Obesity is a risk factor for acute promyelocytic leukemia: evidence from population and cross-sectional studies and correlation with FLT3 mutations and polyunsaturated fatty acid metabolism

Luca Mazzarella,[1,2] Edoardo Botteri,[3] Anthony Matthews,[4] Elena Gatti,[1] Davide Di Salvatore,[5] Vincenzo Bagnardi,[6] Massimo Breccia,[7] Pau Montesinos,[8] Teresa Bernal,[9] Cristina Gil,[10] Timothy J Ley,[11] Miguel Sanz,[8] Krishnan Bhaskaran,[4] Francesco Lo Coco[12†] and Pier Giuseppe Pelicci[1,13]

[1]Department of Experimental Oncology, European Institute of Oncology IRCCS, Milan, Italy; [2]Division of Early Drug Development for Innovative Therapies, European Institute of Oncology IRCCS, Milan, Italy; [3]Cancer Registry of Norway, Oslo, Norway; [4]London School of Hygiene and Tropical Medicine, London, UK; [5]IFOM-FIRC Institute of Molecular Oncology, Milan, Italy; [6]University of Milan-Bicocca, Milan, Italy; [7]Division of Hematology, Department of Cellular Biotechnologies and Hematology, Sapienza University, Rome, Italy; [8]Hematology Department, Hospital Universitario y Politécnico La Fe, Valencia, Spain; [9]Hospital Central de Asturias, Oviedo, Spain; [10]Hospital General, Alicante, Spain; [11]Department of Medicine, Division of Oncology, Washington University in St. Louis, Saint Louis, MO, USA; [12]Department of Hematology, University Tor Vergata, Roma, Italy and [13]Department of Oncology and Hemato-Oncology, University of Milan

 doi:10.3324/haematol.2019.223925

Received: Apri 5, 2019.

Accepted: September 11, 2019.

Pre-published: September 12, 2019.

Correspondence: *LUCA MAZZARELLA* - luca.mazzarella@ieo.it

PIER GIUSEPPE PELICCI - piergiuseppe.pelicci@ieo.it

Supplementary methods

UK population-based study: data collection and statistical methods

Methods for the UK population study were described in depth previously [1]. The study was approved by the London School of Hygiene and Tropical Medicine Ethics Committee. Briefly, data were collected from the UK Clinical Practice Research Datalink (CPRD), which contains computerised primary-care records from general practitioners who use the Vision IT system and have agreed at the practice level to participate (covering about 9% of the UK population). The CPRD dataset is representative of the UK population in terms of age, sex, ethnicity and BMI when compared with census data [2,3] Study entry began 12 months after registration and we assigned BMI records as exposure only 12 months after their recording, to guard against reverse causality (ie, BMI being affected by undiagnosed cancer). We included all people aged 16 years or older with BMI and subsequent eligible follow-up time. BMI was recorded as per local general practice. Individuals with any record of cancer before study entry were excluded. BMI records and diagnosis collected between years 1987-2012 were included in the analysis. To identify outcomes of specific leukaemia sub-types, CPRD clinical records were searched for codes relating to: AML (ICD-10 codes: C92.0, C92.5, C92.6, C93.0, C94.0, C95.0); APL (ICD-10 code C92.4); LL (ICD-10 code C91); and any other leukaemias that were not specifically coded ("other").

Subjects were followed-up from study entry until the earliest of: first cancer diagnoses (any site), death, transfer out of CPRD, or last data collection of the practice. To relate BMI to risk of each type of leukaemia, we fitted Cox regression models with attained age as the underlying

timescale. We fitted fully adjusted models, with BMI as a continuous linear term to estimate the average effect of a 5 kg/m² increase in BMI on leukaemia risk. We also fitted a model including BMI as a 3-knot spline in case of non-linearity in the relationship with leukaemia risk; we tested for evidence of non-linearity by conducting a likelihood ratio test comparing nested models with and without the non-linear terms in the spline basis. We controlled for the following covariates at time of the BMI record(s): age (three-knot restricted cubic spline to allow for non-linearity); smoking status (never smoker, current smoker, ex-smoker); alcohol use (non-drinker, current drinker [light, moderate, heavy, unknown], ex-drinker); previous diabetes diagnosis; index of multiple deprivation (in quintiles, a measure of socioeconomic status); calendar period (<1989, 1990–94, 1995–99, 2000–04, 2005–09, ≥2010); and stratified by sex. We excluded people with missing smoking (49 206/5.24 million [0.9%]) and alcohol status (394 196/5.24 million [7.5%]. All CIs are presented at the 95% level.

Cross-sectional studies: data collection and statistical methods

APL cases from Spain were extracted from the PETHEMA database to include 414 cases diagnosed between 1998 and 2012. APL cases from Italy were 134 adult patients treated with AIDA protocol included in the previously described cohort [4]. APL cases from USA included the entire cohort of the published AML TCGA project [5] (n=20) plus 22 additional APL cases, unselected for any clinical variable, diagnosed at Washington University (Expanded TCGA cohort). For all case cohorts, BMI was measured at the time of diagnosis. Data collection was approved by the Research Ethics Board of each participating institution, as referenced [5–8]

We compared the distribution of BMI observed in the three APL case cohorts to the distribution of BMI expected in the general population of

the same countries. Specifically, to calculate the expected distribution of BMI in Italy we used data from the Italian National Institute of Statistics [9] and we selected the area of Lazio, where the APL cases were diagnosed, in the years 2000-2010. For Spain, we used data from the Eurostat [10] and we selected the general population of Spain in the year 2008, the only year available. For both Italy and Spain, the expected BMI distribution was calculated using the available age- and sex- specific BMI distribution of the general population classified in 3 categories (<25; 25-29.9; ≥30). For USA we used the 2009-2010 data from the American National Health and Nutrition Examination Survey [11]. The expected BMI distribution was calculated using the available race-, age- and sex- specific BMI distribution of the general population classified in 4 categories (<25; 25-29.9; 30.0-34.9; ≥35).

The global null hypothesis that the observed counts did not differ from the expected ones across the BMI categories was tested in a null Poisson regression model, where the observed counts were considered as dependent variable and the expected counts as the offset. We included in the model BMI as an ordinal variable to test the log-linear relationship between BMI and the observed to expected ratio (i.e. to test for linear trend). The Pearson's chi-square goodness of fit test p-value was reported.

Expression data analysis

Expression data (RPKM matrix) were downloaded from the AML TCGA data portal. Cases with available RNAseq, BMI and FAB classification data (177/200) were used in the present study. Cases were classified by FAB in "APL" (FAB="M3") and "non-APL" (FAB ≠ "M3") and by BMI in "obese" (BMI ≥ 30) and "non-obese" (BMI < 30). Genes with < 0.2 RPKM in at least 75% of patients were removed [5]. The Quantitative Set

Analysis for Gene Expression method as implemented in the quSAGE package [12] in the R programming language (v 3.2.3) was used to conduct supervised gene set enrichment analysis. For each expressed gene, the quSAGE algorithm calculates a probability density function (PDF) of differential expression between two groups of samples. For each gene set, it then calculates "activity", ie the mean difference in log-expression of individual genes included in a gene set. Gene sets with False Discovery Rate (FDR) < 0.05 were considered significant. We focused on the KEGG and CGP gene set collections, downloaded from MSigDB (http://software.broadinstitute.org/gsea/msigdb/). The CGP collection was used to confirm enrichment of previously identified APL-specific gene signatures [13] (supplementary table S2). We focused on the KEGG collection as it is enriched for metabolism-associated gene annotations[14]. The script to generate the present results is available upon request.

Mutational data analysis

For the analysis in the TCGA data, mutational data were retrieved from the TCGA AML paper[5] and AML driver genes were downloaded from IntOgen [15]. For each gene, different mutations were conflated so that gene status in each patient was either "mutated" or "wild type". For each gene we then calculated the number of mutated or wild-type patients in the obese or non-obese groups, and calculated Odds Ratios (OR), 95% confidence intervals (CI) and p-values by Fisher's test with Benjamini-Hochberg correction. Only genes with >1 mutation in the dataset were considered, using the fdsm package in R.

For the analysis of the retrospective cohort, FLT3 Internal Tandem Duplication (ITD) mutational data were provided by the referring centers. Logistic regression was employed to calculate ORs with 95% CI.

Supplementary references

1. Bhaskaran K, Douglas I, Forbes H, et al. Body-mass index and risk of 22 specific cancers: a population-based cohort study of 5·24 million UK adults. *Lancet*. 2014;384(9945):755–765.
2. Herrett E, Gallagher AM, Bhaskaran K, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). *Int. J. Epidemiol.* 2015;44(3):827–836.
3. Bhaskaran K, Forbes HJ, Douglas I, Leon DA, Smeeth L. Representativeness and optimal use of body mass index (BMI) in the UK Clinical Practice Research Datalink (CPRD). *BMJ Open*. 2013;3(9):e003389.
4. Breccia M, Mazzarella L, Bagnardi V, et al. Increased BMI correlates with higher risk of disease relapse and differentiation syndrome in patients with acute promyelocytic leukemia treated with the AIDA protocols. *Blood*. 2012;119(1):49–54.
5. Ley TJ, Miller C, Ding L, et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N. Engl. J. Med.* 2013;368(22):2059–2074.
6. Avvisati G, Lo-Coco F, Paoloni FP, et al. AIDA 0493 protocol for newly diagnosed acute promyelocytic leukemia: very long-term results and role of maintenance. *Blood*. 2011;117(18):4716–4725.

7. Lo-Coco F, Avvisati G, Vignetti M, et al. Front-line treatment of acute promyelocytic leukemia with AIDA induction followed by risk-adapted consolidation for adults younger than 61 years: results of the AIDA-2000 trial of the GIMEMA Group. *Blood*. 2010;116(17):3171–3179.
8. Adès L, Sanz MA, Chevret S, et al. Treatment of newly diagnosed acute promyelocytic leukemia (APL): a comparison of French-Belgian-Swiss and PETHEMA results. *Blood*. 2008;111(3):1078 LP – 1084.
9. ISTAT. Health for all. 2015;
10. European Commission. Eurostat population database. 2016;
11. Flegal KM, Carroll MD, Kit BK, Ogden CL. Prevalence of obesity and trends in the distribution of body mass index among US adults, 1999-2010. *JAMA*. 2012;307(5):491–7.
12. Yaari G, Bolen CR, Thakar J, Kleinstein SH. Quantitative set analysis for gene expression: A method to quantify gene set differential expression including gene-gene correlations. *Nucleic Acids Res.* 2013;41(18):1–11.
13. Casorelli I, Tenedini E, Tagliafico E, et al. Identification of a molecular signature for leukemic promyelocytes and their normal counterparts: Focus on DNA repair genes. *Leukemia*. 2006;20(11):1978–88.
14. Morishima K, Tanabe M, Furumichi M, Kanehisa M, Sato Y. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 2016;45(D1):D353–D361.

15. Gonzalez-Perez A, Perez-Llamas C, Deu-Pons J, et al. IntOGen-mutations identifies cancer drivers across tumor types. *Nat. Methods*. 2013;10:1081.

Table S1. quSage activity scores of previously idntified APL-associated signature and PPARG transcriptional targets

	M3vsNonM3.KEGG			ObVSNorm_All.KEGG			ObVSNorm_M3.KEGG			ObVSNorm_nonM3.KEG G		
pathway.name	log.fold.ch ange	p.Va lue	FDR	log.fold.ch ange	p.Va lue	FDR	log.fold.ch ange	p.Va lue	FDR	log.fold.ch ange	p.Va lue	FDR
CASORELLI_ACUTE_PROMYEL OCYTIC_LEUKEMIA_DN	-0.4517	0.00 00	0.0 000	-0.0506	0.29 15	0.9 995	0.3199	0.01 53	0.0 972	-0.0637	0.16 90	0.8 948
CASORELLI_ACUTE_PROMYEL OCYTIC_LEUKEMIA_UP	0.7557	0.00 00	0.0 000	0.1166	0.01 92	0.9 995	0.1573	0.10 24	0.2 172	0.0723	0.07 55	0.8 948
LI_ADIPOGENESIS_BY_ACTIVA TED_PPARG	1.0578	0.00 00	0.0 000	-0.0871	0.47 97	0.9 995	-0.1286	0.68 90	0.7 617	-0.1412	0.24 97	0.8 968
WANG_CLASSIC_ADIPOGENIC_ TARGETS_OF_PPARG	0.6578	0.00 00	0.0 000	0.0057	0.94 87	0.9 995	-0.1266	0.28 91	0.4 158	-0.0167	0.84 42	0.9 787

Table S2. quSage activity scores of KEGG gene sets. Only gene sets with FDR < 0.05 in at least one comparison are shown

	M3vsNonM3.KEGG			ObVSNorm_All.KEGG			ObVSNorm_M3.KEGG			ObVSNorm_nonM3.K EGG		
pathway.name	log.fold.c hange	p.V alue	FD R	log.fold.c hange	p.V alue	FD R	log.fold.c hange	p.V alue	FD R	log.fold.c hange	p.V alue	FD R
KEGG_RENIN_ANGIOTENSIN_SYSTEM	0.6503	0.0 023	0.0 09 3	-0.0187	0.8 880	0.9 85 9	-0.2661	0.5 992	0.6 94 3	-0.0294	0.8 297	0.9 90 1
KEGG_LINOLEIC_ACID_METABOLISM	0.6381	0.0 002	0.0 01 0	0.0537	0.6 317	0.9 85 9	0.1104	0.6 912	0.7 60 7	0.0136	0.9 083	0.9 90 1
KEGG_GLYCOSAMINOGLYCAN_BIOSYN THESIS_HEPARAN_SULFATE	0.4217	0.0 000	0.0 00 0	0.1973	0.0 022	0.4 09 8	0.2753	0.1 104	0.2 23 3	0.1680	0.0 122	0.9 90 1
KEGG_GLYCOSPHINGOLIPID_BIOSYNT HESIS_LACTO_AND_NEOLACTO_SERIES	0.3391	0.0 003	0.0 01 7	0.0490	0.4 612	0.9 85 9	0.0944	0.5 590	0.6 64 1	0.0264	0.7 106	0.9 90 1
KEGG_ALANINE_ASPARTATE_AND_GL UTAMATE_METABOLISM	0.3258	0.0 009	0.0 03 9	0.0366	0.5 529	0.9 85 9	0.2905	0.1 212	0.2 37 2	-0.0065	0.9 168	0.9 90 1

KEGG_ARACHIDONIC_ACID_METABOLI SM	0.3221	0.0 037	0.0 13 0	0.0907	0.2 234	0.9 85 9	0.0785	0.6 402	0.7 26 4	0.0751	0.3 437	0.9 90 1
KEGG_GLYCOSAMINOGLYCAN_DEGRA DATION	0.3208	0.0 000	0.0 00	0.1114	0.0 519	0.9 85	-0.0164	0.9 407	0.9 50	0.1077	0.0 707	0.9 90
			1			9			9			1
KEGG_HISTIDINE_METABOLISM	0.2996	0.0 044	0.0 14 8	0.0633	0.3 112	0.9 85 9	0.1274	0.4 197	0.5 47 5	0.0410	0.5 403	0.9 90 1
KEGG_ARGININE_AND_PROLINE_META BOLISM	0.2582	0.0 001	0.0 00 8	-0.0136	0.7 768	0.9 85 9	0.1283	0.2 240	0.3 65 5	-0.0421	0.4 087	0.9 90 1
KEGG_LIMONENE_AND_PINENE_DEGR ADATION	0.1662	0.0 087	0.0 25 0	0.0088	0.8 790	0.9 85 9	0.0364	0.7 788	0.8 13 8	-0.0030	0.9 581	0.9 90 1
KEGG_CARDIAC_MUSCLE_CONTRACTIO N	0.1475	0.0 084	0.0 24 5	0.0321	0.4 163	0.9 85 9	0.1541	0.0 864	0.1 98 5	0.0120	0.7 777	0.9 90 1
KEGG_PROTEIN_EXPORT	0.1439	0.0 066	0.0 20 4	-0.0374	0.2 974	0.9 85 9	0.1933	0.0 958	0.2 03 1	-0.0688	0.0 632	0.9 90 1
KEGG_PATHWAYS_IN_CANCER	0.1346	0.0 169	0.0 42 8	0.0002	0.9 991	0.9 99 1	0.2432	0.0 112	0.0 92 2	-0.0316	0.4 585	0.9 90 1
KEGG_UBIQUITIN_MEDIATED_PROTEO LYSIS	-0.0801	0.0 163	0.0 42 5	-0.0111	0.6 109	0.9 85 9	0.1777	0.0 128	0.0 92 2	-0.0258	0.2 492	0.9 90 1
KEGG_PANCREATIC_CANCER	-0.0866	0.0 187	0.0 46 3	-0.0124	0.6 339	0.9 85 9	0.1224	0.0 993	0.2 07 4	-0.0213	0.4 381	0.9 90 1
KEGG_INSULIN_SIGNALING_PATHWAY	-0.0874	0.0 170	0.0 42 8	0.0096	0.7 116	0.9 85 9	0.1847	0.0 012	0.0 40 2	-0.0031	0.9 115	0.9 90 1
KEGG_PYRIMIDINE_METABOLISM	-0.0975	0.0 165	0.0 42 5	0.0062	0.8 245	0.9 85 9	0.0774	0.3 112	0.4 63 1	0.0043	0.8 807	0.9 90 1
KEGG_RNA_DEGRADATION	-0.1018	0.0 146	0.0 39 5	-0.0229	0.4 004	0.9 85 9	0.1292	0.1 580	0.2 80 1	-0.0329	0.2 515	0.9 90 1

KEGG_AMYOTROPHIC_LATERAL_SCLE ROSIS_ALS	-0.1077	0.0002	0.001 3	0.0147	0.5221	0.985 9	0.0734	0.2217	0.365 5	0.0147	0.5431	0.990 1
KEGG_NON_SMALL_CELL_LUNG_CANCE R	-0.1128	0.0010	0.004 2	0.0152	0.5572	0.985 9	0.1381	0.0261	0.103 4	0.0090	0.7385	0.990 1
KEGG_CHRONIC_MYELOID_LEUKEMIA	-0.1140	0.0001	0.000 9	0.0155	0.4669	0.985 9	0.2026	0.0126	0.092 2	0.0030	0.8915	0.990 1
KEGG_COLORECTAL_CANCER	-0.1163	0.0002	0.001 0	0.0126	0.5758	0.985 9	0.2530	0.0000	0.001 0	-0.0052	0.8215	0.990 1
KEGG_PEROXISOME	-0.1242	0.0125	0.034 2	0.0283	0.4070	0.985 9	0.1403	0.1243	0.240 8	0.0241	0.5087	0.990 1
KEGG_APOPTOSIS	-0.1347	0.0008	0.003 8	0.0364	0.1948	0.985 9	0.1970	0.0187	0.103 4	0.0279	0.3395	0.990 1
KEGG_PROTEASOME	-0.1371	0.0013	0.005 3	-0.0334	0.2960	0.985 9	0.1156	0.2150	0.363 5	-0.0412	0.2221	0.990 1
KEGG_REGULATION_OF_ACTIN_CYTOS KELETON	-0.1426	0.0154	0.040	-0.0296	0.4645	0.985	0.1678	0.0559	0.153	-0.0419	0.3333	0.990
			9			9			3			1
KEGG_PHOSPHATIDYLINOSITOL_SIGN ALING_SYSTEM	-0.1478	0.0024	0.009 4	0.0232	0.4947	0.985 9	0.1730	0.0077	0.092 2	0.0164	0.6496	0.990 1
KEGG_LONG_TERM_POTENTIATION	-0.1525	0.0053	0.017 4	-0.0171	0.6000	0.985 9	0.1396	0.0108	0.092 2	-0.0247	0.4774	0.990 1
KEGG_VALINE_LEUCINE_AND_ISOLEUC INE_DEGRADATION	-0.1664	0.0021	0.008 5	-0.0054	0.8847	0.985 9	0.1395	0.3371	0.486 0	-0.0109	0.7775	0.990 1
KEGG_PENTOSE_PHOSPHATE_PATHW AY	-0.1698	0.0029	0.010 5	-0.0559	0.1519	0.985 9	-0.0887	0.2304	0.372 6	-0.0438	0.2926	0.990 1
KEGG_PATHOGENIC_ESCHERICHIA_CO LI_INFECTION	-0.1747	0.0028	0.0	0.0001	0.9983	0.999 1	0.2391	0.0248	0.103 4	-0.0145	0.7640	0.990 1

			10 5									
KEGG_WNT_SIGNALING_PATHWAY	-0.1798	0.0 003	0.0 01 8	-0.0202	0.5 735	0.9 85 9	0.2704	0.0 049	0.0 79 8	-0.0399	0.2 867	0.9 90 1
KEGG_MELANOGENESIS	-0.1822	0.0 101	0.0 28 5	0.0017	0.9 737	0.9 99 1	0.2365	0.0 708	0.1 72 8	-0.0120	0.8 199	0.9 90 1
KEGG_ENDOMETRIAL_CANCER	-0.1884	0.0 002	0.0 01 1	0.0121	0.7 239	0.9 85 9	0.1965	0.0 107	0.0 92 2	0.0039	0.9 124	0.9 90 1
KEGG_MISMATCH_REPAIR	-0.1993	0.0 191	0.0 46 8	-0.0195	0.7 019	0.9 85 9	0.3508	0.0 325	0.1 08 0	-0.0461	0.3 817	0.9 90 1
KEGG_RIBOFLAVIN_METABOLISM	-0.2067	0.0 027	0.0 10 2	-0.0038	0.9 237	0.9 87 4	0.0432	0.6 998	0.7 65 7	0.0027	0.9 527	0.9 90 1
KEGG_DNA_REPLICATION	-0.2114	0.0 206	0.0 49 6	-0.0159	0.7 776	0.9 85 9	0.2973	0.0 941	0.2 03 1	-0.0361	0.5 441	0.9 90 1
KEGG_JAK_STAT_SIGNALING_PATHWA Y	-0.2119	0.0 084	0.0 24 5	0.0150	0.7 826	0.9 85 9	0.2723	0.0 166	0.1 01 5	0.0008	0.9 909	0.9 91 8
KEGG_DORSO_VENTRAL_AXIS_FORMA TION	-0.2122	0.0 045	0.0 14 9	0.0181	0.7 070	0.9 85 9	0.2959	0.0 227	0.1 03 4	0.0018	0.9 728	0.9 91 8
KEGG_INOSITOL_PHOSPHATE_METAB OLISM	-0.2128	0.0 000	0.0 00 2	-0.0010	0.9 730	0.9 99 1	0.1404	0.0 113	0.0 92 2	-0.0037	0.9 133	0.9 90 1
KEGG_EPITHELIAL_CELL_SIGNALING_I N_HELICOBACTER_PYLORI_INFECTION	-0.2133	0.0 000	0.0 00 1	0.0413	0.2 478	0.9 85 9	0.1706	0.0 797	0.1 85 3	0.0403	0.2 762	0.9 90 1
KEGG_RIG_I_LIKE_RECEPTOR_SIGNALI NG_PATHWAY	-0.2148	0.0 006	0.0 02 7	0.0316	0.3 340	0.9 85 9	0.1770	0.1 163	0.2 32 5	0.0289	0.3 786	0.9 90 1
KEGG_LONG_TERM_DEPRESSION	-0.2202	0.0 068	0.0 20 7	-0.0218	0.6 805	0.9 85 9	0.1495	0.0 884	0.2 00 5	-0.0272	0.6 329	0.9 90 1
KEGG_BETA_ALANINE_METABOLISM	-0.2380	0.0 059	0.0 19 0	-0.0296	0.5 572	0.9 85 9	0.1582	0.4 291	0.5 50 4	-0.0358	0.4 898	0.9 90 1
KEGG_FRUCTOSE_AND_MANNOSE_ME TABOLISM	-0.2433	0.0 000	0.0 00	-0.0113	0.7 895	0.9 85	0.1312	0.0 715	0.1 72	-0.0124	0.7 805	0.9 90

			1			9			8			1
KEGG_DRUG_METABOLISM_OTHER_EN ZYMES	-0.2535	0.0 082	0.0 24 5	0.0270	0.6 671	0.9 85 9	-0.1247	0.5 606	0.6 64 1	0.0563	0.3 928	0.9 90 1
KEGG_O_GLYCAN_BIOSYNTHESIS	-0.2725	0.0 103	0.0 28 6	0.0280	0.6 628	0.9 85 9	0.4616	0.0 169	0.1 01 5	-0.0006	0.9 918	0.9 91 8
KEGG_CALCIUM_SIGNALING_PATHWA Y	-0.2771	0.0 001	0.0 00 4	-0.0054	0.9 034	0.9 85 9	0.2483	0.0 677	0.1 70 2	-0.0159	0.7 335	0.9 90 1
KEGG_VALINE_LEUCINE_AND_ISOLEUC INE_BIOSYNTHESIS	-0.2834	0.0 044	0.0 14 8	0.0075	0.8 942	0.9 85 9	0.2955	0.1 340	0.2 51 8	-0.0060	0.9 138	0.9 90 1
KEGG_NOD_LIKE_RECEPTOR_SIGNALIN G_PATHWAY	-0.2899	0.0 002	0.0 01 0	0.0682	0.1 886	0.9 85 9	0.2244	0.0 358	0.1 12 8	0.0688	0.2 078	0.9 90 1
KEGG_CYTOSOLIC_DNA_SENSING_PAT HWAY	-0.3003	0.0 000	0.0 00 1	0.0227	0.5 065	0.9 85 9	0.1045	0.3 320	0.4 86 0	0.0310	0.3 465	0.9 90 1
KEGG_B_CELL_RECEPTOR_SIGNALING_ PATHWAY	-0.3036	0.0 000	0.0 00 0	0.0131	0.7 092	0.9 85 9	0.2192	0.0 125	0.0 92 2	0.0090	0.7 968	0.9 90 1
KEGG_ERBB_SIGNALING_PATHWAY	-0.3255	0.0 000	0.0 00 0	0.0105	0.7 342	0.9 85 9	0.2496	0.0 011	0.0 40 2	0.0043	0.8 823	0.9 90 1
KEGG_GLYCEROLIPID_METABOLISM	-0.3303	0.0 001	0.0 00 4	-0.0405	0.4 091	0.9 85 9	0.2382	0.0 564	0.1 53 3	-0.0509	0.3 134	0.9 90 1
KEGG_FC_GAMMA_R_MEDIATED_PHAG OCYTOSIS	-0.3306	0.0 000	0.0 00 0	-0.0166	0.7 137	0.9 85 9	0.1495	0.0 525	0.1 50 3	-0.0154	0.7 414	0.9 90 1
KEGG_SPHINGOLIPID_METABOLISM	-0.3680	0.0 000	0.0 00 0	0.0174	0.6 685	0.9 85 9	0.3152	0.0 416	0.1 24 8	0.0076	0.8 462	0.9 90 1
KEGG_RETINOL_METABOLISM	-0.3804	0.0 029	0.0 10 5	0.0302	0.7 095	0.9 85 9	0.0697	0.7 325	0.7 92 1	0.0472	0.5 811	0.9 90 1
KEGG_GAP_JUNCTION				-0.0223			0.3313			-0.0360		

	-0.4063	0.0 000	0.0 00 1		0.6 911	0.9 85 9		0.0 358	0.1 12 8		0.5 254	0.9 90 1
KEGG_CELL_ADHESION_MOLECULES_C AMS	-0.4328	0.0 008	0.0 03 5	0.0313	0.7 135	0.9 85 9	0.5266	0.0 280	0.1 03 4	0.0053	0.9 546	0.9 90 1
KEGG_NON_HOMOLOGOUS_END_JOINI NG	-0.4837	0.0 005	0.0 02 6	0.0143	0.8 625	0.9 85 9	0.4767	0.0 342	0.1 11 5	-0.0058	0.9 434	0.9 90 1
KEGG_TOLL_LIKE_RECEPTOR_SIGNALI NG_PATHWAY	-0.4982	0.0 000	0.0 00 0	0.0443	0.4 457	0.9 85 9	0.3124	0.0 205	0.1 03 4	0.0448	0.4 386	0.9 90 1
KEGG_OLFACTORY_TRANSDUCTION	-0.5325	0.0 062	0.0 19 7	0.1045	0.4 100	0.9 85 9	0.1150	0.7 770	0.8 13 8	0.1333	0.3 200	0.9 90 1
KEGG_CHEMOKINE_SIGNALING_PATH WAY	-0.5333	0.0 000	0.0 00 0	-0.0104	0.8 722	0.9 85 9	0.1242	0.4 531	0.5 69 4	0.0050	0.9 422	0.9 90 1
KEGG_PENTOSE_AND_GLUCURONATE_ INTERCONVERSIONS	-0.6093	0.0 010	0.0 04	0.0156	0.8 942	0.9 85	-0.1942	0.5 218	0.6 32	0.0700	0.5 586	0.9 90
			2			9			0			1
KEGG_BIOSYNTHESIS_OF_UNSATURAT ED_FATTY_ACIDS	-0.6717	0.0 000	0.0 00 0	-0.0434	0.5 171	0.9 85 9	0.4095	0.0 304	0.1 06 2	-0.0528	0.4 121	0.9 90 1
KEGG_VIRAL_MYOCARDITIS	-0.6738	0.0 000	0.0 00 0	-0.0105	0.8 928	0.9 85 9	0.3705	0.0 201	0.1 03 4	-0.0122	0.8 761	0.9 90 1
KEGG_ANTIGEN_PROCESSING_AND_PR ESENTATION	-0.6982	0.0 000	0.0 00 0	0.0580	0.5 438	0.9 85 9	0.4852	0.0 123	0.0 92 2	0.0535	0.5 840	0.9 90 1
KEGG_SYSTEMIC_LUPUS_ERYTHEMAT OSUS	-0.7637	0.0 000	0.0 00 1	0.0663	0.5 374	0.9 85 9	0.1498	0.6 639	0.7 48 4	0.1000	0.3 574	0.9 90 1
KEGG_LEISHMANIA_INFECTION	-0.7943	0.0 000	0.0 00 0	0.0247	0.7 869	0.9 85 9	0.2493	0.1 456	0.2 66 7	0.0455	0.6 214	0.9 90 1
KEGG_PRIMARY_BILE_ACID_BIOSYNTH ESIS	-0.8093	0.0	0.0	-0.0140	0.8	0.9	0.0694	0.8	0.8	0.0216	0.8	0.9

		000	00 0		994	85 9		325	50 8		437	90 1
KEGG_GRAFT_VERSUS_HOST_DISEASE	-0.9436	0.0 000	0.0 00 1	0.1202	0.3 917	0.9 85 9	0.7427	0.0 248	0.1 03 4	0.1099	0.4 481	0.9 90 1
KEGG_TYPE_I_DIABETES_MELLITUS	-0.9923	0.0 000	0.0 00 0	0.0923	0.4 402	0.9 85 9	0.5909	0.0 210	0.1 03 4	0.0968	0.4 174	0.9 90 1
KEGG_ASTHMA	-0.9993	0.0 000	0.0 00 0	0.0057	0.9 727	0.9 99 1	0.4580	0.2 233	0.3 65 5	0.0146	0.9 282	0.9 90 1
KEGG_ALLOGRAFT_REJECTION	-1.0462	0.0 000	0.0 00 0	0.1173	0.4 038	0.9 85 9	0.6057	0.0 569	0.1 53 3	0.1261	0.3 788	0.9 90 1
KEGG_AUTOIMMUNE_THYROID_DISEA SE	-1.0623	0.0 000	0.0 00 0	0.0830	0.5 357	0.9 85 9	0.6382	0.0 408	0.1 24 3	0.0856	0.5 274	0.9 90 1
KEGG_INTESTINAL_IMMUNE_NETWOR K_FOR_IGA_PRODUCTION	-1.2645	0.0 000	0.0 00 0	-0.0217	0.8 626	0.9 85 9	0.4684	0.1 186	0.2 34 6	-0.0023	0.9 859	0.9 91 8
KEGG_PANTOTHENATE_AND_COA_BIO SYNTHESIS	-1.3380	0.0 000	0.0 00 0	-0.0035	0.9 670	0.9 99 1	0.4184	0.0 275	0.1 03 4	0.0268	0.7 322	0.9 90 1

Table S3. Mutation prevalences in obese ("OB") and non-obese ("NW") patients in the TCGA dataset

ID	OB_MUT	NW_MU T	OB_WT	NW_WT	OR	CI_inf	CI_sup	fisher	mLogPva l	FDR	Sum_Mut
FLT3	33	22	55	88	2.4000	1.2706	4.5335	0.0070	4.9636	0.1607	55
NPM1	23	30	65	80	0.9436	0.5005	1.7791	0.8732	0.1355	1	53
DNMT3A	20	29	68	81	0.8215	0.4269	1.5808	0.6206	0.4771	1	49
IDH2	9	11	79	99	1.0253	0.4049	2.5967	1.0000	0.0000	1	20
IDH1	10	9	78	101	1.4387	0.5577	3.7119	0.4757	0.7430	1	19

RUNX1	7	12	81	98	0.7058	0.2655	1.8758	0.6287	0.4642	1	19
TET2	7	10	81	100	0.8642	0.3150	2.3712	0.8052	0.2167	1	17
TP53	6	10	82	100	0.7317	0.2552	2.0981	0.6100	0.4943	1	16
NRAS	7	8	81	102	1.1019	0.3835	3.1660	1.0000	0.0000	1	15
CEBPA	6	6	82	104	1.2683	0.3944	4.0783	0.7689	0.2628	1	12
WT1	6	6	82	104	1.2683	0.3944	4.0783	0.7689	0.2628	1	12
PTPN11	4	5	84	105	1.0000	0.2604	3.8410	1.0000	0.0000	1	9
KIT	5	3	83	107	2.1486	0.4991	9.2497	0.4703	0.7544	1	8
KRAS	3	5	85	105	0.7412	0.1722	3.1903	0.7348	0.3081	1	8
U2AF1	2	6	86	104	0.4031	0.0793	2.0483	0.3037	1.1916	1	8
STAG2	2	5	86	105	0.4884	0.0924	2.5799	0.4654	0.7649	1	7
PHF6	1	5	87	105	0.2414	0.0277	2.1051	0.2291	1.4735	1	6
ASXL1	2	3	86	107	0.8295	0.1355	5.0763	1.0000	0.0000	1	5
RAD21	1	4	87	106	0.3046	0.0334	2.7754	0.3843	0.9564	1	5
KDM6A	2	2	86	108	1.2558	0.1733	9.0985	1.0000	0.0000	1	4
DIS3	1	2	87	108	0.6207	0.0554	6.9593	1.0000	0.0000	1	3
EZH2	0	3	88	107	0.0000	0.0000	NA	0.2555	1.3645	1	3
SUZ12	0	3	88	107	0.0000	0.0000	NA	0.2555	1.3645	1	3