



Published in final edited form as:

Cancer Epidemiol Biomarkers Prev. 2021 April ; 30(4): 751–756. doi:10.1158/1055-9965.EPI-20-1461.

Gene Expression Pathways in Prostate Tissue Associated with Vigorous Physical Activity in Prostate Cancer

Claire H. Parnar^a, Giovanni Parmigiani^{b,c}, Edward L. Giovannucci^{a,d,e}, Eric B. Rimm^{a,d,e}, Svitlana Tyekucheva^{b,c}, Massimo Loda^f, Stephen P. Finn^{g,h}, Matthew G. Vander Heiden^{i,j}, Michelangelo Fiorentino^{h,k}, Ericka M. Ebot^{a,*}, Lorelei A. Mucci^{a,e,*}

^aDepartment of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA

^bDepartment of Data Sciences, Dana-Farber Cancer Institute, Boston, MA ^cDepartment of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA ^dDepartment of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA ^eChanning Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA ^fDepartment of Pathology and Laboratory Medicine, Weill Cornell Medicine, New York Presbyterian Hospital, New York, NY ^gDepartment of Histopathology, Trinity College, Dublin, Ireland ^hCenter for Molecular Oncologic Pathology, Dana-Farber Cancer Institute, Boston, MA ⁱKoch Institute for Cancer Research and Department of Biology, Massachusetts Institute of Technology, Cambridge, MA ^jDepartment of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA ^kPathology Unit, Addarii Institute, S. Orsola-Malpighi Hospital, Bologna, Italy

Abstract

Background: Men engaged in high physical activity have lower risks of advanced and fatal prostate cancer. Mechanisms underlying this association are not well understood but may include systemic and tumor-specific effects. We investigated potential mechanisms linking physical activity and gene expression in prostate tissue from men with prostate cancer.

Methods: We included a subset of 118 men in the Health Professionals Follow-up Study diagnosed with prostate cancer between 1986–2005 with whole transcriptome gene expression profiling on tumor and adjacent normal prostate tissue and physical activity data. Long-term vigorous physical activity was self-reported as the average time spent engaged in various forms of recreational physical activity at baseline and biennially until prostate cancer diagnosis. Gene set enrichment analysis was performed among KEGG and Hallmark gene sets to identify pathways with differential expression based on vigorous physical activity.

Results: In adjacent normal tissue, we identified 25 KEGG gene sets enriched (down-regulated) in the highest compared to lowest quintile of vigorous physical activity at a false-discovery rate <0.10, including a number of cancer- and immune-related pathways. While no gene sets reached

Corresponding author: Claire H. Parnar, MPH, ScD, Department of Epidemiology, Harvard T.H. Chan School of Public Health, 677 Huntington Ave, Boston, MA 02115, cpernar@mail.harvard.edu.

*Shared last authorship

Conflict of interest disclosure:

The authors declare no potential conflicts of interest.

statistical significance in tumor tissue, top gene sets differentially expressed included transforming growth factor (TGF) beta, apoptosis, and p53 signaling pathways.

Conclusions: These findings suggest that physical activity may influence the tumor microenvironment. Future studies are needed to confirm these findings and further investigate potential mechanisms linking physical activity to lethal prostate cancer.

Impact: Identification of gene expression alterations in the prostate associated with physical activity can improve our understanding of prostate cancer etiology.

Introduction

Physical activity is associated with lower risk of several cancers¹, but much remains unknown regarding the biological mechanisms involved. Epidemiologic studies support a link between physical activity and reduced prostate cancer incidence and progression.^{2,3} In a prospective study, men who engaged in higher amounts of vigorous physical activity had a lower risk of clinically-significant and *TMPRSS2:ERG*-positive prostate cancer.⁴ Physical activity influences several biological processes that may be involved in prostate cancer^{5,6}, and may exert both systemic and local effects.⁷ Specifically, it alters endogenous hormone levels, including testosterone, insulin, and insulin-like growth factor (IGF)-1.⁵⁻⁷ Furthermore, physical activity of different types and duration may have different effects on the tumor microenvironment.⁸

While its effects on the prostate, at this time, are unclear, there is emerging evidence that physical activity is linked to epigenetic changes in prostate tissue.^{9,10} Epigenetic modifications such as DNA methylation and histone modifications are known to disrupt key biological processes in cancers, such as tumor growth, tissue invasion, and metastasis, and contribute to the occurrence of genetic mutations.¹¹ There is growing evidence that lifestyle factors, such as diet and physical activity, have an important role in influencing epigenetic processes.¹² Additional studies by our group and others have shown that activity at higher intensity and over the long-term may be most relevant to decreasing prostate cancer development and improving outcomes.^{4,13,14} There is limited evidence regarding the relationship between physical activity and gene expression in prostate cancer. Identification of gene expression alterations in the prostate that are associated with physical activity would potentially strengthen the biologic plausibility of these associations and may facilitate identification of biomarkers related to prostate cancer development.

The present study aimed to investigate the potential relationship between long-term, pre-diagnosis vigorous activity and gene expression alterations in prostate tumor and adjacent normal tissue. We broadly hypothesized that physical activity may alter gene expression in prostate tissue that may contribute to the beneficial association between physical activity and prostate cancer risk.

Materials and Methods

Study Population

The Health Professionals Follow-up Study (HPFS) is an ongoing prospective cohort of 51,529 US male health professionals aged 40 to 75 years at baseline in 1986. Participants completed questionnaires at baseline and were mailed questionnaires every two years thereafter to ascertain lifestyle, health-related factors, and disease outcomes. The study protocol was approved by institutional review boards of the Brigham and Women's Hospital and Harvard T.H. Chan School of Public Health, and those of participating registries as required.

For HPFS participants diagnosed with prostate cancer during follow-up, we contacted hospital pathology departments to retrieve archival formalin-fixed paraffin-embedded (FFPE) prostate tumor tissue obtained through either radical prostatectomy or transurethral resection of the prostate. Gene expression profiling was performed on a subset of cases diagnosed between 1982–2005 using an extreme case design with the aim of identifying expression signatures to differentiate indolent and lethal disease.¹⁵ These cases included men with prostate cancer who had a metastatic event or died of prostate cancer during follow-up through 2016 (lethal cases), and men who survived at least 8 years without any evidence of metastases during follow-up (indolent cases). Gene expression profiling of tumor tissue was obtained for a total of 254 men in HPFS. Of these men, a subset of 120 men had profiling of both tumor tissue and adjacent normal tissue. To improve the comparability between our analyses of tumor and normal tissue, only men with both tumor tissue and matched adjacent normal tissue were included. After exclusion of 2 men who did not provide baseline physical activity data, our primary analysis included 118 cases (44 lethal and 74 indolent) with tumor and adjacent normal tissue.

Gene expression profiling

Using archival FFPE tissues from this subset of cases, cores were taken from foci highly enriched for cancer and mRNA was extracted from areas selected by study pathologists, as previously described.^{16,17} Adjacent normal tissue was defined as prostatic tissue with histologic features closest to normal prostatic glands and stroma and clearly separated (at least 5 mm) from the cancer counterpart. In cases with inflammation, the area with less inflammation was selected. We performed whole-transcriptome amplification (WT-Ovation FFPE System V2; NuGEN) followed by hybridization to the GeneChip Human Gene 1.0 ST Array (Affymetrix).^{18,19} For the expression profiles, we regressed out technical variables including mRNA concentration, age of the block, batch (96-well plate), percentage of probes detectable above the background, log-transformed average background signal, and the median of the perfect match probes for each probe intensity of the raw data. The residuals were shifted to the original mean expression values and normalized using the robust multichip average (RMA) method.^{20,21} We used NetAffx annotations to map gene names to Affymetrix transcript cluster IDs as implemented in Bioconductor annotation package *pd.hugene.1.0.st.v1*; this resulted in 20,254 unique gene names. Gene expression data are available through Gene Expression Omnibus (GSE79021).

Clinical data ascertainment

Incident prostate cancer was ascertained initially by self-report on questionnaires. Medical records and pathology reports were used to confirm prostate cancer diagnosis and to extract clinical and treatment information. Since 2000, participants diagnosed with prostate cancer were followed through biennial disease-specific questionnaires for development of metastases, treatment, and PSA levels. Prostate cancer-specific death was determined by review of death certificates and medical records by an endpoint committee of physicians. The study pathologists reviewed hematoxylin-and-eosin slides to provide uniform Gleason grade and histopathologic review. Lethal prostate cancer was defined as distant metastasis or death due to the disease with follow-up through December 2016.

Physical activity assessment

In the HPFS, physical activity was assessed through validated questionnaires beginning at baseline in 1986 and every two years thereafter.²² Participants were asked to report in categories the average total time per week engaged in specific recreational activities during the past year. Specific activities included walking or hiking outdoors, jogging (>10 minutes/mile), running (10 minutes/mile), bicycling, lap swimming, tennis, squash or racquetball, and calisthenics or rowing. Participants also reported their usual walking pace and the number of flights of stairs climbed daily. Additional specific activities were included on the questionnaire in subsequent cycles: heavy outdoor work (e.g., digging or chopping) from 1988, weightlifting from 1990, moderate outdoor work (e.g., yardwork or gardening) from 2004.

To quantify intensity of activity, each specific activity was assigned a metabolic equivalent of task (MET) value based on a compendium of physical activities.²³ A measure of MET-hours per week was derived for each specific activity by multiplying the MET value assigned for that activity by the average number of hours per week reported by the participant. Vigorous activity was restricted to activities with a MET value of 6 or greater: walking at a pace of 4 miles per hour or faster, jogging, running, bicycling, lap swimming, tennis, squash or racquetball, calisthenics or rowing, and stair climbing. A validation study in the HPFS showed that vigorous activity may be measured with better validity than lower intensity activity.²²

To examine the long-term association with physical activity in our primary analyses, we used cumulative average physical activity, using the average of all available questionnaire data from baseline until the time of prostate cancer diagnosis. In our primary analysis, we considered physical activity categorized into quintiles and compared the highest to the lowest category of activity. This was based on previous findings by our group of association with prostate cancer risk comparing extreme quintiles of vigorous activity.^{4,13} In analyses stratified by lethal status, we used continuous vigorous physical activity to maximize the statistical power of this analysis. In secondary analyses, we also evaluated the association with recent physical activity, using only physical activity reported on the questionnaire completed most recently prior to cancer diagnosis.

Statistical analysis

To identify predefined sets of functionally related genes associated with long-term, pre-diagnosis vigorous activity, gene set enrichment analysis (GSEA) was performed on mRNA expression profiles of tumor and adjacent normal tissue.²⁴ Prior to performing GSEA, we filtered genes to exclude those with low expression across 50% or more samples, separately for tumor and normal tissue. A total of 6,167 genes were included in the tumor tissue analysis, and 6,215 genes in the adjacent normal tissue analysis. Age is a potential confounder in this study because it may affect gene expression as well as vigorous activity. To remove variation in gene expression due to age, we obtained gene expression residuals from linear regression on age at diagnosis. As secondary analyses, we performed GSEA among lethal cases and among indolent cases, separately.

As determined *a priori*, we included 186 Kyoto Encyclopedia of Genes and Genomes (KEGG) and 50 Hallmark gene sets from the Molecular Signature Database version 7.0 with software from the Broad Institute (<http://software.broadinstitute.org/gsea/index.jsp>). Gene sets with fewer than 15 or more than 500 genes (KEGG $n=79$; Hallmark $n=3$) were excluded. The signal-to-noise metric was used to rank the genes in analyses comparing the highest to lowest quintile of vigorous activity. Pearson correlations were used to rank genes in analyses with continuous vigorous activity. An enrichment score (ES) was calculated for each gene set using a weighted Kolmogorov-Smirnoff statistic. The ES represents how much the gene set is overrepresented at the top or bottom of the ranked list of genes. A positive ES indicated gene set enrichment at the top of the ranked list (upregulated in high vs. low activity) while a negative ES indicated gene set enrichment at the bottom of the ranked list (downregulated in high vs. low activity). The top-ranked subset of genes contributing to the ES were considered the leading edge genes. The significance level was estimated using 10,000 phenotype-based permutations. The ES was normalized to account for variable size of gene sets and multiple testing was accounted for by calculating the false discovery rate (FDR). Cytoscape version 3.7.1 (www.cytoscape.org) was used to visualize results from GSEA. To evaluate differential expression of individual genes by activity level, we used linear regression as implemented in the *limma* Bioconductor package. All other analyses were performed in R version 3.1.0. Gene sets with $FDR < 0.10$ were considered statistically significant.

Results

Table 1 shows the characteristics of the subset of prostate cancer cases at diagnosis among whom gene expression profiling was performed. Men in the lowest quintile of vigorous activity were somewhat older, had slightly higher PSA level at diagnosis, more likely to be diagnosed with advanced stage disease, and more likely to have lethal disease compared to men in the highest quintile. Other clinical characteristics were similar among men in the highest compared to the lowest quintile of vigorous activity. Results from the individual gene analysis were largely non-significant; therefore, we focused on the pathway analysis (Supplemental Table 1).

Overall, we observed a signal in both tumor and normal tissue as illustrated by enrichment of nominally significant GSEA p-values in both tissue types (Supplemental Figure 1). These

p-values indicate that while some pathways are common between tumor and normal tissue, others are not (Figure 1). Controlling for multiple testing using the FDR, we did not identify any significant gene sets in tumor tissue at an $FDR < 0.10$.

In normal tissue, we identified 25 KEGG gene sets down-regulated among men in the highest versus lowest quintile of vigorous activity ($FDR < 0.10$) (Figure 2; Supplemental Table 2). Among the sets identified were eight cancer-defined pathways and several pathways related to cellular immune response, such as B and T cell receptor signaling, and signal transduction. The extensive overlap of genes across these pathways is depicted in Figure 2. Further analysis of the leading-edge genes showed *MAPK1*, *RAF1*, *PIK3R3*, *HRAS*, *AKT2*, *SOS1* and *SOS2* were present in more than half of the 25 gene sets identified in normal tissue (Figure 3).

In analyses stratified by lethal status, we identified 25 significant gene sets in normal tissue of lethal cases but none in that of indolent cases (Supplemental Tables 3–4). Of the 25 gene sets, 12 were identified in the analysis of adjacent normal tissue overall, while 13 were unique to the lethal cases. When we examined recent vigorous activity, we identified 11 gene sets significantly enriched ($FDR < 0.10$) in adjacent normal tissue, fewer gene sets than we identified in association with long-term activity (Supplemental Table 5).

We conducted a sensitivity analysis in adjacent normal tissue excluding 15 cases with gene expression assayed from TURP specimens. There were fewer gene sets (21) that reached significance at $FDR < 0.10$, likely due to the decreased sample size. The top gene sets were largely the same as when both TURP and RP specimens were included (Supplement Table 6). None of the Hallmark gene sets reached statistical significance at $FDR < 0.10$ in tumor or adjacent normal tissue (Supplemental Table 7).

Discussion

Our study evaluated the associations between long-term vigorous activity prior to prostate cancer diagnosis and gene expression alterations in tumor and adjacent normal tissue. We identified 25 KEGG gene sets that were down-regulated in men who reported high compared to low vigorous activity in the tumor-adjacent normal tissue. These gene sets included several cancer-related, immune system, and signal transduction pathways. To our knowledge, our study is the first to examine the associations between long-term, pre-diagnostic vigorous activity and gene expression alterations in prostate tissue.

Physical activity, particularly long-term activity, has been linked to epigenetic changes in breast, gastric, and colorectal cancers.¹² There is limited evidence regarding the relationship between physical activity and gene expression in prostate cancer. One study examining prostate gene expression and post-diagnostic vigorous activity in low-risk prostate cancer identified associations with genes related to cell signaling, metabolism and DNA repair pathways.¹⁰ In a cohort of men with localized prostate cancer, pre-diagnostic vigorous activity was associated with DNA methylation in the *CRACR2A* gene.⁹

We identified several cancer-related pathways that are down-regulated in adjacent normal tissue among men who performed long-term high compared to low amounts of vigorous

activity. These findings suggest that physical activity may affect biological pathways that are common across different cancers. For example, the mitogen-activated protein kinase (MAPK), phosphatidylinositol 3-kinase (PI3K)-Akt, and p53 signaling pathways are common to many of the cancer gene sets identified in our study. The present study also identified several immune system pathways down-regulated in men engaged in high versus low vigorous activity. This is consistent with prior studies showing that physical activity may reduce oxidative stress and improve immune functions.⁷

Our leading-edge analysis showed that many of the gene sets identified by GSEA had leading edge genes in common. Consideration of these individual genes provides additional understanding of the potential biological effects of physical activity. *MAPK1*, a member of the MAPK signaling pathway, and *PIK3R3*, a member of the PI3K pathway, were identified as a leading edge gene in our analysis. Upregulation of MAPK pathway is associated with reduced survival in castration resistant prostate cancer.²⁵ The PI3K pathway regulates cell growth, survival, proliferation, metabolism, and angiogenesis, and is altered in metastatic prostate cancer.²⁶ *TP53*, another leading edge gene, is responsible for the tumor protein p53, a well-known tumor suppressor gene. Regular exercise may alter the IGF-axis and reduce cell proliferation and increase apoptosis through expression of p53 in prostate cancer cells.²⁷

We identified differentially expressed gene sets in adjacent normal tissue while enriched gene sets in tumor tissue did not reach significance. Given its proximity to the tumor, adjacent normal tissue may exhibit similar characteristics or undergo processes similar to those in the tumor. It is possible that, by the time that cancer is detectable, cellular activity is dysregulated to such an extent that relatively small differences in gene expression due to lifestyle factors become more difficult to detect. There may also be an interaction between normal and tumor tissue that affects disease progression.

Strengths of this study include its long-term, prospective physical activity assessment and ability to examine tumor and normal tissue. Compared to individual gene analysis, GSEA improves power when effects of individual genes are small and facilitates interpretation of biological mechanisms.²⁴ There are limitations to this study. Our sample size was moderate and may have limited our power to detect significant associations. This was an observational study; therefore, we cannot conclude that differences in vigorous activity are causally related to epigenetic changes in prostate tissue. Although physical activity was assessed by self-report using a validated questionnaire, there may be misclassification of men according to their physical activity level. Additionally, because participants in the HPFS are primarily Caucasian, our findings may not be generalizable to other racial/ethnic groups.

In conclusion, our study identified sets of functionally related genes differentially expressed in tumor and adjacent normal prostate tissue among men who engage in high compared to low vigorous activity. This exploratory study suggests that vigorous activity may influence prostate tissue through changes in cancer-related, immune system, cell signaling and other pathways in men with prostate cancer.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported in part by National Institutes of Health (T32 ES 007069 and T32 CA 009001 to C. Pernar, P50 CA 090381 to L. Mucci, R01 CA 136578 to L. Mucci, R01 CA 174206 to G. Parmigiani, 4P30CA006516–51 to L. Mucci, G. Parmigiani); the Prostate Cancer Foundation Young Investigator Awards to L. Mucci, S. Finn; the Prostate Cancer Foundation Challenge Award to L.A. Mucci. The Health Professionals Follow-up Study is supported by U01CA167552 from the National Cancer Institute.

We would like to thank the participants and staff of the HPFS for their valuable contributions as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY. In particular, we would like to acknowledge Elizabeth Frost-Hawes, Sioban Saint-Surin, Eleni Konstantis, Liza Gazeeva, Robert Sheahan, Ann Fischer, and Ruifeng Li. The authors assume full responsibility for analyses and interpretation of these data.

References

1. Moore SC, Lee IM, Weiderpass E, Campbell PT, Sampson JN, Kitahara CM, et al. Association of Leisure-Time Physical Activity With Risk of 26 Types of Cancer in 1.44 Million Adults. *JAMA Intern Med.* 2016;176(6):816–825. [PubMed: 27183032]
2. Liu Y, Hu F, Li D, Wang F, Zhu L, Chen W, et al. Does physical activity reduce the risk of prostate cancer? A systematic review and meta-analysis. *Eur Urol.* 2011;60(5):1029–1044. [PubMed: 21802197]
3. Bonn SE, Sjolander A, Lagerros YT, Wiklund F, Stattin P, Holmberg E, et al. Physical activity and survival among men diagnosed with prostate cancer. *Cancer Epidemiol Biomarkers Prev.* 2015;24(1):57–64. [PubMed: 25527697]
4. Pernar CH, Ebot EM, Pettersson A, Graff RE, Giunchi F, Ahearn TU, et al. A Prospective Study of the Association between Physical Activity and Risk of Prostate Cancer Defined by Clinical Features and TMPRSS2:ERG. *Eur Urol.* 2019;76(1):33–40. [PubMed: 30301696]
5. Gann PH, Hennekens CH, Ma J, Longcope C, Stampfer MJ. Prospective study of sex hormone levels and risk of prostate cancer. *J Natl Cancer Inst.* 1996;88(16):1118–1126. [PubMed: 8757191]
6. Barnard RJ, Ngo TH, Leung PS, Aronson WJ, Golding LA. A low-fat diet and/or strenuous exercise alters the IGF axis in vivo and reduces prostate tumor cell growth in vitro. *Prostate.* 2003;56(3):201–206. [PubMed: 12772189]
7. McTiernan A Mechanisms linking physical activity with cancer. *Nat Rev Cancer.* 2008;8(3):205–211. [PubMed: 18235448]
8. Koelwyn GJ, Quail DF, Zhang X, White RM, Jones LW. Exercise-dependent regulation of the tumour microenvironment. *Nat Rev Cancer.* 2017;17(10):620–632. [PubMed: 28943640]
9. Dai JY, Wang B, Wang X, Cheng A, Kolb S, Stanford JL, et al. Vigorous Physical Activity Is Associated with Lower Risk of Metastatic-Lethal Progression in Prostate Cancer and Hypomethylation in the CRACR2A Gene. *Cancer Epidemiol Biomarkers Prev.* 2019;28(2):258–264. [PubMed: 30464020]
10. Magbanua MJ, Richman EL, Sosa EV, Jones LW, Simko J, Shinohara K, et al. Physical activity and prostate gene expression in men with low-risk prostate cancer. *Cancer Causes Control.* 2014;25(4):515–523. [PubMed: 24504435]
11. Timp W, Feinberg AP. Cancer as a dysregulated epigenome allowing cellular growth advantage at the expense of the host. *Nat Rev Cancer.* 2013;13(7):497–510. [PubMed: 23760024]
12. Grazioli E, Dimauro I, Mercatelli N, Wang G, Pitsiladis Y, Di Luigi L, et al. Physical activity in the prevention of human diseases: role of epigenetic modifications. *BMC Genomics.* 2017;18(Suppl 8):802. [PubMed: 29143608]

13. Giovannucci EL, Liu Y, Leitzmann MF, Stampfer MJ, Willett WC. A prospective study of physical activity and incident and fatal prostate cancer. *Arch Intern Med*. 2005;165(9):1005–1010. [PubMed: 15883238]
14. Richman EL, Kenfield SA, Stampfer MJ, Paciorek A, Carroll PR, Chan JM. Physical activity after diagnosis and risk of prostate cancer progression: data from the cancer of the prostate strategic urologic research endeavor. *Cancer Res*. 2011;71(11):3889–3895. [PubMed: 21610110]
15. Lu D, Sinnott JA, Valdimarsdottir U, Fang F, Gerke T, Tyekucheva S, et al. Stress-Related Signaling Pathways in Lethal and Nonlethal Prostate Cancer. *Clin Cancer Res*. 2016;22(3):765–772. [PubMed: 26490316]
16. Penney KL, Sinnott JA, Tyekucheva S, Gerke T, Shui IM, Kraft P, et al. Association of prostate cancer risk variants with gene expression in normal and tumor tissue. *Cancer Epidemiol Biomarkers Prev*. 2015;24(1):255–260. [PubMed: 25371445]
17. Tyekucheva S, Martin NE, Stack EC, Wei W, Vathipadiekal V, Waldron L, et al. Comparing Platforms for Messenger RNA Expression Profiling of Archival Formalin-Fixed, Paraffin-Embedded Tissues. *J Mol Diagn*. 2015;17(4):374–381. [PubMed: 25937617]
18. Hall JS, Leong HS, Armenoult LS, Newton GE, Valentine HR, Irlam JJ, et al. Exon-array profiling unlocks clinically and biologically relevant gene signatures from formalin-fixed paraffin-embedded tumour samples. *British journal of cancer*. 2011;104(6):971–981. [PubMed: 21407225]
19. Kennedy RD, Bylesjo M, Kerr P, Davison T, Black JM, Kay EW, et al. Development and independent validation of a prognostic assay for stage II colon cancer using formalin-fixed paraffin-embedded tissue. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2011;29(35):4620–4626. [PubMed: 22067406]
20. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*. 2003;4(2):249–264. [PubMed: 12925520]
21. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res*. 2003;31(4):e15. [PubMed: 12582260]
22. Chasan-Taber S, Rimm EB, Stampfer MJ, Spiegelman D, Colditz GA, Giovannucci E, et al. Reproducibility and validity of a self-administered physical activity questionnaire for male health professionals. *Epidemiology*. 1996;7(1):81–86. [PubMed: 8664406]
23. Ainsworth BE, Haskell WL, Herrmann SD, Meckes N, Bassett DR Jr., Tudor-Locke C, et al. 2011 Compendium of Physical Activities: a second update of codes and MET values. *Medicine and science in sports and exercise*. 2011;43(8):1575–1581. [PubMed: 21681120]
24. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545–15550. [PubMed: 16199517]
25. Mukherjee R, McGuinness DH, McCall P, Underwood MA, Seywright M, Orange C, et al. Upregulation of MAPK pathway is associated with survival in castrate-resistant prostate cancer. *Br J Cancer*. 2011;104(12):1920–1928. [PubMed: 21559022]
26. Bitting RL, Armstrong AJ. Targeting the PI3K/Akt/mTOR pathway in castration-resistant prostate cancer. *Endocr Relat Cancer*. 2013;20(3):R83–99. [PubMed: 23456430]
27. Leung PS, Aronson WJ, Ngo TH, Golding LA, Barnard RJ. Exercise alters the IGF axis in vivo and increases p53 protein in prostate tumor cells in vitro. *J Appl Physiol (1985)*. 2004;96(2):450–454. [PubMed: 14715676]

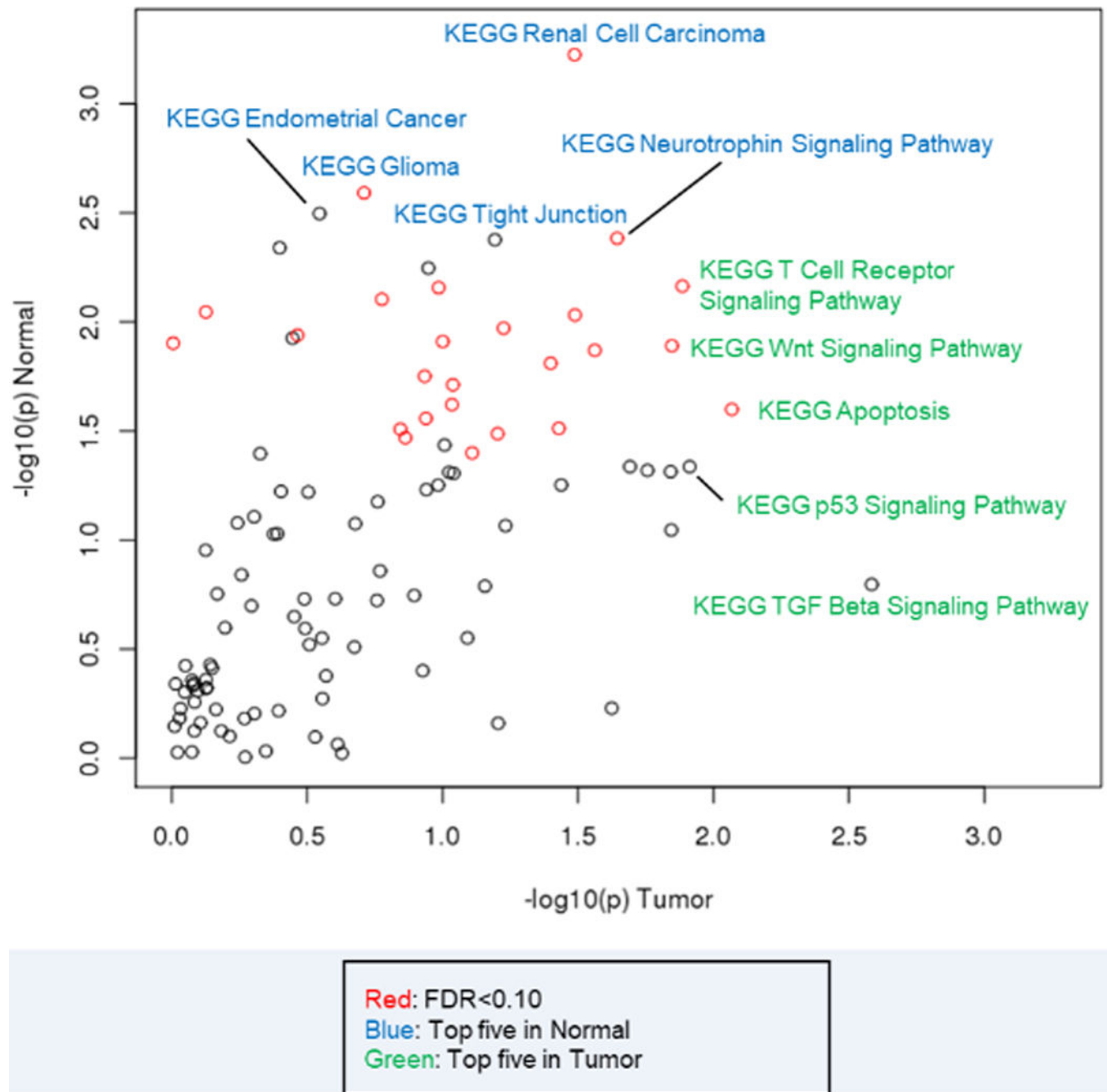


Figure 1.

Significance of enriched KEGG gene sets (n=107) in tumor tissue versus adjacent normal tissue comparing the highest to lowest quintile of vigorous activity in the Health Professionals Follow-up Study (n=48). Red indicates gene sets with FDR < 0.10. The top five gene sets according to nominal p-value in tumor (green) and adjacent normal tissue (blue) are labeled with gene set name. FDR: False discovery rate.

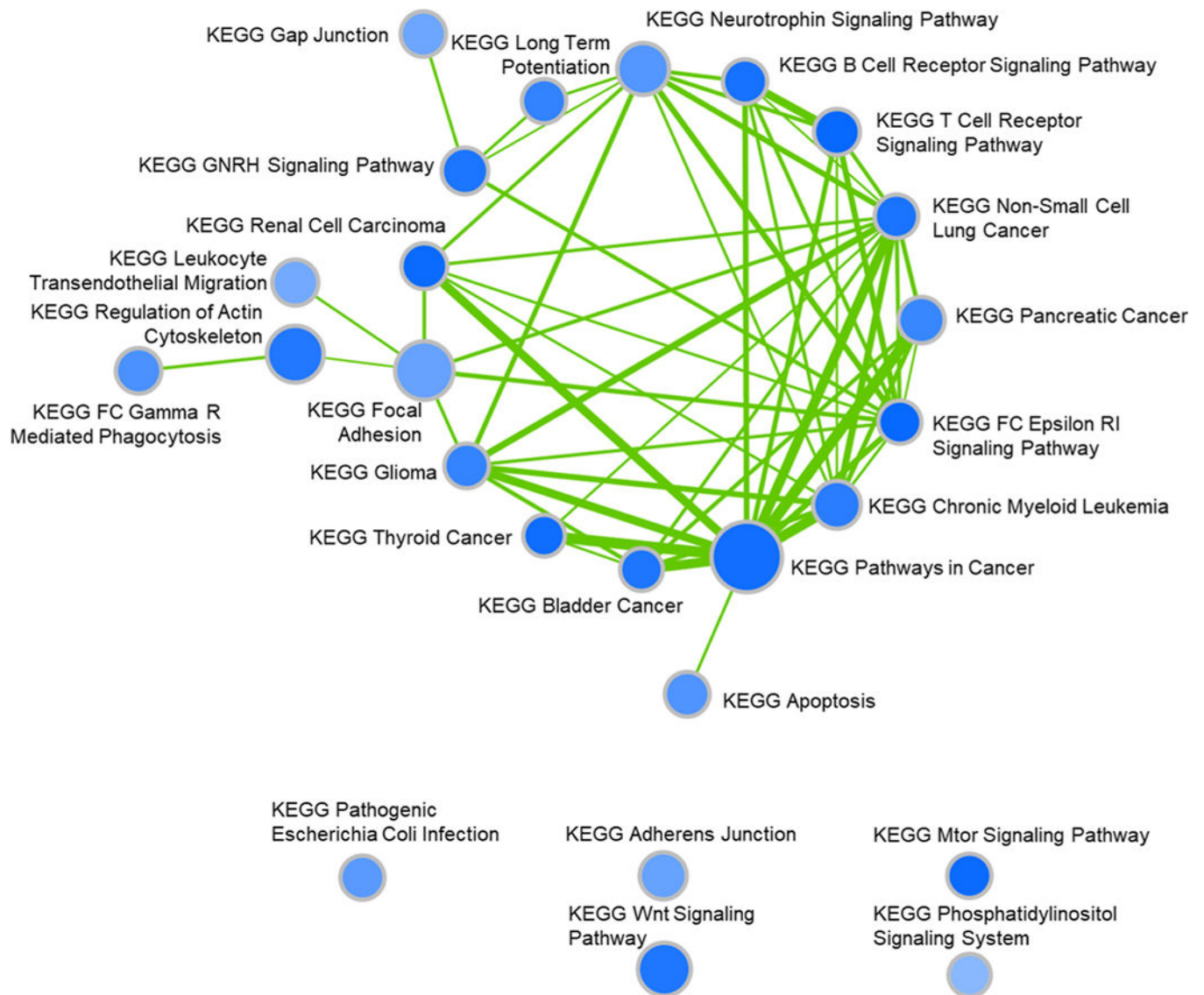


Figure 2.

Enrichment plot including the 25 KEGG gene sets significantly enriched in adjacent normal tissue at FDR <0.10 using a similarity cutoff of 0.5. Blue node color indicates a gene set is downregulated in high versus low vigorous activity. Darker node color indicates a smaller nominal p-value. Node size corresponds with gene set size. Edge width indicates the degree of similarity between gene sets.

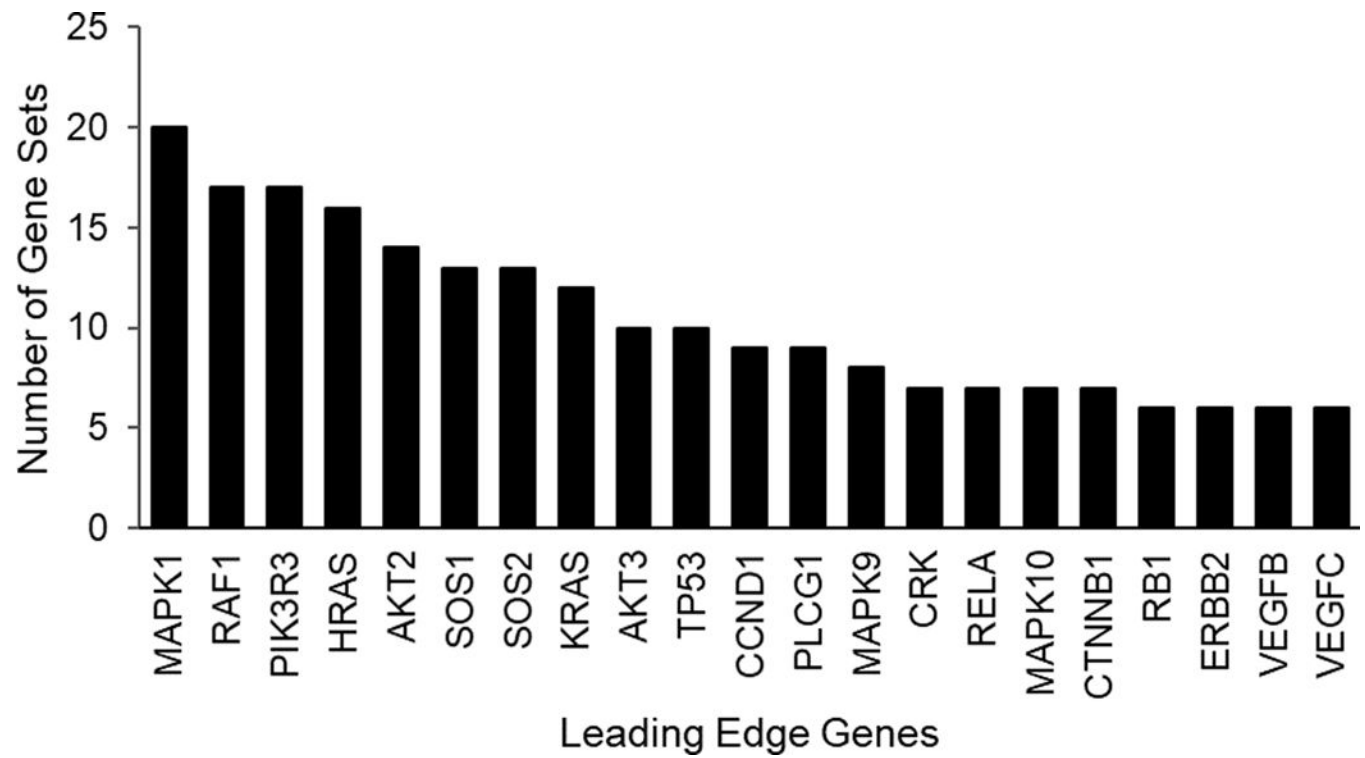


Figure 3.

Number of gene sets containing each leading-edge gene in 25 KEGG gene sets significantly enriched in adjacent normal tissue at FDR <0.10. Genes present in at least six gene sets are shown.

Table 1.

Characteristics of prostate cancer cases (1982–2005) overall and according to quintile of cumulative average vigorous physical activity in the Health Professionals Follow-up Study.

| Characteristics | All cases | Cases According to Vigorous Activity Quintile | | | | |
|---|-------------------|---|------------------|-------------------|-------------------|-------------------|
| | | 1 | 2 | 3 | 4 | 5 |
| Number | 118 | 24 | 23 | 24 | 23 | 24 |
| Age at diagnosis, years (mean, SD) | 65.1 (6.7) | 66.6 (6.1) | 64.7 (6.8) | 65.2 (6.7) | 64.9 (7.0) | 63.9 (7.2) |
| Year of Diagnosis, N (%) | | | | | | |
| Before 1990 | 9 (8) | 4 (17) | 1 (4) | 1 (4) | 1 (4) | 2 (8) |
| 1990–1993 | 20 (17) | 4 (17) | 3 (13) | 5 (21) | 8 (35) | 0 (0) |
| After 1993 | 89 (75) | 16 (67) | 19 (83) | 18 (75) | 14 (61) | 22 (92) |
| PSA Level, ng/mL, median (Q1, Q3) ^a | 7.0 (5.3, 12.7) | 8.8 (5.8, 18.9) | 8.3 (6.3, 10.5) | 7.2 (4.8, 11.1) | 6.0 (5.3, 14.7) | 5.2 (4.0, 7.0) |
| Pathologic Gleason Score, N (%) | | | | | | |
| 6 | 8 (7) | 0 (0) | 0 (0) | 4 (17) | 3 (13) | 1 (4) |
| 7 | 81 (69) | 20 (83) | 15 (65) | 14 (58) | 13 (57) | 19 (79) |
| 8–10 | 29 (25) | 4 (17) | 8 (35) | 6 (25) | 7 (30) | 4 (17) |
| Clinical Stage, N (%) ^b | | | | | | |
| T1 / T2 | 92 (81) | 17 (71) | 19 (83) | 21 (88) | 13 (57) | 22 (92) |
| T3 | 12 (11) | 3 (13) | 2 (9) | 2 (8) | 4 (17) | 1 (4) |
| T4 / N1 / M1 | 9 (8) | 1 (4) | 1 (4) | 1 (4) | 5 (22) | 1 (4) |
| Lethal, N (%) | 44 (37) | 12 (50) | 8 (35) | 7 (29) | 11 (48) | 6 (25) |
| Total activity, MET-h/wk, median (Q1, Q3) ^c | 26.3 (15.2, 41.2) | 14.0 (7.5, 25.2) | 19.4 (7.4, 26.0) | 26.9 (16.4, 36.3) | 28.0 (19.3, 41.2) | 44.6 (35.7, 57.8) |
| Vigorous activity, MET-h/wk, median (Q1, Q3) ^c | 4.4 (0.7, 12.8) | 0.1 (0.0, 0.1) | 1.3 (0.7, 2.4) | 4.4 (4.1, 5.0) | 11.3 (9.0, 12.7) | 28.9 (25.3, 34.5) |

PSA: prostate-specific antigen; Percentages may not sum to 100 due to rounding.

^aPSA at diagnosis was missing for 21 men.

^bClinical TNM stage was missing for 5 men.

^cValues are cumulative averages updated from baseline to prostate cancer diagnosis.