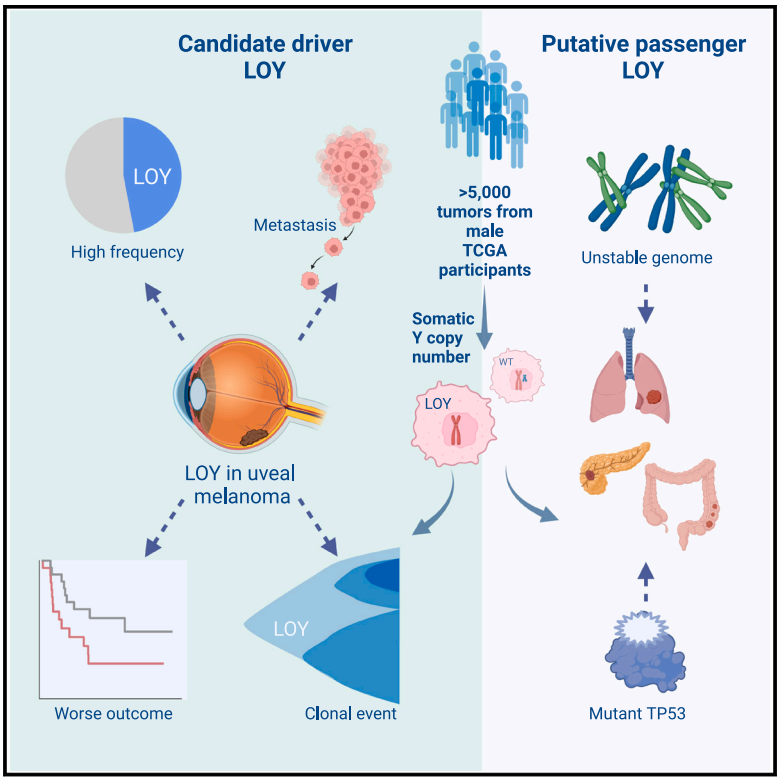
Cell

Loss of chromosome Y in primary tumors

Graphical abstract



Highlights

- Somatic Y copy number variations are identified in >5,000 TCGA male tumors
- LOY is a frequent event, and frequency varies by tumor type
- Evidence that LOY can be either a passenger or driver event in cancer
- LOY is an early and clonal candidate driver in uveal melanoma

Authors

Meifang Qi, Jiali Pang, Irene Mitsiades, Andrew A. Lane, Esther Rheinbay

Correspondence

erheinbay@mgh.harvard.edu

In brief

Analysis of the loss of chromosome Y across 5,000 primary tumors reveals that it can be an oncogenic passenger or driver event depending upon cancer type; for uveal melanoma, it is associated with age and survival and is an independent predictor of poor outcome.

Qi et al., 2023, Cell *186*, 3125–3136
July 6, 2023
https://doi.org/10.1016/j.cell.2023.06.006

CellPress

Cell

CellPress

Resource

Loss of chromosome Y in primary tumors

Meifang Qi,[1,2,3] Jiali Pang,[1,4,7] Irene Mitsiades,[1,2,8] Andrew A. Lane,[2,3,5] and Esther Rheinbay[1,2,3,6,9,*]
[1]Massachusetts General Hospital Center for Cancer Research, Charlestown, MA 02129, USA
[2]Harvard Medical School, Boston, MA 02115, USA
[3]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[4]Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA
[5]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02215, USA
[6]Massachusetts General Hospital Department of Pathology, Boston, MA 02114, USA
[7]Present address: Shanghai Artificial Intelligence Laboratory, Shanghai, China
[8]Present address: College of Arts and Sciences, Boston University, Boston, MA 02215, USA
[9]Lead contact
*Correspondence: erheinbay@mgh.harvard.edu
https://doi.org/10.1016/j.cell.2023.06.006

SUMMARY

Certain cancer types afflict female and male patients disproportionately. The reasons include differences in male/female physiology, effect of sex hormones, risk behavior, environmental exposures, and genetics of the sex chromosomes X and Y. Loss of Y (LOY) is common in peripheral blood cells in aging men, and this phenomenon is associated with several diseases. However, the frequency and role of LOY in tumors is little understood. Here, we present a comprehensive catalog of LOY in >5,000 primary tumors from male patients in the TCGA. We show that LOY rates vary by tumor type and provide evidence for LOY being either a passenger or driver event depending on context. LOY in uveal melanoma specifically is associated with age and survival and is an independent predictor of poor outcome. LOY creates common dependencies on *DDX3X* and *EIF1AX* in male cell lines, suggesting that LOY generates unique vulnerabilities that could be therapeutically exploited.

INTRODUCTION

Gender and biological sex have been implicated in cancer incidence, mortality, and response to therapy. These sex differences are caused by different physiology, the effect of sex hormones, environmental exposure, and intrinsic genetic sex differences caused by the dichotomous sex chromosomes. Sex chromosome loss is the most frequent somatic change in peripheral blood lymphocytes of elderly individuals[1–6] and sporadically occurs in normal tissues.[7] Specifically, loss of the Y chromosome (LOY) in aging men has been linked to increased mortality, incidence of hematologic and solid cancers, and Alzheimer's disease.[8] Along with other copy number alterations, LOY has also been observed in some tumor types, including papillary renal cancer,[9,10] esophageal cancer,[11] and advanced prostate cancer.[12,13] The Y chromosome is comprised of a male-specific region (MSY), unique to this chromosome, and the pseudoautosomal regions (PARs), short genomic stretches near the ends of X and Y that undergo homologous recombination between the two sex chromosomes during meiosis. The MSY contains several ubiquitously expressed genes,[14] some of which are evolutionary conserved ancestral and potentially functional homologs of X inactivation escape genes.[15] Furthermore, somatic mutations in escape from X-inactivation tumor suppressor (EXITS) genes are enriched in tumors that also have loss of the second sex chromosome (X or Y).[16] Little is known about the role of specific Y-linked genes in cancer, yet anecdotal evidence suggests relevance to disease biology. For example, aberrant expression of the Y gene *TSPY1* in females with dysgenetic gonads is associated with gonadoblastoma.[17] In prostate tumors, expression of the Y-linked histone demethylase *KDM5D* is associated with response to chemotherapy,[12] and *KDM5D* loss through LOY increases viability of renal cancer cell lines.[18]

Despite this important evidence for a role of Y in cancer, whole-exome sequencing (WES) and whole-genome sequencing (WGS) cancer studies have nearly universally neglected analysis of the Y chromosome or deliberately excluded it, even in studies focused on cancer sex differences.[19] Reasons include the very small number of genes on Y that are relevant outside of male spermatogenesis, technical challenges caused by its haploid ground state, homology with X regions, and repetitive sequence from gene expansions. Prior studies on chromosome Y loss have focused on assessing the fraction of LOY cells in normal blood from single nucleotide polymorphism (SNP) array,[2,3] evaluated loss of gene expression from Y,[20] or been limited to fewer tumor types.[16,21] However, no comprehensive study of LOY in primary tumor tissue has been performed to date.

Our lack of knowledge around the Y chromosome in cancer creates missed therapeutic opportunities: LOY can expose

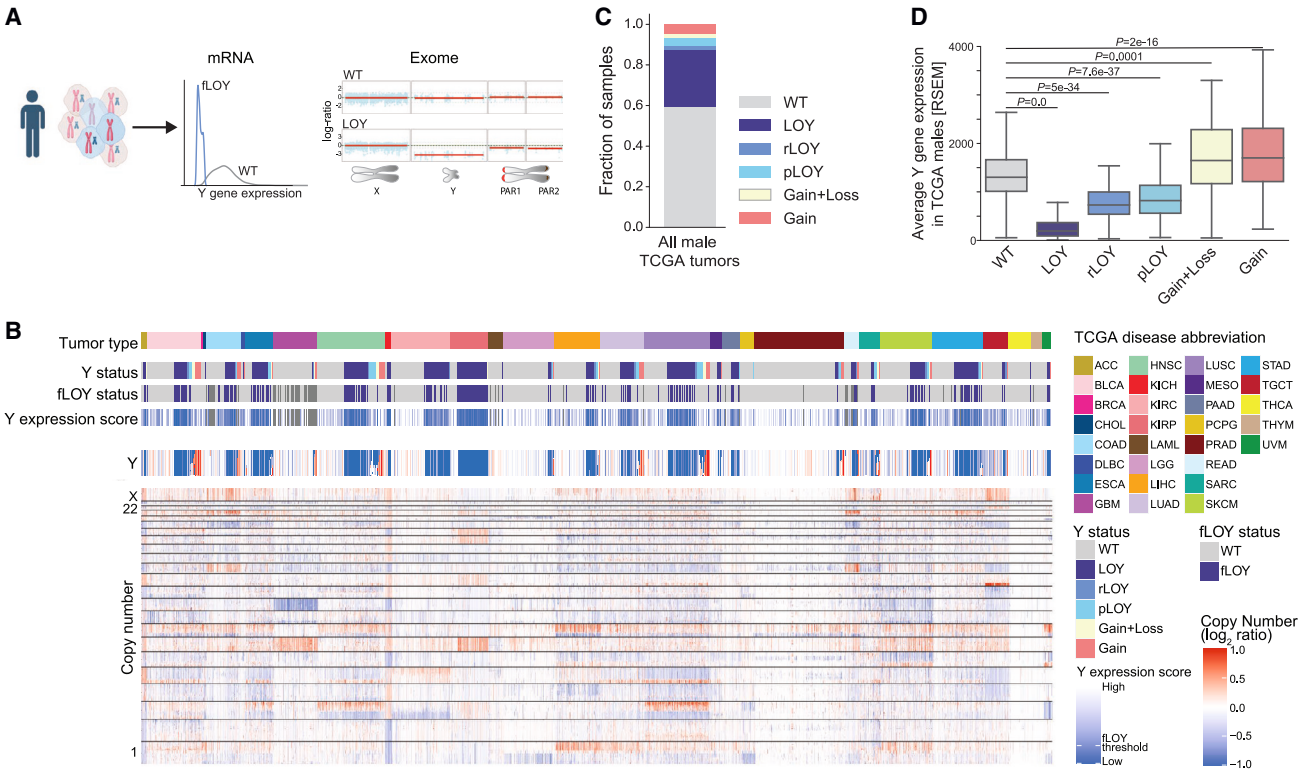


Figure 1. LOY is a frequent somatic event in primary cancers

(A) Schematic of methods to detect LOY in tumors. LOY can be detected from tumor gene expression profiles (mRNA-Seq or microarray) or exome (WES). fLOY, "functional" LOY (expression-based).

(B) Copy number profiles of 5,014 male tumors from TCGA with the Y chromosome copy number shown disproportionally large. Y gene expression score, expression-based loss calls (fLOY), and Y status inferred from exome copy number are shown. Tumor types are labeled with TCGA tumor codes.

(C) Fraction of tumors with indicated Y chromosome alterations among male TCGA tumors. pLOY, partial LOY; rLOY, relative LOY (total Y copy number ≥ 1); LOY, total Y copy number equals zero. Details are described in STAR Methods.

(D) Average Y gene expression for each type of Y alteration corresponds to inferred copy state. p values calculated with the Mann-Whitney U (MWU) test.

specific cellular vulnerabilities caused by loss of genes without "backup copies," loss of those with homologs on X,[16] and loss of heterozygosity and potential haploinsufficiency of genes in the PARs.

Here, we present a comprehensive analysis of the Y chromosome across >5,000 male tumors from The Cancer Genome Atlas (TCGA), suggesting both driver and passenger roles for LOY in different cancer types.

RESULTS

LOY is a frequent somatic event in primary cancers

Many commonly used somatic copy number calling methods do not provide faithful copy number estimates for the Y chromosome. We therefore adapted an established copy number analysis method[22] to analyze X, Y, and the two pseudoautosomal regions (PAR1 and PAR2) separately (Figure 1A; see STAR methods for details) in 5,014 male and 5,394 female primary tumors using WES from TCGA (Table S1). Additionally, we derived a method based on Y gene expression to measure functional LOY (fLOY) from RNA-seq as an alternative method for samples without WES or normal control data, confirming that gene expression can reliably detect LOY[20] (Figures 1A and S1A–S1D; see STAR Methods for details). We observed a range of Y copy events, most frequently complete LOY (28%). Relative LOY in the presence of additional copies (rLOY) also occurred at 2% frequency (Figures 1B and 1C; STAR Methods). Overall, 1,504 of 5,014 male tumors (30%) harbored either complete or relative LOY. As expected, LOY calls from exome were accompanied by loss of expression of Y-linked genes[14] (Figures 1B and 1D) and a median reduction of PAR gene expression by 32% compared to wild-type (WT) Y samples. Indeed, fLOY from TCGA tumor RNA-seq alone gave nearly identical results, with the most differences in lower-purity tumors with considerable normal cell content (Figures 1B, S1E, and S1F; Table S1A). For the subset of 354 samples with WGS available, LOY calls from WES and WGS were consistent in 94% of cases, with incongruencies attributable to low purity, read coverage, and potential somatic mosaic LOY (mLOY) in the normal samples (Figure S1G). In contrast to LOY, loss of the entire X chromosome (LOX) in females was always relative in this cohort, consistent with essentiality of X for cellular survival (Table S1B). LOX in females was accompanied by loss of *XIST* expression, which induces X inactivation when more than one X chromosome is present, and a

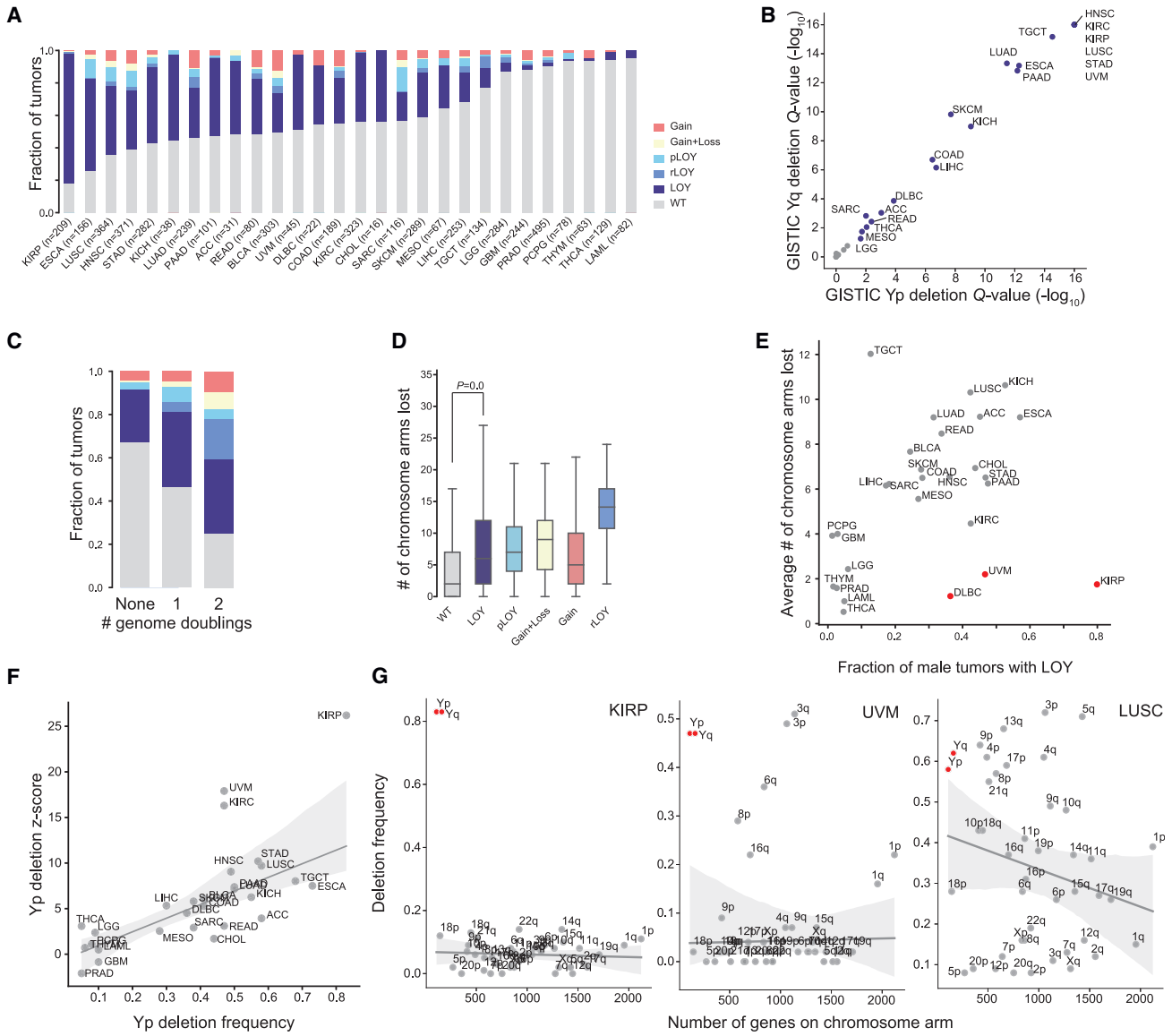


Figure 2. LOY rates across tumor types and association with aneuploidy

(A) Fraction of different Y chromosome alterations for each tumor type.

(B) GISTIC Q values for Yp and Yq chromosome loss confirm recurrent losses in many tumor types. Strong correlation between Yp and Yq further suggests that Y is typically lost as whole chromosome.

(C) Fraction of Y alterations in tumors with zero, one, or two genome doublings.

(D) Number of autosomal chromosome arms lost in tumors with WT, LOY, or other Y alterations. p value calculated with the MWU test.

(E) Correlation between total arm losses (autosomes only) and LOY. Red dots indicate tumor types with high LOY rate and low overall arm-loss rate.

(F) Association between significance of Yp loss under a background model of autosomal alteration frequencies[24] (scores calculated by GISTIC) and Yp loss frequencies for each tumor type.

(G) Chromosome arm loss frequencies for tumor types where LOY is likely selective (KIRP, UVM) or occurs in a background of genomic instability with many chromosomal arm losses (LUSC).

26% decrease of PAR gene expression (Figure S1H). In total, 757 of 5,394 (14%) female tumors harbored LOX.

LOY frequencies across tumor types

Like other patterns of chromosomal gains and losses, frequencies of different Y chromosome events varied by cancer type (Figure 2A; Table S1C). Complete LOY was most frequent in renal papillary cancer (KIRP; 80%) and esophageal cancer (ESCA; 57%) and as low as <2% in pheochromocytoma and paraganglioma (PCPG) and thymoma (THYM), and these fractions were consistent with our fLOY calls (Figures S2A and S2B), frequencies in the independent International Cancer Genome

Consortium (ICGC) subset from the Pan-Cancer Analysis of Whole Genomes (PCAWG) dataset (Figures S2C and S2D; Table S1D), and previously published results based on cytogenetics or WGS.[9–11,21,23] Statistical evaluation of arm-level losses with Genomic Identification of Significant Targets In Cancer (GISTIC; see STAR Methods for details) confirmed significant concomitant loss of Yq and Yp in many tumor types (Figure 2B; Table S2A). In contrast, LOX was most frequent in chromophobe kidney cancer (KICH) (56%) and uveal melanoma (UVM; 43%) but nearly absent in thyroid cancer (THCA; 1%) and THYM (0%) (Figure S2E; Table S1C). In general, LOY occurred much more frequently than LOX, with major differences in KIRP, suggesting a specific selective advantage of LOY in this tumor type (Figure S2F). LOY and LOX rates were nearly equally frequent (>20%) in KICH, UVM, mesothelioma (MESO), and cutaneous melanoma (SKCM; Figure S2F), possibly because loss of a shared PAR or homologous gene on X and Y might drive selection of tumors with sex chromosome loss.

LOY is common in aneuploid tumors

As a small and gene-poor chromosome, Y has a relatively high chance of being lost from cells "by chance,"[24,25] 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.[3] To answer whether somatic LOY is correlated with genomic instability in primary tumors, we compared LOY frequencies with genome doubling events and aneuploidy, measured as arm-level losses of only the autosomes.[26] Tumors that had undergone genome doublings were more likely to harbor various Y copy changes (Figure 2C). LOY tumors also had significantly more arm-level losses (median 6, IQR 2–12) compared to WT tumors (median 2, IQR 0–7; Mann-Whitney U test [MWU] p = 0.0; Figure 2D). As expected, overall arm loss frequencies were highest in tumors with rLOY as a product of genome doubling (Figures 2C and 2D). On a tumor-type level, LOY frequencies were correlated with the mean number of autosomal chromosome arms lost (Pearson's $r = 0.33$; p = 0.08), with UVM, KIRP, and diffuse large B-cell lymphoma (DLBC) having comparatively high fractions of LOY tumors (Figure 2E). Under a background model of overall frequency of arm level alterations (gain and loss) adjusted by number of genes (as a proxy for size and "importance") on each chromosome arm,[24] Yp and Yq loss frequencies remained outliers in KIRP, UVM, and kidney renal clear cell carcinoma (KIRC) (Figures 2F and 2G), suggesting that at least in these tumor types, LOY is not due to size-related random loss and lack of selection.

Inactivation of the tumor suppressor *TP53* causes genomic instability and numerous DNA copy gains and losses.[27,28] Indeed, we found that tumor types where *TP53* damaging mutations are common were enriched in LOY tumors across the TCGA Pan-Cancer cohort (47.7% of LOY tumors compared to 25.7%; Fisher's Exact p = 2.23×10^{-46}; Figure 3A) and confirmed in PCAWG (Figure S3A). LOY was enriched in *TP53*-mutant tumors in several cancer types, with highly aneuploid tumors (colon adenocarcinoma [COAD], ESCA; Figures 2 and 3B) showing significant overlap of the two alterations. Importantly, several tumor types had high LOY fractions in the absence of *TP53* mutations (UVM, DLBC, KIRP; Figures 3B and 3C). These data provide further evidence that LOY is not a consequence of genomic instability in these diseases.

Cancer cell fraction of LOY

The cancer cell fraction (CCF) of a tumor containing a somatic event can suggest whether this event is present in all (clonal) or only a subset of cancer cells (subclonal). Comparing CCF distributions for LOY across tumor types, we found that LOY is present in nearly all cancer cells in UVM, KICH, adrenocortical carcinoma (ACC), and brain lower grade glioma (LGG) (median CCFs 0.9, 0.85, 0.83, and 0.82, respectively; Figure 3D). In KICH and ACC, LOY occurs in a background of many chromosome losses (Figure 2E). In contrast, many other tumor types had wide distributions of LOY CCFs, especially those with prevalent *TP53* mutations, indicating that LOY is subclonal (Figure 3D). *TP53* mutations were nearly universally clonal (median CCF of 1 in LOY tumors)[29] with much higher CCFs than LOY (Figure 3D), suggesting that LOY generally follows *TP53*-inactivation-induced genomic instability in many tumor types.

Association of LOY with point mutation drivers

We next assessed whether LOY was associated with mutations in cancer driver genes.[30] Confirming the link with chromosomal instability discussed above, the most significantly associated mutated driver gene was *TP53* in the pan-cancer set (Figure 3E; Fisher's Exact p = 2.23×10^{-46}), as well as in COAD, head and neck squamous carcinoma (HNSC), and lung adenocarcinoma (LUAD) (Figure 3F; Table S3), in which *TP53* mutations are clonal (Figure 3D) and genomic instability is high (Figure 2E). In addition, we found *KDM5C* and *KDM6A* mutations enriched in the TCGA LOY pan-cancer cohort and in individual tumor types (*KDM5C* in KIRC with p = $1.09E^{-4}$, *KDM6A* in bladder cancer (BLCA) with p = $6.09E^{-3}$, and lung squamous carcinoma [LUSC] with p = 0.033; Figures 3F, S3B, and S3C; Table S3).[16,31,32] Both genes have homologs on Y, indicative of a potential two-hit event.[16]

Driver candidates on the Y chromosome

To identify individual candidate tumor suppressors on Y, we queried our catalog of Y copy number calls for recurrent focal deletions in tumors without LOY (STAR Methods). Only two tumor types, HNSC and LUSC, had a focal deletion on Yq potentially extending into the heterochromatin region (Table S2B). In LUSC, this region includes the ubiquitously expressed[14] tumor suppressor *KDM5D*[13,18] and *EIF1AY*, homolog of the X-linked cancer gene *EIF1AX*. All focal deletions occurred in tumors with many copy number changes; thus, further functional study will be required to determine whether *KDM5D* and/or *EIF1AY* are tumor suppressors in LUSC.

We also searched for MSY and PAR genes with recurrent point mutations (single nucleotide variants [SNVs], short indels) in tumors without LOY as potential tumor suppressors (STAR Methods), reasoning that they might be inactivated by point mutations in the absence of LOY. However, no Y-linked genes were significantly recurrently altered. A caveat to this analysis is the careful filtering of the TCGA pan-cancer mutation calls, including

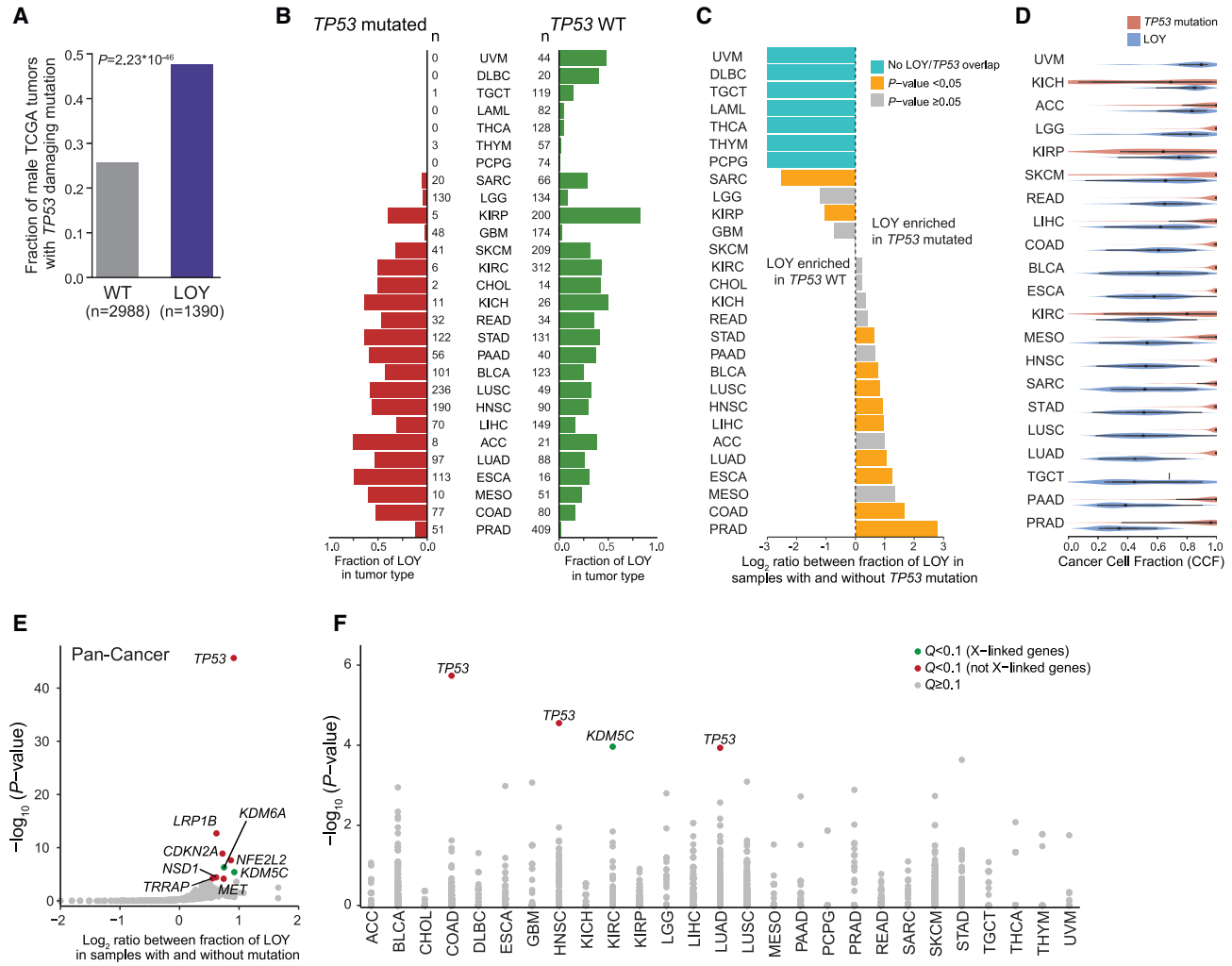


Figure 3. Association of LOY with *TP53* and other point mutation drivers

(A) Frequency of *TP53* mutations in LOY and WT (with Y) tumors. p value calculated with one-sided Fisher's Exact test.

(B) Fraction of LOY tumors stratified by *TP53* mutation status (left, red, *TP53* damaging mutation; right, green, no *TP53* damaging mutation, WT). "n" indicates the number of samples used in the calculation of each bar.

(C) Log_2 ratio of the fraction of LOY tumors with and without *TP53* mutation in each tumor type (red divided by green bar in B). Gold bars indicate significant enrichment for LOY in *TP53* WT or mutated samples by two-sided Fisher's Exact test. Teal bars are capped for diseases where the fraction of LOY in *TP53* mutated samples is zero in (B) (red bars).

(D) Distribution of the CCFs for LOY (blue) and *TP53* (red) for samples with the respective alteration. Tumor types are ordered top to bottom by decreasing median LOY CCF, and only cohorts with at least 10 LOY samples are shown.

(E) Cancer genes (STAR Methods) with somatic mutations enriched in LOY compared to WT male tumors in the TCGA Pan-Cancer cohort. Y axis shows one-sided Fisher's Exact test for enrichment of somatic mutations in LOY tumors. Colors as in (F).

(F) Enrichment of somatic mutations in LOY tumors in individual tumor types (compare to E).

panels-of-normals and read depth requirements,[33] could lead to overconservative removal of mutations on the Y chromosome.

LOY predicts patient outcome

To test whether somatic LOY is associated with patient outcome, we performed survival analysis for all male TCGA cases with progression-free survival (PFS) and chromosome Y status information. Although a pan-cancer analysis is necessarily influenced by tumor type, we observed a poorer outcome in tumors with LOY (Figure 4A; hazard ratio [HR] = 1.17, log rank p value < 0.00117). Stratification by tumor stage, a major determinant of outcome, confirmed the trend of poorer outcomes for LOY tumors (stage I/II, HR = 1.17, p = 0.04; Stage III/IV, HR = 1.13, p = 0.12; Figure S3D). Within individual tumor types, survival regression analysis that accounted for age at diagnosis and overall chromosomal instability through the number of chromosome arms lost in each tumor type showed that LOY was independently significantly ($Q < 0.05$) associated with poor PFS in UVM (HR = 8.03, log likelihood ratio test p = 0.002, $Q = 0.029$), prostate adenocarcinoma (PRAD; HR = 3.36,

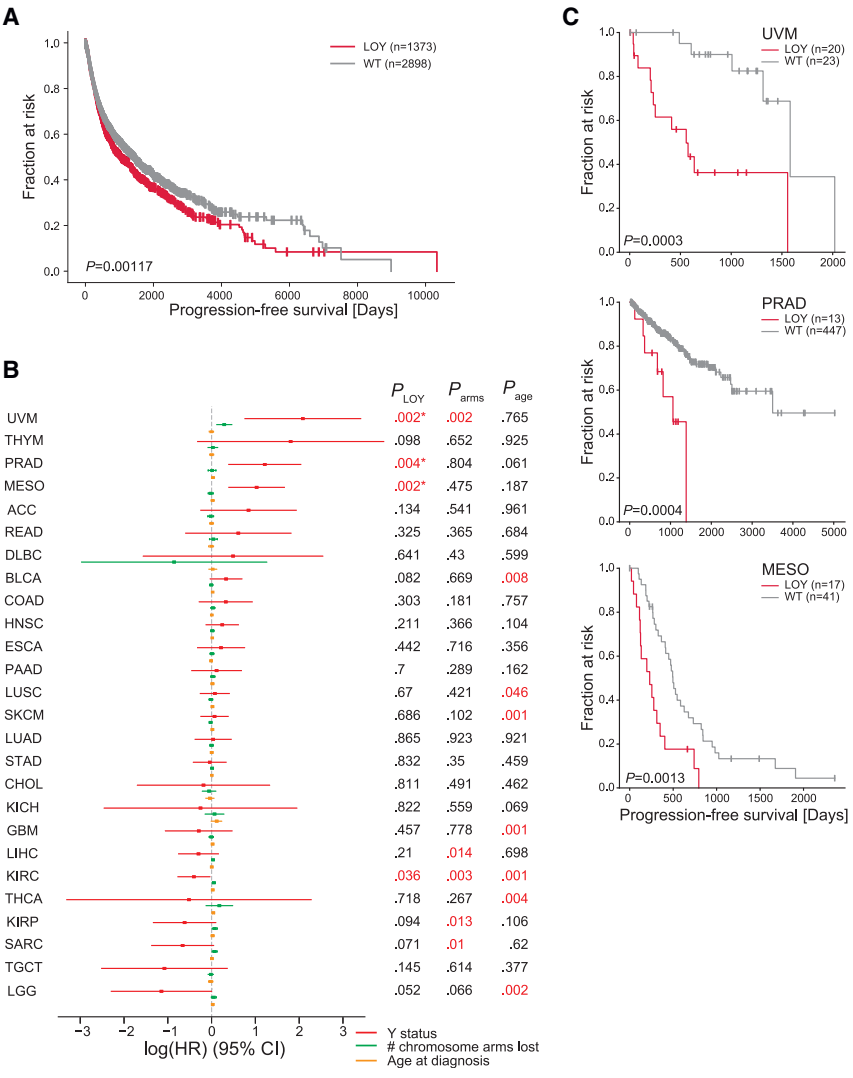


Figure 4. LOY predicts patient outcome

(A) Kaplan-Meier survival statistic depicts progression-free survival for male tumors with complete LOY compared to WT tumors.

(B) HRs (log) and 95% confidence intervals for LOY, the number of chromosome arms lost, and age for each tumor type, sorted by LOY HR. Regression p values marked in red indicate $p < 0.05$, * significant at $Q < 0.1$.

(C) Kaplan-Meier survival curves for tumor types with significant contribution of LOY to survival. UVM, uveal melanoma; PRAD, prostate adenocarcinoma; MESO, mesothelioma.

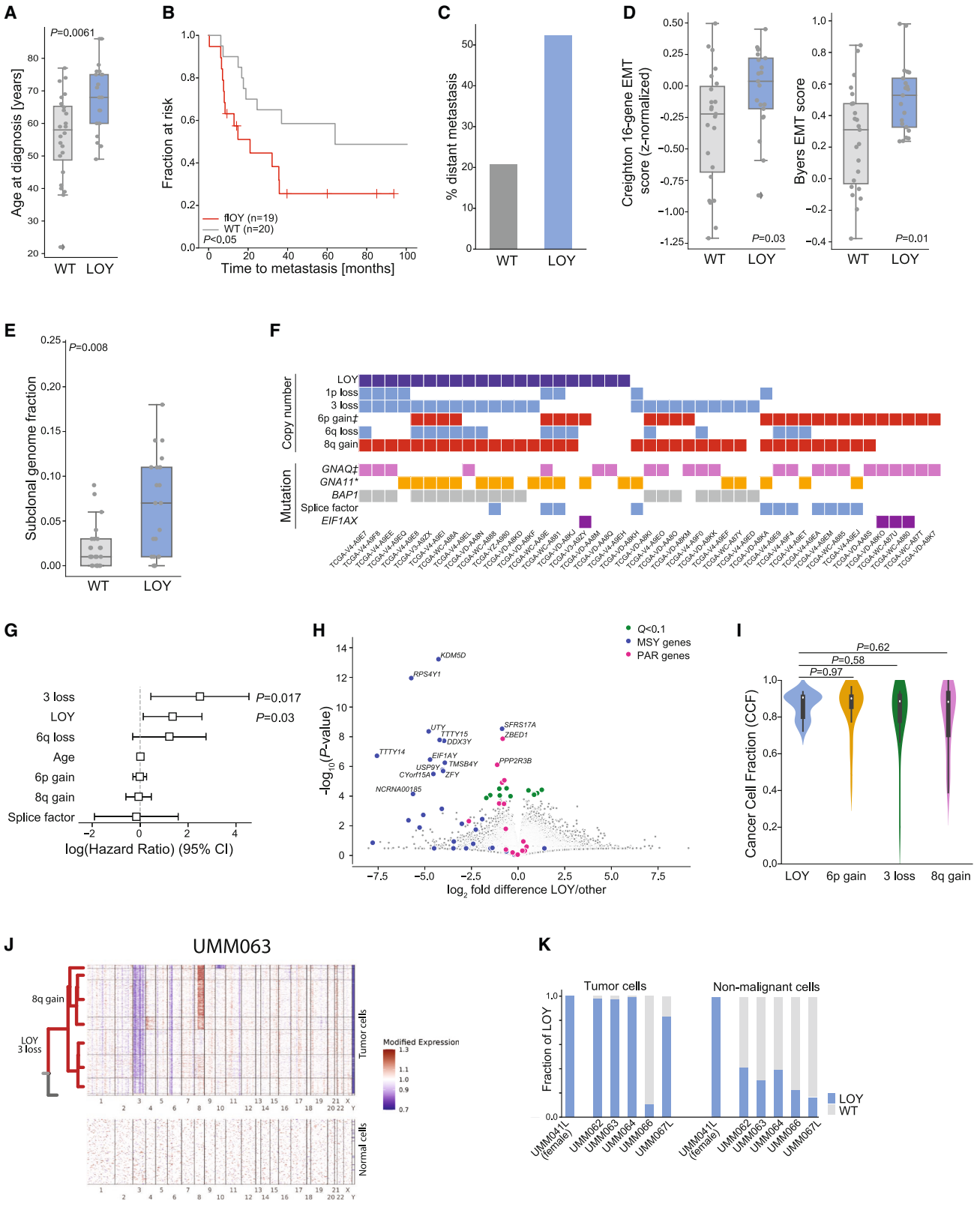
p = 0.004, $Q = 0.038$) and MESO (HR = 2.8, p = 0.002, $Q = 0.029$) (Figures 4B and 4C; Table S4). Importantly, we did not observe a general trend toward older age of diagnosis, as might be expected from known trends of somatic LOY in non-malignant tissues,[7] with only UVM and KIRC LOY patients being significantly older (Figure S3E).

LOY is a candidate driver event in uveal melanoma

The high LOY frequency not associated with genomic instability, *TP53* mutations (Figures 2 and 3), and strong difference in PFS between LOY and WT samples in UVM (Figures 4B and 4C) prompted us to further study the role of LOY in this tumor type. UVM is a rare malignancy of melanocytes in the eye. In contrast to melanoma of the skin, this disease is not associated with ultraviolet light exposure, and its defining genomic aberrations differ from the classic cutaneous melanoma events.[34,35] This includes few well-characterized arm-level copy number changes and rare *TP53* mutations.[35,36] LOY was observed in UVM decades ago, but not investigated much further.[37] For unknown reasons, UVM incidence is biased toward males, and this bias increases with patient age[34] (Figure S4A). Interestingly, we found a 10-year difference in median age at diagnosis between WT and LOY for UVM in the TCGA cohort (median age of 57 vs. 68, MWU p = 0.006; Figures 5A and S3E) and a steady increase in LOY frequency with age (Figure S4B), raising the question of whether age-related or environmentally induced LOY occurs in the cell of origin. Importantly, the observed difference in outcome was not explained by age, with LOY patients faring significantly worse than age-matched WT patients (p = 0.02; Figure S4C; see also Figure 4B).

We validated the remarkable difference in outcome between LOY and WT tumors in an independent cohort of 39 male UVM patients with gene expression profiles[38] (STAR Methods; Figure S4D). Confirming our prior results, 49% (19/39) of these patients had Y-expression-based fLOY (compared to 47% in the TCGA WES cohort), and outcome for these patients was significantly worse (Figure 5B; p = 0.045). Recapitulating the age difference in the TCGA cohort, age at diagnosis was significantly higher for patients with fLOY tumors (68 vs. 60 years; p = 0.01).

Metastatic spread is common in patients with UVM, and once a tumor has metastasized, prognosis is very poor.[34] We found that LOY tumors were significantly more likely to metastasize than WT tumors (52% vs. 21%, p = 0.03, Fisher's Exact test), and this trend was confirmed in the validation dataset (68% vs. 45%, p = 0.2; Figures 5C and S4E). In addition, epithelial-to-mesenchymal (EMT) gene expression scores[39,40] (Figure 5D) and subclonal genome fraction[35] (Figure 5E), associated with metastatic potential and aggressiveness of cancer cells, were significantly higher in LOY compared to WT cases (EMT, MWU $p < 0.03$; subclonal fraction, MWU p = 0.008). Associations with poor outcome factors were not due to overall genomic instability, as—although significant (MWU p value = 0.004)—both the number and difference of arm losses between LOY (median arm loss 1, after correcting for



(legend on next page)

3 loss) and WT tumors (median zero arm losses) are small (Figure S4F).

UVMs are characterized by ubiquitous early mutations in *GNAQ/GNA11*, monosomy/loss of chromosome 3, 8q gain, mutations in and copy loss of the *BAP1* tumor suppressor, and mutations in the *EIF1AX* elongation and the *SF3B1* splicing factors.[35] Monosomy 3, *BAP1* (located on chromosome 3), and 8q have been linked to poor prognosis, while *EIF1AX* predicts better outcome. To test whether the LOY outcome difference seen in the two cohorts could be explained by these underlying, well-described alterations, we calculated the overlap of events across tumors (Figure 5F). There was no association between *BAP1* alterations and LOY (Fisher's Exact p = 0.14), monosomy 3 (p = 0.14), or 8q gain (p = 0.73); however, chromosome 6p gain was depleted from LOY tumors (p = 0.017). Importantly, we observed no significant overlap of *EIF1AX* mutations and loss of its Y-linked homolog *EIF1AY* through LOY (p = 0.61), consistent with their opposite predictive effects on patient outcome. Interestingly, *GNA11* mutations were enriched (p = 0.017) and *GNAQ* mutations were depleted (p = 0.037) in LOY tumors, suggesting a potential selective advantage of LOY in the presence of the *GNA11* but not *GNAQ* genotype. Further supporting an independent role of LOY, a multivariate Cox proportional hazard model that accounted for known outcome predictors in UVM maintained LOY as a significant predictor (Figure 5G; p = 0.03). Aside from a single *DDX3Y* missense mutation, no mutations in Y-linked genes were detected in the UVM TCGA cohort.[35,42]

Lack of somatic point mutations on Y and the fact that Y is lost in its entirety in UVM complicate the identification of a single tumor suppressor. We therefore compared gene expression profiles between male LOY and WT tumors to identify differentially regulated candidate Y-linked drivers. As expected, nearly all significantly differentially expressed (DE) genes are located on Y (either in the MSY or PAR; Figure 5H; Table S5). Among DE genes, chromatin regulators *KDM5D* and *UTY* (*KDM6C*) are among the most significantly downregulated in LOY tumors (Figure S4G). *KDM5D* demethylates the active histone mark H3K4me3. Loss of *KDM5D* interferes with regulation of normal gene expression and has been shown to accelerate cell cycling.[13] This is particularly intriguing in the context of frequent mutations and loss of the histone ubiquitinase *BAP1*, suggesting multifaceted alterations of chromatin regulation in UVM. One potential driver function of LOY might thus be the loss of the *KDM5D* tumor suppressor.

Next, we sought to understand the clonality of LOY in UVM by directly assessing the fraction of LOY in both bulk samples from TCGA and single cells[41] (STAR Methods). Prior work has suggested the key drivers in UVMs occur early and potentially at the same time, resulting in mostly clonal driver alterations.[43] In the TCGA UVM cohort, the very high CCFs for LOY (Figure 3D) were not significantly different than the CCFs of other critical drivers, including 3p/3q loss and *BAP1* mutations[35] (Figure 5I). Phylogenetic reconstruction in single-cell gene expression profiles[41] mapped LOY alongside monosomy 3 and other copy number driver events (Figures 5J and S4H). In total, four out of five male tumors had substantial to complete (83 to >99%) fLOY, with much smaller percentages of fLOY in non-malignant cells (17%–42%; Figure 5K). Single-cell gene expression experiments can suffer from gene detection limits ("drop out"), which likely contributes to the high fLOY rates in non-malignant cells. Yet, fLOY cells were not of lower quality overall than WT cells (Figure S4I), supporting true LOY in these cells. Together, these results suggest that loss of the Y chromosome is an early, clonal event in UVM.

Our sex chromosome calls suggest that in addition to LOY in males, LOX is also common in female UVM with nearly equal frequency (43%; Figures S2E and S2F). Similar to LOY tumors, LOX cases had increased rates of metastasis (47% vs. 15%, Fisher's Exact p = 0.06; Figure S4J), not due to overall chromosomal instability (median arm loss count 1 for LOX and WT groups; Figure S4K). In contrast to LOY tumors, there was no significant age difference (MWU p = 0.8). These results suggest that sex chromosome loss is not simply a consequence of age and implies genetic driver events among shared homologous genes on X and Y.

Differential dependencies in LOY cell lines

Finally, we searched for gene dependencies that are found in male cell lines with LOY as specific, potentially targetable vulnerabilities.[44] Top dependencies among fLOY cell lines in aggregate (Table S6; STAR Methods) were *DDX3X* and *EIF1AX*, X-linked homologs of *DDX3Y* and *EIF1AY* (Figures 6A and 6B). *DDX3X/DDX3Y* are RNA helicases implicated in transcription. *EIF1AX/EIF1AY* are essential and ubiquitously expressed translation initiation factors, and *EIF1AX* is somatically mutated in several cancer types.[42] Concordantly, Y-linked *DDX3Y* and *EIF1AY* were among the top four differentially expressed genes between LOY and WT cell lines (Figure S5). This includes a UVM cell line, OMM.1, that is strongly dependent on these two genes (gene effect scores −1.195755 and −1.326149, respectively). It is important to note that although

Figure 5. LOY is a candidate driver event in uveal melanoma

(A) Distributions of age at diagnosis for LOY and WT male UVM patients from TCGA. p value calculated with the MWU test.

(B) Kaplan-Meier survival curve for male patients from an independent cohort.[38]

(C) Percentage of TCGA patients with distant metastasis with LOY or WT.

(D) Distribution of EMT scores[39,40] for LOY and WT TCGA tumors. p value calculated with the MWU test.

(E) Distribution of subclonal genome fraction[35] for TCGA tumors. p value calculated with the MWU test.

(F) Somatic mutations and LOY for each TCGA UVM tumor. * indicates significant (p < 0.05) enrichment; ‡ indicates depletion within LOY.

(G) Cox proportional hazards and p values for known predictors of poor survival in UVM and LOY.

(H) Volcano plot showing differentially expressed genes between LOY and WT UVM in TCGA. MSY, male-specific region on Y; PAR, pseudoautosomal region.

(I) CCF distributions for LOY and other known driver events in UVM from TCGA. Only samples with an event are included in the distribution. Two low purity cases were removed from the plot. p values calculated with the MWU test.

(J) Representative example of an inferred copy number profile from single cells from a UVM tumor[41] and inferred phylogenetic tree with driver alterations mapped to branches.

(K) Fraction of tumor single cells from five male UVMs[41] with LOY (blue) or intact Y (WT, gray). One female patient (UMM041L) is shown as control.

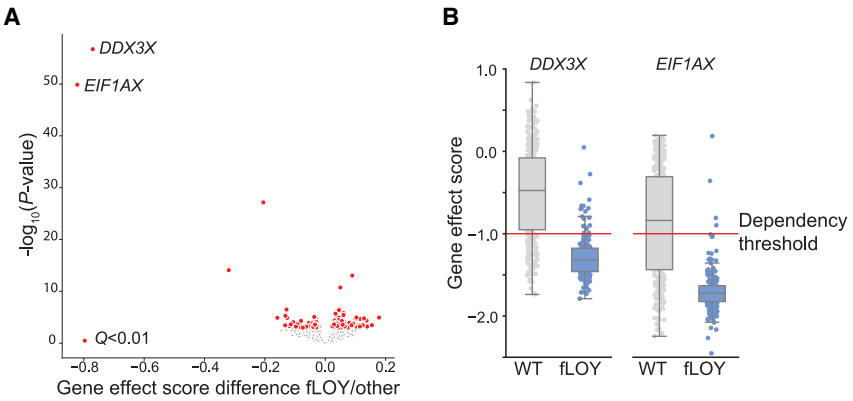


Figure 6. Differential dependencies in LOY cell lines

(A) Volcano plot depicting differential dependency for each gene in LOY vs. WT cell lines from Cancer Cell Line Encyclopedia (CCLE).[44] Red dots indicate significantly different dependencies.

(B) Gene effect scores for *DDX3X* and *EIF1AX* show no dependency on these genes in cell lines with Y and strong dependency in LOY cell lines. Dependency map-defined threshold for dependency (−1) is shown as red line.

LOY cell lines are more dependent on these genes, both genes have low effect scores in non-LOY cell lines, suggesting the possibility of broader dependency or additional alterations affecting the X-linked paralogs. *DDX3X* and *EIF1AX* thus represent two complex vulnerabilities that could be therapeutically exploited in LOY tumors.

DISCUSSION

Anecdotal evidence has suggested that chromosome Y might play a driver role in some cancer types. Through development of specific methods to analyze copy numbers of chromosome Y from genomic and transcriptomic profiles, our study presents a comprehensive analysis of Y copy number variation across 29 tumor types of different lineages. We show that in the TCGA, LOY is extremely common in many tumor types, with frequencies higher than some classic driver genes. LOY can occur in the context of overall genomic instability, where LOY is potentially a passenger, lost due to its small size and low gene content. Recent work has shown that the actual size of Y varies considerably between individuals,[45] and future analyses comparing accurate Y size obtained from long-read sequencing and likelihood of LOY will shed more light on the importance of Y size for LOY in malignant and non-malignant cells. Importantly, even in tumor types where LOY occurs as a "passenger" byproduct of genomic instability, a contribution to tumor biology through Y-linked gene loss cannot be excluded. We also present evidence for a driver role of LOY in UVM, where LOY is a clonal event that is strongly associated with predictors of poor survival, including older patient age, tumor heterogeneity, proliferation, and time to progression. Yet LOY is independent and potentially more predictive than other well-known somatic alterations in TCGA UVM, suggesting it as a prognostic biomarker that is relatively easy to detect through standard clinical assays and warrants functional investigation. Age-related mosaic LOY in the bone marrow and its association with hematologic malignancies[46] raise the question of whether LOY as somatic mutations in clonally expanded cells elsewhere in the body increases the risk of malignant transformation. The clonal nature and association with older age in UVM patients in TCGA suggests that LOY might exist, and potentially be detectable, in precursor lesions.

Because chromosome Y usually does not have a "backup" copy, LOY leads to loss of several unique and ubiquitously expressed genes, with potentially important implications for cell fitness (e.g., *KDM5D*, *UTY/KDM6C*, and *RPS4Y1*). Through loss of these genes, which include several paralogs of X-linked genes, LOY can create unique and targetable vulnerabilities.[47] In addition, loss of a copy of the PARs introduces loss of heterozygosity (LOH) and could cause decreased expression of several immune receptor genes, including the macrophage receptors *CSF2RA* and *IL3RA* and the surface marker *CD99*. *CD99* is highly expressed in glioblastoma cells, where it correlates with migration and invasiveness.[48,49] The dependence of glioblastoma multiforme (GBM) tumors on *CD99* could explain the relatively low LOY rates observed in this disease. Future functional studies in appropriate LOY and WT model systems will be necessary to fully understand the role of LOY in tumors and the interaction with immune and stromal cells of the tumor microenvironment in these contexts.

Limitations of the study

There are several important limitations to our study. First, exome sequencing allowed us to only investigate protein-coding genes on Y, without the power to assess potential driver super-enhancers or other regulatory elements. With only three protein-coding genes in PAR2, the confidence in calling copy numbers in this region is reduced compared to PAR1 and the MSY. Although WGS provides coverage outside of exons, the lower coverage depth limits faithful detection of subclonal LOY and LOY in the presence of low-level mosaic LOY in the normal comparison blood or tissue sample. Deeply sequenced, evenly covered WGS will overcome some of these drawbacks in the future. In addition, the relatively small number of patients in some tumor cohorts, especially the rare cancer UVM, limits our statistical power in LOY vs. WT comparisons. Finally, an important limitation of our study is the lack of evidence for recurrent alteration of a specific tumor suppressor on Y, as could be assessed by focal deletion or recurrent somatic mutation. This might be caused by lack of sensitivity of our methods (especially point mutation calling). It is also possible that whole-chromosome LOY creates a unique fitness advantage through loss of multiple tumor suppressors simultaneously while Y is relatively easily lost from cells. Future functional studies will be required to complement the existing literature on Y-linked tumor suppressors across all tumor types with frequent LOY.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- METHOD DETAILS
  - Somatic copy number for WES
  - Analysis of Y status for PCAWG WGS samples and comparison with TCGA WES datasets
  - Classification of LOY based on gene expression (fLOY)
  - Survival analysis
  - Genomic instability
  - Association of point mutations in TCGA and PCAWG datasets
  - Cell line dependencies
  - Incidence data for uveal melanoma
  - Uveal melanoma validation cohort
  - Classification of single cells
  - Clonality analysis for LOY in single cells
  - Statistics

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2023.06.006.

ACKNOWLEDGMENTS

We thank Dr. Gad Getz for valuable discussions of this work and Dr. Hang Lee from the MGH Biostatistics Core for assistance with survival analyses. E.R. is supported by funds from MGH, the Broad Institute, and the NCI. M.Q. is supported by an MGH ECOR Claflin Award to E.R. A.A.L. is supported by the Mark Foundation for Cancer Research and is a Leukemia & Lymphoma Society Scholar. A.A.L. and E.R. are supported by the Bertarelli Rare Cancer Fund at Harvard Medical School. This work was conducted with support from Harvard Catalyst and financial contributions from Harvard University and its affiliated academic healthcare centers. The content is solely the responsibility of the authors and does not necessarily represent the official views of Harvard Catalyst, Harvard University and its affiliated academic healthcare centers, or the National Institutes of Health.

AUTHOR CONTRIBUTIONS

E.R. and A.A.L. conceived the study. M.Q., J.P., and E.R. developed computational methods and contributed analysis. I.M. contributed analysis. M.Q. and E.R. wrote the manuscript.

DECLARATION OF INTERESTS

A.A.L. is a consultant for Qiagen. A.A.L. received research support from AbbVie and Stemline Therapeutics. J.P. is currently an employee of the Shanghai Artificial Intelligence Laboratory, Shanghai 200232, China. All other authors declare no competing interests.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: September 2, 2022
Revised: January 17, 2023
Accepted: June 8, 2023
Published: June 28, 2023

REFERENCES

1. Forsberg, L.A. (2017). Loss of chromosome Y (LOY) in blood cells is associated with increased risk for disease and mortality in aging men. Hum. Genet. *136*, 657–663. https://doi.org/10.1007/s00439-017-1799-2.
2. Forsberg, L.A., Rasi, C., Malmqvist, N., Davies, H., Pasupulati, S., Pakalapati, G., Sandgren, J., Diaz de Ståhl, T., Zaghlool, A., Giedraitis, V., et al. (2014). Mosaic loss of chromosome Y in peripheral blood is associated with shorter survival and higher risk of cancer. Nat. Genet. *46*, 624–628. https://doi.org/10.1038/ng.2966.
3. Thompson, D.J., Genovese, G., Halvardson, J., Ulirsch, J.C., Wright, D.J., Terao, C., Davidsson, O.B., Day, F.R., Sulem, P., Jiang, Y., et al. (2019). Genetic predisposition to mosaic Y chromosome loss in blood. Nature *575*, 652–657. https://doi.org/10.1038/s41586-019-1765-3.
4. Mitchell, E., Spencer Chapman, M., Williams, N., Dawson, K.J., Mende, N., Calderbank, E.F., Jung, H., Mitchell, T., Coorens, T.H.H., Spencer, D.H., et al. (2022). Clonal dynamics of haematopoiesis across the human lifespan. Nature *606*, 343–350. https://doi.org/10.1038/s41586-022-04786-y.
5. Guttenbach, M., Koschorz, B., Bernthaler, U., Grimm, T., and Schmid, M. (1995). Sex chromosome loss and aging: in situ hybridization studies on human interphase nuclei. Am. J. Hum. Genet. *57*, 1143–1150.
6. Machiela, M.J., Zhou, W., Karlins, E., Sampson, J.N., Freedman, N.D., Yang, Q., Hicks, B., Dagnall, C., Hautman, C., Jacobs, K.B., et al. (2016). Female chromosome X mosaicism is age-related and preferentially affects the inactivated X chromosome. Nat. Commun. *7*, 11843. https://doi.org/10.1038/ncomms11843.
7. Coorens, T.H.H., Moore, L., Robinson, P.S., Sanghvi, R., Christopher, J., Hewinson, J., Przybilla, M.J., Lawson, A.R.J., Spencer Chapman, M., Cagan, A., et al. (2021). Extensive phylogenies of human development inferred from somatic mutations. Nature *597*, 387–392. https://doi.org/10.1038/s41586-021-03790-y.
8. Dumanski, J.P., Lambert, J.C., Rasi, C., Giedraitis, V., Davies, H., Grenier-Boley, B., Lindgren, C.M., Campion, D., Dufouil, C., et al.; European Alzheimer's Disease Initiative Investigators (2016). Mosaic Loss of Chromosome Y in Blood Is Associated with Alzheimer Disease. Am. J. Hum. Genet. *98*, 1208–1219. https://doi.org/10.1016/j.ajhg.2016.05.014.
9. Brunelli, M., Eble, J.N., Zhang, S., Martignoni, G., and Cheng, L. (2003). Gains of chromosomes 7, 17, 12, 16, and 20 and loss of Y occur early in the evolution of papillary renal cell neoplasia: a fluorescent in situ hybridization study. Mod. Pathol. *16*, 1053–1059. https://doi.org/10.1097/01.MP.0000090924.90762.94.
10. Büscheck, F., Fraune, C., Garmestani, S., Simon, R., Kluth, M., Hube-Magg, C., Ketterer, K., Eichelberg, C., Höflmayer, D., Jacobsen, F., et al. (2021). Y-chromosome loss is frequent in male renal tumors. Ann. Transl. Med. *9*, 209. https://doi.org/10.21037/atm-20-3061.
11. Hunter, S., Gramlich, T., Abbott, K., and Varma, V. (1993). Y chromosome loss in esophageal carcinoma: an in situ hybridization study. Genes Chromosomes Cancer *8*, 172–177. https://doi.org/10.1002/gcc.2870080306.
12. Komura, K., Jeong, S.H., Hinohara, K., Qu, F., Wang, X., Hiraki, M., Azuma, H., Lee, G.S.M., Kantoff, P.W., and Sweeney, C.J. (2016). Resistance to docetaxel in prostate cancer is associated with androgen receptor activation and loss of KDM5D expression. Proc. Natl. Acad. Sci. USA *113*, 6259–6264. https://doi.org/10.1073/pnas.1600420113.
13. Komura, K., Yoshikawa, Y., Shimamura, T., Chakraborty, G., Gerke, T.A., Hinohara, K., Chadalavada, K., Jeong, S.H., Armenia, J., Du, S.Y., et al. (2018). ATR inhibition controls aggressive prostate tumors deficient in Y-linked histone demethylase KDM5D. J. Clin. Invest. *128*, 2979–2995. https://doi.org/10.1172/JCI96769.

14. Godfrey, A.K., Naqvi, S., Chmátal, L., Chick, J.M., Mitchell, R.N., Gygi, S.P., Skaletsky, H., and Page, D.C. (2020). Quantitative analysis of Y-Chromosome gene expression across 36 human tissues. Genome Res. *30*, 860–873. https://doi.org/10.1101/gr.261248.120.

15. Bellott, D.W., Hughes, J.F., Skaletsky, H., Brown, L.G., Pyntikova, T., Cho, T.J., Koutseva, N., Zaghlul, S., Graves, T., Rock, S., et al. (2014). Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. Nature *508*, 494–499. https://doi.org/10.1038/nature13206.

16. Dunford, A., Weinstock, D.M., Savova, V., Schumacher, S.E., Cleary, J.P., Yoda, A., Sullivan, T.J., Hess, J.M., Gimelbrant, A.A., Beroukhim, R., et al. (2017). Tumor-suppressor genes that escape from X-inactivation contribute to cancer sex bias. Nat. Genet. *49*, 10–16. https://doi.org/10.1038/ng.3726.

17. Hertel, J.D., Huettner, P.C., Dehner, L.P., and Pfeifer, J.D. (2010). The chromosome Y-linked testis-specific protein locus TSPY1 is characteristically present in gonadoblastoma. Hum. Pathol. *41*, 1544–1549. https://doi.org/10.1016/j.humpath.2010.04.007.

18. Arseneault, M., Monlong, J., Vasudev, N.S., Laskar, R.S., Safisamghabadi, M., Harnden, P., Egevad, L., Nourbehesht, N., Panichnantakul, P., Holcatova, I., et al. (2017). Loss of chromosome Y leads to down regulation of KDM5D and KDM6C epigenetic modifiers in clear cell renal cell carcinoma. Sci. Rep. *7*, 44876. https://doi.org/10.1038/srep44876.

19. Li, C.H., Haider, S., Shiah, Y.-J., Thai, K., and Boutros, P.C. (2018). Sex Differences in Cancer Driver Genes and Biomarkers. Cancer Res. *78*, 5527–5537. https://doi.org/10.1158/0008-5472.CAN-18-0362.

20. Caceres, A., Jene, A., Esko, T., Perez-Jurado, L.A., and Gonzalez, J.R. (2020). Extreme down-regulation of chromosome Y and cancer risk in men. J Natl Cancer Inst. https://doi.org/10.1093/jnci/djz232.

21. Priestley, P., Baber, J., Lolkema, M.P., Steeghs, N., de Bruijn, E., Shale, C., Duyvesteyn, K., Haidari, S., van Hoeck, A., Onstenk, W., et al. (2019). Pan-cancer whole-genome analyses of metastatic solid tumours. Nature *575*, 210–216. https://doi.org/10.1038/s41586-019-1689-y.

22. Shen, R., and Seshan, V.E. (2016). FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. Nucleic Acids Res. *44*, e131. https://doi.org/10.1093/nar/gkw520.

23. Hollows, R., Wei, W., Cazier, J.B., Mehanna, H., Parry, G., Halford, G., and Murray, P. (2019). Association between loss of Y chromosome and poor prognosis in male head and neck squamous cell carcinoma. Head Neck *41*, 993–1006. https://doi.org/10.1002/hed.25537.

24. Beroukhim, R., Mermel, C.H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., Barretina, J., Boehm, J.S., Dobson, J., Urashima, M., et al. (2010). The landscape of somatic copy-number alteration across human cancers. Nature *463*, 899–905. https://doi.org/10.1038/nature08822.

25. Duijf, P.H.G., Schultz, N., and Benezra, R. (2013). Cancer cells preferentially lose small chromosomes. Int. J. Cancer *132*, 2316–2326. https://doi.org/10.1002/ijc.27924.

26. Taylor, A.M., Shih, J., Ha, G., Gao, G.F., Zhang, X., Berger, A.C., Schumacher, S.E., Wang, C., Hu, H., Liu, J., et al. (2018). Genomic and Functional Approaches to Understanding Cancer Aneuploidy. Cancer Cell *33*, 676–689.e3. https://doi.org/10.1016/j.ccell.2018.03.007.

27. Aylon, Y., and Oren, M. (2011). p53: guardian of ploidy. Mol. Oncol. *5*, 315–323. https://doi.org/10.1016/j.molonc.2011.07.007.

28. Donehower, L.A., Soussi, T., Korkut, A., Liu, Y., Schultz, A., Cardenas, M., Li, X., Babur, O., Hsu, T.K., Lichtarge, O., et al. (2019). Integrated Analysis of TP53 Gene and Pathway Alterations in The Cancer Genome Atlas. Cell Rep. *28*, 1370–1384.e5. https://doi.org/10.1016/j.celrep.2019.07.001.

29. Gerstung, M., Jolly, C., Leshchiner, I., Dentro, S.C., Gonzalez, S., Rosebrock, D., Mitchell, T.J., Rubanova, Y., Anur, P., Yu, K., et al. (2020). The evolutionary history of 2,658 cancers. Nature *578*, 122–128. https://doi.org/10.1038/s41586-019-1907-7.

30. Rheinbay, E., Nielsen, M.M., Abascal, F., Wala, J.A., Shapira, O., Tiao, G., Hornshøj, H., Hess, J.M., Juul, R.I., Lin, Z., et al. (2020). Analyses of non-coding somatic drivers in 2,658 cancer whole genomes. Nature *578*, 102–111. https://doi.org/10.1038/s41586-020-1965-x.

31. Dalgliesh, G.L., Furge, K., Greenman, C., Chen, L., Bignell, G., Butler, A., Davies, H., Edkins, S., Hardy, C., Latimer, C., et al. (2010). Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. Nature *463*, 360–363. https://doi.org/10.1038/nature08672.

32. Ricketts, C.J., and Linehan, W.M. (2015). Gender Specific Mutation Incidence and Survival Associations in Clear Cell Renal Cell Carcinoma (CCRCC). PLoS One *10*, e0140257. https://doi.org/10.1371/journal.pone.0140257.

33. Ellrott, K., Bailey, M.H., Saksena, G., Covington, K.R., Kandoth, C., Stewart, C., Hess, J., Ma, S., Chiotti, K.E., McLellan, M., et al. (2018). Scalable Open Science Approach for Mutation Calling of Tumor Exomes Using Multiple Genomic Pipelines. Cell Syst. *6*, 271–281.e7. https://doi.org/10.1016/j.cels.2018.03.002.

34. Kaliki, S., and Shields, C.L. (2017). Uveal melanoma: relatively rare but deadly cancer. Eye (Lond) *31*, 241–257. https://doi.org/10.1038/eye.2016.275.

35. Robertson, A.G., Shih, J., Yau, C., Gibb, E.A., Oba, J., Mungall, K.L., Hess, J.M., Uzunangelov, V., Walter, V., Danilova, L., et al. (2017). Integrative Analysis Identifies Four Molecular and Clinical Subsets in Uveal Melanoma. Cancer Cell *32*, 204–220.e15. https://doi.org/10.1016/j.ccell.2017.07.003.

36. Johansson, P., Aoude, L.G., Wadt, K., Glasson, W.J., Warrier, S.K., Hewitt, A.W., Kiilgaard, J.F., Heegaard, S., Isaacs, T., Franchina, M., et al. (2016). Deep sequencing of uveal melanoma identifies a recurrent mutation in PLCB4. Oncotarget *7*, 4624–4631. https://doi.org/10.18632/oncotarget.6614.

37. Prescher, G., Bornfeld, N., and Becher, R. (1990). Nonrandom chromosomal abnormalities in primary uveal melanoma. J. Natl. Cancer Inst. *82*, 1765–1769. https://doi.org/10.1093/jnci/82.22.1765.

38. Laurent, C., Valet, F., Planque, N., Silveri, L., Maacha, S., Anezo, O., Hupe, P., Plancher, C., Reyes, C., Albaud, B., et al. (2011). High PTP4A3 phosphatase expression correlates with metastatic risk in uveal melanoma patients. Cancer Res. *71*, 666–674. https://doi.org/10.1158/0008-5472.CAN-10-0605.

39. Gibbons, D.L., and Creighton, C.J. (2018). Pan-cancer survey of epithelial-mesenchymal transition markers across the Cancer Genome Atlas. Dev. Dyn. *247*, 555–564. https://doi.org/10.1002/dvdy.24485.

40. Byers, L.A., Diao, L., Wang, J., Saintigny, P., Girard, L., Peyton, M., Shen, L., Fan, Y., Giri, U., Tumula, P.K., et al. (2013). An epithelial-mesenchymal transition gene signature predicts resistance to EGFR and PI3K inhibitors and identifies Axl as a therapeutic target for overcoming EGFR inhibitor resistance. Clin. Cancer Res. *19*, 279–290. https://doi.org/10.1158/1078-0432.CCR-12-1558.

41. Durante, M.A., Rodriguez, D.A., Kurtenbach, S., Kuznetsov, J.N., Sanchez, M.I., Decatur, C.L., Snyder, H., Feun, L.G., Livingstone, A.S., and Harbour, J.W. (2020). Single-cell analysis reveals new evolutionary complexity in uveal melanoma. Nat. Commun. *11*, 496. https://doi.org/10.1038/s41467-019-14256-1.

42. Bailey, M.H., Tokheim, C., Porta-Pardo, E., Sengupta, S., Bertrand, D., Weerasinghe, A., Colaprico, A., Wendl, M.C., Kim, J., Reardon, B., et al. (2018). Comprehensive Characterization of Cancer Driver Genes and Mutations. Cell *174*, 1034–1035. https://doi.org/10.1016/j.cell.2018.07.034.

43. Field, M.G., Durante, M.A., Anbunathan, H., Cai, L.Z., Decatur, C.L., Bowcock, A.M., Kurtenbach, S., and Harbour, J.W. (2018). Punctuated evolution of canonical genomic aberrations in uveal melanoma. Nat. Commun. *9*, 116. https://doi.org/10.1038/s41467-017-02428-w.

44. Tsherniak, A., Vazquez, F., Montgomery, P.G., Weir, B.A., Kryukov, G., Cowley, G.S., Gill, S., Harrington, W.F., Pantel, S., Krill-Burger, J.M., et al. (2017). Defining a Cancer Dependency Map. Cell *170*, 564–576.e16. https://doi.org/10.1016/j.cell.2017.06.010.

45. Hallast, P., Ebert, P., Loftus, M., Yilmaz, F., Audano, P.A., Logsdon, G.A., Bonder, M.J., Zhou, W., Höps, W., Kim, K., et al. (2022). Assembly of 43 diverse human Y chromosomes reveals extensive complexity and variation. bioRxiv. https://doi.org/10.1101/2022.12.01.518658.

46. Ouseph, M.M., Hasserjian, R.P., Dal Cin, P., Lovitch, S.B., Steensma, D.P., Nardi, V., and Weinberg, O.K. (2021). Genomic alterations in patients with somatic loss of the Y chromosome as the sole cytogenetic finding in bone marrow cells. Haematologica *106*, 555–564. https://doi.org/10.3324/haematol.2019.240689.

47. Köferle, A., Schlattl, A., Hörmann, A., Thatikonda, V., Popa, A., Spreitzer, F., Ravichandran, M.C., Supper, V., Oberndorfer, S., Puchner, T., et al. (2022). Interrogation of cancer gene dependencies reveals paralog interactions of autosome and sex chromosome-encoded genes. Cell Rep. *39*, 110636. https://doi.org/10.1016/j.celrep.2022.110636.

48. Úrias, U., Marie, S.K.N., Uno, M., da Silva, R., Evagelinellis, M.M., Caballero, O.L., Stevenson, B.J., Silva, W.A., Jr., Simpson, A.J., and Oba-Shinjo, S.M. (2014). CD99 is upregulated in placenta and astrocytomas with a differential subcellular distribution according to the malignancy stage. J. Neuro Oncol. *119*, 59–70. https://doi.org/10.1007/s11060-014-1462-x.

49. Seol, H.J., Chang, J.H., Yamamoto, J., Romagnuolo, R., Suh, Y., Weeks, A., Agnihotri, S., Smith, C.A., and Rutka, J.T. (2012). Overexpression of CD99 Increases the Migration and Invasiveness of Human Malignant Glioma Cells. Genes Cancer *3*, 535–549. https://doi.org/10.1177/1947601912473603.

50. Mermel, C.H., Schumacher, S.E., Hill, B., Meyerson, M.L., Beroukhim, R., and Getz, G. (2011). GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol. *12*, R41. https://doi.org/10.1186/gb-2011-12-4-r41.

51. Waskom, M. (2021). seaborn: statistical data visualization. J. Open Source Softw. *6*, 3021.

52. Davidson-Pilon, C. (2019). lifelines: survival analysis in Python. J. Open Source Softw. *4*, 1317.

53. Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nat. Biotechnol. *33*, 495–502. https://doi.org/10.1038/nbt.3192.

54. Lawrence, M.S., Stojanov, P., Mermel, C.H., Robinson, J.T., Garraway, L.A., Golub, T.R., Meyerson, M., Gabriel, S.B., Lander, E.S., and Getz, G. (2014). Discovery and saturation analysis of cancer genes across 21 tumour types. Nature *505*, 495–501. https://doi.org/10.1038/nature12912.

55. Tickle, T., Tirosh, I., Georgescu, C., Brown, M., and Haas, B. (2019). inferCNV of the Trinity CTAT Project (Klarman Cell Observatory, Broad Institute of MIT and Harvard).

56. Yu, G., Lam, T.T.Y., Zhu, H., and Guan, Y. (2018). Two Methods for Mapping and Visualizing Associated Data on Phylogeny Using Ggtree. Mol. Biol. Evol. *35*, 3041–3043. https://doi.org/10.1093/molbev/msy194.

57. Geoffroy, V., Herenger, Y., Kress, A., Stoetzel, C., Piton, A., Dollfus, H., and Muller, J. (2018). AnnotSV: an integrated tool for structural variations annotation. Bioinformatics *34*, 3572–3574. https://doi.org/10.1093/bioinformatics/bty304.

58. Amemiya, H.M., Kundaje, A., and Boyle, A.P. (2019). The ENCODE Blacklist: Identification of Problematic Regions of the Genome. Sci. Rep. *9*, 9354. https://doi.org/10.1038/s41598-019-45839-z.

59. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303. https://doi.org/10.1101/gr.107524.110.

60. Carter, S.L., Cibulskis, K., Helman, E., McKenna, A., Shen, H., Zack, T., Laird, P.W., Onofrio, R.C., Winckler, W., Weir, B.A., et al. (2012). Absolute quantification of somatic DNA alterations in human cancer. Nat. Biotechnol. *30*, 413–421. https://doi.org/10.1038/nbt.2203.

61. Frankish, A., Diekhans, M., Ferreira, A.M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J.M., Sisu, C., Wright, J., Armstrong, J., et al. (2019). GENCODE reference annotation for the human and mouse genomes. Nucleic Acids Res. *47*, D766–D773. https://doi.org/10.1093/nar/gky955.

62. Tirosh, I., Izar, B., Prakadan, S.M., Wadsworth, M.H., 2nd, Treacy, D., Trombetta, J.J., Rotem, A., Rodman, C., Lian, C., Murphy, G., et al. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science *352*, 189–196. https://doi.org/10.1126/science.aad0501.

63. Li, Y., Roberts, N.D., Wala, J.A., Shapira, O., Schumacher, S.E., Kumar, K., Khurana, E., Waszak, S., Korbel, J.O., Haber, J.E., et al. (2020). Patterns of somatic structural variation in human cancer genomes. Nature *578*, 112–121. https://doi.org/10.1038/s41586-019-1913-9.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
TCGA WES BAM files	Genomic Data Commons	https://portal.gdc.cancer.gov/legacy-archive/search/f
TCGA WGS BAM files	Genomic Data Commons	https://portal.gdc.cancer.gov/
TCGA Pan-Cancer somatic mutation data	Genomic Data Commons	https://gdc.cancer.gov/about-data/publications/pancanatlas/mc3.v0.2.8.PUBLIC.maf.gz
TCGA aneuploidy data	Taylor et al.[26]	Table S2
TCGA Pan-Cancer purity and ploidy	Genomic Data Commons	https://gdc.cancer.gov/about-data/publications/pancanatlas/TCGA_mastercalls.abs_tables_JSedit.fixed.txt
TCGA Pan-Cancer gene expression data	Genomic Data Commons	https://gdc.cancer.gov/about-data/publications/pancanatlas/EBPlusPlusAdjustPANCAN_IlluminaHiSeq_RNASeqV2.geneExp.tsv
TCGA Pan-Cancer clinical data	Genomic Data Commons	https://gdc.cancer.gov/about-data/publications/pancanatlas/clinical_PANCAN_patient_with_followup.tsv
dbSNP	The Single Nucleotide Polymorphism Database	https://ftp.ncbi.nlm.nih.gov/snp/organisms/human_9606_b151_GRCh37p13/VCF/common_all_20180423.vcf.gz
PCAWG somatic mutations	ICGC Data Portal	https://dcc.icgc.org/releases/PCAWG/consensus_snv_indel
PCAWG somatic copy number	ICGC Data Portal	https://dcc.icgc.org/releases/PCAWG/consensus_cnv
PCAWG somatic structural variants	ICGC Data Portal	https://dcc.icgc.org/releases/PCAWG/consensus_sv
CCLE gene expression	DepMap 22Q2	https://depmap.org/portal/
Incidence data for “eye and orbit”	NCI’s Surveillance, Epidemiology and End Results website	https://seer.cancer.gov/statistics-network/explorer/application.html
Gene expression of 39 individual uveal melanoma cases	Laurent et al.[38]	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE22138
Single-cell expression data of uveal melanoma	Durante et al.[41]	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE139829
Software and algorithms		
FACETS 0.6.2	Shen et al.[22]	https://github.com/mskcc/facets
GISTIC 2.0.23	Mermel et al.[50]	http://software.broadinstitute.org/software/cprg/?q=node/31
Python seaborn package 0.11.2	Waskom et al.[51]	https://seaborn.pydata.org/
Python lifelines package 0.26.5	Davidson-Pilon[52]	https://github.com/CamDavidsonPilon/lifelines/
Seurat 4.0.1	Satija et al.[53]	https://satijalab.org/seurat/
MutSig2CV	Lawrence et al.[54]	https://github.com/getzlab/MutSig2CV
InferCNV 1.15.0	Tickle et al.[55]	https://github.com/broadinstitute/infercnv
ggtree 3.6.2	Yu et al.[56]	https://bioconductor.org/packages/release/bioc/html/ggtree.html
AnnotSV	Geoffroy et al.[57]	https://github.com/lgmgeo/AnnotSV
Other		
Analyses, and resources related to call LOY	This study	https://github.com/rheinbaylab/LOY_in_primary_tumors

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Esther Rheinbay (erheinbay@mgh.harvard.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

This study used TCGA Pan-Cancer Atlas obtained from https://gdc.cancer.gov/about-data/publications/pancanatlas. Code for sex chromosome copy number calling and analysis notebooks are available at https://github.com/rheinbaylab/LOY_in_primary_tumors.

METHOD DETAILS

Somatic copy number for WES

We used the FACETs copy number caller[22] framework as a base for calling Y copy number changes, with several modifications: (i) In the human hg19 genome assembly, the two PAR regions at the ends of the X and Y chromosomes are exclusively assigned to the X chromosome, leading to discordant X copy number and heterozygous SNP calls in these regions in tumors with XY genotype. We therefore separated the PARs from X and treated them as additional autosomes. For simplicity, we assume that in normal control tissue of most males, there will be exactly one copy of each X and Y, and two X and zero Y copies in the majority of female tissues (see below for limitations). All autosomes will typically have two copies regardless of biological sex. (ii) The Y chromosome contains a large heterochromatin region as well as repetitive stretches of sequence that can lead to incorrect copy calls. We therefore identified ambiguously mappable regions on Y by counting read depth in contiguous 500 bp bins in female TCGA normal samples and two female WGS samples. Bins in which at least one position had a read depth $\geq$10 in WGS samples were excluded from analysis. We also excluded positions that had at least read depth $\geq$10 in $\geq$20% female samples from TCGA. Additionally, we excluded regions on the X and Y chromosomes with low mappability[58] (score <0.5); and ubiquitously high coverage[58] and the Y centromere[59] (from GATK). (iii) We used the dbSNP151 set of common germline variants to gather allelic counts on all chromosomes. To increase the number of coverage data points on gene-poor Y, we added additional pseudo SNPs ("pseudo_snps = 100") and coverage count at the middle position for each exon of chromosomes X and Y to the coverage pileup file. (iv) Because the segmentation parameter cval is sensitive to the number of data points (SNPs), we adjusted cval = 50 to increase the copy segment resolution on Y while leaving the default cval = 150 for all other chromosomes.

With these modifications, we used the FACETs framework to estimate total copy number (tcn) for Y. In cases where purity and ploidy estimates diverged substantially (ploidy difference $\geq$1 or purity difference $\geq$0.2) from previously published TCGA values (TCGA Pan-Cancer file TCGA_mastercalls.abs_tables_Jsedit.fixed.txt), we corrected tcn with published TCGA purity and ploidy values. Official TCGA calls are based on the ABSOLUTE method,[60] which relies on copy number as well as somatic mutations as data source; this is advantageous especially for tumors with few copy alterations where purity and ploidy estimates from copy number alone can be less reliable. The correction was calculated as

$$tcn.tcga = round\left(\frac{2^{seqlogr.adj.tcga} - (1 - purity.tcga)}{purity.tcga}\right),$$

Where cnlr.median is the median log ratio, $seqlogr.adj.tcga = cnlr.median - digLogR.tcga$ and $digLogR.tcga = -log\,2\left(\frac{ploidy.tcga*purity.tcga+2*(1-purity.tcga)}{2}\right)$.

If $round(ploidy)$ is odd, then we calculate the expected maximum Y WT copy number (max.Y.WT) as $ceiling\left(\frac{ploidy}{2}\right)$ and the expected minimum Y WT copy number (min.Y.WT) as $floor\left(\frac{ploidy}{2}\right)$. If $round(ploidy)$ is even, then the expected Y copy number (both min.Y.WT and max.Y.WT) is $round\left(\frac{ploidy}{2}\right)$ (half the number of autosomes). A segment is called "gained" if tcn>max.Y.WT and a segment "lost" if tcn<min.Y.WT or tcn = 0. A segment on Y is considered "WT" if min.Y.WT $\leq$ tcn $\leq$ max.Y.WT. Then we classified CNVs of Y chromosome into several groups (Figure S6): (1) WT: All segments on Y are "WT"; (2) Gain: At least one segment on Y is gained and potential other segments are WT; (3) Gain+Loss: at least one segment is gained and at least one segment is lost; (4) pLOY (partial LOY): at least one segment on Y is lost, and all others are WT; (5) rLOY: All segments of the Y ($\geq$99% of the Y chromosome) are lost, and tcn $\geq$1 (e.g. after genome doubling) (6) LOY: All segments on Y ($\geq$99% of the Y chromosome) are lost, and tcn = 0.

X chromosome status classification in tumors from female patients is calculated similarly, except that LOX is compared with $round(ploidy)$ and LOX is defined as $\geq$99% of the X chromosome is lost (Figure S6). We identified 39 tumors with discordant reported gender and genomic sex or low quality by manually review, and these were not included in further analysis (Tables S1A and S1B). Recurrently mutated broad and focal copy number changes on X and Y were identified with GISTIC2[50] with parameter -rx 0 to include the sex chromosomes. BLCA was excluded from Figure 2B due to lack of markers to evaluate Yq. Yq peak regions were manually trimmed back to the boundary with the heterochromatin region after automatic extension with markers by GISTIC to limit the focal peak to regions with measured coverage data points. Significant Y regions were annotated with the GENCODE V40 gene list.[61]

A caveat of our analysis is that LOY rates are likely conservative and underestimated for two reasons: (i) most controls in TCGA are peripheral blood samples that may be subject to age-related somatic LOY, and thus tumor Y copies can be "gained" relative to control; (ii) if the fraction of somatic LOY in the control and tumor are approximately equal, no difference will be detected and LOY is not called.

Analysis of Y status for PCAWG WGS samples and comparison with TCGA WES datasets

Consensus copy number profiles for male PCAWG WGS samples were obtained from https://dcc.icgc.org/releases/PCAWG/consensus_cnv. Inference of Y status was performed as described above for TCGA WES. We extracted 354 male TCGA cases with both WGS in PCAWG and WES for direct comparison. For samples with discordant calls, we applied our method to TCGA WGS and manually investigated the results. Comparison of LOY fraction was performed between our TCGA calls and the non-overlapping, independent ICGC portion of PCAWG. Tumor types were matched by tumor classification and subtype.

Classification of LOY based on gene expression (fLOY)

Gene expression RSEM values were downloaded from the GDC Pan-Cancer Atlas website (File EBPlusPlusAdjustPANCAN_IlluminaHiSeq_RNASeqV2.geneExp.tsv). A list of genes in the male specific region of Y was used to identify a seven-gene signature of genes consistently expressed in normal tissues to score expression from this chromosome (*RPS4Y1, DDX3Y, KDM5D, USP9Y, EIF1AY, UTY, ZFY*; Figures S1A and S1B). A common set of housekeeping genes[62] was used to control for overall expression activity. The ratio of the mean of expression of the seven Y genes to the expression across housekeeping genes was calculated as "Y expression score" for each individual sample. A ratio threshold of 0.035, corresponding to Y gene expression of 3.5% of housekeeping gene expression, was manually identified to classify tumors into "functional" LOY (fLOY) (ratio <0.035) and "WT" (ratio $\geq$0.035) (Figures S1C and S1D). 478 male patients did not have RNA-seq expression calls (mostly GBM) and were not included in the classification or downstream analysis based on Y status. Differential expression p values were calculated with a *t*-test followed by the Benjamini-Hochberg FDR correction.

Survival analysis

Survival analyses were conducted using the lifelines Python package (https://github.com/CamDavidsonPilon/lifelines/). Progression-free survival from the Pan-Cancer Atlas (TCGA-CDR-SupplementalTableS1.xlsx) was used as endpoint. "age_at_initial_pathologic_diagnosis" and "ajcc_pathologic_tumor_stage" from the same data source were included to account for age and tumor stage. Significance for Kaplan-Meier statistics were calculated with the log rank test. Hazard ratios were calculated with the Cox proportional hazards model and significance was assessed with the log likelihood ratio test. The total number of chromosome arms lost in a sample were obtained from Taylor et al.[26] Only tumor types with at least five samples in either LOY or WT group were included. Patient-matching for the age-matched survival analysis was achieved by binning patients into 3-year bins and randomly selecting a similar number of LOY and WT patients from each group. 1000 selections and KM survival analyses were run. Age distribution and KM plot for the sample set with the median p-value across all randomizations is displayed in Figure S4C.

Genomic instability

Aneuploidy scores for TCGA cases were obtained from Taylor et al.[26] Arm-level statistics were calculated for each tumor type using the GISTIC copy number significance software.[50] The relative copy number value used as the input for GISTIC is calculated as $\log_2$(-copy value)-1 from FACETs results. To prevent GISTIC from identifying "losses" of the haploid sex chromosomes in male samples (where the normal state is only one copy of X and Y respectively), we manually doubled the copy number values of X and Y calls before running GISTIC. Default parameters were used, except for: run_broad_analysis = 1, broad_len_cutoff = 0.5, remove_X = 0.

Association of point mutations in TCGA and PCAWG datasets

TCGA Pan-Cancer multi-center somatic mutation calls (mc3.v0.2.8.PUBLIC.maf.gz) were used for mutation analyses. Only variants with the following damaging classifications were included: 'Frame_Shift_Del', 'Frame_Shift_Ins', 'In_Frame_Del', 'Missense_Mutation', 'Nonsense_Mutation', 'Splice_Site', 'Translation_Start_Site'. Point mutations (final_consensus_passonly.snv_mnv_indel.icgc.public.maf.gz and final_consensus_passonly.snv_mnv_indel.tcga.controlled.maf.gz) and structural variation (SV) calls (final_consensus_sv_bedpe_passonly.icgc.public.tgz and final_consensus_sv_bedpe_passonly.tcga.public.tgz) derived from WGS for PCAWG samples were obtained from the ICGC PCAWG DCC. Structural variations were associated with genes using AnnotSV. Association of LOY with point mutations in ~600 known cancer genes[30] or SVs was evaluated with the Fisher's Exact test (alternative = "greater") and Benjamini-Hochberg FDR correction for multiple hypothesis testing. Recurrence analysis for mutated Y chromosome genes was performed with MutSig2CV with default settings on only male cases without LOY and rLOY chromosome status.

Cell line dependencies

Gene expression and CRISPR gene effect scores were obtained from the DepMap portal (https://depmap.org/portal/; version 22Q2). Cell lines were classified with respect to sex and fLOY status by using a combination of Y gene expression of the seven-gene signature described above and *XIST* expression levels. We excluded 12 annotated male cell lines with absent Y expression and high *XIST* expression as potentially female lines (although it is possible that some of these lines are male with LOY and X multisomy, where additional X chromosomes will undergo silencing. Differential gene expression and dependency for all cell lines in aggregate and each disease type were evaluated with a two-sided *t*-test followed by FDR multiple hypothesis correction.

Incidence data for uveal melanoma

Incidence data for cases in the United States in 2019 for cancer site "eye and orbit" were downloaded from the NCI's Surveillance, Epidemiology and End Results website (https://seer.cancer.gov/statistics-network/explorer/application.html) on May 11, 2022.

Uveal melanoma validation cohort

Functional LOY scores were calculated with Affymetrix array gene expression data using the same seven Y-linked genes and housekeeping gene sets for 39 uveal melanoma samples from Laurent et al.[38] Classification into fLOY and WT tumors was obtained from the bimodal distributions of Y gene expression (Figure S4D) and survival analysis was performed as described above.

Classification of single cells

Single-cell expression data from uveal melanoma tumors was obtained from Durante et al.[41] scRNA-seq data were processed and analyzed using R (4.0.5) and Seurat package (4.0.1).[53] Raw counts of 11 samples were read into R using the Read10X function and aggregated into one Seurat object. Several metrics were used to account for dead cells and droplets: 1) The number of unique molecular identifiers (UMIs) per cell; 2) The number of detected genes per cell; 3) The proportion of mitochondrial genes; 4) Number of genes detected per UMI ($\log_{10}$(number of detected genes)/$\log_{10}$(number of UMIs)). Only cells with UMI count greater than 500, 250 to 8000 expressed genes, mitochondrial content less than 10% and >0.8 detected genes/UMI were retained for future analysis. After filtering, 52,294 cells were left. Data were normalized using the NormalizeData function in Seurat with LogNormalize setting and a scaling factor of 10,000. Principal component analysis (PCA) was used to reduce dimensionality with number of variable features set to 2000. Clustering was conducted with FindClusters using the first 20 principal components and 1.5 as resolution parameter. The original Louvain algorithm was utilized for modularity optimization, which resulted in 46 clusters. Identified clusters were visualized using t-distributed stochastic neighbor embedding (t-SNE) and they were annotated as described in the source using the following markers[41]: Tumor cells (*MLANA, MITF*, *DCT*), T Cells (*CD3D*, *CD3E*, *CD8A*), B cells (*CD19, CD79A, MS4A1*), plasma cells (*IGHG1, MZB1, SDC1, CD79A*), monocytes and macrophages (*CD68, CD163, CD14*), NK Cells (*FGFBP2, FCG3RA, CX3CR1*), retinal pigment epithelium (*RPE65*), photoreceptor cells (*RCVRN*), and endothelial cells (*PECAM1, VWF*). To measure functional LOY, we calculated average Y expression for the seven most expressed genes (*DDX3Y, EIF1AY, KDM5D, RPS4Y1, USP9Y, UTY, ZFY*) and determined a threshold from female cell values as

$$mean(average\ Y) + 3 * standard\ deviation(average\ Y) = 0.00079.$$

Cells with average Y expression smaller than the threshold were classified as LOY, "WT" otherwise.

Clonality analysis for LOY in single cells

We used the inferCNV algorithm[55] for inferring copy number changes from RNA-seq data in five male uveal melanoma tumors, following a similar pipeline and setting as in the original publication,[41] filtered and annotated cells (as described above) were used for further analysis. We selected normal reference cells as those with high expression of *CD3E* (>2 standard deviations above the mean expression) and no expression of *PRAME* and *HTR2B* (as described in the original publication), and only retained inferred tumor cells with UMI count greater than 3000. Because the number of expressed Y genes (cutoff: 0.1 in the inferCNV) is insufficient for inferring CNV changes with inferCNV, we manually assigned expression of Y genes with our corresponding predicted Y classification (as described above) in step 7: LOY samples (as identified above) were assigned an expression value at the 10^{th} percentile of the sample distribution. WT samples were assigned the median cell expression value. InferCNV "subcluster" was run with the following parameter settings: HMM_type = "i3", cutoff = 0.1, denoise = True, tumor_subcluster_partition_method = "random_trees". We calculated the average modified expression for each tumor subcluster and the cluster of all normal cells (to define the root), and reconstructed phylogenetic trees for each sample.

Statistics

Non-parametric comparisons were performed using the Mann-Whitney U test implemented in the Python scipy.stats package. Plots were generated in Python using the Seaborn package.

CellPress

Supplemental figures

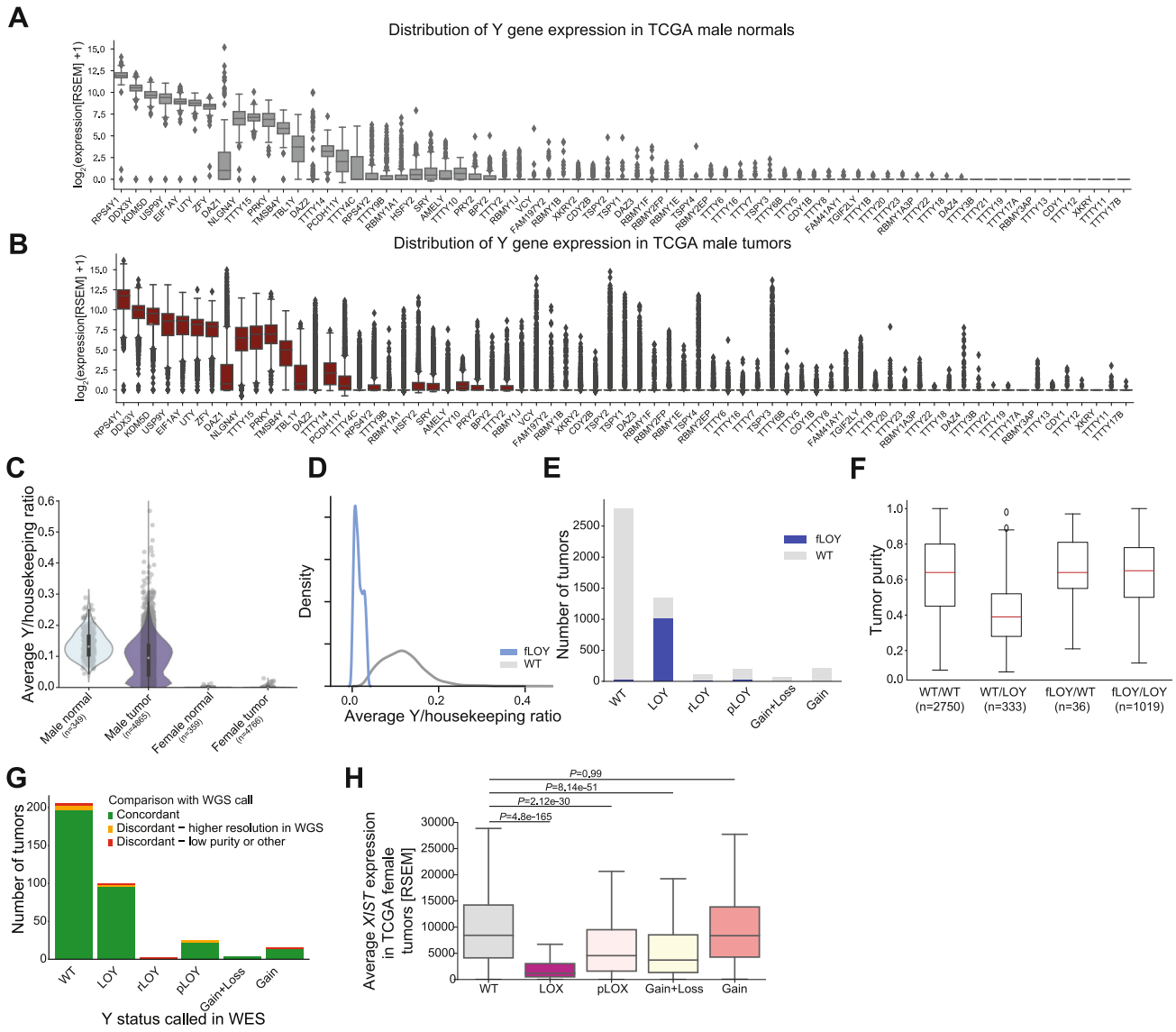


Figure S1. Definition of fLOY and validation of LOY/LOX calls, related to Figure 1

(A) Y gene expression for TCGA male normal samples ordered by median expression value.

(B) Y gene expression for TCGA male tumor samples ordered by median expression value.

(C) Distribution of Y/housekeeping gene expression ratio in TCGA male and female normal and tumor samples.

(D) Y/housekeeping expression ratio for male tumor samples called fLOY (blue) and WT (gray).

(E) Bar chart depicts concordance between exome-based LOY calls and expression-based fLOY calls.

(F) Tumor purity for male tumor samples identified as WT by both methods. Discordant calls and loss of Y (fLOY/LOY) show that LOY calls missed by expression have lower purity.

(G) Comparison of Y status calls from WES and WGS for tumors for which both were available. Green, concordant calls between both sequencing strategies; orange, WGS call is more likely to be correct after manual review, owing to higher resolution; red, discordant calls between WGS and WES caused by low purity, differential coverage, or mosaic LOY in the normal tissues.

(H) Distribution of *XIST* expression by X alteration in female TCGA tumors. p value calculated with the MWU test.

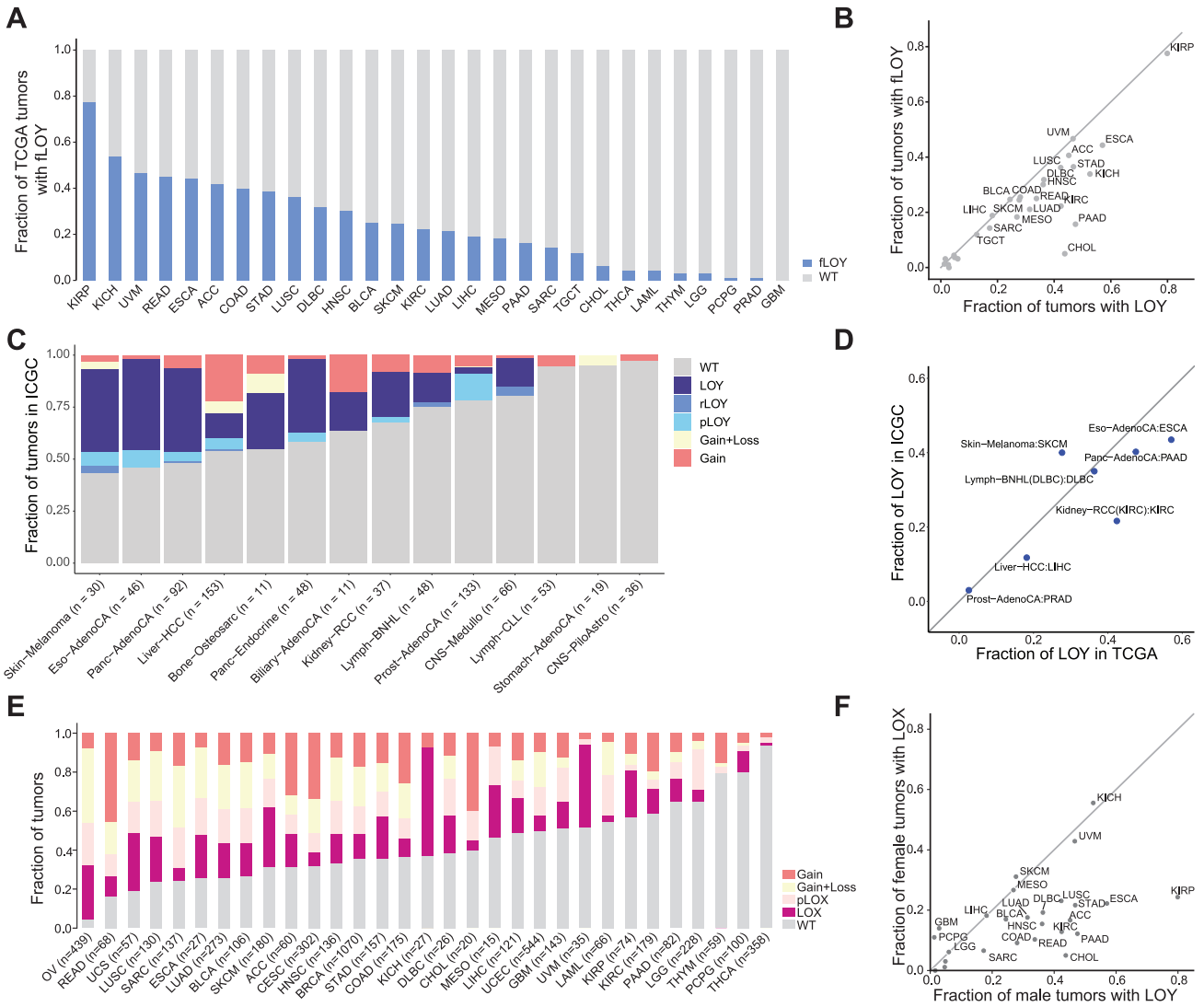


Figure S2. Frequency of LOY/LOX across tumor types, related to Figure 2

(A) Fraction of TCGA tumors by tumor type identified as fLOY (compared to Figure 2A).

(B) Comparison of exome-inferred LOY and expression-based fLOY in male TCGA tumors.

(C) Fraction of Y alterations in 808 male tumors from the ICGC (PCAWG) dataset. Only tumor types with greater than 10 cases are included.

(D) Comparison of the fraction of LOY in ICGC and TCGA. Labels represent (ICGC histology abbreviation):(TCGA study abbreviation). Only the diffuse large B-cell lymphoma subset ("Lymph-BNHL") was selected for matching the corresponding tumor type DLBC in TCGA. Tumor types with at least 20 cases are included.

(E) X alterations in TCGA female tumors by tumor type. Tumor types are ordered by fraction of WT samples.

(F) Distribution of fraction of LOY in male tumors vs. fraction of LOX in female tumors.

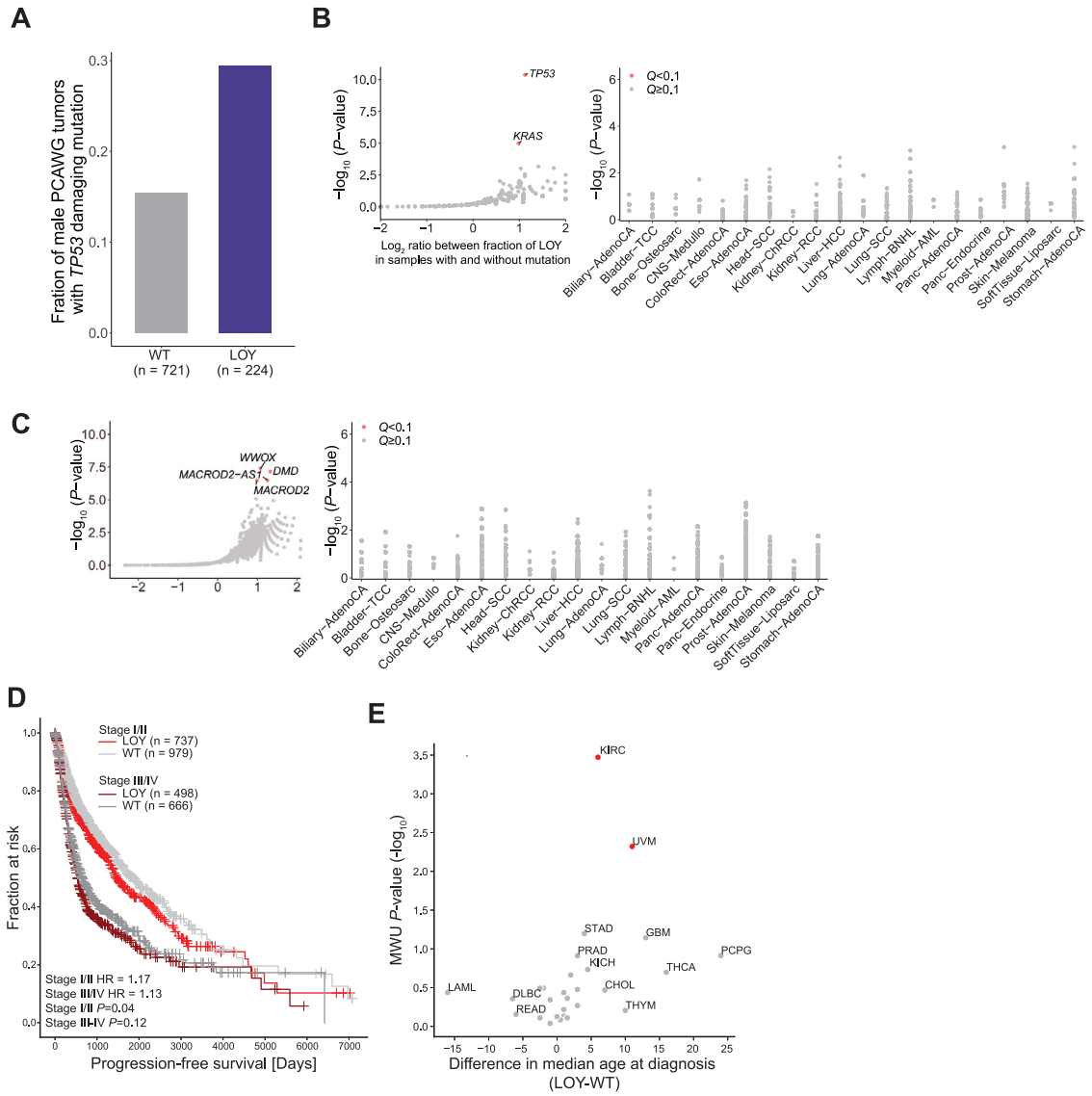


Figure S3. Association of LOY with genomic instability, survival, and age, related to Figures 3 and 4

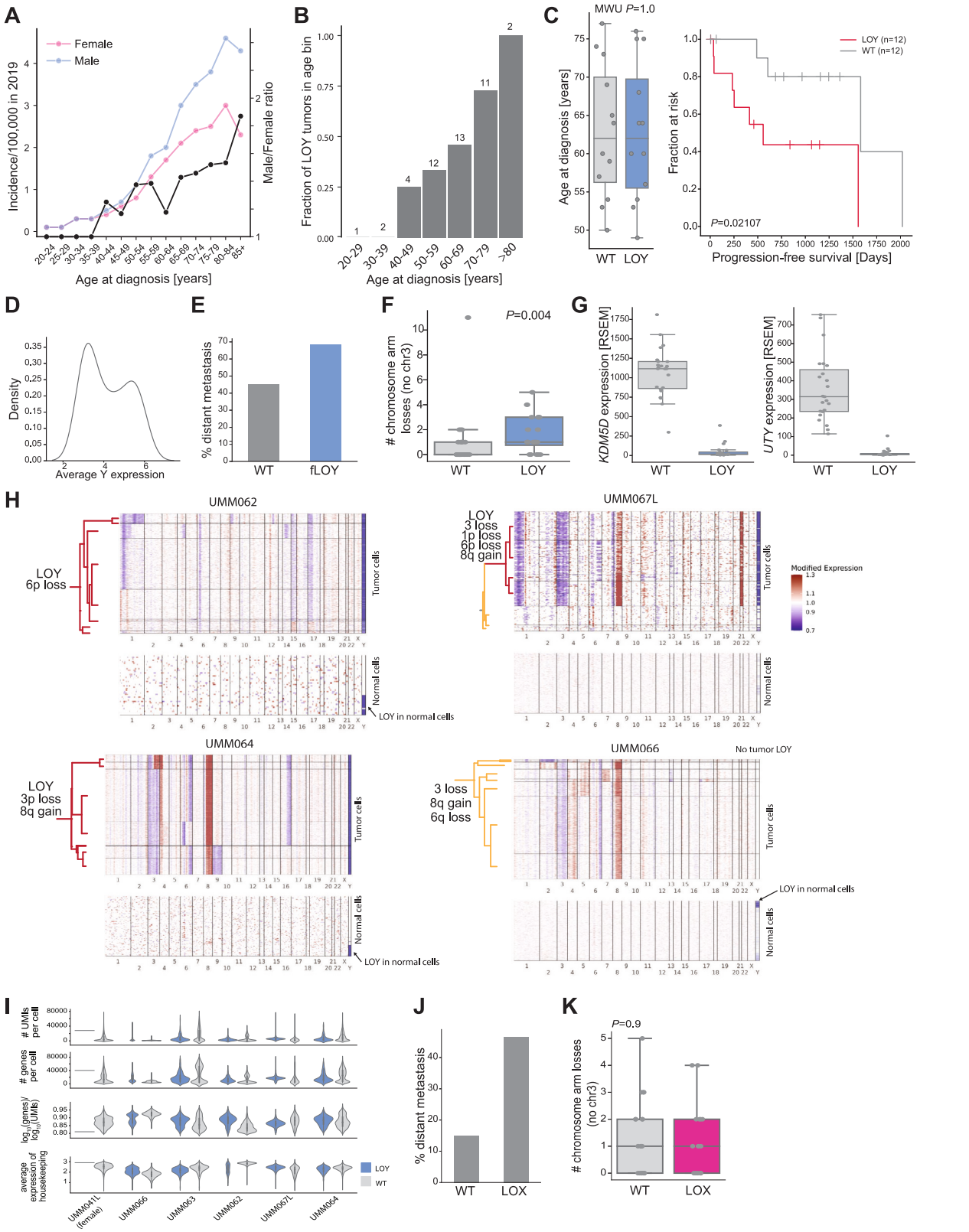
(A) Fraction of LOY and WT (with Y) PCAWG tumors with *TP53* mutation.

(B) Enrichment of cancer gene somatic mutations in LOY over WT tumors for the PCAWG pan-cancer cohort (left) and individual tumor types (right). p values calculated with one-sided Fisher's Exact test.

(C) Enrichment of structural variants in the PCAWG pan-cancer cohort and individual tumor types (p values as in B). *WWOX, DMD*, and *MACROD2* are known fragile sites, confirming association of LOY with genomic instability.[63]

(D) Kaplan-Meier analysis for TCGA LOY/WT tumors by stage.

(E) Difference in age at diagnosis by tumor type. Age difference between WT and LOY in years is shown on the x axis. Y axis indicates $-\log_{10}$ of MWU p value between WT and LOY age distributions. Significant tumor types after FDR correction ($Q < 0.1$) are highlighted in red.



(legend on next page)

Figure S4. LOY and LOX in UVM, related to Figure 5

(A) Surveillance, Epidemiology, and End Results (SEER) incidence for UVM for male (blue) and female (pink) patients. The male/female incidence ratio is shown in black (secondary y axis).

(B) Fraction of TCGA UVM LOY tumors by binned age at diagnosis. The number above each bar reflects the number of cases in each age bin.

(C) Kaplan-Meier survival analysis for UVM after age-matching of patients. Boxplot shows age distribution of patients selected in each group.

(D) Distribution of average Y gene expression among male tumors from Laurent et al.[38]

(E) Percentage of patients who develop distant metastasis from Laurent et al.[38] by tumor LOY status (compared to Figure 5C).

(F) Distribution of arm losses from Taylor et al.[26] in LOY and WT male UVM tumors. Arm loss counts for the driver alteration chromosome 3 (3p, 3q) were subtracted if the sample had chromosome 3 loss.

(G) TCGA gene expression for Y-linked genes *KDM5D* and *UTY* (*KDM6C*) by tumor LOY status.

(H) Copy number profiles for male UVM tumor and normal cells[41] inferred from RNA-seq (STAR Methods). Top-level known UVM drivers are annotated. For detailed evolution analysis, see Durante et al.[41]

(I) Different metrics for single-cell quality (# unique molecular identifiers [UMIs] per cell, # detected genes per cell, $\log_{10}$ [genes per UMI], average housekeeping gene expression level) for WT and LOY cells in UVM show that LOY is not due to fewer overall genes detected and missed Y genes. UMI, unique molecular identifier.

(J) Fraction of patients with distant metastasis in WT and LOX tumors.

(K) Distribution of arm losses from Taylor et al.[26] in LOX and WT male UVM tumors. Arm loss counts for the driver alteration chromosome 3 (3p, 3q) were subtracted if the sample had chromosome 3 loss.

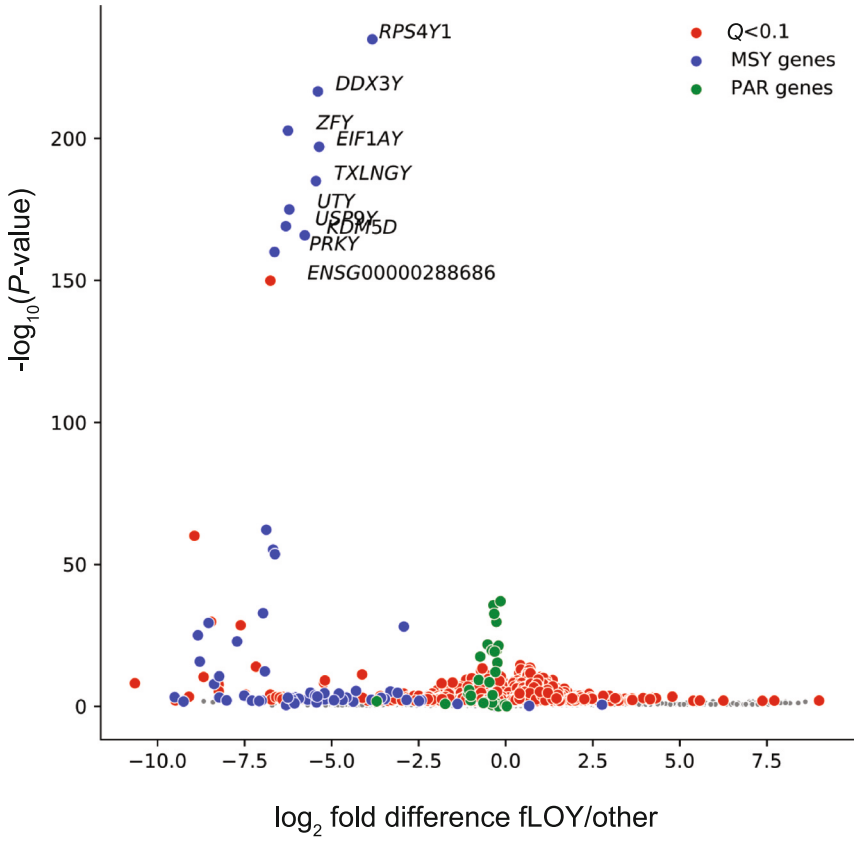


Figure S5. Differential gene expression between LOY and WT male cell lines from the CCLE, related to Figure 6

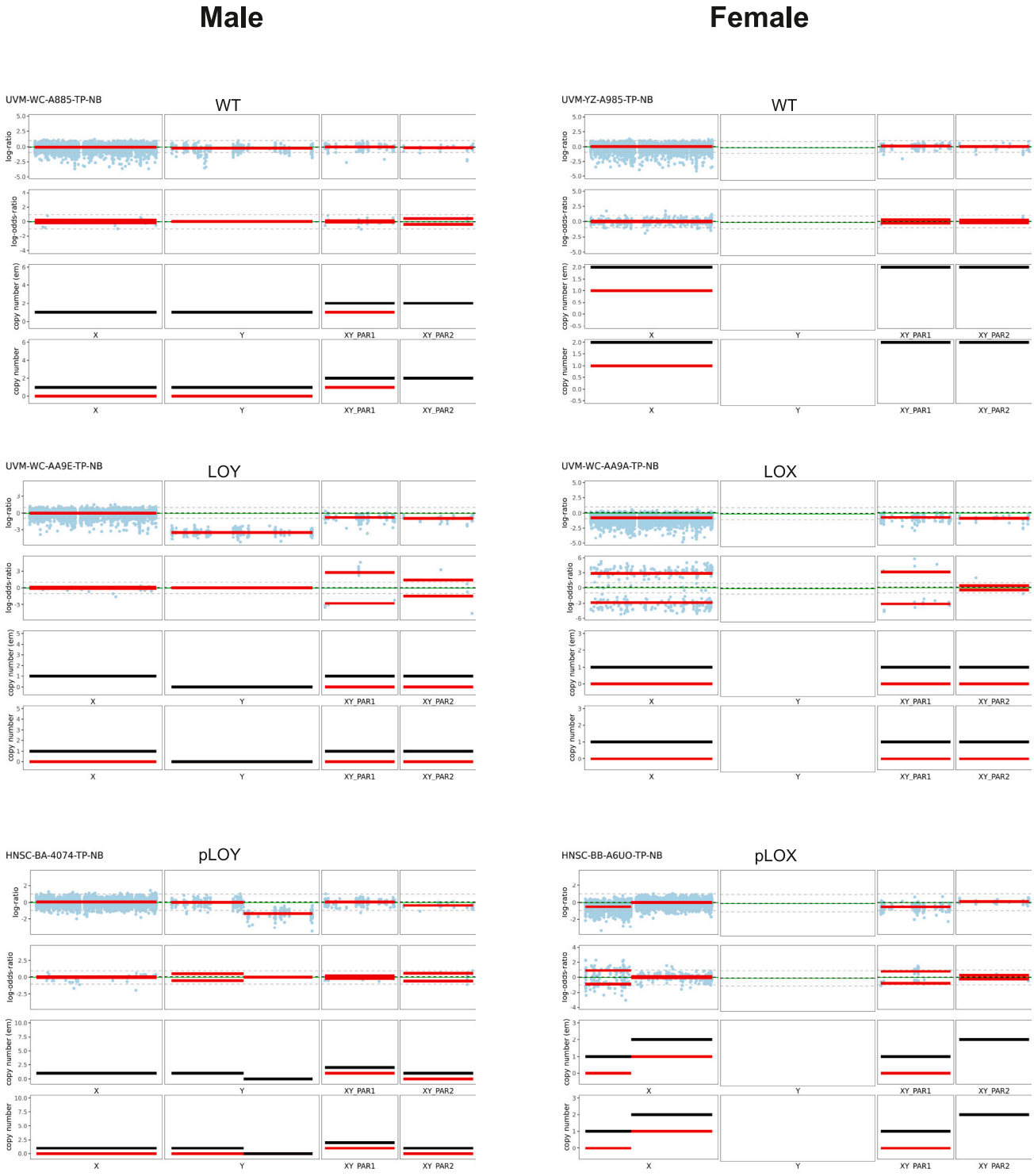


Figure S6. Representative examples for male and female sex chromosome copy number calls, related to STAR Methods
Left: male tumors with WT, LOY, pLOY events. Right: WT, LOX, pLOX in female tumors.