An improved prognostic model for patients with metastatic castration-resistant prostate cancer: results from an open-data, crowdsourced competition.

Authors:
Justin Guinney, PhD[1*], Tao Wang, PhD[2,3*], Teemu D Laajala, MSc[4,5*], Prostate Cancer Challenge DREAM Community, Kimberly Kanigel Winner, PhD[8], Chris Bare[1], Elias Chaibub Neto, PhD[1], Suleiman A Khan, PhD[5], Gopalacharyulu Peddinti, PhD[5], Antti Airola, PhD[6], Tapio Pahikkala, PhD[6], Tuomas Mirtti, MD[5,7], Thomas Yu[1], Brian M Bot, MS[1], Liji Shen, PhD[18], Kald Abdallah, MD[9], Thea Norman, PhD[1], Stephen Friend, MD[1], Gustavo Stolovitzky, PhD[10], Howard Soule, PhD[11], Christopher J Sweeney, MBBS[12], Charles J Ryan, MD[13], Howard I Scher, MD[14], Oliver Sartor, MD[15], Yang Xie, PhD[2,16,17+], Tero Aittokallio, PhD[4,5+], Fang Liz Zhou, MD[18+], James C Costello, PhD[8,19+]

Affiliations:
[1]Sage Bionetworks, Seattle, WA, USA
[2]Quantitative Biomedical Research Center, Department of Clinical Sciences, University of Texas Southwestern Medical Center, Dallas, TX, USA
[3]Center for the Genetics of Host Defense, University of Texas Southwestern Medical Center, Dallas, TX, USA
[4]Department of Mathematics and Statistics, University of Turku, Finland
[5]Institute for Molecular Medicine Finland, FIMM, University of Helsinki, Finland
[6]Department of Information Technology, University of Turku, Finland
[7]Department of Pathology (HUSLAB), Helsinki University Hospital, Finland
[8]Department of Pharmacology & Computational Biosciences Program, University of Colorado, Anschutz Medical Campus, Aurora, CO, USA
[9]AstraZeneca, Gaithersburg, MD, USA
[10]IBM T.J. Watson Research Center, IBM, Yorktown Heights, NY, USA
[11] Prostate Cancer Foundation, Santa Monica, CA, USA
[12]Department of Medical Oncology, Dana-Farber Cancer Institute and Brigham and Women's Hospital, Harvard Medical School, Boston, MA
[13]Genitourinary Medical Oncology Program, Division of Hematology & Oncology, University of California, San Francisco, CA, USA
[14]Genitourinary Oncology Services, Department of Medicine, Sidney Kimmel Center for Prostate and Urologic Cancers, Memorial Sloan-Kettering Cancer Center and Weill Cornell Medical College, New York, NY, USA
[15]Tulane Cancer Center, Tulane University, New Orleans, LA, USA
[16]The Simmons Comprehensive Cancer Center, University of Texas Southwestern Medical Center, Dallas, TX, USA
[17]Lyda Hill Department of Bioinformatics, University of Texas Southwestern Medical Center, Dallas, TX, USA
[18]Sanofi, Bridgewater, NJ, USA
[19]University of Colorado Comprehensive Cancer Center, University of Colorado, Anschutz Medical Campus, Aurora, CO, USA
*equal contribution
[+]corresponding authors

Contact Corresponding Author:
James C Costello, PhD
University of Colorado Anschutz Medical Campus
Mail Stop 8303
12801 E. 17th Ave, Rm L18-6114
Aurora, CO 80045
(303)724-8619
james.costello@ucdenver.edu

Summary

Background

Improved prognostic models in metastatic castration-resistant prostate cancer (mCRPC) have the potential to improve clinical trial design and to guide treatment strategies. In partnership with *Project Data Sphere, LLC* (PDS), a not-for-profit initiative allowing cancer clinical trial data to be shared broadly with researchers, we designed an open-data, crowdsourced DREAM (Dialogue for Reverse Engineering Assessments and Methods) Challenge to both identify a better prognostic model and engage a community of international data scientists to study mCRPC.

Methods

Comparator arm data from four phase III clinical trials in first line mCRPC (n=2070) were obtained from PDS. Datasets comprising over 150 clinical variables were centrally curated, including demographics, lab values, medical history, lesion measures, and prior therapies. Three of the clinical trials were released publicly to be used as training data, and the fourth trial was used for independent model validation. The outcome variables in the validation dataset were hidden from the model builders representing an unbiased and rigorous evaluation.

Findings

A total of 50 independent prognostic models were developed and evaluated through the Challenge. The top-performer, based on an ensemble of penalized Cox regression models uniquely identified predictive interaction effects with immune biomarkers and markers of hepatic and renal function. Overall, it significantly outperformed all other models (iAUC=0·791) and surpassed a recently published benchmark (iAUC=0·743). The model was validated further on a fifth mCRPC placebo-only cohort with robust performance (iAUC=0·77). Meta-analysis across all models confirmed previously identified predictive clinical covariates and revealed aspartate aminotransferase as an important albeit previously under-reported prognostic biomarker.

Interpretation

The results of this effort demonstrate the enhanced value of data sharing when combined with a crowdsourced challenge concerning prognostic modeling in advanced prostate cancer patients. Novel prognostic factors were delineated, and improvement upon the prior state of the art accomplished.

Funding

In-kind contribution from Sanofi US Services Inc., Project Data Sphere, LLC

Introduction

Prostate cancer is the most common cancer among men in developed countries and ranks third in terms of mortality after lung cancer and colorectal cancer[1]. Among the over two million men diagnosed with prostate cancer in the US over the last ten years, roughly 10% presented with metastatic disease. For these men, the mainstay of treatment is androgen deprivation therapy, which results in a high rate of response. However, it is not durable, and nearly all tumors eventually progress to the lethal metastatic castration-resistant (mCRPC) state. Although significant improvements in outcome for men with mCRPC have been achieved from the recent drug approvals of next generation hormonal agents, an immunotherapy, a radiopharmaceutical, and a cytotoxic drug[2–10], how best to deploy them has not been determined. The elucidation of variables associated with patient outcomes independent of treatment will facilitate the design of future trials by homogenizing risk, thus enabling clinical trial questions to be answered more rapidly because smaller samples sizes will be required.

Prognostic models in mCRPC have been previously described utilizing baseline covariates from independent cohort studies[11–13]. A recently reported prognostic model for mCRPC[14] included eight clinical covariates - ECOG performance status, disease site, opioid analgesic use, lactate dehydrogenase, albumin, hemoglobin, PSA, and alkaline phosphatase. An open question is whether model generation could be improved through more systematic search using data-driven approaches, and at the same time provide insights into biological aspects of the disease that affect patient outcomes. An example is interaction effects between clinical variables that are underexplored in contemporary modeling, even though interactions between genetic variants are widely used and known to improve genetic-based risk prediction and prognostic stratification[15,16].

Here we present the results from the Prostate Cancer DREAM (Dialogue for Reverse Engineering Assessments and Methods) Challenge , the first open-data, crowdsourced competition in mCRPC. Critical to the effort was the removal of the privacy and legal barriers associated with open access to phase III clinical trial data[17] by *Project Data Sphere LLC* (PDS) - a not-for-profit initiative of the CEO Roundtable on Cancer's Life Sciences Consortium that broadly shares oncology clinical trial data with researchers. The Challenge was designed to accomplish two goals. The first was to leverage the open-data enabling a community-based approach to identify the best performing prognostic model in a rigorous and unbiased manner.

The second was to use the predictive models built by 50 independent, international teams to both validate previously characterized predictive clinical covariates and to discover new prognostic features. Consistent with the mission of DREAM, all Challenge data, results, and method descriptions from participating teams are publicly available at: https://www.synapse.org/ProstateCancerChallenge.

Methods

Trial selection, data curation and patient population

In April 2014, the Challenge organizing team reviewed all existing and incoming prostate cancer trial datasets (comparator arm only) in PDS and selected four trials with that enrolled 2070 patients as the source of training and validation datasets for the Challenge. All four trials were randomized phase III clinical trials in which the comparator arm consisted of chemotherapy-naïve patients with mCRPC receiving docetaxel regimen[18–21]. These patient-level trial datasets were de-identified by data providers and made available for the Challenge through PDS.

The original datasets from PDS contained patient level raw tables that conformed to either Study Data Tabulation Model (SDTM) standards or company-specific clinical database standards. In an effort to optimize the use of these data for the Challenge, the four sets of raw trial data were compiled into one set of standardized raw tables comprising over 150 clinically defined baseline covariates and outcome variables into a “Core Table”. In addition to the Core Table, five additional tables were made available as standardized raw longitudinal tables (lab, lesion, prior medicine, medical history, vital sign). Full details of the data curation process are contained in the Appendix pp 2-3.

Training datasets were from three trials: ASCENT2[18], a randomized, open-label study evaluating DN-101 in combination with docetaxel in mCRPC; VENICE[19], a randomized, double-blind study comparing efficacy and safety of aflibercept versus placebo in patients treated with docetaxel/prednisone for mCRPC; and MAINSAIL[20], a randomized, double-blind study to evaluate efficacy and safety of docetaxel and prednisone with or without lenalidomide in patients with mCRPC. In ASCENT2, 476 patients received docetaxel and calcitriol in comparator arm; in VENICE, 598 comparator arm patients received docetaxel, prednisone, and placebo; in MAINSAIL, 526 patients were treated with docetaxel, prednisone, and placebo in

comparator arm. The training datasets combining all three trials (n=1600) were made publicly available through the Challenge and were used by teams to build their models.

ENTHUSE 33[21] is a randomized, double-blind study to assess efficacy and safety of 10 mg ZD4054 combined with docetaxel in comparison with docetaxel in patients with mCRPC. The clinical variables derived from 470 men treated with docetaxel and placebo in comparator arm of ENTHUSE 33 were made public through the Challenge. The outcome variables were not released so as to serve as the independent validation dataset for model performance evaluation.

To further validate the top-performing prognostic model independently, data derived from a fifth trial dataset ENTHUSE M1[22], a randomized, double-blind study to assess efficacy and safety of 10 mg ZD4054 versus placebo in patients with mCRPC (specifically bone metastasis) who were pain free or mildly symptomatic was used. The patients in the comparator arm received only placebo treatment. Due to the regulation and privacy environment of certain countries, not all patients in the comparator arm from ENTHUSE 33 and M1 were provided to PDS.

Scoring of Challenge submissions

The Challenge was hosted and fully managed on Synapse (www.synapse.org), a cloud-based platform for collaborative scientific data analysis, through which all model predictions were submitted. Teams were scored and ranked using the ENTHUSE 33 trial. For method evaluation, we utilized the integrated AUC (iAUC)[23] calculated from 6-30 months as our primary scoring method. Robust determination of best performing team(s) was carried out using Bayes Factor (BF) analysis and randomization test based on iAUC (see Appendix p 4-5). In addition, we evaluated model predictions using the log-rank test after dichotomizing the patients for each team by the median risk score using the R coxph function. We also calculated other statistics such as median survival time, one and two year survival rate for the dichotomized high risk and low risk groups.

The reference prognostic model for predicting OS was developed by Halabi et al.[14]; hereafter referred to as the “reference model”. This model was used to compare all other models developed by teams in the Challenge. Reference model coefficients were obtained from hazard

ratios as reported in Table 2.[21] For each team, we calculated the BF to directly compare the performance of a model to the reference.

Top-performing method construction in the training datasets

Based on unsupervised explorative analyses (Appendix pp 5-6,10), two of the most representative datasets (MAINSAIL and VENICE) were utilized in the supervised model learning. The ensemble of models was based on penalized Cox proportional hazards (Eqn 1). The model estimation procedure identified an optimal penalization parameter (λ), which regularizes the number of non-zero coefficients in the model. Simultaneously, the L_1/L_2 norm parameter (α) was explored throughout the full model spectrum, ranging from Ridge Regression ($\alpha = 0$) through Elastic Net ($0 < \alpha < 1$) and LASSO ($\alpha = 1$) in penalized regression with objective function:

$$argmax_{\beta}[\tfrac{2}{n}\sum_{i=1}^{n}(x_{j(i)}^{T}\beta - log\,(\sum_{j\in R_i} e^{x_j^T\beta})) - \lambda(\alpha\sum_{i=1}^{p}|\beta_i| + \tfrac{1}{2}(1-\alpha)\sum_{i=1}^{p}\beta_i^2)]\ (Eqn.\ 1)$$

Here, x are the predictors (clinical covariates or their pairwise interactions), β are the model coefficients subjected to the L_1 and L_2 norm penalization, p is the number of covariates, n is the number of observations, $j(i)$ is the index of the observation event at time t_i, and R_i is the set of indices j with $y_j \geq t_i$ (patients at risk at time t_i). Individual ensemble model components were composed separately for the MAINSAIL and VENICE trials, as well as for a third combined ensemble component, which simultaneously modeled the two trial datasets. To reduce the risk of overfitting and to avoid randomness bias in the binning, the final ensemble models were optimized using 10-fold cross-validation iAUC as well as averaged over multiple cross-validation runs. Each ensemble component resulted in a different optimum in Eqn. 1, although the resulting Elastic Net models favored Ridge Regression. The final ensemble prediction was performed by averaging the ranks from the component-wise predicted risks for the ENTHUSE 33 dataset. Complete details of the method can be found in the Appendix pp 5-6.

Data and Method Availability

The clinical trial data used in the Challenge can be accessed at https://www.projectdatasphere.org/projectdatasphere/html/pcdc. Challenge documentation, including the detailed description of the Challenge design, overall results, scoring scripts, and

the clinical trials data dictionary can be found at: https://www.synapse.org/ProstateCancerChallenge.

Role of Funding Source

Project Data Sphere, LLC (PDS) is an independent, not-for-profit initiative of CEO Round-table on Cancer's Life Science Consortium. The PDS online platform http://www.projectdatasphere.org hosted the clinical trial datasets. Sanofi US Services Inc. provided in-kind contribution of human resources for curation the raw datasets for the Challenge and for clinical and scientific support of Challenge organization at the request of PDS. The corresponding authors had full access to the data in this study and had final responsibility for the decision to submit for publication.

Results

The baseline characteristics of the four clinical trials are shown in Table 1 and the overall Challenge design is shown in Figure 1 with full details of the Challenge in the Appendix pp 3-4,7-8,14. Over 150 clinical baseline variables and longitudinal variables were measured for laboratory values, lesions, prior medicines, medical history and vital signs. When combined, the clinical variables for each trial were highly similar (Appendix p 11), although when binary variables, primarily representing lesion sites, were considered separately, differences in clinical trial were observed (Appendix p 11). ASCENT2 had a much lower frequency of patients with visceral metastases (1·1% liver and 1·7% lung) compared to patients from the other three trials (10-14% liver, 11-15% lung). In contrast, the percentage of patients with bone metastases (72-100%) was high across the four trials. Median follow-up times differed among the four studies with 11·7 months, 21·1 months , 9·2 months, and 15·3 months in the ASCENT2, VENICE, MAINSAIL, and ENTHUSE 33 trials, respectively (Appendix p 12). Risk profiles for each of the trials, specifically the rate of death, were highly similar among the four trials (Appendix p 12) with proportionality of hazards ($p > 0{\cdot}5$).

Fifty international teams - comprising 163 individuals - submitted predictions from their models to the Challenge; a total of 51 when the reference model was included. The distribution of all team scores is depicted in on Appendix p 13 with the top scoring model reporting an iAUC of 0·791 and outscoring all other teams with a BF > 5 (Appendix p 13), surpassing the minimum threshold of BF > 3.

The top performing model was developed by a collaborative team from the Institute for Molecular Medicine Finland (FIMM) and the University of Turku (UTU). The method used was based on an ensemble of penalized Cox regression (ePCR) models (Appendix p 10). The ePCR model was trained using the MAINSAIL and VENICE trials. The ePCR model extended beyond the LASSO-based reference model by using an Elastic Net to select additional, correlated groups of clinical variables and their interactions, modeled as interaction terms. The risk predictions from the trial-specific ensemble components were rank-averaged to produce the final ensemble risk prediction vector that was robust to study-wise effects.

Using the held-out ENTHUSE 33 trial data as validation, the ePCR model achieved an iAUC of 0·791 and the reference model an iAUC of 0·743 (Figure 2A), a significant difference in scores (BF > 20; Appendix p 13). Using a time-dependent AUC metric, the ePCR model consistently outperformed the reference at each time point, and in particular at later time points between 18 to 30 months (Figure 2A). Kaplan Meier analysis using median split of the patients into low and high risk groups revealed that both models were highly significant, with the ePCR model outperforming (Figure 2B; p = 2·4e-14; HR=3·32 (2·39-4·62 CI)) the reference model (Figure 2C; p = 1·4e-9; HR=2·56 (1·85-3·53 CI)). A complete comparison is provided in Appendix p 9.

To help interpret the significant groups of variables and their predictive relationships identified in the ePCR model, we generated a network visualization based on the importance of the model covariates and their interactions (Figure 3). While many of the variables used in the reference model were included in the ePCR model, aspartate aminotransferase (AST) was identified as an important predictor. We also observed covariates that were included as interaction terms, and of particular note were those reflecting the immunological or renal function of the patient. PSA was an independent but weak prognostic factor that interacted strongly with LDH and AST. These results suggest a complex relationship and dependency structure among many of the covariates.

In addition to identifying the top performing model, the Challenge tested 50 independent models; 30 of which outperformed the reference with a BF > 3 (Appendix p 13). The results present a rich set of information that can be mined for additional insights into both patient stratification and the robustness of clinical predictive variables. Accordingly, we first

hierarchically clustered the risk scores from the 51 models to identify three distinct patient risk groups (Figure 4A). Differences in overall survival (OS) among these three groups were highly significant ($p < 1e\text{-}17$, log-rank test) with median OS times of 12·0, 18·3, and 27·8 months (Figure 4B). Next, we surveyed teams directly to identify common clinical covariates incorporated into their final models to which 40 of the 50 teams provided answers. Figure 4C summarizes the frequency that a covariate was reported as being important or significant in a team's model. The results both confirmed the covariates previously identified in reference, but also identified several that were not[14]. Of note, AST was included in over half of the team models. Other novel included total white blood cell count (WBC), absolute neutrophil count (NEU), body mass index (BMI) and creatinine (CREAT).

All mCRPC patients included in the four trials used in the Challenge were chemotherapy-naïve before treatment with docetaxel. We were interested in assessing whether the ePCR model could be used as a risk stratifier for mCRPC patients who received placebo alone and no docetaxel treatment. Through PDS, we obtained data from a fifth phase III mCRPC clinical trial, ENTHUSE M1[22] ($n = 266$) which, in contrast to the ENTHUSE 33 study, included a comparator arm that received placebo alone. The inclusion/exclusion criteria were similar except that patients were pain-free or mildly symptomatic (Table 2). This dataset was curated and normalized the same as the other Challenge datasets (Appendix pp 2-3).

We applied the ePCR and reference models to the ENTHUSE M1 dataset, and computed performance metrics consistent with prior analysis. Notably, we observed model performances comparable to the primary Challenge, with iAUCs of 0·77 and 0·73 for the ePCR and reference models, respectively (Figure 5A). Kaplan-Meier analysis of the ENTHUSE M1 data also showed significant separation of median-split high and low risk predicted patients ($p < 1e\text{-}8$) with median survival times of 15·8, and 27·1 months, respectively (Figure 5B). These results suggest that the ePCR model of OS may be generalized to first-line mCRPC patients undergoing various treatments as a risk stratifier, although additional datasets are needed for further confirmation.

Discussion

The Prostate Cancer DREAM Challenge resulted in a single prognostic model that significantly outperformed all other models, including a recent reference model by Halabi, et al.[14], and led to a network perspective of predictive biological variables and their interactions. The

top-performing team's approach pointed to important interaction effects with immune biomarkers and markers of hepatic (potentially reflected in the increased AST levels) and renal functions. Many of these observed interactions, while not significant as independent variables, may be important modulators of key clinical traits such as heme- and haematopoietic-related measurements (e.g. hemoglobin and hematocrit). While further investigation is necessary to explore the clinical implication of these relationships and provide new insights into tumor/host interaction, it sheds light on the complex and interwoven nature of prognostic factors on patient survival.

Open-data, crowdsourced scientific competitions have been highly effective at drawing large cross-disciplinary teams of experts to solve complex problems[24–29](http://dreamchallenges.org). Partnering with PDS and Sage Bionetworks, this DREAM Challenge represented the first public competition utilizing open-access registration trial datasets in cancer with the intent of improving outcome predictions. A total of 163 individuals comprising 50 teams participated in the Challenge, applying state-of-the-art machine learning and statistical modeling methods. The contribution of five clinical trial datasets from industry and academic institutions to PDS, and their subsequent use in an open Challenge, enabled the advancement of prognostic models in mCRPC that heretofore was not possible, including:

1. Modelers had access to several, independent clinical trial cohorts with subtle differences in eligibility that increased the diversity (heterogeneity) of the total patient population considered for model development.
2. The PDS data tables included over 150 independent and standardized variables compiled from the trials in contrast to only 22 considered in the reference model.
3. Creativity in data mining using the standardized raw longitudinal tables, which are rarely leveraged for prognostic model development, enabled innovative clinical covariates to be derived for modeling. These longitudinal tables were utilized by several teams including the top performing team.
4. The evaluation of 50 methods - validated by an independent and neutral party - provided the most comprehensive assessment of prognostic models in mCRPC. These data presented both a benchmark for future prognostic model development and rich source of clinical predictors.

All together, this study illustrated the benefits of open data access, at a time when clinicians, researchers, and the public are advocating for improved platforms and policies that encourage

sharing of clinical trial data [30,31]. PDS has overcome major barriers of data sharing with support of data providers to allow broad access of cancer clinical trial data. To researchers who are interested in leveraging open-access cancer trial data, this study represented a novel research approach that employed scientific rigor and deep understanding of clinical data through effective collaboration of a multidisciplinary team of experts.

The trials used here represent the standard of care at the time when the trials were performed, a limitation of this study. Since 2010, several therapies have become available in both pre- and post-docetaxel space, and the new trials have changed the way clinicians approach this disease[32]. Abiraterone and enzalutamide, both approved in first line setting of mCRPC, are not included in the scope of this Challenge due to limitation of control arm data – both COU-AA-302[5] and PREVAIL[6] have placebo or prednisone controls – and comparative trials using these agents as the control arm have not been performed. Accordingly, trial sponsors should be encouraged to contribute experimental arm data (particularly for approved drugs) to an active and engaged research community. While there is concern by sponsors that virtual comparisons may be made between treatments in experimental arm of different trials, there is far more benefit in leveraging these data to validate prognostic factors/models and to explore intermediate clinical endpoints predictive of survival.

The DREAM Challenge described here has demonstrated that there is opportunity to further optimize prognostic models in mCRPC using baseline clinical covariates. For significant advances beyond the work presented here, clinical trial data must be made available that reflects current advancements in treatment paradigms, including new data capture techniques such as (immuno)genomics or metabolomics that may more accurately describe the malignant state of the tumor and its micro-environment. Critical to either of these will be the need to share patient-level oncology data with the research community, an effort championed by PDS and facilitated by DREAM, for the development of the next generation of prognostic and predictive models in cancer.

Acknowledgements

This publication is based on research using information obtained from http://www.projectdatasphere.org/, which is maintained by Project Data Sphere, LLC. Neither Project Data Sphere, LLC nor the owner(s) of any information from the web site have

contributed to, approved, or are in any way responsible for the contents of this publication. We thank the Sage Bionetworks Synapse team for the development and design of the Challenge website. This work is supported in part by the following: National Institutes of Health, National Library of Medicine (2T15-LM009451), National Cancer Institute (16X064; 5R01CA152301), Boettcher Foundation, Academy of Finland (grants 292611, 269862, 272437, 279163, 295504), Cancer Society of Finland, Drug Research Doctoral Programme (DRDP) at the University of Turku and Finnish Cultural Foundation.

Author Contributions

J.G., T.W., C.B., E.C.N., B.B., K.A., T.N., S.F., G.S., Y.X., F.L.Z., and J.C.C designed the Challenge. F.L.Z and L.S. led the PDS efforts to collect and process the clinical trial data. T.D.L., S.K., G.P., A.A., T.P., T.M., and T.A. designed the top-performing method. J.G., T.W., T.D.L., K.K.W., C.B., E.C.N., G.S., T.A., F.L.Z., and J.C.C. performed the post-Challenge data analysis and interpretation. H.S., C.J.S., C.J.R., H.I.S., and O.S. assisted in clinical variable interpretation. All members of the Prostate Cancer Challenge DREAM Consortium submitted prognostic models to the Challenge, provided method write-ups and the code to reproduce their predictions. J.G., T.W., T.D.L., H.S., C.J.S., C.J.R., H.I.S., O.S., T.A., F.L.Z., and J.C.C. wrote the paper.

Competing Financial Interests

H.I.S. has consulting or advisory roles for Sanofi and Celgene, and has received travel/accommodation expenses from Sanofi and AstraZenenca. C.J.R. has consulting or advisory roles for AstraZeneca. C.J.S. has consulting or advisory roles for Sanofi and AstraZeneca. O.S. has consulting or advisory roles for Sanofi and AstraZeneca, and has received research funding from Sanofi. H.S. has consulting or advisory roles for Sanofi. F.L.Z. and L.S. are employees and own stock in Sanofi. K.A. is an employee of AstraZeneca and owns stock in AstraZeneca and Sanofi. All other authors have no conflicts to declare.

Figure Legends

Figure 1. Study design. Data was acquired from Project Data Sphere, and centrally curated by the organizing team to provide a harmonized dataset across the 4 studies. Three of the studies were provided as training data, and a fourth, ENTHUSE 33, was held back as the validation dataset. Teams submitted risk scores for the ENTHUSE 33 trail, their predictions were scored and ranked using an integrated time-dependent AUC metric.

Figure 2. Performance of the top-performing model. (A) The time-dependent area under the curve was measured from 6 to 30 months on one month intervals. The top-performing model (ePCR) is shown compared to the Halabi reference model. Overall survival was evaluated using Kaplan-Meier plots for the (A) ePCR model and the (B) Halabi reference model.

Figure 3. Graph projection of the most important covariates in the ePCR model. Automated data-driven network layout of the most significant model variables according to their interconnections with other model variables. Node size and color indicate the importance of the variable alone for OS prediction and its coefficient sign, respectively. Edge color indicates the importance of an interaction between two model variables, with darker color corresponding to stronger interaction effect. Colored sub-network modules annotate the variables based on expert curated categories.

Figure 4. Challenge meta-analysis. (A) Hierarchical clustering of patients (Euclidean distance, complete linkage) by normalized prediction scores from all 51 models (B) Kaplan-Meier plot of survival probability for the 3 patient clusters from (A). (C) Most frequently utilized variables by teams in final models. Starred variables are not used in the Halabi model.

Figure 5. Assessment of top-performing model on ENTHUSE M1 study. (a) Time-dependent AUC comparing the ePCR and Halabi models. (b) Kaplan-Meier plot of ePCR prediction, separated at median risk score.

References

1 Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA Cancer J Clin* 2011; **61**: 69–90.

2 Tanimoto T, Hori A, Kami M. Sipuleucel-T immunotherapy for castration-resistant prostate cancer. *N Engl J Med* 2010; **363**: 1966; author reply 1967–8.

3 Berruti A, Pia A, Terzolo M. Abiraterone and increased survival in metastatic prostate cancer. *N Engl J Med* 2011; **365**: 766; author reply 767–8.

4 Fizazi K, Carducci M, Smith M, *et al.* Denosumab versus zoledronic acid for treatment of bone metastases in men with castration-resistant prostate cancer: a randomised, double-blind study. *Lancet* 2011; **377**: 813–22.

5 Ryan CJ, Smith MR, de Bono JS, *et al.* Abiraterone in metastatic prostate cancer without previous chemotherapy. *N Engl J Med* 2013; **368**: 138–48.

6 Scher HI, Fizazi K, Saad F, *et al.* Increased survival with enzalutamide in prostate cancer after chemotherapy. *N Engl J Med* 2012; **367**: 1187–97.

7 de Bono JS, Oudard S, Ozguroglu M, *et al.* Prednisone plus cabazitaxel or mitoxantrone for metastatic castration-resistant prostate cancer progressing after docetaxel treatment: a randomised open-label trial. *Lancet* 2010; **376**: 1147–54.

8 Parker C, Nilsson S, Heinrich D, *et al.* Alpha emitter radium-223 and survival in metastatic prostate cancer. *N Engl J Med* 2013; **369**: 213–23.

9 Kantoff PW, Higano CS, Shore ND, *et al.* Sipuleucel-T immunotherapy for castration-resistant prostate cancer. *N Engl J Med* 2010; **363**: 411–22.

10 Beer TM, Armstrong AJ, Rathkopf DE, *et al.* Enzalutamide in metastatic prostate cancer before chemotherapy. *N Engl J Med* 2014; **371**: 424–33.

11 Halabi S, Small EJ, Kantoff PW, *et al.* Prognostic model for predicting survival in men with hormone-refractory metastatic prostate cancer. *J Clin Oncol* 2003; **21**: 1232–7.

12 Smaletz O, Scher HI, Small EJ, *et al.* Nomogram for overall survival of patients with progressive metastatic prostate cancer after castration. *J Clin Oncol* 2002; **20**: 3972–82.

13 Armstrong AJ, Garrett-Mayer ES, Yang Y-CO, de Wit R, Tannock IF, Eisenberger M. A contemporary prognostic nomogram for men with hormone-refractory metastatic prostate cancer: a TAX327 study analysis. *Clin Cancer Res* 2007; **13**: 6396–403.

14 Halabi S, Lin C-Y, Kelly WK, *et al.* Updated prognostic model for predicting overall survival in first-line chemotherapy for patients with metastatic castration-resistant prostate cancer. *J Clin Oncol* 2014; **32**: 671–7.

15 Okser S, Pahikkala T, Airola A, Salakoski T, Ripatti S, Aittokallio T. Regularized machine learning in the genetic prediction of complex traits. *PLoS Genet* 2014; **10**: e1004754.

16 Papaemmanuil E, Gerstung M, Bullinger L, *et al.* Genomic Classification and Prognosis in Acute Myeloid Leukemia. *N Engl J Med* 2016; **374**: 2209–21.

17 Longo DL, Drazen JM. Data Sharing. *N Engl J Med* 2016; **374**: 276–7.

18 Scher HI, Jia X, Chi K, *et al.* Randomized, open-label phase III trial of docetaxel plus high-dose calcitriol versus docetaxel plus prednisone for patients with castration-resistant prostate cancer. *J Clin Oncol* 2011; **29**: 2191–8.

19 Tannock IF, Fizazi K, Ivanov S, *et al.* Aflibercept versus placebo in combination with docetaxel and prednisone for treatment of men with metastatic castration-resistant prostate cancer (VENICE): a phase 3, double-blind randomised trial. *Lancet Oncol* 2013; **14**: 760–8.

20 Petrylak DP, Vogelzang NJ, Budnik N, *et al.* Docetaxel and prednisone with or without lenalidomide in chemotherapy-naive patients with metastatic castration-resistant prostate cancer (MAINSAIL): a randomised, double-blind, placebo-controlled phase 3 trial. *Lancet Oncol* 2015; **16**: 417–25.

21 Fizazi K, Fizazi KS, Higano CS, *et al.* Phase III, randomized, placebo-controlled study of docetaxel in combination with zibotentan in patients with metastatic castration-resistant prostate cancer. *J Clin Oncol* 2013; **31**: 1740–7.

22 Nelson JB, Fizazi K, Miller K, *et al.* Phase III study of the efficacy and safety of zibotentan (ZD4054) in patients with bone metastatic castration-resistant prostate cancer (CRPC). In: JOURNAL OF CLINICAL ONCOLOGY. AMER SOC CLINICAL ONCOLOGY 2318 MILL ROAD, STE 800, ALEXANDRIA, VA 22314 USA, 2011.

23 Hung H, Chiang C-T. Estimation methods for time-dependent AUC models with survival data. *Can J Stat* 2010; **38**: 8–26.

24 Costello JC, Stolovitzky G. Seeking the wisdom of crowds through challenge-based competitions in biomedical research. *Clin Pharmacol Ther* 2013; **93**: 396–8.

25 Bender E. Challenges: Crowdsourced solutions. *Nature* 2016; **533**: S62–4.

26 Bansal M, Yang J, Karan C, *et al.* A community computational challenge to predict the activity of pairs of compounds. *Nat Biotechnol* 2014; **32**: 1213–22.

27 Costello JC, Heiser LM, Georgii E, *et al.* A community effort to assess and improve drug sensitivity prediction algorithms. *Nat Biotechnol* 2014; **32**: 1202–12.

28 Margolin AA, Bilal E, Huang E, *et al.* Systematic analysis of challenge-driven improvements in molecular prognostic models for breast cancer. *Sci Transl Med* 2013; **5**: 181re1.

29 Ewing AD, Houlahan KE, Hu Y, *et al.* Combining tumor genome simulation with crowdsourcing to benchmark somatic single-nucleotide-variant detection. *Nat Methods* 2015; **12**: 623–30.

30 Merson L, Gaye O, Guerin PJ. Avoiding Data Dumpsters—Toward Equitable and Useful Data Sharing. *N Engl J Med* 2016. http://www.nejm.org/doi/full/10.1056/NEJMp1605148.

31 Bierer BE, Li R, Barnes M, Sim I. A Global, Neutral Platform for Sharing Trial Data. *N Engl J Med* 2016; **374**: 2411–3.

32 Lewis B, Sartor O. The changing landscape of metastatic prostate cancer. *Am J Hematol* 2015; **11**. http://www.gotoper.com/publications/ajho/2015/2015Aug/The-Changing-Landscape-of-Metastatic-Prostate-Cancer.

Prostate Cancer Challenge DREAM Community

Kald Abdallah[82], Tero Aittokallio[21,22], Antti Airola[23], Catalina Anghel[6], Helia Azima[45], Robert Baertsch[35], Pedro J Ballester[39,78], Chris Bare[78], Vinayak Bhandari[72], Brian M Bot[78], Cuong C Dang[39,78], Maria Bekker-Nielsen Dunbar[34], Ann-Sophie Buchardt[34], Ljubomir Buturovic[77], Da Cao[10], Prabhakar Chalise[28], Junwoo Cho[20], Tzu-Ming Chu[32], R Yates Coley[8], Sailesh Conjeti[13], Sara Correia[15,16], James C Costello[81,87], Ziwei Dai[26], Junqiang Dai[28], Philip Dargatz[3], Sam Delavarkhan[45], Detian Deng[8], Ankur Dhanik[27], Yu Du[8], Aparna Elangovan[14], Shellie Ellis[29], Laura L Elo[53,55], Shadrielle M Espiritu[72], Fan Fan[72], Ashkan B Farshi[45], Ana Freitas[16], Brooke Fridley[28], Stephen Friend[78], Christiane Fuchs[1,4], Eyal Gofer[43], Gopalacharyulu Peddinti[22], Stefan Graw[28], Russ Greiner[41,42], Yuanfang Guan[56], Justin Guinney[78], Jing Guo2[5,65], Pankaj Gupta[13], Anna I Guyer[12], Jiawei Han[47], Niels R Hansen[34], Billy HW Chang[40], Outi Hirvonen[52], Barbara Huang[72], Chao Huang[58], Jinseub Hwang[19], Joseph G Ibrahim[58], Vivek Jayaswal[50], Jouhyun Jeon[6], Zhicheng Ji[8], Deekshith Juvvadi[30], Sirkku Jyrkkiö[52], Kimberly Kanigel-Winner[81], Amin Katouzian[13], Marat D Kazanov[37], Suleiman A Khan[22], Shahin Khayyer[45], Dalho Kim[20], Agnieszka K Golińska[59], Devin Koestler[28], Fernanda Kokowicz[17], Ivan Kondofersky[1,4], Norbert Krautenbacher[1,4], Damjan Krstajic[76,77], Luke Kumar[41], Christoph Kurz[2], Matthew Kyan[74], Teemu D Laajala[21,22], Michael Laimighofer[1,4], Eunjee Lee[58], Wojciech Lesiński[59], Miaozhu Li[11], Ye Li[61,68], Qiuyu Lian[44], Xiaotao Liang[61,62], Minseong Lim[20], Henry Lin[47], Xihui Lin[6], Jing Lu[31], Mehrad Mahmoudian[53], Roozbeh Manshaei[45], Richard Meier[28], Dejan Miljkovic[13], Tuomas Mirtti[22,24], Krzysztof Mnich[60], Nassir Navab[13], Elias C Neto[78], Yulia Newton[35], Thea Norman[78], Tapio Pahikkala[23], Subhabrata Pal[51], Byeongju Park[20], Jaykumar Patel[41], Swetabh Pathak[30], Alejandrina Pattin[13], Donna P Ankerst[4,5], Jian Peng[47], Anne H Petersen[34], Robin Philip[30], Stephen R Piccolo[12], Sebastian Pölsterl[13], Aneta Polewko-Klim[59], Karthik Rao[9], Xiang Ren[47], Miguel Rocha[15,16], Witold R. Rudnicki[59,60,66], Charles J Ryan[71], Hyunnam Ryu[20], Oliver Sartor[67], Hagen Scherb[1], Raghav Sehgal[30], Fatemeh Seyednasrollah[53,55], Jingbo Shang[47], Bin Shao[26], Liji Shen[86], Howard Sher[88], Motoki Shiga[36], Artem Sokolov[35], Julia F Söllner[1], Lei Song[48], Howard Soule[69], Gustavo Stolovitzky[83], Josh Stuart[35], Ren Sun[6,7], Christopher J Sweeney[70], Nazanin Tahmasebi[41], Kar-Tong Tan[25], Lisbeth Tomaziu[34], Joseph Usset[28], Yeeleng S Vang[57], Roberto Vega[41], Vitor Vieira[16], David Wang[72], Difei Wang[49], Junmei Wang[33], Lichao Wang[13], Sheng Wang[47], Tao Wang[79,80], Yue Wang[58], Russ Wolfinger[32], Chris Wong[35], Zhenke Wu[8], Jinfeng Xiao[46], Xiaohui Xie[57], Yang Xie[79,84,85], Doris Xin[47], Hojin Yang[58], Nancy Yu[6], Thomas Yu[78], Xiang Yu[10], Sulmaz Zahedi[73,75], Massimiliano Zanin[38], Chihao Zhang[64], Jingwen Zhang[58], Shihua Zhang[64], Yanchun Zhang[61,68], Fang Liz Zhou[86], Hongtu Zhu[58], Shanfeng Zhu[61,62,63] and Yuxin Zhu[8]

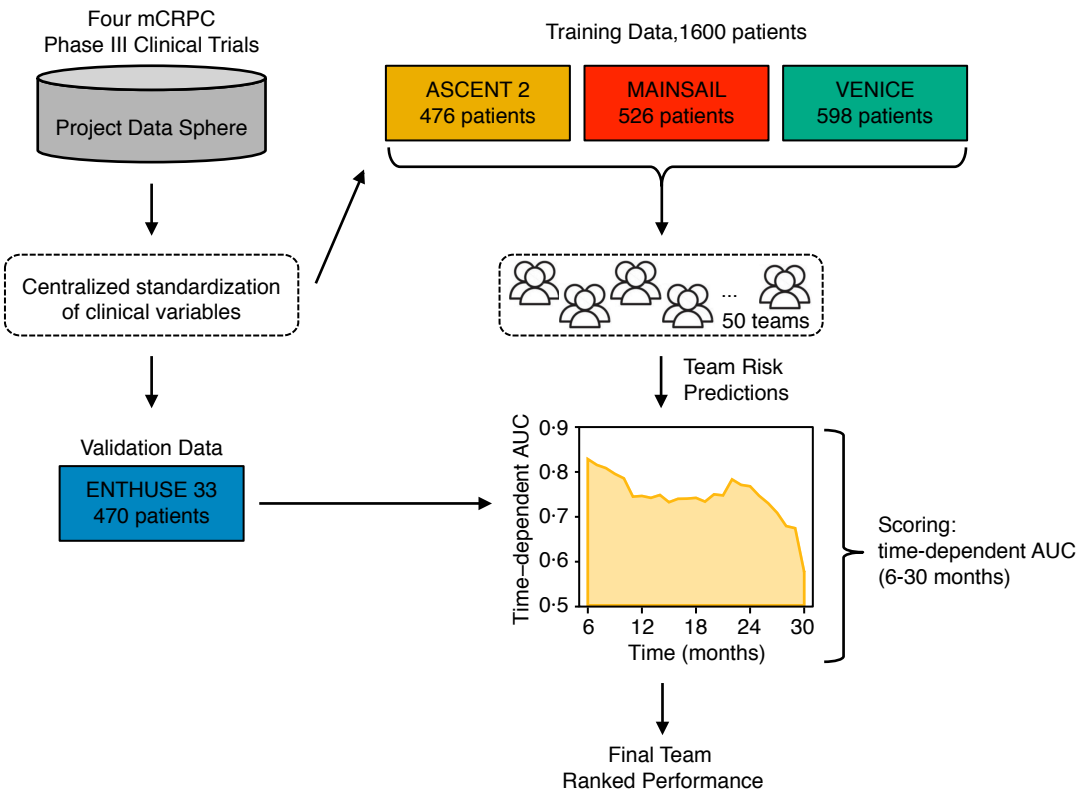
[1]Institute of Computational Biology, Helmholtz Zentrum München, Munich, Germany
[2]Institute of Health Economics and Health Care Management, Helmholtz Zentrum München, Munich, Germany
[3]Department of Hematology and Oncology, Johannes Wesling Klinikum Minden, Germany
[4]Department of Mathematics, Technische Universität München, Munich, Germany
[5]University of Texas Health Science Center at San Antonio, TX, USA
[6]Informatics and Biocomputing Program, Ontario Institute for Cancer Research (OICR), Toronto, Canada
[7]Department of Pharmacology and Toxicology, University of Toronto, Toronto, Canada
[8]Department of Biostatistics, Johns Hopkins University, Baltimore, MD 21205, USA
[9]School of Medicine, Johns Hopkins University, Baltimore, MD 21205, USA
[10]University of Pennsylvania, Philadelphia, PA, USA

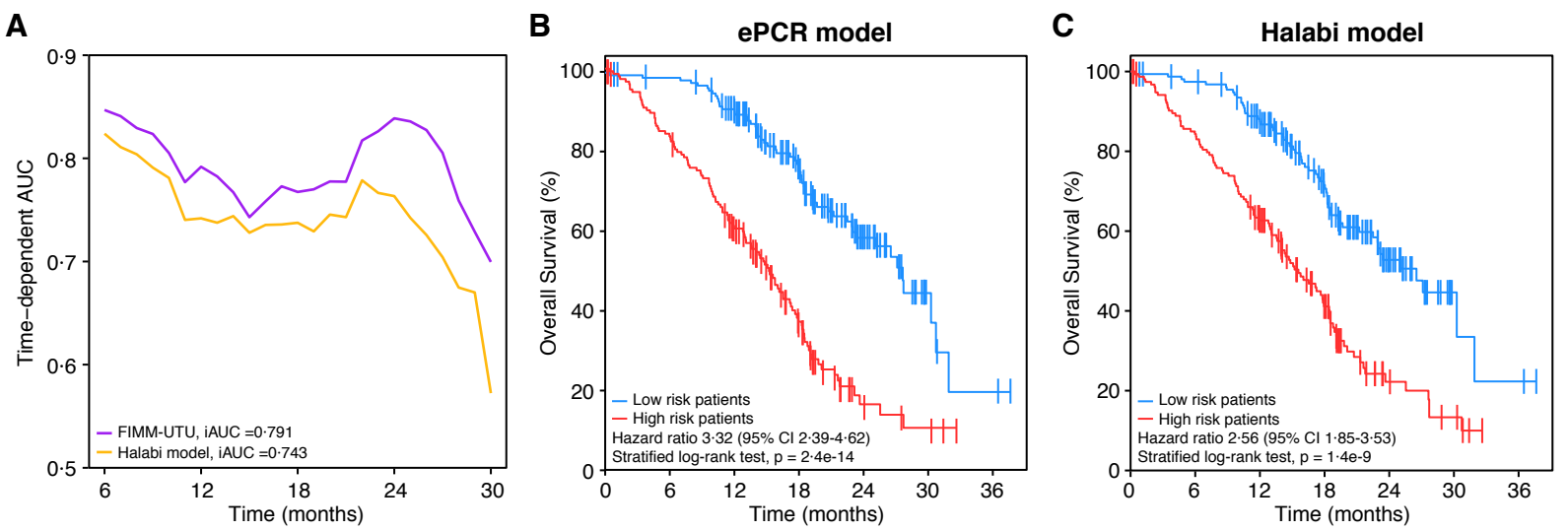
[11]Biodemography of Aging Research Unit, Center for Population Health and Aging, Social Science Research Institute, Duke University, Durham, NC, USA
[12]Department of Biology, Brigham Young University, Provo, UT, USA
[13]Computer Aided Medical Procedures, Technische Universität München, Germany
[14]Computer Science Department, University of Melbourne, Melbourne, Australia
[15]Department of Informatics, University of Minho, Portugal
[16]Centre of Biological Engineering, University of Minho, Portugal
[17]Plant Morphogenesis and Biochemistry Laboratory, Federal University of Santa Catarina, Florianpolis, Brazil
[18]Johns Hopkins University, Baltimore, MA, USA
[19]Department of Computer science and Statistics, Daegu University 712-714, Daegu, South Korea
[20]Department of Statistics, Kyungpook National University, 702-701 Daegu, South Korea
[21]Department of Mathematics and Statistics, University of Turku, Finland
[22]Institute for Molecular Medicine Finland, University of Helsinki, Finland
[23]Department of Information Technology, University of Turku, Finland
[24]Department of Pathology, Helsinki University Hospital, Finland
[25]Cancer Science Institute of Singapore, National University of Singapore, Singapore
[26]Center for Quantitative Biology, Peking University, Beijing 100871, China
[27]Regeneron Pharmaceuticals Inc, Tarrytown, New York, NY, USA
[28]Department of Biostatistics, University of Kansas Medical Center, Kansas City, KS, USA
[29]Department of Health Policy and Management, University of Kansas Medical Center, Kansas City, KS, USA
[30]Jeevomics Pvt. Ltd.
[31]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA
[32]JMP Life Sciences Division, SAS Institute Inc., Cary, NC, USA
[33]UT Southwestern, Dallas, TX, USA
[34]University of Copenhagen, Copenhagen, Denmark
[35]Department of Biomolecular Engineering and Center for Biomolecular Science and Engineering, University of California, Santa Cruz, CA, USA
[36]Department of Electrical, Electronic and Computer Engineering, Gifu University, Gifu, Japan
[37]Research and Training Center on Bioinformatics, Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, Russia
[38]INNAXIS Foundation & Research Institute, Madrid, Spain
[39]Cancer Research Centre of Marseille, Marseille, France
[40]Division of Biostatistics, Jockey Club School of Public Health and Primary Care, The Chinese University of Hong Kong
[41]Department of Computing Science, University of Alberta, Edmonton, Alberta, Canada
[42]Alberta Innovates Centre for Machine Learning, Edmonton, Alberta, Canada
[43]The Rachel and Selim Benin School of Computer Science and Engineering, The Hebrew University, Jerusalem, Israel
[44]Tsinghua University, Beijing 100084, China
[45]Electrical and Computer Engineering Dept., Ryerson University, Toronto, Canada
[46]Center for Biophysics and Quantitative Biology, The University of Illinois at Urbana-Champaign, IL, USA
[47]Department of Computer Science, The University of Illinois at Urbana-Champaign, IL, USA

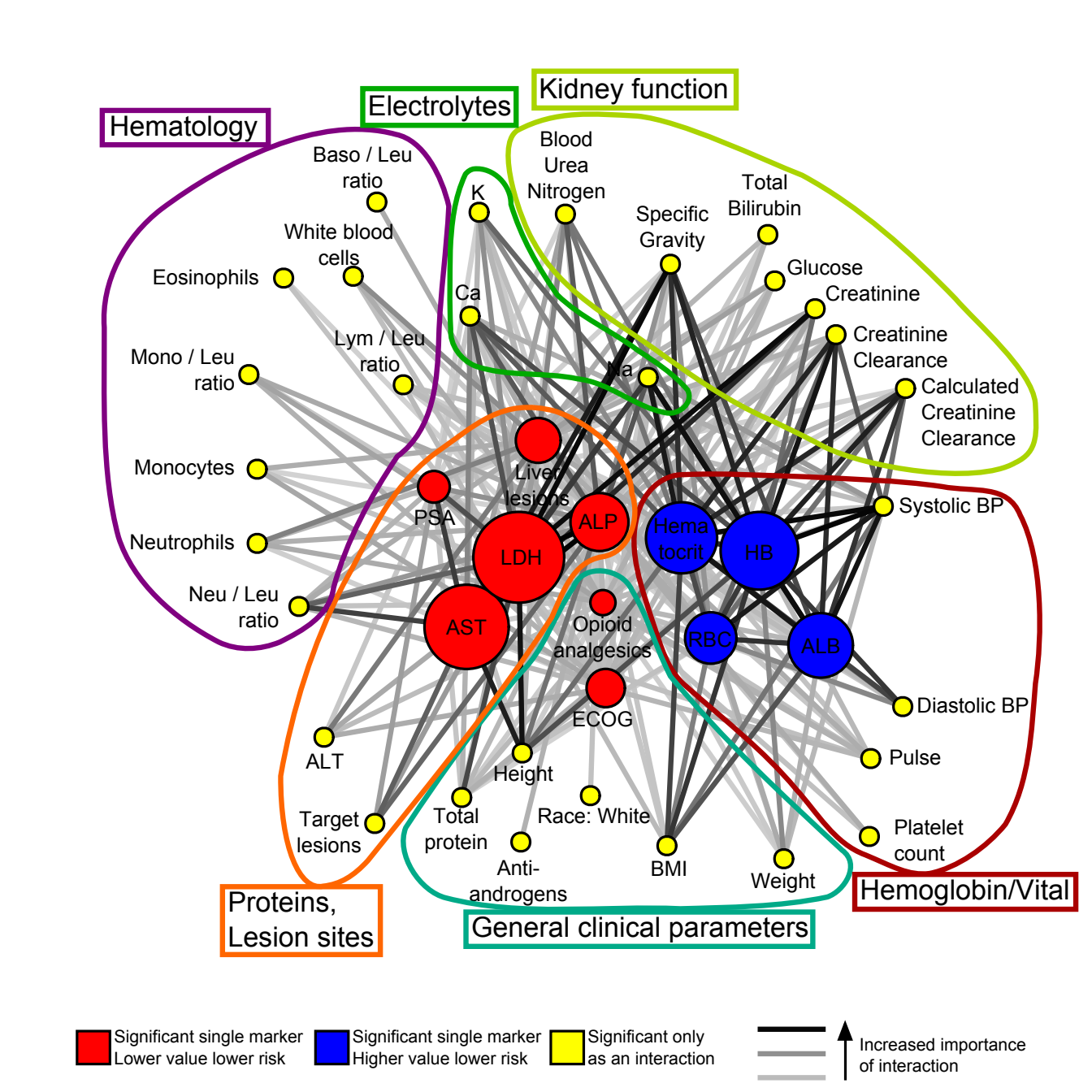
[48]National Cancer Institute, National Institutes of Health, 9609 Medical Center Dr., Rockville, MD, USA
[49]Department of Biochemistry and Molecular and Cellular Biology, Georgetown University Medical Center, 4000 Reservoir Rd NW, Washington DC, USA
[50]Biocon Bristol-Myers Squibb Research Centre, Bangalore, India
[51]Centre for Cellular and Molecular Platforms, Bangalore, India
[52]The Department of Oncology and Radiotherapy, Turku University Central Hospital, Turku, Finland
[53]Turku Centre for Biotechnology, University of Turku and Åbo Akademi University, Turku, Finland
[54]The Department of Clinical Oncology, University of Turku, Turku, Finland
[55]Department of Mathematics and Statistics, University of Turku, Turku, Finland
[56]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA
[57]Department of Computer Science, University of California Irvine, Irvine, CA, USA
[58]Biostatistics and Imaging Analysis Lab, University of North Carolina at Chapel Hill, NC, USA
[59]Faculty of Mathematics and Informatics, University of Bialystok, Poland
[60]Computational Centre, University of Bialystok, Poland
[61]School of Computer Science, Fudan University, Shanghai 200433, China
[62]Shanghai Key Lab of Intelligent Information Processing, Fudan University, Shanghai 200433, China
[63]Centre for Computational Systems Biology,Fudan University, Shanghai 200433, China
[64]National Center for Mathematics and Interdisciplinary Sciences, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, 100190 Beijing, China
[65]Research and development department, Annoroad Gene Technology Co. Ltd, Beijing, China
[66]Interdisciplinary Centre for Mathematical and Computational Modelling, University of Warsaw, Poland
[67]Tulane Cancer Center, Tulane University, New Orleans, LA, USA
[68]Shanghai Key Lab of Data Science, Fudan University, Shanghai 200433, China
[69]Prostate Cancer Foundation, Santa Monica, CA, USA
[70]Department of Medical Oncology, Dana-Farber Cancer Institute and Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA
[71]Genitourinary Medical Oncology Program, Division of Hematology & Oncology, University of California, San Francisco, CA, USA
[72]Ontario Institute for Cancer Research, Toronto, Ontario, Canada
[73]The Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, Canada
[74]Electrical Engineering and Computer Science Dept., York University, Toronto, Canada
[75]iBEST - Li Ka Shing Institute of Knowledge, St. Michael's Hospital, Toronto, Canada
[76]Research Centre for Cheminformatics, Jasenova 7, 11030 Beograd, Serbia
[77]Clinical Persona Inc, 932 Mouton Circle, East Palo Alto, CA, USA
[78]Sage Bionetworks, Seattle, WA, USA
[79]Quantitative Biomedical Research Center, Department of Clinical Sciences, University of Texas Southwestern Medical Center, Dallas, Texas, USA
[80]Center for the Genetics of Host Defense, University of Texas Southwestern Medical Center, Dallas, Texas, USA
[81]Department of Pharmacology & Computational Biosciences Program, University of Colorado,

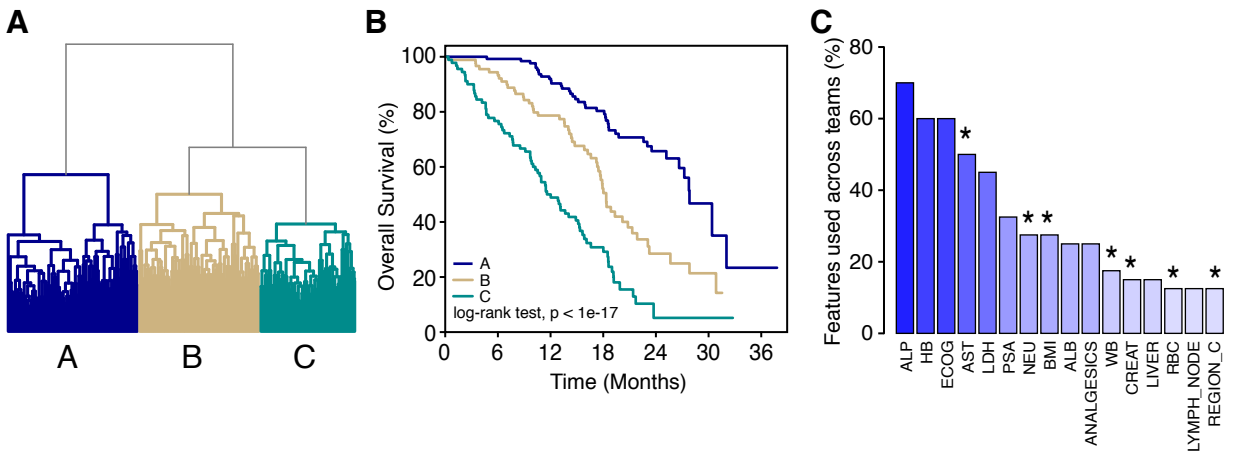
Anschutz Medical Campus, Aurora, CO, USA
[82]AstraZeneca, Gaithersburg, MD, USA
[83]IBM T.J. Watson Research Center, IBM, Yorktown Heights, NY, USA
[84]The Simmons Comprehensive Cancer Center, University of Texas Southwestern Medical Center, Dallas, TX, USA
[85]Lyda Hill Department of Bioinformatics, University of Texas Southwestern Medical Center, Dallas, TX, USA
[86]Sanofi, Bridgewater, NJ, USA
[87]University of Colorado Comprehensive Cancer Center, University of Colorado, Anschutz Medical Campus, Aurora, CO, USA
[88]Sidney Kimmel Center for Prostate and Urologic Cancers, Memorial Sloan-Kettering Cancer Center and Weill Cornell Medical College, New York, NY, USA

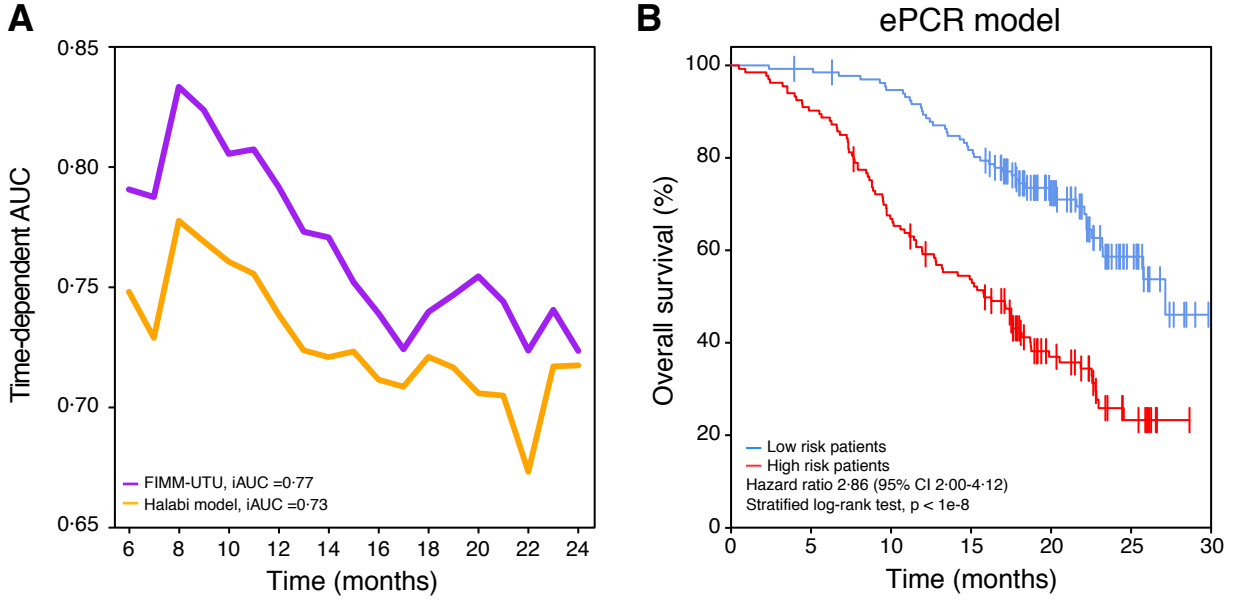
Funding Support:
European Union within the ERC grant LatentCauses supported the work of C.F and I.K. German Research Foundation (DFG) within the Collaborative Research Centre 1243, subproject A17 awarded to C.F. German Federal Ministry of Education and Research (BMBF) through the Research Consortium e:AtheroMED (Systems medicine of myocardial infarction and stroke) under the auspices of the e:Med Programme (grant # 01ZX1313C) supported the work of D.P.A., P.D., C.F., C.K., I.K., N.K., M.L., H.S. and J.F.S. at the Institute of Computational Biology. NIH Grants RR025747-01, MH086633 and 1UL1TR001111, and NSF Grants SES-1357666, DMS-14-07655 and BCS-0826844 supported the work of C.H., J.I., E.L., Y.W., H.Y., H.Z. and J.Z. NSFC Grant Nos. 61332013, 61572139 supported the work of X.L, Y.L, Y.Z., and S.Z. National Natural Science Foundation of China grants [Nos. 61422309, 61379092] was awarded to S.Z. The Patrick C. Walsh Prostate Research Fund and the Johns Hopkins Individualized Health Initiative supported the work of R.Y.C., D.D., Y.D., Z.J., K.R., Z.W. and Y.Z. FCT Ph.D. Grant SFRH/BD/80925/2011 was awarded to S.C. Clinical Persona Inc., East Palo Alto, CA supported the work of L.B. and D.K. The Finnish Cultural Foundation and the Drug Research Doctoral Programme (DRDP) at the University of Turku supported T.D.L. The National Research Foundation Singapore and the Singapore Ministry of Education, under its Research Centres of Excellence initiative, supported the work of J.G. and K.T. A grant from the Russian Science Foundation 14-24-00155 was awarded to M.D.K. A*MIDEX grant (no. ANR-11-IDEX-0001-02) was awarded to P.J.B. NSERC supported the work of R.G. The Israeli Centers of Research Excellence (I-CORE) program (Center No. 4/11) supported the work of E.G. Academy of Finland (grants 292611, 269862, 272437, 279163, 295504), National Cancer Institute (16X064), and Cancer Society of Finland supported the work of T.A.


Four mCRPC
Phase III Clinical Trials
Project Data Sphere
Centralized standardization
of clinical variables
Validation Data
ENTHUSE 33
470 patients
Training Data,1600 patients
ASCENT 2
476 patients
MAINSAIL
526 patients
VENICE
598 patients
...
50 teams
Team Risk
Predictions
Time–dependent AUC
0·9
0·8
0·7
0·6
0·5
6
12
18
24
30
Time (months)
Scoring:
time-dependent AUC
(6-30 months)
Final Team
Ranked Performance


A
Time−dependent AUC
0·9
0·8
0·7
0·6
0·5
6
12
18
24
30
Time (months)
FIMM-UTU, iAUC =0·791
Halabi model, iAUC =0·743
B
ePCR model
Overall Survival (%)
100
80
60
40
20
0
0
6
12
18
24
30
36
Time (months)
Low risk patients
High risk patients
Hazard ratio 3·32 (95% CI 2·39-4·62)
Stratified log-rank test, p = 2·4e-14
C
Halabi model
Overall Survival (%)
100
80
60
40
20
0
0
6
12
18
24
30
36
Time (months)
Low risk patients
High risk patients
Hazard ratio 2·56 (95% CI 1·85-3·53)
Stratified log-rank test, p = 1·4e-9


Hematology
Electrolytes
Kidney function
Proteins, Lesion sites
General clinical parameters
Hemoglobin/Vital
Baso / Leu ratio
White blood cells
Eosinophils
Lym / Leu ratio
Mono / Leu ratio
Monocytes
Neutrophils
Neu / Leu ratio
K
Ca
Na
Blood Urea Nitrogen
Specific Gravity
Total Bilirubin
Glucose
Creatinine
Creatinine Clearance
Calculated Creatinine Clearance
Liver lesions
PSA
LDH
ALP
AST
ALT
Target lesions
Opioid analgesics
ECOG
Height
Total protein
Race: White
Anti-androgens
BMI
Weight
Hema tocrit
HB
RBC
ALB
Systolic BP
Diastolic BP
Pulse
Platelet count
Significant single marker Lower value lower risk
Significant single marker Higher value lower risk
Significant only as an interaction
Increased importance of interaction


A
A
B
C
B
Overall Survival (%)
100
80
60
40
20
0
0
6
12
18
24
30
36
Time (Months)
A
B
C
log-rank test, p < 1e-17
C
Features used across teams (%)
80
60
40
20
0
ALP
HB
ECOG
AST
LDH
PSA
NEU
BMI
ALB
ANALGESICS
WB
CREAT
LIVER
RBC
LYMPH_NODE
REGION_C


A
0·85
0·80
0·75
0·70
0·65
Time-dependent AUC
FIMM-UTU, iAUC =0·77
Halabi model, iAUC =0·73
6
8
10
12
14
16
18
20
22
24
Time (months)
B
ePCR model
100
80
60
40
20
0
Overall survival (%)
Low risk patients
High risk patients
Hazard ratio 2·86 (95% CI 2·00-4·12)
Stratified log-rank test, p < 1e-8
0
5
10
15
20
25
30
Time (months)

	Training Set			Validation set
Characteristics	**ASCENT2 (n=476)**	**MAINSAIL (n=526)**	**VENICE (n=598)**	**ENTHUSE 33 (n=470)**
Age				
18-64	111 (23·3%)	171 (32·5%)	219 (36·6%)	160 (34%)
65-74	211 (44·3%)	246 (46·8%)	254 (42·5%)	217 (46·2%)
>=75	154 (32·4%)	109 (20·7%)	125 (20·9%)	93 (19·8%)
ECOG PS				
0	220 (46·2%)	257 (48·9%)	280 (46·8%)	247 (52·6%)
1	234 (49·2%)	247 (47%)	291 (48·7%)	223 (47·4%)
2	22 (4·6%)	20 (3·8%)	27 (4·5%)	0 (0%)
3	0 (0%)	1 (0·2%)	0 (0%)	0 (0%)
Missing	0 (0%)	1 (0·2%)	0 (0%)	0 (0%)
Metastasis				
Liver	5 (1·1%)	58 (11%)	60 (10%)	64 (13·6%)
Bone	345 (72·5%)	439 (83·5%)	529 (88·5%)	470 (100%)
Lungs	8 (1·7%)	74 (14·1%)	88 (14·7%)	56 (11·9%)
Lymph nodes	163 (34·2%)	298 (56·7%)	323 (54%)	208 (44·3%)
Analgesic use				
No	338 (71%)	347 (66%)	419 (70·1%)	339 (72·1%)
Yes	138 (29%)	179 (34%)	179 (29·9%)	131 (27·9%)
LDH, U/L				
1st Quantile	176	174	NA	181
Median	202	210	NA	213
3rd Quantile	250	267	NA	287
Missing	13 (2·7%)	1 (0·2%)	596 (99·7%)	5 (1·1%)
PSA, ng/mL				
1st Quantile	24·2	32·22	30·82	33·6
Median	68·78	84·9	90·78	99·55
3rd Quantile	188·4	271·2	260·6	236·8
Missing	1 (0·2%)	4 (0·8%)	6 (1%)	12 (2·6%)
Hemoglobin, g/dL				
1st Quantile	11·6	11·5	11·7	11·3
Median	12·6	12·7	12·7	12·5
3rd Quantile	13·6	13·7	13·5	13·5
Missing	3 (0·6%)	10 (1·9%)	0 (0%)	4 (0·9%)
Albumin, g/L				
1st Quantile	NA	41	38	40
Median	NA	43	42	43
3rd Quantile	NA	45	45	46
Missing	476 (100%)	1 (0·2%)	16 (2·7%)	2 (0·4%)
Alkaline phosphatase, U/L				
1st Quantile	80	81	85	97·75
Median	113	124	135	155
3rd Quantile	213	265	270·3	328·2
Missing	2 (0·4%)	0 (0%)	3 (0·5%)	2 (0·4%)
Aspartate aminotransferase, U/L				
1st Quantile	20	19	20	20

Median	24	24	25	25
3rd Quantile	31	31	33	33
Missing	4 (0·8%)	1 (0·2%)	8 (1·3%)	3 (0·6%)

Data are quantiles (1st, median, 3rd) or n (%). ECOG PS=ECOG Performance Status, LD=Lactate dehydrogenase, PSA= Prostate-Specific Antigen. Albumin tests for ASCENT2 trial are all missing and LDH tests for VENICE trial are almost all missing, therefore summary statistics are not computed for them.

Table 1. Baseline Characteristics

Characteristics	ENTHUSE M1 (N=266)
Age	
18-64	58 (21·8%)
65-74	111 (41·7%)
>=75	97 (36·5%)
ECOG PS	
0	196 (73·7%)
1	70 (26·3%)
Metastasis	
Liver	12 (4·5%)
Bone	266 (100%)
Lungs	13 (4·9%)
Lymph nodes	80 (30·1%)
Analgesic use	
No	256 (96·2%)
Yes	10 (3·8%)
LDH, U/L	
1st Quantile	170
Median	188
3rd Quantile	219
Missing	7 (2·6%)
PSA, ng/mL	
1st Quantile	17·3
Median	52·25
3rd Quantile	153·0
Missing	4 (1·5%)
Hemoglobin, g/dL	
1st Quantile	12·2
Median	12·9
3rd Quantile	13·7
Missing	2 (0·8%)
Albumin, g/L	
1st Quantile	41
Median	43
3rd Quantile	45
Missing	1 (0·4%)
Alkaline phosphatase, U/L	
1st Quantile	83
Median	130
3rd Quantile	222
Missing	1 (0·4%)
Aspartate aminotransferase, U/L	
1st Quantile	19
Median	24
3rd Quantile	29
Missing	3 (1·1%)

Data are quantiles (1st, median, 3rd) or n (%). ECOG PS=ECOG Performance Status, LD=Lactate dehydrogenase, PSA= Prostate-Specific Antigen.

Table 2. Baseline Characteristics

Supplement to: Guinney J, Wang T, Laajala TD, et al. An prognostic model to predict overall survival for patients with metastatic castration-resistant prostate cancer: results from an open-data, crowdsourced challenge.

Clinical trial descriptions, curation, and splitting
Challenge design, rules, and web-based resources
Evaluation of best performing team
Top-performing model description
Data-driven network projection for the ePCR model

Supplementary Tables

Supplementary Table 1. Full results across several scoring metrics from the Challenge. Teams are listed with the links to their predictions, methods write-up, and code.

Supplementary Table 2. Comparison of patient risk stratification by the ePCR and Halabi models. Patients were dichotomized at median risk scores.

Supplementary Table 3. Top 15 single and interacting variables. Comprehensive list of evaluated variables is available at: https://www.synapse.org/#!Synapse:syn7113819

Supplementary Table 4. TRIPOD Checklist: Prediction model development and validation

Supplementary Figures

Supplementary Figure 1. Overview of the top-performing ePCR method in comparison to the reference Halabi model. (A) The benchmarking Halabi model explored the LASSO model ($\alpha = 1$) in a training data cohort with respect to the regularization parameter (λ) using cross-validation (CV). (B) The top-performing ePCR approach is based on an ensemble of Penalized Cox Regression models (ePCR), which are optimized separately for each cohort or a combination of cohorts in terms of the regularization parameter (λ) as well as the full range of the L1/L2 regularization parameter ($0 <= \alpha <= 1$). (C) Ensemble predictions were generated by averaging over the predicted risk ranks from each ensemble component.

Supplementary Figure 2. (A) All data across the 4 studies – both binary and continuous data – were used in a PCA. (B) All data across the 4 studies – only binary variables – were used in PCA.

Supplementary Figure 3. (A) Density plot of follow-up times per study. (B) Survival profile for each of the trials.

Supplementary Figure 4. Summary of Challenge results. (A) Performance of submissions. Each submission underwent 1,000 paired bootstrap of final scoring patient set to calculate a Bayes factor against the top-performer and the Halabi model, as well as the P-value from randomization test. X-axis is iAUC and y-axis is submissions ranked by iAUC from high to low. Each team's bootstrapped iAUC scores are shown as horizontal boxplot with the black diamonds representing the point estimate of a team's performance. The colored boxes show the inter-quartile ranges and the whiskers extend to 1.5 times the corresponding interquartile ranges. Top-performer is colored in orange, other teams within Bayes factor of 20 were labeled in blue, and the rest of the teams were labeled in green. The Halabi model is labeled in purple. (B) Bayes factors of all submissions against the top-performer are shown. Bayes factors greater than 20 were rounded to 20. (C) Bayes factors of all submissions against the Halabi model. Bayes factors greater than 20 were rounded to 20.

Supplementary Figure 5. Calibration plots of the ePCR model applied to the validation data set at 18, 24, 30, and 36 months.

Supplementary Figure 6. Timeline for the Challenge. Five submissions were allowed per round, and only a single submission for the final validation round.

Clinical trial description, curation, data splitting

Three datasets were used to create the training dataset for the Challenge (Novacea ASCENT2[1], Sanofi VENICE[2], and Celgene MAINSAIL[3]), while one dataset (AstraZeneca ENTHUSE 33[4]) was held back for leaderboard and blinded validation. The data represented 2,070 first line mCRPC patients in four cancer trials, where all patients received docetaxel treatment in the comparator arm.

In order to perform further validation of the top-performing prognostic model algorithm, the organizing team identified a fifth trial dataset (AstraZeneca ENTHUSE M1[5]) as an independent validation dataset post-Challenge.

Due to the regulation and privacy environment of certain countries, not all patients in the comparator arm from ENTHUSE 33 and M1 were provided to PDS.

ASCENT2 (Novacea, provided by Memorial Sloan Kettering Cancer Center): ASCENT2[1] is a randomized, open-label study evaluating DN-101 in combination with docetaxel in mCRPC. Patients received docetaxel and calcitriol in comparator arm (N = 476; N = 138 events).

VENICE (Sanofi): VENICE[2] is a randomized, double-blind study comparing efficacy and safety of aflibercept versus placebo in patients treated with docetaxel / prednisone for mCRPC. Patients received docetaxel, prednisone, and placebo in comparator arm (N = 598; 433 events).

MAINSAIL (Celgene): MAINSAIL[3] is a randomized, double-blind study to evaluate efficacy and safety of docetaxel and prednisone with or without lenalidomide in patients with mCRPC. Patients received docetaxel, prednisone, and placebo in comparator arm (N = 526; 96 events).

ENTHUSE 33 (AstraZeneca): ENTHUSE 33[4] is a randomized, double-blind study to assess efficacy and safety of 10 mg ZD4054 combined with docetaxel in comparison with docetaxel in patients with mCRPC. Patients received docetaxel and placebo in comparator arm (N = 470; 255 events).

ENTHUSE M1 (AstraZeneca): ENTHUSE M1[5] is a randomized, double-blind study to assess efficacy and safety of 10 mg ZD4054 versus placebo in patients with CRPC and bone metastasis who are pain free or mildly symptomatic. Patients received only placebo in comparator arm (N = 266; 133 events).

The original datasets from PDS contained patient level raw tables that conformed to either Study Data Tabulation Model (SDTM) standards or company-specific clinical database standards. In an effort to optimize the use of these data for the Challenge, 4 sets of raw trial data first needed to be consolidated into one set of standardized raw tables.

During initial analysis scoping, 12-15 key SDTM domains were identified as targets for standardization because they covered majority of necessary information for study subjects. These domains included demographics, trial design, follow-up including survival outcomes, treatment history, lab and lesion measurement, and vital sign. The curation team converted data from each study into a common structure that then can be combined into one dataset for each domain (SDTM). Major efforts were carried out to standardize reference date, capture, and validate survival information through careful evaluation of the data, protocol, and clinical report form (CRF). The process was especially laborious around lab domain due to differences across trials. Lab test names and units could vary; the way information was presented in its original structure could be dramatically different as well. Some studies came with a single table for lab, others used 6-8 tables to capture the same level of information. However, this standardization phase was critical to ensure robustness of the Challenge data.

Once standardized raw tables were in place, clinically important baseline covariates and dependent variables relevant to the draft research questions were then created to form the “Core Table”. A list of prostate cancer related prognostic factors was pre-identified through literature review. The analysis expanded beyond the list to cover more

than 120 variables including patient demographics, risk factors, functional status, prostate cancer treatment history, concomitant medicine, prevalent comorbidity, and condition by body system, major hematology/urology test, lesion measure/location, and vital sign. Variable creation was intended to be extensive yet not exhaustive to encourage independent thinking from the DREAM community.

Six data tables were released for this Challenge. The Core Table was the main table that was summarized at the patient level with dependent variables and clinical covariates. The remaining five tables were standardized raw longitudinal tables (lab, lesion, prior medicine, medical history, vital sign) used to create the Core Table that was at the event level and could be used for additional variable creation and/or exploration.

Challenge design, rules, and web-based resources

The Challenge was hosted on Synapse (www.synapse.org), a cloud-based platform for collaborative scientific data analysis. Synapse was used to allow access to Challenge data and to track participant agreements to the appropriate data use agreements (https://www.synapse.org/#!Synapse:syn3348040) and the Challenge rules (https://www.synapse.org/#!Synapse:syn3348041).

The Challenge was designed to have several rounds, including real-time leaderboard rounds and a final scoring round. A timeline for the Challenge can be found in Supplementary Figure 5. The leaderboard rounds provided teams the ability to build their models, make predictions, submit their predictions, and get real-time feedback on their performance. A total of three leaderboard rounds were run and teams were limited to five submissions per leaderboard round. For every submission made, an email was returned to the team with several performance metrics, including the iAUC, concordance index, and the AUC for 12, 18, and 24 months. At the end of a leaderboard round, a public leaderboard was updated with the best team score for that round.

For final submissions to the final scoring round, Challenge participants created Synapse projects containing predictions from their best model together with the code used to derive them and wikis in which participants describe their methods in text and figures. Teams were only allowed one submission to the final scoring round. To ensure reproducibility of the Challenge results, the Challenge organizers ran the code of the best performing methods and reviewed team write-ups. Team scores were not released until the top performing models were verified to reproduce the predictions that the team submitted. After the final method vetting, final scores were posted publicly on the final scoring leaderboard (Supplementary Table 1).

The ASCENT2, MAINSAIL, and VENICE datasets were used as training datasets, while the ENTHUSE 33 dataset was used as the validation dataset. The ENTHUSE 33 dataset was split in a non-overlapping manner into one 157-patient leaderboard set and one 313-patient final scoring round set. To choose this separation, we generated 100 random splits and manually chose one that yielded moderately different performance accuracy between the two sets. The 157-patient leaderboard set was further split into three overlapping smaller sets for the three leaderboard rounds. Each smaller set had 126 patients. We chose the three groups by generating 100 random splits and manually chose three that were dissimilar in patient membership and each yielded a moderate difference in performance accuracy between the chosen 126 patients and the other 31 patients. Together the three groups covered the whole set of 157 patients in the leaderboard set.

Evaluation of best performing team

Teams were evaluated using several criteria to rank and determine the top-performing team(s). Principally, we were interested in the three following evaluations: a team's prediction is meaningfully (a) better than random, (b) better than the existing reference model, and (c) better than the next best performing team.

Both (b) and (c) were evaluated using the Bayes factor measurement[6,7]. To calculate the Bayes factor, we used paired bootstrap sampling of the final set of patients multiple times and scored each new sample using the designated scoring metrics to obtain a distribution for each submission. Using these distributions, we tested the hypothesis H1 (defined as submission A is better than submission B) vs H0 (defined as submission A is no better than submission B). To be more specific, the Bayes factor of submission B vs. submission A is calculating as the posterior probability of H1 as the fraction of bootstrap replications in which submission A is better than submission B divided by the posterior probability of H0 as the fraction of bootstrap replications in which submission A is no better than submission B. The Bayes factor will decide against H0 if the calculated posterior odds is larger than a pre-specified cutoff (three in this Challenge).

Better than random. To assess whether team predictions were better than random (a), a team's score was compared against an empirical null distribution from 1,000 resamplings of the dependent variable. One-sided P-values were computed and corrected for multiple testing using the Benjamini-Hochberg procedure.

Better than the reference model. The prognostic model from Halabi, et al[8] was used as the reference model for predicting OS in mCRPC. The Halabi model consists of 8 clinical variables: ECOG performance status, disease site, opioid analgesic use, LDH > 1 x ULN, albumin, hemoglobin, PSA, and alkaline phosphatase. Beta coefficients used in implementing this model were obtained from hazard ratios as reported in Table 2 from Halabi, et al.[8] The Bayes factor was calculated from 1,000 resamplings to compare the Halabi model against each submission.

Better than next-best performer. We compared each submission against the top-performing submission using the Bayes factor, calculated using 1,000 resamplings. Submissions within Bayes factor < 3 from the top-performing team were declared indistinguishable from each other. In this Challenge, the top-performing team had a Bayes factor > 3 for when compared to all other teams.

Top-performing method description

The key phases of the team FIMM-UTU method included*:* (*i*) processing of raw data input, imputation of missing values, filtering, and truncation; (*ii*) utilizing unsupervised learning to identify most relevant patterns in the training datasets; (*iii*) fitting study-wise optimized penalized Cox regression models; and (*iv*) constructing the ensemble collection of study-wise optimized components for performing the final predictions.

(*i*) In addition to the refined core table provided by the Challenge organizers, a number of additional variables were manually extracted from the available additional data tables, namely the vital signs and lab values for markers such as blood pressure and hematocrit. After an initial data matrix was composed, imputation of missing data values was carried out using penalized regression model in two steps. In the first phase, missing at random (MAR) variables were imputed, and in the second phase, structural study-wise imputation was conducted for the study-specific variables. All the variables were then truncated where appropriate and log-transformed (Supplementary Fig. 1a)). (*ii*) Study-wise differences or redundancies were observed for some features, which were dealt with by omitting or further transforming the selected variables. Interactions were introduced between the extracted single markers to derive new covariates. Principal Component Analysis (PCA) revealed systematic differences between the four studies (Supplementary Fig. 3)), which was later accounted for by modeling study-specific components through ensemble learning. Further, clinical expertise within the team was utilized by omitting non-relevant or confounding factors. Initial data matrix included 124 variables and after removing clinically irrelevant ones, redundant, or highly-skewed variables, 101 variables were left for use in the predictive modeling. Modeling of non-linearity through pairwise interactions resulted in a final data matrix with 3,422 features. (*iii*) Based on the unsupervised explorative analyses, two of the most representative studies (MAINSAIL and VENICE) were utilized in the supervised model learning. Three separate ensemble components were composed: MAINSAIL-specific ensemble component, VENICE-specific ensemble component and a combined ensemble component, which simultaneously modeled the two selected studies (Supplementary Fig. 1b). To reduce the risk of overfitting and avoid randomness bias in the binning, the final ensemble models were optimized using 10-fold cross-validation as well as averaged

over multiple cross-validation runs. The model estimation procedure identified an optimal penalization parameter (λ), which controlled for the number of non-zero coefficients in the model. Simultaneously, the L_1/L_2 norm regularization parameter (α) was explored throughout the full model spectrum, ranging from Ridge Regression ($\alpha = 0$) to Elastic Net ($0 < \alpha < 1$) and LASSO ($\alpha = 1$) in penalized regression with respect to the objective function:

$$argmax_{\beta}[\frac{2}{n}\sum_{i=1}^{n}(x_{j(i)}^{T}\beta - log\,(\sum_{j\in R_i} e^{x_j^T\beta})) - \lambda(\alpha\sum_{i=1}^{p}|\beta_i| + \frac{1}{2}(1-\alpha)\sum_{i=1}^{p}\beta_i^2)] \quad (Eqn.\ 1)$$

Here, x are the predictors (selected clinical variables or their pairwise interactions), β are the model coefficients subjected to the L_1 and L_2 norm penalization, p is the number of dimensions in the data, n is the number of observations, $j(i)$ is the index of the observation event at time t_i, and R_i is the set of indices j with $y_j \geq t_i$ (those patients at risk at time t_i). Each ensemble component resulted in a different optimum in Eqn. 1, as investigated by 10-fold cross-validated iAUC, although the resulting Elastic Net models closely resembled Ridge Regression. The penalized regression model was based on Cox proportional hazards (Eqn. 1). (*iv*) An ensemble prediction was performed by averaging the ranks over the component-wise predicted risk for the ENTHUSE 33 study (Supplementary Fig. 1c). Overall, the highest and lowest risk patients were concordantly predicted in each component. A few patient cases resulted in a moderate ensemble risk score, even if a particular ensemble component predicted a high or a low risk. Such challenging cases were controlled by not allowing any single study-specific effects to dominate the final predictions, through averaging over all the ensemble components.

Data-driven network projection for the ePCR model

The top-performing model's ensemble dual-study component was summarized by network visualization to create a clinically relevant representation of the most important markers and interaction effects (Figure 3). Each model coefficient β_i was given an importance score by computing the Elastic Net area under or above the regularization curve in the penalization and coefficient $\{\lambda, \beta_i\}$-space. Absolute values of the areas were used to rank each coefficient, which yielded a simultaneous scoring of both the effect size of the covariate as well as the importance of the feature in relation to the penalization. Statistical significance of each coefficient was then assessed by re-fitting to 10,000 bootstrapped datasets, and empirical P-values were computed as the proportion of bootstrapped coefficients where $|\beta_{i,bootstrap}| \leq 10^{-10}$ or where $\beta_{i,bootstrap}$ flipped sign. A stringent threshold of $P < 1e^{-3}$ was used to select the coefficients as network nodes (single marker) or edges (interaction effects). Ensemble p-values were averaged over all the components. Variable and interaction weighting was computed according to the average rank of the integrated regularization area over all ensemble components. The automated network layout was performed using attracting and repelling forces among the vertices, and the physical system (*graphopt*) was simulated until it reached the equilibrium (Figure 3). Top variables and interactions presented in this graph are available in the Supplementary Table 3, with the full variable and interaction list available at (https://www.synapse.org/#!Synapse:syn7113819).

Supplementary Tables

Supplementary Table 1. Full results across several scoring metrics from the Challenge. Teams are listed with the links to their predictions, methods write-up, and code.

Team	Risk score predictions	Method write-up & code	iAUC	c-index	AUC12	AUC18	AUC24
FIMM-UTU (ePCR)	syn4732198	syn4227610	0.7915	0.7307	0.7918	0.7674	0.8388
Team Cornfield	syn4732339	syn4732274	0.7789	0.7263	0.7708	0.7663	0.8147
TeamX	syn4732955	syn4732218	0.7778	0.7157	0.7492	0.7645	0.8369
jls	syn4732934	syn4732827	0.7758	0.7212	0.7713	0.7553	0.8085
PC LEARN	syn4733119	syn3822697	0.7743	0.7205	0.7577	0.762	0.8258
KUstat	syn4741808	syn4260742	0.7732	0.7126	0.7436	0.7533	0.8376
A Bavarian dream	syn4732177	syn5592405	0.7725	0.7237	0.7721	0.7664	0.8019
qiuyulian1994	syn4732213	syn4732205	0.7716	0.711	0.7423	0.7506	0.8297
JayHawks	syn4731663	syn4214500	0.7711	0.7193	0.7717	0.7607	0.8124
Wind	syn4731647	syn4731645	0.771	0.7181	0.7625	0.7688	0.8124
Alvin	syn4732814	syn4229406	0.7707	0.7136	0.7586	0.7568	0.7927
brainstorm	syn4730818	syn3821841	0.7706	0.718	0.7617	0.7614	0.8175
uci-cbcl	syn4731657	syn4227279	0.7704	0.717	0.76	0.7716	0.8206
DreamOn	syn4731710	syn4731708	0.7704	0.712	0.7559	0.7582	0.8245
Clinical Persona	syn4681602	syn4681529	0.7704	0.7149	0.7533	0.7545	0.8328
Murat Dundar	syn4595033	syn4595029	0.7701	0.7305	0.7763	0.7773	0.773
Mistral	syn4622079	syn4622016	0.7689	0.7073	0.7382	0.7624	0.8268
UNC-BIAS	syn4731768	syn4731674	0.7685	0.717	0.7559	0.7568	0.8293
Team Marie	syn4731882	syn4485029	0.7682	0.7142	0.7519	0.7705	0.8151
A Elangovan	syn4643159	syn4212102	0.7677	0.7135	0.7655	0.7461	0.7977
M S	syn4730601	syn4229266	0.7671	0.707	0.7372	0.7652	0.8256
Jeevomics	syn4733845	syn4074987	0.7651	0.719	0.7733	0.7526	0.7917
CAMP	syn4731373	syn3647478	0.7646	0.7077	0.7331	0.758	0.8143
DAL_LAB	syn4731755	syn4731746	0.7642	0.7103	0.7521	0.7486	0.8305
Yuanfang Guan	syn4680160	syn3816890	0.7618	0.7143	0.7545	0.7631	0.8005
Bmore Dream Team	syn4733165	syn3616830	0.761	0.7121	0.7464	0.766	0.7948
Brigham Young University	syn4733391	syn4382527	0.7578	0.7048	0.7381	0.7685	0.7599
Team Simon	syn4733651	syn4732901	0.7573	0.7033	0.7278	0.7611	0.827
alan.saul	syn4731492	syn4587469	0.7568	0.7078	0.7464	0.7606	0.7961
BiSBII-UM	syn4733056	syn4229636	0.7561	0.6992	0.7394	0.7397	0.8007
RUBME6	syn4733262	syn4590933	0.7547	0.6994	0.7419	0.7198	0.7866
Jing Zhou	syn4646618	syn3685423	0.7507	0.6994	0.7361	0.7491	0.803
TYTDreamChallenge	syn4733257	syn4228911	0.748	0.7002	0.7343	0.7402	0.7657

UoB_Prostate	syn4733441	syn4591879	0.7478	0.7057	0.7468	0.7367	0.7699
Junmei Wang	syn4732891	syn4225820	0.7475	0.694	0.7319	0.7332	0.7955
Halabi Model	syn4770841	syn3324780	0.7429	0.6985	0.7418	0.7375	0.7634
Trishna	syn4730580	syn4730570	0.742	0.6922	0.7285	0.7383	0.774
CQB	syn4732202	syn3566822	0.7412	0.6914	0.7185	0.7293	0.7686
Ye Li	syn4731357	syn4731355	0.74	0.6907	0.7258	0.7249	0.806
Zhang Chihao	syn4748861	syn4259433	0.7376	0.7063	0.7561	0.7426	0.745
Guoping Feng	syn4730823	syn4730561	0.7261	0.6781	0.7073	0.707	0.7504
Y P	syn4732913	syn4732909	0.7241	0.6799	0.732	0.7057	0.7594
RainLab	syn4730829	syn4238316	0.7232	0.6708	0.7141	0.7394	0.7821
forPro	syn4707761	syn4707464	0.7219	0.6839	0.7267	0.7249	0.739
Marat Kazanov	syn4731369	syn4730567	0.7215	0.6675	0.7089	0.7112	0.7524
Jing Lu	syn4732498	syn4556277	0.7035	0.6689	0.6931	0.7073	0.7154
orion	syn4733693	syn4732963	0.6837	0.6457	0.717	0.7359	0.7952
limax	syn4732094	syn4721051	0.6756	0.6484	0.7033	0.6685	0.689
ECOP	syn4647266	syn4647259	0.6746	0.6554	0.6774	0.6881	0.6949
Massimiliano Zanin	syn4732241	syn4732239	0.6171	0.6081	0.6206	0.432	0.3852
The Data Wizard	syn4229053	syn4228992	0.5945	0.5815	0.6039	0.5824	0.6085
Compiled set of all predictions		syn7071669					

Supplementary Table 2. Comparison of patient risk stratification by the ePCR and Halabi models. Patients were dichotomized at median risk scores.

ePCR	**Median survival time (month)**	**1 year survival rate**	**2 year survival rate**			
Low risk group	27.6	90.20%	58.60%			
High risk group	15.1	59.90%	15.70%			
Halabi	**Median survival time (month)**	**1 year survival rate**	**2 year survival rate**			
Low risk group	26.5	87.40%	52.80%			
High risk group	15.6	62.70%	22.20%			
	Time (months)	t=6	t=12	t=18	t=24	t=30
	Cases	28	75	121	153	160
	Survivors	279	214	118	41	9
	Censored	6	24	74	119	144
	ePCR	92.89	81.32	72.63	65.86	60.67
Sensitivity (%)	Halabi	85.73	75.94	67.43	61.19	61.21
	ePCR	54.48	60.28	68.64	82.93	66.67
Specificity (%)	Halabi	53.76	57.94	64.41	73.17	44.44
	ePCR	16.96	40.15	64.2	86.31	82.41
PPV (%)	Halabi	15.65	37.17	59.46	78.85	73.93
	ePCR	98.71	90.78	76.41	59.78	39.7
NPV (%)	Halabi	97.41	88.02	71.86	53.57	30.8

Supplementary Table 3. Top 15 single and interacting variables. Comprehensive list of evaluated variables is available at: https://www.synapse.org/#!Synapse:syn7113819

Top 15 single variables		Ensemble *p*-value	Ensemble effect size
Lactate Dehydrogenase (LDH)		< 0.0001	3405.667
Aspartate Aminotransferase (AST)		< 0.0001	3376.667
Hemoglobin (HB)		< 0.0001	3369.667
Hematocrit (HEMAT)		< 0.0001	3354.333
Albumin (ALB)		0.0004	3316.667
Alkaline Phosphatase (ALP)		< 0.0001	3291.333
Red blood cell count (RBC)		< 0.0001	3237.333
Systolic blood pressure (SYSTOLICBP)		0.0012	3192.000
Lesions at liver (LIVER)		< 0.0001	3184.000
Sodium (NA.)		0.0205	3032.000
Lesions at target site (TARGET)		0.0118	3001.000
ECOG performance status (ECOG_C)		0.0003	2923.000
Medical History: Cardiac disorders (MHCARD)		0.1100	2827.667
Lymphocyte/Leukocyte ratio (LYMperLEU)		0.0143	2684.333
Body mass index (BMI)		0.0214	2679.333
Top 15 interactions		**Ensemble *p*-value**	**Ensemble effect size**
AST	LDH	< 0.0001	3408.333
ALP	LDH	< 0.0001	3406.667
ALP	AST	< 0.0001	3404.333
HB	SYSTOLICBP	< 0.0001	3402.333
LDH	Specific Gravity	< 0.0001	3400.667
SYSTOLICBP	HEMAT	< 0.0001	3400.333
Creatinine	LDH	< 0.0001	3397.333
LDH	LDH	< 0.0001	3392.000
HB	ALB	< 0.0001	3387.333
AST	AST	< 0.0001	3384.333
HB	NA.	< 0.0001	3382.667

Height	LDH	< 0.0001	3381.667
ALB	SYSTOLICBP	< 0.0001	3379.333
HB	Creatinine Clearance	< 0.0001	3378.000
ALB	HEMAT	< 0.0001	3377.333

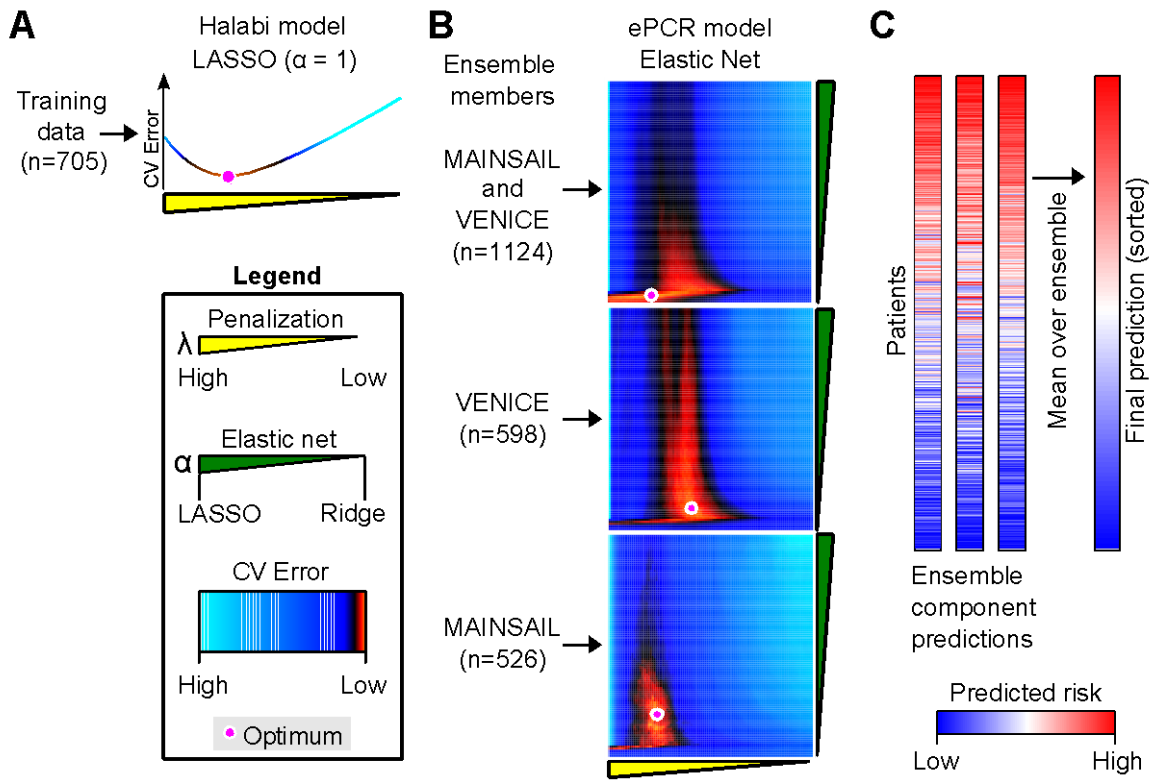
Supplementary Table 4. TRIPOD Checklist: Prediction model development and validation

Section/Topic			Checklist Item	Page
Title and abstract				
Title	1	D;V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1
Abstract	2	D;V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	2
Introduction				
Background and objectives	3a	D;V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	3
	3b	D;V	Specify the objectives, including whether the study describes the development or validation of the model or both.	3/4
Methods				
Source of data	4a	D;V	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	5
	4b	D;V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	6
Participants	5a	D;V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	6
	5b	D;V	Describe eligibility criteria for participants.	6
	5c	D;V	Give details of treatments received, if relevant.	6
Outcome	6a	D;V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	6/7
	6b	D;V	Report any actions to blind assessment of the outcome to be predicted.	7
Predictors	7a	D;V	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	7
	7b	D;V	Report any actions to blind assessment of predictors for the outcome and other predictors.	7
Sample size	8	D;V	Explain how the study size was arrived at.	NA
Missing data	9	D;V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	8
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.	8
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	8
	10c	V	For validation, describe how the predictions were calculated.	8
	10d	D;V	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	7
	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done.	NA
Risk groups	11	D;V	Provide details on how risk groups were created, if done.	10/11
Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	6
Results				
Participants	13a	D;V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	NA
	13b	D;V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	6
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	Tbl1
Model development	14a	D	Specify the number of participants and outcome events in each analysis.	6
	14b	D	If done, report the unadjusted association between each candidate predictor and outcome.	NA
Model specification	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	Fig3
	15b	D	Explain how to the use the prediction model.	5 (URL

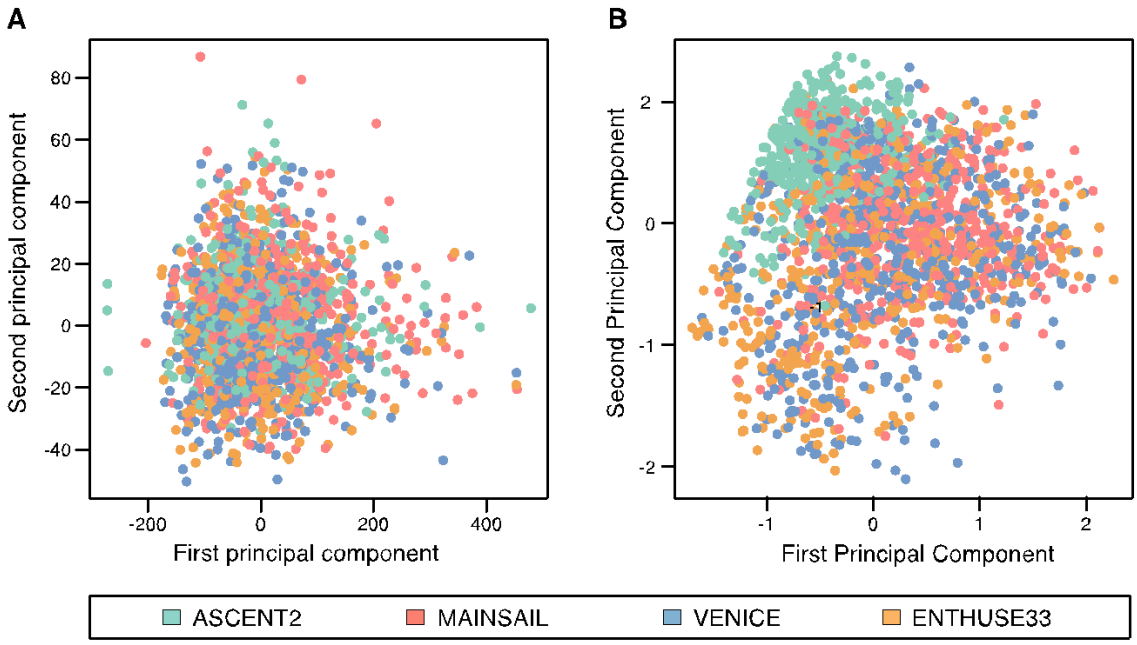
				to model code)
Model performance	16	D;V	Report performance measures (with CIs) for the prediction model.	10
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).	N/A
Discussion				
Limitations	18	D;V	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	13
Interpretation	9a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.	12/13
	9b	D;V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	12/13
Implications	20	D;V	Discuss the potential clinical use of the model and implications for future research.	13
Other information				
Supplementary information	21	D;V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	5
Funding	22	D;V	Give the source of funding and the role of the funders for the present study.	1

*Items relevant only to the development of a prediction model are denoted by D, items relating solely to a validation of a prediction model are denoted by V, and items relating to both are denoted D;V. We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.

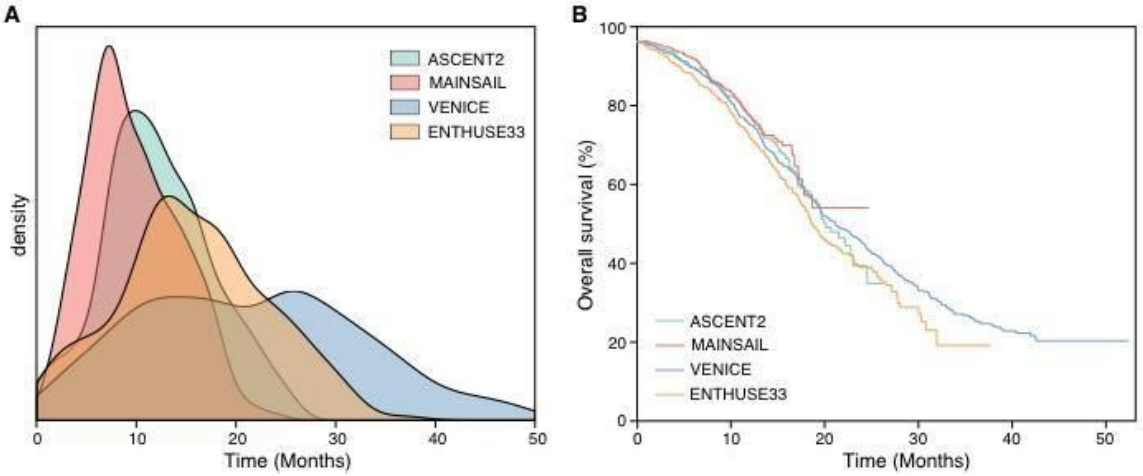
Supplementary Figures



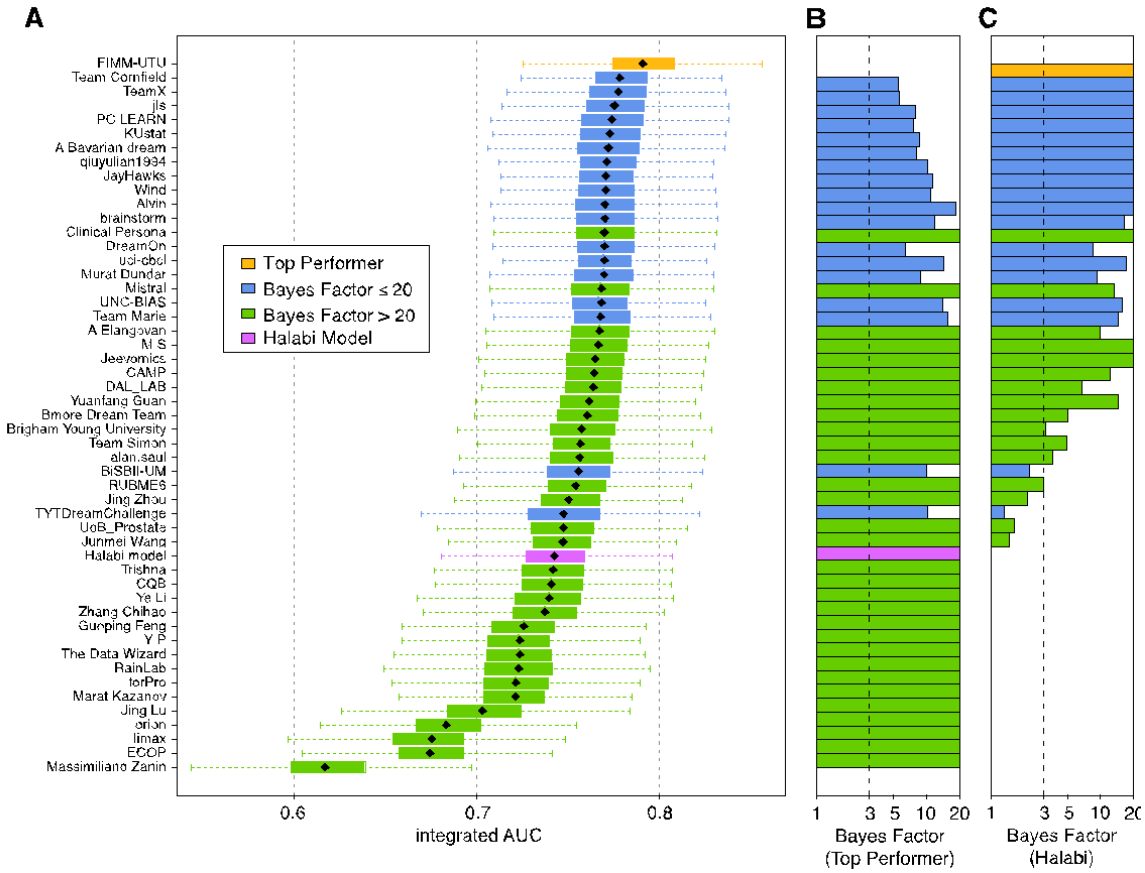
Supplementary Figure 1. Overview of the top-performing ePCR method in comparison to the reference Halabi model. (A) The benchmarking Halabi model explored the LASSO model (α = 1) in a training data cohort with respect to the regularization parameter (λ) using cross-validation (CV). (B) The top-performing ePCR approach is based on an ensemble of Penalized Cox Regression models (ePCR), which are optimized separately for each cohort or a combination of cohorts in terms of the regularization parameter (λ) as well as the full range of the L1/L2 regularization parameter ($0 <= \alpha <= 1$). (C) Ensemble predictions were generated by averaging over the predicted risk ranks from each ensemble component.



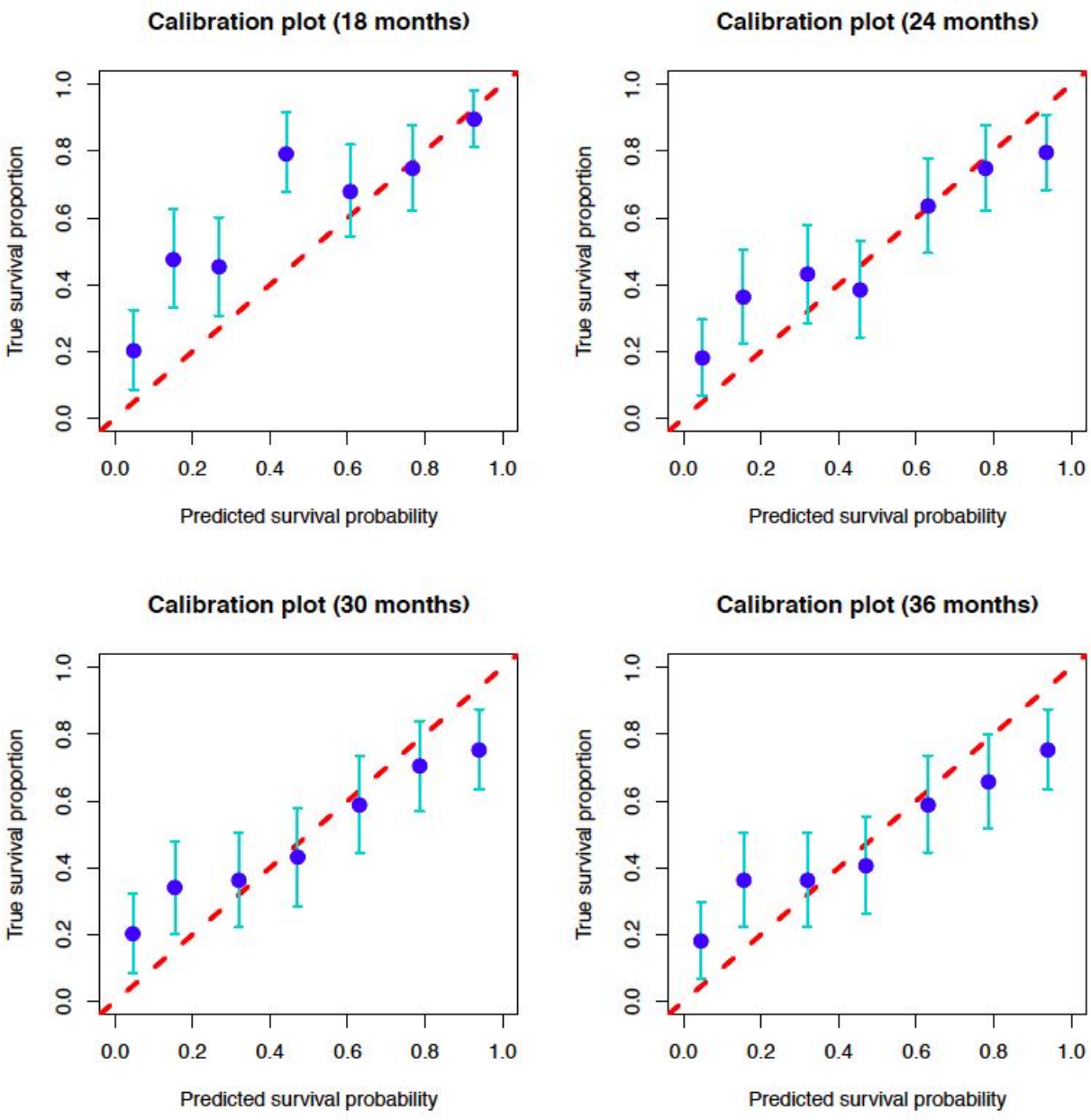
Supplementary Figure 2. (A) All data across the 4 studies – both binary and continuous data – were used in a PCA. (B) All data across the 4 studies – only binary variables – were used in PCA.



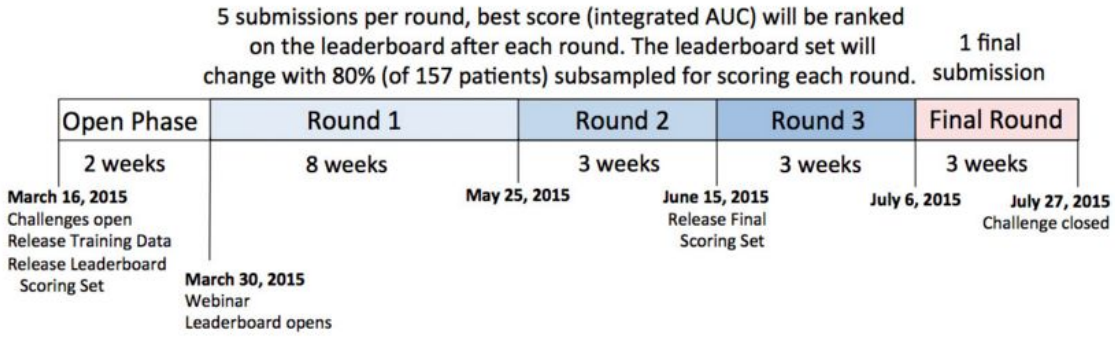
Supplementary Figure 3. (A) Density plot of follow-up times per study. (B) Survival profile for each of the trials.



Supplementary Figure 4. Summary of Challenge results. (A) Performance of submissions. Each submission underwent 1,000 paired bootstrap of final scoring patient set to calculate a Bayes factor against the top-performer and the Halabi model, as well as the P-value from randomization test. X-axis is iAUC and y-axis is submissions ranked by iAUC from high to low. Each team's bootstrapped iAUC scores are shown as horizontal boxplot with the black diamonds representing the point estimate of a team's performance. The colored boxes show the inter-quartile ranges and the whiskers extend to the most extreme data points which are no more than 1.5 times the corresponding interquartile ranges. Top-performer is colored in orange, other teams within Bayes factor of 20 were labeled in blue, and the rest of the teams were labeled in green. The Halabi model is labeled in purple. (B) Bayes factors of all submissions against the top-performer are shown. Bayes factors greater than 20 were rounded to 20. (C) Bayes factors of all submissions against the Halabi model. Bayes factors greater than 20 were rounded to 20.



Supplementary Figure 5. Calibration plots of predicted survival probability vs true survival proportion for the ENTHUSE33 validation dataset at 18, 24, 30, and 36 months.



Supplementary Figure 6. Timeline for the Challenge. Five submissions were allowed per round, and only a single submission for the final validation round.

References

1 Scher HI, Jia X, Chi K, *et al.* Randomized, open-label phase III trial of docetaxel plus high-dose calcitriol versus docetaxel plus prednisone for patients with castration-resistant prostate cancer. *J Clin Oncol* 2011; **29**: 2191–8.

2 Tannock IF, Fizazi K, Ivanov S, *et al.* Aflibercept versus placebo in combination with docetaxel and prednisone for treatment of men with metastatic castration-resistant prostate cancer (VENICE): a phase 3, double-blind randomised trial. *Lancet Oncol* 2013; **14**: 760–8.

3 Petrylak DP, Vogelzang NJ, Budnik N, *et al.* Docetaxel and prednisone with or without lenalidomide in chemotherapy-naive patients with metastatic castration-resistant prostate cancer (MAINSAIL): a randomised, double-blind, placebo-controlled phase 3 trial. *Lancet Oncol* 2015; **16**: 417–25.

4 Fizazi K, Fizazi KS, Higano CS, *et al.* Phase III, randomized, placebo-controlled study of docetaxel in combination with zibotentan in patients with metastatic castration-resistant prostate cancer. *J Clin Oncol* 2013; **31**: 1740–7.

5 Nelson JB, Fizazi K, Miller K, *et al.* Phase III study of the efficacy and safety of zibotentan (ZD4054) in patients with bone metastatic castration-resistant prostate cancer (CRPC). *J Clin Oncol* 2011; 29:abstract 117.

6 Kass RD, Raftery AE. Bayes factors. *JASA*. 1995; **90**: 773-795.

7 Goodman SN. Towards evidence-based medical statistics. 2: The Bayes factor. *Ann Intern Med* 1999; **130**: 1005-13.

8 Halabi S, Lin C-Y, Kelly WK, *et al.* Updated prognostic model for predicting overall survival in first-line chemotherapy for patients with metastatic castration-resistant prostate cancer. *J Clin Oncol* 2014; **32**: 671–7.

Prostate Cancer Challenge DREAM Community

Kald Abdallah[82], Tero Aittokallio[21,22], Antti Airola[23], Catalina Anghel[6], Helia Azima[45], Robert Baertsch[35], Pedro J Ballester[39,78], Chris Bare[78], Vinayak Bhandari[72], Brian M Bot[78], Cuong C Dang[39,78], Maria Bekker-Nielsen Dunbar[34], Ann-Sophie Buchardt[34], Ljubomir Buturovic[77], Da Cao[10], Prabhakar Chalise[28], Junwoo Cho[20], Tzu-Ming Chu[32], R Yates Coley[8], Sailesh Conjeti[13], Sara Correia[15,16], James C Costello[81,87], Ziwei Dai[26], Junqiang Dai[28], Philip Dargatz[3], Sam Delavarkhan[45], Detian Deng[8], Ankur Dhanik[27], Yu Du[8], Aparna Elangovan[14], Shellie Ellis[29], Laura L Elo[53,55], Shadrielle M Espiritu[72], Fan Fan[72], Ashkan B Farshi[45], Ana Freitas[16], Brooke Fridley[28], Stephen Friend[78], Christiane Fuchs[1,4], Eyal Gofer[43], Gopalacharyulu Peddinti[22], Stefan Graw[28], Russ Greiner[41,42], Yuanfang Guan[56], Justin Guinney[78], Jing Guo2[5,65], Pankaj Gupta[13], Anna I Guyer[12], Jiawei Han[47], Niels R Hansen[34], Billy HW Chang[40], Outi Hirvonen[52], Barbara Huang[72], Chao Huang[58], Jinseub Hwang[19], Joseph G Ibrahim[58], Vivek Jayaswal[50], Jouhyun Jeon[6], Zhicheng Ji[8], Deekshith Juvvadi[30], Sirkku Jyrkkiö[52], Kimberly Kanigel-Winner[81], Amin Katouzian[13], Marat D Kazanov[37], Suleiman A Khan[22], Shahin Khayyer[45], Dalho Kim[20], Agnieszka K Golińska[59], Devin Koestler[28], Fernanda Kokowicz[17], Ivan Kondofersky[1,4], Norbert Krautenbacher[1,4], Damjan Krstajic[76,77], Luke Kumar[41], Christoph Kurz[2], Matthew Kyan[74], Teemu D Laajala[21,22], Michael Laimighofer[1,4], Eunjee Lee[58], Wojciech Lesiński[59], Miaozhu Li[11], Ye Li[61,68], Qiuyu Lian[44], Xiaotao Liang[61,62], Minseong Lim[20], Henry Lin[47], Xihui Lin[6], Jing Lu[31], Mehrad Mahmoudian[53], Roozbeh Manshaei[45], Richard Meier[28], Dejan Miljkovic[13], Tuomas Mirtti[22,24], Krzysztof Mnich[60], Nassir Navab[13], Elias C Neto[78], Yulia Newton[35], Thea Norman[78], Tapio Pahikkala[23], Subhabrata Pal[51], Byeongju Park[20], Jaykumar Patel[41], Swetabh Pathak[30], Alejandrina Pattin[13], Donna P Ankerst[4,5], Jian Peng[47], Anne H Petersen[34], Robin Philip[30], Stephen R Piccolo[12], Sebastian Pölsterl[13], Aneta Polewko-Klim[59], Karthik Rao[9], Xiang Ren[47], Miguel Rocha[15,16], Witold R. Rudnicki[59,60,66], Charles J Ryan[71], Hyunnam Ryu[20], Oliver Sartor[67], Hagen Scherb[1], Raghav Sehgal[30], Fatemeh Seyednasrollah[53,55], Jingbo Shang[47], Bin Shao[26], Liji Shen[86], Howard Sher[88], Motoki Shiga[36], Artem Sokolov[35], Julia F Söllner[1], Lei Song[48], Howard Soule[69], Gustavo Stolovitzky[83], Josh Stuart[35], Ren Sun[6,7], Christopher J Sweeney[70], Nazanin Tahmasebi[41], Kar-Tong Tan[25], Lisbeth Tomaziu[34], Joseph Usset[28], Yeeleng S Vang[57], Roberto Vega[41], Vitor Vieira[16], David Wang[72], Difei Wang[49], Junmei Wang[33], Lichao Wang[13], Sheng Wang[47], Tao Wang[79,80], Yue Wang[58], Russ Wolfinger[32], Chris Wong[35], Zhenke Wu[8], Jinfeng Xiao[46], Xiaohui Xie[57], Yang Xie[79,84,85], Doris Xin[47], Hojin Yang[58], Nancy Yu[6], Thomas Yu[78], Xiang Yu[10], Sulmaz Zahedi[73,75], Massimiliano Zanin[38], Chihao Zhang[64], Jingwen Zhang[58], Shihua Zhang[64], Yanchun Zhang[61,68], Fang Liz Zhou[86], Hongtu Zhu[58], Shanfeng Zhu[61,62,63] and Yuxin Zhu[8]

[1]Institute of Computational Biology, Helmholtz Zentrum München, Munich, Germany
[2]Institute of Health Economics and Health Care Management, Helmholtz Zentrum München, Munich, Germany
[3]Department of Hematology and Oncology, Johannes Wesling Klinikum Minden, Germany
[4]Department of Mathematics, Technische Universität München, Munich, Germany
[5]University of Texas Health Science Center at San Antonio, TX, USA
[6]Informatics and Biocomputing Program, Ontario Institute for Cancer Research (OICR), Toronto, Canada
[7]Department of Pharmacology and Toxicology, University of Toronto, Toronto, Canada
[8]Department of Biostatistics, Johns Hopkins University, Baltimore, MD 21205, USA
[9]School of Medicine, Johns Hopkins University, Baltimore, MD 21205, USA
[10]University of Pennsylvania, Philadelphia, PA, USA
[11]Biodemography of Aging Research Unit, Center for Population Health and Aging, Social Science Research Institute, Duke University, Durham, NC, USA
[12]Department of Biology, Brigham Young University, Provo, UT, USA
[13]Computer Aided Medical Procedures, Technische Universität München, Germany
[14]Computer Science Department, University of Melbourne, Melbourne, Australia
[15]Department of Informatics, University of Minho, Portugal
[16]Centre of Biological Engineering, University of Minho, Portugal
[17]Plant Morphogenesis and Biochemistry Laboratory, Federal University of Santa Catarina, Florianpolis, Brazil
[18]Johns Hopkins University, Baltimore, MA, USA
[19]Department of Computer science and Statistics, Daegu University 712-714, Daegu, South Korea
[20]Department of Statistics, Kyungpook National University, 702-701 Daegu, South Korea

[21]Department of Mathematics and Statistics, University of Turku, Finland
[22]Institute for Molecular Medicine Finland, University of Helsinki, Finland
[23]Department of Information Technology, University of Turku, Finland
[24]Department of Pathology, Helsinki University Hospital, Finland
[25]Cancer Science Institute of Singapore, National University of Singapore, Singapore
[26]Center for Quantitative Biology, Peking University, Beijing 100871, China
[27]Regeneron Pharmaceuticals Inc, Tarrytown, New York, NY, USA
[28]Department of Biostatistics, University of Kansas Medical Center, Kansas City, KS, USA
[29]Department of Health Policy and Management, University of Kansas Medical Center, Kansas City, KS, USA
[30]Jeevomics Pvt. Ltd.
[31]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA
[32]JMP Life Sciences Division, SAS Institute Inc., Cary, NC, USA
[33]UT Southwestern, Dallas, TX, USA
[34]University of Copenhagen, Copenhagen, Denmark
[35]Department of Biomolecular Engineering and Center for Biomolecular Science and Engineering, University of California, Santa Cruz, CA, USA
[36]Department of Electrical, Electronic and Computer Engineering, Gifu University, Gifu, Japan
[37]Research and Training Center on Bioinformatics, Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, Russia
[38]INNAXIS Foundation & Research Institute, Madrid, Spain
[39]Cancer Research Centre of Marseille, Marseille, France
[40]Division of Biostatistics, Jockey Club School of Public Health and Primary Care, The Chinese University of Hong Kong
[41]Department of Computing Science, University of Alberta, Edmonton, Alberta, Canada
[42]Alberta Innovates Centre for Machine Learning, Edmonton, Alberta, Canada
[43]The Rachel and Selim Benin School of Computer Science and Engineering, The Hebrew University, Jerusalem, Israel
[44]Tsinghua University, Beijing 100084, China
[45]Electrical and Computer Engineering Dept., Ryerson University, Toronto, Canada
[46]Center for Biophysics and Quantitative Biology, The University of Illinois at Urbana-Champaign, IL, USA
[47]Department of Computer Science, The University of Illinois at Urbana-Champaign, IL, USA
[48]National Cancer Institute, National Institutes of Health, 9609 Medical Center Dr., Rockville, MD, USA
[49]Department of Biochemistry and Molecular and Cellular Biology, Georgetown University Medical Center, 4000 Reservoir Rd NW, Washington DC, USA
[50]Biocon Bristol-Myers Squibb Research Centre, Bangalore, India
[51]Centre for Cellular and Molecular Platforms, Bangalore, India
[52]The Department of Oncology and Radiotherapy, Turku University Central Hospital, Turku, Finland
[53]Turku Centre for Biotechnology, University of Turku and Åbo Akademi University, Turku, Finland
[54]The Department of Clinical Oncology, University of Turku, Turku, Finland
[55]Department of Mathematics and Statistics, University of Turku, Turku, Finland
[56]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA
[57]Department of Computer Science, University of California Irvine, Irvine, CA, USA
[58]Biostatistics and Imaging Analysis Lab, University of North Carolina at Chapel Hill, NC, USA
[59]Faculty of Mathematics and Informatics, University of Bialystok, Poland
[60]Computational Centre, University of Bialystok, Poland
[61]School of Computer Science, Fudan University, Shanghai 200433, China
[62]Shanghai Key Lab of Intelligent Information Processing, Fudan University, Shanghai 200433, China
[63]Centre for Computational Systems Biology,Fudan University, Shanghai 200433, China
[64]National Center for Mathematics and Interdisciplinary Sciences, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, 100190 Beijing, China

[65]Research and development department, Annoroad Gene Technology Co. Ltd, Beijing, China
[66]Interdisciplinary Centre for Mathematical and Computational Modelling, University of Warsaw, Poland
[67]Tulane Cancer Center, Tulane University, New Orleans, LA, USA
[68]Shanghai Key Lab of Data Science, Fudan University, Shanghai 200433, China
[69]Prostate Cancer Foundation, Santa Monica, CA, USA
[70]Department of Medical Oncology, Dana-Farber Cancer Institute and Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA
[71]Genitourinary Medical Oncology Program, Division of Hematology & Oncology, University of California, San Francisco, CA, USA
[72]Ontario Institute for Cancer Research, Toronto, Ontario, Canada
[73]The Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, Canada
[74]Electrical Engineering and Computer Science Dept., York University, Toronto, Canada
[75]iBEST - Li Ka Shing Institute of Knowledge, St. Michael's Hospital, Toronto, Canada
[76]Research Centre for Cheminformatics, Jasenova 7, 11030 Beograd, Serbia
[77]Clinical Persona Inc, 932 Mouton Circle, East Palo Alto, CA, USA
[78]Sage Bionetworks, Seattle, WA, USA
[79]Quantitative Biomedical Research Center, Department of Clinical Sciences, University of Texas Southwestern Medical Center, Dallas, Texas, USA
[80]Center for the Genetics of Host Defense, University of Texas Southwestern Medical Center, Dallas, Texas, USA
[81]Department of Pharmacology & Computational Biosciences Program, University of Colorado, Anschutz Medical Campus, Aurora, CO, USA
[82]AstraZeneca, Gaithersburg, MD, USA
[83]IBM T.J. Watson Research Center, IBM, Yorktown Heights, NY, USA
[84]The Simmons Comprehensive Cancer Center, University of Texas Southwestern Medical Center, Dallas, TX, USA
[85]Lyda Hill Department of Bioinformatics, University of Texas Southwestern Medical Center, Dallas, TX, USA
[86]Sanofi, Bridgewater, NJ, USA
[87]University of Colorado Comprehensive Cancer Center, University of Colorado, Anschutz Medical Campus, Aurora, CO, USA
[88]Sidney Kimmel Center for Prostate and Urologic Cancers, Memorial Sloan-Kettering Cancer Center and Weill Cornell Medical College, New York, NY, USA

Funding Support:
European Union within the ERC grant LatentCauses supported the work of C.F and I.K. German Research Foundation (DFG) within the Collaborative Research Centre 1243, subproject A17 awarded to C.F. German Federal Ministry of Education and Research (BMBF) through the Research Consortium e:AtheroMED (Systems medicine of myocardial infarction and stroke) under the auspices of the e:Med Programme (grant # 01ZX1313C) supported the work of D.P.A., P.D., C.F., C.K., I.K., N.K., M.L., H.S. and J.F.S. at the Institute of Computational Biology. NIH Grants RR025747-01, MH086633 and 1UL1TR001111, and NSF Grants SES-1357666, DMS-14-07655 and BCS-0826844 supported the work of C.H., J.I., E.L., Y.W., H.Y., H.Z. and J.Z. NSFC Grant Nos. 61332013, 61572139 supported the work of X.L, Y.L, Y.Z., and S.Z. National Natural Science Foundation of China grants [Nos. 61422309, 61379092] was awarded to S.Z. The Patrick C. Walsh Prostate Research Fund and the Johns Hopkins Individualized Health Initiative supported the work of R.Y.C., D.D., Y.D., Z.J., K.R., Z.W. and Y.Z. FCT Ph.D. Grant SFRH/BD/80925/2011 was awarded to S.C. Clinical Persona Inc., East Palo Alto, CA supported the work of L.B. and D.K. The Finnish Cultural Foundation and the Drug Research Doctoral Programme (DRDP) at the University of Turku supported T.D.L. The National Research Foundation Singapore and the Singapore Ministry of Education, under its Research Centres of Excellence initiative, supported the work of J.G. and K.T. A grant from the Russian Science Foundation 14-24-00155 was awarded to M.D.K. A*MIDEX grant (no. ANR-11-IDEX-0001-02) was awarded to P.J.B. NSERC supported the work of R.G. The Israeli Centers of Research Excellence (I-CORE) program (Center No. 4/11) supported the work of E.G. Academy of Finland (grants 292611, 269862, 272437, 279163, 295504), National Cancer Institute (16X064), and Cancer Society of Finland supported the work of T.A.

Academy of Finland (grant 268531) supported the work of T.M.