

Conditional Sign Randomization with Estimated Whitening in Gaussian Kinship Models

Siyu Guo

Beijing Gewu Lingxi Technology Co., Ltd., Room 0403, 9th Floor, Building B, Changyang
Entrepreneurship Building, Changyang Town, Fangshan District, Beijing, China

Ruixiang Zhang

Guangdong Basic Research Center of Excellence for Precise Breeding of Future Crops, Guangdong
Laboratory for Lingnan Modern Agriculture, State Key Laboratory for Conservation and Utilization of
Subtropical Agro-Bioresources, College of Life Sciences, College of Agriculture, South China Agricultural
University, Guangzhou 510642, China

Benfan Lin

Department of Dermatology, Linyi People's Hospital, Linyi, China

Yunfan Zhang

Department of Automation, Tsinghua University, No. 30 Shuangqing Road, Haidian District, Beijing
100084, China

yu wang

`yu2025_w@mail.tsinghua.edu.cn`

State Key Laboratory of Membrane Biology, School of Life Sciences, Tsinghua-Peking Center for Life
Sciences, Institute for Immunology, Ministry of Education Key Laboratory of Protein Sciences, Tsinghua
University, Beijing 100084, China

Research Article

Keywords: conditional randomization, kinship mixed model, restricted likelihood, gene-set testing,
invariant preprocessing, weak-signal aggregation

Posted Date: September 9th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10965765/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare potential competing interests as follows: Siyu Guo is an
employee of Beijing Gewu Lingxi Technology Co., Ltd. Certain computational models and software

components used in this study are proprietary to and owned by the company. The company may therefore have a commercial interest in the work. The remaining authors declare no competing interests.

Conditional Sign Randomization with Estimated Whitening in Gaussian Kinship Models

Siyu Guo¹, Ruixiang Zhang², Benfan Lin³, Yunfan Zhang⁴, Yu Wang^{5,*}

¹Beijing Gewu Lingxi Technology Co., Ltd., Room 0403, 9th Floor, Building B, Changyang Entrepreneurship Building, Changyang Town, Fangshan District, Beijing, China; gewucode@gewulx.com

²Guangdong Basic Research Center of Excellence for Precise Breeding of Future Crops, Guangdong Laboratory for Lingnan Modern Agriculture, State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, College of Life Sciences, College of Agriculture, South China Agricultural University, Guangzhou 510642, China; zhangruixiang@stu.scau.edu.cn

³Department of Dermatology, Linyi People's Hospital, Linyi, China; benfan_lin@163.com

⁴Department of Automation, Tsinghua University, No. 30 Shuangqing Road, Haidian District, Beijing 100084, China; zhang-yf23@mails.tsinghua.edu.cn

⁵State Key Laboratory of Membrane Biology, School of Life Sciences, Tsinghua-Peking Center for Life Sciences, Institute for Immunology, Ministry of Education Key Laboratory of Protein Sciences, Tsinghua University, Beijing 100084, China

*Corresponding author: yu2025_w@mail.tsinghua.edu.cn

Abstract

Gene-set contrasts can organize association evidence around biological knowledge, but randomization inference must account for data-dependent preprocessing. We study a pre-specified member-versus-nonmember contrast under a Gaussian kinship model with unknown fixed effects and covariance $aK + bI$. Projection onto error contrasts and rotation into a design-fixed kinship eigenbasis yield independent uniform signs conditional on their magnitudes for every admissible variance parameter. Building on classical finite-group randomization theory, we give an explicit compute-once certificate: restricted-likelihood fitting, marker normalization and finite reference calibration depend on the contrasts only through their squares and can therefore be reused across the sign orbit. The upper-rank Monte Carlo p -value retains finite-sample conditional level control, including ties. For linear covariance families with a freely varying identity component, a common sign-symmetric representation by a fixed invertible linear map exists exactly when the projected covariance components commute. The test has no power above nominal level against alternatives preserving the uniform conditional sign law. In a separate independent-score model, we derive the positive quadratic response of bounded unsigned evidence and a two-way dense-sparse separation between set aggregation and maximum scanning. An explicit construction establishes nonvacuity of the fitted bounded statistic. Exhaustive small-orbit calculations and reproducible implementation checks accompany the proofs. The inferential target is a global structured model null, not competitive enrichment, causal genes or breeding prediction. Literature-guided maize association analysis motivates the setting.

Keywords: conditional randomization; kinship mixed model; restricted likelihood; gene-set testing; invariant preprocessing; weak-signal aggregation.

Introduction

Genome-wide association analysis produces local evidence at many correlated markers, whereas biological knowledge is commonly organized around genes, pathways and regulatory processes. A literature-derived graph can connect these two scales by defining candidate sets or structured annotations before the target response is analyzed. This organization is potentially useful when evidence is distributed across several weak effects. It also creates an inferential question: which aspects of the resulting pipeline can be fitted to the same data and then held fixed during a randomization test?

The question is particularly relevant in structured populations. Mixed models account for relatedness through a covariance component rather than by treating individuals as exchangeable (Yu et al. 2006). Efficient spectral algorithms make variance-component fitting practical (Kang et al. 2008). Once the covariance is estimated, however, a normalized residual vector should not automatically be treated as a sample of exchangeable or standard-normal coordinates. A resampling justification must address the actual transformation group, the unknown parameters, and every intermediate quantity that is reused. The same issue arises when gene windows are normalized against finite simulated references or a learned component is frozen after inspecting the target phenotype.

Our starting point is a Gaussian model with covariance $aK + bI$, where the kinship matrix K and the covariate design are fixed and both variance components are unknown. A single, response-independent basis diagonalizes this entire covariance family after the fixed effects have been removed. Conditional on the absolute values of these error contrasts, their signs are independent and uniform. This supplies a familiar finite-group experiment without requiring the estimated whitening to equal the true whitening. The central practical distinction is between recomputing an arbitrary complete statistic on every orbit element and reusing quantities fitted only at the observation. The former is covered by the group principle; the latter requires an equivalence argument. We provide that argument for a bounded gene-window contrast.

Relation to existing work

Finite-group randomization and Monte Carlo rank corrections are established theory. Hemerik and Goeman (2018) give exact tests using randomly sampled group elements with the observed configuration included. Winkler et al. (2014) study permutation and sign-flip inference for general linear models, including the treatment of nuisance regressors. Eigen sign-flip methods also predate the present analysis: Ferraccioli et al. (2022) develop an asymptotically exact eigen sign-flip score test for spatial regression with differential regularization. Their model, transformation and asymptotic justification differ from the finite-sample Gaussian kinship-contrast experiment studied here; this article does not extend or evaluate their procedure. We do not claim priority for sign flipping, eigendecomposition, restricted likelihood or the upper-rank lemma.

Gene-set testing has a separate distinction between the null model and the visual form of a score. Goeman and Bühlmann (2007) explain why competitive and self-contained null hypotheses should not be conflated. MAGMA provides an established framework for gene and gene-set analysis of GWAS data (Leeuw et al. 2015). A statistic that subtracts a nonmember mean may resemble a competitive contrast, yet the randomization distribution can still be justified only under a global model null. The present article makes this distinction explicit and does not claim to replace tests of different hypotheses. Likewise, combining networks with maize GWAS has already been demonstrated in Camoco (Schaefer et al. 2018); our maize example explains an inferential design, not a new biological discovery.

The distinction between dense and sparse alternatives is also well established, including in the normal-mixture detection literature (Donoho and Jin 2004). Our power calculations are deliberately narrower: they analyze a fixed bounded unsigned score under independence, with an explicit asymptotic threshold, and separately give a nonvacuity construction for the fitted sign statistic. They are not a minimax result or a general transfer theorem for correlated production scores.

Scope and organization

The contribution is a self-contained specification of the compatibility between estimated whitening, reference calibration and a conditional sign experiment, together with the boundaries of that specification. Section 2 states the classical randomization argument in the precise conditional form needed here. Section 3 derives a compute-once certificate for the fitted window statistic and its relation to generalized least squares. Section 4 characterizes admissible fixed-basis covariance families and gives counterexamples for residual correlation and invalid frozen adaptation. Section 5 establishes the separate aggregation and nonvacuity results, and Section 6 reports finite, reproducible calculations. Section 7 describes a literature-guided maize use case and its inferential interpretation.

For completeness, Appendix 8 collects standard derivations for downstream identification, masked fusion, graph propagation, literature selection and prediction risk. These results clarify why such modules cannot inherit the main test’s guarantee merely by sharing features. They are supporting mathematical context, not additional claims of algorithmic novelty.

Conditional randomization and reusable preprocessing

All probability statements are conditional on a fixed design and external information, denoted by $\mathcal{D}$. This information specifies a full-column-rank matrix $C \in \mathbb{R}^{n \times p}$, a symmetric positive-semidefinite kinship matrix K , a marker matrix X , test windows, and external annotations. We suppress $\mathcal{D}$ in notation. The assumption is a conditional model for the response, not probabilistic independence of genotypes and phenotypes. In the primary model,

$$H_0: \quad y = C\gamma + \varepsilon, \quad \varepsilon \sim N_n(0, aK + bI_n), \quad a \geq 0, \quad b > 0.$$

Here γ , a , and b are unknown, and $d = n - p \geq 2$. The null excludes an additional fixed mean contribution outside $\text{col}(C)$; it permits the structured random component encoded by K . It is not a competitive null allowing arbitrary fixed genetic effects with equal enrichment inside and outside a selected set.

Let $Q_0 \in \mathbb{R}^{n \times d}$ have orthonormal columns spanning $\text{col}(C)^\perp$. Choose an eigendecomposition, using the design alone,

$$Q_0^\top K Q_0 = U \Lambda U^\top, \quad H = Q_0 U, \quad q = H^\top y.$$

Then $H^\top C = 0$, $H^\top H = I_d$, and

$$q \sim N_d(0, a\Lambda + bI_d).$$

This single basis diagonalizes the entire unknown two-component covariance family. Repeated eigenvalues cause no difficulty provided the basis choice does not use y .

Write $\mathcal{G} = \{-1, 1\}^d$, with coordinatewise multiplication, and $D_s = \text{diag}(s)$. The following specialization of finite-group randomization theory (Hemerik and Goeman 2018) isolates the required conditional symmetry.

Theorem 1 (Conditional upper-rank calibration). *Let q be a random vector in $\mathbb{R}^d$ with $\Pr(q_i = 0) = 0$ for every i and $D_s q \stackrel{d}{=} q$ for every $s \in \mathcal{G}$. Let ξ be a random element of a measurable space $(\Xi, \mathcal{B}_\Xi)$, independent of q . Let $T: \mathbb{R}^d \times \Xi \rightarrow \mathbb{R}$ be measurable with respect to the product σ -field and finite-valued. For a fixed integer $B \geq 1$, draw $s_1, \dots, s_B$ independently and uniformly from $\mathcal{G}$, independently of (q, ξ) . Define $p_B = \frac{1 + \sum_{b=1}^B \mathbf{1}\{T(D_{s_b} q, \xi) \geq T(q, \xi)\}}{B+1}$. Then, for every $\alpha \in [0, 1]$, $\Pr(p_B \leq \alpha \mid |q|, \xi) \leq \frac{\lfloor \alpha(B+1) \rfloor}{B+1} \leq \alpha$ almost surely. In particular, the result holds under (2.1) for all admissible a, b, γ .*

Proof. Set $m = |q|$. For a measurable subset $A \subset (0, \infty)^d$, group invariance makes $\Pr(m \in A, \text{sign}(q) = u)$ identical for all $u \in \mathcal{G}$. Summing over the 2^d sign patterns shows that $s_0 = \text{sign}(q)$ is uniform conditionally on m . Independence of ξ gives the same conclusion conditionally on (m, ξ) . Thus $q = D_{s_0} m$.

Conditionally on (m, ξ) , $(s_0, s_1, \dots, s_B)$ is uniform on $\mathcal{G}^{B+1}$. The map

$$(s_0, s_1, \dots, s_B) \mapsto (s_0, s_1 s_0, \dots, s_B s_0)$$

is a bijection, with inverse $(u_0, u_1, \dots, u_B) \mapsto (u_0, u_1 u_0, \dots, u_B u_0)$. Its output coordinates are therefore conditionally independent and uniform. Define $W_0 := T(q, \xi)$ and $W_b := T(D_{s_b} q, \xi)$ for $b = 1, \dots, B$. Since $q = D_{s_0} m$, we have $D_{s_b} q = D_{s_b} D_{s_0} m = D_{u_b} m$ and $q = D_{u_0} m$. Thus $W_0 = T(D_{u_0} m, \xi)$ and $W_b = T(D_{u_b} m, \xi)$: all $B + 1$ values apply the same function $u \mapsto T(D_u m, \xi)$ to conditionally independent uniform coordinates, so they are conditionally independent and identically distributed.

For deterministic real numbers $w_0, \dots, w_B$, set $R_i = \#\{j: w_j \geq w_i\}$. At most k indices can have $R_i \leq k$: otherwise the smallest value among more than k such indices would have upper rank greater than k . Hence $\sum_i \mathbf{1}\{R_i \leq k\} \leq k$. Conditional exchangeability and expectation give $\Pr(R_0 \leq k \mid m, \xi) \leq k/(B + 1)$. Since

$p_B = R_0/(B + 1)$, choose $k = \lfloor \alpha(B + 1) \rfloor$. Under (2.2), positive diagonal Gaussian variances imply independent continuous symmetric coordinates and therefore the required group invariance. $\square$

The theorem concerns upper bounds, not continuous uniform p -values. Ties, identity draws, repeated draws, and repeated statistic values are included. Neither the individual coordinates of q nor the fitted residuals need to have unit variance. The conditioning is on $|q|$, not on the full observed vector q . The Monte Carlo budget is fixed independently of the observed resampling outcomes; optional stopping is not covered. Positive Gaussian coordinate variances explicitly ensure $\Pr(q_i = 0) = 0$. Because only finitely many integer rank thresholds occur, the almost-sure inequalities can be chosen simultaneously for all α .

An implemented statistic must satisfy the theorem's domain condition. It can be given a fixed finite value on a whole-orbit refusal set. If a refusal indicator is measurable with respect to $(|q|, \xi)$, conditioning on acceptance preserves the bound, provided acceptance has positive probability. Assigning $T = 0$ on a refused orbit gives $p_B = 1$ there. A sign-dependent filter on which tests are reported does not inherit this selected-report guarantee. In the present construction $q = 0$ is a null event, and markers with $H^\top x_j = 0$ are excluded by a design-only rule. Successful positive-scale fitting and finite evaluation are required; ideal-arithmetic guarantees do not certify arbitrary floating-point extremes.

Proposition 2 (A sufficient compute-once certificate). *Suppose preprocessing is $A = A(q, \xi)$ and a statistic has the complete-map form $T(q, \xi) = F(q, A(q, \xi))$. If $A(D_s q, \xi) = A(q, \xi)$ for every s , then $T(D_s q, \xi) = F(D_s q, A(q, \xi))$ for every orbit element. Thus preprocessing can be evaluated once and reused without changing the randomization statistic.*

Proof. Substitute the stated invariance into the definition of the complete map. $\square$

Magnitude-only preprocessing, $A = A(q_1^2, \dots, q_d^2, \xi)$, satisfies this certificate. Invariance is not necessary for the validity of a fully recomputed measurable statistic: Theorem 1 already permits arbitrary T . Nor does failure of the certificate by itself prove that every frozen shortcut is invalid. It removes this guarantee and requires a separate argument. Quantities extracted from $Q_1^\top y$, where Q_1 spans $\text{col}(C)$, are not automatically admissible auxiliary inputs: $\text{Cov}(Q_1^\top y, q) = a Q_1^\top K H$ need not vanish. Conditioning on them can destroy sign symmetry. A separate argument is required for such inputs.

For a predeclared family of L statistics satisfying the theorem under a common model null, the union bound controls the probability of any rejection at thresholds α/L . This is a global-null family guarantee; it does not establish strong familywise control for arbitrary configurations of distinct competitive or set-specific nulls. Shared resampling draws are allowed, since independence across the L tests is not used.

A fitted-whitening gene-set contrast

We now instantiate Proposition 2 for a gene-set statistic. The algebra is included to identify every quantity that may be reused. Restricted likelihood and spectral mixed-model computation are established methods (Patterson and Thompson 1971; Kang et al. 2008); neither is introduced here as a new estimator.

Restricted likelihood as a magnitude-only fit

For a positive-definite matrix V , define the residual precision operator

$$P_V = V^{-1} - V^{-1}C(C^\top V^{-1}C)^{-1}C^\top V^{-1}.$$

The standard error-contrast identities are

$$P_V = Q_0(Q_0^\top V Q_0)^{-1}Q_0^\top, \\ \det(V)\det(C^\top V^{-1}C) = \det(Q_0^\top V Q_0)\det(C^\top C).$$

For completeness, write $C = Q_1 R_C$, where Q_1 has orthonormal columns and R_C is nonsingular, and partition

$$[Q_1, Q_0]^\top V [Q_1, Q_0] = \begin{pmatrix} A & B_0 \\ B_0^\top & D \end{pmatrix}, \quad F = A - B_0 D^{-1} B_0^\top.$$

Both D and F are positive definite. The block inverse is

$$\begin{pmatrix} F^{-1} & -F^{-1} B_0 D^{-1} \\ -D^{-1} B_0^\top F^{-1} & D^{-1} + D^{-1} B_0^\top F^{-1} B_0 D^{-1} \end{pmatrix}.$$

Substituting $C^\top V^{-1} C = R_C^\top F^{-1} R_C$ into (3.1) cancels every block except the lower-right D^{-1} , proving (3.2). Since $\det(V) = \det(F)\det(D)$ and $\det(C^\top C) = \det(R_C)^2$, (3.3) follows.

Fit the covariance family $\sigma(K + \delta I_n)$ with $\sigma, \delta > 0$. Ignoring additive constants independent of these parameters, the negative twice restricted log likelihood is

$$\ell(\sigma, \delta) = d \log \sigma + \sum_{i=1}^d \log(\lambda_i + \delta) + \sigma^{-1} \sum_{i=1}^d \frac{q_i^2}{\lambda_i + \delta}.$$

For $q \neq 0$, set $Q(\delta) = \sum_i q_i^2 / (\lambda_i + \delta) > 0$. Differentiation gives

$$\partial_\sigma \ell = \frac{d}{\sigma} - \frac{Q(\delta)}{\sigma^2}, \quad \hat{\sigma}(\delta) = \frac{Q(\delta)}{d}.$$

The derivative is negative below and positive above this point, so $\hat{\sigma}(\delta)$ is the unique minimizer. Profiling out σ therefore yields

$$F(\delta) = d \log\{Q(\delta)/d\} + \sum_i \log(\lambda_i + \delta).$$

Any deterministic, measurable fitting rule using only (3.4), the eigenvalues and fixed tuning parameters produces $\hat{\delta}$ as a function of q^2 . For example, the implementation searches a fixed grid and a bounded local interval in $\log \delta \in [-10, 10]$. Define

$$\hat{a} = Q(\hat{\delta})/d, \quad \hat{b} = \hat{a}\hat{\delta}, \quad \hat{v}_i = \hat{a}\lambda_i + \hat{b} > 0.$$

No global optimization claim, consistency assumption, or equality to the true variance is needed for the orbit argument. In particular, the testing theorem includes the true boundary $a = 0$ even though this bounded fitting rule returns $\hat{a} > 0$. If all $\lambda_i = \lambda$, then $\hat{v}_i = \|q\|^2/d$ for all i ; the separate variance components need not be identifiable.

Normalized marker geometry and finite reference scores

Define

$$r_i = \frac{q_i}{\sqrt{\hat{v}_i}}, \quad \tilde{X} = \text{diag}(\hat{v})^{-1/2} H^\top X, \quad P_j = \frac{\sqrt{d} \tilde{X}_j}{\|\tilde{X}_j\|_2}.$$

Assume $H^\top x_j \neq 0$ for all tested markers. A positive diagonal scale preserves this nonzero property. The implementation additionally refuses the whole test if a scaled column norm is at most 10^{-10} ; this acceptance rule depends only on the design and magnitudes, so it is constant on each sign orbit. Each marker score is $z_j = P_j^\top r / \sqrt{d}$. For the same fitted covariance $\hat{V} = \hat{a}K + \hat{b}I_n$, (3.2) gives

$$z_j = \frac{x_j^\top P_{\hat{v}} y}{\sqrt{x_j^\top P_{\hat{v}} x_j}},$$

$$\frac{P_j^\top P_k}{d} = \frac{x_j^\top P_{\hat{v}} x_k}{\sqrt{(x_j^\top P_{\hat{v}} x_j)(x_k^\top P_{\hat{v}} x_k)}}.$$

These are identities at a common covariance estimate. They neither imply standard-normal fitted scores nor equate tests that use different resampling groups.

Let $W_1, \dots, W_G$ be nonempty marker windows; their sizes can differ and windows can overlap. Let $S \subset \{1, \dots, G\}$ be a fixed nonempty proper subset. For each window, generate and sort $R \geq 1$ reference maxima $\mathcal{R}_{g1}, \dots, \mathcal{R}_{gR}$ using an independent raw random source ξ and the correlation matrix $P_{W_g}^\top P_{W_g}/d$. One implementation draws a centered Gaussian vector with that correlation and takes its maximum absolute coordinate. Only fixed-orbit reproducibility of these thresholds is required for the outer test.

At a response r , put

$$t_g(r) = \max_{j \in W_g} |z_j|,$$

$$u_g(r) = \frac{1 + R - \#\{k: \mathcal{R}_{gk} < t_g(r)\}}{R + 1},$$

$$E_g(r) = -\log_{10} u_g(r), \quad T_S(r) = \frac{1}{|S|} \sum_{g \in S} E_g(r) - \frac{1}{G - |S|} \sum_{g \notin S} E_g(r).$$

The strict inequality in the reference count retains ties in the upper tail. The reference quantity u_g is an intermediate score, not an asserted conditionally uniform gene-level p -value. Each E_g lies in $[0, \log_{10}(R + 1)]$. Heterogeneous reference sets can give T_S a nonzero null mean; centering at zero is not a premise of Theorem 1.

Proposition 3 (Exact reuse for the fitted window statistic). *In ideal arithmetic, fix the design, fitting rule, R , and raw reference source ξ . On each valid sign orbit, the fitted scales, normalized markers, window correlation matrices and reference thresholds are unchanged by $q \mapsto D_s q$, whereas $r(D_s q) = D_s r(q)$. Therefore evaluating (3.8) after fitting once is exactly the same statistic as recomputing the full construction at each $D_s q$. Its upper-rank Monte Carlo p -value satisfies (2.4) under (2.1).*

Proof. Equation (3.4) reads q only through its squares, so every fitted scale is unchanged. Equation (3.5) then leaves $\tilde{X}$ and P unchanged and changes only the signs of r . Hence the window correlation matrices are unchanged. Applying the same reference-generating map to these matrices and the same ξ gives the same thresholds. Membership, window definitions, and all remaining scoring conventions are fixed. Proposition 2 and Theorem 1 apply. $\square$

Implementation summary.

The following procedure implements Proposition 3 for the fixed inputs in Section 2.

1. Compute Q_0, U, Λ and H from the design alone. Exclude markers with $H^\top x_j = 0$ and require every retained window to remain nonempty; otherwise return an invalid-design status without testing. Fix a nonempty proper member set S and integers $R, B \geq 1$ before testing.
2. Form $q = H^\top y$. Use a deterministic magnitude-only fitting rule to obtain $\hat{\delta}, \hat{a}, \hat{b}$ and the positive scales $\hat{v}$. Form r and P by (3.5). Apply only whole-orbit refusal rules, including $q = 0$, fitting failure or a scaled column norm at most the fixed tolerance. The mathematical extension assigns $p_B = 1$ on a refused orbit; an implementation may instead return an explicit refusal status.

3. Generate one collection of window references from the raw source ξ , independent of q . Retain the resulting thresholds, P and membership throughout the test.
4. Evaluate $W_0 = T_S(r)$. Draw B independent uniform signs, independently of (q, ξ) , and evaluate $W_b = T_S(D_{s_b} r)$ using the same scoring and numerical reduction as for W_0 .
5. Return $p_B = (1 + \#\{b: W_b \geq W_0\})/(B + 1)$ and report B and its resolution $1/(B + 1)$. Do not choose the budget by inspecting the randomization outcomes.

No variance refitting or reference redrawing is needed within step 4. This shortcut follows from the certificate; it does not authorize freezing arbitrary response-trained features.

Resolution and implementation obligations

For the full group, define

$$p_G = 2^{-d} \#\{s \in \mathcal{G}: T_S(D_s r) \geq T_S(r)\}.$$

Under the null it is super-uniform by the same upper-rank argument. Conditional on the complete observation and reference source, the Monte Carlo upper-tail count is $\text{Bin}(B, p_G)$. Therefore

$$\mathbb{E}[p_B \mid q, \xi] = \frac{1 + B p_G}{B + 1}.$$

The minimum Monte Carlo value is $1/(B + 1)$. Since $T_S(-r) = T_S(r)$, at least two full-group states tie with the observation, giving $p_G \geq 2^{1-d}$. This full-group lower bound is not a lower bound on the finite Monte Carlo value. A Bonferroni rejection at α/L requires $B + 1 \geq L/\alpha$; drawing too few signs can preclude rejection regardless of the observed score.

The implementation must evaluate observed and resampled states through the same numerical reduction rule, count upper-tail ties using $\geq$, and keep reference generation separate from the sign-draw stream. The orchestration layer constructs each window correlation as $P_{W_g}^\top P_{W_g}/d$ and uses separately labelled reference and sign substreams. A shared seed label is not a proof of independent ideal randomness. Numerical symmetrization and clipping of small negative eigenvalues are design- or magnitude-only operations; they do not extend the model to truly indefinite kinship matrices or correlated contrast errors. Finite-precision tests can detect specific discrepancies but cannot establish equivalence for every floating-point input or hardware platform. Coordinate permutations are a different group and do not inherit this proof from the common $+1$ formula.

Covariance and adaptation boundaries

The covariance family in (2.1) is useful because it has a parameter-independent diagonalizing basis, not because two variance components are intrinsically special. The following elementary characterization specifies when the same construction extends. It is an application of the classical simultaneous diagonalization criterion for real symmetric matrices (Horn and Johnson 2013).

Proposition 4 (A fixed-basis characterization). *Let $A_1, \dots, A_r$ be fixed real symmetric $d \times d$ matrices, and suppose $\Sigma(\theta) = \theta_0 I_d + \sum_{j=1}^r \theta_j A_j$ is positive definite for every θ in a nonempty open set $\Theta \subset \mathbb{R}^{r+1}$. There exists a single orthogonal matrix U , independent of θ , such that $D_s U^\top q \stackrel{d}{=} U^\top q$ for all $s \in \mathcal{G}$ and every $q \sim N_d(0, \Sigma(\theta))$, if and only if the A_j commute pairwise.*

Proof. A centered Gaussian vector with covariance $V > 0$ is invariant under every D_s if and only if $D_s V D_s = V$ for every s . Flipping coordinate i changes the sign of every off-diagonal entry V_{ij} with $j \neq i$, so this equality forces all off-diagonal entries to vanish. A diagonal covariance plainly suffices.

If the stated U exists, every off-diagonal entry of $U^\top \Sigma(\theta) U$ is a linear function of θ vanishing on the open set Θ . All its coefficients vanish. Thus $U^\top A_j U$ is diagonal for every j , and the A_j commute. Conversely, commuting real symmetric matrices have a common orthonormal eigenbasis: diagonalize one matrix, note that all the others preserve each of its eigenspaces, and recursively diagonalize their restrictions within those spaces. In that basis every $\Sigma(\theta)$ is diagonal, which proves sign invariance. $\square$

Remark 5 (Fixed invertible transformations). The characterization also covers any fixed invertible linear transformation L . Indeed, diagonality of $L\Sigma(\theta)L^\top$ for every $\theta \in \Theta$ forces $LL^\top =: D$ and each $LA_jL^\top$ to be diagonal. The matrix D is positive definite, and $U^\top = D^{-1/2}L$ is orthogonal with $U^\top A_j U$ diagonal. The freely varying identity coefficient is essential to this argument. Sufficiency also holds at boundary parameter values whenever the resulting covariance remains positive definite.

For a multi-kernel model, apply this proposition to the *projected* matrices $A_j = Q_0^\top K_j Q_0$, not to the unprojected kernels; the identity component projects to I_d . Pairwise commutation can be lost on projection. For example, take $C = \mathbf{1}_3$, $K_1 = \text{diag}(1,2,3)$, $K_2 = \text{diag}(1,1,2)$, and $Q_0 = [(1, -1, 0)^\top/\sqrt{2}, (1, 1, -2)^\top/\sqrt{6}]$. Then $Q_0^\top K_2 Q_0 = \text{diag}(1, 5/3)$, whereas the off-diagonal entry of $Q_0^\top K_1 Q_0$ is $-1/\sqrt{12}$, so the projections do not commute. The conclusion concerns parameter-independent full-rank linear preprocessing and the full coordinate sign group; it is not an impossibility theorem for all other tests, estimated transformations, or smaller randomization groups.

Why diagonal rescaling does not repair residual correlation.

For $q \sim N_2(0, V)$ with $V_{11} = V_{22} = 1$ and $V_{12} = \rho \in (0, 1)$, the conditional probabilities of the four sign patterns at $|q| = (m_1, m_2)$ are proportional to

$$\exp\left\{\frac{\rho s_1 s_2 m_1 m_2}{1 - \rho^2}\right\}.$$

They are not uniform. Positive diagonal scaling, even when estimated from the magnitudes, leaves the signs unchanged. For the complete-group statistic $T(q) = q_1 q_2$, the upper-tail p -value is $1/2$ on positive-product states and 1 otherwise. At nominal level $1/2$, its actual rejection probability is $1/2 + \arcsin(\rho)/\pi > 1/2$. This is a counterexample outside the diagonal sign-symmetric model, not a claim that every statistic is invalid when correlation remains.

Proposition 6 (Information discarded by sign conditioning). *Fix a measurable statistic, auxiliary randomness and sign draws as in Theorem 1; the auxiliary source remains independent of q , and the independent uniform sign draws remain independent of (q, ξ) under the alternative considered. Under any such alternative distribution with $\Pr(q_i = 0) = 0$ for every i and uniform conditional signs given $|q|$, the same upper-rank p_B still satisfies (2.4). In particular, replacing the Gaussian contrast covariance by any positive diagonal matrix cannot yield rejection probability above the stated level bound, even if that matrix is outside $\{a\Lambda + bI_d\}$.*

Proof. Conditionally on each $(|q|, \xi)$, the sign patterns have exactly the distribution used in the proof of Theorem 1. The pointwise upper-rank argument therefore applies regardless of the distribution of $|q|$. Integration gives the conclusion. A centered Gaussian with arbitrary positive diagonal covariance has precisely these uniform conditional signs. $\square$

This is the cost of eliminating the unknown coordinate scales by conditioning. The procedure can respond to departures in conditional sign geometry, but it is not an omnibus test for every departure from the Gaussian covariance family. It also explains why changing a fitted variance estimate alone cannot create power against an alternative that retains the same conditional sign law.

A sharp example of invalid frozen adaptation.

For an observed q with all coordinates nonzero, select $v = \text{sign}(q)$ and then freeze it when evaluating $T_v(x) = |v^\top x|$ over the sign orbit. The observed value is $\sum_i |q_i|$. By the strict triangle inequality for positive

magnitudes, exactly the two common-sign transformations attain that value. Hence the complete-group p -value equals 2^{1-d} for every observation. With $d = 8$, the procedure rejects every observation at level .05, even under independent centered Gaussian coordinates. If the selection is instead recomputed on every orbit element, the complete map is $T(x) = \sum_i |x_i|$, constant on the orbit, and its p -value is 1. This elementary example concerns adaptive directions; it does not assert that the implemented, externally fixed gene sets undergo this selection.

Signal aggregation: a surrogate model and a nonvacuity example

The exact calibration theorem does not establish a general power advantage. We therefore separate two questions: what a bounded unsigned score can aggregate in an independent-score model, and whether the fitted window statistic can have nontrivial power at all. Dense–sparse detection tradeoffs are classical (Donoho and Jin 2004); the calculations below concern a particular bounded contrast and make no minimax optimality claim.

Local response of bounded unsigned evidence

Let $Z \sim N(0,1)$, and let $h: [0, \infty) \rightarrow \mathbb{R}$ be bounded, nondecreasing and not almost everywhere constant under the law of $|Z|$. Write $\mu(\delta) = \mathbb{E}h(|Z + \delta|)$, $\mu_0 = \mu(0)$ and $v_0 = \text{Var}\{h(|Z|)\} > 0$.

Proposition 7 (Positive quadratic response). *As $\delta \rightarrow 0$, $\mu(\delta) = \mu_0 + c\delta^2 + O(\delta^4)$, $c = 1/2 \text{Cov}\{h(|Z|), Z^2\} > 0$. For the finite-reference step score in (3.8), at a fixed scalar set of thresholds $0 < t_1 \leq \dots \leq t_R < \infty$, let $\eta_k = -\log_{10}\{(R+1-k)/(R+1)\}$, $\eta_0 = 0$ and $a_k = \eta_k - \eta_{k-1} > 0$. Thus η_N equals $E_g = -\log_{10}u_g$ in (3.8) when exactly N reference values lie strictly below t_g . Then $h(|z|) = \sum_{k=1}^R a_k \mathbf{1}\{|z| > t_k\}$, $c = \sum_{k=1}^R a_k t_k \phi(t_k) > 0$, where ϕ is the standard-normal density.*

Proof. The convolution $\mu(\delta) = \int h(|x|)\phi(x - \delta) dx$ is smooth and even. Boundedness of h and integrable domination of Gaussian density derivatives permit differentiation under the integral in a neighborhood of zero. Thus $\mu''(0) = \mathbb{E}[h(|Z|)(Z^2 - 1)]$, and Taylor's theorem gives (5.1). For an independent copy Z' , twice the covariance equals

$$\mathbb{E}[(h(|Z|) - h(|Z'|))(Z^2 - Z'^2)].$$

The integrand is nonnegative by monotonicity. Choose a number strictly between the essential infimum and supremum of $h(|Z|)$. By nonconstancy and independence, $h(|Z|)$ lies above that number while $h(|Z'|)$ lies below it with positive probability. On this event monotonicity forces $|Z| > |Z'|$, making both factors strictly positive. Thus $c > 0$, and the same nonconstancy gives $v_0 > 0$.

The step representation is a telescoping sum, also when thresholds repeat. With $\bar{\Phi} = 1 - \Phi$,

$$\mu(\delta) = \sum_k a_k \{\bar{\Phi}(t_k - \delta) + \bar{\Phi}(t_k + \delta)\},$$

whose second derivative at zero is $2 \sum_k a_k t_k \phi(t_k)$. This proves (5.2). $\square$

Evenness alone does not imply a positive response. For the decreasing unsigned score $h(|z|) = e^{-z^2}$, direct Gaussian integration gives $\mu(\delta) = 3^{-1/2} e^{-\delta^2/3}$, with a negative quadratic coefficient. Moreover, Proposition 7 assumes a unit-variance scalar Gaussian input; it does not assert that a production marker score with fitted covariance has that law.

A two-way dense–sparse separation

In this subsection only, let $Z_g \sim N(\delta_g, 1)$ be independent for $g = 1, \dots, G$, all using the same fixed h above. Let $m = |S| \leq G/2$, with $m \rightarrow \infty$, and define

$$D_G = \frac{1}{m} \sum_{g \in S} h(|Z_g|) - \frac{1}{G-m} \sum_{g \notin S} h(|Z_g|), \quad s_0^2 = v_0 \left(\frac{1}{m} + \frac{1}{G-m} \right).$$

Let $L = \sup h - \inf h > 0$ be the oscillation of h , fix $0 < \alpha < 1/2$, and use the asymptotic set rejection rule $D_G > z_{1-\alpha} s_0$. Compare it with the Bonferroni maximum rule $\max_g |Z_g| > c_G$, where $c_G = \Phi^{-1}(1 - \alpha/(2G))$.

Theorem 8 (Separation in the independent-score model). *The following assertions hold.*

1. Under $\delta_g = 0$ for every g , the set rejection probability tends to α .
2. If all members have effects $\pm \delta_G$, all nonmembers have zero effect, and $\delta_G \rightarrow 0$, $\delta_G^2 \sqrt{m} \rightarrow \infty$, then the set rejection probability tends to one. If also $\delta_G \sqrt{\log G} \rightarrow 0$, the maximum rejection probability tends to $1 - e^{-\alpha}$, its own null limit.
3. If at most a fixed number k of members have nonzero effects, of any sizes depending on G , and the other effects are zero, the set rejection probability still tends to α .
4. In particular, with a single affected member of size $2c_G$, the maximum rejection probability tends to one while the set rejection probability tends to α .

Proof. Under the null, write $D_G = \sum_g w_g \{h(|Z_g|) - \mu_0\}$ with $w_g = 1/m$ on S and $w_g = -1/(G-m)$ otherwise. The standardized summands have total variance one and absolute values at most $L/\sqrt{v_0 m} \rightarrow 0$. The Lindeberg condition follows directly, so $D_G/s_0 \Rightarrow N(0,1)$, proving the first assertion.

For the dense alternative, evenness gives $\mathbb{E}D_G = c\delta_G^2 + O(\delta_G^4)$. Independence and the bound $\text{Var}(V) \leq L^2/4$ for a variable of range L yield

$$\text{Var}(D_G) \leq \frac{L^2}{4} \left(\frac{1}{m} + \frac{1}{G-m} \right) \leq \frac{L^2}{2m}.$$

Since $z_{1-\alpha} s_0 = O(m^{-1/2}) = o(\delta_G^2)$, the mean exceeds the threshold by at least $c\delta_G^2/4$ for all sufficiently large G . Chebyshev's inequality bounds the nonrejection probability by $8L^2/(c^2 m \delta_G^4) \rightarrow 0$.

For the maximum statistic, integration by parts gives the normal tail bounds

$$\frac{x\phi(x)}{1+x^2} \leq \bar{\Phi}(x) \leq \frac{\phi(x)}{x} \quad (x > 0).$$

They imply $c_G \rightarrow \infty$ and $c_G = O(\sqrt{\log G})$. The additional assumption gives $c_G \delta_G \rightarrow 0$, and

$$\frac{\phi(c_G \pm \delta_G)}{\phi(c_G)} = \exp(\mp c_G \delta_G - \delta_G^2/2) \rightarrow 1.$$

Combining this ratio with the tail bounds shows that a member's exceedance probability is $(\alpha/G)(1 + o(1))$, compared with α/G for a null coordinate. If p_g denotes these exceedance probabilities, then $\sum_g p_g = \alpha + o(1)$ and $\sum_g p_g^2 = O(G^{-1})$. Independence gives

$$\log \Pr(\max_g |Z_g| \leq c_G) = \sum_g \log(1 - p_g) \rightarrow -\alpha.$$

The resulting rejection limit is not zero and is not exactly α ; it equals the maximum test's own Bonferroni null limit. The dense conditions are compatible, for example with $m = \lfloor G/2 \rfloor$ and $\delta_G = m^{-1/8}$.

For a sparse alternative, couple the shifted and null scores using the same independent Gaussian noises.

Changing at most k member scores changes D_G by at most kL/m . Dividing by s_0 gives a bound $kL/\sqrt{v_0 m} \rightarrow 0$. The null central limit theorem and Slutsky's theorem prove the third assertion, uniformly in the sizes of those

k shifts. Finally, if one coordinate has mean $2c_G$, the event $Z_g > c_G$ has probability $\Phi(c_G) \rightarrow 1$. The maximum rejection probability is at least this probability, proving the fourth assertion. $\square$

This theorem concerns an independent-score contrast with a fixed scoring function and an asymptotic normal threshold. It is not a power theorem for the fitted sign-orbit test with linkage disequilibrium (LD), overlapping windows and random reference thresholds. In particular, increasing R with G changes the scoring function and requires new conditions. The theorem explains complementarity, not uniform superiority over single-marker analysis.

Prior concentration and dependent evidence

For a signal set A with common effect magnitude δ , write $\pi_S = |A \cap S|/|S|$ and $\pi_C = |A \cap S^c|/|S^c|$. Under the same marginal scalar model,

$$\mathbb{E}D_G = (\pi_S - \pi_C)\{\mu(\delta) - \mu_0\}.$$

This follows by separating signal and null coordinates within each mean and does not require independence. With heterogeneous shifts the expression is $\sum_g w_g \{\mu(\delta_g) - \mu_0\}$. Thus the useful feature of a prior is relative signal concentration, not merely the number of annotated genes.

If an evidence vector E has covariance Ω , then

$$\text{Var}(w^\top E) = w^\top \Omega w.$$

For example, $\Omega = \nu I + \sigma^2 \mathbf{1}\mathbf{1}^\top$ adds no common-noise term because $\sum_g w_g = 0$. By contrast, a term $\sigma^2 \mathbf{1}_S \mathbf{1}_S^\top$ adds variance σ^2 , which does not disappear as the member set grows. Correlation can therefore cancel or accumulate; counting genes does not measure the number of independent pieces of evidence.

Nonvacuity of the fitted bounded statistic

The next construction is separate from the independent-score theorem and uses the actual fitted-score form. Take $d = 8$, $Q_0^\top K Q_0 = \lambda I_8$, two normalized marker columns $P_1 = \mathbf{1}_8$, $P_2 = \sqrt{8}e_1$, and two singleton windows, with the first a member. Fix one reference threshold 2.5 in each window. Consider $q = t\mathbf{1}_8 + \epsilon$, $\epsilon \sim N_8(0, I_8)$, as $t \rightarrow \infty$. The equal-eigenvalue fit yields $\hat{v}_i = \|q\|^2/8$, hence $r \rightarrow \mathbf{1}_8$ almost surely under this coupling.

For every $s \in \mathcal{G}$,

$$|z_1(D_s r)| \rightarrow \frac{|\sum_i s_i|}{\sqrt{8}}, \quad |z_2(D_s r)| \rightarrow 1.$$

Because $6/\sqrt{8} < 2.5 < \sqrt{8}$, only the all-positive and all-negative patterns exceed the member threshold, while the nonmember window never exceeds its threshold. There are finitely many states and strict threshold gaps, so their classifications are simultaneously stable for sufficiently large t on almost every noise realization. The observed full-group p -value therefore eventually equals $2/256 = 1/128$. The complete-group test has power tending to one at level .05. For $B = 99$, its Monte Carlo counterpart instead has limiting power

$$\Pr\{\text{Bin}(99, 1/128) \leq 4\} = \sum_{j=0}^4 \binom{99}{j} (1/128)^j (127/128)^{99-j} \approx 0.9988671474.$$

The same argument holds whenever the member reference lies in $(6/\sqrt{8}, \sqrt{8})$ and the nonmember reference exceeds 1; 2.5 is a convenient representative. This event has positive probability for independently generated singleton-window $|N(0,1)|$ references. The stated limiting powers are conditional on such fixed admissible references, not unconditional powers averaged over all reference draws. The design can be constructed as $X = H[P_1, P_2]$, but these residual-coordinate columns need not be realizable as maize allele dosages. This is

an existence result for the statistic, not an empirical breeding result or a transfer of Theorem 8 to the production pipeline.

Finite numerical verification

Scope and fixed settings

All calculations in this section use synthetic designs or analytic probability formulas. The settings were written to a verification contract before the calculations were executed. Three fixed seeds, 61001, 61002 and 61003, define designs with $n = 12$, $p = 4$ and $d = 8$. The first column of C is an intercept and its other columns are independent standard-normal draws. A 12×7 standard-normal matrix B_0 defines $K = B_0 B_0^\top / 7$; ten standard-normal marker columns define X . Responses are generated as $C(1,2,3,4)^\top + \text{chol}(1.3K + 0.7I)z$, with $z \sim N_{12}(0, I)$.

Using one-based marker indices, the six windows are (1), (2,3), (3,4,5), (6), (7,8) and (8,9,10); the first two are members. Each window has 31 reference draws. For each design we enumerate all $2^8 = 256$ signs while holding magnitudes and the raw reference source fixed. The source archive contains the exact draw order, seed convention, versions and file hashes. These small constructions test the mathematics and its implementation correspondence, not the effectiveness of a biological prior.

Conditional ranks and algebraic identities

For each orbit state s , compute $p_G(s)$ by exhaustive ranking. The conditional Monte Carlo rejection probability at budget B is then evaluated without sampling Monte Carlo replicates:

$$\frac{1}{256} \sum_{s \in \mathcal{G}} \Pr \{ \text{Bin}(B, p_G(s)) \leq \lfloor \alpha(B + 1) \rfloor - 1 \}.$$

The expression is zero if the integer threshold is negative. We check $B \in \{19, 99, 999\}$ and $\alpha \in \{.01, .05, .10, .50\}$ for every design. All 36 evaluations satisfy their conditional upper-rank bounds within the stated numerical tolerance. Table 1 shows the $B = 99$ calculations. There is no Monte Carlo sampling error in this table; numerical evaluation still uses finite precision.

Orbit-conditional rejection probabilities from (6.1), with $B = 99$. The rank bound is $\lfloor \alpha(B + 1) \rfloor / (B + 1)$. Each design is one fixed magnitude/reference configuration, not a collection of independently simulated studies.

α	Rank bound	Design 1	Design 2	Design 3
0.010000	0.010000	0.006589	0.006592	0.006592
0.050000	0.050000	0.045407	0.046015	0.046094
0.100000	0.100000	0.094136	0.095167	0.096094
0.500000	0.500000	0.495994	0.495602	0.496094

For every design, the REML fit obtained directly from $D_s q$ is bitwise identical across all 256 sign patterns. At the fitted covariance, the maximum absolute discrepancy between spectral and direct GLS scores is 7.55×10^{-15} , and the corresponding Gram discrepancy is 1.67×10^{-15} . These finite checks support the identities in Section 3; the proof, not the observed discrepancy, carries their general statement.

Replay through the implemented resampler

A supplemental contract specifies a deterministic enumeration adapter for the production resampler. It supplies all 256 sign patterns at chunk widths 1, 17 and 256. Per-state contrasts are exposed through a

comparison channel using the primary membership mask and the same reduction on the same window evidence. All nine design-chunk comparisons reproduce the independently evaluated per-state statistics, the observed statistic and the upper-tail decisions exactly; the primary p -value also equals its enumerated formula. Since the adapter includes all group states *and* the resampler retains its observed +1, the checked output is $(1 + 256p_G)/257$, not p_G . Deterministic enumeration checks the computation; it is not an iid resampling experiment.

We also reconstruct phenotype-space orbit elements as $y_s = y + H(D_s - I)q$, refit the spectral transformation, and regenerate references from the same raw source. This floating-point round trip is not bitwise exact: the largest fit-parameter discrepancy, scaled by $1 + |\text{baseline}|$, is 2.99×10^{-10} ; the largest marker and residual scaled discrepancies are 4.63×10^{-11} and 4.04×10^{-11} , respectively. The maximum reference discrepancy is 1.91×10^{-10} . Nevertheless, all recomputed scalar window statistics agree exactly on these inputs. These differences are retained in the results rather than reclassified as an exact floating-point equivalence theorem.

The existing repository regression suite passes 65 tests, and its current mathematical runner passes 10 check groups. The standalone numerical calculations record Python 3.13.11 and NumPy 2.4.2; the replay additionally records SciPy 1.16.3. The reproduction requirements pin these NumPy and SciPy versions. The repository suite uses a separate environment with NumPy 2.5.2 and SciPy 1.18.1. These evidence sets are identified separately. Successful software checks are not a certification of mathematical proofs or population-level calibration.

Finite illustrations of the distinct power regimes

For an exactly computable surrogate illustration, take $h(|z|) = \mathbf{1}\{|z| > 2\}$, $m = G/2$ and $\alpha = .05$. Let $p_0 = 2\bar{\Phi}(2)$ and $p_\delta = \bar{\Phi}(2 - \delta) + \bar{\Phi}(2 + \delta)$. Under the dense alternative, the member and nonmember exceedance counts are independent $\text{Bin}(m, p_\delta)$ and $\text{Bin}(m, p_0)$ variables. The set rule rejects when their difference is at least

$$k_m = \lfloor z_{.95} \sqrt{2mp_0(1 - p_0)} \rfloor + 1.$$

Its rejection probability is computed by summing the member binomial mass times the corresponding nonmember cumulative probability. Maximum-test probabilities are computed from the product of the Gaussian nonexceedance probabilities.

Table 2 uses $\delta_G = m^{-1/8}$ for all members in the dense column. The sparse column instead contains one member of fixed effect 5 and all other effects zero. The latter illustrates dilution; it is not the diverging $2c_G$ sequence in the fourth assertion of Theorem 8. The set threshold is asymptotic, so its finite null probabilities need not equal .05. The maximum test's null probability is $1 - (1 - .05/G)^G$, approximately .0488 at these sizes.

Rejection probabilities in the independent-score model, computed from binomial sums and Gaussian tails. Dense effects are $\delta_G = (G/2)^{-1/8}$ on every member; sparse effects are fixed at 5 on one member. These are different alternatives and are not fitted-engine or maize-performance measurements.

G	δ_G	Null set	Dense set	Dense max	Sparse set	Sparse max
128	0.5946	0.0667	0.3545	0.1195	0.1347	0.9305
512	0.5000	0.0554	0.4637	0.1076	0.0816	0.8717
2048	0.4204	0.0500	0.6617	0.0960	0.0613	0.7929
8192	0.3536	0.0474	0.8867	0.0859	0.0526	0.6988

The dense maximum rejection rates decrease only slowly toward their null limit; $\delta_G \sqrt{\log G}$ is still of order one at the displayed sizes. The asymptotic separation is therefore not used as a finite-sample approximation. Separately, exhaustive enumeration at the limiting configuration of Section 5.4 gives $p_G = 1/128$ and the binomial probability in (5.5). The convergence to that configuration is proved analytically, not estimated from the table.

Interpretation and a literature-guided maize use case

An admissible knowledge-to-test interface

Consider a pre-specified maize trait, such as flowering time. A literature collection, a genome assembly, gene identifiers and an explicit relation-selection rule can be used to define a set S . Individual-level genotypes, phenotypes and experimental covariates supply X , y and C ; background markers supply a fixed K . Graph edges must retain source and study provenance so that the analyst can assess whether the information defining S is external to the target response. Publication cutoffs and content hashes help document this choice but cannot prove the absence of sample overlap, outcome-informed curation or upstream dependence.

Conditioning on the supplied design and external information must leave the response model in (2.1) valid. This is the statistical assumption; a file being frozen is only an engineering fact. The set can be selected by a complicated rule on genuinely external information, or by a rule proved invariant on the tested sign orbit. Other same-response selection cannot be frozen on the strength of the certificate alone; it requires either recomputation on every orbit element or a separate validity argument. The counterexample in Section 4 shows why.

The construction requires matched individual-level observations. A table of SNP identifiers, estimated effects and marginal GWAS p -values alone does not supply the response contrasts or their fitted geometry. A summary-statistic extension would be a different procedure with its own observation model. The present mathematical analysis does not require single-cell measurements or a drug-response encoder.

What a rejection means

The reported p -value measures the extremeness of a pre-specified member-minus-nonmember contrast under the stated global structured model. It does not establish that members outperform nonmembers under every polygenic alternative, that a specific gene is causal, or that an additional feature improves breeding-value prediction. Gene length, marker density, LD, overlapping windows and heterogeneous reference thresholds can affect the statistic and its power. The conditional orbit calibration accommodates these fixed design features under the model; it does not make them biologically irrelevant.

Matched random sets and off-trait sets may be useful descriptive comparators. Their construction does not automatically supply a competitive null or an exchangeability theorem. Predeclared test families may use the global-null union bound in Section 2; post hoc selection of favorable sets requires additional adjustment. Conversely, a nonsignificant result can reflect lack of signal, dilution, coarse Monte Carlo resolution, or a direction discarded by conditioning. These possibilities should not be reduced to a single assessment of graph quality.

Consequences for modular algorithm design

The analysis suggests three distinct outputs. The calibrated branch returns a fixed-set statistic and its model-specific randomization p -value. An exploratory representation branch may organize literature, expression profiles or fixed expert scores into a candidate ranking. A prediction branch, if developed, must define a held-out target, the available information at deployment, and an out-of-sample risk. Sharing gene identifiers across these branches is useful; sharing inferential labels is not justified by that connection alone.

For a representation to enter the calibrated branch, one must either train it on admissible external data, certify its orbit invariance, or include its full adaptation in every randomization evaluation. A complex nonlinear statistic is not excluded in principle by Theorem 1; the problem is unaccounted-for adaptation or an incorrect transformation law. The fixed-basis characterization further shows why adding an arbitrary second kinship or environment kernel is not a harmless extension: the projected covariance components need not commute.

The knowledge graph’s potential statistical role is correspondingly specific. In the scalar model, it can improve relative signal concentration through (5.3). It can also encode external information not already present in a marker summary. Whether that information resolves an LD ambiguity or improves prediction depends on the observation and conditional-risk models in Appendix 8, rather than on the complexity of the architecture alone.

Conclusion

Conditional sign randomization provides a finite-sample calibration route in a Gaussian kinship model with unknown variance components. The reusable fitted construction follows from an explicit invariance certificate, not from treating fitted residuals as known-whitened observations. Its covariance extension is governed by simultaneous diagonalization after nuisance projection, and its power is limited by the information retained in conditional signs. Bounded unsigned aggregation and maximum scanning can be complementary under explicit alternatives, but their simplified-model separation should not be mistaken for a general GWAS performance guarantee. Together, the proofs and finite checks provide a precise mathematical basis for one testing module. The exploratory ranking and prediction components of a knowledge-guided workflow lie outside this guarantee and are addressed only by the standard identities in Appendix 8.

Reproducibility and use of generative tools

All numerical examples in this article are synthetic or analytic; no new real-maize or wet-laboratory experiment is reported. The accompanying reproducibility materials contain the calculation scripts, frozen numerical settings, source identifiers and machine-readable results. Software checks test specified finite inputs and do not replace proofs.

OpenAI Codex and the API model identified as `claude-fable-5-1` were used for mathematical drafting, adversarial checking, code preparation and English editing, including preparation of the accompanying review and audit notes. Their outputs do not constitute peer review, formal proof certification or human verification. Responsibility for independently verifying and approving every derivation, source correspondence and numerical statement rests with the human authors; neither system is an author.

Standard identities for downstream modules

This appendix preserves mathematical context for an extended knowledge-guided workflow. Notation introduced within each subsection is local to that subsection. The results are standard linear-algebra, differentiation and conditional-expectation arguments, or elementary counterexamples. They are not additional novelty claims and do not extend the null guarantee of Theorem 1 to learned ranking or genomic prediction.

Observation equivalence, LD and untyped effects

Consider the separate observation model $Y_{\text{obs}} \sim N(A\theta, \Sigma)$, where A is fixed and $\Sigma \succcurlyeq 0$ does not depend on θ . Two parameter values give the same observation distribution if and only if

$$A(\theta - \theta') = 0.$$

Necessity follows from equality of means. For sufficiency, the Gaussian characteristic function is $\exp\{it^\top A\theta - 1/2 t^\top \Sigma t\}$, identical at the two parameter values when (A.1) holds. This also covers singular Σ . Any measurable algorithm applied to the same observation law and the same independent randomization law has the same output law. An external prior can prefer one explanation, but cannot make identical likelihoods distinguishable from these observations alone.

When $\Sigma > 0$, let $\Delta = A(\theta - \theta')$ and $d_*^2 = \Delta^\top \Sigma^{-1} \Delta$. Expanding the Gaussian log-density ratio and taking expectation gives

$$\text{KL}(P_\theta \parallel P_{\theta'}) = d_*^2/2.$$

Whitening and rotating reduce the likelihood-ratio decision boundary to the midpoint between two univariate unit-variance normals separated by d_* . Consequently

$$\text{TV}(P_\theta, P_{\theta'}) = 2\Phi(d_*/2) - 1 \leq \frac{d_*}{\sqrt{2\pi}}.$$

The inequality follows by integrating a normal density bounded by $1/\sqrt{2\pi}$. A test of level α under $P_{\theta'}$ has power at most $\min\{1, \alpha + \text{TV}(P_\theta, P_{\theta'})\}$ under P_θ .

For the illustrative summary-score model $z \sim N(\sqrt{n}R\beta, R)$, complete LD with $R = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ makes $(b, 0)$ and $(0, b)$ observationally equivalent. If $R = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ with $|\rho| < 1$, the matrix is nonsingular and the squared separation is $2nb^2(1 - \rho)$. High correlation gives a small separation, not exact nonidentifiability. Also $R(b, -b)^\top = (b(1 - \rho), -b(1 - \rho))^\top$, whereas $R(b, b)^\top = (b(1 + \rho), b(1 + \rho))^\top$. Unsigned marginal evidence is therefore not insensitive to all within-LD effect-sign configurations.

For fixed $V > 0$, consider $y = C\gamma + u + \varepsilon$ with $\varepsilon \sim N(0, V)$. The GLS numerator $s = X^\top P_V y$ has covariance $X^\top P_V X$, because (3.2) implies $P_V V P_V = P_V$, and mean $X^\top P_V u$ because $P_V C = 0$. Comparing this model with the same noise law and $u = 0$, the distribution of s changes if and only if $X^\top P_V u \neq 0$. The unweighted condition $X^\top u = 0$ is insufficient: with no covariates (so $P_V = V^{-1}$), $X = e_1$, $u = e_2$ and $V = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ for $0 < |\rho| < 1$, one has $X^\top u = 0$ but $X^\top V^{-1} u = -\rho/(1 - \rho^2) \neq 0$. This is a separate fixed-covariance calculation, not a guarantee after refitting the full engine.

Likewise, observing E environment-specific effects only through $w^\top \beta$ with $w \neq 0$ leaves an $(E - 1)$ -dimensional null space. One additional linear contrast raises the rank by at most one. Full identification requires the stacked observation matrix to have full column rank; an extra model component cannot substitute for an absent contrast.

Masked fixed-expert fusion and its exact gradient

Let e_{im} be fixed expert scores and x_{im} fixed gate features. All masks, expert scores and features in this subsection are fixed with respect to the gate parameters. For the available modality set A_i , define

$$\ell_{im} = x_{im}^\top v_m + b_m, \quad \alpha_{im} = \frac{e^{\ell_{im}}}{\sum_{j \in A_i} e^{\ell_{ij}}}, \quad f_i = \sum_{m \in A_i} \alpha_{im} e_{im}.$$

Unavailable weights are zero. If A_i is empty, set every α_{im} and f_i to zero, without evaluating an empty softmax. For available modalities the quotient rule gives

$$\frac{\partial \alpha_{ij}}{\partial \ell_{im}} = \alpha_{ij}(\mathbf{1}\{j = m\} - \alpha_{im}), \quad \frac{\partial f_i}{\partial \ell_{im}} = \alpha_{im}(e_{im} - f_i).$$

No encoder derivative appears because the experts are fixed.

For $N \geq 1$ positive rows, each with $J_- \geq 1$ negative scores, let $\tau > 0$, $0 \leq \pi < 1$, $w = 1 - \pi > 0$, and

$$\begin{aligned} s_i^+ &= f_i^+ / \tau, & s_{ik}^- &= f_{ik}^- / \tau, \\ D_i &= e^{s_i^+} + w \sum_{k=1}^{J_-} e^{s_{ik}^-}, \\ \mathcal{L} &= \frac{1}{N} \sum_i (\log D_i - s_i^+) + \frac{\lambda}{2} \sum_m \|v_m\|^2, & \lambda &\geq 0. \end{aligned}$$

Differentiating the log denominator and then dividing by τ yields

$$g_i^+ = \frac{1}{N\tau} \left(\frac{e^{s_i^+}}{D_i} - 1 \right), \quad g_{ik}^- = \frac{w e^{s_{ik}^-}}{N\tau D_i}.$$

Index all positive and negative score occurrences by a , with derivative g_a . The chain rule gives

$$\begin{aligned} \nabla_{v_m} \mathcal{L} &= \sum_a g_a \alpha_{am} (e_{am} - f_a) x_{am} + \lambda v_m, \\ \partial_{b_m} \mathcal{L} &= \sum_a g_a \alpha_{am} (e_{am} - f_a). \end{aligned}$$

Repeated use of the same item requires summing its occurrences. Biases are not regularized in (A.3).

For each nonempty row, f_i is in the convex hull of the available expert scores. With one available expert the data gradient is zero because $e_{im} = f_i$; with none it is zero by the fixed output convention. The penalty gradient can remain nonzero. Adding a common scalar to all logits leaves the weights unchanged, and the sum of the data derivatives with respect to the shared modality biases is zero.

To evaluate (A.3) in the log domain, put

$$M_i = \max\{s_i^+, \max_k (s_{ik}^- + \log w)\}.$$

Factoring e^{M_i} out of the denominator gives the exact identity

$$\log D_i - s_i^+ = (M_i - s_i^+) + \log \left[e^{s_i^+ - M_i} + \sum_k e^{s_{ik}^- + \log w - M_i} \right].$$

Every exponent in brackets is nonpositive and at least one term is one. The algebraic derivatives remain (A.4); no derivative of the chosen maximum is needed to establish this identity. This avoids taking a logarithm of an already underflowed positive probability, but does not certify finite output for every extreme floating-point temperature or input.

Define $\mathcal{L}_{\text{mean}}$ by using the same experts, rows, negative samples, τ and π in (A.3), with equal weights on each nonempty available-modality set and no penalty. Setting $v_m = b_m = 0$ reproduces these weights with zero penalty, so

$$\inf_{v,b} \mathcal{L}_{\text{gate}}(v, b) \leq \mathcal{L}_{\text{gate}}(0, 0) = \mathcal{L}_{\text{mean}}.$$

This is a feasible-set inclusion, not an optimization-convergence or test-risk theorem. A best-single-expert oracle is not an upper performance bound for all fusions: experts -1 and $+1$ have squared error one for target zero, while their equal-weight fusion has zero error. A truly optimal arbitrary convex-weight relaxation is a different comparator whose feasible set does contain the parameterized gate.

Linear graph propagation and its retained component

Let $A = A^\top \geq 0$ be the adjacency matrix of a finite connected undirected non-bipartite graph with at least two nodes and positive degrees d_i . Define $D = \text{diag}(d_i)$, $P = D^{-1}A$, $S = D^{-1/2}AD^{-1/2}$ and $u = \sqrt{d}/\sqrt{\sum_i d_i}$. For $y_i = x_i/\sqrt{d_i}$,

$$\|x\|^2 \mp x^\top Sx = 1/2 \sum_{ij} A_{ij} (y_i \mp y_j)^2 \geq 0.$$

Hence the eigenvalues of the symmetric matrix S lie in $[-1, 1]$. The equality condition for eigenvalue 1 forces y to be constant by connectivity; thus it is simple with eigenvector u . An eigenvalue -1 would force opposite signs along every edge and hence a bipartition, excluded by assumption. With $\rho = \max_{j \geq 2} |\lambda_j(S)| < 1$, orthogonal spectral expansion gives

$$\|S^k f - uu^\top f\|_2 \leq \rho^k \|f - uu^\top f\|_2.$$

Since $P = D^{-1/2}SD^{1/2}$, $P^k f \rightarrow \mathbf{1}\pi^\top f$ with $\pi_i = d_i/\sum_j d_j$, and

$$\|P^k f - \mathbf{1}\pi^\top f\|_\pi \leq \rho^k \|f - \mathbf{1}\pi^\top f\|_\pi, \quad \|f\|_\pi^2 = \sum_i \pi_i f_i^2.$$

This follows by applying the previous bound to $D^{1/2}f$ and dividing by $\sqrt{\sum_i d_i}$. A two-node single-edge graph instead makes $(1, -1)$ alternate forever, explaining the non-bipartite assumption. These are facts about a specified linear operator, not a theorem for arbitrary nonlinear graph networks or normalized candidate rankings.

Literature selection and positive-unlabeled labels

Let $Y \in \{0, 1\}$ be true relevance and $L \in \{0, 1\}$ indicate a recorded positive in the literature. Under the no-false-positive assumption $L = 1 \Rightarrow Y = 1$, put $\eta(x) = \Pr(Y = 1 | x)$ and $e(x) = \Pr(L = 1 | Y = 1, x)$. Then

$$g(x) := \Pr(L = 1 | x) = e(x)\eta(x).$$

This selection mechanism is central to positive-unlabeled learning (Elkan and Noto 2008). Let X take two values, each with positive probability, with the same marginal law in both scenarios. The choices $\eta = (.4, .6)$, $e = (.5, .5)$ and $\eta' = (.8, .4)$, $e' = (.25, .75)$ both give $g = (.2, .3)$, so the joint laws of (X, L) coincide although the true relevance rankings are reversed. Thus the observed (X, L) law does not identify the ranking without a selection restriction. A positive constant e preserves the ranking, but an unknown value need not identify calibrated probabilities. A known positive $e(x)$ permits division in (A.6).

Let the true class prior be $\pi_Y = \Pr(Y = 1)$, and let P and N denote the class-conditional feature laws. The mixture identity $\mathbb{E}_X \ell(f, 0) = \pi_Y \mathbb{E}_P \ell(f, 0) + (1 - \pi_Y) \mathbb{E}_N \ell(f, 0)$ yields

$$R(f) = \pi_Y \mathbb{E}_P \ell(f, 1) + \mathbb{E}_X \ell(f, 0) - \pi_Y \mathbb{E}_P \ell(f, 0).$$

An unbiased empirical version additionally requires appropriate sampling and selection assumptions. It is not the weighted contrastive loss (A.3). For freely varying positive and negative scores a, b , a row of that log-sum loss has mixed derivative $-we^{a+b}/D^2 < 0$; a sum of separate per-sample PU risks has zero mixed derivative between those scores. Putting a factor called $1 - \pi$ inside a logarithmic denominator therefore does not establish unbiased PU risk, calibrated relevance or a mutual-information lower bound.

Information, irreducible risk and prediction increments

Let $Y \in L^2$, let $\mathcal{F}$ contain all admissible information, and let ξ be independent of $(Y, \mathcal{F})$. Any prediction $h \in L^2$ considered here is measurable with respect to $\sigma(\mathcal{F}, \xi)$. Writing $m = \mathbb{E}[Y | \mathcal{F}]$ and expanding $Y - h = (Y - m) + (m - h)$ gives

$$\mathbb{E}(Y - h)^2 = \mathbb{E}\text{Var}(Y | \mathcal{F}) + \mathbb{E}(m - h)^2,$$

because $\mathbb{E}[Y - m | \mathcal{F}, \xi] = 0$ eliminates the cross term. Availability masks, training information and any allowed covariates must be included in $\mathcal{F}$ when applicable. Architecture complexity cannot remove the first term.

For nested information sets $\mathcal{F}_0 \subseteq \mathcal{F}_1$, let $m_j = \mathbb{E}[Y | \mathcal{F}_j]$ and $R(f) = \mathbb{E}(Y - f)^2$. Conditional-expectation orthogonality gives

$$\begin{aligned} R(m_0) - R(m_1) &= \mathbb{E}(m_1 - m_0)^2, \\ R(f) - R(m_0) &= \mathbb{E}(f - m_1)^2 - \mathbb{E}(m_1 - m_0)^2 \end{aligned}$$

for each square-integrable $\mathcal{F}_1$ -measurable f . To see this, expand $Y - m_0 = (Y - m_1) + (m_1 - m_0)$ and then $Y - f = (Y - m_1) + (m_1 - f)$; the cross terms vanish after conditioning on $\mathcal{F}_1$. Strict oracle improvement requires $m_1 \neq m_0$ with positive probability, whereas a fitted predictor can lose that gain through estimation error.

For a jointly Gaussian pair (U, y) with $V_y = \text{Var}(y) > 0$, the predictor

$$m(y) = \mu_U + C_{Uy}V_y^{-1}(y - \mu_y)$$

has residual uncorrelated with y , hence independent of y by joint Gaussianity. It is therefore $\mathbb{E}[U | y]$, with residual variance $\text{Var}(U) - C_{Uy}V_y^{-1}C_{yU}$. This identity motivates a distinct prediction model; it is not implemented by a gene-set significance test.

Finally, let A, δ, ϵ be independent standard normals, $U = A + \epsilon$ and $B = A + \delta$. The features A and B are associated, and B is also associated with U , but $\mathbb{E}[U | A, B] = \mathbb{E}[U | A] = A$. Thus a strong marginal association can coexist with zero incremental prediction value. A set-level p -value is not, by itself, evidence of an improvement in breeding-value prediction or selection response.

Donoho, David, and Jiashun Jin. 2004. "Higher Criticism for Detecting Sparse Heterogeneous Mixtures." *The Annals of Statistics* 32 (3): 962–94. <https://doi.org/10.1214/009053604000000265>.

Elkan, Charles, and Keith Noto. 2008. "Learning Classifiers from Only Positive and Unlabeled Data." *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 213–20. <https://doi.org/10.1145/1401890.1401920>.

Ferraccioli, Federico, Laura M. Sangalli, and Livio Finos. 2022. "Some First Inferential Tools for Spatial Regression with Differential Regularization." *Journal of Multivariate Analysis* 189: 104866. <https://doi.org/10.1016/j.jmva.2021.104866>.

Goeman, Jelle J., and Peter Bühlmann. 2007. "Analyzing Gene Expression Data in Terms of Gene Sets: Methodological Issues." *Bioinformatics* 23 (8): 980–87. <https://doi.org/10.1093/bioinformatics/btm051>.

Hemerik, Jesse, and Jelle Goeman. 2018. "Exact Testing with Random Permutations." *TEST* 27 (4): 811–25. <https://doi.org/10.1007/s11749-017-0571-1>.

Horn, Roger A., and Charles R. Johnson. 2013. *Matrix Analysis*. Second. Cambridge University Press. <https://doi.org/10.1017/CBO9781139020411>.

Kang, Hyun Min, Noah A Zaitlen, Claire M Wade, et al. 2008. "Efficient Control of Population Structure in Model Organism Association Mapping." *Genetics* 178 (3): 1709–23. <https://doi.org/10.1534/genetics.107.080101>.

- Leeuw, Christiaan A. de, Joris M. Mooij, Tom Heskes, and Danielle Posthuma. 2015. "MAGMA: Generalized Gene-Set Analysis of GWAS Data." *PLOS Computational Biology* 11 (4): e1004219. <https://doi.org/10.1371/journal.pcbi.1004219>.
- Patterson, H. D., and R. Thompson. 1971. "Recovery of Inter-Block Information When Block Sizes Are Unequal." *Biometrika* 58 (3): 545–54. <https://doi.org/10.1093/biomet/58.3.545>.
- Schaefer, Robert J., Jean-Michel Michno, Joseph Jeffers, et al. 2018. "Integrating Coexpression Networks with GWAS to Prioritize Causal Genes in Maize." *The Plant Cell* 30 (12): 2922–42. <https://doi.org/10.1105/tpc.18.00299>.
- Winkler, Anderson M., Gerard R. Ridgway, Matthew A. Webster, Stephen M. Smith, and Thomas E. Nichols. 2014. "Permutation Inference for the General Linear Model." *NeuroImage* 92: 381–97. <https://doi.org/10.1016/j.neuroimage.2014.01.060>.
- Yu, Jianming, Gael Pressoir, William H Briggs, et al. 2006. "A Unified Mixed-Model Method for Association Mapping That Accounts for Multiple Levels of Relatedness." *Nature Genetics* 38 (2): 203–8. <https://doi.org/10.1038/ng1702>.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [conditionalrates.tex](#)
- [surrogaterates.tex](#)