



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Genome-Wide Meta-analysis identifies three novel loci associated with stroke

Citation for published version:

MEGASTROKE consortium and the International Stroke Genetics Consortium (ISGC) & Sudlow, C 2018, 'Genome-Wide Meta-analysis identifies three novel loci associated with stroke', *Annals of Neurology*. <https://doi.org/10.1002/ana.25369>

Digital Object Identifier (DOI):

[10.1002/ana.25369](https://doi.org/10.1002/ana.25369)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Annals of Neurology

Publisher Rights Statement:

This is the author's peer-reviewed manuscript as accepted for publication.

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Genome-Wide Meta-analysis identifies three novel loci associated with stroke

Rainer Malik, PhD¹, Kristiina Rannikmäe, MD^{2,3}, Matthew Traylor, PhD⁴, Marios K Georgakis, MD¹, Muralidharan Sargurupremraj, PhD^{5,6}, Hugh S Markus, DM⁴, Jemma C Hopewell, PhD⁷, Stephanie Debette, MD, PhD^{5,6}, Cathie LM Sudlow, DPhil^{2,3}, and Martin Dichgans, MD^{1,8,9} for the MEGASTROKE consortium and the International Stroke Genetics Consortium (ISGC)

Affiliations

¹ Institute for Stroke and Dementia Research (ISD), University Hospital, LMU Munich, Munich, Germany

² Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK

³ Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK

⁴ Stroke Research Group, Division of Clinical Neurosciences, University of Cambridge, Cambridge, UK

⁵ INSERM U1219 Bordeaux Population Health Research Center, University of Bordeaux, France

⁶ Department of Neurology, Institute for Neurodegenerative Disease, Bordeaux University Hospital, Bordeaux, France

⁷ Clinical Trial Service Unit and Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, Oxford, UK

⁸ Munich Cluster for Systems Neurology (SyNergy), Munich, Germany

⁹ German Center for Neurodegenerative Diseases (DZNE), Munich, Germany

Corresponding author:

Martin Dichgans, MD

Institute for Stroke and Dementia Research (ISD), University Hospital, LMU Munich, Munich, Germany

Munich Cluster for Systems Neurology (SyNergy), Munich, Germany

German Center for Neurodegenerative Diseases (DZNE), Munich, Germany

Feodor-Lynen-Str. 17, 81377 Munich, Germany

T: +49-89-4400-46018

E: martin.dichgans@med.uni-muenchen.de

Running head: MEGASTROKE and UK Biobank GWAS

Number of characters in title: 121

Number of characters in running head: 30

Word count: 1,583

Figures: 2, Tables: 1

Abbreviations: UKB, UK Biobank; AS, Any Stroke; AIS, Any Ischemic Stroke; GWAS, Genome-wide association study; MR, Mendelian Randomization; eQTL, expression quantitative trait locus

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/ana.25369

ABSTRACT

We conducted a European-only and trans-ancestral genome-wide association meta-analysis in 72,147 stroke patients and 823,869 controls using data from UK Biobank (UKB) and the MEGASTROKE consortium. We identified an exonic polymorphism in *NOS3* (rs1799983, p.Glu298Asp; $p=2.2E-8$; $OR=1.05[1.04-1.07]$) and variants in an intron of *COL4A1* (rs9521634, $p\text{-value}=3.8E-8$, $OR=1.04[1.03-1.06]$) and near *DYRK1A* (rs720470, $p=6.1E-9$; $OR=1.05[1.03-1.07]$) at genome-wide significance for stroke. Effect sizes of known stroke loci were highly correlated between UKB and MEGASTROKE. Using Mendelian Randomization we further show that genetic variation in the nitric oxide synthase (NOS) – nitric oxide (NO) pathway in part affects stroke risk via variation in blood pressure.

INTRODUCTION

Stroke is the leading cause of disability and the second most common cause of death worldwide. The identification of common genetic variants for vascular conditions has provided mechanistic insights, improved options for risk prediction, and facilitated the development of novel therapeutics.¹ The MEGASTROKE consortium recently reported on the largest genome-wide association study (GWAS) to date in >520,000 subjects from multiple ethnicities.² Aside from finding novel risk loci for any stroke (AS), any ischemic stroke (AIS) and ischemic stroke subtypes, this study demonstrated marked genetic overlap with related vascular traits. Still, much of the heritability of stroke remains unexplained and the biological mechanisms and pathways underlying shared genetic influences with related traits are largely elusive.

The UK Biobank (UKB) was established to improve understanding of common diseases including stroke. Participants were recruited from the general adult population and, in addition to having provided self-reported medical history at recruitment, are followed prospectively, chiefly through linkage to their National Health Service records.³(<http://www.ukbiobank.ac.uk>) UKB recently released genotypes on over 500,000 participants thus adding to available GWAS data.

The current study aimed to identify additional susceptibility loci for stroke and obtain further insights into relevant pathways by meta-analyzing GWAS summary statistics from UKB with data from the MEGASTROKE European-only stratum followed by a trans-ancestral analysis.

SUBJECTS AND METHODS

The protocol for this study received prior approval by all IRBs, and informed consent was obtained from each subject.

UK Biobank

For definition of stroke cases we used UKB fields 42007 and 42009, the algorithmically defined stroke outcome, including both prevalent strokes (stroke prior to recruitment) and incident strokes (first stroke diagnosed during follow-up) (http://biobank.ctsu.ox.ac.uk/crystal/docs/alg_outcome_stroke.pdf). Stroke events that were self-reported only without corroborating evidence from medical records were excluded due to substantial uncertainty about the accuracy of stroke self-report.⁴ Coded hospital admissions and death record data (ICD 9 and 10 coding systems) were included based on previous work showing good accuracy of these data sources for identifying stroke cases.⁵ Participants without a stroke diagnosis were included as controls. Related participants and those of non-White-British descent were excluded, as were SNPs with MAF <0.01. The final UKB dataset consisted of 4,985 AS cases, 3,628 AIS cases and 369,419 controls. The genetics dataset of UKB is described at <https://www.biorxiv.org/content/early/2017/07/20/166298>. We fitted a logistic regression model with stroke as the outcome and each SNP as a dependent variable including age, sex, principal components 1-10, and genotyping chip as covariates.

MEGASTROKE

We used the full summary statistics from MEGASTROKE after filtering as recently described. Analyses included 67,162 AS cases, 60,341 AIS cases, 6,688 large artery stroke (LAS) cases, 9,006 cardioembolic stroke (CES) cases and 11,710 small vessel stroke (SVS) cases.²

Genome-wide association meta-analyses

We first performed a fixed-effects meta-analysis for AS and AIS using summary statistics from UKB and the European stratum from MEGASTROKE. The newly formed European stratum was then meta-analyzed with the East Asian, South Asian, African American, other Asian, and Latin American strata from MEGASTROKE using a fixed-effects model. The final dataset consisted of 72,147 AS cases (45,570 European), 63,969 AIS cases (37,845 European) and 823,869 controls (775,530 European). Single genomic control was applied for all analyses. We report on results from both the new European-only meta-analysis and the trans-ancestral meta-analysis. Genome-wide significance was set at $p < 5 \times 10^{-8}$.

Mendelian Randomization analysis

To evaluate the causal association of recently published variants in the NOS3 pathway⁶ and stroke risk, we performed a two-sample Mendelian Randomisation (MR) analysis using blood pressure (BP) data from UKB as an exposure variable (systolic BP [SBP] and diastolic BP [DBP]: 318,417 subjects analyzed using a BOLT LMM model) and stroke data from the combined European-only analysis as outcome. We used the R package “mendelianRandomization” and report on results obtained from the weighted median, inverse variance weighted, MR-EGGER, and MBE (modal) method.

Expression quantitative trait loci (eQTLs)

For lead variants at the novel loci and SNPs in LD ($r^2 > 0.8$) we queried publicly available eQTL databases GTEx⁷ and GRASP2.⁸

RESULTS

We first compared effect sizes and directions of previously published lead SNPs from MEGASTROKE between UKB and the European MEGASTROKE stratum. Two low-frequency variants at *RGS7* and *TMF4SF4* were not available in UKB leaving 30 loci for analysis. We observed significant positive correlations in the effect sizes for both AS ($r=0.56$ [0.26-0.76], $p=0.0011$) and AIS ($r=0.62$ [0.34-0.80], $p=0.00021$) between the two datasets, thus supporting our approach of meta-analyzing them.

Fixed-effects meta-analysis of the UKB and MEGASTROKE datasets (up to 72,147 cases and 823,869 controls) revealed 3 novel loci reaching genome-wide significance ($p < 5E-8$) for association with AS in the European-only or the full trans-ancestral analysis (**Table 1** and **Figure 1A**). Genomic inflation was estimated to be 1.05 both for AS and AIS (Europeans-only analysis). For the transethnic analyses, we found lambda values of 1.04 and 1.07, respectively. The explained phenotypic variance of the three novel loci is estimated to be 0.06% for AS in Europeans assuming a disease prevalence of 0.055.

The first finding is an exonic SNP in the Nitric Oxide Synthase 3 (*NOS3*) gene (rs1799983, p.Glu298Asp) that was associated with AS in the European-only analysis ($p=2.2E-8$; OR=1.05 [1.04-1.07]; **Table 1** and **Figure 1B**). Functional variants in *NOS3* and other genes of the NOS-NO pathway have been associated with hypertension.⁶ Hence, we explored the possibility that the effects of genetically determined dysregulation of the NOS-NO pathway on stroke risk are mediated via BP using a two-sample MR analysis. We

selected instruments in the genes encoding proteins in the NOS-NO pathway that exert known functional effects on NO signaling, as summarized in a recent review⁶. The genetic instruments included variants in NOS3 (rs1799983, rs2070744, rs3918226), Guanylate Cyclase 1, Soluble Alpha 3 *GUCY1A3* (rs7692387), and the L-arginine transporter gene *SLC7A1* (rs41318021)⁶; rs7539120 in the Nitric oxide synthase 1 adaptor protein gene *NOS1AP* was not available for analysis in UKB. Considering BP in UKB as the exposure and AS as outcome we found significant associations in the weighted median approach (OR=1.11, $p<0.0001$ for SBP; OR=1.14, $p<0.0001$ for DBP) and IVW (OR=1.15, $p=0.005$ for SBP; OR=1.19, $p=0.002$ for DBP), suggesting a causal association of the instruments with stroke via BP (**Figure 2**) with an estimated contribution⁹ of 4%. We found no evidence of heterogeneity in the IVW analyses and the intercept in the MR-Egger analyses were not significant (all $p>0.10$) suggesting absence of significant pleiotropy. After excluding two potentially pleiotropic SNPs identified through PhenoScanner as being associated with CAD (rs3918226, rs7692387), the results remained significant.

We further found variants in *COL4A1* to be associated with AS (rs9521634, trans-ancestral analysis: $p=3.8E-8$; OR=1.04 [1.03-1.06], **Table 1** and **Figure 1C**). rs9521634 is situated in intron 28 but does not act as a known eQTL for any gene in available tissues as is also true for SNPs in linkage disequilibrium (LD, $r^2>0.8$) with rs9521634.

The third locus resides in the intergenic region between *DYRK1A* and *KCNJ6* (rs720470, $p=6.1E-9$; OR=1.05 [1.03-1.07], **Table 1** and **Figure 1D**). rs720470 serves as an eQTL for *DYRK1A* in whole blood ($p=9E-4$)¹⁰ thus identifying *DYRK1A* as a plausible causal gene at this locus.

Association results for ischemic stroke subtypes were available from MEGASTROKE. rs1799983 in *NOS3* showed the strongest association with CES ($p=4.6E-5$, OR=1.09 [1.04-1.13]; $p>0.1$ for both SVS and LAS) whereas rs9521634 in *COL4A1* showed the strongest association with SVS ($p=0.0004$, OR=1.06[1.03-1.10]; $p>0.3$ for CES and LAS). There was no clear association signal for ischemic stroke subtypes at *DYRK1A* (all $p>0.01$).

DISCUSSION

We found rs1799983 encoding p.Glu298Asp in endothelial NOS to reach genome-wide significance for association with AS. Nitric oxide signaling is a key regulator of vascular tone, BP, and platelet aggregation. p.Glu298Asp lowers eNOS activity by disruption of eNOS caveolar localization¹¹. Another variant, rs3918226, that is situated in the *NOS3*

promoter and was found to lower promoter activity¹², has been shown to associate with both hypertension¹² and coronary artery disease.¹³ rs3918226 is in low LD ($r^2=0.17$) with our lead SNP and did not reach genome-wide significance for association with AS or AIS. Still, our MR analysis suggests that the aggregate effects of common variants in the NOS-NO pathway on stroke risk are in part mediated through blood pressure. Similar results have recently been shown for coronary artery disease using a genetic risk score comprised of two common variants in *NOS3* and *GUCY1A3*.¹⁴ Somewhat surprisingly, we found the strongest association with CES. While this might relate to limited power in stroke subtypes, potential mechanisms underlying this association might include prothrombotic effects as well as mechanism that are yet unknown.

Our findings further highlight a role of *COL4A1* in stroke. Collagen type IV $\alpha 1$ is a major constituent of the vascular basement membrane and forms heterotrimers with collagen IV $\alpha 2$. Rare variants in *COL4A1* cause monogenic small vessel disease with hemorrhagic and ischemic stroke being part of the spectrum.^{15, 16} These mutations are associated with structural protein changes or altered expression levels of COL4A1, which interfere with the assembly, secretion, or biological function of COL4A1. While we found no eQTLs for variants in LD with our lead SNPs in GTEX7 and GRASP2 there might be eQTLs in relevant tissues or cell types not captured by these sources. In keeping with the role of *COL4A1* in small vessel disease we found rs9521634 to show the strongest association signal in SVS. Interestingly, common variants in the adjacent *COL4A2* gene associate with SVS^{2, 17}, intracerebral hemorrhage¹⁷, and white matter hyperintensities.¹⁸ Collectively, these findings define COL4A1 and COL4A2 as key molecules in the biology of stroke and small vessel disease.

Our association results in combination with the eQTL data further point to a potential role of *DYRK1A* in stroke. *DYRK1A* encodes a dual-specificity tyrosine-phosphorylation-regulated kinase 1A that has recently been shown to regulate angiogenic responses in vascular endothelial cells.¹⁹ *Dyrk1a* heterozygous mice exhibit defects in retinal vascularization and DYRK1A was found to positively regulate VEGF-dependent transcriptional responses in endothelial cells.¹⁹ We found no association signal with specific ischemic stroke subtypes possibly related to limited power. *DYRK1A* maps to the Down syndrome (DS) critical genetic region and is thought to contribute to the manifestations of DS. Recent work has drawn attention to an increased risk of stroke in DS.²⁰ While this might relate to other factors our findings in conjunction with the above experimental data suggest a link between *DYRK1* and stroke.

Accepted Article

As a limitation, we were not able to perform meta-analyses for LAS, CES, and SVS with UKB since information on etiological subtypes of ischemic stroke is not yet available in UKB. Also, the gain in power compared to MEGASTROKE was limited because of the relatively small number of stroke cases (around 5,000) in UKB, although numbers will increase with further follow-up.¹³ Nonetheless, we found three novel loci providing further insights into relevant stroke pathways and mechanisms. Integration of additional datasets will be key to better understanding the genetic basis of stroke.

ACKNOWLEDGMENTS

The study was conducted using the UK Biobank resource, under UKB project 2532 (“UK Biobank Stroke Study (UKBiSS)”). The MEGASTROKE project received funding from sources specified at <http://www.megastroke.org/acknowledgments.html>.

JCH acknowledges personal funding from the British Heart Foundation (FS/14/55/30806).

AUTHOR CONTRIBUTIONS

R.M., K.R., J.C.H., C.L.M.S., S.D. and M.D. were involved in the conception and the design of the study

R.M., M.T., M.G., D.G., M.S., K.R., J.C.H., H.S.M., C.L.M.S., S.D. and M.D. acquired and analysed the data

R.M. and M.D. drafted the manuscript

R.M., M.D., K.R., C.L.M.S., M.S., S.D. made critical revisions of the manuscript

POTENTIAL CONFLICTS OF INTEREST

None

REFERENCES

1. Khera AV, Kathiresan S. Genetics of coronary artery disease: discovery, biology and clinical translation. *Nature reviews Genetics*. 2017 Jun;18(6):331-44.
2. Malik R, Chauhan G, Traylor M, et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nature genetics*. 2018 Apr;50(4):524-37.
3. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine*. 2015 Mar;12(3):e1001779.
4. Woodfield R, U. K. Biobank Stroke Outcomes Group, U. K. Biobank Follow-up Outcomes Working Group, Sudlow CL. Accuracy of Patient Self-Report of Stroke: A Systematic Review from the UK Biobank Stroke Outcomes Group. *PLoS one*. 2015;10(9):e0137538.
5. Woodfield R, Grant I, U. K. Biobank Stroke Outcomes Group, U. K. Biobank Follow-Up Outcomes Working Group, Sudlow CL. Accuracy of Electronic Health Record Data for Identifying

Stroke Cases in Large-Scale Epidemiological Studies: A Systematic Review from the UK Biobank Stroke Outcomes Group. *PloS one*. 2015;10(10):e0140533.

6. Farah C, Michel LYM, Balligand JL. Nitric oxide signalling in cardiovascular health and disease. *Nature reviews Cardiology*. 2018 May;15(5):292-316.
7. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nature genetics*. 2013 Jun;45(6):580-5.
8. Eicher JD, Landowski C, Stackhouse B, et al. GRASP v2.0: an update on the Genome-Wide Repository of Associations between SNPs and phenotypes. *Nucleic acids research*. 2015 Jan;43(Database issue):D799-804.
9. Hartwig FP, Borges MC, Horta BL, Bowden J, Davey Smith G. Inflammatory Biomarkers and Risk of Schizophrenia: A 2-Sample Mendelian Randomization Study. *JAMA psychiatry*. 2017 Dec 1;74(12):1226-33.
10. Westra HJ, Peters MJ, Esko T, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nature genetics*. 2013 Oct;45(10):1238-43.
11. Joshi MS, Mineo C, Shaul PW, Bauer JA. Biochemical consequences of the NOS3 Glu298Asp variation in human endothelium: altered caveolar localization and impaired response to shear. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*. 2007 Sep;21(11):2655-63.
12. Salvi E, Kuznetsova T, Thijs L, et al. Target sequencing, cell experiments, and a population study establish endothelial nitric oxide synthase (eNOS) gene as hypertension susceptibility gene. *Hypertension*. 2013 Nov;62(5):844-52.
13. Nelson CP, Goel A, Butterworth AS, et al. Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nature genetics*. 2017 Sep;49(9):1385-91.
14. Emdin CA, Khera AV, Klarin D, et al. Phenotypic Consequences of a Genetic Predisposition to Enhanced Nitric Oxide Signaling. *Circulation*. 2018 Jan 16;137(3):222-32.
15. Jeanne M, Gould DB. Genotype-phenotype correlations in pathology caused by collagen type IV alpha 1 and 2 mutations. *Matrix biology : journal of the International Society for Matrix Biology*. 2017 Jan;57-58:29-44.
16. Verdura E, Herve D, Bergametti F, et al. Disruption of a miR-29 binding site leading to COL4A1 upregulation causes pontine autosomal dominant microangiopathy with leukoencephalopathy. *Annals of neurology*. 2016 Nov;80(5):741-53.
17. Rannikmae K, Sivakumaran V, Millar H, et al. COL4A2 is associated with lacunar ischemic stroke and deep ICH: Meta-analyses among 21,500 cases and 40,600 controls. *Neurology*. 2017 Oct 24;89(17):1829-39.
18. Traylor M, Zhang CR, Adib-Samii P, et al. Genome-wide meta-analysis of cerebral white matter hyperintensities in patients with stroke. *Neurology*. 2016 Jan 12;86(2):146-53.
19. Rozen EJ, Roewenstrunk J, Barallobre MJ, et al. DYRK1A Kinase Positively Regulates Angiogenic Responses in Endothelial Cells. *Cell reports*. 2018 May 8;23(6):1867-78.
20. Sobey CG, Judkins CP, Sundararajan V, Phan TG, Drummond GR, Srikanth VK. Risk of Major Cardiovascular Events in People with Down Syndrome. *PloS one*. 2015;10(9):e0137093.

TABLES

rsID	rs1799983	rs9521634	rs720470
Gene(s)	<i>NOS3</i>	<i>COL4A1</i>	<i>DYRK1A</i>
Risk allele/ reference allele	T/G	C/T	T/C
Risk allele frequency, %	32	36	71
Phenotype	AS	AS	AS
Analysis	EUR	TRANS	TRANS
MEGASTROKE OR [95% CI]	1.05 [1.03-1.07]	1.04 [1.03-1.06]	1.04 [1.02-1.06]
UKB OR [95% CI]	1.06 [1.02-1.11]	1.05 [1.01-1.09]	1.11 [1.07-1.17]
Combined OR [95% CI]	1.05 [1.04-1.07]	1.04 [1.03-1.06]	1.05 [1.03-1.07]
Combined P-value	2.2E-08	3.8E-08	6.1E-09

Table 1 Results from the fixed effects (trans-ancestral and European-only) GWAS meta-analyses. For each locus the variant the lowest p-value in the fixed effects trans-ancestral or European-only meta-analysis, respectively, is shown. Chr, chromosome; TRANS, trans-ancestral fixed-effects meta-analysis; EUR, European-only fixed-effects meta-analysis; OR, odds ratio; CI, confidence interval;

FIGURE LEGENDS

Figure 1 Manhattan plot and Regional Association Plots for the three novel risk loci identified in the current study.

(A) Manhattan plot for any stroke in the trans-ancestral meta-analysis. The dotted line marks the threshold of genome-wide significance. Novel loci are marked on the top. Note that *NOS3* reached genome-wide significance in the European-only analysis. (B to D) Regional Association Plots for lead variants at the three novel loci. (B) *NOS3*, (C) *COL4A1* and (D) *DYRK1A*. Shown is the region for the top signal ± 500 kb. The y-axis represents the $-\log(\text{p-value})$ from the European-only fixed-effects GWAS meta-analysis (*NOS3*) or trans-ancestral fixed-effects GWAS meta-analysis (*COL4A1* and *DYRK1A*). Variants in LD with the lead variant are shown in red ($r^2 > 0.6$), orange ($r^2 > 0.4$), yellow ($r^2 > 0.2$) and grey ($r^2 < 0.2$). RAF; risk allele frequency

Figure 2 Association of common variants in the NOS-NO signaling pathway with

blood pressure and with any stroke. For each SNP in the NOS-NO signaling pathway, the effect for systolic blood pressure (SBP) (A) and diastolic blood pressure (DBP) (B) is displayed on the x-axis and the odds ratio (OR) for any stroke from the European-only analysis is displayed on the y-axis together with their respective standard errors. The solid and dashed lines display the inverse variance weighted (IVW) effect estimate and 95% confidence interval (CI), respectively. Estimates for systolic and diastolic blood pressure were derived from UKB. Estimates for stroke were derived from inverse variance fixed effects meta-analyses of MEGASTROKE and UKB.

Figure 1

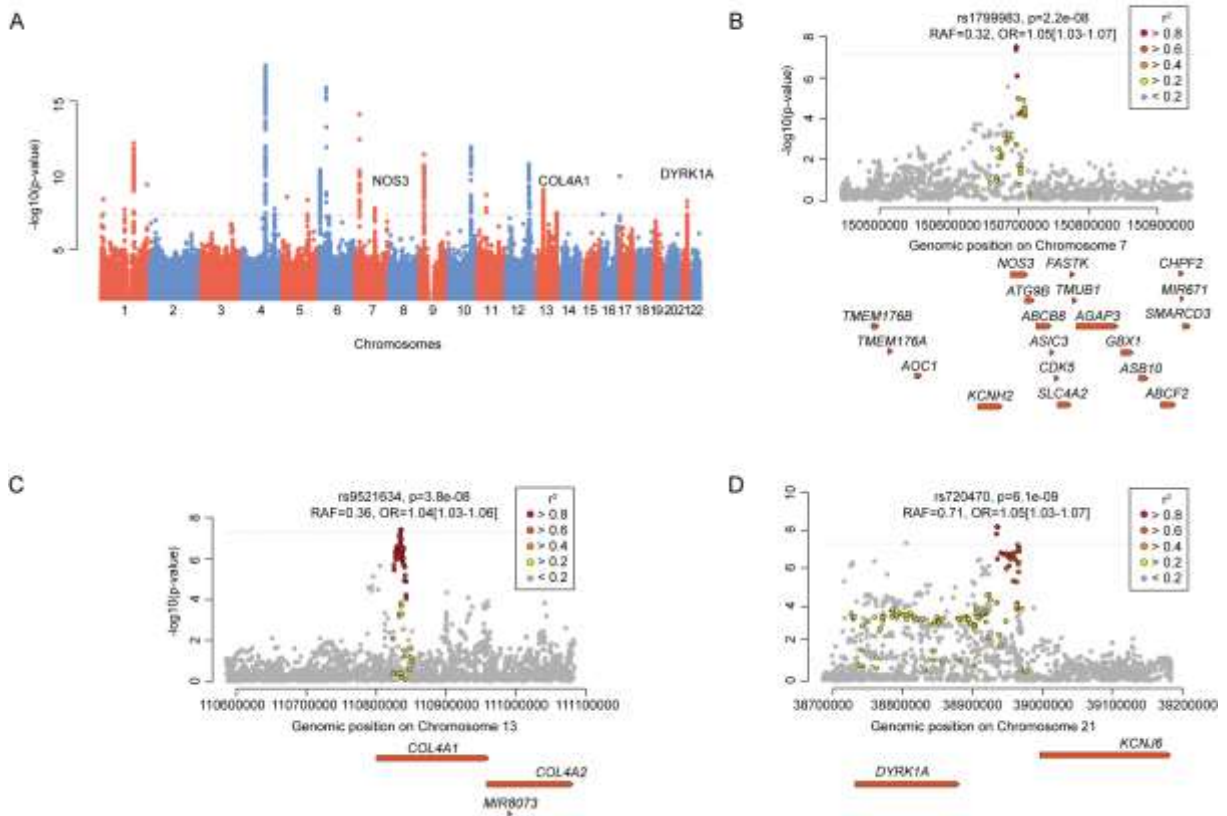


Figure 2

