**Supplementary materials to**

# Rsu1 Regulates Ethanol Consumption in *Drosophila* and Humans

Shamsideen A. Ojelade,[1,2] Tianye Jia,[3,4] Aylin R. Rodan,[1,%] Tao Chenyang,[5,6] Julie L. Kadrmas,[7] Anna Cattrell,[3,4] Barbara Ruggeri,[3,4] Pimphen Charoen,[8] Hervé Lemaitre,[9] Tobias Banaschewski,[10] Christian Büchel,[11] Arun L.W. Bokde,[12] Fabiana M. Carvalho,[3,4] Patricia J. Conrod,[3,13] Herta Flor,[14] Vincent Frouin,[15] Jürgen Gallinat,[16] Hugh Garavan,[12,17] Penny A. Gowland,[18] Andreas Heinz,[16] Bernd Ittermann,[18] Mark Lathrop,[19] Steven Lubbe,[3,4] Jean-Luc Martinot,[9] Tomás Paus,[20-22] Michael N. Smolka,[23,24] Rainer Spanagel,[10] Paul F. O'Reilly,[3,4] Jaana Laitinen,[25] Juha M. Veijola,[26] Jianfeng Feng,[5,6,27] Sylvane Desrivières,[3,4] Marjo-Riitta Jarvelin,[8,28-31] the IMAGEN consortium, Gunter Schumann,[3,4,*] and Adrian Rothenfluh[1,2,*]

[1]Department of Psychiatry

[2]Program in Neuroscience, UT Southwestern Medical Center at Dallas, USA

[3]Institute of Psychiatry, King's College London, United Kingdom

[4]MRC Social, Genetic and Developmental Psychiatry (SGDP) Centre, London, United Kingdom

[5]Center for Computational Systems Biology, Faculty of Methematics, Fudan University, Shanghai, China

[6]Department of Computer Science, Warwick University, Coventry

[7]Department of Oncological Sciences, Huntsman Cancer Institute, University of Utah

[8]Department of Epidemiology and Biostatistics, MRC Health Protection Agency Centre for Environment and Health, School of Public Health, Imperial College London, UK

[9]Institut National de la Santé et de la Recherche Médicale, INSERM CEA Unit 1000 "Imaging & Psychiatry", University Paris Sud, Orsay, and AP-HP Department of Adolescent Psychopathology and Medicine, Maison de Solenn, University Paris Descartes, Paris, France

[10]Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, Germany
[11]Universitätsklinikum Hamburg Eppendorf, Hamburg, Germany
[12]Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland
[13]Department of Psychiatry, Université de Montreal, CHU Ste Justine Hospital, Canada
[14]Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, Germany
[15]Neurospin, Commissariat à l'Energie Atomique, Gif-sur-Yvette, France
[16]Department of Psychiatry and Psychotherapy, Campus Charité Mitte, Charité – Universitätsmedizin Berlin, Germany
[17]Departments of Psychiatry and Psychology, University of Vermont, Burlington, USA
[18]Physikalisch-Technische Bundesanstalt (PTB), Braunschweig und Berlin, Germany
[19]Centre National de Génotypage, Evry, France
[20]School of Psychology, University of Nottingham, United Kingdom
[21]Rotman Research Institute, University of Toronto, Toronto, Canada
[22]Montreal Neurological Institute, McGill University, Canada
[23]Department of Psychiatry and Psychotherapy, Technische Universität Dresden, Germany
[24]Neuroimaging Center, Department of Psychology, Technische Universität Dresden, Germany
[25]Finnish Institute of Occupational Health, Helsinki, Finland
[26]Department of Psychiatry, University of Oulu and Oulu University Hospital, Oulu, Finland
[27]School of Life Science and the Collaborative Innovation Center for Brain Science, Fudan University, Shanghai, China;
[28]Institute of Health Sciences, University of Oulu, Finland
[29]Biocenter Oulu, University of Oulu, Finland
[30]Unit of Primary Care, Oulu University Hospital, Oulu, Finland
[31]Department of Children and Young People and Families, National Institute for Health and Welfare, Oulu, Finland

% current address: Division of Nephrology, Department of Internal Medicine, UT Southwestern Medical Center at Dallas, USA

* these authors contributed equally

## SUPPLEMENTARY MATERIALS AND METHODS

### Fly Stocks and Genetics

*Drosophila melanogaster* were raised in a 12:12 hr L:D cycle on a standard cornmeal/molasses diet at 25°C with 70% humidity, except for temperature sensitive experiments, which used 18 or 29°C as indicated. *w Berlin* served as the genetic background for all experiments (unless explicitly stated) which were done with 2-7 day old flies during the light phase. The Gal4-expressing $ics^{G4}$ line was obtained through a P{GawB} forward genetic screen as described (1). Excisions to obtain $ics^{x5}$ (imprecise excision) and $ics^{x23}$ (precise excision) were carried out through standard genetic crosses using the {delta2-3} jump-starter insertion and were verified by PCR and standard DNA sequencing analyses. UAS-Rsu1 transgenics were generated by PCR amplification of LD43981, introducing an N-terminal BglII site and a C-terminal XhoI site, subcloning into pUASt vector, and injection into embryos (Duke Model System Genomics). The *Rsu1 UAS*-RNAi construct targeting the fourth, and largest exon of *Rsu1* (*UAS-Rsu1* RNAi) was amplified with primers ACAACAAGATCAGCGTAATCAGTCCGGGAA and CTTATAGGTCTCCGTTTTGAGGTAGTCGATG, and cloned into pWIZ (2). This construct was injected using standard procedures. Integrin mutants ($mys^{ts}$) were obtained from M. Grotewiel (3). All other fly lines were obtained from the Bloomington Stock center.

### Fly Ethanol Behaviors

Loss-of-righting (LOR) assay was performed as described previously (1). Twenty males (except in Fig. 4*A*) per tube were exposed to ethanol vapor. The LOR of ethanol-exposed flies was measured during ethanol exposure every 5 min by lightly tapping the tube and then counting the flies unable to right themselves. The time to 50% LOR (ST-50) was calculated for each exposure tube by linear interpolation of the two time points around the median and then averaged over the number of tubes. The data shown in most behavior figures were collected from assays performed on a single day, to eliminate day-to-day variability. However, all experiments were repeated on multiple days, with similar results.

Ethanol preference was performed using the 2-bottle choice Capillary Feeder (CAFÉ) assay as described (4) with some modifications. Our CAFÉ apparatus consisted of a 6 well plate with 4 small holes per well drilled for insertion of truncated pipette tips and 5 µl capillaries (VWR, Radnor, PA), and 2 damp cotton balls in between wells for humidity. Capillaries were filled by capillary action, a small mineral oil overlay was added to reduce evaporation, and the capillaries were measured and replaced daily. Preference assays with 8 males per well were conducted at 25°C and 70% relative humidity, and flies chose between liquid sucrose/yeast food with, or without 15% ethanol. For the *MB-GeneSwitch* experiment, food-deprived flies were fed with 0.5 mM mifepristone for 3 hours prior to the CAFÉ assay.

For measurements of ethanol concentration, flies were frozen in dry ice and homogenized in 50 mM Tris-HCl (pH 7.5), and then assayed using a kit from Genzyme Diagnostics P.E.I Inc (Charlottetown, PE, Canada).

**Ethanol Absorption**

Ethanol concentration in flies was measured using the ethanol reagent kit (# 229-29) from Genzyme Diagnostics. Millimolar ethanol concentration in flies was calculated in assuming the volume of a fly to be 2µl. Flies (a total of n=3 per genotype were tested, where n=1 consisted of 90 flies) were exposed to ethanol vapors (E/A: 150/0) for various times points and sedation was monitored throughout the exposures. At the end of the exposures, flies were frozen in dry ice and homogenized

**Immunohistochemistry**

Immunohistochemistry was performed as described (1). Green fluorescent protein (GFP) was visualized with anti-GFP (rabbit anti-GFP, 1:250); to label relevant architectural features, the presynaptic marker mouse anti-nc82 was used at 1:40 to label general neuropil/brain structure. It was developed by Erich Buchner, and obtained from the Developmental Studies Hybridoma Bank developed under the auspices of the NICHD and maintained by The University of Iowa, Department of Biological Sciences, Iowa City, IA.

**Schneider Cell Culture**

Stably expressing S2-Gal4 cells were transfected with 3µg of plasmids tagged to GFP or Flag for co-immunoprecipitation assay. Gateway plasmids transfected were the constitutively active form of Rac1 (pT.wV RacG12V) and Rho (pT.wV RhoG12V), dominant negative Rac1 (pT.wV RacT17N) and Rho1 (pT.wV RhoT19N), and Flag-

tagged Rsu1 (pT.wF Rsu1). Rac.GTP pull down experiments were performed with GST-PBD bait protein (EMD Millipore, Billerica, MA).

Rsu1 dsRNA was made using two T7 primers to prepare a Rsu1 cDNA template for in vitro transcription using the Ambion T7 MEGAscript Kit (Invitrogen, Grand Island, NY). 50µg of Rsu1 dsRNA was added to S2 cell culture for 72 hours to deplete Rsu1 levels.

**G/F-actin In Vivo Assay**

G/F-actin assay was performed according to the manufacturer's instructions (G/F-actin In Vivo Assay Kit, Cytoskeleton, Denver, CO). G- and F-actin bands on western blots were scanned by densitometry and the ratios of free G-actin to actin present as F-actin were calculated.

**Statistics for *Drosophila* Experiment**

Statistical significance was established with one-way analysis of variance (ANOVA) tests using GraphPad Prism for Mac. Since each measurement was counted based on 20 flies, its normality is automatically fulfilled based on the central limit theorem. For the post-hoc analyses, Dunnett's Test was applied to control for the multiple comparison when several groups were compared to the same control. Error bars in all experiments represent SEM. Significance was only attributed to experimental lines that were statistically different from their respective controls, defined as $p < 0.05$. In all graphs $^{***}$ = $p < 0.001$, $^{**}$ = $p < 0.01$, $^{*}$ = $p < 0.05$.

**Human Cohort**

Participants were tested in eight IMAGEN assessment centers (London, Nottingham, Dublin, Mannheim, Berlin, Hamburg, Paris and Dresden). The study was approved by local ethics research committees at each site. A detailed description of recruitment and assessment procedures, as well as in/exclusion criteria, has previously been published (5). In addition, all participants passed quality control procedures for the behavioral, functional MRI (fMRI), genotyping and gene expression data.

**Monetary Incentive Delay Task and Neuroimaging Analyses**

This version of the MID task has been carried out as previously described (5). *Monetary Incentive Delay (MID) Task*: This version of the MID task consisted of 66 10-second trials. In each trial participants were presented with one of three cues (displayed for 250ms) denoting whether a target (white square) would appear on the left or right side of the screen, and whether 0, 2 or 10 points could be won in that trial. After a variable delay (4000-4500ms) of fixation on a white cross hair participants were instructed to respond with a left or right button press as soon as the target appeared. Feedback on whether and how many points were won during the trial was presented for 1450ms after the response. A tracking algorithm adjusted task difficulty (i.e. target duration varied between 100 and 300ms) so that each participant successfully responded on ~66% of the trials. For every 5 points won the participant received one food snack in the form of chocolate candy. Only successfully hit trials were included for analysis.

*Functional MRI data analysis*: Functional MRI data were analyzed with SPM8 (Statistical Parametric Mapping version 8; http://www.fil.ion.ucl.ac.uk/spm). Slice-time

correction was conducted to adjust for time differences caused by multislice imaging acquisition, all volumes were aligned to the first volume, and nonlinear warping was performed to an echo planar imaging (EPI) template. Images were then smoothed with a Gaussian kernel of 5-mm full width at half-maximum. At the first level of analysis, changes in the BOLD response for each subject were assessed by linear combinations at the individual subject level for each experimental condition, and each trial (i.e. reward anticipation high gain) was convolved with the hemodynamic response function to form regressors that account for variance associated with the processing of reward anticipation. Estimated movement parameters were added to the design matrix in the form of 18 additional columns (3 translation, 3 rotation, 3 quadratic and 3 cubic translation columns, and each 3 translations had a shift of ±1 repetition time). Single-subject contrast images were normalized to Montreal Neurological Institute space. The normalized and smoothed single-subject contrast images were then taken to a second-level random effects analysis.

*Whole Brain Analysis:* As this analysis is exploratory, the voxel-wise height threshold was set at $p < 0.001$ uncorrected. Statistically significant differences between genotype groups are reported as voxel-intensity t-values for clusters at $p < 0.05$ family wise error (FWE) corrected. All analyses control for handedness, gender and imaging site. The beta values from the significant clusters were averaged across all voxels within these clusters using the MarsBaR toolbox (http://marsbar.sourceforge.net) and the data exported for graphical presentation in MS Excel.

Region of Interest (ROI) Analysis: Using the MarsBaR toolbox (http://marsbar.sourceforge.net) the ventral striatum (VS) ROI was extracted from the 'anticipation of high gain vs. anticipation of no gain' contrast. The extracted ROI was based on (xyz ±15 9 -9, sphere radius 9mm; (6). The beta values were averaged across all voxels within the region and these data were exported for statistical analysis in PLINK (http://pngu.mgh.harvard.edu/~purcell/plink/).

**Behavioral Characterization.**

The 'Lifetime Frequency of Drinking' phenotype was defined using an adapted version of the 2007 ESPAD questionnaire (www.espad.org), which assesses "the frequency of alcohol use in your lifetime". The variable is coded in a 7-point scale ranging from 0 ("none drinker") to 6 ("50 times or more").

**Human Genetic Analyses**

For the fMRI data, 1303 baseline adolescents at age 14 years (mean = 14.4, SD = 0.4, range: 12.9-16.4) were included in the SNP analysis.

DNA purification and genotyping were performed by the Centre National de Génotypage in Paris. DNA was extracted from whole blood samples (~10ml) preserved in BD Vacutainer EDTA tubes (Becton, Dickinson and Company, Oxford, UK) using Gentra Puregene Blood Kit (QIAGEN Inc., Valencia, CA) according to the manufacturer's instructions. Genotype information was collected at 582,982 markers using the Illumina HumanHap610 Genotyping BeadChip (Illumina, San Diego, CA). SNPs with call rates of

<98%, minor allele frequency <1% or deviation from the Hardy-Weinberg equilibrium ($P \leq 1\times10^{-4}$) were excluded from the analyses. Individuals with an ambiguous sex code, excessive missing genotypes (failure rate >2%), and outlying heterozygosity (heterozygosity rate of 3 SDs from the mean) were also excluded. Identity-by-state similarity was used to estimate cryptic relatedness for each pair of individuals using PLINK software. Closely related individuals with identity-by-descent (IBD > 0.1875) were eliminated from the subsequent analysis. Population stratification for the GWAS data was examined by principal component analysis (PCA) using EIGENSTRAT software. The four HapMap populations were used as reference groups in the PCA analysis and individuals with divergent ancestry (from CEU) were also excluded.

DNA purification and genotyping was performed by the Centre National de Génotypage in Paris. Details are provided in the supplementary information. In total, 70 SNPs were detected in the human *RSU1* gene, and PLINK (http://pngu.mgh.harvard.edu/~purcell/plink) was implemented in the association analysis between the candidate SNP and phenotypes as well as the corresponding permutation analysis if applicable. All statistical analyses were controlled for gender and site, and handedness was also controlled in case the fMRI data were involved. Presented p-values were all uncorrected unless otherwise specified.

**Kernel-based Association Analysis**

We used the kernel-generalized variance (7) to quantify the dependency between the BOLD response and Genes in the IMAGEN samples. Statistical inference was based on a permutation procedure, both a parametric approximation of the p-value and an

empirically p-value were calculated. For brief, a kernel based canonical correlation analysis (CCA) is to solve the following eigen-problem between two random vectors *Y* and *X*:

$$\begin{pmatrix} (K_Y+\lambda I)^2 & K_Y K_X \\ K_Y K_X & (K_X+\lambda I)^2 \end{pmatrix} \begin{pmatrix} \xi_Y \\ \xi_X \end{pmatrix} = (1+\rho) \begin{pmatrix} (K_Y+\lambda I)^2 & 0 \\ 0 & (K_X+\lambda I)^2 \end{pmatrix} \begin{pmatrix} \xi_Y \\ \xi_X \end{pmatrix}$$

,where f and g are transformation determined by linear combination of kernel function, $K_X$and $K_Y$ are the Gram matrix of the sample calculated using the kernel function, and λ is a small regularization parameter to avoid overfitting. The kernel generalized variance statistics ('regularized kernel association' would be a more preferable name) is then defined as

$$rkassoc^{(k)} = -\sum_{i=1}^{k} \log\left(1-\rho_i^2\right)$$

where $\rho_i$ is the $i^{th}$ leading eigenvalue of the regularized eigenproblem (minus one from the eigenvalue calculated directly from the matrices). This is justified by the fact that the rest eigenvalues converge rapidly to 0, and therefore retaining these eigenvalues will not only contribute little to the association but also sacrifice the numerical stability. The thus defined kernel generalized variance approximates the mutual information between variables *Y* and *X* to the second level when the variables in question follow arbitrary distributions and 'near independency' (see (7) for a detailed proof), and therefore $rkassoc^{(k)} \approx 0$ if *Y* and *X* are independent.

To get the p-value under the NULL hypothesis, we used the permutation procedure. The columns of *X* were permutated for *B* times, and the permuted kernel generalized variance statistics were calculated and recorded. The empirical p-value was then

calculated as the percent of permuted statistics exceeding the original one. We found that a gamma fit of the permuted statistics approximates the NULL distribution of statistics quite well and is especially useful when one is performing screening where statistical correction are required. We truncated the left most 2% tail of the permuted NULL to get a stable fit of the gamma distribution. A diagnostic QQ-plot was also produced along with the approximated p-value in case of deviation from the gamma fit, which happens if inappropriate regularization parameter or kernel function is chosen, or the distribution from the original space is too irregular given a small sample size. Intensive numerical simulations reveal no sign of inflated false positive rate of the kernel association measure. Nevertheless, in this paper we will report both the approximate/theoretical and the empirical p-values.

To control for the covariates, we eliminate the covariate effect from the original space of the data. While stimulation studies indicate that removing covariates from one side is more conservative than removing covariates from two sides, both of them show no sign of inflated false positives.

For our current application, Linear kernels were used, and the determination of kernel bandwidth followed the protocol of (7), the regularization parameter was set to 0.1. For consistence, the number of leading eigenvalues was set to 5 and the step for iterative process was set to $1.0 \times 10^{-5}$ throughout the whole paper. We permuted the sample for 1, 000 to 10000 times to get the empirical p-value. For large sample, accurate linear time approximation algorithm with incomplete Cholesky decomposition is used to speed up the calculation.

We thank Chenyang Tao for his contribution of establishing the kernel association

analyses and the manuscript for a more generalized method is under submission. The toolbox of this generalized method 'NAC' could be found at http://www.dcs.warwick.ac.uk/~feng/nac/.

**North Finnish Birth Cohort 1966**

A sample of 4772 individuals from the Northern Finland Birth Cohort 1966 (NFBC 1966) with genotypic and phenotypic data available was drawn from the population-based NFBC 1966 (see http://www.oulu.fi/nfbc/). Pregnant females with delivery dates in 1966 were recruited from the northern Finish provinces of Oulu and Lapland. Offspring data used here was obtained in 1997, when the cohort was 31 years of age. Frequencies of food and alcohol consumption at 31 y were ascertained as part of the larger postal questionnaire, which the study subjects returned at the clinical examination. Alcohol use questions (AUQ) were designed to measure the average frequency of consumption of beer, wine and spirits during the last year, and the usual amount of each consumed on one occasion. The amount of alcohol consumed per day was calculated using the following estimates of alcohol content (vol%): beer 4.8; light wines 5.0; wines 14.5; spirits 37.0. The subjects were then assigned to four groups by sex-specific quartiles of alcohol intake, those in the highest quartile being regarded as heavy drinkers.

## SUPPLEMENTARY REFERENCES


1. Rothenfluh A et al. (2006) Distinct Behavioral Responses to Ethanol Are Regulated by Alternate RhoGAP18B Isoforms. *Cell* 127:199–211.

2. Sik Lee Y (2003) Making a better RNAi vector for Drosophila: use of intron spacers. *Methods* 30:322–329.

3. Bhandari P, Kendler KS, Bettinger JC, Davies AG, Grotewiel M (2009) An Assay for Evoked Locomotor Behavior in Drosophila Reveals a Role for Integrins in Ethanol Sensitivity and Rapid Ethanol Tolerance. *Alcohol Clin Exp Res* 33:1794–1805.

4. Devineni AV, Heberlein U (2009) Preferential Ethanol Consumption in Drosophila Models Features of Addiction. *Current Biology* 19:2126–2132.

5. Schumann G et al. (2010) The IMAGEN study: reinforcement-related behaviour in normal brain function and psychopathology. *Mol Psychiatry* 15:1128–1139.

6. O'Doherty J et al. (2004) Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* 304:452–454.

7. Bach FR, Jordan MI (2003) Kernel independent component analysis. *J Machine Learn Res* 3:1–48.

## SUPPLEMENTARY DATA

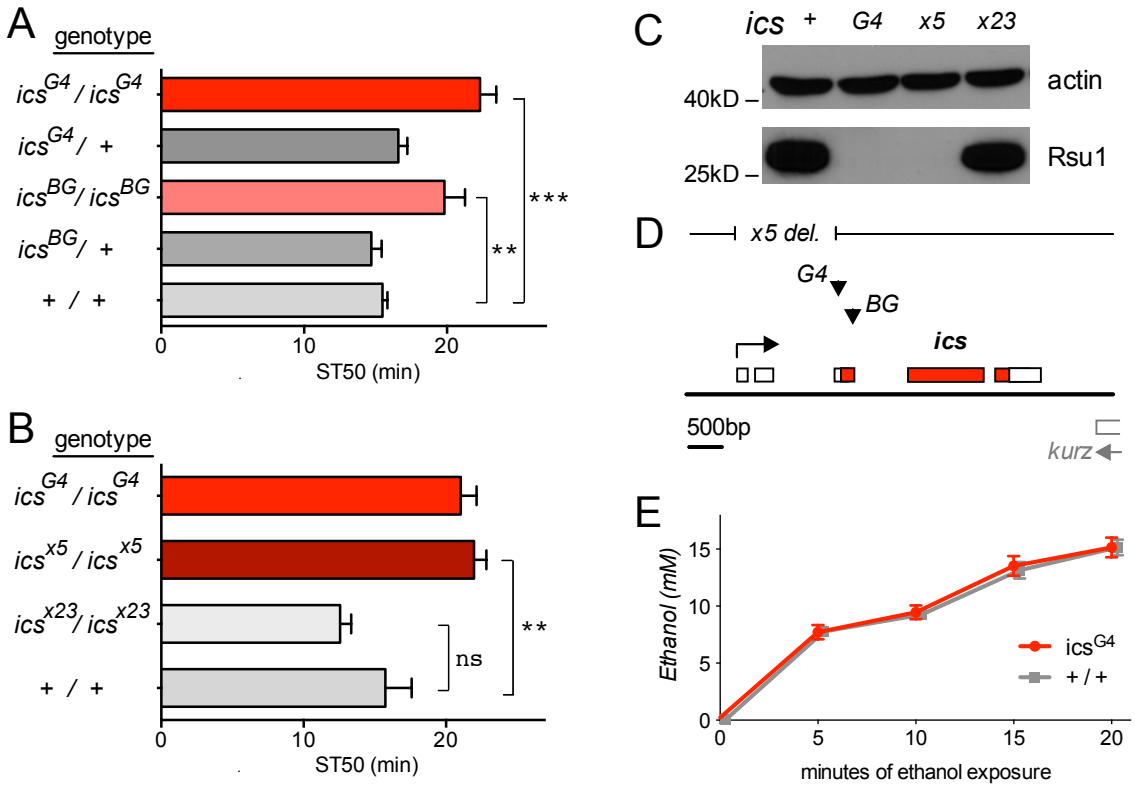



**Fig. S1.** *icarus*, encoding Rsu1, is required for normal ethanol responses. In this, and the following graph, bars represent means ±SEM, flies were exposed to 130/20 ethanol/air flow rate, and the median sedation time (ST-50) was determined as described (1). (*A*) *ics* homozygous mutant flies ($ics^{BG}$ and $ics^{G4}$) are resistant to ethanol-induced sedation when compared to wild type (***$p < 0.001$, n = 10-17, one-way ANOVA with Dunnett's multiple comparison post-hoc test). (*B*) Precise excision of the $ics^{G4}$ P-element, $ics^{x23}$, reverts the sedation resistance back to wild-type levels, while an imprecise excision, resulting in a 1.4 kbp deletion, $ics^{x5}$, retains the ethanol-resistance phenotype (***$p < 0.001$, n = 8-9). (*C*) Western blot analysis showing that Rsu1 protein expression is absent in *ics* mutants ($ics^{G4}$ and $ics^{x5}$), but present in the precise excision

(*ics*$^{x23}$) when compared to controls (*ics*$^{+}$). A representative blot of 3 repeats is shown. (*D*) Schematic representation of the *ics* locus, with exons as boxes, and the open reading frame in red. P-element insertion sites are represented by triangles, and the imprecise excision *ics*$^{x5}$ is depicted by the interruption in the line atop (*x5 del.*). (*E*) *ics* mutants have normal ethanol absorption and metabolism. Flies were exposed to 150:0 ethanol/air, flash frozen, and their internal ethanol concentration was measured. Two-way ANOVA indicates significant ethanol increase over exposure time ($p < 0.001$, n = 4 per genotype), but no effect of genotype ($p > 0.71$)

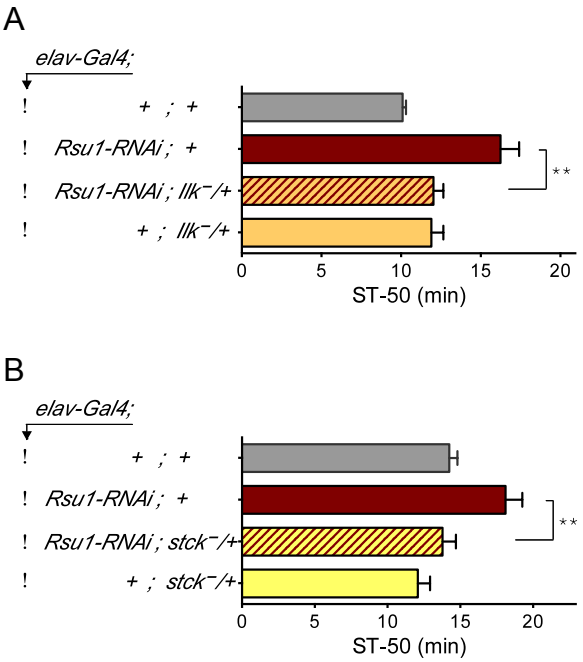



**Fig. S2.** Rsu1 functions in a pathway with PINCH and ILK, and opposes their action. (*A, B*) Flies were exposed to 130/20 ethanol/air and their ST-50 determined. Expressing *UAS-Rsu1-RNAi* in the nervous system, using elav-Gal4 as a driver leads to resistance. Introducing a heterozygous null allele of *Ilk* (encoding integrin linked kinase, (*A*), or *steamer duck* (*stck*, encoding Drosophila PINCH, (*B*) reduces the severity of Rsu1-RNAi resistance, while having no effect on their own. Alleles used were $Ilk^{1}$ and $stck^{3R-17}$.

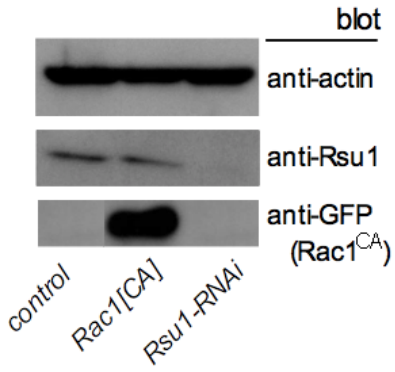



**Fig. S3.** Cell culture expression control blots. Western blot controls showing expression of the indicated proteins from Figs. 4*D, E*. All blots are representative examples of at least 3 replicates.

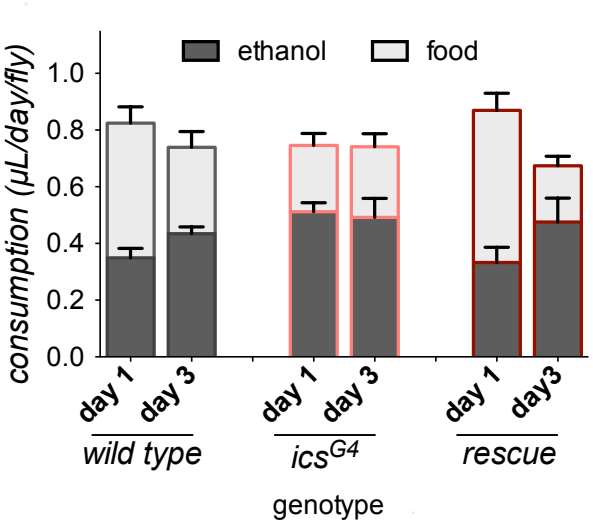



**Fig. S4.** *ics* affects alcohol consumption. Amounts of liquid food with, and without ethanol consumed are shown. Wild-type and (*ics*$^{G4}$*;UAS-Rsu1*) rescue flies shown an increase in ethanol, and decrease in food consumption from day 1 to day 3, while *ics*$^{G4}$ mutants consume high amounts of alcohol from day 1. Total amounts consumed were no different over the days or genotypes (or their interaction; two-way ANOVA, $F < 2.2$, $p > 0.14$), while *ics*$^{G4}$ consumed more ethanol on day 1 than the wild-type control (one-way ANOVA with Dunnett's multiple comparison, *ics* vs. *wt*: $p < 0.01$, $q = 3.16$, n = 20-29 groups of 8 flies per genotype).

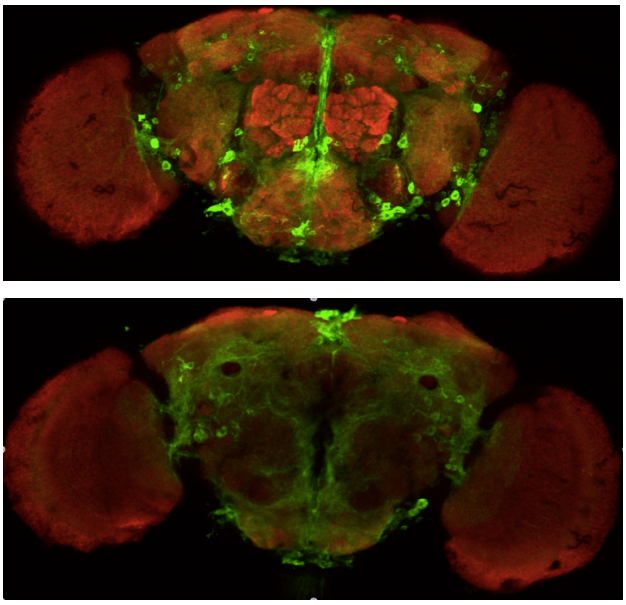


**Fig. S5.** *MB-Gal80* suppresses mushroom body expression in *ics*$^{G4}$ *MB-Gal80/+;UAS-mCD8-GFP* flies. Anterior (top) and posterior (bottom) stacks are shown.

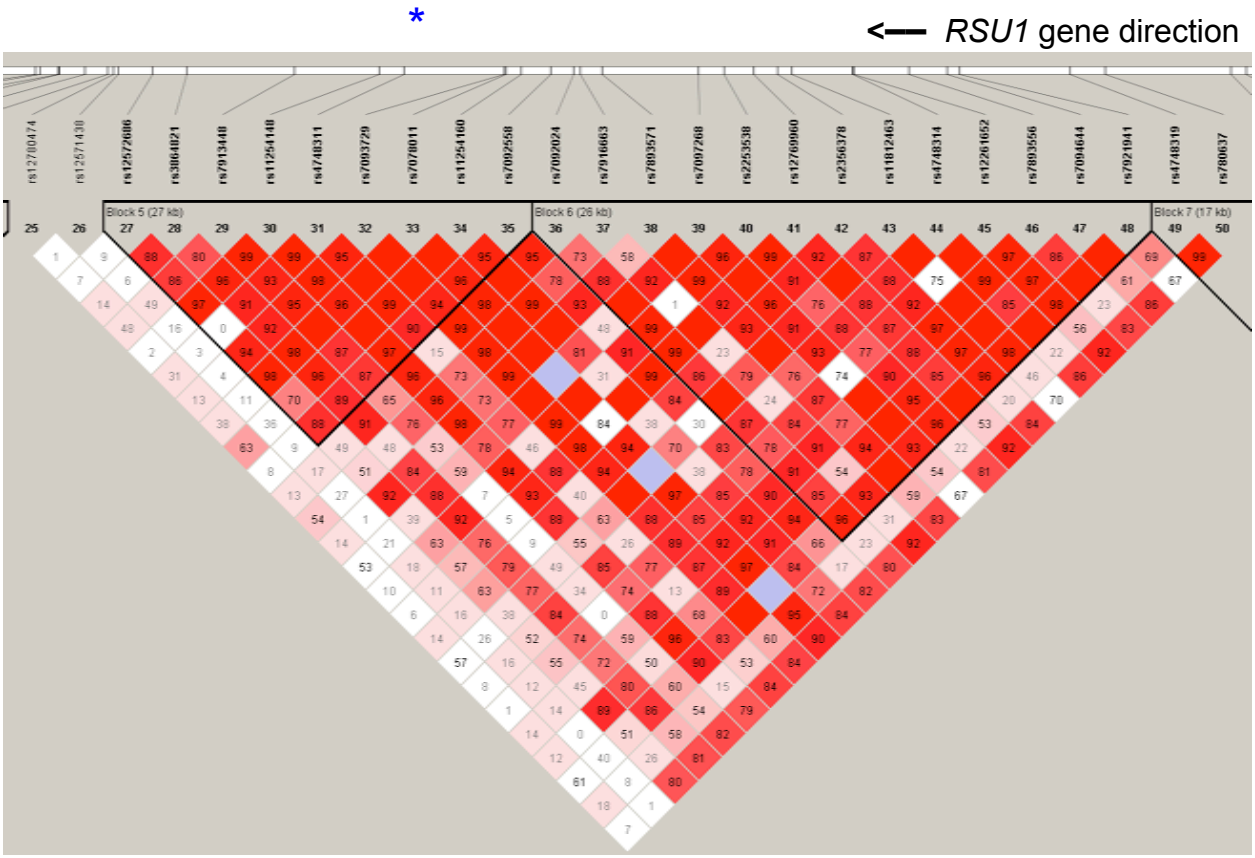
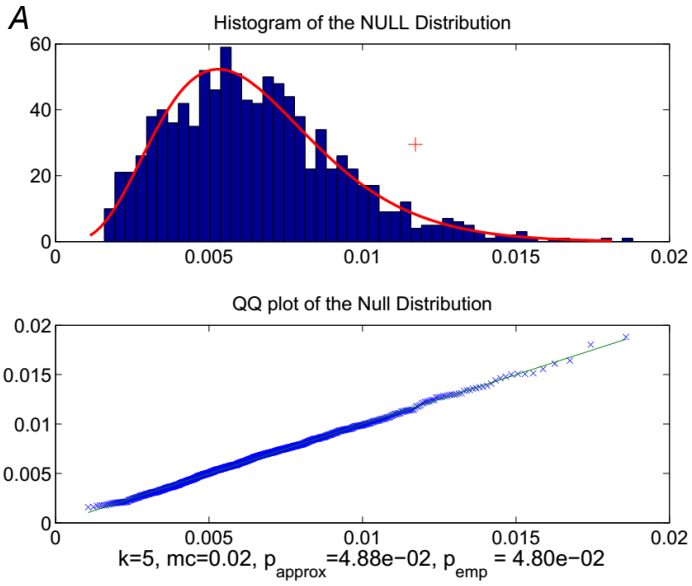
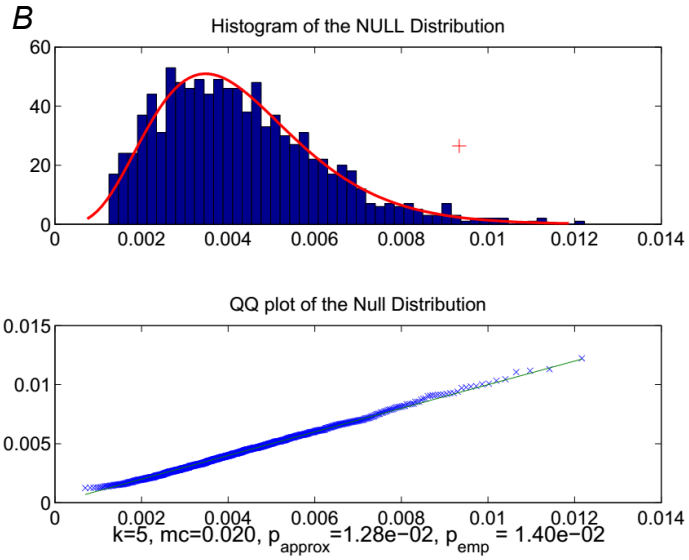



**Fig. S6.** Linkage disequilibrium structure of haplotype blocks 5 and 6 of human *RSU1* around SNP rs7078011 (asterisk atop), encompassing the 22 SNPs grouped in the kernel analysis, where both the adjusted linkage disequilibrium (scales as number) and the R-square (scales as color) are shown. The haplotype blocks are defined through the 'solid spine of LD' with threshold 0.80.


A
Histogram of the NULL Distribution
60
40
20
0
0
0.005
0.01
0.015
0.02
QQ plot of the Null Distribution
0.02
0.015
0.01
0.005
0
0
0.005
0.01
0.015
0.02
k=5, mc=0.02, p_approx=4.88e−02, p_emp = 4.80e−02


B
Histogram of the NULL Distribution
60
40
20
0
0
0.002
0.004
0.006
0.008
0.01
0.012
0.014
QQ plot of the Null Distribution
0.015
0.01
0.005
0
0
0.002
0.004
0.006
0.008
0.01
0.012
0.014
k=5, mc=0.020, p_approx=1.28e−02, p_emp = 1.40e−02

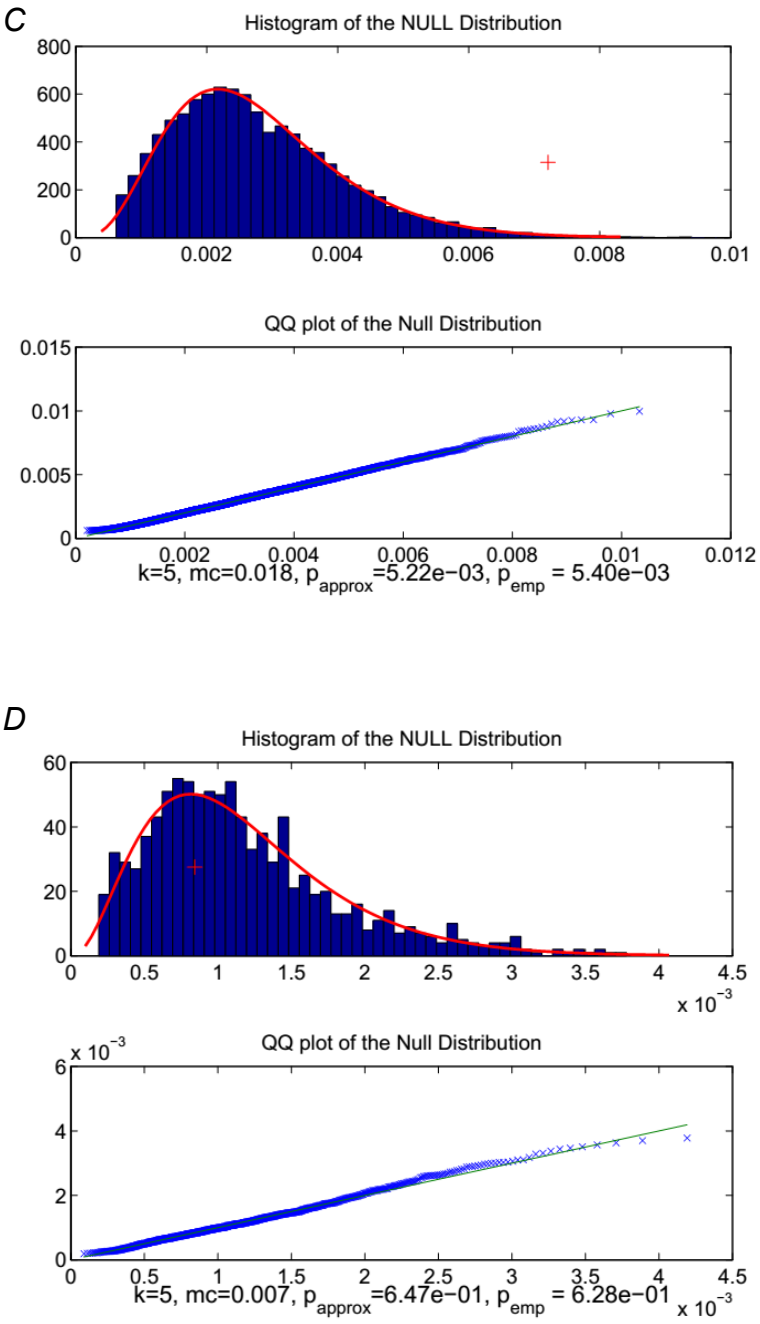
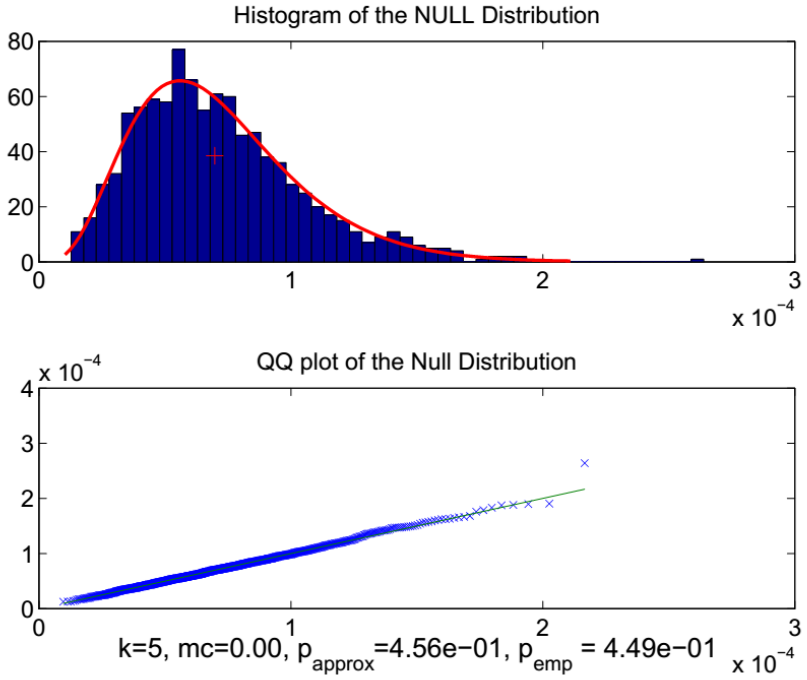


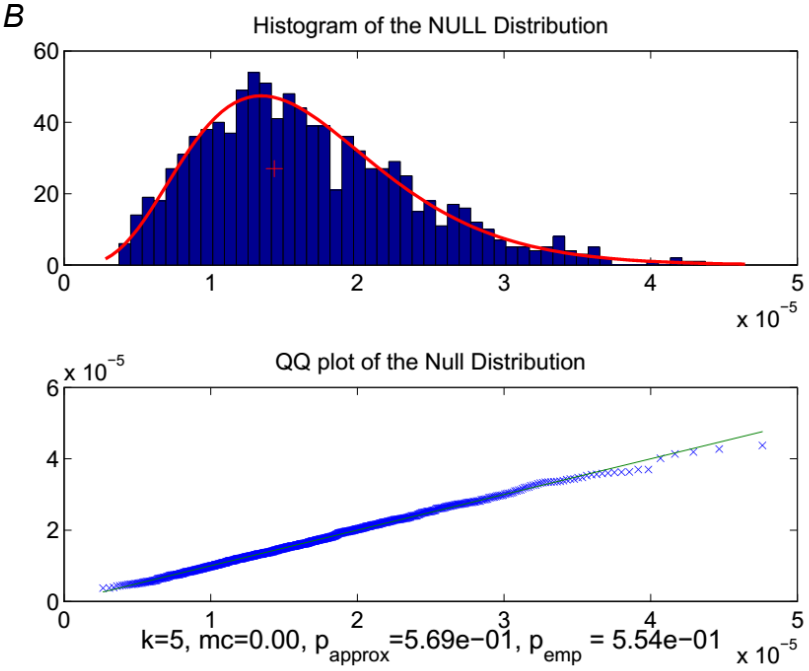

**Fig. S7.** Kernel-based associations of *RSU1* SNPs. (*A-D*) Associations in kernel-based analyses between *RSU1* SNPs and ventral striatum activation in the MID task (*A*), lifetime alcohol consumption frequency in the IMAGEN sample at 14 years old (*B*), alcohol dependence in SAGE Caucasian sample (*C*) and alcohol consumption in the

NFBC sample at 31 years old (*D*). In the histograms (top), the empirical distributions of statistics (column bars) from 1000 or 10000 permutations were plotted along with its theoretical gamma distributions (red lines), and the observed statistics were plotted as red crosses. The further those statistics are from the median, the smaller the observed p-value. The empirical and theoretical distributions were plotted against each other as the Q-Q plot (bottom), where the match between the dots (the observed quantile ratios) and the hard line (the expected quantile ratios) suggests that the observed p-values based on the theoretical distribution was reliable.

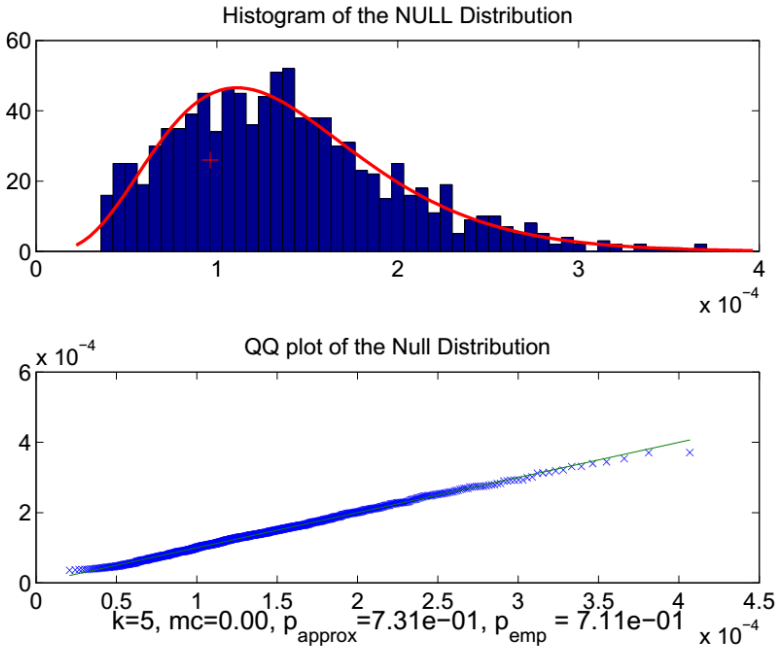
*A*


Histogram of the NULL Distribution
80
60
40
20
0
0
1
2
3
x 10^-4
QQ plot of the Null Distribution
x 10^-4
4
3
2
1
0
0
1
2
3
x 10^-4
k=5, mc=0.00, p_approx=4.56e−01, p_emp = 4.49e−01

*B*


Histogram of the NULL Distribution
60
40
20
0
0
1
2
3
4
5
x 10^-5
QQ plot of the Null Distribution
x 10^-5
6
4
2
0
0
1
2
3
4
5
x 10^-5
k=5, mc=0.00, p_approx=5.69e−01, p_emp = 5.54e−01

*C*




*D*

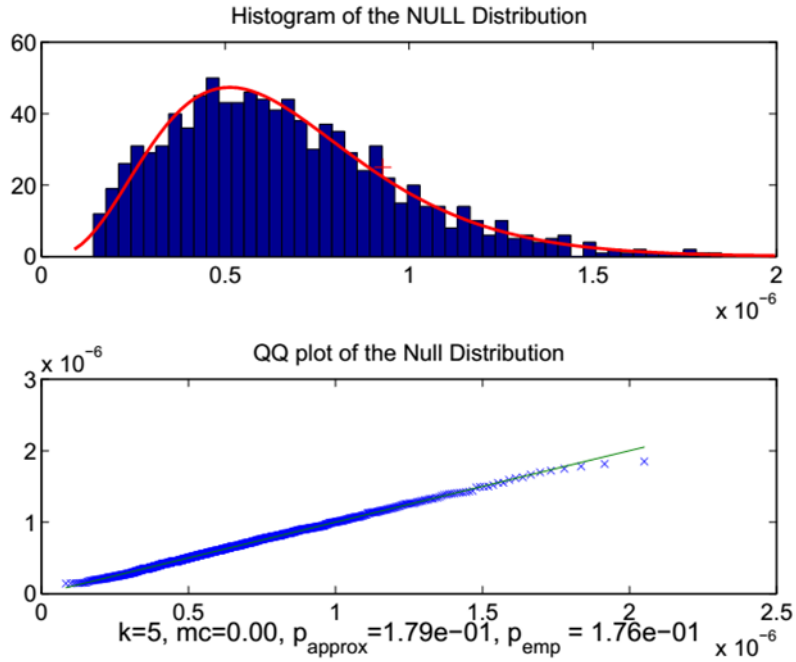



**Fig. S8.** Kernel-based Associations of *RSU1* SNPs with diffusion tensor imaging data. Results for kernel based association analyses between *RSU1* SNPs and ventral striatum (VS) grey matter volume (*A*), fractional anisotropy measures of diffusion tensor

imaging in fornix crescent (*B*), fornix body (*C*) and VS (*D*). In the histograms (top), the empirical distributions of statistics (column bars) from 1000 permutations were plotted along with their theoretical gamma distributions (red lines), and the observed statistics were plotted as the red crosses. The empirical and theoretical distributions were plotted against each other as the Q-Q plot (bottom), where a match between the dots (the observed quantile ratios) and the line (the expected quantile ratios) suggests that the observed p-values based on the theoretical distribution is reliable.

| Miss_Sense SNP | Gene | NCBI Reference | Location_on_Protein | AA1 | AA2 | Probility of Disruption |
|---|---|---|---|---|---|---|
| rs375646999 | RSU1 isoform 1 | NP_036557.1 | 208 | K | E | 0.028 |
| rs144428707 | RSU1 isoform 1 | NP_036557.1 | 230 | V | L | 0.001 |
| rs372364335 | RSU1 isoform 1 | NP_036557.1 | 238 | R | H | 1 |
| rs375416941 | RSU1 isoform 1 | NP_036557.1 | 242 | Y | C | 1 |
| rs375646999 | RSU1 isoform 2 | NP_689937.2 | 155 | K | E | 0.608 |
| rs144428707 | RSU1 isoform 2 | NP_689937.2 | 177 | V | L | 0.001 |
| rs372364335 | RSU1 isoform 2 | NP_689937.2 | 185 | R | H | 0.999 |
| rs375416941 | RSU1 isoform 2 | NP_689937.2 | 189 | Y | C | 1 |

**Table S1.** Predicted function of missense SNPs of 8th exon of RSU1 from PolyPhen2.

| Study | N (% Women) | Age in years (SD) | | Drinking Phenotypes | Phenotype Descriptions | |
|---|---|---|---|---|---|---|
| | | Men | Women | | Men | Women |
| IMAGEN | 1908 (52.0) | 14.4 (0.4) | 14.4 (0.4) | Lifetime Frequency | mean=2.0<br>*stdev=1.8<br>median=2.0 | mean=2.1<br>stdev=1.7<br>median=2.0 |
| NFBC 1966 | 4604 (51.0) | 31.2 (0.4) | 31.2 (0.4) | Quantity (grams/day) | mean=13.8<br>stdev=19.7<br>median=7.8 | mean=4.8<br>stdev=7.9<br>median=2.2 |
| SAGE (Caucasian) | 2509 (56.5) | 38.6 (10.3) | 38.4 (9.1) | Dependence vs Control | case=688<br>control=404 | case=461<br>control=956 |

*stdev stands for standard deviation.

**Table S2.** Phenotypic characteristics of alcohol related behaviors in human datasets.

| | Frequencies | | | Comparison in Chi-square | | |
|---|---|---|---|---|---|---|
| | IMAGEN (N) | SAGE (N) | NFBC (N) | IMAGEN vs SAGE | IMAGEN vs NFBC | SAGE vs NFBC |
| rs12572686_A | 0.440 (1966) | 0.427 (2544) | 0.432 (4772) | 0.80 | 0.36 | 0.19 |
| rs3864821_T | 0.338 (1965) | 0.327 (2544) | 0.278 (4772) | 0.62 | 23.96 | 18.96 |
| rs7913448_C | 0.129 (1967) | 0.128 (2543) | 0.214 (4772) | 0.01 | 64.77 | 80.52 |
| rs11254148_C | 0.386 (1966) | 0.386 (2537) | 0.395 (4772) | 0.00 | 0.46 | 0.47 |
| rs4748311_A | 0.268 (1967) | 0.276 (2543) | 0.356 (4772) | 0.36 | 48.59 | 47.88 |
| rs7093729_A | 0.309 (1963) | 0.306 (2544) | 0.263 (4772) | 0.04 | 14.86 | 15.67 |
| rs7078011_T | 0.041 (1967) | 0.048 (2542) | 0.029 (4772) | 1.32 | 6.02 | 17.05 |
| rs11254160_A | 0.093 (1967) | 0.091 (2544) | 0.133 (4772) | 0.02 | 21.65 | 27.78 |
| rs7092558_G | 0.404 (1966) | 0.395 (2543) | 0.375 (4772) | 0.40 | 4.98 | 2.76 |
| rs7092024_G | 0.487 (1964) | 0.500 (2542) | NA | 0.67 | NA | NA |
| rs7916663_A | 0.137 (1960) | NA | 0.180 (4772) | NA | 18.38 | NA |
| rs7893571_G | 0.332 (1963) | 0.327 (2533) | 0.293 (4772) | 0.14 | 10.17 | 9.08 |
| rs7097268_G | 0.032 (1964) | 0.031 (2544) | 0.027 (4772) | 0.02 | 1.21 | 1.05 |
| rs2253538_A | 0.450 (1967) | 0.439 (2541) | 0.468 (4772) | 0.56 | 1.77 | 5.61 |
| rs12769960_A | 0.154 (1941) | 0.178 (2544) | 0.179 (4772) | 4.38 | 5.71 | 0.01 |
| rs2356378_A | 0.471 (1967) | 0.468 (2543) | 0.475 (4772) | 0.04 | 0.10 | 0.34 |
| rs11812463_G | 0.034 (1963) | 0.032 (2544) | 0.080 (4772) | 0.19 | 46.82 | 64.80 |
| rs4748314_T | 0.478 (1967) | 0.473 (2544) | 0.489 (4772) | 0.08 | 0.72 | 1.64 |
| rs12261652_G | 0.163 (1966) | 0.179 (2541) | 0.197 (4772) | 2.05 | 10.57 | 3.38 |
| rs7893556_T | 0.460 (1966) | 0.473 (2544) | 0.453 (4772) | 0.78 | 0.25 | 2.66 |
| rs7094644_G | 0.320 (1967) | 0.316 (2541) | 0.294 (4772) | 0.10 | 4.37 | 3.57 |
| rs7921941_G | 0.029 (1961) | 0.029 (2543) | 0.026 (4772) | 0.01 | 0.47 | 0.41 |
| Sum-up Chi-square Statistics and P-values | | | | $\chi^2_{df=21}$ = 12.59 P=0.922 | $\chi^2_{df=21}$ = 286.19 P=2.03x$10^{-48}$ | $\chi^2_{df=20}$ = 303.80 P=1.35x$10^{-52}$ |

**Table S3.** Comparison of SNP frequencies among three datasets.

For each pair of SNP frequency comparison (right), the chi-square statistic with 1 degree of freedom (df) was calculated as the square of two sample t-statistic between the SNP frequencies by definition. Under the null hypothesis, as chi-square statistics of each pair of datasets are identical and independent distributed, their sum-up will follow chi-square test with degrees of freedom equal to the number of SNPs in comparison.

| | Haplotype Phase | IMAGEN | | NFBC | | SAGE - Caucasian | |
|---|---|---|---|---|---|---|---|
| | | Freq | P | Freq | P | Freq | P |
| Hap1 | GGTCCGCGG | 0.373 | 0.269 | 0.316 | 0.141 | 0.377 | 0.274 |
| Hap2 | ATTACACGT | 0.247 | 0.697 | 0.224 | 0.727 | 0.235 | 0.0903 |
| Hap3 | AGCAAGCAT | 0.072 | 0.661 | 0.024 | 0.112 | 0.073 | 0.00271 |
| Hap4 | GGTAAGCGT | 0.125 | 0.0343 | 0.018 | 0.0360 | 0.133 | 0.0856 |
| Hap5 | ATTACATGT | 0.041 | 0.768 | 0.028 | 0.144 | 0.046 | 0.979 |

**Table S4.** Haplotype analysis of RSU1 gene with alcohol related behaviours in human. Top five most frequent haplotype block 5 phases are included in the analysis in IMAGEN sample, and then re-validate in NFBC and SAGE-Caucasian samples. The frequency of each haplotype phase and its corresponding P-value are summarized for each dataset. SNPs included in the haplotype blocks are rs12572686, rs3864821, rs7913448, rs11254148, rs4748311, rs7093729, rs7078011, rs11254160 and rs7092558 as indicated as haplotype block 5 in Fig. S6.