bioRxiv preprint doi: https://doi.org/10.1101/765651; this version posted September 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Genome-wide selection scan in the Kuwaiti population identifies a *TNKS* haplotype linked to metabolic traits and hypertension

Muthukrishnan Eaaswarkhanth[1], Andre Luiz Campelo Santos[2,3], Omer Gokcumen[2], Fahd Al-Mulla[1*], Thangavel Alphonse Thanaraj[1*]

[1]Department of Genetics and Bioinformatics, Dasman Diabetes Institute, Kuwait.
[2]Department of Biological Sciences, State University of New York at Buffalo, NY, USA.
[3]Department of Archeology, Federal University of Pernambuco, Brazil.

*Correspondence should be addressed to F.A-M. (fahd.almulla@dasmaninstitute.org) or T.A.T. (alphonse.thangavel@dasmaninstitute.org).

bioRxiv preprint doi: https://doi.org/10.1101/765651; this version posted September 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

Despite the extreme and varying environmental conditions prevalent in the Arabian Peninsula, it has experienced several waves of human migrations following the out-of-Africa diaspora. Eventually, the inhabitants of the peninsula region adapted to the hot and dry environment. The adaptation and natural selection that shaped the extant human populations of the Arabian Peninsula region have been scarcely studied. In an attempt to explore natural selection in the region, we analyzed 662,750 variants in 583 Kuwaiti individuals. We searched for regions in the genome that display signatures of positive selection in the Kuwaiti population using an integrative approach in a conservative manner. We highlight a haplotype overlapping *TNKS* that showed strong signals of positive selection based on the results of the multiple selection tests conducted (integrated Haplotype Score, Cross Population Extended Haplotype Homozygosity, Population Branch Statistics, and log-likelihood ratio scores). Notably, the *TNKS* haplotype under selection potentially conferred a fitness advantage to the Kuwaiti ancestors for surviving in the harsh environment while posing a major health risk to present-day Kuwaitis.

Keywords: Kuwait population, positive selection, metabolic traits, hypertension

bioRxiv preprint doi: https://doi.org/10.1101/765651; this version posted September 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

Archaeological evidence suggests that the Arabian Peninsula played a key role during the dispersal of modern humans out-of-Africa (Cabrera et al. 2010; Rose and Petraglia 2010; Petraglia et al. 2019). Anatomically modern humans have inhabited the Arabian Peninsula since immediately after the out-of-Africa migration; therefore, the resident populations have a long and complex evolutionary history (Petraglia and Alsharekh 2003). Being at the crossroads between Africa and Eurasia, the Arabian Peninsula served as a point of interaction of human populations and trade across the region (Groucutt and Petraglia 2012). The resettlement of peoples and traders facilitated population admixture and increased genetic diversity. Several genetic studies offer insights on how the genetic ancestry, consanguinity, and admixture have structured the genetic diversity of the Arab populations. For example, uniparental genetic examinations demonstrated maternal (Abu-Amero et al. 2007; Rowold et al. 2007; Abu-Amero et al. 2008) and paternal (Abu-Amero et al. 2009; Triki-Fendri et al. 2016) genetic affinities and admixture events. Genome-wide characterization studies (Behar et al. 2010; Hunter-Zinck et al. 2010; Alsmadi et al. 2013) have elaborated the genetic structure and diversity within the peninsula as well as across the continents. In addition, a recent whole-exome-based study revealed the heterogeneous genetic structure of the Middle Eastern populations (Scott et al. 2016).

The population of the State of Kuwait exemplifies the overall heterogeneity of Middle Eastern populations. The State of Kuwait is one of the seven countries located in the Arabian Peninsula on the coastal region of the Arabian Gulf. Also placed at the head of the Persian Gulf, Kuwait is bordered by Saudi Arabia and Iraq to the south and north, respectively. The ancestors of the extant Kuwaiti population were early settlers that migrated from Saudi Arabia (Alghanim 1998; Casey 2007). Until the discovery of oil, they mostly derived their livelihoods from fishing and merchant seafaring (Lienhardt 2001). The maritime trade activities with India and Africa made Kuwait a central hub linking India, Arabia, and Persia to Europe (Slot 2003). The movement of populations from settlements in neighboring regions, particularly Saudi Arabia and Persia (Alenizi et al. 2008), in addition to the consequent admixture between populations and consanguinity (Yang et

bioRxiv preprint doi: https://doi.org/10.1101/765651; this version posted September 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

al. 2014), have potentially shaped the genetic diversity of the Kuwaiti population. One of our previous small-scale genome-wide studies revealed that the genetic structure of the Kuwait population is heterogeneous, comprising three distinct ancestral genetic backgrounds that could be linked roughly to contemporary Saudi Arabian, Persian, and Bedouin populations (Alsmadi et al. 2013). Accordingly, putative ancestors of the extant Kuwait population could be traced to Saudi Arabian tribes, Persian tribes, and the nomadic Bedouin populations. This observation has been further corroborated in some whole-exome and genome-based studies (Alsmadi et al. 2014; John et al. 2015; Thareja et al. 2015; John et al. 2018).

Paleoanthropological studies have recorded the dramatic environmental transformations and extreme climatic conditions in the Arabian Peninsula over time in addition to the subsequent human dispersal into the region (Groucutt and Petraglia 2012). The extreme and varying environmental conditions could have influenced natural selection and triggered adaptation to the hot and dry desert climates (Rose and Petraglia 2010). Additionally, the ramifications of adaptive trends reported for continental populations (e.g., lactose tolerance, skin color, resistance to blood pathogens, etc.) may have implications for the health of Arabian populations. Indeed, genome-wide selection scans have revealed positive selection for lactose tolerance, as well as skin and eye color similar to those in Europeans and malaria resistance as that in Africans (Fernandes et al. 2019). Such studies provide a robust framework for investigating the potential adaptive trends in specific populations in the Arabian Peninsula. However, such focused studies are scarce (Yang et al. 2014; Fernandes et al. 2019). For example, an earlier small-scale exploration of ancestry components suggested that genetic regions associated with olfactory pathways were under natural selection in Kuwaiti populations (Yang et al. 2014). However, the study was limited to a sample of less than 50 individuals. In the present study, we build on the findings of previous studies by increasing the sample size considerably in addition to applying multiple approaches that have become available recently to identify selection in a genome-wide manner in Kuwaiti populations (**Fig. 1A**).

bioRxiv preprint doi: https://doi.org/10.1101/765651; this version posted September 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Results and Discussion

Integration of all selection tests scores: 385 common SNPs under selection

We interrogated 556,188 single nucleotide polymorphisms (SNPs) with all the four selection test results (**Fig. 1B** and **supplementary Table S1**). According to previously published guidelines (Cardona et al. 2014), we subsequently identified 385 SNPs, which had Cross Population Extended Haplotype Homozygosity (XP-EHH), Population Branch Statistics (PBS), and log-likelihood ratio scores (LLRS) values > 0, in addition to integrated Haplotype Score (iHS) values > 2 (**Fig. 1C** and **supplementary Table S2**). We surmised that the filtering would detect specific variants that are more likely to be true positive candidates for classical positive selection in the Kuwaiti population.

As expected, the allele frequencies of the alternate alleles for the 385 SNPs were high in the Kuwaiti population (KWT) in comparison with the allele frequencies for the same SNPs in the continental populations (**supplementary Fig. S1** and **supplementary Table S2**). In fact, we observed that the SNPs had the highest allele frequencies in Kuwaiti populations when compared with other human populations.

Window-based screening and genomic regions under positive selection

We then identified the genomic locations of the 385 SNPs (**Fig. 2**). Specifically, we speculated that if indeed selective sweeps in the Kuwaiti population explain the deviations from neutral expectations in the 385 SNPs, then they are expected to cluster within a smaller number of haplotypes. Therefore, we checked for clustering of the SNPs across the human reference genome in 100-kb windows. We identified approximately 220 windows 100 kb in length with at least one SNP under putative selection. In addition, we observed multiple instances where adjacent 100-kb windows harbored putatively selected SNPs, which we subsequently merged into single regions (**supplementary Table S3**).

To understand the functional implications of the putatively selected regions, we identified all the genes (n = 379) inside the 220 100-Kb windows and then performed a Gene Ontology enrichment analysis using all the GO annotations for *Homo sapiens* as the

bioRxiv preprint doi: https://doi.org/10.1101/765651; this version posted September 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

comparison background. After Bonferroni correction ($p < 0.05$), the analysis returned results associated only with a single group of biological processes: glycosaminoglycan biosynthetic process (aminoglycan biosynthetic process, aminoglycan metabolic process, glycosaminoglycan metabolic process) (**supplementary Table S4**). There were no significant results for molecular functions and cellular components.

TNKS haplotype related to obesity, hypertension, and asthma

To investigate further the functional impact of putatively selected haplotypes, we conducted a more thorough manual investigation of the genomic regions containing five or more SNPs identified by our selection scan. Specifically, we checked if any variant that was in linkage disequilibrium with the haplotypes had been reported in previous studies or in the UK Biobank (Sudlow et al. 2015; Bycroft et al. 2018) (**Table 1** and **supplementary Table S5**).

Based on the manual curation, we highlight a ~400-K haplotype on chromosome 8 (chr8: 9.3–9.7 Mb), which harbors seven putatively selected SNPs. To reveal the haplotype structure of the locus, we first investigated whether the seven putatively selected SNPs were in linkage disequilibrium with each other. Indeed, a single haplotype carried the alternate alleles for all the seven SNPs (1111111) (**Fig. 4C** and **Supplementary Fig. S2**). We further observed that the haplotype exhibited the highest allele frequency in the Kuwaiti population when compared with other populations from different parts of the globe (**supplementary Fig. S3**). In addition, within the Kuwaiti population, the subgroup of individuals with putative Saudi ancestry (Alsmadi et al. 2013) seemed to be driving up the frequency of the haplotype 1111111 (**supplementary Fig. S3**).

Subsequently, we investigated the potential functional impact of the haplotype. The overall 400K haplotype block encompasses a single gene, the poly-ADP-ribosyltransferase 1 (*TNKS*). Three out of the seven SNPs that revealed selection signatures were located upstream of the *TNKS* while the other four were within the introns of the gene (**Fig. 3** and **4C**). The variation in the gene has been associated with adiposity (Lindgren et al. 2009) and asthma (Ober et al. 2000) previously. Indeed, when we searched the UK Biobank database, we found that the specific SNPs within the putatively

bioRxiv preprint doi: https://doi.org/10.1101/765651; this version posted September 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

selected haplotype that we identified using our approach were indeed associated with metabolic disorders, for example, body mass index and limb fat mass were associated with adiposity and eosinophil percentage was associated with asthma (George 2005) (**Fig. 4A** and **supplementary Table S6**). Therefore, both studies published previously, and a reanalysis of UK Biobank dataset revealed that the specific *TNKS* haplotype is potentially associated with both metabolic traits and asthma.

In addition to verifying previous associations, we observed that six out of seven putatively selected SNPs in the region were associated significantly with hypertension. To the best of our knowledge, this is the first report linking the region to hypertension, and considering hypertension is associated with high levels of adiposity (obesity) (Beevers et al. 2001), it further implicates the variation in *TNKS* in the pleiotropic effects on metabolism at the organismal level. In addition, expression data of *TNKS* revealed that depending on which of the seven SNPs of the haplotype a given individual carries, it is upregulated mainly in the spinal cord and tibial nerve, besides the transformed fibroblasts (**Fig. 4B**). According to existing data based on expression quantitative trait loci, *TNKS* is expressed globally and mainly in brain-related tissues. It is critical to note here that most association studies that link genetic to phenotypic variation have been conducted in peoples with European descent; therefore, they have a limited capacity to capture population–specific associations that are currently widespread comprehensively. However, considering the derived haplotype is extremely common in multiple populations, a more general impact could be deduced from already available databases.

Adaptation to faster metabolism and fitness advantage in Kuwaiti ancestors

We identified gene regions that presented strong signals of positive selection in Kuwaiti populations using an integrative approach. Our decision to focus only on putatively selected SNPs that were identified based on multiple tests of selection made our approach highly conservative and, therefore, not prone to a high false-positive rate. For example, some of the top candidates identified using PBS (not exhibiting adequately high signals in other selection tests) cluster into a haplotype block that included *LCT*. The result was not surprising since lactose tolerance is one of the most extensively studied

bioRxiv preprint doi: https://doi.org/10.1101/765651; this version posted September 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

adaptive traits and it has been published elsewhere that *LCT* haplotypes were selected for in the Middle Eastern populations (Enattah et al. 2008; Bayoumi et al. 2016; Liebert et al. 2017). Therefore, by providing data from multiple selection tests, our study establishes a robust framework for the generation of additional adaptive hypotheses for Middle Eastern populations.

In the present study, we chose to highlight one 400-kb haplotype that was detected to be under positive selection based on multiple tests. The haplotype encompasses *TNKS* and we verified its association with diabetes. In addition, we found that the haplotype is associated with hypertension and an increase in *TNKS* expression in the nervous system. The harsh desert climate of the Kuwait region could have driven the selection (Weder 2007; Young 2007). For example, a recent study speculated that natural selection for insulin resistance and the associated hypertension, in addition to increased activity in the sympathetic nervous system could have been beneficial in hunter-gatherer populations conferring a hemodynamic advantage (Lewis et al. 2019). The pleiotropic effects of the *TNKS* haplotypes are consistent with the speculative phenotypes described in the present study. Notably, other gene regions under potential selection (**Table 1**) have also been associated with obesity and hypertension (**supplementary Text** and **supplementary Table S5**). Therefore, it is plausible that the *TNKS* haplotype exemplifies a general trend in which a more rapid metabolism rate and hypertension have been selected in the Kuwaiti population, which increased the allele frequency of multiple haplotypes and conferred some degree of fitness advantage to ancestors of present-day Kuwaiti populations in the extremely dry and hot ecological environments.

Is this selection a major health concern to present-day Kuwaiti population?

In modern Kuwait, however, the effect of the *TNKS* haplotype is potentially detrimental. Indeed, hypertension and obesity are prevalent in the Kuwaiti population, affecting a staggering 25.3% and 48.2% of the population, respectively (Al-Sejari 2018). The World Health Organization has estimated that the mortality rate in Kuwait due to noncommunicable diseases is approximately 72% (https://www.who.int/nmh/publications/ncd-profiles-2018/en/), which is alarmingly high.

bioRxiv preprint doi: https://doi.org/10.1101/765651; this version posted September 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Such mortality and morbidity levels could be attributed largely to the drastic changes in the lifestyles and behaviors associated with westernization following oil discovery. Nevertheless, our results suggest that past adaptive trends have further predisposed Kuwaiti populations to the illnesses above at the genetic level. Overall, the mechanisms through which the *TNKS* haplotype conferred a fitness advantage and how the same haplotype predisposes the population to metabolic diseases remain fascinating areas that could be explore in future research.

Materials and Methods

We genotyped 583 healthy unrelated Kuwaiti individuals from the State of Kuwait using Illumina HumanOmniExpress arrays for 730,525 SNPs. The quality control checks and data filtering were executed in the PLINK command-line program (Chang et al. 2015). The dataset was filtered through standard QC filtering to include only autosomal SNPs, specifically those that had a genotyping success rate greater than 90% and passed the Hardy-Weinberg Exact test with a p-value greater than 0.001. In addition, we eliminated strand-based ambiguous A/T and G/C SNPs. After the above filtering steps, about 662,750 SNPs remained for downstream analyzes (**Fig. 1B**). The SNPs were then phased using Beagle (Browning and Browning 2007).

Collectively, four different selection tests that measure deviations from expected linkage disequilibrium/homozygosity (iHS, XP-EHH) and population differentiation (PBS, LLRS) distributions across the genome were performed. Each of the tests has different statistical power to detect signatures of slightly different types of selection (e.g., complete vs. incomplete sweeps) and are sensitive to the time of selection (Cadzow et al. 2014). Therefore, the integration of multiple approaches facilitated the comprehensive interrogation of the genome for selection signatures and offered a list of candidate regions that could have been evolving under non-neutral evolutionary forces.

We implemented integrated Haplotype Score (iHS) (Voight et al. 2006) and Cross Population Extended Haplotype Homozygosity (XP-EHH) (Sabeti et al. 2007) tests available in selscan v1.2.0a (Szpiech and Hernandez 2014). Both the iHS and XP-EHH

bioRxiv preprint doi: https://doi.org/10.1101/765651; this version posted September 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

tests were performed in the phased dataset to capture haplotype homozygosity-based signals of positive selection. The XP-EHH test was conducted by comparing the Kuwaiti population with the Utah Residents with Northern and Western European Ancestry-CEU; Han Chinese from Beijing, China-CHB; and Yoruba from Ibadan, Nigeria-YRI, available from the 1000 Genomes Project Phase 3 dataset (Auton et al. 2015).

In addition to the haplotype-level analyses, we employed population-differentiation based tests, including PBS as described elsewhere (Yi et al. 2010). For this, we first computed SNP-wise F_{ST} values for all pairwise comparisons between populations (KWT, CEU, CHB, and YRI) in PLINK (Chang et al. 2015). The F_{ST} values were then used to calculate PBS scores in three ways: (i) between Kuwaiti and CEU using YRI as an outgroup; (ii) between Kuwaiti and CHB using YRI as an outgroup, and (iii) between Kuwaiti and YRI using CHB as an outgroup. The PBS scores were expected to pinpoint loci under selection exclusively in the Kuwaiti population.

Finally, we computed log-likelihood ratio scores for positive selection (LLRS) using Ohana (Cheng et al. 2019). This is a recently published selection detection framework that identifies signals of positive selection through population differentiation independent of self-reported ancestry or admixture correction to group individuals into populations.

Gene functions were determined using the UCSC Genome Browser (http://genome.ucsc.edu/) and literature searches. We performed Gene Ontology enrichment analysis using an online tool (http://pantherdb.org/) integrated with the PANTHER classification system (Mi et al. 2019). The association of the putatively selected SNPs with any phenotypic traits were detected using the PheWAS tool of the Gene Atlas database (Canela-Xandri et al. 2018). In addition, the expression patterns of selected genes were detected using the GTEx Portal (accessed on 05/03/19).

The maps presented in Figure 1A were created in R using *rworldmap*, *maptools*, and the corresponding associated tools; Fig. 1C, 2, 3, 4C, supplementary Fig. S1, and supplementary Fig. S3 were created using Python and R packages (*pandas*, *seaborn*,

bioRxiv preprint doi: https://doi.org/10.1101/765651; this version posted September 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

and *matplotlib*; and *venn*, *chromomap*, and *ggplot2*, respectively); Fig. 4A, and 4C were obtained from the GTEx Portal (accessed on 05/03/19); haplotype network Figure S2 was created in PopART (Leigh and Bryant 2015). All the codes related to creation of figures and analyzes are available in https://www.dropbox.com/home/Kuwait_Selection_Codes.

Acknowledgments

The work was supported by the Kuwait Foundation for the Advancement of Sciences research grant (RA 2015-022). We thank the members of the National Dasman Diabetes BioBank Core Facility for sample processing and DNA extraction. We are thankful to Prashantha Hebbar for processing the raw genotype data.

Author Contributions

T.A.T., O.G., and M.E. designed the study; M.E., and A.L.C.S., conducted most analyses; O.G., contributed significantly to the interpretation of the results; M.E., A.L.C.S., and O.G. wrote the main paper; T.A.T., and F.A-M. contributed to the writing of the paper; F.A-M. provided required resources, critically reviewed, and approved the paper; All authors reviewed the paper.

Competing interest

The authors have no competing interests.

bioRxiv preprint doi: https://doi.org/10.1101/765651; this version posted September 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

Abu-Amero KK, González AM, Larruga JM, Bosley TM, Cabrera VM. 2007. Eurasian and African mitochondrial DNA influences in the Saudi Arabian population. BMC Evol. Biol. 7:32.

Abu-Amero KK, Hellani A, González AM, Larruga JM, Cabrera VM, Underhill PA. 2009. Saudi Arabian Y-Chromosome diversity and its relationship with nearby regions. BMC Genet. 10:59.

Abu-Amero KK, Larruga JM, Cabrera VM, González AM. 2008. Mitochondrial DNA structure in the Arabian Peninsula. BMC Evol. Biol. 8:45.

Alenizi M, Goodwin W, Ismael S, Hadi S. 2008. STR data for the AmpFlSTR Identifiler loci in Kuwaiti population. Leg. Med. 10:321–325.

Alghanim S. 1998. The Reign of Mubarak-Al-Sabah: Shaikh of Kuwait 1896-1915. I.B.Tauris

Al-Sejari M. 2018. Impact of Attending the Diwaniyyah on Kuwaiti Men's Health. Am. J. Mens. Health 12:380–387.

Alsmadi O, John SE, Thareja G, Hebbar P, Antony D, Behbehani K, Thanaraj TA. 2014. Genome at juncture of early human migration: a systematic analysis of two whole genomes and thirteen exomes from Kuwaiti population subgroup of inferred Saudi Arabian tribe ancestry. PLoS One 9:e99069.

Alsmadi O, Thareja G, Alkayal F, Rajagopalan R, John SE, Hebbar P, Behbehani K, Thanaraj TA. 2013. Genetic substructure of Kuwaiti population reveals migration history. PLoS One 8:e74913.

Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis, Gonçalo R, 1000 Genomes Project Consortium. 2015. A global reference for human genetic variation. Nature 526:68–74.

Bayoumi R, De Fanti S, Sazzini M, Giuliani C, Quagliariello A, Bortolini E, Boattini A, Al-Habori M, Al-Zubairi AS, Rose JI, et al. 2016. Positive selection of lactase persistence among people of Southern Arabia. Am. J. Phys. Anthropol. 161:676–684.

Beevers G, Lip GY, O'Brien E. 2001. ABC of hypertension. Blood pressure measurement. Part I-sphygmomanometry: factors common to all techniques. BMJ 322:981–985.

bioRxiv preprint doi: https://doi.org/10.1101/765651; this version posted September 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Behar DM, Yunusbayev B, Metspalu M, Metspalu E, Rosset S, Parik J, Rootsi S, Chaubey G, Kutuev I, Yudkovsky G, et al. 2010. The genome-wide structure of the Jewish people. Nature 466:238–242.

Browning SR, Browning BL. 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am. J. Hum. Genet. 81:1084–1097.

Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O'Connell J, et al. 2018. The UK Biobank resource with deep phenotyping and genomic data. Nature 562:203–209.

Cabrera VM, Abu-Amero KK, Larruga JM, González AM. 2010. The Arabian Peninsula: Gate for Human Migrations Out of Africa or Cul-de-Sac? A Mitochondrial DNA Phylogeographic Perspective. In: Petraglia MD, Rose JI, editors. The Evolution of Human Populations in Arabia: Paleoenvironments, Prehistory and Genetics. Dordrecht: Springer Netherlands. p. 79–87.

Cadzow M, Boocock J, Nguyen HT, Wilcox P, Merriman TR, Black MA. 2014. A bioinformatics workflow for detecting signatures of selection in genomic data. Front. Genet. 5:293.

Canela-Xandri O, Rawlik K, Tenesa A. 2018. An atlas of genetic associations in UK Biobank. Nat. Genet. 50:1593–1599.

Cardona A, Pagani L, Antao T, Lawson DJ, Eichstaedt CA, Yngvadottir B, Shwe MTT, Wee J, Romero IG, Raj S, et al. 2014. Genome-wide analysis of cold adaptation in indigenous Siberian populations. PLoS One 9:e98076.

Casey MS. 2007. The History of Kuwait. Greenwood Publishing Group

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4:7.

Cheng JY, Racimo F, Nielsen R. 2019. Ohana: detecting selection in multiple populations by modelling ancestral admixture components. bioRxiv [Internet]:546408. Available from: https://www.biorxiv.org/content/10.1101/546408v1

Enattah NS, Jensen TGK, Nielsen M, Lewinski R, Kuokkanen M, Rasinpera H, El-Shanti H, Seo JK, Alifrangis M, Khalil IF, et al. 2008. Independent introduction of two lactase-

bioRxiv preprint doi: https://doi.org/10.1101/765651; this version posted September 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

persistence alleles into human populations reflects different history of adaptation to milk culture. Am. J. Hum. Genet. 82:57–72.

Fernandes V, Brucato N, Ferreira JC, Pedro N, Cavadas B, Ricaut F-X, Alshamali F, Pereira L. 2019. Genome-Wide Characterization of Arabian Peninsula Populations: Shedding Light on the History of a Fundamental Bridge between Continents. Mol. Biol. Evol. 36:575–586.

George RB. 2005. Chest Medicine: Essentials of Pulmonary and Critical Care Medicine. Lippincott Williams & Wilkins

Groucutt HS, Petraglia MD. 2012. The prehistory of the Arabian Peninsula: deserts, dispersals, and demography. Evol. Anthropol. 21:113–125.

Hunter-Zinck H, Musharoff S, Salit J, Al-Ali KA, Chouchane L, Gohar A, Matthews R, Butler MW, Fuller J, Hackett NR, et al. 2010. Population genetic structure of the people of Qatar. Am. J. Hum. Genet. 87:17–25.

John SE, Antony D, Eaaswarkhanth M, Hebbar P, Channanath AM, Thomas D, Devarajan S, Tuomilehto J, Al-Mulla F, Alsmadi O, et al. 2018. Assessment of coding region variants in Kuwaiti population: implications for medical genetics and population genomics. Sci. Rep. 8:16583.

John SE, Thareja G, Hebbar P, Behbehani K, Thanaraj TA, Alsmadi O. 2015. Kuwaiti population subgroup of nomadic Bedouin ancestry—Whole genome sequence and analysis. Genomics Data 3:116–127.

Leigh JW, Bryant D. 2015. popart: full-feature software for haplotype network construction. Methods Ecol. Evol. 6:1110–1116.

Lewis MT, Lujan HL, Wiseman RW, DiCarlo SE. 2019. The hypertension advantage and natural selection: Since type 2 diabetes associates with co-morbidities and premature death, why have the genetic variants remained in the human genome? Med. Hypotheses 129:109237.

Liebert A, López S, Jones BL, Montalva N, Gerbault P, Lau W, Thomas MG, Bradman N, Maniatis N, Swallow DM. 2017. World-wide distributions of lactase persistence alleles and the complex effects of recombination and selection. Hum. Genet. 136:1445–1453.

Lienhardt P. 2001. Shaikhdoms of Eastern Arabia. (Al-Shahi A, editor.). Palgrave

bioRxiv preprint doi: https://doi.org/10.1101/765651; this version posted September 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Macmillan, London

Lindgren CM, Heid IM, Randall JC, Lamina C, Steinthorsdottir V, Qi L, Speliotes EK, Thorleifsson G, Willer CJ, Herrera BM, et al. 2009. Genome-wide association scan meta-analysis identifies three Loci influencing adiposity and fat distribution. PLoS Genet. 5:e1000508.

Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. 2019. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic Acids Res. 47:D419–D426.

Ober C, Tsalenko A, Parry R, Cox NJ. 2000. A Second-Generation Genomewide Screen for Asthma-Susceptibility Alleles in a Founder Population. Am. J. Hum. Genet. 67:1154–1162.

Petraglia MD, Alsharekh A. 2003. The Middle Palaeolithic of Arabia: Implications for modern human origins, behaviour and dispersals. Antiquity 77:671–684.

Petraglia MD, Breeze PS, Groucutt HS. 2019. Blue Arabia, Green Arabia: Examining Human Colonisation and Dispersal Models. In: Rasul NMA, Stewart ICF, editors. Geological Setting, Palaeoenvironment and Archaeology of the Red Sea. Cham: Springer International Publishing. p. 675–683.

Rose JI, Petraglia MD. 2010. Tracking the Origin and Evolution of Human Populations in Arabia. In: Petraglia MD, Rose JI, editors. The Evolution of Human Populations in Arabia: Paleoenvironments, Prehistory and Genetics. Dordrecht: Springer Netherlands. p. 1–12.

Rowold DJ, Luis JR, Terreros MC, Herrera RJ. 2007. Mitochondrial DNA geneflow indicates preferred usage of the Levant Corridor over the Horn of Africa passageway. J. Hum. Genet. 52:436–447.

Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie X, Byrne EH, McCarroll SA, Gaudet R, et al. 2007. Genome-wide detection and characterization of positive selection in human populations. Nature 449:913–918.

Scott EM, Halees A, Itan Y, Spencer EG, He Y, Azab MA, Gabriel SB, Belkadi A, Boisson B, Abel L, et al. 2016. Characterization of Greater Middle Eastern genetic variation for enhanced disease gene discovery. Nat. Genet. 48:1071–1076.

Slot B. 2003. Kuwait: the growth of a historic identity. London Centre of Arab Studies

bioRxiv preprint doi: https://doi.org/10.1101/765651; this version posted September 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J, Landray M, et al. 2015. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 12:e1001779.

Szpiech ZA, Hernandez RD. 2014. selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. Mol. Biol. Evol. 31:2824–2827.

Thareja G, John SE, Hebbar P, Behbehani K, Thanaraj TA, Alsmadi O. 2015. Sequence and analysis of a whole genome from Kuwaiti population subgroup of Persian ancestry. BMC Genomics 16:92.

Triki-Fendri S, Sánchez-Diz P, Rey-González D, Alfadhli S, Ayadi I, Ben Marzoug R, Carracedo Á, Rebai A. 2016. Genetic structure of the Kuwaiti population revealed by paternal lineages. Am. J. Hum. Biol. 28:203–212.

Voight BF, Kudaravalli S, Wen X, Pritchard JK. 2006. A map of recent positive selection in the human genome. PLoS Biol. 4:e72.

Weder AB. 2007. Evolution and hypertension. Hypertension 49:260–265.

Yang X, Al-Bustan S, Feng Q, Guo W, Ma Z, Marafie M, Jacob S, Al-Mulla F, Xu S. 2014. The influence of admixture and consanguinity on population genetic diversity in Middle East. J. Hum. Genet. 59:615–622.

Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZXP, Pool JE, Xu X, Jiang H, Vinckenbosch N, Korneliussen TS, et al. 2010. Sequencing of 50 human exomes reveals adaptation to high altitude. Science 329:75–78.

Young JH. 2007. Evolution of blood pressure regulation in humans. Curr. Hypertens. Rep. 9:13–18.

bioRxiv preprint doi: https://doi.org/10.1101/765651; this version posted September 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure legends

Fig. 1. Geographic location of the genotyped individuals and materials and methods applied. (A) All the genotyped individuals are residents of Kuwait in the Arabian Peninsula. (B) Schematic diagram of the methods, techniques, and filtering actions applied to the genomic data. (C) The thresholds applied to all the selection data to identify 385 single nucleotide polymorphisms specifically under positive selection in the Kuwait population.

Fig. 2. Regions in the autosomal DNA under positive selection in the Kuwait population. All the 385 SNPs under putative positive selection in the Kuwait population were assigned to 100-Kb windows in the autosomal DNA. Each colored bar is a 25-Mb-segment in the chromosome representing at least one 100-Kb window, and their colors reflect the number of SNPs under positive selection.

Fig. 3. Top regions under positive selection for the Kuwait population: chr8:9.3–9.7 Mb. The values of all the selection scores for all the SNPs around the region are presented in the line plots. All the selection scores for the seven SNPs under positive selection in the region are indicated by gray bars, while their positions relative to *TNKS* are indicated by red small bars in the model at the bottom.

Fig. 4. Phenotypic traits, gene expression, and haplotypes around *TNKS*. (A) PheWAS result for SNP rs13276086. In total, 17 phenotypic traits revealed an association above the threshold (-log10[p-value] > 8, red dashed line). (B) Expression of *TNKS*, mainly in the spinal cord and the tibial nerve (in addition to the transformed fibroblasts), depending on which of the seven SNPs is carried by an individual (numbers in red, according to C). (C) Haplotype clustermap for the Kuwait population in the chr8:9.3–9.7-Mb region. The haplotype composed of the alternate alleles (black) of the seven SNPs under positive selection in the region is the most common even within putative subgroups, KWS - Kuwaitis with putative Saudi ancestry (green); KWP - Kuwaiti with putative Persian ancestry (red); B: Kuwaitis with putative Bedouin ancestry (gray). Each row represents one of the 583 genotyped individuals.

bioRxiv preprint doi: https://doi.org/10.1101/765651; this version posted September 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 1. Regions under selection in the Kuwait population empirically ranked based on the number of SNPs under positive selection.

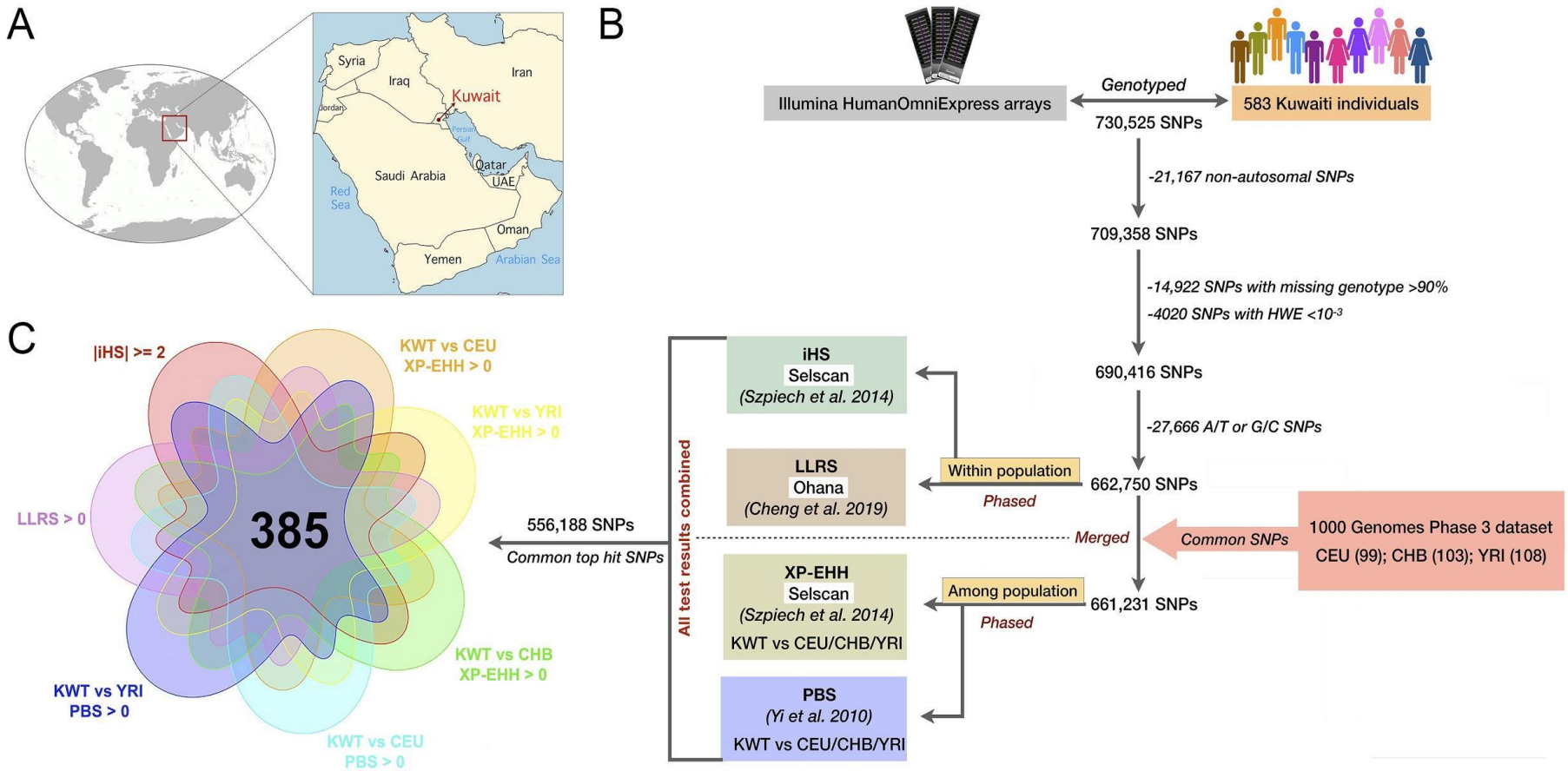
Chr	Region (Mbp)	Number of SNPs	Genes harboring SNPs	Number of Diseases and traits[a]	Number of Phenotypic Traits[b]
7	86,7–87,1	12	*BC035377*, *TMEM243*, *CROT*, *ABCB4*	2	N/A
11	83,7–84	12	*DLG2*	9	N/A
8	9,3–9,7	7	*TNKS*	2	17[c]
7	125,9–126,2	6	*GRM8*	7	N/A
11	45,6–45,7	6	*CHST1*	N/A	N/A
11	128,8–129	6	*TP53AIP1*, *ARHGAP32*	1	1
17	53,8–53,9	6	*PCTP*	1	N/A
16	81–81,2	5	*ATMIN*, *C16orf46*, *GCSH*	3	3

[a]Associated diseases and traits, according to UCSC Genome Browser;

[b]Threshold: -log10 (p-value) > 8, PheWas analysis provided through Gene Atlas;

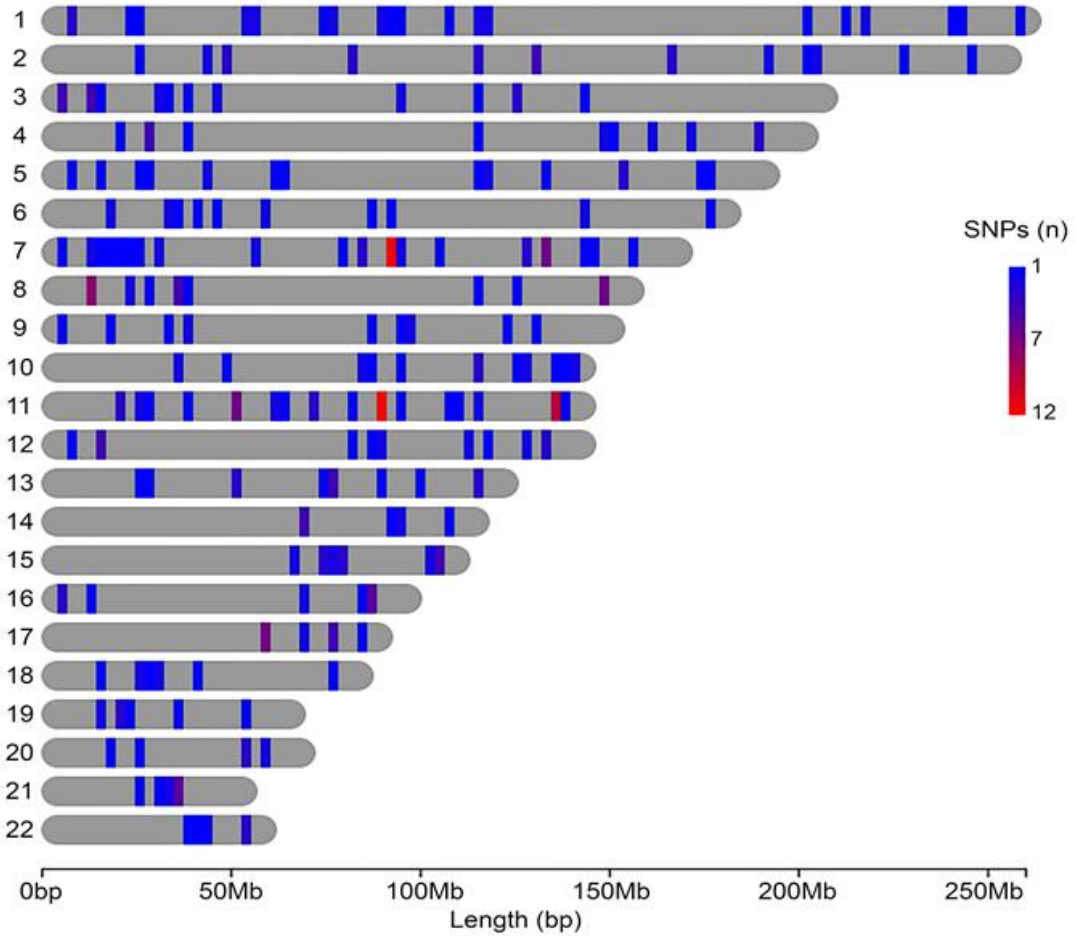
[c]Individual p-values are presented in Table S6.

Figure 1


A
Syria
Iraq
Iran
Kuwait
Jordan
Persian Gulf
Qatar
UAE
Saudi Arabia
Red Sea
Oman
Yemen
Arabian Sea
B
Illumina HumanOmniExpress arrays
Genotyped
583 Kuwaiti individuals
730,525 SNPs
-21,167 non-autosomal SNPs
709,358 SNPs
-14,922 SNPs with missing genotype >90%
-4020 SNPs with HWE <10-3
690,416 SNPs
-27,666 A/T or G/C SNPs
662,750 SNPs
Within population
Phased
Merged
Common SNPs
1000 Genomes Phase 3 dataset
CEU (99); CHB (103); YRI (108)
661,231 SNPs
Among population
Phased
All test results combined
iHS
Selscan
(Szpiech et al. 2014)
LLRS
Ohana
(Cheng et al. 2019)
XP-EHH
Selscan
(Szpiech et al. 2014)
KWT vs CEU/CHB/YRI
PBS
(Yi et al. 2010)
KWT vs CEU/CHB/YRI
556,188 SNPs
Common top hit SNPs
C
|iHS| >= 2
KWT vs CEU
XP-EHH > 0
KWT vs YRI
XP-EHH > 0
LLRS > 0
385
KWT vs CHB
XP-EHH > 0
KWT vs YRI
PBS > 0
KWT vs CEU
PBS > 0

bioRxiv preprint doi: https://doi.org/10.1101/765651; this version posted September 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 2



bioRxiv preprint doi: https://doi.org/10.1101/765651; this version posted September 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 3

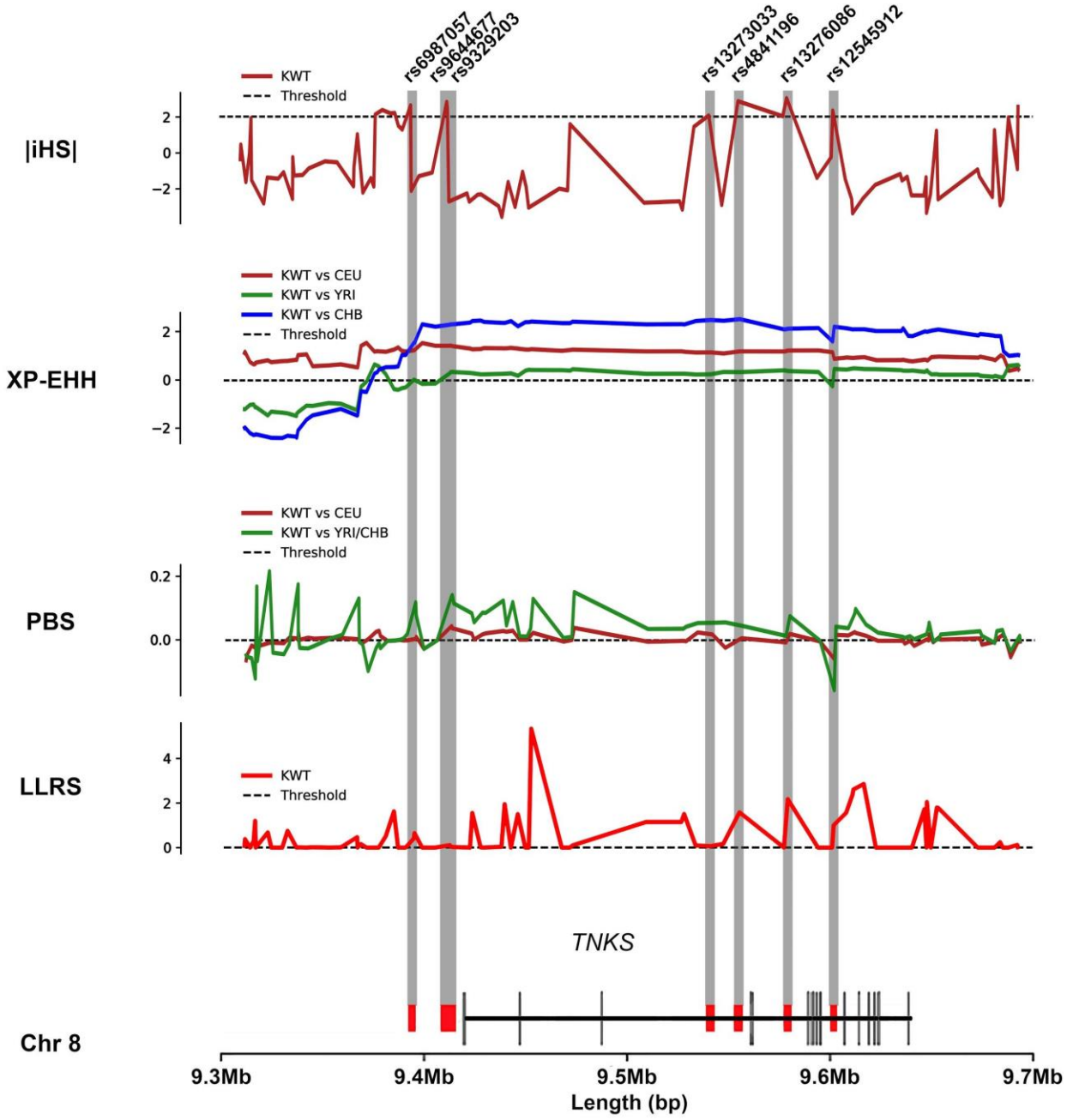
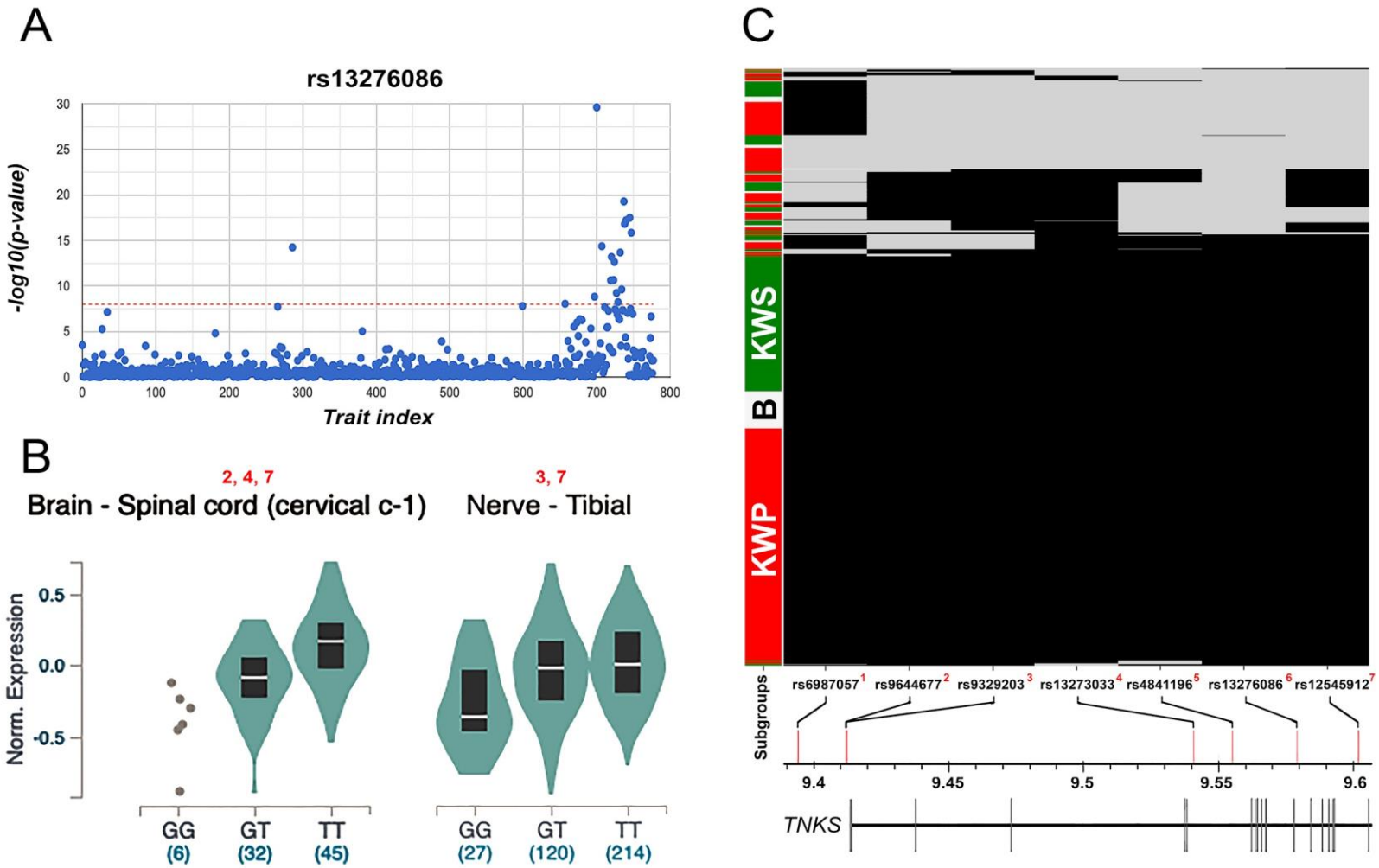


Figure 4


A
rs13276086
-log10(p-value)
Trait index
B
2, 4, 7
Brain - Spinal cord (cervical c-1)
3, 7
Nerve - Tibial
Norm. Expression
GG (6)
GT (32)
TT (45)
GG (27)
GT (120)
TT (214)
C
KWS
B
KWP
Subgroups
rs6987057 1
rs9644677 2
rs9329203 3
rs13273033 4
rs4841196 5
rs13276086 6
rs12545912 7
TNKS