Pan-cancer polygenic risk score associates with cancer susceptibility following kidney transplantation

Jarmo Ritari[1], Kati Hyvärinen[1], Kirsi Jahnukainen[2], FinnGen, Jukka Partanen[1], Ilkka Helanterä[3], Timo Jahnukainen[4]

1) Finnish Red Cross Blood Service, Haartmaninkatu 8, 00290, Helsinki, Finland
2) Division of Hematology-Oncology and Stem Cell Transplantation, New Children's Hospital, Helsinki University Hospital and University of Helsinki, Stenbäckinkatu 9, 00290, Helsinki, Finland
3) Transplantation and Liver Surgery, Helsinki University Hospital and University of Helsinki, Haartmaninkatu 4, 00290, Helsinki, Finland
4) Department of Pediatric Nephrology and Transplantation, New Children's Hospital, Helsinki University Hospital and University of Helsinki, Stenbäckinkatu 9, 00290 Helsinki, Finland

Corresponding author: Timo Jahnukainen, Department of Pediatric Nephrology and Transplantation, New Children's Hospital, Helsinki University Hospital, Stenbäckinkatu 9, 00029 HUS, Finland, email timo.jahnukainen@hus.fi ORCID 0000-0002-1815-7327

Conflict-of-interest statement

The authors have declared that no conflict of interest exists.

Abstract

Background

Cancer accounts for over 20% of late post-transplant mortality, yet the contribution of genetic susceptibility to post-transplant cancer risk remains unclear. This study investigates germline genetic risk factors for post-transplant cancer in the Finnish population using data from the FinnGen cohort.

Methods

A pan-cancer polygenic risk score (PRS) was constructed using genetic variants identified in UK and US populations to assess the influence of common germline variants on time to first cancer diagnosis in 1,802 Finnish kidney transplant recipients (KTRs), of whom 317 developed post-transplant cancer. The PRS was first validated in the FinnGen non-transplantation cohort and subsequently applied to KTRs, with replication in lung and liver transplant recipients (n = 476). Functional relevance was explored by assessing associations between the PRS and expression levels of 2,923 plasma proteins in the UK Biobank (n = 53,013).

Results

Compared to a matched non-transplantation cohort (n = 68,294), KTRs exhibited earlier cancer onset. The PRS was significantly associated with time to first cancer diagnosis in the non-transplantation population (HR 1.04; 95% CI 1.038-1.056; $p = 3.75 \times 10^{-25}$). Among KTRs younger than 40 years, higher PRS was associated with earlier cancer onset (HR, 1.08; 95% CI ,1.01-1.17; p = 0.036), indicating a stronger genetic effect at younger ages. The PRS significantly (Bonferroni < 0.05) altered the regulation of 87 plasma proteins, several of which were known cancer-related markers.

Conclusion

Inherited genetic predisposition, captured by pan-cancer PRS, may contribute to individual susceptibility to cancer after solid organ transplantation, particularly at younger ages.

Kidney transplantation; solid organ transplantation; immunosuppression; cancer risk; polygenic risk score

Introduction

Solid organ transplantation (SOT) is the treatment of choice for severe kidney, liver, and heart failure, with substantially improved short- and long-term outcomes (1–3). Lifelong immunosuppression is essential to prevent allograft rejection. In pediatric and young adult patients, exposure to immunomodulatory drugs may span several decades (4,5). Consequently, SOT recipients face increased risks of post-transplant complications, including diabetes, osteoporosis, cardiovascular disease, fertility problems, and malignancies (6–9).

Cancer is among the most serious long-term complications after SOT, accounting for over 20% of late post-transplant deaths (10–18). Large population-based studies show that transplant recipients without pre-existing cancer have approximately double the risk of malignancy compared to the general population (10–15). Notably, Webster et al. reported (19) that kidney transplant (KT) recipients are diagnosed with cancer earlier than non-transplanted patients. This elevated risk is largely attributed to prolonged immunosuppression (10,20), which impairs immune surveillance and increases susceptibility to infection-associated cancers such as non-Hodgkin lymphoma, liver cancer, and Kaposi's sarcoma. Risks of non–infection-related cancers, including kidney and thyroid cancer, are also elevated. Moreover, immunosuppressive agents like calcineurin inhibitors and azathioprine may promote carcinogenesis via non-immune mechanisms (10).

The contribution of genetic factors to post-transplant cancer risk remains unclear. Some evidence links genetic polymorphisms, especially those affecting skin type, to elevated skin cancer risk in SOT recipients. Specific variants in *interleukin-10* and *transforming growth factor-β* have been associated with post-transplant lymphoproliferative disorder (21,22).

Stapleton et al. demonstrated (23) that a polygenic risk score (PRS) for non-melanoma skin cancer (NMSC) can predict both risk and timing of NMSC after transplant. Another study showed (24) PRSs can stratify SOT recipients into high- and low-risk groups for basal cell carcinoma (BCC) and squamous cell carcinoma (SCC). PRSs have also been applied to post-transplant kidney function (25) and diabetes (26). However, large-scale studies on genetic cancer risk in transplant recipients remain lacking.

In this study, we evaluated genetic risk factors for post-transplant cancer using data from the FinnGen cohort (27), which includes genetic and health registry data from approximately 500,000 Finnish non-transplantation biobank participants and 2,000 solid organ transplantation patients. To assess how common germline variants influence timing of first cancer diagnosis, we developed a pan-cancer PRS encompassing multiple cancer types (Figure 1). We validated the PRS in the non-transplantation FinnGen population and applied it to KT recipients to compare individuals with and without post-transplant malignancies. We hypothesized that KT recipients developing cancer at a young age would exhibit a higher genetic risk, as measured by the PRS, than those who remain cancer-free.

Results

Patient characteristics and data coverage

The FinnGen non-transplantation dataset included 496,641 individuals, while the kidney transplant (KT) dataset comprised 1,546 patients. Liver and lung transplantation (LLT) patients numbered 476. Among the non-transplantation cohort, 115,917 individuals had a cancer diagnosis, and 317 KT patients and 99 LLT patients were diagnosed with cancer post-transplantation. BMI and smoking data were available for 279,283 non-transplantation individuals and 688 KT patients.

Of the 10,626 quality-filtered genetic variants used to construct the pan-cancer polygenic risk score (PRS), 9,818 (92.3%) were present in the FinnGen genotype data. Variant counts per cancer type are listed in Table S1.

Table 1 summarizes key demographic and clinical characteristics of KT recipients with and without post-transplant cancer, including age, follow-up time, and primary renal diagnosis (PRD). KT recipients with post-transplant cancer were significantly older at transplantation than those without cancer. Median follow-up time did not differ significantly. Glomerulonephritis and cystic kidney disease (Table S2) were more common PRDs among patients with cancer, while diabetic nephropathy tended to be more common in those without, however, the difference did not reach statistical significance (Table 1). Diabetes as a comorbidity also differed statistically significantly between groups. Distribution of primary diagnoses over time in the KT cohort is shown by Figure S1. The most common cancer diagnosis class (Table S3) among KT patients was basal cell carcinoma (Figure S2). LTT patients' demographic information is summarized by Table S4.

Polygenic risk score validation and cancer risk in the KT population

Evaluation of cancer type–specific PRSs and four different pan-cancer allele scoring methods in the FinnGen non-transplantation cohort indicated that the sum of allele dosages yielded the best hazard ratio (HR) and p-value (Figure 2A). Based on these findings, the allele dosage sum was selected as the preferred method for constructing the pan-cancer PRS. This PRS followed a normal distribution in the non-transplantation cohort (Figure 2B), allowing for stratification into low- and high-risk groups based on standard deviation (SD) thresholds. HRs for overall cancer risk, calculated using the Cox model, showed that PRS values further from the mean were associated with increasingly extreme HRs (Figure 2C), indicating a stronger association with cancer risk. Cox proportional hazards analysis of full dataset that included both the non-transplantation cohort and the KT cohort did not show evidence of interaction between transplantation status and the PRS (HR 1.47; 95% CI 0.6–3.5; $p = 0.358$) (Table S5).

To assess the PRS's ability to predict cancer severity in the non-transplantation group, we classified ICD-O-3 cancer behavior into benign/semi-malignant versus malignant categories. A Cox model adjusting for sampling year, age at sampling, principal components 1–10 (PC1–10), smoking status, BMI, sex, diabetes, and family history of malignancy showed a significant positive association between the pan-cancer PRS and malignant cancer ($p < 0.001$).

In analyzing the KT and LTT cohorts, we focused on patients younger than 40 years at the time of transplantation. We hypothesized that the higher cancer prevalence in the transplant population may partly reflect prolonged immunosuppression, which could allow underlying genetic predisposition to manifest as cancer at a younger age. This age threshold is consistent with the NIH and American Cancer Society definition of 'young' cancer patients (<40 years; https://www.cancer.gov/types/aya). We divided the young kidney transplant (KT) patients into the high- and low-risk groups, as defined by the pan-cancer PRS ±1 SD, and analyzed their differences in terms of cancer types and number of unique cancer diagnoses. The high-risk group had more diagnoses of lymphoma and basal cell carcinoma compared to the low-risk group (Figure 3A), as well as a higher total number of cancer diagnoses overall (Figure 3B). However, breast cancer diagnoses were less frequent in the high-PRS group than in the low-risk group.

Model covariates

Analysis of the FinnGen non-transplantation cohort with coxph included age at sampling, sampling year, sex, biobank cohort, the first ten genetic principal components, family history of malign neoplasm and diabetes diagnosis as the basic set of analysis covariates. Race and ethnicity were not available in the FinnGen dataset and were therefore not included as variables in the analysis. BMI and smoking information were added for an additional analysis because these data were not available to all individuals (Figure 1). Family history of malignant neoplasm, representing rare high-risk germline variants, was significantly associated with increased cancer risk ($p < 0.001$) (Figure S3), independently of the PRS.

Coxph analysis covariates for KT and LLT cohorts included age at sampling, sampling year, age at transplantation, sex, the first ten genetic principal components, family history of malign neoplasm and diabetes diagnosis. The KT cohort was also analyzed by including BMI and smoking as covariates, although this reduced the number of available samples (Table S6). The most significant predictor was patient age at the time of transplantation ($p < 0.001$) (Table S6).

Polygenic risk score and time between KT and cancer occurrence

We next compared cancer occurrence between the FinnGen non-transplantation cohort and the KT cohort among individuals sampled before age 40 (the age limit of a young cancer patient according to NIH and ACS) and adjusted for age, sex, smoking, BMI and the first ten genetic principal components. The result showed that cancer occurred earlier in KT patients than in non-KT individuals (Figure 4A). Similarly, patient age at the time of transplantation influenced the timing of first cancer occurrence during follow-up (Figure 4B).

When stratifying KT recipients under 40 years of age at KT into high- and low-risk groups based on their pan-cancer PRS values, individuals in the high-risk group (+1 SD) experienced earlier first cancer events compared to those in the low-risk group (−1 SD) (Figure 4C). Further analysis of PRS hazard ratios across cumulative transplantation age showed that HRs were generally higher in younger patients (Figures 4D and 4E), highlighting a stronger genetic contribution to cancer risk in pediatric and young adult KT recipients. Coxph results from different age limits at transplantation for KT patients are shown by Table S7.

PRS in lung and liver transplantation patients

We conducted a replication analysis of the PRS effect observed in young KT patients by applying the same PRS to lung and liver (LLT) SOT recipients in FinnGen. In contrast to the KT cohort, the PRS was not statistically significant among LLT patients younger than 40 years at the time of transplantation ($n = 108$) when using the same covariates and including transplantation type in the Cox proportional hazards model. We additionally tested the interaction between the PRS and diabetes diagnosis within the same model and found that the interaction term was associated with a statistically significant increase in risk (HR, 3.99; 95% CI, 1.38–11.6; $p = 0.0106$) (Figure S4).

Adjustment for multiple testing

After testing multiple PRS candidates in the non-transplantation cohort, assessing PRS and family history of malignancy across age groups in the KT cohort, and evaluating the PRS in the LTT cohort, and other tests, we compiled the resulting p-values for multiple-hypothesis correction, which forms the basis of our conclusions. The Benjamini–Yekutieli–adjusted p-values are presented in Table S8. The corrected p-value for the PRS in KT recipients younger than 40 years at the time of transplantation was 0.23.

Plasma proteome quantitative traits in the UK Biobank

To elucidate the functional basis of the PRS, we examined its association with plasma protein expression levels in the UK Biobank cohort ($n = 53{,}013$). This analysis identified 87 proteins whose levels were significantly regulated by the PRS after Bonferroni correction (adjusted $p < 0.05$; Figure 5, Supplemental Data). Of these, 61 proteins were upregulated and 26 were downregulated. We then performed functional enrichment analyses separately for the upregulated and downregulated sets (Figure S5). Upregulated proteins were predominantly enriched for cell surface components (FDR < 0.05), whereas the downregulated proteins were enriched for lysosomal hydrolases and proteases.

Discussion

Cancer is a major complication after solid organ transplantation (SOT) and contributes considerably to post-transplant mortality (10–15,20). Transplant recipients are known to develop cancer at a younger age than the general population (19,28). To assess the role of common germline variants in cancer risk after kidney transplantation (KT), we developed a pan-cancer polygenic risk score (PRS) using GWAS data from the UK Biobank and US Kaiser Permanente cohorts (29). In the FinnGen non-transplantation population, the PRS showed good transferability and supported earlier cancer onset in KT recipients. The younger age at transplantation was associated with increased cancer risk. The PRS stratified KT recipients by risk when adjusted for clinical and genetic covariates. These results indicate that common genetic polymorphisms influence post-KT cancer risk and may have future clinical utility for risk stratification.

However, interaction between transplant status and PRS was not statistically significant in our analysis, suggesting that the influence of the PRS does not appear to differ meaningfully between transplant and non-transplant individuals in our data set. This finding further highlights the need for a more focused analysis of transplant patients, particularly with respect to age at transplantation because this variable cannot be incorporated into the interaction model as it is absent in the non-transplantation group of the full dataset.

PRSs for specific cancers, such as breast cancer, have been successfully developed in previous studies (30). However, efforts to model general cancer risk independently of cancer type are, to our knowledge, limited. Notably, Zhu and colleagues conducted such an analysis even though their PRS was limited to data from the UK Biobank, and no validation across different populations or cohorts was demonstrated (31). To date post-KT genetic cancer prediction has primarily focused on specific cancer types. Stapleton *et al*. demonstrated that a PRS for NMSC could provide predictive value for post-transplant skin cancer risk (23). In s study by Seviiri *et al*., a PRS generated from the general population was used to stratify the risk of BCC and SCC in SOT recipients under chronic immunosuppression across both low and high ultraviolet exposure environments (24,32). The study showed that transplant recipients in the highest PRS quintile had more than a threefold increased risk of BCC compared to those in the lowest quintile (24). The present results in a more heterogeneous cohort are consistent with these observations. In our cohort, 64% of post-KT cancer patients had at least one NMSC diagnosis, which may have influenced our findings. Therefore, these findings should be validated in a larger KT recipient population with a broader spectrum of post-transplant, non-skin cancers.

Pan-cancer scoring frameworks perform better in lymphomas and basal cell carcinoma (BCC) than in breast cancer because they rely on the assumption of relatively uniform tumor biology. This assumption holds in tumor types where key driver pathways are consistent across patients, such as BCC, which is dominated by activation of the Hedgehog signaling pathway, and many lymphomas, where B-cell receptor and NF-κB–mediated signaling play central roles (33,34). Breast cancer, by contrast, is highly heterogeneous, with intrinsic subtypes that differ markedly at genetic, transcriptomic, and microenvironmental levels. This diversity reduces the performance of tissue-agnostic scoring systems in breast tumors. (35,36)

The risk of cancer following transplantation is likely influenced by multiple factors beyond the transplantation procedure itself. Epidemiological data from the general population indicates that individuals with diabetes mellitus have an elevated risk of developing breast, colorectal, pancreatic, and liver cancers. In contrast, diabetic SOT recipients have been reported to exhibit a lower overall cancer risk (19,37,38). In the current study, the prevalence of diabetic nephropathy as the primary kidney disease did not significantly differ between kidney KT recipients with and without post-transplant malignancies. However, diabetes as a comorbid condition was significantly more frequent among patients without post-transplant cancer. In lung and liver transplantation patients an interaction between diabetes and the PRS showed evidence of increasing post-transplantation cancer risk. This is consistent with a recent observational study in lung transplantation patients which recorded all malignancies post-transplantation and reported diabetes as a risk factor (39).

Polycystic kidney disease (PKD) has been associated with an increased risk of liver, colorectal, and kidney cancers in non-transplantation populations when compared to matched controls from the general population (40). In our cohort, PKD was significantly more common among KT recipients who developed malignancies. The incidence of glomerulonephritis was also significantly higher in the cancer group, possibly due to pre-transplant immunosuppressive treatments. A recent single-center study by Massicotte-Azarniouch *et al*. reported that pre-KT exposure to cyclophosphamide and rituximab, but not calcineurin inhibitors or mycophenolate mofetil, was associated with increased post-transplant cancer risk (41). The relationship between primary kidney disease and post-transplant cancer risk may be confounded by higher mortality rates, particularly among diabetic transplant recipients, which could limit the time available for cancer development. Furthermore, patients with cancers related to their primary disease may have been excluded from transplantation due to pre-existing cancers.

In the FinnGen non-transplantation cohort, a significant association was observed between a family history of malignant neoplasms and a shorter time to cancer diagnosis. However, due to the limited number of KT patients in our study, we lacked the statistical power to properly evaluate the effect of family history on cancer risk within transplant population, although it was included as a model covariate.

The PRS–plasma proteome association analysis highlights potential mechanisms how genetic factors may affect cancer risk. The combined upregulation of cell-surface proteins and downregulation of lysosomal hydrolases/proteases suggests a shift toward sustained receptor-proximal signaling with impaired degradative control, conditions that can lower the threshold for mitogenic and pro-survival pathways, reduce receptor down-regulation and autophagic quality control, and blunt antigen processing. Functionally, this asymmetric increase in signaling input with decreased lysosomal output may favor clonal persistence, accrual of additional alterations, and immune evasion, thereby elevating potential cancer risk without being independently oncogenic (42,43). Notably, several PRS-regulated proteins have established roles in modulating treatment response. For example, GGH (gamma-glutamyl hydrolase), a key regulator of methotrexate polyglutamate levels, has been linked to chemotherapy resistance and prognosis (44). The analysis also points to recently proposed therapeutic targets such as SMPD1 (45) and emerging biomarkers including DPEP1 (46).

Developing a multi-protein score could enhance patient stratification beyond what is achievable with genetics alone. However, causal interpretation requires careful validation and follow-up analyses, including replication, colocalization, Mendelian randomization, and rigorous control for ancestry and other confounders.

This study has several limitations. First, potential sampling bias due to voluntary biobank participation may have led to underrepresentation of certain KT recipients and overrepresentation of cancer patients among non-transplant patients. However, in assessing the validity of the PRS in the non-transplantation group with cox regression, we included detailed cohort information about the biobank source. These adjustments should mitigate most disease risk-derived biases in the PRS validation result. Second, the relatively small number of post-transplant cancer cases within the KT cohort limits statistical power, particularly for subgroup analyses. Third, although our analysis provides evidence for the transferability of polygenic cancer risk across populations, replication in independent cohorts is lacking. Finally, the pan-cancer PRS may not capture all cancer types with equal sensitivity. For instance, post-KT cases of squamous cell carcinoma of the head and neck and breast cancer were less frequent in the high-risk PRS group than in the low-risk group.

Conclusion

Our findings confirm an elevated cancer risk after kidney transplantation and support a role for germline genetic variation in cancer susceptibility. The pan-cancer PRS stratifies young KT recipients by risk, but further validation is required before clinical application

Methods

In the present study, sex was not considered as a biological variable.

Data acquisition and study population

The data for cancer risk association analyses was achieved from the FinnGen data freeze R12, which consisted of genotype data from the samples of 498,187 Finnish biobank participants. The genetic FinnGen data were combined with longitudinal data from Finnish health registries, including the Finnish Registry for Kidney Diseases and the Finnish Cancer Registry. All KT recipients transplanted in Finland were identified from the FinnGen database and the information about possible post-transplant cancer was achieved by linking these individuals to the Finnish Cancer Registry within FinnGen. A total of 1,801 KT recipients were identified, out of which 255 recipients were excluded because of cancer occurrence before KT. Of the remaining 1,546 KT recipients, 317 had been diagnosed with cancer after the KT, and 1,229 recipients had no cancer diagnosis within their follow-up time frame. The primary kidney diagnoses were classified to the main categories shown in Table 1 according to diagnosis categories listed in Table S2. Cancer diagnoses in low- and high-risk groups in KT were classified into categories according to Table S1.

To identify SOTs other than KT in FinnGen, we searched longitudinal data using operation codes JJC00 and JJC01 corresponding with allogenic transplantation of liver and allogenic partial transplantation of liver. Furthermore, we searched endpoint data for terms “lung transplant” and “lung transplantation”. All samples overlapping with KT or showing cancer diagnosis prior to transplantation were excluded.

Genotyping

FinnGen samples were genotyped with Illumina (Illumina Inc., San Diego, USA) and Affymetrix (Thermo Fisher Scientific, Santa Clara, CA, USA) microarrays. Genotype calls were made with GenCall/zCall and AxiomGT1 algorithms for Illumina and Affymetrix data, respectively. Genotype data produced with older array platforms and reference genome builds were lifted over to GRCh38/hg38. Genotype quality control criteria included removing samples with a mismatch between genetically inferred sex and the reported sex in registry data, >5% genotype missingness, and heterozygosity exceeding four standard deviations. Furthermore, variants with >2% missingness, significant deviation from Hardy–Weinberg equilibrium (p-value <1e–6), and minor allele count < 3 were removed. Genotype imputation was performed with the population-specific SISu v.3 imputation panel comprising 3,775 Finns with 25–30X coverage whole-genome sequencing data using Beagle 4.1 (v.08Jun17.d8b, https://faculty.washington.edu/browning/beagle/b4_1.html). Quality control of imputed variants involved a conformity test against the imputation panel and exclusion of variants with imputation INFO scores <0.6 or MAF values <0.0001. Details of the FinnGen project and data analysis pipeline are previously described in the FinnGen flagship paper (27).

Genetic analyses

The pan-cancer PRS was constructed from common polymorphisms identified by a meta-analysis on UK Biobank and the US Kaiser Permanente cohort (https://github.com/Wittelab/pancancer_pleiotropy) (29). The summary statistics result data were filtered for variants that had a fixed effects p-value $<5 \times 10^{-8}$ and low variance between cohorts as measured by the I^2 heterogeneity index value less than five (47). Next, for each cancer type assessed by the meta-analysis (bladder, breast, cervix, colon, gastroesophageal, kidney, leukemia, lung, melanoma, non-Hodgkin's lymphoma, prostate, rectum, thyroid), variants fulfilling the above inclusion criteria were identified and extracted from FinnGen genotype data, and converted to allele dosage format (i.e., 0, 1, or 2 alleles) using plink2 (48) (v2.00a6LM, www.cog-genomics.org/plink/2.0/) ´recode A´ command. The dosages were then re-oriented to risk allele dosages. The risk allele dosages were averaged within each cancer type, scaled to mean zero and unit variance, and combined into a candidate PRS. Candidate PRSs were constructed in four different ways: 1) ´sum´ computes a sum over the different cancer type-specific risk allele mean scores; 2) ´max´ selects the highest from the different cancer type-specific scores; 3) ´min´ selects the lowest from the different cancer type-specific scores; 4) ´pos´ selects the cancer type-specific scores reaching over the population mean and computes a sum of those. The scores were computed for all FinnGen individuals including the KT recipients, and the impact of the scores on time to the first cancer diagnosis were first analyzed in the non-transplantation FinnGen cohort to validate and compare the scores. Finally, the best performing score was evaluated within the FinnGen KT cohort for its ability to predict the time to the first cancer diagnosis after KT.

Statistical analysis

Statistical analyses and data management were performed with R software v4.3.2 (49) (https://www.R-project.org) using RStudio v2023.03.1 (50). Cox proportional hazards (coxph) models for time-to-cancer analyses were performed with the R package survival v3.2-7 (51,52) function coxph, and Kaplan-Maier plots were drawn using function survfit. Plotting and data manipulation was performed with the R package tidyverse v1.3.0 (53). Multiple testing for hypotheses addressing the role the PRS and genetics in cancer risk was performed using the Benjamini-Yekutieli procedure (Table S8). Analysis scripts are available in GitHub (https://github.com/orgs/FRCBS/post-KT_cancer).

Coxph model covariates

Analysis of the FinnGen non-transplantation cohort for PRS validation with coxph included age at sampling, sampling year, sex, biobank cohort, the first ten genetic principal components, family history of malign neoplasm and diabetes diagnosis as analysis covariates. BMI and smoking information were added for an additional analysis because these data were not available to all individuals.

Coxph analyses comparing FinnGen KT and non-TX cohorts included diabetes diagnosis, smoking, BMI, family history of malign neoplasm, sex, age at sampling, sampling year and the first ten genetic PCs as model covariates.

Coxph analyses within the KT and LLT cohorts included age at sampling, sampling year, age at transplantation, sex, the first ten genetic principal components, family history of malign neoplasm and diabetes diagnosis as model covariates.

Analysis of cancer diagnoses

We compared the cumulative sums of unique cancer diagnoses recorded for each patient between the high and low PRS risk groups under 40 years of age at KT. Prior to the analysis we removed broad, non-informative categories present in every patient with any type of cancer diagnosis, such as “cancer” or “neoplasm”, and focused on specific diagnoses only. Since the order of patients for the calculation of the cumulative sum is arbitrary, we randomized the order of patients 100 times and computed an average of the cumulative curves over these randomizations. 95% confidence intervals were calculated based on variation within these randomizations.

Plasma proteomics PRS-pQTL analysis in the UK Biobank

The pan-cancer PRS was computed in the UKB using the same approach applied in FinnGen, as described above. We then merged the PRS values with plasma proteomics measurements and analysis covariates for 53,013 individuals with available proteomics data. Age at recruitment (data-field 21022), sex (data-field 22001), and the first ten genetic principal components (data-field 22009) were included as covariates in regression models assessing the association between PRS and protein expression level. The models were fitted using the bayesglm function from the R package arm (v1.14-4). Data processing and management were performed using dplyr (v1.1.4) and data.table (v1.18.2.1). This research has been conducted using the UK Biobank Resource under Application Number 74245.

We accepted proteins below Bonferroni adjusted p-value of 0.05 as statistically significant and analyzed functional enrichment of significantly upregulated and downregulated proteins separately using the STRING database (https://string-db.org/) (55) with default settings.

Study approval

Study subjects in FinnGen provided informed consent for biobank research, based on the Finnish Biobank Act. Alternatively, separate research cohorts, collected prior the Finnish Biobank Act came into effect (in September 2013) and start of FinnGen (August 2017), were collected based on study-specific consents and later transferred to the Finnish biobanks after approval by Fimea (Finnish Medicines Agency), the National Supervisory Authority for Welfare and Health. Recruitment protocols followed the biobank protocols approved by Fimea. The Coordinating Ethics Committee of the Hospital District of Helsinki and Uusimaa (HUS) statement number for the FinnGen study is Nr HUS/990/2017. The present study was approved by the FinnGen administration team (F_2023_043). Further information on FinnGen ethics approvals is available in Supplemental material.

The UK Biobank (UKB) (54) is a prospective cohort study comprising genotypic and phenotypic data from over 500,000 voluntary participants aged 37–73 years at recruitment. All participants provided informed consent. For the PRS–pQTL analyses, we used plasma proteomics data generated with the Olink Explore 3072 platform as part of the UK Biobank Pharma Proteomics Project Consortium (UKB-PPP).

Data availability statement

FinnGen summary statistics data from >5 individuals from release 12 (R12) are publicly available at https://r12.finngen.fi/. Individual-level data cannot be made public under data protection regulations of GDPR. UK Biobank data are available for use by eligible researchers from academic, charity, government, and commercial organizations from around the world, for health-related research that is in the public interest. Analysis scripts are available at GitHub:

https://github.com/FRCBS/KT_cancer. All data presented in the article, except individual-level data protected by GDPR regulations, are included in the Supporting Data Values file and in Supplementary Data Tables.

Author contributions

J.R., K.H., J.P. and T.J. conceived and designed the study, J.R. did data analysis and statistical analysis, and J.R. and T.J. drafted the manuscript, K.H., K.J., J.P., and I.H. participated in interpretation of the results and critically revised the manuscript, and all authors accepted the final version of the manuscript.

Funding support

This study was supported by the State research funding (Helsinki and Uusimaa Health District), Foundation for Pediatric Research (Timo Jahnukainen), and by Sigrid Jusélius Foundation (Jarmo Ritari and Kati Hyvärinen).

Acknowledgements

We want to acknowledge the participants and investigators of the FinnGen study. The FinnGen project is funded by two grants from Business Finland (HUS 4685/31/2016 and UH 4386/31/2016) and the following industry partners: AbbVie Inc., AstraZeneca UK Ltd, Biogen MA Inc., Bristol Myers Squibb (and Celgene Corporation & Celgene International II Sàrl), Genentech Inc., Merck Sharp & Dohme LCC, Pfizer Inc., GlaxoSmithKline Intellectual Property Development Ltd., Sanofi US Services Inc., Maze Therapeutics Inc., Janssen Biotech

Inc, Novartis AG, and Boehringer Ingelheim International GmbH. Following biobanks are acknowledged for delivering biobank samples to FinnGen: Auria Biobank (www.auria.fi/biopankki), THL Biobank (www.thl.fi/biobank), Helsinki Biobank (www.helsinginbiopankki.fi), Biobank Borealis of Northern Finland (https://www.ppshp.fi/Tutkimus-ja-opetus/Biopankki/Pages/Biobank-Borealis-briefly-in-English.aspx), Finnish Clinical Biobank Tampere (www.tays.fi/en-US/Research_and_development/Finnish_Clinical_Biobank_Tampere), Biobank of Eastern Finland (www.ita-suomenbiopankki.fi/en), Central Finland Biobank (www.ksshp.fi/fi-FI/Potilaalle/Biopankki), Finnish Red Cross Blood Service Biobank (www.veripalvelu.fi/verenluovutus/biopankkitoiminta), Terveystalo Biobank (www.terveystalo.com/fi/Yritystietoa/Terveystalo-Biopankki/Biopankki/) and Arctic Biobank (https://www.oulu.fi/en/university/faculties-and-units/faculty-medicine/northern-finland-birth-cohorts-and-arctic-biobank). All Finnish Biobanks are members of BBMRI.fi infrastructure (https://www.bbmri-eric.eu/national-nodes/finland/). Finnish Biobank Cooperative -FINBB (https://finbb.fi/) is the coordinator of BBMRI-ERIC operations in Finland. The Finnish biobank data can be accessed through the Fingenious services (https://site.fingenious.fi/en/) managed by FINBB.

Figure 1 was partly drawn by adapting images from Servier Medical Art (https://smart.servier.com/), licensed under a Creative Commons Attribution 4.0 Unported License (https://creativecommons.org/licenses/by/4.0/).

References

1. Hariharan S, et al. Long-Term Survival after Kidney Transplantation. *N Engl J Med.* 2021;385:729-743.
2. Kwong AJ, et al. OPTN/SRTR 2022 Annual Data Report: Liver. *Am J Transplant* 2022;24(2S1):S176-S265.
3. Khush KK, et al. The International Thoracic Organ Transplant Registry of the International Society for Heart and Lung Transplantation: Thirty-eighth adult heart transplantation report - 2021; Focus on recipient characteristics. *J Heart Lung Transplant* 2021;40:1035-1049.
4. Chinnakotla S, et al. Outcomes and Risk Factors for Graft Loss: Lessons Learned from 1,056 Pediatric Kidney Transplants at the University of Minnesota. *J Am Coll Surg* 2017;224:473-486.
5. Van Arendonk KJ, et al. National trends over 25 years in pediatric kidney transplant outcomes. *Pediatrics* 2014;133:594-601.
6. Kim JJ, Marks SD. Long-term outcomes of children after solid organ transplantation. *Clinics (Sao Paulo)* 2014;69 (Suppl 1):28-38.
7. Brar S, et al. Incidence of cardiovascular disease and mortality in childhood solid organ transplant recipients: a population-based study. *Pediatr Nephrol* 2023;38:801-810.
8. Sharif A, et al. International consensus on post-transplantation diabetes mellitus. Nephrol Dial Transplant 2024;39:531-549.
9. Tainio J, et al. Testicular function, semen quality, and fertility in young men after renal transplantation during childhood or adolescence. *Transplantation* 2014;98:987-993.

10. Vajdic CM, et al. Cancer incidence before and after kidney transplantation. *JAMA* 2006;296:2823-2831.

11. Vajdic CM, Van Leeuwen MT. Cancer incidence and risk factors after solid organ transplantation. *Int J Cancer* 2009;125:1747-1754.

12. Engels EA, et al. Spectrum of cancer risk among US solid organ transplant recipients. *JAMA*. 2011;306:1891-1901.

13. Grulich AE, et al. Incidence of cancers in people with HIV/AIDS compared with immunosuppressed transplant recipients: a meta-analysis. *Lancet* 2007;370:59-67.

14. Hall EC, et al. Cumulative incidence of cancer after solid organ transplantation. *Cancer* 2013;119:2300-2308.

15. Krynitz B, et al. Risk of skin cancer and other malignancies in kidney, liver, heart and lung transplant recipients 1970 to 2008--a Swedish population-based study. *Int J Cancer* 2013;132:1429-1438.

16. Au E, et al. Cancer in kidney transplant recipients. *Nat Rev Nephrol* 2018;14:508-520.

17. Endén K, et al. Cancer morbidity and mortality after pediatric solid organ transplantation-a nationwide register study. *Pediatr Nephrol* 2020;35:1719-1728.

18. Friman TK, et al. Cancer risk and mortality after solid organ transplantation: A population-based 30-year cohort study in Finland. *Int J Cancer* 2022;150:1779-1791.

19. Webster AC, et al. Identifying high risk groups and quantifying absolute risk of cancer after kidney transplantation: a cohort study of 15,183 recipients. *Am J Transplant* 2007;7:2140-2151.

20. Sapir-Pichhadze R, et al. Immunosuppression and cancer risk in kidney transplant recipients: A retrospective cohort study. Int J Cancer 2024;154:2043-2053.

21. Babel N, et al. Evidence for genetic susceptibility towards development of posttransplant lymphoproliferative disorder in solid organ recipients. *Transplantation* 2007;84:387-391.

22. Laing ME, et al. Genetic factors associated with skin cancer in renal transplant patients. *Photodermatol Photoimmunol Photomed*.2007;23:62-67.

23. Stapleton CP, et al. Polygenic risk score as a determinant of risk of non-melanoma skin cancer in a European-descent renal transplant cohort. *Am J Transplant 2019;*19:801-810.

24. Seviiri M, et al. Polygenic Risk Scores Stratify Keratinocyte Cancer Risk among Solid Organ Transplant Recipients with Chronic Immunosuppression in a High Ultraviolet Radiation Environment. *J Invest Dermatol* 2021;141:2866-2875.e2.

25. Stapleton CP, et al. The relationship between donor-recipient genetic distance and long-term kidney transplant outcome. *HRB Open Res* 2020;3:47.

26. Shaked A, et al. Donor and recipient polygenic risk scores influence the risk of post-transplant diabetes. *Nat Med* 2022;28:999-1005.

27. Kurki MI, et al. FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* 2023;613:508-518.

28. Nimmo A, et al. Increased cancer risk in kidney transplant patients in Scotland: a national registry linkage study. *Br J Cancer* 2025 doi: 10.1038/s41416-025-03086-2.

29. Rashkin SR, et al. Pan-cancer study detects genetic risk variants and shared genetic basis in two large cohorts. *Nat Commun* 2020;11:4423.

30. Mars N, et al. The role of polygenic risk and susceptibility genes in breast cancer over the course of life. *Nat Commun* 2020;11:6383.

31. Zhu M, et al. Genetic Risk for Overall Cancer and the Benefit of Adherence to a Healthy Lifestyle. *Cancer Res* 2021;81:4618-4627. 40.

32. Seviiri M, et al. Polygenic Risk Scores Allow Risk Stratification for Keratinocyte Cancer in Organ-Transplant Recipients. *J Invest Dermatol* 2021;141:325-333.

33. Liu C, Zeinomar N, Chung WK, et al. Generalizability of Polygenic Risk Scores for Breast Cancer Among Women With European, African, and Latinx ancestry. *JAMA Netw Open.*2021;4:e2119084.

34. Almasmoum HA. Molecular complexity of diffuse large B-cell lymphoma: a molecular perspective and therapeutic implications. *J Appl Genet*. 2024;65:57-72.

35. Jia G, Lu Y, Wen W, Long J, Liu Y, Tao R, Li B, Denny JC, Shu XO, Zheng W. Evaluating the Utility of Polygenic Risk Scores in Identifying High-Risk Individuals for Eight Common Cancers. *JNCI Cancer Spectr*. 2020; doi:10.1093/jncics/pkaa021.

36. Hoadley KA, Yau C, Hinoue T, Wolf DM, Lazar AJ, Drill E, Shen R, Taylor AM, Cherniack AD, Thorsson V, Akbani R, Bowlby R, Wong CK, Wiznerowicz M, Sanchez-Vega F, Robertson AG, Schneider BG, Lawrence MS, Noushmehr H, Malta TM; Cancer Genome Atlas Network; Stuart JM, Benz CC, Laird PW. Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell.* 2018;173:291-304.e6.

37. Vigneri P, et al. Diabetes and cancer. *Endocr Relat Cancer*. 2009;16:1103-1123.

38. Kasiske BL, et al. Cancer after kidney transplantation in the United States. *Am J Transplant* 2004;4:905-913.

39. Gitto S, Magistri P, Marzi L, Mannelli N, De Maria N, Mega A, Vitale G, Valente G, Vizzutti F, Villa E, Marra F, Andreone P, Falcini M, Catellani B, Guerrini GP, Serra V, Di Sandro S, Ballarin R, Piai G, Schepis F, Margotti M, Cursaro C, De Simone P, Petruccelli S, Carrai P, Forte P, Campani C, Zoller H, Di Benedetto, F, MEDITRA© group. Predictors of solid extra-hepatic non-skin cancer in liver transplant recipients and analysis of survival: A long-term follow-up study. *Ann Hepatol*. 2022;27(3):100683.

40. Yu TM, et al. Risk of cancer in patients with polycystic kidney disease: a propensity-score matched analysis of a nationwide, population-based cohort study. *Lancet Oncol.* 2016;17:1419-1425.

41. Massicotte-Azarniouch D, et al. Malignancy risk in kidney transplant recipients exposed to immunosuppression pre-transplant for the treatment of glomerulonephritis. *Nephrol Dial Transplant.* 2023;38:2009-2018.

42. Li S, Sampson C, Liu C, Piao HL, Liu HX. Integrin signaling in cancer: bidirectional mechanisms and therapeutic opportunities. *Cell Commun Signal*. 2023;21:266.

43. Xu Y, Shao B and Zhang Y. The significance of targeting lysosomes in cancer immunotherapy. *Front. Immunol*. 2024;15:1308070.

44. Cole PD, Kamen BA, Gorlick R, Banerjee D, Smith AK, Magill E, Bertino JR. Effects of overexpression of gamma-Glutamyl hydrolase on methotrexate metabolism and resistance. Cancer Res. 2001;61(11):4599–4604.
45. Shi L, Mao N, Zheng Z, Liu J, Zhou H, Hou J, Su Y. Glycosphingolipids-Dependent Phospholipid Metabolism Enhances Cancer Initiation and Progression through SMPD1/GLTP/B3GALT4/ST8SIA6 Signaling Axis: A Novel Therapeutic Target. Int J Med Sci. 2025;22(3):604-615.
46. Glass SE, Bechard ME, Cao Z, Aramandla R, Zhao P, Ellis ST, Green EH, Fisher EG, Smith RT, Sievers CK, Irudayam MJ, Revetta F, Washington MK, Ayers GD, Heiser CN, Simmons AJ, Xu Y, Wang Y, Windon A, Shrubsole MJ, Markham NO, Liu Q, Lau KS, Coffey RJ. Dipeptidase-1-knockout mice develop invasive tumors with features of microsatellite-unstable colorectal cancer. JCI Insight. 2025;10(9):e186938.
47. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002;21:1539-1558.
48. Chang CC, et al. Second-generation PLINK: rising to the challenge of larger and richer datasets . *Gigascience* 2015;4:7.
49. R Core Team, Team RDC. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; 2023. http://www.r-project.org
50. Posit team. RStudio: Integrated Development Environment for R. Posit Software, PBC, Boston, MA.; 2023. http://www.posit.co/
51. Therneau, T.M. A Package for Survival Analysis in R. R package version 3.7-0; 2024. https://CRAN.R-project.org/package=survival

52. Therneau, T.M., Grambsch, P.M. Modeling Survival Data: Extending the Cox Model. Springer; 2000.

53. Wickham, H., Averick, M., Bryan, J., Chang, W., D'Agostino McGowan, L., Francois, R. et al. Welcome to the Tidyverse. *J. Open Source Softw.* 4, 1686 (2019).

54. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J, Landray M, Liu B, Matthews P, Ong G, Pell J, Silman A, Young A, Sprosen T, Peakman T, Collins R. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 2015;12(3):e1001779.

55. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, Gable AL, Fang T, Doncheva NT, Pyysalo S, Bork P, Jensen LJ, von Mering C. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res. 2023;51(D1):D638-D646.

Table 1. Patient demographics including all KT recipients, patients with diagnosed posttransplant cancer, and patients without cancer. Some patients may have more than one diagnosis. Primary diagnoses are obtained from kidney disease register.

	All patients	Patients with cancer	Patients without cancer	
	n = 1546	n = 317	n = 1229	*p*-value [a]
Male sex, n (%)	933 (60.4)	187 (59)	746 (60.7)	0.62
Median age at KT, years (IQR)	49.54 (22.9)	54.19 (19.1)	48.20 (23.9)	<0.001[b]
Median time from KT, years (IQR)	7.30 (10.4)	7.57 (7.8)	7.24 (11)	0.025 [b]
Primary diagnosis, n (%),				
Chronic tubulo-interstitial nephritis	65 (4.2)	22 (6.9)	43 (3.5)	0.01
Congenital malformations	56 (3.6)	11 (3.5)	45 (3.7)	1.00
Cystic kidney disease	298 (19)	86 (27)	212 (17)	<0.001
Glomerular disorders in diabetes mellitus	553 (36)	100 (32)	453 (37)	0.09
Glomerulonephritis	175 (11)	53 (17)	122 (9.9)	<0.001
Other	80 (5.2)	15 (4.7)	65 (5.3)	0.80
Unknown	319 (21)	30 (9.5)	289 (23)	<0.001
Reported co-morbidities before KT, n (%)				
Diabetes	545 (35.3)	84 (26.5)	461 (37.5)	< 0.001
Hypertension	1353 (87.5)	298 (94.0)	1055 (85.8)	< 0.001

IQR, interquartile range; KT; Kidney transplantation
[a] Patients with vs. without cancer
[b] Two-way t-test; otherwise, two-way Z-test of proportions

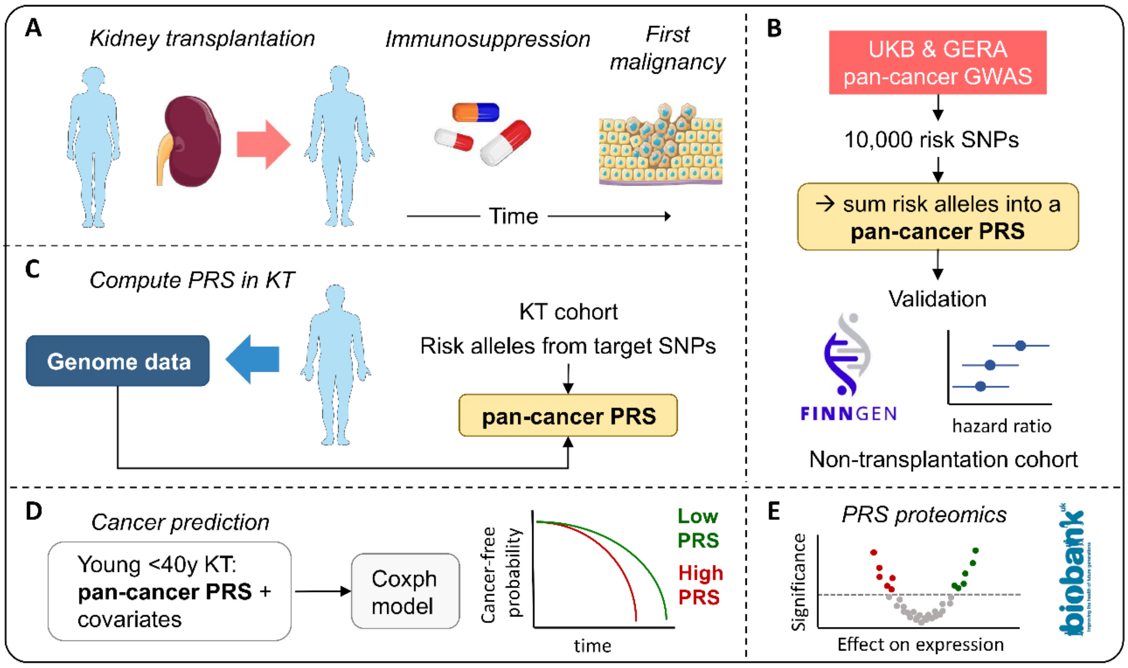

Figure 1. Schematic overview of the study. (**A**) Kidney transplantation (KT) patients receive long-term immunosuppression which elevates post-KT cancer risk especially in young patients. (**B**) A pan-cancer polygenic risk score (PRS) is constructed by selecting statistically significant variants from an UK Biobank (UKB) and US Kaiser Permanente (GERA) GWAS meta-analysis[28] and extracting the risk alleles of the selected variants from FinnGen non-transplantation population (n = 496,641). The PRS is a sum of the number of risk variants normalized for cancer type. The pan-cancer PRS is then validated against longitudinal electronic health record data of cancer diagnoses by fitting a multivariate cox proportional hazards (coxph) survival model for time-to-first cancer occurrence in FinnGen non-transplantation cohort. (**C**) The pan-cancer PRS target variants are extracted from the FinnGen KT cohort and the PRS is similarly computed for the KT patients. (**D**) The PRS in the KT cohort is combined with longitudinal cancer diagnosis data to fit coxph models. The ability of the PRS to stratify KT patients into low- and high-risk groups is evaluated. (**E**) The effect of the PRS on plasma protein expression levels is measured in the UK Biobank.

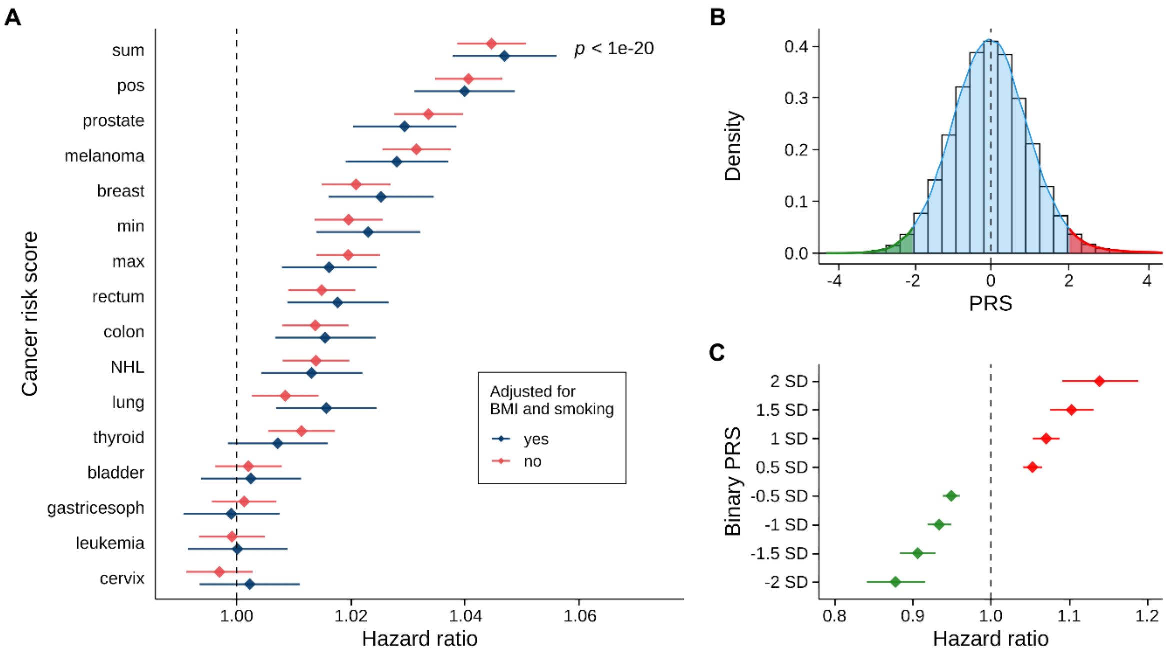

Figure 2. Validation of the pan-cancer polygenic risk score (PRS) in non-transplantation FinnGen cohort. (**A**) Hazard ratios for overall cancer risk of the various cancer type-specific risk scores and the four pan-cancer PRS candidates (sum, max, min, and pos; see Methods for details) obtained from multivariate cox proportional hazards (coxph) survival models. The coxph models analyze the time to the first cancer diagnosis with and without adjusting for body mass index (BMI) and smoking along with other model covariates such as sampling age and population stratification (Methods). (**B**) Continuous population distribution of the best performing score, ´sum´, that is used as the pan-cancer PRS of choice in the follow-up analyses. The x-axis unit is the standard deviation of the mean (SD). The distribution tails, that are colored according to protective (green) and risk (red) PRS directions, are most relevant for cancer risk analysis and stratification. (**C**) PRS distribution in the FinnGen non-transplantation cohort categorized into two classes according to SD (binary PRS) and analyzed with coxph survival models. The distribution tails behave as expected by increasing or lowering the risk according to the distanced from the PRS mean. In (**A**) and (**C**), the error bars represent 95% confidence intervals. Survival analyses in (**A**) were performed with multivariable Coxph regression.

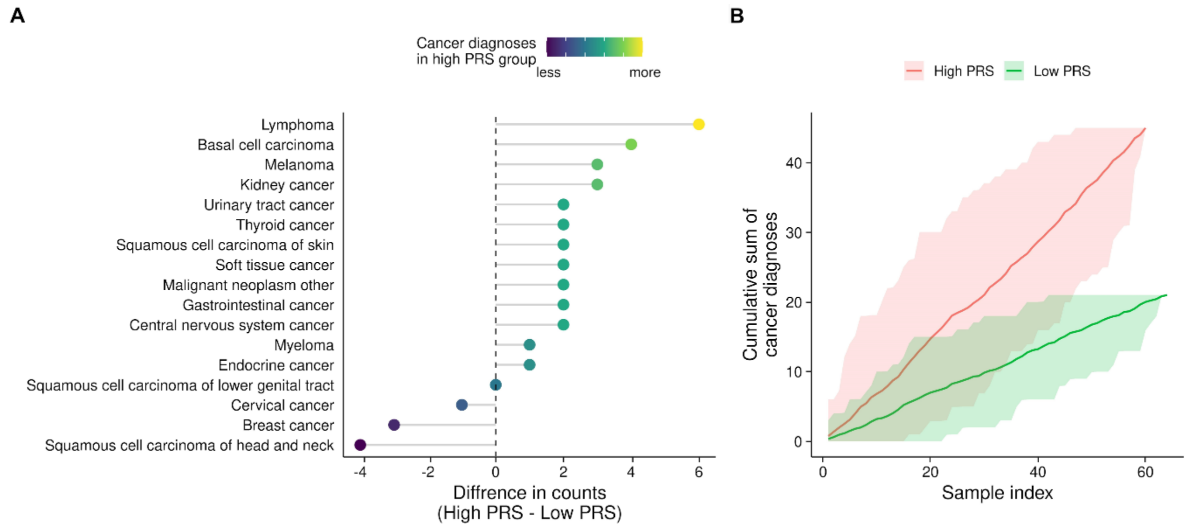

Figure 3. Cancer diagnoses in the high and low pan-cancer PRS groups under 40 years of age at KT. (**A**) Difference in numbers of cancer diagnoses between the high (+1 SD) and low (-1 SD) PRS groups. (**B**) Cumulative sum of unique cancer diagnoses over the follow-up period after KT. A given cancer type can be represented by multiple diagnosis entries. The lines represent averages over 100 sample order randomizations (see Methods).

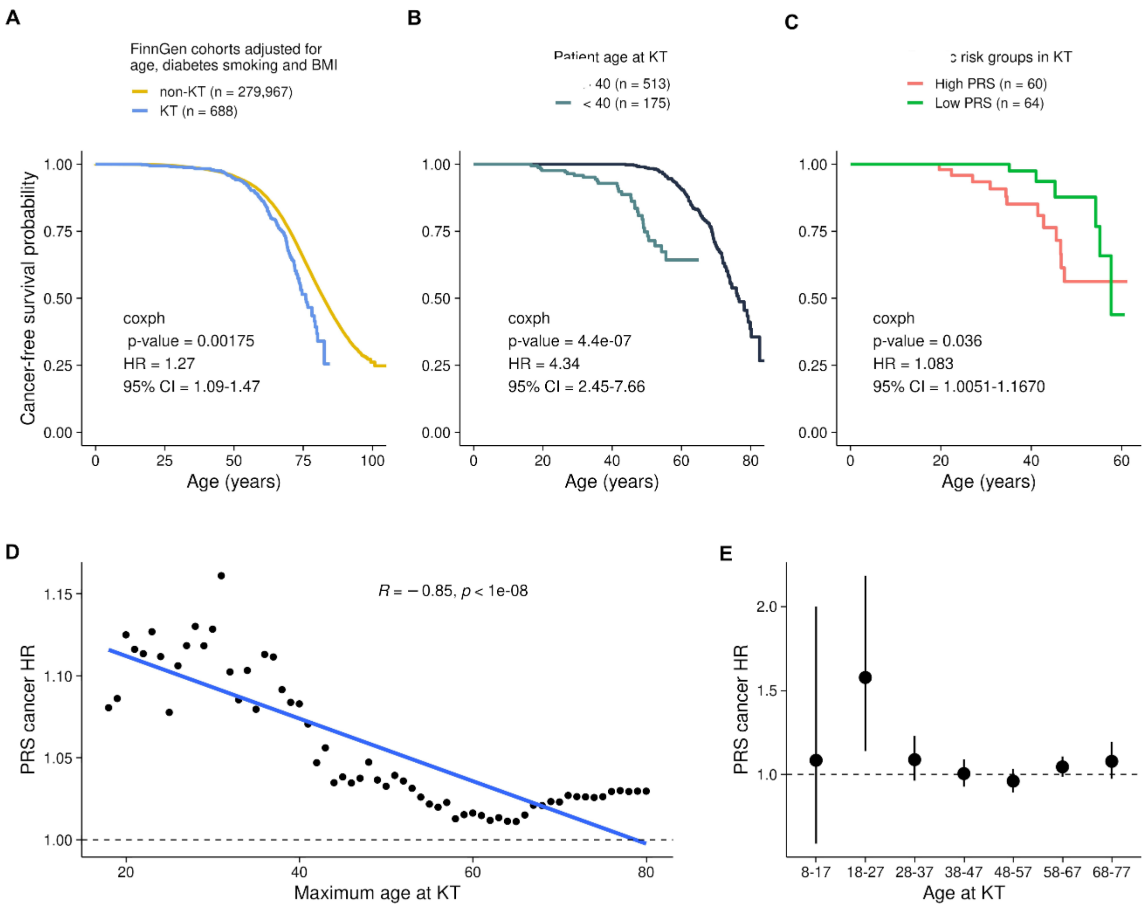

Figure 4. Cancer risk in KT recipients. (**A**) Kaplan-Meier survival curves for the time to the first cancer diagnosis in KT cohort vs. non-transplantation cohort in FinnGen, limited to individuals sampled below 40 years of age. The p-value is obtained from coxph model (Methods) adjusted for age, sex, population stratification, genetic cancer risk, BMI, and smoking. (**B**) Kaplan-Meier curves for the time to the first cancer diagnosis by binned patient age at KT. The p-value is obtained from a coxph model as described above. (**C**) Kaplan-Meier curves for the time to the first cancer diagnosis in KT recipients under 40 years of age at the time of KT, stratified for high (+1 SD) and low (-1 SD) pan-cancer PRS. The p-value is obtained from a coxph model for continuous PRS as described above, including transplantation age as a covariate, but excluding BMI and smoking. (**D**) Pan-cancer PRS coxph model hazard ratio for the first cancer diagnosis as a function of cumulative maximum age at KT, ranging from 18 to 80 years of age at KT. The models are adjusted as described above, including age at KT, but excluding BMI and smoking. (**E**) Pan-cancer PRS hazards ratios by KT age bins of ten years, showing higher risk in young patients. The first bin was defined by the youngest age at KT for which the cox model could be fitted

without errors. Survival analyses in (**A-C**) were performed with multivariable Coxph regression. Analysis of trend in (**D**) was performed with univariate linear regression.

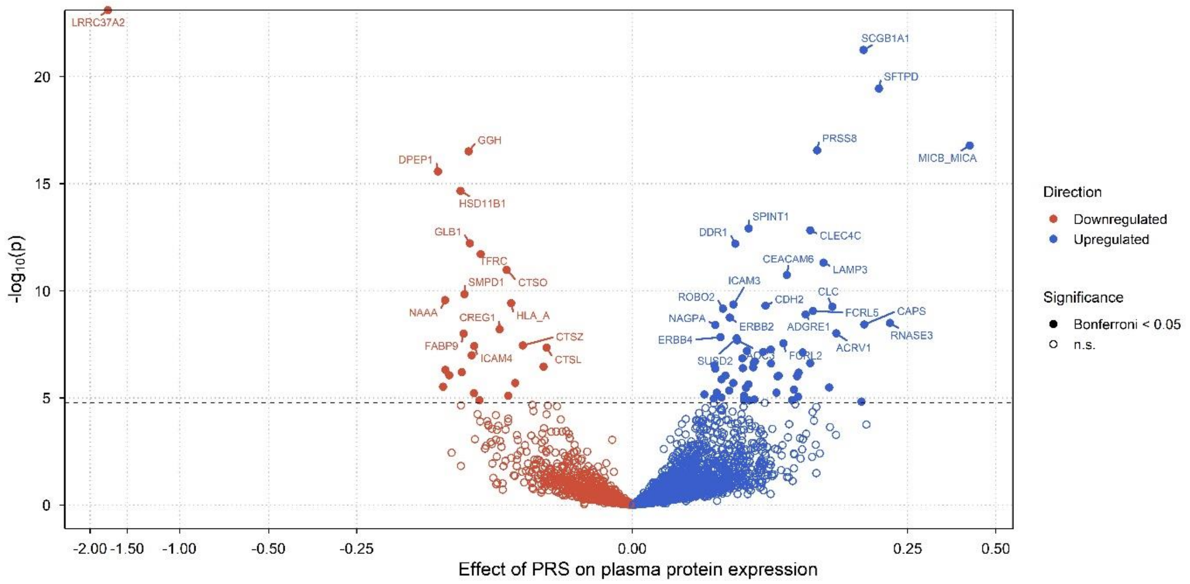

Figure 5. Regulation of plasma protein levels by the pan-cancer PRS in the UK Biobank. The X-axis displays the effect size of the PRS on plasma protein levels, and the Y-axis shows the corresponding association p-values. The horizontal dashed line marks the Bonferroni-corrected significance threshold ($p < 1.71 \times 10^{-5}$). Analysis of expression levels was performed with multivariable logistic regression.