

Preparation composition confounds treatment effects in cochlear synaptosomal proteomics: a nested machine learning reanalysis of cisplatin and MnTBAP exposure

Monazza Shahab

fv9248@wayne.edu

Wayne State University

Research Article

Keywords: Computational Biology; Machine Learning; Computational Proteomics; Experimental Confounding; High-Dimensional Data Analysis; Sample Composition Heterogeneity; Nested Cross-Validation; Feature Selection Bias

Posted Date: October 20th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11340135/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Abstract

Cisplatin causes permanent hearing loss in a large proportion of treated patients, and the ribbon synapse between inner hair cells and spiral ganglion neurons is an early target of this injury. Quantitative proteomics of cochlear synaptosomes can identify proteins associated with cisplatin-induced synaptic dysfunction. Because cochlear tissue is limited, these studies typically include four or fewer animals per group, which makes both machine learning performance and differential abundance vulnerable to selection bias and to differences in the tissue content of each preparation. This study tests the hypothesis that cisplatin treatment produces a cochlear synaptosomal proteome signature that is classifiable, robust to the composition of each preparation, and reversed by co-treatment with MnTBAP, a peroxynitrite scavenger. A tandem mass tag (TMT) 16-plex dataset from four groups of mice (Control, Cisplatin, MnTBAP+Cisplatin and MnTBAP, four biological replicates each) was reanalyzed from the raw reporter ion abundances using channel-level quality control, empirical Bayes moderated statistics, nested leave-one-out classification with exact permutation tests, and projection of the held-out MnTBAP groups onto the cisplatin axis. Additionally, tissue composition was quantified with six protein marker panels and an unsupervised principal component. One channel (MnTBAP+Cisplatin replicate 1) contained 10% of the median channel signal and was excluded. Of 4,667 quantified proteins, one protein differed between Cisplatin and Control at a false discovery rate (FDR) below 0.05, and 147 differed at an FDR below 0.20, a signal well above all balanced label permutations. Nested classifiers reached an area under the ROC curve of 0.75 to 0.94, and no classifier was significant by exact permutation test ($p \geq 0.086$). Furthermore, the first principal component of the proteome (31.3% of variance) captured synaptic protein enrichment ($r = 0.91$ with a synaptic marker score) and differed between groups ($p = 0.026$). Cisplatin preparations were the most synapse-enriched, MnTBAP+Cisplatin preparations contained the most bone matrix and plasma protein, and MnTBAP preparations contained the most marrow and immune-cell protein. Among the 4,630 proteins with complete values, adjustment for this composition axis reduced the proportion of proteins altered by cisplatin ($p < 0.05$) from 15.9% to 6.6% (5% expected by chance). Moreover, classifier scores correlated with composition ($r = 0.76$ to 0.90). Because composition itself differed between groups ($R^2 = 0.55$), the result depended on the covariate used, and the cisplatin effect could not be separated from composition. Collectively, these findings suggest that differences in preparation composition are confounded with treatment in this dataset and may account for much of the apparent treatment effect, and that composition should be measured and modeled before synaptosomal proteomics can support treatment-specific conclusions.

1. Introduction

Inner hair cells transmit acoustic signals to the auditory nerve through ribbon synapses, which connect each inner hair cell to type I spiral ganglion neurons. Therefore, ribbon synapses are essential for conducting sound signals to the upstream auditory pathway. These synapses are particularly vulnerable to cochlear injury. Studies of noise exposure established that ribbon synapses can be lost without a permanent change in hearing thresholds (Kujawa and Liberman, 2009). Moreover, this cochlear

synaptopathy reduces the amplitude of wave I of the auditory brainstem response and is thought to impair hearing in noisy environments (Liberman and Kujawa, 2017).

Cisplatin is a cornerstone of chemotherapy for testicular, ovarian, head and neck, and pediatric solid tumors. Permanent, dose-dependent sensorineural hearing loss is one of its most common long-term adverse effects (Rybak et al., 2019). Furthermore, platinum is retained in the cochlea for months to years after treatment, with the highest levels in the stria vascularis (Breglio et al., 2017), and this retention is likely to contribute to the progressive hearing loss observed in many survivors. The mechanisms of cisplatin ototoxicity include the generation of reactive oxygen and nitrogen species, DNA damage and the activation of cell death pathways in outer hair cells, the stria vascularis and spiral ganglion neurons (Rybak et al., 2019). In addition to these effects, cisplatin treatment alters the abundance of cochlear synaptosomal proteins and causes cochlear synaptic dysfunction in mice, and targeting oxidative stress with the manganese porphyrin MnTBAP may prevent this dysfunction (Shahab et al., 2024, 2025). MnTBAP is a weak superoxide dismutase mimic but an efficient scavenger of peroxynitrite (Batinić-Haberle et al., 2009, 2010).

Quantitative proteomics with isobaric tandem mass tags (TMT) allows 16 samples to be labeled, pooled and measured as a single multiplexed batch, which removes run-to-run variation. Two features of cochlear synaptosome studies make their analysis difficult. First, the small amount of tissue in a mouse cochlea limits designs to a few animals per group. With far more proteins than animals, selecting features on the same samples used to evaluate a classifier produces optimistic and sometimes perfect accuracy, even without any true signal (Ambrose and McLachlan, 2002, Varma and Simon, 2006). This error is the machine learning form of circular analysis (Kriegeskorte et al., 2009). Additionally, small samples inflate the effect sizes of features that pass a significance threshold (Button et al., 2013). Second, a synaptosome preparation from the cochlea is a mixture. Dissection of the cochlea carries over otic capsule bone, bone marrow, blood, vasculature, myelinated nerve and non-synaptic cochlear structures, and the proportions of these components vary from preparation to preparation. In other omics fields, unmeasured variation in sample composition is a well-recognized source of false associations (Leek et al., 2010, Shen-Orr and Gaujoux, 2013). Similarly, plasma contamination produces systematic bias in clinical proteomics (Geyer et al., 2019). To date, the effect of composition on cochlear synaptosomal proteomics has not been assessed.

This study tests the hypothesis that cisplatin treatment produces a cochlear synaptosomal proteome signature that is classifiable, robust to preparation composition, and reversed by MnTBAP co-treatment. A TMT dataset generated for a doctoral dissertation (Shahab, 2023) and reported previously (Shahab et al., 2024) was reanalyzed from the raw reporter ion abundances, with no values carried over from the original processing. The analysis combined channel-level quality control, empirical Bayes moderated statistics (Smyth, 2004), nested cross-validation with exact permutation tests, consensus biomarker ranking, and projection of both held-out MnTBAP groups onto the cisplatin axis. Furthermore, tissue composition was quantified with an unsupervised principal component and with six marker panels for synaptic, bone, blood, marrow, myelin and nuclear proteins, and treatment effects were then reassessed after adjustment for composition.

2. Materials and methods

2.1 Dataset and reuse

Animal experiments and sample preparation were performed at Wayne State University as part of the author's doctoral research (Shahab, 2023). Mass spectrometry was performed at the Wayne State University Proteomics Core. The detailed experimental methods are described in Shahab (2023) and Shahab et al. (2024) and are summarized here.

Five-week-old male CBA/J mice (Jackson Laboratory) with normal baseline auditory thresholds were used in this study. Cisplatin was administered at 3 mg/kg/day by slow intraperitoneal infusion for 5 days. MnTBAP was administered intraperitoneally at 10 mg/kg/day for 7 days, one hour before cisplatin on the days both drugs were given. Control mice received an equal volume of saline. Additionally, all mice received subcutaneous saline for hydration and were euthanized on day 8. Cochlear tissue (lateral wall, sensory epithelium and modiolus) was homogenized in Syn-PER reagent (Thermo Fisher). Debris was removed at $1,200 \times g$, and the synaptosomal fraction was collected as the pellet of a $15,000 \times g$ centrifugation. Proteins were solubilized in lithium dodecyl sulfate, reduced, alkylated, precipitated with methanol and digested with trypsin. Twenty micrograms of each digest was labeled with one TMTpro 16-plex reagent, and the pooled sample was separated into 13 fractions by high-pH reversed-phase chromatography. The fractions were analyzed by data-dependent liquid chromatography and tandem mass spectrometry on an Orbitrap Eclipse mass spectrometer with FAIMS, and reporter ions were quantified in MS2 spectra. Furthermore, spectra were searched with Proteome Discoverer 2.4 (Sequest HT and Percolator) against the UniProt mouse proteome (UP000000589, 17,035 entries), with a false discovery rate of 1% for high-confidence identifications.

The experiment contained four groups of four biological replicates: Control, Cisplatin, MnTBAP+Cisplatin and MnTBAP alone. Each group was assigned four consecutive TMT channels: Control 126, 127N, 127C and 128N, Cisplatin 128C, 129N, 129C and 130N, MnTBAP+Cisplatin 130C, 131N, 131C and 132N, and MnTBAP 132C, 133N, 133C and 134N. Samples are named by group and replicate number in channel order, from Control 1 to Control 4, Cisplatin 1 to Cisplatin 4, MnTBAP+Cisplatin 1 to MnTBAP+Cisplatin 4 and MnTBAP 1 to MnTBAP 4.

Protein-level reporter ion abundances were obtained as a Proteome Discoverer 2.4 export (Sequest HT search engine) from the Wayne State University Proteomics Core, and raw spectra were not reprocessed. The export contained 8,306 protein groups. Spreadsheet working rows in the export file (channel sums, loading factors and blank rows) were removed before analysis. Only the raw grouped reporter ion abundances for the 16 channels and the protein identification fields were used. Therefore, normalized abundances, group statistics, fold changes and p values computed in the original analysis were not used at any step.

2.2 Channel quality control

Before normalization, each channel was summarized by its total reporter abundance, its median $\log_2$ abundance across high-confidence proteins, and its number of missing values. A channel was excluded if its total reporter abundance was below 25% of the median channel total. This threshold was set before inspection of the treatment groups.

2.3 Protein filtering and normalization

Proteins with a Proteome Discoverer protein FDR confidence of "High" were retained. Furthermore, proteins were required to be quantified in at least three retained replicates of every group. Abundances were $\log_2$ -transformed, and each channel was median-centered using proteins quantified in all retained channels, with the grand median added back. No missing values were imputed. Differential abundance testing used all retained proteins with their available values, whereas machine learning, principal component analysis and co-expression analysis used the proteins with complete values across all retained channels.

2.4 Differential abundance

Differential abundance was assessed with empirical Bayes moderated t-statistics for a four-group means model (Smyth, 2004). This approach estimates a common prior for protein-wise variances and shrinks the variance of each protein toward it, which increases power in small experiments. The prior degrees of freedom and prior variance were estimated by the method of moments, and per-protein residual degrees of freedom accounted for missing values. The implementation was written in Python and validated against the limma algorithm (Ritchie et al., 2015) as implemented in InMoose (Python port of limma). For all five contrasts, prior estimates were identical, and moderated t and p values agreed to within 3×10^{-14} for all retained proteins (including missing values) as well as for proteins with complete values. Five contrasts were tested: Cisplatin versus Control, MnTBAP+Cisplatin versus Control, MnTBAP+Cisplatin versus Cisplatin, MnTBAP versus Control, and MnTBAP+Cisplatin versus MnTBAP. P values were adjusted by the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995), and proteins with FDR below 0.05 were considered differentially abundant. Additionally, an exploratory "cisplatin signature" was defined before downstream analyses as the proteins with a Cisplatin versus Control FDR below 0.20, for descriptive analyses of the cisplatin response.

2.5 Global signal and permutation null

The overall Cisplatin versus Control signal was assessed in three ways. First, the distribution of p values and the proportion of true null hypotheses, π_0 , were estimated by Storey's method at $\lambda = 0.5$ (Storey and Tibshirani, 2003). Second, the number of proteins at $p < 0.01$ and at $\text{FDR} < 0.20$ was determined. Third, an empirical null distribution of these counts was generated. The eight Control and Cisplatin samples can be split into two groups of four in 35 distinct ways, one of which is the true labeling, and the moderated model was refitted for each of the 34 other partitions. The 18 balanced partitions, with two Control and two Cisplatin samples on each side, form the cleanest null because the true contrast cancels.

2.6 Classification with nested feature selection

Control and Cisplatin samples ($n = 8$) were classified by leave-one-out cross-validation. In each fold, protein selection, standardization and model fitting used only the seven training samples. Protein filtering and

median normalization (Section 2.3) were applied once to all retained samples before classification. Proteins were ranked on the seven training samples by a two-group moderated t-test, and the 50 most significant proteins were retained, a value fixed in advance. Features were standardized with training-fold statistics, a classifier was fitted, and the held-out sample was scored. Five classifiers with fixed hyperparameters were compared: L2-penalized logistic regression ($C = 0.1$), a linear support vector machine ($C = 1$) (Cortes and Vapnik, 1995), a random forest of 300 trees (Breiman, 2001), gradient boosting with 100 trees of depth 2 and a learning rate of 0.1 (Friedman, 2001), and an equal-weight signature score. The signature score standardizes each selected protein by its pooled within-group standard deviation around the midpoint of the two training-class means, signs it by the direction of change, and averages across proteins. This rule is diagonal linear discriminant analysis with sign weights and suits very small training sets (Dudoit et al., 2002, Tibshirani et al., 2002). Performance was summarized as accuracy, balanced accuracy, F1 score and the area under the ROC curve (AUROC) computed from pooled held-out scores, and all models used scikit-learn (Pedregosa et al., 2011).

Significance was assessed with an exact permutation test (Ojala and Garriga, 2010). The full nested procedure, including in-fold protein selection, was repeated for all 70 assignments of four Cisplatin labels among the eight samples, and the p value is the fraction of assignments reaching at least the observed performance. For every classifier, the mirror image of the true labeling also reached the observed value, so the smallest attainable p value was $2/70$ (0.029). Moreover, to measure the optimism of the conventional protocol, each classifier was also evaluated after selecting the top 50 proteins on all eight samples before cross-validation. The stability of in-fold selection was summarized by the Jaccard overlap of the selected sets across folds. In addition, a signature-size curve repeated the nested procedure with 5 to 500 proteins and with no selection, for all classifiers except gradient boosting.

2.7 Consensus ranking of the cisplatin signature

Each protein in the cisplatin signature was scored on four criteria in the eight Control and Cisplatin samples: selection stability, the absolute coefficient in an L2-penalized logistic regression fitted to all signature proteins, random forest impurity importance (5,000 trees), and the absolute bias-corrected standardized mean difference, Hedges' g (Hedges, 1981). Selection stability was defined as the fraction of the 16 balanced three-versus-three subsamples in which the protein reached $p < 0.05$ by moderated t-test, computed on all proteins with complete values, with the same direction of change (Meinshausen and Bühlmann, 2010). Each criterion was converted to a percentile rank, and the consensus score was the mean of the four percentiles. This ranking uses all eight samples and is therefore descriptive.

2.8 Held-out projection of the MnTBAP groups

MnTBAP+Cisplatin and MnTBAP samples were never used to fit the binary models or to select the top 50 and top 150 proteins. The full cisplatin signature (FDR below 0.20) was defined from the four-group model, whose variance estimates include all retained samples. These groups carry neither the Control nor the Cisplatin label, and therefore this analysis does not measure classifier accuracy. Instead, it measures how much of the cisplatin signature each held-out sample carries. Moreover, the MnTBAP group served as a specificity check, because a treatment that does not produce a cisplatin-like proteome should project near

the Control reference. Three approaches assessed where these samples fall relative to Control and Cisplatin. First, each classifier was trained on all eight Control and Cisplatin samples, with either the top 50 proteins or the full signature, and each MnTBAP sample was classified as Control-like or Cisplatin-like. Second, each sample was projected onto the vector joining the Control and Cisplatin centroids in $\log_2$ space, $s = \sum d_j (x_j - a_j) / \sum d_j^2$, where a_j is the Control mean of protein j and d_j is the difference between Cisplatin and Control means. A score of 0 marks the Control centroid and 1 marks the Cisplatin centroid. Furthermore, because proteins selected for a large Cisplatin versus Control difference regress toward the middle in any independent sample, Control and Cisplatin samples were scored leave-one-out, with the held-out sample excluded from protein selection and from both centroids. These held-out scores provide calibrated references for independent control and cisplatin-exposed animals. Scores were computed for the top 50 and top 150 proteins. Third, the $\log_2$ fold changes of MnTBAP+Cisplatin versus Control were compared with those of Cisplatin versus Control across the signature, with the slope of a regression through the origin interpreted against the held-out Cisplatin reference.

2.9 Tissue composition

Composition was assessed in two independent ways. First, principal component analysis of all proteins with complete values was performed without group labels, and the first principal component (PC1) was examined. Second, six marker panels were assembled from canonical proteins for each component: synaptic and neuronal proteins (for example SYP, SNAP25, STXBP1, GRIA1 to GRIA4, DLG4, CTBP2 and OTOF), bone and extracellular matrix proteins (for example COL1A1, COL1A2, IBSP, SPP1, SPARC and TNFRSF11B), blood and plasma proteins (for example hemoglobins, albumin, apolipoproteins, fibrinogens, complement C3 and MUG1), marrow and immune-cell proteins (for example PTPRC, LCP1, PTPN6, S100A8, S100A9, MPO and LTF), myelin and glial proteins (for example MBP, PLP1, MPZ and PMP22), and nuclear proteins (lamins, nucleoporins, heterogeneous nuclear ribonucleoproteins and all quantified histones). For each panel, the $\log_2$ abundances of quantified markers were z-scored across samples, and the per-sample panel score was their median. The complete marker lists are provided in the code. Panel scores were compared between groups with two-sample t-tests without correction for multiple comparisons, and these comparisons are descriptive. Panel scores were also correlated with PC1.

To assess treatment effects independent of composition, the moderated linear model was refitted with PC1 as a covariate in addition to the four group means. PC1 was derived without group labels. Because PC1 also differed between groups, adjustment for PC1 removes any treatment effect aligned with composition. Therefore, the variance in PC1 explained by group (R^2) was reported, and two alternative covariates were examined for the Cisplatin versus Control contrast: the synaptic marker score, and PC1 after removal of the group means, a group-orthogonal covariate related to surrogate variable and unwanted-variation methods (Leek and Storey, 2007, Gagnon-Bartsch and Speed, 2012). The covariate model was validated against limma in the same way as the group model, with identical prior degrees of freedom and moderated t and p values agreeing to within 1.5×10^{-13} for all five contrasts. Additionally, the relationship between classifier scores and composition was examined by correlating the held-out score of each classifier with PC1, by classifying samples with PC1 alone, and by repeating nested classification and the exact permutation test after removing PC1 from the data.

2.10 Four-group classification and co-expression

All retained samples were classified into the four groups by nested leave-one-out cross-validation with logistic regression, the linear support vector machine, random forest and gradient boosting, and with the 300 proteins with the largest one-way ANOVA F statistic selected within each fold. For co-expression analysis, the signature proteins were z-scored across all retained samples and clustered by average linkage on $1 - \text{Pearson } r$, with the number of modules (two to eight, with at least five proteins per module) chosen by the silhouette coefficient. Module eigengenes were the first principal component of each module, signed to agree with the module mean (Zhang and Horvath, 2005, Langfelder and Horvath, 2008). Furthermore, hub proteins were ranked by intramodular connectivity using the soft-threshold adjacency $|r|^6$. Because the signature was selected on the Cisplatin versus Control contrast, eigengene differences between these two groups are expected by construction, whereas comparisons involving the MnTBAP groups are independent of selection.

2.11 Sensitivity analyses

The influence of each Control and Cisplatin sample was assessed by removing it and recomputing the Cisplatin versus Control signal. In addition, all 35 partitions of the eight samples were ranked by the number of proteins at $p < 0.01$. An exploratory analysis repeated the core analyses without the Control and Cisplatin samples whose removal most increased the proportion of proteins at $p < 0.05$, which were identified from the data and therefore do not provide an independent test. Finally, the excluded channel was restored to measure its effect on the MnTBAP+Cisplatin contrasts.

2.12 Software and code availability

Analyses used Python 3.11 with NumPy 2.4, pandas 2.3, SciPy 1.17, scikit-learn 1.8, matplotlib 3.10 and adjustText 1.4, and the moderated statistics were validated against InMoose 0.9. The complete pipeline (p1_analysis.py, p1_sensitivity.py, p1_figures.py, limma_stats.py, relabel_outputs.py and validate_limma.py, Python 3.11.15) regenerates every number, table and figure in this article from the Proteome Discoverer export. It is openly available at <https://github.com/Monazzashahab/cochlear-synaptosome-composition-ml>.

3. Results

3.1 Channel quality control identifies one under-loaded sample

Fifteen channels contained between 258,000 and 550,000 total reporter abundance units. In contrast, channel 130C, which carried MnTBAP+Cisplatin 1, contained 32,000 units, which is 10.0% of the median channel total (Fig. 2A, Table S1). Moreover, its median $\log_2$ abundance was 2.58, compared with 4.94 to 6.14 in the other channels, and it lacked values for 1,565 high-confidence proteins, compared with 795 to 848 in the other channels. MnTBAP+Cisplatin 1 was the only channel below the quality control threshold and was therefore excluded, which left three replicates in the MnTBAP+Cisplatin group. Additionally, in a

principal component analysis of all 16 channels, MnTBAP+Cisplatin 1 was separated from all other samples on the third principal component (score 74.4, compared with -28.9 to 19.1, Figure S1A).

Of 8,306 protein groups, 5,552 were identified with high confidence, and 4,667 were quantified in at least three retained replicates of every group. Only 45 of 70,005 values (0.06%) were missing among these proteins, and 4,630 proteins had complete values across the 15 retained channels.

3.2 Cisplatin produces a modest and diffuse change in the synaptosomal proteome

The empirical Bayes prior had 3.73 degrees of freedom. Only one protein, MUG1 (murinoglobulin 1), was decreased in Cisplatin relative to Control at FDR below 0.05 (fold change 0.32, $q = 0.013$). Furthermore, 14 proteins reached FDR below 0.10, and 147 reached FDR below 0.20 (78 increased and 69 decreased, Fig. 2B, Table S3). Absolute $\log_2$ fold changes in this signature ranged from 0.54 to 2.72. The p value distribution was enriched near zero (Fig. 2C), and 15.8% of proteins had $p < 0.05$, compared with 5% expected under the null, with a π_0 of 0.61.

This signal exceeded the empirical null (Fig. 2D). Across the 18 balanced relabelings, the number of proteins at $p < 0.01$ had a median of 34.5 and a maximum of 135, compared with 220 under the true labels. Similarly, the number of proteins at FDR below 0.20 had a median of 0 and a maximum of 75, compared with 147. Moreover, among all 35 partitions, the true labeling ranked third by the number of proteins at $p < 0.01$.

The other contrasts produced larger differences (Table S2). MnTBAP+Cisplatin differed from Cisplatin in 209 proteins at FDR below 0.05 (127 increased and 82 decreased), from Control in 70 proteins, and from MnTBAP in 103 proteins. In addition, MnTBAP alone differed from Control in 57 proteins.

Unsupervised principal component analysis of the 4,630 complete proteins did not separate Control from Cisplatin cleanly (Fig. 2E). PC1 and PC2 explained 31.3% and 17.3% of the variance, respectively. Cisplatin 1, Cisplatin 2 and Cisplatin 4 lay at high PC1 values, MnTBAP+Cisplatin samples lay at low PC2 values, and MnTBAP 1 to MnTBAP 3 lay at low PC1 values. Furthermore, Control 3 and Control 4 lay between the Control and Cisplatin samples (Figure S1B and C).

3.3 Nested cross-validation removes the apparent perfect classification

When the top 50 proteins were selected on all eight Control and Cisplatin samples before cross-validation, all five classifiers classified every sample correctly (accuracy 1.00, AUROC 1.00, Fig. 3B, Table 1). This is the result a conventional workflow would report.

Table 1

Control versus Cisplatin classification by leave-one-out cross-validation (n = 8). Nested, top 50 proteins re-selected by moderated t-test within each fold. Selection outside CV, top 50 proteins selected on all eight samples before cross-validation (biased protocol, shown for comparison). Permutation p values are exact over all 70 labelings.

Classifier	Accuracy, nested	AUROC, nested	Permutation p, accuracy	Permutation p, AUROC	Accuracy, selection outside CV	AUROC, selection outside CV
Signature score	0.625	0.875	0.257	0.114	1.000	1.000
Logistic regression	0.625	0.750	0.286	0.171	1.000	1.000
Linear SVM	0.875	0.938	0.114	0.086	1.000	1.000
Random forest	0.500	0.750	0.486	0.143	1.000	1.000
Gradient boosting	0.625	0.750	0.186	0.129	1.000	1.000

In contrast, with protein selection repeated inside each fold, accuracy was 0.875 for the linear support vector machine, 0.625 for the signature score, logistic regression and gradient boosting, and 0.500 for the random forest (Table 1), and AUROC ranged from 0.75 to 0.94 (Fig. 3A). No classifier reached significance by exact permutation test. The linear support vector machine came closest (AUROC 0.94, $p = 0.086$, accuracy $p = 0.114$), and accuracy p values for the other classifiers ranged from 0.186 to 0.486 (Fig. 3C). Moreover, the top 50 proteins selected in different folds overlapped poorly, with a median Jaccard index of 0.35, and only 15 proteins were selected in all eight folds. Across signature sizes from 5 to 500 proteins and without selection, nested AUROC ranged from 0.63 to 0.94 (Fig. 3D, Table S5). Because the signature size was not tuned, this curve is descriptive.

The errors were consistent across classifiers (Fig. 3E, Table S6). Control 3 was misclassified by all five classifiers, and Control 4 and Cisplatin 3 were misclassified by four of five. The remaining five samples were classified correctly by at least four classifiers.

3.4 Preparation composition is the dominant source of variation

PC1, which explained 31.3% of the variance, corresponded to the synaptic content of each preparation. Its highest positive loadings were synaptic and neuronal proteins, including WASF1, CAMK2D, the AP-2 adaptor subunits AP2A2 and AP2B1, SRGAP2, RGS8, CACNB4 and DLGAP3 (Fig. 4F). In contrast, its highest negative loadings included vascular, adipose, plasma and stress proteins such as ANXA11, plasminogen (PLG), CRYAB, TIMP3 and adiponectin (ADIPOQ). PC1 correlated with the synaptic and neuronal marker score ($r = 0.91$) and inversely with the blood and plasma ($r = -0.74$), marrow and immune ($r = -0.74$), bone and extracellular matrix ($r = -0.45$) and myelin and glia ($r = -0.43$) scores (Fig. 4A, Table 5). Furthermore, PC1 differed between groups (one-way ANOVA $p = 0.026$, Fig. 4B).

Table 5

Tissue composition of the preparations. Panel score, median z-score of $\log_2$ abundance across the quantified markers of each panel. Group means and two-sample t-test p values. $r(\text{PC1})$, correlation of the panel score with the first principal component across 15 samples. Ctrl, Control. Cis, Cisplatin. M + C, MntBAP+Cisplatin. Mn, MntBAP.

Panel	Markers	Ctrl	Cis	M + C	Mn	Cis vs Ctrl p	M + C vs Cis p	Mn vs Ctrl p	$r(\text{PC1})$
Synaptic/neuronal	35	0.29	0.86	-0.58	-0.69	0.15	0.055	0.053	0.91
Bone/ECM	22	-0.07	-0.75	1.28	-0.40	0.037	0.002	0.21	-0.45
Blood/plasma	32	0.07	-0.84	0.58	0.23	0.023	0.004	0.49	-0.74
Marrow/immune	28	-0.26	-0.95	-0.03	1.26	0.11	0.064	0.002	-0.74
Myelin/glia	9	0.30	-0.28	0.54	-0.26	0.46	0.14	0.51	-0.43
Nuclear	45	-0.66	0.13	0.30	0.15	0.036	0.50	0.059	0.15

Each treatment group had a distinct composition profile (Fig. 4A, Table 5). Cisplatin preparations had the highest synaptic score (mean 0.86, compared with 0.29 in Control, $p = 0.15$) and the lowest bone (-0.75 compared with -0.07, $p = 0.037$) and blood scores (-0.84 compared with 0.07, $p = 0.023$). In contrast, MntBAP+Cisplatin preparations had the highest bone score (1.28 compared with -0.75 in Cisplatin, $p = 0.002$) and a higher blood score than Cisplatin (0.58, $p = 0.004$). Additionally, MntBAP preparations had the highest marrow and immune score (1.26 compared with -0.26 in Control, $p = 0.002$). The nuclear score was lower in Control (-0.66) than in Cisplatin (0.13, $p = 0.036$) and MntBAP+Cisplatin (0.30, $p = 0.022$), and showed the same trend relative to MntBAP (0.15, $p = 0.059$). The nuclear score was unrelated to PC1 ($r = 0.15$).

The atypical samples followed composition. Control 3 had a synaptic score of 0.76, close to the Cisplatin mean, whereas Cisplatin 3 had a synaptic score of 0.04, close to the Control mean (Figure S2C). Moreover, the held-out scores of the classifiers correlated with PC1 ($r = 0.76$ to 0.90 , Fig. 4C), and PC1 alone classified six of eight Control and Cisplatin samples correctly by leave-one-out.

3.5 The cisplatin effect cannot be separated from composition

Adding PC1 as a covariate reduced the Cisplatin versus Control signal to near the null level (Fig. 4D and E, Table 6). Covariate models require complete values, and therefore these comparisons use the 4,630 proteins with complete values. Thus, the unadjusted counts differ from those for all 4,667 proteins in Section 3.2 (for example, 92 rather than 70 proteins for MntBAP+Cisplatin versus Control). The proportion of proteins at $p < 0.05$ decreased from 15.9% to 6.6%, π_0 increased from 0.61 to 0.94, the number of proteins at $p < 0.01$ decreased from 219 to 81, and the number at FDR below 0.20 decreased from 159 to nine. Similarly, the MntBAP+Cisplatin versus Cisplatin contrast decreased from 209 to five proteins at FDR below 0.05, and the five remaining proteins were the bone and plasma proteins IBSP, SERPINC1, AEBP1,

prothrombin (F2) and TNFRSF11B. In contrast, the MnTBAP versus Control contrast was not reduced by adjustment (57 proteins at FDR below 0.05 before and 67 after). The increased proteins in this contrast included the hematopoietic proteins PTPN6, LCP1, ARHGDIB, pleckstrin (PLEK) and SAMHD1, consistent with the marrow and immune composition of these preparations, which PC1 does not fully capture. Additionally, the contrasts between MnTBAP+Cisplatin and Control or MnTBAP increased after adjustment. MnTBAP+Cisplatin samples lay outside the bone and plasma composition range of the other groups, which PC1 does not capture, and therefore these adjusted estimates are not interpreted.

Table 6

Differential abundance with and without adjustment for composition (4,630 proteins with complete values). Unadjusted counts differ from those reported for all 4,667 proteins in Section 3.2 because proteins with any missing value are excluded here. Alternative covariates for Cisplatin versus Control are given in Table S16.

Contrast	Model	FDR < 0.05	FDR < 0.20	p < 0.01	% p < 0.05	π_0
Cisplatin vs Control	Group only	1	159	219	15.9	0.61
Cisplatin vs Control	+ PC1	0	9	81	6.6	0.94
MnTBAP+Cisplatin vs Cisplatin	Group only	209	1,268	510	26.3	0.54
MnTBAP+Cisplatin vs Cisplatin	+ PC1	5	222	224	15.9	0.68
MnTBAP vs Control	Group only	57	303	269	15.1	0.72
MnTBAP vs Control	+ PC1	67	463	319	16.8	0.70
MnTBAP+Cisplatin vs Control	Group only	92	632	368	18.8	0.65
MnTBAP+Cisplatin vs Control	+ PC1	423	1,346	622	27.2	0.56
MnTBAP+Cisplatin vs MnTBAP	Group only	104	390	303	15.1	0.72
MnTBAP+Cisplatin vs MnTBAP	+ PC1	192	583	386	17.4	0.73

Within the cisplatin signature, 52 of 146 proteins correlated with PC1 at $|r| \geq 0.7$ (Fig. 6B). Decreased proteins tracked composition more closely (mean $r = -0.67$) than increased proteins (mean $r = 0.46$). Moreover, the single protein significant at FDR below 0.05, MUG1, is a plasma protein.

The effect of adjustment depended on the covariate (Table S16). Group explained 55% of the variance in PC1 ($R^2 = 0.55$), so adjustment for PC1 also removes any cisplatin effect aligned with composition. With the synaptic marker score as the covariate, 29 proteins reached FDR below 0.20, 11.6% of proteins had $p < 0.05$, and π_0 was 0.76. In contrast, with PC1 after removal of the group means, which retains any treatment

effect aligned with composition, 474 proteins reached FDR below 0.05, 28.0% of proteins had $p < 0.05$, and π_0 was 0.55.

Removing PC1 from the data before nested classification did not abolish classification. Accuracy was 0.75 for the signature score, logistic regression and linear support vector machine, with AUROC of 0.88 to 0.94. These results remained non-significant by exact permutation test (accuracy $p = 0.143$, AUROC $p = 0.086$).

3.6 Consensus biomarkers and an exploratory nuclear RNA-processing module

Of the 146 signature proteins with complete values, 113 were selected in at least 75% of the balanced subsamples and 79 in all 16 (Table S7). The 20 proteins with the highest consensus scores are listed in Table 2 and shown in Fig. 6A. Of the top 20 proteins, 11 were decreased and nine were increased in Cisplatin. Figure 6C shows the top 25 proteins, and Figure S3 shows the top eight. The decreased proteins included TACC1, WDR47, AKAP9, ADSL, TRIP12, AMACR, GPT2, ANXA7, OGFR and the plasma protein MUG1, and several of these correlated strongly with composition, including AMACR ($r = -0.82$), GPT2 ($r = -0.77$) and ADSL ($r = -0.72$). In contrast, the increased proteins were dominated by nuclear RNA-processing and chromatin proteins, including FIP1L1, RBM39, NUFIP2, TCOF1, CNOT3, EIF5B and CBX5, together with the proteasome inhibitor PSMF1 (PI31). With the exception of AP3S2 ($r = 0.73$) and EIF5B ($r = 0.62$), these proteins correlated less strongly with PC1 ($r = 0.02$ to 0.45).

Table 2

Top 20 consensus-ranked proteins of the cisplatin signature. FC, ratio of geometric mean abundance, Cisplatin/Control. p and q, moderated t-test and Benjamini-Hochberg FDR. Stability, fraction of 16 balanced three-versus-three subsamples with $p < 0.05$ in the same direction. r(PC1), correlation with the composition axis across 15 samples. Recovery, descriptive fraction of the cisplatin $\log_2$ change absent in MnTBAP+Cisplatin.

Rank	Protein	Name	FC	p	q	Stability	r(PC1)	Recovery
1	TACC1	Transforming acidic coiled-coil protein 1	0.18	2.3 × 10 ⁻⁵	0.053	1.00	-0.45	0.68
2	WDR47	WD repeat-containing protein 47	0.52	1.4 × 10 ⁻³	0.145	1.00	-0.49	1.23
3	ALDH1L2	Mitochondrial 10-formyltetrahydrofolate dehydrogenase	0.53	9.2 × 10 ⁻⁴	0.128	1.00	-0.48	0.90
4	PSMF1	Proteasome inhibitor PI31 subunit	1.97	2.4 × 10 ⁻³	0.158	1.00	0.02	0.48
5	FIP1L1	Pre-mRNA 3'-end-processing factor FIP1	1.80	1.6 × 10 ⁻³	0.145	1.00	0.39	-0.04
6	ADSL	Adenylosuccinate lyase	0.50	7.1 × 10 ⁻⁴	0.124	1.00	-0.72	0.43
7	TRIP12	E3 ubiquitin-protein ligase TRIP12	0.48	4.9 × 10 ⁻³	0.186	1.00	-0.66	0.74
8	AMACR	Alpha-methylacyl-CoA racemase	0.54	4.6 × 10 ⁻³	0.180	1.00	-0.82	0.89
9	CBX5	Chromobox protein homolog 5	1.72	7.1 × 10 ⁻⁴	0.124	1.00	0.45	0.30
10	MUG1	Murinoglobulin 1	0.32	2.9 ×	0.013	1.00	-0.39	0.89

Rank	Protein	Name	FC	p	q	Stability	r(PC1)	Recovery
				10^{-6}				
11	AP3S2	AP-3 complex subunit sigma-2	2.72	4.7×10^{-5}	0.055	1.00	0.73	0.36
12	ANXA7	Annexin A7	0.57	2.3×10^{-3}	0.158	1.00	-0.67	1.46
13	NUFIP2	Nuclear FMR1-interacting protein 2	2.26	1.5×10^{-4}	0.070	1.00	0.21	-0.31
14	RBM39	RNA-binding protein 39	1.59	1.9×10^{-3}	0.150	1.00	0.35	-0.18
15	OGFR	Opioid growth factor receptor	0.32	6.7×10^{-5}	0.060	1.00	-0.28	0.48
16	EIF5B	Eukaryotic translation initiation factor 5B	1.80	1.9×10^{-3}	0.150	1.00	0.62	0.32
17	AKAP9	A-kinase anchor protein 9	0.34	1.0×10^{-4}	0.060	1.00	-0.55	0.62
18	GPT2	Alanine aminotransferase 2	0.50	8.0×10^{-4}	0.124	1.00	-0.77	0.67
19	TCOF1	Treacle protein	1.89	2.6×10^{-4}	0.099	1.00	0.07	0.00
20	CNOT3	CCR4-NOT transcription complex subunit 3	2.07	1.4×10^{-4}	0.070	1.00	0.29	0.10

Co-expression clustering of the signature produced two modules (silhouette 0.75, Figure S5, Table 4). Module M1 contained the 78 increased proteins, with splicing and nuclear speckle proteins among its hubs (BOD1L, WBP4, SRSF2, PPP1R10, CDKN2AIP and SRSF5). Module M2 contained the 68 decreased proteins, with GGT5, MAGT1, CD163, LY6C2, LRRC8C and PGM5 as hubs. Furthermore, the M1 eigengene was also higher in MnTBAP than in Control ($p = 0.017$). After adjustment for PC1, all 78 M1 proteins remained higher in Cisplatin than in Control (mean $\log_2$ fold change 0.67, 61 proteins at $p < 0.05$), and after adjustment for both PC1 and the nuclear score, 42 proteins remained at $p < 0.05$. Because these proteins were selected for a higher abundance in Cisplatin, part of this persistence is expected from selection alone.

Table 4

Co-expression modules of the cisplatin signature. p values from two-sample t-tests on module eigengenes. The Cisplatin versus Control comparison is expected by construction. Recovery, (Cisplatin – MnTBAP+Cisplatin) / (Cisplatin – Control) on the eigengene. Ctrl, Control. Cis, Cisplatin. M + C, MnTBAP+Cisplatin. Mn, MnTBAP.

Module	Proteins	Change in Cis	Cis vs Ctrl p	M + C vs Cis p	Mn vs Ctrl p	Recovery	Hub proteins
M1	78	Increased	0.001	0.18	0.017	0.24	BOD1L, WBP4, SRSF2, PPP1R10, CDKN2AIP, SRSF5
M2	68	Decreased	0.001	0.006	0.57	0.78	GGT5, MAGT1, CD163, LY6C2, LRRC8C, PGM5

3.7 MnTBAP+Cisplatin samples occupy an intermediate position that composition confounds

In the held-out projection, the classifiers trained on Control and Cisplatin samples called one of three MnTBAP+Cisplatin samples (MnTBAP+Cisplatin 3) and one of four MnTBAP samples (MnTBAP 4) Cisplatin-like (Fig. 5C, Table 3). All five classifiers agreed, with both the top 50 proteins and the full signature. Under nested leave-one-out, the same classifiers called one or two of four Control samples and two to four of four Cisplatin samples Cisplatin-like, and therefore the held-out calls should be read against that baseline.

Table 3

Held-out projection of the MnTBAP groups. Number of samples per group called Cisplatin-like. Control and Cisplatin, nested leave-one-out with the top 50 proteins. MnTBAP+Cisplatin (n = 3) and MnTBAP (n = 4), models trained on all eight Control and Cisplatin samples. Cisplatin-axis score, mean $\pm$ SD (0, Control centroid, 1, Cisplatin centroid), with Control and Cisplatin scored leave-one-out.

Signature score	Control	Cisplatin	MnTBAP+Cisplatin	MnTBAP
	2/4	3/4	1/3	1/4
Logistic regression	2/4	3/4	1/3	1/4
Linear SVM	1/4	4/4	1/3	1/4
Random forest	2/4	2/4	1/3	1/4
Gradient boosting	2/4	3/4	1/3	1/4
Cisplatin-axis score, top 50	0.24 $\pm$ 0.36	0.70 $\pm$ 0.28	0.51 $\pm$ 0.25	0.41 $\pm$ 0.30
Cisplatin-axis score, top 150	0.24 $\pm$ 0.41	0.68 $\pm$ 0.35	0.43 $\pm$ 0.24	0.26 $\pm$ 0.34

On the calibrated cisplatin axis, held-out Control samples scored 0.24 ± 0.36 (mean $\pm$ SD) and held-out Cisplatin samples scored 0.70 ± 0.28 (Fig. 5B, Table S8). The Cisplatin reference is well below 1, which reflects selection-induced shrinkage. MnTBAP+Cisplatin samples scored 0.51 ± 0.25 , about 42% of the calibrated distance from Cisplatin back toward Control, and this position was not significantly different from either Cisplatin ($p = 0.40$) or Control ($p = 0.33$). Similarly, MnTBAP samples scored 0.41 ± 0.30 ($p = 0.52$ versus Control). The 150-protein version gave the same pattern (Control 0.24, Cisplatin 0.68, MnTBAP+Cisplatin 0.43, MnTBAP 0.26, MnTBAP+Cisplatin versus Cisplatin $p = 0.33$). Moreover, across the signature, the changes in MnTBAP+Cisplatin relative to Control correlated with those in Cisplatin ($r = 0.81$), with a slope of 0.42 against a held-out Cisplatin reference of 0.68.

The MnTBAP+Cisplatin versus Cisplatin contrast was the largest in the dataset, but it followed composition (Fig. 5D). The most increased proteins were bone and extracellular matrix proteins (IBSP, AEBP1, TNFRSF11B, CRTAP, CILP and SPP1) and plasma proteins (SERPINC1, F2, ITIH4, IGHM, MUG1, CFB and HABP2). In contrast, the most decreased proteins included neural and synaptic proteins, such as the astrocytic water channel AQP4, PDE1C, hippocalcin (HPCA), RAB37, FGF12, SLC8A1, STXBP1 (MUNC18-1), the AMPA receptor subunit GRIA1 and synaptotagmin 7 (SYT7). Furthermore, this contrast decreased from 209 to five proteins after adjustment for PC1 (Fig. 5E).

In four-group nested classification, accuracy was 0.60 to 0.67 against a chance level of 0.25 (Figure S4, Table S10). Control samples were recalled at 0.50, Cisplatin at 0.50 to 0.75, MnTBAP+Cisplatin at 0.67 and MnTBAP at 0.75.

3.8 Sensitivity analyses

Removing Control 1, Control 2, Cisplatin 2 or Cisplatin 4 sharply reduced the number of proteins at FDR below 0.20 for Cisplatin versus Control, to one, three, six and nine proteins, respectively (Figure S2A, Table

S12). Removing Control 1, Cisplatin 2 or Control 2 also produced the largest decreases in the proportion of proteins at $p < 0.05$. Among the Control and Cisplatin samples, Control 1 and Control 2 had the lowest PC1 scores and Cisplatin 2 the highest. In contrast, removing Control 3, Control 4 or Cisplatin 3 strengthened the signal, and after removal of Cisplatin 3, 957 proteins reached FDR below 0.20. Moreover, among the 35 partitions of the eight samples, the partition that exchanged Control 3 and Cisplatin 3 produced the largest contrast, with 339 proteins at $p < 0.01$, 1,167 at FDR below 0.20 and 25.0% of proteins at $p < 0.05$ (Figure S2B, Table S13). This partition groups samples by composition rather than by treatment.

In an exploratory analysis without Control 3 and Cisplatin 3, 1,429 proteins reached FDR below 0.20, and all 147 proteins of the original signature were among them with the same direction of change. Additionally, nested classification of the remaining six samples reached an accuracy of 0.67 and an AUROC of 0.78 for all five classifiers, with exact permutation p values from 0.10 to 0.30, where 0.10 is the smallest attainable value for six samples.

Restoring MnTBAP+Cisplatin 1 reduced the MnTBAP+Cisplatin versus Cisplatin contrast from 209 to 74 proteins at FDR below 0.05 and the MnTBAP+Cisplatin versus Control contrast from 70 to seven proteins.

4. Discussion

Cochlear synaptosome proteomics provides a direct view of the proteins that support ribbon synapse function and of how ototoxic drugs such as cisplatin alter them (Shahab et al., 2024). The value of this approach depends on whether differences between groups reflect treatment or the content of each preparation. This study provides evidence that differences in preparation composition form the dominant axis of variation between samples in a cochlear synaptosomal TMT dataset and align with the treatment groups. Although the cisplatin signal exceeded every balanced relabeling of the samples, it was small for individual proteins, unstable across folds and samples, and largely removed by adjustment for a composition axis defined without group labels.

Nested cross-validation changed the conclusion of the machine learning analysis. When proteins were selected on all eight samples before cross-validation, every classifier reached perfect accuracy, whereas no classifier exceeded chance by exact permutation test when selection was repeated inside each fold. This difference is consistent with the selection bias described in gene expression studies (Ambroise and McLachlan, 2002, Varma and Simon, 2006) and with circular analysis in neuroscience (Kriegeskorte et al., 2009). Proteins selected on all eight samples already encode the label of each held-out sample, which produces the perfect result. Moreover, the proteins selected in different folds overlapped poorly, with a median Jaccard index of 0.35, which indicates that the selected list depends strongly on which animal is held out. Additionally, all classifiers made the same errors on the same animals. Together, these observations suggest that, at four animals per group, the information in the samples, rather than the capacity of the model, limits classification.

The central finding of this study is that tissue composition dominates the proteome of these preparations. The first principal component, derived without group labels, tracked synaptic protein content closely ($r = 0.91$) and separated synaptic and neuronal proteins from bone, blood, marrow and vascular proteins.

Furthermore, each treatment group had a distinct composition profile. Cisplatin preparations were the most synapse-enriched, MnTBAP+Cisplatin preparations contained the most bone and plasma protein, MnTBAP preparations contained the most marrow and immune-cell protein, and Control preparations contained the least nuclear protein. These differences are likely to arise during dissection and synaptosome isolation, which carry over variable amounts of otic capsule bone, marrow, blood and nuclei. Similar differences in sample composition have been reported as a source of false associations in other omics fields (Leek et al., 2010, Shen-Orr and Gaujoux, 2013) and of systematic bias in plasma proteomics (Geyer et al., 2019). In contrast, a biological explanation is less consistent with the data. Cisplatin causes cochlear synaptopathy, which would be expected to reduce, not increase, the synaptic protein content of a preparation (Liberman and Kujawa, 2017), yet the cisplatin preparations had the highest synaptic scores. Moreover, the atypical animals also followed composition. Control 3 had a cisplatin-like synaptic content, Cisplatin 3 had a control-like content, and the partition that exchanged them produced a far larger contrast than the true labels.

Adjustment for PC1 removed most of the cisplatin signal. The proportion of proteins at $p < 0.05$ decreased to 6.6%, close to the 5% expected by chance, and π_0 increased to 0.94. Moreover, the held-out classifier scores correlated strongly with PC1, which indicates that the models largely read the synaptic content of each preparation. These results do not indicate that cisplatin leaves the synaptosomal proteome unchanged. Group explained 55% of the variance in PC1, so adjustment for PC1 also removes any true cisplatin effect that is aligned with composition. Furthermore, the result depended on the covariate. A group-orthogonal version of PC1 increased the signal to 474 proteins at FDR below 0.05, and the synaptic marker score gave an intermediate result. The group-orthogonal covariate cannot remove composition differences between groups, and PC1 cannot preserve a treatment effect aligned with composition, so the two results bracket the possibilities. Therefore, this dataset cannot attribute its group differences to treatment rather than to composition. The previous analysis of this dataset applied unadjusted thresholds and reported 249 increased and 102 decreased proteins (Shahab et al., 2024). Thus, the present findings suggest that a substantial part of that list reflects composition and that the list should be revisited with composition taken into account.

In addition to the composition effect, a residual cisplatin signal may remain after adjustment. Nested classification after removal of PC1 reached an AUROC of 0.94, close to the significance threshold ($p = 0.086$). The increased proteins in the signature were dominated by nuclear RNA-processing, splicing and chromatin proteins, such as FIP1L1, RBM39, NUFIP2, TCOF1, CNOT3, CBX5, SRSF2 and SRSF5, and these proteins correlated weakly with the synaptic axis. Cisplatin forms DNA adducts and inhibits transcription, and therefore changes in the abundance of RNA-processing proteins are a plausible response. Furthermore, this module remained higher in Cisplatin after adjustment for both the synaptic axis and nuclear content. The proteins were selected for being higher in Cisplatin, so this persistence partly reflects selection, and the module was also higher in MnTBAP alone. Therefore, this module is a hypothesis for a larger, composition-controlled study rather than an established finding.

The MnTBAP results require particular caution. MnTBAP+Cisplatin samples occupied an intermediate position on the calibrated cisplatin axis, about 42% of the way back toward Control, but did not differ

significantly from either group. This position cannot be interpreted as protection alone. The cisplatin axis captures only the direction of the cisplatin effect, and MnTBAP has its own actions as a peroxynitrite scavenger and superoxide dismutase mimetic (Batinić-Haberle et al., 2009, 2010). Moreover, MnTBAP alone scored 0.41 on the same axis, close to the 0.51 of MnTBAP+Cisplatin. Therefore, part of the intermediate position may reflect effects of MnTBAP itself, or the composition of the MnTBAP preparations, rather than attenuation of cisplatin injury. Calibration with held-out Control and Cisplatin samples was essential, because any independent sample regresses toward the middle of an axis defined by selected proteins. Additionally, the largest contrast in the dataset, MnTBAP+Cisplatin versus Cisplatin, consisted mainly of bone and plasma proteins that were higher and synaptic proteins that were lower in MnTBAP+Cisplatin. This pattern reflects the composition of the MnTBAP+Cisplatin preparations, and only five bone and plasma proteins remained after adjustment. The under-loaded MnTBAP+Cisplatin 1 sample was excluded, and including it reduced this contrast further. Furthermore, the MnTBAP-alone contrast survived adjustment for the synaptic axis but was led by hematopoietic proteins, consistent with the marrow content of those preparations. Together, the present data do not support a proteomic conclusion about rescue by MnTBAP. These data do not contradict the physiological and histological evidence for its protective effect (Shahab et al., 2024, 2025), which rests on different measurements.

The consensus biomarker panel should be interpreted in light of composition. Several top-ranked decreased proteins, including AMACR, GPT2 and ADSL, correlated strongly with the synaptic axis, and MUG1 is a plasma protein. Therefore, the proteins least related to composition, such as TACC1, WDR47, PSMF1 and the nuclear RNA-processing proteins, are the more promising candidates for validation. Immunolabeling of synaptic markers in whole-mount cochleae and targeted proteomics with composition controls would be the natural next steps.

This study has several limitations. With four animals per group and three in the MnTBAP+Cisplatin group, every inferential test has limited power, and the exact permutation tests are coarse. Moreover, the marker panels were assembled from canonical proteins after inspection of the data. Their conclusions agree with the label-free PC1, but a pre-specified panel would be stronger. Additionally, composition adjustment cannot separate a true treatment effect that coincides with composition from a composition artifact. Finally, the dataset comes from a single TMT multiplex, which removes batch effects but prevents assessment of reproducibility across batches. In addition, the comparisons of panel scores between groups were not corrected for multiple testing and are descriptive.

These findings lead to practical recommendations for small-cohort tissue proteomics. Every feature selection step should be nested inside cross-validation, and performance should be compared with an exact or Monte Carlo permutation null. Furthermore, comparisons of new groups against a selected signature should be calibrated with held-out samples. More importantly, the composition of each preparation should be measured and reported, with marker panels for the expected contaminating tissues and an unsupervised check of the dominant axes of variation, and treatment effects should be tested with and without adjustment. In addition, balancing the timing and operator of dissections across groups, and recording protein yield and enrichment for each preparation, would reduce composition confounding at the

design stage. Moreover, pairing each proteome with the auditory brainstem response of the same animal would allow the proteome to be tested directly against synaptic function.

5. Conclusions

In conclusion, this study applied nested machine learning, exact permutation testing and composition analysis to the cochlear synaptosomal proteome after cisplatin and MnTBAP treatment. Differences in preparation composition formed the dominant axis of variation between samples and were confounded with the apparent cisplatin and MnTBAP effects. Collectively, the findings of this study suggest that tissue composition should be measured and modeled in synaptosomal proteomics before treatment-specific protein changes are interpreted.

Declarations

Animal Ethics Statement

This study is a secondary computational reanalysis of previously generated proteomics data. No new animal experiments were performed. The original animal procedures were carried out at Wayne State University and were approved and overseen by the Wayne State University Institutional Animal Care and Use Committee (IACUC) under protocol number [animal protocol # 17–02-214]. All procedures involving [Five-week-old male CBA/J mice] followed the NIH Guide for the Care and Use of Laboratory Animals. Full details of the animal procedures are published in Shahab et al., 2024 (Hearing Research) and in the doctoral dissertation of Shahab, 2023 (Wayne State University).

Author summary

Cisplatin is an effective anticancer drug, but it causes permanent hearing loss in many patients. The ribbon synapses that transmit sound signals from the inner hair cells to the auditory nerve are an early target of this damage. This study reanalyzed the abundance of thousands of proteins in synaptic fractions from mouse cochleae to determine whether cisplatin treatment can be recognized from the cochlear synaptic proteome. In addition, this study examined whether co-treatment with MnTBAP, a peroxynitrite scavenger, restores a normal protein profile. Each analysis step was repeated without the animal being predicted, and each result was compared against all possible relabelings of the animals. Under these conditions, the classification models did not recognize cisplatin treatment better than chance. Moreover, the largest source of variation between samples was the amount of synaptic tissue in each preparation relative to bone, blood and bone marrow. These differences in tissue content aligned with the treatment groups, and the apparent effect of cisplatin largely disappeared after they were taken into account. Because tissue content and treatment were linked, the data cannot fully separate the two. Together, these findings indicate that tissue composition is a measurable source of error in small inner ear proteomics studies and should be modeled in their analysis. The analysis code is openly available so that other groups can apply the same checks.

Data availability

The data analyzed in this study were generated at Wayne State University and are described in Shahab (2023) and Shahab et al. (2024). The doctoral dissertation is available at https://digitalcommons.wayne.edu/oa_dissertations/3846. Requests for access to the underlying data may be directed to the Office of the Vice President for Research, Wayne State University, Detroit, Michigan, USA (<https://research.wayne.edu>). All results derived from the data in this study are provided as Supplementary Tables S1 to S16.

Code availability

The analysis code (Python) that reproduces every number, table and figure in this article is openly available at <https://github.com/Monazzashahab/cochlear-synaptosome-composition-ml> under the MIT License.

Ethics statement

No new animal experiments were performed for this study. All animal procedures were completed for the original study and are described in Shahab (2023) and Shahab et al. (2024). This reanalysis did not involve human participants.

Acknowledgments

The author thanks Dr. Paul M. Stemmer and the Wayne State University Proteomics Core for performing the mass spectrometry and providing the Proteome Discoverer export analyzed in this study.

Funding

This study was conducted independently. The computational analysis and the development of the machine learning models were funded by the author's personal funds. No external funding was received for this study.

Conflict of interest

The author is the sole author of this study and declares no conflict of interest.

Consent for publication

Not applicable.

Author contributions

Monazza Shahab: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing (original draft, review and editing).

References

1. Ambroise C, McLachlan GJ (2002) Selection bias in gene extraction on the basis of microarray gene-expression data. *Proc Natl Acad Sci USA* 99:6562–6566
2. Batinić-Haberle I, Cuzzocrea S, Rebouças JS, Ferrer-Sueta G, Mazzon E, Di Paola R et al (2009) Pure MnTBAP selectively scavenges peroxynitrite over superoxide: comparison of pure and commercial MnTBAP samples to MnTE–2-PyP in two models of oxidative stress injury. *Free Radic Biol Med* 46:192–201
3. Batinić-Haberle I, Rebouças JS, Spasojević I (2010) Superoxide dismutase mimics: chemistry, pharmacology, and therapeutic potential. *Antioxid Redox Signal* 13:877–918
4. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B* 57:289–300
5. Breglio AM, Rusheen AE, Shide ED, Fernandez KA, Spielbauer KK, McLachlin KM et al (2017) Cisplatin is retained in the cochlea indefinitely following chemotherapy. *Nat Commun* 8:1654
6. Breiman L (2001) Random forests. *Mach Learn* 45:5–32
7. Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, Munafò MR (2013) Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci* 14:365–376
8. Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20:273–297
9. Dudoit S, Fridlyand J, Speed TP (2002) Comparison of discrimination methods for the classification of tumors using gene expression data. *J Am Stat Assoc* 97:77–87
10. Friedman JH (2001) Greedy function approximation: a gradient boosting machine. *Ann Stat* 29:1189–1232
11. Gagnon-Bartsch JA, Speed TP (2012) Using control genes to correct for unwanted variation in microarray data. *Biostatistics* 13:539–552
12. Geyer PE, Voytik E, Treit PV, Doll S, Kleinhempel A, Niu L et al (2019) Plasma proteome profiling to detect and avoid sample-related biases in biomarker studies. *EMBO Mol Med* 11:e10427
13. Hedges LV (1981) Distribution theory for Glass's estimator of effect size and related estimators. *J Educ Stat* 6:107–128
14. InMoose a Python port of limma, edgeR and related tools. <https://github.com/epigenelabs/inmoose> (version 0.9.1)
15. Kriegeskorte N, Simmons WK, Bellgowan PSF, Baker CI (2009) Circular analysis in systems neuroscience: the dangers of double dipping. *Nat Neurosci* 12:535–540
16. Kujawa SG, Liberman MC (2009) Adding insult to injury: cochlear nerve degeneration after temporary noise-induced hearing loss. *J Neurosci* 29:14077–14085

17. Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559
18. Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE et al (2010) Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet* 11:733–739
19. Leek JT, Storey JD (2007) Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet* 3:e161
20. Liberman MC, Kujawa SG (2017) Cochlear synaptopathy in acquired sensorineural hearing loss: manifestations and mechanisms. *Hear Res* 349:138–147
21. Meinshausen N, Bühlmann P (2010) Stability selection. *J R Stat Soc B* 72:417–473
22. Ojala M, Garriga GC (2010) Permutation tests for studying classifier performance. *J Mach Learn Res* 11:1833–1863
23. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O et al (2011) Scikit-learn: machine learning in Python. *J Mach Learn Res* 12:2825–2830
24. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43:e47
25. Rybak LP, Mukherjea D, Ramkumar V (2019) Mechanisms of cisplatin-induced ototoxicity and prevention. *Semin Hear* 40:197–204
26. Shahab M Role of nitrate stress in cisplatin induced ototoxicity and synaptic dysfunction [doctoral dissertation]. Detroit (MI): Wayne State University, 2023. Wayne State University Dissertations 3846. https://digitalcommons.wayne.edu/oa_dissertations/3846
27. Shahab M, Rosati R, Stemmer PM, Dombkowski A, Jamesdaniel S (2024) Quantitative profiling of cochlear synaptosomal proteins in cisplatin-induced synaptic dysfunction. *Hear Res* 447:109022. 10.1016/j.heares.2024.109022
28. Shahab M, Rosati R, Bhatia P, Doyon-Reale N, Jamesdaniel S (2025) MnTBAP, a peroxynitrite scavenger, attenuates cisplatin-induced apoptosis and cytotoxicity in organ of Corti cells. *Toxicol Rep* 14:101967
29. Shen-Orr SS, Gaujoux R (2013) Computational deconvolution: extracting cell type-specific information from heterogeneous samples. *Curr Opin Immunol* 25:571–578
30. Smyth GK (2004) Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 3:Article3
31. Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. *Proc Natl Acad Sci USA* 100:9440–9445
32. Tibshirani R, Hastie T, Narasimhan B, Chu G (2002) Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc Natl Acad Sci USA* 99:6567–6572
33. Varma S, Simon R (2006) Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics* 7:91
34. Zhang B, Horvath S (2005) A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* 4:Article17

Figures

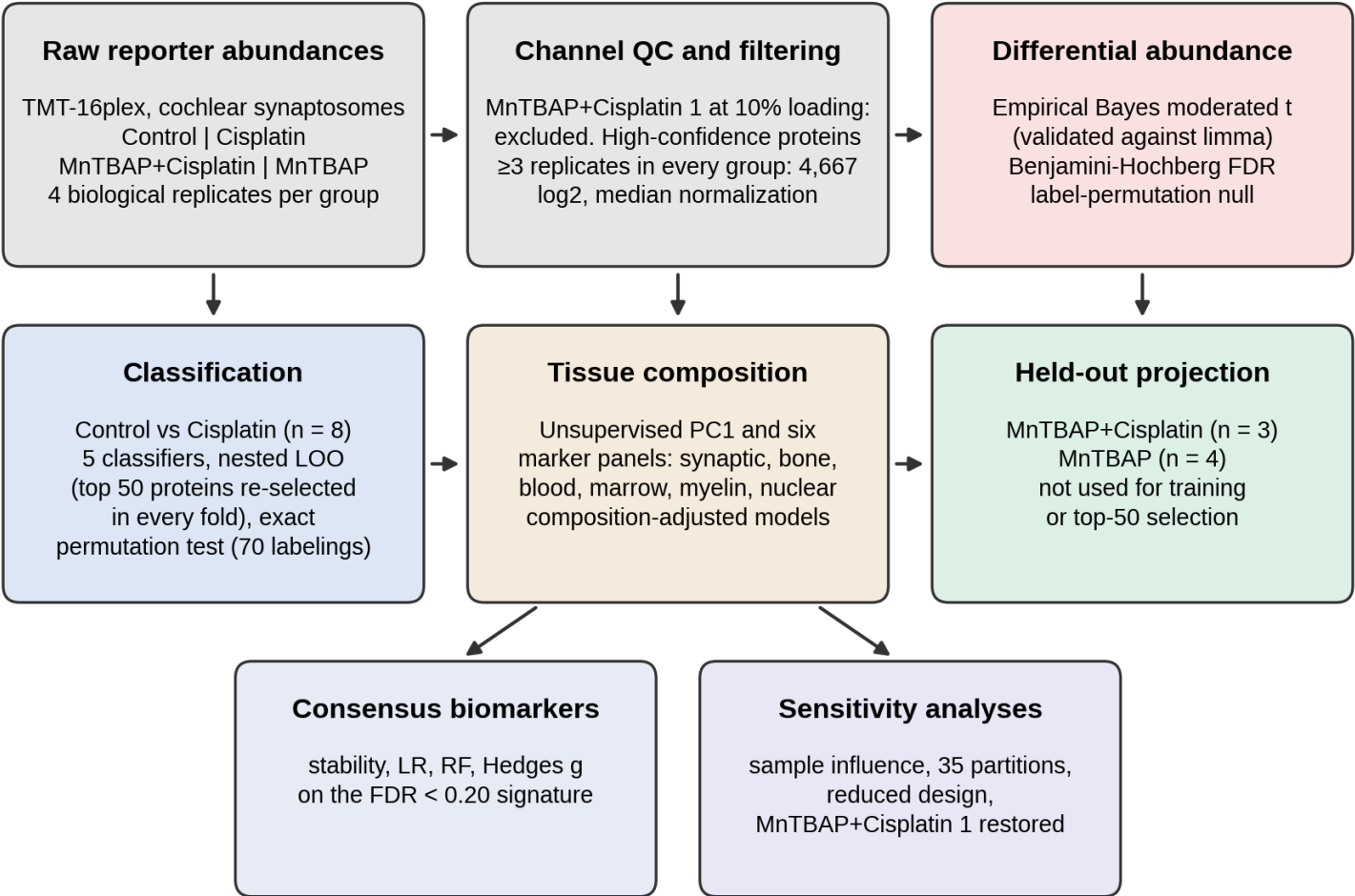


Figure 1

Study design and analysis workflow. Raw TMT reporter abundances from cochlear synaptosomes of four groups (Control, Cisplatin, MnTBAP+Cisplatin and MnTBAP, four biological replicates each) underwent channel quality control, filtering and median normalization. Differential abundance used empirical Bayes moderated t-tests with Benjamini-Hochberg FDR and a label-permutation null. Control and Cisplatin samples formed the discovery set for nested classification and consensus ranking. Tissue composition was quantified with an unsupervised principal component and six marker panels. The MnTBAP groups were held out and projected onto the cisplatin axis.

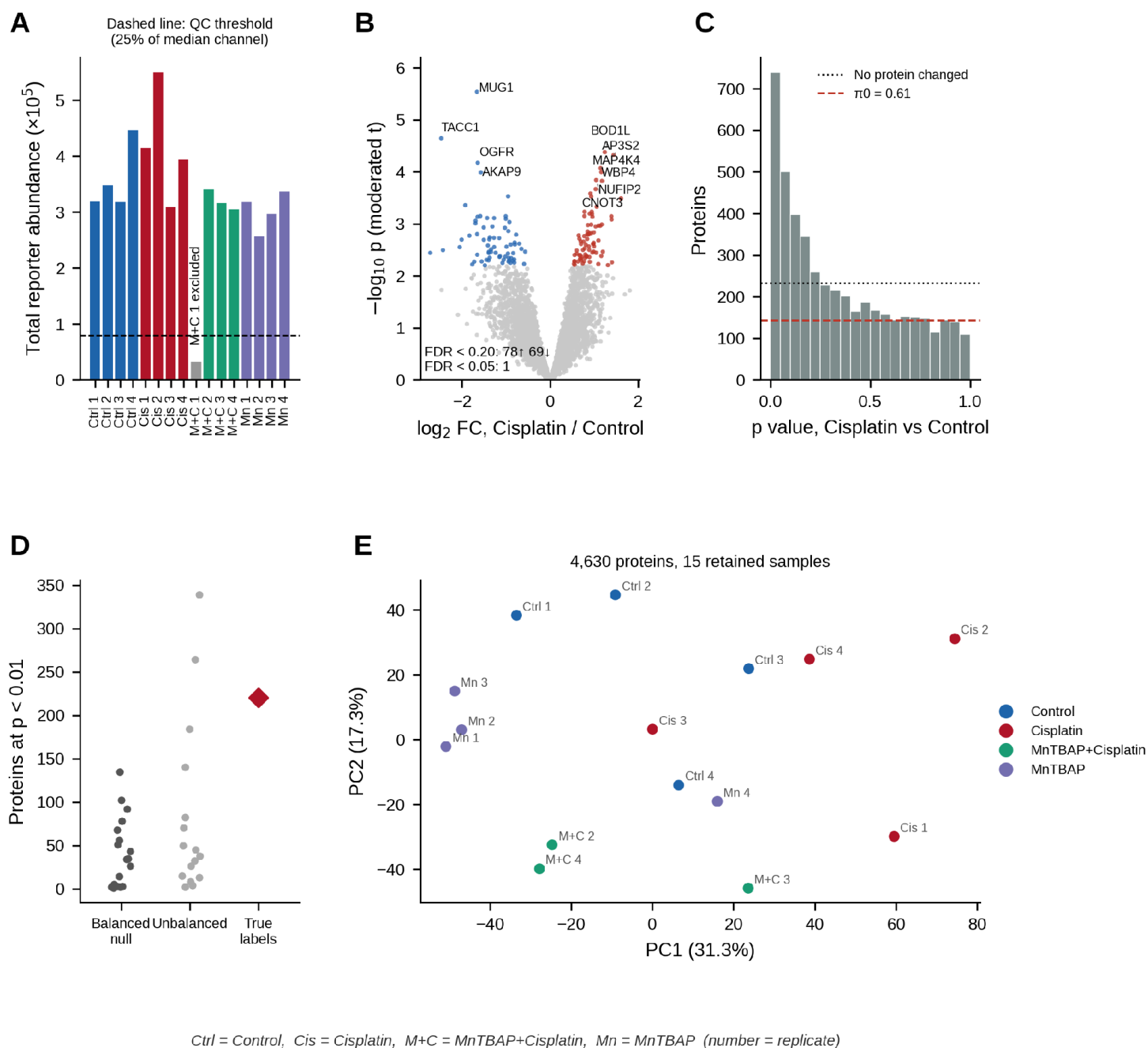


Figure 2

Quality control and the global Cisplatin versus Control signal. (A) Total reporter abundance per channel. Channel 130C (sample MnTBAP+Cisplatin 1) contained 10% of the median channel total and fell below the pre-specified threshold (dashed line, 25% of the median) and was excluded. (B) Volcano plot of 4,667 proteins, Cisplatin versus Control, moderated t-test. Red and blue, proteins increased (78) or decreased (69) at FDR < 0.20. Labels mark the ten most significant. (C) Distribution of Cisplatin versus Control p values. Dotted line, density expected if no protein changed. Dashed line, estimated null density ($\pi_0 = 0.61$). (D) Number of proteins at $p < 0.01$ under the true labels (diamond) and under the 34 other partitions of the eight samples into two groups of four. Dark grey, 18 balanced partitions (two Control and two Cisplatin samples on each side). Light grey, 16 unbalanced partitions. (E) Principal component analysis of 4,630

proteins with complete values across the 15 retained samples. Sample labels: Ctrl, Control. Cis, Cisplatin. M+C, MnTBAP+Cisplatin. Mn, MnTBAP. The number is the replicate.

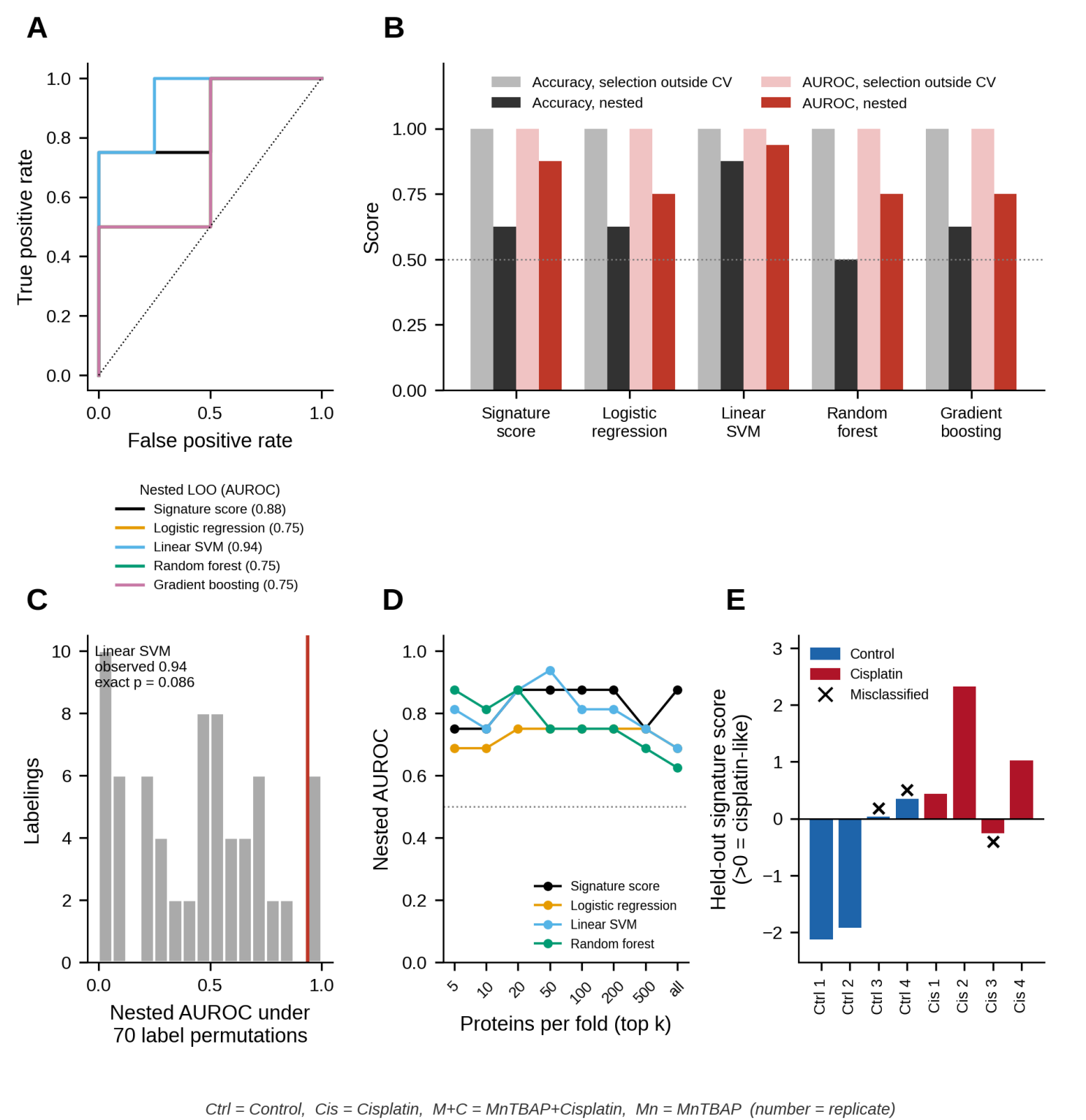
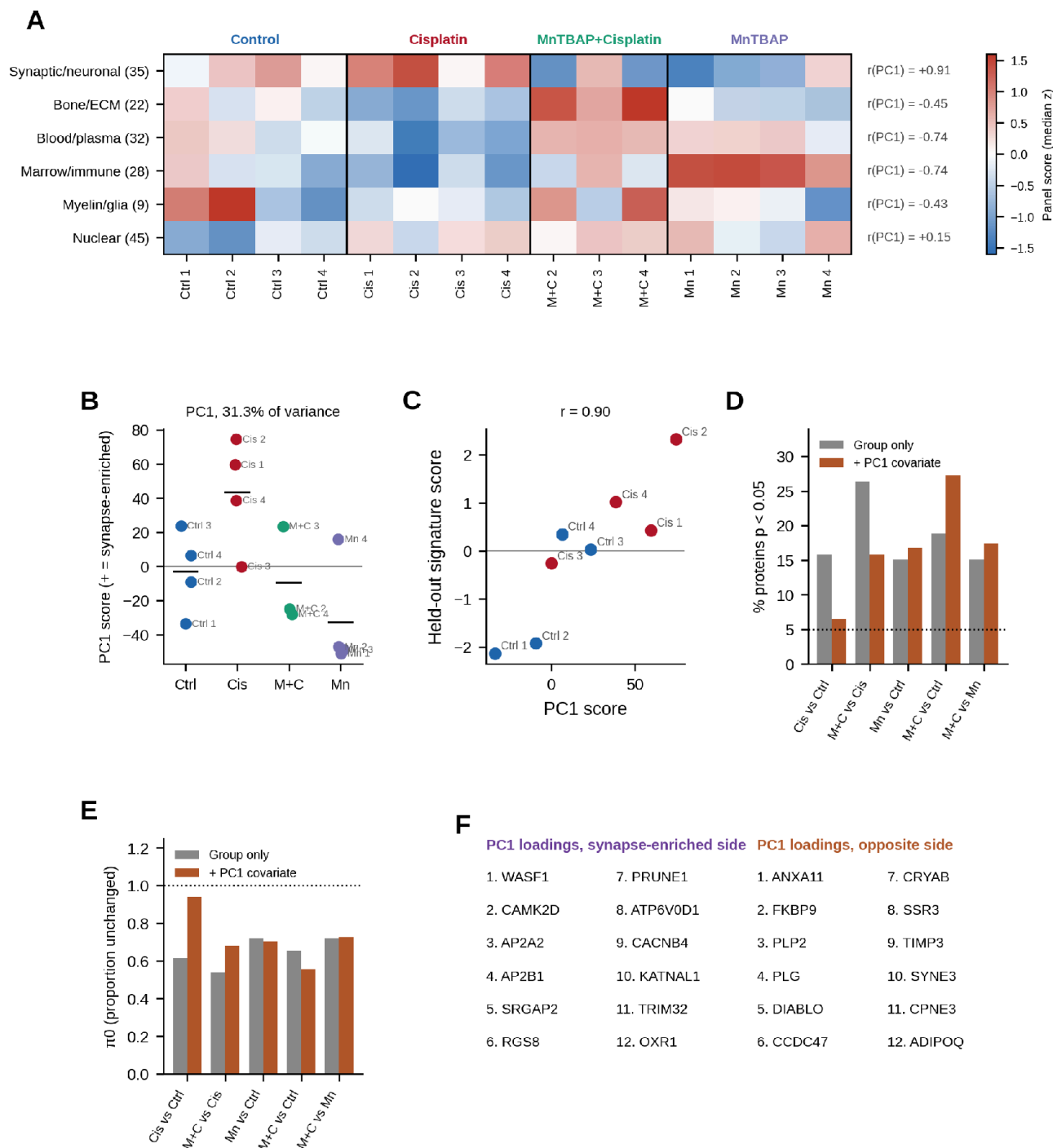


Figure 3

Nested cross-validation removes the apparent perfect classification of Cisplatin versus Control. (A) ROC curves from nested leave-one-out cross-validation with the top 50 proteins re-selected in each fold, for five

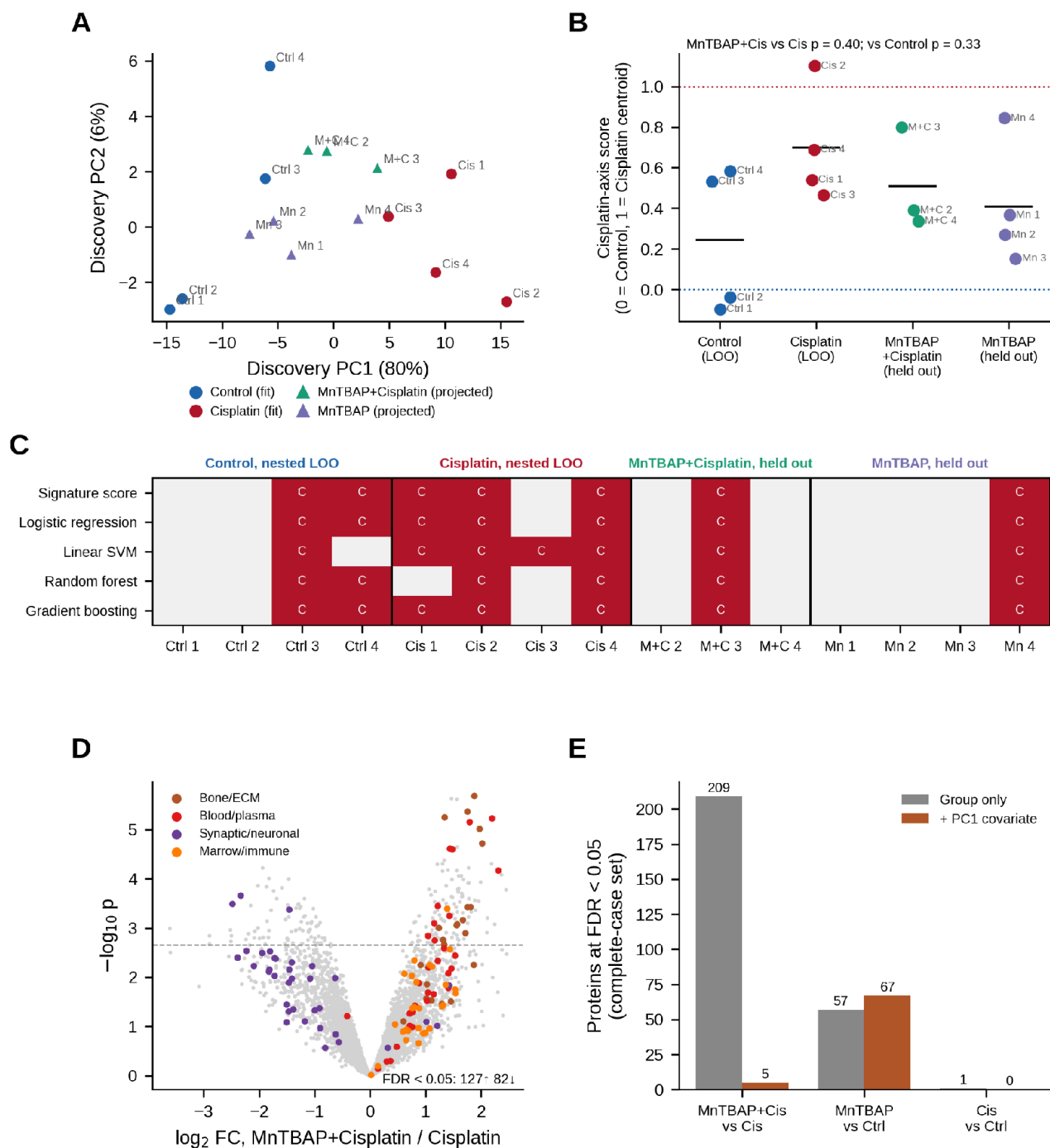
classifiers. AUROC in brackets. (B) Accuracy and AUROC with proteins selected on all eight samples before cross-validation (light bars) and with selection nested in each fold (dark bars). Dotted line, chance. (C) Null distribution of nested AUROC for the linear support vector machine over all 70 assignments of four Cisplatin labels among eight samples. Red line, observed AUROC. (D) Nested AUROC against the number of proteins retained per fold. "All", no selection. (E) Held-out signature score for each sample (positive, Cisplatin-like). Crosses mark misclassified samples. Sample labels: Ctrl, Control. Cis, Cisplatin. M+C, MnTBAP+Cisplatin. Mn, MnTBAP. The number is the replicate.



Ctrl = Control, Cis = Cisplatin, M+C = MnTBAP+Cisplatin, Mn = MnTBAP (number = replicate)

Figure 4

Preparation composition is the dominant source of variation. (A) Per-sample scores for six marker panels (median z-score of log2 abundance). Numbers of quantified markers in brackets. $r(\text{PC1})$, correlation of each panel score with the first principal component. (B) PC1 score by group, oriented so that positive values indicate synaptic enrichment. Lines, group means. (C) Held-out signature score of Control and Cisplatin samples against PC1 score. (D) Percentage of proteins at $p < 0.05$ for each contrast, with group terms only (grey) and with PC1 as a covariate (brown). Dotted line, 5% expected by chance. (E) Estimated proportion of unchanged proteins, π_0 , for the same models. (F) Proteins with the highest PC1 loadings on each side. Ctrl, Control. Cis, Cisplatin. M+C, MnTBAP+Cisplatin. Mn, MnTBAP. In sample labels, the number is the replicate.



Ctrl = Control, Cis = Cisplatin, M+C = MnTBAP+Cisplatin, Mn = MnTBAP (number = replicate)

Figure 5

Held-out projection of the MnTBAP groups. (A) Principal component analysis fitted on the signature proteins in Control and Cisplatin samples (circles), with MnTBAP+Cisplatin and MnTBAP samples projected (triangles). (B) Cisplatin-axis score, the projection onto the vector from the Control to the Cisplatin centroid using the top 50 proteins. Control and Cisplatin samples were scored leave-one-out and provide calibrated references. MnTBAP groups were never used for training or for top 50 selection. Lines,

group means. (C) Classification of every sample by each classifier ("C", Cisplatin-like). Control and Cisplatin by nested leave-one-out. MnTBAP groups by models trained on all eight Control and Cisplatin samples. (D) Volcano plot of MnTBAP+Cisplatin versus Cisplatin, with marker panel proteins colored. Dashed line, FDR = 0.05. (E) Number of proteins at FDR < 0.05 with group terms only (grey) and with PC1 as a covariate (brown). Sample labels: Ctrl, Control. Cis, Cisplatin. M+C, MnTBAP+Cisplatin. Mn, MnTBAP. The number is the replicate.

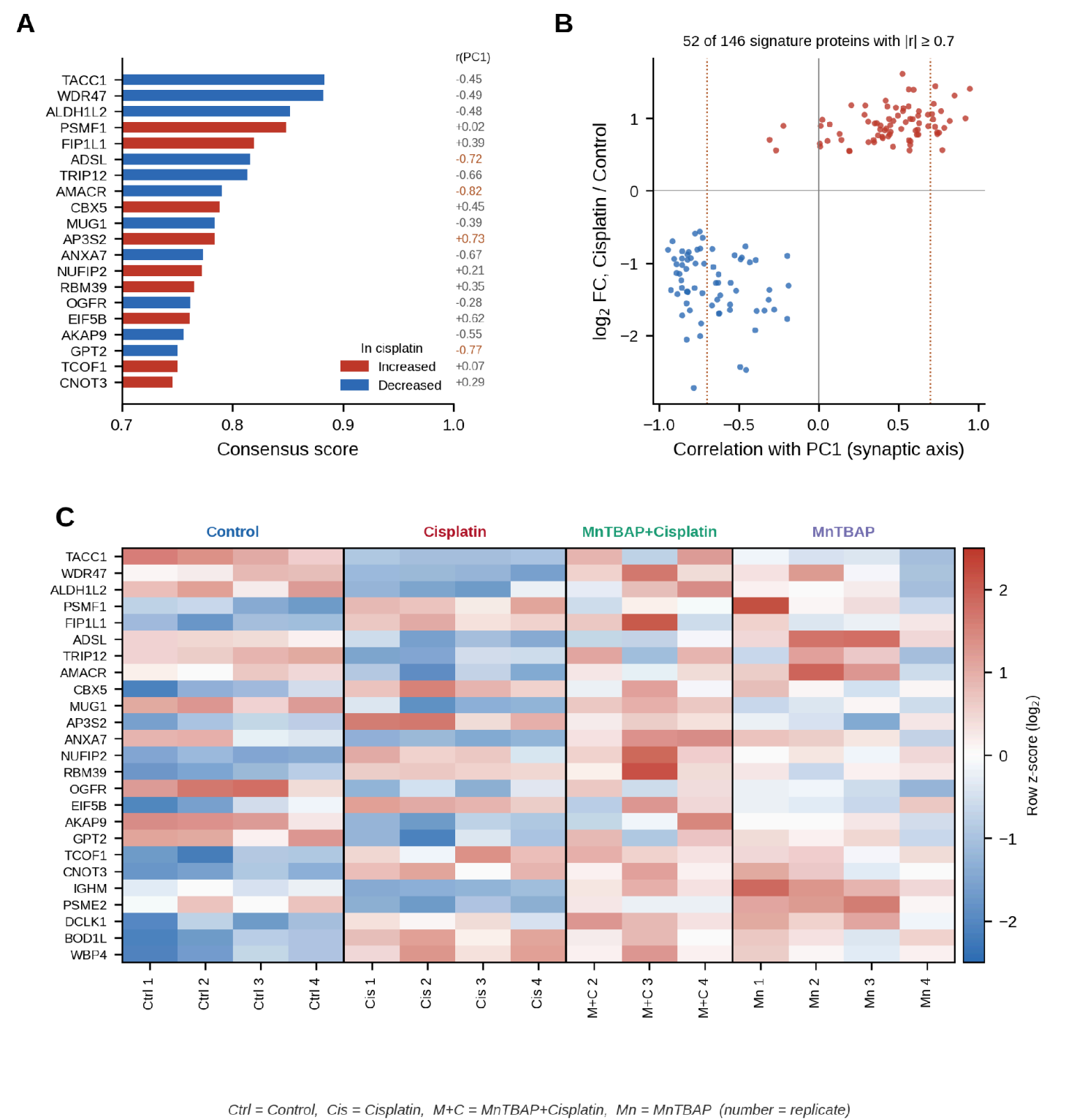


Figure 6

Consensus biomarkers of the cisplatin signature. (A) The 20 proteins with the highest consensus score (mean percentile rank across selection stability, logistic regression coefficient, random forest importance and Hedges' g). Red, increased. Blue, decreased in Cisplatin. Numbers at right give the correlation with PC1 (brown, $|r| \geq 0.7$). (B) Correlation with PC1 against log2 fold change for the 146 signature proteins. Dotted lines, $|r| = 0.7$. (C) Heatmap of the 25 top consensus proteins across the 15 retained samples (log2 abundance, z-scored per protein). Sample labels: Ctrl, Control. Cis, Cisplatin. M+C, MnTBAP+Cisplatin. Mn, MnTBAP. The number is the replicate.

Supplementary Files

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

- [FigureS1.png](#)
- [FigureS2.png](#)
- [FigureS3.png](#)
- [FigureS4.png](#)
- [FigureS5.png](#)
- [Table1.png](#)
- [Table2.png](#)
- [Table3.png](#)
- [Table4.png](#)
- [Table5.png](#)
- [Table6.png](#)