Clonal Lineage Tracing with Somatic Delivery of Recordable Barcodes Reveals Migration Histories of Metastatic Prostate Cancer

Ryan N. Serio[1], Armin Scheben[3], *Billy Lu[2], *Domenic V. Gargiulo[2], Lucrezia Patruno[4], Caroline L. Buckholtz[1], Ryan J. Chaffee[1], Megan C. Jibilian[1], Steven G. Persaud[1], Stephen J. Staklinski[3], Rebecca Hassett[3], Lise M. Brault[1], Daniele Ramazzotti[4,5], Christopher E. Barbieri[1,6], Adam C. Siepel[3, #], Dawid G. Nowak[1,2,7 #]

[1] Meyer Cancer Center, Weill Cornell Medicine, New York, NY, USA
[2] Department of Pharmacology, Weill Cornell Medicine, New York, NY, USA
[3] Simons Center for Quantitative Biology, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA
[4] Department of Informatics, Systems and Communication, University of Milano-Bicocca, Milan, Italy
[5] School of Medicine and Surgery, University of Milano-Bicocca, Milan, Italy
[6] Department of Urology, Weill Cornell Medicine, New York, NY, USA
[7] Division of Hematology and Medical Oncology, Department of Medicine, New York Presbyterian Hospital, Weill Cornell Medicine, New York, NY, USA

[*] These Authors Contributed Equally to This Work
[#] Corresponding Authors

Migration Histories of Metastatic Prostate Cancer

Corresponding Authors:

Dr. Dawid G. Nowak
413 East 69 Street, BB952-A01, New York, NY 10065
dgn2001@med.cornell.edu
(646)962-6169

Dr. Adam C. Siepel
1 Bungtown Rd, Cold Spring Harbor, NY 11724
asiepel@cshl.edu
(516)367-6922

The authors declare no potential conflicts of interest.

ABSTRACT

The patterns by which primary tumors spread to metastatic sites remain poorly understood. Here, we define patterns of metastatic seeding in prostate cancer (PCa) using a novel injection-based mouse model — *EvoCaP* (Evolution in Cancer of the Prostate), featuring aggressive metastatic cancer to bone, liver, lungs, and lymph nodes. To define migration histories between primary and metastatic sites, we used our *EvoTraceR* pipeline to track distinct tumor clones containing recordable barcodes. We detected widespread intratumoral heterogeneity from the primary tumor in metastatic seeding, with few clonal populations (CPs) instigating most migration. Metastasis-to-metastasis seeding was uncommon, as most cells remained confined within the tissue. Migration patterns in our model were congruent with human PCa seeding topologies. Our findings support the view of metastatic PCa as a systemic disease driven by waves of aggressive clones expanding their niche, infrequently overcoming constraints that otherwise keep them confined in the primary or metastatic site.

STATEMENT OF SIGNIFICANCE

Defining the kinetics of prostate cancer metastasis is critical for developing novel therapeutic strategies. This study uses CRISPR/Cas9-based barcoding technology to accurately define tumor clonal patterns and routes of migration in a novel somatically engineered mouse model (*EvoCaP*) that recapitulates human prostate cancer using an in-house developed analytical pipeline (*EvoTraceR*).

INTRODUCTION

The prognosis of prostate cancer (PCa) is mainly determined by the presence of metastases, with the five-year survival rate of metastatic PCa at only 36.6%, compared to nearly 100% for organ-confined disease (RRID:SCR_006902). Patients with PCa are at the highest risk of developing bone metastases amongst all cancer types, occurring in 90% of metastatic PCa cases (1), followed by visceral sites, such as the lungs (46%) and liver (25%), and also to lymph nodes (LN; 45%) (2). Lethal metastases often arise from rare, elusive subclones of heterogeneous tumors, making them difficult to treat conventionally (3).

Metastatic subclones may be inconsequential in forming the bulk tumor yet are highly invasive, possessing the capacity to seed distal organs and lie dormant over extended periods of time before expansion (3-7). Tumor clones have differing propensities for cell dissemination, colonization, and metastatic growth, which affect their capacity to metastasize (8, 9). Cancer cells may disseminate not only from the primary tumor to separate organs, but also from other secondary sites in a metastatic cascade (10-12). Additionally, a xenograft model demonstrated that primary tumors may be re-colonized by metastatic clones (13). Furthermore, metastatic seeding can occur bidirectionally between two separate organs (9).

Studying the evolution of PCa metastases is challenging due to several factors: *(i)* a lack of longitudinal clinical samples from the same patients with different metastatic sites; *(ii)* the confounding effect of PCa therapies on the trajectory of disease; and *(iii)* a lack of the appropriate models to delineate seeding patterns of metastatic cells. Resolving routes that cancer cells traverse in the body has conventionally been a difficult task but recent lineage tracing strategies, in particular using CRISPR/Cas9 technology (14-18), are opening new channels for probing directionality of metastatic spread (9, 19, 20). In recent years, more sensitive tools to trace tumor evolution in experimental models have been designed (6, 7, 9, 19, 20). Here, we adopt a Cas9-based recordable barcode strategy to determine clonal evolution and metastatic seeding patterns using our *in vivo* lineage tracing system.

In this work we demonstrate the applicability of our newly generated somatically engineered mouse model (SEMM) (21), *EvoCaP* (*Evo*lution in *Ca*ncer of the *P*rostate), where

genetic alterations are induced postnatally in the prostate using lentiviral injections. We initiate PCa in our model by co-deleting *Pten* and *Trp53* genes, a combination that is often enriched in patients with metastatic disease (22, 23). This strategy robustly induces an aggressive, metastatic pathology, recapitulating human PCa by derivation from a primary focal lesion and metastasis to relevant sites, including bones, LN, and visceral organs such as liver and lungs (22, 24). Inclusion of a recordable barcode (BC10) that accumulates heritable edits over time enables the tracking of cancer evolution over disease-relevant timescales under normal physiologic selection pressures during the life cycles of mice with intact immune systems.

Using quantitative analysis of cancer clonal populations (CPs) in the *EvoCaP* model *via* our *EvoTraceR* R package, we found a high degree of clonal heterogeneity, with many shared Cas9 edits (truncal mutations), defining founder CPs between the prostate and metastatic sites. Subsequent barcode edits then define subclonal populations within each CP. Seeding of distal sites occurred from the primary tumor or from clones in other metastases. Global metastasis-to-metastasis seeding was a less common source of metastatic fractions by comparison to cells that reached the site of destination directly from the primary tumor. We detected one instance of a rare primary tumor re-seeding, defined as bidirectional movement of a clone between the prostate and a metastatic site. We did not, however, observe metastatic re-seeding between secondary sites on a subclonal level. Most cells remained in a colonized organ after seeding, often lying potentially dormant, considering the relatively small size of several metastatic lesions. These patterns closely resemble the trajectories observed in human metastatic PCa (10, 25). Taken together, our experimental data suggest that a small subset of aggressive, dominant clones are most likely to spread to secondary sites, utilizing both the organ of origin and, to a lesser extent, additional metastases, as starting points en route to distant sites.

RESULTS

EvoCaP — a somatically engineered mouse model enabling lineage tracing.

Metastatic PCa regularly harbors copy number alterations (CNAs) (4, 26, 27), with *PTEN*/*TP53* among the most frequently co-deleted gene pairs. We generated a conditional *Pten*$^{loxP/loxP}$; *Trp53*$^{loxP/loxP}$ double knockout SEMM, with Cre-inducible Cas9-eGFP inserted in the *Rosa26* locus (*R26*) (28). Introduction of a lentivirus (LV) expressing Cre recombinase and firefly luciferase (FLuc) postnatally leads to: *(i)* focal loss of *Pten* and *Trp53* genes; *(ii)* activation of Cas9-eGFP; and *(iii)* delivery of a synthetic, recordable barcode (BC10) into the prostate (LV.GECPL.BC10) **(Fig. 1A-B)**. A marking guide (MG) then targets Cas9 to the barcode **(Fig. 1B-C)**. Once integrated into the mouse genome, BC10 becomes cumulatively edited over time, causing heritable modifications that enable the tracing of CPs within and between tissues as the cells harboring the barcode metastasize.

Barcode edits can be efficiently tracked with high sensitivity in scarce cell populations.

We first tested our lentivirus in mouse embryonic fibroblasts (MEFs) derived from mice carrying the *Pten*$^{loxP/loxP}$; *Trp53*$^{loxP/loxP}$ genotype. Cells were cultured over 28 days to investigate Cas9 editing capacity and efficiency **(Fig. 1D)**. Loss of Pten/p53 proteins coincided with the expression of markers for plasmid integration (Cre, eGFP) at day 7 compared to non-transduced cells **(Fig. 1E)**. Growth of eGFP^{+} cells showed similar rates between lentivirus-transduced groups up to day 28 **(Fig. 1F, Supplementary Fig. S1A-B)**, whereas non-transduced MEFs expire between 12-14 days. Cells transduced with lentivirus containing either MG or a non-marking guide (NMG), a non-editable control, also showed comparably high positive bioluminescence (BL^{+}) from FLuc after 28 days in culture **(Supplementary Fig. S1C)**. Together, these results validate the technical efficiency of our system *in vitro*.

In a SEMM where PCa is frequently instigated from small, focal lesions, and micrometastases are distinguishable by fluorescence, the capacity to detect diverse CPs from a small amount of starting material is vital. BC10 editing displayed similar editing patterns in DNA samples from 10^3, 10^4, 10^5, and 10^6 cultured MEFs harvested at day 28 after initial sorting for eGFP$^+$ cells at day 3 **(Supplementary Fig. S2A-B)**. Editing analysis of the sample containing 10^3 cells maintained a Jaccard index of 0.75 compared to the sample containing 10^6 cells, demonstrating a high concordance in edits detected between the samples despite a three order of magnitude difference in the number of cells analyzed **(Supplementary Fig. S2B-C)**. The *EvoCaP* system is therefore equipped for detecting BC10 edits *in vivo*, including in micrometastases (29), which represent a small fraction of total cells in a given organ, and for discerning scarce yet aggressive clones as they populate organs and spread to local or distal sites.

EvoCaP mice develop traceable metastatic lesions in multiple secondary sites.

We tracked PCa evolution over the lifetimes in immunocompetent mice after injecting BC10-containing lentivirus into the prostates of 12-week-old mice. Tumor progression was longitudinally monitored by BL$^+$ cells expressing FLuc, up to week 60 post-injection, when organs were collected from sacrificed mice for downstream analysis **(Fig. 2A)**.

The *EvoCaP* SEMM is a tunable system amenable to lentiviral dosage adjustments without the limitations of an inducible CreER system (30-32). Infection rates can thus be controlled, unlike in traditional Pb-Cre4 models where majority of cells in prostate receive a specific modification (33, 34). We functionally titrated LV.BC10.MG to scale dosage using $1x10^5$ and $2x10^5$ fluorescence forming units (FFU). Disease penetrance based on BLI by week 24 was 36.4% (8/22) for mice injected with $1x10^5$ FFU, and 62.5% (10/16) for the cohort receiving $2x10^5$ FFU. Therefore, we divided our cohort into groups based on low penetrance (*EvoCaP*-LP) and high penetrance (*EvoCaP*-HP), with *EvoCaP*-HP serving as a more efficient tool for exceeding the threshold for PCa tumorigenesis.

In total, we diagnosed primary PCa in 18 mice (8-*EvoCaP*-LP; 10-*EvoCaP*-HP), based on eGFP$^+$ signal in BL$^+$ mice, excluding mice expressing only BL but no eGFP in the dissected prostate (n = 2, 1-*EvoCaP*-LP; 1-*EvoCaP*-HP). Most tumor-bearing mice (7/8, 87.5% — *EvoCaP*-LP; 9/10, 90% — *EvoCaP*-HP*)* progressed to metastatic disease and were further evaluated. Mortality rates for mice monitored for up to 60 weeks (n = 16) trended higher before week 60 in the HP cohort (6/9; 67%) compared to the LP mice (2/7; 29%), but the difference was not significant **(Fig. 2B)**. Injected mice demonstrated a basal evolution of metastatic disease over long temporal periods after initial tumorigenesis **(Fig. 2C)**. Increases in distal BL$^+$ were indicative of distant metastases that were evident in late disease progression, particularly ≥ week 48 **(Fig. 2D)**, consistent with slow growth and spread as PCa evolves from an indolent to aggressive disease. Importantly, both *EvoCaP*-LP and *EvoCaP*-HP cohorts developed similar patterns of BLI intensity with time but with LP having lower disease penetrance. It suggests that possibly initial number of infected and surviving cells were different between cohorts, also confirmed with data coming from early weeks post-injection **(Fig. 2D)**.

Tumor-bearing mice exhibited overlapping BL$^+$ and eGFP$^+$ in dissected organs in high concordance with Cre detection in DNA at the time of harvest, demonstrating a high degree of immune evasion **(Fig. 2E-F, Supplementary Fig. S3A-K, Supplementary Fig. S4A-E).** Tumor cells derived from prostate epithelium were eGFP$^+$ and Pten$^-$ after Cre-mediated gene excision **(Supplementary Fig. 5A-B)**. eGFP signal overlapped with cytokeratin 8 (Ck8) **(Supplementary Fig. 5A-B)**, indicating prostate luminal epithelial cell origin, and was exclusive of stromal cells (α-SMA$^-$) **(Supplementary Fig. 5A, C)**. Taken together, these data show the robustness of the *EvoCaP* SEMM for identifying *Pten*/*Trp53*-loss cells in primary and secondary sites with minimal silencing of foreign transgenes by the immune systems of injected mice.

The destinations of metastases faithfully replicated human PCa spread (35) based on the locations of eGFP$^+$ signal **(Fig. 2F)**. Based on fluorescent (eGFP$^+$) microscopy analysis of mouse tissues across both cohorts, initial focal primary disease progressed to local seminal vesicle invasion (43.8%) and distally to other sites, with eGFP$^+$ lesions identified and isolated from the bones (92.9%), liver (37.5%), lungs (33.3%), LN (18.8%), oral cavity (7.7%), and bladder (6.3%). **(Fig. 2F)**. Bone metastases have been difficult to model in GEMMs (36) due to the long dormancy period between seeding of tumor cells in bone and expansion into overt metastases (37). However, BL$^+$ combined with eGFP$^+$ samples enabled the discovery of multiple bone micrometastases for downstream analysis of seeding patterns **(Fig. 2E-F)**. Notably, 6/9 (66.7%) of *EvoCaP*-HP mice developed liver metastasis, compared to zero mice in the *EvoCaP*-LP cohort, with only one mouse with liver metastasis surviving for the full 60 weeks **(Fig. 2F)**. This is consistent with liver metastasis signifying aggressive, lethal disease (38). *EvoCaP* thus provides a robust platform for analyzing PCa progression, organ seeding and colonization during the extended periods appropriate for investigating cancer evolution *in vivo*.

Barcode tracing reveals connections between clonal populations.

BC10 was engineered as a synthetic array of ten 20-nucleotide Cas9 target sites (TS) followed by a protospacer adjacent motif (PAM) sequence and a 3-nucleotide spacer, separating target sites. We designed target sites in BC10 using a cutting frequency determination (CFD) score algorithm **(Fig. 3A, Supplementary Fig. S6A)**, flanked by sequencing adapters **(Supplementary Fig. S6B)** (39, 40). Target sites are ordered in BC10 by decreasing activity (1.0: maximum editing; 0.1: minimal editing). Heritable, Cas9-induced marks annotate CPs and allow phylogenetic reconstruction of unique amplicon sequence variants (ASV), from which detailed lineage relationships can be inferred using maximum parsimony.

To assess clonal architecture, we developed an R package, *EvoTraceR*, that can summarize all truncal edits in BC10 by detecting edits from amplicon sequencing data and inferring clonal trees using algorithms implemented in the *Cassiopeia* suite (41). Using simulations, we showed that *EvoTraceR* accurately recovers induced mutation patterns and reconstructs ASVs **(Supplementary Fig. S7A)**. Negligible rates of alignment errors were recovered with *EvoTraceR* based on 100 simulations each of mutation rates of 10-TS simulated barcode sites **(Supplementary Fig. S7B)**. F1 score, a measure reflecting the harmonic mean of precision and recall parameters, was comparable for *EvoTraceR* (>0.9) with established methods designed primarily for use with one TS (42, 43), despite the required scaling to ten TS in the *EvoCaP* model **(Fig. 3B)**.

The use of *EvoTraceR* in lineage tracing analysis enables the identification of truncal edits, or shared ASVs that define a unique CP from which a genealogical history can be described based on subsequent edits that may be shared between ASVs (subclonal populations) or occurring in a single ASV only (terminal subclone) **(Fig. 3C)**. Migration paths are then called using *MACHINA* (*M*etastatic *a*nd *C*lonal *H*istory *In*tegrative *A*nalysis) (44) with high accuracy at low migration frequency **(Supplementary Fig. S7C)**. Using the information obtained from the *EvoTraceR* pipeline, it is possible to determine the frequency of ASVs in different organs, the clonal architecture, and the routes of dissemination for clonal and subclonal populations with high sensitivity through shared editing patterns **(Fig. 3D)**.

Primary and metastatic tumors share truncal edits that define clonal populations and enable phylogenetic tree construction.

Metastases are difficult to study due to their separation in time and space from the primary tumor. A high degree of clonal heterogeneity adds another layer of complexity in elucidating migratory patterns of metastatic cells. Precisely resolving the connections of individual CPs between different organs would thus illuminate the dynamics of tumor spread. Toward this endeavor, we annotated all marks in ASVs compared to the unedited BC10, tallied frequencies,

and assigned them to the organs from which their CPs were isolated based on truncal mutations using our *EvoTraceR* package **(Supplementary Fig. S8A)**. We noted a median of 29.5 CPs per mouse (29, *EvoCaP*-LP; 30, *EvoCaP*-HP; range = 10-74), which is consistent with ~10-50 cells initially infected previously using our injection-based technique (45).

We compared the editing efficiency of our evolving barcode *in vivo* in two *EvoCaP*-HP mice that were imaged for BL^+ to track cancer progression **(Supplementary Fig. S9A-B)**. MMUS1458 developed a primary tumor before its sacrifice at week 24 post-injection. We monitored a second mouse for double this amount of time, until its expiration at week 48 (MMUS1457). Upon sequencing, we identified 21 ASVs, excluding the non-marked barcode, in MMUS1458, constituting just 2.2% of barcodes **(Supplementary Fig. S9C-D)**. In contrast, MMUS1457 displayed 194 total ASVs, with an 18.5% saturation rate in the prostate **(Supplementary Fig. S9E-F)**. Whereas the majority (2.1%/2.2%) of ASVs in MMUS1458 were singletons with one edit only, 56% of ASVs in MMUS1457 (7.5%/13.4%) were part of a CP of 2+ edits **(Supplementary Fig. S9G-H)**. We therefore observed a gradual accumulation of BC10 edits over time *in vivo* without early saturation of the barcode that would prevent accurate lineage reconstruction.

We first focused on the most information-rich mouse with the most unique lineages, MMUS1495 (CP = 74), for further scrutiny. MMUS1495 succumbed to PCa at week-52 post-injection with aggressive metastasis to bone and liver. To analyze regions containing BC10, we first dissected $eGFP^+$ tissue from mouse organs that progressed to metastatic PCa (MMUS1495) **(Fig. 4A)**. We then extracted and sequenced genomic DNA containing integrated plasmid from the prostate left lobe (PRL), right and median liver lobes (LVR, LVM), and left rib (RBL), to assess whether Cas9-induced marks were heritable and shared between primary and metastatic sites *in vivo*. An accumulation of alterations, particularly deletions, was observed in ASVs from different organs **(Fig. 4B)**. Most marks in BC10 were short (<30 bp deletions, 1-6 bp insertions) **(Fig. 4C)**. We did not observe many large deletions that potentially could create overlapping, nested marks that would complicate analysis of genealogies **(Fig. 4C)**.

Tumor clones can be characterized by various metrics, including: *(i)* ASV richness (*S*); *(ii)* Shannon's Entropy (*H*); *(iii)* Pielou's evenness (*J'*) and (*iv*) Bray-Curtis dissimilarity (BC_{jk}). We subjected all CPs from MMUS1495 to eco-statistical analysis **(Supplementary Fig. S10A)**. These metrics allow us to track clonal architecture, changes in heterogeneity and continuity between primary tumors and metastases. The total number of ASVs (*S*) was highest in the liver lobes and lowest in the primary tumor **(Supplementary Fig. S10B)**. This is consistent with increasing editing diversity as tumor cells spread to different sites, where cells would continue to expand in their new niche and acquire further edits, once they overcome any microenvironmental constraints that would result in dormancy. It would thus be expected that heterogeneity is greater in macro-metastatic sites than in the primary tumor. Intra-organ (*S*) heterogeneity is indeed both lower in the prostate than in the three metastatic sites **(Fig. 4D)**, and *J'* is lowest in the prostate, indicating greater evenness of clonal landscape based on the abundances of identical clones in metastases **(Supplementary Fig. S10C)**. Edits are evenly dispersed amongst organ sites without approaching saturation **(Supplementary Fig. S10D-E)**. Inter-organ heterogeneity is highest in the median liver lobe sample, indicating furthest evolutionary divergence in the cells colonizing this site compared to cells in the primary tumor **(Fig. 4E)**.

The first three CPs in MMUS1495 (CP01, CP02, CP03) were composed of 33-74 distinct ASVs compared to non-marked BC10, generating diverse subclones displaying high heterogeneity (*H*) with high total ASV frequency and above average dispersal scores (9) **(Supplementary Fig. S10A)**. Because they contained the highest ASV richness, these clones were further categorized genealogically. We constructed phylograms based on three parameters: *(i)* truncal edits, which define a given founder CP; *(ii)* subsets of shared edits along with the truncal

modifications (branched subclonal populations); and *(iii)* edits unique to an individual ASV (private edits) **(Fig. 4F-H)**.

Barcode tracing enables identification of migration topologies between primary prostate tumors and metastases.

A tumor clone has several possible fates as it evolves through additional adaptations that exhibit high metastatic potential (9), assuming it does not undergo primary tumor extinction. We focused in our analysis on specific topologies between primary and metastatic sites by applying *MACHINA* software (44) to infer migration histories of metastatic cells. Using *MACHINA*, we integrated phylogenetic data for each CP and its ASVs as well as the organ-specificity of each ASV. *MACHINA* infers parsimonious migration trajectories by jointly resolving migration routes and ambiguous polytomous branches in the phylogeny for the heterogeneous populations characteristic of cancer clones (44).

Ample barcode resolution enables tracking the degree to which a tumor clone expands within a specific organ versus its frequency of disseminating and colonizing another site. We first analyzed migration patterns of each CP identified in MMUS1495. If metastases resulted mainly from the linear spread of a dominant clone in the primary tumor, we would expect to see few shared routes between target organs. However, we instead observed a broad distribution of founder clones between the four tissues **(Fig. 5A-C)**. Such a high occurrence of founder clones in distal regions likely indicates early and probable polyclonal migration early during tumor formation.

Though metastasis-to-metastasis seeding was present, it was a rare event, occurring in just 2.9% of all transitions. Most (1.8%) metastatic seeding events occurred in parallel from one secondary tissue to two additional sites, mainly originating from clones that colonized the right liver lobe **(Fig. 5D, Supplementary Fig. S11A)**. We focused on the first three CPs as they exhibited the greatest ASV diversity. Metastasis-to-metastasis seeding occurred in each of CP01, CP02, and CP03, albeit at a low frequency **(Fig. 5A-C)**. Transition matrices show the source and directionality of seeding, both in the entire mouse when all CPs are considered **(Fig. 5D)**, and in individual CPs **(Fig. 5E-G)**. In both CP01 and CP02, seeding occurred from the prostate to one metastatic site, which was the right liver lobe (LVR) in CP01 and the left rib (RBL) in CP02. From the secondary site, metastatic seeding occurred to additional sites in parallel, including the median liver lobe in both cases **(Fig. 5E-F)**. Considering the median liver lobe was the terminal metastatic site in most clones with metastasis-to-metastasis seeding, its furthest evolutionary divergence from the prostate as inferred by its highest BC_{jk} score is further supported by migration trajectory analysis. Similar to CP01, metastatic parallel seeding occurred in CP03, but into the RBL and LVM from the LVR, illuminating the high metastatic potential of clones that initially seeded the right liver lobe. Bidirectional spreading of tumor cells was apparent between different organs (RBL, LVR) but not by cells that were a part of the same CP. A third primary migration route was undertaken by CP03, in which two metastatic sites were seeded in parallel by the primary tumor **(Fig. 5C, 5G)**. Hence, despite occurrence of the three CPs in all tumor sites, we delineated three unique migration paths taken by each clone.

Clonal architecture and dynamics were next compared. Performing MACHINA analysis on ten mice pooled from both the *EvoCaP*-HP and *EvoCaP*-LP cohorts enabled us to investigate routes and trajectories from a potentially wider distribution of phenotypes. We sampled the seeding topologies of all CPs in sequenced mice and quantified each intra-organ expansion (“Primary Confined” and “Metastatic Confined”) and migratory seeding event (“Primary Mono-Seeding,” “Primary Parallel Seeding,” “Primary Re-Seeding,” “Metastatic Mono-Seeding,” “Metastatic Parallel Seeding” and “Metastatic Re-Seeding”) **(Fig. 6A-B)**. Primary seeding events, from mono-seeding and parallel transitions, were less common in all mice compared to prostate confinement when evaluating the primary tumor, with most clonal expansion events occurring within the prostate **(Fig. 6C-D)**. Confinement of a CP to a single organ, whether primary or

metastatic, dominated the overall clonal architecture similarly in both cohorts (LP: 86.1-94.1%; HP: 80.4-94.5%), with inefficient cumulative metastatic spread from the prostate **(Fig. 6C-D, Supplementary Fig. S11B-C)**. Metastasis-to-metastasis seeding occurred in 60% of mice (3/5, LP; 3/5, HP), but was an uncommon event at a frequency of ≤2.9%. Metastatic re-seeding of the primary tumor was rare, occurring in just one CP (CP01) of a single mouse (MMUS1492; HP), where we inferred a migration from the lymph node back to the prostate. Although we found that the liver had the highest mean clonal heterogeneity based on *H*, we did not find significant differences in clonal heterogeneity between tissues **(Supplementary Fig. S12A)**. Because visceral metastasis to lungs and liver results in poor prognosis (38), we assessed whether routes of spread between viscera and non-viscera (LN, bone) were stochastic or non-random. We discovered no preferred route between visceral and non-visceral tissues, although most primary seeding events were expectedly to non-visceral sites **(Supplementary Fig. 12B-C)**. More seeding of visceral tissue occurred in *EvoCaP*-HP samples (41.4%) compared to those of *EvoCaP*-LP (27.4%) **(Supplementary Fig. S12B-C)**, likely due to the preponderance of liver metastases in this cohort. In summary, by analyzing 10 mice with metastases, we found that these patterns of seeding were broadly consistent across individuals, with metastatic seeding, if present, occurring as a fraction of primary seeding.

The calling of migration pathways in metastatic prostate cancer are consistent between mouse and human datasets.

To test how closely the calling of migratory seeding patterns detailed in *EvoCaP* mice with metastatic PCa reliably correspond to human PCa migration patterns, we performed *MACHINA* analysis on a public human dataset first analyzed by Gundem *et*. *al*., (10) and later scrutinized using *MACHINA* (44). Seeding events delineated from barcode editing in our SEMM were comparable to seeding topologies in four human PCa samples that we re-analyzed using *MACHINA* and our topology definitions. Seeding topology frequencies were similarly distributed between the groups of human, LP mice, and HP mice. In human dataset analyses, primary seeding (mono-seeding and parallel seeding) occurred at a mean of 13.3%, and metastatic seeding was either absent or present in low abundance ≤10% **(Fig. 6E)**. Primary seeding occurred at slightly higher percentages in mice (LP_{Mean} = 8.2%, HP_{Mean}= 10.7%) **(Fig. 6C-D)**. No re-seeding of either the primary tumor or a metastatic site that served as a tumor reservoir for another metastasis was noted in human data (0/4), and was detectable in just one of the ten mice sequenced, highlighting the rare nature of this seeding trajectory **(Fig. 6D)**. In both the human and mouse sample sets, a small fraction of cancer cells metastasized from the prostate, and a smaller percentage migrated from a metastatic site (when applicable), compared to the total frequency of clones confined in their host organ, demonstrating that metastatic spread is a rare event. Furthermore, more clonal expansion occurred within metastatic sites ("Metastatic Confined") compared to the primary tumor in 6/9 mice and 2/3 human tumors with multiple metastases, likely due to an aggressive proliferating phenotype of metastatic clones compared to those that remain confined to the primary organ.

Our results identified probable metastasis-to-metastasis seeding in two (A31, A32) of four patient datasets we analyzed **(Fig. 6E)**, consistent with the conclusions reached by the original authors using Bayesian methodology (10). Our own analysis was in agreement with previous re-analysis of patient dataset A10 using *MACHINA* (44), concluding that a plausible alternative pattern to metastasis-to-metastasis seeding was most likely. By combining the migration path-building function of *MACHINA* with a robust phylogenetic tree-building approach, *EvoTraceR* accurately predicts cancer migration trajectories in DNA sequenced from mouse tumor samples, with the migratory seeding patterns in mice using our *EvoCaP* SEMM comparable to those observed in humans with metastatic PCa **(Fig. 6F)**.

Overall, the identification of CPs with their specific ASVs that make the tumor susceptible to clonal expansion and migration demonstrates the ability of our lineage tracing technology to

reveal the metastatic properties of prostate tumors at adequate subclonal resolution. Together, these data reveal an environment whereby distinct tumor clones evolve metastatic and migratory potential to systemically populate permissive sites, overcoming the major immune and microenvironmental barriers that normally hinder tumor cell spread (46). Despite polyclonal primary seeding, metastasis is an inefficient process in the context of all analyzed CP transitions **(Fig. 6C-F)**. Invasive capacity is largely retained in the initial founder clones, suggesting that *Pten*/*Trp53*-loss genetics, a dominant co-deletion occurrence in human PCa, intrinsically primes tumor cells for metastatic spread from an early stage. Spread from one metastatic site to another is an infrequent event, and largely limited to specific dominant subclones exhibiting the greatest capacity for clonal expansion.

A variety of prostate cancer-associated genotypes may be investigated using evolvable barcodes.

Introducing different starting gene combinations could lead to distinct trajectories as different genes might influence adaptations that alter the spread of metastatic cells. We sought to adapt our lentiviral construct for modeling alternative genotypes by including an additional targeting guide driven by the hH1 promoter **(Supplementary Fig. S13A)**. *EvoCaP* MEFs were transduced and cultured for 28 days to investigate differences in BC10 editing and the emergence of ASVs in more aggressive PCa genotypes such as *Pten/p53/Rb1*-loss and *Pten/p53/Smad4*-loss **(Supplementary Fig. S13B)**. Cas9-targeted deletion of *Rb1* and *Smad4* corresponded with eGFP expression and Pten/p53 loss **(Supplementary Fig. S13C-E)**, with an increasing diversity of edits over time **(Supplementary Fig. S13F)**. Alpha- and beta-diversity measures both increased over time **(Supplementary Fig. S13G-I)**. Comparison of all genotypes juxtaposed revealed that the 2PR MEFs clustered apart from 2P and 2PS samples in beta-diversity successively from day 7 to day 28 **(Supplementary Fig. S13J)**. Based on *EvoTraceR* analysis of additional genotypes associated with PCa, our model predicts that loss of *Rb1* in addition to *Pten*/*Trp53* co-deletions may lead to increased tumor heterogeneity, which could result in more aggressive pathology that could be amenable to further dissection of clonal and migration dynamics by barcode-based lineage tracing strategies *in vivo*.

DISCUSSION

The organ site of metastases impacts patient overall survival (38), making the dynamics of PCa spread to secondary sites critical to discover new avenues for therapeutic interventions by targeting mechanisms of metastatic spread. Unlike traditional modes for PCa monitoring that are insufficient for detecting small lesions (47), the *EvoCaP* SEMM enables inference of precise distribution of metastatic cells even in the absence of overt macro-metastases, as small lesions can be detected post-mortem by bioluminescence and fluorescence, and analyzed using our lineage tracing methods. Adjusting viral titer provides tunability by altering the threshold for tumorigenesis without the limitations of other inducible systems (32, 48). Adjusting tamoxifen dosage to control CreER nuclear translocation may require repeated administration, cause drug-related adverse events, result in different effects in disparate cell types, or lead to incomplete knock-out in some cells (30-32). Furthermore, doxycycline administration in a Tet-inducible system may confound results by affecting cellular proliferation in the prostate (48). Despite using identical starting tumor drivers, we enhanced phenotypic variability in the *EvoCaP*-HP cohort by increasing viral titer. This modification increased tumor penetrance and established liver metastasis, without substantially affecting tumor focality and growth or BL^+ intensity over time **(Fig. 2B-F)**. Clonal architecture and seeding topologies were consistent between cohorts **(Fig. 6C-D)**, denoting comparable tumor dynamics.

Our system is designed to minimize Cas9 and eGFP expression by using only one allele in the *R26* locus. Expression of foreign transgenes can activate immune surveillance (49). Our

transgenic markers are stable for up to 60 weeks and are not silenced during disease progression **(Supplementary Fig. S3)**. The concordance between BL^+ and $eGFP^+$ suggests that cancer progressed with colonization of multiple organs, thereby largely avoiding immune clearance, and demonstrating the robustness of *EvoCaP* in modeling metastasis.

The inclusion of the BC10 reporter in our system provides the capacity to dissect spatiotemporal patterns of metastatic seeding. We designed the *EvoTraceR* R-package-based pipeline to specifically analyze lineage tracing patterns and resolve clonal architecture from the DNA sequences of mouse tumor cells containing integrated lentivirus. Appreciable barcode resolution was noted, as CPs enriched with multiple ASVs were resolved through additional marks following the initial truncal edit, denoting the emergence of subclones from a founder CP. Several clones in disparate metastases shared common genealogy with the original tumor. Most clonal founders were widespread throughout all metastases **(Fig. 5A-C, Supplementary Fig. S11)**, indicating early spread from multiple clones, though dormancy may prevent overt colonization for some time (36). Metastases derived from clusters of clonal aggregates disseminating in unison have been previously observed using multicolored reporters in breast and pancreatic cancers (50-52). Such circulating clusters are associated with a high metastatic seeding capacity (53). Thus, it is possible that distinct CPs clustered and simultaneously arrived in the organ.

Deactivation of *PTEN/TP53* genes represents one of the most commonly co-occurring gene deletions in human metastatic PCa (4, 26). We identified ecosystems wherever metastases seeded that were composed of founder clones that formed the initial prostate tumor **(Fig. 5, Supplementary Fig. S11B-C)**, consistent with metastasis being driven primarily by bulk primary tumor drivers (54). We propose that the combination of *Pten*/*Trp53* loss primed tumors for metastatic disease by igniting the conditions required for punctuated evolution of tumor progression, facilitating early seeding of distant sites wherever a permissive niche presented itself, followed by outgrowth of these colonies that largely maintained the presence of founder CPs. Heterogeneous clonal expansion in the primary tumor with early, polyclonal metastatic seeding contrasts with models whereby metastases arise from subclonal selection and expansion of a single pro-metastatic clone within the primary tumor (20, 55). Sources of these differences may include advanced disease genotype (*Pten*/*Trp53*-loss) and multifocality of tumor initiation to model a more aggressive stage of PCa progression with the *EvoCaP* platform. Genotypes with a less severe prognosis may therefore present with different patterns of evolution that may instead favor a scenario where one dominant subclone in the primary tumor drives all metastasis (25). The widespread metastatic seeding by multiple CPs we observed could indicate why almost half of metastatic PCa patients present with *PTEN/TP53* co-deletions (4, 26).

Most tumor cells remained confined within an organ, possibly dormant in the case of micrometastases **(Fig. 6D-E)**. This behavior could be partially consistent with the "go or grow" hypothesis, claiming that invasion/migration and proliferation show rather mutually exclusive spatiotemporal patterns (56), with cell cycle arrest preceding the acquisition of invasive phenotype (57, 58). We detected metastasis-to-metastasis seeding by a small number of progeny, which is consistent with human PCa reports that show seeding between metastatic sites after the initial metastatic events (9, 10, 25, 59). These were infrequent events when taking all CPs into consideration, revealing the importance of 1-2 dominant, highly invasive clones in facilitating metastatic spread. In contrast to dormant cells comprising micrometastases, clones that were able to colonize and then expand in metastatic sites are likely to account for most of the expansion and further spread of CPs that dominate large metastases, such as CP01-03 in MMUS1495 **(Fig. 5)**.

Alternative routes of spread include primary organ re-seeding by a secondary site and bidirectional seeding (9). We identified one case of prostate re-seeding (CP01; MMUS1492) and no bidirectional metastatic seeding. Detection of primary re-seeding by a metastatic clone indicates that our *EvoTraceR* pipeline is sufficiently sensitive to detect rare events. Still, the

possibility remains that the scarcity of alternative migration routes in our analysis is due to conservative maximum parsimony migration inference. Therefore, to further validate our analysis, we re-analyzed a human dataset previously analyzed using *MACHINA* (44) in which metastasis-to-metastasis seeding was validated in 5 of 8 patients that were originally identified as exhibiting metastasis-to-metastasis spread. Importantly, our analyses comparing our mouse datasets generated here *via* the *EvoCaP* platform, show similar patterns of metastatic spread to that of public human datasets that have been scrutinized previously by multiple methods of analysis. In each case, migration from one site to another was uncommon, as clones were predominantly confined to the prostate or a secondary metastasis following their spread. Seeding from the primary tumor never exceeded 15.4% of inferred transitions in human datasets or 19.6% in mouse datasets. Metastasis-to-metastasis seeding likewise was uncommon, occurring in 67% of mouse (6/9) and human (2/3) cancers with multiple metastases, but at rates of ≤10% of all transitions. The similarities in distribution and metastatic seeding routes and trajectories between mouse and human datasets help to illustrate the robustness of our model and pipeline in accurately calling migration paths.

Slight alterations to our flexibly designed lentivirus would enable a broader range of biologic inquiries to be probed. For example, combining scRNA-Seq with a transcribable barcode would allow us to study clonal architecture more precisely in metastases and the surrounding milieu by analyzing gene expression corresponding to specific clones. Further, adding a unique molecular indicator would prevent the occurrence of homoplasy, the same mutation in different lineages, from obscuring evolutionary histories. Additionally, our lineage tracing data were limited to CPs exhibiting 3-4 unique edits per ASV, which may restrict analysis. Improving BC10 design by optimizing target sites and their surrounding sequences to facilitate editing efficiency will enhance tree depth, improve power to detect rare events, such as metastatic re-seeding, resolve inconclusive metastasis-to-metastasis seeding events (see human datasets A31 and A32) (10, 44), and possibly improve the reconstruction of metastatic seeding trajectories. Moreover, introducing agents driving greater evolutionary selection pressure, such as androgen deprivation therapy in PCa, may potentially alter migratory trajectories (60-62). Therefore, the *EvoCaP* platform can be a valuable tool in analyzing the mechanisms driving resistance to therapeutic interventions and the resultant evolution toward more aggressive, lethal forms of PCa resulting from failed therapies (61-64). Efficient delivery of Cas9 with guides against disease-associated genes directly into somatic cells of an organ of interest democratizes our model so it can be easily and readily reproduced and applied to a wide variety of cancers and other diseases amenable to somatic genetic manipulations.

In summary, barcode-mediated lineage tracing enables the construction of detailed phylogenies to probe mechanisms driving clonal diversity at the level of the whole organism rather than being restricted to the confines of the cancer origin. As a result, we demonstrated a scenario in which punctuated evolution characterizes PCa that is driven by the frequent gene pair deletions of *PTEN*/*TP53*, with dynamic spread induced by polyclonal seeding between tissue sites in the tumor microenvironment. We delineated the quantitative spread of specific cancer clonal populations with high sensitivity using our lineage tracing tool, resolving precise clonal topology and intratumoral heterogeneity. The most frequent topology we observed was clonal expansion within a tissue, with fewer seeding events from the prostate and atypical spread from a metastatic lesion to another site. Additionally, our model demonstrates similar patterns of metastatic seeding topologies to human disease. Hence, using our *EvoCaP* model, we can view PCa holistically by elucidating highly pathogenic clones in whatever region they arise and spread in the body after initial tumorigenesis.

METHODS

PLASMID DESIGN AND CLONING

Barcode plasmids were generated from parent pGECPL (Guide RNA-EF1a-Cre-P2A-FLuc) plasmids. We first designed pGECPL plasmids using a pECPV (EF1a-Cas9-P2A-Venus) backbone containing a gRNA scaffold upstream of the EF1a promoter sequence (65). We deleted Cas9-P2A-Venus using AgeI and BamHI restriction enzymes. We then inserted a printed Cre-P2A-FLuc construct flanked by an AgeI digestion site near the 5' end and a BamHI digestion site on the 3' end. The 323 bp Barcode sequence was PCR amplified with addition of EcoRI restriction sites using the following primers: F: 5'-TCGGAATTCAATGATACGGCGACCACCGA-3'; R: 5'-TCGGAATTCGTGACTGGAGTTCAGACGTG-3'. The PCR product was gel purified and cloned into the pGECPL.Filler vector using EcoRI restriction enzyme. Cut vectors were dialyzed for two hours prior to ligation and transformation. Incorporation of BC10 was validated by Sanger sequencing. The filler downstream of the hU6 promoter was removed by BsmBI (New England BioLabs, Ipswich, MA) digestion (3 hours). The guide specific to Barcode (marking guide — MG or g.Neg.Mm.01) used was 5'-TCTACACGCGCGTTCAACCG-3'. The negative control guide with no sequence specificity to either the BC or any locus in the mouse genome, or non-marking guide (NMG/g.Neg.Hs.Mm) used was 5'-GACCGGAACGATCTCGCGTA-3'. The MG and NMG were each cloned into separate pGECPL.BC10 vectors following PCR amplification to obtain the pGECPL.BC10.MG (experimental) and pGECPL.BC10.NMG (control) plasmids.

To create the p2GECPL "double guide" vector, a site downstream of hU6_filler was added with hH1 promoter sequence and a second filler. This filler was removed using PaqCI and guide RNA sequences against *Rb1* (5'-TGCGCGGGGTCGTCCTCCCG-3') and *Smad4* (5'-AGACGGGCATAGATCACATG-3') were cloned in using the same protocol as the removal of the BsmBI site, except with an added deactivation step required for PaqCI digestion.

WESTERN BLOTTING, ANTIBODIES, AND REAGENTS

The following antibodies and dilutions were used for Western blotting experiments: Pten (Sigma-Aldrich, Burlington, MA, Cat#04-035), 1:2,000; p53 (Cell Signaling Technology, Danvers, MA, Cat#32532S), 1:1,000; Cre (Cell Signaling Technology, Cat#15036S), 1:2,000; GFP (Cell Signaling Technology, Cat#2956T), 1:2,000; Rb1 (Abcam, Cat#ab181616), 1:1,000; Smad4 (Cell Signaling Technology, Cat#46535), 1:1,000; Actin (Sigma-Aldrich, St. Louis, MO, Cat#A3854, RRID:AB_262011), 1:20,000-1:30,000. For Western blotting, MEFs were harvested in 1x RIPA Lysis Buffer (EMD Millipore, Billerica, MA) with protease inhibitors (cOmplete, EDTA-free Mini®, Sigma-Aldrich) and phosphatase inhibitors (PhosSTOP®, Roche, Basel, CH), boiled, and resolved on SDS-PAGE gels. Band sizes of probed proteins were compared against a Precision Plus Protein™ Dual Color Standards Protein Ladder (Bio-Rad, Philadelphia, PA, Cat#161-0374), 10-250 kDa, containing ten recombinant proteins for molecular weight determination.

CELL CULTURE EXPERIMENTS

Primary mouse embryonic fibroblasts (MEFs) were derived from *Pten*$^{loxP/loxP}$; *Trp53*$^{loxP/loxP}$ embryos containing eGFP-Cas9 (2PECAS genotype). To obtain MEFs, we removed heads, limbs, and internal organs from embryos and minced the remaining portion for 4 minutes. We digested each minced embryo in 2 mL trypsin for 30 minutes at 37ºC, with mixing by pipet every 10 minutes. We added 10 mL of full DMEM (10% FBS, 1X pen/strep) and centrifuged at 1,200 RPM for 10 minutes before aspirating 10 mL of Trypsin/DMEM mixture and plating in 12 mL total DMEM, growing and expanding in cell culture for one week, and freezing in liquid nitrogen.

For cell culture experiments using single guide-containing plasmids, we transduced HEK293-FT cells (Thermo Fisher Scientific) with 1 μg/μL of each plasmid (pGECPL.BC10.NMG, pGECPL.BC10.MG), 1 μg/μL of each packaging construct (psPAX2, RRID:Addgene_12260; pMD2.G, RRID:Addgene_12259) and 1 mg/mL polyethyleneimine (PEI 25k, linear; Polysciences, Inc, Warrington, PA) (1:3, DNA/PEI). We changed medium after 6 hours and allowed cells to grow

for 72 hours total. We collected virus from supernatant, centrifuged for 5 minutes at 1,000 rpm, filter-sterilized, and diluted 1:1 in DMEM (2% FBS, 1x pen/strep) before infecting 2PECAS MEFs and changing medium after 8 hours. We analyzed fluorescence by microscopy (EVOS® FL Auto Imaging System, Thermo Fisher Scientific, Waltham, MA) prior to collecting cells for flow cytometry at days 3, 7, 14, and 28. Flow cytometry was performed using the Attune NxT Flow Cytometer (Thermo Fisher Scientific), with data obtained from side scattering taken for analysis. We collected 10^6 cells for DNA extraction at days 3, 7, 14, 21, and 28. Cells were harvested for protein analysis by Western blotting.

Transduction was performed similarly when comparing different genotypes using a double guide-containing plasmid, but instead using 1 *μ*g/*μ*L of p2GECPL.BC10.MG.NMG, p2GECPL.BC10.MG.g.*Rb1*.70, and p2GECPL.BC10.MG.g.*Smad4*.277, where the terminal number after the gene name represents the nucleotide cut site for Cas9. Cells were passaged from day 3 to day 7, and then once weekly every 7 days until day 28. A portion of cells (one million cells each) were collected for DNA extraction and protein extraction at days 7, 14, 21, and 28. All cell lines were routinely checked for Mycoplasma infection by semi-quantitative PCR analysis as previously described (66).

LENTIVIRUS PRODUCTION AND TITRATION

We routinely produce high quality lentivirus in bulk for *in vivo* injections. We transduced three 15 cm plates of HEK293-FT (RRID:CVCL_6911) cells seeded at 4 x 10^5 cells/mL for 72 hours with 1 mg/mL PEI and 1 μg/μL target, psPAX2, and pMD2.G plasmids at an optimal molar ratio of 1:1:1. Supernatants were centrifuged for 5 minutes at 1,000 rpm, 0.45 μm filter-sterilized, ultracentrifuged at 25,000 rpm for two hours with a 20% sucrose solution cushion, and resuspended in Ca^{2+} and Mg^{2+} free PBS. Each produced lentivirus was titrated using a functional titer assay based on fluorescence and flow cytometry (Attune NxT Flow Cytometer, Thermo Fisher Scientific), yielding arbitrary fluorescence forming units (FFU) allowing us to strictly control efficiency and quality of the virus. To perform the assay, a cell line was created from primary MEFs dissected from mouse embryos bred for heterozygous tdTomato transgene expression for a fluorescence readout (45). These “ET” MEF cells were seeded at 40,000 cells/mL and incubated with the serially diluted lentivirus for 8 hours. TdTomato (Texas red filter = 620/15 nm wavelength) fluorescence was measured *via* flow cytometry 72 hours after infection to determine FFU/μL titer of the lentiviral solution. We standardized our viral dosing scheme by later using Cre-Sensor cells, which we produced by infecting 22Rv1 (RRID:CVCL_1045) cells with the pLV-CMV-LoxP-DsRed-LoxP-eGFP (RRID:Addgene_65726) vector and puromycin selection, ensuring accuracy in titration and reproducibility between different cell types. Animals were injected with a total of 1x10^5 FFU and 2x10^5 FFU in 20 μL of lentiviral solution. Production of the lentivirus in one batch per plasmid was sufficient for injections of ~50 animals, ensuring that we reduce the probability of lentivirus batch effects in our models.

EVOCAP HUSBANDRY

Our research in mice was covered under animal use protocol (Mouse Models of the Genitourinary Cancers, 2018-0017) that has been approved by the Institutional Animal Care and Use Committee (IACUC) of Weill Cornell Medical College of Cornell University, where all research was conducted, using the Research Animal Resource Center (RARC). For these experiments, ~20× mice per experimental group were calculated to be necessary to achieve appropriate power analysis. Our cohort exceeded the number of animals for attaining statistical significance of $p<0.05$ with the 80% power and effect size equal to 1.25 (Cohen's *d*).

To produce our animal cohort, *Pten*$^{loxP/loxP}$; *Trp53*$^{loxP/loxP}$ mice were crossed with Rosa26-^{loxP}STOP^{loxP} *SpCas9*-eGFP (R26-Cas9-eGFP) mice. This generated mice that express Cas9 and eGFP from one *R26* locus following the addition of Cre from the lentivirus injection, which also

releases the loxP sites to simultaneously delete both *Pten* and *Trp53* genes. Only male mice could be used for experiments.

EVOCAP ANIMAL INJECTIONS

EvoCaP mice were injected on week 12 in the left anterior prostate with 20 μL of lentivirus containing Cre and the Barcode (pGECPL.BC10) and consisting of $1x10^5$ - $2x10^5$ fluorescence forming units (FFU). Briefly, the animal was anesthetized using isoflurane and administered the analgesics meloxicam (2 mg/kg) and buprenorphine (0.5 mg/kg). A 0.5 cm incision was made in the lower abdominal region and bupivocaine was applied to the tissue before the muscle layer was cut. Sterile forceps were used to probe the internal organs and locate the bladder and seminal vesicles, exposing the prostate. The left anterior prostate was penetrated with a 30G, 0.3 cc, one-half inch syringe with needle and lentiviral solution was injected. The organs were then tucked back into the viscera and the muscle fascia layer was sealed with absorbable sutures and skin closed with sterile wound clips. Post-operative meloxicam was given daily for 2 days while the animal is under observation in a hazard room. The animal was returned to the vivarium seven days after surgery and subsequently monitored.

BIOLUMINESCENCE IMAGING (BLI)

Injected mice were followed every four weeks using bioluminescence imaging (BLI), which was performed using IVIS® SpectrumCT preclinical imaging system and resolved using Living Image software (RRID:SCR_014247). We injected mice anesthetized with isoflurane with 150 mg/kg of D-Luciferin. After ten minutes incubation time, we imaged the injected mice over one minute inside the imager and compared the signal to a control mouse of the same genotype. Images were processed using Living Image software using minimum and maximum radiance counts of 4,000 and 25,000 photons, respectively. This was performed every four weeks until week 60 or until the mouse expired. After all time points were completed, we retrospectively took region of interest (ROI) measurements for the abdominal region (primary tumor, local metastases) and torso (distal metastases) for every week, covering the same area from week 4 (first BLI) until expiration for each mouse. For quantitative analysis, the total flux of radiance in photons/second was plotted at each time point for all mice.

TUMOR HARVESTING AND ANALYSIS

BLI was performed prior to each sacrifice. Following carbon dioxide exposure and cervical dislocation, euthanized animals were sprayed with 70% ethanol and cut open from a midline incision beneath the rib cage.

We removed the fat pads on the lateral edges of the peritoneal cavity to expose the ventral prostate. We cut the vas deferens and urethra to detach the prostate, bladder, and seminal vesicles from the animals, which were subsequently washed in PBS ($-Ca^{2+}$, $-Mg^{2+}$) and placed in a 6-well plate on ice. We then surgically removed the spleen, liver, lungs, brain, and tongue, and washed each organ in PBS before placing in a 6-well plate on ice. Enlarged lymph nodes were removed along with other tissues. The dead animal was then imaged using the IVIS® SpectrumCT preclinical imaging system to search for signal in additional lymph nodes. Lymph nodes showing BLI signal were excised. All isolated organs and lymph nodes were then imaged. Bones were next dissected, beginning with long bones (femur, tibia/fibula, humerus, ulna/radius) and then ribs and spine. Long bones were not washed with PBS but kept on ice in a 12-well plate, while spine and ribs were collected into a 6-well plate. All bones were imaged using BLI upon isolation.

When all organs, lymph nodes, and bones were removed, we took photographs of each individual organ or bone on a DispoCut dissection board (Simport Scientific, Saint-Mathieu-de-Beloeil, QC). We then imaged each sample in both anterior and posterior positions using bright-field and eGFP fluorescence on an inverted fluorescence microscope (Nikon SMZ1500). Where fluorescence signal was observed, we used gross dissection to cut out $eGFP^+$ lesions under the microscope and continued imaging the minced portions. Each sample was transferred to an Eppendorf tube

and stored at -80°C until DNA was extracted using phenol/chloroform/isoamyl alcohol. Where IHC was performed, half of the sample was placed in biopsy paper in an immunohistochemistry cassette and stored in 10% formalin overnight at room temperature before transferring to 70% ethanol.

COHORT INCLUSION AND EXCLUSION CRITERIA

Initially, 52 mice were injected with LV.GECPL.BC10.MG. Thirty-four mice total were injected with $1x10^5$ FFU of lentivirus (LP cohort), and eighteen mice were injected with $2x10^5$ FFU (HP cohort). Presence of $eGFP^+$ lesions was the final determinant in collecting samples for analysis of Cre and BC10. In most cases, eGFP presence aligned with BLI positivity, but in several cases spleens of mice were BL^+ but $eGFP^-$ and not further analyzed. In a single case (MMUS1544; LVL), eGFP status was questionable and BLI, Cre, and BC10 were analyzed and found to be positive. DNA was therefore sequenced by Amp-Seq. Mice were excluded for the following reasons: post-surgery death within 4 weeks (n = 6); husbandry death (n = 1); misinjection (n = 2); non-timeline interventions (n = 3); lack of BL^+ signal by week 28 (n = 20; 14 LP, 6 HP); lack of eGFP in the prostate upon dissection (n = 2). Therefore, 18 mice were included. Of these mice, 16 displayed metastasis and were further analyzed. Of these 16 mice with one or more metastasis, seven belonged to the LP cohort and 9 to the HP cohort.

FLOW CYTOMETRY AND SORTING

Cells were FACS sorted by eGFP on day 3 and passaged until day 28, when DNA was collected at quantities of 10^3, 10^4, 10^5, and 10^6 cells per sample for analysis of barcode editing. We sorted cells using a BD FACSAria II Cell Sorter (BD Biosciences, Franklin Lakes, NJ). Prior to sorting, cells were washed with 1x PBS, trypsinized, and neutralized with FACS Wash Buffer (phenol red, calcium, and magnesium-free DMEM with 0% FBS, 1x PS). The cells were then centrifuged, resuspended in FACS medium (DMEM, 2% FBS, 20mM HEPES, 2 mM EDTA, 10 U/mL DNaseI), and diluted to 10^6 cells/mL. We next filtered the cells through a 40 micron filter (BD Falcon) to create a single cell suspension. Following the sort, cells were collected in DMEM containing 20% FBS, 20 mM HEPES, and 1x Penicillin/Streptomycin, centrifuged, and reseeded in complete DMEM plus 100 μg/mL Normocin (Invivogen). Normocin was added to medium for 1 week post-sort before only complete DMEM was used.

DNA AMPLICON (BARCODE 10 — BC10) PREPARATION AND AMPLICON SEQUENCING

We digested cells in 500 μL digestion buffer (100 mM NaCl, 10 mM Tris-HCl pH 7.5, 25 mM EDTA pH 8.0, 0.5X sodium dodecyl sulfate, 0.1 mg/mL proteinase K) at 55ºC with shaking for 3 hours, before extracting DNA using phenol/chloroform/isoamyl alcohol (25:24:1). Tissues were digested in the same manner but left to digest overnight. We normalized DNA to 100 ng/μL for PCR loading in a 50 μL total reaction, and used 200-400 ng per sample, depending on tumor purity, for analysis. Prior to BC10 analysis, we added Taq 2x Master Mix (M0270L, New England Biolabs) and performed PCR on extracted DNA (95ºC for 30 seconds, 54ºC for 30 seconds, 68ºC for 10 seconds) for 28 cycles to assay for the presence of Cre recombinase (Cre_F, 5'-CGAGTGATGAGGTTCGCAAG-3'; Cre_R, 5'-ATCTTCAGGTTCTGCGGGAA-3'; 155-bp product). Our Barcode sequence is flanked by Read1/2 and P5/7 Next Generation Sequencing adapters to enable in-house indexing for sequencing. We therefore used 0.5 μM of partial Read1 (RD1, 5'-ACACTCTTTCCCTACACGAC-3') and Read2 (RD2, 5'-CTGGAGTTCAGACGTGTGCT-3') primers to amplify DNA containing BC10. The partial RD1/2 sequences were designed to minimize potential sequencing errors and enhance DNA purity resulting from lack of wide separation n degrees between annealing and extension temperatures. We added Q5 Hot Start High-Fidelity 2x Master Mix (M0494S, New England BioLabs). We then amplified DNA (98ºC for 20 seconds, 65ºC for 20 seconds, 72ºC for 30 seconds) for up to a maximum of 34 cycles due to scarcity of starting $eGFP^+$ material in some samples (rare cancer

cells in micrometastases). We detected bands ranging from ~100-400 bp showing edits (~323-bp expected size of non-edited barcode), and then used AMPure XP beads (Beckman Coulter, Brea, CA) to remove residual impurities. We next performed Amplicon Sequencing (Amp-Seq) on 20 ng/μL of the purified barcode-containing DNA. DNA sequencing of amplicons was performed by Azenta Life Sciences (South Plainfield, NJ), using the Amplicon-EZ (150-500 bp) Next Generation Sequencing (NGS) service. Purified DNA was quantified using the dsDNA Quantitation Kit (Thermo Scientific) and measured using the Qubit 4 Fluorometer (Thermo Scientific) using the 1x dsDNA: High Sensitivity feature. A total of 500 ng of DNA containing Illumina partial adapters (RD1/RD2) from each sample was then sequenced using the Illumina 2x250PE platform configuration after adding P5/P7 with different indexes, with at least 50,000 reads per sample.

IMMUNOHISTOCHEMISTRY

Prostates were microdissected under a fluorescence microscope (Nikon SMZ1500) to ensure eGFP$^+$ cells were enriched in our samples. Samples were placed in tissue biopsy paper and stored in cassettes. Cassettes were stored in 10% formalin for 24 hours before long-term storage in 70% ethanol. Samples were processed and paraffin-embedded by the Translational Research Program (TRP) Core Facility in WCM prior to immunostaining. The following antibodies were used for immunohistochemistry: PTEN (Cell Signaling Technology, Cat#9559S), 1:100; GFP (Abcam, Cat#ab183735), 1:100; Ki67 (Abcam, Cat#16667), 1:500. We took images using the EVOS® FL Auto Imaging System (Thermo Fisher Scientific).

IMMUNOFLUORESCENCE

Unstained slides containing paraffin-embedded prostate sections were provided by the TRP Core Facility. Sections were deparaffinized in two 15-minute washes with CitriSolv (Fisher Scientific) and rehydrated by consecutive washes with 100% EtOH, 95% EtOH, 75% EtOH, 50% EtOH, and ddH_2O. The slides were transferred to boiling antigen retrieval solution (10 mM citrate buffer pH = 6.0 + 0.05% Tween-20) (Vector Laboratories; Newark, CA) for 40 minutes. Once at room temperature, we washed the slides with PBS and incubated in permeabilization solution (1x PBS + 0.5% Triton X-100) for 10 minutes. We then washed slides with PBS again and incubated in blocking solution (1 x PBS + 0.2% Triton X-100 + 0.05% Tween-20 + 1% Bovine Serum Albumin + 5% Normal Goat Serum) for 1 hour. The following antibodies and dilutions were used for immunofluorescence: GFP (Thermo Fisher, Cat#A10262), 1:50; Ck8 (Abcam, Cat#ab133273), 1:200; α-SMA (Abcam, Cat#ab7817), 1:200. Sections were kept in a humidified chamber at 4°C overnight with antibody exposure. The next day, we removed primary antibodies and washed sections with IF buffer (1x PBS + 0.2% Triton X-100 + 0.05% Tween-20). We then added secondary antibodies in blocking solution to the slide in a humidified chamber and incubated for 1 hour at room temperature. We used the following secondary antibodies for IF (Thermo Fisher): anti-chicken (for GFP): 488 nm (green); anti-rabbit (for Ck8): 594 nm (red); anti-mouse (for α-SMA): 647 nm (far red). After secondary antibody staining, we added 1 μg/mL DAPI to each section for five minutes, washed with IF buffer, and added ProLong Gold Antifade Mounting (Invitrogen) to each section before sealing on a coverslip. We imaged all slides on a Zeiss AXIO Observer Z.1 spinning disc confocal microscope and analyzed used *FIJI* software (RRID:SCR_002285).

COMPUTATIONAL METHODS

DATA ANALYSIS AND VISUALIZATION

Data analyses and graph visualizations were performed with the *tidyverse* collection of R packages including *dplyr*, *purrr*, *stringr*, *tidyr*, *ggplot2, tidygraph* and also *GraphPad Prism* v9.5.1 (RRID:SCR_002798) on Apple Macintosh computers. Figures were created with the assistance of BioRender (*www.BioRender.com*, RRID:SCR_018361).

EPIDEMIOLOGICAL ASSESSMENT

Epidemiology of human prostate cancer was performed using SEER Stat Fact Sheets 2022, available from: http://seer.cancer.gov/statfacts/ (RRID:SCR_006902).

DESIGNING AMPLICON USING CFD (CUTTING FREQUENCY DETERMINING) SCORE

BC10 is a synthetic array of 10 Cas9 target sites totaling in 260-base pairs (bp) including: (1) target site (20-bp); (2) protospacer adjacent motif (PAM, 3-bp); and (3) spacers (3-bp). BC10, which is included in the plasmid injected into mice, contains target sites ordered by decreasing activity. BC10 was designed using a cutting frequency determination (CFD) score algorithm implemented in *CRISPRseek* (67). CFD scores are based on mismatches of each possible type at each position within the guide sequence (40). We assigned CFD scores based on four mismatched nucleotides (maximum cut: 1.0, minimum cut 0.1). Cas9 edits enable the identification of clonal frequency and lineage relationships through shared mutational patterns (amplicon sequence variants, ASVs).

EVOTRACER CODE

Code for the *EvoTraceR* package is available *via* GitHub (*https://github.com/Nowak-Lab/EvoTraceR*).

FASTQ FILES PROCESSING WITH EVOTRACER

The output of the amplicon sequencing experiment consists of two fastq files containing paired-end reads. For each mouse the fastq files were demultiplexed according to the sample, which can be either a metastatic site or the primary tumor. We processed the files to extract all ASVs present in each sample with a standard bioinformatics pipeline. *Trimmomatic* (RRID:SCR_011848) (68) v0.39 was used to trim the adapters and remove low quality bases, and *Flash* (69) v2.2.00 was then used to merge the paired-end reads. After this preliminary step we obtained merged reads, and computed the frequency of identical sequences.

AMPLICON SEQUENCE VARIANT PROCESSING WITH EVOTRACER

After the preliminary filters described above, we performed multiple steps to take into account possible sequencing errors and identify the final pool of ASVs to employ in the analyses.

Hamming Distance

We started by pooling together the reads characterized by Hamming distance equal to or less than 2. To do so, we first grouped reads with identical length and then we performed clustering and pooling of the counts using the *UMIClusterer* class from UMI-tools package (RRID:SCR_017048) (70). This package is designed to pool together UMIs in a single-cell experiment, and it implements multiple methods to perform this task. We employed the network-based method that employs a directed graph, where each node corresponds to an ASV and edges are created in the following way: for each pair of nodes (A, B), there is an edge from A to B if the Hamming distance is equal to or less than a threshold (in our case we set it to 2) and if $|A| \geq (2 * |B| - 1)$, where |A| and |B| refer to the corresponding sequence counts. Then, each connected component in the graph is treated as an ASV group, where the sequence with the highest number of counts is chosen as the representative and counts from every group member are pooled together.

Alignment and Merging

After this pooling, we aligned every sequence to the original non-marked barcode BC10, using the Needleman-Wunsch algorithm implemented in the PairwiseAlignment function from the R package *Biostrings* v4.3. For this step we employed the penalty scheme which was tested and selected in Labun et al., 2019 (43). After alignment, we analyzed all mutations and discarded indels too distal from any target site to have been caused by Cas9. Cas9 is mainly responsible

for deletions and insertions. We thus ignored substitutions, which are more likely than indels to be caused by sequencing errors, pooling the counts of those sequences that were identical in terms of indels.

Finally, to remove possible contamination artifacts we leveraged a property of the barcode: from its design we know that the last 10 nucleotides (right flanking sequence) should not be mutated by Cas9. Additionally, the first 5 nucleotides can be used to identify barcode sequences from the pool. However, these nucleotides lie in proximity to the first target site, which is characterized by the highest probability of being cleaved by Cas9 (See Paragraph: DESIGNING AMPLICON USING CFD SCORE), and we observed that they may also be affected by indels. Thus, for the analysis presented in this manuscript we exploited the first 5 nucleotides or last 10 nucleotides to perform the contamination filtering step, discarding all sequences whose right flanking sequences did not match the non-marked barcode.

PHYLOGENETIC TREE RECONSTRUCTION

Once all ASVs had been identified, we reconstructed the phylogenetic tree based on the accumulation of edits. To perform this inference, we built a binary mutation matrix, where each row corresponds to an ASV and each column corresponds to a mutation. Note that deletions are identified through the start and end position, while insertions correspond to the position where the insertion is added, with the number of nucleotides inserted. Given the input binary matrix, to reconstruct the phylogeny we used the greedy algorithm implemented in *Cassiopeia* (41) v2.0.0. This algorithm proceeds iteratively by splitting sequences in two groups based on the most common mutation present in the current set and keeps recursively applying the same procedure until a set is composed by only one sequence.

SIMULATING BC10 BARCODE DATA TO EVALUATE EVOTRACER

To simulate BC10 barcode sequences with known clone phylogeny and tissue migration paths, we relied on *Cassiopeia 2.0.0* and *ART Illumina Q Version 2.5.8*. We simulated a phylogenetic tree with 100 leaves, overlaying an indel matrix based on variable mutation rates per target site using the *Cassiopeia Cas9LineageTracingDataSimulator* module. Each row in the indel matrix represents the barcode sites of an individual. The size of simulated indels was drawn from an exponential distribution with a scale parameter β of 6, which best fit observed indel sizes in our empirical data. We model insertions and deletions to have equal probabilities of being induced. We also allow for the dropout of target sites caused by large neighboring deletions. Inserted sequences are random nucleotide strings. The simulated mutated sequences are then used to simulate paired-end Illumina reads in fastq format using ART. To evaluate *EvoTraceR*'s ability to detect mutations and compare it with similar tools, we then inferred mutations using *EvoTraceR 1.0*, *ampliCan 1.20.0*, and *CRISPResso2 2.2.7*. Finally, by overlaying tissue labels onto each node in the simulated phylogeny based on a variable migration probability matrix, we simulated tissue-to-tissue migrations to evaluate migration inference using *MACHINA v1.2*. The code to reproduce all simulations and benchmarking is provided *via* Github (*https://github.com/ascheben/evotracer_machina*).

STATISTICAL ANALYSIS OF BENCHMARKING EVOTRACER INDEL DETECTION AND MIGRATION INFERENCE WITH SIMULATED BC10 DATA

The analysis of *EvoTraceR*'s ability to retain simulated BC10 information (10-target sites) involved calculating alignment errors by comparing the length of the ASV in base pairs with the best alignment score for the pool of simulated barcodes. To assess the performance of *EvoTraceR*, *ampliCan*, and *CRISPResso2* in calling specific mutations, BC10 was simulated 100 times with only one active target site as *ampliCan* and *CRISPResso2* are not designed to easily handle multiple sites. The mutations called by each software were directly compared to a list of ground truth simulated mutations to identify true positives, false positives, and false negatives. Precision was calculated as true positives / (true positives + false positives), recall as true positives / (true

positives + false negatives), and F1 score as 2 * (precision * recall) / (precision + recall) for each software's mutation calling performance. For the analysis of migration inference by MACHINA, the same statistical measures were used for 1,100 simulations (100 simulations for each of the 11 migration rates tested). The true positive, false positive, and false negative results were defined by independently comparing the number of true simulated counts to inferred counts for each migration path between tissues.

INFERRING MIGRATION HISTORIES WITH MACHINA

We used the parsimonious migration history with tree resolution (PMH-TR) mode in the *MACHINA v1.2* software (44) to infer clonal migration histories for our populations, setting the prostate as the site of the primary tumor. Using the indel-based phylogeny inferred by *Cassiopeia* and the observed tissue labels at the tree tips, *MACHINA* infers migration histories while jointly resolving ambiguous polytomous branches in the phylogeny. These polytomies frequently arise in phylogenies of clonal populations, which often do not have sufficient mutations to be fully resolved. The employed maximum parsimony approach in *MACHINA* relies on minimizing the number of migrations between tissues across the clonal phylogeny to resolve polytomies and assign ancestral tissue states. *MACHINA* PMH-TR produces a single optimal migration history, providing limited information on uncertainty in its inference.

We selected *MACHINA* for migration history inference because this parsimony-based approach is well suited to the simple datasets used here. *MACHINA* has been applied to analyze various cancers, including PCa (10), breast (71), ovarian (72), and colorectal cancer (73), and has been shown to perform similarly to Bayesian methods such as PathFinder (74). To quantify uncertainty in the migration history, we applied a bootstrapping approach with 100 replicates to the most information-rich samples, MMUS1469 and MMUS1495. Using *Cassiopeia*, we randomly sampled characters in the mutation matrix with replacement and then recalculated the phylogeny and migration history. Although this approach is not able to address all sources of uncertainty, we found that our migration inferences are robust to resampling of the mutation matrix **(Supplementary Data Tables 1-2)**.

We inferred migration histories for a total of ten mice, five from the *EvoCaP*-LP cohort (MMUS1466, MMUS1467, MMUS1469, MMUS1874, MMUS1875), and five from the *EvoCaP*-HP cohort (MMUS1457, MMUS1492, MMUS1495, MMUS1544, MMUS1588), excluding MMUS1542 and MMUS1463 due to lack of metastases and others due to lack of available DNA samples or low DNA purity.

TISSUE MIGRATION TRANSITIONS

We aimed at studying the transitions of ASVs across samples to obtain information concerning directionality of spread between different CPs. Using the inferred migration graphs, we calculated conditional probabilities of transitioning from a source based on the frequency of observed transitions between all tissues in each CP and across the sum of all CPs.

For each tissue, we also compared clonal heterogeneity based on the Shannon's *H* (75), which has been widely used to assess tumor heterogeneity (76). We used an unpaired two-sample Wilcoxon test to compare Shannon diversity indices calculated for 10 mice between pairs of tissues. To simplify comparison between all mice, subtissue categories were merged to the main tissue groups consisting of prostate, lymph nodes, liver, lung, and bone. Furthermore, we assessed the rate of migration between primary, visceral (liver and lung) and non-visceral (lymph nodes and bone) tissue groups.

TISSUE SEEDING TOPOLOGIES

We classified the inferred migration histories using a set of seeding topology types based on previously defined metastatic topologies (9), including trajectories such as parallel seeding and metastatic seeding. To assign seeding topologies to each CP, we traversed the polytomy-

resolved trees with ancestral tissue states for each node inferred by MACHINA from the tips to the root. The root ancestral tissue state is known due to the initiation of the primary tumor in the prostate. Seeding topologies can then be inferred using a simple algorithm. For example, parallel seeding is detected during this traversal when a parent node has multiple child nodes with different metastatic tissue states. Similarly, when child nodes share the same tissue as their parent node, we infer the presence of a "confined" trajectory. Using this simple approach for classifying seeding topologies relies on the accuracy of the ancestral tissues states inferred by *MACHINA*, which does not capture uncertainty in ancestral tissue states but has nevertheless been shown to perform well at identifying seeding patterns (76). To investigate seeding topologies in human prostate cancer using the approach described above, we used *MACHINA* with the same settings used for our previous analyses to re-analyze human prostate cancer data from the samples A10, A29, A31, and A32 generated by Gundem et al., 2015 (10) and processed by El-Kebir et al., 2018 (44). Samples for these four patients were chosen based on their availability in a public repository (https://github.com/raphael-group/machina/). By traversing the trees with inferred ancestral states from the tips to the root, we counted the frequency of all seeding topologies for each CP.

TISSUE DISPERSAL SCORE

We next sought to characterize clones by their dispersion across tissues, with the goal of understanding which clones are more (or less) metastatic: a group of ASVs that are closely related in the phylogenetic tree but are found dispersed across different sites is assumed to be more metastatic than a clone found in a limited number of samples (9). For this purpose, we used the tissue dispersal score described in Jones *et al.*, 2020 (41). This score serves as an indicator of how well the distribution of one clone over the tissues matches the background distribution obtained by pooling together all counts from any other cluster. Quinn *et al.*, 2021(9) demonstrated how this tissue dispersal score correlates with the metastatic rate of cells, thus we employed it to obtain quantitative insights about the metastatic potential of clonal populations. For each clone *c* we followed the same procedure described by Quinn *et al.*, 2021 (9). Briefly, we created a contingency table by pooling together the counts of all ASVs in *c* in each tissue, and we also summed all counts across all the other clones, to obtain a background distribution. Then, we scaled the background counts so that the sum of both columns in the contingency table (*c* and background) is the same. We performed a chi-squared test on the contingency table, and we computed the Cramer's V test statistic, which was finally inverted to obtain the tissue dispersal score.

CODE AND DATA AVAILABILITY

Data were generated by the authors and included in the article. Publicly available data generated by others were used by the authors. The human data analyzed in this study were obtained from a secondary analysis of data generated by Gundem *et al*., 2015 (10) that was performed by El-Kebir *et al*., 2018 (44). A compute capsule containing code used in this manuscript is available through Code Ocean using the following url: https://codeocean.com/capsule/3348925.

ACKNOWLEDGMENTS

We would like to thank Dr. Sébastien Monette for his pathology evaluation discussion. We thank Dr. Matthias Stadtfeld for his generous gift of a fluorescent microscope to our laboratory. We thank Research Assistant Francesca Jereis for her constructive criticism of this manuscript. We would like to thank Lizzie Kaminoff for her assistance in mouse breeding and husbandry. We would like to thank Dr. Olivier Elemento for his consultations regarding bioinformatics approaches. Financial support for D.G.N. was provided by the Research Scholar Grant from the American Cancer Society (ACS), National Cancer Institute (NCI) R01-CA272466 and Weill Cornell Medicine, Walter B. Wriston Research Scholar. Financial support for A.C.S., and to the Simons

Center for Quantitative Biology, was provided by the U.S. National Institute of Health (NIH) grant R35-GM127070. Financial support for R.N.S. was provided by the Prostate Cancer Research Program (PCRP) of the Department of Defense (DoD). This is the Early Investigator Research Award, number W81XWH-22-1-0068, project number PC210035. Additional funding support was provided by the National Cancer Institute (NCI) Molecular and Translational Oncology Research (MTOR) Award (T32CA203702). Financial support for B.L. was provided by National Institutes of Health (NIH) T32 Training Grant 1T32GM141949-01. Financial support for D.G. was provided by National Institutes of Health (NIH) T32 Training Grant 5T32GM141949-02. L.P. was supported by the Cancer Research UK and Associazione Italiana per la Ricerca sul Cancro (CRUK/AIRC) “Accelerator Award” (award #22790). S.J.S. was supported by a Starr Centennial Scholarship endowed to Cold Spring Harbor Laboratory from the Starr Foundation. We would like to thank Dr. David Wilkes and Thomas Caiazza from the Englander Institute for Precision Medicine (EIPM) for access to their equipment and reagents. Illustrations were created using a BioRender.com license.

BIBLIOGRAPHY

1. Hernandez RK, Wade SW, Reich A, Pirolli M, Liede A, Lyman GH. Incidence of bone metastases in patients with solid tumors: analysis of oncology electronic medical records in the United States. BMC Cancer. 2018;18(1):44.
2. Disibio G, French SW. Metastatic patterns of cancers: results from a large autopsy study. Arch Pathol Lab Med. 2008;132(6):931-9.
3. Massagué J, Ganesh K. Metastasis-Initiating Cells and Ecosystems. Cancer Discov. 2021;11(4):971-94.
4. Armenia J, Wankowicz SAM, Liu D, Gao J, Kundra R, Reznik E, et al. The long tail of oncogenic drivers in prostate cancer. Nat Genet. 2018;50(5):645-51.
5. Stanta G, Bonin S. Overview on Clinical Relevance of Intra-Tumor Heterogeneity. Front Med (Lausanne). 2018;5:85.
6. Heyde A, Reiter JG, Naxerova K, Nowak MA. Consecutive seeding and transfer of genetic diversity in metastasis. Proc Natl Acad Sci U S A. 2019;116(28):14129-37.
7. Hu Z, Ding J, Ma Z, Sun R, Seoane JA, Scott Shaffer J, et al. Quantitative evidence for early metastatic seeding in colorectal cancer. Nat Genet. 2019;51(7):1113-22.
8. Obenauf AC, Massagué J. Surviving at a Distance: Organ-Specific Metastasis. Trends Cancer. 2015;1(1):76-91.
9. Quinn JJ, Jones MG, Okimoto RA, Nanjo S, Chan MM, Yosef N, et al. Single-cell lineages reveal the rates, routes, and drivers of metastasis in cancer xenografts. Science. 2021;371(6532).
10. Gundem G, Van Loo P, Kremeyer B, Alexandrov LB, Tubio JMC, Papaemmanuil E, et al. The evolutionary history of lethal metastatic prostate cancer. Nature. 2015;520(7547):353-7.
11. Brown D, Smeets D, Székely B, Larsimont D, Szász AM, Adnet PY, et al. Phylogenetic analysis of metastatic progression in breast cancer using somatic mutations and copy number aberrations. Nat Commun. 2017;8:14944.
12. Ullah I, Karthik GM, Alkodsi A, Kjällquist U, Stålhammar G, Lövrot J, et al. Evolutionary history of metastatic breast cancer reveals minimal seeding from axillary lymph nodes. J Clin Invest. 2018;128(4):1355-70.
13. Kim MY, Oskarsson T, Acharyya S, Nguyen DX, Zhang XH, Norton L, et al. Tumor self-seeding by circulating cancer cells. Cell. 2009;139(7):1315-26.
14. McKenna A, Findlay GM, Gagnon JA, Horwitz MS, Schier AF, Shendure J. Whole-organism lineage tracing by combinatorial and cumulative genome editing. Science. 2016;353(6298):aaf7907.
15. Rogers ZN, McFarland CD, Winters IP, Naranjo S, Chuang CH, Petrov D, et al. A quantitative and multiplexed approach to uncover the fitness landscape of tumor suppression in vivo. Nat Methods. 2017;14(7):737-42.
16. Kalhor R, Kalhor K, Mejia L, Leeper K, Graveline A, Mali P, et al. Developmental barcoding of whole mouse via homing CRISPR. Science. 2018;361(6405).
17. Chan MM, Smith ZD, Grosswendt S, Kretzmer H, Norman TM, Adamson B, et al. Molecular recording of mammalian embryogenesis. Nature. 2019;570(7759):77-82.
18. Weinreb C, Rodriguez-Fraticelli A, Camargo FD, Klein AM. Lineage tracing on transcriptional landscapes links state to fate during differentiation. Science. 2020;367(6479).
19. Simeonov KP, Byrns CN, Clark ML, Norgard RJ, Martin B, Stanger BZ, et al. Single-cell lineage tracing of metastatic cancer reveals selection of hybrid EMT states. Cancer Cell. 2021;39(8):1150-62.e9.

20. Yang D, Jones MG, Naranjo S, Rideout WM, Min KHJ, Ho R, et al. Lineage tracing reveals the phylodynamics, plasticity, and paths of tumor evolution. Cell. 2022;185(11):1905-23.e25.
21. Lima A, Maddalo D. SEMMs: Somatically Engineered Mouse Models. A New Tool for. Front Oncol. 2021;11:667189.
22. Nowak DG, Cho H, Herzka T, Watrud K, DeMarco DV, Wang VM, et al. MYC Drives Pten/Trp53-Deficient Proliferation and Metastasis due to IL6 Secretion and AKT Suppression via PHLPP2. Cancer Discov. 2015;5(6):636-51.
23. Armenia J, Wankowicz SAM, Liu D, Gao J, Kundra R, Reznik E, et al. The long tail of oncogenic drivers in prostate cancer. Nat Genet. 2018;50(5):645-51.
24. Cho H, Herzka T, Zheng W, Qi J, Wilkinson JE, Bradner JE, et al. RapidCaP, a novel GEM model for metastatic prostate cancer analysis and therapy, reveals myc as a driver of Pten-mutant metastasis. Cancer Discov. 2014;4(3):318-33.
25. Woodcock DJ, Riabchenko E, Taavitsainen S, Kankainen M, Gundem G, Brewer DS, et al. Prostate cancer evolution from multilineage primary to single lineage metastases with implications for liquid biopsy. Nat Commun. 2020;11(1):5070.
26. Taylor BS, Schultz N, Hieronymus H, Gopalan A, Xiao Y, Carver BS, et al. Integrative genomic profiling of human prostate cancer. Cancer Cell. 2010;18(1):11-22.
27. Labbe DP, Nowak DG, Deblois G, Lessard L, Giguere V, Trotman LC, et al. Prostate cancer genetic-susceptibility locus on chromosome 20q13 is amplified and coupled to androgen receptor-regulation in metastatic tumors. Mol Cancer Res. 2014;12(2):184-9.
28. Platt RJ, Chen S, Zhou Y, Yim MJ, Swiech L, Kempton HR, et al. CRISPR-Cas9 knockin mice for genome editing and cancer modeling. Cell. 2014;159(2):440-55.
29. Dabbs DJ, Fung M, Landsittel D, McManus K, Johnson R. Sentinel lymph node micrometastasis as a predictor of axillary tumor burden. Breast J. 2004;10(2):101-5.
30. Reinert RB, Kantz J, Misfeldt AA, Poffenberger G, Gannon M, Brissova M, et al. Tamoxifen-Induced Cre-loxP Recombination Is Prolonged in Pancreatic Islets of Adult Mice. PLoS One. 2012;7(3):e33529.
31. Donocoff RS, Teteloshvili N, Chung H, Shoulson R, Creusot RJ. Optimization of tamoxifen-induced Cre activity and its effect on immune cell populations. Sci Rep. 2020;10(1):15244.
32. Ilchuk LA, Stavskaya NI, Varlamova EA, Khamidullina AI, Tatarskiy VV, Mogila VA, et al. Limitations of Tamoxifen Application for In Vivo Genome Editing Using Cre/ER. Int J Mol Sci. 2022;23(22).
33. Wang S, Gao J, Lei Q, Rozengurt N, Pritchard C, Jiao J, et al. Prostate-specific deletion of the murine Pten tumor suppressor gene leads to metastatic prostate cancer. Cancer Cell. 2003;4(3):209-21.
34. Martin P, Liu YN, Pierce R, Abou-Kheir W, Casey O, Seng V, et al. Prostate epithelial Pten/TP53 loss leads to transformation of multipotential progenitors and epithelial to mesenchymal transition. Am J Pathol. 2011;179(1):422-35.
35. Gandaglia G, Abdollah F, Schiffmann J, Trudeau V, Shariat SF, Kim SP, et al. Distribution of metastatic sites in patients with prostate cancer: A population-based analysis. Prostate. 2014;74(2):210-6.
36. Arriaga JM, Abate-Shen C. Genetically Engineered Mouse Models of Prostate Cancer in the Postgenomic Era. Cold Spring Harb Perspect Med. 2019;9(2).
37. Arriaga JM, Panja S, Alshalalfa M, Zhao J, Zou M, Giacobbe A, et al. A MYC and RAS co-activation signature in localized prostate cancer drives bone metastasis and castration resistance. Nat Cancer. 2020;1(11):1082-96.
38. Gandaglia G, Karakiewicz PI, Briganti A, Passoni NM, Schiffmann J, Trudeau V, et al. Impact of the Site of Metastases on Survival in Patients with Metastatic Prostate Cancer. Eur Urol. 2015;68(2):325-34.

39. Zhu LJ, Holmes BR, Aronin N, Brodsky MH. CRISPRseek: a bioconductor package to identify target-specific guide RNAs for CRISPR-Cas9 genome-editing systems. PLoS One. 2014;9(9).
40. Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat Biotechnol. 2016;34(2):184-91.
41. Jones MG, Khodaverdian A, Quinn JJ, Chan MM, Hussmann JA, Wang R, et al. Inference of single-cell phylogenies from lineage tracing data using Cassiopeia. Genome Biol. 2020;21(1):92.
42. Clement K, Rees H, Canver MC, Gehrke JM, Farouni R, Hsu JY, et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. Nat Biotechnol. 2019;37(3):224-6.
43. Labun K, Guo X, Chavez A, Church G, Gagnon JA, Valen E. Accurate analysis of genuine CRISPR editing events with ampliCan. Genome Res. 2019;29(5):843-7.
44. El-Kebir M, Satas G, Raphael BJ. Inferring parsimonious migration histories for metastatic cancers. Nat Genet. 2018;50(5):718-26.
45. Taranda J, Mathew G, Watrud K, El-Amine N, Lee MF, Elowsky C, et al. Combined whole-organ imaging at single-cell resolution and immunohistochemical analysis of prostate cancer and its liver and brain metastases. Cell Rep. 2021;37(7):110027.
46. Massagué J, Obenauf AC. Metastatic colonization by circulating tumour cells. Nature. 2016;529(7586):298-306.
47. Erdi YE. Limits of Tumor Detectability in Nuclear Medicine and PET. Mol Imaging Radionucl Ther. 2012;21(1):23-8.
48. van den Bogert C, Dontje BH, Holtrop M, Melis TE, Romijn JC, van Dongen JW, et al. Arrest of the proliferation of renal and prostate carcinomas of human origin by inhibition of mitochondrial protein synthesis. Cancer Res. 1986;46(7):3283-9.
49. Grzelak CA, Goddard ET, Lederer EE, Rajaram K, Dai J, Shor RE, et al. Elimination of fluorescent protein immunogenicity permits modeling of metastasis in immune-competent settings. Cancer Cell. 2022;40(1):1-2.
50. Aceto N, Bardia A, Miyamoto DT, Donaldson MC, Wittner BS, Spencer JA, et al. Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. Cell. 2014;158(5):1110-22.
51. Maddipati R, Stanger BZ. Pancreatic Cancer Metastases Harbor Evidence of Polyclonality. Cancer Discov. 2015;5(10):1086-97.
52. Cheung KJ, Padmanaban V, Silvestri V, Schipper K, Cohen JD, Fairchild AN, et al. Polyclonal breast cancer metastases arise from collective dissemination of keratin 14-expressing tumor cell clusters. Proc Natl Acad Sci U S A. 2016;113(7):E854-63.
53. Aceto N, Toner M, Maheswaran S, Haber DA. En Route to Metastasis: Circulating Tumor Cell Clusters and Epithelial-to-Mesenchymal Transition. Trends Cancer. 2015;1(1):44-52.
54. Reiter JG, Makohon-Moore AP, Gerold JM, Heyde A, Attiyeh MA, Kohutek ZA, et al. Minimal functional driver gene heterogeneity among untreated metastases. Science. 2018;361(6406):1033-7.
55. Tang YJ, Huang J, Tsushima H, Ban GI, Zhang H, Oristian KM, et al. Tracing Tumor Evolution in Sarcoma Reveals Clonal Origin of Advanced Metastasis. Cell Rep. 2019;28(11):2837-50.e5.
56. Giese A, Loo MA, Tran N, Haskett D, Coons SW, Berens ME. Dichotomy of astrocytoma migration and proliferation. Int J Cancer. 1996;67(2):275-82.
57. Matus DQ, Lohmer LL, Kelley LC, Schindler AJ, Kohrman AQ, Barkoulas M, et al. Invasive Cell Fate Requires G1 Cell-Cycle Arrest and Histone Deacetylase-Mediated Changes in Gene Expression. Dev Cell. 2015;35(2):162-74.

58. Kohrman AQ, Matus DQ. Divide or Conquer: Cell Cycle Regulation of Invasive Behavior. Trends Cell Biol. 2017;27(1):12-25.
59. Naxerova K, Reiter JG, Brachtel E, Lennerz JK, van de Wetering M, Rowan A, et al. Origins of lymphatic and distant metastases in human colorectal cancer. Science. 2017;357(6346):55-60.
60. Beltran H, Prandi D, Mosquera JM, Benelli M, Puca L, Cyrta J, et al. Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. Nat Med. 2016;22(3):298-305.
61. Ceder Y, Bjartell A, Culig Z, Rubin MA, Tomlins S, Visakorpi T. The Molecular Evolution of Castration-resistant Prostate Cancer. Eur Urol Focus. 2016;2(5):506-13.
62. Hu Z, Li Z, Ma Z, Curtis C. Multi-cancer analysis of clonality and the timing of systemic spread in paired primary tumors and metastases. Nat Genet. 2020;52(7):701-8.
63. Aggarwal R, Huang J, Alumkal JJ, Zhang L, Feng FY, Thomas GV, et al. Clinical and Genomic Characterization of Treatment-Emergent Small-Cell Neuroendocrine Prostate Cancer: A Multi-institutional Prospective Study. J Clin Oncol. 2018;36(24):2492-503.
64. Beltran H, Hruszkewycz A, Scher HI, Hildesheim J, Isaacs J, Yu EY, et al. The Role of Lineage Plasticity in Prostate Cancer Therapy Resistance. Clin Cancer Res. 2019;25(23):6916-24.
65. Nowak DG, Katsenelson KC, Watrud KE, Chen M, Mathew G, D'Andrea VD, et al. The PHLPP2 phosphatase is a druggable driver of prostate cancer progression. J Cell Biol. 2019;218(6):1943-57.
66. Uphoff CC, Drexler HG. Detecting mycoplasma contamination in cell cultures by polymerase chain reaction. Methods Mol Biol. 2011;731:93-103.
67. Zhu LJ, Holmes BR, Aronin N, Brodsky MH. CRISPRseek: a bioconductor package to identify target-specific guide RNAs for CRISPR-Cas9 genome-editing systems. PLoS One. 2014;9(9):e108424.
68. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114-20.
69. Magoč T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics. 2011;27(21):2957-63.
70. Smith T, Heger A, Sudbery I. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. Genome Res. 2017;27(3):491-9.
71. Hoadley KA, Siegel MB, Kanchi KL, Miller CA, Ding L, Zhao W, et al. Tumor Evolution in Two Patients with Basal-like Breast Cancer: A Retrospective Genomics Study of Multiple Metastases. PLoS Med. 2016;13(12):e1002174.
72. McPherson A, Roth A, Laks E, Masud T, Bashashati A, Zhang AW, et al. Divergent modes of clonal spread and intraperitoneal mixing in high-grade serous ovarian cancer. Nat Genet. 2016;48(7):758-67.
73. Chen HN, Shu Y, Liao F, Liao X, Zhang H, Qin Y, et al. Genomic evolution and diverse models of systemic metastases in colorectal cancer. Gut. 2022;71(2):322-32.
74. Kumar S, Chroni A, Tamura K, Sanderford M, Oladeinde O, Aly V, et al. PathFinder: Bayesian inference of clone migration histories in cancer. Bioinformatics. 2020;36(Suppl_2):i675-i83.
75. Shannon CE. The mathematical theory of communication. 1963. MD Comput. 1997;14(4):306-17.
76. Kashyap A, Rapsomaniki MA, Barros V, Fomitcheva-Khartchenko A, Martinelli AL, Rodriguez AF, et al. Quantification of tumor heterogeneity: from data acquisition to metric generation. Trends Biotechnol. 2022;40(6):647-76.

FIGURE LEGENDS

Figure 1. The *EvoCaP* platform for defining cancer evolution. A, The BC10 platform is compatible with loxP technology. Lentivirus is injected into prostate of the *EvoCaP* mice. *EvoTraceR* is an R package for specific analysis of the barcode. **B**, BC10 and MG or NMG are driven by an hU6 promoter. Amplicon DNA can be extracted, amplified and sequenced for analysis of edits in the barcode (Amp-Seq). **C**, Plasmid components used in BC10 plasmid technology. **D**, Diagram of experimental procedure. Primary MEFs extracted from mice harboring the *Pten*$^{loxP/loxP}$;*Trp53*$^{loxP/loxP}$ genotype are infected with lentivirus at day 0. At days 7, 14, and 28, DNA is collected for amplicon analysis. Additional readouts include eGFP fluorescence, FLuc bioluminescence, and protein expression by Western blot. **E**, KO of Pten/p53 and activation of Cas9-eGFP with Cre expression in virus-infected cells at day 7. NVC = no virus control (negative). **F**, The number of eGFP$^+$ cells increases over time and is consistent between viruses. Individual groups were analyzed at each time point using parametric unpaired t-tests with two-tailed p-values ± SD (95% confidence interval). (Created with BioRender.com).

Figure 2. *EvoCaP* mice develop metastatic PCa to relevant secondary sites. A, Timeline for *EvoCaP* injections and analysis. Primary tumors and metastases may be tracked longitudinally during the course of live imaging based on BL$^+$ signal. **B**, Kaplan-Meier survival curve displaying rate of death in mice injected with BC10-containing lentivirus that developed metastatic prostate cancer over the 60 week period of monitoring. **C**, Longitudinal BL$^+$ analysis of BC10-injected mice shows advancing tumor progression and metastasis over 60 weeks. The top three panels show BL$^+$ progression in *EvoCaP*-LP cohort. Bottom four panels show BL$^+$ progression in *EvoCaP*-HP cohort. Black line on border indicates that mouse did not survive through 60 weeks. **D**, Box plot displaying quantitation of BL$^+$ radiance in mice that survived up to a 60 week period. After week 28, animals that did not develop BL$^+$ signal were no longer followed and penetrance was determined. Only mice with primary and metastatic eGFP$^+$ signal are included in the analysis. MMUS1466 was sacrificed between week 56-60 and final BL$^+$ data was included for week 60. Percentages listed under week 28 indicate only those mice with metastases that exhibit BL$^+$ signal. Data are displayed in log scale. **E**, Bioluminescence imaging (BLI) enables post-mortem visualization of primary tumor and metastases and fluorescence enables robust isolation of eGFP$^+$ cells using microscopy for downstream analysis. Scaling for BL$^+$ signal is identical to that used in 2C**. F**, Frequency of eGFP$^+$ signal post-mortem in analyzed tissues from all mice exhibiting metastasis. FL* in some lymph nodes indicates that fluorescence was not measured but the lack of BL$^+$ signal or enlargement makes presence of tumor cells unlikely. The numbers in each green box represent the number of different sites that has at least one metastasis. In MMUS1874, two metastases were isolated in the same bone but counted as one site (FMR). Local invasion into seminal vesicle is also depicted. Fluorescence shows metastasis distribution in different organs (frequency of metastases: bones > liver > lungs > lymph nodes). (Created with BioRender.com)

Figure 3. Scheme for BC10 design. A, CFD scores were developed based on assessment of types of mismatches at every position within the guide sequence. We assigned CFD scores based on four mismatched nucleotides and positioned the 10x TS on the barcode in order of decreasing predicted activity (maximum cut TS01: 1.0; minimum cut TS10: 0.1). Cas9 creates heritable marks (insertions and deletions) in target sites with likelihood predicted by CFD score. ASVs are identified based on deletions (rectangles) and insertions (diamonds) in the BC10 and transformed to the Boolean matrix (color: mark or white: absent mark in BC10). **B**, *EvoTraceR* comparison to *CRISPResso2* and *ampliCan.* **C**, ASVs are grouped based on common marks with nucleotide resolution (truncal mutations) and are considered as related clonal populations (CPs). ASVs could be found in different metastatic sites. **D**, BC10 analysis in the *EvoCaP* enables descriptions of CPs and routes of metastatic spread. (Created with BioRender.com)

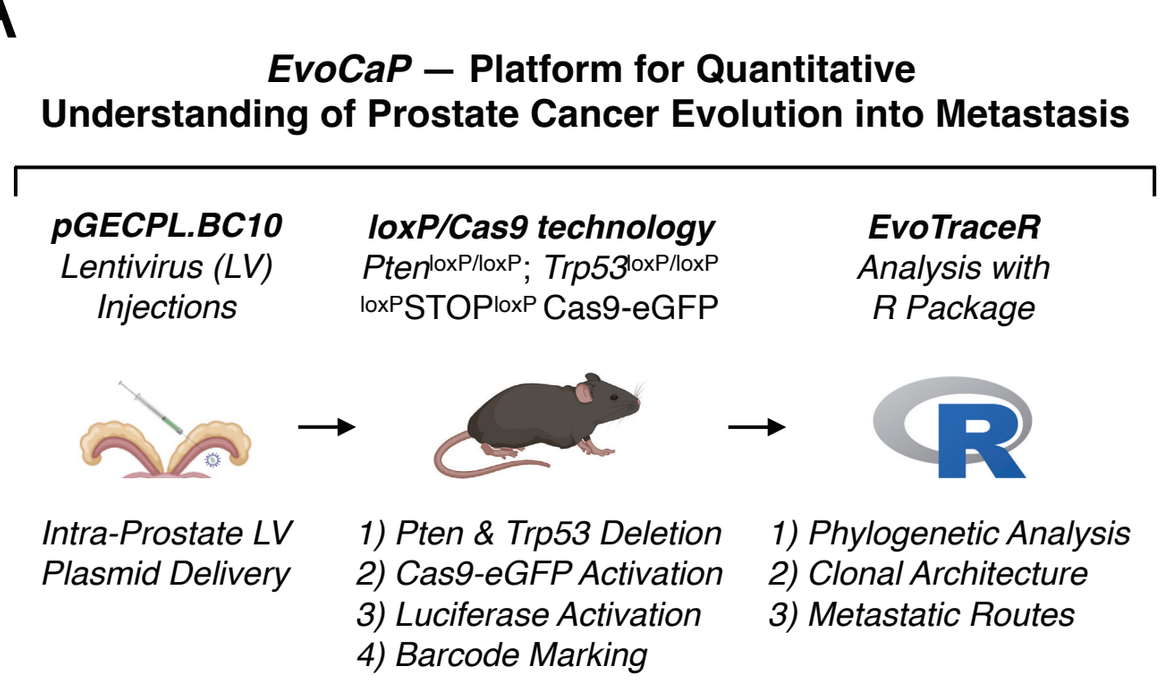
Figure 4. Barcode analysis reveals high shared editing patterns between primary and metastatic sites. A, Fluorescence images in prostate (PRL), median and right liver lobes (LVM, LVR), and left rib (RBL) in MMUS1495. **B**, BC10 edits are characterized by deletions and insertions caused by Cas9, mainly at 1.0/0.9 and 0.5/0.4 target sites as predicted by CFD score. Frequency (y-axis) of deletions (blue) and insertions (red) across the entire amplicon (260 bp, x-axis) in different organs. **C**, The majority of marks in the BC10 are short (<30 bp deletions and <6 bp insertions) affecting from 1-4x different sites (x-axis: 1-52, light blue and light red bars). By comparison, larger indels are rarely observed (x-axis: >130, darker blue bars). X-axis signifies the number of base pairs within the BC10 regions that a specific edit spans. Edits greater than 26 bp constitute a region larger than one TS. **D-E**, Eco-statistical measures of heterogeneity; **D**, Shannon's index measures alpha-diversity, which quantifies intraclonal heterogeneity. **E**, Beta-diversity is calculated using Bray-Curtis dissimilarity scores, and is representative on inter-organ heterogeneity. Organs with scores closest to 1 are the least similar to one another, while a score of zero would indicate that the tissues are exactly homogeneous in terms of their edits. **F-H**, Phylograms are shown for the most information-rich CPs with the highest number of counts. **F**, CP01, **G**, CP02, and **H**, CP03, We depict genealogies of selected CPs composed of their intrinsic ASVs. Phylogenetic analysis (phylogram) performed using *Cassiopeia* suite with the greedy algorithm. ASVs are expressed as lengths of BC10 deletion (blue) or insertion (red). Individual CPs are shown with descendent subclones (branches and leaves).

Figure 5. Migration histories of mouse with advanced prostate cancer. A-F, Reconstructed migration histories of CP01, CP02, and CP03 from MMUS1495. **A**-**C**, *MACHINA*-generated metastatic trajectories for **A**, CP01, **B**, CP02, and **C**, CP03. Circles represent ASV frequency within a particular organ. Intensity of edge color denotes degree of expansion of an ASV within the defined organ. Diamonds represent ancestral states inferred by *MACHINA*. **D-F**, Individual transition matrices for **D**, Transition matrix for all CPs in MMUS1495. **E-G**, Individual transition matrices for **E**, CP01, **F**, CP02, and **G**, CP03, showing trajectories of metastatic spread between organs and extent of expansion within an organ. Numbers range from 0 (absent in site) to 1 (confinement within site). Transition events occurring at rates below 1% fall below the threshold and not shown. Seeding trajectories are displayed above each matrix.

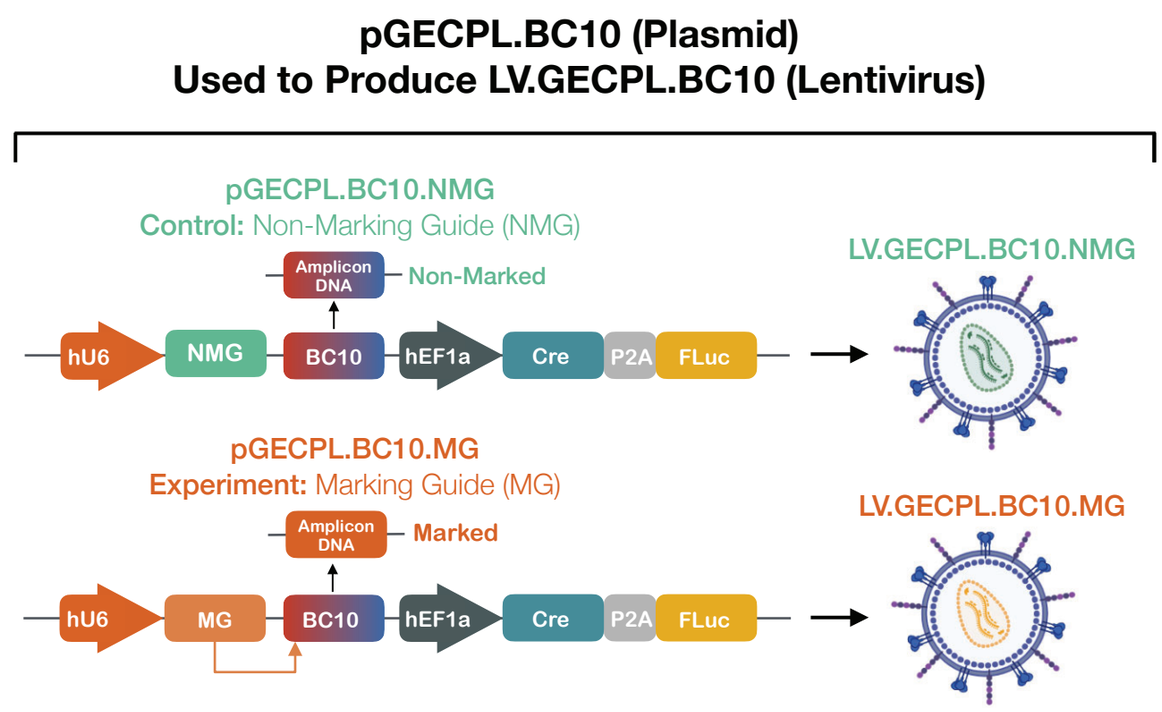
Figure 6. Tissue seeding topologies of clonal populations. A, Possible seeding topologies delineating the routes of tumor cell expansion and spread within and between different organs. **B**, Example of expected migration patterns based on possible seeding topologies. **C-E**, Actual seeding topologies exhibited by mice in *EvoCaP*-LP **(C)** and *EvoCaP*-HP **(D)** cohorts compared to seeding topologies of clones analyzed from a human PCa dataset **(E)**. Pie charts indicate percentages of each possible type of seeding topology based on total number of CPs exhibiting each topology. Total CPs undergoing a specific type of topology are indicated by percentages. Topologies were considered for CPs in the primary tumor that were confined to the primary site, disseminated to a secondary site (either seeded individually or in parallel with more than one site), or re-seeded from a secondary site. Topologies of CPs from metastatic sites include those that remained confined to the same site, seeded another secondary organ, or traveled bidirectionally between different metastatic sites. **F**, Frequency of seeding topologies observed across human patients from Gundem *et al*., 2015 (10) (n=4), HP mice (n=5), and LP mice (n=5). Percentages for all CPs of each mouse and human patient are shown. No significant differences between groups were detected for any seeding topology (unpaired two-sample Wilcoxon test, Bonferroni-adjusted p>0.05).

Figure 1.

A



B



C

Components of pGECPL.BC10 Plasmids

hEF1a Cre **hEF1a** (human Elongation Factor 1-alpha) ubiquitous promoter driven-Cre

P2A **P2A** Self-Cleaving Peptide

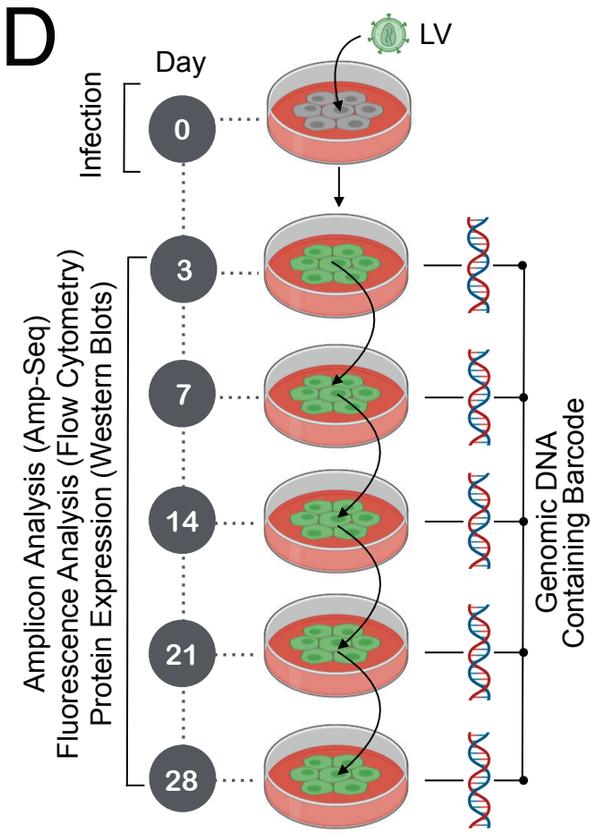
FLuc **FLuc** (Firefly Luciferase), marker used for *in vivo* longitudinal Bioluminescence Imaging (BLI)

hU6 NMG **NMG** (Non-Marking Guide), Control guide

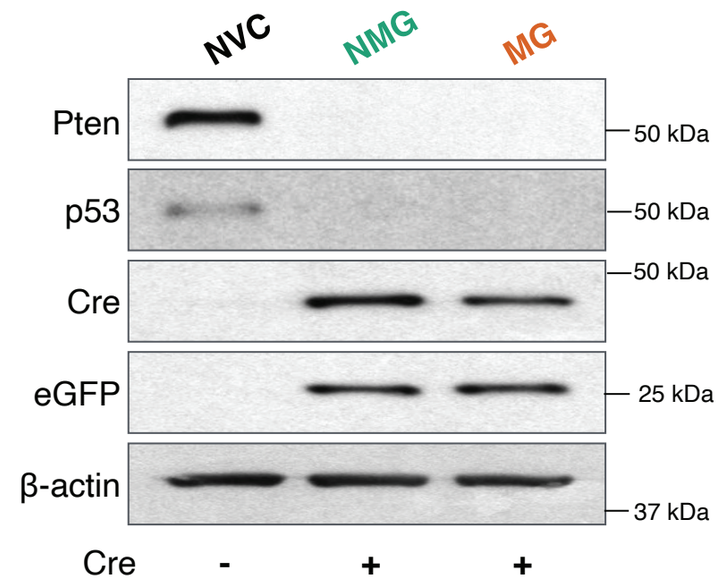
hU6 MG **MG** (Marking Guide), specifically targets **BC10**; driven by **hU6** promoter

BC10 **BC10** (Barcode 10) synthetic array of 10× Cas9 target sites with decreasing activity for cumulative editing over time

D



E



F

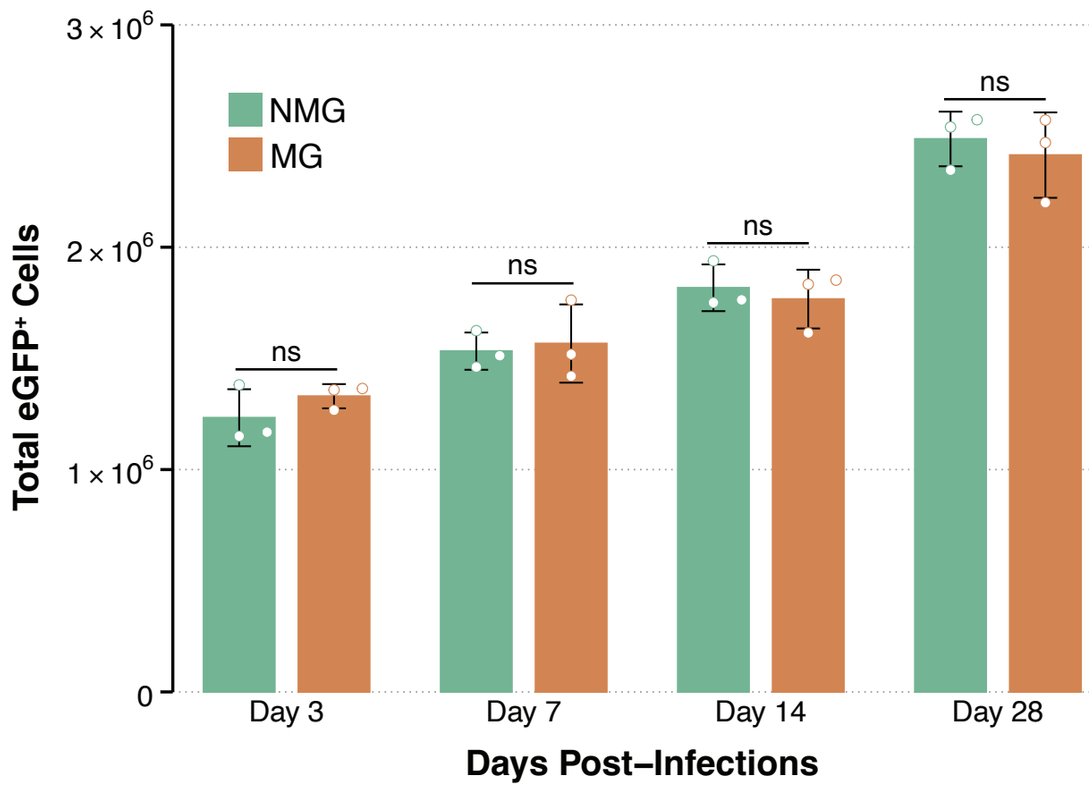
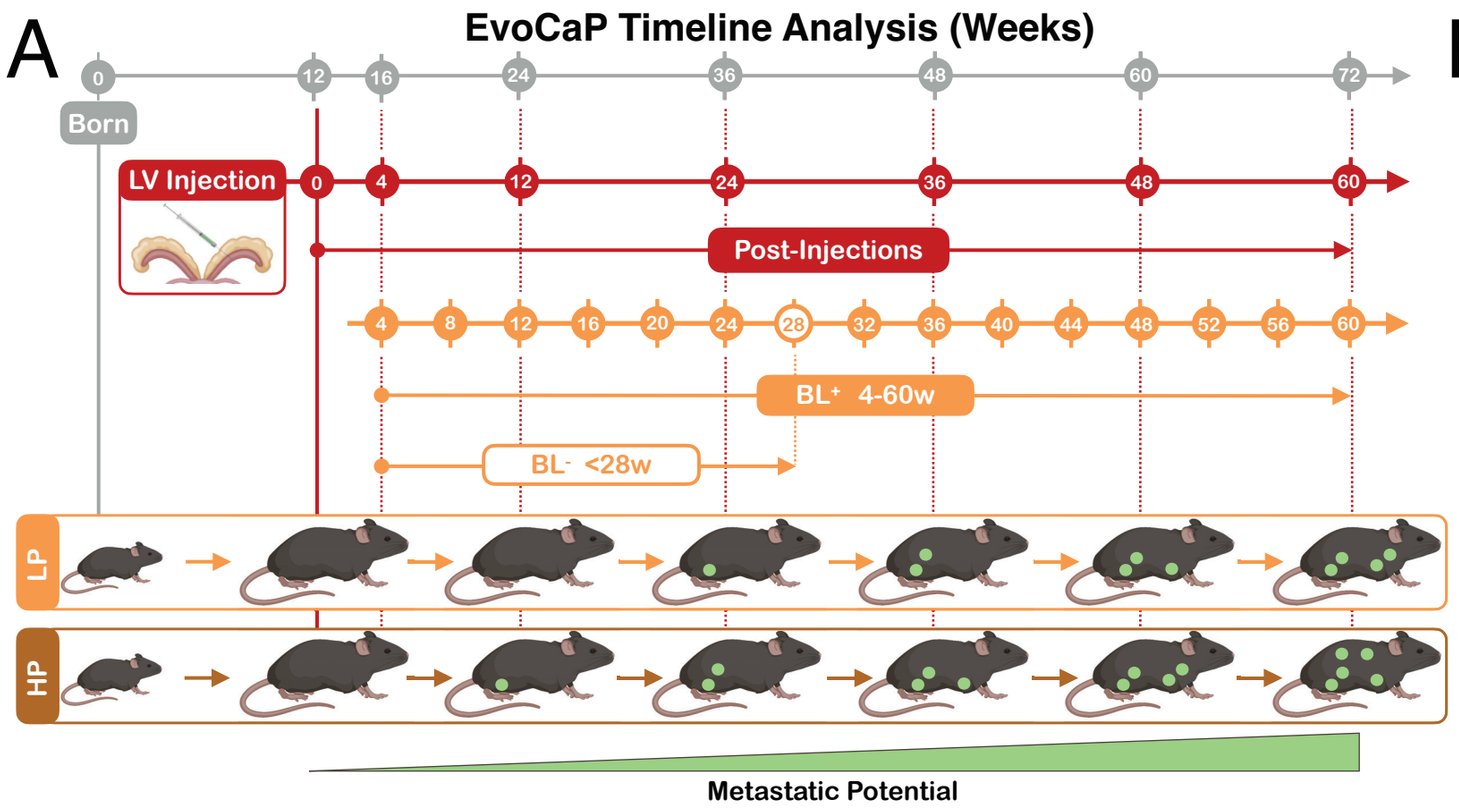
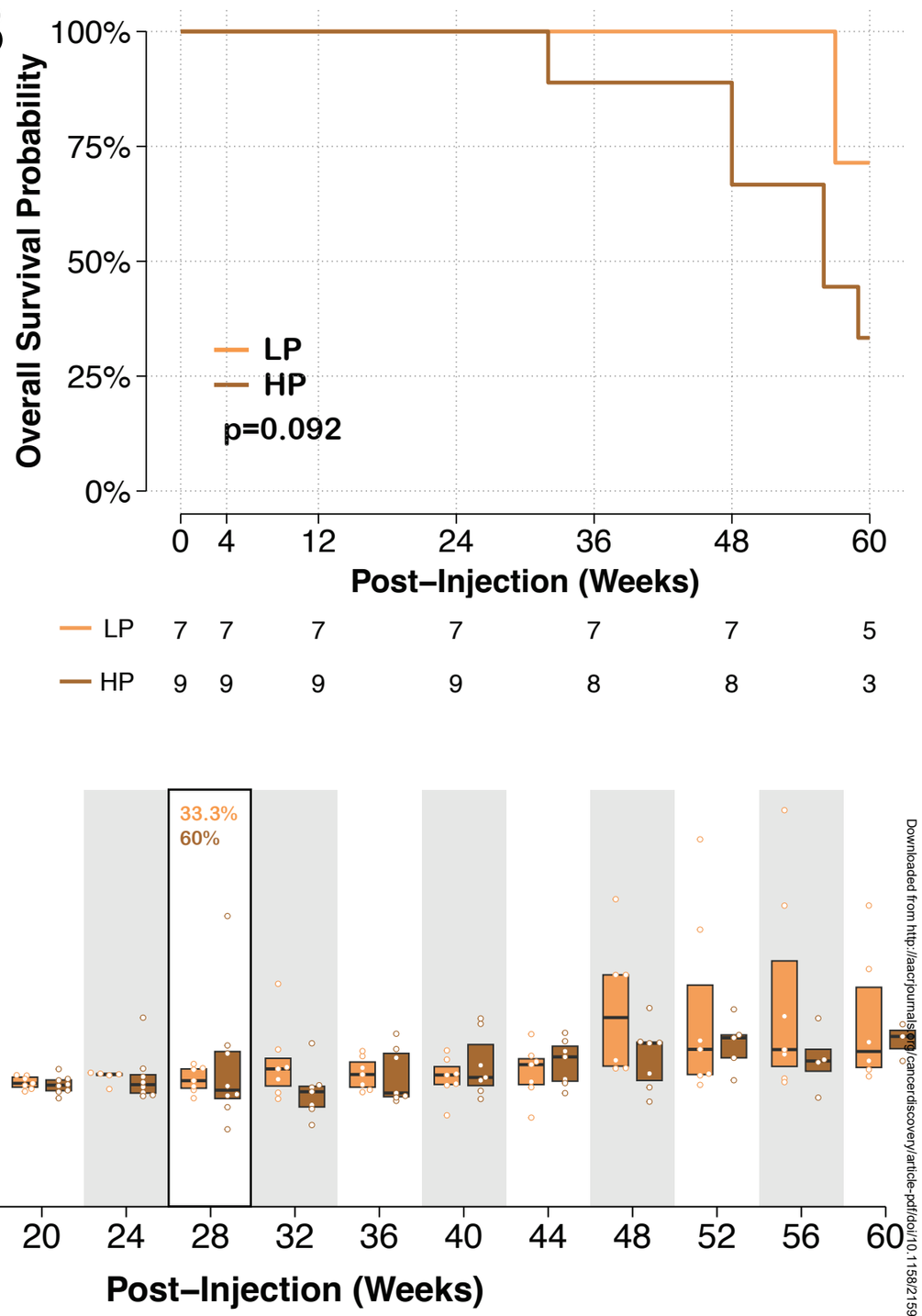
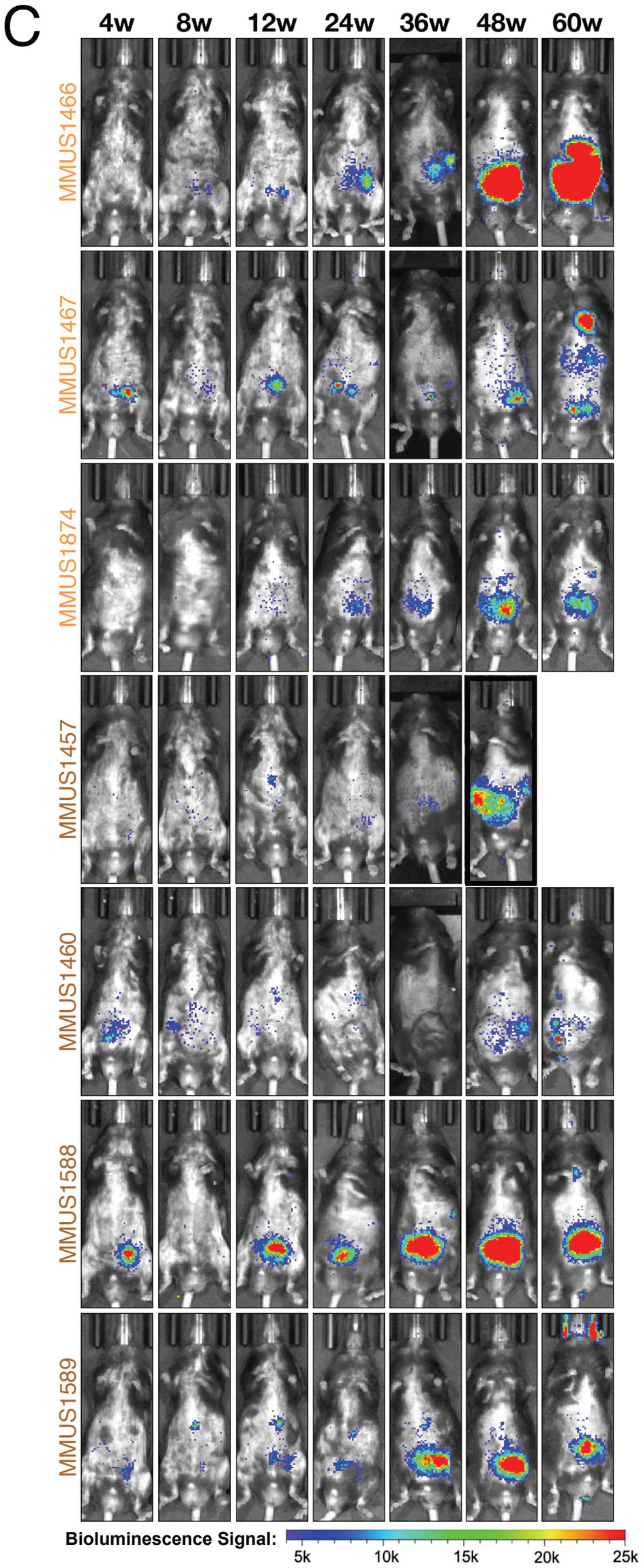
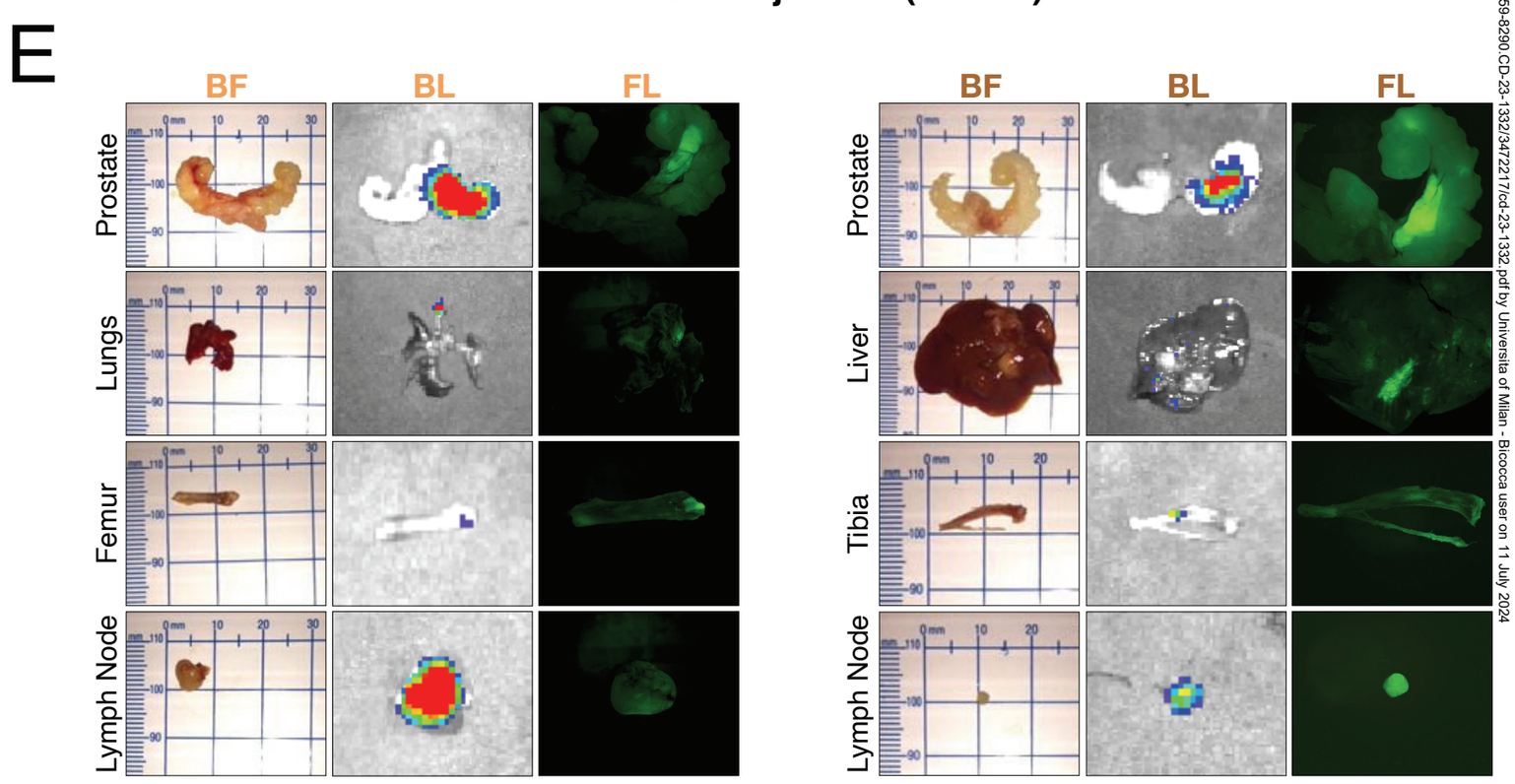


Figure 2.









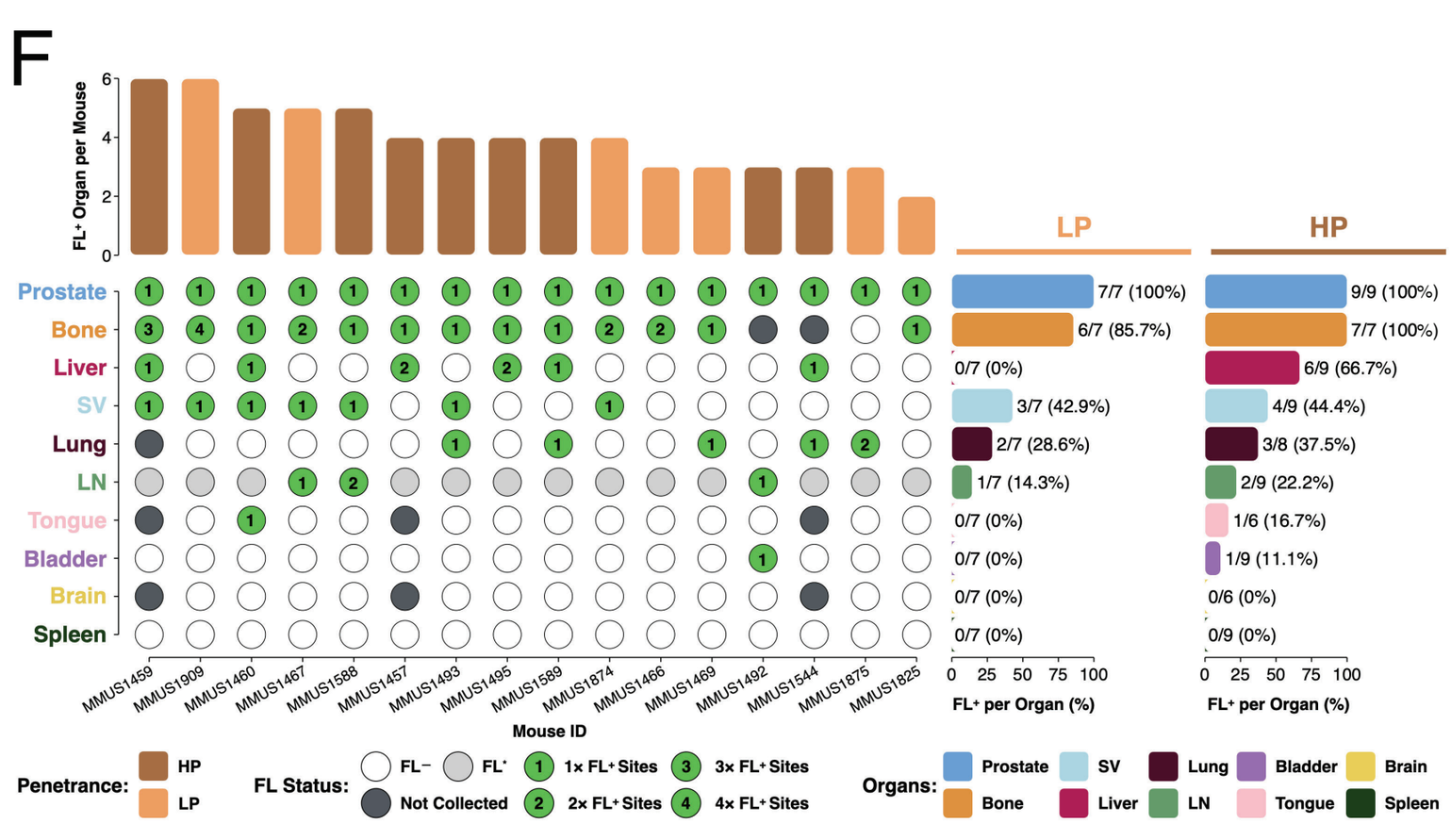
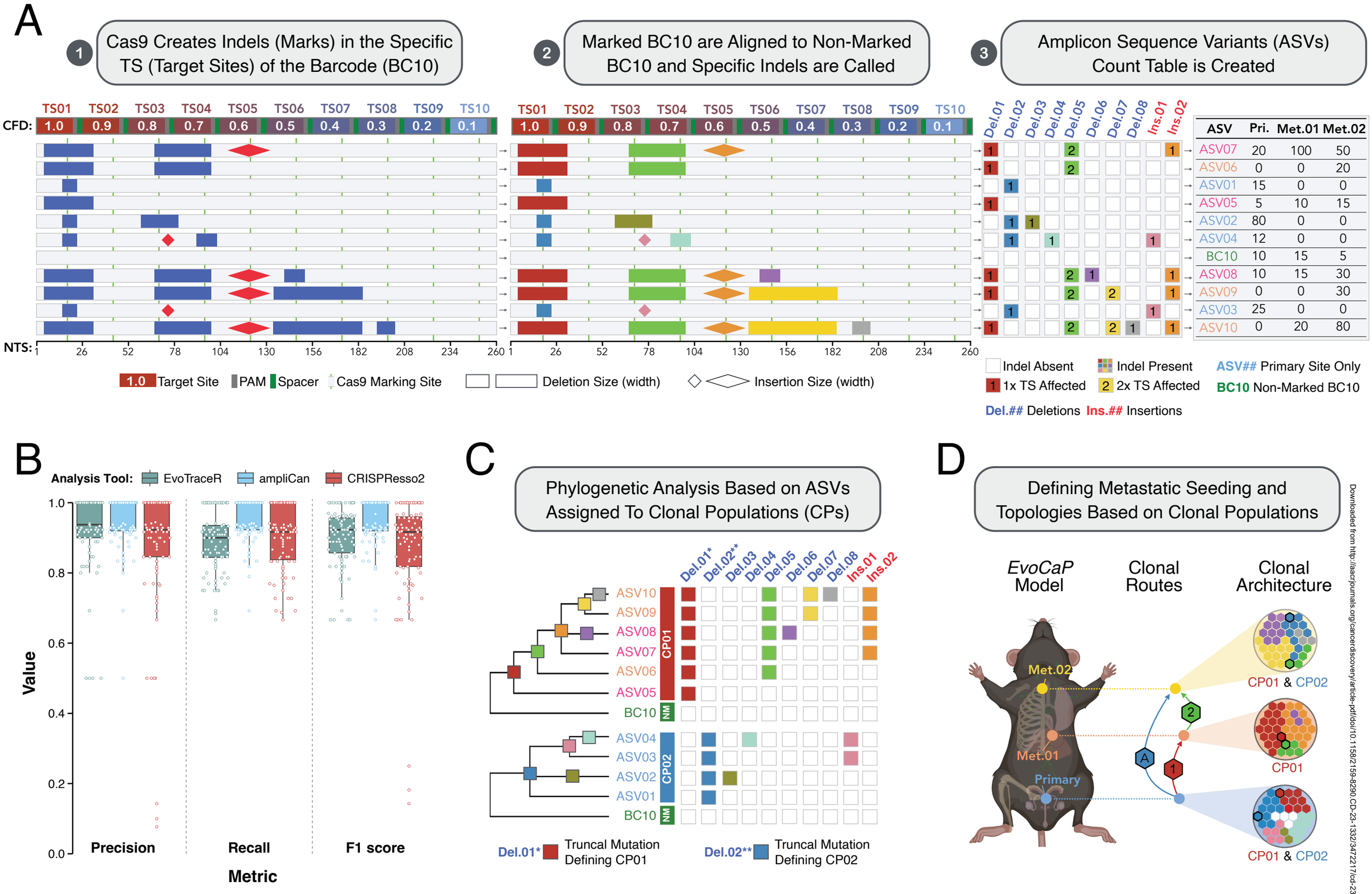


Figure 3.

A

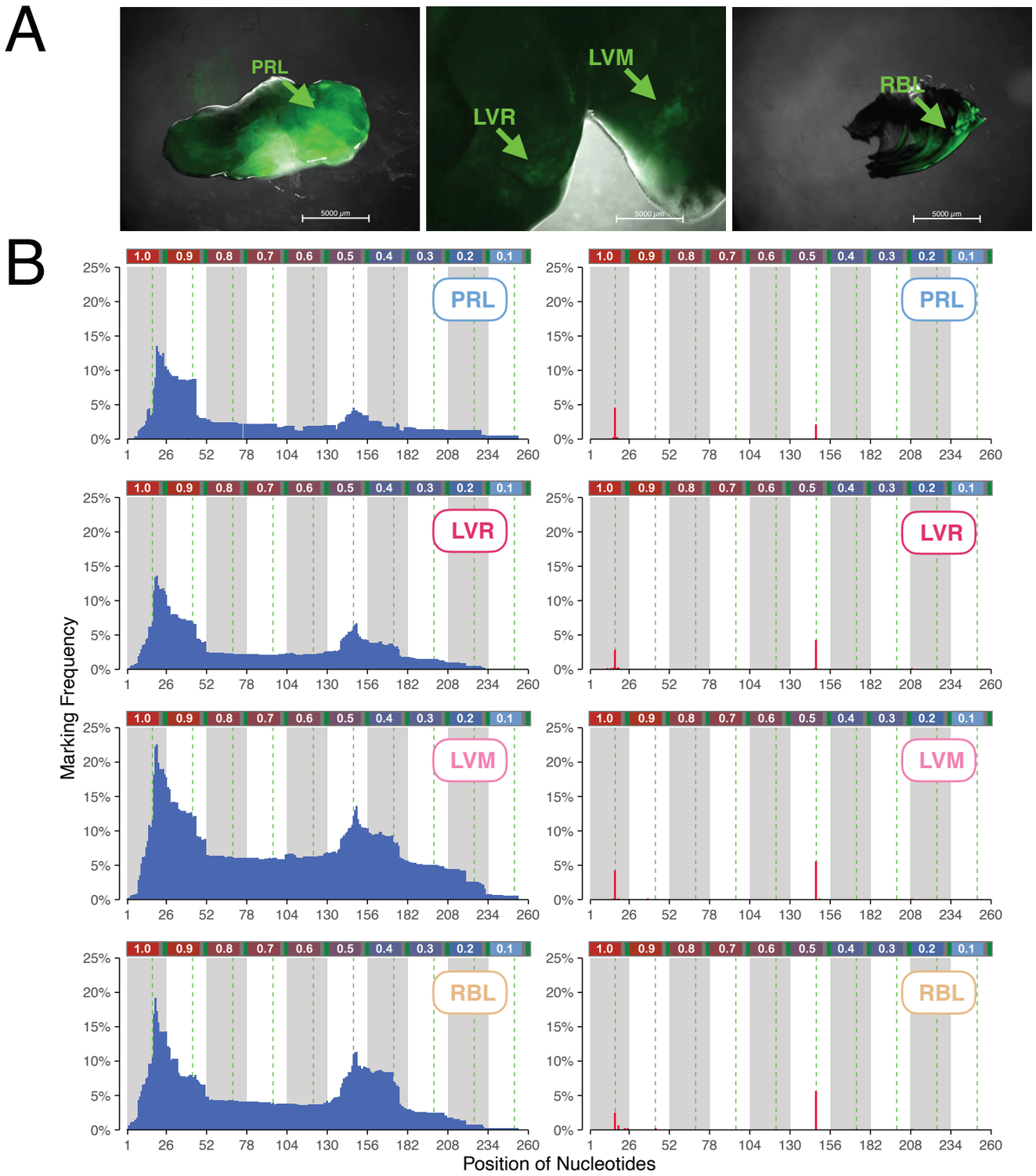


B

C

D

Figure 4.



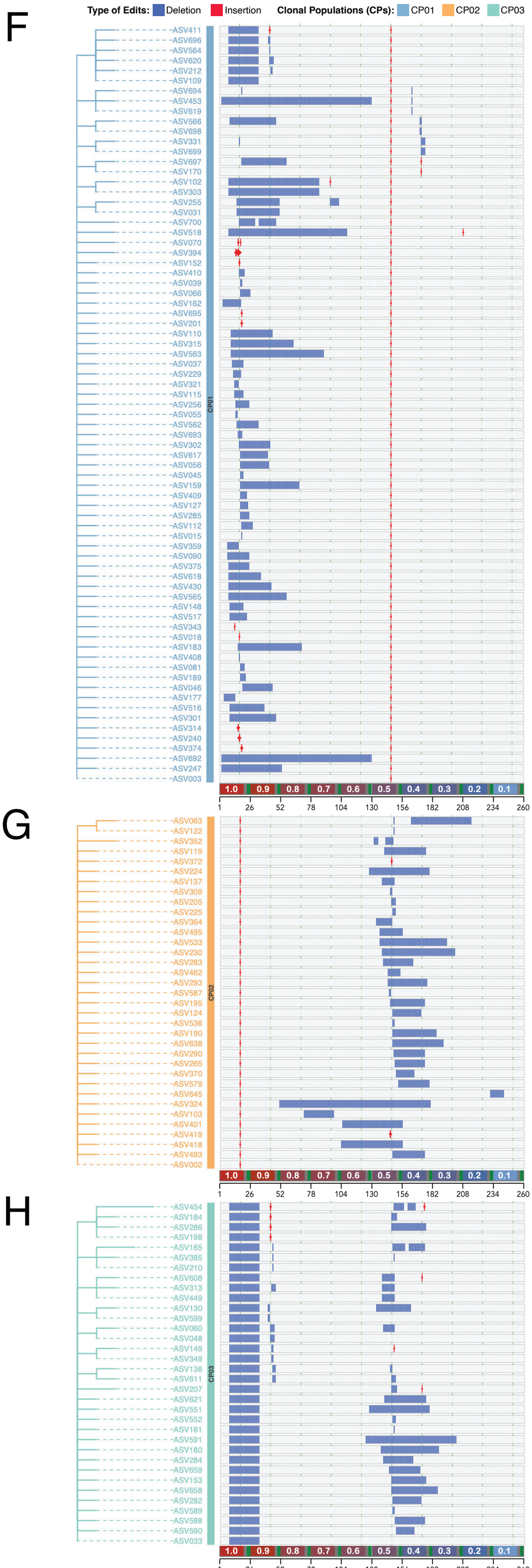
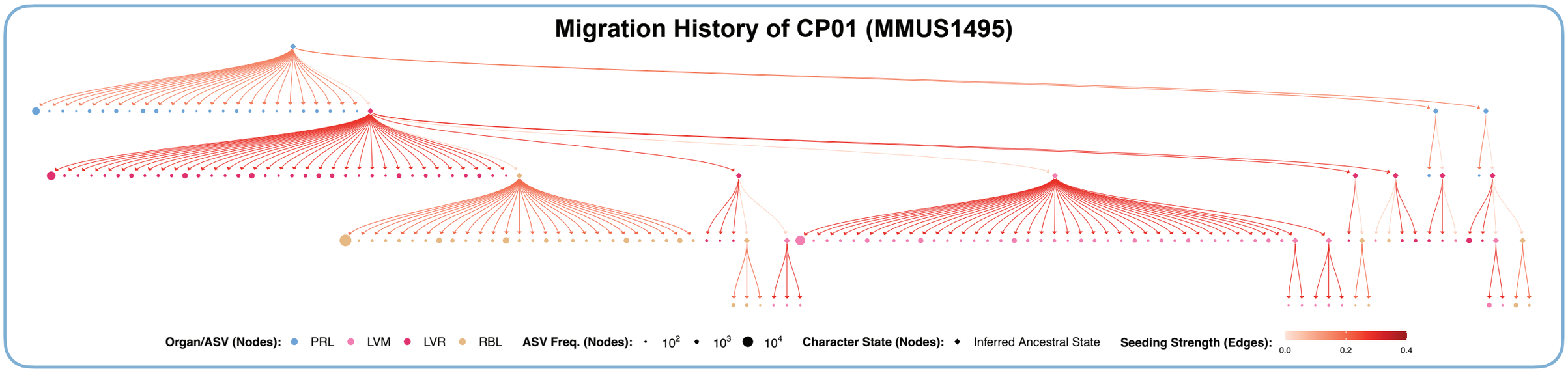
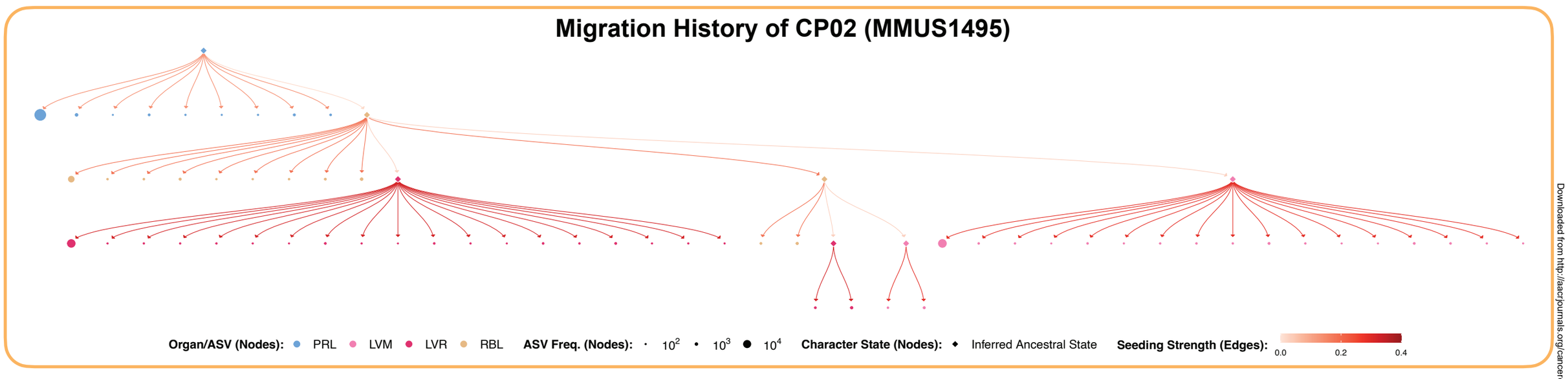


Figure 5.

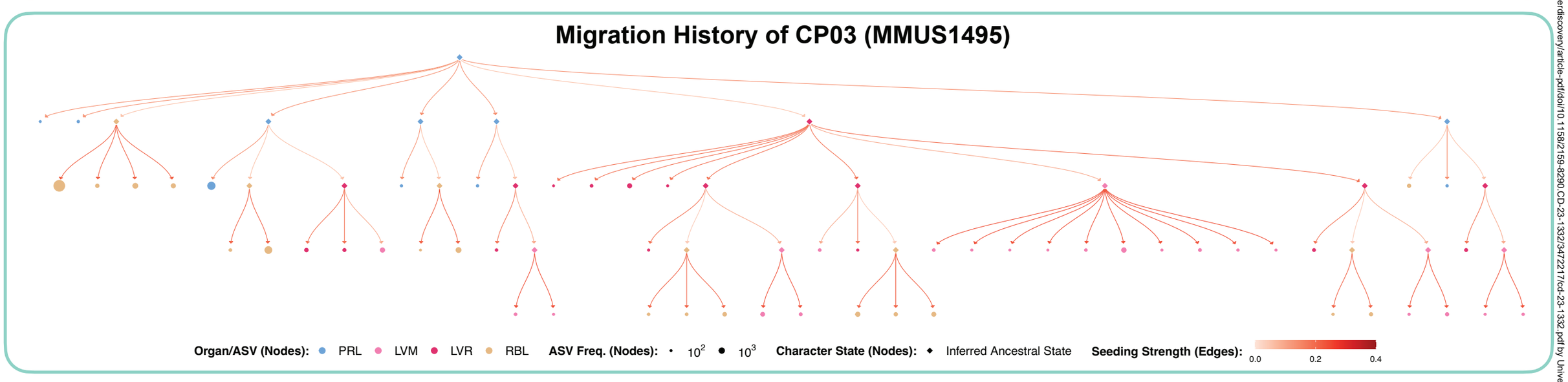
A



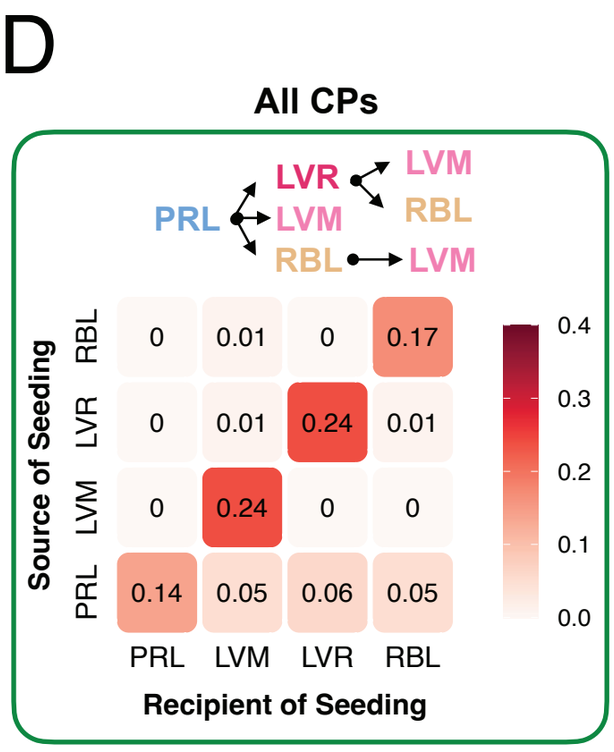
B



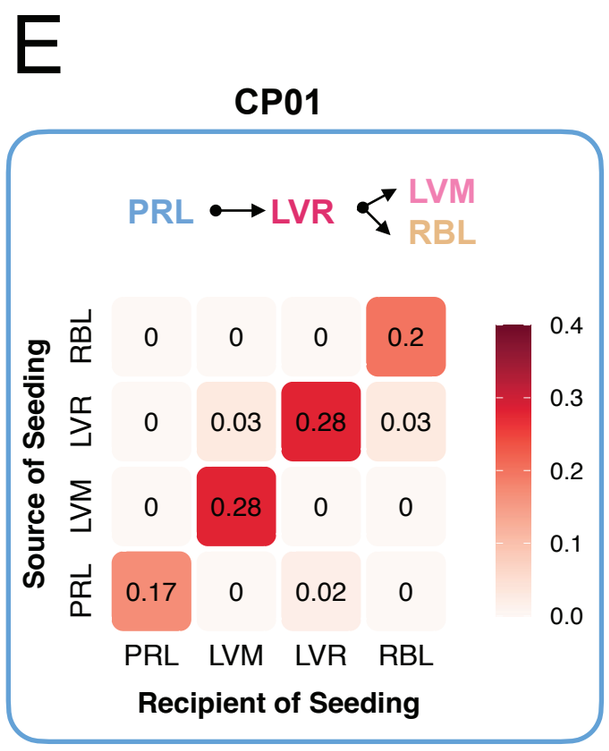
C



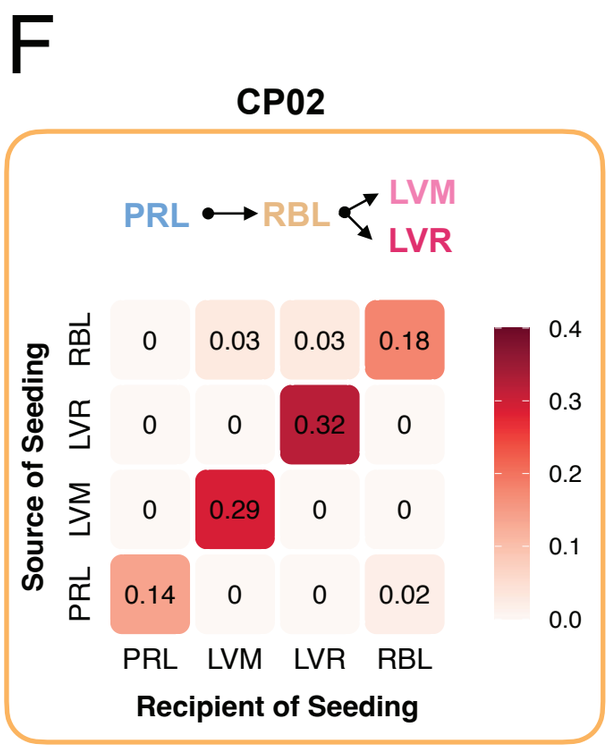
D



E



F



G

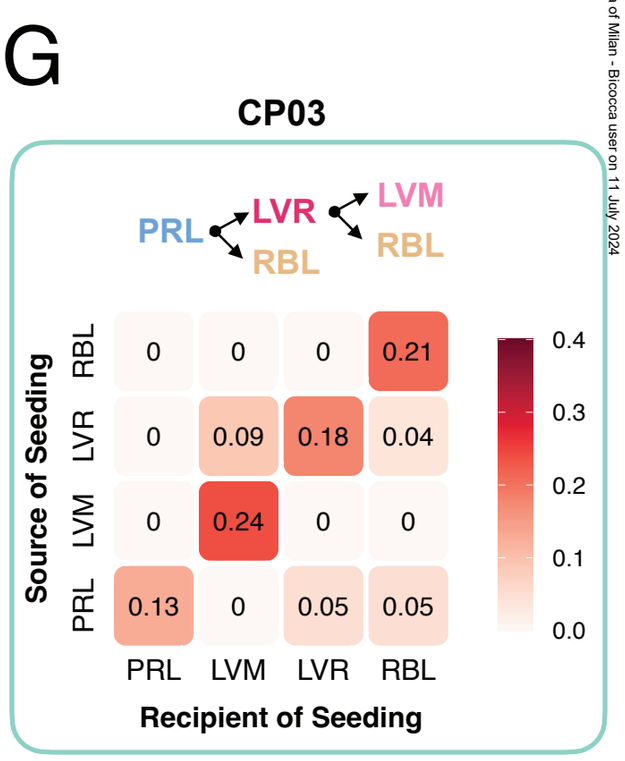


Figure 6.