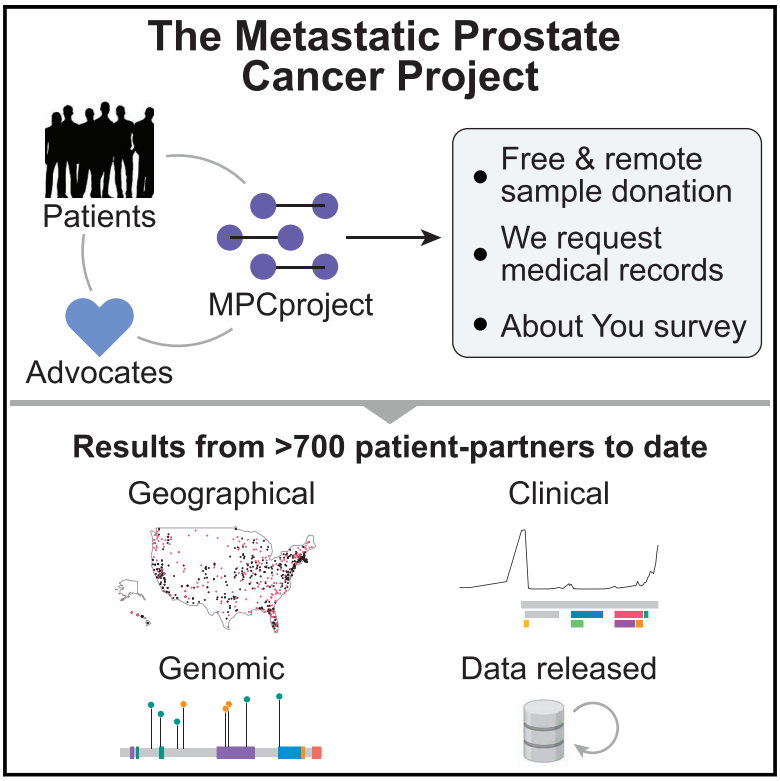
Cell Genomics

Article

A patient-driven clinicogenomic partnership for metastatic prostate cancer

Graphical abstract



Authors

Jett Crowdis, Sara Balch, Lauren Sterlin, ..., Corrie A. Painter, Nikhil Wagle, Eliezer M. Van Allen

Correspondence

eliezerm_vanallen@dfci.harvard.edu

In brief

Crowdis et al. describe the MPCproject (mpcproject.org), a decentralized initiative to partner with patients with metastatic prostate cancer in the US and Canada to accelerate molecular research. The authors describe clinicogenomic results from the first 706 geographically diverse patient partners and lay the foundation for sustained and inclusive partnership in this disease.

Highlights

- MPCproject partners with metastatic prostate cancer patients for molecular research
- Over 1,000 patient partners to date are from across the US and Canada
- 41% of patient partners are from rural or medically underserved areas
- Remotely donated samples from real-world settings recapitulate genomic findings

Crowdis et al., 2022, Cell Genomics *2*, 100169
September 14, 2022
https://doi.org/10.1016/j.xgen.2022.100169

CellPress

Cell Genomics

CellPress
OPEN ACCESS

Article

A patient-driven clinicogenomic partnership for metastatic prostate cancer

Jett Crowdis,[1,2,16] Sara Balch,[1,2,16] Lauren Sterlin,[2,3] Beena S. Thomas,[2,3] Sabrina Y. Camp,[1,2] Michael Dunphy,[2,3] Elana Anastasio,[2,3] Shahrayz Shah,[2,3] Alyssa L. Damon,[2,3] Rafael Ramos,[1,2,3] Delia M. Sosa,[2,3] Ilan K. Small,[2,3] Brett N. Tomson,[2,3] Colleen M. Nguyen,[2,3] Mary McGillicuddy,[2,3] Parker S. Chastain,[2,3] Meng Xiao He,[1,2,4] Alexander T.M. Cheung,[1,2,5] Stephanie Wankowicz,[6,7] Alok K. Tewari,[1] Dewey Kim,[1,2] Saud H. AlDubayan,[1,2,8] Ayanah Dowdye,[2,3] Benjamin Zola,[2,3] Joel Nowak,[9] Jan Manarite,[9] Idola Henry Gunn,[10] Bryce Olson,[11] Eric S. Lander,[12,13,14] Corrie A. Painter,[2,3] Nikhil Wagle,[1,2,3,17] and Eliezer M. Van Allen[1,2,15,17,18,*]

[1]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02115, USA
[2]Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA
[3]Count Me In, Cambridge, MA 02142, USA
[4]Harvard Graduate Program in Biophysics, Boston, MA 02115, USA
[5]NYU Grossman School of Medicine, New York, NY 10016, USA
[6]Biophysics Graduate Program, University of California San Francisco, San Francisco, CA 94143, USA
[7]Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco, CA 94143, USA
[8]College of Medicine, King Saud bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia
[9]Cancer ABCs, Brooklyn, NY 11201, USA
[10]Founder, MomVetWifeWidow.Org
[11]Global Strategist, Health and Life Sciences, Intel Corp, Patient Advocate, SequenceMe, Santa Clara, CA 95054, USA
[12]Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA
[13]Department of Biology, MIT, Cambridge, MA 02142, USA
[14]Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA
[15]Center for Cancer Genomics, Dana-Farber Cancer Institute, Boston, MA 02115, USA
[16]These authors contributed equally
[17]Senior author
[18]Lead contact
*Correspondence: eliezerm_vanallen@dfci.harvard.edu
https://doi.org/10.1016/j.xgen.2022.100169

SUMMARY

Molecular profiling studies have enabled discoveries for metastatic prostate cancer (MPC) but have predominantly occurred in academic medical institutions and involved non-representative patient populations. We established the Metastatic Prostate Cancer Project (MPCproject, mpcproject.org), a patient-partnered initiative to involve patients with MPC living anywhere in the US and Canada in molecular research. Here, we present results from our partnership with the first 706 MPCproject participants. While 41% of patient partners live in rural, physician-shortage, or medically underserved areas, the MPCproject has not yet achieved racial diversity, a disparity that demands new initiatives detailed herein. Among molecular data from 333 patient partners (572 samples), exome sequencing of 63 tumor and 19 cell-free DNA (cfDNA) samples recapitulated known findings in MPC, while inexpensive ultra-low-coverage sequencing of 318 cfDNA samples revealed clinically relevant *AR* amplifications. This study illustrates the power of a growing, longitudinal partnership with patients to generate a more representative understanding of MPC.

INTRODUCTION

Prostate cancer is the second most diagnosed cancer in men, with nearly 200,000 men diagnosed in 2020 alone in the US.[1] Survival rates for localized disease are high, but the 5-year survival rate for the over 300,000 men currently living with metastatic prostate cancer (MPC) is only 31%, representing the third leading cause of death for men.[1,2] Genomic sequencing studies have enabled new therapeutic targets for MPC, but obtaining large cohorts of tumor biopsies for molecular study has been difficult, as MPC often spreads to bone and requires technically challenging procedures to sample.[3–6] Because prostate cancer can shed cell-free DNA (cfDNA) into the bloodstream, blood biopsies that sample this circulating tumor DNA have proven to be a useful alternative for the study of MPC.[7,8]

Historically, quaternary care academic medical institutions have had the necessary infrastructure to lead clinically integrated MPC sequencing studies. However, the resulting clinical and genomic data is often siloed within these institutions, leading many to push for mandatory data sharing.[9,10] These efforts, while important, do not directly improve access to molecular research programs and do not address underlying ethnic,

socioeconomic, and geographic patient disparities in such studies, which threaten to bias findings and eventually care toward select patient populations.[11–14] Commercial sequencing options for prostate cancer are emerging but are often proprietary, only available with appropriate insurance, and regularly inaccessible for research use.[15–17] Indeed, despite growing interest from patients with MPC in clinical and research-based genomic sequencing, there are only limited mechanisms for these patients to partner with the research community to accelerate discoveries.[18–20]

We hypothesized that a patient-partnered framework that empowers patients with MPC to share their biological samples, clinical histories, and lived experiences directly with researchers regardless of geographic location or hospital affiliation would lead to new clinicogenomic discoveries and begin to address demographic inequities and data-access barriers in molecular studies for this disease. Thus, we established the Metastatic Prostate Cancer Project (MPCproject, mpcproject.org), a research model that leverages patient advocacy and social media to enable patients with MPC to participate in genomic research remotely at no personal cost.

RESULTS

Development of a patient-partnered MPC research model

Working with patients, loved ones, and advocates, we developed an MPCproject enrollment process for men living with MPC in the US and Canada (Figure 1A). The MPCproject outreach model is community centered and utilizes advocacy partnerships, social media campaigns, and educational initiatives to engage patients (Figure S1). To enroll, patient partners complete an online survey describing their experience with MPC, followed by signing electronic consent and release forms, which allow the MPCproject team to contact their hospitals to request medical records and optionally archival tumor tissue for research-grade genomic sequencing (Figure S2). Enrolled patient partners can also use a mailed kit to donate saliva and/or blood at routine blood draws at no cost, and these samples are sequenced to assess germline DNA and cfDNA, respectively (Figures S3 and S4).

Patient partners and advocates are involved in every step of the project's design and execution—we respond directly to their feedback and keep them informed of our progress and findings (supplemental information; Figure S5). Patient advocates help design the website and all patient-facing enrollment material, lead patient information sessions about the project, and advise the project's mission. We also work with patient partners who continue donating blood to help the research community understand the evolution of MPC, and we regularly release prepublication, deidentified genomic, patient-reported, and clinical data in public repositories for research use.

Partnering with a demographically distinct patient population

To date, the MPCproject has partnered with over 1,000 patients in the US and Canada and has orchestrated three public data releases (Figure 1B). The analyses presented here are based on the 706 men from the US and Canada who had enrolled (completed consent forms) as of June 1, 2020 (Figure S6).

Using patient-reported survey data, we assessed the geographical diversity of our patient partners. Hailing from 49 US states and six Canadian provinces, patient partners reported receiving care for their prostate cancer at over 1,000 distinct medical institutions, 91% of which were reported by two or fewer patients (Figure 1C). We found that 56% of patient partners have never received care at an NCI-designated cancer center, where genomic research is traditionally conducted (Table S1). These patient partners were three times less likely to report participating in a clinical trial (7% versus 20%, $p = 1 \times 10^{-6}$, Fisher's exact test).

We then used patient-reported data to identify residential census tracts and their geographic characteristics (n = 628/706 participants had identifiable census tracts; STAR Methods). We found that 13% of patient partners live in rural areas defined by the USDA, a proportion consistent with patients with MPC in the US (11%).[21] We additionally found that 30% of patient partners live in health-physician-shortage areas (HPSAs) and that 24% live in medically underserved areas (MUAs) as defined by the Health Resources and Services Administration (Figure 1D; STAR Methods).[22] These proportions could not be compared with patients with MPC in the US or with other sequencing efforts due to a lack of published data but are significantly enriched compared with the US population (25% HPSAs, 5% MUAs, $p = 0.03$ and 1×10^{-82}, respectively, Fisher's exact test).[23,24] While living in a rural area was associated with being in an MUA or HPSA, 28% of MPCproject patient partners live in urban primary care MUAs or HPSAs ($p = 5.7 \times 10^{-13}$, Fisher's exact test). We additionally found that patient partners living in rural areas compared with urban areas lived a median of 160 km farther from institutions where they reported receiving treatment, suggesting that they may travel farther for cancer care ($p < 10^{-11}$, Mann-Whitney U test; Figure S7).

We next examined the socioeconomic traits of patient partner residential areas using the national Area Deprivation Index (ADI), a 0–100 ranking that includes factors of income, education, employment, and housing quality, where 100 indicates the most disadvantage.[25] The average ADI of patient partner residential areas was lower than the age- and race-matched national average (31 versus 46), which may reflect the relative success of patient partner engagement via social media outreach, the usage of which is correlated with socioeconomic status, compared with our community-driven efforts to date (Figure S7).[26] Notably, we cannot compare this average with patient populations from existing sequencing studies due to a lack of published data. We also found that patient partners living in more disadvantaged areas were less likely to attend NCI cancer centers for treatment, even after controlling for rural, MUA, and HPSA status (ADI = 35 versus 27, NCI treated versus not, $p < 0.001$, logistic regression) (Figure 1E). We are cautious, however, in interpreting the results of these geographic analyses. Patient partners may not currently live in their reported locations, we do not directly survey their income or socioeconomic status, and their experiences may not be represented by their residential area. We did not observe significant associations in baseline clinical factors, therapies

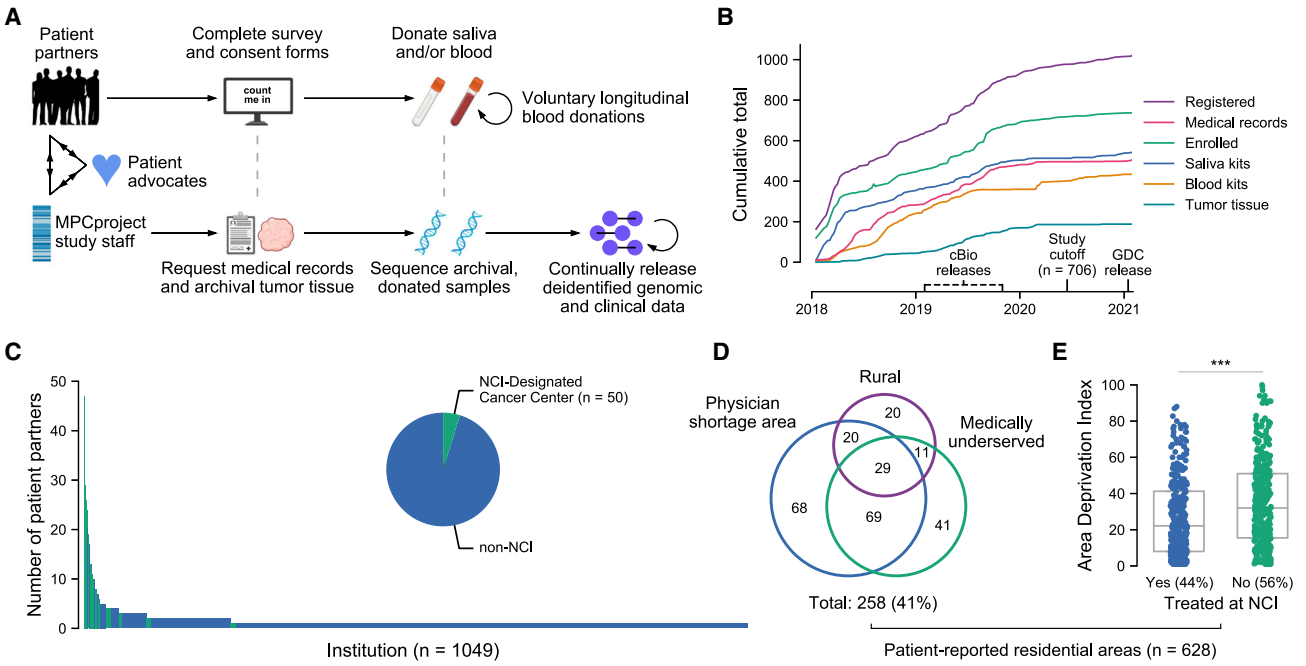


Figure 1. Partnering with diverse patients to enhance our understanding of metastatic prostate cancer

(A) Summary of MPCproject enrollment process. Patients learn about the project primarily through outreach and partnered advocacy groups. If they register, patient partners complete online intake, consent, and medical release forms, then can opt into donating saliva via a mailed kit and/or blood at routine blood draws at no charge. In parallel, MPCproject staff request medical records and archival tumor samples from patients' medical institutions, then abstract medical information from obtained records and sequence archival tumor tissue and/or donated blood and saliva (STAR Methods). Deidentified clinical, genomic, and patient-reported data are released on a continual, prepublication basis and deposited in public repositories.

(B) Enrollment statistics and timeline for the MPCproject. Depicted are the cumulative number of patients that began the registration process (registered), patients that completed the survey and consent forms (enrolled), patients with at least one medical record received (medical records), and blood kits, saliva kits, and archival tumor tissue received at the Broad Institute for sequencing (blood kits, saliva kits, and tumor tissue, respectively). 706 patient partners enrolled before "study cutoff," June 1, 2020, and are included in this study's analyses. cBioPortal (cbioportal.org) releases include summary abstracted medical, genomic, and patient-reported data; Genomic Data Commons (GDC) releases include raw sequencing files and demographic data.

(C) Represented medical institutions among patient partners living in the US and Canada. Shown are the 1,049 unique institutions (x axis) where patient partners report receiving care for their prostate cancer, with the number of distinct patient partners at each institution (y axis). NCI-designated cancer centers are shown in green. Patient partners that did not complete this survey question (n = 36) and institutions outside the US and Canada (n = 56) are not shown.

(D) Access to medical care among patient partners living in the US. Patient-reported data were used to identify residential census tracts that were overlapped with primary care health-physician-shortage areas (HPSAs), medically underserved population/areas (MUAs), and rural areas obtained from the Health Resources and Services Administration and US Census. Patient partners that live in Canada (n = 30) who did not provide residential data (n = 40) or who provided only a P.O. box (n = 8) are not shown.

(E) Patient partners living in more disadvantaged areas are less likely to attend NCI cancer centers. The Area Deprivation Index, a metric that assesses neighborhood disadvantage, was assessed for each residential census block group. Higher values indicate more disadvantage. The x axis reflects whether patient partners reported receiving care at an NCI-designated cancer center. *** $p < 0.001$ in a logistic regression model that adjusts for rural, MUA, and HPSA status.

received, or likelihood to participate in a clinical trial with ADI or across patient partners in rural areas, MUAs, or HPSAs.

The combination of the MPCproject's online enrollment and patient-centered outreach through advocacy partnerships enabled the creation of a geographically distinct prostate cancer research program. Despite the project's geographical diversity, however, fewer than 10% of patient partners self-identify as non-White (Table S2). While similar to existing studies, this representation remains well below the proportion of minority patients with prostate cancer generally (20%).[21] The lack of racial diversity in our study is a critical flaw that is thus far insufficient to accelerate research for communities of color, and it has spurred new, community-driven MPCproject initiatives to connect with these patients, as detailed in the limitations of the study.

Patient-reported data augment medical records to amplify patient stories

Through the patient-reported data, we sought to understand the real-world experiences of those living with MPC. 45% of patient partners report being diagnosed with *de novo* metastatic disease, with bone (48%) and lymph node (39%) lesions as the most common metastatic sites (Figures 2A and 2B). 48% of patient partners reported a family history of prostate or breast cancer, while 24% reported having at least one other cancer diagnosis in their lifetime, 30% of which was a non-skin form of cancer (Figures 2C and 2D). The average age at diagnosis was significantly younger than the national average (61 versus 65 years old, $p < 10^{-39}$, t-test), and 24% of participants were diagnosed with early-onset prostate cancer ($\leq$55 years at diagnosis; Table S2).[27] We note that these characteristics of

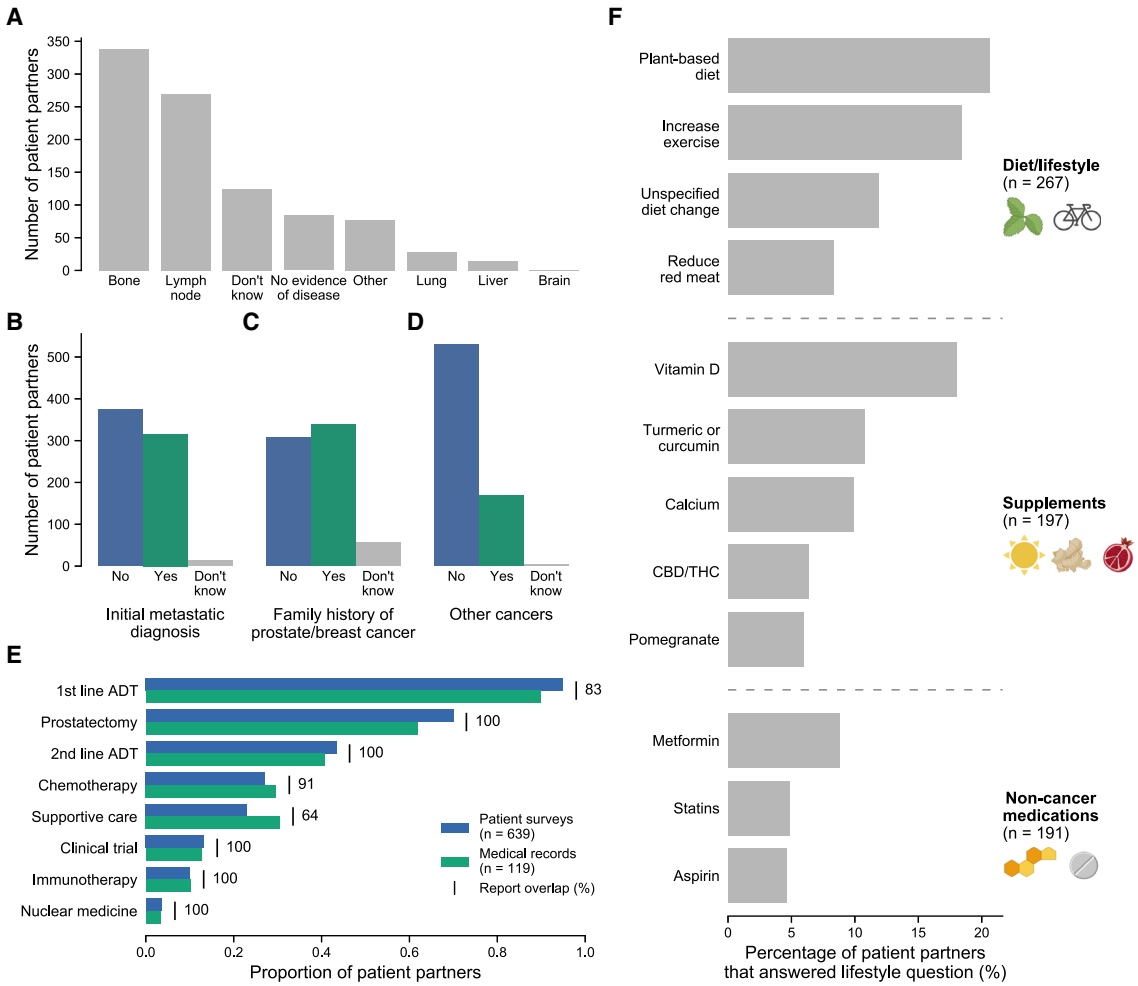


Figure 2. Patient voices reveal the landscape of living with metastatic prostate cancer

(A–D) Self-reported data of 706 patient partners related to their prostate cancer.

(A) Patient partners were asked for the current location of their cancer. Participants were free to choose multiple if their cancer had metastasized to multiple locations.

(B–D) Responses were tabulated from questions asking patient partners if their initial prostate cancer diagnosis was metastatic (B), if they have a family history of prostate/breast cancer (C), or if they have ever had another cancer diagnosis (D). Patient partners who did not complete these questions ($n < 5$ for all questions) are not shown.

(E) Self-reported therapies show strong overlap with medical records. Therapy categories are shown on the y axis, with the proportion of patient partners from each data type (patient surveys and medical records) receiving therapies of that category shown on the x axis. In the online survey, patient partners selected therapies they received for their metastatic prostate cancer from a list. 639/706 patient partners reported at least one therapy and are shown. 119 of these participants also had abstracted therapy data from medical records. Report overlap refers to how often patient partners report receiving a therapy when their medical records show that they have received that therapy as a percentage. Only therapies available for selection in the patient survey were used in this comparison (Table S4).

(F) Landscape of lifestyle changes for patient partners. Participants were asked to list additional medications, alternative medications, or lifestyle changes since their diagnosis of prostate cancer. Free-text responses were manually abstracted and categorized into diet/lifestyle changes, supplements, and non-cancer medications. The y axis shows individual instances of diet/lifestyle changes, supplements, or medications. The x axis shows the percentage of patient partners with that lifestyle change or that were taking that supplement/therapy out of all patient partners that responded to the lifestyle question ($n = 456$). CBD/THC, cannabidiol/tetrahydrocannabinol (oils, medical marijuana, etc.).

our patient partners are likely influenced by participation bias and may differ from other prostate cancer studies as a result.

We used the MPCproject's comprehensive abstracted medical records together with patient-reported data to evaluate the treatments received in this real-world cohort (STAR Methods; Figure 2E). Patient partners reported taking an average of 2.8 therapies (range 1–13) to treat their prostate cancer. 119 (17%) patient partners had abstracted medical records at the time of writing, and there was 90% concordance between therapies noted in formal medical records and therapies reported by these patient partners. The overlap was lowest for

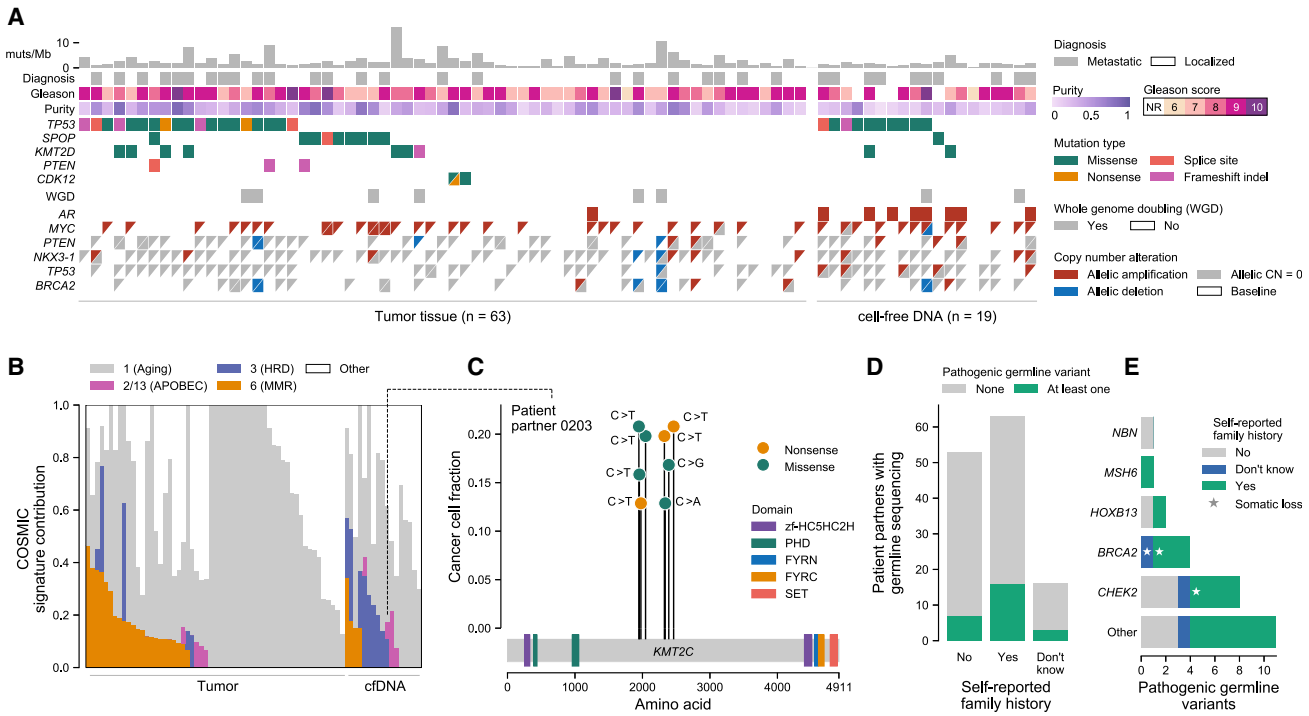


Figure 3. Remotely donated tumor and cell-free DNA samples obtained through patient partnership recapitulate known genomic findings in metastatic prostate cancer

(A) Genomic and clinical landscape of 82 sequenced samples. Columns represent samples, separated into tumor (prostate, left) and cell-free DNA (cfDNA; donated blood, right) samples, while rows represent select clinical and genomic features. Gleason scores for tumor samples are taken from the pathology report received with the sample (n = 58) or the patient partner's medical records (n = 5) if Gleason scores were not provided in the report. Gleason scores for cfDNA were taken from pathology reports in the medical record, with NR representing cases where a Gleason score was not reported in the medical record. Diagnosis refers to whether the initial diagnosis of prostate cancer was localized or metastatic. Multiple mutations in the same gene are represented as triangles. WGD refers to whole-genome doubling. Copy-number calls are allelic and defined with respect to baseline allelic ploidy (2 for samples with WGD, one for those without), with calls for the two alleles indicated by two triangles (except for *AR*, which has only one allele in men and so is shown as a single box). Allelic CN = 0 refers to complete allelic deletions. Allelic deletions that are not complete deletions are possible in samples with WGD. Figure created with CoMut.[29]

(B) Mutational signature analysis of sequenced samples. The relative contribution of select COSMIC v.2.0 mutational signatures are shown, separated by tumor and cfDNA (donated blood) sample type.[30] APOBEC refers to signatures associated with activity of APOBEC family of cytidine deaminases (signatures 2 and 13); MMR to the signature associated with deficient DNA mismatch repair (signature 6); and HRD to the signature associated with homologous recombination deficiency (signature 3). To be denoted as present, a signature cutoff of 6% was used. Samples with too few mutations for signature analysis (<50 mutations, n = 5 samples) are not shown.

(C) Instance of localized hypermutation (kataegis) of *KMT2C* in cfDNA from a donated blood sample. The y axis shows the cancer cell fraction of each mutation, while the x axis shows their amino acid within *KMT2C*. Domains taken from Pfam.[31] The dotted line connects to this sample's mutational signature profile.

(D and E) Germline pathogenic alterations and their overlap with patient-reported family history. Pathogenic germline alterations (as annotated by ClinVar) in genes from a select panel of genes previously implicated in cancer heritability were detected in patient partners with sequenced saliva or blood buffy coat (n = 132) (STAR Methods; Tables S3 and S5).[32] Survey responses to a question asking about a family history of prostate or breast cancer were tabulated and overlapped with this genomic data. Stars in (E) indicate instances where a somatic deletion also affected that gene in a tumor or cfDNA sample from that patient partner, suggesting biallelic inactivation.

treatments typically given earlier in the therapeutic timeline (first-line androgen deprivation therapy, 83%), supportive care therapies (64%), or treatments abandoned quickly due to side effects (Figure 2E).

We also used the patient-reported data to assess how living with prostate cancer has changed the daily lives of our patient partners. 56% of patient partners reported a lifestyle change because of living with their cancer, with the most common being a change in diet or exercise (Figure 2F). Common nutritional supplements reported include vitamin D and antioxidant-based supplements, while common non-cancer medications included metformin and statins.

Whole-exome sequencing of a real-world MPC patient cohort

To complement the demographic, patient-reported, and clinical data, we have completed molecular profiling of 572 samples from 333 patient partners to date, including ultra-low-pass whole-genome sequencing (ULP-WGS; average depth of 0.1×) of cfDNA from 318 donated blood samples; whole-exome sequencing (WES) of cfDNA from 47 of those blood samples; WES of 106 tumor samples; and WES of 148 germline samples from donated saliva or blood buffy coat. In total, 82 exome-sequenced samples (63 tumor and 19 cfDNA) from 79 patient partners enrolled before June 1, 2020, were included in

downstream genomic analyses after assessment of sufficient tumor purity (≥10%) and coverage (STAR Methods).

Exome sequencing from the tumor and cfDNA samples recapitulated known genomic patterns in MPC (Figure 3A). *TP53* and *SPOP* were recurrently altered, consistent with previous studies of both metastatic and primary prostate cancer ($q < 0.1$ via MutSig2CV).[3,4,6] In primary tumor samples from this cohort, the mutation frequency of *TP53* (29%) was more consistent with metastatic cohorts than those of primary prostate cancer.[3,6] Twenty-four (38%) primary tumor samples were from men diagnosed with *de novo* metastatic disease, and samples from these patient partners were more likely to carry *TP53* mutations ($p = 0.04$, Fisher's exact test). We also observed known patterns of copy-number alteration in prostate cancer, including recurrent amplifications of androgen receptor (*AR*) and *FOXA1*, as well as recurrent deletions of *PTEN* ($q < 0.1$ via GISTIC2.0; Figure 3A).[28] Whole-genome doubling was present in 6/63 tumor samples and 2/19 cfDNA samples, including in two tumor samples from patient partners initially diagnosed with localized prostate cancer. Both patient partners were diagnosed with metastatic disease within a few months of their initial diagnosis.

To understand the mutational processes in this cohort's exome-sequenced samples, we used a mutation-based method (deconstructSigs) to determine the contribution of COSMIC v.2.0 signatures to each sample[30,33] (Figure 3B; STAR Methods). We detected the presence of aging-associated clock-like signature one in all samples and the presence of signature 3 (associated with homologous recombination deficiency [HRD]) and signature 6 (associated with mismatch repair deficiency [MMR]) in a subset of samples. These results are consistent with previous studies implicating these signatures in prostate cancer, although they likely overestimate the prevalence of signature six in tumor samples due to formalin-induced deamination artifacts.[34,35] We found that the presence of signature three was enriched in metastasis-associated samples (cfDNA and primary tumors obtained in the metastatic setting) relative to tumor tissue from patient partners with strictly localized tumors at time of resection ($p = 0.04$, Fisher's exact test). While some samples with signature three had at least one alteration in *BRCA1* or *BRCA2* ($n = 9/16$), this association was not statistically significant, highlighting the potential role of other homologous repair defects in the etiology of signature 3, as noted in prior studies of prostate and breast cancer.[5,36–39] All samples with signature 3, however, had at least one alteration in a DNA-repair pathway gene, and biallelic *BRCA2* alterations were associated with copy-number-based estimations of HRD (STAR Methods; Figure S8).[40]

In 10% of samples (8/82), we observed contributions from COSMIC signatures 2 and 13, which are driven by APOBEC cytidine deaminases and are known to operate at a baseline level in prostate cancer.[34,41] APOBEC-driven mutagenesis has been implicated in kataegis—rare, localized hypermutation in specific nucleotide contexts that is associated with genomic instability and increased Gleason score in prostate cancer.[42,43] In one patient partner's cfDNA sample, we detected eight distinct mutations within a 2-kB window in *KMT2C*, a known prostate cancer driver (Figure 3C).[3] Six of these mutations were in a T(C>T)A nucleotide context, and this sample had a detectable contribution from COSMIC signature 13. We found that two pairs of the mutations, p.S1947F/p.S1954F and p.Q2325*/p.S2337Y, were each present on individual sequencing reads, confirming that these mutations existed within the same cell and strongly implicating *KMT2C* disruption through kataegis (Figure S9).

Given the strong heritability of prostate cancer, we assessed inherited germline alterations and their overlap with patient-reported family history of cancer.[44] We found that among the 132 patient partners (19%) with WES of donated saliva or blood buffy coat, 15 and 11 had pathogenic germline alterations in select genes implicated in prostate cancer and other cancers, respectively.[45] Men that self-reported a family history of prostate or breast cancer were more likely to have a pathogenic germline alteration associated with cancer, although this difference was not statistically significant (25% versus 13%, $p = 0.11$, Fisher's exact test; Figure 3D). The most mutated gene was *CHEK2* (8 patient partners), followed by *BRCA2* (4 patient partners). In three cases, we detected an accompanying somatic loss of a germline-mutated gene (Figures 3E and S10).

Longitudinal blood biopsies enable study of tumor evolution in a patient-partnered model

Ten patient partners had WES from both tumor tissue and cfDNA, and three patient partners had both samples pass quality-control metrics. Using the molecular data and abstracted medical records, we sought to explore the evolutionary relationships between these longitudinal samples in the context of patient clinical trajectories. Like most men with MPC, one participant, patient partner 0495, received a diverse range of treatments between biopsy timepoints (Figure 4A). After responding to first-line anti-androgen therapy (leuprolide + bicalutamide), they took second-generation anti-androgen inhibitors (abiraterone, enzalutamide), as well as experimental radiotherapy and immunotherapy. To explore the relationship between samples, we utilized PhylogicNDT, an algorithm that clusters mutations based on their prevalence in the tumor (cancer cell fraction) into evolutionarily related subclones (STAR Methods).[46] In the cfDNA sample of patient partner 0495, but not the primary tumor, we observed two distinct frameshift mutations in *ASXL2*, a gene implicated in castration-resistant MPC, as well as an amplification of *AR*, a known resistance mechanism to abiraterone and enzalutamide.[47,48] Patient partner 0093's tumor had clonal mutations in *TP53* and *KMT2D* but harbored an *NF2* mutation solely in the cfDNA sample. Patient partner 0213's tumor had a *TP53* mutation and APOBEC-associated COSMIC signature 13 detected exclusively in the cfDNA sample.

Two of these patient partners, 0495 and 0093, were initially diagnosed with primary prostate cancer (Gleason score 4 + 3 and 5 + 4, respectively), while patient partner 0213 was diagnosed with *de novo* metastatic disease. Their donated blood samples were separated from their primary tissue biopsies by a range of years (2–10 years). Despite these varied disease presentations, clinical trajectories, and biopsy timelines, we observed similar patterns of a "clonal switch" between the primary tumor and cfDNA, wherein different subclones were dominant in each sample (Figures 4B and S11). We did not, however, observe primary tumor-specific copy-number alterations, bolstering previous claims that subclonal diversification in

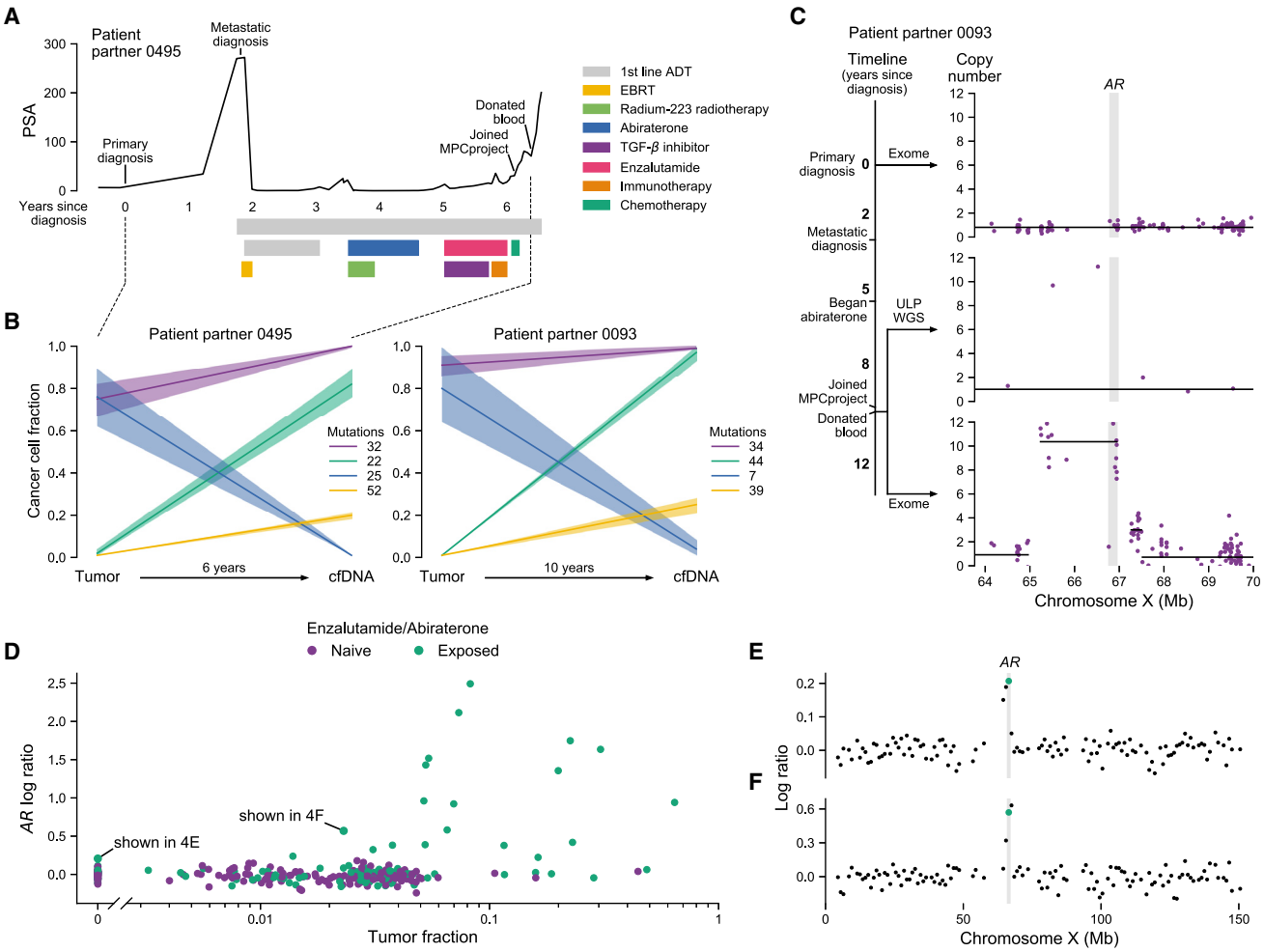


Figure 4. cfDNA from donated blood reveals patterns of clonal dynamics and clinically relevant genomic changes

(A) Clinical trajectory of patient partner 0495. This patient partner's prostate-specific antigen (PSA) trajectory is shown on the y axis, time in years since initial diagnosis is shown on the x axis, and bars denote the beginning and end of therapies. EBRT, external beam radiation therapy; first-line androgen deprivation therapy (ADT), leuprolide and bicalutamide; immunotherapy, nivolumab; chemotherapy, cisplatin and etoposide.

(B) Tumor evolution from primary tumor to metastatic cfDNA samples. The y axis shows the cancer cell fraction (CCF) of clonal clusters identified between tumor and cfDNA samples (x axis). Time between samples shown on the x axis. Colors indicate how many mutations were identified in each clone, with a 95% confidence interval around the estimated CCF. Purple represents the truncal/ancestral clone. Clusters with CCF <0.10 across all biopsies are omitted. The clinical trajectory of patient partner 0495 (left) is shown in (A), while the trajectory of patient partner 0093 (right) is shown in (C).

(C) Emergence of *AR* amplification in patient partner 0093 induced by anti-androgen therapy. The timeline depicts this patient's clinical trajectory, while the plots show the absolute copy number (y axis) of the genomic region around *AR* (x axis, gene body shown in gray). The first plot depicts exome sequencing from the patient's archival tumor tissue; the second and third plots depict ultra-low-pass whole-genome sequencing (ULP-WGS) and exome sequencing of cfDNA from the patient's donated blood, respectively. Individual points represent copy number of target regions (exome) or copy number of 1 Mb genomic windows (ULP-WGS). Black lines represent discrete copy-number segments.

(D–F) ULP-WGS reveals clinically relevant *AR* amplifications even at low tumor fraction. In (D), tumor fraction of 318 cfDNA samples from donated blood of 300 patient partners with ULP-WGS sequencing is shown on the x axis, while the log copy ratio (logR) of the genomic interval containing *AR* is shown on the y axis. Points are colored by whether patient partners self-reported taking enzalutamide or abiraterone. 89 samples are shown with tumor fraction of 0 (undetectable), while 229 have non-zero tumor fractions. Two samples, one at a tumor fraction of 0 and another at a tumor fraction of 0.023, have chromosome X log copy ratio profiles shown in (E) and (F), respectively. The green points represent the values shown in (D), with the genomic interval containing *AR* highlighted in gray.

MPC via mutations may happen after acquisition of ancestral copy-number alterations (Figure S12).[49] Furthermore, we observed likely primary-tumor-specific mutations across all seven other patient partners with both tumor and cfDNA samples, although the samples had low purity (Figure S13). While we cannot account for the sampling bias of tumor biopsies, these results suggest that such clonal switches may be common in the development of metastatic disease.

In several cases, we detected the emergence of an amplification in the *AR* between the initial diagnosis and metastatic blood sample that was captured using ULP-WGS of cfDNA (example patient partner shown in Figure 4C). This led us to examine *AR*

copy number using ULP-WGS of cfDNA samples across the entire cohort (n = 300 patient partners, 318 samples; Figures 4D and S14). We found that patient partners who reported taking enzalutamide or abiraterone had significantly higher *AR* log copy ratios across a range of tumor fractions ($p < 0.001$, linear regression). Men who had taken enzalutamide or abiraterone also had significantly higher tumor fractions, likely reflecting a more advanced disease state and subsequent higher tumor burden in blood ($p < 0.001$, Mann-Whitney U test).[50] We observed that *AR* amplifications are often detectable in ULP-WGS of cfDNA even when the tumor fraction is below 0.03 (Figures 4E and 4F). For one patient partner, the tumor fraction within their donated blood was inferred as undetectable, but we nevertheless observed a clear *AR* amplification (Figure 4E). This highlights the potential efficacy of cfDNA to reveal clinically relevant changes in MPC, even in cases of very low or undetectable tumor burden. Attempts to identify other common copy-number changes were limited by tumor fraction (Figure S15). Broadly, these sequencing results illustrate the feasibility of identifying relevant genomic and evolutionary alterations from both archival tumor tissue and donated blood samples irrespective of geographical source site, enabling patient partners to participate in genomic research at no cost and with little effort.

DISCUSSION

Here, we describe the MPCproject, a patient-driven framework for partnering with patients with MPC in the US and Canada to increase access to genomics research and strengthen our understanding of this disease. The online enrollment process was jointly created with patient partners and advocates to emphasize simplicity, requiring only the completion of online consent and survey forms, along with optional mailed saliva and blood kits. To our knowledge, no previous effort in MPC has used patient partnership to integrate demographic, clinical, patient-reported, and genomic data from patients at a national level.

To that end, we demonstrated the feasibility of working with over 700 patient partners, 41% of whom live in rural areas, MUAs, or HPSAs, a metric unreported in previous molecular profiling efforts. We found that 56% of our patient partners have never received care at an NCI-designated cancer center and that patient partners living in more disadvantaged areas were less likely to attend those institutions for treatment. Taken together with previous studies showing disparities in standard treatment and clinical trial outcomes by socioeconomic status, these results highlight existing barriers in access to care and sequencing studies.[51–53] Furthermore, a recent study found that incomplete medical records are associated with shorter overall survival for patients with MPC, particularly for those with complicated clinical histories or whose care is fragmented between institutions.[54] Our analysis of abstracted medical record data revealed a strong overlap between clinical histories represented in medical records and patient-reported data, even for patient partners with complex treatment trajectories or who had received treatment at multiple hospitals, supporting the use of patient surveys to improve care in this disease.

We also demonstrated that tumor tissue collected from archival samples and cfDNA from donated blood samples from across the US and Canada accurately recapitulate known genomic findings in MPC and place findings in the context of both patient-reported and abstracted medical record data. There has been substantial effort in the field to identify molecular features associated with selective response to therapies like PARP inhibition and immunotherapy, including the use of mutational signatures to assess targetable HRD, MMR, and APOBEC deficiencies in cases without a causative molecular alteration.[36,55] Our results strengthen previous findings that such signatures can be detected using cfDNA and, combined with our ability to obtain cfDNA from participants nationwide, demonstrate the scalability of a patient-partnered approach to identify and validate such genomic findings within a real-world cohort in parallel to existing molecular approaches.[56,57]

Moreover, we used archival tumor tissue and cfDNA from donated blood to reconstruct tumor phylogenetic profiles, revealing polyclonality between primary and metastatic diagnosis. Despite well-known findings of heterogeneity in both primary and MPC, there is a paucity of matched primary-metastatic studies, owing mostly to the invasiveness and logistical challenges of longitudinal biopsy studies.[34,58] Our project enables such studies paired with comprehensive clinical histories with minimal patient effort. To that end, we also found clinically relevant *AR* amplifications via low-pass WGS of cfDNA from donated blood, even at very low or undetectable tumor fractions. This result provides additional inexpensive utility to the suggested use of cfDNA tumor fraction as a clinically relevant biomarker in MPC.[50,56] We are working with patient partners who continue to donate blood and have been able to collect multiple secondary blood biopsy kits for future longitudinal analysis.

New approaches in molecular cancer research are needed to address an increased desire from patients to actively participate in research and a pressing need for equity in the clinic. Paired with emerging open-access clinical trials, patient-driven studies hold great promise to achieve equity and accelerate discovery in genomic research.[59] The MPCproject is part of a wider "Count Me In" patient-partnered initiative (joincountmein.org) that has already yielded new findings in angiosarcoma and has expanded to metastatic breast cancer and osteosarcoma, among others.[60–62] The achievements of the MPCproject are based entirely on the courage and altruism of the men with whom we partner, who, in the words of one participant, hope that their "participation will help other men […] and lead eventually to a cure."

Limitations of the study

Despite the geographic diversity of our patient partners, we acknowledge that they do not reflect the racial diversity of patients with MPC, a critical issue given substantial disparities in both cancer care and genomics research by race and ethnicity.[11,63,64] These unmet disparities demand that we rethink our models of outreach and patient engagement, and our effort cannot be considered a success until sustained and equitable partnership is achieved.[65] Recognizing that building trust in marginalized communities takes time, we must continue to work longitudinally with community-based advocacy organizations to partner with Black communities. Since the launch of our project, we have worked to build an engagement model

that meets patients in their communities, including churches, barbershops, and fraternities. Using the longitudinal model of this study, we will continue to iteratively learn from community engagement successes and failures. We received feedback, for example, that Black patients and their cancer stories are rarely heard—in response, we are building a campaign to amplify the voices of Black patients with cancer and their lived experiences (www.BlackCancerVoices.org). Additionally, a common request is for the project to return clinically relevant sequencing results to patient partners and their physicians. We are working with regulatory, clinical, and sequencing experts to build the infrastructure necessary to fulfill this request.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - MPCproject website
  - Informed consent
  - Patient-reported data
  - Acquisition of medical records
  - Acquisition of patient samples
  - Medical record abstraction
  - Geographic analysis
  - Whole exome sequencing analysis
  - Whole exome sequencing quality control
  - Ultra-low pass whole genome sequencing analysis
  - Mutational signature analysis and kataegis
  - Germline variant discovery
  - Association of DNA-repair alterations and presence of signature 3
  - Phylogenetic analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Statistical analysis
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xgen.2022.100169.

ACKNOWLEDGMENTS

We thank our patient partners, caregivers, loved ones, project advisory council, and advocacy partners, without whom this project would not be possible. We would like to pay our respects to the late Jack Whelan, a patient with MPC and advocate, who was instrumental in developing the MPCproject. We also thank the staff of the MPCproject, the engineering team from the Data Sciences Platform at the Broad Institute (A. Zimmer, E. Baker, S. Maiwald, P. Taheri, D. Kaplan, J. Lapan, S. Sutherland), and all members of Count Me In who work daily to ensure that all patients with MPC can participate in research. Finally, we would like to express our gratitude to the Broad Institute Cancer Program, the Broad Institute Genomics Platform, Broad Institute Communications and Development teams, and the compliance team at the Broad Institute for their support of the project. Figure 1A and parts of Figure 2F were created with BioRender.com. This work was funded by the following: Count Me In, Inc., Fund for Innovation in Cancer Informatics (E.M.V.A.); PCF-Movember Challenge Award (E.M.V.A.); NIH R01CA227388 and U01CA233100 (E.M.V.A.); Mark Foundation Emerging Leader Award (E.M.V.A.); Participant Engagement and Cancer Genome Sequencing (U2CCA252974); US Department of Defense (W81XWH-21-1-0084 and PC200150, S.H.A.); Prostate Cancer Foundation (S.H.A.); Conquer Cancer Foundation of the American Society of Clinical Oncology (S.H.A.); National Science Foundation (GRFP DGE1144152, M.X.H.); and National Institutes of Health (T32 GM008313).

AUTHOR CONTRIBUTIONS

N.W., C.A.P., and E.M.V.A. conceived and designed the MPCproject with support from E.S.L. J.C., S.B., L.S., and E.M.V.A. designed and prepared the study and interpreted the data. J.C. wrote the manuscript and performed the analyses. S.B. and L.S. led study operations including tumor sample and medical record acquisition, sample sequencing, and patient coordination. L.S., B.S.T., M.D., E.A., S.S., A.L.D., R.R., D.M.S., and I.K.S. oversaw medical record abstraction. S.Y.C. provided feedback on various analyses of the study and completed germline variant calling with oversight from S.H.A. S.B., L.S., J.C., B.N.T., M.D., M.M., and P.S.C. coordinated data releases. M.M., P.S.C., A.D., and B.Z. led recent project operations. M.D. supervised early project operations. C.M.N. and E.A. led patient advocacy and outreach efforts. A.T.M.C. and S.W. oversaw early project sequencing analyses. M.X.H. provided feedback of study analyses. A.K.T. provided feedback on medical record abstractions and tissue sample collection. D.K. enabled electronic medical record searching. J.N., J.M., I.H.G., and B.O. contributed to survey design, project development, assessment of patient criteria, and outreach strategy.

DECLARATION OF INTERESTS

M.X.H. has been a consultant to Amplify Medicines and Ikena Oncology and is a current employee of Genentech/Roche. E.S.L. is currently in the process of divesting any relevant holdings. N.W. reports advisory relationships and consulting with Eli Lilly and Co.; advising and stockholding interest in Relay Therapeutics; and grant support from Puma Biotechnology. E.M.V.A. reports advisory relationships and consulting with Tango Therapeutics, Genome Medical, Invitae, Illumina, Enara Bio, Mani-fold Bio, and Janssen; research support from Novartis and BMS; equity in Tango Therapeutics, Genome Medical, Syapse, Mani-fold Bio, and Enara Bio; and travel reimbursement from Roche and Genentech, outside the submitted work.

Received: November 22, 2021
Revised: April 25, 2022
Accepted: July 26, 2022
Published: August 19, 2022

REFERENCES

1. Siegel, R.L., Miller, K.D., and Jemal, A. (2020). Cancer statistics, 2020. CA. Cancer J. Clin. *70*, 7–30. https://doi.org/10.3322/caac.21590.
2. Litwin, M.S., and Tan, H.-J. (2017). The diagnosis and treatment of prostate cancer: a review. JAMA *317*, 2532–2542. https://doi.org/10.1001/jama.2017.7248.
3. Cancer Genome Atlas Research Network (2015). The molecular taxonomy of primary prostate cancer. Cell *163*, 1011–1025. https://doi.org/10.1016/j.cell.2015.10.025.
4. Armenia, J., Wankowicz, S.A.M., Liu, D., Gao, J., Kundra, R., Reznik, E., Chatila, W.K., Chakravarty, D., Han, G.C., Coleman, I., et al. (2018). The long tail of oncogenic drivers in prostate cancer. Nat. Genet. *50*, 645–651. https://doi.org/10.1038/s41588-018-0078-z.
5. de Bono, J., Mateo, J., Fizazi, K., Saad, F., Shore, N., Sandhu, S., Chi, K.N., Sartor, O., Agarwal, N., Olmos, D., et al. (2020). Olaparib for

metastatic castration-resistant prostate cancer. N. Engl. J. Med. *382*, 2091–2102. https://doi.org/10.1056/NEJMoa1911440.

6. Abida, W., Cyrta, J., Heller, G., Prandi, D., Armenia, J., Coleman, I., Cieslik, M., Benelli, M., Robinson, D., Van Allen, E.M., et al. (2019). Genomic correlates of clinical outcome in advanced prostate cancer. Proc. Natl. Acad. Sci. USA *116*, 11428–11436. https://doi.org/10.1073/pnas.1902651116.
7. Annala, M., Vandekerkhove, G., Khalaf, D., Taavitsainen, S., Beja, K., Warner, E.W., Sunderland, K., Kollmannsberger, C., Eigl, B.J., Finch, D., et al. (2018). Circulating tumor DNA genomics correlate with resistance to abiraterone and enzalutamide in prostate cancer. Cancer Discov. *8*, 444–457. https://doi.org/10.1158/2159-8290.CD-17-0937.
8. Sonpavde, G., Agarwal, N., Pond, G.R., Nagy, R.J., Nussenzveig, R.H., Hahn, A.W., Sartor, O., Gourdin, T.S., Nandagopal, L., Ledet, E.M., et al. (2019). Circulating tumor DNA alterations in patients with metastatic castration-resistant prostate cancer. Cancer *125*, 1459–1469. https://doi.org/10.1002/cncr.31959.
9. Siu, L.L., Lawler, M., Haussler, D., Knoppers, B.M., Lewin, J., Vis, D.J., Liao, R.G., Andre, F., Banks, I., Barrett, J.C., et al. (2016). Facilitating a culture of responsible and effective sharing of cancer genome data. Nat. Med. *22*, 464–471. https://doi.org/10.1038/nm.4089.
10. Joly, Y., Dove, E.S., Knoppers, B.M., Bobrow, M., and Chalmers, D. (2012). Data sharing in the post-genomic world: the experience of the international cancer genome consortium (ICGC) data access compliance office (DACO). PLoS Comput. Biol. *8*, e1002549. https://doi.org/10.1371/journal.pcbi.1002549.
11. Spratt, D.E., Chan, T., Waldron, L., Speers, C., Feng, F.Y., Ogunwobi, O.O., and Osborne, J.R. (2016). Racial/ethnic disparities in genomic sequencing. JAMA Oncol. *2*, 1070–1074. https://doi.org/10.1001/jamaoncol.2016.1854.
12. Feyman, Y., Provenzano, F., and David, F.S. (2020). Disparities in clinical trial access across US urban areas. JAMA Netw. Open *3*, e200172. https://doi.org/10.1001/jamanetworkopen.2020.0172.
13. Huey, R.W., Hawk, E., and Offodile, A.C. (2019). Mind the Gap: precision Oncology and its potential to widen disparities. J. Oncol. Pract. *15*, 301–304. https://doi.org/10.1200/JOP.19.00102.
14. Mamun, A., Nsiah, N.Y., Srinivasan, M., Chaturvedula, A., Basha, R., Cross, D., Jones, H.P., Nandy, K., and Vishwanatha, J.K. (2019). Diversity in the era of precision medicine - from bench to bedside implementation. Ethn. Dis. *29*, 517–524. https://doi.org/10.18865/ed.29.3.517.
15. Messner, D.A., Al Naber, J., Koay, P., Cook-Deegan, R., Majumder, M., Javitt, G., Deverka, P., Dvoskin, R., Bollinger, J., Curnutte, M., et al. (2016). Barriers to clinical adoption of next generation sequencing: perspectives of a policy Delphi panel. Appl. Transl. Genom. *10*, 19–24. https://doi.org/10.1016/j.atg.2016.05.004.
16. American Cancer Society Cancer Action Network (2020). Payer Coverage Policies of Tumor Biomarker Testing (American Cancer Society Cancer Action Network).
17. Chakradhar, S. (2014). Tumor sequencing takes off, but insurance reimbursement lags. Nat. Med. *20*, 1220–1221. https://doi.org/10.1038/nm1114-1220.
18. McGuire, A.L., Oliver, J.M., Slashinski, M.J., Graves, J.L., Wang, T., Kelly, P.A., Fisher, W., Lau, C.C., Goss, J., Okcu, M., et al. (2011). To share or not to share: a randomized trial of consent for data sharing in genome research. Genet. Med. *13*, 948–955. https://doi.org/10.1097/GIM.0b013e3182227589.
19. Husedzinovic, A., Ose, D., Schickhardt, C., Fröhling, S., and Winkler, E.C. (2015). Stakeholders' perspectives on biobank-based genomic research: systematic review of the literature. Eur. J. Hum. Genet. *23*, 1607–1614. https://doi.org/10.1038/ejhg.2015.27.
20. Sequence me - demand genomic & DNA testing for cancer treatments. (2022). https://sequenceme.org/.
21. Surveillance Research Program. (2022). National Cancer Institute SEER*Stat Software (National Cancer Institute).
22. Economic Research Service (2013). Rural-Urban Continuum Codes - Documentation (U.S. Department of Agriculture).
23. https://data.hrsa.gov/tools/shortage-area; 2022.
24. Population Clock, (2022). https://www.census.gov/popclock/.
25. Maroko, A.R., Doan, T.M., Arno, P.S., Hubel, M., Yi, S., and Viola, D. (2016). Integrating social determinants of health with treatment and prevention: a new tool to assess local area deprivation. Prev. Chronic Dis. *13*, E128. https://doi.org/10.5888/pcd13.160221.
26. Anderson, M., and Perrin, A. (2017). Technology Use Among Seniors (Pew Research Center Internet Science Tech). https://www.pewresearch.org/internet/2017/05/17/technology-use-among-seniors/.
27. Rawla, P. (2019). Epidemiology of prostate cancer. World J. Oncol. *10*, 63–89. https://doi.org/10.14740/wjon1191.
28. Mermel, C.H., Schumacher, S.E., Hill, B., Meyerson, M.L., Beroukhim, R., and Getz, G. (2011). GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol. *12*, R41. https://doi.org/10.1186/gb-2011-12-4-r41.
29. Crowdis, J., He, M.X., Reardon, B., and Van Allen, E.M. (2020). CoMut: visualizing integrated molecular information with comutation plots. Bioinformatics *36*, 4348–4349. https://doi.org/10.1093/bioinformatics/btaa554.
30. Tate, J.G., Bamford, S., Jubb, H.C., Sondka, Z., Beare, D.M., Bindal, N., Boutselakis, H., Cole, C.G., Creatore, C., Dawson, E., et al. (2019). COSMIC: the catalogue of somatic mutations in cancer. Nucleic Acids Res. *47*, D941–D947. https://doi.org/10.1093/nar/gky1015.
31. Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., et al. (2014). Pfam: the protein families database. Nucleic Acids Res. *42*, D222–D230. https://doi.org/10.1093/nar/gkt1223.
32. Landrum, M.J., Lee, J.M., Benson, M., Brown, G.R., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Jang, W., et al. (2018). ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res. *46*, D1062–D1067. https://doi.org/10.1093/nar/gkx1153.
33. Rosenthal, R., McGranahan, N., Herrero, J., Taylor, B.S., and Swanton, C. (2016). deconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome Biol. *17*, 31. https://doi.org/10.1186/s13059-016-0893-4.
34. Gerhauser, C., Favero, F., Risch, T., Simon, R., Feuerbach, L., Assenov, Y., Heckmann, D., Sidiropoulos, N., Waszak, S.M., Hübschmann, D., et al. (2018). Molecular evolution of early-onset prostate cancer identifies molecular risk markers and clinical trajectories. Cancer Cell *34*, 996–1011.e8. https://doi.org/10.1016/j.ccell.2018.10.016.
35. Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A.J.R., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Børresen-Dale, A.L., et al. (2013). Signatures of mutational processes in human cancer. Nature *500*, 415–421. https://doi.org/10.1038/nature12477.
36. Mateo, J., Carreira, S., Sandhu, S., Miranda, S., Mossop, H., Perez-Lopez, R., Nava Rodrigues, D., Robinson, D., Omlin, A., Tunariu, N., et al. (2015). DNA-repair defects and Olaparib in metastatic prostate cancer. N. Engl. J. Med. *373*, 1697–1708. https://doi.org/10.1056/NEJMoa1506859.
37. Pritchard, C.C., Mateo, J., Walsh, M.F., De Sarkar, N., Abida, W., Beltran, H., Garofalo, A., Gulati, R., Carreira, S., Eeles, R., et al. (2016). Inherited DNA-repair gene mutations in men with metastatic prostate cancer. N. Engl. J. Med. *375*, 443–453. https://doi.org/10.1056/NEJMoa1603144.
38. Polak, P., Kim, J., Braunstein, L.Z., Karlic, R., Haradhavala, N.J., Tiao, G., Rosebrock, D., Livitz, D., Kübler, K., Mouw, K.W., et al. (2017). A mutational signature reveals alterations underlying deficient homologous recombination repair in breast cancer. Nat. Genet. *49*, 1476–1486. https://doi.org/10.1038/ng.3934.
39. Sztupinszki, Z., Diossy, M., Krzystanek, M., Borcsok, J., Pomerantz, M.M., Tisza, V., Spisak, S., Rusz, O., Csabai, I., Freedman, M.L., and Szallasi, Z. (2020). Detection of molecular signatures of homologous recombination

deficiency in prostate cancer with or without BRCA1/2 mutations. Clin. Cancer Res. *26*, 2673–2680. https://doi.org/10.1158/1078-0432.CCR-19-2135.

40. Sztupinszki, Z., Diossy, M., Krzystanek, M., Reiniger, L., Csabai, I., Favero, F., Birkbak, N.J., Eklund, A.C., Syed, A., and Szallasi, Z. (2018). Migrating the SNP array-based homologous recombination deficiency measures to next generation sequencing data of breast cancer. Npj Breast Cancer *4*, 16. https://doi.org/10.1038/s41523-018-0066-6.

41. Swanton, C., McGranahan, N., Starrett, G.J., and Harris, R.S. (2015). APOBEC enzymes: mutagenic fuel for cancer evolution and heterogeneity. Cancer Discov. *5*, 704–712. https://doi.org/10.1158/2159-8290.CD-15-0344.

42. Nik-Zainal, S., Alexandrov, L.B., Wedge, D.C., Van Loo, P., Greenman, C.D., Raine, K., Jones, D., Hinton, J., Marshall, J., Stebbings, L.A., et al. (2012). Mutational processes molding the genomes of 21 breast cancers. Cell *149*, 979–993. https://doi.org/10.1016/j.cell.2012.04.024.

43. Fraser, M., Sabelnykova, V.Y., Yamaguchi, T.N., Heisler, L.E., Livingstone, J., Huang, V., Shiah, Y.-J., Yousif, F., Lin, X., Masella, A.P., et al. (2017). Genomic hallmarks of localized, non-indolent prostate cancer. Nature *541*, 359–364. https://doi.org/10.1038/nature20788.

44. Mucci, L.A., Hjelmborg, J.B., Harris, J.R., Czene, K., Havelick, D.J., Scheike, T., Graff, R.E., Holst, K., Möller, S., Unger, R.H., et al. (2016). Familial risk and heritability of cancer among twins in nordic countries. JAMA *315*, 68–76. https://doi.org/10.1001/jama.2015.17703.

45. AlDubayan, S.H. (2019). Considerations of multigene test findings among men with prostate cancer - knowns and unknowns. Can. J. Urol. *26*, 14–16.

46. Leshchiner, I., Livitz, D., Gainor, J.F., Rosebrock, D., Spiro, O., Martinez, A., Mroz, E., Lin, J.J., Stewart, C., Kim, J., et al. (2018). Comprehensive analysis of tumour initiation, spatial and temporal progression under multiple lines of treatment. Preprint at bioRxiv, 508127. https://doi.org/10.1101/508127.

47. Grasso, C.S., Wu, Y.-M., Robinson, D.R., Cao, X., Dhanasekaran, S.M., Khan, A.P., Quist, M.J., Jing, X., Lonigro, R.J., Brenner, J.C., et al. (2012). The mutational landscape of lethal castration-resistant prostate cancer. Nature *487*, 239–243. https://doi.org/10.1038/nature11125.

48. Tucci, M., Zichi, C., Buttigliero, C., Vignani, F., Scagliotti, G.V., and Di Maio, M. (2018). Enzalutamide-resistant castration-resistant prostate cancer: challenges and solutions. OncoTargets Ther. *11*, 7353–7368. https://doi.org/10.2147/OTT.S153764.

49. Espiritu, S.M.G., Liu, L.Y., Rubanova, Y., Bhandari, V., Holgersen, E.M., Szyca, L.M., Fox, N.S., Chua, M.L.K., Yamaguchi, T.N., Heisler, L.E., et al. (2018). The evolutionary landscape of localized prostate cancers drives clinical aggression. Cell *173*, 1003–1013.e15. https://doi.org/10.1016/j.cell.2018.03.029.

50. Choudhury, A.D., Werner, L., Francini, E., Wei, X.X., Ha, G., Freeman, S.S., Rhoades, J., Reed, S.C., Gydush, G., Rotem, D., et al. (2018). Tumor fraction in cell-free DNA as a biomarker in prostate cancer. JCI Insight *3*, e122109. https://doi.org/10.1172/jci.insight.122109.

51. Onega, T., Duell, E.J., Shi, X., Demidenko, E., and Goodman, D. (2009). Determinants of NCI Cancer Center attendance in Medicare patients with lung, breast, colorectal, or prostate cancer. J. Gen. Intern. Med. *24*, 205–210. https://doi.org/10.1007/s11606-008-0863-y.

52. Unger, J.M., Moseley, A.B., Cheung, C.K., Osarogiagbon, R.U., Symington, B., Ramsey, S.D., and Hershman, D.L. (2021). Persistent disparity: socioeconomic deprivation and cancer outcomes in patients treated in clinical trials. J. Clin. Oncol. *39*, 1339–1348. https://doi.org/10.1200/JCO.20.02602.

53. Singh, G.K., and Jemal, A. (2017). Socioeconomic and racial/ethnic disparities in cancer mortality, incidence, and survival in the United States, 1950-2014: over six decades of changing patterns and widening inequalities. J. Environ. Public Health *2017*, 2819372. https://doi.org/10.1155/2017/2819372.

54. Yang, D.X., Khera, R., Miccio, J.A., Jairam, V., Chang, E., Yu, J.B., Park, H.S., Krumholz, H.M., and Aneja, S. (2021). Prevalence of missing data in the national cancer database and association with overall survival. JAMA Netw. Open *4*, e211793. https://doi.org/10.1001/jamanetworkopen.2021.1793.

55. Abida, W., Campbell, D., Patnaik, A., Shapiro, J.D., Sautois, B., Vogelzang, N.J., Voog, E.G., Bryce, A.H., McDermott, R., Ricci, F., et al. (2020). Non-BRCA DNA damage repair gene alterations and response to the PARP inhibitor rucaparib in metastatic castration-resistant prostate cancer: analysis from the phase II TRITON2 study. Clin. Cancer Res. *26*, 2487–2496. https://doi.org/10.1158/1078-0432.CCR-20-0394.

56. Adalsteinsson, V.A., Ha, G., Freeman, S.S., Choudhury, A.D., Stover, D.G., Parsons, H.A., Gydush, G., Reed, S.C., Rotem, D., Rhoades, J., et al. (2017). Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. Nat. Commun. *8*, 1324. https://doi.org/10.1038/s41467-017-00965-y.

57. Ritch, E., Fu, S.Y.F., Herberts, C., Wang, G., Warner, E.W., Schönlau, E., Taavitsainen, S., Murtha, A.J., Vandekerkhove, G., Beja, K., et al. (2020). Identification of hypermutation and defective mismatch repair in ctDNA from metastatic prostate cancer. Clin. Cancer Res. *26*, 1114–1125. https://doi.org/10.1158/1078-0432.CCR-19-1623.

58. Gundem, G., Van Loo, P., Kremeyer, B., Alexandrov, L.B., Tubio, J.M.C., Papaemmanuil, E., Brewer, D.S., Kallio, H.M.L., Högnäs, G., Annala, M., et al. (2015). The evolutionary history of lethal metastatic prostate cancer. Nature *520*, 353–357. https://doi.org/10.1038/nature14347.

59. McKay, R.R., Gold, T., Zarif, J.C., Chowdhury-Paulino, I.M., Friedant, A., Gerke, T., Grant, M., Hawthorne, K., Heath, E., Huang, F.W., et al. (2021). Tackling diversity in prostate cancer clinical trials: a report from the diversity working group of the ironman registry. JCO Glob. Oncol. *7*, 495–505. https://doi.org/10.1200/GO.20.00571.

60. Painter, C.A., Jain, E., Tomson, B.N., Dunphy, M., Stoddard, R.E., Thomas, B.S., Damon, A.L., Shah, S., Kim, D., Gómez Tejeda Zañudo, J., et al. (2020). The Angiosarcoma Project: enabling genomic and clinical discoveries in a rare cancer through patient-partnered research. Nat. Med. *26*, 181–187. https://doi.org/10.1038/s41591-019-0749-z.

61. Count Me In. Count Me In. (2022). https://joincountmein.org/.

62. Wagle, N., Painter, C., Krevalin, M., Oh, C., Anderka, K., Larkin, K., Lennon, N., Dillon, D., Frank, E., Winer, E.P., et al. (2016). The Metastatic Breast Cancer Project: a national direct-to-patient initiative to accelerate genomics research. J. Clin. Oncol. *34*, LBA1519. https://doi.org/10.1200/JCO.2016.34.18_suppl.LBA1519.

63. Ward, E., Jemal, A., Cokkinides, V., Singh, G.K., Cardinez, C., Ghafoor, A., and Thun, M. (2004). Cancer disparities by race/ethnicity and socioeconomic status. CA. Cancer J. Clin. *54*, 78–93. https://doi.org/10.3322/canjclin.54.2.78.

64. Rebbeck, T.R. (2018). Prostate cancer disparities by race and ethnicity: from nucleotide to neighborhood. Cold Spring Harb. Perspect. Med. *8*, a030387. https://doi.org/10.1101/cshperspect.a030387.

65. Bailey, Z.D., Krieger, N., Agénor, M., Graves, J., Linos, N., and Bassett, M.T. (2017). Structural racism and health inequities in the USA: evidence and interventions. Lancet *389*, 1453–1463. https://doi.org/10.1016/S0140-6736(17)30569-X.

66. USDA ERS - Documentation. (2020). https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes/documentation/.

67. Kind, A.J.H., and Buckingham, W.R. (2018). Making neighborhood-disadvantage metrics accessible — the neighborhood atlas. N. Engl. J. Med. *378*, 2456–2458. https://doi.org/10.1056/NEJMp1802313.

68. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., and Cunningham, F. (2016). The ensembl variant effect predictor. Genome Biol. *17*, 122. https://doi.org/10.1186/s13059-016-0974-4.

69. Sondka, Z., Bamford, S., Cole, C.G., Ward, S.A., Dunham, I., and Forbes, S.A. (2018). The COSMIC Cancer Gene Census: describing genetic

dysfunction across all human cancers. Nat. Rev. Cancer *18*, 696–705. https://doi.org/10.1038/s41568-018-0060-1.

70. Python Software, Foundation. (2022). Python, *3* (Python Software Foundation).
71. R Core Team (2021). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing).
72. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.
73. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303. https://doi.org/10.1101/gr.107524.110.
74. Cibulskis, K., Lawrence, M.S., Carter, S.L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel, S., Meyerson, M., Lander, E.S., and Getz, G. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat. Biotechnol. *31*, 213–219. https://doi.org/10.1038/nbt.2514.
75. Saunders, C.T., Wong, W.S.W., Swamy, S., Becq, J., Murray, L.J., and Cheetham, R.K. (2012). Strelka: accurate somatic small-variant calling from sequenced tumor–normal sample pairs. Bioinformatics *28*, 1811–1817. https://doi.org/10.1093/bioinformatics/bts271.
76. Ramos, A.H., Lichtenstein, L., Gupta, M., Lawrence, M.S., Pugh, T.J., Saksena, G., Meyerson, M., and Getz, G. (2015). Oncotator: cancer variant annotation tool. Hum. Mutat. *36*, E2423–E2429. https://doi.org/10.1002/humu.22771.
77. Lawrence, M.S., Stojanov, P., Mermel, C.H., Robinson, J.T., Garraway, L.A., Golub, T.R., Meyerson, M., Gabriel, S.B., Lander, E.S., and Getz, G. (2014). Discovery and saturation analysis of cancer genes across 21 tumour types. Nature *505*, 495–501. https://doi.org/10.1038/nature12912.
78. Carter, S.L., Cibulskis, K., Helman, E., McKenna, A., Shen, H., Zack, T., Laird, P.W., Onofrio, R.C., Winckler, W., Weir, B.A., et al. (2012). Absolute quantification of somatic DNA alterations in human cancer. Nat. Biotechnol. *30*, 413–421. https://doi.org/10.1038/nbt.2203.
79. Shen, R., and Seshan, V.E. (2016). FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. Nucleic Acids Res. *44*, e131. https://doi.org/10.1093/nar/gkw520.
80. Taylor-Weiner, A., Stewart, C., Giordano, T., Miller, M., Rosenberg, M., Macbeth, A., Lennon, N., Rheinbay, E., Landau, D.-A., Wu, C.J., and Getz, G. (2018). DeTiN: overcoming tumor-in-normal contamination. Nat. Methods *15*, 531–534. https://doi.org/10.1038/s41592-018-0036-9.
81. Cibulskis, K., McKenna, A., Fennell, T., Banks, E., DePristo, M., and Getz, G. (2011). ContEst: estimating cross-contamination of human samples in next-generation sequencing data. Bioinformatics *27*, 2601–2602. https://doi.org/10.1093/bioinformatics/btr446.
82. Poplin, R., Chang, P.-C., Alexander, D., Schwartz, S., Colthurst, T., Ku, A., Newburger, D., Dijamco, J., Nguyen, N., Afshar, P.T., et al. (2018). A universal SNP and small-indel variant caller using deep neural networks. Nat. Biotechnol. *36*, 983–987. https://doi.org/10.1038/nbt.4235.
83. Tesseract, OCR (2022). https://github.com/tesseract-ocr/tesseract.
84. Geocoder - U.S. Census Bureau. (2022). https://geocoding.geo.census.gov/.
85. FilterByOrientationBias, GATK (2019). https://gatk.broadinstitute.org/hc/en-us/articles/360037060232-FilterByOrientationBias-EXPERIMENTAL-.
86. CrosscheckFingerprints (Picard), – GATK (2021). https://gatk.broadinstitute.org/hc/en-us/articles/360037594711-CrosscheckFingerprints-Picard-.
87. Prentice, L.M., Miller, R.R., Knaggs, J., Mazloomian, A., Aguirre Hernandez, R., Franchini, P., Parsa, K., Tessier-Cloutier, B., Lapuk, A., Huntsman, D., et al. (2018). Formalin fixation increases deamination mutation signature but should not lead to false positive mutations in clinical practice. PLoS One *13*, e0196434. https://doi.org/10.1371/journal.pone.0196434.
88. AlDubayan, S.H., Conway, J.R., Camp, S.Y., Witkowski, L., Kofman, E., Reardon, B., Han, S., Moore, N., Elmarakeby, H., Salari, K., et al. (2020). Detection of pathogenic variants with germline genetic testing using deep learning vs standard methods in patients with prostate cancer and melanoma. JAMA *324*, 1957–1969. https://doi.org/10.1001/jama.2020.20457.
89. Huang, K.-L., Mashl, R.J., Wu, Y., Ritter, D.I., Wang, J., Oh, C., Paczkowska, M., Reynolds, S., Wyczalkowski, M.A., Oak, N., et al. (2018). Pathogenic germline variants in 10, 389 adult cancers. Cell *173*, 355–370.e14. https://doi.org/10.1016/j.cell.2018.03.039.
90. Rahman, N. (2014). Realizing the promise of cancer predisposition genes. Nature *505*, 302–308. https://doi.org/10.1038/nature12981.
91. Stachler, M.D., Taylor-Weiner, A., Peng, S., McKenna, A., Agoston, A.T., Odze, R.D., Davison, J.M., Nason, K.S., Loda, M., Leshchiner, I., et al. (2015). Paired exome analysis of Barrett's esophagus and adenocarcinoma. Nat. Genet. *47*, 1047–1055. https://doi.org/10.1038/ng.3343.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Raw sequencing files	This paper	dbGaP study accession phs001939.v3.p1
Raw sequencing files (processed by GDC)	This paper	https://portal.gdc.cancer.gov/projects/CMI-MPC
Processed and deidentified sequencing and clinical data	This paper	https://www.cbioportal.org/study/summary?id=mpcproject_broad_2021
Processed and deidentified figure data and code	This paper	https://github.com/vanallenlab/mpcproject-paper
Study information and materials seen by patients	This paper	https://mpcproject.org/
Rural-area continuum codes (2010)	USDA[66]	https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes.aspx
Information on MPC patients nationwide (2018)	SEER[21]	https://seer.cancer.gov/data-software/
Medically underserved and health-physician shortage areas (accessed Dec 2021)	HRSA[23]	https://data.hrsa.gov/tools/shortage-area
National Area Deprivation Index 2019 data	Kind and Buckingham, 2018[67]	https://www.neighborhoodatlas.medicine.wisc.edu/
ClinVar (2019)	Landrum et al., 2018[32]	https://www.ncbi.nlm.nih.gov/clinvar/
Variant Effect Predictor GRCh37 Cache	McLaren et al., 2016[68]	https://useast.ensembl.org/info/docs/tools/vep/script/vep_cache.html
COSMIC germline cancer census gene set v86	Sondka et al., 2018[69]	https://cancer.sanger.ac.uk/census
Software and algorithms		
Python 3.8	Python Software Foundation, 2021[70]	https://www.python.org/
R 3.5.1	R Core Team, 2021[71]	https://www.r-project.org/
BWA	Li and Durbin, 2009[72]	http://bio-bwa.sourceforge.net/
GATK 3.7	McKenna et al., 2010[73]	https://github.com/broadinstitute/gatk/releases
Sequence alignment and alteration calling (component algorithms detailed below)	The Getz Laboratory	https://portal.firecloud.org/#methods/getzlab/CGA_WES_Characterization_Pipeline_v0.1_Dec2018/
Mutect v1.1.6	Cibulskis et al., 2013[74]	http://archive.broadinstitute.org/cancer/cga/mutect
FilterByOrientationBias	McKenna et al., 2010[73]	https://gatk.broadinstitute.org/hc/en-us/articles/360037060232
Strelka v2.8.0	Saunders et al., 2012[75]	https://github.com/Illumina/strelka
Oncotator v1.9.9.0	Ramos et al., 2015[76]	https://github.com/broadinstitute/oncotator
MutSig2CV	Lawrence et al., 2014[77]	https://github.com/getzlab/MutSig2CV
GATK 3.7 (CNV)	McKenna et al., 2010[73]	https://gatk.broadinstitute.org/hc/en-us/articles/360035531092
ABSOLUTE v1.5	Carter et al., 2012[78]	https://software.broadinstitute.org/cancer/cga/absolute_download
FACETS v0.6.2	Shen and Seshan, 2016[79]	https://github.com/mskcc/facets
GISTIC2.0 v2.0.23	Mermel et al., 2011[28]	https://github.com/broadinstitute/gistic2
DeTiN v2.0.1	Taylor-Weiner et al., 2018[80]	https://github.com/getzlab/deTiN
ContEst	Cibulskis et al., 2011[81]	https://software.broadinstitute.org/cancer/cga/contest
CrossCheckFingerprints (GATK 3.7)	McKenna et al., 2010[73]	https://gatk.broadinstitute.org/hc/en-us/articles/360037594711
ichorCNA	Adalsteinsson et al., 2017[56]	https://github.com/broadinstitute/ichorCNA
deconstructSigs (COSMIC v2 signatures, v1.9.0)	Rosenthal et al., 2016[33]	https://genomebiology.biomedcentral.com/articles/10.1186/s13059-016-0893-4

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
DeepVariant v0.8.0	Poplin et al., 2018[82]	https://github.com/google/deepvariant
PhylogicNDT	Leshchiner et al., 2018[46]	https://github.com/broadinstitute/PhylogicNDT
Other		
Repository for regenerating main study findings and figures of this paper	This paper	https://github.com/vanallenlab/mpcproject-paper, https://doi.org/10.5281/zenodo.6816267

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Eliezer M. Van Allen (Eliezerm_vanallen@dfci.harvard.edu).

Materials availability

This study did not generate any new unique reagents.

Data and code availability

The MPCproject releases deidentified clinical, patient-reported and research-grade genomic data into public repositories, such as cBioPortal: mpcproject_broad_2021 (https://www.cbioportal.org/study/summary?id=mpcproject_broad_2021), the Genomic Data Commons: CMI-MPC (https://portal.gdc.cancer.gov/projects/CMI-MPC), and dbGaP: phs001939.v3.p1 (https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001939.v3.p1) at regular intervals and prepublication. Data is processed and formatted as required by each repository's guidelines. All patient identifiers are stripped prior to data deposition to protect patient privacy. On the MPCproject data release webpage (https://mpcproject.org/data-release), patients can access project data, additional information about the data, a list of common terms used in research, methods used to generate the data, and an e-mail address for any additional data-related questions. All other data used in this paper are from publicly available resources. The code used to generate most main figures, central analyses, and supplementary figures can be found at can be found at https://github.com/vanallenlab/mpcproject-paper, except for figures and analyses requiring sample-level germline data. An unchanging version of the code at time of publication is also available at Zenodo: https://doi.org/10.5281/zenodo.6816267. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Patients who chose to enroll in this research study provided informed consent using a web-based consent form approved by the Dana-Farber/Harvard Cancer Center Institutional Review Board (DF/HCC Protocol 15-057B). Patient partners can exit the study at any time. All patient partners were male, with age and other features detailed in Table S2. If patient partners consented, FFPE exomes were requested from hospitals where they received treatment. Germline DNA was collected using mailed saliva collection kits. cfDNA from blood biopsies was collected through blood draws by medical providers or Quest Diagnostics (with a complimentary voucher), received by mail (Method details).

METHOD DETAILS

MPCproject website

The MPCproject utilizes a website (https://mpcproject.org/) to enroll patients through an online consent and release form. The website provides information about the project and advocacy groups that have partnered with the study. The website design, messaging, and workflow were developed with direct input from patient partners and advocates.

Informed consent

A link to the electronic informed consent document for formal enrollment in the study (https://mpcproject.org/ConsentAndRelease.pdf) was sent to registrant emails, and upon signing, a copy of the completed form was shared. At minimum, informed consent enabled study staff to request and abstract medical records, send a saliva kit directly to patients, perform sequencing on any returned saliva samples, and release de-identified integrated clinical, genomic, and patient-reported data for research use. Patient partners had the additional option to consent to study staff obtaining a portion of archived tumor tissue and/or a blood sample for further sequencing analysis.

Patient-reported data

After registering, patient partners completed a 17-question survey asking them about themselves and their disease (https://mpcproject.org/AboutYouSurvey.pdf). All questions were optional. Information on how question responses were standardized and categorized can be found in the supplemental methods.

Acquisition of medical records

Medical records were obtained for patient partners from the U.S. and Canada who completed the consent and medical release forms. Later in project development, a donated saliva or blood sample was also required. Study staff submitted medical record requests to all institutions and physician offices at which the patient reported receiving clinical care for their prostate cancer. A detailed medical record request form, along with the consent and release forms, were electronically faxed to each facility listed in a patient's release form. Medical records were returned to the project via mail, fax, or secure online portals. If a record request was not fulfilled in six months, study staff called the hospital, and a second request was submitted, with up to three requests made. Patient partners that communicated with study staff about changes in their treatment could request a medical record update, in which case their current hospital was again contacted for medical records. All medical records were saved in an electronic format to a secure drive at the Broad Institute.

Acquisition of patient samples

All consented patient partners living in the United States or Canada were mailed saliva kits with appropriate instructions, a sample tube labeled with a unique barcode, and a prepaid return box to send back the saliva sample. Samples were returned to the Broad Institute Genomics Platform, logged, and stored at room temperature (25 °C) until further sequencing.

If a consented patient partner opted into the blood biopsy component of the study, they were sent a blood kit with instructions (https://mpcproject.org/BloodSampleInstructions.pdf, Figure S4). Participants could take this kit to their next blood draw and request a courtesy draw by their medical provider; if a courtesy draw was not possible, patient partners could go to Quest Diagnostics with a complimentary voucher to have their blood drawn. Blood kits were returned free of charge to the Broad Institute Genomics Platform where they were fractionated into plasma and buffy coats and stored at −80 °C. If a patient partner did not provide a saliva sample, buffy coats were used to extract germline DNA for WES. Plasma samples continued to WES if ultra-low pass WGS detected a tumor fraction of circulating tumor DNA greater than 0.03. Some patient partners were selected to provide additional blood samples and were sent a new consent form. If they agreed to submit another blood sample, a new blood kit was shipped.

For patient partners that provided a germline sample and consented to the acquisition of some of their archival tumor tissue, study staff reviewed each patient's medical records and identified available tissue supplemental methods). Patient partners were screened by the study staff to determine if they had metastatic or advanced prostate cancer based on the definition by our study. If a patient partner had a sample that met the project's strict requesting criteria, study staff coordinated with that hospital's pathology department to fax a request for one H&E-stained slide as well as either 5–20 5-μm unstained slides or one formalin-fixed paraffin-embedded tissue block. Requests explicitly asked that the pathology department should not exhaust a sample to fulfill the request. Samples were sent to the MPCproject by mail. Tissue samples received as slides were labeled with unique barcode identifiers and submitted for whole exome sequencing. Tissue samples received as blocks were cut into three 30-μm scrolls per block, labeled with unique barcode identifiers, and then submitted for whole exome sequencing.

Medical record abstraction

A data dictionary comprising 60 clinical fields with possible options was curated by trained study staff working with prostate oncologists. Electronic health records were converted to searchable PDF files using the Optical Character Recognition (OCR) engine known as Tesseract.[83] Three study staff abstractors were involved in the abstraction and QC process for each record (supplemental methods). If a field had lack of concordance between abstractors or there were outstanding questions, a prostate cancer oncologist reviewed the content. Whenever possible, clinical data was abstracted directly from the records. For information that's not found, it was abstracted as 'NOT FOUND IN RECORD'. In instances where ambiguity or incomplete data was present, inferences were made considering the whole narrative of the medical record. Incomplete dates missing the day or month are abstracted as the first day of the month or first month of the year, respectively. While all medical records will eventually be abstracted, medical records from patient partners that received molecular sequencing of some form were prioritized for this study, resulting in 125 patient partners with medical record abstractions, 119 of which had at least one therapy noted. In examining the overlap between patient surveys and medical record therapies, we only considered therapies that were given for metastatic prostate cancer at least one week before the patient enrolled.

Geographic analysis

Using patient-reported data and secure Census Bureau geocoding, we identified residential census tracts for 628/706 patient partners.[84] To identify patient partners living in rural areas, this information was overlapped with rural-area continuum (RUCA) codes from the United States Department of Agriculture (USDA).[66] Census tracts with a secondary RUCA code greater than 3 were designated as rural. For comparison, the proportion of metastatic prostate cancer patients within each RUCA code from 2004 – 2017 was taken from Surveillance, Epidemiology, and End Results (SEER) using SEER*stat with the following selection table: {Site and

Morphology.Site recode ICD-O-3/WHO 2008} = 'Prostate' AND {Stage - Summary/Historic.SEER Combined Summary Stage 2000 (2004-2017)} != 'In situ', 'Localized only', 'Not applicable', 'Unknown/unstaged/unspecified/DCO', 'Blank(s)'.[21] To identify patient partners living in medical shortage areas, census tracts were overlapped with primary care health physician shortage areas (HPSA) and medically underserved areas (MUA) defined by the Health Resources and Services Administration (HRSA).[23] Census tracts were labelled as existing within a MUA or HPSA if they were designated as within a medically underserved area/population or within a primary care HPSA, respectively. Published geographic datasets of cancer patients (e.g., SEER, NPCR) do not contain census-tract resolved data or summary results of MUA/HPSA status, so for comparison we instead used the total U.S. population living in HPSAs and MUAs, taken from HRSA, divided by the entire U.S. population taken from the U.S. Census.[23,24] To calculate appointment distances, we calculated the round-trip Haversine distances between residential zip codes and the zip code of reported institutions. To assess socioeconomic advantage, we used secure Census Bureau geocoding to identify residential census block groups (12 digit FIPS codes) and cross-referenced them with a publicly available dataset of Area Deprivation Index (https://www.neighborhoodatlas.medicine.wisc.edu/download).[67] We used the National ADI, which ranks neighborhoods by percentiles (1–100), with 100 indicating the highest level of disadvantage.

To protect privacy, geographic locations in the graphical abstract do not represent real patient partner residential areas. Random counties from the state of each reported residential area are shown instead.

Whole exome sequencing analysis

Whole exome sequences were captured using Illumina technology and the sequence data processing and analysis was performed using Picard and FireCloud pipelines on Terra (https://terra.bio/) (supplemental methods). The Picard pipeline (http://picard.sourceforge.net) was used to produce a BAM file with aligned reads. This includes alignment to the GRCh37 human reference sequence using BWA[72] and estimation and recalibration of base quality score with the Genome Analysis Toolkit (GATK).[73] Somatic alterations for tumor samples were called using a customized version of the Getz Lab CGA WES Characterization pipeline (https://portal.firecloud.org/#methods/getzlab/CGA_WES_Characterization_Pipeline_v0.1_Dec2018/) developed at the Broad Institute. Briefly, MuTect v1.1.6 algorithm was used to identify somatic mutations.[74] Somatic mutation calls were filtered using a panel of normals (PoN), oxoG filter and an FFPE filter to remove artifacts introduced during the sequencing or formalin fixation process.[85] Small somatic insertions and deletions were detected using the Strelka algorithm.[75] Somatic mutations were annotated using Oncotator.[76] Recurrently altered mutations were identified using MutSig2CV.[77] To define somatic copy ratio profiles, we used GATK CNV.[73] To generate allele-specific copy number profiles and assess tumor purity and ploidy, we used ABSOLUTE and FACETS.[78,79] Final segmentation calls were taken from ABSOLUTE, except for the X chromosome, which was taken from FACETS. We utilized GISTIC2.0 to identify significantly recurrent amplification and deletion peaks.[28] For determining allele-specific copy number alterations, we assessed the absolute allelic copy numbers of the segment containing each gene. Mutation burden was calculated as the total number of mutations (non-synonymous + synonymous) detected for a given sample divided by the length of the total genomic target region captured with appropriate coverage from whole exome sequencing.

Whole exome sequencing quality control

Samples with average coverage below 55x in the tumor sample or below 30x in the normal sample were excluded. Samples with purity <0.10 from both ABSOLUTE and FACETS were excluded. DeTiN was applied to samples to estimate the amount of tumor contamination in the normal samples; samples with TiN (tumor in normal) > 0.25 were excluded.[80] ContEst was applied to measure the amount of cross-sample contamination in samples; samples with contamination >0.04 were excluded.[81] The Picard task CrossCheckFingerprints was applied to determine sample mixups; samples with Fingerprints LOD value <0 were excluded.[86] Two FFPE samples that failed sequence processing and were noted to have extensive segment fragmentation and allelic imbalance were also excluded due to suspicion of poor sequencing. A table of samples with quality control metrics for each sample can be found in the Supplementary Data. Samples which passed quality control were submitted to cBioPortal and GDC.

Ultra-low pass whole genome sequencing analysis

ichorCNA was used to assess the tumor fraction in cfDNA samples that completed ultra-low pass whole genome sequencing.[56] The log copy ratio of *AR* was assessed by the log copy ratio of the genomic interval containing *AR.* This value could not consistently be converted to absolute copy number due to the low tumor fractions of many samples.

Mutational signature analysis and kataegis

Mutational processes in our cohort were determined using deconstructSigs with default parameters applying COSMIC v2 signatures as the reference with a maximum number of signatures of 6[29,30]. A signature was assessed as present if the signature contribution was greater than 6%. Because tumor samples were formalin-fixed and paraffin embedded (FFPE), a process known to introduce stranded mutational artifacts in specific nucleotide contexts, we used a filter to remove likely FFPE artifacts according to nucleotide context and strand bias before using deconstructSigs.[87] We also tried to assess the colocalization of the kataegis event with structural variant breakpoints but were limited by targeted sequencing in exomes and low coverage in ULP-WGS. *KMT2C* and its surrounding region were not copy number altered in the sample with kataegis. Kataegis was not identified in any other sample.

Germline variant discovery

To call short germline single-nucleotide polymorphisms, insertions, and deletions from germline WES data, we used DeepVariant (v0.8.0).[82,88] Specifically, we used the publicly-released WES model (https://console.cloud.google.com/storage/browser/deepvariant/models/DeepVariant/0.8.0/DeepVariant-inception_v3-0.8.0+data-wes_standard/) to generate single-sample germline variant call files using the human genome reference GRCh37(b37). We filtered variants with bcftools v1.9 to only keep high-quality variants annotated as "PASS" in the "FILTER" column. The high-quality variants were merged into single-sample Variant Call Format (VCF) files using CombineVariants from GATK 3.7 (https://github.com/broadinstitute/gatk/releases). To decompose multiallelic variants and normalize variants, we used the computational package vt v3.13 (https://github.com/atks/vt). Lastly, germline variants were annotated using the VEP v92 with the publicly-released GRCh37 cache file (https://github.com/Ensembl/ensembl-vep).[68] An alteration was also considered if there was a pathogenic germline alteration, denoted by "Pathogenic", "Pathogenic/Likely_pathogenic", "Likely_pathogenic", "_risk_factor", or "Conflicting_interpretations_of_pathogenicity" (if at least one expert source indicated "Likely_pathogenic" or "Pathogenic") in ClinVar (Dec 2019 version).[32] An alteration was also considered if it had an "HIGH" predicted impact on protein function and had a maximum allele fraction of <0.01 in all populations. The germline cancer predisposition genes were selected based on the level of evidence supporting their Mendelian disease susceptibility. This is composed of the well-curated COSMIC germline cancer census gene set (v86; http://cancer.sanger.ac.uk/census) and the germline cancer gene set listed in Huang et al. 2018 and Rahman 2014.[30,69,89,90]

Association of DNA-repair alterations and presence of signature 3

Alterations in a select list of genes previously implicated in DNA-repair were examined (Table S3). An alteration was considered if there was a somatic single-copy deletion, double deletion, nonsense mutation, missense mutation, frameshift indel, or splice site mutation. An alteration was also considered if there was a pathogenic germline alteration. An alteration was considered biallelic for Figure S7 if there was a double somatic deletion, a pathogenic germline/protein-altering somatic variant plus a somatic loss, or more than one mutation in the same gene, although we cannot confirm the biallelic nature of multiple mutations.

Phylogenetic analysis

To compare mutations between distinct samples (tumor and cfDNA) from the same patient, we used a previously described method designed to recover evidence for mutations called in one sample in all other samples derived from the same individual.[91] In brief, the 'force-calling' method uses the strong prior of the mutation being present in at least one sample in the patient to detect and recover mutations that might otherwise be missed. A mutation was deemed tumor/cfDNA specific if there were no force-called reads that supported the mutation in the other sample, although this process underestimates the proportion of shared mutations in low purity tumors. The cancer cell fraction (CCF) of mutations were defined using ABSOLUTE, which calculates the CCF based on variant allele frequency, purity, and local allelic copy number.[78] To reconstruct tumor phylogenies, we used PhylogicNDT, which clusters mutations into subclones across multiple samples based on their underlying similar CCFs.[46]

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis

Except where otherwise specified, analysis and data visualization were performed with Python 3.8, SciPy v.1.5.2, Matplotlib v.3.3.2, seaborn v.0.11.0 and R v.3.5.1.[90,91] The code used to generate most main figures, analyses, and supplementary figures can be found at https://github.com/vanallenlab/mpcproject-paper or Zenodo: https://doi.org/10.5281/zenodo.6816267, except for figures and analyses requiring sample-level germline data. Between-group comparisons of continuous variables were performed with the Mann-Whitney U test (Wilcoxon rank sum test) or Student's *t*-test. Contingency table tests were performed with Fisher's exact test. All tests were two-sided.

ADDITIONAL RESOURCES

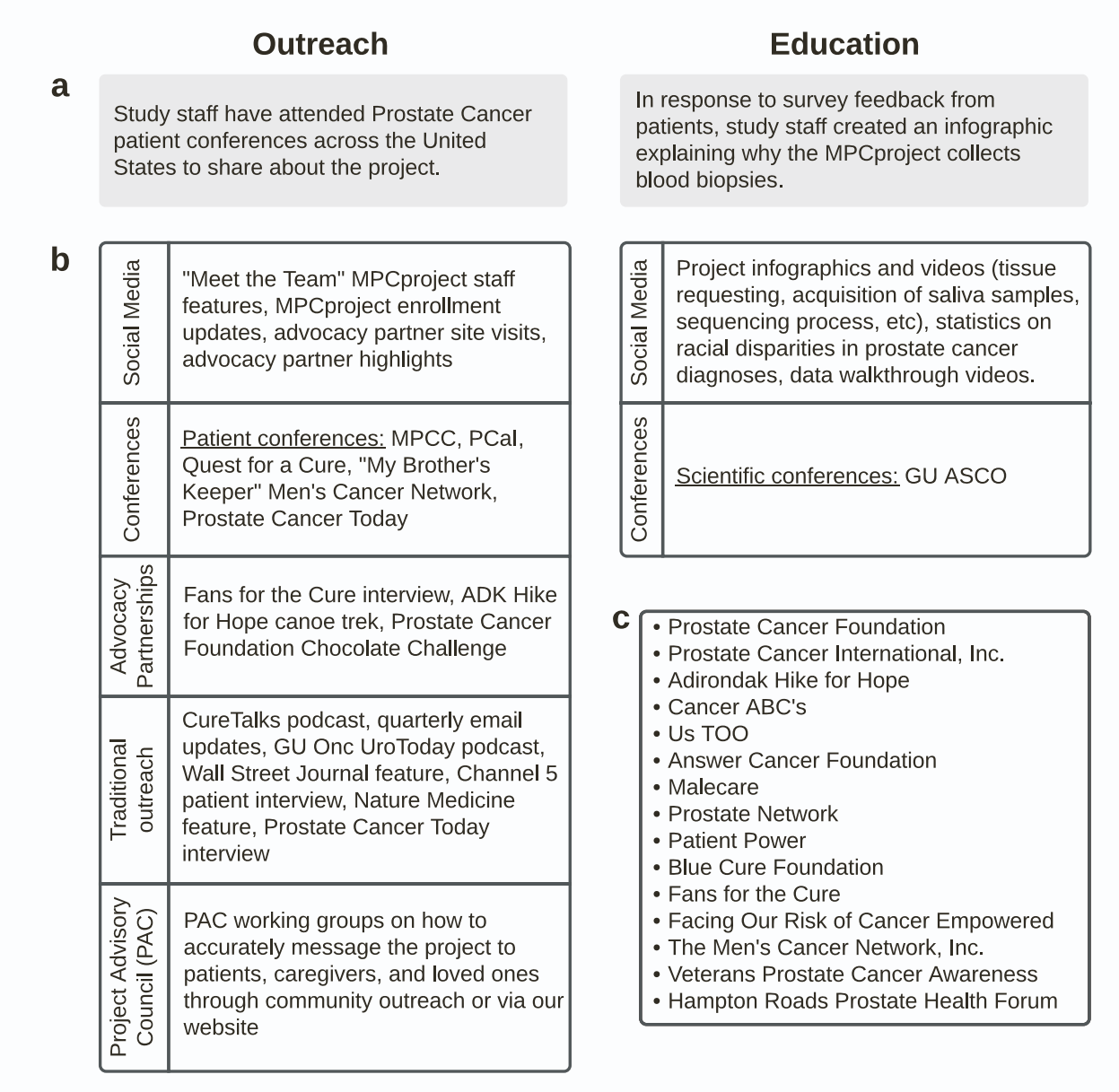
MPCproject website: https://mpcproject.org/.

Cell Genomics, Volume *2*

Supplemental information

A patient-driven clinicogenomic partnership for metastatic prostate cancer

Jett Crowdis, Sara Balch, Lauren Sterlin, Beena S. Thomas, Sabrina Y. Camp, Michael Dunphy, Elana Anastasio, Shahrayz Shah, Alyssa L. Damon, Rafael Ramos, Delia M. Sosa, Ilan K. Small, Brett N. Tomson, Colleen M. Nguyen, Mary McGillicuddy, Parker S. Chastain, Meng Xiao He, Alexander T.M. Cheung, Stephanie Wankowicz, Alok K. Tewari, Dewey Kim, Saud H. AlDubayan, Ayanah Dowdye, Benjamin Zola, Joel Nowak, Jan Manarite, Idola Henry Gunn, Bryce Olson, Eric S. Lander, Corrie A. Painter, Nikhil Wagle, and Eliezer M. Van Allen



Supplementary Figure S1. MPCproject education and outreach initiatives reach patient-partners across the country, related to Fig. 1A.

a) Education and outreach spotlights. Study staff attend and present at patient conferences to share information about the MPCproject with the extended prostate cancer community. Conference tables have example sample kits, brochures, and a mailing list sign-up to learn more. For patients who follow the MPCproject on social media, study staff create online polls to

identify educational content important to the community. One such poll revealed interest in learning about the biological significance of liquid biopsies and why the project collects them. **b)** Select examples of outreach and education initiatives. As a result of the decentralized, online nature of the study, the MPCproject uses diverse modes of education and outreach to reach patient-partners. **c)** The MPCproject partners with patient advocacy groups across the United States and Canada. Advocacy partners help encourage patient participation in the project as well provide ongoing input regarding the design and implementation of the project overall.

a

Please fill out as much as you can. All questions are optional. You can return at any time with the link sent to you by email.

1. When were you first diagnosed with prostate cancer? If you do not remember the month, you can enter just the year.

   Choose month... | Choose year...

2. When you were first diagnosed, were you diagnosed with advanced or metastatic prostate cancer (prostate cancer that has spread beyond the prostate, including biochemical recurrence)?

   - (•) Yes
   - ( ) No
   - ( ) I don't know

3. Did you receive local treatment to your prostate when you were first diagnosed (local treatment includes surgery, radiation, or cryotherapy)?

   - (•) Yes
   - ( ) No
   - ( ) I don't know

4. Have you had your entire prostate surgically removed (known as a prostatectomy)?

   - (•) Yes
   - ( ) No
   - ( ) I don't know

5. Where is your prostate cancer currently located (check all that apply)?

   - [x] Lymph Node
   - [ ] Bone
   - [ ] Liver
   - [ ] Lung
   - [ ] Brain
   - [x] Other

   Please provide details

   - [ ] No Evidence of Disease (NED)
   - [ ] I don't know

6. For your advanced prostate cancer (prostate cancer that is outside of the prostate), please check off all therapies that you have previously received or are currently receiving (Check all that apply)

   Hormones
   ...
   Chemotherapy
   ...
   Other Therapy
   ...

   See Supplementary Table 4 for therapy list

   Experimental/Clinical Trial

   - [x] Experiment/Clinical Trial

   Please provide details

   - [x] Other

   Please provide details

7. Please list additional medications, alternative medications, you've taken or lifestyle changes that you've made since your diagnosis with prostate cancer.

8. Have you had any other types of cancer?

   - (•) Yes
   - ( ) No
   - ( ) I don't know

9. What other cancer(s) have you had?

10. Do you have any family history of prostate and/or breast cancer?

    - (•) Yes
    - ( ) No
    - ( ) I don't know

11. How did you find out about this project?

12. Is there anything else you would like us to know about your prostate cancer?

13. Do you consider yourself Hispanic, Latino or Spanish?

    - (•) Yes
    - ( ) No
    - ( ) I don't know

14. What is your race (select all that apply)?

    - [ ] American Indian or Native American
    - [ ] Japanese
    - [ ] Chinese
    - [ ] Other East Asian
    - [ ] South East Asian or Indian
    - [ ] Black or African American
    - [ ] Native Hawaiian or other Pacific Islander
    - [ ] White
    - [ ] I prefer not to answer
    - [x] Other

    Please provide details

15. In what year were you born?

    Choose year...

16. What country do you live in?

    Choose country...

17. What is your ZIP or postal code?

    Zip Code

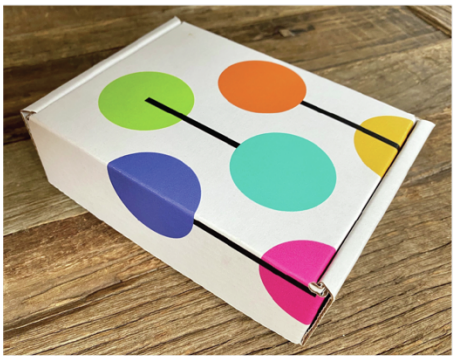
I understand that the information I entered here will be stored in a secure database and may be used to match me to one or more research studies conducted by the Metastatic Prostate Cancer Project. If the information that I entered matches a study being conducted by the Metastatic Prostate Cancer Project, either now or in the future, I agree to be contacted about possibly participating. I understand that if I would like my information deleted from the database, now or in the future, I can email info@mpcproject.org and my information will be removed from the database.

SUBMIT

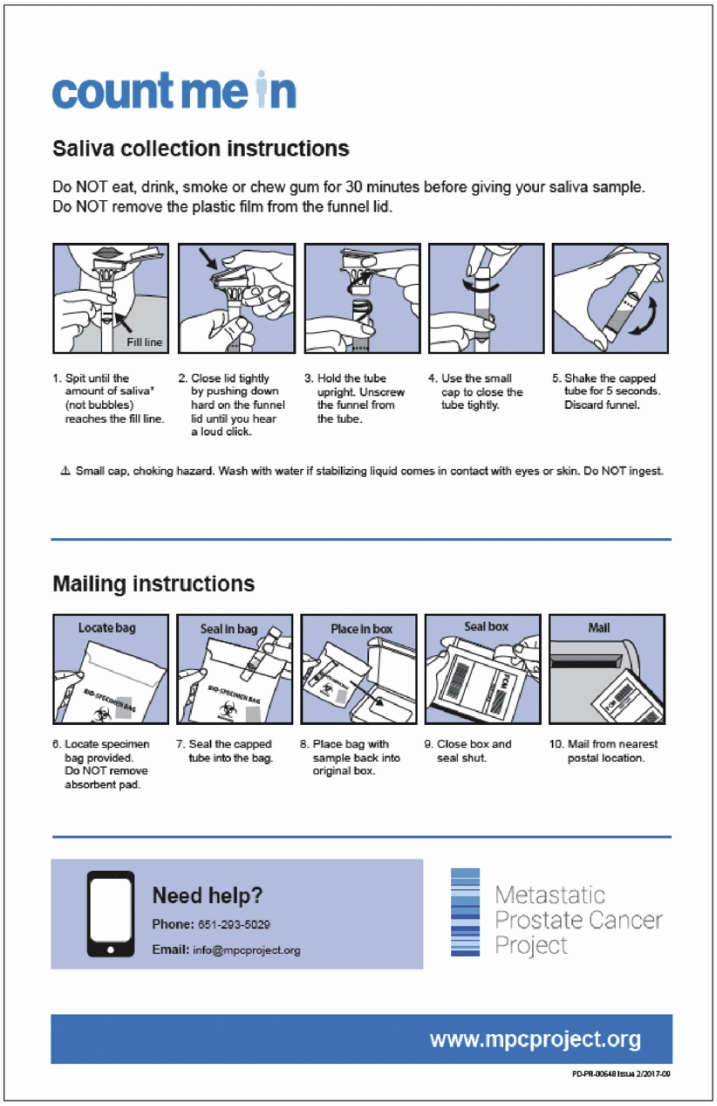
Supplementary Figure S2. MPCproject About You Intake Survey, related to STAR Methods.

a) After registering, patient-partners complete an online intake survey detailing their experience with metastatic prostate cancer (https://mpcproject.org/AboutYouSurvey.pdf). All questions are optional. Questions were developed in collaboration with patient-partners and practicing prostate cancer oncologists. For a full list of therapies for question 6, see Supplementary Table S4. The survey responses above are shown as an example and do not represent any specific patient-partner's responses.

a



b



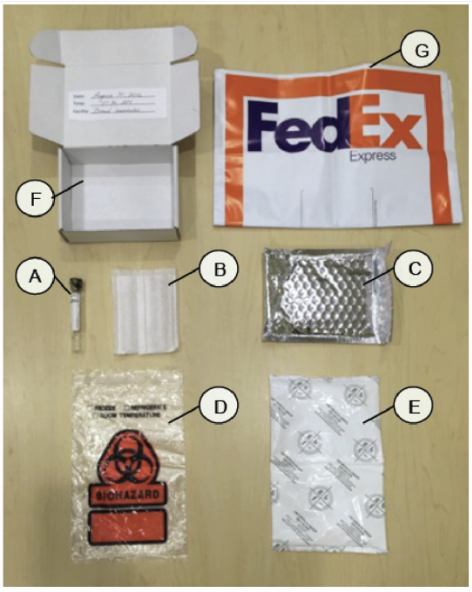
Supplementary Figure S3. MPCproject remote saliva donation kit, related to STAR Methods.

a) Enrolled patients in the U.S. and Canada are mailed a saliva kit. Each kit comes with a tube for saliva donation and a prepaid FedEx return envelope. All components of the kit, including the box itself, contain a unique, nonidentifiable barcode associated with the patient-partner. Acting on feedback about privacy from patient-partners and advocates, boxes are kept nondescript to avoid identifying the recipient as a patient with prostate cancer.

b) Saliva kit instructions. These instructions are included in the box itself, and patient-partners can contact the MPCproject study team for additional assistance if necessary.

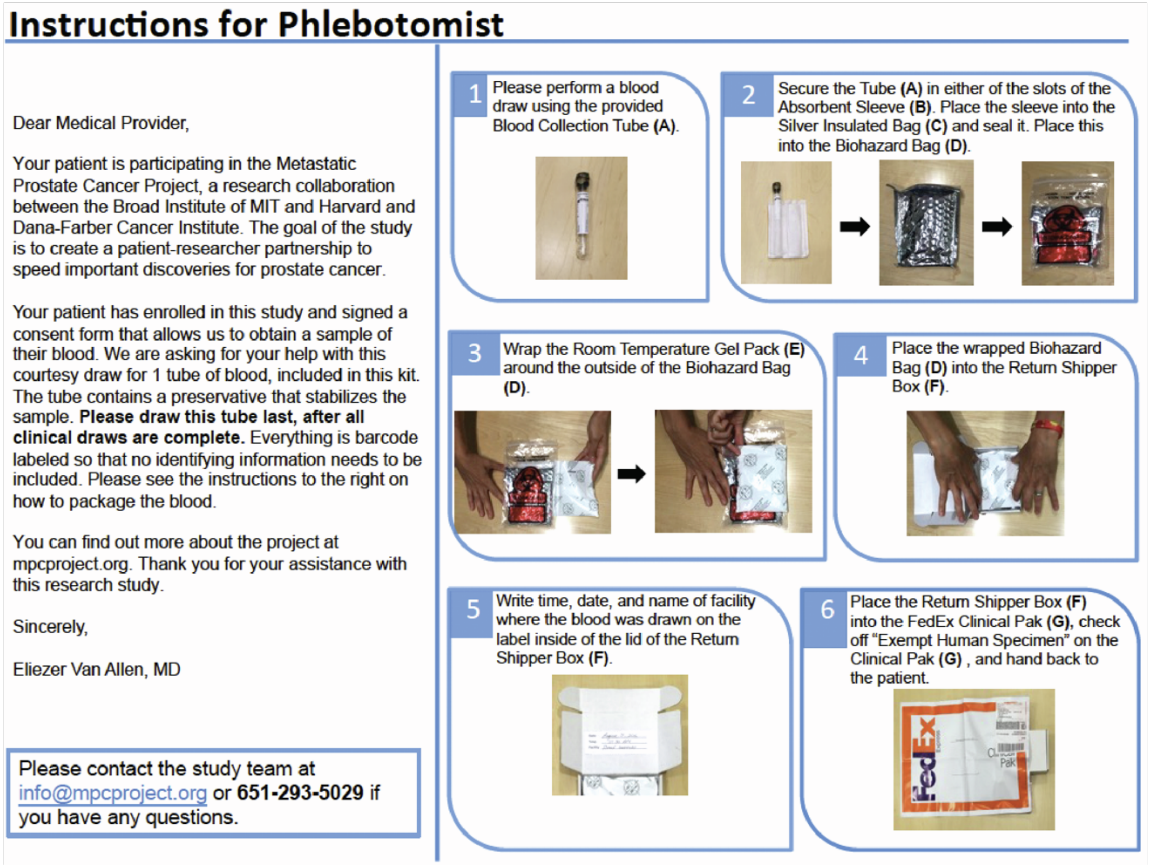
a

b



A: Blood collection tube
B: Absorbent sleeve
C: Silver insulated bag
D: Biohazard bag
E: Room temperature gel pack
F: Return box
G: FedEx clinical pack

c



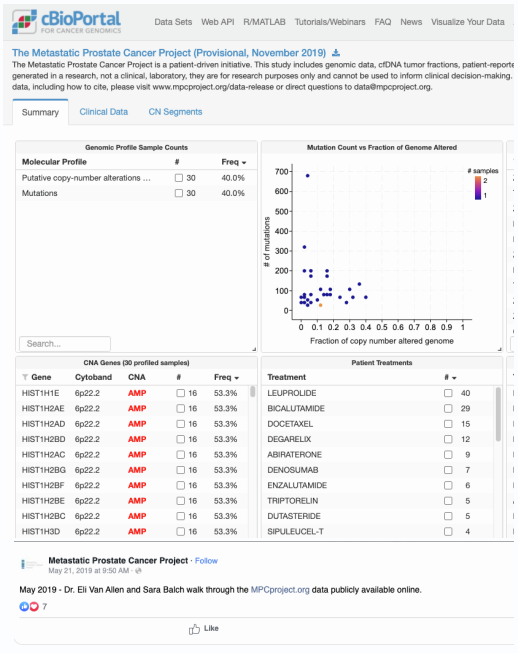
Supplementary Figure S4. MPCproject blood donation kit, related to STAR Methods.

a) If they consented to donate blood on their online survey, patient-partners are mailed a blood kit. Each kit comes with a tube for blood donation, instructions for use, and a unique, nonidentifiable barcode. Acting on feedback about privacy from patient-partners and advocates, boxes are kept nondescript to avoid identifying the recipient as a patient with prostate cancer. **b)** Composition of blood donation kit. This graphic is included within the blood donation kit. **c)** Instruction for healthcare providers. Patient-partners provide these instructions to their healthcare provider or phlebotomist at regular, standard of care blood draws. A courtesy draw is requested, free of charge, but if this is not available, patient-partners can also visit a local Quest Diagnostics lab with a free voucher for a blood draw. After completion, the kit is placed within the prepaid FedEx envelope and mailed to the Broad Institute where it is kept for sequencing.

a

Patient-partner Concern/Feedback	MPCproject Team Response
I would like to donate tissue, but I am starting a trial that may need it in the future. *Email*	Worked directly with hospital pathologist to ensure tissue remained, kept regular communication with patient throughout request process.
I cannot get the online form to work. *Email, Phone*	Talked with patient on phone, sent paper versions of forms with prepaid envelopes to patient's home.
How do I get my blood drawn? My doctor would not give a courtesy draw. What is my blood used for? *Email*	Patient was walked through process of free Quest Diagnostic blood draw. Graphics created to explain how donated blood is used.
I recently had a large change in my treatment regimen. Can you update my medical records? *Email*	Medical records rerequested from patient's current hospital.
I want to participate, but I don't want those close to me to know I have prostate cancer. *Email*	Working with patients and advocates, blood and saliva kits redesigned to be nondescript for privacy.

b



c

Metastatic Prostate Cancer Project

Dear MPCproject Mailing List,

We are writing to share another quarterly update with the patients, caregivers, scientists, and advocacy partners engaged in the Metastatic Prostate Cancer Project (MPCproject.org). We are tremendously grateful for your continued support. Below you will find some recent news from the MPCproject.

The Numbers

Over 745 patients have clicked "Count Me In" to register for the project. Thanks to your generosity, we have received:

- 662 medical records
- 417 saliva kits
- 329 blood samples

Understanding Our Data

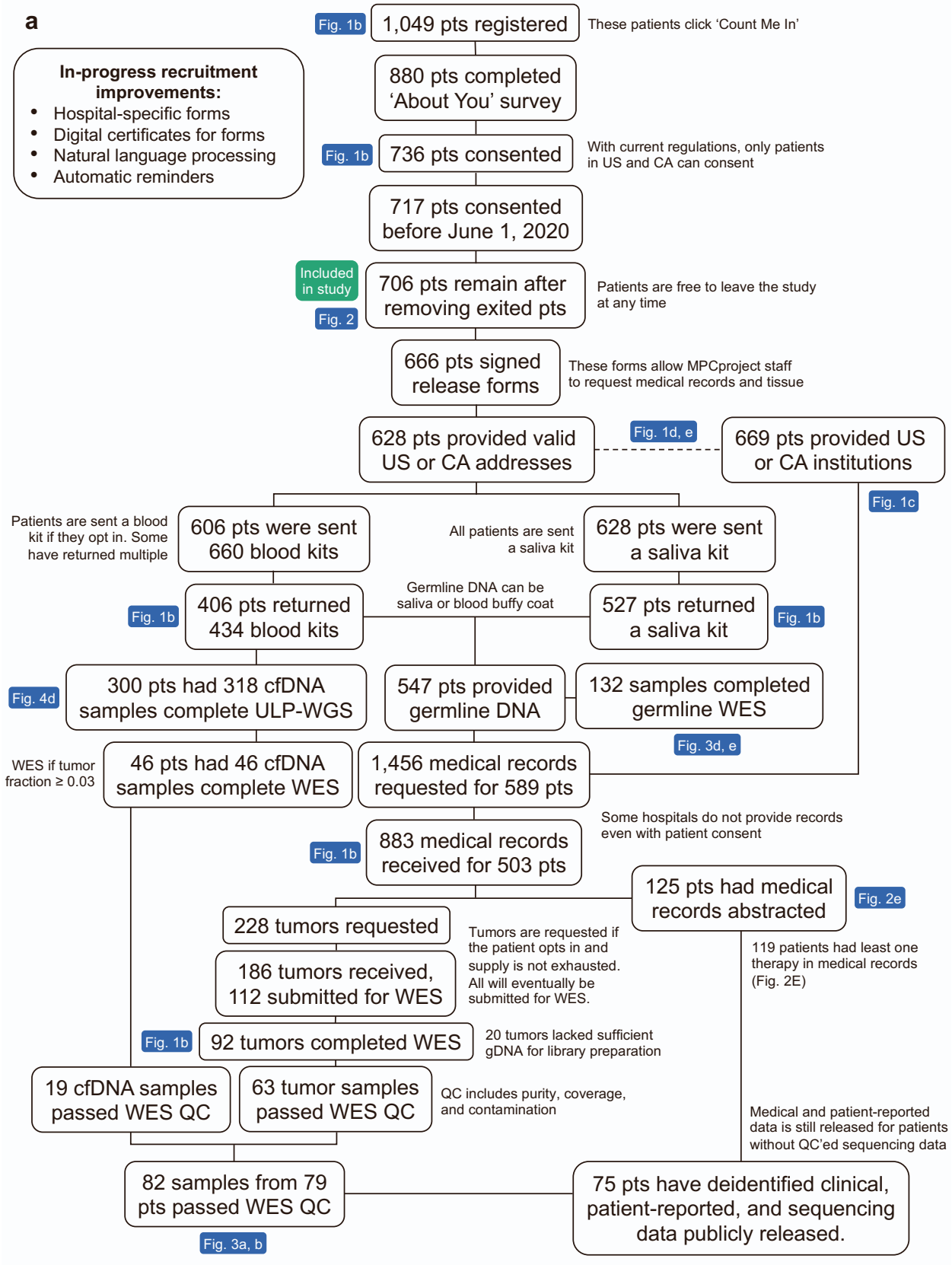
The MPCproject recently released a video of Dr. Eli Van Allen walking through the genomic and clinical data we have released thus far on cBioPortal. We will continue to release data as it is generated, pre-publication. If you have any questions regarding the data currently available, please email us at data@mpcproject.org.

Supplementary Figure S5. Working directly with patients in the MPCproject, related to Fig. 1A.

a) Examples of feedback from patient-partners and the response of the project team. In each case, patient-partners contacted the MPCproject office with concerns, questions, or feedback. The MPCproject study staff maintains regular contact with patient-partners that have questions and creates infographics and educational materials based on common questions.

b) Walkthrough of initial MPCproject data on cBioPortal. When the project's first data release happened on cBioPortal, Dr. Van Allen and the study team recorded a walkthrough (https://m.facebook.com/watch/?v=471939353546532) explaining the shared MPCproject data to patient-partners.

c) Quarterly email updates. An example of a quarterly update sent four times a year to patient-partners, loved ones, and advocates on the MPCproject mailing list. These emails explain study progress, how to interpret data releases, and new project initiatives.

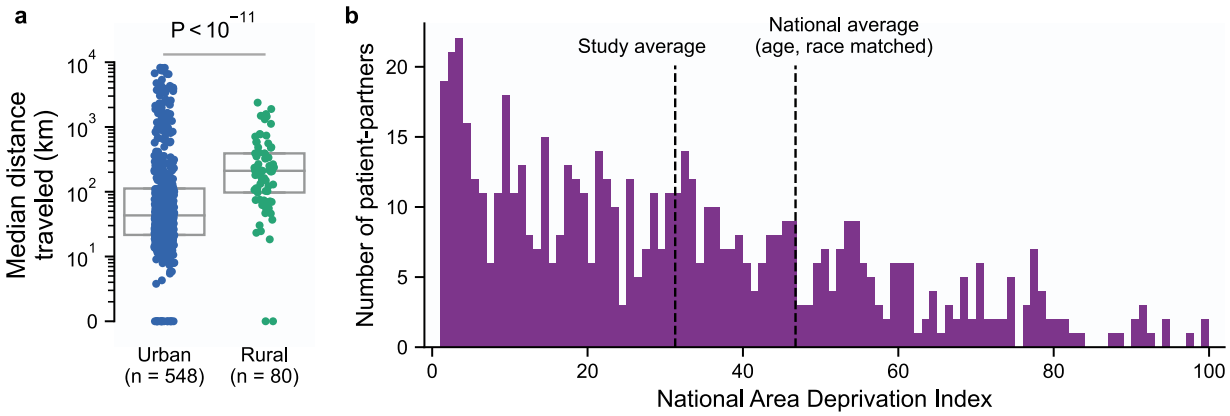

a
In-progress recruitment improvements:
• Hospital-specific forms
• Digital certificates for forms
• Natural language processing
• Automatic reminders
Fig. 1b
1,049 pts registered
These patients click 'Count Me In'
880 pts completed 'About You' survey
Fig. 1b
736 pts consented
With current regulations, only patients in US and CA can consent
717 pts consented before June 1, 2020
Included in study
Fig. 2
706 pts remain after removing exited pts
Patients are free to leave the study at any time
666 pts signed release forms
These forms allow MPCproject staff to request medical records and tissue
628 pts provided valid US or CA addresses
Fig. 1d, e
669 pts provided US or CA institutions
Fig. 1c
Patients are sent a blood kit if they opt in. Some have returned multiple
606 pts were sent 660 blood kits
All patients are sent a saliva kit
628 pts were sent a saliva kit
Germline DNA can be saliva or blood buffy coat
Fig. 1b
406 pts returned 434 blood kits
527 pts returned a saliva kit
Fig. 1b
Fig. 4d
300 pts had 318 cfDNA samples complete ULP-WGS
547 pts provided germline DNA
132 samples completed germline WES
Fig. 3d, e
WES if tumor fraction ≥ 0.03
46 pts had 46 cfDNA samples complete WES
1,456 medical records requested for 589 pts
Some hospitals do not provide records even with patient consent
Fig. 1b
883 medical records received for 503 pts
125 pts had medical records abstracted
Fig. 2e
228 tumors requested
Tumors are requested if the patient opts in and supply is not exhausted. All will eventually be submitted for WES.
186 tumors received, 112 submitted for WES
119 patients had least one therapy in medical records (Fig. 2E)
Fig. 1b
92 tumors completed WES
20 tumors lacked sufficient gDNA for library preparation
19 cfDNA samples passed WES QC
63 tumor samples passed WES QC
QC includes purity, coverage, and contamination
Medical and patient-reported data is still released for patients without QC'ed sequencing data
82 samples from 79 pts passed WES QC
Fig. 3a, b
75 pts have deidentified clinical, patient-reported, and sequencing data publicly released.

59

Supplementary Figure S6. MPCproject attrition chart, related to Fig. 1B.

a) Chart detailing project attrition for patient-partners that consented as of June 1, 2020. The chart represents data collected on June 7, 2021. Patient recruitment, sample acquisition, medical record abstraction, sequencing, and data releases are ongoing processes, so these values will grow as the project continues. Colored boxes indicate the figures that use those values in analysis and visualization. Values for Fig. 1B shown in this attrition chart may be greater than those shown in Fig. 1b at the study cutoff date, as Fig. 1B is a snapshot showing values collected as of June 1, 2020, while this attrition chart includes steps that may have been completed by consented patient-partners after June 1, 2020.

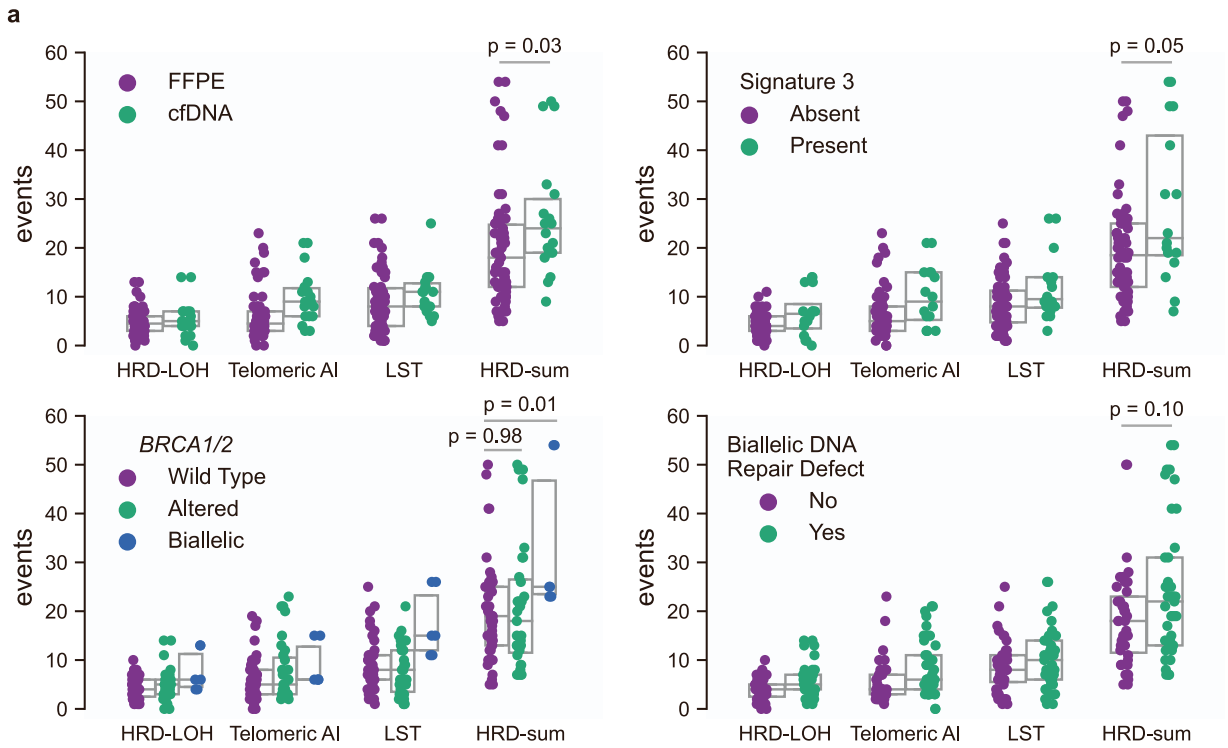
b) Real-time and proposed improvements to improve participant recruitment. Several steps in the study process have been targeted for improvement, with implement and potential changes listed.



Supplementary Figure S7. Additional geographic characteristics of patient-partner residential areas, related to Fig. 1D, E.

a) Patient-partners in rural areas travel farther for clinical care. Patient-partner residential areas were categorized as rural or urban based on USDA rural-urban continuum codes. For each patient-partner, the median Haversine round-trip distance between the zip code of their home address and that of institutions they visited was calculated (Methods). Patient-partners that live in Canada (n = 30), did not provide a residential area (n = 40), or provided only a P.O. box (n = 8) are not shown. *P*-value calculated via two sided Mann-Whitney U test.

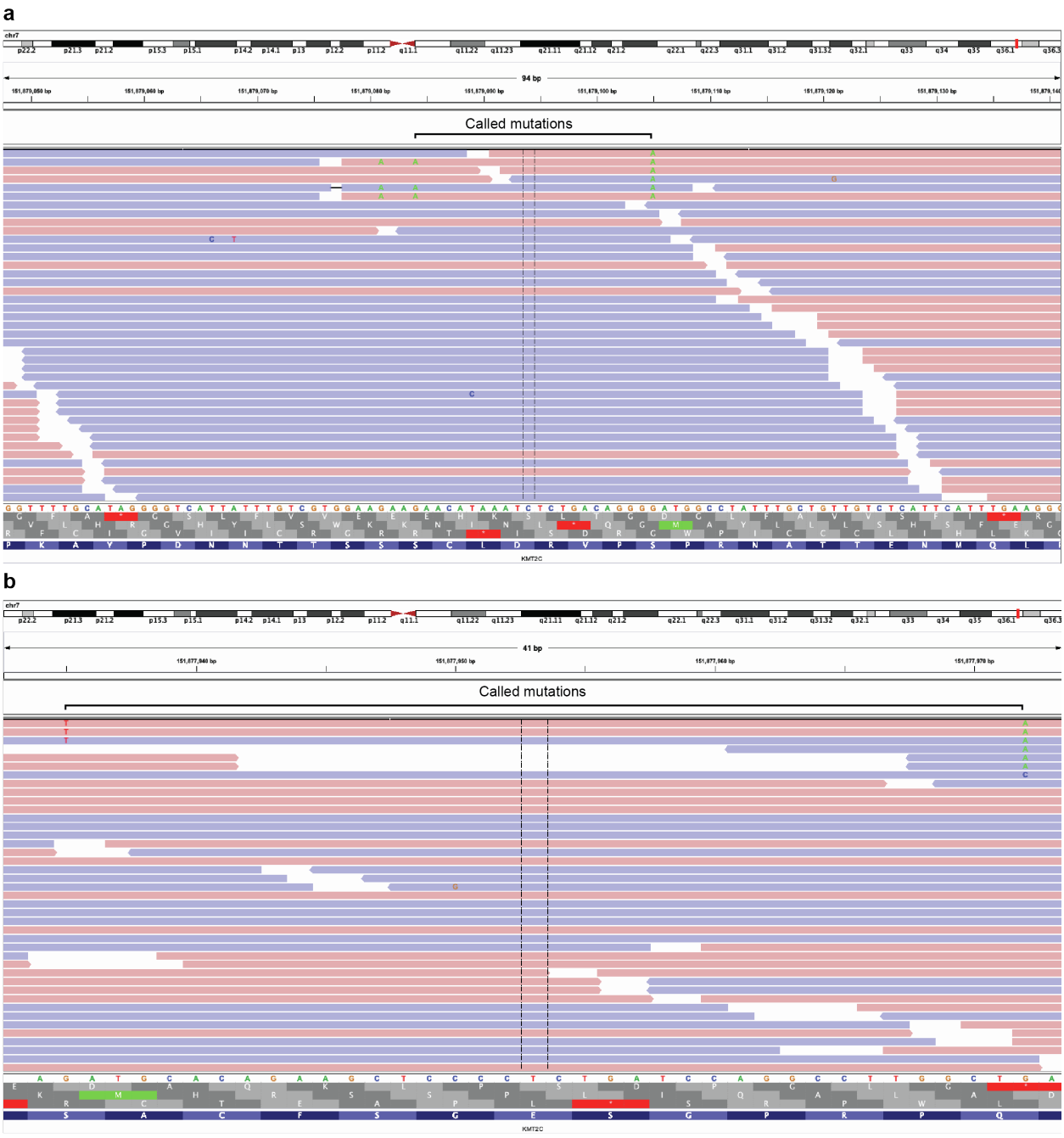
b) National Area Deprivation Index (ADI) distribution of patient-partner residential areas. Higher values indicate higher degrees of disadvantage. The national average was calculated by overlapping census block group populations counts from the U.S. census with public ADI data, weighting by age and race matched to the distribution of age and race of patient-partners.



Supplementary Figure S8. Associations between sample characteristics and scarHRD score, related to Fig. 3B.

a) Copy number based events associated with homologous recombination deficiency were assessed using scarHRD[1] (Methods). These include the number of large (> 15 Mb) loss-of-heterozygosity events (HRD-LOH), the number of allelic imbalances that extend to the end of the chromosome (Telomeric AI), the number of chromosomal breaks between adjacent regions of at least 10 Mb, with a distance between them not larger than 3Mb (LST), and the sum of all the previous listed events (HRD-sum). Biallelic DNA repair defects were classified as 1) a double deletion, 2) a loss with a protein-altering somatic or pathogenic germline mutation, or 3) more than one protein-altering somatic/pathogenic germline mutation, although we cannot confirm the biallelic nature of double mutations. Genes used in the biallelic DNA repair defect association found in Supplementary Table S3. The association with cfDNA is consistent with

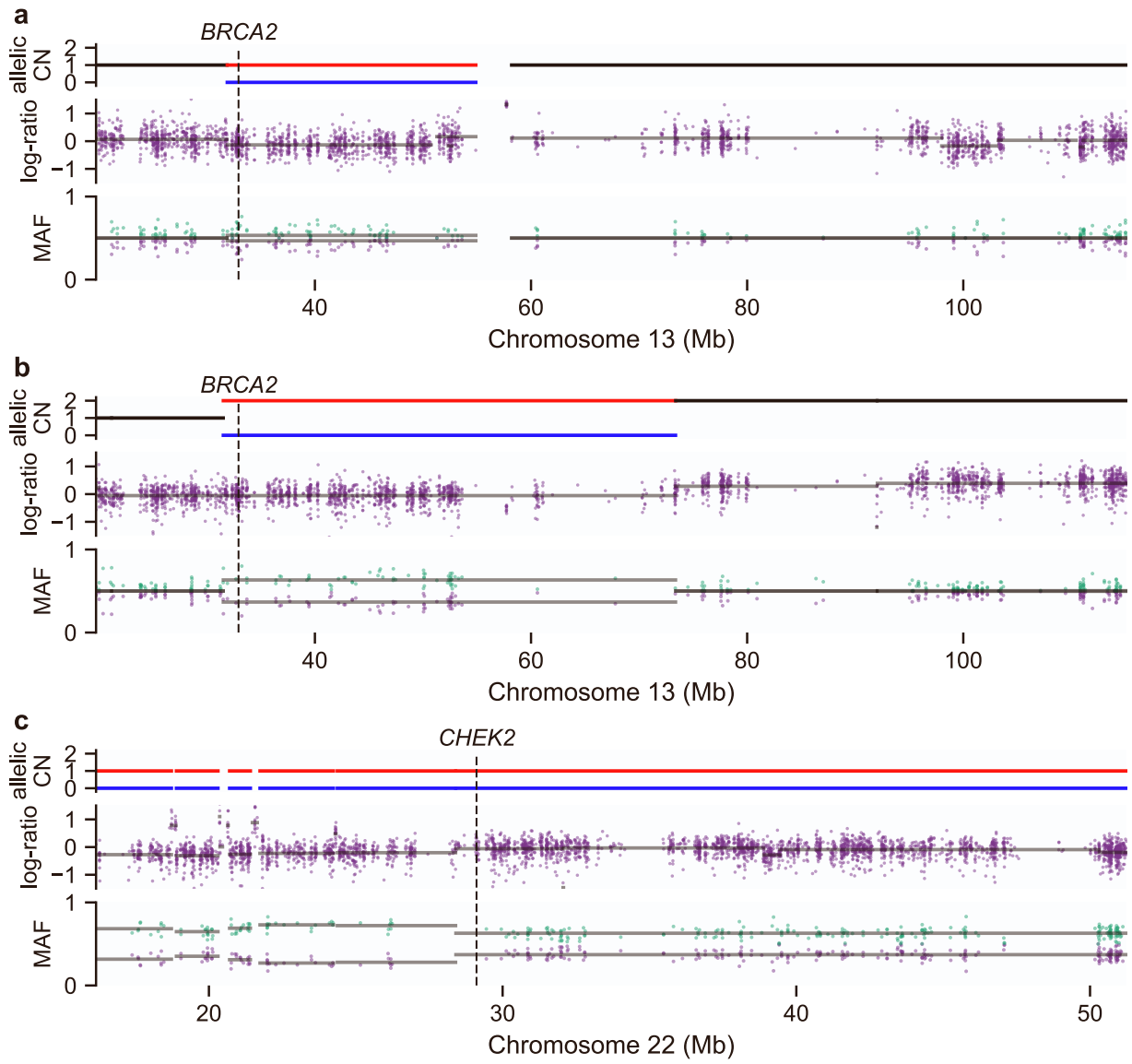
prior literature, as the cfDNA samples represent metastatic disease, although we cannot rule out the confounding influence of analyte type.



Supplementary Figure S9. IGV screenshots of *KMT2C* mutation-sharing reads, related to Fig. 3C.

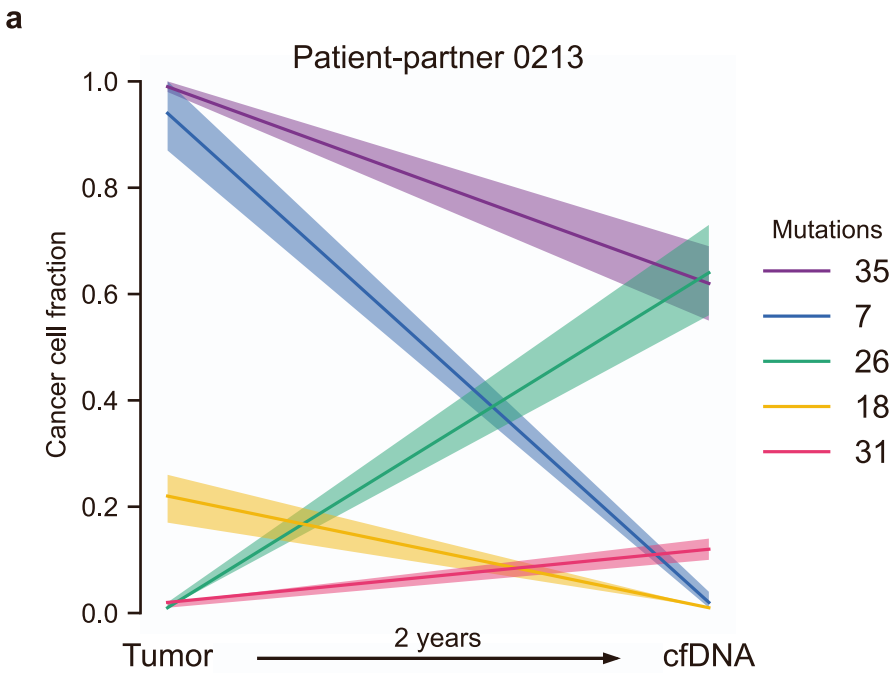
a) IGV screenshot containing reads that span somatic *KMT2C* mutations (chr7:151879084/p.S1947F and chr7:151879105/p.S1954F) in the cfDNA sample of patient-

partner 0203. A mutation may also be present at chr7:151879081 but was rejected by Mutect's internal filters as it is close to an inferred gap event. Coloring of reads indicates strand. **b)** IGV screenshot containing reads that span somatic *KMT2C* mutations (chr7:151877972/ p.Q2325* and chr7:151877935/p.S2337Y). Coloring of reads indicates strand.



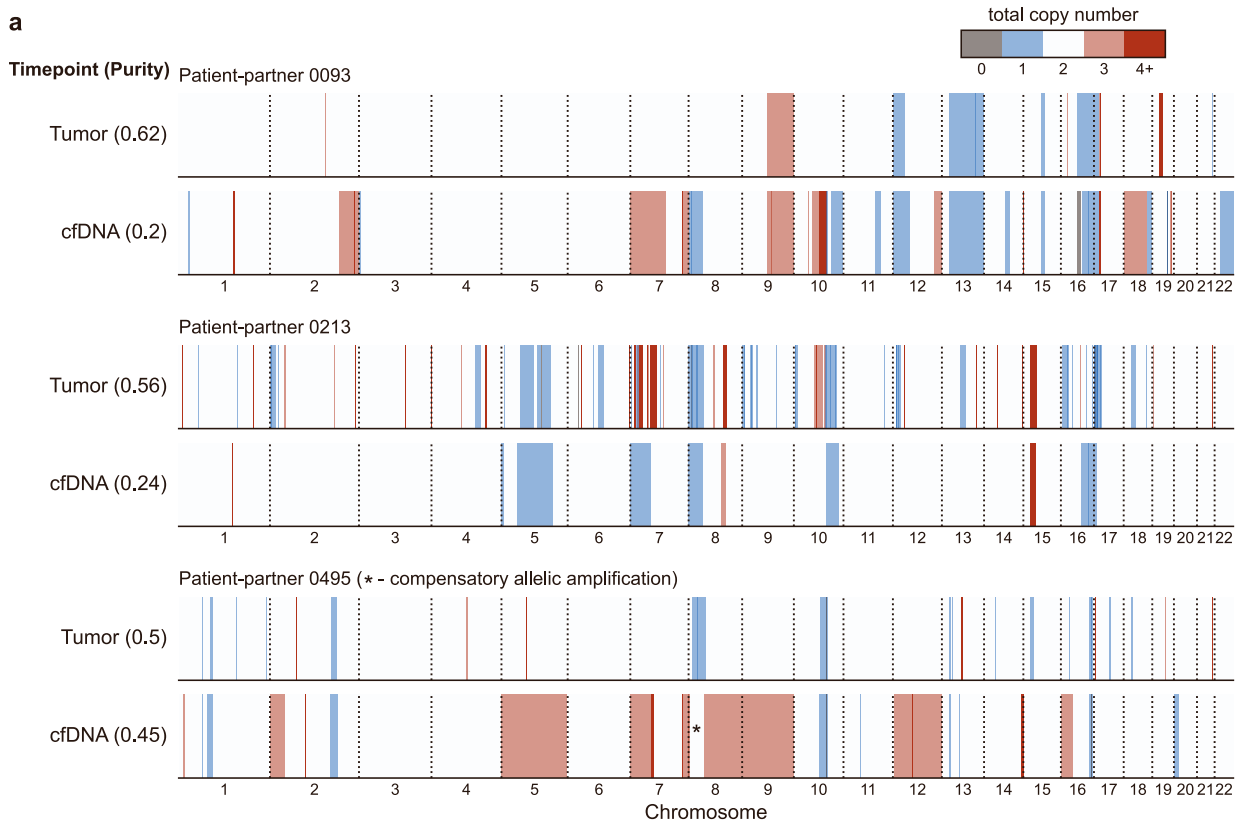
Supplementary Figure S10. Somatic losses affecting genes with pathogenic germline mutations, related to Fig. 3E.

a-c) Copy number profiles for three samples with pathogenic germline mutations are shown. The top profile of each panel reflects the integer allelic copy number segments, the middle reflects log-ratio coverage, and the bottom reflects minor allele fraction (MAF), an indicator of allelic balance. The pathogenic germline alterations of these samples are depicted in Supplementary Table S5.



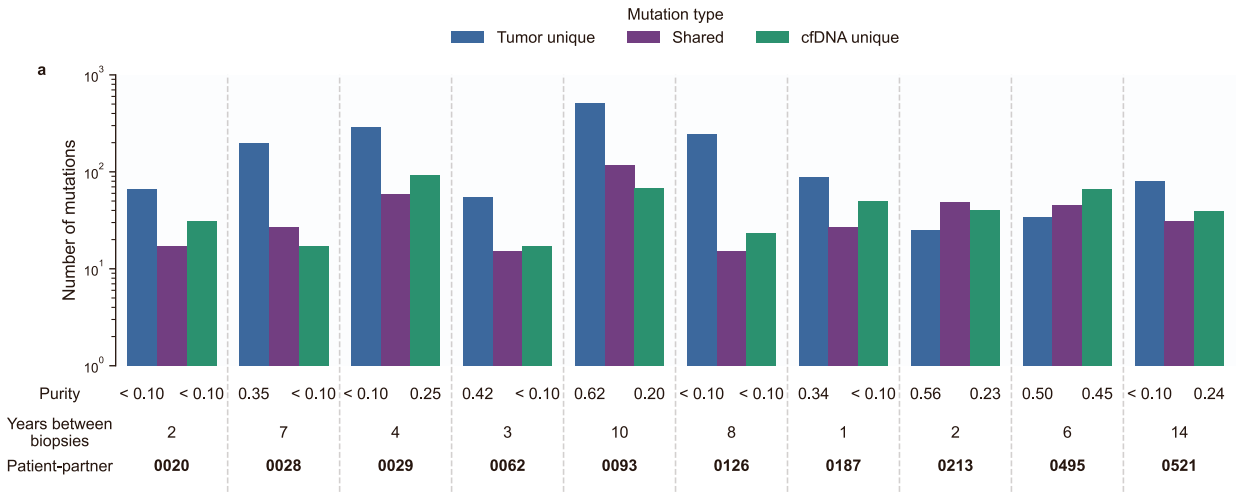
Supplementary Figure S11. Phylogenetics of samples from patient-partner 0213, related to Fig. 4B.

a) The y-axis shows the cancer cell fraction (CCF) of clonal clusters identified between primary tumor and cfDNA from donated blood (x-axis). Colors indicate how many mutations were identified in each clone, with a 95% confidence interval around the estimated CCF. Purple represents the truncal/ancestral clone. The ancestral clone does not reach a CCF of 1 in the cfDNA sample because its inferred purity (0.20) is low, which confounds the ability to accurately quantify CCF.



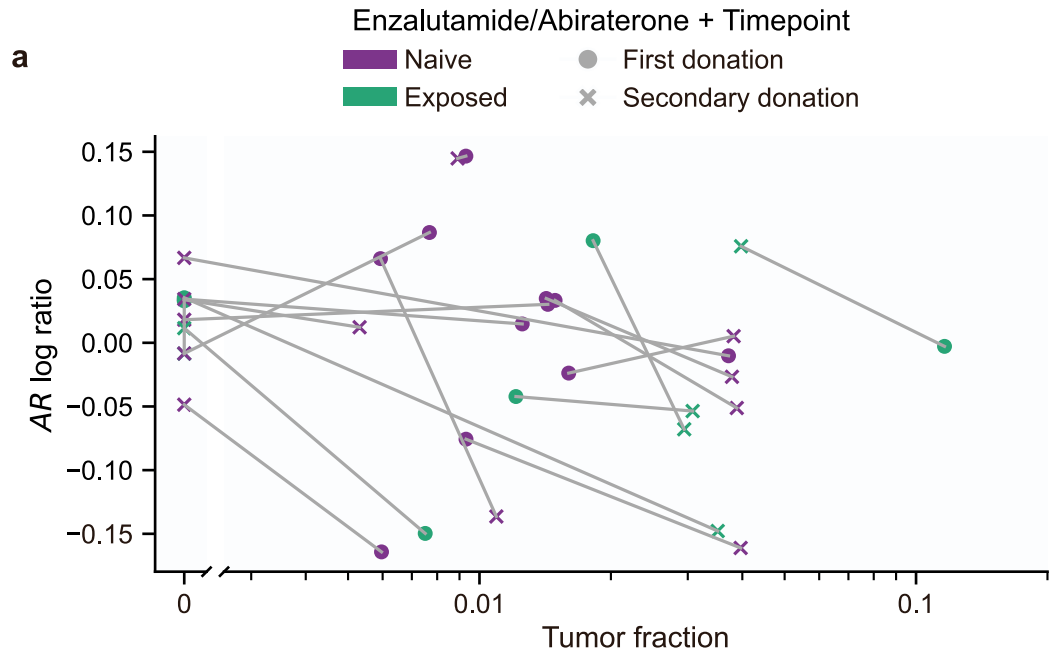
Supplementary Figure S12. Copy number profiles of shared tumor and cfDNA samples, related to Fig. 4B.

a) Concordance of copy number profiles between archival primary tumors and donated cfDNA samples. The x-axis depicts chromosomal location, with coloring representing copy number alterations and their absolute copy number. In general, there are no archival-specific copy number alterations, with the potential exception of chr7p amplification in patient-partner 0213. When sample purity is below 0.30, focal copy number amplifications can be undetectable. In patient-partner 0495's samples, an arm-level deletion of 8p acquired a compensatory amplification on the other allele that restored diploid copy number.



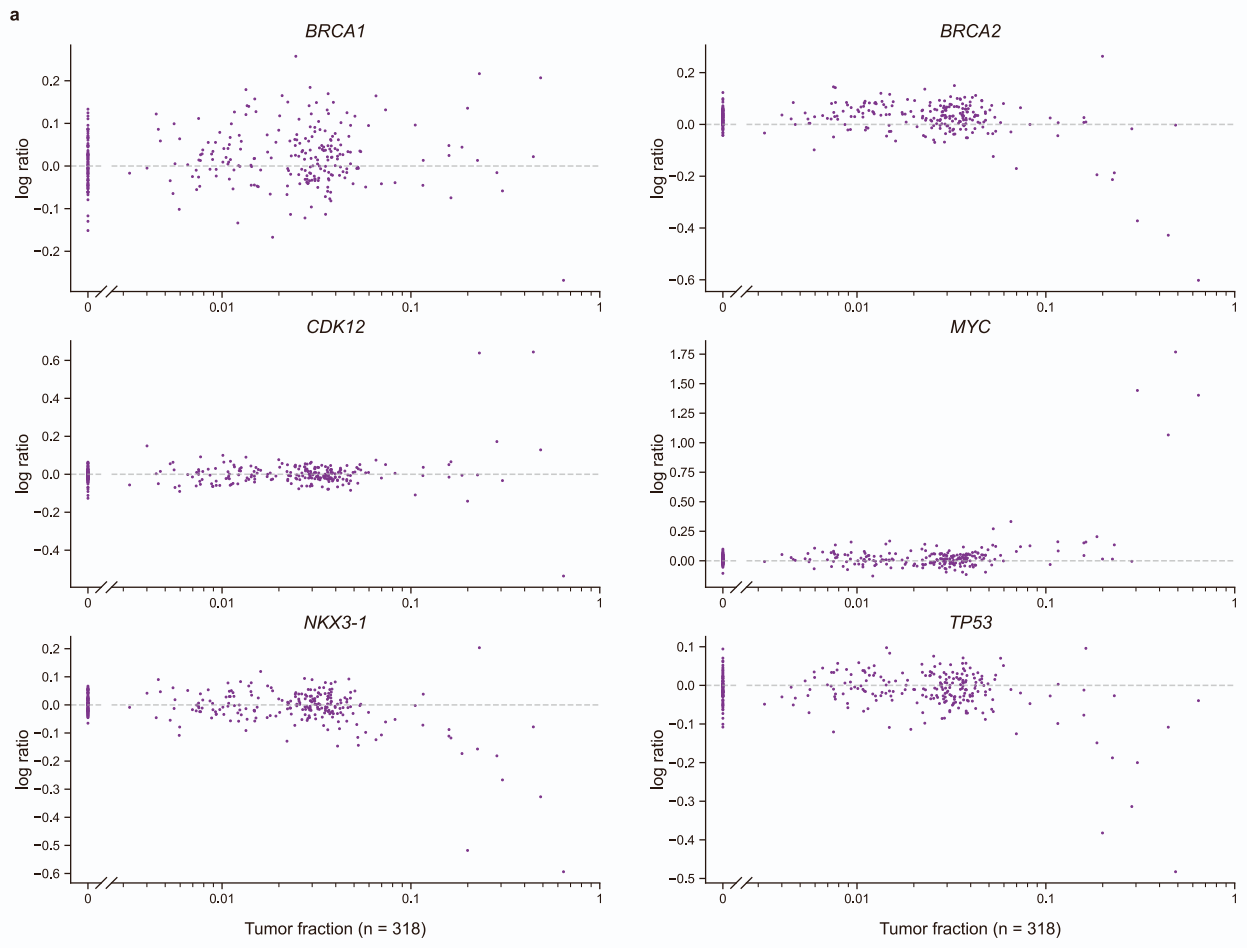
Supplementary Figure S13. Mutation exclusivity between tumor and cfDNA samples from the same patient, related to Fig. 4B.

a) Number of mutations for each sample type for ten patient-partners with both archival tumor and donated cfDNA samples. The y-axis shows number of mutations, while the x-axis shows each patient. The purple and blue bars represent mutations identified exclusively in the archival tumor and cfDNA samples, respectively. The green bars represent mutations that had at least one supporting read in both tumor and cfDNA samples within the union of all mutations called in tumor and cfDNA samples (see Methods – *Phylogenetic analysis*). The purities and amount of time between samples are shown below each bar. Purities below 0.10 cannot be accurately estimated.



Supplementary Figure S14. Shared tumor fraction and *AR* copy-number log-ratios, related to Fig. 4D.

a) Tumor fraction of 36 cfDNA samples from donated blood of 18 patient-partners with ULP-WGS sequencing is shown on the x-axis, while the log copy-ratio (logR) of the genomic interval containing *AR* is shown on the y-axis. Samples from the same patient-partner are connected with a line. The first sample donated is shown with a circle and the second (usually donated months later) is shown with an 'x'.



Supplementary Figure S15. Ultra-low pass WGS signal for other common copy number changes in metastatic prostate cancer, related to Fig. 4D.

a) Tumor fraction of 318 cfDNA samples from donated blood of 300 patient-partners with ULP-WGS sequencing is shown on the x-axis, while the log copy-ratio (logR) of the genomic interval containing the gene is shown on the y-axis. In general, deletions are difficult to detect at low tumor fractions, as are single or double amplifications (e.g. *MYC*). Signal from *AR* is detectable because its amplification often generates dozens of copies. See Adalsteinsson et al. for a more in-depth analysis of the sensitivity and sensitivity of ULP-WGS[2].

Institution	Patient count	Institution	Patient count
DANA-FARBER CANCER INSTITUTE	47	UC HEALTH - UNIVERSITY OF COLORADO CANCER CENTER	4
UT M. D. ANDERSON CANCER CENTER	29	MASSEY	3
HELEN DILLER FAMILY COMPREHENSIVE CANCER CENTER	26	CARBONE	3
MAYO CLINIC HOSPITAL ROCHESTER	24	NORRIS COTTON	3
SIDNEY KIMMEL CANCER CENTER	24	COLUMBIA UNIVERSITY IRVING MEDICAL CENTER	3
MEMORIAL SLOAN	19	PERLMUTTER CANCER CENTER	3
HUTCHINSON	17	ROGEL	3
MOUNT SINAI HOSPITAL	13	STEPHENSON CANCER CENTER	2
SMILOW CANCER	12	ROSWELL PARK	2
KNIGHT	11	FOX CHASE	2
SITEMAN	10	CITY OF HOPE	2
MOORES CANCER CENTER	10	MASONIC	2
INGRAM CANCER	8	HOLLING	2
SIMON COMPREHENSIVE CANCER	8	SYLVESTER	2
NORTHWESTERN	8	HOLDEN	2
DUKE CANCER	7	OHIO STATE UNIVERSITY COMPREHENSIVE CANCER CENTER - THE JAMES	1
THE UNIVERSITY OF CHICAGO COMPREHENSIVE CANCER CENTER	6	HILLMAN	1
MOFFITT	5	LINEBERGER	1
UC DAVIS HEALTH - COMPREHENSIVE CANCER CENTER	4	CHAO FAMILY COMPREHENSIVE CANCER CENTER	1
UNIVERSITY OF KANSAS CANCER CENTER	4	UNIVERSITY OF NEW MEXICO	1
STANFORD CANCER INSTITUTE	4	BAYLOR	1
RUTGERS CANCER	4	MAYS	1
SIMMONS COMPREHENSIVE CANCER CENTER	4	UK MARKEY CANCER CENTER	1
WINSHIP	4	THOMAS JEFFERSON	1
KECK HOSPITAL OF USC - NORRIS CANCER CENTER	4	LOMBARDI	1

Supplementary Table S1. List of NCI-designated cancer centers, related to Fig. 1C.

List of NCI-designated cancer centers along with unique patient-partner attendance counts. For institutions that have satellite locations, only the main location was considered in tabulating patient attendance and NCI-designated status. These institutions are depicted in green in Fig. 1c.

Patient-reported data	Number of patient-partners (%)
Age at initial diagnosis (mean: 61)	
Did not respond	1 (0.1%)
≤ 40 years	4 (0.6%)
> 40, ≤ 50 years	62 (8.7%)
> 50, ≤ 60 years	256 (35.8%)
≥ 60 years	383 (54.8%)
What is your race? (Select all that apply)	
White	657 (93.1%)
Black or African American	12 (1.7%)
Other (Not specified)	10 (1.4%)
Japanese	4 (0.6%)
Chinese	4 (0.6%)
American Indian	3 (0.4%)
Prefer to not respond	3 (0.4%)
Did not respond	4 (0.4%)
Southeast Asian or Indian	2 (0.3%)
American Indian and White	2 (0.3%)
White, Other (Not specified)	2 (0.3%)
Japanese and White	2 (0.3%)
Japanese, Chinese, Hawaiian, and White	1 (0.1%)
Do you consider yourself Hispanic or Latino?	
Yes	12 (1.7%)
No	689 (97.6%)
Did not respond	5 (0.7%)

Supplementary Table S2. Additional patient reported data, related to STAR Methods.

Patient-partner reported demographic data for patient-partners enrolled before June 1, 2020 (n = 706). Age at initial prostate cancer diagnosis is calculated based on the patient reported date of birth and month/year of initial prostate cancer diagnosis. Patient-partners were free to select as many racial identities as they identified with.

Gene	Info	Gene	Info	Gene	Info
ABCB11	O	*FANCI*	O	*RAD51*	R
ACD	O	*FANCL*	RO	*RAD51B*	R
AIP	O	*FANCM*	RO	*RAD51C*	RO
ALK	O	*FH*	O	*RAD51D*	RO
APC	O	*FLCN*	O	*RAD54L*	R
ATM	RPO	*GATA2*	O	*RAF1*	O
ATR	RO	*GEN1*	R	*RB1*	O
AXIN2	O	*GPC3*	O	*RECQL*	O
BAP1	RO	*HDAC2*	R	*RECQL4*	O
BARD1	RO	*HFE*	O	*RET*	O
BLM	O	*HMBS*	O	*RFWD3*	O
BMPR1A	O	*HNF1A*	O	*RHBDF2*	O
BRAF	O	*HOXB13*	PO	*RTEL1*	O
BRCA1	RPO	*HRAS*	O	*RUNX1*	O
BRCA2	RPO	*KIT*	O	*SBDS*	O
BRIP1	RO	*KRAS*	O	*SDHA*	O
BUB1B	O	*LZTR1*	O	*SDHAF2*	O
CBL	O	*MAP2K1*	O	*SDHB*	O
CDC73	O	*MAP2K2*	O	*SDHC*	O
CDH1	O	*MAX*	O	*SDHD*	O
CDK12	R	*MEN1*	O	*SETBP1*	O
CDK4	O	*MET*	O	*SH2D1A*	O
CDKN1B	O	*MITF*	O	*SLC25A13*	O
CDKN1C	O	*MLH1*	RPO	*SLX4*	O
CDKN2A	O	*MLH3*	R	*SMAD4*	O
CDKN2B	O	*MPL*	O	*SMARCA4*	O
CEBPA	O	*MRE11A*	R	*SMARCB1*	O
CHEK1	R	*MSH2*	RPO	*SMARCE1*	O
CHEK2	RPO	*MSH6*	RPO	*SOS1*	O
CYLD	O	*MTAP*	O	*SPRTN*	O
DDB2	O	*MUTYH*	O	*SRP72*	O
DDX41	O	*NBN*	RPO	*STAT3*	O
DICER1	O	*NF1*	O	*STK11*	O
DIS3L2	O	*NF2*	O	*SUFU*	O
DKC1	O	*NHP2*	O	*TERT*	O
EGFR	O	*NRAS*	O	*TGFBR1*	O
EPCAM	O	*NTHL1*	O	*TINF2*	O
ERCC1	O	*PALB2*	RO	*TMEM127*	O
ERCC2	O	*PDGFRA*	O	*TP53*	O
ERCC3	RO	*PHOX2B*	O	*TRIM37*	O
ERCC4	O	*PMS2*	RO	*TSC1*	O
ERCC5	O	*POLD1*	O	*TSC2*	O
ETV6	O	*POLE*	O	*TSHR*	O
EXT1	O	*POLH*	O	*UROD*	O
EXT2	O	*POT1*	O	*VHL*	O
FAH	O	*PPP2R2A*	R	*WRN*	O
FAM175A	R	*PRF1*	O	*WT1*	O
FANCA	RO	*PRKAR1A*	O	*XPA*	O
FANCC	O	*PTCH1*	O	*XPC*	O
FANCD2	O	*PTCH2*	O	*XRCC2*	R
FANCE	O	*PTEN*	O	*XRCC3*	O
FANCF	O	*PTPN11*	O		
FANCG	O	*RAD50*	R		

173

Supplementary Table S3. DNA repair and germline cancer susceptibility gene list, related to Fig 3D.

List of genes used for germline analysis in this study. Genes with "R" denote the genes used in the analysis of the association between the presence of COSMIC2.0 signature 3 and DNA-repair alterations, taken from Mateo et al. 2015, de Bono et al. 2020, and Pritchard et al. 2016[3–5]. Genes with "P" denote the list of genes used to evaluate germline alterations in prostate cancer susceptibility genes, taken from Aldubayan 2019[6]. Genes with "O" denote the list of genes used to evaluate germline alterations in other cancer susceptibility genes, taken from COSMIC, Huang et al. 2018, and Rahman 2014[7–9]. See Methods for the specifics of these analyses.

Therapy brand name (Generic name)	Category	Number of patient-partners (% of 639)
Hormones		
Lupron (Leuprolide)	1st line ADT	538 (84.2%)
Casodex (Bicalutamide)	1st line ADT	326 (51.0%)
Zytiga (Abiraterone)	2nd line ADT	220 (34.4%)
Firmagon (Degarelix)	1st line ADT	109 (17.1%)
Xtandi (Enzalutamide)	2nd line ADT	107 (16.7%)
Zoladex (Goserelin)	1st line ADT	38 (5.9%)
Drogenil (Flutamide)	1st line ADT	5 (0.8%)
Nilandron (Nilutamide)	1st line ADT	5 (0.8%)
Decapeptyl (Triptorelin)	1st line ADT	3 (0.4%)
Prostap (Leuprorelin)	1st line ADT	1 (0.2%)
Suprefact (Buserelin)	1st line ADT	0 (0.0%)
Chemotherapy		
Taxotere (Docetaxel)	Chemotherapy	168 (26.3%)
Jevtana (Cabazitaxel)	Chemotherapy	18 (2.8%)
Paraplatin (Carboplatin)	Chemotherapy	17 (2.6%)
Etopophos / Toposar (Etoposide)	Chemotherapy	5 (0.8%)
Emcyt (Estramustine)	Chemotherapy	3 (0.5%)
Taxol (Paclitaxel)	Chemotherapy	2 (0.3%)
Novantrone (Mitoxantrone)	Chemotherapy	1 (0.2%)
Other Therapy		
Xgeva/Prolia (Denosumab)	Supportive care	103 (16.2%)
Provenge (Sipuleucel-T)	Immunotherapy	59 (9.2%)
Zometa (Zoledronic Acid)	Supportive care	50 (7.9%)
Xofigo (Radium-223)	Nuclear medicine	23 (3.6%)
Keytruda (Pembrolizumab)	Immunotherapy	10 (1.5%)
Lynparza (Olaparib)	PARP inhibitor	6 (0.9%)
Yervoy (Ipilimumab)	Immunotherapy	3 (0.5%)
Opdivo (Nivolumab)	Immunotherapy	2 (0.3%)
Tecentriq (Atezolizumab)	Immunotherapy	0 (0.0%)
Rubraca (Rucaparib)	PARP inhibitor	0 (0.0%)
Quadramet (Samarium SM 153 lexidronam)	Supportive care	0 (0.0%)
Metastron (Strontium-89)	Supportive care	0 (0.0%)
Experimental/Clinical Trial		
Experimental/Clinical Trial	Clinical trial	87 (13.6%)

Supplementary Table S4. Therapies available for selection on patient survey, related to Fig. 2E.

List of therapies available for selection on patient survey (Supplementary Figure S2). Only these therapies were used to determine the overlap between patient-reported therapies and medical

187 record therapies. Percentage defined relative to the number of patient-partners that provided at
188 least one therapy on the survey (n = 639/706).

gene	consequence	ClinVar 2019 annotation	purity	t_ref	t_alt	n_ref	n_alt
BRCA2 (**S10-a**)	splice_acceptor	Pathogenic/Likely_pathogenic	0.25	31	37	25	23
BRCA2 (**S10-b**)	stop_gained	Pathogenic	0.24	5	10	21	27
BRCA2	frameshift	Pathogenic	0.2	17	7	6	6
BRCA2	frameshift	Pathogenic	-	183	181	127	101
BUB1B	stop_gained	Pathogenic	0.54	351	324	103	108
CHEK2	frameshift	Conflicting	0.73	89	42	56	40
CHEK2	missense	Conflicting	0.52	75	38	69	58
CHEK2	frameshift	Conflicting	0.5	48	33	74	52
CHEK2	frameshift	Conflicting	-	89	59	196	155
CHEK2	missense	Conflicting	-	110	93	84	61
CHEK2 (**S10-c**)	missense	Pathogenic/Likely_pathogenic	0.62	88	151	95	129
CHEK2	missense	Conflicting	-	5	4	72	64
ERCC2	missense	Pathogenic	-	150	201	93	96
FANCD2	stop_gained	-	-	197	122	117	72
FANCL	inframe_deletion	Conflicting	0.53	28	17	102	94
FH	missense	Conflicting	0.56	136	132	96	70
FH	inframe_insertion	Conflicting	0.3	201	76	138	84
HOXB13	missense	_risk_factor	-	139	103	52	49
HOXB13	missense	_risk_factor	-	238	222	52	49
HOXB13	missense	_risk_factor	0.5	177	143	96	85
NBN	frameshift	Pathogenic	-	30	6	35	25
NF1	missense	Pathogenic	-	241	124	142	63
NF1	splice_donor	Pathogenic	-	173	140	136	103
SBDS	splice_donor	Pathogenic	0.2	91	40	80	45
SBDS	splice_donor	Pathogenic	-	180	75	125	53
CHEK2	frameshift	Pathogenic/Likely_pathogenic	-	-	-	63	51
SBDS	splice_donor	Pathogenic	-	-	-	84	46

Supplementary Table S5. Pathogenic germline mutations associated with cancer heritability and their presence in tumor samples, related to Fig. 3D, E.

26 pathogenic germline mutations from 25 germline DNA samples were identified among 132 sequenced samples, using a curated list of genes associated with cancer heritability (Methods). The tumor reference, tumor alt, normal reference, and normal alt read counts are shown as t_ref, t_alt, n_ref, n_alt, respectively. Pathogenic germline mutations with an accompanying somatic deletion depicted in Supplementary Fig. S10 are listed under "gene". "Conflicting" refers to "Conflicting_interpretations_of_pathogenicity". Dashes indicate missing data, either because the tumor purity was too low to pass quality control or because no tumor was sequenced.

Supplementary Methods

Table of Contents

Patient Enrollment and Study Material Acquisition

Establishing patient partnership

Patients and the extended metastatic prostate community have been directly involved in the creation and development of the Metastatic Prostate Cancer Project (MPCproject) since the project's conception. During the initial development of the project, a patient advisory council (PAC) comprised of patients, loved ones, and advocates met frequently with study staff to determine the study's approach for outreach, patient enrollment, study website design, and sample collection, among other details of project operations. Study staff from the project continue to meet regularly with the PAC. In addition to working with members of the PAC, the MPCproject leverages the expertise of the many prostate cancer advocacy group partners to improve outreach and project operations. Finally, patients that are not directly involved in the PAC or an advocacy group, can learn about and partner with the project through various social media platforms, newsletters, or educational materials generated by study staff to provide input or feedback.

This study includes as authors patient advocates who were instrumental in survey design, project development, assessment of patient criteria, and outreach strategy. The MPCproject glossary included with the study was reviewed by practicing oncologists, patient advocates, and study staff.

Patient Enrollment and Informed Consent

The MPCproject is a decentralized, online patient-partnered genomics research study. Patients anywhere in the United States and Canada can visit the project website (https://mpcproject.org) to learn about the research initiative and register for the study. If a patient is interested in participating, the online registration process has four steps: registration, an

optional intake survey, an electronic consent form, and a medical record release form. For the study, we consider any patient that completes the consent form to be enrolled.

On the study registration page, a patient provides their first and last name, email, and confirmation of their metastatic or advanced prostate cancer diagnosis as well as acknowledgement of their willingness to provide further information on their medical care and experience with the disease. The registration page prompts patients to create a password protected account to save provided information and to allow patients to revisit their completed survey and forms at any time. Once the account has been created, registrants are taken to an optional intake survey (Supplementary Fig. S2) where they are asked to provide basic demographic information as well as answer questions about their experience with prostate cancer via a 17-question survey that was developed in partnership with clinicians, researchers, and patients. Each question is optional and survey responses can be revisited. To submit the survey, patients agree to the MPCproject saving their survey information, and, if they live in the U.S. or Canada, agree to study staff reaching out if the MPCproject conducts future studies. The minimum requirement to submit the survey is providing country of origin and a zip code.

Registrants that choose to submit the survey and who reside in the U.S. or Canada are then taken to an electronic consent form. Patients provide informed consent using a web-based consent form as approved by the Dana-Farber/Harvard Cancer Center Institutional Review Board (DF/HCC Protocol 15-057B). To formally enroll in the study, patients provide their electronic signature on the consent form. The consent form provides various levels of participation. The minimum consent enables study staff to request and abstract medical records, send the patient a saliva kit, perform germline sequencing analysis if a saliva sample is returned, and release de-identified clinical and genomic data into public repositories. Patients have the additional option

of consenting to study staff obtaining archived tumor tissue and/or blood sample(s) for further somatic and germline sequencing analyses. Email reminders are sent to registrants who have not completed the consent process (weekly for three weeks, and again at six weeks). A copy of the completed consent form is saved in the patient's account and emailed to them.

Upon submission of the consent form, the final step in the study enrollment process is to complete a medical release form. On this form, patients provide their contact information and information about any physician or hospital involved in the care of their prostate cancer. By submitting the release form, patients agree to study staff reaching out to the listed institutions to requested medical records and, if elected on the electronic consent form, archived tissue samples. Email reminders are sent weekly for three weeks, and again at six weeks, to registrants who have not completed the release form. A copy of the completed release form is saved in the patient's account and emailed to them.

Medical Records

After patients complete the consent and release forms and provide institutions where they received care for their prostate cancer, the study staff requests their medical records. Study staff call each institution's medical record departments to obtain copies of the patient's records starting at the date of diagnosis of prostate cancer through the day of the faxed request. Requests are faxed to the respective departments after phone confirmation of the fax number. Medical records are returned to the project via mail, fax, or online portals. Once a medical record arrives, it is saved in an electronic format in a secure database. If a record request is not fulfilled in 6 months, a second request is submitted. If the medical records department requires additional paperwork or signatures per the specific institution's release requirements, the patient is contacted and asked to provide the additional required forms. When patients are contacted for

this purpose, study staff are clear that this additional step is optional for patients. Study staff can also request subsequent medical records after an initial request had been fulfilled if the need arises.

Samples

All patients that complete the electronic consent form are sent a saliva kit to provide a saliva sample. In addition, patients can opt-in to providing archival tumor tissue and/or one or more blood samples.

Saliva

Saliva kits are sent to patients who complete the consent and medical release form and provide a valid mailing address in the United States or Canada. Staff at the Broad Institute Genomics Platform prepare each unique patient's kit by assigning it a unique barcode and prepaid business reply-label and packaging the kit with instructions for the patient on how to provide at least 2 mL of saliva in a DNA Genotek Oragene Discover (OGR-600) tube labeled with a matching barcode. All kits are affixed with a prepaid business-reply label. Samples are mailed back to the Broad Institute by patients after collection, and then logged and stored at room temperature by study staff upon receipt. Saliva samples are eventually pushed for whole exome sequencing to obtain germline DNA once matched tumor samples are also received and submitted for sequencing.

Archived Tumor Tissue

Once a patient's medical record and normal normal sample (saliva or blood) are received, study staff review the record to confirm the patient has had a clinical diagnosis of metastatic or advanced prostate cancer. Surgical and pathology records are used to develop a patient's surgical history and identify archived formalin-fixed paraffin embedded (FFPE) prostate cancer tumor

tissue that may be requested. Study staff, in collaboration with oncologists and pathologists, developed strict guidelines for selecting which tumor sample to request to obtain the minimal amount of tissue that will not interfere with the patient's future clinical care. For each patient, a specific sample is requested only there are at least three blocks with prostatic adenocarcinoma and at minimum two of those blocks are actively being stored in the source pathology department. If a sample meets the requesting criteria, study staff coordinate with the sending pathology department to fax a request and obtain the sample via mail. The tissue request form requests that pathology departments send an H&E slide along with either an entire block from the surgical case or 5-20 5-micron unstained slides from a block. All tissue requests submitted by the MPCproject state that no sample should be exhausted to fulfill the request. Tissue samples received as blocks are labeled with unique numerical identifiers and sent to the Dana-Farber/Harvard Cancer Center Specialized Histopathology Services (SHS) Core to be cut into three 30- micron scrolls per block and an accompanying H&E for tumor confirmation. Scrolls, unstained slides, and H&Es are labeled with unique barcode identifiers. Archived tumor tissue with a matched germline sample (from either saliva or a blood sample's buffy coat) are sent to the Broad Institute's Genomics Platform for whole exome sequencing.

Primary and Secondary Blood Samples

Blood sample acquisition and sequencing preparation are performed as described in Painter et al. 2020 except in the additional steps of sending secondary blood kits to patients[10]. The MPCproject was awarded a grant to send a cohort of selected patients second blood kits to obtain an additional blood sample to study tumor evolution. Patients are selected based on a combination of criteria including date of registration, date of primary blood draw, primary blood sample containing sufficient ctDNA quantity for whole exome sequencing, and successful

acquisition of medical records. An email is sent to selected patients describing the intent and optional nature of the second blood kit. The email contains a link to a new consent form and asks if they would be willing to provide an additional blood sample. If the patient selects 'Yes' on the consent form, another round of the blood sample acquisition process is triggered: a new blood kit is sent to their home, returned to the Broad Institution, and processed using the same procedure outlined for their primary kit.

Data Generation

Medical Record Abstraction

Medical records are requested for any consented patient in the US and Canada that listed any institution(s) from which they received care on their medical release form. Medical records arrive in various formats and all are eventually transferred to an electronic format and stored on a secure internal server. Scanned medical records are run through the Optical Character Recognition (OCR) engine known as Tesseract (LSTM model inside Tesseract version 4.0; (https://github.com/tesseract-ocr/tesseract)) to facilitate manual abstraction by study staff.

Three separate abstractors on the study staff team are involved in the abstraction and quality control process of the clinical data from each searchable record. To begin, two abstractors independently read and isolate the same clinical information for each patient. A third abstractor aligns the separate abstractions and identifies field-specific discrepancies between the two abstractions. The third abstractor attempts to resolve any lack of concordance by returning to the patient's medical record to identify the correct data. At any point in the process, abstractors can work with clinical oncologists to answer questions or address lack of concordance.

The abstractors use a clinical data dictionary comprising 60 fields that was curated by prostate cancer oncologists. For information that's not found, it was abstracted as 'NOT FOUND

IN RECORD'. In instances where ambiguity or incomplete data was present, inferences were made considering the whole narrative of the medical record. The dictionary includes possible responses for each field. For date-type fields, incomplete dates missing either the month or day, are abstracted as the first month of the year and/or first day of the month, respectively. All time related fields are anchored from the date of primary prostate cancer diagnoses. For example, a patient's metastatic diagnosis date is represented as the calculated number of days from the primary diagnosis date to the metastatic diagnosis date. This was done to protect patient privacy.

Patient-Reported Data

Study inclusion

Survey responses were cleaned for patients that completed their consent and release forms and submitted a survey by June 1, 2020. 706 of these patients reported being located within the U.S. and Canada and were thus included in downstream analyses.

Cleaning/categorization of medical institutions

Patients were asked in their medical release form to report all physicians with whom they received care for their prostate cancer, institutions where they received an initial prostatic biopsy or prostatectomy, and institutions where they received another surgery related to their prostate cancer. Institutions of reported physicians were gathered based on the most recent affiliation identified from affiliated websites. Satellite locations of larger institutions were considered separate institutions. Names were standardized by three separate reviewers manually. For Fig 1c, only unique institutions for each patient are shown. The NCI designated cancer center list was taken from cancer.gov/research/infrastructure/cancer-centers/find.

Cleaning/categorization of therapies

Patients selected all therapies that they had received for their prostate cancer in the intake

survey. Therapies were categorized by prostate cancer oncologists into broad treatment categories according to their primary therapeutic function (See Supplementary Table S4).

Cleaning/categorization of alternative lifestyles

Patient responses to question 7 on the intake survey (Supplementary Fig. S2) were categorized into four broad categories: Diet/lifestyle, Supplements, and Non-Cancer Therapies. Except for plant-based diet and unspecified diet change, responses were not mutually exclusive. Different methods of taking similar supplements (e.g., turmeric paste, turmeric capsules, turmeric powder) were considered the same supplement. Brand name products were converted to generic forms (e.g., Pomi-T was considered "pomegranate"). Manual classification was conducted by two separate reviewers.

Genomic Sequencing

All samples were received and sequenced at the Broad Institute's Genomics Platform. Due to changes in sequencing methods as a function of improved technologies and the longitudinal nature of this project, certain sequencing methods are subset by date to indicate what was applied for samples received within the specific timeframe.

DNA Isolation in Saliva

DNA was extracted via the Chemagic MSM I with the Chemagic DNA Blood Kit-96 from Perkin Elmer. This kit combines a chemical and mechanical lysis with magnetic bead-based purification. Saliva samples were incubated at 50°C for 2 hours. The saliva was then transferred to a deep well plate placed on the Chemagic MSM I. The following steps were automated on the MSM I.

M-PVA Magnetic Beads were added to the saliva. Lysis buffer was added to the solution and mixed. The bead-bound DNA was then removed from solution via a 96-rod magnetic head

and washed in three Ethanol-based wash buffers. The beads were then washed in a final water wash buffer. Finally, the beads were dipped in elution buffer to resuspend the DNA sample in solution. The beads were then removed from solution, leaving purified DNA eluate. DNA samples were quantified using a fluorescence based PicoGreen assay.

cfDNA Extraction from Whole Blood

Whole blood was collected in EDTA, CellSave, or Streck tubes and processed for plasma fractionation. Blood tubes were centrifuged at 1900 g for 10 minutes and plasma was transferred to second tube before further centrifugation at 15000 g for 10 minutes. Supernatant plasma was stored at -80C until cfDNA extraction. cfDNA was extracted using the QIAsymphony DSP Circulating DNA Kit according to the manufacturer's instructions, with 6.3 mL of plasma as input and with a 60 uL DNA elution (Qiagen, 2017).

Ultra-Low Pass Whole Genome Sequencing (ULP-WGS)

non-UMI ULP-WGS sequencing [dates: 2017-2/11/2018]:

1. Library Construction

Initial DNA input is normalized to be within the range of 25-52.5 ng in 50 uL of TE buffer (10mM Tris HCl 1mM EDTA, pH 8.0) according to picogreen quantification. For adapter ligation, Illumina paired end adapters were replaced with palindromic forked adapters, purchased from Integrated DNA Technologies, with unique dual-indexed molecular barcode sequences to facilitate downstream pooling. With the exception of the palindromic forked adapters, the reagents used for end repair, A-base addition, adapter ligation, and library enrichment PCR were purchased from KAPA Biosciences in 96-reaction kits. In addition, during the post-enrichment SPRI cleanup, elution volume was reduced to 30μL to maximize library concentration, and a vortexing step was added to maximize the amount of template eluted.

2. Post Library Construction Quantification and Normalization

Library quantification was performed using the Invitrogen Quant-It broad range dsDNA quantification assay kit (Thermo Scientific Catalog: Q33130) with a 1:200 PicoGreen dilution. Following quantification, each library is normalized to a concentration of 25 ng/μL, using a 1X Low TE pH 7.0 solution.

3. Library Pool Creation for Ultra-low Pass Sequencing

In preparation for the sequencing of the ultra-low pass libraries (ULP), approximately 4 μL of the normalized library is transferred into a new receptacle and further normalized to a concentration of 2ng/μL using Tris-HCl, 10mM, pH 8.0. Following normalization, up to 95 ultra-low pass WGS samples are pooled together using equivolume pooling. The pool is quantified via qPCR and normalized to the appropriate concentration to proceed to sequencing.

4. Cluster amplification and sequencing

Cluster amplification of library pools was performed according to the manufacturer's protocol (Illumina) using Exclusion Amplification cluster chemistry and HiSeq X flowcells. Flowcells were sequenced on v2 Sequencing-by-Synthesis chemistry for HiSeq X flowcells. The flowcells are then analyzed using RTA v.2.7.3 or later. Each pool of ultra-low pass whole genome libraries is run on one lane using paired 151bp runs.

UMI ULP-WGS sequencing [dates: 2/12/2018-6/1/2020]:

1. Library Construction

Initial DNA input is normalized to be within the range of 25-52.5 ng in 50 uL of TE buffer (10mM Tris HCl 1mM EDTA, pH 8.0) according to picogreen quantification. Library preparation is performed using a commercially available kit provided by KAPA Biosystems

(KAPA HyperPrep Kit with Library Amplification product KK8504) and IDT's duplex UMI adapters. Unique 8-base dual index sequences embedded within the p5 and p7 primers (purchased from IDT) are added during PCR. Enzymatic clean-ups are performed using Beckman Coultier AMPure XP beads with elution volumes reduced to 30μL to maximize library concentration.

2. Post Library Construction Quantification and Normalization

Library quantification was performed using the Invitrogen Quant-It broad range dsDNA quantification assay kit (Thermo Scientific Catalog: Q33130) with a 1:200 PicoGreen dilution. Following quantification, each library is normalized to a concentration of 35 ng/μL, using Tris-HCl, 10mM, pH 8.0.

3. Library Pool Creation for Ultra-low Pass Sequencing

In preparation for the sequencing of the ultra-low pass libraries (ULP), approximately 4 μL of the normalized library is transferred into a new receptacle and further normalized to a concentration of 2ng/μL using Tris-HCl, 10mM, pH 8.0. Following normalization, up to 95 ultra-low pass WGS samples are pooled together using equivolume pooling. The pool is quantified via qPCR and normalized to the appropriate concentration to proceed to sequencing.

4. Cluster amplification and sequencing

Cluster amplification of library pools was performed according to the manufacturer's protocol (Illumina) using Exclusion Amplification cluster chemistry and HiSeq X flowcells. Flowcells were sequenced on v2 Sequencing-by-Synthesis chemistry for HiSeq X flowcells. The flowcells are then analyzed using RTA v.2.7.3 or later. Each pool of ultra-low pass whole genome libraries is run on one lane using paired 151bp runs.

Whole Exome Sequencing (WES)

Express WES for saliva and tissue [dates: 8/13/2017 - 4/15/2018]:

1. Library Construction

Library construction was performed as described in Fisher et al., with the following modifications DNA input into shearing was reduced from 3µg to 10-100ng in 50µL of solution[11]. For adapter ligation, Illumina paired end adapters were replaced with palindromic forked adapters, purchased from Integrated DNA Technologies, with unique dual-indexed molecular barcode sequences to facilitate downstream pooling. Kapa HyperPrep reagents in 96-reaction kit format were used for end repair/A-tailing, adapter ligation, and library enrichment PCR. In addition, during the post-enrichment SPRI cleanup, elution volume was reduced to 30µL to maximize library concentration, and a vortexing step was added to maximize the amount of template eluted.

2. In-solution hybrid selection

After library construction, hybridization and capture were performed using the relevant components of Illumina's TruSeq Rapid Exome Kit and following the manufacturer's suggested protocol, with the following exceptions: first, all libraries within a library construction plate were pooled prior to hybridization. Second, the Midi plate from Illumina's TruSeq Rapid Exome Kit was replaced with a skirted PCR plate to facilitate automation. All hybridization and capture steps were automated on the Agilent Bravo liquid handling system.

3. Preparation of libraries for cluster amplification and sequencing

After post-capture enrichment, library pools were quantified using qPCR (automated assay on the Agilent Bravo), using a kit purchased from KAPA Biosystems with probes specific to the ends of the adapters. Based on qPCR quantification, libraries were normalized to 2nM,

then denatured using 0.1 N NaOH on the Hamilton Starlet. After denaturation, libraries were diluted to 20pM using hybridization buffer purchased from Illumina.

4. Cluster amplification and sequencing

Cluster amplification of denatured templates was performed according to the manufacturer's protocol (Illumina) using HiSeq 4000 cluster chemistry and HiSeq 4000 flowcells. Flowcells were sequenced on v1 Sequencing-by-Synthesis chemistry for HiSeq 4000 flowcells. The flowcells are then analyzed using RTA v.1.18.64 or later. Each pool of whole exome libraries was run on paired 76bp runs, reading the dual-indexed sequences to identify molecular indices and sequenced across the number of lanes needed to meet coverage for all libraries in the pool.

Express WES for saliva and tissue [dates: 4/15/2018-6/1/2020]:

1. Library Construction

Library construction was performed as described in Fisher et al., with the following modifications: initial genomic DNA input into shearing was reduced from 3µg to 10-100ng in 50µL of solution. For adapter ligation, Illumina paired end adapters were replaced with palindromic forked adapters, purchased from Integrated DNA Technologies, with unique dual-indexed molecular barcode sequences to facilitate downstream pooling. Kapa HyperPrep reagents in 96-reaction kit format were used for end repair/A-tailing, adapter ligation, and library enrichment PCR. In addition, during the post-enrichment SPRI cleanup, elution volume was reduced to 30µL to maximize library concentration, and a vortexing step was added to maximize the amount of template eluted.

2. In-solution hybrid selection

After library construction, hybridization and capture were performed using the relevant components of Illumina's TruSeq Rapid Exome Kit and following the manufacturer's suggested protocol, with the following exceptions: first, all libraries within a library construction plate were pooled prior to hybridization. Second, the Midi plate from Illumina's TruSeq Rapid Exome Kit was replaced with a skirted PCR plate to facilitate automation. All hybridization and capture steps were automated on the Agilent Bravo liquid handling system.

3. Preparation of libraries for cluster amplification and sequencing

After post-capture enrichment, library pools were quantified using qPCR (automated assay on the Agilent Bravo), using a kit purchased from KAPA Biosystems with probes specific to the ends of the adapters. Based on qPCR quantification, libraries were normalized to 2nM, then denatured using 0.2 N NaOH on the Hamilton Starlet. After denaturation, libraries were diluted to 20pM using hybridization buffer purchased from Illumina.

4. Cluster amplification and sequencing

Cluster amplification of denatured templates was performed according to the manufacturer's protocol (Illumina) using exclusion amplification cluster chemistry and HiSeq X flowcells. Flowcells were sequenced on v2.5 Sequencing-by-Synthesis chemistry for HiSeq X flowcells. The flowcells are then analyzed using RTA v.2.7.0 or later. Each pool of whole exome libraries was run on paired 76bp runs, reading the dual-indexed sequences to identify molecular indices and sequenced across the number of lanes needed to meet coverage for all libraries in the pool.

Deep ICE Exome from Non-UMI Enabled ULP Libraries Methods [dates: previous to 8/13/2017]:

1. Library Construction

Initial DNA input is normalized to be within the range of 25-52.5 ng in 50 uL of TE buffer (10mM Tris HCl 1mM EDTA, pH 8.0) according to picogreen quantification. For adapter ligation, Illumina paired end adapters were replaced with palindromic forked adapters, purchased from Integrated DNA Technologies, with unique dual-indexed molecular barcode sequences to facilitate downstream pooling. With the exception of the palindromic forked adapters, the reagents used for end repair, A-base addition, adapter ligation, and library enrichment PCR were purchased from KAPA Biosciences in 96-reaction kits. In addition, during the post-enrichment SPRI cleanup, elution volume was reduced to 30μL to maximize library concentration, and a vortexing step was added to maximize the amount of template eluted.

2. In-solution hybrid selection

After library construction, hybridization and capture were performed using the relevant components of Illumina's Nextera Rapid Capture Exome Kit and following the manufacturer's suggested protocol, with the following exceptions: first, all libraries within a library construction plate were pooled prior to hybridization. Second, the Midi plate from Illumina's Nextera Rapid Capture Exome Kit was replaced with a skirted PCR plate to facilitate automation. All hybridization and capture steps were automated on the Agilent Bravo liquid handling system.

3. Preparation of libraries for cluster amplification and sequencing

After post-capture enrichment, library pools are quantified using qPCR (automated assay on the Agilent Bravo), using a kit purchased from KAPA Biosystems with probes specific to the

ends of the adapters. Based on qPCR quantification, pools are normalized using a Hamilton Starlet to 2nM and sequenced using Illumina sequencing technology.

4. Cluster amplification and sequencing

Cluster amplification of library pools was performed according to the manufacturer's protocol (Illumina) using Exclusion Amplification cluster chemistry and HiSeq X flowcells. Flowcells were sequenced on v2 Sequencing-by-Synthesis chemistry for HiSeq X flowcells. The flowcells are then analyzed using RTA v.2.7.3 or later. Each pool of libraries was run on paired 151bp runs, reading the dual-indexed sequences to identify molecular indices and sequenced across the number of lanes needed to meet coverage for all libraries in the pool.

Deep ICE Exome from UMI-Enabled ULP Libraries [dates: 8/13/2017-6/1/2020]:

1. Library Construction

Initial DNA input is normalized to be within the range of 25-52.5 ng in 50 uL of TE buffer (10mM Tris HCl 1mM EDTA, pH 8.0) according to picogreen quantification. Library preparation is performed using a commercially available kit provided by KAPA Biosystems (KAPA HyperPrep Kit with Library Amplification product KK8504) and IDT's duplex UMI adapters. Unique 8-base dual index sequences embedded within the p5 and p7 primers (purchased from IDT) are added during PCR. Enzymatic clean-ups are performed using Beckman Coultier AMPure XP beads with elution volumes reduced to 30μL to maximize library concentration.

2. Post Library Construction Quantification and Normalization

Library quantification was performed using the Invitrogen Quant-It broad range dsDNA quantification assay kit (Thermo Scientific Catalog: Q33130) with a 1:200 PicoGreen dilution.

Following quantification, each library is normalized to a concentration of 25 ng/μL, using Tris-HCl, 10mM, pH 8.0.

3. In-solution hybrid selection

After library construction, hybridization and capture were performed using the relevant components of Illumina's TruSeq Rapid Exome Kit and following the manufacturer's suggested protocol, with the following exceptions: first, all libraries within a library construction plate were pooled prior to hybridization. Second, the Midi plate from Illumina's TruSeq Rapid Exome Kit was replaced with a skirted PCR plate to facilitate automation. All hybridization and capture steps were automated on the Agilent Bravo liquid handling system.

4. Preparation of libraries for cluster amplification and sequencing

After post-capture enrichment, library pools are quantified using qPCR (automated assay on the Agilent Bravo), using a kit purchased from KAPA Biosystems with probes specific to the ends of the adapters. Based on qPCR quantification, pools are normalized using a Hamilton Starlet to 2nM and sequenced using Illumina sequencing technology.

5. Cluster amplification and sequencing

Cluster amplification of library pools was performed according to the manufacturer's protocol (Illumina) using Exclusion Amplification cluster chemistry and HiSeq X flowcells. Flowcells were sequenced on v2 Sequencing-by-Synthesis chemistry for HiSeq X flowcells. The flowcells are then analyzed using RTA v.2.7.3 or later. Each pool of libraries was run on paired 151bp runs, reading the dual-indexed sequences to identify molecular indices and sequenced across the number of lanes needed to meet coverage for all libraries in the pool.

SUPPLEMENTAL REFERENCES

1. Sztupinszki, Z., Diossy, M., Krzystanek, M., Reiniger, L., Csabai, I., Favero, F., Birkbak, N.J., Eklund, A.C., Syed, A., and Szallasi, Z. (2018). Migrating the SNP array-based homologous recombination deficiency measures to next generation sequencing data of breast cancer. Npj Breast Cancer *4*, 16, 10.1038/s41523-018-0066-6.
2. Adalsteinsson, V.A., Ha, G., Freeman, S.S., Choudhury, A.D., Stover, D.G., Parsons, H.A., Gydush, G., Reed, S.C., Rotem, D., Rhoades, J., et al. (2017). Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. Nat. Commun. *8*, 1324, 10.1038/s41467-017-00965-y.
3. Mateo, J., Carreira, S., Sandhu, S., Miranda, S., Mossop, H., Perez-Lopez, R., Nava Rodrigues, D., Robinson, D., Omlin, A., Tunariu, N., et al. (2015). DNA-Repair Defects and Olaparib in Metastatic Prostate Cancer. N. Engl. J. Med. *373*, 1697–1708, 10.1056/NEJMoa1506859.
4. de Bono, J., Mateo, J., Fizazi, K., Saad, F., Shore, N., Sandhu, S., Chi, K.N., Sartor, O., Agarwal, N., Olmos, D., et al. (2020). Olaparib for Metastatic Castration-Resistant Prostate Cancer. N. Engl. J. Med. *382*, 2091–2102, 10.1056/NEJMoa1911440.
5. Pritchard, C.C., Mateo, J., Walsh, M.F., De Sarkar, N., Abida, W., Beltran, H., Garofalo, A., Gulati, R., Carreira, S., Eeles, R., et al. (2016). Inherited DNA-Repair Gene Mutations in Men with Metastatic Prostate Cancer. N. Engl. J. Med. *375*, 443–453, 10.1056/NEJMoa1603144.
6. AlDubayan, S. (2019). Considerations of multigene test findings among men with prostate cancer - knowns and unknowns. Can. J. Urol. *26*, 14–16.

7. Tate, J.G., Bamford, S., Jubb, H.C., Sondka, Z., Beare, D.M., Bindal, N., Boutselakis, H., Cole, C.G., Creatore, C., Dawson, E., et al. (2019). COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res. *47*, D941–D947, 10.1093/nar/gky1015.
8. Huang, K.-L., Mashl, R.J., Wu, Y., Ritter, D.I., Wang, J., Oh, C., Paczkowska, M., Reynolds, S., Wyczalkowski, M.A., Oak, N., et al. (2018). Pathogenic Germline Variants in 10,389 Adult Cancers. Cell *173*, 355-370.e14, 10.1016/j.cell.2018.03.039.
9. Rahman, N. (2014). Realizing the promise of cancer predisposition genes. Nature *505*, 302–308, 10.1038/nature12981.
10. Painter, C.A., Jain, E., Tomson, B.N., Dunphy, M., Stoddard, R.E., Thomas, B.S., Damon, A.L., Shah, S., Kim, D., Gómez Tejeda Zañudo, J., et al. (2020). The Angiosarcoma Project: enabling genomic and clinical discoveries in a rare cancer through patient-partnered research. Nat. Med. *26*, 181–187, 10.1038/s41591-019-0749-z.
11. Fisher, S., Barry, A., Abreu, J., Minie, B., Nolan, J., Delorey, T.M., Young, G., Fennell, T.J., Allen, A., Ambrogio, L., et al. (2011). A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. Genome Biol. *12*, R1, 10.1186/gb-2011-12-1-r1.