PNAS
www.pnas.org

Supplementary Information for

Aneuploidy drives lethal progression in prostate cancer

Konrad H. Stopsack, Charles A. Whittaker, Travis A. Gerke, Massimo Loda, Philip W. Kantoff, Lorelei A. Mucci, Angelika Amon

Correspondence: Lorelei Mucci or Angelika Amon
Email: lmucci@hsph.harvard.edu or angelika@mit.edu

This PDF file includes:

www.pnas.org/cgi/doi/10.1073/pnas.1902645116

Supplementary Methods

Genomic and histologic data. Tumor tissue in all three cohorts underwent histopathologic re-review by genitourinary pathologists (1). In TCGA, samples were retained if nucleic acid content was sufficient; from HPFS and PHS, tumor areas with >80% tumor cellularity were selected.

TCGA investigators normalized genome-wide copy number estimates against non-cancer normal samples and segmented using circular binary segmentation, effectively filtering out germline variants, and identified focal alterations using GISTIC (2). RNA sequencing used the Illumina TruSeq RNA protocol and the Illumina HiSeq platform.

For tumors from HPFS and PHS, the Affymetrix GeneChip Human Gene 1.0 ST array was used (Gene Expression Omnibus: GSE62872) (3). Tissue microarrays were assayed using genomically validated immunohistochemistry for ERG protein expression (a proxy for *TMPRSS2:ERG* gene fusion), nuclear MYC protein expression, absence of PTEN protein expression (a proxy for *PTEN* loss), the Ki-67 proliferative index (percentage positively stained tumor cells), and the TUNEL index indicating tumor cells undergoing apoptosis (positively stained area) (4-8).

Further information including the procedures to obtain and access data from HPFS is described at https://sites.sph.harvard.edu/hpfs/for-collaborators.

Statistical analysis. To assess how well aneuploidy predicted from the transcriptome discriminated for measured aneuploidy, the AUC was calculated with bootstrapped 95% CIs from 5000 permutations. Calibration was assessed using a Bland-Altman plot. Quantitative tissue biomarkers (Ki-67, TUNEL) were quantile-normalized between tissue microarrays and log-transformed. Associations between markers were assessed using Pearson correlation. Binomial proportions were reported with Wilson CIs. Logistic regression was used to estimate ORs and 95% CIs for predicted aneuploidy and lethal disease. Models for causal effects were adjusted for age at diagnosis (linear), year of diagnosis (categorical, all categorizations as in Table 1), and Gleason score (categorical); prognostic models were additionally adjusted for PSA (categorical, plus missingness indicator), clinical stage (categorical), and primary treatment (binary; prostatectomy *vs*. other). Associations between single-arm events (e.g., 10q), single-gene events (e.g., PTEN), and lethal disease were assessed using logistic regression.

Supplementary Figures

Fig. S1. Distribution of copy number alterations of all chromosome arms, including those less frequently affected by aneuploidy.

Fig. S2. Test characteristics of the transcriptome in predicting DNA copy number-defined aneuploidy.

(A) Discrimination analysis for predicting any aneuploidy (one or more altered chromosome arms) using the transcriptome (yellow line), compared with random chance (gray line). A larger area under the curve (AUC) indicates better performance.

(B) Bland-Altman plot for systematic differences in numbers of altered chromosome arms, comparing measured aneuploidy using DNA copy number alterations and aneuploidy predicted from the transcriptome. The size of the dots indicates the number of tumors with identical values (*n*). Frequency distributions are reflected in the histograms on the top (for the *x* axis) and right (*y* axis). With higher absolute aneuploidy (*x* axis), expressed by the mean between measured and predicted aneuploidy scores, the absolute differences between measured and predicted values (*y* axis) increase; there is no systematic deviation towards more positive or more negative differences (*y* axis) with higher absolute values (*x* axis). Dashed lines indicate 95% CIs of the difference.

Fig. S3. Aneuploidy as predicted from the tumor transcriptome (Health Professionals Follow-up Study, HPFS, and Physicians' Health Study, PHS).
Shown is the distribution of each tumor's sum of mRNA expression levels, normalized in standard deviations, per chromosome arm. Predicted chromosome arm gains (yellow) and losses (blue) are highlighted. Compare to Fig. 2C, predicted aneuploidy in TCGA.

Fig. S4. Gleason score and aneuploidy scores.
In all plots, the boxes extend from the first to the third quartile; the horizontal line indicates the median; whiskers extend 1.5 times the interquartile range.

(A) Gleason score and aneuploidy measured through copy number data in TCGA (compare to Fig. 3D).

(B) Gleason score and aneuploidy predicted from the transcriptome in TCGA.

(C) Gleason score and aneuploidy predicted from the transcriptome in HPFS and PHS. Two outliers with aneuploidy scores of 30 in HPFS/PHS (Gleason score, 3+4) were not plotted but were included in all analyses.

Figs. S5. Distribution of aneuploidy by *ERG* status.
Aneuploidy was defined using copy number alterations (from –2, indicating homozygous deletion, to +2, indicating amplification [or two-copy gain]), and gene-level DNA copy number alterations were summed for each tumor and chromosome arm from The Cancer Genome Atlas (TCGA). Plotted are distributions of these sums for each chromosome arm and tumor for tumors carrying the *TMPRSS2:ERG* fusion (A) or lacking it (B). The labels indicate the number of genes per chromosome arm. Copy number sums that were more extreme than the number of genes were defined as chromosome arm gains (yellow) and losses (blue). Compare to Fig. 2A.

Fig. S6. Number of altered chromosome arms, by *ERG* status in (A) TCGA (A) and (B) HPFS and PHS.

Fig. S7. Discrimination analysis for predicting aneuploidy from the tumor transcriptome with quantile cut-offs defined separately by *TMPRSS2:ERG* fusion status.
The outcome is aneuploidy (five or more altered chromosome arms). Predictors are the transcriptome quantile cut-offs per chromosome arm defined irrespective of *TMPRSS2:ERG* fusion status (gray line) and the transcriptome with quantile cut-offs per chromosome arm define separately for fusion-positive and fusion-negative tumors (blue line). A larger area under the curve (AUC) indicates better performance. Compare to Fig. 3B.

Fig. S8. Potential consequences of aneuploidy.
(A) Correlation between the number of aneuploid chromosome arms and mRNA levels of *MKI67* in prostate cancers from TCGA.
(B) Correlation between the number of aneuploid chromosome arms and apoptosis as assessed by TUNEL staining in HPFS and PHS.

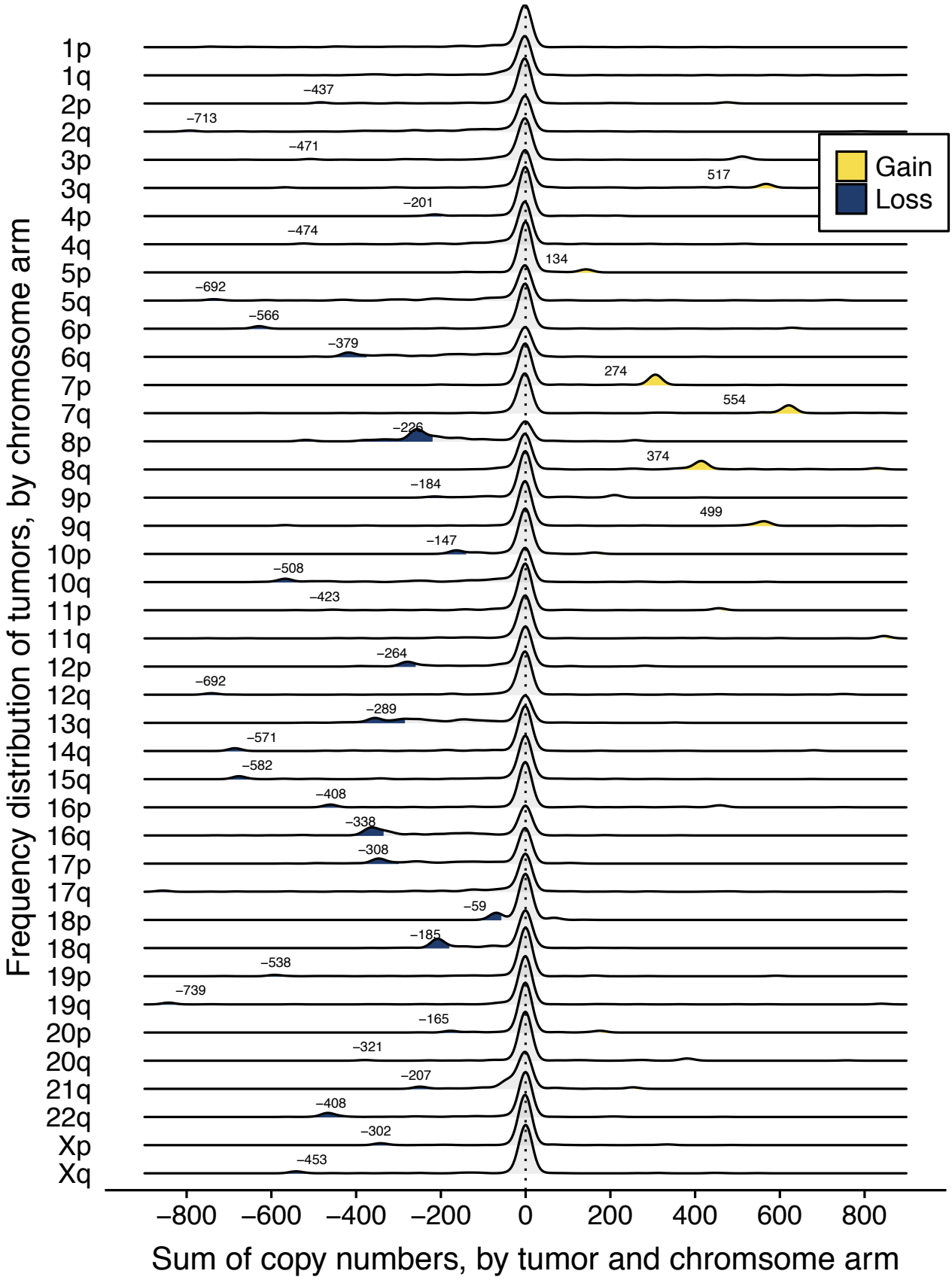
Supplementary Tables

Table S1. Characteristics of patients with primary prostate cancer from The Cancer Genome Atlas, by number of altered chromosome arms as derived from copy number data.

Table S2. Difference in median gene expression of chromosome arms 8q and 8p and risk of lethal prostate cancer over long-term follow-up in the Health Professionals Follow-up Study (HPFS) and the Physicians' Health Study (PHS).

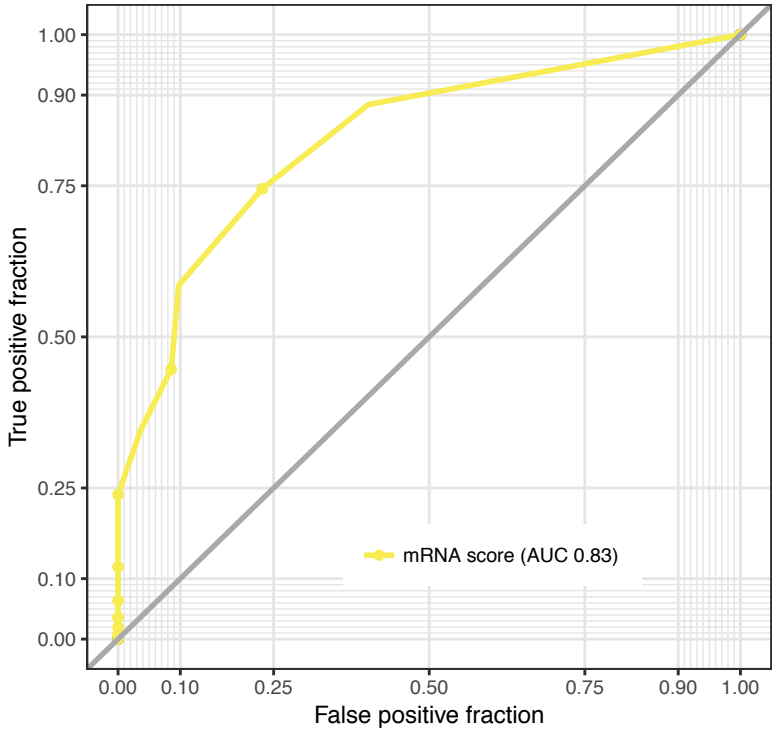
Table S3. A comparison of predicted losses or gains of entire chromosome arms and losses or gains of single tumor suppressors or oncogenes on those arms.

Supplementary Figure 1

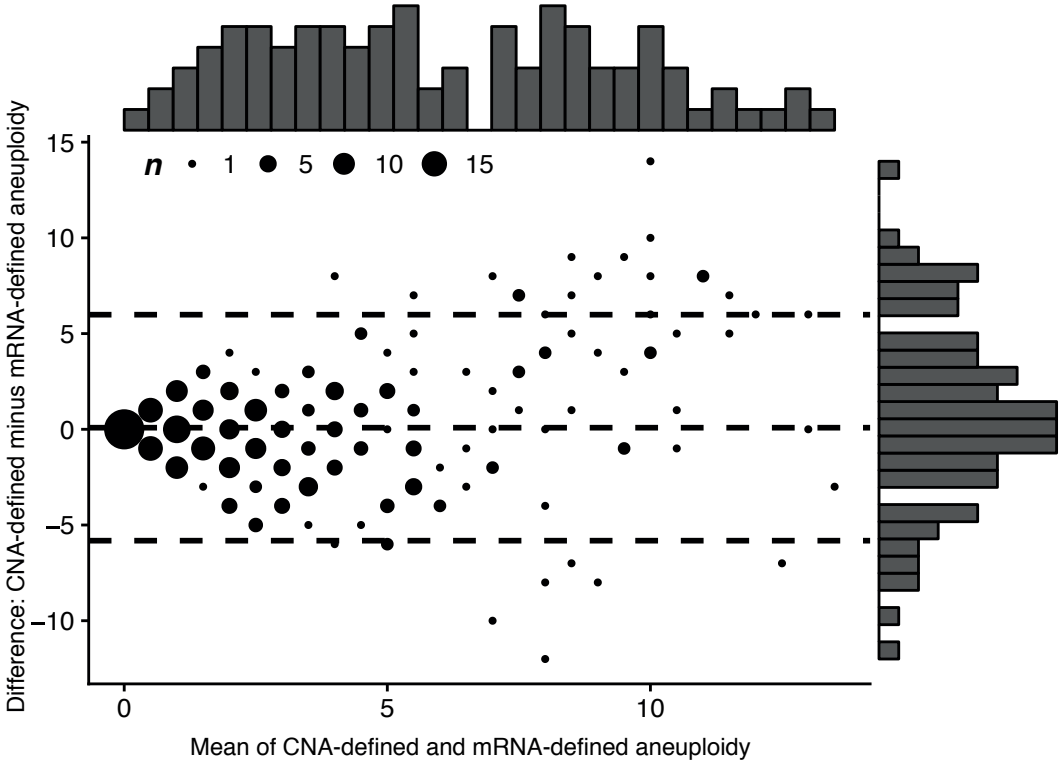


Supplementary Figure 2

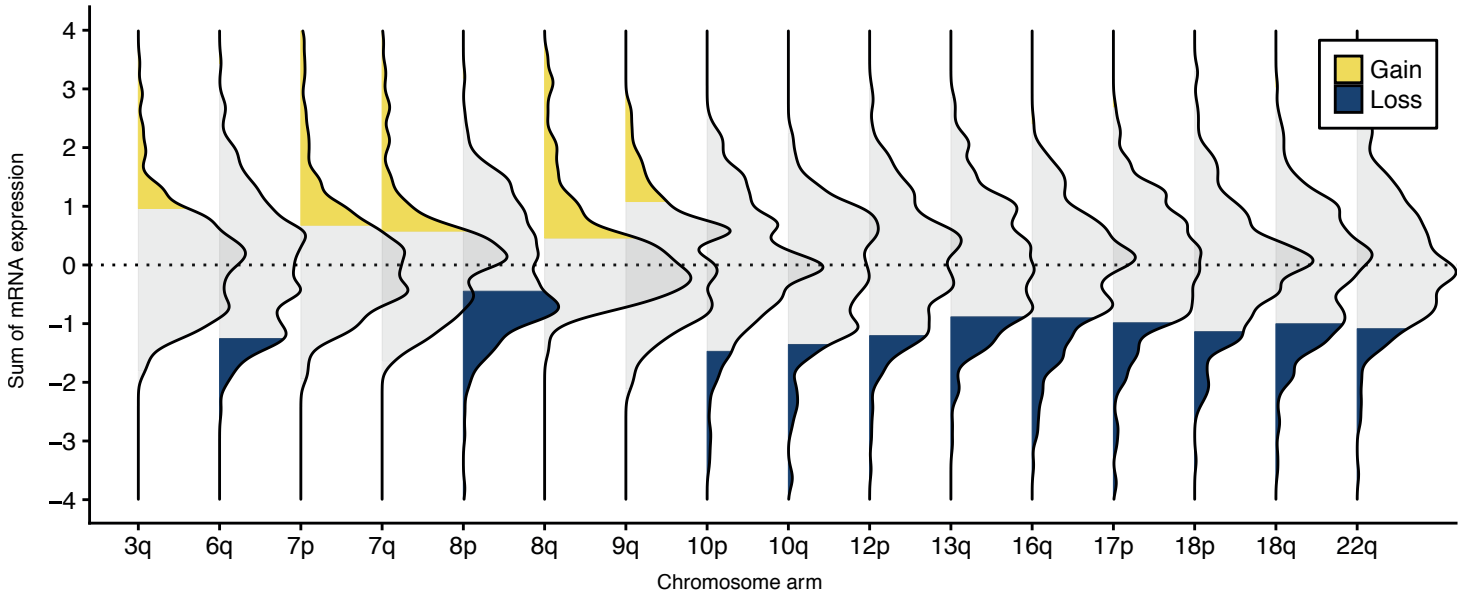
A



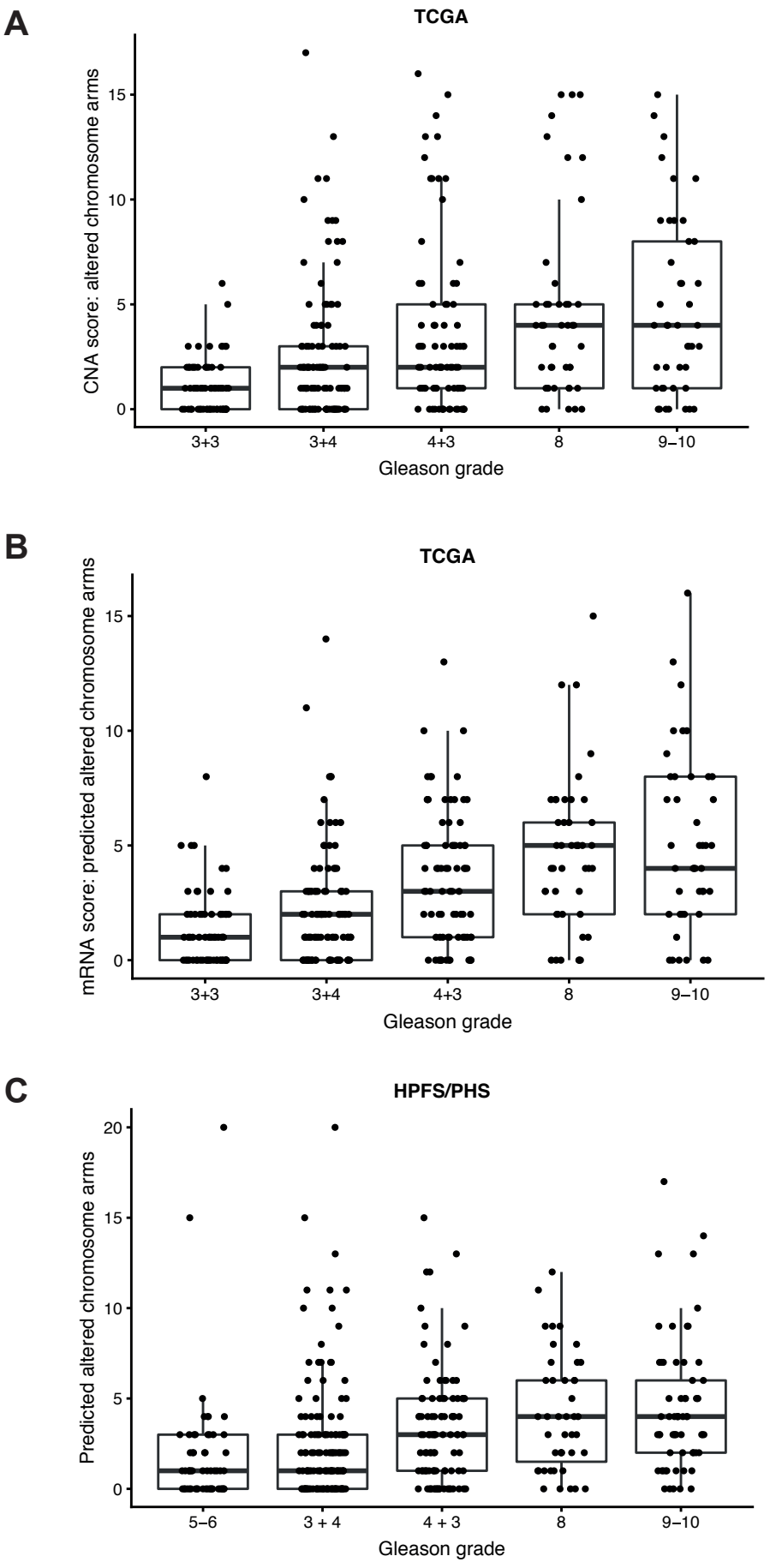
B



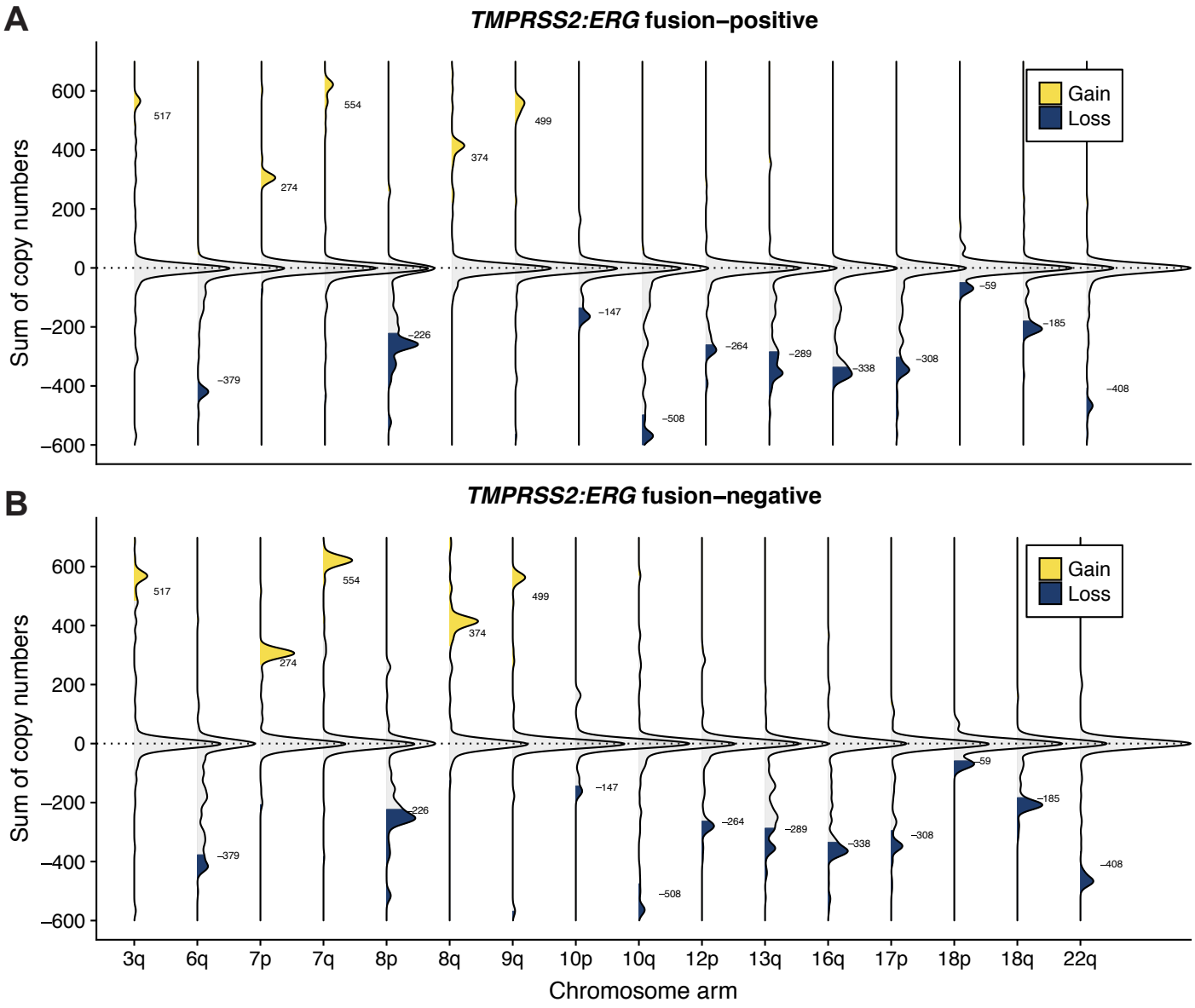
Supplementary Figure 3



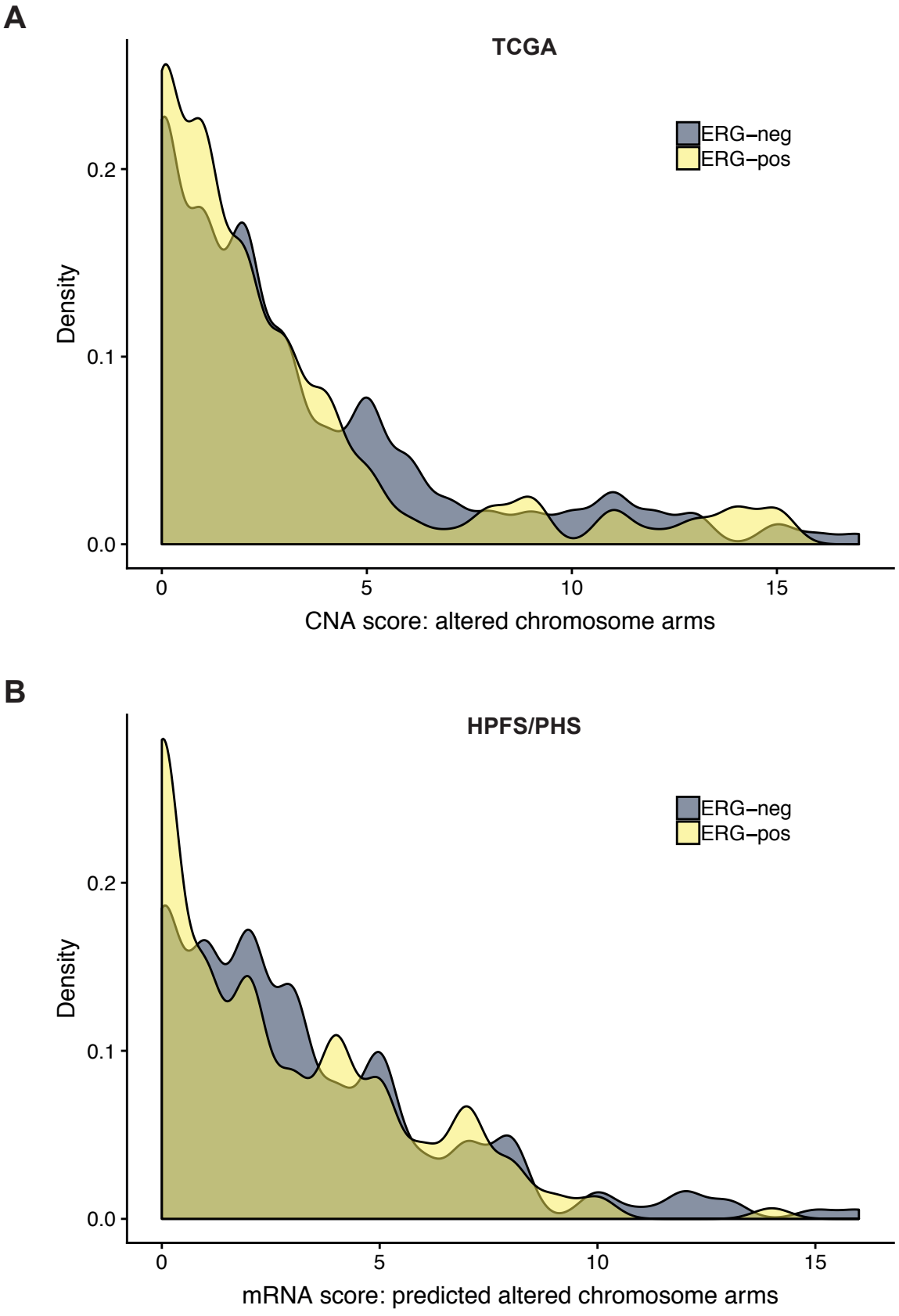
Supplementary Figure 4



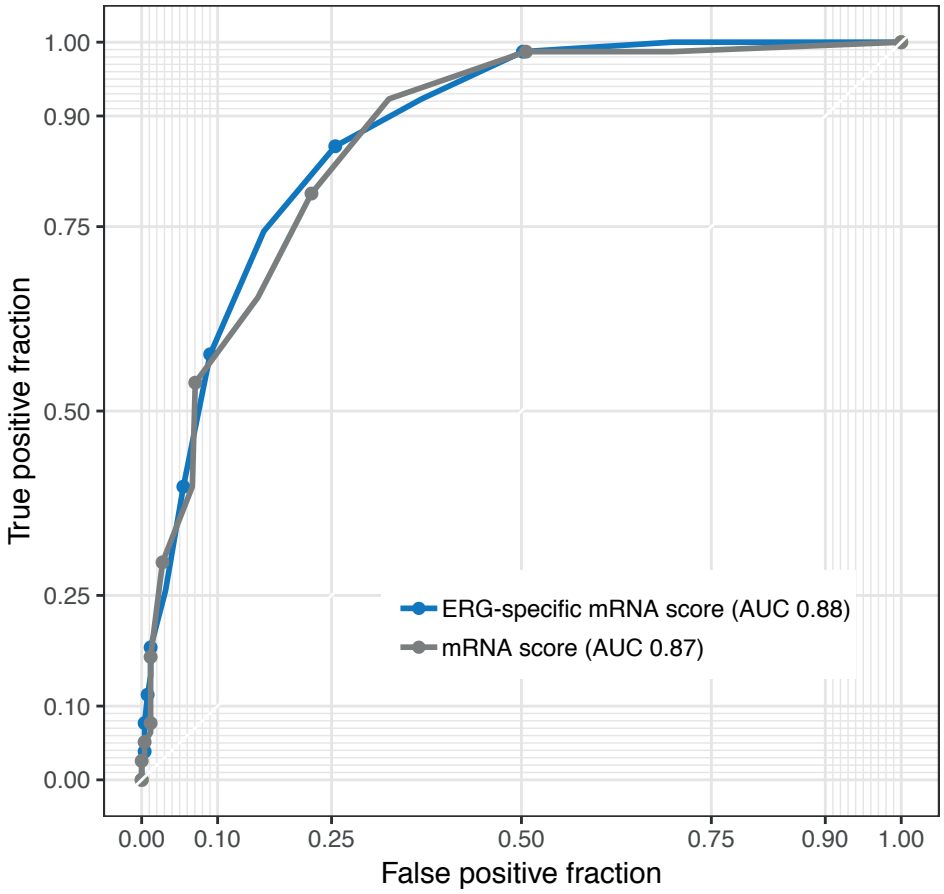
Supplementary Figure 5



Supplementary Figure 6



Supplementary Figure 7



Supplementary Figure 8

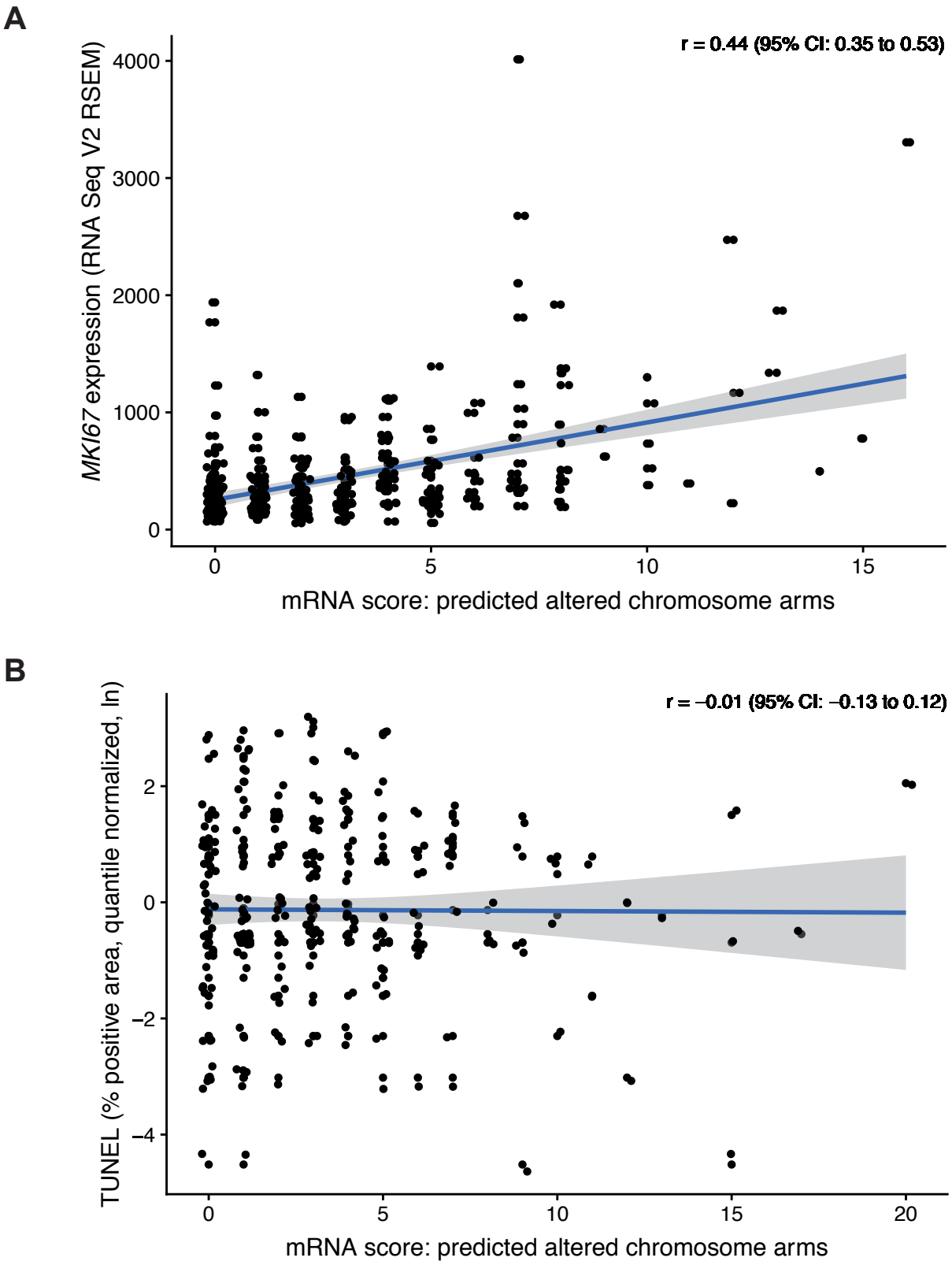


Table S1. Characteristics of patients with primary prostate cancer from The Cancer Genome Atlas, by number of altered chromosome arms as derived from copy number data.

Number of altered chromosome arms	*0*	*1–2*	*3–4*	*≥5*
Patients, *n*	82	116	57	78
Aneuploidy score based on copy number data, median (interquartile range)	0 (0–0)	1 (1–2)	3 (3–4)	8 (5–11)
Aneuploidy score predicted from mRNA, median (interquartile range)	0 (0–1)	2 (1–3)	3 (2–5)	6 (4–8)
Age at diagnosis [years], median (range)	60 (47–73)	60 (43–74)	64 (44–72)	62 (44–76)
Missing, *n* (%)	11 (13)	14 (12)	11 (19)	18 (23)
Race, *n* (%)				
White	66 (86)	89 (81)	49 (88)	66 (85)
African-American	10 (13)	17 (15)	6 (11)	10 (13)
Asian	1 (1)	4 (4)	1 (2)	2 (3)
Missing	5	6	1	0
Gleason score, n (%)				
6	27 (33)	29 (25)	7 (12)	2 (3)
3+4	27 (33)	39 (34)	16 (28)	20 (26)
4+3	16 (20)	26 (22)	15 (26)	21 (27)
8	5 (6)	13 (11)	10 (18)	17 (22)
9–10	7 (9)	9 (8)	9 (16)	18 (23)
PSA at diagnosis [ng/ml], n (%)				
<4	5 (10)	10 (15)	2 (6)	4 (11)
4–10	35 (71)	40 (59)	14 (41)	18 (50)
10–20	7 (14)	14 (21)	11 (32)	4 (11)
>20	2 (4)	4 (6)	7 (21)	10 (28)
Missing	33	48	23	42
TMPRSS2:ERG status, n (%)				
ERG fusion-negative	42 (51)	60 (52)	29 (51)	50 (64)
ERG fusion-positive	40 (49)	56 (48)	28 (49)	28 (36)

Abbreviations: PSA, prostate-specific antigen.

Table S2. Difference in median gene expression of chromosome arms 8q and 8p and risk of lethal prostate cancer over long-term follow-up in the Health Professionals Follow-up Study (HPFS) and the Physicians' Health Study (PHS).

8q-8p difference (quartiles)	*1st (lowest)*	*2nd*	*3rd*	*4th (highest)*
Lethal : non-lethal cases, *n*	15 : 86	22 : 79	27 : 74	49 : 52
OR (95% CI), unadjusted	1 (ref.)	1.60 (0.77–3.29)	2.09 (1.04–4.23)	5.40 (2.76–10.6)
OR (95% CI), adjusted for age, year, Gleason score	1 (ref.)	1.40 (0.59–3.30)	1.53 (0.66–3.55)	3.37 (1.50–7.55)
OR (95% CI), adjusted for aneuploidy score	1 (ref.)	1.29 (0.61–2.76)	1.47 (0.70–3.08)	2.78 (1.32–5.87)

Altered chromosome arms	*0*	*1–2*	*3–4*	*≥5*
Lethal : non-lethal cases, *n*	9 : 86	25 : 93	28 : 69	51 : 43
OR (95% CI), unadjusted	1 (ref.)	2.57 (1.14–5.81)	3.88 (1.72–8.76)	11.3 (5.11–25.2)
OR (95% CI), adjusted for 8q-8p difference	1 (ref.)	2.36 (1.04–5.39)	1.71 (1.16–6.40)	7.62 (3.29–17.7)

Abbreviations: CI, confidence interval; OR, odds ratio; ref., reference category.

Table S3. A comparison of predicted losses or gains of entire chromosome arms and losses or gains of single tumor suppressors or oncogenes on those arms.

The table shows how strongly arm-level alterations are associated with gene-level alterations; how arm-level alterations and gene-level alterations are associated with lethal disease individually (univariable models); and how arm-level and gene-level alterations are associated with lethal disease in a model that mutually adjusts for both factors (multivariable models). All measures of association are OR (95% CI).

	10q loss: PTEN loss[*]	*8q gain: high* MYC[†]	*8q gain: high* SQLE[‡]
Chromosome arm alteration and single gene alteration	3.97 (1.46–10.8)	1.62 (0.88–3.02)	3.13 (1.67–5.86)
Univariable models for lethal disease[§]			
Arm alteration	3.78 (1.39–10.3)	6.21 (3.78–10.2)	6.21 (3.78–10.2)
Single gene alteration	2.22 (1.19–4.12)	0.80 (0.44–1.44)	3.39 (1.81–6.34)
Multivariable model for lethal disease[§]			
Arm alteration	3.14 (1.13–8.75)	6.20 (3.74–10.3)	5.38 (3.23–8.98)
Single gene alteration	1.97 (1.04–3.73)	0.62 (0.32–1.18)	2.53 (1.29–4.96)

[*] Loss of PTEN protein expression compared to intact PTEN staining in at least one of the assessed cores per tumor.

[†] Highest quartile of *MYC* mRNA expression compared to lowest quartile. There was no association between 8q gain and extent of nuclear MYC staining (proportion of nuclei with positive staining). See reference (8).

[‡] Highest quartile of *SQLE* mRNA expression compared to lowest quartile. See reference (9).

[§] The outcome is lethal prostate cancer over long-term follow-up in the Health Professionals Follow-up Study (HPFS) and the Physicians' Health Study (PHS), as in Table 2.

References

1. Stark (Rider) JR, *et al.* (2009) Gleason score and lethal prostate cancer: does 3 + 4 = 4 + 3? *J. Clin. Oncol.* 27(21):3459-3464.
2. Cancer Genome Atlas Research Network (2015) The Molecular Taxonomy of Primary Prostate Cancer. *Cell* 163(4):1011-1025.
3. Penney KL, *et al.* (2015) Association of prostate cancer risk variants with gene expression in normal and tumor tissue. *Cancer Epidemiol. Biomarkers Prev.* 24(1):255-260.
4. Zu K, *et al.* (2014) Dietary lycopene, angiogenesis, and prostate cancer: a prospective study in the prostate-specific antigen era. *J. Natl. Cancer Inst.* 106(2):djt430.
5. Rider JR, *et al.* (2015) Tumor expression of adiponectin receptor 2 and lethal prostate cancer. *Carcinogenesis* 36(6):639-647.
6. Ahearn TU, *et al.* (2016) A Prospective Investigation of PTEN Loss and ERG Expression in Lethal Prostate Cancer. *J. Natl. Cancer Inst.* 108(2).
7. Pettersson A, *et al.* (2012) The TMPRSS2:ERG rearrangement, ERG expression, and prostate cancer outcomes: a cohort study and meta-analysis. *Cancer Epidemiol. Biomarkers Prev.* 21(9):1497-1509.
8. Pettersson A, *et al.* (2018) MYC Overexpression at the Protein and mRNA Level and Cancer Outcomes among Men Treated with Radical Prostatectomy for Prostate Cancer. *Cancer Epidemiol. Biomarkers Prev.* 27(2):201-207.
9. Stopsack KH, *et al.* (2016) Cholesterol Metabolism and Prostate Cancer Lethality. *Cancer Res.* 76(16):4785-4790.