1 atrt

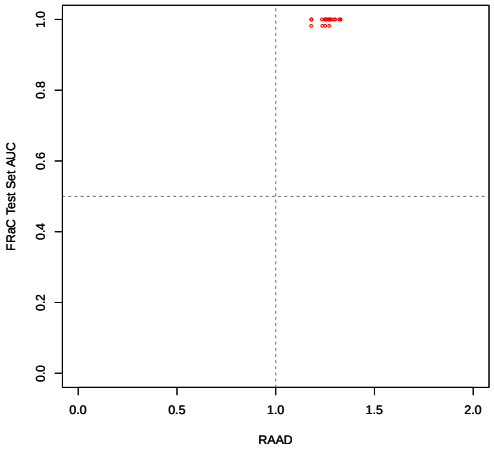

Figure 1: RAAD scatter plot for data set ATRT[1]: Rhabdoid and non-rhabdoid brain tumors. Task: Identify AT/RT (Atypical teratoid rhabdoid tumor) among brain tumor samples. Number of samples: 48. Tissue type: brain tumor. Normal: non-rhabdoid tumors, Anomaly: atypical rhabdoid tumors

2 bcat

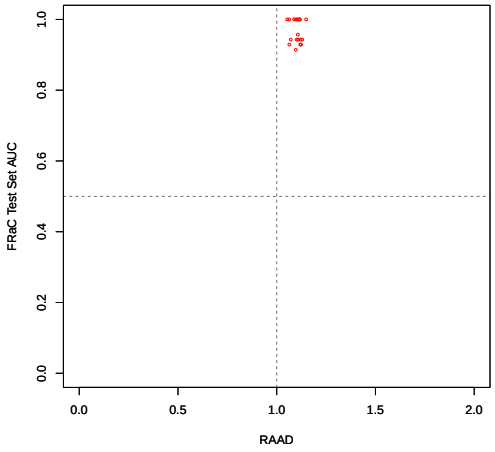

Figure 2: RAAD scatter plot for data set BCAT[2]: Medulloblastoma data. Task: Identify beta-catenin (BCAT) deletion 6q (monosomy 6). Number of samples: 45. Tissue type: Medulloblastoma. Normal: other medulloblastoma, Anomaly: beta-catenin deletion 6q (monosomy 6)

3 bild

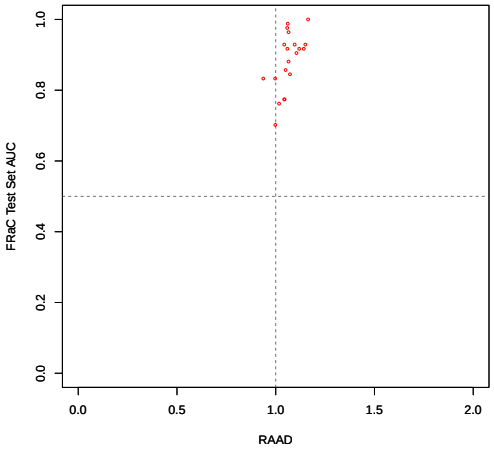

Figure 3: RAAD scatter plot for data set BILD[3]: RNA from frozen tissue of primary lung tumors. Task: Distinguish Src from Myc, Ras, E2F3, beta-catenin. Number of samples: 55. Normal: Myc, beta-catenin, E2F3, Ras, and GFP controls, Anomaly: Src activity in human cancers. Availability: `http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE3151`

4 biomarkers

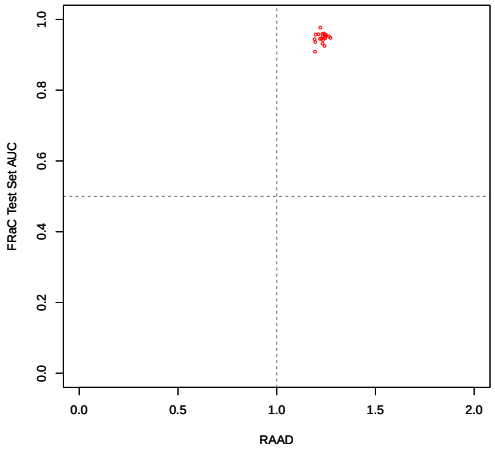

Figure 4: RAAD scatter plot for data set BIOMARKERS[4]: Breast cancer data. Task: Distinguish ER- from ER+. Number of samples: 127. Normal: ER+, Anomaly: ER-. Availability: `http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE5460`

5 breast.basal

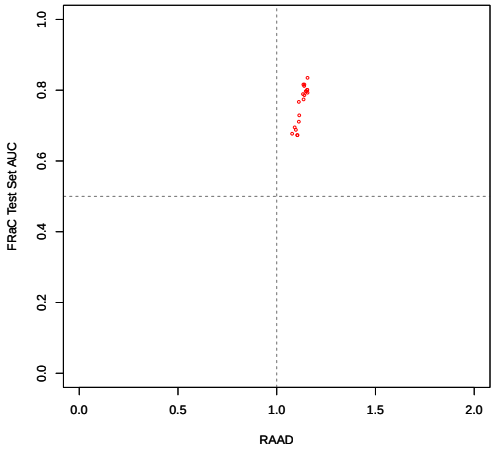

Figure 5: RAAD scatter plot for data set BREAST.BASAL[5]: Breast samples. Task: Distinguish basal-like samples from: ERRB2-overexpressing; luminal-like (two related classes A and B), normal-breast-tissue-like subgroup. Number of samples: 75. Normal: ERRB2-overexpressing, luminal-like, or normal-breast-tissue-like subgroup, Anomaly: Basal-like. Availability: `http://smd.stanford.edu/cgi-bin/publication/viewPublication.pl?pub_no=248`

6 breast.er

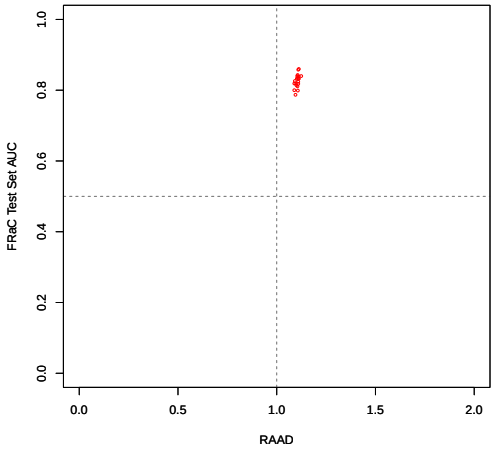

Figure 6: RAAD scatter plot for data set BREAST.ER[6]: Breast cancer data. Task: Distinguish ER- from ER+. Number of samples: 286. Normal: ER+, Anomaly: ER-. Availability: `http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE2034`. See `http://www.ncbi.nlm.nih.gov/projects/geo/query/acc.cgi?view=data&acc=GSE2034&id=40089&db=GeoDb_blob26`

7 desmoplastic

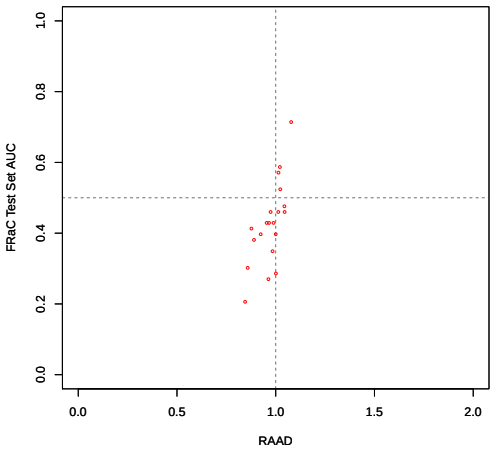

Figure 7: RAAD scatter plot for data set DESMOPLASTIC[2]: Medulloblastoma data. Task: Distinguish desmoplastic from classic medulloblastomas. Number of samples: 34. Normal: classic, Anomaly: desmoplastic

8 diabetes

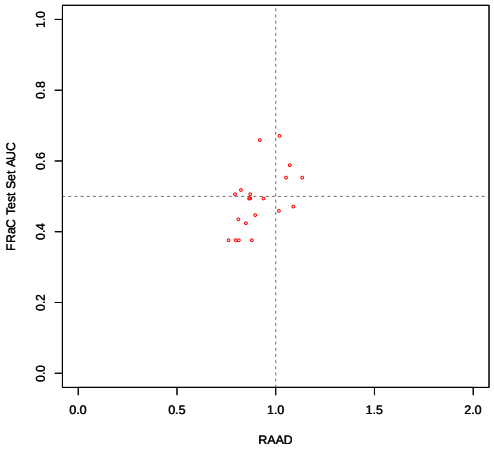

Figure 8: RAAD scatter plot for data set DIABETES[7]: Transcriptional profiles of smooth muscle biopsies of diabetic and normal individuals. Task: Identify diabetic samples among normal individuals. Number of samples: 34. Tissue type: Smooth muscle. Normal: non-diabetic, Anomaly: diabetic. Availability: `http://www.broadinstitute.org/gsea/datasets.jsp`

9 downs

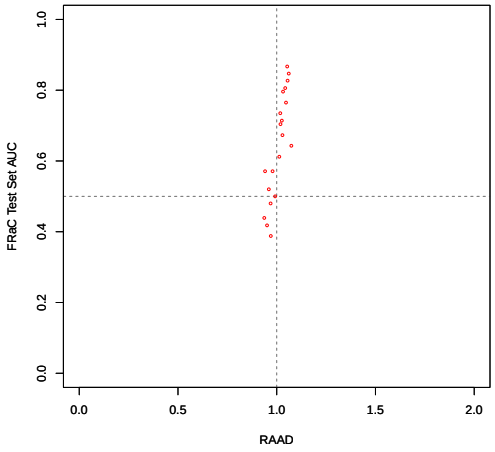

Figure 9: RAAD scatter plot for data set DOWNS[8]: Expression profiling of acute megakaryoblastic leukemia. Task: Identify Down Syndrome. Number of samples: 39. Tissue type: Megakaryoblastic leukemia. Normal: euploid, Anomaly: Down syndrome. Note: This data set is limited to the following 39 samples: 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 24, 25, 26, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 44, 45, 46, 47, 48, 49, 50, 52, 53, 54, 57. Availability: `http://www.ncbi.nlm.nih.gov/projects/geo/query/acc.cgi?acc=GSE4119`

10 ethnic

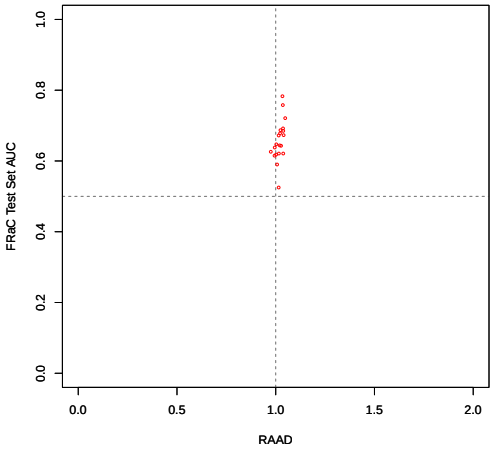

Figure 10: RAAD scatter plot for data set ETHNIC[9, 10, 11, 12, 13, 14, 15]: Lymphoblastoid cell lines from various ethnicities. Task: Identify Caucasian (vs. African American). Number of samples: 191. Normal: African American, Anomaly: Caucasian. Note: This anomaly detection task does not use the Han Chinese American samples. Availability: `http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE23120`

11 gender

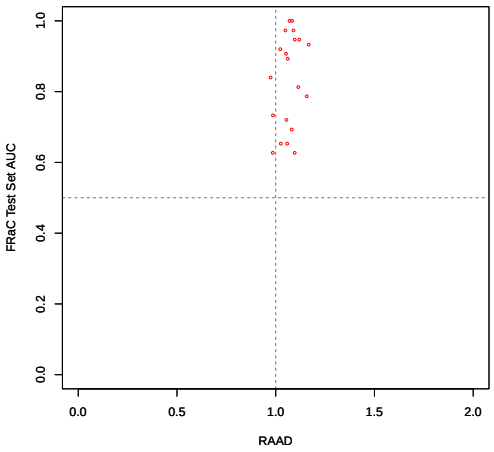

Figure 11: RAAD scatter plot for data set GENDER: Lymphoblastoid cell lines. Task: Distinguish gender. Number of samples: 32. Normal: female, Anomaly: male. Note: Unpublished. See `http://www.broadinstitute.org/gsea/datasets.jsp`

12 hematopoiesis

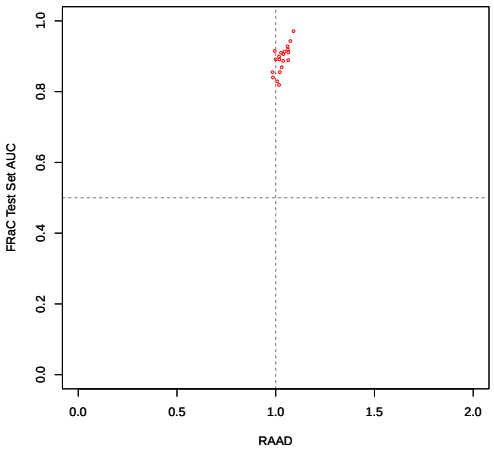

Figure 12: RAAD scatter plot for data set HEMATOPOIESIS[16]: Human hematopoiesis data from various (myeloid, lymphoid, and progenitor) blood cell types, including: basophil, myeloid or plasmacytoid dendritic cell, erythroid, granulocyte/monocyte progenitor, hematopoietic stem cell, CFU-MK and megakaryocyte, NK cell, B cell, and T cell. Task: Distinguish myeloid from lymphoid types (common progenitor samples are removed). Number of samples: 188. Normal: myeloid, Anomaly: lymphoid. Availability: `http://www.ncbi.nlm.nih.gov/projects/geo/query/acc.cgi?acc=GSE24759`

13 leukemia

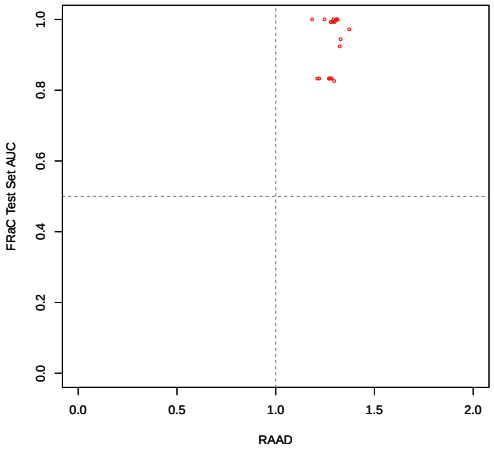

Figure 13: RAAD scatter plot for data set LEUKEMIA[17]: Transcriptional profiles from leukemias - ALL and AML. Task: Distinguish ALL and AML. Number of samples: 48. Normal: AML, Anomaly: ALL. See `http://www.broadinstitute.org/gsea/datasets.jsp`

14 lymphomas

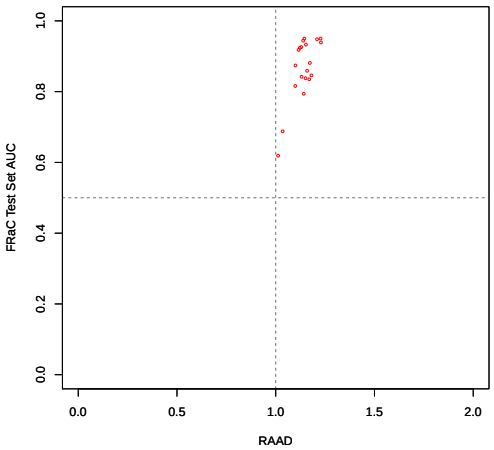

Figure 14: RAAD scatter plot for data set LYMPHOMAS[18]: Lymphoma types, taken on a special "lymphochip" (see Alizadeh et al., Nature, 2000). Task: Distinguish diffuse large B-cell lymphoma (DLBCL) from other types, follicular lymphoma (FL), and chronic lymphocytic leukaemia (CLL). Number of samples: 98. Normal: FL or CLL, Anomaly: one of two types of DLBCL. Note: Samples not labeled DLBCL, FL or CLL were omitted from this data set. Availability: `http://smd.stanford.edu/cgi-bin//publication/viewPublication.pl?pub_no=79`. See `http://www.ncbi.nlm.nih.gov/projects/geo/query/acc.cgi?acc=GSE60`. See `ftp://smd-ftp.stanford.edu/pub/smd/publications/79/183/exptset_183.meta`. See `ftp://smd-ftp.stanford.edu/pub/smd/publications/79/183/exptsetno_183.tar.gz`

15 meningiomas

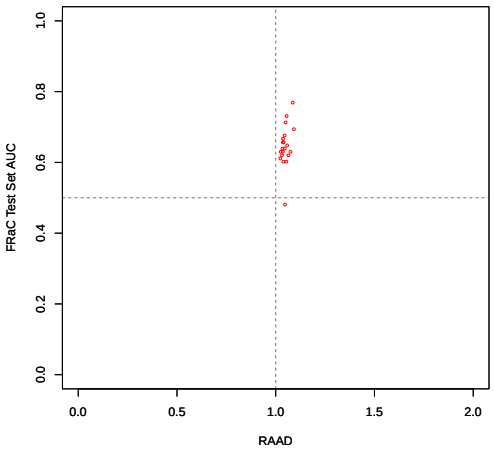

Figure 15: RAAD scatter plot for data set MENINGIOMAS: Meningioma data. Task: Distinguish recurrent meningioma (Grade 1,2 or 3) from primary (Grade 1,2 or 3). Number of samples: 56. Normal: primary, Anomaly: recurrent. Note: The samples have been public since 2006, but the GEO series page (`http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE4780`) does not include a citation, and its web link (`http://arrayconsortium.tgen.org/np2/viewProject.do?action=viewProject&projectId=183`) is dead. Availability: `http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE4780`

16 meta.1.2

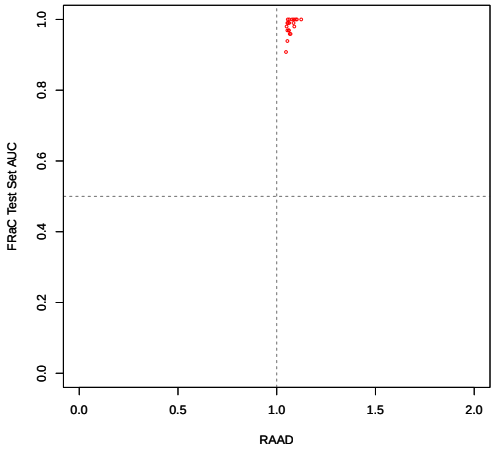

Figure 16: RAAD scatter plot for data set META.1.2[19]: (i) Expression data from ERBB2 over-expression and EGF stimulation in MCF10A (non-tumorigenic epithelial) cells (GEO series GSE14987); (ii) Expression data from DHT stimulation vs. control in LNCaP (androgen-sensitive human prostate adenocarcinoma) cells (GEO series GSE14988). Task: Identify prostate cell type among samples from breast, colon, lung, ovary, uterus. Number of samples: 61. Normal: breast, colon, lung, ovary, or uterus, Anomaly: prostate. Availability: `http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE14990`. See `http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE14987`. See `http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE14988`. See `http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE14990`

17 mind.body

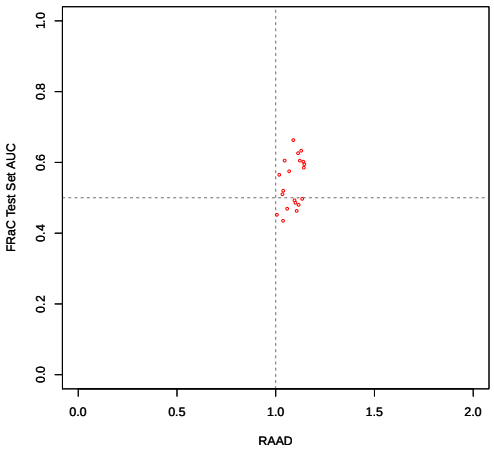

Figure 17: RAAD scatter plot for data set MIND.BODY[20]: Blood samples in various states of relaxation from people with various practice. Task: Identify samples from relaxed people. Number of samples: 72. Tissue type: blood. Normal: non-relaxed, Anomaly: relaxed. Availability: `http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE10041`

18 multitumor

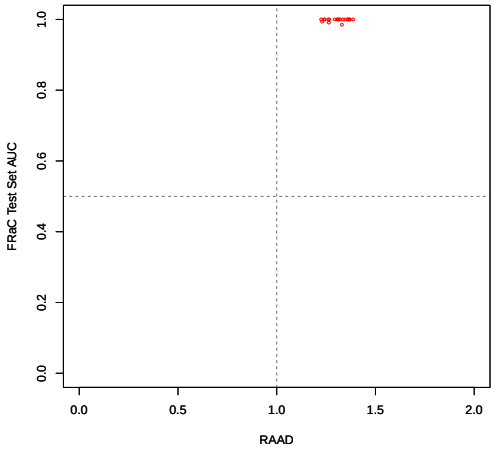

Figure 18: RAAD scatter plot for data set MULTITUMOR[21, 22, 23, 24, 25, 26]: Cancer samples from various tissue, including breast, skin, kidney, colon and blood. Task: Distinguish leukemia from other cancers. Number of samples: 134. Normal: other cancers, Anomaly: leukemia. Availability: `http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE2138`. Availability: `http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE781`. Availability: `http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE3189`. Availability: `http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE3744`

19 revlimid

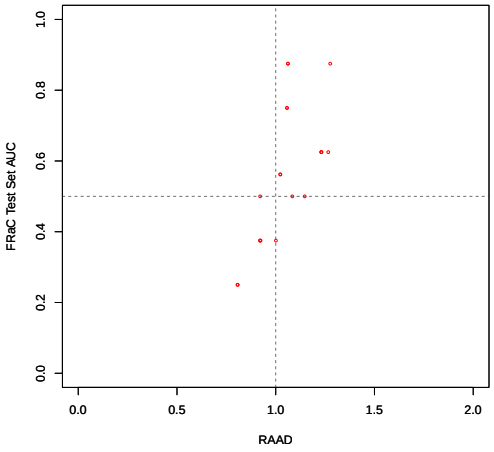

Figure 19: RAAD scatter plot for data set REVLIMID[27]: Blood samples, including Revlimid responders and non-responders. Task: Identify Revlimid responders. Number of samples: 16. Normal: non-responders, Anomaly: revlimid responders

20 ross2

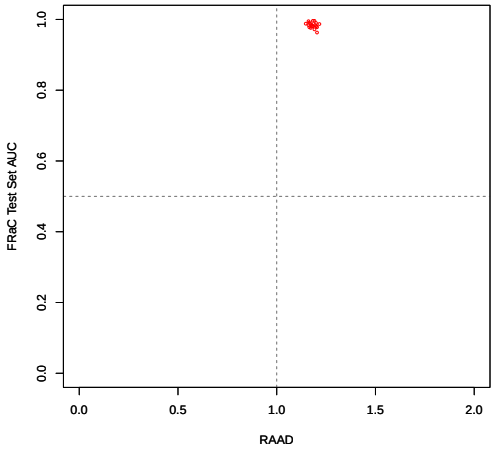

Figure 20: RAAD scatter plot for data set ROSS2[22]: AML and ALL Leukemia. Task: Distinguish AML from ALL. Number of samples: 170. Normal: ALL, Anomaly: AML

21 ross3

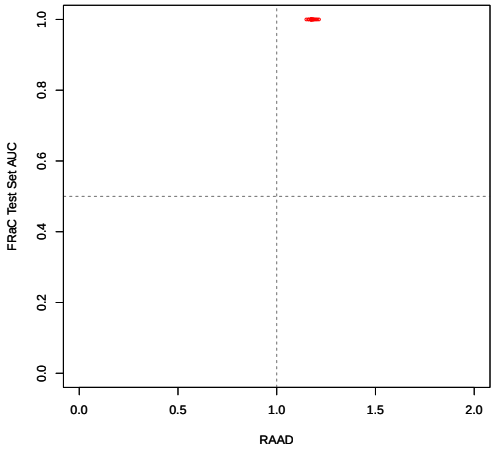

Figure 21: RAAD scatter plot for data set ROSS3[21]: Leukemia samples including AML and subtypes of ALL. Task: Distinguish AML from (all subtypes of) ALL leukemia samples. Number of samples: 30. Normal: any ALL subtype, Anomaly: AML. Note: GEO series GSE17703 contains a superset of these arrays. See `http://www.ncbi.nlm.nih.gov/projects/geo/query/acc.cgi?acc=GSE17703`

22 roth07

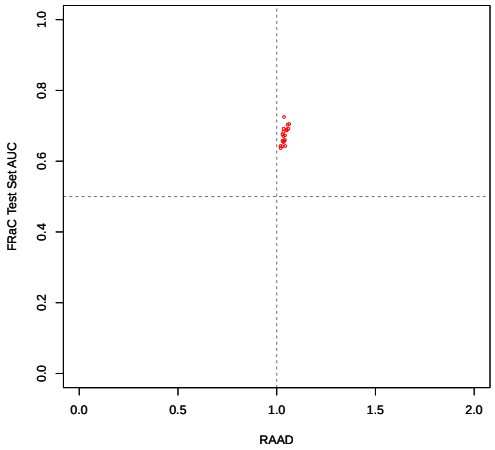

Figure 22: RAAD scatter plot for data set ROTH07: Various tissue types, normal and diseased, including: Accumbens, Adipose Tissue, Amygdala, Aorta, B Cells, Bone Marrow, Breast, Bronchus, CD4+, CD8+, Cerebellar Hemisphere AB, Cerebellar Vermis AB, Cerebellum, Cerebellum BD, Cerebral Cortex, Cervix, Colon BD, Colon Cecum, Coronary Artery, Corpus Callosum, Deltoid Muscle, Dorsal Root Ganglia, Endometrium, Endometrium/Ovary, Fallopian Tube, Fetal Brain BD, Fetal Liver BD, Frontal Cortex, Frontal Lobe, HUVEC Cell Line, Heart Atrium, Heart BD, Heart Ventricle, HepG2, Hippocampus, Hypothalamus, Joint Tissue Synovium, Kidney BD, Lung, MDA, Mammary Gland, Medulla, Midbrain, Monocytes, Myometrium, Nipple Cross Section, Nodose Nucleus, Occipital Lobe, Ovary, PBMC, Pancreas SG1, Paravertebral Muscle, Parietal Lobe, Penis, Pericardium, Peritneum, Pharyngeal Mucosa, Pituitary Gland, Placenta, Pons, Prefrontal Cortex, Prostate, Prostate Gland, Putamen, Retrocervical Infiltrate, Salivary Gland, Saphenous Vein, Skeletal Muscle, Skeletal Muscle Superior Quadracep, Skin, Small Intestine, Small Intestine Duodenum, Small Intestine Jejunum, Spinal Cord, Spleen, Stomach, Stomach, Substantia Nigra, Subthalamic Nucleus, Synovial Membrane, T Cells , Temporal Lobe, Testes, Thalamus, Thymus Gland, Thyroid Gland BD, Tongue, Tongue Main Corpus, Tongue Superior With Papillae, Tonsil, Trachea, Trigeminal Ganglia, Urethra, Uterus, Vagina, Vena Cava, Ventral Tegmental Area, Vestibular Nuclei Superior, Vulva. Task: Distinguish diseased human tissues from normal. Number of samples: 623. Normal: normal, Anomaly: diseased. Note: "Control" and "Treated" samples are removed in favor of only "Disease" and "Normal" samples. Note: Public since 2007, but citation is missing. Availability: `http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE7307`

23 sepsis

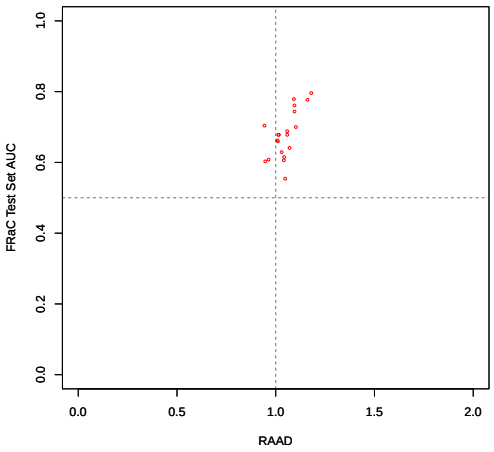

Figure 23: RAAD scatter plot for data set SEPSIS[28]: Leukocytes (some from patients with sepsis). Task: Identify sepsis, as opposed to experimental control (non-sepsis) samples. Number of samples: 94. Normal: control, Anomaly: sepsis. Note: There are many missing values, which are replaced automatically. Availability: `http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE5772`

24 shakes

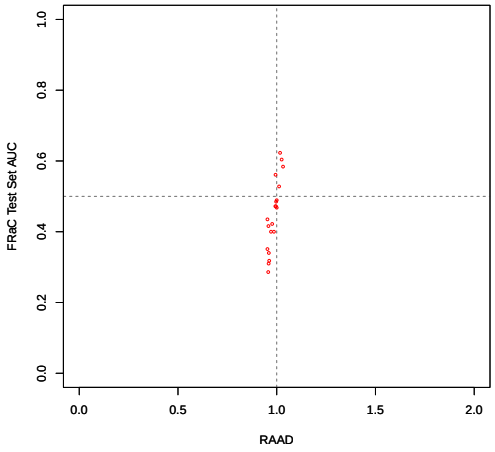

Figure 24: RAAD scatter plot for data set SHAKES[29]: Fasting venous peripherial blood mononuclear cell samples were collected at baseline and every 2 hours after intake of shakes containing SFA, MUFA or PUFA, up to eight hours after intake. All samples male age 18-27. Task: Identify experimental samples (intake of shakes containing SFA or PUFA) from baseline. Number of samples: 84. Normal: baseline, Anomaly: after taking SFA, MUFA or PUFA. Availability: `http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE13466`

25 smokers2

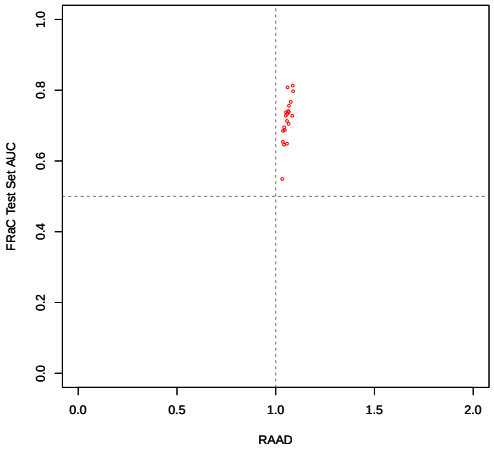

Figure 25: RAAD scatter plot for data set SMOKERS2[30]: Buccal mucosa from smokers and subjects who never smoked. Task: Distiguish smokers from (age- and gender-matched) never-smokers. Number of samples: 79. Normal: never smoked, Anomaly: smokers. Availability: `http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE17913`

26 smokers

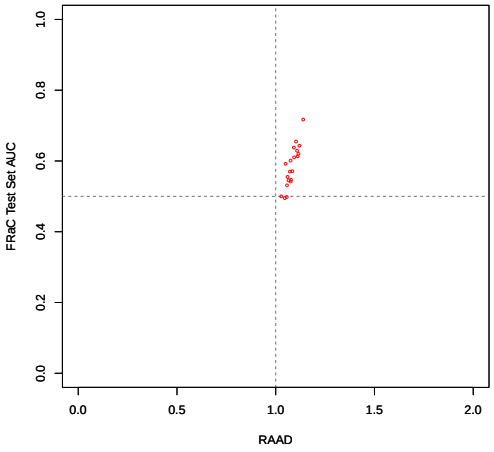

Figure 26: RAAD scatter plot for data set SMOKERS[31, 32]: Bronchial epithelium from smokers with and without lung cancer. Task: Distinguish lung cancer samples from those not diagnosed. Number of samples: 192. Normal: negative diagnosis, Anomaly: positive lung cancer diagnosis. Availability: `http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE4115`

27 survey

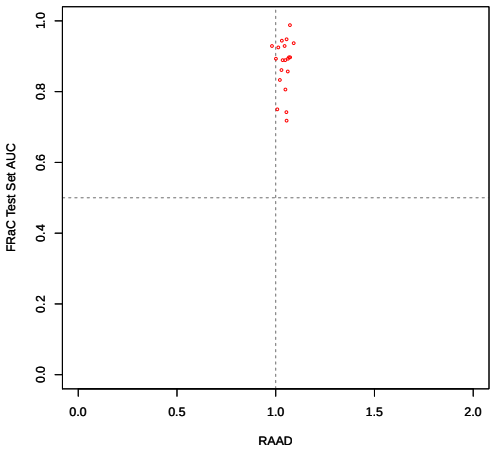

Figure 27: RAAD scatter plot for data set SURVEY[33]: Whole genome survey of 32 Human Tissues. Task: Distinguish fetal tissue from non-fetal. Number of samples: 96. Normal: non-fetal, Anomaly: fetal. Note: Tissues are: PBLs, UHR, adrenal gland, bone marrow, brain, colon, fetal brain, fetal kidney, fetal liver, fetal thymus, heart, kidney, liver, lung, mammary gland, ovary, pancreas, placenta, prostate, retina, salivary gland, skeletal muscle, skin, small intestine, spinal cord, spleen, testis, thymus, thyroid, tonsil, trachea, uterus. Availability: `http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE7905`

28 tzd

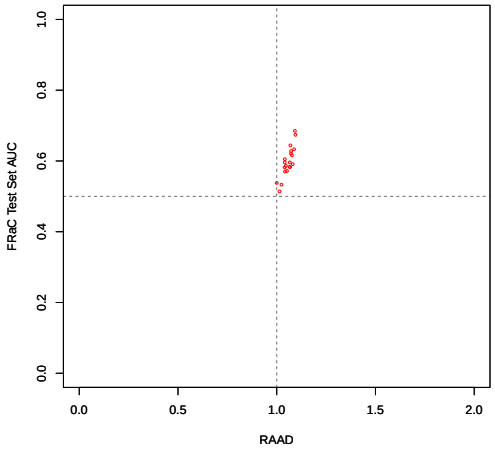

Figure 28: RAAD scatter plot for data set TZD[34]: Skeletal muscle from vastus lateralis, before and after hyperinsulinemic-euglycemic clamp, at baseline and after three-month thiazolidinedione (TZD) treatment. Task: Distinguish pre-TZD treatment from post-TZD. Number of samples: 254. Normal: post-TZD, Anomaly: pre-TZD treatment. Availability: `http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE13070`

References

[1] Jagani, Z., Mora-Blanco, E. L., Sansam, C. G., McKenna, E. S., Wilson, B., Chen, D., Klekota, J., Tamayo, P., Nguyen, P. T. L., Tolstorukov, M., Park, P. J., Cho, Y.-J., Hsiao, K., Buonamici, S., Pomeroy, S. L., Mesirov, J. P., Ruffner, H., Bouwmeester, T., Luchansky, S. J., Murtie, J., Kelleher, J. F., Warmuth, M., Sellers, W. R., Roberts, C. W. M., and Dorsch, M. Loss of the tumor suppressor snf5 leads to aberrant activation of the hedgehog-gli pathway. *Nature medicine* **16**, 1429–1433 (2010).

[2] Tamayo, P., Cho, Y.-J., Tsherniak, A., Greulich, H., Ambrogio, L., van Meeteren, N. S., Zhou, T., Buxton, A., Kool, M., Meyerson, M., Pomeroy, S. L., and Mesirov, J. P. Predicting relapse in patients with medulloblastoma by integrating evidence from clinical and genomic features. *Journal of Clinical Oncology* **29**, 1415–1423 (2011).

[3] Bild, A. H., Yao, G., Chang, J. T., Wang, Q., Potti, A., Chasse, D., Joshi, M., Harpole, D., Lancaster, J. M., Berchuck, A., J. A. Olson, J. R. M., Dressman, H. K., M.West, and Nevins, J. R. Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* **439**, 353–357 (2006).

[4] Lu, X., Lu, X., Wang, Z. C., Iglehart, J. D., Zhang, X., and Richardson, A. L. Predicting features of breast cancer with gene expression patterns. *Breast Cancer Research and Treatment* **108**, 191–201 (2008).

[5] Sørlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J. S., Nobel, A., Deng, S., Johnsen, H., Pesich, R., Geisler, S., Demeter, J., Perou, C. M., Lønning, P. E., Brown, P. O., Børresen-Dale, A.-L., and Botstein, D. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proceedings of the National Academy of Sciences, USA* **100**, 8418–8423 (2003).

[6] Wang, Y., Klijn, J. G. M., Zhang, Y., Sieuwerts, A. M., Look, M. P., Yang, F., Talantov, D., Timmermans, M., van Gelder, M. E. M., Yu, J., Jatkoe, T., Berns, E. M. J. J., Atkins, D., and Foekens, J. A. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* **365**, 671–679 (2005).

[7] Mootha, V., Lindgren, C., Eriksson, K.-F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstråle, M., Laurila, E., Houstis, N., Daly, M., Patterson, N., Mesirov, J., Golub, T. R., Tamayo, P., Spiegelman, B., Lander, E. S., Hirschhorn, J. N., Altshuler, D., and Groop, L. C. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genetics* **34**(3), 267–73 (2003).

[8] Bourquin, J.-P., Subramanian, A., Langebrake, C., Reinhardt, D., Bernard, O., Ballerini, P., Baruchel, A., Cavé, H., Dastugue, N., Hasle, H., Kaspers, G. L., Lessard, M., Michaux, L., Vyas, P., van Wering, E., Zwaan, C. M., Golub, T. R., and Orkin, S. H. Identification of distinct molecular phenotypes in acute megakaryoblastic leukemia by gene expression profiling. *Proceedings of the National Academy of Sciences, USA* **103**, 3339–3344 (2006).

[9] Niu, N., Qin, Y., Fridley, B. L., Hou, J., Kalari, K. R., Zhu, M., Wu, T.-Y., Jenkins, G. D., Batzler, A., and Wang, L. Radiation pharmacogenomics: a genome-wide association approach to identify radiation response biomarkers using human lymphoblastoid cell lines. *Genome research* **20**, 1482–1492 (2010).

[10] Kalari, K. R., Hebbring, S. J., Chai, H. S., Li, L., Kocher, J.-P. A., Wang, L., and Weinshilboum, R. M. Copy number variation and cytidine analogue cytotoxicity: a genome-wide association approach. *BMC Genomics* **11**, 357 (2010).

[11] Li, L., Fridley, B. L., Kalari, K., Jenkins, G., Batzler, A., Weinshilboum, R. M., and Wang, L. Gemcitabine and arabinosylcytosin pharmacogenomics: genome-wide association and drug response biomarkers. *PloS one* **4**, e7765 (2009).

[12] Fridley, B. L., Jenkins, G., Schaid, D. J., and Wang, L. A Bayesian hierarchical nonlinear model for assessing the association between genetic variation and drug cytotoxicity. *Statistics in medicine* **28**, 2709–2722 (2009).

[13] Li, L., Fridley, B., Kalari, K., Jenkins, G., Batzler, A., Safgren, S., Hildebrandt, M., Ames, M., Schaid, D., and Wang, L. Gemcitabine and cytosine arabinoside cytotoxicity: association with lymphoblastoid cell expression. *Cancer research* **68**, 7050–7058 (2008).

[14] Ingle, J. N., Schaid, D. J., Goss, P. E., Liu, M., Mushiroda, T., Chapman, J.-A. W., Kubo, M., Jenkins, G. D., Batzler, A., Shepherd, L., Pater, J., Wang, L., Ellis, M. J., Stearns, V., Rohrer, D. C., Goetz, M. P., Pritchard, K. I., Flockhart, D. A., Nakamura, Y., and Weinshilboum, R. M. Genome-wide associations and functional genomic studies of musculoskeletal adverse events in women receiving aromatase inhibitors. *Journal of clinical oncology* **28**, 4674–4682 (2010).

[15] Tan, X.-L., Moyer, A. M., Fridley, B. L., Schaid, D. J., Niu, N., Batzler, A. J., Jenkins, G. D., Abo, R. P., Li, L., Cunningham, J. M., Sun, Z., Yang, P., and Wang, L. Genetic variation predicting cisplatin cytotoxicity associated with overall survival in lung cancer patients receiving platinum-based chemotherapy. *Clinical cancer research* **17**, 5801–5811 (2011).

[16] Novershtern, N., Subramanian, A., Lawton, L. N., Mak, R. H., Haining, W. N., McConkey, M. E., Habib, N., Yosef, N., Chang, C. Y., Shay, T., Frampton, G. M., Drake, A. C. B., Leskov, I., Nilsson, B., Preffer, F., Dombkowski, D., Evans, J. W., Liefeld, T., Smutko, J. S., Chen, J., Friedman, N., Young, R. A., Golub, T. R., Regev, A., and Ebert, B. L. Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell* **144**, 296–309 (2011).

[17] Armstrong, S. A., Staunton, J. E., Silverman, L. B., Pieters, R., den Boer, M. L., Minden, M. D., Sallan, S. E., Lander, E. S., Golub, T. R., and Korsmeyer, S. J. MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. *Nature genetics* **30**, 41–47 (2002).

[18] Alizadeh, A. A., Eisen, M. B., Davis, R. E., Ma, C., Lossos, I. S., Rosenwald, A., Boldrick, J. C., Sabet, H., Tran, T., Yu, X., Powell, J. I., Yang, L., Marti, G. E., Moore, T., Hudson, J., Lu, L., Lewis, D. B., Tibshirani, R., Sherlock, G., Chan, W. C., Greiner, T. C., Weisenburger, D. D., Armitage, J. O., Warnke, R., Levy, R., Wilson, W., Grever, M. R., Byrd, J. C., Botstein, D., Brown, P. O., and Staudt, L. M. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**, 503–511 (2000).

[19] Wolfer, A., Wittner, B. S., Irimia, D., Flavin, R. J., Lupien, M., Gunawardane, R. N., Meyer, C. A., Lightcap, E. S., Tamayo, P., Mesirov, J. P., Liu, X. S., Shioda, T., Toner, M., Loda, M., Brown, M., Brugge, J. S., and Ramaswamy, S. MYC regulation of a "poor-prognosis" metastatic cancer cell state. *Proceedings of the National Academy of Sciences, USA* **107**, 3698–3703 (2010).

[20] Dusek, J. A., Otu, H. H., Wohlhueter, A. L., Bhasin, M., Zerbini, L. F., Joseph, M. G., Benson, H., and Libermann, T. A. Genomic counter-stress changes induced by the relaxation response. *PloS one* **3**, e2576 (2008).

[21] Ross, M. E., Zhou, X., Song, G., Shurtleff, S. A., Girtman, K., Williams, W. K., Liu, H.-C., Mahfouz, R., Raimondi, S. C., Lenny, N., Patel, A., and Downing, J. R. Classification of pediatric acute lymphoblastic leukemia by gene expression profiling. *Blood* **102**, 2951–2959 (2003).

[22] Ross, M. E., Mahfouz, R., Onciu, M., Liu, H.-C., Zhou, X., Song, G., Shurtleff, S. A., Pounds, S., Cheng, C., Ma, J., Ribeiro, R. C., Rubnitz, J. E., Girtman, K., Williams, W. K., Raimondi, S. C., Liang, D.-C., Shih, L.-Y., Pui, C.-H., and Downing, J. R. Gene expression profiling of pediatric acute myelogenous leukemia. *Blood* **104**, 3679–3687 (2004).

[23] Koinuma, K., Yamashita, Y., Liu, W., Hatanaka, H., Kurashina, K., Wada, T., Takada, S., Kaneda, R., Choi, Y. L., Fujiwara, S.-I., Miyakura, Y., Nagai, H., and Mano, H. Epigenetic silencing of AXIN2 in colorectal carcinoma with microsatellite instability. *Oncogene* **25**, 139–146 (2006).

[24] Lenburg, M. E., Liou, L. S., Gerry, N. P., Frampton, G. M., Cohen, H. T., and Christman, M. F. Previously unidentified changes in renal cell carcinoma gene expression identified by parametric analysis of microarray data. *BMC Cancer* **3**, 31 (2003).

[25] Talantov, D., Mazumder, A., Yu, J. X., Briggs, T., Jiang, Y., Backus, J., Atkins, D., and Wang, Y. Novel genes associated with malignant melanoma but not benign melanocytic lesions. *Clinical Cancer Research* **11**, 7234–7242 (2005).

[26] Alimonti, A., Carracedo, A., Clohessy, J. G., Trotman, L. C., Nardella, C., Egia, A., Salmena, L., Sampieri, K., Haveman, W. J., Brogi, E., Richardson, A. L., Zhang, J., and Pandolfi, P. P. Subtle variations in pten dose determine cancer susceptibility. *Nature genetics* **42**, 454–458 (2010).

[27] Ebert, B. L., Galili, N., Tamayo, P., Bosco, J., Mak, R., Pretz, J., Tanguturi, S., Ladd-Acosta, C., Stone, R., Golub, T. R., and Raza, A. An erythroid differentiation signature predicts response to lenalidomide in myelodysplastic syndrome. *PLoS medicine* **5**, e35 (2008).

[28] Tang, B. M. P., McLean, A. S., Dawes, I. W., Huang, S. J., and Lin, R. C. Y. The use of gene-expression profiling to identify candidate genes in human sepsis. *American journal of respiratory and critical care medicine* **176**, 676–684 (2007).

[29] Bouwens, M., Bromhaar, M. G., Jansen, J., Müller, M., and Afman, L. A. Postprandial dietary lipid-specific effects on human peripheral blood mononuclear cell gene expression profiles. *The American journal of clinical nutrition* **91**, 208–217 (2010).

[30] Boyle, J. O., Gümüs, Z. H., Kacker, A., Choksi, V. L., Bocker, J. M., Zhou, X. K., Yantiss, R. K., Hughes, D. B., Du, B., Judson, B. L., Subbaramaiah, K., and Dannenberg, A. J. Effects of cigarette smoke on the human oral mucosal transcriptome. *Cancer prevention research* **3**, 266–278 (2010).

[31] Spira, A., Beane, J. E., Shah, V., Steiling, K., Liu, G., Schembri, F., Gilman, S., Dumas, Y.-M., Calner, P., Sebastiani, P., Sridhar, S., Beamis, J., Lamb, C., Anderson, T., Gerry, N., Keane, J., Lenburg, M. E., and Brody, J. S. Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer. *Nature medicine* **13**, 361–366 (2007).

[32] Gustafson, A. M., Soldi, R., Anderlind, C., Scholand, M. B., Qian, J., Zhang, X., Cooper, K., Walker, D., McWilliams, A., Liu, G., Szabo, E., Brody, J., Massion, P. P., Lenburg, M. E., Lam, S., Bild, A. H., and Spira, A. Airway PI3K pathway activation is an early and reversible event in lung cancer development. *Science translational medicine* **2**, 26ra25 (2010).

[33] Dezso, Z., Nikolsky, Y., Sviridov, E., Shi, W., Serebriyskaya, T., Dosymbekov, D., Bugrim, A., Rakhmatulin, E., Brennan, R. J., Guryanov, A., Li, K., Blake, J., Samaha, R. R., and Nikolskaya, T. A comprehensive functional analysis of tissue specificity of human gene expression. *BMC Biology* **6**, 49 (2008).

[34] Sears, D. D., Hsiao, G., Hsiao, A., Yu, J. G., Courtney, C. H., Ofrecio, J. M., Chapman, J., and Subramaniam, S. Mechanisms of human insulin resistance and thiazolidinedione-mediated insulin sensitization. *Proceedings of the National Academy of Sciences, USA* **106**, 18745–18750 (2009).