Title:

Prospective functional classification of all possible missense variants in *PPARG*

Amit R. Majithia[1,2,3,4], Ben Tsuda[1¶], Maura Agostini[5¶], Keerthana Gnanapradeepan[1¶], Robert Rice[1], Gina Peloso[1,6], Kashyap A. Patel[7], Xiaolan Zhang[1], Marjoleine F. Broekema[8], Nick Patterson[1], Marc Duby[1], Ted Sharpe[1], Eric Kalkhoven[8], Evan D. Rosen[4,9], Inês Barroso[5], Sian Ellard[7,10], UK Monogenic Diabetes Consortium[11], Sekar Kathiresan[1,3,4,12] Myocardial Infarction Genetics Consortium[11], Stephen O'Rahilly[5], UK Congenital Lipodystrophy Consortium[11], Krishna Chatterjee[5], Jose C. Florez[1,2,3,4], Tarjei Mikkelsen[1#], David B. Savage[5**], and David Altshuler[1,2,3,4**#]

[1] Program in Medical & Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA

[2] Diabetes Research Center, Diabetes Unit, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA

[3] Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA

[4] Department of Medicine, Harvard Medical School, Boston, MA, USA

[5] University of Cambridge Metabolic Research Laboratories, Wellcome Trust-Medical Research Council Institute of Metabolic Science, Cambridge CB2 0QQ, United Kingdom

[6] Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA

[7] Institute of Biomedical and Clinical Science, University of Exeter Medical School, Exeter, UK

[8] Molecular Cancer Research and Center for Molecular Medicine, University Medical Centre Utrecht, Universiteitsweg 100, 3584 CG, Utrecht, The Netherlands

[9] Division of Endocrinology and Metabolism, Beth Israel Deaconess Medical Center, 330 Brookline Avenue, Boston, MA 02115, USA

[10] Department of Molecular Genetics, Royal Devon and Exeter National Health Service Foundation Trust, Exeter, UK

[11] A list of members and affiliations appears in the Supplementary Note.

[12] Cardiovascular Research Center, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA

Present addresses: 10X Genomics Inc., Pleasanton, CA 94566, USA (T.M.), Vertex Pharmaceuticals, Northern Ave, Boston, MA (D.A.)

¶ Equal contribution

**Co-senior author

Corresponding author:

Amit R. Majithia MD

Program in Medical & Population Genetics

Broad Institute of Harvard and MIT

75 Ames Street, 10105C

Cambridge, MA 02142

T: (617)-680-8002

E: amajithia@mgh.harvard.edu

47 **Introductory paragraph**

48

49 Clinical exome sequencing routinely identifies missense variants in disease-related genes, but
50 functional characterization is rarely undertaken, leading to diagnostic uncertainty[1,2]. For
51 example, mutations in *PPARG* cause Mendelian lipodystrophy[3],[4] and increase risk of type 2
52 diabetes (T2D)[5]. While approximately one in 500 people harbor missense variants in *PPARG,*
53 most are of unknown consequence. To prospectively characterize PPARγ variants we used
54 highly parallel oligonucleotide synthesis to construct a library encoding all 9,595 possible single
55 amino acid substitutions. We developed a pooled functional assay in human macrophages,
56 experimentally evaluated all protein variants, and used the experimental data to train a variant
57 classifier by supervised machine learning (http://miter.broadinstitute.org). When applied to 55
58 novel missense variants identified in population-based and clinical sequencing, the classifier
59 annotated six as pathogenic; these were subsequently validated by single-variant assays.
60 Saturation mutagenesis and prospective experimental characterization can support immediate
61 diagnostic interpretation of newly discovered missense variants in disease-related genes.

62

63

A major challenge in clinical exome sequencing is determining pathogenicity of missense variants incidentally found in genes previously implicated in a severe genetic disease [1,2,6]. Every exome contains ~200 missense variants that have never before been seen[7]. Few of these are in fact pathogenic, but functional testing is too slow and resource intensive for clinical use, leading to many Variants of Uncertain Significance (VUS)[8]. The lack of functional data and failure to explicitly incorporate information about ascertainment and prior probability can lead both to misdiagnosis[6,9] (if a benign variant is presumed pathogenic) and overestimation of penetrance (if modestly functional variants are systematically excluded from disease databases).

The peroxisome proliferator-activated receptor γ (PPARγ) exemplifies the challenge of classifying newly identified variants even in a well-studied disease gene. Rare mutations in *PPARG* cause familial partial lipodystrophy 3 (FPLD3)[3,4] and a common missense variant p.P12A, along with linked non-coding variants, associates with risk of T2D[10,11]. Molecular functions of PPARγ are well characterized[12,13] including its role as the target of anti-diabetic thiazolidinedione medications. Approximately 0.2% of the general population carries a rare missense variant in *PPARG,* but only 20% of these variants are functionally significant and associated with metabolic disease[5].

In order to enable functional interpretation of PPARγ variants identified in exome sequencing we constructed a cDNA library consisting of all possible amino acid substitutions in the protein (Figure 1A and Supplementary Figure 1). Based on the observation that primary human blood monocytes from patients with FPLD3 exhibit blunted PPARG response when stimulated with agonists ex vivo[13], the construct library was introduced into human macrophages edited to lack the endogenous *PPARG* gene (Supplementary Figure 2). After stimulation with PPARγ agonists, cells were FACS sorted according to the level of expression of CD36, a canonical target of PPARγ in multiple tissues[14,15] (Figure 1A). The sorted CD36+ and

CD36- cell populations were sequenced to determine the distribution of each PPARG variant in relation to CD36 activity.

"Function scores" were generated for each amino acid substitution at each site in PPARγ (see Methods, Figure 1B, Figure 2A) based on the partitioning of variants into CD36+/- FACS populations. Over 99% of all possible amino acid substitutions in the protein were covered. Of the twenty possible amino acid substitutions at each site, change to proline was most likely to reduce function, and to cysteine was best tolerated, consistent with the known conformational effects of amino acid side chains on protein structure[16]. Each of the 505 amino positions in PPARγ was assigned a "tolerance score" by combining function scores of the 19 alternative amino acids at that position (Figure 1B). Tolerance scores were overlaid on the known crystal structure of PPARγ (Figure 2B)[17],[18] demonstrating that amino acid positions that are intolerant of substitution cluster at residues that contact DNA, co-activating proteins, and ligands (rosiglitazone) (Figure 1B, 2B).

We next examined the function scores derived from the CD36/macrophage assay for those mutations previously reported in patients with lipodystrophy/insulin resistance and known to diminish PPARγ activity (Figure 2A). These pathogenic variants (Figure 2A, 2C), clustered in the PPARγ ligand-binding and DNA-binding domains[19],[4] and had function scores demonstrating enrichment in the CD36-"low" activity bin. In contrast, higher frequency variants including the common P12A variant had function scores demonstrating enrichment in the CD36-"high" activity bin (Figure 2C, Supplementary Table 1). The distribution of function scores for the pathogenic and common variants were significantly different ($p < 6x10^{-7}$, KS test).

Linear discriminant analysis was used to combine function scores for each of the 9,595 variants across multiple agonist conditions (Figure 2C) into a classifier that maximized discrimination between the set of lipodystrophy-associated variants and the set of high frequency variants described above. The classifier emits the likelihood of each variant being

drawn from either of the two classes (pathogenic or benign) and can be expressed as a continuous integrated function score (IFS) (Figure 2C-D).

As above and described in the Methods, the classifier was trained on pathogenic variants obtained from the published literature and benign variants from population-based sequencing[20]. In order to evaluate the performance of the model on independent data, we turned to novel variants obtained in population-based exome sequencing and sequencing of *PPARG* in patients referred to specialty clinics for possible lipodystrophy and early-onset diabetes. Specifically, we tested the predictions of functionality emitted by the classifier using standard assays and correlation to clinical phenotypes.

The classifier was applied to data from exome sequencing of 22,106 case/controls selected for study of early-onset myocardial infarction (MIGEN[21]). In total, 57 missense variants in *PPARG* were observed with minor allele frequency < 0.1%. Of these, 74% (n=42/57) were novel and thus had not previously been functionally characterized (Supplementary Table 1). In order to calculate a posterior probability of pathogenicity relevant to the clinical context in which the carriers were identified we combined the IFS of these variants with the estimated prevalence of FPLD3 in the general population (1:100,000-1:1,000,000[19]). One variant, p.R194Q, was estimated pathogenic with high posterior odds (benign:pathogenic) of 1:10,000. The individual who was heterozygous for p.R194Q carried a diagnosis of T2D and had fasting triglyceride levels in the 99th percentile (Supplementary Table 2)**.** As described below, p.R194Q was independently identified in a separate individual referred for clinical features of lipodystrophy (Figure 3, and Supplementary Table 3) who similarly manifested T2D and severe hypertriglyceridemia. Moreover, the p.R194Q variant abolished PPARγ transactivation activity in standard assays (Figure 3C). The combination of clinical and functional data indicate that p.R194Q is likely pathogenic, and that the individual from MIGEN may have undiagnosed FPLD3.

We next applied the classifier to variants ascertained from 335 patients referred to UK centers specializing in monogenic forms of diabetes and/or insulin resistance. Thirteen individuals were identified as carrying novel missense variants in *PPARG* (Supplementary Table 2 and 3), of whom 77% (10/13) had clinical features suggestive of lipodystrophy and associated metabolic derangement including severe insulin resistance, non-alcoholic fatty liver, dyslipidaemia and low serum adiponectin (Supplementary Table 3). The IFS for these thirteen variants were lower than those found in the population-based cohort (above and Figure 3A) (P<0.005 Student's t-test). For each variant, the posterior probability of pathogenicity was calculated by combining the IFS for that variant and the prevalence of FPLD3 in patients ascertained in these specialty clinics (~1:7 as estimated from the Cambridge national lipodystrophy clinic records).

Three variants (p.E54Q, p.D92N, p.D230N) were found in patients without clinical features of lipodystrophy who had been referred for sequencing based on suspected monogenic diabetes. Despite a higher prior probability based on ascertainment in specialty clinics, these three variants were classified as benign with high confidence (posterior odds benign:pathogenic = 200:1) (Supplementary Table 2). Moreover, when tested individually in standard PPARγ reporter assays these variants showed function indistinguishable from wild-type PPARγ (Figure 3C). Thus, the rate of benign variant identification in individuals ascertained in specialty clinics (~1:110, n=335) was similar to the rate of benign variants identified in the MIGEN cohort (~1:200, n=22,106).

Three variants (p.M31L, p.R308P, p.R385Q) classified as benign with high confidence were found in individuals with clinical features of partial lipodystrophy. The p.M31L variant was found in a female proband with features of lipodystrophy and metabolic derangement (Supplementary Table 3); critically, her daughter had a very similar fat distribution and metabolic phenotype but did not carry the p.M31L variant. Thus, in this case, the phenotype did not

segregate with genotype at *PPARG*. An individual with partial lipodystrophy carried p.R385Q, which was independently identified in a woman from the population-based cohort who had not developed T2D at age 61 (Supplementary Table 2). When tested in PPARγ reporter assays, these variants retained reporter activity, albeit subtly diminished under some conditions (Figure 3). The combination of functional testing, clinical data, and segregation / epidemiology suggests that p.M31L, p.R308P, and p.R385Q are likely incidental findings, although it is not possible to rule out that they act as partial risk-factors for metabolic phenotypes.

Six variants (p.R194Q, p.A417V, p.R212W, p.P387S, p.M203I, p.T356R) were found in patients with lipodystrophy and classified as pathogenic with high probability (posterior-odds benign:pathogenic = 1:>25,000). Five of the six were confirmed as defective in classical transactivation assays. The exception was p.R212W, where transactivation function when tested using a synthetic PPARγ response element (PPRE) was normal. However, R212W showed less activity in a reporter assay with an endogenous promoter (Figure 4A), and reduced in vitro binding to three PPREs (Figure 4B). The R212 side-chain forms multiple hydrogen-bond contacts in the minor-groove-bound DNA (Figure 4C), outside the main PPRE binding motif. These data indicate that R212W is likely a pathogenic variant despite not showing decreased activity in the traditional functional assay using a synthetic promoter.

Finally, p.T468K, found in a single patient with partial lipodystrophy, was classified by IFS as pathogenic with low confidence (posterior-odds benign:pathogenic = 2:3): its score fell in the overlapping tails of the benign and lipodystrophy-associated variant distributions. In PPARγ reporter assays, this variant demonstrated severely decreased function (Figure 3), supporting that p.T468K is likely a pathogenic variant.

We previously reported that rare missense variants in *PPARG* that impair function in a single-variant adipocyte differentiation assay confer increased risk of T2D in the general population [5]. We re-examined this relationship using functional annotation emitted by the

189 classifier (i.e. IFS) for the original sample of 118 *PPARG* variant carriers ascertained from
190 19,752 T2D case/controls (Figure 5A). We observe a long tail of variants with low IFS in T2D
191 cases but not controls (P =0.024, two-sample Kolmogorov-Smirnov test). We quantified this
192 inverse relationship between IFS and T2D case status (logistic regression beta = -0.49 +/- SE
193 0.15, *P*=0.002). The odds ratio for T2D in carriers of variants with the lowest tertile of IFS (as
194 compared to carriers of variants in the highest tertile) was 6.5 (95%CI 1.9 – 41) consistent with
195 our previously published estimate[5]. The odds ratio for the middle vs highest tertile of IFS was
196 2.0 (95%CI 1.3 – 3.1) suggesting that *PPARG* variants with even moderately reduced IFS
197 confer a modest increase in T2D risk. By contrast, a conventional predictor of mutation
198 deleteriousness (CONDEL score[22]) failed to distinguish between likely pathogenic and benign
199 variants (Figure 5b; P > 0.1 two-sample Kolmogorov-Smirnov test) by misclassifying many likely
200 benign variants as pathogenic (Figure 5C).
201 These data show that it is possible to experimentally characterize all possible missense
202 variants in a mammalian gene and use the information to guide interpretation of VUS, a concept
203 that has been previously applied to single protein domains[23],[24]. Testing variants prospectively
204 (that is, prior to their discovery in patients) overcomes barriers of time and scalability that have
205 thus far made it impractical to incorporate experimental data into routine clinical variant
206 interpretation. Furthermore, by simultaneously and consistently evaluating all variants in a single
207 experiment, more valid comparisons can be made across variants as compared to data on
208 different variants generated in different labs at different times.
209 The PPARG classifier annotated as benign nearly all variants (56/57) incidentally
210 identified in a study of myocardial infarction. The one variant classified as pathogenic with high
211 confidence (and confirmed by single variant laboratory experiments) was observed in an
212 individual with hypertriglyceridemia and T2D, and independently observed in a patient with
213 lipodystrophy, likely indicating FPLD3[25]. In 12/13 cases referred for suspected lipodystrophy or
214 monogenic diabetes and carrying a PPARG variant, the classifier provided immediate, high

confidence information regarding the likelihood of a functional defect and a molecular diagnosis of FPLD3. In only a single case (p.T468K) did the classifier not provide a high confidence estimate and low-throughput laboratory assays fail to corroborate the pooled assay data[13]. Systematic variant construction, pooled experimental characterization in relevant assays, and statistical integration with epidemiological data offer a generalizable approach to enable genome interpretation at clinically important genes, reducing overdiagnosis[6,9] and diagnostic uncertainty[8]. Fully realizing such comprehensive approaches will require a complementary array of methods[26]. The *PPARG* construct library is easily shared so that others can generate and contribute function scores in other assays[27], but as a transgene library it is not ideally suited for detecting functional effects of coding variation on splicing efficiency. Given the limitations on the library and because CD36 expression is unlikely to report on all the functions of PPARγ, we have made the PPARγ classifier available as a web application (http://miter.broadinstitute.org) that can be updated as new genetic and functional data become available. Broadening this approach to other genes and diseases will require cellular assays that read out disease relevant characteristics, are robust and scalable, and the availability of training sets of pathogenic and benign variants. Such assays and variants exist for a number of Mendelian disease genes, making it possible to apply a similar approach to help interpret VUS for many other clinical situations.

URLs

http://miter.broadinstitute.org; PPARG missense variant lookup table

www.cuh.org.uk/national-severe-insulin-resistance-service

www.diabetesgenes.org

http://www.broadinstitute.org/rnai/public/resources/protocols; lentivirus

http://www.broadinstitute.org/achilles; cell lines

Acknowledgements

Supported by grants from the National Institute of Diabetes, Digestive, and Kidney Diseases (1K08DK102877-01, to Dr. Majithia; 1R01DK097768-01, to Dr. Altshuler), NIH/Harvard Catalyst (1KL2 TR001100-01, to Dr. Majithia), Broad Institute (SPARC award, to Dr. Majithia and Dr. Mikkelsen), and Wellcome Trust (#095564, to Dr. Chatterjee; #107064, to Dr. Savage).

We thank John Doench, Cong Zhu, Daniel O'Connell, Glenn Cowley, Meagan Sullender, Daniel MacArthur, Eric Minkel, Brendan Bulik-Sullivan and Joseph Avruch for helpful discussions, laboratory assistance and manuscript review.

Dedicated to the memory of Promila Nandi April 30, 1933 – December 27, 2013

Author Contributions

A.R.M, T.M. and D.A. designed the study. A.R.M., B.T., M.A., and K.G. performed experiments with help from R.R., X.Z., M.F.B. and E.K. A.R.M and N.P. analyzed the data with help from B.T., T.S., G.P., K.A.P., M.D., and T.M. I.B., S.E., S.K., S.O.R., K.C. and D.B.S. contributed clinical data and genotypes. A.R.M and D.A. wrote the manuscript. D.B.S., S.O.R., K.C., E.D.R., and J.C.F revised the manuscript.

Competing financial interests

No competing financial interests

Figure Legends

Figure 1. Comprehensive functional testing of 9,595 PPARγ amino acid variants.

a) A library of 9,595 PPARG constructs was synthesized, each construct containing one amino acid substitution. The construct library was introduced into THP-1 monocytes (edited to lack the endogenous *PPARG* gene) such that each cell received a single construct. This polyclonal population of THP-1 monocytes was differentiated to macrophages and stimulated with PPARγ agonists (rosiglitazone, PGJ2); the stimulated macrophages were separated via fluoresence activated cell sorting according to expression of the PPARγ response gene CD36 into low (-) and high (+) activity bins. Each bin of cells was subject to next-generation sequencing at the transgenic PPARG locus to identify and tabulate introduced variants. PPARγ variant counts in the CD36 low and CD36 high bins were used to calculate a functional score for all 9,595 variants. b) Raw PPARγ function scores for each of the 9,595 variants plotted according to amino acid position along the PPARγ sequence. “Blue” denotes that any amino acid change away from reference results in low CD36 function score, whereas ”white” denotes that amino acid changes do not alter function; “grey” denotes the reference amino acid. Function scores summed by amino acid position are plotted to the right, denoting tolerance for any amino acid substitution away from reference.

Figure 2. Integrating experimental function to construct a PPARγ classification table.

a) Raw PPARγ function scores ranked for all 9,595 PPARγ variants tested. Highlighted in red are raw function scores of known lipodystrophy causing mutations if they reside in the DNA-binding domain (DBD) or in orange if they reside in the Ligand-binding domain (LBD). The

common P12A variant is shown in blue. b) Mutation tolerance scores as described in Figure1 are shown color-coded and mapped onto the known crystal structure of PPARγ with RXRα, NCoA and Rosiglitazone. “Red” denotes that amino acid changes away from reference results in low CD36 function score, whereas ”white” denotes that amino acid changes do not alter function. c) Raw PPARγ function scores were obtained for 9,595 variants under four experimental conditions: 1) 1 μM Rosiglitazone, 2) 0.1 μM Rosiglitazone, 3) 10 μM Prostaglandin J2, and 4) 0.1 μM Prostaglandin J2. The function of known benign (n=13) and lipodystrophy-causing (n=11) variants are highlighted in blue and red respectively with their overall distributions overlaid. The raw function scores were combined into an integrated function score (IFS) after classifier training using linear discriminant analysis (LDA).

Figure 3. Experimental and clinical classification of novel missense *PPARG* variants identified in sequenced individuals.

a) Variants identified in patients plotted according to their integrated function score (IFS) alongside the IFS distributions of known benign, and lipodystrophy associated variants. b) Diagnostic classification for Familial Partial Lipodystrophy 3 (FPLD3) expressed as posterior probability of non-pathogenicity of *PPARG* variants shown in (a). Posterior probability was calculated by combining IFS with prevalence of lipodystrophy in the general population (1:100,000) or from patients referred for lipodystrophy/familial diabetes (1:7). c) The variants identified in patients were individually recreated and tested for their ability to activate luciferase reporter constructs containing three, tandemly-repeated, copies of the PPRE from the Acyl-CoA oxidase gene linked to the thymidine kinase promoter under varying doses of pharmacologic (rosiglitazone) or endogenous (prostaglandin J2; PGJ2) ligands (mean +/- S.E.M n =5). Variants are grouped according to not-pathogenic/pathogenic designation in (b).

Figure 4 Ability of PPARγ p.R212W to transactivate gene expression and bind DNA at endogenous enhancers

a) Ability of PPARγ2 WT or R212W mutant to activate luciferase reporter constructs containing FABP4 promoter under varying doses of pharmacologic (rosiglitazone 0-1μM) or endogenous (prostaglandin J2; PGJ2 0-10μM) ligands (mean +/- S.E.M n = 5). b) Comparison of the DNA binding properties of in vitro translated wild type or mutant PPARγ proteins, tested in electrophoretic mobility shift assays using either γ1 (R184W) or γ2 (R212W) mutants and radiolabelled PPREs from the acyl coenzyme A oxidase (AcCoA: 5' ggaccAGGA-CAaAGGTCAcgtt 3'), fatty acid binding protein 4 (FABP4: 5'aaacaCAGGCAaAGGTCAgagg 3') or muscle carnitine palmitoyl transferase 1 (CPT1: 5' atcggTGACCTtTTCCCTaca 3') promoters with retinoid X receptor (RXR) and increasing concentrations of ligand (Rosiglitazone 0 to 10uM). RL, reticulocyte lysate. c) PPARγ colored by mutation tolerance scores obtained under stimulation with 1μM Rosiglitazone in THP-1 cells. As in Figure 2b, red represents sites that exhibited low CD36 response when mutated away from WT. Arginine 212 is highlighted which occurs in the 'hinge' region of PPARγ connecting the DNA binding and ligand binding domains. The positively charged arginine side chain extends into the minor groove of DNA forming multiple hydrogen bonds with bases.

Figure 5. Relationship of PPARγ function to T2D risk in the general population.

a) Missense PPARγ variants identified from 19,752 sequenced type 2 diabetes (T2D) case/controls plotted according to IFS (integrated functional score) from the PPARγ classification table alongside the IFS distributions of known benign, and lipodystrophy associated variants. Each point represents a missense variant; point size denote the number of individuals carrying that variant. Among the 118 individuals carrying missense PPARγ variants T2D cases contained a long tail of low-functioning missense variants, which was notably absent from the distribution of variants observed in T2D controls (p = 0.024 two-sample Kolmogorov-

Smirnov test). b) When the same 118 individuals were plotted according to computational prediction of deleteriousness no difference is distributions of functional variants is seen among T2D cases vs controls ($p > 0.1$ two-sample Kolmogorov-Smirnov test). c) Scatterplot of IFS vs computational prediction scores for PPARγ missense variants from T2D case/controls as described above.

References

1. Majewski, J., Schwartzentruber, J., Lalonde, E., Montpetit, A. & Jabado, N. What can exome sequencing do for you? *J Med Genet* **48**, 580-9 (2011).
2. Gahl, W.A. *et al.* The National Institutes of Health Undiagnosed Diseases Program: insights into rare diseases. *Genet Med* **14**, 51-9 (2012).
3. Barroso, I. *et al.* Dominant negative mutations in human PPARgamma associated with severe insulin resistance, diabetes mellitus and hypertension. *Nature* **402**, 880-3 (1999).
4. Jeninga, E. & Gurnell, M. Functional implications of genetic variation in human PPAR [gamma]. *Trends in Endocrinology & ...* (2009).
5. Majithia, A.R. *et al.* Rare variants in PPARG with decreased activity in adipocyte differentiation are associated with increased risk of type 2 diabetes. *Proc Natl Acad Sci U S A* **111**, 13127-32 (2014).
6. Flannick, J. *et al.* Assessing the phenotypic effects in the general population of rare variants in genes for a dominant Mendelian form of diabetes. *Nature genetics* **45**, 1380-1385 (2013).
7. Tennessen, J.A. *et al.* Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* **337**, 64-9 (2012).
8. McLaughlin, H.M. *et al.* A systematic approach to the reporting of medically relevant findings from whole genome sequencing. *BMC Med Genet* **15**, 134 (2014).
9. Manrai, A.K. *et al.* Genetic Misdiagnoses and the Potential for Health Disparities. *N Engl J Med* **375**, 655-65 (2016).
10. Altshuler, D. *et al.* The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat Genet* **26**, 76-80 (2000).
11. Claussnitzer, M. *et al.* Leveraging cross-species transcription factor binding site patterns: from diabetes risk loci to disease mechanisms. *Cell* **156**, 343-58 (2014).
12. Tontonoz, P. & Spiegelman, B.M. Fat and beyond: the diverse biology of PPARgamma. *Annu Rev Biochem* **77**, 289-312 (2008).
13. Agostini, M. *et al.* Non-DNA binding, dominant-negative, human PPAR [gamma] mutations cause lipodystrophic insulin resistance. *Cell metabolism* **4**, 303-311 (2006).
14. Yu, S. *et al.* Adipocyte-specific gene expression and adipogenic steatosis in the mouse liver due to peroxisome proliferator-activated receptor gamma1 (PPARgamma1) overexpression. *J Biol Chem* **278**, 498-505 (2003).

15. Tontonoz, P., Nagy, L., Alvarez, J.G., Thomazy, V.A. & Evans, R.M. PPARgamma promotes monocyte/macrophage differentiation and uptake of oxidized LDL. *Cell* **93**, 241-52 (1998).
16. Barnes, M.R. & Gray, I.C. *Bioinformatics for geneticists*, xiv, 408 p., 8 p. of plates (Wiley, Chichester, West Sussex, England ; Hoboken, N.J., 2003).
17. Chandra, V. *et al.* Structure of the intact PPAR-gamma-RXR- nuclear receptor complex on DNA. *Nature* **456**, 350-356 (2008).
18. Schrodinger, LLC. The PyMOL Molecular Graphics System, Version 1.8. (2015).
19. Garg, A. Acquired and inherited lipodystrophies. *New England Journal of Medicine* (2004).
20. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* http://dx.doi.org/10.1038/nature19057(2016).
21. Myocardial Infarction Genetics Consortium, I. *et al.* Inactivating mutations in NPC1L1 and protection from coronary heart disease. *N Engl J Med* **371**, 2072-82 (2014).
22. Gonzalez-Perez, A. & Lopez-Bigas, N. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *Am J Hum Genet* **88**, 440-9 (2011).
23. Fowler, D.M. *et al.* High-resolution mapping of protein sequence-function relationships. *Nat Methods* **7**, 741-6 (2010).
24. Starita, L.M. *et al.* Massively Parallel Functional Analysis of BRCA1 RING Domain Variants. *Genetics* **200**, 413-22 (2015).
25. Demir, T. *et al.* Familial partial lipodystrophy linked to a novel peroxisome proliferator activator receptor -gamma (PPARG) mutation, H449L: a comparison of people with this mutation and those with classic codon 482 Lamin A/C (LMNA) mutations. *Diabet Med* (2016).
26. Findlay, G.M., Boyle, E.A., Hause, R.J., Klein, J.C. & Shendure, J. Saturation editing of genomic regions by multiplex homology-directed repair. *Nature* **513**, 120-3 (2014).
27. Fowler, D.M. & Fields, S. Deep mutational scanning: a new style of protein science. *Nat Methods* **11**, 801-7 (2014).
28. Melnikov, A., Rogov, P., Wang, L., Gnirke, A. & Mikkelsen, T.S. Comprehensive mutational scanning of a kinase in vivo reveals substrate-dependent fitness landscapes. *Nucleic Acids Res* **42**, e112 (2014).
29. Brinkman, E.K., Chen, T., Amendola, M. & van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res* **42**, e168 (2014).
30. Forman, B.M. *et al.* 15-Deoxy-delta 12, 14-prostaglandin J2 is a ligand for the adipocyte determination factor PPAR gamma. *Cell* **83**, 803-12 (1995).
31. Ellard, S. *et al.* Improved genetic testing for monogenic diabetes using targeted next-generation sequencing. *Diabetologia* **56**, 1958-63 (2013).

Methods

Synthesis and assembly of 9,595 PPARG variant constructs

A library of all 9,595 possible single amino acid variants in PPARG was synthesized using a site-directed, multiplexed method (Mutagenesis by Integrated TilEs (MITE)[28]) adapted to render it suitable for saturation mutagenesis in mammalian cells. Detail is provided below where methodologic advancements were made permitting saturation mutagenesis of *PPARG*. First, the PPARG cDNA sequence (CCDS2609.1) was recoded (see Supplementary Table 4) to eliminate susceptibility to restriction enzymes and CRISPR/CAS9 targeting sgRNAs (see below) to enable a "delete and replace" strategy. As described previously, DNA oligonucleotides were synthesized on a programmable microarray, each oligonucleotide encoding a desired amino acid change but otherwise homologous to the template un-mutated *PPARG* in all other respects. Oligonucleotides were organized into 'tiles', where those within each tile differ in a central variable region but share identical 5' and 3' ends (see Supplementary Table 4). Tiles were staggered such that their variable regions collectively span the entire template. To ensure uniform amplification and reduce chimera formation for the longer PPARG template, the protocol was modified to amplify each tile by emulsion PCR (MICELLULA DNA Emulsion & Purification Kit; EURx). The resulting products were inserted into linearized plasmids (Phusion® High-Fidelity DNA Polymerase NEB M0530) that carry the remaining template sequence using multiplexed Gibson assembly (NEBuilder® HiFi DNA Assembly Master Mix, NEB, cat E2621L) according to the manufacturer's protocol. A "frameshift cleaning" procedure was introduced given that the most common error mode during library construction (25-30% of constructs; data not shown) resulted from oligo synthesis errors causing 1-2 bp indels. The PPARG template vector was designed such that all PPARG constructs terminated with amber stop codons (i.e. TAG) and bore an in-frame zeocin resistance cassette (pUC57-PPARG-zeo; GenScript). Constructs bearing frame-shifting indels were depleted by transforming into an amber suppressor cloning host (TG1, Lucigen) and selecting the construct library under zeocin and kanamycin dual selection. Library plasmids were purified from $>10^6$ colonies to preserve complexity and the frameshift depleted PPARG transgenes excised from the zeocin resistance

cassette. To enable mammalian cell transduction, the transgene library was transferred into a lenti-viral expression vector by simple restriction cloning and transfected into a packaging cell line to produce pooled lenti-virus according to standard protocols (pLXI_TRC401; http://www.broadinstitute.org/rnai/public/resources/protocols)[5].

Deletion of endogenous PPARG in THP-1 monocytes using CRISPR/CAS9

The endonuclease Cas9 and sgRNAs targeting exon 6 of PPARG (CCCAAACCTGATGGCTATAG) and exon 8 of a control gene, PHACTR1 (CTATCATTCTGCAGCCCGAG), were introduced into THP1 cells by lenti-viral transduction. To quantify modification of the endogenous gene, genomic DNA was extracted at multiple time points, amplified by PCR around the PPARG sgRNA target site (forward primer: GGAGAGCACAGT, reverse primer: AATCCAGAGTCCGCTGACCT) and Sanger sequenced. Cutting efficiency was determined using the TIDE web tool for decomposition analysis of the sequencing traces[29].

Twenty-one days after transduction of CRISPR/Cas9 with PPARG or control sgRNAs, cells were tested for PPARG response by gene (FABP4) and protein (CD36) expression to validate lack of functional endogenous *PPARG*. PPARG targeting sgRNA and control sgRNA treated THP1 cells were stimulated with 1 μM Rosiglitazone in THP1 growth media (RPMI 1640 + 10% heat-inactivated FBS + 1% PenStrep + 0.1% BME) for 72 hours. mRNA was then extracted and quantified for FABP4 gene expression(nanoString Technologies). For CD36 protein expression, THP1 cells were stimulated with 50 ng/mL PMA and 1 μM of Rosiglitazone in growth media for 72 hours. Cells were then detached from the plate, washed and stained with a monoclonal antibody to CD36 according to the manufacturer's protocol (Miltenyi 130-100-149) and subjected to flow cytometry.

Simultaneous testing of 9,595 PPARG variants in experimental assays

The PPARG construct library was introduced into a human monocytic cell line (THP-1: obtained from http://www.broadinstitute.org/achilles and tested mycoplasma negative) engineered through CRISPR/CAS9 to lack endogenous *PPARG* (Supplementary Figure 2) by pooled infection. While isoform 1 of PPARG is dominantly expressed in monocyte/macrophages, we expressed isoform 2, which is identical in sequence but encodes a protein with an additional 28 N-terminal amino acids. Both isoforms demonstrated identical ligand dependent activity. The pooled virus was diluted such that the multiplicity of infection (number of viral particles per cell) was 0.3 so that each monocyte would receive zero or a single PPARG variant. Uninfected cells were eliminated by selection with puromycin 2 μg/mL. Expression of the PPARG transgene was controlled by a doxycycline inducible promoter[5]. At least 10^7 cells were infected to ensure that each *PPARG* variant was independently represented in 1000 monocytes. The resulting polyclonal population of THP-1 monocytes containing the PPARG variant library was stimulated for 72 hours with 1) 50 μM phorbol ester (PMA) to induce differentiation into macrophages, 2) doxycycline 1 μg/mL to induce expression of *PPARG* constructs, and 3) low/high doses (based on ranges used in prior studies [13]) of thiazolidinedione (Roziglitazone 0.1 μM/1 μM) or proposed natural ligand[30] (Prostaglandin J2 (PGJ2) 0.1 μM/10 μM) to stimulate PPARG activity. The population of stimulated THP-1 macrophages was immuno-stained for CD36 (Miltenyi: 130-095-472), a cell surface protein that is a direct transcriptional target of PPARG [15]. Using fluorescence activated cell sorting, stained cells were grouped into two activity bins separated by at least 5-10 fold expression of CD36 and selected to encompass equal numbers of cells (Supplementary Figure 3). For each stimulation condition, at least three replicates were generated, each with at least 5×10^6 cells sorted. To re-identify and quantitate the PPARG variants in the CD36 'high' and 'low' bins, genomic DNA was extracted from the cells in each bin and the integrated proviral PPARG transgenes amplified by PCR and shotgun sequenced (Nextera, Illumina). Raw sequencing reads were aligned to the reference *PPARG* cDNA

sequence (see Supplementary Table 4) and the number of occurrences of each amino acid at each position along the coding region counted and tabulated with a custom aligner. To minimize erroneous mutation calls, only codons that matched designed mutations and consisted of high quality base calls (Phred score > 30) were tabulated. Over 99 percent of the designed amino acid substitutions were observed at least 50 times for a given experimental condition (see Supplementary Figure 1). A raw function score was calculated based on the ratio of observed frequencies of each mutant amino acid in the two CD36 activity bins (see Figure 1).

Calculation of raw function score

Control experiments showed that variants deleterious to PPARG function were enriched in the CD36 low fraction and benign variants enriched in the CD36 high fraction. We constructed a likelihood function based on the log-odds of an amino acid variant in the CD36 high and low fractions. The log-odds for each amino acid variant was estimated by maximizing a likelihood function based on the observed counts of each amino acid variant in the CD36 high and low fractions as well as the total read depth at that amino acid position. Data were combined across experimental replicates after determining replicate variability (see Supplementary Figure 4). To avoid spuriously high or low log-odds estimates for any given variant, we constrained the log-odds estimate with a Gaussian prior whose parameters were estimated from data combined across all variants. See “Supplemental Note: Supplementary Analytic Methods” for detailed specification.

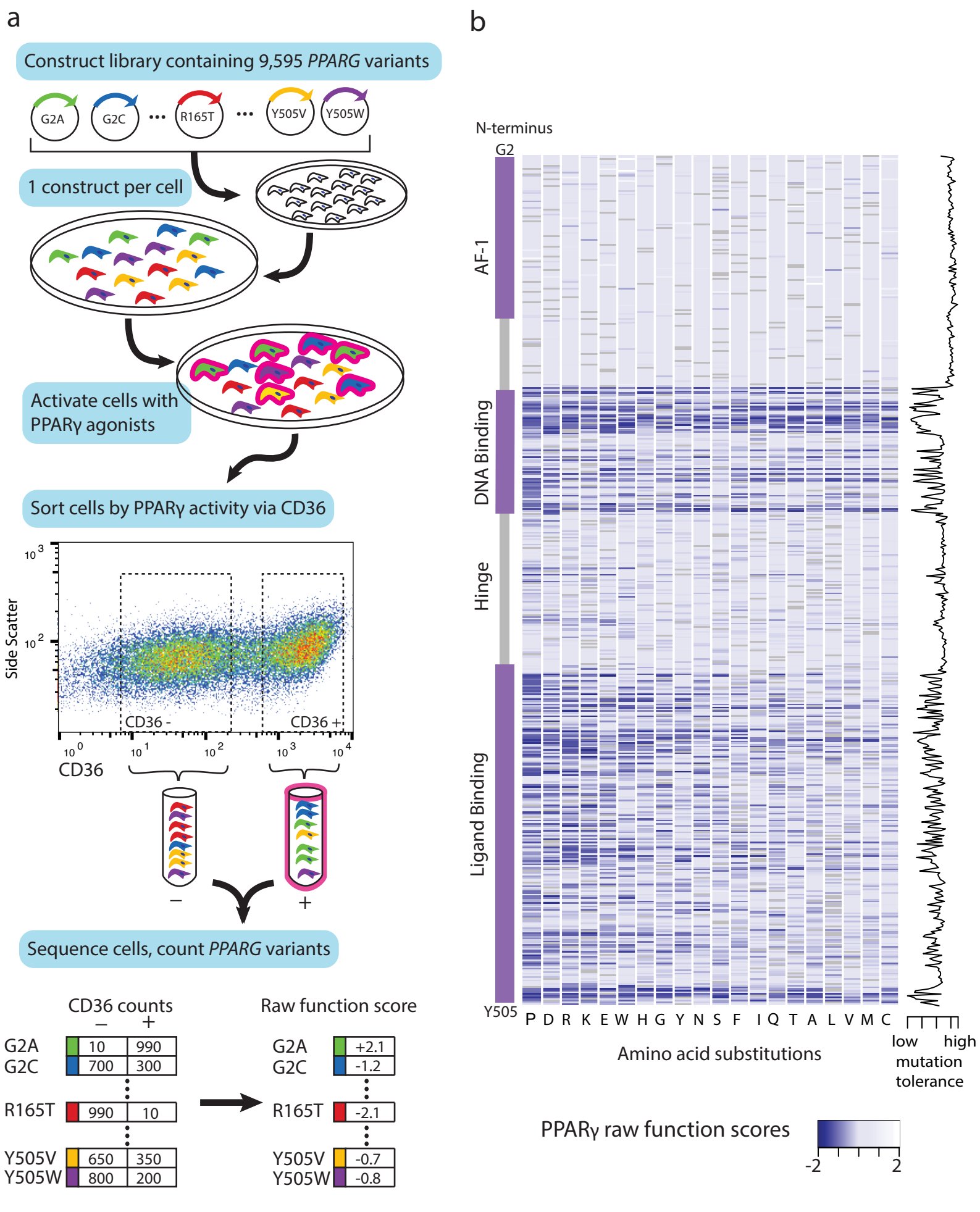
Construction of a PPARG classifier by supervised machine learning

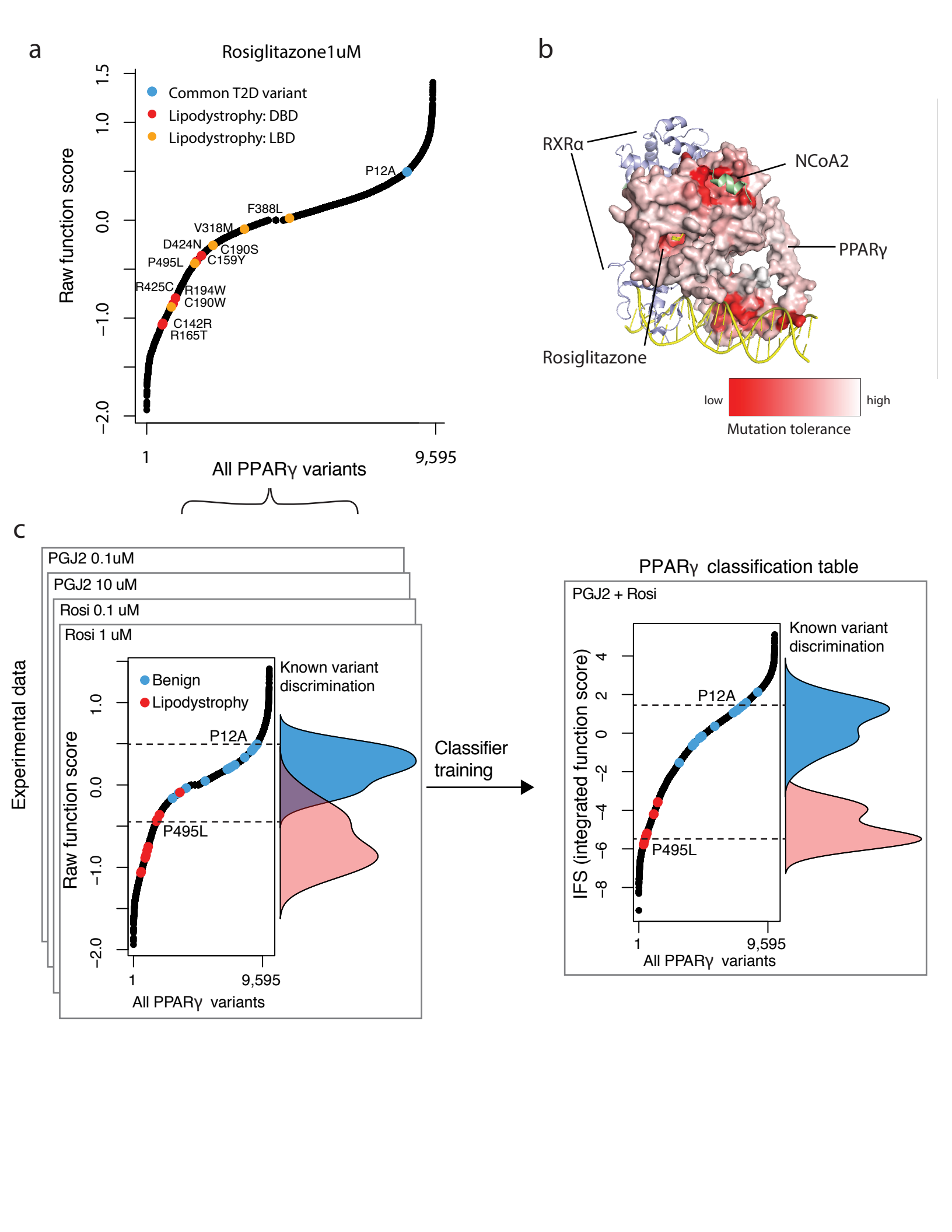
To predict the likelihood of novel variants being benign and pathogenic, we developed a classifier based on raw function scores obtained across various experimental conditions. The synthesis of multiple experimental conditions was intended to span a greater range of possible activities of PPARγ than would be queried using a single condition. Specifically, we used linear

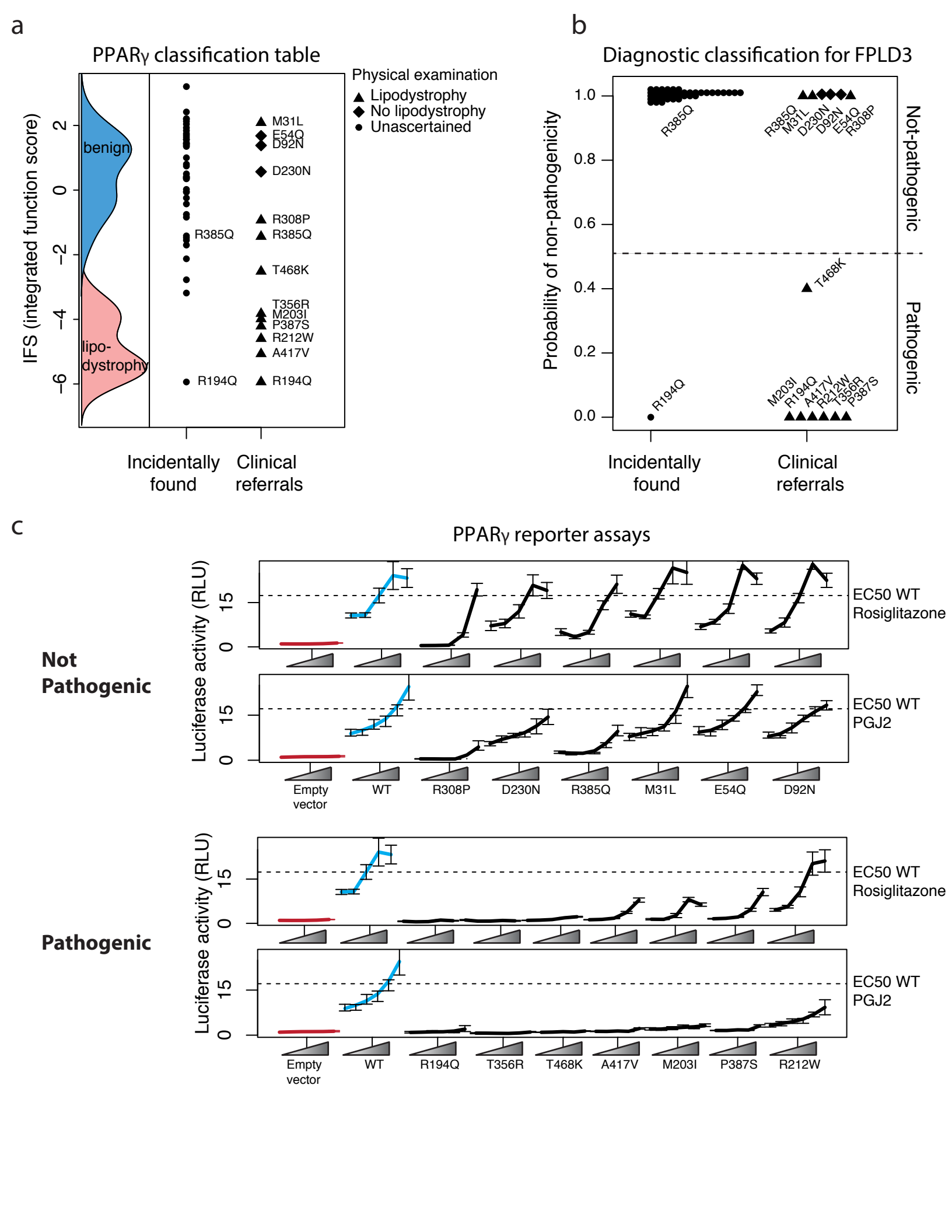
discriminant analysis (MASS package in R 3.0) to train the classifier, adopting a two-class model. The model incorporates as parameters (a) raw function scores for each PPARγ variant as measured across the four experimental conditions (i.e. rosiglitazone (Rosi) and Prostaglandin J2 (PGJ2) at high and low doses) and (b) mutation tolerance scores calculated for each position in PPARG as measured across the four experimental conditions (see Figure 1B). Potential classifiers were systematically constructed on linear combinations of four of these eight parameters, with a requirement that one parameter be included from each experimental condition. Classifier models were built for each the 16 possible combinations of four parameters using a training set of pathogenic and benign PPARγ variants (see Supplementary Table 1). Pathogenic variants used to train the classifier were selected based on (a) segregation with FPLD3 and (b) prior demonstration of loss-of-function in cellular assays. Benign variants used to train the classifier were selected from among variants identified in 60,706 aggregated exome sequences[20] at an allele frequency rendering them very unlikely to be causal for FPLD3 under a dominant model of inheritance and prevalence estimate ranging from 1:100,000 to 1:1,000,000 (P<0.05 1-tailed binomial probability n=121,412 chromosomes, p=10^{-5}) (see Supplementary table 1). The performance of these 16 models was compared using a leave-one-out cross-validation (LOOCV) protocol with each model scored by its aggregate ability to correctly classify the "left-out" variant over all the cycles of LOOCV. The highest scoring model consisted of raw function scores for each possible variant obtained from three conditions (Rosi 1μM, Rosi 0.1μM, PGJ2 10μM) and mutation tolerance score for each position in PPARG obtained from PGJ2 0.1μM. This model was fit to the full training dataset for prospective evaluation of novel *PPARG* variants. The weighted sum of the four parameters in the final model, as fit by the LDA algorithm, is denoted as the integrated function score (IFS) (see Figure 2C and Supplementary Figure 5) and represents an aggregate measure of variant function over the four experimental conditions. For clinical prediction, the IFS was expressed as an odds (benign:pathogenic),

546 which when multiplied by the estimated prior odds of FPLD3 based on the clinical situation (i.e.
547 prevalence) yielded an estimated probability of pathogenicity. Because the final model was
548 trained on the full set of available pathogenic and benign variants, its performance next required
549 prospective evaluation on a completely independent set of variants. These variants were
550 obtained from the population and clinic data described below, and evaluated as described in
551 Figure 3.
552
553 *Missense PPARG variants identified in population based exomes and clinically referred*
554 *individuals*
555 The study was conducted in accordance with the Declaration of Helsinki, and approved
556 by research ethics committees; written informed consent was obtained from all participants.
557 Missense *PPARG* variants were extracted from 22,106 exomes (8,400 with early-onset
558 coronary artery disease and 12,804 controls) sequenced by the Myocardial Genetics
559 Consortium (MIGEN) as described elsewhere[21]. Study participants were ascertained from the
560 following studies: ATVB, DHM, DUKE, JHS, ESP-EOMI, MedStar, OHS, PennCath,
561 PROCARDIS, PROMIS, and REGICOR. Participants were of European ancestry (n=12,849;
562 58%), Asian ancestry (n=6,823; 31%), African ancestry (n=2,399; 11%), and "other or unknown"
563 self-reported ethnicity (n=34; 0.2%). Twenty-two percent (n=4,258) reported a diagnosis of T2D.
564 Patients were referred to one of two UK centers (Cambridge: www.cuh.org.uk/national-
565 severe-insulin-resistance-service or Exeter: www.diabetesgenes.org) which specialize in
566 syndromes of severe insulin resistance and/or monogenic forms of diabetes. In clinically
567 suspected FPLD3 cases, mutations in *PPARG* were identified in genomic DNA extracted from
568 peripheral-blood leukocytes using *PPARG* amplification and sequencing. In patients for whom
569 FPLD3 was not the primary clinical diagnosis, *PPARG* was sequenced as part of a targeted
570 next-generation panel of 29 genes[31] selected to improve diagnostic yield for suspected
571 monogenic diabetes. Mutations were confirmed in index patients and, where possible, from

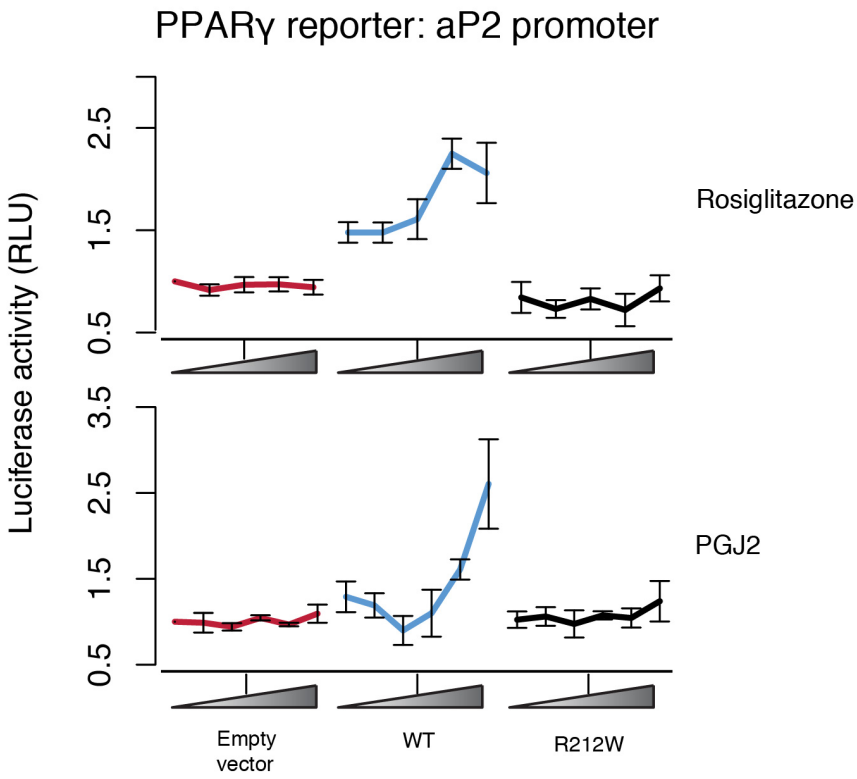
572 family members. In all instances, the nomenclature used for missense variants is for isoform 2
573 of *PPARG* (transcript accession: NM_015869.4; protein accession: NP_056953.2).
574
575 *Individual testing of PPARG variant function by transcriptional activity*
576 The novel variants identified in patients with suspected familial lipodystrophy or diabetes
577 were characterized using a well-established PPARG reporter containing three, tandemly-
578 repeated, copies of the PPRE from the Acyl-CoA oxidase (AcCoA: 5'
579 ggaccAGGACAaAGGTCAcgtt 3') gene upstream of the thymidine kinase (TK) promoter and
580 luciferase. In brief, 293EBNA cells, cultured in DMEM/10%FCS were transfected with
581 Lipofectamine2000 in 24-well plates and assayed for luciferase and β-galactosidase activity as
582 described previously[13] following a 36-hour incubation with or without ligand.
583
584


a
Construct library containing 9,595 PPARG variants
G2A
G2C
R165T
Y505V
Y505W
1 construct per cell
Activate cells with PPARγ agonists
Sort cells by PPARγ activity via CD36
Side Scatter
CD36 -
CD36 +
CD36
–
+
Sequence cells, count PPARG variants
CD36 counts
Raw function score
G2A 10 990 +2.1
G2C 700 300 -1.2
R165T 990 10 -2.1
Y505V 650 350 -0.7
Y505W 800 200 -0.8
b
N-terminus
G2
AF-1
DNA Binding
Hinge
Ligand Binding
Y505
P D R K E W H G Y N S F I Q T A L V M C
Amino acid substitutions
low mutation tolerance
high
PPARγ raw function scores
-2
2

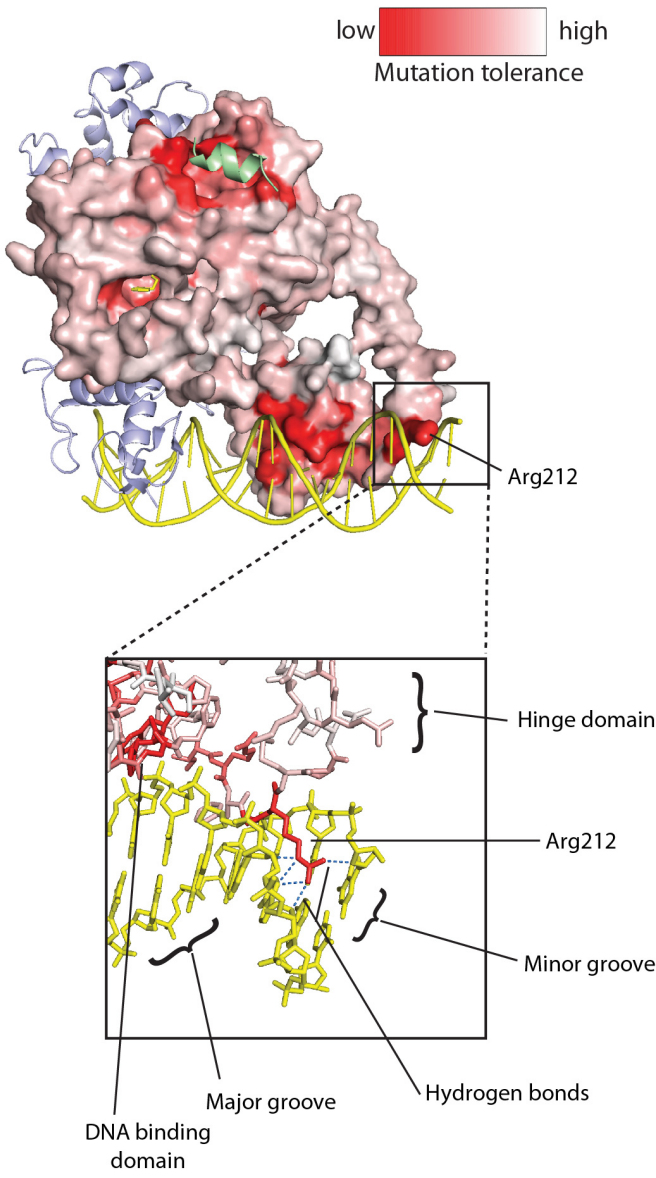

a
Rosiglitazone1uM
Common T2D variant
Lipodystrophy: DBD
Lipodystrophy: LBD
Raw function score
1.5
1.0
0.0
−1.0
−2.0
P12A
F388L
V318M
D424N
C190S
C159Y
P495L
R425C
R194W
C190W
C142R
R165T
1
9,595
All PPARγ variants
b
RXRα
NCoA2
PPARγ
Rosiglitazone
low
high
Mutation tolerance
c
Experimental data
PGJ2 0.1uM
PGJ2 10 uM
Rosi 0.1 uM
Rosi 1 uM
Benign
Lipodystrophy
Known variant discrimination
P12A
P495L
Raw function score
1.0
0.0
−1.0
−2.0
1
9,595
All PPARγ variants
Classifier training
PPARγ classification table
PGJ2 + Rosi
Known variant discrimination
IFS (integrated function score)
4
2
0
−2
−4
−6
−8
P12A
P495L
1
9,595
All PPARγ variants


a
PPARγ classification table
IFS (integrated function score)
benign
lipo-
dystrophy
Physical examination
Lipodystrophy
No lipodystrophy
Unascertained
M31L
E54Q
D92N
D230N
R308P
R385Q
R385Q
T468K
T356R
M203I
P387S
R212W
A417V
R194Q
R194Q
Incidentally found
Clinical referrals
b
Diagnostic classification for FPLD3
Probability of non-pathogenicity
Not-pathogenic
Pathogenic
R385Q
R385Q
M31L
D230N
D92N
E54Q
R308P
T468K
R194Q
M203I
R194Q
A417V
R212W
T356R
P387S
Incidentally found
Clinical referrals
c
PPARγ reporter assays
Not Pathogenic
Pathogenic
Luciferase activity (RLU)
EC50 WT Rosiglitazone
EC50 WT PGJ2
Empty vector
WT
R308P
D230N
R385Q
M31L
E54Q
D92N
Empty vector
WT
R194Q
T356R
T468K
A417V
M203I
P387S
R212W

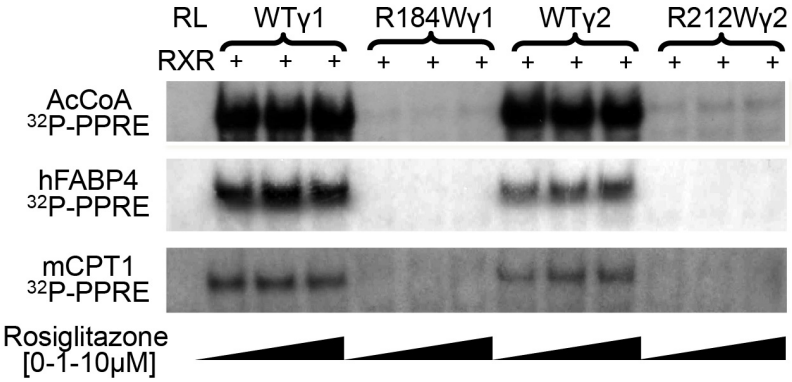
a

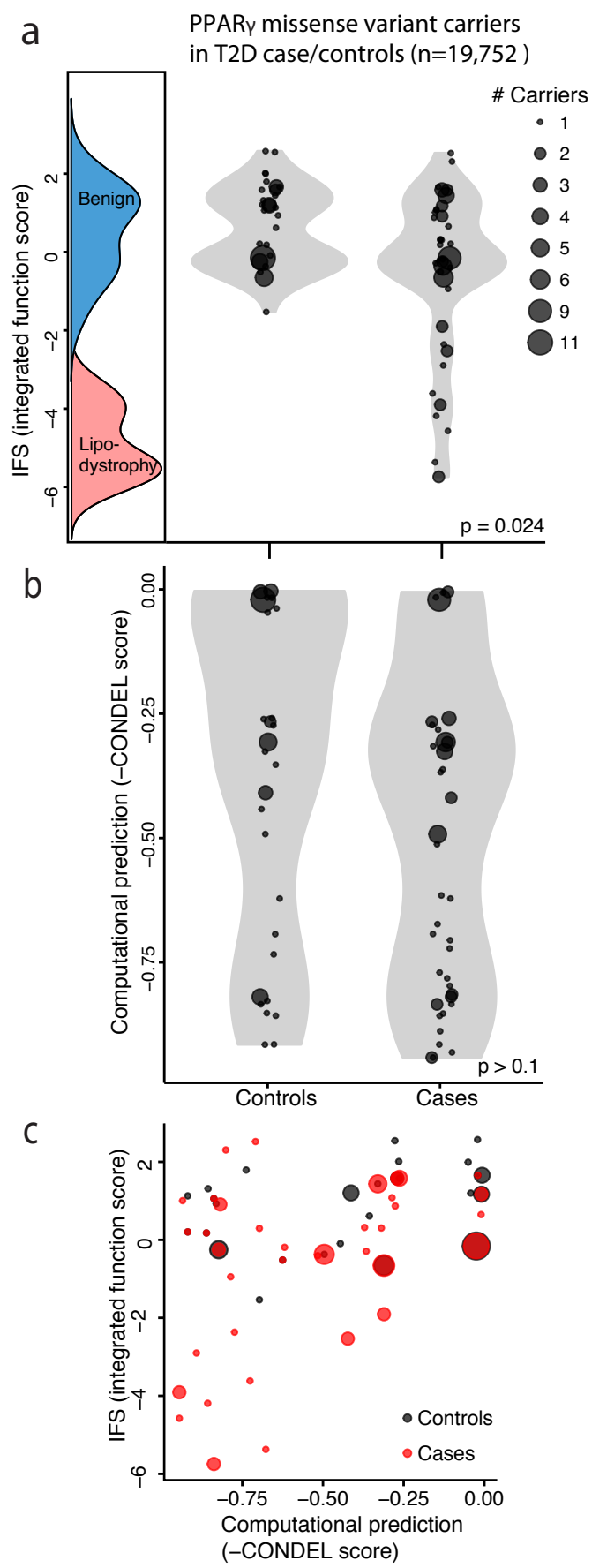

PPARγ reporter: aP2 promoter
Luciferase activity (RLU)
0.5
1.5
2.5
3.5
Rosiglitazone
PGJ2
Empty vector
WT
R212W

b


low
high
Mutation tolerance
Arg212
Hinge domain
Arg212
Minor groove
Major groove
Hydrogen bonds
DNA binding domain

c


RL
WTγ1
R184Wγ1
WTγ2
R212Wγ2
RXR
AcCoA 32P-PPRE
hFABP4 32P-PPRE
mCPT1 32P-PPRE
Rosiglitazone [0-1-10μM]


a
PPARγ missense variant carriers
in T2D case/controls (n=19,752)
Carriers
1
2
3
4
5
6
9
11
IFS (integrated function score)
Benign
Lipo-
dystrophy
p = 0.024
b
Computational prediction (–CONDEL score)
p > 0.1
Controls
Cases
c
IFS (integrated function score)
Controls
Cases
Computational prediction
(–CONDEL score)

Supplementary Note:

Author list

Supplementary Acknowledgements

Supplementary Tables 1-4

Author List

UK Congenital Lipodystrophy Consortium
Maura Agostini[1], Liang Dong[1], Erik Schoenmakers[1], Irene Campi[2], Alessandra Gambineri[3], Alison Sleigh[4], Iona Isaac[1], Anna Stears[1], Rachel Williams[1], Claire Adams[1], Rinki Murphy[5], Sudhesh Kumar[6], Naveed Sattar[7], John Petrie[7], Sebastian Kummer[8], Thomas Meissner[8], Verónica Mericq G.[9], Felicity Payne[10], Stephen O'Rahilly[1,13,14,16,17], Inês Barroso[10,12], Robert K. Semple[1, 14,18], Krishna Chatterjee[1,14,15], David B. Savage[1,11,13,14]

1. University of Cambridge Metabolic Research Laboratories, Wellcome Trust-Medical Research Council Institute of Metabolic Science, Cambridge CB2 0QQ, United Kingdom
2. Endocrine Unit, Fondazione IRCCS Ca' Granda, Milan, Italy
3. Endocrinology Unit, Department of Medical and Surgical Science, University of Bologna, Italy
4. Wolfson Brain Imaging Centre, University of Cambridge School of Clinical Medicine, Cambridge Biomedical Campus, Cambridge, CB20QQ, United Kingdom
5. Department of Medicine Faculty of Medical and Health Sciences, University of Auckland, New Zealand
6. University Hospital Coventry and Warwickshire & Warwick Medical School, University of Warwick, Coventry CV47AL, United Kingdom
7. BHF Glasgow Cardiovascular Research Centre, University of Glasgow, Glasgow, G128TA, United Kingdom
8. Department of General Pediatrics, Neonatology and Pediatric Cardiology, University Children's Hospital, Moorenstr. 5, 40225 Dusseldorf, Germany
9. Institute of Maternal and Child Research, Faculty of Medicine, University of Chile, Chile
10. Wellcome Trust Sanger Institute, CB10 1SA, UK
11. Wellcome Trust Grant Number 107064
12. Wellcome Trust Grant Number WT098051
13. U.K. NIHR Cambridge Biomedical Research Centre Grant
14. Core Facilities at the MRC Metabolic Diseases Unit Grant MC_UU_12012/5
15. Wellcome Trust Grant Number 095564
16. Wellcome Trust Grant Number 095515
17. EU/EFPIA Innovative Medicines Initiative Joint Undertaking EMIF 115372
18. Wellcome Trust Grant Number WT098498

UK Monogenic Diabetes Consortium
Kashyap A Patel[1], Kevin Colclough[2], Ewan Pearson[3], Sian Ellard[2,2]

1. Institute of Biomedical and Clinical Science, University of Exeter Medical School, Exeter, UK
2. Department of Molecular Genetics, Royal Devon and Exeter National Health Service Foundation Trust, Exeter, UK
3. School of Medicine, University of Dundee, Dundee, UK

Myocardial Infarction Genetics Consortium
Nathan O. Stitziel, M.D., Ph.D.[1,2], Hong-Hee Won, Ph.D.[3,4,5,6], Alanna C. Morrison, Ph.D.[7], Gina M. Peloso, Ph.D.[3,4,5,6], Ron Do, Ph.D.[3,4,5,6], Leslie A. Lange, Ph.D.[8], Pierre Fontanillas, Ph.D.[6], Namrata Gupta, Ph.D.[6], Stefano Duga, Ph.D.[9], Anuj Goel, M.Sc.[10], Martin Farrall, F.R.C.Path.[10], Danish Saleheen, MBBS, Ph.D.[11], Paola Ferrario, Ph.D.[12], Inke König, Ph.D.[12], Rosanna Asselta, Ph.D.[9], Piera Angelica Merlini, M.D.[13,14], Nicola Marziliano, Ph.D.[13], Maria Francesca Notarangelo, M.D.[13], Ursula Schick, M.S.[15], Paul Auer, Ph.D.[16], Themistocles L. Assimes, M.D., Ph.D.[17], Muredach Reilly, M.D.[18], Robert Wilensky, M.D.[18], Daniel J. Rader, M.D.[19], G. Kees Hovingh, M.D., Ph.D.[20], Thomas Meitinger, M.D.[21,22], Thorsten Kessler, M.D.[23], Adnan Kastrati, M.D.[22,23], Karl-Ludwig Laugwitz, M.D.[22,24], David Siscovick, M.D., M.P.H.[25], Jerome I. Rotter, M.D.[26], Stanley L. Hazen, M.D., Ph.D.[27], Russell Tracy, Ph.D.[28], Sharon Cresci, M.D.[1,29], John Spertus, M.D., M.P.H.[30], Rebecca Jackson, M.D.[31], Stephen M. Schwartz, Ph.D.[15,32], Pradeep Natarajan, M.D.[3,4,5,6], Jacy Crosby, Ph.D.[7], Donna Muzny, M.S.[33], Christie Ballantyne, M.D.[34], Stephen S. Rich, Ph.D.[35], Christopher J. O'Donnell, M.D.[5,36,37], Goncalo Abecasis, Ph.D.[38], Shamil Sunyaev, Ph.D.[6,39], Deborah A. Nickerson, Ph.D.[40], Julie E. Buring, Sc.D.[41], Paul M. Ridker, M.D.[41], Daniel I. Chasman, Ph.D.[41], Erin Austin, Ph.D.[42], Zi Ye, M.D., Ph.D.[42], Iftikhar J. Kullo, M.D.[42], Peter E. Weeke, M.D.[43,44], Christian M. Shaffer, B.S.[43], Lisa A. Bastarache, M.S.[45],

Joshua C. Denny, M.D., M.S.[43,45], Dan M. Roden, M.D.[43,46], Colin Palmer, Ph.D.[47], Panos Deloukas, Ph.D.[48], Dan-Yu Lin, Ph.D.[49], Zheng-zheng Tang, Ph.D.[50], Jeanette Erdmann, Ph.D.[51,52], Heribert Schunkert, M.D.[22,23], John Danesh, M.B., Ch.B., D.Phil.[53], Jaume Marrugat, M.D., Ph.D.[54], Roberto Elosua, M.D., Ph.D.[54], Diego Ardissino, M.D.[13,14], Ruth McPherson, M.D.[55], Hugh Watkins, M.D., Ph.D.[10], Alex P. Reiner, M.D., M.Sc.[15,32], James G. Wilson, M.D.[56], David Altshuler, M.D., Ph.D. [3,5,6], Richard A. Gibbs, Ph.D.[33], Eric S. Lander, Ph.D.[6], Eric Boerwinkle, Ph.D.[7,33], Stacey Gabriel, Ph.D.[6], Sekar Kathiresan, M.D.[3,4,5,6,37]

1. Cardiovascular Division, Department of Medicine, Washington University School of Medicine, Saint Louis, MO, USA
2. Division of Statistical Genomics, Washington University School of Medicine, Saint Louis, MO, USA
3. Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA
4. Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA
5. Department of Medicine, Harvard Medical School, Boston, MA, USA
6. Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA
7. Human Genetics Center, The University of Texas Health Science Center at Houston, Houston, TX, USA
8. Department of Genetics, University of North Carolina, Chapel Hill, NC, USA
9. Dipartimento di Biotecnologie Mediche e Medicina Traslazionale, Università degli Studi di Milano, Milan, Italy
10. Division of Cardiovascular Medicine, Radcliffe Department of Medicine and The Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK
11. Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, PA, USA; Center for Non-Communicable Diseases, Karachi, Pakistan
12. Institut für Medizinische Biometrie und Statistik, Universität zu Lübeck, Lübeck, Germany
13. Azienda Ospedaliero-Universitaria di Parma, Parma, Italy
14. Associazione per lo Studio della Trombosi in Cardiologia, Pavia, Italy
15. The Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA
16. School of Public Heath, University of Wisconsin-Milwaukee, Milwaukee, WI, USA
17. Stanford Cardiovascular Institute and the Division of Cardiovascular Medicine, Stanford University, Stanford, CA, USA
18. Cardiovascular Institute, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA
19. Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
20. Department of Vascular Medicine, Academic Medical Center, Amsterdam, The Netherlands
21. Institut für Humangenetik, Helmholtz Zentrum, Neuherberg, Germany and Institut für Humangenetik, Klinikum rechts der Isar, Technische Universität München, Munich, Germany
22. German Centre for Cardiovascular Research (DZHK), partner site Munich Heart Alliance, Munich, Germany
23. Deutsches Herzzentrum München, Technische Universität München, Munich, Germany
24. 1. Medizinische Klinik, Klinikum rechts der Isar, Technische Universität, München, Munich, Germany
25. Cardiovascular Health Research Unit, Department of Medicine, and Department of Epidemiology, University of Washington, Seattle, WA, USA
26. Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA, USA

27. Cardiovascular Medicine, Cleveland Clinic, Cleveland, OH, USA □
28. Departments of Pathology and Biochemistry, University of Vermont College □of Medicine, Burlington, VT, USA □
29. Department of Genetics, Washington University School of Medicine, Saint □Louis, MO, USA □
30. St. Luke's Mid America Heart Institute, University of Missouri-Kansas City, □Kansas City, MO, USA □
31. Division of Endocrinology, Diabetes and Metabolism, Department of □Medicine, Ohio State University, Columbus, OH, USA □
32. Department of Epidemiology, University of Washington, Seattle, □Washington, USA □
33. Human Genome Sequencing Center, Baylor College of Medicine, Houston, □TX, USA □
34. Section of Atherosclerosis and Vascular Medicine, Baylor College of □Medicine, Houston, TX, USA □
35. Center for Public Health Genomics, University of Virginia, Charlottesville, □VA, USA □
36. National Heart, Lung, and Blood Institute's Framingham Heart Study, □Framingham, MA, USA □
37. Cardiology Division, Massachusetts General Hospital, Boston, MA, USA □
38. Center for Statistical Genetics, Department of Biostatistics, University of □Michigan, Ann Arbor, MI, USA. □
39. Division of Genetics, Brigham and Women's Hospital, Harvard Medical □School, Boston, MA, USA □
40. Department of Genome Sciences, University of Washington, Seattle, WA, □USA □
41. Division of Preventive Medicine, Brigham and Women's Hospital and □Harvard Medical School, Boston, MA, USA □
42. Division of Cardiovascular Diseases, Mayo Clinic, Rochester, MN, USA □
43. Department of Medicine, Vanderbilt University, Nashville, TN, USA □
44. Department of Cardiology, Laboratory of Molecular Cardiology, □Copenhagen University Hospital Rigshospitalet, Denmark □
45. Department of Biomedical Informatics, Vanderbilt University, Nashville, TN, □USA □
46. Department of Pharmacology, Vanderbilt University, Nashville, TN, USA □
47. Medical Research Institute, University of Dundee, Dundee, UK □
48. Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA UK; □William Harvey Research Institute, Barts and The London School of □ Medicine and Dentistry, Queen Mary University of London, EC1M 6BQ UK; and King Abdulaziz University, Jeddah 21589, Saudi Arabia
49. Department of Biostatistics, University of North Carolina, Chapel Hill, NC, □USA □
50. Department of Biostatistics, Vanderbilt University, Nashville, TN, USA □
51. Institut für Integrative und Experimentelle Genomik, Universität zu Lübeck, □23562 Lübeck, Germany □
52. DZHK (German Research Centre for Cardiovascular Research) partner site □Hamberg/Lübeck /Kiel, Lübeck, Germany □
53. Public Health and Primary Care, University of Cambridge, Cambridge, UK □
54. Grupo de Epidemiología y Genética Cardiovascular, Institut Hospital del □Mar d'Investigacions Mèdiques (IMIM), Barcelona, Spain □
55. Division of Cardiology, University of Ottawa Heart Institute, Ottawa, ON, □Canada □
56. Department of Physiology and Biophysics, the University of Mississippi □Medical Center, Jackson, MS, USA □

Additional Acknowledgements
We would like to acknowledge The CUHFT Clinical Genetics Department and the UK10K Consortium. A full list of the investigators who contributed to the generation of data for the UK10K is available online (http://www.UK10K.org). Funding for UK10K was provided by the Wellcome Trust under award WT91310.

Supplementary Analytic Methods

1 Data

For each reference amino acid position x along PPARG2 and alternative amino acid t and run r we are given counts

$$C(x,t,r,k)\ k=0,1$$

One value of t is distinguished as wild-type (WT). Here '0' corresponds to '-' and '1' to '+' ; Fix x,t where t is not WT. We write

$$\begin{aligned} A(r,k) &= C(x,t,r,k) \\ W(r,k) &= C(x,WT,r,k) \end{aligned}$$

Here for the W (Wild-type) counts, we use the total coverage.

2 Calculation of raw function score

We wish to calculate a maximum likelihood estimator (MLE) of the offset to WT log-odds. for a probability p we have a map $ptoz(p) = \log(p/(1-p)))$ with inverse map $ztop(z) = \frac{e^z}{1+e^z}$ The wild type log-odds for run r is

$$w(r) = ptoz(W(r,1)/(W(r,0)+W(r,1)))$$

For offset θ we take the log-odds for run r to be $w(r)+\theta$ and the probability of 1 as:

$$p(r,\theta) = ztop(w(r)+\theta)$$

We now can define the log-likelihood:

$$\mathcal{L}(\theta) = \sum_r A(r,0)\log(1-p(r,\theta)) + A(r,1)\log p(r,\theta)$$

Set $\hat{\theta}$ to be the arg max of $\mathcal{L}(\theta)$.
We maximized $\mathcal{L}(\theta)$ on the interval $[-5,5]$ (R stats package).

Supplementary Analytic Methods

1 Data

For each reference amino acid position x along PPARG2 and alternative amino acid t and run r we are given counts

$$C(x,t,r,k)\ k=0,1$$

One value of t is distinguished as wild-type (WT). Here '0' corresponds to '-' and '1' to '+' ; Fix x,t where t is not WT. We write

$$\begin{aligned} A(r,k) &= C(x,t,r,k) \\ W(r,k) &= C(x,WT,r,k) \end{aligned}$$

2.1 Introduction of a prior

Especially if the counts A are small, the MLE for θ may be unreasonably extreme. We will assume that there is a distribution Q on θ and for a given mutant class set

$$\mathcal{L}'(\theta) = \log Q(\theta) + \mathcal{L}(\theta)$$

We took

$$Q(\theta,\sigma) = N(\theta;0,\sigma)$$

where N is the normal distribution with mean 0, variance σ^2, and set σ by maximizing the the global likelihood, using the following *empirical Bayes* procedure.

Set

$$\mathcal{M}(\theta,\sigma) = \exp\left(\log Q(\theta,\sigma) + \mathcal{L}(\theta)\right)$$

The likelihood for mutant class i is now obtained by integrating over θ:

$$\mathcal{K}_i(\sigma) = \int_{-\infty}^{\infty} \mathcal{M}(\theta_i,\sigma)\, d\theta_i \tag{1}$$

Then the overall log-likelihood is:

$$\mathcal{S}(\sigma) = \sum_i \log \mathcal{K}_i(\sigma) \tag{2}$$

Summarizing, We estimate σ by maximizing the log-likelihood $\mathcal{S}(\sigma)$, and then plug σ in to our estimation procedure for θ_i.

2 Calculation of raw function score

We wish to calculate a maximum likelihood estimator (MLE) of the offset to WT log-odds. for a probability p we have a map $ptoz(p) = \log(p/(1-p)))$ with inverse map $ztop(z) = \frac{e^z}{1+e^z}$ The wild type log-odds for run r is

$$w(r) = ptoz(W(r,1)/(W(r,0)+W(r,1)))$$

For offset θ we take the log-odds for run r to be $w(r)+\theta$ and the probability of 1 as:

$$p(r,\theta) = ztop(w(r)+\theta)$$

We now can define the log-likelihood:

$$\mathcal{L}(\theta) = \sum_r A(r,0)\log(1-p(r,\theta)) + A(r,1)\log p(r,\theta)$$

Set $\hat{\theta}$ to be the arg max of $\mathcal{L}(\theta)$.
We maximized $\mathcal{L}(\theta)$ on the interval $[-5,5]$ (we use standard usage).

Supplementary Table 1: Known variants used to train PPARG classification table and novel PPARG variants incidentally identified from 22,106 exomes

PPARG Variant	Classification
P12A	classifier training:benign
V276I	classifier training:benign
I331V	classifier training:benign
V335L	classifier training:benign
L361F	classifier training:benign
I437V	classifier training:benign
I45F	classifier training:benign
P454A	classifier training:benign
V48L	classifier training:benign
K486T	classifier training:benign
V52I	classifier training:benign
D55V	classifier training:benign
E79K	classifier training:benign
C142R	classifier training:lipodystrophy
Y151C	classifier training:lipodystrophy
C159Y	classifier training:lipodystrophy
R165T	classifier training:lipodystrophy
C190S	classifier training:lipodystrophy
C190W	classifier training:lipodystrophy
R194W	classifier training:lipodystrophy
V318M	classifier training:lipodystrophy
R425C	classifier training:lipodystrophy
P495L	classifier training:lipodystrophy
S104R	novel exome sequencing
D11Y	novel exome sequencing
S117A	novel exome sequencing
Q121R	novel exome sequencing
N132S	novel exome sequencing
S14G	novel exome sequencing
V141I	novel exome sequencing
K170N	novel exome sequencing
R181G	novel exome sequencing
R194Q	novel exome sequencing
I208V	novel exome sequencing
N233S	novel exome sequencing
P234S	novel exome sequencing
R240Q	novel exome sequencing
L246M	novel exome sequencing
A263V	novel exome sequencing
F292L	novel exome sequencing
T296N	novel exome sequencing
R308H	novel exome sequencing

V335I	novel exome sequencing
V350G	novel exome sequencing
M357V	novel exome sequencing
M36T	novel exome sequencing
N363S	novel exome sequencing
I369M	novel exome sequencing
R385Q	novel exome sequencing
W39R	novel exome sequencing
T41A	novel exome sequencing
S410G	novel exome sequencing
F43L	novel exome sequencing
I434V	novel exome sequencing
I437V	novel exome sequencing
V48L	novel exome sequencing
T487M	novel exome sequencing
S51F	novel exome sequencing
D55V	novel exome sequencing
I62N	novel exome sequencing
I62V	novel exome sequencing
T84P	novel exome sequencing
D92N	novel exome sequencing
K94E	novel exome sequencing
K98T	novel exome sequencing

Variant numbering is with respect to amino acid position on PPARG isoform 2

Supplementary Table 2. Biochemical Findings In Patients With Confirmed PPARG Mutations ascertained from targeted/exome sequencing.

PPARG mutation	Glu54Gln	Asp92Asn	Arg194Gln	Asp230Asn	Arg385Gln		Normal range
Clinical phenotype	Laurence-Moon-Biedl syndrome	Early onset diabetes in lean adult	unascertained	Morbidly obese with possible limb lipodystrophy	Partial lipodystrophy	unascertained	
Gender	Male	Male	Female	Female	Female	Female	
Age at time of assessment - yr	40	34	58	63	38	61	
Age at first presentation - yr	36	27	58	60	23	61	
Height – m	1.68	1.78	1.55	NA	1.85	1.65	
Weight – kg	154	61.8	58	NA	93.9	64.5	
BMI* - kg/m^2	54	20	24	43	27	24	
Hypertension	Yes	No	Yes	NA	Yes	Yes	
T2DM or IGT¶	Yes	Yes	Yes	Yes	Yes	No	
PCOS§	No	No	NA	NA	Yes	NA	
NAFLD♯	Yes (US)	No	NA	NA	Yes (US)	NA	
Triglyceride - mg/dL	327	168	613	133	230	**	<200
HDL-Cholesterol - mg/dL	39	66	30	60	31	**	>40
Total-Cholesterol - mg/dL	143	154	225	143	131	**	<150
Glycated hemoglobin -mmol/mol	62	140	NA	53	53	NA	20-41
Functional score	1.679	1.381	-5.938	0.568	-1.416		

NA denotes not available.

*The body-mass index (BMI) is the weight in kilograms divided by the height squared

** Treated hyperlipidemia

of the height in meters.

***Functional score as derived from http://miter.broadinstitute.org/

¶Type 2 diabetes mellitus (T2DM) or impaired glucose tolerance (IGT) – yes or no indicates the presence or absence of either of these conditions.

§Polycystic ovary syndrome (PCOS) - yes or no indicates the presence or absence of this syndrome.

♯Non-alcoholic fatty liver disease (NAFLD) – yes or no indicates presence of absence as confirmed by biopsy or ultrasound(US).

To convert the values for triglycerides to millimoles per liter multiply by 0.0113. To convert the values for cholesterol and HDL to millimoles per liter multiply by 0.0259.

Fatty liver was assessed by ultrasound.

Supplementary Table 3. Biochemical Findings In Patients With A Clinical Diagnosis Of Suspected Familial Partial Lipodystrophy Type 3 (FPLD3) And With Confirmed PPARG Mutations.

PPARG mutation	Met31Leu	Arg194Gln	Met203Ile	Arg212Trp	Arg308Pro	Thr356Arg	Pro387Ser	Ala417Val	Thr468Lys	Normal range
Gender	Female	Female	Female	Female	Female	Female	Female	Female	Female	
Age at time of assessment - years	56	46	17	31	16	19	13	40	15	
Age at first presentation - years	24	24	10	15	16	19	8	39	7	
Height – m	1.83	1.72	1.54	1.63	1.46	1.85	1.49	1.56	1.78	
Weight – kg	110.7	71.8	60.9	74.5	48.0	118.6	43.6	87.5	98.0	
BMI* - kg/m^2	33	24	26	28	23	34	20	36	31	
Fat distribution	Central obesity with limb and femorogluteal lipodystrophy	Limb and femorogluteal lipodystrophy	Limb and femorogluteal lipodystrophy	Limb and femorogluteal lipodystrophy	Limb and femorogluteal lipodystrophy	Central obesity with limb and femorogluteal lipodystrophy	Limb and femorogluteal lipodystrophy	Central obesity with limb and femorogluteal lipodystrophy	Central obesity with limb and femorogluteal lipodystrophy	
Total body fat - %	NA	NA	30	21	20	41	NA	NA	NA	
Predicted body fat - %**	42	29	31	34	27	43	NA	NA	NA	
Truncal fat - %	NA	NA	35	25	22	47	NA	NA	NA	
Leg fat - %	NA	NA	28	15	18	32	NA	NA	NA	
Fat mass ratio***	NA	NA	1.25	1.67	1.22	1.47	NA	NA	NA	>1.2
Hypertension	Yes	Yes	No	Yes	No	No	No	No	Yes	
T2DM or IGT¶	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
PCOS§	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	NA	
NAFLD♯	Yes (US)	Yes (biopsy)	Yes (US)	12 (MRS)	Yes (US)	18 (MRS)	Yes	Yes (US)	NA	<5 (MRS)
Triglyceride - mg/dL	221	5106	5000	212	1168	434	150	593	965	<150
HDL-Cholesterol - mg/dL	39	35	NA	19	15	15	46	35	31	>40

Total-Cholesterol - mg/dL	135	544	981	151	181	158	212	201	NA	<200
Insulin - pmol/L	240	128	177	306	405	476	160	165	310	<60
Glucose - mg/dL	111.7	304.5	122.5	82.9	131.5	120.7	66.7	167.6	94.5	<110
Glycated hemoglobin -mmol/mol	49	94	48	37	61	44	44	68	66	20-41
HOMA IR§§	4.50	3.86	3.45	5.24	7.46	7.35	2.67	3.44	5.49	<1.18
HOMA %S	22.2	25.9	29.0	19.1	13.4	13.6	37.5	29.1	18.2	100
Leptin - ug/L	18.7	4.3	5.4	5.2	NA	22.6	NA	18.7	NA	2.4-60.2
Adiponectin - mg/L	3.5	2.5	4.0	0.8	NA	2.9	NA	2.1	NA	2.6-17.7
Familial co-segregation	Obese daughter with PCOS is mutation negative.	Affected sibling is mutation positive	Unaffected mother is mutation negative and father with high fasting insulin level is mutation positive	NA	Unaffected mother and sibling are mutation negative	Affected father is mutation positive	Father with high triglycerides is mutation positive	NA	Mother with gestational diabetes is mutation positive	
****Functional score**	2.085	-6.267	-3.982	-4.598	-0.932	-3.823	-4.211	-5.068	-2.516	

NA denotes not available. Biochemical tests were all done after an overnight fast.

*The body-mass index (BMI) is the weight in kilograms divided by the square of the height in meters.

**Predicted body fat = (1.48*BMI)-7

***Fat mass ratio (FMR) is the trunk fat % divided by the leg fat %. FMR>1.2 in women is consistent with lipodystrophy, though not diagnostic in itself.

¶Type 2 diabetes mellitus (T2DM) or impaired glucose tolerance (IGT) – yes or no indicates the presence or absence of either of these conditions.

§Polycystic ovary syndrome (PCOS) - yes or no indicates the presence or absence of this syndrome.

♯Non-alcoholic fatty liver disease (NAFLD) – yes or no indicates presence of absence as confirmed by biopsy, ultrasound(US) or nuclear magnetic resonance spectroscopy (MRS) which is expressed as a ratio of the CH2/(CH2 + water) values.

§§HOMA IR (Homeostatic model assessment) and HOMA %S (% sensitivity) were calculated using the Oxford Homa Calculator https://www.dtu.ox.ac.uk/homacalculator/. Note that this calculator does not accept insulin values >400 pmol/L so in cases where fasting insulin levels were >400 pmol/L, we defaulted to 400 pmol/L. The HOMA estimates will thus over-estimate insulin sensitivity in these patients.

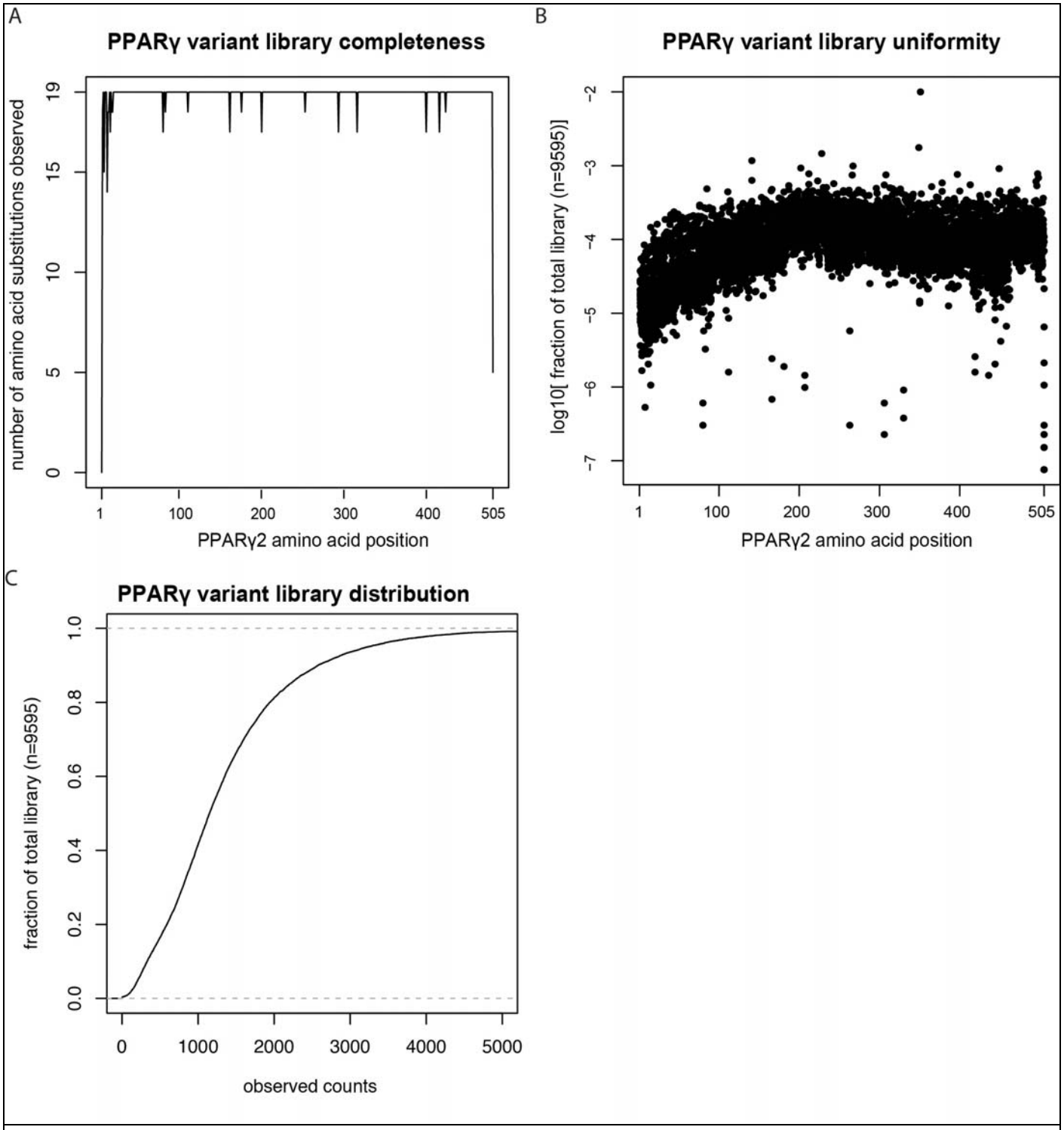
****Functional score as derived from http://miter.broadinstitute.org/

To convert the values for glucose into millimoles per liter multiply by 0.0555. To convert the values for insulin to microinternational units per milliliter divide by 6.945. To convert the values for triglycerides to millimoles per liter multiply by 0.0113. To convert the values for cholesterol and HDL to millimoles per liter multiply by 0.0259.

Fatty liver was assessed by ultrasound.

Supplementary Table 4: Template and Primer sequences for Mutagenesis by Integrated Tiles	
PPARG recoded cDNA template	ATGGGCGAGACCCTGGGCGACAGCCCCATCGACCCCGAGAGCG ACAGCTTCACCGACACCCTGAGCGCCAACATCAGCCAGGAGATG ACCATGGTGGACACCGAGATGCCCTTCTGGCCCACCAACTTCGG CATCAGCAGCGTGGACCTGAGCGTGATGGAGGACCACAGCCACA GCTTCGACATCAAGCCCTTCACCACCGTGGACTTCAGCAGCATCA GCACCCCCCACTACGAGGACATCCCCTTCACCCGCACCGACCCC GTGGTGGCCGACTACAAGTACGACCTGAAGCTGCAGGAGTACCA GAGCGCCATCAAGGTGGAGCCCGCCAGCCCCCCCTACTACAGC GAGAAGACCCAGCTGTACAACAAGCCCCACGAGGAGCCCAGCAA CAGCCTGATGGCCATCGAGTGCCGCGTGTGCGGCGACAAGGCC AGCGGCTTCCACTACGGCGTGCACGCCTGCGAGGGCTGCAAGG GCTTCTTCCGCCGCACCATCCGCCTGAAGCTGATCTACGACCGC TGCGACCTGAACTGCCGCATCCACAAGAAGAGCCGCAACAAGTG CCAGTACTGCCGCTTCCAGAAGTGCCTGGCCGTGGGCATGAGCC ACAACGCCATCCGCTTCGGCCGCATGCCCCAGGCCGAGAAGGA GAAGCTGCTGGCCGAGATCAGCAGCGACATCGACCAGCTGAACC CCGAGAGCGCCGACCTGCGCGCCCTGGCCAAGCACCTGTACGA CAGCTACATCAAGAGCTTCCCCCTGACCAAGGCCAAGGCCCGCG CCATCCTGACCGGCAAGACCACCGACAAGAGCCCCTTCGTGATC TACGACATGAACAGCCTGATGATGGGCGAGGACAAGATCAAGTT CAAGCACATCACCCCCCTGCAGGAGCAGAGCAAGGAGGTGGCC ATCCGCATCTTCCAGGGCTGCCAGTTCCGCAGCGTGGAGGCCGT GCAGGAGATCACCGAGTACGCCAAGAGCATCCCCGGCTTCGTGA ACCTGGACCTGAACGACCAGGTGACCCTGCTGAAGTACGGCGTG CACGAGATCATCTACACCATGCTGGCCAGCCTGATGAACAAGGA CGGCGTGCTGATCAGCGAGGGCCAGGGCTTCATGACCCGCGAG TTCCTGAAGAGCCTGCGCAAGCCCTTCGGCGACTTCATGGAGCC CAAGTTCGAGTTCGCCGTGAAGTTCAACGCCCTGGAGCTGGACG ACAGCGACCTGGCCATCTTCATCGCCGTGATCATCCTGAGCGGC GACCGCCCCGGCCTGCTGAACGTGAAGCCCATCGAGGACATCCA GGACAACCTGCTGCAGGCCCTGGAGCTGCAGCTGAAGCTGAACC ACCCCGAGAGCAGCCAGCTGTTCGCCAAGCTGCTGCAGAAGATG ACCGACCTGCGCCAGATCGTGACCGAGCACGTGCAGCTGCTGCA GGTGATCAAGAAGACCGAGACCGACATGAGCCTGCACCCCCTGC TGCAGGAGATCTACAAGGACCTGTAC
Tile amplification primers	
NR1C3_A1_amF	GTGGTGGAATTCTGCAGATATGC
NR1C3_B1_amF	TTCTGGCCCACCAACTTCGGCAT
NR1C3_A2_amF	ACCCGCACCGACCCCGTGGTGGC
NR1C3_B2_amF	CCCAGCAACAGCCTGATGGCCAT
NR1C3_A3_amF	TGCGACCTGAACTGCCGCATCCA
NR1C3_B3_amF	CTGGCCGAGATCAGCAGCGACAT
NR1C3_A4_amF	AAGACCACCGACAAGAGCCCCTT
NR1C3_B4_amF	CAGTTCCGCAGCGTGGAGGCCGT
NR1C3_A5_amF	AGCCTGATGAACAAGGACGGCGT
NR1C3_B5_amF	GAGCTGGACGACAGCGACCTGGC
NR1C3_A6_amF	CCGAGAGCAGCCAGCTGTTCGCC
NR1C3_A1_amR	TGGCTGTGGTCCTCCATCACGCT
NR1C3_B1_amR	CTCTGGTACTCCTGCAGCTTCAG

NR1C3_A2_amR	TGGAAGCCGCTGGCCTTGTCGCC
NR1C3_B2_amR	TGGAAGCGGCAGTACTGGCACTT
NR1C3_A3_amR	AGGGCGCGCAGGTCGGCGCTCTC
NR1C3_B3_amR	TCCTCGCCCATCATCAGGCTGTT
NR1C3_A4_amR	AAGCCGGGGATGCTCTTGGCGTA
NR1C3_B4_amR	AACTCGCGGGTCATGAAGCCCTG
NR1C3_A5_amR	GGGCGGTCGCCGCTCAGGATGAT
NR1C3_B5_amR	TGGCGCAGGTCGGTCATCTTCTG
NR1C3_A6_amR	AGGTCAGCAGGGACCCCCTTCCC
Template linearization primers	
NR1C3_A1_lnF	CATGGTGGCATATCTGCAGAATT
NR1C3_B1_lnF	GCTGCTGATGCCGAAGTTGGTGG
NR1C3_A2_lnF	GTAGTCGGCCACCACGGGGTCGG
NR1C3_B2_lnF	GCACTCGATGGCCATCAGGCTGT
NR1C3_A3_lnF	CTTCTTGTGGATGCGGCAGTTCA
NR1C3_B3_lnF	CTGGTCGATGTCGCTGCTGATCT
NR1C3_A4_lnF	GATCACGAAGGGGCTCTTGTCGG
NR1C3_B4_lnF	CTCCTGCACGGCCTCCACGCTGC
NR1C3_A5_lnF	GATCAGCACGCCGTCCTTGTTCA
NR1C3_B5_lnF	GAAGATGGCCAGGTCGCTGTCGT
NR1C3_A6_lnF	GGCGAACAGCTGGCTGCTCTCGG
NR1C3_A1_lnR	GGACCTGAGCGTGATGGAGGACC
NR1C3_B1_lnR	GTACGACCTGAAGCTGCAGGAGT
NR1C3_A2_lnR	CGTGTGCGGCGACAAGGCCAGCG
NR1C3_B2_lnR	CCGCAACAAGTGCCAGTACTGCC
NR1C3_A3_lnR	GAACCCCGAGAGCGCCGACCTGC
NR1C3_B3_lnR	CGACATGAACAGCCTGATGATGG
NR1C3_A4_lnR	CACCGAGTACGCCAAGAGCATCC
NR1C3_B4_lnR	CGAGGGCCAGGGCTTCATGACCC
NR1C3_A5_lnR	CGCCGTGATCATCCTGAGCGGCG
NR1C3_B5_lnR	GCTGCTGCAGAAGATGACCGACC
NR1C3_A6_lnR	GGGAAGGGGGTCCCTGCTGACCT

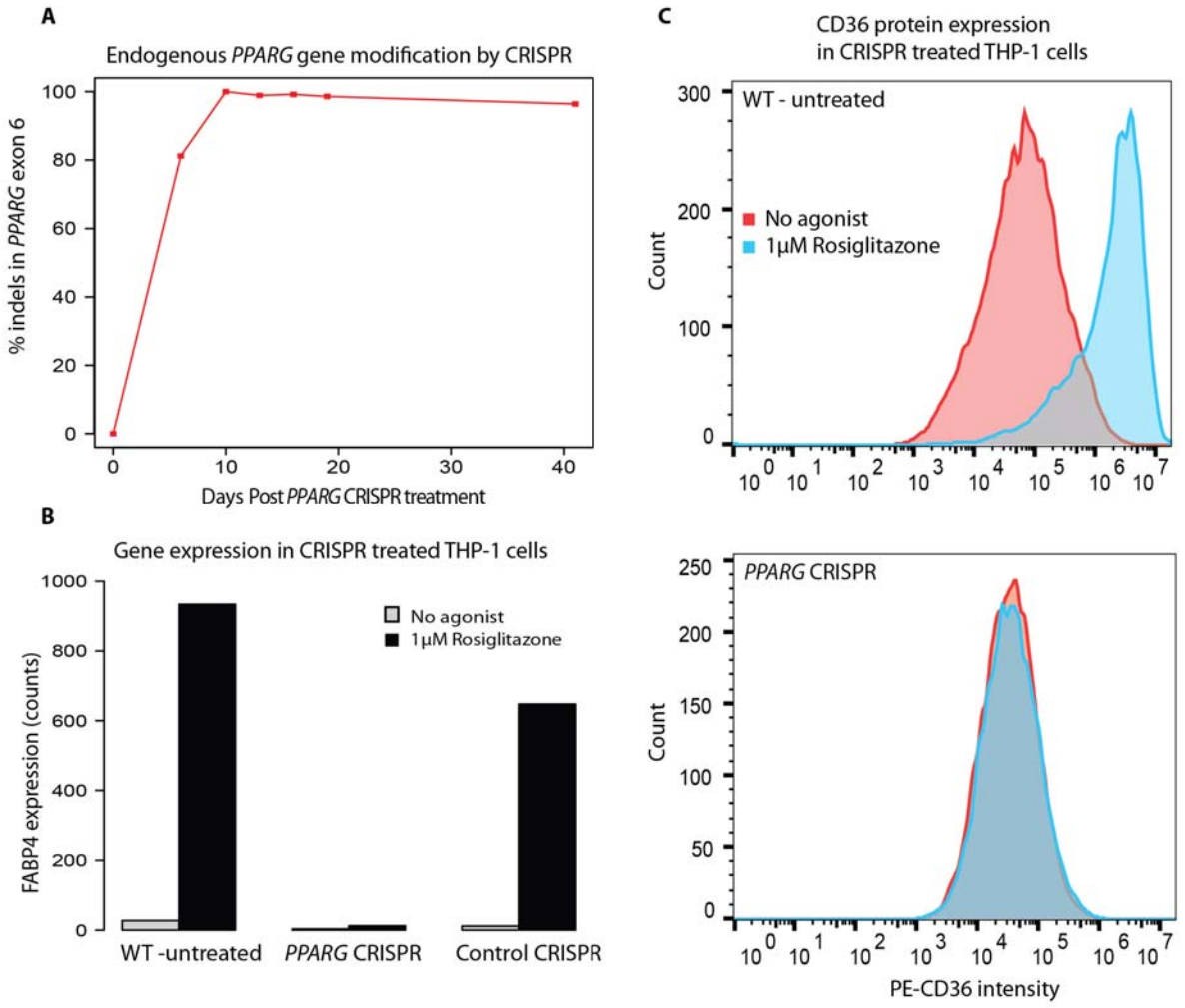


Supplementary Figure 1

PPARγ variant library completeness and uniformity

After transduction into THP1 macrohages, cell sorting and transgene recovery, the library of 9,595 PPARγ variant transgenes was shotgun sequenced (Nextera- Illumina) to assess completeness and balance. A) For each amino acid position along the PPARγ2

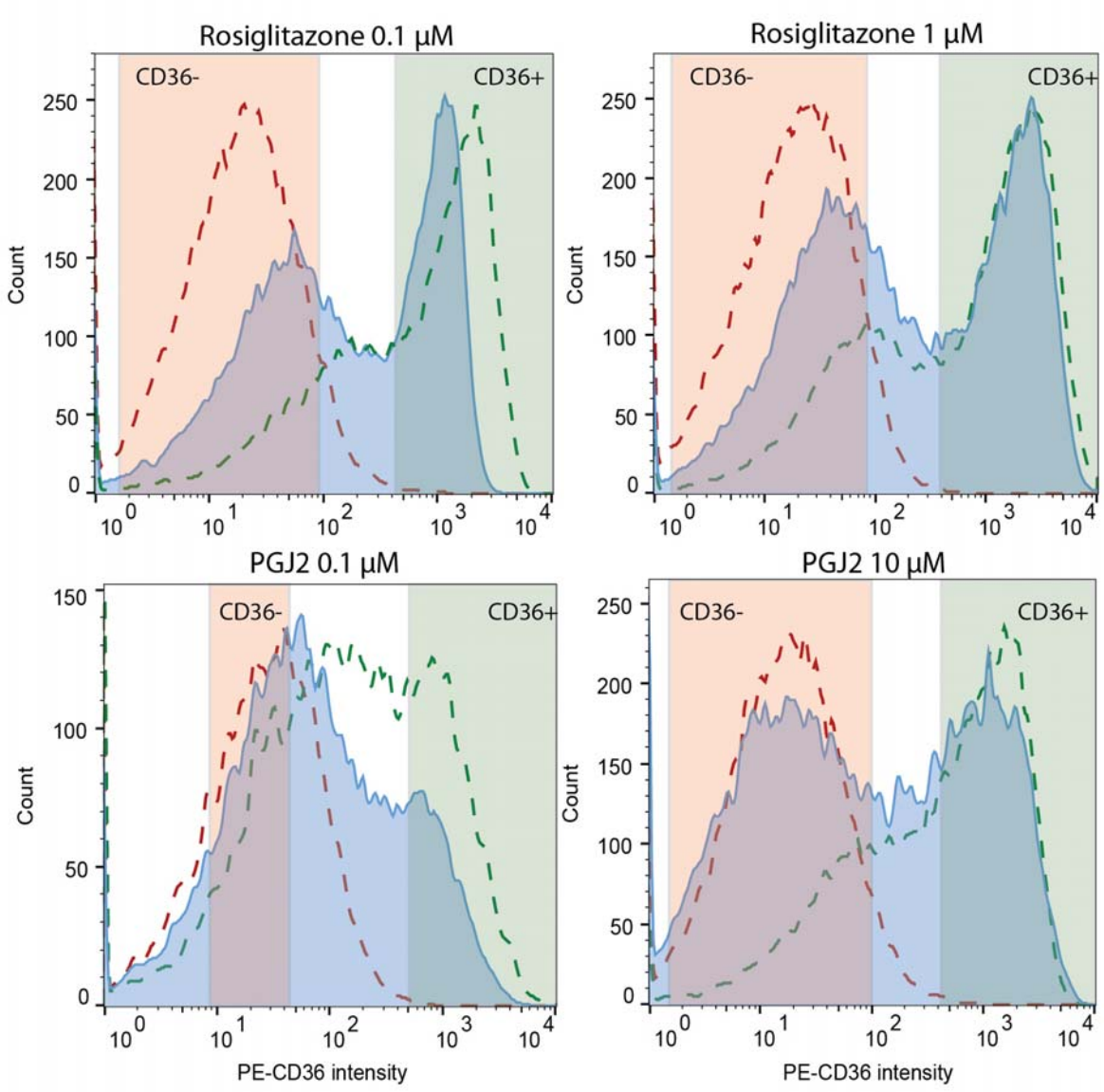
protein, the number of observed amino acid changes out of the 19 possible changes was quantified; 99.3% of the total 9,595 possible missense variants were observed. B) The distribution of missense variant frequency is plotted by position along the PPARγ2 protein. No single variant comprised more than 0.9% of the variant construct library. C) The cumulative distribution of counts observed for each variant is shown.



Supplementary Figure 2

Deletion of endogenous *PPARG* in THP-1 monocytes

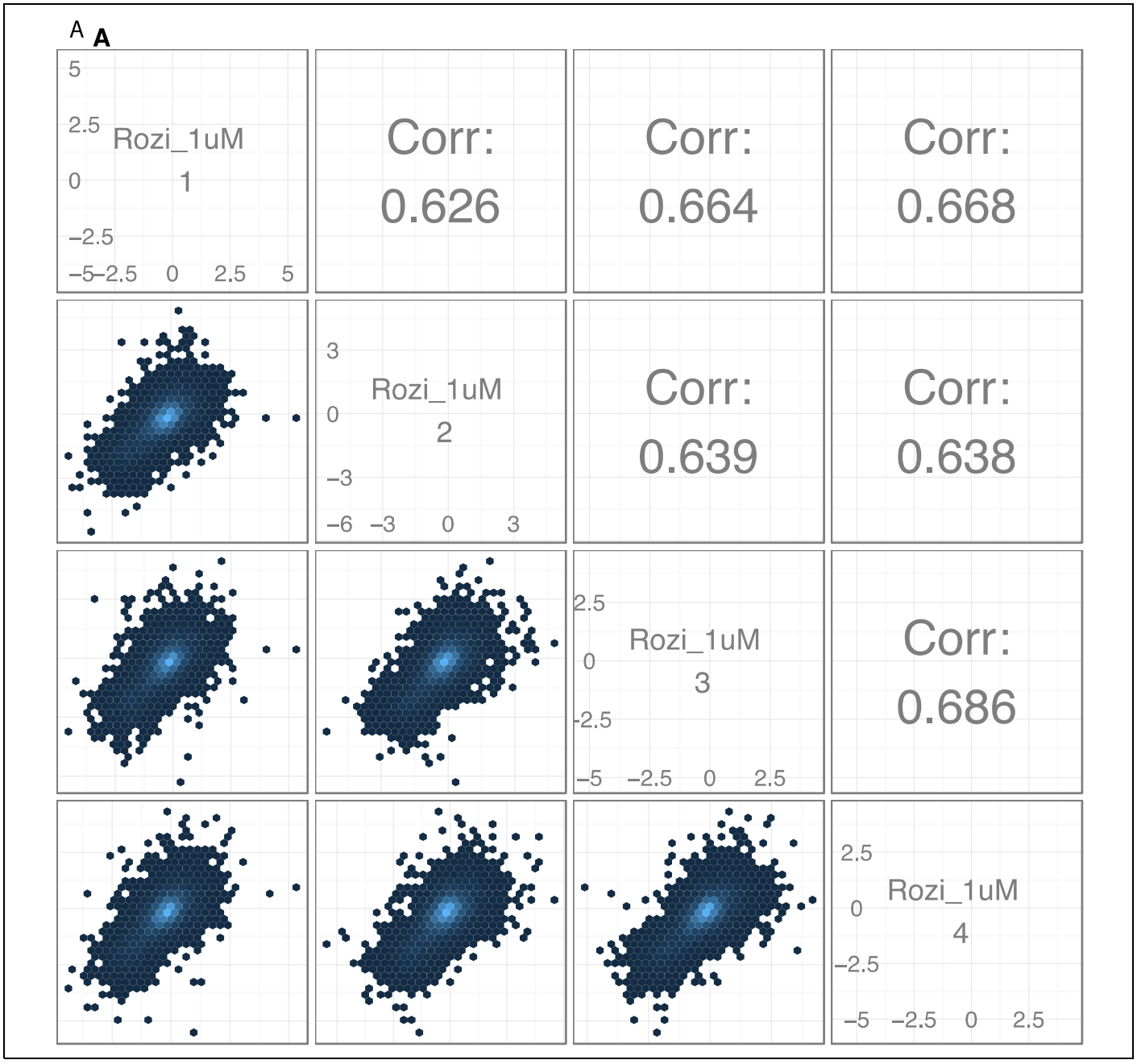
Endogenous *PPARG* was deleted in THP-1 monocytes by lentivirus-transduced CRISPR-Cas9. WT THP-1 cells were transduced with CRISPR-Cas9 and an sgRNA targeting either exon 6 of *PPARG* (NM_015869.4) or exon 8 of a control gene, PHACTR1. Samples of the resulting cell populations were collected at various time points. A) Percent of genomic DNA sequences with indels in *PPARG* over time. *PPARG* modification appears to saturate by the 10th day after introduction of CRISPR-Cas9. B) FABP4 gene expression assayed by nCounter in THP-1 cells stimulated with Rosiglitazone for 3 days. C) Cell surface CD36 protein expression in THP-1 cells stimulated with Rosiglitazone. Cells were treated with Rosiglitazone for 3 days, stained with PE-conjugated CD36 and subjected to flow cytometry. The response of WT cells is shown in the top panel and the response of PPARG CRISPR treated cells is shown in the bottom panel.

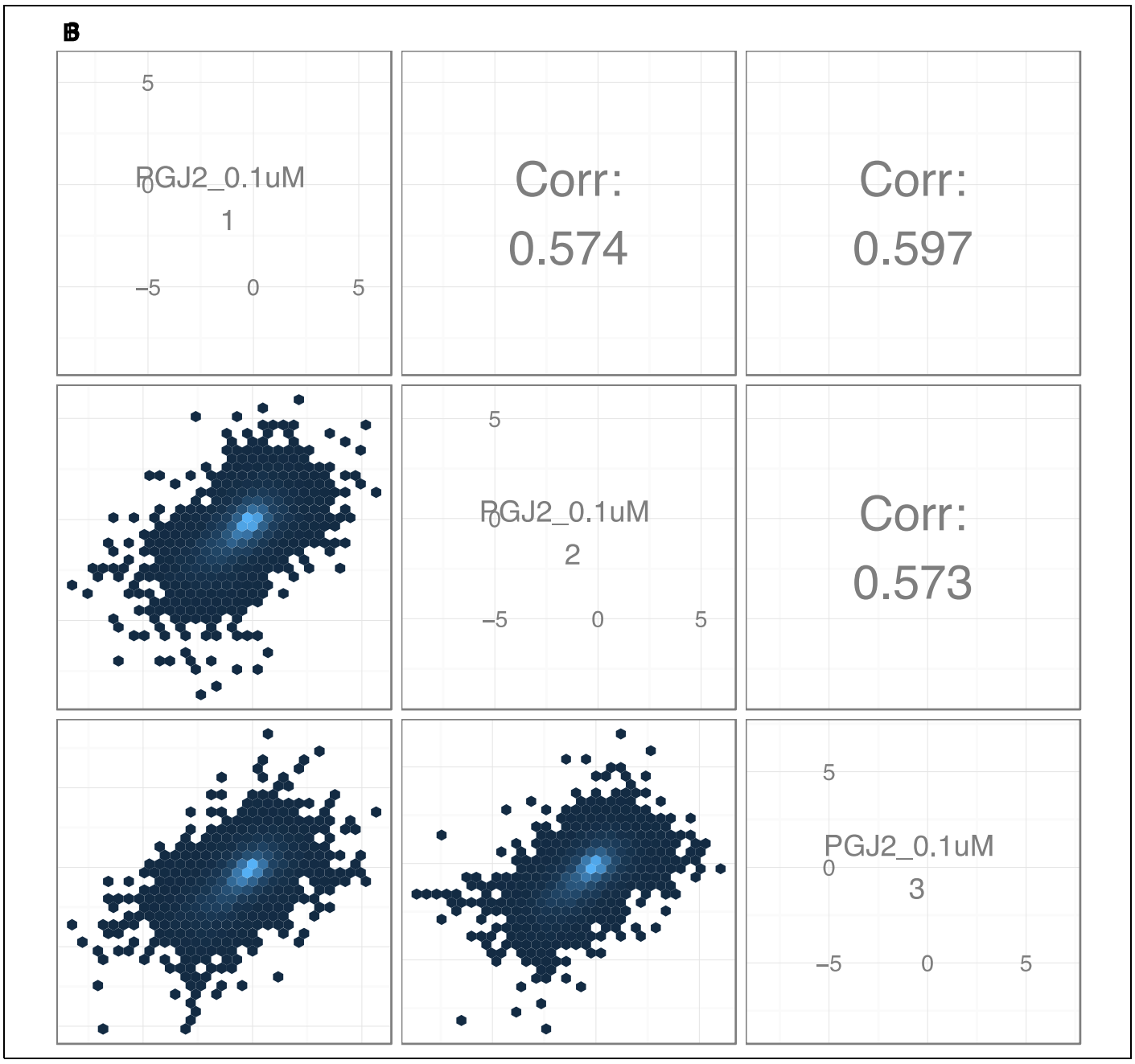


Supplementary Figure 3

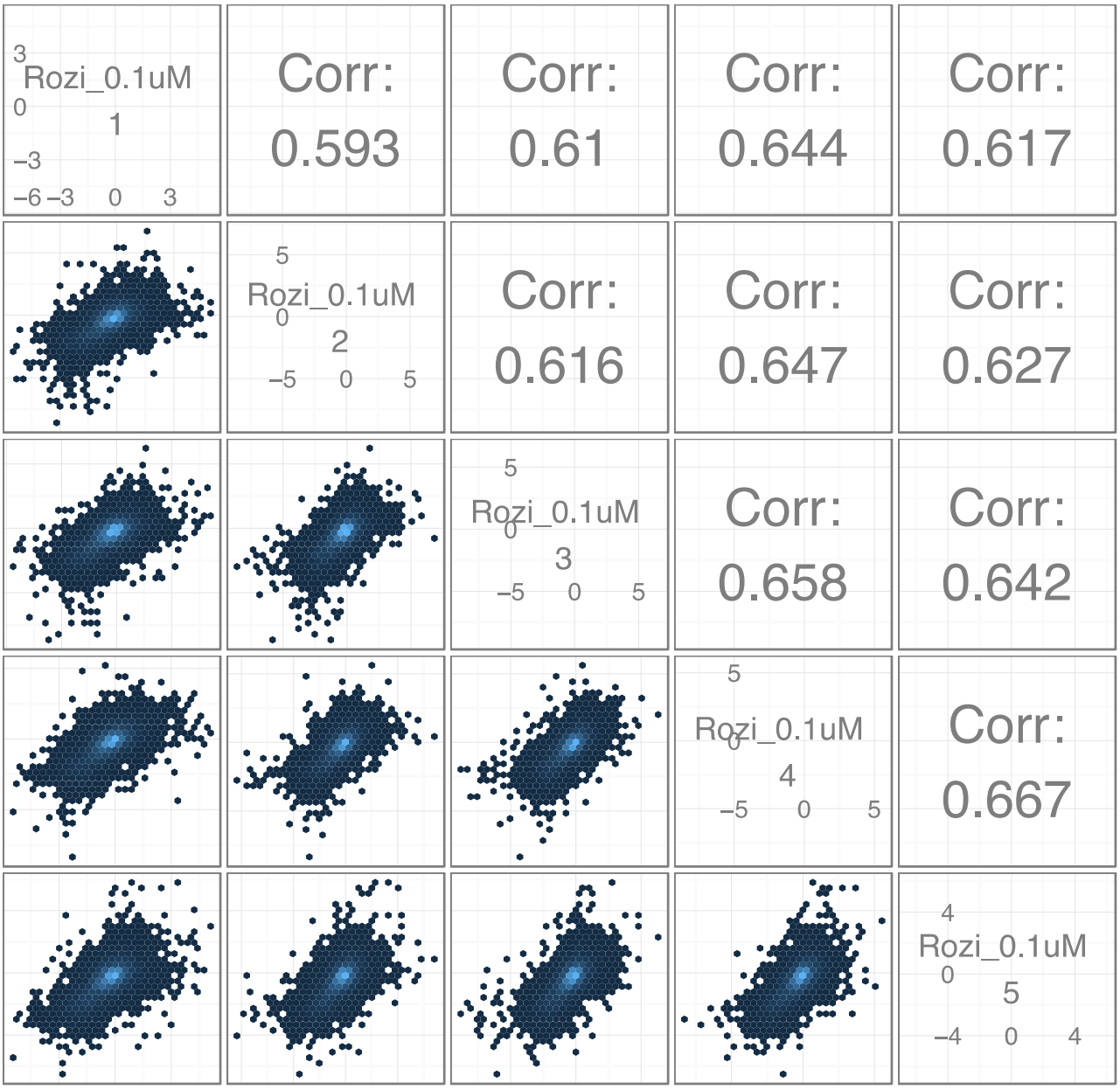
Sorting of PPARγ variants by CD36 activity

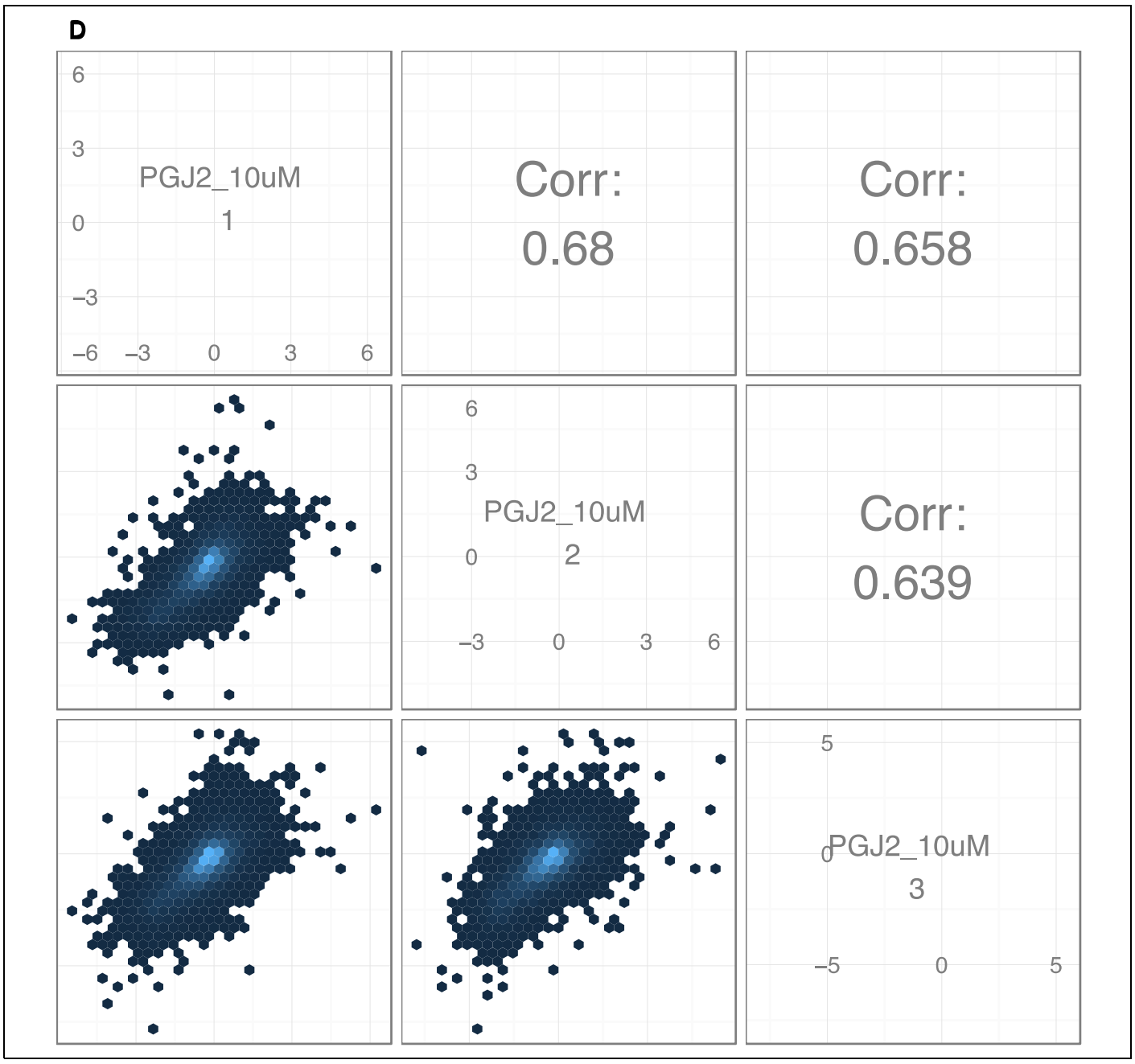
The library of 9595 PPARγ variant transgenes was introduced into a THP-1 cell line deleted for endogenous PPARγ. These cells expressed CD36 at a background level as shown (dashed red lines). Upon stimulation with PPARγ agonists (Rosiglitazone/PGJ2), a shift in the distribution of CD36 expression is seen (shaded blue lines). WT PPARγ was introduced into a separate population of PPARγ-/- THP-1 cells and stimulated with the same PPARγ agonists as above. The distribution of CD36 expression in cells bearing only WT PPARγ transgenes is shown (green dashed lines). The stimulated population of THP-1 cells containing the PPARγ transgene library was sorted by FACS into two subpopulations based on CD36 activity (red and green shaded areas). CD36 activity bins were selected to contain equal proportions of the shaded blue distribution and to be separated by a 5-10 fold difference in CD36 activity.


A
A
Rozi_1uM
1
Corr:
0.626
Corr:
0.664
Corr:
0.668
Rozi_1uM
2
Corr:
0.639
Corr:
0.638
Rozi_1uM
3
Corr:
0.686
Rozi_1uM
4


B
5
PGJ2_0.1uM
1
0
−5
0
5
Corr:
0.574
Corr:
0.597
5
PGJ2_0.1uM
2
0
−5
0
5
Corr:
0.573
5
PGJ2_0.1uM
3
0
−5
0
5

c


Rozi_0.1uM
Corr: 0.593
Corr: 0.61
Corr: 0.644
Corr: 0.617
Rozi_0.1uM
Corr: 0.616
Corr: 0.647
Corr: 0.627
Rozi_0.1uM
Corr: 0.658
Corr: 0.642
Rozi_0.1uM
Corr: 0.667
Rozi_0.1uM


D
PGJ2_10uM
1
Corr:
0.68
Corr:
0.658
PGJ2_10uM
2
Corr:
0.639
PGJ2_10uM
3

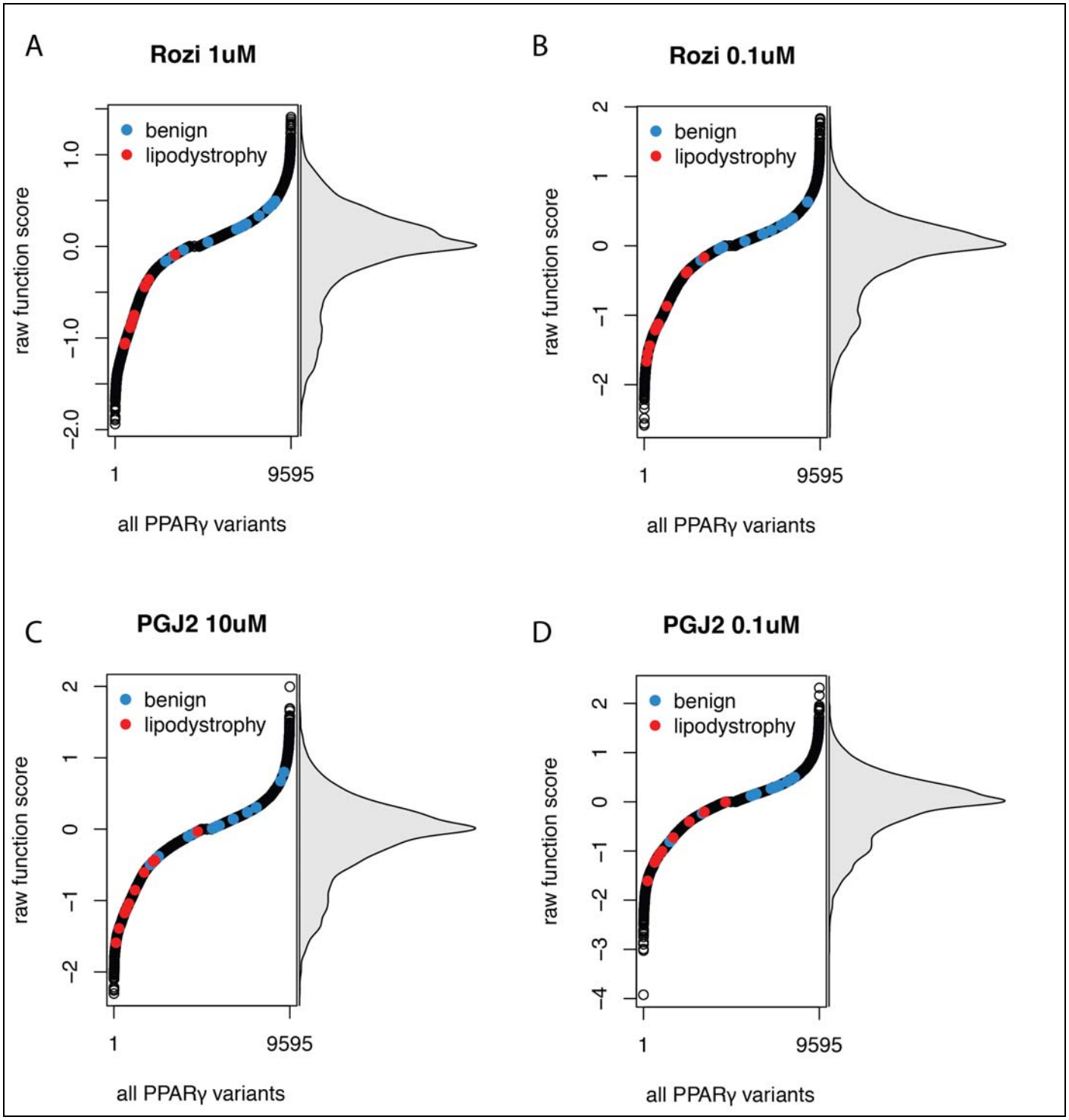
E

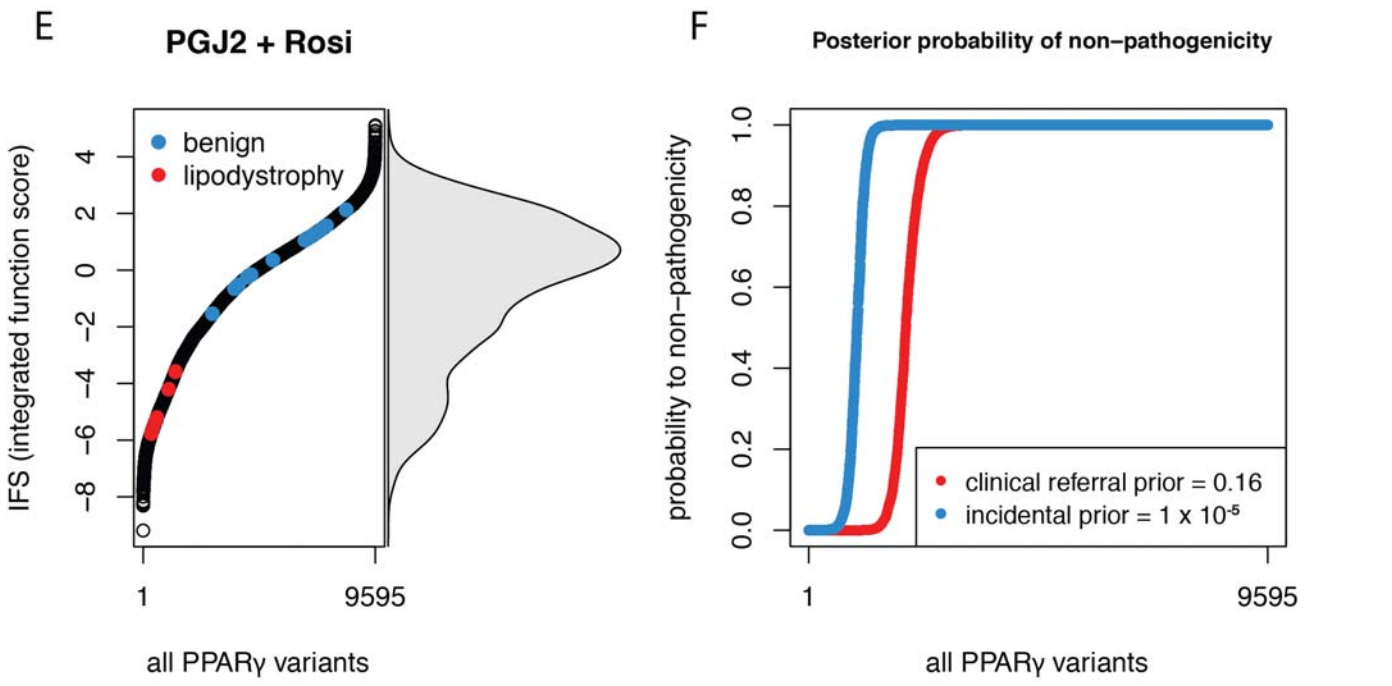

2
1
0
−1
−2
−3
−2
−1
0
1
2
Rozi 1uM, transduction 1
Corr:
0.746
1
0
−1
−2
−1
0
1
Rozi 1uM, transduction 2

Supplementary Figure 4

Replicate variability from simultaneous testing of 9,595 PPARG variants.

(A, B, C, D) Pairwise hexbin scatterplots and correlation coefficients of log2(CD36+/ CD36-) counts for FACS sorting, sequencing and counting of each PPARγ variant in the library stimulated with A) Rosiglitazone 1uM B) PGJ2 0.1uM C) Rosiglitazone 0.1uM and D) PGJ2 10uM. The elements of each row/column represent an independent sorting run into CD36+ and CD36- fractions performed on the same day. For each variant the log2 ratio of the coverage corrected counts in the CD36+ and CD36- bins is plotted. D) Pairwise hexbin scatterplot and correlation coefficient of raw function scores from two separate transductions of THP-1 cells with the PPARγ variant library, followed by Rosiglitazone stimulation, FACS sorting, sequencing and counting as above. The raw function scores plotted on each axis represent the synthesis of multiple sorting runs (A: all panels) (see Methods) and correspond to log2(CD36+/CD36-) counts for a single sorting run (A: single panel).


A
Rozi 1uM
benign
lipodystrophy
raw function score
1.0
0.0
−1.0
−2.0
1
9595
all PPARγ variants
B
Rozi 0.1uM
benign
lipodystrophy
raw function score
2
1
0
−1
−2
1
9595
all PPARγ variants
C
PGJ2 10uM
benign
lipodystrophy
raw function score
2
1
0
−1
−2
1
9595
all PPARγ variants
D
PGJ2 0.1uM
benign
lipodystrophy
raw function score
2
1
0
−1
−2
−3
−4
1
9595
all PPARγ variants



Supplementary Figure 5

Distribution of PPARγ function scores by experimental condition and following integration with clinically relevant priors

Raw PPARγ function scores were obtained for 9595 variants under four experimental conditions: A) 1uM Rosiglitazone, B) 0.1uM Rosiglitazone, C) 10uM Prostaglandin J2, and D) 0.1uM Prostaglandin J2. The overall distribution of scores is overlaid to the right and the function of known benign and lipodystrophy causing variants are highlighted in green and red respectively. E) Integrated functional scores (IFS) after classifier training using iterative linear discriminant analysis. F) Posterior probability of non-pathogenicity of 9595 PPARγ variants. This was calculated by combining IFS with prevalence of lipodystrophy in the general population (1:100,000) or from patients referred for lipodystrophy/familial diabetes (1:7).

Click inside this box and insert a single image for Supplementary Figure 6. For best results, use Insert menu to select a saved file; do not paste images. Source images must be JPEGs (no larger than 10 MB) saved in RBG color profile, at a resolution of 150–300 dpi. Optimize panel arrangement to a 2:3 height-to-width ratio; maximum online display is 600h x 900w pixels. Reduce empty space between panels and around image. Keep each image to a single page.

Delete these instructions before inserting the image.

Supplementary Figure 6

Insert figure title here by deleting or overwriting this text; keep title to a single sentence; use Symbol font for symbols and Greek letters.

Insert figure caption here by deleting or overwriting this text; captions may run to a second page if necessary. To ensure accurate appearance in the published version, please use the Symbol font for all symbols and Greek letters.

Click inside this box and insert a single image for Supplementary Figure 7. For best results, use Insert menu to select a saved file; do not paste images. Source images must be JPEGs (no larger than 10 MB) saved in RBG color profile, at a resolution of 150–300 dpi. Optimize panel arrangement to a 2:3 height-to-width ratio; maximum online display is 600h x 900w pixels. Reduce empty space between panels and around image. Keep each image to a single page.

Delete these instructions before inserting the image.

Supplementary Figure 7

Insert figure title here by deleting or overwriting this text; keep title to a single sentence; use Symbol font for symbols and Greek letters.

Insert figure caption here by deleting or overwriting this text; captions may run to a second page if necessary. To ensure accurate appearance in the published version, please use the Symbol font for all symbols and Greek letters.

Click inside this box and insert a single image for Supplementary Figure 8. For best results, use Insert menu to select a saved file; do not paste images. Source images must be JPEGs (no larger than 10 MB) saved in RBG color profile, at a resolution of 150–300 dpi. Optimize panel arrangement to a 2:3 height-to-width ratio; maximum online display is 600h x 900w pixels. Reduce empty space between panels and around image. Keep each image to a single page.

Delete these instructions before inserting the image.

Supplementary Figure 8

Insert figure title here by deleting or overwriting this text; keep title to a single sentence; use Symbol font for symbols and Greek letters.

Insert figure caption here by deleting or overwriting this text; captions may run to a second page if necessary. To ensure accurate appearance in the published version, please use the Symbol font for all symbols and Greek letters.