

SUPPLEMENTARY APPENDIX TO: Seyednasrollah F, Koestler DC, Wang T, et al. A community-based collaboration to build prediction models for short-term discontinuation of docetaxel in metastatic castration-resistant prostate cancer patients

TABLE OF CONTENTS

ADDITIONAL SUPPORTING INFORMATION.....	2
Clinical trial description	2
Data curation	2
Data splitting into training and validation sets	3
Creation of the dependent variable	3
Challenge design, rules, and web-based resources.....	3
Model evaluation and determination of top performing teams/models	4
Post-Challenge ensemble-based prediction model	5
SUPPLEMENTARY TABLES	6
Supplementary Table S1. Criteria defining the five curated discontinuation categories	6
Supplementary Table S2. Study-specific distributions over curated discontinuation categories.....	7
Supplementary Table S3. Preprocessing and analytical strategies used by Challenge participants.....	8
Supplementary Table S4. Full results from the Challenge.....	10
Supplementary Table S5. Lift chart analysis results across Challenge submissions	11
Supplementary Table S6. Data dictionary and features abbreviation definition	12
SUPPLEMENTARY FIGURES	16
Supplementary Figure S1. Schematic diagram illustrating the various phases of the Challenge	16
Supplementary Figure S2. Conceptual diagram for the training and evaluation of the post-Challenge ensemble prediction model	17
Supplementary Figure S3. Scatterplot of the first two principal components computed using baseline clinical features collected on trial participants.....	18
Supplementary Figure S4. Summary of Challenge results.....	19
Supplementary Figure S5. Spearman correlation of the risk-scores submitted by the Challenge top performing models	20
Supplementary Figure S6. Performance of the post-Challenge ensemble prediction model compared to the Challenge top performing models	21
Supplementary Figure S7. Results from the clinical trial simulation analysis	22
Supplementary Figure S8. Patient clustering and clinical variables for webtool	23
Supplementary Figure S9. Challenge Timeline	24
REFERENCES	25
Prostate Cancer Challenge DREAM Community.....	26

ADDITIONAL SUPPORTING INFORMATION

Clinical trial description

The data used in this challenge were collated data based on de-identified comparator arm data sets of four Phase III prostate cancer clinical trials hosted on Project Data Sphere (PDS) (<https://www.projectdatasphere.org/projectdatasphere/html/pcdc>). Three data sets were used to create the training data set for the Challenge (Novacea ASCENT2, Sanofi VENICE, and Celgene MAINSAIL), and one data set (AstraZeneca ENTHUSE 33) was withheld for leaderboard scoring and validation (Supplementary Figure 1). In total, the data consisted of 2,070 first line mCRPC patients enrolled in one of four different cancer trials.

- 1. ASCENT2¹ (Novacea, provided by Memorial Sloan Kettering Cancer Center).** ASCENT2 is a randomized, open-Label study evaluating DN-101 in combination with docetaxel in mCRPC. Patients received docetaxel and calcitriol in the comparator arm (n = 476; 105 patients discontinued docetaxel within three months due to AE or possible AE). Detailed inclusion/exclusion criteria is described on page 2192 from the published study.
- 2. VENICE² (Sanofi).** VENICE is a randomized, double-blind study comparing efficacy and safety of aflibercept versus placebo in mCRPC patients treated with docetaxel and prednisone. Patients received docetaxel, prednisone and placebo in the comparator arm (n = 598; 51 patients discontinued docetaxel within three months due to AE or possible AE). Detailed inclusion/exclusion criteria is described on pages 761-762 from the published study.
- 3. MAINSAIL³ (Celgene).** MAINSAIL is a randomized, double-blind study to evaluate efficacy and safety of docetaxel and prednisone with or without lenalidomide among mCRPC patients. Subjects received docetaxel, prednisone and placebo in the comparator arm (n = 526; 41 patients discontinued docetaxel within three months). Detailed inclusion/exclusion criteria is described on page 418 from the published study.
- 4. ENTHUSE 33⁴ (AstraZeneca).** ENTHUSE 33 is a randomized, double-blind study to assess efficacy and safety of 10 mg ZD4054 combined with docetaxel in comparison with docetaxel only among mCRPC patients. Subjects received docetaxel and placebo in the comparator arm (n = 470; 49 patients discontinued docetaxel within three months due to AE or possible AE). Detailed inclusion/exclusion criteria is described on page 1741 from the published study.

All patients received a docetaxel-based treatment regimen in the comparator arm. Due to regulation and privacy environment of certain countries, not all patients in the comparator arm from ENTHUSE 33 were provided to PDS.

Data curation

The original data sets from PDS contained patient level raw tables, which 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 this Challenge, we first consolidated the four sets of raw trial data into a single set of standardized raw tables.

During initial analysis scoping, 12-15 key SDTM domains were identified as targets for standardization as they covered the majority of necessary information across study subjects. These domains included: patient demographics, trial design, follow-up information (i.e., survival, time-to-discontinuation), treatment history, lab and lesion measurements, and vital signs. The Challenge organizing team converted data from each study into a common structure, which was then combined into a single data set conforming to SDTM standards. Significant effort was made to standardize reference dates, capture and validate survival information through careful evaluation of the data, protocol and clinical report forms (CRF). The process was especially laborious for lab and lesion measurements due to differences across trials. Lab test names

and units, as well as the way the information was presented, often varied across trials. Some studies included a single table containing lab values, whereas others used anywhere from 6-8 tables to capture the same level of information. While considerable effort was expended with regard to the standardization of data across trials, this phase was critical to ensure robustness of data used in this Challenge.

Once the standardized raw tables were in place, clinically important baseline covariates and dependent variables relevant to the study challenges were created to form the “Core Data Table” to improve accessibility and consistency in the data. Prostate cancer specific prognostic factors were pre-identified through literature review. Our analysis expanded beyond this 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 tests, lesion measures/locations, and vital signs. Variable creation was intended to be extensive, yet not exhaustive, to encourage independent thinking among challenge participants.

In total, six data tables were released to the challenge participants. The Core Data Table was summarized at the patient level and included both dependent variables and clinical covariates. The remaining five tables were standardized raw longitudinal tables at the event level (lab, lesion, prior medicine, medical history, vital signs) and were used to create Core Data Table. These five tables were also given to teams for the purpose of additional variable creation and/or data exploration. All data used in this Challenge is available for download from Synapse (<https://synapse.org/#!/Synapse:syn3325825>).

Data splitting into training and validation sets

The ASCENT2, MAINSAIL, and VENICE datasets were used as training datasets (total n = 1,600), whereas the ENTHUSE 33 dataset was used as an independent validation dataset (n = 470). The ENTHUSE 33 dataset was split into two non-overlapping data sets. The first data set consisted of 30% of the patients in the validation data (n = 157) and was provided to Challenge participants to assist statistical modeling. Only covariate data for patients in this set was provided to participating teams; the outcome variables (i.e., survival time and discontinuation status) were blinded to teams throughout the Challenge. The second set was comprised of 70% of the patients in the validation data (n = 313) and was completely blinded to teams throughout the majority of Challenge. Specifically, six weeks prior to the Challenge deadline the covariate data for patients in the second set was released so that teams could submit risk-scores for final scoring. To determine the allocation of patients in the above two data sets, the validation data was randomly split 100 times and the split resulting in the minimum difference in the data distribution between the two sets was selected.

Creation of the dependent variable

The dependent variable “DISCONT” was derived from two factors: (1) reason for treatment discontinuation (i.e., “discontinue reason”) and (2) the time from treatment initiation to discontinuation (i.e., “discontinue time”). Using the raw trial data, a patient’s reason for treatment discontinuation was first assigned to one of five major categories: (1) discontinuation due to an adverse event (AE), (2) discontinuation possibly due to an AE, (3) death or progression, (4) completed treatment, and (5) a miscellaneous group (Supplementary Table 1). Supplementary Table 2 contains the number and percentage of patients assigned to the above categories within each of the four trial data sets.

Patients were labeled as DISCONT=1 if and only if they discontinued treatment due to AE or possible AE within 3 months (3x30.5 days) after beginning treatment. Otherwise, patients were labeled as DISCONT=0 (Supplementary Table 2). Challenge participants were permitted to make adjustments to the criteria used for defining DISCONT if they wished, but the definition of DISCONT in the validation set used for scoring was fixed based on the description above. Patients in “miscellaneous groups” and patients with missing discontinue time information in the validation set (n = 16) were not included in the scoring of models. While challenge participants were encouraged to discard “miscellaneous” patients and patients with missing discontinue time in the training set, it was left for teams to decide how to best handle these patients in their analyses.

Challenge design, rules, and web-based resources

A schematic diagram of the Challenge is given in Supplementary Figure 1. 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>).

Teams were tasked with developing models for individualized risk predictions of docetaxel discontinuation (due to AE or possible AE) within three months after beginning treatment. A timeline for the Challenge can be found in Supplementary Figure 7. Six weeks prior to the Challenge deadline, teams were given access to the covariate data collected on patients in the validation set, i.e., ENTHUSE 33. Using this data, teams submitted risk-scores for each patient for up to two prediction models. For final submissions, Challenge participants were asked to create Synapse projects containing their predictions, corresponding code, and wikis describing their analytical approach. To assure reproducibility of the challenge, the organizers of submissions ran the code of the best performing methods. 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 4).

Model evaluation and determination of top performing teams

Teams were scored and ranked based on their prediction performance in ENTHUSE 33 trial data set ($n = 470$). Performance was assessed based on the area under the precision-recall curve (AUPRC), computed using ROCR package in R (<https://cran.r-project.org/web/packages/ROCR/>). AUPRC values range between 0 and 1, with larger values indicating improved prediction performance. For teams that submitted predictions for two models, the model achieving the best (highest) AUPRC score was used to determine their final ranking in the leaderboard.

The rationale for using AUPRC as opposed to the area under the receiver operating characteristic curve (AUCROC) was due to the large skew in the distribution of the dependent variable within the ENTHUSE 33 trial (only 10.4% of patients were labeled as DISCONT = 1)⁵. Since 10.4% of patients in ENTHUSE 33 validation set discontinued docetaxel within 3 months due to an AE or possible AE, the expected AUPRC for a random prediction model is equal to 0.104; thus providing a performance benchmark for Challenge submissions.

To determine the top-performing teams, the following two evaluations were considered: (1) predictions were significantly better than a random prediction model and (2) predictions from the top-performing team(s) were better than next best performing team by a statistically meaningful margin.

1. Significantly better than random. To assess whether team predictions were better than random, a team's AUPRC score was compared to the empirical null distribution, generated through 5,000 permutations of the dependent variable. For each team, one-sided p-values were computed as the probability of observing an AUPRC under the empirical null distribution at least as large as the AUPRC computed for that team. P-values were corrected for multiple testing using the Benjamini-Hochberg procedure⁶ and multiple-testing adjusted p-values less than 10% ($P < 0.10$) were considered statistically significant.

2. Better than next-best. To assess whether consecutively ranked teams were measurably distinguishable in their scores, each submission was compared against the first ranked submission using the Bayes factor^{7,8} and submissions within Bayes factor of three from the first ranked submission (i.e., Bayes factor ≤ 3) were declared statistically indistinguishable from one another. To calculate the Bayes factor, we used paired bootstrap sampling of the ENTHUSE 33 data set (5000 replications) and scored each new sample using the designated scoring metric to obtain a distribution of AUPRC for each submission. Using these distributions, we tested the null hypothesis H0 (defined as submission A is no better than submission B) against the alternative hypothesis H1 (defined as submission A is better than submission B). More specifically, Bayes factor was computed as the ratio of the posterior probability of H1 - calculated as the

fraction of bootstrap replications in which submission A was better than submission B - by the posterior probability of H_0 , the fraction of bootstrap replications in which submission A is no better than submission B. The Bayes factor provides evidence against H_0 if the calculated posterior odds is larger than a pre-specified cutoff (three in this Challenge).

Although not used in determining a team's placement in the leaderboard, a cumulative lift chart analysis was performed on each model submitted to the Challenge as an alternative assessment of the clinical utility of prediction models (Supplementary Table 5). For a given model, patients in the ENTHUSE 33 data set were first rank-ordered according to their predicted risk. Among the top $M\%$ of patients with the highest predicted risk, the fraction of patients discontinued docetaxel treatment within three months was recorded and used to compute the lift ratio at the selected cutoff percentage (i.e., M). The lift ratio represents enrichment of truly discontinued patients over patients predicted to discontinue.

Post-Challenge ensemble-based prediction model

Following the completion of the Challenge, an ensemble-based prediction model⁹ was generated as a function of the models submitted by the top teams (Supplementary Figure 2). As seven teams were declared top-performers in this Challenge, this model was consequently constructed using the individual prediction models submitted by each of these teams. To construct the ensemble-based model, top-performing teams first ran their learning algorithm $L_i(\cdot)$, $i = 1, \dots, 7$ on the full training data, D ($N = 1,600$), to produce the following predictors: $\{C_1(r), \dots, C_7(r)\}$. Each predictor, $C_i(r) \in \mathfrak{R}$, represents an estimate of the risk that patient r will discontinue docetaxel-treatment due to AE or possible AE within three months of beginning treatment. Thus, $C_i(r) > C_i(s)$ indicates that for the i^{th} predictor, the predicted risk for short-term treatment discontinuation is larger for patient r compared to patient s . Using the predictors generated by each of the top teams, an ensemble-based prediction model was generated as the following simple weighted average:

$$C_w(r) = \sum_{i=1}^7 w_i C_i(r) \text{ (Eqn. 1)}$$

with weights, w_i , $i = 1, \dots, 7$, proportional to the prediction accuracy of $C_i(\cdot)$. To learn these weights, the training data, D , was randomly split into two independent sets: D^{70} , which contained 70% of the patients in the training data ($N = 1,120$) and D^{30} , which contained the remaining 30% ($N = 480$). The learning algorithms $L_i(\cdot)$ generated by the top teams were first trained on D^{70} to produce seven new predictors $\{C_i^{70}(\cdot), i = 1, \dots, 7\}$. Each of these predictors were then applied D^{30} to obtain an estimate of risk for early treatment discontinuation across each of the patients in D^{30} . Given that discontinuation status (i.e., DISCONT) was observed for all patients in D , and consequently D^{30} , the prediction accuracy associated with each of the classifiers, A_i , was computed as the fraction of patients that were correctly predicted to discontinue treatment. Finally, weights were set as: $w_i^{70} = A_i$, leading to the following predictor:

$$C_w^{70}(r) = \sum_{i=1}^7 w_i^{70} C_i^{70}(r) \text{ (Eqn. 2)}$$

As our analyses demonstrated favorable performance of $C_w^{70}(r)$ when applied to D^{30} , these weights were used to construct the ensemble-based prediction model in Eqn. 1, with $w_i = w_i^{70}$. The ensemble-based prediction model, $C_w(r)$, was then applied to the ENTHUSE 33 data set and the resulting AUPRC was computed and compared scores achieved by individual models submitted to the Challenge.

To determine if the AUPRC score generated from the ensemble-based prediction model represented an improvement over the scores obtained from individual model submissions, bootstrap sampling was used to approximate the distribution of AUPRC for each team and for the ensemble-based model. For each bootstrap sample (5000 total replications), the difference in the AUPRC scores between the ensemble-based model and individual submissions were computed, allowing us to estimate the fraction of times the ensemble-based model outperformed each of the individual model submissions. The Bayes factor between each team and the ensemble-based model was also calculated using the procedure described above.

SUPPLEMENTARY TABLES

Supplementary Table S1. Criteria defining the five curated discontinuation categories.

Curated category	Raw category
Discontinued due to adverse event	ADVERSE EVENT
	ADVERSE EVENT (COMPLETE AE FORM)
	DOCETAXEL TOXICITY (COMPLETE GI, TE, AND/OR AE FORMS AS APPROPRIATE)
Discontinued possibly due to adverse event	AT THE REQUEST OF SUBJECT (WILL CONTINUE ON STUDY)
	DEVELOPMENT OF STUDY SPECIFIC DISCONTINUATION CRITERIA
	OTHER REASON
	INVESTIGATORS DECISION
	OTHER
	VOLUNTARY DISCONTINUATION BY SUBJECT
	WITHDRAWAL OF CONSENT
Completed study	EMPTY STRING (one EFC6546 patient)
	CONDITION UNDER INVESTIGATION IMPROVED / SUBJECT RECOVERED
	COMPLETED 30 WEEKS OF STUDY TREATMENT PHASE
	MAXIMUM CYCLE OF CHEMOTHERAPY REACHED
Death or progression	LACK OF THERAPEUTIC RESPONSE
	CONDITION UNDER INVESTIGATION WORSENERD
	DEATH
	DEATH (COMPLETE SURVIVAL FORM)
Miscellaneous groups	LOST TO FOLLOW UP
	SUBJECT LOST TO FOLLOW-UP
	POOR COMPLIANCE TO PROTOCOL
	PROTOCOL VIOLATION

Supplementary Table S2. Summary of the study-specific distributions over the five curated discontinuation categories.

	Training datasets						Validation dataset	
	ASCENT2 (n=476)		MAINSAIL (n=526)		VENICE (n=598)		ENTHUSE 33 (n=470)	
Discontinuation class	All patients	Discontinue <3 months	All patients	Discontinue <3 months	All patients	Discontinue <3 months	All patients	Discontinue <3 months
Discontinued due to adverse event	43 (9.0%)	11 (2.3%)	70 (13.3%)	18 (3.4%)	126 (21.1%)	33 (5.5%)	116 (24.7%)	30 (6.4%)
Discontinued possibly due to adverse event	223 (46.8%)	94 (19.7%)	238 (45.2%)	23 (4.4%)	133 (22.2%)	18 (3.0%)	72 (15.3%)	19 (4.0%)
Completed study	115 (24.2%)		95 (18.1%)		1 (0.2%)		162 (34.5%)	
Death or progression	95 (20.0%)		112 (21.3%)		333 (55.7%)		104 (22.1%)	
Miscellaneous groups	0 (0.0%)		11 (2.1%)		5 (0.8%)		16 (3.0%)	

Patient counts are shown as n (%). The table shows the detailed discontinuation status for each study. For patients who discontinued due to AE or possibly due to AE, the subset that discontinued before 3 months is also shown.

Supplementary Table S3. Summary of preprocessing and analytical strategies used by Challenge participants

Team name	Preprocessing and Imputation	Outcome distribution	Model building and evaluation
1. Decision trees			
Yuanfang Guan (Y G)	Preprocessing: Adverse event was transformed to early death and hazard ratios from Halabi et al. model ¹⁰ were induced to the model.	Binary	Decision trees regression model evaluated by CV.
2. Logistic regression			
Jayhawks	Preprocessing: Largely missed variables were removed and variables were scaled to have similar distributions across datasets. Imputation: Median.	Binary	Logistic regression evaluated by CV.
3. Ensemble techniques			
TYTDreamChallenge	Preprocessing: Largely missed and redundant variables were removed. Log transformation was performed for highly skewed variables.	TTE modeling	Gradient Boosting Machines evaluated by CV.
PC-LEARN	Preprocessing: Irrelevant variables were removed. Imputation: Median/Mode.	Binary	Random Forest evaluated by CV.
Brigham Young University	Preprocessing: Largely missed variables were removed and performed variable standardization. Imputation: Median/Mode.	Binary	Balanced Random Forest evaluated by CV.
jls	Preprocessing: Log transformation was performed and variables were standardized to normal mean and variance. Imputation: Full Conditional Specification (FCS) method.	Binary	Ensemble of regression models, KNN and ridge evaluated by CV.
A Bavarian dream	Preprocessing: Largely missed and redundant variables were removed. Log transformations, scaling between trials and corrections and assignments into categories based on clinical expertise were performed. Imputation: 5 fold MICE.	Binary	Random Forest (tree base) evaluated by CV.
orion	Preprocessing: Largely missed variables and categorical variables with sporadic distribution were removed. Imputation: Median.	Binary	Gradient Boosting Machines evaluated by CV.
Zhang Chihao	Preprocessing: Largely missed variables were removed. Imputation: Random forest was utilized to impute missing data.	TTE modeling	Random Forest evaluated by CV.
ProsperousCat	Preprocessing: Log transformation was performed and categorical variable were transformed to binary.	Binary + TTE modeling	Ensemble of Cox proportional hazard model, Gradient boosting and SVM evaluated by CV.
Mistral	Preprocessing: Redundant variables were removed. Imputation: Multiple imputation utilizing expectation-maximization with bootstrapping (EMB) algorithm.	Binary + Regression	Ensemble of Elastic Net and Random Forest (Kuhn's technique).
CAMP	Preprocessing: Lab values were standardized and Log transformations were performed. Imputation: Study-specific Random survival forest imputation.	Binary	Ensemble of Gradient Boosting and Random Forest evaluated by CV.
Alvin	Preprocessing: Linear and log transformation were performed. Imputation: Random Forest was utilized to predict the missing values.	Not Reported	Ensemble of Random Forest and Lasso classifier.
M S	Preprocessing: Largely missed variables were removed.	Binary	Random Forest evaluated by CV.
A Elangovan	Not Reported	Not Reported	Random Forest.
Team Simon	Preprocessing: Largely missed and redundant variable were removed. Log transformation was performed. Imputation: Mode.	TTE modeling	General Boosting Machines

Jing Lu	Preprocessing: YeoJohnson transformation was performed. Imputation: Median.	Binary	Ensemble of GLM, SVM, Random Forest and Neural Network evaluated by CV.
yoda	Preprocessing: Largely missed and redundant variables were removed. Imputation: First Study-specific KNN imputation and then whole data median/mean imputation.	Binary	Random Forest evaluated by CV.
Trishna	Preprocessing: Largely missed variables were removed. Imputation: Multiple imputation using Fully Conditional Specification (FCS) using MICE algorithm.	Binary	Ensemble of SVM and linear regression.
Wind	Preprocessing: Largely missed variables and patients with large proportion of missing features were removed. Continuous variables were normalized. Imputation: KNN method.	Binary	Random Forest evaluated by CV.
FIMM-UTU	Preprocessing: Redundant variables were removed. Skewed variables were truncated and Log transformation was performed. Imputation: Model based structural imputation.	Binary	Ensemble of penalized regression models evaluated by CV.
UoB_Prostate	Not Reported	Binary + TTE modeling	Ensemble of Random Forest and Cox proportional hazard model evaluated by CV.
The Data Wizard	Imputation: Blood tests were imputed with normal ranges and continuous features were imputed using entropy-based method.	Binary	Ensemble of Random Forest and Gradient Boosting evaluated by CV.
4. Cox Regression			
Junmei Wang	Imputation: Mean.	TTE modeling	Cox proportional hazard model
TeamX	Preprocessing: Largely missed and redundant variable were removed. Imputation: Median.	TTE modeling	Cox proportional hazard model
RUBME6	Preprocessing: Largely missed variables and Ascent2 dataset were removed.	TTE modeling	Cox proportional hazard model evaluated by CV.
brainstorm	Preprocessing: Largely missed variables were removed and performed log transformation. Imputation: Multiple imputation using Fully Conditional Specification (FCS) using MICE algorithm.	TTE modeling	Cox regression model evaluated by CV.
forPro	Preprocessing: Largely missed, redundant and variables absent in the validation dataset were removed.	TTE modeling	Cox proportional hazard model.
5. Support Vector Machines (SVM)			
Marat Kazanov	Not Reported	Binary	SVM regression model

* No reports were received from the following teams: Clinical Persona, DreamOn, UNC-BIAS, qiuyulian1994, and Y P

* Time-to-event (TTE)

Supplementary Table S4. Full scoring results from the Challenge evaluated using the ENTHUSE 33 trial data. Teams are listed with the links to their predictions, methods write-up, and code.

Team	Subchallenge 2 Predictions	Method Write-up & Code	AUPRC	Bayes factor	P-value	Adjusted P-value
Yuanfang Guan (Y G)	syn7152465	syn7152438	0.178	1.000	0.004	0.071
TYTDreamChallenge	syn4733449	syn4228911	0.175	1.036	0.007	0.071
PC LEARN	syn4730743	syn3822697	0.173	1.358	0.008	0.071
JayHawks	syn4732927	syn4731624	0.172	1.361	0.008	0.071
Brigham Young University	syn4733287	syn4382527	0.165	1.588	0.014	0.080
jls	syn4732940	syn4732827	0.165	1.589	0.014	0.080
A Bavarian dream	syn4733431	syn4599473	0.164	1.369	0.017	0.082
orion	syn4733495	syn4732967	0.151	2.544	0.039	0.163
Clinical Persona	syn4732636	syn4689890	0.151	2.062	0.045	0.163
Marat Kazanov	syn4730732	syn4730567	0.146	2.674	0.060	0.163
Zhang Chihao	syn4750004	syn4259433	0.145	4.252	0.057	0.163
Junmei Wang	syn4732889	syn4225820	0.144	2.546	0.061	0.163
ProsperousCat	syn4732921	syn4228681	0.144	2.805	0.064	0.163
TeamX	syn4732960	syn4732218	0.142	10.574	0.067	0.163
Mistral	syn4634543	syn4622030	0.142	4.995	0.072	0.163
DreamOn	syn4731549	syn4228561	0.141	3.748	0.077	0.163
CAMP	syn4751083	syn3647478	0.138	5.944	0.097	0.194
Alvin	syn4732818	syn4229406	0.137	3.448	0.105	0.195
Motoki Shiga	syn4730602	syn4229266	0.136	3.310	0.126	0.214
A Elangovan	syn4732108	syn4212102	0.135	4.291	0.109	0.195
Team Simon	syn4763388	syn4732901	0.131	8.074	0.154	0.250
UNC-BIAS	syn4744568	syn4744560	0.128	4.807	0.187	0.289
Jing Lu	syn4732948	syn4674923	0.126	4.688	0.211	0.311
yoda	syn4733454	syn4601848	0.122	10.628	0.264	0.374
Trishna	syn4730579	syn4730570	0.119	7.818	0.314	0.421
RUBME6	syn4731671	syn4590933	0.118	7.361	0.326	0.421
brainstorm	syn4730044	syn4730587	0.118	8.058	0.335	0.421
Wind	syn4731649	syn4731645	0.118	32.784	0.348	0.423
FIMM-UTU	syn4733268	syn4227610	0.115	7.460	0.405	0.461
UoB_Prostate	syn4733781	syn4591879	0.114	7.881	0.407	0.461
The Data Wizard	syn4733282	syn4228992	0.109	39.984	0.515	0.565
qiuyulian1994	syn4733251	syn4732205	0.105	18.763	0.638	0.678
Y P	syn4732915	syn4732909	0.103	21.026	0.692	0.713
forPro	syn4707765	syn4707464	0.088	237.06	0.974	0.974

Supplementary Table S5. Results from the Lift chart analysis of Challenge submissions evaluated using the ENTHUSE 33 trial data.

Team	Area Under the LR curve	LR among patients with the highest predicted risk		
		Top 5%	Top 10%	Top 20%
Yuanfang Guan (Y G)	1.404	2.450	2.041	1.633
TYTDreamChallenge	1.380	2.042	2.530	1.633
PC LEARN	1.356	2.361	1.866	1.914
JayHawks	1.302	1.633	1.713	1.225
Brigham Young University	1.368	2.456	1.838	1.626
jls	1.287	2.450	1.633	1.531
A Bavarian dream	1.323	2.042	2.245	1.531
orion	1.227	1.225	1.134	1.633
Clinical Persona	1.267	2.450	2.041	1.531
Marat Kazanov	1.258	2.042	1.837	1.531
Zhang Chihao	1.269	1.633	1.713	1.510
Junmei Wang	1.261	2.042	1.837	1.327
ProsperousCat	1.232	2.042	2.245	1.429
TeamX	1.246	1.915	2.041	1.429
Mistral	1.175	1.633	1.225	1.122
DreamOn	1.235	1.507	1.429	1.633
CAMP	1.220	1.225	1.633	1.510
Alvin	1.165	2.450	1.837	1.531
Motoki Shiga	1.158	1.225	1.305	1.122
A Elangovan	1.206	0.817	1.839	1.327
Team Simon	1.136	2.858	1.633	1.020
UNC-BIAS	1.149	0.817	1.021	1.429
Jing Lu	1.137	1.633	1.225	1.122
yoda	1.111	1.226	1.225	1.225
Trishna	1.071	0.817	0.816	1.122
RUBME6	1.063	2.042	1.432	0.903
brainstorm	1.064	1.507	1.633	1.225
Wind	1.068	1.225	1.225	1.225
FIMM-UTU	1.017	1.225	0.612	1.021
UoB_Prostate	1.034	1.225	1.429	1.122
The Data Wizard	0.995	0.817	0.816	1.020
qiuyulian1994	0.952	1.633	1.021	0.918
Y P	0.936	0.817	1.225	1.020
forPro	0.767	1.099	0.612	0.510

* Abbreviations: Lift ratio (LR)

Supplementary Table S6. Data dictionary and clinical feature definitions across the ASCENT2, MAINSAIL, VENICE, and ENTHUSE 33 trials.

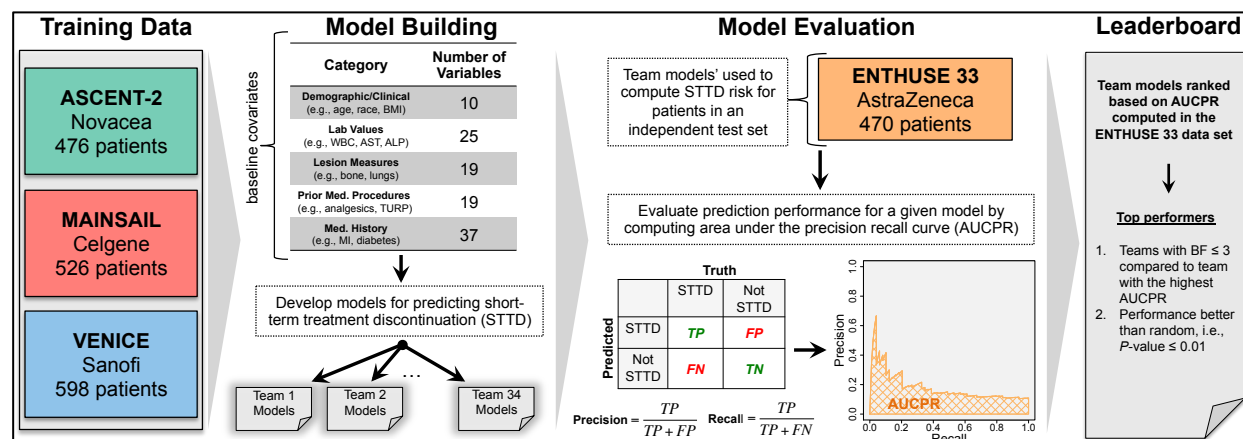
Variable Name	Definition
DOMAIN	Domain (table) Name Abbreviation
STUDYID	Study Identifier
RPT	Patient ID (or dummy id)
LKADT_P	Last Known Alive Dt Period in days
DEATH	Patient Died Flag
DISCONT	Discontinue Flag
ENDTRS_C	Discontinue Trt Reason (category)
ENTRT_PC	Discontinue Trt Date Period in days
PER_REF	Reference Day used for ENTRT_PC
LKADT_REF	Reference Day used for LKADT_P
LKADT_PER	Period unit for period values, LKADT_P
GLEAS_DX	Gleason Score at Initial diagnosis, see detail definition in wiki data description
TSTAG_DX	Primary Tumor-stage Score at Initial diagnosis, see detail definition in wiki data description
AGEGRP	Age Group
AGEGRP2	Age Group (3 category)
RACE_C	Race
BMI	Baseline Body Mass Index (kg/m ²)
HEIGHTBL	Baseline Height in cm
HGTBLCAT	Baseline Height in cm category
WEIGHTBL	Baseline Weight in kg
WGTBLCAT	Baseline Weight in kg category
REGION_C	Region of the World
SMOKE	Ever Smoked
SMOKFREQ	Smoking Frequency
SMOKSTAT	Current Smoking Status
ECOG_C	Baseline Patient Performance Status
TRT1_ID	Treatment 1 product
TRT2_ID	Treatment 2 product (docetaxel)
TRT3_ID	Treatment 3 product (prednisone except AZ)
ALP	BASELINE LAB VALUE: ALKALINE PHOSPHATASE U/L
ALT	BASELINE LAB VALUE: ALANINE TRANSAMINASE U/L
AST	BASELINE LAB VALUE: ASPARTATE AMINOTRANSFERASE U/L
CA	BASELINE LAB VALUE: CALCIUM MMOL/L
CREAT	BASELINE LAB VALUE: CREATININE UMOL/L
HB	BASELINE LAB VALUE: HEMOGLOBIN G/DL
LDH	BASELINE LAB VALUE: LACTATE DEHYDROGENASE U/L
NEU	BASELINE LAB VALUE: NEUTROPHILS 10 ⁹ /L
PLT	BASELINE LAB VALUE: PLATELET COUNT 10 ⁹ /L
PSA	BASELINE LAB VALUE: PROSTATE SPECIFIC ANTIGEN NG/ML
TBILI	BASELINE LAB VALUE: TOTAL BILIRUBIN UMOL/L

TESTO	BASELINE LAB VALUE: TESTOSTERONE NMOL/L
WBC	BASELINE LAB VALUE: WHITE BLOOD CELLS 10 ⁹ /L
CREACL	BASELINE LAB VALUE: CREATININE CLEARANCE ML/MIN
NA	BASELINE LAB VALUE: SODIUM MMOL/L
MG	BASELINE LAB VALUE: MAGNESIUM MMOL/L
PHOS	BASELINE LAB VALUE: PHOSPHORUS MMOL/L
ALB	BASELINE LAB VALUE: ALBUMIN G/L
TPRO	BASELINE LAB VALUE: TOTAL PROTEIN G/L
RBC	BASELINE LAB VALUE: RED BLOOD CELLS 10 ¹² /L
LYM	BASELINE LAB VALUE: LYMPHOCYTES 10 ⁹ /L
BUN	BASELINE LAB VALUE: BLOOD UREA NITROGEN MMOL/L
CCRC	BASELINE LAB VALUE: CALCULATED CREATININE CLEARANCE ML/MIN
GLU	BASELINE LAB VALUE: GLUCOSE MMOL/L
CREACLCA	BASELINE LAB VALUE: CREATININE CLEARANCE CALCUL. (COCKCROFT AND GAULT) ML/MIN
NON_TARGET	Baseline Non-Target Lesion(s), target vs. non-target lesion definition see wiki data description
TARGET	Baseline Target Lesion(s), target vs. non-target lesion definition see wiki data description
BONE	Baseline Bone Lesion(s)
RECTAL	Baseline Rectal Lesion(s)
LYMPH_NODES	Baseline Lymph Node Lesion(s)
KIDNEYS	Baseline Kidney Lesion(s)
LUNGS	Baseline Lung Lesion(s)
LIVER	Baseline Liver Lesion(s)
PLEURA	Baseline Pleura Lesion(s)
OTHER	Baseline Other Lesion(s)
PROSTATE	Baseline Prostate Lesion(s)
ADRENAL	Baseline Adrenal Lesion(s)
BLADDER	Baseline Bladder Lesion(s)
PERITONEUM	Baseline Peritoneum Lesion(s)
COLON	Baseline Colon Lesion(s)
HEAD_AND_NECK	Baseline Head and Neck Lesion(s)
SOFT_TISSUE	Baseline Soft Tissue Lesion(s)
STOMACH	Baseline Stomach Lesion(s)
PANCREAS	Baseline Pancreas Lesion(s)
THYROID	Baseline Thyroid Lesion(s)
ABDOMINAL	Baseline Abdominal Lesion(s)
ORCHIDECTOMY	Prior Orchiectomy(includes bilateral)
PROSTATECTOMY	Prior Prostatectomy
TURP	Prior Turp
LYMPHADENECTOMY	Prior Bilateral Lymphadenectomy
SPINAL_CORD_SURGERY	Prior Spinal Cord Surgery
BILATERAL_ORCHIDECTOMY	Prior Bilateral Orchiectomy
PRIOR_RADIOOTHERAPY	Prior Radiotherapy
ANALGESICS	Prior analgesics

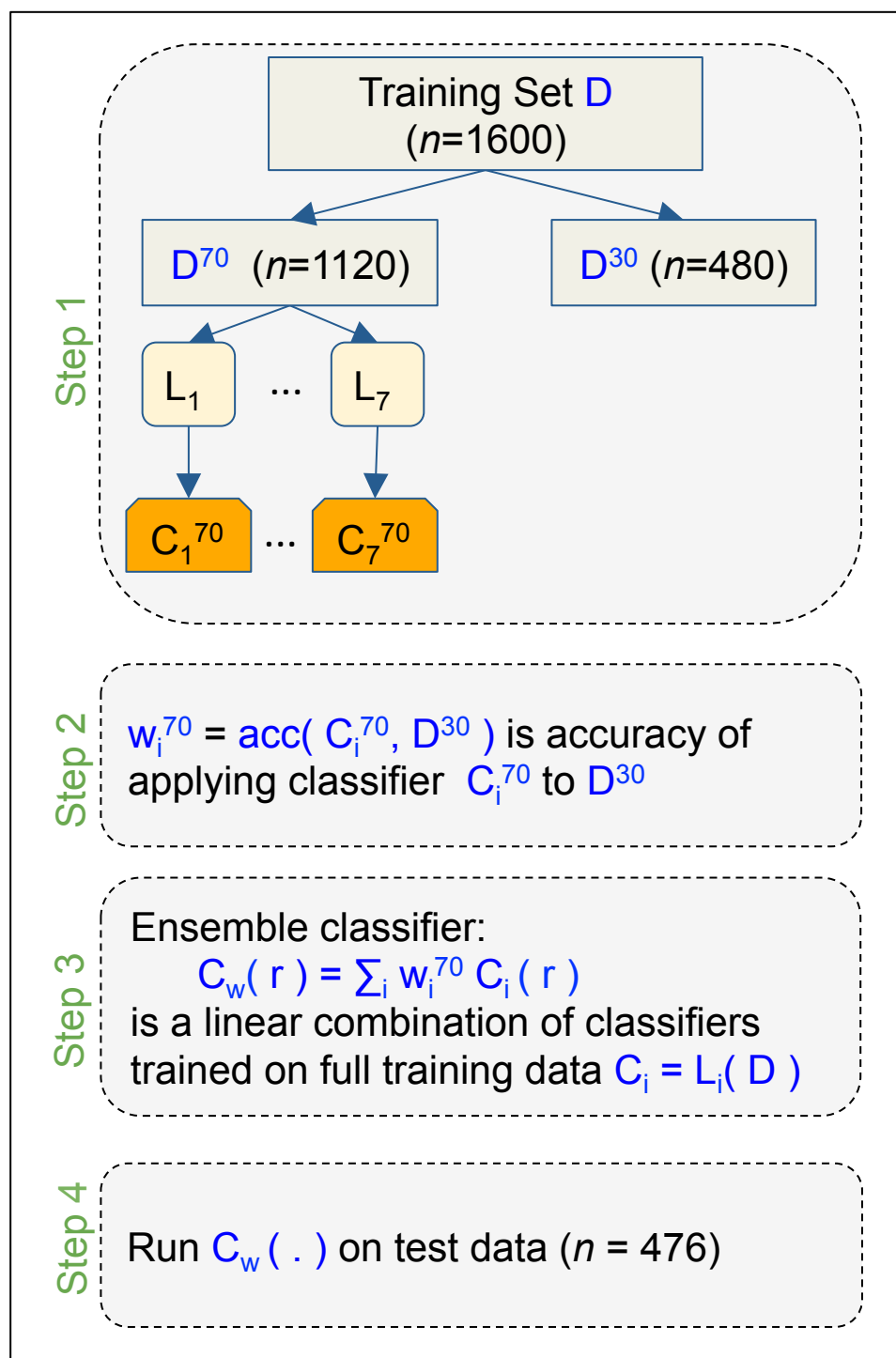
ANTI_ANDROGENS	Prior Anti-Androgens
GLUCOCORTICOID	Prior Glucocorticoids
GONADOTROPIN	Prior Gomadotropin
BISPHOSPHONATE	Prior Bisphosponate
CORTICOSTEROID	Prior Corticosteroid
IMIDAZOLE	Prior Imidazole
ACE_INHIBITORS	Prior ACE Inhibitors
BETA_BLOCKING	Prior Beta Blocking Agents
HMG_COA_REDUCT	Prior HMG COA Reductase Inhibitors
ESTROGENS	Prior Estrogens
ANTI_ESTROGENS	Prior Anti-Estrogens
ARTTHROM	MEDICAL HISTORY: ARTERIAL THROMBOSIS
CEREBACC	MEDICAL HISTORY: CEREBROVASCULAR ACCIDENT (HEMORRHAGIC AND/OR ISCHEMIC)
CHF	MEDICAL HISTORY: CONGESTIVE HEART FAILURE
DVT	MEDICAL HISTORY: DEEP VENOUS THROMBOSIS (DVT)
DIAB	MEDICAL HISTORY: DIABETES
GASTREFL	MEDICAL HISTORY: GASTROESOPHAGEAL REFLUX DISEASE (GERD)
GIBLEED	MEDICAL HISTORY: GASTROINTESTINAL (GI) BLEED
MI	MEDICAL HISTORY: MYOCARDIAL INFARCTION (MI)
PUD	MEDICAL HISTORY: PEPTIC ULCER DISEASE (PUD)
PULMEMB	MEDICAL HISTORY: PULMONARY EMBOLISM (PE)
PATHFRAC	MEDICAL HISTORY: PATHOLOGICAL BONE FRACTURES
SPINCOMP	MEDICAL HISTORY: SPINAL CORD COMPRESSION
COPD	MEDICAL HISTORY: CHRONIC OBSTRUCTIVE PULMONARY DISEASE
MHBLOOD	MEDICAL HISTORY (Body system): BLOOD & LYMPHATIC SYSTEM
MHCARD	MEDICAL HISTORY (Body system): CARDIAC DISORDERS
MHCONGEN	MEDICAL HISTORY (Body system): CONGENITAL, FAMILIAL & GENETIC
MHEAR	MEDICAL HISTORY (Body system): EAR & LABYRINTH
MHENDO	MEDICAL HISTORY (Body system): ENDOCRINE DISORDERS
MHEYE	MEDICAL HISTORY (Body system): EYE DISORDERS
MHGASTRO	MEDICAL HISTORY (Body system): GASTROINTESTINAL DISORDERS
MHGEN	MEDICAL HISTORY (Body system): GEN DISORD & ADMIN SITE
MHHEPATO	MEDICAL HISTORY (Body system): HEPATOBILIARY DISORDERS
MHIMMUNE	MEDICAL HISTORY (Body system): IMMUNE SYSTEM DISORDERS
MHINFECT	MEDICAL HISTORY (Body system): INFECTIONS & INFESTATIONS
MHINJURY	MEDICAL HISTORY (Body system): INJURY, POISON & PROCEDURAL
MHINVEST	MEDICAL HISTORY (Body system): INVESTIGATIONS
MHMETAB	MEDICAL HISTORY (Body system): METABOLISM & NUTRITION
MHMUSCLE	MEDICAL HISTORY (Body system): MUSC/SKELETAL & CONNECT TISSUE
MHNEOPLA	MEDICAL HISTORY (Body system): NEOPLASMS BENIGN, MALIG & UNSPEC
MHNERV	MEDICAL HISTORY (Body system): NERVOUS SYSTEM DISORDERS
MHPSYCH	MEDICAL HISTORY (Body system): PSYCHIATRIC DISORDERS
MHRENAL	MEDICAL HISTORY (Body system): RENAL & URINARY DISORDERS

MHRESP	MEDICAL HISTORY (Body system): RESP, THORACIC & MEDIASTINAL
MHSKIN	MEDICAL HISTORY (Body system): SKIN & SUBCUTANEOUS TISSUE
MHSOCIAL	MEDICAL HISTORY (Body system): SOCIAL CIRCUMSTANCES
MHSURG	MEDICAL HISTORY (Body system): SURGICAL & MEDICAL PROCEDURES
MHVASC	MEDICAL HISTORY (Body system): VASCULAR DISORDERS

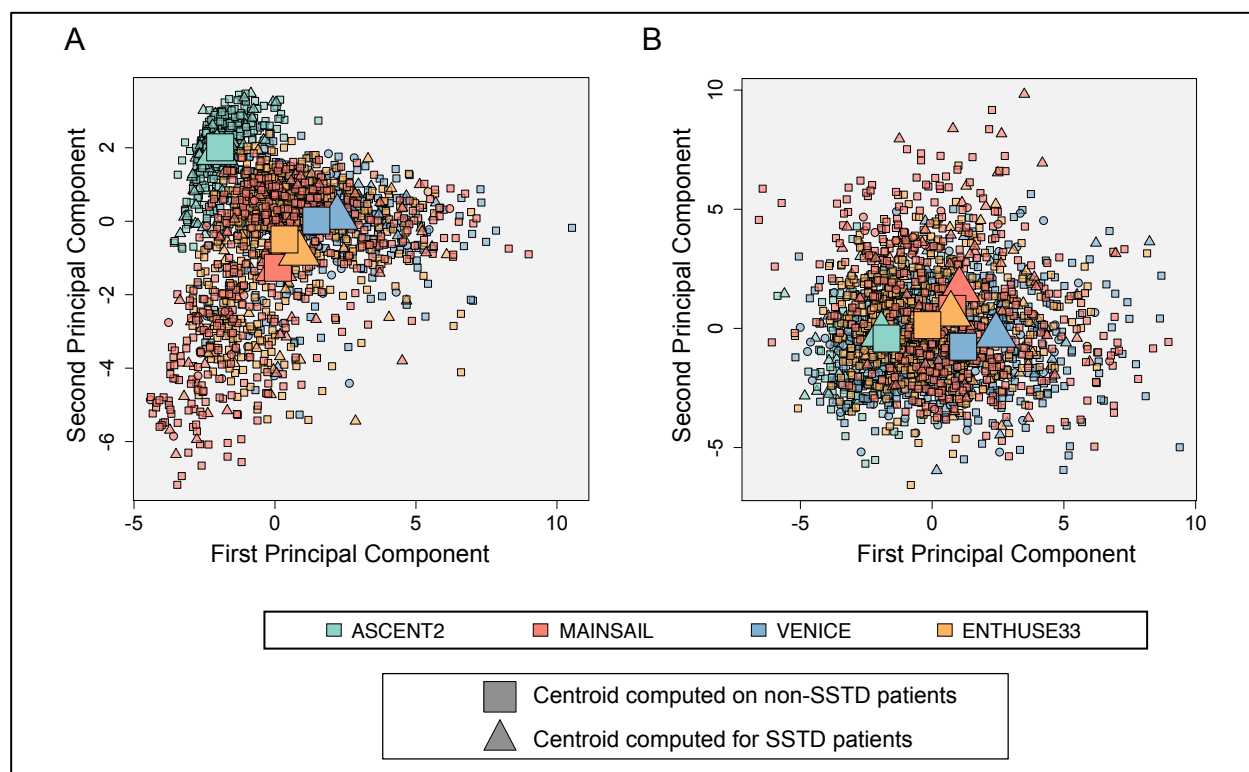
SUPPLEMENTARY FIGURES



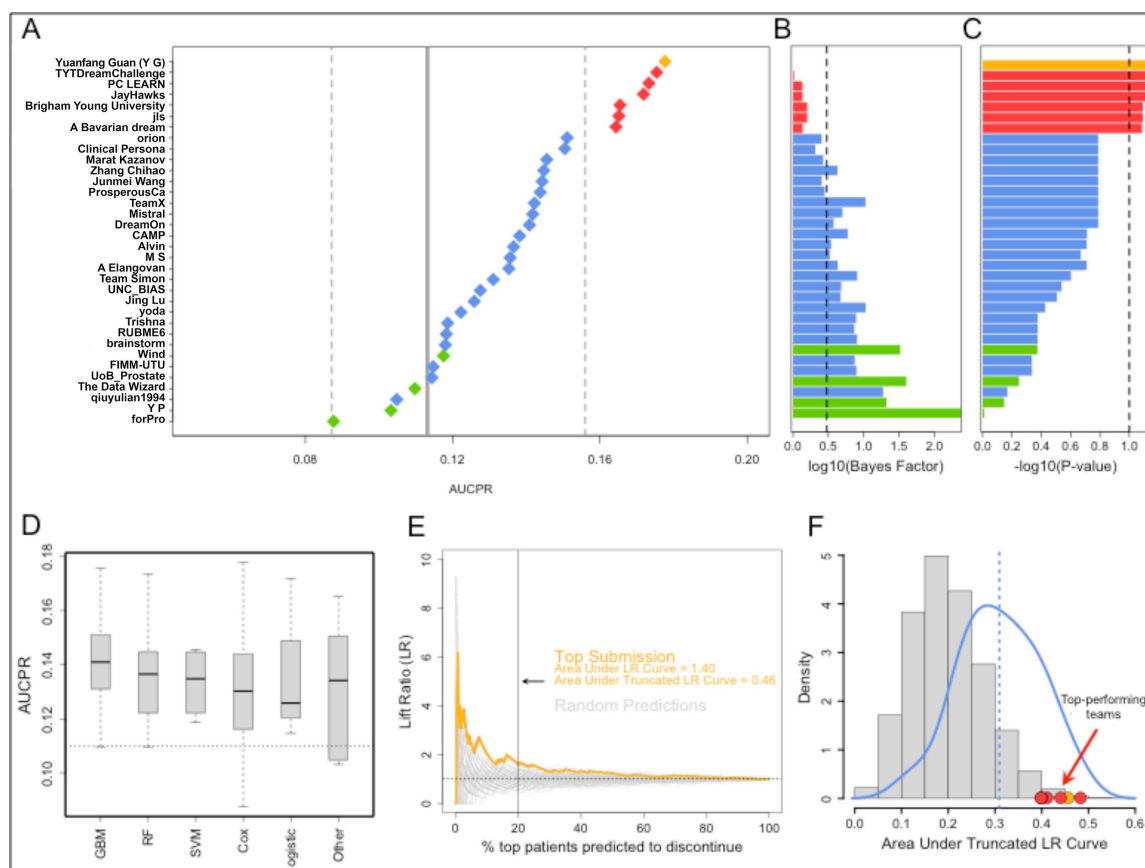
Supplementary Figure S1. Schematic diagram illustrating the various phases of the Challenge. The ASCENT2, MAINSAIL, and VENICE trial data were made available for method training. The ENTHUSE 33 trial data was used for evaluation.



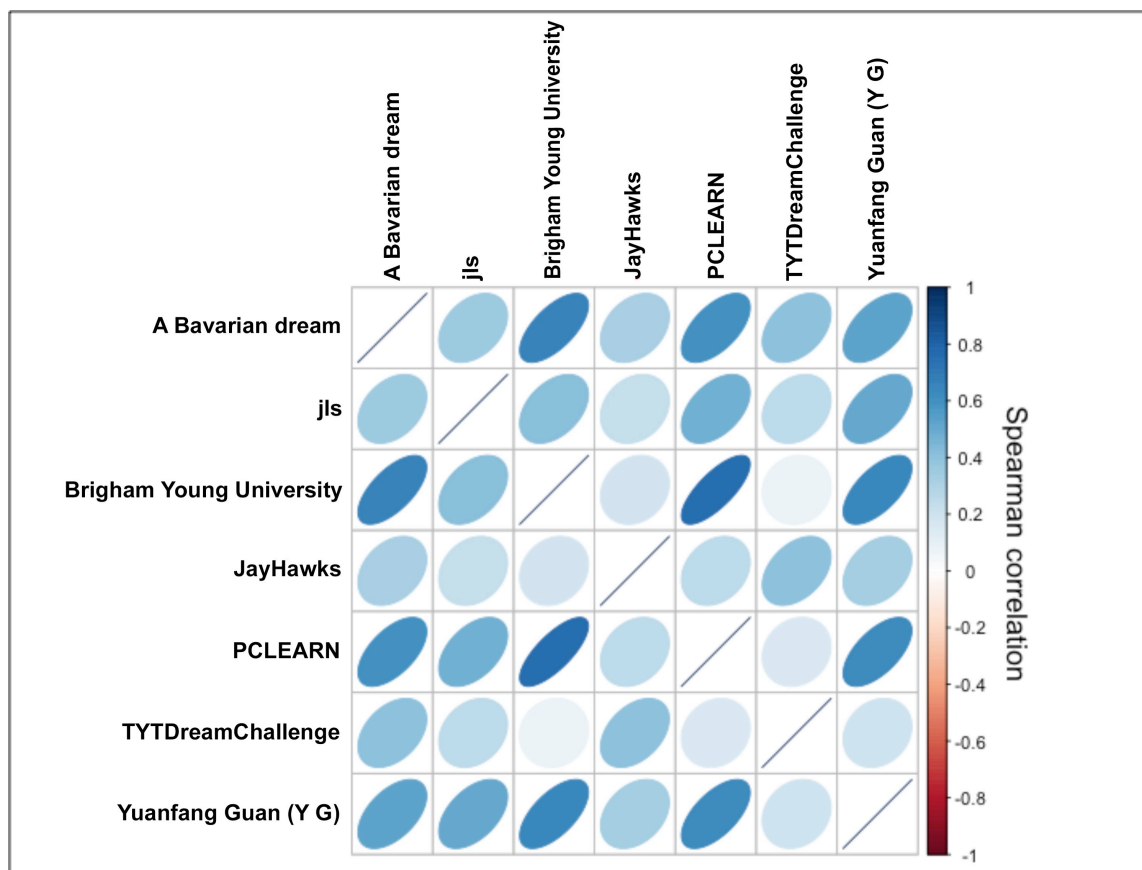
Supplementary Figure S2. Schematic diagram illustrating the various steps involved in the training, calibration, and evaluation of the post-Challenge ensemble-based prediction model.



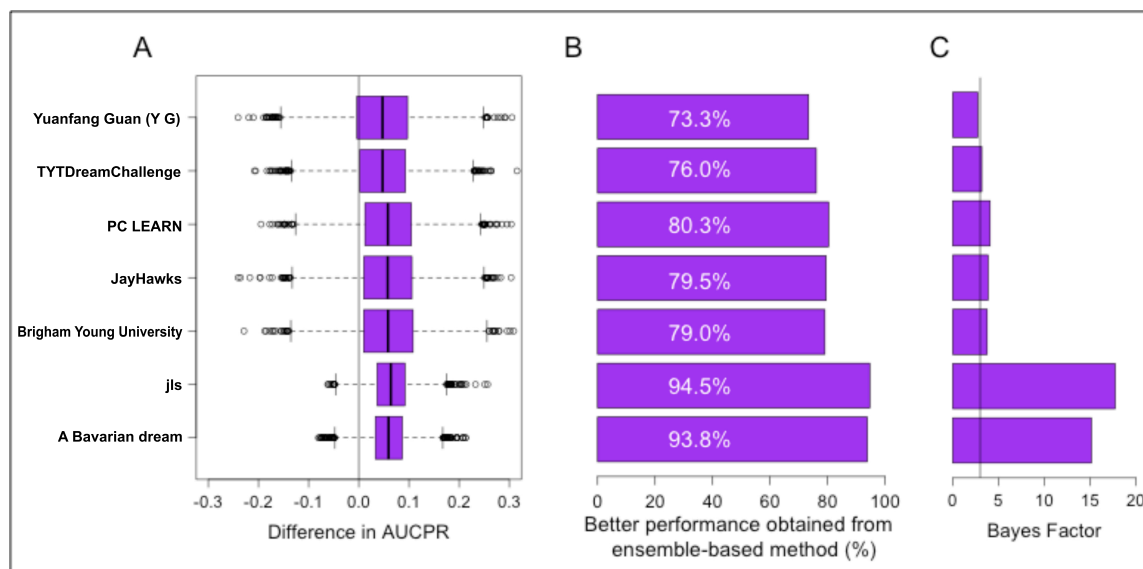
Supplementary Figure S3. Scatterplot of the first two principal components computed using (A) only binary features and (B) all 131 features in the Core Data Table for the ASCENT2, MAINSAIL, VENICE, and ENTHUSE 33 trials. Triangles represent patients that discontinued treatment due to an adverse event (AE) or possible AE within 3 months (STTD) and squares represent patients that either: completed treatment, discontinued treatment after 3 months, or died during follow up (non-STTD). Large square and triangle points represent study-specific centroids based on the first two principal components.



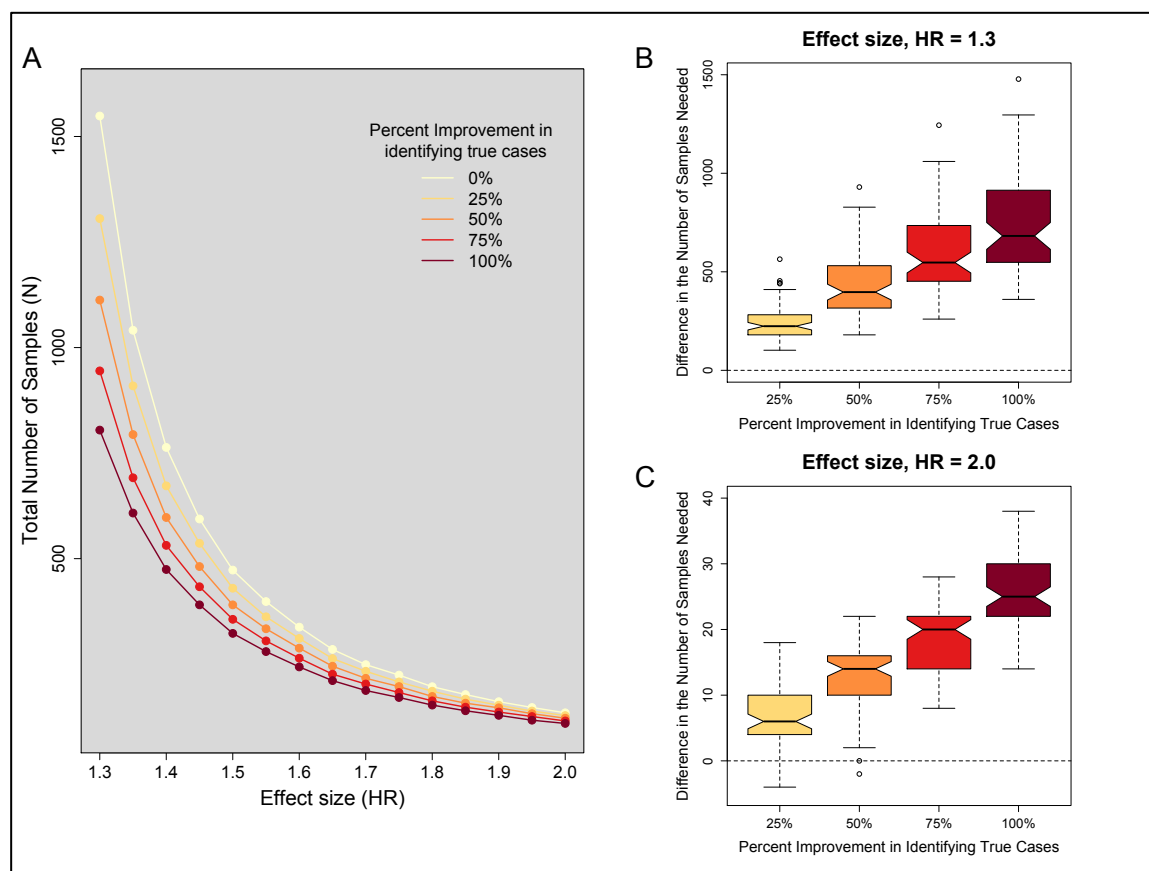
Supplementary Figure S4. Summary of Challenge results. (A) Leaderboard ranking of teams based on their prediction performance in the ENTHUSE 33 data set. Diamonds indicate the estimated area under the precision recall curve (AUPRC) for each team. The vertical solid line represents the AUPRC expected for a random prediction model and vertical dotted lines represent the upper and lower 95% bootstrap confidence intervals based on 10,000 bootstrap samples generated from a random prediction model. (B) $\log_{10}(\text{Bayes factor})$ computed for each team by comparing its AUPRC to the AUPRC from the top-submission. Vertical dotted line represents threshold used to delineate top-performing teams. (C) $-\log_{10}(P\text{-value})$ obtained by computing the likelihood of observing each team's AUPRC, or more extreme, under a random prediction model. Vertical dotted line represents threshold used to delineate top-performing teams. (D) Box-and-whisker plot of AUPRC as a function of the machine learning methodology used by each of the participating teams. (E) Lift-ratio (LR) curve for the top-submission with grey lines representing the LR-curves generated for 100 random prediction models. (F) Distribution of the area under the truncated LR curve at 20% based on random prediction models (grey), all participating teams (blue), and the top-performing teams in this Challenge (red and orange points).



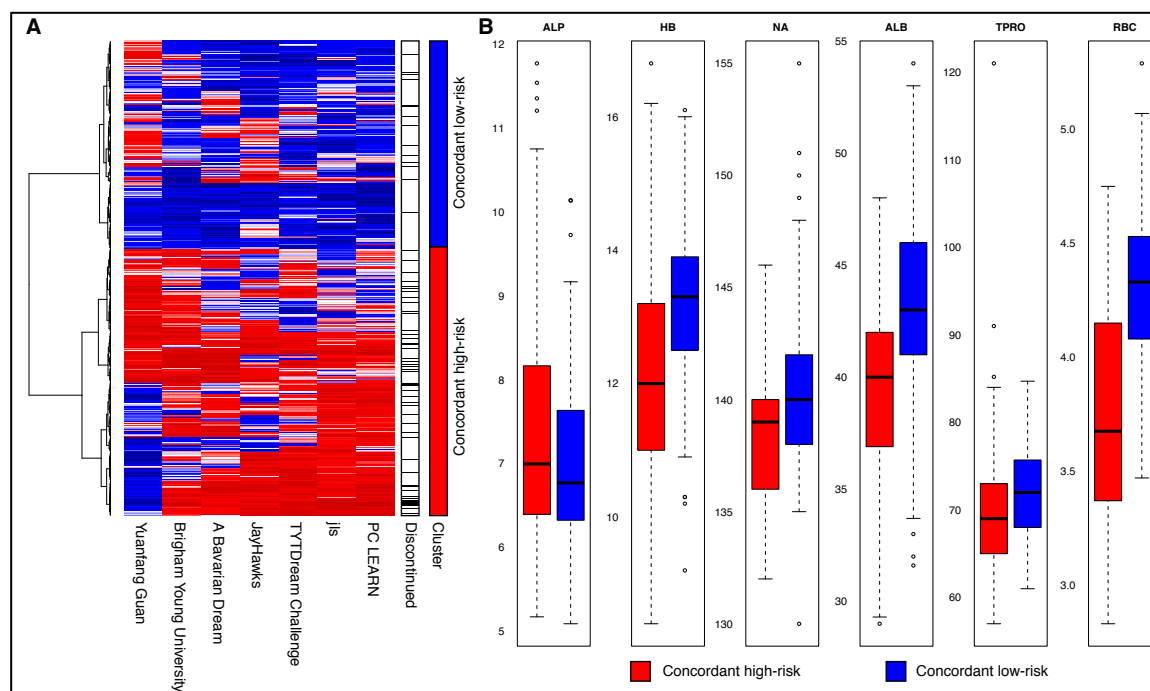
Supplementary Figure S5. Spearman correlation in the risk-scores (computed in the ENTHUSE 33 data set) submitted by the Challenge top-performers.



Supplementary Figure S6. Performance of the post-Challenge ensemble-based prediction model compared to the Challenge top-performers. (A) Box and whisker plot of the difference in AUPRC between the ensemble-based prediction model and each of the Challenge top-performers across 5,000 bootstrap samples of the ENTHUSE 33 data set. Values to the right of the vertical line at zero indicate that the ensemble-based prediction model achieved a higher AUPRC. (B) Percent of bootstrap samples where the ensemble-based prediction model achieved a higher AUPRC. (C) Bayes factor computed between each of the Challenge top-performers and the ensemble-based prediction model.

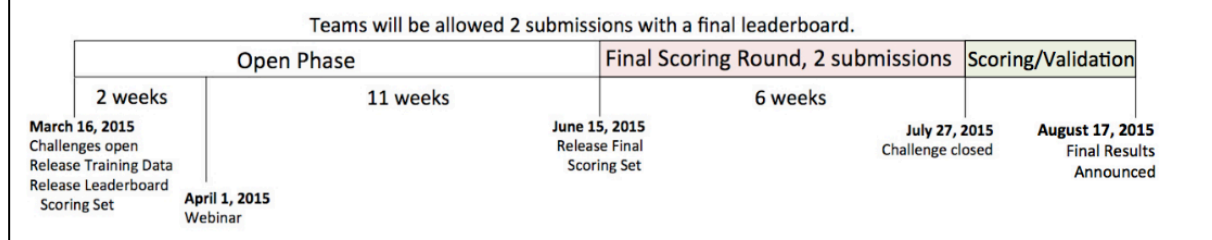


Supplementary Figure S7: Results from the clinical trial simulation analysis. (A) Mean sample size (across 100 simulations) as a function of the desired effect size (hazard ratio between treatment and control group) when prediction models of varying accuracy (0% – 100%) were employed to identify and exclude patients at high risk for early treatment discontinuation. A type 1 error-rate of 5% and statistical power of 80% were used to compute sample size. (B) Distribution (across 100 simulations) of the difference in the number of samples needed to detect a hazard ratio of 1.30 between treatment and control groups when no model was used to inform patient selection (i.e., 0% improvement in identifying true cases) and when the percent improvement in identifying true cases was 25%, 50%, 75%, and 100%. Points falling above the dotted line at zero indicate a reduction in the number of samples needed. (C) Distribution (across 100 simulations) of the difference in the number of samples needed to detect a hazard ratio of 2.00 between treatment and control groups when no model was used to inform patient selection (i.e., 0% improvement in identifying true cases) and when the percent improvement in identifying true cases was 25%, 50%, 75%, and 100%. Points falling above the dotted line at zero indicate a reduction in the number of samples needed



Supplementary Figure S8: Patient clustering and clinical variables for webtool. (A) Hierarchical clustering heat map of patients in the training data set ($n = 1,600$) based on their normalized ranked risk score, computed across the seven top performing teams. (B) Baseline clinical features that were identified as significantly different between the concordant high and low-risk groups in a subset of the training data, D^{30} ; Wilcoxon rank-sum tests were used to test continuous features and Fisher's exact tests were used for binary and categorical features.

Predict treatment discontinuation for metastatic, castration-resistant prostate cancer patients treated with docetaxel due to adverse events (AE) at early time points (<3 months).



Supplementary Figure S9. Timeline for the Challenge. Teams were allowed 2 submissions for the final scoring and 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. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 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. *The Lancet Oncology* 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 (MANSAIL): a randomised, double-blind, placebo-controlled phase 3 trial. *The Lancet Oncology* 2015;16:417-25.
4. Fizazi K, Higano CS, Nelson JB, et al. Phase III, randomized, placebo-controlled study of docetaxel in combination with zibotentan in patients with metastatic castration-resistant prostate cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2013;31:1740-7.
5. Davis JaG, M. The Relationship Between Precision-Recall and ROC Curves. *International Conference on Machine Learning*; 2006; Pittsburgh, PA.
6. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *J Roy Stat Soc B Met* 1995;57:289-300.
7. Sweeney CJ, Chen YH, Carducci M, et al. Chemohormonal Therapy in Metastatic Hormone-Sensitive Prostate Cancer. *The New England journal of medicine* 2015;373:737-46.
8. Kass RDR, A.E. Bayes factors. *Journal of the American Statistical Association* 1995;90:773-95.
9. Rokach L. Ensemble-based classifiers. *Artif Intell Rev* 2010;33:1-39.
10. Halabi S, Lin CY, Kelly WK, et al. Updated prognostic model for predicting overall survival in first-line chemotherapy for patients with metastatic castration-resistant prostate cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2014;32:671-7.

Prostate Cancer Challenge DREAM Community

Kald Abdallah⁸¹, Tero Aittokallio^{1,22}, Antti Airola²³, Catalina Anghel⁶, Helia Azima⁴⁵, Robert Baertsch³⁵, Pedro J Ballester^{39,77}, Chris Bare⁷⁷, Vinayak Bhandari⁷¹, Brian M Bot⁷⁷, Cuong C Dang^{39,77}, Maria Bekker-Nielsen Dunbar³⁴, Ann-Sophie Buchardt³⁴, Ljubomir Buturovic⁷⁶, Da Cao¹⁰, Prabhakar Chalise²⁸, Junwoo Cho²⁰, Tzu-Ming Chu³², R Yates Coley⁸, Sailesh Conjeti¹³, Sara Correia^{15,16}, James C Costello^{80,84}, Ziwei Dai²⁶, Junqiang Dai²⁸, Philip Dargatz³, Sam Delavarkhan⁴⁵, Detian Deng⁸, Ankur Dhanik²⁷, Yu Du⁸, Aparna Elangovan¹⁴, Shellie Ellis²⁹, Laura L Elo^{3,55}, Shadrielle M Espiritu⁷¹, Fan Fan⁷¹, Ashkan B Farshi⁴⁵, Ana Freitas¹⁶, Brooke Fridley²⁸, Christiane Fuchs^{1,4}, Eyal Gofer⁴³, Gopalacharyulu Peddinti²², Stefan Graw²⁸, Russ Greiner^{41,42}, Justin Guinney⁷⁷, Jing Guo^{5,65}, Pankaj Gupta¹³, Anna I Guyer¹², Jiawei Han⁴⁷, Niels R Hansen³⁴, Billy HW Chang⁴⁰, Outi Hirvonen⁵², Barbara Huang⁷¹, Chao Huang⁵⁷, Jinseub Hwang¹⁹, Joseph G Ibrahim⁵⁷, Vivek Jayaswal⁵⁰, Jouhyun Jeon⁶, Zhicheng Ji⁸, Deekshith Juvvadi³⁰, Sirkku Jyrkkio⁵², Kimberly Kanigel-Winner⁸⁰, Amin Katouzian¹³, Marat D Kazanov³⁷, Suleiman A Khan²², Shahin Khayyer⁴⁵, Dalho Kim²⁰, Agnieszka K Golińska⁵⁸, Devin Koestler²⁸, Fernanda Kokowicz¹⁷, Ivan Kondofersky^{1,4}, Norbert Krautenbacher^{1,4}, Damjan Krstajic^{75,76}, Luke Kumar⁴¹, Christoph Kurz², Matthew Kyan⁷³, Teemu D Laajala^{21,22}, Michael Laimighofer^{1,4}, Eunjee Lee⁵⁷, Wojciech Lesiński⁵⁸, Miao Zhu Li¹¹, Ye Li^{60,67}, Qiuyu Lian⁴⁴, Xiaotao Liang^{60,61}, Minseong Lim²⁰, Henry Lin⁴⁷, Xihui Lin⁶, Jing Lu³¹, Mehrad Mahmoudian⁵³, Roozbeh Manshaei⁴⁵, Richard Meier²⁸, Dejan Miljkovic¹³, Tuomas Mirtti^{22,24}, Krzysztof Mnich⁵⁹, Nassir Navab¹³, Elias C Neto⁷⁷, Yulia Newton³⁵, Thea Norman⁷⁷, Tapio Pahikkala²³, Subhabrata Pal⁵¹, Byeongju Park²⁰, Jaykumar Patel⁴¹, Swetabh Pathak³⁰, Alejandrina Pattin¹³, Donna P Ankerst⁴, Jian Peng⁴⁷, Anne H Petersen³⁴, Robin Philip³⁰, Stephen R Piccolo¹², Sebastian Pölsterl¹³, Aneta Polewko-Klim⁵⁸, Karthik Rao⁹, Xiang Ren⁴⁷, Miguel Rocha^{15,16}, Witold R. Rudnicki^{58,59,65}, Charles J Ryan⁷⁰, Hyunnam Ryu²⁰, Oliver Sartor⁶⁶, Hagen Scherb¹, Raghav Sehgal³⁰, Fatemeh Seyednasrollah^{53,55}, Jingbo Shang⁴⁷, Bin Shao²⁶, Liji Shen⁸³, Howard Sher⁸⁵, Motoki Shiga³⁶, Artem Sokolov³⁵, Julia F Söllner¹, Lei Song⁴⁸, Howard Soule⁶⁸, Gustavo Stolovitzky⁸², Josh Stuart³⁵, Ren Sun^{6,7}, Christopher J Sweeney⁶⁹, Nazanin Tahmasebi⁴¹, Kar-Tong Tan²⁵, Lisbeth Tomaziu³⁴, Joseph Usset²⁸, Yeeleng S Vang⁵⁶, Roberto Vega⁴¹, Vitor Vieira¹⁶, David Wang⁷¹, Difei Wang⁴⁹, Junmei Wang³³, Lichao Wang¹³, Sheng Wang⁴⁷, Tao Wang^{78,79}, Yue Wang⁵⁷, Russ Wolfinger³², Chris Wong³⁵, Zhenke Wu⁸, Jinfeng Xiao⁴⁶, Xiaohui Xie⁵⁶, Doris Xin⁴⁷, Hojin Yang⁵⁷, Nancy Yu⁶, Thomas Yu⁷⁷, Xiang Yu¹⁰, Sulmaz Zahedi^{72,74}, Massimiliano Zanin³⁸, Chihao Zhang⁶³, Jingwen Zhang⁵⁷, Shihua Zhang⁶³, Yanchun Zhang^{60,67}, Fang Liz Zhou⁸³, Hongtu Zhu⁵⁷, Shanfeng Zhu^{60,61,62} and Yuxin Zhu⁸

Affiliations

- 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 KOMMA 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, Florianopolis, 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 Computer Science, University of California Irvine, Irvine, CA, USA
- 57: Biostatistics and Imaging Analysis Lab, University of North Carolina at Chapel Hill, NC, USA
- 58: Faculty of Mathematics and Informatics, University of Bialystok, Poland
- 59: Computational Centre, University of Bialystok, Poland
- 60: School of Computer Science, Fudan University, Shanghai 200433, China
- 61: Shanghai Key Lab of Intelligent Information Processing, Fudan University, Shanghai 200433, China
- 62: Centre for Computational Systems Biology, Fudan University, Shanghai 200433, China
- 63: National Center for Mathematics and Interdisciplinary Sciences, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, 100190 Beijing, China
- 64: Research and development department, Annoroad Gene Technology Co. Ltd, Beijing, China
- 65: Interdisciplinary Centre for Mathematical and Computational Modelling, University of Warsaw, Poland
- 66: Tulane Cancer Center, Tulane University, New Orleans, LA, USA
- 67: Shanghai Key Lab of Data Science, Fudan University, Shanghai 200433, China
- 68: Prostate Cancer Foundation, Santa Monica, CA, USA
- 69: Department of Medical Oncology, Dana-Farber Cancer Institute and Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA
- 70: Genitourinary Medical Oncology Program, Division of Hematology & Oncology, University of California, San Francisco, CA, USA
- 71: Ontario Institute for Cancer Research, Toronto, Ontario, Canada
- 72: The Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, Canada
- 73: Electrical Engineering and Computer Science Dept., York University, Toronto, Canada
- 74: iBEST - Li Ka Shing Institute of Knowledge, St. Michael's Hospital, Toronto, Canada
- 75: Research Centre for Cheminformatics, Jasenova 7, 11030 Beograd, Serbia
- 76: Clinical Persona Inc, 932 Mouton Circle, East Palo Alto, CA, USA
- 77: Sage Bionetworks, Seattle, WA, USA
- 78: Quantitative Biomedical Research Center, Department of Clinical Sciences, University of Texas Southwestern Medical Center, Dallas, Texas, USA
- 79: Center for the Genetics of Host Defense, University of Texas Southwestern Medical Center, Dallas, Texas, USA
- 80: Department of Pharmacology & Computational Biosciences Program, University of Colorado, Anschutz Medical Campus, Aurora, CO, USA
- 81: AstraZeneca, Gaithersburg, MD, USA
- 82: IBM T.J. Watson Research Center, IBM, Yorktown Heights, NY, USA
- 83: Sanofi, Bridgewater, NJ, USA
- 84: University of Colorado Comprehensive Cancer Center, University of Colorado, Anschutz Medical Campus, Aurora, CO, USA
- 85: 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 C.F., 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 Doctoral Programme in Mathematics and Computer Sciences at the University of Turku and Sigrid Juselius Foundation supported F.S. 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.