scientific reports

Check for updates

OPEN

Epigenetic conservation infers that colorectal cancer progenitors retain the phenotypic plasticity of normal colon

Kelly Street[1], Yifan Zhang[1], Kimberly Siegmund[1] & Darryl Shibata[2]✉

Plasticity, or the ability to rapidly and reversibly change phenotypes, may help explain how a single progenitor cell eventually generates a tumor with many different cell phenotypes. Normal colon plasticity is characterized by a conserved and broadly permissive epigenome, where expression and phenotype are determined by the microenvironment instead of epigenetic remodeling. To determine whether this stem-like plasticity is retained during progression, gene expression was measured with spatial transcriptomics and compared with gene-level DNA methylation in two colorectal cancers (CRCs). Like normal colon, genes that were differentially expressed between regions, subclones, and phenotypes (superficial, invasive, and metastatic) tended to have lower DNA methylation variability. We propose a quantitative signal of plasticity that correlates gene epigenetic variability with gene expression variability. In this framework, negative correlation implies phenotypic plasticity, as more variably expressed genes tend to have less epigenetic variability. We verify the presence of this signal in multiple external single-cell RNA-Seq datasets, in both normal colon and CRC samples. Therefore, the plasticity of normal colon appears to be retained during progression. A CRC progenitor with a preconfigured plastic phenotype is poised for rapid growth because it expresses, as needed, transcripts required for progression with minimal epigenetic remodeling.

Keywords Colorectal cancer, Phenotypic plasticity, Spatial transcriptomics, DNA methylation, Wound healing

Cancers start from single progenitor cells that eventually form tumors with heterogeneous phenotypes and microenvironments. It is uncertain how cancer cells acquire so many different phenotypes during growth. Most driver mutations are clonal or present in all tumor cells, and most subclonal mutations that differ between tumor cells are passengers[1]. Subclonal mutations are unlikely causes of substantial phenotypic heterogeneity because cell phenotype differences can be discerned even in small (1 mm^2) tumor regions[2,3].

The epigenome is a likely source of phenotypic heterogeneity because mammalian expression is modulated by epigenetics[4]. Epigenetic remodeling can rapidly and irreversibly change gene expression and cell phenotypes, as exemplified by hematopoiesis, where multiple cell types differentiate from stem cells in a few weeks. For example, blood cell types can be predicted by differences in gene DNA methylation[5].

Another epigenetic mechanism is plasticity, or the ability of cells to rapidly and reversibly change phenotypes, a feature commonly observed in cancers[6–10]. Phenotypic plasticity is present in intestinal crypts, where expression is determined by the microenvironment and transcription factors, but the epigenome of functionally important genes is static and broadly permissive, and similar between differentiated cell types[11,12]. For example, it has been shown that DNA methylation does not substantially change during crypt stem cell differentiation[13,14]. Besides tissue maintenance, crypt stem cells are also poised, as needed, for tissue repair, with observations that cancers resemble wounds that do not heal[15,16]. Rather than being acquired or "unlocked" as a new capability during progression[9], the remarkable plasticity of crypt stem cells may be retained by colorectal cancer (CRC) progenitors.

Multiregional sampling reveals that epigenetic intratumoral heterogeneity is very common[15–21] although few epigenetic changes are seen in CRCs[22,23]. Here we characterize colon phenotypic plasticity based on the

[1]Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, United States. [2]Department of Pathology, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, United States. ✉email: dshibata@usc.edu

correlation between gene methylation variability and gene expression variability. By measuring DNA methylation along with expression measures from multiple transcriptomics platforms, we show that more variably expressed genes tend to have more conserved methylation profiles in normal colon crypts, adenomas, and CRCs. This relationship indicates that the plasticity of normal crypts is retained during tumor growth and epigenetic configurations permissive for variable gene expression are present at the start of growth. A CRC progenitor with a phenotype of plasticity is poised for rapid growth because its progeny can adapt, as needed, to the many microenvironments encountered during tumorigenesis (Fig. 1A).

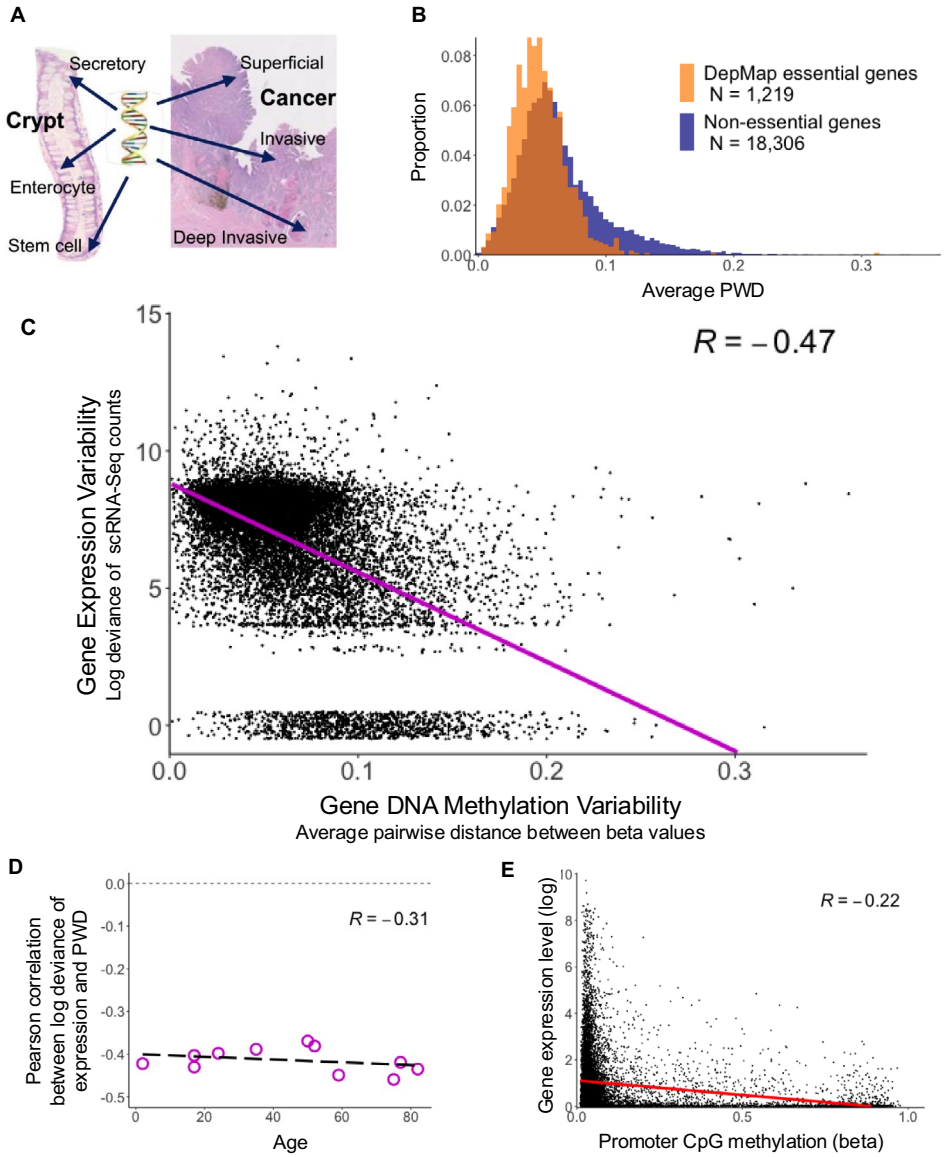

Fig. 1. Normal colon phenotypic plasticity. **(A)** Crypt stem and differentiated cell epigenomes are similar with phenotypic changes dependent on transcription factors induced by the microenvironment. A similar plasticity may be retained during CRC progression, where cell phenotypes depend on microenvironments and not on epigenetic remodeling or new driver mutations. A CRC progenitor cell starting with all its driver mutations and plasticity could rapidly grow into a large tumor with phenotypic diversity. **(B)** Histogram of average pairwise distances (PWD) between DNA methylation beta values annotated to the same gene, across 11 normal colon samples. DepMap essential genes (orange) appear to be preferentially conserved, with lower PWD values. **(C)** Scatter plot comparing gene DNA methylation variability (PWD) with scRNAseq gene expression variability, quantified by the deviance of a binomial model, adjusted for between-sample variation. Consistent with static but permissive epigenomes, more variably expressed colon genes have more conserved methylation profiles. **(D)** Scatter plot showing the Pearson correlations between DNA methylation variability and expression variably (as in C) as a function of age. Evidence of plasticity is present by 2 years of age and maintained with aging. **(E)** Scatter plot of average gene expression by promotor DNA methylation level. As we would expect, higher expression levels are generally seen with unmethylated promoters and conversely, highly methylated promoters correspond to lowly expressed genes.

Results

Plasticity in normal colon crypts

For genes that play an important role in any of a wide range of functions, colon epigenomes tend to be open, conserved, and broadly permissive, with expression variability depending on transcription factors induced by the microenvironment[11,12]. To characterize this plasticity, we compared gene expression variability with gene methylation variability. Methylation was measured by DNA methylation arrays and expression was measured by publicly-available single-cell RNA sequencing (scRNAseq) data[24]. Normal colon crypt DNA methylation was measured in 11 individuals (4 crypts each). Epigenetic variability is quantified by a pairwise distance (PWD) or the average absolute difference in methylation (beta values) between all CpG sites annotated with a gene or its promoter region. This distance is calculated between all 6 possible crypt pairs and averaged for each individual. If expression is determined by epigenetic remodeling, rather than plasticity, more variably expressed genes should show more methylation variability because "on" and "off" configurations are different. By contrast, with the plasticity of normal colon, expression variability occurs with an open and conserved, broadly permissive epigenome with little methylation variability[12].

To understand the functional importance of these conserved DNA methylation profiles, genes were subdivided into "essential" and "non-essential" categories, as determined by CRISPR gene essentiality data. Constitutive "essential" genes here are defined as the consensus DepMap-essential genes[25]. As previously observed[26], essential genes had more conserved DNA methylation patterns, as evidenced by their low PWD scores (Fig. 1B, mean = 0.0484). By contrast, non-essential genes had more variable methylation (mean PWD = 0.0645).

For comparison, normal colon scRNAseq data were obtained from two publicly available data sets: Wang, et al.[24], (GSE125970) and Lee, et al.[27], (GSE132465). Within the subset of cells annotated as normal colon and epithelial tissue, gene expression variability was quantified with deviance[28], a model-based measure of variation that accounts for the non-Gaussian distribution of scRNA-Seq read counts and adjusts for between-sample variation. Consistent with our model of plasticity, we find a significant negative correlation (r = −0.47) between gene expression deviance and DNA methylation variability (PWD), indicating that more variably expressed genes tend to have more conserved methylation profiles (Fig. 1C). This signal contradicts the expectations of a remodeling paradigm, in which higher epigenetic variability would correspond to higher expression variability. We also observe a similar pattern when epigenetic conservation is calculated based on PWDs between CpG sites within the TSS200 (promoter) regions of genes, rather than all CpG sites annotated with a gene (Fig S1A). Full-gene PWD is subsequently used to infer plasticity because more gene-annotated CpG sites are present on EPIC arrays.

This colon crypt plasticity signal may reflect epigenetic drift within a colon, including age-related changes in DNA methylation[29]. However, we observed that the plasticity signal is largely consistent across samples ranging in age from infancy (age 2 years) to 82 years (Fig. 1D). The preferential conservation of methylation patterns in essential genes and genes with high expression variability may reflect the importance of maintaining a robust, open or permissive configuration at these functionally important genes. Consistent with open or permissive epigenomes, most promoters of highly expressed genes were unmethylated and conserved, although a small number of expressed genes had high promoter methylation (Fig. 1E). The preferential conservation of more variably expressed genes is consistent with static but permissive crypt epigenomes where expression depends on transcription factors induced by the microenvironment, and not on epigenetic remodeling[12].

There was little correlation between essential gene expression variability and gene conservation, and as previously observed[30], most essential genes were conserved (Fig S1B). For genes classified as non-essential, there was less consistent expression, but we again observed a negative correlation between expression variability and DNA methylation variability (Fig S1C).

Plasticity in a localized CRC

Comparisons between gene methylation variability and expression variability were used to characterize the plasticity of two CRCs (referred to as "M" and "I"). Epigenetic conservation was measured by comparing DNA methylation sampled from opposite tumor sides (2 samples each). PWD values were calculated as before, by averaging absolute differences between beta values for all CpG sites that share a gene annotation, but with only one comparison per individual, rather than 6. Average conservation was higher between CRC sides than between normal crypts (Fig S2), likely reflecting their recent clonal expansions. Expression variability was measured with spatial transcriptomics (GeoMx Digital Spatial Profiler) of multiple small ~ 200–400 keratin positive cell spots. The CRCs were also analyzed with saturation microdissection and targeted resequencing[31] to map subclone boundaries. Hence, each spot is defined by a transcriptional profile, location, phenotype, and ancestry. A total of 49 spots were analyzed for tumor M and 48 spots for tumor I.

CRC "M" was locally invasive (Stage 2) with superficial and invasive regions on two sections (Fig. 2A). An adenoma subclone arose before the cancer subclones (Fig. 2B). Cancer subclones were oriented vertically and have both superficial and invasive regions. As in normal colon, essential genes had more conserved DNA methylation profiles (mean PWD = 0.0220, vs. 0.0394 for non-essential genes) (Fig. 2C). Expression patterns from 49 spots were similar between phenotypes, with greater correlation between superficial and invasive phenotypes than with adenoma spots (Fig S3A). Expression was slightly more correlated between physically adjacent spots compared to more distantly spaced spots (Fig S4A).

The variability of gene expression was quantified in multiple ways to examine the relationship with gene methylation variability. In order to account for spatial patterns, we quantified spatial gene expression variability using Moran's I statistic and observed a negative correlation (r = −0.11) with DNA methylation variability (Fig. 2D). We also tested for differential expression between annotated regions and found genes that were differentially expressed between any pair of regions had more conserved DNA methylation profiles (mean PWD = 0.0272, vs. 0.0404 for non-differentially expressed genes) (Fig. 2E). When quantifying expression

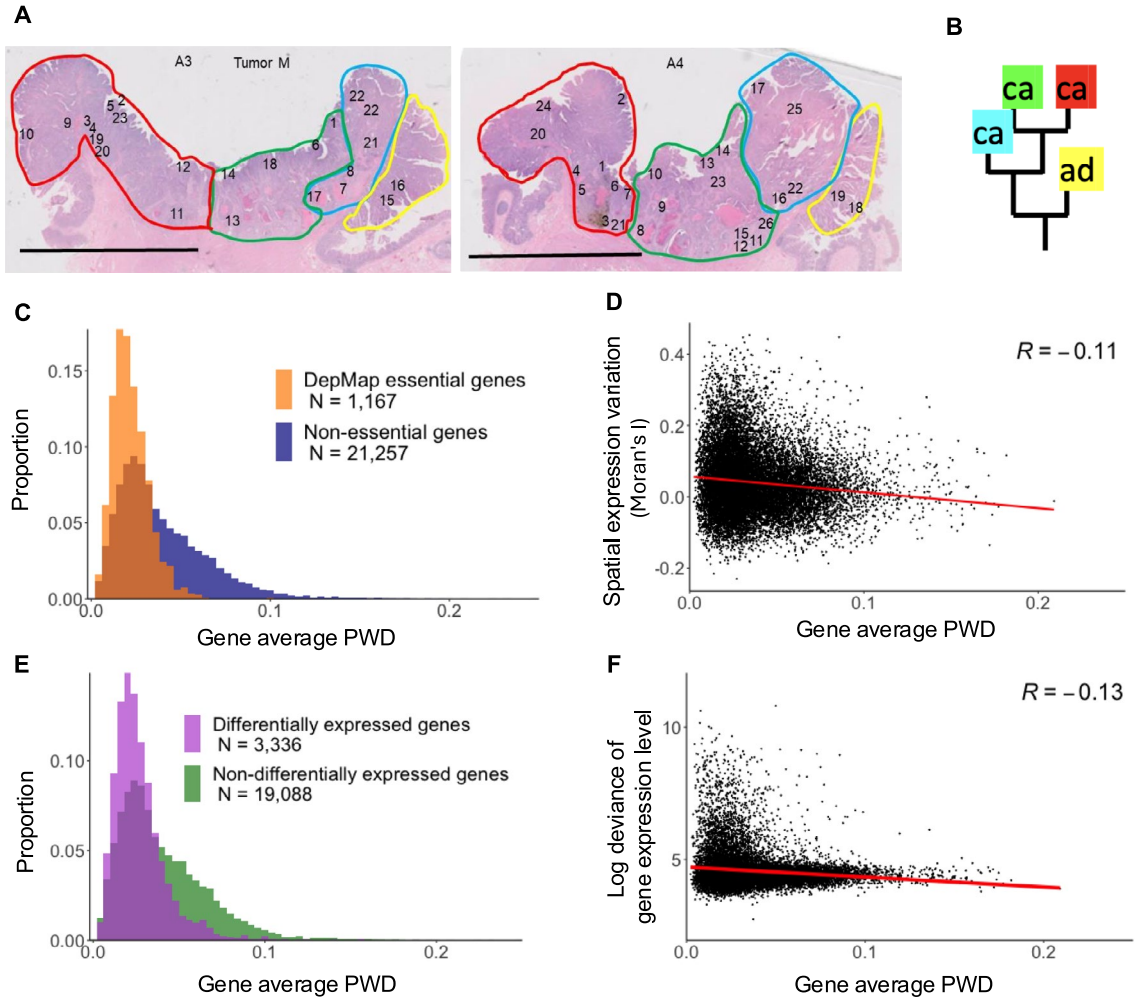

Fig. 2. CRC phenotypic plasticity. **(A)** Subclones defined by subclonal mutations are mapped onto two microscopic sections from a Stage 2 CRC with local invasion. **(B)** Ancestral tree indicates early branching of the adenoma subclone followed by the three subclones with superficial and invasive phenotypes. **(C)** As in normal colon, DepMap essential genes (orange) have lower overall DNA methylation PWD, indicating that they are preferentially conserved. **(D)** Scatter plot comparing gene DNA methylation variability (PWD) with spatial expression variability, quantified by Moran's I statistic. The negative relationship indicates the continued presence of phenotypic plasticity. **(E)** As in (C), but genes are split by differential expression testing. Genes significantly differentially expressed ($p < 0.05$ by DESeq2) between any pair of phenotypes were more conserved. **(F)** Scatter plot comparing gene DNA methylation variability (PWD) with expression variability, quantified by deviance, as in Fig. 1C, but expression is taken from spatial transcriptomics rather than public scRNAseq data. The "L"-shaped pattern indicates that genes tend to have either highly variable expression or highly variable DNA methylation, but not both.

variability with deviance, we again see a negative correlation with methylation variability ($r = -0.13$) (Fig. 2F), similar to the pattern observed with the single-cell expression data. All of these findings are consistent with plasticity, as more variable genes tend to be more epigenetically conserved.

Plasticity in a metastatic CRC

CRC "I" was metastatic (Stage 3). There were three major subclones (Fig. 3A) with one superficial and invasive (green), one superficial, invasive, deeply invasive, and metastatic (blue), and one deeply invasive and metastatic (red). The red metastatic subclone arose early and before the green subclone with only superficial and invasive phenotypes (Fig. 3B). Like normal colon and the localized CRC ("M"), genes categorized as essential had more conserved DNA methylation profiles (mean PWD = 0.0187, vs. 0.0366 for non-essential genes) (Fig. 3C). Expression was again similar between phenotypes and subclones with only slight differences between the primary and metastatic spots (Fig S3B). There was again only a slight relationship between the correlation of gene expression measures between spots and their physical distance (Fig S4B).

While more genes (10,048) were differentially expressed between phenotypic regions in tumor "I" than tumor "M" (3,336), these genes did not have a noticeably distinct distribution of PWD values (Fig. 3E), though they did have a slightly lower mean PWD (0.0332 vs. 0.0376 for non-differentially expressed genes). However,

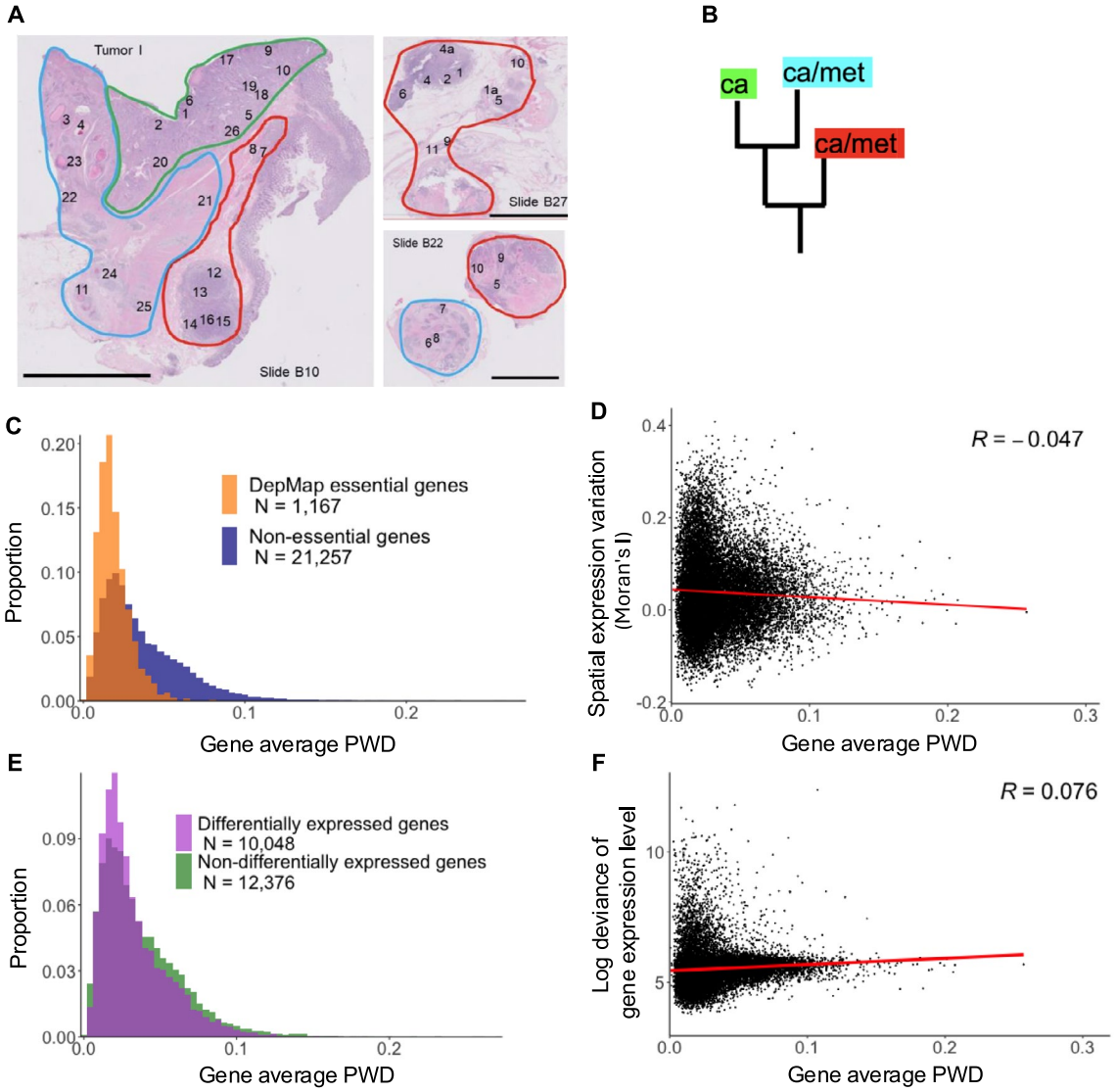

Fig. 3. Metastatic CRC phenotypic plasticity. **(A)** Subclones defined by subclonal mutations are mapped onto primary and nodal metastatic microscopic sections. **(B)** Ancestral tree indicates early metastatic branching (red subclone). **(C)** As in normal colon, DepMap essential genes (orange) have lower overall DNA methylation PWD, indicating that they are preferentially conserved. **(D)** Scatter plot comparing gene DNA methylation variability (PWD) with spatial expression variability, quantified by Moran's I statistic. The negative relationship indicates the continued presence of phenotypic plasticity. **(E)** As in (C), but genes are split by differential expression (DE) testing. Genes that were significantly DE ($p < 0.05$ by DESeq2) between any pair of phenotypes are shown in purple, non-DE genes are shown in green. **(F)** Scatter plot comparing gene DNA methylation variability (PWD) with expression variability, quantified by deviance, as in Fig. 1C, but expression is taken from spatial transcriptomics rather than public scRNAseq data. Although the correlation is positive due to a large number of conserved and lowly expressed genes, the overall "L"-shaped pattern indicates that genes tend to have either highly variable expression or highly variable DNA methylation, but not both.

we observed a similar patterns in the relationship between gene expression variability and DNA methylation variability. When quantifying expression variability with Moran's I statistic to account for spatial trends, there was a weak negative correlation ($r = -0.047$) (Fig. 3D). When using deviance, we again note that highly variable genes tend to be more conserved and genes with high methylation variability tend to have low expression variability (Fig. 3F). While we did observe a weak, positive correlation value ($r = 0.076$), this can be attributed to a preponderance of seemingly unimportant genes with low methylation variability and very low expression variability. Both CRCs appear to retain colon crypt phenotypic plasticity because more variably expressed genes have more conserved DNA methylation profiles.

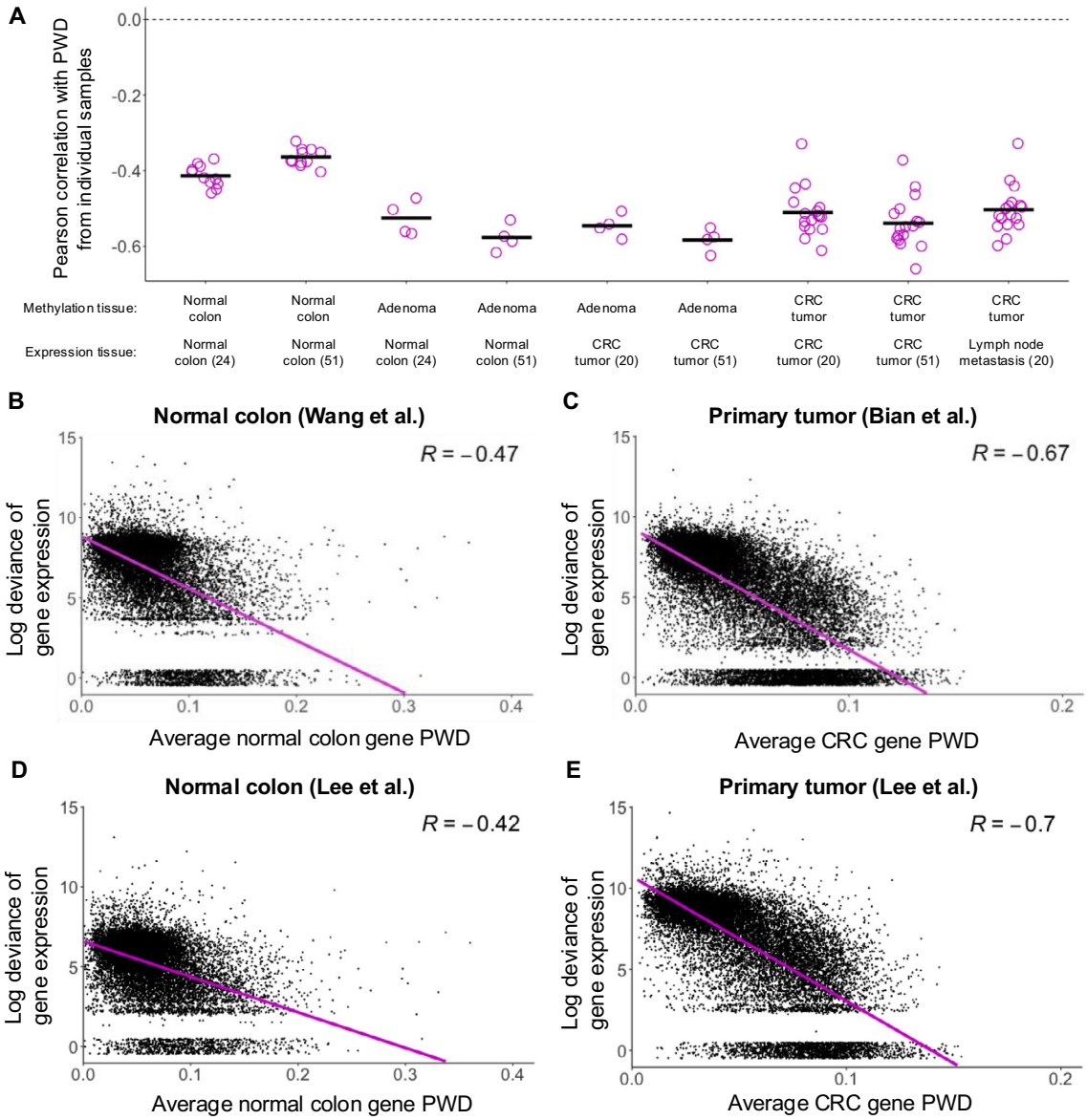

Fig. 4. Normal colon and CRC phenotypic plasticity. **(A)** Scatter plot showing the Pearson correlations obtained by comparing gene PWD values (DNA methylation variability) from individual samples to deviance values (expression variability) obtained from scRNAseq samples of similar cell types[20,24,27]. We find consistent negative correlation between methylation variability and expression variability across a range of colon and CRC cell types, consistent with a plasticity model of expression. **(B)** As in Fig. 1C, there is a negative correlation between the average PWD across all 11 normal colon samples and the expression deviance from normal colon cells in a public scRNAseq dataset[24]. This correlation (−0.47) is lower than the average in Fig. 4 A because PWDs were averaged before the correlation was calculated, rather than calculating 11 individual correlation values and then averaging them. **(C)** There is a negative correlation (−0.67) between the average gene methylation PWD across all 17 CRC samples and the expression deviance calculated from "Primary Tumor" samples in[20]. **(D)** Despite limited cell counts, there is a negative correlation (−0.42) between the average gene methylation PWD across all 11 normal colon samples and the expression deviance calculated from "Normal" samples in[27]. **(E)** There is a negative correlation (−0.70) between the average DNA methylation PWD across all 17 CRC samples and the expression deviance calculated from "Primary Tumor" samples in[27].

CRC plasticity with scRNAseq data

The lack of spot expression variability within the two CRCs may reflect the cell purity achievable with spatial transcriptomics, and is consistent with recent studies that indicate CRC expression can be patient-specific even with scRNAseq data[32]. Spatial transcriptomics may also dampen variation when averaging the expression of 200–400 cells per spot. Therefore, scRNAseq data[20,24,27] from matching cell types was compared with the gene conservation measures (Fig. 4) from normal colon and 19 addition CRC tumors (4 adenomas and 15 new CRCs, Table S1).

We find that the same signature of plasticity described above is widely present in normal and neoplastic colon. Across all normal colon samples in our cohort, there was a negative correlation between DNA methylation variability (PWD) and the expression variability (deviance) of normal colon cells from scRNA-Seq (Fig. 4A "Normal colon" values). Taking the average PWD across all normal colon methylation samples showed similar negative correlations (Fig. 4B,D). The magnitude of these correlations tended to increase when using average PWD values across individuals, perhaps indicating that functionally important genes are epigenetically conserved across individuals, but others may change their methylation profiles in a stochastic manner.

Comparing the average PWD of our 4 adenoma samples to the expression variability from normal colon cells showed a similar pattern (Fig. 4A, Fig S5A). To compare with CRC DNA methylation samples, we measured the gene expression deviance in CRC cells from Bian, et al.[20] (GSE97693) and Lee, et al.[27] (GSE132465). Again, we observed a consistent negative correlation, indicative of plasticity, across all samples (Fig. 4A, "CRC tumor" samples) and the average PWD distribution (Fig. 4C,E). Similar negative correlations were observed when comparing these gene expression deviance values to PWD values taken from our adenoma samples (Fig. 4A, Fig S5B) and with scRNAseq data from lymph node metastases (Fig. 4A, Fig S5C). These results were robust to the choice of correlation measure (Fig S6 replicates Fig. 4A using Spearman correlation). We note that our deviance calculation fits separate models per individual, so that the scores represent a summary of within-patient variability, not between-patient variability, although measures of deviance are broadly similar across patients (Fig S7). In all comparisons, we find that more variably expressed genes tended to have less DNA methylation variability, indicating the preservation of normal crypt-like phenotypic plasticity throughout CRC progression.

Discussion

Multiregional sampling can infer the genome of a progenitor cell. Clonal mutations detected throughout the tumor reside in the progenitor cell whereas subclonal mutations are acquired during growth. Typically, a progenitor starts with all its driver mutations because most subclonal mutations are passengers[1]. Similarly, the methylation profile of a progenitor cell can be inferred by comparing DNA methylation between regions. Conserved configurations were likely present in the progenitor cell, and some characteristics of the progenitor cell phenotype can be inferred because mammalian gene expression is modulated epigenetically[4].

This multiregional strategy could recover the phenotype of normal stem cells by comparing DNA methylation profiles from multiple colon crypts. Stem cells are morphologically undifferentiated but have a functional phenotype of plasticity because their progeny differentiate into multiple cell types. Crypt stem cells and their progeny have static or conserved epigenomes that are broadly permissive across a range of functionally important genes, allowing expression to be modulated by the microenvironment instead of epigenetic remodeling[11,12]. Comparisons between colon crypts were consistent with this stem cell phenotype because more variably expressed genes have more conserved DNA methylation profiles. Expression plasticity enables stem cells to rapidly differentiate into multiple different distinct crypt cell types, and conversely, differentiated cells can, as needed, reverse back to an undifferentiated stem cell state[29]. This preferential conservation likely reflects purifying or negative selection, where conservation provides empirical evidence of essential function in genomic regions[33].

Recent multiregional studies have found evidence of widespread plasticity in CRCs by noting poor correlations between epigenomes and expression patterns[22,23], including DNA methylation[34]. CRC plasticity resembles normal crypt plasticity because more variably expressed CRC genes also have more conserved methylation profiles. Retention of crypt stem cell-like plasticity with permissive and conserved epigenomes would allow CRC cells to rapidly change phenotypes and adapt to different microenvironments with minimal epigenetic remodeling. Stem cell repertoire is extensive, and as needed, normal stem cells produce cells with morphologic features of neoplasia, notably during regeneration, as tumors resemble "wounds that do not heal"[35,36].

This crypt stem-like plasticity can be maintained by organoids with specialized media[37] but is lost in most settings. Limited plasticity is observed in large tumor populations and typically only a minority of cells can change their phenotypes[38]. This plasticity may reflect selection of random genetic and epigenetic differences that are common in large human tumors[6,7]. Tumors have distinct and diverse epigenomes that modulate transcription at the single-cell level[39,40]. Glial tumors also retain some stem-like plasticity because some individual lineages differentiate and dedifferentiate[40]. A dependence on drift and diversity is consistent with plasticity being an acquired cancer hallmark[8,9]. However, this type of population plasticity is limited at the start of growth when tumor cells are few and diversity is inherently low.

Instead, retention of crypt plasticity, where epigenomes are preconfigured to respond to different microenvironments, would facilitate rapid early phenotypic diversification during "Big Bang" tumorigenesis where growth occurs as single expansions[41]. CRC mutations are consistent with single expansion with public driver mutations[1] and subclonal passengers with neutral mutation frequency distributions[42]. Instead of subclones with distinct phenotypes, plasticity allows subclones to have multiple morphological phenotypes with both superficial and invasive locations[31], and gradual phenotypic gradients in three-dimensional CRC reconstructions[43]. Progenitor cell plasticity may also facilitate the early metastasis commonly inferred from lineage studies, with estimates that metastatic CRC cells migrate when the primary tumor is as small as 10,000 cells[44].

A weakness of this study is that retention of CRC progenitor plasticity, like in the normal colon, is inferred from the lack of epigenetic changes. Normal crypt plasticity is tolerant to minor epigenetic differences, but it is uncertain to what extent small chromatin changes modulate expression in CRCs. We also note that the negative correlations indicative of plasticity are strongest in the cases when we aggregate (average) PWDs from multiple samples, and weakest when only use a single sample, as with tumor M and I. This may be due to random variations in epigenetic drift between patients, such that aggregating data across patients provides a clearer picture of which genes are preferentially conserved.

A CRC progenitor that retains crypt stem-like plasticity and has all its driver mutations is poised for rapid growth. Early metastasis and the difficulties of defining metastatic specific driver mutations[45] may indicate that the degree of plasticity at the start of growth and subsequent loss during growth may help determine how fast and far a CRC may spread.

Methods

Samples

Samples are 11 normal colons (4 crypts each), 4 adenomas, and 17 CRCs (2 sides each) obtained at the Norris Cancer Center as excess tissue during routine clinical care (Table S1). The study was approved by the Institutional Review Board (IRB) at the University of Southern California Keck School of Medicine (#HS-18–00,043), and all methods were performed in accordance with the relevant guidelines and regulations. No sample size calculations were performed, as inclusion was based on availability of tissues. Informed consent was waived by the Institutional Review Board (IRB) at the University of Southern California Keck School of Medicine due to the use of de-identified clinical samples. All samples were anonymized to maintain patient confidentiality. Single colon crypts and tumor glands were isolated from fresh bulk regions (0.5 cm^3) dissected from normal colon, or from opposite tumor sides with an EDTA washout method[46].

DNA methylation between tumor sides (hundreds of glands) was measured with Illumina EPIC arrays, with single crypts measured with the Restore protocol[47]. IDAT files were processed using the noob normalization function in the minfi R package[48]. Only autosomal genes with at least 4 annotated CpG sites were analyzed. Gene-level conservation was measured as a PWD (average absolute difference) of the CpG site beta values for all its annotated CpG sites. For normal colon samples, all six possible pairwise PWDs were averaged. Gene promoter methylation (TSS200) used all available annotated CpG sites. Pathway enrichment analysis was performed on the Reactome website[49]. The list of common DepMap essential genes (Dataset S2) was obtained from DepMap data[25] of single CRISPR-Cas9 gene disruptions, downloaded from the DepMap website (https://depmap.org/portal/ Public 20Q4, Archilles_gene_effect.csv).

Spatial transcriptomics

Expression of keratin positive cell spots (~ 200–400 cells per spot, 49 spots for tumor M and 48 spots for tumor I) were measured from formalin fixed tissue microscope slides from two CRCs using the GeoMx Digital Spatial Profiler (NanoString) and the GeoMx Human Whole Transcriptome Atlas (about 18,000 genes). Expression levels (log2) were from 12,829 (tumor M) and 12,407 (tumor I) genes after filtering and quantile-3 normalization with GeoMx Software v2.5 without background subtraction. Processed data are provided in Dataset S3.

Differential expression testing between phenotypic regions was performed with DESeq2[50]. Moran's I statistic[28] was used as a metric of spatial expression variability, with weights determined by the inverse distances between spots. Positive values of Moran's I indicate the expression of a gene varies according to a spatial pattern, with nearby spots having more similar levels of expression. Values of Moran's I close to zero indicate random distribution of expression levels, and negative values indicate local anti-correlation, with high- and low-expression spots finely interspersed. Expression variability was also measured without spatial consideration by using the deviance of a binomial model, as implemented in the scry package (https://www.bioconductor.org/packages/release/bioc/html/scry.html)[28].

scRNAseq analysis

scRNAseq data were obtained from three prior studies. Data from normal colon cells were obtained from Wang, et al.[24], (GSE125970) and all cells from colon samples ("Colon 1" and "Colon 2") were included in our analysis. scRNAseq data from CRC cells originating from both primary tumor and lymph node metastasis samples were obtained from Bian, et al.[20], (GSE97693). Data from both normal colon and CRC primary tumor samples were obtained from Lee, et al.[27], (GSE132465). In all three datasets, gene expression variability was measured using the deviance of a binomial model, as implemented in the scry package (https://www.bioconductor.org/packages/release/bioc/html/scry.html)[28]. In order to focus on between-cell type variability rather than between-sample variability, each sample was treated as a separate batch, meaning unique model parameters were fit for each individual.

Spatial CRC subclone analysis

Subclones were mapped onto microsections from the same paraffin block as the spatial transcriptomic studies using saturation microdissection and targeted resequencing, as previously described for tumor M[31]. Data for tumor I are new and provided in Dataset S4.

Data availability

The colon crypt methylation data are available from the European Genome– Phenome Archive (EGA) (accession EGAS00001005514). The colorectal cancer data are available at the NCBI GEO Database, under accession code GSE166212.

Code availability

Analysis was performed in the R programming language and all code is available at: https://github.com/kstreet13/plasticity.

Received: 20 June 2024; Accepted: 15 October 2025
Published online: 20 November 2025

References

1. Reiter, J. G. et al. An analysis of genetic heterogeneity in untreated cancers. *Nat. Rev. Cancer* **19**(11), 639–650 (2019).
2. Schürch, C. M. et al. Coordinated cellular neighborhoods orchestrate antitumoral immunity at the colorectal cancer invasive front. *Cell* **182**, 1341–1359 (2020).
3. Jackson, H. W. et al. The single-cell pathology landscape of breast cancer. *Nature* **578**, 615–620 (2020).
4. Allis, C. D. & Jenuwein, T. The molecular hallmarks of epigenetic control. *Nat. Rev. Genet.* **17**, 487–500 (2016).
5. Farlik, M. et al. DNA methylation dynamics of human hematopoietic stem cell differentiation. *Cell Stem Cell* **19**(6), 808–822 (2016).
6. Easwaran, H., Tsai, H. C. & Baylin, S. B. Cancer epigenetics: Tumor heterogeneity, plasticity of stem- like states, and drug resistance. *Mol. Cell* **54**, 716–727 (2014).
7. Flavahan, W. A., Gaskell, E. & Bernstein, B. E. Epigenetic plasticity and the hallmarks of cancer. *Science* **357**, eaal2380 (2017).
8. Mathur, R. & Costello, J. F. Epigenomic contributions to tumor cell heterogeneity and plasticity. *Nat. Genet.* **53**, 1403–1404 (2021).
9. Hanahan, D. Hallmarks of cancer: New dimensions. *Cancer Discov.* **12**(1), 31–46 (2022).
10. Burkhardt, D. B., San Juan, B. P., Lock, J. G., Krishnaswamy, S. & Chaffer, C. L. Mapping phenotypic plasticity upon the cancer cell state landscape using manifold learning. *Cancer Discov.* **12**(8), 1847–1859 (2022).
11. Kim, T. H. et al. Broadly permissive intestinal chromatin underlies lateral inhibition and cell plasticity. *Nature* **506**, 511–515 (2014).
12. Saxena, M. & Shivdasani, R. A. Epigenetic signatures and plasticity of intestinal and other stem cells. *Annu. Rev. Physiol.* **83**, 405–427 (2021).
13. Kaaij, L. T. et al. DNA methylation dynamics during intestinal stem cell differentiation reveals enhancers driving gene expression in the villus. *Genome Biol.* **14**, R50 (2013).
14. Kazakevych, J., Sayols, S., Messner, B., Krienke, C. & Soshnikova, N. Dynamic changes in chromatin states during specification and differentiation of adult intestinal stem cells. *Nucleic Acids Res.* **45**, 5770–5784 (2017).
15. Martínez-Cardús, A. et al. Epigenetic homogeneity within colorectal tumors predicts shorter relapse-free and overall survival times for patients with locoregional cancer. *Gastroenterology* **151**, 961–972 (2016).
16. Hua, X. et al. Genetic and epigenetic intratumor heterogeneity impacts prognosis of lung adenocarcinoma. *Nat. Commun.* **11**, 2459 (2020).
17. Brocks D, Assenov Y, Minner S, Bogatyrova O, Simon R, Koop C, Oakes C, Zucknick M, Lipka DB, Weischenfeldt J, Feuerbach L, Cowper-Sal Lari R, Lupien M, Brors B, Korbel J, Schlomm T, Tanay A, Sauter G, Gerhäuser C, Plass C; ICGC Early Onset Prostate Cancer Project. Intratumor DNA methylation heterogeneity reflects clonal evolution in aggressive prostate cancer. *Cell Rep.* **8**, 798–806 (2014).
18. Assenov, Y., Brocks, D. & Gerhäuser, C. Intratumor heterogeneity in epigenetic patterns. *Semin. Cancer Biol.* **51**, 12–21 (2018).
19. Ryser, M. D., Yu, M., Grady, W., Siegmund, K. & Shibata, D. Epigenetic heterogeneity in human colorectal tumors reveals preferential conservation and evidence of immune surveillance. *Sci. Rep.* **8**, 17292 (2018).
20. Bian, S. et al. Single-cell multiomics sequencing and analyses of human colorectal cancer. *Science* **362**, 1060–1063 (2018).
21. Mazor, T., Pankov, A., Song, J. S. & Costello, J. F. Intratumoral heterogeneity of the epigenome. *Cancer Cell* **29**, 440–451 (2016).
22. Heide, T. et al. The co-evolution of the genome and epigenome in colorectal cancer. *Nature* **611**(7937), 733–743 (2022).
23. Househam, J. et al. Phenotypic plasticity and genetic control in colorectal cancer evolution. *Nature* **611**(7937), 744–753 (2022).
24. Wang, Y. et al. Single-cell transcriptome analysis reveals differential nutrient absorption functions in human intestine. *J. Exp. Med.* **217**(2), e20191130 (2020).
25. Behan, F. M. et al. Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. *Nature* **568**, 511–516 (2019).
26. Rud, D., Marjoram, P., Siegmund, K. & Shibata, D. Functional human genes typically exhibit epigenetic conservation. *PLoS ONE* **16**(9), e0253250 (2021).
27. Lee, H. O. et al. Lineage-dependent gene expression programs influence the immune landscape of colorectal cancer. *Nat. Genet.* **52**(6), 594–603 (2020).
28. Moran, P. A. P. Notes on continuous stochastic phenomena. *Biometrika* **37**(1), 17–23 (1950).
29. Li, H. et al. Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors. *Nat. Genet.* **49**, 708–718 (2017).
30. Shivdasani, R. A., Clevers, H. & de Sauvage, F. J. Tissue regeneration: Reserve or reverse?. *Science* **371**, 784–786 (2021).
31. Ryser, M. D. et al. Minimal barriers to invasion during human colorectal tumor growth. *Nat. Commun.* **11**, 1280 (2020).
32. Joanito, I. et al. Single-cell and bulk transcriptome sequencing identifies two epithelial tumor cell states and refines the consensus molecular classification of colorectal cancer. *Nat. Genet.* **54**, 963–975 (2022).
33. Haerty, W. & Ponting, C. P. No gene in the genome makes sense except in the light of evolution. *Annu. Rev. Genomics Hum. Genet.* **15**, 71–92 (2014).
34. Becker, W. R. et al. Single-cell analyses define a continuum of cell state and composition changes in the malignant transformation of polyps to colorectal cancer. *Nat. Genet.* **54**(7), 985–995 (2022).
35. MacCarthy-Morrogh, L. & Martin, P. The hallmarks of cancer are also the hallmarks of wound healing. *Sci. Signal.* **13**, eaay8690 (2020).
36. Ge, Y. & Fuchs, E. Stretching the limits: From homeostasis to stem cell plasticity in wound healing and cancer. *Nat. Rev. Genet.* **19**, 311–325 (2018).
37. Sato, T. et al. Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. *Nature* **459**(7244), 262–265 (2009).
38. Sacchetti, A. Phenotypic plasticity underlies local invasion and distant metastasis in colon cancer. *Elife* **10**, e61461 (2021).
39. Johnson, K. C. et al. Single-cell multimodal glioma analyses identify epigenetic regulators of cellular plasticity and environmental stress response. *Nat. Genet.* **53**, 1456–1468 (2021).
40. Chaligne, R. et al. Epigenetic encoding, heritability and plasticity of glioma transcriptional cell states. *Nat. Genet.* **53**, 1469–1479 (2021).
41. Sottoriva, A. et al. A Big Bang model of human colorectal tumor growth. *Nat. Genet.* **47**, 209–216 (2015).
42. Williams, M. J. et al. Quantification of subclonal selection in cancer from bulk sequencing data. *Nat. Genet.* **50**(6), 895–903 (2018).
43. Lin, J. R. et al. Multiplexed 3D atlas of state transitions and immune interaction in colorectal cancer. *Cell* **186**(2), 363-381.e19 (2023).
44. Hu, Z. et al. Quantitative evidence for early metastatic seeding in colorectal cancer. *Nat. Genet.* **51**(7), 1113–1122 (2019).
45. Bernards Weinberg RA. A progression puzzle Nature. 418:823 (2002).
46. Nakamura, S., Goto, J., Kitayama, M. & Kino, I. Application of the crypt-isolation technique to flow- cytometric analysis of DNA content in colorectal neoplasms. *Gastroenterology* **106**, 100–107 (1994).
47. Illumina. Infinium HD FFPE Restore Protocol. (Illumina). 541 (2011).
48. Aryee, M. J. et al. Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363–1369 (2014).
49. Jassal, B. et al. The reactome pathway knowledgebase. *Nucleic Acids Res.* **48**, D498–D503 (2020).
50. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
51. Townes, F. W., Hicks, S. C., Aryee, M. J. & Irizarry, R. A. Feature selection and dimension reduction for single-cell RNA-Seq based on a multinomial model. *Genome Biol.* **20**, 295 (2019).

Acknowledgements
We are grateful for the investigators that provided well-annotated data in GEO. We thank Lukas Weber for suggesting Moran's I statistic as a metric for spatial expression variability.

Author contributions
DS collected samples and datasets. K Siegmund analyzed DNA methylation data. K Street analyzed scRNA-seq data. K Street and YZ prepared figures.

Funding
National Institutes of Health, P01CA196569, P01CA196569, P01CA196569.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-24703-3.

Correspondence and requests for materials should be addressed to D.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.