

Web service catalogue for Biomedical Grid infrastructure

Maximiliano García, Johan Karlsson and Oswaldo Trelles

University of Malaga

maxgarcia@uma.es

Towards Personalized Medicine



- (Sensible) Data Integration
- Distributed Infraestructure
- Diverse Data Processing



Laboratory



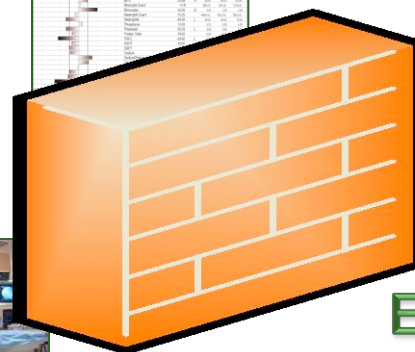
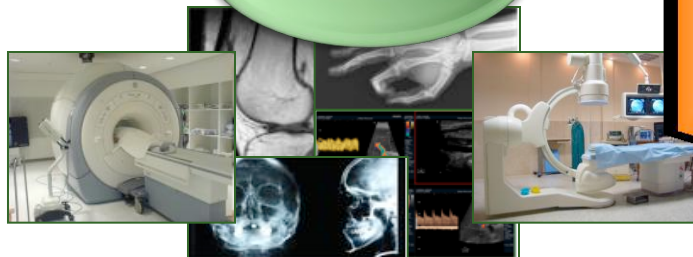
Medical
Doctor



Clinical
Records



Imaging

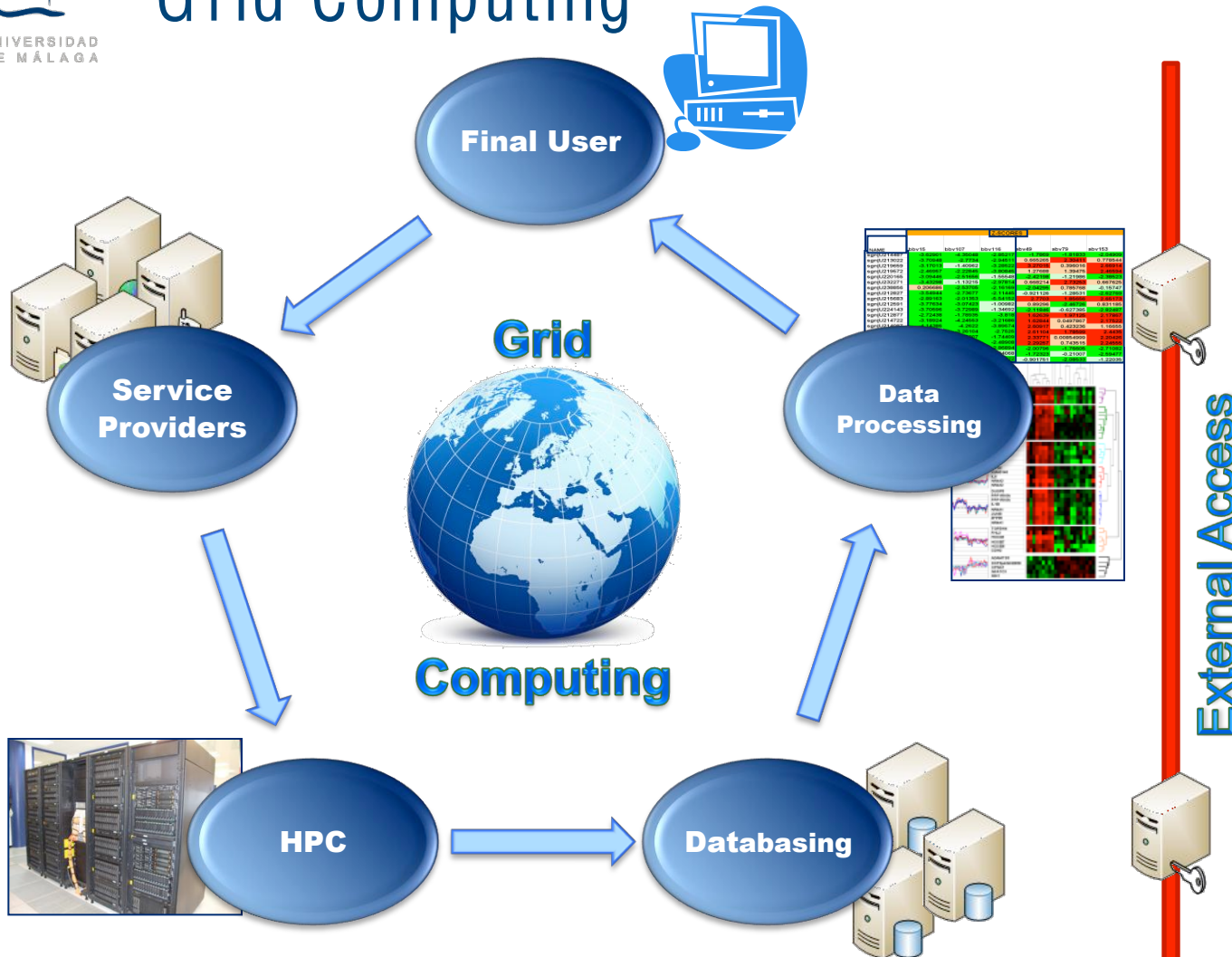


PubMed

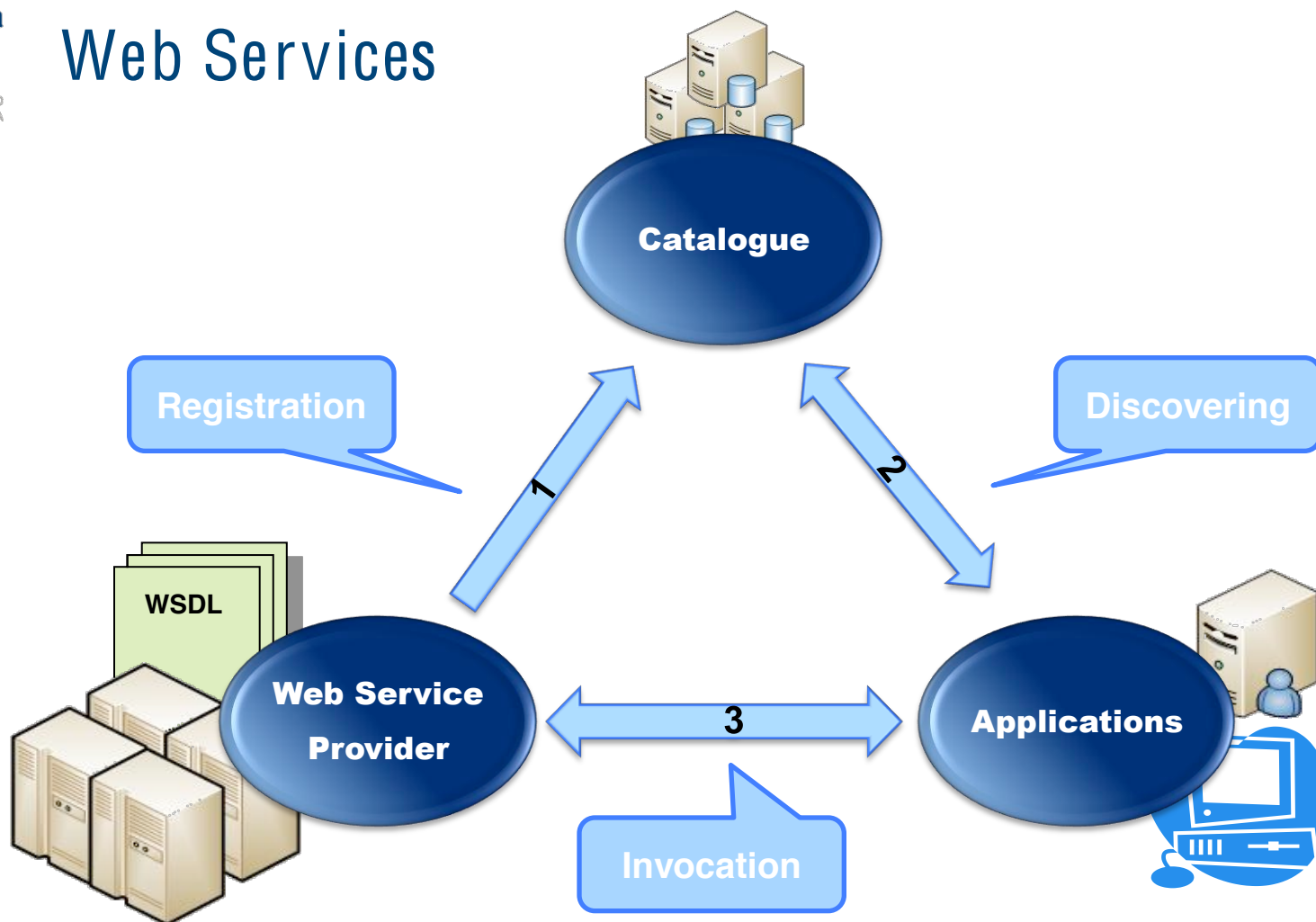
External
Resources
OMIM



Online Mendelian Inheritance in Man

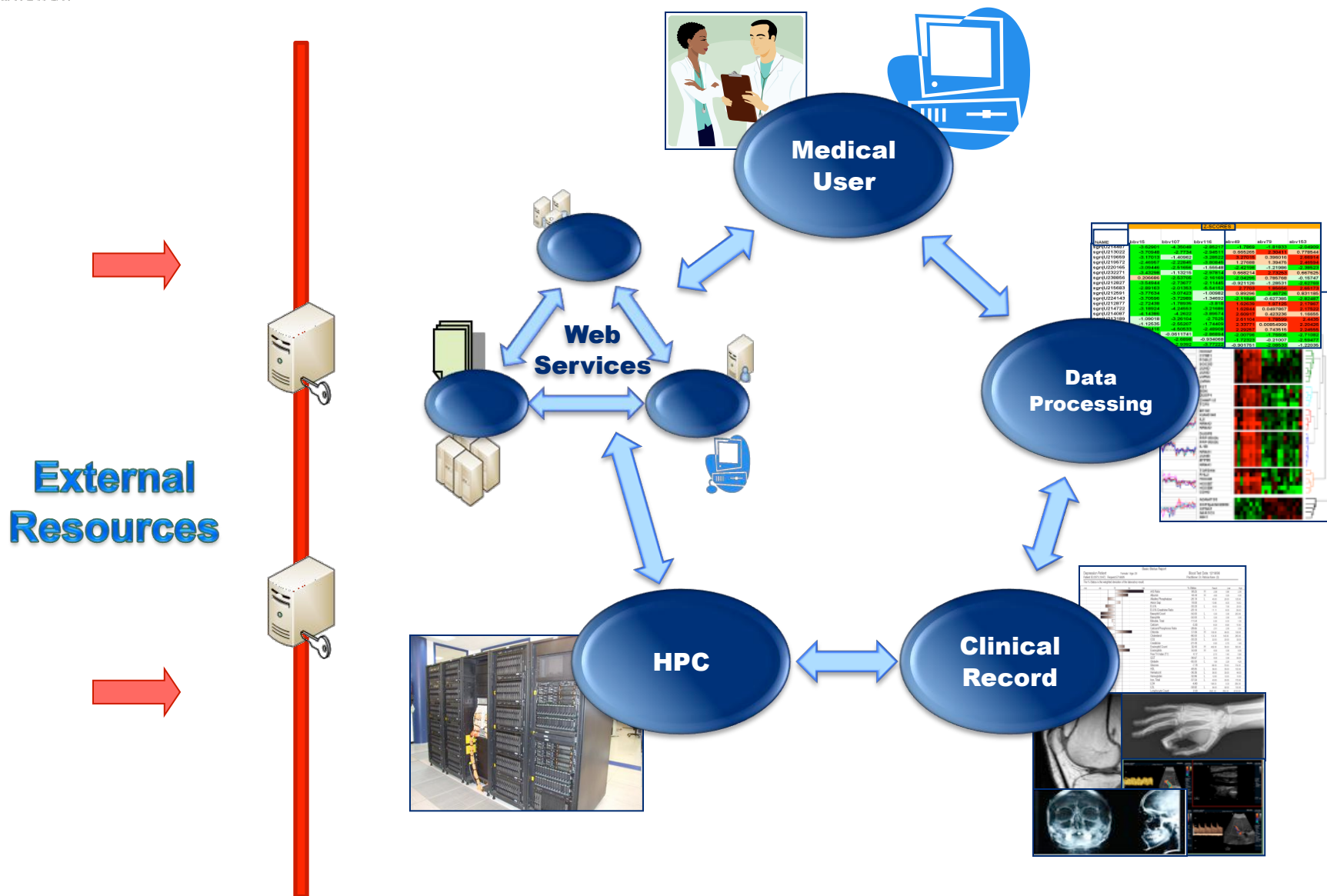


- Secure access
- Massive data management
- Distributed Computing



- Huge resources available
- Easy deployment
- Standardization and interoperability
- Discovering

Biomedicine & Grid Computing



Tool Details

Parent	Pairwise_Global_Aligning
Name	AligCons
ID	urn:eu-acgt.org:tool:aligcons
Description	Aligns a set of sequences against a consensus sequence.
Documentation	Aligns a set of sequences against a consensus sequence.
Version	1.0
WSDL	https://chirimoyo.ac.uma.es:8444/acgt/services/CLExec
Author	maxgarcia
Author e-mail	maxgarcia@ac.uma.es
Authority	BITLAB
Authority e-mail	bitlab@uma.es
Save tool Cancel Delete tool	

Available operations

Name	Actions
AligCons	Execute Edit Delete
Add operation	

Available tool locations

Host	Tool location type	Tool location status	Actions
chirimoyo.ac.uma.es	GRIDCLSERVICE	ONLINE	
Add tool location			

Parameters Definition

Tool name	AligCons
Operation name	AligCons
Operation description	runs alignment

Input parameters

Index	Parameter name	Parameter type	Parameter data type	Parameter description	
1	Contig	Simple	FASTA	Consensus Sequence	Delete
2	Common Data Set	Simple	FASTA	Fasta with reads of cons	Delete
3	InitGapPenalty	Secondary	Integer	Contribute to the overall	Delete
4	ExtendGapPenalty	Secondary	Integer	Contribute to the overall	Delete

[Add Input Parameter](#)

Output parameters

Index	Parameter name	Parameter type	Parameter data type	Parameter description	
1	Multiple Alignment	Simple	FASTA	Multiple Alignment between	Delete
2	Alignment Info	Simple	text-plain	Offsets of data alignment	Delete

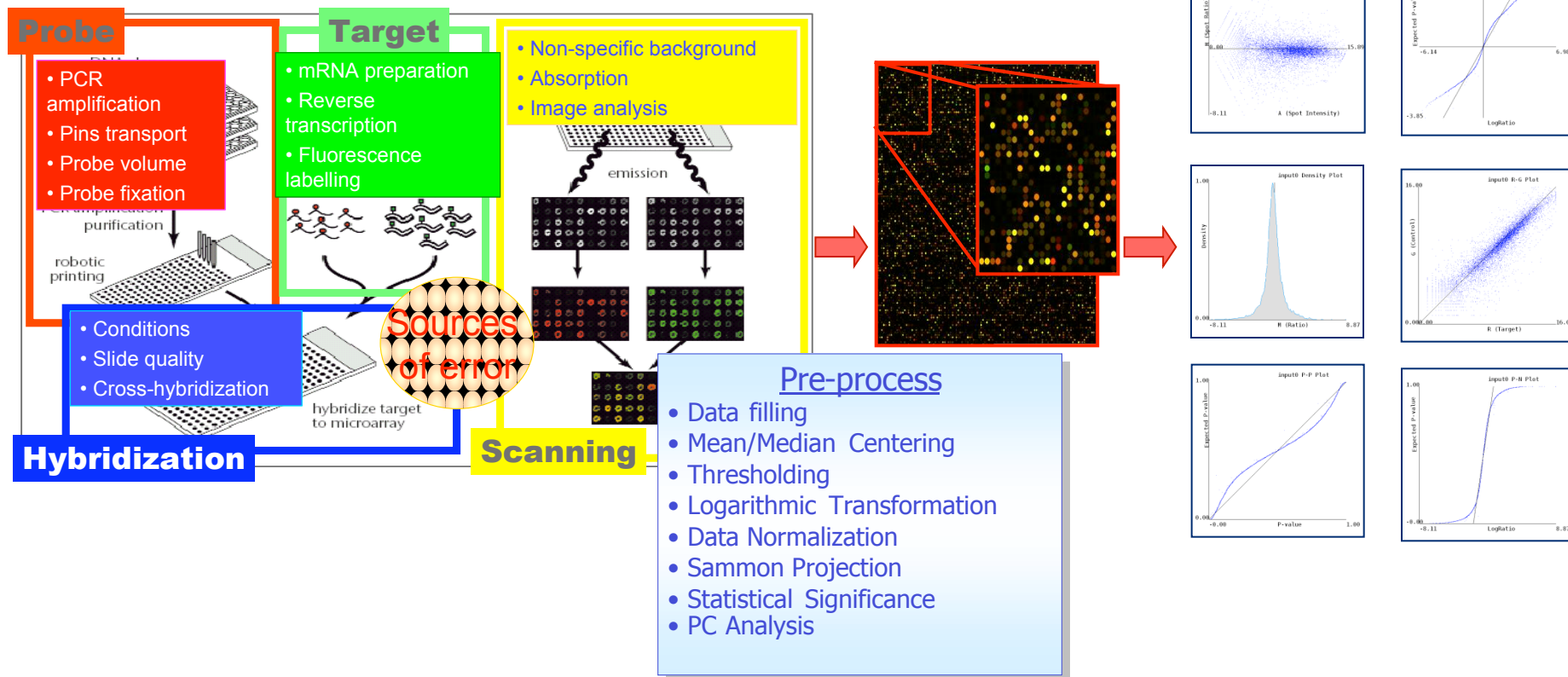
[Add Output Parameter](#)

Did you know: You can easily change the order of the parameters by clicking and dragging the index field!

Type of Service	# available	scope
BioMOBY services	31	General biomedical and bioinformatics Web services.
CLP	30	CLP for Gene Expression data.
Mediator queries	15	Predefined queries on Ontology search.
R-Scripts	50	Statistical treatment of data and data mining methods.

- Metadata for Service Discovering: Interconnection-Standardization, invocation, etc.

- Data quality services
- Graphical representations (AM, QQ, Density, RG, PP, PN, ...)
- Formats supported: One channel and dual channel



- Analysis tools to visualizing, clustering and post-processing expression profiles

Clustering

- K Means
- Fuzzy K Means
- Double Threshold
- Hierarchical
- Self Organizing Map (SOM)
- Batch SOM
- Fuzzy SOM
- KerDenSOM
- Fuzzy Kohonen Clustering

Distances

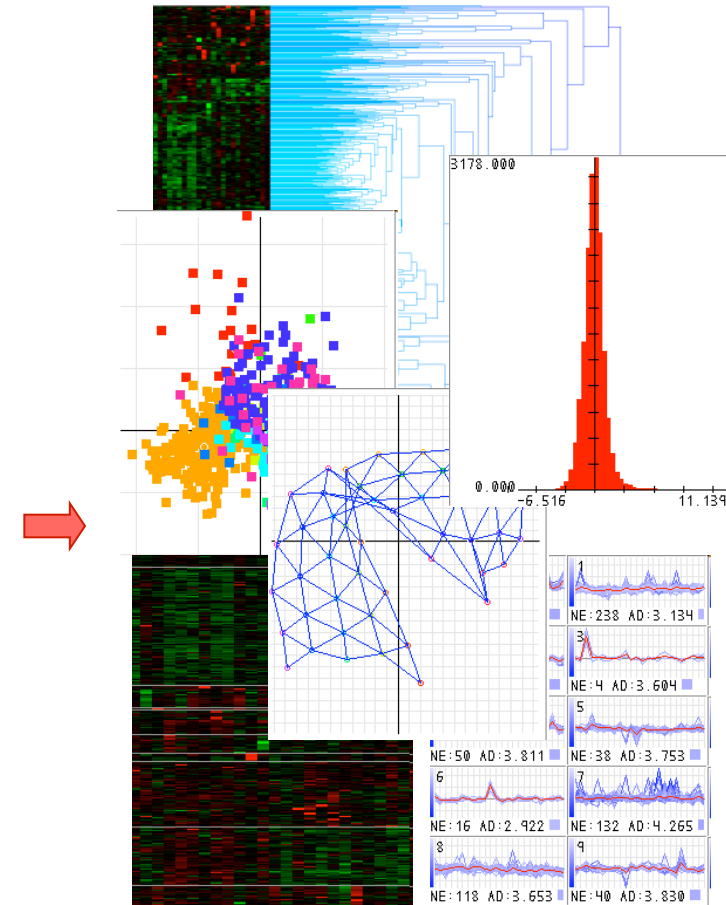
- Euclidean
- Correlation
- Absolute Correlation
- Manhattan
- Weighted

Visualization

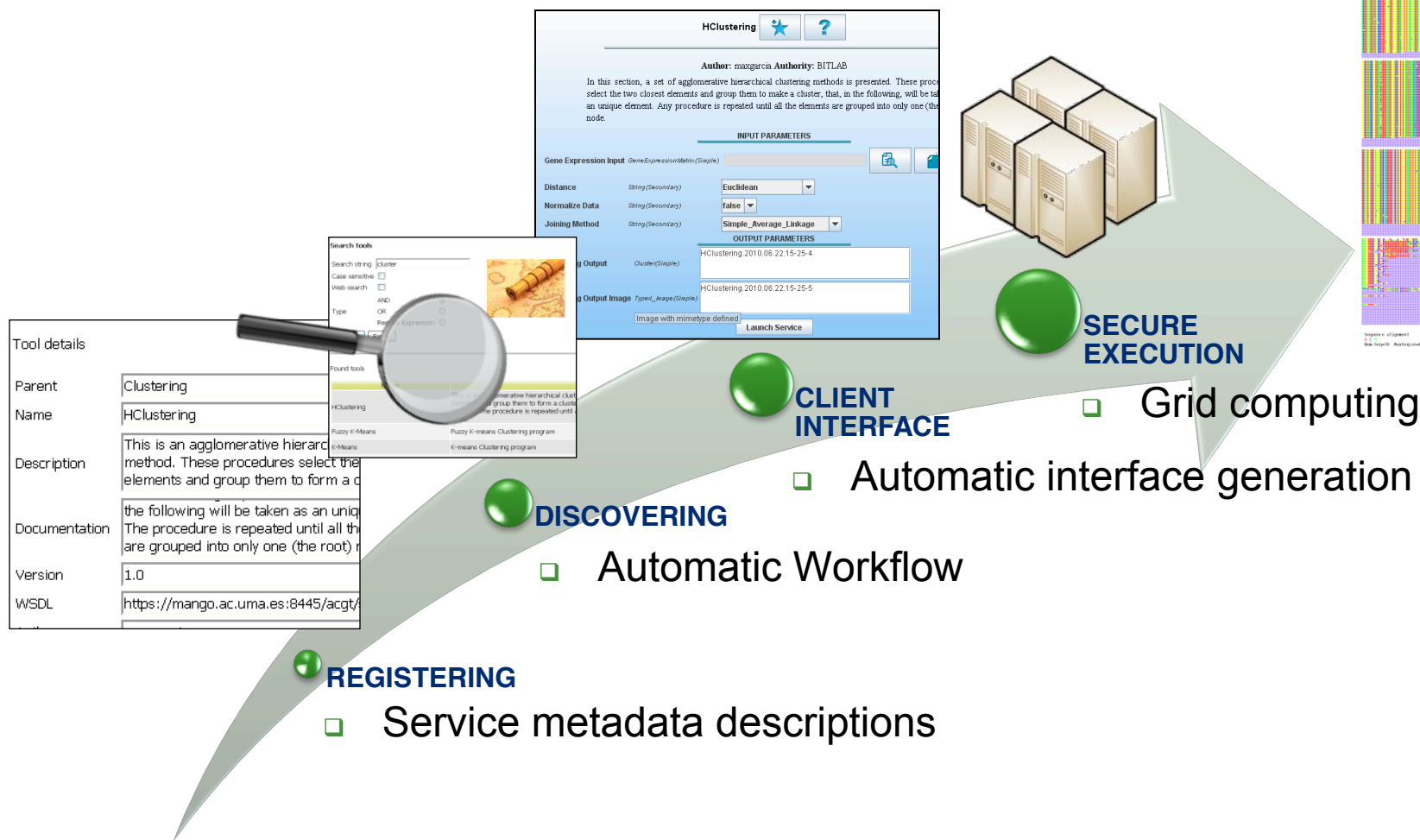
- Heat Map
- Cluster profiles
- Projection Maps
- Cluster Silhouettes
- Hierarchical Tree Vis.
- U-Matrix
- Histograms

Post-processing

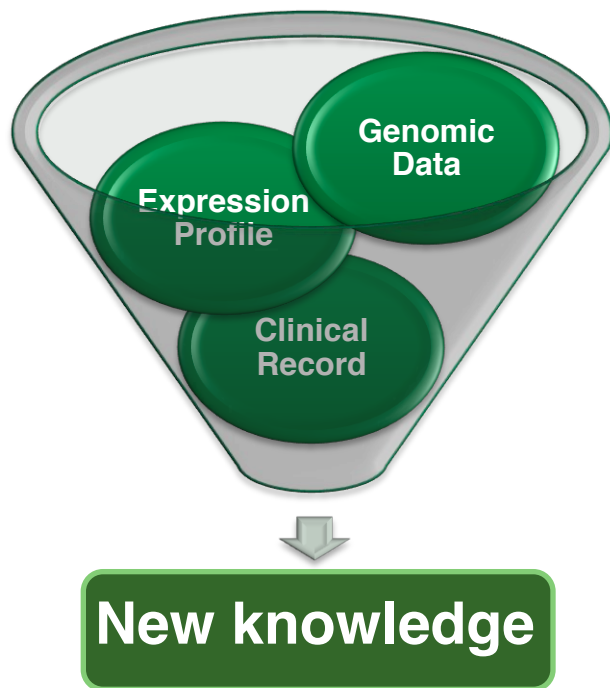
- Silhouette value optimisation
- KDD, association rules



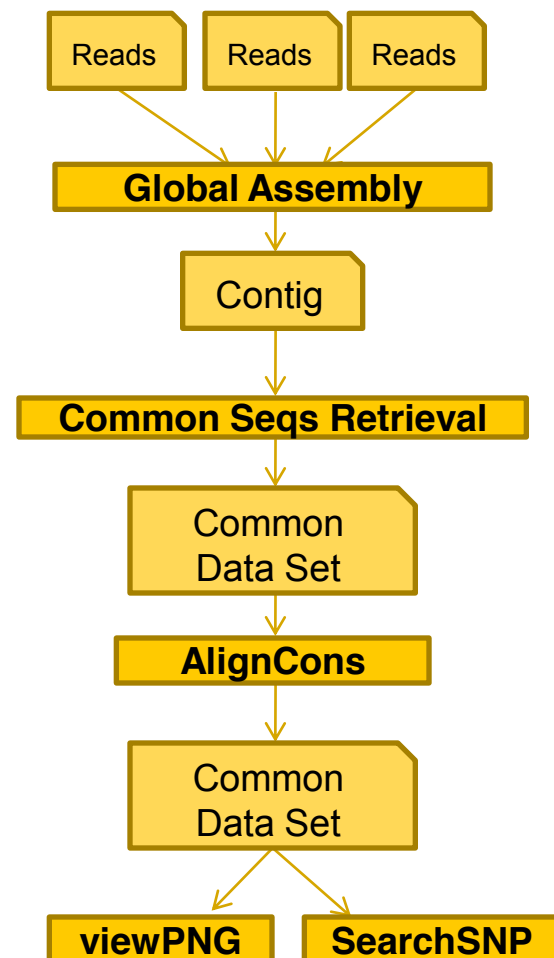
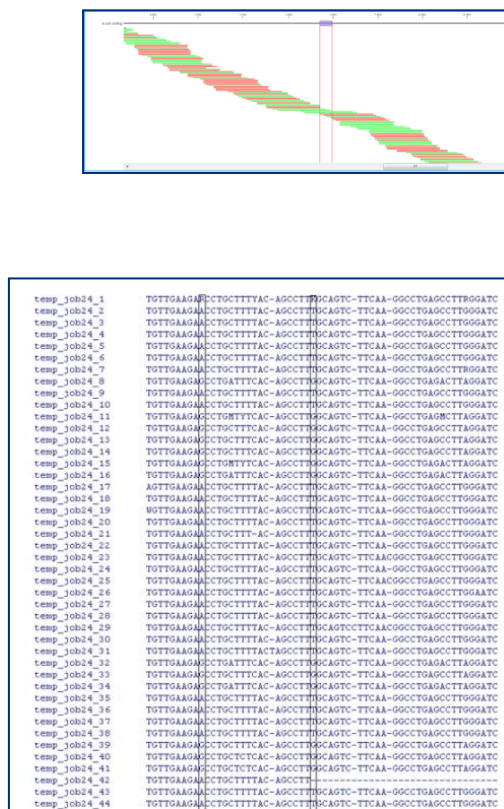
Life cycle of Services



Full experiment



SNPs Search



SNP Search - jORCA client interface



GlobalAssembly ★ ?

Author: maxgarcia Authority: BITLAB

runs assembler

INPUT PARAMETERS

Reads collection *FASTA(Array)* D:\Data\Reads_Collection.fasta

OUTPUT PARAMETERS

Contig *FASTA(Simple)* Contig.FASTA

Contig Info *text-plain(Simple)* Contig.INFO

Launch Service



Common sequences Retrieval ★ ?

Author: maxgarcia Authority: BITLAB

runs retrieval

INPUT PARAMETERS

Reads collection *FASTA(Array)* D:\Data\Reads_Collection.fasta

Contig *FASTA(Simple)* D:\Data\Contig.FASTA

Contig Info *text-plain(Simple)* D:\Data\Contig.INFO

OUTPUT PARAMETERS

Data_Set.FASTA

Launch Service

Enter your grid credentials

Grid User Name: maxgarcia

Grid User Password: ●●●●●●

Grid Proxy Host: moss1.man.poznan

Grid Proxy Port: 7512

OK Cancel



AlignCons ★ ?

Author: maxgarcia Authority: BITLAB

runs alignment

INPUT PARAMETERS

Contig *FASTA(Simple)* D:\Data\Contig.FASTA

Common Data Set *FASTA(Simple)* D:\Data\Data_Set.FASTA

InitGapPenalty *Integer(Secondary)* 2

ExtendGapPenalty *Integer(Secondary)* 0

OUTPUT PARAMETERS

Multiple Alignment *FASTA(Simple)* Alignment.FASTA

Alignment Info *text-plain(Simple)* Alignment.INFO

Launch Service

SearchSNP ★ ?

Author: maxgarcia Authority: BITLAB

INPUT PARAMETERS

Multiple Alignment *FASTA(Simple)* D:\Data\Alignment.FASTA

Alignment Info *text-plain(Simple)* D:\Data\Alignment.INFO

PCol *Integer(Secondary)* 50

PDif *Integer(Secondary)* 10

OUTPUT PARAMETERS

SNPs *text-plain(Simple)* SNPs.INFO

Image *Image(Simple)* SNPs.PNG

Launch Service

- Uniform & automatic interfaces
- I-O Interconnection

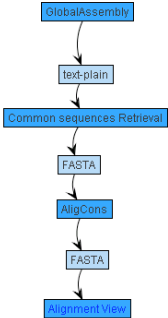
- User friendly clients
- Secure environment

Executed tasks

Executed tasks in JORCA

Tool	Date	Where
Tool: CreateAMPlot	Date: 11 Jun 2010 09:59:01 GMT	Where: chirimoyo.ac.uma.es
Repository: ACGT Stable Repository chirimoyo.ac.uma.es		
Parameters: [null,D:\2010\FileSystem\250PRRead]		
Status: FINISHED		
Tool: GlobalAssembly	Date: 23 Jun 2010 09:14:38 GMT	Where: chirimoyo.ac.uma.es
Repository: ACGT Stable Repository chirimoyo.ac.uma.es		
Parameters: [null,D:\2010\FileSystem\FASTA_Text\2010.05.31.16-35]		
Status: FINISHED		
Tool: Common sequences Retrieval	Date: 23 Jun 2010 09:16:19 GMT	Where: chirimoyo.ac.uma.es
Repository: ACGT Stable Repository chirimoyo.ac.uma.es		
Parameters: [null,D:\2010\FileSystem\FASTA_Text\2010.05.31.16-35],[null,D:\2010\FileSystem\FASTA_Text\2010.05.31.16-35],[null,D:\2010\FileSystem\formatAffy.txt]		
Status: FINISHED		
Tool: AlignCons	Date: 23 Jun 2010 09:20:56 GMT	Where: chirimoyo.ac.uma.es
Repository: ACGT Stable Repository chirimoyo.ac.uma.es		
Parameters: [null,D:\2010\FileSystem\FASTA_Text\2010.05.31.16-35],[null,D:\2010\FileSystem\FASTA_Text\2010.05.31.16-35],[null,D:\2010\FileSystem\FASTA_Text\2010.05.31.16-35],[null,D:\2010\FileSystem\FASTA_Text\2010.05.31.16-35]		
Status: FINISHED		
Tool: Alignment View	Date: 23 Jun 2010 09:29:59 GMT	Where: chirimoyo.ac.uma.es
Repository: ACGT Stable Repository chirimoyo.ac.uma.es		
Parameters: [null,D:\2010\FileSystem\FASTA_Text\2010.05.31.16-35],[null,D:\2010\FileSystem\formatAffy.txt],[null,200],[null,4]		
Status: FINISHED		
Tool: SearchSNP	Date: 23 Jun 2010 09:24:24 GMT	Where: chirimoyo.ac.uma.es
Repository: ACGT Stable Repository chirimoyo.ac.uma.es		
Parameters: [null,D:\2010\FileSystem\FASTA_Text\2010.05.31.16-35]		

Drag a tool from the green icon to this Canvas (Tools must match I/O)



```

graph TD
    GlobalAssembly --> text-plain
    text-plain --> Common_sequences_Retrieval[Common sequences Retrieval]
    Common_sequences_Retrieval --> FASTA1[FASTA]
    FASTA1 --> AlignCons
    AlignCons --> FASTA2[FASTA]
    FASTA2 --> Alignment_View[Alignment View]
  
```

- Workflow: Service composition for automatic enactment

My ACGT My Files Metadata Registration Query Tool **Workflow Editor** GridR Session Data Access Ontology My Settings How to

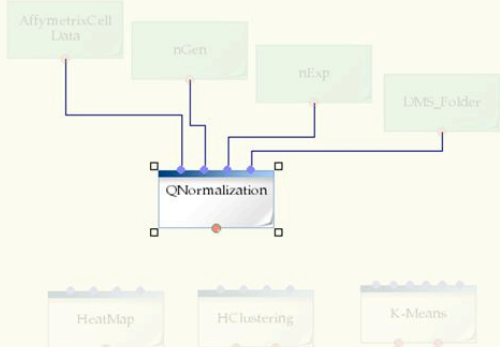
File Edit Tools Enactments Help

Welcome "Jane Doe" !

Services Biomoby My Files

- HCustering
 - K-Means
 - K-Means
- GE Matrix PreProc
 - HeatMap
- PreProcessing
 - Transpose
- Projection Method
- Statistical Analysis
- Microarray PreProc
 - GE Data Normaliz
 - MaxTest1
 - MaxTest2
 - MaxTest4
 - MicroarrayDat
 - MicroarrayDat
 - MicroarrayDat
 - MicroarrayDat
 - MicroarrayDat
 - MicroarrayDat
 - MicroarrayDat
 - QNormalization

Workflow Area (MaxWTest5)



```

graph TD
    AffymetrixCell_Data[AffymetrixCell Data] --> QNormalization
    nGen --> QNormalization
    nExp --> QNormalization
    DMS_Folder[DMS_Folder] --> QNormalization
    QNormalization --> HeatMap
    QNormalization --> HCustering
    QNormalization --> K-Means
  
```

Properties Panel

QNormalization

Inputs :

- AffymetrixCellData : List of text-plain
- nGen : Integer
- nExp : Integer
- DMS_Folder : String

Outputs :

- Gene Expression Output : GeneExpressionMatrix

Description :

QNormalization process.

SNP SEARCH – Workflow enactment



Pipeline execution  

☒ Save intermediate results

GlobalAssembly parameters : _____

Reads collection *FASTA(Array)* 10FileSystem\FASTA.2010.06.03.10-27   

Common sequences Retrieval parameters :  _____

AlignCons parameters :  _____

Alignment View parameters :  _____

Multiple Alignment Sequences *FASTA(Simple)* **This input will be the output of the above tool**

Offsets *text-plain(Simple)* 10FileSystem\FASTA.2010.06.23.10-59   

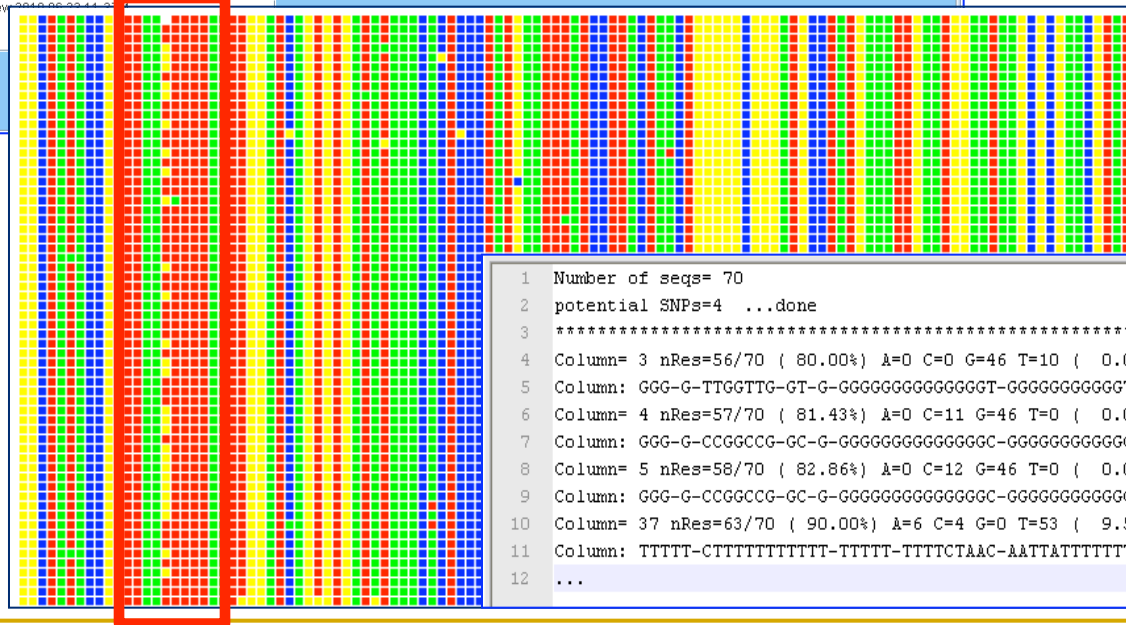
NElemXrow *Integer(Secondary)*

ElemSize *Integer(Secondary)*

Output parameters : _____

Image *Image(Simple)* Alignment View 2010.06.23.11-3

Launch pipeline



```

1 Number of seqs= 70
2 potential SNPs=4 ...done
3 *****
4 Column= 3 nRes=56/70 ( 80.00%) A=0 C=0 G=46 T=10 ( 0.00% 0.00% 82.14% 17.86%)
5 Column: GGG-G-TTGGTTG-GT-G-GGGGGGGGGGGGGT-GGGGGGGGGGGTGGGGG---GGGGG---TGTG-T-
6 Column= 4 nRes=57/70 ( 81.43%) A=0 C=11 G=46 T=0 ( 0.00% 19.30% 80.70% 0.00%)
7 Column: GGG-G-CCGGCCG-GC-G-GGGGGGGGGGGGGC-GGGGGGGGGGGCGGGGG---CGGGG---CGCG-C-
8 Column= 5 nRes=58/70 ( 82.86%) A=0 C=12 G=46 T=0 ( 0.00% 20.69% 79.31% 0.00%)
9 Column: GGG-G-CCGGCCG-GC-G-GGGGGGGGGGGGGC-GGGGGGGGGGGCGGGGG---CGGGG---CCGCG-C-
10 Column= 37 nRes=63/70 ( 90.00%) A=6 C=4 G=0 T=53 ( 9.52% 6.35% 0.00% 84.13%)
11 Column: TTTT-CTTTTTTTTTTT-TTTT-TTTTCTAAC-AATTATTTTTTTTTTT-TATTTT--TTTTTTC
12 ...

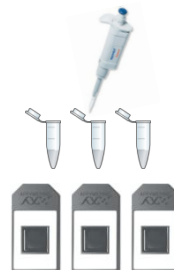
```


SNP SEARCH – Association rules



ID	Patient1	Patient2
200000_s_at	317.8	217
200001_at	678	726
200002_at	1647.5	1115
200003_s_at	3981	2297
200004_at	982.3	1027
200005_at	461	382.8
200006_at	1222	1298
200007_at	951	1394.8
200008_s_at	387	361
200009_at	1246.5	1163.8
200010_at	3175	1553.3
200011_s_at	391	825
200012_x_at	3473	2337
200013_at	2522	1897.8
200014_s_at	475	428
200015_s_at	808.5	611.3
200016_x_at	1868	1262
200017_at	1682	1322
200018_at	5454.5	3058
200019_s_at	2435.8	1565
200020_at	314	396
...	...	...

Gene expr.
+ metadata



Microarrays

Bg Correction

GEDatanormalization

CreateGEMatrix

GetAnnotationsFromDBs

AssignMetadata

ExtractAssociationRules

Gender	Male	Male	Female	Male	Female	Female	Male	Female	Female	Female	Male
Age	49	57	52	61	54	52	71	66	52	74	
Family History	Cancer	Cancer	Cancer	Unknown	Cancer	Cancer	Unknown	Cancer	Cancer	Unknown	Cancer
Location	Renal Medulla	Cortex	Nephron	Nephron	Renal Medulla	Renal Medulla	Cortex	Renal Medulla	Nephron	Cortex	Renal M
Severity	2	1	1	3	2	2	3	4	1	3	
Variant SNP-2348	A	A	T	T	A	A	T	A	A	T	A
Variant SNP-3562	C	C	G	C	A	C	G	C	C	G	C
Variant SNP-3891	T	T	T	T	T	T	C	C	T	T	C
Variant SNP-4113	G	G	A	A	A	G	A	A	G	A	G

GO terms	InterPro	IDs	Pat.1	Pat.2	Pat.3	Pat.4	Pat.5	Pat.6	Pat.7	Pat.8	Pat.9	Pat.10	Pat.11
GO:0003824,IPR008147,IPR008146		222666_s_at	-0.321806	0.216248	0.536375	-0.356016	-0.0894655	0.00102498	0.208238	-0.0714255	-0.451755	-0.26292	0.4
GO:0004356,IPR006082,IPR006083		222812_s_at	0.202284	-0.278358	-0.417992	0.118334	-2.27961	-1.39733	-0.464161	-0.100261	-0.215401	0.191589	0.3
GO:0005524,IPR000073,IPR008262		224228_s_at	-0.429797	-0.0130022	1.02271	0.222158	0.221091	-0.290343	-0.605924	-0.0721774	0.328234	-0.160366	-0.1
GO:0016829,IPR001344		225957_at	1.14764	0.552223	0.260031	-0.209298	0.257798	0.175835	-0.0443977	0.871135	0.291432	0.359633	1
IPR001547,IPR013781		224861_at	-0.222963	-0.109308	-0.492854	-0.489262	0.451481	0.313038	-0.455047	-0.538378	-0.33425	-0.462914	-0.4
GO:0043169,IPR000741,IPR013785		224004_at	-0.5284	-1.73712	0.785261	0.0574439	-0.328043	-0.0991853	0.930249	0.378547	0.872716	-0.0770358	0.5
GO:0004332,IPR000894		224707_at	0.0710741	-2.15892	2.47439	1.2962	1.12293	1.91462	2.57933	2.86568	2.39217	0.5276	0.1
GO:0009573,IPR005478,IPR005476,IPR005475		224638_at	-2.12601	0.0941617	0.367413	-0.483302	0.377169	0.832026	0.779874	0.381155	-0.319561	-0.312835	-0.3
GO:0009535,IPR001344,IPR010999		222853_at	-0.22092	0.0319082	0.265628	0.0856191	-0.243858	-0.156252	-0.00742786	-0.18464	-0.293908	-0.4034	0.4
IPR006139,IPR006140		223687_s_at	-0.156055	-0.141488	-0.770779	0.036068	-0.137053	-0.00810592	-0.605514	0.0355932	-0.315148	-0.0872605	-0.1
GO:0008465,IPR000173,IPR006424		224253_at	-2.07067	-0.0975094	-1.13136	-0.312433	1.74596	1.3998	-0.0513603	-1.29803	-0.221016	-0.503389	-0
GO:0047100,IPR000131		222448_s_at	-0.972248	-0.175185	0.0374878	-0.221862	-1.28975	-0.0984862	-0.123132	-0.969214	-0.107961	-0.664729	-0
IPR001128,IPR002403		226145_s_at	0.363917	-0.0110114	-0.0463903	-0.410662	-0.944265	-0.343608	-0.534332	-0.314075	-0.15549	0.0419386	0.4

ID	Pat.1	Pat.2	Pat.3	Pat.4	Pat.5	Pat.6	Pat.7	Pat.8
Gender	Male	Male	Female	Male	Female	Female	Male	Female
Age	49	57	52	61	54	52	71	66
Family History	Cancer	Cancer	Cancer	Unknown	Cancer	Cancer	Unknown	Cancer
Location	Renal Medulla	Cortex	Nephron	Nephron	Renal Medulla	Renal Medulla	Cortex	Renal
Severity	2	1	1	3	2	2	3	4
Variant SNP-2348	A	A	T	T	A	A	T	A
Variant SNP-3562	C	C	G	C	A	C	G	C
Variant SNP-3891	T	T	T	T	T	T	C	C
Variant SNP-4113	G	G	A	A	A	G	A	A

GO terms	InterPro	IDs	Pat.1	Pat.2	Pat.3	Pat.4	Pat.5	Pat.6	Pat.7	Pat.8
GO:0003824,IPR008147,IPR008146		222666_s_at	-0.322	0.216	0.536	-0.356	-0.089	0.001	0.208	-0.071
GO:0004356,IPR006082,IPR006083		222812_s_at	0.202	-0.278	-0.418	0.118	-2.28	-1.397	-0.464	-0.1
GO:0005524,IPR000073,IPR008262		224228_s_at	-0.43	-0.013	1.023	-0.222	0.221	-0.29	-0.606	-0.072
GO:0016829,IPR001344		225957_at	1.148	0.552	0.26	-0.209	0.258	0.176	-0.044	0.871
IPR001547,IPR013781		224861_at	-0.223	-0.109	-0.493	-0.489	0.451	0.313	-0.455	-0.536

Association
rules

GO:0007049+Overexpression→Cancer

Variant SNP-4113 [A]+Female→High Severity

Conclusions

- Biomedical applications features:
 - Huge and diverse data sets
 - Distributed computational resources
 - Sensible data
- Deploy and use of secure Grid services in Biomedicine
- Command line programs as traditional grid components
- Profusion of Web Services
 - Standardization
 - Interconnectivity
 - Discovering
- Use of Web services catalogue in the Grid.
- Use case
 - Fast deploy of about 30 CLP
 - Clinical data + SNP + gene expression → Association rules