



# Body mass index as a modifiable risk factor for type 2 diabetes: Refining and understanding causal estimates using Mendelian randomisation.

Journal:	Diabetes
Manuscript ID	DB16-0418.R1
Manuscript Type:	Brief Report
Date Submitted by the Author:	n/a
Complete List of Authors:	Corbin, Laura; University of Bristol, MRC IEU Richmond, Rebecca; University of Bristol, CRUK ICEP Wade, Kaitlin; University of Bristol, CRUK ICEP Burgess, Stephen; University of Cambridge, Department of Public Health and Primary Care Bowden, Jack; University of Bristol, MRC IEU Davey Smith, George ; University of Bristol, MRC IEU Timpson, Nicholas; University of Bristol, MRC IEU

SCHOLARONE<sup>™</sup> Manuscripts

1	Title: Body mass index as a modifiable risk factor for type 2 diabetes: Refining and understanding
2	causal estimates using Mendelian randomisation.
3	
4	Authors:
5	Laura J Corbin <sup>1</sup> ; Rebecca C Richmond <sup>1</sup> ; Kaitlin H Wade <sup>1</sup> ; Stephen Burgess <sup>1,3</sup> ; Jack Bowden <sup>1,2</sup> ;
6	George Davey Smith <sup>1</sup> ; Nicholas J Timpson <sup>1</sup>
7	
8	Affiliations:
9	1) MRC Integrative Epidemiology Unit (IEU) at University of Bristol, Bristol, UK
10	2) MRC Biostatistics Unit, Cambridge Institute of Public Health, Cambridge, UK
11	3) Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
12	
13	Running title: Body mass index and type 2 diabetes; Mendelian randomisation methods
14	
15	
16	
17	
18	
19	Corresponding author: Nicholas J Timpson, Email: N.J.Timpson@bristol.ac.uk, Address: MRC
20	Integrative Epidemiology Unit, University of Bristol, Oakfield House, Oakfield Grove, Bristol, BS8
21	2BN, Tel: 0117 3310131.

#### 22 ABSTRACT

23 This study focused on resolving the relationship between body mass index (BMI) and type 2 24 diabetes. The availability of multiple variants associated with BMI offers a new chance to resolve 25 the true causal effect of BMI on T2D, however the properties of these associations and their 26 validity as genetic instruments need to be considered alongside established and new methods for 27 undertaking Mendelian randomisation. We explore the potential for pleiotropic genetic variants to 28 generate bias, revise existing estimates and illustrate value in new analysis methods. A two-29 sample Mendelian randomisation (MR) approach with 96 genetic variants was employed using 30 three different analysis methods, two of which (MR-Egger and the weighted median) have been 31 developed specifically to address problems of invalid instrumental variables. We estimate an odds 32 ratio for type 2 diabetes per unit increase in BMI (kg/m<sup>2</sup>) of between 1.19 and 1.38, with the most 33 stable estimate using all instruments and a weighted median approach (1.26 95%CI (1.17, 1.34)). 34 TCF7L2(rs7903146) was identified as a complex effect or pleiotropic instrument and removal of 35 this variant resulted in convergence of causal effect estimates from different causal analysis 36 methods. This indicated the potential for pleiotropy to affect estimates and differences in 37 performance of alternative analytical methods. In a real type 2 diabetes focused example, this 38 study demonstrates the potential impact of invalid instruments on causal effect estimates and the 39 potential for new approaches to mitigate the bias caused.

40 Observational studies have shown body mass index (BMI) to be associated with risk of type 2 41 diabetes as well as with a range of diabetes-related metabolic traits (1; 2). However, it is well 42 known that confounding, reverse causation and biases can generate such associations and that 43 even with careful study design, incorrect inference is possible (3). One approach to circumventing 44 these problems is to use genetic association results within a Mendelian randomization (MR) 45 framework (3; 4). In MR analyses, genetic variants act as proxies for an exposure in a manner 46 independent of confounders. If in addition the variants only affect an outcome of interest through 47 the chosen exposure, then they are said to be valid instrumental variables (IVs). This enables 48 evaluation of the causal effect of the exposure on the outcome, escaping some of the limitations of 49 observational epidemiology; (5).

50

51 Following the success of genome-wide association studies (GWASs), the number of MR analyses 52 using large numbers of mostly uncharacterized variants associated with complex health outcomes 53 or intermediates is rapidly increasing (6; 7). In the case of BMI, there are now 97 genetic variants 54 reliably associated and there are examples where multiple variants have been used as a 55 composite IV to estimate the causal impact of BMI on health (8). Although using many IVs can 56 increase the power of MR analyses, it brings with it the concern that enlarged sets of genetic 57 variants are more likely to contain invalid IVs due to violations of the assumptions necessary for 58 valid causal inference using traditional methods (9). In particular, horizontal pleiotropy – where a 59 genetic variant affects the outcome via more than one biological pathway (10) – is a concern. 60 Importantly, the properties of these associations and their validity as genetic instruments need to 61 be considered alongside established and new methods for undertaking Mendelian randomisation. 62

In response to the general issue of using multiple genetic variants in MR, Bowden *et al.* (9)
propose both MR-Egger regression, an approach developed from the original Egger regression
technique for assessing small study bias in meta-analysis and a weighted weighted median
approach (11) as alternatives to the standard MR analysis. The MR-Egger and weighted weighted
median approaches both operate using distinct, but critically weaker, versions of the IV
assumptions, and therefore have the potential to deliver robust causal effect estimates. The MR-

69 Egger method also provides a formal statistical test as to whether or not the average pleiotropic

70 effect of the genetic variants is equal to zero (9).

71

#### 72 **Research Design and Methods**

73 With increasing evidence for multiple biological pathways underlying type 2 diabetes (12; 13) and 74 increasing numbers of genetic variants available as IVs for BMI, we set out to test the potential for 75 bias in causal estimates from MR using these state-of-the-art approaches. We compared results 76 from MR-Egger regression (9) and weighted weighted median (11) approaches to a traditional 77 inverse-variance weighted (IVW) method (which makes the strong assumption that all variants are 78 valid IVs) (14) in an investigation of the causal relationship between BMI and type 2 diabetes. 79 These methods all undertake two-sample Mendelian randomisation whereby the GWAS results for 80 a disease outcome are unified with those of an exposure of interest and together used to estimate 81 the causal impact of that exposure on disease. We used published data in a two-sample analysis 82 strategy taking SNP-exposure and SNP-outcome associations from different sources (15; 16).

83

84 The effect sizes for BMI-associated SNPs with associated standard errors from a mixed-sex cohort 85 of European ancestry were taken from the Genetic Investigation of ANthropometric Traits (GIANT) 86 consortium (17) along with results for type 2 diabetes from the DIAbetes Genetics Replication And 87 Meta-analysis (DIAGRAM) Consortium. To avoid sample overlap, GIANT estimates were re-88 calculated in the absence of DIAGRAM cohorts yielding a maximum sample size at any given 89 locus of 189,079. To aid interpretation of the effects of BMI on type 2 diabetes, effect sizes were transformed to BMI units prior to analysis, assuming one standard deviation (SD) = 4.5kg/m<sup>2(17)</sup>. 90 91 For the corresponding SNP-outcome association, we took odds ratios (ORs) and confidence 92 intervals from a GWAS meta-analysis conducted by the DIAGRAM Consortium. This genome-wide 93 meta-analysis includes data from 12,171 type 2 diabetes cases and 56,862 controls of mainly 94 European descent imputed at up to 2.5 million autosomal SNPs (DIAGRAMv3) (18). All but one 95 (rs4787491, *INO80E*) of the BMI-associated SNPs ( $p < 5x10^8$ ) from GIANT had results listed in the 96 DIAGRAMv3 dataset so 96 SNPs with results in both datasets were taken forward for analysis.

97

98	SNP-exposure and SNP-outcome associations were combined using the three different
99	approaches outlined above. All analyses were conducted in R 3.2.0 (19). First, an inverse-variance
100	weighted (IVW) method was implemented to provide a weighted average of the causal effect
101	estimates (14). This method assumes that all genetic variants (i.e. 100%) satisfy the IV
102	assumptions (including zero pleiotropy) and uses weights that assume the gene-exposure
103	association estimates are measured without error (the No Measurement Error (NOME)
104	assumption).
105	
106	Second, we performed MR-Egger regression (9), which assumes NOME but allows each variant to
107	exhibit pleiotropy. MR-Egger estimates remain consistent only if the magnitude of the gene
108	exposure associations across all variants are independent of their pleiotropic effects (the InSIDE
109	assumption) (9). As recommended by Bowden et al (9), the extent to which pleiotropy was
110	balanced across the set of instruments as a whole was visually assessed by plotting the causal

111 effect estimates against their precision, using a funnel plot and checking for asymmetry (Figure

112 1A). The NOME assumption was assessed for MR-Egger via an adaptation of the  $l^2$  statistic ( $l_{GX}^2$ )

113 (20) and adjusted for by combining MR-Egger with the method of Simulation Extrapolation

114 (SIMEX) (21). Using SIMEX, new data sets are created by simulating gene-exposure association

estimates under increasing violations of NOME and recording the amount of attenuation in the

116 estimate that occurs. The set of attenuated estimates are then used to extrapolate back to the

117 estimate that would have been obtained if NOME had been satisfied.

118

Finally, a weighted weighted median estimation method was applied (11). The weighted median provides a consistent estimate of causal effect if at least 50% of the information in the analysis comes from variants that are valid IVs. For a more detailed description of the three methods applied, see Online Appendix (Supplementary methods). A leave-one-out permutation analysis was conducted across all methods to assess the influence of potentially pleiotropic SNPs on the causal estimates (22). In the case of the linear models (IVW and MR-Egger) two additional analyses were conducted (23; 24). Firstly, the extent to which the causal estimate from each SNP

in the set could be considered an outlier was assessed using studentized residuals. Secondly,

127 Cook's distance (25) was used as a measure of the aggregate impact of each SNP on the model.

128

#### 129 Results

130 All three approaches provide evidence of a positive causal relationship between BMI and type 2 131 diabetes. This is demonstrated in Figure 1B where the slope of the lines show the causal effect 132 estimates as predicted by the IVW, MR-Egger and m weighted median approaches. Estimates 133 correspond to an OR for type 2 diabetes per unit increase in BMI (kg/m<sup>2</sup>) of 1.19, 1.26 and 1.38 for 134 the IVW, weighted median and MR-Egger analyses, respectively and are in line with a previous 135 MR estimate of 1.27 (95%CI 1.18, 1.36) (2) (Table 1). Assessment of the NOME assumption with 136 respect to the MR-Egger estimate gave  $I_{GX}^2$ =0.83, suggesting an approximate 15% attenuation of 137 the causal estimate towards zero. Bias adjustment via SIMEX gave a corrected MR-Egger causal 138 estimate of 1.46 (95%CI 1.16, 1.84) for type 2 diabetes per unit increase in BMI (kg/m<sup>2</sup>).

139

Considering the individual SNP-based contributions to MR analysis, there is one clear outlier in the distribution of effects shown in Figure 1 and that is *TCF7L2*(rs7903146). *TCF7L2*(rs7903146) shows an association with BMI that is in the opposite direction to the overall trend (and weak relative to its effect on type 2 diabetes), resulting in a large negative causal estimate from this SNP alone. The presence of at least some unbalanced pleiotropy was detected within the set of variants, as reflected by the intercept estimate of -0.019 (*p*=0.10) in the MR-Egger analysis.

146

147 To illustrate the impact of TCF7L2(rs7903146) on causal estimates, we performed a sensitivity 148 analysis in which each SNP in turn was removed from the set in a leave-one-out permutation 149 analysis. We saw a shift in the causal estimates from the IVW (an increase) and MR-Egger (a 150 decrease) as a result of the removal of TCF7L2(rs7903146) but no difference in the estimate from 151 the weighted median approach (Table 1; Figure 2). The results of the leave-one-out permutation 152 analysis showed that the impact of removing TCF7L2(rs7903146) from the variant set on the IVW 153 and MR-Egger estimates was greater than that of removing almost any other variant, with the 154 exception of FTO(rs1558902) (Figure 2A & B). When FTO(rs1558902) was removed, causal

155 estimates from both the IVW and MR-Egger analysis decreased (Table 1; Figure 2). In this 156 instance we also observed movement in the causal effect estimate from the weighted median 157 (Table 1; Figure 2C). The estimate of the intercept from MR-Egger moved closer to zero following 158 both the removal of TCF7L2(rs7903146) and FTO(rs1558902) (Figure 2D). TCF7L2(rs7903146) 159 was also identified as an outlier in both IVW and MR-Egger (studentized residuals, Bonferroni 160 corrected  $p < 1 \times 10^{-19}$ ) but FTO(rs1558902) was not (Online Appendix (Supplementary Results, 161 Figures S1A/B)). Calculation of Cook's distance showed both variants to have a disproportionate 162 level of influence on the model compared to other variants in the set (Online Appendix 163 (Supplementary Results, Figures S2A/B)). 164 165 These results suggest *TCF7L2*(rs7903146) may be pleiotropic with respect to the outcome, i.e. 166 that it influences type 2 diabetes through an alternative pathway (other than BMI). Evidence from 167 existing literature supports this assertion as the type 2 diabetes risk increasing allele at 168 TCF7L2(rs7903146) has been associated with both increased fasting glucose (26) and decreased 169 BMI (17). Under the assumption that TCF7L2(rs7903146) demonstrates horizontal pleiotropy with 170 respect to type 2 diabetes, we would expect its inclusion in the variant set to bias the causal 171 estimate predicted by the IVW approach, but not that predicted by MR-Egger or the weighted 172 median. Removing TCF7L2(rs7903146) from the variant set causes a slight shift in the causal 173 estimates from the IVW and MR-Egger approaches, bringing them more in line with one another 174 and also with the weighted median estimate which remained stable in this instance. Also of note is 175 the reduction in the 95% confidence interval of the MR-Egger estimate following removal of the

the set is another potentially favourable quality of this estimator. The relatively small changes

TCF7L2(rs7903146). This increase in precision following removal of a likely invalid instrument from

178 observed across all methods as a result of removing *TCF7L2*(rs7903146) are in line with the

179 relatively weak effect of the SNP as shown in Figure 1B.

180

176

In contrast, the effect of removing *FTO*(rs1558902) is more noticeable. Regardless of the method
 used, removing this variant results in a lower causal estimate (Table 1; Figure 2). The substantial
 influence of *FTO*(rs1558902) was predicable given the strength of its effect relative to the other

184 variants (Figure 1B), though properties of this effect are not in line with other variants used to 185 instrument BMI as reported elsewhere for physical activity (27), thyroid function (28) and 186 depression (29). The concomitant increase in standard error associated with the estimates here 187 point towards increased uncertainty moving the estimates towards the null in the absence of 188 FTO(rs1558902). The weighted median appears robust, even to the removal of FTO(rs1558902), 189 as demonstrated by the relatively tight distribution of estimates returned from the leave-one-out 190 permutation analysis (Figure 2C). This is as expected given the tolerance of weighted median 191 approaches to outliers.

192

#### 193 **Discussion**

194 By applying new analytical techniques to an old question – the causal relationship between BMI 195 and type 2 diabetes – we have explored the potential for invalid instruments to bias causal 196 estimates in MR. In this case where BMI is the exposure, the opportunity to use a large instrument 197 list in causal analyses presents both opportunity, through variance explained, but also cost, 198 through complications generated by instrument properties or methods employed. Results here 199 suggest that both TCF7L2 and FTO appear to have genetic variation which predicts BMI reliably, 200 but for which associations with type 2 diabetes do not fully align with that for other variants (given 201 BMI effects and assumed causality).

202

203 For TCF7L2, only recently suggested to be associated with BMI directly (17), this is not surprising 204 and reinforces the important point that the validity of a specific method's MR estimate depends on 205 whether the genetic variants collectively satisfy its assumptions. In this case, it is possible that the 206 negative association with BMI observed in GIANT is the product of a form of bias where the risk of 207 type 2 diabetes is leading to effective treatment, health benefit and BMI reduction. This is 208 supported by the apparently causal negative relationship between type 2 diabetes and BMI seen in 209 a reciprocal analysis where BMI is the outcome of interest (Online Appendix (Supplementary 210 Results, Figure S3)), though is likely to be more a comment on study design than biological effect.

211

212 In this example, the use of recently derived methods (9; 11) designed to overcome problems 213 caused by directional pleiotropy, yields estimates which are more stable in the presence or 214 absence of potentially invalid instruments and confirm the likely magnitude of the average effect of 215 BMI on type 2 diabetes (i.e. from the most likely and stable estimate, an elevation of odds of 216 disease of ~26% for each additional unit of BMI). The comparison of results from different methods 217 for any set of potential instruments is important when assessing the reliability of causal inferences 218 and important for downstream interpretation. In this case, whilst it is impossible to model precisely, 219 one can estimate the hypothetical impact of an average population level change in lifecourse BMI 220 on type 2 diabetes. Given a population size of 64.1 million in the UK in mid 2013(30) and a 221 modelled prevalence of type 2 diabetes (including non-diagnosed cases) of 7.4%(31; 32), the 222 estimated reduction in odds for a 1kg/m2 reduction would potentially yield a reduction in the 223 number of cases from  $\sim$ 4.7-3.6 million (a shift in prevalence to 5.6%). 224 225 Acknowledgements 226 Author contributions: NJT conceived and supervised the study, in discussion with GDS, LJC, 227 KHW and RCR. JB and SB contributed to method and script development. LJC conducted the 228 analysis and wrote the manuscript. JB prepared the (Online Appendix (Supplementary methods)). 229 NJT is the guarantor of this work and, as such, had full access to all the data in the study and 230 takes responsibility for the integrity of the data and the accuracy of the data analysis. 231 232 Funding: NJT, LC, RCR, KHW, SB, GDS and JB work in the Medical Research Council 233 Integrative Epidemiology Unit (IEU) at the University of Bristol which is supported by the Medical

Research Council (MC\_UU\_12013/1, MC\_UU\_12013/2, MC\_UU\_12013/3) and the University of

Bristol. SB is supported by the Wellcome Trust (grant number 100114). JB is funded by a Medical

236 Research Council Methodology Research Fellowship (grant number MR/N501906/1). RCR and

237 KHW are supported by CRUK (C18281/A19169).

238

- 239 **GIANT:** Data on BMI have been contributed by GIANT (Genetic Investigation of ANthropometric
- 240 Traits). With special thanks to Adam Locke for conducting the additional analyses required for this
- 241 work.
- 242
- 243 **DIAGRAM:** Data on type 2 diabetes have been contributed by the DIAGRAM (**DIA**betes **G**enetics
- 244 Replication And Meta-analysis) consortium and were downloaded from <u>here</u>.
- 245
- There are **no** conflicts to declare.

247 248	References
249 250 251	1. Prospective Studies C: Body-mass index and cause-specific mortality in 900 000 adults: collaborative analyses of 57 prospective studies. The Lancet 373:1083-1096
252 253 254 255 256 257 258 259 260	2. Holmes Michael V, Lange Leslie A, Palmer T, Lanktree Matthew B, North Kari E, Almoguera B, Buxbaum S, Chandrupatla Hareesh R, Elbers Clara C, Guo Y, Hoogeveen Ron C, Li J, Li Yun R, Swerdlow Daniel I, Cushman M, Price Tom S, Curtis Sean P, Fornage M, Hakonarson H, Patel Sanjay R, Redline S, Siscovick David S, Tsai Michael Y, Wilson James G, van der Schouw Yvonne T, FitzGerald Garret A, Hingorani AD, Casas Juan P, de Bakker Paul IW, Rich Stephen S, Schadt Eric E, Asselbergs Folkert W, Reiner Alex P, Keating Brendan J: Causal Effects of Body Mass Index on Cardiometabolic Traits and Events: A Mendelian Randomization Analysis. The American Journal of Human Genetics 2014;94:198-208
261 262 263	3. Davey Smith G, Hemani G: Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Human Molecular Genetics 2014;23:R89-R98
264 265 266 267	4. Davey Smith G, Ebrahim S: 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? International Journal of Epidemiology 2003;32:1-22
268 269 270 271	5. Lawlor DA, Harbord RM, Sterne JAC, Timpson N, Davey Smith G: Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Stat Med 2008;27:1133-1163
272 273 274	6. Sleiman PMA, Grant SFA: Mendelian Randomization in the Era of Genomewide Association Studies. Clinical Chemistry 2010;56:723-728
275 276 277 278 279 280	<ol> <li>Boef AGC, Dekkers OM, le Cessie S: Mendelian randomization studies: a review of the approaches used and the quality of reporting. International Journal of Epidemiology 2015;</li> <li>Nordestgaard BG, Palmer TM, Benn M, Zacho J, Tybjærg-Hansen A, Davey Smith G, Timpson NJ: The Effect of Elevated Body Mass Index on Ischemic Heart Disease Risk: Causal Estimates from a Mendelian Randomisation Approach. PLoS Med 2012;9:e1001212</li> </ol>
280 281 282 283 284	9. Bowden J, Davey Smith G, Burgess S: Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. International Journal of Epidemiology 2015;44:512-525
284 285 286 287	10. Didelez V, Sheehan N: Mendelian randomization as an instrumental variable approach to causal inference. Statistical Methods in Medical Research 2007;16:309-330
288 289 290 291	11. Bowden J, Davey Smith G, Haycock PC, Burgess S: Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. Genetic Epidemiology 2016;40:304-314
292 293 294	12. Yaghootkar H, Frayling T: Recent progress in the use of genetics to understand links between type 2 diabetes and related metabolic traits. Genome Biol 2013;14:1-7
295 296 297 298 299 300 301 302	13. Dimas AS, Lagou V, Barker A, Knowles JW, Mägi R, Hivert M-F, Benazzo A, Rybin D, Jackson AU, Stringham HM, Song C, Fischer-Rosinsky A, Boesgaard TW, Grarup N, Abbasi FA, Assimes TL, Hao K, Yang X, Lecoeur C, Barroso I, Bonnycastle LL, Böttcher Y, Bumpstead S, Chines PS, Erdos MR, Graessler J, Kovacs P, Morken MA, Narisu N, Payne F, Stancakova A, Swift AJ, Tönjes A, Bornstein SR, Cauchi S, Froguel P, Meyre D, Schwarz PEH, Häring H-U, Smith U, Boehnke M, Bergman RN, Collins FS, Mohlke KL, Tuomilehto J, Quertemous T, Lind L, Hansen T, Pedersen O, Walker M, Pfeiffer AFH, Spranger J, Stumvoll M, Meigs JB, Wareham NJ, Kuusisto J, Laakso M, Langenberg C, Dupuis J, Watanabe RM, Florez JC, Ingelsson E, McCarthy MI, Prokopenko I,

Investigators obotM: Impact of Type 2 Diabetes Susceptibility Variants on Quantitative Glycemic
 Traits Reveals Mechanistic Heterogeneity. Diabetes 2014;63:2158-2171

305

306 14. Burgess S, Butterworth A, Thompson SG: Mendelian Randomization Analysis With Multiple
 307 Genetic Variants Using Summarized Data. Genetic Epidemiology 2013;37:658-665
 308

- 309 15. Burgess S, Scott R, Timpson N, Davey Smith G, Thompson S: Using published data in
- 310 Mendelian randomization: a blueprint for efficient identification of causal risk factors. Eur J
- 311 Epidemiol 2015:1-10
- 312

16. Pierce BL, Burgess S: Efficient Design for Mendelian Randomization Studies: Subsample and
 2-Sample Instrumental Variable Estimators. American Journal of Epidemiology 2013;178:1177 1184

315 316

317 17. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, Powell C, Vedantam S, 318 Buchkovich ML, Yang J, Croteau-Chonka DC, Esko T, Fall T, Ferreira T, Gustafsson S, Kutalik Z, 319 Luan Ja, Magi R, Randall JC, Winkler TW, Wood AR, Workalemahu T, Faul JD, Smith JA, Hua 320 Zhao J, Zhao W, Chen J, Fehrmann R, Hedman AK, Karjalainen J, Schmidt EM, Absher D, Amin 321 N, Anderson D, Beekman M, Bolton JL, Bragg-Gresham JL, Buyske S, Demirkan A, Deng G, Ehret 322 GB, Feenstra B, Feitosa MF, Fischer K, Goel A, Gong J, Jackson AU, Kanoni S, Kleber ME, 323 Kristiansson K, Lim U, Lotay V, Mangino M, Mateo Leach I, Medina-Gomez C, Medland SE, Nalls 324 MA, Palmer CD, Pasko D, Pechlivanis S, Peters MJ, Prokopenko I, Shungin D, Stancakova A, 325 Strawbridge RJ, Ju Sung Y, Tanaka T, Teumer A, Trompet S, van der Laan SW, van Setten J, Van 326 Vliet-Ostaptchouk JV, Wang Z, Yengo L, Zhang W, Isaacs A, Albrecht E, Arnlov J, Arscott GM, 327 Attwood AP, Bandinelli S, Barrett A, Bas IN, Bellis C, Bennett AJ, Berne C, Blagieva R, Bluher M, 328 Bohringer S, Bonnycastle LL, Bottcher Y, Boyd HA, Bruinenberg M, Caspersen IH, Ida Chen Y-D, 329 Clarke R, Warwick Daw E, de Craen AJM, Delgado G, Dimitriou M, Doney ASF, Eklund N, Estrada 330 K, Eury E, Folkersen L, Fraser RM, Garcia ME, Geller F, Giedraitis V, Gigante B, Go AS, Golay A, 331 Goodall AH, Gordon SD, Gorski M, Grabe H-J, Grallert H, Grammer TB, Graszler J, Gronberg H, 332 Groves CJ, Gusto G, Haessler J, Hall P, Haller T, Hallmans G, Hartman CA, Hassinen M, Hayward 333 C, Heard-Costa NL, Helmer Q, Hengstenberg C, Holmen O, Hottenga J-J, James AL, Jeff JM, 334 Johansson A, Jolley J, Juliusdottir T, Kinnunen L, Koenig W, Koskenvuo M, Kratzer W, Laitinen J, 335 Lamina C, Leander K, Lee NR, Lichtner P, Lind L, Lindstrom J, Sin Lo K, Lobbens S, Lorbeer R, 336 Lu Y, Mach F, Magnusson PKE, Mahajan A, McArdle WL, McLachlan S, Menni C, Merger S, 337 Mihailov E, Milani L, Moayyeri A, Monda KL, Morken MA, Mulas A, Muller G, Muller-Nurasyid M, 338 Musk AW, Nagaraja R, Nothen MM, Nolte IM, Pilz S, Rayner NW, Renstrom F, Rettig R, Ried JS, 339 Ripke S, Robertson NR, Rose LM, Sanna S, Scharnagl H, Scholtens S, Schumacher FR, Scott 340 WR, Seufferlein T, Shi J, Vernon Smith A, Smolonska J, Stanton AV, Steinthorsdottir V, Stirrups K, 341 Stringham HM, Sundstrom J, Swertz MA, Swift AJ, Syvanen A-C, Tan S-T, Tayo BO, Thorand B, 342 Thorleifsson G, Tyrer JP, Uh H-W, Vandenput L, Verhulst FC, Vermeulen SH, Verweij N, Vonk JM, 343 Waite LL, Warren HR, Waterworth D, Weedon MN, Wilkens LR, Willenborg C, Wilsgaard T, 344 Wojczynski MK, Wong A, Wright AF, Zhang Q, The LifeLines Cohort S, Brennan EP, Choi M, 345 Dastani Z, Drong AW, Eriksson P, Franco-Cereceda A, Gadin JR, Gharavi AG, Goddard ME, 346 Handsaker RE, Huang J, Karpe F, Kathiresan S, Keildson S, Kiryluk K, Kubo M, Lee J-Y, Liang L, 347 Lifton RP, Ma B, McCarroll SA, McKnight AJ, Min JL, Moffatt MF, Montgomery GW, Murabito JM, 348 Nicholson G, Nyholt DR, Okada Y, Perry JRB, Dorajoo R, Reinmaa E, Salem RM, Sandholm N, 349 Scott RA, Stolk L, Takahashi A, Tanaka T, van/t Hooft FM, Vinkhuyzen AAE, Westra H-J, Zheng 350 W, Zondervan KT, The AC, The A-BMIWG, The CDC, The CC, The G, The I, The MI, The Mu TC, 351 The MC, The PC, The ReproGen C, The GC, The International Endogene C, Heath AC, Arveiler D, 352 Bakker SJL, Beilby J, Bergman RN, Blangero J, Bovet P, Campbell H, Caulfield MJ, Cesana G, 353 Chakravarti A, Chasman DI, Chines PS, Collins FS, Crawford DC, Adrienne Cupples L, Cusi D, 354 Danesh J, de Faire U, den Ruijter HM, Dominiczak AF, Erbel R, Erdmann J, Eriksson JG, Farrall 355 M, Felix SB, Ferrannini E, Ferrieres J, Ford I, Forouhi NG, Forrester T, Franco OH, Gansevoort 356 RT, Gejman PV, Gieger C, Gottesman O, Gudnason V, Gyllensten U, Hall AS, Harris TB, 357 Hattersley AT, Hicks AA, Hindorff LA, Hingorani AD, Hofman A, Homuth G, Kees Hovingh G, 358 Humphries SE, Hunt SC, Hypponen E, Illig T, Jacobs KB, Jarvelin M-R, Jockel K-H, Johansen B, 359 Jousilahti P, Wouter Jukema J, Jula AM, Kaprio J, Kastelein JJP, Keinanen-Kiukaanniemi SM,

Kiemeney LA, Knekt P, Kooner JS, Kooperberg C, Kovacs P, Kraja AT, Kumari M, Kuusisto J, 360 361 Lakka TA, Langenberg C, Le Marchand L, Lehtimaki T, Lyssenko V, Mannisto S, Marette A, Matise 362 TC, McKenzie CA, McKnight B, Moll FL, Morris AD, Morris AP, Murray JC, Nelis M, Ohlsson C, 363 Oldehinkel AJ, Ong KK, Madden PAF, Pasterkamp G, Peden JF, Peters A, Postma DS, 364 Pramstaller PP, Price JF, Qi L, Raitakari OT, Rankinen T, Rao DC, Rice TK, Ridker PM, Rioux JD, 365 Ritchie MD, Rudan I, Salomaa V, Samani NJ, Saramies J, Sarzynski MA, Schunkert H, Schwarz 366 PEH, Sever P, Shuldiner AR, Sinisalo J, Stolk RP, Strauch K, Tonjes A, Tregouet D-A, Tremblay 367 A, Tremoli E, Virtamo J, Vohl M-C, Volker U, Waeber G, Willemsen G, Witteman JC, Carola 368 Zillikens M, Adair LS, Amouyel P, Asselbergs FW, Assimes TL, Bochud M, Boehm BO, Boerwinkle 369 E, Bornstein SR, Bottinger EP, Bouchard C, Cauchi S, Chambers JC, Chanock SJ, Cooper RS, de 370 Bakker PIW, Dedoussis G, Ferrucci L, Franks PW, Froquel P, Groop LC, Haiman CA, Hamsten A, 371 Hui J, Hunter DJ, Hveem K, Kaplan RC, Kivimaki M, Kuh D, Laakso M, Liu Y, Martin NG, Marz W, 372 Melbye M, Metspalu A, Moebus S, Munroe PB, Njolstad I, Oostra BA, Palmer CNA, Pedersen NL, 373 Perola M, Perusse L, Peters U, Power C, Quertermous T, Rauramaa R, Rivadeneira F, Saaristo 374 TE. Saleheen D. Sattar N. Schadt EE, Schlessinger D, Eline Slagboom P, Snieder H, Spector TD, 375 Thorsteinsdottir U. Stumvoll M. Tuomilehto J. Uitterlinden AG. Uusitupa M. van der Harst P. 376 Walker M, Wallaschofski H, Wareham NJ, Watkins H, Weir DR, Wichmann HE, Wilson JF, Zanen 377 P, Borecki IB, Deloukas P, Fox CS, Heid IM, O/Connell JR, Strachan DP, Stefansson K, van Duijn 378 CM, Abecasis GR, Franke L, Frayling TM, McCarthy MI, Visscher PM, Scherag A, Willer CJ, 379 Boehnke M, Mohlke KL, Lindgren CM, Beckmann JS, Barroso I, North KE, Ingelsson E, Hirschhorn 380 JN, Loos RJF, Speliotes EK: Genetic studies of body mass index yield new insights for obesity 381 biology. Nature 2015;518:197-206 382 383 18. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segrè AV, Steinthorsdottir V, Strawbridge RJ, 384 Khan H, Grallert H, Mahajan A, Prokopenko I, Kang HM, Dina C, Esko T, Fraser RM, Kanoni S, 385 Kumar A, Lagou V, Langenberg C, Luan Ja, Lindgren CM, Müller-Nurasyid M, Pechlivanis S, 386 Rayner NW, Scott LJ, Wiltshire S, Yengo L, Kinnunen L, Rossin EJ, Raychaudhuri S, Johnson AD,

387 Dimas AS, Loos RJF, Vedantam S, Chen H, Florez JC, Fox C, Liu C-T, Rybin D, Couper DJ, Kao 388 WHL, Li M, Cornelis MC, Kraft P, Sun Q, van Dam RM, Stringham HM, Chines PS, Fischer K, 389 Fontanillas P, Holmen OL, Hunt SE, Jackson AU, Kong A, Lawrence R, Meyer J, Perry JRB, 390 Platou CGP, Potter S, Rehnberg E, Robertson N, Sivapalaratnam S, Stančáková A, Stirrups K, 391 Thorleifsson G, Tikkanen E, Wood AR, Almgren P, Atalay M, Benediktsson R, Bonnycastle LL, 392 Burtt N, Carey J, Charpentier G, Crenshaw AT, Doney ASF, Dorkhan M, Edkins S, Emilsson V, 393 Eury E, Forsen T, Gertow K, Gigante B, Grant GB, Groves CJ, Guiducci C, Herder C, Hreidarsson 394 AB, Hui J, James A, Jonsson A, Rathmann W, Klopp N, Kravic J, Krjutškov K, Langford C, 395 Leander K, Lindholm E, Lobbens S, Männistö S, Mirza G, Mühleisen TW, Musk B, Parkin M, 396 Rallidis L, Saramies J, Sennblad B, Shah S, Sigurðsson G, Silveira A, Steinbach G, Thorand B, 397 Trakalo J, Veglia F, Wennauer R, Winckler W, Zabaneh D, Campbell H, van Duijn C, Uitterlinden 398 AG, Hofman A, Sijbrands E, Abecasis GR, Owen KR, Zeggini E, Trip MD, Forouhi NG, Syvänen A-399 C, Eriksson JG, Peltonen L, Nöthen MM, Balkau B, Palmer CNA, Lyssenko V, Tuomi T, Isomaa B, 400 Hunter DJ, Qi L, Wellcome Trust Case Control C, Investigators M, Consortium G, Consortium A-401 TD, Consortium SD, Shuldiner AR, Roden M, Barroso I, Wilsgaard T, Beilby J, Hovingh K, Price 402 JF, Wilson JF, Rauramaa R, Lakka TA, Lind L, Dedoussis G, Njølstad I, Pedersen NL, Khaw K-T, 403 Wareham NJ, Keinanen-Kiukaanniemi SM, Saaristo TE, Korpi-Hyövälti E, Saltevo J, Laakso M, 404 Kuusisto J, Metspalu A, Collins FS, Mohlke KL, Bergman RN, Tuomilehto J, Boehm BO, Gieger C, 405 Hveem K, Cauchi S, Froguel P, Baldassarre D, Tremoli E, Humphries SE, Saleheen D, Danesh J, 406 Ingelsson E, Ripatti S, Salomaa V, Erbel R, Jöckel K-H, Moebus S, Peters A, Illig T, de Faire U, 407 Hamsten A, Morris AD, Donnelly PJ, Frayling TM, Hattersley AT, Boerwinkle E, Melander O, 408 Kathiresan S, Nilsson PM, Deloukas P, Thorsteinsdottir U, Groop LC, Stefansson K, Hu F, Pankow 409 JS, Dupuis J, Meigs JB, Altshuler D, Boehnke M, McCarthy MI: Large-scale association analysis 410 provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nature 411 genetics 2012;44:981-990

412

413 19. Team RC: R: A language and environment for statistical computing., 3.2.0 ed. Vienna, Austria,

- 414 R Foundation for Statistical Computing, 2014
- 415

416 20. Higgins JPT, Thompson SG, Deeks JJ, Altman DG: Measuring inconsistency in meta-

- 417 analyses. BMJ : British Medical Journal 2003;327:557-560
- 418

419 21. Bowden J: Assessing the suitability of summary data for Mendelian randomization analyses
420 using MR-Egger regression: the role of the I<sup>A</sup>2 statistic. Technical Report (available on request),
421 University of Bristol 2015

422

423 22. Stone M: Cross-Validatory Choice and Assessment of Statistical Predictions. Journal of the
 424 Royal Statistical Society Series B (Methodological) 1974;36:111-147

425 426

426 23. Venables WN, Ripley BD: *Modern Applied Statistics with S*. New York, Springer, 2002 427

428 24. Fox J, Weisberg S: An R Companion to Applied Regression. Los Angeles, Sage, 2011

429

430 25. Cook RD: Detection of Influential Observation in Linear Regression. Technometrics431 1977;19:15-18

432

433 26. Manning AK, Hivert M-F, Scott RA, Grimsby JL, Bouatia-Naji N, Chen H, Rybin D, Liu C-T, 434 Bielak LF, Prokopenko I, Amin N, Barnes D, Cadby G, Hottenga J-J, Ingelsson E, Jackson AU, 435 Johnson T, Kanoni S, Ladenvall C, Lagou V, Lahti J, Lecoeur C, Liu Y, Martinez-Larrad MT, 436 Montasser ME, Navarro P, Perry JRB, Rasmussen-Torvik LJ, Salo P, Sattar N, Shungin D, 437 Strawbridge RJ, Tanaka T, van Duijn CM, An P, de Andrade M, Andrews JS, Aspelund T, Atalay 438 M, Aulchenko Y, Balkau B, Bandinelli S, Beckmann JS, Beilby JP, Bellis C, Bergman RN, Blangero 439 J, Boban M, Boehnke M, Boerwinkle E, Bonnycastle LL, Boomsma DI, Borecki IB, Bottcher Y, 440 Bouchard C, Brunner E, Budimir D, Campbell H, Carlson O, Chines PS, Clarke R, Collins FS, 441 Corbaton-Anchuelo A, Couper D, de Faire U, Dedoussis GV, Deloukas P, Dimitriou M, Egan JM, 442 Eiriksdottir G, Erdos MR, Eriksson JG, Eury E, Ferrucci L, Ford I, Forouhi NG, Fox CS, Franzosi MG, Franks PW, Frayling TM, Froguel P, Galan P, de Geus E, Gigante B, Glazer NL, Goel A, 443 444 Groop L, Gudnason V, Hallmans G, Hamsten A, Hansson O, Harris TB, Hayward C, Heath S, 445 Hercberg S, Hicks AA, Hingorani A, Hofman A, Hui J, Hung J, Jarvelin M-R, Jhun MA, Johnson 446 PCD, Jukema JW, Jula A, Kao WH, Kaprio J, Kardia SLR, Keinanen-Kiukaanniemi S, Kivimaki M, 447 Kolcic I, Kovacs P, Kumari M, Kuusisto J, Kyvik KO, Laakso M, Lakka T, Lannfelt L, Lathrop GM, 448 Launer LJ, Leander K, Li G, Lind L, Lindstrom J, Lobbens S, Loos RJF, Luan Ja, Lyssenko V, Magi 449 R, Magnusson PKE, Marmot M, Meneton P, Mohlke KL, Mooser V, Morken MA, Miljkovic I, Narisu 450 N, O'Connell J, Ong KK, Oostra BA, Palmer LJ, Palotie A, Pankow JS, Peden JF, Pedersen NL, 451 Pehlic M, Peltonen L, Penninx B, Pericic M, Perola M, Perusse L, Peyser PA, Polasek O, 452 Pramstaller PP, Province MA, Raikkonen K, Rauramaa R, Rehnberg E, Rice K, Rotter JI, Rudan I, 453 Ruokonen A, Saaristo T, Sabater-Lleal M, Salomaa V, Savage DB, Saxena R, Schwarz P, Seedorf 454 U, Sennblad B, Serrano-Rios M, Shuldiner AR, Sijbrands EJG, Siscovick DS, Smit JH, Small KS, 455 Smith NL, Smith AV, Stancakova A, Stirrups K, Stumvoll M, Sun YV, Swift AJ, Tonjes A, 456 Tuomilehto J, Trompet S, Uitterlinden AG, Uusitupa M, Vikstrom M, Vitart V, Vohl M-C, Voight BF, 457 Vollenweider P, Waeber G, Waterworth DM, Watkins H, Wheeler E, Widen E, Wild SH, Willems 458 SM, Willemsen G, Wilson JF, Witteman JCM, Wright AF, Yaghootkar H, Zelenika D, Zemunik T, 459 Zgaga L, Wareham NJ, McCarthy MI, Barroso I, Watanabe RM, Florez JC, Dupuis J, Meigs JB, 460 Langenberg C: A genome-wide approach accounting for body mass index identifies genetic 461 variants influencing fasting glycemic traits and insulin resistance. Nat Genet 2012;44:659-669 462 463 27. Richmond RC, Davey Smith G, Ness AR, den Hoed M, McMahon G, Timpson NJ: Assessing 464 Causality in the Association between Child Adiposity and Physical Activity Levels: A Mendelian 465 Randomization Analysis. PLoS Med 2014;11:e1001618 466 467 28. Taylor PN, Richmond R, Davies N, Sayers A, Stevenson K, Woltersdorf W, Taylor A, Groom A,

- 468 Northstone K, Ring S, Okosieme O, Rees A, Nitsch D, Williams GR, Smith GD, Gregory JW,
- 469 Timpson NJ, Tobias JH, Dayan CM: Paradoxical Relationship Between Body Mass Index and
- 470 Thyroid Hormone Levels: A Study Using Mendelian Randomization. J Clin Endocrinol Metab
- 471 2016;101:730-738
- 472

- 473 29. Walter S, Kubzansky LD, Koenen KC, Liang L, Tchetgen Tchetgen EJ, Cornelis MC, Chang S-
- 474 C, Rimm E, Kawachi I, Glymour MM: Revisiting mendelian randomization studies of the effect of
- body mass index on depression. American Journal of Medical Genetics Part B: Neuropsychiatric
   Genetics 2015;168:108-115
- 477
- 478 30. Office of National Statistics(UK): Annual Mid-year Population Estimates, 2013. Statistical
   479 *bulletin*, 2014
- 480
- 481 31. The National Cardiovascular Intelligence Network: Cardiovascular disease key facts. In
- 482 Diabetes: Public Health England, 2013
- 483
- 484 32. Adult obesity and type 2 diabetes. Public Health England, 2014
- 485

## 486 Tables

- 487 Table 1 Estimates from the application of inverse-variance weighted, MR-Egger and weighted
- 488 median Mendelian randomisation methodologies. Estimates represent the estimated causal effect
- 489 of body mass index on type 2 diabetes.

Method	Estimate	95% CI	p-value
Complete variant set (n=96 S	NPs)		
IVW	1.20	1.09, 1.30	8.00 x 10 <sup>-05</sup>
MR-Egger	1.39	1.14, 1.68	1.53 x 10 <sup>-03</sup>
MR-Egger <sup>(α)</sup>	-0.019	-0.041, 0.004	0.10
Weighted median	1.26	1.17, 1.34	5.26 x 10 <sup>-9</sup>
TCF7L2(rs7903146) removed	from the variant set (r	1=95 SNPs)	
IVW	1.22	1.16, 1.28	1.49 x 10 <sup>-11</sup>
MR-Egger	1.34	1.17, 1.51	9.71 x 10 <sup>-06</sup>
MR-Egger <sup>(<i>a</i>)</sup>	-0.011	-0.024, -0.024	0.13
Weighted median	1.26	1.19, 1.32	3.29 x 10 <sup>-10</sup>
FTO(rs1558902) removed fro	m the variant set (n=98	5 SNPs)	
IVW	1.16	1.06, 1.27	1.31 x 10 <sup>-03</sup>
MR-Egger	1.30	1.01, 1.65	0.04
MR-Egger <sup>(<i>α</i>)</sup>	-0.012	-0.038, 0.014	0.34
Weighted median	1.21	1.13, 1.28	6.81 x 10 <sup>-08</sup>

490 Intercept coefficients MR-Egger<sup>( $\alpha$ )</sup> represent the average pleiotropic effect of a genetic variant on

491 type 2 diabetes risk. "IVW" refers to inverse variance weighted estimates, SNP refers to single492 nucleotide polymorphism.

493 Figures

494 Figure 1 – Genetic associations with body mass index (BMI) and type 2 diabetes from 96 variants measured in GIANT (17) and DIAGRAM (18),

- 495 respectively. *TCF7L2*(rs7903146) and *FTO*(rs1558902) are marked with a 'X' and labelled.
- 496 A funnel plot of minor allele frequency corrected genetic associations with BMI (interpreted as instrument strength) against causal estimates based
- 497 on each genetic variant individually, where the causal effect is expressed in logs odds ratio of type 2 diabetes for each unit increase in BMI. The
- 498 overall causal estimates (*β* coefficients) of BMI on type 2 diabetes estimated by inverse-variance weighted (solid black line), MR-Egger (dashed black
- 499 line) and weighted median (dotted black line) methods are also shown. Grey solid line represent x=0, that is a causal estimate of zero.
- 500 B scatter plot of genetic associations with type 2 diabetes against associations with BMI, with causal estimates (β coefficients) of BMI on type 2
- 501 diabetes estimated by inverse-variance weighted (solid line), MR-Egger (dashed line) and weighted median (dotted line) methods.
- 502
- 503 Figure 2 Distributions of regression estimates resulting from leave-one-out permutation analysis. Solid line = estimate from main analysis (n=96
- variants); dashed line = estimate with *TCF7L2*(rs7903146) removed; dotted line = estimate with *FTO*(rs1558902) removed.
- 505 A Causal estimates ( $\beta$  coefficients) of BMI on type 2 diabetes estimated by an inverse-variance weighted method
- 506 B Causal estimates ( $\beta$  coefficients) of BMI on type 2 diabetes estimated by MR-Egger
- 507 C Causal estimates ( $\beta$  coefficients) of BMI on type 2 diabetes estimated by a weighted median method
- 508 D Estimates of the intercept by MR-Egger











Figure 2







В



Median-based causal estimate

С



MR-Egger estimate of the intercept

D

# **Online Appendix**

## Supplementary methods

#### **Mendelian randomization framework**

Let  $\hat{\Gamma}_j$  equal the gene-outcome association estimate for variant j = 1, ..., J, with associated standard error  $\sigma_{y_j}$ . Let  $\hat{\gamma}_j$  equal the gene-exposure association estimate for variant j, with associated standard error  $\sigma_{x_j}$ . Let the causal effect of the exposure on the outcome be denoted by  $\beta$ . An estimate for  $\beta$  based on variant j alone can be obtained via the ratio method as

$$\hat{\beta}_j = \frac{\hat{\Gamma}_j}{\hat{\gamma}_j}$$

Two forms for the variance of  $\hat{\beta}_j$  are often used:

\_2

(i) 
$$Var(\hat{\beta}_j) = \frac{\sigma_{y_j}^2}{\hat{\gamma}_j^2}$$

(ii) 
$$Var(\hat{\beta}_j) = \frac{\delta_{Yj}}{\hat{\gamma}_j^2} + \frac{\Gamma_j \delta_{Xj}}{\hat{\gamma}_j^4}$$

Using either a first order (i) or second order (ii) Taylor series expansion. We use the variance from (i). This is equivalent to assuming that the gene-exposure association estimates are measured without error and is referred to as the No Measurement Error (NOME) assumption. NOME is equivalent to the assumption  $\sigma_{Xj}^2 = 0$  for all j, so that  $\hat{\gamma}_j = \gamma_j$  for all j.

The inverse variance weighted (IVW) method for the overall causal effect estimate Let  $w_j = 1/\operatorname{var}(\hat{\beta}_j)$  where  $\operatorname{var}(\hat{\beta}_j)$  is defined as in (i) under NOME. The inverse variance weighted (IVW) estimate for the causal effect is given by the standard meta-analytic formula

$$\frac{\sum_{j} w_{j} \hat{\beta}_{j}}{\sum_{j} w_{j}}$$

The  $w_j$  terms derived under NOME are also referred to as 'Toby Johnson' weights. The IVW estimate assumes that all genetic variants satisfy the instrumental variable assumptions. If this is not true then it could give a biased estimate for  $\beta$ . The IVW estimate for  $\beta$  is consistent even if all genetic variants are invalid, provided that:

- Across all variants, the magnitude of the gene exposure associations are independent of their pleiotropic effects (the InSIDE assumption)
- NOME is satisfied
- The pleiotropic effects have zero mean

The weighted median method for the overall causal effect estimate

Let  $\hat{\beta}_{(1)},...,\hat{\beta}_{(J)}$  equal the J causal effect estimates ordered from smallest ( $\hat{\beta}_{(1)}$ ) to largest ( $\hat{\beta}_{(J)}$ ). Now define

$$w_{(j)}^* = \frac{w_j}{S_j}$$
, where  $S_j = \sum_j w_j$ ,

and equate  $\hat{\beta}_{(i)}$  with a quantile,  $p^w_{(i)}$  , defined as

$$p_{(j)}^{w} = \frac{100}{S_{j}} \left( S_{(j)} - \frac{w_{(j)}}{2} \right).$$

 $p_{(j)}^{w}$  represents the quantile from the weighted empirical distribution function of the ordered estimates  $\hat{\beta}_{(1)}, \dots, \hat{\beta}_{(J)}$ . The weighted median estimate,  $\hat{\beta}_{WM}$  is defined as the 50th percentile of this weighted distribution. Typically the 50th percentile will lie between two estimates ( $\hat{\beta}_{(l)}$  and  $\hat{\beta}_{(m)}$ , say), in which case  $\hat{\beta}_{WM}$  is found by linear interpolation.

 $\hat{\beta}_{_{W\!M}}$  is a consistent estimate for  $\beta$  provided that at least 50% of the `weight' making up  $S_{_J}$  comes from genetic variants that are valid instruments.

## The MR-Egger method for the overall causal effect estimate

The MR-Egger method performs a weighted linear regression of the gene-outcome coefficients on the gene-exposure coefficients:

$$\frac{\hat{\Gamma}_{j}}{\sigma_{y_{j}}} = \frac{\beta_{0E}}{\sigma_{y_{j}}} + \beta_{1E} \frac{\hat{\gamma}_{j}}{\sigma_{y_{j}}}$$

The weights used are also derived under the NOME assumption. If all genetic variants are valid instruments, then  $\beta_{0E} = 0$ . The value of  $\hat{\beta}_{0E}$  can be interpreted as an estimate of the average pleiotropic effect across the genetic variants. An intercept term that differs from zero is indicative of overall directional pleiotropy. The MR-Egger estimate for  $\beta$ ,  $\hat{\beta}_{1E}$ , is consistent even if all genetic variants are invalid, provided that:

• Across all variants, the magnitude of the gene exposure associations are independent of their pleiotropic effects (the InSIDE assumption)

NOME is satisfied.

If NOME is violated then the MR-Egger estimate of causal effect will be attenuated towards the null. We can assess the strength of NOME violation for MR-Egger through the  $I_{GX}^2$ 

statistic:  $I_{GX}^2 = \frac{Q-df}{Q}$ , where  $Q = \sum_{J=1}^{J} \frac{\left(\frac{\hat{\gamma}_j}{\sigma_{Y_j}^2} - \bar{\gamma}\right)^2}{\frac{\sigma_{X_j}^2}{\sigma_{Y_j}^2}}$  and where  $\bar{\gamma}$  equals the arithmetic mean of

the  $\hat{\gamma}_j / \sigma_{\gamma_j}^2$  terms . Specifically, the  $I_{GX}^2$  statistic quantifies the proportion of the total variation

between the  $\hat{\gamma}_j / \sigma_{\gamma_j}^2$  terms that is due to `true' variation between the  $\gamma_j / \sigma_{\gamma_j}^2$  terms.

Consequently, when NOME is satisfied  $\hat{\gamma}_1, ..., \hat{\gamma}_J = \gamma_1, ..., \gamma_J$ ,  $I_{GX}^2$  equals 1, and no attenuation occurs. When  $I_{GX}^2 = 0.9$  we can expect the MR-Egger estimate to be only 90% of its value had NOME been satisfied. A crude correction for NOME violation would be  $\frac{\hat{\beta}_{1E}}{I_{GX}^2}$ , however this can be unstable as  $I_{GX}^2$  can sometimes be estimated as zero, even when it is truly large. We used the established method of Simulation Extrapolation (SIMEX) (1) instead, as implemented using the R package simex() (2). Under SIMEX, new data sets are created by simulating gene-exposure association estimates under increasing violations of NOME and recording the amount of attenuation in the estimate that occurs. The set of attenuated estimates are then used to extrapolate back to the estimate that would have been obtained if NOME had been satisfied.

# **Supplementary Results**

# **Outlier analysis – Studentized residuals**



Figure S1A – Studentised residuals applied to the IVW method.



Figure S1B – Studentised residuals applied to the MR-Egger method.s





Figure S2A – Cook's distance applied to the IVW method.



Figure S2B – Cook's distance applied to the MR-Egger method.



# Reciprocal analysis of type 2 diabetes and BMI



A - scatter plot of genetic associations with BMI against associations with type 2 diabetes, with causal estimates ( $\beta$  coefficients) of type 2 diabetes on BMI estimated by inverse-variance weighted (red line), MR-Egger (blue line) and median-based (green line) methods. For this analysis, all 115 confirmed type 2 diabetes associated loci with OR not equal to 1 from Morris et al (2012)(3) downloaded from DIAGRAM http://diagram-consortium.org/downloads.html) were used.

A - scatter plot of genetic associations with BMI against associations with type 2 diabetes, with causal estimates ( $\beta$  coefficients) of type 2 diabetes on BMI estimated by inverse-variance weighted (red line), MR-Egger (blue line) and median-based (green line) methods. For this analysis, 110 confirmed type 2 diabetes associated loci with OR not equal to 1 and no overlapping known BMI loci (excluding *FTO*, *MC4R* and *TCF7L2*) from Morris et al (2012)(3) were again used.

## References

1. Cook JR, Stefanski LA: Simulation-Extrapolation Estimation in Parametric Measurement Error Models. Journal of the American Statistical Association 1994;89:1314-1328

2. Lederer W, Küchenhoff H: simex: SIMEX- and MCSIMEX-Algorithm for measurement error models., 2013

3. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segrè AV, Steinthorsdottir V, Strawbridge RJ, Khan H, Grallert H, Mahajan A, Prokopenko I, Kang HM, Dina C, Esko T, Fraser RM, Kanoni S, Kumar A, Lagou V, Langenberg C, Luan Ja, Lindgren CM, Müller-Nurasyid M, Pechlivanis S, Rayner NW, Scott LJ, Wiltshire S, Yengo L, Kinnunen L, Rossin EJ, Ravchaudhuri S, Johnson AD, Dimas AS, Loos RJF, Vedantam S, Chen H, Florez JC, Fox C, Liu C-T, Rybin D, Couper DJ, Kao WHL, Li M, Cornelis MC, Kraft P, Sun Q, van Dam RM, Stringham HM, Chines PS, Fischer K, Fontanillas P, Holmen OL, Hunt SE, Jackson AU, Kong A, Lawrence R, Meyer J, Perry JRB, Platou CGP, Potter S, Rehnberg E, Robertson N, Sivapalaratnam S, Stančáková A, Stirrups K, Thorleifsson G, Tikkanen E, Wood AR, Almgren P, Atalay M, Benediktsson R, Bonnycastle LL, Burtt N, Carey J, Charpentier G, Crenshaw AT, Doney ASF, Dorkhan M, Edkins S, Emilsson V, Eury E, Forsen T, Gertow K. Gigante B, Grant GB, Groves CJ, Guiducci C, Herder C, Hreidarsson AB, Hui J, James A, Jonsson A, Rathmann W, Klopp N, Kravic J, Krjutškov K, Langford C, Leander K, Lindholm E, Lobbens S, Männistö S, Mirza G, Mühleisen TW, Musk B, Parkin M, Rallidis L, Saramies J, Sennblad B, Shah S, Sigurðsson G, Silveira A, Steinbach G, Thorand B, Trakalo J, Veglia F, Wennauer R, Winckler W, Zabaneh D, Campbell H, van Duijn C, Uitterlinden AG, Hofman A, Sijbrands E, Abecasis GR, Owen KR, Zeggini E, Trip MD, Forouhi NG, Syvänen A-C, Eriksson JG, Peltonen L, Nöthen MM, Balkau B, Palmer CNA, Lyssenko V, Tuomi T, Isomaa B, Hunter DJ, Qi L, Wellcome Trust Case Control C, Investigators M, Consortium G, Consortium A-TD, Consortium SD, Shuldiner AR, Roden M, Barroso I, Wilsgaard T, Beilby J, Hovingh K, Price JF, Wilson JF, Rauramaa R, Lakka TA, Lind L, Dedoussis G, Njølstad I, Pedersen NL, Khaw K-T, Wareham NJ, Keinanen-Kiukaanniemi SM, Saaristo TE, Korpi-Hyövälti E, Saltevo J, Laakso M, Kuusisto J, Metspalu A, Collins FS, Mohlke KL, Bergman RN, Tuomilehto J, Boehm BO, Gieger C, Hveem K, Cauchi S, Froquel P, Baldassarre D, Tremoli E, Humphries SE, Saleheen D, Danesh J, Ingelsson E, Ripatti S, Salomaa V, Erbel R, Jöckel K-H, Moebus S, Peters A, Illig T, de Faire U, Hamsten A, Morris AD, Donnelly PJ, Frayling TM, Hattersley AT, Boerwinkle E, Melander O, Kathiresan S, Nilsson PM, Deloukas P, Thorsteinsdottir U, Groop LC, Stefansson K, Hu F, Pankow JS, Dupuis J, Meigs JB, Altshuler D, Boehnke M, McCarthy MI: Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nature genetics 2012;44:981-990

1	Title: Body mass index as a modifiable risk factor for type 2 diabetes: Refining and understanding	Deleted: Resolving
2	causal estimates using Mendelian randomisation.	
3		
4	Authors:	
5	Laura J Corbin <sup>1</sup> ; Rebecca C Richmond <sup>1</sup> ; Kaitlin H Wade <sup>1</sup> ; Stephen Burgess <sup>1,3</sup> ; Jack Bowden <sup>1,2</sup> ;	
6	George Davey Smith <sup>1</sup> ; Nicholas J Timpson <sup>1</sup>	
7		
8	Affiliations:	
9	1) MRC Integrative Epidemiology Unit (IEU) at University of Bristol, Bristol, UK	
10	2) MRC Biostatistics Unit, Cambridge Institute of Public Health, Cambridge, UK	
11	3) Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK	
12		
13	Running title: Body mass index and type 2 diabetes; Mendelian randomisation methods	
14		
15		
16		
17		
18		
19	Corresponding author: Nicholas J Timpson, Email: <u>N.J.Timpson@bristol.ac.uk</u> , Address: MRC	
20	Integrative Epidemiology Unit, University of Bristol, Oakfield House, Oakfield Grove, Bristol, BS8	

21 2BN, Tel: 0117 3310131.

## Page 29 of 45

39

#### Diabetes

#### 23 ABSTRACT

24	This study focused on resolving the relationship between body mass index (BMI) and type 2	Deleted: is a worked example of causal analysis
25	diabetes. The availability of multiple variants associated with BMI offers a new chance to resolve	Formatted: Not Highlight
26	the true causal effect of BMI on T2D, however the properties of these associations and their	
27	validity as genetic instruments need to be considered alongside established and new methods for	
28	undertaking Mendelian randomisation. We explore the potential for pleiotropic genetic variants to	
29	generate bias, revise existing estimates and illustrate value in new analysis methods. A two-	
30	sample Mendelian randomisation (MR) approach with 96 genetic variants was employed using	
31	three different analysis methods, two of which (MR-Egger and the weighted median) have been	Deleted: å
32	developed specifically to address problems of invalid instrumental variables. We estimate an odds	
33	ratio for type 2 diabetes per unit increase in BMI ( $kg/m^2$ ) of between 1.19 and 1.38, with the most	
34	stable estimate using all instruments and a weighted median approach (1.26 <u>95%CI (1.17, 1.34)</u> ).	
35	TCF7L2(rs7903146) was identified as a complex effect or pleiotropic instrument and removal of	Deleted: potentially
36	this variant resulted in convergence of causal effect estimates from different causal analysis	Deleted: locus
37	methods. This indicated the potential for pleiotropy to affect estimates and differences in	
38	performance of alternative analytical methods. In a real type 2 diabetes focused example, this	

40 potential for new approaches to mitigate the bias caused.

study demonstrates the potential impact of invalid instruments on causal effect estimates and the

45	Observational studies have shown body mass index (BMI) to be associated with risk of type 2		
46	diabetes as well as with a range of diabetes-related metabolic traits (1; 2). However, it is well		
47	known that confounding, reverse causation and biases can generate such associations and that		
48	even with careful study design, incorrect inference is possible (3). One approach to circumventing		
49	these problems is to use genetic association results within a Mendelian randomization (MR)		Deleted: improve causal inference
50	framework (3; 4). In MR analyses, genetic variants act as proxies for an exposure in a manner		
51	independent of confounders, If in addition the variants only affect an outcome of interest through		Deleted: ,
52	the chosen exposure, then they are said to be valid instrumental variables (IVs). This enables		Deleted:
53	evaluation of the causal effect of the exposure on the outcome, escaping some of the limitations of		Deleted: ing
55		~	Deleted: that
54	observational epidemiology: (5).		Deleted: an
١		$\langle     \rangle$	Deleted:
55		$\langle \rangle \rangle$	Deleted: of interest
56	Following the success of genome-wide association studies (GWASs), the number of MR analyses		<b>Deleted:</b> . Here, genetic variation fulfils the role of an instrumental variable (IV) which can be thought of as one able to predict an exposure of interest, but which
57	using large numbers of mostly uncharacterized variants associated with complex health outcomes		Deleted: es
50		, I)	Deleted: i.e. confounding, bias and reverse causation
58	or intermediates is rapidly increasing (6; 7). In the case of BMI, there are now 97 genetic variants	$\mathbb{N}$	Deleted:
59	reliably associated and there are examples where multiple variants have been used as a	Ì	Deleted: s
			Deleted: already
60	composite IV to estimate the causal impact of BMI on health (8). Although using many IVs can		
61	increase the power of MR analyses, it brings with it the concern that enlarged sets of genetic		Deleted: in
62	variants are more likely to contain invalid IVs due to violations of the assumptions necessary for	$\mathbb{N}$	Deleted: is
63	valid causal inference using traditional methods (9). In particular, horizontal pleiotropy – where a	$\mathbb{N}$	Deleted: has some benefit in explaining more variance in the exposure of interest
	······································		Deleted: this kind of approach
64	genetic variant affects the outcome via more than one biological pathway (10) <u>– is a concern</u> .		
65	Importantly, the properties of these associations and their validity as genetic instruments need to		Formatted: Not Highlight
66	be considered alongside established and new methods for undertaking Mendelian randomisation.		
67			
68	In response to the general issue of using multiple genetic variants in MR, Bowden et al. (9)		Deleted:
69	propose both MR-Egger regression, an approach developed from the original Egger regression		Formatted: Font:Bold
70	technique for assessing small study bias in meta-analysis and a weighted weighted median		Deleted: median-based
71	approach (11) as alternatives to the standard MR analysis The MR-Edger and weighted weighted	<u></u>	Deleted: It has been suggested that both the
, 1	wrphosen ( ) / as allothearter to the standard in Caller old, and in Caller and Woldflice		Deleted: median-based
72	median, approaches both operate using distinct, but critically weaker, versions of the IV	· /	Deleted:
72	accumptions, and therefore have the notential to deliver rejust equal effect estimates. The MD	1	Deleted: ,
13	assumptions, and therefore have the potential to deliver robust causal effect estimates. The MR-		Deleted: with t
1	3		

130

Diabetes

102	Egger method also provides, a formal statistical test as to whether or not the average pleiotropic		Deleted: ing
103	effect of the genetic variants is equal to zero (9).		
104			
104			
105	Research Design and Methods		Formatted: Font:Bold
106	With increasing evidence for multiple biological pathways underlying type 2 diabetes (12; 13) and		Deleted: amounts of
107	increasing numbers of genetic variants available as IVs for BMI, we set out to test the notential for		Deleted: pointing towards
107		1	<b>Deleted:</b> instruments for exposures of interest, such as
108	bias in causal estimates from MR using these state-of-the-art approaches. We compared results		Deleted: hypothesise
109	from MR-Egger regression (9) and weighted weighted median (11) approaches to a traditional		Deleted: that
110		M/	Deleted:
110	inverse-variance weighted (IVW) method (which makes the strong assumption that all variants are		Deleted: best current methods
111	valid IVs) (14) in an investigation of the causal relationship between BMI and type 2 diabetes.		Deleted: is non-negligible
112	These methods all undertake two servels Mandalian reademisation whereby the CMAS results for		Deleted: median-based
112	These methods all undertake two-sample Mendelian randomisation whereby the GWAS results for		
113	a disease outcome are unified with those of an exposure of interest and together used to estimate		
114	the causal impact of that exposure on disease. We used published data in a two-sample analysis		
115	strategy taking SNP-exposure and SNP-outcome associations from different sources (15: 16).		
44.0			
116	۲		Deleted: . ( [1])
116 117	The effect sizes for BMI-associated SNPs with associated standard errors from a mixed-sex cohort		Deleted:[1] Deleted: their effect sizes and
116 117 118	The effect sizes for BMI-associated SNPs with associated standard errors from a mixed-sex cohort		Deleted:
116 117 118	The effect sizes for BMI-associated SNPs with associated standard errors from a mixed-sex cohort of European ancestry, were taken from the Genetic Investigation of ANthropometric Traits (GIANT)		Deleted:
116 117 118 119	The effect sizes for BMI-associated SNPs with associated standard errors from a mixed-sex cohort of European ancestry were taken from the Genetic Investigation of ANthropometric Traits (GIANT) consortium (17) along with results for type 2 diabetes from the DIAbetes Genetics Replication And		Deleted:       [1]         Deleted:       their effect sizes and         Deleted:       estimated in         Deleted:       . Results are taken         Deleted:       . Results are taken         Deleted:       with cohorts which also contributed data to
116 117 118 119 220	The effect sizes for BMI-associated SNPs with associated standard errors from a mixed-sex cohort of European ancestry, were taken from the Genetic Investigation of ANthropometric Traits (GIANT) consortium (17) along with results for type 2 diabetes from the DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium. To avoid sample overlap, GIANT estimates were re-		Deleted:       [1]         Deleted:       their effect sizes and         Deleted:       estimated in         Deleted:       . Results are taken         Deleted:       . Results are taken         Deleted:       with cohorts which also contributed data to
116 117 118 119 120 221	<u>The effect sizes for BMI-associated SNPs with associated standard errors from a mixed-sex cohort</u> of European ancestry <u>were taken</u> from the Genetic Investigation of ANthropometric Traits (GIANT) consortium (17) <u>along with results for type 2 diabetes from the DIAbetes Genetics Replication And</u> Meta-analysis (DIAGRAM) Consortium <u>. To avoid sample overlap, GIANT estimates were re-</u> <u>calculated in the absence of DIAGRAM cohorts yielding a maximum sample size at any given</u>		Deleted:
116 117 118 119 120 121	The effect sizes for BMI-associated SNPs with associated standard errors from a mixed-sex cohort of European ancestry, were taken from the Genetic Investigation of ANthropometric Traits (GIANT) consortium (17) along with results for type 2 diabetes from the DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium. To avoid sample overlap, GIANT estimates were re- calculated in the absence of DIAGRAM cohorts yielding a maximum sample size at any given		Deleted:      [1]         Deleted:       their effect sizes and         Deleted:       estimated in         Deleted:       . Results are taken         Deleted:       . Results are taken         Deleted:       with cohorts which also contributed data to         Deleted:       removed to avoid sample overlap         Deleted:       (
116 117 118 119 120 121 122	The effect sizes for BMI-associated SNPs with associated standard errors from a mixed-sex cohort of European ancestry, were taken from the Genetic Investigation of ANthropometric Traits (GIANT) consortium (17) along with results for type 2 diabetes from the DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium. To avoid sample overlap, GIANT estimates were re- calculated in the absence of DIAGRAM cohorts yielding a maximum sample size at any given locus of 189,079. To aid interpretation of the effects of BMI on type 2 diabetes, effect sizes were		Deleted:      [1]         Deleted:       their effect sizes and         Deleted:       estimated in         Deleted:       . Results are taken         Deleted:       . Results are taken         Deleted:       with cohorts which also contributed data to         Deleted:       removed to avoid sample overlap         Deleted:       (         Deleted:       was
116 117 118 119 120 121 122 123	The effect sizes for BMI-associated SNPs with associated standard errors from a mixed-sex cohort of European ancestry were taken from the Genetic Investigation of ANthropometric Traits (GIANT) consortium (17) along with results for type 2 diabetes from the DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium. To avoid sample overlap, GIANT estimates were re- calculated in the absence of DIAGRAM cohorts yielding a maximum sample size at any given locus of 189,079. To aid interpretation of the effects of BMI on type 2 diabetes, effect sizes were transformed to BMI units prior to analysis, assuming one standard deviation (SD) = 4.5kg/m <sup>2(17)</sup> .		Deleted:      [1]         Deleted:       their effect sizes and         Deleted:       estimated in         Deleted:       . Results are taken         Deleted:       . Results are taken         Deleted:       with cohorts which also contributed data to         Deleted:       removed to avoid sample overlap         Deleted:       (         Deleted:       was         Deleted:       )         Deleted:
116 117 118 119 120 121 122 123 124	The effect sizes for BMI-associated SNPs with associated standard errors from a mixed-sex cohort of European ancestry, were taken from the Genetic Investigation of ANthropometric Traits (GIANT) consortium (17) along with results for type 2 diabetes from the DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium. To avoid sample overlap, GIANT estimates were re- calculated in the absence of DIAGRAM cohorts yielding a maximum sample size at any given locus of 189,079. To aid interpretation of the effects of BMI on type 2 diabetes, effect sizes were transformed to BMI units prior to analysis, assuming one standard deviation (SD) = 4.5kg/m <sup>2(17)</sup> . For the corresponding SNP-outcome association, we took odds ratios (ORs) and confidence		Deleted:      [1]         Deleted:       their effect sizes and         Deleted:       estimated in         Deleted:       . Results are taken         Deleted:       . Results are taken         Deleted:       with cohorts which also contributed data to         Deleted:       removed to avoid sample overlap         Deleted:       (         Deleted:          Deleted:
116 117 118 119 120 121 122 123 124 125	The effect sizes for BMI-associated SNPs with associated standard errors from a mixed-sex cohort of European ancestry, were taken from the Genetic Investigation of ANthropometric Traits (GIANT) consortium (17) along with results for type 2 diabetes from the DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium. To avoid sample overlap, GIANT estimates were re- calculated in the absence of DIAGRAM cohorts yielding a maximum sample size at any given locus of 189,079. To aid interpretation of the effects of BMI on type 2 diabetes, effect sizes were transformed to BMI units prior to analysis, assuming one standard deviation (SD) = 4.5kg/m <sup>2(17)</sup> . For the corresponding SNP-outcome association, we took odds ratios (ORs) and confidence intervals from a GWAS meta-analysis conducted by the DIAGRAM Consortium. This genome-wide		Deleted:      [1]         Deleted:       their effect sizes and         Deleted:       estimated in         Deleted:
116 117 118 119 120 121 122 123 124 125 126	The effect sizes for BMI-associated SNPs with associated standard errors from a mixed-sex cohort of European ancestry, were taken from the Genetic Investigation of ANthropometric Traits (GIANT) consortium (17) along with results for type 2 diabetes from the DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium. To avoid sample overlap, GIANT estimates were re- calculated in the absence of DIAGRAM cohorts yielding a maximum sample size at any given locus of 189,079. To aid interpretation of the effects of BMI on type 2 diabetes, effect sizes were transformed to BMI units prior to analysis, assuming one standard deviation (SD) = 4.5kg/m <sup>2(17)</sup> . For the corresponding SNP-outcome association, we took odds ratios (ORs) and confidence intervals from a GWAS meta-analysis conducted by the DIAGRAM Consortium. This genome-wide meta-analysis includes data from 12,171 type 2 diabetes cases and 56,862 controls of mainly		Deleted:      [1]         Deleted:       their effect sizes and         Deleted:       estimated in         Deleted:
116 117 118 119 120 121 122 123 124 125 126 127	The effect sizes for BMI-associated SNPs with associated standard errors from a mixed-sex cohort of European ancestry, were taken from the Genetic Investigation of ANthropometric Traits (GIANT) consortium (17) along with results for type 2 diabetes from the DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium. To avoid sample overlap, GIANT estimates were re- calculated in the absence of DIAGRAM cohorts yielding a maximum sample size at any given locus of 189,079. To aid interpretation of the effects of BMI on type 2 diabetes, effect sizes were transformed to BMI units prior to analysis, assuming one standard deviation (SD) = 4.5kg/m <sup>2(17)</sup> . For the corresponding SNP-outcome association, we took odds ratios (ORs) and confidence intervals from a GWAS meta-analysis conducted by the DIAGRAM Consortium. This genome-wide meta-analysis includes data from 12,171 type 2 diabetes cases and 56,862 controls of mainly European descent imputed at up to 2.5 million autosomal SNPs (DIAGRAMv3) (18). All but one		Deleted: .      [1]         Deleted: their effect sizes and
116 117 118 119 120 121 122 123 124 125 126 127 128	The effect sizes for BMI-associated SNPs with associated standard errors from a mixed-sex cohort of European ancestry were taken from the Genetic Investigation of ANthropometric Traits (GIANT) consortium (17) along with results for type 2 diabetes from the DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium. To avoid sample overlap, GIANT estimates were re- calculated in the absence of DIAGRAM cohorts yielding a maximum sample size at any given locus of 189,079. To aid interpretation of the effects of BMI on type 2 diabetes, effect sizes were transformed to BMI units prior to analysis, assuming one standard deviation (SD) = 4.5kg/m <sup>2(17)</sup> . For the corresponding SNP-outcome association, we took odds ratios (ORs) and confidence intervals from a GWAS meta-analysis conducted by the DIAGRAM Consortium. This genome-wide meta-analysis includes data from 12,171 type 2 diabetes cases and 56,862 controls of mainly European descent imputed at up to 2.5 million autosomal SNPs (DIAGRAMv3) (18). All but one (rs4787491, <i>INO80E</i> ) of the BMI-associated SNPs ( <i>p</i> <5x10 <sup>-8</sup> ) from GIANT had results listed in the		Deleted:      [1]         Deleted:       their effect sizes and         Deleted:       estimated in         Deleted:

154	SNP-exposure and SNP-outcome associations were combined using the three different		
155	approaches outlined above, All analyses were conducted in R 3.2.0 (19). First, an inverse-variance		Deleted: suitable for estimating the overall causal
156	weighted (IVW) method was implemented to provide a weighted average of the causal effect	A A A A A A A A A A A A A A A A A A A	effect of an exposure on an outcome when multiple instruments are available in a two-sample setting
157	estimates (14). This method assumes that all constitution variants (i.e. $1000$ ) satisfy the $W$		Deleted: traditional
137			Deleted:
158	assumptions (including zero pleiotropy) and uses weights that assume the gene-exposure		
159	association estimates are measured without error (the No Measurement Error (NOME)		
160	assumption).		
161			
162	Second, we performed MR-Egger regression (9), which assumes NOME but allows each variant to		
163	exhibit pleiotropy. MR-Egger estimates remain consistent only if the magnitude of the gene		
164	exposure associations across all variants are independent of their pleiotropic effects (the InSIDE		
165	assumption) (9). As recommended by Bowden et al (9), the extent to which pleiotropy was		
166	balanced across the set of instruments as a whole was visually assessed by plotting the causal		
167	effect estimates against their precision, using a funnel plot and checking for asymmetry (Figure		Formatted: Not Highlight
168	1A). The NOME assumption was assessed for MR-Egger via an adaptation of the $l^2$ statistic ( $l_{GX}^2$ )		
169	(20) and adjusted for by combining MR-Egger with the method of Simulation Extrapolation		
170	(SIMEX) (21). Using SIMEX, new data sets are created by simulating gene-exposure association		Deleted: , as previously described
171	estimates under increasing violations of NOME and recording the amount of attenuation in the		Deleted: Under Formatted: Font:(Default) Arial, 11 pt
172	estimate that occurs. The set of attenuated estimates are then used to extrapolate back to the	1	Formatted: Font:(Default) Arial, 11 pt
173	estimate that would have been obtained if NOME had been satisfied.		
174			
175	Finally, a weighted weighted median estimation method was applied (11). The weighted median		Deleted: median-based
176	provides a consistent estimate of causal effect if at least 50% of the information in the analysis		
177	comes from variants that are valid IVs. For a more detailed description of the three methods		
178	applied, see Online Appendix (Supplementary methods), A leave-one-out permutation analysis		Deleted: Supplementary Data (Technical Appendix)
179	was conducted across all methods to assess the influence of potentially pleiotropic SNPs on the		
180	causal estimates (22). In the case of the linear models (IVW and MR-Egger) two additional		
181	analyses were conducted (23; 24). Firstly, the extent to which the causal estimate from each SNP		

192	in the set could be considered an outlier was assessed using studentized residuals. Secondly,	
193	Cook's distance (25) was used as a measure of the aggregate impact of each SNP on the model.	
194		
195	<u>Results</u>	Formatted: Font:(Default) Arial, 11 pt, Bold
196	All three approaches provide evidence of a positive causal relationship between BMI and type 2	Formatted: Font:Bold
197	diabetes. This is demonstrated in Figure 1B where the slope of the lines show the causal effect	Formatted: Not Highlight
198	estimates as predicted by the IVW, MR-Egger and m weighted median approaches. Estimates	Deleted: edian-based
199	correspond to an OR for type 2 diabetes per unit increase in BMI ( $kg/m^2$ ) of 1.19, 1.26 and 1.38 for	
200	the IVW, weighted median and MR-Egger analyses, respectively and are in line with a previous	Deleted: median-based
201	MR estimate of 1.27 (95%CI 1.18, 1.36) (2) (Table 1). Assessment of the NOME assumption with	
202	respect to the MR-Egger estimate gave $I_{GX}^2$ =0.83, suggesting an approximate 15% attenuation of	
203	the causal estimate towards zero. Bias adjustment via SIMEX gave a corrected MR-Egger causal	
204	estimate of 1.46 (95%CI 1.16, 1.84) for type 2 diabetes per unit increase in BMI (kg/m <sup>2</sup> ).	<b>Deleted:</b> 0.38 (0.15, 0.61), that is, an OR for type 2 diabetes per unit increase in BMI (kg/m <sup>2</sup> ) of
205		
206	Considering the individual SNP-based contributions to MR analysis, there is one clear outlier in the	
207	distribution of effects shown in Figure 1 and that is TCF7L2(rs7903146). TCF7L2(rs7903146)	
208	shows an association with BMI that is in the opposite direction to the overall trend (and weak	
209	relative to its effect on type 2 diabetes), resulting in a large negative causal estimate from this SNP	
210	alone. The presence of at least some unbalanced pleiotropy was detected within the set of	Deleted: (or `directional')
<b>2</b> 11	variants, as reflected by the intercept estimate of -0.019 (p=0.10) in the MR-Egger analysis.	Deleted: within
 212		Deleted: was also Deleted: The extent to which pleiotropic SNPs are
213	To illustrate the impact of TCF7L2(rs7903146) on causal estimates, we performed a sensitivity	their strength and the extent to which pleiotropy is balanced across the set of instruments as a whole.
214	analysis in which each SNP in turn was removed from the set in a leave-one-out permutation	Here we see only slight asymmetry (Figure 1B), suggesting the impact of pleiotropy on our estimates is likely to be relatively small
215	analysis. We saw a shift in the causal estimates from the IVW (an increase) and MR-Egger (a	
216	decrease) as a result of the removal of TCF7L2(rs7903146) but no difference in the estimate from	
217	the weighted median approach (Table 1; Figure 2). The results of the leave-one-out permutation	Deleted: median-based
218	analysis showed that the impact of removing TCF7L2(rs7903146) from the variant set on the IVW	
219	and MR-Egger estimates was greater than that of removing almost any other variant, with the	
220	exception of FTO(rs1558902) (Figure 2A & B). When FTO(rs1558902) was removed, causal	
	6	

<ul> <li>237</li> <li>238</li> <li>239</li> <li>240</li> <li>241</li> <li>41</li> <li>40</li> </ul>	instance we also observed movement in the causal effect estimate from the <u>weighted median</u> (Table 1; Figure 2C). The estimate of the intercept from MR-Egger moved closer to zero following both the removal of <i>TCF7L2</i> (rs7903146) and <i>FTO</i> (rs1558902) (Figure 2D). <i>TCF7L2</i> (rs7903146) was also identified as an outlier in both IVW and MR-Egger (studentized residuals, Bonferroni corrected $p$ <1x10 <sup>-19</sup> ) but <i>FTO</i> (rs1558902) was not ( <u>Online Appendix (Supplementary Results</u> , Figures S1A/B)). Calculation of Cook's distance showed both variants to have a disproportionate	Deleted: median-based approach
238 239 240 241	(Table 1; Figure 2C). The estimate of the intercept from MR-Egger moved closer to zero following both the removal of <i>TCF7L2</i> (rs7903146) and <i>FTO</i> (rs1558902) (Figure 2D). <i>TCF7L2</i> (rs7903146) was also identified as an outlier in both IVW and MR-Egger (studentized residuals, Bonferroni corrected $p$ <1x10 <sup>-19</sup> ) but <i>FTO</i> (rs1558902) was not ( <u>Online Appendix (Supplementary Results,</u> )	Deleted: Supplementary Data
239 240 241	both the removal of <i>TCF7L2</i> (rs7903146) and <i>FTO</i> (rs1558902) (Figure 2D). <i>TCF7L2</i> (rs7903146) was also identified as an outlier in both IVW and MR-Egger (studentized residuals, Bonferroni corrected $p$ <1x10 <sup>-19</sup> ) but <i>FTO</i> (rs1558902) was not ( <u>Online Appendix (Supplementary Results,</u> )	Deleted: Supplementary Data
240 241	was also identified as an outlier in both IVW and MR-Egger (studentized residuals, Bonferroni corrected $p$ <1x10 <sup>-19</sup> ) but <i>FTO</i> (rs1558902) was not ( <u>Online Appendix (Supplementary Results,</u>	Deleted: Supplementary Data
241	corrected $p < 1 \times 10^{-19}$ ) but <i>FTO</i> (rs1558902) was not (Online Appendix (Supplementary Results, Figures S1A/B)) Calculation of Cook's distance showed both variants to have a disproportionate	Deleted: Supplementary Data
	Figures S1A/B)) Calculation of Cook's distance showed both variants to have a disproportionate	
242		
243	level of influence on the model compared to other variants in the set (Online Appendix	
244	(Supplementary Results, Figures S2A/B)),	Deleted: (Supplementary Data)
245		
246	These results suggest TCF7L2(rs7903146) may be pleiotropic with respect to the outcome, i.e.	
247	that it influences type 2 diabetes through an alternative pathway (other than BMI). Evidence from	
248	existing literature supports this assertion as the type 2 diabetes risk increasing allele at	
249	TCF7L2(rs7903146) has been associated with both increased fasting glucose (26) and decreased	
250	BMI (17). Under the assumption that TCF7L2(rs7903146) demonstrates horizontal pleiotropy with	
251	respect to type 2 diabetes, we would expect its inclusion in the variant set to bias the causal	
252	estimate predicted by the IVW approach, but not that predicted by MR-Egger or the <u>weighted</u>	Deleted: median-based approaches
253	median. Removing TCF7L2(rs7903146) from the variant set causes a slight shift in the causal	
254	estimates from the IVW and MR-Egger approaches, bringing them more in line with one another	
255	and also with the weighted median estimate which remained stable in this instance. Also of note is	Deleted: median-based
256	the reduction in the 95% confidence interval of the MR-Egger estimate following removal of the	
257	TCF7L2(rs7903146). This increase in precision following removal of a likely invalid instrument from	
258	the set is another potentially favourable quality of this estimator. The relatively small changes	
259	observed across all methods as a result of removing TCF7L2(rs7903146) are in line with the	
260	relatively weak effect of the SNP as shown in Figure 1B.	
261		
262	In contrast, the effect of removing FTO(rs1558902) is more noticeable. Regardless of the method	
263	used, removing this variant results in a lower causal estimate (Table 1; Figure 2). The substantial	
264	influence of FTO(rs1558902) was predicable given the strength of its effect relative to the other	Deleted: can be predicted because of

# Page 35 of 45

## Diabetes

271	variants (Figure 1B), though properties of this effect are not in line with other variants used to	
272	instrument BMI as reported elsewhere for physical activity (27), thyroid function (28) and	 <b>Deleted:</b> and has been observed previously, for example, in the context of
273	depression (29). The concomitant increase in standard error associated with the estimates here	Deleted:
274	point towards increased uncertainty moving the estimates towards the null in the absence of	Deleted:
<b>1</b> 75	ETO(ro1559002) The weighted modion appears reput over to the removal of ETO(ro1559002)	Formatted: Not Highlight
1/5	Profis 1556902). The weighted median appears robust, even to the removal of Profis 1556902),	 Deleted: median-based approach
276	as demonstrated by the relatively tight distribution of estimates returned from the leave-one-out	
277	permutation analysis (Figure 2C). This is as expected given the tolerance of <u>weighted median</u>	
278	approaches to outliers.	 Deleted: median-based
 279		
280		 Formatted: Font:Bold
		Formatted: Normal
281	By applying new analytical techniques to an old question – the causal relationship between BMI	 <b>Deleted:</b> "To what extent does increased BMI predictors individuals to type 2 diabetes?"
282	and type 2 diabetes – we have explored the potential for invalid instruments to bias causal	predispose individuals to type 2 diabetes?
283	estimates in MR. In this case where BMI is the exposure, the opportunity to use a large instrument	
284	list in causal analyses presents both opportunity, through variance explained, but also cost,	
285	through complications generated by instrument properties or methods employed. Results here	
286	suggest that both TCF7L2 and FTO appear to have genetic variation which predicts BMI reliably,	
287	but for which associations with type 2 diabetes do not fully align with that for other variants (given	
288	BMI effects and assumed causality).	
289		
 290	For TCF7L2, only recently suggested to be associated with BMI directly (17), this is not surprising	
291	and reinforces the important point that the validity of a specific method's MR estimate, depends on	 Deleted: s
292	whether the genetic variants collectively satisfy its assumptions. In this case, it is possible that the	 <b>Deleted:</b> the validity of the genetic variants chosen as
293	negative association with BMI observed in GIANT is the product of a form of bias where the risk of	
294	type 2 diabetes is leading to effective treatment, health benefit and BMI reduction. This is	
295	supported by the apparently causal negative relationship between type 2 diabetes and BMI seen in	 Formatted: Not Highlight
296	a reciprocal analysis where BMI is the outcome of interest (Online Appendix (Supplementary	 Formatted: Not Highlight
297	Results, Figure S3)), though is likely to be more a comment on study design than biological effect.	
298		

310	In this example, the use of recently derived methods (9; 11) designed to overcome problems		
311	caused by directional pleiotropy, vields estimates which are more stable in the presence or		Deleted: appears to
312	absence of potentially invalid instruments and confirm the likely magnitude of the average effect of		
313	BMI on type 2 diabetes (i.e. from the most likely and stable estimate, an elevation of odds of		
314	disease of ~26% for each additional unit of BMI). The comparison of results from different methods		
315	for any set of potential instruments is important when assessing the reliability of causal inferences		
316	and important for downstream interpretation. In this case, whilst it is impossible to model precisely,		Deleted:
317	one can estimate the hypothetical impact of an average population level change in lifecourse BMI		
318	on type 2 diabetes. Given a population size of 64.1 million in the UK in mid 2013(30) and a		
319	modelled prevalence of type 2 diabetes (including non-diagnosed cases) of 7.4%(31; 32), the		
320	estimated reduction in odds for a 1kg/m2 reduction would potentially yield a reduction in the		
321	number of cases from ~4.7-3.6 million (a shift in prevalence to 5.6%),		Deleted: The comparison of results from different
322			worthwhile approach for assessing the reliability of causal inferences.
323	Acknowledgements		Deleted: ACKNOWLEDGMENTS
	A the sector is discussion of a sector is discussion with ODO 1 10		
324	Author contributions: NJ1 conceived and supervised the study, in discussion with GDS, LJC,		
324 325	KHW and RCR. JB and SB contributed to method and script development. LJC conducted the		
324 325 326	KHW and RCR. JB and SB contributed to method and script development. LJC conducted the analysis and wrote the manuscript. JB prepared the (Online Appendix (Supplementary methods)).		Deleted: technical appendix
324 325 326 327	Author contributions: NJT conceived and supervised the study, in discussion with GDS, LJC, KHW and RCR. JB and SB contributed to method and script development. LJC conducted the analysis and wrote the manuscript. JB prepared the <u>(Online Appendix (Supplementary methods))</u> , NJT is the guarantor of this work and, as such, had full access to all the data in the study and	[	Deleted: technical appendix Formatted: Font:(Default) Arial, 11 pt
324 325 326 327 328	Author contributions: NJT conceived and supervised the study, in discussion with GDS, LJC, KHW and RCR. JB and SB contributed to method and script development. LJC conducted the analysis and wrote the manuscript. JB prepared the <u>(Online Appendix (Supplementary methods))</u> , NJT is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.	(	Deleted: technical appendix Formatted: Font:(Default) Arial, 11 pt
324 325 326 327 328 329	Author contributions: NJT conceived and supervised the study, in discussion with GDS, LJC, KHW and RCR. JB and SB contributed to method and script development. LJC conducted the analysis and wrote the manuscript. JB prepared the <u>(Online Appendix (Supplementary methods))</u> , <u>NJT is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.</u>	(	Deleted: technical appendix Formatted: Font:(Default) Arial, 11 pt
324 325 326 327 328 329 330	Author contributions: NJT conceived and supervised the study, in discussion with GDS, LJC, KHW and RCR. JB and SB contributed to method and script development. LJC conducted the analysis and wrote the manuscript. JB prepared the <u>(Online Appendix (Supplementary methods))</u> , <u>NJT is the guarantor of this work and, as such, had full access to all the data in the study and</u> takes responsibility for the integrity of the data and the accuracy of the data analysis. <b>Funding:</b> NJT, LC, RCR, KHW, SB, GDS and JB work in the Medical Research Council	(	Deleted: technical appendix Formatted: Font:(Default) Arial, 11 pt
<ul> <li>324</li> <li>325</li> <li>326</li> <li>327</li> <li>328</li> <li>329</li> <li>330</li> <li>331</li> </ul>	Author contributions: NJT conceived and supervised the study, in discussion with GDS, LJC, KHW and RCR. JB and SB contributed to method and script development. LJC conducted the analysis and wrote the manuscript. JB prepared the <u>(Online Appendix (Supplementary methods))</u> , <u>NJT is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.</u> <b>Funding:</b> NJT, LC, RCR, KHW, SB, GDS and JB work in the Medical Research Council Integrative Epidemiology Unit (IEU) at the University of Bristol which is supported by the Medical	(	Deleted: technical appendix Formatted: Font:(Default) Arial, 11 pt
<ul> <li>324</li> <li>325</li> <li>326</li> <li>327</li> <li>328</li> <li>329</li> <li>330</li> <li>331</li> <li>332</li> </ul>	Author contributions: NJT conceived and supervised the study, in discussion with GDS, LJC, KHW and RCR. JB and SB contributed to method and script development. LJC conducted the analysis and wrote the manuscript. JB prepared the <u>(Online Appendix (Supplementary methods))</u> , NJT is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Funding: NJT, LC, RCR, KHW, SB, GDS and JB work in the Medical Research Council Integrative Epidemiology Unit (IEU) at the University of Bristol which is supported by the Medical Research Council (MC_UU_12013/1, MC_UU_12013/2, MC_UU_12013/3) and the University of	(	Deleted: technical appendix Formatted: Font:(Default) Arial, 11 pt
<ul> <li>324</li> <li>325</li> <li>326</li> <li>327</li> <li>328</li> <li>329</li> <li>330</li> <li>331</li> <li>332</li> <li>333</li> </ul>	Author contributions: NJT conceived and supervised the study, in discussion with GDS, LJC, KHW and RCR. JB and SB contributed to method and script development. LJC conducted the analysis and wrote the manuscript. JB prepared the <u>(Online Appendix (Supplementary methods))</u> , NJT is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. <b>Funding:</b> NJT, LC, RCR, KHW, SB, GDS and JB work in the Medical Research Council Integrative Epidemiology Unit (IEU) at the University of Bristol which is supported by the Medical Research Council (MC_UU_12013/1, MC_UU_12013/2, MC_UU_12013/3) and the University of Bristol. SB is supported by the Wellcome Trust (grant number 100114). JB is funded by a Medical	(	Deleted: technical appendix Formatted: Font:(Default) Arial, 11 pt
<ul> <li>324</li> <li>325</li> <li>326</li> <li>327</li> <li>328</li> <li>329</li> <li>330</li> <li>331</li> <li>332</li> <li>333</li> <li>334</li> </ul>	Author contributions: NJT conceived and supervised the study, in discussion with GDS, LJC, KHW and RCR. JB and SB contributed to method and script development. LJC conducted the analysis and wrote the manuscript. JB prepared the <u>(Online Appendix (Supplementary methods))</u> , NJT is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. <b>Funding:</b> NJT, LC, RCR, KHW, SB, GDS and JB work in the Medical Research Council Integrative Epidemiology Unit (IEU) at the University of Bristol which is supported by the Medical Research Council (MC_UU_12013/1, MC_UU_12013/2, MC_UU_12013/3) and the University of Bristol. SB is supported by the Wellcome Trust (grant number 100114). JB is funded by a Medical Research Council Methodology Research Fellowship (grant number MR/N501906/1). RCR and	_(	Deleted: technical appendix Formatted: Font:(Default) Arial, 11 pt
324 325 326 327 328 329 330 331 332 333 334 335	Author contributions: NJT conceived and supervised the study, in discussion with GDS, LJC, KHW and RCR. JB and SB contributed to method and script development. LJC conducted the analysis and wrote the manuscript. JB prepared the <u>(Online Appendix (Supplementary methods))</u> , NJT is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. <b>Funding:</b> NJT, LC, RCR, KHW, SB, GDS and JB work in the Medical Research Council Integrative Epidemiology Unit (IEU) at the University of Bristol which is supported by the Medical Research Council (MC_UU_12013/1, MC_UU_12013/2, MC_UU_12013/3) and the University of Bristol. SB is supported by the Wellcome Trust (grant number 100114). JB is funded by a Medical Research Council Methodology Research Fellowship (grant number MR/N501906/1). RCR and KHW are supported by CRUK (C18281/A19169).	(	Deleted: technical appendix Formatted: Font:(Default) Arial, 11 pt

# Page 37 of 45

#### Diabetes

345	GIANT: Data on BMI have been contributed by GIANT (Genetic Investigation of ANthropometric
346	Traits). With special thanks to Adam Locke for conducting the additional analyses required for this
347	work.
348	
349	DIAGRAM: Data on type 2 diabetes have been contributed by the DIAGRAM (DIAbetes Genetics
350	Replication And Meta-analysis) consortium and were downloaded from here.
351	
352	There are <b>no</b> conflicts to declare.

33

References	Deleted:Section Break (Next Page)		
4. Deserve stiller Ox De du mane index and serve and site mantality in 000,000 adultar	Formatted: Font:Bold		
collaborative analyses of 57 prospective studies. The Lancet 373:1083-1096	Formatted: Font:(Default) Arial, 11 pt		
2. Holmes Michael V, Lange Leslie A, Palmer T, Lanktree Matthew B, North Kari E, Almoguera B,			
Buxbaum S, Chandrupatla Hareesh R, Elbers Clara C, Guo Y, Hoogeveen Ron C, Li J, Li Yun R,			
Swerdlow Daniel I, Cushman M, Price Tom S, Curtis Sean P, Fornage M, Hakonarson H, Patel			
Sanjay R, Redline S, Siscovick David S, Tsai Michael Y, Wilson James G, van der Schouw			
Yvonne T, FitzGerald Garret A, Hingorani AD, Casas Juan P, de Bakker Paul IW, Rich Stephen S,			
Schadt Eric E, Asselbergs Folkert W, Reiner Alex P, Keating Brendan J: Causal Effects of Body			
Mass Index on Cardiometabolic Traits and Events: A Mendelian Randomization Analysis. The			
American Journal of Human Genetics 2014;94:198-208			
3. Davey Smith G, Hemani G: Mendellan randomization: genetic anchors for causal inference in pridemic large turbles. Human Maleuro Cauties 2014 902 D00			
epidemiological studies. Human Molecular Genetics 2014;23:R89-R98			
4 Devey Smith C. Ebrohim S. (Mandalian randomization), can gapatia anidemialagy contribute to			
4. Davey Smith G, Ebrahim S. Mendelian randomization : can genetic epidemiology contribute to			
2003,32.1-22			
5 Lawlor DA Harbord RM Sterne JAC Timpson N Davey Smith G Mendelian randomization			
using genes as instruments for making causal inferences in epidemiology. Stat Med 2008:27:1133-	-		
1163			
6. Sleiman PMA, Grant SFA: Mendelian Randomization in the Era of Genomewide Association			
Studies. Clinical Chemistry 2010;56:723-728			
7. Boef AGC, Dekkers OM, le Cessie S: Mendelian randomization studies: a review of the			
approaches used and the quality of reporting. International Journal of Epidemiology 2015;			
8. Nordestgaard BG, Palmer TM, Benn M, Zacho J, Tybjærg-Hansen A, Davey Smith G, Timpson			
NJ: The Effect of Elevated Body Mass Index on Ischemic Heart Disease Risk: Causal Estimates			
from a Mendelian Randomisation Approach. PLoS Med 2012;9:e1001212			
9. Bowden J, Davey Smith G, Burgess S: Mendellan randomization with invalid instruments: effect			
estimation and bias detection through Egger regression. International Journal of Epidemiology			
2015,44.512-525			
10 Didelez V. Sheehan N: Mendelian randomization as an instrumental variable approach to			
causal inference. Statistical Methods in Medical Research 2007;16:309-330			
11. Bowden J, Davey Smith G, Haycock PC, Burgess S: Consistent estimation in Mendelian			
randomization with some invalid instruments using a weighted median estimator. Genetic			
Epidemiology 2016;40:304-314			
12. Yaghootkar H, Frayling T: Recent progress in the use of genetics to understand links between			
type 2 diabetes and related metabolic traits. Genome Biol 2013;14:1-7			
13. Dimas AS, Lagou V, Barker A, Knowles JW, Mägi R, Hivert M-F, Benazzo A, Rybin D, Jackson			
AU, Stringham HM, Song C, Fischer-Rosinsky A, Boesgaard TW, Grarup N, Abbasi FA, Assimes			
IL, Hao K, Yang X, Lecoeur C, Barroso I, Bonnycastle LL, Bottcher Y, Bumpstead S, Chines PS,			
Eluus wirk, Glaessier J, Kovacs P, Morkell WA, Narisu N, Payne F, Stancakova A, Switt AJ, Tonjes A, Bornetoin SP, Cauchi S, Eroquel P, Morkell D, Schwarz DEL, Höring H, L. Smith H, Dashala M	5		
A, DUITISTEILI ON, CAUGHI O, FIOYUELF, MEYLE D, OCHWALZ FER, Rahning R-D, Stillett U, BOEITIKE M, Bergman RN, Collins FS, Mohlke KI, Tuomilahto L, Ouertemous T, Lind L, Hansen T, Bodorson O,			
Walker M. Pfeiffer AFH. Spranger J. Stumvoll M. Meigs JB. Wareham NJ. Kuusisto J. Laakso M.	,		
Langenhers C. Dunuis I. Wetensbe D. Elensor I.C. Ingelegence McCarthy ML Drekensor I.			

-15

#### **Diabetes**

10	Investigators obotM: Impact of Type 2 Diabetes Susceptibility Variants on Quantitative Glycemic
-11	Traits Reveals Mechanistic Heterogeneity. Diabetes 2014;63:2158-2171

14. Burgess S, Butterworth A, Thompson SG: Mendelian Randomization Analysis With Multiple
 Genetic Variants Using Summarized Data. Genetic Epidemiology 2013;37:658-665

15. Burgess S, Scott R, Timpson N, Davey Smith G, Thompson S: Using published data in
 Mendelian randomization: a blueprint for efficient identification of causal risk factors. Eur J
 Epidemiol 2015:1-10

Herce BL, Burgess S: Efficient Design for Mendelian Randomization Studies: Subsample and
 2-Sample Instrumental Variable Estimators. American Journal of Epidemiology 2013;178:1177 1184

422 423 424 425 17. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, Powell C, Vedantam S, Buchkovich ML, Yang J, Croteau-Chonka DC, Esko T, Fall T, Ferreira T, Gustafsson S, Kutalik Z, 26 Luan Ja, Magi R, Randall JC, Winkler TW, Wood AR, Workalemahu T, Faul JD, Smith JA, Hua 27 Zhao J, Zhao W, Chen J, Fehrmann R, Hedman AK, Karjalainen J, Schmidt EM, Absher D, Amin 128 N, Anderson D, Beekman M, Bolton JL, Bragg-Gresham JL, Buyske S, Demirkan A, Deng G, Ehret 120 129 130 GB, Feenstra B, Feitosa MF, Fischer K, Goel A, Gong J, Jackson AU, Kanoni S, Kleber ME, Kristiansson K, Lim U, Lotay V, Mangino M, Mateo Leach I, Medina-Gomez C, Medland SE, Nalls 131 MA, Palmer CD, Pasko D, Pechlivanis S, Peters MJ, Prokopenko I, Shungin D, Stancakova A, 132 Strawbridge RJ, Ju Sung Y, Tanaka T, Teumer A, Trompet S, van der Laan SW, van Setten J, Van Vliet-Ostaptchouk JV, Wang Z, Yengo L, Zhang W, Isaacs A, Albrecht E, Arnlov J, Arscott GM, 133 134 Attwood AP, Bandinelli S, Barrett A, Bas IN, Bellis C, Bennett AJ, Berne C, Blagieva R, Bluher M, 135 Bohringer S, Bonnycastle LL, Bottcher Y, Boyd HA, Bruinenberg M, Caspersen IH, Ida Chen Y-D, 136 Clarke R, Warwick Daw E, de Craen AJM, Delgado G, Dimitriou M, Doney ASF, Eklund N, Estrada -37 K, Eury E, Folkersen L, Fraser RM, Garcia ME, Geller F, Giedraitis V, Gigante B, Go AS, Golay A, Goodall AH, Gordon SD, Gorski M, Grabe H-J, Grallert H, Grammer TB, Graszler J, Gronberg H, 138 Groves CJ, Gusto G, Haessler J, Hall P, Haller T, Hallmans G, Hartman CA, Hassinen M, Hayward C, Heard-Costa NL, Helmer Q, Hengstenberg C, Holmen O, Hottenga J-J, James AL, Jeff JM, 139 40 Johansson A, Jolley J, Juliusdottir T, Kinnunen L, Koenig W, Koskenvuo M, Kratzer W, Laitinen J, 41 42 Lamina C, Leander K, Lee NR, Lichtner P, Lind L, Lindstrom J, Sin Lo K, Lobbens S, Lorbeer R, 43 Lu Y, Mach F, Magnusson PKE, Mahajan A, McArdle WL, McLachlan S, Menni C, Merger S, 44 Mihailov E, Milani L, Moayyeri A, Monda KL, Morken MA, Mulas A, Muller G, Muller-Nurasyid M, 45 Musk AW, Nagaraja R, Nothen MM, Nolte IM, Pilz S, Rayner NW, Renstrom F, Rettig R, Ried JS, Ripke S, Robertson NR, Rose LM, Sanna S, Scharnagl H, Scholtens S, Schumacher FR, Scott 46 47 WR, Seufferlein T, Shi J, Vernon Smith A, Smolonska J, Stanton AV, Steinthorsdottir V, Stirrups K, 448 449 Stringham HM, Sundstrom J, Swertz MA, Swift AJ, Syvanen A-C, Tan S-T, Tayo BO, Thorand B, Thorleifsson G, Tyrer JP, Uh H-W, Vandenput L, Verhulst FC, Vermeulen SH, Verweij N, Vonk JM, 150 Waite LL, Warren HR, Waterworth D, Weedon MN, Wilkens LR, Willenborg C, Wilsgaard T, Wojczynski MK, Wong A, Wright AF, Zhang Q, The LifeLines Cohort S, Brennan EP, Choi M, 51 452 453 454 Dastani Z, Drong AW, Eriksson P, Franco-Cereceda A, Gadin JR, Gharavi AG, Goddard ME, Handsaker RE, Huang J, Karpe F, Kathiresan S, Keildson S, Kiryluk K, Kubo M, Lee J-Y, Liang L, Lifton RP, Ma B, McCarroll SA, McKnight AJ, Min JL, Moffatt MF, Montgomery GW, Murabito JM, ł55 Nicholson G, Nyholt DR, Okada Y, Perry JRB, Dorajoo R, Reinmaa E, Salem RM, Sandholm N, 156 Scott RA, Stolk L, Takahashi A, Tanaka T, van/'t Hooft FM, Vinkhuyzen AAE, Westra H-J, Zheng W, Zondervan KT, The AC, The A-BMIWG, The CDC, The CC, The G, The I, The MI, The Mu TC, ł57 158 The MC, The PC, The ReproGen C, The GC, The International Endogene C, Heath AC, Arveiler D, 159 Bakker SJL, Beilby J, Bergman RN, Blangero J, Bovet P, Campbell H, Caulfield MJ, Cesana G, 60 Chakravarti A, Chasman DI, Chines PS, Collins FS, Crawford DC, Adrienne Cupples L, Cusi D, Danesh J, de Faire U, den Ruijter HM, Dominiczak AF, Erbel R, Erdmann J, Eriksson JG, Farrall 61 62 M, Felix SB, Ferrannini E, Ferrieres J, Ford I, Forouhi NG, Forrester T, Franco OH, Gansevoort 463 RT, Gejman PV, Gieger C, Gottesman O, Gudnason V, Gyllensten U, Hall AS, Harris TB, Hattersley AT, Hicks AA, Hindorff LA, Hingorani AD, Hofman A, Homuth G, Kees Hovingh G 64 Humphries SE, Hunt SC, Hypponen E, Illig T, Jacobs KB, Jarvelin M-R, Jockel K-H, Johansen B, 65 Jousilahti P, Wouter Jukema J, Jula AM, Kaprio J, Kastelein JJP, Keinanen-Kiukaanniemi SM, 66

67 Kiemeney LA, Knekt P, Kooner JS, Kooperberg C, Kovacs P, Kraja AT, Kumari M, Kuusisto J, Lakka TA, Langenberg C, Le Marchand L, Lehtimaki T, Lyssenko V, Mannisto S, Marette A, Matise 68 69 TC, McKenzie CA, McKnight B, Moll FL, Morris AD, Morris AP, Murray JC, Nelis M, Ohlsson C, Oldehinkel AJ, Ong KK, Madden PAF, Pasterkamp G, Peden JF, Peters A, Postma DS, Pramstaller PP, Price JF, Qi L, Raitakari OT, Rankinen T, Rao DC, Rice TK, Ridker PM, Rioux JD, 71 Ritchie MD, Rudan I, Salomaa V, Samani NJ, Saramies J, Sarzynski MA, Schunkert H, Schwarz PEH, Sever P, Shuldiner AR, Sinisalo J, Stolk RP, Strauch K, Tonjes A, Tregouet D-A, Tremblay A, Tremoli E, Virtamo J, Vohl M-C, Volker U, Waeber G, Willemsen G, Witteman JC, Carola Zillikens M, Adair LS, Amouyel P, Asselbergs FW, Assimes TL, Bochud M, Boehm BO, Boerwinkle ł76 E, Bornstein SR, Bottinger EP, Bouchard C, Cauchi S, Chambers JC, Chanock SJ, Cooper RS, de 477 478 Bakker PIW, Dedoussis G, Ferrucci L, Franks PW, Froguel P, Groop LC, Haiman CA, Hamsten A, Hui J, Hunter DJ, Hveem K, Kaplan RC, Kivimaki M, Kuh D, Laakso M, Liu Y, Martin NG, Marz W, ł79 Melbye M, Metspalu A, Moebus S, Munroe PB, Njolstad I, Oostra BA, Palmer CNA, Pedersen NL, 180 Perola M, Perusse L, Peters U, Power C, Quertermous T, Rauramaa R, Rivadeneira F, Saaristo TE, Saleheen D, Sattar N, Schadt EE, Schlessinger D, Eline Slagboom P, Snieder H, Spector TD, 81 182 Thorsteinsdottir U, Stumvoll M, Tuomilehto J, Uitterlinden AG, Uusitupa M, van der Harst P, 483 Walker M, Wallaschofski H, Wareham NJ, Watkins H, Weir DR, Wichmann HE, Wilson JF, Zanen 84 P, Borecki IB, Deloukas P, Fox CS, Heid IM, O//Connell JR, Strachan DP, Stefansson K, van Duijn CM, Abecasis GR, Franke L, Frayling TM, McCarthy MI, Visscher PM, Scherag A, Willer CJ, 185 186 Boehnke M, Mohlke KL, Lindgren CM, Beckmann JS, Barroso I, North KE, Ingelsson E, Hirschhorn 87 JN, Loos RJF, Speliotes EK: Genetic studies of body mass index yield new insights for obesity 188 biology. Nature 2015;518:197-206 189 190 18. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segrè AV, Steinthorsdottir V, Strawbridge RJ,

91 Khan H, Grallert H, Mahajan A, Prokopenko I, Kang HM, Dina C, Esko T, Fraser RM, Kanoni S, 92 Kumar A, Lagou V, Langenberg C, Luan Ja, Lindgren CM, Müller-Nurasyid M, Pechlivanis S, 93 Rayner NW, Scott LJ, Wiltshire S, Yengo L, Kinnunen L, Rossin EJ, Raychaudhuri S, Johnson AD, 94 Dimas AS, Loos RJF, Vedantam S, Chen H, Florez JC, Fox C, Liu C-T, Rybin D, Couper DJ, Kao WHL, Li M, Cornelis MC, Kraft P, Sun Q, van Dam RM, Stringham HM, Chines PS, Fischer K, 95 Fontanillas P, Holmen OL, Hunt SE, Jackson AU, Kong A, Lawrence R, Meyer J, Perry JRB, 96 97 Platou CGP, Potter S, Rehnberg E, Robertson N, Sivapalaratnam S, Stančáková A, Stirrups K, 198 Thorleifsson G, Tikkanen E, Wood AR, Almgren P, Atalay M, Benediktsson R, Bonnycastle LL, 99 Burtt N, Carey J, Charpentier G, Crenshaw AT, Doney ASF, Dorkhan M, Edkins S, Emilsson V, Eury E, Forsen T, Gertow K, Gigante B, Grant GB, Groves CJ, Guiducci C, Herder C, Hreidarsson 500 501 AB, Hui J, James A, Jonsson A, Rathmann W, Klopp N, Kravic J, Krjutškov K, Langford C, 502 Leander K, Lindholm E, Lobbens S, Männistö S, Mirza G, Mühleisen TW, Musk B, Parkin M, 503 Rallidis L, Saramies J, Sennblad B, Shah S, Sigurðsson G, Silveira A, Steinbach G, Thorand B, 504 Trakalo J, Veglia F, Wennauer R, Winckler W, Zabaneh D, Campbell H, van Duijn C, Uitterlinden 505 AG, Hofman A, Sijbrands E, Abecasis GR, Owen KR, Zeggini E, Trip MD, Forouhi NG, Syvänen A-506 C, Eriksson JG, Peltonen L, Nöthen MM, Balkau B, Palmer CNA, Lyssenko V, Tuomi T, Isomaa B, 507 Hunter DJ, Qi L, Wellcome Trust Case Control C, Investigators M, Consortium G, Consortium A-TD, Consortium SD, Shuldiner AR, Roden M, Barroso I, Wilsgaard T, Beilby J, Hovingh K, Price 508 509 JF, Wilson JF, Rauramaa R, Lakka TA, Lind L, Dedoussis G, Njølstad I, Pedersen NL, Khaw K-T, Wareham NJ, Keinanen-Kiukaanniemi SM, Saaristo TE, Korpi-Hyövälti E, Saltevo J, Laakso M, 510 Kuusisto J, Metspalu A, Collins FS, Mohlke KL, Bergman RN, Tuomilehto J, Boehm BO, Gieger C, 511 Hveem K, Cauchi S, Froguel P, Baldassarre D, Tremoli E, Humphries SE, Saleheen D, Danesh J, 512 Ingelsson E, Ripatti S, Salomaa V, Erbel R, Jöckel K-H, Moebus S, Peters A, Illig T, de Faire U, 513 Hamsten A, Morris AD, Donnelly PJ, Frayling TM, Hattersley AT, Boerwinkle E, Melander O, 514 Kathiresan S, Nilsson PM, Deloukas P, Thorsteinsdottir U, Groop LC, Stefansson K, Hu F, Pankow 515 JS, Dupuis J, Meigs JB, Altshuler D, Boehnke M, McCarthy MI: Large-scale association analysis 516 provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nature 517 genetics 2012;44:981-990 518

19. Team RC: R: A language and environment for statistical computing., 3.2.0 ed. Vienna, Austria,
 R Foundation for Statistical Computing, 2014

19

423 20. Higgins JPT, Thompson SG, Deeks JJ, Altman DG: Measuring inconsistency in metaanalyses. BMJ : British Medical Journal 2003;327:557-560
525

21. Bowden J: Assessing the suitability of summary data for Mendelian randomization analyses
using MR-Egger regression: the role of the I<sup>^</sup>2 statistic. Technical Report (available on request),
University of Bristol 2015

\$30 22. Stone M: Cross-Validatory Choice and Assessment of Statistical Predictions. Journal of the
 \$31 Royal Statistical Society Series B (Methodological) 1974;36:111-147

4. Solar Provide Galactical Control of Con

\$34
\$35 24. Fox J, Weisberg S: An R Companion to Applied Regression. Los Angeles, Sage, 2011

36
37 25. Cook RD: Detection of Influential Observation in Linear Regression. Technometrics
38 1977;19:15-18

539 540 26. Manning AK, Hivert M-F, Scott RA, Grimsby JL, Bouatia-Naji N, Chen H, Rybin D, Liu C-T, 541 Bielak LF, Prokopenko I, Amin N, Barnes D, Cadby G, Hottenga J-J, Ingelsson E, Jackson AU, 542 543 Johnson T, Kanoni S, Ladenvall C, Lagou V, Lahti J, Lecoeur C, Liu Y, Martinez-Larrad MT, Montasser ME, Navarro P, Perry JRB, Rasmussen-Torvik LJ, Salo P, Sattar N, Shungin D, 644 Strawbridge RJ, Tanaka T, van Duijn CM, An P, de Andrade M, Andrews JS, Aspelund T, Atalay 545 M, Aulchenko Y, Balkau B, Bandinelli S, Beckmann JS, Beilby JP, Bellis C, Bergman RN, Blangero 546 J, Boban M, Boehnke M, Boerwinkle E, Bonnycastle LL, Boomsma DI, Borecki IB, Bottcher Y, Bouchard C, Brunner E, Budimir D, Campbell H, Carlson O, Chines PS, Clarke R, Collins FS, Corbaton-Anchuelo A, Couper D, de Faire U, Dedoussis GV, Deloukas P, Dimitriou M, Egan JM, 647 548 649 Eiriksdottir G, Erdos MR, Eriksson JG, Eury E, Ferrucci L, Ford I, Forouhi NG, Fox CS, Franzosi 550 MG, Franks PW, Frayling TM, Froguel P, Galan P, de Geus E, Gigante B, Glazer NL, Goel A, 551 Groop L, Gudnason V, Hallmans G, Hamsten A, Hansson O, Harris TB, Hayward C, Heath S, 552 Hercberg S, Hicks AA, Hingorani A, Hofman A, Hui J, Hung J, Jarvelin M-R, Jhun MA, Johnson 553 PCD, Jukema JW, Jula A, Kao WH, Kaprio J, Kardia SLR, Keinanen-Kiukaanniemi S, Kivimaki M, 554 Kolcic I, Kovacs P, Kumari M, Kuusisto J, Kyvik KO, Laakso M, Lakka T, Lannfelt L, Lathrop GM, 555 Launer LJ, Leander K, Li G, Lind L, Lindstrom J, Lobbens S, Loos RJF, Luan Ja, Lyssenko V, Magi R, Magnusson PKE, Marmot M, Meneton P, Mohlke KL, Mooser V, Morken MA, Miljkovic I, Narisu 556 557 N, O'Connell J, Ong KK, Oostra BA, Palmer LJ, Palotie A, Pankow JS, Peden JF, Pedersen NL, 558 Pehlic M, Peltonen L, Penninx B, Pericic M, Perola M, Perusse L, Peyser PA, Polasek O, Pramstaller PP, Province MA, Raikkonen K, Rauramaa R, Rehnberg E, Rice K, Rotter JI, Rudan I, 559 560 Ruokonen A, Saaristo T, Sabater-Lleal M, Salomaa V, Savage DB, Saxena R, Schwarz P, Seedorf U, Sennblad B, Serrano-Rios M, Shuldiner AR, Sijbrands EJG, Siscovick DS, Smit JH, Small KS, 561 562 Smith NL, Smith AV, Stancakova A, Stirrups K, Stumvoll M, Sun YV, Swift AJ, Tonjes A, Tuomilehto J, Trompet S, Uitterlinden AG, Uusitupa M, Vikstrom M, Vitart V, Vohl M-C, Voight BF, 63 Vollenweider P, Waeber G, Waterworth DM, Watkins H, Wheeler E, Widen E, Wild SH, Willems 564 565 SM, Willemsen G, Wilson JF, Witteman JCM, Wright AF, Yaghootkar H, Zelenika D, Zemunik T, Zgaga L, Wareham NJ, McCarthy MI, Barroso I, Watanabe RM, Florez JC, Dupuis J, Meigs JB, 566 \$67 Langenberg C: A genome-wide approach accounting for body mass index identifies genetic 568 variants influencing fasting glycemic traits and insulin resistance. Nat Genet 2012;44:659-669 569

27. Richmond RC, Davey Smith G, Ness AR, den Hoed M, McMahon G, Timpson NJ: Assessing
 Causality in the Association between Child Adiposity and Physical Activity Levels: A Mendelian
 Randomization Analysis. PLoS Med 2014;11:e1001618

28. Taylor PN, Richmond R, Davies N, Sayers A, Stevenson K, Woltersdorf W, Taylor A, Groom A,
28. Taylor PN, Richmond R, Davies N, Sayers A, Stevenson K, Woltersdorf W, Taylor A, Groom A,
Northstone K, Ring S, Okosieme O, Rees A, Nitsch D, Williams GR, Smith GD, Gregory JW,
Timpson NJ, Tobias JH, Dayan CM: Paradoxical Relationship Between Body Mass Index and
Thyroid Hormone Levels: A Study Using Mendelian Randomization. J Clin Endocrinol Metab
2016;101:730-738

580 581	29. Walter S, Kubzansky LD, Koenen KC, Liang L, Tchetgen Tchetgen EJ, Cornelis MC, Chang S- C, Rimm E, Kawachi I, Glymour MM: Revisiting mendelian randomization studies of the effect of		
582	body mass index on depression. American Journal of Medical Genetics Part B: Neuropsychiatric		
583	Genetics 2015;168:108-115		
584			
585	30. Office of National Statistics(UK): Annual Mid-year Population Estimates, 2013. Statistical		Deleted: Statistics OfN
586	bulletin, 2014		Deleted: In
587			
588	31. The National Cardiovascular Intelligence Network; Cardiovascular disease key facts. In		Deleted: Network TNCI
589	Diabetes: Public Health England, 2013		
590			
591	32. Adult obesity and type 2 diabetes. Public Health England, 2014		
592		<b>.</b>	Formatted: Left

#### **5**96 Tables

597 Table 1 – Estimates from the application of inverse-variance weighted, MR-Egger and weighted

598 median, Mendelian randomisation methodologies. Estimates represent the estimated causal effect Deleted: median-based

599 of body mass index on type 2 diabetes.

Method	Estimate	95% CI	p-value	
Complete variant set (n=	96 SNPs)			Formatted: Left
IVW	1.20	1.09, 1.30	8.00 x 10 <sup>-05</sup>	Formatted: Left
MR-Egger	1.39	1.14, 1.68	1.53 x 10 <sup>-03</sup>	Formatted: Left
MR-Egger <sup>(_)</sup>	-0.019	-0.041, 0.004	0.10	Formatted: Left
Weighted median,	1.26	1.17, 1.34	5.26 x 10 <sup>-9</sup>	Formatted: Left
TCF7L2(rs7903146) removed from the variant set (n=95 SNPs)				Deleted: Median-based Formatted: Left
IVW	1.22	1.16, 1.28	1.49 x 10 <sup>-11</sup>	Formatted: Left
MR-Egger	1.34	1.17, 1.51	9.71 x 10 <sup>-06</sup>	Formatted: Left
MR-Egger <sup>(_)</sup>	-0.011	-0.024, -0.024	0.13	Formatted: Left
Weighted median,	1.26	1.19, 1.32	3.29 x 10 <sup>-10</sup>	Formatted: Left
FTO(rs1558902) remove	ed from the variant set	(n=95 SNPs)		Deleted: Median-based Formatted: Left
IVW	1.16	1.06, 1.27	1.31 x 10 <sup>-03</sup>	Formatted: Left
MR-Egger	1.30	1.01, 1.65	0.04	Formatted: Left
MR-Egger <sup>(_)</sup>	-0.012	-0.038, 0.014	0.34	Formatted: Left
Weighted median,	1.21	1.13, 1.28	6.81 x 10 <sup>-08</sup>	Formatted: Left
ntercept coefficients MR-	Egger <sup>()</sup> represent the a	average pleiotropic effec	t of a genetic variant	On Formatted: Left, Line spacing: single
ype 2 diabetes risk <u>. "IVW</u> aucleotide polymorphism	" reters to inverse varia	ance weighted estimates	s, SNP refers to single	

608	Figures	For	matted: Left
609	Figure 1 – Genetic associations with body mass index (BMI) and type 2 diabetes from 96 variants measured in GIANT (17) and DIAGRAM (18),		
610	respectively. TCF7L2(rs7903146) and FTO(rs1558902) are marked with a 'X' and labelled.		
611	A - funnel plot of minor allele frequency corrected genetic associations with BMI (interpreted as instrument strength) against causal estimates based		
612	on each genetic variant individually, where the causal effect is expressed in logs odds ratio of type 2 diabetes for each unit increase in BMI. The		
613	overall causal estimates (β coefficients) of BMI on type 2 diabetes estimated by inverse-variance weighted (solid black line), MR-Egger (dashed black		
614	line) and weighted median (dotted black line) methods are also shown. Grey solid line represent x=0, that is a causal estimate of zero.	Dele	eted: median-based
615	B - scatter plot of genetic associations with type 2 diabetes against associations with BMI, with causal estimates (β coefficients) of BMI on type 2		
616	diabetes estimated by inverse-variance weighted (solid line), MR-Egger (dashed line) and weighted median (dotted line) methods.	Dele	eted: median-based
617			
618	Figure 2 – Distributions of regression estimates resulting from leave-one-out permutation analysis. Solid line = estimate from main analysis (n=96		
619	variants); dashed line = estimate with TCF7L2(rs7903146) removed; dotted line = estimate with FTO(rs1558902) removed.		
620	A - Causal estimates ( $\beta$ coefficients) of BMI on type 2 diabetes estimated by an inverse-variance weighted method		
621	B - Causal estimates ( $\beta$ coefficients) of BMI on type 2 diabetes estimated by MR-Egger		
622	C - Causal estimates (β coefficients) of BMI on type 2 diabetes estimated by a weighted median method	Dele	eted: median-based
623	D – Estimates of the intercept by MR-Egger		
624			
1			
	17		

Page 4: [1] Deleted

NJ Timpson

6/6/16 11:33:00 AM

A genetic instrument for BMI was constructed using