AWARD NUMBER: W81XWH-18-1-0199

TITLE: Mutational Landscape of the Y Chromosome and Prostate Cancer

PRINCIPAL INVESTIGATOR: Nikolaus D. Schultz, PhD

CONTRACTING ORGANIZATION: Sloan Kettering Institute for Cancer Research

REPORT DATE: OCTOBER 2022

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Development Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
OCTOBER 2022	Final	15JUN2018 - 14JUN2022

4. TITLE AND SUBTITLE

Mutational Landscape of the Y Chromosome and Prostate Cancer

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-18-1-0199

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Nikolaus Schultz

E-Mail: schultzn@mskcc.org

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Sloan Kettering Institute for
Cancer Research
1275 York Avenue
New York, NY 10065

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Development Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Mortality from prostate cancer (PC), an estimated 33,330 deaths in 2020, is associated with development of aggressive and treatment-insensitive metastatic castration-resistant prostate cancer (mCRPC). We will investigate the status and role of Y chromosome (ChrY) genes in regulating drug sensitivity and mCRPC development and progression. Though ChrY loss in men is associated with increased risk of disease and mortality, the role of ChrY genes in regulating PC progression is poorly understood. To investigate the clinical impact of ChrY gene expression, we developed new methodology to analyze mutational variants of ChrY genes in PC patient cohorts, previously unsuccessful due to the high number of repetitive sequences and paralog families. Using a custom reference for each paralog family, our method increased ChrY read depth coverage to be on par with whole-exome sequencing allowing for normal/tumor variant calling. We also generated the first CRISPR/Cas9 library targeting human ChrY to further understand the role of individual ChrY genes in regulating antiandrogen treatment sensitivity and mCRPC development in PC models in vitro and in vivo. This multifaceted approach will potentially identify predictive markers for treatment sensitivity based on ChrY. These markers will allow for development of tailored therapies and serve as targets for drug development.

15. SUBJECT TERMS

Metastatic castration-resistant prostate cancer, Y chromosome, antiandrogen therapy, epigenetics, biomarkers, tumor suppressor, mutations, CRISPR/Cas9 library screening

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRDC
Unclassified	Unclassified	Unclassified	Unclassified	33	19b. TELEPHONE NUMBER *(include area code)*

TABLE OF CONTENTS

1. INTRODUCTION:

Prostate cancer (PC) is the second most common cancer and second leading cause of cancer death among men in the United States with an estimated 33,330 deaths in 2020. PC associated mortality is attributed to the development of metastatic castration-resistant prostate cancer (mCRPC) which is characterized by its aggressiveness and poor response to treatment. Though loss of the Y chromosome (ChrY) in men has been associated with increased risk of disease and mortality, the role of ChrY genes in disease progression is poorly understood (Dumanski et al., 2016; Forsberg, 2017; Noveski et al., 2016). Our team presented the first report of a ChrY gene, *KDM5D*, which regulates tumor growth and docetaxel sensitivity through epigenetic modification of key cell cycle regulators and androgen receptor signaling (Komura et al., 2016; Komura et al., 2018). The study also reported the loss of *KDM5D* to be associated with increased mortality and aggressive disease in patient cohorts suggesting its role as a potential biomarker for mCRPC. Together, these studies highlight the urgency to further explore the role of ChrY genes in PC progression and further determine its mutational landscape to develop therapeutic targets as well as biomarkers and gene expression signatures which will allow physicians to predict drug response in patients and thereby prescribe effective treatment regiments. This multidisciplinary approach will help determine the clinical impact of ChrY genes on PC progression and treatment resistance.

2. KEYWORDS: Prostate cancer, metastatic castration-resistant prostate cancer, Y chromosome, antiandrogen therapy, drug insensitivity, docetaxel, epigenetics, biomarkers, tumor suppressor, precision medicine, mutations, CRISPR/Cas9 library screening

3. ACCOMPLISHMENTS:

What were the major goals of the project?

The major goals of the project as outlined in the SOW are:

SPECIFIC AIM 1: To determine the mutational landscape of the Y chromosome (ChrY) in men with prostate cancer in the SU2C/PCF, TCGA, and other cohorts

Major Task 1: Structural analysis of the Y chromosome (ChrY). This goal is 100% complete, in accordance with the SOW (1–36 months, responsible PIs and sites: Schultz, MSK; Van Allen, DFCI).

Major Task 1, Subtask 1: Identify the samples with ChrY loss. This goal is 100% complete, in accordance with the SOW (1–36 months, responsible PIs and sites: Schultz, MSK; Van Allen, DFCI).

Major Task 1, Subtask 2: Quantify the focality of ChrY loss. This goal is 100% complete, in accordance with the SOW (1–36 months, responsible PI and site: Schultz, MSK).

Major Task 1, Subtask 3: Assess mutual exclusivity of ChrY loss with genomic lesions in prostate cancer pathways. This goal is 100% complete, in accordance with the SOW (1–36 months, responsible PI and site: Schultz, MSK).

Major Task 1, Milestones: Define the extent of ChrY loss in metastatic prostate cancer and evaluate the association with clinically actionable signaling pathways This goal is 100% complete, in accordance with the SOW (1–36 months, responsible PI and site: Schultz, MSK).

Major Task 2: Determine functional features associated with ChrY mutations. This goal is 100% complete, in accordance with the SOW (1–36 months, responsible PI and site: Schultz, MSK).

Major Task 2, Subtask 1: Identify the putative tumor suppressors that are inactivated on ChrY. This goal is 1000% complete, in accordance with the SOW (1–36 months, responsible PI and site: Schultz, MSK).

Major Task 2, Subtask 2: Assess differential AR activity between samples that show ChrY loss and samples without alterations on ChrY. This goal is 100% complete, in accordance with the SOW (1–36 months, responsible PI and site: Schultz, MSK).

Major Task 2, Subtask 3: Correlation of ChrY loss with Gleason score and sample type. This goal is 100% complete, in accordance with the SOW (1–36 months, responsible PI and site: Schultz, MSK).

Major Task 2, Milestones: Define the mutational landscape of ChrY and determine if the LOY is significantly associated with disease risk. This goal is 100% complete, in accordance with the SOW (1–36 months, responsible PI and site: Schultz, MSK).

What was accomplished under these goals?

Major progress has been made towards the aims outlined in the original application, following the timeline indicated in the SOW.

SPECIFIC AIM 1: To determine the mutational landscape of the Y chromosome (ChrY) in men with prostate cancer in the SU2C/PCF, TCGA, and other cohorts

Major Activities

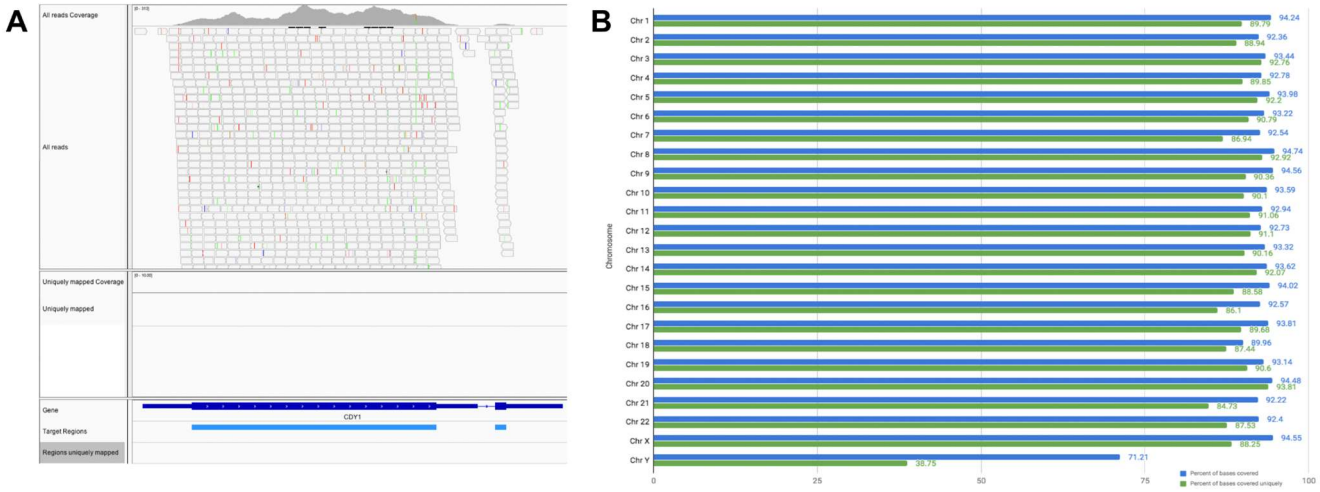
The analysis of the landscape of somatic copy-number alterations and somatic mutations of the ChrY in PC were supervised by the Schultz/Van Allen groups at MSK and DFCI. To circumvent the problem of numerous paralogs on ChrY, we first developed a new method to specifically call mutations in paralogous genes, allowing us to map the landscape of mutations on the ChrY. We then switched our focus from mutation calling to determining copy-number alterations, which was another challenge on ChrY, due to the fact that there is only a single copy, and all existing copy-number methods were written for chromosomes that exist in two copies. Lastly, we assessed the correlation between ChrY loss and outcome across cancer types in multiple cohorts.

Specific Objectives

The specific objectives proposed in the SOW were to 1) identify the samples with ChrY loss 2) quantify the focality of ChrY loss, 3) assess mutual exclusivity of ChrY loss with genomic lesions in prostate cancer pathways, 4) identify the putative tumor suppressors that are inactivated on ChrY, 5) assess differential AR activity between samples that show ChrY loss and samples without alterations on ChrY, 6) correlate ChrY alterations with Gleason score and sample type, and 7) define the mutational landscape of ChrY and determine if loss of the Y chromosome is significantly associated with disease risk.

Significant Results or Key Outcomes:

Among the 44 protein-coding genes present on the Y chromosome, the TCGA prostate cancer data set contained only 11 non-silents mutations in 6 genes. The most frequently mutated gene on the Y chromosome was *PCDH11Y* (n=5, 1.2%) followed by *KDM5D (n=2, 0.5%).* A total of 11 patients (2.6%) had at least one non-silent mutation. However, 29 of these 44 genes belong to a paralog family. Among the six mutated genes reported by TCGA, only one belongs to a paralog family (*TSPY2*), which was lower than expected by chance (P = 0.018, Monte-Carlo test). The numerous Y chromosome genes belonging to paralog families are known to be challenging for aligning using short-read sequencing technologies. Many genes in the human genome were duplicated during evolution and are still transcriptionally and translationally active. These duplicated regions distributed along the genome pose a problem for short-reads alignment. For example, the Y chromosome gene *CDY1* has 99% homology with *CDY1B*, *CDY2A,* and *CDY2B*, making variant calling impossible. We first sought to investigate the extent of the problem at the nucleotide and gene level across chromosomes. As shown in **Figure 1**, the Y chromosome is most impacted by the drop in coverage due to non-uniquely mapped reads. Only ~38% of the Y chromosome region is covered uniquely, making variant calling difficult, as aligners will map reads randomly to one of these regions and the mapping quality (MAPQ) of the read will be assigned as 0, which is then ignored by mutation callers.



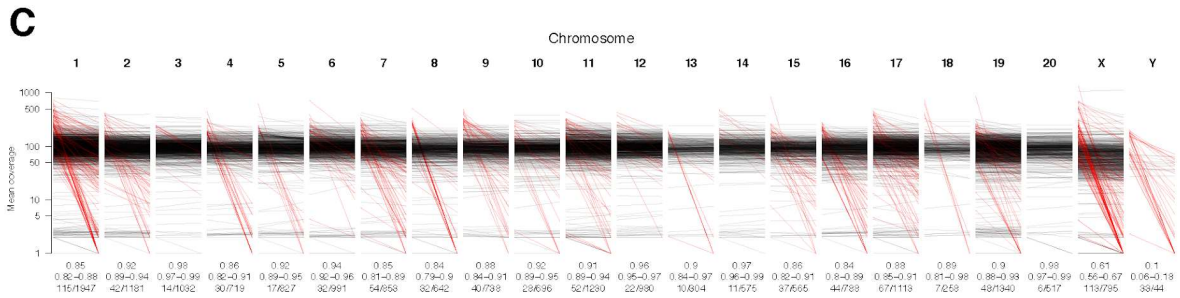


Figure 1. Somatic variant calling of most Y chromosome genes is challenging.
A) An example of reads that do not uniquely map to *CDY1* due to the presence of multiple paralogs distributed along the Y chromosome (white filled reads indicate reads with low mapping quality). B) Percent of bases covered before removing low quality reads (in blue) and in percent of bases covered uniquely (in green) from one representative sample from the TCGA. C) ladder-plot showing the mean coverage of every gene before and after removing reads that do not uniquely map from one representative sample from the TCGA cohort. Each ladder-plot represents a chromosome. Each line represents a gene. Lines are color-coded according to whether the corresponding gene can (black) or cannot (red) be accurately assessed for variant calling. Geometric mean changes, 95% CI, and the proportion of genes that cannot be accurately assessed for variant calling due to a major drop in coverage are shown below each ladder-plot.

To avoid this prob,lem we developed a new method to specifically call mutations in paralogs. Briefly, for each paralog family, we realigned the reads using a custom reference, which is the longest representative gene from its paralog family while masking the other genes of the family (**Fig. 2**).

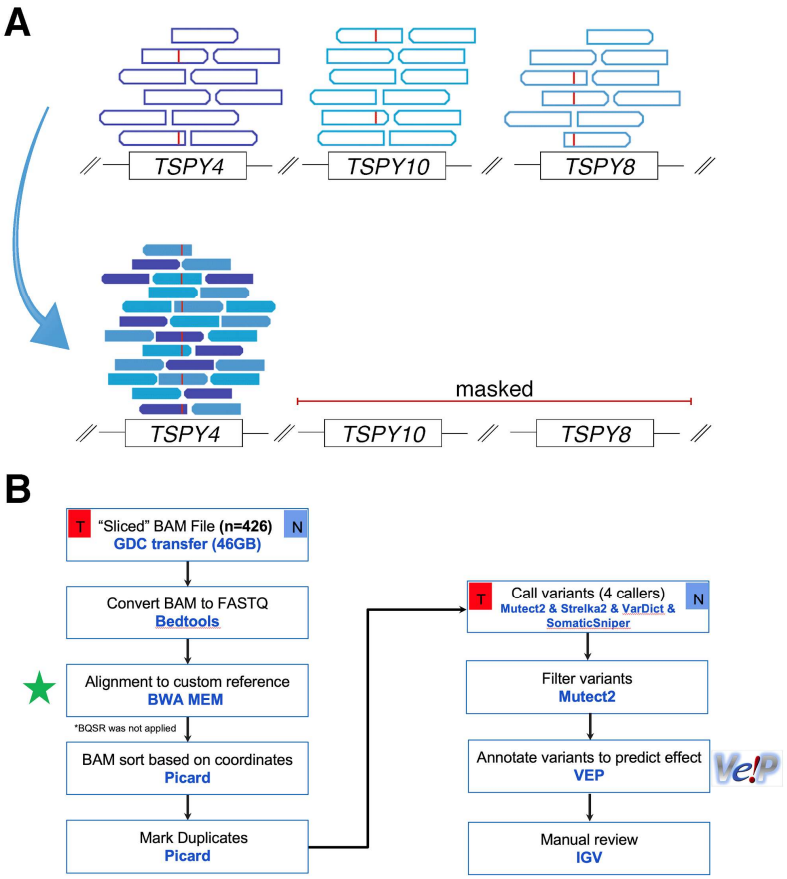


Figure 2. Strategy employed to call mutations present in paralogs.
A, Schematic representation of the crucial step of the strategy employed to call mutations in Y chromosome paralogs (eg. *TSPY4*). *TSPY4* has many paralogs within the Y chromosome. Standard aligners will randomly align reads across these paralogs and assign a low mapping quality and will be discarded for subsequent variant calling (represented by white filled reads). Red lines indicate an individual's mutation that exists in one of these paralogs but reads containing this mutation also get scattered and assigned a low mapping quality. The crucial step of our method is to realign all the reads to one custom reference representative of the paralogs family. B, Flowchart representing the workflow used for paralogs tumor/normal variant calling. Each bioinformatics tool used is colored blue. The crucial step is denoted with the green star.

On the Y chromosome, there are eight paralog families representing a total of 29 genes. For each family, we selected a representative gene and employed our tailored strategy for variant calling. After filtering the variants that passed we retrieved a total of 227 paralog variants (**Fig. 3**). The observed read depth coverage and the variant allele frequency distribution was in the range of whole-exome sequencing technologies (**Fig. 3A, B**). We found a majority of intronic variants in *RBMY1D* (n = 73) and *TSPY4* (n = 42) that include large intronic regions (**Fig. 3C**). We found 27 missense variants and 9 truncating variants in 34 patients (34/426, 8%), of which two were previously identified which brings the total number of patients with Y chromosome mutations to 43 (43/426, 10%). Of note, we found several hotspot mutations (defined as present in at least two patients) in *TSPY4*, *BPY2*, *RBMY1D*, *HSFY1*, and *CDY1* as shown in the lollipop plots (**Fig. 3D**).

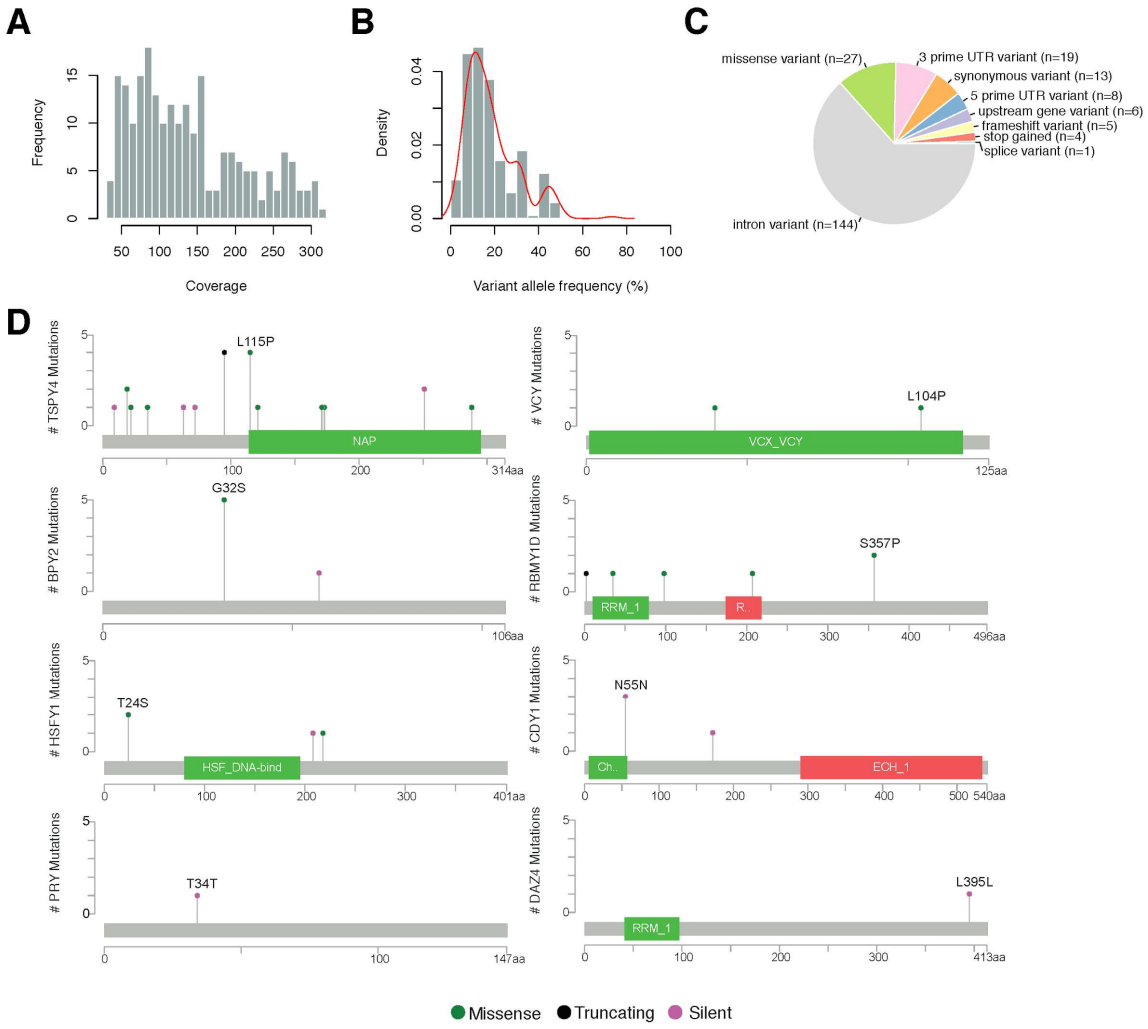


Figure 3. Paralog variants called from the 426 TCGA prostate cancer patients.
A) Histogram showing the coverage distribution form the 227 paralog variants called.B) Distribution of the variant allele frequency of the 227 paralog variants called. C) Pie chart representing the frequency of variants according to protein consequence. D) Variants identified in the 8 representative genes from the 8 paralog families present on the Y chromosome.

We next sought to analyze loss of chromosome Y, first in TCGA samples, then in targeted sequencing data generated locally at MSK (MSK-IMPACT). FacetsY, an extension of the allele-specific copy number analysis tool Facets (Shen et al., 2016) was used to investigate Y-chromosome losses in PC samples. To methodologically justify the utilization of FacetsY, we compared the segmentation profiles of 333 primary prostate cancer samples from The Cancer Genome Atlas (TCGA) obtained with FacetsY and Affymetrix Genome-Wide Human 6.0 Array data. Breakpoint locations as well as signal intensities obtained from both methods are highly similar (**Fig. 4A**). Moreover, chromosome-

arm alteration (defined as >80% of respective chromosome-arm length with segmentation mean > abs(0.2)) frequencies show a high concordance (**Fig. 4B**).

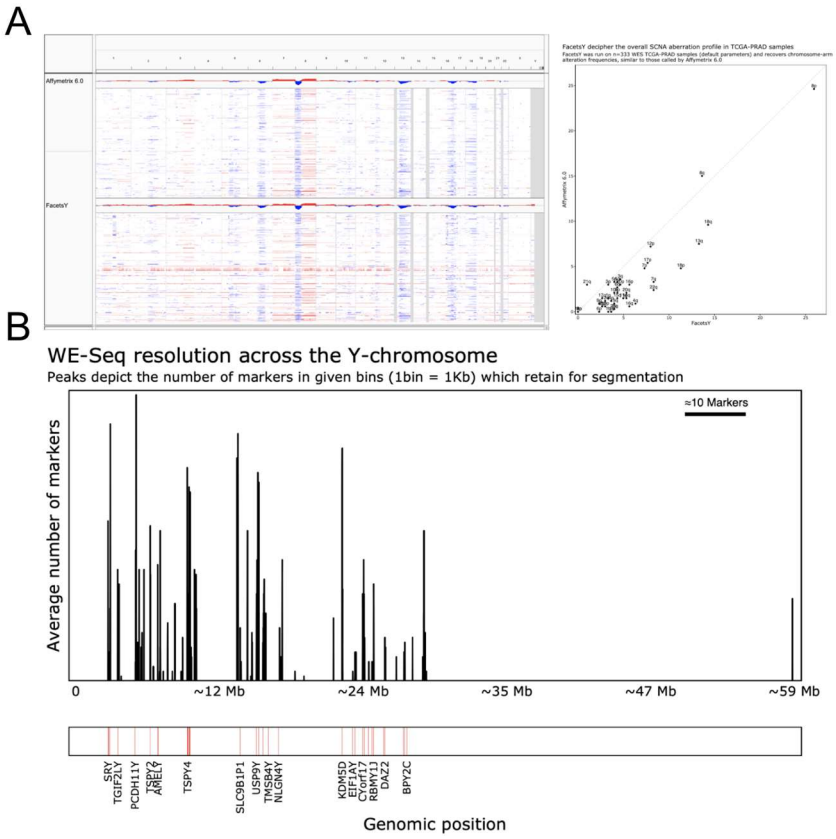


Figure 4. Chromosome Y copy-number analysis.
A) FacetsY is a reliable tool for copy number profiling. Comparison of continuous segmentation profiles obtained by Affymetrix 6.0 and FacetsY on prostate cancer samples from TCGA (n=333). The right panel shows the relationship between chromosome arm alteration frequencies called in either Affymetrix 6.0 (y-axis) or FacetsY (x-axis). B) Whole exome sequencing data provide sufficient information for the assessment of the Y-chromosome status in individual PC patients. Peaks along the x-axis depict the average number of markers, which were retained for segmentation analysis using FacetsY. Red bars in the bottom annotation lane highlight protein coding genes where putative losses can directly be assessed as the marker density is adequate for predictions.

Integer copy-number calls obtained from FacetsY were used to determine the presence of chromosome loss. We defined a sample to have chromosome loss if a chromosome had greater than 50% of total copy number (TCN) to be equal to zero. Due to an integer copy-number transformation (using the median copy-number log ratio of respective called segments) a direct comparison between the two sequencing approaches was enabled. We observed a strong linear relationship ($r = 0.834$; $p < 2.2e\text{-}16$) on a per-sample basis, suggesting that targeted sequencing via MSK-IMPACT provides sufficient information to accurately call Y-chromosome losses (**Fig. 5A**). In addition, we investigated genomic correlates with Y-chromosome loss. A total of 133 patients with available WES sequencing and

clinical annotations were included (one sample selected per patient), out of which 23% of patients (n=30) had ChrY loss and 77% of patients (n=103) had an intact ChrY (intact) (**Fig. 5B**). While fraction genome altered (FGA) along with chr16q and 13q alteration portend to a greater likelihood of Y-chromosome loss, none of the mentioned variables showed significance (**Fig. 5C**).

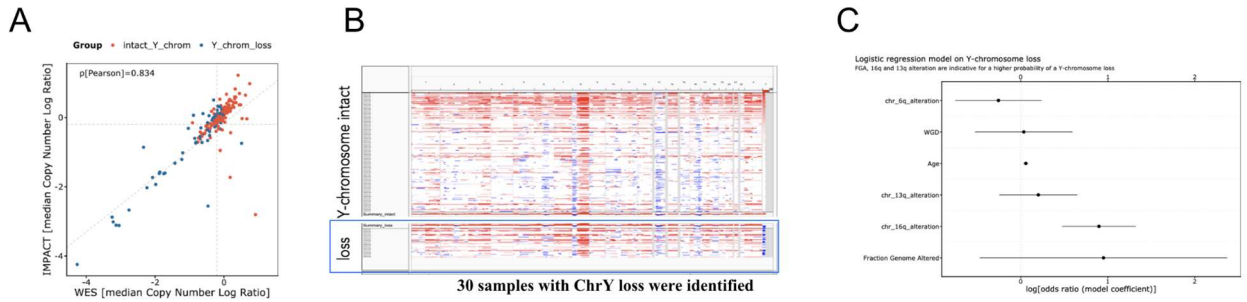


Figure 5. Targeted sequencing provides sufficient information to call ChrY loss.
MSK-IMPACT targeted panel sequencing provides reliable data for ChrY copy-number calling. FacetsY was used in a cohort of 238 PC patients where both MSK-IMPACT and WES data is available. Binary Y-chromosome calls of both sequencing strategies were transferred into a continuous scale and compared against each other. Blue dots represent samples where the Y-chromosome is lost whereas red denotes an intact Y-chromosome. B) Segmentation profiles of the two respective groups. C) A logistic regression model suggests that fraction genome altered (FGA) and specific chromosome-arm alteration may be indicative of a Y-chromosome loss in PC.

To explore the association of known clinical features according to ChrY loss, we examined several features such as sample type, tissue localization, Gleason grade, disease extent and prostate-specific antigen (PSA) (**Fig. 6A**). We did not observe any significant association with any of the above clinical features. In addition, we also performed a genomic analysis to compare the intact with the ChrY loss groups. The loss group was associated with a slightly higher fraction genome alteration (p=0.05); however, it was not associated with TMB (**Fig. 6B**).

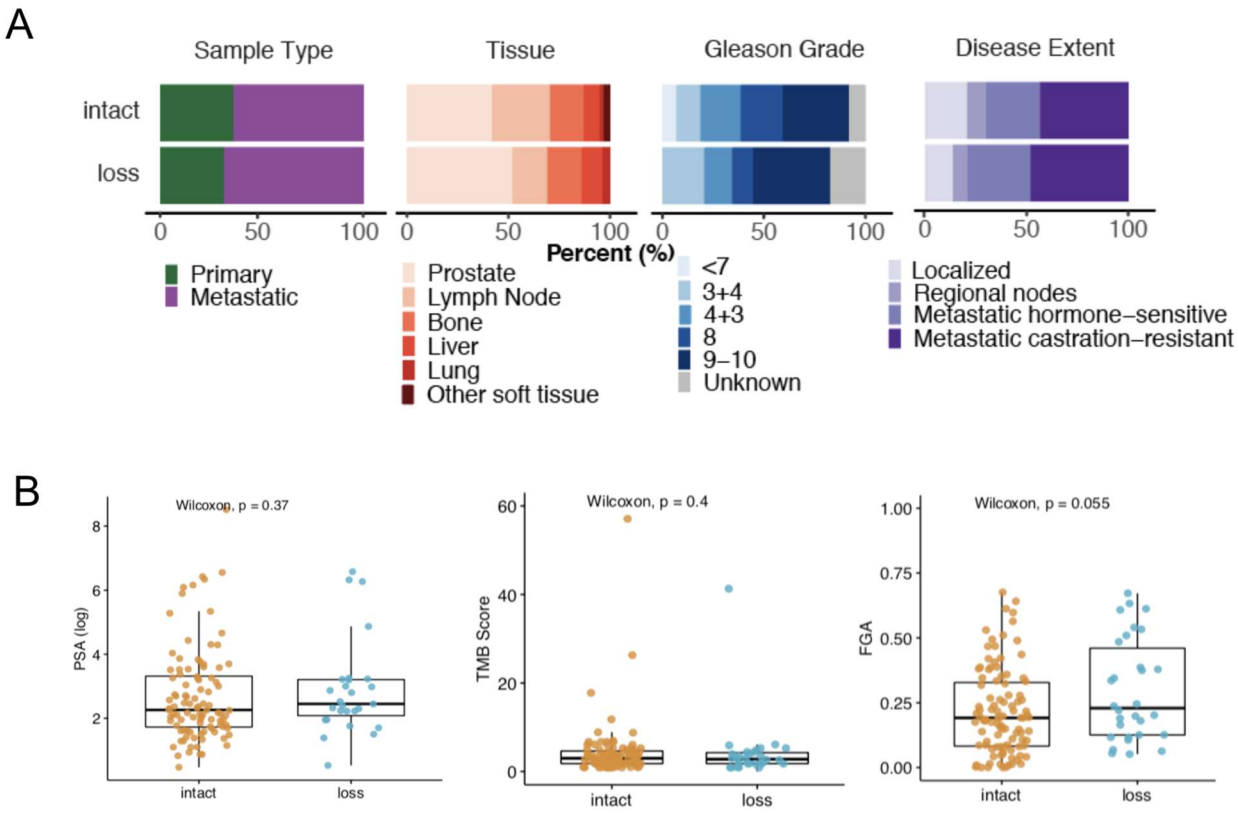


Figure 6. Clinical and pathological associations with Y loss.
A) Comparison of categorical clinical features such as sample type, tissue, gleason grade and disease extent is intact vs loss groups. No significant differences between the groups. B) Comparison of continuous clinical variables such as PSA, tumor mutation burden (TMB) and fraction genome altered (FGA) in intact vs no loss groups. No significant difference between the groups. There is an association of FGA with the chrY loss group.

To evaluate the differences in the prevalence of genomic alterations by ChrY loss status, we selected for gene alterations annotated as oncogenic or likely oncogenic function based on the curated precision oncology database OncoKB, version 2.8 (Sep 17, 2020). We found that samples present in the loss group had a slightly higher frequency of *RB1*, *ZFHX3*, *CCNE1*, *CHEK2*, *FLT3* alterations (p-value < 0.05), however those association were not significant after correcting for multiple testing **(Fig. 7A)**. We also analyzed the prevalence of eleven oncogenic signaling pathways (cell cycle, Myc, Hippo, Notch, PI3K, TGF-Beta, WNT, p53, Nrf2, RTK-RAS and epigenetic pathways (Sanchez-Vega et al., Cell 2018)). We did not observe any significant difference based on pathway level alterations between the two groups (**Fig. 7B**).

We interrogated the association between chrY loss status and overall survival (OS) of prostate cancer patients. For OS, follow-up started at the time of sequencing and ended with patient death, with censorship occurring at last patient contact. Patients with chrY loss had a shorter OS (36.4 months) than those of intact group (44.3 months), and this difference was significant in an unadjusted model (hazard ratio [HR]1.9; with 95% CI (1-3.5), p=0.04) However, when we accounted for known prognostic factor (FGA), the trend remained the same. (HR: 1.3; with 95% CI (0.69-2.5), p=0.49)

(**Fig. 4C**). All analyses were performed using R v3.5.2 (www.R-project.org).

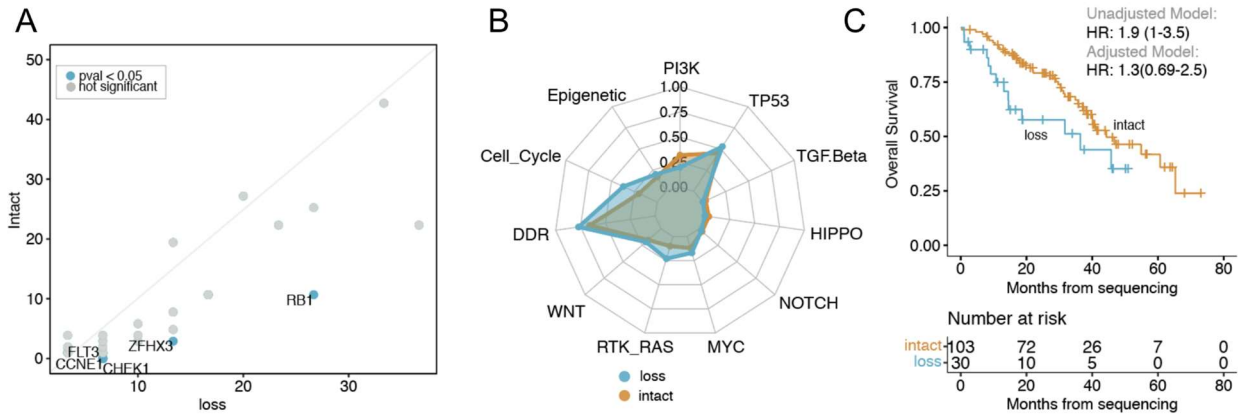


Figure 7. Genomic and survival analysis of ChrY loss samples in MSK-IMPACT WES recapture cohort.

A) Gene enrichment analysis between intact vs loss groups. RB1, ZFHX3,CCNE1 etc., alterations were more frequently observed in samples with ChrY loss. (not significant when corrected for multiple testing) B) Oncogenic pathway level alteration comparison between intact vs loss groups using curated pathways templates. No significant pathway level enrichment between the two groups. C) Kaplan-Meier curve showing unadjusted and adjusted hazard ratios with 95% C.I estimated using Cox proportional hazard models displaying overall survival of patients harboring chrY loss vs intact.

Expansion to other cancer types

Chromosome Y loss is a frequent somatic event among different cancers

A total of 14,322 male tumor samples were eligible for studying chromosome Y loss in cancer. We observed a range of Y copy events, most frequently complete LOY (33.6%). Relative LOY (rLOY), i.e., losing a copy while the overall ploidy is >2, occurred in 1.4% of male tumors. Overall, 5010 male tumors (35%) harbored either complete or relative LOY (**Fig. 8A**). Like other patterns of chromosomal gains and losses, rates of different Y chromosome events varied by cancer type (**Fig. 8B**). Complete LOY was most frequent in esophagogastric cancer (67.3%) and pancreatic cancer (54,9%), and as low as 14.2% in prostate and 11.3% in glioma.

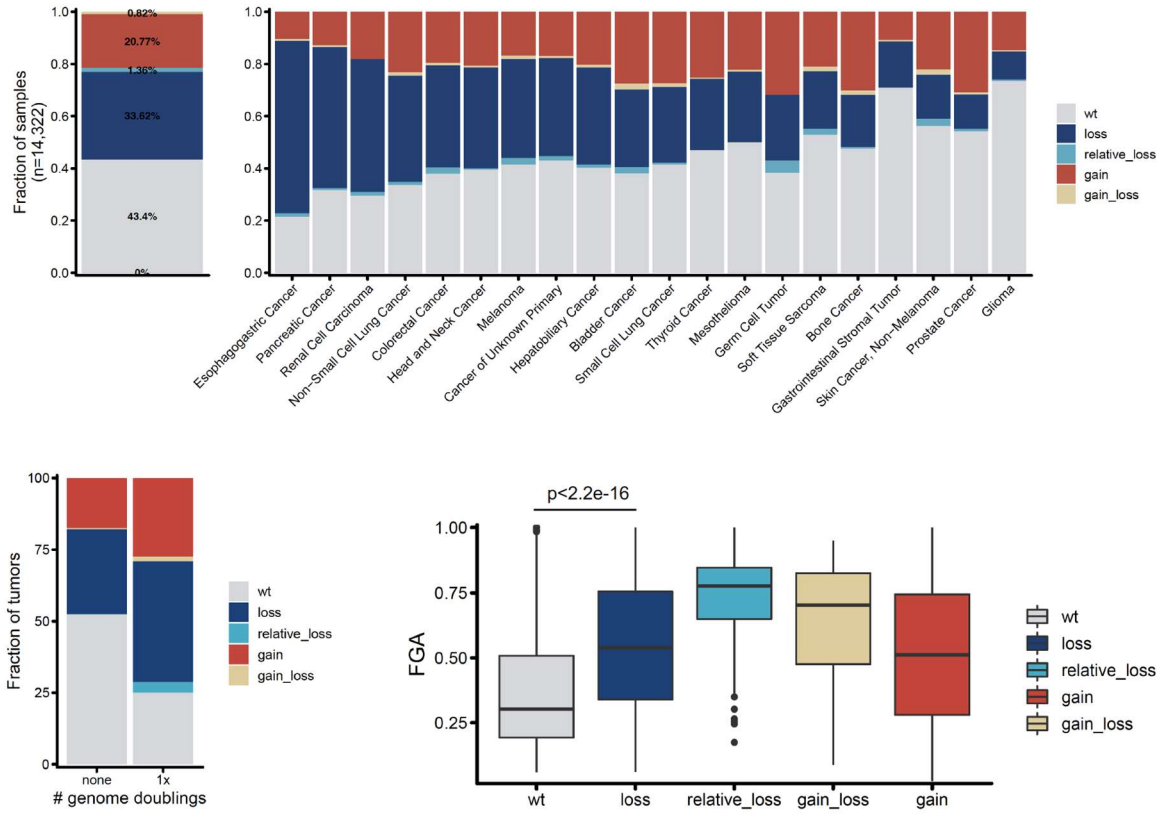


Figure 8. LOY rates across tumor types.
A) Fraction of different Y chromosome alteration for each tumor type. B) Fraction of Y alterations in tumors with no or one genome doubling C) FGA (y-axis) of male tumor samples with WT, Y loss or other Y alterations.

As a small, gene-poor chromosome, Y has a relatively high chance of being lost from cells "by chance", and the paucity of genes on this chromosome that are expressed outside of the male reproductive system suggests that there could be little selective pressure for its retention. In peripheral blood, somatic LOY has been associated with generally increased genomic instability. To assess whether somatic LOY is correlated with genomic instability in male tumors, we compared LOY rates with genome doubling events and aneuploidy, measured as FGA (i.e., fraction of genome altered) of only the autosomes. Concordant with unstable genomes, tumors that had undergone a genome doubling were more likely to harbor various Y copy changes. LOY tumors also had significantly higher FGA estimates (median 0.54; Wilcoxon rank sum test, P<2.2e-16) compared to WT tumors (median 0.3). As expected, FGA were highest in tumors with rLOY as a product of genome doubling (**Fig. 8B, C**).

Focal alteration signals are rarely seen on the Y-chromosome

Our copy number calling method (FacetsY) concentrates on 25.3 Mb (min: 2654800, max: 28000000 bp) which comprises the majority of the Yp arm and about half the Yq arm. FacetsY called one copy

number segment for the majority of male tumor samples (98%; **Fig. 9A**). For 269 tumor samples (1.9%) we obtained 2 segments with breakpoints centering in between the centromere and Yq11.221 (**Fig. 9B**). Interestingly, tumors that were classified as gain_loss (i.e., one segment gained while the other segment was lost) showed that the copy-number gains predominantly occurred on the shorter Yp arm (**Fig. 9C, D**).

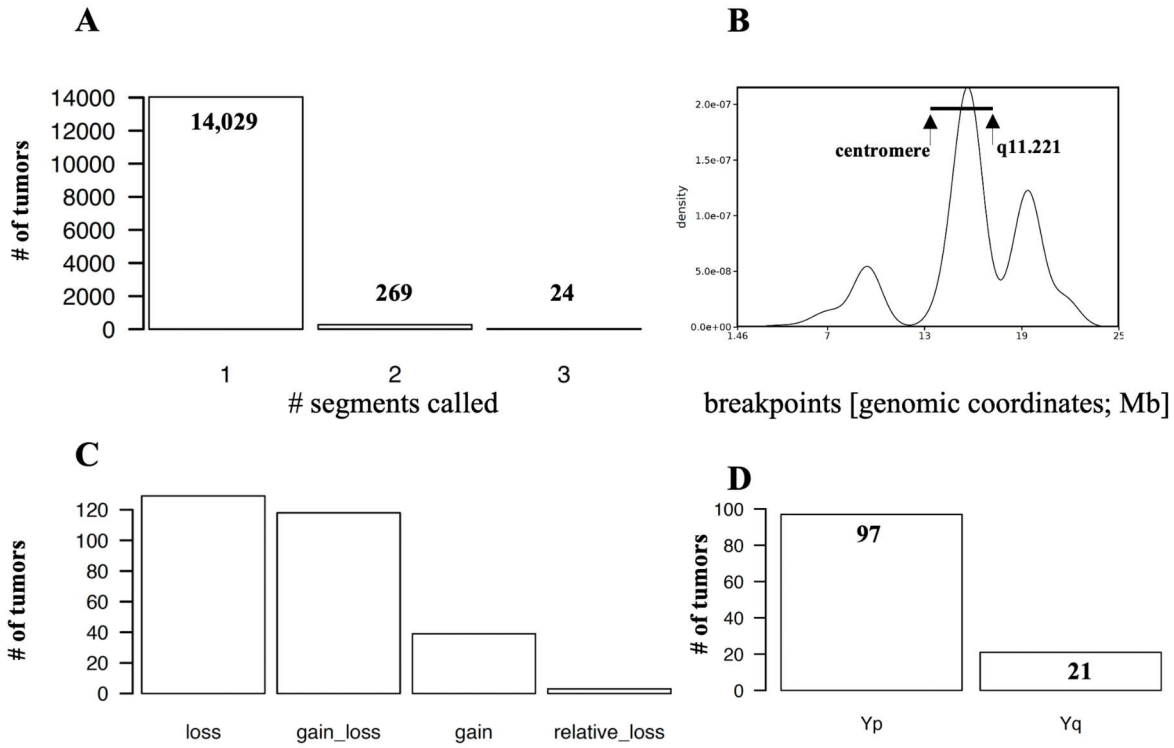


Figure 9. Focality of copy-number calls

A) Number of male tumors where either 1, 2 or 3 segments are called (FacetsY) for the Y chromosome. B) Density plot of breakpoints for samples where >1 segments were called. The black horizontal line indicates the coordinates for the peak in the density plot (min: 12958100; max: 18958100bp). C) Number of tumors (y-axis) for different Y chromosome alterations are plotted where we obtained >1 segment. D) For tumors that have gained and lost one segment respectively, the number of gained segments (y-axis) is plotted for chromosome arms Yp and Yq (x-axis) separately.

Chromosome Y loss rates across primary and metastatic samples

Next, we concentrated on selected cancer types to see whether chromosome Y loss rates differ among sample types (primary and metastasis). We observed that LOY incidences are highest in esophagogastric cancer, specifically primary cancer samples (67%). Furthermore, there are two cancer types (Head and Neck Cancer and Mesothelioma) where the LOY rate in primary tumors differ by > 13% compared to metastatic samples. Interestingly, Gastrointestinal Stromal tumors showed an incidence of Y-chromosome loss in metastatic samples (32.8%) that was much higher than its

associated primaries (11.0%); a difference which can't be seen in any other cancer type. Overall, although there is a higher incidence rate of LOY seen in metastatic samples, no significant difference can be reported (**Fig. 10B**). This suggests that LOY is cancer type specific rather than affected by sample type. Finally, we examined whether varying LOY rates can be explained by the age of male participants. While we see a relative plateau of ~20% LOY in patients younger than 30 years, there is a subtle increase of LOY rates towards the age of 50. This is expected and in line with previous reports. Patient groups older than 50 years steadily show similar LOY rates, suggesting that chromosome Y loss cannot be fully explained by the current patient's age.

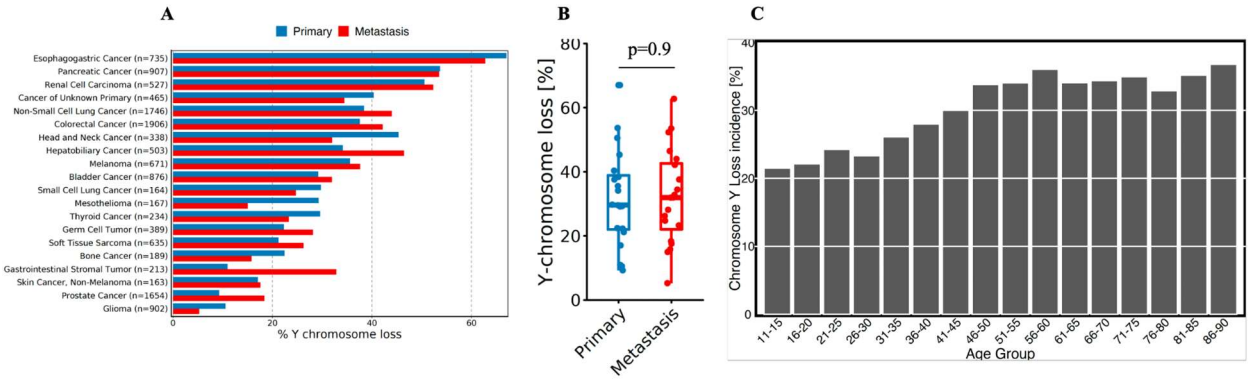


Figure 10. LOY rates across primary and metastatic samples
A) Cancer types listed according to the median LOY incidences and stratified by tissue type. B) Frequency of LOY across primary and metastatic tumor samples. C) LOY incidences (y-axis) grouped by various age groups starting from age 11. Age groups include 5 years each.

Association of ChrY loss with oncogenic alterations and pathway
We sought to identify associations between ChrY loss and somatic driver alterations. A multivariate logistic regression was performed. The binary alteration status of a gene is defined as 1 when the gene has any somatic SNV, indel, focal CNA or fusion event, and 0 otherwise. Only oncogenic somatic alterations defined by OncoKB were used in this analysis. Additionally, we excluded genes or pathways that had an overall alteration frequency less than 3%. P-values were adjusted for multiple hypothesis testing with false discovery rate (FDR) procedure.

In each cancer type, we tested the alteration status of each gene while controlling for sample type (primary or metastasis), fraction of genome altered (FGA) and MSI status (stable, in stable and indeterminate). In the pan-caner analysis setting, we tested the alteration status of each gene or a pathway while controlling for all the above confounders as well as cancer type.

Altogether, we identified 23 significant associations across 13 different cancer types (**Fig. 11A**). The most frequently associated gene with ChrY loss across several cancer types was *TP53* (bladder, hepatobiliary, colorectal, non-small cell lung, melanoma and esophagogastric cancers with an adjusted

p value < 0.05). Additionally, we identified chromosome X genes that have homologs on chromosome Y, *KDM5C* (renal cell carcinoma; adjusted p-value 1.78E-07) and KDM6A (bladder cancer; adjusted p-value 0.009), to be associated with ChrY loss. *FOXA1*, a known driver in prostate cancer onset and progression, was associated with ChrY loss in this disease type (adjusted p-value < 0.01). Genes such as *CDKN1A, CTNNB1, KRAS* etc. were enriched in samples with intact chromosome Y in specific cancer types with adjusted p-value < 0.05 (shown in red).

We repeated the analysis, this time additionally controlling for cancer type in a pan-cancer setting (**Fig. 11B**). We found three recurrent somatic alterations that differed in prevalence by Y loss status loss status (p-adjusted < 0.05): *TP53*, *MYC* and *KDM6A* were frequently altered in the ChrY loss group.

Next, we interrogated the alteration prevalence of ten canonical oncogenic signaling pathways with chrY loss in a pan-cancer setting (Sanchez-Vega et al., Cell 2018). Alterations in the p53 (adjusted p-value = 1.02E-47) and MYC (adjusted p-value = 0.0037) pathways were more frequently altered in ChrY loss than in intact Y tumors, whereas alterations in the β-catenin/Wnt (adjusted p-value = 0.0020) pathway was less frequently observed in ChY loss than in intact Y tumors **(Fig. 11C).**

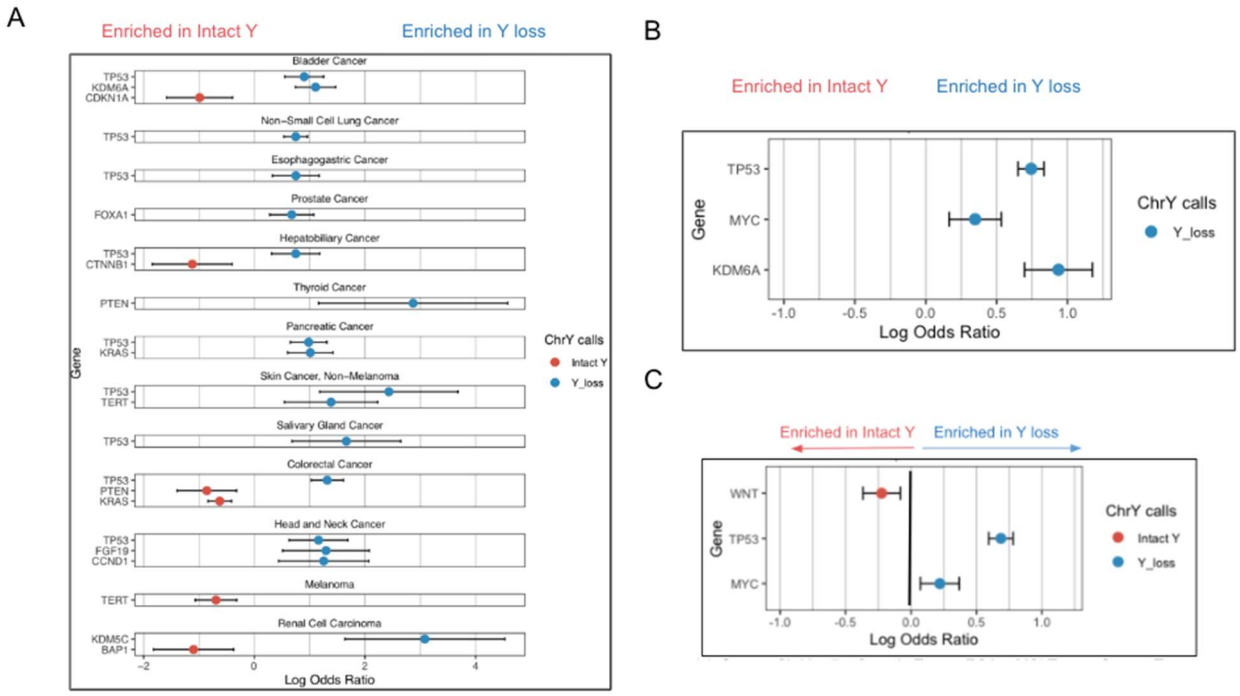


Figure 11. ChrY loss and oncogenic gene and pathway level alterations
A) ChrY loss - gene level associations in different cancer types. X axis indicates log odds ratios obtained from multivariate logistic regression models with 95% CI. Associations with adjusted p-value < 0.05 are highlighted (after adjusting for age for sample type, FGA, MSI status). Genes highlighted in blue are enriched in samples with ChrY loss and red are enriched in intact Y B) Chromosome Y loss - gene level associations in pan-cancer setting. X axis indicates log odds ratios obtained from multivariate logistic regression models with 95% CI.

Associations with adjusted p-value < 0.05 are highlighted (after adjusting for age for sample type, FGA, MSI status as well as cancer type). Genes highlighted in blue are enriched in samples with ChrY loss and red are enriched in intact Y C) Chromosome Y loss - pathway level associations in pan-cancer setting. X axis indicates log odds ratios obtained from multivariate logistic regression models with 95% CI. Associations with adjusted p-value < 0.05 are highlighted (after adjusting for age for sample type, FGA, MSI status as well as cancer type). Genes highlighted in blue are enriched in samples with ChrY loss and red are enriched in intact Y.

Association of ChrY loss with outcomes

To test whether ChrY loss was associated with outcomes, we performed survival analysis among males profiled with MSK-IMPACT. For survival analysis at pan-cancel level and across individual cancer types, we followed patients for overall survival OS, with follow-up starting at sequencing and ending at death with censorship occurring at last patient contact, with stratification by sample type status (primary and metastasis). To assess associations between ChrY loss and OS, we estimated each hazard ratio using Cox proportional-hazards regression model and p-values were compared via the Mantel-Cox log rank test.

We noticed ChrY loss to be associated with shorter OS in the pan cancer analysis (HR **1.36** , 95CI;1.29-1.44 , P < 2e-16, **Fig. 12A**). To account for influence by cancer type, we adjusted the model to account for cancer type, sample type, FGA and MSI status (adjusted HR;**1.22** , 95CI; 1.14-1.28, P = 4.84e-11), a similar trend was observed.

At an individual cancer type level, we found that ChrY loss was significantly associated with shorter OS in prostate cancer (**HR: 1.95**, 95% CI; 1.53 -2.51, P = 6.57e-08, **Fig. 12B**), pancreatic cancer (**HR 1.53**, 95% CI; 1.28-1.82, P = 2.89e-06, **Fig 12C.**), non-small cell lung cancer (HR **1.27**, 95CI; 1.12-1.44, P =0.0001, **Fig. 12D.**).

ChrY loss was associated with poorer outcomes even after we adjusted for sample type, FGA and MSI status in prostate cancer (**HR: 1.43**, 95% CI; 1.12-1.85, P =0.0043), pancreatic cancer (**HR 1.50**, 95% CI; 1.25-1.80, P =8.04e-06), non-small cell lung cancer (HR **1.19**, 95CI; 1.04-1.35, P = 0.009).

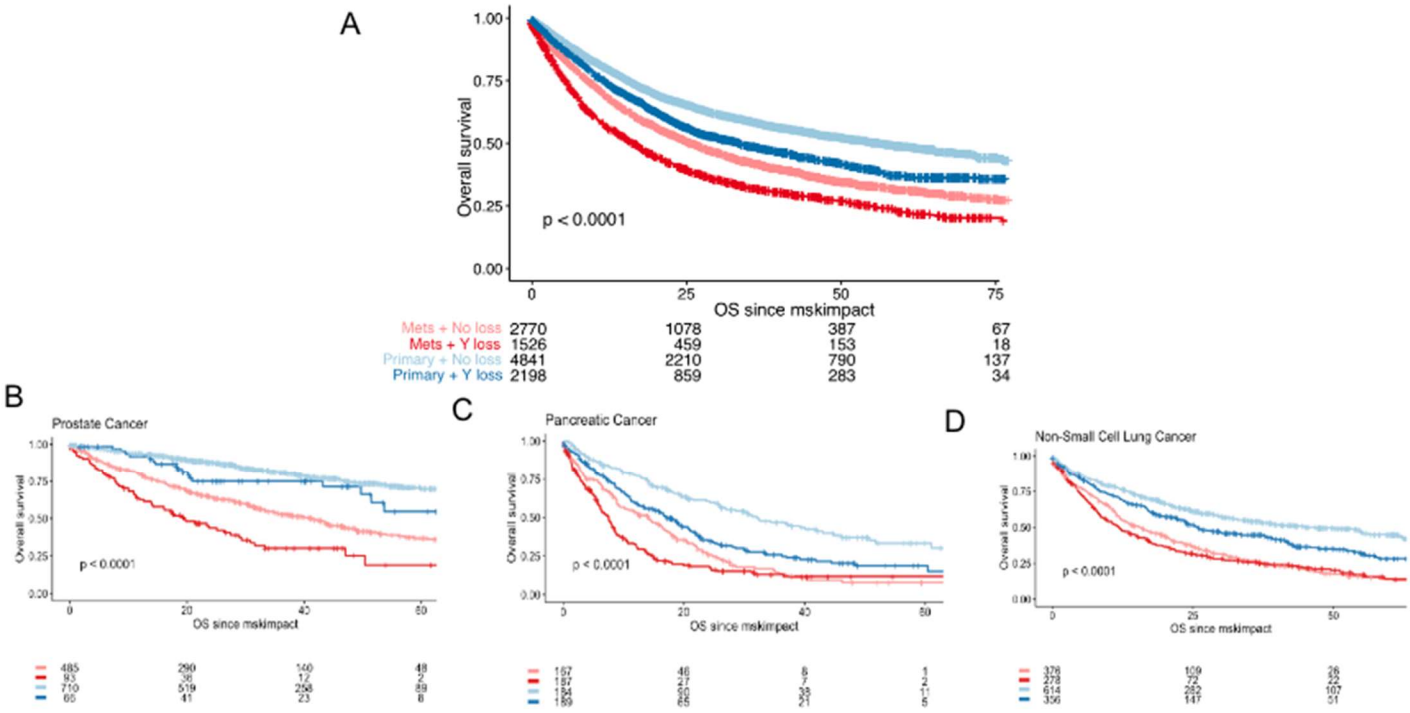


Figure 12. ChrY loss and outcomes

A) Kaplan Meir curves displaying overall survival status at pan cancer level. Samples with ChrY loss in primary and metastatic samples are shown in darker shades of blue and red and those with intact Y are shower in lighter shades of blue and red respectively. B) Kaplan Meir curves displaying overall survival status in prostate cancer C) Kaplan Meir curves displaying overall survival status in prostate cancer D) Kaplan Meir curves displaying overall survival status in non-small cell lung cancer.

SPECIFIC AIM 2: To perform genetic screening by CRISPR to identify ChrY genes that are of importance in the development of castration-resistant prostate cancer (CRPC) or resistance to androgen receptor (AR)–targeted therapies.

Major Activities

Genetic screening by CRISPR was conducted by the Kantoff group at MSK. Sequencing analysis of sgRNA and barcodes to identify target genes were conducted by the Kantoff and Schultz groups. As proposed in the SOW (1-24 months), we have completed the positive genetic screens for identifying genes regulating enzalutamide sensitivity.

Specific Objectives

The objectives proposed in the SOW were to: 1) establish barcoded cell line model systems; 2) design and construct ChrY CRISPR/Cas9 library; 3) optimize the ChrY CRISPR/Cas9 library in target cell lines; and 4) conduct positive selection screens with the ChrY CRISPR/Cas9 library to identify genes responsible for mCRPC development and antiandrogen resistance.

Significant Results or Key Outcomes

We have successfully generated the first, to our knowledge, ChrY-targeting CRISPR/Cas9 library. The pooled library is constructed to be used in a lentiviral system allowing high transduction efficiency. The CRISPR library (Figure 5A) contains 4 sgRNAs/gene and is targeting 45 protein coding genes (17 unique protein coding genes and 28 paralogous genes), 53 non-coding genes, and 188 pseudogenes. In addition, we have included 150 AASV1 (Adeno-Associated Virus Integration Site 1) controls, 150 negative controls, and 72 positive controls, resulting in a total of 1,519 sgRNAs.

LNCaP, LNCaP-Abl, and RWPE-1 cell lines expressing the ChY CRISPR/Cas9 library were created (Figure 5B). LNCaP-Abl and RWPE-1 cell lines were used as controls for enzalutamide-insensitive cell lines. Following enzalutamide treatment, the surviving population was analyzed to identify populations with sgRNA enrichment. Enrichment and depletion at the sgRNA and gene level were determined using the MagecK algorithm (version 0.5.6) (Li et al., 2014).

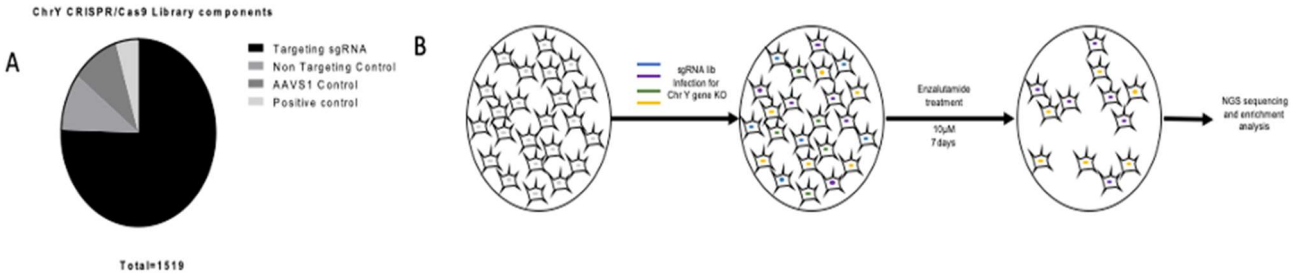


Figure 13. ChrY-targeting CRISPR/Cas9 library generation and experimental design.
A) Library composition. B) Protocol for generating cell lines expressing ChrY-targeting CRISPR/Cas9 library. AAVS1, Adeno-Associated Virus Integration Site 1; ChrY, Y chromosome; NGS, next generation sequencing; sgRNA, single guide RNA.

The gene targets obtained were TMSB4Y, DDX3Y, TTTY17C, TTTY6, GPR143P, and ZNF736P9Y. The protein coding gene TMSB4Y was excluded from further analysis, as it was not expressed in prostate cancer cell line systems. TTTY17C and TTTY6 genes were excluded from the analysis, as their expression profiles could not be discerned from overall long non-coding RNA (lncRNA) family expression.
The protein coding gene, DDX3Y, and pseudogenes, GPR143P and ZNF736P9Y sgRNA, were significantly enriched in enzalutamide-treated samples compared to control samples in castration conditions, indicating their loss conferred a growth advantage under enzalutamide treatment (**Fig. 14**).

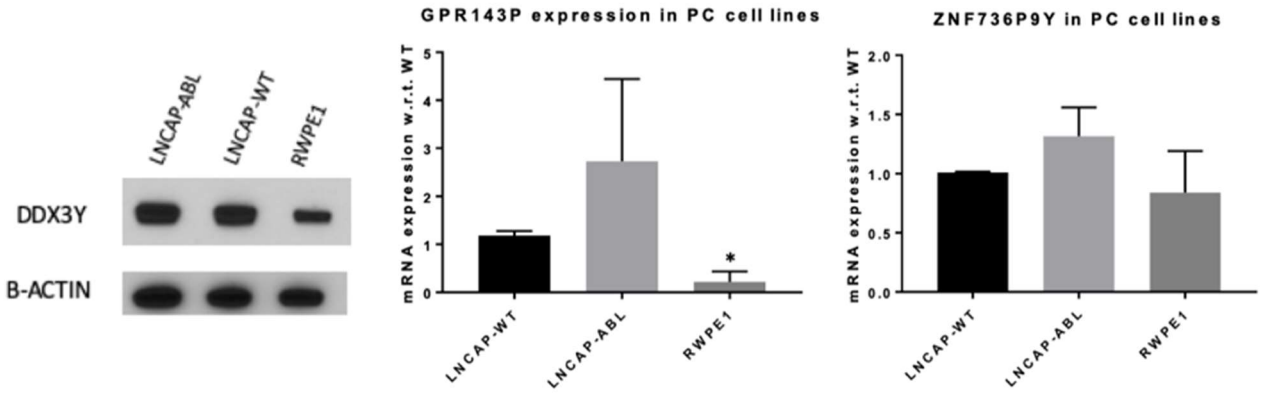


Figure 14. Target gene expression in prostate cancer (PC) cell line models.
A) DDX3Y baseline expression in LNCAP-ABL, LNCAP-WT was analyzed using immunoblotting. B) GPR143P expression in prostate cancer cell line models. *GPR143P baseline expression was significantly lower in RWPE-1 cell line model compared to LNCAP-WT cell line model ($P \leq 0.05$). C) ZNF736P9Y expression in prostate cancer cell line models.

Other Achievements
In addition to spectral karyotyping in target cell lines, we also conducted XY FISH (fluorescence in situ hybridization) on patient-derived prostate cancer organoids (Gao et al., 2014). All organoids were derived from metastatic prostate cancer tissue and cultured according to conditions outlined by Gao et al (2014). We screened 13 patient-derived prostate cancer organoid cultures. The FISH probe system marked the X chromosome as well as the euchromatic Yp (short arm) and heterochromatic Yq (long arm) regions of the ChrY with orange, green, and red fluorochromes, respectively (**Fig. 15**). Using this

system, we were able to identify prevalent ChrY loss in organoids, presenting the first report on the status of ChrY in patient-derived prostate cancer organoids.

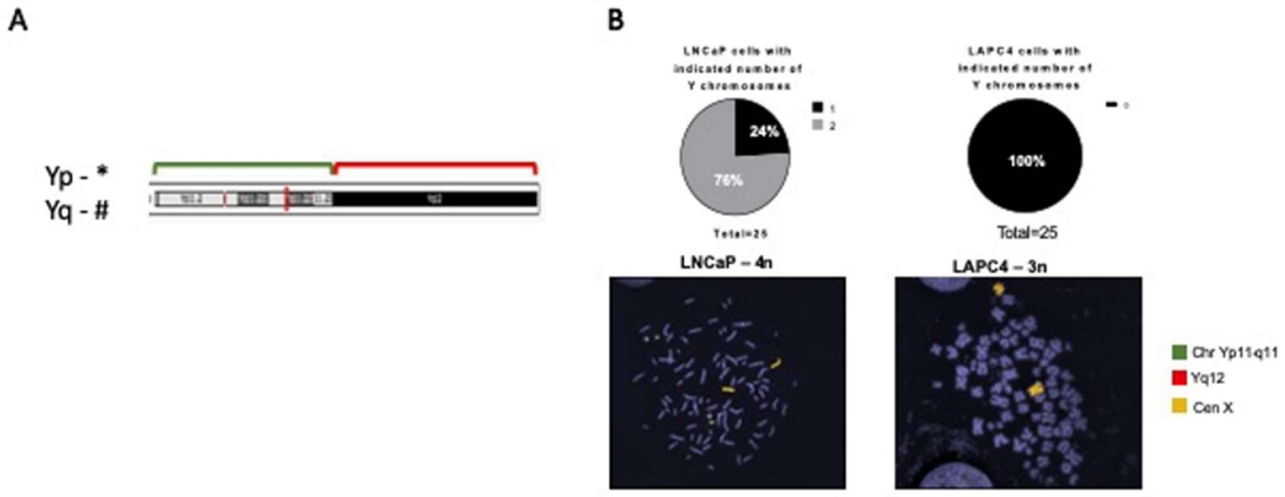


Figure 15. XY paint probe design and validation.
A) Chr Y probe design. B) FISH probe validated in prostate cancer cell lines. LNCaP-WT cells had heterogenous ChrY status. LAPC4 cells, which do not have ChrY, showed no signal with the probe. ChrY, Y chromosome; FISH, fluorescence in situ hybridization; WT, wildtype.

Six out of 13 (46.2%) organoid samples were heterogeneous in terms of ChrY status. We classified all organoid samples with >40% ChrY or Yq loss as having ChrY or Yq loss, respectively. Using this cutoff to classify chromosomal loss, approximately 30% of the tested organoid samples were classified with loss of the ChrY, indicating a potential role for the anomaly in disease progression. Interestingly, Yq loss was observed in 38.5% of our sample set, occurring at a more frequent rate than total ChrY loss (**Fig. 16**).

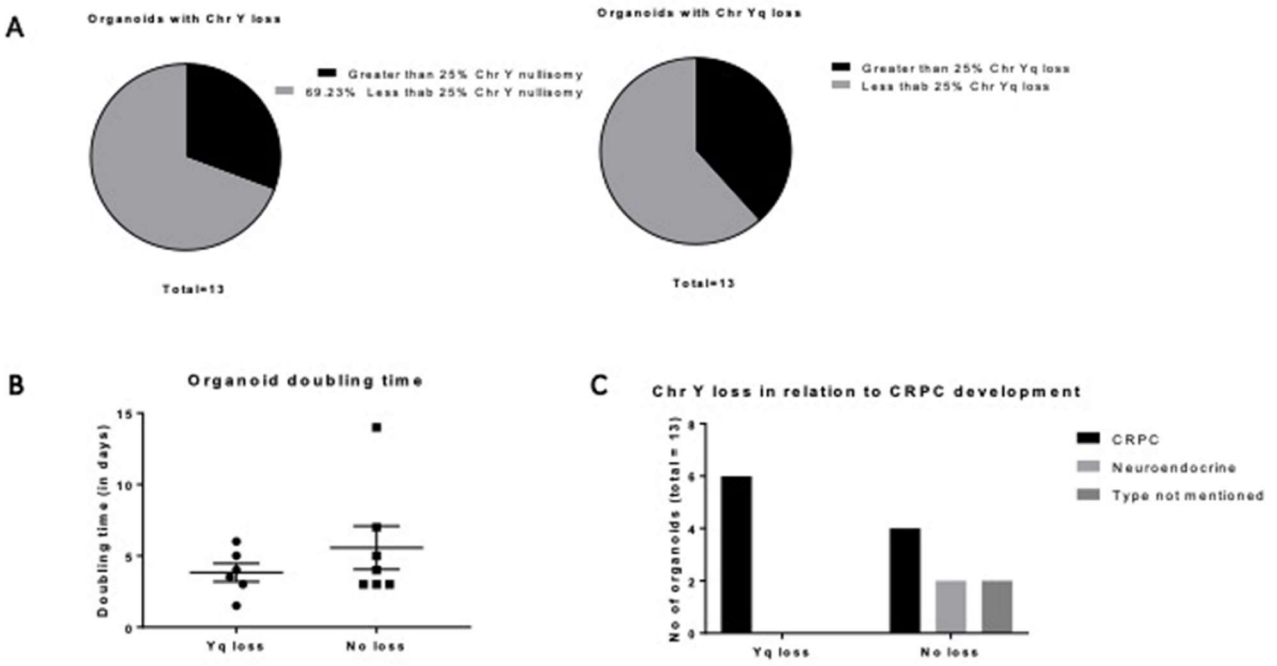
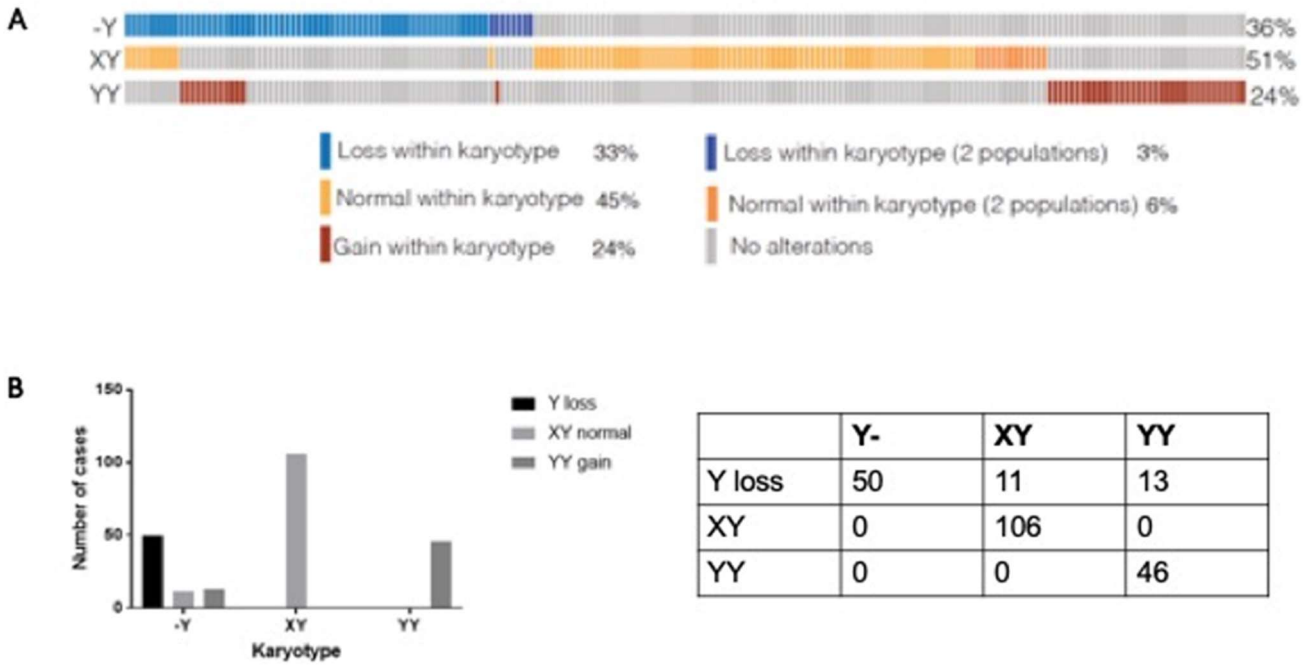


Figure 16. Characteristics of prostate cancer organoids. A) Percentages of prostate cancer organoids with ChrY and Yq loss. B) Doubling times of prostate cancer organoids. C) Breakdown of prostate cancer organoids and cancer tissue type used to derive the organoids.

We also queried Mitelman's database to analyze the status of ChrY loss in primary prostate adenocarcinoma samples. Interestingly, ChrY loss occurred at 36% frequency in primary prostate samples (**Fig. 17A**). Additionally, heterogeneity within prostate cancer samples in terms of ChrY status was observed only in samples with ChrY loss and not in patients without loss, indicating that ChrY loss could be an indicator for genomic instability in prostate cancer (**Fig. 17B**).



	Y-	XY	YY
Y loss	50	11	13
XY	0	106	0
YY	0	0	46

Figure 17. Y chromosome (ChrY) status.
A) ChrY status in primary prostate cancer samples. B) ChrY heterogeneity in primary prostate adenocarcinoma samples. Chi-square test reporting significance.

SPECIFIC AIM 3:
To characterize the functional significance of genes involved in resistance in cell culture and animal models. We will confirm the functional importance of KDM5D and UTY in progression to CRPC. The clinical significance of these genes will be corroborated with an evaluation of the ChrY landscape in prostate cancer specimens.
Major Activities
As per the timeline in the SOW, work on Specific Aim 3 began in month 24, following identification of candidate genes from Specific Aim 2.

Specific Objectives
The specific objectives described in the SOW were to: 1) perform functional validation of candidate genes including KDM5D and UTY in cell lines models; 2) identify pathways/mechanisms that a specific gene is involved in leading to the observed phenotype by RNA-seq, CHIP-seq, or phosphor-kinase screening; 3) generate mouse xenografts using stable cell lines with inducible knockdown or overexpression of candidate genes and treat them with drug or vehicle to measure tumors followed by molecular characterization, and 4) assess the clinical impact of KDM5D, UTY, and candidate gene expression on prostate cancer outcomes in clinical datasets from the following patient cohorts: Stand Up To Cancer/Prostate Cancer Foundation (SUC2/PCF), TCGA, Harvard Prostate Tumor (Health Professionals Follow-Up Study [HPFS] and Physician's Health Study [PHS]) study cohorts.

Significant Results or Key Outcomes

Gene targets DDX3Y, GPR143P, and ZNF736P9Y were selected for functional analysis studies. Small interfering RNA (siRNA)-mediated knockdown in LNCaP cells resulted in ~50% silencing (**Fig. 18**). However, this silencing did not result in any change in short-term growth.

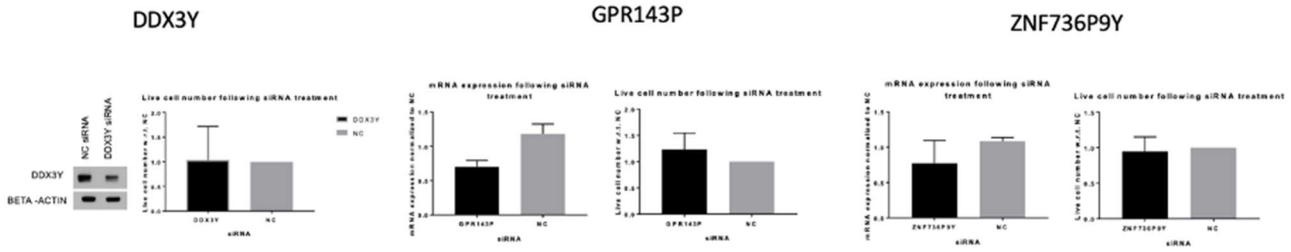


Figure 18. DDX3Y, GPR143P, and ZNF736P9Y were identified from the CRISPR screen.
Silencing expression using siRNA did not result in decreased cell growth at 48 hours.

Short hairpin RNA (shRNA)-mediated knockdown of DDX3Y was achieved using two constructs, DD6 and DD7. DD6 was used for functional studies, as it resulted in greater knockdown. DDX3Y knockdown rescued cells from enzalutamide-mediated growth inhibition when stimulated with dihydrotestosterone (DHT; **Fig. 19A**). DDX3Y knockdown also resulted in increased basal expression of PSA in LNCaP cells compared to control (**Fig. 19B**). DHT stimulation coupled with enzalutamide treatment did not abolish PSA expression in DDX3Y knockdown models compared to control, indicating a heightened androgen receptor (AR) response (**Fig. 19C**).

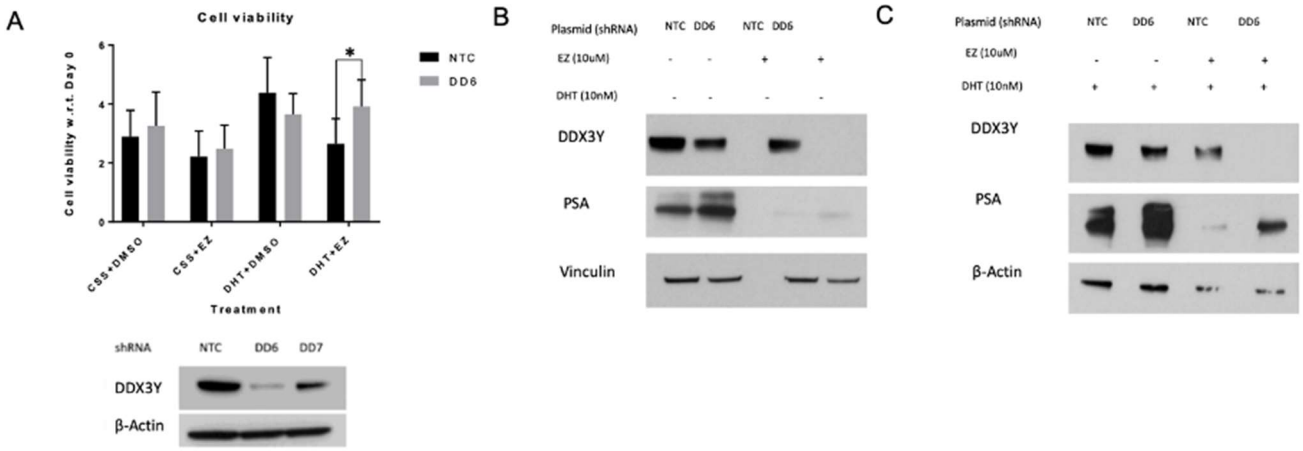


Figure 19. *DDX3Y* knockdown augmented enzalutamide-resistant phenotype in cell line models.
A) Knockdown of *DDX3Y* conferred resistance from enzalutamide-mediated growth inhibition. B) DDX3Y knockdown resulted in increased expression of prostate-specific antigen (PSA). C) dihydrotestosterone (DHT) stimulation (10 nM) coupled with enzalutamide (EZ; 10 μM) treatment did not abolish PSA expression in *DDX3Y* knockdown models compared to control, indicating a heightened androgen receptor (AR) response.

Other Achievements

Nothing to Report.

- **What opportunities for training and professional development has the project provided?**

  Nothing to Report.

- **How were the results disseminated to communities of interest?**

  Nothing to Report.

- **What do you plan to do during the next reporting period to accomplish the goals?**

  Nothing to Report.

REFERENCES

1. Dumanski JP, Lambert JC, Rasi C, et al. Mosaic Loss of Chromosome Y in Blood Is Associated with Alzheimer Disease. *Am J Hum Genet.* 2016;98(6):1208-1219. PMCID: PMC4908225.
2. Forsberg LA, Rasi C, Malmqvist N, et al. Mosaic loss of chromosome Y in peripheral blood is associated with shorter survival and higher risk of cancer. *Nat Genet*. 2014;46(6):624-628. PMCID: PMC5536222.
3. Noveski P, Madjunkova S, Stefanovska ES, et al. Loss of Y chromosome in peripheral blood of colorectal and prostate cancer patients. *PLoS One*. 2016;11(1):e0146264. PMCID: PMC4706411.
4. Komura K, Jeong SH, Hinohara K, et al. Resistance to docetaxel in prostate cancer is associated with androgen receptor activation and loss of KDM5D expression. *Proc Natl Acad Sci U S A*. 2016;113(22):6259-6264. PMCID: PMC4896695.
5. Komura K, Yoshikawa Y, Shimamura T, et al. ATR inhibition controls aggressive prostate tumors deficient in Y-linked histone demethylase KDM5D. *J Clin Invest*. 2018;128(7):2979-2995. PMCID: PMC6025984.
6. Shen R, Seshan VE. FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucleic Acids Res*. 2016;44(16):e131. PMCID: PMC5027494.
7. Sanchez-Vega F, Mina M, Armenia J, et al. Oncogenic signaling pathways in The Cancer Genome Atlas. *Cell*. 2018;173(2):321-337.e10. PMCID: PMC6070353.

8. Li W, Xu H, Xiao T, et al. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol*. 2014;15(12):554. PMCID: PMC4290824.

9. Gao D, Vela I, Sboner A, et al. Organoid cultures derived from patients with advanced prostate cancer. *Cell*. 2014;159(1):176-187. PMCID: PMC4237931.

4. IMPACT

- **What was the impact on the development of the principal discipline(s) of the project?**

  Our work presented the first evidence of ChrY genes regulating enzalutamide sensitivity in prostate cancer cell line systems. We also were the first to quantify and evaluate ChrY loss in patient-derived prostate cancer organoids and primary prostate cancer tissue. This work has uncovered novel insights into the role of ChrY loss in prostate cancer and as a potential surrogate marker for chromosomal instability.

- **What was the impact on other disciplines?**

  Our ChrY-targeting CRISPR/Cas9 library can be utilized to screen for multiple phenotypes in disease models other than prostate cancer. The library is a versatile tool used to investigate the role of ChrY genes in male responses in diseases that affect both sexes. The FISH probes developed for XY paint may also be used to quantify X or Y chromosome status in disease/non-disease models, not limited to cancer. Overall, our developed resources and protocols can be easily adapted to investigate the role of ChrY loss in any disease model.

- **What was the impact on technology transfer?**

  Nothing to Report.

- **What was the impact on society beyond science and technology?**

  Nothing to Report.

5. CHANGES/PROBLEMS

- **Changes in approach and reasons for change**

  The initial narrative proposed RWPE-1, LNCaP, LAPC4, LNCaP-Abl, and VCaP as cell line systems. Our efforts to optimize lentiviral transduction protocols in VCaP and LAPC4 cells have resulted in sub-optimal selection. The low transfection efficiency of these cells resulted in models that were unsuitable to conduct shRNA and CRISPR screens. We have therefore excluded VCaP and LAPC4 cells from our analysis. This will not affect the power of our analysis, as cell lines, RWPE-1, LNCaP, and LNCaP-Abl, are sufficient and present enough ChrY heterogeneity for our purposes.

The target generation and functional analysis were delayed due to COVID-related halts in sequencing analysis. We were therefore unable to conduct studies in xenograft and other *in vivo* models.

Due to delays in the experimental work and a PI leaving MSK, we shifted some of the work to more computational analyses during a NCE period, which resulted in significantly expanded analyses on a larger sample set and across multiple cancer types.

- **Actual or anticipated problems or delays and actions or plans to resolve them**

  We experienced a delay of 3 months because of a mandated shutdown of facilities due to the COVID-19 pandemic, related lockdowns, and decreased personnel. This has resulted in a lack of sufficient time to generate and conduct studies in *in vivo* models. However, lack of xenograft studies (activities related to Major Task 5) does not compromise the relevance of our functional analysis, as validation studies are conducted in cell line models and publicly available clinical cohorts.

- **Changes that had a significant impact on expenditures**

  Nothing to Report.

- **Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents**

  Nothing to Report.

- **Significant changes in use or care of human subjects**

  Nothing to Report.

- **Significant changes in use or care of vertebrate animals.**

  Nothing to Report.

- **Significant changes in use of biohazards and/or select agents**

  Nothing to Report.

6. PRODUCTS

- **Publications, conference papers, and presentations**
  - **Journal publications.**

    Nothing to Report.

  - **Books or other non-periodical, one-time publications.**
    Nothing to Report.

  - **Other publications, conference papers, and presentations.**

    Nothing to Report.

- **Website(s) or other Internet site(s)**

  Nothing to Report.

- **Technologies or techniques**

  Nothing to Report.

- **Inventions, patent applications, and/or licenses**

  Nothing to Report.

- **Other Products**.

  Nothing to Report.

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

- **What individuals have worked on the project?**

  Subhiksha Nandakumar, Bioinformatic Engineer, no change

Name:	Philip Kantoff
Project Role:	Former Initiating Principal Investigator
Researcher Identifier (e.g. ORCID ID):	0000-0001-7275-0597
Nearest person month worked:	0.6
Contribution to Project:	Dr. Kantoff served as the Initiating Principal Investigator (PI) until June 2021, at which time he departed from MSK. Dr. Kantoff was replaced with Dr. Schultz, who served as the PI for these Collaborative Awards.

Funding Support:	NIH, DoD

Name:	Nikolaus Schultz
Project Role:	Initiating Principal Investigator
Researcher Identifier (e.g. ORCID ID):	N/A
Nearest person month worked:	1.2 months
Contribution to Project:	Dr. Schultz oversaw all efforts related to this proposal in close collaboration with Drs. Kantoff and Van Allen. Dr. Schultz replaced Dr. Kantoff as the PI for these Collaborative Awards due to Dr. Kantoff's departure from MSK.
Funding Support:	NIH, DOD, Cholangiocarcinoma Foundation, Prostate Cancer Foundation, MSK, The Gray Foundation, Functional Genomics Initiative

Name:	Eliezer Van Allen
Project Role:	Site Principal Investigator
Researcher Identifier (e.g. ORCID ID):	vanallen81
Nearest person month worked:	0.36 months
Contribution to Project:	Dr. Van Allen oversaw all efforts related to this proposal in close collaboration with Dr. Schultz through June 2021.
Funding Support:	NIH, DOD, Damon Runyon Cancer Research Foundation, Novartis, Prostate Cancer Foundation, Bristol Myers Squibb, Movember Foundation, Brown Performance Group, Leidos Biomedical Research, Inc, Starr Consortium

Name:	Travis Gerke
Project Role:	Site Principal Investigator
Researcher Identifier (e.g. ORCID ID):	0000-0002-9500-8907
Nearest person month worked:	0
Contribution to Project:	Dr. Gerke is the Moffitt Cancer Center (MCC) Principal Investigator (site location 1). In collaboration with Drs. Kantoff and Schultz, he was responsible for the analysis of gene expression data from the

	HPFS and the PHS cohorts and broadly assisting with the epidemiologic and statistical interpretation of findings. Effective as of 5/16/2020, this subaward was terminated (revised subaward period of performance: 6/15/2018 – 5/16/2020; effective date of amendment number 3: 6/15/19); Dr. Gerke did not contribute to the project during this period and did not continue as a site PI/key personnel.
Funding Support:	H. Lee Moffitt Cancer Center and Research Institute, Inc.

Name:	Sai Harisha Rajanala
Project Role:	Research Fellow
Researcher Identifier (e.g. ORCID ID):	0000-0002-7096-3756
Nearest person month worked:	3
Contribution to Project:	Dr. Rajanala has replaced Yuki Yoshikawa, MD. She will conduct genetics screens to identify regulators of antiandrogen therapy sensitivity using generated CRISPR/Cas9 screens. Dr. Rajanala will also conduct functional validation of candidate genes following positive genetic screens.
Funding Support:	Institutional

Name:	Rahim Hirani
Project Role:	Research Technician
Researcher Identifier (e.g. ORCID ID):	0000-0002-9304-9916
Nearest person month worked:	6
Contribution to Project:	Mr. Hirani has replaced Mohammad Atiq, MD. He will assist Dr. Rajanala with the functional validation of candidate genes.
Funding Support:	DOD

Name:	Bastien Nguyen
Project Role:	Research Fellow
Researcher Identifier (e.g. ORCID ID):	N/A

Nearest person month worked:	4
Contribution to Project:	Dr. Nguyen will collaborate with Dr. Van Allen's (DFCI) laboratory on sequence analysis.
Funding Support:	DOD

Name:	Anisha Luthra
Project Role:	Bioinformatics Software Engineer
Researcher Identifier (e.g. ORCID ID):	N/A
Nearest person month worked:	3
Contribution to Project:	Ms. Luthra will assist Dr. Schultz on bioinformatic software engineering.
Funding Support:	DOD

Name:	Sabrina Camp
Project Role:	Computational Biologist
Researcher Identifier (e.g. ORCID ID):	N/A
Nearest person month worked:	2
Contribution to Project:	Ms. Camp has replaced Dr. Eric Kofman as the Computational Biologist assisting Dr. Van Allen. Ms. Camp's focus is on the analysis, method development, and application pertaining to identifying genomic features that correlate with mutational signature analysis of prostate cancers.
Funding Support:	DOD

8. SPECIAL REPORTING REQUIREMENTS

- **QUAD CHARTS**

  Not applicable.