evotec

'RESEARCH NEVER STOPS'

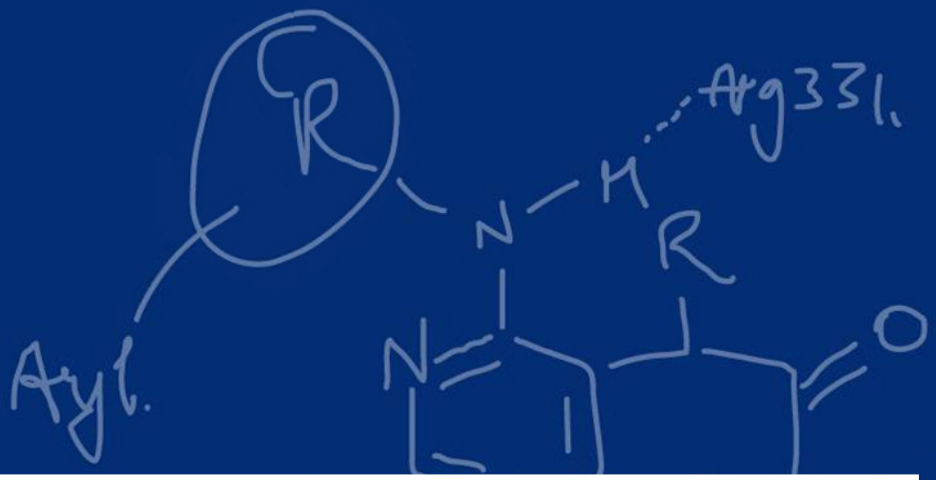
Structural Data Analysis workflow in the context of industrial uHTS

Creating integrated drug discovery innovation alliances

KNIME

3rd KNIME Users Meeting and Workshop

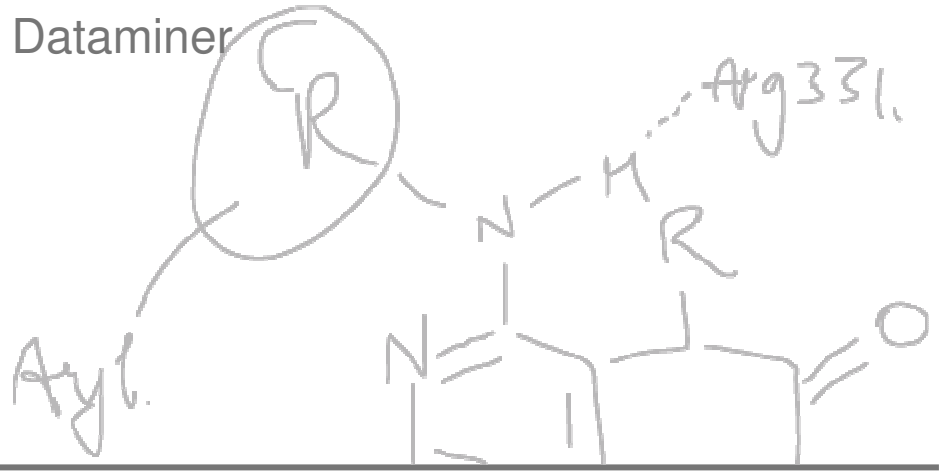
February 22-27, 2010, Zurich, Switzerland



Evotec AG, Knime, Evotec spectral clustering, Tripos HTS Benchware explorer, 22th -27th February 2010

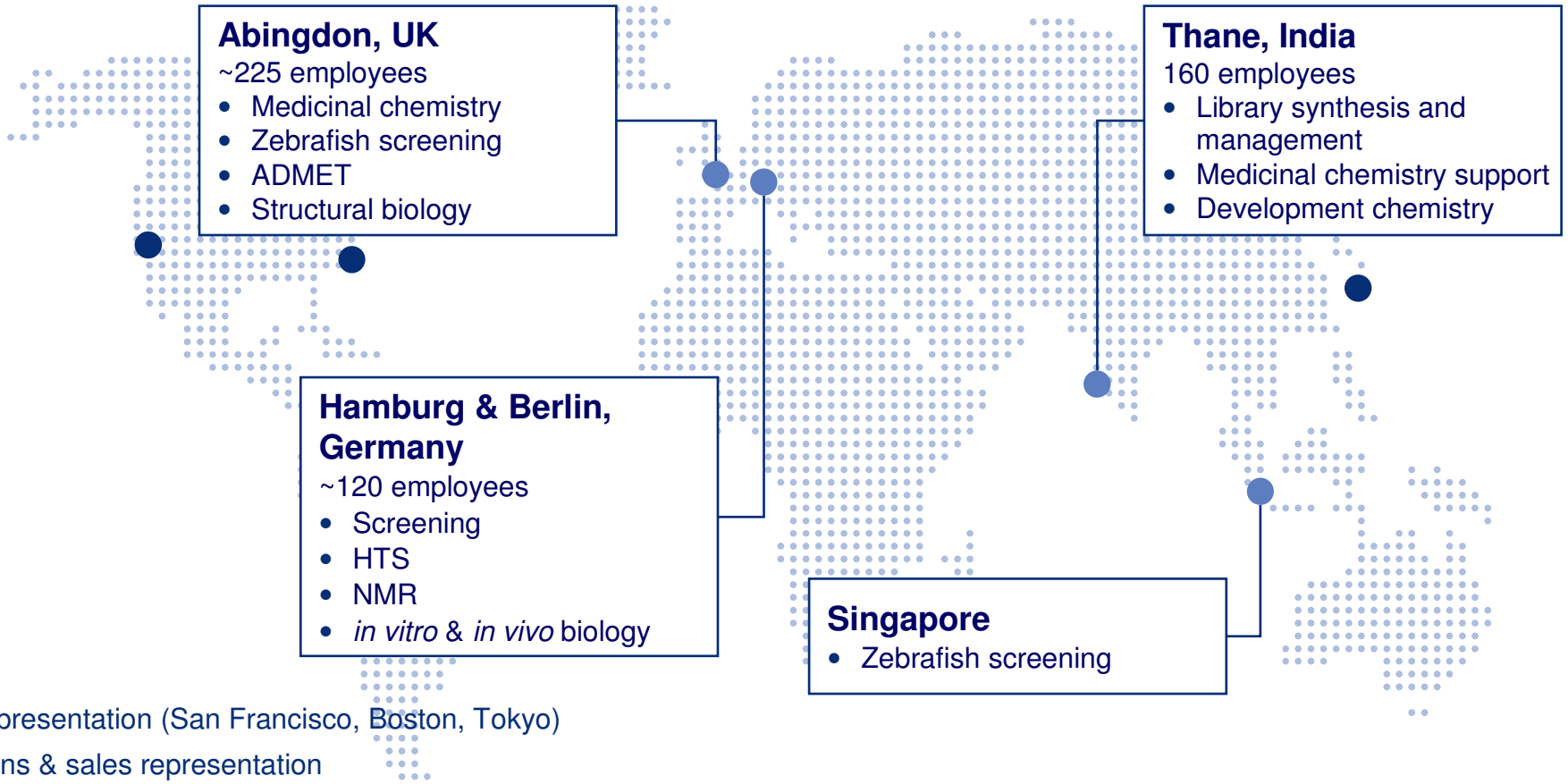
Agenda

- Introduction
- Data analysis in the context of uHTS
  - Knime workflow
  - Spectral Clustering
  - HTS Benchware Dataminer
- Examples
- Summary



Global reach for global projects

Ca. 500 Evotec employees worldwide



We create innovation alliances to discover drugs

Evotec business model

Discovery & technology platform

- Strong track record after >15 years of drug discovery history
  - >100 lead series
  - 20 preclinical candidates
  - 16 clinical compounds
- Leading technology platforms
  1. Assay development & screening
  2. Fragment-based drug discovery
  3. Comp. chem./structural based drug discovery
  4. Medicinal chemistry
  5. Early ADMET& zebrafish

Discovery alliances with the top pharma & biotech companies

Boehringer Ingelheim
NOVARTIS
Roche
ONO
CHDI FOUNDATION, Inc.
biogen idec
Pfizer

Product franchises

- Core expertise with clinical candidates in
  - Treatment-Resistant Depression
  - Others
- Wealth of preclinical assets for partnering or own development

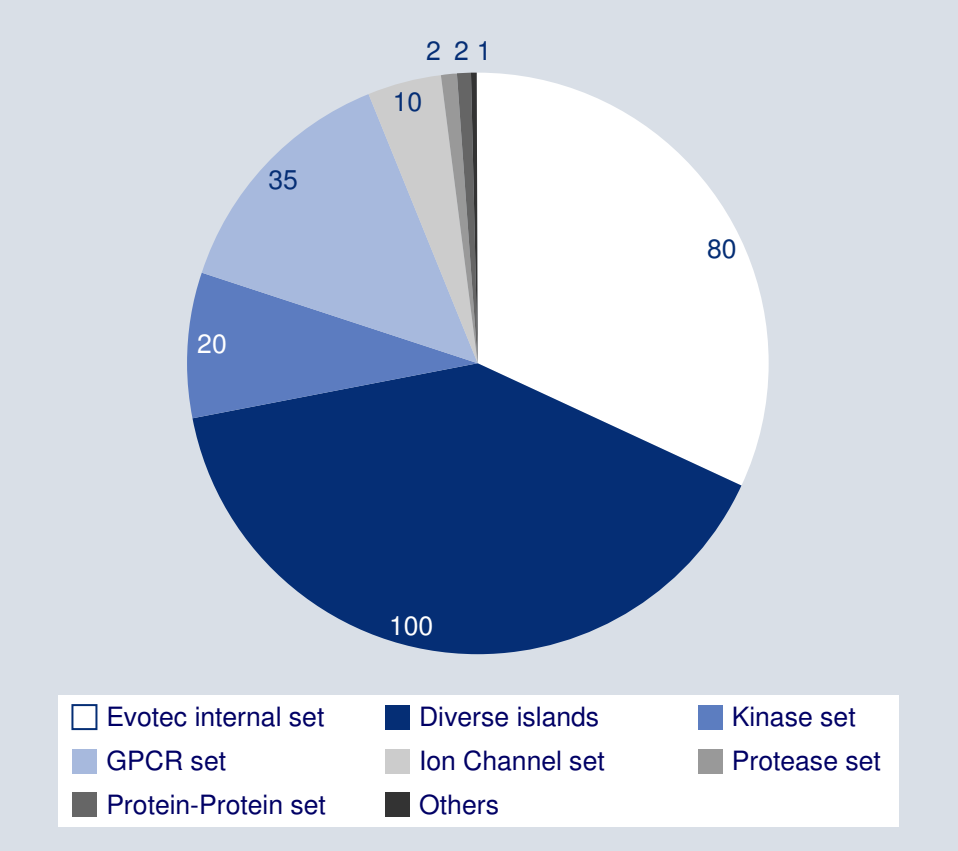
Our value proposition

- Expand highly profitable and cash generating discovery alliance business into globally leading Contract Research Organization
- Finance innovative R&D through
  - DAB profits
  - Development partnerships

Evotec‘s Screening Library – 250K HTS screening collection

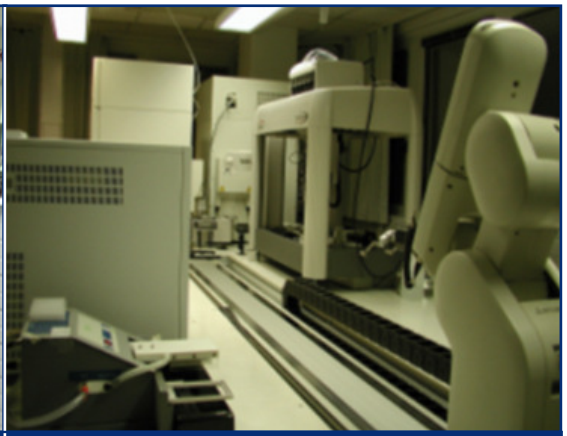
HTS collections

- 250,000 diverse lead-like compound for uHTS
- Optimised compound storage for long term stability
- LC-MS based continuous quality management
- Proven enhanced hit-rate from focused sets
- Diverse ‘islands’ of similar compounds facilitate hit series selection and generate SAR from screening data
- 30,000 fragments for fragment screening



A powerful and multipurpose technology platform

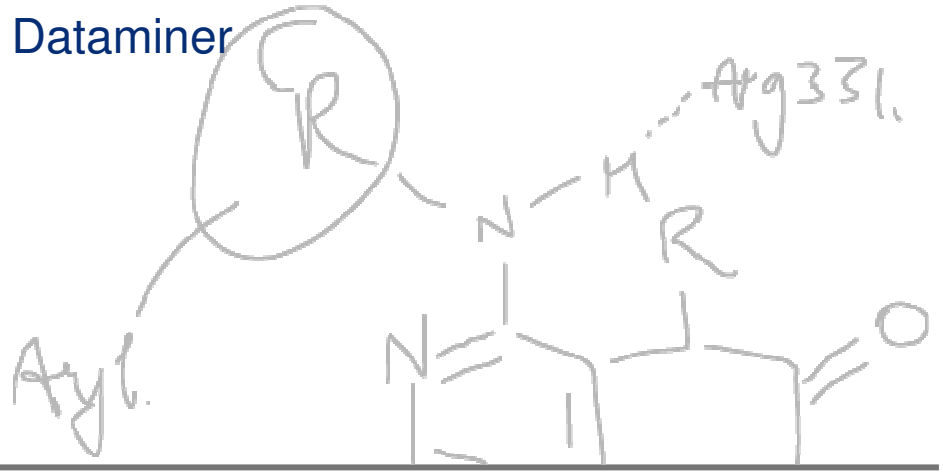
Screening hardware - proprietary and commercial



	EVOscreen Mark II	EVOscreen Mark III	384-HTS System
No. of Systems	2	1	1
Volume	1µl/well	1-50µl/well	10-50µl/well
Plates	2080-well	384, 1536, 2080-well	384-well
Formats	Biochemical	Cellular + Biochemical	Cellular + ELISA
Readers	FCS	EnVision, Opera, FCS etc	Acquest, FLIPR3, Safire
Detection	FCS+*plus*	FCS, fluo, lumi, abs	fluo, lumi, abs
Throughput	100,000 per day	100,000 per day	30,000 per day

Agenda

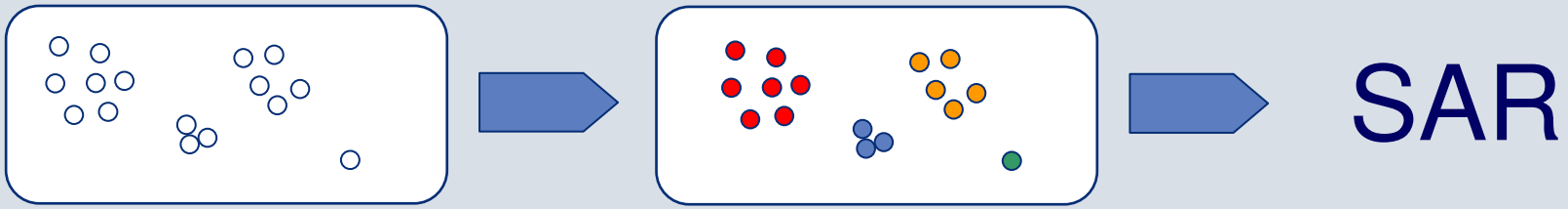
- Introduction
- **Data analysis in the context of uHTS**
  - Knime workflow
  - Spectral Clustering
  - HTS Benchware Dataminer
- Examples
- Summary



Analyze large hit populations and understand SAR

Industry standard - clustering

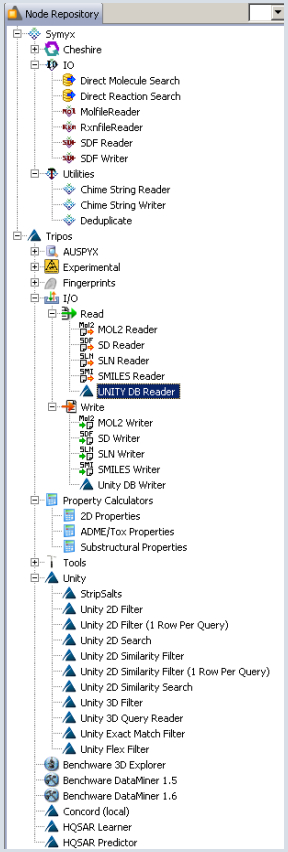
- Our interest: develop an automated procedure to classify 2D molecular structures
  - assess the number and representation of different scaffolds



- Assignment of descriptors
  - assess the number and representation of different scaffolds
  - one dimensional properties (e.g. molecular weight)
  - more complex properties (e.g. molecular electrostatic potential)
- Quantify molecular similarity
  - highly dependent on the molecular descriptors
- Clustering algorithm
  - compute intermolecular similarities and arrange the molecules into clusters

Excellent case study for Knime workflow

All required nodes implemented/available



- Obtain activity data from Evotec analysis platform “A+”
  - tabulated text file
- Split IDs and retrieve structures from Symyx ISIS host database via ISIS direct
  - alternative: read SDfile
- Convert structures to Sybyl line notation
- Calculate Tripos UNITY fingerprints
- Feed fingerprints to cluster algorithm (Evotec spectral clustering)
- Feed merged dataset to HTS Benchware explorer 1.6

- Manual analysis and review of data

- Feed final product to Knime reporting functionality
- Write pdf / Word document

Knime Workflow

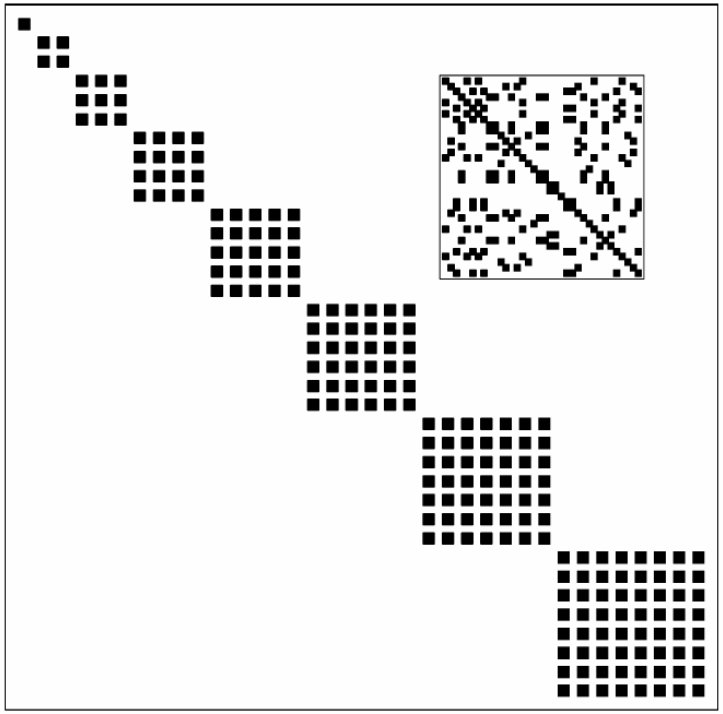
advantages / caveats

- Automated process, easy modification of workflow
- Full documentation via graphical frontend
- Integration of all Evotec available SW vendors (Tripos, Symyx, R-Project, Moe, …)
- Automated reporting available
- No server deployment yet
- Licensing of target applications might prevent company wide role out
- Evotec final platform decision (Linux, Windows) pending

Spectral clustering*

Eigenvalue and eigenvector decomposition of similarity matrix

- *N* molecules produce a similarity matrix *S* (*N x N*)
  - ideally: *k* distinct clusters
  - molecules with similarity $S_{xy} = 1$ belong to one cluster
  - not similar $S_{xy} = 0$ belong to different clusters

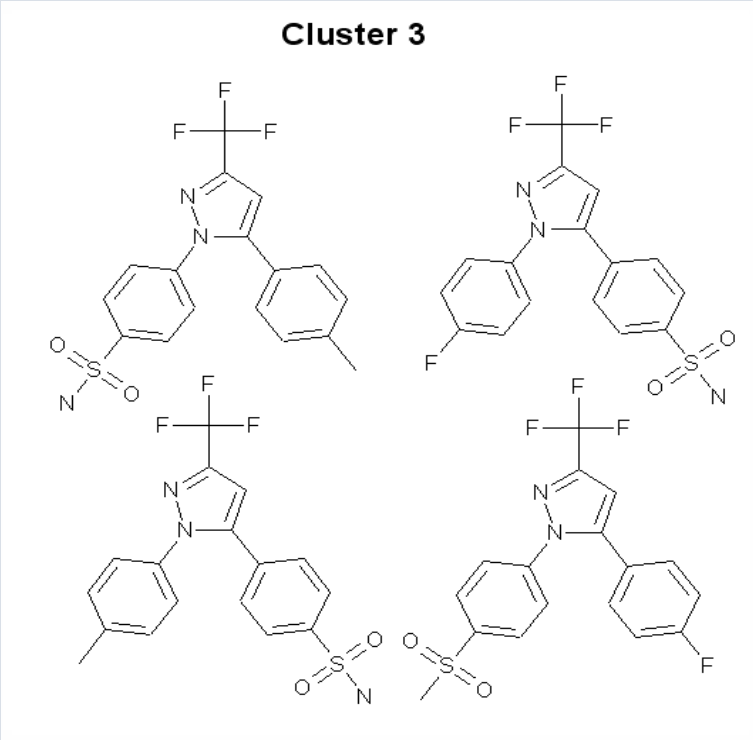


* Brewer, ML J Chem Inf Model. 2007 Sep-Oct;47(5):1727-33

Spectral clustering*

Eigenvalue and eigenvector decomposition of similarity matrix

- Reorder similarity matrix in block diagonal form
  - by definition *S* will have *k* non zero eigenvalues; values are equal to the cluster populations p_i
  - each nonzero eigenvalue has an associated eigenvector that has p_i nonzero elements and the indices of these elements correspond to the molecules belonging to that cluster
- Conclusion: with the eigenvectors at hand it will be possible to reorder the matrix to a block form and reveal the distinct clusters
  - validated on a published test dataset (COX-2)
  - method widely used in image analysis

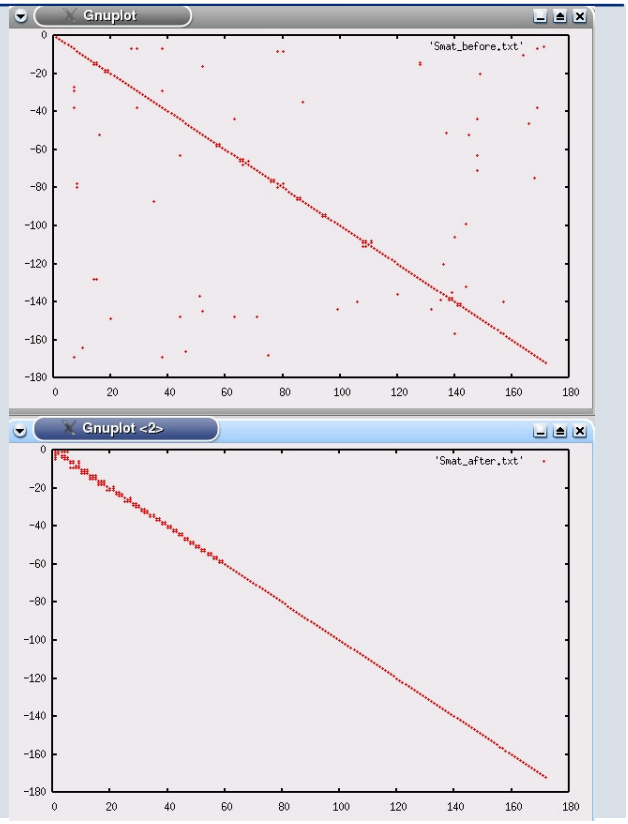


* Brewer, ML J Chem Inf Model. 2007 Sep-Oct;47(5):1727-33

Workflow integration and extensive testing

Pre-calculate clusters at different similarity thresholds

- Automated procedure (initially perl script now Knime workflow)
  - retrieve structures from database (by list of compound IDs)
  - convert to UNITY database and calculate 2D fingerprints
  - calculate similarity matrix
  - reorder matrix and identify clusters (spectral clustering)
  - merge to compound IDs, potency data and compound properties (Lipinski and ADMET)
- Manual review of cluster
  - ranking of clusters by relevance / chemical interest
- Extensive testing on real datasets



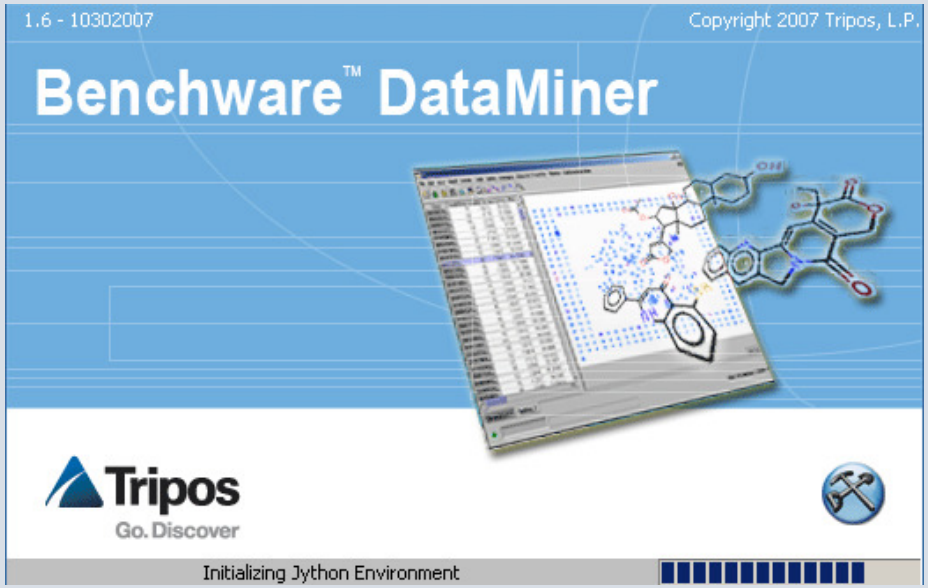
Spectral clustering

advantages / caveats

- Fast and reliable calculation
- Free to integrate into Evotec software tools
- Fully automated cluster assignment
- Applicable to a variety of similarity matrices (different fingerprints)
- No visualization of chemistry space
- No substructures integration
- Cluster ranking independent of target potency information

Tripos HTS Benchware Dataminer

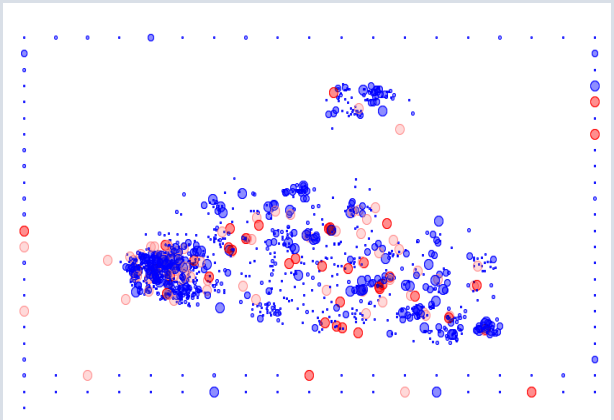
- Workflow integration
- Examples



HTS Benchware Dataminer

Practical Workflow

- Compile dataset (prepare SDFile via Knime Workflow)
  - target potency data (log IC50 values / % inhibition)
  - selectivity information
  - readout artefact behaviour
  - compound purity information
  - predicted Lipinski or ADMET properties

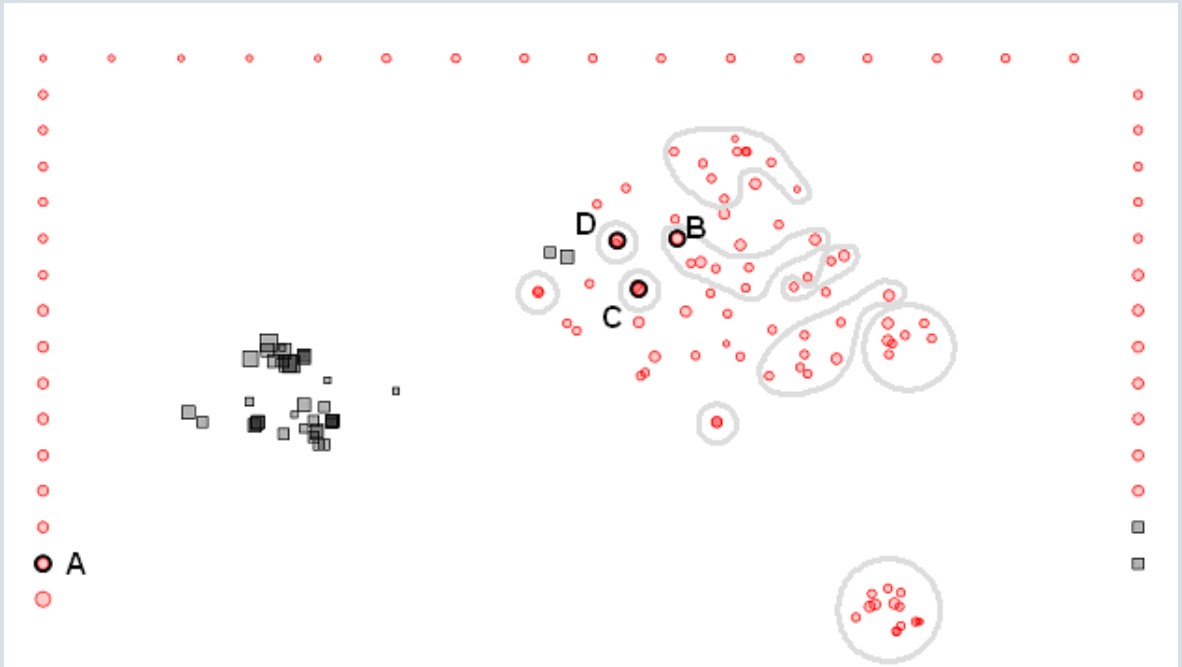


- 2D projection of chemistry space (PCA & NLM)
- Manually identify and rank clusters
- Calculate mcs and add substructure search results to cluster
- Characterize identified clusters

Identification of novel Kv1.3 blockers using a fluorescent cell-based ion channel assay

Comparison to literature known actives

- HTS hits (red circles)
- Literature known actives (black squares)
- Size = potency
- A-D confirmed by patch clamp method
- Singletons at the border of projection



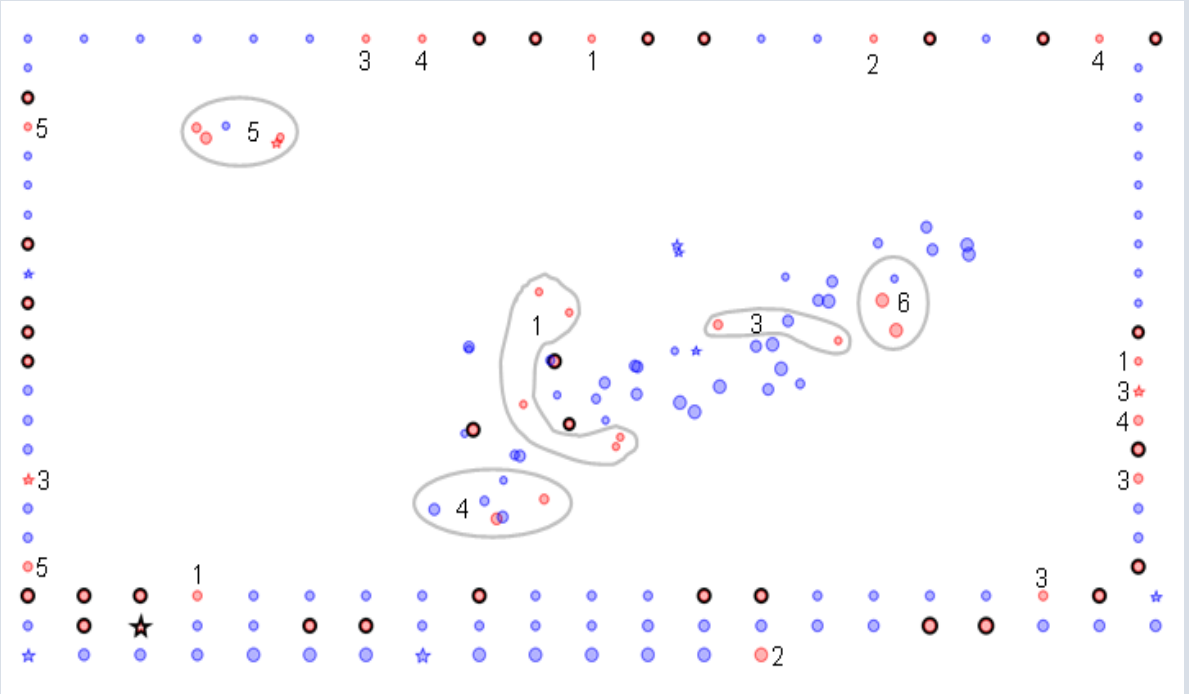
* Slack M et al J Biomol Screen. 2006 Feb;11(1):57-64. Epub 2005 Nov 28

Identification of selective kinase modulators

Inhibitors not interacting with ATP binding pocket

- Difficult target

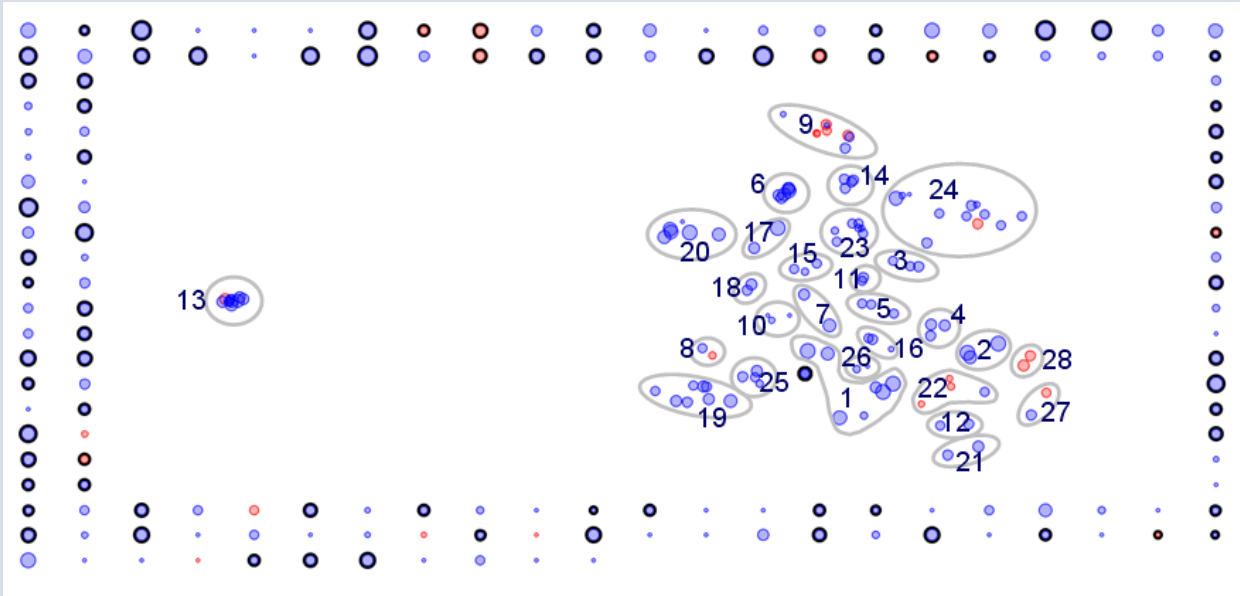
- Size = kinase potency
- red = selective
- blue = ATP active
- star = QC failures



Identification of enzyme inhibitors

Target in field of epigenetics / HTRF assay

- Potency (size)
- Counter screen active (red)
- 30 singletons (black circles)



HTS Benchware dataminer

advantages / caveats

- Easy and intuitive to use application; endpoint for Knime workflow
- Full integration of required workflows
  - Structure import, clustering, visualization, determine mcs, substructure search
- Excellent 2D projection of chemical space
- Manual review and ranking required
- Reporting functionality could be improved (export plots), no Knime input node
- Improvement of spreadsheet functionality
- Only applicable to UNITY fingerprints
- Small user base

Acknowledgements

Discovery Informatics (Hamburg)

Dr. Pierre Ilouga

Dr. Alexander Böcker

Dr. Nicolas Fay (Novartis)

„Pharmaziepraktikant“ Jan Löhr

Discovery Services (Biology Hamburg)

Dr. Thomas Hesterkamp

Dr. Birgit Hecks

Dr. Mark Slack

Dr. Clemens Möller

and colleagues

Computational Chemistry (Oxford)

Dr. Mark Brewer (Schroedinger)

Dr. Oliver Barker

Dr. Richard Law

and colleagues

Medicinal Chemistry (Oxford)

Dr. Tom Coulter

Dr. Mark Whittaker

and colleagues

Knime.com

Organizing committee

evotec

'RESEARCH NEVER STOPS'

KNIME

3rd KNIME Users Meeting and Workshop

February 22-27, 2010, Zurich, Switzerland

Your contact:

Dr. Christian Kirchhoff
Discovery Informatics

+49 (0) 40 56081 390
Christian.Kirchhoff@evotec.com