

# Quadratic Unconstrained Binary Optimization With Simulated Annealing for Combinatorial Genomic Feature Selection in Gynecologic Cancer

Robert Smith <sup>1</sup>, Stephen Shang <sup>2</sup>, Frank D. Cirisano <sup>3</sup>

Review began 09/27/2026

Review ended 10/15/2026

Published 10/19/2026

© Copyright 2026

Smith et al. This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

DOI: 10.7759/cureus.118168

1. Medical School, Lake Erie College of Osteopathic Medicine, Erie, USA 2. Research, South Florida Proton Therapy Institute, Delray Beach, USA 3. Gynecologic Oncology, Women's Cancer Care Center, Delray Beach, USA

**Corresponding author:** Robert Smith, smithr21.hsc@gmail.com

## Abstract

### Background

Combinations of somatic alterations can carry information that is not evident from any one variant, and a discovery procedure is more useful if it recovers relevant features without flooding later validation with large candidate sets. We evaluated a pairwise-aware Quadratic Unconstrained Binary Optimization (QUBO) formulation, solved by classical simulated annealing, for genomic feature selection in gynecologic cancer.

### Methodology

The retrospective cohort comprised 183 patients with somatic tumor profiling, represented by 34 gene-by-variant classification indicators. High tumor grade was available for 174 patients, including 101 with grade 3 tumors. QUBO was compared with LASSO, Elastic Net, and XGBoost. On the clinical outcome, each method's selected feature count was compared with the distribution obtained from 100 label permutations, with p-values reported using the add-one Monte Carlo convention. Because the clinical cohort has no independently established true feature set, we planted a two-feature joint-risk signal on a modified copy of the clinical matrix and scored recovery of the two designated columns by recall, precision, and F1 over 50 replications per signal setting. We also recorded how often those designated columns were still selected when the planted combination carried no extra risk (target-wise null selection rate). Larger datasets were generated by resampling the original matrix.

### Results

QUBO achieved higher feature-level F1 than LASSO in each of the four tabulated synthetic settings: 0.40 versus 0.24, 0.64 versus 0.41, 0.76 versus 0.65, and 0.79 versus 0.62. Under the null construction, the target-wise selection rate was 0.10 for QUBO compared with 0.57 for LASSO, 0.68 for Elastic Net, and 0.75 for XGBoost, with mean selected feature counts of 2.0 versus 15.4, 14.5, and 14.1. In the clinical application, QUBO selected nine features compared with a permutation mean of 1.4 ( $p = 0.010$ ), including features in *ARID1A*, *PTEN*, *PIK3CA*, *TP53*, and *KRAS*. A co-occurrence support rule expressed as a fraction of cohort size, rather than a fixed count of 10, was required to preserve this compact null behavior at resampled sizes up to 1,000.

### Conclusions

Under the tested configurations, QUBO recovered the two constituents of a planted joint-risk pattern with higher F1 than LASSO, selected the designated columns less often under a null outcome, and produced a nine-feature clinical set concentrated in established gynecologic cancer genes. The study is a clinically motivated methods evaluation. It does not report a validated predictive biomarker, conventional false discovery rate control, or a quantum hardware advantage.

**Categories:** Obstetrics/Gynecology, Genetics, Oncology

**Keywords:** arid1a, combinatorial feature selection, endometrial cancer, gynecologic oncology, machine learning, quantum computing, tumor genomic profiling, variant of uncertain significance

## Introduction

The modern understanding of cancer genetics rests on the description by Vogelstein and colleagues of the cancer genome as a structured accumulation of driver mutations across a small number of core pathways [1]. In this model, a typical solid tumor carries only two to eight driver mutations, each acquired over time and each giving the cell a selective growth advantage, set against a much larger background of passenger mutations that confer no such advantage [1]. A mutation is a change in DNA sequence. A driver helps the cancer grow, while passengers confer no selective growth advantage. A central difficulty in connecting a tumor's genome to its behavior is that individual alterations and their combinations can both contribute, and information carried by a combination may not be evident from considering each variant separately.

#### How to cite this article

Smith R, Shang S, Cirisano F D (October 19, 2026) Quadratic Unconstrained Binary Optimization With Simulated Annealing for Combinatorial Genomic Feature Selection in Gynecologic Cancer. Cureus 18(10): e118168. DOI 10.7759/cureus.118168

The combinatorial character of cancer motivates attention to pathways in gynecologic malignancies. In endometrial carcinoma, recurring drivers include *PTEN*, *ARID1A*, *PIK3CA*, and *TP53*, involving PI3K signaling, chromatin remodeling, and other regulatory processes [2]. *ARID1A* is a subunit of the switch/sucrose non-fermentable (SWI/SNF) complex and acts as a tumor suppressor, with frequent loss-of-function mutations in endometrioid and clear cell tumors [3–5]. Experimental work supports functional cooperation between *ARID1A* and *p53* in regulating shared targets [4]. Activating *PIK3CA* mutations and alterations of pathway regulators such as *PTEN* and *PIK3R1* provide additional biological context. *PIK3CA* mutations have been studied for prognostic relevance in endometrial cancer; the cited meta-analysis reported a non-significant tendency toward poorer survival (relative risk (RR) = 1.28, 95% confidence interval (CI) = 0.84–1.94,  $p = 0.25$ ) rather than a statistically significant pooled association [6]. High-grade serous ovarian tumors are characterized by near-universal *TP53* mutation and frequent DNA repair abnormalities [7]. These recurring genes and pathways motivate examination of joint genomic information.

The inherited background adds another layer. Pathogenic germline variants in DNA repair genes can shape tumor development and treatment vulnerability. *BRCA1* and *BRCA2* are examples: *BRCA*-associated repair deficiency can create selective vulnerability to PARP inhibition through synthetic lethality [8]. Tumors without a *BRCA* alteration that share aspects of this repair-deficient phenotype have been described as showing BRCAness [9], a context relevant to treatment response and resistance [10]. The 2020 American Society of Clinical Oncology (ASCO) guideline recommends germline testing in epithelial ovarian cancer and specifies indications for additional somatic testing [11]. Variants are classified according to supporting evidence, from benign and likely benign through uncertain significance to likely pathogenic and pathogenic. The present feature selection analysis concerns somatic alterations.

Variants of uncertain significance (VUS) provide an important motivation for this work. A VUS lacks sufficient evidence for a definitive classification, which is different from evidence that it has no biological effect. Clinical decisions should not be based on a VUS [11]. For research, however, a VUS-classified alteration alongside other alterations may be considered as a candidate source of information about tumor behavior. The classical two-hit model illustrates the importance of combinations affecting tumor suppressor function [12]. This motivates investigation of joint genomic information without presupposing the function of any particular VUS combination. Feature selection can prioritize candidates for subsequent variant-level and functional evaluation.

Standard predictive models and combinatorial selectors address this selection problem differently. LASSO and Elastic Net estimate coefficients jointly; with main-effects inputs, their linear predictors are additive on the log-odds scale. Explicit product terms allow regression models to represent interactions, and tree-based models can capture conditional relationships. The number of possible combinations nevertheless grows rapidly, and the selected set depends on the scoring function, tuning objective, and retention rule. In a discovery setting, every candidate may require follow-up validation. A method that recovers relevant features while retaining a manageable candidate set can therefore have practical value separate from its predictive accuracy.

Combinatorial optimization offers a way to incorporate these priorities directly into the selection objective. In Quadratic Unconstrained Binary Optimization (QUBO), each feature is a binary choice: included in a candidate set or left out. The formulation evaluated here combines feature outcome association scores, a penalty for inclusion, and rewards for eligible pairwise co-occurrence signals. The objective therefore incorporates both individual and joint information while searching over feature subsets. QUBO is also the native problem format for quantum annealers [13–15]. In this study, we did not use quantum hardware. We used classical simulated annealing on an ordinary computer, so that what we measure is the behavior of the implemented selection procedure. Establishing that behavior on classical hardware is the necessary step before asking whether quantum annealing adds anything on top of it.

We approached this as a clinical genomics application with a controlled feature recovery benchmark. We first applied QUBO, LASSO, Elastic Net, and gradient-boosted trees to our cohort of 183 patients, comparing selected feature counts under the observed high-grade outcome with counts under outcome permutation. Because the clinical cohort has no independently established true feature set, we also planted a two-feature joint-risk signal in a modified version of the clinical matrix and measured recovery of its constituent features. The benchmark was evaluated at the original cohort size and at larger simulated dataset sizes. The clinical analysis provides disease context; the synthetic experiments provide a known-target comparison of feature selection performance.

## Materials And Methods

### Study design and cohort

This was a retrospective study of 183 patients with gynecologic malignancies treated at a single gynecologic oncology practice. Every patient had undergone somatic tumor profiling on the CARIS molecular intelligence panel as part of routine clinical care, and for each patient, we collected the genes reported as altered together with the classification of each variant as pathogenic, likely pathogenic, or a VUS. Patients were included when their profiling report contained tumor DNA sequencing results with variant-level

interpretations. Older reports that predated next-generation sequencing, and reports on which testing could not be completed for the submitted specimen, carried no variant classifications to encode and were excluded. Of the 220 patients with available profiling reports, 183 met these criteria. We also collected the clinical variables recorded for each patient, including primary tumor site, histology, stage, tumor grade, recurrence, progression, age, residual disease, and lymph node status. A subset of 40 patients also had germline testing available. All clinical and genomic data were de-identified before analysis. The study was reviewed by the BRANY Institutional Review Board and received exempt status.

## Feature matrix construction

We encoded the data as a binary matrix with one patient per row and one gene-by-variant classification indicator per feature column. For example, pathogenic PTEN, likely pathogenic PTEN, and PTEN-VUS were separate columns. We retained features recorded in at least nine patients, leaving 34 somatic indicators: 26 pathogenic or likely pathogenic and eight VUS-classified indicators. The full data matrix contained 148 columns, including clinical variables, outcomes, and precomputed pairwise co-mutation columns. The four primary selection methods operated on the 34 somatic indicators only; the interaction-augmented LASSO additionally used constructed pairwise products. A feature represents a gene and classification category, not necessarily one specific nucleotide or protein change. Accordingly, ARID1A-VUS denotes the classification-specific feature throughout.

## Computational environment

All analyses were run in Python and JupyterLab using Anaconda. We used scikit-learn for regularized logistic regression [16], XGBoost for gradient-boosted trees [17], and the D-Wave Ocean framework for QUBO construction and classical simulated annealing, using dimod and the dwave-samplers simulated annealing sampler [18]. Statistical testing used SciPy and statsmodels. Library versions were recorded at runtime and random seeds were fixed throughout. Identical solutions across solver seeds demonstrate repeatability of the tested annealing runs; they do not by themselves establish stability under resampling of patients.

## Feature selection methods

We compared four primary feature selection procedures: QUBO, LASSO logistic regression, Elastic Net logistic regression, and XGBoost. Each returned a set of retained features. An interaction-augmented LASSO provided an additional synthetic benchmark comparator.

The QUBO model represented each candidate feature by a binary inclusion variable. Its linear cost was a fixed penalty minus the strength of the feature's correlation with the outcome; the penalty was set to two divided by the square root of the cohort size. For eligible feature pairs, we assessed whether their co-occurrence tracked the outcome more strongly than either feature alone and rewarded joint selection in proportion to that excess association. Pair eligibility used a minimum co-occurrence count of approximately 5.5% of the cohort, corresponding to 10 patients at  $n = 183$ . This association-based score incorporated pairwise information into selection rather than estimating an interaction coefficient in a fitted outcome model. We used a simulated annealing sampler with 100 reads and selected the features in its lowest observed energy solution. Optimization was repeated across five random seeds to examine solver repeatability. The cohort-relative threshold arose from the development experiment described below.

LASSO and Elastic Net fitted penalized logistic regressions. We used five-fold cross-validation to select penalty strength by maximizing the area under the receiver operating characteristic curve (ROC-AUC), and retained features with nonzero coefficients. LASSO used an L1 penalty, while Elastic Net used an equal mix of L1 and L2 penalties.

XGBoost is a gradient-boosted tree classifier. We used 200 trees with a maximum depth of three and a learning rate of 0.1, and we kept the features whose importance was above the average across all features.

As an additional comparator, we constructed an interaction-augmented LASSO. We added a product term for each pair of features co-occurring in at least 10 patients, fitted the cross-validated LASSO, and mapped selected product terms back to their constituent genes. This provided regression with explicit pair terms alongside the main effects. It does not automatically equate penalty selection, candidate-set size, or scoring with the QUBO configuration.

## Application to study cohort

The clinical outcome was recorded as high tumor grade, defined as grade 3 versus grades 1-2. Grade was available for 174 of the 183 patients, including 101 with grade 3 tumors. Molecular and histologic context can be relevant to grading [2,19]; the present analysis pooled the recorded endpoint across tumor sites. For each primary method, we recorded the number of selected features under the observed outcome and under 100 randomly shuffled outcomes. Let  $b$  be the number of shuffled runs with at least as many selected features as the observed run. We report the add-one Monte Carlo P-value  $(b + 1)/(100 + 1)$  [20]. This assesses whether the observed selected feature count is extreme relative to that method's own permutation distribution. It is

not a test of predictive accuracy and does not establish that any particular selected gene is a causal driver. We also described co-occurring pathogenic alterations among carriers of the selected ARID1A-VUS feature.

## Synthetic plant and recover benchmark

To measure feature recovery against a known target, we generated synthetic outcomes on a modified version of the clinical feature matrix. In each run, two VUS-classified columns were chosen at random and their existing values replaced. At  $n = 183$ , the construction assigned 20 joint carriers and 60 single-only carriers for each target. Joint carriers received an elevated event probability, termed signal strength, while all other patients had a baseline probability of 0.15. We used 50 replications per signal setting. Because each target column includes the elevated-risk joint-carrier group, each constituent feature carries residual marginal information. Recovery therefore concerns selection of the two designated columns of a planted joint-risk pattern, summarized by recall, precision, and F1. It does not require selection of an explicit interaction term, and it is not an interaction-only recovery experiment.

For the null experiment, the joint-carrier event probability was set equal to the baseline of 0.15 and selection of the two designated columns was recorded. We report this quantity as a target-wise null selection rate: the frequency with which the two designated columns were selected when they conferred no extra risk. This is distinct from the false-discovery rate in the Benjamini-Hochberg sense, which is the expected proportion of false selections among all selected features [21]. A supplementary comparison varied LASSO penalty strength to match QUBO's target-wise null selection rate and compared feature recovery at that matched setting. Mean selected feature counts under the null were summarized over 30 replications.

## Behavior across simulated dataset sizes

To examine behavior across dataset sizes, we resampled rows of the original feature matrix with replacement to  $n = 183, 300, 600$ , and  $1,000$ , using 20 replications at each size. This samples the empirical background distribution; frequencies can vary between resamples. We then rebuilt the target columns and generated outcomes as before, scaling joint-carrier and single-only-carrier counts with  $n$ . Rebuilding those columns changes their original prevalences and relationships with other features. These experiments used synthetic outcomes only and were separate from the clinical outcome analysis. Larger  $n$  increased simulated observations rather than the number of independently recruited patients or candidate feature columns.

## Co-occurrence threshold

During the dataset-size experiments, a fixed minimum joint-carrier count of 10 became permissive at larger  $n$ . At  $n = 1,000$ , this configuration produced a target-wise null selection rate of 0.95 and selected 28 of the 34 features under the null. We therefore expressed support as approximately 5.5% of  $n$ , corresponding to 10 at  $n = 183$ . The revised rule left the  $n = 183$  results unchanged and improved the reported null selection behavior at larger sizes. The threshold was revised after inspecting benchmark behavior, making this a method-development experiment rather than independent validation of a prespecified rule.

## Statistical analysis

Differences between methods on synthetic data were assessed using Wilcoxon signed-rank tests across replications. Matched and dataset size comparisons used paired permutation tests. Benjamini-Hochberg adjustment was applied within families of comparisons, with adjusted values reported as q-values [21]. This adjustment applies to between-method tests and is distinct from the error rate of selected features. Clinical p-values are within-method selected-count tests as defined above.

# Results

## Patient population and cohort characteristics

The cohort comprised 183 patients with gynecologic malignancies. The primary tumor sites were endometrium (116 patients, 63.4%), ovary (43 patients, 23.5%), fallopian tube (seven patients, 3.8%), cervix (six patients, 3.3%), tubo-ovarian (four patients, 2.2%), uterus (four patients, 2.2%), and other sites (three patients, 1.6%). Among the 174 patients with staging data, 69 (39.7%) had advanced-stage disease (International Federation of Gynecology and Obstetrics (FIGO) III or IV). Among the 174 with grade data, 101 (58.0%) had high-grade (grade 3) tumors. Recurrence was documented in 45 of 175 evaluable patients (25.7%) and progression in 46 of 175 (26.3%). Germline testing was available for 40 (21.9%) patients. Somatic profiling contributed 1,449 calls across 191 unique genes.

## Feature selection in the study cohort

For the high-grade outcome, QUBO selected nine features compared with a mean of 1.4 under label permutation ( $p = 0.010$ ). LASSO selected six features (permutation mean 14.4,  $p = 0.475$ ), Elastic Net three (18.6,  $p = 0.723$ ), and XGBoost 11 (14.6,  $p = 0.980$ ) (Table 1). QUBO was the only procedure whose observed count exceeded its permutation mean by a large margin. The comparator p-values indicate that those observed counts were not extreme relative to each method's own permutation distribution. Because LASSO

and Elastic Net selected fewer features on the observed outcome than under permutation, a count-based upper-tail test is not a matched comparison of predictive signal among methods. Table 1 is therefore a within-method sparsity contrast, not evidence that the regression methods found no structure in the cohort.

| Method      | Selected | Mean permuted selections | P-value |
|-------------|----------|--------------------------|---------|
| QUBO        | 9        | 1.4                      | 0.010   |
| LASSO       | 6        | 14.4                     | 0.475   |
| Elastic Net | 3        | 18.6                     | 0.723   |
| XGBoost     | 11       | 14.6                     | 0.980   |

TABLE 1: Selected feature counts and permutation summaries for high grade (101 grade 3 outcomes among 174 patients).

The test statistic is the number of selected features.  $P = (b + 1)/(100 + 1)$ , where b is the number of label-shuffled runs with a selected count at least as large as the observed count [20]. These within-method tests do not compare predictive performance.

Biological interpretation of the selected feature set

The nine selected gene-by-classification features included genes already implicated in gynecologic cancer [2]. PTEN, PIK3CA, and PIK3R1 participate in PI3K signaling, providing a biologically recognizable grouping [2,6]. ARID1A represents the SWI/SNF chromatin-remodeling complex and has a tumor-suppressor role in endometrioid and clear cell disease [3,4]. TP53 is prominent in high-grade serous disease [2,7], while KRAS, CTNNB1, and FGFR2 are recurrently altered endometrial cancer genes [2].

The selected set therefore has face validity on a prevalence-filtered cancer panel. We interpret this as descriptive biological coherence rather than a new driver discovery, a formal pathway-enrichment result, or proof that QUBO outperforms the other methods on real outcomes. The direction and independence of each selected association, and any interaction between selected features, require separate evaluation.

The selected ARID1A-VUS feature and co-occurring alterations

One of the nine selected features was ARID1A-VUS; the other eight were pathogenic or likely pathogenic feature columns. The reported unadjusted correlation of the ARID1A-VUS feature with high grade was -0.219. All 14 carriers had a pathogenic alteration somewhere in the profiled genes, and 11 also had a pathogenic ARID1A alteration. Another 153 patients had a pathogenic alteration without the ARID1A-VUS feature (Table 2). Thus, the VUS indicator and the indicator for any pathogenic alteration are not identical: the 153 pathogenic-without-VUS patients show that the two variables can disagree. What the cohort does not provide is an isolated VUS contrast. These counts describe the molecular backgrounds in which the VUS-classified feature occurred; an adjusted association or variant-specific functional effect was not established.

| Quantity                                                             | Value  |
|----------------------------------------------------------------------|--------|
| Patients carrying the ARID1A-VUS feature                             | 14     |
| Carriers without any co-occurring pathogenic mutation                | 0      |
| Carriers with a co-occurring pathogenic ARID1A mutation              | 11     |
| Patients with a pathogenic alteration but not the ARID1A-VUS feature | 153    |
| Correlation of the ARID1A-VUS feature with high grade                | -0.219 |

TABLE 2: Descriptive co-occurrence counts for the selected ARID1A-VUS feature.

Counts are cohort-level summaries and are descriptive; no hypothesis test is applied. The reported correlation is unadjusted.

VUS = variants of uncertain significance

Feature recovery in synthetic data

The synthetic benchmark provided a known-target comparison of feature recovery. The principal comparative result is feature-level F1. QUBO achieved higher F1 than LASSO in each of the four tabulated settings: 0.40 versus 0.24 (weak), 0.64 versus 0.41 (moderate), 0.76 versus 0.65 (strong), and 0.79 versus 0.62 (very strong) (Table 3). All four contrasts were significant by Wilcoxon signed-rank test after Benjamini-Hochberg adjustment (W = 190.0, 121.5, 141.5, and 96.0, respectively; all q ≤ 0.0052) (Table 3). LASSO was the strongest of the three primary standard methods; Elastic Net and XGBoost recovered less than LASSO at every tabulated signal strength.

| Signal strength | QUBO F1 | LASSO F1 | W     | P-value | q      |
|-----------------|---------|----------|-------|---------|--------|
| Weak            | 0.40    | 0.24     | 190.0 | 0.0052  | 0.0052 |
| Moderate        | 0.64    | 0.41     | 121.5 | 0.0002  | 0.0002 |
| Strong          | 0.76    | 0.65     | 141.5 | 0.0044  | 0.0047 |
| Very strong     | 0.79    | 0.62     | 96.0  | 0.0001  | 0.0002 |

TABLE 3: Feature-level F1 for the two planted columns: QUBO versus LASSO by signal setting.

Weak, moderate, strong, and very strong are the reported scenario labels. F1 summarizes the recovery of the two designated columns. LASSO is shown as the strongest primary comparator. W is the Wilcoxon signed-rank statistic across 50 paired replications per setting; q-values are Benjamini-Hochberg adjusted within the family of between-method comparisons [21].

QUBO = Quadratic Unconstrained Binary Optimization

Separately, at n = 183, the target-wise null selection rate was 0.10 for QUBO, compared with 0.57 for LASSO, 0.68 for Elastic Net, and 0.75 for XGBoost (Table 4). Mean selected feature counts in the same table were 2.0, 15.4, 14.5, and 14.1, respectively. The tested QUBO configuration therefore combined higher tabulated F1 with both a lower target-wise null selection rate and a smaller reported mean candidate set. The null-selection contrasts against QUBO were reported with mean paired differences of -0.47, -0.58, and -0.65, and Benjamini-Hochberg adjusted q-values of 0.0016, 0.0011, and 0.0002 for LASSO, Elastic Net, and XGBoost, subject to the metric definition above. These two Table 4 columns are reported as implemented; they should not be read as a conventional false-discovery rate for the entire selected set.

| Method      | Target-wise null selection rate | Mean features selected | Mean paired difference | P-value | q      |
|-------------|---------------------------------|------------------------|------------------------|---------|--------|
| QUBO        | 0.10                            | 2.0                    | —                      | —       | —      |
| LASSO       | 0.57                            | 15.4                   | -0.47                  | 0.0016  | 0.0016 |
| Elastic Net | 0.68                            | 14.5                   | -0.58                  | 0.0007  | 0.0011 |
| XGBoost     | 0.75                            | 14.1                   | -0.65                  | <0.0001 | 0.0002 |

TABLE 4: Target-wise null selection rate and reported selection counts at n = 183.

The null selection rate is the selection frequency of the two designated columns when the joint-carrier event probability equals the baseline of 0.15. It is not a Benjamini-Hochberg false-discovery rate for the entire selected set [21]. Null selection rates are means over 20 paired replications of the dataset-size experiment at n = 183; mean features selected are means over 30 replications of the planted benchmark. The test statistic is the mean paired difference in null selection rate across those same 20 replications, computed as QUBO minus the comparator, assessed by a paired permutation test with Benjamini-Hochberg adjustment across the three comparisons. Dashes mark QUBO as the reference condition for those comparisons.

QUBO = Quadratic Unconstrained Binary Optimization

In the supplementary comparison matching LASSO to QUBO’s target-wise null selection rate, higher QUBO recovery was still reported, with q ≤ 0.0005. The interaction-augmented LASSO comparison was also reported as favoring QUBO on null selection and strong signal recovery. These results argue against explaining the tabulated F1 advantage solely as a side effect of selecting fewer features, or solely as a side effect of denying regression access to pair terms. They do not fully isolate pair scoring, sparsity, and solver choice as separate mechanisms, because complete matched estimates and comparison tables are not presented here. The QUBO solution was identical across five random seeds, demonstrating repeatability of the tested solver runs.

Behavior across simulated dataset sizes

QUBO’s target-wise null selection rate was 0.10 at n = 183 and 0.05 at n = 300, 600, and 1,000. Comparator values ranged from 0.38 to 0.75 across the displayed sizes and varied non-monotonically (Table 5). The dataset-size comparisons were reported as favoring QUBO, with the narrowest corrected contrast against Elastic Net at n = 1,000 (q = 0.0395). These findings describe the specified resampling and planting procedure. They are not a clinical sample size calculation and do not show that conventional methods are incapable of improving with additional independent patients.

| Method      | n = 183 | n = 300           | n = 600           | n = 1,000         |
|-------------|---------|-------------------|-------------------|-------------------|
| QUBO        | 0.10    | 0.05              | 0.05              | 0.05              |
| LASSO       | 0.57    | 0.38 (q = 0.0138) | 0.55 (q = 0.0012) | 0.50 (q = 0.0013) |
| Elastic Net | 0.68    | 0.40 (q = 0.0185) | 0.57 (q = 0.0012) | 0.57 (q = 0.0009) |
| XGBoost     | 0.75    | 0.70 (q = 0.0001) | 0.65 (q = 0.0002) | 0.72 (q = 0.0001) |

**TABLE 5: Target-wise null selection rate by method and simulated dataset size.**

Datasets are with-replacement resamples of the original 183-patient matrix with rebuilt target columns. The null selection rate has the definition given in Table 4. Each q-value is from a paired permutation test of that method against QUBO at that dataset size, with the mean paired difference in null selection rate as the test statistic and a Benjamini-Hochberg adjustment across the dataset-size comparisons. Contrasts at n = 183 are given in Table 4.

QUBO recall was reported to rise from 0.78 at the original dataset size toward 1.0 in the larger size experiments. Recall is not tabulated by signal setting and dataset size, so this summary should not be read as a uniform recovery trend across every condition.

The threshold experiment identified a marked configuration sensitivity. At n = 1,000, the fixed-count rule produced a target-wise null selection rate of 0.95 and selected 28 of 34 features under the null. With the cohort-relative rule, the reported null selection rate was approximately 0.05 at the larger sizes, while the n = 183 results were unchanged. Pair eligibility is therefore a consequential part of the implemented selection procedure, not a neutral preprocessing choice.

Discussion

Principal findings

The principal comparative finding is higher QUBO feature-level F1 than LASSO in each of the four tabulated synthetic settings. That recovery advantage is accompanied by a lower target-wise null selection rate and a smaller reported mean selected set under the tested configurations. In the clinical application, QUBO selected nine gene-by-classification features containing established gynecologic cancer genes [2,7], and that count was extreme relative to QUBO’s own permutation distribution (9 versus 1.4, p = 0.010). A further practical finding was the sensitivity of null behavior to the co-occurrence threshold as simulated dataset size increased. Together, these results provide benchmark evidence and a disease-motivated application of pairwise-aware genomic feature selection.

Biological interpretation

The clinical feature set spans PI3K signaling, chromatin remodeling, and other processes already implicated in gynecologic cancer [2-5,7]. Its value here is a biologically interpretable set of candidates generated from a real profiling cohort. This coherence is descriptive: the input was itself a prevalence-filtered cancer panel, and no formal pathway-enrichment analysis or equivalent biological comparison of all methods was reported. The clinical application and synthetic benchmark make different contributions. The former identifies candidates in a relevant disease setting; the latter evaluates recovery against a known target.

Why feature recovery and candidate burden matter

A genomic selection procedure is useful when it recovers relevant features without generating an unnecessarily large set for validation. The higher tabulated F1 values indicate favorable target recovery as summarized by that metric, while the separately reported compact selections are relevant to the intended candidate prioritization task. The lower target-wise null selection rates describe another aspect of the tested procedure’s behavior: under a null outcome, the designated columns were selected less often. Together, these observations support a favorable recovery candidate burden profile under the tested generator and configurations. They do not demonstrate a reduction in clinical false discoveries or validation costs, and they are not conventional false-discovery rate control [21].

## The co-occurrence threshold as a scaling caveat

The co-occurrence threshold is a practical lesson from the development experiments. A fixed count represented a progressively smaller fraction of the simulated dataset, whereas the cohort-relative rule retained a comparable support requirement as  $n$  increased. The improvement in the reported null selection rate shows why pair eligibility should be treated as part of the method configuration. Because the rule was refined after observing benchmark behavior, the favorable fractional setting calls for independent assessment. The 5.5% threshold is a tested development choice, not a universally optimal value; any support criterion also limits access to sufficiently rare pairs.

## The VUS question and the limits of this cohort

The selection of ARID1A-VUS preserves the motivating question of whether uncertain alterations can contribute information in a broader genomic context. All 14 carriers also had a pathogenic alteration, but the converse was not true. This pattern is consistent with the two-hit model of tumor-suppressor inactivation [12], and guidelines are explicit that a variant of uncertain significance should not drive clinical decisions [11]. Zero isolated-VUS carriers restrict a VUS-only contrast; it does not prove that every adjusted association is unidentifiable. Estimability depends on the complete outcome and covariate structure, and no such adjusted analysis is reported here. The finding identifies a gene-by-classification feature for further investigation, while variant-level annotation, informative molecular backgrounds, and functional evidence would address the biological meaning of the underlying sequence variants.

## Limitations

The clinical analysis is a retrospective single-practice series with grade available for 174 of 183 patients. Grade was pooled across tumor sites and histologies, and independence from site, histology, or molecular subtype was not established [19]. No held-out predictive evaluation was reported; the clinical results are candidate selections rather than a validated prediction rule.

The benchmark concerns two reconstructed targets among 34 prevalence-filtered features and contains both joint-risk and residual marginal signal. Resampling adds simulated observations, not independent source patients, and rebuilding the targets changes their original VUS prevalences. The findings therefore apply to the tested generator and selection configurations, whose tuning objectives and retention rules differ. The comparison does not isolate the contribution of pair scoring, sparsity, or solver choice. Repeatability across solver seeds addresses a different source of variability from patient-sampling stability.

## Future directions

Broader evaluation should retain the distinction between feature recovery, interaction detection, and clinical prediction. Prespecified thresholds, explicit recovery endpoints, and comparator calibration matched to the selection objective would strengthen methodological comparisons. Interaction-only generators and exact-pair recovery would address stronger interaction claims than those supported here. Independent multi-site data could evaluate clinical transportability, while variant-level annotation and functional studies would address VUS biology. Clinical sample-size planning should specify the intended effect, joint-carrier prevalence, covariates, and error criterion rather than extrapolating directly from the present resampling experiment. Clinical and imaging characteristics could also be integrated with molecular features in future multimodal models. Endometrial thickness and uterine artery Doppler parameters have been evaluated as predictors in endometrial cancer [22], and imaging contributes to the assessment of uterine sarcoma [23]. Additional data types, including tumor mutational burden and copy number, and comparisons among classical solvers are further directions; quantum hardware is a separate subsequent question.

## Conclusions

QUBO-based combinatorial feature selection achieved higher feature-level F1 than LASSO in the four tabulated synthetic scenarios, selected the designated columns less often under a null outcome, and selected a nine-feature set containing established gynecologic cancer genes in the clinical application. The dataset-size experiments further identified co-occurrence support as an important determinant of null-selection behavior. The study provides a clinically motivated implementation and benchmark evidence for pairwise-aware genomic candidate selection. It claims no quantum-hardware advantage and no clinically validated biomarker.

## Additional Information

### Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

**Concept and design:** Robert Smith, Frank D. Cirisano, Stephen Shang

**Acquisition, analysis, or interpretation of data:** Robert Smith, Stephen Shang

**Drafting of the manuscript:** Robert Smith

**Critical review of the manuscript for important intellectual content:** Robert Smith, Frank D. Cirisano, Stephen Shang

**Supervision:** Frank D. Cirisano

## Disclosures

**Human subjects:** Informed consent for treatment and open access publication was obtained or waived by all participants in this study. Biomedical Research Alliance of New York (BRANY) Institutional Review Board issued approval 26-12-269-2404. This retrospective analysis was conducted under BRANY IRB File #26-12-269-2404 (Investigator Initiated Protocol FDC2026-QA), which received an Exempt determination under Category 4(ii) of 45 CFR 46.104(d)(4)(ii). A waiver of HIPAA authorization was granted under 45 CFR 164.512(i)(2)(ii). All clinical and genomic data were de-identified prior to analysis. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

## Acknowledgements

De-identified data and analysis code are available from the authors on reasonable request, subject to institutional data-use governance. Requests should be directed to the corresponding author.

## References

1. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr, Kinzler KW: Cancer genome landscapes. *Science*. 2013, 339:1546-58. [10.1126/science.1235122](https://doi.org/10.1126/science.1235122)
2. Kandoth C, Schultz N, Cherniack AD, et al.: Integrated genomic characterization of endometrial carcinoma. *Nature*. 2013, 497:67-73. [10.1038/nature12113](https://doi.org/10.1038/nature12113)
3. Wiegand KC, Shah SP, Al-Agha OM, et al.: ARID1A mutations in endometriosis-associated ovarian carcinomas. *N Engl J Med*. 2010, 363:1532-43. [10.1056/NEJMoa1008433](https://doi.org/10.1056/NEJMoa1008433)
4. Guan B, Wang TL, Shih IeM: ARID1A, a factor that promotes formation of SWI/SNF-mediated chromatin remodeling, is a tumor suppressor in gynecologic cancers. *Cancer Res*. 2011, 71:6718-27. [10.1158/0008-5472.CAN-11-1562](https://doi.org/10.1158/0008-5472.CAN-11-1562)
5. Wu RC, Wang TL, Shih IeM: The emerging roles of ARID1A in tumor suppression. *Cancer Biol Ther*. 2014, 15:655-64. [10.4161/cbt.28411](https://doi.org/10.4161/cbt.28411)
6. Bredin HK, Krakstad C, Hoivik EA: PIK3CA mutations and their impact on survival outcomes of patients with endometrial cancer: a systematic review and meta-analysis. *PLoS One*. 2023, 18:e0283203. [10.1371/journal.pone.0283203](https://doi.org/10.1371/journal.pone.0283203)
7. Integrated genomic analyses of ovarian carcinoma. *Nature*. 2011, 474:609-15. [10.1038/nature10166](https://doi.org/10.1038/nature10166)
8. Farmer H, McCabe N, Lord CJ, et al.: Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature*. 2005, 434:917-21. [10.1038/nature03445](https://doi.org/10.1038/nature03445)
9. Turner N, Tutt A, Ashworth A: Hallmarks of 'BRCAness' in sporadic cancers. *Nat Rev Cancer*. 2004, 4:814-9. [10.1038/nrc1457](https://doi.org/10.1038/nrc1457)
10. Lord CJ, Ashworth A: Mechanisms of resistance to therapies targeting BRCA-mutant cancers. *Nat Med*. 2013, 19:1381-8. [10.1038/nm.3369](https://doi.org/10.1038/nm.3369)
11. Konstantinopoulos PA, Norquist B, Lacchetti C, et al.: Germline and somatic tumor testing in epithelial ovarian cancer: ASCO guideline. *J Clin Oncol*. 2020, 38:1222-45. [10.1200/JCO.19.02960](https://doi.org/10.1200/JCO.19.02960)
12. Knudson AG Jr: Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A*. 1971, 68:820-3. [10.1073/pnas.68.4.820](https://doi.org/10.1073/pnas.68.4.820)
13. Nałęcz-Charkiewicz K, Charkiewicz K, Nowak RM: Quantum computing in bioinformatics: a systematic review mapping. *Brief Bioinform*. 2024, 25:bbae391. [10.1093/bib/bbae391](https://doi.org/10.1093/bib/bbae391)
14. Willsch D, Willsch M, De Raedt H, Michielsen K: Support vector machines on the D-Wave quantum annealer. *Comput Phys Commun*. 2020, 248:107006. [10.1016/j.cpc.2019.107006](https://doi.org/10.1016/j.cpc.2019.107006)
15. Li RY, Di Felice R, Rohs R, Lidar DA: Quantum annealing versus classical machine learning applied to a simplified computational biology problem. *npj Quantum Inf*. 2018, 4:14. [10.1038/s41534-018-0060-8](https://doi.org/10.1038/s41534-018-0060-8)
16. Pedregosa F, Varoquaux G, Gramfort A, et al.: Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011, 12:2825-30.
17. Chen T, Guestrin C: XGBoost: a scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016, 785-94. [10.1145/2939672.2939785](https://doi.org/10.1145/2939672.2939785)
18. D-Wave Ocean software documentation. (2026). Accessed: June 15, 2026: <https://docs.ocean.dwavesys.com>.
19. Talhouk A, McConechy MK, Leung S, et al.: A clinically applicable molecular-based classification for endometrial cancers. *Br J Cancer*. 2015, 113:299-310. [10.1038/bjc.2015.190](https://doi.org/10.1038/bjc.2015.190)
20. Phipson B, Smyth GK: Permutation P-values should never be zero: calculating exact P-values when

- permutations are randomly drawn. *Stat Appl Genet Mol Biol*. 2010, 9:Article39. [10.2202/1544-6115.1585](https://doi.org/10.2202/1544-6115.1585)
21. Benjamini Y, Hochberg Y: Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol*. 1995, 57:289-300. [10.1111/j.2517-6161.1995.tb02031.x](https://doi.org/10.1111/j.2517-6161.1995.tb02031.x)
  22. Nguyen PN, Nguyen VT: Endometrial thickness and uterine artery Doppler parameters as soft markers for prediction of endometrial cancer in postmenopausal bleeding women: a cross-sectional study at tertiary referral hospitals from Vietnam. *Obstet Gynecol Sci*. 2022, 65:430-40. [10.5468/ogs.22053](https://doi.org/10.5468/ogs.22053)
  23. Nguyen XL, Huynh QH, Nguyen PN: Assessing the clinical characteristics and the role of imaging modalities in uterine sarcoma: a single-center retrospective study from Vietnam. *J Clin Ultrasound*. 2025, 53:1527-37. [10.1002/jcu.24046](https://doi.org/10.1002/jcu.24046)