J Biochem Technol (2026) 17(4): 30-40
https://doi.org/10.51847/v2hFoOjdoQ
ISSN: 0974-2328

Connecting Noncoding Cancer Variants to Effector Genes: A Scoping Review of Functional Perturbation Strategies, Regulatory Context, and Evidence Supporting Causal Gene Assignment

Mariana Silva*, Rafael Duarte, Camila Rocha, Paulo Ricardo, Ana Santos, Lucas Ferreira

Received: 19 August 2026 / Received in revised form: 19 September 2026, Accepted: 22 September 2026, Published online: 12 November 2026

Abstract

Noncoding cancer-associated variants are frequently located in regulatory regions whose biological consequences cannot be inferred reliably from genomic position or statistical association alone. Establishing an effector gene requires evidence that a candidate variant alters regulatory function in an appropriate cellular context and that this alteration can be connected to a specific gene or downstream phenotype. This scoping review mapped experimental and computational strategies used between 2017 and 2025 to resolve that problem, with particular emphasis on massively parallel reporter assays, CRISPR-based regulatory perturbation, endogenous nucleotide editing, chromatin-interaction mapping, single-cell perturbation profiling, and integrative multi-omic approaches. Forty-eight unique records entered screening, 39 proceeded to eligibility assessment, and 32 evidence units were included. The evidence landscape showed that individual technologies interrogate different parts of the causal chain. Reporter assays efficiently detect allele-dependent regulatory activity but usually detach variants from endogenous chromatin; chromatin maps nominate physically plausible targets without establishing regulatory necessity; CRISPR-based element perturbation tests endogenous regulatory dependency but may not resolve the causal nucleotide; and base or prime editing provides more direct allele-specific evidence while remaining dependent on cellular state and assay context. Across the mapped literature, confidence in effector-gene assignment increased when fine mapping, regulatory-element activity, endogenous perturbation, target-gene response, and tissue-relevant chromatin evidence converged. The review therefore supports a conditional, evidence-layered interpretation of noncoding cancer variants in which causal gene assignment is strongest when alternative linked variants, neighboring genes, cell states, and regulatory routes have been experimentally discriminated.

Keywords: Noncoding cancer variants, Effector genes, Regulatory genomics, CRISPR perturbation, Enhancer-gene mapping, Functional genomics

Introduction

Noncoding variation constitutes a substantial part of the genomic landscape in human cancer, yet the biological interpretation of such variation remains considerably more difficult than the interpretation of protein-altering coding mutations. Somatic and inherited noncoding variants can affect promoters, enhancers, untranslated regions, transcription-factor binding sites, chromatin architecture, RNA processing, or other regulatory mechanisms, but recurrent occurrence or proximity to a gene does not by itself establish driver function (Elliott & Larsson, 2021). Genome-wide analyses of somatic noncoding mutation patterns further demonstrate that candidate regulatory events must be distinguished from background mutational processes and other sources of regional mutation enrichment before biological meaning is assigned (Dietlein *et al.*, 2022).

The central interpretive problem is therefore not simply whether a noncoding locus is associated with cancer, but how evidence should be connected across three distinct questions: whether a particular variant is functionally active, which regulatory element or process it alters, and which gene mediates the downstream consequence. Contemporary approaches to noncoding-variant prioritization increasingly emphasize experimental validation because statistical association, epigenomic annotation, and predictive modeling can nominate plausible mechanisms without directly demonstrating them (Chin *et al.*, 2024). Large-scale

Mariana Silva*
Department of Genetics, University of Campinas, Campinas, Brazil.

Rafael Duarte
Department of Molecular Biology, University of São Paulo, São Paulo, Brazil.

Camila Rocha
Department of Biochemistry, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil.

Paulo Ricardo
Department of Functional Genomics, University of São Paulo, São Paulo, Brazil.

Ana Santos
Department of Molecular Genetics, University of Campinas, Campinas, Brazil.

Lucas Ferreira
Department of Biochemistry, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil.

***E-mail:** mariana.silva@unicamp.br



© 2026 The Author(s). This is an **Open Access** article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0).
https://creativecommons.org/licenses/by/4.0/deed.en

functional analyses of cancer-associated germline variants illustrate how high-throughput assays can narrow broad sets of candidate alleles to variants with measurable regulatory consequences (Kellman *et al.*, 2025). In colorectal cancer, systematic integration of fine mapping and functional evidence has similarly been used to prioritize both candidate causal variants and effector genes at susceptibility loci (Law *et al.*, 2024). Lung-cancer susceptibility studies have shown that regulatory effects can be heterogeneous across loci and experimental contexts, limiting the assumption that a statistically prioritized allele will behave uniformly across cellular models (Long *et al.*, 2024).

Enhancer-to-gene mapping adds another layer to this problem because the closest gene to a risk variant is not necessarily the regulatory target. Genome-wide cancer-oriented enhancer-gene maps can link candidate variants to genes using regulatory activity, chromatin relationships, and genomic context (Ying *et al.*, 2023). Multi-ancestry fine mapping of breast-cancer risk regions can further reduce credible variant sets and reveal biological heterogeneity created by linkage disequilibrium and population structure (Jia *et al.*, 2025). Functional annotation studies of breast-cancer susceptibility loci reinforce the need to integrate association data with experimentally informative regulatory evidence rather than treating positional annotation as definitive target assignment (Romualdo Cardoso *et al.*, 2022). More broadly, enhancer maps developed across human cellular systems demonstrate that distal regulatory elements can connect disease-associated variants to genes located beyond the nearest transcriptional unit (Nasser *et al.*, 2021).

Functional perturbation has consequently become central to attempts to move from candidate regulation to causal inference. The activity-by-contact model combined enhancer activity with chromatin contact information and was calibrated against thousands of CRISPR perturbations, demonstrating how predictive regulatory maps can be anchored to experimental evidence (Fulco *et al.*, 2019). Pooled CRISPR screens have extended this principle by directly perturbing regulatory elements while measuring gene-expression consequences at scale (Gasperini *et al.*, 2019). CRISPR-based epigenome editing provides a complementary strategy in which regulatory activity can be altered without deleting the underlying DNA sequence, allowing element-level function to be tested in endogenous chromatin (Klann *et al.*, 2017). Massively parallel reporter technologies, including high-throughput assays capable of evaluating millions of candidate regulatory single-nucleotide variants, provide exceptional throughput for detecting allele-dependent regulatory activity (van Arensbergen *et al.*, 2019). Their principal interpretive limitation is that reporter activity can be measured outside the native chromosomal environment in which chromatin accessibility, three-dimensional contacts, transcriptional competition, and cellular state influence the true regulatory effect. Sequence-based predictive models add still another layer by learning relationships between genomic sequence and transcriptional output, but model prediction remains mechanistically different from direct perturbational evidence (Linder *et al.*, 2025).

These distinctions motivate a scoping review rather than a single-effect systematic synthesis. The relevant literature spans cancer genetics, regulatory genomics, chromatin biology, functional screening, single-cell genomics, and computational gene prioritization, and the evidence units do not estimate a common intervention effect. The purpose of this review is therefore to map how noncoding cancer variants have been connected to effector genes, identify what each functional strategy can and cannot establish, and clarify the conditions under which multiple forms of evidence collectively support a stronger causal assignment.

Materials and Methods

Scoping Review Design, Information Sources, Eligibility, and Evidence Charting

The review was designed and reported according to principles of the PRISMA Extension for Scoping Reviews, with transparent separation of evidence identification, screening, eligibility assessment, charting, and synthesis (Tricco *et al.*, 2018). Methodological decisions were also aligned with updated JBI guidance for scoping reviews, particularly the definition of the review question, evidence-unit eligibility, data charting, and presentation of mapped evidence (Peters *et al.*, 2020). Because the objective was to characterize evidence architectures rather than estimate a pooled intervention effect, no meta-analysis was planned.

The eligibility framework was organized around the concept of variant-to-effector-gene assignment in human cancer. Primary cancer-focused studies were eligible when they investigated noncoding variants, regulatory elements, or cancer-risk loci and generated functional, perturbational, reporter-based, chromatin-interaction, expression-linked, or integrative evidence relevant to assignment of one or more effector genes. Human-cell functional-genomics studies outside a single cancer type were also eligible when they provided directly transferable experimental methods for testing regulatory-element-to-gene or variant-to-gene relationships. Studies centered exclusively on protein-coding mutations, gene dependency without a noncoding-regulatory question, or non-cancer regulatory biology without sufficient transfer value were excluded from the evidence synthesis.

Searches were executed on 22 October 2026 with the evidence window restricted to publications from 1 January 2017 through 31 December 2025. PubMed-indexed records formed the principal literature-discovery source, supplemented by publisher metadata and backward and forward citation chaining. Search concepts combined cancer or neoplasm terms with noncoding variation, enhancer/promoter/UTR terminology, functional perturbation terms, and target- or effector-gene terminology. The search families incorporated CRISPR, CRISPR interference or activation, massively parallel reporter assays, base editing, prime editing, Perturb-seq, enhancer-gene mapping, chromatin-interaction assays, expression quantitative trait loci, and activity-by-contact approaches. Records were de-duplicated during retrieval before formal screening.

The unit of inclusion was an eligible evidence report or study, not an individual variant, gene, locus, assay, or bibliography entry. Reviews, reporting guidance, and broad reference atlases could support interpretation and methodology in the article but were not

counted as included evidence units. This distinction prevented the fixed bibliography size from being used as a surrogate for the number of included studies.

Evidence charting captured study design, cancer or biological context, variant or regulatory element, perturbation type, measurement modality, target-gene evidence, alternative explanations, contextual limitations, and the degree of directness between the manipulated genomic feature and the proposed effector gene. Evidence directness was used descriptively rather than as a numerical risk-of-bias score. Endogenous perturbation with gene-level or phenotypic readout and orthogonal corroboration represented the strongest direct evidence class; regulatory assays or mapping studies with an incomplete causal chain were treated as intermediate evidence; transferable computational or methodological benchmarks were distinguished from cancer-specific validation. Null, discordant, tissue-dependent, and assay-dependent results were retained rather than collapsed into a single direction of effect.

The operational distinctions used during screening and charting are summarized in **Table 1**.

Table 1. Operational eligibility and evidence-charting structure for the scoping review

Review component	Operational rule	Cancer-specific implementation	Evidence recorded	Principal exclusion safeguard
Biological concept	Noncoding variation with potential regulatory consequences	Germline or somatic variants, enhancers, promoters, UTRs, eRNAs, distal regulatory loci	Variant/locus, regulatory element, putative gene	Coding-only mechanisms excluded
Evidence unit	One eligible primary report or study	Cancer-focused studies plus transferable human functional-genomics benchmarks	Study design and biological system	Bibliography entries not counted as studies
Functional requirement	Evidence must inform variant-, element-, or locus-to-gene interpretation	Reporter activity, CRISPR perturbation, endogenous editing, chromatin mapping, expression response, multi-omics	Perturbation, readout, target-gene evidence	Annotation alone insufficient for inclusion
Context requirement	Human cancer relevance or direct methodological transferability	Tumor type, relevant cell state, primary cancer tissue, or transferable human-cell model	Tissue/cell context and transfer boundary	Unrelated non-cancer biology excluded
Search window	Publication between 2017 and 2025	Applied uniformly to evidence units	Publication year	Earlier or later evidence excluded from synthesis
Information sources	PubMed-indexed literature, publisher metadata, citation chaining	Cancer and functional-genomics terminology combined	Search family and retrieval path	No inferred database counts
Directness charting	Separate association, regulatory activity, element dependency, allele-specific effect, and effector-gene response	Applied to every included report	Evidence-directness category	No automatic conversion of association into causality
Alternative explanations	Record competing variants, genes, regulatory routes, assay artifacts, and context effects	Retained when reported or biologically material	Competing explanation and boundary condition	Discordant evidence not suppressed
Appraisal approach	Descriptive evidence-directness appraisal	No pooled risk-of-bias score or meta-analysis	Directness and contextual limitations	Scoping design not converted into effectiveness review

The resulting selection pathway preserved the distinction between contextual sources retained for article framing and primary evidence units entering the scoping synthesis **(Figure 1)**.

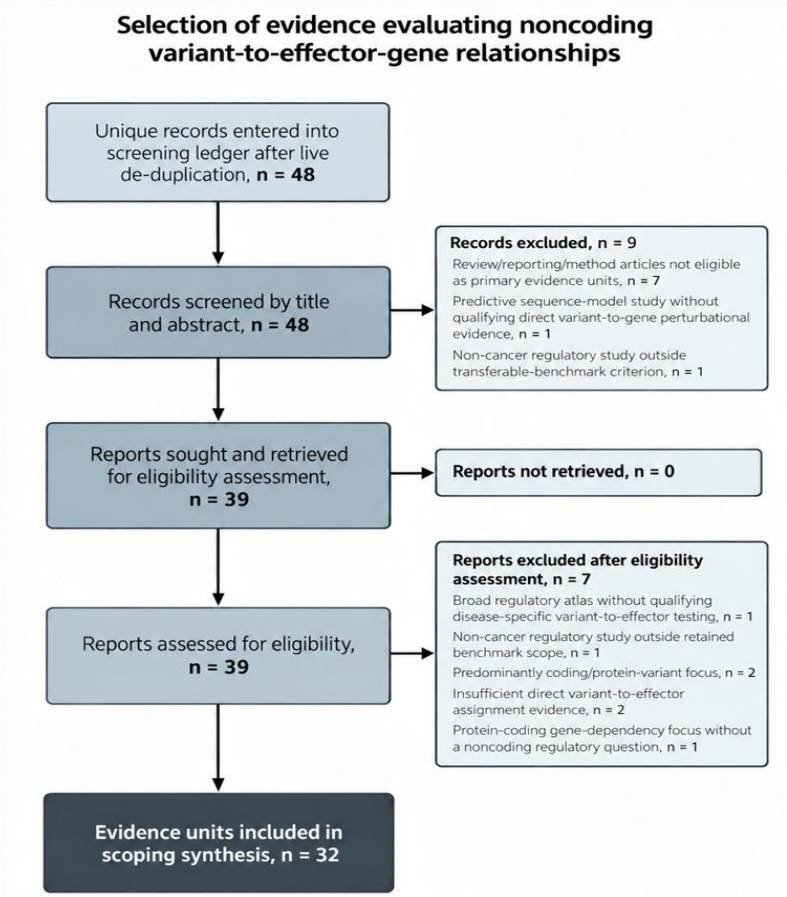


Figure 1. Selection of evidence evaluating noncoding variant-to-effector-gene relationships

Results and Discussion

Evidence Identification and Distribution of Included Studies

The reconciled screening process yielded 32 included evidence units from 48 unique records. Twenty included studies directly addressed cancer-associated noncoding variants, cancer-risk loci, cancer regulatory elements, or tumor-specific regulatory architecture. Twelve additional studies were retained as transferable human functional-genomics benchmarks because they directly tested technologies or analytical relationships required for variant-to-gene interpretation, including enhancer perturbation, endogenous editing, single-cell perturbation, or regulatory-element-to-gene mapping. This distinction was maintained throughout charting so that method-development studies did not acquire disease-specific evidentiary weight merely because their techniques were applicable to cancer.

The cancer-focused evidence covered several distinct disease contexts rather than a single regulatory system. Melanoma studies used massively parallel reporter assays and complementary validation to identify functional susceptibility variants and candidate target genes (Choi *et al.*, 2020). A subsequent melanoma analysis emphasized that apparent variant effects can change across cellular contexts, making cell-type specificity a substantive component of causal interpretation rather than an experimental nuisance (Long *et al.*, 2022). Chromatin-interaction mapping at melanoma susceptibility loci further demonstrated that regulatory targets may lie at substantial genomic distance from the associated variant, while simultaneously illustrating why physical contact requires functional corroboration before it can be treated as causal regulation (Thakur *et al.*, 2025). In breast cancer, locus-specific functional analysis at 2q35 implicated structural variation in long-range enhancer activity (Baxter *et al.*, 2021), whereas CRISPR-based screening across breast-cancer risk loci provided a more direct strategy for identifying genes whose expression depends on regulatory perturbation (Tuano *et al.*, 2023).

Across the complete evidence set, five review-defined primary evidence modes were identified. These categories describe the principal analytical role of each included study and are not intended as mutually exclusive descriptions of every assay used within a paper. Reporter and massively parallel allelic assays provided high-throughput evidence of regulatory activity; CRISPR-based regulatory and gene perturbation studies tested endogenous dependencies; precision editing approaches examined nucleotide-specific effects in situ; chromatin and regulatory mapping studies established plausible physical or statistical links; and integrative/fine-mapping studies combined multiple evidence layers around candidate loci. The distribution is summarized in **Table 2**.

Table 2. Distribution of the 32 included evidence units by primary functional-evidence mode

Primary evidence mode	Included evidence units, n	Typical experimental or analytical unit	Principal contribution to variant-to-gene interpretation	Main limitation retained in synthesis
Reporter and massively parallel allelic assays	5	Individual noncoding allele or regulatory sequence	Detects allele-dependent regulatory activity at high throughput	Often separates sequence from endogenous chromatin and native gene context
CRISPR-based regulatory or gene perturbation, including single-cell perturbation	9	Regulatory element, candidate enhancer, or perturbed gene	Tests whether endogenous regulatory activity is necessary or sufficient for a gene-level response	Element perturbation may not identify the causal nucleotide; downstream effects may be indirect
Precision endogenous nucleotide editing	2	Defined nucleotide substitution	Tests allele-specific effects within endogenous sequence context	Editing efficiency, by-products, and cellular state can affect interpretation
Chromatin, enhancer-gene, eRNA, or three-dimensional regulatory mapping	6	Regulatory element–gene relationship or chromatin contact	Narrows physically or statistically plausible effector genes	Contact or association alone does not establish regulatory necessity
Integrative fine mapping and locus-focused functional studies	10	Cancer-risk locus, candidate variant set, or regulatory region	Combines genetic, epigenomic, functional, and locus-specific evidence	Strength depends on concordance among evidence layers and completeness of alternative-gene testing
Total	32	—	—	—

The mapped literature therefore did not converge on a single preferred experimental strategy. Instead, the studies interrogated different parts of a causal chain extending from statistical localization of a variant, through regulatory activity and chromatin context, to gene-level response and downstream cellular consequence. Reporter assays were particularly efficient for narrowing large variant sets, whereas endogenous perturbation studies offered greater directness for testing regulatory dependence. Chromatin-contact and enhancer-gene maps expanded the set of plausible target genes beyond the nearest-gene assumption, but their inferential strength depended on whether corresponding perturbational evidence was available. Precision editing and single-cell perturbation approaches addressed narrower parts of the evidence set but provided a route toward testing allelic effects within native sequence or cell-state contexts.

The evidence distribution also revealed a recurring mismatch between technological scale and causal completeness. Genome-scale screening can test thousands of variants or regulatory elements, yet high throughput often comes at the cost of endogenous context or detailed mechanistic validation. Conversely, locus-focused editing and perturbation can produce more direct evidence while remaining limited in breadth, disease representation, or cell-state diversity. These complementary strengths and weaknesses establish the basis for the subsequent comparison of functional perturbation strategies and for evaluating how regulatory context changes the strength of an effector-gene assignment.

Functional Perturbation Strategies for Noncoding Variant Interpretation

Functional perturbation methods differ most importantly in what part of the proposed causal chain they intervene on. Some assays test whether a sequence has regulatory potential, whereas others manipulate an endogenous element, replace a nucleotide, or perturb a candidate effector gene and observe downstream transcriptional consequences. Pooled single-cell CRISPR screening has expanded this landscape by connecting regulatory perturbations with transcriptome-wide responses, thereby allowing candidate GWAS loci to be evaluated against multiple genes and pathways simultaneously (Morris *et al.*, 2023). Such screens increase mechanistic resolution, but a transcriptional response following perturbation may still be indirect and therefore does not automatically establish that the manipulated sequence and the responding gene form a direct regulatory pair.

Endogenous sequence editing provides a more allele-specific test. High-throughput prime-editing strategies can install defined nucleotide changes within their chromosomal context and compare the resulting cellular phenotype with the unedited state (Ren *et al.*, 2023). This approach addresses a major limitation of reporter assays because the tested variant remains embedded within local chromatin and native regulatory sequence. Interpretation nonetheless depends on editing efficiency, the distribution of editing products, cellular fitness effects, and the relevance of the experimental model to the cancer state under investigation. Base-editing screens confront similar issues. Joint modeling of genotype and phenotype can improve estimation of variant effects where editing creates heterogeneous cellular outcomes, but the resulting inference remains dependent on the completeness of genotype measurement and the assumptions of the analytical model (Ryu *et al.*, 2024).

Not all functionally important noncoding variation operates through classical distal enhancers. Massively parallel interrogation of 3′ untranslated-region variants has demonstrated that rare noncoding changes can modify the abundance of transcripts encoded by cancer-relevant genes (Fu *et al.*, 2024). These findings broaden the variant-to-gene problem beyond enhancer-promoter

communication because the altered gene is already defined by transcript structure, while the functional mechanism may involve RNA stability, processing, translation, or binding of regulatory molecules. Cancer-associated alterations in ultraconserved noncoding elements provide another distinct regulatory setting in which sequence conservation and recurrent mutation can motivate functional investigation, although conservation itself does not identify the affected gene or demonstrate driver status (Bayraktar *et al.*, 2025).

The native three-dimensional genome places an additional constraint on perturbational interpretation. Maps of primary human cancers show that tumor-associated regulatory architecture can organize candidate enhancers and promoters in configurations that differ across cancer types and states (Yost *et al.*, 2025). Consequently, a variant that is active in a reporter construct may fail to influence the predicted gene if the relevant chromatin interaction is absent in the disease-relevant cell. Conversely, regulatory relationships can be missed by assays performed in a state in which the required contact or transcriptional program is inactive. Pan-cancer analyses of enhancer RNAs illustrate a complementary strategy for detecting genetically associated regulatory activity, but transcriptome-wide association signals remain subject to colocalization, linkage and pleiotropic explanations and therefore require experimental validation before being interpreted as variant-to-effector causality (Chen, Wang, *et al.*, 2025).

Perturbation of larger regulatory units can identify dependencies that are invisible to variant-by-variant screening. CRISPR screening of transcribed super-enhancers in triple-negative breast cancer has been used to identify regulatory regions associated with disease progression and downstream gene programs (Lewis *et al.*, 2024). At another mechanistic scale, selective enhancer gain-of-function can deregulate an established oncogenic effector such as MYC, demonstrating that regulatory activation itself can produce a biologically consequential gene-expression change (Rahmat *et al.*, 2024). These studies provide strong evidence for enhancer dependency but do not necessarily determine which naturally occurring nucleotide variant initiates that dependency.

Integrative studies therefore increasingly combine perturbation with genetic and regulatory mapping. Multi-omic analyses of ovarian-cancer risk regions illustrate how fine mapping, regulatory annotation, expression-related evidence and functional characterization can narrow candidate variants and genes within the same susceptibility interval (Dareng *et al.*, 2024). In glioma, enhancer-connectome mapping paired with functional evidence provides a route from inherited risk variation through candidate regulatory elements toward specific gene targets (Bi *et al.*, 2025). Massively parallel variant-to-function mapping in non-small-cell lung cancer similarly demonstrates how large candidate-variant sets can be reduced using experimental measurements of regulatory activity before target-gene hypotheses are pursued (Chen, Li, *et al.*, 2025).

Single-cell perturbation technologies add another level of discrimination because they measure heterogeneous cellular responses rather than only bulk gene-expression averages. Genome-scale Perturb-seq has demonstrated that genetic interventions can generate information-rich genotype-phenotype landscapes across many cellular states (Replogle *et al.*, 2022). Targeted Perturb-seq increases the efficiency of this approach by concentrating sequencing depth on selected transcriptional readouts while retaining large perturbation libraries (Schraivogel *et al.*, 2020). For noncoding cancer variants, these approaches are most informative when the perturbation corresponds to the regulatory hypothesis being tested; perturbing an entire gene can clarify downstream consequences but should not be treated as equivalent to perturbing a risk nucleotide or enhancer.

The functional distinctions underlying these technologies are summarized in **Table 3**.

Table 3. Experimental reach of major perturbation strategies used to interpret noncoding cancer variants

Perturbation strategy	Primary manipulated unit	Typical direct readout	Native chromatin retained?	Variant-specific inference	Effector-gene inference	Principal unresolved issue
Massively parallel reporter assay	Candidate allele or short regulatory sequence	Reporter activity	Usually no	Strong for sequence-dependent regulatory activity	Usually indirect unless combined with target-gene evidence	Episomal or ectopic regulatory context
CRISPRi/CRISPRa regulatory-element perturbation	Endogenous enhancer/promoter region	Target or transcriptomic response	Yes	Usually element-level rather than nucleotide-level	Stronger when a specific gene responds reproducibly	Element perturbation may exceed the effect of a natural allele
CRISPR deletion	Endogenous regulatory interval	Gene expression or phenotype	Yes, but local sequence is removed	Low-to-moderate for individual nucleotide inference	Potentially strong for element dependency	Deletion can disrupt spacing, chromatin or multiple motifs simultaneously
Prime editing	Defined nucleotide	Allele-specific molecular or cellular response	Yes	High	High when gene response is measured directly	Editing efficiency and cellular-model dependence

Base editing	Editable nucleotide class	Allele/genotype -associated phenotype	Yes	High within editing constraints	High when coupled to gene-level readout	Mixed editing outcomes and by-products
Perturb-seq	Gene or regulatory element	Single-cell transcriptomic response	Yes	Depends on perturbation design	High for response-network mapping	Direct versus secondary transcriptional effects
Chromatin-linked perturbation	Regulatory element within mapped contact architecture	Gene expression plus contact/context information	Yes	Moderate unless nucleotide is edited	Strong when contact and perturbation converge	Contact may be permissive rather than causal
Super-enhancer perturbation	Large regulatory domain	Gene-expression or growth dependency	Yes	Low for individual variant assignment	Strong for regulatory dependency	Difficult to localize causality to a specific variant

Taken together, the included perturbational literature supports a graded interpretation of functional evidence rather than a universal assay hierarchy. Reporter assays are efficient for candidate reduction, endogenous element perturbation tests regulatory necessity, precision editing approaches allele-specific causality, and single-cell measurements expose context-dependent downstream effects. No one method resolves every step simultaneously.

This layered relationship between a noncoding variant, its regulatory environment, perturbational evidence and the proposed effector gene is represented in **Figure 2**.

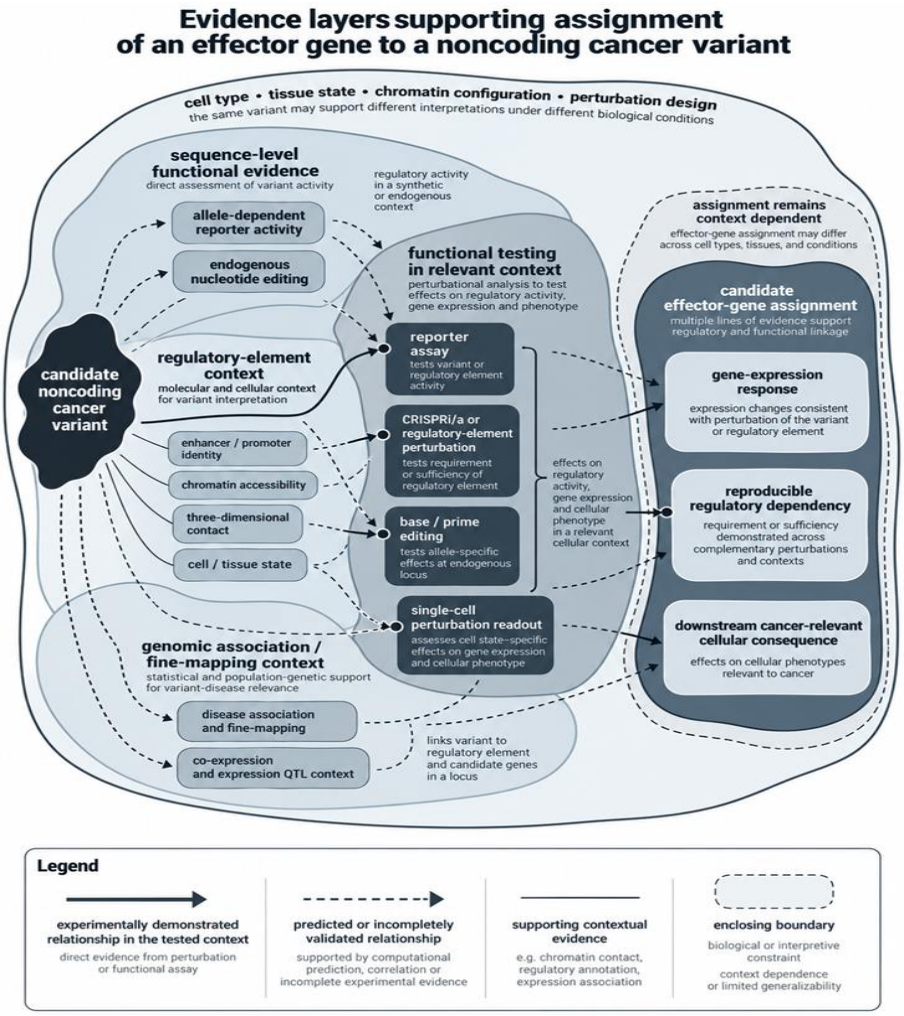


Figure 2. Evidence layers supporting assignment of an effector gene to a noncoding cancer variant

Regulatory Context and the Strength of Variant-to-Effector Gene Assignment

The strength of an effector-gene assignment depends on the extent to which the experimental system reproduces the biological context in which the variant acts. Regulatory effects can change during cell-state transitions, and CRISPRi screening has shown that enhancer-gene relationships may be constrained by chromatin loops and may behave nonlinearly across cellular states (Luo *et al.*, 2023). A negative perturbation result in one state can therefore reflect absence of the required regulatory configuration rather than absence of biological function.

Reference epigenomic atlases provide a broad framework for locating candidate regulatory elements, transcription-factor occupancy, chromatin accessibility and other functional annotations (ENCODE Project Consortium *et al.*, 2020). Such information is valuable for prioritization, but an annotation describes biochemical or chromatin features and should not be converted directly into evidence that a cancer-associated allele causes a specific gene-expression change. Similarly, physical enhancer-promoter contact narrows the plausible target set but cannot distinguish productive regulation from proximity unless gene expression changes when the regulatory hypothesis is perturbed.

The strongest assignments in the mapped literature consequently arise through **c**onvergence across mechanistically different evidence types. Fine mapping narrows the candidate allele set; reporter testing establishes that allele sequence can alter regulatory activity; endogenous CRISPR perturbation determines whether the regulatory element influences a gene; precision editing tests whether the implicated nucleotide produces the effect in situ; and chromatin or cell-state measurements establish whether the regulatory relationship is biologically plausible in the relevant tumor context. Failure of one layer need not invalidate all others, but discordance changes the claim that can safely be made.

Table 4 formalizes the evidence levels used to interpret this convergence.

Table 4. Evidence patterns governing the strength of noncoding variant-to-effector-gene assignment

Evidence pattern	What is directly supported	Remaining alternative explanation	Strength of permissible assignment	Additional evidence that would most reduce uncertainty
Statistical association or fine mapping only	Variant or locus is associated with cancer susceptibility	Linked variant, pleiotropy, population structure	Candidate locus/variant	Functional allele testing
Regulatory annotation only	Variant overlaps a biochemically annotated element	Annotation may not be active in the relevant cancer state	Candidate regulatory mechanism	Context-specific functional assay
Allele-dependent reporter activity	Variant sequence can alter regulatory output in the assay	Episomal context, absent native chromatin, uncertain target gene	Functional regulatory variant candidate	Endogenous perturbation plus gene readout
Chromatin contact with candidate gene	Variant-containing element is physically compatible with distal regulation	Nonproductive contact or multiple contacted genes	Plausible target gene	Element perturbation and target-gene response
Endogenous regulatory-element perturbation	Element contributes to regulation of a measured gene	Natural variant may not reproduce full perturbation effect	Strong element-to-gene support	Allele-specific endogenous editing
Endogenous nucleotide editing with concordant gene response	Specific allele affects gene-level output in native sequence context	Model-specific or cell-state-specific effect	Strong variant-to-gene assignment	Replication in independent disease-relevant context
Convergent allele editing, regulatory context and downstream phenotype	Variant, regulatory element and effector-gene response form a coherent experimental chain	Residual context transfer and alternative downstream pathways	Highest support represented in this review	Independent replication and disease-relevant validation

This hierarchy is not a numerical scoring system. Its purpose is to prevent evidence obtained at one biological scale from being interpreted as proof at another. A chromatin contact is evidence about spatial organization; reporter activity is evidence about sequence-dependent regulatory potential; a CRISPR perturbation is evidence about endogenous regulatory dependence; and allele editing can provide direct evidence about the nucleotide itself. Causal confidence increases when these observations converge on the same effector gene under a disease-relevant biological state.

Evidence Gaps, Causal Attribution, and Research Priorities

The scoping map shows that the field has moved well beyond nearest-gene annotation, yet complete experimental traversal from variant to regulatory element to effector gene to cancer-relevant phenotype remains uncommon. Contemporary association-to-gene methods integrate expression quantitative trait loci, regulatory annotations, enhancer-gene maps and network information, but different approaches can prioritize different genes because they encode different assumptions about regulation (Qi *et al.*, 2024). Functional testing is therefore most valuable when it is designed to

discriminate among competing target genes rather than simply confirm the initially favored interpretation.

A second gap concerns biological transfer. Many high-throughput technologies are optimized in tractable cell lines, whereas the regulatory state of a primary tumor may reflect lineage, differentiation, microenvironmental signals, acquired genomic alterations and therapy exposure. Negative results should consequently be interpreted against assay sensitivity and regulatory-state compatibility. Positive results also require qualification when the perturbation is much stronger than the naturally occurring allele.

Clinical interpretation of noncoding variation provides an instructive external benchmark. Systematic evaluation of promoter and untranslated-region variants in large rare-disease cohorts demonstrates that noncoding pathogenicity can be identified when genomic evidence, regulatory context and functional interpretation are assembled explicitly (Martin-Geary *et al.*, 2025). Cancer differs because somatic evolution, tissue heterogeneity and susceptibility architecture introduce additional causal layers, but the same principle applies: a noncoding finding becomes clinically or mechanistically persuasive only when the route from sequence change to gene consequence is testable.

Three research priorities follow from the mapped evidence. First, large-scale variant screens should increasingly incorporate endogenous sequence perturbation where technically feasible, because this reduces uncertainty created by reporter context. Second, candidate target genes should be tested in the same cellular state in which regulatory activity and chromatin relationships are measured. Third, studies should report negative and discordant evidence explicitly, including variants with reporter activity that fail endogenous validation, chromatin contacts without measurable regulatory consequences, and perturbations whose effects change across cellular states. These discordances are informative because they identify where apparently plausible causal chains break.

The review also has limitations. The literature is methodologically heterogeneous, and the scoping design maps evidence architectures rather than estimating a pooled effect. Transferable functional-genomics benchmarks were included when they directly informed variant-to-gene inference, but they were kept distinct from cancer-specific evidence. The search window ended in 2025, and technological advances after that date were outside eligibility. Finally, evidence directness was charted descriptively rather than converted into a numerical quality score; such a score would imply a level of comparability that the diverse assay systems do not support.

Conclusion

Connecting a noncoding cancer variant to an effector gene requires more than genomic association, regulatory annotation, or physical proximity. The mapped evidence supports a layered causal interpretation in which each experimental strategy answers a different question. Reporter assays establish sequence-dependent regulatory potential; chromatin and enhancer maps define plausible regulatory context; CRISPR-based perturbation tests endogenous element dependency; precision editing approaches the causal nucleotide directly; and single-cell measurements expose context-dependent gene responses.

The strongest assignments emerge when these evidence types converge in a disease-relevant cellular state and when plausible alternative variants, regulatory elements and target genes have been actively tested. Where such convergence is absent, the appropriate conclusion is a calibrated candidate relationship rather than a definitive causal gene assignment. Future progress will therefore depend less on generating larger lists of regulatory variants than on designing experiments that connect sequence, regulatory context, effector-gene response and cancer-relevant phenotype within the same interpretable causal chain.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

References

Baxter, J. S., Johnson, N., Tomczyk, K., Gillespie, A., Maguire, S., Brough, R., Fachal, L., Michailidou, K., Bolla, M. K., Wang, Q., et al. (2021). Functional annotation of the 2q35 breast cancer risk locus implicates a structural variant in influencing activity of a long-range enhancer element. *The American Journal of Human Genetics, 108*(7), 1190-1203. doi:10.1016/j.ajhg.2021.05.013

Bayraktar, R., Tang, Y., Dragomir, M. P., Ivan, C., Peng, X., Fabris, L., Zhang, J., Carugo, A., Aneli, S., Liu, J., et al. (2025). The mutational landscape and functional effects of noncoding ultraconserved elements in human cancers. *Science Advances, 11*(8), eado2830. doi:10.1126/sciadv.ado2830

Bi, J., Mo, W., Liu, M., Song, Y., Xiao, Q., Fan, S., Wang, W., Shi, T., Zheng, Y., Lian, J., et al. (2025). Systematic decoding of functional enhancer connectomes and risk variants in human glioma. *Nature Cell Biology, 27*(10), 1838-1847. doi:10.1038/s41556-025-01737-3

Chen, C., Li, Y., Gu, Y., Zhai, Q., Guo, S., Xiang, J., Xie, Y., An, M., Li, C., Qin, N., et al. (2025). Massively parallel variant-to-function mapping determines functional regulatory variants of non-small cell lung cancer. *Nature Communications, 16*(1), 1391. doi:10.1038/s41467-025-56725-w

Chen, W., Wang, Z., Wang, Y., Lin, J., Chen, S., Chen, H., Ma, X., Zou, X., Li, X., Qin, Y., et al. (2025). Enhancer RNA transcriptome-wide association study reveals a distinctive class of pan-cancer susceptibility eRNAs. *Advanced Science, 12*(13), e2411974. doi:10.1002/advs.202411974

Chin, I. M., Gardell, Z. A., & Corces, M. R. (2024). Decoding polygenic diseases: Advances in noncoding variant prioritization and validation. *Trends in Cell Biology, 34*(6), 465-483. doi:10.1016/j.tcb.2024.03.005

Choi, J., Zhang, T., Vu, A., Ablain, J., Makowski, M. M., Colli, L. M., Xu, M., Hennessey, R. C., Yin, J., Rothschild, H., et al.

(2020). Massively parallel reporter assays of melanoma risk variants identify MX2 as a gene promoting melanoma. *Nature Communications, 11*(1), 2718. doi:10.1038/s41467-020-16590-1

Dareng, E. O., Coetzee, S. G., Tyrer, J. P., Peng, P.-C., Rosenow, W., Chen, S., Davis, B. D., Dezem, F. S., Seo, J.-H., Nameki, R., et al. (2024). Integrative multi-omics analyses to identify the genetic and functional mechanisms underlying ovarian cancer risk regions. *The American Journal of Human Genetics, 111*(6), 1061-1083. doi:10.1016/j.ajhg.2024.04.011

Dietlein, F., Wang, A. B., Fagre, C., Tang, A., Besselink, N. J. M., Cuppen, E., Li, C., Sunyaev, S. R., Neal, J. T., & Van Allen, E. M. (2022). Genome-wide analysis of somatic noncoding mutation patterns in cancer. *Science, 376*(6589), eabg5601. doi:10.1126/science.abg5601

Elliott, K., & Larsson, E. (2021). Non-coding driver mutations in human cancer. *Nature Reviews Cancer, 21*(8), 500-509. doi:10.1038/s41568-021-00371-z

ENCODE Project Consortium, Moore, J. E., Purcaro, M. J., Pratt, H. E., Epstein, C. B., Shoresh, N., Adrian, J., Kawli, T., Davis, C. A., Dobin, A., et al. (2020). Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature, 583*(7818), 699-710. doi:10.1038/s41586-020-2493-4

Fu, T., Amoah, K., Chan, T. W., Bahn, J. H., Lee, J.-H., Terrazas, S., Chong, R., Kosuri, S., & Xiao, X. (2024). Massively parallel screen uncovers many rare 3' UTR variants regulating mRNA abundance of cancer driver genes. *Nature Communications, 15*(1), 3335. doi:10.1038/s41467-024-46795-7

Fulco, C. P., Nasser, J., Jones, T. R., Munson, G., Bergman, D. T., Subramanian, V., Grossman, S. R., Anyoha, R., Doughty, B. R., Patwardhan, T. A., et al. (2019). Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nature Genetics, 51*(12), 1664-1669. doi:10.1038/s41588-019-0538-0

Gasperini, M., Hill, A. J., McFaline-Figueroa, J. L., Martin, B., Kim, S., Zhang, M. D., Jackson, D., Leith, A., Schreiber, J., Noble, W. S., et al. (2019). A genome-wide framework for mapping gene regulation via cellular genetic screens. *Cell, 176*(1-2), 377-390.e19. doi:10.1016/j.cell.2018.11.029

Jia, G., Chen, Z., Ping, J., Cai, Q., Tao, R., Li, C., Bauer, J. A., Xie, Y., Ambs, S., Barnard, M. E., et al. (2025). Refining breast cancer genetic risk and biology through multi-ancestry fine-mapping analyses of 192 risk regions. *Nature Genetics, 57*(1), 80-87. doi:10.1038/s41588-024-02031-y

Kellman, L. N., Neela, P. H., Srinivasan, S., Siprashvili, Z., Shanderson, R. L., Hong, A. W., Rao, D., Porter, D. F., Reynolds, D. L., Meyers, R. M., et al. (2025). Functional analysis of cancer-associated germline risk variants. *Nature Genetics, 57*(3), 718-728. doi:10.1038/s41588-024-02070-5

Klann, T. S., Black, J. B., Chellappan, M., Safi, A., Song, L., Hilton, I. B., Crawford, G. E., Reddy, T. E., & Gersbach, C. A. (2017). CRISPR-Cas9 epigenome editing enables high-throughput screening for functional regulatory elements in the human genome. *Nature Biotechnology, 35*(6), 561-568. doi:10.1038/nbt.3853

Law, P. J., Studd, J., Smith, J., Vijayakrishnan, J., Harris, B. T., Mandelia, M., Mills, C., Dunlop, M. G., & Houlston, R. S. (2024). Systematic prioritization of functional variants and effector genes underlying colorectal cancer risk. *Nature Genetics, 56*(10), 2104-2111. doi:10.1038/s41588-024-01900-w

Lewis, M. W., King, C. M., Wisniewska, K., Regner, M. J., Coffey, A., Kelly, M. R., Mendez-Giraldez, R., Davis, E. S., Phanstiel, D. H., & Franco, H. L. (2024). CRISPR screening of transcribed super-enhancers identifies drivers of triple-negative breast cancer progression. *Cancer Research, 84*(21), 3684-3700. doi:10.1158/0008-5472.CAN-23-3995

Linder, J., Srivastava, D., Yuan, H., Agarwal, V., & Kelley, D. R. (2025). Predicting RNA-seq coverage from DNA sequence as a unifying model of gene regulation. *Nature Genetics, 57*(4), 949-961. doi:10.1038/s41588-024-02053-6

Long, E., Patel, H., Golden, A., Antony, M., Yin, J., Funderburk, K., Feng, J., Song, L., Hoskins, J. W., Amundadottir, L. T., et al. (2024). High-throughput characterization of functional variants highlights heterogeneity and polygenicity underlying lung cancer susceptibility. *The American Journal of Human Genetics, 111*(7), 1405-1419. doi:10.1016/j.ajhg.2024.05.021

Long, E., Yin, J., Funderburk, K. M., Xu, M., Feng, J., Kane, A., Zhang, T., Myers, T., Golden, A., Thakur, R., et al. (2022). Massively parallel reporter assays and variant scoring identified functional variants and target genes for melanoma loci and highlighted cell-type specificity. *The American Journal of Human Genetics, 109*(12), 2210-2229. doi:10.1016/j.ajhg.2022.11.006

Luo, R., Yan, J., Oh, J. W., Xi, W., Shigaki, D., Wong, W., Cho, H. S., Murphy, D., Cutler, R., Rosen, B. P., et al. (2023). Dynamic network-guided CRISPRi screen identifies CTCF-loop-constrained nonlinear enhancer gene regulatory activity during cell state transitions. *Nature Genetics, 55*(8), 1336-1346. doi:10.1038/s41588-023-01450-7

Martin-Geary, A. C., Blakes, A. J. M., Dawes, R., Findlay, S. D., Lord, J., Dong, S., Walker, S., Talbot-Martin, J., Wieder, N., D'Souza, E. N., et al. (2025). Systematic identification of disease-causing promoter and untranslated region variants in 8040 undiagnosed individuals with rare disease. *Genome Medicine, 17*(1), 40. doi:10.1186/s13073-025-01464-2

Morris, J. A., Caragine, C., Daniloski, Z., Domingo, J., Barry, T., Lu, L., Davis, K., Ziosi, M., Glinos, D. A., Hao, S., et al. (2023). Discovery of target genes and pathways at GWAS loci by pooled single-cell CRISPR screens. *Science, 380*(6646), eadh7699. doi:10.1126/science.adh7699

Nasser, J., Bergman, D. T., Fulco, C. P., Guckelberger, P., Doughty, B. R., Patwardhan, T. A., Jones, T. R., Nguyen, T. H., Ulirsch, J. C., Lekschas, F., et al. (2021). Genome-wide enhancer maps link risk variants to disease genes. *Nature, 593*(7858), 238-243. doi:10.1038/s41586-021-03446-x

Peters, M. D. J., Marnie, C., Tricco, A. C., Pollock, D., Munn, Z., Alexander, L., McInerney, P., Godfrey, C. M., & Khalil, H. (2020). Updated methodological guidance for the conduct of scoping reviews. *JBI Evidence Synthesis, 18*(10), 2119-2126. doi:10.11124/JBIES-20-00167

Qi, T., Song, L., Guo, Y., Chen, C., & Yang, J. (2024). From

genetic associations to genes: Methods, applications, and challenges. *Trends in Genetics*, *40*(8), 642-667. doi:10.1016/j.tig.2024.04.008

Rahmat, M., Clement, K., Alberge, J.-B., Sklavenitis-Pistofidis, R., Kodgule, R., Fulco, C. P., Heilpern-Mallory, D., Nilsson, K., Dorfman, D., Engreitz, J. M., et al. (2024). Selective enhancer gain-of-function deregulates MYC expression in multiple myeloma. *Cancer Research*, *84*(24), 4173-4183. doi:10.1158/0008-5472.CAN-24-1440

Ren, X., Yang, H., Nierenberg, J. L., Sun, Y., Chen, J., Beaman, C., Pham, T., Nobuhara, M., Takagi, M. A., Narayan, V., et al. (2023). High-throughput PRIME-editing screens identify functional DNA variants in the human genome. *Molecular Cell*, *83*(24), 4633-4645.e9. doi:10.1016/j.molcel.2023.11.021

Replogle, J. M., Saunders, R. A., Pogson, A. N., Hussmann, J. A., Lenail, A., Guna, A., Mascibroda, L., Wagner, E. J., Adelman, K., Lithwick-Yanai, G., et al. (2022). Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq. *Cell*, *185*(14), 2559-2575.e28. doi:10.1016/j.cell.2022.05.013

Romualdo Cardoso, S., Gillespie, A., Haider, S., & Fletcher, O. (2022). Functional annotation of breast cancer risk loci: Current progress and future directions. *British Journal of Cancer*, *126*(7), 981-993. doi:10.1038/s41416-021-01612-6

Ryu, J., Barkal, S., Yu, T., Jankowiak, M., Zhou, Y., Francoeur, M., Phan, Q. V., Li, Z., Tognon, M., Brown, L., et al. (2024). Joint genotypic and phenotypic outcome modeling improves base editing variant effect quantification. *Nature Genetics*, *56*(5), 925-937. doi:10.1038/s41588-024-01726-6

Schraivogel, D., Gschwind, A. R., Milbank, J. H., Leonce, D. R., Jakob, P., Mathur, L., Korbel, J. O., Merten, C. A., Velten, L., & Steinmetz, L. M. (2020). Targeted Perturb-seq enables genome-scale genetic screens in single cells. *Nature Methods*, *17*(6), 629-635. doi:10.1038/s41592-020-0837-5

Thakur, R., Xu, M., Sowards, H., Yon, J., Jessop, L., Myers, T., Zhang, T., Chari, R., Long, E., Rehling, T., et al. (2025). Mapping chromatin interactions at melanoma susceptibility loci uncovers distant cis-regulatory gene targets. *The American Journal of Human Genetics*, *112*(7), 1625-1648. doi:10.1016/j.ajhg.2025.04.015

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., et al. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine*, *169*(7), 467-473. doi:10.7326/M18-0850

Tuano, N. K., Beesley, J., Manning, M., Shi, W., Perlaza-Jimenez, L., Malaver-Ortega, L. F., Paynter, J. M., Black, D., Civitarese, A., McCue, K., et al. (2023). CRISPR screens identify gene targets at breast cancer risk loci. *Genome Biology*, *24*(1), 59. doi:10.1186/s13059-023-02898-w

van Arensbergen, J., Pagie, L., FitzPatrick, V. D., de Haas, M., Baltissen, M. P., Comoglio, F., van der Weide, R. H., Teunissen, H., Võsa, U., Franke, L., et al. (2019). High-throughput identification of human SNPs affecting regulatory element activity. *Nature Genetics*, *51*(7), 1160-1169. doi:10.1038/s41588-019-0455-2

Ying, P., Chen, C., Lu, Z., Chen, S., Zhang, M., Cai, Y., Zhang, F., Huang, J., Fan, L., Ning, C., et al. (2023). Genome-wide enhancer-gene regulatory maps link causal variants to target genes underlying human cancer risk. *Nature Communications*, *14*(1), 5958. doi:10.1038/s41467-023-41690-z

Yost, K. E., Zhao, Y., Hung, K. L., Zhu, K., Xu, D., Corces, M. R., Shams, S., Louie, B. H., Sarmashghi, S., Sundaram, L., et al. (2025). Three-dimensional genome landscape of primary human cancers. *Nature Genetics*, *57*(5), 1189-1200. doi:10.1038/s41588-025-02188-0