1 **Non-invasive detection of clinically significant prostate cancer using circulating tumor cells**

2

3 Lei Xu[1,2][*], Xueying Mao[1][*], Alistair Grey[3,4,5], Glenda Scandura[1], Tianyu Guo[1], Edwina Burke[1],

4 Jacek Marzec[1], Semah Abdu[1], Elzbieta Stankiewicz[1], Caitlin R Davies[1], Prabhakar Rajan[1,3,4,6],

5 Karen Tipples[3], John Hines[3,6], Pui Ying Chan[7], Diane Campbell[3], Karen Wilkinson[3,6], Sakunthala

6 Kudahetti[1], Jonathan Shamash[7], Tim Oliver[1], Daniel Berney[1], Greg Shaw[3,4,6], Yong-Jie Lu[1,8]

7

8 [1]Centre for Molecular Oncology, Barts Cancer Institute, Queen Mary University of London,

9 London, UK.

10 [2]Department of Urology, Zhongshan Hospital, Fudan University, Shanghai, China.

11 [3]Department of Urology, Barts Health NHS, London, UK.

12 [4]Division of Surgery and Interventional Sciences, University College London, UK

13 [5]Department of Surgery and Cancer, Imperial College London, UK

14 [6]Department of Urology, University College London NHS Foundation Trust, UK

15 [7]Department of Medical Oncology, Barts Health NHS, London, UK.

16 [8]First Affiliated Hospital & Academy of Medical Sciences, Zhengzhou University, Zhengzhou,

17 China

18 [*]These authors contributed equally to this work.

19

20 Corresponding Author: Yong-Jie Lu, Centre for Molecular Oncology, Barts Cancer Institute,

21 Queen Mary University of London, John Vane Science Centre, Charterhouse Square, London,

22 EC1M 6BQ, UK. Phone: +44 (0)20 7882 3597; Fax: +44 (0)20 7882 3884; Email:

23 y.j.lu@qmul.ac.uk

24

25 Word account: 2490

26 **Running title**: CTCs in prostate cancer diagnosis

27

28   **Abstract**

29   **Purpose:** PSA testing results in unnecessary biopsy and over-diagnosis with consequent over-

30   treatment. Tissue biopsy is an invasive procedure, associated with significant morbidity. More

31   accurate non- or minimum-invasive diagnostic approaches should be developed to avoid

32   unnecessary prostate biopsy and over-diagnosis. We investigated the potential of using

33   circulating tumor cell analysis in cancer diagnosis, particularly in predicting clinically

34   significant prostate cancer in pre-biopsy patients.

35   **Material and methods:** We enrolled 155 treatment naïve prostate cancer patients and 98

36   pre-biopsy patients for circulating tumor cell numeration. RNA was extracted from circulating

37   tumor cells from 184 patients for gene expression analysis. Kruskal-Wallis, Spearman's rank,

38   multivariate logistic regression and random forest were applied to assess the association of

39   circulating tumor cells with aggressive prostate cancer.

40   **Results:** In localized prostate cancer patients, 54% were scored as circulating tumor cell

41   positive, which was associated with higher Gleason score ($p=0.0003$), risk group ($p<0.0001$)

42   and clinically significant prostate cancer ($p<0.0001$). In pre-biopsy group, positive circulating

43   tumor cell score in combination with PSA predicted clinically significant prostate cancer with

44   AUC=0.869. A 12-gene panel prognostic for clinically significant prostate cancer was also

45   identified. Combining PSA level, circulating tumor cell-score and the 12-gene panel, AUC for

46   clinically significant prostate cancer prediction was 0.927 and in cases with multi-parametric

47   MRI data, adding these to multi-parametric MRI significantly increased the prediction

48   accuracy (AUC 0.936 vs 0.629).

49   **Conclusions:** Circulating tumor cell analysis has the potential to significantly improve patient

50   stratification by PSA and/or multi-parametric MRI for biopsy and treatment.

51

52 **Key words:** circulating tumor cell, gene expression, non-invasive diagnosis, clinically

53 significant prostate cancer, pre-biopsy triage

54

1. Introduction

PSA testing lacks specificity **in the detection of prostate cancer (PCa), the most common cancer in Western men**[1,2] **and over half** patients with elevated PSA levels do not have cancer on biopsy**, an invasive procedure** with significant risks of urinary retention, bleeding and infection. In addition more than50% of the patients diagnosed with early stage PCa will not die of the disease[3-6] **suggesting** PSA may lead to unnecessary biopsies, over-diagnosis, and overtreatment[7]. Histological grading by Gleason score (GS) from biopsy specimens is currently needed for risk stratification, allowing the offer of appropriate therapeutic options[7,8]. An accurate, non-invasive test for clinically significant PCa (csPCs) might provide a safer, more efficient means of diagnosis.

Multiparametric MRI (mpMRI) has recently shown value in the detection of csPCA, with specificities of 23-87% and sensitivies of 58-96% reported[9-11]. **The PROMIS trial of** 576 men demonstrated a 93% sensitivity and 41% specificity of mpMRI, compared to 48% and 96% respectively for untargeted transrectal biopsy[9] **suggesting 27% could avoid biopsy using mpMRI as triage but with an accurate pre-biopsy biomarker a further 50% might do so**.

Circulating tumor cell (CTC) analysis using CellSearch system has been approved by FDA for prognostics in advanced, metastatic PCa[12]. The study of CTCs in non-metastatic PCa has been predominantly in locally advanced disease[13-18]. Most studies used CellSearch, concluding that CTCs are rare in patients with non-metastatic PCa[13-15]. Recent studies using new CTC isolation systems demonstrate greater CTC capture efficiency than CellSearch in locally advanced PCa[16-18]. Most methods detect CTCs with epithelial cell features, missing CTCs undergoing epithelial-mesenchymal transition (EMT), an important process in metastasis development.

79 We demonstrated that the Parsortix system, which uses cell size and deformability to capture

80 CTCs, harvested different subtypes in greater numbers than CellSearch[19, 20]. Here we

81 investigate its efficiency in capturing CTCs from patients with localized PCa and in PCa

82 diagnosis and risk stratification.

83

84 **2. Materials and methods**

85 **2.1 Study patient cohorts**

86 155 treatment-naïve, histologically confirmed localized PCa patients and 98 biopsy-naïve

87 patients with concerning PSA levels and/or abnormal digital rectal examination were enrolled

88 at St Bartholomew's Hospital. MpMRI was performed before biopsy. Ultrasound guided

89 transrectal or transperineal biopsy was performed with targeted biopsy on **suspicious (Likert**

90 **3+) mpMRI lesions**. Two pre-biopsy patients had bone metastases demonstrated by bone

91 scintigraphy. Control samples were collected from 12 healthy volunteers.

92

93 Clinical data including age, PSA, radiological results, biopsy results and TNM stage were

94 collected (**Supplementary Table 1**). Patients were classified into low-, intermediate-, and

95 high-risk tumor following EAU guidelines[7] and favorable disease or csPCa were defined based

96 on previous publications[21, 22] shown in **Supplementary Table 2**. The primary outcome was

97 men diagnosed with PCa, including risk stratification into favorable/clinically significant

98 disease.

99

100 **2.2 Cell lines**

101 Three PCa cell lines, PC3, LNCaP and VCaP from ATCC were used with authentication by short

102 tandem repeat testing.

103

2.3 CTC isolation, enumeration and RNA extraction

7.5 mL of whole blood was used for CTC isolation and enumeration as described

previously[19,20]. Positive CTC-score was defined as any epithelial CTC (CK+/VIM-/CD45-), any

EMTing CTC (CK+/VIM+/CD45-), and/or >3 mesenchymal CTCs (CK-/VIM+/CD45-) based on

our previous analysis of 24 age-matched male healthy control samples and the confirmation

of the malignant nature of CTCs in PCa cases by fluorescence *in situ* hybridization analysis of

multiple genomic regions commonly altered in PCa cells[20]. 97/155 PCa patients and 87/98

pre-biopsy patients had an extra 7.5 mL blood for CTC mRNA analysis **harvested from cassette**.

Total RNA was extracted using miRNeasy micro kit (Qiagen) following manufacturer's

instructions but eluted with a final volume of 11.5 μL. cDNA synthesis was performed using

SuperScript™ II Reverse Transcriptase (ThermoFisher Scientific).

115

2.4 Quantitative RT-PCR (qRT-PCR) for analytical validation

Gene expression was determined either using ABI 7500 Real-Time PCR system (Life

technologies) or Fluidigm multiplex PCR.

119

2.5 Statistical analyses

Kruskal-Wallis test was applied to assess the equality of CTCs between subgroups based on

CTC-score and different clinical features, such as mpMRI data, primary GS, and risk

classification. Data was shown as median (interquartile range [IQR]). Spearman's rank

correlation was used to assess the association between CTC counts and concurrent PSA level.

Receiver operating characteristic (ROC) curve analysis was performed to test the ability of

MRI, PSA, CTCs and different combined risk scores (CRSs) to predict patients with PCa and

127 csPCa. Regression coefficients for individual variables in CRSs were computed by multivariate

128 logistic regression. Optimal cut-off point was calculated to provide best available sensitivity

129 and specificity. Random forest classification algorithm[23] was applied to **rank prediction**

130 **abilities of** CTC expression genes and the final gene set selection was conducted by

131 comparing out-of-bag error rates of random forest models composed of decreasing number

132 of genes. Bonferroni correction method was applied to adjust p values (p_{adj}) for multiple

133 testing. Statistical analyses were performed using Stata 13.0 and R3.3.1.

134

135 **3. Results**

136 **3.1 Detection of CTCs in patients with localized PCa and their correlations with risk groups**

137 We first investigated the ability of CTCs, analyzed in three categories: epithelial (CK+/VIM-

138 /CD45-), EMTing (CK+/VIM+/CD45-) and mesenchymal (CK-/VIM+/CD45-) CTCs (**Fig. 1A**) for a

139 CTC score, in distinguishing clinically insignificant and significant cancers in diagnosed

140 localized PCa patients. In 155 patients with localized PCa, at least one traditional epithelial

141 CTC (all CK+ CTCs) were detected in 30% (46/155) of patients, at least one of any subtypes of

142 our defined CTCs in 78% (121/155) of patients and 54% (84/155) of patients were CTC-score

143 positive. In the 64 GS 3+3 and 40 low-risk cancer patients, CTCs were scored positive in 34%

144 (22/64) and 25% (10/40) of cases respectively, indicating that cancer cells are released into

145 the circulation at an early development stage.

146

147 Considering subtypes of CTCs, epithelial, EMTing and mesenchymal CTCs all showed trends of

148 correlations (Spearman's ϱ=0.15, 0.24 and 0.11, respectively) with serum PSA levels (*p*=0.07,

149 0.0029 and 0.17 respectively), although only EMTing CTCs are significant. Epithelial, EMTing

150 and mesenchymal CTC counts generally increased from low to high GS groups (3+3, 3+4, 4+3,

151   and ≥4+4) but without statistical significance (p_{adj}=0.16, 0.06 and 0.24 respectively,

152   **Supplementary Fig. 1**). Positive CTC-score was significantly associated with high GS

153   (p_{adj}=0.0012, **Table 1**).

154

155   If the 155 patients were divided into low-, intermediate- and high-risk groups, EMTing and

156   mesenchymal CTCs significantly increased with higher risk (p_{adj}=0.0136 and 0.016 respectively)

157   but not epithelial CTCs (p_{adj}=0.44, **Table 1, Fig. 1B**). CTC-score positivity associated more

158   significantly (p_{adj}<0.0001) with high-risk disease. Dividing into clinically significant and

159   favorable disease, high PSA level (p=0.0001), positive CTC-score (p_{adj}<0.0001), epithelial

160   (p_{adj}=0.0264), EMTing (p_{adj}=0.01) and mesenchymal (p_{adj}=0.0384) CTC counts were all

161   significantly correlated to csPCa (**Table 1, Fig. 1C**). Combining CTC-score with PSA, we

162   generated the combined risk score (CRS-PC) by 0.233xPSA + 1.548xCTC-score, which

163   discriminated csPCa better than PSA alone (AUC: 0.826 vs 0.764, ***p*=0.03, Fig. 2A**). In the 115

164   patients with mpMRI data at diagnosis, a significantly higher MRI positive (using Likert=3 as

165   threshold) rate was found in csPCa (P=0.0001) than favorable patients (**Table 1**). The AUC

166   using Likert 1-5 was 0.753 (95%CI: 0.663-0.842, with a cut-off point ≥3 to reach sensitivity of

167   98.59% and specificity of 47.73%, or a cut-off point ≥5 to reach sensitivity of 7.04% and

168   specificity of 100%).

169

170   **3.2 Predicting csPCa in pre-biopsy patients using serum PSA and CTC positivity**

171   We then assessed the potential of using CTCs to predict csPCa in 98 pre-biopsy patients.

172   Positive CTC-score was significantly associated with a positive biopsy results (p_{adj}<0.0001) and

173   csPCa (p_{adj}<0.0001, **Table 2**). Positive MRI (Likert≥3) had similar distribution in benign and

174   malignant patients (p=0.52), but was significantly more frequent in csPCa than in favorable

9

175 disease (*p*=0.0002, **Table 2**) and favourable combined with benign patients (*p*=0.0017, **Table**

176 **2**). The AUC to identify csPCa by PSA level was 0.733 (95%CI: 0.630-0.835, with an optimal

177 cut-off point ≥15 ng/mL to reach 44.19% sensitivity and 96.36% specificity), by CTC-score was

178 0.811 (95%CI: 0.732-0.890 with 76.74% sensitivity and 85.45% specificity) and by CRS-PC was

179 0.869 (95%CI: 0.792-0.945, with an optimal cut-off point ≥2.87 to reach 87.27% sensitivity

180 and 83.67% specificity)(**Fig. 2B**), using the model developed previously in localized patients,

181 significantly (*p*=0.0008) better than PSA alone. In the 87 pre-biopsy patients with pre-biopsy

182 MRI data, the AUC to predict csPCa using Likert 1-5 was 0.698 (95%CI: 0.588-0.808, with a

183 cut-off point ≥3 to reach sensitivity of 97.2% and specificity of 29.4%, or a cut-off point ≥5 to

184 reach sensitivity of 47.2% and specificity of 90.2%), PSA 0.739 and CTC-score 0.783 (**Fig. 2C**).

185 Various combinations of these three factors were produced; CRS-PM (combining PSA and MRI

186 Likert as 0.201xPSA + 0.550xMRI Likert), CRS-PC (combining PSA and CTC-score as 0.179xPSA

187 + 2.798xCTC-score), CRS-MC (combining MRI Likert and CTC-score as 0.593xMRI Likert +

188 2.528xCTC-score) and CRS-PMC (combining PSA, MRI Likert and CTC-score as 0.207xPSA +

189 2.477xCTC-score + 0.551xMRI Likert), in predicting csPCa. Each combination increased the

190 prediction value **(*p*<0.01 for all combinations including CTC score compared to PBS or MRI)**.

191 AUC for the combination of all three factors (CRS-PMC) reached 0.891 (**Fig. 2C**).

192

193 **3.3 Using CTC gene expressions to improve the prediction of csPCa**

194 **We identified** 74 PCa prognostic genes through extensive bioinformatic analysis of **all**

195 **available** transcriptome sequencing data and 50 reported PCa-specific and/or prognostic

196 genes by **literature search (Supplementary Table 3)**. 38 **of the 124 genes** were selected based

197 on their relatively high expression in prostate and low/zero expression in WBC/whole blood

198 using Genecards data (http://www.genecards.org/, **Supplementary Table 3**) for experimental

199 validation by qRT-PCR in PC3, LNCaP, VCaP and PBMC samples (**Supplementary Fig. 2**). Out of

200 the 38 genes, 30 with low expression in PBMC (minimum median Ct of 33.9, **Table 3,**

201 **Supplementary Fig. 2**) were finally selected for further analysis together with housekeeping

202 genes *GAPDH* and *MRFAP1*. Good qRT-PCR amplification efficiency was achieved both for the

203 ABI 7500 and Fluidigm systems using *FOLH1* (*PSMA*) assay in 1, 5 10, and 20 spiked LNCaP

204 samples (**Supplementary Fig. 3**). Minimum detectability of spiked cells for each gene using

205 the Fluidigm system were shown in **Supplementary Table 4**. All the 30 genes were negative

206 in PBMC controls. *CDH12, CHGA, CSMD3, GRHL2, KLK2,* and *PART1* were only positive in

207 cancer patients and csPCa cases were more frequently (17/108, 15.7%) with >6 gene positive

208 than the remaining patients (6/76, 8%)(*p*=0.049).

209

210 Using random forest classifier, we identified a 12-gene panel (**Table 3**, Supplementary **Fig. 4)**

211 to distinguish csPCa from favorable disease with an AUC of 0.707 (95%CI: 0.634-0.779, with

212 an optimal cut-off point ≥0.442, sensitivity 51.85%, specificity 80.26%). When we combined

213 PSA level, CTC-score and 12-gene panel score as CRS-PCG (0.200xPSA + 2.082xCTC-score +

214 1.035x12-gene panel score) the AUC increased from CRS-PC AUC=0.844 to 0.881 (95%CI:

215 0.832-0.929 with an optimal cut-off point ≥3.154 to reach 83.33% sensitivity and 80.26%

216 specificity, **Fig. 2D**)(***p*=0.024**) in above 184 samples and it increased to 0.927 (95%CI: 0.870-

217 0.985, with an optimal cut-off point ≥3.095 to reach 87.5% sensitivity and 89.36% specificity)

218 from a CRS-PC of 0.899 in the 87 pre-biopsy patients with CTC gene expression data (**Fig.**

219 **2E)(*p*=0.23)**. In the 78 samples with both MRI results and RNA samples, adding PSA and CTC

220 data to mpMRI (valued as 1 if Likert≥3 and 0 otherwise) as CRS-PCGM=3.127xMRI likert +

221 0.276xPSA + 3.014xCTC-score + 1.174x12-gene panel) dramatically increased AUC from 0.629

222 to 0.936 (***p*<0.0001**)(**Fig. 2F**).

223

3.4 Clinical implications

Modelling CTC score use in the 98 pre-biopsy men, 85% of biopsies were avoided, but 23% of csPCas were missed, reflecting a high specificity but low sensitivity (**Table 4**). Combining PSA and CTC score increased biopsies avoided to 87% while missing 23% of csPCa (**Table 4**). With the additional 12-gene panel, 91% biopsies were avoided with 18% csPCa missed. mpMRI predicted csPCa at a high sensitivity (94% negative predictive value) but lower specificity compared to CTC-score, avoiding 27% vs 85% biopsies (**Table 4**). Adding PSA and CTC data to mpMRI (CRS-PMC), 89% biopsies were avoided with only 15% csPCa missed (**Table 4**). With an alternative cut-off point, CRS-PCGM could avoid 42% biopsies without missing csPCa, doubling that by MRI alone.

234

4. Discussion

The recent development of efficient CTC capture systems permits study of CTCs in non-metastatic PCa, but its value in PCa detection is yet to be evaluated[16-18]. Using a cell size and deformability-based CTC isolation system in a large cohort of localized PCa, we detected CTCs at a high frequency and in low GS and low-risk cancer patients. Most importantly, we showed that CTC analysis in combination with serum PSA can efficiently detect csPCa, potentially avoiding prostate biopsy, and bringing major benefits to the PCa diagnostics. Cancer can invade the blood circulation at early development stages, including cancer precursor conditions[24]. However, due to their rarity and challenges in capturing CTCs, their potential for cancer detection has only been explored in lung malignancy[12, 25, 26]. Our study further supports the application of CTCs to early cancer detection.

246

247 **High (>50%) negative biopsy rates in abnormal PSA (>4ng/ml) highlight its limitation as a**

248 **biopsy trigger**[9]. Additionally, many early-stage PCas are indolent, do not affect mortality[3]. A

249 non-invasive biomarker, which can be used to avoid unnecessary biopsies, over-diagnosis,

250 and over-treatment, would be a useful addition to the diagnostic pathway, allowing resources

251 to be focused on patients with csPCa[9]. mpMRI shows promise in triaging patients with

252 suspected PCa for prostate biopsy and play an increasing role[9-11]. Here, we show that CTCs

253 may efficiently predict biopsy results, particularly for csPCa, and improve csPCa prediction

254 value of mpMRI. Further study in large cohorts is warranted to establish the roles of CTCs in

255 csPCa prediction alongside mpMRI, to improve patient biopsy triage and cancer prognosis.

256

257 The prognostic value of cancer RNA expression has been demonstrated[27, 28] and AR-V7

258 expression in CTCs has been used to predict the response to androgen deprivation therapy.

259 Here we demonstrate that, in addition to CTC enumeration, CTC gene expression analysis may

260 provide further prognostic information and bypass the problem of tumor heterogeneity

261 which occurs when analysing prostate biopsy samples[29]. Future CTC analysis in combination

262 of both CTC enumeration and gene expression level may significantly increase the potential

263 of using CTCs for cancer diagnosis and prognosis.

264

265 Including mesenchymal CTCs, our study significantly increased the CTC positive cases in both

266 the localized PCa and pre-biopsy cohorts of cancer cases. Mesenchymal cancer cells show

267 invasive growth properties and may cause spread at early stage of cancer development[30]. In

268 our localized disease cohort, only EMTing and mesenchymal CTCs were significantly

269 associated with GS.

270

271   There are limitations to this study. Firstly, our CTC analysis may miss small CTCs. Secondly,

272   **The CTC gene expression panel is yet to be validated. Finally,** this is a single centre study,

273   which requires validation by independent research centres.

274

275   **5. Conclusion**

276   In a large series of localized PCa, we detected using our novel CTC analysis method, a high

277   CTC positive rate which was correlated with higher GS and aggressive cancer. Importantly,

278   positive CTC-score was associated with csPCa. In the pre-biopsy cohort, CTCs in combination

279   with PSA efficiently predict csPCa. A CTC 12-gene prognostic panel was also identified to

280   further increase the prediction accuracy of csPCa, which can be used to improve mpMRI

281   prediction value. Therefore, we demonstrate the value of CTCs in PCa detection and

282   prognostication.

283

288

295

References

1. Siegel R, Ma J, Zou Z, *et al.* Cancer statistics, 2014. CA Cancer J Clin 2014; **64**:9-29.

2. Society AC. Cancer Facts & Figures 2016. Atlanta: American Cancer Society: 2016.

3. Wolf AM, Wender RC, Etzioni RB, *et al.* American Cancer Society guideline for the early detection of prostate cancer: update 2010. CA Cancer J Clin; **60**:70-98.

4. Wilt TJ, Brawer MK, Jones KM, *et al.* Radical prostatectomy versus observation for localized prostate cancer. N Engl J Med 2012; **367**:203-13.

5. Stattin P, Holmberg E, Johansson JE, *et al.* Outcomes in localized prostate cancer: National Prostate Cancer Register of Sweden follow-up study. J Natl Cancer Inst 2010; **102**:950-8.

6. Draisma G, Etzioni R, Tsodikov A, *et al.* Lead time and overdiagnosis in prostate-specific antigen screening: importance of methods and context. J Natl Cancer Inst 2009; **101**:374-83.

7. Mottet N, Bellmunt J, Bolla M, *et al.* EAU-ESTRO-SIOG Guidelines on Prostate Cancer. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent. Eur Urol 2017; **71**:618-29.

8. Hurley P, Dhir A, Gao Y, *et al.* A Statewide Intervention Improves Appropriate Imaging in Localized Prostate Cancer. J Urol 2017; **197**:1222-8.

9. Ahmed HU, El-Shater Bosaily A, Brown LC, *et al.* Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study. Lancet 2017; **389**:815-22.

10. Futterer JJ, Briganti A, De Visschere P, *et al.* Can Clinically Significant Prostate Cancer Be Detected with Multiparametric Magnetic Resonance Imaging? A Systematic Review of the Literature. Eur Urol 2015; **68**:1045-53.

11. Pokorny MR, de Rooij M, Duncan E, *et al.* Prospective study of diagnostic accuracy comparing prostate cancer detection by transrectal ultrasound-guided biopsy versus

319　magnetic resonance (MR) imaging with subsequent MR-guided biopsy in men without

320　previous prostate biopsies. Eur Urol 2014; **66**:22-29.

321　12. Moon DH, Lindsay DP, Hong S*, et al.* Clinical indications for, and the future of, circulating

322　tumor cells. Adv Drug Deliv Rev 2018; **125**:143-50.

323　13. Davis JW, Nakanishi H, Kumar VS*, et al.* Circulating tumor cells in peripheral blood samples

324　from patients with increased serum prostate specific antigen: initial results in early prostate

325　cancer. J Urol 2008; **179**:2187-91.

326　14. Loh J, Jovanovic L, Lehman M*, et al.* Circulating tumor cell detection in high-risk non-

327　metastatic prostate cancer. J Cancer Res Clin Oncol 2014; **140**:2157-62.

328　15. Khurana KK, Grane R, Borden EC*, et al.* Prevalence of circulating tumor cells in localized

329　prostate cancer. Curr Urol 2013; **7**:65-69.

330　16. Kuske A, Gorges TM, Tennstedt P*, et al.* Improved detection of circulating tumor cells in

331　non-metastatic high-risk prostate cancer patients. Sci Rep 2016; **6**:39736.

332　17. Todenhofer T, Park ES, Duffy S*, et al.* Microfluidic enrichment of circulating tumor cells in

333　patients with clinically localized prostate cancer. Urol Oncol 2016; **34**:483.

334　18. Theil G, Fischer K, Weber E*, et al.* The Use of a New CellCollector to Isolate Circulating

335　Tumor Cells from the Blood of Patients with Different Stages of Prostate Cancer and Clinical

336　Outcomes - A Proof-of-Concept Study. Plos One 2016; **11**:e0158354.

337　19. Xu L, Mao X, Imrali A*, et al.* Optimization and Evaluation of a Novel Size Based Circulating

338　Tumor Cell Isolation System. PLoS One 2015; **10**:e0138032.

339　20. Xu L, Mao X, Guo T*, et al.* The Novel Association of Circulating Tumor Cells and Circulating

340　Megakaryocytes with Prostate Cancer Prognosis. Clin Cancer Res 2017; **23**:5112-22.

341 21. Zumsteg ZS, Spratt DE, Pei I, *et al.* A new risk classification system for therapeutic decision

342 making with intermediate-risk prostate cancer patients undergoing dose-escalated external-

343 beam radiation therapy. Eur Urol 2013; **64**:895-902.

344 22. Zumsteg ZS, Zelefsky MJ. Short-term androgen deprivation therapy for patients with

345 intermediate-risk prostate cancer undergoing dose-escalated radiotherapy: the standard of

346 care? Lancet Oncol 2012; **13**:e259-69.

347 23. Breiman L. Random Forests. Machine Learning. Kluwer Academic Publishers 2001; **45**:5-

348 32.

349 24. Kang Y, Pantel K. Tumor cell dissemination: emerging biological insights from animal

350 models and cancer patients. Cancer Cell 2013; **23**:573-81.

351 25. Fiorelli A, Accardo M, Carelli E, *et al.* Circulating Tumor Cells in Diagnosing Lung Cancer:

352 Clinical and Morphologic Analysis. Ann Thorac Surg 2015; **99**:1899-905.

353 26. Ilie M, Hofman V, Long-Mira E, *et al.* "Sentinel" circulating tumor cells allow early diagnosis

354 of lung cancer in patients with chronic obstructive pulmonary disease. PLoS One 2014;

355 **9**:e111597.

356 27. Cuzick J, Swanson GP, Fisher G, *et al.* Prognostic value of an RNA expression signature

357 derived from cell cycle proliferation genes in patients with prostate cancer: a retrospective

358 study. Lancet Oncol 2011; **12**:245-55.

359 28. Bostrom PJ, Bjartell AS, Catto JW, *et al.* Genomic Predictors of Outcome in Prostate Cancer.

360 Eur Urol 2015; **68**:1033-44.

361 29. Boyd LK, Mao X, Lu YJ. The complexity of prostate cancer: genomic alterations and

362 heterogeneity. Nat Rev Urol 2012; **9**:652-64.

363 30. Peng Z, Wang CX, Fang EH, *et al.* Role of epithelial-mesenchymal transition in gastric

364 cancer initiation and progression. World J Gastroenterol 2014; **20**:5403-10.

365

Table 1. Summary of CTC count in 155 treatment-naïve prostate cancer patients by risk groups

	Spearman's ϱ with PSA (p)	GS = 3+3 (n = 64)	GS = 3+4 (n = 51)	GS = 4+3 (n = 22)	GS ≥ 4+4 (n = 18)	p (p_adj[a])	Low-risk (n = 40)	Intermediate-risk (n = 78)	High-risk (n = 37)	p (p_adj[a])	FD (n = 63)	csPCa (n = 92)	p (p_adj[a])
Age, y						0.0078				0.0021			0.0076
Median (IQR)		61 (56-68)	66 (56-71)	63.5 (57-71)	69 (62-76.5)		60 (55-67)	65 (56-71)	68 (61.5-73)		61 (55-68)	66 (58-72)	
PSA, ng/mL						0.0001				0.0001			0.0001
Median (IQR)		6.7 (5-9.9)	8.3 (5.6-12)	12 (7.7-18.3)	18.8 (8.9-26)		5.5 (4.7-7.4)	9.1 (6.2-12)	17.6 (8.1-26)		6.5 (5-9.0)	10.5 (7.0-17.5)	
mpMRI, n (%)						0.0001				0.0001			0.0001
1,2		22 (34)	0 (0)	0 (0)	0 (0)		16 (40)	6 (8)	0 (0)		21 (33)	1 (1)	
3,4,5		21 (33)	43 (84)	17 (77)	12 (67)		10 (25)	60 (77)	23 (62)		23 (37)	70 (76)	
n/a		21 (33)	8 (16)	5 (23)	6 (33)		14 (35)	12 (15)	14 (38)		19 (30)	21 (23)	
Epithelial CTC	0.15 (0.07)					0.0425 (0.17)				0.11 (0.44)			0.0066 (0.0264)
Median (IQR) (% detected)		0 (0-0) (14%)	0 (0-1) (29.4%)	0 (0-0.5) (23%)	0 (0-2) (44%)		0 (0-0) (15%)	0 (0-0) (23%)	0 (0 -1) (35%)		0 (0-0) (13%)	0 (0-1) (32%)	
EMTing CTC	0.24 (0.0029)					0.0155 (0.06)				0.0034 (0.0136)			0.0025 (0.01)
Median (IQR) (%)		0 (0-0) (6%)	0 (0-0) (12%)	0 (0-0) (18%)	0 (0-1.3) (67%)		0 (0-0) (2.5%)	0 (0-0) (12%)	0 (0-1) (27%)		0 (0-0) (3%)	0 (0-0) (20%)	
mesenchymal CTC	0.11 (0.17)					0.0608 (0.25)				0.0040 (0.016)			0.0096 (0.0384)
Median (IQR) (%)		1 (0-2) (61%)	1 (0-4) (66%)	3 (0-10.5) (68%)	3 (0.75-7) (94%)		1 (0-2) (55%)	2 (0-4.3) (64%)	4 (0-7.5) (73%)		1 (0-2) (57%)	2 (0-5) (68%)	
CTC-score, n (%)						0.0003[b] (0.0012)				<0.0001[b] (<0.0001)			<0.0001[b] (<0.0001)
Negative		42 (66)	19 (37)	5 (23)	5 (28)		30 (75)	33 (42)	8 (22)		44 (70)	27 (29)	
Positive		22 (34)	32 (63)	17 (77)	13 (72)		10 (25)	45 (58)	29 (78)		19 (30)	65 (71)	

[a] p value adjusted for multiple testing using Bonferroni correction method; [b] Fisher's exact test.

PSA: prostate specific antigen; GS: Gleason score; n: number; FD: favorable disease; csPCa: clinically significant prostate cancer; IQR: interquartile range; mpMRI: Multi-Parametric magnetic resonance imaging; n/a: data not available; CTC: circulating tumor cell; EMTing: during epithelial-mesenchymal transition.

Table 2. Summary of CTC count in 98 pre-biopsy patients by biopsy results

	Benign biopsy (n = 33)	Malignant biopsy (n = 65)	p (p_adj[a])	FD (n = 22)	csPCa (n = 43)	p (p_adj[a])	FD+Benign biopsy (n=55)	csPCa (n=43)	p (p_adj[a])
Age, y			0.06			0.021			0.0097
Median (IQR)	65 (56-69)	65 (59.5-70)		63 (56-66)	68 (63-71)		63 (57-68)	68 (63-71)	
PSA, ng/mL			0.0173			0.0017			0.0001
Median (IQR)	6.5 (5.2-10.2)	9.3 (6.4-17)		7.2 (6.0-9.4)	11 (7.0-23)		7.2 (5.4-10)	11 (7.0-23)	
Abnormal PSA, n (%)			1.0[b]			0.0108[b]			0.0333
> 4 ng/mL	31 (94)	61(94)		18 (82)	43 (100)		49 (89)	43 (100)	
≤ 4 ng/mL	2 (6)	4 (6)		4 (18)	0 (0)		6 (11)	0 (0)	
mpMRI, n (%)			0.52			0.0002			0.0017
1,2	7 (21)	9 (14)		8 (36)	1 (2)		15 (27)	1 (2)	
3,4,5	25 (76)	46 (71)		11 (50)	35 (82)		36 (66)	35 (82)	
n/a	1 (3)	10 (15)		3 (14)	7 (16)		4 (7)	7 (16)	
Epithelial CTC			0.0146 (0.06)			0.0147 (0.06)			0.0002 (0.0008)
Median (IQR)(%)	0 (0-0)**(3%)**	0 (0-0)**(22%)**		0 (0-0)**(5%)**	0 (0-1)(30%)		0 (0-0)(4%)	0 (0-1)(30%)	
EMTing CTC			0.0181 (0.07)			0.0806 (0.32)			0.0019 (0.0076)
Median (IQR)(%)	0 (0-0)**(0%)**	0 (0-0)**(15%)**		0 (0-0)(5%)	0 (0-0)(21%)		0 (0-0)(2%)	0 (0-0)(21%)	
Mesenchymal CTC			0.0022 (0.0088)			0.0105 (0.042)			0.0001 (0.0004)
Median (IQR)(%)	0 (0-1.5)**(36%)**	2 (0-6)**(63%)**		0 (0-2.25)(45%)	3 (0-7)(72%)		0 (0-2)(40%)	3 (0-7)(72%)	
CTC-score, n (%)			<0.0001[b] (<0.0001)			<0.0001[b] (0.0002)			<0.0001[b] (<0.0001)
Negative	30 (91)	27 (41.5)		17 (77)	10 (23)		47 (85)	10 (23)	
Positive	3 (9)	38 (58.5)		5 (23)	33 (77)		8 (15)	33 (77)	

[a] p value adjusted for multiple test; [b] Fisher's exact test.

N: number; FD: favorable disease; csPCa: clinically significant prostate cancer; IQR: interquartile range; PSA: prostate specific antigen; mpMRI: Multi-Parametric magnetic resonance imaging; n/a: data not available; CTC: circulating tumor cell; EMTing: during epithelial-mesenchymal transition.

Table 3. Threshold cycle of candidate genes in PBMC and regression coefficients of genes in 12-gene panel

Genes in 12-gene panel	C_T in PBMC, median (range)	Regression coefficient in panel	Rest genes in test	Ct in PBMC, median (range)	Genes not included in test	Ct in PBMC, median (range)
AOX1	35.2 (34.9-36.1)	0.854	AR-V7	undetermined	**CPLX1**	33.7 (33.0-35.1)
ACOX2	34.8 (33.4-36.9)	-1.89	CDH12	Undetermined	**COL5A2**	33.0 (32.7-34.1)
EYA4	36.3 (34.5-undetermined)	1.25	CHGA	Undetermined	**ACTG2**	33.0 (32.4-33.9)
FAT1	34.8 (34.1-36.5)	0.265	CSMD3	undetermined	**WNT5A**	33.3 (32.6-36.0)
FOXA1	34.9 (32.8-36.2)	-0.389	CYP3A5	Undetermined	**FRMD6**	32.4 (32.3-33.1)
GRHL2	Undetermined	0.934	LCE2B	undetermined	**SYP**	32.3 (31.8-32.7)
HOXB13	36.1 (35.8-36.5)	-0.146	MSMB	Undetermined	**AR**	31.8 (29.9-32.6)
KLK2	35.8(35.4-37.2)	0.71	PART1	Undetermined	**CDH1**	31.5 (30.2-33.1)
MNX1	35.6 (34.6-37.1)	-7.8	ROBO2	undetermined		
FOLH1(PSMA)	36.6 (35.4-37.2)	0.078	TMPRSS2:ERG	undetermined		
RAB3B	34.5 (34.0-36.5)	0.693	KLK3 (PSA)	37.3 (37.2-37.9)		
SRD5A2	Undetermined	-16.708	TWIST2	36.3 (35.3-37.8)		
			SPOCK3	35.9 (35.0-36.4)		
			FAM107A	35.3(35.0-undetermined)		
			HSPB8	37.0 (35.6-undetermined)		
			PCDH18	34.7 (32.2-37.1)		
			PCA3	34.5 (32.7-35.0)		
			TBX3	33.9 (30.8-37.1)		

PBMC: peripheral blood mononuclear cell.
Genes in bold were those not selected due to relative lower Ct value.

Table 4. Clinical implications of CTC enumeration and gene expression in 98 pre-biopsy patients

	Benign biopsies avoided (n = 33) (%)	prostate cancers missed (n = 65) (%)	PPV	NPV	FD+benign biopsy diagnosis avoided (n = 55) (%)	csPCa missed (n = 43) (%)	PPV	NPV
PSA > 4 ng/mL	2 (6)	4 (6)	66%	33%	6 (11)	0 (0)	47%	100%
CTC-score (positive)	30 (91)	27 (42)	93%	53%	47 (85)	10 (23)	80%	82%
CRS-PC ≥ 2.87	29 (88)	29 (45)	90%	50%	48 (87)	10 (23)	83%	83%
	n = 28	n = 59			n = 47	n = 40		
CRS-PCG ≥ 3.154	25 (89)	25 (38)	92%	50%	43 (91)	7 (18)	89%	86%
CRS-PCG ≥ 1.072	-	-	-	-	11 (23)	0 (0)	53%	100%
	n = 32	n = 55			n = 47	n = 40		
MRI positive (likert ≥ 3)	7 (22)	9 (16)	65%	44%	15 (27)	1 (3)	49%	94%
	n = 28	n = 50			n = 45	n = 33		
CRS-PCGM ≥ 7.327	25 (89)	20 (40)	91%	56%	40 (89)	5 (15)	85%	89%
CRS-PCGM ≥ 4.582	-	-	-	-	19 (42)	0 (0)	56%	100%

N: number; PPV: positive predictive value; NPV: negative predictive value; FD: favorable disease; csPCa: clinically significant prostate cancer; CRS: combined risk score; CRS-PC: combining PSA and CTC-score; CRS-PCG: combining PSA, CTC-score, and 12-gene panel score; CRS-PCGM: combining MRI, PSA, CTC-score and 12-gene panel score.

Figure legend

Figure 1. Representative CTC images and the distribution of subtypes of CTCs in PCa patient groups with different progression risk. (A) Representative CTC images identified by immunofluorescence. (B) The distribution of epithelial, EMTing and mesenchymal CTCs in patient groups with low, intermedium and high progression risk PCa. (C) The distribution of epithelial, EMTing and mesenchymal CTCs in patient groups with favorable cancer and csPCa. In B and C, data are expressed as mean (middle horizontal bar) ± SD (top and bottoms). X-axis: Gleason score groups; Y-axis: CTC numbers in each patient.

Figure 2. AUCs of CTCs and in combinations with other parameters for csPCa prediction in treatment-naïve prostate cancer and pre-biopsy patients. AUCs in predicting csPCa in 155 localized PCa patients (A), 98 pre-biopsy patients (B), 87 patients with MRI, PSA and CTC-score data(C), 184 patients with CTC gene expression data(D), 87 pre-biopsy patients(E) and the 78 samples with MRI, PSA, CTC-score and gene expression data (F). CRS-PC: PSA combined with CTC; CRS-PM: PSA combined with MRI likert; CRS-MC: MRI likert combined with CTC-score; CRS-PMC: PSA combined with MRI likert and CTC-score; CRS-PCG: PSA combined with CTC count and 12-gene panel score; CRS-PCGM: MRI combined with PSA, CTC-score and 12-gene panel score.

Key of Definitions for Abbreviations

AUC   Area under the ROC curve

CRS   combined risk score

csPCa   clinically significant prostate cancer

CTC   circulating tumor cell

EMT   epithelial-mesenchymal transition

GS   Gleason score

IQR   interquartile range

mpMRI   Multi-Parametric MRI

PBMC   peripheral blood mononuclear cells

PCa   prostate cancer

QRT-PCR   quantitative RT-PCR

ROC   receiver operating characteristic