A low-frequency inactivating *AKT2* variant enriched in the Finnish population is associated with fasting insulin levels and type 2 diabetes risk.

Short title: AKT2 coding variant affects fasting insulin levels

SUPPLEMENTARY MATERIAL

Table of Contents

Ethics Statements

All human research was approved by the relevant institutional review boards, and conducted according to the Declaration of Helsinki and all patients provided written informed consent. FIN-D2D 2007, DPS, DR's EXTRA, FINRISK 2007, FUSION, and METSIM were approved by the University of Michigan Health Sciences and Behavioral Sciences Institutional Review Board (ID: H03-00001613-R2). The Danish studies (Health 2006, Inter99, and Vejle Biobank) were approved by the local Ethical Committees of Capital Region (approval # H-3-2012-155, KA 98155 and KA-20060011) and Region of Southern Denmark (approval # S-20080097). The GoDARTS study was approved by EoS REC 09/S1402/44. The Twins UK study was approved by EC04/015. The OBB study was approved by South Central, Oxford C, 08/H0606/107+5, IRAS project 136602. The PIVUS study is approved by 00–419 and ULSAM study by 251/90 and 2007/338. The PPP study was approved by the Committee On the Use of Humans as Experimental Subjects at MIT (IRB 0912003615). T2D-GENES and GoT2D exome sequencing was approved by local institutional review boards. The study protocol of the Health 2000 survey was approved by the Epidemiology Ethics Committee of the Hospital District of Helsinki and Uusimaa. All participants gave signed informed consent. The YFS study was approved by local ethics committees. The HBCS study was approved by the Ethics Committee of Hospital District of Helsinki and Uusimaa and conducted according to the guidelines in the Declaration of Helsinki. The EuroBATS study was approved by St Thomas' Hospital Research Ethics Committee (ref. EC04/015).

Additional Acknowledgements

Individuals (in author order):

Alisa K. Manning was supported by American Diabetes Association grant #7-12-MN-02. Taru Tukiainen was supported by Orion-Farmos Research Foundation and the Finnish Cultural Foundation. Manuel A Rivas received the NDM Prize Studentship, Clarendon Award. Tune H Pers is supported by the Benzon Foundation and the Lundbeck Foundation.

Ana Viñuela has been funded by the EU FP7 grant EuroBATS (Grant No. 259749).

Andrew Anand Brown has been funded by the EU FP7 grant EuroBATS (Grant No. 259749) and by the South East Norway Health Authority (Grant No. 2011060). Eric. R. Gazamon was supported by NIH Grants for GTEx: R01 MH101820 and MH090937. Hae Kyung Im was in part funded by R01MH107666 and K12CA139160, and travel was funded by P30DK020595. John R B Perry was supported by the Sir Henry Wellcome Postdoctoral Fellowship. Martijn van de Bunt is supported by

the NDM Prize Studentship. Martin Hrabe de Angelis was supported by the German Center for Diabetes Research (DZD). Reedik Magi was supported by the Estonian Research Council (grant IUT20-60), the Development Fund of the University of Tartu (grant SP1GVARENG), EU structural support through Archimedes Foundation, grant no: 3.2.1001.11-0033, EU 7FP grant 278913, and H2020 grants 633589, 676550, 654248. Panos Deloukas's work forms part of the research themes contributing to the translational research portfolio of Barts Cardiovascular Biomedical Research Unit, which is supported and funded by the National Institute for Health Research. Katharine R Owen is a NIHR Clinician Scientist. Andrew Farmer is a NIHR Senior Investigator. Gilean McVean is a Wellcome Trust Senior Investigator. Eleftheria Zeggini was supported by The Wellcome Trust (098051). Heikki A. Koistinen has received funding from Academy of Finland (support for clinical research careers, grant no 258753). Veikko Salomaa is funded by the Finnish Foundation for Cardiovascular Research and the Academy of Finland (grant # 139635). Andrew P Morris is a Wellcome Trust Senior Fellow in Basic Biomedical Science (under award WT098017). Fredrik Karpe is supported by NIHR Oxford Biomedical Research Centre and NIHR National Bioresource. Graeme I. Bell was supported by NIH P30DK020595 (for genotyping, and analysis). James B. Meigs was supported by National Institute for Diabetes and Digestive and Kidney Diseases (NIDDK) R01 DK078616, NIDDK K24 DK080140. Mark I McCarthy is a Wellcome Trust Senior Investigator and a NIHR Senior Investigator. Anna L Gloyn is a Wellcome Trust Senior Fellow in Basic Biomedical Science. Cecilia Lindgren is supported in part by Wellcome Trust (WT086596/Z/08/A and 086596/Z/08/Z) and the Li Ka Shing Foundation.

Study and Cohort Acknowledgements

Funding for the GoT2D and T2D-GENES studies was provided by grants: NIH U01s DK085526, DK085501, DK085524, DK085545, and DK085584 (Multiethnic Study of Type 2 Diabetes Genes) and DK088389 (Low-Pass Sequencing and High-Density SNP Genotyping for Type 2 Diabetes). The work at the University of Oxford, UK was supported by the European Commission (ENGAGE: HEALTH-F4-2007-201413; Marie-Curie Fellowship PIEF-GA-2012-329156), MRC (G0601261, G0900747-91070), National Institutes of Health (RC2-DK088389, DK085545, DK098032), and Wellcome Trust (064890, 083948, 085475, 086596, 090367, 090532, 092447, 095101, 095552, 098017, 098381). The work at the Wellcome Trust Sanger Institute, UK was supported by the National Institute for Health Research and the Wellcome Trust (098051).

The Jackson Heart Study is supported by contracts HHSN268201300046C, HHSN268201300047C, HHSN268201300048C, HHSN268201300049C, HHSN268201300050C from the National Heart, Lung, and Blood Institute and the National Institute on Minority Health and Health Disparities.

The work at Wake Forest School of Medicine (WFSM) was supported by NIH grant R01 DK066358 (DWB).

The Korea Association Research Project was supported at Center for Genome Science, National Institute of Health, Republic of Korea by the Korea National Institute of Health (2012-N73002-00) and the Korea National Institute of Health and Korea Centers for Disease Control and Prevention (4845–301); and at Hallym University Chuncheon, Republic of Korea by theNational Research Foundation of Korea (NRF-2012R1A2A1A03006155). This study was provided with biospecimens and data from the Korean Genome Analysis Project (4845-301), the Korean Genome and Epidemiology Study (4851-302), and the Korea Biobank Project (4851-307, KBP-2013-11 and KBP-2014-68) that were supported by the Korea Centers for Disease Control and Prevention, Republic of Korea.

The work at the University of Texas Health Science Center at Houston, USA was supported by the National Institutes of Health (U01DK085501, R01HL102830, R01DK073541)

The work at Imperial College London, UK was supported by Action on Hearing Loss (G51), the British Heart Foundation (SP/04/002), European Union FP7 (EpiMigrant, 279143), Medical Research Council (G0601966, G0700931), MRC-PHE Centre for Environment and Health, The National Institute for Health Research (NIHR) (RP-PG-0407-10371), NIHR Biomedical Research Centre at Imperial College Health Care NHS Trust, NIHR Health Protection Research Unit on Health Impact of Environmental Hazards, and the Wellcome Trust (084723). Personal support includes Paul Elliot: NIHR Senior Investigator.

The LOLIPOP study is supported by the National Institute for Health Research (NIHR) Comprehensive Biomedical Research Centre Imperial College Healthcare NHS Trust. The work was carried out in part at the NIHR/Wellcome Trust Imperial Clinical Research Facility. We thank the participants and research staff who made the study possible.

The work at the National University of Singapore was supported by Biomedical Research Council (BMRC) Individual Research Gran, National Medical Research Council (NMRC) Individual Research Grant, NMRC Centre Grant. Personal support includes: Ching-Yu Cheng: NMRC Clinician Scientist award; E Shyong Tai: NMRC Clinician Scientist award; YY Teo: National Research Foundation Fellowship; TY Wong: NMRC Singapore Translational Research Investigator award.

The work at Helmholtz Zentrum München – German Research Center for Environmental Health, Germany was supported by The German Center for Diabetes Research (DZD), Helmholtz Zentrum München (German Research Center for Environmental Health), which is supported by the German Federal Ministry of Education and Research (BMBF) and by the State of Bavaria, and the Munich Center of Health Sciences (MC-Health), Ludwig-Maximilians-Universität, as part of LMUinnovativ.

The KORA research platform (KORA, Cooperative Research in the Region of Augsburg) was initiated and financed by the Helmholtz Zentrum München – German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research and by the State of Bavaria. Furthermore, KORA research was supported within the Munich Center of Health Sciences (MC Health), Ludwig-Maximilians-Universität, as part of LMUinnovativ.

The work at Lund University, Sweden was supported by the Academy of Finland, a European Research Council Advanced Research Grant, the Folkhälsan Research Foundation, Novo Nordisk, the Påhlssons Foundation, the Sigrid Juselius Foundation, the Skåne Regional Health Authority, the Swedish Heart-Lung Foundation, and the Swedish Research Council (Linné and Strategic Research Grant).

TwinsUK was funded by the Wellcome Trust; European Community's Seventh Framework Programme (FP7/2007-2013). The study also receives support from the National Institute for Health Research (NIHR) funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London. SNP Genotyping was performed by The Wellcome Trust Sanger Institute and National Eye Institute via NIH/CIDR.

Some computations were performed at the Vital-IT (http://www.vital-it.ch) Center for high-performance computing of the SIB Swiss Institute of Bioinformatics; and at the ACEnet, the regional

high performance computing consortium for universities in Atlantic Canada. ACEnet is funded by the Canada Foundation for Innovation (CFI), the Atlantic Canada Opportunities Agency (ACOA), and the provinces of Newfoundland and Labrador, Nova Scotia, and New Brunswick.

FIN-D2D was supported by financing from the hospital districts of Pirkanmaa, Southern Ostrobothnia, North Ostrobothnia, Central Finland, and Northern Savo; the Finnish National Public Health Institute; the Finnish Diabetes Association; the Ministry of Social Affairs and Health in Finland; Finland's Slottery Machine Association; the Academy of Finland (grant No. 129293) and Commission of the European Communities, Directorate C-Public Health (grant agreement No. 2004310) in cooperation with the FIN-D2D Study Group.

The Finnish DPS study was supported by the Academy of Finland (grants 128315, 129330, 131593).

The METSIM study was supported by the Academy of Finland (contract 124243), the Finnish Heart Foundation, the Finnish Diabetes Foundation, Tekes (contract 1510/31/06), and the Commission of the EuropeanCommunity(HEALTH-F2-2007-201681), and the US National Institutes of Health grants DK093757, DK072193, DK062370, and 1Z01 HG000024. Genotyping of the METSIM and DPS studies was conducted at the Genetic Resources Core Facility (GRCF) at the Johns Hopkins Institute of Genetic Medicine.

The DR's EXTRA Study was supported by grants to RR by the Ministry of Education and Culture of Finland (627;2004-2011), Academy of Finland (102318; 123885), Kuopio University Hospital, Finnish Diabetes Association, Finnish Heart Association, Päivikki and Sakari Sohlberg Foundation and by grants from European Commission FP6 Integrated Project (EXGENESIS); LSHM-CT-2004-005272, City of Kuopio and Social Insurance Institution of Finland (4/26/ 2010).

The National FINRISK 2007 study was supported by Finnish Foundation for Cardiovascular Research, the Academy of Finland (grant # 139635).

The FUSION study was supported by DK093757, DK072193, DK062370, and 1Z01 HG000024.

The Inter99 study Data collection in the Inter99 study was supported economically by The Danish Medical Research Council, The Danish Centre for Evaluation and Health Technology Assessment,

Novo Nordisk, Copenhagen County, The Danish Heart Foundation, The Danish Pharmaceutical Association, Augustinus foundation, Ib Henriksen foundation and Becket foundation. The Danish studies (Inter99, Health2006, and Vejle Biobank) were supported by the Lundbeck Foundation (Lundbeck Foundation Centre for Applied Medical Genomics in Personalised Disease Prediction, Prevention and Care (LuCamp); http://www. lucamp.org/) and the Danish Council for Independent Research. The Novo Nordisk Foundation Center for Basic Metabolic Research is an independent Research Center at the University of Copenhagen, partially funded by an unrestricted donation from the Novo Nordisk Foundation (http://www.metabol.ku.dk/).

GoDARTS study was funded by The Wellcome Trust Study Cohort Wellcome Trust Functional Genomics Grant (2004-2008) (Grant No: 072960/2/ 03/2) and The Wellcome Trust Scottish Health Informatics Programme (SHIP) (2009-2012) (Grant No: 086113/Z/08/Z). Analysis and genotyping of the British UK cohorts was supported by Wellcome Trust funding 090367, 098381, 090532, 083948, 085475, MRC (G0601261), EU (Framework 7) HEALTH-F4-2007-201413, and NIDDK DK098032.

TwinsUK study was funded by the Wellcome Trust; European Community's Seventh Framework Programme (FP7/2007–2013). The study also receives support from the National Institute for Health Research (NIHR) BioResource Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London.

The Oxford Biobank is supported by the Oxford Biomedical Research Centre and part of the National NIHR Bioresource.

The PIVUS/ULSAM cohort was supported by Wellcome Trust Grants WT098017, WT064890, WT090532, Uppsala University, Uppsala University Hospital, the Swedish Research Council and the Swedish Heart-Lung Foundation.

The Botnia study has been financially supported by grants from the Sigrid Juselius Foundation, Folkhälsan Research Foundation, Nordic Center of Excellence in Disease Genetics, an EU grant (EXGENESIS), Signe and Ane Gyllenberg Foundation, Swedish Cultural Foundation in Finland, Finnish Diabetes Research Foundation, Foundation for Life and Health in Finland, Finnish Medical Society, Paavo Nurmi Foundation, Helsinki University Central Hospital Research Foundation, Perklén Foundation, Ollqvist Foundation, Närpes Health Care Foundation and Ahokas Foundation. The study

has also been supported by the Ministry of Education in Finland, Municipal Heath Care Center and Hospital in Jakobstad and Health Care Centers in Vasa, Närpes and Korsholm.

The Cardiovascular Risk in Young Finns Study was financially supported by the Academy of Finland (grants 121584, 126925, 124282, and 129378), the Social Insurance Institution of Finland, the Turku University Foundation, special federal grants for University Hospitals, the Juho Vainio Foundation, Paavo Nurmi Foundation, the Finnish Foundation of Cardiovascular Research, Orion-Farmos Research Foundation, and the Finnish Cultural Foundation.
The Hesinki Birth Cohort Study was supported by Emil Aaltonen Foundation, Finnish Foundation for Diabetes Research, Novo Nordisk Foundation, Signe and Ane Gyllenberg Foundation, Samfundet Folkhälsan, Finska Läkaresällskapet, Liv och Hälsa, Finnish Foundation for Cardiovascular Research.

Additional Acknowledgements

We thank the High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics for the generation of array and sequencing data. The High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics is funded by a Wellcome Trust grant (reference 090532/Z/09/Z)

The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health. Additional funds were provided by the NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. Donors were enrolled at Biospecimen Source Sites funded by NCI\SAIC-Frederick, Inc. (SAIC-F) subcontracts to the National Disease Research Interchange (10XS170), Roswell Park Cancer Institute (10XS171), and Science Care, Inc. (X10S172). The Laboratory, Data Analysis, and Coordinating Center (LDACC) was funded through a contract (HHSN268201000029C) to The Broad Institute, Inc. Biorepository operations were funded through an SAIC-F subcontract to Van Andel Institute (10ST1035). Additional data repository and project management were provided by SAIC-F (HHSN261200800001E). The Brain Bank was supported by a supplements to University of Miami grants DA006227 & DA033684 and to contract N01MH000028. Statistical Methods development grants were made to the University of Geneva (MH090941 & MH101814), the University of Chicago (MH090951, MH090937, MH101820, MH101825), the University of North Carolina – Chapel Hill (MH090936 & MH101819), Harvard University (MH090948), Stanford University (MH101782), Washington University St Louis (MH101810), and the

University of Pennsylvania (MH101822). The data used for the analyses described in this manuscript were obtained from dbGaP (accession number phs000424.v3.p1).

Funding support for “Building on GWAS for NHLBI-diseases: the U.S. CHARGE consortium” was provided by the NIH through the American Recovery and Reinvestment Act of 2009 (ARRA) (5RC2HL102419). Sequence data for “Building on GWAS for NHLBI-diseases: the U.S. CHARGE consortium” was provided by Eric Boerwinkle on behalf of the Atherosclerosis Risk in Communities (ARIC) Study, L. Adrienne Cupples, principal investigator for the Framingham Heart Study, and Bruce Psaty, principal investigator for the Cardiovascular Health Study. Sequencing was carried out at the Baylor Genome Center (U54 HG003273). Further support came from HL120393, “Rare variants and NHLBI traits in deeply phenotyped cohorts” (Bruce Psaty, principal investigator). Supporting funding was also provided by NHLBI with the CHARGE infrastructure grant HL105756.

Supplementary Notes

SUPPLEMENTARY NOTE 1: SUMMARY OF ASSOCIATION RESULTS AT KNOWN AND NOVEL LOCI.

The exome-wide single variant association results are displayed in **Supplementary Table 2**. We first partitioned the significant ($P < 5 \times 10^{-7}$) and suggestive ($P < 5 \times 10^{-6}$) single variant association results into two sets: variants in previously reported associated regions [**Supplementary Table 2A**] and variants with potentially novel association signals [**Supplementary Table 2B**].

Of the 57 loci with common variants associated with FG or FI in multiple ancestries[4,5,10,58,72-80], twenty-one regions contained significant or suggestive association signals in our analysis. Of the seven regions harboring significant associations with non-synonymous variants, five (*GCKR, G6PC2, SLC30A8, PCSK1*, and *GLP1R*) were described previously by our group[10], where, when possible, conditional analyses and functional experiments are utilized to illuminate functional transcripts. In the *MADD* locus, a missense variant *ACP2* p.Arg29Gln showed significant association with FG levels ($P = 1.91 \times 10^{-7}$, MAF = 38%). This variant is in low LD ($r^2 = 0.138$) with the reported variant, rs7944584 ($P = 2.62 \times 10^{-11}$, MAF = 39%), but after conditioning on rs7944584 the association was not significant (P = 0.003). An additional association with a low-frequency variant was observed at the *MTNR1B* locus. A variant upstream of *MTNR1B*, rs7950811, (effect = 0.057; $P = 6.8 \times 10^{-11}$), has a MAF of 4.5% and in low LD with the index SNP, rs10830963 ($r^2 = 0.002$), in 1000 Genomes data[121]. After conditioning on the index SNP, the association of rs7950811 with FG remained significant ($P = 3.07 \times 10^{-7}$). For FI, five regions contained significant or suggestive association signals. All of the insulin-associated variants were common with MAF > 25%. Two of these regions, the *GCKR* and *GRB14/COBLL1* loci, harbor significant missense variants and were previously described[10].

Association results at previously reported variants from genome-wide association studies are presented in **Supplementary Table 2C**. Of the 68 previously published common variant associations with FG and FI, we were able to carry out association tests at 36 FG and 16 FI variants. Thirty of the FG association loci showed P < 0.05, with 100 % having a consistent direction of effect. Thirteen FI associated loci had P < 0.05, with 100% demonstrating a consistent direction of effect.

Potentially novel association signals

We observed five and seven variants passing suggestive level of significance for FI and FG, respectively **[Supplementary Table 2B]**. As this analysis focused on coding variation, we took the three coding variants forward to a replication analysis in four independent Finnish studies (N = 5,747)[82-85]. The p.Pro50Thr variant in *AKT2* was present and well-imputed in the 1000 Genomes reference panel (imputation score: 0.886 to 0.957). The correlation between imputed and directly genotyped genotypes was high ($r^2 > 0.88$), and the association of this variant with FI levels replicated, ($P_{replication} = 0.00054$, N = 5,747) resulting in a combined (discovery and replication) sample P value of 9.98×10^{-10} [**Supplementary Table 2E**]. *MMEL1* p.Glu323Gln, which has a MAF of only 0.2% (seven minor allele carriers in the HBCS subset), was poorly imputed and not tested for

association (imputation score: 0.718 to 0.945, r^2 = 0.57). *TP53BP1* p.Thr1278Ile was not observed in the studies.

Summary of exome-wide significant gene based association results

The suggestive and significant gene based association signals from each ancestry group in the exome sequencing data and the exome chip data, as well as combined results, are displayed in **Supplementary Table 2D**. The *AKT2* gene based association with FI is described in the main text.

In gene-based tests using the PTV+NS_{broad} mask, *NDUFAF1* was significantly associated with FI levels (P_{Burden} = 1.10×10^{-6}). This association was driven by a single missense variant (p.His309Asp, rs199599633, P = 9.3×10^{-5}, N = 1,673) that was not associated with FI levels in exome array data (P = 0.018, N = 19,569). NADH dehydrogenase (ubiquinone) complex I, assembly factor 1, or *NDUFAF1*, encodes for a complex I assembly factor protein, which is part of the first step of the respiratory chain. Mutations in both copies of this gene are reported to cause mitochondrial complex I deficiency, which manifests as cardioenphalomypathy or fatal hypertrophic cardiomyopathy while heterozygous parents were reported as healthy[122,123].

Additionally, a third gene, *GIMAP8*, was associated with FG levels in the PTV-only mask (P_{Burden} = 2.30×10^{-6}). This association was driven by singleton and doubleton variants. This gene encodes a GTPase of the immunity-associated protein family[124]

SUPPLEMENTARY NOTE 2: EXPRESSION PROFILE OF *AKT2*

To gain further insights into the tissues relevant for AKT2 function we explored gene and transcript expression patterns of *AKT2* (ENSG00000105221) from multiple (N = 44) human tissues using RNA sequencing (RNA-seq) data from the Genotype Tissue Expression (GTEx) Project[108].

In the GTEx data *AKT2* is ubiquitously expressed **[Supplementary Fig. 10A,B]**; the gene is present in all the available tissues (median expression across individuals RPKM[125] (reads per kb per million reads) > 7 in all tissues, [**Supplementary Table 4**] and in all individuals, in agreement with previous studies examining *AKT2* expression via RT-PCR, Western blot, and Northern Blot analysis[37,38,126,127], and documented essential role of AKT isoforms in biological processes throughout the body[39]. No enrichment of *AKT2* expression is present in insulin sensitive tissues (i.e. pancreas, skeletal muscle, adipose tissue (both subcutaneous and visceral), liver and kidney cortex) via RNA sequencing as proposed in mouse and rat models, however, this is consistent with previous examination of *AKT2* mRNA in human tissues[38,126-128]. This GTEx RNA sequencing data does not address insulin-sensitive tissue enrichment seen at the level of AKT2 protein, yet in general mRNA levels correlate with protein abundance[129-131].

AKT2 has multiple alternatively spliced transcripts, yet little is known of their specific roles, and therefore we investigated which of the transcripts are the most abundant and which tissues these are active in Gencode version 12 used in the gene and transcript annotations lists 28 *AKT2* transcripts and 17 of these transcripts are expressed (mean RPKM > 1) in at least one of the studied tissues [**Supplementary Fig. 10C,D**]. However, majority of the expression appears to be due to three *AKT2* transcripts: *AKT2-004* (processed transcript) and *AKT2-001* (protein-coding) that span the full length of the gene, and *AKT2-008* (protein-coding), which does not include the downstream exons. Together these three transcripts constitute on average 44% (range 18-65%) of *AKT2* expression in the GTEx tissues. The two longer *AKT2* transcripts, *AKT2-004* and *AKT2-001*, follow similar expression pattern to the gene, while the shorter one, *AKT2-008*, shows more specific pattern of expression being most expressed in uterus, kidney cortex and esophagus mucosa.

The exon containing the p.Pro50Thr variant is included in 14 out of 28 expressed transcripts (all the 28 *AKT2* transcripts are expressed at a detectable level in at least one individual in at least one tissue), including in all the three most highly expressed transcripts [**Supplementary Fig. 10D**]. The expression profile of the exon containing p.Pro50Thr is similar to the whole *AKT2* gene with the tissues showing highest *AKT2* expression generally having the higher levels of expression of the exon containing p.Pro50Thr [**Supplementary Fig. 10B**]**.** Notably, the exon is expressed in all tissues and all individuals, further suggesting that the exon likely encodes part of the protein integral for its function.

Similarly to *AKT2*, the two other members of the *AKT* gene family, *AKT1* and *AKT3*, are expressed in all the tissues available in the GTEx data with the exception of rather low expression of *AKT3* in liver and whole blood. Of the three genes, *AKT1* is generally the most and *AKT3* the least abundant in all tissues. *AKT2* is the

most highly expressed of the three homologs ($P < 0.05$ for all comparisons using one-sided paired Student's t-test and log2 transformed expression values) only in skeletal muscle, pituitary and cerebellum/cerebellar hemisphere, with the higher *AKT2* expression being most pronounced in skeletal muscle [**Supplementary Fig. 11]**.

SUPPLEMENTARY NOTE 3: PATHWAY ANALYSES

Methods

We used biological knowledge to test for enrichment of signal in pathways. Pathways and networks were selected from MSigDB[132], which includes Gene Ontology, pathways from KEGG, Ingenuity, Reactome, and Biocarta; and the manually curated monogenic pathways previously considered. We carried out a two-stage enrichment analysis: step one calculates gene aggregation scores using a function of single variant statistics; and step two calculates gene set scores using a function of aggregation scores from each gene in the set. In step one, we make use of a range of gene aggregation functions, including the minimum p-value (or maximum Bayes' factor) for single-variant association (within ancestry or trans-ethnic) in the gene (with correction for the number of variants in the gene). In step two, we apply a pre-ranked GSEA method[132], which consists of a sensitive-improved Kolmogorov-Smirnov (random bridge) statistic, and which provides better correction of the null distribution for highly correlated gene sets (as we see for our hand curated gene sets). Additionally, we performed a biologically enhanced pathway analyses with DEPICT[133], an integrative tool that we used to highlight enriched pathways and identify tissues/cell types where genes from associated loci are highly expressed.

Gene set definitions: We assembled pre-defined, hand-curated lists to create four gene sets: "Monogenic All" (N = 81), including any gene with reported mutations that result in a disease or syndrome leading to either increased prevalence of diabetes or changes in glycemic traits. We further prioritized two subsets of genes, "Monogenic Glucose" (N = 41) and "Monogenic Insulin" (N = 37) including any gene with mutations leading to changes in respective glycemic traits as a primary feature. The list contains genes identified before September 2013. The fourth gene set, "Insulin Receptor Signaling," was created using Ingenuity Pathway Analysis (IPA) tools[134] by merging the insulin receptor signaling, IGF-1 signaling, and PI3K/AKT signaling pathways and adding all downstream phosphylated substrates of AKT.

Association Analysis: SKAT and burden tests were performed after aggregating functional variants (according to the previously described criteria) across all the genes in each gene set. Conditional analyses were performed using features implemented in RareMETALS[67,68].

Enrichment of association signals: Empirical enrichment for the number of gene based tests with $P < 0.001$ and the number of single variant tests with $P < 0.001$ in each gene set was determined by first counting the number of tests below the threshold. For a particular gene set, let $N_{observed}$ denote the number of tests with $P < 0.001$. A pool of similar genes was assigned to each gene in the gene set, according to the quartile of exon length and quintiles of the number of the nonsynonymous and synonymous variants in the gene. For each gene set, 1,000 matched gene sets were created. An empirical distribution of N_i (the number of tests with $P < 0.001$ in matched set i) was constructed for each of the matched sets. The empirical enrichment P-value was calculated by observing the proportion of matched sets with $N_i \geq N_{observed}$.

Additional traits related to insulin resistance: We examined the single variant association of fasting adiponectin level (log-transformed, age, sex and BMI adjusted, and inverse-normalized), 2 hour glucose level (age, sex and BMI-adjusted, and inverse-normalized) and 2 hour insulin level (log-transformed, age, sex and BMI adjusted, and inverse-normalized) in these pathways using exome array data when available from the discovery cohorts (D2D2007, DPS, DRSEXTRA, FINRISK, FUSION, Health2008, Inter99, METSIM, ULSAM).

Results

To further assess the evidence of enriched signals in biologically related genes, we looked for enrichment across pathways using both hand curated and publically available pathways. This was conducted using GSEA[132,135,136]. While no gene-set was significant after multiple testing correction, there is enrichment for several pathways, including adipocytokine signaling, glucose transport, galactose metabolism, glycolysis and gluconeogenesis, and starch and sucrose metabolism pathways, all of which include both *G6PC2* and *G6PC*. While the *G6PC2* association with FG has previously been described [10], we note that *G6PC* mutations result in glycogen storage disorders[137].

Since *AKT2* lies in the insulin receptor signaling pathway and *AKT2* mutations are a known cause of both familial lipodystophy, severe insulin resistance and hypoglycemia [23-26] we next explored whether there was an enrichment of rare and low frequency variants in these gene sets ("Monogenic Genes," and "Insulin Receptor Signaling Genes") [**Supplementary Table 8A**]. First, we tested for global enrichment by aggregating all variants predicted to be deleterious using the annotation masks previously described for gene based testing (PTV-only, PTV+NS_{strict}, PTV+NS_{broad}, PTV+Missense)[136]. We found a significant enrichment of deleterious variants (protein truncating, splice site and non-synonymous) in the monogenic genes ($P = 2 \times 10^{-4}$) in exome array data [**Supplementary Table 8B**) but no such enrichment in an analysis of the exome sequencing data set (P = 0.87) [**Supplementary Table 8C**]. Conditional analyses demonstrated that in addition to *AKT2* p.Pro50Thr (P conditional on *AKT2* p.Pro50Thr = 0.0017), seven additional top ranked variants contribute to this signal (P conditional on *AKT2* p.Pro50Thr, *CFTR* p.Asp1270Asn, *INSR* p.Val1012Met, *ZMPSTE24* p.Arg178His, *ZFP57* p.Arg178His, *CFTR* splice donor variant rs78756941 and *PCNT* p.Glu1785Lys jointly = 0.0104) [**Supplementary Table S8D,E**]. No other novel associations were detected with the other gene sets and variant masks, although when comparing the effects of the burden tests across the four variant aggregation categories, we observed a positive trend of effect as we examined the category containing the least predicted deleterious (PTV+missense) to the most predicted deleterious (PTV-only), although the confidence intervals widen as the number of included variants decrease [**Supplementary Fig. 13**].

To find specific genes harboring an enrichment of association with either FG or FI levels, we next focused on association results from the monogenic genes, testing each set for empirical enrichment. We found that a gene implicated in congenital generalized lipodystrophy, *CAV1*[138], showed enrichment of association with FG levels when considering the set of glucose-specific monogenic genes from the exome sequencing analysis (enrichment P = 0.03; *CAV1* $P = 1.9 \times 10^{-4}$ with protein truncating and low-frequency missense variants and P

= 7.0 × 10^{-4} with protein truncating and predicted deleterious variants). Mutations in *CAV1* are characterized by extreme insulin resistance and lipodystrophy [138] but in our data no association of *CAV1* variants with FI levels was observed. We also observed a borderline enrichment for fasting insulin level with a gene-based burden test in the insulin receptor signaling pathway (enrichment P = 0.06; (*PTGS2* burden P = 1.1 × 10^{-4} with protein truncating and low-frequency missense variants; [**Supplementary Fig. 14**, **Table S9A,B**).

We further examined the association of three quantitative traits related to insulin resistance: fasting adiponectin level, and 2 hour glucose and 2 hour insulin levels after an oral glucose tolerance test. Besides a nominally significance Other than the *AKT2* p.Pro50Thr allele association with 2 hour insulin level (Effect = 26% increase, 95% confidence interval = 16% - 38%, P = 7.86 × 10^{-8}), no other associations were observed [**Supplementary Fig. 14C**].

SUPPLEMENTARY NOTE 4: PRIMERS FOR FUNCTIONAL WORK

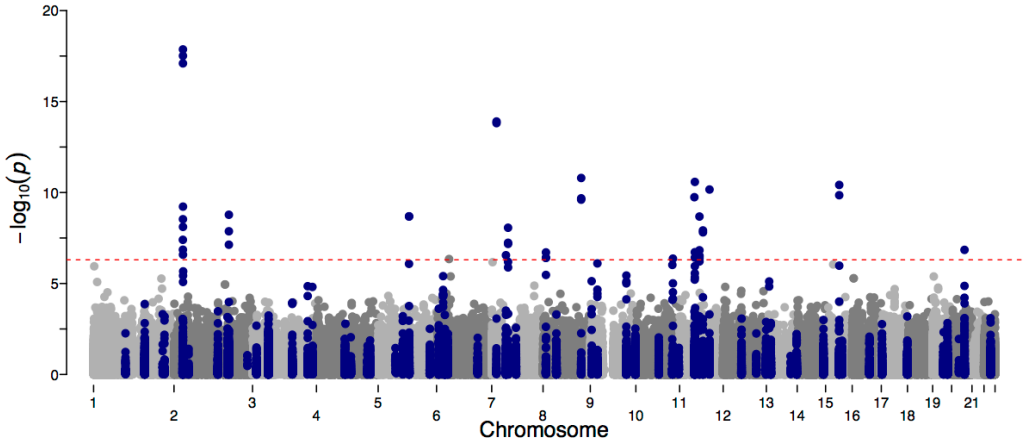
The generation of the AKT2 allelic series was initiated by the production of pDONR223- AKT2 through PCR of the human AKT2 open reading frame with the integration of terminal attR sites using primers FWD: 5’ -GGGGACAAGTTTGTACAAAAAAGTTGGCACCATGAATGAGGTGTCTGTCATC -3’ REV: 5’-GGGGACCACTTTGTACAAGAAAGTTGGCAACTCGCGGATGCTG -3’, and subsequent Gateway BP reaction into pDONR223 obtained from The Broad Institute Genetics Perturbation Platform. Site-directed mutagenesis was then performed to generate AKT2.E17K (AKT2.Lys17), AKT2.P50T (AKT2.Thr50), AKT2.R208K (AKT2.Lys208), AKT2.R274H (AKT2.His274), AKT2.R467W (AKT2.Trp467) with the following primers:

- AKT2.E17K: FWD: 5'- GGCTCCACAAGCGTGGTAAATACATCAAGACCTGG -3' REV: 5'-CCAGGTCTTGATGTATTTACCACGCTTGTGGAGCC -3'
- AKT2.P50T: FWD: 5'- AGGCCCCTGATCAGACTCTAACCCCCTTAAAC -3' REV: 5'-GTTTAAGGGGGTTAGAGTCTGATCAGGGGCCT -3'
- AKT2.R208K: FWD: 5'- GTCCTCCAGAACACCAAGCACCCGTTCC -3' REV: 5'-GGAACGGGTGCTTGGTGTTCTGGAGGAC -3'
- AKT2.R274H: FWD: 5'- GGGACGTGGTATACCACGACATCAAGCTGGA -3'REV3'REV: 5'-TCCAGCTTGATGTCGTGGTATACCACGTCCC -3'
- AKT2.R467W: FWD: 5'- GGAGCTGGACCAGTGGACCCACTTCCC -3' REV: 5'-GGGAAGTGGGTCCACTGGTCCAGCTCC -3'

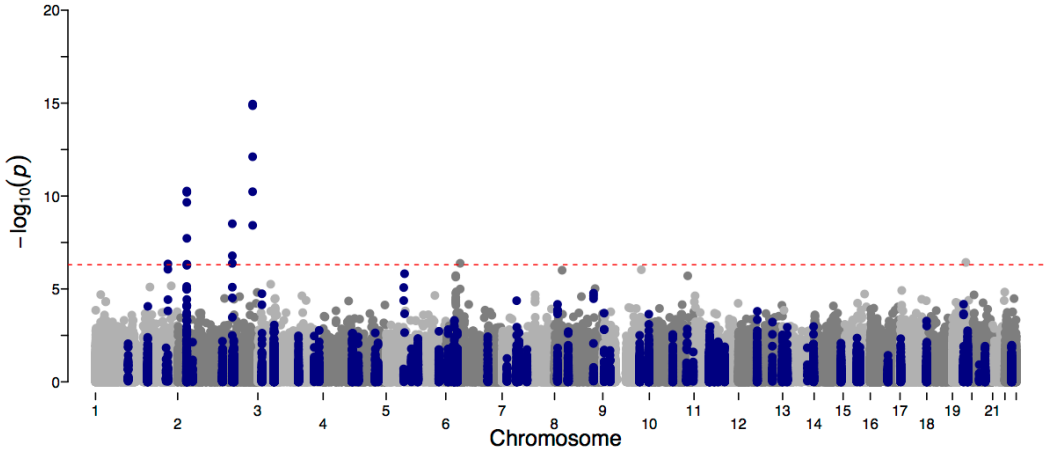
C-terminal, V5-tagged lentiviral pLX304-AKT2.E17K, pLX304-AKT2.P50T, pLX304- AKT2.R208K, pLX304-AKT2.R274H, and pLX304- AKT2.R467W were each generated by subsequent Gateway LR reactions with pDONR223-AKT2.E17K, pDONR223-AKT2.P50T, pDONR223-AKT2.R208K, pDONR223-AKT2.R274H, and pDONR223-AKT2.R467W, respectively, and pLX304 obtained from The Broad Institute Genetics Perturbation Platform. Control plasmid pLX304- empty vector was additionally acquired from The Broad Institute Genetics Perturbation Platform.

Supplementary Figures

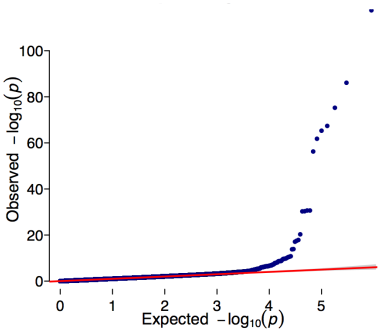
A. Fasting Plasma Glucose *



B. Fasting Insulin



C. Fasting Plasma Glucose



D. Fasting Insulin

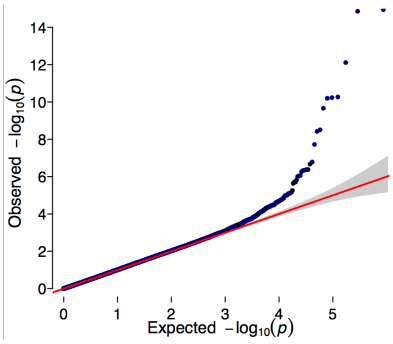
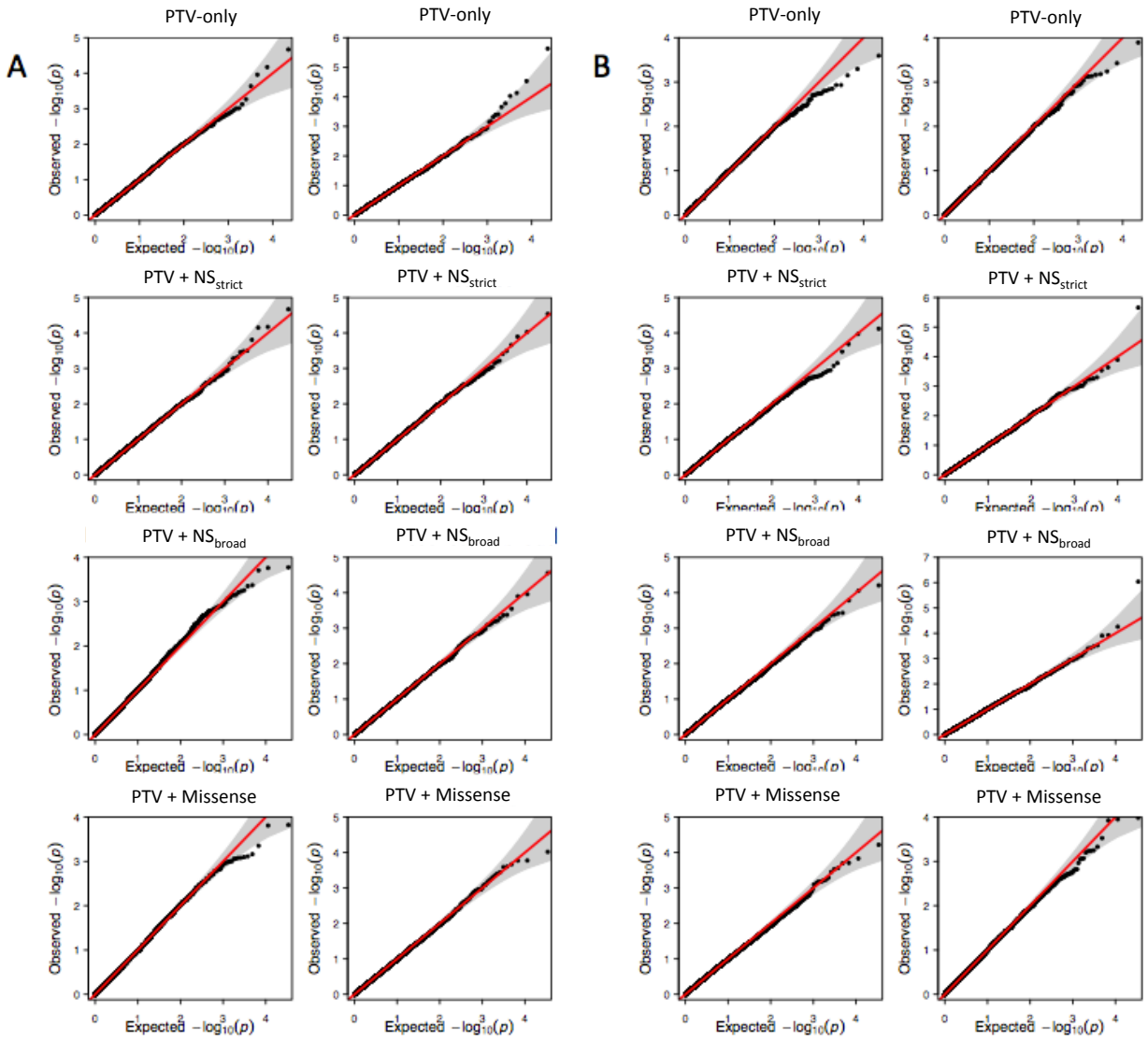


Fig. S1. Manhattan and quantile-quantile (QQ) plots for exome-wide association analysis with FG (A and C) and FI levels (B and D). A. Manhattan plot for FI, **B**. Manhattan plot for FG, C. QQ plot for FI, **D**. QQ plot for FI. On the manhattan plots, variants within regions of known association are colored in dark blue, and variants outside those regions are colored in gray. The red horizontal line in the manhattan plots represents the exome-wide significance threshold for single variant associations ($P < 2.5 \times 10^{-7}$). In the QQ plots the grey shaded area shows the 95% confidence interval. * For readability, the FG manhattan plot is truncated at –log10(P) = 20, although variants in the *G6PC2* region on chromosome 2 have –log10(P) values > 20.


A
B
PTV-only
PTV-only
PTV + NS$_{strict}$
PTV + NS$_{strict}$
PTV + NS$_{broad}$
PTV + NS$_{broad}$
PTV + Missense
PTV + Missense
Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$

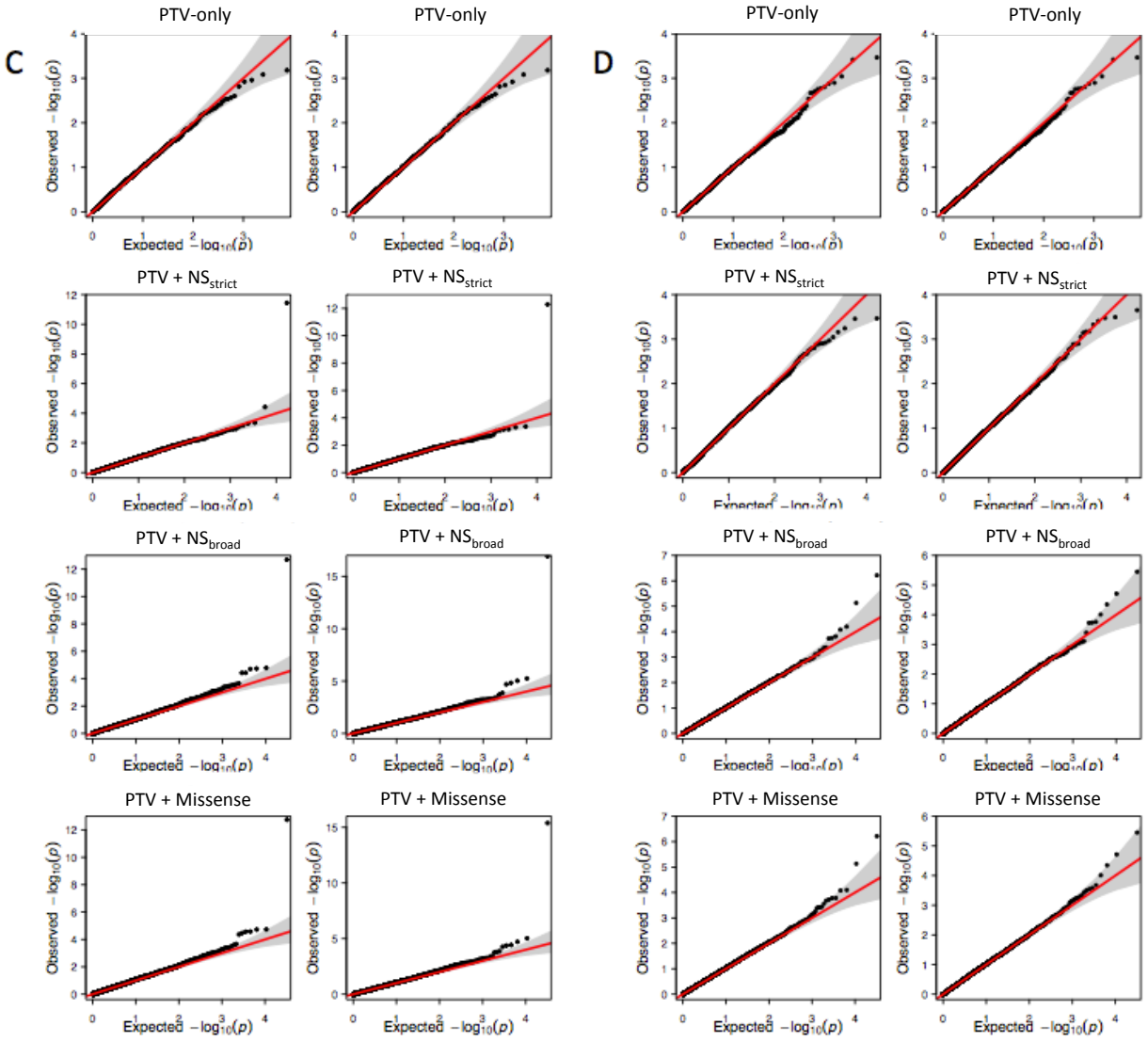


Fig S2. QQ plots from the gene based association tests for FI and FG. Two tests were applied, SKAT (left column) and Burden (right column) to four annotation masks (PTV, PTV+NS$_{Broad}$, PTV+NS$_{Strict}$, PTV+Missense, see **Methods** for description). **A.** FI with variants in exome sequencing data set. **B.** FG with variants in exome sequencing data set. **C.** FI with variants in exome chip data set. The point deviating from the diagonal is the association test for *AKT2*; see Table S2a for association details. **D.** FG with variants in exome chip data set.

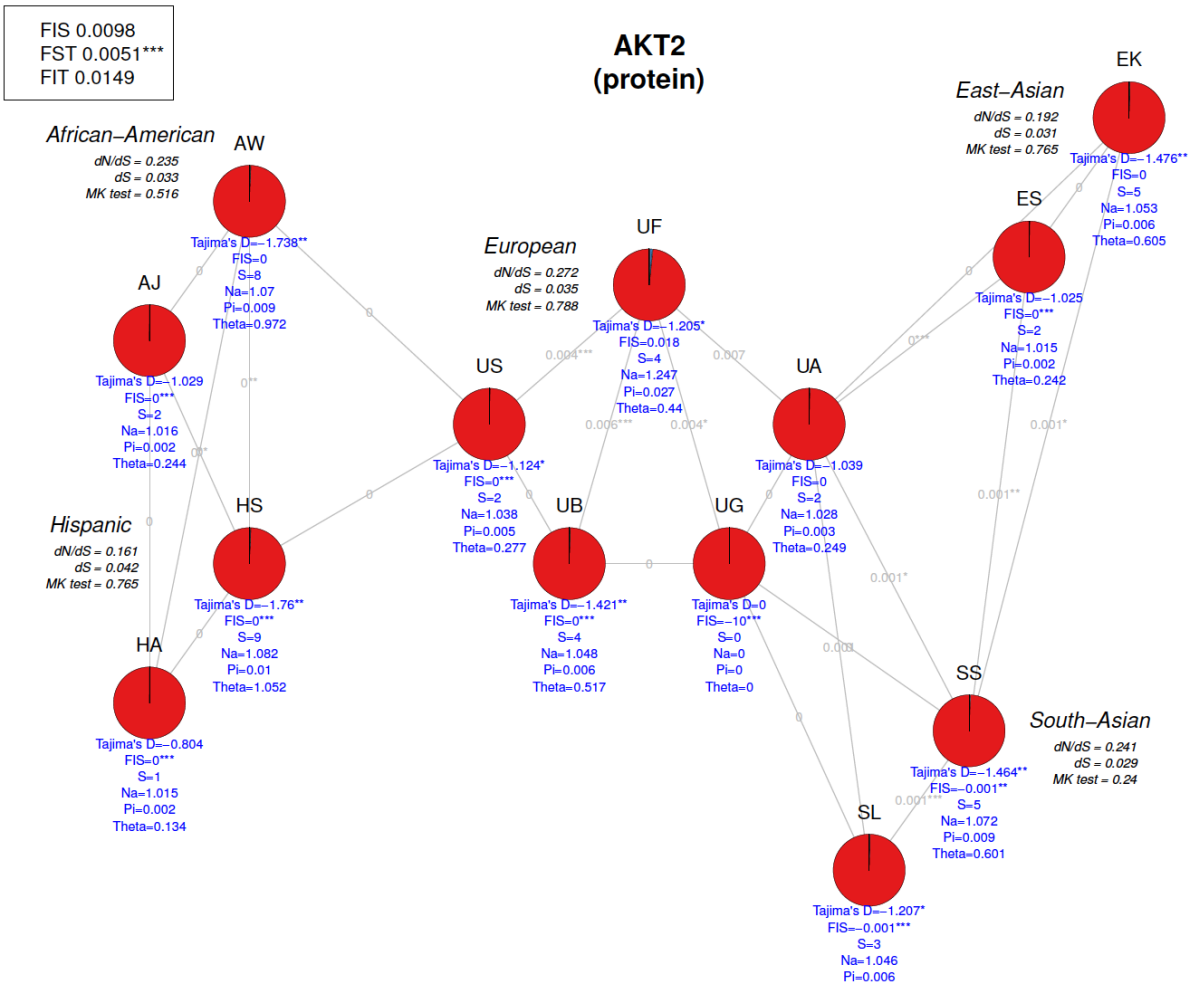


Fig. S3: Population structure and diversity indices of AKT2 protein in the exome sequencing data set. Each pie represents the frequency of different haplotypes, estimated from phased exome sequencing data in the five continental ancestries (grouped by study or country of origin). Significance of Tajima's D and F-statistics (global F_{ST}, F_{IS}, F_{IT}, and pairwise F_{ST} (gray line), and within population F_{IS}) are indicated with asterisk: * P-value < 0.05; ** P-value < 0.01; *** P-value < 0.001. S: Number of segregating sites; Na: expected number of alleles; Pi (π): Mean number of pairwise differences; Theta (θ): Watterson's θ estimate; MK: McDonald-Kreitman test. **African-American**: AJ – Jackson Heart Study, AW – Wake Forest School of Medicine Study; **East-Asian**: EK – Korea Association Research Project, ES – Singapore Diabetes Cohort Study and Singapore Prospective Study Program; **European**: UA – Ashkenazi (US, Israel), UB – UKT2D Consortium (UK) , UF (Finland) – Metabolic Syndrome in Men Study, Finland-United States Investigation of NIDDM Genetics (FUSION) Study, Malmo-Botnia Study, UG (Germany) – KORA-gen (Germany), US (Sweden) – Malmo-Botnia Study; **Hispanic**: HA – San Antonio Family Heart Study, San Antonio Family Diabetes/ Gallbladder Study, Veterans Administration Genetic Epidemiology Study, and the Investigation of Nephropathy and Diabetes Study family component, HS – Starr County, Texas; **South-Asian**: SL – London Life Sciences Population Study, SS – Singapore Indian Eye Study.

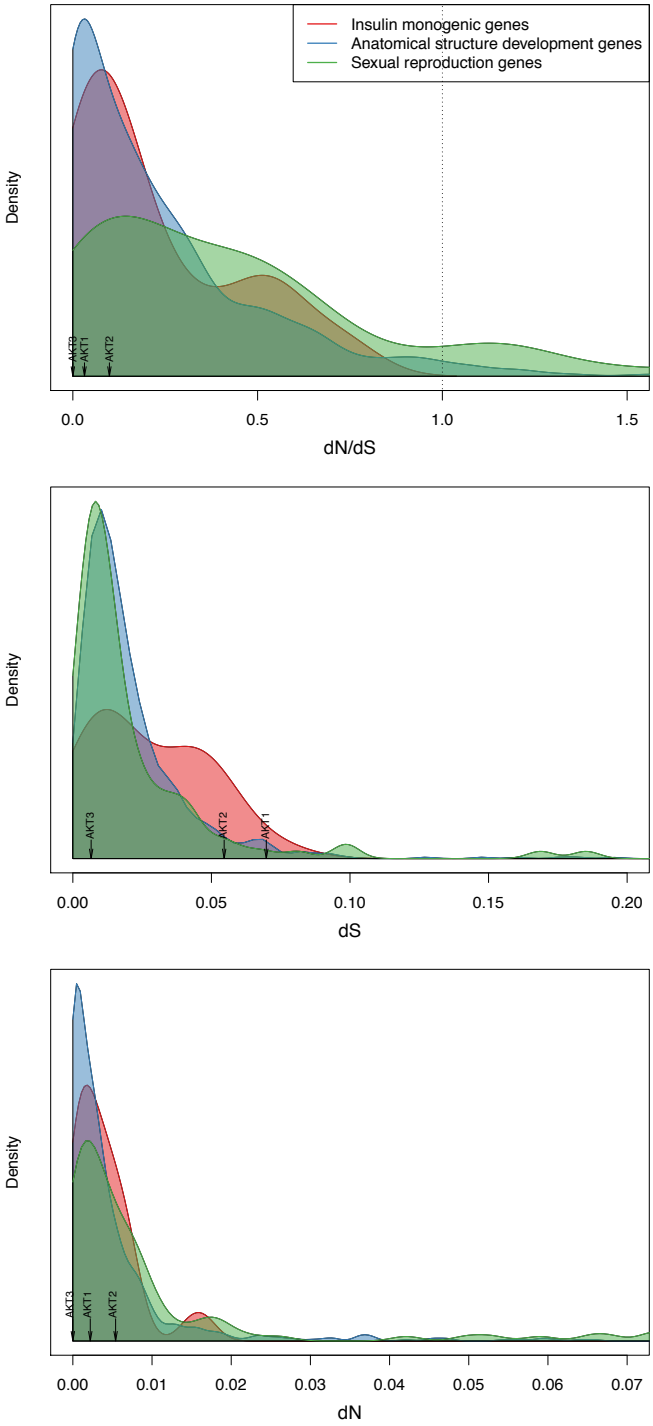


Fig. S4: *AKT* family conservation compared to other genes. The dN/dS ratio (ratio of the number of non-synonymous nucleotide substitutions per non-synonymous site and number of synonymous nucleotide substitutions per synonymous site) is calculated by comparing homologous coding sequences between human and chimpanzee. It indicates the degree to which selection is acting on a gene: ratio < 1 points to negative selection/purifying selection, i.e. evolutionary pressure to conserve the sequence in ancestral state, ratio > 1 to positive selection, and ratio = 1 to neutral evolution. The three *AKT* homologs, highlighted with arrows in the plot, are highly conserved when compared to the set of "Insulin monogenic" genes (37 genes), to which *AKT2* belongs, and two other gene sets: 1,002 anatomical structure development genes ("conserved"), and 132 sexual reproduction genes ("fast evolving").

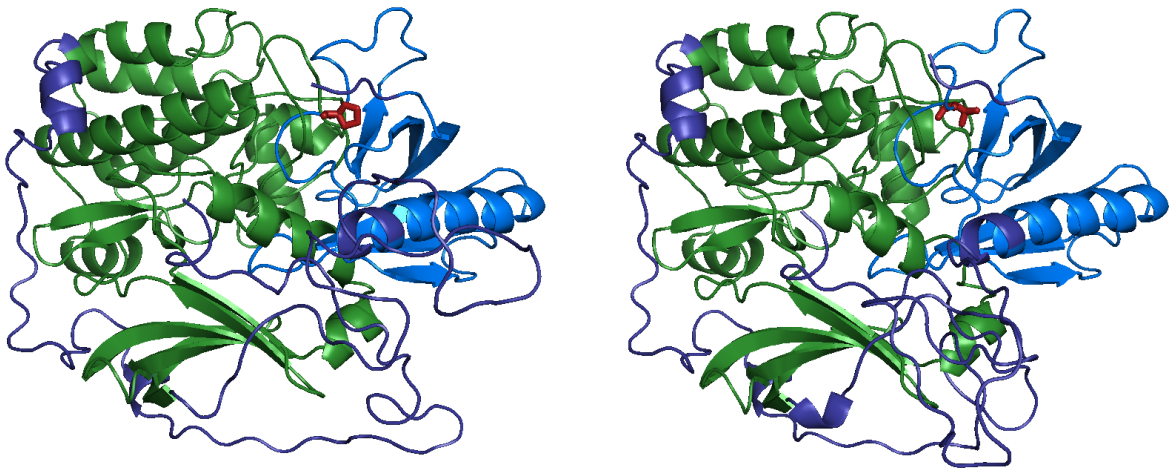


Fig S5. Predicted structure change in AKT2 due to *AKT2* p.Pro50Thr. The right plot shows the predicted structure of wild-type AKT2. The right plot shows the predicted structure of AKT2.Thr50.

A. General linear analysis

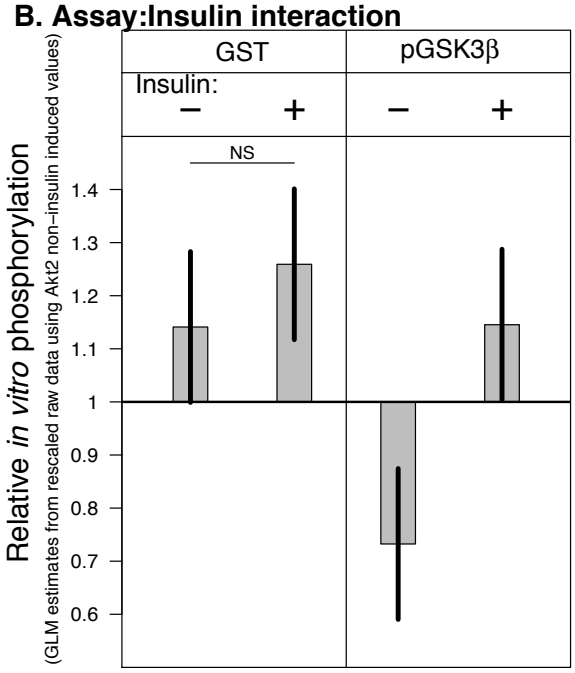
"Round" model:

Variables	*DF*	*Variance explained (%)*	*F*	*Pr(>F)*
Round	2	2.73%	1.228	0.300
Assay	1	8.42%	7.572	**0.008**
Insulin induction	1	12.38%	11.125	**0.001**
Round:Assay	2	1.60%	0.718	0.492
Round:Insulin	2	4.52%	2.033	0.140
Assay:Insulin	1	3.34%	2.999	0.088
Round:Assay:Insulin	2	0.27%	0.121	0.887

Full model:

Variables	*DF*	*Variance explained (%)*	*F*	*Pr(>F)*
Assay	1	8.42%	14.71	**3.12E-04**
Insulin induction	1	12.38%	21.61	**1.98E-05**
Variants	5	23.52%	8.21	**6.49E-06**
Assay:Insulin	1	3.34%	5.83	**1.90E-02**
Assay:Variant	5	19.13%	6.68	**5.64E-05**

B. Assay:Insulin interaction



C. Assay:Variants interaction

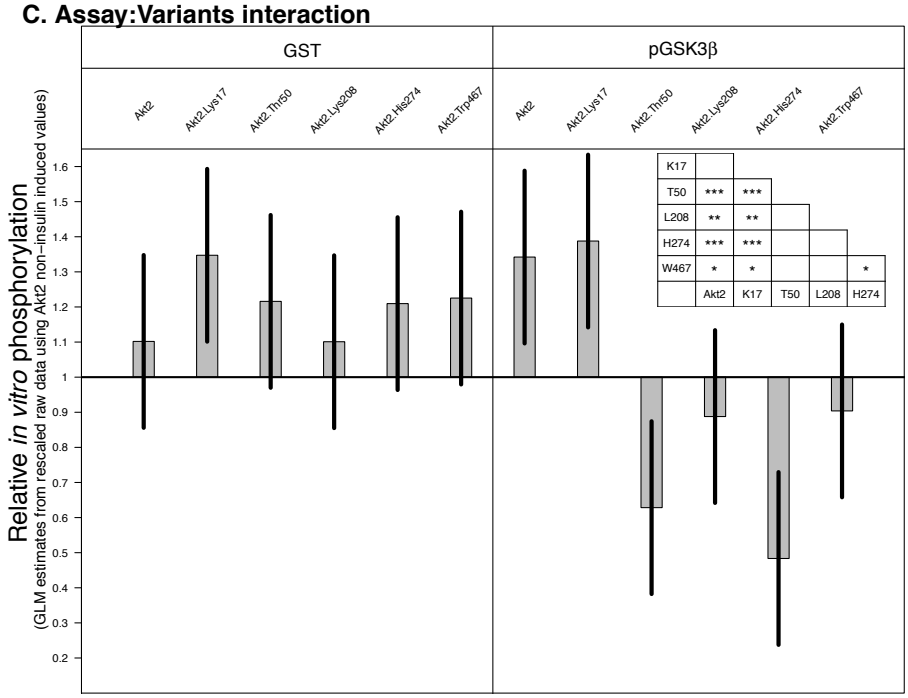
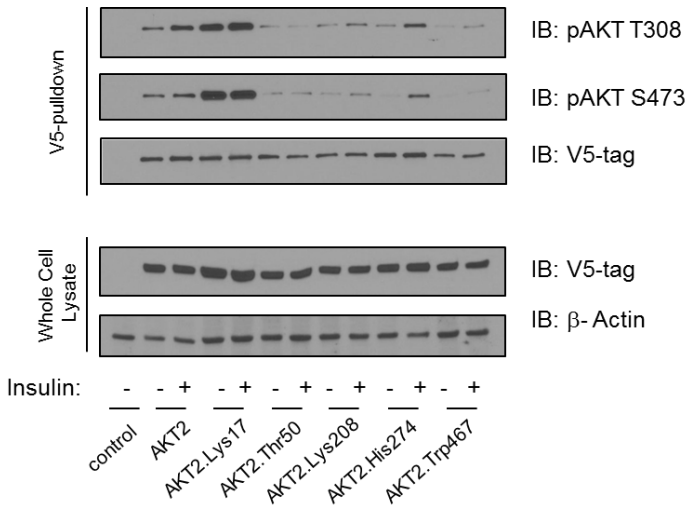
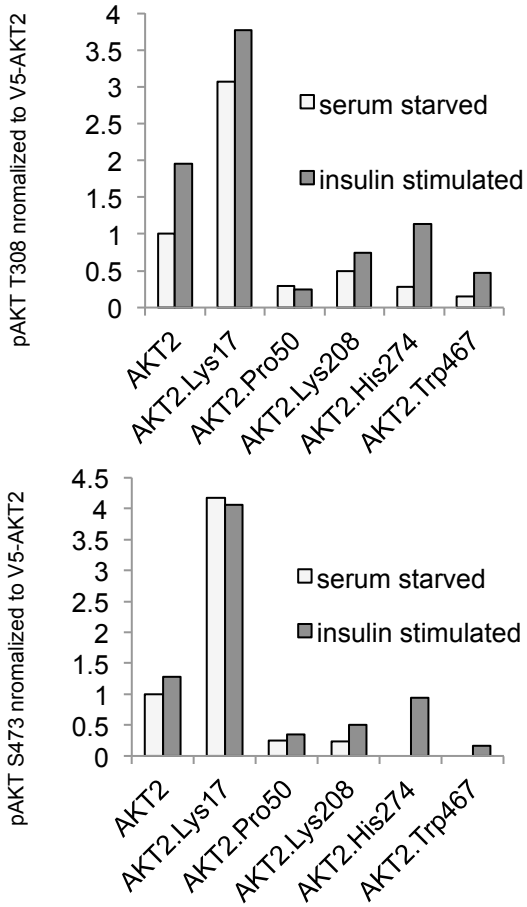


Fig. S6. In vitro kinase (IVK) assay. Results of a generalized linear model (GLM) applied on rescaled raw data. The relative substrate phosphorylation values were generated by dividing each value in each round of analysis with the value for non-stimulated, serum-starved AKT2. A first GLM ("Round" model) was analyzed including the Round as variable; the three independent rounds were not significant: we used them as replicate in the Full model. The plots represent the GLM estimates (and 95% CI) in the Full model for the two significant interactions: **A**. Assay:Insulin. **B**. Assay:Variants. **C**. For the Glycogen Synthase Kinase 3 b (GSK3b), the different AKT2 variants show significant relative phosphorylation (pairwise comparison p-values from contrast analysis reported in inset table). For GST-GSK3 peptide, none of the AKT2 variants showed different relative phosphorylation values. * P < 0.05, ** P < 0.01, *** P < 0.001. DF: degrees of freedom, F: statistic testing the importance of the grouping term, Pr(>F): P value of the F statistic.

A.


V5-pulldown
IB: pAKT T308
IB: pAKT S473
IB: V5-tag
Whole Cell Lysate
IB: V5-tag
IB: β- Actin
Insulin: - - + - + - + - + - + - +
control
AKT2
AKT2.Lys17
AKT2.Thr50
AKT2.Lys208
AKT2.His274
AKT2.Trp467

B.


pAKT T308 nromalized to V5-AKT2
4
3.5
3
2.5
2
1.5
1
0.5
0
serum starved
insulin stimulated
AKT2
AKT2.Lys17
AKT2.Pro50
AKT2.Lys208
AKT2.His274
AKT2.Trp467
pAKT S473 nromalized to V5-AKT2
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
serum starved
insulin stimulated
AKT2
AKT2.Lys17
AKT2.Pro50
AKT2.Lys208
AKT2.His274
AKT2.Trp467

C.

General linear analysis

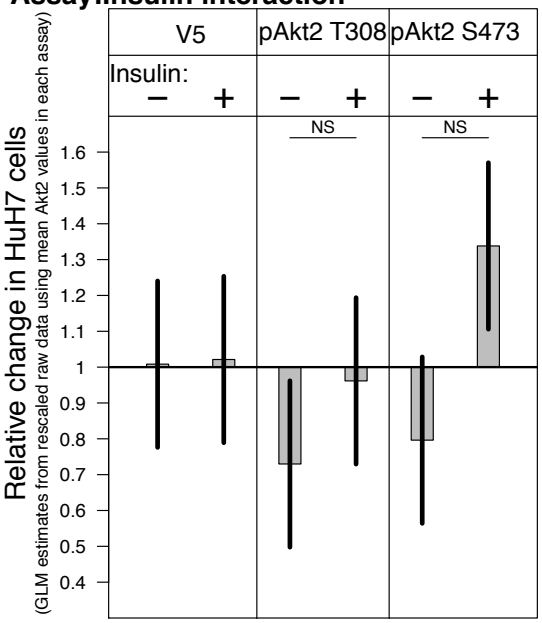
"Round" model:

Variables	*df*	*Variance explained (%)*	*F*	*Pr(>F)*
Round	2	1.86%	0.903	0.409
Assay	2	1.04%	0.504	0.606
Insulin induction	1	2.00%	1.941	0.167
Round:Assay	4	0.20%	0.049	0.995
Round:Insulin	2	0.11%	0.055	0.946
Assay:Insulin	2	1.37%	0.664	0.517
Round:Assay:Insulin	4	0.63%	0.152	0.962

Full model:

Variables	*df*	*Variance explained (%)*	*F*	*Pr(>F)*
Assay	2	1.04%	1.96	**1.47E-01**
Variants	5	46.52%	35.13	**2.20E-16**
Insulin induction	1	2.00%	7.56	**7.28E-03**
Assay:Variant	10	26.02%	9.83	**8.39E-11**
Assay:Insulin	2	1.37%	2.59	8.11E-02

Assay:Insulin interaction



Assay:Variants interaction

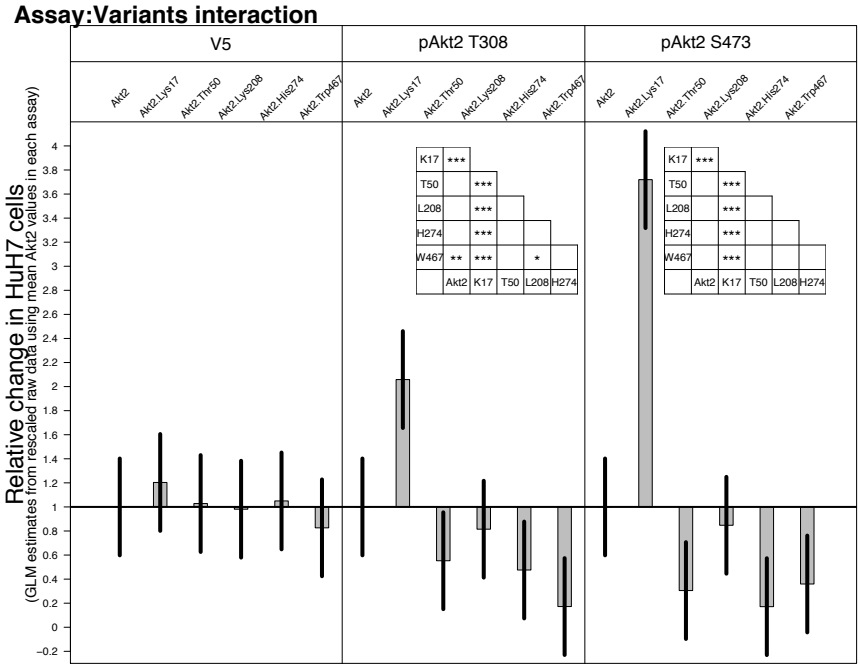


Fig. S7: **Phosphorylation of AKT2 activation sites in HuH7 liver cells** (A) HuH7 cells cells were infected with lentiviral V5-AKT2, V5-AKT2-Lys17, V5-AKT2-Thr50, V5-AKT2-Lys208, V5-AKT2-His274, V5-AKT2-Trp467, starved for 18 hr (white bar), and stimulated for 20 min with 100nm insulin (grey bar). V5-tagged AKT2 was isolated from cell lysates with anti-V5 agarose beads and immunoblots (IB) were probed with indicated antibodies. (B) Phosphorylated AKT2 Thr308 and Ser473 were quantified and normalized to total by V5-AKT2. (C) Linear model for the statistical analysis of quantified pAKT2. The "Round" model tests for significant differences between the three rounds of analysis. The Full model examines significance of assay (V5, pAKT2 T308 and pAKT2 S473) and variants (AKT2, AKT2.Lys17, AKT2.Thr50, AKT2.Lys208, AKT2.His274 and AKT2.Trp467) and their interactions. * P < 0.05, ** P < 0.01, *** P < 0.001. DF: degrees of freedom, F: statistic testing the importance of the grouping term, Pr(>F): P value of the F statistic.

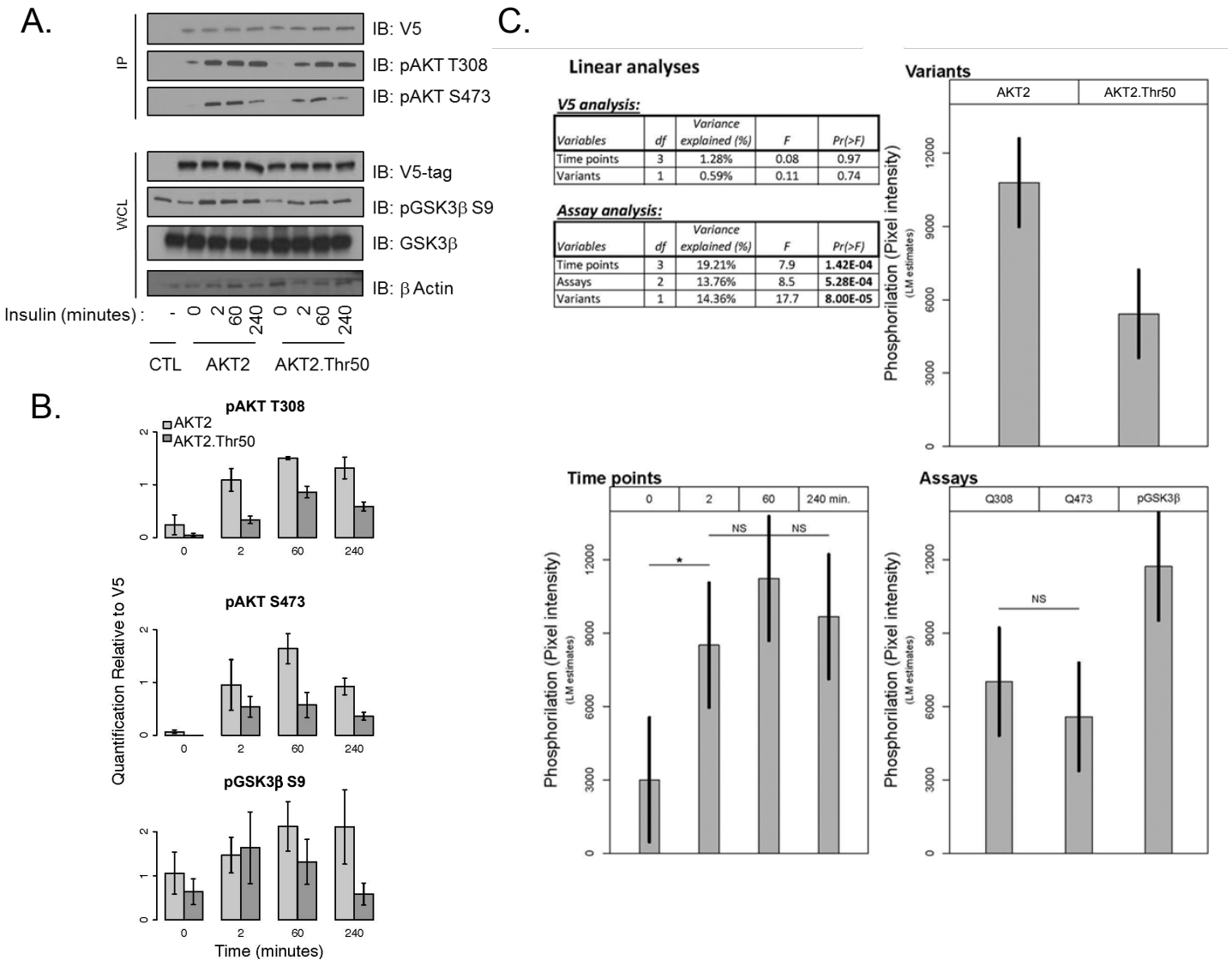


Fig. S8. Time-course analysis of AKT2 phosphorylation (A) HeLa cells were infected with lentiviral V5-AKT2, V5-AKT2-Thr50, or control pLX304, starved for 18 hours and then stimulated for 0, 2, 60, and 240 minutes with 100nm insulin. V5-tagged AKT2 was isolated from cell lysates with anti-V5 agarose beads. Immunoprecipitated (IP) V5-AKT2 and whole cell lysates (WCL) were immunoblotted (IB) with the indicated antibodies. Immunoblots are representative of three independent replicates. (B) Quantification of the three replicates of indicated immunoblots relative to total V5-AKT2. (C) Linear Model (LM) statistical analysis across all three independent replicates. Error bars represent the standard deviation (SD). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

A. General linear analysis

Variables	*df*	*Variance explained (%)*	*F*	*Pr(>F)*
Round	2	33.41%	1186.3	2.20E-16
Variants	2	28.95%	1028.2	**2.20E-16**
Round:Variants	4	37.13%	659.3	**2.20E-16**

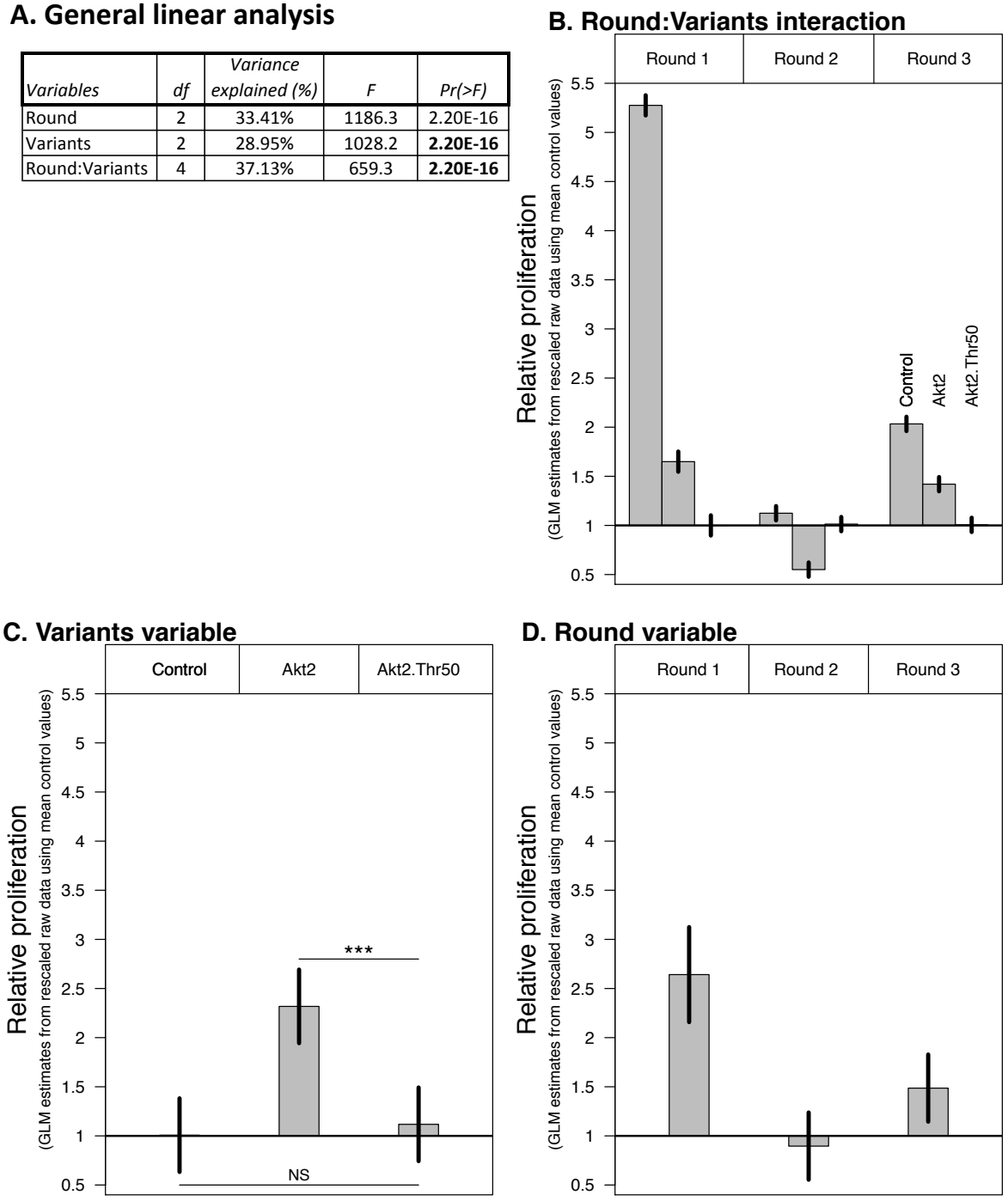
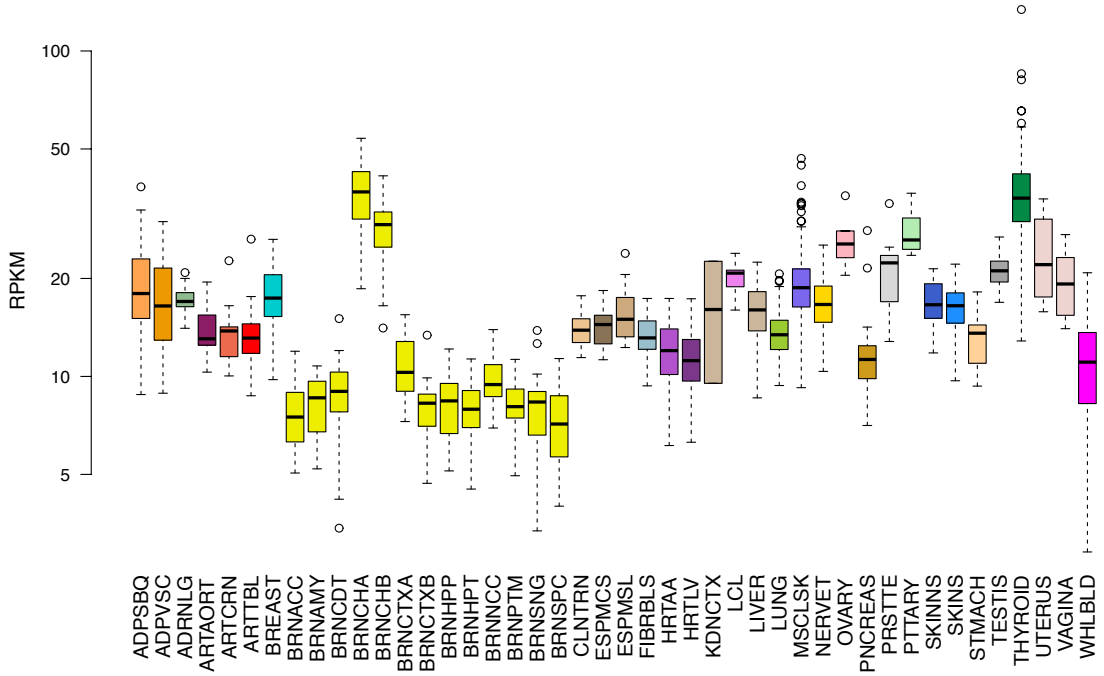
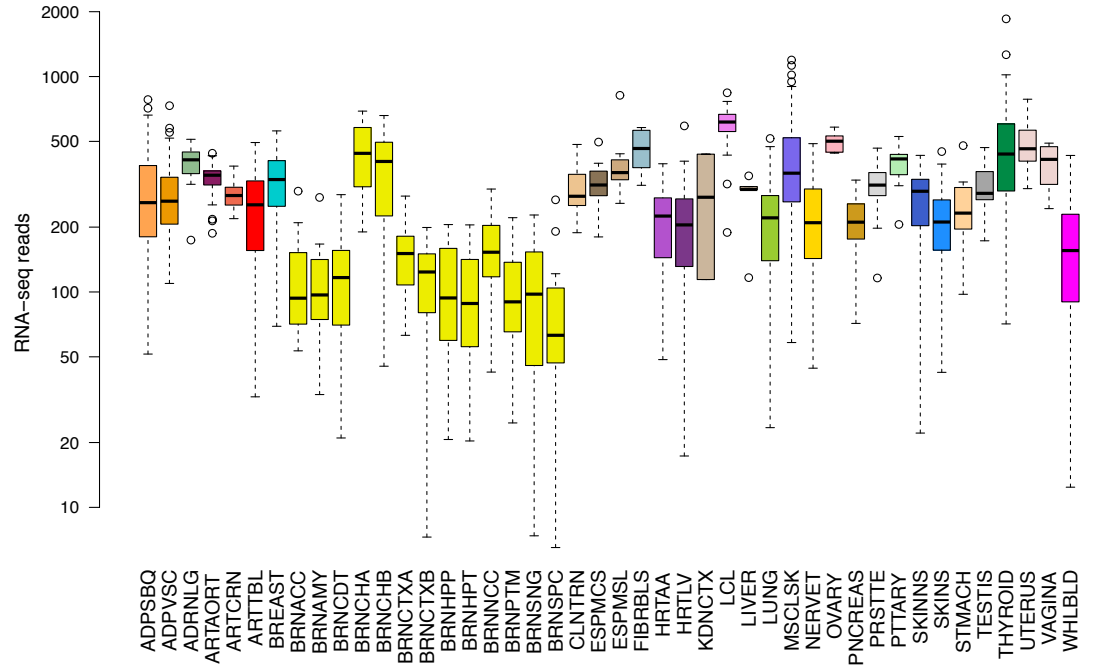


Fig. S9. Proliferation assay. A. Results of a generalized linear model (GLM) applied on rescaled raw data (absorbance value) to test for significant difference in proliferation between the three rounds of analysis, the three variants and an interaction between round and variants. The rescaling was performed by dividing all the values in each round by the average absorbance in controls. The plots represent the GLM estimates (and 95% CI) for the **B.** Round:Variant interaction and individual variables: **C.** Round and **D.** Variants. * P < 0.05, ** P < 0.01, *** P < 0.001. DF: degrees of freedom, F: statistic testing the importance of the grouping term, Pr(>F): P value of the F statistic.

A


RPKM
100
50
20
10
5
ADPSBQ
ADPVSC
ADRNLG
ARTAORT
ARTCRN
ARTTBL
BREAST
BRNACC
BRNAMY
BRNCDT
BRNCHA
BRNCHB
BRNCTXA
BRNCTXB
BRNHPP
BRNHPT
BRNNCC
BRNPTM
BRNSNG
BRNSPC
CLNTRN
ESPMCS
ESPMSL
FIBRBLS
HRTAA
HRTLV
KDNCTX
LCL
LIVER
LUNG
MSCLSK
NERVET
OVARY
PNCREAS
PRSTTE
PTTARY
SKINNS
SKINS
STMACH
TESTIS
THYROID
UTERUS
VAGINA
WHLBLD

B


RNA-seq reads
2000
1000
500
200
100
50
20
10
ADPSBQ
ADPVSC
ADRNLG
ARTAORT
ARTCRN
ARTTBL
BREAST
BRNACC
BRNAMY
BRNCDT
BRNCHA
BRNCHB
BRNCTXA
BRNCTXB
BRNHPP
BRNHPT
BRNNCC
BRNPTM
BRNSNG
BRNSPC
CLNTRN
ESPMCS
ESPMSL
FIBRBLS
HRTAA
HRTLV
KDNCTX
LCL
LIVER
LUNG
MSCLSK
NERVET
OVARY
PNCREAS
PRSTTE
PTTARY
SKINNS
SKINS
STMACH
TESTIS
THYROID
UTERUS
VAGINA
WHLBLD

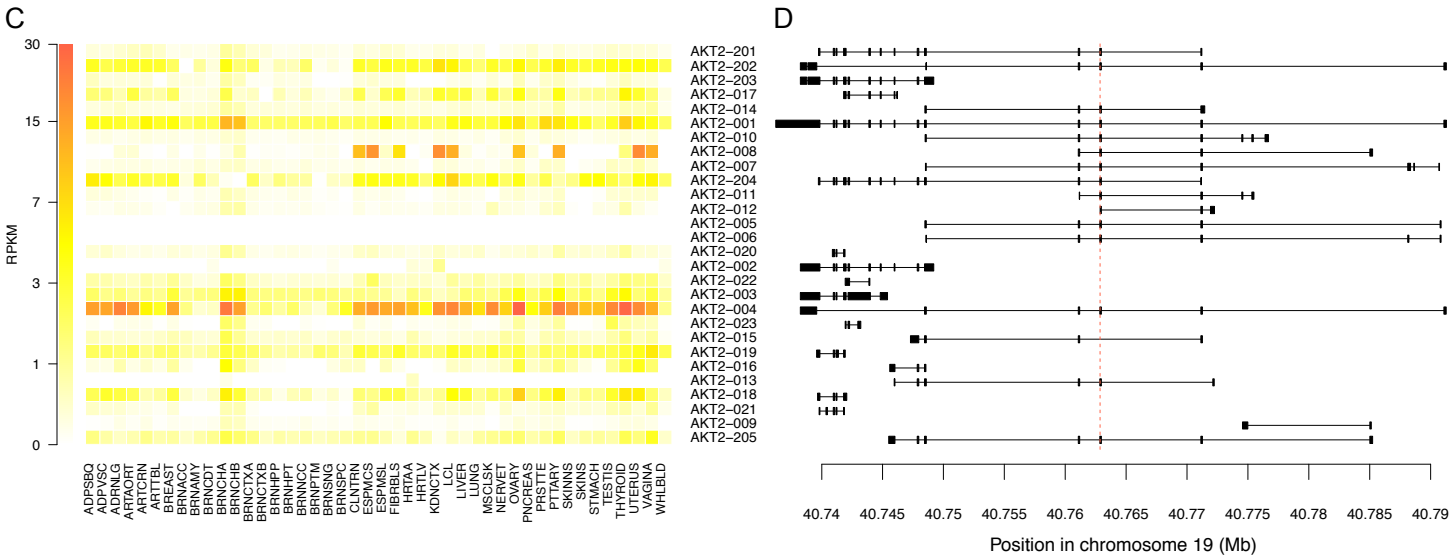


Fig. S10: *AKT2* expression in human tissues. A. Boxplot displaying the level and distribution of *AKT2* gene expression (in reads per kilobase per million mapped reads, RPKM) in 44 human tissues available in the GTEx RNA-seq data. **B**. Box plot of the expression (in RNA-seq reads) of the *AKT2* exon of affected by the p.Pro50Thr variant. Read counts are not normalized by the total number of reads per sample, resulting in larger variance in the expression within each tissue. **C**. Heat map of expression patterns of the 28 *AKT2* transcripts in the GTEx tissues, as annotated in Gencode version 12. Intensity of color in each cell represents the expression of the transcript in that tissue; white indicating no expression, and red indicating higher expression. **D**. Visualization of the transcript structure of *AKT2* (Gencode v12). The affected exon, highlighted with the red dashed line, is included in the majority of the *AKT2* transcripts and in all the three most highly expressed transcripts. The tissues are presented in the same order across panels A-C, and colored similarly in panels A and B. Tissue abbreviations are listed in **Table S4**.

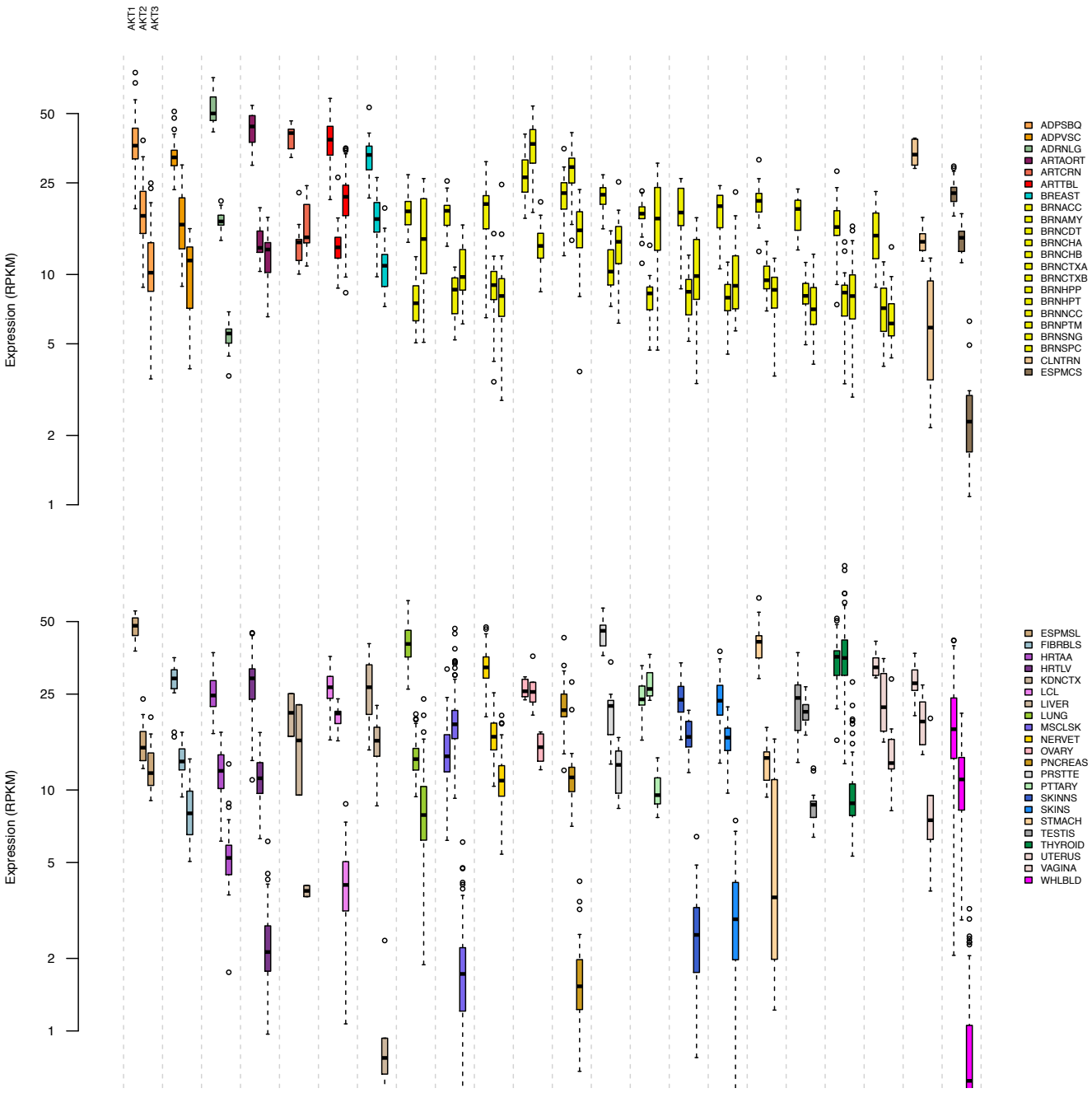
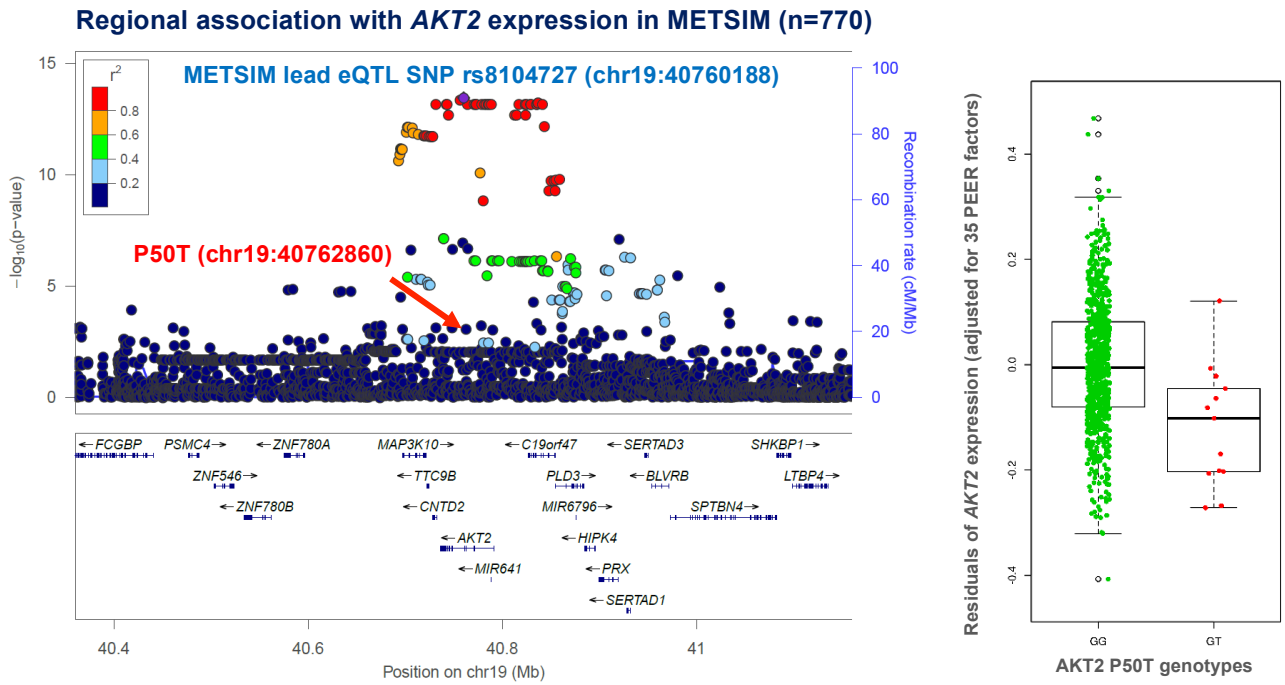


Fig. S11. Expression of the *AKT* gene family across human tissues. Each cluster of three boxplots represents the expression of *AKT1* (**left**), *AKT2* (**middle**) and *AKT3* (**right**) in each tissue. *AKT2* is the isoform with the highest expression (P-value < 0.05) in BRNCHA (Brain – Cerebellum), BRNCHB (Brain - Cerebellar Hemisphere), MSCLSK (Muscle – Skeletal) and PTTARY (Pituitary). Tissue abbreviations are listed in **Table S4**.



	Increasing allele / decreasing alleles	Frequency of decreasing allele	Initial Effect of decreasing allele	P	Conditional Effect of decreasing allele	Conditional P
AKT2 Pro50Thr	G/T	0.0083	-0.980	8.9E-04	-0.754	8.4E-03
Lead eSNP rs8104727	T/C	0.647	-0.403	3.6E-14	-0.391	1.9E-13

Fig S12: Expression analysis with common eQTL SNP and *AKT2* p.Pro50Thr. Top left plot: The regional association plot of variants in the *AKT2* region testing association with *AKT2* expression. The SNP showing the most significant signal in this plot, rs8104727, is a proxy for rs11880261 (r^2 = 1, D' = 1 in the 1000 Genomes phase 3 Finnish sample). Top right plot: observed AKT2 expression levels for the two *AKT2* p.Pro50Thr genotypes observed in the METSIM cohort. Bottom table: eQTL statistics and reciprocal conditional analysis with the two SNPs: rs8104727and *AKT2* p.Pro50Thr. The "Beta conditional" and "P conditional" columns highlight the associations with *AKT2* expression after conditioning on the other SNP.

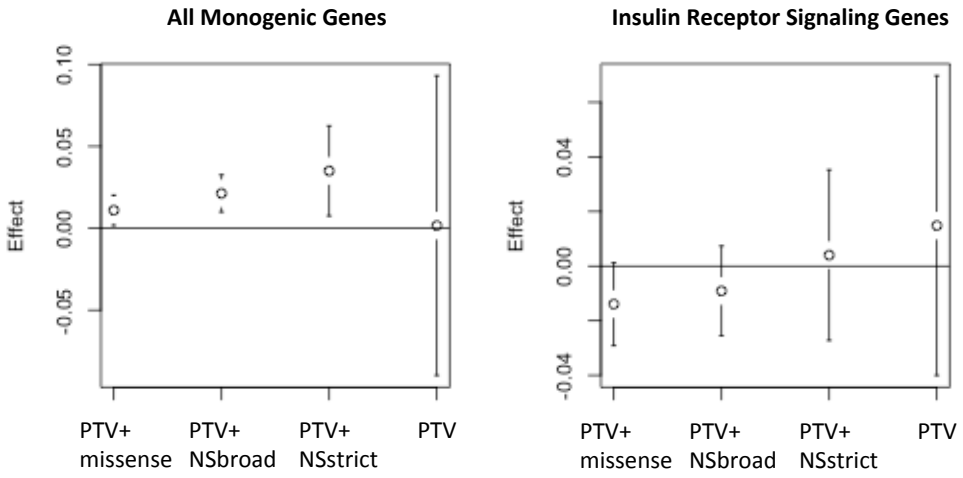


Fig. S13: The trend in the estimate of the effect size of the global gene burden test for the four variant aggregation categories. The effect estimates (and 95% confidence interval) were provided as output of the burden test result in the RareMETALS package in R.

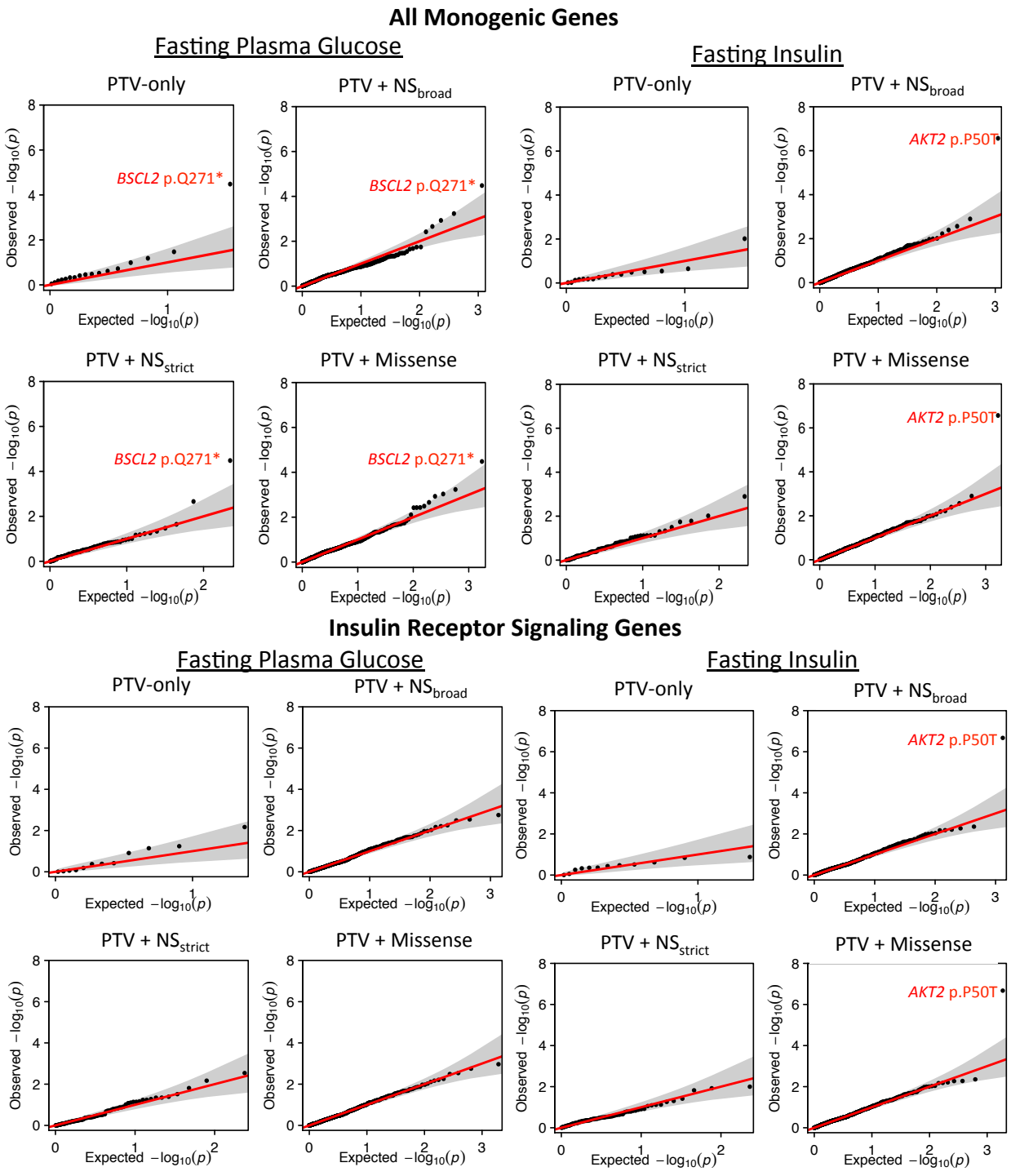


Fig. S14A: Monogenic enrichment in single variant association tests. Single variant association results from the FG and FI association analysis for variants in the four masks in the monogenic gene sets (top) and the insulin receptor signaling genes (bottom).

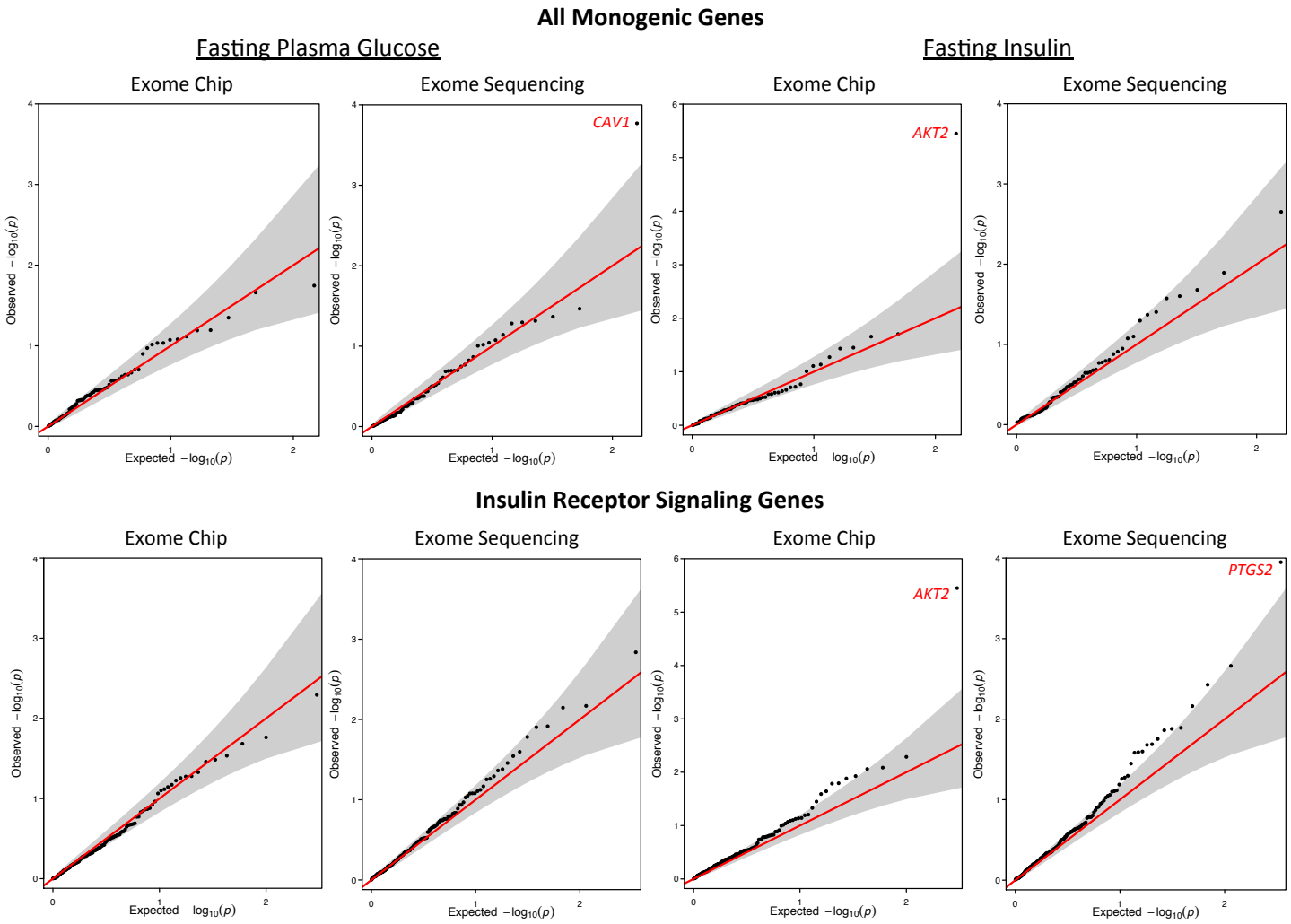


Fig. S14B: Pathway enrichment in gene-based tests. Gene burden association results from the fasting glucose and fasting insulin analysis for variants in the PTV+Missense mask in the monogenic gene sets (top) and the insulin receptor signaling genes (bottom).

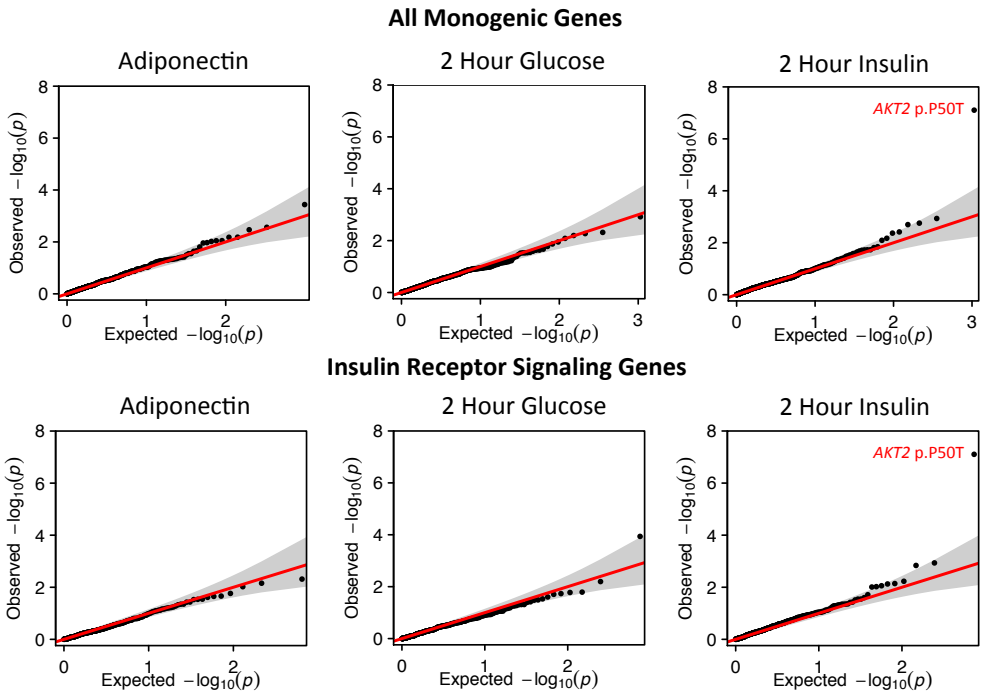


Fig. S14C: Pathway associations in traits related to insulin resistance. Single variant association results for three traits related to insulin resistance: fasting adiponectin levels, 2 hour glucose level and 2 hour insulin level after an oral glucose tolerance test. The variants in these plots are in the PTV+Missense annotation category, with results from variants in the Monogenic gene sets (top) and the insulin receptor signaling genes (bottom).

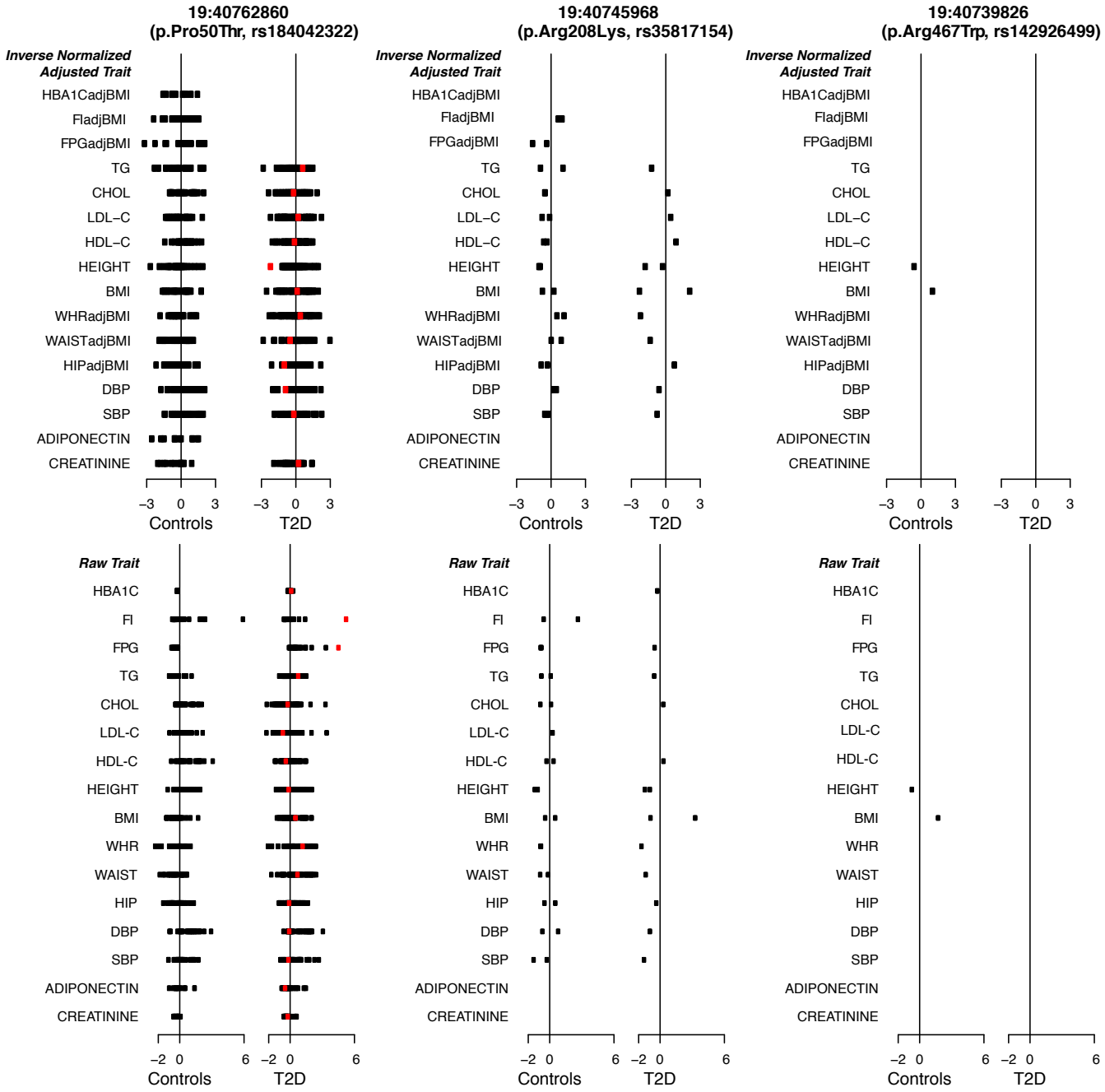


Fig. S15A: Trait values among *AKT2* variant carriers. Profile of the inverse normalized, adjusted metabolic trait values (top plot) and scaled raw trait values (bottom plot) of carriers of three *AKT2* variants: *AKT2* p.Pro50Thr, *AKT2* p.Arg208Lys and *AKT2* p.Arg467Trp from the T2D-GENES whole exome sequencing data set. Points on the graph are observed trait values for heterozygous (black) and homozygous (red) carriers of the variants, split by type 2 diabetes status. Trait abbreviations: HBA1C- glycated hemoglobin, FAST_INS- fasting insulin, FAST_GLU- fasting plasma glucose, TG- triglycerides, CHOL- total cholesterol, LDL-C, low-density lipoprotein cholesterol, HDL-C- high-density lipoprotein cholesterol, BMI- body mass index, WHR- waist to hip ratio, WASITC- waist circumference, HIPC- hip circumference, DBP- diastolic blood pressure, SBP- systolic blood pressure. adjBMI- trait adjusted for BMI

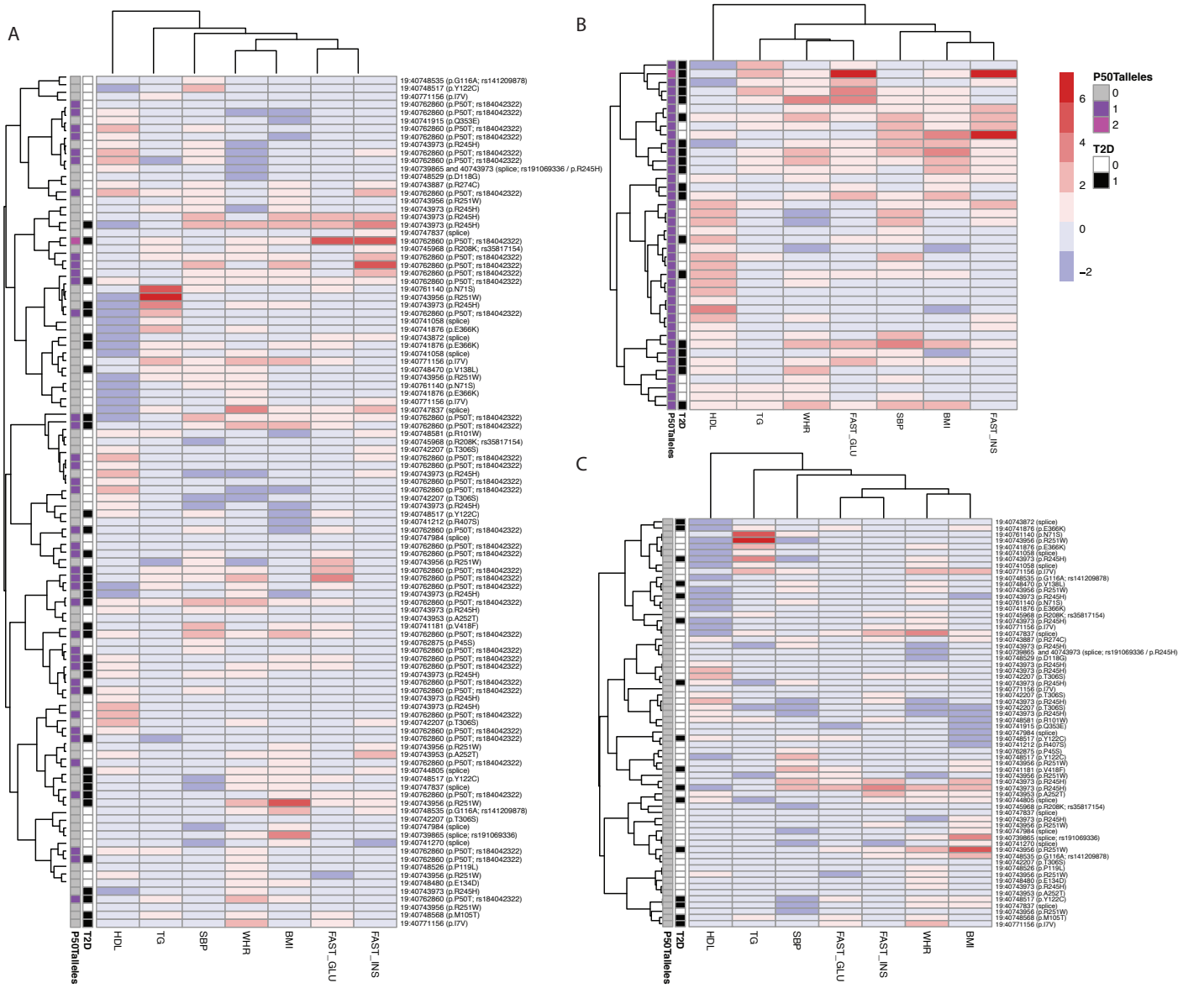


Fig. S15B: Phenotype clustering of *AKT2* missense variant carriers in the T2D-GENES whole exome sequencing dataset on seven metabolic traits: all missense carriers (**A**), carriers of *AKT2* p.Pro50Ala variant (**B**), and carriers of the other variants (**C**), (see Supplementary Table 5). The row labels indicate the variant carried by an individual. P50Talleles: the number of Ala alleles carried; T2D: 0 for controls and 1 for type 2 diabetics.

Additional References

121. Johnson, A.D. *et al.* SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* **24**, 2938-9 (2008).
122. Dunning, C.J. *et al.* Human CIA30 is involved in the early assembly of mitochondrial complex I and mutations in its gene cause disease. *EMBO J* **26**, 3227-37 (2007).
123. Fassone, E. *et al.* Mutations in the mitochondrial complex I assembly factor NDUFAF1 cause fatal infantile hypertrophic cardiomyopathy. *J Med Genet* **48**, 691-7 (2011).
124. Krucken, J. *et al.* Comparative analysis of the human gimap gene cluster encoding a novel GTPase family. *Gene* **341**, 291-304 (2004).
125. Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5**, 621-8 (2008).
126. Konishi, H., Shinomura, T., Kuroda, S., Ono, Y. & Kikkawa, U. Molecular cloning of rat RAC protein kinase alpha and beta and their association with protein kinase C zeta. *Biochem Biophys Res Commun* **205**, 817-25 (1994).
127. Yang, Z.Z. *et al.* Dosage-dependent effects of Akt1/protein kinase Balpha (PKBalpha) and Akt3/PKBgamma on thymus, skin, and cardiovascular and nervous system development in mice. *Mol Cell Biol* **25**, 10407-18 (2005).
128. Altomare, D.A., Lyons, G.E., Mitsuuchi, Y., Cheng, J.Q. & Testa, J.R. Akt2 mRNA is highly expressed in embryonic brown fat and the AKT2 kinase is activated by insulin. *Oncogene* **16**, 2407-11 (1998).
129. Greenbaum, D., Colangelo, C., Williams, K. & Gerstein, M. Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biol* **4**, 117 (2003).
130. Ning, K., Fermin, D. & Nesvizhskii, A.I. Comparative analysis of different label-free mass spectrometry based protein abundance estimates and their correlation with RNA-Seq gene expression data. *J Proteome Res* **11**, 2261-71 (2012).
131. Schwanhausser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337-42 (2011).
132. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-50 (2005).
133. Pers, T.H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun* **6**, 5890 (2015).
134. The pathways were generated through the use of QIAGEN's Ingenuity Pathway Analysis (IPA® , QIAGEN Redwood City, http://www.qiagen.com/ingenuity).

135. Mootha, V.K. *et al.* PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* **34**, 267-73 (2003).
136. Materials and methods are available as supplementary materials on Science Online.
137. Lei, K.J., Shelly, L.L., Pan, C.J., Sidbury, J.B. & Chou, J.Y. Mutations in the glucose-6-phosphatase gene that cause glycogen storage disease type 1a. *Science* **262**, 580-3 (1993).
138. Cao, H., Alston, L., Ruschman, J. & Hegele, R.A. Heterozygous CAV1 frameshift mutations (MIM 601047) in patients with atypical partial lipodystrophy and hypertriglyceridemia. *Lipids Health Dis* **7**, 3 (2008).