life

MDPI

Review

Connecting Cancer Susceptibility with Somatic Evolution in Gastrointestinal and Hepatobiliary Cancers

Mikhail B. Potievskiy [1,*], Ivan V. Korobov [1], Ilya D. Klabukov [1], Petr A. Shatalov [1], Alexei A. Traspov [1], Kamilla R. Khalitova [2], Vladislava D. Garshina [2], Valerii V. Strarinskii [1], Petr V. Shegai [1] and Andrei D. Kaprin [1,3]

[1] FSBI "National Medical Research Radiological Center", Ministry of Health of the Russian Federation, 4 Koroleva Street, Obninsk 249036, Russia; i.korobov99@mail.ru (I.V.K.); ilya.klabukov@gmail.com (I.D.K.); shatalov.peter@nmicr.ru (P.A.S.); traspovalex@gmail.com (A.A.T.); valstar40@mail.ru (V.V.S.); dr.shegai@mail.ru (P.V.S.); kaprin@mail.ru (A.D.K.)

[2] Institute of clinical medicine, Sechenov University, 8 Trubetskaya Street, Moscow 119048, Russia; khalitova_k_r@student.sechenov.ru (K.R.K.); garshina_v_d@student.sechenov.ru (V.D.G.)

[3] Institute of medicine, Peoples' Friendship University of Russia—RUDN University, 6 Miklukho-Maklaya Street, Moscow 117198, Russia

* Correspondence: mikhailpotievsky@yandex.ru; Tel.: +7-916-035-82-55

Abstract

Inherited cancer susceptibility and tumor microevolution concern different levels of cancer biology. This critical narrative review examines their connections in gastric, colorectal, hepatocellular, biliary tract and pancreatic cancers. We organize the evidence around four processes: mutation acquisition, early clonal establishment, microenvironmental selection and the expansion of resistant clones during treatment. For each process, we distinguish population associations from functional experiments and measurements of somatic lineages within patients. The clearest mechanistic evidence concerns mismatch repair loss in hereditary predisposition, microbial mutagenesis, mutation-specific expansion in normal tissue and repair-associated resistance in pancreatic cancer. Broad ancestry differences and blood methylation associations usually leave the intervening clonal process unresolved. A revised evidence matrix identifies what each study directly measures and what remains to be tested. We describe the assumptions needed to estimate clone abundance and interpret serial ctDNA, then consider applications in surveillance, pretreatment profiling, monitoring and resistance investigation. Methylation and fragmentation measurements may provide complementary information, but their value for guiding treatment requires validation. Connecting susceptibility to tumor evolution will require paired host and tumor measurements that test a specific somatic transition.

Keywords: cancer susceptibility; population differences; genetic ancestry; tumor microevolution; clonal selection; treatment resistance; gastrointestinal cancers

Check for updates

Academic Editor: Jian-Hua Mao

Received: 24 October 2026
Revised: 12 November 2026
Accepted: 13 November 2026
Published: 15 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Inherited risk studies identify factors associated with developing cancer. Tumor phylogenies reconstruct relationships among somatic cell lineages. An association with cancer occurrence does not establish how quickly a tumor diversifies, when it disseminates or which clones survive treatment. Even a difference in driver-mutation frequency between patient groups can reflect several processes: mutation acquisition, subtype composition, selection or differences in sampling. This review examines the evidence needed to distinguish these explanations.

We focus on gastric cancer, colorectal cancer (CRC), hepatocellular carcinoma (HCC), biliary tract cancers (BTCs) and pancreatic ductal adenocarcinoma (PDAC). These diseases offer contrasting examples of hereditary predisposition, infection and chronic inflammation, anatomical heterogeneity and repair-associated treatment response. Each also has population datasets and studies of tumor lineages, although these measurements are seldom available in the same patients. Esophageal and other gastrointestinal cancers are outside the scope of this review.

Population labels introduce a second interpretive problem. Genetic ancestry is estimated from similarity to reference populations and depends on the markers, panels, and analytical methods used. Self-identified race or ethnicity and geographic recruitment record other aspects of population history, social conditions, exposure, and health care. These variables may be correlated, but they are not interchangeable. In particular, a broad ancestry category does not define a fixed biological tumor type [1]. An analysis should therefore retain the population measure used in the original study.

We examine four mechanisms that could connect host susceptibility to somatic change: mutation acquisition, early clonal establishment, microenvironment-dependent selection, and treatment-driven expansion of resistant clones. Within each mechanism, population evidence is distinguished from tumor microevolution within an individual. Population evidence concerns variation in inherited risk, exposure, tumor profiles or outcomes between people; microevolution concerns the origin, persistence and changing representation of somatic lineages. Their connection is evaluated only where a study measures the relevant intermediate process.

Related reviews address complementary questions. Saoudi González and colleagues examine heterogeneity and clinical applications in metastatic CRC [2]; Toghey and colleagues integrate CRC cell biology, epigenetic regulation, and clonal evolution [3]. Fan and Zheng compare multi-omic evidence, microenvironmental selection, and translation in HCC, esophageal squamous cell carcinoma, and gastric cancer [4]. The present review concentrates instead on connecting host susceptibility and etiologic exposures to measured somatic processes. It uses hereditary syndromes, normal epithelium, mutagenic exposures, and serial resistance measurements to identify where that connection is supported and where an incidence or case-only association remains the principal evidence.

2. Review Methods and Evidence Interpretation

2.1. Literature Identification and Scope

This is a critical narrative review. The original targeted searches of web-indexed PubMed/MEDLINE records, Europe PMC, and primary journal websites were updated in October 2026. Terms combined the five cancer groups with genetic ancestry, germline and somatic variation, GWAS, ExWAS, rare variants, EWAS, exposure-wide association, TWAS, RWAS, mutational signatures, normal epithelium, multiregion sequencing, clonal evolution, metastasis, and resistance. For the present revision, targeted searches on 11 November 2026 assessed reviewer-suggested publications and methodological sources on clonal quantification, MSI testing, error-suppressed ctDNA sequencing, and cfDNA fragmentation. These procedures did not constitute exhaustive systematic screening.

Peer-reviewed human primary studies were prioritized when they measured susceptibility, a somatic feature, or within-patient architecture relevant to the review question. Large association studies established the population context; paired tumor–normal, multiregion, normal-tissue, and serial studies supplied the mechanistic examples. Experimental organoid evidence was included to assess a specific mutagenic mechanism. Reviews provided context and comparison, not independent replication. Conference abstracts and preprints were not

used as central mechanistic evidence. No new genomic data were analyzed and no effect estimates were pooled.

Bibliographic checks used DOI-linked records in Europe PMC, PubMed and journal websites, including titles, authors, journals and available publication details. Targeted update queries combined pancreatic cancer with BRCA, reversion, rucaparib, and cell-free DNA, and colorectal cancer with colibactin, pks, organoid, and mutational signature. Abstracts were inspected for study design and principal results, with available full texts used to check key denominators and mechanistic claims. Verification depth varied with access. We extracted patient and sample counts separately, the biological compartment, population definition, temporal design, molecular measurement, main finding, and the untested step in a proposed evolutionary explanation. The tables are a purposively selected evidence map, not a complete registry or a formal risk-of-bias assessment.

2.2. Population Features and Tumor Microevolution

We distinguish two analytical levels and three outcomes. At the population level, studies measure cancer occurrence or the molecular features of tumors across individuals. At the within-patient level, studies reconstruct somatic lineages or measure their change over space and time; we refer to this as tumor microevolution. Spatial architecture can inform an evolutionary reconstruction, whereas a trajectory requires temporal evidence or explicit modeling. Population differences are not themselves evidence of evolutionary change in human populations. Case–control, case-only and within-patient comparisons therefore remain separate throughout the four mechanistic sections.

For each major study, we considered the number of patients, assay, population definition, timing of sampling, and evidence for an evolutionary process. Patient counts were kept distinct from counts of cells, glands, or regions. Dense sampling can resolve variation within a patient without increasing the number of independent patients available for comparison. We also distinguished observed temporal change from spatial coexistence or computational pseudotime.

We distinguish documented constitutional variants from inferred ancestry, self-reported population labels, and clinical syndrome diagnoses. Germline pathogenic variants require constitutional testing and variant classification; tumor-only calls or a family history alone do not establish inherited status. In the FAP study discussed below, five of six clinically diagnosed individuals had a detected germline APC truncating variant. The rucaparib cohort included both germline and somatic repair-gene variants and is not exclusively a hereditary cohort. Broad population profiling studies are treated as associations unless the proposed molecular transition was directly tested.

2.3. Interpretation of Study Designs

We evaluated functional experiments, lineage reconstructions and longitudinal observations according to the process each measured. A cellular experiment may establish a growth effect without showing its contribution to clinical progression. We did not apply a formal evidence-grading instrument. Table 1 summarizes the study designs and the inferences they support; Figure 1 distinguishes their sampling units and outcomes.

We assess mechanistic evidence at the level of the process actually observed: a functional experiment, a reconstructed lineage, or a measured change over time. Statistical co-occurrence is treated as association evidence. Sections 3–6 each address one mechanism and distinguish population findings, within-patient microevolution and the evidence connecting them. A lineage-resolved observation can directly establish clonal architecture without establishing that inherited variation caused it. Differences in study design and endpoints preclude a common causal effect estimate.

Table 1. Study designs and the outcomes they can support.

Design and Material	Principal Outcome	Permissible Inference	Additional Evidence Needed for Evolution
GWAS or PRS; constitutional DNA	Cancer occurrence or subtype risk	Inherited susceptibility in the sampled population	Tumor lineages, timing, or clonal change measured against genotype
ExWAS; germline coding variants	Variant or gene-burden association with cancer	Rare coding contribution to susceptibility	Paired tumor analysis, allelic state, and lineage evidence
TWAS, methylation-based association, or RWAS; genetically predicted traits	Association of predicted expression, methylation, or accessibility with disease	Candidate regulatory mechanism; causality requires further tests	Clonal or temporal measurements beyond the risk endpoint
Epigenome-wide association; measured methylation	CpG or regional association with risk, diagnosis, or prognosis	Compartment- and time-specific epigenetic association	Heritable state variation, lineage linkage, and differential expansion
Exposure-wide analysis; measured exposures or proxies	Incidence, prevalence, or outcome association	Epidemiologic association with the specified endpoint	Validated individual exposure, mechanism, and somatic measurements
Case-only tumor profiling	Somatic alteration or subtype prevalence	Molecular differences among sampled cancer cases	Mutation timing, clone dynamics, or validated selection analysis
Multiregion or serial genomic profiling	Relatedness, clonal architecture, or change over time	Evolutionary reconstruction conditional on sampling and models	Independent replication and genotype/exposure comparison to bridge levels
Single-cell RNA or spatial transcriptomics	Cell states and ecological organization	Phenotypic or spatial heterogeneity	Genomic lineage support for claims about genetic clonality

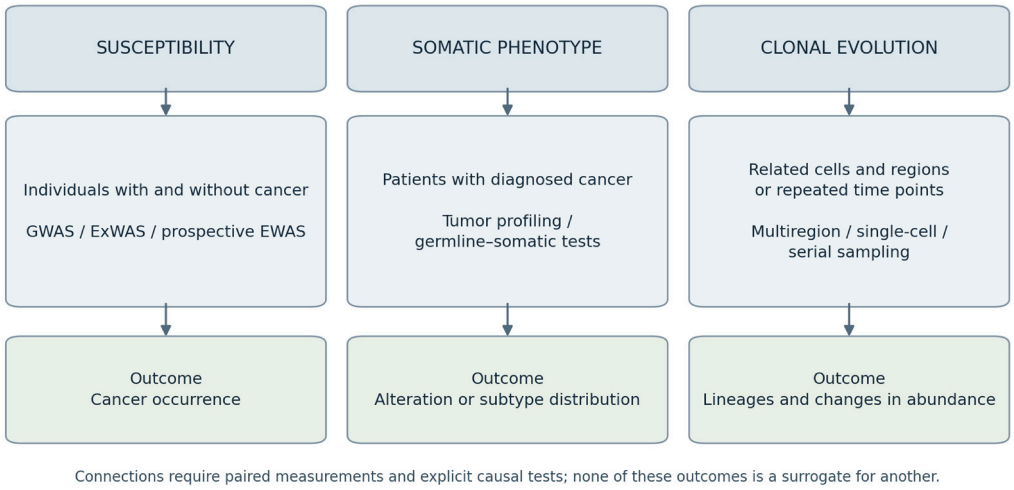


Figure 1. Analytical levels and outcomes used throughout the review. Susceptibility and between-patient somatic profiles constitute population-level evidence; clonal architecture and change concern microevolution within a patient. These outcomes are compared within each mechanistic section. The figure is conceptual and contains no quantitative data.

2.4. Interpreting Population Associations

GWAS tests common inherited variants against disease occurrence, whereas ExWAS examines coding variants individually or in gene-level aggregates. Sequencing a tumor

exome alone does not constitute an association study. TWAS and regulome-wide analyses relate genetically predicted expression or accessibility to disease; epigenome-wide studies usually analyze measured methylation. The latter depend strongly on the sampled compartment and timing. Here, ExWAS denotes exome-wide association; exposure-wide studies are written out to avoid ambiguity. Differences in allele frequency, linkage disequilibrium, imputation, assay coverage and recruitment affect comparisons across populations.

Colocalization can assess whether a molecular association and a disease association share a causal signal. Mendelian randomization addresses an exposure–outcome hypothesis under assumptions about instrument relevance, independence, and exclusion restriction. Both can sharpen mechanistic hypotheses, but the outcome being analyzed remains decisive. An incidence association, however well supported, does not supply a measurement of clonal expansion or metastatic competence.

Interpreting EWAS requires attention to cell composition, technical batch effects, the interval from sampling to diagnosis, and reverse causation. Replication should match the original biospecimen and clinical setting. A functionally important methylation change may be transient, disease-induced, or unrelated to lineage expansion. To study epigenetic evolution, investigators must additionally establish heritable state variation and differential expansion, ideally through repeated sampling or linkage to genomic lineages.

Gene–environment analyses ask whether an exposure association varies with inherited susceptibility. They require adequate power, a specified interaction scale, multiple-testing correction, and replication. Even a reproducible statistical interaction needs mechanistic investigation before it can be interpreted as cooperation within a cancer cell. Table 2 summarizes representative susceptibility studies; their mechanistic relevance is assessed in Sections 3–6.

John and colleagues documented germline variants and clinicopathological associations in 192 Indian patients with triple-negative breast cancer [5]. Although outside the digestive-cancer scope, this study illustrates the contribution of constitutional testing in an underrepresented population. Its relevance here concerns ascertainment of inherited susceptibility; tumor lineage dynamics were not examined.

Population comparisons address different questions: the overall distribution of somatic alterations among diagnosed cases, differences within a molecular subtype, or the effect of a particular inherited determinant. These questions require different analyses. Essential study descriptors include the ancestry inference method and reference panel, self-identified ethnicity, recruitment geography, age, sex, family history and germline testing. Tumor site, histology, stage, sampling location and date, prior treatment, and access to testing describe the clinical context; assay platform, callable regions, depth, purity and copy number describe measurement. CRC comparisons also require MSI/MMR status, hypermutation and sidedness, whereas HCC comparisons depend on HBV, HCV, alcohol, metabolic liver disease and relevant aflatoxin exposure.

The role of each covariate depends on the proposed causal pathway and its timing. Platform and callable regions require technical harmonization; demographic, exposure and recruitment factors may require adjustment or standardization. Molecular subtype, repair loss, inflammation and stage can lie downstream of the determinant under study. Adjusting for these variables may remove part of the association of interest or introduce selection bias. A prespecified causal diagram helps distinguish these roles. Reporting overall and subtype-specific estimates separately, with sensitivity analyses for missing etiology and cohort selection, makes the interpretation more transparent. Estimating a direct effect additionally requires mediation assumptions that are often unverifiable in these datasets.

Table 2. Population association studies and the limits of their contribution to mechanistic inference.

Study and Design	Cohort or Material	Contribution	Boundary of Interpretation
CRC multi-ancestry GWAS [6]	100,204 cases; 154,587 controls; European and East Asian ancestries	Common-variant susceptibility architecture with molecular integration	Incidence association; no direct clonal endpoint
HBV-associated HCC GWAS [7]	1161 cases and 1353 carrier controls in discovery; additional carrier cohorts	STAT4 and HLA-DQ susceptibility associations	Risk conditional on HBV carriage; no tumor lineage comparison
CCA GWAS [8]	2366 cases; 11,750 controls; European ancestry	Inherited susceptibility and subtype analyses	Anatomical and etiologic composition affect transferability
PDAC whole-exome association [9]	1591 cases; 2134 controls; European ancestry	ATM gene-level signal; SIK3 candidate at an FDR threshold	Rare carrier counts and broad effect uncertainty; no trajectory measurement
CRC ExWAS [10]	146 high-risk Middle Eastern cases; 1395 controls	APC and additional candidate coding associations	Selected high-risk cohort; candidates require independent replication
Gastric QTL and TWAS [11]	262 gastric biopsy donors; GWAS of 10,254 cases and 10,914 controls	Expression, splicing, and polyadenylation mechanisms	Regulatory susceptibility; no direct measurement of clonal selection
PDAC TWAS and RWAS [12,13]	TWAS: 8803 cases and 67,523 controls; RWAS: 8275 and 6723	Candidate expression and accessibility mechanisms	Predicted molecular traits; potential source-cohort overlap
HCC blood methylation EWAS [14]	48 discovery case–control pairs in an HBV-carrier cohort; validation datasets	Prediagnostic leukocyte methylation associations	Blood compartment, immune composition, and chronic liver disease matter
CRC blood methylation EWAS [15]	955 women; 29 incident CRCs; tissue follow-up datasets	Prediagnostic CpG and regional signals	Few events; tissue confirmation is not external prospective blood validation
Cancer exposure-wide analysis [16]	NHANES; 55,021 participants for prevalence and 5163 survivors for mortality	Screening of environmental and biological associations	Broad cancer outcomes; prevalence and mortality do not measure evolution

2.5. Measuring Microevolution Within Patients

Somatic evolution involves heritable variation and changes in the representation of cell lineages through mutation, selection, drift, and tissue population structure. Mutation burden describes only part of this process. A large clone may have arisen early, grown rapidly, survived competing lineages, or occupied a favorable niche. Its size alone cannot distinguish those explanations. In bulk sequencing, purity, copy number, ploidy, and sampling further complicate the interpretation of variant allele fractions.

Multiregion sequencing reconstructs relationships among sampled lineages, while serial samples reveal changes over time. Both leave unsampled or extinct clones unresolved, and molecular-clock estimates remain model-dependent. Single-cell DNA sequencing can improve resolution at the cost of coverage and error constraints. RNA and spatial assays contribute information about cellular function and tissue organization; common expression states, however, need not imply common clonal ancestry.

Clonal quantification depends on allele counts, tumor purity, allele-specific local copy number, ploidy and mutation multiplicity. Under a simplified mixture with diploid normal cells, expected VAF = $p \times f \times m / [p \times C + 2(1 - p)]$, where p is purity, f is cancer cell fraction (CCF), m is mutant-copy multiplicity and C is local tumor copy number. The expression assumes a shared copy-number state; subclonal copy-number alterations require a joint model. Methods such as PyClone estimate cellular prevalence while accounting for normal contamination and allelic imbalance [17]. Confidence or posterior intervals should be reported, including alternative copy-number solutions where these change the clonal assignment. Benchmarking across reconstruction tasks shows that performance depends on the task and input data, supporting explicit assessment of uncertainty [18].

Phylogenetic reconstruction is more convincing when mutation clusters recur consistently across regions or time points. Mutation loss and copy-number changes can support alternative trees, which should be retained when the data cannot distinguish them. Mutation multiplicity can place a mutation before or after a copy-number gain; conversion to calendar time requires further assumptions about mutation rates [19]. Longitudinal comparisons should document treatment intervals, use consistent processing and include pretreatment material and matched anatomical sites where feasible. Genomic lineage information is needed to relate expression pseudotime or spatial cell states to these histories.

An increase in CCF indicates a change in clonal representation only when it exceeds measurement uncertainty. Inferring selection additionally requires comparison with explanations involving drift, bottlenecks, migration and sampling. Model-based estimates depend on growth and measurement assumptions; weak selection may remain undetectable [20]. Concordant changes across lineage markers, independent resistant lineages and functional competition experiments strengthen the inference. In a two-clone population with constant net growth rates and no migration, the change in $\ln[f/(1 - f)]$ per unit time estimates the difference between those rates. This relationship assumes that f measures the same cell population at both times. Differential shedding or changing biopsy composition can violate that assumption. No single VAF or CCF threshold establishes selection across diseases and assays.

3. Genetic Background and Mutation Acquisition

Inherited variation may affect the production of somatic mutations through DNA repair or other cellular processes. The relevant distinction is between a population association with cancer or a mutation profile and evidence that a defined defect changes mutagenesis within a lineage. Environmental mutagenesis provides a useful comparison but is considered separately from inherited effects.

3.1. Population Associations with Repair Genes and Somatic Profiles

An early PDAC ExWAS compared 437 cases with 1922 controls. Rare inactivating BRCA2 variants showed the strongest enrichment, with other genes reaching suggestive thresholds [21]. In a later whole-exome analysis of 1591 cases and 2134 controls of European ancestry, ATM was the leading gene-level association; SIK3 reached an FDR-based threshold but not the Bonferroni threshold [9]. These results differ in statistical support and should not be presented as an equally established set of risk genes.

Another CRC study, involving 2327 cases and 2966 controls, identified ST6GALNAC2 as its leading gene-level result and reported Bayesian evidence for DNA-repair pathway involvement [22]. This supports pathway-directed tumor studies without assigning all associated genes the same repair phenotype. Interpretation depends on annotation, aggregation masks, frequency cutoffs, coverage, and multiple-testing strategy; a pathway signal does not establish a shared mutational signature in carriers.

Somatic profiling compares tumors from people who have already developed cancer. The observed alteration frequencies depend on who develops the disease and who enters the sequencing cohort. Tumor site, stage, subtype, prior treatment, testing access, and survival can all influence that composition. When an alteration is more common in one group, its prevalence alone cannot distinguish more frequent acquisition from greater subsequent expansion or preferential sampling.

Amuzu and colleagues meta-analyzed two panel-sequencing cohorts containing 275,605 samples across 14 cancer types. They identified ancestry-associated somatic differences and raised concerns about panels developed largely from European-ancestry tumors [23]. The analysis offers broad evidence of molecular differences among sampled cases. It lacks the temporal or lineage measurements needed to resolve how those differences arose. Source-cohort overlap also needs checking before related publications are counted as independent replication.

Among 46,140 profiled colorectal adenocarcinomas, ancestry analyses identified 5301 African-descent and 33,770 European-descent patients. KRAS, APC, and PIK3CA alterations were more frequent in the former group, whereas MSI-high tumors were less frequent [24]. In a separate series of 8454 predominantly late-stage patients, greater continuous African ancestry was also associated with higher odds of KRAS, APC, and PIK3CA mutations [25].

The Latino Colorectal Cancer Consortium extended this work to 718 tumor–normal pairs and tested findings in an independent MSK-IMPACT dataset [26]. Several replication estimates had the same direction without reaching individual significance. Thus, the studies show recurring associations, but not uniform replication. Age, tumor location, MSI, hypermutation, treatment, and panel coverage complicate causal interpretation, particularly for uncommon alterations.

The meta-analysis of TP53, APC and PIK3CA prevalence documents considerable variation between CRC studies [27]. Its pooled estimates provide background for population comparisons, although heterogeneity limits their use as ancestry-specific reference rates. Mutation frequency combines acquisition with the subsequent survival and expansion of affected cells. The allele-specific KRAS experiment discussed in Section 4.2 illustrates why function must be considered alongside these frequencies.

Paired host and tumor data can test a specific inherited determinant against a specific somatic event. In 5954 tumors, Carter and colleagues reported germline–somatic interactions, including an RBFOX1-associated enrichment of SF3B1 mutation and a 19p13.3 association with PTEN mutation, with functional support for selected relationships [28]. These findings are more specific than broad ancestry contrasts. They nevertheless leave mutation acquisition and subsequent selection partly entangled: enrichment among sampled cancers is compatible with more than one route to the observed phenotype.

3.2. Microevolution Following Mismatch Repair Loss

Lynch syndrome separates inherited predisposition from the somatic transition that changes mutagenesis. Whole-genome sequencing of 152 normal and neoplastic crypts or glands from 10 carriers found near-baseline mutation processes in most normal epithelia; a morphologically normal MMR-deficient crypt was an exception [29]. The tumor crypt phylogenies placed MMR deficiency before the most recent common ancestor of sampled tumor cells. The inherited state therefore creates an opportunity for a later repair-deficient lineage rather than imposing uniform hypermutation across normal tissue.

A comparison of 14 hereditary and 29 sporadic MSI-CRCs found similar mutation landscapes but greater allele-fraction-based heterogeneity in sporadic tumors [30]. Median ages were 41.5 and 74 years, respectively. The contrast is therefore not a controlled estimate

of an inherited effect on evolutionary rate. Together, these studies support a link through MMR-deficient mutagenesis while showing that the origin of deficiency does not uniquely specify the eventual landscape or architecture.

MSI is a tumor phenotype and is not synonymous with Lynch syndrome. CRC assessment uses MMR immunohistochemistry for MLH1, PMS2, MSH2 and MSH6 and/or validated PCR-based MSI testing; validated NGS-based MSI assays are also an option in CRC [31]. With MLH1/PMS2 loss, MLH1 promoter methylation and, in CRC, BRAF V600E testing help evaluate a sporadic origin; unexplained deficiency warrants germline assessment and, when appropriate, paired tumor analysis for double-somatic events [32]. A negative BRAF result alone does not establish Lynch syndrome. Clinical context remains important, including rare constitutional MLH1 methylation.

Rajoua and colleagues compared HSP110 T17, CAT-25 and a mononucleotide pentaplex panel in 96 sporadic CRCs, including nine MSI cases [33]. HSP110 T17 matched the pentaplex classification in this cohort, whereas CAT-25 performed less well. Larger validation cohorts are needed before generalizing this performance. These assays identify the MSI phenotype; its hereditary or sporadic origin requires separate investigation.

3.3. Environmental Mutagenesis as a Mechanistic Comparison

EWAS can also mean environment-wide association study, in which many exposures or biological measurements are screened together. A recent NHANES analysis examined hundreds of factors in relation to cancer prevalence and mortality [16]. It illustrates this approach across broad cancer outcomes rather than within a specific gastrointestinal malignancy. Interpretation depends on the endpoint: prevalence, incidence, cancer mortality, and all-cause mortality differ, and measurements obtained after diagnosis may reflect treatment or disease consequences.

Methylation proxies may help investigate exposures when direct histories are unavailable, but their endpoint must be specified. Maas and colleagues compared early-onset CRC with CRC diagnosed at age 70 years or older, using tumor methylation scores as exposure proxies and county-level pesticide data for ecological follow-up [34]. This is principally a contrast between age-defined cancer groups, not a prospective comparison of exposed and unexposed cancer-free individuals. A proxy association cannot identify an individual's dose, establish that methylation preceded the tumor, or reconstruct clonal history.

Experimental exposure studies can resolve the acquisition side of the problem. Pleguezuelos-Manzano and colleagues repeatedly exposed human intestinal organoids to colibactin-producing pks+ E. coli and compared their genomes with organoids exposed to an isogenic pks-mutant strain. Exposure produced a characteristic mutational pattern that was also found in human cancers [35]. This supports a causal connection between a microbial exposure and mutation supply. It does not, by itself, measure a selective advantage of the resulting mutant cells or the probability that an exposed individual develops cancer.

A complementary whole-genome study of 981 CRCs from 11 countries linked colibactin-associated SBS88 and ID18 signatures to geographic and age variation and to early APC driver events [36]. In conjunction with the organoid experiment, this connects a defined mutagenic process to mutations present in human tumors. The remaining steps include individual exposure timing, progression of the affected precursor, and the relative contribution of altered mutation supply and later selection. Geography in these analyses should not be recast as an inherited ancestry mechanism.

3.4. What Connects Population Findings to Mutation Acquisition

MMR loss and colibactin exposure provide evidence for altered mutation production. They leave a further question: which affected lineages persist and progress in human tissue?

The organoid experiment directly tests exposure-induced mutagenesis, and the Lynch crypt genomes locate repair loss within a somatic history. Repair-gene burden tests and ancestry-associated mutation frequencies identify candidate relationships but cannot resolve these transitions. Linking these levels requires paired constitutional and tumor data, confirmation of the relevant somatic repair transition, and measurements of mutation accumulation or signatures within comparable tissue and disease settings.

4. Genetic Background and Early Clonal Establishment

A susceptibility variant may influence whether a newly altered cell persists and contributes to a lesion. This mechanism concerns establishment after an alteration has occurred. Population associations nominate candidate genes, functional experiments test cellular effects, and lineage studies reveal how many populations contribute to a precursor; each resolves a different part of the proposed connection.

4.1. Population Susceptibility and Regulatory Candidates

In CRC, Fernandez-Rozadilla and colleagues combined 100,204 cases and 154,587 controls of European and East Asian ancestries and integrated molecular data to investigate susceptibility loci [6]. This scale improves risk-locus resolution and biological interpretation. To ask whether a polygenic score also predicts tumor diversity, one would need a separate analysis among cases with consistent subtype, sampling, and clonal endpoints.

A 2026 CCA study included 2366 cases and 11,750 controls of European ancestry and examined inherited susceptibility and clinical subtypes through GWAS and post-GWAS analyses [8]. This work extends association mapping to a less common malignancy. Its findings should be interpreted alongside anatomical subtype and primary sclerosing cholangitis. Geographic variation in somatic IDH1 or FGFR2 alterations addresses a separate question and cannot be assumed to arise from the susceptibility loci identified in this cohort.

Polygenic risk scores (PRS) could contribute to prevention or surveillance if their calibration, discrimination, and clinical value hold in the population where they are used. Performance may change across populations because of both genetic differences and study design. Validation for incidence would still leave their value for predicting tumor progression or resistance unresolved.

A Chinese exome-array study of PDAC used 943 cases and 3908 controls for discovery and 2142 cases and 4697 controls for replication [37]. The PKN1 coding variant was supported by experiments linking it to increased phosphorylation and FAK/PI3K/AKT signaling. In CRC, a multistage study linked a TCF7L2 missense risk variant to a MYC enhancer variant and increased cellular proliferation [38]. These examples identify a functional route from inherited variation to a growth-related phenotype. Whether carriers' tumors preferentially acquire or expand particular somatic lineages remains a separate patient-level question.

Regional studies can uncover rare variants in populations poorly represented in international datasets. A Middle Eastern high-risk CRC ExWAS, for example, compared 146 cases with 1395 controls and reported APC and additional candidate associations [10]. Its selected familial or early-onset population differs from an unselected CRC series. Sparse carrier counts and differences in sequencing or annotation can also influence burden estimates. Replication, segregation analysis where appropriate, and functional assessment are therefore especially important; discovery in a regional cohort does not establish that a variant is unique to that population.

A gastric study mapped regulatory QTLs in 262 Chinese individuals and integrated them with GWAS data from 10,254 cases and 10,914 controls. Expression, splicing, and alternative polyadenylation analyses nominated candidate genes; selected mechanisms,

including DIP2B and RMC1, underwent experimental follow-up [11]. For PDAC, a TWAS included 8803 cases and 67,523 controls [12]. A regulome-wide study of 8275 cases and 6723 controls used pan-cancer chromatin-accessibility prediction models rather than direct measurements of pancreatic precursor evolution [13]. Tissue relevance and shared source cohorts therefore matter when interpreting these functional extensions of risk mapping.

A GENIE analysis compared recorded racial groups among 4327 patients with PDAC. Most genomic profiles were similar, with fewer CDKN2A mutations in the Asian group [39]. This result concerns race as recorded in the dataset and mutations as called by the assay. It does not encompass every form of CDKN2A inactivation, such as deletion, or establish a corresponding difference in tumor aggressiveness.

4.2. Microevolution of Precursors and the Persistence of Early Lineages

A study of clinically diagnosed familial adenomatous polyposis analyzed 123 tissue samples from six individuals, five with a detectable germline truncating APC variant [40]. Genomic reconstruction supported polyclonal origins in subsets of benign and dysplastic polyps; additional single-crypt sequencing supplied corroborating evidence. This demonstrates that hereditary predisposition can coexist with early lesions containing distinct lineages. The analysis was cross-sectional, and dense polyposis, lesion collision, and purity remain relevant to interpretation.

These lesions raise an allele-specific question: do particular APC variants change the probability that several lineages become established in one precursor? Comparison with matched sporadic lesions would help separate inherited effects from the tissue context of dense polyposis.

The Big Bang model was developed from analysis of 349 glands in 15 colorectal cancers. It explains how substantial heterogeneity can arise early and persist through tumor expansion without repeated population-wide selective sweeps [41]. It need not describe every CRC or every phase of progression. Later spatial and multi-omic mapping of 31 colorectal specimens found individualized progression trajectories with coordinated epithelial and microenvironmental changes [42].

At the allele level, experimental work linked KRAS p.Ala146Thr to CDK1-mediated ERK signaling and CRC cell proliferation [43]. This provides functional support for a growth-related effect after mutation acquisition. Whether the same effect determines clone fixation in normal mucosa or explains population differences in mutation prevalence remains untested.

Metastatic studies provide context for the persistence of established lineages, rather than direct tests of early clone establishment. Multiregion profiling of 106 samples from 10 patients with advanced gastric cancer identified monoclonal and polyclonal seeding [44]. PDAC reconstructions estimated intervals before dissemination [45], while sequencing showed that untreated metastases often shared known driver mutations, consistent with a multicancer study [46,47]. These observations show that dissemination can preserve earlier lineages, but neither their architecture nor model-dependent timing isolates an inherited effect on initiation.

4.3. From Inherited Growth Effects to Clone Establishment

The PKN1 and TCF7L2 experiments support susceptibility-related signaling and proliferation [37,38]; FAP reconstructions demonstrate multiple lineages in some lesions [40]. The missing comparison is the probability of clone establishment in carriers and noncarriers under comparable tissue conditions. Genotype-defined cohorts and matched sporadic precursors could measure persistence while distinguishing independent initiation from lesion collision.

5. Tissue Microenvironment and Clonal Selection

Inflammation, immune recognition and stromal interactions may change which somatic lineages persist. Population measures of exposure or immune context suggest where selection might differ; within-patient genomic and spatial measurements are needed to identify the cells undergoing that selection.

5.1. Population Differences in Etiology and Tissue Context

For HCC, the underlying liver disease helps define the relevant risk comparison. A study of chronic HBV carriers identified HCC associations at STAT4 and HLA-DQ and replicated them in additional carrier cohorts [7]. Using carrier controls reduced variation in the major infectious exposure. The remaining association could involve host control of infection, inflammation, or liver injury, as well as other processes leading to HCC. Distinguishing those mechanisms requires more than the case–control result.

In a prospective HBV-carrier cohort, 48 matched case–control pairs formed the discovery set for leukocyte methylation patterns associated with subsequent HCC, with additional validation datasets [14]. The analysis considered viral load, hepatitis status, and leukocyte composition. Sampling before diagnosis improves the temporal interpretation, although immune state, exposure, or occult disease may still explain an association. The assay measures blood methylation rather than competition among hepatocyte clones.

A CRC study analyzed prediagnostic leukocyte methylation in 955 Women's Health Initiative participants, of whom 29 subsequently developed CRC, and examined selected signals in independent tissue datasets [15]. The small event count limits precision and model stability. The tissue analyses add biological context, but they do not constitute prospective validation of a blood-based predictor in a second population.

The TCGA and ACRG classifications divide gastric adenocarcinoma into molecular subtypes with differing biological features [48,49]. Population comparisons must account for these subtypes alongside histology, tumor location, infection, and exposure. Otherwise, an apparent ancestry association may largely reflect different proportions of the constituent diseases.

A study of 1042 patients with inferred ancestry reported differences involving CCNE1, KRAS, and PI3K-pathway alterations [50]. An analysis of 1335 gastric cancers found driver patterns associated with subtype, infection, and ancestry, including a suggested link between alcohol-related factors and RHOA alterations in East Asian ancestry [51]. The latter illustrates a useful mechanistic question: does exposure or metabolism change mutation acquisition, or does the tissue favor cells carrying that alteration? Case-only frequencies do not distinguish these explanations. Microbial-read associations also depend on contamination control and assay validation.

A profiling study of 2372 HCCs and 150 combined HCC–CCA cases reported ancestry-associated alterations, but detailed clinical etiologic information was limited [52]. This is a substantial constraint in liver cancer, where HBV, HCV, metabolic disease, alcohol, and aflatoxin contribute unevenly across settings. An East Asian ancestry association with TP53, for example, cannot separate ancestry from these contributors. Combined HCC–CCA also requires separate interpretation because of its distinct diagnostic and molecular context.

Anatomical composition is equally important for BTCs. An analysis of GENIE and other datasets found geographic and site-related variation involving TP53, IDH1, and FGFR2 [53]. Western and Eastern recruitment indicate geography rather than measured genetic ancestry. Small-duct and large-duct iCCA, extrahepatic CCA, and gallbladder cancer also differ biologically. Comparisons must retain both anatomical categories and alteration classes: all FGFR2 alterations cannot be used as an estimate of therapeutically relevant FGFR2 fusions.

5.2. Microevolution in Normal Tissue and Tumor Niches

Normal epithelium provides a different kind of evidence. Gastric sequencing included 238 microdissections from 30 individuals, 18 of whom had gastric cancer, and an additional targeted survey of 829 microbiopsies [54]. Mutant-clone prevalence was associated with age and severe inflammation. In normal colon, targeted analysis of 76,800 glands from 56 patients, supported by additional sampling approaches, distinguished preferential crypt fixation of FBXW7 R465C from enhanced expansion of KRAS G12 clones [55]. These are tissue-dependent somatic effects. The specimens were not a prospective cohort followed from a cancer-free state, so their findings do not directly estimate the subsequent cancer risk of an individual clone.

A multiregion study combined genome, transcriptome, and chromatin-accessibility measurements from 495 glands in 29 CRCs, with spatial neoantigen and imaging analyses in a further set of microbiopsies [56]. It implicated early antigen-presentation and neoantigen silencing in immune escape. The relevant evidence is tumor regulatory change placed within genomic architecture, whereas leukocyte EWAS measure a different compartment and outcome. Both concern epigenetics, but only the former design directly addresses its position in tumor evolution.

Integrated multiregion profiling established genetic, transcriptional, and immune heterogeneity within HCC [57]. A prospective treatment-naive cohort of 123 patients, providing more than 600 genome and transcriptome samples, subsequently identified convergent evolutionary paths [58]. The most aggressive tumor fraction could be more informative for prognosis than overall transcriptional heterogeneity, indicating that the identity of a heterogeneous component may matter more than diversity alone.

Analysis of 111 HCC samples combined digital pathology and molecular methods to examine hyperploid foci and intratumoral diversification [59]. This links tissue morphology to genomic structure. The cohort was not designed to isolate ancestry effects from geography or underlying liver disease; longitudinal treatment-associated state changes are considered in Section 6.2.

In iCCA, multiregion WES of 69 regions from six tumors revealed pronounced branching and frequent branch-associated drivers [60]. Single-cell and spatial studies have since characterized immune niches and the leading edge of the tumor [61,62]. These observations relate tissue organization to the genomic findings without making the assays interchangeable. Their applicability to extrahepatic CCA and gallbladder cancer remains to be tested.

Shared drivers do not imply uniform cellular behavior. Spatial transcriptomics of 55 primary and metastatic samples from rapid autopsies of 13 patients with treatment-refractory PDAC found organ-associated variation in cellular states and microenvironments [63]. A separate center-versus-front study profiled four well-differentiated PDACs, with immunohistochemical assessment in 17 tumors [64]. These small, differently selected cohorts explain why apparent state diversity should be interpreted alongside treatment and histology, rather than attributed solely to dissemination or ancestry.

Combining spatial transcriptomics with genomic and imaging data can place subclones in their immune and stromal surroundings [65]. Copy-number alterations, whole-genome doubling, and extrachromosomal DNA add further routes to diversification beyond point mutation [66]. Such mechanisms broaden the study of somatic evolution, although a pan-cancer observation still needs disease-specific testing before it can explain a population difference. In early gastric lesions, spatial multi-omics and experiments in cell lines, organoids and mice implicated NAMPT- and AREG-related niche interactions [67]. These experiments support local progression mechanisms. Their dependence on inherited background or population ancestry was not tested.

5.3. Evidence for Selection and the Unresolved Host Contribution

Normal-tissue sequencing identifies mutation-specific fixation or expansion, whereas spatial assays locate candidate immune and stromal niches. To test whether a niche changes competition, studies need to place genomic lineages in that niche and measure their persistence or growth, preferably with a functional perturbation. Inflammation and leukocyte methylation associations help define the population context, but do not identify which epithelial lineage benefits.

6. Treatment Pressure and Resistant Clone Expansion

Treatment changes the environment in which tumor lineages compete and may also introduce new mutations. Inherited repair or immune variation can influence response, but a population-level association with outcome does not identify the lineage responsible for resistance. This section separates those associations from serial measurements of tumors under therapy.

6.1. Population Variation in Treatment Context and Outcome

POLO trial screening and a subsequent meta-analysis of germline BRCA1/2 variants highlight how testing access and recruitment influence geographic differences in observed carrier prevalence [68,69]. These are important issues for equitable genetic testing, but they are separate from somatic evolution. The population-profile datasets are compared in Table 3, with their original group definitions retained.

Table 3. Population features of somatic profiles across the five disease groups. These between-patient comparisons do not directly measure within-patient microevolution.

Cancer and Source	Cohort and Grouping	Main Contribution	Principal Limitation
Gastric cancer [50,51]	1042 and 1335 patients; inferred ancestry	Ancestry-, subtype-, and infection-associated driver patterns	Etiologic and subtype composition; no direct ancestry–trajectory test
CRC [24,25]	46,140 total profiled tumors; separate 8454-patient cohort; inferred ancestry	Recurring differences involving KRAS, APC, PIK3CA, BRAF, and MSI	Late-stage and panel ascertainment; covariate adjustment and overlap matter
CRC [26]	718 tumor–normal pairs; ethnicity and continuous ancestry; MSK replication	Extends analysis to admixed and Latino populations	Some replication estimates were directionally consistent but nonsignificant
HCC [52]	2372 HCCs; 150 combined HCC–CCAs analyzed separately; ancestry	Large real-world molecular landscape	Limited clinical etiology data; mixed histology must remain distinct
BTCs [53]	GENIE and other datasets; anatomical site and geographic recruitment	Geographic and site-related alteration patterns	Geography is not ancestry; alteration classes and subtypes differ
PDAC [39]	4327 patients; recorded racial groups	Lower CDKN2A mutation prevalence in the Asian group; broad similarity otherwise	Race is not inferred ancestry; mutations are not all inactivation events

Repair-associated PDAC links host genotype to tumor phenotype only when the relevant alleles and gene are specified. In 262 patients, Park and colleagues found that core homologous recombination genes and biallelic alterations refined associations with genomic instability and platinum outcomes [70]. Tumors from 25 germline ATM carriers showed frequent biallelic ATM inactivation and a distinct molecular landscape, emphasizing the need for gene-specific interpretation [71]. These studies describe repair-associated phenotypes; they do not imply that all repair-gene carriers share the same selective trajectory.

A retrospective study of 84 immunotherapy-treated gastrointestinal cancer patients, with an independent validation cohort, associated germline HLA-B evolutionary divergence with outcome and an inflamed tumor phenotype [72]. Here divergence describes inherited allele sequence differences. The study nominates host antigen-presentation capacity as a modifier of treatment response, but does not track the replacement of defined tumor lineages. Distinguishing host immune context from tumor immune escape requires measurements of both.

Inherited susceptibility may also be associated with treatment tolerance. Among 161 chemotherapy-treated CRC patients evaluated for suspected Lynch syndrome, a retrospective study reported more adverse reactions in those without a pathogenic Lynch-associated variant [73]. Differences in regimen, dose, age and ascertainment complicate interpretation. The result warrants further study of host determinants of toxicity; it provides no basis for syndrome-specific dose changes or conclusions about tumor-clonal evolution.

6.2. Microevolution During Therapy

Treatment creates a different selective setting. Whole-genome sequencing of 58 single-cell-derived tumoroids and 18 matched bulk samples from six patients with metastatic CRC identified heterogeneous treatment-associated mutational processes and late structural changes [74]. Culture sampling and the small patient count restrict generalization. Nevertheless, the study shows why a model of early expansion may not describe late treated disease: therapy can alter both mutation supply and the relative abundance of existing lineages.

Brown and colleagues examined paired cell-free DNA in 42 patients receiving maintenance rucaparib for BRCA1-, BRCA2-, or PALB2-associated pancreatic cancer; eligibility included germline or somatic variants. Reversion variants were newly detected in five of the 30 patients with plasma available at progression [75]. The changes were consistent with restoration of the affected reading frame and were associated with resistance. These serial observations connect repair-gene context to a plausible resistance mechanism, although they do not establish when the reversion first arose.

The clinical sequence depends on the tumor's repair phenotype and treatment exposure: repair deficiency may create a therapeutic vulnerability, while treatment can favor a population with restored repair function. Germline status, current tumor repair function and resistance history therefore provide distinct information [70,75].

Single-cell analysis of gastric cancer with peritoneal metastasis included matched primary, metastatic and pretreatment or progression ascites samples and identified resistance-associated states and immune interactions [76]. Longitudinal liver cancer transcriptomics likewise linked changes in cell states and ecology to treatment response [77]. These studies capture temporal phenotypic change; expression-derived trajectories require genomic support before they can be interpreted as replacement of somatic lineages.

6.3. Interpreting Serial ctDNA and Newly Detected Resistance

A reversion first detected at progression may have been present below baseline sensitivity or in an unsampled lesion. Increased shedding can also bring it above the detec-

tion threshold. Interpreting baseline nondetection requires locus-specific sensitivity and input-molecule counts. Relevant reporting includes plasma volume, cfDNA input, unique molecular depth, the error model, validated limits of blank and detection, allele counts with uncertainty, and confirmation criteria for low-frequency calls. The threshold is assay- and sample-dependent.

Unique molecular identifiers (UMIs), consensus reads and background-error suppression improve detection of rare variants [78]. Sensitivity still depends on the number of informative molecules sampled. Under ideal independent sampling and perfect detection, observing zero mutant molecules among N informative molecules gives an approximate 95% upper allele-fraction bound of $3/N$. Actual assays require validated error and recovery models. Matched leukocyte sequencing helps identify hematopoietic variants, while replicate or orthogonal testing can support uncertain low-frequency calls.

Plasma VAF is a molecular proportion whose relationship to tumor-cell abundance depends on copy number, total ctDNA fraction, lesion-specific shedding, cell death and clearance. Serial measurements of several stable truncal markers, supplemented by absolute mutant-molecule counts where validated, can aid interpretation alongside imaging and tissue sequencing. Differential shedding remains a limitation even after normalization. Evidence of resistance selection is stronger when serial genomic changes agree with tissue findings and functional restoration of repair.

6.4. Evidence Connecting Host Background to Resistant Expansion

Repair-gene studies associate a defined molecular context with response, and serial reversion measurements identify a plausible resistance route [70,75]. Functional restoration and expansion of the reversion-bearing lineage under treatment would complete the proposed connection. HLA-associated outcomes and Lynch-associated toxicity instead concern host response; neither directly measures resistant-clone dynamics. Studies addressing inherited effects on resistance need host genotype, baseline tumor phenotype, treatment exposure and serial clonal measurements in the same patients.

7. Integrating Population Features and Tumor Microevolution

7.1. Mechanistic Synthesis Across Analytical Levels

The evidence is strongest when a defined host or exposure determinant is linked to a measured somatic transition. Population comparisons establish how often the determinant or tumor feature occurs; lineage and longitudinal studies examine what happens within the affected tissue. Table 3 retains the population comparisons. Table 4 distinguishes experiments, genomic reconstructions, spatial observations and serial measurements for the four mechanisms, with the unresolved connection stated separately. Figure 2 summarizes the evidence needed to connect the two analytical levels.

Table 4. Four mechanisms linking population features to tumor microevolution. Direct evidence applies to the measured process and does not establish the entire causal chain.

Mechanism	Population Findings	Evidence and Study Design	Unresolved Connection
Mutation acquisition	Repair-gene and germline–somatic associations; ancestry-associated profiles [9,21,22,24–28].	Experiment: colibactin-induced mutations in organoids [35]. Genomic reconstruction: MMR loss in crypt lineages [29,30]. Tumor signatures: timing and geographic associations [36].	Does the determinant change mutation production in comparable human lineages, independently of later selection?

Table 4. *Cont.*

Mechanism	Population Findings	Evidence and Study Design	Unresolved Connection
Early clonal establishment	Risk loci, regulatory candidates and regional APC associations [6,8,10–13,37,38].	Functional assays: growth-related effects [37,38,43]. Genomic reconstruction: multiple lineages in FAP lesions and early CRC diversity [40,41].	Does an inherited allele alter clone persistence or establishment after accounting for tissue context and lesion collision?
Microenvironmental selection	HBV-conditional risk, blood methylation and etiologic composition [7,14,15,50–53].	Experiment: gastric niche interactions [67]. Genomic evidence: normal-tissue clone distributions and immune-escape architecture [54–56]. Spatial states: candidate niches, without proof of lineage selection [61–64].	Does host or exposure context cause differential lineage survival? Spatial coexistence alone cannot answer this.
Resistance during treatment	Repair phenotype, HLA-associated outcome and Lynch-associated toxicity [70–73].	Serial plasma: newly detected repair-gene reversions [75]. Genomic reconstruction: treatment-associated changes in CRC [74]. Neither dates the first resistant cell.	Was the lineage present at baseline, did repair function recover, and did treatment favor its expansion?

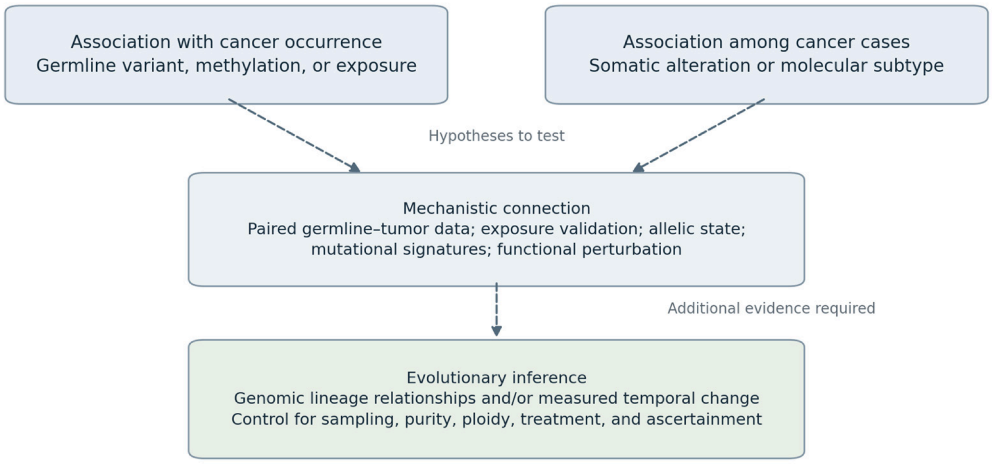


Figure 2. Connecting population-level susceptibility and somatic-profile associations to within-patient microevolution. Dashed arrows identify steps requiring additional evidence. Functional support for a cellular mechanism and a reconstructed tumor lineage answer different questions; both may be needed to connect a host determinant to clonal change.

The available evidence differs across cancers (Table 5). CRC provides links through hereditary precursors, MMR deficiency and microbial mutagenesis. PDAC supplies a serial repair–resistance example. Gastric studies connect inflammation and regulatory mechanisms to early tissue change. In HCC and BTCs, etiology, population profiles and tumor architecture have more often been measured separately. This difference reflects study design and coverage, not a demonstrated difference in how strongly susceptibility affects evolution.

Table 5. Disease-specific connections between population features and microevolution, with the principal missing comparison.

Disease Group	Most Informative Observed Connection	Remaining Separation	Priority Comparison
Gastric cancer [11,50,51,54]	Regulatory risk mechanisms; inflammation-associated normal clones	Host risk and later tumor branching mostly measured separately	Paired host genotype, exposure, normal glands, and precursor lineages
CRC [29,30,35,36,40,55]	MMR transition, early polyclonality, microbial mutagenesis, normal crypt selection	Mutational opportunity differs from progression probability	Matched hereditary/sporadic and exposure-defined precursor cohorts
HCC [7,52,57–59,77]	HBV-conditional risk; etiologic context and diverse tumor architectures	Ancestry, liver disease, and sampling remain partly entangled	Etiology-matched host–tumor cohorts with serial or multifocal sampling
BTCs [8,53,60–62]	Subtype-specific susceptibility and branching in iCCA	Anatomical subtype may account for a population contrast	Anatomically stratified paired germline and multiregion tumor analyses
PDAC [70,71,75]	Repair-associated phenotype and reversion-associated resistance	Gene, allelic state, and therapy determine the relevant comparison	Serial functional and genomic assays within defined repair-gene groups

7.2. Study Designs and Open Questions

A useful prospective design would begin with a defined determinant and a prespecified transition. For an inherited MMR variant, the endpoint could be the first detectable repair-deficient crypt, followed by its expansion or progression. For a microbial exposure, the endpoint could be acquisition of its signature and a driver event in the same precursor lineage. These designs separate time to initiation from subsequent growth and allow competing outcomes, including persistent nonprogressing clones.

Case-only host–tumor comparisons need an explicit account of selection into the cohort. Conditioning on diagnosis, survival, or clinical sequencing can induce associations between independent determinants. Molecular subtype may be a confounder in one question and a mediator in another; adjustment should follow the specified causal question rather than a universal covariate list. Age, therapy, assay coverage, ancestry estimation, and recruitment should be harmonized, with cells and regions treated as observations nested within patients.

The proposed endpoint should determine the assay and analysis. Mutation supply calls for signatures or measured mutation accumulation; selection calls for lineage expansion or validated competition; resistance calls for paired treatment and progression measurements. Genomic scars, expression pseudotime, and an incidence PRS cannot substitute for these endpoints. Negative results are informative if the study has sufficient carrier counts, sampling depth, and power for the clonal outcome rather than for cancer incidence alone.

Three comparisons could help separate susceptibility from subsequent evolution. An incidence PRS could be tested against clonal diversity within cases, with the estimand and the role of molecular subtype specified before adjustment. Repair-genotype cohorts could be followed for the appearance and persistence of repair-deficient lineages; conditioning on second-hit status would address a narrower question than the total effect of inherited predisposition. Serial resistance variants could be tested for information about progression beyond baseline repair phenotype and circulating tumor burden. Null results would also be informative if sampling and carrier numbers were adequate for the clonal endpoint.

7.3. Clinical Scenarios and Validation Requirements

These distinctions have practical consequences. Inherited susceptibility informs counseling and surveillance, whereas validated tumor biomarkers inform treatment in their established disease settings. Clonal measurements may explain resistance or the limitations of one biopsy, but their clinical use requires validation against outcomes. Population ancestry does not replace any of these individual assessments.

A truncal alteration is attractive as a target because many tumor cells share it. Its therapeutic value still depends on tumor dependence, an effective drug, resistance pathways, tissue context, and clinical evidence. Conversely, conservation of major driver genotypes does not make all lesions functionally equivalent. The PDAC spatial studies show how different cellular states and microenvironments can accompany shared drivers.

Ancestry-associated biomarker frequencies may change the proportion of patients eligible for an intervention without implying different efficacy in patients who carry the same biomarker. Diagnosis, testing, and treatment access also contribute to cohort differences. Measuring these factors directly offers a more informative account of disparities than assigning their combined effects to a population label.

Early detection in individuals at high genetic risk would combine constitutional testing with prospective surveillance. Investigational plasma assays could measure mutations, methylation or fragmentation, with abnormal findings confirmed by the relevant diagnostic procedure or tissue examination. Validation cohorts should include unaffected carriers, noncarriers and patients with benign disease. Stage-specific sensitivity, specificity, positive predictive value, interval cancers and false-positive procedures would establish diagnostic performance; a subsequent assessment of clinical benefit would determine whether the assay improves existing surveillance.

Before treatment, paired tumor and normal sequencing, supplemented by plasma and clinically feasible multiregion sampling, could characterize actionable alterations and baseline clonal architecture. MSI/MMR status, repair-gene allelic state and copy-number-adjusted clone estimates would define the relevant biological context. Independent cohorts receiving a specified regimen could then test whether clonal information improves prediction of response or progression-free survival beyond established biomarkers. Treatment selection would require evidence that acting on this information improves outcomes.

During treatment, baseline and serial plasma collected at prespecified intervals can be compared with imaging and clinical response. Error-suppressed sequencing or validated targeted assays, with matched leukocyte DNA, would help distinguish tumor variants from other sources. Prospective cohorts would need to quantify progression-detection sensitivity, lead time and false alarms, including performance in patients with low ctDNA shedding. Whether a molecular signal should trigger a treatment change is an interventional question.

After treatment failure, progression plasma and, where feasible, a biopsy of a progressing lesion can be compared with baseline material to investigate reversions, secondary mutations, copy-number changes and alternative pathways. The assay must cover the alteration class of interest, and functional evidence can help distinguish a resistance mechanism from an incidental finding. In biomarker-defined cohorts, the relevant endpoints are response, progression-free survival and quality of life after resistance-matched treatment. A newly detected variant is useful only insofar as its interpretation can inform such decisions.

7.4. Multidimensional Liquid Biopsy and Mechanistic Fragmentomics

Methylation and fragmentation can complement sequence-based plasma assays by providing information about tissue composition and the release and processing of DNA. Dong and colleagues tested a two-locus methylation assay in 103 patients with diverse cancers and 40 healthy participants, reporting sensitivity of 47.57% and specificity of 90.00% [79].

This exploratory result requires validation in the intended screening population, including hereditary-risk cohorts. Tissue-of-origin inference would additionally depend on suitable reference methylomes and independent evaluation.

Mechanistic fragmentomics seeks to explain fragment lengths, end motifs and nucleosome-related patterns through the processes that generate them [80]. Experimental studies implicating DNASE1, DNASE1L3 and DFFB provide a biological basis for this approach [81]. Interpretation in cancer remains challenging because inflammation, nonmalignant cell turnover, clearance and sample handling also affect fragment profiles. Their value as indicators of resistance or clonal change has yet to be established.

A prospective study could combine serial mutation and copy-number measurements with methylation-based estimates of tissue composition and fragment features linked to nuclease activity or chromatin organization. Somatic variants could identify the lineages followed over time, while the other measurements characterize changes in tissue contribution or cell state. An integrated model would need external validation against a model using clinical variables and mutation data alone. Relevant comparisons include calibration, sensitivity at a fixed specificity, and information gained beyond tumor burden and imaging. All serial samples from one patient must remain in the same training or validation group. Ultimately, treatment-guiding use would require prospective evidence that, acting on these measurements, benefits patients.

8. Conclusions

Inherited susceptibility is most clearly connected to somatic evolution at specific transitions. MMR loss and colibactin exposure change mutation production; normal-tissue studies identify mutation-dependent fixation or expansion; serial PDAC samples reveal repair-associated resistance. These findings do not establish a general ancestry-dependent trajectory, and direct evidence for an inherited effect on early clone establishment remains limited.

Progress requires studies that measure the proposed determinant and its somatic consequences together. Population comparisons define the distribution of risk and tumor features. Genomic lineages, repeated samples and functional experiments then test whether a determinant changes mutation acquisition, clone persistence or survival during treatment. Keeping those questions separate makes the connection between them testable.

Author Contributions: Conceptualization: M.B.P., I.V.K., I.D.K., P.A.S. and A.A.T.; writing—original draft preparation: M.B.P., I.V.K., K.R.K. and V.D.G.; supervision: V.V.S., P.V.S. and A.D.K.; writing—review and editing: all authors. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Russian Science Foundation, grant No. 25-24-20176.

Institutional Review Board Statement: Not applicable. This narrative review did not involve new research with human participants or animals.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article. The sources discussed in this review are cited in the reference list.

Acknowledgments: During preparation of this manuscript, the authors used GenAI OpenAI ChatGPT for English-language editing. The authors reviewed and edited the output and take full responsibility for the content of the publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ACRG	Asian Cancer Research Group
BTC	biliary tract cancer
CCA	cholangiocarcinoma
CRC	colorectal cancer
EWAS	epigenome-wide association study (also used in the literature for environment-wide association)
ExWAS	exome-wide association study
FAP	familial adenomatous polyposis
FDR	false discovery rate
GENIE	Genomics Evidence Neoplasia Information Exchange
GWAS	genome-wide association study
HBV	hepatitis B virus
HCC	hepatocellular carcinoma
HCV	hepatitis C virus
HLA	human leukocyte antigen
HR	homologous recombination
iCCA	intrahepatic cholangiocarcinoma
MMR	mismatch repair
MSI	microsatellite instability
PDAC	pancreatic ductal adenocarcinoma
PRS	polygenic risk score
QTL	quantitative trait locus
RNA-seq	RNA sequencing
RWAS	regulome-wide association study
TCGA	The Cancer Genome Atlas
TWAS	transcriptome-wide association study
WES	whole-exome sequencing
WGS	whole-genome sequencing
cfDNA	cell-free DNA
HRD	homologous recombination deficiency
PARP	poly(ADP-ribose) polymerase
PV	pathogenic variant
CCF	cancer cell fraction
ctDNA	circulating tumor DNA
UMI	unique molecular identifier
VAF	variant allele fraction
IHC	immunohistochemistry
NGS	next-generation sequencing
PCR	polymerase chain reaction

References

1. Khan, A.T.; Gogarten, S.M.; McHugh, C.P.; Stilp, A.M.; Sofer, T.; Bowers, M.L.; Wong, Q.; Cupples, L.A.; Hidalgo, B.; Johnson, A.D.; et al. Recommendations on the use and reporting of race, ethnicity, and ancestry in genetic research: Experiences from the NHLBI TOPMed program. *Cell Genom.* **2022**, *2*, 100155. [CrossRef] [PubMed]
2. Saoudi González, N.; Salvà, F.; Ros, J.; Baraibar, I.; Rodríguez-Castells, M.; García, A.; Alcaráz, A.; Vega, S.; Bueno, S.; Tabernero, J.; et al. Unravelling the Complexity of Colorectal Cancer: Heterogeneity, Clonal Evolution, and Clinical Implications. *Cancers* **2023**, *15*, 4020. [CrossRef] [PubMed]
3. Toghey, S.; Harvey-Jones, E.J.; Towler, J.D.; Hafkamp, C.J.H.; Chong, I.Y. Advances in Colorectal Cancer Cell Biology and Clonal Evolution. *Int. J. Mol. Sci.* **2026**, *27*, 953. [CrossRef] [PubMed]
4. Fan, W.; Zheng, C. Clonal evolution in gastrointestinal cancers: Multi-omics insights into tumor heterogeneity, microenvironmental selection, and translational biomarkers. *Front. Oncol.* **2026**, *16*, 1907210. [CrossRef] [PubMed]
5. John, J.; Bapat, A.; Gahlaut, S.; Luke, N.; Kumar, R.; Thakur, Y.; Mathew, C.; Konnur, A.; Namewar, N.; Reddy, R.; et al. Assessing germline mutational profile and its clinicopathological associations in Triple Negative Breast Cancer. *Cancer Genet.* **2025**, *296–297*, 65–75. [CrossRef] [PubMed]

6. Fernandez-Rozadilla, C.; Timofeeva, M.; Chen, Z.; Law, P.; Thomas, M.; Schmit, S.; Díez-Obrero, V.; Hsu, L.; Fernandez-Tajes, J.; Palles, C.; et al. Deciphering colorectal cancer genetics through multi-omic analysis of 100,204 cases and 154,587 controls of European and east Asian ancestries. *Nat. Genet.* **2023**, *55*, 89–99. [CrossRef] [PubMed]
7. Jiang, D.K.; Sun, J.; Cao, G.; Liu, Y.; Lin, D.; Gao, Y.Z.; Ren, W.-H.; Long, X.-D.; Zhang, H.; Ma, X.-P.; et al. Genetic variants in STAT4 and HLA-DQ genes confer risk of hepatitis B virus-related hepatocellular carcinoma. *Nat. Genet.* **2013**, *45*, 72–75. [CrossRef] [PubMed]
8. Han, Y.; Byun, J.; VanLith, C.J.; Cooley, M.A.; Shaw, V.R.; Hatia, R.I.; Ahmed, F.Y.; Giama, N.H.; Ali, H.M.; Campbell, N.A.; et al. Uncovering the genetic landscape of cholangiocarcinoma and its subtypes via GWAS and integrative analyses. *Hepatology* **2026**, *84*, 1335–1351. [CrossRef] [PubMed]
9. Yu, Y.; Chang, K.; Chen, J.S.; Bohlender, R.J.; Fowler, J.; Zhang, D.; Huang, M.; Chang, P.; Li, Y.; Wong, J.; et al. A whole-exome case-control association study to characterize the contribution of rare coding variation to pancreatic cancer risk. *HGG Adv.* **2022**, *3*, 100078. [CrossRef] [PubMed]
10. Siraj, A.K.; Bu, R.; Azam, S.; Qadri, Z.; Iqbal, K.; Parvathareddy, S.K.; Al-Dayel, F.; Al-Kuraya, K.S. Whole Exome-Wide Association Identifies Rare Variants in APC Associated with High-Risk Colorectal Cancer in the Middle East. *Cancers* **2024**, *16*, 3720. [CrossRef] [PubMed]
11. Hu, B.; Zhang, M.; Yu, Y.; Liu, Y.; Wang, J.; Zhang, A.; Li, Q.; Gao, Y.; Li, Z.; Bian, L.; et al. Integrative analysis of gastric tissue transcriptomes and gastric cancer GWAS implicates candidate susceptibility genes. *Am. J. Hum. Genet.* **2026**, *113*, 1521–1542. [CrossRef] [PubMed]
12. Tan, M.C.B.; Isom, C.A.; Liu, Y.; Trégouët, D.A.; Lindstrom, S.; Wang, L.; Smith, E.; Gordon, W.; Vlieg, A.V.H.; De Andrade, M.; et al. Transcriptome-wide association study and Mendelian randomization in pancreatic cancer identifies susceptibility genes and causal relationships with type 2 diabetes and venous thromboembolism. *EBioMedicine* **2024**, *106*, 105233. [CrossRef] [PubMed]
13. Liu, S.; Zhong, H.; Zhu, J.; Wu, Y.; Deng, Y.; Wu, L. Regulome-wide association study identifies genetically driven accessible regions associated with pancreatic cancer risk. *Int. J. Cancer* **2024**, *154*, 670–678. [CrossRef] [PubMed]
14. Kao, W.Y.; Yang, S.H.; Liu, W.J.; Yeh, M.Y.; Lin, C.L.; Liu, C.J.; Huang, C.; Lin, S.; Lee, S.; Chen, P.; et al. Genome-wide identification of blood DNA methylation patterns associated with early-onset hepatocellular carcinoma development in hepatitis B carriers. *Mol. Carcinog.* **2017**, *56*, 425–435. [CrossRef] [PubMed]
15. Jung, S.Y.; Yu, H.; Tan, X.; Pellegrini, M. Novel DNA methylation-based epigenetic signatures in colorectal cancer from peripheral blood leukocytes. *Am. J. Cancer Res.* **2024**, *14*, 2253–2271. [CrossRef] [PubMed]
16. Li, Z.; Yu, C.; Hao, J.; Shu, Y.; Zhang, J.; Zhao, K.; Pu, Q.; Liu, L. Identifying environmental factors and biological metrics associated with cancer prevalence and mortality: An environment-wide association study. *Cancer Epidemiol.* **2025**, *97*, 102828. [CrossRef] [PubMed]
17. Roth, A.; Khattra, J.; Yap, D.; Wan, A.; Laks, E.; Biele, J.; Ha, G.; Aparicio, S.; Bouchard-Côté, A.; Shah, S.P. PyClone: Statistical inference of clonal population structure in cancer. *Nat. Methods* **2014**, *11*, 396–398. [CrossRef] [PubMed]
18. Salcedo, A.; Tarabichi, M.; Buchanan, A.; Espiritu, S.M.G.; Zhang, H.; Zhu, K.; Yang, T.-H.O.; Leshchiner, I.; Anastassiou, D.; Guan, Y.; et al. Crowd-sourced benchmarking of single-sample tumor subclonal reconstruction. *Nat. Biotechnol.* **2025**, *43*, 581–592. [CrossRef] [PubMed]
19. Gerstung, M.; Jolly, C.; Leshchiner, I.; Dentro, S.C.; Gonzalez, S.; Rosebrock, D.; Mitchell, T.J.; Rubanova, Y.; Anur, P.; Yu, K.; et al. The evolutionary history of 2,658 cancers. *Nature* **2020**, *578*, 122–128. [CrossRef] [PubMed]
20. Williams, M.J.; Werner, B.; Heide, T.; Curtis, C.; Barnes, C.P.; Sottoriva, A.; Graham, T.A. Quantification of subclonal selection in cancer from bulk sequencing data. *Nat. Genet.* **2018**, *50*, 895–903. [CrossRef] [PubMed]
21. Grant, R.C.; Denroche, R.E.; Borgida, A.; Virtanen, C.; Cook, N.; Smith, A.L.; Connor, A.A.; Wilson, J.M.; Peterson, G.; Roberts, N.J.; et al. Exome-Wide Association Study of Pancreatic Cancer Risk. *Gastroenterology* **2018**, *154*, 719–722.e3. [CrossRef] [PubMed]
22. Matejcic, M.; Shaban, H.A.; Quintana, M.W.; Schumacher, F.R.; Edlund, C.K.; Naghi, L.; Pai, R.K.; Haile, R.W.; Levine, A.J.; Buchanan, D.D.; et al. Rare Variants in the DNA Repair Pathway and the Risk of Colorectal Cancer. *Cancer Epidemiol. Biomark. Prev.* **2021**, *30*, 895–903. [CrossRef] [PubMed]
23. Amuzu, S.; Xie, A.X.; Bai, X.; Pekala, K.R.; Pickersgill, N.A.; Ma, D.; Perea-Chamblee, T.; Arora, K.; Chatila, W.K.; Derkach, A.; et al. Meta-analysis reveals differences in somatic alterations by genetic ancestry across common cancers. *Nat. Genet.* **2025**, *57*, 2655–2660. [CrossRef] [PubMed]
24. Myer, P.A.; Lee, J.K.; Madison, R.W.; Pradhan, K.; Newberg, J.Y.; Isasi, C.R.; Klempner, S.J.; Frampton, G.M.; Ross, J.S.; Venstrom, J.M.; et al. The Genomics of Colorectal Cancer in Populations with African and European Ancestry. *Cancer Discov.* **2022**, *12*, 1282–1293. [CrossRef] [PubMed]
25. Rhead, B.; Hein, D.M.; Pouliot, Y.; Guinney, J.; De La Vega, F.M.; Sanford, N.N. Association of genetic ancestry with molecular tumor profiles in colorectal cancer. *Genome Med.* **2024**, *16*, 99. [CrossRef] [PubMed]

26. Matejcic, M.; Teer, J.K.; Hoehn, H.J.; Diaz, D.B.; Shankar, K.; Gong, J.; Nguyen, N.T.; Loroña, N.C.; Coppola, D.; Fulmer, C.G.; et al. Colorectal Tumors in Diverse Patient Populations Feature a Spectrum of Somatic Mutational Profiles. *Cancer Res.* **2025**, *85*, 1928–1944. [CrossRef] [PubMed]
27. Gheibi, M.; Sadeghi, S.; Hessami, A.; Erfani, Z.; Ghadirzadeh, E.; Moosazadeh, M. The prevalence of TP53, APC, and PIK3CA gene mutations in colorectal cancer patients: A Systematic Review and Meta-analysis. *Cancer Genet.* **2025**, *298–299*, 198–207. [CrossRef]
28. Carter, H.; Marty, R.; Hofree, M.; Gross, A.M.; Jensen, J.; Fisch, K.M.; Wu, X.; DeBoever, C.; Van Nostrand, E.L.; Song, Y.; et al. Interaction Landscape of Inherited Polymorphisms with Somatic Events in Cancer. *Cancer Discov.* **2017**, *7*, 410–423. [CrossRef] [PubMed]
29. Lee, B.C.H.; Robinson, P.S.; Coorens, T.H.H.; Yan, H.H.N.; Olafsson, S.; Lee-Six, H.; Sanders, M.A.; Siu, H.C.; Hewinson, J.; Yue, S.S.K.; et al. Mutational landscape of normal epithelial cells in Lynch Syndrome patients. *Nat. Commun.* **2022**, *13*, 2710. [CrossRef] [PubMed]
30. Martin, S.; Katainen, R.; Taira, A.; Välimäki, N.; Ristimäki, A.; Seppälä, T.; Renkonen-Sinisalo, L.; Lepistö, A.; Tahkola, K.; Mattila, A.; et al. Lynch syndrome-associated and sporadic microsatellite unstable colorectal cancers: Different patterns of clonal evolution yield highly similar tumours. *Hum. Mol. Genet.* **2024**, *33*, 1858–1872. [CrossRef] [PubMed]
31. College of American Pathologists. Mismatch Repair and Microsatellite Instability Testing for Immune Checkpoint Inhibitor Therapy. Guideline and Summary of Recommendations, 2022. Available online: https://www.cap.org/cap-guidelines/mismatch-repair-and-microsatellite-instability-testing-for-immune-checkpoint-inhibitor-therapy/ (accessed on 11 November 2026).
32. National Cancer Institute. Genetics of Colorectal Cancer (PDQ). Health Professional Version. Available online: https://www.cancer.gov/types/colorectal/hp/colorectal-genetics-pdq (accessed on 11 November 2026).
33. Rajoua, N.; Daunay, A.; Triki, W.; Baraket, O.; Bouchoucha, S.; Maghrebi, H.; Mabrouk, A.; Deleuze, J.-F.; How-Kit, A.; Kharrat, M. HSP110 T17 marker matches the pentaplex panel and outperforms CAT-25 for detecting microsatellite instability in sporadic colorectal cancer. *Cancer Genet.* **2025**, *294–295*, 21–26. [CrossRef] [PubMed]
34. Maas, S.C.E.; Baraibar, I.; Lemler, L.; Butjosa-Espín, M.; Blanco Irazuegui, O.; Tabernero, J.; Elez, E.; Seoane, J.A. Epigenetic fingerprints link early-onset colon and rectal cancer to pesticide exposure. *Nat. Med.* **2026**, *32*, 1827–1837. [CrossRef] [PubMed]
35. Pleguezuelos-Manzano, C.; Puschhof, J.; Rosendahl Huber, A.; van Hoeck, A.; Wood, H.M.; Nomburg, J.; Gurjao, C.; Manders, F.; Dalmasso, G.; Stege, P.B.; et al. Mutational signature in colorectal cancer caused by genotoxic pks^{+} *E. coli*. *Nature* **2020**, *580*, 269–273. [CrossRef] [PubMed]
36. Díaz-Gay, M.; Dos Santos, W.; Moody, S.; Kazachkova, M.; Abbasi, A.; Steele, C.D.; Vangara, R.; Senkin, S.; Wang, J.; Fitzgerald, S.; et al. Geographic and age variations in mutational processes in colorectal cancer. *Nature* **2025**, *643*, 230–240. [CrossRef] [PubMed]
37. Chang, J.; Tian, J.; Zhu, Y.; Zhong, R.; Zhai, K.; Li, J.; Ke, J.; Han, Q.; Lou, J.; Chen, W.; et al. Exome-wide analysis identifies three low-frequency missense variants associated with pancreatic cancer risk in Chinese populations. *Nat. Commun.* **2018**, *9*, 3688. [CrossRef] [PubMed]
38. Chang, J.; Tian, J.; Yang, Y.; Zhong, R.; Li, J.; Zhai, K.; Ke, J.; Lou, J.; Chen, W.; Zhu, B.; et al. A Rare Missense Variant in TCF7L2 Associates with Colorectal Cancer Risk by Interacting with a GWAS-Identified Regulatory Variant in the MYC Enhancer. *Cancer Res.* **2018**, *78*, 5164–5172. [CrossRef] [PubMed]
39. El Moheb, M.; Cummins, K.C.; Shen, C.; Bitar, E.R.; Kim, S.J.; Mayhew, M.M.; Zhang, H.; Witt, R.G.; Ruff, S.M.; Tsung, A. Comprehensive Analysis of Genomic Variations in Pancreatic Adenocarcinoma by Race. *Cancer Med.* **2025**, *14*, e71346. [CrossRef] [PubMed]
40. Van Egeren, D.; Schenck, R.O.; Khan, A.; Horning, A.M.; Mo, S.; Weiß, C.L.; Esplin, E.D.; Becker, W.R.; Wu, S.; Hanson, C.; et al. Polyclonal origins of human premalignant colorectal lesions. *Nature* **2026**, *650*, 1017–1024. [CrossRef] [PubMed]
41. Sottoriva, A.; Kang, H.; Ma, Z.; Graham, T.A.; Salomon, M.P.; Zhao, J.; Marjoram, P.; Siegmund, K.; Press, M.F.; Shibata, D.; et al. A Big Bang model of human colorectal tumor growth. *Nat. Genet.* **2015**, *47*, 209–216. [CrossRef] [PubMed]
42. Heiser, C.N.; Simmons, A.J.; Revetta, F.; McKinley, E.T.; Ramirez-Solano, M.A.; Wang, J.; Kaur, H.; Shao, J.; Ayers, G.D.; Wang, Y.; et al. Molecular cartography uncovers evolutionary and microenvironmental dynamics in sporadic colorectal tumors. *Cell* **2023**, *186*, 5620–5637.e16. [CrossRef] [PubMed]
43. Li, W.; Xu, Z.; Dong, X.; Yang, X.; Wang, G.; He, G. The KRAS p.Ala146Thr mutation promotes the proliferation of colorectal cancer cells via CDK1-mediated ERK signaling. *Cancer Genet.* **2025**, *296–297*, 154–162. [CrossRef] [PubMed]
44. Hirotsu, Y.; Hada, M.; Amemiya, K.; Oyama, T.; Mochizuki, H.; Omata, M. Multi-regional sequencing reveals clonal and polyclonal seeding from primary tumor to metastases in advanced gastric cancer. *J. Gastroenterol.* **2020**, *55*, 553–564. [CrossRef] [PubMed]
45. Yachida, S.; Jones, S.; Bozic, I.; Antal, T.; Leary, R.; Fu, B.; Kamiyama, M.; Hruban, R.H.; Eshleman, J.R.; Nowak, M.A.; et al. Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature* **2010**, *467*, 1114–1117. [CrossRef] [PubMed]
46. Makohon-Moore, A.P.; Zhang, M.; Reiter, J.G.; Bozic, I.; Allen, B.; Kundu, D.; Chatterjee, K.; Wong, F.; Jiao, Y.; Kohutek, Z.A.; et al. Limited heterogeneity of known driver gene mutations among the metastases of individual patients with pancreatic cancer. *Nat. Genet.* **2017**, *49*, 358–366. [CrossRef] [PubMed]

47. Reiter, J.G.; Makohon-Moore, A.P.; Gerold, J.M.; Heyde, A.; Attiyeh, M.A.; Kohutek, Z.A.; Tokheim, C.J.; Brown, A.; DeBlasio, R.M.; Niyazov, J.; et al. Minimal functional driver gene heterogeneity among untreated metastases. *Science* **2018**, *361*, 1033–1037. [CrossRef] [PubMed]
48. The Cancer Genome Atlas Research Network; Bass, A.J.; Thorsson, V.; Shmulevich, I.; Reynolds, S.M.; Miller, M.; Bernard, B.; Hinoue, T.; Laird, P.W.; Curtis, C.; et al. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* **2014**, *513*, 202–209. [CrossRef] [PubMed]
49. Cristescu, R.; Lee, J.; Nebozhyn, M.; Kim, K.M.; Ting, J.C.; Wong, S.S.; Liu, J.; Yue, Y.G.; Wang, J.; Yu, K.; et al. Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. *Nat. Med.* **2015**, *21*, 449–456. [CrossRef] [PubMed]
50. Abate, M.; Walch, H.; Arora, K.; Vanderbilt, C.M.; Fei, T.; Drebin, H.; Shimada, S.; Maio, A.B.; Kemel, Y.; Stadler, Z.K.; et al. Unique Genomic Alterations and Microbial Profiles Identified in Patients With Gastric Cancer of African, European, and Asian Ancestry: A Novel Path for Precision Oncology. *Ann. Surg.* **2023**, *278*, 506–518. [CrossRef] [PubMed]
51. Totoki, Y.; Saito-Adachi, M.; Shiraishi, Y.; Komura, D.; Nakamura, H.; Suzuki, A.; Tatsuno, K.; Rokutan, H.; Hama, N.; Yamamoto, S.; et al. Multiancestry genomic and transcriptomic analysis of gastric cancer. *Nat. Genet.* **2023**, *55*, 581–594. [CrossRef] [PubMed]
52. Gerdes, C.; Rengarajan, S.; Murugesan, K.; Ross, J.S.; Bartels, S.; Vogel, A.; Saborowski, A. Mutational patterns and ancestry-linked profiles in a large hepatocellular carcinoma and combined hepatocellular-cholangiocarcinoma cohort. *ESMO Open* **2026**, *11*, 106048. [CrossRef] [PubMed]
53. Tsilimigras, D.I.; Stecko, H.; Moris, D.; Pawlik, T.M. Genomic Profiling of Biliary Tract Cancers: Comprehensive Assessment of Anatomic and Geographic Heterogeneity, Co-Alterations and Outcomes. *J. Surg. Oncol.* **2025**, *131*, 1352–1361. [CrossRef] [PubMed]
54. Coorens, T.H.H.; Collord, G.; Jung, H.; Wang, Y.; Moore, L.; Hooks, Y.; Mahbubani, K.; Law, S.Y.K.; Yan, H.H.N.; Yuen, S.T.; et al. The somatic mutation landscape of normal gastric epithelium. *Nature* **2025**, *640*, 418–426. [CrossRef] [PubMed]
55. Skoufou-Papoutsaki, N.; Kemp, R.; Adler, S.; Marks, K.; Girard, A.C.; Mehmed, S.; Lourenço, F.C.; Moutin, E.B.; Hoopen, R.T.; Morrissey, E.; et al. Clonal biases dictate availability of colonic cancer driver mutations for transformation. *Nat. Commun.* **2026**, *17*, 5295. [CrossRef] [PubMed]
56. Lakatos, E.; Gunasri, V.; Zapata, L.; Househam, J.; Heide, T.; Trahearn, N.; Swinyard, O.; Cisneros, L.; Lynn, C.; Mossner, M.; et al. Epigenetically driven and early immune evasion in colorectal cancer evolution. *Nat. Genet.* **2025**, *57*, 3039–3049. [CrossRef] [PubMed]
57. Losic, B.; Craig, A.J.; Villacorta-Martin, C.; Martins-Filho, S.N.; Akers, N.; Chen, X.; Ahsen, M.E.; von Felden, J.; Labgaa, I.; D'avola, D.; et al. Intratumoral heterogeneity and clonal evolution in liver cancer. *Nat. Commun.* **2020**, *11*, 291. [CrossRef] [PubMed]
58. Chen, J.; Kaya, N.A.; Zhang, Y.; Kendarsari, R.I.; Sekar, K.; Lee Chong, S.; Seshachalam, V.P.; Ling, W.H.; Phua, C.Z.J.; Lai, H.; et al. A multimodal atlas of hepatocellular carcinoma reveals convergent evolutionary paths and 'bad apple' effect on clinical trajectory. *J. Hepatol.* **2024**, *81*, 667–678. [CrossRef] [PubMed]
59. Cordier, P.; Hirsch, T.Z.; Caruso, S.; Pan, L.; Galy-Fauroux, I.; Klein, C.; De La Cruz-Ojeda, P.; Maille, P.; Poupel, L.; Imbeaud, S.; et al. Intratumour ploidy heterogeneity and clonal evolution in hepatocellular carcinoma. *J. Hepatol.* **2026**, *84*, 776–792. [CrossRef] [PubMed]
60. Dong, L.Q.; Shi, Y.; Ma, L.J.; Yang, L.X.; Wang, X.Y.; Zhang, S.; Wang, Z.-C.; Duan, M.; Zhang, Z.; Liu, L.-Z.; et al. Spatial and temporal clonal evolution of intrahepatic cholangiocarcinoma. *J. Hepatol.* **2018**, *69*, 89–98. [CrossRef] [PubMed]
61. Zuyin, L.; Zhao, L.; Qian, C.; Changkun, Z.; Delin, M.; Jialing, H.; Zhuomiaoyu, C.; Yuzi, L.; Jiaxi, Z.; Jie, G.; et al. Single-Cell and Spatial Transcriptomics Delineate the Microstructure and Immune Landscape of Intrahepatic Cholangiocarcinoma in the Leading-Edge Area. *Adv. Sci.* **2025**, *12*, e2412740. [CrossRef] [PubMed]
62. Li, F.; Li, Y.; Wang, L.; Xu, L.; Xue, H.; Wei, W.; Xia, Y.; Wang, L.; Shen, F.; Wang, K. Tumor microenvironment heterogeneity and progression mechanisms in intrahepatic cholangiocarcinoma: A study based on single-cell and spatial transcriptomic sequencing. *Hepatology* **2026**, *83*, 1082–1097. [CrossRef] [PubMed]
63. Pei, G.; Min, J.; Rajapakshe, K.I.; Branchi, V.; Liu, Y.; Selvanesan, B.C.; Thege, F.; Sadeghian, D.; Zhang, D.; Cho, K.S.; et al. Spatial mapping of transcriptomic plasticity in metastatic pancreatic cancer. *Nature* **2025**, *642*, 212–221. [CrossRef] [PubMed]
64. Alver, T.N.; Bergholtz, H.; Holm, M.B.; Dorg, L.T.; Skrede, M.L.; Kure, E.H.; Verbeke, C.S. Spatial Transcriptomics Reveals Cancer and Stromal Cell Heterogeneity Between Center and Invasive Front of Pancreatic Cancer. *Mod. Pathol.* **2025**, *38*, 100726. [CrossRef] [PubMed]
65. Mo, C.K.; Liu, J.; Chen, S.; Storrs, E.; Targino da Costa, A.L.N.; Houston, A.; Wendl, M.C.; Jayasinghe, R.G.; Iglesia, M.D.; Ma, C.; et al. Tumour evolution and microenvironment interactions in 2D and 3D space. *Nature* **2024**, *634*, 1178–1186. [CrossRef] [PubMed]
66. Bailey, C.; Pich, O.; Thol, K.; Watkins, T.B.K.; Luebeck, J.; Rowan, A.; Stavrou, G.; Weiser, N.E.; Dameracharla, B.; Bentham, R.; et al. Origins and impact of extrachromosomal DNA. *Nature* **2024**, *635*, 193–200. [CrossRef] [PubMed]

67. Gao, P.; Zuo, C.; Yuan, W.; Cai, J.; Chai, X.; Gong, R.; Yu, J.; Yao, L.; Su, W.; Liu, Z.; et al. Spatiotemporal multi-omics analysis uncovers NAD-dependent immunosuppressive niche triggering early gastric cancer. *Signal Transduct. Target. Ther.* **2025**, *10*, 313. [CrossRef] [PubMed]
68. Golan, T.; Kindler, H.L.; Park, J.O.; Reni, M.; Macarulla, T.; Hammel, P.; Van Cutsem, E.; Arnold, D.; Hochhauser, D.; McGuinness, D.; et al. Geographic and Ethnic Heterogeneity of Germline BRCA1 or BRCA2 Mutation Prevalence Among Patients With Metastatic Pancreatic Cancer Screened for Entry Into the POLO Trial. *J. Clin. Oncol.* **2020**, *38*, 1442–1454. [CrossRef] [PubMed]
69. Paiella, S.; Azzolina, D.; Gregori, D.; Malleo, G.; Golan, T.; Simeone, D.; Davis, M.; Vacca, P.; Crovetto, A.; Bassi, C.; et al. A systematic review and meta-analysis of germline BRCA mutations in pancreatic cancer patients identifies global and racial disparities in access to genetic testing. *ESMO Open* **2023**, *8*, 100881. [CrossRef] [PubMed]
70. Park, W.; Chen, J.; Chou, J.F.; Varghese, A.M.; Yu, K.H.; Wong, W.; Capanu, M.; Balachandran, V.; McIntyre, C.A.; El Dika, I.; et al. Genomic Methods Identify Homologous Recombination Deficiency in Pancreas Adenocarcinoma and Optimize Treatment Selection. *Clin. Cancer Res.* **2020**, *26*, 3239–3247. [CrossRef] [PubMed]
71. Yadav, S.; Bao, R.; Graham, R.P.; Hu, C.; Hart, S.N.; Na, J.; Boddicker, N.; Gnanaolivu, R.D.; Smadbeck, J.; Ding, L.; et al. Genomic and Immune Landscape of Pancreatic Ductal Adenocarcinoma Associated with Germline Pathogenic Variants in ATM. *Clin. Cancer Res.* **2025**, *31*, 4463–4474. [CrossRef] [PubMed]
72. Lu, Z.; Chen, H.; Jiao, X.; Wang, Y.; Wu, L.; Sun, H.; Li, S.; Gong, J.; Li, J.; Zou, J.; et al. Germline HLA-B evolutionary divergence influences the efficacy of immune checkpoint blockade therapy in gastrointestinal cancer. *Genome Med.* **2021**, *13*, 175. [CrossRef] [PubMed]
73. Alencar, A.V.S.; Barreto, G.A.V.; Luciano, M.C.D.S.; Albuquerque, C.P.; de Sant'ANa, R.O.; Silva, P.G.d.B.; Bittencourt, F.d.S.; Lima, V.P.; Bezerra, M.J.B. Influence of lynch syndrome on the incidence of side effects to antineoplastic treatment of patients with colorectal cancer. *Cancer Genet.* **2025**, *298–299*, 169–179. [CrossRef] [PubMed]
74. Lee, W.H.; Kim, B.; Nam, C.H.; Yeo, H.Y.; Lee, D.W.; Park, S.C.; Oh, J.H.; Han, S.-S.; Lim, M.C.; Kim, H.; et al. Clonal evolution and mutational trajectories of metastatic colorectal cancer shaped by anticancer therapies. *Nat. Commun.* **2026**, *17*, 7509. [CrossRef] [PubMed]
75. Brown, T.J.; Yablonovitch, A.; Till, J.E.; Yen, J.; Kiedrowski, L.A.; Hood, R.; O'HAra, M.H.; Teitelbaum, U.; Karasic, T.B.; Schneider, C.; et al. The Clinical Implications of Reversions in Patients with Advanced Pancreatic Cancer and Pathogenic Variants in BRCA1, BRCA2, or PALB2 after Progression on Rucaparib. *Clin. Cancer Res.* **2023**, *29*, 5207–5216. [CrossRef] [PubMed]
76. Peng, H.; Jiang, L.; Yuan, J.; Wu, X.; Chen, N.; Liu, D.; Liang, Y.; Xie, Y.; Jia, K.; Li, Y.; et al. Single-cell characterization of differentiation trajectories and drug resistance features in gastric cancer with peritoneal metastasis. *Clin. Transl. Med.* **2024**, *14*, e70054. [CrossRef] [PubMed]
77. Revsine, M.; Wang, L.; Forgues, M.; Behrens, S.; Craig, A.J.; Liu, M.; Tran, B.; Kelly, M.; Budhu, A.; Monge, C.; et al. Lineage and ecology define liver tumor evolution in response to treatment. *Cell Rep. Med.* **2024**, *5*, 101394. [CrossRef] [PubMed]
78. Newman, A.M.; Lovejoy, A.F.; Klass, D.M.; Kurtz, D.M.; Chabon, J.J.; Scherer, F.; Stehr, H.; Liu, C.L.; Bratman, S.V.; Say, C.; et al. Integrated digital error suppression for improved detec Integrated digital error suppression for improved detection of circulating tumor DNA. *Nat. Biotechnol.* **2016**, *34*, 547–555. [CrossRef] [PubMed]
79. Dong, W.; Lau, C.H.; Li, J.; Huang, Z.; Li, J.; Wu, W.; Chen, X.; Huang, Y.; Huang, X.; Xu, M.; et al. Pan-cancer methylation analysis of circulating cell-free DNA. *Cancer Genet.* **2025**, *296–297*, 182–195. [CrossRef] [PubMed]
80. Liu, H. Mechanistic fragmentomics. *Cancer Genet.* **2026**, *306–307*, 237–239. [CrossRef] [PubMed]
81. Han, D.S.C.; Ni, M.; Chan, R.W.Y.; Chan, V.W.; Lui, K.O.; Chiu, R.W.; Lo, Y.D. The Biology of Cell-free DNA Fragmentation and the Roles of DNASE1, DNASE1L3, and DFFB. *Am. J. Hum. Genet.* **2020**, *106*, 202–214. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.