cSVI-DM: a single-cell-guided, vascular-centered framework for layer-specific drug mapping in cerebral small vessel disease

Yuqian Li

yuqian-li@hotmail.com

Independent Researcher https://orcid.org/0009-0003-9256-9420

Method Article

Keywords:

Posted Date: June 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9937194/v1

Additional Declarations: The authors declare potential competing interests as follows: The authors declare potential competing interests as follows: The author is the developer of the cSVI-DM framework and the csviDM R package described in this preprint, and has filed related software copyright records. The package is publicly available for research use. The author declares no current financial competing interests.

cSVI-DM: a single-cell-guided, vascular-centered framework for layer-specific drug mapping in cerebral small vessel disease

Yuqian Li[1*]

[1] Independent Researcher, Beijing, China

*Corresponding author:

Yuqian Li, Independent Researcher, Beijing, China. Email: yuqian-li@hotmail.com

Abstract

Motivation:

Existing drug-screening workflows are commonly signature-first, target-first, response-prediction-first or structure-first. These strategies are effective for candidate ranking but rarely preserve the biological layer from which disease evidence originates. This is limiting for cerebral small vessel disease (cSVD), where vascular cell-state, source-to-vascular communication and terminal receptor/protein signals may correspond to different intervention scales.

Results:

We present cSVI-DM, a single-cell-guided, vascular-centered framework for layer-specific drug mapping in cSVD. Instead of using the final subtype barcode as the direct drug-screening input, cSVI-DM uses subtype-resolved CSS, CIS and LRS outputs as structured disease-side evidence and maps them to Macro, Balanced and Micro intervention channels. In the current v1.0 implementation, CSS-, CIS- and LRS-derived inputs were constructed for A1/N3KO-like and B1/BCAS-like contexts, whereas full pharmacological database mapping and structural validation were implemented for Micro. LRS-derived target seeds were linked to ChEMBL and BindingDB to generate subtype-specific pre-MD rankings, followed by docking, 5 ns molecular dynamics and MM/GBSA-based prioritization. The workflow identified an A1/NCL-centered local candidate pattern and a B1/FLT1-KDR-centered Micro candidate cluster.

Availability and implementation:

cSVI-DM is implemented in R and released as the csviDM workflow repository. Source code, documentation, example input files, and quick-start instructions are freely available at https://github.com/YuqianLii/csviDM. Future versions will further package cSVI-DM as an R module interoperable with cSVI-subtype, enabling CSS/CIS/LRS outputs generated by cSVI-subtype to be directly transferred into downstream Macro, Balanced and Micro drug-mapping workflows.

Contact:

yuqian-li@hotmail.com

Supplementary information:

Supplementary data are available at https://github.com/YuqianLii/csviDM. The cSVI-DM framework is archived at Zenodo under DOI: 10.5281/zenodo.20465204.

Introduction

Computational drug screening and repurposing are commonly supported by perturbational transcriptomic resources, drug-target knowledgebases, single-cell drug-response predictors, and structure-based validation workflows. Transcriptomic connectivity approaches, such as the Connectivity Map(Michnick, 2006), LINCS L1000(Subramanian, et al., 2017) and L1000CDS ² (Duan, et al., 2016), match disease-associated expression signatures with drug-induced perturbational signatures. Drug-target resources, including ChEMBL(Zdrazil, et al., 2024), DrugBank(Knox, et al., 2024), BindingDB(Liu, et al., 2025), and DGIdb(Cotto, et al., 2018), provide curated or experimentally supported compound-gene, compound-protein and binding-activity information. Single-cell-based tools, such as scDrug+(Sun, et al., 2024) and scDEAL(Chen, et al., 2022), extend drug-response prediction toward single-cell resolution, while docking, molecular-dynamics simulation and MM/GBSA estimation can evaluate specific protein-ligand hypotheses. Together, these methods have enabled large-scale candidate generation, target-level annotation and structural prioritization. However, they are usually optimized for a specific entry point, such as a disease signature, nominated target, predicted drug response or docked complex, and do not explicitly preserve the biological layer from which disease evidence is derived.

This limitation is relevant for heterogeneous vascular brain disorders such as cerebral small vessel disease (cSVD), where disease evidence may arise from vascular

cell states(Li, et al., 2025), source-to-vascular communication(Chen, et al., 2021), or terminal receptor/protein-level signals(Duering, et al., 2023; Hilkens, et al., 2024; Li, et al., 2025; Li, et al., 2022; Markus and Joutel, 2025). These different evidence layers may imply different intervention scales. Therefore, cSVD drug mapping requires a layer-aware strategy that translates disease-side evidence into therapeutic hypotheses without collapsing all signals into a single gene list or merged drug score.

Here, we present cSVI-DM, a single-cell-guided, vascular-centered framework for layer-specific drug mapping in cSVD. cSVI-DM was designed as the downstream drug-mapping module of cSVI-subtype(Li, 2026). Instead of using the final subtype barcode or footprint as the direct drug-screening input, cSVI-DM takes subtype-resolved CSS, CIS and LRS outputs as structured disease-side evidence. CSS represents vascular cell-state and abundance evidence, CIS represents source-to-vascular relay and communication-burden evidence, and LRS represents ligand-receptor-to-gene/protein terminal evidence. These layers are mapped to three intervention channels: Macro for state-remodeling hypotheses, Balanced for communication- or network-rebalancing hypotheses, and Micro for target-level small-molecule hypotheses. Thus, cSVI-DM generates layer-specific drug panels rather than a universal drug ranking.

In the current v1.0 implementation, CSS-, CIS- and LRS-derived inputs were constructed for two supported subtype contexts, A1/N3KO-like and B1/BCAS-like, while full pharmacological database mapping and structural validation were

implemented for the Micro layer. LRS-derived receptor/protein seeds were linked to ChEMBL(Zdrazil, et al., 2024) and BindingDB(Liu, et al., 2025) resources to generate subtype-specific pre-MD drug rankings. Prioritized Micro candidates were then advanced to docking, short-timescale molecular-dynamics simulation and MM/GBSA-based structural prioritization. This implementation identified an A1/NCL-centered local candidate pattern and a B1/FLT1-KDR-centered Micro candidate cluster, illustrating cSVI-DM as a reusable framework for translating vascular-centered single-cell disease evidence into subtype-aware therapeutic hypotheses.

Materials and methods

Overview of cSVI-DM and layer-specific input definition

cSVI-DM was designed as a downstream drug-mapping module of cSVI-subtype. For each supported subtype branch s, the direct cSVI-DM input was defined as a three-layer vascular-centered single-cell evidence set:

$$X_s = \{CSS_s, CIS_s, LRS_s\}$$

where CSS_s denotes vascular cell-state and abundance evidence, CIS_s denotes source-to-vascular relay and communication-burden evidence, and LRS_s denotes ligand-receptor-to-gene/protein terminal evidence. The final subtype barcode or footprint was not used as the primary drug-mapping input; it was retained only as a posterior interpretive anchor.

The three evidence layers were modeled as independent intervention channels:

$$CSS_s \Longrightarrow Macro$$

$$CIS_s \Longrightarrow Balanced$$

$$LRS_s \Longrightarrow Micro$$

Accordingly, Macro, Balanced, and Micro outputs were treated as layer-specific panels rather than components of a single merged drug score. The framework defines three independent outputs: Top20 Macro, Top20 Balanced, and Top20 Micro panels. In the current v1.0 implementation, CSS-, CIS-, and LRS-derived inputs were constructed for all three layers, whereas full pharmacological database mapping and structural validation were implemented for the Micro layer.

For input construction, cSVI-DM was restricted to the two supported subtype branches, A1/N3KO-like and B1/BCAS-like. Derived contrast groups, such as B1_minus_A1, were excluded from DM outputs. The preprocessing workflow generated group-level CSS/Macro input, sender-to-vascular CIS/Balanced input, pair-level LRS/Micro input, and receptor/protein-level Micro target tables. Receptor-complex information was retained at the pair-level LRS table, whereas the v1.0 Micro target table prioritized single-protein receptor or target seeds. This matches the documented preprocessing logic, where A1/B1 are retained, B1_minus_A1 is excluded, and the final barcode is not used as the DM main input.

Layer-specific object definitions

For a fixed subtype s, cSVI-DM defines three different object spaces.

Macro layer: CSS-derived state modules and state-remodeling drug mapping

The Macro layer operates on CSS-derived state modules. Let:

$$M_s = \{m_1, m_2, \cdots, m_M\}$$

denote the set of state modules derived from CSS_s. Each module m contains:

$$m = \{G_m, \omega_m, \delta_m, sev_m\}$$

where each state module m contains a module gene set G_m, a module weight ω_m, a module direction δ_m, and a module severity score sev_m. Module weights were normalized within each subtype s:

$$\sum_{m \in M_s} \omega_m = 1$$

Thus, Macro does not directly operate on a single receptor or ligand-receptor pair. It operates on vascular state modules derived from CSS.

Balanced-layer objects: CIS-derived relay axes

The Balanced layer operates on CIS-derived dominant relay axes. Let:

$$A_s = \{a_1, a_2, \cdots, a_A\}$$

denote the set of relay axes derived from CIS_S. Each relay axis a contains:

$$a = \{K_a, \omega_a, \delta_a, burden_a, fam_a\}$$

where K_a is the key-node set of the relay axis, ω_a is the axis weight, δ_a is the abnormal relay direction, $burden_a$ is the communication-burden score, and fam_a is the relay or pathway family. Axis weights are normalized within subtype s:

$$\sum_{a \in A_s} \omega_a = 1$$

Thus, Balanced does not target one terminal protein. It operates on source-to-vascular communication axes and asks which interventions may reduce abnormal relay burden or rebalance the network.

Micro-layer objects: LRS-derived targetable proteins

The Micro layer operates on LRS-derived terminal proteins or receptors. Let:

$$P_s = \{p_1, p_2, \cdots, p_P\}$$

denote the set of targetable proteins derived from LRS_s. Each protein p contains:

$$p = \{\omega_p, \delta_p, \tau_p, \rho_p\}$$

where ω_p is the protein weight, δ_p is the abnormal target direction, τ_p is the tractability score, and ρ_p is the structure-availability score. Protein weights are normalized within subtype s:

$$\sum_{p \in P_s} \omega_p = 1$$

Thus, Micro is the only layer designed to directly connect to target-level small molecules and protein-ligand structural validation.

Macro layer: CSS-to-Macro state-remodeling drug mapping

Macro primarily corresponds to state-remodeling resources, including natural active compounds, traditional-medicine-derived active ingredients, large-molecule or

phenotype-reversal interventions, and other resources that may remodel disease-associated vascular state.

Let the Macro candidate set be:

$$D_{Macro} = \{d_1, d_2, \cdots, d_D\}$$

Each Macro candidate d contains:

$$d = \left\{T_d^{Macro}, E_d^{Macro}, R_d\right\}$$

where T_d^{Macro} is the target or response gene set associated with the candidate, E_d^{Macro} is the evidence strength, and R_d denotes external resource or phenotype-reversal support when available.

For candidate d and CSS-derived state module m, module coverage is defined as:

$$Cov_M(d, m) = \frac{\sum_{g \in G_m \cap T_d^{Macro}} \omega_g}{\sum_{g \in G_m} \omega_g + \epsilon}$$

where G_m is the gene set of module mmm, T_d^{Macro} is the candidate-associated target or response gene set, ω_g is the within-module gene importance weight, and ϵ is a small constant used to avoid division by zero.

Direction consistency is defined as:

$$Dir_M(d, m) = \begin{cases} 1, & if\ d\ is\ expected\ to\ reverse\ module\ m \\ 0.5, & if\ the\ direction\ is\ unknown \\ 0, & if\ d\ is\ expected\ to\ reinforce\ the\ disease\ direction \end{cases}$$

If external phenotype-reversal evidence is available, it is represented as:

$$Rev(d, m) \in [0, 1]$$

If no external phenotype-reversal evidence is available, the default value is:

$$Rev(d, m) = 1$$

The Macro score is then defined as:

$$S_{Macro}(d) = \sum_{m \in M_s} \omega_m \cdot Cov_M(d, m) \cdot Dir_M(d, m) \cdot Rev(d, m) \cdot E_d^{Macro}$$

Candidates are ranked only within the Macro candidate set:

$$Rank_{Macro}(d) = rank_{d \in D_{Macro}}\left(S_{Macro}(d)\right)$$

and the output is the Top20 Macro panel.

In the current v1.0 implementation, the CSS-derived input strength used for Figure 2A, such as:

$$MacroInput_s = MeanVascularFraactions_s \times MeanAbsLogFC_s$$

was used only to summarize the strength of the CSS-derived Macro entry point. It is not the formal Macro drug score. The formal Macro drug score is $S_{Macro}(d)$ as defined above.

Balanced layer: CIS-to-Balanced relay-balancing drug mapping

Balanced primarily corresponds to hormone-like drugs, systemic regulatory drugs, and network-balancing intervention resources.

Let the Balanced candidate set be:

$$D_{Balanced} = \{d_1, d_2, \cdots, d_D\}$$

Each Balanced candidate d contains:

$$d = \left\{T_d^{Balanced}, E_d^{Balanced}, Reg_d\right\}$$

where $T_d^{Balanced}$ is the known action-node set of candidate d, $E_d^{Balanced}$ is the system-regulation evidence strength, and Reg_d represents the broad regulatory coefficient or prior.

For candidate d and relay axis a , relay-axis coverage is defined as:

$$Cov_B(d, a) = \frac{\sum_{k \in K_a \cap T_d^{Balanced}} \eta_k}{\sum_{k \in K_a} \eta_k + \epsilon}$$

where K_a is the key-node set of relay axis a, $T_d^{Balanced}$ is the action-node set of candidate d, η_k is the node-level weight, and ϵ is a small constant.

Relay-direction consistency is defined as:

$$Dir_B(d, a) = \begin{cases} 1, & if\ d\ is\ expected\ to\ weaken\ the\ abnormal\ relay \\ 0.5, & if\ the\ direction\ is\ unknown \\ 0, & if\ d\ may\ reinforce\ the\ abnormal\ relay \end{cases}$$

If a candidate covers multiple dominant relay axes, a broad multi-axis regulation reward is applied:

$$Broad_B(d) = 1 + \lambda \cdot log\big(1 + n_a(d)\big)$$

where $n_a(d)$ is the number of relay axes covered by candidate d, and λ is the broad-regulation parameter.

The Balanced score is defined as:

$$S_{Balanced}(d) = Broad_B(d) \cdot \sum_{a \in A_s} \omega_a \cdot Cov_B(d, a) \cdot Dir_B(d, a) \cdot E_d^{Balanced}$$

Candidates are ranked only within the Balanced candidate set:

$$Rank_{Balanced}(d) = rank_{d \in D_{Balanced}}\big(S_{Balanced}(d)\big)$$

and the output is the Top20 Balanced panel.

In the current v1.0 implementation, quantities such as $\text{BalancedAxisScore}_a$, Relay_a, and Burden_a were used to construct and prioritize CIS-derived relay-axis inputs. They are not themselves Balanced drug scores. The formal Balanced drug score is $S_{Balanced}(d)$, which requires candidate-level action-node coverage and evidence.

Micro layer: LRS-to-Micro target-level small-molecule drug mapping

Micro primarily corresponds to target-level small molecules and is the only layer that directly proceeds to protein-ligand docking and molecular-dynamics validation.

Let the Micro candidate set be:

$$D_{Micro} = \{d_1, d_2, \cdots, d_D\}$$

For each candidate d and target protein p, the Micro layer uses:

$$d, p = \left\{A_{d,p}, E_d^{Micro}, M_{d,p}\right\}$$

where $A_{d,p}$ is the activity value between candidate d and protein p, E_d^{Micro} is the drug-target evidence strength, and $M_{d,p}$ is the mechanism annotation, such as inhibitor, antagonist, agonist, or other mechanism classes.

Hit function

If candidate d hits protein p, the activity-normalized hit score is defined as:

$$Hit_X(d, p) = \begin{cases} min\left(1, \dfrac{4 - log_{10} A_{d,p,}}{4}\right), & if\ quantitative\ activity\ is\ available \\ 0.4, & if\ only\ qualitative\ hit\ annotation\ is\ available \\ 0, & if\ no\ hit\ is\ available \end{cases}$$

where $A_{d,p}$ is interpreted on the nM scale. This transformation gives stronger scores to lower activity values and caps the hit score at 1.

Mechanism-direction consistency

Direction consistency in the Micro layer is defined as:

$$Dir_X(d,p) = \begin{cases} 1, & if\ the\ drug\ mechanism\ opposes\ the\ disease - protein\ direction \\ 0.5, & if\ the\ direction\ is\ unknown \\ 0, & if\ the\ drug\ mechanism\ may\ reinforce\ the\ disease - protein\ direction \end{cases}$$

Structural-verifiability term

The structure-verifiability term is defined as:

$$Struct(p) = \tau_p \cdot \rho_p$$

where τ_p is the tractability score and ρ_p is the structure-availability score. Both are normalized to $[0, 1]$.

Micro score

The Micro score for candidate *d* is defined as:

$$S_{Micro}(d) = \sum_{p \in P_s} \omega_p \cdot Hit_X(d,p) \cdot Dir_X(d,p) \cdot E_{d,p}^{Micro} \cdot Struct(p)$$

Candidates are ranked only within the Micro candidate set:

$$Rank_{Micro}(d) = rank_{d \in D_{Micro}}\left(S_{Micro}(d)\right)$$

and the output is the Top20 Micro panel.

In the current Micro-v1 implementation, LRS-derived receptor/protein seeds were connected to ChEMBL and BindingDB resources to generate subtype-specific

pre-MD Micro rankings. Structural readiness was not incorporated into the pre-MD pharmacological ranking step; instead, structure readiness and structural validation were handled downstream during MD-ready candidate selection and molecular simulation.

MD-ready candidate selection and molecular-dynamics validation

MD was implemented as a downstream structural-validation module for the Micro layer only. Macro and Balanced panels were retained for state-level and relay-level interpretation and were not used as direct MD inputs. This follows the cSVI-DM design in which only Micro candidates, derived from LRS-supported targetable proteins and small-molecule hits, are eligible for protein-ligand structural validation.

MD entry rule

A candidate drug-target pair (d, p) was eligible for MD only if it was derived from the Micro candidate set:

$$d \in TopK(D_{Micro})$$

and satisfies three entry conditions:

$$Hit_X(d, p) > \theta_h$$

$$Struct(p) > \theta_s$$

$$E_{d,p}^{Micro} > \theta_e$$

where $Hit_X(d, p)$ is the normalized drug-target hit score, $Struct(p)$ is the structural-verifiability score of protein p, $E_{d,p}^{Micro}$ is the drug-target evidence

strength, and θ_h, θ_s, and θ_e are predefined thresholds for hit strength, structure readiness, and evidence strength.

The formal MD priority score is defined as:

$$MDPriority(d, p) = \omega_p \cdot Hit_X(d, p) \cdot Struct(p) \cdot E_{d,p}^{Micro}$$

where ω_p is the LRS-derived protein weight. Candidate pairs are ranked by $MDPriority(d, p)$, and the top-ranked pairs are selected for docking and molecular-dynamics simulation. The original formula document defines MD as a Micro-only downstream step and specifies this priority rule using protein weight, hit strength, structural-verifiability, and Micro evidence strength.

v1.0 MD-ready candidate selection used in this study

In the current v1.0 implementation, MD-ready candidate selection was implemented as a pragmatic Micro-only proxy of the formal entry rule. Candidate pairs were required to originate from LRS-derived Micro targets, have positive subtype-level Micro target support, be mapped to pharmacological drug-target records, and have a positive Micro drug-target score.

Specifically, a candidate pair *(d, p)* was retained for MD consideration when:

$$MicroTargetScore_p > 0$$

$$S_{MicroTarget(d,p)} > 0$$

and the target was annotated as an actionable Micro target in the current v1.0 target table. The Micro drug-target score had the form:

$$S_{MicroTarget(d,p)} = TargetSignal_p \cdot Hit(d, p) \cdot Direction(d, p) \cdot Evidence(d, p)$$

Thus, $S_{MicroTarget(d,p)} > 0$ served as a positive-score proxy for non-zero target support, drug-target hit evidence, direction compatibility, and database evidence. Candidate pairs were then prioritized using Micro drug-target score, drug-level Micro score, and activity strength. Formal numeric $Struct(p)$ -thresholding and explicit $MDPriority(d, p)$ scoring were not fully automated in this v1.0 implementation; instead, structural readiness was handled during downstream protein-fragment selection, structure availability assessment, and docking feasibility screening.

This distinction is important: the present study follows the Micro-only entry logic of cSVI-DM, but implements MD input selection as a positive-score, structure-ready v1.0 proxy rather than a fully thresholded MDPriority implementation.

Candidate selection for structural validation

MD-ready candidates were selected from the subtype-specific Micro ranking and positive Micro drug-target pairs. Candidate selection considered LRS-derived target support, pharmacological activity evidence, database mappability, and structural feasibility.

For the A1/N3KO-like branch, NCL-centered candidates were selected:

$$NCL - CHEMBL1232461$$

$$NCL - CHEMBL3752910$$

For the B1/BCAS-like branch, FLT1/KDR-centered candidates were selected from the VEGFR-like Micro candidate cluster:

$$FLT1 - CHEMBL535$$

$$KDR - CHEMBL535$$

$$FLT1 - CHEMBL587723$$

$$KDR - CHEMBL587723$$

These candidates represented the transition from LRS-derived Micro target seeds to pharmacologically mapped, docking-ready drug-target hypotheses. The project workflow records this structural-validation sequence as cSVI-subtype CSS/CIS/LRS output, DM input construction, Macro/Balanced/Micro mapping, Micro target seed, ChEMBL drug-target ranking, DiffDock pose, GROMACS 5 ns MD, RMSD/RMSF/Rg/SASA/H-bond/PCA analysis, and MMPBSA.

Protein and ligand preparation

Protein structures or fragments were selected according to target size, domain interpretability, and structure availability. For NCL, available ligand-binding or RRM-like structural regions were prioritized over full-length NCL. For FLT1 and KDR, full-length receptors were not used directly because of their large size and domain complexity. Instead, extracellular-like and kinase-domain-like fragments were evaluated. The project records define FLT1 1-687 and KDR 1-678 as AlphaFoldDB-derived fragments, and KDR 806-1171 as a PDB-derived kinase-domain-like fragment; FLT1 kinase-domain-like structures were also derived from PDB resources, although the exact final residue range should be kept consistent with the simulated structure used in Figure 4.

In the structural-validation set, FLT1 was represented by:

$$FLT1_{1-687} \quad and \quad FLT1_{801-1185}$$

KDR was represented by:

$$KDR_{1-678} \quad and \quad KDR_{806-1171}$$

PDB-derived structures were repaired when necessary by adding missing heavy atoms and removing heterogens before topology generation. Ligand structures were prepared from the corresponding ChEMBL identifiers. Ligand SDF files were converted to PDB format using OpenBabel, and ligand topologies were generated using ACPYPE in GROMACS-compatible format.

Docking

Protein-ligand docking poses were generated using DiffDock(Corso, et al., 2022). Docking was used as the pose-generation step for downstream molecular dynamics and was not interpreted as standalone evidence of stable binding. Selected docked poses were used as initial conformations for MD simulations.

Molecular-dynamics simulation setup

Molecular dynamics simulations were performed using GROMACS(Pronk, et al., 2013). Ligand topology generation was performed using ACPYPE(Kagami, et al., 2023) with a neutral ligand charge setting and GROMACS-compatible output. Protein topologies were generated using pdb2gmx with the amber99sb-ildn(Lindorff-Larsen, et al., 2010) force field. In the uploaded production workflow, pdb2gmx used the SPC/E water model(Berendsen, et al., 2002), and systems were solvated with spc216

water. Protein-ligand coordinates were merged while preserving the docked ligand pose.

Each protein-ligand complex was placed in a cubic simulation box with a 1.0 nm margin, solvated, and neutralized using Na+ or Cl− ions. Ion and temperature-coupling group definitions were checked against the actual system composition before production simulation. Energy minimization was performed using the steepest-descent algorithm for up to 50,000 steps, with an energy minimization tolerance of: $emtol = 1000\ kJ\ mol^{-1}\ nm^{-1}$, and a step size of: $emstep = 0.01$.

Each system then underwent two equilibration stages. The first equilibration stage used: $nsteps = 50{,}000$, $dt = 0.0005\ ps$. The second equilibration stage used: $nsteps = 50{,}000$, $dt = 0.002\ ps$. Production MD was performed for: $nsteps = 2{,}500{,}000$ with $dt = 0.002\ ps$, corresponding to: $5\ ns$.

Production simulations used V-rescale temperature coupling at: $T = 300\ K$, and Parrinello-Rahman isotropic pressure coupling at: $P = 1\ bar$. Bonds involving hydrogen atoms were constrained using LINCS. Long-range electrostatics were treated using PME, with a short-range electrostatic cutoff of: $r_{coulomb} = 1.2\ nm$, Van der Waals interactions were treated using a cutoff scheme with force-switching from: $r_{vdw-switch} = 1.0\ nm$ to $r_{vdw} = 1.2\ nm$, Production coordinates, energies, and log files were saved every $10\ ps$.

Trajectory analysis

Trajectory-level stability was evaluated using: Protein RMSD, Ligand RMSD, Residue RMSF, Radius of gyration(Rg), Solvent-accessible surface area(SASA), Hydrogen bond count, PCA projection, Eigenvalue-based motion summaries.

Ligand-pocket distance was additionally evaluated when pocket-distance output was available. These trajectory metrics were used to classify each complex as dynamically stable, flexible, rearranged, or cautionary. The same categories were then summarized in Figure 4 as the MD stability and caution matrix.

MM/GBSA binding-energy estimation

Binding-energy estimates were obtained using gmx_MMPBSA v1.4.3 based on MMPBSA.py v16.0. MM/GBSA calculations were performed using the Generalized Born model. The mmpbsa.in file specified: $startframe = 100$, $endframe = 300$, $interval = 5$, $igb = 5$, $saltcon = 0.150$, $probe = 1.4$. This corresponds to analysis of frames 100 to 300 at an interval of 5, yielding 41 frames per calculation when the full frame range was available, using a 0.150 M salt concentration and a 1.4 Å probe radius.

The reported DELTA TOTAL value was interpreted as the MM/GBSA binding-energy estimate:

$$\Delta G_{bind}^{MM/GBSA} = \Delta G_{TOTAL}$$

These binding-energy estimates were used for relative computational prioritization across simulated complexes. They were not interpreted as experimental binding affinities. Candidate prioritization was based on the combined interpretation of:

$$\Delta G_{bind}^{MM/GBSA}$$

trajectory stability, hydrogen-bond behavior, pocket-distance caution, and target-level subtype support.

Results

cSVI-DM establishes a layer-specific drug-mapping framework downstream of cSVI-subtype

To extend vascular-centered subtype modeling toward therapeutic hypothesis generation, we developed cSVI-DM as a downstream drug-mapping module of the cSVI-subtype framework (Figure 1). Rather than using the final subtype barcode as a single drug-screening input, cSVI-DM preserves the upstream CSS-CIS-LRS structure generated by cSVI-subtype and reuses these three layers as distinct intervention entry points.

As shown in Figure 1, the input side of cSVI-DM consists of vascular-centered single-cell RNA-seq evidence derived from cSVI-subtype. The CSS layer summarizes vascular cell-state evidence, including differentially expressed genes, expression prevalence, and vascular cell-ratio information. The CIS layer captures cell-cell communication evidence, including source-to-vascular interaction weight and interaction number. The LRS layer represents ligand-receptor allocation evidence and provides terminal receptor/protein-level signals. Therefore, the direct input of

cSVI-DM is not a compressed subtype footprint, but the structured CSS, CIS, and LRS outputs generated during subtype construction.

cSVI-DM then maps these three evidence layers to three corresponding drug-mapping strategies. CSS-derived evidence is extended to the Macro mapping layer, which is designed to prioritize state-remodeling interventions. CIS-derived evidence is extended to the Balanced mapping layer, which is designed to prioritize network-balancing or communication-modulating interventions. LRS-derived evidence is extended to the Micro mapping layer, which is designed to prioritize target-level small-molecule drug candidates. This layer-determined mapping strategy allows each disease-evidence layer to be interpreted according to its corresponding intervention scale, instead of being forced into a single integrated drug score.

The middle part of Figure 1 further shows that cSVI-DM connects subtype-derived evidence to both pharmacological and structural resources. Drug databases, including ChEMBL(Zdrazil, et al., 2024), and BindingDB(Liu, et al., 2025) are used for drug-target evidence retrieval and candidate ranking, while protein and ligand resources provide the basis for downstream structural validation. Through this design, cSVI-DM links single-cell-derived disease evidence with drug-target knowledge and protein-ligand validation in one continuous workflow.

The primary output of cSVI-DM is a set of layer-specific drug panels rather than a universal drug list. The Macro, Balanced, and Micro layers each generate an independent Top20 panel, corresponding respectively to state-remodeling,

network-balancing, and target-level intervention hypotheses. Among these layers, Micro candidates are further advanced to docking, molecular dynamics simulation, and MM/GBSA-based binding-energy estimation, producing subtype-aware drug-target hypotheses that can be prioritized for future validation.

Together, Figure 1 defines the overall logic of cSVI-DM: cSVI-subtype constructs a vascular-centered disease model, whereas cSVI-DM extends this model toward intervention-side interpretation. In this framework, CSS, CIS, and LRS are no longer treated only as subtype-defining evidence layers, but are repurposed as Macro, Balanced, and Micro drug-mapping channels. This establishes cSVI-DM as a layer-specific and subtype-aware computational framework for therapeutic hypothesis generation in cerebral small vessel disease.

Construction of CSS-CIS-LRS-derived cSVI-DM inputs from vascular-centered scRNA evidence

We next organized the upstream cSVI-subtype outputs into layer-specific inputs for cSVI-DM (Figure 2). According to the preprocessing design, the DM input was restricted to the two supported subtype branches, A1/N3KO-like and B1/BCAS-like. This step generated CSS-, CIS-, and LRS-derived intermediate files, including CSS input tables, CIS input and top-axis tables, LRS pair-level input, and LRS target-level input. The final subtype barcode was not used as the main DM input at this stage; instead, the Micro layer was first compressed toward single-protein receptor seeds,

while receptor-complex information was retained at the pair-level table. For the CSS-derived Macro input, A1 showed a higher vascular fraction and higher state-layer score than B1 (Figure 2A). In the processed CSS input, A1 had a mean vascular fraction of 0.304467 and a macro_state_score of 0.326819, whereas B1 had a mean vascular fraction of 0.033746 and a macro_state_score of 0.058504. This result was consistent with the subtype-level CSS interpretation that A1/N3KO-like contains higher vascular abundance than B1/BCAS-like. The vascular subcompartment summary also showed that A1 was composed mainly of endothelial cells, whereas B1 had a relatively higher mural-cell proportion. For the CIS-derived Balanced input, the dominant sender-to-vascular axis in both subtype branches was glia-to-vascular communication (Figure 2B). A1 showed a glia-derived balanced_axis_score of 0.780779, while B1 showed a higher glia-derived balanced_axis_score of 1.009534. B1 also retained a weak immune-to-vascular axis with a balanced_axis_score of 0.082859. In contrast, the immune and neuron axes in A1 were zero-score axes, and the neuron axis in B1 was also zero. Therefore, the top-axis table should be interpreted with this limitation: although the technical output contains top-ranked sender axes, some lower-ranked axes are zero-score placeholders rather than effective relay axes. For the LRS-derived Micro input, pair-level evidence was generated by combining ligand-receptor pairs with CIS-axis information and receptor-level subtype feature support (Figure 2C). The LRS pair score was computed from pair_weight_norm, relay_score, receptor_feature_norm, and receptor_is_deg, with

direction_match and receptor_feature_norm acting as upstream gates. This produced a conservative pre-Micro seed score. Among the pair-level signals, the strongest pair was B1/Cadm1-Cadm1 with an lrs_pair_score of 0.223439. Other visible signals included A1/Ptn-Ncl, B1/Ptprm-Ptprm, and B1/Vegfa-Flt1. A1/Vegfb-Flt1 was retained in the pair-level table but had an lrs_pair_score of zero because it lacked feature support or direction matching. The LRS pair-level table was then aggregated into receptor/protein-level Micro target seeds (Figure 2D). In total, eight single-protein receptor seeds were retained. A1 contained two target seeds, Ncl and Flt1, but the effective pre-Micro signal was concentrated in Ncl, which had a micro_target_score of 0.162535, whereas A1/Flt1 had a score of zero. B1 contained six target seeds: Cadm1, Ptprm, Flt1, Plxna2, Kdr, and Jam2. Their micro_target_scores were 0.223439, 0.048801, 0.046163, 0.001616, 0.000565, and 0.000244, respectively. Thus, the current pre-Micro target layer separated the retained seeds into stronger candidates, including A1/Ncl and B1/Cadm1, Ptprm, and Flt1, and weaker or marginal seeds, including B1/Plxna2, Kdr, Jam2 and A1/Flt1. Together, these results show that the cSVI-DM preprocessing step converted vascular-centered scRNA evidence into CSS-, CIS-, and LRS-derived DM inputs. At this stage, A1 showed a stronger CSS-derived vascular-state input and a concentrated Ncl Micro seed, whereas B1 showed a stronger glia-to-vascular CIS axis and a larger set of LRS-derived target seeds. These inputs provided the immediate basis for the subsequent Micro drug-mapping step.

Micro-v1 pharmacological mapping generated subtype-specific pre-MD drug rankings

Based on the receptor/protein-level Micro seeds generated from the LRS-derived input layer, we next performed Micro-v1 pharmacological mapping to generate subtype-specific pre-MD drug rankings (Figure 3). This step linked the Micro target seed table to pharmacological resources and produced a unified drug-target database, followed by subtype-specific ranking of candidate small molecules. Because structural availability and MD readiness were not incorporated into the ranking score at this stage, the resulting table was interpreted as a pre-MD Micro drug ranking rather than a final MD entry list.

The Micro database mapping build showed that the workflow started from eight DM targets and generated a ChEMBL target mapping for four targets (Figure 3A). ChEMBL retrieval produced 65,824 activity records and 70 mechanism records, which were further processed into 15,704 ChEMBL drug-target records. BindingDB did not contribute retained drug-target records under the current filtering rules, and the unified drug-target database therefore contained 15,704 records. The downstream Micro target-level table contained 18,036 records, the Micro drug-ranking table contained 15,935 records, and the subtype-specific Micro Top20 output contained 40 entries. Overall, five targets had current database coverage, whereas three retained biological seeds had no current drug-database hit.

Target coverage analysis showed that database mappability was uneven across the retained Micro seeds (Figure 3B). In A1, Ncl had three mapped compounds, whereas Flt1 had 2,332 mapped compounds. In B1, Kdr and Flt1 showed extensive database coverage, with 13,341 and 2,332 mapped compounds, respectively. Ptprm showed moderate coverage with 28 mapped compounds. In contrast, Cadm1, Plxna2, and Jam2 had no mapped compounds in the current database layer. Therefore, several LRS-derived biological seeds were retained by the subtype model but were not yet translated into pharmacological candidates in this version.

The relationship between Micro target score and pharmacological mappability further separated biological seed strength from database coverage (Figure 3C). A1/Ncl had a relatively strong Micro target score but limited compound coverage. B1/Cadm1 showed the highest Micro target score among the retained B1 seeds, but lacked current drug-database support. Conversely, Kdr and Flt1 had extensive compound coverage but differed in their Micro target scores. This distinction indicates that Micro-v1 ranking was constrained by two partially independent factors: subtype-derived biological seed strength and available drug-target database coverage.

Subtype-specific Micro drug ranking revealed different candidate structures for A1 and B1 (Figure 3D). In A1, the top-ranked candidates were concentrated on the Ncl signal, with CHEMBL1232461 ranked first and CHEMBL3752910 ranked second. The remaining A1 top candidates were mainly Flt1-associated entries with lower or near-zero Micro drug scores under the current scoring rule. In B1, the top-ranked

candidates formed a more continuous Flt1/Kdr-centered group. CHEMBL535 ranked first, followed by CHEMBL587723, CHEMBL1289926, CHEMBL3939307, CHEMBL273187, CHEMBL272938, CHEMBL374044, CHEMBL1993996, CHEMBL1986979, and CHEMBL1994638. These B1 top candidates were consistently annotated with Flt1 and Kdr as their target hits.

Together, Figure 3 shows that Micro-v1 pharmacological mapping successfully transformed a subset of LRS-derived receptor/protein seeds into subtype-specific pre-MD drug rankings. A1 produced a localized Ncl-centered ranking pattern, whereas B1 produced a broader Flt1/Kdr-centered ranking pattern. At the same time, the unmapped status of Cadm1, Plxna2, and Jam2 indicates that not all biological seeds can be pharmacologically translated using the current database resources. These results defined the candidate space for the subsequent structural validation step.

Structural validation prioritized a B1/FLT1-KDR Micro axis supported by favorable binding-energy estimates and short-timescale MD stability

To further evaluate whether Micro-v1 candidates could be translated into structurally testable drug-target hypotheses, we performed protein-ligand structural validation on selected A1/NCL and B1/FLT1-KDR candidates (Figure 4). Candidate selection was based on the pre-MD Micro ranking and structure-readiness considerations. The structural workflow included DiffDock pose generation, 5 ns GROMACS molecular dynamics simulations, trajectory-level stability assessment, and MMPBSA-derived

MM/GBSA binding-energy estimation. The purpose of this step was not to establish experimental binding affinity, but to prioritize computationally plausible protein-ligand hypotheses for future validation.

Across the simulated complexes, the strongest binding-energy estimates were observed in B1/FLT1-KDR kinase-domain-like fragments (Figure 4A). The most favorable estimate was observed for FLT1 801-1185-CHEMBL587723, with a ΔGbind of -53.5724 ± 2.6089 kcal/mol. KDR 806-1171-CHEMBL587723 showed a similarly strong estimate of -50.5850 ± 3.0168 kcal/mol. FLT1 801-1185-CHEMBL535 also showed a strongly favorable estimate of -44.5570 ± 2.6418 kcal/mol. KDR 806-1171-CHEMBL535 showed a more moderate but still favorable estimate of -31.1939 ± 3.1994 kcal/mol. These results indicate that the B1/FLT1-KDR axis, previously identified by Micro-v1 drug ranking, was also supported at the structural prioritization level. Candidate-level MM/GBSA binding-energy estimates for all structurally validated Micro drug-target pairs are provided in Supplementary Table S1.

The MD stability summary further separated high-energy candidates from dynamically conditional candidates (Figure 4B). The three Tier 1 B1 candidates, FLT1-CHEMBL587723, KDR-CHEMBL587723, and FLT1-CHEMBL535 on kinase-domain-like fragments, showed favorable or very favorable energy estimates together with stable RMSD and stable Rg/SASA behavior. Their hydrogen-bond support was variable, ranging from weak to intermittent or moderate, indicating that

the favorable binding-energy profiles were not primarily driven by persistent hydrogen-bond networks. Instead, the results are more consistent with van der Waals, hydrophobic, and nonpolar burial-supported binding modes. Among the supportive candidates, KDR 806-1171-CHEMBL535 showed a moderate binding-energy estimate but clean short-timescale dynamics, supporting its placement as a secondary B1/KDR structural hypothesis.

Fragment-level comparisons showed that kinase-domain-like FLT1 and KDR structures were more suitable for short-timescale MD-based prioritization than extracellular-like fragments in this exploratory workflow. FLT1 1-687-CHEMBL587723 and FLT1 1-687-CHEMBL535 produced favorable or moderate binding-energy estimates, but both were classified as dynamically flexible or rearranged. Similarly, KDR 1-678-CHEMBL587723 was energetically favorable compared with KDR 1-678-CHEMBL535, but remained dynamically conditional, whereas KDR 1-678-CHEMBL535 showed pocket drift and was retained only as an exploratory candidate. These results suggest that the extracellular-like fragments provided supportive or exploratory evidence, but the main B1 structural signal was concentrated in kinase-domain-like fragments.

The A1/NCL candidates provided cross-branch proof-of-concept support but were not prioritized as main structural hits. NCL-CHEMBL3752910 showed a favorable Δ Gbind estimate of -30.5582 $\pm$ 3.3466 kcal/mol, but the trajectory summary indicated conformational rearrangement, compressed Rg/SASA behavior, weak hydrogen-bond

support, and ligand-pocket distance drift. NCL-CHEMBL1232461 showed a more moderate favorable estimate of -16.3931 ± 2.2994 kcal/mol and was also associated with dynamic rearrangement. These results support the existence of structurally plausible A1/NCL candidates, but indicate that their dynamic behavior requires caution and further validation.

Integrated prioritization placed FLT1 801-1185-CHEMBL587723, KDR 806-1171-CHEMBL587723, and FLT1 801-1185-CHEMBL535 into Tier 1, with KDR 806-1171-CHEMBL535 and flexible extracellular-like B1 complexes placed into supportive tiers (Figure 4C). A1/NCL complexes and KDR 1-678-CHEMBL535 were retained as cautionary, pilot, or exploratory candidates. Together, these structural results support the B1/FLT1-KDR axis as the main cSVI-DM v1.0 Micro structural-validation branch, while positioning A1/NCL as a secondary proof-of-concept branch rather than the primary structural claim.

Discussion

Computational drug screening workflows differ in the type of disease evidence they accept as input. Perturbational transcriptomic methods, including CMap- and LINCS-like resources, are effective when disease biology can be represented as an expression signature. Drug-target knowledgebases, such as ChEMBL(Zdrazil, et al., 2024), BindingDB(Liu, et al., 2025), DrugBank(Knox, et al., 2024), and DGIdb(Cotto, et al., 2018), are most useful when a nominated gene or protein can be linked to

known compounds or binding activities. Single-cell drug-response tools extend prediction toward cell-level contexts, whereas docking, molecular dynamics and MM/GBSA workflows provide downstream structural evaluation for defined protein-ligand pairs. These approaches are complementary, but each is typically organized around a specific entry point: a signature, a target, a predicted response or a docked complex.

The cSVI-DM distinction changes the role of existing resources. In a conventional signature-reversal workflow, vascular state evidence might be reduced to a single up/down gene signature. In cSVI-DM, CSS-derived evidence is assigned to Macro mapping, where the relevant question is state remodeling. Similarly, CIS-derived communication evidence is not forced into a single target list, but is assigned to Balanced mapping, where the relevant question is whether abnormal source-to-vascular relay can be modulated. LRS-derived evidence, by contrast, is naturally closer to receptor/protein-level intervention and is therefore assigned to Micro mapping, where drug-target databases and structural validation are most appropriate.

The current v1.0 implementation most fully realizes the Micro channel. ChEMBL and BindingDB were used as pharmacological translators of LRS-derived target seeds, and DiffDock, GROMACS and MM/GBSA analysis were used as downstream structural prioritization tools. This ordering is important: structural validation was not

used as a blind screening engine, but as a final evaluation step for subtype-resolved Micro candidates derived from vascular-centered single-cell evidence.

The separation between biological seed generation and pharmacological mappability is also important. Some LRS-derived seeds may be biologically supported but poorly represented in current drug databases, whereas database-rich targets may dominate candidate output once they pass subtype-level filters. This was observed in the current implementation, where several retained LRS seeds lacked drug-database translation, while the B1/FLT1-KDR axis produced a more pharmacologically mappable Micro cluster. This illustrates a general advantage of the cSVI-DM design: lack of database coverage can be distinguished from lack of biological support, and target mappability can be interpreted within a subtype-aware evidence framework.

Compared with single-cell drug-response prediction tools, cSVI-DM also has a different objective. Tools such as scDrug or scDEAL aim to estimate drug response or sensitivity at single-cell resolution, often using pharmacological response references. cSVI-DM does not directly predict drug sensitivity. Instead, it uses subtype-resolved single-cell disease evidence to define which therapeutic layer a candidate belongs to. In future versions, drug-response prediction could be incorporated as additional downstream evidence, but it would not replace the CSS/CIS/LRS-based organization of disease-side signals.

Several limitations define the current v1.0 scope. First, Macro and Balanced are formalized as layer-specific mapping frameworks and supported by CSS/CIS-derived input construction, but they are not yet implemented with the same pharmacological depth as Micro. Future versions should connect Macro to perturbational signature-reversal resources and Balanced to pathway-, cytokine-, hormone- or network-modulating drug resources. Second, Micro output remains dependent on database coverage, which may favor well-annotated targets. Third, the structural validation used short-timescale MD simulations and MM/GBSA estimates, which are appropriate for computational prioritization but not for claiming experimental binding affinity. Fourth, protein-fragment selection can influence MD stability and should be further refined using longer simulations, replicate trajectories and experimental validation.

Overall, cSVI-DM should be viewed as a layer-aware organizing framework for drug mapping rather than a standalone replacement for existing drug-screening tools. Its main contribution is to preserve the structure of disease-side evidence and assign different evidence layers to different therapeutic interpretation channels.

Conclusion

In conclusion, cSVI-DM establishes a layer-aware route from vascular-centered subtype evidence to subtype-aware therapeutic hypothesis generation, with Micro-v1 providing the first fully implemented pharmacological and structural validation branch.

Acknowledgements

None.

Author contributions

Y.L. conceived and designed the study, developed the methodology, performed the analyses, generated the figures, and wrote the manuscript. **Y.L.** approved the final version of the manuscript.

Funding

None

Data Availability

Processed data and analysis outputs generated in this study are available in the accompanying GitHub repository and archived Zenodo release. These include cSVI-DM CSS/CIS/LRS input tables, Micro target-seed tables, ChEMBL/BindingDB-derived drug-target mapping outputs, subtype-specific Micro drug rankings, target-coverage summaries, figure source data, MM/GBSA result summaries, and supplementary tables. Structural-validation files will include selected protein/ligand inputs, docking-ready candidate information, DiffDock pose outputs where appropriate, GROMACS configuration files, trajectory-derived summary metrics, and processed MD/MMGBSA analysis outputs. Large binary MD trajectory files will be deposited in Zenodo when file-size limits permit; otherwise, processed trajectory summaries and final analysis outputs will be provided, with raw trajectories

available from the author upon reasonable request. Raw public datasets and database records should be accessed from their original sources, including ChEMBL, BindingDB, AlphaFoldDB and PDB. The reported pharmacological mapping results correspond to the processed ChEMBL/BindingDB outputs used for this manuscript snapshot. Live database queries may return slightly different record counts as database content is updated. The upstream cSVI-subtype framework is archived at Zenodo under DOI: 10.5281/zenodo.19198768, and at Research square under DOI: 10.21203/rs.3.rs-9937210/v1.

The csviDM workflow repository, including example files and reproducibility scripts, is publicly available at https://github.com/YuqianLii/csviDM. Additional processed intermediate outputs used in the current manuscript are available from the corresponding author upon reasonable request.

Declarations

Consent to participate

Not applicable.

Consent for publication

Not applicable.

Conflict of Interest

The authors declare that they have no competing interests.

References

Berendsen, H.J.C., Grigera, J.R. and Straatsma, T.P. The missing term in effective pair potentials. *The Journal of Physical Chemistry* 2002;91(24):6269-6271.

Chen, J.*, et al.* Deep transfer learning of cancer drug responses by integrating bulk and single-cell RNA-seq data. *Nat Commun* 2022;13(1):6494.

Chen, J.*, et al.* Dysregulation of Principal Circulating miRNAs in Non-human Primates Following Ischemic Stroke. *Front Neurosci* 2021;15:738576.

Corso, G.*, et al.* DiffDock: Diffusion Steps, Twists, and Turns for Molecular Docking. 2022.

Cotto, K.C.*, et al.* DGIdb 3.0: a redesign and expansion of the drug-gene interaction database. *Nucleic Acids Res* 2018;46(D1):D1068-D1073.

Duan, Q.*, et al.* L1000CDS(2): LINCS L1000 characteristic direction signatures search engine. *NPJ Syst Biol Appl* 2016;2.

Duering, M.*, et al.* Neuroimaging standards for research into small vessel disease-advances since 2013. *Lancet Neurol* 2023;22(7):602-618.

Hilkens, N.A.*, et al.* Stroke. *Lancet* 2024;403(10446):2820-2836.

Kagami, L.*, et al.* The ACPYPE web server for small-molecule MD topology generation. *Bioinformatics* 2023;39(6).

Knox, C.*, et al.* DrugBank 6.0: the DrugBank Knowledgebase for 2024. *Nucleic Acids Res* 2024;52(D1):D1265-D1275.

Li, X.*, et al.* Molecular Mediators of Neutrophil Primary Granule Release Following Acute Ischemic Stroke and their Associated Epigenetic Modulation by HDAC2. *Mol Neurobiol* 2025;62(5):6544-6561.

Li, Y. cSVI-subtype: a GWAS-anchored, vascular-centered framework for molecular subtype barcode discovery in cerebral small vessel disease. *Research Square* 2026.

Li, Y.*, et al.* The High-Affinity IL-2 Receptor Affects White Matter Damage after Cerebral Ischemia by Regulating CD8 + T Lymphocyte Differentiation. *J Neuroimmune Pharmacol* 2025;20(1):8.

Li, Y.*, et al.* The IL-2A receptor pathway and its role in lymphocyte differentiation and function. *Cytokine Growth Factor Rev* 2022;67:66-79.

Lindorff-Larsen, K.*, et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* 2010;78(8):1950-1958.

Liu, T.*, et al.* BindingDB in 2024: a FAIR knowledgebase of protein-small molecule binding data. *Nucleic Acids Res* 2025;53(D1):D1633-D1644.

Markus, H.S. and Joutel, A. The pathogenesis of cerebral small vessel disease and vascular cognitive impairment. *Physiol Rev* 2025;105(3):1075-1171.

Michnick, S.W. The connectivity map. *Nat Chem Biol* 2006;2(12):663-664.

Pronk, S.*, et al.* GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 2013;29(7):845-854.

Subramanian, A.*, et al.* A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* 2017;171(6):1437-1452 e1417.

Sun, Y.Y., *et al.* scDrug+: predicting drug-responses using single-cell transcriptomics and molecular structure. *Biomed Pharmacother* 2024;177:117070.
Zdrazil, B., *et al.* The ChEMBL Database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. *Nucleic Acids Res* 2024;52(D1):D1180-D1192.

Figure Legends

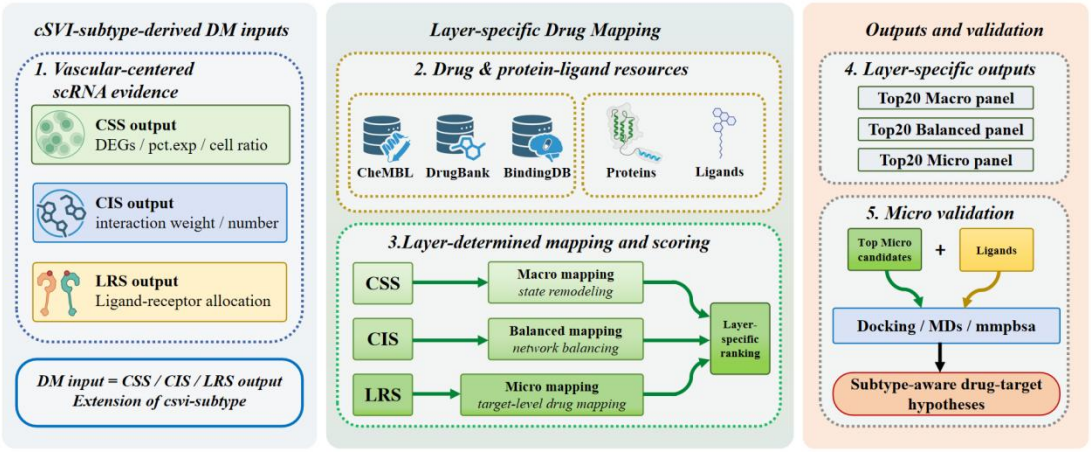

Fig. 1. Overview of the cSVI-DM framework.

Schematic overview of cSVI-DM as a downstream drug-mapping module of cSVI-subtype. The framework uses CSS, CIS, and LRS outputs as layer-specific intervention inputs rather than using the final subtype barcode as the direct drug-screening input. CSS is mapped to Macro state-remodeling strategies, CIS to Balanced network-regulation strategies, and LRS to Micro target-level small-molecule strategies. Micro candidates are further advanced to docking, molecular-dynamics simulation, and MM/GBSA-based structural prioritization.

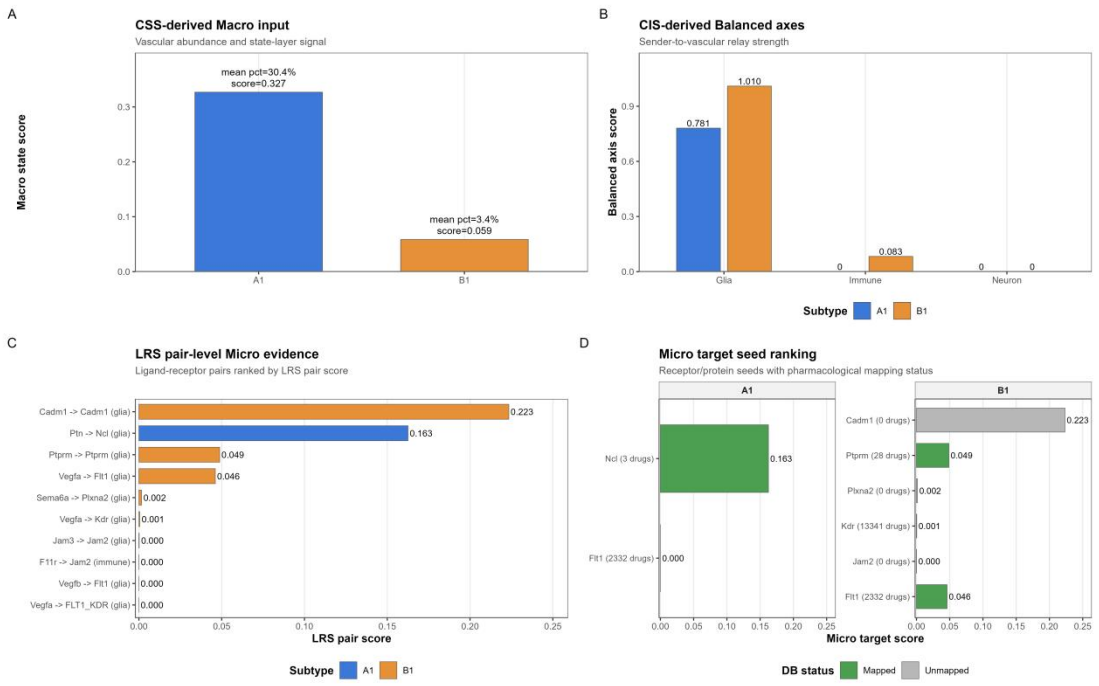

Fig. 2. Construction of CSS-, CIS-, and LRS-derived cSVI-DM inputs.

Layer-specific input construction from vascular-centered scRNA evidence. (A) CSS-derived Macro input summarizing vascular abundance and state-layer strength in A1/N3KO-like and B1/BCAS-like branches. (B) CIS-derived Balanced input summarizing dominant sender-to-vascular relay axes. (C) LRS pair-level Micro input integrating ligand-receptor evidence, relay support, receptor feature support, and direction matching. (D) Aggregated receptor/protein-level Micro target seeds used for downstream pharmacological mapping.

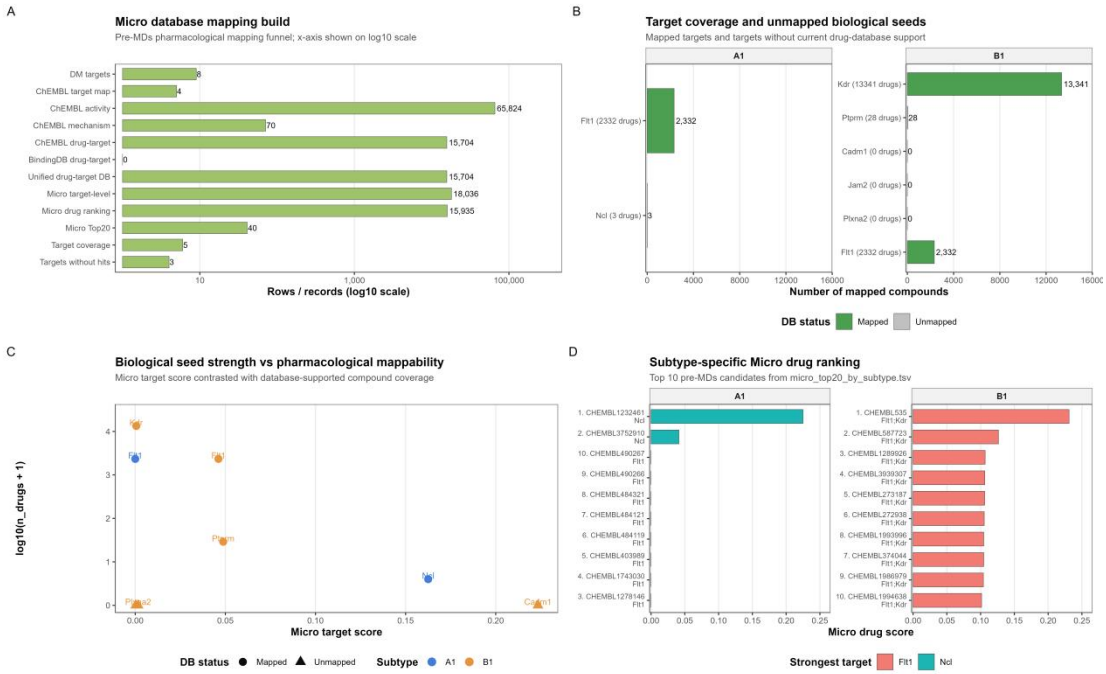

Fig. 3. Micro-v1 pharmacological mapping and subtype-specific pre-MD rankings.

Micro-layer drug mapping from LRS-derived target seeds to pharmacological candidate rankings. (A) Summary of ChEMBL/BindingDB-based drug-target database construction and downstream Micro ranking outputs. (B) Target-level database coverage across retained Micro seeds. (C) Relationship between subtype-derived Micro target support and pharmacological mappability. (D) Subtype-specific Micro Top20 rankings, showing an A1/NCL-centered pattern and a B1/FLT1-KDR-centered candidate cluster.

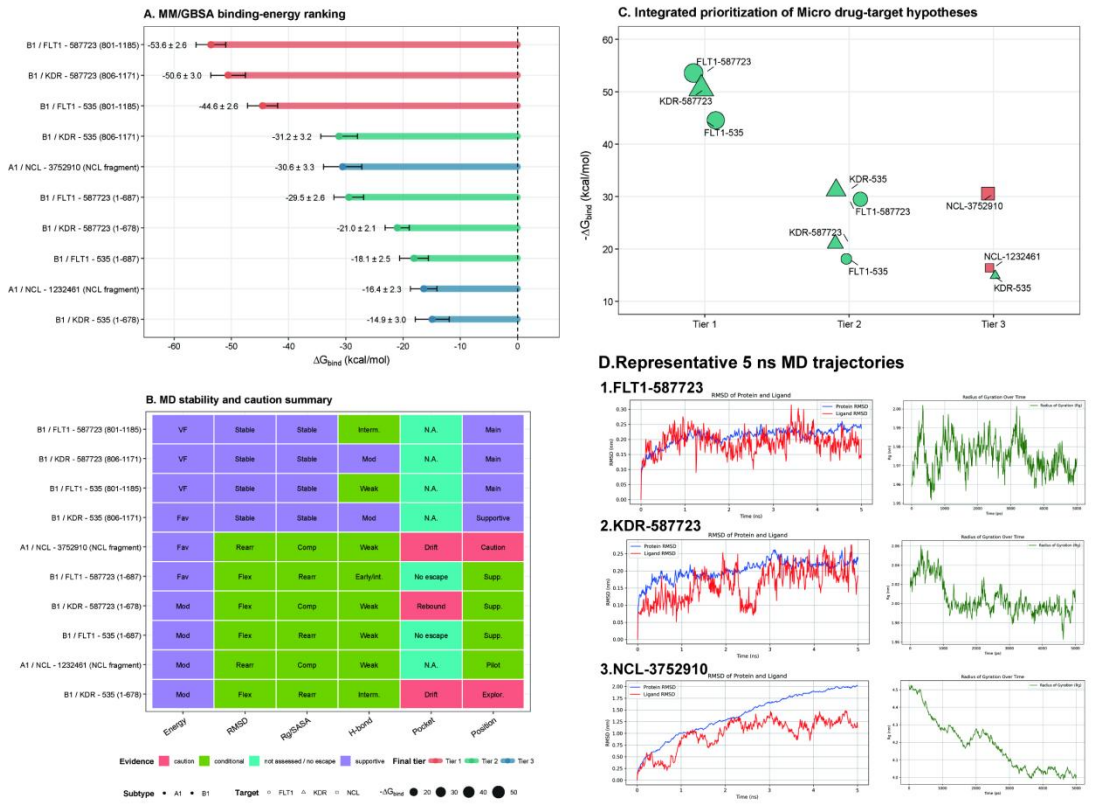

Fig. 4. Structural validation of prioritized Micro drug-target candidates.

Protein-ligand structural validation of selected A1/NCL and B1/FLT1-KDR Micro candidates. (A) MM/GBSA binding-energy estimates across simulated complexes. (B) MD stability and caution summary integrating RMSD, Rg, SASA, hydrogen-bond behavior, PCA trajectory patterns, and pocket-distance information when available. (C) Integrated prioritization of structural candidates, highlighting the B1/FLT1-KDR kinase-domain-like branch as the main cSVI-DM v1.0 structural-validation axis and A1/NCL as a secondary proof-of-concept branch.

Supplementary Table S1. MM/GBSA binding-energy estimates for structurally validated Micro drug-target candidates

Candidate	Subtype/target	Fragment	ΔG_bind / kcal·mol⁻¹
B1_FLT1_CHEMBL587723	B1 / FLT1	801-1185	-53.5724 ± 2.6089
B1_KDR_CHEMBL587723	B1 / KDR	806-1171	-50.5850 ± 3.0168
B1_FLT1_CHEMBL535	B1 / FLT1	801-1185	-44.5570 ± 2.6418
B1_KDR_CHEMBL535	B1 / KDR	806-1171	-31.1939 ± 3.1994
A1_NCL_CHEMBL3752910	A1 / NCL	NCL fragment	-30.5582 ± 3.3466
B1_FLT1_CHEMBL587723	B1 / FLT1	1-687	-29.4727 ± 2.5968
B1_KDR_CHEMBL587723	B1 / KDR	1-678	-21.0210 ± 2.1064
B1_FLT1_CHEMBL535	B1 / FLT1	1-687	-18.0990 ± 2.5249
A1_NCL_CHEMBL1232461	A1 / NCL	NCL fragment	-16.3931 ± 2.2994
B1_KDR_CHEMBL535	B1 / KDR	1-678	-14.9109 ± 2.9557