Genome-wide scan identifies novel genetic loci regulating salivary metabolite levels

Abhishek Nag[1,2], Yuko Kurushima[1], Ruth C. E. Bowyer[1], Philippa M. Wells[1], Stefan Weiss[3], Maik Pietzner[4], Thomas Kocher[5], Johannes Raffler[6], Uwe Völker[3], Massimo Mangino[1], Tim D. Spector[1], Michael V. Milburn[7], Gabi Kastenmüller[6], Robert P. Mohney[7], Karsten Suhre[8], Cristina Menni[1], Claire J. Steves[1]*

[1]Department of Twin Research and Genetic Epidemiology, King's College London, London, UK

[2]Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK

[3]Interfaculty Institute of Genetics and Functional Genomics, University Medicine Greifswald & University of Greifswald, Greifswald, Germany

[4]Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany

[5]Department of Restorative Dentistry, Periodontology, Endodontology, and Preventive and Pediatric Dentistry, University Medicine Greifswald, Greifswald, Germany

[6]Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany

[7]Discovery and Translational Sciences, Metabolon, Inc., Morrisville, North Carolina, USA

[8]Department of Physiology and Biophysics, Weill Cornell Medicine-Qatar, Doha, Qatar

Abstract

Identification of hundreds of genetic loci involved in the regulation of blood metabolites has provided unprecedented insights into the pathogenesis of complex human diseases. Saliva, as a biofluid, is inexpensive and non-invasive to obtain, and provides a vital tool to investigate the interplay between oral health and systemic diseases. Here we report a genome-wide association analysis exploring 476 salivary metabolites in 1,419 subjects of European ancestry from the TwinsUK cohort (discovery phase). A total of 14 salivary metabolites were significantly associated ($p<10^{-10}$) with genetic variants that mapped to 11 distinct loci, findings that by and large replicated in an independent cohort (the Study of Health in Pomerania (SHIP-2)). A number of the locus-metabolite associations that were identified, including the ones for the *AGMAT* (with the metabolites 4-guanidinobutanoate and beta-guanidinopropanoate), *ATP13A5* (with the metabolite creatinine) and *DPYS* (with the metabolites 3-ureidopropionate and 3-ureidoisobutyrate) loci, were novel i.e. they have not been reported previously for blood

metabolites. Our study, thus, suggests that there might exist biological processes that are specific to the regulation of the salivary metabolome. In addition, some of our findings might have clinical relevance, such as the utility of the pyrimidine (uracil) degradation metabolites in predicting 5-fluorouracil toxicity and the role of the agmatine pathway metabolites as biomarkers of oral health.

Introduction

Metabolic reactions pervade every aspect of human physiology, abnormalities in which underlie a plethora of human diseases[1]. Investigating the genetic underpinnings of population-wide variation of metabolites can offer novel insights into human metabolism and diseases, in addition to providing potential therapeutic targets to modulate metabolite levels. Large-scale genetic association studies have so far identified loci that regulate the levels of hundreds of metabolites in blood[2–6], and to a lesser extent in other biospecimens as well[7–9]. Previous studies have shown that genetic variants on average explain a greater proportion of trait variance for metabolites compared to what is generally observed for complex traits[3,4], highlighting the utility of metabolites as intermediate traits for dissecting the genetics of complex diseases.

Saliva is an abundantly produced biofluid, and it can be obtained in an inexpensive and non-invasive manner, without the need for healthcare professionals. It is mainly composed of water (>99%) and several other

minor constituents such as mucous, digestive enzymes, cytokines, immunoglobulins, antibacterial peptides and low molecular weight metabolites[10].

Recent advances in metabolomic profiling allow quantification of hundreds of metabolites belonging to diverse biochemical pathways in large population samples[4,11]. In 2015, the Human Metabolome Database (HMDB) incorporated data on the 'salivary metabolome' which included 853 salivary metabolites that were systematically characterized using a multiplatform approach[12]. Since saliva is separated from the systemic circulation by just a thin layer of cells, which allows passive and active exchange of substances[13], it provides a reflection of not just oral health but the functioning of other organ systems as well[14]. Indeed, a number of studies have previously reported associations between oral health and systemic conditions such as cardiovascular diseases, diabetes, autoimmune diseases, mental health disorders, dementia, among others[15–19]. Therefore, investigation of salivary metabolites could not only provide novel biomarkers but also further our understanding of biological pathways underlying oral as well as general health conditions.

Here we report a genome-wide association analysis (based on 1000 Genomes imputed data) for 476 salivary metabolites in the population-based TwinsUK study, followed by replication in the population-based Study of Health in Pomerania (SHIP-2) cohort.

Methods

(I) Discovery phase

Study population

The discovery phase of the study was conducted in the TwinsUK cohort, an adult twin registry comprising of healthy volunteers, based at St. Thomas' Hospital in London[20]. Twins gave fully informed consent under a protocol reviewed by the St. Thomas' Hospital Local Research Ethics Committee. Subjects of European ancestry with available genotype data and for whom salivary metabolite profiling was done on a fasting state sample were included in our study (N = 1,419; mean age = 62.2 years; % females = 92.7).

Genotyping, imputation and QC

Subjects were genotyped in two different batches of approximately the same size, using two genotyping platforms from Illumina: 300K Duo and HumanHap610-Quad arrays. Whole genome imputation of the genotypes was performed using the 1000 Genomes reference

haplotypes[21], further details of which are provided in Moayyeri *et al* (2013). Stringent QC measures, including minimum genotyping success rate (>95%), Hardy-Weinberg equilibrium ($p>10^{-6}$), minimum MAF (>0.5%) and imputation quality score (INFO>0.5), retained ~9.6 million variants for genome-wide analysis.

Saliva sample collection

Saliva samples were obtained by asking the fasted volunteer to spit as much saliva as possible into an empty sterile pot over a period of 10 minutes. The saliva samples were immediately refrigerated and then frozen at -80°C (usually within 4 hours of sample collection) before further processing. Following that, the samples were shipped on dry ice for metabolite profiling at Metabolon Inc., Durham, USA (See **Supplementary methods (I)** for further details on sample processing).

Metabolic profiling of saliva samples

Metabolite concentrations in the saliva samples were estimated using Ultrahigh Performance Liquid Chromatography-Tandem Mass

Spectroscopy (UPLC-MS) i.e. chromatographic separation, followed by full-scan mass spectroscopy, to record all detectable ions in the samples (see **Supplementary methods (II)** for further details). Based on their unique ion signatures (chromatographic and mass spectral), 997 distinct metabolites were identified, of which 823 had known chemical identity at the time of analysis. The 823 known metabolites were broadly classified into 8 metabolic groups (amino acids, peptides, carbohydrates, energy, lipids, nucleotides, cofactors and vitamins, and xenobiotics) as described in the KEGG (Kyoto Encyclopedia of Genes and Genomes) database[22]. The 8 metabolic groups were further subdivided into 99 distinct biochemical pathways.

Raw metabolite values were normalised for the volume and osmolality measurement of the saliva samples. The normalised metabolite values were then log-transformed, and scaled to uniform mean 0 and standard deviation 1. Of the 823 known metabolites, 476 were retained for analysis based on presence of measurement in more than 80% samples. For the samples with missing values for these 476 metabolites, data were imputed using the run day minimum value for the metabolite. The

resulting imputed dataset of the 476 metabolites was used for further analysis.

Genome-wide association analysis of salivary metabolites

(i) Primary genome-wide association analysis

For each of the 476 metabolites, a linear mixed-model was fitted to test the association between the metabolite (dependent variable) and genome-wide variants (independent variable). Age, sex and time of saliva sample collection were included as covariates in the model. The score test implemented in GEMMA[23], which utilises a sample kinship matrix (estimated using a subset of ~500,000 variants) to account for the twin structure or relatedness in the TwinsUK data, was used to assess significance of the associations. A genome-wide and metabolome-wide significance cut-off of $p<10^{-10}$ (corresponding to the conventional genome-wide significance threshold of $5x10^{-8}$, corrected for 476 metabolites) was used to identify significant variant-metabolite

associations. For each locus that was significantly associated with a metabolite, we reported the variant with the lowest association p-value.

(ii) Testing loci identified in the primary analysis for additional metabolite associations

Next, we focused just on the loci that were identified in the primary stage of association testing to look for additional variant-metabolite associations for those loci. For each locus that was identified in the primary analysis, we clumped all variants located within a 100 Mb block and with LD (r^2) > 0.2, and checked for additional metabolite associations at a significance threshold of $p<10^{-6}$ (corresponding to p=0.05, corrected for 476 metabolites and a prior assumption of about 100 associated loci).

(iii) Testing the significantly associated loci using metabolite data from other biospecimens

For each locus-metabolite association, we further assessed the most significant variant-metabolite pair by using measurements for the

respective metabolite in serum[4] and faecal samples[7] of the TwinsUK subjects (provided the metabolite was measured in that biospecimen). We tested only those serum and faecal samples that overlapped with the ones in saliva and were collected within 5 years of the saliva samples (in order to ensure that, for a given individual, samples from the different biospecimens being tested were obtained within a certain period of one another). Association testing for serum and faecal metabolites was done using an identical model to that described for the analysis of salivary metabolites.

(iv) Conditional analysis for the significantly associated loci

(a) Detection of secondary association signals

We used approximate conditional analysis, as implemented on GCTA[24], to test whether any of the associated loci had multiple distinct i.e. secondary association signals, at a "locus-wide" significance threshold of $p<10^{-5}$. For each associated locus, all variants that surpassed the study-wide significance threshold ($p<10^{-10}$) were conditioned on the most significantly associated variant at that locus (using the association

summary statistics). For the conditional analysis, we used genotype data from the complete TwinsUK dataset (N=5,654) to model LD patterns between variants.

(b) Adjustment for periodontal disease status

Since salivary metabolites are known to be associated with periodontal disease (PD)[11], for each locus-metabolite association, we additionally adjusted the most significant variant-metabolite pair for PD status. Self-reported gingival bleeding, and history of gum disease or tooth mobility were used as indicators of PD in TwinsUK[25] (270 PD cases; 1,083 controls).

Expression quantitative trait locus (eQTL) analysis

We used the version 7 data release of the Genotype-Tissue Expression (GTEx) project (accessed 15 April 2019), based on RNA-Seq data obtained from 48 non-diseased tissue sites across ~1,000 individuals, to test whether the most significant variant at each associated locus had an eQTL effect on transcripts located within a 1 Mb window of the variant.

Annotation of associations using reference databases

We searched the NHGRI GWAS catalogue (accessed 15 April 2019) for previous disease associations for the significantly associated loci that were identified in our study. We also searched the OMIM database (accessed 15 April 2019) to check the candidate genes at the associated loci for a causal link with inborn errors of metabolism. Moreover, we also queried the HMDB[12] and KEGG[22] databases to identify biochemical pathways and known disease associations for the associated metabolites.

(II) Replication phase

The replication phase was performed in the Study of Health in Pomerania (SHIP-2), a population-based study comprising of European ancestry subjects, conducted in the northeastern area of Germany. Further details of SHIP-2, including cohort details, genotyping and imputation, and saliva sample collection are provided in **Supplementary methods (III)**. Metabolic profiling was performed for 1,000 SHIP-2 saliva samples using an identical process to that described for TwinsUK.

Since the method of saliva sample collection in SHIP-2 (chewing on a piece of cotton) meant that the sample thus obtained represented stimulated saliva, normalisation of the metabolite measurements for sample osmolality was not considered necessary. The fact that the salivary osmolality values in SHIP-2 had a much narrower distribution compared to that in TwinsUK verified our rationale (**Supplementary Figure 1**).

A linear regression model was fitted on R (version 3.5.2) to test the most significant variant-metabolite pair for each locus-metabolite association that was identified in the discovery phase. Covariates used in the association model were similar to those used for the discovery phase.

(III) Testing the associated salivary metabolites with phenotypes of interest

We wanted to test how salivary metabolites that were regulated by genetic loci related to relevant phenotypes. For that, we selected the metabolites that were uniquely associated in saliva i.e. the ones for which a genetic association had not been previously reported in blood,

and tested them with phenotypes (diseases / traits / adverse drug effects) relating to the metabolite or its associated biochemical pathway. We obtained the relevant phenotype information from the TwinsUK database, selecting one twin per pair (N=1,426). The phenotype association analysis was performed on R (version 3.5.2) by fitting a linear regression model to test the association between the salivary metabolite and the disease / trait / adverse drug effect (adjusted for age and sex).

Results

Identification of novel genetic loci regulating salivary metabolite levels

Primary genome-wide discovery analysis in TwinsUK identified 13 metabolites that were significantly associated with genetic loci after correcting for multiple testing ($p<10^{-10}$). Furthermore, when we narrowed our analysis to just the loci that were identified in the primary stage, one additional metabolite was found associated ($p<10^{-6}$). Consequently, a total of 14 distinct locus-metabolite associations (hereafter, referred to as 'mQTLs') were identified in the discovery phase of our study. The set of significantly associated variants mapped to 11 distinct genetic loci, which have been referred to by the name(s) of the overlapping or the nearest gene(s) (**Figure 1**). Three of those loci (*AGMAT, SLC2A9* and *DPYS*) were associated with two metabolites each (**Table 1**). In all three instances, the two metabolites regulated by the same locus were correlated (Pearson's r^2 for the metabolite pairs ranged between 0.42 - 0.84) (**Supplementary Figure 2**). On the other

hand, none of the 14 metabolites that was associated in our study had more than one significant locus. Quantile-quantile (QQ) plots for the 14 associated metabolites are provided in **Supplementary Figures 3 (i-xvi)**.

Of the 11 genetic loci that were identified in the discovery phase, four of them (*SLC2A9, DMGDH, FADS2* and *ACADS*) have previously been reported in association with blood metabolites[2,4,5]; the remaining seven loci were novel i.e. there were no previously known associations for them with metabolites in blood or any other biospecimens.

The *AGMAT* locus, one of the novel loci identified, was associated with the metabolites 4-guanidinobutanoate and beta-guanidinopropanoate. These metabolites are generated as intermediate products in the polyamine synthesis pathway, the main site of action for the enzyme agmatinase (encoded by the *AGMAT* gene)[26]. The most significantly associated variants for the two metabolites i.e. rs10927806 and rs6690813, respectively, are in high LD with one another (r^2=0.99), suggesting a shared underlying genetic regulation for the two metabolites by the *AGMAT* locus. Likewise, the association between the

ATP13A5 locus and creatinine (a widely used measure of renal function), observed in our study, has also not been reported previously. Another interesting novel association that we identified pertained to the metabolism of the pyrimidine uracil - the *DPYS* locus (encodes for the enzyme dihydropyrimidinase, involved in uracil degradation) was associated with the metabolites 3-ureidopropionate and 3-ureidoisobutyrate (breakdown products of uracil metabolism). The associations of the *SLC2A9* locus with allantoin, the *ABO* locus with N-acetylglucosamine/N-acetylgalactosamine, the *UGCG* locus with glycosyl-N-stearoyl-sphinganine, the *MGP* locus with gamma-carboxyglutamate, and the *TYMS / ENOSF1* locus with ribonate were the remaining mQTLs observed in our study that have not been reported previously.

Of the 14 distinct mQTLs that were identified in our study, it was possible to test the most significant variant-metabolite pair for nine mQTLs in both serum and faecal metabolite data in TwinsUK (the remaining metabolites were not present in the serum and faecal metabolite datasets). Of them, the associations for the *ATP13A5* and

DPYS loci did not replicate in serum ($p>0.05$); while, none of the associations, barring the one for the *ACADS* locus, replicated in the faecal data ($p>0.05$) (**Supplementary Table 1**). Thus, a comparison across all three biospecimens (for the significantly associated salivary metabolites which were also measured in serum and faecal samples in TwinsUK) demonstrates that the effects of the *ATP13A5* and *DPYS* loci were specific to saliva (**Supplementary Figure 4**).

For none of the mQTLs that was identified, did we find any additional independent association signals after conditioning for the most significant variant at the locus (conditional $p>10^{-5}$ for all variants tested at each locus), a finding which was verified by regional association plots (**Supplementary Figures 5 (i-xiii)**). The observation that a single genetic association signal underlies each of the significantly associated metabolites might partly be due to our lack of power to detect secondary signals at these loci.

Moreover, the strength of association for the most significant variant-metabolite pair did not change much on adjusting for PD status for any of the mQTLs (**Supplementary Table 2**). Hence, it does not appear that

the presence of underlying PD has a significant effect on the associations observed in our study.

For 8 of the 11 associated loci, it was observed that the most significant variant demonstrated an eQTL effect on at least one transcript in one of the tissues in the GTEx database (the most significant variants at the *ATP13A5*, *SLCA2A9* and *ABO* loci did not show any significant eQTL effects) (**Supplementary Table 3**). In case of 7 of those 8 loci (except *DPYS*), a significant eQTL effect was observed on the overlapping or the nearest gene transcript; and for 5 of those 7 loci (*AGMAT*, *FADS2*, *DMGDH*, *TYMS/ENOSF1* and *UGCG*), a significant eQTL effect was observed in one of the gut-related tissues. While eQTL data for transcripts assayed in the minor salivary glands was available for a small number of donors (N=97) in the GTEx database, none of the significant eQTL effects that we observed was in the salivary tissue.

Replication of the discovery phase findings

In SHIP-2, we could attempt replication for 9 of the 14 mQTLs that were identified in the discovery phase (metabolites corresponding to the

remaining five mQTLs were not measured in SHIP-2). In the initial replication analysis, which was performed using salivary metabolite data that was not normalised for sample osmolality, 8 of the 9 discovery phase associations were replicated ($p<0.05$), with the direction of effect consistent with that observed in TwinsUK (**Table 2**). The association between the *ABO* locus and N-acetylglucosamine/N-acetylgalactosamine was the only finding from the discovery phase that did not replicate in SHIP-2. When we repeated the replication analysis in SHIP-2 with metabolite data that was normalised for the sample osmolality, the strength of all the associations was comparatively reduced (**Supplementary Table 4**).

Phenotypic associations for salivary metabolites

Majority of the loci that were associated in our study have been reported in relation with GWAS traits, inborn errors of metabolism and / or clinically relevant biochemical pathways (**Table 3**). We further tested the salivary metabolites associated with the *DPYS*, *AGMAT* and

ATP13A5 loci in relation with specific phenotypes using information available in the TwinsUK database (**Table 4**).

The *AGMAT*-associated metabolites (4-guanidinobutanoate and beta-guanidinopropanoate) are generated in the polyamine synthesis pathway (https://www.genome.jp/kegg-bin/show_pathway?hsa00330). This pathway also produces the compound putrescine, which has been implicated in poor oral health and foul breath[27,28]. Consequently, we evaluated the significance of the *AGMAT*-associated metabolites in oral health by testing them with PD status - the levels of both 4-guanidinobutanoate and beta-guanidinopropanoate were significantly higher in PD cases compared to controls ($p=0.0003$ and $p=0.0006$, respectively). Since eGFR (estimated glomerular filtration rate) is calculated on the basis of serum creatinine, and these two commonly used measures of renal function are negatively correlated, we wanted to investigate the association between salivary creatinine and eGFR. We observed a similar strong negative relationship between salivary creatinine and eGFR ($p=1.6 \times 10^{-11}$), which was indicative of a homeostasis between creatinine concentrations in serum and saliva.

Furthermore, creatinine is also known to be a marker of muscle mass and strength[29]. We, therefore, tested salivary creatinine in relation with grip strength (a measure of muscle strength), which suggested a positive correlation between them (p=0.01).

Apart from these findings, the remaining phenotypic associations that we tested were largely negative, as follows:

The enzyme agmatinase (encoded by the *AGMAT* gene), which acts on the substrate agmatine, has been implicated in the pathophysiology of mood disorders[30]. Moreover, studies have also proposed agmatine as a novel neuromodulator[31]. We, therefore, tested the *AGMAT*-associated metabolites in relation with a diagnosis of clinical depression or anxiety disorder, and responses (on a ordinal scale) to questions in the Hospital Anxiety and Depression Scale (HADS) questionnaire[32]. But, neither analysis showed any significant association (**Supplementary Table 5**).

ATP13A5, the locus that was associated with salivary creatinine, belongs to the family of ATPases that regulate the activity of HMG-CoA reductase[33], the main site of action of the cholesterol-lowering class of drugs called statins. Statins are known to cause muscle dysfunction

(myopathy) in a small fraction of patients[34]. Given that *ATP13A5* and statins both act in the same biochemical pathway, we assessed whether salivary creatinine is also associated with statin usage, and could therefore be used as a biomarker for statin-induced myopathy. There was, however, no association between salivary creatinine and statin usage (p=0.22).

5-Fluorouracil (5-FU) is a pyrimidine analogue that is a commonly used anticancer drug. Since it is eliminated from the body via the pyrimidine degradation pathway, variants in genes coding for the pyrimidine degradation enzymes (for instance, *DPYS*) are known to be associated with the development of 5-FU toxicity[35], which mainly manifests as gastrointestinal side effects. In order to understand whether the levels of the *DPYS*-associated metabolites (3-ureidopropionate and 3-ureidoisobutyrate) could be utilised as predictors of 5-FU toxicity, we tested their association with a commonly observed phenotype of gastrointestinal dysfunction, irritable bowel syndrome or IBS (ascertained using the ROME-III questionnaire[36]). We observed that for both metabolites, 3-ureidopropionate and 3-ureidoisobutyrate, the levels

were not significantly different in IBS “cases” compared to “controls” ($p>0.05$, for both metabolites). However, this negative finding does not negate the possibility that these metabolites could be of clinical use in predicting 5-FU toxicity.

Discussion

Here we report a genome-wide association analysis of 476 metabolites measured in saliva samples of healthy population-based studies of European descent. We identified a total of 11 distinct genetic loci that regulate the level of 14 salivary metabolites, of which three loci were associated with more than one metabolite each.

The fact that saliva is reflective of the concentration of biochemicals in blood forms the basis for certain clinical applications of saliva that others have proposed previously such as therapeutic monitoring of drugs[37], cortisol measurement[38] and renal function monitoring[39]. Using salivary metabolite data, we replicated the associations for certain well-established genetic loci that are known to regulate the level of blood metabolites. Thus, our findings add further credence to the notion that, as biofluids, a degree of homeostasis exists between blood and saliva.

Additionally, we identified some novel associations in our study, which have expanded our knowledge of genetic influences on human

metabolites. In particular, the association between the *DPYS* locus and pyrimidine metabolites is intriguing because of its clinical relevance. Mutations in the pyrimidine catabolism genes such as *DPYD* (encodes dihydropyrimidine dehydrogenase) and *DPYS* (encodes dihydropyrimidinase) have been linked to inborn errors of metabolism[40,41] as well as development of severe toxicity to the chemotherapeutic agent 5-FU[27,42,43]. Studies have previously demonstrated the applicability of salivary measurement of certain pyrimidine pathway metabolites (uracil and dihydrouracil) for evaluating 5-FU toxicity due to deficient *DPYD* activity[44]. In our study, variants in the *DPYS* gene correlated with the levels of specific pyrimidine metabolites (3-ureidopropionate and 3-ureidoisobutyrate). Therefore, studies to explore the utility of salivary measurements of these metabolites as non-invasive tools for predicting 5-FU toxicity resulting from mutations that affect *DPYS* activity are warranted. While we could not test 3-ureidopropionate and 3-ureidoisobutyrate in relation to the gastrointestinal side effects of 5-FU toxicity, we did not find any association between these metabolites and a phenotype relevant to gut

dysfunction (IBS phenotype). The other novel finding of note was the association between the *AGMAT* locus and the metabolites 4-guanidinobutanoate and beta-guanidinopropanoate. The *AGMAT*-associated metabolites are produced in the polyamine synthesis pathway, which also generates the compound putrescine that has been implicated in oral health. Moreover, in our study, the *AGMAT*-associated metabolites were correlated with PD status, a disease related to poor oral health. Together, these findings suggest that the *AGMAT*-associated metabolites might serve as potential biomarkers for oral health. On the other hand, though the agmatine pathway has been previously implicated in mood disorders, we did not find any association between the *AGMAT*-associated metabolites and either symptoms of or a prior diagnosis of clinical depression or anxiety disorder. Thus, evidence based on the *AGMAT*-associated metabolites does not lend support to the hypothesis regarding a potential link between pathways involved in maintaining oral health and regulation of mood[15,18]. On the other hand, there were a few other novel associations of interest in our study, such as the metabolite gamma-carboxyglutamate (associated with the *MGP* locus,

which is known to cause abnormal vascular calcification in patients with cardiovascular disease[45]), for which we could not assess the clinical significance since specific phenotypic information was not available in sufficient numbers in those with salivary metabolite data.

Given that we identified some novel genetic associations for salivary metabolites i.e. genetic loci that had not been reported in association with blood metabolites, it is indicative of the presence of biological processes that might be specific to the regulation of certain metabolites in the saliva. For the novel genetic loci, in most cases, the overlapping or the nearest gene transcript was expressed in one or more gut-related tissues, including salivary glands. However, for these loci, we did not find much evidence for cis-eQTL effects specific to salivary or other gut tissues, whereby we could attribute their association with the respective salivary metabolite(s) to transcriptional regulation of neighbouring genes. Interestingly, there is growing evidence to suggest that the human metabolome is a reflection of an interaction between the host and the gut microbiome[7,46,47] – it will therefore be worth exploring whether the

salivary metabolites that were associated in our study correlate with the composition of the salivary microbiome. Attempting to elucidate, in this manner, biological processes that might mediate the observed associations between genetic loci and salivary metabolites will help broaden our understanding of the underlying pathways.

In summary, our study has provided an initial map of genetic loci that regulate the salivary metabolome. Based on what has been observed for other complex human traits, future studies with larger sample sizes are expected to uncover additional genetic loci with much smaller effects on salivary metabolites. Furthermore, our study also offered insights into hitherto unknown biological pathways involved in maintaining the levels of salivary metabolites. While we did explore, to an extent, the clinical relevance of a few salivary metabolites of interest, a more comprehensive analysis with a wider range of phenotypic domains will help broaden our understanding of the relationship between the salivary metabolome and systemic health conditions.

URLs

1000 Genomes project: http://www.internationalgenome.org/

Metabolon: https://www.metabolon.com/

GEMMA: http://www.xzlab.org/software.html

KEGG: https://www.genome.jp/kegg/

GCTA: http://cnsgenomics.com/software/gcta/#COJO

GTEx: https://gtexportal.org/home/

NHGRI GWAS catalogue: https://www.ebi.ac.uk/gwas/

OMIM: http://omim.org/

HMDB: http://www.hmdb.ca/

LocusZoom: http://locuszoom.org/

References

1. DeBerardinis, R. J. & Thompson, C. B. Cellular metabolism and disease: what do metabolic outliers teach us? *Cell* **148,** 1132–1144 (2012).
2. Illig, T. *et al.* A genome-wide perspective of genetic variation in human metabolism. *Nat. Genet.* **42,** 137–141 (2010).
3. Kettunen, J. *et al.* Genome-wide association study identifies multiple loci influencing human serum metabolite levels. *Nat. Genet.* **44,** 269–276 (2012).
4. Shin, S.-Y. *et al.* An atlas of genetic influences on human blood metabolites. *Nat. Genet.* **46,** 543–550 (2014).
5. Long, T. *et al.* Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites. *Nat. Genet.* **49,** 568–578 (2017).
6. Gieger, C. *et al.* Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum. *PLoS Genet.* **4,** e1000282 (2008).
7. Zierer, J. *et al.* The fecal metabolome as a functional readout of the gut microbiome. *Nat. Genet.* **50,** 790–795 (2018).
8. Suhre, K. *et al.* A genome-wide association study of metabolic traits in human urine. *Nat. Genet.* **43,** 565–569 (2011).
9. Luykx, J. J. *et al.* Genome-wide association study of monoamine metabolite levels in human cerebrospinal fluid. *Mol. Psychiatry* **19,** 228–234 (2014).
10. Soini, H. A. *et al.* Analysis of volatile organic compounds in human saliva by a static sorptive extraction method and gas chromatography-mass spectrometry. *J. Chem. Ecol.* **36,** 1035–1042 (2010).
11. Barnes, V. M. *et al.* Global metabolomic analysis of human saliva and plasma from healthy and diabetic subjects, with and without periodontal disease. *PLoS One* **9,** e105181 (2014).
12. Wishart, D. S. *et al.* HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res.* **46,** D608–D617 (2018).
13. Spielmann, N. & Wong, D. T. Saliva: diagnostics and therapeutic perspectives. *Oral Dis.* **17,** 345–354 (2011).
14. Chiappin, S., Antonelli, G., Gatti, R. & De Palo, E. F. Saliva specimen: a new laboratory tool for diagnostic and basic investigation. *Clin. Chim. Acta.* **383,** 30–40 (2007).
15. Kisely, S., Baghaie, H., Lalloo, R., Siskind, D. & Johnson, N. W. A systematic review and meta-analysis of the association between poor oral health and severe mental illness. *Psychosom. Med.* **77,** 83–92 (2015).
16. Kaufman, E. & Lamster, I. B. The diagnostic applications of saliva--a review. *Crit. Rev. Oral Biol. Med.* **13,** 197–212 (2002).
17. de Oliveira, C., Watt, R. & Hamer, M. Toothbrushing, inflammation, and risk of cardiovascular disease: results from Scottish Health Survey. *BMJ* **340,** c2451 (2010).
18. O'Neil, A. *et al.* The association between poor dental health and depression: findings from a large-scale, population-based study (the NHANES study). *Gen. Hosp. Psychiatry* **36,** 266–270 (2014).
19. Noble, J. M. *et al.* Periodontitis is associated with cognitive impairment among older adults: analysis of NHANES-III. *J. Neurol. Neurosurg. Psychiatry* **80,** 1206–1211 (2009).
20. Moayyeri, A., Hammond, C. J., Valdes, A. M. & Spector, T. D. Cohort Profile: TwinsUK and healthy ageing twin study. *Int. J. Epidemiol.* **42,** 76–85 (2013).

21. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526,** 68–74 (2015).
22. Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. & Tanabe, M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* **40,** D109–14 (2012).
23. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **44,** 821–824 (2012).
24. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44,** 369–75, S1–3 (2012).
25. Kurushima, Y. *et al.* Epigenetic findings in periodontitis in UK twins: a cross-sectional study. *Clin. Epigenetics* **11,** 27 (2019).
26. Dudkowska, M. *et al.* Agmatine modulates the in vivo biosynthesis and interconversion of polyamines and cell proliferation. *Biochim. Biophys. Acta* **1619,** 159–166 (2003).
27. Goldberg, S. *et al.* Cadaverine as a putative component of oral malodor. *J. Dent. Res.* **73,** 1168–1172 (1994).
28. Porter, S. R. & Scully, C. Oral malodour (halitosis). *BMJ* **333,** 632–635 (2006).
29. Patel, S. S. *et al.* Serum creatinine as a marker of muscle mass in chronic kidney disease: results of a cross-sectional study and review of literature. *J. Cachexia. Sarcopenia Muscle* **4,** 19–29 (2013).
30. Bernstein, H.-G. *et al.* Agmatinase, an inactivator of the putative endogenous antidepressant agmatine, is strongly upregulated in hippocampal interneurons of subjects with mood disorders. *Neuropharmacology* **62,** 237–246 (2012).
31. Regunathan, S. Agmatine: biological role and therapeutic potentials in morphine analgesia and dependence. *AAPS J.* **8,** E479–84 (2006).
32. Stern, A. F. The hospital anxiety and depression scale. *Occup. Med. (Lond).* **64,** 393–394 (2014).
33. Roitelman, J. & Simoni, R. D. Distinct sterol and nonsterol signals for the regulated degradation of 3-hydroxy-3-methylglutaryl-CoA reductase. *J. Biol. Chem.* **267,** 25264–25273 (1992).
34. Sathasivam, S. & Lecky, B. Statin induced myopathy. *BMJ* **337,** a2286 (2008).
35. Akai, F., Hosono, H., Hirasawa, N. & Hiratsuka, M. Novel single nucleotide polymorphisms of the dihydropyrimidinase gene (DPYS) in Japanese individuals. *Drug Metab. Pharmacokinet.* **30,** 127–129 (2015).
36. Longstreth, G. F. *et al.* Functional bowel disorders. *Gastroenterology* **130,** 1480–1491 (2006).
37. Drobitch, R. K. & Svensson, C. K. Therapeutic drug monitoring in saliva. An update. *Clin. Pharmacokinet.* **23,** 365–379 (1992).
38. Aardal-Eriksson, E., Karlberg, B. E. & Holm, A. C. Salivary cortisol--an alternative to serum cortisol determinations in dynamic function tests. *Clin. Chem. Lab. Med.* **36,** 215–222 (1998).
39. Goll, R. D. & Mookerjee, B. K. Correlation of biochemical parameters in serum and saliva in chronic azotemic patients and patients on chronic hemodialysis. *J. Dial.* **2,** 344–399 (1978).
40. Hamajima, N. *et al.* Dihydropyrimidinase deficiency: structural organization, chromosomal localization, and mutation analysis of the human dihydropyrimidinase gene. *Am. J. Hum. Genet.* **63,** 717–726 (1998).
41. Berger, R. *et al.* Dihydropyrimidine dehydrogenase deficiency leading to thymine-

uraciluria. An inborn error of pyrimidine metabolism. *Clin. Chim. Acta.* **141,** 227–234 (1984).
42. Wei, X., McLeod, H. L., McMurrough, J., Gonzalez, F. J. & Fernandez-Salguero, P. Molecular basis of the human dihydropyrimidine dehydrogenase deficiency and 5-fluorouracil toxicity. *J. Clin. Invest.* **98,** 610–615 (1996).
43. van Kuilenburg, A. B. *et al.* Clinical implications of dihydropyrimidine dehydrogenase (DPD) deficiency in patients with severe 5-fluorouracil-associated toxicity: identification of new mutations in the DPD gene. *Clin. Cancer Res.* **6,** 4705–4712 (2000).
44. Neto, O. V. *et al.* DPD functional tests in plasma, fresh saliva and dried saliva samples as predictors of 5-fluorouracil exposure and occurrence of drug-related severe toxicity. *Clin. Biochem.* **56,** 18–25 (2018).
45. Mizobuchi, M., Towler, D. & Slatopolsky, E. Vascular calcification: the killer of patients with chronic kidney disease. *J. Am. Soc. Nephrol.* **20,** 1453-1464 (2009).
46. Chen, M. X., Wang, S.-Y., Kuo, C.-H. & Tsai, I.-L. Metabolome analysis for investigating host-gut microbiota interactions. *J. Formos. Med. Assoc.* **118 Suppl 1,** S10–S22 (2019).
47. Peisl, B. Y. L., Schymanski, E. L. & Wilmes, P. Dark matter in host-microbiome metabolomics: Tackling the unknowns-A review. *Anal. Chim. Acta* **1037,** 13–27 (2018).
48. Alfonso, P. *et al*. Characterization of variants in the glucosylceramide synthase gene and their association with type 1 Gaucher disease severity. *Hum. Mutat.* **34,** 1396-1403 (2013).
49. Hirano, H. *et al.* Association of natural tooth loss with genetic variation at the human matrix Gla protein locus in elderly women. *J. Hum. Genet.* **48,** 288-292 (2003).

Acknowledgements

C.J.S. acknowledges funding from the Chronic Disease Research Foundation and the Wellcome Trust (grant WT081878MA). K.S. was supported by the Biomedical Research Program at Weill Cornell Medicine in Qatar, a program funded by the Qatar Foundation.

TwinsUK is funded by the Wellcome Trust, Medical Research Council, European Union, the National Institute for Health Research (NIHR)-funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St. Thomas' NHS Foundation Trust in partnership with King's College London.

The SHIP cohort study is part of the Community Medicine Research Net (http://www.medizin.uni-greifswald.de/cm) of the University of Greifswald, Germany, which is funded by the German Federal Ministry of Education and Research (BMBF, grant no. 01ZZ96030, 01ZZ0701); the Ministry for Cultural Affairs and the Ministry for Social Affairs of the Federal State of Mecklenburg-West Pomerania (SHIP; http://www.medizin.uni-greifswald.de/cm).

Author contributions

Conceived and designed the experiments: C.J.S., C.M., T.D.S., G.K., R.P.M., M.V.M, J.R., U.V.

Performed the experiments: R.P.M.

Analysed the data: A.N., K.S., Y.K., P.M.W., R.C.E.B., T.K., M.P., S.W., M.M.

Wrote the manuscript: A.N., C.J.S., C.M., K.S., R.P.M.

All authors revised the manuscript

Competing interests

R.P.M. is an employee of Metabolon, Inc. and, as such, has affiliations with or financial involvement with Metabolon, Inc.

Figure legends

Figure 1: Manhattan plot illustrating the findings of the discovery phase (TwinsUK) of the genome-wide association study for salivary metabolites (mGWAS)

The red horizontal line demarcates the study-wide significance threshold of 10^{-10}. Eleven loci surpassed the study-wide significance threshold. The loci are referred to by the name(s) of the overlapping or the nearest gene(s).

Supplementary figure 1: Comparison of the distribution of saliva sample osmolality in TwinsUK and SHIP-2

The distribution of saliva sample osmolality observed in SHIP-2 was much narrower in comparison to TwinsUK, indicative of the fact that, in contrast to TwinsUK, the saliva samples in SHIP-2 represented stimulated saliva.

Supplementary figure 2: Plots demonstrating the correlation between the pairs of metabolites that were associated with the same locus in the discovery phase

The figure demonstrates the correlation between the following three pairs of metabolites that were regulated by the same locus:

(i) beta-guanidinopropanoate and 4-guanidinobutanoate (*AGMAT*)

(ii) urate and allantoin (*SLC2A9*)

(iii) 3-ureidopropionate and 3-ureidoisobutyrate (*DPYS*)

Supplementary figure 3: Quantile-quantile (QQ) plots of the genome-wide association analysis for the 14 salivary metabolites that were associated in the discovery phase

The figure demonstrates the QQ plots for the following salivary metabolites:

(i) 4-guanidinobutanoate

(ii) beta-guanidinopropanoate

(iii) creatinine

(iv) urate

(v) allantoin

(vi) dimethylglycine

(vii) 3-ureidopropionate

(viii) 3-ureidoisobutyrate

(ix) N-acetylglucosamine/N-acetylgalactosamine

(x) glycosyl-N-stearoyl-sphinganine

(xi) 1-(1-enyl-palmitoyl)-2-arachidonoyl-GPC

(xii) ethylmalonate

(xiii) gamma-carboxyglutamte

(xiv) ribonate

Supplementary figure 4: A comparison of the effect of the index variant (from the salivary mGWAS) using measurements of the corresponding metabolite in faecal, serum and saliva samples of the TwinsUK subjects

The plots demonstrate a comparison of the effect of the index variant (from the salivary mGWAS) for the following significantly associated salivary metabolites that were measured in all three biospecimens (faeces, serum and saliva):

(i) 4-guanidinobutanoate

(ii) creatinine

(iii) urate

(iv) dimethylglycine

(v) 3-ureidopropionate

(vi) ethylmalonate

(vii) ribonate

Supplementary figure 5: Regional association plots for the genetic loci that were significantly associated with a salivary metabolite in the discovery phase

Regional association plots were created using the LocusZoom tool for the following locus-metabolite associations (mQTLs) that were identified in the discovery phase:

(i) *AGMAT* and 4-guanidinobutanoate

(ii) *AGMAT* and beta-guanidinopropanoate

(iii) *ATP13A5* and creatinine

(iv) *SLC2A9* and urate

(v) *SLC2A9* and allantoin

(vi) *DMGDH* and dimethylglycine

(vii) *DPYS* and 3-ureidopropinoate

(viii) *DPYS* and 3-ureidoisobutyrate

(ix) *ABO* and N-acetylglucosamine/N-acetylgalactosamine

(x) *UGCG* and glycosyl-N-stearoyl-sphinganine

(xi) *FADS2* and 1-(1-enyl-palmitoyl)-2-arachidonoyl-GPC

(xii) *ACADS* and ethylmalonate

(xiii) *TYMS/ENOSF1* and ribonate

*The regional association plot for the *MGP* locus has not been included since LD information for variants at that locus was not well characterised in the 1000 Genomes dataset

Supplementary figure 6: Boxplots demonstrating the distribution of the associated salivary metabolites, stratified by the genotypes of the corresponding index variant

The 14 salivary metabolites that were identified in the discovery phase of our study mapped to 11 distinct genetic loci: of which, three loci (*AGMAT*, *SLC2A9* and *DPYS*) were associated with two metabolites each; and the remaining eight loci were each associated with one metabolite. The plots demonstrate the distribution of each associated salivary metabolite with respect to the genotypes of the corresponding index variant (for each index variant, the number of observations for each genotype is provided).

Tables

Locus	Index variant	Chr	Position (build 37 bp)	Biochemical (Metabolite)	EA	EAF	Beta	SE	P-value
AGMAT	rs10927806	1	15,909,239	4-guanidinobutanoate	C	0.45	0.252	0.037	4.4×10^{-11}
	rs6690813	1	15,861,073	beta-guanidinopropanoate	C	0.45	0.708	0.034	4.6×10^{-66}
ATP13A5	rs55918334	3	193,086,314	creatinine	G	0.53	0.305	0.039	3.9×10^{-14}
SLC2A9	rs13129697	4	9,926,967	urate	T	0.72	0.337	0.042	1.3×10^{-14}
	*rs7675964	4	9,941,434	allantoin	C	0.73	0.239	0.043	4.1×10^{-8}
DMGDH	rs248386	5	78,330,227	dimethylglycine	C	0.81	-0.349	0.048	1.1×10^{-12}
DPYS	rs80274300	8	105,470,882	3-ureidopropionate	C	0.81	-0.764	0.043	8.1×10^{-54}
	rs80274300	8	105,470,882	3-ureidoisobutyrate	C	0.81	-0.482	0.048	5.8×10^{-22}
ABO	rs201298979	9	136,145,424	N-acetylglucosamine/ N-acetylgalactosamine	AC	0.77	-0.314	0.046	2.6×10^{-11}
UGCG	rs10981098	9	114,591,623	glycosyl-N-stearoyl-sphinganine	T	0.79	-0.308	0.046	4.6×10^{-11}
FADS2	rs174564	11	61,588,305	1-(1-enyl-palmitoyl)-2-arachidonoyl-GPC	A	0.67	0.324	0.039	1.9×10^{-15}
ACADS	rs34673751	12	121,171,891	ethylmalonate	G	0.74	-0.539	0.041	5.1×10^{-34}
MGP	rs5796614	12	15,053,094	gamma-carboxyglutamate	G	0.6	-0.29	0.038	1.9×10^{-13}
TYMS/ ENOSF1	rs2790	18	673,086	ribonate	A	0.81	0.325	0.049	5.8×10^{-11}

Table 1: Summary of genetic loci that were significantly associated with salivary metabolite(s) in the discovery phase (TwinsUK)

(Chr = Chromosome; EA = Effect Allele; EAF = Effect Allele Frequency; SE = Standard Error)

For the primary stage of association analysis in the discovery phase (TwinsUK), locus-metabolite associations (mQTLs) were identified using a genome-wide and metabolome-wide significance threshold of $p<10^{-10}$.

*This mQTL was identified when the analysis was restricted to only the loci that were identified in the primary stage of association testing (significance threshold of $p<10^{-6}$)

Locus	Chr	Index variant	Biochemical (Metabolite)	EA	Beta	SE	P-value
AGMAT	1	rs10927806	4-guanidinobutanoate	C	0.115	0.044	8.9×10^{-3}
ATP13A5	3	rs55918334	creatinine	G	0.239	0.04	2.8×10^{-9}
SLC2A9	4	rs13129697	urate	T	0.576	0.045	7.7×10^{-35}
	4	rs7675964	allantoin	C	0.209	0.049	2.3×10^{-5}
DMGDH	5	rs248386	dimethylglycine	C	-0.336	0.055	1.4×10^{-9}
DPYS	8	rs80274300	3-ureidopropionate	C	-1.035	0.049	3.1×10^{-82}
ABO[¶]	9	rs9411378	N-acetylglucosamine/ N-acetylgalactosamine	A	0.032	0.048	0.5
ACADS	12	rs34673751	ethylmalonate	G	-0.132	0.049	7.3×10^{-3}
TYMS/ ENOSF1	18	rs2790	ribonate	A	0.137	0.055	0.013

Table 2: Summary of the discovery phase associations that were tested in the replication study (SHIP-2)

(Chr = Chromosome; EA = Effect Allele; SE = Standard Error)

[¶]Since information for the index variant (rs201298979) at the *ABO* locus was not available in the replication study, a proxy variant (rs9411378) that was in high LD (r^2=0.95) with the index variant was used for the replication analysis.

Replication in SHIP-2 could be attempted for 9 of the 14 mQTLs that were identified in the discovery phase (metabolites corresponding to the remaining five mQTLs were not measured in SHIP-2). For each of these nine mQTLs, the most significant variant-metabolite pair identified in the discovery phase was tested in SHIP-2 using salivary metabolite data that was not normalised for osmolality (since SHIP-2 saliva samples represented stimulated saliva).

Locus	Biochemical (Metabolite)	Metabolite super-pathway[1]	Metabolite sub-pathway[1]	Known GWAS disease associations for the locus[2]	Known OMIM phenotype for the locus[3]	Clinical significance of the locus, the metabolite or the biochemical pathway involved
AGMAT	4-guanidinobutanoate	Amino acid	Guanidino and Acetamido metabolism	Glomerular filtration rate, alcoholic chronic pancreatitis	-	Putrescine, a compound that is generated in the pathway which involves the *AGMAT* gene i.e. polyamine synthesis pathway, has been implicated in bad breath / oral health[27,28]
	beta-guanidinopropanoate	Xenobiotics	Plant (food) component			
ATP13A5	creatinine	Amino acid	Creatine metabolism	-	Kufor-Rakeb syndrome	The *ATP13A5* gene encodes the family of proteins that regulate the activity of HMG-CoA reductase[33], the rate-limiting enzyme in cholesterol synthesis
SLC2A9	urate	Nucleotide	Hypo(Xanthine) / Inosine (purine) metabolism	Gout	Renal hypouricemia type 2	The *SLC2A9* gene encodes a carrier protein that is involved in urate transport in the proximal convoluted tubules of kidneys
	allantoin	Nucleotide	Hypo(Xanthine) / Inosine (purine) metabolism			
DMGDH	dimethylglycine	Amino acid	Glycine, Serine and Threonine metabolism	-	Dimethylglycine dehydrogenase deficiency	The *DMGDH* gene encodes an enzyme (dimethylglycine dehydrogenase), which catalyses the conversion of dimethylglycine to sarcosine
DPYS	3-ureidopropionate	Nucleotide	Uracil (pyrimidine) metabolism	Paget's disease	Dihydropyrimidinuria	Variants in the *DPYS* gene, which encodes an enzyme in the pyrimidine degradation pathway, have been associated with 5-fluorouracil toxicity[35]
	3-ureidoisobutyrate	Nucleotide	Uracil (pyrimidine) metabolism			
ABO	N-acetylglucosamine/ N-acetylgalactosamine	Carbohydrate	Aminosugar metabolism	Gastric carcinoma, stroke, venous thromboembolism, ovarian cancer, malaria, blood cholesterol level, type 2 diabetes	-	The *ABO* gene encodes a protein with glycosyltransferase activity, which forms the basis of the ABO blood group system
UGCG	glycosyl-N-stearoyl-sphinganine	Lipid	Ceramide metabolism	-	-	The *UGCG* gene is involved in glycosphingolipid synthesis, defects in which are known to cause Gaucher's disease[48], a lysosomal storage disorder
FADS2	1-(1-enyl-palmitoyl)-2-arachidonoyl-GPC	Lipid	Plasmalogen	ADHD, blood cholesterol level	-	The *FADS2* gene encodes an enzyme (fatty acid desaturase) which catalyses the rate-limiting step in the desaturation of polyunsaturated fatty acids (PUFA)
ACADS	ethylmalonate	Amino acid	Leucine, Isoleucine and Valine metabolism	-	Ethylmalonic aciduria	The *ACADS* gene encodes the acyl-CoA dehydrogenase enzyme, which catalyses the initial step in the fatty acid of oxidation pathway
MGP	gamma-carboxyglutamate	Amino acid	Glutamate metabolism	-	Keutel syndrome	The *MGP* gene, a regulator of physiologic tissue calcification, is associated with calcification of vasculature in patients with cardiovascular disease[45]. The *MGP* gene has also been associated with natural tooth loss in elderly women[49]
TYMS/ ENOSF1	ribonate	Carbohydrate	Pentose metabolism	Hypertension	-	The *TYMS* gene encodes the enzyme thymidylate synthase, the main site of action of 5-fluorouracil. The *ENOSF1* gene encodes the antisense RNA of *TYMS*

Table 3: Annotation of the genetic loci and the corresponding metabolites that were significantly associated in the discovery phase

[1]obtained from the KEGG database

[2]accessed from the NHGRI GWAS catalogue

[3]accessed from the OMIM database

The mQTLs that were identified in the discovery phase were annotated for the metabolite pathway, known disease associations for the locus from the GWAS and the OMIM databases, and the clinical significance of the locus, the metabolite or the biochemical pathway involved.

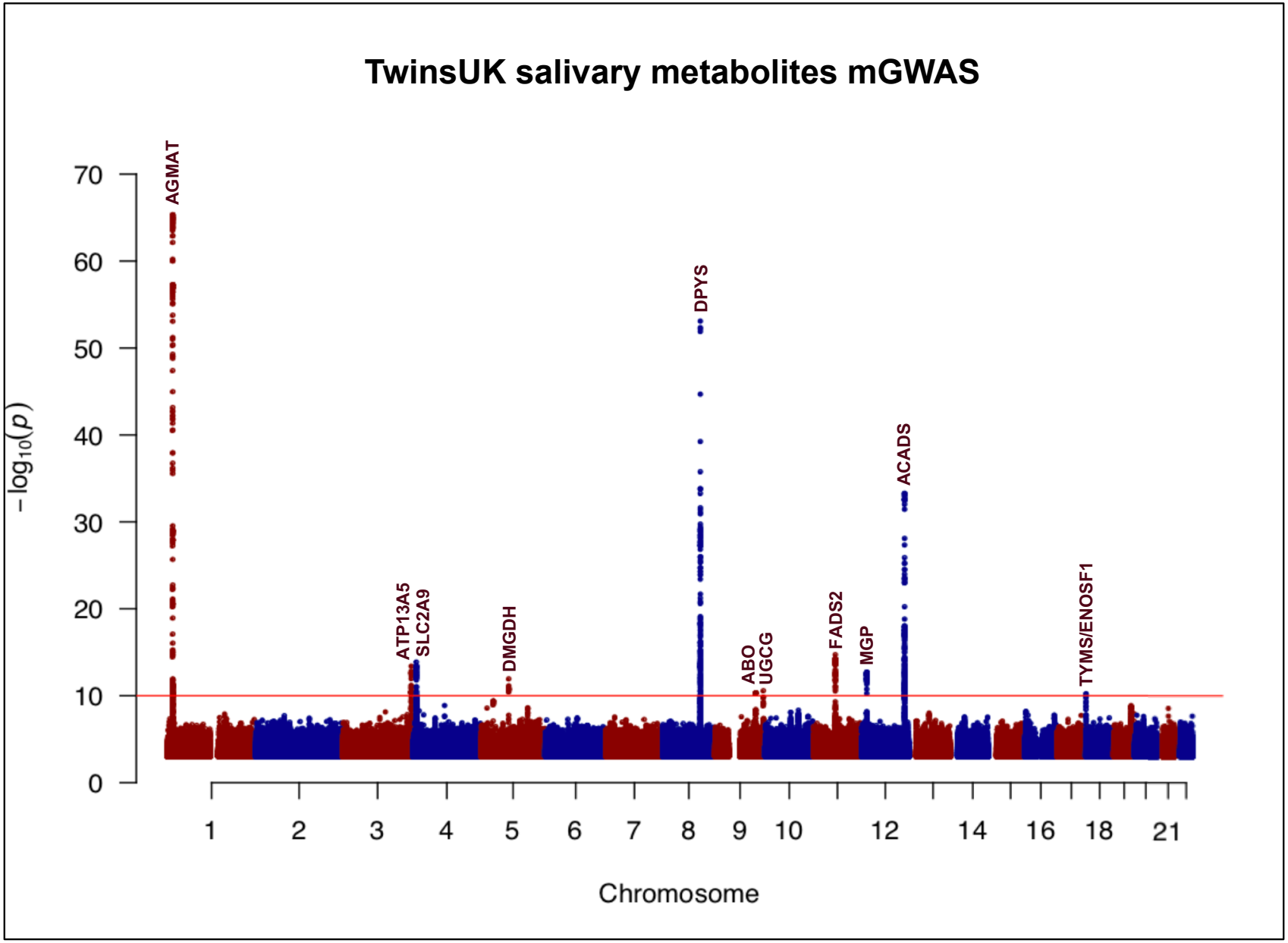
Locus	Chr	Biochemical (Metabolite)	Phenotypic association(s) tested
AGMAT	1	4-guanidinobutanoate	1. Periodontal disease 2. Clinical depression or anxiety disorder
		beta-guanidinopropanoate	
ATP13A5	3	creatinine	1. eGFR (measure of renal function) 2. Grip strength (measure of muscle strength) 3. Statin usage
DPYS	8	3-ureidopropionate	1. Irritable bowel syndrome (based on the ROME-III criteria[36])
		3-ureidoisobutyrate	

Table 4: Summary of the phenotypes that were tested in relation with specific salivary metabolites

(Chr = Chromosome; eGFR = effective Glomerular Filtration Rate)

The metabolites that were uniquely associated in saliva i.e. the ones for which a genetic association had not been previously reported in blood, were tested with phenotypes (diseases / traits / adverse drug effects) relating to the metabolite or its associated biochemical pathway, using information from the TwinsUK database.

Figure 1


TwinsUK salivary metabolites mGWAS
AGMAT
ATP13A5
SLC2A9
DMGDH
DPYS
ABO
UGCG
FADS2
MGP
ACADS
TYMS/ENOSF1
70
60
50
40
30
20
10
0
$-\log_{10}(p)$
1
2
3
4
5
6
7
8
9
10
12
14
16
18
21
Chromosome