Supplementary Information
Cross-species transcriptomic analysis identifies candidate genes associated with diquat poisoning

Qiqi Liu[a],[1], Tingzhan Lv[a],[1], Yong Li[a], Min Yan[a], Ruijie Li[a], Kun Huang[a], Hongbo Liu[a],* and Yihong Yang[a],*

This file contains expanded methods, primer information (Supplementary Table S1), serum effect estimates (Supplementary Table S2), participant-level serum measurements (Supplementary Table S3), and dataset provenance (Supplementary Table S4).

Supplementary Methods

Human serum RNA sequencing and contrasts

Human serum RNA-seq profiles originated from the previously reported cohort of 17 participants [1]. The groups comprised six healthy controls (JKDZ), four poisoned patients without CNSI during hospitalization (NCNSI), four who developed CNSI after admission (CNSI1), and three with CNSI at admission (CNSI2). These are groups of different individuals, not longitudinal stages. Low-input RNA was used to prepare strand-specific poly(A)-selected libraries, sequenced as paired-end 150-bp reads. FastQC was used for read-quality inspection, fastp for adapter and low-quality trimming [2], HISAT2 for alignment to GRCh38 [3], and featureCounts for gene-level counting [4] with GENCODE annotation [5]. Low-count genes were removed before DESeq2 normalization and group-based modeling [6].

The contrasts were CNSI1 versus NCNSI and CNSI2 versus NCNSI. Differential expression was defined by Benjamini–Hochberg-adjusted $P < 0.05$ and absolute log2 fold change ≥ 1 [7]. This was a secondary transcriptomic analysis; prior metabolomic and proteomic features were not entered into the cross-species candidate-selection pipeline.

Zebrafish RNA sequencing and cross species mapping

The zebrafish dataset comprised control (C), 100 mg/L DQ (TL), and 175 mg/L DQ (TH) groups, with three independently prepared pools per group and 15 larvae per pool. Embryos were allocated at 1 hpf, exposed from 72 to 96 hpf, and collected at 96 hpf. Reads were summarized against GRCz11 annotation, and edgeR was used for differential expression [8]. TH versus C was the primary contrast; TL was retained for expression visualization. Of 532 genes meeting FDR < 0.05 and absolute log2 fold change ≥ 1, 237 mapped to human orthologs through Ensembl Compara/BioMart. One-to-many relationships were retained during direction checking and collapsed after mapping. Candidates required concordant directions of change in TH versus C and both human contrasts. The dataset is identified by BioProject PRJNA1358919 and was generated within the broader project; the present analysis is a secondary use of that dataset.

Functional annotation and association networks

Over-representation analyses used Gene Ontology Biological Process, KEGG, and Reactome through clusterProfiler and ReactomePA [9,10]. The background comprised unique mapped genes eligible for analysis. Benjamini–Hochberg-adjusted $P < 0.05$ defined significant enrichment; displayed terms above this threshold were treated as descriptive annotations. DQ-associated gene records were queried in CTD using the chemical term “Diquat” [11]. STRING v12.0 networks used Homo sapiens, a minimum combined score of 0.700, all active evidence channels, and up to five first-shell interactors per seed [12]. Database associations were not treated as evidence of physical interactions under DQ exposure.

Exploratory molecular docking

The DQ dication was obtained from PubChem (CID 6795) [13], with counterions omitted. Human ANXA2 (PDB 2HYW) and PRDX1 (PDB 9B7A) were retrieved from the RCSB Protein Data Bank [14]. Crystallographic water and non-target ligands were removed, polar hydrogen atoms were added, and receptor and ligand structures were converted to PDBQT. Five candidate regions were screened using AutoDock Vina v1.2.7, with nine poses generated per region [15,16]. Lowest-energy poses were visualized using PyMOL and Discovery Studio. The deposited PRDX1 structure contains rosmarinic acid. No redocking benchmark, negative-control docking, or experimental affinity measurement was performed; docking therefore provided structural hypotheses only.

PCNA imaging and zebrafish endpoint assays

All current targeted experiments used allocation at 1 hpf, exposure to 0, 100, or 175 mg/L DQ from 72 to 96 hpf, and collection at 96 hpf. Twenty Tg(elavl3:EGFP) larvae per group were fixed in 4% paraformaldehyde. After

dehydration, external-skin removal, rehydration, and acetone treatment at −20 °C for 6 h, larvae were washed in PBS containing 3% Triton X-100 and blocked with 4% BSA and 0.02% sodium azide. Anti-PCNA antibody (1:100) was followed by a fluorophore-conjugated secondary antibody and DAPI counterstaining. Images were acquired using a Leica TCS SP8 with consistent settings. Eight larvae per group were quantified in matched regions, with counts normalized to imaged area. The imaging cohort was separate from the earlier zebrafish work and from destructive endpoint assays.

MDA, SOD activity, and reduced GSH were measured in three independently prepared larval pools per group using commercial colorimetric kits. For qPCR, RNA from three independent pools was reverse-transcribed, and reactions were run in technical duplicate with SYBR Green chemistry. β-actin was the reference, and relative expression was calculated using the $2^{-\Delta\Delta Ct}$ method [17]. Exposure batches, pools, and measurements were independent of previous publications. The five newly supplied primer pairs were designed using Ensembl GRCz11 transcripts and checked against corresponding NCBI records (Supplementary Table S1).

Serum protein analysis

ANXA2, PRDX1, IL-6, and TNF-α were newly measured in archived serum from all 17 participants using commercial ELISA kits. Each value was normalized to the JKDZ mean for that marker. Effects were expressed as log2[mean(CNSI group)/mean(NCNSI)]. Unadjusted two-sided 95% BCa confidence intervals used 10,000 independent within-group resamples and seed 20260817 [18]. Exact independent-label permutation tests enumerated 70 allocations for CNSI1 versus NCNSI and 35 for CNSI2 versus NCNSI. Two-sided P values were twice the smaller tail probability, including ties, capped at 1; the observed allocation was included [19]. The eight P values were jointly adjusted using the Benjamini–Hochberg procedure [7].

Bootstrap intervals were unadjusted and were not obtained by inverting the permutation tests. For three-versus-four comparisons, the smallest possible two-sided P value under the stated convention is 2/35 (approximately 0.0571). This finite-sample constraint can produce positive effect estimates and intervals excluding zero alongside nonsignificant exact tests. No comparison was significant after multiple testing correction. These assays provide follow-up within the source cohort, not external clinical validation.

Supplementary Table S1

Primer sequences and available amplicon information

Gene	Forward primer 5′–3′	Reverse primer 5′–3′	Length (bp)	Predicted Tm F/R (°C)	Source
β-actin	AGCACGGTATTGTGACTAACTG	TCGAACATGATCTGTGTGATC	NR	NR	Prior project protocol
tnfa	TTTACCGCTGGTGATGGTGT	TGTTGATTGCCCTGGGTCTTA	NR	NR	Prior project protocol
il1b	GGCATGCGGGCAATATGAAG	TGTAGCTCATTGCAAGCCGA	116	60.04 / 60.04	Current primer record
il6	TCCGCTCAGAAAACAGTCCT	CATACTGCTGAACACGGGGA	90	59.89 / 59.75	Current primer record
anxa2a	ACGCCATGAAGAGCAAAGGA	GTCTGGTGCAGAGACTTCCC	127	59.96 / 60.04	Current primer record
prdx1	GCGGCAGATTACCATCAACG	TCCGGCTGGACAAACTTCTC	112	59.70 / 59.97	Current primer record
gna15.2	CCAATCCACAGAGCGAGGTT	CGCCGACTGTAGCAACTCTT	139	60.04 / 60.39	Current primer record

F, forward; R, reverse; NR, not recorded in the supplied primer record. Product lengths and predicted melting temperatures were available for five newly supplied pairs. Predicted Tm values are not experimentally measured amplification efficiencies.

Supplementary Table S2

Serum effect estimates and multiple testing results

Marker	Contrast	Log2 mean ratio	95% BCa CI	Exact P	BH-adjusted P
ANXA2	CNSI1 vs. NCNSI	0.4583	0.0654 to 0.8023	0.1143	0.1306
ANXA2	CNSI2 vs. NCNSI	0.6464	0.0564 to 1.2588	0.1714	0.1714
PRDX1	CNSI1 vs. NCNSI	0.6248	0.0678 to 1.0110	0.0857	0.1143
PRDX1	CNSI2 vs. NCNSI	0.8416	0.5137 to 1.1797	0.0571	0.0914
IL-6	CNSI1 vs. NCNSI	0.6795	0.3432 to 0.9955	0.0286	0.0914
IL-6	CNSI2 vs. NCNSI	0.9366	0.5212 to 1.3342	0.0571	0.0914
TNF-α	CNSI1 vs. NCNSI	0.4955	0.1046 to 0.7834	0.0571	0.0914
TNF-α	CNSI2 vs. NCNSI	0.7275	0.3772 to 0.9680	0.0571	0.0914

The effect estimate is log2[mean(CNSI group)/mean(NCNSI)]. The 95% BCa intervals are unadjusted; BH correction was applied jointly across the eight marker–group comparisons. Positive intervals and nonsignificant adjusted tests reflect different inferential procedures and should not be presented as equivalent criteria.

Group sizes were JKDZ, n = 6; NCNSI, n = 4; CNSI1, n = 4; and CNSI2, n = 3. Exact P values, effect estimates, and adjusted P values agree with calculations from the normalized observations in Supplementary Table S3 to the precision shown. The reported BCa intervals were retained from the original analysis.

Supplementary Table S3

De identified normalized serum measurements

Group	Subject	ANXA2	PRDX1	IL-6	TNF-α
JKDZ	JKDZ_01	1.270	1.281	1.104	0.936
JKDZ	JKDZ_02	0.918	1.195	0.943	0.988
JKDZ	JKDZ_03	0.981	0.834	0.991	1.120
JKDZ	JKDZ_04	0.809	0.678	0.908	1.134
JKDZ	JKDZ_05	0.881	0.855	0.963	0.753
JKDZ	JKDZ_06	1.142	1.156	1.091	1.070
NCNSI	NCNSI_01	1.504	1.049	1.154	1.211
NCNSI	NCNSI_02	0.980	1.321	1.011	1.097
NCNSI	NCNSI_03	1.210	1.546	1.616	1.462
NCNSI	NCNSI_04	0.906	0.883	1.339	1.110
CNSI1	CNSI1_01	1.886	2.062	1.677	1.227
CNSI1	CNSI1_02	1.767	2.293	2.479	1.552
CNSI1	CNSI1_03	1.227	1.135	1.878	2.053
CNSI1	CNSI1_04	1.440	1.910	2.166	2.048
CNSI2	CNSI2_01	2.576	1.881	2.388	1.623
CNSI2	CNSI2_02	1.180	2.432	3.075	2.206
CNSI2	CNSI2_03	1.644	2.137	1.887	2.231

Values were normalized to the JKDZ mean for each marker. Subject labels are study identifiers, not personal identifiers. Each row represents one participant. Rounded normalized values need not average to exactly 1.000.

JKDZ, healthy controls; NCNSI, poisoned patients without CNSI during hospitalization; CNSI1, patients who developed CNSI after admission; CNSI2, patients with CNSI at admission. Group labels do not denote longitudinal stages.

Supplementary Table S4

Dataset provenance and relationship to earlier work

Component	Provenance	Interpretation
Human serum RNA-seq	The same 17-participant cohort reported by Liu et al. [1].	Secondary analysis of CNSI-group contrasts; no new cohort was recruited.
Archived-serum ELISA	The same archived cohort samples; ANXA2, PRDX1, IL-6, and TNF-α assays are new to this article.	Exploratory follow-up within the source cohort; not independent clinical validation.
Zebrafish RNA-seq	BioProject PRJNA1358919, generated within the broader project.	Secondary cross-species ortholog analysis; not presented as newly generated sequencing data.
PCNA imaging	New images from an independently conducted exposure batch.	Current-study data; 20 larvae prepared and eight analyzed per group.
Biochemical assays and qPCR	New exposure batches, biological pools, and measurements.	Distinct from previous publication samples and measurements; three biological pools per group.
Docking, CTD, and STRING analyses	Analyses not reported in the supplied published cohort article.	Exploratory biological and structural context; not causal or binding evidence.

The current targeted zebrafish experiments used allocation at 1 hpf and exposure from 72 to 96 hpf. Their nominal concentrations matched the earlier project, but the previous longer exposure protocol and LC50 cannot be transferred directly to the current 24 h protocol. No metabolomic or proteomic feature from the earlier publication was used in the present candidate-selection pipeline.

References

References cited in this Supplementary Information are listed on the following page. Their numbering is independent of the main manuscript.

Supplementary References

1. Liu, Q. et al. Diquat-induced central nervous system injury: serum multi-omics and zebrafish evidence for histidine–methylmalonate imbalance and neuroinflammation. Chem. Biol. Interact. 432, 112046 (2026). https://doi.org/10.1016/j.cbi.2026.112046

2. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: An ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34, i884–i890 (2018). https://doi.org/10.1093/bioinformatics/bty560

3. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. 37, 907–915 (2019). https://doi.org/10.1038/s41587-019-0201-4

4. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923–930 (2014). https://doi.org/10.1093/bioinformatics/btt656

5. Frankish, A. et al. GENCODE reference annotation for the human and mouse genomes. Nucleic Acids Res. 47, D766–D773 (2019). https://doi.org/10.1093/nar/gky955

6. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550 (2014). https://doi.org/10.1186/s13059-014-0550-8

7. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. R. Stat. Soc. B 57, 289–300 (1995). https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

8. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140 (2010). https://doi.org/10.1093/bioinformatics/btp616

9. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: An R package for comparing biological themes among gene clusters. OMICS 16, 284–287 (2012). https://doi.org/10.1089/omi.2011.0118

10. Yu, G. & He, Q. Y. ReactomePA: An R/Bioconductor package for Reactome pathway analysis and visualization. Mol. BioSyst. 12, 477–479 (2016). https://doi.org/10.1039/C5MB00663E

11. Davis, A. P. et al. Comparative Toxicogenomics Database (CTD): Update 2023. Nucleic Acids Res. 51, D1257–D1262 (2023). https://doi.org/10.1093/nar/gkac833

12. Szklarczyk, D. et al. The STRING database in 2023: Protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res. 51, D638–D646 (2023). https://doi.org/10.1093/nar/gkac1000

13. Kim, S. et al. PubChem 2023 update. Nucleic Acids Res. 51, D1373–D1380 (2023). https://doi.org/10.1093/nar/gkac956

14. Burley, S. K. et al. RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. Nucleic Acids Res. 49, D437–D451 (2021). https://doi.org/10.1093/nar/gkaa1038

15. Eberhardt, J., Santos-Martins, D., Tillack, A. F. & Forli, S. AutoDock Vina 1.2.0: New docking methods, expanded force field, and Python bindings. J. Chem. Inf. Model. 61, 3891–3898 (2021). https://doi.org/10.1021/acs.jcim.1c00203

16. Trott, O. & Olson, A. J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J. Comput. Chem. 31, 455–461 (2010). https://doi.org/10.1002/jcc.21334

17. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta Ct}$ method. Methods 25, 402–408 (2001). https://doi.org/10.1006/meth.2001.1262

18. Efron, B. Better bootstrap confidence intervals. J. Am. Stat. Assoc. 82, 171–185 (1987). https://doi.org/10.1080/01621459.1987.10478410

19. Phipson, B. & Smyth, G. K. Permutation P-values should never be zero: Calculating exact P-values when permutations are randomly drawn. Stat. Appl. Genet. Mol. Biol. 9, Article 39 (2010). https://doi.org/10.2202/1544-6115.1585