Multivariable Regression

For multivariable survival modelling of baseline prognostic markers using the Cox proportional-hazards model (**Figure 1**), the following variables were included: tumour fraction (Log10-transformed), cell-free DNA concentration (cfDNA, ng/ml, Log10-transformed), Eastern Cooperative Oncology Group status (ECOG, 0 vs 1/2), presence of visceral metastases (No/Yes), bone metastases only (No/Yes), serum albumin (ALB, g/dl), haemoglobin (HB, g/dl), lactate dehydrogenase (LDH, U/l Log10-transformed), prostate specific antigen (PSA, ng/ml Log10-transformed), alkaline phosphatase (ALP, U/l Log10-transformed), neutrophil-lymphocyte ratio (NLR). In cases with multiple baseline samples (SCR and C1D1), the median value for continuous variables was used for concordance with previous prognostic studies. These variables were applied identically for overall survival (OS), radiographic progression-free survival (RPFS) and PSA progression-free survival (PSAPFS). These multivariable regressions were stratified by FIRSTANA/PROSELICA study inclusion, due to the underlying differences in baseline hazard observed between the two trials.

For logistic regression to study the association of LST score with prior Abiraterone/Enzalutamide status (Figure 3), the following variables were included: large-scale transition score (LST), tumour fraction (Log10-transformed), Eastern Cooperative Oncology Group status (ECOG, 0 vs 1/2), presence of visceral metastases (No/Yes), bone metastases only (No/Yes). The response variable was prior Abi/Enza treatment status, encoded as 0=treatment-naïve and 1=treatment-exposed.

For linear regression to study the association of several genomic loci with LST score, the copy number of these loci were used as input variables, encoded as numeric (-1=deletion, 0=neutral, 1=gain). Prior Abi/Enza treatment status (0=treatment-naïve and 1=treatment-exposed), tumour fraction (Log10-transformed), and study inclusion (FIRSTANA/PROSELICA clinical trial). The response variable was large-scale transition score (LST).

Copy-Number and Large-Scale Transition Scores

Following processing of lpWGS read-depth data by ichorCNA, several heuristics were applied to generate final CN calls. The assigned integer copy number of each segment was deducted from the sample ploidy value to remove spurious copy-gains due to variable ploidy states (common in mCRPC). Following this calculation, CN values of -1 were treated as hemizygous (shallow) deletions, <-1 were treated as homozygous (deep) deletions, values of >0 were defined as copy gains, and amplifications as >6. Deep deletions were re-

classified as shallow if the median log2-ratio of the segment was >-0.4. Per-sample LST values for lpWGS data were calculated by summing segment boundaries of adjacent segments >10Mb in size, omitting small gaps of <3Mb (20).

Elastic Net Regression

The following parameters were set for the Elastic Net Regression: 5-fold cross-validation, alpha values between 0 and 1 in increments of 0.05, 25 null-model permutations, and 100 cross-validated iterations. Input to the model was bin-level CN data, with values assigned to gain (1), neutral (0) and loss (-1), with LST scores (numeric) used as the response variable. Prior abiraterone/enzalutamide status was included, encoded as 0 (treatment-naïve) or 1 (treatment-exposed). Final model was chosen by highest quality function as recommended by the eNetXplorer package (21).