Rev Endocr Metab Disord (2016) 17:209–219
DOI 10.1007/s11154-016-9354-3

A Genome-Wide Association Study in isolated populations reveals new genes associated to common food likings

Nicola Pirastu[1,2] · **Maarten Kooyman**[3] · **Michela Traglia**[5] · **Antonietta Robino**[1] · **Sara M. Willems**[3] · **Giorgio Pistis**[5] · **Najaf Amin**[3] · **Cinzia Sala**[5] · **Lennart C. Karssen**[3,6] · **Cornelia Van Duijn**[3,4] · **Daniela Toniolo**[5] · **Paolo Gasparini**[1,2]

Published online: 30 April 2016

Abstract Food preferences are the first factor driving food choice and thus nutrition. They involve numerous different senses such as taste and olfaction as well as various other factors such as personal experiences and hedonistic aspects. Although it is clear that several of these have a genetic basis, up to now studies have focused mostly on the effects of polymorphisms of taste receptor genes. Therefore, we have carried out one of the first large scale (4611 individuals) GWAS on food likings assessed for 20 specific food likings belonging to 4 different categories (vegetables, fatty, dairy and bitter). A two-step meta-analysis using three different isolated populations from Italy for the discovery step and two populations from The Netherlands and Central Asia for replication, revealed 15 independent genome-wide significant loci ($p < 5 \times 10^{-8}$) for 12 different foods. None of the identified genes coded for either taste or olfactory receptors suggesting that genetics impacts in determining food likings in a much broader way than simple differences in taste perception. These results represent a further step in uncovering the genes that underlie liking of common foods that in the end will greatly help understanding the genetics of human nutrition in general.

Keywords Food preferences · Food consumption · Food choice · GWAS · Association study · Isolated populations

Electronic supplementary material The online version of this article (doi:10.1007/s11154-016-9354-3) contains supplementary material, which is available to authorized users.

✉ Nicola Pirastu
nicola.pirastu@burlo.trieste.it

[1] Institute for Maternal and Child Health - IRCCS "Burlo Garofolo", Trieste, Italy

[2] University of Trieste, Trieste, Italy

[3] Genetic Epidemiology Unit, Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands

[4] Centre for Medical Systems Biology, Leiden University Medical Center, Leiden, The Netherlands

[5] Division of Genetics and Cell Biology, San Raffaele Scientific Institute, Milano, Italy

[6] PolyOmica, Groningen, The Netherlands

1 Introduction

Diet and nutrition play a key role as risk factors for several chronic diseases such as diabetes, obesity, dyslipidemia and cardiovascular diseases [1]. It is well known that dietary choices differ among individuals and are influenced by physiologic, social, psychological and, most likely, genetic factors [2]. In the last two decades Genome-Wide Association Studies (GWAS) have led to the discovery of many genetic variants associated with chronic diseases. However, to date few studies have focused on the effect of human genetic differences on eating behavior and nutrient intake, and ultimately their impact on health outcomes. Recently, Tanaka and collaborators reported association of variants in the *FTO* gene with higher carbohydrate and lower fat consumption, independent of BMI, supporting the role of genetic variants on macronutrient consumption in humans [3].

A limitation of most studies is that they measure diet by using food-frequency or dietary records questionnaires. These classical intake measures all suffer from reporting biases and attenuation effects that can lead to inaccurate conclusions about diet-disease relationships [4].

In a developed society where food is readily available, food preferences are the first factor influencing food choice [5]. For

this reason, it has been proposed that food hedonics may be better predictors of health outcomes than food consumption, and thus may provide a good alternative to assess dietary intake. Recently, Duffy et al. have shown that food preferences, in particular for fat and fibers, predict adiposity and blood pressure better than their reported consumption health [6]. Similar results have been also obtained when looking at blood metabolites [7] and on children [8], children [8] demonstrating that preferences may be more reliable markers than food frequency questionnaires in estimating the impact of nutrition on health.

Finally, evidence from genetic studies has shown that food preferences are genetically determined, with high estimated heritability (up to 70 %) [9, 10].

In this light, understanding the genetic factors driving food preferences could be important for addressing how they affect food choices and intake and thus health parameters. Several studies have tried to link food liking to specific genetic variants, mostly looking at the effect of taste receptor genes [10–17].

Here, we report the results of a GWAS on liking of 20 different foods belonging to 4 different categories (vegetables, fatty, dairy and bitter) using three different isolated populations from Italy. Data have been replicated in other populations from The Netherlands and Central Asia.

2 Materials and methods

2.1 Study populations

Participants have been collected from Europe and Central Asia, including: 381 individuals from INGI-CARL recruited from Carlantino, a small village located in Puglia (Southern Italy); 744 from INGI-FVG, recruited from 6 villages situated in the Friuli-Venezia Region in Northern-Eastern Italy, and 1115 from INGI-VB, recruited from the Val Borbera Valley in Northern-Western Italy. DNA from 1261 individuals was collected in the Erasmus Rucphen Family (ERF) study, a cross-sectional cohort including 3000 living descendants of 22 couples who had at least 6 children baptized in the community church around 1850–1900. Finally, DNA samples from 335 individuals was obtained from the Silk Road (SR) [12] cohort of ~1000 individuals from 20 communities located along the Silk Road (Armenia, Azerbaijan, Georgia, Uzbekistan, Tajikistan and Kazakhstan).

2.2 Assessing food liking

Liking/disliking for 20 foods were ascertained through a questionnaire administered by an operator. The foods can be conceptually grouped into 4 categories: vegetables (artichokes, broccoli, chicory, spinach, and mushrooms), fatty (bacon, ham, lamb meat, oil or butter on bread, and pork chops), dairy (goat cheese, blue cheese, ice cream, mozzarella, plain yogurt and whole milk) and bitter (dark beer, dark chocolate, liver and coffee).

Participants were asked to rate their liking for each food on a scale ranging from 1 (dislike extremely) to 9 (like extremely) [18] or to indicate never having tasted the particular food. The SR population was in excepption because a 5-point facial hedonic scale was used. This scale is commonly used to minimize linguistic barriers or when working with (mostly) illiterate people as was the case of the SR population [19]. Given the differences in the two hedonic scales, the responses have been numerically standardized by dividing each score by the number of scale categories (*i.e.*, 9 for the European populations, 5 for the SR study). This standardization is the same as the "simple proportion method" described in Colman et al. 2007 [20] and similar to the equations used in Preston and Colman 2000 [21] and Dawes 2002 [22]. The standardization, while important for interpreting regression analyses and its intercept, would not influence the interpretation of the meta-analysis used in the present study. In meta-analysis we are interested only in effect sizes (betas) and since association testing is performed within each group, this difference is of no consequence. Finally, some foods (broccoli, chicory, bacon, oil or butter on bread, blue cheese, mozzarella cheese, and dark beer) were unknown in the Silk Road communities and consequently were not assessed in this population.

2.3 Genotyping and imputation

Genotyping was carried out as previously described [15, 23, 24]. Briefly, the Italian and Silk Road samples were genotyped with Illumina high density SNP arrays and data imputed, after standard QC, using SHAPEIT2 [25] for the phasing step and IMPUTE2 [26] for the imputation using the 1000 Genomes phase I v3 reference set [26]. ERF has been genotyped on several genotyping platforms: Illumina 318 k, 350 k, 610 k and Affymetrics 200 k. Genotypes were pooled after QC, phased and imputed to the 1000 Genomes phase I v3 reference set [27] using MaCH and minimac [27]. After imputation SNPs with MAF < 0.01 or IMPUTE2 Info metric < 0.4 were excluded from the statistical analyses for all populations, except for ERF for which minimac $R^2 < 0.3$ was used instead.

2.4 Association analysis

Association analysis was conducted using mixed model linear regression where the standardized food liking of individuals foods were the dependent variable and the SNP dosages as the independent variable. Sex and age were used as covariates. The kinship matrix based on all available genotyped SNPs

was used as the random effect. For ERF, the kinship matrix was estimated on 14.4 k SNPs common to the various genotyping platform used. Association analysis was conducted using the GRAMMAR+ method [28] as implemented in the GenABEL 1.7–2 [29] R package in order to eliminate the effect of familial relatedness from the trait. MixABEL [29] was used for the actual association of the imputed SNPs. SNPs that did not pass quality control for more than one population were discarded.

It is common practice in GWA studies to test for association presuming the genetic variants have an additive effect on the phenotype, however this is not necessarily true [30, 31]. Therefore, we decided to also use non-additive genetic models, in particular dominant, recessive and over-dominant. For the association analysis, a two-step joint replication approach was used. Association analysis was conducted separately for each INGI cohort and results pooled together using the inverse-variance weighting method. Given that no meta-analysis software support non-additive genetic models, we developed custom R scripts. After meta-analysis in the INGI cohorts, genomic control was used to eliminate any residual stratification and all SNPs, with $p < 1 \times 10^{-5}$ where taken forward to the replication step using the ERF and SR cohorts. We considered as replicated all SNPs with nominal p-value $< 5 \times 10^{-8}$ at the combined analysis and whose p-value was lower at the replication step compared to the discovery phase. These criteria are very similar to those used in a similar study on food consumption [3].

3 Results

Table 1 shows sex, age and food liking distributions for each of the participating cohorts. In general, the ranking of the different food groups was consistent across the different populations, with the fatty foods being the most liked and the bitter ones the least. Surprisingly, in all populations vegetables are preferred over dairy products. However, after removing the two strong-tasting cheeses (dairy w/o) the two categories switch ranks in all studies except INGI-CARL in which vegetables are still preferred over dairy products. All items in the bitter foods category had low mean scores except coffee which is among the highest ranking foods in all populations. All single foods showed a good variance with standard deviations ranging from 0.15 to 0.35. Variance is maximized for those foods whose mean is close to 0.5 while it decreases for those foods which have either higher or lower means.

Regarding the association analysis for the discovery step, we first performed a GWAS in each separate cohort and then pooled the results. As the first phase of the discovery step, only the Italian populations were analyzed. From these analyses, we subsequently selected all SNPs with nominal p-value after genomic control $< 1 \times 10^{-5}$. These SNPs were then used for replication in a joint meta-analysis.

Table S1 (See additional materials) lists the number of SNPs used for each analysis, lambda (genomic control) values, the number of SNPs brought forward for replication and the number of independent loci. QQ-plots and complete results for all SNP used for replication can be found in the supplemental material.

Overall, 15 replicated loci for the 20 analyzed traits were detected (see Table 2). Four out of the 15 discovered loci show an additive genetic effect, six an over-dominant one, two a dominant one, and three follow a recessive model. Two loci show significant associations under more than one genetic model: rs6661761, which is associated with “oil or butter on bread” (both additive and dominant models), and rs28849980, associated with artichokes (both over-dominant and additive models).

Seven loci were associated with vegetables. In particular, three with artichokes liking: rs10050951 on chromosome 5, located in between the *ADAMTS19* and *CHSY3* genes ($p = 4.5 \times 10^{-8}$), rs8034691, on chromosome 15 located within *LOC100128714* gene ($p = 1.9 \times 10^{-8}$), and rs28849980 on chromosome 4 close to the *CCRNL4* gene ($p = 4.4 \times 10^{-8}$). Two loci were identified for broccoli, one located in a gene desert region on chromosome 17 (rs2530184, $p = 4.5 \times 10^{-9}$) and the second one on chromosome 3 close to the *RYBP* gene (rs9832668, $p = 4.4 \times 10^{-8}$). Furthermore, we have observed a locus significantly associated with chicory liking on chromosome 8, very close to the *CSMD1* gene (rs138369603, $p = 2.6 \times 10^{-9}$). Finally, a locus associated with mushrooms liking close to C9orf123 (rs6477241, $p = 1.6 \times 10^{-8}$) was also detected.

In the group of “fatty foods,” two GWAS significant loci were identified—one for bacon liking (rs140738262, $p = 5.9 \times 10^{-9}$) and the second one, located within the *BPNT1* gene on chromosome 1, for “oil or butter on bread” liking (rs6661761, $p = 3.6 \times 10^{-10}$).

In the category “dairy food,” 3 loci were identified—one on chromosome 2, between the *KCMF1* and *TCF7L1* genes, associated with blue cheese liking (rs12994253, $p = 8.8 \times 10^{-9}$), one on chromosome 5, close to the *IRX4* gene associated with ice cream liking (rs2035613, $p = 3.9 \times 10^{-8}$), and one on chromosome 22, inside the *IGL* gene associated to plain yogurt liking (rs4239891, $p = 3.8 \times 10^{-8}$).

For “bitter foods,” we identified 3 loci, one for dark chocolate (rs73082019, $p = 4.1 \times 10^{-8}$), one for coffee (rs145671205, $p = 3.1 \times 10^{-8}$), and one for liver liking (rs34088951, $p = 3.4 \times 10^{-8}$). Figures 1, 2 and 3 show the regional plots of the most significant SNP in each locus.

Finally we verified whether any of the identified SNPs showed an effect on the other food likings. Figure 4 summarizes the results from this analysis. Clearly although for most of the identified SNPs we can detect weak association also to

Table 1 Descriptive statistics for each population

	Age	Male percentage	Artichokes	Broccoli	Chicory	Spinach
INGI-CARL	52.56 (17.26)	0.42	0.87 (0.20)	0.83 (0.24)	0.81 (0.26)	0.79 (0.25)
INGI-FVG	50.84 (15.80)	0.41	0.75 (0.25)	0.71 (0.27)	0.69 (0.28)	0.80 (0.21)
INGI-VB	53.23 (16.62)	0.37	0.77 (0.24)	0.64 (0.28)	0.60 (0.29)	0.75 (0.23)
ERF	47.4 (13.23	0.45	0.40 (0.26)	0.69 (0.21)	0.75 (0.20)	0.67 (0.24)
SR	39.12 (15.86)	0.41	0.55 (0.28)	NA (NA)	NA (NA)	0.72 (0.28)
	Mushrooms	Bacon	Ham	Lamb	Oil or Butter on Bread	Pork Chops
INGI-CARL	0.82 (0.22)	0.84 (0.22)	0.90 (0.13)	0.81 (0.26)	0.76 (0.24)	0.88 (0.17)
INGI-FVG	0.75 (0.24)	0.75 (0.22)	0.85 (0.16)	0.58 (0.31)	0.75 (0.24)	0.77 (0.21)
INGI-VB	0.77 (0.24)	0.76 (0.22)	0.83 (0.16)	0.60 (0.31)	0.80 (0.22)	0.75 (0.22)
ERF	0.50 (0.30)	0.71 (0.18)	0.74 (0.17)	0.58 (0.26)	0.71 (0.20)	0.76 (0.13)
SR	0.76 (0.28)	NA (NA)	0.74 (0.29)	0.83 (0.25)	NA (NA)	0.70 (0.31)
	Goat Cheese	Blue Cheese	Ice Cream	Mozzarella	Plain Yogurt	Whole Milk
INGI-CARL	0.65 (0.31)	0.64 (0.34)	0.85 (0.19)	0.84 (0.19)	0.58 (0.34)	0.69 (0.33)
INGI-FVG	0.58 (0.30)	0.70 (0.27)	0.87 (0.17)	0.79 (0.20)	0.66 (0.28)	0.74 (0.26)
INGI-VB	0.58 (0.31)	0.76 (0.24)	0.85 (0.20)	0.74 (0.23)	0.52 (0.30)	0.73 (0.28)
ERF	0.42 (0.27)	0.37 (0.26)	0.82 (0.15)	0.49 (0.26)	0.69 (0.21)	0.65 (0.22)
SR	NA (NA)	NA (NA)	0.87 (0.21)	NA (NA)	NA (NA)	0.79 (0.27)
	Dark Beer	Dark Chocolate	Liver	Coffee		
INGI-CARL	0.51 (0.32)	0.58 (0.32)	0.50 (0.32)	0.83 (0.18)		
INGI-FVG	0.51 (0.29)	0.72 (0.27)	0.55 (0.32)	0.81 (0.18)		
INGI-VB	0.43 (0.30)	0.76 (0.27)	0.56 (0.28)	0.82 (0.21)		
ERF	0.58 (0.25)	0.68 (0.25)	0.56 (0.27)	0.75 (0.21)		
SR	NA (NA)	0.74 (0.29)	0.69 (0.29)	0.76 (0.28)		
	Fatty	Vegetables	Dairy	Dairy w/o	Bitter	
INGI-CARL	0.83	0.82	0.71	0.74	0.53	
INGI-FVG	0.74	0.74	0.72	0.76	0.59	
INGI-VB	0.75	0.70	0.7	0.71	0.58	
ERF	0.7	0.60	0.57	0.66	0.61	
SR	0.76	0.68	NA	NA	0.63	

For each food the mean is reported, with the standard errors between parenthesis. Sex has been reported as the proportion of male subjects in each cohort. For food groups only means have been reported. Given that for SR a 5-point scale has been used values for all cohorts have been standardized as explained in the Materials and Methods section. All reported values refer to the scores after standardization. As can be seen from the table, the order of the groups is the same in all populations : Fatty, Dairy w/o, Vegetables, Dairy and Bitter. The only exception is INGI-CARL in which Vegetables are still preferred as compared to dairy products even after removing the strong tasting cheeses from the group (Dairy w/o). For the SR population it was impossible to establish a mean of Dairy products liking since only two items were in common with the other populations

foods similar to the one for which the SNP was identified, most of them seem to be rather specific and no significantly additional strong signals have been detected.

4 Discussion

Here, we report the first GWAS on reported liking for individual foods conceptually grouped as vegetables, fatty foods and bitter foods and beverages. We have successfully identified 15 new novel genetic variants that determine differences in liking of 12 different food by using three Italian populations for the discovery step and two additional populations from Netherland and Central Asia to confirm our results. Given the novelty of the topic of the present study, a detailed discussion will be mainly focused on the genes for which it was possible to hypothesize a link between its function and its role in food hedonics.

The strongest detected association was between the A allele of rs6661761 and lower liking "oil or butter on bread." This SNP is located within the *BPNT1* gene that is a magnesium-dependent phosphomonoesterase. Although the function of *BPNT1* is still unknown, it is widely expressed in the brain and is strongly inhibited by Lithium [32], a drug largely used in bipolar disorder treatment. Lithium has been shown to restore hedonic responses to palatable foods in rats conditioned

Table 2 The most significant SNP for each identified locus

Locus	SNP	Trait	Chr	Pos Mb	Coded/Other Allele	First step *p*	Combined *p*	Dir	Beta	Model
Bitter										
FIBIN	rs145671205	Coffee	11	27.0	C/T	1.15×10^{-6}	3.13×10^{-8}	———	–0.056	Overdominant
DFNA5	rs73082019	Dark Chocolate	7	24.8	G/A	8.38×10^{-6}	4.09×10^{-8}	+++++	0.063	Dominant
RNU6-66	rs34088951	Liver	19	46.8	T/C	4.21×10^{-6}	3.41×10^{-8}	+++?+	0.188	Recessive
Dairy										
KCMF1-TCF7L1	rs12994253	Blue Cheese	2	85.3	A/G	1.84×10^{-6}	8.81×10^{-9}	——	–0.078	Overdominant
IRX4	rs2035613	Ice Cream	5	2.0	T/C	6.54×10^{-7}	3.92×10^{-8}	———	–0.038	Dominant
IGL	rs4239891	Plain Yogurt	22	22.5	A/G	2.79×10^{-7}	3.81×10^{-8}	++++	0.054	Overdominant
Fatty										
CNTN5	rs140738262	Bacon	11	99.3	L/S	1.94×10^{-6}	5.93×10^{-9}	——	–0.042	Overdominant
BPNT1	rs6661761	Oil or butter on bread	1	220.3	A/G	3.10×10^{-6}	3.62×10^{-10}	——	–0.062	Additive
Vegetables										
CCRN4L	rs28849980	Artichokes	4	139.9	G/A	5.66×10^{-6}	4.40×10^{-8}	———	–0.052	Overdominant
ADAMTS19-CHSY3	rs10050951	Artichokes	5	129.2	G/T	1.21×10^{-6}	4.54×10^{-8}	+++++	0.031	Additive
LOC100128714	rs8034691	Artichokes	15	26.2	C/A	2.39×10^{-8}	1.93×10^{-8}	+++?+	0.040	Additive
NA	rs2530184	Broccoli	17	51.4	C/A	1.76×10^{-6}	4.50×10^{-9}	——	–0.048	Additive
RYBP	rs9832668	Broccoli	3	72.4	A/G	7.37×10^{-7}	4.36×10^{-8}	–?-	–0.127	Recessive
CSMD1	rs138369603	Chicory	8	2.7	C/T	3.20×10^{-6}	2.56×10^{-9}	++++	0.084	Overdominant
C9orf123	rs6477241	Mushrooms	9	7.8	G/C	2.31×10^{-7}	1.57×10^{-8}	+++++	0.062	Recessive

The "Locus" column shows the gene closest to the most significant SNP; the "SNP" column shows the name polymorphism; "Trait" lists the associated food liking; "Chr "the chromosome number; "Pos Mb" the position of the SNP expressed in mega-basepairs. For indel alleles, we have used L for the long allele and S for the shorter allele. "Dir" indicates direction of the effect in the different populations in the following order: INGI-CARL, INGI-FVG, INGI-VB, ERF and SR. The question mark indicates missing data

to anhedonic responses through the nucleus accumbens [33]. Our result potentially suggests that liking of "oil or butter on bread", might be linked to the reward of palatable foods trough the nucleus accumbens. There is a linear inverse correlation between the activation of this nucleus and obesity related traits [34]. Thus, *BPNT1* seems a good candidate for understanding the physiology underlying the liking of palatable foods and the activation of the reward system. It is unclear why we detect such an effect only on such a specific trait such as "oil or butter on bread", however we must consider that the range of examined foods is limited and further studies on broader sets of foods are needed to clarify this point.

With respect to the other fatty foods we have detected a significant association between bacon and rs140738262 inside the *CNTN5* gene. This gene is mostly expressed in the brain and has been associated with traits ranging from blood viscosity [35] to HDL cholesterol [36]. More interestingly, rs140738262 has been associated to *anorexia nervosa* in a recent GWAS [37]. Although the physiopathology of *anorexia nervosa* is extremely complex and not completely understood, it is clear that in patients suffering from this disease one of the deregulated mechanisms is the reward response linked to food. It is thus plausible that *CNTN5* gene is linked to fatty food liking through the reward related to palatable foods. This finding is not surprising considering the presence of a positive correlation between BMI and fat liking [6] and that obese people have been shown to have an hypo-functioning reward system in fMRI studies [34]. Looking at the pleiotropy data in Fig. 4 it seems that while *BPNT1* is specifically linked to "oil or butter on bread", *CNTN5* shows marginal association also with lamb ($p = 5.8\times10^{-3}$), pork chops ($p = 1.0\times10^{-3}$) and goat cheese ($p = 1.6\times10^{-3}$) suggesting that its role is linked to strong tasting fatty foods. Further studies are clearly necessary to clarify the role of this gene in determining food liking.

Another significant association for dairy foods was found between ice cream and the *IRX4* gene. Mutations in this gene, mostly expressed in the heart, have been correlated to cardiomyopathy [38, 39]. Nevertheless, *IRX4* and *IRX3* belong to the same gene family, which is responsible for the association between the non-coding *FTO* variants and obesity [40]. *IRX3* elicits its effect on obesity in the thalamus [40] by modulating the relationship between food reinforcement and energy intake [41]. This same brain area is activated with palatable foods, especially high sugar ones [42] as is the case with ice

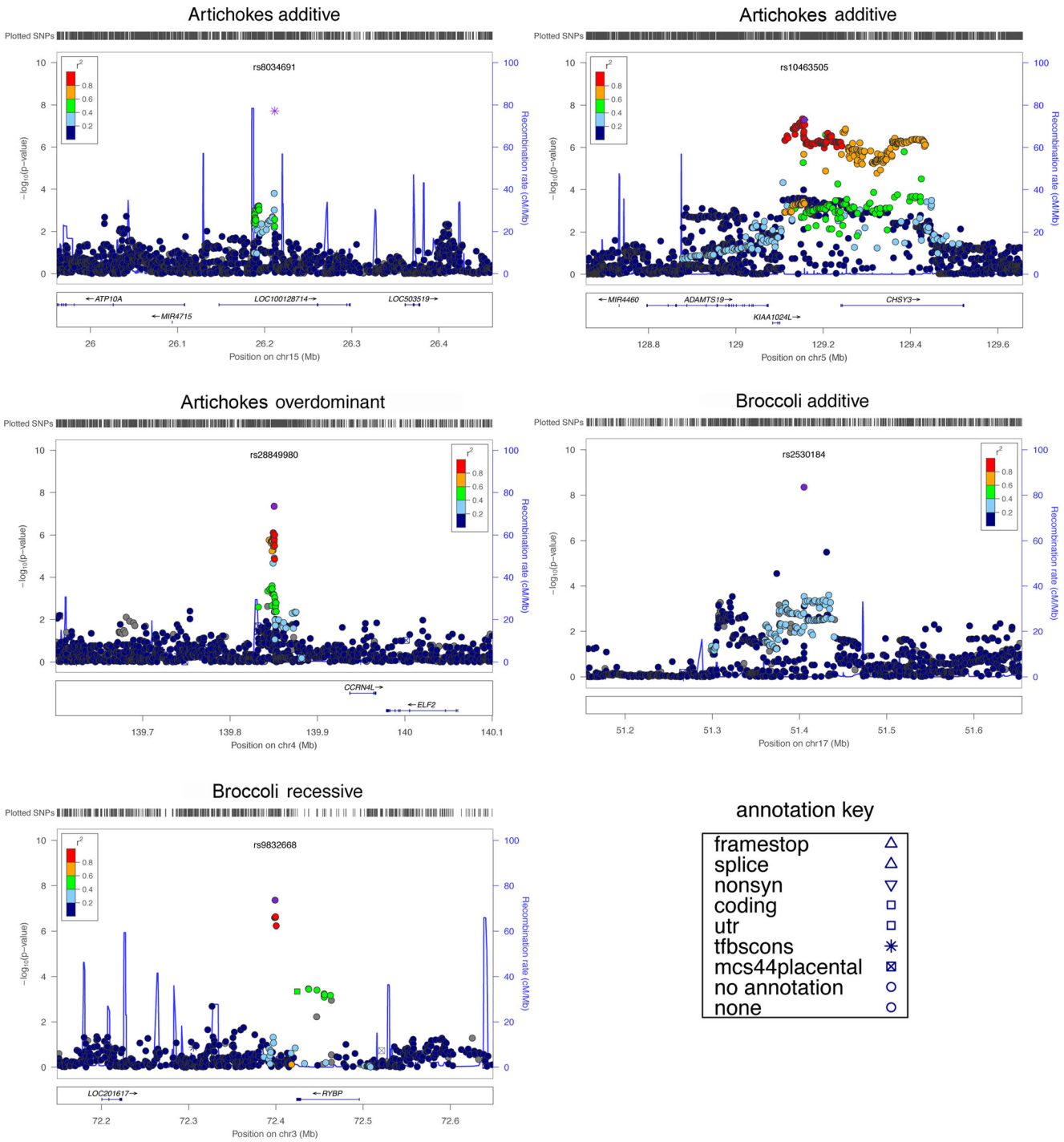

Fig. 1 Regional association plots of the identified loci. On the y-axis – $\log_{10}$ of *p*-values are shown, with the genomics position on the x-axis. *Colors* represent the linkage disequilibrium (LD) with the most significantly associated SNP in the locus and *shape* represents the SNP class. The analysed trait with the corresponding genetic model is reported in the title of each plot

cream. Thus, *IRX4* could have a similar role linking this gene to the reward aspects of food as well.

For vegetables, an interesting association was detected between the *CSMD1* gene and chicory liking. This gene has been previously associated in large GWAS studies to schizophrenia [43] and to metabolic syndrome [44]. Rose et al. have also shown that variants from this gene lead to differential activation in the cuneus [45], which in turn, has been described as being differentially activated in women suffering from *bulimia nervosa* in response to food stimuli, suggesting a role of this area in the food related reward system [46]. These data indirectly suggest that *CSMD1* might also regulate the

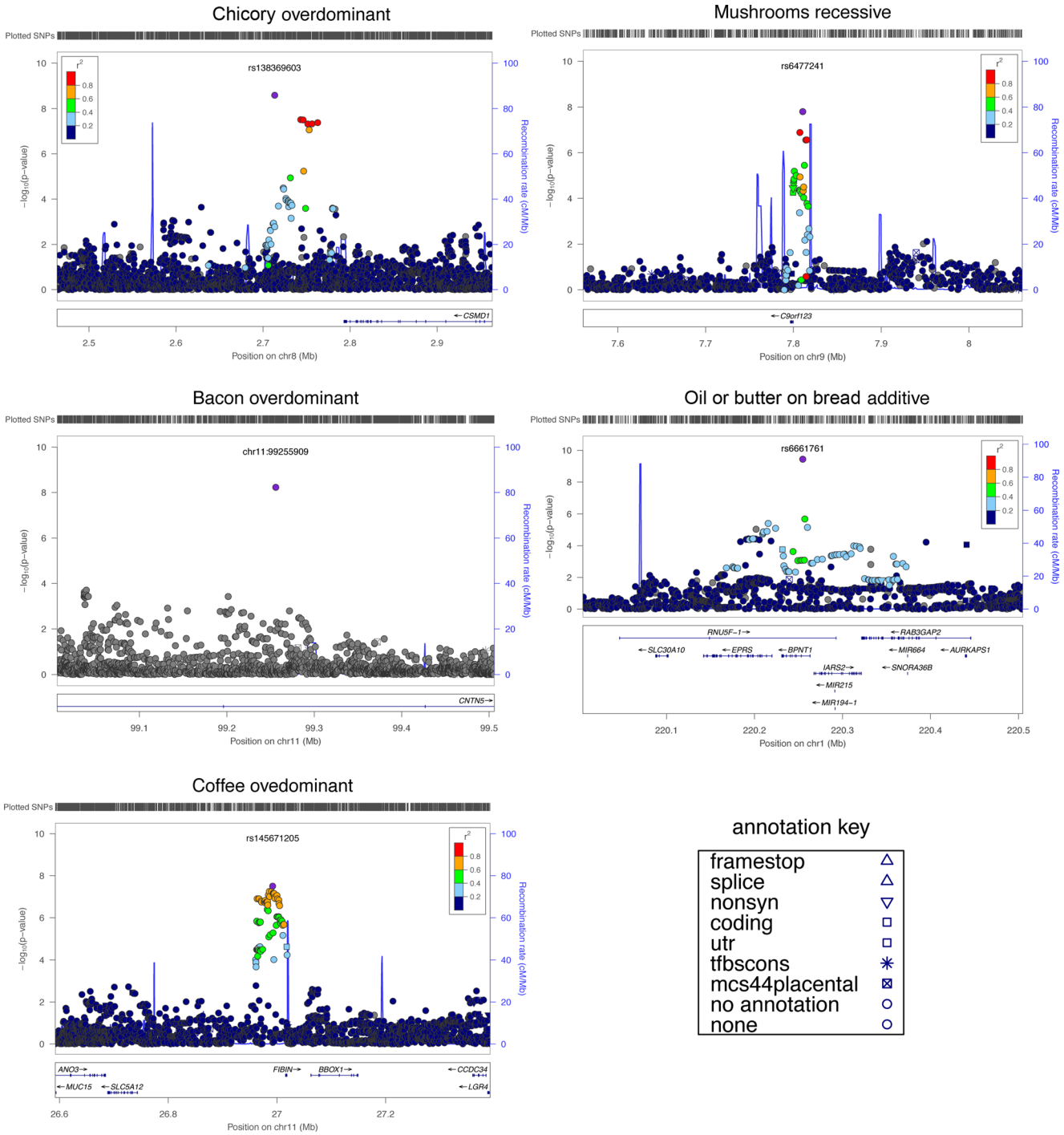

Fig. 2 Regional association plots of the identified loci. On the y-axis – $\log_{10}$ of *p*-values are shown, with the genomics position on the x-axis. *Colors* represent the linkage disequilibrium (LD) with the most significantly associated SNP in the locus and *shape* represents the SNP class. The analyzed trait with the corresponding genetic model is reported in the title of each plot

reward response to food by regulating the activation of the cuneus.

For all the other genes identified we were not able to find enough data to explain a link with food liking. This was mainly due to the current poor knowledge on their roles and functions. However, this being the first GWAS on common food likings such lack of data on candidate genes/loci was to be expected. Finally, none of the identified genes code for either taste or olfactory receptors. This observation is consistent with the fact that although some polymorphisms of taste and olfactory receptors genes have been shown to explain differences in specific compound perception which in some cases translate

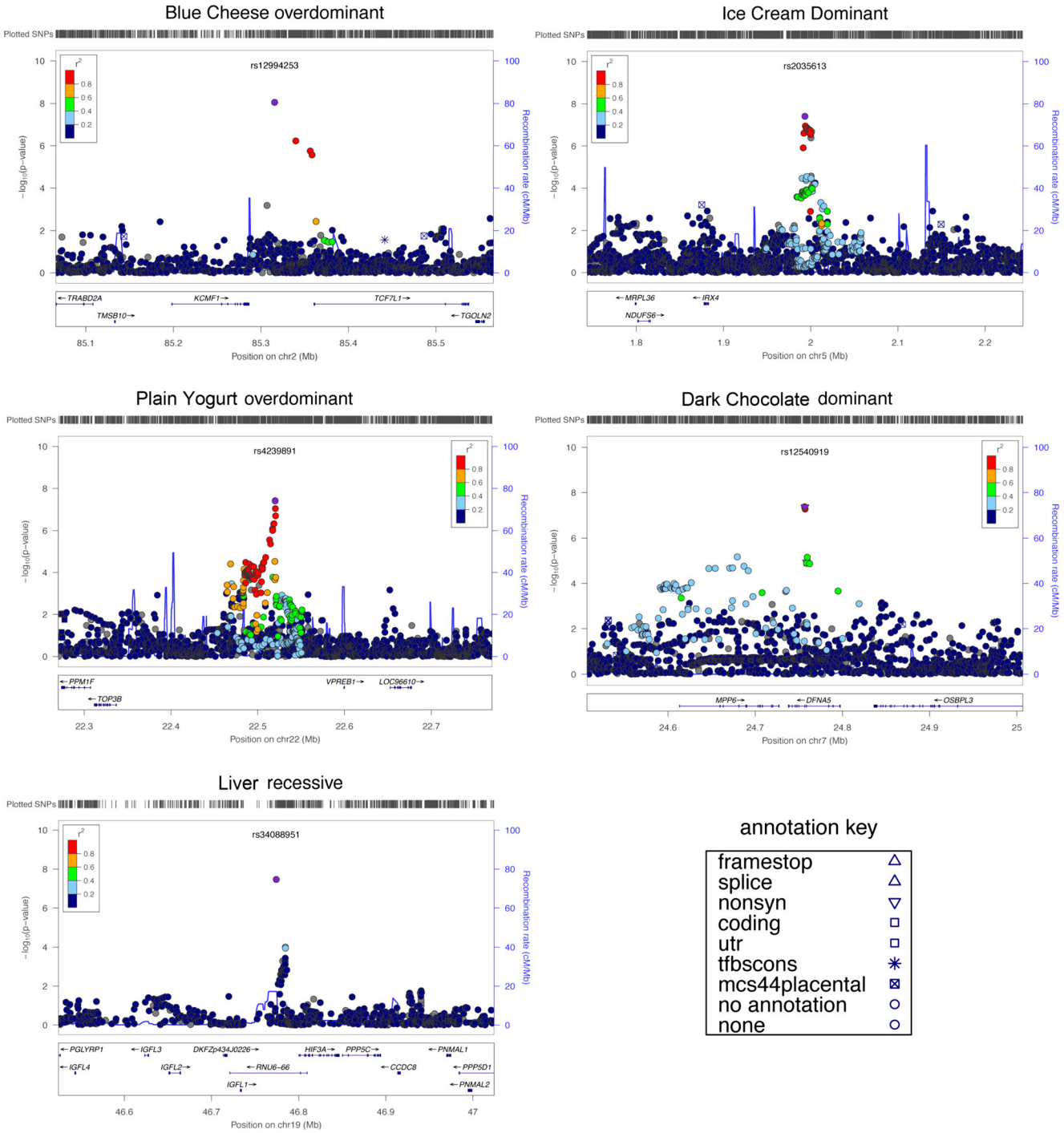

Fig. 3 Regional association plots of the identified loci. On the y-axis – $\log_{10}$ of *p*-values are shown, with the genomic position on the x-axis. *Colors* represent the linkage disequilibrium (LD) with the most significantly associated SNP in the locus and *shape* represents the SNP class. The analyzed trait with the corresponding genetic model is reported in the title of each plot

in differences in real food liking [12, 15, 17] or consumption [10, 13, 14, 16], in most cases the proportion of explained variance is relatively small and thus undetectable on a genome-wide study of this size.

Many of the identified loci do not follow the usually used additive genetic inheritance model. Although it is very well known that miss-specifying the correct genetic model in association studies leads to loss of power [31] other models are rarely tested. In our case we would have been able to identify only 4 out of the 15 loci, and none of the non-additive model ones would have given significant associations under an additive model. This

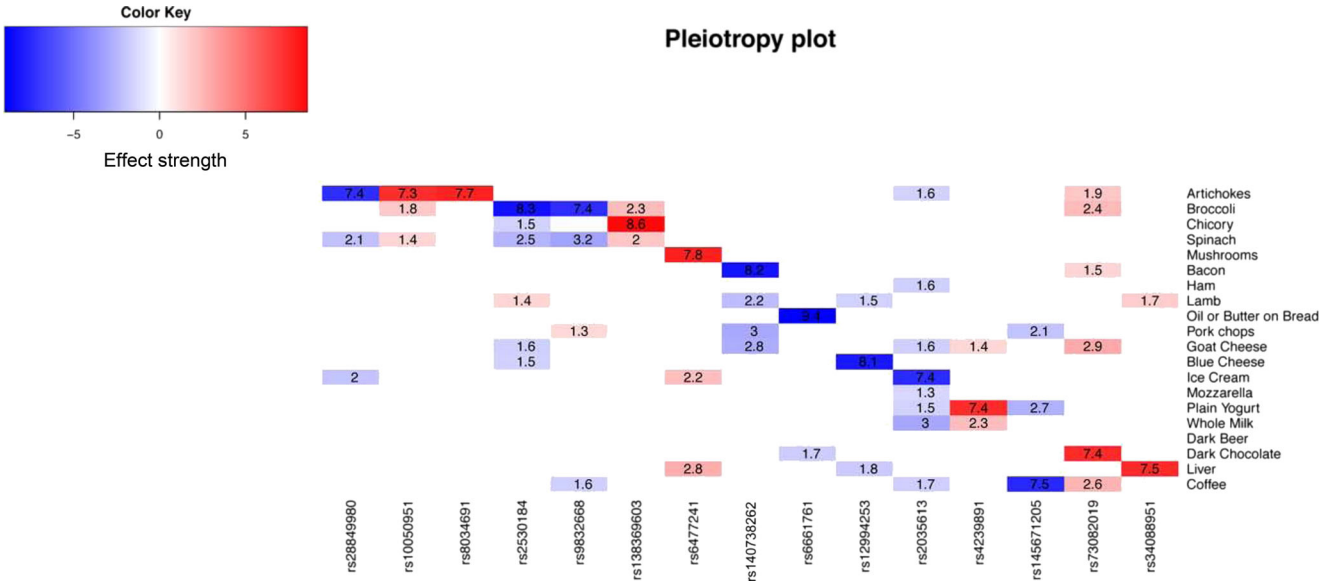

Fig. 4 The heat map reports the *p*-values and direction of effect for the association between each identified SNP and all of the studied traits. The *number* represents the $-\log_{10}$ of the *p*-value while the *color* indicate the direction and strength of the effect. Results with *p*-value larger than 0.05 have been omitted

observation suggests that the non-additive models should be tested in all GWAS studies.

Many loci (6 out of 15) were identified as following an "over-dominant" model, in which a heterozygote expresses a different phenotype from the two homozygotes. This particular genetic model is rarely tested but it has been described in human genetics for schizophrenia [47] and cervical cancer [48]. Moreover, one should consider that the power to detect an association using an additive model is very low, when the true model is over-dominant. Consequently, it is unclear if the heterozygous effect is limited to food liking or if it extends also to other quantitative traits. Reanalysis of existing data is needed to clarify this issue.

The present study clearly shows some limitations. First of all the sample size, despite being the largest one ever used for such a study, is relatively small when considering genome-wide association studies and independent replication will be needed to confirm the present results. Another limitation is that we have considered a relatively small range of foods due to the difference in the items used to assess likings in very different populations. Future studies will need to use more standardized questionnaires and look also at food groups instead of specific items.

Despite its limitations the present study represent a very important step in understanding the biology determining food liking beyond mere taste. These knowledge will help in creating new products and drugs to help people comply to healthier diet. Moreover it could lead to the design of personalized dietary interventions for people suffering from food related diseases, helping increase the efficacy of the treatments.

In conclusion, our results represent one of the first successful attempts at uncovering the genetic factors underlying food liking. Several of the identified genes are good candidates for linking food hedonics to actual consumption. This study opens new perspectives in understanding the relationship between genes and food hedonics, and ultimately their impact on health.

Acknowledgments We would like to thank all the participants in the various studies for their contribution and support.

The SR study has been founded by the Region Friuli Venezia Giulia grant number 35\09 Linea 2 "Sulle tracce di Marco Polo: geni, gusto e loro implicazioni sulla salute lungo la Via della Seta".

The INGI-FVG study was founded through the Italian Ministry of health.

We thank the inhabitants and the administrators of the Val Borbera for their participation in the study. A special thanks to Professor Clara Camaschella, Dr Silvia Bione, Dr Laura Crocco, Ms Maria Rosa Biglieri, Dr Diego Sabbi for help with the data collection. We thank Fondazione Compagnia di San Paolo, Torino, Fondazione Cariplo, Milano and Health Ministry (Progetto Finalizzato and Italian Centre for Disease Prevention and Control) for financial support.

The ERF study as a part of EUROSPAN (European Special Populations Research Network) was supported by European Commission FP6 STRP grant number 018947 (LSHG-CT-2006-01947) and also received funding from the European Community's Seventh Framework Programme (FP7/2007–2013)/grant agreement HEALTH-F4-2007-201413 by the European Commission under the programme "Quality of Life and Management of the Living Resources" of 5th Framework Programme (no. QLG2-CT-2002-01254). High-throughput analysis of the ERF data was supported by joint grant from Netherlands Organisation for Scientific Research and the Russian Foundation for Basic Research (NWO-RFBR 047.017.043). We are grateful to all study participants and their relatives, general practitioners and neurologists for their contributions and to P. Veraart for her help in genealogy, J. Vergeer for the supervision of the laboratory work and P. Snijders for his help in data collection. Statistical analyses were partly carried out on the Genetic Cluster Computer (http://www.geneticcluster.org) which is financially supported by the Netherlands Scientific Organization (NWO 480-05-003 PI: Posthuma) along with a supplement from the Dutch Brain Foundation and the VU University Amsterdam. This research was

financially supported by BBMRI-NL, a Research Infrastructure financed by the Dutch Government (NWO 184.021.007). The work of LCK was partially funded by the FP7 projects MIMOmics (grant no. 305280) and PainOmics (grant no. 602736).

Author contribution Conceived and designed the experiments: NP LN PG. Performed the experiments: NP SMW MT GP LCK AR PDA AD. Analyzed the data: NP MK CS. Contributed reagents/materials/analysis tools: CVD DT PG NA. Wrote the paper: NP MK LCK PG.

Compliance with ethical standards

Ethics statement Consent forms for clinical and genetic studies were signed by each participant and all research was conducted according to the ethical standards defined by the Helsinki declaration. The INGI-CARL, INGI-FVG and SR studies have been approved by the Institutional Review Board of IRCCS Burlo Garofolo PROT CE/v-78 in Trieste Italy. The INGI-VB study was approved by San Raffaele Hospital and Regione Piemonte ethical committees. The ERF study was approved by the Erasmus institutional medical-ethics committee in Rotterdam, The Netherlands

Conflict of interest The authors declare no conflict of interest.

Competing financial interests The author(s) declare no competing financial interests.

References

1. Diet, nutrition and the prevention of chronic diseases. World Health Organ. Tech. Rep. Ser. [Internet]. 2003 [cited 2014 Jun 2];916:i—viii, 1–149, backcover. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12768890.
2. Story M, Neumark-Sztainer D, French S. Individual and environmental influences on adolescent eating behaviors. J Am Diet Assoc. 2002;102:S40–51. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11902388.
3. Tanaka T, Ngwa JS, van Rooij FJA, Zillikens MC, Wojczynski MK, Frazier-Wood AC. Genome-wide meta-analysis of observational studies shows common genetic variants associated with macronutrient intake. Am J Clin Nutr. 2013;97:1395–402. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3652928&tool=pmcentrez&rendertype=abstract.
4. Schatzkin A, Kipnis V, Carroll RJ, Midthune D, Subar AF, Bingham S. A comparison of a food frequency questionnaire with a 24-hour recall for use in an epidemiological cohort study: results from the biomarker-based Observing Protein and Energy Nutrition (OPEN) study. Int J Epidemiol. 2003;32:1054–62. Available from: http://www.ncbi.nlm.nih.gov/pubmed/14681273.
5. International Food Information Council Foundation. The 2015 food & health survey: consumer attitudes toward food safety, nutrition & health. Washington, D.C.; 2015.
6. Duffy VB, Hayes JE, Sullivan BS, Faghri P. Surveying food and beverage liking: a tool for epidemiological studies to connect chemosensation with health outcomes. Ann N Y Acad Sci. 2009;1170:558–68. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19686193.
7. Pallister T, Sharafi M, Lachance G, Pirastu N, Mohney RP, MacGregor A, et al. Food preference patterns in a UK Twin Cohort. Twin Res Hum Genet. 2015.
8. Sharafi M, Peracchio H, Scarmo S, Huedo-Medina TB, Mayne ST, Cartmel B. Preschool-Adapted Liking Survey (PALS): a brief and valid method to assess dietary quality of preschoolers. Child Obes. 2015;11:530–40. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26322657.
9. Keskitalo K, Tuorila H, Spector TD, Cherkas LF, Knaapila A, Silventoinen K. Same genetic components underlie different measures of sweet taste preference. Am J Clin Nutr. 2007;86:1663–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18065584.
10. Duffy VB, Hayes JE, Davidson AC, Kidd JR, Kidd KK, Bartoshuk LM. Vegetable intake in college-aged adults is explained by oral sensory phenotypes and TAS2R38 genotype. Chemosens Percept. 2010;3:137–48. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3000691&tool=pmcentrez&rendertype=abstract.
11. Hayes JE, Wallace MR, Knopik VS, Herbstman DM, Bartoshuk LM, Duffy VB. Allelic variation in TAS2R bitter receptor genes associates with variation in sensations from and ingestive behaviors toward common bitter beverages in adults. Chem Senses. 2011;36:311–9. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3038275&tool=pmcentrez&rendertype=abstract.
12. Pirastu N, Robino A, Lanzara C, Athanasakis E, Esposito L, Tepper BJ, et al. Genetics of food preferences: a first view from silk road populations. J Food Sci. 2012. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22888812.
13. Colares-Bento FCJ, Souza VC, Toledo JO, Moraes CF, Alho CS, Lima RM. Implication of the G145C polymorphism (rs713598) of the TAS2r38 gene on food consumption by Brazilian older women. Arch Gerontol Geriatr. 2016;54:e13–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21763010.
14. Hayes JE, Feeney EL, Allen AL. Do polymorphisms in chemosensory genes matter for human ingestive behavior? Food Qual Prefer. 2013;30:202–16. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23878414.
15. Pirastu N, Kooyman M, Traglia M, Robino A, Willems SM, Pistis G. Association analysis of bitter receptor genes in five isolated populations identifies a significant correlation between TAS2R43 variants and coffee liking. PLoS One Publ Lib Sci. 2014;9:e92065.
16. Sandell M, Hoppu U, Mikkilä V, Mononen N, Kähönen M, Männistö S. Genetic variation in the hTAS2R38 taste receptor and food consumption among Finnish adults. Genes Nutr. 2014;9:433. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4235829&tool=pmcentrez&rendertype=abstract.
17. Hayes JE, Feeney EL, Nolden AA, McGeary JE. Quinine bitterness and grapefruit liking associate with allelic variants in TAS2R31. Chem Senses. 2015;40:437–43. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26024668.
18. Jones LV, Peryam DR, Thurstone LL. development of a scale for measuring soldiers'food preferences b. J Food Sci. 1955;20:512–20. Available from: http://doi.wiley.com/10.1111/j.1365-2621.1955.tb16862.x.
19. Coetzee H, Taylor JRN. The use and adaptation of the paired-comparison method in the sensory evaluation of hamburger-type patties by illiterate/semi-literate consumers. Food Qual Prefer. 1996;7:81–5. Available from: http://www.sciencedirect.com/science/article/pii/0950329395000399.
20. Colman AM, Norris CE. PCC. comparing rating scales of different lengths: equivalence of scores from 5-point and 7-point scales. Psychol Rep. 1997;80:355–62.
21. Preston CC, Colman AM. Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. Acta Psychol (Amst). 2000;104:1–15. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10769936.
22. J.G. D. Five point vs. eleven point scales: does it make a difference to data characteristics? Australas. J Mark Res. 2002;10:39–47.

23. Aulchenko YS, Heutink P, Mackay I, Bertoli-Avella AM, Pullen J, Vaessen N. Linkage disequilibrium in young genetically isolated Dutch population. Eur J Hum Genet. 2004;12:527–34. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15054401.
24. Pardo LM, MacKay I, Oostra B, van Duijn CM, Aulchenko YS. The effect of genetic drift in a young genetically isolated population. Ann Hum Genet. 2005;69:288–95. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15845033.
25. Delaneau O, Zagury J-F, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. Nat Methods. 2013;10:5–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23269371.
26. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012;491:56–65. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3498066&tool=pmcentrez&rendertype=abstract.
27. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet. 2012;44:955–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22820512.
28. Belonogova NM, Svishcheva GR, van Duijn CM, Aulchenko YS, Axenovich TI. Region-based association analysis of human quantitative traits in related individuals. PLoS One. 2013;8:e65395. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3684601&tool=pmcentrez&rendertype=abstract.
29. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide association analysis. Bioinformatics. 2007;23:1294–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17384015.
30. González JR, Carrasco JL, Dudbridge F, Armengol L, Estivill X, Moreno V. Maximizing association statistics over genetic models. Genet Epidemiol. 2008;32:246–54. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18228557.
31. Lettre G, Lange C, Hirschhorn JN. Genetic model testing and statistical power in population-based association studies of quantitative traits. Genet Epidemiol. 2007;31:358–62. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17352422.
32. Spiegelberg BD, Xiong JP, Smith JJ, Gu RF, York JD. Cloning and characterization of a mammalian lithium-sensitive bisphosphate 3′-nucleotidase inhibited by inositol 1,4-bisphosphate. J Biol Chem. 1999;274:13619–28. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10224133.
33. Marchese G, Scheggi S, Secci ME, De Montis MG, Gambarana C. Anti-anhedonic activity of long-term lithium treatment in rats exposed to repeated unavoidable stress. Int J Neuropsychopharmacol. 2013;16:1611–21. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23363811.
34. Green E, Jacobson A, Haase L, Murphy C. Reduced nucleus accumbens and caudate nucleus activation to a pleasant taste is associated with obesity in older adults. Brain Res. 2011;1386:109–17. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3086067&tool=pmcentrez&rendertype=abstract.
35. Yang Q, Kathiresan S, Lin J-P, Tofler GH, O'Donnell CJ. Genome-wide association and linkage analyses of hemostatic factors and hematological phenotypes in the Framingham Heart Study. BMC Med Genet. 2007;8 Suppl 1:S12. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1995619&tool=pmcentrez&rendertype=abstract.
36. Kathiresan S, Manning AK, Demissie S, D'Agostino RB, Surti A, Guiducci C. A genome-wide association study for blood lipid phenotypes in the Framingham Heart Study. BMC Med Genet. 2007;8 Suppl 1:S17. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1995614&tool=pmcentrez&rendertype=abstract.
37. Nakabayashi K, Komaki G, Tajima A, Ando T, Ishikawa M, Nomoto J. Identification of novel candidate loci for anorexia nervosa at 1q41 and 11q22 in Japanese by a genome-wide association analysis with microsatellite markers. J Hum Genet. 2009;54:531–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19680270.
38. Bayrak F, Kömürcü-Bayrak E, Mutlu B, Kahveci G, Erginel-Unaltuna N. Genetic analysis of the Irx4 gene in hypertrophic cardiomyopathy. Türk Kardiyol Derneği arşivi Türk Kardiyol Derneğinin yayın organıdır. 2008;36:90–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18497553.
39. Cheng Z, Wang J, Su D, Pan H, Huang G, Li X. Two novel mutations of the IRX4 gene in patients with congenital heart disease. Hum Genet. 2011;130:657–62. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21544582.
40. Smemo S, Tena JJ, Kim K-H, Gamazon ER, Sakabe NJ, Gómez-Marín C. Obesity-associated variants within FTO form long-range functional connections with IRX3. Nature. 2014;507:371–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24646999.
41. Scheid JL, Carr KA, Lin H, Fletcher KD, Sucheston L, Singh PK. FTO polymorphisms moderate the association of food reinforcement with energy intake. Physiol Behav. 2014;132:51–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24768648.
42. Stice E, Burger KS, Yokum S. Relative ability of fat and sugar tastes to activate reward, gustatory, and somatosensory regions. Am J Clin Nutr. 2013;98:1377–84. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24132980.
43. Ripke S, Sanders AR, Kendler KS, Levinson DF, Sklar P, Holmans PA, et al. R GP. Genome-wide association study identifies five new schizophrenia loci. Nat Genet. 2011;43:969–76. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3303194&tool=pmcentrez&rendertype=abstract.
44. Nock NL, Wang X, Thompson CL, Song Y, Baechle D, Raska P. Defining genetic determinants of the Metabolic Syndrome in the Framingham Heart Study using association and structural equation modeling methods. BMC Proc. 2009;3 Suppl 7:S50. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2795950&tool=pmcentrez&rendertype=abstract.
45. Rose EJ, Morris DW, Hargreaves A, Fahey C, Greene C, Garavan H. Neural effects of the CSMD1 genome-wide associated schizophrenia risk variant rs10503253. Am J Med Genet B Neuropsychiatr Genet. 2013;162B:530–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23839771.
46. Van den Eynde F, Giampietro V, Simmons A, Uher R, Andrew CM, Harvey P-O. Brain responses to body image stimuli but not food are altered in women with bulimia nervosa. BMC Psychiatr. 2013;13:302. Available from: http://www.biomedcentral.com/1471-244X/13/302.
47. Zhou DH, Yan QZ, Yan XM, Li CB, Fang H, Zheng YL. The study of BDNF Val66Met polymorphism in Chinese schizophrenic patients. Prog Neuropsychopharmacol Biol Psychiatry. 2010;34:930–3. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20420877.
48. Chen H, Zhu J. C677T polymorphism of methylenetetrahydrofolate reductase may contribute to cervical cancer risk in complete overdominant model. Med Hypotheses. 2013;80:679–83. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23490201.