BioProfile - Extract Knowledge About Compound Cross Reactivities from Corporate Databases

6th Joint Sheffield Conference on Chemoinformatics
24th July, 2013

Dr. Bernd Beck

Boehringer Ingelheim

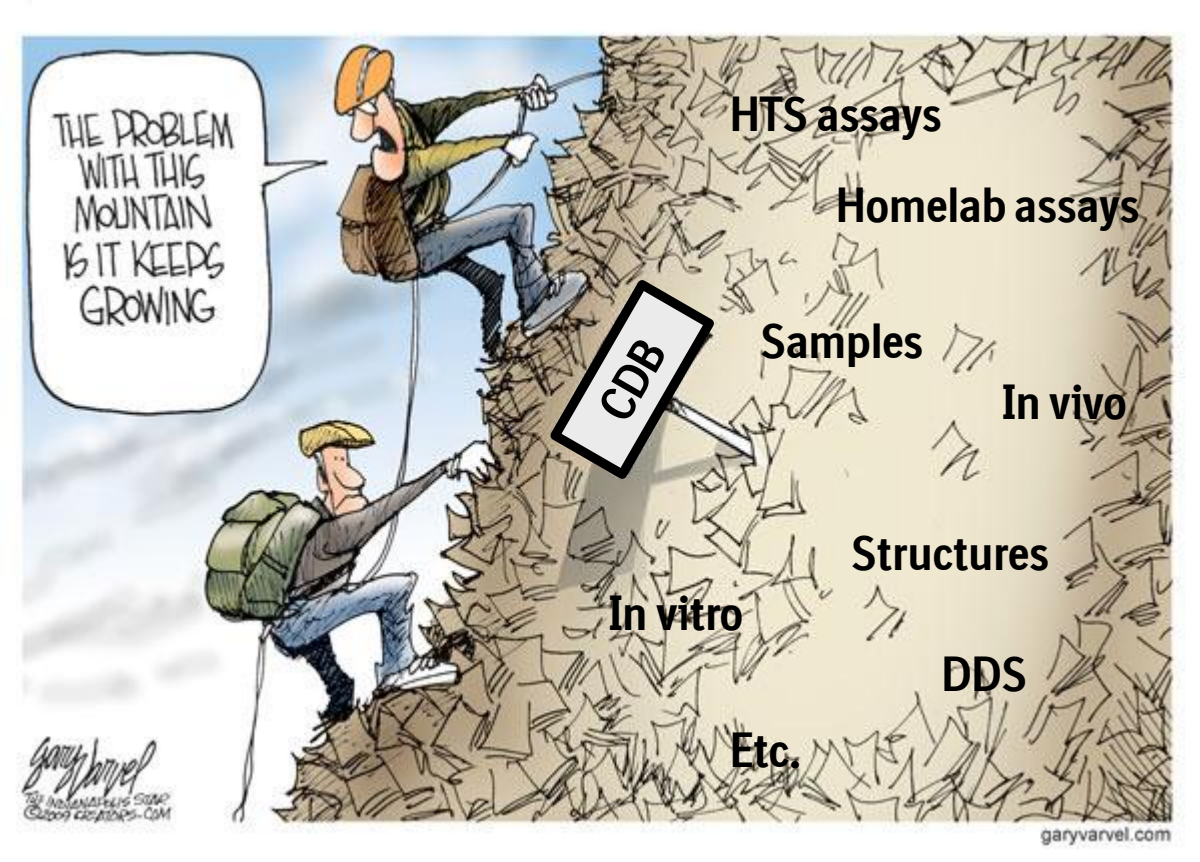
Outline

Boehringer Ingelheim

- Introduction
- Data Preprocessing
  - Data sources
  - Primary Screen Profile
  - Dose Response Profile
  - CRO Screens
- Application Examples
  - Project database
  - BioProfile within the HTS data analysis
  - Frequent hitter analysis
- Summary and Outlook
- Acknowledgement

Introduction

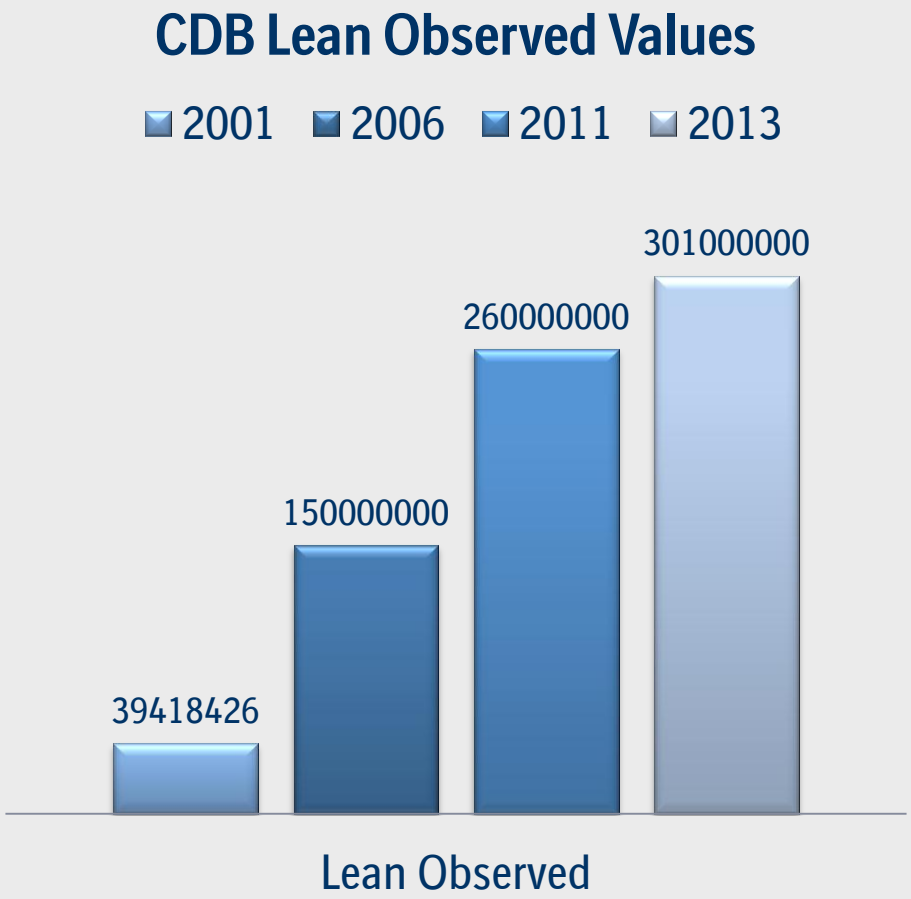
The Corporate Database (CDB)

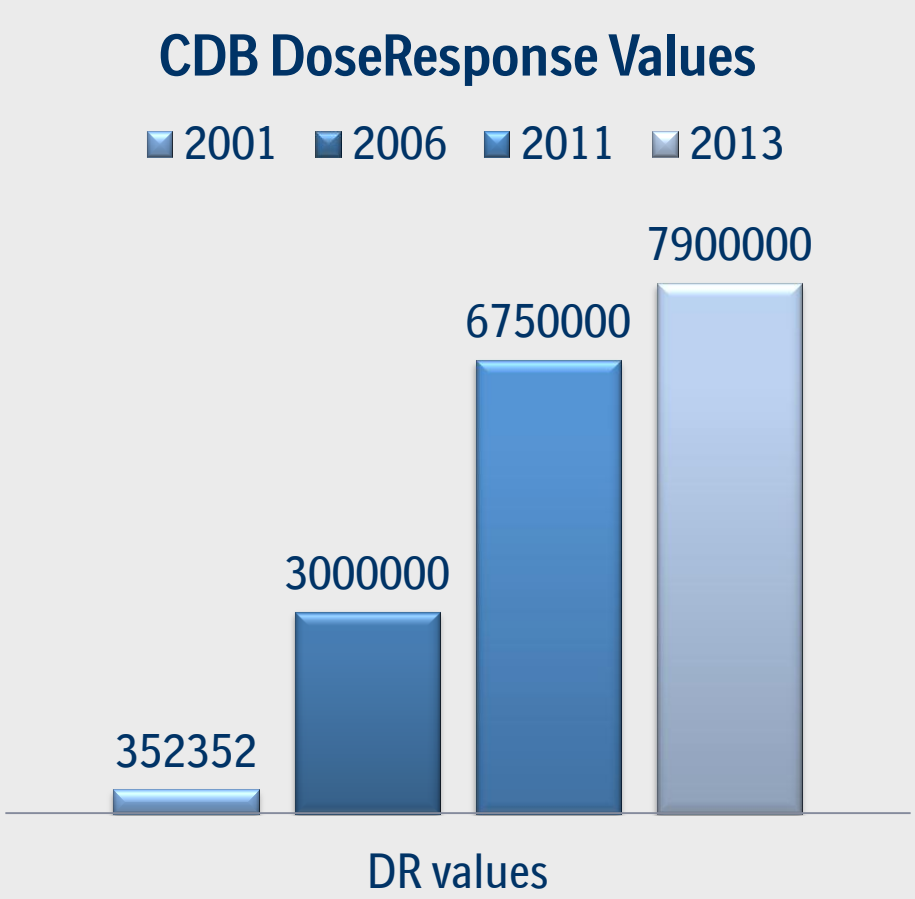


By Gary Varvel,
The Indianapolis Star

A growing data mountain

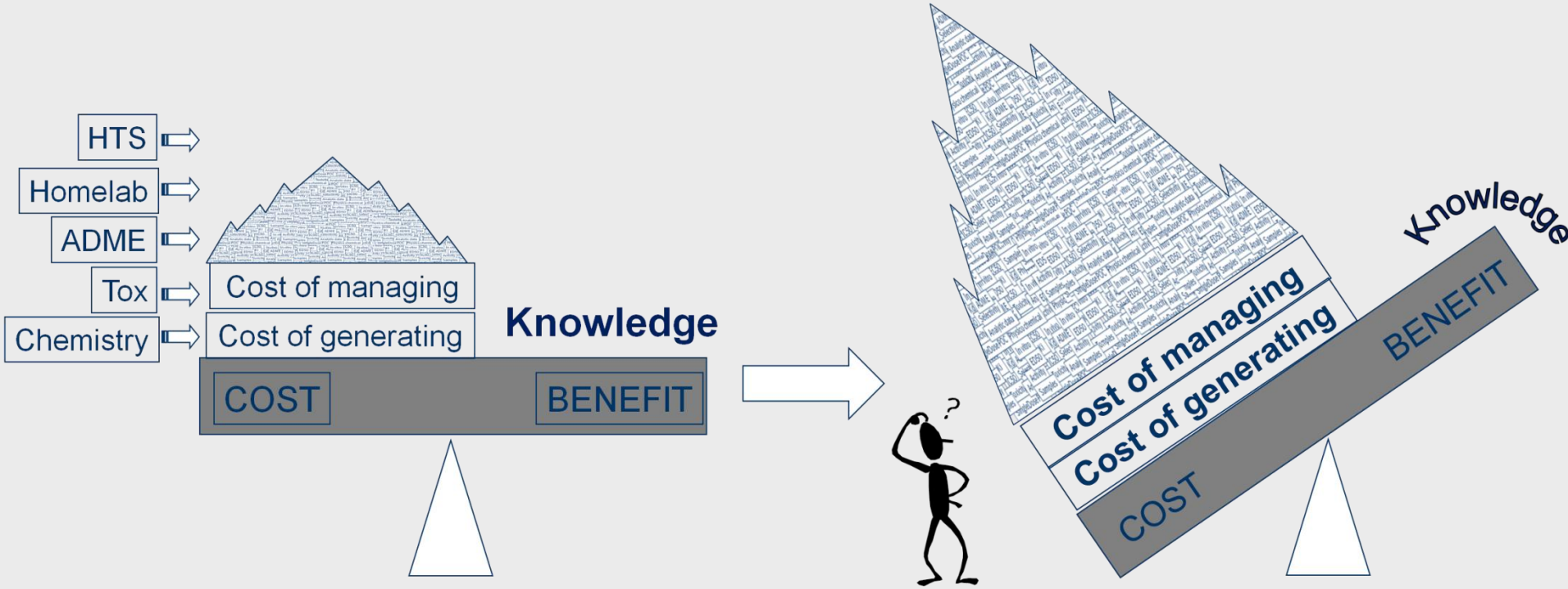
The Corporate Database





Development of the stored single dose and dose response values in the BI corporate database

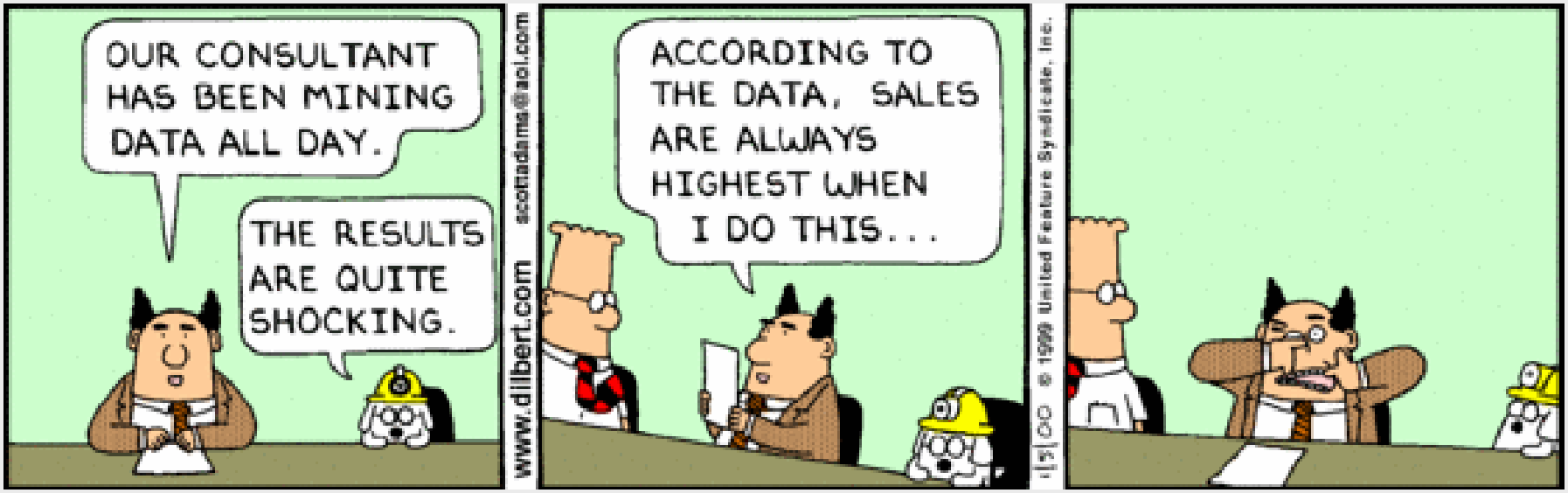
Data mountain – Too much of a good thing?



Data mining

Boehringer Ingelheim

"Data mining is a powerful tool for digging deep into enterprise data to reveal underlying patterns and relationships." (Source WWW)



BioProfile - Motivation

Boehringer Ingelheim

Driven by the question how to prioritize hit classes from HTS or VS we want to find out:

- Is a given compound
  - a real hit or a frequent hitter or an assay technology artifact?
  - Can we identify selectivity targets not known in advance for the HTS hit set or an interesting hit class?
- Are there selectivity issues for a given compound/compound class?
- Are there known toxic effects?

The data is in principle available from CDB; how can we easily access and use it?

Is all necessary information available? Easy accessible?

Data Preprocessing

BioProfile – Data Sources

Boehringer Ingelheim

- Focus on single dose measurements from HTS Screening campaigns from CRO screens and on DoseResponse data
- Get the assay data from CDB
- Get additional data for primary screens directly from the HTS group in Biberach via DB transfer or via files from Vienna and Ridgefield (additional info for about **250** primary screens)
  - Agonist/Antagonist info
  - Hit threshold, mean value and stddev of the screen
  - Target type and technology information
- Target type and technology information is also annotated for DR data and single dose data from CROs (approx. 3000 methods already annotated)
- All the additional data is stored in the CompChem DB

BioProfile - Prepare Data

Boehringer Ingelheim

- retrieve new data from CDB
- process the data, join additional information and store the preprocessed data in the CompChem DB
- automated update every weekend using KNIME workflows

Primary screen data

- generate minimum/maximum values for multiple measurements per compound per method
- approx. **22** million agonist data points and **188** million antagonist data points

DoseRepsonse data

- Convert all values to µM and calculate a median value for multiple measurements per compound per method
- approx. **5.4** million data points

CRO data

- Standardize units to %CTL and test concentration to µM
- approx. **1.6** million data points

Application Examples

BioProfile – Application Examples

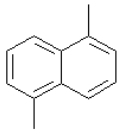
Boehringer Ingelheim

- Project database
- BIMESH HTS data analysis
  - Primary hit filtering
  - Hit set/compound class profile (PrimScreen and DoseResponse data)
- Frequent hitter analysis

BioProfile Data in Project DB – ISIS Base example

Boehringer Ingelheim

Code XYZ | CDB Browser | Lab Sample ID: CCKWAC00620B1



Salt	B.	ID	amou.	unit
BS	1	15015458	36	mg
BS	2	15144715	250	mg

Purpose

Run Date 09.02.2004 | Subm. 25.02.2004 | Expir.

Molformula Sample | Weight

Molformula | Molweight

CDB Comments

Availability

Site	Sample ID	mg Disp
BIDE	15144715	116.8
BIUS	15144715	22.6

Site	Sample ID	Vails
BIDE	15144715	2
BIUS	15144715	

Main	Bioprofile
ChemoInf	BICEPS
PK invitro	RxR
PK invivo	X
CMC&Safety	Y
ExtScreens	Z

Project XYZ

Some information above may change with selected compound batch

Text1 | Activity on target: Avg 9.0

Text2 | LE 0.20

HTS-Analytics / Calculated Properties: M? Area Source Comment

CLOGD 7.4: 0.8

CLogP	MW	TPSA	#Acc	#Don	#Ro5	Andr.
3.6	483	100.6	7	2	0	23

Cluster No	Cluster Size	Cluster Comment
4	26	ExSphere065-Manuell

Dose Response Compound Profile Summary

Meths participated	Avg activity [µM]	min activity [µM]	Meths > 10.0 µM	Meths <=10.0 and >1.0 µM	Meths <=1.0 and >0.1 µM	Meths <= 0.1 µM	Meths with Op
13	7.940	0.0005	2	5		4	2

Dose Response Profile

METHOD ID	Method abbreviation	OP	Avg potency [µM]	RESULT_CLASS
15081	Assay 1		0.0005	IC50
51561	Assay 2		0.0012	IC50
26111	Assay 3		0.0171	IC50
15101	Assay 4		0.0303	IC50
30821	Assay 5		1.0100	IC50
26011	Assay 6		3.8120	IC50
69751	Assay 7		5.5050	IC50
105091	Assay 8		8.9550	IC50

Antagonist - HTS Profile Summary

# of Methods	# of Methods for Compound	# of Methods below hit criteria
212	127	17

Antagonist - HTS Profile

Abbreviation	Unit	Mean Value for Compound	Mean Value for Method	Std Dev for Method	Op Hitcriterion	Hitcriterion for Method
Assay 1	%CTL	38.70	97.40	3.60	<	50.00
Assay 2	%CTL	14.10	96.10	13.90	<	50.00
Assay 3	%CTL	59.40	93.10	6.50	<=	50.00
Assay 4	%CTL	12.40	83.70	14.50	<=	50.00
Assay 5	%CTL	9.00	104.30	27.01	<=	45.00
Assay 6	%CTL	10.00	94.20	27.09	<=	50.00
Assay 7	%CTL	32.00	100.35	19.04	<=	40.00
Assay 8	%CTL	57.20	102.00	14.00	<=	60.00
Assay9	%CTL	3.00	108.19	44.72	<=	40.00

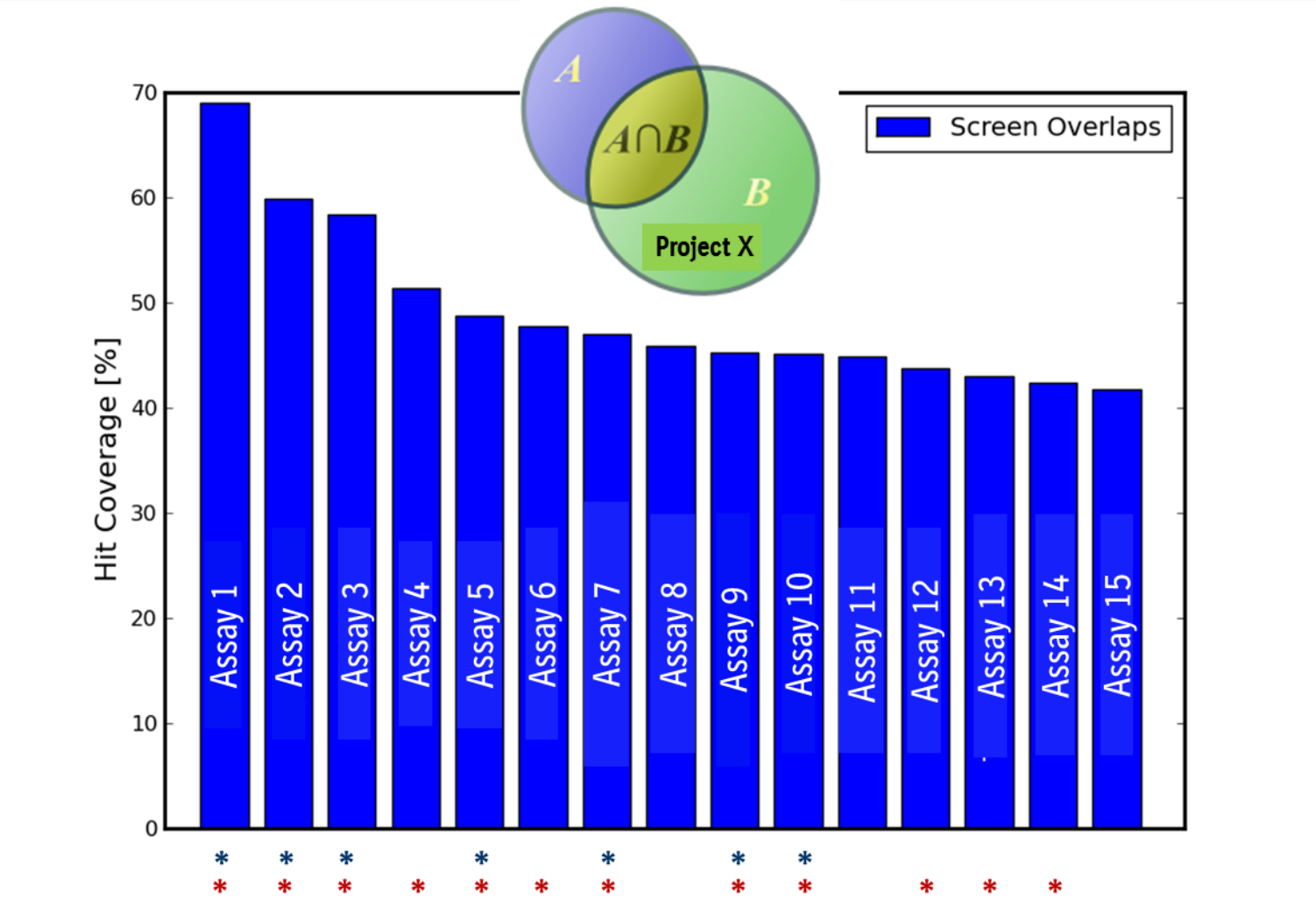
Agonist - HTS Profile Summary

# of Methods	# of Methods for Compound	# of Methods above hit criteria
24	14	1

Agonist - HTS Profile

Abbreviation	Unit	Mean Value for Compound	Mean Value for Method	Std Dev for Method	Op Hitcriterion	Hitcriterion for Method
Assay 1	%CTL	154.00	105.00	13.00	>	150.00
Assay 2	%CTL	115.00	86.50	9.70	>=	120.00
Assay 3	%CTL	125.00	113.00	8.00	>	150.00
Assay 4	%CTL	105.18	100.96	3.66	>	130.00
Assay 5	%CTL	105.00	97.10	4.40	>	110.00
Assay 6	%CTL	104.00	102.00	2.00	>	125.00
Assay 7	%CTL	16.00	1.38	9.42	>=	40.00
Assay 8	%CTL	96.70	98.60	2.20	>	120.00
Assay9	%CTL	83.30	92.00	23.00	>	150.00

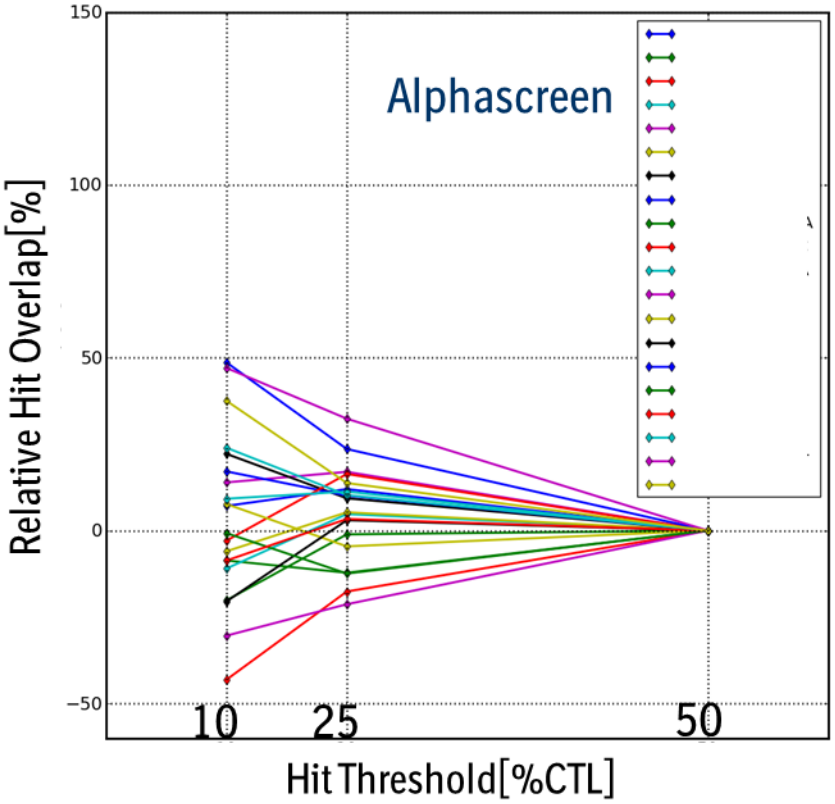
Hit overlaps with previous BI screens (Bioprofile)
(Project X Hitset@25 %CTL, 158000 cpds)

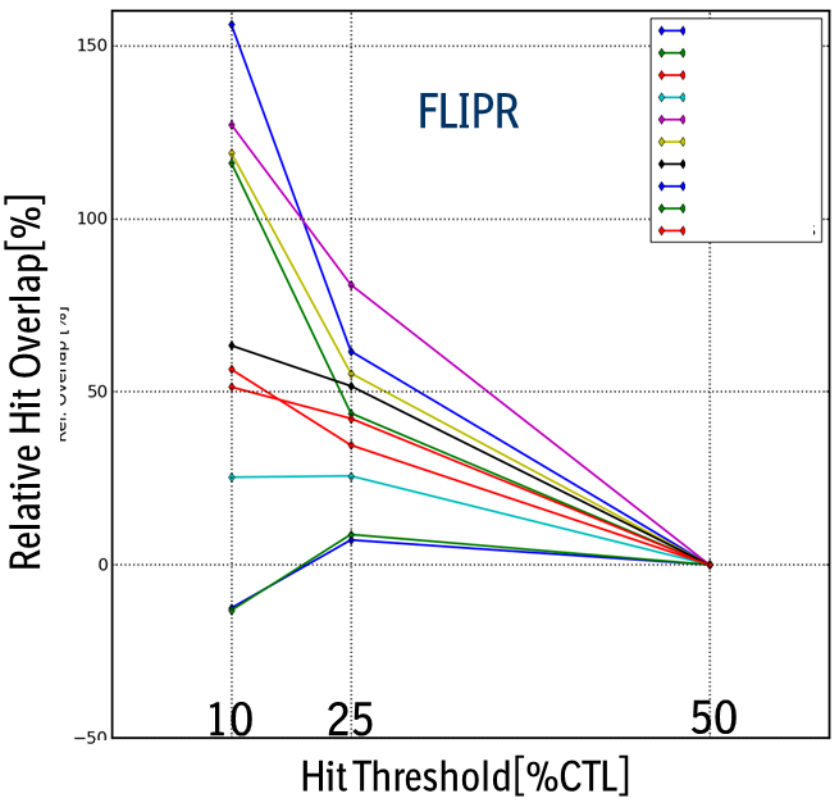


* FLIPR assays; *GPCR/Ion Channel Targets

Technology dependence of hit overlaps

Boehringer Ingelheim



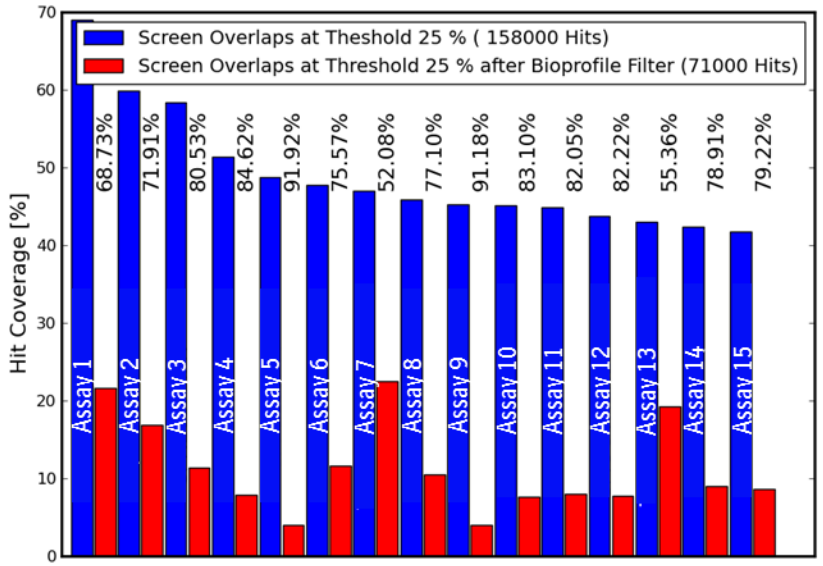


Reducing hit threshold increases relative hit overlaps with other FLIPR screens.
-> Many potent FLIPR hits act unspecific

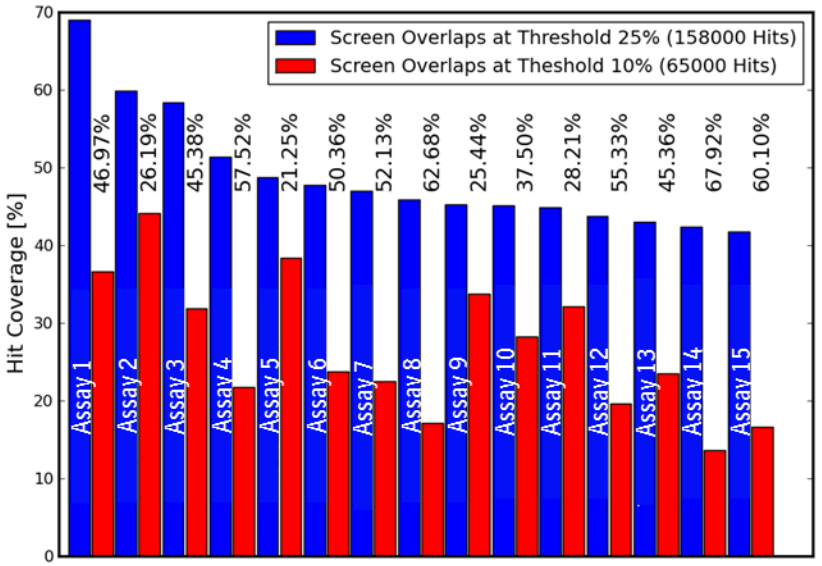
Bioprofile: Filtered Hitset

Boehringer Ingelheim

Overlap reduction through Bioprofile filtering (71000 cpds, 55 % reduction)



Overlap reduction through Hit Threshold adaption to 10 %CTL (65000 cpds, 59 % reduction)



71000 compounds selected for the counterScreen

Hit set Profile – Primary Screen Data

Boehringer Ingelheim

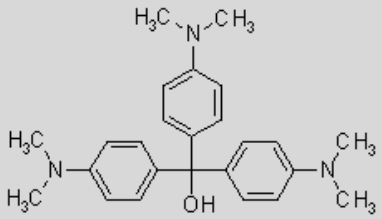
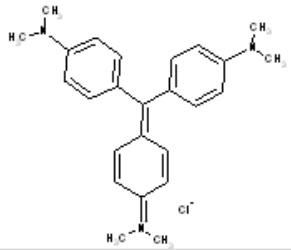
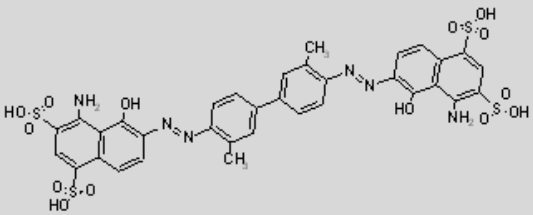
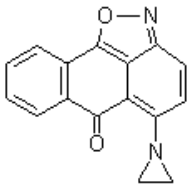
Assay	Assay technology	Assay target type	Hitcrit.	Hitcrit. OP	Hits from Hit set	Hits Assay	Hit rate Assay
Assay24	AlphaScreen	Protease	50	<	5362	41081	4.2
Assay 21	AlphaScreen	Protein binding	50	<=	3556	99957	10.6
Assay 14	SPA red	Transporter	50	<=	2542	91197	10.5
Assay 1	SPA red	Kinase	50	<	1991	13090	2.1
Assay 100	AlphaScreen	Kinase	50	<	1771	34970	4.3
Assay 34	FLINT	Other	50		1512	31553	3.5
Assay 65	SPA red	Kinase	50	<	1490	10194	1.7
Assay 87	Luminescence	Kinase	50	<	1087	14726	1.7
Assay 213	FlashPlate blue	Kinase	50	<	907	4914	0.7
Assay 17	SPA red	Enzyme	20	<	804	26588	3.8
Assay 33	AlphaScreen	Kinase	50	<	757	15001	2.5
Assay 45	FLINT	Enzyme	50	<	588	8394	0.8
Assay 99	LANCE	Kinase	50	<	526	15410	1.6
Assay156	Luminescence	Protein binding	50	<	492	32521	5.5
Assay 108	AlphaScreen	Protein binding	50	<=	436	14117	1.5

Hit set Profile – DoseResponse Data

It easily can be analyzed which other assays overlap with the given hits. The color coding is from green (inactive) to dark red (very active) compounds

Example output for the frequent hitter analysis

Boehringer Ingelheim

ID	Structure	Screens Partic.	# Screeens found as Primary Hits	Primary hit in # different technologies	Primary hit in # different Target Typs	Score	NewScore
MFCD00071920		159	78	9	10	214	19295
MFCD00011750		204	79	10	10	193	19333
MFCD00004021		125	56	11	10	148	16268
MFCD00602221		135	56	8	9	142	10256

Summary & Outlook

Summary

Boehringer Ingelheim

- We have implemented a procedure to collect additional information for all screening campaigns
- Additional information also available for all DR assays and assays from CROs via manual annotation (first automated procedure is under evaluation)
- Automatically extract and preprocess assay data from CDB and join it with the additional data (weekly update)

- All procedures are implemented as KNIME workflows

- BioProfile enables every research project team to easily access information about cross reactivities from corporate databases
- Data can be used to prioritize compounds/compound classes or to check for frequent hitters or counter targets

- BioProfile is part our HTS data analysis (BIMESH) and is used in global research

Outlook

- Further automation for the target type/technology type annotations
- Include external data (e.g. CHEMBL DB)
- Use the BioProfile data for additional analysis tools
  - MMP approach
  - Chemogenomics typ analysis (like Assay Related Target Similarity (ARTS)) *

  ...

* M. Bieler et al J.Chem.Inf.Model. 2011, 51, 1897

Acknowledgements

- Hans Schubert (former colleague from the HTS group Biberach)
- HTS group Biberach
- Renate Schnitzer (Vienna) and Richard Nelson (Ridgefield)

- Daniel Seeliger Computational Chemistry Biberach
- IS support Biberach

Thank you for your attention

- B. Beck, Bioorganic & Medicinal Chemistry 20 (2012) 5428–5435