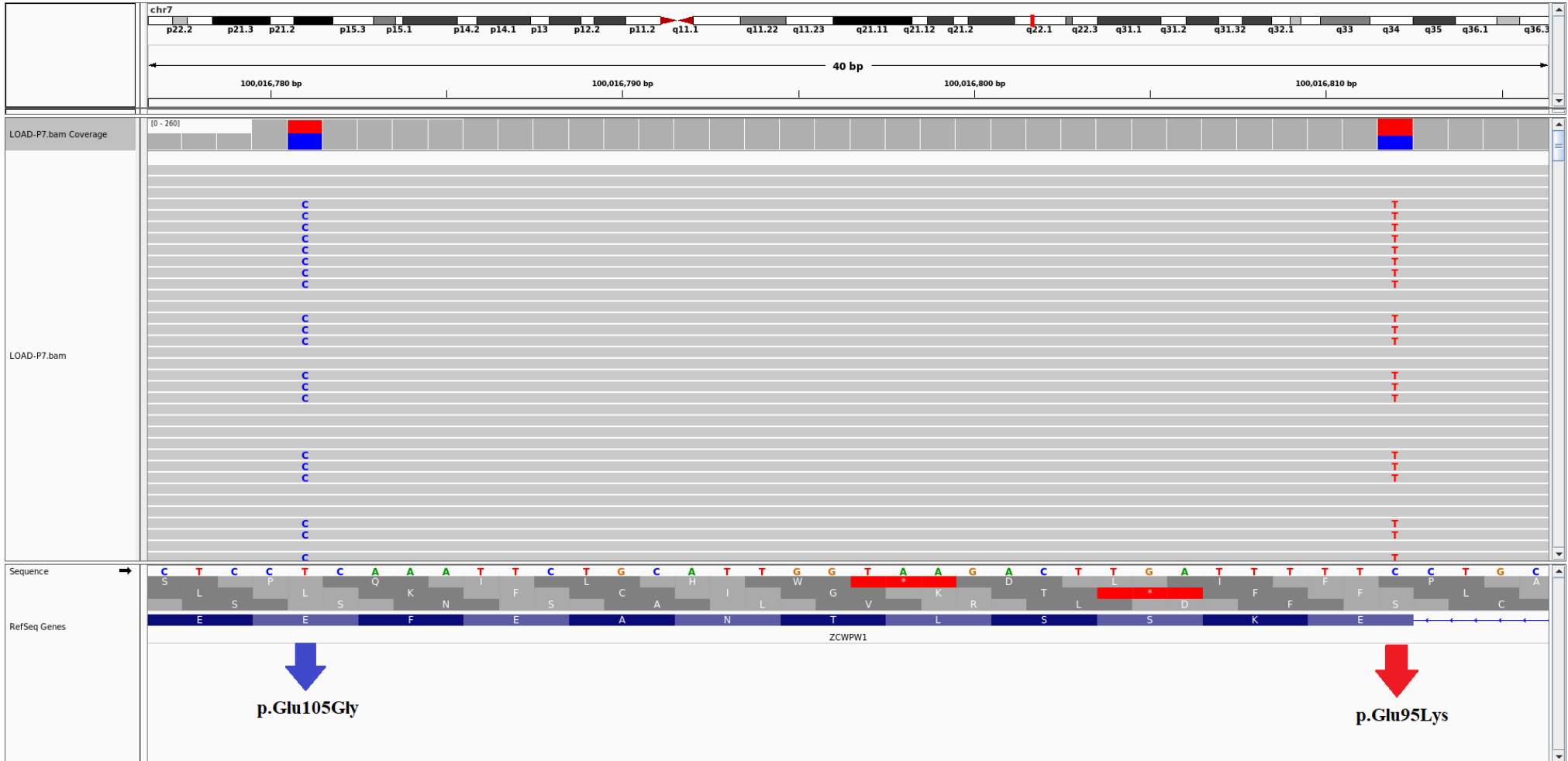
Distance to LoF	STR or SNP	<i>ZCWPWI</i> c.631+1G>T				Common STR Allele Frequencies or <i>gnomAD</i> <i>NFE MAF</i>
		LOAD-P2		LOAD-P3		
		Allele 1	Allele 2	Allele 1	Allele 2	
-4.87 Mb	D7S2431	116	118	102	114	-
-4.39 Mb	<i>rs1488514</i>	A	A	A	A	G (23.3%)
-3.46 Mb	D7S1796	287	292	287	284	287 (28.4%)
-2.68 Mb	D7S554	260	254	260	262	260 (33.7%)
-2.09 Mb	<i>rs6970990</i>	T	T	G	G	G (47.7%)
-1.46 Mb	D7S651	178	182	178	178	178 (36.6%)
-745 Kb	D7S647	178	180	178	178	178 (38.6%)
-9.48 Kb	<i>rs1476679</i>	T	C	T	C	C (29.7%)
-1.59 Kb	<i>rs34919929</i>	A	G	A	G	G (29.7%)
0	<i>rs774275324</i> (c.631+1G>T)	A	C	A	C	A (0.0023%)
+42 Kb	D7S2480	213	223	213	223	213 (8.4%) & 223 (1.4%)
+763 Kb	D7S477	224	226	224	224	224 (37.2%)
+1.73 Mb	D7S666	161	165	161	161	161 (43.8%)
+2.09 Mb	D7S2448	249	237	249	251	249 (29.9%)
+3.31 Mb	<i>rs727708</i>	T	T	C	C	T (48.5%)
+3.45 Mb	D7S2504	200	208	194	212	-
+3.53 Mb	<i>rs10245317</i>	G	A	A	G	G (16.6%)
+4.03 Mb	D7S2494	286	292	292	296	292 (6.1%)

Supplementary Figure 3 continued

B.

Distance to LoF	STR or SNP	ZCWPWI c.-29-1G>A						Common STR Allele Frequencies or <i>gnomAD</i> NFE MAF
		LOAD-P4		LOAD-P5		LOAD-P6		
		Allele 1	Allele 2	Allele 1	Allele 2	Allele 1	Allele 2	
-4.87 Mb	D7S2431	102	118	114	122	112	118	118 (17.7%)
-4.39 Mb	<i>rs1488514</i>	G	G	?	?	A	A	G (23.3%)
-3.46 Mb	D7S1796	287	296	287	296	287	287	287 (28.4%) & 296 (15.4%)
-2.69 Mb	D7S554	254	260	254	254	254	258	254 (34%)
-2.09 Mb	<i>rs6970990</i>	T	G	?	?	T	T	G (47.7%)
-1.46 Mb	D7S651	182	178	182	182	182	182	182 (28.4%)
-749 Kb	D7S647	178	176	178	178	178	180	178 (38.6%)
-13.85 Kb	<i>rs1476679</i>	T	C	T	T	T	C	C (29.7%)
-5.96 Kb	<i>rs34919929</i>	A	G	A	A	A	G	G (29.7%)
0	<i>rs1180932049</i> (c.-29-1G>A)	T	C	T	C	T	C	T (0.0023%)
+37 Kb	D7S2480	202	221	202	217	202	221	202 (21.2%) & 221 (20%)
+758 Kb	D7S477	226	226	226	226	226	224	226 (18.8%)
+1.73 Mb	D7S666	159	165	165	161	159	161	159 (21.2%) & 161 (43.8%) & 165 (11.9%)
+2.08 Mb	D7S2448	247	249	247	243	247	249	247 (13%)
+3.31 Mb	<i>rs727708</i>	T	C	?	?	T	C	T (48.5%)
+3.44 Mb	D7S2504	194	204	194	206	194	210	194 (12.7%)
+3.53 Mb	<i>rs10245317</i>	A	A	?	?	G	G	G (16.6%)
+4.03 Mb	D7S2494	292	286	294	298	292	296	292 (6.1%)

Supplementary Figure 4. IGV snapshot of Illumina short-read sequencing reads to show phasing between *ZCWP1* p.Glu105Gly and *ZCWP1* p.Glu95Lys predicted deleterious rare missense variants in LOAD-P7.



Supplementary Figure 5. IGV snapshot of Nanopore long-read sequencing reads of LOAD-P2 and LOAD-P3 harboring rs34919929 and rs774275324 (*ZCWPWI* c.631+1G>T).



Supplementary Figure 6. IGV snapshot of Nanopore long-read sequencing reads of LOAD-P4, LOAD-P5 and LOAD-P6 harboring rs34919929 and rs1180932049 (*ZCWPWI* c.-29-1G>A).



Supplementary Figure 7. IGV snapshot of Nanopore long-read sequencing reads of three negative control samples having all possible genotypes for rs34919929 and not harboring any predicted LoF mutations of interest.

