Conclusion

The presence of phenolic hydroxyl groups is strongly associated with a coordinated cellular stress response in TBBPA metabolites, with molecules containing phenolic OH groups (n=15) showing significantly elevated predicted probabilities across all five stress response pathways compared to molecules without phenolic OH groups (n=6): SR-ARE (20.7-fold, $p=7.37\times10^{-5}$), SR-MMP (55.7-fold, $p=3.69\times10^{-5}$), SR-HSE (49.2-fold, $p=3.69\times10^{-5}$), SR-p53 (97.0-fold, $p=2.58\times10^{-4}$), and SR-ATAD5 (27.8-fold, $p=4.42\times10^{-4}$), all remaining significant after Bonferroni correction ($\alpha=0.01$).

Methods

The analysis utilized the phenolic_oh_sr_are_analysis.csv artifact and smiles2_toxicity_results.csv to investigate associations between phenolic hydroxyl group presence and five stress response endpoints (SR-ARE, SR-MMP, SR-HSE, SR-p53, SR-ATAD5). Datasets were merged on SMILES identifiers using pandas (version in standard Python environment). Molecules were classified into two groups based on phenolic OH count (Present: count > 0, n=15; Absent: count = 0, n=6). For each endpoint, Mann-Whitney U tests (two-sided) were performed using scipy.stats.mannwhitneyu to compare probability distributions between groups, as this non-parametric test is appropriate for the small sample size and does not assume normal distributions. Descriptive statistics (mean, median, quartiles, min/max) were calculated using numpy. Fold-changes were computed as the ratio of mean probabilities (Present/Absent). Multiple testing correction was applied using both Bonferroni method ($\alpha=0.05/5=0.01$) and Benjamini-Hochberg FDR correction ($\alpha=0.05$) via statsmodels.stats.multitest.multipletests. Results were visualized in a grouped bar plot showing mean probabilities for each endpoint by group, with significance markers and fold-change annotations. The final comprehensive results table was saved as phenolic_oh_stress_response_analysis.csv.

Results

All 21 TBBPA metabolites were successfully analyzed, with 15 containing phenolic OH groups (12 with one OH, 3 with two OH) and 6 lacking phenolic OH groups. Mann-Whitney U tests revealed highly significant associations (all $p<0.001$) between phenolic OH presence and elevated stress response probabilities across all five endpoints:

1. **SR-ARE (Oxidative Stress)**: Mean Present=0.329, Mean Absent=0.016, Fold-change=20.7×, U=89.0, $p=7.37\times10^{-5}$
2. **SR-MMP (Mitochondrial Disruption)**: Mean Present=0.535, Mean Absent=0.010, Fold-change=55.7×, U=90.0, $p=3.69\times10^{-5}$
3. **SR-HSE (Heat Shock Response)**: Mean Present=0.165, Mean Absent=0.003, Fold-change=49.2×, U=90.0, $p=3.69\times10^{-5}$
4. **SR-p53 (DNA Damage Response)**: Mean Present=0.146, Mean Absent=0.002, Fold-change=97.0×, U=87.0, $p=2.58\times10^{-4}$
5. **SR-ATAD5 (DNA Damage Response)**: Mean Present=0.011, Mean Absent=0.0004, Fold-change=27.8×, U=86.0, $p=4.42\times10^{-4}$

All five associations remained significant after Bonferroni correction ($p<0.01$) and FDR correction ($q<0.001$). Distribution analysis revealed minimal or no overlap between groups for SR-MMP and SR-HSE (no overlap), while SR-ARE, SR-p53, and SR-ATAD5 showed slight overlap at distribution extremes. The effect was strongest for SR-p53 (97-fold increase) and SR-MMP (56-fold increase).

Challenges

The primary challenge was the small sample size (n=21 total, with unequal groups of 15 vs 6), which limits statistical power despite the highly significant findings. The unequal group sizes (15:6 ratio) could theoretically affect test sensitivity, though the Mann-Whitney U test is robust to this imbalance. The high dimensionality of the stress response probability space relative to sample size prevented multivariate analysis approaches. Three molecules in the "Present" group showed distribution overlap with the "Absent" group for certain endpoints (SR-ARE, SR-p53, SR-ATAD5), suggesting potential mechanistic heterogeneity or influence of other structural features. The absence of intermediate phenolic OH counts between the groups (no molecules with exactly 0.5 OH groups, by definition) prevented dose-response modeling. Column naming inconsistencies between datasets (_Prob vs _Probability suffix) required careful identification and correction during merging.

Discussion

This analysis provides compelling evidence that phenolic hydroxyl groups serve as a structural alert for coordinated cellular stress response in TBBPA metabolites, extending beyond the previously identified SR-ARE association to encompass mitochondrial disruption, heat shock response, and DNA damage pathways. The magnitude of associations (20-97 fold increases) and statistical significance (all $p<0.001$ after multiple testing correction) despite the small sample size (n=21) indicate a robust and biologically meaningful relationship. The coordinated elevation across multiple stress pathways suggests that phenolic OH groups may trigger a common upstream mechanism, possibly related to quinone formation, redox cycling, or direct protein modification. The strongest associations were observed for SR-p53 (97-fold) and SR-MMP (56-fold), suggesting particular relevance for DNA damage and mitochondrial toxicity pathways. Mechanistically, phenolic compounds are well-known to undergo metabolic activation to electrophilic quinones, which can form protein and DNA adducts, generate reactive oxygen species, and disrupt cellular redox homeostasis—all processes that activate the observed stress response pathways. The lack of distribution overlap for SR-MMP and SR-HSE indicates particularly clean separation between phenolic and non-phenolic metabolites for these endpoints. This holistic stress signature strengthens the case for phenolic OH as a key toxicophore in TBBPA metabolism and provides a mechanistic framework for understanding why certain transformation pathways (e.g., methylation that blocks OH groups, dimerization that eliminates phenolic structures) may reduce toxicity while others (e.g., debromination preserving OH groups, glucuronidation retaining phenolic character) maintain toxic potential.

Proposed Next Hypotheses

1. The number of phenolic hydroxyl groups (0, 1, or 2) shows a dose-dependent relationship with stress response pathway activation, with molecules containing two phenolic OH groups exhibiting even higher predicted toxicity probabilities than those with one OH group across all five stress response endpoints.
2. Metabolic transformation pathways that preserve phenolic hydroxyl groups (glucuronidation, sulfation, debromination) result in metabolites with significantly higher predicted stress response probabilities compared to pathways that eliminate or block phenolic groups (methylation, coupling/dimerization), providing a mechanistic basis for detoxification vs bioactivation classification of TBBPA metabolic routes.

Artifacts

Artifact 1:

File name: phenolic_oh_stress_response_analysis.csv

Artifact description: Comprehensive statistical analysis table containing Mann-Whitney U test results comparing five stress response endpoints (SR-ARE, SR-MMP, SR-HSE, SR-p53, SR-ATAD5) between TBBPA metabolites with and without phenolic hydroxyl groups. Includes mean and median probabilities for each group (n=15 Present, n=6 Absent), fold-changes, U statistics, uncorrected p-values, FDR-corrected p-values, and significance flags for both Bonferroni and FDR correction methods. Created through non-parametric statistical testing of merged toxicity prediction and phenolic OH classification data.