ChromMine™
Intuitive datamining for complex data.

ChromMine™ 2026R2 Specifications

Highlights

- Intuitive, fast interface – all processing is web-based, no software installation required.
- Unique data import system (completely software agnostic) with automatic data cleansing:
    - Duplicate compound/sample renaming
    - Consistency checking
    - Metadata validation
- Simple dataset organisation and filtering (see Data Cleaning below).
- Remove unwanted samples, compounds, or peaks.
- Focus analysis on selected subsets.
- Multiple built-in analysis methods – from single-compound tracking to full dataset comparisons:
    - Heatmaps, bubble plots, reconstituted chromatograms
    - Sample correlations and profile comparisons
    - Fully interpretable cluster plots with automatic calculations
    - Sample Similarity (compound-based) and/or Profile Similarity (RT-based)
    - Volatile Concentration Similarity
    - Time-series sample plotting
- Comprehensive export options – all data, plots, and a fully detailed automated report are downloadable.
- Can be used with any type of chromatographic data
    - Compound identified, e.g., GC-MS or LC-MS
    - Compound agnostic, e.g. GC-FID, LC-UV, LC-RI, GPC-UV etc.

ChromMine™
Intuitive datamining for complex data.

Full Specifications

Data Security

- Secure website with encrypted (hashed) passwords
- Imported data remains unchanged during use
- All data deleted upon logout

Usage

- Online web app – no installation required
- Fully dynamic, interactive interface

Data Import

- Accepts software-agnostic CSVs from any GC-MS package
- Supports both:
  - Flat data (one row per peak per sample)
  - Tabulated data (samples in columns, compounds in rows)
- Import multiple files and choose/change which ones to analyse at any time
- Import metadata for Samples, Compounds, and Peaks
- Automatic checks and alignment for:
  - Unique sample/file name pairs
  - Consistent metadata
  - Missing data
  - Compounds with the same formula at similar retention times
  - Compounds detected at significantly different retention times
- Renaming of duplicate compounds within or across samples, using unique retention times or grouping by median retention time
  - Automatic
  - Manual – choose from suggested list or assign a custom compound name, formula and CAS number
- Prevents data loss or averaging – a unique ChromMine feature

Data Cleaning

- Peak abundance scaling:
  - Raw
  - log10
  - normalised (by sample, by compound, or to internal standard)
- Global filters:
  - Minimum cutoff by area/abundance
  - Minimum cutoff by match factor
  - Retention time inclusion/exclusion windows
- Sample filters:
  - Include/exclude specific samples
  - Include/exclude by metadata
- Compound filters:
  - Include/exclude by saturation
  - Include/exclude by element(s) (Br, Cl, D, F, I, N, O, P, S, Si)
  - Include by element, and exclude by element from subset (e.g., include all N-compounds, but exclude all that also contain O)
  - Include/exclude specific compounds
  - Include/exclude by categorical metadata

    - Apply cutoffs to numerical metadata
- Peak filters:
    - Include/exclude by categorical metadata
    - Apply cutoffs to numerical metadata

Sample Comparison

- Identify which samples contain a given compound(s)
- Compare compound profiles of up to 10 samples:
    - Direct comparisons
    - Relative to a reference sample
- Calculate correlations:
    - Choose correlation basis (Sample Similarity (compound-based) or Profile Similarity (RT-based))
    - Single sample vs all others
    - Full dataset correlation
    - Sort samples in alphabetical order, by sample metadata or correlation similarity
    - Annotate/group by sample metadata (including multiple levels)
- Dataset visualisations: bubble plots, heatmaps
- Reconstitute GC-MS chromatograms in 2D or 3D as Gaussian peaks or as stick plots
- Compare sample repeats
    - check which compounds are in all samples and their variance
    - check the overall similarity of the repeat samples
- Compare compound concentrations over time
    - normalise samples with respect to the first time-point, as % of maximum and more
    - view as bubble plot, multiple plots or trajectory clusters (see which compounds increase/decrease together in subsequent samples)

Clustering

- 2D and 3D clustering with interpretable plots
- Standard plotting variables:
    - Sample Similarity (compound-based),
    - Profile Similarity (RT-based), and
    - Volatile Concentration Similarity.
- Use Sample categorical or numeric metadata as clustering variables (plotting axes)
- Proprietary automated clustering algorithm (ChromMine Clustering Algorithm™ – CMCA™)
- Cluster using the CMCA™ and highlight samples using their metadata
- Cluster using Sample metadata – choose which metadata variables you want to use and automatically generate all possible combinations
- Automatic calculation of:
    - Distance of sample from cluster centroid
    - Cluster centroid with downloadable profile
- Direct comparison of samples' or clusters' profiles
- Automated clustering reporting – understand why a cluster of samples are different from another
    - show the top compounds that most distinguish one cluster from the others
    - summary of the differences between clusters, in plain language and in numbers

Reporting

- Export all raw/filtered data as CSV
- Export all plot data as CSV (with full settings summary)
- Export plots as PNG or interactive HTML
- Generate a self-contained, portable dynamic HTML report including:
    - Imported file details
    - Compound renaming log
    - Filter/selection log
    - All plots and data links
    - Lists of renamed and excluded compounds

Multi-language

- English (UK)
- English (US)
- Italian
- French
- German
- Spanish
- Portuguese
- Japanese
- Korean
- Chinese (Simplified)
- Arabic