DNA barcoded competitive clone-initiating cell analysis reveals novel features of metastatic growth in a cancer xenograft model

Syed Mohammed Musheer Aalam[1,†], Xiaojia Tang[2,†], Jianning Song[1,†], Upasana Ray[1], Stephen J. Russell[3], S. John Weroha[4,5], Jamie Bakkum-Gamez[6], Viji Shridhar[1], Mark E. Sherman[7], Connie J. Eaves[8,9], David J.H.F. Knapp[1,10,‡], Krishna R. Kalari[2,‡] and Nagarajan Kannan [1,11,12,*,‡]

[1]Division of Experimental Pathology and Laboratory Medicine, Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA, [2]Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA, [3]Department of Molecular Medicine, Mayo Clinic, MN 55905, USA, [4]Department of Oncology, Mayo Clinic, Rochester, MN, USA, [5]Department of Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic, Rochester, MN, USA, [6]Division of Gynecologic Oncology Surgery, Department of Obstetrics and Gynecology, Mayo Clinic, Rochester, MN, USA, [7]Department of Quantitative Health Sciences, Mayo Clinic, Jacksonville, FL, USA, [8]Terry Fox Laboratory, British Columbia Cancer Research Institute, Vancouver, BC, Canada, [9]Departments of Medical Genetics and School of Biomedical Engineering, University of British Columbia, Vancouver, BC, Canada, [10]Institut de Recherche en Immunologie et Cancérologie, and Département de Pathologie et Biologie Cellulaire, Université de Montréal, Montreal, QC, Canada, [11]Mayo Clinic Cancer Center, Mayo Clinic, Rochester, MN, USA and [12]Center for Regenerative Medicine, Mayo Clinic, Rochester, MN, USA

ABSTRACT

A problematic feature of many human cancers is a lack of understanding of mechanisms controlling organ-specific patterns of metastasis, despite recent progress in identifying many mutations and transcriptional programs shown to confer this potential. To address this gap, we developed a methodology that enables different aspects of the metastatic process to be comprehensively characterized at a clonal resolution. Our approach exploits the application of a computational pipeline to analyze and visualize clonal data obtained from transplant experiments in which a cellular DNA barcoding strategy is used to distinguish the separate clonal contributions of two or more competing cell populations. To illustrate the power of this methodology, we demonstrate its ability to discriminate the metastatic behavior in immunodeficient mice of a well-established human metastatic cancer cell line and its co-transplanted *LRRC15* knockdown derivative. We also show how the use of machine learning to quantify clone-initiating cell (CIC) numbers and their subsequent metastatic progeny generated in different sites can reveal previously unknown relationships between different cellular genotypes and their initial sites of implantation with their subsequent respective dissemination patterns. These findings underscore the potential of such combined genomic and computational methodologies to identify new clonally-relevant drivers of site-specific patterns of metastasis.

GRAPHICAL ABSTRACT
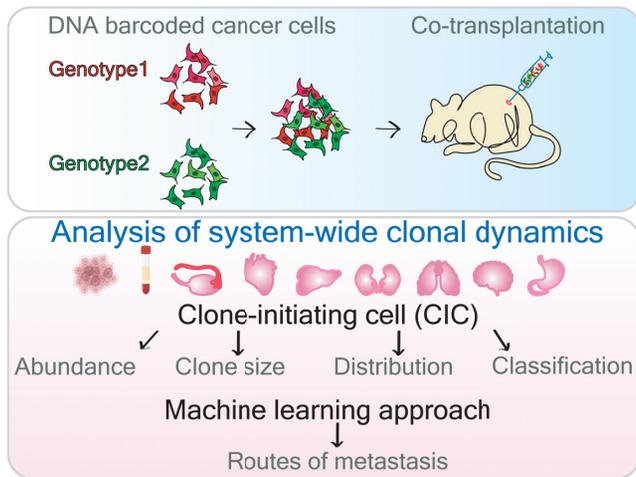
INTRODUCTION

Metastases are the primary cause of cancer fatality, and a primary aim of adjuvant systemic treatments is to eliminate cancer at metastatic sites (1,2). However, the complex processes that determine the metastatic behavior of individual cells within human tumors and their ability to invade and proliferate in heterotopic sites remain poorly understood. Indeed, it seems that the relevant factors may differ significantly within as well as between patients, confounding the effectiveness of current precision therapy approaches (1). Thus, much interest is focused on developing new methods to analyze the metastatic process at a clonal level across different target organ sites to facilitate the identification and causes of their differential behavior and ultimate therapeutic vulnerabilities.

Cellular DNA barcoding is emerging as a powerful next-generation sequencing (NGS)-based technology for tracking clonal growth properties displayed *in vivo* and *in vitro* wherein the genome of single cells are uniquely labelled with semi-random nucleic acid sequences which serves as barcodes (3–9). Several 'proof-of-concept' studies have already shown the ability of this technology to delineate at a clonal level, the heterogeneity of normal stem cell (5,10–12) and cancer cell growth properties (3,4,13–18) and resistance of the latter to defined therapeutics (7,8,13,14,19). However, a lack of standardized analytical and visualization tools, and methods for reporting DNA-barcoded cell data continue to pose significant challenges to the widespread use of this methodology (5,10,12,20,21).

To develop a methodology for clonal investigation of metastatic spread, we chose a human epithelial carcinoma cell line, OVCAR5, established from the ascitic fluid of an untreated patient with progressive ovarian adenocarcinoma, as a transplantable metastatic cancer model (22). OVCAR5 cells overexpress claudin-4 and carry a homozygous G12V mutation in the KRAS oncogene that are key to the aggressive growth properties they display *in vitro* and in xenograft models (23,24). We therefore chose this model to assess the capacity of DNA-barcoded cells to provide informative 'competitive clone-initiating cell (CIC)' data under conditions anticipated to alter metastatic activity. Given a recent report that expression of type-1 transmembrane LRRC15 in primary tumors is associated with an increased frequency of bowel and omental metastasis (23), we designed and analyzed the results of competitive transplant experiments in which OVCAR5 cells with high endogenous expression of LRRC15 were paired with OVCAR5 cells in which LRRC15 had been suppressed. The results demonstrate the utility of this methodology to quantify the metastatic spread of xenografts of malignant human cells at clonal resolution and the related effects of their genetic perturbation on their clonal patterns of growth at two different sites of injection and their consequent sites of metastasis.

MATERIALS AND METHODS

OVCAR5 cancer cell line derivatives

OVCAR5 cells were obtained from Fox Chase Cancer Center, Philadelphia and were maintained using RPMI-1640 media supplemented with 10% FBS and 1% penicillin–streptomycin. LRRC15 knockdown was performed in the OVCAR5 cells with shLRRC15 (Sigma-Aldrich) targeting the 3′UTR [Sequence: GCTATGAAAGAGAGAAGGAAA] using standard transfection guidelines and reagents. Whole cell lysates of OVCAR5-shControl and OVCAR5-shLRRC15 cell lines were subjected to western blot analysis against LRRC15 (ab150376) and GAPDH (sc-47724) antibodies. The blots were visualized using fluorophore-conjugated secondary antibodies (LICOR) and scanned by LI-COR OdysseyFc Imaging System (Nebraska, USA).

Cellular DNA barcoding

OVCAR5-shControl and OVCAR5-shLRRC15 cell lines were barcoded by lentiviral infection with EGFP_index 1.4 and Ruby3_index 1.2 libraries respectively using 0.8 μg/ml polybrene. After a 24-h incubation, the virus containing medium was removed and cells were cultured for an additional 24-h in fresh medium. To ensure that the majority of cells were labeled with a single barcode per cell, for lentiviral infection we used a predetermined dilution of virus (1:500–1:1000) corresponding to infectivity rate <37% based on Poisson statistical modelling. The cells were dissociated using 0.05% trypsin and sorted using stringent gating procedures. The barcoded OVCAR5-shControl (also referred to as LRRC15+) and OVCAR5-shLRRC15 (LRRC15–) cells were sorted and mixed in equal numbers (up to 1 million cells) in PBS, loaded on cell injection pipettes and placed on ice until injection. A small portion of sorted cells were verified for purity by FACS analysis. Further, 500–1500 sorted cells were also plated into 6-well plates in 2D clonogeneic assay. After 10 days, colonies were fixed with cold methanol/acetone (1:1) and stained with Giemsa stain. The clonogenic frequency was enumerated and the colony was imaged using Cytation 5 imager and colony area measured using Biotek's Gen5 software.

Ovarian and oviductal xenotransplantation

The oviductal and ovarian xenotransplant studies in female NSG mice were specifically approved for the study by the

Mayo Clinic Institutional Animal Care and Use Committee (IACUC, Protocol # A00004713-19). Surgery was performed on one mouse at a time. Mouse weight was measured to properly dose intraperitoneal ketamine and xylazine. Following anesthesia induction, fur was removed, and two incisions were made:(i) between last rib and hip and (ii) between dorsal side and abdomen. The fat pad linked to the ovary and oviduct was exposed using a stereomicroscope and ovarian fat pad was grabbed and held outside the abdominal cavity in order to minimize the contact with reproductive tract and avoid luteolysis.

Oviductal xenotransplantation. Under the stereomicroscope, the ampulla portion of the oviduct (identifiable as a highly coiled structure) was placed horizontally on the surgical platform and a small transversal cut nearly three fourth of the circumference on the distal part of ampulla was made using a micro-scissor. The glass needle was slowly inserted through the cut site in the direction of uterus and cell suspension in trypan blue buffer was injected. Few air bubbles were also injected to indicate successful injection.

Ovarian transplantation. The characteristic reddish ovary was held using forceps. Subsequently, a cell injection pipette was inserted into the bursa of ovary and cells were injected. After removal of the pipette, compression exerted by forceps prevented fluid regurgitation or leakage into the peritoneal cavity. The ovary and oviduct were returned to the abdominal cavity and muscle and skin was closed using absorbable sutures. Ibuprofen (40 mg/kg, 0.2 mg/ml of water) was given as analgesic in drinking water 48 h before and after the surgical procedure.

Bioluminescence imaging

The tumor growth and volume were non-invasively monitored using Xenogen IVIS optical imaging system. Briefly, animals were anesthetized using 2% isoflurane in an induction chamber and injected with 0.2 ml (150 mg/kg) of luciferin substrate through intraperitoneal route using 25-gauge needles. Subsequently, the animals were placed on a pre-warmed stage in IVIS imaging chamber and 2% isoflurane anesthesia was maintained during the course of procedure via nose cone. Living Image® Software from Xenogen was used for image acquisition. The animals were returned to respective cages following the procedure and observed for recovery.

Preparation of 'spike in' controls

A small library of 17 known barcode (Sanger sequenced) plasmids of index 1.6 was created and individual plasmids were quantified using Nanophotometer. The copy number/ng of plasmid DNA was calculated from the size the of plasmid (11.1 kb); ∼1.25 ng of plasmid DNA corresponds to 1.04×10^8 copies. Subsequently, each plasmid was serially diluted and combined to obtain a plasmid pool with 1–10,000 copies for sequencing reactions.

Barcode amplification

At the endpoint, primary tumor and other organs were harvested and snap frozen. Genomic DNA was extracted from the frozen tissues (10 mg) with a MasterPure Complete DNA and RNA purification Kit (Lucigen). We used Q5® High-Fidelity 2X Master Mix (NEB) to amplify the barcode sequence for NGS by introducing Illumina adaptors and 5-bp-long index sequences. The sampling of sufficient template coverage was ensured by parallel PCR reactions. For each PCR reaction, 1 µg of genomic DNA was spiked with 'spike in' controls and used as a template. Labeling each sample with 1 of the 20 unique sample indexes (referred to as index 2) enabled us to multiplex and sequence up to 20 samples at once.

NGS barcode sequencing

Completed libraries were spiked with 10% commercially prepared PhiX library (Illumina) to increase base diversity for improved sequencing quality. Samples were sequenced at one sample per lane to generate ∼85–95 million total reads per sample. The single read flow cells were sequenced as 100 base single end reads on an Illumina HiSeq 2500 using TruSeq Rapid SBS sequencing kit version 1 and HCS version 2.0.12.0 data collection software. Base-calling is performed using Illumina's RTA version 1.17.21.3.

Barcode data processing and threshold determination

In the preprocessing step, the barcode sequences were extracted from the raw FASTQ files based on the sequences of library index (index 1) and sample index (index 2) at either ends of the amplicon. The number of sequences sharing the same unique barcode were then counted and barcodes from the same library with no more than 2 mismatched base pairs were merged as these likely arose due to sequencing errors (5,11). A barcode count table was generated with respect to all identified clones with unique index 1, index 2 and barcode sequence combination. Raw counts were then normalized to fractional read values (FRVs) by dividing the raw clone counts by the total sample size: FRV = raw Counts/total raw counts in sample.

The CIC Calculator filtered the noise using the dedicated spike-in index 1.6 (GTCA) library in each sample as a ground truth. To do so, the true positive rate (TPR) and false positive rate (FPR) were calculated as below.

$$TPR = \frac{\# \, Spikein \, clusters \, after \, filtering}{\# \, total \, spikein \, clusters},$$

$$FPR = \frac{\# \, NonSpikein \, clusters \, after \, filtering}{\# \, total \, NonSpikein \, clusters} \quad (1)$$

The fractional read value (FRV) threshold was then determined based on the optimal accuracy from the receiver operating characteristic (ROC) curves. This was done either using the local spike-in controls (those from each individual sample) or based on a pool of spike-in controls from all samples (global). For our downstream analyses, the global setting was used. Small barcode clones with FRV lower than the threshold were filtered out. Next a linear fit was performed between the $\log_{10}$(FRVs) of the spike-in controls and the known input copy numbers (ignoring single copy inputs). The resulting fit was then used to convert FRVs into absolute cell counts which could be used in subsequent analyses.

Analyses of clonal patterns

All unique clones appearing above detection threshold were first identified for each mouse and all sites with detected values for each identified. As we know that all remaining sites these clones were below the threshold of detection, and many identical values can interfere with various tests, we interpolated values for each unique clone at below threshold sites. To do so we fit a left-censored Poisson distribution to each site for each mouse based on observed clones at that site. Lambda values between 0 and 100 were tested in increments of 0.1 and the value selected which minimized the sum of squared error. From this lambda value, we could then infer the expected frequency of cells at below threshold values. Clones were then assigned random cell numbers below thresholds with the probability based on the Poisson fit (i.e. filled in the below threshold values of the best-fit Poisson distribution). For those sites that lacked sequencing data due to a failure to amplify the barcode, we once again know that they must have below threshold levels. As no distribution could be fit, a uniform random distribution was used to assign below threshold values in these cases. Missing data for which tissues had not been collected were left as missing. These interpolated cell number matrices were used for subsequent analyses. Notably, this interpolation did not alter the relative relationships between clones (as assessed by their 2D patterns following dimensionality reduction), but simply eliminated technical artifacts in subsequent analyses caused by repeated 0 measurements. All analyses and statistics for this and following sections were performed in R (version 3.6.1).

To determine whether there was a relationship in how CIC types distribute across different metastatic sites, we first calculated pairwise Spearman's correlations between sites for each mouse based on the detected cell numbers of each clone at each metastatic site. These correlations thus represent an overview of the similarity of clonal outputs for each clone between the sites. These matrices were then converted to distance matrices based on the cosine theorem: $d = \sqrt{2(1 - \rho)}$. As an initial measure of whether these mice showed similar patterns (and thus it was reasonable to combine distance matrices), we performed a congruence among distance matrices test using the R package 'ape' (version 5.4.1). We next took a mean of the distance matrices across mice ignoring missing values to get the average pattern of relationships. This was also done for only ovarian injections, and for oviductal mice only, including only those sites which were measured in all mice in that set. Each of these methods was used to perform a classical multi-dimensional scaling (MDS). In order to obtain a measure of the biological variability of these patterns, we first calculated a principal component analysis (PCA) on the mean distance matrix (approximately equivalent MDS). The PCAs were then used to embed the matrices from each individual mouse into the same PCA space.

Given the apparent association of blood and ascites to specific metastatic sites, we next set out to estimate the relative importance of blood and ascites to the establishment of metastases in each site. For this, we analyzed the fit of a series of different multivariate linear regression models. In each case, a model was designed to predict one site given the cell numbers in Blood, Ascites, and the Primary Tumor as predictors: $\text{asinh}(\frac{site}{10}) = 0 + \beta_1 * \text{asinh}(\frac{Blood}{10}) + \beta_2 \text{asinh}(\frac{Ascites}{10}) + \beta_3 \text{asinh}(\frac{Primary\ Tumour}{10})$. Clones with missing measurements in any predictor or in the site being predicted were excluded from that given model. Adjusted R^2 and p-values were calculated for each fit, and a false-discovery rate was then calculated across all models/predictors to correct for multiple testing.

In order to identify common patterns of metastatic spread we next performed a clustering analysis on the overall dataset. This was done in several stages. First, we removed all clones which had less than 10 cells detected at all sites and assigned these to a low ('L') cluster. From the remaining clones, we randomly sampled 100 of these from each mouse and calculated a pair-wise Pearson's correlation between the asinh(cell #/10) of each cell. Where measurements were missing, the correlation was calculated between those cells including only the values measured in both clones (ie. pair-wise complete measurements). A *k*-nearest neighbor (KNN) graph was then constructed based with a maximum of 10 nearest neighbors per cell and including edges only if the Pearson's correlation was 0.9 or greater using the R packages 'igraph' (version 1.2.6). The Leiden clustering algorithm [25] was then run on the KNN graph with a resolution parameter of 0.5 to generate an initial set of clusters using the R package 'leiden' (version 0.3.6). Mean values for each site were then calculated for each cluster and cells which were not part of the initial clustering assigned to their most correlated cluster mean (pair-wise complete Pearson's correlation of the asinh(cell #/10)) if that most correlated cluster was at least 0.7. If not, these cells were left unassigned. Next, all clusters that were only detected in a single mouse and those with 10 clones or less were removed to eliminate possible artifacts. Finally, to minimize over-clustering, clusters with cluster means with a Pearson's correlation of 0.7 or more were merged. This resulted in a total of 17 clusters, a cluster of clones which were borderline detectable in all sites, and 295 unassigned clones. As a visual assessment of cluster relationships and cluster solution success, dimensionality reduction was performed on a distance matrix calculated from the pair-wise complete Pearson's correlation matrix of asinh(cell #/10) between all clones (converted as: $d = \sqrt{2(1 - r)}$). This was done both by classical MDS, as well as by UMAP using the R package 'umap' (version 0.2.7.0).

To identify the effects of LRRC15 and injection site on clone numbers both overall and for each cluster, a series of Aligned Rank Transform (ART) 2-factor ANOVAs were performed [26] using the R package 'ARTool' (version 0.10.8). In all cases, clone number was used as response, while genotype (LRRC15+ vs LRRC15–) and injection site (ovarian vs oviductal) and the possible interaction between them used as predictors. For comparisons of clone numbers across clusters, *P*-values were multiple testing corrected by false discovery rate (FDR).

As blood was not measured in most oviductal mice but was a very strong predictor of engraftment at distant metastatic sites, we performed an elastic net (EN) logistic regression to predict whether a clone would be present in the blood or not. This model used asinh (cell #/10) for all

sites measured in all mice (other than blood and ascites) as predictors for the binary variable of detectable or not in blood. To implement this model, we first split the data into training and validation sets, with 70% of the clones present in the blood from each of the ovarian mice included in the training set, and double that number of clones which were not present in the blood from those mice. This over-representation of blood clones was done to ensure that a sufficient number of positives were included in the training data and prevent specificity (as most clones were not in blood) from being the driving force of the scoring. The remaining 30% of blood clones and the remaining non-blood clones were kept back as a validation set. We then performed hyperparameter fitting with a fixed alpha value of 0.5 (to balance between model sparsity and grouping) with varying lambda and gamma regularization penalties using 5-fold cross validation with area under the receiver-operator curve (AUC) as the scoring parameter using the R package 'glmnet' (version 4.0.2). Overall prediction accuracy, sensitivity, and specificity was then scored on a per-mouse basis in the training and validation sets. Another model was then generated to predict whether a clone was present in ascites or not. In this case, the training set included 70% of clones present in ascites from the ovarian mice and an equal number (as ascites clones were more frequent than blood) which were not present in ascites. Hyperparameter optimization, training and validation was then performed as was done for blood.

CIC calculator

The CIC calculator was implemented with R and Python languages.

RESULTS

Multiplexed DNA barcoding libraries with spike-in barcodes

To generate a technology that would enable the clonal tracking of concurrently generated progeny of multiple sources of tumor populations, we first introduced into a MNDU3 plasmid lentivector, a red-shifted firefly luciferase (Luc) fused to the cDNA of either mRuby3 or EGFP, via a 'self-cleaving' P2A linker peptide sequence. A scaffold to facilitate insertion of an individual non-coding 37 bp semi-random barcode from a previously described plasmid library (10) was then inserted downstream of the Luc-2A-FP cassette. To expand the DNA barcode library repertoire, we added a unique 4 bp library index (referred as index 1.1, 1.2, 1.3, 1.4 or 1.5; Supplementary Figure S1A) to the 5′ end of the barcode sequences and adapted the forward and reverse primer binding sequences to our design (adopted from (7)) for PCR-based amplification of the barcode sequences and NGS library preparation of labelled cells. In this study, we assigned the library indexes 1.2 and 1.4 to OVCAR5-shLRRC15 (referred to as LRRC15–) and OVCAR5-shControl (LRRC15+) cell lines respectively (Figure 1A).

The possible theoretical estimate of unique barcodes in the 25 libraries thus created is 4×10^6. Transformed bacterial cell plating yielded values from 1 to 3×10^6 with a median diversity of 2×10^6 (Supplementary Figure S1B) and

barcodes isolated from 250 individual bacterial clones confirmed no sequence redundancy (data not shown). Titers of concentrated lentiviral supernatants ranged between 4 and 28×10^{12} (median of 21×10^{12}) transduction units/ml (Supplementary Figure S1B). To control variables associated with NGS analysis of test preparations and to set a threshold for data filtering, we also created a 1.6 indexed plasmid library of 17 known and Sanger sequence validated barcodes (Supplementary Figure S2A). These plasmids were then mixed at precise copy numbers to serve as internal calibration controls. This 'spike in' control barcode library was added to each sample at the time of starting the preparation of every test PCR-based amplicon sequencing library (Supplementary Figure S2B). At the time of amplification, a unique sample index was added per sample (index 2.1 to index 2.20). This approach enabled 20 test samples to be multiplexed for sequencing on an Illumina HiSeq 2500 sequencer (Supplementary Figure S2B), to decrease sequencing costs.

Design of the clone-initiating cell (CIC) calculator

The flow chart of the CIC calculator is shown in Figure 1B. Briefly, indexes were first identified from each read, unique barcodes from each library were identified and counted, and barcodes from the same library with ≤2 mismatches were pooled (5,11) to identify the yield counts per unique index (∼500K unique barcodes/flow cell). In practice, we found low frequency repeated reads (<10) accounted for only 0.5% of the total labelled cells (total read counts). Receiver-operator characteristics (ROC) based on the known 'spike-in' controls were then used to filter out such likely technical artifacts (Figure 1C). As predicted by Poisson sampling, the reads obtained for the known barcode copy numbers in the spike-in samples gave a log-linear relationship except for the single copy inputs (Figure 1D). Removal of the latter gave an improved correlation with known barcodes, and sensitivity (Figure 1E, Supplementary Table S1).

This CIC Calculator also takes advantage of two different steps for thresholding clone sizes: one at the level of the sample, and a separate one at the global level of clone analysis across all samples being analyzed (see flowchart, Figure 1B). The local sample thresholding calculates the clone size threshold using only the spike-in in the specific sample of interest, which is more accurate and specific to the sample. The global sample thresholding, in contrast, pools all spike-in control barcodes across separate groups of samples, to obtain a shared threshold for all samples. This is particularly useful for generating ROC curves if individual samples fail for technical reasons. The two strategies can be set by the user before execution.

Overall growth of OVCAR5 LRRC15+ and LRRC15- cells *in vitro* and in co-transplanted mice

To test the utility of the above barcode system, we chose to apply it to compare the growth of a well-studied human ovarian cancer cell line (OVCAR5) in competition with a derivative we engineered that we anticipated would reduce its metastatic activity in transplanted immunodeficient mice. For the latter we chose to knockdown LRRC15,
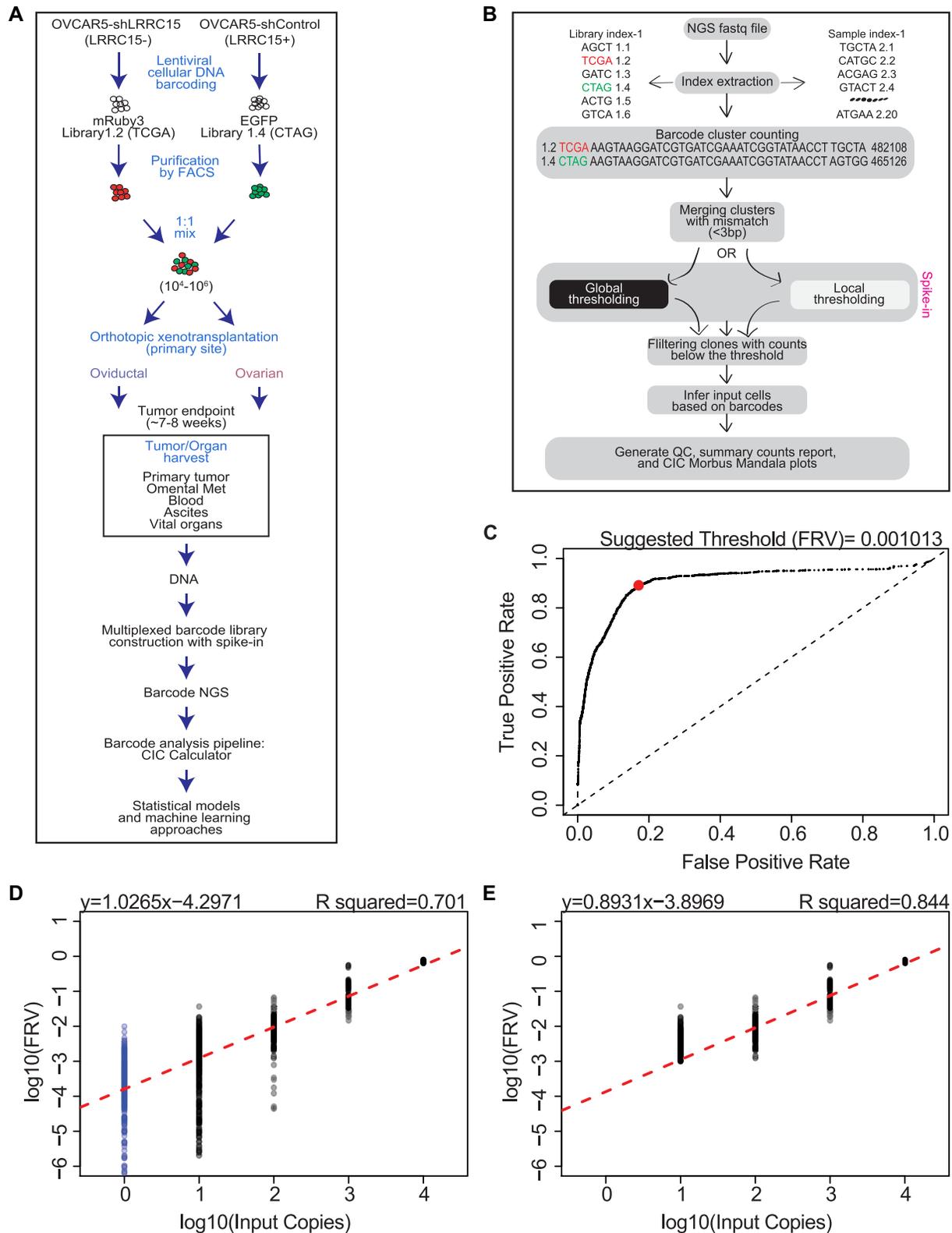
Figure 1. Barcoded competitive clone-Initiating cell (BC-CIC) assay and clone-initiating cell (CIC) calculator. (**A**) experimental design for BC-CIC assay. LRRC15+ and LRRC15- were tagged with EGFP _index 1.4 and Ruby3_index 1.2 libraries and FACS sorted. The FACS sorted cells were mixed in equal numbers and co-transplanted at oviductal or ovarian sites in NSG mice. Mice were monitored for tumor development till study endpoint. At necropsy primary tumor and tissues/organs were harvested and DNA was extracted. Multiplexed sequencing libraries were prepared with 'spike-in' control barcodes by PCR and subjected to NGS. The data was analyzed using CIC Calculator. (**B**) flow chart showing the data deconvolution steps following NGS sequencing. (**C**) receiver operator characteristics (ROC) based on the known 'spike-in' controls were used to filter out artifacts. (**D**) plot showing log-linear relationship between the known copy numbers (1–10,000) in the spike-in control barcodes and sequencing reads. (**E**) removal of single-copy spike-ins improved the R2 values.

as its expression has been strongly associated with bowel metastases in ovarian cancer patients (23,24). The two shRNA LRRC15 knockdown constructs namely shRNA-1 and shRNA-2 when transduced in OVCAR5 cell line led to suppression of LRRC15 protein levels compared to a non-targeting shRNA control or parental cells (Supplementary Figure S3Ai-ii). We used shRNA-1 transduced cells (referred as LRRC15–) as a genetic model of LRRC15 suppression. *In vitro* colony-forming assays of FACS-purified LRRC15+ and LRRC15– cells isolated 48 hours post-transduction showed no effect of LRRC15 knockdown on their immediate colony-forming efficiency, although the colonies derived from the LRRC15- cells were noticeably smaller than the LRRC15+ control cells (Supplementary Figure S3B–E). We then compared the tumorigenic activity of the co-injected LRRC15+ and LRRC15– cells <2 h post-FACS purification (ratio 1:1) into ~8 week old NOD/SCID/IL2rγ−/− (NSG) mice either under the ovarian bursa (ovary, $0.06–1.25 \times 10^6$ total cells into each of 10 mice) or within the ampullary oviducts (oviduct, $0.10–1.25 \times 10^6$ total cells into each of 10 mice) (Supplementary Table S2) (27). All animals survived the surgical procedure and bioluminescence imaging showed 100% tumor take at both injection sites (Supplementary Figure S4A) and confirmed at necropsy.

Both groups of NSG mice co-injected LRRC15+ and LRRC15– cells either oviductally or directly into the ovary developed distended abdomens, reduced ambulatory activity and body weight with evidence at autopsy at 7–8 weeks post-injection of bulky disease (median survival of 48 days for both groups (Supplementary Figure S4B, C). The groups also showed similar bioluminescence signals (Supplementary Figure S4D), tumor (Supplementary Figure S4E), and organ weights (Supplementary Figure S4F–M). Gel electrophoresis analysis of PCR products of genomic DNA from primary tumors, omental metastatic tumors, ascites, blood and other organs confirmed the presence of widespread metastases in both groups (data not shown). All ovarian and oviductal injected mice showed multiple macro-metastases (Supplementary Figure S5).

Detection of different CIC distributions of co-transplanted LRRC15+ and LRRC15– cells

We next asked whether the spectrum of separately analyzed CICs in similarly combined equal transplants of LRRC15+ and LRRC15– cells would also be similar or different. Genomic DNA was then isolated from various anatomical sites including the primary tumors from all 20 transplanted mice. We proceeded with barcode sequencing of various sites in a subset of mice (seven ovary and five oviduct transplants). Amplicon libraries from barcode signal 'positive' sites, independently verified by gel analysis (data not shown), were prepared with 'spike in control barcodes', multiplexed and sequenced. Clones were then identified and analyzed using the CIC Calculator strategy (Figure 1B). The results showed a highly significant decrease in clonal contribution of the LRRC15- cells at both primary injection sites (Figure 2A), as well as at their various metastatic sites (Figure 2B–E). For example, the mean number of LRRC15+ CIC in primary sites (ovary) was 89.5 ±

3.23, as compared to 7.80 ± 1.21 for the LRRC15- cells ($P < 0.0005$ paired student t-test, Figure 2D). The data also showed sporadic incidences of LRRC15– CIC numbers higher than LRRC15+ CIC in organs such as bowel mets (Figure 6B). Similarly, for the oviduct injection group, the calculated clone sizes (corrected for total tissue/organ weights) identified large differences in the largest clones produced by the LRRC15– and LRRC15+ cells (Supplementary Figure S6). For example, in mouse #48, the largest clone produced from a LRRC15– cell contained 45,000 cells in contrast to a clone of 1.2×10^6 LRRC15+ cells found in a kidney of the same mouse (Supplementary Table S3). Taken together, the size as well as the CIC numbers dominating the primary and metastatic sites of tumor growth demonstrated a superior fitness of CICs dependent on LRRC15 expression (Supplementary Figure S6, Table S3).

The CIC Calculator was also used to generate radial chord diagrams, also called Morbus–Mandala (MM) plots. These provide a visual comprehensive description of the multiple individual clonal trajectories detected, the inter-relationships revealed in their respective metastatic behavior, and hence the complex relationship patterns exhibited by different CICs (Figure 3). The examples shown here illustrate both CICs from multiple or individual experiments conducted in a 'single' mouse as well as system-wide activities of individual clones (Figure 3A, B). To simplify analysis of the different metastatic behaviors represented in the MM plots, we categorized each CIC according to the activity it displayed as follows: (i) CICs of indeterminate potential (clones detected only in the primary site), (ii) CICs of mono-metastatic potential (clones detected in only one metastatic site as well as the primary site); (iii) CICs of pluri-metastatic potential (clones detected in more than one, but not in all metastatic sites) and (iv) CICs of pan-metastatic potential (clones detected in all metastatic sites). Examples of clones representing each of these categories are shown in Figure 4A and their complete distributions in Figure 4B–D and Supplementary Table S4. These distributions highlight the predominance of LRRC15+ clones in the disseminating disease, but with notable inter-mouse differences in metastatic behavior of clones emanating from the two primary sites tested (Supplementary Figures S7 and S8).

Identification of unique and complex patterns of LRRC15+ and LRRC15– CIC dissemination and growth

Interestingly, in contrast to the lack of a difference in the overall metastatic distribution of either LRRC15+ or LRRC15- clones, the representation of specific clones in the different sites examined showed reproducible differences and these were consistent between both primary injection sites (Congruence Among Distance Matrices [CADM] test $P = 0.001$, Figure 5A and B, Supplementary Figure S9A and B). These patterns included one group that appeared generally confined to organs within the peritoneal cavity (i.e. adjacent kidney, contralateral ovary, and omentum, with ascites), and another group of distal sites (i.e. salivary gland, brain, and heart, with blood), and others that we classified as intermediate (i.e. lung, spleen, liver, contralateral kidney) (Figure 5A and B). Importantly, these relationships were highly consistent across individual mice as indi-
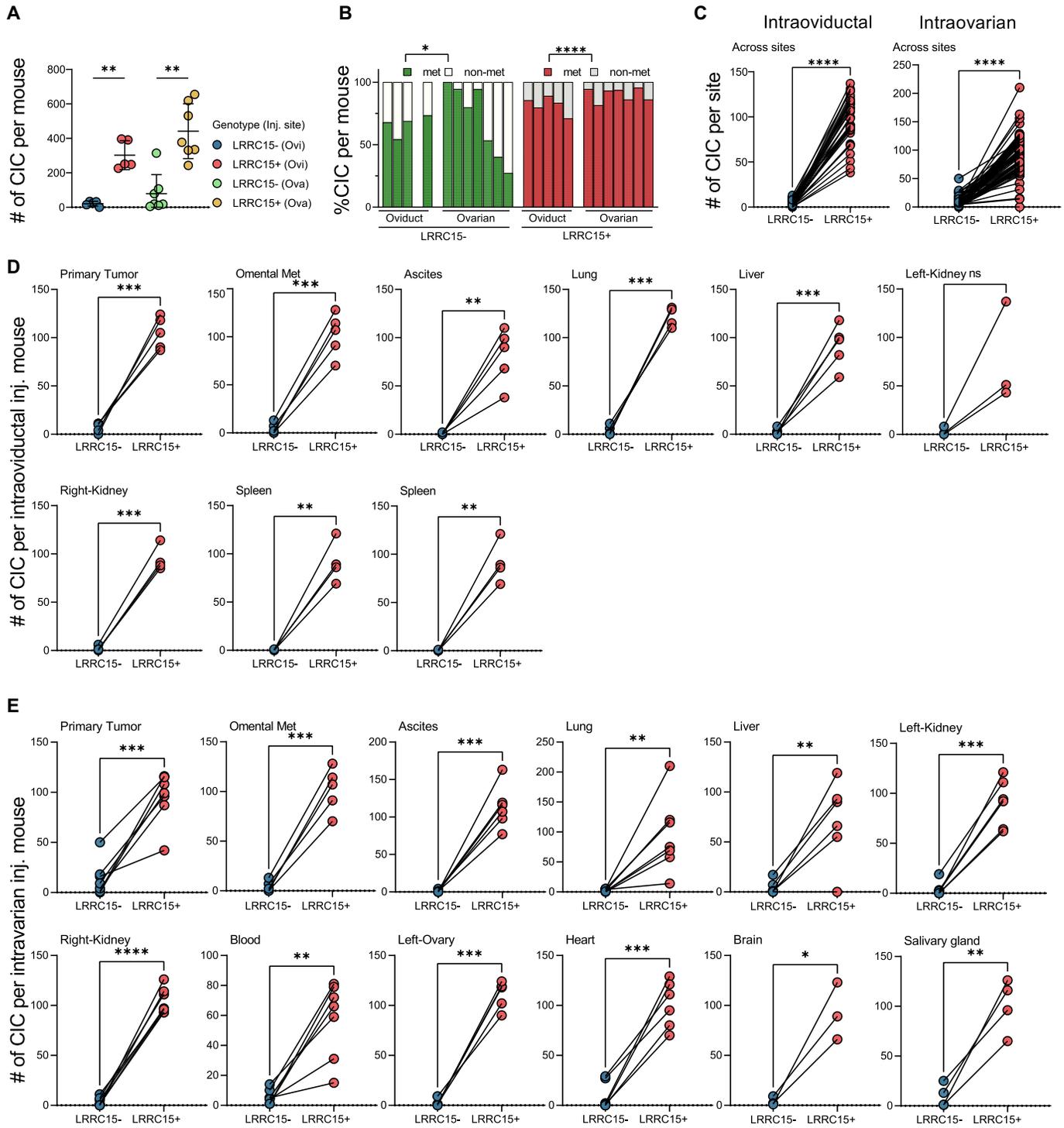
Figure 2. CIC numbers across sites are influenced by LRRC15 gene. (**A**) total CIC per mouse split by LRRC15 status in both groups. (**B**) percentage of metastatic and non-metastatic CICs per mouse split in both groups. CIC clones were calculated based on clone counts. Groups in panel A and B were compared using one-way ANOVA and Tukey test, **$P < 0.005$ and ****$P < 0.0001$. (**C**) CIC numbers across sites in both groups. (**D**) CIC numbers in each site in oviductal co-injected group. (**E**) CIC numbers in each site in ovarian co-injected group. LRRC15– and LRRC15+ groups were compared using paired two-tailed t-test, *$P < 0.05$, **$P < 0.005$, ***$P < 0.0005$ and ****$P < 0.0001$. See also, Supplementary Figure S5.
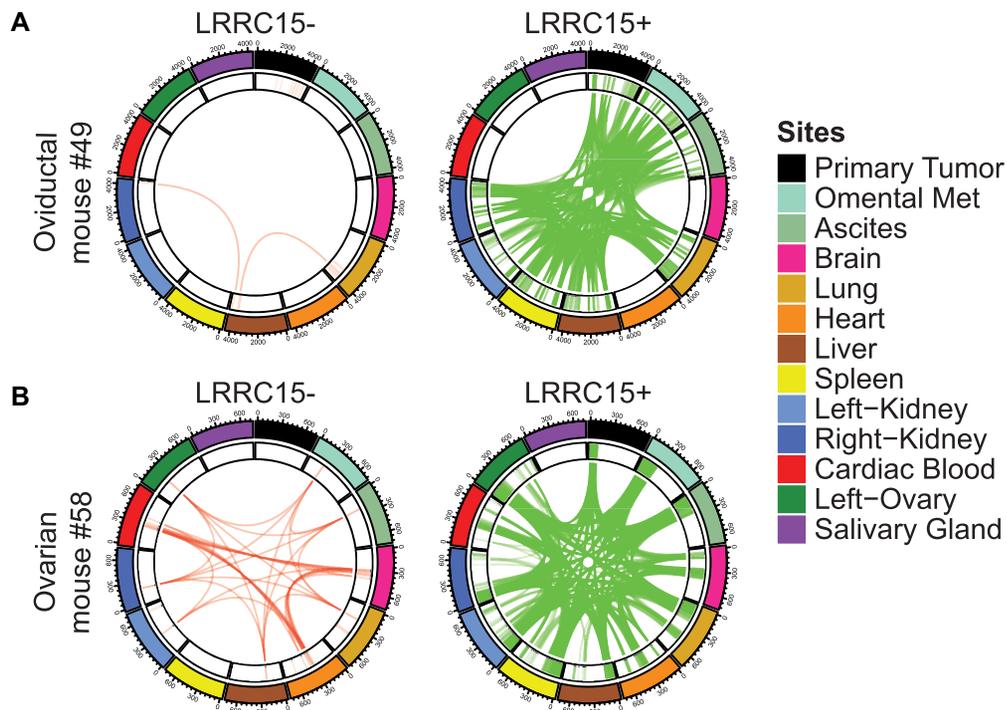
Figure 3. Morbus Mandala (MM) plots pictorially depict system-wide metastasis. (**A**) a representative MM plot depicting system-wide metastasis of CICs in mouse#49 co-injected with barcoded LRRC15+ and LRRC15- cells in oviductal site. (**B**) a similar representative MM plot of CICs in mouse #58 co-injected in ovarian site. See also, Supplementary Figures S6 and S7.

cated by the CADM test (Figure 5B, Supplementary Figure S9C and D).

Since clone frequencies in the blood were positively correlated with distal metastases, with a similar correlation between their presence in ascites and peritoneal sites, we hypothesized that these patterns might be indicative of distinct mechanisms of metastasis. To quantify the relationship between blood and ascites to each of their preferred metastatic sites, we constructed a series of linear models using clone abundance in the two paired sites and also the primary tumor as an outgroup to predict spread likelihood to the non-associated metastatic site. This showed distal sites were strongly and almost exclusively predicted by blood CICs, and similarly, peritoneal sites were predicted mainly by corresponding CICs in the ascites (Figure 5C; Supplementary Figures S9E and S10; Table S5). In spite of the fact that cells were injected in right ovary, clone presence in the blood strongly predicted left (FDR 9.58E–148) than right (FDR 4.17E–69) kidney metastasis likely owing to ovarian venous drainage variations, i.e. left ovary's venous drains into the left renal vein and the right ovary's venous drainage directly into the inferior vena cava (Figure 5C; Supplementary Table S6). Together these data accounted for half of the variability in metastatic activity observed (r^2 in the 0.4–0.6 range).

To determine whether the CIC Calculator could be used to detect previously inaccessible site-specific patterns of clonal metastatic activity, we focused on the bowel metastases, associated with lethal bowel obstruction (23), generated in mouse #61 initially co-injected with 0.12×10^6 cells at the ovary site and organs harvested 7 weeks later. MM

plots generated from the CIC detected in the 8 bowel metastases identified in this mouse showed the LRRC15+ clones in all eight sites were derived from CIC.Pluri cells, whereas those derived from the co-injected LRRC15– cells reflected the activity of CIC.Pluri in only one of the eight bowel sites and CIC.Mono were active in six of these (Figure 6A; Supplementary Figure S11). Moreover, the metastatic distributions of all the CIC.Pluri clones, irrespective of their LRRC15 genotypes, were nearly identical (Supplementary Figure S11). In addition, analysis of all the clone sizes in these eight sites indicated that both LRRC15+ and LRRC15– clones were equally competitive in most (Figure 6B, C). Overall, these findings indicate how a genetic change may alter the competitive metastatic growth ability of cells in site-specific manner.

To obtain a more comprehensive picture of the spectrum of metastatic clone behavior of the cells tested in this study, we next analyzed the metastatic distributions on a per-clone basis. Cluster analysis identified distinct patterns (labelled as Clusters I to XVII) of clonal behavior that were reproducible across similarly injected mice. Visualization of the identified clusters using multidimensional scaling (MDS) and uniform manifold approximation and projection (UMAP) for dimensionality reduction confirmed that the identified clusters were relatively distinct and compact (Figure 7A). Notably, CICs behavior patterns ranged from poorly metastatic (Cluster I) to highly metastatic with high levels at blood-associated sites but without blood detection (possibly due to earlier blood-based dissemination (Cluster X; Figure 7B). Cluster IV type CICs displayed peritoneal
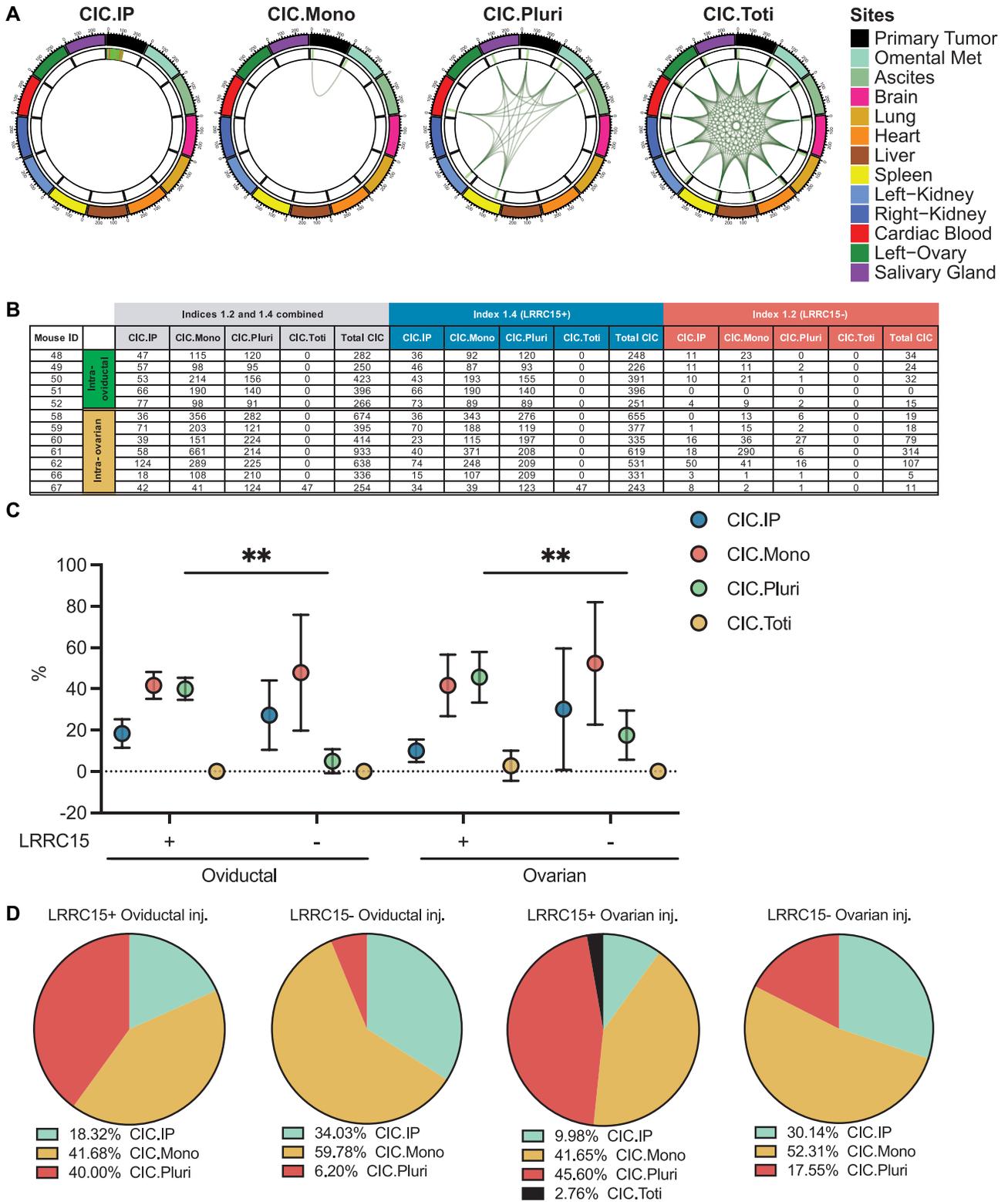
Figure 4. CIC classes and genetic influence. (**A**) representative MM plots of CIC.IP, CIC.Mono, CIC.Pluri and CIC.Toti. (**B**) Table summarizing CIC classes by clone sizes based on LRRC15 status and injection sites. (**C**) CIC numbers per class split by LRRC15 status and injection site. Comparison between LRRC15+ and LRRC15– classes within oviductal and ovarian groups was done using two-way ANOVA and Tukey test, **$P < 0.005$. (**D**) Pie charts showing % of CIC classes. Note: CIC.Toti class was observed only in LRRC15+ CICs in a single ovarian injected mouse.
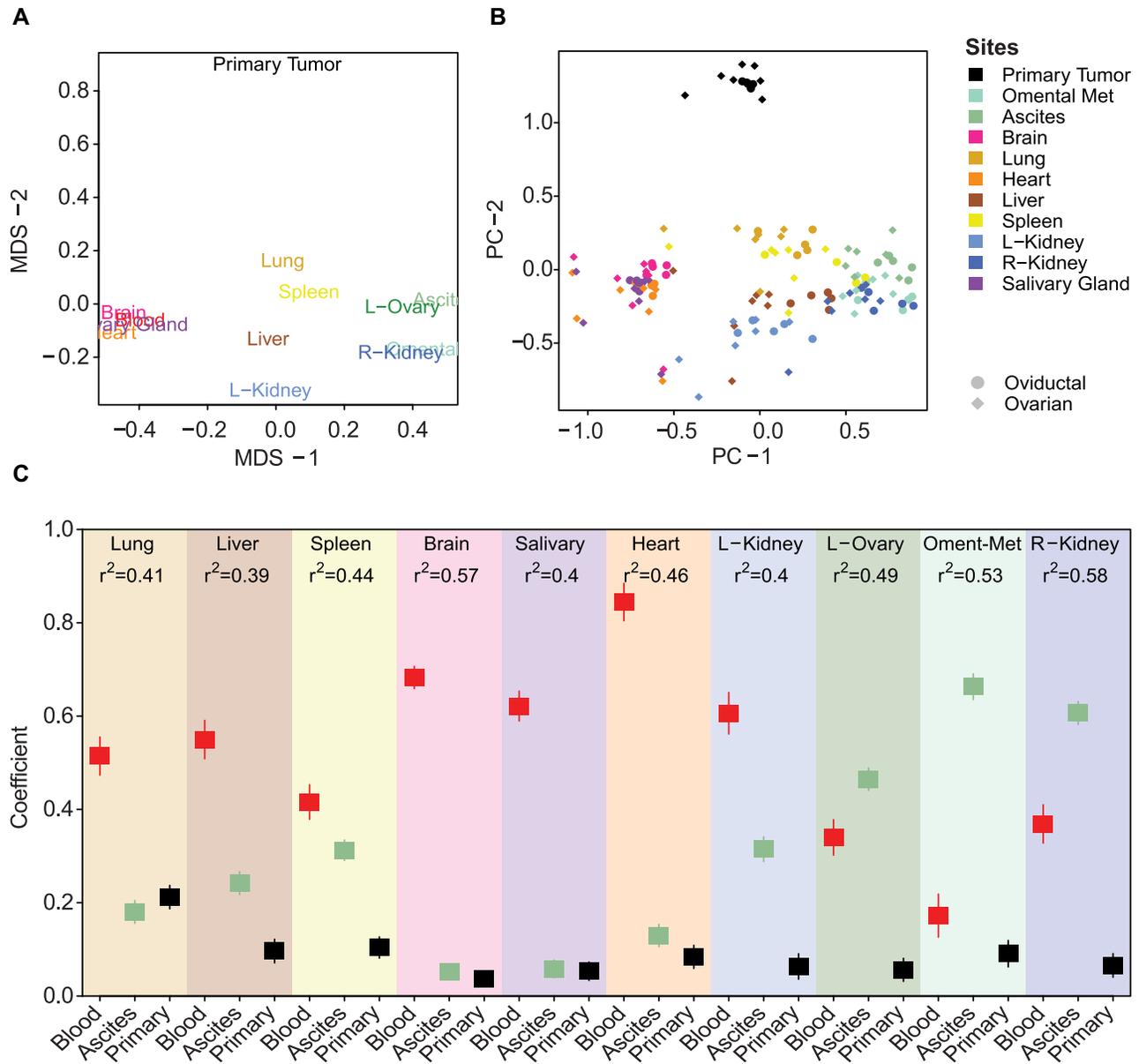
Figure 5. Clones detected in blood are associated with metastasis to different organs than those detected in ascites. (**A**) relative relationships between sites of tumor growth. Multi-dimensional scaling (MDS) dimensionality reduction based on the mean distance matrix across mice is shown. Site names displayed at their embedded locations. The closer together sites are displayed on the plot, the more correlated their growth. (**B**) per-mouse relationships between sites of tumor growth. The plot shows a PCA embedding based on the averaged distance matrix across mice (approximately equivalent to MDS) followed by embedding the per-mouse relationships into this space as a measure of how consistent these relationships are across individual animals. As with the MDS plots in (A), the closer the points, the more similar is their patterns of clonal growth. Injection site is indicated by point shape and site of growth is indicated by color as shown in the legend. Only sites measured in all mice are shown in this plot. (**C**) coefficients of are shown for a series of multivariate linear regression models using clone size detected in the blood, ascites, and primary tumor to predict levels in each other site (all asinh(x/10) transformed to maintain linearity). Higher coefficients (values on the y-axis) represent a greater predictive power of that site for the overall clone size at the indicated site. Lines showing 2× the standard error are shown for each coefficient. See also, Supplementary Figures S8 and S9.

spread with significant enrichment for LRRC15+ clones that had been injected into the oviduct (ART ANOVA FDR = 0.01 for injection site, 0.004 for LRRC15 status, and 0.004 for the interaction of the two, Figure 7B-C). Cluster VIII CICs with detection in blood and distal sites was also significantly enriched in similarly injected mice (ART ANOVA FDR = 0.02, Figure 7B, C, Supplementary Table

S6). Most of the LRRC15− CICs generated relatively small clones (Figure 7C; ART ANOVA FDR = 0.01). Multiple patterns characterized by large clonal outputs at a single metastatic site suggested their origin from an early emigrant from the primary site (Supplementary Figure S10). A few notable patterns of clonal metastasis are shown in Figure 6C (see also Supplementary Figure S12, Table S6).
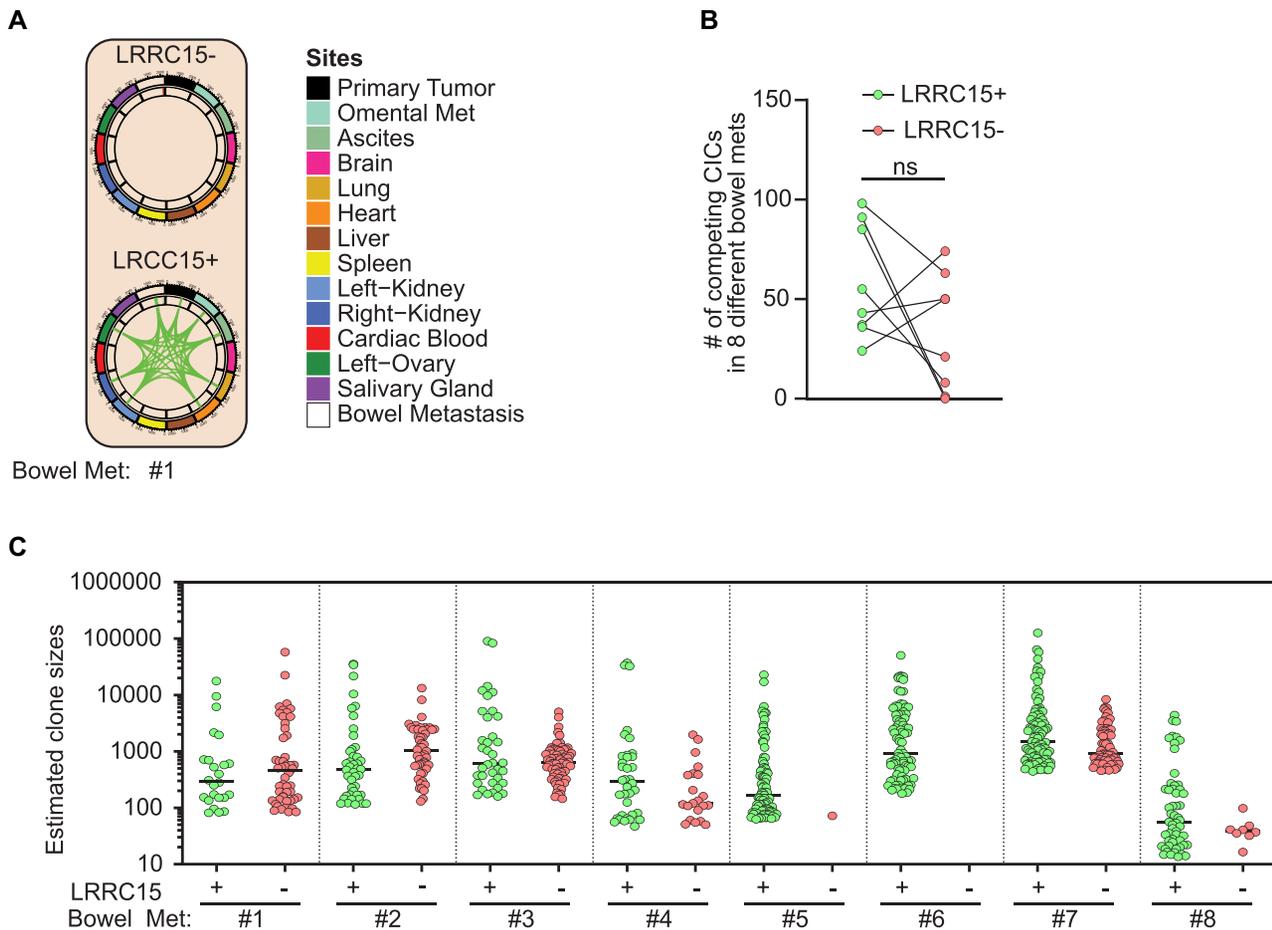
Figure 6. Clonal analysis of bowel macro-metastases (bowel met) and their system-wide linkages. (**A**) Representative MM plot of harvested macroscopic bowel metastasis from mouse #61 depicting LRRC15– and LRRC15+ clones respectively. (**B**) plot showing number of LRRC15+ and LRRC15– CICs in bowel mets. (**C**) Estimated clone sizes across in LRRC15+ and LRRC15– bowel metastasis. See also, Supplementary Figure S10.

Use of machine learning to identify patterns of CIC dissemination

To investigate the utility of machine learning to predict CIC dissemination patterns, we used an elastic net (EN) model to predict on a per clone basis whether or not it would be detectable in the blood. Following training on a randomly selected subset of the clones generated in recipients of cells injected into the oviduct, this model showed an overall per mouse median prediction accuracy of 0.92, with a sensitivity of 0.71 and a specificity of 0.95 when tested on the remaining clones that had not been included in the training set (Figure 8A, Supplementary Figure S13A, B). One of the benefits of an EN model is that the final parameters contributing to each decision are explicit, unlike more powerful but less interpretable deep learning approaches. In our case, the EN model was able to show that the primary indicators of a clone being detectable in the blood were its presence in the brain and heart (Figure 8B), two of the distal sites associated with blood from both the bulk (Figure 5), and the Cluster VIII data (Figure 7C). Finally, when applied across all mice, we observed a significantly lower frequency of clones predicted to be in the blood for recipients of cells injected into the oviduct as compared to the ovary

(Wilcoxon rank-sum test $P = 0.01$, Figure 8C). This finding was further supported by the measured clone numbers in the cells injected into the ovary compared to the 0 observed and 0 predicted in the cells injected in the single oviduct mouse for which blood measurements were made (Supplementary Figure S13B). This suggests that injection into the ovary may have conferred a greater ability to execute blood-mediated metastasis.

We also trained an EN model to predict whether a clone would be detectable in ascites (Supplementary Figure S13C). However, this model proved less accurate with a median accuracy of 0.82, and a sensitivity of 0.40, and specificity of 0.86 in the corresponding validation set (Supplementary Figure S13D). This decreased accuracy could have been caused by an earlier ascites-mediated spread than would be detected by an end-point measure. Predictions in this case could also be impacted by the fact that blood-borne cells have the potential to reach many more sites via different routes, and hence decrease the prediction accuracy. Nevertheless, the model was still able to show that omental metastasis and the local kidney represented the primary predictors of ascites detection (Supplementary Figure S13E). While no significant differences were observed in as-
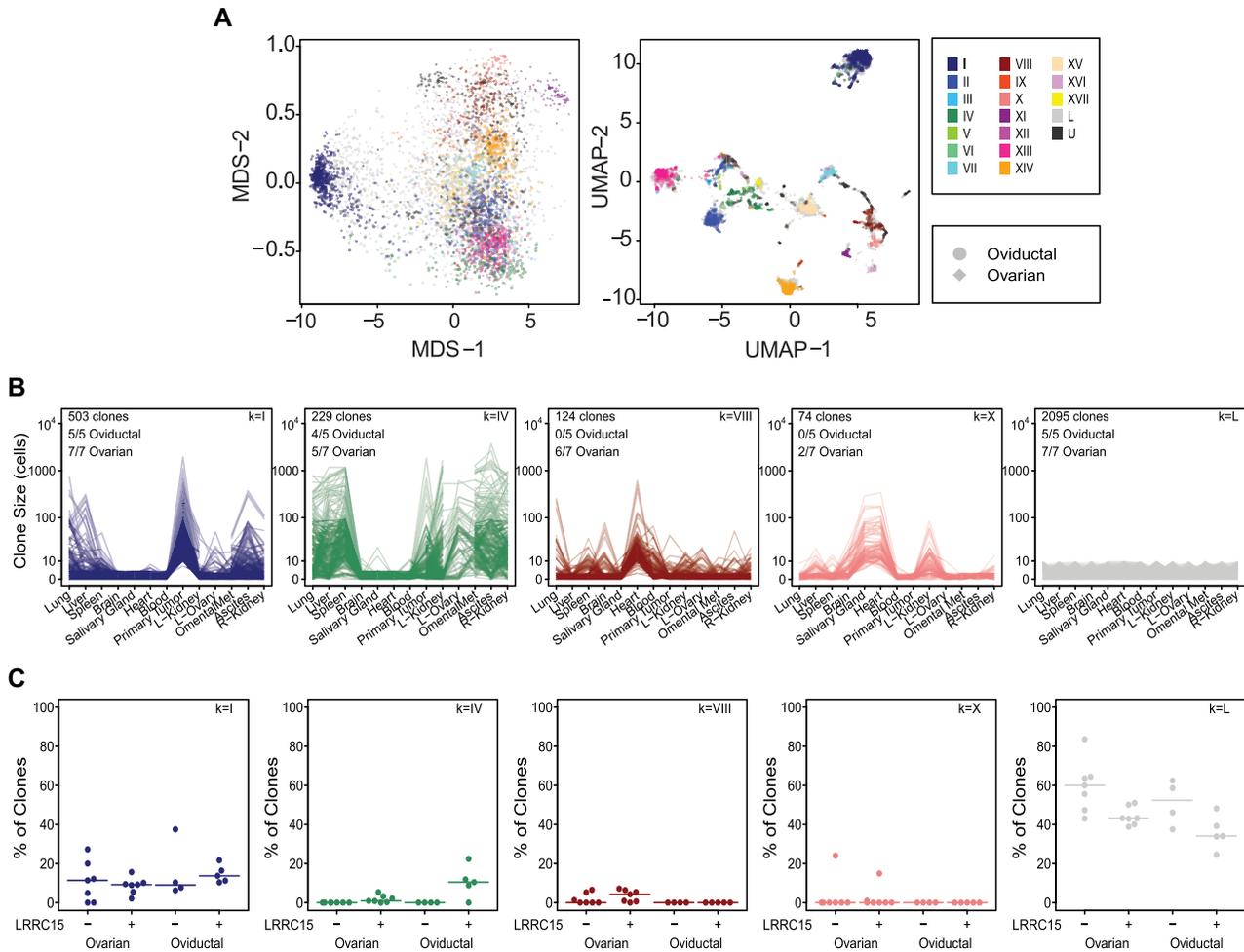
Figure 7. Common patterns of clonal growth exist and are influenced by injection site and LRRC15 status. (**A**) dimensionality reduction by MDS (left) and UMAP (middle) based on the pair-wise distance transformed Pearson's correlations between all clones. Injection sites are indicated by point shape, and cluster membership is indicated by point color as indicated in the legend (right). (**B**) clonal engraftment patterns are shown for selected clusters. Lines connect cell numbers detected for each clone between sites. Missing values for a given clone are shown as gaps between adjacent line segments. Clone numbers are asinh(cell #/10) transformed with the linear numbers indicated. The total number of clones which were part of a cluster, and the number of mice with at least one clone that was a member of that cluster for each injection site are shown in the top left. The cluster identifier for each is shown in the top right. (**C**) the percentage of detectable clones which are part of selected clusters is shown for each mouse separated by injection site and LRRC15 status ('–' = LRRC15 shRNA, '+' = LRRC15 wild-type). See Supplementary Figures S11 and S12 for engraftment patterns and injection site/LRRC15 comparisons for remaining clusters.

cites by injection site at the time of measurement (Wilcoxon rank sum test $P = 1$, Figure 8D), the predicted ascites clones were significantly higher in the primary tumors arising from cells injected into the oviduct (Wilcoxon rank sum test $P = 0.01$, Supplementary Figure S13F), an observation in keeping with the observed enrichment of Cluster IV (peritoneally-confined) tumors derived from cells injected into the oviduct (Figure 7D).

Perhaps the most interesting results from the EN models come from a comparison of their coefficients. Omental metastasis was a counter predictor of blood detection (Figure 8B), while brain and heart countered predicted ascites spread (Supplementary Figure S13E). As these were in both cases the primary positive predictors for the other type (brain and heart predicted blood, gut predicted ascites), this finding implies that ascites and blood likely represent independent routes of metastasis.

DISCUSSION

We report here several new features of metastatic dissemination properties made possible by the combined use of cellular DNA barcoding to track clonal growth of competing malignant transplant populations and a computational approach to derive complex relationships between the clonal composition and patterns of dissemination of cells in both the primary and metastatic lesions produced. The new features of the barcoding methodology used include a simplified illustration of system-wide complex spatial linkages of clones using MM plots and two self-reinforcing methods to characterize clone dissemination patterns. One of these allowed thousands of CICs to be subclassified into four general groupings (referred to as CIC.IP, CIC.Mono, CIC.Pluri and CIC.Toti). The other made use of a combination of dimensionality reduction and clustering to infer additional biological explanations for the observed dissem-
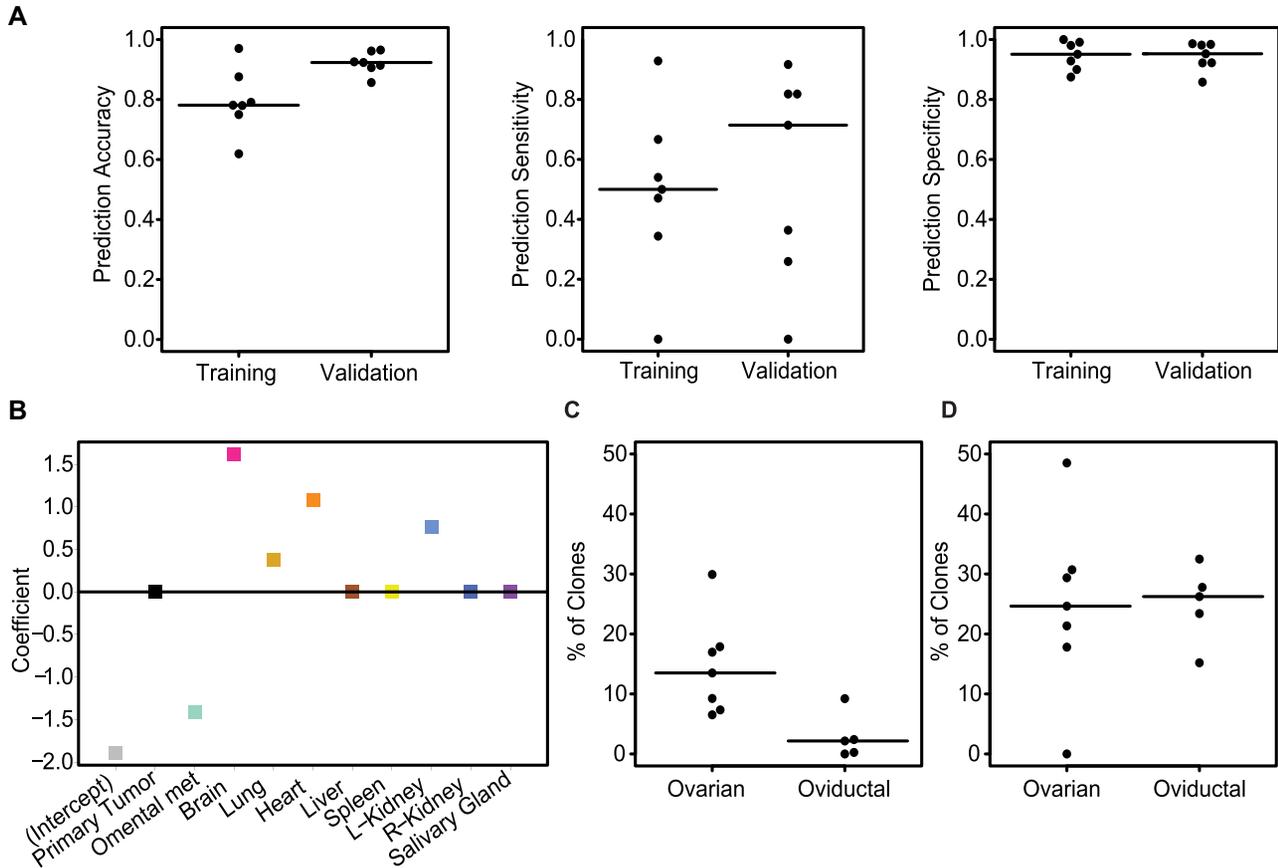
Figure 8. Primary tumors in the oviduct are deficient in blood metastasis than those that arise in the ovary. (**A**) overall accuracy (left), sensitivity (middle) and specificity (right) of an elastic net (EN) logistic regression to predict whether a given clone is present (detectable) in blood. The EN model was trained on 70% of blood detectable clones from each mouse in the ovarian cohort and double that number of non-blood clones with remaining clones kept back for the validation set. Each point represents the predictions for clones from one mouse. Each plot shows the relevant metric for the training and validation sets. (**B**) the coefficients retained for each site in the final EN model for predicting whether a clone will be detectable in blood or not and their magnitude. Positive numbers indicate a value that is used to predict that a clone is present in blood, while negative numbers indicate sites in which the presence there means the clone is less likely to be in blood. The negative intercept value represents the threshold of positives needed to indicate a clone is likely present in blood. (**C**) percentage of clones per mouse predicted to be detectable in blood by the EN model separated by injection site. (**D**) percentage of clones per mouse with measured detection in ascites separated by injection site. See Supplementary Figure S12 for model hyperparameter fitting, measured percentage of clones detectable in blood for those mice which had measurements, and an EN predictive model for detection in ascites.

ination patterns. We also demonstrate how the CIC Calculator output data can be integrated with a suite of complementary statistical and machine learning approaches that use CIC numbers, frequencies, and system-wide clone size distributions to infer clone classes, and their system-wide metastatic properties. We expect the sample indexing and multiplexing approach described here will enable inclusion of a more extensive range to understand how clones exploit different tissue environments. Further adoption of these analytical tools should facilitate comparison of findings from independently generated clonal barcode studies.

Notwithstanding the limitation of using a pair of isogenic cell lines, our statistical methods to analyze the growth patterns of thousands of clones generated in multiple mice injected with the same initial pair of genetically matched cells revealed 17 different clonal patterns of system-wide metastases. These findings suggest some support for historic evidence of intrinsically determined preferred metastatic environments originally derived from studies of clonally marked malignant mouse cells (28,29).

The hematogenous metastasis, an important mode of ovarian cancer metastasis (30), is an understudied area. Our barcode approach not only detected clones in blood and blood accessible organ sites, but our analysis demonstrated distinct clonal behaviors that exploit peritoneal and hematogenous routes for OVCAR5 cell metastasis. Although the coefficient was higher for blood in some organs like liver, spleen and contralateral (L) kidney, ascites also showed a significant predictive ability, suggesting that both routes can contribute to metastasis at these sites. Given the main trend of an increasing contribution of blood and decreasing contribution of ascites with distance from the primary tumor site, we believe the most parsimonious explanation is that both mechanisms can directly seed these sites, however, this by no means rules out a multi-step metastasis process. The additional extensive of diversity metastatic spread observed here is also likely to reflect the origin of the parental OVCAR5 cell line obtained from the ascites of an untreated ovarian cancer patient (22). A study using gene expression compositional assignment (GECA)

statistical method identified commonly used ovarian cell line OVCAR5 as non-ovarian, being gastrointestinal in origin (31). The lack of information regarding the primary gastrointestinal cancer in the ascites donor precluded the study from clinically confirming the findings of GECA method. The problem of primary gastrointestinal cancers metastasizing to the ovary and being misclassified as ovarian primary cancers and contributing to further disease heterogeneity is legendary (32), therefore, it is prudent to be cautious while attributing the metastatic patterns observed using the barcoding technology to the tissue origin of OVCAR5.

In the present study we investigated the influence of LRRC15 genotype on CIC fitness. LRRC15 was previously identified as frequently elevated in omental metastatic lesions in comparison to unmatched ovarian serous carcinoma (23). We have shown its involvement in regulating cell-cell and cell-matrix interactions and conferring an invasive mesenchymal-like phenotype increasing risk for ovarian metastasis (24). LRRC15 gene is poorly expressed in normal tissues and frequently upregulated in multiple cancer types including mesenchymal-derived tumors, providing target specificity and efficacy with low toxicity (33). We acknowledge that the shRNA approach used here may not have sufficiently silenced in all LRRC15- CIC but the scale of clonal analysis offered unprecedented resolution in growth and dissemination differences between LRRC15– versus LRRC15+ CICs.

It is noteworthy that our results also suggest that two distinct routes of metastasis reflect clonally distinct sets of cells that are released as ascites into the local peritoneal cavity, in contrast to those that penetrate the vasculature and then reach distal organs via the circulation. This exclusivity may be affected by microenvironmental factors but was also influenced by LRRC15 expression. It is therefore interesting to speculate that the blood and ascites content of malignant ovarian cells in patients might be similar predictors of probable sites of metastasis. A similarly interesting association of metastatic properties was seen when LRRC15 genotypes were compared. In this case the LRRC15+ cells produced metastatic clones in many organs whereas the LRRC15– cells metastases were largely restricted to the bowel, a pattern consistent with cells metastasizing as oligoclonal clusters reported previously (34).

The surgical implant sites were chosen here to mimic the competitive generation of primary ovarian tumors from paired cells differing only in the LRRC15 expression and hence enable a comparison of their subsequent ability to metastasize to different sites. This strategy was thus designed to potentially minimize aberrant growth responses obtained with transplants of single cell types injected in saturating numbers (5). The lack of an immune system in the immunodeficient mice used here to analyze the clonal patterns of system-wide metastatic activity of a model of human ovarian cancer remains a confounding variable in interpreting the significance of the results obtained. Nevertheless, they provide a baseline for adapting this approach to test the clonal growth, metastatic and drug response behavior of transplantable tumor cells from ovarian cancer patients.

In breast cancer research, barcoding approach has identified tumor cell clonal spread via circulation to lung (4,13).

A limitation of the DNA barcoding studies is the inability to isolate clones at single cell resolution and trace their molecular states which could be accomplished by integrating barcoding technology with scRNA sequencing. Recent studies have provided proof-of-concept use of 'expressed' DNA barcodes in detecting clones in bulk (35) or single cell RNA sequences (36). Further improvements in these cellular barcoding technologies will help understand tumor cell metastasis and ultimately control their fate. Moreover, the presently described competitive DNA barcoded transplant protocol and the computational framework to assist quantitative investigation of system-wide cancer metastasis at clonal resolution should facilitate further development of many such models in the future.

DATA AVAILABILITY

The codes to run the CIC Calculator that support the findings of this study are openly available in github (https://github.com/tangxj98/CIC.Calculator). All other data generated or analyzed during this study are included in this article and its supplementary files.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Cancer Online.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Riggi,N., Aguet,M. and Stamenkovic,I. (2018) Cancer metastasis: a reappraisal of its underlying mechanisms and their relevance to treatment. *Annu. Rev. Pathol.*, **13**, 117–140.
2. Fares,J., Fares,M.Y., Khachfe,H.H., Salhab,H.A. and Fares,Y. (2020) Molecular principles of metastasis: a hallmark of cancer revisited. *Signal Transduct Target Ther.*, **5**, 28.
3. Lan,X., Jorg,D.J., Cavalli,F.M.G., Richards,L.M., Nguyen,L.V., Vanner,R.J., Guilhamon,P., Lee,L., Kushida,M.M., Pellacani,D. *et al.* (2017) Fate mapping of human glioblastoma reveals an invariant stem cell hierarchy. *Nature*, **549**, 227–232.
4. Nguyen,L.V., Cox,C.L., Eirew,P., Knapp,D.J., Pellacani,D., Kannan,N., Carles,A., Moksa,M., Balani,S., Shah,S. *et al.* (2014) DNA barcoding reveals diverse growth kinetics of human breast tumour subclones in serially passaged xenografts. *Nat. Commun.*, **5**, 5871.
5. Nguyen,L.V., Makarem,M., Carles,A., Moksa,M., Kannan,N., Pandoh,P., Eirew,P., Osako,T., Kardel,M., Cheung,A.M. *et al.* (2014) Clonal analysis via barcoding reveals diverse growth and differentiation of transplanted mouse and human mammary stem cells. *Cell Stem Cell*, **14**, 253–263.
6. Nguyen,L.V., Pellacani,D., Lefort,S., Kannan,N., Osako,T., Makarem,M., Cox,C.L., Kennedy,W., Beer,P., Carles,A. *et al.* (2015) Barcoding reveals complex clonal dynamics of de novo transformed human mammary cells. *Nature*, **528**, 267–271.
7. Bhang,H.E., Ruddy,D.A., Krishnamurthy Radhakrishna,V., Caushi,J.X., Zhao,R., Hims,M.M., Singh,A.P., Kao,I., Rakiec,D., Shaw,P. *et al.* (2015) Studying clonal dynamics in response to cancer therapy using high-complexity barcoding. *Nat. Med.*, **21**, 440–448.
8. Hata,A.N., Niederst,M.J., Archibald,H.L., Gomez-Caraballo,M., Siddiqui,F.M., Mulvey,H.E., Maruvka,Y.E., Ji,F., Bhang,H.E., Krishnamurthy Radhakrishna,V. *et al.* (2016) Tumor cells can follow distinct evolutionary paths to become resistant to epidermal growth factor receptor inhibition. *Nat. Med.*, **22**, 262–269.
9. Kebschull,J.M. and Zador,A.M. (2018) Cellular barcoding: lineage tracing, screening and beyond. *Nat. Methods*, **15**, 871–879.
10. Gerrits,A., Dykstra,B., Kalmykowa,O.J., Klauke,K., Verovskaya,E., Broekhuis,M.J., de Haan,G. and Bystrykh,L.V. (2010) Cellular barcoding tool for clonal analysis in the hematopoietic system. *Blood*, **115**, 2610–2618.
11. Cheung,A.M., Nguyen,L.V., Carles,A., Beer,P., Miller,P.H., Knapp,D.J., Dhillon,K., Hirst,M. and Eaves,C.J. (2013) Analysis of the clonal growth and differentiation dynamics of primitive barcoded human cord blood cells in NSG mice. *Blood*, **122**, 3129–3137.
12. Naik,S.H., Schumacher,T.N. and Perie,L. (2014) Cellular barcoding: a technical appraisal. *Exp. Hematol.*, **42**, 598–608.
13. Merino,D., Weber,T.S., Serrano,A., Vaillant,F., Liu,K., Pal,B., Di Stefano,L., Schreuder,J., Lin,D., Chen,Y. *et al.* (2019) Barcoding reveals complex clonal behavior in patient-derived xenografts of metastatic triple negative breast cancer. *Nat. Commun.*, **10**, 766.
14. Seth,S., Li,C.Y., Ho,I.L., Corti,D., Loponte,S., Sapio,L., Del Poggetto,E., Yen,E.Y., Robinson,F.S., Peoples,M. *et al.* (2019) Pre-existing functional heterogeneity of tumorigenic compartment as the origin of chemoresistance in pancreatic tumors. *Cell Rep.*, **26**, 1518–1532.
15. Porter,S.N., Baker,L.C., Mittelman,D. and Porteus,M.H. (2014) Lentiviral and targeted cellular barcoding reveals ongoing clonal dynamics of cell lines in vitro and in vivo. *Genome Biol.*, **15**, R75.
16. Echeverria,G.V., Powell,E., Seth,S., Ge,Z., Carugo,A., Bristow,C., Peoples,M., Robinson,F., Qiu,H., Shao,J. *et al.* (2018) High-resolution clonal mapping of multi-organ metastasis in triple negative breast cancer. *Nat. Commun.*, **9**, 5079.
17. Wagenblast,E., Soto,M., Gutierrez-Angel,S., Hartl,C.A., Gable,A.L., Maceli,A.R., Erard,N., Williams,A.M., Kim,S.Y., Dickopf,S. *et al.* (2015) A model of breast cancer heterogeneity reveals vascular mimicry as a driver of metastasis. *Nature*, **520**, 358–362.
18. Roh,V., Abramowski,P., Hiou-Feige,A., Cornils,K., Rivals,J.P., Zougman,A., Aranyossy,T., Thielecke,L., Truan,Z., Mermod,M. *et al.* (2018) Cellular Barcoding Identifies Clonal Substitution as a Hallmark of Local Recurrence in a Surgical Model of Head and Neck Squamous Cell Carcinoma. *Cell Rep.*, **25**, 2208–2222.
19. Echeverria,G.V., Ge,Z., Seth,S., Zhang,X., Jeter-Jones,S., Zhou,X., Cai,S., Tu,Y., McCoy,A., Peoples,M. *et al.* (2019) Resistance to neoadjuvant chemotherapy in triple-negative breast cancer mediated by a reversible drug-tolerant state. *Sci. Transl. Med.*, **11**, eaav0936.
20. Aalam,S.M.M., Beer,P.A. and Kannan,N. (2019) Assays for functionally defined normal and malignant mammary stem cells. *Adv. Cancer. Res.*, **141**, 129–174.
21. Kokkaliaris,K.D., Lucas,D., Beerman,I., Kent,D.G. and Perie,L. (2016) Understanding hematopoiesis from a single-cell standpoint. *Exp. Hematol.*, **44**, 447–450.
22. Johnson,S.W., Laub,P.B., Beesley,J.S., Ozols,R.F. and Hamilton,T.C. (1997) Increased platinum-DNA damage tolerance is associated with cisplatin resistance and cross-resistance to various chemotherapeutic agents in unrelated human ovarian cancer cell lines. *Cancer Res.*, **57**, 850–856.
23. Mariani,A., Wang,C., Oberg,A.L., Riska,S.M., Torres,M., Kumka,J., Multinu,F., Sagar,G., Roy,D., Jung,D.B. *et al.* (2019) Genes associated with bowel metastases in ovarian cancer. *Gynecol. Oncol.*, **154**, 495–504.
24. Ray,U., Jung,D.B., Jin,L., Xiao,Y., Dasari,S., Sarkar Bhattacharya,S., Thirusangu,P., Staub,J.K., Roy,D., Roy,B. *et al.* (2022) Targeting LRRC15 inhibits metastatic dissemination of ovarian cancer. *Cancer Res.*, **82**,1038–1054.
25. Traag,V.A., Waltman,L. and van Eck,N.J. (2019) From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.*, **9**, 5233.
26. Wobbrock,J.O., Findlater,L., Gergle,D. and Higgins,J.J. (2011) In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Vancouver, BC, Canada, pp. 143–146.
27. Bermejo-Alvarez,P., Park,K.E. and Telugu,B.P. (2014) Utero-tubal embryo transfer and vasectomy in the mouse model. *J. Vis. Exp.*, **84**, e51214.
28. Nicolson,G.L. and Custead,S.E. (1982) Tumor metastasis is not due to adaptation of cells to a new organ environment. *Science*, **215**, 176–178.
29. Talmadge,J.E., Wolman,S.R. and Fidler,I.J. (1982) Evidence for the clonal origin of spontaneous metastases. *Science*, **217**, 361–363.
30. Pradeep,S., Kim,S.W., Wu,S.Y., Nishimura,M., Chaluvally-Raghavan,P., Miyake,T., Pecot,C.V., Kim,S.J., Choi,H.J., Bischoff,F.Z. *et al.* (2014) Hematogenous metastasis of ovarian cancer: rethinking mode of spread. *Cancer Cell*, **26**, 77–91.
31. Blayney,J.K., Davison,T., McCabe,N., Walker,S., Keating,K., Delaney,T., Greenan,C., Williams,A.R., McCluggage,W.G., Capes-Davis,A. *et al.* (2016) Prior knowledge transfer across transcriptional data sets and technologies using compositional statistics yields new mislabelled ovarian cell line. *Nucleic Acids Res.*, **44**, e137.
32. Gaudet,M.M., Sherman,M.E. and Thun,M.J. (2012) Learning from disease heterogeneity. *Lancet Oncol.*, **13**, 862–863.
33. Ray,U., Pathoulas,C.L., Thirusangu,P., Purcell,J.W., Kannan,N. and Shridhar,V. (2022) Exploiting LRRC15 as a novel therapeutic target in cancer. *Cancer Res.*, **82**, 1675–1681.
34. Aceto,N., Bardia,A., Miyamoto,D.T., Donaldson,M.C., Wittner,B.S., Spencer,J.A., Yu,M., Pely,A., Engstrom,A., Zhu,H. *et al.* (2014) Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. *Cell*, **158**, 1110–1122.
35. Jin,X., Demere,Z., Nair,K., Ali,A., Ferraro,G.B., Natoli,T., Deik,A., Petronio,L., Tang,A.A., Zhu,C. *et al.* (2020) A metastasis map of human cancer cell lines. *Nature*, **588**, 331–336.
36. Weinreb,C., Rodriguez-Fraticelli,A., Camargo,F.D. and Klein,A.M. (2020) Lineage tracing on transcriptional landscapes links state to fate during differentiation. *Science*, **367**, eaaw3381 .