nature portfolio

https://doi.org/10.1038/s41588-021-00986-w

Supplementary information

The *UGT2A1/UGT2A2* locus is associated with COVID-19-related loss of smell or taste

In the format provided by the authors and unedited

Supplementary Text

Conditional Associations

In the main association presented in Figure 1b, we noted the presence of a number of genome-wide significant SNPs that appear to be low linkage disequilibrium ($r^2 < 0.2$) with the index SNP. This appears suggestive of the presence of an independent association within the same locus. To test this hypothesis, we conducted a conditional analysis within the European population by retesting for association within the locus, but having included the index SNP (rs7688383) within the regression model.

The results of this analysis are shown in the figure below (Supplementary Figure 2). On the left hand side, we show the European association signal prior to conditioning. On the right hand side, we show the association signal having included rs7688383 in the model. Including rs7688383 within the model eliminates the genetic association within the region, with no SNP having a p-value < 1e-4 in the conditioned model. From this we conclude that there is little evidence supporting the presence of a secondary association within the locus.

eQTL and colocalization analyses

In order to investigate if eQTLs colocalized with the GWAS association, we performed a targeted eQTL search in a window 500kb up- and downstream of rs7688383 for eQTLs for *UGT2A1*, *UGT2A2*, *UGT2B4* and *SULT1B1* in 5 eQTL datasets; GTEx 8 [1], Parkinson's Progression Marking Initiative (PPMI) [2], Database of Immune Cell Expression (DICE) [3], GEUVADIS [4] and Depression Genes Network (DGN) [5] datasets. This resulted in 93 tissue-by-gene combinations tested at 5,946 genetic variants, resulting in 447,035 statistical tests. Using a conservative p-value threshold of 5×10^{-8}, 20 tissue-gene pairs exhibited at least one significant eQTL association. Of the detected eQTLs, those of UGT2A1 are physically proximal to the GWAS association, but the GWAS index SNP has moderately low LD with the eQTL index SNPs (Supplementary Figure 3). For all tissue-gene pairs where we only observed a significant effect, we performed a colocalization analysis [6] between a given eQTL signal and our GWAS signal, using default prior probabilities. In all cases, we observed the posterior probability of the GWAS and eQTL associations sharing a common causal variant (“H4”) to be less than 10%.

Preliminary inspection of the eQTL associations suggested the presence of independent associations in some instances. We therefore wanted to check if any of the conditionally independent signals colocalized with the GWAS association. To do so, we conducted a conditional stepdown analysis for each tissue-gene pair with a significant eQTL, conditioning on the variant with the smallest p-value within each step. For the 20 tissue-gene pairs with a detected eQTL, we observed a total of 6 conditionally distinct associations for which the most significant variant passed a nominal p-value threshold of 10^{-6}. For each of these performed a conditional-leave-each-out (CLEO) analysis using both the unconditional and conditional

variants. Again, we performed colocalization analysis with the resulting signals. In all cases, we observed the posterior probability of the GWAS and eQTL associations sharing a common causal variant (“H4”) to be less than 10%.

Supplementary Table 1: Logistic regression model of age, sex, and ancestry, predicting loss of smell or taste among those with a SARS-CoV-2 positive test.

	Model results		
	Odds Ratio	95% confidence interval	p-value
European	1.0 (REF)	NA	NA
Latino	1.05	1.003-1.09	0.033
African American or Black	0.88	0.81-0.96	0.003
East Asian	0.80	0.70-0.92	0.002
South Asian	0.84	0.70-1.03	0.09
Other	1.01	0.93-1.11	0.79
Female sex	1.56	1.51-1.61	4.10E-155
Age < 25	0.99	0.94-1.06	0.97
26-35	1.14	1.09-1.20	5.50E-08
36-45	1.0 (REF)	NA	NA
46-55	0.87	0.83-0.92	2.50E-07
56-65	0.69	0.65-0.73	1.30E-39
66-75	0.47	0.43-0.50	2.50E-97
75 and older	0.37	0.33-0.42	9.00E-62

Supplementary Table 2: Genome-wide association statistics (odds ratios and p-values) for rs7688383 for each ancestry, including allele frequencies and sample sizes for each GWAS. Note that each GWAS sample only includes individuals with complete data at the time the study was run, and excludes related individuals.

Population	Case / Control N	Alleles (protective / risk)	Risk allele frequency	Imputation r^2	Odds ratio	95% CI	p-value
Europe	27,786 / 13,903	C/T	36.8%	0.989	1.11	[1.08, 1.15]	1.78E-11
South Asian	230 / 143	C/T	32.2%	0.964	1.26	[0.88, 1.82]	0.206
African American	1,556 / 804	C/T	31.4%	0.958	1.13	[0.98, 1.29]	0.0895
Latino	7,841 / 3,419	C/T	30.1%	0.984	1.35	[1.05, 1.19]	0.0007
East Asian	557 / 314	C/T	19.2%	0.957	1.18	[0.91, 1.53]	0.213
Meta-analysis	**37,790 / 18,583**	**C/T**			**1.12**	**[1.09, 1.15]**	**4.17E-15**

Supplementary Table 3: The 99% credible set for the UGT2A1/2 association locus, as estimated from the trans-ethnic meta-analysis. The credible set contains 28 variants, covering 44.594kb of chromosome 4 from 69,571,521 bp to 69,616,115 bp.

rsID	Chr4 position	Alleles	p-value	OR	95% CI	Posterior probability
rs7688383	69615987	C/T	4.20E-15	1.115	[1.085,1.146]	0.97825
rs62306197	69616115	C/T	4.70E-12	0.884	[0.854,0.916]	0.00123
rs750402645	69572977	D/I	6.00E-12	0.878	[0.846,0.911]	0.00101
rs11249451	69609750	A/C	8.80E-12	1.128	[1.090,1.168]	0.00066
rs7675531	69571733	A/G	9.80E-12	0.887	[0.857,0.918]	0.00059
rs7655867	69612465	C/T	1.00E-11	0.886	[0.856,0.918]	0.00057
rs4148303	69595443	C/T	8.30E-12	1.097	[1.068,1.127]	0.00055
rs12644575	69613693	C/T	1.10E-11	1.128	[1.089,1.168]	0.00054
rs10033801	69599908	A/G	9.00E-12	0.912	[0.888,0.936]	0.00051
rs1319811	69597175	A/C	9.20E-12	0.912	[0.888,0.936]	0.00050
rs2010207	69597085	G/T	9.40E-12	0.912	[0.888,0.936]	0.00049
rs1432336	69610877	C/T	1.20E-11	0.887	[0.857,0.918]	0.00049
rs4148297	69599723	C/T	9.50E-12	0.912	[0.888,0.936]	0.00048
rs72637448	69578030	A/G	1.40E-11	0.887	[0.857,0.918]	0.00043

rs62306196	69615039	A/C	1.60E-11	1.127	[1.089,1.167]	0.00036
rs11721934	69600568	A/G	2.00E-11	1.126	[1.088,1.166]	0.00030
rs7696606	69571747	A/G	2.00E-11	1.126	[1.088,1.166]	0.00029
rs7696068	69571521	A/G	2.10E-11	1.128	[1.089,1.168]	0.00028
rs11730451	69600658	C/T	2.10E-11	0.888	[0.858,0.919]	0.00028
rs11729380	69574313	C/T	2.20E-11	1.126	[1.088,1.166]	0.00027
rs10004047	69579782	A/G	1.80E-11	0.913	[0.889,0.938]	0.00026
rs72637470	69613194	C/T	2.30E-11	1.127	[1.088,1.168]	0.00026
rs17616815	69572555	C/T	2.40E-11	0.888	[0.858,0.920]	0.00025
rs17616887	69572818	A/T	2.40E-11	1.125	[1.087,1.165]	0.00025
rs17616972	69573045	C/T	2.50E-11	1.125	[1.087,1.165]	0.00023
rs17147413	69572428	G/T	2.60E-11	0.889	[0.859,0.920]	0.00023
rs72637445	69572578	A/G	2.60E-11	0.889	[0.859,0.920]	0.00023
rs1432318:C	69582167	C/G	2.00E-11	0.913	[0.889,0.938]	0.00023

Supplementary Table 4: Phenome-wide Association Study (PheWAS) of rs7688383. The table below shows four phenotypes associated with rs7688383 in the 23andMe database at a significance threshold of $p < 1e\text{-}6$.

Phenotype name	Description	p-value	OR / effect	95% CI
covid_symptoms_loss_of_smell_or_taste	Cases reported a SARS-CoV-2 positive test and loss of smell or taste. Controls reported a SARS-CoV-2 positive test but reported no loss of smell or taste.	4.20e-15	1.115 (OR)	[1.085, 1.146]
ice_vanilla	Preferred ice cream flavor = Vanilla	1.47e-14	1.019 (OR)	[1.014, 1.024]
iqb.ability_to_smell	How would you rate your current ability to smell? (very_poor/poor/fair/good/very_good)	3.05e-11	0.0075 (effect)	[0.0053, 0.0098]
ever_tobacco_user	Ever tobacco user: either regular use or at least 100 cigarettes.	9.74e-10	0.990 (OR)	[0.987, 0.993]
ha_smell_rating_cc	Cases report moderate or worse sense of smell, controls report they smell everything.	1.24e-09	0.966 (OR)	[0.955, 0.977]

Supplementary References

[1] The GTEx Consortium atlas of genetic regulatory effects across human tissues. The GTEx Consortium. Science. 369 (1318-1330), 10 Sep 2020.
[2] The Parkinson Progression Marker Initiative (PPMI). Parkinson Progression Marker Initiative. Prog. Neurobiol. 2011 Dec;95(4):629-35.
[3] Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. Benjamin J. Schmiedel, Divya Singh, Ariel Madrigal, Alan G. Valdovino-Gonzalez, Brandie M. White, Jose Zapardiel-Gonzalo, Brendan Ha, Gokmen Altay, Jason A. Greenbaum, Graham McVicker, Grégory Seumois, Anjana Rao, Mitchell Kronenberg, Bjoern Peters, Pandurangan Vijayanand. Cell 2018;175(6):1701-1715.e16
[4] Lappalainen, T., Sammeth, M., Friedländer, M. R., 't Hoen, P. A., Monlong, J., Rivas, M. A., Gonzàlez-Porta, M., Kurbatova, N., Griebel, T., Ferreira, P. G., Barann, M., Wieland, T., Greger, L., van Iterson, M., Almlöf, J., Ribeca, P., Pulyakhina, I., Esser, D., Giger, T., Tikhonov, A., … Dermitzakis, E. T. (2013). Transcriptome and genome sequencing uncovers functional variation in humans. Nature, 501(7468), 506–511.
[5] Battle, A., Mostafavi, S., Zhu, X., Potash, J. B., Weissman, M. M., McCormick, C., Haudenschild, C. D., Beckman, K. B., Shi, J., Mei, R., Urban, A. E., Montgomery, S. B., Levinson, D. F., & Koller, D. (2014). Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. Genome research, 24(1), 14–24.
[6] Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, Plagnol V. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. 2014 May 15;10(5)