UNIVERSITÄT LEIPZIG

Interdisciplinary Centre for Bioinformatics

Affymetrix Analysis Data Model (AADM) and data files

Toralf Kirsten

Agenda

- Introduction
- Affy‘s data files
- AADM – Affy analysis data model
    - Affy‘s MicroDB
    - Overview AADM
    - Dimensions / Facts
    - Stars
- Matching Affy Suite facts to AADM
- Access to AADM
- Conclusion

Introduction

AADM - Affymetrix Analysis Data Model

Standard ???

GATC - Genetic Analysis Technology Consortium

Molecular Dynamics and Affymetrix

AADM Notation

Hierarchy

chip → scheme
↓
probe set → biological item
↓
probe pair → atom
↓
probe → cell

---

unit – subset of chip where cells have some similar characteristics

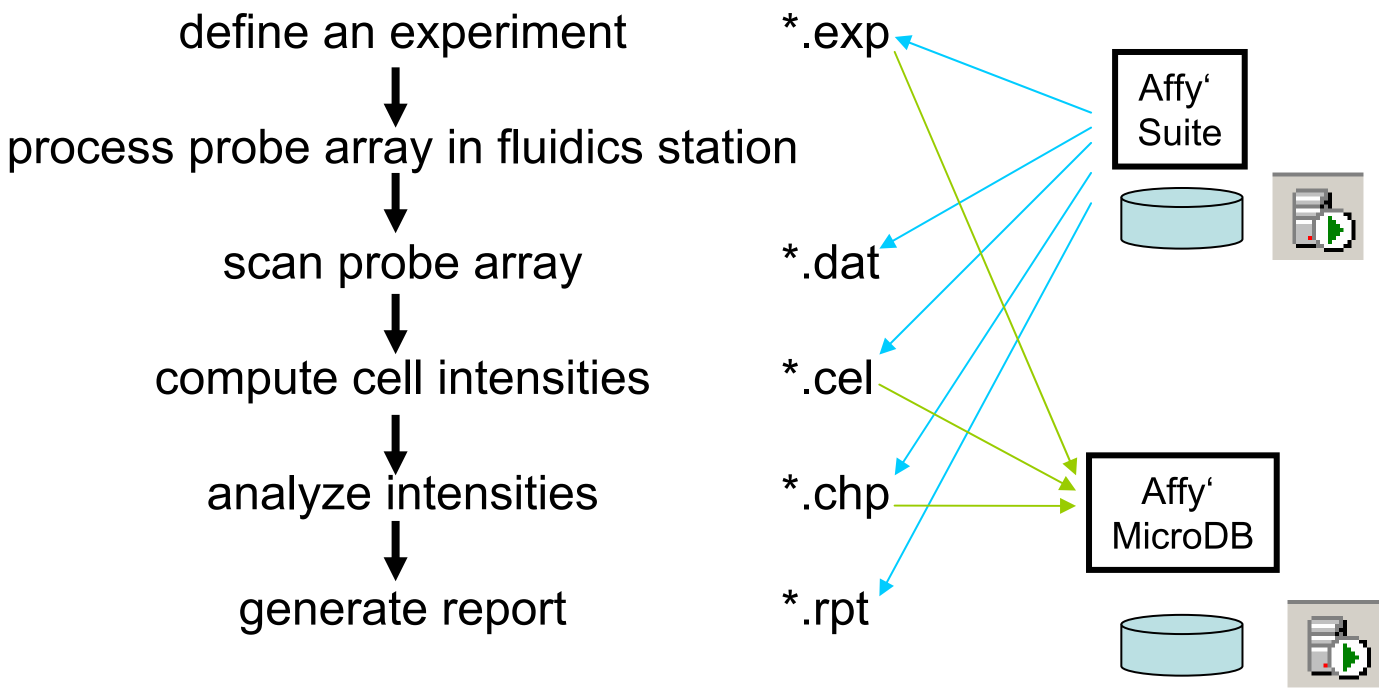
block – subset of a unit where cells have similar characteristics

→ in gene expression studies **block = biological item / block = unit**

Affy‘s data files I

❑ Back end and data logistics

Data files from experimental process



define an an experiment → process probe array in fluidics station → scan probe array → compute cell intensities → analyze intensities → generate report

Library data files

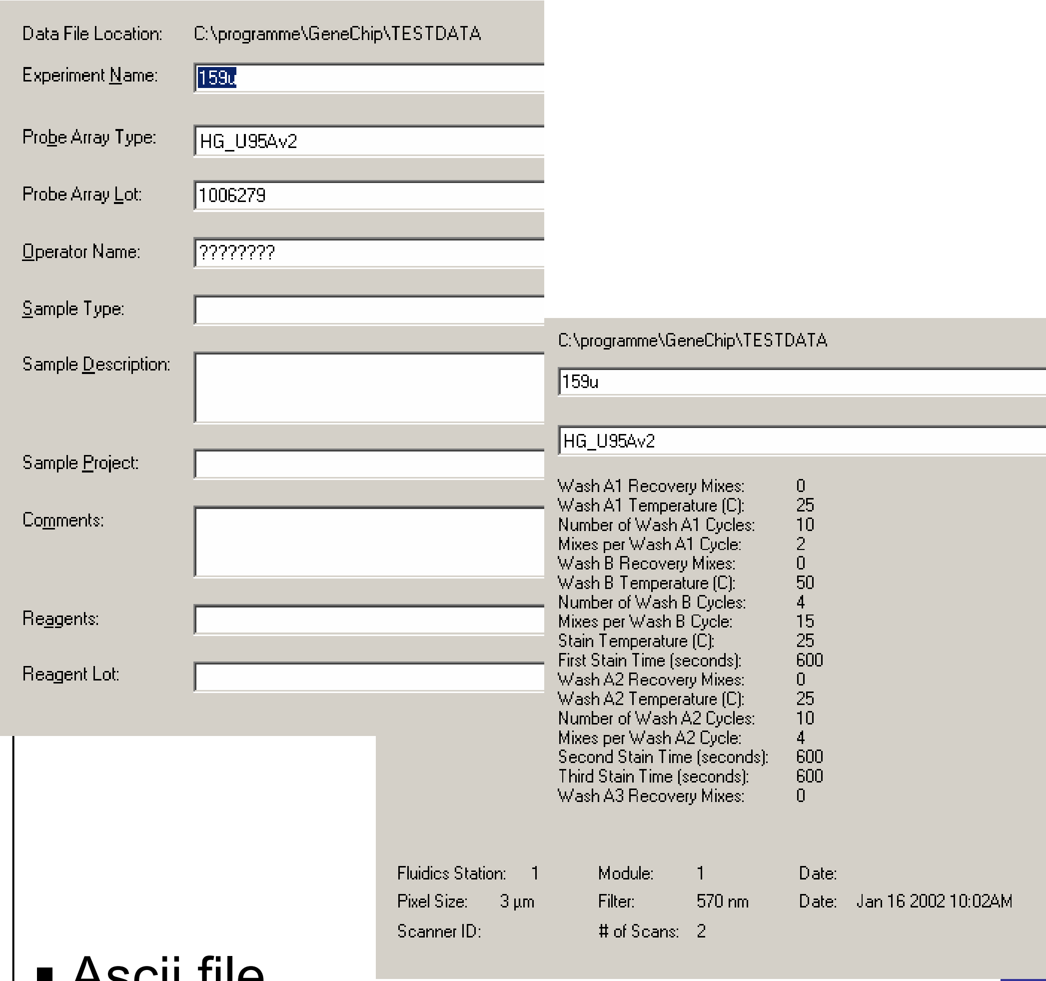
*.cif
*.psi
*.cdf

- MSDE – Microsoft Desktop Engine
- SQL-Server 2000 DB with specific access
- specific file format

Affy‘s data files II

❑Experimental file (*.exp)

```
Affymetrix GeneChip Experiment Information
Version          1

[Sample Info]
Chip Type                                         HG_U95Av2
Chip Lot                                          1006279
Operator                                          ??????????
Sample Type
Description
Project
Comments
Solution Type
Solution Lot

[Fluidics]
Protocol                                          EukGE-WS2v3
Wash A1 Recovery Mixes                            0
Wash A1 Temperature (C)                           25
Number of Wash A1 Cycles                          10
Mixes per Wash A1 Cycle                           2
Wash B Recovery Mixes                             0
Wash B Temperature (C)                            50
Number of Wash B Cycles                           4
Mixes per Wash B Cycle                            15
Stain Temperature (C)                             25
First Stain Time (seconds)                        600
Wash A2 Recovery Mixes                            0
Wash A2 Temperature (C)                           25
…
```



Data File Location: C:\programme\GeneChip\TESTDATA
Experiment Name: 159u
Probe Array Type: HG_U95Av2
Probe Array Lot: 1006279
Operator Name: ????????
Sample Type:
Sample Description:
Sample Project:
Comments:
Reagents:
Reagent Lot:

C:\programme\GeneChip\TESTDATA
159u
HG_U95Av2

Wash A1 Recovery Mixes: 0
Wash A1 Temperature (C): 25
Number of Wash A1 Cycles: 10
Mixes per Wash A1 Cycle: 2
Wash B Recovery Mixes: 0
Wash B Temperature (C): 50
Number of Wash B Cycles: 4
Mixes per Wash B Cycle: 15
Stain Temperature (C): 25
First Stain Time (seconds): 600
Wash A2 Recovery Mixes: 0
Wash A2 Temperature (C): 25
Number of Wash A2 Cycles: 10
Mixes per Wash A2 Cycle: 4
Second Stain Time (seconds): 600
Third Stain Time (seconds): 600
Wash A3 Recovery Mixes: 0

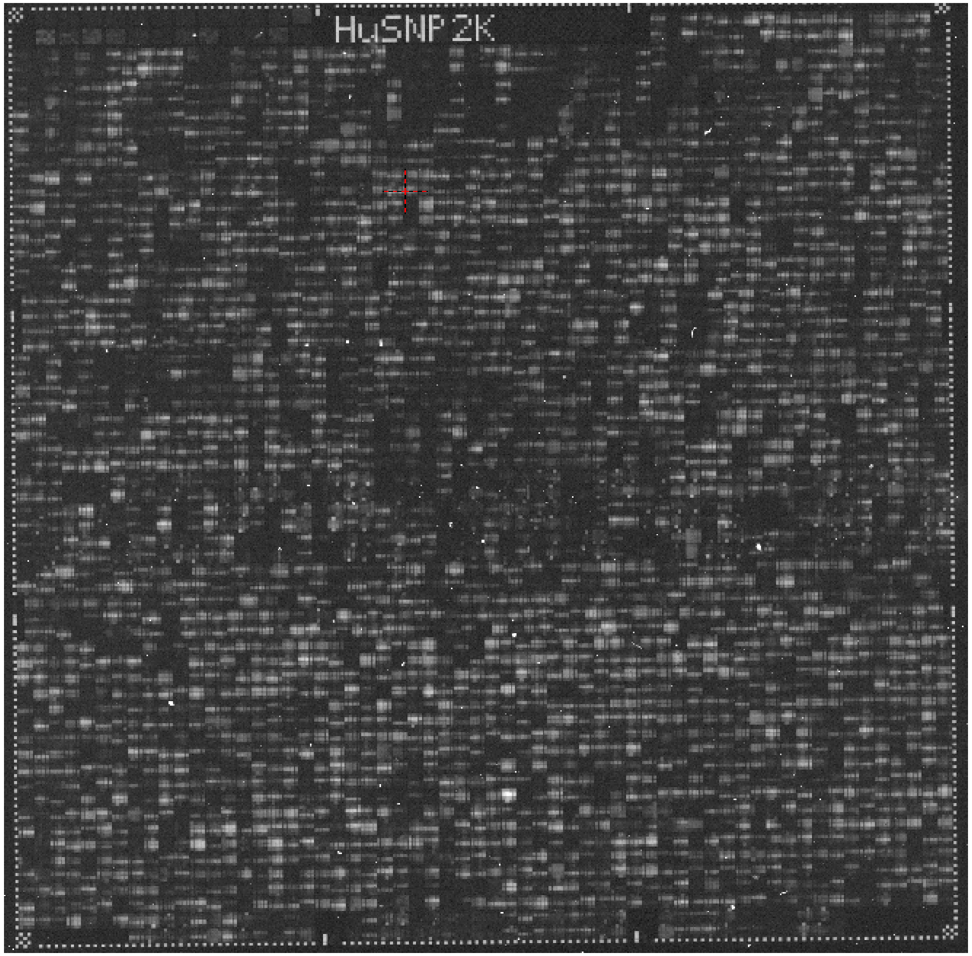
Fluidics Station: 1
Module: 1
Date:
Pixel Size: 3 µm
Filter: 570 nm
Date: Jan 16 2002 10:02AM
Scanner ID:
# of Scans: 2

- Ascii file
- Affy Suite creation
- Control the experiment procedure

Affy‘s data files III

❑Image file (*.dat)

- scan image
- export to *.dat/*.tif
- basis for intensities
- very large file (ca. 40MB)



Affy‘s data files IV

❑Cell intensities (*.cel)

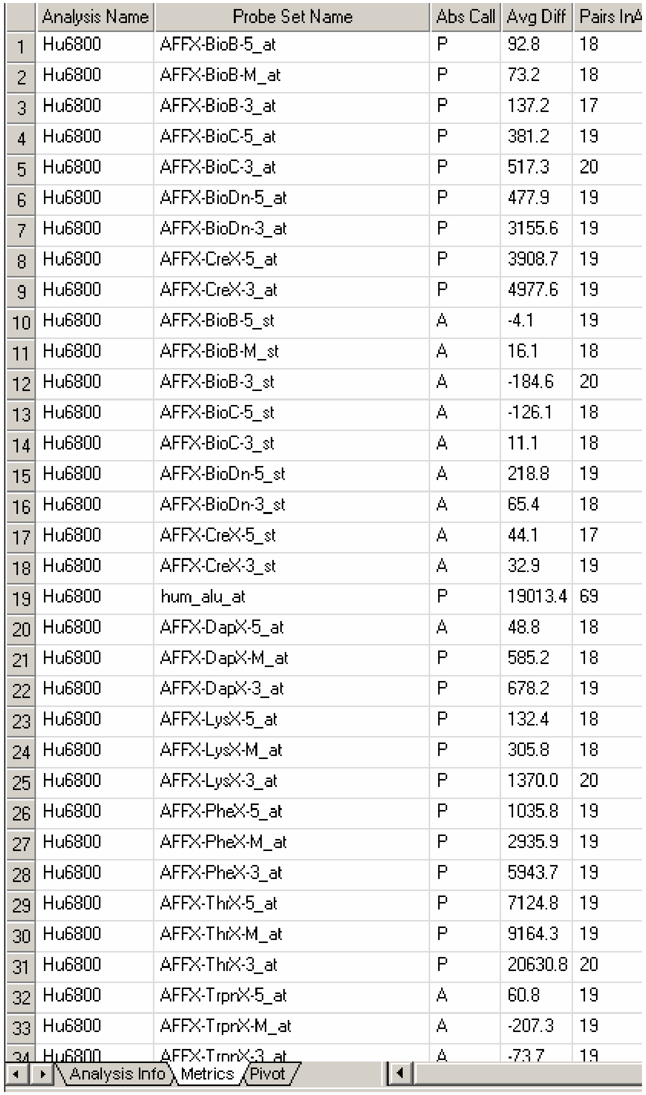
```
[CEL]
Version=3

[HEADER]
Cols=242
Rows=248
TotalX=242
TotalY=248
OffsetX=0
OffsetY=0
GridCornerUL=42 58
GridCornerUR=1385 45
GridCornerLR=1396 1387
GridCornerLL=53 1399
Axis-invertX=0
AxisInvertY=0
swapXY=0
Algorithm=Percentile
AlgorithmParameters=Percentile:75;CellMargin:2;OutlierHigh:1.500;OutlierLow:1.004

[INTENSITY]
NumberCells=60016
CellHeader=X           Y          MEAN       STDV       NPIXELS
             0           0          3179.5     311.6       12
             1           0          46167.0    0.0          9
             2           0          3633.3     410.6       12
             3           0          46167.0    0.0          9
             4           0          2684.5     223.0       12
             5           0          3476.0     205.0        9
             6           0          46167.0    0.0         12
…
```

```
240          247          46167.0     0.0            9
241          247          3457.8      354.8         12

[MASKS]
NumberCells=0
CellHeader=X              Y

[OUTLIERS]
NumberCells=5059
CellHeader=X              Y
             1            0
             3            0
             6            0
             8            0
             10           0
```

- Ascii file
- Affy Suite creation
- XY coordinates without probe set desc.

Affy‘s data files V

❑Chip file (*.chp)



	Analysis Name	Probe Set Name	Abs Call	Avg Diff	Pairs In
1	Hu6800	AFFX-BioB-5_at	P	92.8	18
2	Hu6800	AFFX-BioB-M_at	P	73.2	18
3	Hu6800	AFFX-BioB-3_at	P	137.2	17
4	Hu6800	AFFX-BioC-5_at	P	381.2	19
5	Hu6800	AFFX-BioC-3_at	P	517.3	20
6	Hu6800	AFFX-BioDn-5_at	P	477.9	19
7	Hu6800	AFFX-BioDn-3_at	P	3155.6	19
8	Hu6800	AFFX-CreX-5_at	P	3908.7	19
9	Hu6800	AFFX-CreX-3_at	P	4977.6	19
10	Hu6800	AFFX-BioB-5_st	A	-4.1	19
11	Hu6800	AFFX-BioB-M_st	A	16.1	18
12	Hu6800	AFFX-BioB-3_st	A	-184.6	20
13	Hu6800	AFFX-BioC-5_st	A	-126.1	18
14	Hu6800	AFFX-BioC-3_st	A	11.1	18
15	Hu6800	AFFX-BioDn-5_st	A	218.8	19
16	Hu6800	AFFX-BioDn-3_st	A	65.4	18
17	Hu6800	AFFX-CreX-5_st	A	44.1	17
18	Hu6800	AFFX-CreX-3_st	A	32.9	19
19	Hu6800	hum_alu_at	P	19013.4	69
20	Hu6800	AFFX-DapX-5_at	A	48.8	18
21	Hu6800	AFFX-DapX-M_at	P	585.2	18
22	Hu6800	AFFX-DapX-3_at	P	678.2	19
23	Hu6800	AFFX-LysX-5_at	P	132.4	18
24	Hu6800	AFFX-LysX-M_at	P	305.8	18
25	Hu6800	AFFX-LysX-3_at	P	1370.0	20
26	Hu6800	AFFX-PheX-5_at	P	1035.8	19
27	Hu6800	AFFX-PheX-M_at	P	2935.9	19
28	Hu6800	AFFX-PheX-3_at	P	5943.7	19
29	Hu6800	AFFX-ThrX-5_at	P	7124.8	19
30	Hu6800	AFFX-ThrX-M_at	P	9164.3	19
31	Hu6800	AFFX-ThrX-3_at	P	20630.8	20
32	Hu6800	AFFX-TrpnX-5_at	A	60.8	19
33	Hu6800	AFFX-TrpnX-M_at	A	-207.3	19
34	Hu6800	AFFX-TrpnX-3_at	A	-73.7	19

- very large files (>10 MB)
- specific / proprietary file format
- not readable with other programs
- export to MS Excel / *.txt files

- contains measurement

Affy‘s data files VI

❑Report file (*.rpt)

```
Report Type:    Expression Report
Date:           01:30PM 01/17/2002
_________________________________________________

Filename:                  159U.chp
Probe Array Type:          HG_U95Av2
Algorithm:                 Expression
Probe Pair Thr:  8
Controls:                  Antisense
_________________________________________________

Absolute Thresholds:
                Difference (SDT):          38.6(4.00Q)
                Ratio (SRT):               1.50
Absolute Decision Matrix:                  {{3.0,4.0}
Scaled Noise (Q):                          9.657
Scale Factor (SF):                         5.518
Norm Factor (NF):                          1.000
__________________________________________________

Background:
                Avg: 52.68    Std: 0.97   Min: 51.52   Max: 54.77
Corner+
                Avg: 65                    Count: 32
Corner-
                Avg: 7699                  Count: 32
Central-
                Avg: 7731                  Count: 9
```

```
__________________________________________________

Housekeeping Controls:
Probe Set                 AD(5')     Call(5')    AD(M')     Call(M')
HUMISGF3A/M97935          19.3       A           172.6      P
HUMRGE/M10098             34.9       A           -23.0      A
HUMGAPDH/M33197           681.6      P           807.7      P
HSAC07/X00351             617.5      P           938.7      P
HUMTFRR/M11507            -10.0      A           3.7        A
M27830                    45.5       A           489.9      P


_______________________________________

Spike Controls:
Probe Set                 AD(3')     Call(3')    AD(all)    AD(3'/5')
BIOB                      750.2      P           882.29     1.05
BIOC                      2614.4     P           2520.81    1.08
BIODN                     9373.5     P           5501.90    5.75
CREX                      22098.2    P           20005.11   1.23
DAPX                      4984.9     P           3134.85    3.18
LYSX                      29800.7    P           21886.76   2.18
PHEX                      13933.4    P           10421.32   2.18
THRX                      1248.2     P           958.20     1.66
TRPNX                     260.4      P           89.42      55.70
```

- Ascii file
- Affy Suite creation
- Hybridization qualities

Affy‘s data files VII

- Library files (…/genechip/library/)

cell information file (*.cif)

```
[Chip]
Rows=640
Cols=640
…

[HP]
XOrigin=-7100
YOrigin=8140
…

[TileTypes]
Type1=Expression

[Chip Servers]
BaseCallProgID=GeneChip.CallGEBaseCall.1
CellAvgProgID=GeneChip.PercentileCellAvg.1
ViewProgID1=GeneChip.GESeqView.1

[CellAverage]
Percentile=75
PercentileDefault=75
PercentileMin=0
PercentileMax=100
RejectFactor=6
RejectFactorDefault=6
RejectFactorMin=1
…
```

probe set information (*.psi)

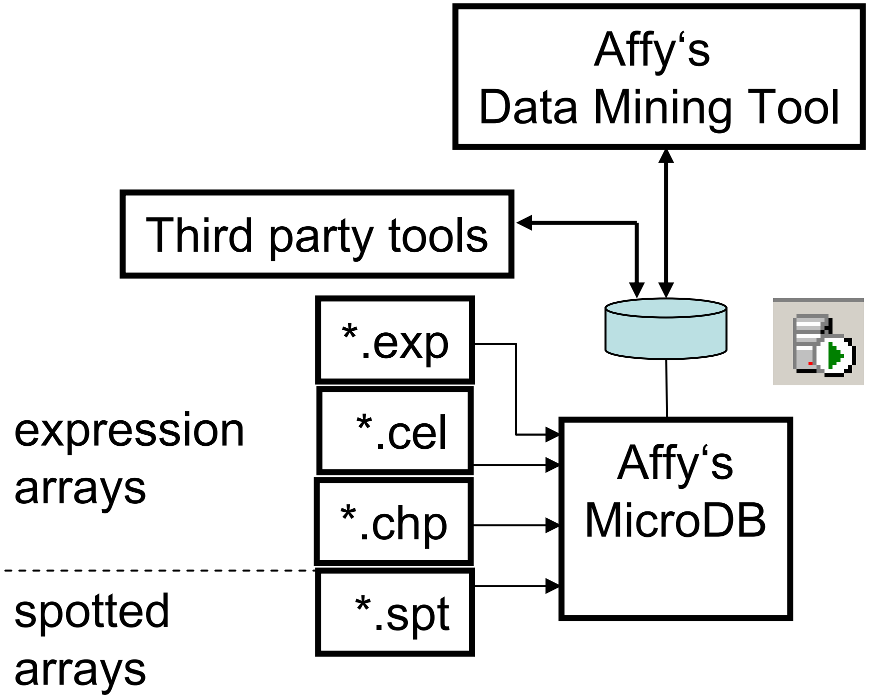
```
#Probe Sets: 12625
1       AFFX-MurIL2_at                        20
2       AFFX-MurIL10_at                       20
3       AFFX-MurIL4_at                        20
4       AFFX-MurFAS_at                        20
10      AFFX-BioB-5_at                        20
11      AFFX-BioB-M_at                        20
12      AFFX-BioB-3_at                        20
13      AFFX-BioC-5_at                        20
…
85      AFFX-HUMGAPDH/M33197_5_st             20
86      AFFX-HUMGAPDH/M33197_M_st             20
87      AFFX-HUMGAPDH/M33197_3_st             20
88      AFFX-HSAC07/X00351_5_st               20
89      AFFX-HSAC07/X00351_M_st               20
96      AFFX-YEL002c/WBP1_at                  20
97      AFFX-YEL018w/_at                      20
98      AFFX-YEL024w/RIP1_at                  20
99      AFFX-YEL021w/URA3_at                  20
100     31307_at                              16
101     31308_at                              16
102     31309_r_at                            16
103     31310_at                              16
104     31311_at                              16
105     31312_at                              16
106     31313_at                              16
107     31314_at                              16
…
```

cell data file (*.cdf)

```
[CDF]
Version=GC3.0

[Chip]
Name=HG_U95Av2
Rows=640
Cols=640
NumberOfUnits=12625
MaxUnit=102119
NumQCUnits=13
ChipReference=

[QC1]
Type=10
NumberCells=300
CellHeader=X    Y     PROBE   PLEN   ATOM   INDEX   MATCH   BG
Cell1=167       80    N       20     0      51367   0       0
Cell2=167       81    N       20     0      52007   1       0
Cell3=167       82    N       20     0      52647   0       0
Cell4=167       83    N       20     0      53287   0       0
Cell5=167       84    N       1      0      53927   -1      1
Cell6=168       80    N       20     1      51368   0       0
Cell7=168       81    N       20     1      52008   1       0
Cell8=168       82    N       20     1      52648   0       0
Cell9=168       83    N       20     1      53288   0       0
Cell10=168      84    N       1      1      53928   -1      1
Cell11=169      80    N       20     2      51369   0       0
Cell12=169      81    N       20     2      52009   1       0
Cell13=169      82    N       20     2      52649   0       0
Cell14=169      83    N       20     2      53289   0       0
Cell15=169      84    N       1      2      53929   -1      1
```

AADM I

❑Affy‘s MicroDB



Affy‘s Data Mining Tool

Third party tools

*.exp

*.cel

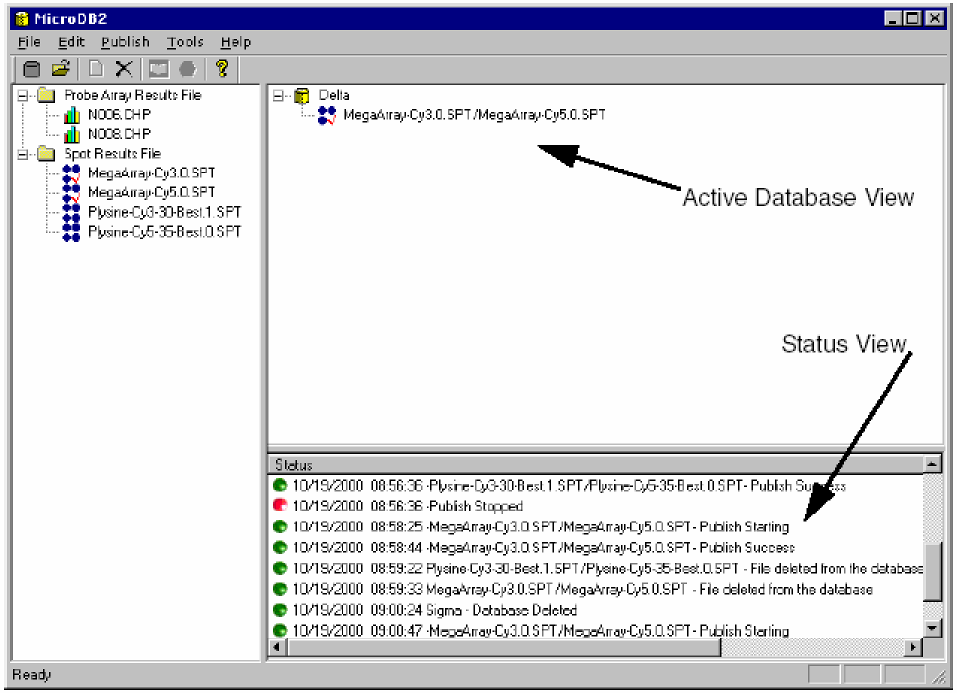
*.chp

*.spt

Affy‘s MicroDB

expression arrays

spotted arrays



Active Database View

Status View

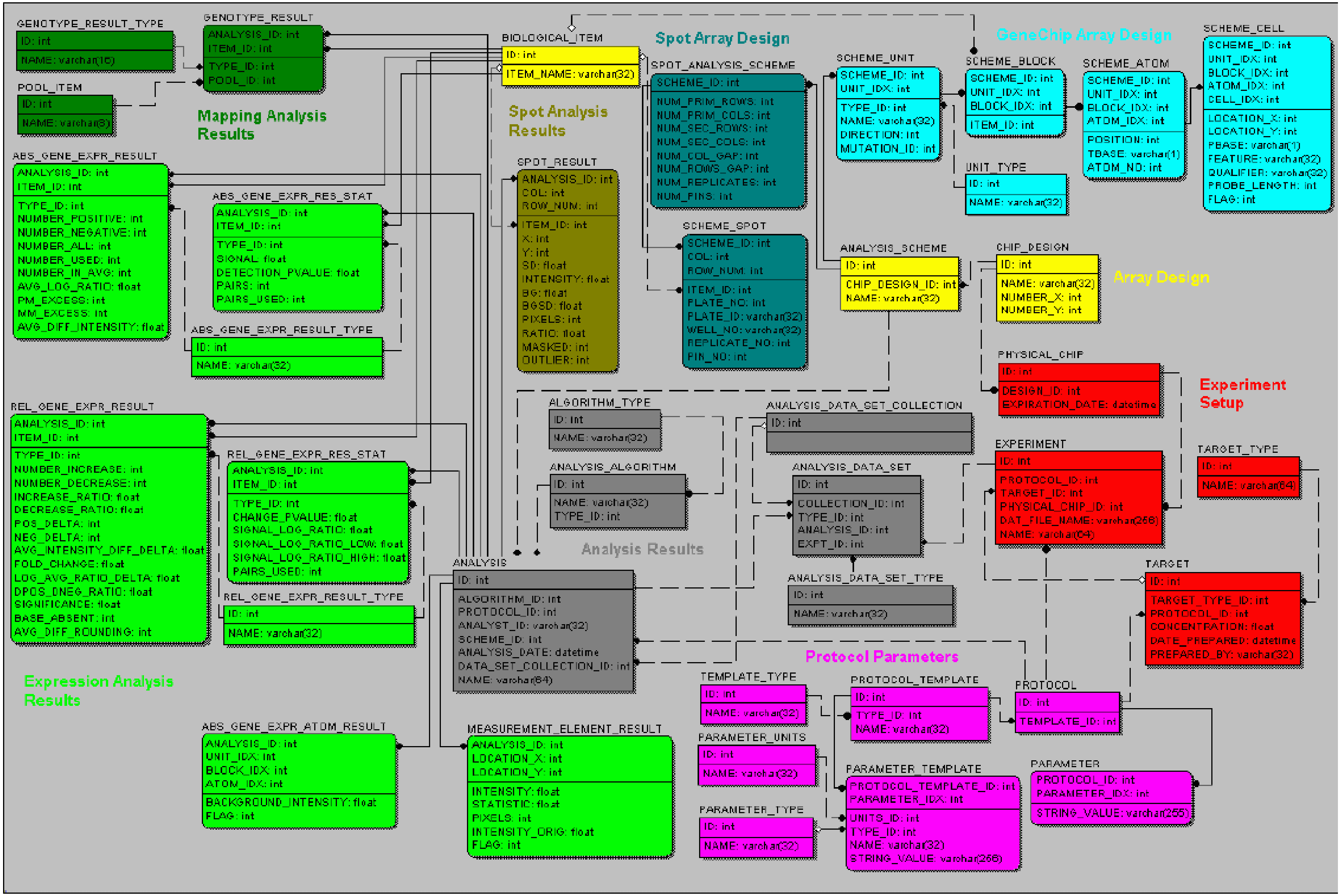
```
10/19/2000  08:56:36 -Plysine-Cy3-30-Best.1.SPT/Plysine-Cy5-35-Best.0.SPT- Publish Success
10/19/2000  08:56:36 -Publish Stopped
10/19/2000  08:58:25 -MegaArray-Cy3.0.SPT/MegaArray-Cy5.0.SPT - Publish Starting
10/19/2000  08:58:44 -MegaArray-Cy3.0.SPT/MegaArray-Cy5.0.SPT - Publish Success
10/19/2000  08:59:22 Plysine-Cy3-30-Best.1.SPT /Plysine-Cy5-35-Best.0.SPT - File deleted from the database
10/19/2000  08:59:33 MegaArray-Cy3.0.SPT /MegaArray-Cy5.0.SPT  - File deleted from the database
10/19/2000  09:00:24 Sigma - Database Deleted
10/19/2000  09:00:47 -MegaArray-Cy3.0.SPT/MegaArray-Cy5.0.SPT - Publish Starting
```

- MSDE – Microsoft Desktop Engine
- SQL-Server 2000 DB with specific access
- specific file format

- one publish db can be opened at the same time
- max. 128 experiments in one db !!!

AADM II

- Overview AADM (subset)



Mapping Analysis Results

GENOTYPE_RESULT_TYPE
ID: int
NAME: varchar(10)

GENOTYPE_RESULT
ANALYSIS_ID: int
ITEM_ID: int
TYPE_ID: int
POOL_ID: int

POOL_ITEM
ID: int
NAME: varchar(8)

Spot Analysis Results

BIOLOGICAL_ITEM
ID: int
ITEM_NAME: varchar(32)

SPOT_RESULT
ANALYSIS_ID: int
COL: int
ROW_NUM: int
ITEM_ID: int
X: int
Y: int
SD: float
INTENSITY: float
BG: float
BGSD: float
PIXELS: int
RATIO: float
MASKED: int
OUTLIER: int

Spot Array Design

SPOT_ANALYSIS_SCHEME
SCHEME_ID: int
NUM_PRIM_ROWS: int
NUM_PRIM_COLS: int
NUM_SEC_ROWS: int
NUM_SEC_COLS: int
NUM_COL_GAP: int
NUM_ROWS_GAP: int
NUM_REPLICATES: int
NUM_PINS: int

SCHEME_SPOT
SCHEME_ID: int
COL: int
ROW_NUM: int
ITEM_ID: int
PLATE_NO: int
PLATE_ID: varchar(32)
WELL_NO: varchar(32)
REPLICATE_NO: int
PIN_NO: int

GeneChip Array Design

SCHEME_UNIT
SCHEME_ID: int
UNIT_IDX: int
TYPE_ID: int
NAME: varchar(32)
DIRECTION: int
MUTATION_ID: int

SCHEME_BLOCK
SCHEME_ID: int
UNIT_IDX: int
BLOCK_IDX: int
ITEM_ID: int

SCHEME_ATOM
SCHEME_ID: int
UNIT_IDX: int
BLOCK_IDX: int
ATOM_IDX: int
POSITION: int
TBASE: varchar(1)
ATOM_NO: int

SCHEME_CELL
SCHEME_ID: int
UNIT_IDX: int
BLOCK_IDX: int
ATOM_IDX: int
CELL_IDX: int
LOCATION_X: int
LOCATION_Y: int
PBASE: varchar(1)
FEATURE: varchar(32)
QUALIFIER: varchar(32)
PROBE_LENGTH: int
FLAG: int

UNIT_TYPE
ID: int
NAME: varchar(32)

Array Design

ANALYSIS_SCHEME
ID: int
CHIP_DESIGN_ID: int
NAME: varchar(32)

CHIP_DESIGN
ID: int
NAME: varchar(32)
NUMBER_X: int
NUMBER_Y: int

Experiment Setup

PHYSICAL_CHIP
ID: int
DESIGN_ID: int
EXPIRATION_DATE: datetime

EXPERIMENT
ID: int
PROTOCOL_ID: int
TARGET_ID: int
PHYSICAL_CHIP_ID: int
DAT_FILE_NAME: varchar(256)
NAME: varchar(64)

TARGET_TYPE
ID: int
NAME: varchar(64)

TARGET
ID: int
TARGET_TYPE_ID: int
PROTOCOL_ID: int
CONCENTRATION: float
DATE_PREPARED: datetime
PREPARED_BY: varchar(32)

Expression Analysis Results

ABS_GENE_EXPR_RESULT
ANALYSIS_ID: int
ITEM_ID: int
TYPE_ID: int
NUMBER_POSITIVE: int
NUMBER_NEGATIVE: int
NUMBER_ALL: int
NUMBER_USED: int
NUMBER_IN_AVG: int
AVG_LOG_RATIO: float
PM_EXCESS: int
MM_EXCESS: int
AVG_DIFF_INTENSITY: float

ABS_GENE_EXPR_RES_STAT
ANALYSIS_ID: int
ITEM_ID: int
TYPE_ID: int
SIGNAL: float
DETECTION_PVALUE: float
PAIRS: int
PAIRS_USED: int

ABS_GENE_EXPR_RESULT_TYPE
ID: int
NAME: varchar(32)

REL_GENE_EXPR_RESULT
ANALYSIS_ID: int
ITEM_ID: int
TYPE_ID: int
NUMBER_INCREASE: int
NUMBER_DECREASE: int
INCREASE_RATIO: float
DECREASE_RATIO: float
POS_DELTA: int
NEG_DELTA: int
AVG_INTENSITY_DIFF_DELTA: float
FOLD_CHANGE: float
LOG_AVG_RATIO_DELTA: float
DPOS_DNEG_RATIO: float
SIGNIFICANCE: float
BASE_ABSENT: int
AVG_DIFF_ROUNDING: int

REL_GENE_EXPR_RES_STAT
ANALYSIS_ID: int
ITEM_ID: int
TYPE_ID: int
CHANGE_PVALUE: float
SIGNAL_LOG_RATIO: float
SIGNAL_LOG_RATIO_LOW: float
SIGNAL_LOG_RATIO_HIGH: float
PAIRS_USED: int

REL_GENE_EXPR_RESULT_TYPE
ID: int
NAME: varchar(32)

ABS_GENE_EXPR_ATOM_RESULT
ANALYSIS_ID: int
UNIT_IDX: int
BLOCK_IDX: int
ATOM_IDX: int
BACKGROUND_INTENSITY: float
FLAG: int

MEASUREMENT_ELEMENT_RESULT
ANALYSIS_ID: int
LOCATION_X: int
LOCATION_Y: int
INTENSITY: float
STATISTIC: float
PIXELS: int
INTENSITY_ORIG: float
FLAG: int

Analysis Results

ALGORITHM_TYPE
ID: int
NAME: varchar(32)

ANALYSIS_ALGORITHM
ID: int
NAME: varchar(32)
TYPE_ID: int

ANALYSIS
ID: int
ALGORITHM_ID: int
PROTOCOL_ID: int
ANALYST_ID: varchar(32)
SCHEME_ID: int
ANALYSIS_DATE: datetime
DATA_SET_COLLECTION_ID: int
NAME: varchar(64)

ANALYSIS_DATA_SET_COLLECTION
ID: int

ANALYSIS_DATA_SET
ID: int
COLLECTION_ID: int
TYPE_ID: int
ANALYSIS_ID: int
EXPT_ID: int

ANALYSIS_DATA_SET_TYPE
ID: int
NAME: varchar(32)

Protocol Parameters

TEMPLATE_TYPE
ID: int
NAME: varchar(32)

PROTOCOL_TEMPLATE
ID: int
TYPE_ID: int
NAME: varchar(32)

PROTOCOL
ID: int
TEMPLATE_ID: int

PARAMETER_UNITS
ID: int
NAME: varchar(32)

PARAMETER_TYPE
ID: int
NAME: varchar(32)

PARAMETER_TEMPLATE
PROTOCOL_TEMPLATE_ID: int
PARAMETER_IDX: int
UNITS_ID: int
TYPE_ID: int
NAME: varchar(32)
STRING_VALUE: varchar(256)

PARAMETER
PROTOCOL_ID: int
PARAMETER_IDX: int
STRING_VALUE: varchar(255)

AADM III

☐ Categories

- Chip design tables
  - gene chip description (name, number of rows/columns, …)
  - spot array description
  - unit description
  - data equivalent to CDF files (library installation)

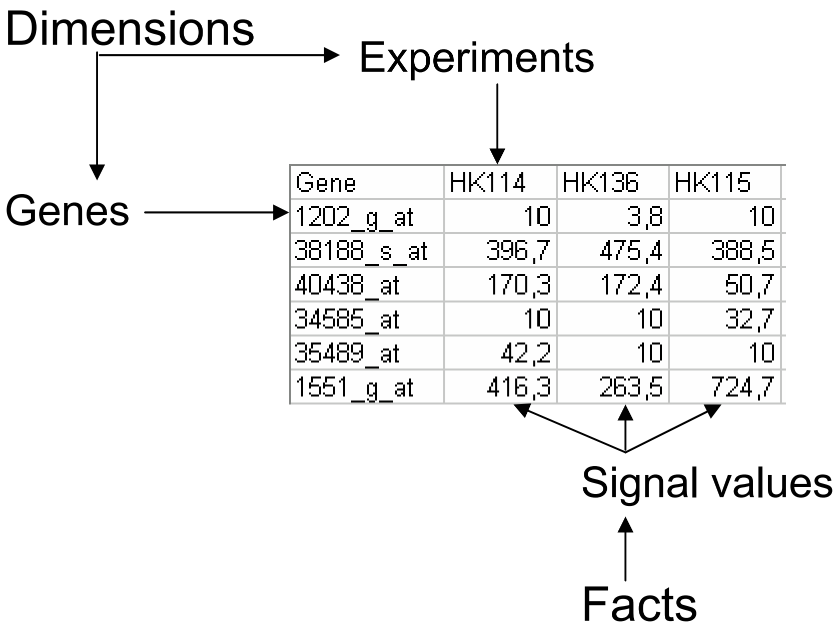
- Experiment setup tables
  - experiment desc. (file name
  - physical chip desc. (relation between experiment and chip design
  - target desc. (concentration, date prepared)

- Analysis result tables
  - cell intensities
  - absolute gene expression ← !!!
  - comparative gene expression

- Protocol parameter tables
  - target preparation
  - experiment setup

AADM IV

❑ Notation of dimensional modeling



Dimensions

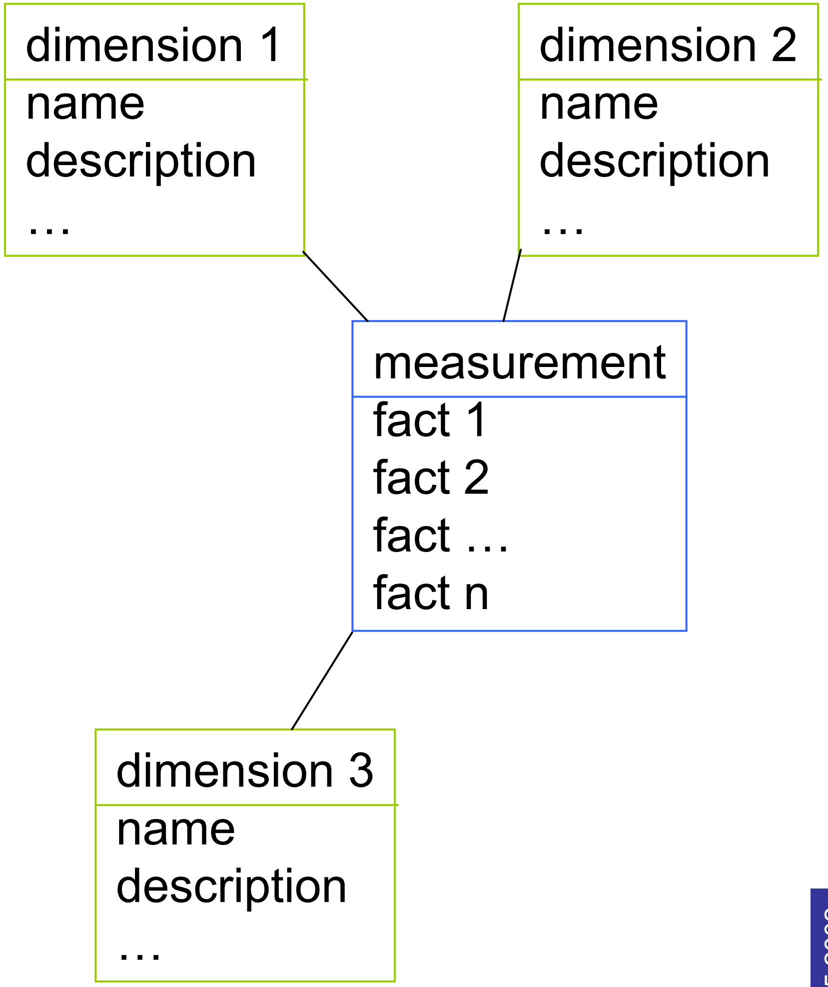
Experiments

Genes

Gene	HK114	HK136	HK115
1202_g_at	10	3,8	10
38188_s_at	396,7	475,4	388,5
40438_at	170,3	172,4	50,7
34585_at	10	10	32,7
35489_at	42,2	10	10
1551_g_at	416,3	263,5	724,7

Signal values

Facts



dimension 1
name
description
…

dimension 2
name
description
…

measurement
fact 1
fact 2
fact …
fact n

dimension 3
name
description
…

- Dimensions
  - more static character
  - descriptions
- Facts
  - measurements (numbers and values)

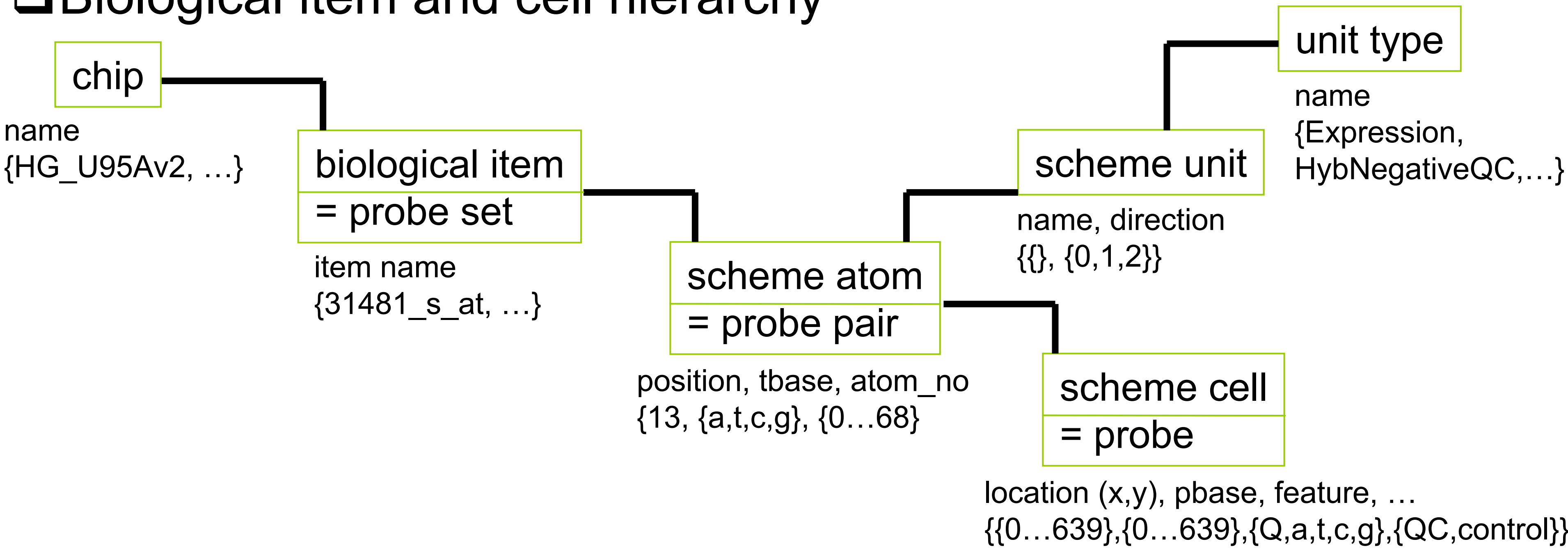
Demodata vs. Real experiment data

- ☐ database analysis

	demodata	**real experiment data**
Absolute gene expression results	filled	empty
Absolute gene expression statistical results	filled	filled
Relative gene expression results	filled	empty
Relative gene expression statistical results	filled	filled
Background intensities	empty	empty
Cell intensities	empty	empty

AADM – Dimensions I

❑Biological item and cell hierarchy



chip

name
{HG_U95Av2, …}

biological item
= probe set

item name
{31481_s_at, …}

scheme atom
= probe pair

position, tbase, atom_no
{13, {a,t,c,g}, {0…68}

scheme unit

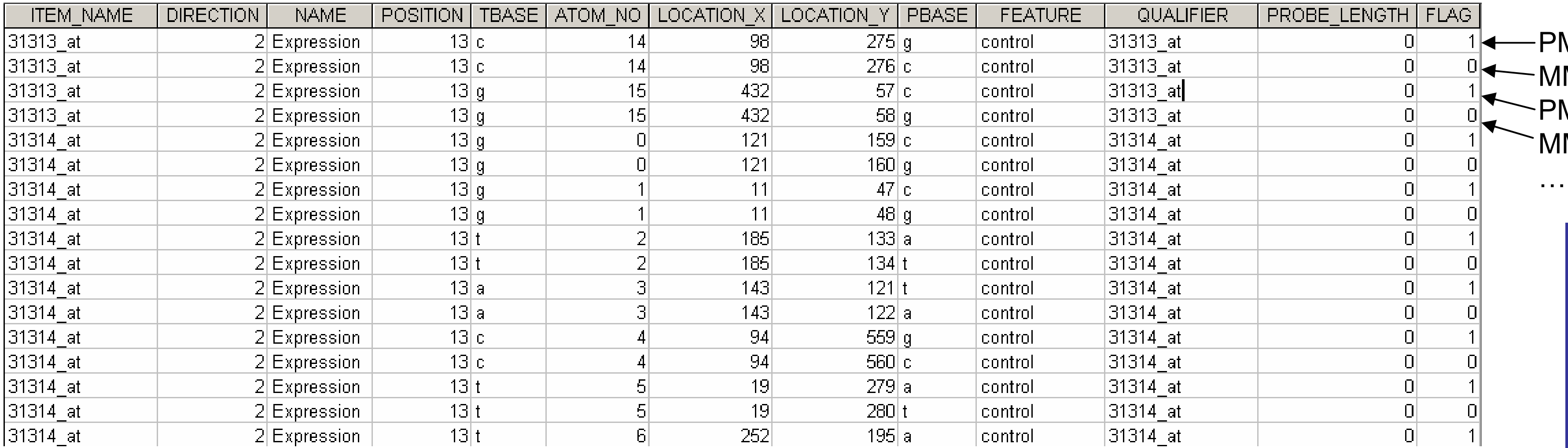
name, direction
{{}, {0,1,2}}

unit type

name
{Expression,
HybNegativeQC,…}

scheme cell
= probe

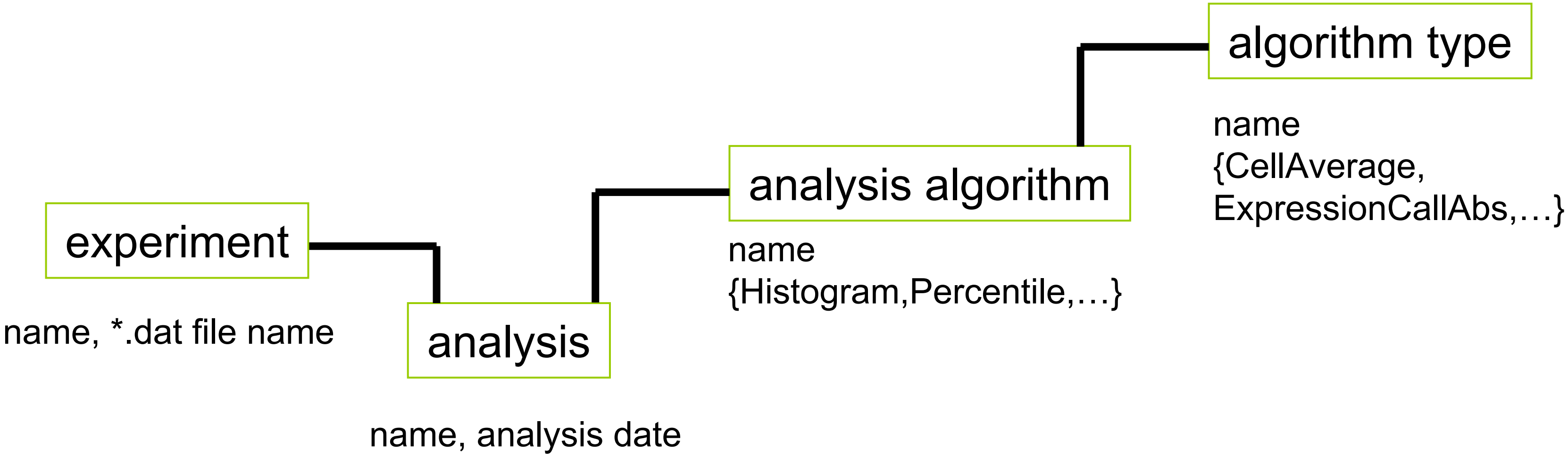
location (x,y), pbase, feature, …
{{0…639},{0…639},{Q,a,t,c,g},{QC,control}}



ITEM_NAME	DIRECTION	NAME	POSITION	TBASE	ATOM_NO	LOCATION_X	LOCATION_Y	PBASE	FEATURE	QUALIFIER	PROBE_LENGTH	FLAG
31313_at	2	Expression	13	c	14	98	275	g	control	31313_at	0	1
31313_at	2	Expression	13	c	14	98	276	c	control	31313_at	0	0
31313_at	2	Expression	13	g	15	432	57	c	control	31313_at	0	1
31313_at	2	Expression	13	g	15	432	58	g	control	31313_at	0	0
31314_at	2	Expression	13	g	0	121	159	c	control	31314_at	0	1
31314_at	2	Expression	13	g	0	121	160	g	control	31314_at	0	0
31314_at	2	Expression	13	g	1	11	47	c	control	31314_at	0	1
31314_at	2	Expression	13	g	1	11	48	g	control	31314_at	0	0
31314_at	2	Expression	13	t	2	185	133	a	control	31314_at	0	1
31314_at	2	Expression	13	t	2	185	134	t	control	31314_at	0	0
31314_at	2	Expression	13	a	3	143	121	t	control	31314_at	0	1
31314_at	2	Expression	13	a	3	143	122	a	control	31314_at	0	0
31314_at	2	Expression	13	c	4	94	559	g	control	31314_at	0	1
31314_at	2	Expression	13	c	4	94	560	c	control	31314_at	0	0
31314_at	2	Expression	13	t	5	19	279	a	control	31314_at	0	1
31314_at	2	Expression	13	t	5	19	280	t	control	31314_at	0	0
31314_at	2	Expression	13	t	6	252	195	a	control	31314_at	0	1

AADM – Dimensions II

- Experiment and analysis hierarchy



experiment

name, *.dat file name

analysis

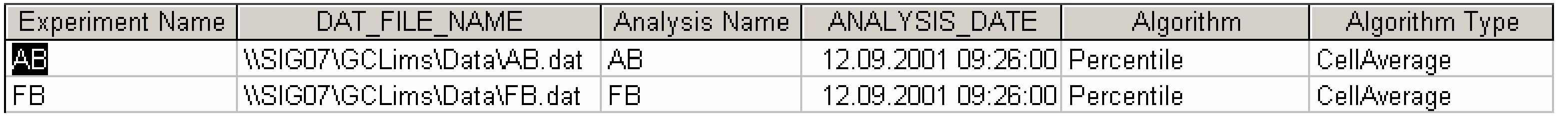
name, analysis date

analysis algorithm

name
{Histogram,Percentile,…}

algorithm type

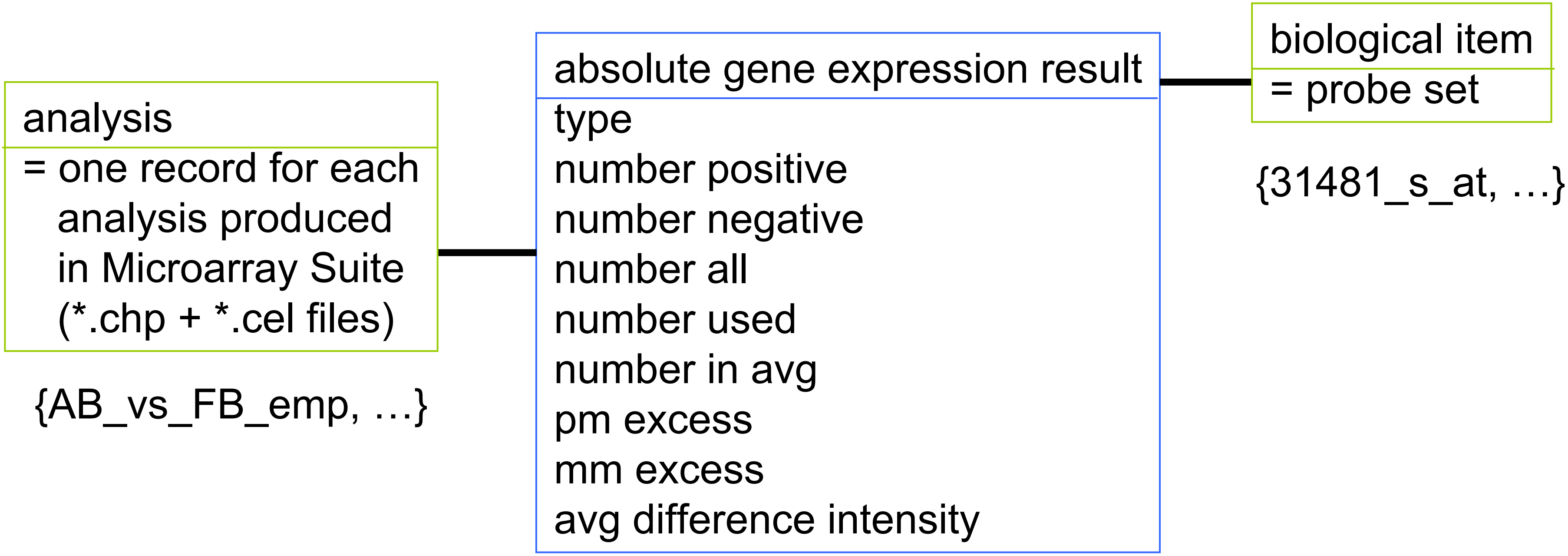
name
{CellAverage,
ExpressionCallAbs,…}



Experiment Name	DAT_FILE_NAME	Analysis Name	ANALYSIS_DATE	Algorithm	Algorithm Type
AB	\\SIG07\GCLims\Data\AB.dat	AB	12.09.2001 09:26:00	Percentile	CellAverage
FB	\\SIG07\GCLims\Data\FB.dat	FB	12.09.2001 09:26:00	Percentile	CellAverage

AADM – Stars I

❑Absolute gene expression results



analysis

= one record for each analysis produced in Microarray Suite (*.chp + *.cel files)

{AB_vs_FB_emp, …}

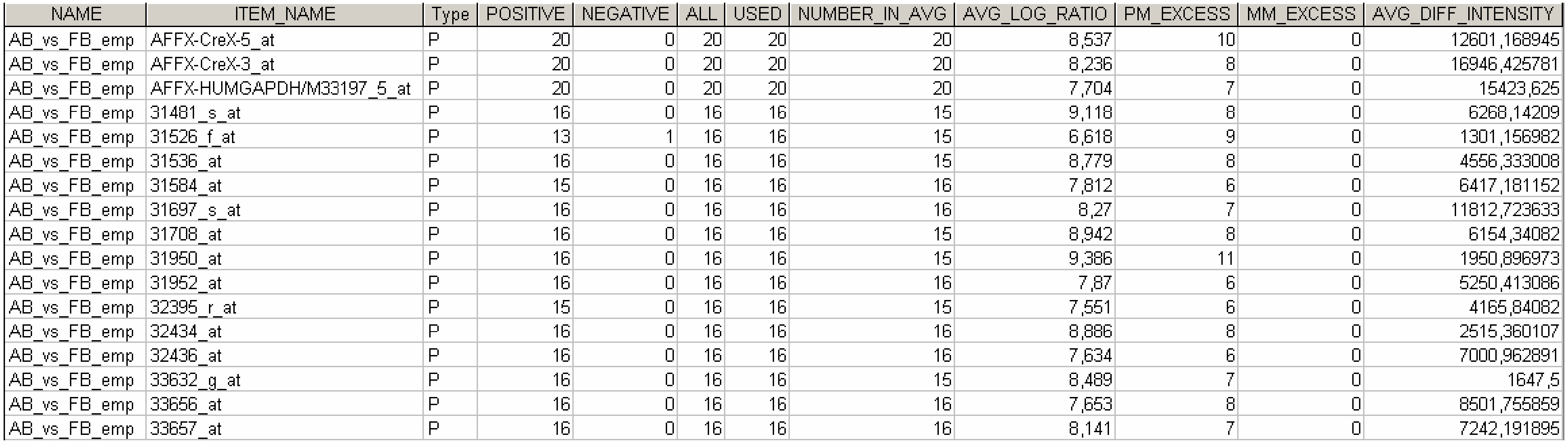
absolute gene expression result

type
number positive
number negative
number all
number used
number in avg
pm excess
mm excess
avg difference intensity

biological item

= probe set

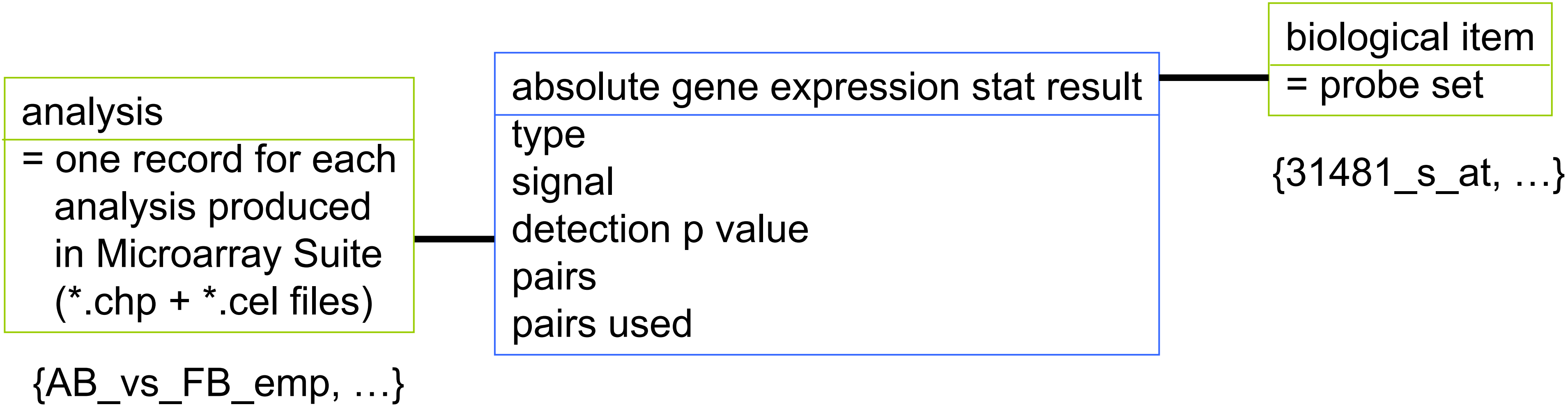
{31481_s_at, …}



NAME	ITEM_NAME	Type	POSITIVE	NEGATIVE	ALL	USED	NUMBER_IN_AVG	AVG_LOG_RATIO	PM_EXCESS	MM_EXCESS	AVG_DIFF_INTENSITY
AB_vs_FB_emp	AFFX-CreX-5_at	P	20	0	20	20	20	8,537	10	0	12601,168945
AB_vs_FB_emp	AFFX-CreX-3_at	P	20	0	20	20	20	8,236	8	0	16946,425781
AB_vs_FB_emp	AFFX-HUMGAPDH/M33197_5_at	P	20	0	20	20	20	7,704	7	0	15423,625
AB_vs_FB_emp	31481_s_at	P	16	0	16	16	15	9,118	8	0	6268,14209
AB_vs_FB_emp	31526_f_at	P	13	1	16	16	15	6,618	9	0	1301,156982
AB_vs_FB_emp	31536_at	P	16	0	16	16	15	8,779	8	0	4556,333008
AB_vs_FB_emp	31584_at	P	15	0	16	16	16	7,812	6	0	6417,181152
AB_vs_FB_emp	31697_s_at	P	16	0	16	16	16	8,27	7	0	11812,723633
AB_vs_FB_emp	31708_at	P	16	0	16	16	15	8,942	8	0	6154,34082
AB_vs_FB_emp	31950_at	P	16	0	16	16	15	9,386	11	0	1950,896973
AB_vs_FB_emp	31952_at	P	16	0	16	16	16	7,87	6	0	5250,413086
AB_vs_FB_emp	32395_r_at	P	15	0	16	16	15	7,551	6	0	4165,84082
AB_vs_FB_emp	32434_at	P	16	0	16	16	16	8,886	8	0	2515,360107
AB_vs_FB_emp	32436_at	P	16	0	16	16	16	7,634	6	0	7000,962891
AB_vs_FB_emp	33632_g_at	P	16	0	16	16	15	8,489	7	0	1647,5
AB_vs_FB_emp	33656_at	P	16	0	16	16	16	7,653	8	0	8501,755859
AB_vs_FB_emp	33657_at	P	16	0	16	16	16	8,141	7	0	7242,191895

AADM – Stars II

- ❑ Absolute gene expression statistical results



analysis

= one record for each analysis produced in Microarray Suite (*.chp + *.cel files)

{AB_vs_FB_emp, …}

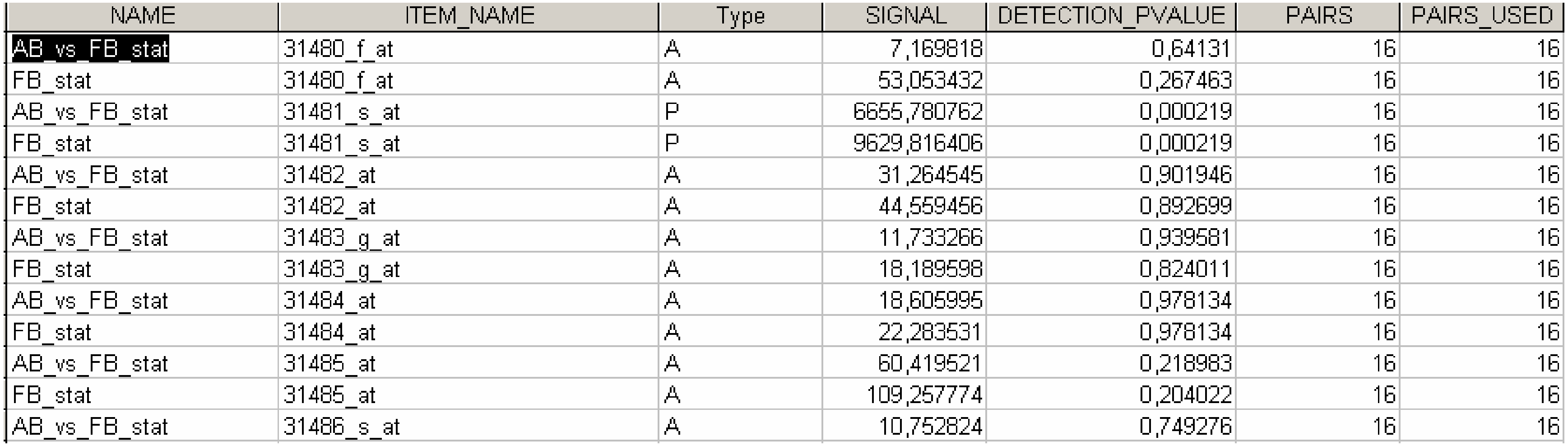
absolute gene expression stat result

type
signal
detection p value
pairs
pairs used

biological item

= probe set

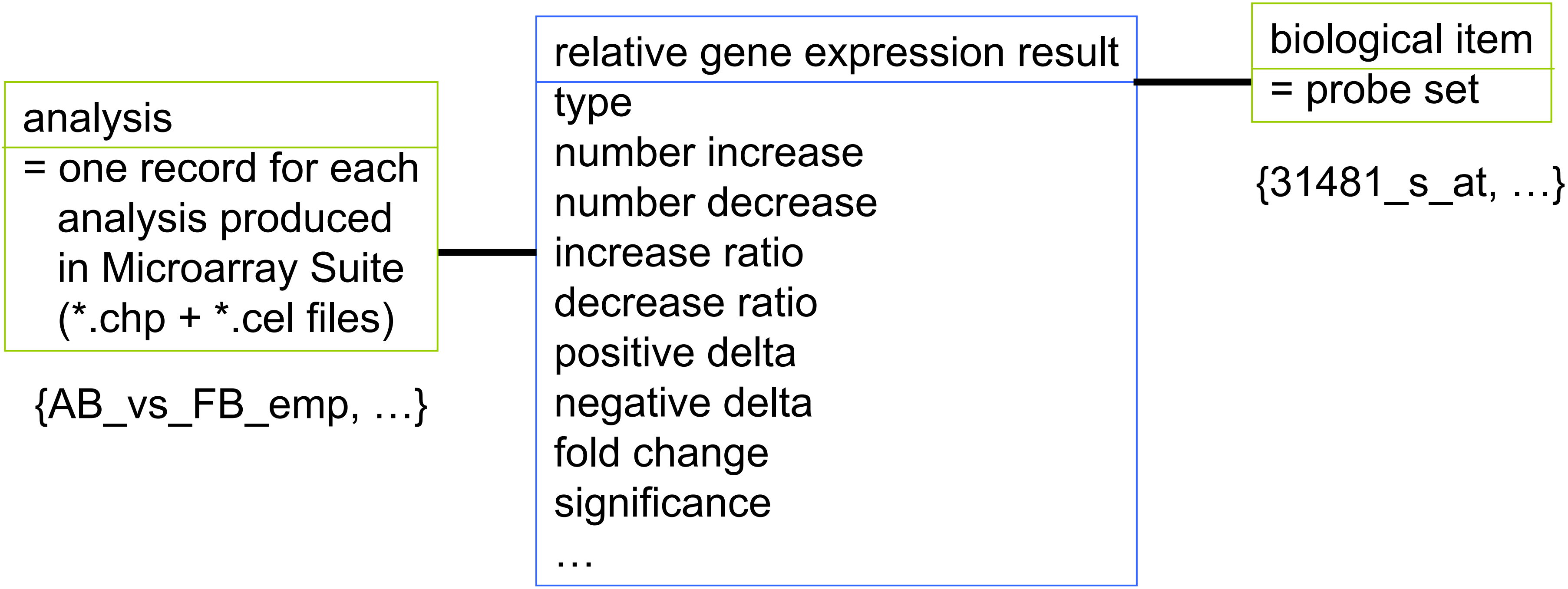
{31481_s_at, …}



NAME	ITEM_NAME	Type	SIGNAL	DETECTION_PVALUE	PAIRS	PAIRS_USED
AB_vs_FB_stat	31480_f_at	A	7,169818	0,64131	16	16
FB_stat	31480_f_at	A	53,053432	0,267463	16	16
AB_vs_FB_stat	31481_s_at	P	6655,780762	0,000219	16	16
FB_stat	31481_s_at	P	9629,816406	0,000219	16	16
AB_vs_FB_stat	31482_at	A	31,264545	0,901946	16	16
FB_stat	31482_at	A	44,559456	0,892699	16	16
AB_vs_FB_stat	31483_g_at	A	11,733266	0,939581	16	16
FB_stat	31483_g_at	A	18,189598	0,824011	16	16
AB_vs_FB_stat	31484_at	A	18,605995	0,978134	16	16
FB_stat	31484_at	A	22,283531	0,978134	16	16
AB_vs_FB_stat	31485_at	A	60,419521	0,218983	16	16
FB_stat	31485_at	A	109,257774	0,204022	16	16
AB_vs_FB_stat	31486_s_at	A	10,752824	0,749276	16	16

AADM – Stars III

- Relative gene expression results (comparative results)



analysis

= one record for each analysis produced in Microarray Suite (*.chp + *.cel files)

{AB_vs_FB_emp, …}

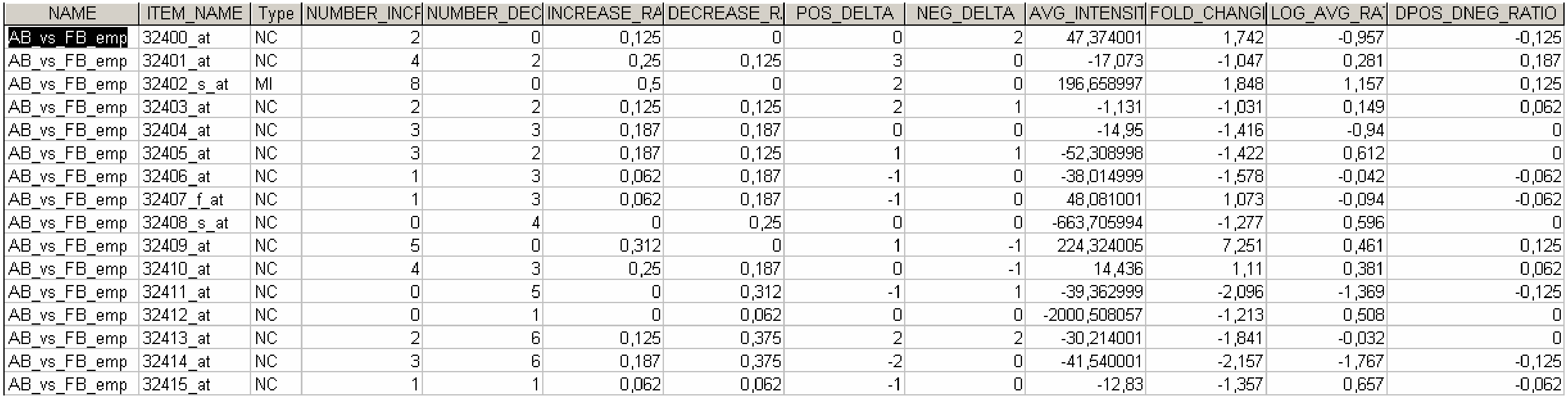
relative gene expression result

type
number increase
number decrease
increase ratio
decrease ratio
positive delta
negative delta
fold change
significance
…

biological item

= probe set

{31481_s_at, …}

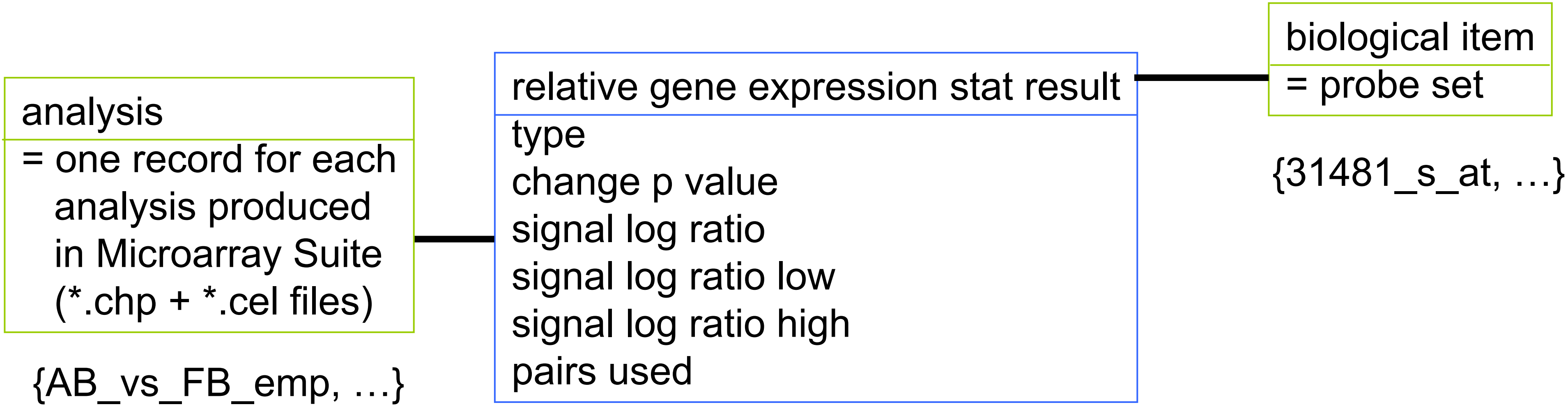


NAME	ITEM_NAME	Type	NUMBER_INCR	NUMBER_DEC	INCREASE_RA	DECREASE_R.	POS_DELTA	NEG_DELTA	AVG_INTENSIT	FOLD_CHANGI	LOG_AVG_RA	DPOS_DNEG_RATIO
AB_vs_FB_emp	32400_at	NC	2	0	0,125	0	0	2	47,374001	1,742	-0,957	-0,125
AB_vs_FB_emp	32401_at	NC	4	2	0,25	0,125	3	0	-17,073	-1,047	0,281	0,187
AB_vs_FB_emp	32402_s_at	MI	8	0	0,5	0	2	0	196,658997	1,848	1,157	0,125
AB_vs_FB_emp	32403_at	NC	2	2	0,125	0,125	2	1	-1,131	-1,031	0,149	0,062
AB_vs_FB_emp	32404_at	NC	3	3	0,187	0,187	0	0	-14,95	-1,416	-0,94	0
AB_vs_FB_emp	32405_at	NC	3	2	0,187	0,125	1	1	-52,308998	-1,422	0,612	0
AB_vs_FB_emp	32406_at	NC	1	3	0,062	0,187	-1	0	-38,014999	-1,578	-0,042	-0,062
AB_vs_FB_emp	32407_f_at	NC	1	3	0,062	0,187	-1	0	48,081001	1,073	-0,094	-0,062
AB_vs_FB_emp	32408_s_at	NC	0	4	0	0,25	0	0	-663,705994	-1,277	0,596	0
AB_vs_FB_emp	32409_at	NC	5	0	0,312	0	1	-1	224,324005	7,251	0,461	0,125
AB_vs_FB_emp	32410_at	NC	4	3	0,25	0,187	0	-1	14,436	1,11	0,381	0,062
AB_vs_FB_emp	32411_at	NC	0	5	0	0,312	-1	1	-39,362999	-2,096	-1,369	-0,125
AB_vs_FB_emp	32412_at	NC	0	1	0	0,062	0	0	-2000,508057	-1,213	0,508	0
AB_vs_FB_emp	32413_at	NC	2	6	0,125	0,375	2	2	-30,214001	-1,841	-0,032	0
AB_vs_FB_emp	32414_at	NC	3	6	0,187	0,375	-2	0	-41,540001	-2,157	-1,767	-0,125
AB_vs_FB_emp	32415_at	NC	1	1	0,062	0,062	-1	0	-12,83	-1,357	0,657	-0,062

…

AADM – Stars IV

- Relative gene expression statistical results (comparative results)



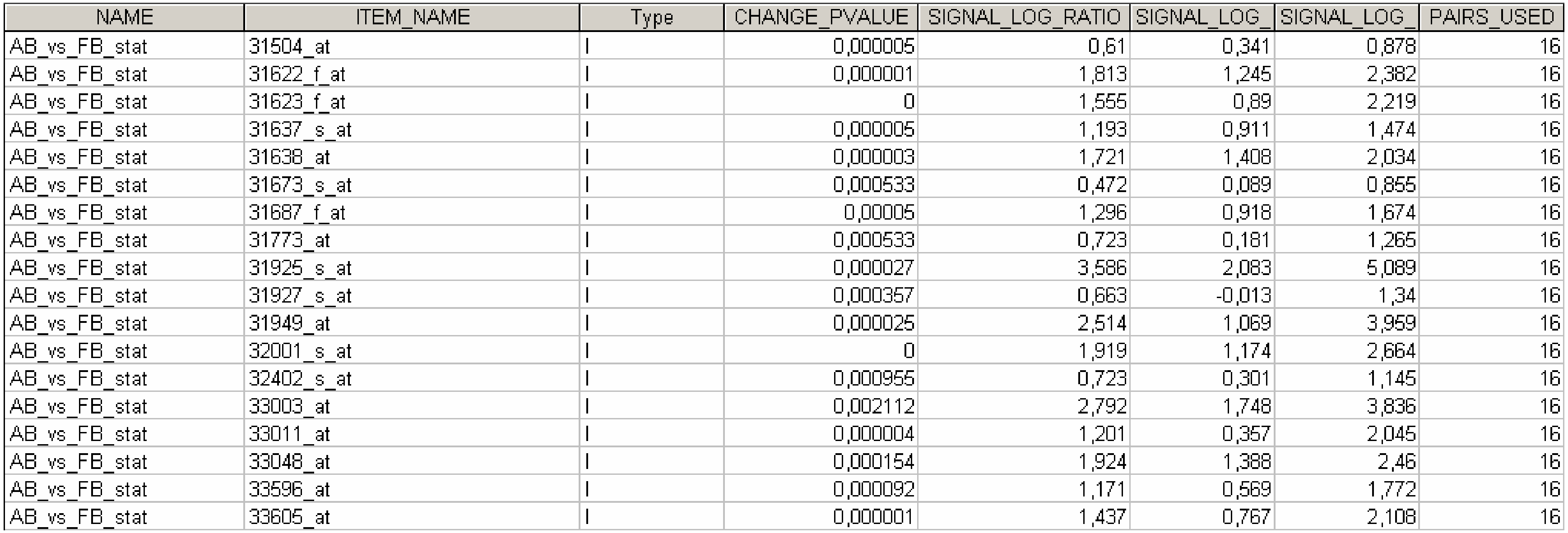
analysis
= one record for each analysis produced in Microarray Suite (*.chp + *.cel files)

{AB_vs_FB_emp, …}

relative gene expression stat result
type
change p value
signal log ratio
signal log ratio low
signal log ratio high
pairs used

biological item
= probe set

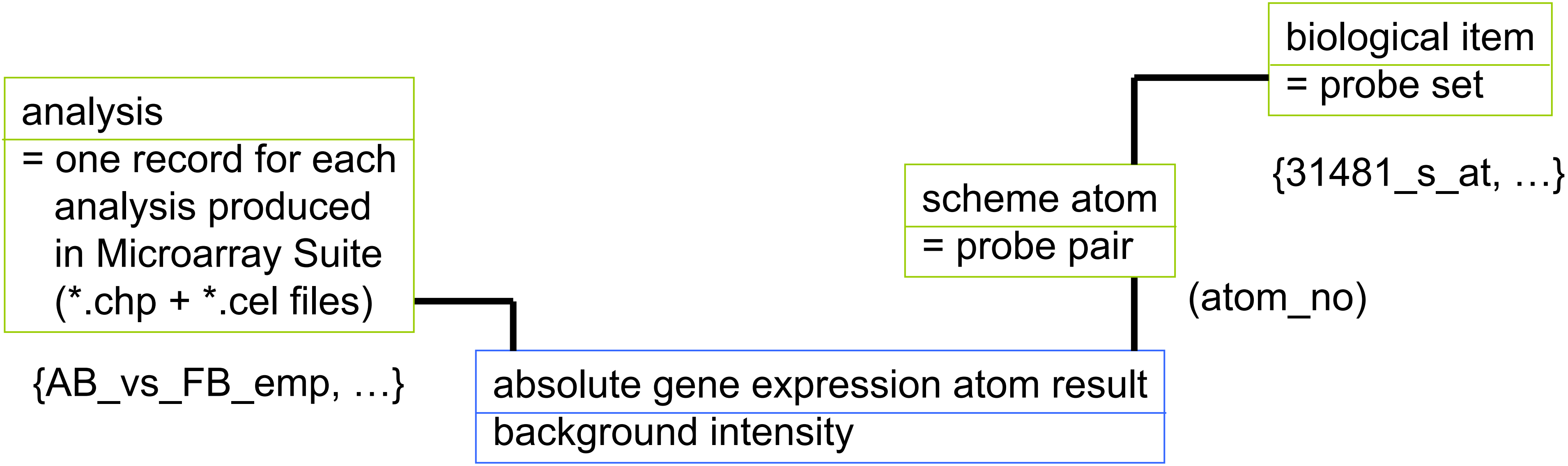
{31481_s_at, …}



NAME	ITEM_NAME	Type	CHANGE_PVALUE	SIGNAL_LOG_RATIO	SIGNAL_LOG_	SIGNAL_LOG_	PAIRS_USED
AB_vs_FB_stat	31504_at	I	0,000005	0,61	0,341	0,878	16
AB_vs_FB_stat	31622_f_at	I	0,000001	1,813	1,245	2,382	16
AB_vs_FB_stat	31623_f_at	I	0	1,555	0,89	2,219	16
AB_vs_FB_stat	31637_s_at	I	0,000005	1,193	0,911	1,474	16
AB_vs_FB_stat	31638_at	I	0,000003	1,721	1,408	2,034	16
AB_vs_FB_stat	31673_s_at	I	0,000533	0,472	0,089	0,855	16
AB_vs_FB_stat	31687_f_at	I	0,00005	1,296	0,918	1,674	16
AB_vs_FB_stat	31773_at	I	0,000533	0,723	0,181	1,265	16
AB_vs_FB_stat	31925_s_at	I	0,000027	3,586	2,083	5,089	16
AB_vs_FB_stat	31927_s_at	I	0,000357	0,663	-0,013	1,34	16
AB_vs_FB_stat	31949_at	I	0,000025	2,514	1,069	3,959	16
AB_vs_FB_stat	32001_s_at	I	0	1,919	1,174	2,664	16
AB_vs_FB_stat	32402_s_at	I	0,000955	0,723	0,301	1,145	16
AB_vs_FB_stat	33003_at	I	0,002112	2,792	1,748	3,836	16
AB_vs_FB_stat	33011_at	I	0,000004	1,201	0,357	2,045	16
AB_vs_FB_stat	33048_at	I	0,000154	1,924	1,388	2,46	16
AB_vs_FB_stat	33596_at	I	0,000092	1,171	0,569	1,772	16
AB_vs_FB_stat	33605_at	I	0,000001	1,437	0,767	2,108	16

AADM – Stars V

- Background intensities



analysis

= one record for each analysis produced in Microarray Suite (*.chp + *.cel files)

{AB_vs_FB_emp, …}

absolute gene expression atom result

background intensity

scheme atom

= probe pair

(atom_no)

biological item

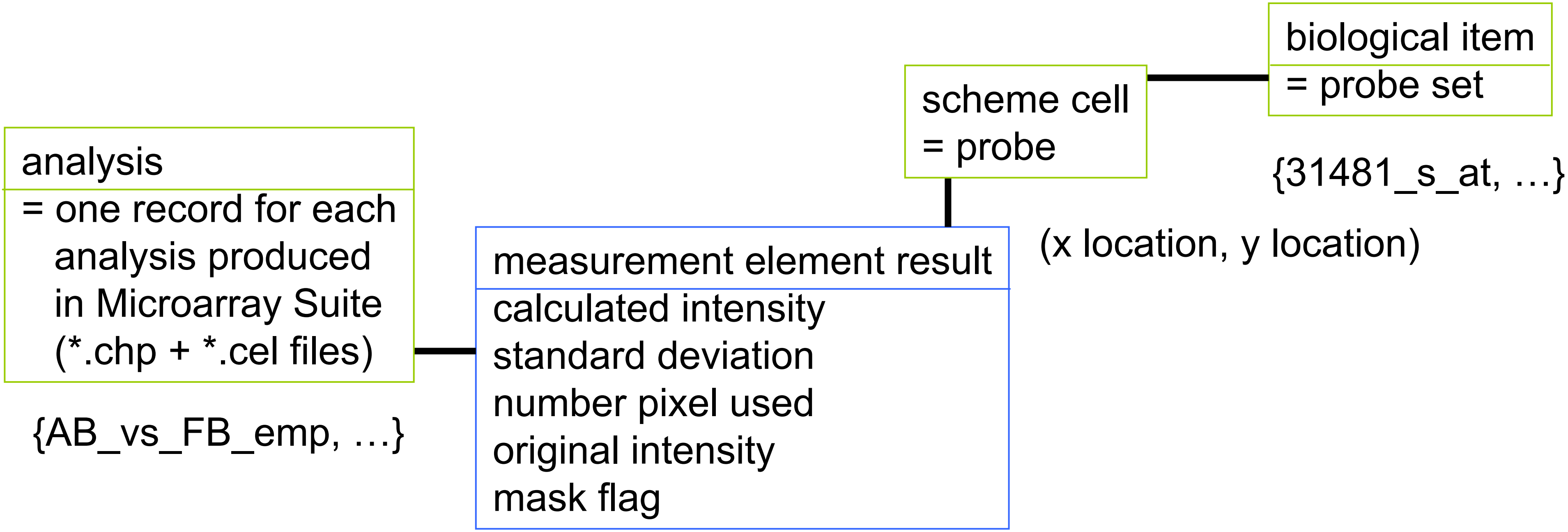
= probe set

{31481_s_at, …}

no data available

AADM – Stars VI

- Cell intensities



biological item

= probe set

{31481_s_at, …}

scheme cell

= probe

(x location, y location)

analysis

= one record for each analysis produced in Microarray Suite (*.chp + *.cel files)

{AB_vs_FB_emp, …}

measurement element result

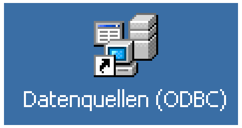
calculated intensity
standard deviation
number pixel used
original intensity
mask flag

no data available

Access to AADM

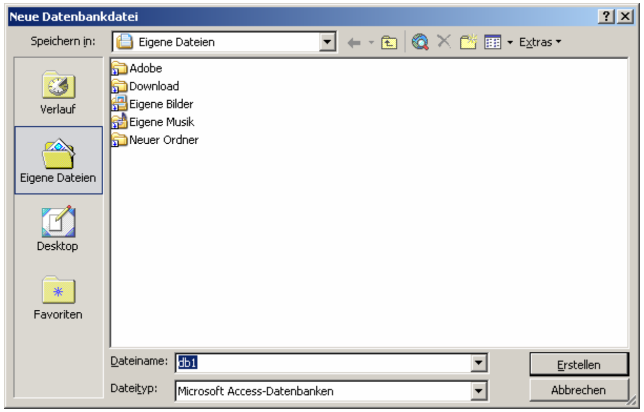
1. Create and fill a publish database
   - open MicroDB
   - select file locations
   - create a publish database
   - select specific experiments files
   - publish experiment files

2. Create a odbc connection



   - file dsn
   - database type: SQL Server
   - specify the db data file

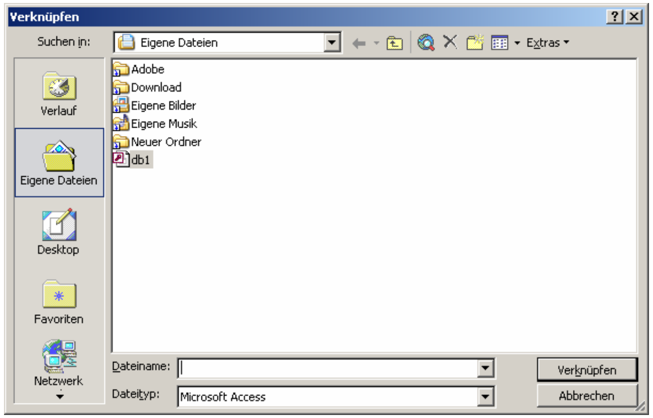
3. Create a MS Access database



Dateiname: db1

Dateityp: Microsoft Access-Datenbanken

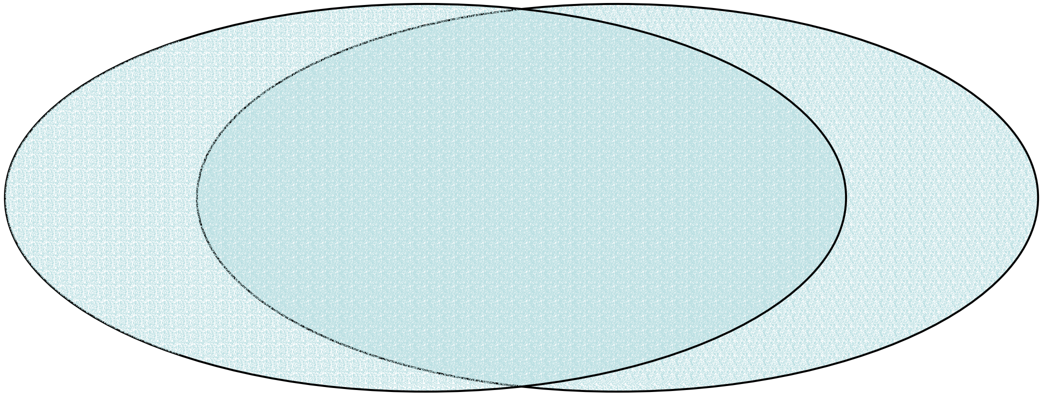
4. Link to the publish database
   - „Datei → externe Daten → verknüpfen“
   - data typ: ODBC
   - select your connection



Dateityp: Microsoft Access

5. Use
   - open table / create a view
   - login / password

Matching Affy Suite fields to AADM



Affy Suite fields

AADM fields

Calculations
e.g. inc / dec

e.g. pm_excess, mm_excess

Conclusion

- ❑ Affy' data files
  - *.cel file contains intensities, but without probe set desc.
  - proprietary *.chp file
- ❑ Affymetrix Analysis Data Model
  - structures for original and derived data
  - structures for absolute and relative (comparison) values
- ❑ MicroDB
  - stores data locally and use the AADM structures
  - stores no intensity data
  - can store max. 128 experiments