# 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*

Alisa Manning*[1,2,3], Heather M Highland*[4,5], Jessica Gasser*[1], Xueling Sim*[6,7], Taru Tukiainen*[1,8,9], Pierre Fontanillas*[1,10], Niels Grarup[11], Manuel A Rivas[12], Anubha Mahajan[12], Adam E Locke[6], Pablo Cingolani[13,14], Tune H Pers[1,11,15,16], Ana Viñuela[17,18,19], Andrew A Brown[20,21], Ying Wu[22], Jason Flannick[1,23], Christian Fuchsberger[6], Eric R Gamazon[24,25], Kyle J Gaulton[12,26], Hae Kyung Im[24], Tanya M Teslovich[6], Thomas W Blackwell[6], Jette Bork-Jensen[11], Noël P Burtt[1], Yuhui Chen[12], Todd Green[1], Christopher Hartl[1], Hyun Min Kang[6], Ashish Kumar[12,27], Claes Ladenvall[28], Clement Ma[6], Loukas Moutsianas[12], Richard D Pearson[12], John R B Perry[12,29,30], N William Rayner[12,31,32], Neil R Robertson[12,31], Laura J Scott[6], Martijn van de Bunt[12,31], Johan G Eriksson[33,34,35,36,37], Antti Jula[37], Seppo Koskinen[37], Terho Lehtimäki[38], Aarno Palotie[1,2,39], Olli T Raitakari[40,41], Suzanne BR Jacobs[1], Jennifer Wessel[42,43], Audrey Y Chu[44], Robert A Scott[30], Mark O Goodarzi[45,46], Christine Blancher[47], Gemma Buck[47], David Buck[47], Peter S Chines[48], Stacey Gabriel[1], Anette P Gjesing[11], Christopher J Groves[31], Mette Hollensted[11], Jeroen R Huyghe[6], Anne U Jackson[6], Goo Jun[6], Johanne Marie Justesen[11], Massimo Mangino[49], Jacquelyn Murphy[1], Matt Neville[31], Robert Onofrio[1], Kerrin S Small[49], Heather M Stringham[6], Joseph Trakalo[47], Eric Banks[1], Jason Carey[1], Mauricio O Carneiro[1], Mark DePristo[1], Yossi Farjoun[1], Timothy Fennell[1], Jacqueline I Goldstein[1,8], George Grant[1], Martin Hrabé de Angelis[50,51,52], Jared Maguire[1], Benjamin M Neale[1,8], Ryan Poplin[1], Shaun Purcell[1,2,53], Thomas Schwarzmayr[54], Khalid Shakir[1], Joshua D Smith[55], Tim M Strom[54,56], Thomas Wieland[54], Jaana Lindstrom[57], Ivan Brandslund[58,59], Cramer Christensen[60], Gabriela L Surdulescu[49], Timo A Lakka[61,62,63], Alex S F Doney[64], Peter Nilsson[65], Nicholas J Wareham[30], Claudia Langenberg[30], Tibor V Varga[66], Paul W Franks[66,67,68], Olov Rolandsson[68], Anders H Rosengren[28], Vidya S Farook[69], Farook Thameem[70], Sobha Puppala[69], Satish Kumar[69], Donna M Lehman[70], Christopher P Jenkinson[70,71], Joanne E Curran[69], Daniel Esten Hale[72], Sharon P Fowler[70], Rector Arya[72], Ralph A DeFronzo[70], Hanna E Abboud[70], Ann-Christine Syvänen[73], Pamela J Hicks[74,75,76], Nicholette D Palmer[74,75,76], Maggie C Y Ng[74,75], Donald W Bowden[74,75,76], Barry I Freedman[77], Tõnu Esko[1,9,78,79], Reedik Mägi[79], Lili Milani[79], Evelin Mihailov[79], Andres Metspalu[79], Narisu Narisu[48], Leena Kinnunen[37], Lori L Bonnycastle[48], Amy Swift[48], Dorota Pasko[29], Andrew R Wood[29], João Fadista[28], Toni I Pollin[80], Nir Barzilai[81], Gil Atzmon[81], Benjamin Glaser[82], Barbara Thorand[51,83], Konstantin Strauch[84,85], Annette Peters[51,83,86], Michael Roden[87,88], Martina Müller-Nurasyid[84,85,86,89], Liming Liang[90,91], Jennifer Kriebel[51,83,92], Thomas Illig[92,93,94], Harald Grallert[51,83,92], Christian Gieger[84], Christa Meisinger[83], Lars Lannfelt[95], Solomon K Musani[96], Michael Griswold[97], Herman A Taylor Jr[98], Gregory Wilson Sr[99], Adolfo Correa[98], Heikki Oksa[100], William R Scott[101], Uzma Afzal[101], Sian-Tsung Tan[102,103], Marie Loh[101,104,105], John C Chambers[101,103,106], Jobanpreet Sehmi[102,103], Jaspal Singh Kooner[102], Benjamin Lehne[101], Yoon Shin Cho[107], Jong-Young Lee[108], Bok-Ghee Han[109], Annemari Käräjämäki[110,111], Qibin Qi[67,112], Lu Qi[67,113], Jinyan Huang[90], Frank B Hu[67,90], Olle Melander[114], Marju Orho-Melander[115], Jennifer E Below[116], David Aguilar[117], Tien Yin Wong[118,119], Jianjun Liu[7,120], Chiea-Chuen Khor[7,118,119,120,121], Kee Seng Chia[7], Wei Yen Lim[7], Ching-Yu Cheng[7,118,119,122], Edmund Chan[123], E Shyong Tai[7,123,124], Tin Aung[118,119], Allan Linneberg[125,126,127], Bo Isomaa[128,129], Thomas Meitinger[54,56,86], Tiinamaija Tuomi[129,130], Liisa Hakaste[35], Jasmina Kravic[28], Marit E Jørgensen[131], Torsten Lauritzen[132], Panos Deloukas[32], Kathleen E Stirrups[133,134], Katharine R Owen[31,135], Andrew J Farmer[136], Timothy M Frayling[29], Stephen P O'Rahilly[137], Mark Walker[138], Jonathan C Levy[31], Dylan Hodgkiss[49], Andrew T Hattersley[139], Teemu Kuulasmaa[140], Alena Stančáková[140], Inês Barroso[32,137], Dwaipayan Bharadwaj[141], Juliana Chan[142,143,144], Giriraj R Chandak[145], Mark J Daly[8], Peter J Donnelly[12,146], Shah B Ebrahim[147], Paul Elliott[101,148], Tasha Fingerlin[149], Philippe Froguel[150], Cheng Hu[151], Weiping Jia[151], Ronald C W Ma[142,143,144], Gilean McVean[12], Taesung Park[152,153], Dorairaj Prabhakaran[147], Manjinder Sandhu[32,154], James Scott[102], Rob Sladek[14,155,156], Nikhil Tandon[157], Yik Ying Teo[7,158,159], Eleftheria Zeggini[32], Richard M Watanabe[160,161,162], Heikki A Koistinen[37,163,164], Y Antero Kesaniemi[165], Matti Uusitupa[166], Timothy D Spector[49], Veikko Salomaa[37], Rainer Rauramaa[167], Colin N A Palmer[168], Inga Prokopenko[12,31,169], Andrew D Morris[170], Richard N Bergman[171], Francis S Collins[48], Lars Lind[172], Erik Ingelsson[173,174], Jaakko Tuomilehto[57,175,176,177], Fredrik Karpe[31,135], Leif Groop[28], Torben Jørgensen[125,178], Torben Hansen[11,179], Oluf Pedersen[11], Johanna Kuusisto[140,180], Gonçalo Abecasis[6], Graeme I

Bell[181], John Blangero[69], Nancy J Cox[24], Ravindranath Duggirala[69], Mark Seielstad[182,183], James G Wilson[184], Josee Dupuis[185,186], Samuli Ripatti[20,39,187], Craig L Hanis[116], Jose C Florez[1,2,3,188], Karen L Mohlke[22], James B Meigs[1,3,189], Markku Laakso[140,180], Andrew P Morris[12,79,190], Michael Boehnke[6], David Altshuler[1,3,9,23,188,191], Mark I McCarthy[12,31,135], Anna L Gloyn[#12,31,135], Cecilia M Lindgren[#1,12,192]

**Affiliations:**

1. Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, USA.
2. Center for Human Genetic Research, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA.
3. Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA.
4. Human Genetics Center, The University of Texas Graduate School of Biomedical Sciences at Houston, The University of Texas Health Science Center at Houston, Houston, Texas, USA.
5. Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA.
6. Department of Biostatistics and Center for Statistical Genetics, School of Public Health, University of Michigan, Ann Arbor, Michigan, USA.
7. Saw Swee Hock School of Public Health, National University of Singapore, National University Health System, Singapore.
8. Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA.
9. Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA.
10. 23andMe, Mountain View, California, USA.
11. The Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.
12. Wellcome Trust Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, UK.
13. School of Computer Science, McGill University, Montreal, Quebec, Canada.
14. McGill University and Génome Québec Innovation Centre, Montreal, Quebec, Canada.
15. Divisions of Endocrinology and Genetics and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Boston, Massachusetts, USA.
16. Department of Epidemiology Research, Statens Serum Institut, Copenhangen, Denmark.
17. Twin Research and Genetic Epidemiology, King's College London, London, UK.
18. Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland.
19. Institute of Genetics and Genomics in Geneva, University of Geneva, Geneva, Switzerland.
20. Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK.
21. NORMENT, KG Jebsen Center for Psychosis Research, Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway.
22. Department of Genetics, University of North Carolina, Chapel Hill, North Carolina, USA.
23. Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts, USA.
24. Department of Medicine, Section of Genetic Medicine, The University of Chicago, Chicago, Illinois, USA.
25. Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands.
26. Department of Pediatrics, University of California San Diego, La Jolla, California, USA.
27. Chronic Disease Epidemiology, Swiss Tropical and Public Health Institute, University of Basel, Basel, Switzerland.
28. Department of Clinical Sciences, Diabetes and Endocrinology, Lund University Diabetes Centre, Malmö, Sweden.
29. Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Exeter, UK.
30. MRC Epidemiology Unit, Institute of Metabolic Science, University of Cambridge, Cambridge, UK.

31. Oxford Centre for Diabetes, Endocrinology and Metabolism, Radcliffe Department of Medicine, University of Oxford, Oxford, UK.
32. Department of Human Genetics, Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, UK.
33. Department of General Practice and Primary Healthcare, University of Helsinki, Helsinki, Finland.
34. Unit of General Practice, Helsinki University Central Hospital, Finland.
35. Folkhälsan Research Center, Helsinki, Finland.
36. Vaasa Central Hospital, Vaasa, Finland.
37. Department of Health, National Institute of Health and Welfare, Helsinki, Finland.
38. Department of Clinical Chemistry, Fimlab Laboratories, University of Tampere School of Medicine, Tampere, Finland.
39. Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland.
40. Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland.
41. Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland.
42. Department of Epidemiology, Fairbanks School of Public Health, Indianapolis, IN, USA.
43. Department of Medicine, Indiana University School of Medicine, Indianapolis, IN, USA.
44. Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA, USA.
45. Division of Endocrinology, Diabetes and Metabolism, Department of Medicine, Cedars-Sinai Medical Center, Los Angeles, California, USA.
46. Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, California, USA.
47. High Throughput Genomics, Oxford Genomics Centre, Wellcome Trust Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, UK.
48. National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, USA.
49. Department of Twin Research and Genetic Epidemiology, King's College London, London, UK.
50. Institute of Experimental Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany.
51. German Center for Diabetes Research (DZD), Neuherberg, Germany.
52. Chair of Experimental Genetics, School of Life Science Weihenstephan, Technische Universität München, Freising, Germany.
53. Department of Psychiatry, Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, USA.
54. Institute of Human Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany.
55. Department of Genome Sciences, University of Washington School of Medicine, Seattle, Washington, USA.
56. Institute of Human Genetics, Technische Universität München, Munich, Germany.
57. Diabetes Prevention Unit, National Institute for Health and Welfare, Helsinki, Finland.
58. Institute of Regional Health Research, University of Southern Denmark, Odense, Denmark.
59. Department of Clinical Biochemistry, Vejle Hospital, Vejle, Denmark.
60. Department of Internal Medicine and Endocrinology, Vejle Hospital, Vejle, Denmark.
61. Institute of Biomedicine, Physiology, University of Eastern Finland, Kuopio, Finland.
62. Kuopio Research Institute of Exercise Medicine, Kuopio, Finland.
63. Department of Clinical Physiology and Nuclear Medicine, Kuopio University Hospital, Kuopio, Finland.
64. Division of Cardiovascular and Diabetes Medicine, Medical Research Institute, Ninewells Hospital and Medical School, Dundee, UK.
65. Department of Clinical Sciences, Medicine, Lund University, Malmö, Sweden.
66. Department of Clinical Sciences, Lund University Diabetes Centre, Genetic and Molecular Epidemiology Unit, Lund University, Malmö, Sweden.
67. Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts, USA.
68. Department of Public Health and Clinical Medicine, Umeå University, Umeå, Sweden.
69. Department of Genetics, Texas Biomedical Research Institute, San Antonio, Texas, USA.

70. Department of Medicine, University of Texas Health Science Center, San Antonio, Texas, USA.
71. Research, South Texas Veterans Health Care System, San Antonio, Texas, USA.
72. Department of Pediatrics, University of Texas Health Science Center, San Antonio, Texas, USA.
73. Department of Medical Sciences, Molecular Medicine and Science for Life Laboratory, Uppsala University, Uppsala, Sweden.
74. Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA.
75. Center for Diabetes Research, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA.
76. Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA.
77. Department of Internal Medicine, Section on Nephrology, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA.
78. Division of Endocrinology, Boston Children's Hospital, Boston, Massachusetts, USA.
79. Estonian Genome Center, University of Tartu, Tartu, Estonia.
80. Department of Medicine, Program in Personalized and Genomic Medicine, University of Maryland, Baltimore, Maryland, USA.
81. Departments of Medicine and Genetics, Albert Einstein College of Medicine, New York, USA.
82. Endocrinology and Metabolism Service, Hadassah-Hebrew University Medical Center, Jerusalem, Israel.
83. Institute of Epidemiology II, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany.
84. Institute of Genetic Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany.
85. Institute of Medical Informatics, Biometry and Epidemiology, Chair of Genetic Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany.
86. Deutsches Forschungszentrum für Herz-Kreislauferkrankungen (DZHK), Partner Site Munich Heart Alliance, Munich, Germany.
87. Institute of Clinical Diabetology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine University, Düsseldorf, Germany.
88. German Center for Diabetes Research, Partner Düsseldorf, Germany.
89. Department of Medicine I, University Hospital Grosshadern, Ludwig-Maximilians-Universität, Munich, Germany.
90. Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA.
91. Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, USA.
92. Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany.
93. Hannover Unified Biobank, Hannover Medical School, Hannover, Germany.
94. Institute of Human Genetics, Hannover Medical School, Hannover, Germany.
95. Department of Public Health and Caring Sciences, Geriatrics, Uppsala University, Uppsala, Sweden.
96. Jackson Heart Study, University of Mississippi Medical Center, Jackson, Mississippi, USA.
97. Center of Biostatistics and Bioinformatics, University of Mississippi Medical Center, Jackson, Mississippi, USA.
98. Department of Medicine, University of Mississippi Medical Center, Jackson, Mississippi, USA.
99. College of Public Services, Jackson State University, Jackson, Mississippi, USA.
100. Pirkanmaa Hospital District, Tampere, Finland.
101. Department of Epidemiology and Biostatistics, Imperial College London, London, UK.
102. National Heart and Lung Institute, Cardiovascular Sciences, Hammersmith Campus, Imperial College London, London, UK.
103. Department of Cardiology, Ealing Hospital NHS Trust, Southall, Middlesex, UK.
104. Institute of Health Sciences, University of Oulu, Oulu, Finland.
105. Translational Laboratory in Genetic Medicine (TLGM), Agency for Science, Technology and Research (A*STAR), Singapore.

106. Imperial College Healthcare NHS Trust, Imperial College London, London, UK.
107. Department of Biomedical Science, Hallym University, Chuncheon, Republic of Korea.
108. Ministry of Health and Welfare, Seoul, Republic of Korea.
109. Center for Genome Science, Korea National Institute of Health, Chungcheongbuk-do, Republic of Korea.
110. Vasa Health Care Center, Vaasa, Finland.
111. Department of Primary Health Care, Vasa Central Hospital, Vasa, Finland.
112. Department of Epidemiology and Population Health, Albert Einstein College of Medicine, New York, USA.
113. Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA.
114. Department of Clinical Sciences, Hypertension and Cardiovascular Disease, Lund University, Malmö, Sweden.
115. Department of Clinical Sciences, Diabetes and Cardiovascular Disease, Genetic Epidemiology, Lund University, Malmö, Sweden.
116. Human Genetics Center, School of Public Health, The University of Texas Health Science Center at Houston, Houston, Texas, USA.
117. Cardiovascular Division, Baylor College of Medicine, Houston, Texas, USA.
118. Singapore Eye Research Institute, Singapore National Eye Centre, Singapore.
119. Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, National University Health System, Singapore.
120. Division of Human Genetics, Genome Institute of Singapore, A*STAR, Singapore.
121. Department of Paediatrics, Yong Loo Lin School of Medicine, National University of Singapore, National University Health System, Singapore.
122. Centre for Quantitative Medicine, Office of Clinical Sciences, Duke-NUS Graduate Medical School Singapore, Singapore.
123. Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, National University Health System, Singapore.
124. Cardiovascular & Metabolic Disorders Program, Duke-NUS Graduate Medical School Singapore, Singapore.
125. Research Centre for Prevention and Health, Glostrup University Hospital, Glostrup, Denmark.
126. Department of Clinical Experimental Research, Rigshospitalet, Glostrup, Denmark.
127. Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.
128. Department of Social Services and Health Care, Jakobstad, Finland.
129. Folkhälsan Research Centre, Helsinki, Finland.
130. Department of Endocrinology, Helsinki University Central Hospital, Helsinki, Finland.
131. Steno Diabetes Center, Gentofte, Denmark.
132. Department of Public Health, Section of General Practice, Aarhus University, Aarhus, Denmark.
133. William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK.
134. Department of Haematology, University of Cambridge, Cambridge, UK.
135. Oxford NIHR Biomedical Research Centre, Oxford University Hospitals Trust, Oxford, UK.
136. Department of Primary Care Health Sciences, University of Oxford, Oxford, UK.
137. Metabolic Research Laboratories, Institute of Metabolic Science, University of Cambridge, Cambridge, UK.
138. The Medical School, Institute of Cellular Medicine, University of Newcastle, Newcastle, UK.
139. University of Exeter Medical School, University of Exeter, Exeter, UK.
140. Faculty of Health Sciences, Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland, Kuopio, Finland.

141. Functional Genomics Unit, CSIR-Institute of Genomics & Integrative Biology (CSIR-IGIB), New Delhi, India.
142. Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Hong Kong, China.
143. Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Hong Kong, China.
144. Hong Kong Institute of Diabetes and Obesity, The Chinese University of Hong Kong, Hong Kong, China.
145. CSIR-Centre for Cellular and Molecular Biology, Hyderabad, Andhra Pradesh, India.
146. Department of Statistics, University of Oxford, Oxford, UK.
147. Centre for Chronic Disease Control, New Delhi, India.
148. MRC-PHE Centre for Environment and Health, Imperial College London, London, UK.
149. Department of Epidemiology, Colorado School of Public Health, University of Colorado, Aurora, Colorado, USA.
150. Genomics and Molecular Physiology, CNRS (Institut de Biologie de Lille), Lille, France.
151. Department of Endocrinology and Metabolism, Shanghai Diabetes Institute, Shanghai Jiao Tong University Affiliated Sixth People's Hospital, Shanghai, China.
152. Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul, Republic of Korea.
153. Department of Statistics, Seoul National University, Seoul, Republic of Korea.
154. Department of Public Health and Primary Care, Institute of Public Health, University of Cambridge, Cambridge, UK.
155. Department of Human Genetics, McGill University, Montreal, Quebec, Canada.
156. Division of Endocrinology and Metabolism, Department of Medicine, McGill University, Montreal, Quebec, Canada.
157. Department of Endocrinology and Metabolism, All India Institute of Medical Sciences, New Delhi, India.
158. Life Sciences Institute, National University of Singapore, Singapore.
159. Department of Statistics and Applied Probability, National University of Singapore, Singapore.
160. Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California, USA.
161. Department of Physiology & Biophysics, Keck School of Medicine, University of Southern California, Los Angeles, California, USA.
162. Diabetes and Obesity Research Institute, Keck School of Medicine, University of Southern California, Los Angeles, California, USA.
163. University of Helsinki and Helsinki University Central Hospital, Department of Medicine and Abdominal Center, Endocrinology, Helsinki, Finland.
164. Minerva Foundation Institute for Medical Research, Helsinki, Finland.
165. Institute of Clinical Medicine, Department of Medicine, University of Oulu and Medical Research Center, Oulu University Hospital, Oulu, Finland.
166. Institute of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio, Finland.
167. Foundation for Research in Health, Exercise and Nutrition, Kuopio Research Institute of Exercise Medicine, Kuopio, Finland.
168. Pat Macpherson Centre for Pharmacogenetics and Pharmacogenomics, Medical Research Institute, Ninewells Hospital and Medical School, Dundee, UK.
169. Department of Genomics of Common Disease, School of Public Health, Imperial College London, London, UK.
170. Clinical Research Centre, Centre for Molecular Medicine, Ninewells Hospital and Medical School, Dundee, UK.
171. Cedars-Sinai Diabetes and Obesity Research Institute, Los Angeles, California, USA.
172. Department of Medical Sciences, Uppsala University, Uppsala, Sweden.
173. Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden.
174. Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, California, USA.

175. Center for Vascular Prevention, Danube University Krems, Krems, Austria.
176. Diabetes Research Group, King Abdulaziz University, Jeddah, Saudi Arabia.
177. Instituto de Investigacion Sanitaria del Hospital Universario LaPaz (IdiPAZ), University Hospital LaPaz, Autonomous University of Madrid, Madrid, Spain.
178. Faculty of Medicine, University of Aalborg, Aalborg, Denmark.
179. Faculty of Health Sciences, University of Southern Denmark, Odense, Denmark.
180. Kuopio University Hospital, Kuopio, Finland.
181. Departments of Medicine and Human Genetics, The University of Chicago, Chicago, Illinois, USA.
182. Department of Laboratory Medicine & Institute for Human Genetics, University of California, San Francisco, San Francisco, California, USA.
183. Blood Systems Research Institute, San Francisco, California, USA.
184. Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, Mississippi, USA.
185. Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA.
186. National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, Massachusetts, USA.
187. Hjelt Institute, University of Helsinki, Helsinki, Finland.
188. Diabetes Research Center (Diabetes Unit), Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA.
189. Division of General Internal Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA.
190. Department of Biostatistics, University of Liverpool, Liverpool, UK.
191. Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.
192. Li Ka Shing Centre for Health Information and Discovery, The Big Data Institute, University of Oxford, Oxford, UK.

* These authors contributed equally to this work.
# These authors jointly directed this research.

**Corresponding authors**
Prof Anna L Gloyn
Oxford Centre for Diabetes Endocrinology & Metabolism
University of Oxford
Churchill Hospital
Headington
Oxford
OX3 7LE
United Kingdom
anna.gloyn@drl.ox.ac.uk

Prof Cecilia M Lindgren
The Wellcome Trust Centre for Human Genetics
University of Oxford
Roosevelt Drive
Oxford
OX3 7BN
United Kingdom
celi@well.ox.ac.uk

**Abstract**

To identify novel coding association signals and facilitate characterization of mechanisms influencing glycemic traits and type 2 diabetes (T2D) risk, we analyzed 109,215 variants derived from exome array genotyping together with an additional 390,225 variants from exome sequence in up to 39,339 normoglycemic individuals from five ancestry groups. We identified a novel coding variant (Pro50Thr) in *AKT2*, a gene in which rare fully-penetrant mutations are responsible for monogenic glycemic disorders. The low-frequency allele is associated with a 12% increase in fasting plasma insulin (FI) levels ($P$=9.2×10$^{-7}$). This variant is present at a 1.1% frequency in Finns but is virtually absent in individuals from other ancestries. Carriers of the FI-increasing allele had increased 2-hour insulin values ($P$=7.9x10$^{-8}$), decreased insulin sensitivity (Matsuda Index, $P$=1.1x10$^{-6}$), and increased risk of T2D (odds ratio=1.05, $P$=8.1×10$^{-5}$). In cellular studies, the AKT2-Thr50 protein exhibited a partial loss of function. Here we extend the allelic spectrum for coding variants in *AKT2* associated with glucose homeostasis, further supporting a role of genetic variants in the insulin-signaling pathway on insulin resistance and T2D risk.

**Introduction**

The increasing prevalence of type 2 diabetes (T2D) is a global health crisis, making it critical to promote development of more efficient strategies for prevention and treatment. Individuals with T2D display both pancreatic beta-cell dysfunction and insulin resistance[1,2]. Genetic studies of surrogate measures of these glycemic traits can identify variants that influence these central features of T2D[3] highlighting potential pathways for therapeutic manipulation. Comprehensive surveys of the influence of common genetic variants on fasting plasma glucose (FG) and fasting plasma insulin (FI) have highlighted defects in pathways involved in glucose metabolism, and insulin processing, secretion, and action[4,5]. Recent studies have identified T2D associated alleles that are common in one population but rare or absent in others[6-8]. These associations were observed either due to an increase in frequency of older alleles based on population dynamics and demography[7], or the emergence of population-specific alleles[6,8]. We set out to identify and characterize low-frequency allele (minor allele frequency [MAF]<5%) glycemic traits associations by meta-analysis of exome sequence and exome array genotype data in a multi-ancestry sample.

**Results**

We assessed the associations of FI and FG with 390,225 variants (minor allele count, MAC>5) from exome sequence (average 80x coverage) in 5,108 non-diabetic African American, European, Hispanic, East Asian, and South Asian individuals from the GoT2D and T2D-GENES consortia with 109,215 variants (MAC>5) derived from exome array genotyping in 33,231 non-diabetic individuals of European ancestry from 14 studies

[**Supplementary Table 1**]. Due to the design of the exome array to include common non-coding variants associated with a number of complex traits[9], the combined exome sequence and exome array data set includes 28 FG and 14 FI loci with known associations[5,10]. Of these, 13 FG and four FI loci show directionally consistent significant associations ($P<5\times10^{-7}$). Among the remaining 15 FG and 10 FI loci, we observe directionally consistent associations in 14 FG and 9 FI loci ($P_{enrichment}$ for FG=$5\times10^{-4}$ and for FI=0.01). In addition, we identified a novel exome wide single variant significant association between FI and the Pro50Thr (rs184042322, MAF = 1.1%) coding variant in *AKT2* (*V-AKT Murine Thymoma Viral Oncogene Homolog 2*) ($P=1.2\times10^{-7}$) [**Fig. 1A; Supplementary Note 1**; **Supplementary Fig. 1**; **Supplementary Tables 2A,B,C**]. The same allele contributed to a significant FI signal for *AKT2* in gene-based analysis ($P=1.2\times10^{-7}$) where we used the sequence kernel association test (SKAT)[11] and a frequency-weighted burden test using four variant masks (protein truncating variants (PTV), PTV+non-synonymous$(NS)_{broad}$, PTV+$NS_{strict}$, and PTV+missense, see **Methods**) and combinations of allele frequency and variant prediction algorithms[12,13]. We discovered two additional gene-based exome-wide significant ($P<2.5\times10^{-6}$) associations between *GIMAP8* ($P_{PTV}=2.3\times10^{-6}$) and FG, and between *NDUFAF1* ($P_{PTV+NSBroad}=9.2\times10^{-7}$) and FI [**Supplementary Note 1; Supplementary Fig. 2**; **Supplementary Table 2D**].

We aggregated our allele frequency observations with data from 46,658 individuals from the CHARGE consortium[14] (where each individual cohort MAC<5). *AKT2* Pro50Thr was observed at a much higher frequency in Finnish individuals (MAF=1.1%) than other European (MAF=0.2%), African American (MAF=0.1%), Asian (MAF=0%), or Hispanic (MAF=0%) individuals [**Figure 1B**]. There was modest heterogeneity across regions of Finland, with North Karelia (MAF=1.7%) different ($0.0011<$pairwise $F_{ST}<0.0027$; $P<0.01$) from all other tested regions, except Central Finland (MAF=1.3%, pairwise $F_{ST}=0.00041$, $P=0.08$). Although there is evidence for constraint against missense changes across the gene as described below, these geographical differences in allele frequencies are consistent with long-term drift[15] with no evidence of selection pressure differences at *AKT2* across Finland (dN/dS [Finland] = 0.13; 0.08 < dN/dS [European] < 0.43).

Given this, we replicated the association between FI and *AKT2* Pro50Thr by meta-analyzing four independent Finnish studies ($P=5.4\times10^{-4}$, N=5,833) with the discovery studies ($P_{combined}=1.0\times10^{-9}$, N=25,316). We did not observe evidence of effect-size heterogeneity between studies. This allele is associated with a 12% (95% confidence interval [CI]=7%-18%) increase in FI levels in the discovery and replication studies which corresponds to a per rare allele effect of 10.4 pmol/l insulin (95%CI=6.6-14.3) [**Fig. 1A; Supplementary Note 1; Supplementary Table 1; Supplementary Table 2E**].

The serine/threonine protein kinases AKT1, AKT2, and AKT3 are conserved across all vertebrates [**Fig. 2A, B**]. Interestingly, Pro50 and the seven preceding residues in the pleckstrin homology (PH) domain appear to be specific for the AKT2 isoform. In the complete GoT2D and T2D-GENES exome sequence data of 12,940 individuals (6,504 with T2D), *AKT2* displayed evidence of purifying selection (dN/dS = 0.099) [**Supplementary Fig. 3; Supplementary Fig. 4]**. We observed only 36 non-synonymous variants in *AKT2* (35 with a MAC<6 and Pro50Thr with MAC=61) [**Supplementary Table 3**]. No additional protein-altering variants had frequency >0.3% in the 60,706 individuals (including the exome sequence samples in this study) in the Exome Aggregation Consortium (ExAC) data[16,17]. *AKT2* also showed fewer missense variants than expected and extreme constraint to loss-of-function (LoF) variants in ExAC, the missense constraint metric, Z=3.48, is in the 94$^{th}$ percentile of all genes and the estimated probability of being LoF intolerant (pLI)=1 (see **Methods** for details).

AKT2 is a primary transducer of phosphoinositide 3-kinase (PI3K) signaling downstream of the insulin receptor and is responsible for mediating the physiological effects of insulin in tissues including liver, skeletal muscle, and adipose[18-20]. *Akt2* null mice are characterized by hyperglycemia and compensatory hyperinsulinemia, and some develop diabetes[21,22]. In humans, fully-penetrant rare alleles in *AKT2* are known to cause familial partial lipodystrophy and hypoinsulinemic hypoglycemia with hemihypertrophy (Glu17Lys)[23,24] or a syndrome featuring severe insulin resistance, hyperinsulinemia, and diabetes mellitus (Arg274His)[25] and additional rare alleles have been observed in patients with severe insulin resistance (Arg208Lys, and Arg467Trp)[26]. To understand the functional consequences of the *AKT2* Pro50Thr variant on the protein, we investigated protein expression, activation, kinase activity, and downstream effector phosphorylation.

First, we used *in silico* classifiers[27,28] to assess the potential functional consequence of *AKT2* Pro50Thr on protein function. Two of the five classifiers predicted *AKT2* Pro50Thr to be deleterious [**Supplementary Table 3**]. Next, we used 3D models of *AKT2* viewed in the PyMol software, which predicted that the Pro50Thr amino acid change causes a change in the conformations of the lipid binding PH domain [**Fig. 3A, Supplementary Fig. 5**]. We hypothesized that the variant protein is inefficiently recruited to the membrane thereby impacting AKT2 phosphorylation and downstream activity.

To assess the molecular and cellular consequence of the *AKT2* Thr50 variant on protein function, we performed a comparative analysis of AKT2-Thr50 with inactivating and activating variants implicated in monogenic disorders of insulin signaling. Analysis of AKT2-Thr50 expression showed that while AKT2 protein levels remained unchanged, there was a partial loss of AKT2-Thr50 phosphorylation at its activation sites

(Thr308 and Ser473) in HeLa cells, indicative of impaired AKT2 signalling[29-33] [**Fig. 3B; Supplementary Fig. 6**]. Similar effects were observed in human liver derived HuH7 cells [**Supplementary Fig. 7**]. AKT2-Thr50 also showed a reduced ability to phosphorylate its downstream target glycogen synthase kinase 3 beta (GSK3β). These defects in AKT2-Thr50 activity were confirmed through an *in vitro* kinase assay ($P$<0.01) [**Fig. 3B; Supplementary Fig. 6**]. AKT2-Thr50 showed a similar decrease in kinase function to the lipodystrophy-causing AKT2-His274 variant [**Fig. 3B; Supplementary Fig. 6**]. We set out to test if the variant protein reached maximal levels and duration of phosphorylation and activity over time. Using a four-hour time course analysis of AKT2 activity we verified a reduction in both maximally phosphorylated Thr308 and Ser473 in AKT2-Thr50 [**Supplementary Fig. 8**]. To understand how this loss of activity could manifest as a defect in a known cellular function of AKT2[34,35], we determined the impact of AKT2-Thr50 on cell proliferation in HuH7 cells. While the addition of AKT2 stimulated hepatocyte proliferation, the response to AKT2-Thr50 was reduced (effect=-1.2 of relative proliferation, P < 0.001) [**Fig. 3C; Supplementary Fig. 9**].

To investigate potential tissue-specific regulation and function of *AKT2* and its transcript levels, we queried RNA sequencing data from the Genotype Tissue Expression (GTEx) Project[36]. In agreement with previous studies[37,38], *AKT2* is highly and ubiquitously expressed across all tissues (44 tissue types, 3-156 individuals/tissue). Notably the *AKT2* Pro50Thr containing exon is expressed in all tissues and individuals [**Supplementary Fig. 10**], reflecting the importance of the PH domain to AKT2 function[39]**.** Of the three *AKT* homologs, *AKT2* had 1.4-fold higher expression in skeletal muscle than *AKT1* ($P$=1.5×10$^{-19}$) and 11-fold higher expression than *AKT3* ($P$=7.8×10$^{-91}$). Skeletal muscle was the only tested tissue displaying such pronounced *AKT2* enrichment [**Fig. 2C; Supplementary Note 2; Supplementary Fig. 11; Supplementary Table 4**]**.**

Motivated by the age-related loss of adipose tissue in *Akt2* null mice[21,22] and the growth and lipodystrophy phenotypes in carriers of fully-penetrant alleles[23-26], we examined associations of expression levels of *AKT2* with body mass index (BMI), FI, and age in three adipose tissue data sets: METSIM[40], EuroBATS[41,42], and GTEx[36]. We found a nominal association with age in one cohort (METSIM effect=0.02 standard deviations (SD) of normalized expression values, see **Methods**; P=0.004; N=770) and we observed an association between lower BMI levels and higher *AKT2* expression in two cohorts (EuroBATS effect=-0.075 SD; $P$=6.1×10$^{-28}$; N=720; METSIM effect=-0.06 SD; $P$=8.1×10$^{-8}$; N=770). In the two data sets that had FI available, we also observed that higher *AKT2* expression was associated with lower log-transformed FI (METSIM, effect=-0.42 SD; $P$=3.3×10$^{-11}$; N=770; EuroBATS, effect=-0.038 SD; $P$=1.1×10$^{-3}$; N=710) [**Supplementary Table 5**].

We then expanded the genetic region for expression quantitative trait locus (eQTL) analysis to a one megabase (Mb) interval centered on Pro50Thr in available adipose tissue data. In this interval we found an eQTL located near *AKT2* (locus tagged by rs11880261, a 5' UTR variant; Effect Allele Frequency (EAF)=34%) with the major allele associated with lower *AKT2* expression levels (METSIM effect = -0.40 SD; $P$=6.9×10$^{-14}$; N=770; EuroBATS effect=-0.21 SD; $P$=2.3×10$^{-8}$; N=720; GTEx effect=0.19 SD; $P$=0.08; N=94) [**Supplementary Fig. 12; Supplementary Table 5**]. The coding Pro50Thr variant is associated with lower *AKT2* expression in adipose tissue in Finns (METSIM effect=-0.98 SD; $P$=8.9×10$^{-4}$, EAF=0.83%). *AKT2* Pro50Thr sits on the same haplotype as the eQTL allele (rs11880261; $r^2$=0.002, D'=0.47 in the 1000 Genomes Finnish sample) that is associated with lower expression. However, a reciprocal conditional analysis showed that they were independent signals (Pro50Thr: $P_{conditional}$=8.4×10$^{-3}$; eQTLs: $P_{conditional}$=1.9×10$^{-13}$) [**Supplementary Fig. 12**]. No association was detected between rs11880261 and FI levels (METSIM $P$=0.3, N=10,081; EuroBATS $P$=0.78, N=710), suggesting that the common variant eQTL does not drive the initial FI association. Also, Mendelian randomization with rs9749056 as an instrumental variable for *AKT2* expression indicated no causal relationship between *AKT2* expression and FI ($P$=0.41) **[Supplementary Table 6]**.

Given the spectrum of diseases and traits linked to *AKT2 (*cancer, insulin resistance, and lipodystrophy[25,26,43]), we used electronic health records available in the Finnish cohorts (METSIM and FINRISK) to further characterize the impact of *AKT2* Pro50Thr on disease risk. We found no obvious patterns of association [**Supplementary Table 7**]. Nor did we find evidence for enrichment of low-frequency associations in any *AKT2* related pathways or genes implicated in monogenic forms of glycemic disease [**Supplementary Note 3**; **Supplementary Table 8**; **Supplementary Table 9**; **Supplementary Fig. 13; Supplementary Fig. 14**]. However, in QTL analysis in the initial discovery and replication cohorts, we did observe a constellation of features indicative of a milder 'lipodystrophy-like phenotype' associated with the rare allele: an association with increased 2-hour insulin values (effect=0.23 SD of log-transformed 2-hour insulin, 95% CI=0.095-0.37; $P$=7.9×10$^{-8}$, N=14,150), a significant lowering of insulin sensitivity (effect=-0.34 SD of the log-transformed Matsuda index[44], 95% CI=-0.48 to -0.21, $P$=1.2×10$^{-6}$, N=8,566), and an increased risk of T2D (odds ratio (OR)=1.05, 95% CI=1.01-1.09, $P$=8.1×10$^{-5}$; 9,783 T2D cases; 22,662 controls), although no effects on either fasting nor postprandial glucose measures were detected [**Supplementary Table 10; Supplementary Fig. 15**].

**Discussion**

Meta-analyses of exome sequence and array genotyping data in up to 38,339 normoglycemic individuals enabled the discovery, characterization, and functional validation of a FI association with a low-frequency *AKT2* coding variant. This work extends the understanding of the genetic and molecular architecture of glucose

homeostasis. Rare, penetrant variants in genes encoding components of the insulin signaling pathway, including *AKT2*, cause monogenic but heterogeneous glycemic disorders[45]. The *AKT2* Pro50Thr association also shows an effect 5-10 times larger than those of previous published common-variant FI associations[4,5]. These results expand the allelic spectrum for *AKT2* coding variants associated with glucose homeostasis into the low-frequency range and highlight the effects of both locus and allelic heterogeneity [**Fig. 4**].

Individuals of one particular ancestry, the Finnish, drove the AKT2 Pro50Thr association signal. This demonstrates the value of association studies in different ancestries where frequencies of rare alleles may increase due to selective pressure or stochastic changes from population bottlenecks and genetic drift. Since there was no evidence of selection on the Pro50Thr variant, it most likely rose to a higher frequency due to genetic drift and exists within the unique spectrum of rare and low-frequency variation observed in Finland[46], the excess of which facilitates the study of complex trait associations[47]

The *AKT2* Pro50Thr allele shows no effect on fasting or postprandial glucose measures, but a strong effect on fasting and postprandial insulin measures, and modest increased T2D risk (OR=1.05). Similarly, all reported carriers of the lipodystrophy causing *AKT2* Arg274His allele are hyperinsulinemic, and three of the four carriers have diabetes mellitus[25]. This is similar to observed associations in *TBC1D4* (which encodes a protein that acts as a substrate immediately downstream of AKT2 in the PI3K pathway) of a population specific, protein-truncating variant (Arg684Ter) that is associated with massively increased T2D risk (OR = 10.3), increased postprandial glucose and insulin levels, and a modest decrease in FI and FG levels[8] [**Figure 4**]. Another rare stop codon allele in *TBC1D4,* Arg363Ter (not observed in the ExAC database[16]) has been reported in a family in which the proband presented with a modest elevation in FI levels but extreme postprandial hyperinsulinemia and acanthosis nigricans[48]. siRNA-mediated gene knock-down of *AKT2* in human primary myotubes completely abolishes insulin action on glucose uptake and glycogen synthesis[49], which highlights the importance of intact AKT2-TBC1D4 signaling pathway in the regulation of insulin sensitivity in humans. *TBC1D4* is ubiquitously expressed with adipose and skeletal muscle tissue ranking among the tissues with highest expression[36]. *TBC1D4* Arg363Ter seems to have an effect in adipocytes[48], while Arg684Ter falls in an exon that is exclusively expressed in skeletal and heart muscle[8,50]. This is a likely cause of the *TBC1D4* Arg684Ter tissue specificity, which appears to differ from the other *TBC1D4* Arg363Ter variant as well as the *AKT2* variants.

The phenotypes exhibited by carriers of rare, penetrant *AKT2* alleles reflect differential AKT2 activation with kinetically inactivating variants resulting in hyperinsulinemia and lipodystrophy while kinetically activating

variants lead to hypoglycemia[23-25]. Our data support *AKT2* Thr50 as a *partial* loss of function variant. The decrease of cellular proliferation demonstrates that the downstream signaling changes caused by AKT2-Thr50 are sufficient in hepatocytes to impair AKT2 function at the cellular level. As this loss of function variant does not completely lose its signaling potential or regulation of cellular processes but maintains varying portions of regulatory capacity, it is more appropriately titled *partial* loss of function.

Our data showcase bidirectional variant effects within the pleckstrin homology domain of AKT2, and we identify *AKT2* Pro50Thr as a homolog-specific inactivating *AKT2* variant. While the Pro50 residue is conserved in AKT2 throughout all vertebrates the variant lies within the pleckstrin homology domain that is not conserved between AKT isoforms [**Fig. 2**]. These residues, harboring the Pro50 variant, may thus distinguish AKT2 from AKT1 and AKT3 function. Although AKT isoforms are activated in the same mechanism within the PI3K pathway downstream of insulin, the *Akt2*$^{-/-}$ mouse is the only knockout of the gene family to be characterized by insulin resistance and diabetes[21,51-53]. A deeper understanding of what makes the AKT2 isoform distinct could offer potential sites for therapeutic intervention and enable more targeted approaches to disease prevention.

## Figures

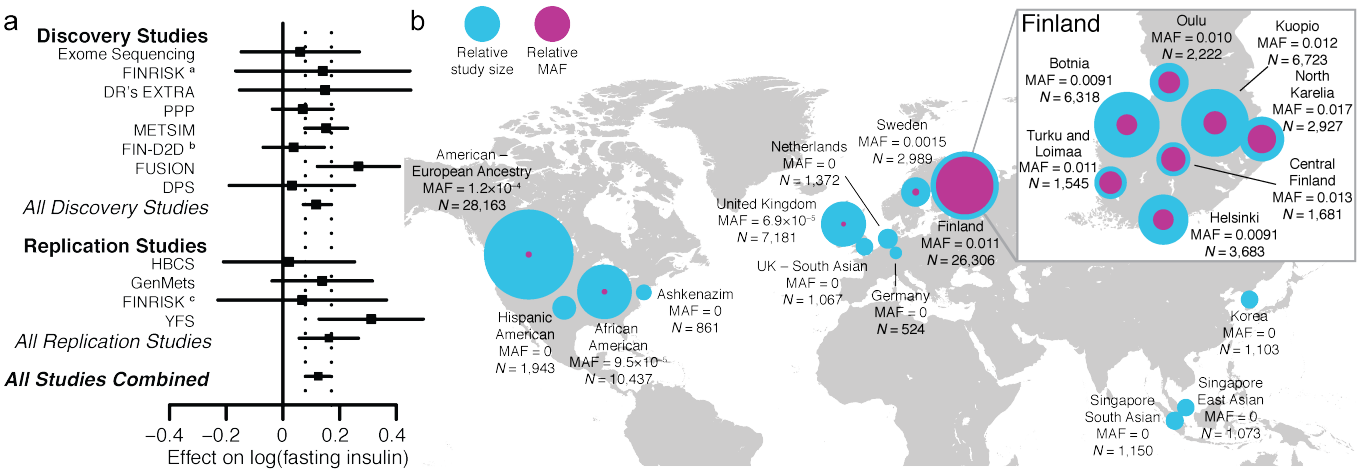


**Fig. 1. *AKT2* Pro50Thr association with FI levels**. (A) For each study, the square represents the estimate of the additive genetic effect for the association of the *AKT2* Pro50Thr allele with log-transformed FI levels and the horizontal line gives the corresponding 95% confidence interval of the estimate. Inverse-variance meta-analyses were performed for *All Discovery Studies*, *All Replication Studies* and *All Studies Combined*. The vertical dashed lines indicate the 95% confidence interval for the estimate obtained in the meta-analysis of *All Studies Combined*. (B) Using all available genotype information on the *AKT2* Pro50Thr variant, MAF was estimated for each available region and ancestry [**Supplementary Table 1**]. Across the world, the MAF ranges from 0% to 1.1%. The relative sample sizes (N) for each region/ancestry are displayed with the blue circles and the relative minor allele frequencies of *AKT2* Pro50Thr are displayed with the purple circles, with the size of the circles showing comparative differences. Within Finland (inset), where the MAF ranges from 0.9% to 1.7%, birthplace and study center data were used to show the allele distribution across the country. [a] FINRISK 2007; [b] FIN-D2D 2007; [c] FINRISK 1997 and 2002

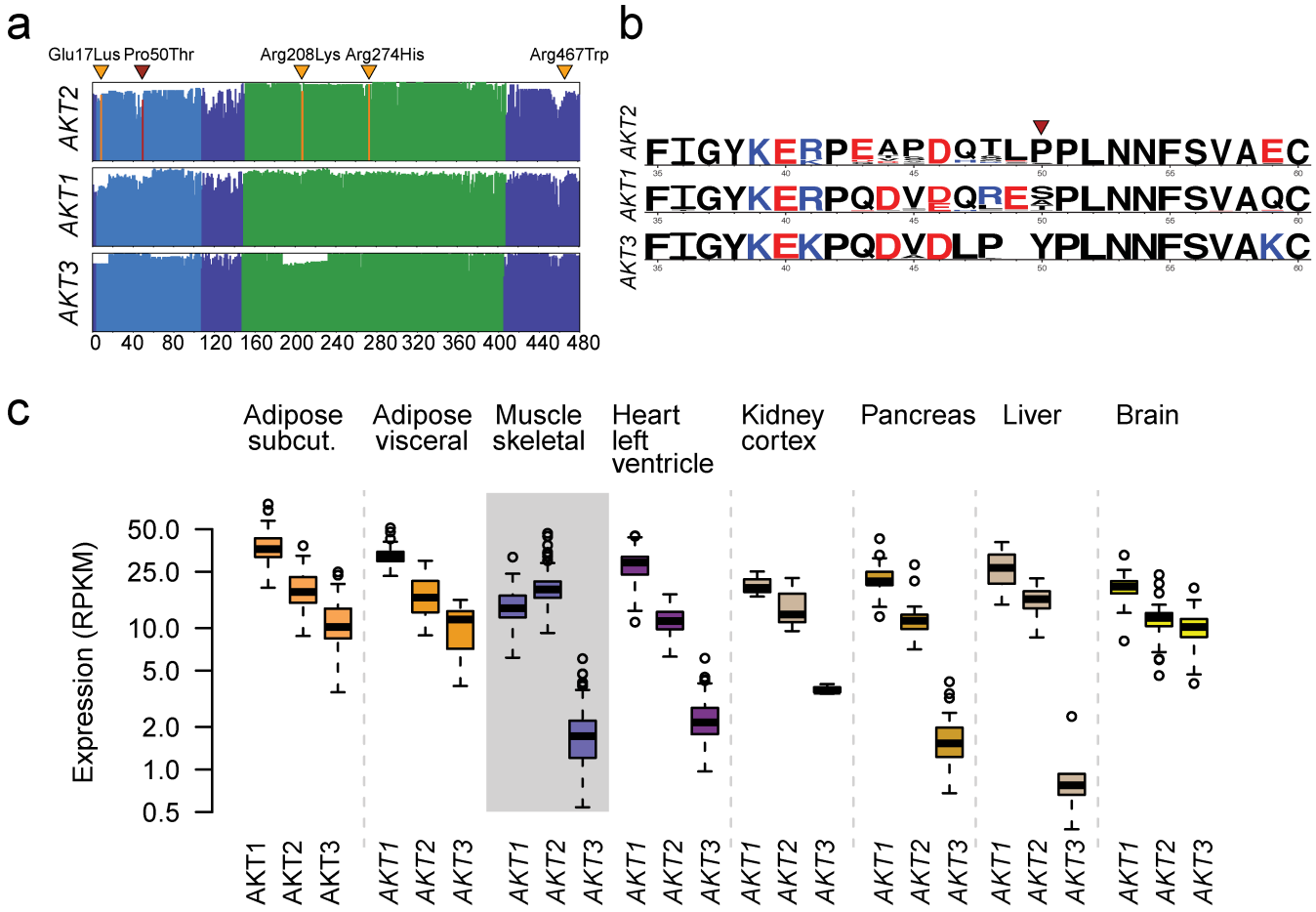


**Fig. 2. Expression and conservation properties.** (A) Amino-acid alignment and conservation of the three AKT proteins in vertebrates. The X-axis gives the amino acid position and the height of the lines depicts the conservation score across 100 vertebrate genome alignments. The functional domains are the pleckstrin homology (PH) domain (blue) and the kinase domain (green). The position of AKT2 Pro50Thr is shown in red while the locations of the other *AKT2* disease-causing mutations[23-26] are shown in orange: Glu17Lys, Arg208Lys, Arg274His and Arg467Trp. (B) WebLogo plots of amino acids 35-60 are shown for AKT2, AKT1, and AKT3 contrasting the homology of the three isoforms. The height of letters gives the relative frequency of different amino acids across the 100 vertebrate species, with the colors showing amino acids with similar charge. (C) Expression of *AKT1*, *AKT2*, and *AKT3* in eight insulin-sensitive tissues using RNA sequencing data from the GTEx consortium[a]. Observations per tissue range from 3 to 156 [see Supplementary Fig. 11 for all 44 tissues]. The Y-axis shows the gene expression in reads per kilobase of mapped reads (RPKM) units. [a] The brain expression displayed represents the mean expression per individual across the available brain subregions (maximum 11 regions).

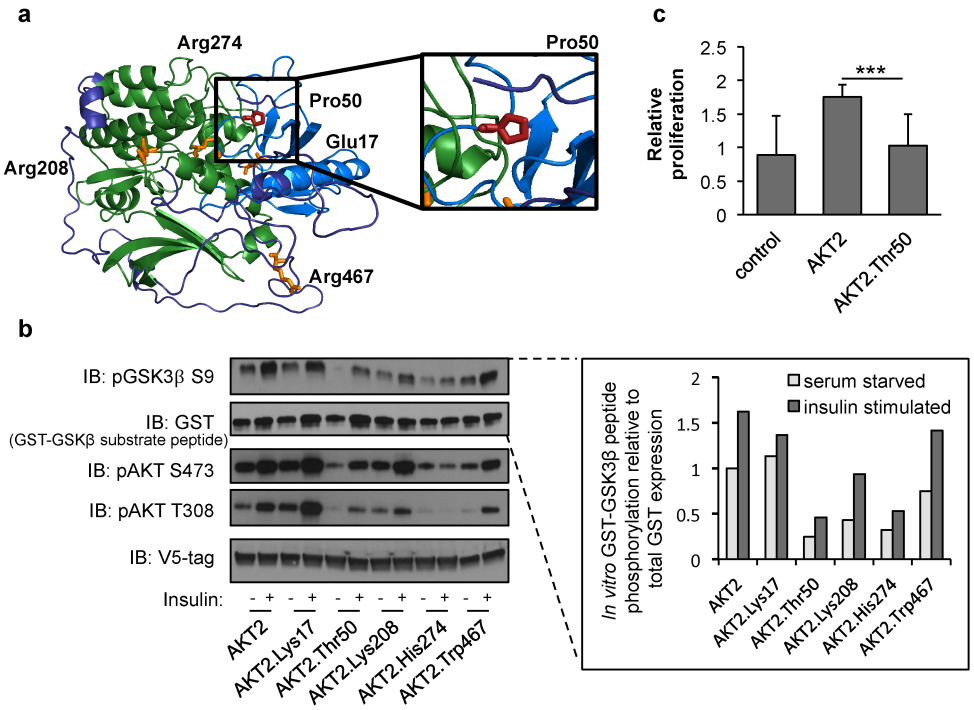


**Fig 3. Functional properties of AKT2-Thr50** (A) Predicted protein structure of AKT2. Domain and variants are highlighted as in Fig. 2. The relative spatial positioning of the AKT2-Pro50 residue is magnified within the inset. (B) HeLa cells were infected with lentiviral V5-AKT2, V5-AKT2-Lys17, V5-AKT2-Thr50, 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 incubated with GSK3*β*-GST peptide in an *in vitro* kinase (IVK) assay. Quantification of phosphorylated substrate peptide (pGSK3β) relative to total peptide (GST-GSK3*β*-) is shown at the inset. Immunoblots and quantification shown are representative of three independent replicates. Linear model (LM) statistical analyses across all three independent replicates are available in Supplementary Fig. 6. The IVK was immunoblotted (IB) with the indicated antibodies. (C) HuH7 cells were infected with lentiviral V5-AKT2, V5-AKT2-Thr50, or control pLX304. At 72 hr. relative cellular proliferation was determined with WST-1 assay of HuH7 cells. Error bars represent the standard deviation (SD). * P<0.05, ** P<0.01, *** P<0.001.

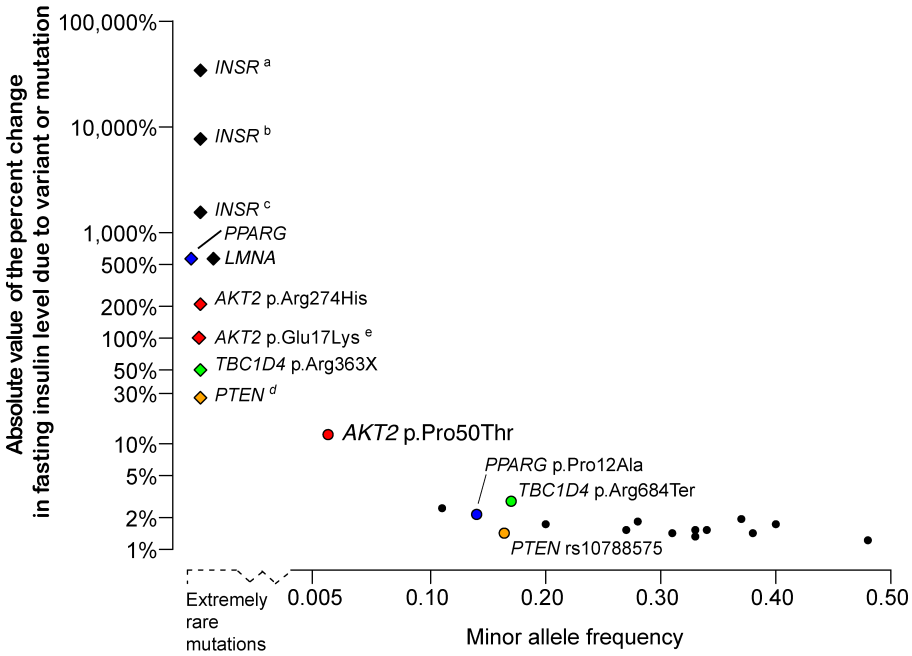


**Fig. 4. Genetic architecture of rare, low frequency, and common variants associated with FI levels**. In this plot, the absolute values of the percent change in fasting insulin level due to rare monogenic mutations (diamonds) and common genetic variants (circles) are plotted against the minor allele frequency of the variant. The extremely rare monogenic mutations (above the dashed line to the left of the x-axis) were observed in 2 to 18 individuals[23-26,48,54-56] with the height of the point indicating the percent change in fasting insulin levels of mutation carriers from 40 pmol/L, an estimate of population mean fasting insulin level. Mutations in *INSR* and *AKT2* p.Arg274His cause compensatory hyperinsulinemia, individuals with *TBC1D4* p.Arg363Ter show normal fasting insulin levels but postprandial hyperinsulinemia, and mutations in *PTEN* cause enhanced insulin sensitivity providing protection against T2D. For common variants, the additive genetic effects on log-transformed fasting insulin levels are plotted above the solid horizontal axis. These observations are from sequencing[8] and array-based GWAS[5]. For several genes, the effects from rare mutations can be compared to the effects of common variants in or near the gene: *PPARG* (blue), *TBC1D4* (green), *PTEN* (orange), and *AKT2* (red). [a] Donohue syndrome: Biallelic loss-of-function mutations in *INSR*[56]. [b] Rabson-Mendenhall syndrome: Biallelic loss-of-function mutations in *INSR*[56]. [c] Post-pubertal severe IR: Heterozygous or homozygous loss-of-function mutations in *INSR*[56]. [d] Loss of function *PTEN* mutations cause Cowden Syndrome in which carriers exhibit a *lowered* fasting insulin level (mean=29 pmol/l) compared to matched controls[54]. [e]Carriers with the *AKT2* p.Glu17Lys mutation were described with hypoinsulinemic hypoketotic hypoglycemia and hemihypertrophy with undetectable serum insulin[23,24].

## Methods

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. Detailed description of ethical permissions is provided in the Supplementary Material.

DISCOVERY ANALYSIS

*Contributing studies*

The Genetics of Type 2 Diabetes (GoT2D) Consortium included 14 studies contributing exome array information on 33,231 non-diabetic individuals of European ancestry [**Supplementary Table 1**]. We further analysed GoT2D and Type 2 Diabetes Genetic Exploration by Next-generation sequencing in Ethnic Samples (T2D-GENES) studies with exome sequence data in five ancestral groups comprised of 12,940 individuals (6,504 with T2D, 6,436 without). The GoT2D and T2D-GENES were initially designed to evaluate the contribution of coding variants to T2D risk and measured fasting glucose (FG) or fasting insulin (FI) levels were available in 1,826 European, 508 South Asian, 1,105 East Asian, 853 Hispanic, and 508 African American non-diabetic individuals [**Supplementary Table 1**].

*Phenotypes*

For the discovery analysis with FI and FG, we excluded individuals from the analysis if they had a diagnosis of T2D, were currently receiving oral or injected diabetes treatment, had FG measures ≥ 7 mmol/L, had 2-hour post-load glucose (2hrG) measures ≥ 11.1 mmol/L[57], or hadHbA1c measures ≥6.5%[57]. Additional exclusions occurring at the study level include pregnancy, non-fasting at time of exam, type 1 diabetes (T1D), or impaired glucose tolerance (IGT) [**Supplementary Table 1A**].

**Trait transformations**: Within each study, we adjusted FG and log transformed FI levels were adjusted for age, sex, body mass index (BMI), and any additional study specific covariates. Due to observing increased power for FI discovery in genome-wide association studies (GWAS), we included BMI as a covariate[5,58]. We applied rank-based inverse-normal transformations to study- or ancestry-specific residuals to obtain satisfactory asymptotic properties of the exome-wide association tests.

*Genetic data processing and quality control (QC)*

**Exome array**: All studies genotyped samples using the Illumina HumanExome BeadChip array (http://genome.sph.umich.edu/wiki/Exome_Chip_Design). See **Supplementary Table 1** for detailed genotype calling and QC procedures for each study.

**Exome sequencing**: Exome hybrid capture was performed using the Agilent SureSelect All Exon Kit v.2. Sequence reads were processed and aligned to the reference genome (hg19) with Picard (http://picard.sourceforge.net) and achieved 82-fold mean coverage, and an estimated genotyping rate of 99.2%, across the coding sequence of 18,281 genes. Polymorphic sites and genotypes were called with GATK[59]. We removed samples and variants based on multiple QC metrics, including: concordance with array data, mean heterozygosity and homozygosity across samples, high singleton counts per sample, Variant Quality Score Recalibration (VSQR) for SNVs, and hard filtering for INDELs. Additional QC was done prior to association analysis: within each ancestry group, we excluded variants with call rate < 90%, deviation from Hardy Weinberg Equilibrium ($P < 1 \times 10^{-6}$), or differential case-control call rate ($P < 1 \times 10^{-4}$). We identified ethnic outliers for exclusion through the visual inspection of principal components (PCs) from the EIGENSTRAT[60] software package.

*Single-variant association analysis*

We performed association analyses within each study for the exome array data sets and within ancestry for the exome sequence data sets. We used linear mixed models implemented in the EMMAX[61] to account for relatedness. Within each study/ancestry, we excluded variants with fewer than five rare alleles. The discovery analysis proceeded in two steps. First, we meta-analyzed the single-variant results from the (European-ancestry) exome array studies using the fixed-effects inverse variance meta-analysis approach implemented in METAL[62]. These results were combined with the European ancestry exome sequence results with another fixed-effects inverse variance meta-analysis. Second, we combined summary statistics across ancestries using inverse variance meta-analysis and a Bayesian method (MANTRA)[63]. MANTRA accounts for heterogeneity in allelic effects between populations and presents statistical evidence for association as a Bayes' Factor (BF). We used $P<5\times10^{-7}$ or $\log_{10}BF>5$ as exome-wide statistical significance thresholds for the single variant tests. This threshold is derived from power calculations showing adequate power to detect an effect at least 0.3 standard deviations (SD)[10] with the ~100,000 variants with MAC>5 available for analysis. We observed consistent sets of variants with statistically significant associations across the two analysis methods. When reporting the association statistics for loci with previously reported associations with FG or FI, we calculated an enrichment P-value for the number of non-significant variants with consistent direction of effects using the binomial distribution.

*Gene-based association analysis*

For variants with MAF < 1% within every ancestry group and all protein truncating variants regardless of MAF, we aggregated multiple variants across defined gene units[64].

**Variant grouping into 'masks'**: We mapped variants to Ensembl 66 (GRCh37.66) and annotated using CHAoS (v 0.6.3), SnpEff (v 3.1), VEP (v 2.7) and dbNSFP (v2)[27,28,65,66]. Using consensus transcript mapping by

these three annotation methods, we identified variants predicted to be protein-truncating (PTV; e.g. nonsense, frameshift, and essential splice site variants) or protein-altering (NS; e.g. missense, in-frame indel, and non-essential splice site variants). We defined four different variant groupings, or masks[12,13]) "PTV-only", containing only variants predicted to severely impair protein function, 2) "PTV+missense", containing PTV and NS variants with MAF <1%, 3) "PTV+$NS_{strict}$" composed of PTV and NS variants predicted damaging by 5 prediction algorithms (SIFT, LRT, MutationTaster, polyphen2 HDIV, and polyphen2 HVAR), and 4) "PTV+$NS_{broad}$" composed of PTV and NS variants with MAF<1% and predicted damaging by at least one prediction algorithm above, with the last two being analogous those described by Purcell *et al.*[12].

**Association analysis**: We conducted exome array gene-based meta-analyses in an unrelated subset of individuals using RareMETALS[67,68] using study-specific score statistics and covariance matrices generated by rvtests or RareMetalWorker. We implemented the sequence kernel association test (SKAT)[69] and BURDEN tests and adjusted for population structure with PCs. We conducted exome sequence gene-based analyses within ancestry (assuming homogeneous allele frequency and effects) using a linear mixed model to account for relatedness and combined results across ancestries with MetaSKAT[70], which accounts for heterogeneous effects. We further combined gene-based results from exome array and exome sequences using Stouffer's method with equal weights. For gene-based tests, we have considered $P< 2.5x10^{-6}$ as exome-wide significant, corresponding to Bonferroni correction for 20,000 genes in the genome[10].

## EXPLORING ASSOCIATION IN GWAS REGIONS

Through the NHGRI GWAS database[71] and literature searches for genetic associations with glycemic traits in diverse ancestries[4,5,10,58,72-80] we identified 72 loci containing variants within regions with reported association reaching genome-wide significance ($P<5\times10^{-8}$) with FG or FI. We defined these regions (using 1000 Genomes Phase I data[81]) by extending upstream and downstream from the GWAS signal first to the two most distant flanking variants with an $r^2>0.5$ from the index SNP, then 0.02 cM to capture recombination hotspots, and finally an additional 300 kb.

## ALLELE FREQUENCY DISTRIBUTION OF *AKT2* PRO50THR

### *Independent Finnish studies*

**Study descriptions:** We included data from 5,948 individuals in four Finnish studies: Cardiovascular Risk in Young Finns Study (YFS)[82], Helsinki Birth Cohort (HBCS)[83], Health 2000 GenMets Study (GenMets)[84], and National FINRISK Study 1997 and 2002 (FR)[85]. Ascertainment criteria for each study are listed in **Supplementary Table 1A**.

**Genotyping, imputation and imputation quality:** The genotyped SNPs (**Supplementary Table 1**) were augmented by genotype imputation using IMPUTE (v2.2.2)[86,87] and the 1000 Genomes project version 3 reference panel (March 2012). Genotype data were imputed in genomic regions of ~5 Mb with additional 1 Mb region on either side using effective population size 11,418 and k value 80. We compared the imputed and the directly genotyped exome array genotypes available in a subset of the HBCS cohort (N=1,570) to verify imputation accuracy. For the *AKT2* variant, the correlation between imputed and directly genotyped was high ($r^2 > 0.88$).

*CHARGE studies*

We compiled information on *AKT2* p.Pro50Thr allele frequency in CHARGE studies with available exome array data[14]. The minor allele of *AKT2* Pro50Thr was observed in four of the 20 studies with European ancestry (MAC=10, range per study 1 to 4; total N = 37,798) and one of the seven studies with African American ancestry (MAC=2, total N = 8,860). None of the studies passed our QC filter of a MAC>5 for inclusion in meta-analysis (Fig. 1B).

*Geographic distribution of AKT2 p.Pro50Thr:*

We calculated the allele frequency of *AKT2* Pro50Thr using data from discovery data, CHARGE cohorts, and independent Finnish cohorts. Individuals with T2D were included in these calculations. Within Finland, individuals were further categorized by birthplace when available (FIN-D2D 2007, FINRISK, FUSION, and METSIM) or study site.

REPLICATION OF *AKT2* PRO50THR ASSOCIATION WITH FI

**Single-variant association analysis:** The *AKT2* Pro50Thr variant was observed at sufficient frequency in the independent Finnish cohorts to perform single-variant association test of association with FI. We tested association in SNPTEST[88] (v.2.4.0) in each study with the same additive linear model used in the discovery analysis. Covariate adjustments for FI levels were sex, age, and ten PCs, and models were run with and without adjustment for BMI. As in the discovery analysis, we used rank-based inverse normalized residuals.

**Estimate of effect on raw FI level and variance explained**. To characterize the association between AKT2 Pro50Thr and FI, we examined full regression models with raw FI in three studies (FUSION, METSIM, and YFS). We estimated the raw effect on FI levels with a fixed-effects meta-analysis. The variance in log-transformed FI explained by *AKT2* Pro50Thr was estimated by a weighted average of the narrow-sense heritability of *AKT2* Pro50Thr seen in these three studies.

POPULATION GENETICS OF *AKT2*

*Protein-altering Genetic Variation in AKT2*

**T2D-GENES Exome Sequence Data:** We catalogued all protein altering variants observed in the T2D-GENES exome sequence data (N=12,940) with: the observed minor allele counts; lookups of allele counts in the 1000 Genomes data set[89] and the ExAC database[16]; predicted deleteriousness obtained from dbNSFP (SIFT, LRT, MutationTaster, and Polyphen 2)[27,28]; the tissue of a reported somatic mutation in cBio[90,91]; phenotype for previously reported monogenic mutation, and functional domain from UniProt[92].

*Conservation and Diversity*

**Constraint metric:** The ExAC consortium provides constraint metrics for each gene (ftp://ftp.broadinstitute.org/pub/ExAC_release/release0.3/functional_gene_constraint/)[16] . The constraint metric is defined as a signed Z score for the deviation of observed missense counts in a gene from the expected number of missense counts in the gene. Positive Z scores indicate increased constraint (intolerance to variation) and therefore that the gene had fewer variants than expected. Negative Z scores are given to genes that had a more variants than expected. To determine constraint on loss of function (LoF) variation within a gene, genes were categorized into three categories (tolerant of LoF; recessive; and haploinsufficient (or LoF intolerant). The probability that a gene is LoF intolerant (pLI) is calculated using the observed and expected variants counts. The closer pLI is to one, the more LoF intolerant the gene appears to be. The set of genes with pLI >= 0.9 as an extremely LoF intolerant set of genes.

**Conservation score and logo:** We obtained sequence alignments for AKT proteins and mRNAs in 100 vertebrates from the UCSC Genome Browser[93]. We used Shannon's entropy (normalized K=21) as a conservation score[94] and plotted the sequence logos in R using the RWebLogo library[95].

**Haplotype and diversity indices:** We phased variants in proteins or genes (including non-coding variants) using SHAPEIT[96] and calculated population statistics and diversity indices with Arlequin (v 3.5)[97], grouped by country of origin. Finally, we built the haplotype network using the pegas and igraph libraries in R.

**dN/dS and Mcdonald-Kreiman test:** dN/dS for Human-Chimpanzee alignments were extracted from ENSEMBL database[98]. We computed the "within-human" dN/dS with codeml (PAML)[99] using hg19 sequence as reference and alternative sequence containing all the observed segregating sites. The dN/dS was computed for the three *AKT* genes, but also for three lists of genes: Monogenic Insulin, used in monogenic analyses, 1,002 anatomical structure development genes ("conserved"), and 132 sexual reproduction genes ("fast evolving"). The Mcdonald-Kreitman test[100] for *AKT2* was computed in Bioperl (Bio::PopGen::Statistics) using *AKT3* (hg19) as an outgroup.

FUNCTIONAL WORK

*3D modeling*

The 3D structure of AKT2 with the full allelic series was predicted using IntFOLD[101,102] and visualized in PyMOL[103].

*Plasmids and Cell Lines*

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 [**Supplementary Note 4**].

HeLa, HuH7, and 293T cells were obtained within The Broad Institute and maintained in 10%FBS DMEM, 100 U/ml penicillin and 100 µg/ml streptomycin, and documented mycoplasma-free. HeLa and HuH7 cells were starved for 18 hr and stimulated for 15 min with 100 nM insulin for activation analyses.

**Antibodies:** Anti-Akt (#4685), anti-phospho-Akt S473 (#4060), anti-phospho-Akt T308 (#9275), anti-β Actin (#4970), anti-GSK3β (#9315), anti-phospho-GSK3β (#9336), anti-GST (#2625), and anti-V5 (#13202) were purchased from Cell Signaling Technologies (product numbers listed for each). Horseradish peroxidase-conjugated anti-rabbit and anti-mouse immunoglobulin G (IgG) antibodies were purchased from Millipore.

***In vitro* kinase assays:** We isolated pLX304 control, V5-AKT2, V5-AKT2.Lys17, V5-AKT2.Thr50, V5-AKT2.Lys208, V5-AKT2.His274, and V5-AKT2.Trp467 lentivirally infected variants from HeLa cell lysate with V5 agarose beads (SIGMA) and incubated with 150ng GST-GSK3β substrate peptide (Cell Signaling Technologies) and 250 mM cold ATP in kinase assay buffer (Cell Signaling Technologies) for 35 min at 30°C.

**Proliferation assay:** We cultured lentiviral pLX304 V5-AKT2 variants and control empty vector infected HuH7 cells in 24 well plate for 72 hr in 10%FBS /phenol red-free DMEM for 72 hr. We added WST-1 (Takara Clontech) to each well at the manufacture recommended 1:10 ratio and incubated for 4 hr at at 37°C prior to absorbance measurement at 450 nm with BioTek Synergy H4 plate reader.

**Immunoblots:** We washed cells with phosphate buffered saline and lysed in EBC buffer (120mM NaCl, 50mM TRIS-HCl (pH7.4), 50nM calyculin, cOmplete protease inhibitor cocktail (Roche), 20mM sodium fluoride, 1mM sodium pyrophosphate, 2mM ethylene glycol tetraacetic acid, 2mM ethylenediaminetetraacetic acid, and 0.5%NP-40) for 20 min on ice. To preclear cell lysates, we centrifuged at 12,700 rmp at 4°C for 15 min. We measured protein concentration with Pierce BCA protein assay kit using a BioTek Synergy H4 plate reader. We resolved Lysates on BioRad any kD mini-PROTEAN TGX polyacrylamide gels by SDS-PAGE and transferred by electrophoresis to nitrocellulose membrane (Life Technologies) at 100V for 70 min. We blocked membranes in 5% nonfat dry milk/ TBST (10 mM Tris-HCl, 150 mM NaCl, 0.2% Tween 20) buffer pH 7.6 for 30 min. We incubated blots with indicated antibody overnight at 4°C. The membrane was then washed in TBST, three times at 15 minute intervals, before 1 hour secondary horseradish peroxidase-conjugated antibody

incubation at room temperature. We again washed nitrocellulose membranes in TBST, three times for 15 minutes, prior to enhanced chemiluminescent substrate detection (Pierce).

**Numerical analysis:** We performed numerical analysis of the quantified assay results with general linear models in R. The graphical representation was produced using functions in the effects (v 3.0-3) package.

GENE EXPRESSION

*GTEx*

We compared the expression pattern of *AKT2* to the two other members of the *AKT* gene family, *AKT1* and *AKT3,* using multi-tissue RNA sequencing (RNA-seq) data from the pilot phase of the GTEx project. Detailed procedures for sample collection, RNA extraction, RNA-seq and gene and transcript quantifications have been previously described[36]. Briefly, in the pilot phase, a total of 9,365 tissue samples targeting more than 30 distinct human tissues were collected from 237 post-mortem donors. RNA was extracted, and 1,749 unique samples that passed QC (RIN value of 6.0 or higher and at least 1µg of total RNA), were selected for RNA-seq. Non strand-specific RNA sequencing after poly-A selection was performed using Illumina TruSeq RNA Sample Preparation protocol on the Illumina HiSeq 2000, and aligned with Tophat (v 1.4.1)[104] to UCSC hg19. Gencode (v 12)[105] was used as a transcriptome model for the alignment, and gene and isoform quantifications. Gene and exon level expression was quantified using RNA-SeQC[106] and the Flux Capacitor (v 1.2.3, http://flux.sammeth.net) was used in the quantification of the expression of several transcriptional elements including gene transcript, splice junctions and introns. In total, 44 tissues had data from more than one individual and were used in the analyses.

The comparison of expression levels of *AKT2* versus *AKT1*, and *AKT2* versus *AKT3* was performed using log2-transformed reads per kilobase per million mapped reads (RPKMs). The percent increase in AKT2 expression was calculated with the following formula: 2^log-fold-change(AKT2 vs AKT1).

**Genotyping and imputation:** Samples were genotyped on the Illumina HumanOmni5-4v1_B SNP array and imputed to the 1,000 Genomes Phase 1 reference (an updated data freeze version from 19 April 2012, release v3) using IMPUTE2[86,87] as recently described[36].

**Age and BMI associations:** We studied BMI and age associations using a linear mixed model as implemented in the lmer function in the lme4 R package[107]. Sex, age, BMI, and three PCs were included in the model as fixed covariates and the date of sequencing and the date of nucleic acid isolation as random covariates. The gene expression RPKM values were inverse variance rank normalized for these analyses.

**eQTL analysis**: The cis-eQTL for AKT2 in subcutaneous adipose tissue was extracted from the eQTL data generated during the pilot phase of the GTEx project. The methods have been previously described in detail[108]. Briefly, the association of common (MAF ≥ 5%) SNPs with gene expression levels was studied using a linear

model in MatrixEQTL[109] including sex, three genotyping PCs, and 15 expression PEER factors[110] as covariates. The cis-window was defined as one megabase (Mb) up- and down-stream of the transcription start site of each transcript. Prior to the eQTL analysis the RPKM values were quantile normalized across genes within each tissue and transformed into a standard normal based on rank.

*EuroBATs*

**EuroBATs RNA-seq samples:** Samples from photo protected subcutaneous adipose tissue from 766 twins were extracted (131 monozygotic twin pairs, 187 dizygotic twin pairs and 130 unrelated individuals) and processed as previously described[41,111]. In short, samples were prepared for sequencing with the Illumina TruSeq sample preparation kit (Illumina, San Diego, CA) according to manufacturer's instructions and were sequenced on a HiSeq2000 machine. Afterwards, the 49-bp sequenced paired-end reads were mapped to the GRCh37reference genome[112] with BWA v0.5.9[113]. We use genes defined in the GENCODE 10 annotation[105], removing genes with more than 10% zero read count. RPKM values were root mean transformed.

**Genotyping and imputation**: Samples were genotyped on a combination of the HumanHap300, HumanHap610Q, 1M-Duo and 1.2MDuo 1M Illumnia arrays, as described in Grundberg *et. al*[114]. Samples were imputed into the 1000 Genomes Phase 1 reference panel (data freeze, 10/11/2010)[5] using IMPUTE2[86,87] and filtered (removing variants with MAF<1%, IMPUTE info value<0.8). Samples with both genotypes and expression values (N=720) were used in the subsequent analyses.

**Gene-age, gene-BMI and insulin associations:** Rank normalized RPKM values were used to assess the variables age and BMI effect on gene expression. We fit a linear mixed model to examine age and BMI effects on gene expression in R[115] with the lmer function in the lme4 package[107]. Confounding factors in all models included fixed effects (primer insert size, GC content mean) and random effects (primer index, date of sequencing, family relationship and zygosity). The P values to assess significance for age and BMI effects were calculated from the Chi-square distribution with 1 degree of freedom using likelihood ratio as the test statistic. The full models for comparison were for age: expression ~ age + BMI + fixed covariates + random covariates; and for BMI: expression ~ BMI + age + fixed covariates + random covariates. FI was measured at the same time point as the fat biopsies, following a previously described protocol [116]. Natural log transformed FI were adjusted for age or for age and BMI and the residuals were inverse rank normalized. FI-SNP and FI-*AKT2* association was tested with a linear model using the lm function in R.

**eQTL analysis**: We ran the eQTL analysis on residuals from a mixed model including the first 20 PCs as fixed effects and as random effects family relationship and zygosity. SNP-expression association was performed with a t-test statistic using the software NP-GWAS (in preparation). We assessed statistical significance through 100,000 permutations.

*METSIM*

**METSIM RNA samples:** Subcutaneous fat biopsy samples were obtained from a sample of the participants of the baseline METSIM study. Total RNA was isolated from these samples using Qiagen miRNeasy Kit according to the manufacturer's instructions. RNA integrity number values were assessed with the Agilent Bioanalyzer 2100. High-quality samples (RNA integrity number>7.0) were used for transcriptional profiling with the Affymetrix Human Genome U219 Array. Genome Studio software (2010.v3) was used to obtain fluorescent intensities.

**eQTL analysis and gene-age, gene-BMI and insulin associations:** The SNP-gene associations were studied for all SNP within 1 Mb of a given gene. The RNA normalized expression data were adjusted for 35 PEER factors and inverse normal transformed PEER processed residuals were for used eQTL mapping[117]. Linear mixed model EMMAX[61] accounts for sample relatedness and was implemented in EPACTS (http://genome.sph.umich.edu/wiki/EPACTS). The sample size for eQTL-mapping was N=770. BMI and age associations, as well as FI associations (with and without adjustment for BMI) were studied using the mixed linear model implemented in lme4[107] in R. The fixed covariates including age and BMI were used as random covariates. Association between the SNPs associated with *AKT2* expression (eSNPs) and FI was tested with a linear model using the lm() function in R. The natural log transformed FI levels were adjusted for age and BMI and the residuals were inverse rank normalized. All analyses using expression data were conducted in 770 METSIM individuals, while for the tests of eSNP and FI association the sample size for analysis was 10,081.

*Effect estimates for expression analyses*

The association effects between *AKT2* expression and age, BMI, FI, and the SNPs from the eQTL analysis are reported in standard deviation (SD) units of normalized RPKM expression values.

*Mendelian Randomization*

To elaborate the potential causality behind the association between *AKT2* expression and FI association we applied a Mendelian randomization based approach using the discovered eSNPs as instrumental variables (IV) following a similar procedure as described recently[118]. The association data for the SNP-gene, gene-FI and SNP-FI analyses from EuroBATS and METSIM were first combined in a fixed-effects inverse-variance-weighted meta-analysis. Using the meta-analyzed results the IV estimator was derived by taking the ratio of the regression coefficients from the SNP-FI and SNP-*AKT2* analyses, and the standard error for the estimator was estimated using the delta method. The significance of the IV estimator and the difference between the IV estimator and the observational estimator, i.e. the regression coefficient for the *AKT2*-FI association, were tested using a Z test. The Mendelian randomization analysis was conducted separately for both eSNPs using both the BMI-adjusted and BMI-unadjusted analysis results.

ASSOCIATION OF *AKT2* PRO50THR WITH RELATED TRAITS AND DISORDERS

To expand the phenotypic profile of *AKT2* Pro50Thr carriers, we looked at the association with T2D, hypertension, and other related quantitative traits in the Finnish studies. We analyzed lipid levels (total cholesterol, high-density lipoprotein cholesterol (HDL), low-density lipoprotein cholesterol (LDL), and triglycerides (TG)), blood pressure (systolic (SBP) and diastolic (DBP) blood pressures and hypertension (HTN)), height, BMI, central adiposity measures (waist-to-hip ratio (WHR), waist circumference, hip circumference), T2D, adiponectin level, 2-hour insulin level and Matsuda index, which provides an estimate of insulin sensitivity in the liver and peripheral tissues such as skeletal muscle and adipose tissue and is known to correlate with the hyperinsulinemic euglycemic clamp ($r$=0.73, $P$<0.0001)[44]. For quantitative traits and HTN, we adjusted for age, sex, BMI (for glycemic, blood pressure and central adiposity traits), stratified by T2D status (and sex for central adiposity measures) within study. We adjusted LDL and total cholesterol for use of lipid lowering medication, by dividing total cholesterol by 0.8 if on lipid lowering medication, prior to calculating LDL using the Friedewald equation[119]. SBP and DBP were adjusted for use of blood pressure lowering medication[120]. Matsuda Index was log transformed and analyzed in non-diabetics only. After adjusting for covariates, traits were inverse-normalized within strata. Genetic associations were conducted within studies as above for both discovery studies as well as the independent Finnish studies used for replication. P values for T2D and HTN came from EMMAX[61] or the Wald test from logistic regression (Finnish replication data sets) and meta-analyzed using an N weighted meta-analysis. Odds ratios (OR) were obtained from logistic regression adjusting for age, sex, with and without BMI, and PCs and meta-analyzed using an inverse variance meta-analysis.

*Effect estimates for the associations with related traits*

These association analyses were performed with rank-normalized covariate-adjusted outcomes, so the effect of *AKT2* Pro50Thr is reported in units of standard deviation (SD) change on the trait (or log-transformed trait in the case of 2-hour insulin levels, adiponectin levels, and the Matsuda Index.)

*Clinical phenotypes*

For METSIM and FINRISK 1997 and 2002, we used ICD codes to categorize affection status for diseases *AKT2* has previously been shown to affect. To increase sample size, we included all samples with genotype and electronic medical record data available, regardless of T2D status. Because of the low number of allele carriers among individuals with the clinical phenotypes, we did not perform statistical tests of association with the *AKT2* Pro50Thr variant.

**Links to metabolic phenotypes:** We constructed a phenotype profile across 16 metabolic traits for three *AKT2* variants (Pro50Thr, p.Arg208Lys, and p.Arg467Trp) was compiled using inverse normalized, covariate adjusted traits and Z-transformed raw trait values. Traits were transformed and adjusted as previously

described. We examined the clustering of all *AKT2* missense variant carriers across seven insulin-resistance related traits (fasting insulin, fasting glucose, systolic blood pressure, body mass index, waist-to-hip ratio, fasting triglycerides and HDL-cholesterol level). Clustering was performed in R using the pheatmap libraries and Pearson correlation to calculate distances.

**References**


1. Kahn, S.E. Clinical review 135: The importance of beta-cell failure in the development and progression of type 2 diabetes. *J Clin Endocrinol Metab* **86**, 4047-58 (2001).
2. Kahn, S.E., Hull, R.L. & Utzschneider, K.M. Mechanisms linking obesity to insulin resistance and type 2 diabetes. *Nature* **444**, 840-6 (2006).
3. Phillips, D.I., Clark, P.M., Hales, C.N. & Osmond, C. Understanding oral glucose tolerance: comparison of glucose or insulin measurements during the oral glucose tolerance test with specific measurements of insulin resistance and insulin secretion. *Diabet Med* **11**, 286-92 (1994).
4. Dupuis, J. *et al.* New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet* **42**, 105-16 (2010).
5. Scott, R.A. *et al.* Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat Genet* **44**, 991-1005 (2012).
6. Sigma Type 2 Diabetes Consortium. Association of a low-frequency variant in HNF1A with type 2 diabetes in a Latino population. *JAMA* **311**, 2305-14 (2014).
7. Sigma Type 2 Diabetes Consortium. Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. *Nature* **506**, 97-101 (2014).
8. Moltke, I. *et al.* A common Greenlandic TBC1D4 variant confers muscle insulin resistance and type 2 diabetes. *Nature* **512**, 190-3 (2014).
9. Exome Chip Design; http://genome.sph.umich.edu/wiki/Exome_Chip_Design; Accessed 03/2016.
10. Mahajan, A. *et al.* Identification and Functional Characterization of G6PC2 Coding Variants Influencing Glycemic Traits Define an Effector Transcript at the G6PC2-ABCB11 Locus. *PLoS Genet* **11**, e1004876 (2015).
11. Wu, M.C. *et al.* Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet* **89**, 82-93 (2011).
12. Purcell, S.M. *et al.* A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* **506**, 185-90 (2014).
13. Zuk, O. *et al.* Searching for missing heritability: designing rare variant association studies. *Proc Natl Acad Sci U S A* **111**, E455-64 (2014).
14. Grove, M.L. *et al.* Best practices and joint calling of the HumanExome BeadChip: the CHARGE Consortium. *PLoS One* **8**, e68095 (2013).
15. Palo, J.U., Ulmanen, I., Lukka, M., Ellonen, P. & Sajantila, A. Genetic markers and population history: Finland revisited. *Eur J Hum Genet* **17**, 1336-46 (2009).
16. Exome Aggregation Consortium (ExAC), Cambridge, MA (URL: http://exac.broadinstitute.org) Accessed 02/2015.
17. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *bioRxiv* (2015).
18. Burgering, B.M. & Coffer, P.J. Protein kinase B (c-Akt) in phosphatidylinositol-3-OH kinase signal transduction. *Nature* **376**, 599-602 (1995).
19. Franke, T.F. *et al.* The protein kinase encoded by the Akt proto-oncogene is a target of the PDGF-activated phosphatidylinositol 3-kinase. *Cell* **81**, 727-36 (1995).
20. Kohn, A.D., Kovacina, K.S. & Roth, R.A. Insulin stimulates the kinase activity of RAC-PK, a pleckstrin homology domain containing ser/thr kinase. *EMBO J* **14**, 4288-95 (1995).

21. Cho, H. *et al.* Insulin resistance and a diabetes mellitus-like syndrome in mice lacking the protein kinase Akt2 (PKB beta). *Science* **292**, 1728-31 (2001).
22. Garofalo, R.S. *et al.* Severe diabetes, age-dependent loss of adipose tissue, and mild growth deficiency in mice lacking Akt2/PKB beta. *J Clin Invest* **112**, 197-208 (2003).
23. Hussain, K. *et al.* An activating mutation of AKT2 and human hypoglycemia. *Science* **334**, 474 (2011).
24. Arya, V.B. *et al.* Activating AKT2 mutation: hypoinsulinemic hypoketotic hypoglycemia. *J Clin Endocrinol Metab* **99**, 391-4 (2014).
25. George, S. *et al.* A family with severe insulin resistance and diabetes due to a mutation in AKT2. *Science* **304**, 1325-8 (2004).
26. Tan, K. *et al.* Analysis of genetic variation in Akt2/PKB-beta in severe insulin resistance, lipodystrophy, type 2 diabetes, and related metabolic phenotypes. *Diabetes* **56**, 714-9 (2007).
27. Liu, X., Jian, X. & Boerwinkle, E. dbNSFP: a lightweight database of human nonsynonymous SNPs and their functional predictions. *Hum Mutat* **32**, 894-9 (2011).
28. Liu, X., Jian, X. & Boerwinkle, E. dbNSFP v2.0: a database of human non-synonymous SNVs and their functional predictions and annotations. *Hum Mutat* **34**, E2393-402 (2013).
29. Franke, T.F., Kaplan, D.R., Cantley, L.C. & Toker, A. Direct regulation of the Akt proto-oncogene product by phosphatidylinositol-3,4-bisphosphate. *Science* **275**, 665-8 (1997).
30. Mora, A., Komander, D., van Aalten, D.M. & Alessi, D.R. PDK1, the master regulator of AGC kinase signal transduction. *Semin Cell Dev Biol* **15**, 161-70 (2004).
31. Sarbassov, D.D., Guertin, D.A., Ali, S.M. & Sabatini, D.M. Phosphorylation and regulation of Akt/PKB by the rictor-mTOR complex. *Science* **307**, 1098-101 (2005).
32. Stokoe, D. *et al.* Dual role of phosphatidylinositol-3,4,5-trisphosphate in the activation of protein kinase B. *Science* **277**, 567-70 (1997).
33. Woodgett, J.R. Recent advances in the protein kinase B signaling pathway. *Curr Opin Cell Biol* **17**, 150-7 (2005).
34. Cheung, M. & Testa, J.R. Diverse mechanisms of AKT pathway activation in human malignancy. *Curr Cancer Drug Targets* **13**, 234-44 (2013).
35. Lawlor, M.A. & Alessi, D.R. PKB/Akt: a key mediator of cell proliferation, survival and insulin responses? *J Cell Sci* **114**, 2903-10 (2001).
36. GTEx Consortium. Multi-tissue transcriptome analysis in a population sample: the Genotype-Tissue Expression (GTEx) pilot study. *submitted* (2015).
37. Jones, P.F., Jakubowicz, T. & Hemmings, B.A. Molecular cloning of a second form of rac protein kinase. *Cell Regul* **2**, 1001-9 (1991).
38. Zinda, M.J. *et al.* AKT-1, -2, and -3 are expressed in both normal and tumor tissues of the lung, breast, prostate, and colon. *Clin Cancer Res* **7**, 2475-9 (2001).
39. Peng, X.D. *et al.* Dwarfism, impaired skin development, skeletal muscle atrophy, delayed bone development, and impeded adipogenesis in mice lacking Akt1 and Akt2. *Genes Dev* **17**, 1352-65 (2003).
40. Stancakova, A. *et al.* Changes in insulin sensitivity and insulin release in relation to glycemia and glucose tolerance in 6,414 Finnish men. *Diabetes* **58**, 1212-21 (2009).
41. Buil, A. *et al.* Gene-gene and gene-environment interactions detected by transcriptome sequence analysis in twins. *Nat Genet* **47**, 88-91 (2015).
42. Chiappa, S., Winn, J., Vinuela, A., Tipney, H. & Spector, T.D. A probabilistic model of biological ageing of the lungs for analysing the effects of smoking, asthma and COPD. *Respir Res* **14**, 60 (2013).
43. Parikh, C. *et al.* Disruption of PH-kinase domain interactions leads to oncogenic activation of AKT in human cancers. *Proc Natl Acad Sci U S A* **109**, 19368-73 (2012).
44. Matsuda, M. & DeFronzo, R.A. Insulin sensitivity indices obtained from oral glucose tolerance testing: comparison with the euglycemic insulin clamp. *Diabetes Care* **22**, 1462-70 (1999).
45. O'Rahilly, S. Human genetics illuminates the paths to metabolic disease. *Nature* **462**, 307-14 (2009).
46. Peltonen, L., Jalanko, A. & Varilo, T. Molecular genetics of the Finnish disease heritage. *Hum Mol Genet* **8**, 1913-23 (1999).

47. Lim, E.T. *et al.* Distribution and medical impact of loss-of-function variants in the Finnish founder population. *PLoS Genet* **10**, e1004494 (2014).
48. Dash, S. *et al.* A truncation mutation in TBC1D4 in a family with acanthosis nigricans and postprandial hyperinsulinemia. *Proc Natl Acad Sci U S A* **106**, 9350-5 (2009).
49. Bouzakri, K. *et al.* siRNA-based gene silencing reveals specialized roles of IRS-1/Akt2 and IRS-2/Akt1 in glucose and lipid metabolism in human skeletal muscle. *Cell Metab* **4**, 89-96 (2006).
50. Baus, D. *et al.* Identification of a novel AS160 splice variant that regulates GLUT4 translocation and glucose-uptake in rat muscle cells. *Cell Signal* **20**, 2237-46 (2008).
51. Cho, H., Thorvaldsen, J.L., Chu, Q., Feng, F. & Birnbaum, M.J. Akt1/PKBalpha is required for normal growth but dispensable for maintenance of glucose homeostasis in mice. *J Biol Chem* **276**, 38349-52 (2001).
52. Toker, A. & Marmiroli, S. Signaling specificity in the Akt pathway in biology and disease. *Adv Biol Regul* **55**, 28-38 (2014).
53. Tschopp, O. *et al.* Essential role of protein kinase B gamma (PKB gamma/Akt3) in postnatal brain development but not in glucose homeostasis. *Development* **132**, 2943-54 (2005).
54. Pal, A. *et al.* PTEN mutations as a cause of constitutive insulin sensitivity and obesity. *N Engl J Med* **367**, 1002-11 (2012).
55. Savage, D.B. *et al.* Human metabolic syndrome resulting from dominant-negative mutations in the nuclear receptor peroxisome proliferator-activated receptor-gamma. *Diabetes* **52**, 910-7 (2003).
56. Semple, R.K. *et al.* Postreceptor insulin resistance contributes to human dyslipidemia and hepatic steatosis. *J Clin Invest* **119**, 315-22 (2009).
57. International Expert Committee report on the role of the A1C assay in the diagnosis of diabetes. *Diabetes care* **32**, 1327-34 (2009).
58. Manning, A.K. *et al.* A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat Genet* **44**, 659-69 (2012).
59. DePristo, M.A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* **43**, 491-8 (2011).
60. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904-9 (2006).
61. Kang, H.M. *et al.* Variance component model to account for sample structure in genome-wide association studies. *Nat Genet* **42**, 348-54 (2010).
62. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).
63. Morris, A.P. Transethnic meta-analysis of genomewide association studies. *Genet Epidemiol* **35**, 809-22 (2011).
64. Lee, S., Abecasis, G.R., Boehnke, M. & Lin, X. Rare-variant association analysis: study designs and statistical tests. *American journal of human genetics* **95**, 5-23 (2014).
65. McLaren, W. *et al.* Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* **26**, 2069-70 (2010).
66. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)* **6**, 80-92 (2012).
67. Feng, S., Liu, D., Zhan, X., Wing, M.K. & Abecasis, G.R. RAREMETAL: fast and powerful meta-analysis for rare variants. *Bioinformatics* **30**, 2828-9 (2014).
68. Liu, D.J. *et al.* Meta-analysis of gene-level tests for rare variant association. *Nat Genet* **46**, 200-4 (2014).
69. Wu, M.C. *et al.* Rare-variant association testing for sequencing data with the sequence kernel association test. *American journal of human genetics* **89**, 82-93 (2011).
70. Lee, S., Teslovich, T.M., Boehnke, M. & Lin, X. General framework for meta-analysis of rare variants in sequencing association studies. *Am J Hum Genet* **93**, 42-53 (2013).
71. Welter, D. *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res* **42**, D1001-6 (2014).

72. Chambers, J.C. *et al.* Common genetic variation near melatonin receptor MTNR1B contributes to raised plasma glucose and increased risk of type 2 diabetes among Indian Asians and European Caucasians. *Diabetes* **58**, 2703-8 (2009).
73. Prokopenko, I. *et al.* Variants in MTNR1B influence fasting glucose levels. *Nat Genet* **41**, 77-81 (2009).
74. Xing, C., Cohen, J.C. & Boerwinkle, E. A weighted false discovery rate control procedure reveals alleles at FOXA2 that influence fasting glucose levels. *Am J Hum Genet* **86**, 440-6 (2010).
75. Kim, Y.J. *et al.* Large-scale genome-wide association studies in East Asians identify new genetic loci influencing metabolic traits. *Nat Genet* **43**, 990-5 (2011).
76. Chen, G. *et al.* Genome-wide association study identifies novel loci association with fasting insulin and insulin resistance in African Americans. *Hum Mol Genet* **21**, 4530-6 (2012).
77. Comuzzie, A.G. *et al.* Novel genetic loci identified for the pathophysiology of childhood obesity in the Hispanic population. *PLoS One* **7**, e51954 (2012).
78. Kristiansson, K. *et al.* Genome-wide screen for metabolic syndrome susceptibility Loci reveals strong lipid gene contribution but no evidence for common genetic basis for clustering of metabolic syndrome traits. *Circ Cardiovasc Genet* **5**, 242-9 (2012).
79. Go, M.J. *et al.* New susceptibility loci in MYL2, C12orf51 and OAS1 associated with 1-h plasma glucose as predisposing risk factors for type 2 diabetes in the Korean population. *J Hum Genet* **58**, 362-5 (2013).
80. Wessel, J. *et al.* Low-frequency and rare exome chip variants associate with fasting glucose and type 2 diabetes susceptibility. *Nat Commun* **6**, 5897 (2015).
81. Genomes Project, C. *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061-73 (2010).
82. Raitakari, O.T. *et al.* Cohort profile: the cardiovascular risk in Young Finns Study. *Int J Epidemiol* **37**, 1220-6 (2008).
83. Eriksson, J.G. Epidemiology, genes and the environment: lessons learned from the Helsinki Birth Cohort Study. *J Intern Med* **261**, 418-25 (2007).
84. Perttila, J. *et al.* OSBPL10, a novel candidate gene for high triglyceride trait in dyslipidemic Finnish subjects, regulates cellular lipid metabolism. *J Mol Med (Berl)* **87**, 825-35 (2009).
85. Vartiainen, E. *et al.* Thirty-five-year trends in cardiovascular risk factors in Finland. *Int J Epidemiol* **39**, 504-18 (2010).
86. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nature genetics* **39**, 906-13 (2007).
87. Howie, B.N., Donnelly, P. & Marchini, J. A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies. *PLoS Genet* **5**, e1000529 (2009).
88. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* **39**, 906-13 (2007).
89. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65 (2012).
90. Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* **6**, pl1 (2013).
91. Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* **2**, 401-4 (2012).
92. UniProt, C. Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res* **42**, D191-8 (2014).
93. Karolchik, D. *et al.* The UCSC Genome Browser database: 2014 update. *Nucleic Acids Res* **42**, D764-70 (2014).
94. Valdar, W.S. Scoring residue conservation. *Proteins* **48**, 227-41 (2002).
95. Crooks, G.E., Hon, G., Chandonia, J.M. & Brenner, S.E. WebLogo: a sequence logo generator. *Genome Res* **14**, 1188-90 (2004).
96. Delaneau, O., Marchini, J. & Zagury, J.F. A linear complexity phasing method for thousands of genomes. *Nat Methods* **9**, 179-81 (2012).
97. Excoffier, L. & Lischer, H.E. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* **10**, 564-7 (2010).

98. Vilella, A.J. *et al.* EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res* **19**, 327-35 (2009).
99. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**, 1586-91 (2007).
100. McDonald, J.H. & Kreitman, M. Adaptive protein evolution at the Adh locus in Drosophila. *Nature* **351**, 652-4 (1991).
101. Roche, D.B., Buenavista, M.T., Tetchner, S.J. & McGuffin, L.J. The IntFOLD server: an integrated web resource for protein fold recognition, 3D model quality assessment, intrinsic disorder prediction, domain prediction and ligand binding site prediction. *Nucleic Acids Res* **39**, W171-6 (2011).
102. Buenavista, M.T., Roche, D.B. & McGuffin, L.J. Improvement of 3D protein models using multiple templates guided by single-template model quality assessment. *Bioinformatics* **28**, 1851-7 (2012).
103. Schrodinger, LLC. The PyMOL Molecular Graphics System, Version 1.3r1. (2010).
104. Trapnell, C., Pachter, L. & Salzberg, S.L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-11 (2009).
105. Harrow, J. *et al.* GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* **22**, 1760-74 (2012).
106. DeLuca, D.S. *et al.* RNA-SeQC: RNA-seq metrics for quality control and process optimization. *Bioinformatics* **28**, 1530-2 (2012).
107. Bates, D., Maechler, M. & Bolker, B. lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-41. (2011).
108. Consortium, G.T. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580-5 (2013).
109. Shabalin, A.A. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* **28**, 1353-8 (2012).
110. Stegle, O., Parts, L., Piipari, M., Winn, J. & Durbin, R. Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat Protoc* **7**, 500-7 (2012).
111. Brown, A.A. *et al. Genetic interactions affecting human gene expression identified by variance association mapping*, (2014).
112. The International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).
113. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-60 (2009).
114. Grundberg, E. *et al.* Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat Genet* **44**, 1084-1089 (2012).
115. http://www.r-project.org/. The R Project for Statistical Computing.
116. Falchi, M., Wilson, S.G., Paximadas, D., Swaminathan, R. & Spector, T.D. Quantitative Linkage Analysis for Pancreatic B-cell Function and Insulin Resistance in a Large Twin Cohort. *Diabetes* **57**, 1120-1124 (2008).
117. Stegle, O., Parts, L., Durbin, R. & Winn, J. A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies. *PLoS Comput Biol* **6**, e1000770 (2010).
118. Fall, T. *et al.* The role of adiposity in cardiometabolic traits: a Mendelian randomization analysis. *PLoS Med* **10**, e1001474 (2013).
119. Peloso, G.M. *et al.* Association of low-frequency and rare coding-sequence variants with blood lipids and coronary heart disease in 56,000 whites and blacks. *Am J Hum Genet* **94**, 223-32 (2014).
120. Tobin, M.D., Sheehan, N.A., Scurrah, K.J. & Burton, P.R. Adjusting for treatment effects in studies of quantitative traits: antihypertensive therapy and systolic blood pressure. *Statistics in medicine* **24**, 2911-35 (2005).

**Acknowledgements / Funding**

We thank the more than 44,412 volunteers who participated in this study. Academy of Finland (129293, 128315, 129330, 131593, 139635, 139635, 121584, 126925, 124282, and 129378, 258753); Action on Hearing Loss (G51); Ahokas Foundation; American Diabetes Association (#7-12-MN-02); Atlantic Canada Opportunities Agency; Augustinus foundation; Becket foundation; Benzon Foundation; Biomedical Research Council; British Heart Foundation (SP/04/002); Canada Foundation for Innovation; Commission of the European Communities, Directorate C-Public Health (2004310); Copenhagen County; Danish Centre for Evaluation and Health Technology Assessment; Danish Council for Independent Research; Danish Heart Foundation (07-10-R61-A1754-B838-22392F); Danish Medical Research Council; Danish Pharmaceutical Association; Emil Aaltonen Foundation; European Research Council Advanced Research Grant; European Union FP7 (EpiMigrant, 279143; FP7/2007-2013; 259749); Finland's Slottery Machine Association; Finnish Cultural Foundation; Finnish Diabetes Research Foundation; Finnish Foundation for Cardiovascular Research; Finnish Foundation of Cardiovascular Research; Finnish Medical Society; Finnish National Public Health Institute; Finska Läkaresällskapet; Folkhälsan Research Foundation; Foundation for Life and Health in Finland; German Center for Diabetes Research (DZD) ; German Federal Ministry of Education and Research; Health Care Centers in Vasa, Närpes and Korsholm; Health Insurance Foundation (2012B233) ; Helsinki University Central Hospital Research Foundation; Hospital districts of Pirkanmaa, Southern Ostrobothnia, North Ostrobothnia, Central Finland, and Northern Savo; Ib Henriksen foundation; Juho Vainio Foundation; Korea Centers for Disease Control and Prevention (4845–301); Korea National Institute of Health (2012-N73002-00); Li Ka Shing Foundation; Liv och Hälsa; Lundbeck Foundation; Marie-Curie Fellowship (PIEF-GA-2012-329156); Medical Research Council (G0601261, G0900747-91070, G0601966, G0700931); Ministry of Education in Finland; Ministry of Social Affairs and Health in Finland; MRC-PHE Centre for Environment and Health;Municipal Heath Care Center and Hospital in Jakobstad; Närpes Health Care Foundation; National Institute for Health Research (RP-PG-0407-10371); National Institutes of Health (U01 DK085526, U01 DK085501, U01 DK085524, U01 DK085545, U01 DK085584, U01 DK088389, RC2-DK088389, DK085545, DK098032, HHSN268201300046C, HHSN268201300047C, HHSN268201300048C, HHSN268201300049C, HHSN, R01MH107666 and K12CA139160268201300050C, R01 DK066358, U01DK085501, R01HL102830, R01DK073541, PO1AG027734, R01AG046949, 1R01AG042188, P30AG038072, R01 MH101820, R01MH090937, P30DK020595, R01 DK078616, NIDDK K24 DK080140, 1RC2DK088389, T32GM007753); National Medical Research Council; National Research Foundation of Korea (NRF-2012R1A2A1A03006155); Nordic Center of Excellence in Disease Genetics; Novo Nordisk; Ollqvist Foundation; Orion-Farmos Research Foundation; Paavo Nurmi Foundation; Perklén Foundation; Samfundet Folkhälsan; Signe and Ane Gyllenberg Foundation; Sigrid Juselius Foundation; Social Insurance Institution of Finland; South East Norway Health Authority (2011060) ; Swedish Cultural Foundation in Finland; Swedish Heart-Lung Foundation; Swedish Research Council; Swedish Research Council (Linné and Strategic Research Grant) ; The American Federation for Aging Research; The Einstein Glenn Center; The European Commission (HEALTH-F4-2007-201413) ; The Finnish Diabetes Association; The Folkhälsan Research Foundation; The Påhlssons Foundation; The provinces of Newfoundland and Labrador, Nova Scotia, and New Brunswick; The Sigrid Juselius Foundation; The Skåne Regional Health Authority; The Swedish Heart-Lung Foundation; Timber Merchant Vilhelm Bang's Foundation; Turku University Foundation; Uppsala University; Wellcome Trust (064890, 083948, 085475, 086596, 090367, 090532, 092447, 095101/Z/10/Z, 095552, 098017, 098381, 098051, 084723, 072960/2/ 03/2, 086113/Z/08/Z, WT098017, WT064890, WT090532, WT098017, 098051, WT086596/Z/08/A and 086596/Z/08/Z). Detailed acknowledgment of funding sources is provided in the Supplementary Material.

**Author Contributions**

**Sample Collection and Phenotyping:** NG, A Mahajan, NPB, C Ladenvall, JB-J, NRR, NWR, RAS, APG, AUJ, CJG, CB, D Buck, GB, GJ, HMS, JRH, J Murphy, JMJ, J Trakalo, KSS, MM, MN, M Hollensted, RO, SG, ARW, ATH, HEA, AC, RAD, A Stančáková, AHR, A Metspalu, AJF, A-CS, A Käräjämäki, YAK, RA, A Swift, TA, BL, BG, BIF, B-GH, C Meisinger, CG, C Langenberg, D Pasko, D Aguilar, D Bowden, DH, EST, EC, C-YC, WYL,

EM, SPF, FBH, G Atzmon, GWSr, DEH, HG, HK, HO, HATJr, TI, JSK, J Sehmi, J Lindstrom, J Kravic, JEC, CPJ, JEB, J Kriebel, JH, J Li, J Fadista, JCC, JCL, KRO, KSC, C-CK, LLB, J-YL, LK, DML, LH, L Milani, J Liu, L Liang, M Loh, MO-M, MW, MM-N, TM, MG, MR, MCYN, NDP, NN, LQ, NJW, NB, OM, OR, PJH, PWF, PN, A Peters, QQ, RM, S-TT, S Kumar, SKM, SPO'R, S Puppala, K Strauch, TMF, TK, TE, FT, BT, TVV, TYW, TAL, T Lauritzen, T Forsén, TIP, UA, VSF, WRS, YSC, ADM, ASFD, AL, BI, CNAP, FSC, CC, EI, FK, GLS, I Brandslund, J Tuomilehto, J Kuusisto, L Lannfelt, L Lind, LG, MEJ, MU, OP, RR, RNB, T Tuomi, TDS, TH, TJ, VS, GIB, JGW, JB, NJC, RD, KLM, M Laakso, CLH, APM, MB, D Altshuler, MIM

**Replication and Expression Studies:** T Tukiainen, AV, AAB, YW, A Palotie, AJ, JGE, OTR, S Koskinen, T Lehtimäki, JW, AYC, RAS, MOG, VS, JD, SR, JCF, JBM, ML, KLM

**Data Production (Sequencing and Genotyping):** XS, NG, A Mahajan, CF, NPB, C Hartl, C Ladenvall, JB-J, NRR, NWR, APG, AUJ, CJG, CB, D Buck, GB, GJ, HMS, JRH, J Murphy, JMJ, J Trakalo, KSS, MM, MN, M Hollensted, RO, PSC, SG, MOC, MD, EB, YF, MHdA, K Shakir, RP, T Fennell, TS, TW, TMS, K Stirrups, TM, PD, MB, MIM

**Variant Calling & Panel Generation:** MAR, KJG, HMK, GJ, BMN, GG, J Maguire, J Carey, JDS, JIG, S Purcell

**Statistical Analysis:** A Manning, HMH, JG, XS, T Tukiainen, P Fontanillas, NG, MAR, A Mahajan, AEL, P Cingolani, T Pers, J Flannick, CF, ERG, KJG, HKI, TMT, A Kumar, NPB, C Hartl, C Ladenvall, HMK, JB-J, YC, JRBP, LJS, C Ma, MvdB, L Moutsianas, NRR, RDP, TWB, TG, NWR, APG, AUJ, CJG, CB, D Buck, GB, GJ, HMS, JRH, J Murphy, JMJ, J Trakalo, KSS, MM, MN, M Hollensted, RO, SG, JBM, APM, MB, MIM, CML

**Functional Studies:** JG, SJ, ALG

**Wrote the paper:** A Manning, HMH, JG, XS, T Tukiainen, P Fontanillas, JCF, MB, MIM, ALG, CML

**Study Design:** XS, LJS, ATH, HEA, RAD, BG, EST, G Atzmon, JSK, CPJ, JCC, KSC, J-YL, DML, TM, TMF, TIP, YSC, C Hu, GRC, D Bharadwaj, PJD, D Prabhakaran, EZ, I Barroso, J Scott, J Chan, GM, MJD, M Sandhu, NT, PE, P Froguel, RCWM, RS, SBE, YYT, T Park, T Fingerlin, WJ, RMW, J Tuomilehto, LG, GIB, G Abecasis, JGW, JB, M Seielstad, NJC, RD, JD, IP, JCF, KLM, M Laakso, JBM, CLH, APM, MB, D Altshuler, MIM, ALG, CML

**Study Supervision:** GIB, G Abecasis, JGW, JB, M Seielstad, NJC, RD, JCF, KLM, JBM, CLH, APM, MB, D Altshuler, MIM, ALG, CML