This document contains support data for “Functional variants in the sucrase-isomaltase gene associate with increased risk of irritable bowel syndrome”.

Maria Henström, Lena Dieckmann, Ferdinando Bonfiglio, Fatemeh Hadizadeh, Eva-Maria Kuech, Maren von Köckritz-Blickwede, Louise B Thingholm, Tenghao Zheng, Ghazaleh Assadi, Claudia Dierks, Martin Heine, Ute Philipp, Ottmar Distl, Mary E Money, Meriem Belheouane, Femke-Anouska Heinsen, Joseph Rafter, Gerardo Nardone, Rosario Cuomo, Paolo Usai-Satta, Francesca Galeazzi, Matteo Neri, Susanna Walter, Magnus Simren, Pontus Karling, Bodil Ohlsson, Peter T Schmidt, Greger Lindberg, Aldona Dlugosz, Lars Agreus, Anna Andreasson, Emeran Mayer, John F Baines, Lars Engstrand, Piero Portincasa, Massimo Bellini, Vincenzo Stanghellini, Giovanni Barbara, Lin Chang, Michael Camilleri, Andre Franke, Hassan Y Naim, Mauro D’Amato

Items included:

METHODS

Study subjects

IBS families. Eight Caucasian individuals (seven postprandial IBS-D cases (Rome Criteria) and one asymptomatic relative) from four unrelated families were selected for re-sequencing of the SI gene (Supplementary Figure 1). Patients had post-prandial IBS-D (according to Rome criteria) but no evidence of inflammatory bowel disease or celiac disease at endoscopy, and experienced meal-related symptoms for more than 2 years and were currently using digestive enzyme supplements to reduce postprandial symptoms. At the time of genetic testing, patients were regularly using either an over the counter enzyme preparation (Essential Enzymes 500 mg by Source Natural containing a combination of different digestive enzymes), or a prescribed pancreatic enzyme supplement with a frequency ranging from 3 times/day to 3-4 times/week. All patients were experiencing improved postprandial IBS-D symptoms. Study participants with IBS-D included five adults (three females and two males aged from 47-92 years) with symptom onset from age 5 to 30,

and two teenagers (one girl and one boy aged 13 and 16 years, respectively) with symptoms onset from age 5. In addition, an asymptomatic parent (male, aged 46) of two affected probands was also included in the sequencing experiments. Each participant signed an informed consent, and the study protocol was approved by the Meritus Medical Center Institutional Review Board, Hagerstown, Maryland, USA.

*IBS case-control cohorts***.** We studied a total of 1031 unrelated IBS cases and 856 unrelated controls of self-reported European ancestry (white) from four independent cohorts from Sweden, Italy and US, who have been all described in detail elsewhere, and already included in previous genetic studies. Detailed information about these subjects is provided in Table S1. The Swedish cohort is a multicenter study with Rome III defined IBS patients recruited at tertiary care centers and clinics throughout the country, together with asymptomatic healthy controls.[1-4] The Italian multicenter case-control cohort includes IBS cases and asymptomatic controls enrolled at seven different clinics across the country, as previously described.[3,4] The two cohorts from US include Rome II IBS cases and healthy volunteers enrolled at the Mayo Clinic in Rochester, Minnesota described before,[1,3,4] and Rome III patients and asymptomatic controls recruited at University of California Los Angeles (UCLA), also reported in previous publications (Hispanic/Latino whites were excluded).[5,6] Subjects with inflammatory bowel disease or celiac disease were excluded in all cohorts. Establishment of a diagnosis of IBS and specific subtypes according to the Rome Criteria was obtained in all cases based on electronic medical record, validated bowel symptom questionnaires, and physicians' examination. Informed consent was obtained from all IBS patients and controls from the respective centers, and the study protocols were approved by local ethics committees.

*PopCol***.** The Population-based Colonoscopy study (PopCol) is a general population-based cohort with randomly selected participants from Stockholm, Sweden, previously described in detail.[7,8] A total of 1186 men and women, aged 18-70 years and born in Sweden, filled in questionnaires (containing Rome criteria bowel symptom modules) and visited a physician for a physical exam and routine blood tests. Of these, 745 subjects volunteered for colonoscopy, including 268 who also provided a stool sample and/or kept a detailed one-

week diary recording defecation patterns and abdominal symptoms. At follow-up, on average 6 years later, 190 subjects filled in a second identical diary for 14 days, with an overlap of 143 subjects who filled in both diaries. For the purpose of this study, we exploited questionnaire (Rome criteria), diary, genotype and fecal microbiota data from a total of 250 subjects, whose demographics are reported in Table 1 in the main text. Confirmed cases of inflammatory bowel disease (IBD) or celiac disease were excluded from all analyses. Informed consent was obtained from all participants and the PopCol study protocol was approved by the local Research Ethics Committee at Karolinska Institutet, Stockholm.

Resequencing of the *SI* gene in IBS family members

SI sequencing in IBS families. Saliva specimens were collected according to the protocol from Oragene (DNAgenotek, Ottawa, Canada) and DNA isolated using prepIT·L2P (DNAgenotek, Ottawa, Canada). The coding sequences of all 48 exons and relative boundaries of the human *SI* gene (NG_017043.1, NM_001041.3) were amplified as 42 separate PCR fragments as previously described.[9] The quality and yield of all PCR products were verified on agarose gels. Single stranded primers were digested by treatment of the PCR products with exonuclease I and alkaline phosphatase and purified PCR products were processed using the ABI PRISM BigDye Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific, Braunschweig, Germany). The DNA of the sequencing reactions was precipitated with ethanol containing EDTA pH 4.4, resuspended in 10 µl Hi-Di Formamide and separated on an Applied Biosystems 3500 Genetic Analyzer (Applied Biosystems, Foster City, USA). Sequence data were analyzed using the software Sequencher 4.8 (Gene Codes Corporation, Ann Arbor, USA).

Genotyping

Genotyping of CSDI mutations and p.Val15Phe was carried out at the Mutation Analysis Facility (MAF) at Karolinska Institute, using the iPLEX chemistry on Sequenom MassARRAY platform (SEQUENOM USA). Only cases and controls successfully genotyped for all CSID mutations and the p.Val15Phe SNPs were included in the analyses (N=38=2% removed). For PopCol participants, p.Val15Phe genotypes were extracted from available Illumina

HumanOmniExpressExome-8v1 array data, obtained from the genomics core facility at the Department of Medical Sciences (SNP&SEQ Technology Platform) of Uppsala University, Uppsala, Sweden.

16S sequencing from fecal microbiota

Bacterial DNA was extracted from fecal samples (stored at -80°C) using Qiagen QIAamp DNA Stool Mini Kit according to manufacturer's instructions. V1-V2 hypervariable regions from 16S ribosomal DNA were amplified using universal 27F (5'-AGAGTTTGATCCTGGCTCAG-3') and 338R (5'-TGCTGCCTCCCGTAGGAGT-3') primers, and sequenced using MiSeq next generation sequencing platform (Illumina. USA). For data processing and quality control, raw paired-end reads were merged using fastq_mergepairs function of Usearch (v7.0.1090)[10]. maxdiffs, minlen, minovlen, minmergelen and maxmergelen were considered as 2, 200, 150, 270 and 330, respectively. Fastq_filter function (maxee: 0.1) from the same software was used to quality filter FASTQ reads and convert them to FASTA. Chimeric sequences were specified using the UCHIME_ref function of the Usearch software[11] applying the SILVA gold database[12]. Bacterial taxonomy was assigned to each sequence from phylum to genus levels using classify.seqs command of Mothur[13]. Sequences classified as chloroplasts, mitochondria, archea, eukaryota and unknown were removed using remove.lineage command. Sequences were aligned to the SILVA alignment database using align.seq command of mothur and non- aligned sequences were trimmed manually. To further de-noise the sequences pre.cluster command was used and finally subsample command was applied to rarify data to 10,000 sequences per sample.

Functional characterization of the 15Phe variant *in vitro*

Transient transfections, biosynthetic labeling, immunoprecipitation and quantification. A Val to Phe change at residue 15 in the SI coding vector was obtained through site-directed mutagenesis via PCR using primers: 5´-ATTTAGTGGATTGGAGATCTCTCTGATTTTCCTTTTTGTCATAG-3´ and 5´-CTATGACAAAAAGGAAAATCAGAGAGATCTCCAATCCACTAAAT-3´ on the original pSG8 SI-expressing clone.[14] COS-1 cells (cultured with 5% CO_2 at 37°C in Dulbecco´s Modified

Eagle´s Medium (DMEM), 10% FCS, 5% penicillin/streptomycin) at 30-40% confluency were transfected with 5 µg of 15Val or 15Phe vectors with the diethylaminoethyl-dextran method.[15] Metabolic labeling was performed with 40 µCi [^{35}S] methionine either for 6 h continuously or in a pulse chase experiment for 1 h and different chase time points, in methionine-free MEM medium. Cell lysates were immunoprecipitated with a mixture of anti-SI monoclonal antibodies (mAb) and protein-A-sepharose. The immunoprecipitates were treated with endoglycosidase H (endo H) for 90 min at 37 °C as previously described,[16] to check the glycosylation pattern, or with trypsin (50 µg/ml) for 60 min at 37°C to determine the folding properties. Afterwards the samples were analyzed using a 6% SDS-PAGE and the signals visualized using a phosphorimager (Bio-Rad, Munich, Germany). Intensities on immunoblots or phosphorimages were quantified using the Quantity One® software from Bio Rad Laboratories GmbH (Munich, Germany).

DRM analysis. The association of 15Val or 15Phe SI to sphingolipid/cholesterol-rich microdomains was determined in detergent extractability assays using TX-100 or Lubrol WX.[17] Transiently transfected COS-1 cells were solubilized with lysis buffer containing 1% of the respective detergent and ultracentifuged at 100000g for 1 h at 4°C. The supernatant and the pellet were immunoprecipitated, loaded on a 6% SDS-PAGE and analyzed by immunoblotting against SI.

Biotin assay. Transiently transfected COS-1 cells were treated with Sulfo-NHS-SS-Biotin (1.5 mg/ml) for 30 min at 4°C, quenched two times with 0.1% BSA for 10 min at 4 °C, solubilized with lysis buffer containing 1% TX-100 and immunoprecipitated using a mixture of anti-SI mAbs. Each immunoprecipitate was split into two samples, one for immunoblotting against total SI and one without DTT for the immunoblot to detect the biotin-labeled SI by utilizing a horseradish peroxidase-conjugated streptavidin antibody. The amount of quantified biotinylated SI was related to the amount of total SI on the corresponding blot.

Confocal Fluorescence Microscopy. Transiently transfected COS-1 cells expressing 15Val or 15Phe were grown on cover slips and confocal laser scanning microscopy performed using the Leica TCS SP5 microscope and the x63 oil planachromat lens (Leica Microsystems,

Bensheim, Germany). For co-localization studies the cells were immunostained with a mixture of anti-SI antibodies and anti-GM130 as primary antibodies and anti-mouse IgG conjugated to Alexa Fluor® 488 and anti-rabbit IgG conjugated to Alexa Fluor® 633 as secondary antibodies. Detection of cell surface localized SI was achieved by treatment of live cells with the primary antibodies at 4°C, followed by extensive washing, fixation, and exposure to the Alexa Fluor® 488-coupled secondary antibody.

Antibodies. The monoclonal mouse anti-SI antibodies hSI2, HBB2/614/88 and HBB3/705/60 were generously provided by Prof. Dr. H. P. Hauri (Biocenter, Basel, Switzerland) and Prof. Dr. E. Sterchi (University of Bern, Bern, Switzerland). For immunoprecipitation a mixture of all antibodies was used to detect all different conformations and glycoforms of SI, whereas for immunoblot analysis only a mixture of HBB2/614/88 and HBB3/705/60 was applied. The secondary horseradish peroxidase-conjugated anti-mouse antibody and the monoclonal mouse anti-Streptavidin antibody were purchased from Thermo Fisher Scientific (Braunschweig, Germany). The secondary antibody coupled to Alexa Fluor dyes was obtained from Invitrogen (Karlsruhe, Germany).

***S*tatistical analyses**

PHRED scores from the Combined Annotation Dependent Depletion (CADD) database v1.3 (http://cadd.gs.washington.edu/) were used for predicting functional effects of known CSID mutations and *SI* common coding polymorphisms. CADD is a computational tool developed to predict and rank/score the deleteriousness of SNP and insertion/deletion variants in the human genome, through support vector machine (SVM) integration of different data sources and the production of a combined SVM score (C-score). Rather than “likelihood of pathogenicity”, the C-scores may be interpreted as “likelihood of deleteriousness”, but it is often adopted to prioritize causal variation in research and clinical settings.[18] Association testing of the SI variants comparing IBS cases and controls was performed in PLINK v1.9 (http://pngu.mgh.harvard.edu/~purcell/plink/). Only subjects who were successfully genotyped for all four mutations and the p.Val15Phe SNP were included in the analyses. Hardy-Weinberg equilibrium was calculated in the control groups using Fischer’s exact test to

confirm (P>0.05) allelic equilibrium at each mutation/SNP site in the study populations. The sample size had >80% statistical power to detect significant association with genetic risk effects OR>2 and OR>1.5, respectively, for CSID mutations (cumulative) and the p.Val15Phe variant. One-sided (testing predisposing risk effects) association analysis of CSID mutations carriage was performed using 2x2 contingency table statistics (Fisher's exact test and Chi-square as appropriate) in IBS cases vs i) controls and ii) reference genotypes from publicly available large-scale sequence data for individuals of European descent (http://exac.broadinstitute.org/).[19] Association with the p.Val15Phe variant was tested on pooled data using logistic regression under a (best powered) additive genetic model adjusting for sex and batch (recruitment center). A meta-analysis of results from individual cohorts was also carried out (using a fixed-effects model with the --meta command in PLINK) to verify absence of statistical heterogeneity. To test the correlation between 15Phe genotype and average bowel movement frequency, we used 2-tailed non-parametric Spearman's correlation in SPSS (v.22.0.0.0). A total of 133 subjects were included for this, all of which had valid recordings from both 7- and 14-day diaries, and the mean stool frequency (number of bowel movements per day) was calculated using the average over the two diaries. The analyses of PopCol microbiota composition in relation to 15Phe were performed in R (v.3.1.2) using Spearman's rank statistics (to test for the correlation between genera abundance and 15Phe copy number).

References

1. Zucchelli M, Camilleri M, Andreasson AN, et al. Association of TNFSF15 polymorphism with irritable bowel syndrome. Gut 2011;60:1671-7.
2. Wouters MM, Lambrechts D, Knapp M, et al. Genetic variants in CDC42 and NXPH1 as susceptibility factors for constipation and diarrhoea predominant irritable bowel syndrome. Gut 2014;63:1103-11.

3. Ek WE, Reznichenko A, Ripke S, et al. Exploring the genetics of irritable bowel syndrome: a GWA study in the general population and replication inmultinational case - control cohorts. Gut 2015;64(11):1774-82.
4. Beyder A, Mazzone A, Strege PR, et al. Loss-of-function of the voltage-gated sodium channel NaV1.5 (channelopathies) in patients with irritable bowel syndrome. Gastroenterol 2014;146:1659-68.
5. Grasberger H, Chang L, Shih W, et al. Identification of a functional TPH1 polymorphism associated with irritable bowel syndrome bowel habit subtypes. Am J Gastroenterol 2013;108:1766-74.
6. Orand A, Gupta A, Shih W, et al. Catecholaminergic Gene Polymorphisms Are Associated with GI Symptoms and Morphological Brain Changes in Irritable Bowel Syndrome. PLoS One 2015;10:e0135910. doi: 10.1371/journal.
7. Walter SA, Kjellström L, Nyhlin H, Talley NJ, Agréus L. Assessment of normal bowel habits in the general adult population: the Popcol study. Scand J Gastroenterol 2010;45:556-66.
8. Kjellström L, Molinder H, Agréus L, Nyhlin H, Talley NJ, Andreasson A. A randomly selected population sample undergoing colonoscopy: prevalence of the irritable bowel syndrome and the impact of selection factors. Eur J Gastroenterol Hepatol 2014;26:268-75.
9. Sander P, Alfalah M, Keiser M et al. Novel mutations in the human sucrase-isomaltase gene (SI) that cause congenital carbohydrate malabsorption. Hum Mutat 2006;27:119.
10. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 2010;26:2460-1.
11. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 2011;27:2194-200.
12. Pruesse E, Quast C, Knittel K, et al. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res 2007;35:7188-96.

13. Schloss PD, Westcott SL, Ryabin T, et al. Introducing mothur: opensource, platform-independent, community-supported software for describing andcomparing microbial communities. Appl Environ Microbiol 2009;75:7537-41.
14. Hoefsloot LH, Hoogeveen-Westerveld M, Reuser AJ, et al. Characterization of the human lysosomal alpha-glucosidase gene. Biochem J 1990; 272:493-7.
15. Naim HY, Lacey SW, Sambrook JF, et al. Expression of a full-length cDNA coding for human intestinal lactase-phlorizin hydrolase reveals an uncleaved, enzymatically active, and transport-competent protein. J Biol Chem 1991; 266:12313-20.
16. Naim HY, Sterchi EE, Lentze MJ. Biosynthesis and maturation of lactase-phlorizin hydrolase in the human small intestinal epithelial cells. Biochem J 1987; 241:427-34.
17. Alfalah M, Jacob R, Preuss U, et al. O-linked glycans mediate apical sorting of human intestinal sucrase-isomaltase through association with lipid rafts. Curr Biol 1999;9:593-6.
18. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet 2014;46:310–5.
19. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature 2016;536:285-91.

Table S1. Cohorts included in the study, extended information and analytical pipeline

	Sweden		Italy		US (Mayo)		US (UCLA)		PopCol	
	Cases	**CTRLs**	**Cases**	**CTRLs**	**Cases**	**CTRLs**	**Cases**	**CTRLs**	**Cases**	**CTRLs**
Study setting	Tertiary centers		Tertiary centers		Tertiary centers		Tertiary centers		Random selection from Stockholm area, Sweden (population registries)	
Study type	Case-control (unmatched)		Case-control (unmatched)		Case-control (unmatched)		Case-control (unmatched)		Prospective population-based (case-control unmatched for this study)	
Inclusion criteria	GI specialist IBS diagnosis (Rome III criteria)	Asymptomatic healthy individuals	GI specialist IBS diagnosis (Rome III criteria)	Asymptomatic healthy individuals	GI specialist IBS diagnosis (Rome II criteria)	Asymptomatic healthy individuals	GI specialist IBS diagnosis (Rome III criteria)	Asymptomatic healthy individuals	Rome III criteria (questionnaire)	Asymptomatic healthy individuals (questionnaire)
Exclusion criteria	IBD or Celiac disease	IBS, IBD or Celiac disease	IBD or Celiac disease	IBS, IBD or Celiac disease	IBD or Celiac disease	IBS, IBD or Celiac disease	IBD or Celiac disease	IBS, IBD or Celiac disease	IBD or Celiac disease	IBS, IBD or Celiac disease
Race/ethnicity [1]	white	white	white	white	white	white	white	white	white	white
Ancestry	self-reported	self-reported	self-reported	self-reported	self-reported	self-reported	self-reported	self-reported	self-reported	self-reported
DNA source	Peripheral blood		Peripheral blood		Peripheral blood		Peripheral blood		Peripheral blood	
Genotyping center	Genotyping Facility (MAF), Karolinska Institutet, Stockholm		Genotyping Facility (MAF), Karolinska Institutet, Stockholm		Genotyping Facility (MAF), Karolinska Institutet, Stockholm		Genotyping Facility (MAF), Karolinska Institutet, Stockholm		Genotyping Facility (SNP&SEQ) Uppsala University, Uppsala	
Platform	Sequenom MassARRAy (iPLEX)		Sequenom MassARRAy (iPLEX)		Sequenom MassARRAy (iPLEX)		Sequenom MassARRAy (iPLEX)		Illumina HumanOmniExpressExome	
N at inclusion	400	359	327	263	187	115	141	133	30	164
Success rate [2]	387 (96.8%)	355 (99.0%)	319 (97.6%)	255 (97.0%)	187 (100%)	114 (99.1%)	138 (98.9%)	132 (99.7%)	30 (100%)	163 (99.4%)
Mean age, yr	42	42.5	39.6	34.8	47.9	46.6	35	33.5	51.6	54.7
% F:M	82:18:00	43:57:00	73:27:00	68:32:00	91:09:00	67:33:00	75:25:00	58:42:00	60:40:00	59:41:00
IBS-D	127		159		89		42		11	
IBS-C	95		128		56		45		5	
IBS-M	162		32		41		34		9	
IBS-U	3		0		1		17		5	
CSID statistics	One-tailed Fischer's exact test for carriership of rare CSID mutations								–	
15Phe statistics	Logistic regression, additive genetic model, adjusted for sex (individual analyses) and study center (pooled analysis) Meta-analysis: inverse-variance method, fixed-effects								Logistic regression, additive genetic model, adjusted for sex and center	
Software	PLINK v1.9		PLINK v1.9		PLINK v1.9		PLINK v1.9		PLINK v1.9	
Study reference	1,2,3,4		3,4		1,3,4		5,6		7,8	

[1] Non-Hispanic white excluded from US-based cohorts; [2] Calculated on individuals successfully genotyped for all CSID mutation and the 15Phe variant

Table S2. Spearman correlation analysis of 15Phe genotype vs 20 most abundant genera (fecal microbiota)

Genera	P	Rho	Relative abundance
Bacteroides	0.97	0.003	0.154084
Unclassified Ruminococcaceae	0.65	0.039	0.148871
Unclassified Lachnospiraceae	0.94	0.006	0.09082
Unclassified Clostridiales	0.52	0.055	0.081297
Unclassified	0.11	0.138	0.048686
Alistipes	0.75	-0.027	0.046443
Faecalibacterium	0.33	-0.083	0.045885
Oscillibacter	0.41	0.071	0.037953
Prevotella	0.71	-0.032	0.024371
Escherichia/Shigella	**0.020**	0.201	0.023515
Parabacteroides	**0.0024***	-0.225	0.019322
Unclassified Bacteroidales	0.77	-0.025	0.019224
Barnesiella	0.62	-0.042	0.017369
Unclassified Enterobacteriaceae	0.21	0.107	0.015992
Roseburia	0.25	-0.100	0.01535
Unclassified Erysipelotrichaceae	0.43	-0.068	0.014239
Ruminococcus	0.51	-0.056	0.012001
Phascolarctobacterium	0.060	-0.160	0.011429
Unclassified Prevotellaceae	0.37	0.077	0.011343
Blautia	0.74	-0.028	0.010081

* Significant after Bonferroni correction for multiple testing (20 genera)

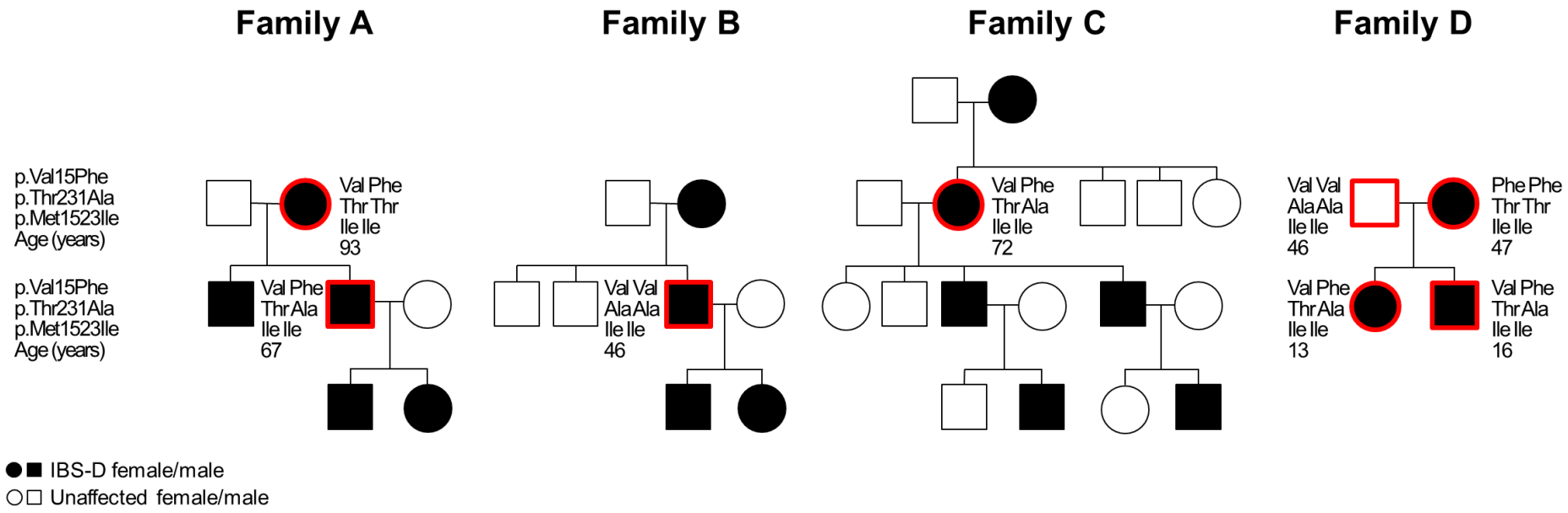

Figure S1. Individuals included in the *SI* re-sequencing experiment (highlighted in red), are reported with age and genotype at relevant *SI* coding SNPs. Respective family trees are also reported for descriptive purposes only.

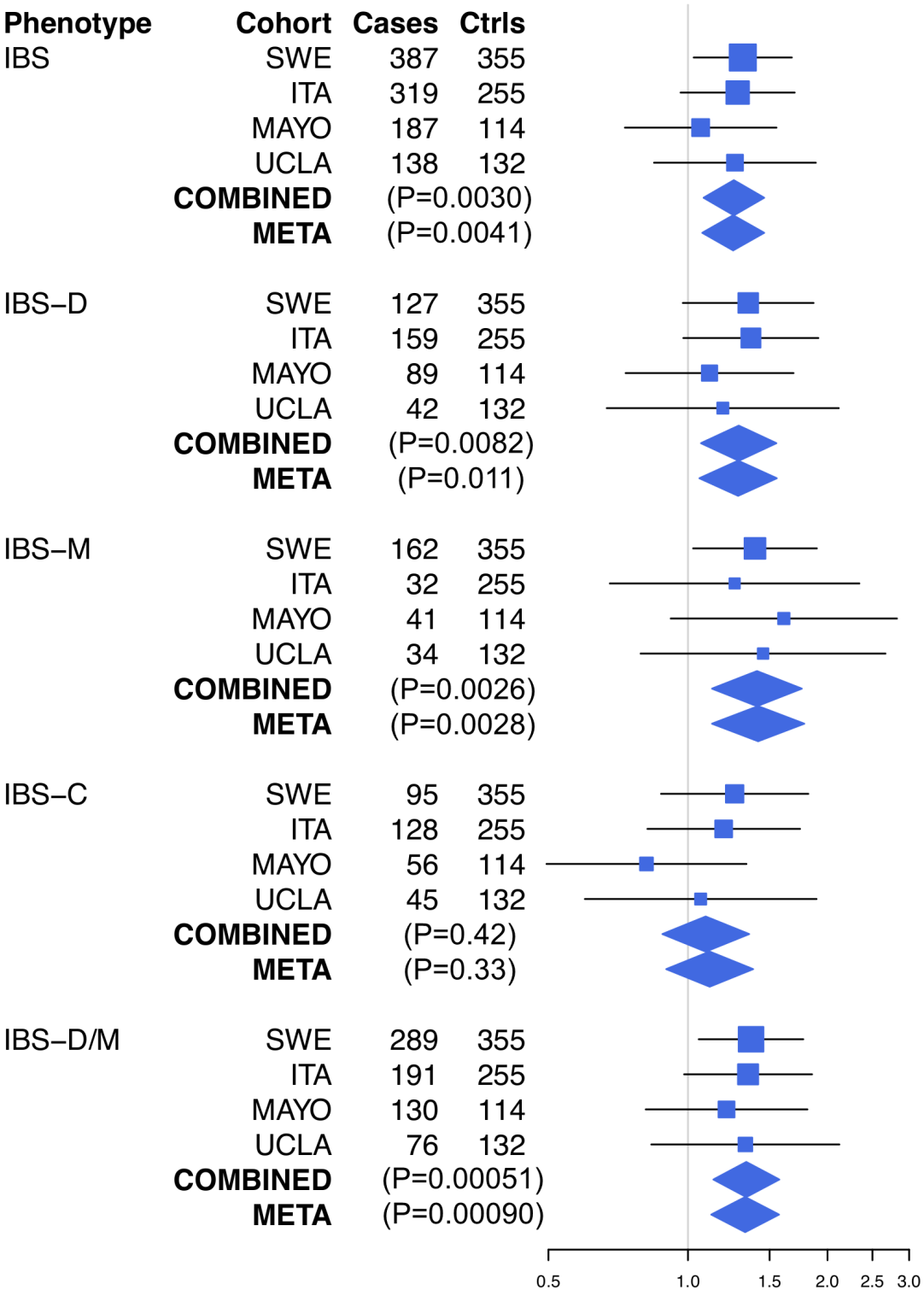

Figure S2. Forrest plot summarizing 15Phe IBS risk effects in our study. Odds ratios and 95% confidence intervals from sex and batch (recruitment center) corrected logistic regression analysis are reported for IBS and IBS subtypes, relative to individual cohorts and the total sample, together with the respective number of cases and controls. Results from a meta-analysis of individual cohorts are also shown for comparison with pooled statistics.

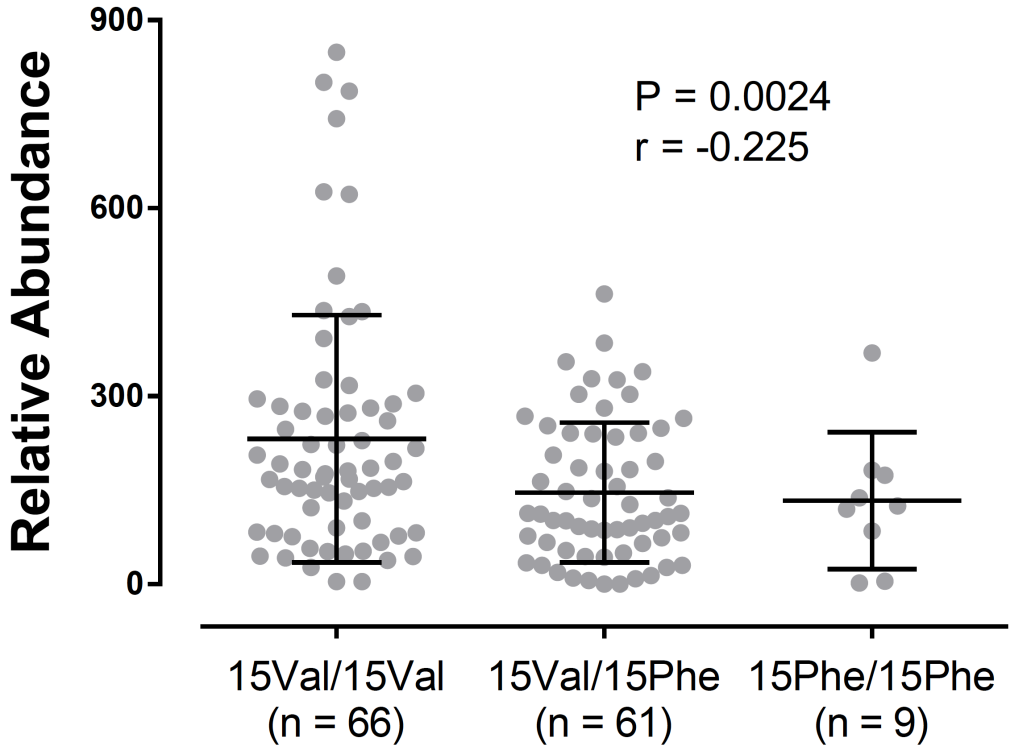

Figure S3. Fecal microbiota abundance of *Parabacterides* correlates with p.Val15Phe genotype. *Parabacterides* sequence counts/10000 reads (Y axis) are reported for PopCol individuals (dots) stratified according to genotype at the p.Val15Phe SNP site (X axis). Mean (middle) and SD (top and bottom) are reported as horizontal bars for each genotype group. Spearman's P value and correlation coefficient are also shown.