Human Heredity

Hum Hered 2013;76:194–200
DOI: 10.1159/000360762

Published online: May 21, 2014

Implications of Population History of European Romani on Genetic Susceptibility to Disease

Isabel Mendizabal[a] Oscar Lao[b] Urko M. Marigorta[a] Manfred Kayser[b] David Comas[a]

[a]Institut de Biologia Evolutiva (CSIC-UPF), Departament de Ciències de la Salut i de la Vida, Universitat Pompeu Fabra, Barcelona, Spain; [b]Department of Forensic Molecular Biology, Erasmus MC University Medical Center Rotterdam, Rotterdam, The Netherlands

Key Words
Romani · Disease risk · Population history · Homozygosity · Founder population

Abstract
Objectives: The population history of European Romani is characterized by extensive bottleneck and admixture events, but the impact of this unique demographic history on the genetic risk for disease remains unresolved. ***Methods:*** Genome-wide SNP data on Romani, non-Romani Europeans and Indians were analyzed. The excess of homozygous variants in Romani genomes was assessed according to their potential functional effect. We also explored the frequencies of risk variants associated with five common diseases which are present at an increased prevalence in Romani compared to other Europeans. ***Results:*** Slightly deleterious variants are present at increased frequencies in European Romani, likely a result of relaxed purifying selection due to bottlenecks in their population history. The frequencies of SNPs associated with common metabolic and cardiovascular diseases are also increased compared to their European hosts. ***Conclusions:*** As observed in other founder populations, we confirm the impact of bottlenecks on the abundance of slightly deleterious variants in Romani groups, probably including metabolic and cardiovascular risk variants.

Introduction

Human populations that prefer consanguineous marriages or have undergone episodes of population bottlenecks and isolation typically show elevated levels of background parental relatedness [1]. High levels of autozygosity have detrimental effects, as it increases the genetic risk for disease [1]. The analysis of population isolates, i.e. those human groups that have experienced a founder effect and extensive genetic drift associated with bottlenecks during their population history, has allowed the discovery of genetic variants associated with Mendelian disorders and complex traits [2]. Some

I. M., O. L. and U.M. M. contributed equally to this work.
M. K. and D. C. contributed equally to this work.

KARGER

E-Mail karger@karger.com
www.karger.com/hhe

0001–5652/14/0764–0194$39.50/0

David Comas
Institut de Biologia Evolutiva
CSIC-Universitat Pompeu Fabra
ES–08003 Barcelona (Spain)
E-Mail david.comas@upf.edu

examples of such population studies are also on European isolates, e.g. Finns [3] and Sardinians [4], although their applicability to common diseases has been challenged [5].

One of the founder populations is the Romani (or Roma), also known by the misnomer Gypsies. Insight into the origins and demographic history of the Romani has been obtained from genetic data [6–12]. These studies have located the origins of the European Romani in Northwest India around 1,500 years ago and they have identified strong reductions in population sizes and endogamy during their population history. In addition, the Romani genomes show signatures of admixture with non-Romani Europeans as well as extensive footprints of recent inbreeding.

In agreement with this observation, the Romani have high prevalences of various Mendelian diseases caused by mutations in single genes that are rare among other populations [reviewed by Kalaydjieva et al. 13, 14]. These include rare neurological diseases – such as hereditary motor and sensory neuropathy Lom [15] or congenital cataracts facial dysmorphism neuropathy [16] – and other diseases such as primary congenital glaucoma [17, 18], galactokinase deficiency [19] or autosomal dominant polycystic kidney disease [20]. They are all caused by mutations that are private to several (sometimes geographically dispersed) Romani groups, denoting a clear common origin and founder effect [13, 14, 21]. Moreover, individuals of Romani ancestry present differences in their health status as compared to the majority of the European population [22, 23]. These differences might be environmental or genetic by origin or they exist due to social discrimination and poorer access to health care [24, 25], less prevention of noncommunicable diseases [26], different lifestyle-induced risks [27, 28] or, more likely, a combination of some or all of those. The demographic events in the Romani open up the possibility that their population history contributed partially to the disparity in the prevalence of complex diseases between Romani and other Europeans.

The efficiency of purifying selection is undermined in populations that faced episodes of extensive drift [29, 30]. In this study, we surveyed the impact of Romani demographic events (namely population bottlenecks and endogamy) on the genetic risk for diseases by evaluating the presence of low-frequency probably damaging variants in the Romani genomes in comparison to their parental populations from India and Europe. In addition, we studied the susceptibility of the European Romani to different complex diseases that are present at a higher prevalence than in non-Romani Europeans. We reason that the variation in the frequencies of risk alleles between the Romani and non-Romani Europeans may account for the differential genetic risk of disease(s).

Material and Methods

Data

We used genome-wide SNP data generated for 152 European Romani individuals from 13 European countries [9] (up to 807,002 SNPs, genotyped on an Affymetrix 6.0 array platform). The genotypes for the corresponding SNPs on HapMap 3 [31] GIH, TSI and CEU individuals were obtained from Ensembl (GRCh37.p12). We selected the other two populations – Europeans and Indians – because they have been shown to proxy the Indian and European parental ancestry components found in the European Romani [9].

The functional effects of the Affymetrix 6.0 SNPs were assigned by Ensembl. We retrieved the consequence to transcript information ('Exon: synonymous coding', 'Exon: nonsynonymous coding', 'Intron', '5′ UTR', '3′ UTR'). All SNPs not within these categories were considered 'Intergenic'. We also recovered the ancestral/derived status for every allele. The merged database including the genotypic data comprised 523,799 polymorphic SNPs. We additionally downloaded the PolyPhen-2 [32] and SIFT [33] functional predictions for the nonsynonymous SNPs, which are based on structural impact and conservation scores. In the case of PolyPhen-2, we merged the 'possibly damaging' and 'probably damaging' categories into a broader category of 'damaging' effect. Finally, we also obtained consensus deleteriousness (Condel) scores for the nonsynonymous SNPs [34]. This tool combines five different prediction methods and has been shown to outperform predictions based on individual tools. In total, 338, 280 and 143 nonsynonymous SNPs remained available for PolyPhen-2 ('possibly damaging' and 'probably damaging'), SIFT ('deleterious') and Condel ('deleterious') classifications, respectively.

Homozygosity Analyses

To weight the presence of slightly deleterious variants in the Romani compared to the parental populations, we computed the derived allele frequency (DAF) for each SNP. In the case of the Romani population, 2 different DAFs were calculated: (i) the frequency observed from the data (DAF_{OBS}) and (ii) the frequency expected given the parental derived frequencies (DAF_{MIX}). The parameters to compute the DAF_{MIX} were estimated by means of the lm function in R [35], fitting the following linear model: $DAF_{Romani} \sim DAF_{EUR} + DAF_{GIH}$, where EUR was either the TSI or CEU populations. The coefficients corresponding to the contributions of the European and Indian parental populations ($\beta_{TSI} = 0.63$ and $\beta_{GIH} = 0.37$; $\beta_{CEU} = 0.61$ and $\beta_{GIH} = 0.39$) agree with previous results [9]. Assuming those admixture proportions, we considered that any deviation from the DAF_{MIX} is probably due to genetic drift specific to the Romani population history and posterior to the admixture event [36]. Specifically, we compared the derived homozygous counts expected from the DAF_{MIX} (according to the Hardy-Weinberg equilibrium; HWE) to the counts observed in Romani genomes across functional categories. To test the significance of these deviations, we performed 10,000 random samplings and applied a permutation test based on drawing from the least functional category ('Intergenic') the same number of SNPs that were present in the tested category (e.g. 'damaging nonsynonymous SNPs').

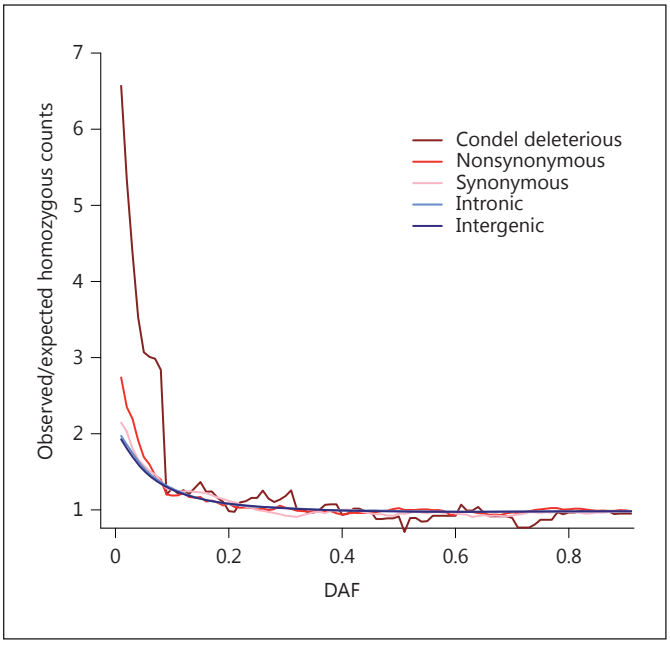
Since the European Romani constitute a heterogeneous population composed of highly differentiated endogamous groups, it can be expected that the mean allele frequency of a SNP computed between Romani populations resembles the one in the ancestral proto-Romani. However, considering all European Romani as a single uniform population can artificially increase the amount of homozygous counts observed (Wahlund effect). To evaluate the extent to which genetic substructure was affecting the number of observed homozygous counts in our joined Romani sample, we computed the ratio of the observed homozygous counts versus those expected from the DAF_{OBS}, assuming HWE. As a further validation, we also repeated the analysis considering more homogeneous Romani groups. Specifically, we chose the Bulgarian (n = 18) and Iberian Romani (n = 19), which are relatively homogeneous Romani groups located at the core and edges of the Romani migration route within Europe [9, 10], respectively. Thus they can be considered good representatives for the diversity of the European Romani.

Complex Disease-Associated SNPs

We used PubMed to retrieve studies that describe differences in the prevalence of disease(s) between the Romani and non-Romani Europeans. Although several diseases show a larger prevalence in Romani Europeans [37–41], we focused on five metabolic and cardiovascular disorders that are (still) more prevalent in Romani Europeans [42–44] even after controlling for several environmental risk factors [45]. In particular, we focused our analysis on insulin response, diabetes type 2, hyperlipidemia, the body mass index (obesity) and coronary artery disease. We aimed to test whether the differences in the average risk allele frequency (RAF) between the Romani and non-Romani Europeans followed the same pattern as that observed for prevalence (i.e. a larger prevalence as well as a larger RAF in the Romani population).

We selected the SNPs associated with disease as follows: first, we downloaded the NHGRI Catalog of Published Genome-Wide Association Studies (http://www.genome.gov/gwastudies/, accessed February 25, 2012). At that time, the Catalog contained records for up to 5,846 different SNPs present in 7,199 associations, with a p value threshold of 10^{-5} between a SNP and a given phenotype. As we aimed to test the differences in RAF among populations, we selected only those 3,964 SNPs (and 5,026 associations) for which information about the risk allele was available in the Catalog. Next, we checked the availability of the selected SNPs in the merged database of European Romani, CEU and GIH populations. In total, we recovered 1,491 SNPs which were assigned to six categories, corresponding to the five ascertained diseases and an additional (sixth) 'other disease' category that pooled SNPs associated with any other diseases (online suppl. table 1; for all online suppl. material, see www.karger.com/doi/10.1159/000360762). For genomic regions with more than one SNP associated with the same disease within a distance of 100 kb, we only considered the signal corresponding to the associations with the smallest p value (online suppl. table 3).

For each disease, we averaged the RAF and performed 3 pairwise population comparisons: (i) Romani versus CEU, (ii) Romani versus GIH and (iii) CEU versus GIH. The significance of each comparison was computed by a permutation test (10,000 random samplings). For each permutation, the same number of SNPs associated with each disease was randomly sampled from the 'other disease' category. Given that all five diseases are related to the 'metabolic syndrome' phenotype, we performed an additional test by pooling all five diseases together (n = 160 associated SNPs).



Color version available online

Fig. 1. Observed/expected homozygous counts in the Romani for different functional SNPs for different DAF windows. The expected frequency of homozygous counts was computed by assuming that the Romani DAF (DAF_{MIX}) was the result of admixture of European (TSI) and Indian (GIH) parental populations and assuming HWE. The lines represent running average levels in sliding windows of 0.02 DAF units with steps of 0.01.

A possible caveat of our approach is using allele frequencies from HapMap instead of those in the original GWAS, which are more reliable because of their larger sample sizes. We repeated the analyses above using 102 SNPs for which the RAF was available in the Catalog (13 and 45 SNPs were discarded because the value was missing or from a non-European population, respectively).

Results

Testing Excess of Homozygosity in Romani

We checked the efficiency of purifying selection in Romani genomes by testing whether functional SNPs (e.g. genic SNPs) presented an increase in DAF compared to the admixture expectation from their putative parental populations, i.e. non-Romani Europeans and Indians. For this aim, we used two proxy-parental populations (HapMap TSI and GIH) to compute the Romani DAF expected from the contribution of each parental gene pool to the Romani gene pool (DAF_{MIX}, see Material and Methods). This simple approach allowed us to obtain a rough estimate of the expected level of

Table 1. Mean RAF differences

Disease /trait	SNPs, n[a]	Average RAF			Test of differences in RAF[b]		
		European Romani	non-Romani Europeans	Indians	Δ_{R_E}	Δ_{R_I}	Δ_{I_E}
Coronary artery disease	33	0.55	0.531	0.535	0.019 (p = 0.19)	0.015 (p = 0.52)	0.004 (p = 0.28)
Diabetes type 2	39	0.503	0.474	0.506	0.029 (p = 0.06)	–0.003 (p = 0.80)	0.032 (p = 0.04)
Hyperlipidemia (triglycerides)	17	0.375	0.341	0.329	0.034 (p = 0.11)	0.046 (p = 0.19)	–0.012 (p = 0.52)
Insulin response	17	0.58	0.534	0.578	0.047 (p = 0.03)	0.003 (p = 0.64)	0.044 (p = 0.07)
Obesity (body mass index)	54	0.508	0.519	0.513	–0.011 (p = 0.89)	–0.004 (p = 0.87)	–0.007 (p = 0.41)
Pooled analysis	160	0.509	0.493	0.503	0.016 (p = 0.07)	0.006 (p = 0.85)	0.010 (p = 0.04)
'Other disease'	1,627	0.400	0.405	0.387	0.005 (–)	0.018 (–)	–0.013 (–)

Δ_{R_E}, Δ_{R_I} and Δ_{I_E} = Mean RAF differences between European Romani and non-Romani Europeans, between European Romani and Indians and between Indians and non-Romani Europeans, respectively.

[a] SNPs associated with each phenotype (see online suppl. table 3).

[b] Differences in the average RAFs between populations. The permutation p value after random resampling of all SNPs from the 'other disease' category is given in parentheses.

homozygosity for the derived alleles under a pure admixture model and assuming no Romani-specific genetic drift. For all tested functional categories, the excess of observed versus expected homozygous counts was only apparent at low DAF values (fig. 1). The excess increased according to the selective constrains on the SNP function, and it was maximal for the most detrimental category composed of deleterious nonsynonymous SNPs. The pattern was consistently replicated when considering another proxy for the European parental population (e.g. CEU, see online suppl. fig. 1) or when using the Romani DAF values from more homogeneous Romani groups, such as Bulgarian and Iberian Romani (online suppl. fig. 2). In addition, the results were robust to different functional classifications for nonsynonymous SNPs (Condel, SIFT and PolyPhen-2, see online suppl. fig. 3). The enrichment for deleterious variants was highly statistically significant (the test for the weakest signal, considering CEU and PolyPhen-2, provided a p value = 0.0001 after 10,000 permutations; online suppl. fig. 4).

An excess of homozygotes for the derived allele could result from an increase in allele frequencies in the Romani after the parental admixture event ($DAF_{OBS} > DAF_{MIX}$) or simply due to HWE deviations (e.g. even if $DAF_{OBS} = DAF_{MIX}$, the homozygous counts could increase due to a substructure in the Romani sample). To rule out the latter, we tested if the Romani showed an excess of homozygotes considering their actual DAF and HWE (DAF_{OBS}). Interestingly, the excess of observed over expected homozygous counts at lower DAFs was discrete and lacked a clear pattern across the functional categories (online suppl. fig. 5). This result indicates that the observed excess of homozygotes is probably due to a post-admixture increase in the DAF at rare SNPs ($DAF_{OBS} > DAF_{MIX}$) and not attributable to a substructure within the Romani sample.

Complex Disease-Associated SNPs in European Romani

The Catalog was used to ascertain SNPs associated with five metabolic and cardiovascular disorders that are present at a larger prevalence in the Romani versus non-Romani Europeans even after controlling for environmental factors [45]. To test a possible genetic origin of these differences, we computed and compared the average allele frequency of the ascertained risk alleles in the Romani populations to that in the parental European and Indian populations (using the CEU and GIH data from HapMap, respectively). The Romani showed a higher average RAF when compared to other Europeans for four of the considered cardiovascular/metabolic conditions (table 1), but the difference was only significant for insulin response ($\Delta_{RAF} = 0.047$, p value = 0.03). Considering these five phenotypes are all part of the metabolic syndrome, we pooled all associated SNPs. Overall, the genetic risks mimicked the pattern of a higher prevalence in the European Romani (p = 0.07; table 1).

A similar but weaker tendency was observed when comparing the European Romani to the Indians (three out of five conditions; table 1). To evaluate if the trend can be ascribed to the Indian origin of the European Romani, we performed the corresponding comparison between Indians and non-Romani Europeans. The Indian GIH population presented a higher average RAF for three of the five conditions and a significant tendency in the pooled analysis ($p = 0.04$; table 1). Finally, we repeated the same analysis but substituting the RAF values from the HapMap's CEU population with those reported in the Catalog (European discovery studies; see Material and Methods). Yet, the RAFs from HapMap were consistent with those from the actual papers (Spearman's $\rho = 0.992$, $p = 10^{-16}$; see online suppl. fig. 6), and the same patterns were observed (online suppl. table 2).

Discussion

The proportion of slightly deleterious genetic variants accumulates during bottleneck events as the efficiency of purifying selection is diminished in small populations. This phenomenon has shaped the extant distribution of detrimental variants across human populations, as it is apparent for events distant in time [29, 30]. Moreover, the increase in population sizes after bottleneck events fuels appearance of new mutations further, which can have a disparate effect on the proportion of slightly deleterious SNPs [46, but see also Simons et al. 47]. If this population growth has been recent, purifying selection may not have had enough time yet to adjust the proportion of nonsynonymous variants to the equilibrium of the new population size, as it has been shown recently for the founder population of French-Canadians [48].

Previous studies have described serial bottlenecks in the recent demographic history of the European Romani [6, 9, 10]. In addition, despite the low population sizes inferred for the proto-Romani prior to their arrival in Europe, the current Romani population in Europe is around 10,000,000, suggesting a considerable recent population growth [13]. Given this recent history as a founder population and probable recent population expansion, an excess of deleterious genetic variants at the lowest range of the allele frequency distribution is expected in the Romani compared to their parental populations from India and Europe. Since extensive homozygosity can result in a higher frequency of harmful recessive mutations [1], we evaluated possible medical implications due to the unique demographic history of the Romani.

The excess of homozygotes observed at low DAFs in the Romani population clearly surpasses the expectation under a pure admixture model. Nonetheless, the assumed simple admixture model has some limitations. First, we did not consider that admixture proportions vary among different Romani groups, as previously described [9]. However, the excess of homozygotes observed across different functional categories should not be affected by admixture heterogeneity. Second, although the pooling of the Romani groups could have resulted in an artificial excess of homozygous counts at RAFs, our analyses showed that any deviation from HWE could only account for a small fraction of the observed excess. Finally, the same results were obtained when more homogeneous Romani groups were analyzed.

Overall, our results indicate that rare alleles have drifted upwards in frequency in the Romani, including deleterious genetic variants; whereas purifying selection has maintained them at low frequencies in both parental populations. Consequently, a higher rate of homozygous possibly damaging variants is found in the Romani than in their parental populations. To which extent these alleles represent a putative health threat is unknown, and only future whole-genome or exome resequencing projects that include Romani individuals could evaluate the actual health risk.

Regarding complex diseases, in four out of five studied cardiovascular/metabolic conditions, the average RAFs of associated SNPs were higher in the European Romani than in non-Romani Europeans. This result indicates that the genetic risk for these conditions matches the known patterns of morbidity, suggesting that common risk alleles discovered by genome-wide association studies (GWAS; minor allele frequency >5%) might, at least partially, explain the higher disease prevalence in the Romani compared to other Europeans.

Noteworthy, signatures of recent positive selection have been found in some genes involved in the lipid metabolism and diabetes type 2 in Indians [49]. Considering that the same pattern of higher RAFs was observed in Indians compared to non-Romani Europeans, it is tempting to speculate that these disparities in the prevalence of metabolic diseases in the Romani may be explained by 'thrifty' genetic variants [50] of Indian origin that are detrimental under a Western lifestyle. However, one needs to be cautious, as the bulk of heritability for these diseases remains undiscovered.

A possible drawback of the latter analysis is that most SNPs associated with these diseases have been discovered in European populations. Therefore, this analysis could

be noninformative in case these SNPs do not participate in the disease etiology of Indians and Romani individuals. The actual extent of risk allele sharing cannot be tested until GWAS are performed on these populations directly. Nevertheless, multi-ethnic GWAS show a clear pattern of replication across populations of different ancestries [51–53], and a careful analysis [54], focused on 25 diseases, revealed that around 90% of all SNPs 'discovered' in Europeans replicate in East Asians, for instance. In any case, the extension of GWAS to populations of different ancestries than Europeans is urgent to properly address the putative differences in the genetic architecture of complex diseases across populations [55].

Conclusions

In this study, we explored the impact of the unique recent demographic history of the Romani to their potential genetic risk of rare and common diseases at the population level. We showed that genetic variants likely to be functional, such as nonsynonymous SNPs predicted to be deleterious, have drifted up in their allele frequencies in the Romani likely due to a relaxation in the purifying selection regime. Notably, the initial and subsequent serial bottleneck events in the history of the Romani population are likely to have shaped the genetic architecture of some cardiovascular/metabolic diseases that show a higher prevalence in the Romani. However, one needs to be cautious when extrapolating these results to actual individual risks of disease: complex diseases are affected by multiple genetic and nongenetic factors, and the population encloses only a fraction of the genetic risk. In this context, our study confirms the impact of population history on the distribution of moderately deleterious genetic variants in humans.

Acknowledgement

This study was supported by the Spanish Government grant CGL2010-14944/BOS.

References

1 McQuillan R, Leutenegger AL, Abdel-Rahman R, et al: Runs of homozygosity in European populations. Am J Hum Genet 2008;83:359–372.

2 Kristiansson K, Naukkarinen J, Peltonen L: Isolated populations and complex disease gene identification. Genome Biol 2008;9:109.

3 Peltonen L, Jalanko A, Varilo T: Molecular genetics of the Finnish disease heritage. Hum Mol Genet 1999;8:1913–1923.

4 Tenesa A, Wright AF, Knott SA, Carothers AD, Hayward C, Angius A, Persico I, Maestrale G, Hastie ND, Pirastu M, Visscher PM: Extent of linkage disequilibrium in a Sardinian sub-isolate: sampling and methodological considerations. Hum Mol Genet 2004;13:25–33.

5 Eaves IA, Merriman TR, Barber RA, Nutland S, Tuomilehto-Wolf E, Tuomilehto J, Cucca F, Todd JA: The genetically isolated populations of Finland and Sardinia may not be a panacea for linkage disequilibrium mapping of common disease genes. Nat Genet 2000;25:320–323.

6 Gresham D, Morar B, Underhill PA, Passarino G, Lin AA, Wise C, Angelicheva D, Calafell F, Oefner PJ, Shen P, Tournev I, de Pablo R, Kucinskas V, Perez-Lezaun A, Marushiakova E, Popov V, Kalaydjieva L: Origins and divergence of the Roma (gypsies). Am J Hum Genet 2001;69:1314–1331.

7 Gusmão A, Valente C, Gomes V, Alves C, Amorim A, Prata MJ, Gusmao L: A genetic historical sketch of European gypsies: the perspective from autosomal markers. Am J Phys Anthropol 2010;141:507–514.

8 Kalaydjieva L, Calafell F, Jobling MA, Angelicheva D, de Knijff P, Rosser ZH, Hurles ME, Underhill P, Tournev I, Marushiakova E, Popov V: Patterns of inter- and intra-group genetic diversity in the Vlax Roma as revealed by Y chromosome and mitochondrial DNA lineages. Eur J Hum Genet 2001;9:97–104.

9 Mendizabal I, Lao O, Marigorta UM, Wollstein A, Gusmao L, Ferak V, Ioana M, Jordanova A, Kaneva R, Kouvatsi A, Kucinskas V, Makukh H, Metspalu A, Netea MG, de Pablo R, Pamjav H, Radojkovic D, Rolleston SJ, Sertic J, Macek M Jr, Comas D, Kayser M: Reconstructing the population history of European Romani from genome-wide data. Curr Biol 2012;22:2342–2349.

10 Mendizabal I, Valente C, Gusmao A, Alves C, Gomes V, Goios A, Parson W, Calafell F, Alvarez L, Amorim A, Gusmao L, Comas D, Prata MJ: Reconstructing the Indian origin and dispersal of the European Roma: a maternal genetic perspective. PLoS One 2011;6:e15988.

11 Moorjani P, Patterson N, Loh PR, Lipson M, Kisfali P, Melegh BI, Bonin M, Kadasi L, Riess O, Berger B, Reich D, Melegh B: Reconstructing Roma history from genome-wide data. PLoS One 2013;8:e58633.

12 Pamjav H, Zalan A, Beres J, Nagy M, Chang YM: Genetic structure of the paternal lineage of the Roma people. Am J Phys Anthropol 2011;145:21–29.

13 Kalaydjieva L, Gresham D, Calafell F: Genetic studies of the Roma (gypsies): a review. BMC Med Genet 2001;2:5.

14 Kalaydjieva L, Morar B, Chaix R, Tang H: A newly discovered founder population: the Roma/gypsies. Bioessays 2005;27:1084–1094.

15 Kalaydjieva L, Gresham D, Gooding R, Heather L, Baas F, de Jonge R, Blechschmidt K, Angelicheva D, Chandler D, Worsley P, Rosenthal A, King RH, Thomas PK: N-myc downstream-regulated gene 1 is mutated in hereditary motor and sensory neuropathy-Lom. Am J Hum Genet 2000;67:47–58.

16 Angelicheva D, Turnev I, Dye D, Chandler D, Thomas PK, Kalaydjieva L: Congenital cataracts facial dysmorphism neuropathy (CCFDN) syndrome: a novel developmental disorder in gypsies maps to 18qter. Eur J Hum Genet 1999;7:560–566.

17 Plasilova M, Ferakova E, Kadasi L, Polakova H, Gerinec A, Ott J, Ferak V: Linkage of autosomal recessive primary congenital glaucoma to the GLC3A locus in Roms (Gypsies) from Slovakia. Hum Hered 1998;48:30–33.

18 Plasilova M, Stoilov I, Sarfarazi M, Kadasi L, Ferakova E, Ferak V: Identification of a single ancestral CYP1B1 mutation in Slovak Gypsies (Roms) affected with primary congenital glaucoma. J Med Genet 1999;36:290–294.

19 Kalaydjieva L, Perez-Lezaun A, Angelicheva D, Onengut S, Dye D, Bosshard NU, Jordanova A, Savov A, Yanakiev P, Kremensky I, Radeva B, Hallmayer J, Markov A, Nedkova V, Tournev I, Aneva L, Gitzelmann R: A founder mutation in the GK1 gene is responsible for galactokinase deficiency in Roma (Gypsies). Am J Hum Genet 1999;65:1299–1307.

20 Veldhuisen B, Saris JJ, de Haij S, Hayashi T, Reynolds DM, Mochizuki T, Elles R, Fossdal R, Bogdanova N, van Dijk MA, Coto E, Ravine D, Norby S, Verellen-Dumoulin C, Breuning MH, Somlo S, Peters DJ: A spectrum of mutations in the second gene for autosomal dominant polycystic kidney disease (PKD2). Am J Hum Genet 1997;61:547–555.

21 Morar B, Gresham D, Angelicheva D, et al: Mutation history of the Roma/gypsies. Am J Hum Genet 2004;75:596–609.

22 Carrasco-Garrido P, Lopez de Andres A, Hernandez Barrera V, Jimenez-Trujillo I, Jimenez-Garcia R: Health status of Roma women in Spain. Eur J Public Health 2011;21:793–798.

23 Hajioff S, McKee M: The health of the Roma people: a review of the published literature. J Epidemiol Community Health 2000;54:864–869.

24 Kosa Z, Szeles G, Kardos L, Kosa K, Nemeth R, Orszagh S, Fesus G, McKee M, Adany R, Voko Z: A comparative health survey of the inhabitants of Roma settlements in Hungary. Am J Public Health 2007;97:853–859.

25 Parekh N, Rose T: Health inequalities of the Roma in Europe: a literature review. Cent Eur J Public Health 2011;19:139–142.

26 Koupilova I, Epstein H, Holcik J, Hajioff S, McKee M: Health needs of the Roma population in the Czech and Slovak Republics. Soc Sci Med 2001;53:1191–1204.

27 Garcia de Cortazar AR, Cabrera Leon A, Hernan Garcia M, Jimenez Nunez JM: Attitudes of adolescent Spanish Roma toward noninjection drug use and risky sexual behavior. Qual Health Res 2009;19:605–620.

28 Niksic D, Kurspahic-Mujcic A: The presence of health-risk behaviour in Roma family. Bosn J Basic Med Sci 2007;7:144–149.

29 Fu W, O'Connor TD, Jun G, Kang HM, Abecasis G, Leal SM, Gabriel S, Rieder MJ, Altshuler D, Shendure J, Nickerson DA, Bamshad MJ, Project NES, Akey JM: Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. Nature 2013;493:216–220.

30 Lohmueller KE, Indap AR, Schmidt S, Boyko AR, Hernandez RD, Hubisz MJ, Sninsky JJ, White TJ, Sunyaev SR, Nielsen R, Clark AG, Bustamante CD: Proportionally more deleterious genetic variation in European than in African populations. Nature 2008;451:994–997.

31 Altshuler DM, Gibbs RA, Peltonen L, et al: Integrating common and rare genetic variation in diverse human populations. Nature 2010;467:52–58.

32 Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR: A method and server for predicting damaging missense mutations. Nat Methods 2010;7:248–249.

33 Kumar P, Henikoff S, Ng PC: Predicting the effects of coding non-synonymous variants on protein function using the sift algorithm. Nat Protoc 2009;4:1073–1081.

34 Gonzalez-Perez A, Lopez-Bigas N: Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. Am J Hum Genet 2011;88:440–449.

35 R Development Core Team: R: A Language and Environment for Statistical Computing. Vienna, R Foundation for Statistical Computing, 1991.

36 Long JC: The genetic structure of admixed populations. Genetics 1991;127:417–428.

37 Fernandez O, Fernandez V, Martinez-Cabrera V, Mayorga C, Alonso A, Leon A, Arnal C, Hens M, Luque G, de Ramon E, Caballero A, Leyva L: Multiple sclerosis in gypsies from southern Spain: prevalence, mitochondrial DNA haplogroups and HLA class II association. Tissue Antigens 2008;71:426–433.

38 Kálmán B, Takács K, Gyódi E, et al: Sclerosis multiplex in gypsies. Acta Neurol Scand 1991;84:181–185.

39 Karlinger K, Gyorke T, Mako E, Mester A, Tarjan Z: The epidemiology and the pathogenesis of inflammatory bowel disease. Eur J Radiol 2000;35:154–167.

40 Milanov I, Kmetski TS, Lyons KE, Koller WC: Prevalence of Parkinson's disease in Bulgarian Gypsies. Neuroepidemiology 2000;19:206–209.

41 Peterka M, Peterkova R, Likovsky Z: The lack of isolated palatal clefts in Czech Gypsies. Acta Chir Plast 2006;48:97–102.

42 Beljic Zivkovic T, Marjanovic M, Prgomelja S, Soldatovic I, Koprivica B, Ackovic D, Zivkovic R: Screening for diabetes among Roma people living in Serbia. Croat Med J 2010;51:144–150.

43 Hujova Z, Desatnikova J, Gabor D: Some cardiovascular risk factors in Gypsy children and adolescents from Central Slovakia. Bratisl Lek Listy 2009;110:233–239.

44 Macekova S, Bernasovsky I, Gabrikova D, Bozikova A, Bernasovska J, Boronova I, Behulova R, Svickova P, Petrejcikova E, Sotak M, Sovicova A, Carnogurska J: Association of the FTO rs9939609 polymorphism with obesity in Roma/Gypsy population. Am J Phys Anthropol 2012;147:30–34.

45 Vozarova de Courten B, de Courten M, Hanson RL, Zahorakova A, Egyenes HP, Tataranni PA, Bennett PH, Vozar J: Higher prevalence of type 2 diabetes, metabolic syndrome and cardiovascular diseases in gypsies than in non-gypsies in Slovakia. Diabetes Res Clin Pract 2003;62:95–103.

46 Kryukov GV, Pennacchio LA, Sunyaev SR: Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. Am J Hum Genet 2007;80:727–739.

47 Simons YB, Turchin MC, Pritchard JK, Sella G: The deleterious mutation load is insensitive to recent population history. Nat Genet 2014;46:220–224.

48 Casals F, Hodgkinson A, Hussin J, Idaghdour Y, Bruat V, de Maillard T, Grenier JC, Gbeha E, Hamdan FF, Girard S, Spinella JF, Lariviere M, Saillour V, Healy J, Fernandez I, Sinnett D, Michaud JL, Rouleau GA, Haddad E, Le Deist F, Awadalla P: Whole-exome sequencing reveals a rapid change in the frequency of rare functional variants in a founding population of humans. PLoS Genet 2013;9:e1003815.

49 Metspalu M, Romero IG, Yunusbayev B, Chaubey G, Mallick CB, Hudjashov G, Nelis M, Magi R, Metspalu E, Remm M, Pitchappan R, Singh L, Thangaraj K, Villems R, Kivisild T: Shared and unique components of human population structure and genome-wide signals of positive selection in South Asia. Am J Hum Genet 2012;89:731–744.

50 Shah AM, Tamang R, Moorjani P, Rani DS, Govindaraj P, Kulkarni G, Bhattacharya T, Mustak MS, Bhaskar LV, Reddy AG, Gadhvi D, Gai PB, Chaubey G, Patterson N, Reich D, Tyler-Smith C, Singh L, Thangaraj K: Indian Siddis: African descendants with Indian admixture. Am J Hum Genet 2011;89:154–161.

51 Shriner D, Adeyemo A, Gerry NP, Herbert A, Chen G, Doumatey A, Huang H, Zhou J, Christman MF, Rotimi CN: Transferability and fine-mapping of genome-wide associated loci for adult height across human populations. PLoS One 2009;4:e8398.

52 Waters KM, Henderson BE, Stram DO, Wan P, Kolonel LN, Haiman CA: Association of diabetes with prostate cancer risk in the multiethnic cohort. Am J Epidemiol 2009;169:937–945.

53 Waters KM, Stram DO, Hassanein MT, Le Marchand L, Wilkens LR, Maskarinec G, Monroe KR, Kolonel LN, Altshuler D, Henderson BE, Haiman CA: Consistent association of type 2 diabetes risk variants found in Europeans in diverse racial and ethnic groups. PLoS Genet 2010;6:e1001078.

54 Marigorta UM, Navarro A: High trans-ethnic replicability of GWAS results implies common causal variants. PLoS Genet 2013;9:e1003566.

55 Bustamante CD, Burchard EG, De la Vega FM: Genomics for the world. Nature 2011;475:163–165.