All happy families are alike: Transcriptomic homogeneity in indolent prostate tumors is a useful prognostic biomarker

Gerke TA[a], Tyekucheva S[b,c], Creed JH[a], Penney KL[c,d], Sinnott JA[e], Ebot E[c], Berglund AE[a], Loda M[b], Stampfer MJ[c,d], Kraft P[c], Parmigiani G[b,c], Mucci LA[c,d]

[a] Moffitt Cancer Center, Tampa, FL, [b] Dana-Farber Cancer Institute, Boston, MA, [c] Harvard T.H. Chan School of Public Health, Boston, MA, [d] Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, [e] The Ohio State University, Columbus, OH

Overview

Biomarkers are needed to complement current clinicopathologic variables towards distinguishing prostate tumors that are likely to be metastatic from those that are indolent. In the present study, we develop a measure of transcriptomic heterogeneity using gene expression profiles of noncancerous prostate tissue from GTEx. We then show that the measure is strongly associated with Gleason score and is an independent predictor of lethal disease in TCGA and two large prospective cohorts of men with prostate cancer.

Introduction and Objectives

Previous evidence suggests that gene expression aberrations are homogeneous in non-aggressive tumors, while transcriptomic profiles are disregulated in widely varying patterns across metastasis-prone tumors. The present study develops a measure of transciptomic heterogeneity and investigates whether the score has prognostic utility.

Participants and Data

- RNA-seq expression profiles from noncancerous prostate tissue were obtained from GTEx V7 release ($n = 152$)
- RNA-seq profiles and Gleason grades from prostate tumor tissue were obtained from TCGA ($n = 333$)

- Nested case-control study of men with prostate cancer from the Health Professionals Follow-Up Study (HPFS; $n = 254$) and Physicians' Health Study (PHS; $n = 150$)
- Diagnoses between 1982–2005; median follow-up of 14 years
- 113 men with metastatic or lethal prostate cancer and 291 men with indolent disease (8 or more years of metastasis-free follow-up after diagnosis)
- Archival formalin-fixed, paraffin-embedded tumor tissue acquired at prostatectomy (92%) or TURP (8%)
- Whole-transcriptome gene expression quantified using the Human Gene 1.0 ST Array (Affymetrix)

Methods

Develop a panel of 250 stably expressed genes in GTEx:

- Filter genes not expressed or expressed at low levels in prostate
- Place genes into decile bins on the basis of median expression across samples
- Select the 25 lowest variance genes within each bin

Calculate heterogeneity scores in HPFS/PHS and TCGA:

- For each patient, calculate standard deviations of the 25 genes in each decile bin
- Compute score as the sum of the 10 standard deviations
- Associations between the score and lethal disease were assessed through logistic regression and AUC analyses

Results

- Transcriptomic heterogeneity associated with higher Gleason in TCGA ($r^2 = 0.33$, $p < 0.001$) and HPFS/PHS (Table 1)
- Heterogeneity score was a strong prognostic indicator for lethal disease in crude (Table 2) and adjusted (Table 3) models
- Heterogeneity score improved AUC of the clinical factor model from 0.84 (95% CI: 0.80-0.88) to 0.85 (95% CI: 0.81-0.89)

	Heterogeneous	Homogeneous	p-value
Cohort			0.06
HPFS – n (%)	136 (54)	118 (46)	
PHS – n (%)	66 (44)	84 (56)	
Age at diagnosis – median (Q1-Q3)	66.0 (61.0-69.0)	66.8 (62.0-70.0)	0.18
Pathologic Gleason score – n(%)			< 0.001
2–6	25 (44)	32 (56)	
3+4	46 (33)	93 (67)	
4+3	61 (60)	41 (40)	
8	27 (63)	16 (37)	
9–10	43 (68)	20 (32)	
Clinical TNM stage – n(%)			0.18
T1/T2 N0/Nx M0/Mx	167 (48)	183 (52)	
T3 N0/Nx M0/Mx	16 (59)	11 (41)	
T4/N1/M1	8 (65)	7 (35)	
PSA ng/mL – n(%)			0.16
0–4	16 (39)	25 (61)	
4–10	98 (50)	98 (50)	
10–20	25 (40)	38 (60)	
>20	24 (59)	17 (41)	
Case status – n(%)			< 0.001
Lethal	81 (72)	32 (28)	
Indolent	121 (42)	170 (58)	

Table 1: Clinical characteristics of HPFS and PHS cohorts stratified by median transcriptomic heterogeneity score

Variable	OR (95% CI)	p-value	AUC (95% CI)
High heterogeneity	3.56 (2.24–5.76)	< 0.001	0.69 (0.64–0.75)

Table 2: Univariate analysis in HPFS/PHS. Heterogeneity score split at median for OR and modeled continuously for AUC.

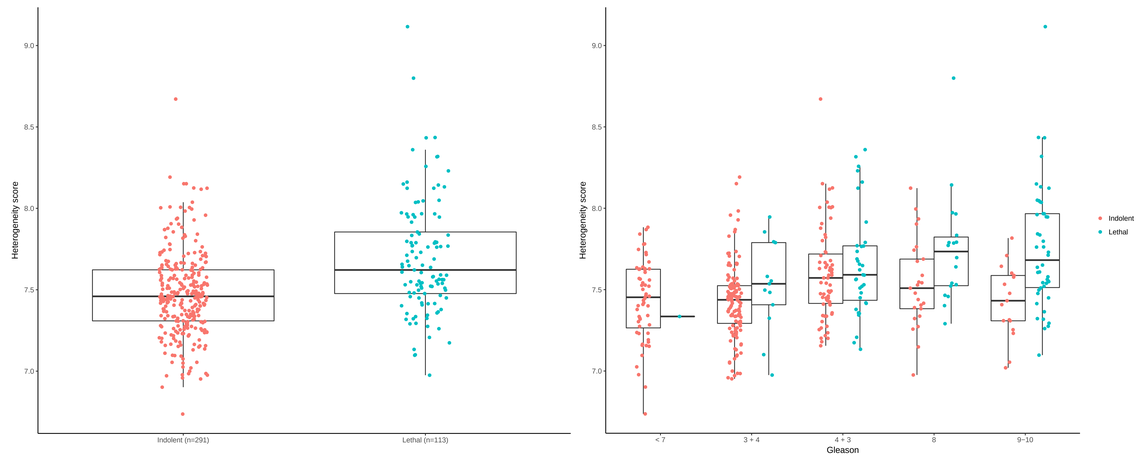

Figure 1: Heterogeneity score is higher in lethal tumors compared to those with an indolent course in HPFS/PHS, both overall and within Gleason categories.

Variable	OR (95% CI)	p-value	AUC (95% CI)
High heterogeneity	2.75 (1.56–4.96)	< 0.001	0.85 (0.81–0.89)
Age (continuous)	1.04 (0.99–1.09)	0.09	
RP Gleason			
2–6	1.00 (ref)	—	
3+4	6.43 (1.22–119)	0.08	
4+3	20.4 (4.09–371)	0.004	
8	26.7 (4.88–500)	0.002	
9–10	87.2 (16.5–1618)	< 0.001	
Clinical stage			
T1/T2	1.00 (ref)	—	
T3	1.67 (0.65–4.15)	0.28	
T4	23.8 (4.44–443)	0.003	

Table 3: Multivariable analysis in HPFS/PHS. Heterogeneity score split at median for OR and modeled continuously for AUC.

Conclusions

- Transcriptomic heterogeneity among genes that are typically stable in noncancerous prostate tissue is prognostic for lethal prostate cancer, and our novel measure of such heterogeneity may improve existing clinical risk models.
- Future work will evaluate whether the heterogeneity score may be correlated/integrated with other molecular biomarkers to enhance prognostic and mechanistic understanding; for example, *MYC Targets* is the most highly overlapping MSigDB Hallmark gene set with our 250-gene panel (FDR $q = 3.32e^{-10}$).
- Prognostic validation in other cohorts, particularly those with biopsy specimens, is warranted.

Funding

National Cancer Institute: 2P50CA090381-11A1, P01 CA055075, CA133891, CA141298, CA136578, UM1 CA167552, and T32 CA009001. We are grateful to the participants and staff of the Physicians' Health Study and Health Professionals Follow-Up Study for their valuable contributions.

MOFFITT CANCER CENTER

HARVARD T.H. CHAN SCHOOL OF PUBLIC HEALTH

THE OHIO STATE UNIVERSITY

DANA-FARBER CANCER INSTITUTE