Accepted manuscript

**A Genome-wide Association Study on Confection Consumption in a Japanese Population- The Japan Multi-Institutional Collaborative Cohort study**

Taro Suzuki[1], Yasuyuki Nakamura[2, 3], Yukio Doi[1], Akira Narita[4], Atsushi Shimizu[5], Nahomi Imaeda[6,7], Chiho Goto[7,8], Kenji Matsui[9], Aya Kadota[3, 10], Katsuyuki Miura[3, 10], Masahiro Nakatochi[11], Keitaro Tanaka[12], Megumi Hara[12], Hiroaki Ikezaki[13, 14], Masayuki Murata[14], Toshiro Takezaki[15], Daisaku Nishimoto[16], Keitaro Matsuo[17], Isao Oze[17], Nagato Kuriyama[18], Etsuko Ozaki[18], Haruo Mikami[19], Yohko Nakamura[19], Miki Watanabe[8], Sadao Suzuki[8], Sakurako Katsuura-Kamano[20], Kokichi Arisawa[20], Kiyonori Kuriki[21], Yukihide Momozawa[22], Michiaki Kubo[22], Kenji Takeuchi[23], Yoshikuni Kita[3, 24], Kenji Wakai[23], for the J-MICC Research Group

[1]Department of Food Science and Human Nutrition, Ryukoku University, Otsu, Japan;

[2]Yamashina Racto Clinic and Medical Examination Center, Kyoto, Japan;

[3]Department of Public Health, Shiga University of Medical Science, Otsu, Japan;

Accepted manuscript

[4]Department of Integrative Genomics, Tohoku Medical Megabank Organization, Tohoku University, Sendai, Japan;

[5]Division of Biomedical Information Analysis, Institute for Biomedical Sciences, Iwate Medical University, Shiwa-gun, Iwate, Japan;

[6]Department of Nutrition, Faculty of Wellness, Shigakkan University, Obu, Japan;

[7]Department of Public Health, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan；

[8]Department of Health and Nutrition, School of Health and Human Life, Nagoya Bunri University, Inazawa, Japan;

[9]Division of Bioethics and Health Law, Center for Public Health Sciences, the National Cancer Center, Tokyo, Japan;

[10]Center for Epidemiologic Research in Asia, Shiga University of Medical Science, Otsu, Japan;

[11]Public Health Informatics Unit, Department of Integrated Health Sciences, Nagoya University Graduate School of Medicine, Nagoya, Japan；

[12]Department of Preventive Medicine, Faculty of Medicine, Saga University, Saga, Japan；

[13]Department of Comprehensive General Internal Medicine, Faculty of Medical Sciences, Kyushu University Graduate School , Fukuoka, Japan；

[14]Department of General Internal Medicine, Kyushu University Hospital, Fukuoka, Japan；

Accepted manuscript

[15]Department of International Island and Community Medicine, Kagoshima University Graduate School of Medical and Dental Sciences, Kagoshima, Japan；

[16]School of Health Sciences, Faculty of Medicine, Kagoshima University, Kagoshima, Japan；

[17]Division of Cancer Epidemiology and Prevention, Aichi Cancer Center, Nagoya, Japan；

[18]Department of Epidemiology for Community Health and Medicine, Kyoto Prefectural University of Medicine、Kyoto, Japan;

[19]Cancer Prevention Center, Chiba Cancer Center Research Institute, Chiba, Japan；

[20]Department of Preventive Medicine, Institute of Biomedical Sciences, Tokushima University Graduate School, Tokushima, Japan；

[21]Laboratory of Public Health, Division of Nutritional Sciences, School of Food and Nutritional Sciences, University of Shizuoka, Shizuoka, Japan；

[22]Laboratory for Genotyping Development, RIKEN Center for Integrative Medical Sciences, Kanagawa, Japan;

[23]Department of Preventive Medicine, Nagoya University Graduate School of Medicine, Nagoya, Japan;

[24]Faculty of Nursing Science, Tsuruga Nursing University, Tsuruga, Japan.

Accepted manuscript

Address correspondence and reprint requests to:

Yasuyuki Nakamura, MD,

Department of Public Health, Shiga University of Medical Science, Otsu, Japan

Phone: +81-77-548-2191, Fax: +81-77-543-9732

E-mail: nakamura@belle.shiga-med.ac.jp

Accepted manuscript

## Abstract

Differences in individual eating habits may be influenced by genetic factors, in addition to cultural, social, or environmental factors. Previous studies suggested that genetic variants within sweet taste receptor genes family were associated with sweet taste perception and the intake of sweet foods. The aim of this study was to conduct a genome-wide association study (GWAS) to find genetic variations that affect confection consumption in a Japanese population. We analyzed GWAS data on sweets consumption using 14,073 participants from the Japan Multi-Institutional Collaborative Cohort study. We used a semi-quantitative food frequency questionnaire to estimate food intake that was validated previously. Association of the imputed variants with sweets consumption was performed by linear regression analysis with adjustments for age, sex, total energy intake and principal component analysis components 1 to 3. Furthermore, the analysis was repeated adjusting for alcohol intake (g/day) in addition to the above-described variables. We found 418 single nucleotide polymorphisms (SNPs) located in 12q24 that were associated with sweets consumption. SNPs with the 10 lowest P-values were located on nine genes including at the *BRAP*, *ACAD10*, and *ALDH2* regions on 12q24.12-13. After adjustment for alcohol intake, no variant was associated with sweets intake with genome-wide significance. In conclusion, we found a significant number of SNPs located on 12q24 genes that were associated with sweets intake before adjustment for alcohol intake. However, all of them lost statistical significance after adjustment for alcohol intake.

Accepted manuscript

**Introduction**

Differences in individual eating habits may be influenced by genetic factors, in addition to cultural, social, or environmental factors. The sweet taste receptor gene family, including *TAS1R2* and *TAS1R3*, is located on chromosome 1 in humans [1-3]. Previous candidate gene studies suggested that genetic variants of this receptor gene family were associated with sweet taste perception and the intake of sweet foods [4-6]. Some of these studies showed the differential effects of SNPs on sucrose sensitivity and sugar intake depending on the presence or absence of being overweight [5, 6]. Other genetic variants that candidate gene studies reported to be associated with sweet consumption were *GNAT3,* (which encodes the G protein alpha subunit that plays a prominent role in bitter and sweet taste transduction as well as in umami taste transduction) [7], *GLUT2* (a glucose transporter gene) [8], and *FGF21* (a fibroblast growth factor gene) [9, 10]. All these variants suggested by candidate gene studies could not be replicated by subsequent GWAS in both samples of European ancestry and Japanese, probably because the previous candidate gene studies did not consider population stratification [11, 12]. Hwang et al., in a GWAS of the perception and intake of sweet substances, found rs11642841 within the *FTO* gene on chromosome 16 was strongly associated with the intake of total sugar in a UK Biobank sample [11].

Recently, the results of several GWAS on the dietary habits of the Japanese population have shown interesting pleiotropic effects of rs671 across dietary habits on foods and beverages and we and others found that an effective rs671 allele was inversely associated with fish consumption, and was directly associated with coffee consumption [12-15]. Kawafune, et al. found a strong direct association

between an effective rs671 allele and a five-point scale of sweet taste preference [12]. All these

associations were profoundly attenuated by adjustment for alcohol drinking.

For more than a century, people in Japan have eaten two main types of confectionaries:

traditional Japanese style confectionaries (*wagashi*), and European style confectionaries, including

patisseries and *confiseries* (confectionaries), which arrived in Japan in the Meiji era. *Wagashi* are

traditional Japanese confectioneries often served with green tea, and made of *mochi* (sticky rice cake),

*anko* (*azuki* bean paste), and fruit. In the present GWAS, we took a semi-quantitative amount of *wagashi*

and European style confectionaries as a more objective measure of sweets liking, rather than a

self-assessed scale of sweet taste preference. The purposes of this study were (1) to replicate GWAS on

sweets consumption performed in samples of European ancestry in a Japanese population, and (2) to use

a more objective measure of sweets liking, rather than a self-assessed scale of sweet taste preference.

**Methods**

*Study population*

The study was conducted with participants aged from 35 to 69 years as a cross-sectional study within

the Japan Multi-Institutional Collaborative Cohort (J-MICC) study, a large cohort study that was

started in 2005 to investigate gene environment interactions in lifestyle-related diseases. The age range

was set for the following reasons: The youngest age for inclusion was set at 35 years old, anticipating

easier recruitment from health check examinees and others. Also, incidences of non-communicable

Accepted manuscript

diseases, including cancer, are thought to increase after age 35. The oldest age for inclusion was set at

70 years old, anticipating an easier understanding of participants for genomic study. For this study,

about 100,000 volunteers provided their blood and lifestyle data based on a questionnaire. The

participants in the J-MICC study were recruited from 14 different areas throughout Japan between

2004 and 2014. Of 47,163 respondents enrolled between 2004 and 2013, 14,539 participants were

randomly selected considering the distribution of study areas for the present GWAS. We used data ver.

20180112. Details of the J-MICC study were reported elsewhere [13, 14, 16]. Briefly, participants

completed a questionnaire about lifestyle and medical information, and had blood sampled at the time

of the baseline survey. The study participants lived in communities in four regions (Chiba, Daiko,

Fukuoka, and Saga), were health check examinees in seven regions (Shizuoka-Sakuragaoka, Shizuoka,

Okazaki, Iga, Takashima, Kagoshima, and KOPS), and first-visit patients at a cancer hospital in the

Aichi Cancer Center region. In Kyoto and Tokushima regions, the study participants were inhabitants

in communities and health check examinees as well as employees of companies or local governments.

All participants in this study gave written informed consent prior to enrollment. The study protocol

was approved by the ethics committees of Nagoya University Graduate School of Medicine, and other

institutions participating in the J-MICC study. The present study was conducted according to the

principles expressed in the World Medical Association Declaration of Helsinki.

Of 14,539 participants, 448 were excluded by the GWAS screening described below in the

section on *Genotyping and quality control filtering*. Of the remaining 14,091 participants, 12 with daily

energy intake less than 500 kcal or greater than 5,000 kcal were excluded. As a result, we analyzed the

data of 14,073 participants in the present study.

*Questionnaire and measurements*

The questionnaire for the J-MICC studies included questions on medical history, height, weight,

smoking and drinking habits, and dietary habits. The questionnaire was checked by experienced staff to

confirm completeness and consistency. Height and weight measurements and blood sampling were

conducted as part of a health checkup or for research purposes at the institutions participating in the

J-MICC study [16]. Question items were collected using a scientifically validated self-administered

questionnaire [17-21]. Body mass index (BMI) was calculated by dividing body weight in kilograms by

the square of height in meters.

*Dietary Assessment*

We used a semi-quantitative food frequency questionnaire (FFQ) to estimate food intake that was

validated previously [17-21]. We chose 20 foods/food groups and beverages, the details of which are

given in **Supplementary Information on FFQ.** Food intake frequencies were classified into eight

categories (never or seldom, 1-3 times per month, 1-2 times per week, 3-4 times per week, 5-6 times per

week, once a day, twice a day, and three or more times a day, which were converted to 0, 0.1, 0.2, 0.5, 0.8,

1, 2, and 3, respectively, before analysis). Confections in the questionnaire were grouped in two

categories as follows: 1. Western confections (cakes or cream puffs, etc.), 2. Japanese confections (manju [steamed sweet bean buns], etc.). For each category of confections, the frequency was multiplied by the portion size, and the total confections intake amount calculated. Unfortunately, we did not have any data on intakes of added sugar or sugar sweetened beverages. Total alcohol intake was estimated as the sum of pure alcohol intake. The frequency of alcohol intake was obtained using six categories (never or seldom, 1-3 times per month, 1-2 times per week, 3-4 times per week, 5-6 times per week, and every day). Total alcohol consumption (g/day) was estimated as the summed amount of pure alcohol consumption.

*Genotyping and quality control filtering*

Buffy coat fractions and DNA were prepared from blood samples and stored at −80°C at the central J-MICC Study office. DNA was extracted from all buffy coat fractions using a BioRobot M48 Workstation (Qiagen Group, Tokyo, Japan) at the central study office. For the samples from two areas (Fukuoka and Kyushu and the Okinawa Population Study [KOPS]), DNA was extracted locally from samples of whole blood using an automatic nucleic acid isolation system (NA-3000, Kurabo, Co., Ltd, Osaka, Japan). The 14,539 study participants from the 13 areas of the J-MICC study were genotyped at the RIKEN Center for Integrative Medicine Sciences using a HumanOmniExpressExome-8v1.2 BeadChip array (Illumina Inc., San Diego, CA, USA). Twenty-six participants with inconsistent sex information between the questionnaire and the estimate from genotyping were excluded. The

identity-by-descent method in the PLINK 1.9 software [22, 23] identified 388 close relationship pairs (pi-hat > 0.1875) and one sample from each pair of the 388 was excluded. Principal component analysis (PCA) [24] with a 1000 Genomes reference panel (phase 3) [25] detected 34 participants whose estimated ancestries were outside the Japanese population [26]. These 34 participants were excluded. In the remaining 14,091 participants, SNPs with a genotype call rate <0.98 and/or a Hardy-Weinberg equilibrium exact test P-value $< 1 \times 10^{-6}$, a low minor allele frequency (MAF) < 0.01, or a departure from the allele frequency computed from the 1000 Genomes Phase 3 EAS samples were excluded. The quality control filtering resulted in 14,091 individuals and 570,162 SNPs.

*Genotype imputation*

Genotype imputation was performed using SHAPEIT [27] and Minimac3 software [28] based on the 1000 Genomes reference panel (phase 3) [25]. After genotype imputation, strict quality control filters were applied; namely, variants with an $R^2 < 0.3$ were excluded, resulting in 12,617,547 variants. Finally, 4,112,564 variants with MAF <0.01 in patients were removed, resulting in 8,504,983 variants for the analysis. We used DosageConvertor software [29] to convert dosage files in VCF format from Minimac3 to PLINK formats.

*Power calculations to test of an association between confection intake and SNPs*

Accepted manuscript

Statistical power to detect a true association was calculated by the method of Delongchamp et al. [30], based on the number of participants and genetic data of the discovery

phase J-MICC study. The required non-centrality parameter was obtained by the equation A4 listed in

the Appendix A by Visscher et al. [31]. When then the number of participants is 14,073, with 14

covariates for adjustment, 0.8 for linkage disequilibrium (LD) $R^2$, 0.2 for MAF, 0.02 for squared

standardized beta estimate, and 8,500,000 for variants analyzed, the statistical power is calculated as

0.991 according to the method proposed by Deongchamp et al. [30].

*Association analyses between genetic variants and confections intake*

Associations between all imputed variants and confections intake were analyzed by linear regression

assuming the additive effects of the allele dosage on confections intake (g/day) adjusted for age, sex,

total energy intake, and principal component analysis (PCA) components 1 to 3 using PLINK 1.9

software [22, 23]. Furthermore, the analysis was repeated adjusting for alcohol intake (ethanol g/day) in

addition to the above-described variables. Variants achieving genome-wide significance ($P<5\times10^{-8}$)

were considered to be confections intake-associated variants. An R package for creating a Q-Q plot,

GWAS Tools was used [32]. For scatter plots of P-values derived from genome-wide scan results for

confections intake, qqman software was used [33]. To visualize regions of interest, we used the

LocusZoom program [34]. For sensitivity analyses, we performed GWAs on Western and Japanese

confections intakes separately.

Accepted manuscript

In addition, replication analysis adjusted for age, sex, total energy intake, and PCA

components 1 to 3 using the J-MICC samples for 20 SNPs that were previously reported to be associated

with sweet taste perception or confections intake [1-11] was performed.

Student's t-tests were used to compare means between men and women.

**Results**

*Baseline characteristics*

Baseline characteristics of the all, male and female participants are shown in Table 1. The mean age of

the participants was $54.8\pm 9.4$ y and the percentage of women was 55.0%. The mean confections intake

was $198\pm142$ KJ/day. Total energy intake was $7,135\pm1,490$ KJ/day. The means for protein, fat,

carbohydrate (% of total energy), and alcohol intake (g/day) were $12.7\pm 2.0$%, $23.8\pm 6.4$%, $55.8\pm 6.2$%;

$9.4\pm 16.3$ g/day, respectively. The mean BMI level was $23.1\pm 3.3$ kg/m$^2$. The mean age, total energy

intake, %carbohydrate intake, alcohol intake, and BMI were significantly larger in men than in women.

The mean confections intake, %protein and %fat intake were significantly smaller in men than in

women.

*Association analyses between confections intake and genetic variants*

Genome-wide analyses among the 8,504,983 variants adjusted for age, sex, total energy intake and PCA

1-3, identified 418 SNPs on chromosome 12 that were associated with confections intake with

genome-wide significance ($P<5\times10^{-8}$). SNPs with the 10 lowest P-values that were associated with

confections intake are shown in Table2. The quantile-quantile plot of the observed P-values is shown in

Figure 1. The inflation factor of the genome-wide scan was 1.019 (95% confidence interval:

1.018-1.021), indicating that the population structure was well-adjusted. Figure 2 shows scatter plots of

P-values derived from genome-wide scan results for confections intake, which found 418 SNPs at 12q24

with genome-wide significance ($P<5\times10^{-8}$). SNPs with the 10 lowest P-values were located on nine

genes, namely the *BRAP, ACAD10, ALDH2, ADAM1A, MAPKAPK5, TMEM116, NAA25, TRAFD1,* and

*HECTD4* regions (Supplemental Figure). After adjustment for alcohol intake (g/day), no variant was

associated with confections intake with genome-wide significance ($P<5\times10^{-8}$). The quantile-quantile

plot of the observed P-values is shown in Figure 3. The inflation factor of the genome-wide scan was

1.017 (95% confidence interval: 1.016-1.019), indicating that the population structure was well-adjusted.

The results of sensitivity analyses on Western and Japanese confections intakes separately yielded

similar results.


*Replication of previously reported SNPs*

The results of a replication study in our J-MICC GWAS data, which adjustment for age, sex, PCA 1 to 3,

and total energy intake on the 20 SNPs that were previously reported to be associated with sweet taste

perception or confections intake are shown in Table 3. None of the SNPs reported were statistically

significant ($P<0.05/20=0.0025$) in our J-MICC data.

Accepted manuscript

**Discussion**

In the present GWAS adjusted for age, sex, total energy intake and PCA 1-3 in a Japanese population, we found 418 SNPs located on 12q24 genes; the *BRAP, ACAD10, ALDH2, ADAM1A, MAPKAPK5, TMEM116, NAA25, TRAFD1, and HECTD4* regions on 12q24.12-13 were significantly associated with confections intake before adjustment for alcohol intake. However, after adjustment for alcohol intake (g/day), no variants were associated with confections intake with genome-wide significance.

Previous studies suggested several SNPs within *TAS1R2, GNAT3, GLUT2, FGF21* and *FTO* were associated with the perception of confections and sugar [1-11]. TAS1R2 (taste receptor type 1 member 2) is a protein encoded by the *TAS1R2* gene in humans. Sweet taste receptor gene family members *TAS1R2* and *TAS1R3* are G-coupled protein genes located on chromosome 1. However, from our GWAS results, no SNPs on chromosome 1, as well as others on chromosomes 3, 4, 7, 16, and 19 were significantly associated with confections intake in replication analysis. Regarding other genes found to be apparently associated with confections intake, analysis with adjustment for alcohol intake in addition indicated they could not be causal genes; descriptions are shown in **Supplementary Information on Genes.**

In our GWAS results, rs671 was strongly associated with confections intake, with the lowest P-value before adjustment for alcohol intake. Rs 671 is a missense mutation in the aldehyde dehydrogenase 2 (*ALDH2*) gene. *ALDH2* is an important enzyme in alcohol metabolism that catalyzes the reaction of acetaldehyde oxidization to acetic acid. In Asian populations, an rs671 mutation replaces

Accepted manuscript

Glu at codon position 504 with Lys.[35] This mutation reduces the activity of ALDH2 [36-39]. In

previous studies, ALDH2 was associated with ischemic stroke [40], chronic kidney diseases [41], blood

pressure [42], primary biliary cirrhosis [43], esophageal cancer [37, 39], and hip fracture [44].

In the present study, most SNPs on chromosome 12 found to be associated with confections

consumption were in strong LD with the *ALDH2* gene. After adjustment for alcohol intake (g/day),

however, no variants were associated with confections intake with genome-wide significance

$(P<5\times10^{-8})$. The fact that the associations between SNPs on chromosome 12 and confections intake were

strongly confounded by alcohol drinking status implies the importance of gene-environmental factor

interactions. The rs671 mutation directly affects the amount of alcohol consumed. Although previous

studies reported genetic variants associated with sweet taste perception and the intake of sweet foods,

the present study suggested that in Japanese alcohol drinking has a much larger influence on eating

habits related to sweet foods than the taste receptor. Kawafune, et al. found a strong direct association

between the effective allele of rs671 and a five-point scale of sweet taste preference. The relation was

attenuated by adjusting for alcohol consumption. However, in that study, the association between rs671

and sweet taste preference was not significant among participants who hardly drank alcohol at all. If

rs671 is the true causal SNP for sweet taste preference, the relationship should have been seen in

non-drinkers. Interestingly, interaction between alcohol drinking and foods intake was noted in a

European population. According to the study in a French population by Kesse, et al. increasing alcohol

consumption was associated with a higher total energy intake, a higher percentage of energy intake as

Accepted manuscript

protein and lipids. Conversely, energy provided by carbohydrates and intake of fruits decreased with

increasing alcohol consumption [45].

*BRAP-ALDH2* loci, including rs671, were found to be associated with sweet foods

consumption before adjustment for alcohol intake in the present study, and it has been reported that this

region is among several identified loci with strong, and very recent (around the past 2000-3000 years),

evolutionary selection signatures in Japanese people in a recent whole-genome sequencing study by

Okada et al. [46]. Using high-depth whole genome

sequence data of 2,234 Japanese individuals, they calculated singleton density score (SDS), which can

detect the signatures of very recent natural selection in a single population. *BRAP-ALDH2* loci was the

one of four loci in which genome-wide SDS P-values demonstrated significant natural selection

pressure that satisfied the genome-wide significance threshold.

This study has several limitations. We did not perform a replication study in a different

Japanese population because the present results were negative from a GWAS point of view. A

replication study in a European population, however, would probably yield quite different results from

those we found in the present study, since ALDH2 polymorphism is restricted in Eastern Asian

populations. Second, although we used a semi-quantitative FFQ to estimate food intake as reported

previously [17-21], the number of sweet foods included in the FFQ was small, and data on the intakes

of added sugar or sugar sweetened beverages, as well as sweetness intensity elicited by the "sweet

foods" were not included, thus some difference from actual sweet foods intake is inevitable. Although

Accepted manuscript

fruit intake may be of interest, a separate GWAS should be performed reflecting a number of cohort studies on the associations between fruit or vegetable consumption and outcomes, such as cancer or cardiovascular disease [47, 48]. Third, in the present study, confections contributed to 2.8% of the total calory intake. Although this contribution appears small, we know that previous observational studies of several foods and beverages that had small contribution to total calorie intake, such as sugar, fruit and vegetable, reported significant impact on outcomes [47-49]. Thus, we the significance of our findings is not small.

In conclusion, we found a significant number of SNPs located on 12q24 genes that were associated with confections intake before adjustment for alcohol intake. However, all lost significance after adjustment for alcohol intake.

Accepted manuscript

**Conflict of Interest**

None of us have any conflict of interests.

**Authorship**

The authors' responsibilities were as follows—TS, YaN, YK and KW: designed the research;

 YaN, AS, NI, CG, KenM, AK, KatM, KT, MH, HI, MM, TT, DN, KeiM, IO, NK, EO, HM, YoN, MW, SS, SKK, KA, KK, YM, MK, KT, YK and KW: conducted the research; TS, YaN, AN, AS and NI : analyzed data and performed statistical analysis; TS, YaN and YD: wrote the manuscript and had primary responsibility for final content; and all authors: read and approved the final manuscript.

Accepted manuscript

**Supplementary material**

For supplementary material referred to in this article, please visit

https://*******

Accepted manuscript

**References**

1    Dias AG, Eny KM, Cockburn M, *et al*. (2015) Variation in the TAS1R2 gene, sweet taste perception and intake of sugars. *J Nutrigenet Nutrigenomics* **8**,81–90.

2    Fushan AA, Simons CT, Slack JP, *et al*. (2009) Allelic polymorphism within the TAS1R3 promoter is associated with human taste sensitivity to sucrose. *Curr Biol* **19**,1288–1293.

3    Han P, Keast RSJ, Roura E. (2017) Salivary leptin and TAS1R2/TAS1R3 polymorphisms are related to sweet taste sensitivity and carbohydrate intake from a buffet meal in healthy young adults. *Br J Nutr* **118**, 763-770.

4    Chamoun E, Hutchinson JM, Krystia O, *et al.* (2018) Single nucleotide polymorphisms in taste receptor genes are associated with snacking patterns of preschool-aged children in the Guelph Family Health Study: a pilot study. *Nutrients* **10**(2), 153.

5    Knaapila A, Hwang LD, Lysenko A, *et al*. (2012) Genetic analysis of chemosensory traits in human twins. *Chem Senses* **37**, 869–881.

6    Eny KM, Wolever TM, Corey PN, *et al.* (2010) Genetic variation in TAS1R2 (Ile191Val) is associated with consumption of sugars in overweight and obese individuals in 2 distinct populations. *Am J Clin Nutr* **92**, 1501–1510.

7    Fushan AA, Simons CT, Slack JP, *et al.* (2010) Association between common variation in genes encoding sweet taste signaling components and human sucrose perception. *Chem Senses* **35,** 579–92.

Accepted manuscript

8    Eny KM, Wolever TM, Fontaine-Bisson B, *et al.* (2008) Genetic variant in the glucose transporter type 2 is associated with higher intakes of sugars in two distinct populations. *Physiol Genomics* **33**, 355–360.

9    Chu AY, Workalemahu T, Paynter NP, *et al.* (2013) Novel locus including FGF21 is associated with dietarymacronutrient intake. *HumMol Genet* **22**, 1895–1902.

10    Holstein-Rathlou S, Grevengoed TJ, Christensen KB, et al. (2017) FGF21 is a sugar-induced hormone associated with sweet intake and preference in humans. *Cell Metab* **25**, 1045–1053.

11    Hwang LD, Lin C, Gharahkhani P, *et al.* (2019) New insight into human sweet taste: a genome-wide association study of the perception and intake of sweet substances. *Am J Clin Nutr* **109**, 1724-1737.

12    Kawafune K, Hachiya T, Nogawa S, *et al.* (2020) Strong association between the 12q24 locus and sweet taste preference in the Japanese population revealed by genome-wide meta-analysis. *J Hum Genet* **65**, 939-947

13    Suzuki T, Nakamura Y, Matsuo K, et al. (2020) A genome-wide association study on fish consumption in a Japanese population-the Japan Multi-Institutional Collaborative Cohort study. *Eur J Clin Nutr*. doi: 10.1038/s41430-020-00702-7. Online ahead of print.

14    Nakagawa-Senda H, Hachiya T, Shimizu A, *et al.* (2018) A genome-wide association study in the Japanese population identifies the 12q24 locus for habitual coffee consumption: The J-MICC Study. *Sci Rep* **8,** 1493.

15    Igarashi M, Nogawa S, Kawafune K, et al. (2019) Identification of the 12q24 locus associated with fish intake frequency by genome-wide meta-analysis in Japanese populations. *Genes Nutr* **14,** 21.

16    Hamajima N (2007) The Japan Multi-Institutional Collaborative Cohort Study (J-MICC Study) to detect gene-environment interactions for cancer. *Asian Pac J Cancer Prev* **8**, 317-323.

17    Tokudome S, Goto C, Imaeda N, *et al.* (2004) Development of a data-based short food frequency questionnaire for assessing nutrient intake by middle-aged Japanese. *Asian Pac J Cancer Prev* **5**, 40-43.

18    Imaeda N, Fujiwara N, Tokudome Y*, et al.* (2002) Reproducibility of a semi-quantitative food frequency questionnaire in Japanese female dietitians. *J Epidemiol* **12**, 45-53.

19    Goto C, Tokudome Y, Imaeda N*, et al.* (2006) Validation study of fatty acid consumption assessed with a short food frequency questionnaire against plasma concentration in middle-aged Japanese. *Scand J Nutr* **2**, 77-82.

20    Tokudome Y, Goto C, Imaeda N*, et al.* (2005) Relative validity of a short food frequency questionnaire for assessing nutrient intake versus three-day weighed diet records in middle-aged Japanese. *J Epidemiol* **15**, 135-145.

21    Wakai K. (2009) A Review of Food Frequency Questionnaires Developed and Validated in Japan. *Journal of Epidemiology* **19**, 1-11.

Accepted manuscript

22  Purcell S, Neale B, Todd-Brown K, *et al.* (2007) PLINK:a tool set for whole-genome Association

and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75.

23  Chang CC, Chow CC, Tellier LC, *et al.* (2015) Second-generation PLINK: rising to the challenge

of larger and richer datasets. *Gigascience* **4**, 7.

24  Price AL, Patterson NJ, Plenge RM, *et al.* (2006) Principal components analysis corrects for

stratification in genome-wide association studies. *Nature Genet* **38**, 904-909.

25  1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, *et al.* (2015) A global

reference for human genetic variation. *Nature* **526**, 68-74.

26  Yamaguchi-Kabata Y, Nakazono K, Takahashi A, *et al.* (2008) Japanese population structure,

based on SNP genotypes from 7003 individuals compared to other ethnic groups: effects on

population-based association studies. *Am J Hum Genet* **83**, 445-456.

27  Delaneau O, Marchini J, Zagury J. (2011) A linear complexity phasing method for thousands of

genomes. *Nature Methods* **9**, 179-181.

28  Das S, Forer L, Schönherr S, *et al.* (2016) Next-generation genotype imputation service and

methods. *Nature Genet* **48**, 1284-1287.

29  Das S. DosageConvertor https://genome.sph.umich.edu/wiki/DosageConvertor.

30  Delongchamp R, Faramawi MF, Feingold E, *et al.* (2018) The Association between SNPs and a

Quantitative Trait: Power Calculation. *Eur J Environ Public Health* **2**, 1-7.

31  Visscher PM, Wray NR, Zhang Q, *et al.* (2017) 10 Years of GWAS Discovery: Biology, Function,

Accepted manuscript

and Translation. *Am J Hum Genet* **101**, 5-22

32    Gogarten SM, Bhangale T, Conomos MP, *et al.* (2012) GWAS Tools: an R/Bioconductor package

for quality control and analysis of genome-wide association studies. *Bioinformatics* **28**,

3329-3331.

33    Turner, S.D. qqman: an R package for visualizing GWAS results using Q-Q and Manhattan plots.

biorXiv 2014 DOI: 10.1101/005165.

34    Pruim RJ, Welch RP, Sanna S, *et al.* (2010) LocusZoom: regional visualization of genome-wide

association scan results. *Bioinformatics* **26**, 2336-2337.

35    Li H, Borinskaya S, Yoshimura K, *et al.* (2009) Refined Geographic Distribution of the Oriental

ALDH2∗504Lys (nee 487Lys) Variant. *Annals of Human Genetics* **73,**335–345

36    Koganebuchi K, Haneji K, Toma T, *et al.* (2017) The allele frequency of ALDH2 Glu504Lys and

ADH1B Arg47His for the Ryukyu islanders and their history of expansion among East Asians.

*Am. J. Hum. Biol* **29**, e22933.

37    Cui R, Kamatani Y, Takahashi A, *et al.* (2009) Functional variants in ADH1B and ALDH2

coupled with alcohol and smoking synergistically enhance esophageal cancer risk.

*Gastroenterology* **137**, 1768-1775.

38    Takeuchi F, Isono M, Nabika T, *et al.* (2011) Confirmation of ALDH2 as a major locus of drinking

behavior and of its variants regulating multiple metabolic phenotypes in a Japanese population.

*Circ J* **75**, 911-918.

39    Matsuo K, Hamajima N, Shinoda M, *et al.* (2001) Gene-environment interaction between an aldehyde dehydrogenase-2 (ALDH2) polymorphism and alcohol consumption for the risk of esophageal cancer. *Carcinogenesis* **22**, 913-916.

40    NINDS Stroke Genetics Network (SiGN) and International Stroke Genetics Consortium (ISGC), Pulit SL, McArdle PF, Wong Q, *et al.* (2016) The NINDS Stroke Genetics Network: a genome-wide association study of ischemic stroke and its subtypes. *Lancet Neurol* **15**, 174-184.

41    Lee J, Lee Y, Park B, *et al.* (2018) Genome-wide association analysis identifies multiple loci associated with kidney disease-related traits in Korean populations. *PLoS One* **13**, e0194044.

42    Kato N, Loh M, Takeuchi F, *et al.* (2015) Trans-ancestry genome-wide association study identifies 12 genetic loci influencing blood pressure and implicates a role for DNA methylation. *Nature genetics* **47**, 1282-1293.

43    Cordell HJ, Han Y, Mells GF, *et al.* (2015) International genome-wide meta-analysis identifies new primary biliary cirrhosis risk loci and targetable pathogenic pathways. *Nature Communications* **6**, 8019.

44    Takeshima K, Nishiwaki Y, Suda Y, *et al.* (2017) A missense single nucleotide polymorphism in the ALDH2 gene, rs671, is associated with hip fracture. *Scientific Reports* **7**, 428.

45    E Kesse, F Clavel-Chapelon, N Slimani, *et al*. (2001) Do Eating Habits Differ According to Alcohol Consumption? Results of a Study of the French Cohort of the European Prospective Investigation Into Cancer and Nutrition (E3N-EPIC) *Am J Clin Nutr* **74**, 322-327

46     Okada Y, Momozawa Y, Sakaue S, *et al.* (2018) Deep whole-genome sequencing reveals recent

selection signatures linked to evolution and disease risk of Japanese. *Nature Communications* **9**,

1631.

47     Wang X, Ouyang Y, Liu J, *et al.* (2014) Fruit and vegetable consumption and mortality from all

causes, cardiovascular disease, and cancer: systematic review and dose-response meta-analysis of

prospective cohort studies. *BMJ* **349**, g4490:1-14

48     Aune D, Giovannucci E, Boffetta P, *et al*. (2017) Fruit and vegetable intake and the risk of

cardiovascular disease, total cancer and all-cause mortality-a systematic review and

dose-response meta-analysis of prospective studies. *Int J Epidemiol* **46**, 1029-1056.

49     Malik VS, Popkin BM, Bray GA, *et al.* (2010) Sugar-sweetened beverages and risk of metabolic

syndrome and type 2 diabetes: a meta-analysis. *Diabetes Care* **33**, 2477-2483.

Accepted manuscript

**Table 1 Baseline characteristics of the study participants**

|  | J-MICC (all) | Men | Women | P |
|---|---|---|---|---|
| Number | 14,073 | 6,329 | 7,744 | |
| Women (%) | 55.0 | | | |
| Age (years) | 54.8±9.4 | 55.4±9.3 | 54.3±9.4 | <0.001 |
| Confections intake (KJ/day) | 198±142 | 179±162 | 243±177 | <0.001 |
| Total energy intake (KJ/day) | 7,135±1,490 | 7,974±1,514 | 6,431±1,074 | <0.001 |
| Protein intake (% energy) | 12.7±2.0 | 11.8±1.8 | 13.3±1.9 | <0.001 |
| Fat intake (% energy) | 23.8± 6.4 | 20.5±5.4 | 26.6±5.9 | <0.001 |

Accepted manuscript

| | | | |
|---|---|---|---|
| Carbohydrate intake (% energy) | 55.8±6.2 | 56.8±6.7 | 54.9±5.6 | <0.001 |
| Alcohol intake (g/day) | 9.4±16.3 | 17.3±20.2 | 2.9±7.6 | <0.001 |
| BMI (kg/m$^2$) | 23.1± 3.3 | 23.8±3.2 | 22.5±3.4 | <0.001 |

Values are shown as the mean ± SD, or as percentages. P values are by Student's t-tests.

BMI=Body Mass Index, J-MICC= the Japan Multi-Institutional Collaborative Cohort study

**Table 2 SNPs with the 10 lowest P-values that were associated with confections intake in discovery samples (J-MICC study, N=14,073), adjusted for age, sex, PCA.**

| SNP | Chr | Position | Gene | EA | NEA | EAFR | BETA | SE | P |
|---|---|---|---|---|---|---|---|---|---|
| rs11066001 | 12 | 112119171 | BRAP | C | T | 0.2652 | 1.6672 | 0.1471 | 1.20E-29 |
| rs11066015 | 12 | 112168009 | ACAD10 | A | G | 0.2559 | 1.7044 | 0.1477 | 1.13E-30 |
| rs4646776 | 12 | 112230019 | ALDH2 | C | G | 0.256 | 1.6999 | 0.1476 | 1.43E-30 |
| rs671 | 12 | 112241766 | ALDH2 | A | G | 0.2563 | 1.7045 | 0.1474 | **8.62E-31** |
| rs78069066 | 12 | 112337924 | ADAM1A/ MAPKAPK5/ TMEM116 | A | G | 0.2604 | 1.6924 | 0.1492 | 1.09E-29 |
| rs11066132 | 12 | 112468206 | NAA25 | T | C | 0.2526 | 1.7512 | 0.1531 | 3.63E-30 |
| rs116873087 | 12 | 112511913 | NAA25 | C | G | 0.2562 | 1.7193 | 0.1524 | 2.12E-29 |

Accepted manuscript

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| rs12231737 | 12 | 112574616 | TRAFD1 | T | C | 0.2619 | 1.6963 | 0.1509 | 3.46E-29 |
| rs144504271 | 12 | 112627350 | HECTD4 | A | G | 0.2581 | 1.6797 | 0.1503 | 6.86E-29 |
| rs2074356 | 12 | 112645401 | HECTD4 | A | G | 0.2325 | 1.6933 | 0.1524 | 1.43E-28 |

Genome-wide analyses among the 8,504,983 variants adjusted for age, sex, total energy intake, and PCA 1-3, identified 418 SNPs on chromosome 12 that were associated with confections intake with genome-wide significance ($P<5\times10^{-8}$).

SNP=single nucleotide polymorphism, Chr=chromosome; Chromosomal position (GRCh37/hg19), EA=effect allele, NEA=non-effect allele, EA FR=effect allele frequency; Beta=effect size; SE=standard error of effect size, J-MICC= the Japan Multi-Institutional Collaborative Cohort study,

**Table 3 Replication analysis using the J-MICC samples for SNPs that were associated with confections intake in previous studies**

| Gene | SNP | Chr:Posotion | EA | NEA | EAFR | BETA | SE | P |
|------|-----|--------------|----|----|------|------|----|----|
| *TAS1R2* | rs12033832 | 1:19166294 | A | G | 0.473 | 0.330 | 0.130 | 0.011 |
| | rs3935570 | 1:19167371 | T | G | 0.066 | -0.042 | 0.261 | 0.871 |
| | rs35874116 | 1:19181393 | C | T | 0.113 | -0.102 | 0.203 | 0.616 |
| | rs121377303 | NA | | | | | | |
| | rs75346183 | 4:162614852 | A | G | 0.208 | 0.211 | 0.160 | 0.187 |
| | rs97017963 | NA | | | | | | |
| *TAS1R3* | rs307355 | 1:1265154 | C | T | 0.806 | -0.117 | 0.161 | 0.467 |
| | rs35744813 | 1:1265460 | C | T | 0.805 | -0.120 | 0.161 | 0.457 |
| *GNAT3* | rs7792845 | 7:80151369 | C | T | 0.785 | 0.228 | 0.163 | 0.163 |
| | rs940541 | 7:80150594 | G | A | 0.828 | 0.076 | 0.170 | 0.653 |

Accepted manuscript

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | rs1107660 | 7:80150131 | G | T | 0.821 | 0.061 | 0.167 | 0.713 |
| | rs1107657 | 7:80150018 | C | T | 0.821 | 0.061 | 0.167 | 0.713 |
| | rs1524600 | 7:80138303 | A | G | 0.114 | 0.006 | 0.202 | 0.975 |
| | rs6467217 | 7:80138178 | C | T | 0.114 | 0.006 | 0.201 | 0.976 |
| | rs6970109 | 7:80138074 | A | C | 0.114 | 0.006 | 0.201 | 0.976 |
| | rs6975345 | 7:80123999 | C | T | 0.111 | -0.047 | 0.203 | 0.817 |
| | rs10242727 | 7:80119730 | G | A | 0.111 | -0.042 | 0.204 | 0.837 |
| | rs6467192 | 7:80107798 | A | G | 0.111 | -0.037 | 0.204 | 0.856 |
| | rs6961082 | 7:80100969 | A | C | 0.099 | -0.001 | 0.214 | 0.995 |
| *GLUT2* | rs5400 | 3:170732300 | A | G | 0.020 | -0.712 | 0.455 | 0.118 |
| *FGF21* | rs838133 | 19:49259529 | G | A | 0.973 | -0.138 | 0.676 | 0.839 |
| *FTO* | rs11642841 | 16:53845487 | A | C | 0.098 | -0.016 | 0.217 | 0.940 |

Accepted manuscript

We carried out a replication study on the 20 identified loci (including sweet taste receptor gene family) associated with sweets intake.

SNP=single nucleotide polymorphism, Chr=chromosome, Position=chromosomal position (GRCh37/hg19), EA=effect allele, NEA=non-effect allele,

EAF=effect allele frequency, Beta=effect size, SE=standard error of effect size

Accepted manuscript

**Figure legends**

**Figure 1.**

A quantile-quantile plot (black) of genome-wide association tests. The $x$-axis indicates the expected

-$\log_{10}$ $P$-values under the null hypothesis. The $y$-axis shows the observed-$\log_{10}P$-values calculated by a

linear regression model using PLINK [23]. The

line represents $y=x$, which corresponds to the null hypothesis. The gray shaded area shows the 95%

confidence interval of the null hypothesis. The inflation factor ($\lambda$) is the median of the observed test

statistics divided by the median of the expected test statistics. An R package for creating the Q-Q plot,
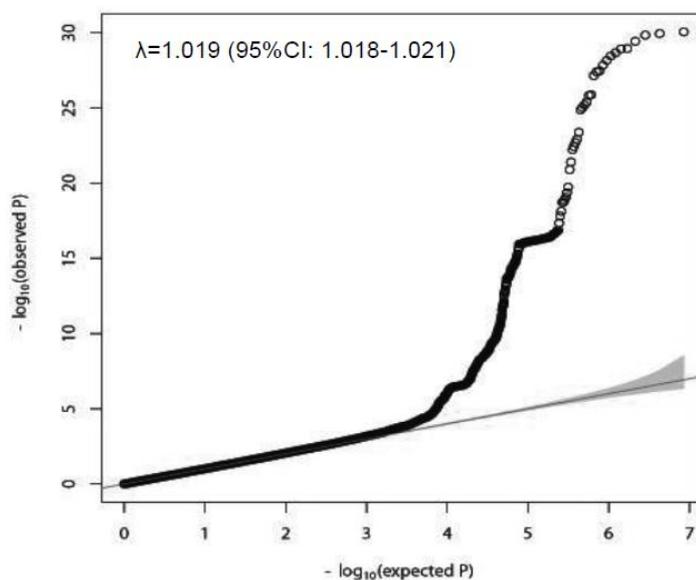
GWAS Tools, was used [32].

**Figure 2.**

Genome-wide association signals. The *x*-axis represents chromosomal positions and the *y*-axis represents -$\log_{10}P$-values calculated by a linear model association analysis. The software, qqman, was used [33].
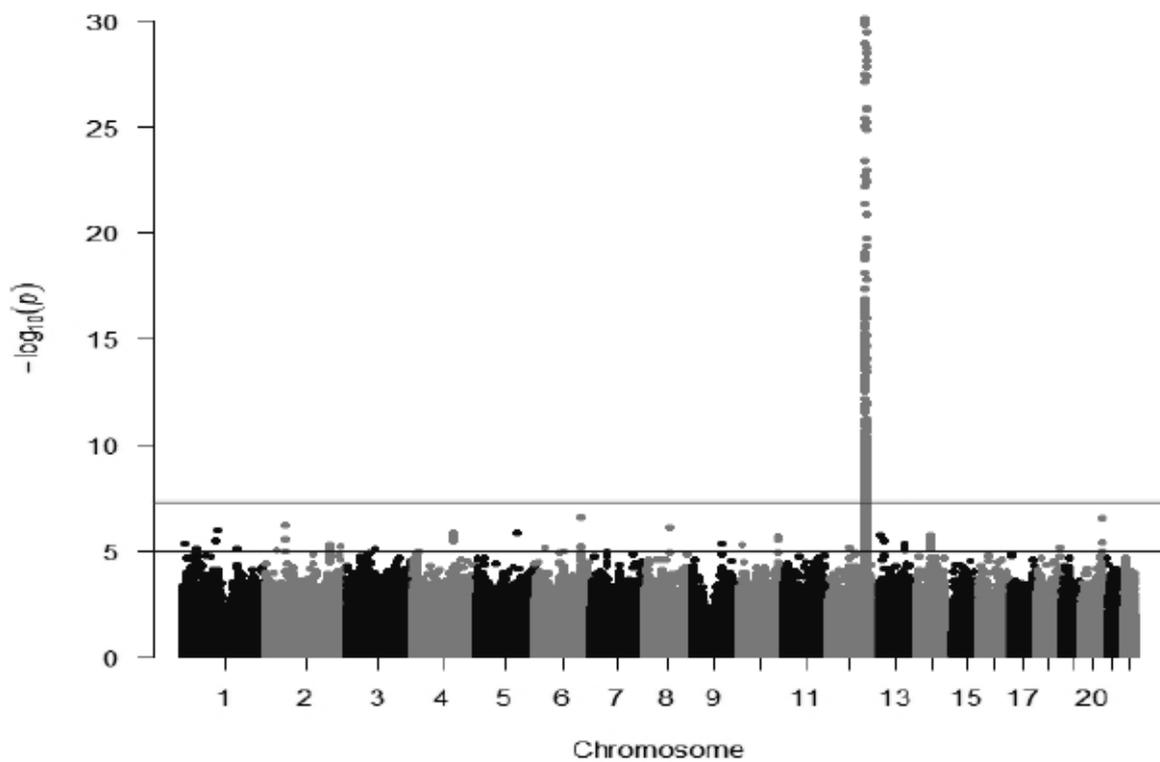
Accepted manuscript

**Figure 3.**

A quantile-quantile plot (black) of genome-wide association tests. The *x*-axis indicates the expected

-$\log_{10}$ *P*-values under the null hypothesis. The *y*-axis shows the observed-$\log_{10}$*P*-values calculated by a

linear regression model using PLINK [23]. The line represents *y=x*, which corresponds to the null

hypothesis. The gray shaded area shows the 95% confidence interval of the null hypothesis. The

inflation factor (λ) is the median of the observed test statistics divided by the median of the expected test

statistics. An R package for creating the Q-Q plot, GWAS Tools, was used [32].