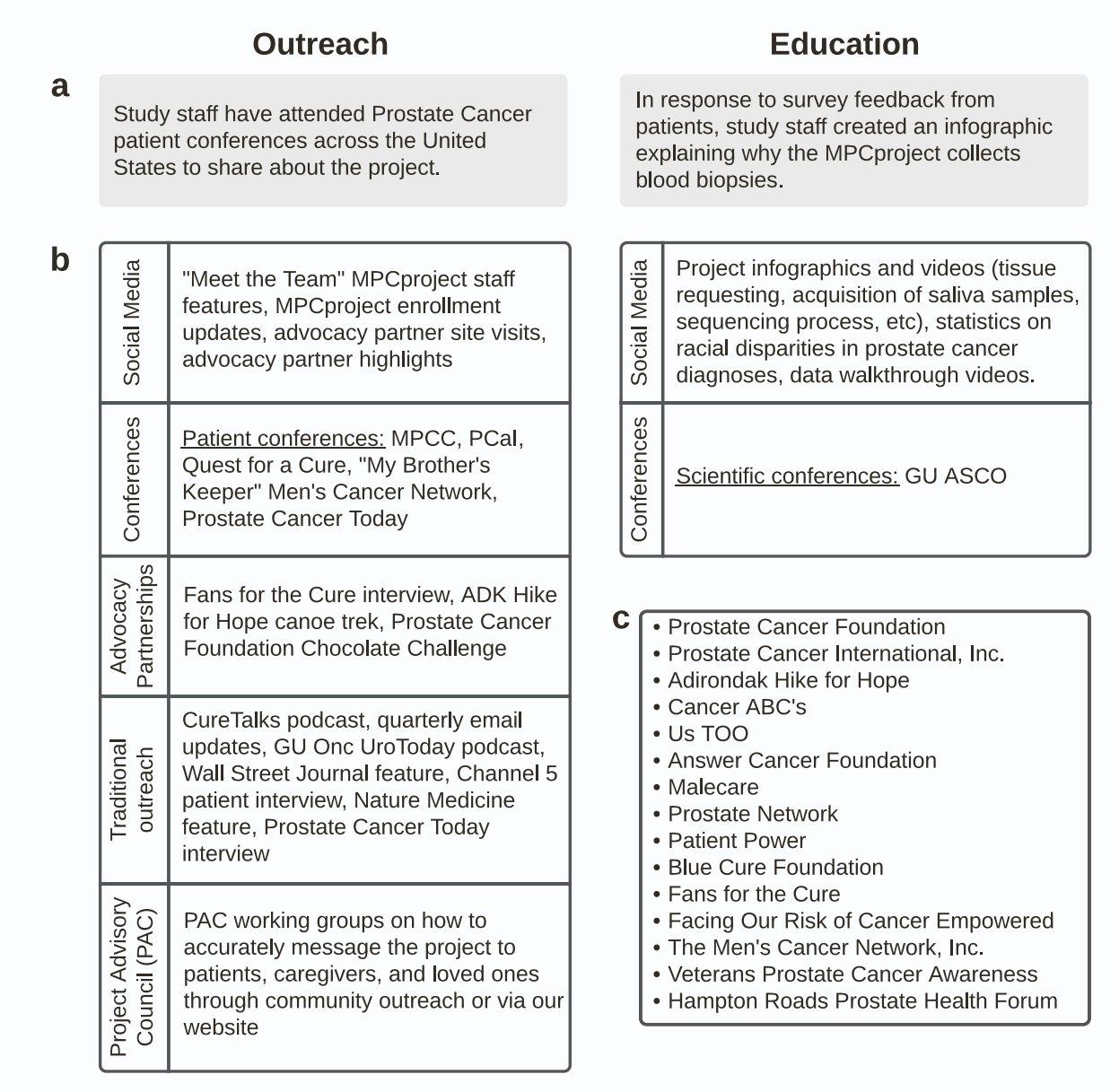
Cell Genomics, Volume *2*

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

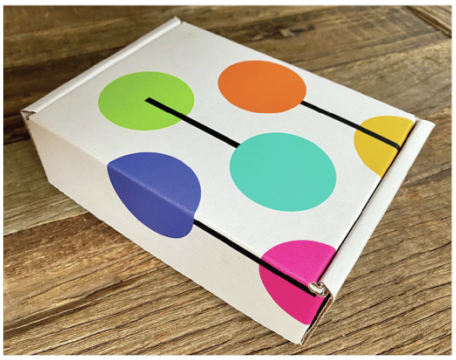
I understand that the information I entered here will be stored in a secure database and may be used to match me to one or more research studies conducted by the Metastatic Prostate Cancer Project. If the information that I entered matches a study being conducted by the Metastatic Prostate Cancer Project, 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

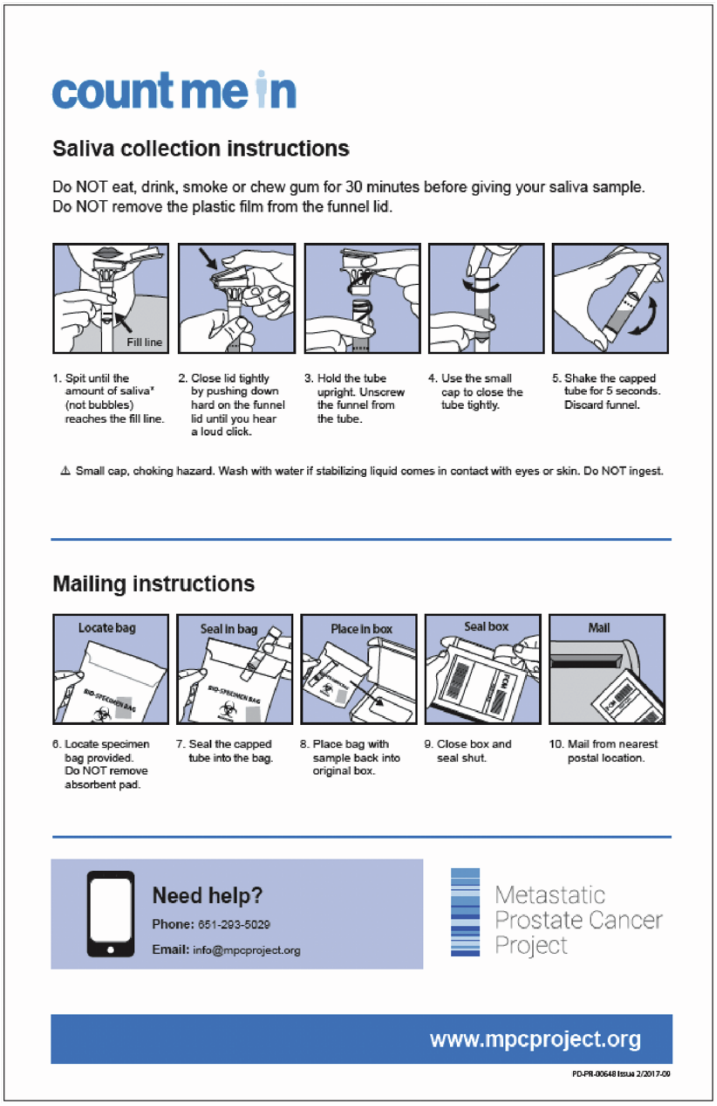
Supplementary Figure S2. MPCproject About You Intake Survey, related to STAR Methods.

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



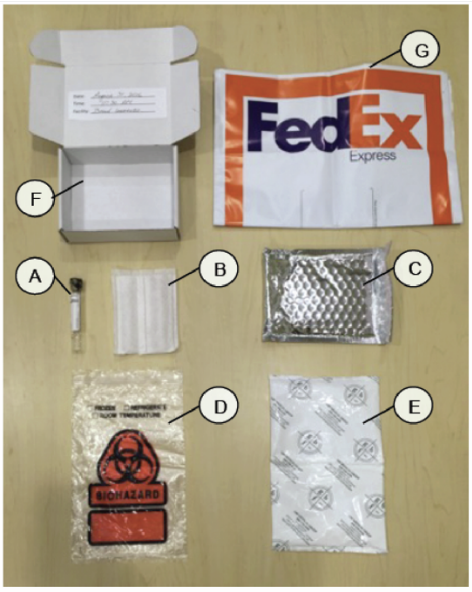
Supplementary Figure S3. MPCproject remote saliva donation kit, related to STAR Methods.

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.

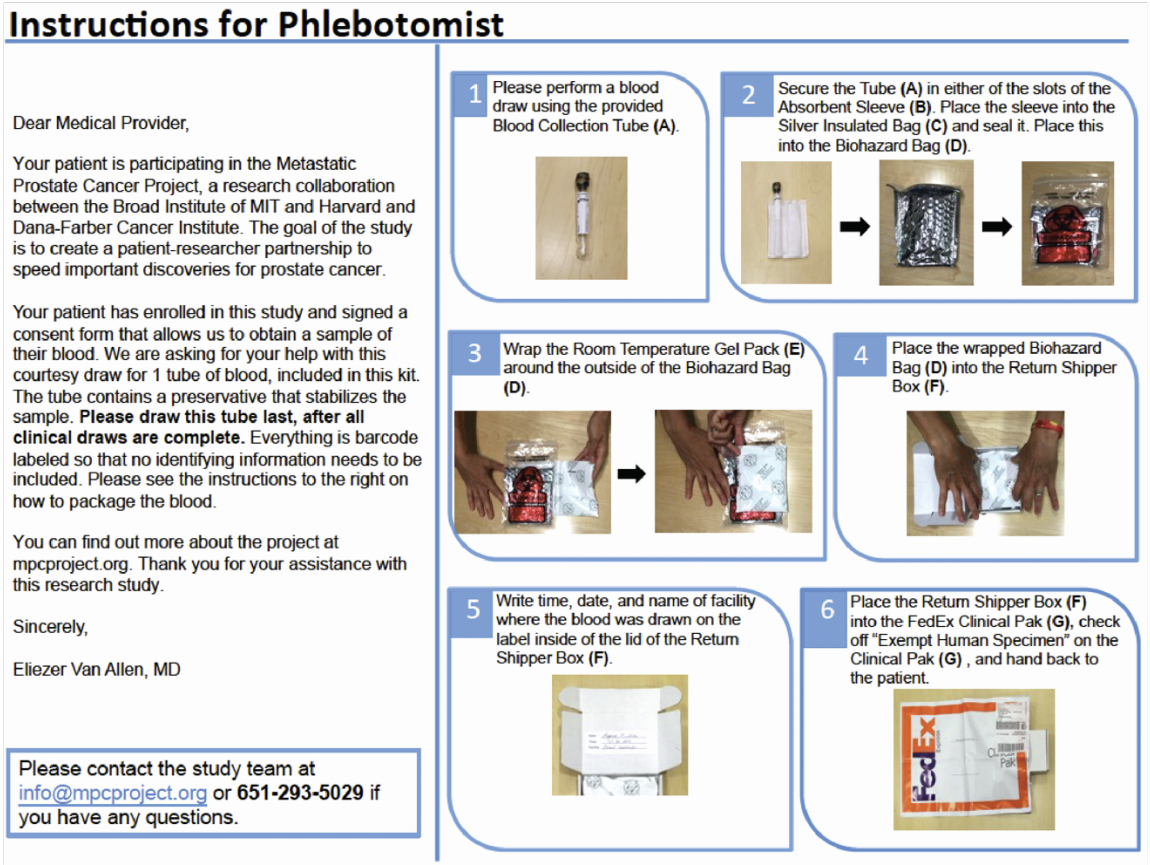
a

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



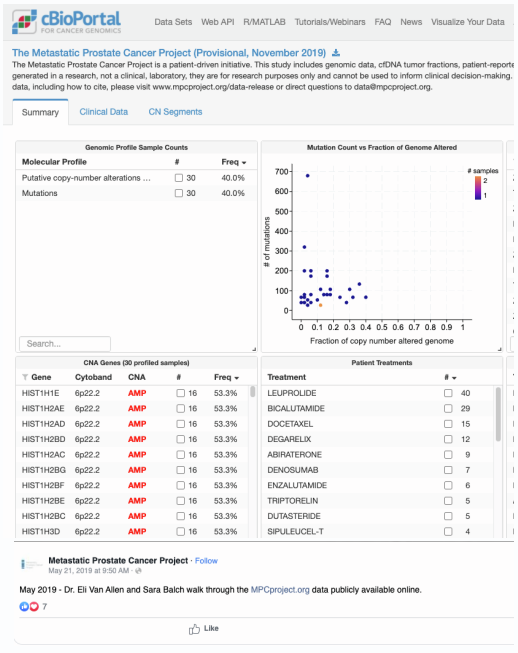
Supplementary Figure S4. MPCproject blood donation kit, related to STAR Methods.

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

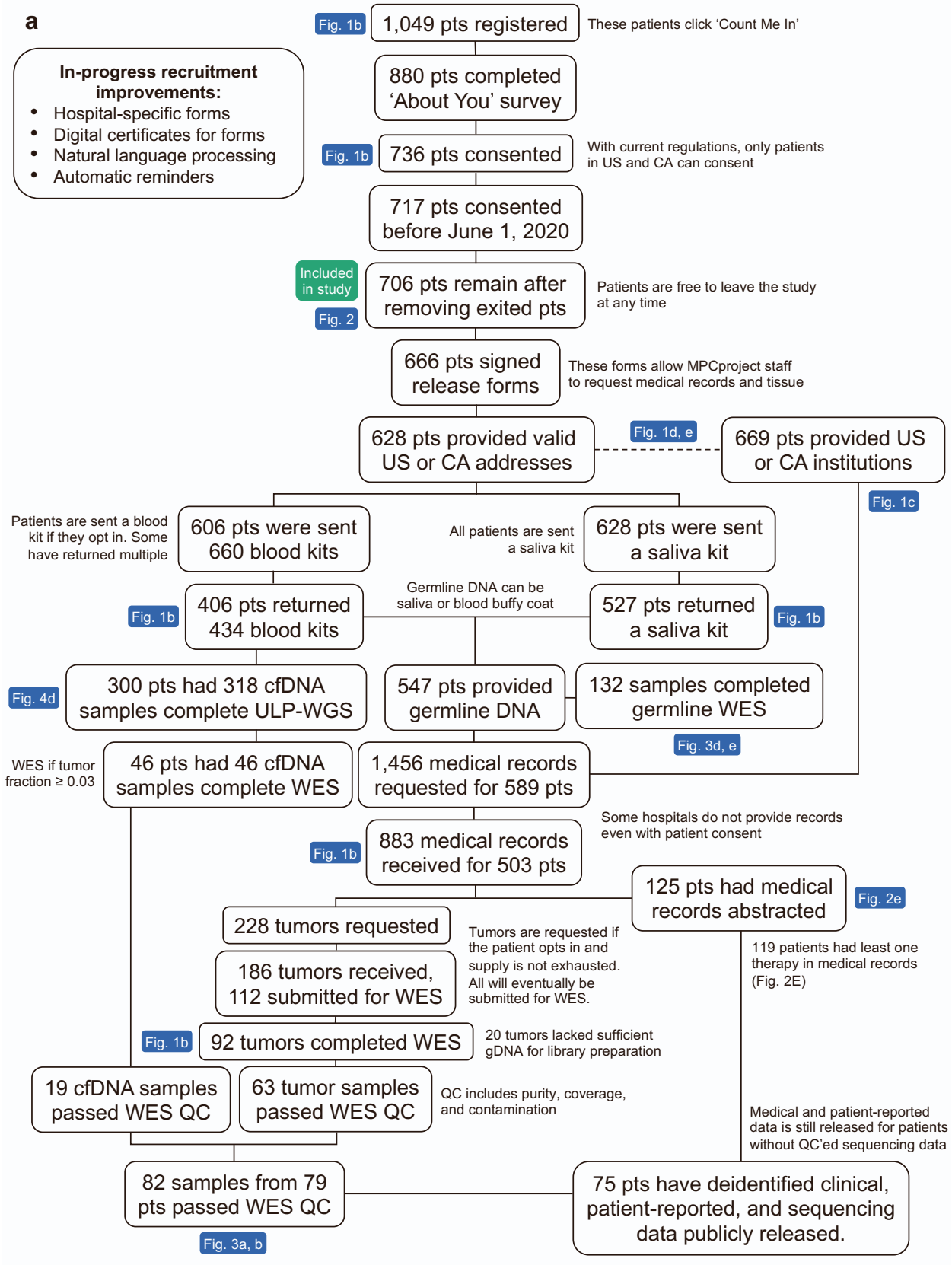
The MPCproject recently released a video of Dr. Eli Van Allen walking through the genomic and clinical data we have released thus far on cBioPortal. We will continue to release data as it is generated, pre-publication. If you have any questions regarding the data currently available, please email us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

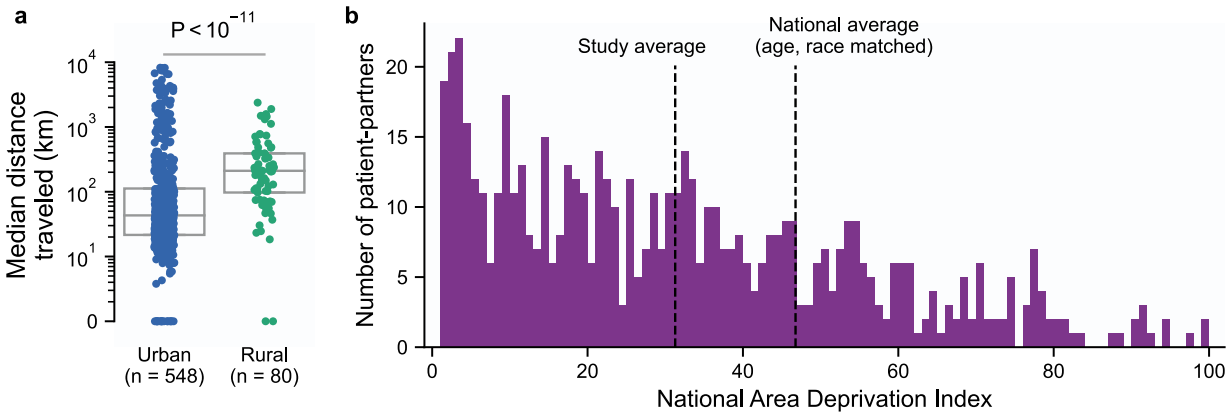

a
In-progress recruitment improvements:
• Hospital-specific forms
• Digital certificates for forms
• Natural language processing
• Automatic reminders
Fig. 1b
1,049 pts registered
These patients click 'Count Me In'
880 pts completed 'About You' survey
Fig. 1b
736 pts consented
With current regulations, 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.

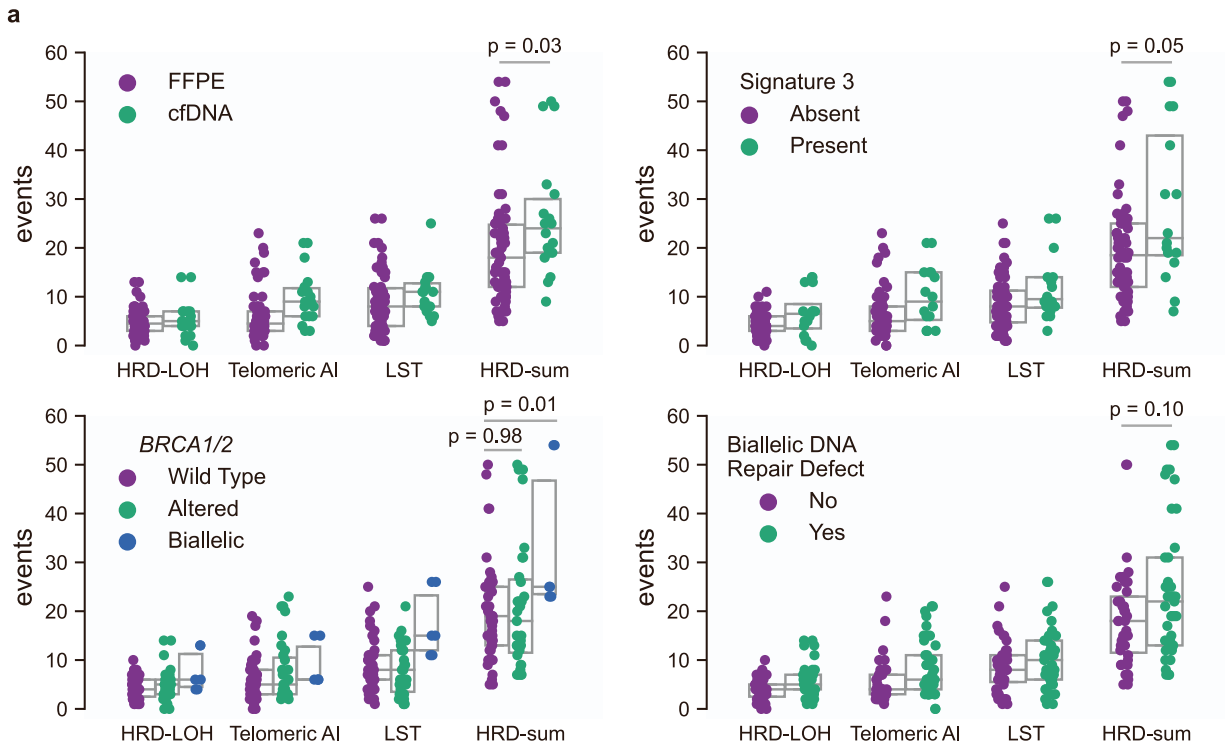
b) Real-time and proposed improvements to improve participant recruitment. Several steps in the study process have been targeted for improvement, with implement and potential changes listed.



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.

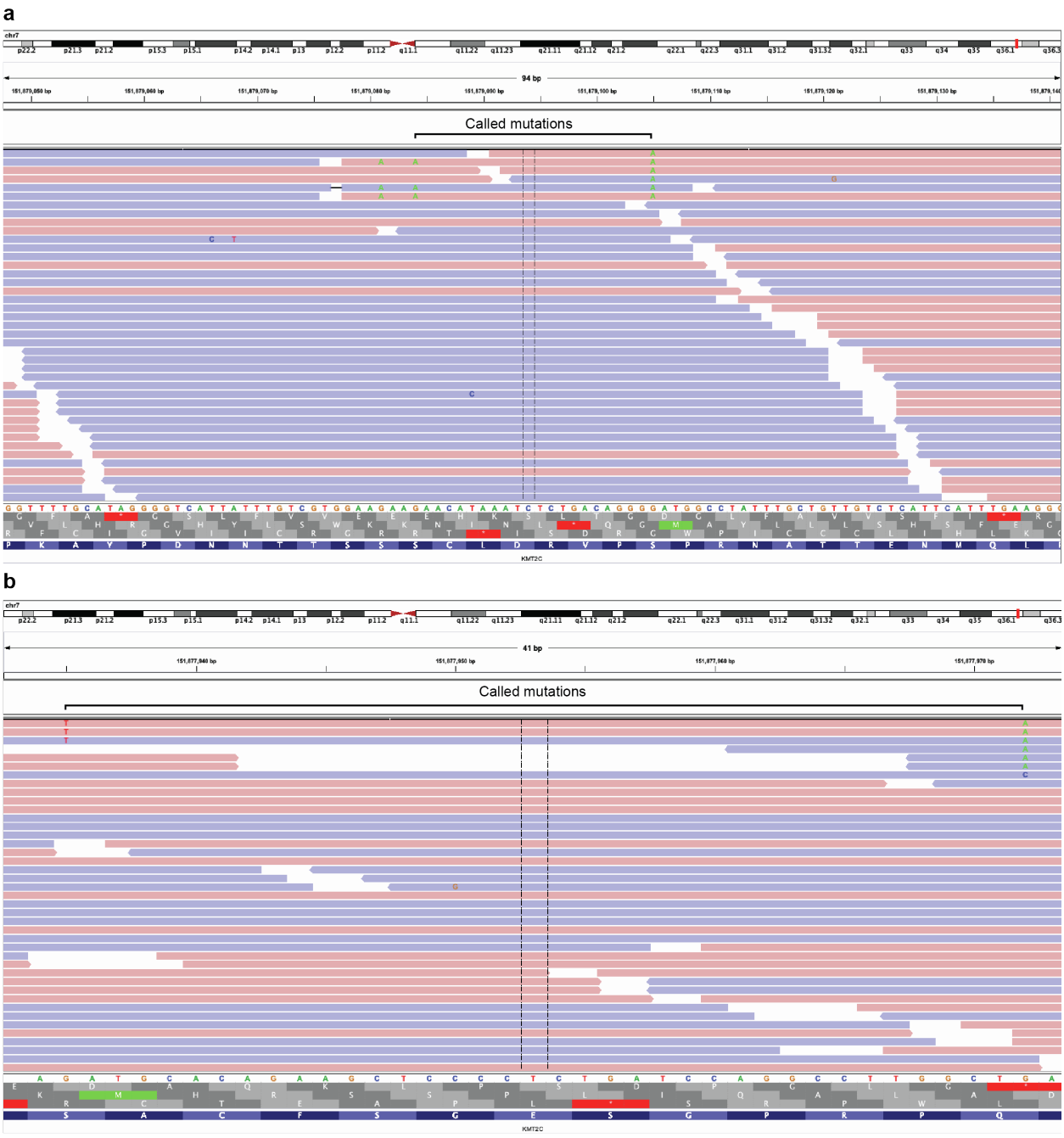
b) National Area Deprivation Index (ADI) distribution of patient-partner residential areas. Higher values indicate higher degrees of disadvantage. The national average was calculated by overlapping census block group populations counts from the U.S. census with public ADI data, weighting by age and race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

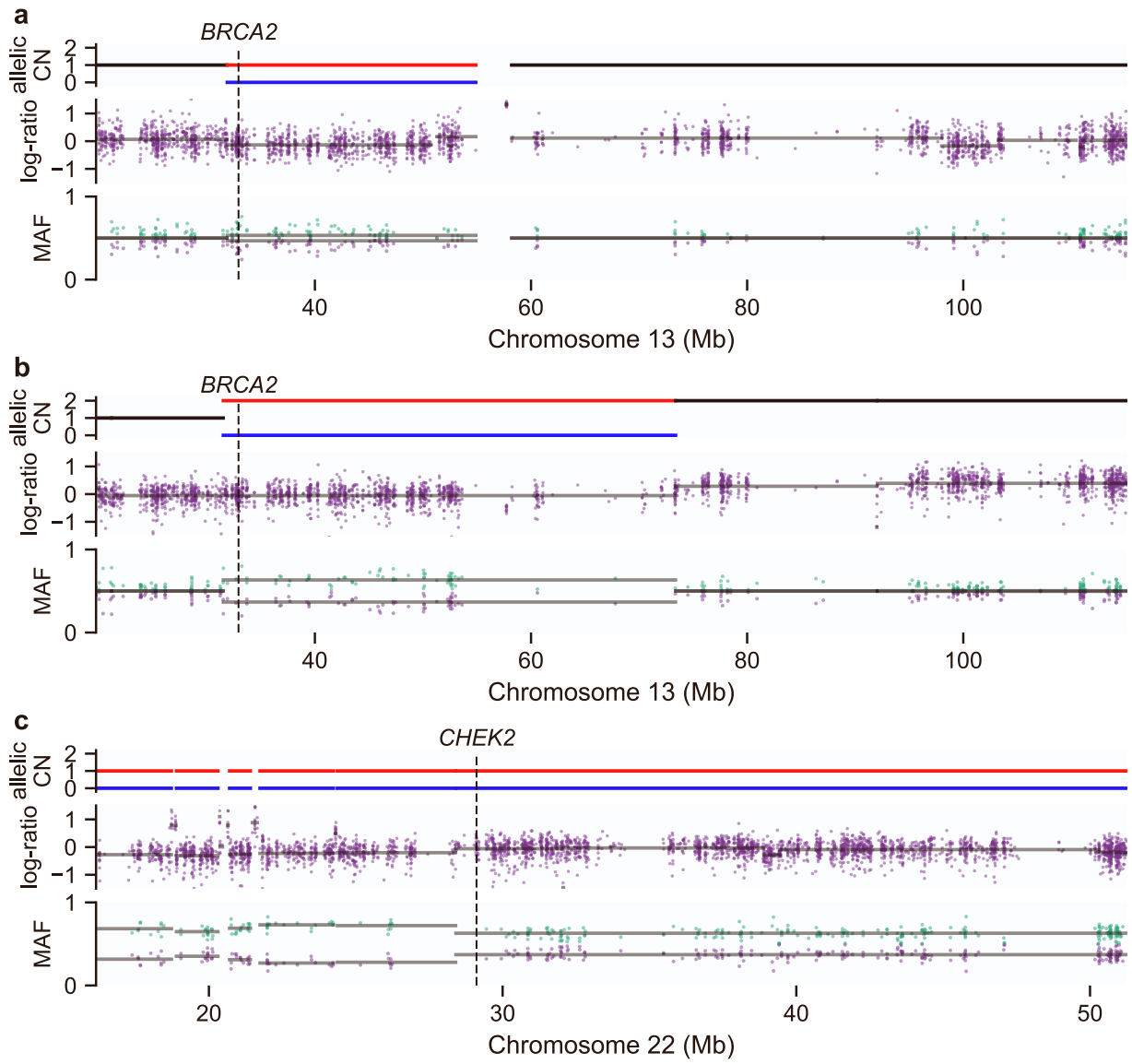
prior literature, as the cfDNA samples represent metastatic disease, although we cannot rule out the confounding influence of analyte type.



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

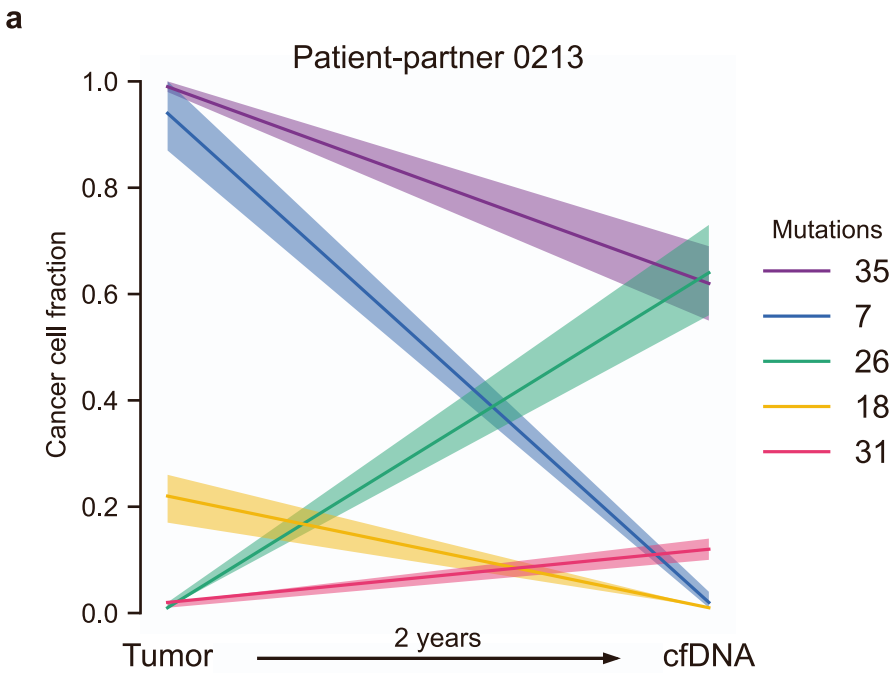
a) IGV screenshot containing reads that span somatic *KMT2C* mutations (chr7:151879084/p.S1947F and chr7:151879105/p.S1954F) in the cfDNA sample of patient-

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.



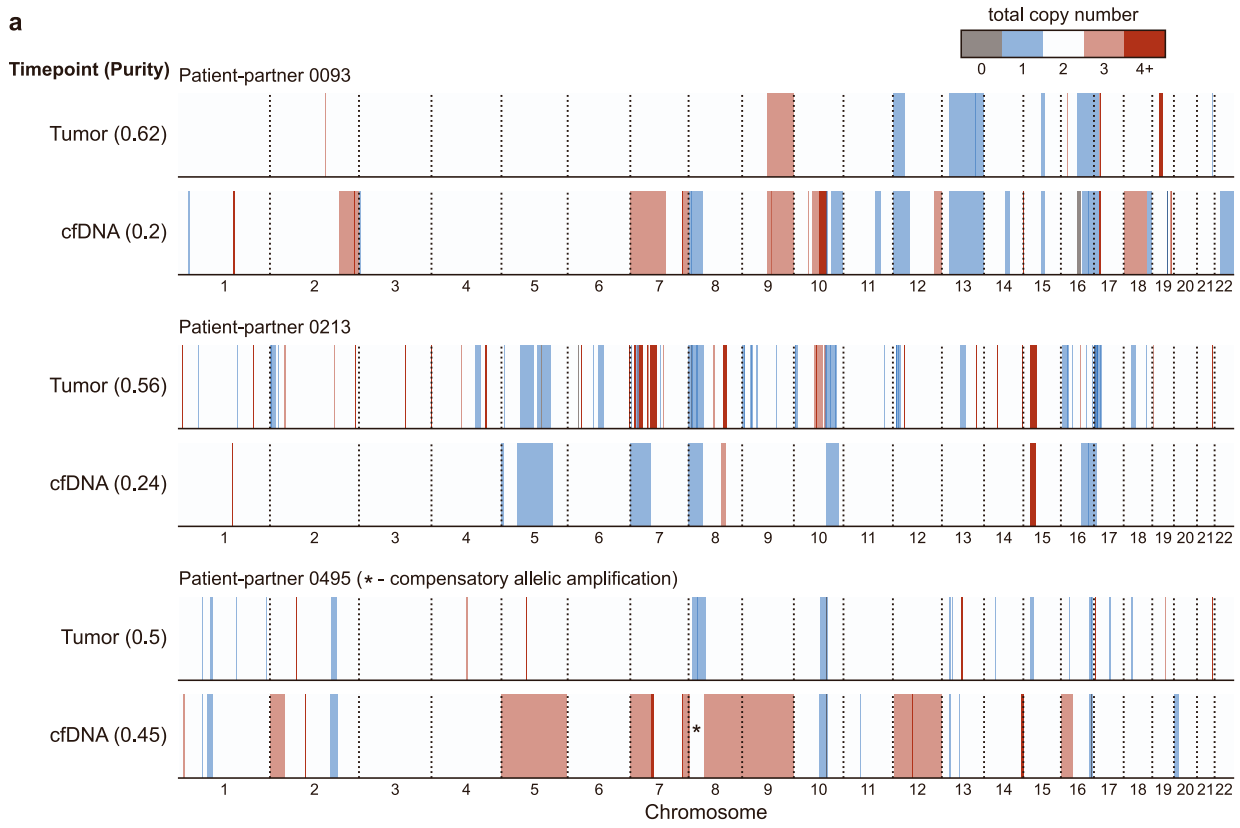
Supplementary Figure S10. Somatic losses affecting genes with pathogenic germline mutations, related to Fig. 3E.

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.



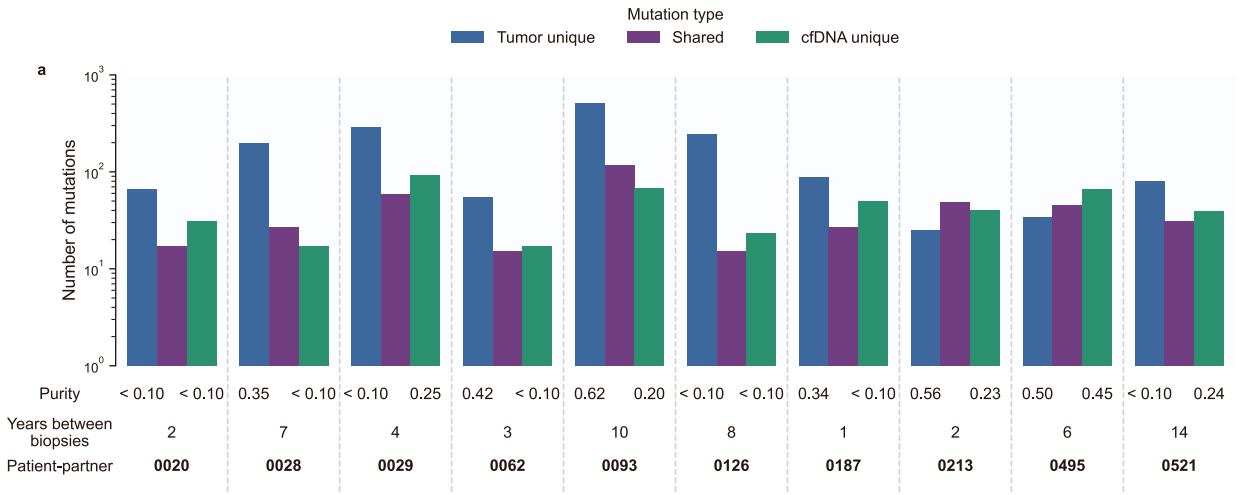
Supplementary Figure S11. Phylogenetics of samples from patient-partner 0213, related to Fig. 4B.

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.



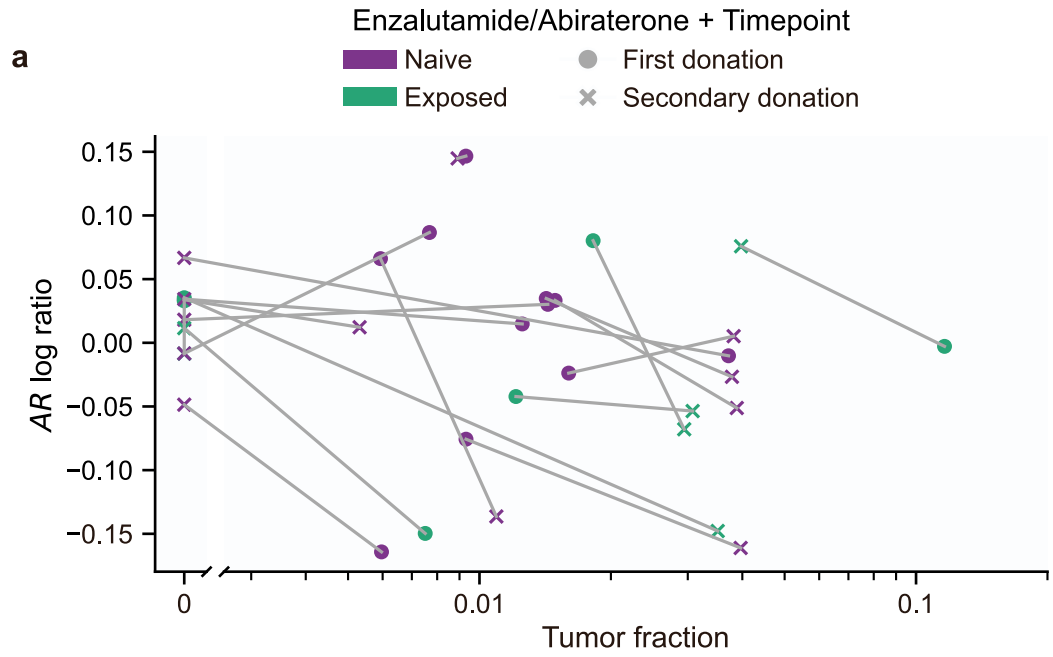
Supplementary Figure S12. Copy number profiles of shared tumor and cfDNA samples, related to Fig. 4B.

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.



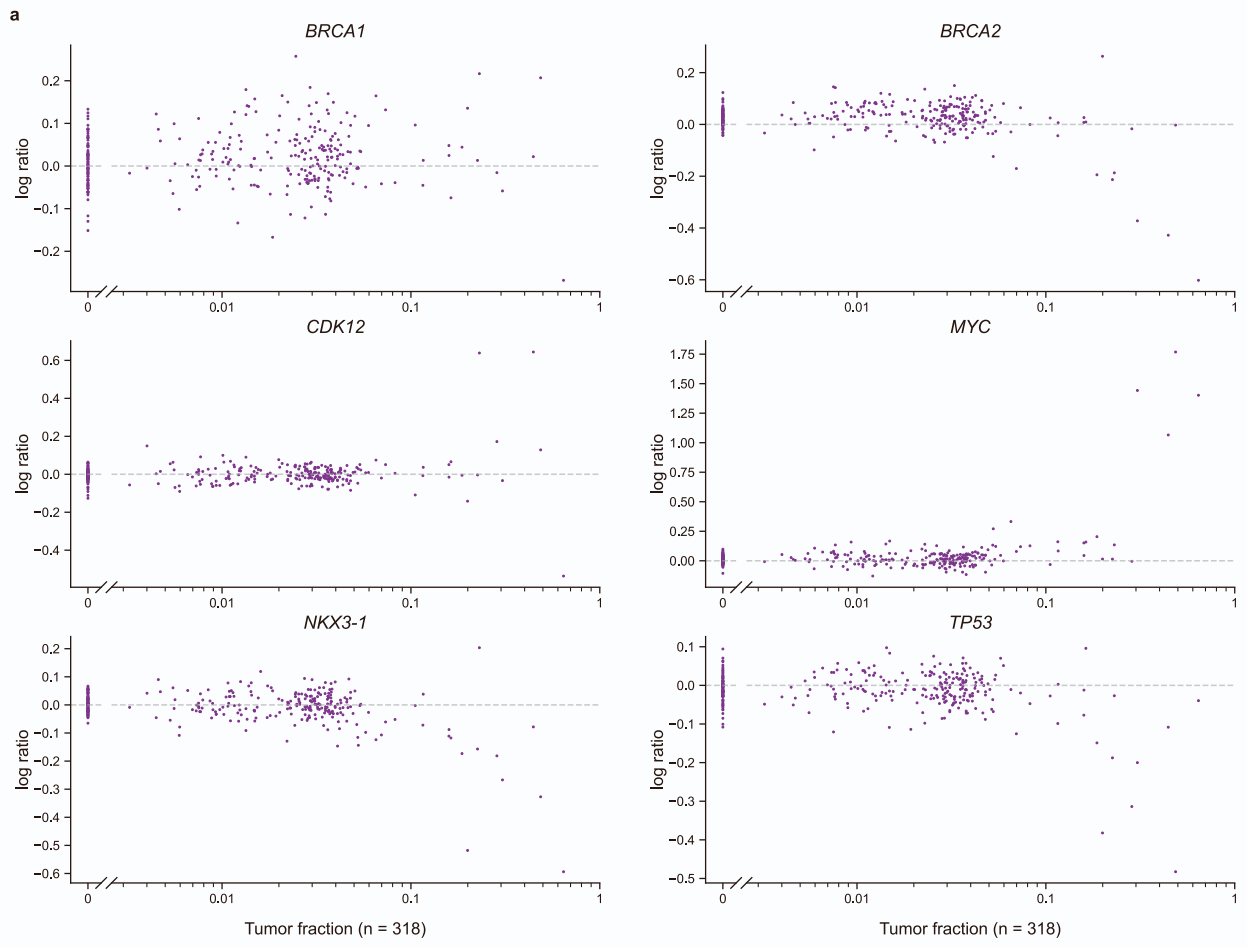
Supplementary Figure S13. Mutation exclusivity between tumor and cfDNA samples from the same patient, related to Fig. 4B.

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 33

Establishing patient partnership 33

Patient Enrollment and Informed Consent 33

Medical Records 35

Samples 36

Saliva 36

Archived Tumor Tissue 36

Primary and Secondary Blood Samples 37

Data Generation 38

Medical Record Abstraction 38

Patient-Reported Data 5

Study inclusion 5

Cleaning/categorization 5

Medical Institutions 5

Therapies 5

Alternative lifestyles 38

Genomic Sequencing 40

DNA Isolation 40

Saliva 40

cfDNA Extraction from Whole Blood 41

Ultra-Low Pass Whole Genome Sequencing (ULP-WGS) 41

non-UMI ULP-WGS sequencing [dates: 2017-2/11/2018] 41

UMI ULP-WGS sequencing [dates: 2/12/2018-6/1/2020] 42

Whole Exome Sequencing (WES) 44

Express WES for saliva and tissue [dates: 8/13/2017 - 4/15/2018] 44

Express WES for saliva and tissue [dates: 4/15/2018-6/1/2020] 45

Deep ICE Exome from Non-UMI Enabled ULP Libraries Methods [dates: previous to 8/13/2017] 47

Deep ICE Exome from UMI-Enabled ULP Libraries [dates: 8/13/2017-6/1/2020] 48

References 49

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.