# Transcriptional landscape of PTEN loss in primary prostate cancer

**Eddie Luidy Imada[1,2,3], Diego Fernando Sanchez[2], Wikum Dinalankara[1,2], Thiago Vidotto[4,] Ericka M Ebot[5], Svitlana Tyekucheva[5], Gloria Regina Franco[3], Lorelei Mucci[1], Massimo Loda[1], Edward M Schaeffer[6], Tamara Lotan[2,3], and Luigi Marchionni[1,5,*]**

[1] Department of Pathology and Laboratory Medicine, Weill Cornell Medicine, New York, NY, USA

[2] Department of Oncology, Johns Hopkins University School of Medicine, Baltimore, MD, USA

[3] Departamento de Bioquímica e Imunologia, ICB, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil

[4] Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, MD, USA

[5] Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA

[6] Department of Urology, Northwestern University, Evanston, IL, US

[*] Correspondence to: lum4003@med.cornell.edu

## ABSTRACT

PTEN is the most frequently lost tumor suppressor in primary prostate cancer (PCa) and its loss is associated with aggressive disease. However, the transcriptional changes associated with PTEN loss in PCa have not been described in detail. Here, we applied a meta-analysis approach, leveraging two large PCa cohorts with experimentally validated PTEN and ERG status, to derive a transcriptomic signature of *PTEN* loss, while also accounting for potential confounders due to *ERG* rearrangements. Strikingly, the signature indicates a strong activation of both innate and adaptive immune systems upon *PTEN* loss, as well as an expected activation of cell-cycle genes. Moreover, we made use of our recently developed FC-R2 expression atlas to expand this signature to include many non-coding RNAs recently annotated by the FANTOM consortium. With this resource, we analyzed the TCGA-PRAD cohort, creating a comprehensive transcriptomic landscape of *PTEN* loss in PCa that comprises both the coding and an extensive non-coding counterpart.

## Introduction

Previous molecular studies have explored the genomic heterogeneity of prostate adenocarcinomas (PCa) revealing distinct molecular subsets characterized by common genome alterations (1–3). Among these molecular alterations, loss of the tumor suppressor gene phosphatase and tensin homolog (*PTEN*) – which is implicated in the negative-regulation of the PI3K-AKT-mTOR pathway – has been identified as one of the most common genomic drivers of primary PCa (4,5). Since alterations in the PI3K pathway are present in more than 30% of human cancers, the identification of an expression signature associated with *PTEN* loss has been investigated in different tumor contexts, including breast, bladder, lung, and PCa (6,7).

Assessment of *PTEN* status by fluorescence in situ hybridization (FISH) and immunohistochemistry (IHC) in large clinical PCa cohorts has shown a consistent association with adverse pathological features such as high Gleason score, extra-prostatic extension, as well as prognostic value for biochemical recurrence and cancer-related death (4,8). IHC-based assessment of *PTEN* status has been shown to correlate tightly with genomic alterations of the *PTEN* locus and captures not only loss of the gene, but also mutation and epigenetic changes that lead to *PTEN* functional inactivation(4,9,10) and the potential clinical utility of PTEN IHC as a valuable prognostic marker has been demonstrated previously (11–14).

Though PTEN is involved in a myriad of cellular processes spanning cellular proliferation to tumor microenvironment interactions (5), the transcriptional landscape related to *PTEN* expression has not yet been explored in depth, and the role of long non-coding RNAs (lncRNAs) remains elusive (15). These observations, added to the evidence that subtle *PTEN* downregulation can lead to cancer susceptibility (16), demonstrate the important role of *PTEN* in cancer biology but also highlight the need for additional studies.

Similarly, gene rearrangements of the ETS transcription factor, *ERG*, with the androgen-regulated gene Transmembrane Serine Protease 2 (*TMPRSS2*) are present in ~50% of PCa from patients of European descent. *TMPRSS2-ERG* fusion (herein denoted as $ERG^{+}$ for fusion present and $ERG^{-}$ for absence of fusion) has been shown to activate the PI3K-kinase pathway similarly to *PTEN* loss (17), leading to increased proliferation and invasion. Importantly, tumors harboring *TMPRSS2-ERG* rearrangements show an enrichment for *PTEN* loss (17,18). The co-

53 occurrence of these two genomic alterations makes it challenging to dissect the contributions of each to the
54 transcriptomic landscape.

55 The goal of this study was to elucidate the transcriptional landscape of *PTEN* loss in PCa through the analysis
56 of two large and very well clinically-curated cohorts, for which *PTEN* and *ERG* status was assessed by clinical-grade IHC:
57 The Natural History (NH) cohort, in which patients that underwent radical prostatectomy for clinically localized PCa did
58 not receive neoadjuvant therapy or adjuvant hormonal therapy prior to documented distant metastases (19); and the
59 Health Professionals Follow-up Study (HPFS) cohort in which the patients were followed for over 25 years (20). Based
60 on IHC-assessed *PTEN* status for these cohorts, we built a *PTEN*-loss signature highly concordant across the
61 independent datasets, in both presence and absence of *TMPRSS2-ERG* fusion. Overall, this *PTEN*-loss signature was
62 associated with cellular processes associated with aggressive tumor behavior (*e.g.*, increased motility and proliferation)
63 and, surprisingly, with increases in gene sets related to the immune response. In addition, through our recently
64 developed FANTOM-CAT/recount2 (FC-R2) resource (21) and copy-number-variation data, we expanded this signature
65 beyond coding genes and report the non-coding RNA repertory resulting from *PTEN* loss.
66

## Methods

### Data collection and Immunostaining

All expression data used in this work were gathered from public domain databases. In this work, we made use of three cohorts: FC-R2 TCGA, Natural History (NH), and Health Professionals Follow-up Study (HPFS). Information about each cohort is summarized in Table 1. Information about *PTEN* status by immunohistochemistry for the HPFS cohort was readily available and therefore obtained from the public domain. For NH cohort samples, IHC staining for *PTEN* and *ERG* were performed using a previously validated protocol (22). Last, for TCGA we used the Copy Number Variation (CNV) called by the GISTIC algorithm to define *PTEN* status and expectation-maximization algorithm to define ERG status.

### Meta-analysis of NH and HPFS cohorts

We performed a meta-analysis approach using a Bayesian hierarchical multi-level model (BHM) for cross-study detection of differential gene expression implemented in the Bioconductor package XDE (23) on microarray-based cohorts to obtain a *PTEN*-null signature from *PTEN* IHC validated samples. The model was fitted using the delta gp model with empirical starting values and 1000 bootstraps were performed. All remaining parameters were set to default values. This analysis was also performed stratifying the samples by *ERG* status to evaluate the impact of the *ERG* rearrangement in the signature.

### Differential expression analysis in the TCGA cohort

A generalized linear model (GLM) approach coupled with empirical Bayes moderation of standard errors and voom precision weigths (24,25) was used to detect differentially expressed genes in the TCGA cohort. The models were adjusted for surrogate variables with the SVA package (26). Adjusted p-values controlling for multiple hypothesis testing were performed using the Benjamini-Hochberg method and genes with false discovery rate (FDR) equal or less than 0.1 were reported (27).

### Gene set enrichment analysis (GSEA)

The results from the meta-analysis performed in the NH and HPFS cohort were ranked by the weighted size effect (average of the posterior probability of concordant differential expression multiplied by the Bayesian effect size of

each cohort). The results from the TCGA cohort were ranked by t-statistics. Ranked lists were tested for gene set enrichment. Gene set enrichment analysis (GSEA) was performed using a Monte Carlo adaptive multilevel splitting approach, implemented in the fgsea (28) package. A collection of gene sets (Hallmarks, REACTOME, and GO Biological Processes) were obtained from the Broad Institute MSigDB database. The androgen response gene set was obtained from Scheaffer et al (29). Gene sets with less than 15 and more than 1500 genes were removed from the analysis, except for the GO biological processes whose max size was set to 300 to avoid overly generic gene sets. The enriched pathways were collapsed to maintain only independent ones using the function collapsePathways from fgsea.

## Results

### Meta-analysis of Natural History and Health Professionals Follow-Up Study cohorts

We sought to obtain a consensus signature of *PTEN* loss that could be reproduced across independent cohorts. We utilized a meta-analysis approach leveraging a multi-level model for cross-study detection of differential gene expression (DGE). We fitted a Bayesian hierarchical model (BHM) for analysis of differential expression across multiple studies that allowed us to aggregate data from two previously described tissue microarray-based cohorts where *PTEN* and *ERG* status was determined by IHC (Table 1 and Figure 1) and we derived a *PTEN*-loss signature (Figure 2). In this analysis, we observed 813 genes for which the differential expression was highly concordant (Bayesian Effect Size (BES) $\geq 1$, posterior probability of concordant differential expression (PPCDE) $\geq 0.95$) (Table S1).

The consequences of *PTEN* loss on cell cycle regulation and tumor cell invasion has been extensively reported previously (4,30,31). Accordingly, beyond *PTEN* itself, the top DEG genes in our signature reflected this profile (Figure 2 and Table S1). Dermatopontin (*DPT*) (BES = -2.59, PPCDE = 1) and Alanyl membrane aminopeptidase (*ANPEP*) (BES = -2.53, PPCDE = 1) were found down-regulated upon *PTEN* loss. Leucine-Rich Repeat Neuronal 1 (*LRRN1*) was among the genes up-regulated upon *PTEN* loss (BES = 3.36, PPCDE = 1). These and other genes found differentially expressed upon *PTEN* loss have all been shown to be associated with a more aggressive phenotype in several cancer types (5) .

Notably, we found *ERG* among the top upregulated genes in the signature (Figure 2). As expected (18,32,33), *ERG* rearrangement was more common among cases with *PTEN* loss compared to intact *PTEN* in all cohorts (Fisher exact test, $p \leq 0.001$). Given this enrichment, it was not surprising that *ERG* was among the most up-regulated genes in the BHM signature, as well as *PLA2G7*, which has been shown to be among the most highly overexpressed genes in *ERG*-rearranged PCa compared to those lacking *ERG* rearrangements (34). The presence of ERG and ERG-regulated transcripts in the *PTEN*-loss signature suggested that this signature might be confounded by enrichment of *ERG* rearranged tumors among the tumors with PTEN loss.

Since *ERG* rearrangements represent a major driver event in PCa and *PTEN* loss is enriched in *ERG*-rearranged tumors, we next investigated the role of ERG in our PTEN-loss signature. To this end, we repeated the Bayesian hierarchical model for the analysis of differential expression by stratifying the samples by *ERG* status. In the background with *ERG* rearrangement, we observed a similar signature to the previous overall *PTEN*-loss signature, but without the aforementioned *ERG*-associated genes (Supplementary figure S1 and Supplementary table S2). However, in the

absence of *ERG* rearrangement, we could not find any significant differences between samples with or without *PTEN* loss. This was unexpected given that *PTEN* is a powerful tumor suppressor capable of triggering multiple molecular changes.

## Extending the PTEN-loss signature

To validate our PTEN loss signatures in an orthogonal cohort, we next examined the TCGA PRAD cohort (35), where *PTEN* status was estimated by genomic copy number (CN) assessment, which was closely aligned with PTEN gene expression (Figure S3). We recently developed a comprehensive expression atlas based on the FANTOM-CAT annotations. This meta-assembly is currently the broadest collection of the human transcriptome (21,36). These gene models include many novel lncRNA categories such as enhancers and promoters, allowing the signature to be further expanded beyond the coding repertoire. We used TCGA expression data from the FC-R2 expression atlas (21) to perform DGE analysis stratified by the *PTEN* status as derived from CN analysis. We also performed the same analysis in a stratified manner as in the HPFS and NH cohorts, using the ERG expression with expectation maximization (EM) algorithm to define ERG status given the bimodal nature of ERG expression in PCa. Interestingly, we were able to detect differential expression between *PTEN*-null and PTEN-intact samples without *ERG* rearrangement in the TCGA cohort, which used high-throughput sequencing as opposed to gene expression microarrays, suggesting that there the lack of signal in the previous analysis can be a reflection of the potential limitations with the later technology.

We observed 521 differentially expressed genes (DEG) when comparing *PTEN*-null and *PTEN*-wild-type samples (FDR $\leq$ 0.01, LogFC $\geq$ 1), of which 257 were coding genes and 264 were non-coding genes (Supplementary Table S3). When stratifying the samples by ERG status, we obtained 435 and 364 DEG in the background with and without ERG rearrangement (Supplementary Table S4 and S5), respectively, with similar proportions of coding and non-coding genes. Using Correspondence-at-the-top (CAT) analysis of the coding genes, we observed a higher concordance than expected by chance between the TCGA PTEN-loss signature and that from the BHM (Figure S4). This confirmed that CN is a reasonable proxy to IHC-staining in TCGA which allowed us to expand this signature beyond coding RNAs.

In this analysis, we were able to detect a variety of lncRNAs that are already known to be involved in PCa development and progression. Notably, several differentially expressed lncRNAs were already reported to be

associated with PCa (37–46) (e.g. *PCA3*, *PCGEM1*, *SCHLAP1*, *KRTAP5-AS1*, *Mir-596*) (Supplementary Table S3-S5). *PCA3* is a prostate-specific lncRNA overexpressed in PCa tissue. Similarly, lncRNA *PCGEM1* expression is increased and highly specific in PCa where it promotes cell growth and it has been associated with high-risk PCa patients (41,42). On the other hand, *KRTAP5-AS1* expression has not been directly associated with PCa.

Also ranked high among lncRNAs differentially expressed were the lncRNAs *SChLAP1* and its uncharacterized antisense neighbor *AC009478.1*. *SchLAP1* is overexpressed in a subset of PCa where it antagonizes the tumor-suppressive function of the SWI/SNF complex and can independently predict poor outcomes (45,46). On the other hand, the role of *AC009478.1* in PCa development is still unknown. Interestingly, *SchLAP1* and *AC009478.1* expression is strongly correlated in the TCGA datasets only in PCa ($R = 0.94$, $p < 2.2e-26$) and bladder cancer ($R = 0.85$, $p < 2.2e-26$) (Figure S5).

Strikingly, a substantial proportion of lncRNAs associated with *PTEN* loss were not yet associated with PCa. Out of the 264 DE non-coding genes, 134 were novel and annotated only in the FANTOM-CAT meta-assembly annotation (Table 2). Among the FANTOM-CAT exclusive genes, those with the highest fold change in close proximity with coding genes were *CATG00000038715*, *CATG00000079217*, and *CATG00000117664* (Figure S6). These genes were mostly expressed in PCa as opposed to other cancer types in the TCGA dataset (Figure 3).

Among the downregulated genes were *CATG00000038715* and *CATG00000079217*. *CATG00000038715* is in close proximity to *CYP4F2* and *CYP4F11*, encoding members of the cytochrome P450 enzyme superfamily. Expression of *CATG00000038715* and *CYP4F2* are highly correlated ($R=0.91$, $p < 2.2e-16$) in PCa, and expression of the former was highly specific for PCa (Figure S7). *CATG00000079217* is in close proximity to the coding gene *FBXL7*, an F-box gene which is a component of the E3 ubiquitin ligase complex. While expression of *FBXL7* and *CATG00000079217* showed only a weak correlation ($R=0.14$, $p < 7.4e-4$), *CATG00000079217* expression was notably higher in PCa and breast cancer than in other cancers, and it was moderately correlated with several PCa biomarkers (e.g. *KLK2*, *KLK3*, *STEAP2*, *PCGEM1*, *SLC45A3*) (41,42,47–51) ($R=0.37-0.57$, $p < 2.2e-16$) in TCGA.

*CATG00000117664* was among the most upregulated lncRNA and it is located near *GPR158*, a G protein coupled receptor highly expressed in brain. The expression between *GPR158* were correlated ($R=0.54$, $p < 2.2e-16$), and *CATG00000117664* expression was shown to be highly specific to PCa (52) (Figure S7).

182

183 **PTEN loss induces the innate and adaptive immune system**

184 We performed Gene Set Enrichment Analysis (GSEA) using fgsea (28) and tested both the BHM- and TCGA-generated
185 molecular signatures for enrichment in three collections of the Molecular Signature Database (MSigDB) (53,54):
186 HALLMARKS, REACTOME, and GO Biological Processes (BP). Results were similar in both signatures, with positive
187 enrichment of proliferation and cell cycle-related gene sets (e.g. MYC1 targets, MTORC1 signaling, cell cycle
188 checkpoints, and DNA repair) and both innate and adaptive immune system associated gene sets (e.g. Neutrophil
189 degranulation, MHC antigen presentation, interferon-alpha, and gamma) (Figure 4-5 and Supplementary Table S6-
190 S20). The positive enrichment of MHC antigen presentation, interferon-alpha and -gamma in PTEN-null tumors is
191 consistent with our previous study showing that the absolute density of T-cells is increased in PCa with PTEN loss (55).

192 Since *PTEN*-null tumors are known to have decreased androgen output, which is a strong suppressor of
193 inflammatory immune cells (29,56,57), we hypothesized that this decrease in androgen levels could activate an
194 immune response. We, therefore, performed a GSEA analysis using a collection of androgen-regulated genes from
195 Schaeffer et al. (29) to test if the *PTEN*-null signature was enriched in this gene set. Both the TCGA- and BHM-signature
196 were shown to be positively enriched in genes that were shown to be repressed upon dihydrotestosterone treatment
197 (NES =1.39-155, FDR $\leq$ 0.05) (Figure S8).

198

## Discussion

With an estimated prevalence of up to 50%, *PTEN* loss is recognized as one of the major driving events in PCa (58). *PTEN* antagonizes PI3K-AKT/PKB and is a key modulator of the AKT-mTOR signaling pathways which are important in regulating cell growth and proliferation. Accordingly, *PTEN* loss is consistently associated with more aggressive disease features and poor outcomes. Saal and collaborators previously generated a transcriptomic signature of *PTEN* loss in breast cancer (6). While this signature was correlated with worse patient outcomes in breast and other independent cancer datasets, including PCa, the signature unsurprisingly fails to capture key characteristics of PCa such as *ERG*-rearrangement (6,11). Significantly, a transcriptomic signature reflecting the landscape of *PTEN* loss in PCa has not been described to date.

Immunohistochemistry (IHC) assay is a clinically utilized technique to determine the status of the *PTEN* gene, with high sensitivity and specificity for underlying genomic deletions (59) (Figure 1). Therefore, we analyzed transcriptome data from two large PCa cohorts – the Health Professional Follow-up Study (HPFS) and the Natural History (NH) study – for which IHC-based PTEN and ERG status was available (n = 390 and 207, respectively), deriving a *PTEN*-loss gene expression signature specific to PCa (Figure 2 and Supplementary Table S1). Genes that are associated with increased proliferation and invasion in several cancer types, such as *DPT*, *ANPEP* and *LRRN1*, were among the most concordant DEG in this signature. *DPT* has been shown to inhibit cell proliferation through MYC repression and to be down-regulated in both oral and thyroid cancer (60,61). It has also been shown to control cell adhesion and invasiveness, with low expression leading to a worst prognosis (61,62). *ANPEP* is known to play an important role in cell motility, invasion, and metastasis progression (62,63), and lower expression of this gene has been associated with the worst prognosis (64). *LRRN1* is a direct transcriptional target of *MYCN*, and an enhancer of EGFR and IGRF signaling pathway (65). Higher levels of *LRRN1* expression promote tumor cell proliferation, inhibiting cell apoptosis, and play an important role in preserving pluripotency-related proteins through *AKT* phosphorylation (65–67), leading to a poor clinical outcome in gastric and brain cancer.

Notably, *ERG* was shown to be upregulated in our signature, which led us to perform a stratified analysis to avoid capturing signals driven mostly by *ERG* overexpression. Surprisingly, we were not able to detect significant differences by *PTEN* status in the HPFS and NH cohorts, which were quantified by gene expression microarrays, in the $ERG^-$ samples. Conversely, when analyzing the TCGA cohort, we were able to detect significant changes by *PTEN* status

226 in the *ERG*$^{-}$ samples (Supplementary Tables S3-S5). However, given the known limitations of gene expression
227 microarrays performed on formalin fixed material, such as the limited dynamic range of expression values (68), we
228 believe that the HPFS and NH datasets were limited by the technology employed. Nevertheless, concordance between
229 the BHM- and TCGA- cohorts were similar in both the overall and the *ERG*$^{+}$ background comparison (Supplementary
230 Figure S4).

231 We observed in the TCGA cohort several lncRNAs that have already been associated with PCa progression were
232 found in our signature. PCA3 acts by a variety of mechanisms such as down-regulation of the oncogene *PRUNE2* and
233 up-regulation of the *PRKD3* gene by acting as a miRNA sponge for *mir-1261* leading to increase proliferation and
234 migration(37,38). Conversely, knockdown of *PCA3* can lead to partial reversion of epithelial-mesenchymal transition
235 (EMT) (39) which can lead to increased cell invasion, motility, and survival (40). Although *KRTAP5-AS1* has not been
236 associated with PCa, it has recently shown that *KRTAP5-AS1* can act as a miRNA sponge for miRNAs, such as *mir-596*,
237 which targets the oncogene *CLDN4* which enhances the invasion capacity of cancer cells and promote EMT (40,43),
238 thereby overexpression of *KRTAP5-AS1* can lead increased levels of *CLDN4* (44). *Mir-596* has also been shown to be
239 overexpressed in response to androgen signaling and associated with anti-androgen therapy resistance (44).

240 Moreover, many lncRNAs exclusively annotated in the FANTOM-CAT were associated with PTEN-loss and were
241 shown to be expressed mostly in PCa (Figure 3). Since these genes are novel genes without elucidated function, we
242 analyzed potential roles for these genes by looking at coding genes located in the same loci. Among the top DE lncRNAs,
243 genes within proximity to coding genes were *CATG00000038715*, *CATG00000079217*, and *CATG00000117664* (Figure
244 S6) which are positioned in the same loci as *CYP4F2*, *FBXL7*, and *GPR158*, respectively. *CYP4F2* is involved in the process
245 of inactivating and degrading leukotriene B4 (*LTB4*). *LTB4* is a key gene in the inflammatory response that is produced
246 in leukocytes in response to inflammatory mediators and can induce the adhesion and activation of leukocytes on the
247 endothelium.(69). *FBXL7* regulates mitotic arrest by degradation of *AURKA*, which is known to promote inflammatory
248 response and activation of NF-κB (70,71). Likewise, increase expression of *GPR158* is reported to stimulate cell
249 proliferation in PCa cell lines, and it is linked to neuroendocrine differentiation (72).

250 We consistently observed a strong enrichment in immune response genes and gene sets upon *PTEN* loss
251 (Figure 4 and Supplementary Tables S6-S20). Immune-associated genes (i.e. *GP2* and *PLA2G2A*) were found amongst

the top up-regulated genes in our signature (Figure 2). Positive enrichment of Interferon-alpha- and gamma-response genes (FDR ≤ 0.01) further suggests that a strong immuno-responsive environment, with both innate and adaptive systems activated, is developed in *PTEN*-null tumors (Figure 5). The positive enrichment of MHC class II antigen presentation, neutrophil degranulation, vesicle-mediated transport, and FC receptor pathway-related genes suggests that *PTEN*-null tumors may be immunogenic (Figure 4). This finding was particularly surprising given that *PTEN* is itself a key positive regulator of innate immune response, controlling the import of *IRF3*, which is responsible for IFN production. Accordingly, disruption of PTEN expression has previously been reported to lead to decreased innate immune response (73). Conversely, it has also been hypothesized that the increased genomic instability caused by, or associated with, *PTEN* loss can increase immunogenicity in the tumor micro-enviroment (TME) (74). This finding is of particular interest given that immune-responsive tumors can be good candidates for immunotherapy-based approaches.

Remarkably, despite loss of *PTEN* being associated with higher expression of the immune checkpoint gene programmed death ligand-1 (*PD-L1*) in several cancer types (75,76) this is not true in PCa (77). So far, current immunotherapeutic interventions, such as *PD-1* blockade, in PCa have not been successful. One of the possible reasons is the lack of *PD-L1* expression (77). Therefore, alternative targets must be considered for immunotherapy in PCa. One alternative target is the checkpoint molecule *B7-H3* (*CD276*), whose expression has already been associated with PCa progression and worse prognosis (78) and has been suggested as a target for immunotherapy (79,80). *CD276* was one of the most concordant up-regulated genes in our signature (Figure 2) suggesting that its expression is associated with *PTEN* loss. Interestingly, *B7-H3* expression may be down-regulated by androgens (81).

The effects of androgen on the immune system has already been extensively studied and reviewed (56). Androgens are known to suppress inflammatory immune cells and to impair the development and function of B- and T-cells (57). We, therefore, hypothesized that the decreased levels of androgen in *PTEN*-null TME could lead to an unsuppressed immune system. By testing our signature for enrichment in androgen-related genes (AR) derived from Schaeffer et al. (29), we observed that upon *PTEN*-loss, androgen-sensitive genes that are typically suppressed by DHT are positively enriched, indicating that androgen levels or androgen response in *PTEN*-null tumors may be lower than in their *PTEN*-intact counterparts (Figure S8). This decrease in AR-signaling has been described in *PTEN*-null tumors, in

278 which activation of PI3K pathway inhibits AR activity. (82). Furthermore, AR inhibition activates AKT signaling by
279 inhibiting AKT phosphatase levels further boosting cell proliferation (82), which has also been noted in this study
280 (Figure 3). Finally, in the non-coding repertoire, both *PCA3* and *PCGEM1* are modulated by androgen (83,84) and were
281 down-regulated upon *PTEN* loss which tracks with the observed decreased androgen response in *PTEN*-null tumors
282 (Figure S6 and S8).
283

## Conclusion

Altogether, we have generated a highly concordant gene signature for *PTEN* loss in PCa across three independent datasets. We show that this signature was highly enriched in proliferation and cell cycle genes, leading to a more aggressive phenotype upon *PTEN* loss, which is concordant with the literature. Moreover, we have shown that *PTEN* loss is associated with an increase in both innate and adaptive immune response. Although the literature shows that *PTEN* loss usually leads to immuno-suppression, we find evidence that this finding may be reversed in PCa. This observation has potential implications in the context of precision medicine since immune responsive tumors are more likely to respond to immunotherapies. Therefore, *PTEN*-null tumors might benefit more from this approach than *PTEN*-intact tumors. Potentially, *PTEN* status can guide immunotherapy combination with other approaches such as androgen ablation.

Finally, by leveraging the FC-R2 resource, we were able to highlight many lncRNAs that may be associated with PCa progression. Although functional characterization these lncRNAs is beyond the scope of this study, we have shown that these novel lncRNAs are highly specific to PCa and track with several coding mRNAs and lncRNAs already reported to be involved in PCa development and progression, most notably, genes involved in immune response. By providing a PCa-specific signature for PTEN loss, as well as highlighting potential new players, we hope to empower further studies on the mechanisms leading to the development of PCa as well its more aggressive subtypes aiding in the future development of potential biomarkers, drug targets and guide therapies choice.

## Acknowledgments

This publication was made possible through support from the National Institutes of Health–National Cancer Institute (NIH-NCI) grants U01CA196390, and R01CA200859; and the U.S. Department of Defense (DoD) office of the Congressionally Directed Medical Research Programs (CDMRP) award W81XWH-16-1-0739 and W81XWH-16-1-0737; and Fundação de Amparo a Pesquisa do Estado de Minas Gerais award BDS-00493-16.

## Author contributions statement

L.M. and T.L. conceived the idea; L.M., E.L.I. and T.L. designed the study; E.L.I., D.F.S., W.D., T.L., and L.M. performed the analysis; E.L.I., D.F.S., T.L., T.V., G.R.F., and L.M. interpreted the results; T.L., E.M.E., S.T., L.M., M.L., and E.M.S. provided data and tools; E.L.I., D.F.S., T.L., and L.M. wrote the manuscript; all authors reviewed and approved the manuscript.

## References

1. Cancer Genome Atlas Research Network, TCGA. The Molecular Taxonomy of Primary Prostate Cancer. Cell [Internet]. 2015 Nov 5;163(4):1011–25. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26544944

2. Taylor BS, Schultz N, Hieronymus H, Gopalan A, Xiao Y, Carver BS, et al. Integrative genomic profiling of human prostate cancer. Cancer Cell. 2010;18(1):11–22.

3. Baca S, Garraway L. The genomic landscape of prostate cancer. Front Endocrinol (Lausanne) [Internet]. 2012;3:69. Available from: https://www.frontiersin.org/article/10.3389/fendo.2012.00069

4. Jamaspishvili T, Berman DM, Ross AE, Scher HI, De Marzo AM, Squire JA, et al. Clinical implications of PTEN loss in prostate cancer. Nat Rev Urol [Internet]. 2018;15(4):222–34. Available from: http://dx.doi.org/10.1038/nrurol.2018.9

5. Lee Y-R, Chen M, Pandolfi PP. The functions and regulation of the PTEN tumour suppressor: new modes and prospects. Nat Rev Mol Cell Biol [Internet]. 2018;19(9):1–16. Available from: http://dx.doi.org/10.1038/s41580-018-0015-0

6. Saal LH, Johansson P, Holm K, Gruvberger-Saal SK, She Q-BBQ-B, Maurer M, et al. Poor prognosis in carcinoma is associated with a gene expression signature of aberrant PTEN tumor suppressor pathway activity. Proc Natl Acad Sci [Internet]. 2007 May 1;104(18):7564–9. Available from: http://www.pnas.org/content/104/18/7564

7. Ong CW, Maxwell P, Alvi MA, McQuaid S, Waugh D, Mills I, et al. A gene signature associated with PTEN activation defines good prognosis intermediate risk prostate cancer cases. J Pathol Clin Res. 2018;4(2):103–13.

8. Morais CL, Han JS, Gordetsky J, Nagar MS, Anderson AE, Lee S, et al. Utility of PTEN and ERG immunostaining for distinguishing high-grade PIN from intraductal carcinoma of the prostate on needle biopsy. Am J Surg Pathol [Internet]. 2015 Feb;39(2):169–78. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25517949

9. Lotan TL, Wei W, Ludkovski O, Morais CL, Guedes LB, Jamaspishvili T, et al. Analytic validation of a clinical-grade PTEN immunohistochemistry assay in prostate cancer by comparison with PTEN FISH. Nat Genet. 2016;29(8):904–14.

10. Lotan TL, Heumann A, Rico SD, Hicks J, Lecksell K, Koop C, et al. PTEN loss detection in prostate cancer: comparison of PTEN immunohistochemistry and PTEN FISH in a large retrospective prostatectomy cohort. Oncotarget. 2017;8(39):65566–76.

11. Han B, Mehra R, Lonigro RJ, Wang L, Suleman K, Menon A, et al. Fluorescence in situ hybridization study shows association of PTEN deletion with ERG rearrangement during prostate cancer progression. Mod Pathol [Internet]. 2009 Aug 1;22(8):1083–93. Available from: http://www.nature.com/articles/modpathol200969

12. Leapman MS, Nguyen HG, Cowan JE, Xue L, Stohr B, Simko J, et al. Comparing Prognostic Utility of a Single-marker Immunohistochemistry Approach with Commercial Gene Expression Profiling Following Radical Prostatectomy. Eur Urol [Internet]. 2018;74(5):668–75. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30181067

13. Ahearn TU, Pettersson A, Ebot EM, Gerke T, Graff RE, Morais CL, et al. A Prospective Investigation of PTEN Loss and ERG Expression in Lethal Prostate Cancer. J Natl Cancer Inst [Internet]. 2016 Feb;108(2). Available from: http://www.ncbi.nlm.nih.gov/pubmed/26615022

14. Lotan TL, Tomlins SA, Bismar TA, Van der Kwast TH, Grignon D, Egevad L, et al. Report From the International Society of Urological Pathology (ISUP) Consultation Conference on Molecular Pathology of Urogenital Cancers. I. Molecular Biomarkers in Prostate Cancer. Am J Surg Pathol [Internet]. 2020 Jul;44(7):e15–29. Available from: http://www.ncbi.nlm.nih.gov/pubmed/32044806

15. Misawa A, Takayama KI, Inoue S. Long non-coding RNAs and prostate cancer. Cancer Sci. 2017;108(11):2107–14.

16. Alimonti A, Carracedo A, Clohessy JG, Trotman LC, Nardella C, Egia A, et al. Subtle variations in Pten dose determine cancer susceptibility. Nat Genet. 2010;42(5):454–8.

17. King JC, Xu J, Wongvipat J, Hieronymus H, Carver BS, Leung DH, et al. Cooperativity of TMPRSS2-ERG with PI3-kinase pathway activation in prostate oncogenesis. Nat Genet [Internet]. 2009 May;41(5):524–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19396167

18. Carver BS, Tran J, Gopalan A, Chen Z, Shaikh S, Carracedo A, et al. Aberrant ERG expression cooperates with loss of PTEN to promote cancer progression in the prostate. Nat Genet [Internet]. 2009 May;41(5):619–24. Available from: http://www.nature.com/doifinder/10.1038/ng.370

19. Ross AE, Johnson MH, Yousefi K, Davicioni E, Netto GJ, Marchionni L, et al. Tissue-based Genomics Augments Post-prostatectomy Risk Stratification in a Natural History Cohort of Intermediate- and High-Risk Men. Eur Urol [Internet]. 2016 Jan;69(1):157–65. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26058959

20. 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 [Internet]. 2015 Jan;24(1):255–60. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25371445

21. Imada EL, Sanchez DF, Collado-Torres L, Wilks C, Matam T, Dinalankara W, et al. Recounting the FANTOM CAGE--Associated Transcriptome. Genome Res. 2020;30(7):gr--254656.

22. Lotan TL, Gurel B, Sutcliffe S, Esopi D, Liu W, Xu J, et al. PTEN Protein Loss by Immunostaining: Analytic Validation and Prognostic Indicator for a High Risk Surgical Cohort of Prostate Cancer Patients. Clin Cancer Res [Internet]. 2011 Oct 15;17(20):6563–73. Available from: http://clincancerres.aacrjournals.org/cgi/doi/10.1158/1078-0432.CCR-11-1244

23. Scharpf RB, Tjelmeland H, Parmigiani G, Nobel AB. A Bayesian model for cross-study differential gene expression. J Am Stat Assoc. 2009;104(488):1295–310.

372 24. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments.
373 Stat Appl Genet Mol Biol [Internet]. 2004;3:Article3. Available from:
374 http://www.ncbi.nlm.nih.gov/pubmed/16646809?dopt=abstract
375 25. Law CW, Chen Y, Shi W, Smyth GK. voom: precision weights unlock linear model analysis tools for RNA-seq read counts.
376 Genome Biol [Internet]. 2014;15(2):R29. Available from: http://genomebiology.biomedcentral.com/articles/10.1186/gb-
377 2014-15-2-r29
378 26. Leek JT, Storey JD. Capturing heterogeneity in gene expression studies by surrogate variable analysis. PLoS Genet [Internet].
379 2007 Sep;3(9):1724–35. Available from:
380 http://www.plosgenetics.org/article/info%3Adoi%2F10.1371%2Fjournal.pgen.0030161
381 27. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R
382 Stat Soc [Internet]. 1995;Series B,(1):289–300. Available from:
383 http://www.dm.uba.ar/materias/analisis_expl_y_conf_de_datos_de_exp_de_marrays_Mae/2006/1/teoricas/FDR
384 1995.pdf
385 28. Sergushichev AA. An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation.
386 BioRxiv [Internet]. 2016;60012. Available from: https://www.biorxiv.org/content/10.1101/060012v1
387 29. Schaeffer EM, Marchionni L, Huang Z, Simons B, Blackman A, Yu W, et al. Androgen-induced programs for prostate
388 epithelial growth and invasion arise in embryogenesis and are reactivated in cancer. Oncogene [Internet]. 2008 Dec
389 4;27(57):7180–91. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18794802
390 30. Leinonen KA, Saramaki OR, Furusato B, Kimura T, Takahashi H, Egawa S, et al. Loss of PTEN Is Associated with Aggressive
391 Behavior in ERG-Positive Prostate Cancer. Cancer Epidemiol Biomarkers Prev [Internet]. 2013 Dec 1;22(12):2333–44.
392 Available from: http://cebp.aacrjournals.org/cgi/doi/10.1158/1055-9965.EPI-13-0333-T
393 31. Yoshimoto M, Ludkovski O, DeGrace D, Williams JL, Evans A, Sircar K, et al. PTEN genomic deletions that characterize
394 aggressive prostate cancer originate close to segmental duplications. Genes, Chromosom Cancer. 2012;51(2):149–60.
395 32. Mehra R, Salami SS, Lonigro R, Bhalla R, Siddiqui J, Cao X, et al. Association of ERG/PTEN status with biochemical recurrence
396 after radical prostatectomy for clinically localized prostate cancer. Med Oncol [Internet]. 2018 Oct 5;35(12):152. Available
397 from: http://www.ncbi.nlm.nih.gov/pubmed/30291535
398 33. Krohn A, Freudenthaler F, Harasimowicz S, Kluth M, Fuchs S, Burkhardt L, et al. Heterogeneity and chronology of PTEN
399 deletion and ERG fusion in prostate cancer. Mod Pathol. 2014;27(12):1612.
400 34. Massoner P, Kugler KG, Unterberger K, Kuner R, Mueller LAJ, Fälth M, et al. Characterization of transcriptional changes in

ERG rearrangement-positive prostate cancer identifies the regulation of metabolic sensors such as neuropeptide Y. PLoS One [Internet]. 2013;8(2):e55207. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23390522

35. Network CGAR, Weinstein JN, Collisson EA, Mills GB, Shaw KRM, Ozenberger BA, et al. The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet [Internet]. 2013 Oct;45(10):1113–20. Available from: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=24071849&retmode=ref&cmd=prlinks

36. Hon C-C, Ramilowski JA, Harshbarger J, Bertin N, Rackham OJLL, Gough J, et al. An atlas of human long non-coding RNAs with accurate 5′ ends. Nature [Internet]. 2017 Mar 1;543(7644):199–204. Available from: http://dx.doi.org/10.1038/nature21374

37. Salameh A, Lee AK, Cardó-Vila M, Nunes DN, Efstathiou E, Staquicini FI, et al. PRUNE2 is a human prostate cancer suppressor regulated by the intronic long noncoding RNA PCA3. Proc Natl Acad Sci [Internet]. 2015;112(27):8403–8. Available from: http://www.pnas.org/lookup/doi/10.1073/pnas.1507882112

38. He JH, Li BX, Han ZP, Zou MX, Wang L, Lv YB, et al. Snail-activated long non-coding RNA PCA3 up-regulates PRKD3 expression by miR-1261 sponging, thereby promotes invasion and migration of prostate cancer cells. Tumor Biol [Internet]. 2016;37(12):16163–76. Available from: http://dx.doi.org/10.1007/s13277-016-5450-y

39. Lemos AEGEGEG, Ferreira LB, Batoreu NM, de Freitas PP, Bonamino MH, Gimba ERP. PCA3 long noncoding RNA modulates the expression of key cancer-related genes in LNCaP prostate cancer cells. Tumor Biol [Internet]. 2016;37(8):11339–48. Available from: http://dx.doi.org/10.1007/s13277-016-5012-3

40. Agarwal R, D'Souza T, Morin PJ. Claudin-3 and claudin-4 expression in ovarian epithelial cells enhances invasion and is associated with increased matrix metalloproteinase-2 activity. Cancer Res. 2005;65(16):7378–85.

41. Srikantan V, Zou Z, Petrovics G, Xu L, Augustus M, Davis L, et al. PCGEM1, a prostate-specific gene, is overexpressed in prostate cancer. Proc Natl Acad Sci [Internet]. 2000;97(22):12216–21. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.97.22.12216

42. Petrovics G, Zhang W, Makarem M, Street JP, Connelly R, Sun L, et al. Elevated expression of PCGEM1, a prostate-specific gene with cell growth-promoting function, is associated with high-risk prostate cancer patients. Oncogene. 2004;23(2):605.

43. Lin X, Shang X, Manorek G, Howell SB. Regulation of the epithelial-mesenchymal transition by claudin-3 and claudin-4. PLoS One. 2013;8(6):e67496.

44. Song YX, Sun JX, Zhao JH, Yang YC, Shi JX, Wu ZH, et al. Non-coding RNAs participate in the regulatory network of CLDN4 via ceRNA mediated miRNA evasion. Nat Commun [Internet]. 2017;8(1):1–16. Available from: http://dx.doi.org/10.1038/s41467-017-00304-1

45. Prensner JR, Iyer MK, Sahu A, Asangani IA, Cao Q, Patel L, et al. The long noncoding RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. Nat Genet. 2013;45(11):1392–403.

46. Mehra R, Udager AM, Ahearn TU, Cao X, Feng FY, Loda M, et al. Overexpression of the long non-coding RNA SChLAP1 independently predicts lethal prostate cancer. Eur Urol. 2016;70(4):549–52.

47. Kumar-Sinha C, Tomlins SA, Chinnaiyan AM. Recurrent gene fusions in prostate cancer. Nat Rev Cancer. 2008;8(7):497.

48. Nam RK, Zhang WW, Klotz LH, Trachtenberg J, Jewett MAS, Sweet J, et al. Variants of the hK2 protein gene (KLK2) are associated with serum hK2 levels and predict the presence of prostate cancer at biopsy. Clin cancer Res. 2006;12(21):6452–8.

49. Cicek MS, Liu X, Casey G, Witte JS. Role of androgen metabolism genes CYP1B1, PSA/KLK3, and CYP11$α$ in prostate cancer risk and aggressiveness. Cancer Epidemiol Prev Biomarkers. 2005;14(9):2173–7.

50. Whiteland H, Spencer-Harty S, Morgan C, Kynaston H, Thomas DH, Bose P, et al. A role for STEAP2 in prostate cancer progression. Clin \& Exp metastasis. 2014;31(8):909–20.

51. Perner S, Rupp NJ, Braun M, Rubin MA, Moch H, Dietel M, et al. Loss of SLC45A3 protein (prostein) expression in prostate cancer is associated with SLC45A3-ERG gene rearrangement and an unfavorable clinical course. Int J cancer. 2013;132(4):807–12.

52. Patel N, Itakura T, Jeong S, Liao C-P, Roy-Burman P, Zandi E, et al. Expression and Functional Role of Orphan Receptor GPR158 in Prostate Cancer Growth and Progression. Robson CN, editor. PLoS One [Internet]. 2015 Feb 18;10(2):e0117758. Available from: https://dx.plos.org/10.1371/journal.pone.0117758

53. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. Bioinformatics [Internet]. 2011 Jun 15;27(12):1739–40. Available from: https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btr260

54. 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 [Internet]. 2005 Oct 25;102(43):15545–50. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16199517

55. Kaur HB, Guedes LB, Lu J, Maldonado L, Reitz L, Barber JR, et al. Association of tumor-infiltrating T-cell density with molecular subtype, racial ancestry and clinical outcomes in prostate cancer. Mod Pathol [Internet]. 2018;31(10):1539–52. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29849114

56. Trigunaite A, Dimo J, Jørgensen TN. Suppressive effects of androgens on the immune system. Cell Immunol [Internet]. 2015 Apr;294(2):87–94. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25708485

159 57. Ylitalo EB, Thysell E, Jernberg E, Lundholm M, Crnalic S, Egevad L, et al. Subgroups of Castration-resistant Prostate Cancer
160 Bone Metastases Defined Through an Inverse Relationship Between Androgen Receptor Activity and Immune Response.
161 Eur Urol [Internet]. 2017;71(5):776–87. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27497761
162 58. Wise HM, Hermida MA, Leslie NR. Prostate cancer, PI3K, PTEN and prognosis. Clin Sci. 2017;131(3):197–210.
163 59. Lotan TL, Gurel B, Sutcliffe S, Esopi D, Liu W, Xu J, et al. PTEN protein loss by immunostaining: analytic validation and
164 prognostic indicator for a high risk surgical cohort of prostate cancer patients. Clin Cancer Res [Internet]. 2011 Oct
165 15;17(20):6563–73. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21878536
166 60. Guo Y, Li H, Guan H, Ke W, Liang W, Xiao H, et al. Dermatopontin inhibits papillary thyroid cancer cell proliferation through
167 MYC repression. Mol Cell Endocrinol. 2019;480:122–32.
168 61. Yamatoji M, Kasamatsu A, Kouzu Y, Koike H, Sakamoto Y, Ogawara K, et al. Dermatopontin: a potential predictor for
169 metastasis of human oral cancer. Int J cancer. 2012;130(12):2903–11.
170 62. Ishii K, Usui S, Sugimura Y, Yoshida S, Hioki T, Tatematsu M, et al. Aminopeptidase N regulated by zinc in human prostate
171 participates in tumor cell invasion. Int J cancer. 2001;92(1):49–54.
172 63. Hashida H, Takabayashi A, Kanai M, Adachi M, Kondo K, Kohno N, et al. Aminopeptidase N is involved in cell motility and
173 angiogenesis: its clinical significance in human colon cancer. Gastroenterology. 2002;122(2):376–86.
174 64. Sørensen KD, Abildgaard MO, Haldrup C, Ulhøi BP, Kristensen H, Strand S, et al. Prognostic significance of aberrantly
175 silenced ANPEP expression in prostate cancer. Br J Cancer. 2013;108(2):420–8.
176 65. Hossain S, Takatori A, Nakamura Y, Suenaga Y, Kamijo T, Nakagawara A. LRRN1 enhances EGF-mediated MYCN induction
177 in neuroblastoma and accelerates tumor growth in vivo. Cancer Res. 2012;72(17):4587–96.
178 66. Hossain MS, Ozaki T, Wang H, Nakagawa A, Takenobu H, Ohira M, et al. N-MYC promotes cell proliferation through a direct
179 transactivation of neuronal leucine-rich repeat protein-1 (NLRR1) gene in neuroblastoma. Oncogene. 2008;27(46):6075–
180 82.
181 67. Liao C-H, Wang Y-H, Chang W-W, Yang B-C, Wu T-J, Liu W-L, et al. Leucine-Rich Repeat Neuronal Protein 1 Regulates
182 Differentiation of Embryonic Stem Cells by Post-Translational Modifications of Pluripotency Factors. Stem Cells.
183 2018;36(10):1514–24.
184 68. Wilhelm BT, Landry J-R. RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing.
185 Methods [Internet]. 2009 Jul;48(3):249–57. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19336255
186 69. Hardwick JP. Cytochrome P450 omega hydroxylase (CYP4) function in fatty acid metabolism and metabolic diseases.
187 Biochem Pharmacol. 2008;75(12):2263–75.

488 70. Coon TA, Glasser JR, Mallampalli RK, Chen BB. Novel E3 ligase component FBXL7 ubiquitinates and degrades Aurora A,
489 causing mitotic arrest. Cell cycle. 2012;11(4):721–9.
490 71. Katsha A, Soutto M, Sehdev V, Peng D, Washington MK, Piazuelo MB, et al. Aurora kinase A promotes inflammation and
491 tumorigenesis in mice and human gastric neoplasia. Gastroenterology [Internet]. 2013 Dec;145(6):1312-22.e1-8. Available
492 from: http://www.ncbi.nlm.nih.gov/pubmed/23993973
493 72. Fenner A. Prostate cancer: Orphan receptor GPR158 finds a home in prostate cancer growth and progression. Nat Rev Urol
494 [Internet]. 2015 Apr;12(4):182. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25753095
495 73. Li S, Zhu M, Pan R, Fang T, Cao Y-YY, Chen S, et al. The tumor suppressor PTEN has a critical role in antiviral innate immunity.
496 Nat Immunol. 2016;17(3):241.
497 74. Vidotto T, Melo CM, Castelli E, Koti M, dos Reis RB, Squire JA. Emerging role of PTEN loss in evasion of the immune response
498 to tumours. Br J Cancer [Internet]. 2020;122(12):1732–43. Available from: http://dx.doi.org/10.1038/s41416-020-0834-6
499 75. Lastwika KJ, Wilson W, Li QK, Norris J, Xu H, Ghazarian SR, et al. Control of PD-L1 Expression by Oncogenic Activation of the
500 AKT-mTOR Pathway in Non-Small Cell Lung Cancer. Cancer Res [Internet]. 2016 Jan 15;76(2):227–38. Available from:
501 http://www.ncbi.nlm.nih.gov/pubmed/26637667
502 76. Berghoff AS, Kiesel B, Widhalm G, Rajky O, Ricken G, Wöhrer A, et al. Programmed death ligand 1 expression and tumor-
503 infiltrating lymphocytes in glioblastoma. Neuro Oncol [Internet]. 2015 Aug;17(8):1064–75. Available from:
504 http://www.ncbi.nlm.nih.gov/pubmed/25355681
505 77. Martin AM, Nirschl TR, Nirschl CJ, Francica BJ, Kochel CM, van Bokhoven A, et al. Paucity of PD-L1 expression in prostate
506 cancer: innate and adaptive immune resistance. Prostate Cancer Prostatic Dis [Internet]. 2015 Dec;18(4):325–32. Available
507 from: http://www.ncbi.nlm.nih.gov/pubmed/26260996
508 78. Yuan H, Wei X, Zhang G, Li C, Zhang X, Hou J. B7-H3 over expression in prostate cancer promotes tumor cell progression. J
509 Urol [Internet]. 2011 Sep;186(3):1093–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21784485
510 79. Papanicolau-Sengos A, Yang Y, Pabla S, Lenzo FL, Kato S, Kurzrock R, et al. Identification of targets for prostate cancer
511 immunotherapy. Prostate [Internet]. 2019;79(5):498–505. Available from:
512 http://www.ncbi.nlm.nih.gov/pubmed/30614027
513 80. Yang S, Wei W, Zhao Q. B7-H3, a checkpoint molecule, as a target for cancer immunotherapy. Int J Biol Sci [Internet].
514 2020;16(11):1767–73. Available from: http://www.ncbi.nlm.nih.gov/pubmed/32398947
515 81. Benzon B, Zhao SG, Haffner MC, Takhar M, Erho N, Yousefi K, et al. Correlation of B7-H3 with androgen receptor, immune
516 pathways and poor outcome in prostate cancer: an expression-based analysis. Prostate Cancer Prostatic Dis [Internet].

517 2017;20(1):28–35. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27801901

518 82. Carver BS, Chapinski C, Wongvipat J, Hieronymus H, Chen Y, Chandarlapaty S, et al. Reciprocal feedback regulation of PI3K
519 and androgen receptor signaling in PTEN-deficient prostate cancer. Cancer Cell. 2011;19(5):575–86.

520 83. Gezer U, Tiryakioglu D, Bilgin E, Dalay N, Holdenrieder S. Androgen stimulation of PCA3 and miR-141 and their release from
521 prostate cancer cells. Cell J. 2015;16(4):488.

522 84. Parolia A, Crea F, Xue H, Wang Y, Mo F, Ramnarine VR, et al. The long non-coding RNA PCGEM1 is regulated by androgen
523 receptor activity in vivo. Mol Cancer. 2015;14(1):46.

524

525 ## Figures and Tables

| Cohort | PTEN-null | PTEN-intact | N |
|---|---|---|---|
| TCGA | 95 | 321 | 416 |
| HPFS | 91 | 299 | 390 |
| Natural History | 56 | 151 | 207 |
| Total | 242 | 771 | 1,013 |

526

527 **Table 1.** Cohorts summary Table shows cohorts summary for the 3 cohorts used in this study: TCGA (only primary tumor samples
528 with high Gistic scores were used); Health Professional Follow-up Study (all); and Natural History cohort (samples with IHC call
529 available). PTEN-null represents samples with PTEN deletion and PTEN-intact regular primary tumors.

530

| | PTEN-null vs PTEN-intact overall | PTEN-null vs PTEN-intact in ERG+ | PTEN-null vs PTEN-intact in ERG- |
|---|---|---|---|
| Coding genes | 257 (13) | 226 (7) | 185 (10) |
| Non-coding genes | 264 (134) | 209 (117) | 179 (82) |
| Total | 521 (137) | 435 (124) | 364 (92) |

**Table 2.** Summary of differentially expressed genes between PTEN-null and PTEN-intact with logFC $\geq 1$ and FDR $\leq 0.01$ across different ERG backgrounds. Number in parenthesis shows the number of genes exclusive to the FANTOM-CAT annotations.

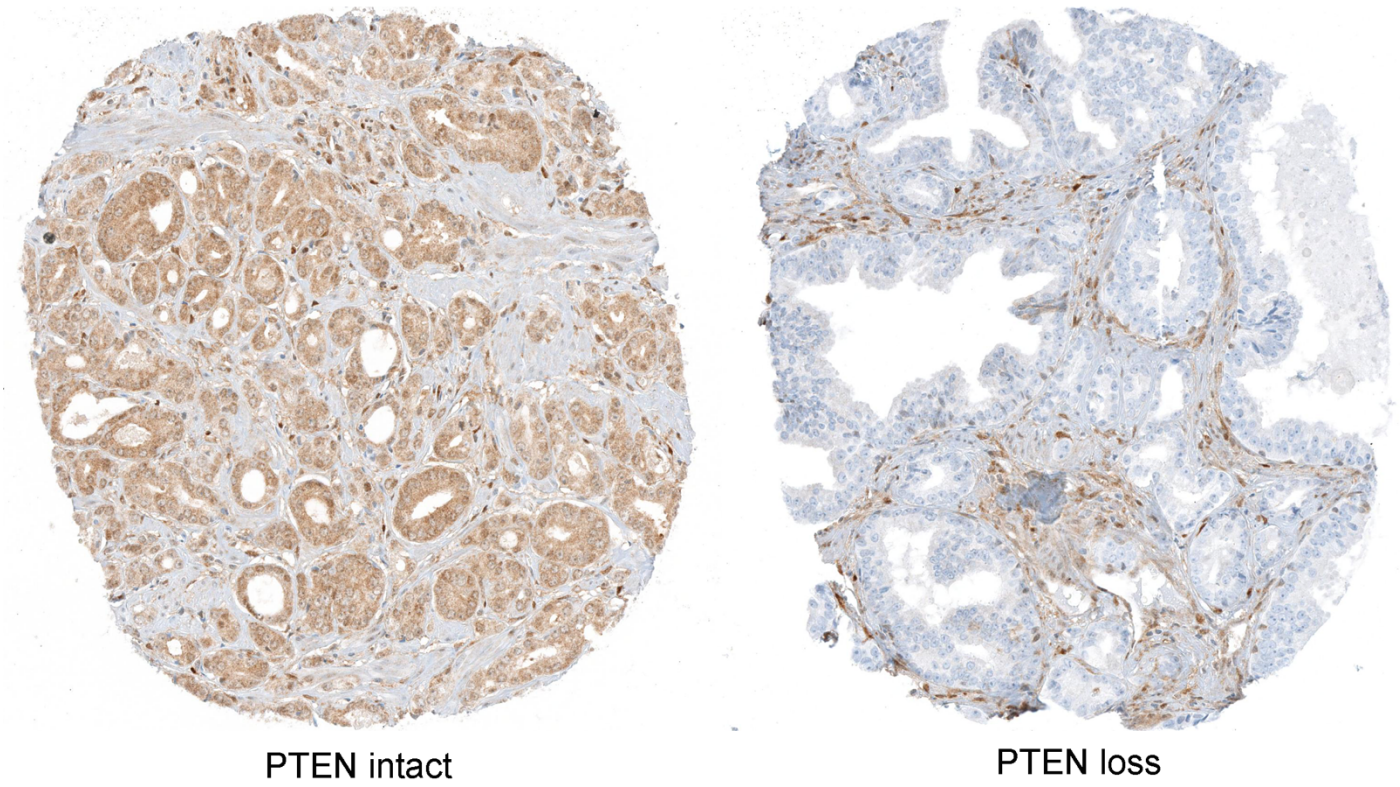



**Figure 1.** PTEN immunostaining in tissue microarray (TMA) spots from the Natural History Cohort. Left panel: intact PTEN protein is present in all sampled tumor glands (brown chromogen). Right panel: PTEN loss in all sampled tumor glands. Images reduced from 40X.

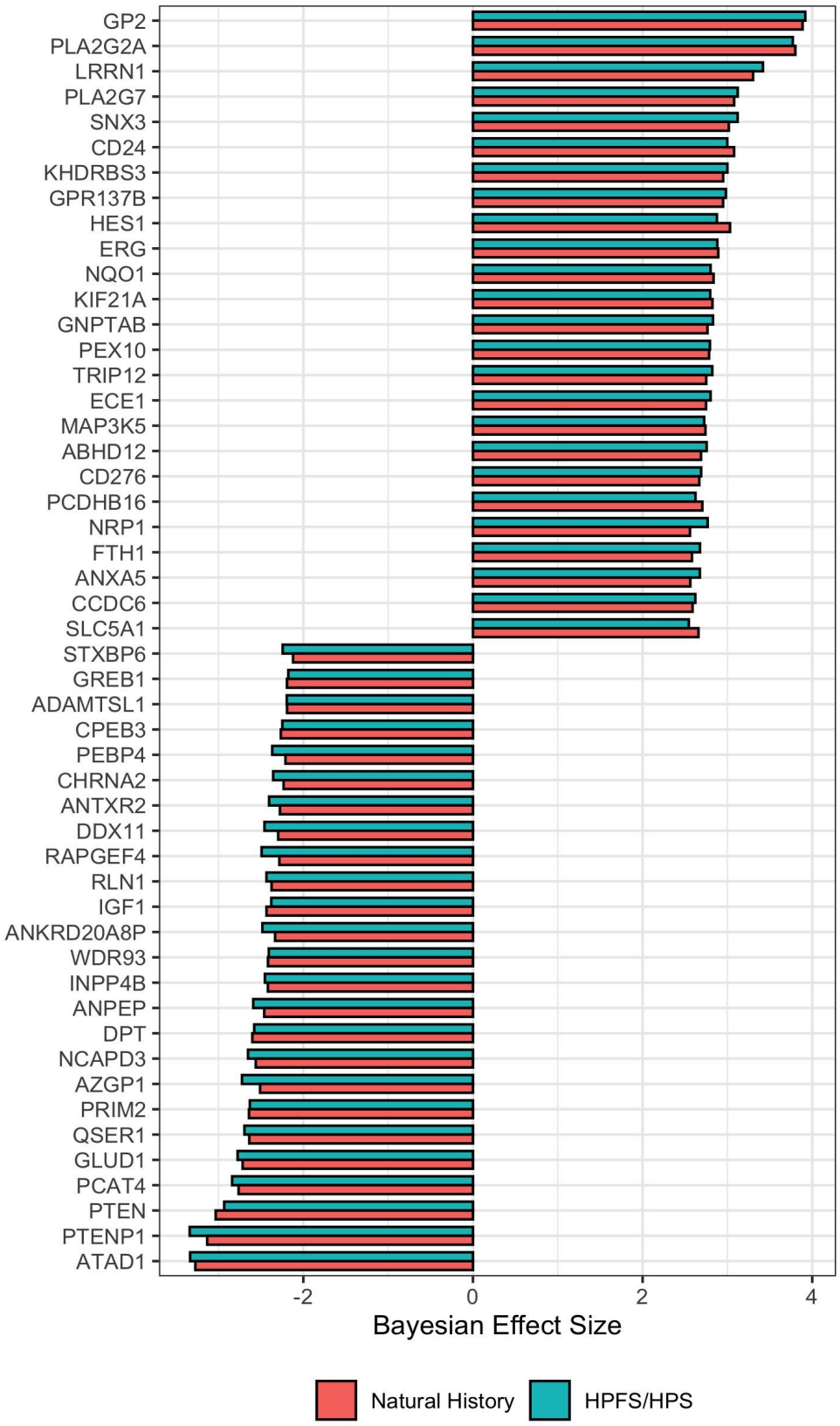



**Figure 2.** Cross-study meta-analysis of differential gene expression. Genes in the same loci as PTEN such as RLN1 and ATAD1 were found down-regulated. PTEN-null vs PTEN-intact meta-analysis of HPFS/PHS and NH cohorts with Bayesian Hierarchical

542 Model for DGE using XDE showing the top 25 most concordant differentially up- and down-regulated genes. PTEN status were

543 based on IHC assays.

544

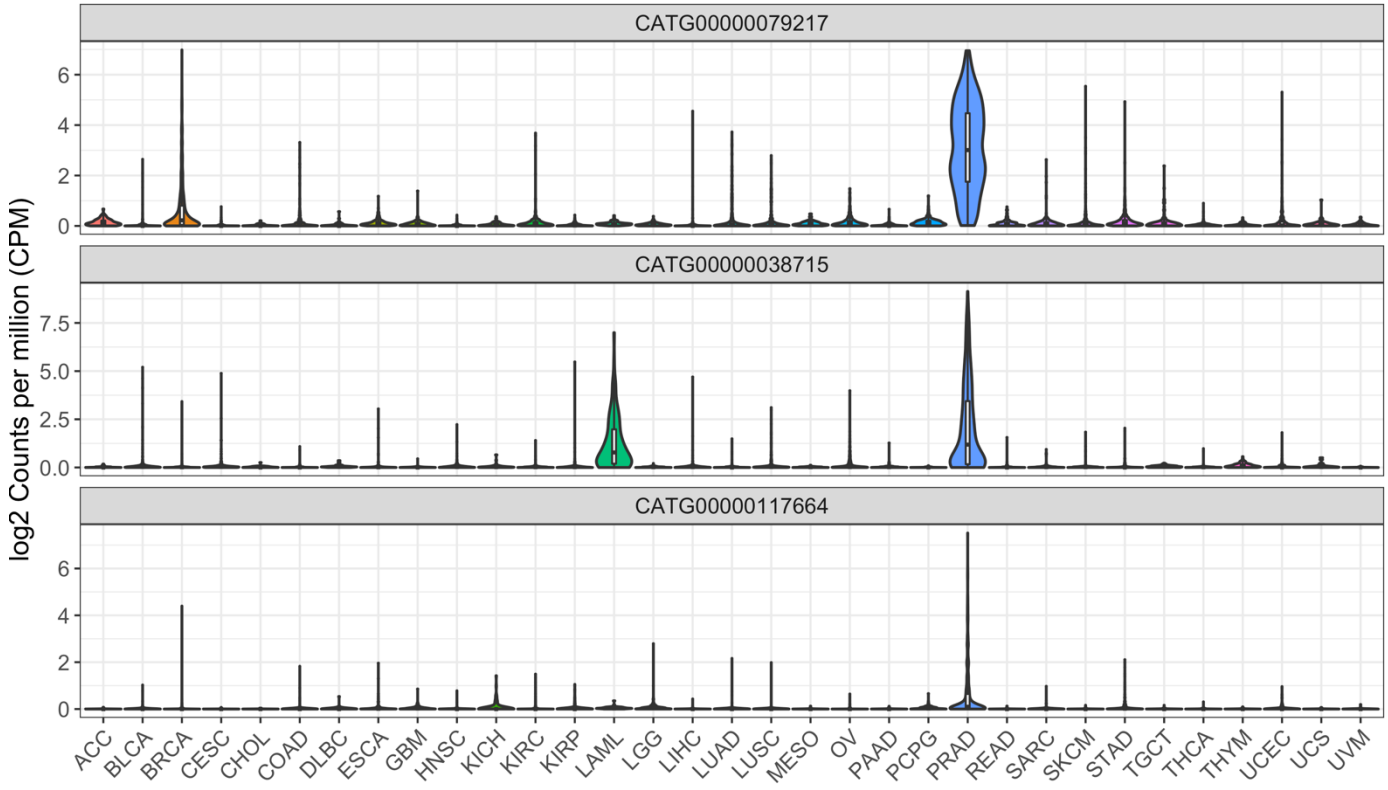



**Figure 3.** Expression profiles of novel FANTOM-CAT genes CATG00000038715, CATG00000079217 and CATG00000117664 across 33 cancer types. Violin-plots shows expression ($log_2$ CPM+1) distribution.

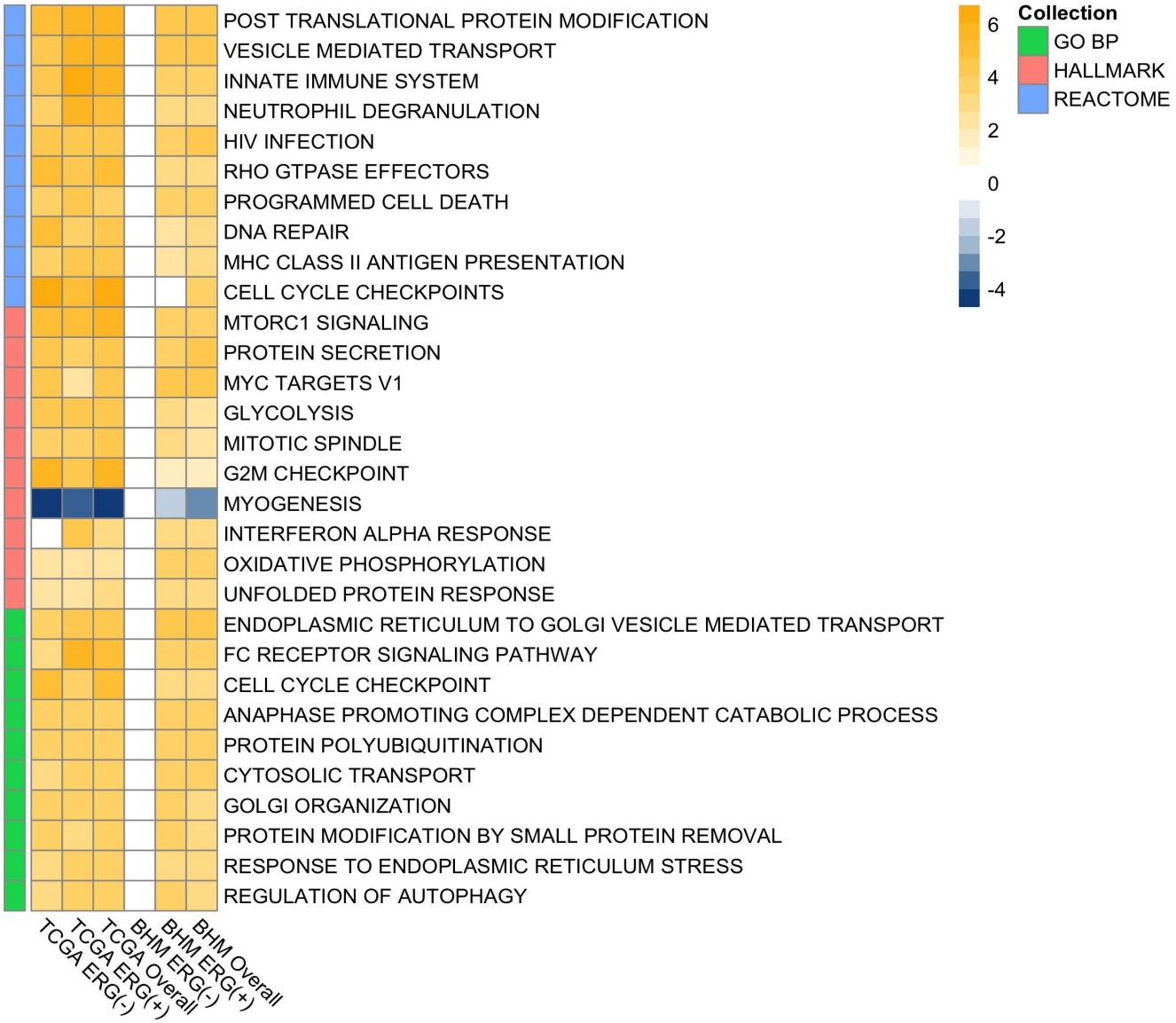



**Figure 4.** Top enriched gene sets enriched across PTEN-null and PTEN-intact in the TCGA and BHM cohorts stratified by ERG status and overall. Heatmap of mean-centered $\log_2$ signed p-values (normalized enriched score multiplied by $\log_{10}$ of p-value) showing the top 10 enriched gene sets of each collection (ranked by signed p-value).

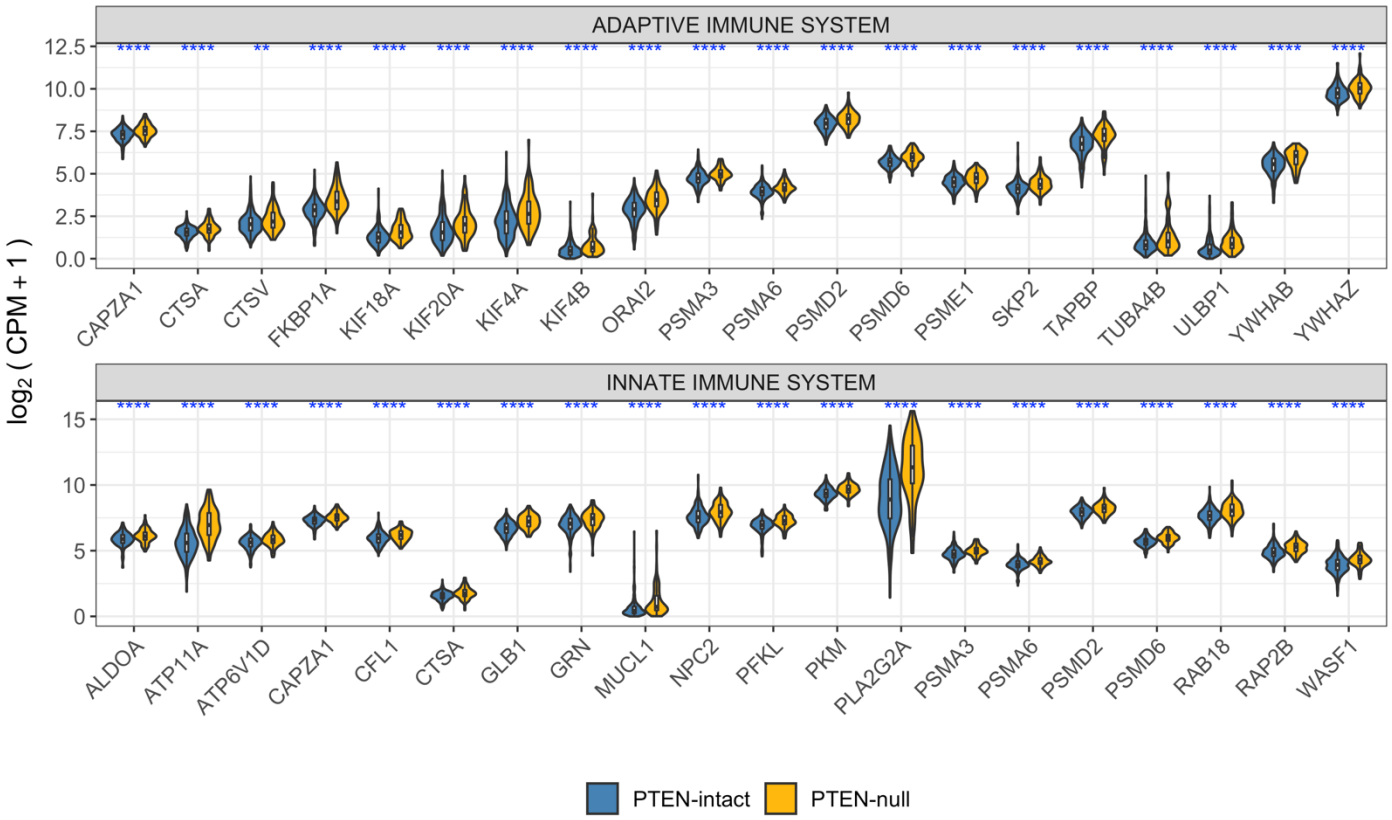
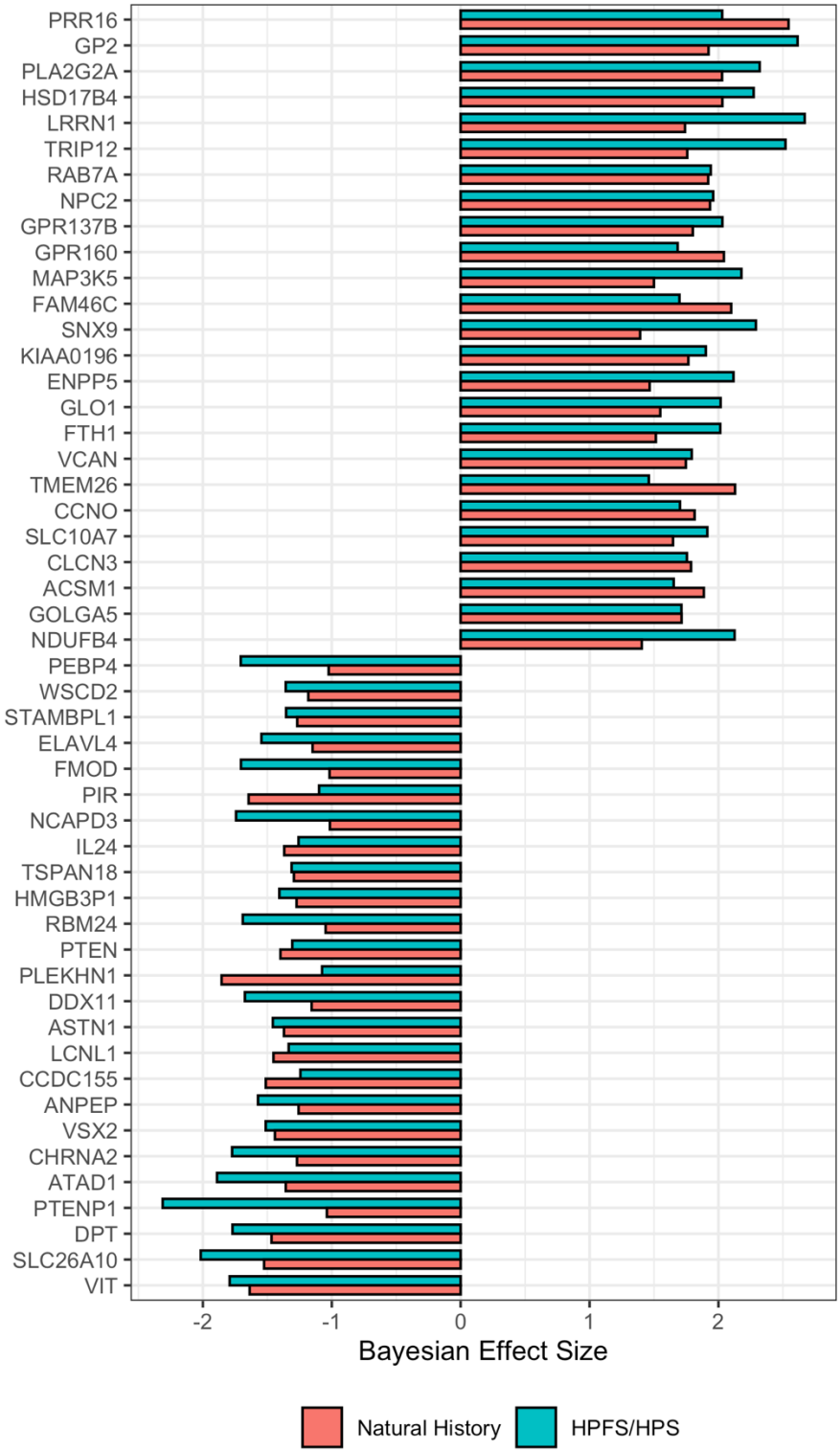



**Figure 5.** Expression of immune-related genes stratified by PTEN status. Top 20 were selected based on the leading edge of the GSEA of the adaptive and innate immune system gene sets from REACTOME. Significances based on t-test between PTEN-null and PTEN-intact using $\log_2$ CPM+1 values. Significance cutoffs: *=≤ 0.05; **≤0.01; *** ≤ 0.001; ****≤0.0001.

## Supplementary Figures and Tables



564 **Figure S1.** Cross-study of differential gene expression in PTEN-null vs PTEN-intact in $ERG^{+}$ samples. Meta-analysis of HPFS/PHS

565 and NH cohorts with Bayesian Hierarchical Model for DGE using XDE showing the top 25 most concordant differentially up- and

566 down-regulated genes. PTEN status were based on IHC assays.

567

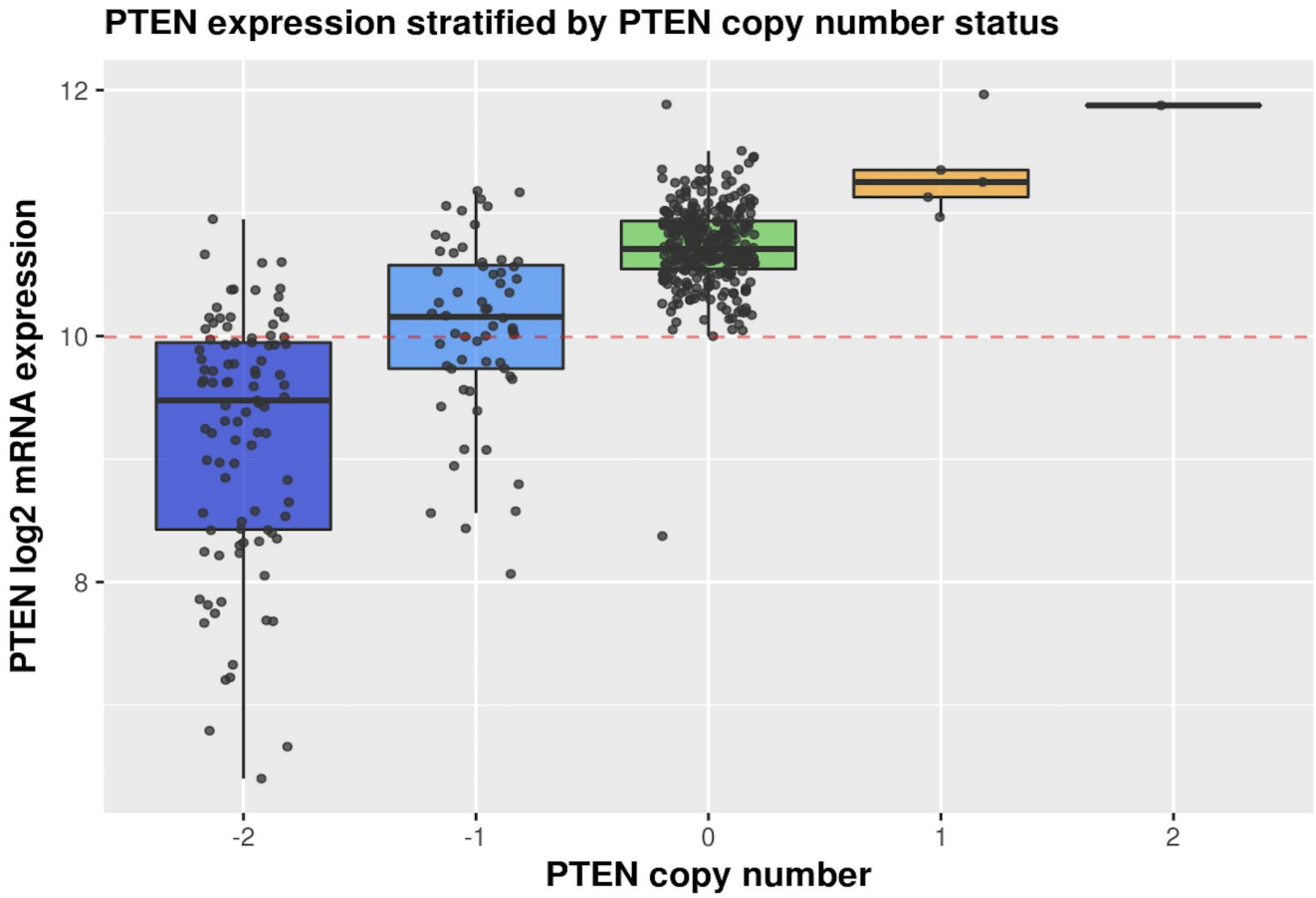
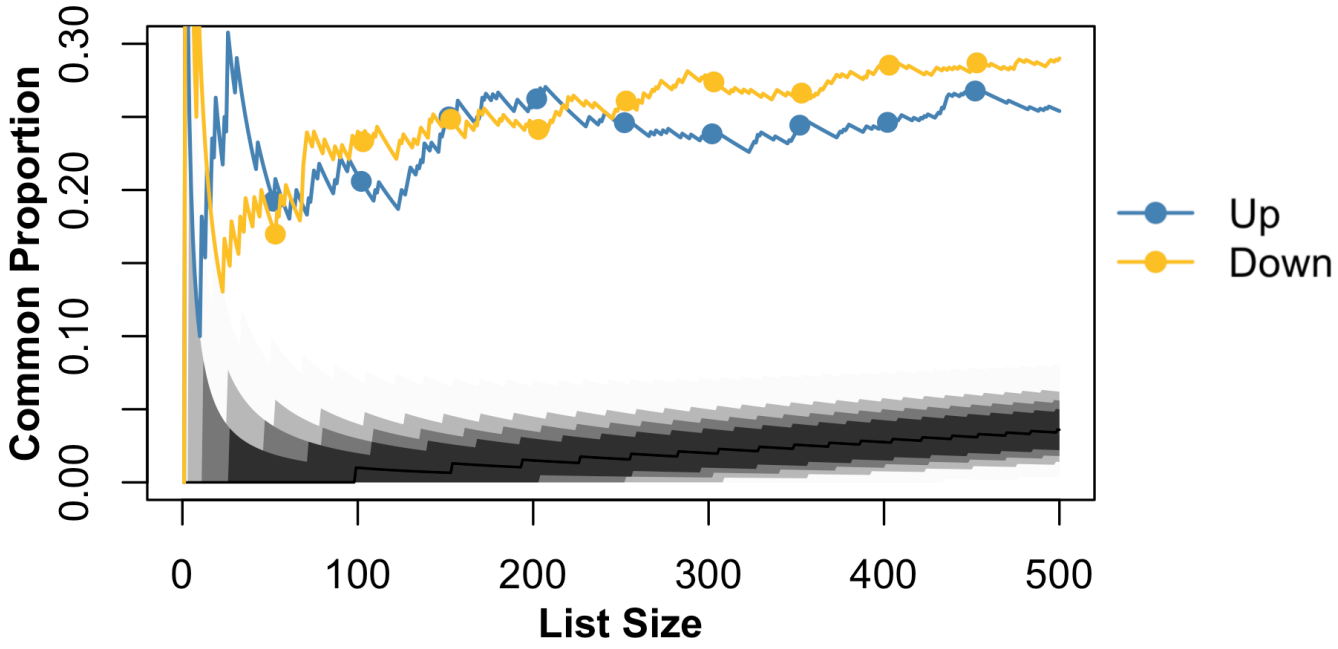



**Figure S3.** PTEN expression levels stratified by CNV. Figure shows PTEN expression levels distribution by copy number variation (CNV), called by GISTIC algorithm.




572

573 **Figure S4.** Correspondence-at-the-top (CAT) plot between TCGA CNV-based calls and the Bayesian Hierarchical Model approach
574 (BHM). Agreement of genes ranked by t-statistics (TCGA) and average Bayesian Effect Size (BHM). Lines represent agreement
575 between tested cohorts for PTEN-intact vs PTEN-null. Black-to-light grey shades represent the decreasing probability of agreeing
576 by chance based on the hypergeometric distribution, with intervals ranging from 0.999999 (light grey) to 0.95 (dark grey). Lines
577 outside this range represent agreement in different cohorts with a higher agreement than expected by chance.

578

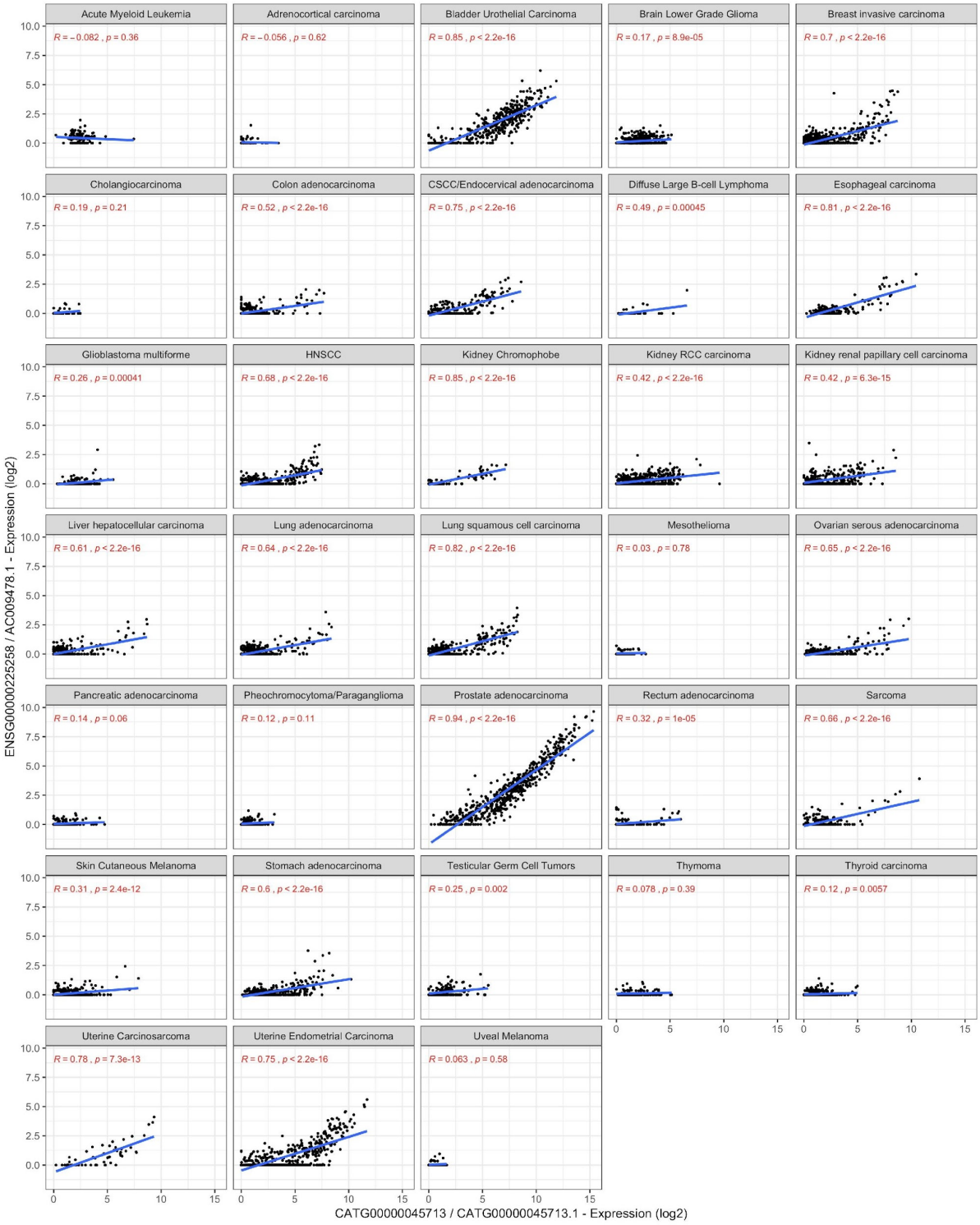



**Figure S5.** Expression of AC009478.1 is shown to be highly specific to PRAD, BLCA, to a lesser extent in UECA and BRCA. Figure shows raw expression values of SchLAP1 and AC009478.1 across cancer types. Pearson correlations and p-values are shown in red.

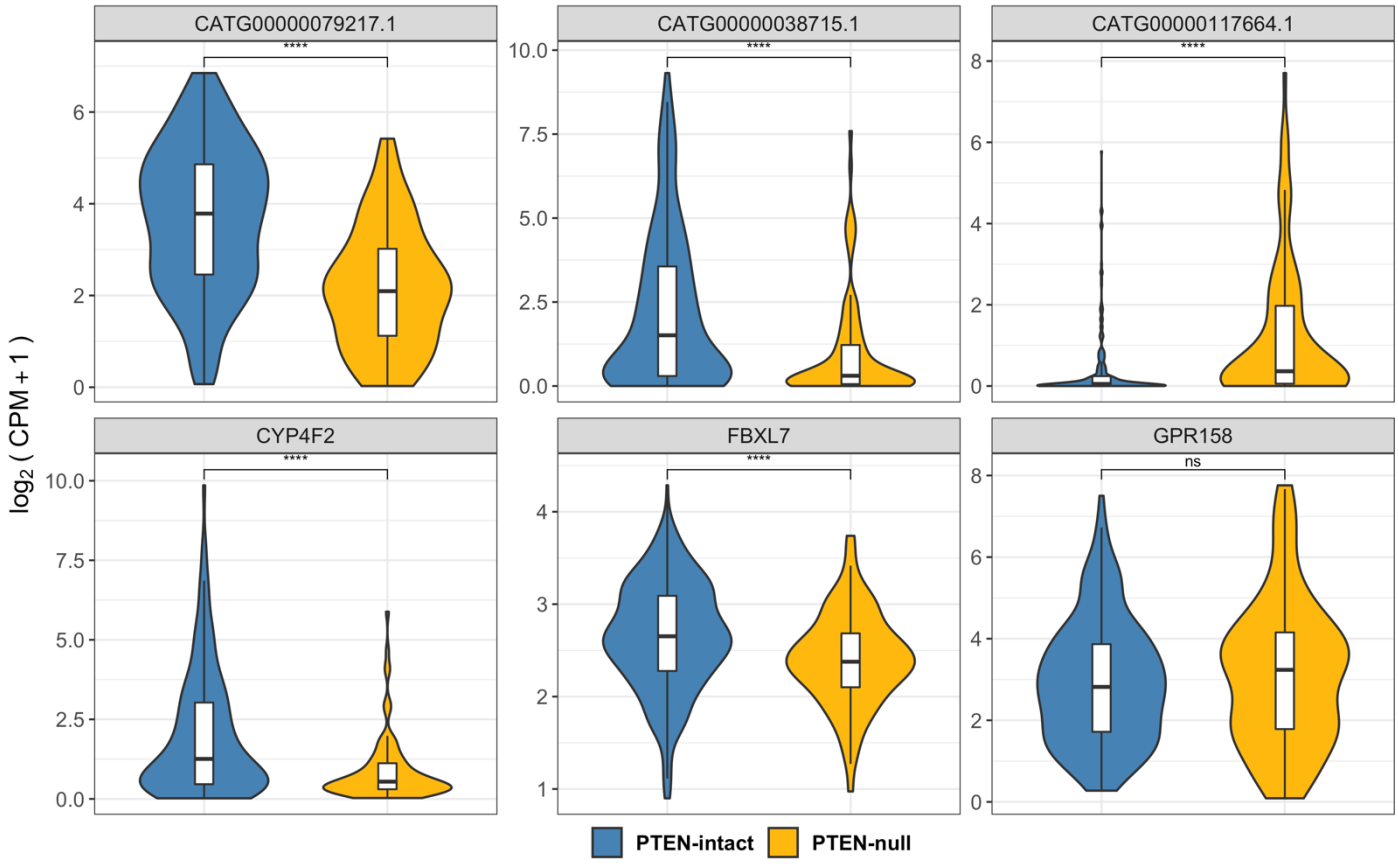



**Figure S6.** Expression of FANTOM-CAT lncRNAs genes (top) and close coding genes (bottom) stratified by PTEN status. Significances based on t-test between PTEN-null and PTEN-intact using $log_2$ CPM+1 value. Significance cutoffs: *=≤ 0.05; **≤0.01; *** ≤ 0.001; ****≤0.0001.

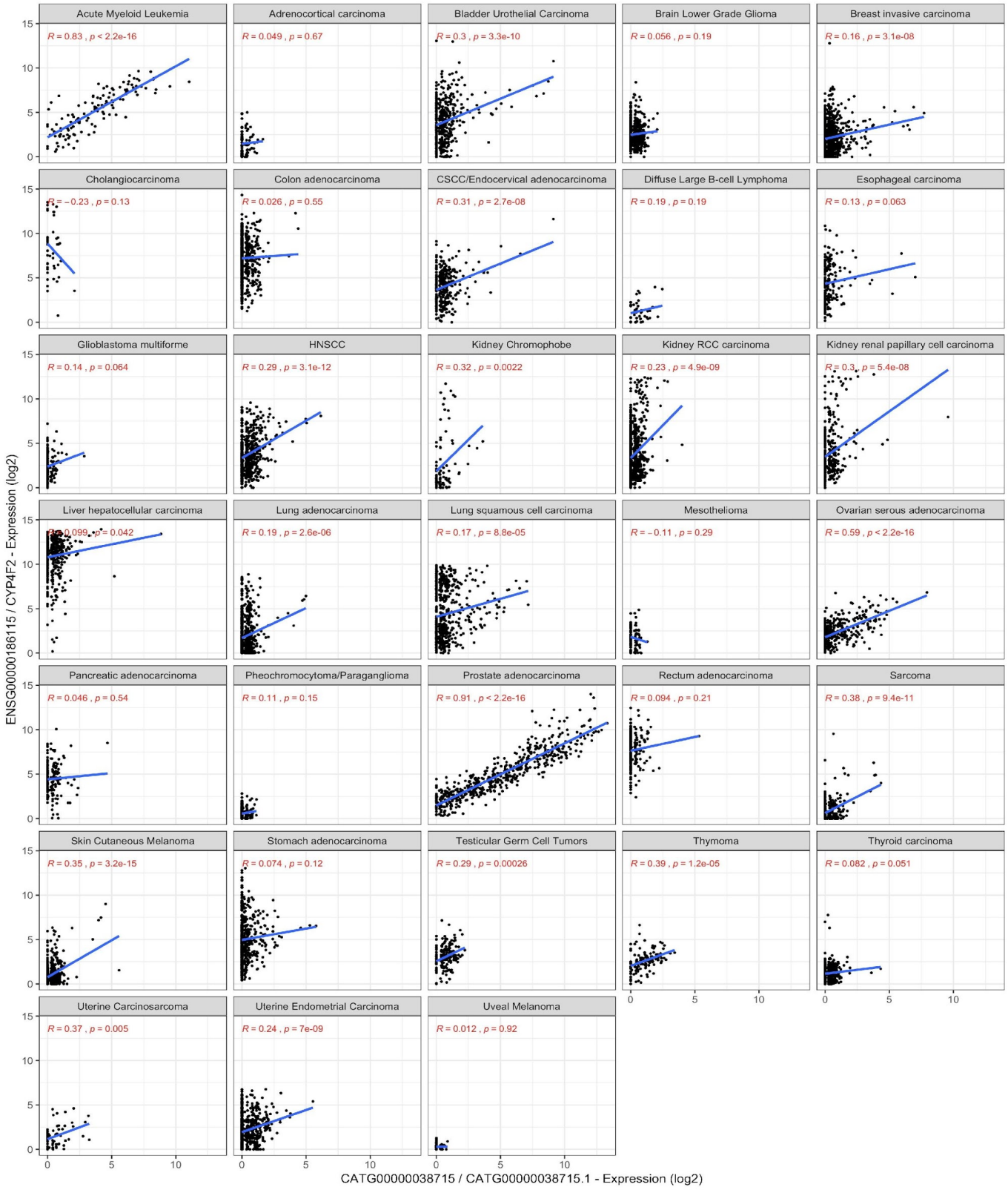



**Figure S7.** Person correlation gene CATG00000038715 and CYP4F2 across cancer types. CATG00000038715 and CYP4F2 expression are shown to be highly correlated in PCa. Moreover, CATG00000038715 expression is shown to be highly specific to PCa. With exception of leukemia cells, none of the other tumors expressed high levels of CATG00000038715.

594

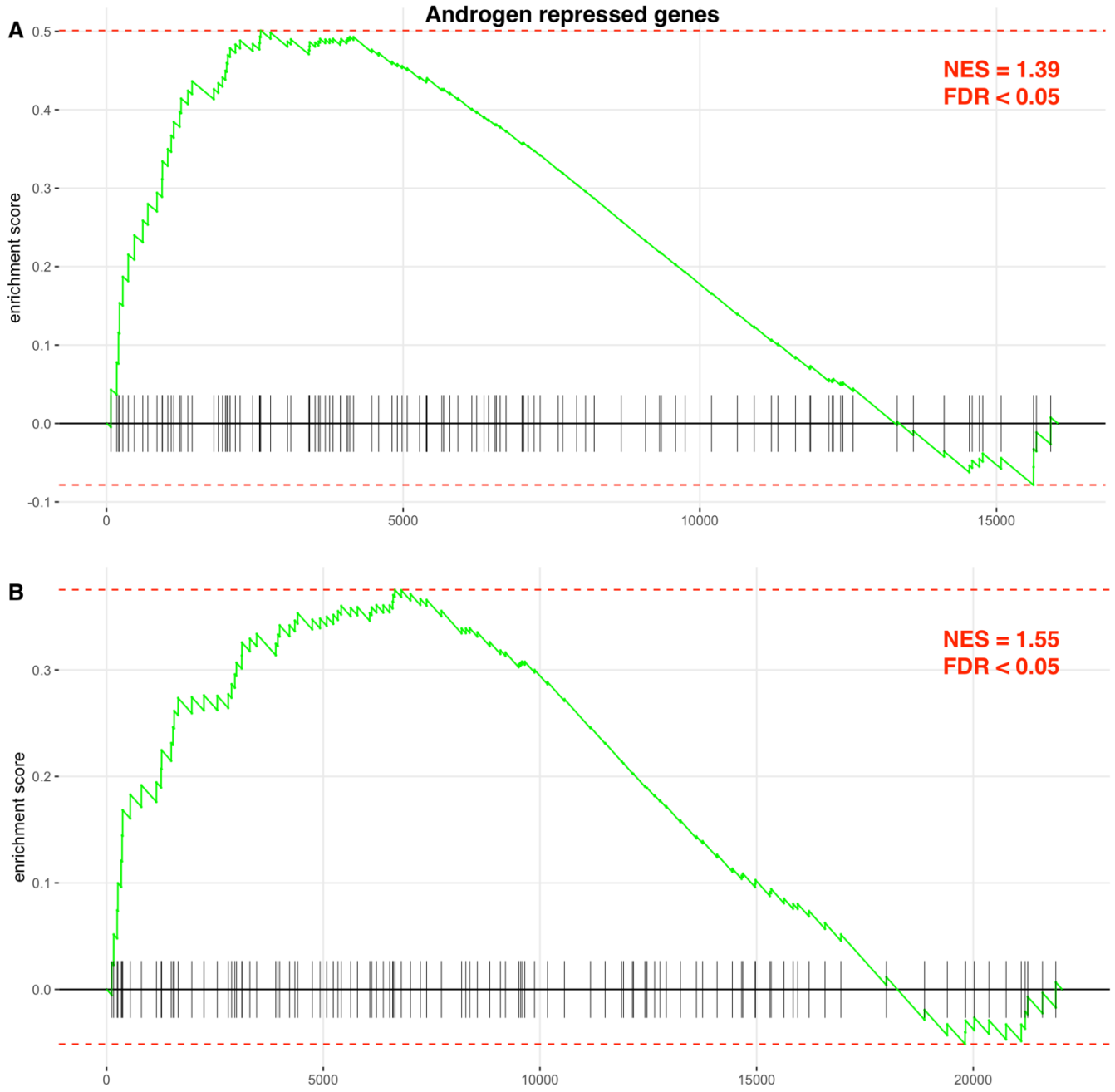



**Figure S8.** Gene set enrichment for Androgen repressed genes. Gene set enrichment analysis of gene signature showing positive enrichment of genes repressed by dihydrotestosterone after 6 hours of exposure obtained from Schaeffer et al.[48]. Enrichment for BHM-signature is shown in panel A and TCGA-signature in panel B.

501