J Biochem Technol (2026) 17(3): 65-72
https://doi.org/10.51847/jgB3TO66tm
ISSN: 0974-2328

From Sequence Variation to Conditional Dysfunction: A Conceptual Model of Gene Dosage, Paralog Compensation, and Cellular Context in the Interpretation of Molecular Genetic Effects

David Chen*, Sarah Thompson, Mark Johnson

Received: 18 July 2026 / Received in revised form: 21 August 2026, Accepted: 21 August 2026, Published online: 14 October 2026

Abstract

Sequence variation is often treated as the initiating unit of molecular genetic interpretation, yet the biological consequences of a variant emerge through processes that extend beyond nucleotide or amino-acid identity. Functional effects can depend on how strongly a locus tolerates perturbation, whether a change modifies protein abundance or biochemical activity, the effective dosage of the affected gene product, the availability of functionally overlapping paralogs, and the cellular conditions under which compensation is possible. This conceptual article develops an evidence-bounded account of conditional molecular dysfunction by integrating recent work in variant-effect mapping, dosage sensitivity, paralog interactions, transcriptional adaptation, and context-dependent genetic dependency. The central proposition is that sequence-level perturbation and realized dysfunction should be treated as related but separable biological states. A variant may impose an intrinsic molecular disturbance while its expressed consequence is amplified, buffered, redirected, or rendered experimentally invisible by gene dosage, compensatory capacity, cellular state, and the measurement system used to observe it. This account distinguishes constitutive functional redundancy from induced compensation and separates observed assay phenotypes from broader mechanistic interpretation. It also generates falsifiable expectations: context dependence should disappear when the proposed modifiers are experimentally neutralized, while purported compensation should fail to explain phenotypic rescue when the relevant paralog or adaptive response is absent. The resulting model supports more disciplined interpretation of genetic effects without assuming that every variant is context sensitive or that compensation is universally available.

David Chen*
Department of Biochemistry, University of Toronto, Toronto, Canada.

Sarah Thompson
Department of Molecular Genetics, University of Toronto, Toronto, Canada.

Mark Johnson
Department of Cell Biology, McGill University, Montreal, Canada.

***E-mail:** david.chen@utoronto.ca

Keywords: Variant effect, Gene dosage, Paralog compensation, Genetic robustness, Functional genomics, Cellular context

Introduction

Interpreting sequence variation requires more than identifying whether a nucleotide or amino-acid substitution differs from a reference sequence. Contemporary functional-genomics programs increasingly frame variant interpretation as a problem of connecting genomic perturbation to molecular function, cellular state, and phenotype, while community benchmarking has shown that predictive performance changes across variant classes, biological endpoints, and evaluation settings (Critical Assessment of Genome Interpretation Consortium [CAGI], 2024; IGVF Consortium, 2024). This widening of the interpretive problem matters because the same sequence-level description can coexist with very different biological consequences. A substitution may alter folding, abundance, catalytic activity, regulatory occupancy, transcript processing, or none of these to a degree sufficient to produce cellular dysfunction.

Large computational models demonstrate how much biological information is already encoded in sequence and evolutionary constraint. Protein language models can predict many missense effects using statistical regularities learned from protein sequences (Brandes *et al*., 2023), while AlphaMissense combines structural and evolutionary information to estimate the pathogenic potential of amino-acid substitutions across the proteome (Cheng *et al*., 2023). Such methods are valuable because they transform sequence information into experimentally testable hypotheses at scales that direct laboratory characterization cannot presently match. Their outputs, however, remain inferential. They do not directly establish which molecular mechanism is operating in a specific cellular environment, nor whether the measured or predicted perturbation exceeds the buffering capacity of that environment.

The same distinction is visible in experimental functional genomics. Updated comparisons of variant-effect predictors against deep mutational scanning data show that estimates of predictor quality depend on benchmark composition and evaluation design (Livesey & Marsh, 2023). Multiplexed assays of variant effect similarly require explicit reporting of assay



© 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

architecture, molecular readout, calibration and uncertainty if results are to be interpreted or compared across systems (Claussnitzer *et al*., 2024). In base-editing screens, modelling editing efficiency, possible editing outcomes and chromatin accessibility improves estimation of variant effects, illustrating that the observed phenotype partly reflects the perturbation and measurement process through which it is produced (Ryu *et al*., 2024). Context therefore enters interpretation both biologically and technically.

This article develops a conceptual account of that distinction. Its focus is not a universal assertion that sequence information is weak, nor a claim that every genetic effect requires extensive contextual correction. Some variants can impose severe molecular disruption across many environments. The narrower question is how interpretation should proceed when intrinsic variant effects, effective gene dosage, paralogous functions, cellular state and assay design plausibly contribute differentially to the observed phenotype. The resulting argument separates the perturbation encoded by sequence variation from the biological conditions that determine whether that perturbation becomes detectable molecular dysfunction. In this view, conditionality is a property to be demonstrated mechanistically rather than assumed by default.

Materials and Methods

Why Sequence Variation Alone Does Not Determine Molecular Dysfunction

Sequence-level information acquires biological meaning within systems that differ greatly in tolerance to perturbation. Population genetic constraint provides one demonstration. Analysis of variation in 141,456 humans revealed large differences among genes in their tolerance of predicted loss-of-function and missense variation, indicating that the biological consequences of sequence disruption cannot be inferred from variant class alone (Karczewski *et al*., 2020). Constraint is not itself an individual-level functional measurement, but it provides evidence that genetic systems differ in how strongly variation is filtered by selection. A substitution arising in a highly constrained locus therefore enters a different prior biological landscape from an apparently similar substitution in a more tolerant locus.

Direct functional assays expose additional heterogeneity. Saturation genome editing of BRCA1 showed that thousands of individual sequence substitutions could be assigned functional scores in an endogenous cellular setting and that these measurements corresponded substantially with established clinical classifications (Findlay *et al*., 2018). Yet molecular dysfunction need not be expressed through the same readout in every protein. VAMP-seq demonstrated that many variants in PTEN and TPMT reduce intracellular protein abundance, identifying destabilization or altered proteostasis as one mechanism through which coding variation can act (Matreyek *et al*., 2018). A variant that preserves abundance may still disrupt biochemical function, whereas a variant that lowers abundance may produce an effect before any change in catalytic properties is considered.

The distinction becomes especially clear when multiple functional dimensions are measured for the same protein. A comprehensive map of glucokinase activity showed that variants can produce hypoactive or hyperactive states through different structural and dynamic mechanisms (Gersing *et al*., 2023; Gersing *et al*., 2024). Subsequent measurement of glucokinase abundance demonstrated that decreased abundance accounted for only part of the loss-of-function landscape; many hypoactive variants retained near-normal abundance. These findings make it difficult to treat "variant effect" as a single molecular quantity. Even within one protein, sequence variation can alter stability, conformational equilibrium and catalytic function in partly independent ways.

The same principle extends from protein properties to the stages of gene expression. Integrated multiplexed assays of COMT variants identified effects at the levels of RNA abundance, ribosome loading and protein abundance, with changes at one layer not inevitably propagated to the next (Hoskins *et al*., 2024). Computational work likewise indicates that evolutionary conservation cannot always distinguish residues whose importance arises from structural stability from those whose importance arises from a more specific functional role (Cagiada *et al*., 2023). DeMAG further illustrates that missense interpretation can benefit from epistatic structural and evolutionary relationships between residues rather than treating each substituted position as biologically independent (Luppino *et al*., 2023). Together, these studies support a layered interpretation in which the molecular meaning of a sequence change depends on what biological property is measured, which interactions are preserved, and which assay conditions expose the perturbation.

A useful consequence of this evidence is to separate several questions that are often collapsed into one. Does a sequence change perturb a molecular property? Which property is altered? How much functional reserve remains after that alteration? Does the affected cell depend strongly on the perturbed activity? Finally, does the assay observe the relevant biological dimension with sufficient fidelity? **Table 1** organizes these distinct routes without assigning an unsupported universal hierarchy among them.

Table 1. Mechanistic routes through which sequence variation can acquire different functional consequences

Biological unit being interrogated	Variant-sensitive property	Observable experimental signature	Context that modifies interpretation	Inference that is justified	Inference that remains unsupported
Gene under population constraint	Tolerance to disruptive variation	Depletion or relative tolerance of variant classes	Selection, coverage and gene-specific architecture	Loci differ in population-level intolerance to variation	Constraint alone identifies an individual molecular mechanism

Endogenous gene function	Cellular consequence of defined substitutions	Variant-specific functional score	Cell model, editing outcome and endogenous regulation	Individual substitutions can differ sharply in measured function	The assay captures all tissue-specific consequences
Protein abundance	Folding, stability or proteostatic handling	Reduced or preserved intracellular abundance	Protein quality control and cellular environment	Destabilization is one route from coding change to dysfunction	Preserved abundance guarantees normal function
Enzymatic activity	Catalysis and conformational regulation	Hypoactive, near-reference or hyperactive functional state	Substrate environment and assay system	Activity can vary independently of simple variant class	Activity score alone establishes organismal disease severity
Expression cascade	RNA, translation and protein-level propagation	Concordant or discordant effects across molecular layers	Regulatory architecture and cell state	Effects at different expression layers can diverge	A disturbance at one layer must propagate unchanged
Residue interaction network	Stability, functional-site participation and epistatic constraint	Context-sensitive computational or experimental effect	Structural contacts and evolutionary dependencies	Residue context contributes information beyond substitution identity	A predicted residue interaction proves a causal disease mechanism
Measurement system	Accessibility and fidelity of the chosen readout	Assay-dependent effect estimate	Editing efficiency, reporter design and model assumptions	Measured effect depends partly on observation conditions	Every assay discrepancy represents biological compensation

These routes also clarify why an apparent disagreement between two variant-effect measurements should not automatically be treated as error. A protein-abundance assay and an activity assay may be interrogating different causal components of the same substitution. Conversely, concordance between sequence-based prediction and a cellular functional score does not establish that intermediate mechanisms are irrelevant. The interpretive task is to identify which component of the perturbation is being observed and how strongly that component is coupled to the phenotype of interest.

This distinction can be compressed into a multilevel representation that keeps intrinsic variant effects separate from the processes that determine whether those effects become expressed dysfunction. **Figure 1** places variant-level evidence on one side of that problem and effective dosage and compensatory capacity on the other, while maintaining explicit boundaries between measured observations, inferential links and context-dependent relations.

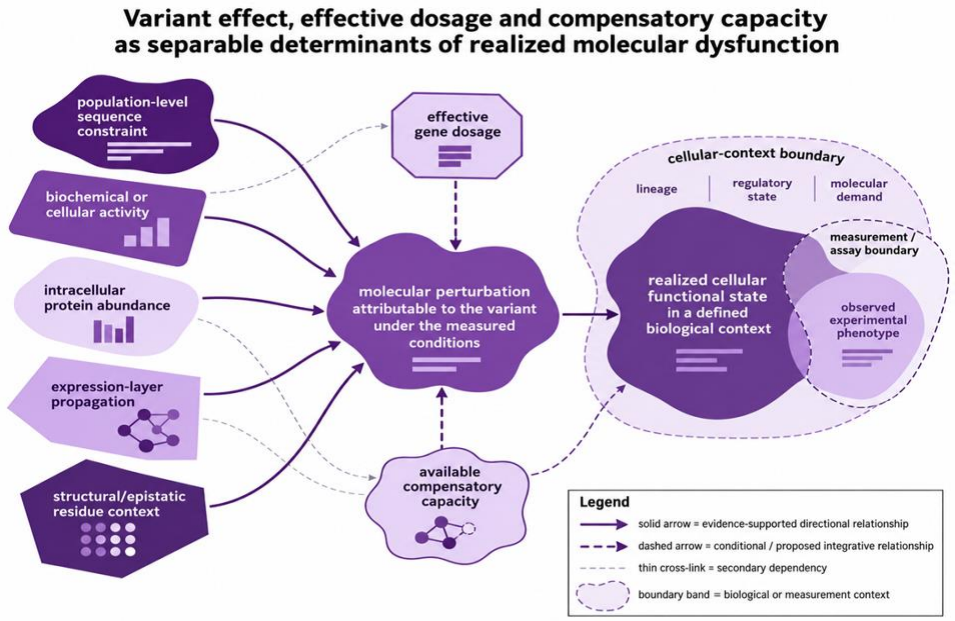


Figure 1. Variant effect, effective dosage and compensatory capacity as separable determinants of realized molecular dysfunction

Gene Dosage, Functional Redundancy, and Paralog Compensation

Variant-level molecular perturbation is only one determinant of realized function because cells also differ in how much gene product is required and how much functional reserve remains when that product is altered. Large-scale analysis of rare copy-number variation has shown substantial heterogeneity in human gene dosage sensitivity, including separable patterns of haploinsufficiency and triplosensitivity (Collins *et al*., 2022). The biological interpretation of dosage therefore cannot be reduced to copy number alone. What matters is how a particular regulatory and molecular system translates altered gene product abundance into downstream function.

Controlled perturbation provides direct evidence for that principle. Modulation of SOX9 levels in human facial progenitor cells showed that regulatory elements and downstream targets differ in their sensitivity to changes in transcription-factor abundance, with some responses buffered across modest dosage changes and others responding more strongly (Naqvi *et al*., 2023). Comparative analyses of pathogenic and benign copy-number variants likewise support dosage sensitivity as an important determinant of CNV pathogenicity while leaving substantial variation among loci and genomic contexts (Rice & McLysaght, 2017). Effective dosage is therefore better treated as a relationship between gene-product supply and system demand than as a uniform consequence of gaining or losing one genomic copy.

Paralogous genes add a second source of functional reserve. Analyses across cancer cell lines showed that the expression and functional availability of paralogs contribute to variation in whether individual genes appear essential (De Kegel & Ryan, 2019; Dede *et al*., 2020; Parrish *et al*., 2021). Dual-gene perturbation is particularly informative because loss of one paralog may produce little phenotype while simultaneous disruption of its partner reveals a strong fitness defect. The absence of a phenotype after single-gene perturbation can therefore reflect genuine redundancy, but it can also reflect limited assay sensitivity or a cellular background in which the relevant dependency is not engaged. The distinction requires direct perturbational evidence.

Even when buffering is real, it is rarely safe to treat it as a fixed property of a paralog pair. Genetic dependency screens demonstrate that paralog interactions change across cellular backgrounds and genomic states (Köferle *et al*., 2022). A pair can behave redundantly in one context and asymmetrically in another because expression, regulatory control, protein stoichiometry or pathway demand differs. Functional redundancy consequently describes a conditional relationship between genes rather than a permanent annotation attached to sequence similarity.

Compensation can also arise through mechanisms that are more active than pre-existing redundancy. Experimental work in zebrafish and mouse systems demonstrated that degradation of mutant mRNA can trigger transcriptional responses involving related genes and that alleles unable to generate the relevant transcript can produce different phenotypes (El-Brolosy *et al*., 2019). This process, termed transcriptional adaptation, provides a mechanistic route through which the form of a mutation influences the response to that mutation (Sztal & Stainier, 2020). It is conceptually distinct from simple paralog coexpression: the compensatory state may be induced by the perturbation itself.

Evidence from other biological systems reinforces the need to keep these categories separate. Plant genetics shows that paralog compensation can involve regulated expression changes rather than passive interchangeability (Iohannes & Jackson, 2023). In mammals, the HAPSTR2 retrogene provides a more specific example in which a paralog can stabilize or partially substitute for its partner under defined tissue and stress conditions (Amici *et al*., 2023). Neither example establishes that paralog compensation is ubiquitous. Instead, they demonstrate that compensatory capacity depends on whether an appropriate substitute exists, is expressed, can interact with the affected system, and is available at the biological time and location in which function is required.

Table 2 therefore distinguishes dosage sensitivity from several mechanistically different forms of redundancy and compensation. The categories are intentionally non-equivalent: each predicts a different experimental signature and different reasons why a sequence perturbation might remain clinically or cellularly silent.

Table 2. Distinct biological routes linking altered gene dosage to preserved or lost function

Mechanistic state	Quantity primarily perturbed	Source of residual functional capacity	Expected perturbational signature	Principal context dependence	Interpretation within this article
Dosage-sensitive locus	Gene-product abundance	Limited intrinsic reserve	Modest dosage change produces measurable downstream response	Regulatory architecture and pathway demand	Sequence effect can be amplified when reserve is low
Regulatory buffering	Gene-product input or transcription-factor level	Nonlinear regulatory response	Output remains relatively stable across part of the dosage range	Target-enhancer architecture and cell lineage	Nominal dosage change need not equal proportional functional change
Constitutive paralog redundancy	Loss of one gene product	Pre-existing activity of a related paralog	Single loss tolerated; combined loss more deleterious	Relative expression and biochemical interchangeability	An apparently silent variant may remain buffered
Conditional paralog buffering	Gene loss in a particular cellular background	Context-available paralog function	Dependency emerges only in selected backgrounds	Cell state, genomic alterations and expression	Compensatory capacity is a state variable rather than a permanent property

Synthetic interaction	Simultaneous disruption of related genes	None sufficient after combined perturbation	Dual loss causes phenotype not produced by either single loss	Network architecture and assay setting	Single-gene experiments can underestimate functional dependence
Transcriptional adaptation	Mutation producing an unstable transcript	Induced expression of related genes	Compensation depends on mutant-transcript processing	Allele structure, RNA surveillance and transcriptional response	Mutation mechanism can influence the phenotype independently of lost protein function
Tissue-limited paralog substitution	Loss or destabilization of a primary gene product	Paralogue expressed in selected tissues or conditions	Rescue or attenuation occurs only where the substitute is available	Tissue distribution, stress state and temporal expression	The same genetic perturbation can be buffered in one context and exposed in another

These distinctions narrow the meaning of "compensation." A tolerated perturbation does not by itself demonstrate that another gene has compensated for it, and paralog similarity does not demonstrate functional interchangeability. Stronger evidence requires showing that removal or reduction of the proposed compensatory mechanism converts a previously buffered state into dysfunction, or that restoring that mechanism reverses the effect under comparable conditions. This criterion also creates a bridge to cellular context: compensatory capacity can influence phenotype only where the relevant paralog, regulatory response or reserve mechanism is actually available. The next stage of the argument therefore shifts from identifying possible buffering mechanisms to determining the biological settings in which they are expressed.

Cellular Context and the Conditional Expression of Genetic Effects

The availability of a compensatory mechanism does not guarantee that it will operate in every biological setting. Genetic interactions emerge within cells that differ in transcriptional programs, paralog expression, chromatin configuration, metabolic demand, developmental history, and other background features. Genome-scale combinatorial perturbation illustrates this dependence directly. Multiplex CRISPR/Cas12a screening can reveal interactions among paralogs that are missed by individual-gene perturbation, while the magnitude and even presence of those interactions can vary among cellular backgrounds (Esmaeili Anvar *et al*., 2024). A genetic effect observed after one perturbation therefore represents the combination of the manipulated locus and the state of the system in which that locus was tested.

Molecular background can expose dependencies that remain inconspicuous elsewhere. In SMARCB1-deficient cancer models, simultaneous suppression of the paralogs CBP and p300 produces a context-specific dependency involving altered regulation of KREMEN2 (Sasaki *et al*., 2024). The relevant point for variant interpretation is not the cancer-specific vulnerability itself but the conditional structure of the interaction. Loss of functional reserve becomes consequential because another molecular state changes the requirement for that reserve. A similar sequence perturbation in a background without the relevant state need not produce an equivalent phenotype.

Compensatory capacity also depends on whether the potential substitute is available in the tissue or condition under consideration. HAPSTR2 can stabilize and functionally buffer HAPSTR1, but its distribution and activity are not uniform across biological contexts (Amici *et al*., 2023). Thus, the existence of a paralog capable of substitution at the biochemical level does not establish that the substitution will occur wherever the primary gene is perturbed. Cellular expression, localization, interaction partners and stress conditions determine whether potential redundancy becomes effective redundancy.

Time introduces another dimension of conditionality. Experimental alteration of bicoid dosage in Drosophila produced substantial initial disruption, yet populations subsequently evolved changes that restored aspects of gene expression, morphology and reproductive performance across generations (Li *et al*., 2024). Such adaptation differs mechanistically from immediate paralog buffering or mutation-triggered transcriptional adaptation. It nevertheless demonstrates that the phenotype associated with a defined dosage perturbation may change as the surrounding biological system changes.

These observations suggest that context should not be appended to variant interpretation as an unspecified residual category. It can be decomposed into experimentally tractable variables: cellular genotype, regulatory state, tissue-specific availability of compensatory factors, developmental state, molecular demand, and the time scale over which the perturbation is observed. The same distinction applies to measurement. A phenotype that appears only in one cellular model may reflect genuine biological conditionality, an assay-specific limitation, or both. Separating these possibilities requires perturbing the proposed contextual modifier rather than inferring its role from phenotypic disagreement alone.

This multilevel dependence is summarized in **Figure 2**, which treats dysfunction as a conditional state arising within intersecting molecular, cellular and observational environments rather than as the terminal node of a universal genotype-to-phenotype pipeline.

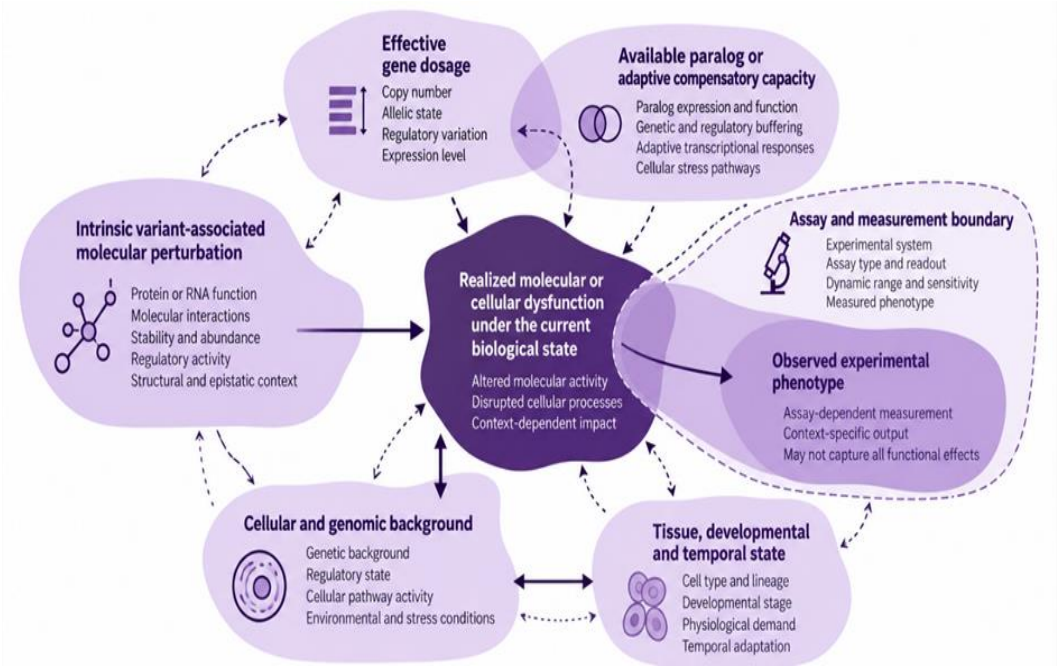


Figure 2. Context-dependent realization of genetic dysfunction across molecular, cellular and temporal states

A Conceptual Model for Interpreting Context-Dependent Molecular Dysfunction

The evidence assembled here supports a layered interpretation rather than a single deterministic translation from sequence to phenotype. Sequence and evolutionary information can provide powerful estimates of intrinsic variant effect, as illustrated by contemporary protein-level prediction methods (Frazer *et al*., 2021). That information describes an important component of biological perturbation, but it does not specify the entire state through which dysfunction becomes expressed.

A second component is effective dosage. A sequence alteration occurring in a dosage-sensitive system may have consequences that differ from a similar alteration in a system with substantial regulatory reserve. A third component is compensatory capacity, including constitutive paralog redundancy and inducible responses such as transcriptional adaptation. These mechanisms are not interchangeable: they differ in what triggers them, when they operate and how they could be experimentally abolished.

The fourth component is cellular realization. The phenotype associated with a perturbation depends on the genetic background, regulatory state, tissue environment and time scale in which the affected molecular function is required. Finally, observation itself must remain conceptually distinct from underlying biology. Methods such as MAVE-NN formalize the importance of separating latent genotype-phenotype relationships from nonlinearities and noise introduced by experimental measurement (Tareen *et al*., 2022). An assay may therefore detect only one projection of a broader biological perturbation.

The model proposed here consequently makes no claim that these components can be combined into a universal numerical score. Its contribution is instead a conditional ordering of questions. What molecular property is altered by the variant? How does the affected system translate that alteration into effective dosage? Which compensatory mechanisms are available in the relevant cell? Which cellular conditions determine demand for the perturbed function? What measurement process generates the observed phenotype? A mechanistic interpretation becomes stronger when these questions can be answered separately and when perturbing one proposed modifier changes the phenotype in the direction predicted by the account.

Results and Discussion

Variant Interpretation, Falsifiability, and Model Limitations

The principal implication of this model is epistemic as well as biological. High predictive accuracy does not automatically identify the mechanism responsible for a genetic effect. Blind benchmarking has repeatedly shown that computational methods vary across phenotypes and challenge structures (CAGI, 2024). Deep mutational scanning benchmarks likewise demonstrate that apparent performance can depend on how experimental datasets are assembled and evaluated (Livesey & Marsh, 2023). Evolutionary models can provide strong pathogenicity information while remaining largely silent about whether a particular effect arises through abundance, catalytic function, dosage sensitivity, compensation or another cellular mechanism (Frazer *et al*., 2021).

Measurement uncertainty creates a related problem. Minimum-information standards for multiplexed variant assays emphasize that experimental design, calibration and metadata are integral to interpretation (Claussnitzer *et al*., 2024). Quantitative modelling of MAVE data further shows that the observed assay signal may be a nonlinear and noisy representation of an underlying genotype-

phenotype relationship (Tareen *et al*., 2022). Consequently, disagreement between two assays cannot immediately be assigned to biological context. Technical alternatives should first be examined, especially when the assays measure different molecular layers.

The model is falsifiable because its proposed modifiers can, in principle, be removed or altered independently. A claim of paralog buffering is weakened if disruption of the proposed paralog fails to expose the phenotype attributed to compensation. A claim of transcriptional adaptation is weakened if phenotypic rescue persists when the mutation no longer produces the RNA-processing event proposed to trigger adaptation (El-Brolosy *et al*., 2019). Conversely, copy-number observations cannot by themselves establish such an adaptive mechanism, because dosage sensitivity may arise without induced compensation (Rice & McLysaght, 2017). Mechanistically distinct explanations therefore require distinct perturbations.

Transfer between biological systems is another limitation. Combinatorial genetic screens demonstrate dependencies in particular cell lines and environmental conditions (Esmaeili Anvar *et al*., 2024), while the CBP/p300 example shows how a dependency can emerge from a defined tumour-genetic background (Sasaki *et al*., 2024). These observations justify a general expectation of context dependence but do not license direct transfer of individual interactions to unrelated tissues or organisms. The conceptual model should therefore be applied most confidently when the relevant modifier—dosage, paralog availability, cellular state or adaptive response—has been observed or experimentally manipulated in the system under interpretation.

The framework also has boundaries at the opposite extreme. Some sequence alterations may impose effects sufficiently severe that compensation contributes little to the final phenotype. Others may have negligible intrinsic molecular consequences, leaving little perturbation to buffer. Context dependence should therefore be treated as an empirical hypothesis whose importance varies among variants, genes and biological systems. The strongest use of the model is not to make every genetic effect appear conditional, but to prevent untested assumptions about dosage, redundancy or measurement from being folded invisibly into variant interpretation.

Conclusion

Sequence variation initiates molecular genetic interpretation, but it does not by itself specify the biological state in which dysfunction is expressed. Evidence from functional assays, dosage-sensitive systems, paralog perturbation, transcriptional adaptation and context-dependent genetic interactions supports a distinction between intrinsic molecular perturbation and realized dysfunction. Effective dosage, compensatory capacity, cellular background, tissue or developmental state, and the conditions of measurement can each modify the relationship between these states.

The proposed account is therefore conditional rather than universal. Compensation should be inferred only when its mechanism and availability are demonstrable, and context dependence should be distinguished from assay-specific disagreement. By separating these elements, variant interpretation can move from a static description of sequence difference toward testable explanations of when, where and through which mechanisms a genetic perturbation becomes functionally consequential.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

References

Amici, D. R., Cingoz, H., Alasady, M. J., Alhayek, S., Phoumyvong, C. M., Sahni, N., Yi, S. S., & Mendillo, M. L. (2023). The HAPSTR2 retrogene buffers stress signaling and resilience in mammals. *Nature Communications, 14*, Article 152. doi:10.1038/s41467-022-35697-1

Brandes, N., Goldman, G., Wang, C. H., Ye, C. J., & Ntranos, V. (2023). Genome-wide prediction of disease variant effects with a deep protein language model. *Nature Genetics, 55*(9), 1512–1522. doi:10.1038/s41588-023-01465-0

Cagiada, M., Bottaro, S., Lindemose, S., Schenstrøm, S. M., Stein, A., Hartmann-Petersen, R., & Lindorff-Larsen, K. (2023). Discovering functionally important sites in proteins. *Nature Communications, 14*, Article 4175. doi:10.1038/s41467-023-39909-0

Cheng, J., Novati, G., Pan, J., Bycroft, C., Žemgulytė, A., Applebaum, T., Pritzel, A., Wong, L. H., Zielinski, M., Sargeant, T., et al. (2023). Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science, 381*(6664), Article eadg7492. doi:10.1126/science.adg7492

Claussnitzer, M., Parikh, V. N., Wagner, A. H., Arbesfeld, J. A., Bult, C. J., Firth, H. V., Muffley, L. A., Nguyen Ba, A. N., Riehle, K., Roth, F. P., et al. (2024). Minimum information and guidelines for reporting a multiplexed assay of variant effect. *Genome Biology, 25*, Article 100. doi:10.1186/s13059-024-03223-9

Collins, R. L., Glessner, J. T., Porcu, E., Lepamets, M., Brandon, R., Lauricella, C., Han, L., Morley, T., Niestroj, L.-M., Ulirsch, J., et al. (2022). A cross-disorder dosage sensitivity map of the human genome. *Cell, 185*(16), 3041–3055.e25. doi:10.1016/j.cell.2022.06.036

Critical Assessment of Genome Interpretation Consortium. (2024). CAGI, the Critical Assessment of Genome Interpretation, establishes progress and prospects for computational genetic variant interpretation methods. *Genome Biology, 25*, Article 53. doi:10.1186/s13059-023-03113-6

De Kegel, B., & Ryan, C. J. (2019). Paralog buffering contributes to the variable essentiality of genes in cancer cell lines. *PLOS Genetics, 15*(10), Article e1008466. doi:10.1371/journal.pgen.1008466

Dede, M., McLaughlin, M., Kim, E., & Hart, T. (2020). Multiplex

enCas12a screens detect functional buffering among paralogs otherwise masked in monogenic Cas9 knockout screens. *Genome Biology, 21*, Article 262. doi:10.1186/s13059-020-02173-2

El-Brolosy, M. A., Kontarakis, Z., Rossi, A., Kuenne, C., Günther, S., Fukuda, N., Kikhi, K., Boezio, G. L. M., Takacs, C. M., Lai, S.-L., et al. (2019). Genetic compensation triggered by mutant mRNA degradation. *Nature, 568*(7751), 193–197. doi:10.1038/s41586-019-1064-z

Esmaeili Anvar, N., Lin, C., Ma, X., Wilson, L. L., Steger, R., Sangree, A. K., Colic, M., Wang, S. H., Doench, J. G., et al. (2024). Efficient gene knockout and genetic interaction screening using the in4mer CRISPR/Cas12a multiplex knockout platform. *Nature Communications, 15*, Article 3577. doi:10.1038/s41467-024-47795-3

Findlay, G. M., Daza, R. M., Martin, B., Zhang, M. D., Leith, A. P., Gasperini, M., Janizek, J. D., Huang, X., Starita, L. M., & Shendure, J. (2018). Accurate classification of BRCA1 variants with saturation genome editing. *Nature, 562*(7726), 217–222. doi:10.1038/s41586-018-0461-z

Frazer, J., Notin, P., Dias, M., Gomez, A., Min, J. K., Brock, K., Gal, Y., & Marks, D. S. (2021). Disease variant prediction with deep generative models of evolutionary data. *Nature, 599*(7883), 91–95. doi:10.1038/s41586-021-04043-8

Gersing, S., Cagiada, M., Gebbia, M., Gjesing, A. P., Coté, A. G., Seesankar, G., Li, R., Tabet, D., Weile, J., Stein, A., et al. (2023). A comprehensive map of human glucokinase variant activity. *Genome Biology, 24*(1), Article 97. doi:10.1186/s13059-023-02935-8

Gersing, S., Schulze, T. K., Cagiada, M., Stein, A., Roth, F. P., Lindorff-Larsen, K., & Hartmann-Petersen, R. (2024). Characterizing glucokinase variant mechanisms using a multiplexed abundance assay. *Genome Biology, 25*(1), Article 98. doi:10.1186/s13059-024-03238-2

Hoskins, I., Rao, S., Tante, C., & Cenik, C. (2024). Integrated multiplexed assays of variant effect reveal determinants of catechol-O-methyltransferase gene expression. *Molecular Systems Biology, 20*(5), 481–505. doi:10.1038/s44320-024-00018-9

IGVF Consortium. (2024). Deciphering the impact of genomic variation on function. *Nature, 633*(8028), 47–57. doi:10.1038/s41586-024-07510-0

Iohannes, S. D., & Jackson, D. (2023). Tackling redundancy: Genetic mechanisms underlying paralog compensation in plants. *New Phytologist, 240*(4), 1381–1389. doi:10.1111/nph.19267

Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., Collins, R. L., Laricchia, K. M., Ganna, A., Birnbaum, D. P., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature, 581*(7809), 434–443. doi:10.1038/s41586-020-2308-7

Köferle, A., Schlattl, A., Hörmann, A., Thatikonda, V., Popa, A., Spreitzer, F., Ravichandran, M. C., Supper, V., Oberndorfer, S., Puchner, T., et al. (2022). Interrogation of cancer gene dependencies reveals paralog interactions of autosome and sex chromosome-encoded genes. *Cell Reports, 39*(2), Article 110636. doi:10.1016/j.celrep.2022.110636

Li, X. C., Gandara, L., Ekelöf, M., Richter, K., Alexandrov, T., & Crocker, J. (2024). Rapid response of fly populations to gene dosage across development and generations. *Nature Communications, 15*, Article 4551. doi:10.1038/s41467-024-48960-4

Livesey, B. J., & Marsh, J. A. (2023). Updated benchmarking of variant effect predictors using deep mutational scanning. *Molecular Systems Biology, 19*(8), Article e11474. doi:10.15252/msb.202211474

Luppino, F., Adzhubei, I. A., Cassa, C. A., & Toth-Petroczy, A. (2023). DeMAG predicts the effects of variants in clinically actionable genes by integrating structural and evolutionary epistatic features. *Nature Communications, 14*, Article 2230. doi:10.1038/s41467-023-37661-z

Matreyek, K. A., Starita, L. M., Stephany, J. J., Martin, B., Chiasson, M. A., Gray, V. E., Kircher, M., Khechaduri, A., Dines, J. N., Hause, R. J., et al. (2018). Multiplex assessment of protein variant abundance by massively parallel sequencing. *Nature Genetics, 50*(6), 874–882. doi:10.1038/s41588-018-0122-z

Naqvi, S., Kim, S., Hoskens, H., Matthews, H. S., Spritz, R. A., Klein, O. D., Hallgrímsson, B., Swigut, T., Claes, P., Pritchard, J. K., et al. (2023). Precise modulation of transcription factor levels identifies features underlying dosage sensitivity. *Nature Genetics, 55*(5), 841–851. doi:10.1038/s41588-023-01366-2

Parrish, P. C. R., Thomas, J. D., Gabel, A. M., Kamlapurkar, S., Bradley, R. K., & Berger, A. H. (2021). Discovery of synthetic lethal and tumor suppressor paralog pairs in the human genome. *Cell Reports, 36*(9), Article 109597. doi:10.1016/j.celrep.2021.109597

Rice, A. M., & McLysaght, A. (2017). Dosage sensitivity is a major determinant of human copy number variant pathogenicity. *Nature Communications, 8*, Article 14366. doi:10.1038/ncomms14366

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

Sasaki, M., Kato, D., Murakami, K., Yoshida, H., Takase, S., Otsubo, T., & Ogiwara, H. (2024). Targeting dependency on a paralog pair of CBP/p300 against de-repression of KREMEN2 in SMARCB1-deficient cancers. *Nature Communications, 15*, Article 4770. doi:10.1038/s41467-024-49063-w

Sztal, T. E., & Stainier, D. Y. R. (2020). Transcriptional adaptation: A mechanism underlying genetic robustness. *Development, 147*(15), Article dev186452. doi:10.1242/dev.186452

Tareen, A., Kooshkbaghi, M., Posfai, A., Ireland, W. T., McCandlish, D. M., & Kinney, J. B. (2022). MAVE-NN: Learning genotype-phenotype maps from multiplex assays of variant effect. *Genome Biology, 23*, Article 98. doi:10.1186/s13059-022-02661-7