

Exome array analysis identifies *GPR35* as a novel susceptibility gene for anthracycline-induced cardiotoxicity in childhood cancer

Sara Ruiz-Pinto¹, Guillermo Pita¹, Ana Patiño-García, PhD ², Purificación García-Miguel, MD³, Javier Alonso, MD⁴, Antonio Pérez-Martínez, MD, PhD³, Antonio J Cartón, MD, PhD ⁵, Federico Gutiérrez-Larraya, MD, PhD ⁵, María R Alonso¹, Daniel R. Barnes, PhD⁶, Joe Dennis⁷, Kyriaki Michailidou, PhD ^{6,8}, Carmen Gómez-Santos⁹, Deborah J. Thompson, PhD ⁷, Douglas F. Easton, PhD ^{6,7}, Javier Benítez, PhD ^{1,10}, Anna González-Neira, PhD ¹

1 Human Genotyping Unit-CeGen, Human Cancer Genetics Programme. Spanish National Cancer Research Centre (CNIO), Madrid, 28029, Spain

2 Department of Pediatrics, Universidad de Navarra, University Clinic of Navarra, Pamplona, 31008, Spain

3 Department of Pediatric Hemato-Oncology, Hospital Universitario La Paz, Madrid, 28046, Spain

4 Pediatric solid tumor laboratory. Human Genetic Department. Research Institute of Rare Diseases. Instituto de Salud Carlos III, Majadahonda, 28220, Madrid, Spain

5 Department of Pediatric Cardiology, Hospital Universitario La Paz, Madrid, 28046, Spain

6 Department of Public Health and Primary Care, Centre for Cancer Genetic Epidemiology, Cambridge, CB1 8RN, UK

7 Department of Oncology, Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge, CB1 8RN, UK

8 Department of Electron Microscopy/Molecular Pathology, Cyprus Institute of Neurology and Genetics, Nicosia, 1683, Cyprus

9 Department of Pediatrics, Hospital Universitario Infanta Elena, 28342, Madrid, Spain

10 Human Genetics Group, Human Cancer Genetics Programme, Spanish National Cancer Research Centre (CNIO), Madrid, 28029, Spain

CORRESPONDING AUTHOR INFORMATION

Dr Anna González-Neira

Human Genotyping Unit-CeGen, Human Cancer Genetics Programme, Spanish National Cancer Centre, Melchor Fernández Almagro 3, Madrid 28029, Spain.

E-mail: agonzalez@cnio.es

Phone: +3491 2246974

Fax: +3491 2246923

RUNNING TITLE

Biomarkers for anthracycline-induced cardiotoxicity

ABSTRACT

Pediatric cancer survivors are a steadily growing population; however chronic anthracycline-induced cardiotoxicity (AIC) is a serious long-term complication leading to substantial morbidity. We aimed to identify new genes and low-frequency variants influencing the susceptibility to AIC for pediatric cancer patients. We studied the association of variants on the Illumina HumanExome BeadChip array in 93 anthracycline-treated pediatric cancer patients. In addition to single-variant association tests we carried out a gene-based analysis to investigate the joint effects of common and low-frequency variants on chronic AIC. Although no single-variant showed an association with chronic AIC which was statistically significant after correction for multiple testing, we identified a novel significant association for *GPR35* (G protein-coupled receptor 35) by gene-based testing, a gene with potential roles in cardiac physiology and pathology ($P=4.0 \times 10^{-6}$, $P_{\text{FDR}}=0.02$). The greatest contribution to this observed association was made by rs12468485, a missense variant (p.Thr253Met, c.758C>T, minor allele frequency=0.04), the T allele being associated with increased risk of chronic AIC and more severe symptomatic cardiac manifestations at low anthracycline doses. This study reveals *GPR35* as a novel susceptibility gene associated with chronic AIC in pediatric cancer patients and the results emphasize the need for very large studies to reveal additional susceptibility variants.

INTRODUCTION

Anthracycline chemotherapeutic agents are highly effective and widely used to treat childhood cancers¹, but their clinical use is compromised by cardiotoxicity, especially for late-onset disease^{2,3}. Children typically develop chronic dilated cardiomyopathy, which can progress with time to congestive heart failure and eventual demise (up to 7.3% of patients after 30-years)^{3,4}, although survivors treated with anthracyclines can be asymptomatic for several years (up to 57% of patients)⁵.

Several studies⁶⁻²¹ have identified genetic variants associated with anthracycline-induced cardiotoxicity (AIC). Most have been hypothesis-driven candidate-gene studies and have focused on common genetic variants (minor allele frequency (MAF) $\geq 5\%$). Still, a substantial portion of genetic susceptibility to AIC remains unexplained. There is increasing evidence that low-frequency variants (MAF $< 5\%$) also play an important role in the genetic architecture of complex traits²², but the role of this type of variation in AIC has not been explored thus far.

We performed a genome-wide association analysis of 93 anthracycline-treated pediatric cancer patients using the Illumina HumanExome Beadchip, which is enriched for low-frequency coding variants ($>80\%$ variants with MAF $\leq 1\%$)²³. Due to the low power to detect individual low-frequency AIC susceptibility alleles we explored the joint associations of variants using gene-based tests. We also performed gene-enrichment and pathway analyses.

MATERIALS AND METHODS

Patients

283 anthracycline-treated pediatric cancer patients at the *La Paz* University Hospital and the *Niño Jesús* University Hospital in Madrid and at the University Clinic of Navarra in Pamplona were reviewed between 2010 and 2014 and relevant clinical information was abstracted from medical records (Supplementary Table 4). All patients were treated with doxorubicin, daunorubicin or epirubicin as part of their chemotherapy protocol.

AIC was defined as early-onset (occurring within 1 year after anthracycline treatment completion) or late-onset (occurring >1 year after anthracycline therapy completion) chronic left ventricle (LV) dysfunction assessed by echocardiogram measurements and evidenced by shortening fraction (SF) or symptoms/signs of severe mitral valve insufficiency, pericardial effusion, left ventricular hypertrophy or pulmonary hypertension. The criteria for determining a symptomatic event were established by pediatric cardiologists. To avoid inter-observer and intra-observer variability in echocardiographic evaluations and to better differentiate between cases and controls, we defined as cases patients with $SF \leq 27\%$ at any time after anthracycline treatment completion. Controls were patients who had no symptoms or signs of cardiac complications and had normal echocardiograms ($SF \geq 35\%$) during and after therapy. To rule out acute AIC, only echocardiograms obtained 30 days or more after an anthracycline dose were considered. All pediatric cancer patients had normal cardiac function before commencing anthracycline therapy. 93 out of 283 (33%) patients fulfilled the above-mentioned criteria.

Written informed consent was obtained from adult patients and the parents or legal guardians of children. The study was approved by the ethics committees of all participating universities and hospitals.

Methods

To determine the role of low-frequency coding variants, we genotyped 93 anthracycline-treated pediatric patients for the 247,870 variants on the Illumina HumanExome Beadchip and we evaluated single-variant associations with risk of chronic AIC and also performed gene-based testing and gene-enrichment and pathway analysis. Details of genotyping, statistical analyses and *in silico* prediction are provided in the **Supplementary Information**.

RESULTS

The demographic and clinical characteristics of the 93 anthracycline-treated pediatric cancer patients are shown in Table 1. Controls were significantly younger than cases at diagnosis (median age 5.1 and 10.4 years, respectively, $P=0.004$). Cumulative anthracycline dose was higher in cases than in controls (360 mg/m^2 v 130 mg/m^2 $P<0.001$), with doxorubicin being the most frequent anthracycline drug administered. There were fewer cases than controls diagnosed with leukemia (37% v 88%; $P<0.001$) but more with pediatric bone tumors (34% v 5.2% with osteosarcoma; $P<0.001$ and 29% v 6.9% with Ewing sarcoma; $P=0.007$). Regarding concomitant therapy, bleomycin was more often administered in cases than in controls (37% v 3.4%; $P<0.001$).

Three patients (1 case and 2 controls) failed genotyping (call rate <0.95) and 7 patients (4 cases and 3 controls) were excluded as ethnic outliers based on

inspection of plots of the two first principal components, leaving 83 patients for further analysis (Supplementary Figure 1).

Single-variant associations

Of the 247,870 variants on the array, 246,060 passed quality control and 53,136 were polymorphic. After adjusting for age at diagnosis and cumulative anthracycline dose, none of these variants reached statistical significance after correcting for multiple testing ($P_{\text{FDR}} > 0.05$).

Gene-based test

We carried out gene-based analysis to further investigate the role of common and low-frequency variants in chronic AIC using the optimized sequence kernel association test (SKAT-O)^{24,25}. We identified *GPR35* (G protein-coupled receptor 35) as the gene most significantly associated with chronic AIC ($P = 4.0 \times 10^{-6}$) and this association remained statistically significant after correction for multiple testing (corrected $P_{\text{FDR}} = 0.02$). Sensitivity analyses revealed variant rs12468485 (p.Thr253Met; c.758C>T) (MAF=0.04) made the greatest contribution to the observed association (Figure 1). The minor T allele of this variant was almost exclusively present in cases (MAF_{CASES}=0.09 v MAF_{CONTROLS}=0.009), with only one CT carrier among controls; no TT carriers were found in our series of anthracycline-treated pediatric cancer patients. The majority (67%) of cases carrying the CT genotype had an extreme chronic AIC phenotype: LV dysfunction, mostly symptomatic, evidenced after treatment with anthracycline doses well below the average for cases (CT cases=155mg/m² v all other cases=360 mg/m²). To assess whether the model with the variant rs12468485 was more informative than the model with only non-genetic

variables (age at diagnosis and cumulative anthracycline dose) we used likelihood ratio tests. We obtained that the model including the variant rs12468485 and clinical factors provides a significant improvement over the model with only clinical variables ($P=8.6 \times 10^{-5}$).

In order to evaluate the impact of the missense variant rs12468485 (p.Thr253Met) on *GPR35* protein structure or function, we applied six in silico prediction algorithms (Supplementary Table 1). p.Thr253Met was classified as pathogenic by PolyPhen-2 and MutPred, and was predicted by F-SNPs to have a potentially regulatory role in splicing (Supplementary Table 1).

***GPR35* sequencing**

Due to the incomplete coverage in the exome array of coding variants in *GPR35*, we sequenced the exonic region of the gene in our series of anthracycline-treated pediatric cancer patients. We identified 17 coding variants, 6 of which had been genotyped on the exome array. Of the other 11, 2 were in complete linkage disequilibrium with the variant rs12468485 and 2 had call rate < 0.90 and were not analysed further. Of the remaining 7 coding variants identified ($r^2 < 0.36$ with variant rs12468485) in *GPR35* (Supplementary Table 2), 4 were synonymous and 3 missense, and all had MAF < 5%.

Of these 7 new *GPR35* variants, only rs35155396 was associated with risk independently of clinical factors ($P=5.16 \times 10^{-3}$), but not independently of rs12468485 ($P=0.99$).

Gene-enrichment and pathway analysis

To gain further insight into the nature of the biological pathways impacting on AIC we performed a gene enrichment analysis using the bioinformatics tool DAVID based on the list of significant genes ($P < 0.05$) with at least 3 variants identified in the SKAT-O analysis. Ten clusters with an enrichment score (ES) ≥ 1.3 (indicating biological significance) were found (Supplementary Table 3). The cluster with the highest ES (3.1) revealed overrepresentation of glycoproteins, receptors, N-linked glycosylation sites and cell and plasma membrane components. No pathways were associated with AIC risk after multiple testing correction ($P \geq 0.05$).

DISCUSSION

Despite the demonstrated therapeutic effectiveness of anthracyclines in cancer treatment²⁶, AIC continues to be a serious problem in survivors long after their treatments have ended and cancer has been cured^{1,2} and clinicians remain unable to accurately stratify patients into high or low-risk groups^{3,4}. Considering that children are particularly vulnerable to the cardiotoxic effect of anthracyclines, even more than adults and/or at lower doses^{3,5}, the identification of biomarkers of AIC risk in this specific group of patients seems crucial to maximize clinical benefit and minimize harm.

In recent decades, candidate-gene studies and genome-wide association studies (GWAS) have been extensively applied to dissect the common genetic architecture of many complex traits^{27,28}. Candidate-gene studies are limited by our understanding of the underlying molecular and biological mechanisms, and common variants discovered in GWAS studies explain only a small proportion

of the total heritability of complex traits²⁹. New evidence has emerged over the past few years that additional heritability might be explained by low-frequency variants^{22,30}, and in particular coding variants, which *a priori* are more likely to have larger impact on protein function³¹. Here, we have explored the contribution of low-frequency coding variants to susceptibility to AIC in children. To the best of our knowledge, this is the first study where this type of variation has been comprehensively analyzed using exome array data for children treated with anthracyclines and with long-term follow-up.

Using standard single-variant association tests, none of the 246,060 variants analyzed was found to be statistically significantly associated with chronic AIC after correction for multiple comparisons; however, this result may be explained by low statistical power, due to small sample size and few patient cases³². We performed gene-based tests that have greater statistical power to detect associations with rare variation and can evaluate the cumulative effect of multiple genetic variants within a gene³³. Some genes/regions may have a high proportion of causal variants and influence the phenotype in the same direction while others may have few causal variants or the causal variants may have different directions of association. Therefore, the use of methods optimal for both scenarios, such as the combined gene-based test SKAT-O applied in this study, is desirable²⁴. Using this approach, we have identified *GPR35* as the gene most significantly associated with chronic AIC in children. While the SKAT-O does not provide any parameter estimates, we assessed the individual contribution of variants within *GPR35* and found that rs12468485, a missense low-frequency variant (p.Thr253Met), strongly associated with the most severe

cardiac manifestations, was most influential. Although in silico prediction algorithms yielded discrepant results, rs12468485 was predicted to have a potentially regulatory role in splicing. Due to the incomplete coverage of coding variants on the exome array, we sequenced the coding region of *GPR35* in our series of 93 anthracycline-treated pediatric cancer patients but we were unable to identify additional independent association signals at the gene.

GPR35 belongs to the G-protein-coupled receptor family, which are membrane proteins mediating a wide range of physiological processes³⁴. Although the exact functions of *GPR35* are not known, several lines of evidence strongly suggest potential roles for this receptor in cardiac physiology and pathology. Sun et al³⁵ were the first to report a cardiovascular role for *GPR35*, with a non-synonymous single nucleotide polymorphism (rs3749172) significantly associated with the burden of coronary artery calcification. *GPR35* was later found to be up-regulated in failing myocardium of patients with severe chronic heart, and in the same study *GPR35* knock-out mice showed higher systolic blood³⁶. More recently, *GPR35* has been characterized as an early marker of progressive cardiac failure³⁷. In vitro functional assays in cardiomyocytes demonstrated that *GPR35* overexpression reduced cell viability and promoted morphological changes^{36,37}. Several studies have linked *GPR35* to inflammatory regulation^{38,39} and there is ample evidence to support the hypothesis that inflammation, as well as hypoxia, play a significant role in the pathogenesis and development of chronic cardiac complications, including cardiomyopathy^{40,41}. These findings provide a possible explanation for the involvement of this receptor in cardiovascular disease. On the other hand, the cellular and

biological effects of *GPR35* on cardiovascular pathophysiology could be largely mediated by downstream signaling pathways such as $G\alpha_{13}$, $G\alpha_{i/o}$ and RhoA, following receptor activation^{42,43}. Putative endogenous ligands of *GPR35* have also been linked to cardiovascular disease in animals models^{44,45} and in patients with chronic advanced heart failure⁴⁶ or acute myocardial infarction⁴⁷.

In conclusion, apart from the well-established clinical risk factors for AIC, such as age at diagnosis and cumulative anthracycline dose, *GPR35*, a gene with important roles in cardiac physiology and pathology, and in particular rs12468485, appear to be an independent risk factor for chronic AIC. We demonstrated that the inclusion of this variant significantly improved the risk prediction model and may enhance the ability of physicians to identify high-risk patients in clinical practice. However, further replication of the association and resequencing of the gene in larger cohorts of patients to identify additional *GPR35* rare variants associated with risk of AIC are required.

ACKNOWLEDGEMENTS

This work was supported by the Spanish Association against Cancer (AECC: Asociación Española contra el Cáncer). Human Genotyping lab is a member of CeGen, PRB2-ISCI and is supported by grant PT13/0001, of the PE I+D+i 2013-2016, funded by ISCI and FEDER (Fondo Europeo de Desarrollo Regional). Sara Ruiz-Pinto is a predoctoral fellow supported by the Severo Ochoa Excellence Programme (Project SEV-2011-0191).

CONFLICT OF INTEREST STATEMENT

The authors indicated no potential conflicts of interest.

Supplementary information is available at The Pharmacogenomics Journal's website.

REFERENCES

1. Smith LA, Cornelius VR, Plummer CJ, Levitt G, Verrill M, Canney P, *et al.* Cardiotoxicity of anthracycline agents for the treatment of cancer: systematic review and meta-analysis of randomised controlled trials. *BMC Cancer* **10**: 337.
2. Trachtenberg BH, Landy DC, Franco VI, Henkel JM, Pearson EJ, Miller TL, *et al.* Anthracycline-associated cardiotoxicity in survivors of childhood cancer. *Pediatr Cardiol* **32**(3): 342-353.
3. Lipshultz SE, Lipsitz SR, Sallan SE, Dalton VM, Mone SM, Gelber RD, *et al.* Chronic progressive cardiac dysfunction years after doxorubicin therapy for childhood acute lymphoblastic leukemia. *J Clin Oncol* 2005; **23**(12): 2629-2636.
4. van der Pal HJ, van Dalen EC, van Delden E, van Dijk IW, Kok WE, Geskus RB, *et al.* High risk of symptomatic cardiac events in childhood cancer survivors. *J Clin Oncol* **30**(13): 1429-1437.
5. Lipshultz SE, Colan SD, Gelber RD, Perez-Atayde AR, Sallan SE, Sanders SP. Late cardiac effects of doxorubicin therapy for acute

- lymphoblastic leukemia in childhood. *N Engl J Med* 1991; **324**(12): 808-815.
6. Wojnowski L, Kulle B, Schirmer M, Schluter G, Schmidt A, Rosenberger A, *et al.* NAD(P)H oxidase and multidrug resistance protein genetic polymorphisms are associated with doxorubicin-induced cardiotoxicity. *Circulation* 2005; **112**(24): 3754-3762.
 7. Blanco JG, Leisenring WM, Gonzalez-Covarrubias VM, Kawashima TI, Davies SM, Relling MV, *et al.* Genetic polymorphisms in the carbonyl reductase 3 gene CBR3 and the NAD(P)H:quinone oxidoreductase 1 gene NQO1 in patients who developed anthracycline-related congestive heart failure after childhood cancer. *Cancer* 2008; **112**(12): 2789-2795.
 8. Rajic V, Aplenc R, Debeljak M, Prestor VV, Karas-Kuzelicki N, Mlinaric-Rascan I, *et al.* Influence of the polymorphism in candidate genes on late cardiac damage in patients treated due to acute leukemia in childhood. *Leuk Lymphoma* 2009; **50**(10): 1693-1698.
 9. Blanco JG, Sun CL, Landier W, Chen L, Esparza-Duran D, Leisenring W, *et al.* Anthracycline-related cardiomyopathy after childhood cancer: role of polymorphisms in carbonyl reductase genes--a report from the Children's Oncology Group. *J Clin Oncol* **30**(13): 1415-1421.
 10. Lipshultz SE, Lipsitz SR, Kutok JL, Miller TL, Colan SD, Neuberg DS, *et al.* Impact of hemochromatosis gene mutations on cardiac status in doxorubicin-treated survivors of childhood high-risk leukemia. *Cancer* **119**(19): 3555-3562.

11. Semsei AF, Erdelyi DJ, Ungvari I, Csagoly E, Hegyi MZ, Kiszal PS, *et al.* ABCC1 polymorphisms in anthracycline-induced cardiotoxicity in childhood acute lymphoblastic leukaemia. *Cell Biol Int* **36**(1): 79-86.
12. Cascales A, Sanchez-Vega B, Navarro N, Pastor-Quirante F, Corral J, Vicente V, *et al.* Clinical and genetic determinants of anthracycline-induced cardiac iron accumulation. *Int J Cardiol* **154**(3): 282-286.
13. Cascales A, Pastor-Quirante F, Sanchez-Vega B, Luengo-Gil G, Corral J, Ortuno-Pacheco G, *et al.* Association of anthracycline-related cardiac histological lesions with NADPH oxidase functional polymorphisms. *Oncologist* **18**(4): 446-453.
14. Volkan-Salanci B, Aksoy H, Kiratli PÖ, Tülümen E, Güler N, Öksüzoglu B, *et al.* The relationship between changes in functional cardiac parameters following anthracycline therapy and carbonyl reductase 3 and glutathione S transferase Pi polymorphisms. *J. Chemother*, 2012; **24**: 285–291.
15. Visscher H, Ross CJ, Rassekh SR, Barhdadi A, Dube MP, Al-Saloos H, *et al.* Pharmacogenomic prediction of anthracycline-induced cardiotoxicity in children. *J Clin Oncol* **30**(13): 1422-1428.
16. Visscher H, Ross CJ, Rassekh SR, Sandor GS, Caron HN, van Dalen EC, *et al.* Validation of variants in SLC28A3 and UGT1A6 as genetic markers predictive of anthracycline-induced cardiotoxicity in children. *Pediatr Blood Cancer* **60**(8): 1375-1381.

17. Lubieniecka JM, Graham J, Heffner D, Mottus R, Reid R, Hogge D, *et al.*
A discovery study of daunorubicin induced cardiotoxicity in a sample of acute myeloid leukemia patients prioritizes P450 oxidoreductase polymorphisms as a potential risk factor. *Front Genet* **4**: 231.
18. Wang X, Liu W, Sun CL, Armenian SH, Hakonarson H, Hageman L, *et al.*
Hyaluronan synthase 3 variant and anthracycline-related cardiomyopathy: a report from the children's oncology group. *J Clin Oncol* **32**(7): 647-653.
19. Krajinovic M, Elbared J, Drouin S, Bertout L, Rezgui A, Ansari M, *et al.*
Polymorphisms of ABCC5 and NOS3 genes influence doxorubicin cardiotoxicity in survivors of childhood acute lymphoblastic leukemia. *Pharmacogenomics J.*
20. Visscher H, Rassekh SR, Sandor GS, Caron HN, van Dalen EC, Kremer LC, *et al.* Genetic variants in SLC22A17 and SLC22A7 are associated with anthracycline-induced cardiotoxicity in children. *Pharmacogenomics* **16**(10): 1065-1076.
21. Aminkeng F, Bhavsar AP, Visscher H, Rassekh SR, Li Y, Lee JW, *et al.*
A coding variant in RARG confers susceptibility to anthracycline-induced cardiotoxicity in childhood cancer. *Nat Genet* **47**(9): 1079-1084.
22. Cirulli ET, Goldstein DB. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat Rev Genet* **11**(6): 415-425.
23. http://genome.sph.umich.edu/wiki/Exome_Chip_Design.

24. Lee S, Wu MC, Lin X. Optimal tests for rare variant effects in sequencing association studies. *Biostatistics* **13**(4): 762-775.
25. Lee S, Emond MJ, Bamshad MJ, Barnes KC, Rieder MJ, Nickerson DA, *et al.* Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am J Hum Genet* **91**(2): 224-237.
26. Weiss, RB. The anthracyclines: will we ever find a better doxorubicin? *Semin Oncol* 1992, **19**: 670-686.
27. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. *Am J Hum Genet* **90**(1): 7-24.
28. Tabor HK, Risch NJ, Myers RM. Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nat Rev Genet* 2002; **3**(5): 391-397.
29. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ, *et al.* Finding the missing heritability of complex diseases. *Nature* 2009; **461**(7265): 747-753.
30. Rivas MA, Beaudoin M, Gardet A, Stevens C, Sharma Y, Zhang CK, *et al.* International Inflammatory Bowel Disease Genetics Consortium Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nat. Genet* 2011, **43**:1066–1073.
31. Gibson G. Rare and common variants: twenty arguments. *Nat Rev Genet* **13**(2): 135-145.

32. Li B, Leal SM. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am J Hum Genet* 2008; **83**(3): 311-321.
33. Lee S, Abecasis GR, Boehnke M, Lin X. Rare-variant association analysis: study designs and statistical tests. *Am J Hum Genet* **95**(1): 5-23.
34. Pierce KL, Premont RT, Lefkowitz RJ. Seven-transmembrane receptors. *Nat Rev Mol Cell Biol* 2002; **3**(9): 639-650.
35. Sun YV, Bielak LF, Peyser PA, Turner ST, Sheedy PF, 2nd, Boerwinkle E, *et al.* Application of machine learning algorithms to predict coronary artery calcification with a sibship-based design. *Genet Epidemiol* 2008; **32**(4): 350-360.
36. Min KD, Asakura M, Liao Y, Nakamaru K, Okazaki H, Takahashi T, *et al.* Identification of genes related to heart failure using global gene expression profiling of human failing myocardium. *Biochem Biophys Res Commun* **393**(1): 55-60.
37. Ronkainen VP, Tuomainen T, Huusko J, Laidinen S, Malinen M, Palvimö JJ, *et al.* Hypoxia-inducible factor 1-induced G protein-coupled receptor 35 expression is an early marker of progressive cardiac remodeling. *Cardiovasc Res* **101**(1): 69-77.
38. Wang J, Simonavicius N, Wum X, Swaminath G, Reagan J, Tian H, *et al.* Kynurenic acid as a ligand for orphan G protein-coupled receptor GPR35. *J Biol Chem* 2006, **281**: 22021–22028.
39. Barth MC, Ahluwalia N, Anderson TJ, Hardy GJ, Sinha S, Alvarez-Cardona JA, *et al.* Kynurenic acid triggers firm arrest of leukocytes to

- vascular endothelium under flow conditions. *J Biol Chem* 2009; **284**(29): 19189-19195.
40. Geft IL, Fishbein MC, Ninomiya K, Hashida J, Chauxm E, Yano J, *et al*. Intermittent brief period of ischemia have a cumulative effect and may cause myocardial necrosis. *Circulation* 1982, **66**:1150–1153.
41. Watanabe Y, Kusuoka H, Fukuchi K, Fujiwara T, Nishimura T. Contribution of hypoxia to the development of cardiomyopathy in hamsters. *Cardiovasc Res* 1997; **35**(2): 217-222.
42. Offermanns S, Mancino V, Revel JP, Simon MI. Vascular system defects and impaired cell chemokinesis as a result of Galpha13 deficiency. *Science* 1997; **275**(5299): 533-536.
43. Maruyama Y, Nishida M, Sugimoto Y, Tanabe S, Turner JH, Kozasa T, *et al*. Galpha(12/13) mediates alpha(1)-adrenergic receptor-induced cardiac hypertrophy. *Circ Res* 2002; **91**(10): 961-969.
44. Schober A, Siess W. Lysophosphatidic acid in atherosclerotic diseases. *Br J Pharmacol* **167**(3): 465-482.
45. Mizutani K, Sugimoto K, Okuda T, Katsuya T, Miyata T, Tanabe T, *et al*. Kynureninase is a novel candidate gene for hypertension in spontaneously hypertensive rats. *Hypertens. Res* 2002, **25**: 135–140.
46. Hamilton MA, Stevenson LW, Luu M, Walden JA. Altered thyroid hormone metabolism in advanced heart failure. *J Am Coll Cardiol* 1990; **16**(1): 91-95.
47. Friberg L, Werner S, Eggertsen G, Ahnve S. Rapid down-regulation of thyroid hormones in acute myocardial infarction: is it cardioprotective in patients with angina? *Arch Intern Med* 2002; **162**(12): 1388-1394.

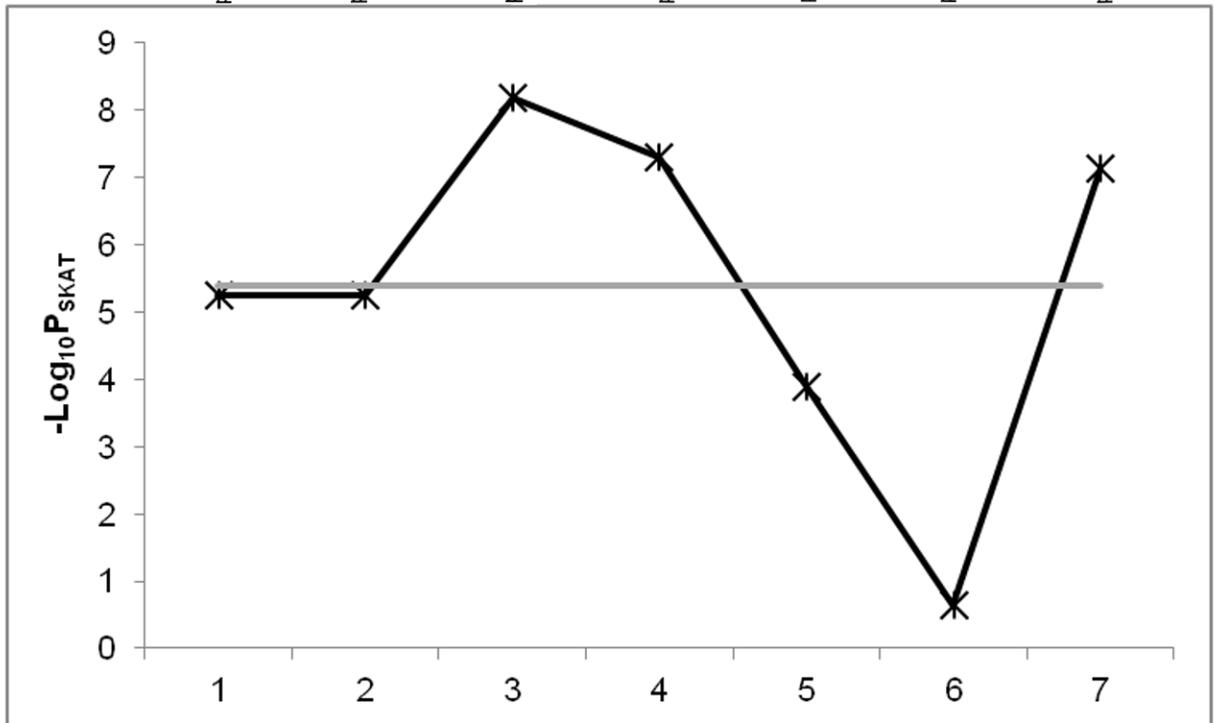
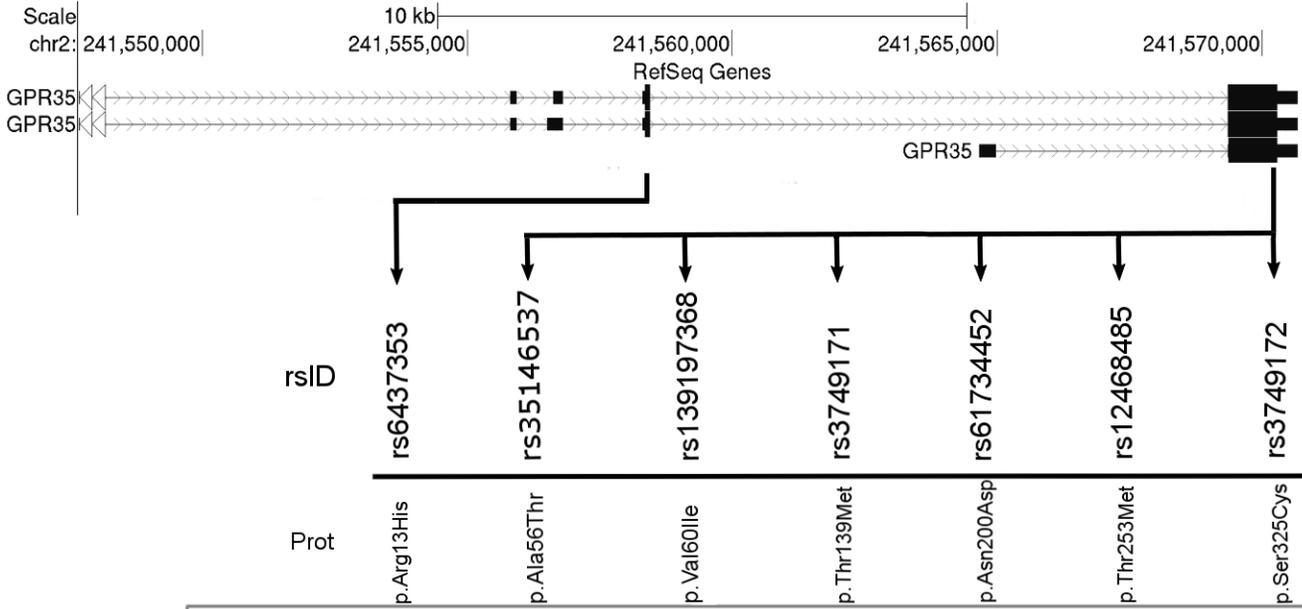
48.Children's Oncology Group Long-Term Follow-Up Guidelines for Survivors of Childhood, Adolescent, and Young Adult Cancers: www-survivorshipguidelines.org.

FIGURE LEGENDS

Figure 1. Contribution of individual *GPR35* variants on statistical significances for the *GPR35* gene. Top: genomic location of *GPR35* displayed in the UCSC Genome Browser. Exon location and amino acid substitution of each of the 7 coding polymorphic variants included in the Illumina HumanExome BeadChip array are depicted. Bottom: *P*-values for the *GPR35* association in SKAT-O gene-based tests after removing one variant (black line) at a time and recalculating the association. Grey line indicates the *P*-value for the *GPR35* association with chronic AIC including all 7 coding variants ($P=4.0 \times 10^{-6}$). Abbreviations: Prot, protein.

Characteristic	Controls (N=58)		Cases (N=35)		P
	N	%*	N	%*	
Age at diagnosis (years)					0.004
Median		5.1		10.4	
Range		1.4-16.9		1.2-21.1	
Gender					0.19
Female	25	43	10	29	
Male	33	57	25	71	
Primary diagnosis (tumor type)					
Leukemia	51	88	13	37	<0.001
Osteosarcoma	3	5.2	12	34	<0.001
Ewing Sarcoma	4	6.9	10	29	0.007
Family history of cardiovascular disease	3	5.2	4	17	0.21
Radiotherapy involving the heart ^a	3	6.7	6	19	0.15
Cumulative anthracycline dose (mg/m ²) ^b					<0.001
Mean		130		360	
Range		49.2-562		105-780	
≤ 200	43	74	8	23	
> 200	15	26	27	77	
Anthracycline type					
Doxorubicin	50	86	34	97	0.31
Daunorubicin	8	14	1	2.9	0.15
Epirubicin	5	8.6	-	-	0.15
Concomitant therapy					
Cyclophosphamide	55	95	27	77	0.06
Vincristine	54	93	32	91	0.42
Etoposide	14	24	13	37	0.24
Bleomycin	2	3.4	13	37	<0.001
Follow-up (years)					0.06
Median		8.3		10.5	
Range		1-24.1		1-27.5	

Age, cumulative anthracycline dose and follow-up were analyzed by Wilcoxon-Mann-Whitney U test. Gender, tumor type, family history of cardiovascular disease, radiotherapy involving the heart, anthracycline type and concomitant therapy were analyzed by Fisher's exact test. * Percentages are computed based on the total number of non-missing values. a Radiotherapy involving the heart includes mediastinal and mantle radiation and total body irradiation. b Cumulative anthracycline dose was calculated using doxorubicin equivalents. Bold fold indicates statistically significant P-values (P<0.05).



Supplementary Table 1. In silico prediction of the functional effect of rs12468485 (p.Thr253Met)

							F-SNP prediction			
Variant	SIFT prediction	Polyphen-2 prediction	MutPred prediction	SNPs&GO Prediction	PON-P2 prediction	PredictSNP prediction	ESEfinder prediction	ESRSearch prediction	PESX prediction	RESCUE_ESE prediction
rs12468485 (p.Thr253Met)	Tolerated	Possibly damaging	Pathogenic	Neutral	Neutral	Neutral	Changed	Changed	Changed	Changed

Bold type indicates a likely pathogenic effect or a change in splicing predicted by each in silico algorithm. Predictions are on GPR35 protein with Uniprot identifier Q9HC97

Supplementary Table 2. Additional *GPR35* coding variants identified by sequencing

Variant	Position	MAF	Function
rs138283952	chr2:241569810	0.007	Synonymous
rs142918765	chr2:241570284	0.007	Synonymous
rs147336244	chr2:241569669	0.006	Synonymous
rs34778053	chr2:241569742	0.01	Missense (p.Arg156Ser)
rs35155396	chr2:241569585	0.02	Synonymous
rs61734453	chr2:241569745	0.006	Missense (p.Gly157Arg)
rs763867971	chr2:241570132	0.007	Missense (p.Arg286Cys)

Chromosome positions are based on Genome Reference Consortium Human Build 37 (GRCh37/hg19). Abbreviations: MAF, minor allele frequency

Supplementary Table 3. Functional Annotation Clustering from the DAVID tool (Enrichment score ≥ 1.3).

Supplementary Table 3. Functional Annotation Clustering from the DAVID tool (Enrichment score ≥ 1.3).					
Annotation cluster 1		Enrichment score: 3.1			
Category	Term	Count	P	Genes	P-Benjamini
SP_PIR_KEYWORDS	Glycoprotein	125	9.2×10^{-7}	<i>FSTL4, OR8S1, GRIN3B, SLC7A4, MMRN2, LPHN2, OR4D2, CHRNA9, CD44, LRRC52, PI16, ODZ3, GRID1, CLCA1, OR10S1, SLCO4A1, SPARCL1, PTPRN2, PLXNB2, CDHR2, SLC22A20, HLA-C, PNPLA3, TNFAIP6, LRP11, OR8B4, CHGB, DST, FGFR4, PANX1, AMTN, OR2T1, KEL, ENPP3, ITGAM, CRB2, P2RY2, ENTPD7, BAI2, SFTPD, B3GNT3, EGF, GCNT1, ADAM28, EPB41, ITGA3, OR5AC2, GPR35, SLC4A11, P2RX3, NOTCH4, TSC2, OR51A7, UTP14C, COL20A1, CHRNG, ARSB, ARSD, CLSTN2, CLSTN3, TMEM161A, CDSN, FCRL3, RSPO4, C14ORF135, SMPDL3A, NMUR2, POMT1, OR6P1, ANO2, CTBS, USH2A, OR2AE1, OR10J1, NCR3, CD86, TNFRSF10C, CLECL1, TAS2R19, ZPBP2, GRM6, ERN1, KCNH8, C4ORF29, PLA2G3, WFIKKN2, LRRC32, CPN2, NUP214, C1QTNF6, FCN2, COL6A2, FUT3, UGT2A1, COL6A1, CCBP2, HRG, PLA2R1, DKKL1, SLC39A4, QSOX1, SOAT1, TMC6, MOGAT1, CES1, OR8G2, FBN1, FUCA2, MAN1C1, COL5A1, PLG, OR51M1, CDH13, PROM2, GPR110, SLC17A2, ANXA11, GPR113, MEP1B, GPR111, MEGF6, ABCC8, GFRA2, ABCC6, GPR116</i>	1.7×10^{-4}
UP_SEQ_FEATURE	N-linked glycosylation sites (GlcNAc...)	116	1.7×10^{-5}	<i>FSTL4, OR8S1, GRIN3B, SLC7A4, MMRN2, LPHN2, OR4D2, CHRNA9, CD44, LRRC52, PI16, ODZ3, GRID1, CLCA1, OR10S1, SPARCL1, SLCO4A1, PTPRN2, PLXNB2, CDHR2, SLC22A20, HLA-C, PNPLA3, TNFAIP6, LRP11, OR8B4, DST, FGFR4, PANX1, OR2T1, KEL, ENPP3, ITGAM, CRB2, P2RY2, ENTPD7, BAI2, SFTPD, B3GNT3, EGF, GCNT1, ADAM28, ITGA3, OR5AC2, GPR35, SLC4A11, P2RX3, NOTCH4, OR51A7, UTP14C, COL20A1, CHRNG, ARSB, ARSD, CLSTN2, CLSTN3, TMEM161A, CDSN, FCRL3, RSPO4, C14ORF135, SMPDL3A, NMUR2, POMT1, OR6P1, ANO2, CTBS, USH2A, OR2AE1,</i>	1.2×10^{-2}

				OR10J1, NCR3, CD86, CLECL1, TAS2R19, ZBPB2, GRM6, ERN1, KCNH8, C4ORF29, PLA2G3, WFIKKN2, LRRC32, CPN2, C1QTNF6, FCN2, COL6A2, FUT3, UGT2A1, COL6A1, CCBP2, HRG, PLA2R1, DKKL1, SLC39A4, QSOX1, TMC6, MOGAT1, CES1, OR8G2, FBN1, FUCA2, MAN1C1, PLG, OR51M1, CDH13, PROM2, GPR110, SLC17A2, GPR113, MEP1B, GPR111, MEGF6, ABCC8, GFRA2, ABCC6, GPR116	
SP_PIR_KEYWORDS	Receptor	55	3.3×10^{-5}	TRPV3, GRIN3B, OR8S1, FCRL3, LPHN2, OR4D2, CHRNA9, CD44, NMUR2, OR6P1, GRID1, OR10S1, PLXNB2, PTPRN2, PPARGC1A, OR2AE1, OR10J1, TRPM2, NCR3, TNFRSF10C, CD86, TAS2R19, LILRB4, GRM6, LRP11, GPR50, OR8B4, FGFR4, OR2T1, PAQR7, ITGAM, P2RY2, BAI2, CCBP2, PLA2R1, TRIP11, OR8G2, ITGA3, PTPN12, OR5AC2, OR51M1, GPR35, EPS8, GPR110, P2RX3, NOTCH4, GPR113, GPR111, MS4A10, ABL1, ABCC8, OR51A7, GFRA2, CHRNG, GPR116	2.5×10^3
GOTERM_CC_FAT	Plasma membrane	107	5.3×10^{-5}	LMO7, OR8S1, GRIN3B, LPHN2, OR4D2, CHRNA9, CD44, WNK4, GRID1, CLCA1, OR10S1, MYH1, PTPRN2, CDHR2, HLA-C, CTNNA3, LILRB4, GPR50, OR8B4, KBTBD10, DST, SNTG2, HDLBP, FGFR4, PANX1, AMTN, OR2T1, KEL, ENPP3, MAP4K2, PAQR7, ITGAM, KCNS3, CRB2, PLCH2, P2RY2, TRO, BAI2, B3GNT3, EGF, SLC28A3, ADAM28, EPB41, LPP, KCNB1, ITGA3, OR5AC2, GPR35, SLC4A11, EPS8, P2RX3, NOTCH4, TSC2, MAP7, OR51A7, PLA2G4E, CHRNG, TM7SF4, CLSTN2, CLSTN3, KCNJ12, CDSN, FCRL3, NMUR2, SDPR, OR6P1, KCNG4, RHOF, USH2A, OR2AE1, TRPM2, OR10J1, NCR3, CD86, TNFRSF10C, CLECL1, SLC26A9, GRM6, PLA2G3, NKD2, LRRC32, TAP2, COL6A2, COL6A1, CCBP2, PLA2R1, SLC39A4, OR8G2, PPP1R9B, COG4, OR51M1, CDH13, PROM2, ERBB2IP, NRAP, GPR110, SLC17A2, GPR113, MEP1B, ANXA13, GPR111, SYNM, TAPBPL, ABCC8, GFRA2, ABCC6, GPR116	1.5×10^2
UP_SEQ_FEATURE	Topological domain: cytoplasmic	94	2.3×10^{-4}	TM7SF4, CLSTN2, CLSTN3, TRPV3, TMEM161A, GRIN3B, OR8S1, KCNJ12, FCRL3, LPHN2, OR4D2, CHRNA9, CD44, NMUR2, LRRC52, OR6P1, ANO2, PI16, KCNG4, ODZ3, USH2A, GRID1, OR10S1, PTPRN2, PLXNB2, CDHR2, SLC22A20, HLA-C, OR2AE1, PNPLA3, OR10J1, TRPM2, NCR3, CD86, TAS2R19, CLECL1, LILRB4, LRP11, GRM6, ERN1, GPR50, KCNH8, OR8B4,	0.078

				<i>FGFR4, PANX1, OR2T1, LRRC32, ENPP3, KEL, PAQR7, ITGAM, KCNS3, CRB2, P2RY2, ENTPD7, TAP2, BAI2, FUT3, UGT2A1, B4GALNT3, CCBP2, B3GNT3, PLA2R1, EGF, SLC39A4, GCNT1, SLC28A3, ADAM28, TMC6, OR8G2, KCNB1, SPPL2A, ITGA3, MAN1C1, OR5AC2, CPT1A, OR51M1, GPR35, SLC4A11, PROM2, PLSCR4, GPR110, P2RX3, NOTCH4, GPR113, MEP1B, MS4A10, GPR111, TAPBPL, OR51A7, ABCC8, ABCC6, GPR116, CHRNG</i>	
SP_PIR_KEYWORDS	Cell membrane	65	5.6×10^{-4}	<i>TM7SF4, CLSTN2, CLSTN3, GRIN3B, OR8S1, FCRL3, LPHN2, OR4D2, CHRNA9, NMUR2, OR6P1, ANO2, RHOF, KCNG4, USH2A, GRID1, CLCA1, OR10S1, CDHR2, HLA-C, OR2AE1, OR10J1, NCR3, TNFRSF10C, CLECL1, LILRB4, GRM6, GPR50, OR8B4, PLA2G3, SNTG2, NKD2, PANX1, OR2T1, KEL, MAP4K2, PAQR7, KCNS3, CRB2, P2RY2, PLCH2, BAI2, CCBP2, PLA2R1, SLC39A4, ADAM28, OR8G2, LPP, OR5AC2, OR51M1, CDH13, GPR35, PROM2, SLC4A11, GPR110, NOTCH4, GPR113, ANXA13, GPR111, MAP7, TAPBPL, OR51A7, GFRA2, CHRNG, GPR116</i>	0.030
UP_SEQ_FEATURE	Topological domain: extracellular	76	1.1×10^{-3}	<i>LSTN2, CLSTN3, TRPV3, TMEM161A, GRIN3B, OR8S1, KCNJ12, FCRL3, LPHN2, OR4D2, CHRNA9, CD44, NMUR2, LRRC52, OR6P1, ANO2, PI16, ODZ3, USH2A, GRID1, OR10S1, PLXNB2, PTPRN2, CDHR2, HLA-C, SLC22A20, OR2AE1, TRPM2, OR10J1, NCR3, CD86, TAS2R19, CLECL1, LILRB4, LRP11, GRM6, GPR50, OR8B4, FGFR4, PANX1, OR2T1, LRRC32, ENPP3, KEL, PAQR7, ITGAM, CRB2, P2RY2, BAI2, UGT2A1, CCBP2, PLA2R1, SLC39A4, EGF, SLC28A3, ADAM28, OR8G2, ITGA3, OR5AC2, OR51M1, GPR35, SLC4A11, PROM2, PLSCR4, GPR110, P2RX3, NOTCH4, GPR113, MEP1B, GPR111, MS4A10, ABCC8, OR51A7, CHRNG, GPR116, ABCC6</i>	0.245
SP_PIR_KEYWORDS	Transmembrane	116	0.017	<i>OR8S1, GRIN3B, SLC7A4, SLC26A10, LPHN2, OR4D2, CHRNA9, CD44, LRRC52, PI16, ODZ3, GRID1, SMCR7, OR10S1, SLC04A1, PTPRN2, PLXNB2, CDHR2, SLC22A20, HLA-C, PPP1R3A, PNPLA3, RNF133, LILRB4, LRP11, GPR50, OR8B4, AKAP1, FGFR4, PANX1, OR2T1, SLC38A8, KEL, ENPP3, PAQR7, ITGAM, KCNS3, CRB2, P2RY2, ENTPD7, BAI2, B3GNT3, EGF, GCNT1, SLC28A3, ADAM28, KCNB1, SPPL2A, ITGA3,</i>	0.37

				OR5AC2, GPR35, SLC4A11, P2RX3, NOTCH4, MS4A10, OR51A7, UTP14C, CHRNG, LRIT3, TM7SF4, CLSTN2, CLSTN3, TRPV3, TMEM161A, KCNJ12, FCRL3, C14ORF135, NMUR2, POMT1, OR6P1, ANO2, KCNG4, USH2A, OR2AE1, TRPM2, OR10J1, NCR3, CD86, CLECL1, TAS2R19, GRM6, SLC26A9, ERN1, KCNH8, C6ORF191, LRRC32, C2ORF85, TAP2, FUT3, UGT2A1, B4GALNT3, ILVBL, CCBP2, PLA2R1, SLC39A4, QSOX1, SOAT1, TMC6, SOAT2, MOGAT1, OR8G2, MAN1C1, CPT1A, OR51M1, PROM2, PLSCR4, CYP26C1, GPR110, SLC17A2, GPR113, MEP1B, GPR111, TAPBPL, ABCC8, ABCC6, GPR116	
UP_SEQ_FEATURE	Transmembrane region	116	0.019	OR8S1, GRIN3B, SLC7A4, SLC26A10, LPHN2, OR4D2, CHRNA9, CD44, LRRC52, PI16, ODZ3, GRID1, SMCR7, OR10S1, SLCO4A1, PTPRN2, PLXNB2, CDHR2, SLC22A20, HLA-C, PPP1R3A, PNPLA3, RNF133, LILRB4, LRP11, GPR50, OR8B4, FGFR4, PANX1, OR2T1, SLC38A8, KEL, ENPP3, PAQR7, ITGAM, KCNS3, CRB2, P2RY2, ENTPD7, BAI2, B3GNT3, EGF, GCNT1, SLC28A3, ADAM28, KCNB1, SPPL2A, ITGA3, OR5AC2, GPR35, SLC4A11, P2RX3, NOTCH4, MS4A10, OR51A7, UTP14C, CHRNG, LRIT3, TM7SF4, CLSTN2, CLSTN3, TRPV3, TMEM161A, KCNJ12, FCRL3, C14ORF135, NMUR2, POMT1, OR6P1, ANO2, KCNG4, USH2A, OR2AE1, TRPM2, OR10J1, NCR3, CD86, CLECL1, TAS2R19, GRM6, SLC26A9, ERN1, KCNH8, C6ORF191, LRRC32, C2ORF85, TAP2, FUT3, UGT2A1, B4GALNT3, ILVBL, CCBP2, PLA2R1, SLC39A4, QSOX1, SOAT1, TMC6, SOAT2, MOGAT1, OR8G2, MAN1C1, CPT1A, OR51M1, PROM2, PLSCR4, CYP26C1, GPR110, SLC17A2, GPR113, MEP1B, GPR111, TAPBPL, ABCC8, ABCC6, GPR116	0.802
SP_PIR_KEYWORDS	Membrane	140	0.029	OR8S1, GRIN3B, SLC26A10, SLC7A4, LPHN2, OR4D2, CHRNA9, CD44, LRRC52, PI16, ODZ3, GRID1, SMCR7, CLCA1, OR10S1, SLCO4A1, PTPRN2, PLXNB2, CDHR2, SLC22A20, HLA-C, PPP1R3A, PNPLA3, RNF133, PITPNM3, LILRB4, LRP11, USO1, GPR50, OR8B4, AKAP1, SNTG2, FGFR4, PANX1, SLC38A8, OR2T1, KEL, ENPP3, MAP4K2, PAQR7, ITGAM, KCNS3, CRB2, PLCH2, P2RY2, ENTPD7, BAI2, B3GNT3, EGF, GCNT1, SLC28A3, ADAM28, LPP, KCNB1, SPPL2A, ITGA3, OR5AC2,	0.41

				GPR35, SLC4A11, P2RX3, TSC2, NOTCH4, MS4A10, MAP7, ABL1, OR51A7, PLA2G4E, UTP14C, CHRNG, UQCRC2, LRIT3, TM7SF4, CLSTN2, CLSTN3, TRPV3, TMEM161A, KCNJ12, FCRL3, C14ORF135, SDPR, POMT1, NMUR2, OR6P1, ANO2, KCNG4, RHOF, USH2A, OR2AE1, TRPM2, OR10J1, NCR3, TNFRSF10C, CD86, CLECL1, TAS2R19, GRM6, SLC26A9, ERN1, KCNH8, PLA2G3, C6ORF191, NKD2, LRRC32, NLRX1, C2ORF85, TAP2, COL6A2, FUT3, UGT2A1, B4GALNT3, ILVBL, CCBP2, PLA2R1, SLC39A4, QSOX1, TRIP11, SOAT1, SOAT2, TMC6, MOGAT1, OR8G2, MAN1C1, CPT1A, CDH13, OR51M1, COG4, PROM2, CYP26C1, PLSCR4, GPR110, SLC17A2, GPR113, MEP1B, ANXA13, GPR111, TAPBPL, ABCC8, GFRA2, ABCC6, GPR116	
GOTERM_CC_FAT	Integral to membrane	123	0.036	OR8S1, GRIN3B, SLC7A4, SLC26A10, LPHN2, OR4D2, CHRNA9, CD44, LRRC52, PI16, ODZ3, GRID1, SMCR7, CLCA1, OR10S1, SLCO4A1, PTPRN2, PLXNB2, CDHR2, SLC22A20, HLA-C, PPP1R3A, PNPLA3, RNF133, PITPNM3, LILRB4, LRP11, GPR50, OR8B4, AKAP1, FGFR4, PANX1, OR2T1, SLC38A8, KEL, ENPP3, PAQR7, ITGAM, KCNS3, CRB2, TRO, P2RY2, ENTPD7, BAI2, B3GNT3, EGF, GCNT1, SLC28A3, ADAM28, KCNB1, SPPL2A, ITGA3, OR5AC2, GPR35, SLC4A11, EPS8, P2RX3, NOTCH4, MS4A10, OR51A7, UTP14C, CHRNG, LRIT3, TM7SF4, CLSTN2, CLSTN3, TRPV3, TMEM161A, KCNJ12, FCRL3, C14ORF135, NMUR2, POMT1, OR6P1, ANO2, KCNG4, USH2A, OR2AE1, TRPM2, OR10J1, NCR3, CD86, TNFRSF10C, CLECL1, TAS2R19, NUP205, GRM6, SLC26A9, ERN1, KCNH8, C6ORF191, LRRC32, C2ORF85, NUP214, TAP2, FUT3, UGT2A1, B4GALNT3, ILVBL, CCBP2, PLA2R1, SLC39A4, QSOX1, SOAT1, TMC6, SOAT2, MOGAT1, OR8G2, MAN1C1, CPT1A, OR51M1, PROM2, CYP26C1, PLSCR4, GPR110, SLC17A2, GPR113, MEP1B, GPR111, TAPBPL, ABCC8, ABCC6, GPR116	0.58
GOTERM_CC_FAT	Intrinsic to membrane	125	0.059	OR8S1, GRIN3B, SLC7A4, SLC26A10, LPHN2, OR4D2, CHRNA9, CD44, LRRC52, PI16, ODZ3, GRID1, SMCR7, CLCA1, OR10S1, SLCO4A1, PTPRN2, PLXNB2, CDHR2, SLC22A20, HLA-C, PPP1R3A, PNPLA3, RNF133, PITPNM3, LILRB4, LRP11, GPR50, OR8B4, AKAP1, FGFR4, PANX1, OR2T1, SLC38A8, KEL, ENPP3,	0.58

				PAQR7, ITGAM, KCNS3, CRB2, TRO, P2RY2, ENTPD7, BAI2, B3GNT3, EGF, GCNT1, SLC28A3, ADAM28, KCNB1, SPPL2A, ITGA3, OR5AC2, GPR35, SLC4A11, EPS8, P2RX3, NOTCH4, MS4A10, OR51A7, UTP14C, CHRNG, LRIT3, TM7SF4, CLSTN2, CLSTN3, TRPV3, TMEM161A, KCNJ12, FCRL3, C14ORF135, NMUR2, POMT1, OR6P1, ANO2, KCNG4, USH2A, OR2AE1, TRPM2, OR10J1, NCR3, CD86, TNFRSF10C, CLECL1, TAS2R19, NUP205, GRM6, SLC26A9, ERN1, KCNH8, C6ORF191, LRRC32, C2ORF85, NUP214, TAP2, FUT3, UGT2A1, B4GALNT3, ILVBL, CCBP2, PLA2R1, SLC39A4, QSOX1, SOAT1, TMC6, SOAT2, MOGAT1, OR8G2, MAN1C1, CPT1A, OR51M1, CDH13, PROM2, CYP26C1, PLSCR4, GPR110, SLC17A2, GPR113, MEP1B, GPR111, TAPBPL, ABCC8, GFRA2, ABCC6, GPR116	
Annotation cluster 2	Enrichment score: 2.35				
Category	Term	Count	P	Genes	P-Benjamini
UP_SEQ_FEATURE	Domain:GPS	6	6.3x10 ⁻⁴	LPHN2, GPR110, GPR113, BAI2, GPR111, GPR116	0.17
INTERPRO	GPS	6	8.8x10 ⁻⁴	LPHN2, GPR110, GPR113, BAI2, GPR111, GPR116	0.45
PIR_SUPERFAMILY	PIRSF800007:secretin receptor-like G protein-coupled receptors	6	9x10 ⁻⁴	LPHN2, GPR110, GPR113, BAI2, GPR111, GPR116	0.13
SMART	GPS	6	1.3x10 ⁻³	LPHN2, GPR110, GPR113, BAI2, GPR111, GPR116	0.21
INTERPRO	GPCR, family 2, secretin-like	6	2.1x10 ⁻³	LPHN2, GPR110, GPR113, BAI2, GPR111, GPR116	0.51
INTERPRO	GPCR, family 2, secretin-like, conserved site	6	2.6x10 ⁻³	LPHN2, GPR110, GPR113, BAI2, GPR111, GPR116	0.44
INTERPRO	GPCR, family 2-like	6	6.3x10 ⁻³	LPHN2, GPR110, GPR113, BAI2, GPR111, GPR116	0.65

GOTERM_BP_FAT	Neuropeptide signaling pathway	7	0.011	<i>LPHN2, GPR110, NMUR2, GPR113, BAI2, GPR111, GPR116</i>	1.00
INTERPRO	GPCR, family 2, extracellular region	3	0.10	<i>LPHN2, GPR113, BAI2</i>	0.95
SMART	HormR	3	0.12	<i>LPHN2, GPR113, BAI2</i>	0.94
Annotation cluster 3	Enrichment score: 2.19				
Category	Term	Count	P	Genes	P-Benjamini
SP_PIR_KEYWORDS	Glycoprotein	125	9.2×10^{-7}	<i>FSTL4, OR8S1, GRIN3B, SLC7A4, MMRN2, LPHN2, OR4D2, CHRNA9, CD44, LRRC52, PI16, ODZ3, GRID1, CLCA1, OR10S1, SLCO4A1, SPARCL1, PTPRN2, PLXNB2, CDHR2, SLC22A20, HLA-C, PNPLA3, TNFAIP6, LRP11, OR8B4, CHGB, DST, FGFR4, PANX1, AMTN, OR2T1, KEL, ENPP3, ITGAM, CRB2, P2RY2, ENTPD7, BAI2, SFTPD, B3GNT3, EGF, GCNT1, ADAM28, EPB41, ITGA3, OR5AC2, GPR35, SLC4A11, P2RX3, NOTCH4, TSC2, OR51A7, UTP14C, COL20A1, CHRNG, ARSB, ARSD, CLSTN2, CLSTN3, TMEM161A, CDSN, FCRL3, RSPO4, C14ORF135, SMPDL3A, NMUR2, POMT1, OR6P1, ANO2, CTBS, USH2A, OR2AE1, OR10J1, NCR3, CD86, TNFRSF10C, CLECL1, TAS2R19, ZPBP2, GRM6, ERN1, KCNH8, C4ORF29, PLA2G3, WFIKKN2, LRRC32, CPN2, NUP214, C1QTNF6, FCN2, COL6A2, FUT3, UGT2A1, COL6A1, CCBP2, HRG, PLA2R1, DKKL1, SLC39A4, QSOX1, SOAT1, TMC6, MOGAT1, CES1, OR8G2, FBN1, FUCA2, MAN1C1, COL5A1, PLG, OR51M1, CDH13, PROM2, GPR110, SLC17A2, ANXA11, GPR113, MEP1B, GPR111, MEGF6, ABCC8, GFRA2, ABCC6, GPR116</i>	1.7×10^{-4}
UP_SEQ_FEATURE	N-linked glycosylation sites (GlcNAc...)	116	1.7×10^{-5}	<i>FSTL4, OR8S1, GRIN3B, SLC7A4, MMRN2, LPHN2, OR4D2, CHRNA9, CD44, LRRC52, PI16, ODZ3, GRID1, CLCA1, OR10S1, SPARCL1, SLCO4A1, PTPRN2, PLXNB2, CDHR2, SLC22A20, HLA-C, PNPLA3, TNFAIP6, LRP11, OR8B4, DST,</i>	1.2×10^{-2}

				<p>FGFR4, PANX1, OR2T1, KEL, ENPP3, ITGAM, CRB2, P2RY2, ENTPD7, BAI2, SFTPD, B3GNT3, EGF, GCNT1, ADAM28, ITGA3, OR5AC2, GPR35, SLC4A11, P2RX3, NOTCH4, OR51A7, UTP14C, COL20A1, CHRNG, ARSB, ARSD, CLSTN2, CLSTN3, TMEM161A, CDSN, FCRL3, RSPO4, C14ORF135, SMPDL3A, NMUR2, POMT1, OR6P1, ANO2, CTBS, USH2A, OR2AE1, OR10J1, NCR3, CD86, CLECL1, TAS2R19, ZPBP2, GRM6, ERN1, KCNH8, C4ORF29, PLA2G3, WFIKKN2, LRRC32, CPN2, C1QTNF6, FCN2, COL6A2, FUT3, UGT2A1, COL6A1, CCBP2, HRG, PLA2R1, DKKL1, SLC39A4, QSOX1, TMC6, MOGAT1, CES1, OR8G2, FBN1, FUCA2, MAN1C1, PLG, OR51M1, CDH13, PROM2, GPR110, SLC17A2, GPR113, MEP1B, GPR111, MEGF6, ABCC8, GFRA2, ABCC6, GPR116</p>	
SP_PIR_KEYWORDS	Signal	83	6.1×10^{-3}	<p>ARSB, ARSD, CLSTN2, CLSTN3, TMEM161A, FSTL4, GRIN3B, CDSN, MMRN2, FCRL3, APOA4, LPHN2, RSPO4, CHRNA9, CD44, SMPDL3A, LRRC52, PI16, CTBS, USH2A, GRID1, CLCA1, SPARCL1, PLXNB2, PTPRN2, CDHR2, HLA-C, CECR5, NCR3, TNFAIP6, AMBN, TNFRSF10C, CD86, ZPBP2, TXNDC5, LILRB4, LRP11, GRM6, ERN1, C4ORF29, PLA2G3, CHGB, WFIKKN2, FGFR4, AMTN, LRRC32, ITGAM, CPN2, C1QTNF6, CRB2, FCN2, SFTPD, BAI2, COL6A2, UGT2A1, COL6A1, HRG, PLA2R1, SLC39A4, DKKL1, EGF, QSOX1, DEFB119, ADAM28, FGF3P3, CES1, FBN1, ITGA3, FUCA2, COL5A1, PLG, CDH13, PROM2, GPR110, NOTCH4, GPR113, MEP1B, TAPBPL, MEGF6, GFRA2, COL20A1, GPR116, CHRNG</p>	0.23
UP_SEQ_FEATURE	Signal peptide	83	7.2×10^{-3}	<p>ARSB, ARSD, CLSTN2, CLSTN3, TMEM161A, FSTL4, GRIN3B, CDSN, MMRN2, FCRL3, APOA4, LPHN2, RSPO4, CHRNA9, CD44, SMPDL3A, LRRC52, PI16, CTBS, USH2A, GRID1, CLCA1, SPARCL1, PLXNB2, PTPRN2, CDHR2, HLA-C, CECR5, NCR3, TNFAIP6, AMBN, TNFRSF10C, CD86, ZPBP2, TXNDC5, LILRB4, LRP11, GRM6, ERN1, C4ORF29, PLA2G3, CHGB, WFIKKN2, FGFR4, AMTN, LRRC32, ITGAM, CPN2, C1QTNF6, CRB2, FCN2, SFTPD, BAI2, COL6A2, UGT2A1, COL6A1, HRG, PLA2R1, SLC39A4, DKKL1, EGF, QSOX1, DEFB119, ADAM28, FGF3P3, CES1, FBN1, ITGA3, FUCA2,</p>	0.61

				COL5A1, PLG, CDH13, PROM2, GPR110, NOTCH4, GPR113, MEP1B, TAPBPL, MEGF6, GFRA2, COL20A1, GPR116, CHRNG	
GOTERM_CC_FAT	Extracellular region part	28	0.052	DLBP, AMTN, MMRN2, APOA4, CD44, SMPDL3A, FCN2, COL6A2, SFTPD, COL6A1, PLA2R1, EGF, DKKL1, QSOX1, USH2A, CLCA1, SPARCL1, FBN1, COL5A1, PLG, CDH13, AMBN, ERBB2IP, NOTCH4, MEP1B, PLA2G3, DST, COL20A1	0.59
SP_PIR_KEYWORDS	Disulfide bond	67	0.10	ARSB, LRIT3, FSTL4, OR8S1, MMRN2, FCRL3, LPHN2, RSPO4, OR4D2, CHRNA9, CD44, NMUR2, OR6P1, ODZ3, USH2A, OR10S1, SPARCL1, HLA-C, OR2AE1, OR10J1, NCR3, TNFAIP6, TNFRSF10C, CD86, TXNDC5, LILRB4, LRP11, ERN1, GPR50, OR8B4, PLA2G3, CHGB, WFIKKN2, FGFR4, OR2T1, ENPP3, KEL, CPN2, ITGAM, CRB2, FCN2, P2RY2, SFTPD, BAI2, HRG, CCBP2, PLA2R1, EGF, GCNT1, DEFB119, ADAM28, FGF3P3, CES1, OR8G2, FBN1, ITGA3, OR5AC2, PLG, OR51M1, GPR35, NOTCH4, MEP1B, TAPBPL, MEGF6, OR51A7, CHRNG, GPR116	0.67
GOTERM_CC_FAT	Extracellular region	49	0.12	WFIKKN2, HDLBP, AMTN, TUFT1, ENPP3, FSTL4, CDSN, CPN2, MMRN2, APOA4, RSPO4, C1QTNF6, CD44, CRB2, SMPDL3A, FCN2, SFTPD, COL6A2, COL6A1, HRG, PLA2R1, PI16, EGF, CASP1, DKKL1, QSOX1, USH2A, DEFB119, FGF3P3, ADAM28, CLCA1, SPARCL1, FBN1, HLA-C, FUCA2, COL5A1, PLG, CDH13, AMBN, ERBB2IP, ZBP2, NOTCH4, MEP1B, C4ORF29, PLA2G3, MEGF6, DST, CHGB, COL20A1	0.76
SP_PIR_KEYWORDS	Secreted	39	0.19	WFIKKN2, AMTN, TUFT1, ENPP3, FSTL4, CDSN, CPN2, MMRN2, APOA4, RSPO4, C1QTNF6, CRB2, SMPDL3A, FCN2, COL6A2, SFTPD, COL6A1, HRG, PLA2R1, DKKL1, QSOX1, USH2A, DEFB119, FGF3P3, ADAM28, CLCA1, SPARCL1, FBN1, HLA-C, FUCA2, COL5A1, PLG, AMBN, ZBP2, C4ORF29, PLA2G3, MEGF6, CHGB, COL20A1	0.77
GOTERM_CC_FAT	Extracellular space	18	0.23	HDLBP, CLCA1, FBN1, PLG, MMRN2, APOA4, CDH13, SMPDL3A, FCN2, NOTCH4, SFTPD, MEP1B, PLA2G3, PLA2R1, EGF, DKKL1, QSOX1, COL20A1	0.86
Annotation cluster 4	Enrichment score: 1.79				

Category	Term	Count	P	Genes	P-Benjamini
GOTERM_CC_FAT	Plasma membrane part	64	2.2x10 ⁻³	<i>LMO7, GRIN3B, KCNJ12, CDSN, CHRNA9, CD44, WNK4, SDPR, RHOF, KCNG4, GRID1, CLCA1, MYH1, PTPRN2, CDHR2, HLA-C, CTNNA3, OR10J1, TRPM2, NCR3, TNFRSF10C, CD86, GRM6, SLC26A9, GPR50, KBTBD10, DST, SNTG2, FGFR4, PANX1, AMTN, LRRC32, ENPP3, KEL, ITGAM, KCNS3, P2RY2, TRO, TAP2, CCBP2, B3GNT3, PLA2R1, SLC39A4, LPP, KCNB1, ITGA3, PPP1R9B, COG4, CDH13, GPR35, SLC4A11, ERBB2IP, EPS8, NRAP, P2RX3, TSC2, SLC17A2, NOTCH4, MEP1B, MAP7, SYNM, PLA2G4E, CHRNG, ABCC6</i>	0.26
GOTERM_CC_FAT	Integral to plasma membrane	34	0.039	<i>FGFR4, LRRC32, KEL, ENPP3, GRIN3B, ITGAM, KCNS3, CHRNA9, CD44, P2RY2, TAP2, TRO, B3GNT3, CCBP2, PLA2R1, KCNG4, CLCA1, PTPRN2, KCNB1, ITGA3, HLA-C, TRPM2, OR10J1, NCR3, GPR35, TNFRSF10C, EPS8, P2RX3, NOTCH4, SLC17A2, GRM6, MEP1B, GPR50, CHRNG</i>	0.55
GOTERM_CC_FAT	Intrinsic to plasma membrane	34	0.050	<i>FGFR4, LRRC32, KEL, ENPP3, GRIN3B, ITGAM, KCNS3, CHRNA9, CD44, P2RY2, TAP2, TRO, B3GNT3, CCBP2, PLA2R1, KCNG4, CLCA1, PTPRN2, KCNB1, ITGA3, HLA-C, TRPM2, OR10J1, NCR3, GPR35, TNFRSF10C, EPS8, P2RX3, NOTCH4, SLC17A2, GRM6, MEP1B, GPR50, CHRNG</i>	0.60
Annotation cluster 5	Enrichment score: 1.62				
Category	Term	Count	P	Genes	P-Benjamini
SP_PIR_KEYWORDS	Cell adhesion	16	0.019	<i>CLSTN2, AMTN, CLSTN3, LPP, CDHR2, ITGA3, ITGAM, CTNNA3, CDH13, TNFAIP6, CD44, TRO, COL6A2, COL6A1, ABL1, DST</i>	0.36
GOTERM_BP_FAT	Cell-cell adhesion	12	0.025	<i>OA4, CDH13, CLSTN2, CD44, CLSTN3, TRO, CDHR2, COL6A2, LMO7, CDSN, ITGAM, CTNNA3</i>	1.00
GOTERM_BP_FAT	Biological adhesion	23	0.026	<i>CLSTN2, AMTN, CLSTN3, LPP, CDHR2, LMO7, ITGA3, CDSN, COL5A1, ITGAM, CTNNA3, APOA4, CDH13, TNFAIP6, AMBN, ERBB2IP, CD44, TRO, COL6A2, COL6A1, ABL1, DST,</i>	1.00

				<i>COL20A1</i>	
GOTERM_BP_FAT	Cell adhesion	23	0.027	<i>CLSTN2, AMTN, CLSTN3, LPP, CDHR2, LMO7, ITGA3, CDSN, COL5A1, ITGAM, CTNNA3, APOA4, CDH13, TNFAIP6, AMBN, ERBB2IP, CD44, TRO, COL6A2, COL6A1, ABL1, DST, COL20A1</i>	1.00
Annotation cluster 6	Enrichment score: 1.61				
Category	Term	Count	P	Genes	P-Benjamini
INTERPRO	Kinesin, motor region, conserved site	5	9.5×10^{-3}	<i>KIF1B, KIF6, KIF17, CENPE, KIF16B</i>	0.66
INTERPRO	Kinesin, motor region	5	9.5×10^{-3}	<i>KIF1B, KIF6, KIF17, CENPE, KIF16B</i>	0.66
UP_SEQ_FEATURE	Domain: kinesin-motor	5	9.7×10^{-3}	<i>KIF1B, KIF6, KIF17, CENPE, KIF16B</i>	0.69
SMART	KISc	5	0.013	<i>KIF1B, KIF6, KIF17, CENPE, KIF16B</i>	0.68
SP_PIR_KEYWORDS	Motor protein	8	0.015	<i>DNALI1, KIF1B, DNAI1, MYH1, KIF6, KIF17, CENPE, KIF16B</i>	0.35
GOTERM_MF_FAT	Microtubule motor activity	6	0.020	<i>DNALI1, KIF1B, KIF6, KIF17, CENPE, KIF16B</i>	0.59
GOTERM_MF_FAT	Motor activity	8	0.026	<i>DNALI1, KIF1B, DNAI1, MYH1, KIF6, KIF17, CENPE, KIF16B</i>	0.62
GOTERM_BP_FAT	Microtubule-based movement	7	0.027	<i>KIF1B, KIF6, KIF17, CENPE, KIF16B, DST, HAP1</i>	1.00
GOTERM_CC_FAT	Microtubule cytoskeleton	19	0.031	<i>KIF17, CENPE, AKAP9, VPS41, WRN, AKAP11, KIF16B, KIAA1009, DNALI1, CDC45, KIF1B, DNAI1, KIF6, ANXA11, NINL, MAP4, MAP7, MAP9, DST</i>	0.55
SP_PIR_KEYWORDS	Microtubule	10	0.037	<i>KIF1B, DNAI1, KIF6, KIF17, NINL, MAP4, CENPE, MAP7, KIF16B, MAP9</i>	0.45
GOTERM_CC_FAT	Microtubule	11	0.053	<i>KIF1B, DNAI1, KIF6, KIF17, NINL, MAP4, VPS41, CENPE, MAP7, KIF16B, MAP9</i>	0.57
GOTERM_CC_FAT	Microtubule associated complex	6	0.056	<i>DNALI1, KIF1B, DNAI1, KIF17, MAP4, MAP7</i>	0.58
GOTERM_BP_FAT	Microtubule-based process	9	0.14	<i>KIF1B, KIF6, KIF17, CENPE, MAP7, KIF16B, TACC3, DST, HAP1</i>	1.00
Annotation cluster 7	Enrichment score: 1.59				

Category	Term	Count	P	Genes	P-Benjamini
INTERPRO	Forkhead-associated	4	0.023	<i>KIF1B, MKI67, KIF16B, FHAD1</i>	0.86
UP_SEQ_FEATURE	Domain: FHA	4	0.024	<i>KIF1B, MKI67, KIF16B, FHAD1</i>	0.84
SMART	FHA	4	0.030	<i>KIF1B, MKI67, KIF16B, FHAD1</i>	0.83
Annotation cluster 8	Enrichment score: 1.42				
Category	Term	Count	P	Genes	P-Benjamini
GOTERM_CC_FAT	Basolateral plasma membrane	11	8.5×10^{-3}	<i>SLC4A11, NRAP, MYH1, ERBB2IP, CD44, LPP, P2RY2, NOTCH4, MAP7, KBTBD10, DST</i>	0.38
GOTERM_CC_FAT	Cell junction	20	9.2×10^{-3}	<i>PANX1, AMTN, MYH1, LPP, CDHR2, LMO7, GRIN3B, CDSN, CTNNA3, PPP1R9B, NRAP, CHRNA9, ERBB2IP, CD44, WNK4, SYNM, KBTBD10, DST, GRID1, CHRNG</i>	0.35
GOTERM_CC_FAT	Anchoring junction	9	0.024	<i>NRAP, MYH1, CD44, LPP, LMO7, SYNM, KBTBD10, CDSN, CTNNA3</i>	0.58
GOTERM_CC_FAT	Cell-substrate junction	7	0.026	<i>NRAP, MYH1, ERBB2IP, CD44, LPP, KBTBD10, DST</i>	0.57
GOTERM_CC_FAT	Adherens junction	8	0.04	<i>NRAP, MYH1, CD44, LPP, LMO7, SYNM, KBTBD10, CTNNA3</i>	0.57
GOTERM_CC_FAT	Cell-substrate adherens junction	5	0.17	<i>NRAP, MYH1, CD44, LPP, KBTBD10</i>	0.79
GOTERM_CC_FAT	Focal adhesion	4	0.34	<i>MYH1, CD44, LPP, KBTBD10</i>	0.92
Annotation cluster 9	Enrichment score: 1.39				
Category	Term	Count	P	Genes	P-Benjamini
GOTERM_CC_FAT	Basement membrane	7	5×10^{-3}	<i>AMTN, ERBB2IP, FBN1, DST, COL5A1, USH2A, MMRN2</i>	0.37
GOTERM_CC_FAT	Extracellular matrix part	8	9.7×10^{-3}	<i>AMTN, ERBB2IP, FBN1, COL6A1, DST, COL5A1, USH2A, MMRN2</i>	0.33
GOTERM_CC_FAT	Extracellular matrix	13	0.049	<i>AMBN, AMTN, ERBB2IP, CD44, SPARCL1, FBN1, COL6A2, SFTPD, COL6A1, DST, COL5A1, USH2A, MMRN2</i>	0.61
GOTERM_CC_FAT	Proteinaceous extracellular matrix	12	0.061	<i>AMBN, AMTN, ERBB2IP, SPARCL1, FBN1, COL6A2, SFTPD, COL6A1, DST, COL5A1, USH2A, MMRN2</i>	0.57

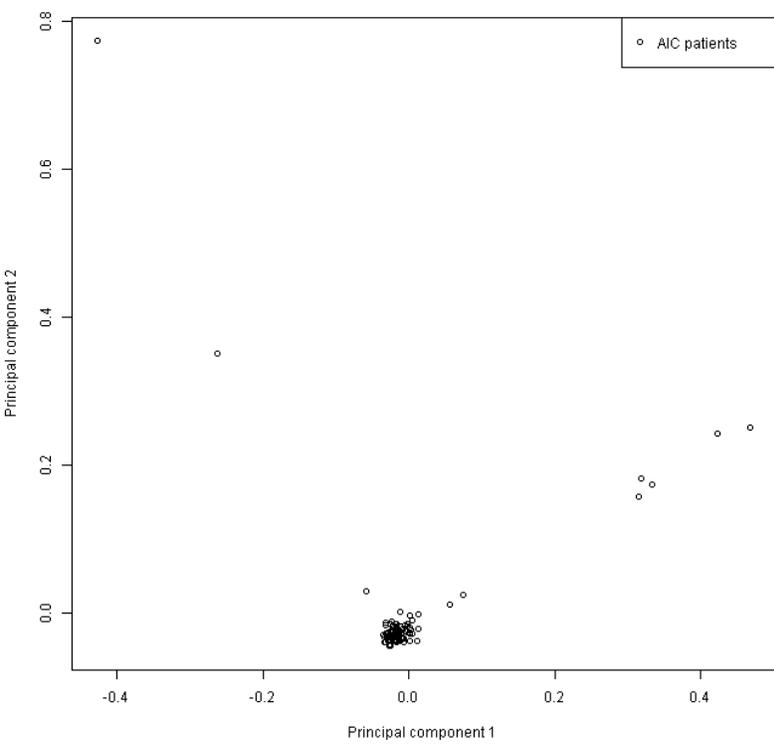
SP_PIR_KEYWORDS	Extracellular matrix	9	0.098	AMBN, SPARCL1, FBN1, COL6A2, SFTPD, COL6A1, DST, COL5A1, MMRN2	0.67
GOTERM_MF_FAT	Integrin binding	3	0.34	ERBB2IP, DST, COL5A1	0.98
Annotation cluster 10	Enrichment score: 1.31				
Category	Term	Count	P	Genes	P-Benjamini
SP_PIR_KEYWORDS	Hydroxyproline	5	5.5x10 ⁻³	FCN2, COL6A2, SFTPD, COL6A1, COL5A1	0.23
INTERPRO	Collagen triple helix repeat	7	6.5x10 ⁻³	C1QTNF6, FCN2, COL6A2, SFTPD, COL6A1, COL5A1, COL20A1	0.59
SP_PIR_KEYWORDS	Collagen	7	0.010	C1QTNF6, FCN2, COL6A2, SFTPD, COL6A1, COL5A1, COL20A1	0.28
SP_PIR_KEYWORDS	Hydroxylysine	4	0.022	COL6A2, SFTPD, COL6A1, COL5A1	0.39
UP_SEQ_FEATURE	Domain: collagen-like	4	0.033	C1QTNF6, FCN2, SFTPD, COL20A1	0.89
SP_PIR_KEYWORDS	Hydroxylation	5	0.049	FCN2, COL6A2, SFTPD, COL6A1, COL5A1	0.52
KEGG_PATHWAY	ECM-receptor interaction	5	0.059	CD44, COL6A2, COL6A1, ITGA3, COL5A1	1.00
UP_SEQ_FEATURE	Region of interest: triple-helical region	3	0.073	COL6A2, COL6A1, COL5A1	0.94
SP_PIR_KEYWORDS	Trimer	3	0.09	COL6A2, COL6A1, COL5A1	0.65
SP_PIR_KEYWORDS	Triple helix	3	0.12	COL6A2, COL6A1, COL5A1	0.7X
SP_PIR_KEYWORDS	Pyroglutamic acid	3	0.22	COL6A2, COL6A1, COL5A1	0.82
SP_PIR_KEYWORDS	Blocked amino end	4	0.26	COL6A2, SFTPD, COL6A1, ANXA13	0.83
KEGG_PATHWAY	Focal adhesion	6	0.28	COL6A2, COL6A1, ITGA3, EGF, COL5A1, MYLK	0.99

Gene-set enrichment analysis of gene-based *P*-values from SKAT-O was performed using the functional annotation clustering analysis module of the bioinformatic tool DAVID. The DAVID tool provides a comprehensive set of functional annotation tools to understand biological meaning (i.e., to discover enriched functional-related gene groups, enriched biological themes) from large list of genes. Each annotation term group is assigned an enrichment score (ES) to rank overall importance. Only annotation clusters with ES≥1.3 (indicating biological significance) are shown. The significance of gene-term enrichment was assessed with a modified Fisher's exact test and *P*-values are corrected using Benjamini-Hochberg's by false discovery rate (FDR-BH) procedure.

Supplementary Table 4. Clinical information recorded from anthracycline-treated pediatric cancer patients

Clinical information	
Age at diagnosis (years)*	
Gender	
Primary tumor site	
Metastasis at diagnosis	
Metastasis during/after completion of treatment	
Relapse	
Family history of cardiovascular disease	
Anthracycline type	
Cumulative anthracycline dose (mg/m²)^a	
Radiotherapy involving the heart ^b	
Concomitant therapy^c	
Pre-existing cardiac risk factors ^d	
Hematopoietic progenitor cell transplantation	
Echocardiographic information	
Status (alive/dead)	
Last date of follow-up	
<p>Information regarding patient demographic and clinical characteristics and therapy was abstracted from medical records. Clinical factors with associated $P < 0.05$ in univariable analyses are indicated in bold type. * Covariates for single-variant, gene-based and pathway tests. a Cumulative anthracycline dose was calculated using doxorubicin equivalents according to the Children's Oncology Group Long-Term Follow-Up Guidelines for Survivors of Childhood, Adolescent, and Young Adult Cancers [48]. b Radiotherapy involving the heart includes mediastinal and mantle radiation and total body irradiation. c Bleomycin concomitant administration was significantly associated in univariable analysis. d Pre-existing cardiac risk factors include hypertension, diabetes and obesity.</p>	

a)



b)

