KATHOLIEKE UNIVERSITEIT LEUVEN

Data mining in molecular karyotyping: linked analysis of Array-CGH data and biomedical text

Public PhD defense
September 28th, 2009

Steven Van Vooren

Department of Electrical Engineering
Katholieke Universiteit Leuven

Examination board:

prof. dr. ir. E. Aernoudt (chairman)
prof. dr. ir. B. De Moor (promotor)
prof. dr. ir. Y. Moreau (promotor)
prof. dr. ir. J. Vermeesch
prof. dr. ir. M. Bruynooghe
prof. dr. ir. S. Vanhuffel
prof. dr. K. Devriendt
prof. dr. F. Speleman (UGent)
Dr. D. Rebholz-Schuhmann (EBI)

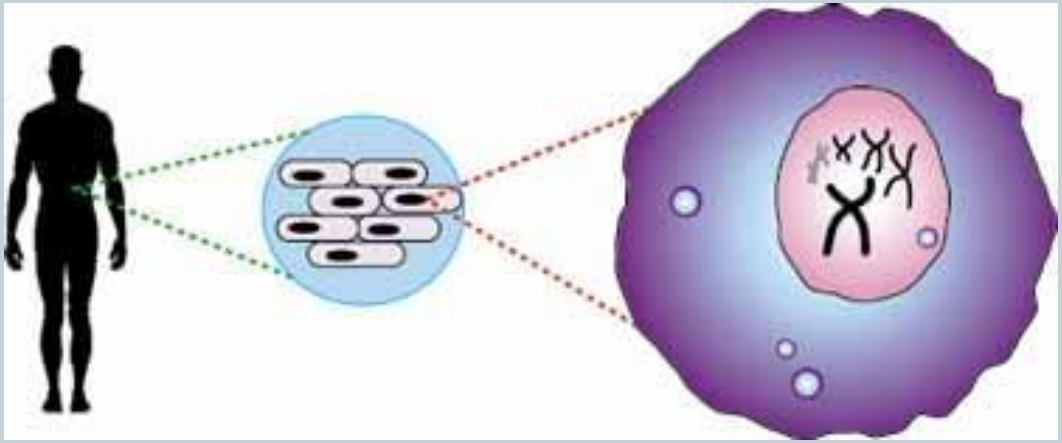
Overview of presentation

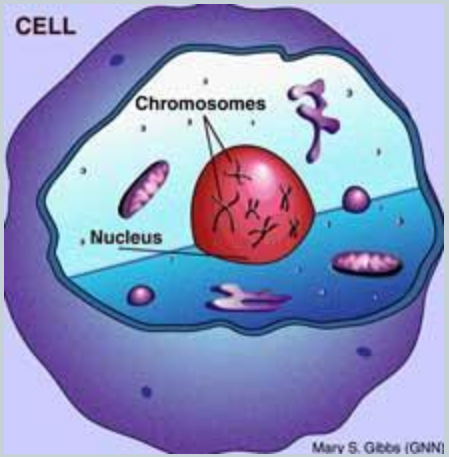
- **Introduction** – Genome variation
- **Chapter 1** – Introduction and context
- **Chapter 2** – ArrayCGH data management and analysis
- **Chapter 3** – Functional interpretation of genetic information
- **Chapter 4** – Intelligent databases for constitutional cytogenetics
- **Chapter 5** – Genome annotation
- **Chapter 6** – Data fusion on biomedical text & gene expression arrays
- **Chapter 7** – Conclusions and perspectives

Introduction to Genome Variation

AN OVERVIEW

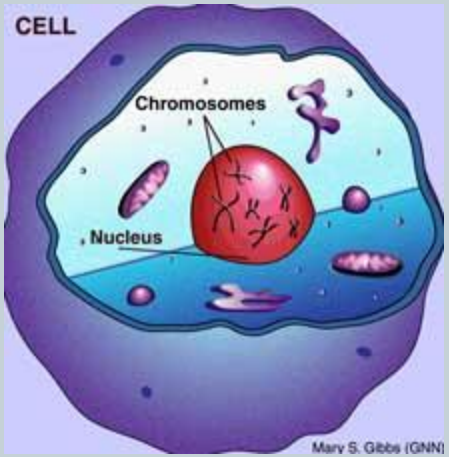
People – tissues – cells



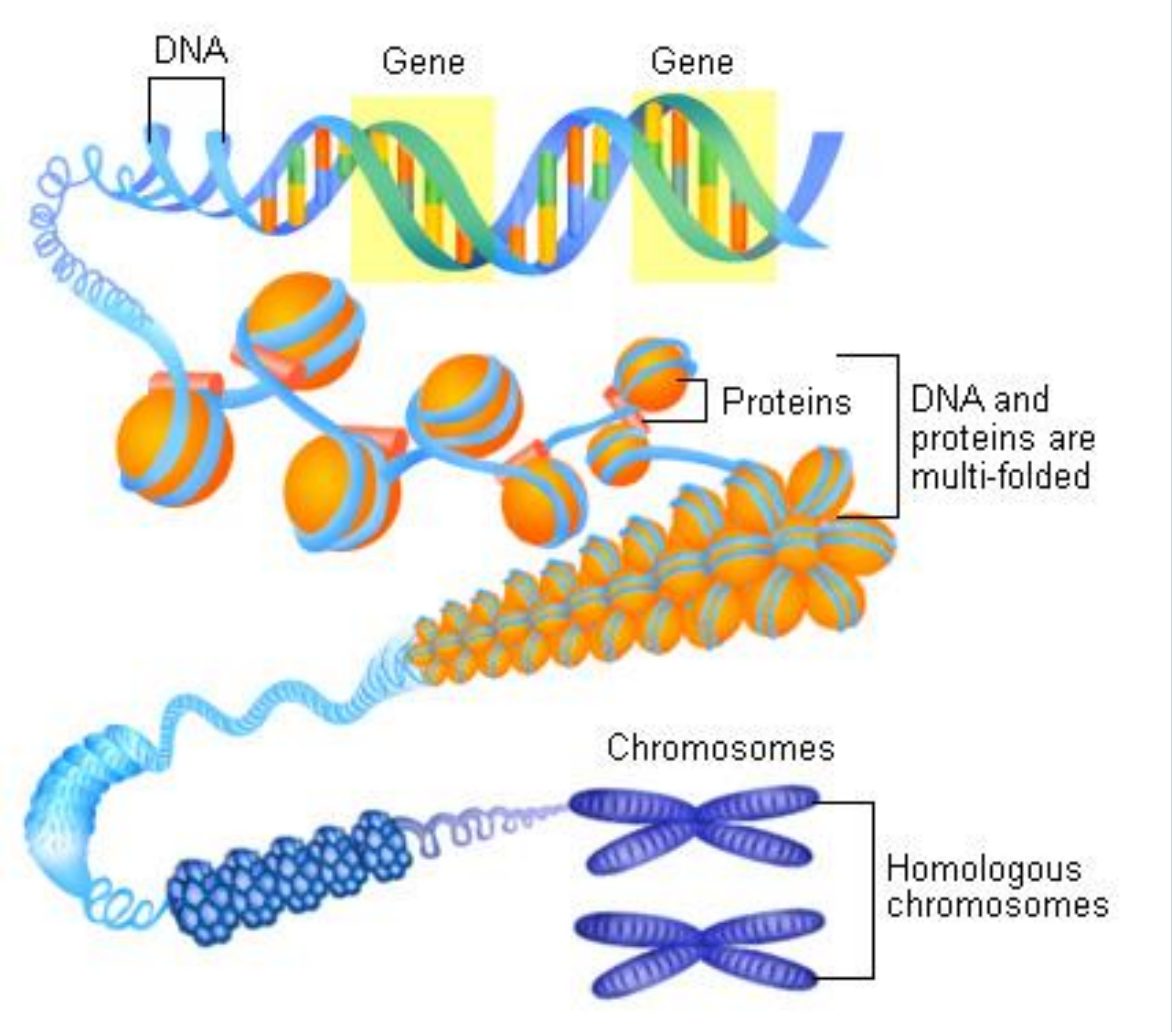


People – tissues - cells

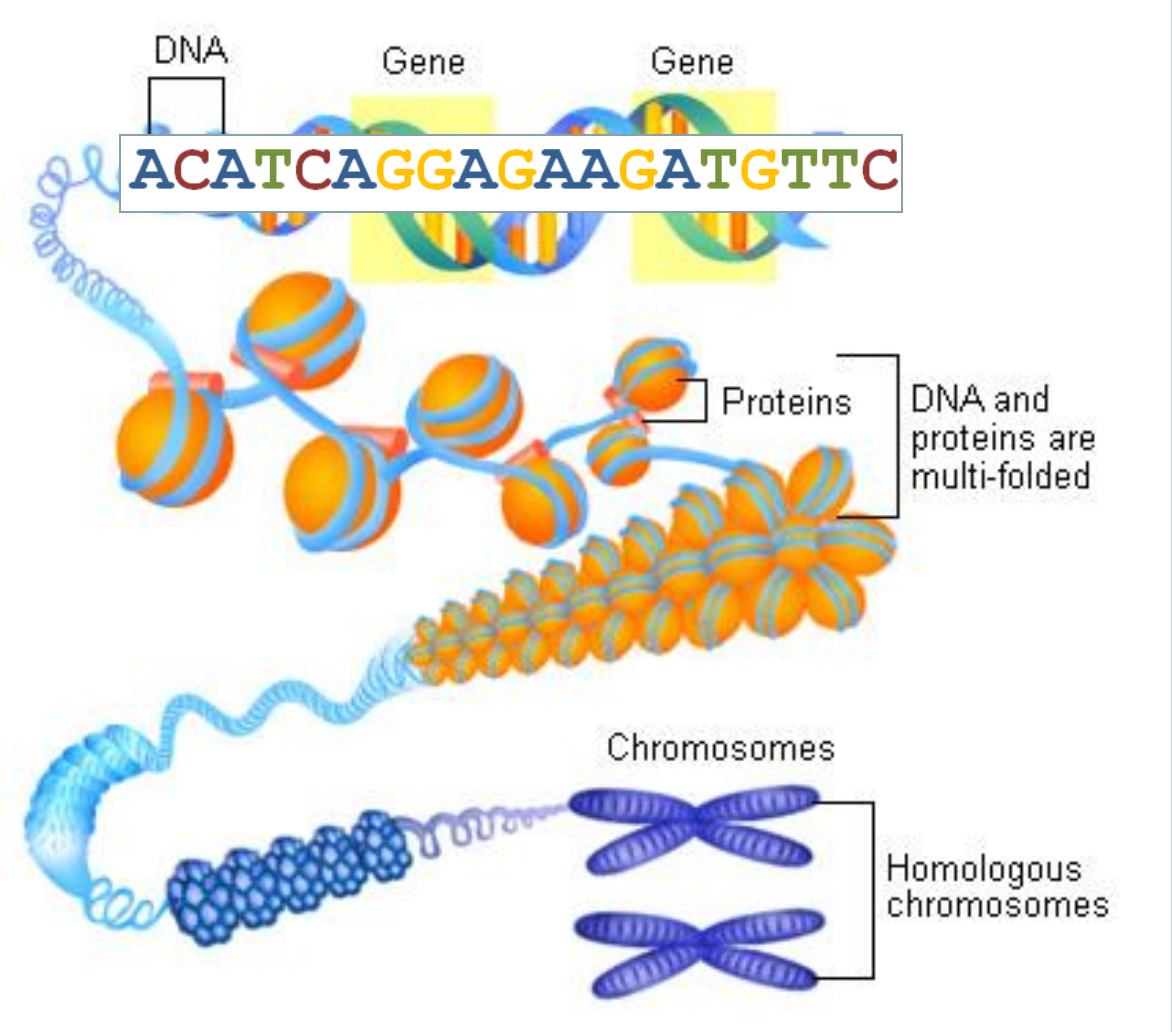
- Basic unit of function: cell
- 10 000 000 000 000 (10^{13})
- genetic material in nucleus: 23 chromosome pairs
- Very long molecules that encode information



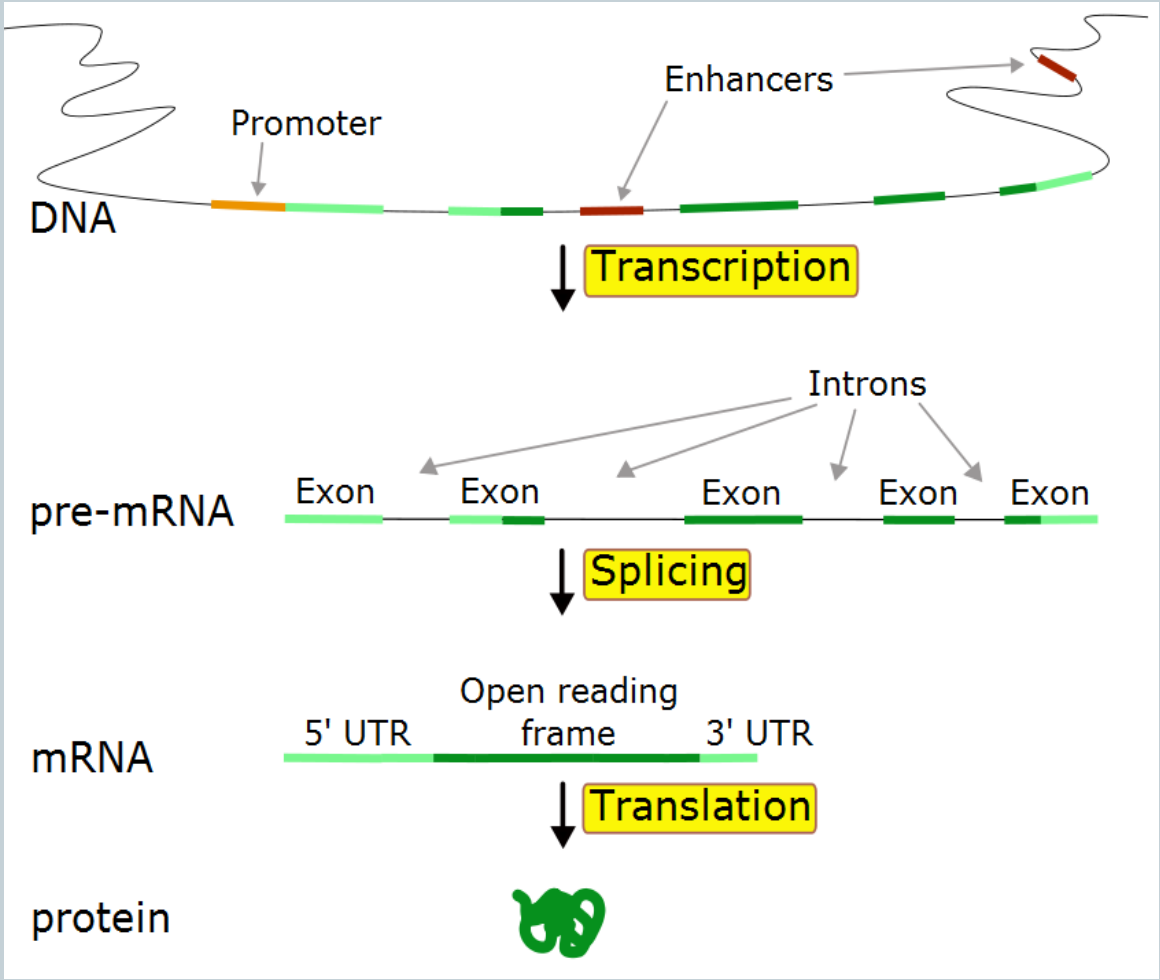
Genetic material

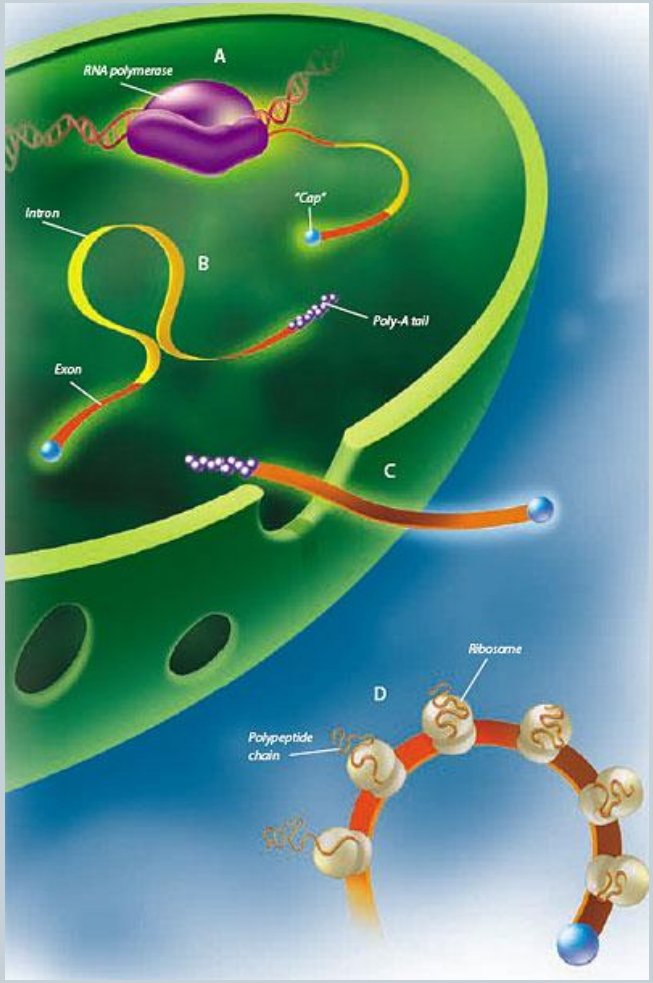


Genetic material

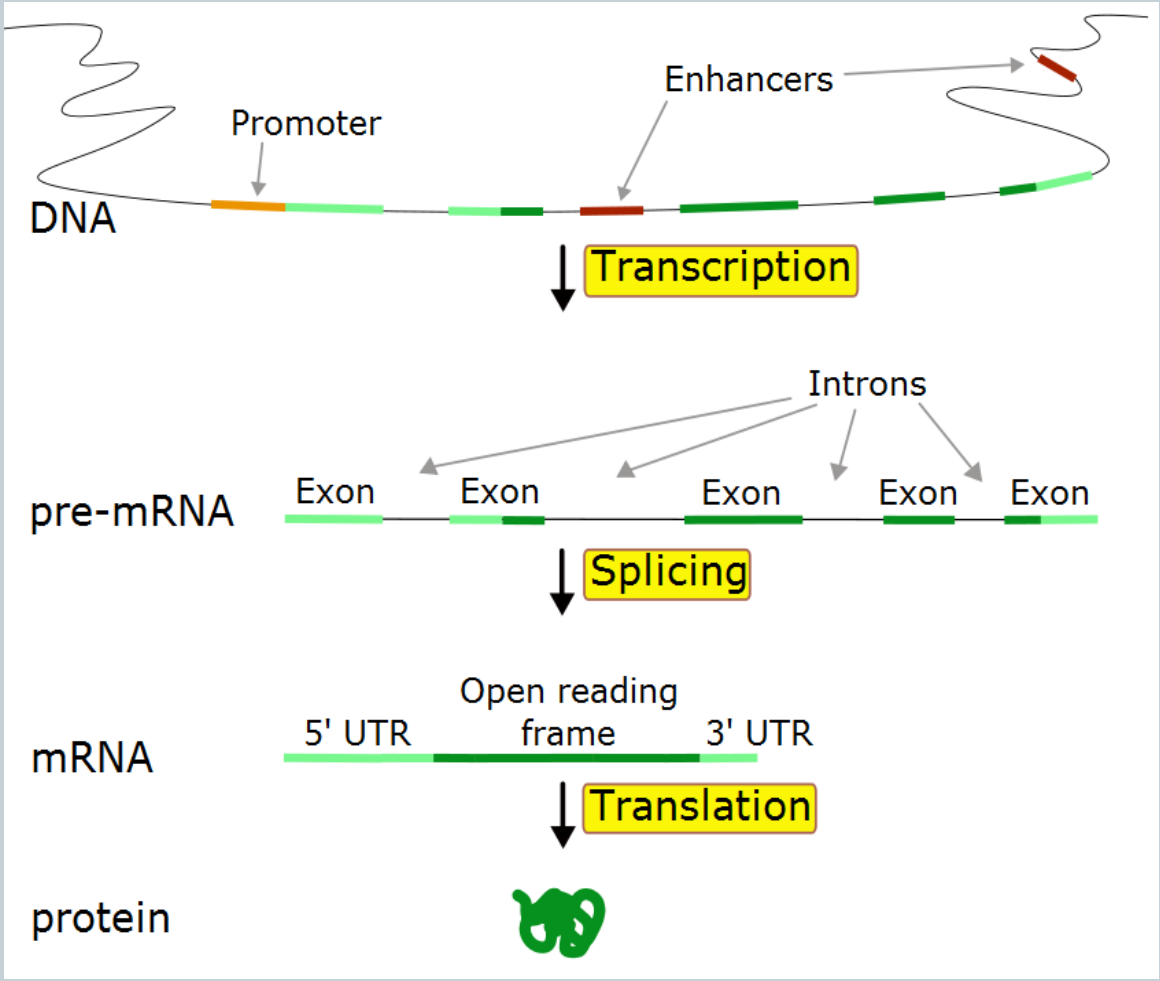


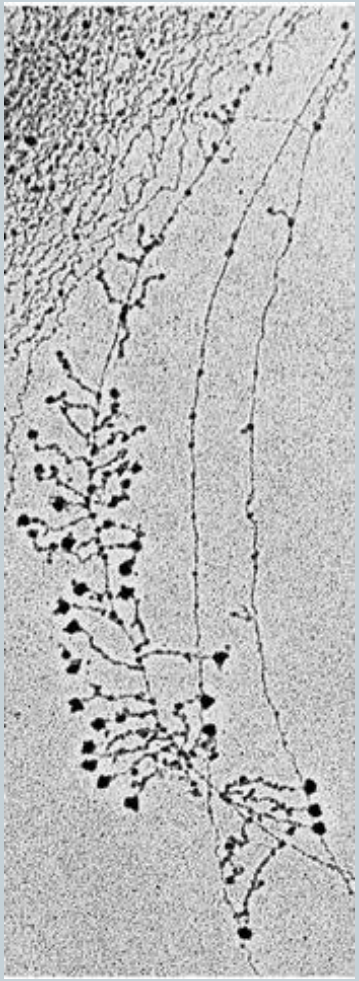
Genetic material



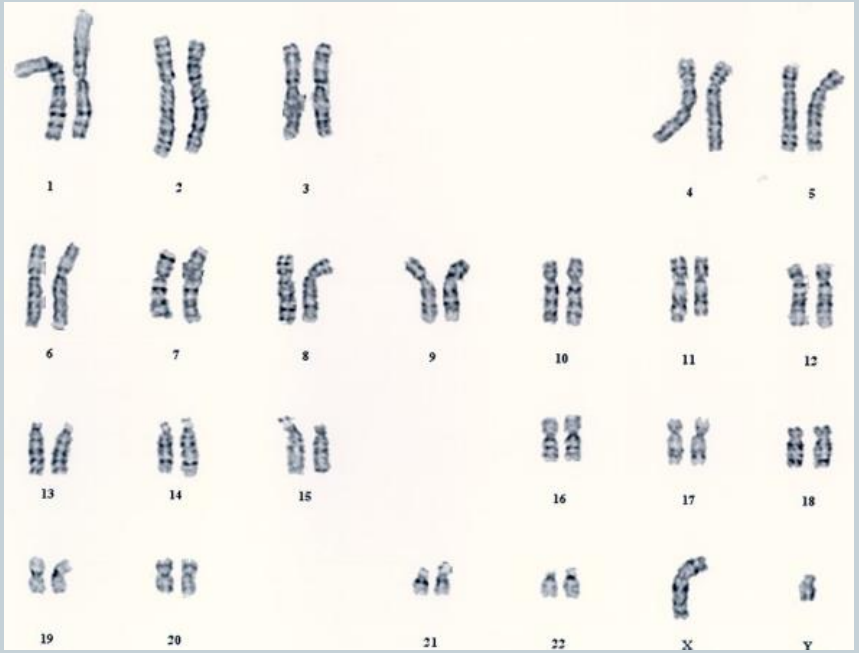


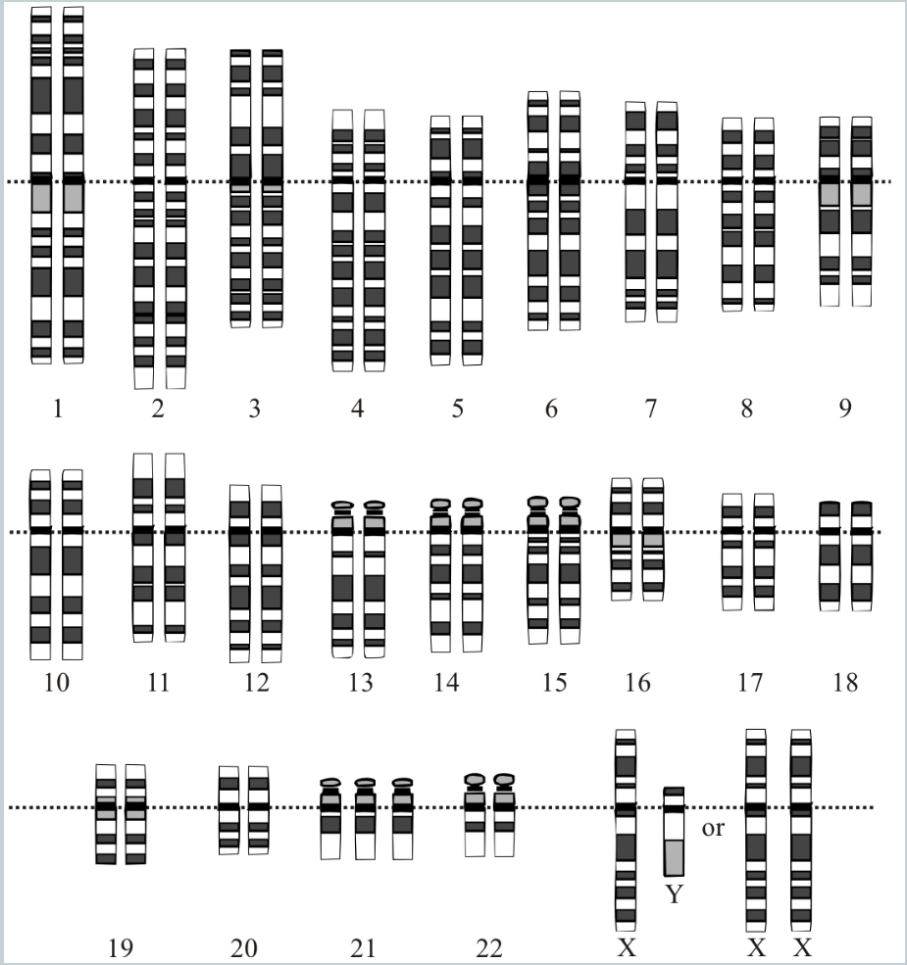
Genetic material

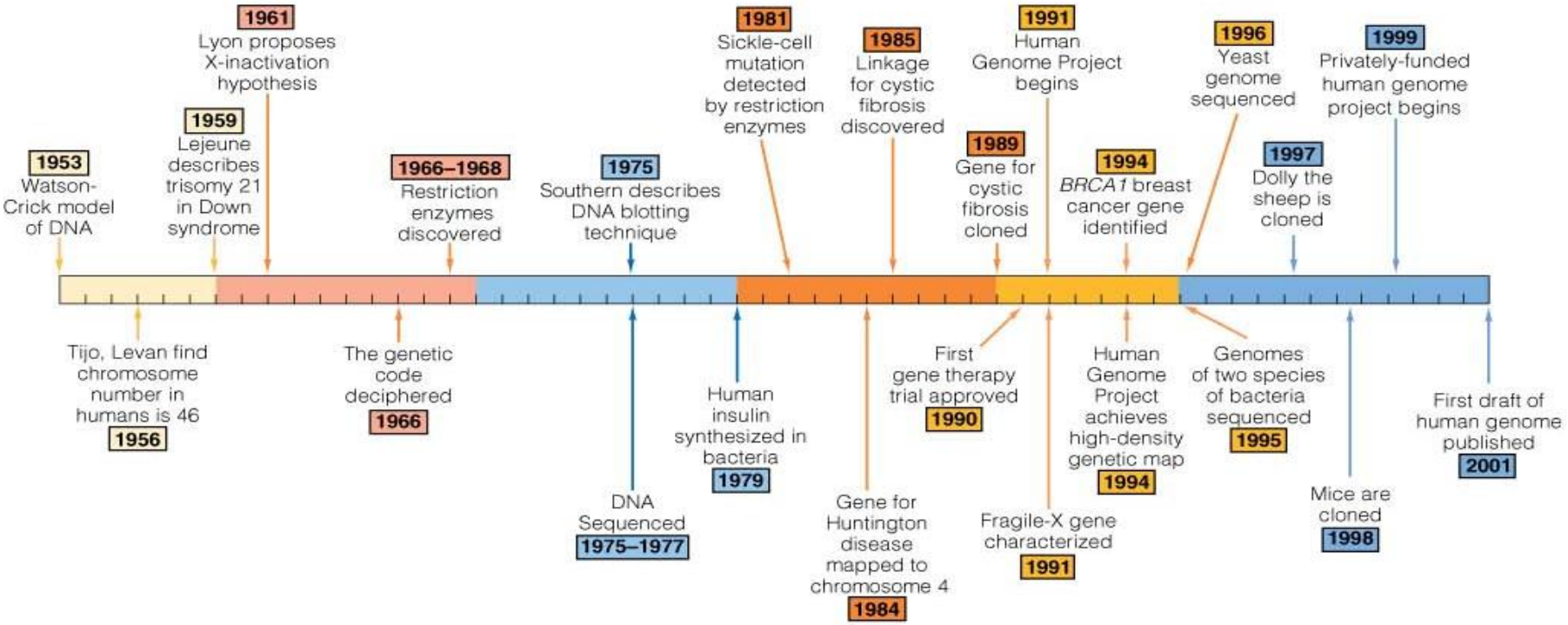




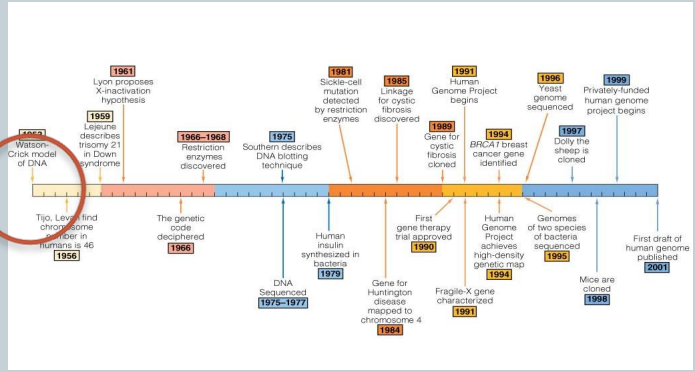
Genetic material

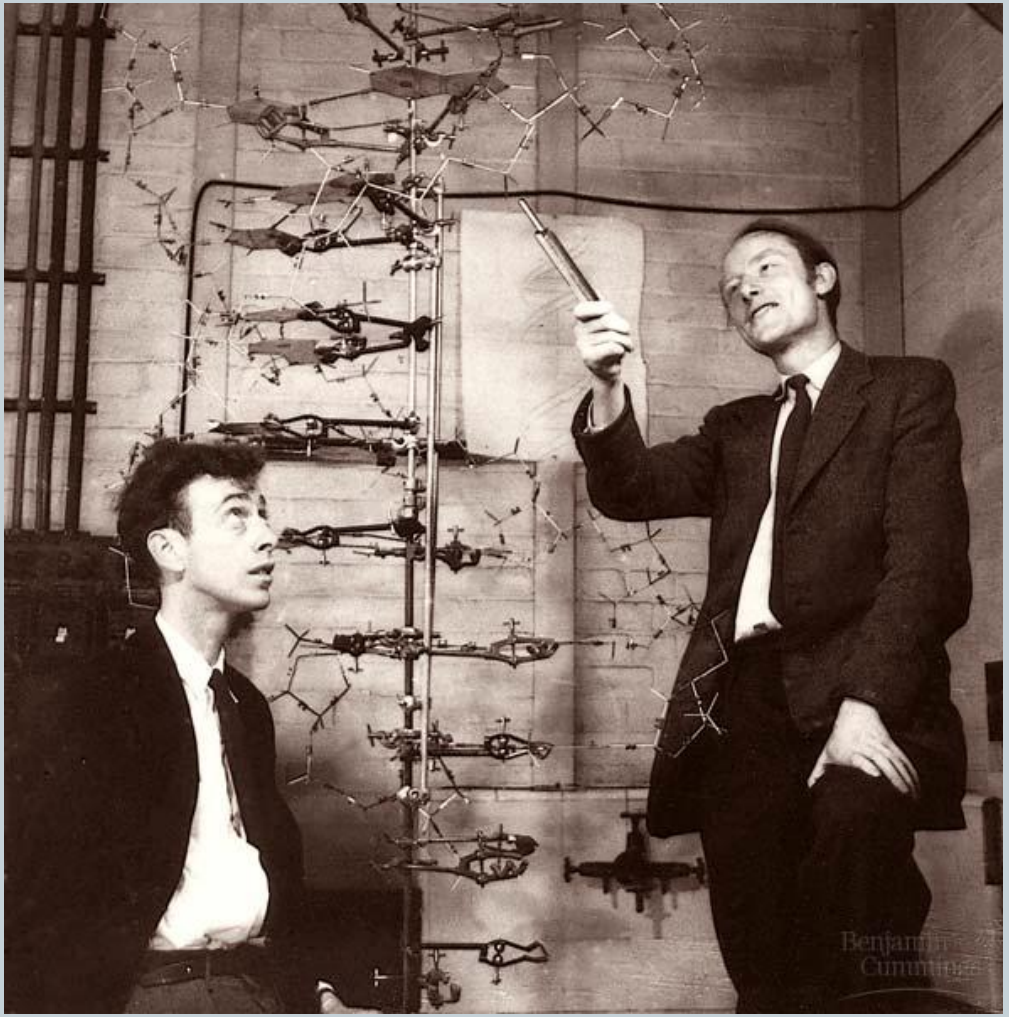




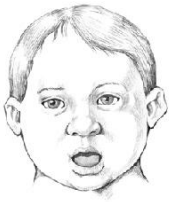

1953
Watson-Crick model of DNA
1956
Tijo, Levan find chromosome number in humans is 46
1959
Lejeune describes trisomy 21 in Down syndrome
1961
Lyon proposes X-inactivation hypothesis
1966
The genetic code deciphered
1966–1968
Restriction enzymes discovered
1975
Southern describes DNA blotting technique
1975–1977
DNA Sequenced
1979
Human insulin synthesized in bacteria
1981
Sickle-cell mutation detected by restriction enzymes
1984
Gene for Huntington disease mapped to chromosome 4
1985
Linkage for cystic fibrosis discovered
1989
Gene for cystic fibrosis cloned
1990
First gene therapy trial approved
1991
Human Genome Project begins
1991
Fragile-X gene characterized
1994
BRCA1 breast cancer gene identified
1994
Human Genome Project achieves high-density genetic map
1995
Genomes of two species of bacteria sequenced
1996
Yeast genome sequenced
1997
Dolly the sheep is cloned
1998
Mice are cloned
1999
Privately-funded human genome project begins
2001
First draft of human genome published

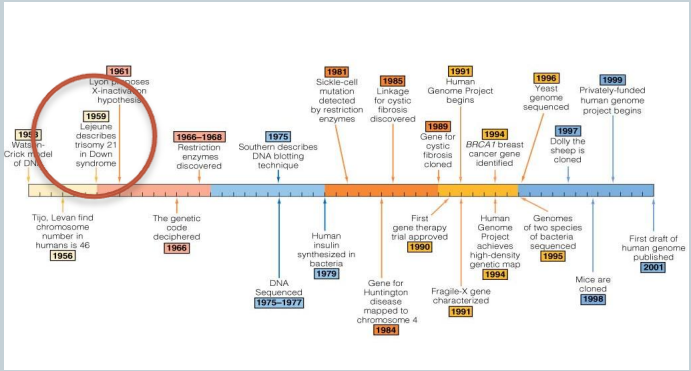
Discoveries

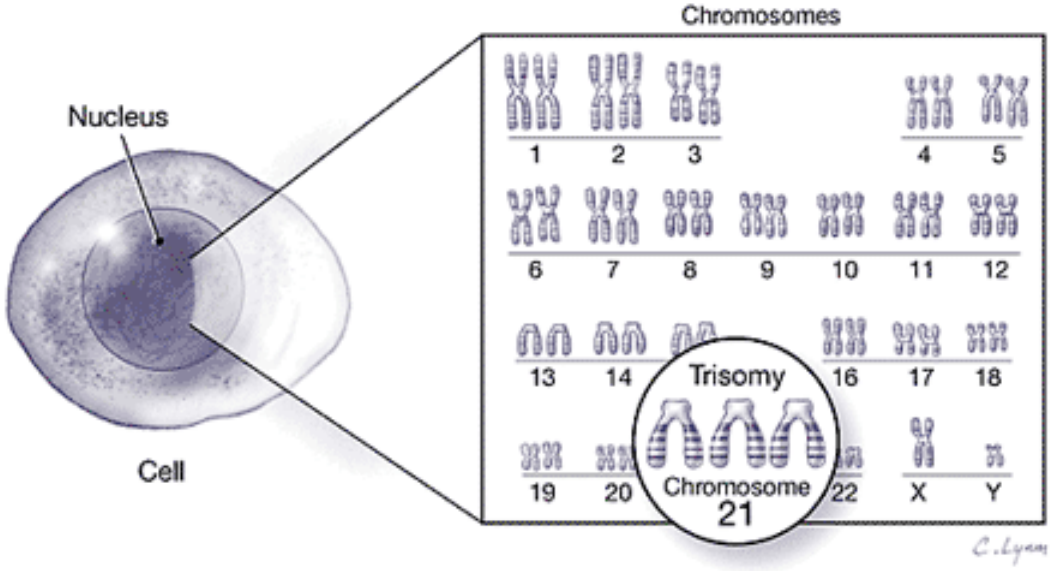




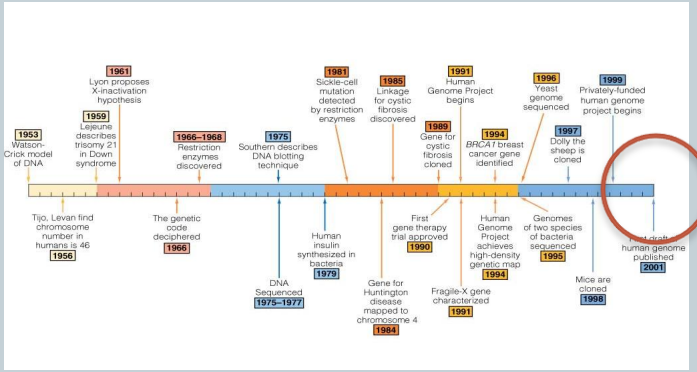
Discoveries







Discoveries



The Human Genome

- “First draft” of the human genome
- International, collabroative research project
- 90 percent of 3 billion base pairs sequenced
  - Initial draft: June 2000
  - Final draft: April 2003
  - 13 year project
  - $300 million value

The Human Genome

- “Human Genome Sequence” of just a few people
- We are all genetically different
  - From other species
  - From our ancestors
  - Between population groups
  - From the person sitting next to you

The Human Genome

- “Human Genome Sequence” of just a few people
- We are all genetically different
  - **From other species**
  - From our ancestors
  - Between population groups
  - From the person sitting next to you

Sequence similarity with other species

Homo Sapiens

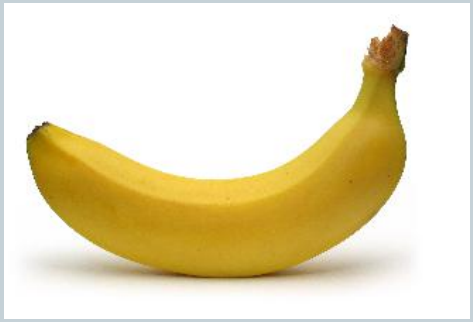
Sequence similarity with other species

Chimpanzee vs. human
sequence identity
for 130 genes (CDS):
98,30%

Sequence similarity with other species

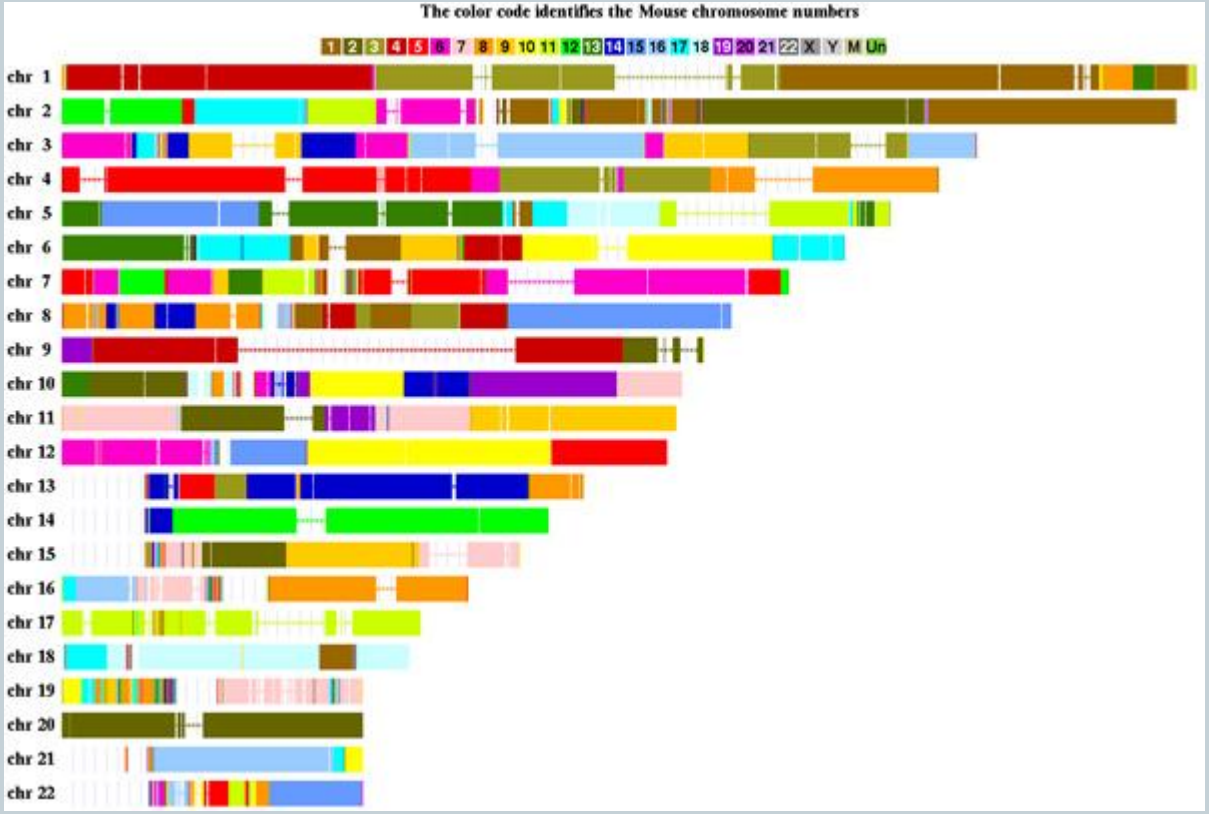
Chimpanzee vs. human
sequence identity
for 130 genes (CDS):
98,30%

50%



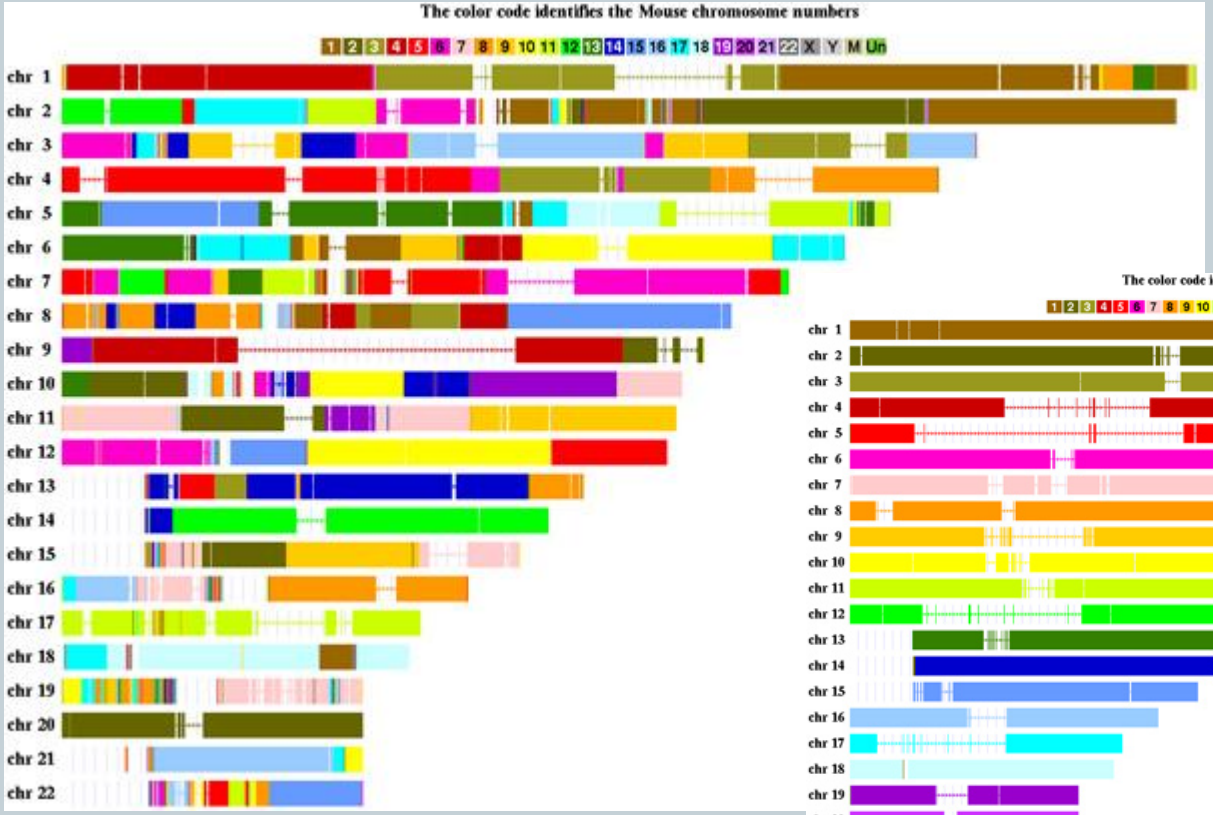
Genomes

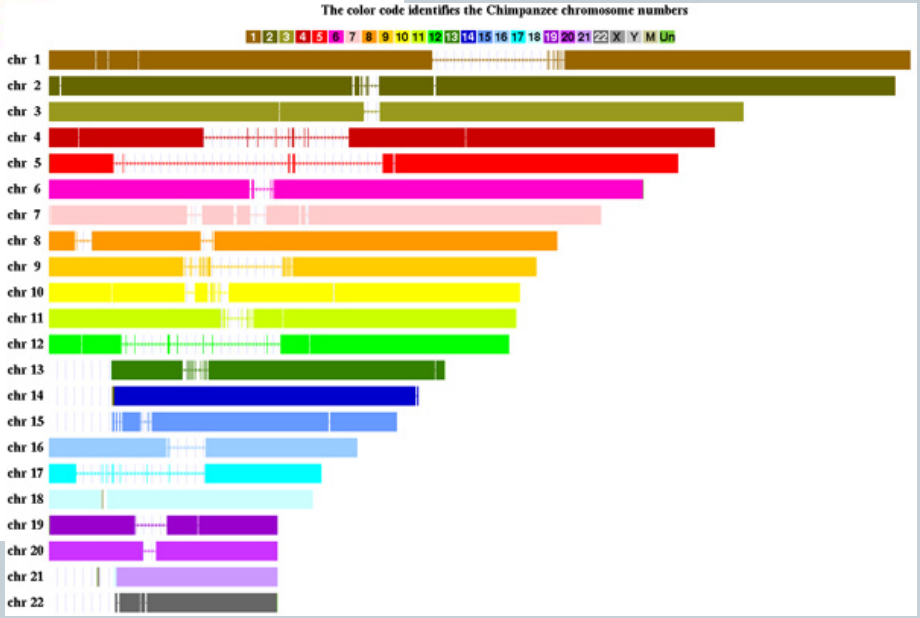
- Human vs. mouse



Genomes

- Human vs. mouse; human vs. chimpanzee





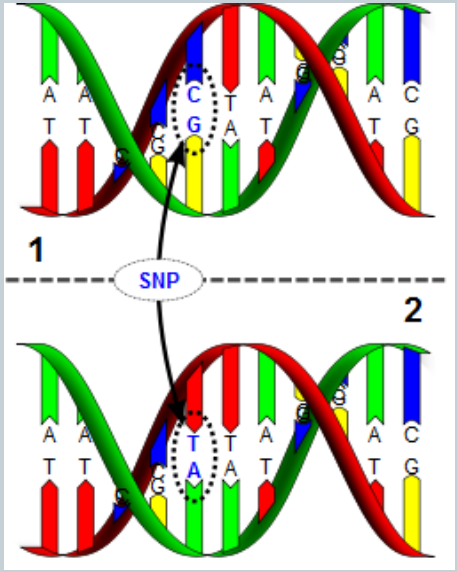
The Human Genome

- “Human Genome Sequence” of just a few people
- We are all genetically different
  - From other species
  - From our ancestors
  - Between population groups
  - **From the person sitting next to you**

Variation in the Human Genome

- **Sequence variation**

```
ACATCAGGAGAAGATGTTC GAGACTTTGCCA
ACATCAGGAGAAGATGTTT GAGACTTTGCCA
ACATCAGGAGAAGATGTT  GAGACTTTGCCA
ACATCAGGAGAAGATGTTCCGAGACTTTGCCA
```



Variation in the Human Genome

- **Sequence variation**

```
ACATCAGGAGAAGATGTTC GAGACTTTGCCA
ACATCAGGAGAAGATGTTT GAGACTTTGCCA
ACATCAGGAGAAGATGTT  GAGACTTTGCCA
ACATCAGGAGAAGATGTTCCGAGACTTTGCCA
```

  - nucleotide diversity between two people: about 0.1% (1/1000)
    - human genome: about 3 billion nucleotides
    - two random people: +/- 3 million nucleotide differences

Variation in the Human Genome

- **Sequence variation**

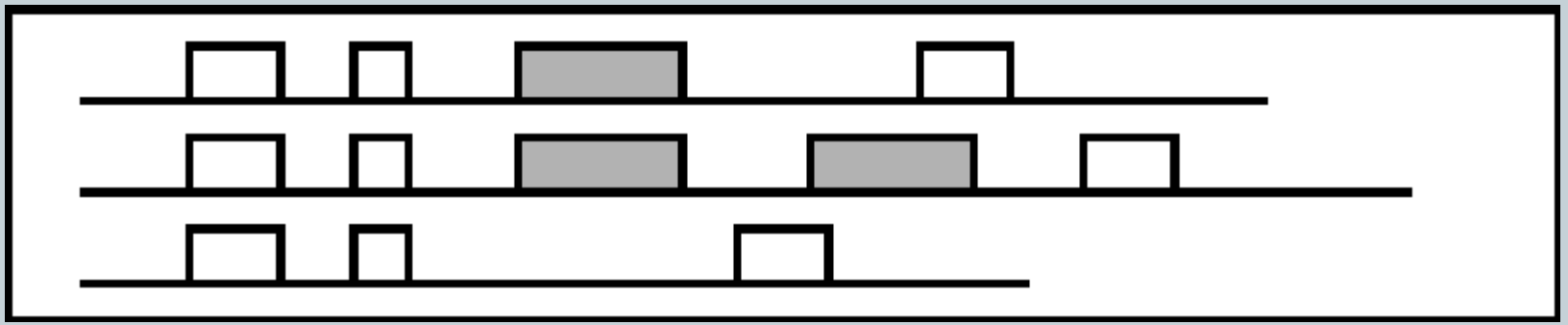
```
ACATCAGGAGAAGATGTTC GAGACTTTGCCA
ACATCAGGAGAAGATGTTT GAGACTTTGCCA
ACATCAGGAGAAGATGTT  GAGACTTTGCCA
ACATCAGGAGAAGATGTTCCGAGACTTTGCCA
```

Variation in the Human Genome

- **Sequence variation**

```
ACATCAGGAGAAGATGTTC GAGACTTTGCCA
ACATCAGGAGAAGATGTTT GAGACTTTGCCA
ACATCAGGAGAAGATGTT  GAGACTTTGCCA
ACATCAGGAGAAGATGTTCCGAGACTTTGCCA
```

- **Copy number variation**



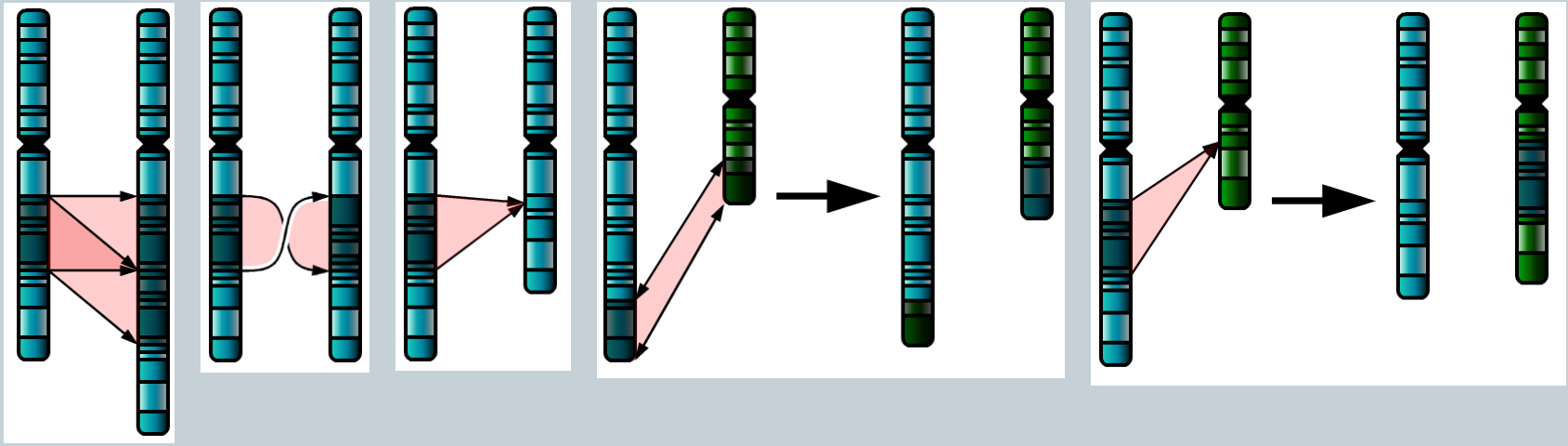
  - human to human genetic variation is estimated to be at least 0.5% (99.5% similarity)

Variation in the Human Genome

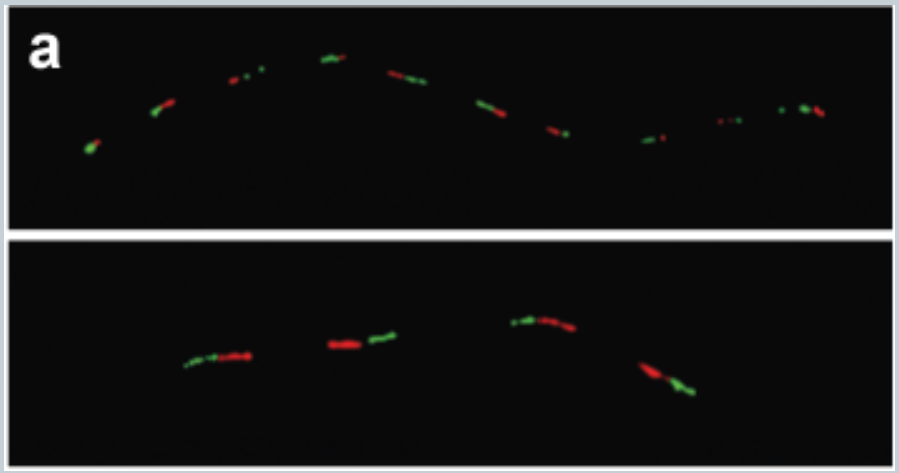
- **Sequence variation**

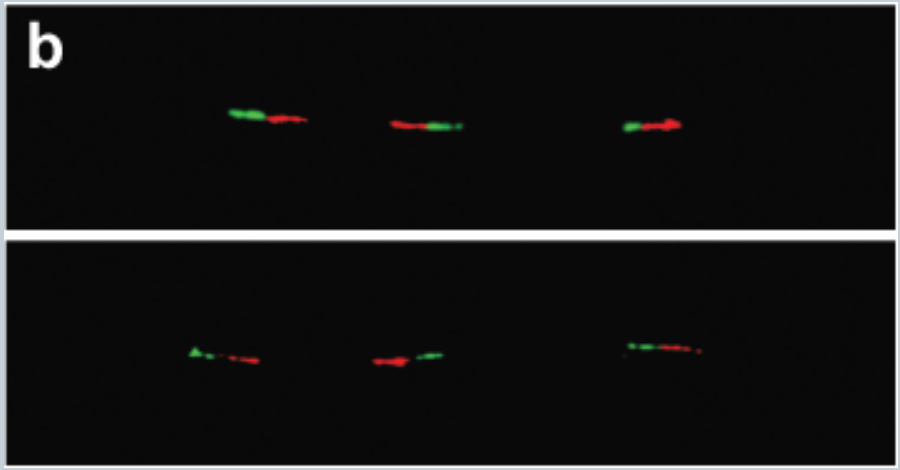
```
ACATCAGGAGAAGATGTTC GAGACTTTGCCA
ACATCAGGAGAAGATGTTT GAGACTTTGCCA
ACATCAGGAGAAGATGTT  GAGACTTTGCCA
ACATCAGGAGAAGATGTTCCGAGACTTTGCCA
```

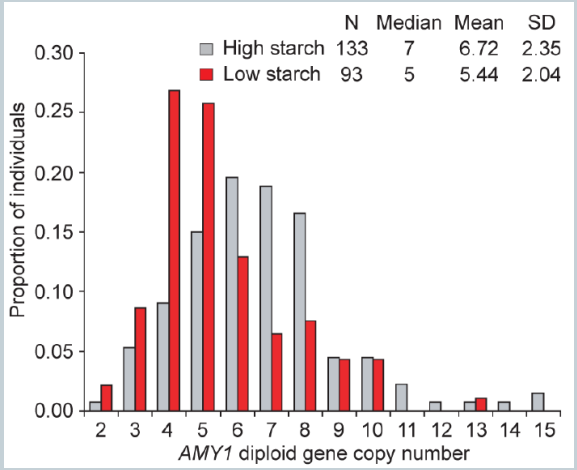
- **Copy number variation**



Human Genome Copy Number Variation







Perry *et al.*

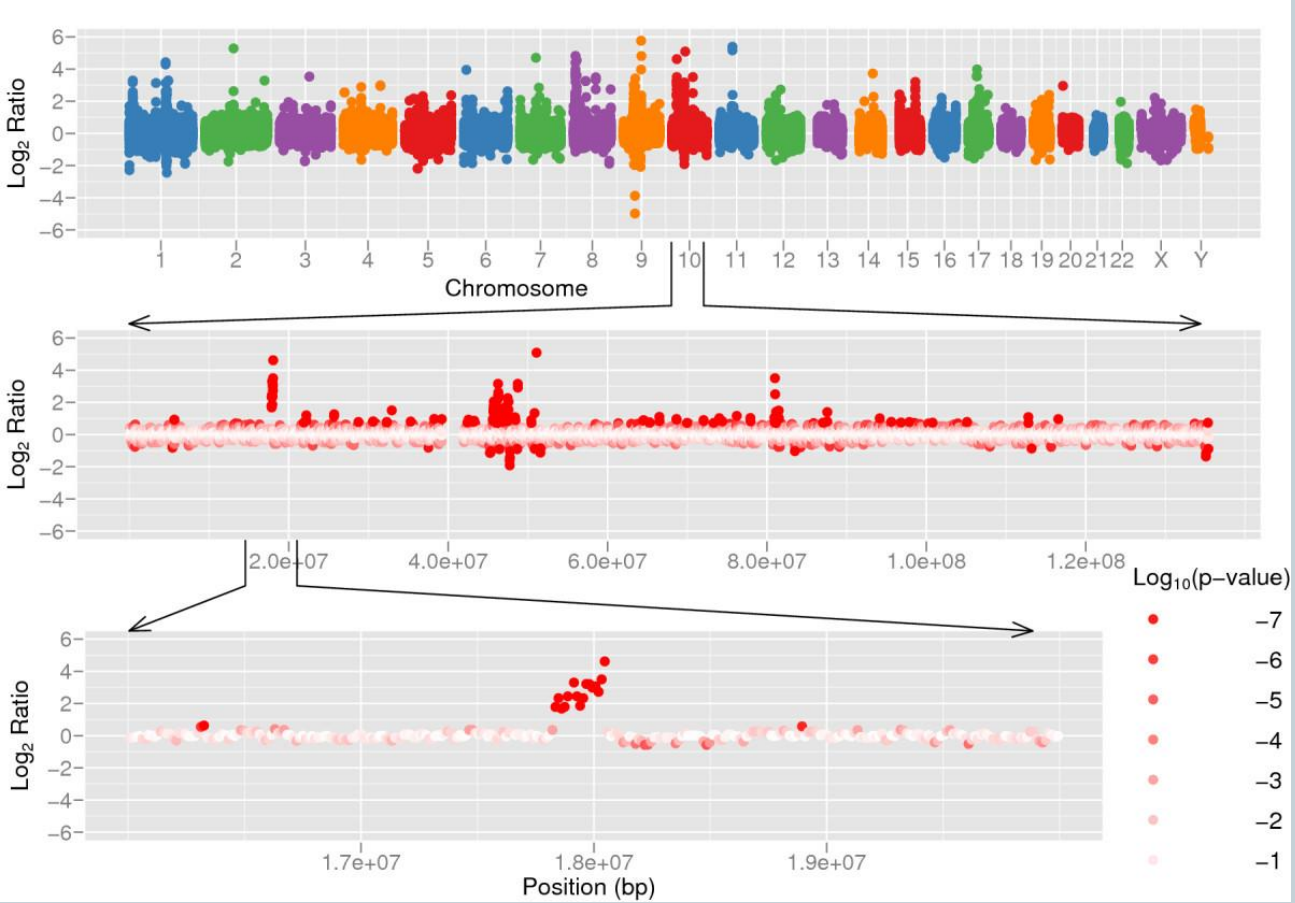
Variation in the Human Genome

- Because of the higher resolution of the new systems, we will uncover larger numbers of smaller variants

Wellcome Trust Sanger Institute

Variation in the Human Genome

- Venter vs. Watson



Xie *et al.*

Chapter 1 – introduction and context

TECHNIQUE, CHALLENGES AND DATA SOURCES

Chapter 1 – introduction and context

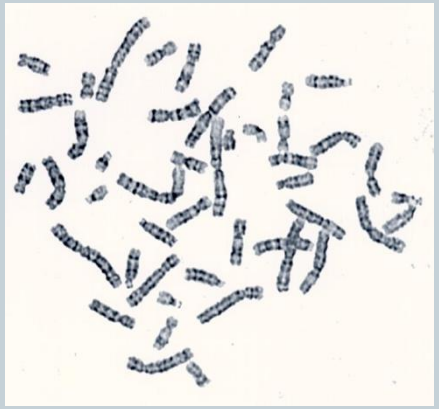
- The Array-CGH technique
- Challenges and data sources
  - Raw data analysis
  - Downstream analysis: interpretation in context
    - Gene function
    - Phenotypes
    - Information from text mining
    - …

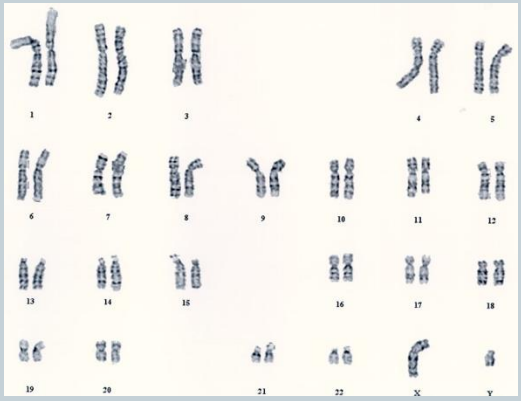
Chapter 1 – introduction and context

- **The Array-CGH technique**
- Challenges and data sources
  - Raw data analysis
  - Downstream analysis: interpretation in context
    - Gene function
    - Phenotypes
    - Information from text mining
    - …

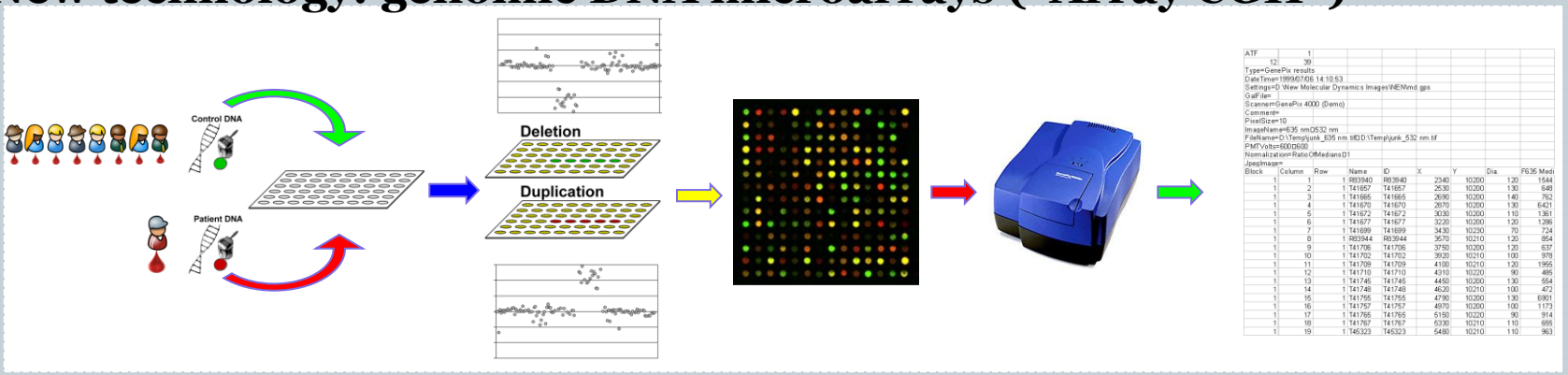
Clinical Cytogenetics

Classical technique: band stain karyotyping

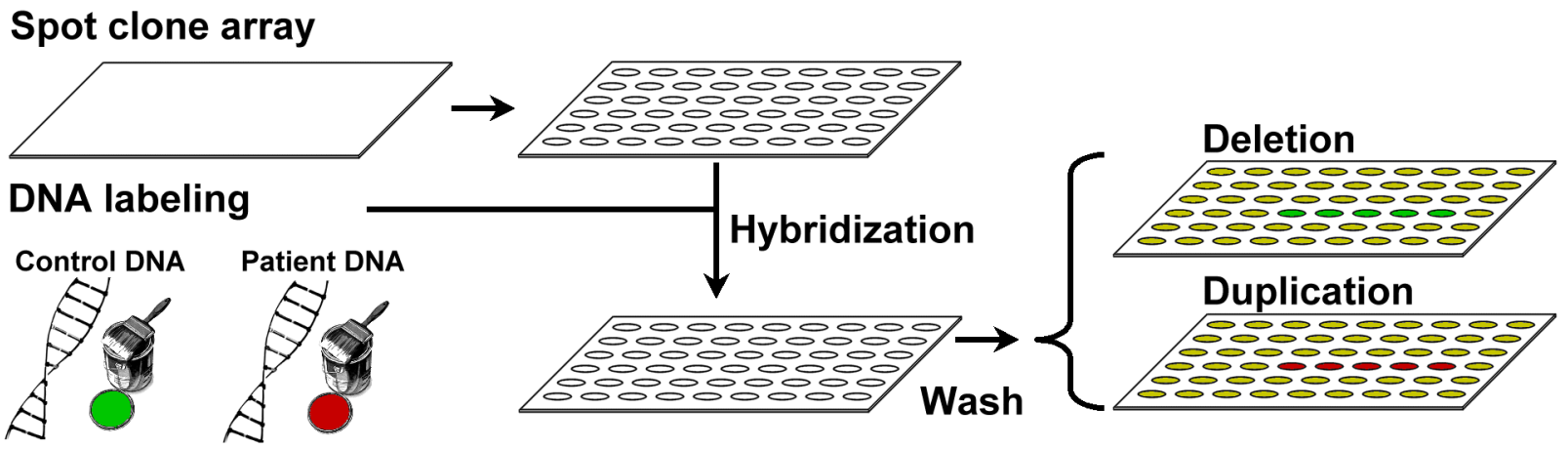


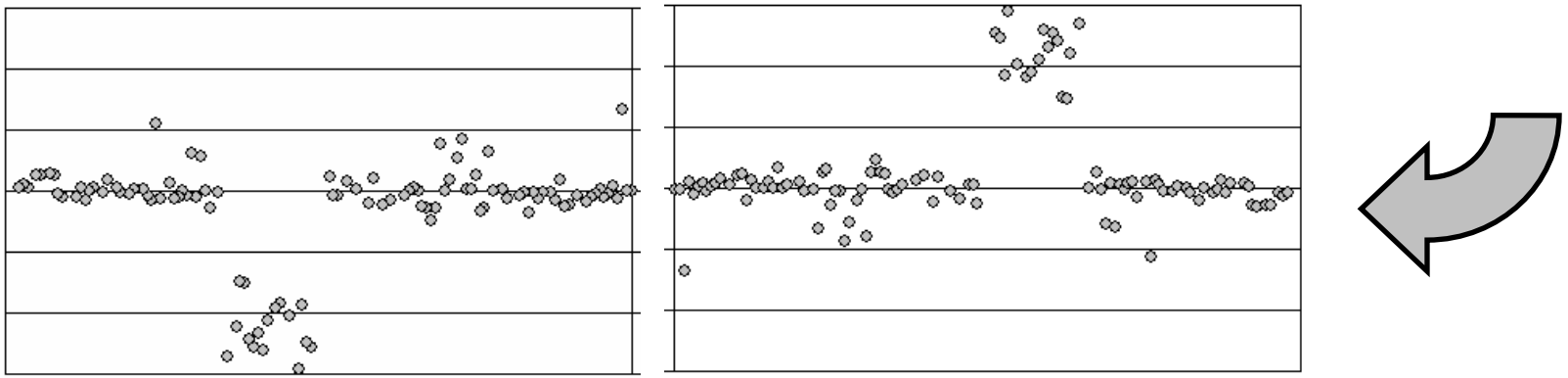


New technology: genomic DNA microarrays ("Array CGH")



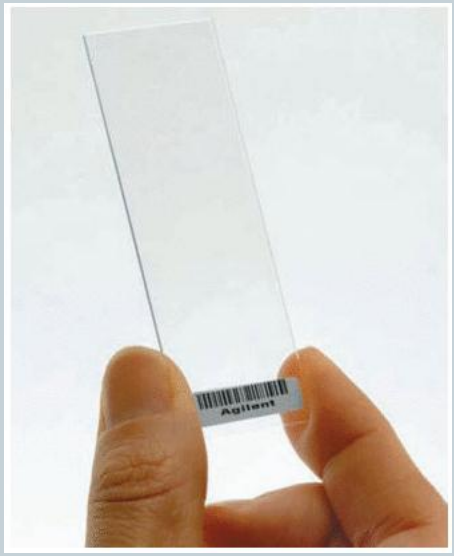
DNA microarray experiments





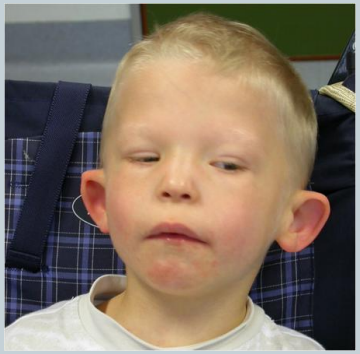
DNA microarray experiments

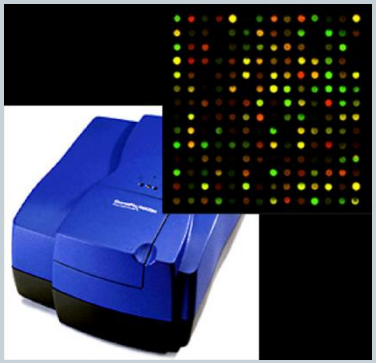
- Microarray slides

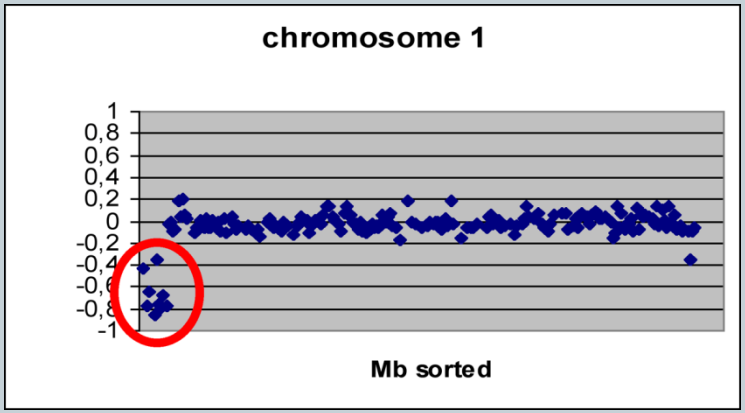


Cytogenetics Diagnosis

Patient with developmental anomaly

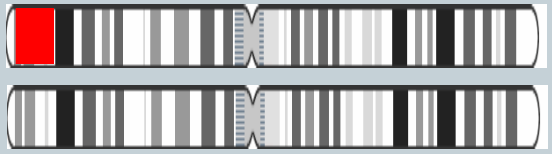






- 2-3% of live births with major congenital anomaly
- Medical & social impact of diagnosis
  - Reduce family distress – end of "diagnostic odyssey"
  - Estimate risk of recurrence for next pregnancy
  - Prevent clinical complications – life planning

- Array CGH is superior to previous technology
  - ++ resolution, + speed, +/- cost
  - Rapid adoption by cytogenetics labs
- ! Need for tools to manage efficiently the flood of genetic information in a *routine clinical* setting



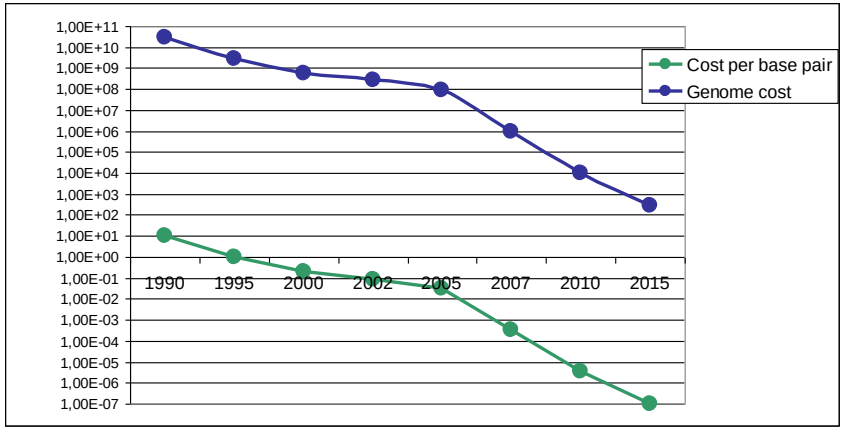
Detected genetic anomaly

Chapter 1 – introduction and context

- The Array-CGH technique
- **Challenges and data sources**
  - Raw data analysis
  - Downstream analysis: interpretation in context
    - Gene function
    - Phenotypes
    - Information from text mining
    - …

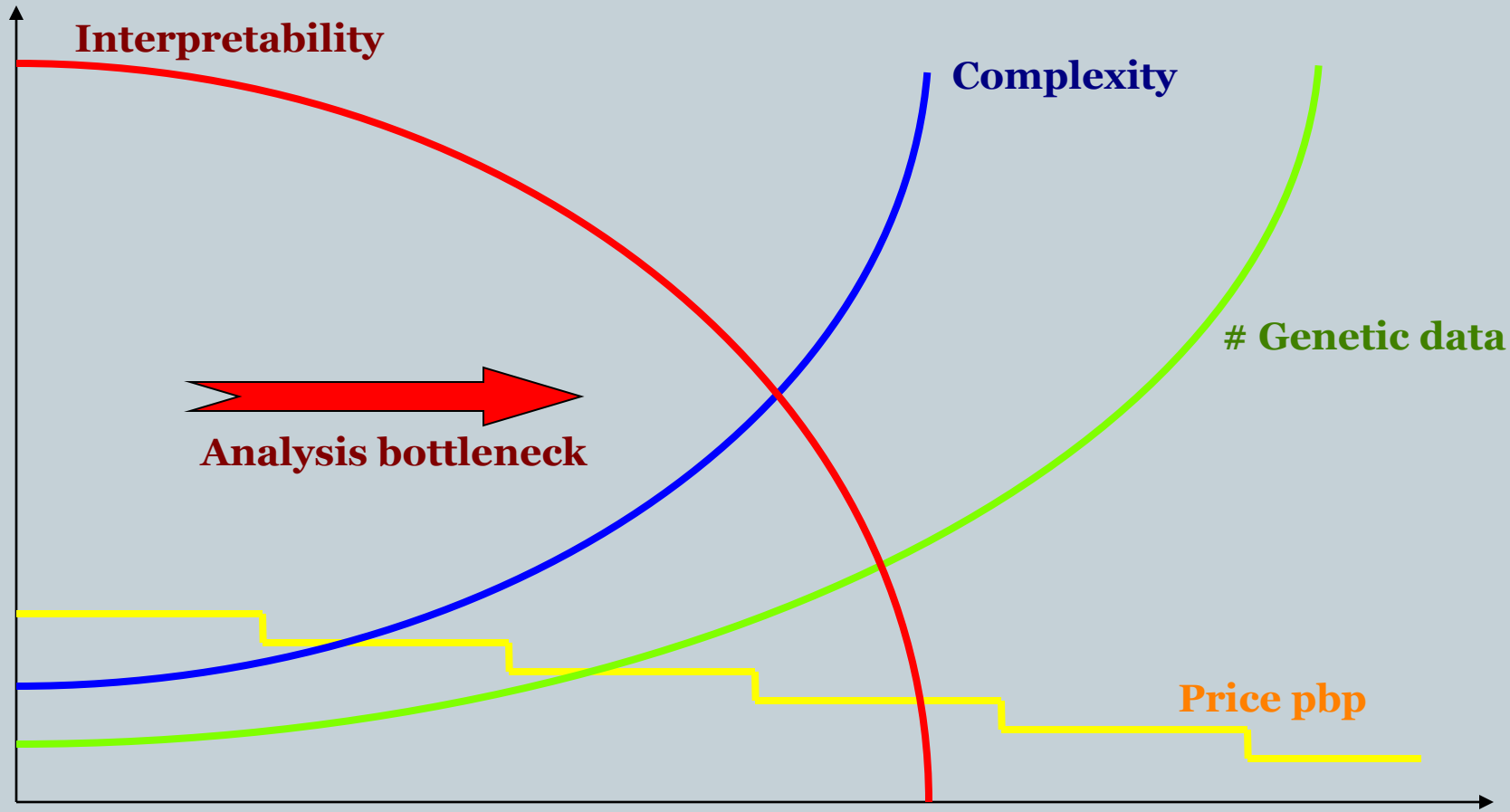
The 1000 $ Genome

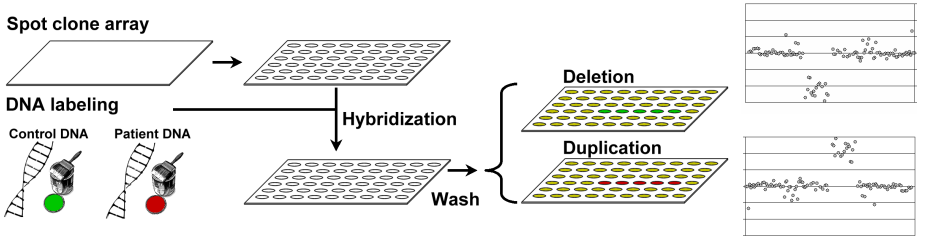
- **Human genome project**
  - Initial draft: June 2000
  - Final draft: April 2003
  - 13 year project
  - $300 million value with 2002 technology
- **Personal genome**
  - June 1, 2007
  - Genome of James Watson, co-discoverer of DNA double helix, is sequenced
    - $1.000.000
    - Two months
- **€50.000-genome**
  - Available (CG: 5000 €)
- **€1000-genome**
  - Expected



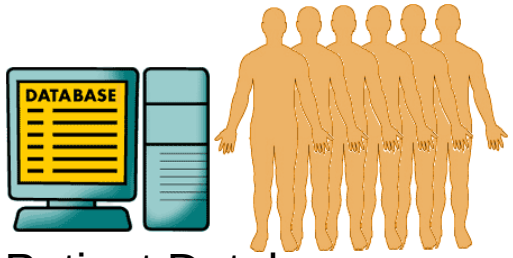
Year	Cost per base pair	Genome cost
1990	10	3E+10
1995	1	3.000.000.000
2000	0.2	600.000.000
2002	0.09	270.000.000
2005	0.03	90.000.000
2007	0.000333333	1.000.000
2010	3.33333E-06	10000
2015	0.0000001	300

Interpretability of Genomic Data

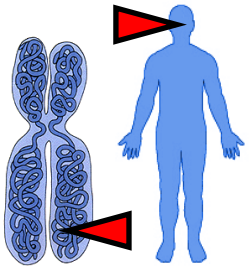




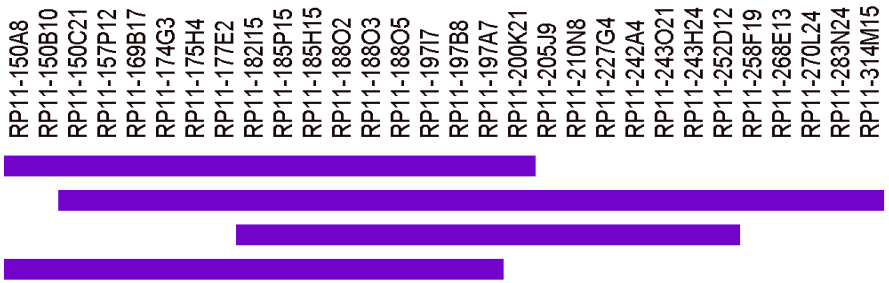
DNA Microarray Experiments



Patient Database

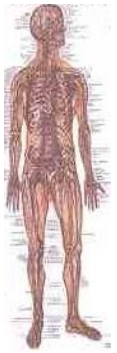


Biomed Literature



Patient Genotype

170520_Pulmonary stenosis
170403_Congenital hernia of diaphragm
100100_Face, general abnormalities
130100_Teeth, general abnormalities
280504_Isolated growth hormone deficiency
120401_Cleft palate
030105_Microcephaly
140100_Voice, general abnormalities
090101_Broad base to nose
340104_Dimples



Patient Phenotype

Chapter 2 – Array CGH data management and analysis

RAW ANALYSIS, EXPERIMENT DESIGN, AND GENOME POLYMORPHISM

Chapter 2 – Array CGH data management and analysis

- Raw data analysis : quality criteria and pitfalls
  - Data normalization
  - Spot and aberration calling (segmentation)
  - Threshold definition
- Alternative array design (loop analysis)
  - Joke Allemeersch credited for statistical models
  - LOOP tool: web architecture for data management
- Benign copy number variation analysis (clone wars)
- Importance of data management
  - Data repositories that feature analysis tools
  - Examples: ArrayCGHBase, Decipher, Store+Bench, Database of Genomic Variants, ECARUCA, …

Chapter 2 – Array CGH data management and analysis

- Raw data analysis : quality criteria and pitfalls
  - Data normalization
  - Spot and aberration calling (segmentation)
  - Threshold definition
- Alternative array design (loop analysis)
  - Joke Allemeersch credited for statistical models
  - **LOOP tool: web architecture for data management**
- Benign copy number variation analysis (clone wars)
- Importance of data management
  - Data repositories that feature analysis tools
  - Examples: ArrayCGHBase, Decipher, Store+Bench, Database of Genomic Variants, ECARUCA, …

LOOP and Clone Wars

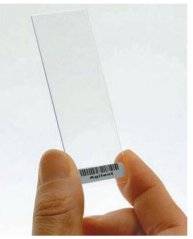
- Database for storing experiments
- Slide Annotation
- Visualisation
- Normalisation
- Statistical method for analysis (LIMMA)

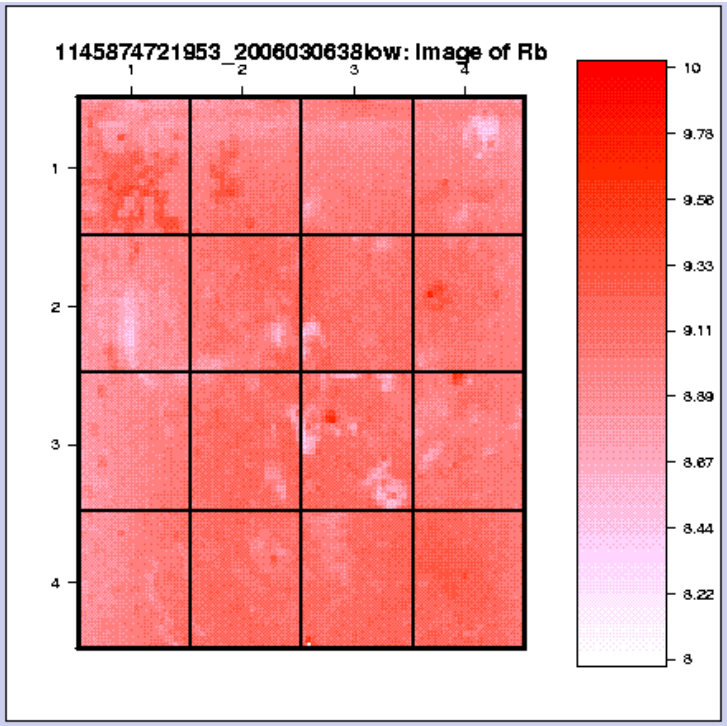

Upload .gpr
Add .gpr · View Slides · View Loop Designs · Log on · Log off
On this page, you can add a single .gpr file to the slide repository. Files are added one at a time. You can combine multiple slides into a loop design later on.
Adding a slide to the repository
Question 1 - Where can I find the .gpr-file?
d:\431457.gpr
Bladeren...
Question 2 - Who conducted the experiment?
Cindy Melotte
Cindy
Question 3 - When was the experiment conducted?
May 1, 2006
Question 4 - Please provide a reference for your Cy5 sample[1].
patient 4687
Question 5 - Please provide a reference for your Cy3 sample[1].
patient 4645
Question 6 - Below, you can add comments to this slide, if necessary.
hybr ok
To store the slide into the repository[2], press Add.
Add
[1] It is important for this reference to be correct. The system will use it to infer how loop designs are set up.
[2] Please note that entering a slide into the repository can take some time, as .gpr-files are rather large.
Add .gpr File
View Slides
View Loop Designs

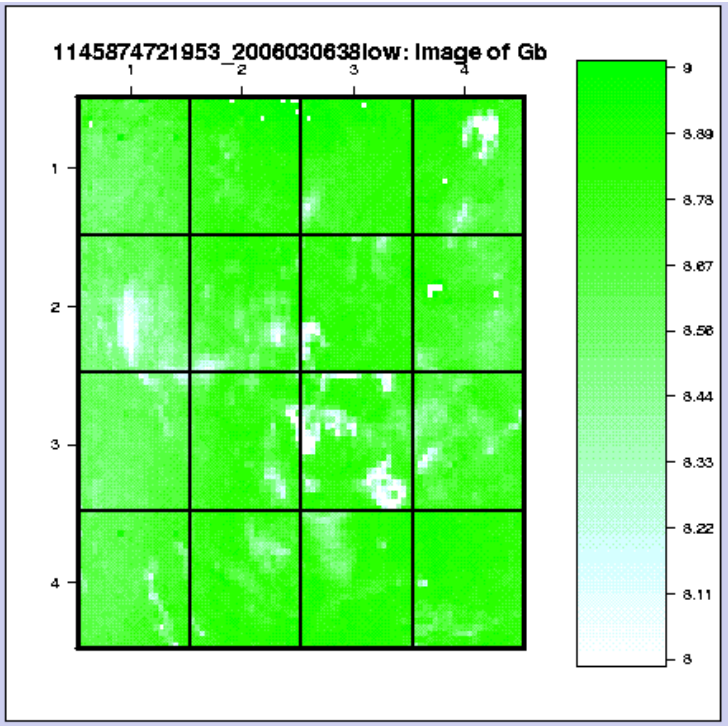

Slide List
Add .gpr - View Slides - View 2way Results - View Loop Designs - Log on - Log off
Slide name
search
id name owner added cy5 cy3 detail edit check
5300 2009052627low vicky Jun 22, 2009 399881 XXY-ref
5299 2009052627high vicky Jun 22, 2009 399970 XXY-ref
5298 2009052626low vicky Jun 18, 2009 XY-ref 399376
5297 2009052626high vicky Jun 18, 2009 399376 XY-ref
5296 2009052625low vicky Jun 18, 2009 XY-ref 400390
5295 2009052625high vicky Jun 18, 2009 400390 XY-ref
5294 2009052619low vicky Jun 18, 2009 XX-ref 328961
5293 2009052619high vicky Jun 18, 2009 328961 XX-ref
5292 2009052613low pascale Jun 17, 2009 384957 357911
5291 2009052613high pascale Jun 17, 2009 384778 384957
5290 2009052612low pascale Jun 17, 2009 357911 384778
5289 2009052612high pascale Jun 17, 2009 124380 44184
5288 2009052611low pascale Jun 17, 2009 394918 124380
5287 2009052611high pascale Jun 17, 2009 44184 394918
5286 2009052624low Pascal Jun 15, 2009 baboon xxy
5285 2009052624high2 pascal Jun 15, 2009 G2 xxy
5284 2009052623low PASCAL Jun 15, 2009 S xxy
5283 2009052623high pascal Jun 15, 2009 G1 XXY
Show older slides
Perform 2way Analysis
Create Loop Design
R version 2.0.1, 2004-11-15 is active.
Add .gpr File
View Slides
View Loop Designs
ESAT

Loop Design List

id	date	details	status	Results
243	Apr 24, 2006	Slide set: [**2006030638low 2006030638high 2006022739low**]	**Done.** (12:34:17)	overview
		Slide **2006030638low**: 79162 (cy5) vs. 318719 (cy3) Slide **2006030638high**: 282913 (cy5) vs. 79162 (cy3) Slide **2006022739low**: 318719 (cy5) vs. 282913 (cy3)		Patient 282913 Patient 318719 Patient 79162
242	Apr 24, 2006	Slide set: [**2006022739high 2006022738low 2006022738high**]	**Done.** (12:09:35)	overview
		Slide **2006022739high**: 232160 (cy5) vs. 303310 (cy3) Slide **2006022738low**: 80931 (cy5) vs. 232160 (cy3) Slide **2006022738high**: 303310 (cy5) vs. 80931 (cy3)		Patient 80931 Patient 303310 Patient 232160
241	Apr 14, 2006	Slide set: [**2006022737low 2006022737high 2006022736low**]	**Done.** (14:23:25)	overview
		Slide **2006022737low**: 159192 (cy5) vs. 303310 (cy3) Slide **2006022737high**: 255280 (cy5) vs. 159192 (cy3) Slide **2006022736low**: 303310 (cy5) vs. 255280 (cy3)		Patient 255280 Patient 303310 Patient 159192
240	Apr 14, 2006	Slide set: [**2006022736high 2006022735low 2006022735high**]	**Done.** (14:23:20)	overview
		Slide **2006022736high**: 75022 (cy5) vs. 165531 (cy3) Slide **2006022735low**: 74537 (cy5) vs. 75022 (cy3) Slide **2006022735high**: 165531 (cy5) vs. 74537 (cy3)		Patient 74537 Patient 165531 Patient 75022
239	Apr 12, 2006	Slide set: [**2006022724low 2006022724high 2006022723low**]	**Done.** (12:39:09)	overview
		Slide **2006022724low**: 259497 (cy5) vs. 302222 (cy3) Slide **2006022724high**: 329822 (cy5) vs. 259497 (cy3) Slide **2006022723low**: 302222 (cy5) vs. 329822 (cy3)		Patient 329822 Patient 302222 Patient 259497
238	Apr 12, 2006	Slide set: [**2006022723high 2006022722low 2006022722high**]	**Done.** (12:35:12)	overview
		Slide **2006022723high**: 12211 (cy5) vs. 316554 (cy3) Slide **2006022722low**: 67043 (cy5) vs. 12211 (cy3) Slide **2006022722high**: 316554 (cy5) vs. 67043 (cy3)		Patient 67043 Patient 316554 Patient 12211
237	Apr 11, 2006	Slide set: [**2006022719b 2006022720a_corr 2006022719a**]	**Done.** (12:55:44)	overview
		Slide **2006022719b**: 292023 (cy5) vs. 198782 (cy3) Slide **2006022720a_corr**: 93433 (cy5) vs. 292023 (cy3) Slide **2006022719a**: 198782 (cy5) vs. 93433 (cy3)		Patient 93433 Patient 198782 Patient 292023
236	Apr 11, 2006	Slide set: [**2006022720b_corr 2006022721a 2006022721b**]	**Done.** (12:52:58)	overview
		Slide **2006022720b_corr**: 158303 (cy5) vs. 213093 (cy3) Slide **2006022721a**: 203174 (cy5) vs. 158303 (cy3) Slide **2006022721b**: 213093 (cy5) vs. 203174 (cy3)		Patient 203174 Patient 213093 Patient 158303
86	Aug 16, 2005	Slide set: [**2005042030 2005042032 2005042031**]	**Done.** (10:34:21)	overview
		Slide **2005042030**: tris13 (cy5) vs. trisX (cy3) Slide **2005042032**: normal female (cy5) vs. tris13 (cy3) Slide **2005042031**: trisX (cy5) vs. normal female (cy3)		Patient tris13 Patient normal_female Patient trisX

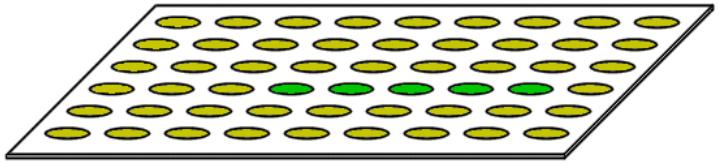






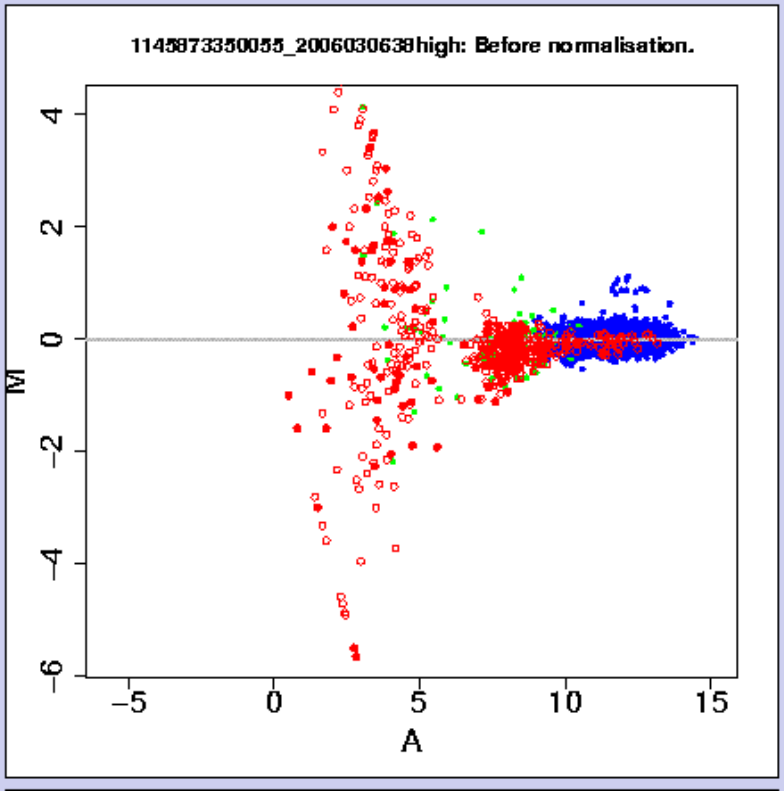
A second indication for the array quality is the number of spots above background. Here, we call a spot above background if Fg>Bg+2*sd(Bg). In the table, the number and percentage of the spots that are above background, is shown; for both channels combined, and for the green and red channel separately.

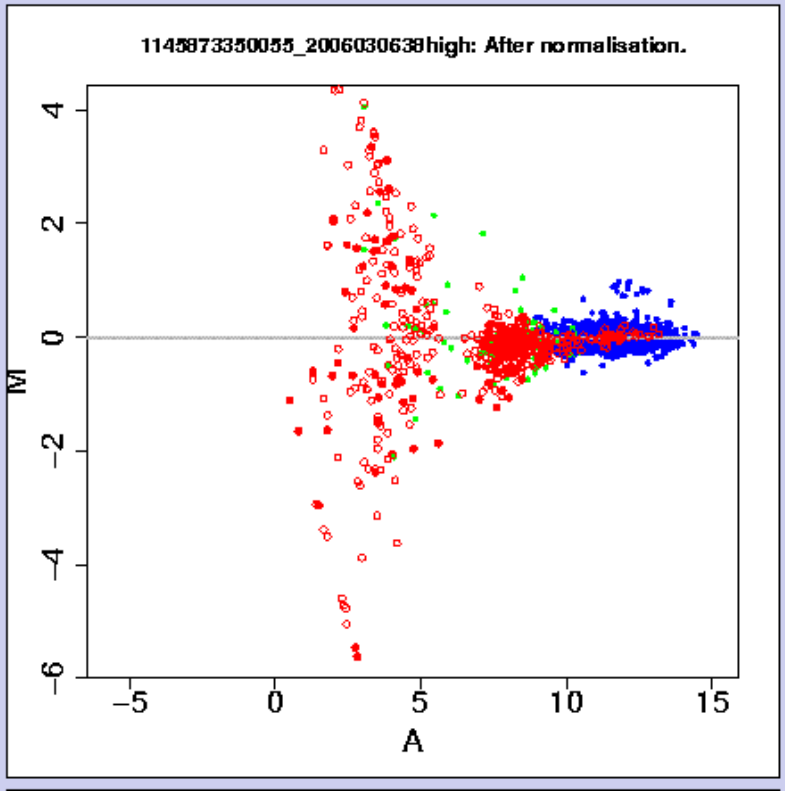
Hybridisation	# in both Channels	% in both channels	# in Red channel	% in Red channel	# in Green channel	% in Green channel
1145873350055_2006030638high	7237	97.903	7273	98.39	7255	98.147
1145873308564_2006022739low	7237	97.903	7252	98.106	7274	98.404
1145874721953_2006030638low	6931	93.764	6964	94.21	7008	94.805
For 282913 combined	7156	96.807	7203	97.443	7193	97.308
For 318719 combined	6855	92.735	6902	93.371	6944	93.939
For 79162 combined	6861	92.817	6925	93.682	6936	93.831



Spatial Loess normalization.

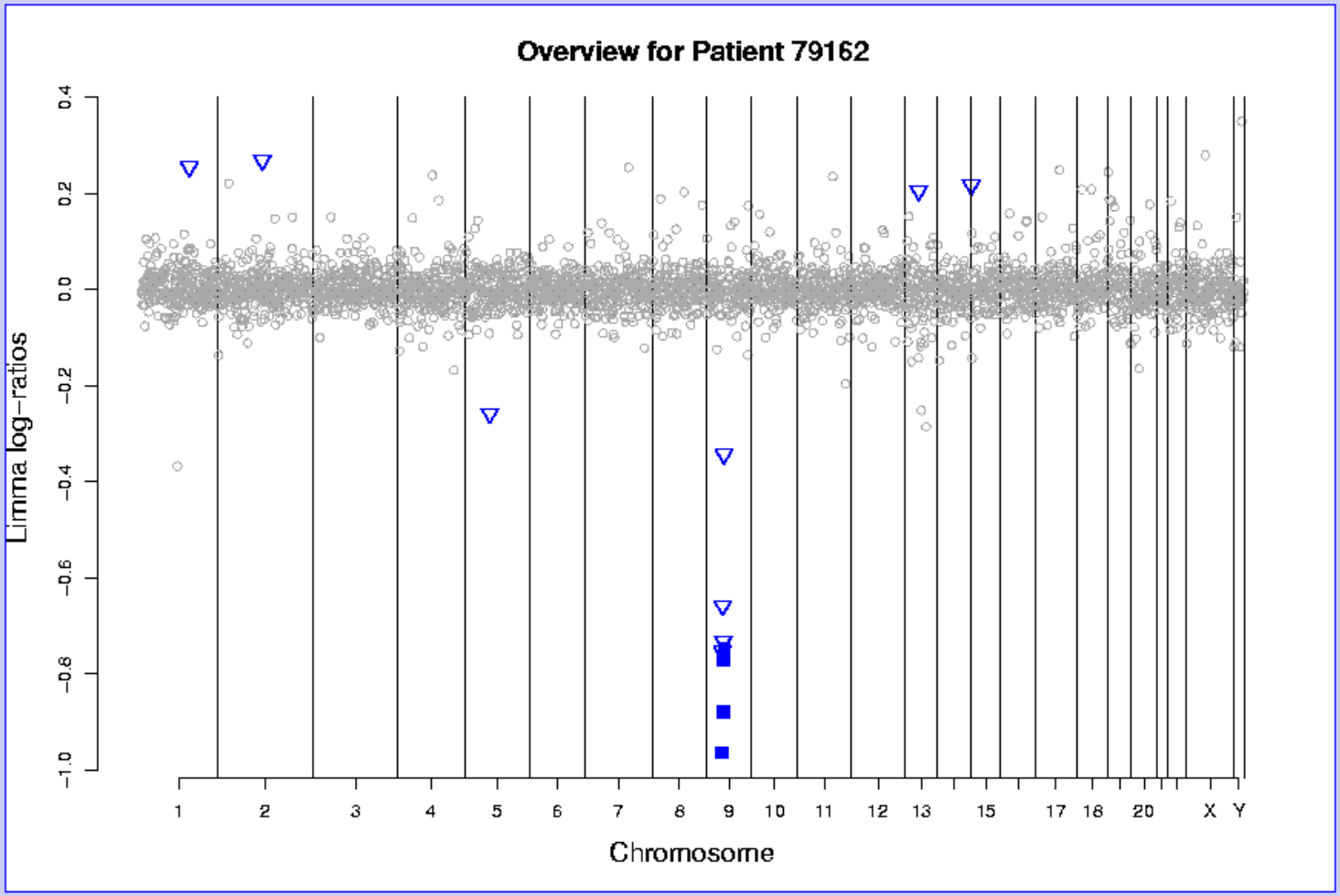
For all hybridisations, the MA plots are shown before and after the spatial Loess normalisation is performed. The measurements in the MA-plots are colored blue if they are above background and green if they are below. The empty spots are colored red.





The significant clones obtained with the Limma techniques:

This plot shows the average of the Limma-estimates for the different clones, ordered according to their chromosome location. Grey spots are not classified as deleted or duplicated for this specific patient. The colored clones are duplicated or deleted for patient 79162. In case they are likely to be partially deleted/duplicated, i.e. significantly below 0.86 or 0.54, respectively, in absolute value, they are drawn as a triangle.



Chapter 2 – Array CGH data management and analysis

- Raw data analysis : quality criteria and pitfalls
  - Data normalization
  - Spot and aberration calling (segmentation)
  - Threshold definition
- Alternative array design (loop analysis)
  - Joke Allemeersch credited for statistical models
  - LOOP tool: web architecture for data management
- **Benign copy number variation analysis (clone wars)**
- Importance of data management
  - Data repositories that feature analysis tools
  - Examples: ArrayCGHBase, Decipher, Store+Bench, Database of Genomic Variants, ECARUCA, …

Clone Profiler - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://tomcatbackup.esat.kuleuven.be/clonewars/?clonename=RP11-289D12&leftboundary=-

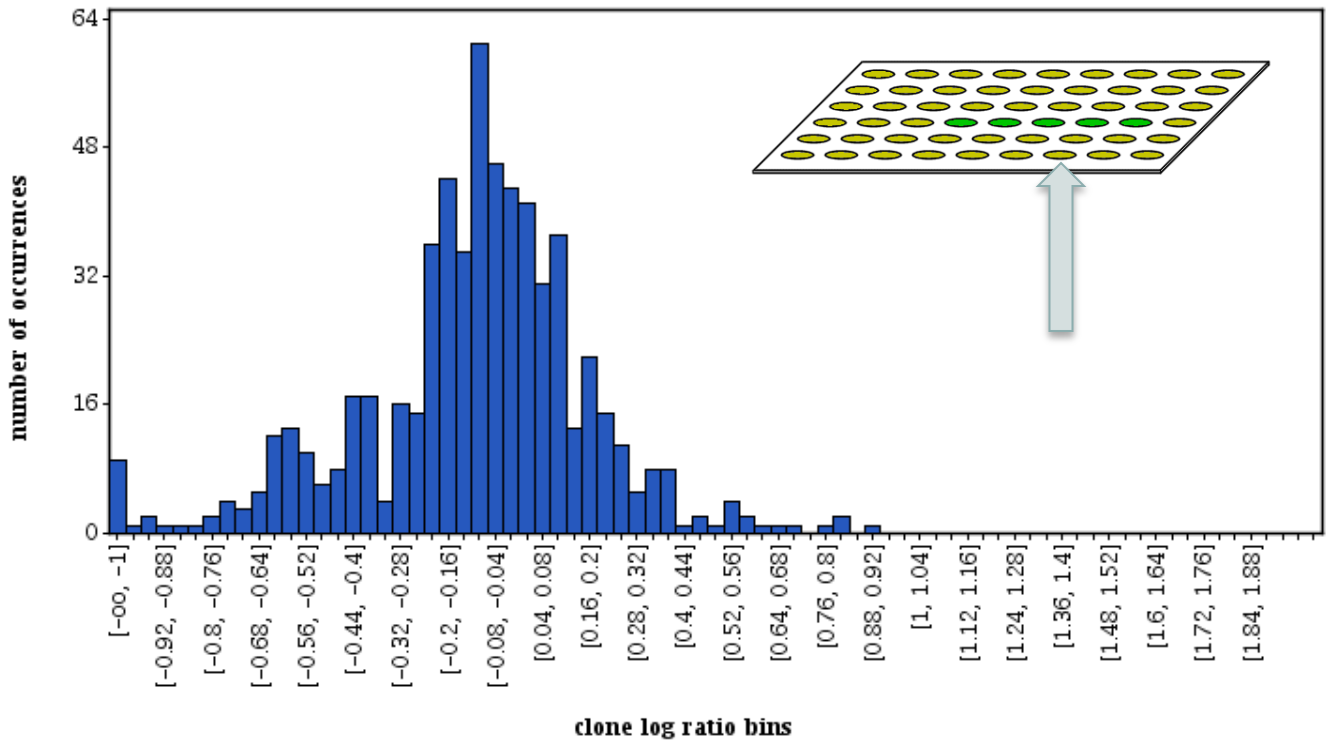
Google

Clone ID	RP11-289D12	(e.g. RP11-289D12)
Start log ratio	-1	(e.g. -1.0)
End log ratio	2	(e.g. 2.0)
Number of bins	75	(e.g. 100)

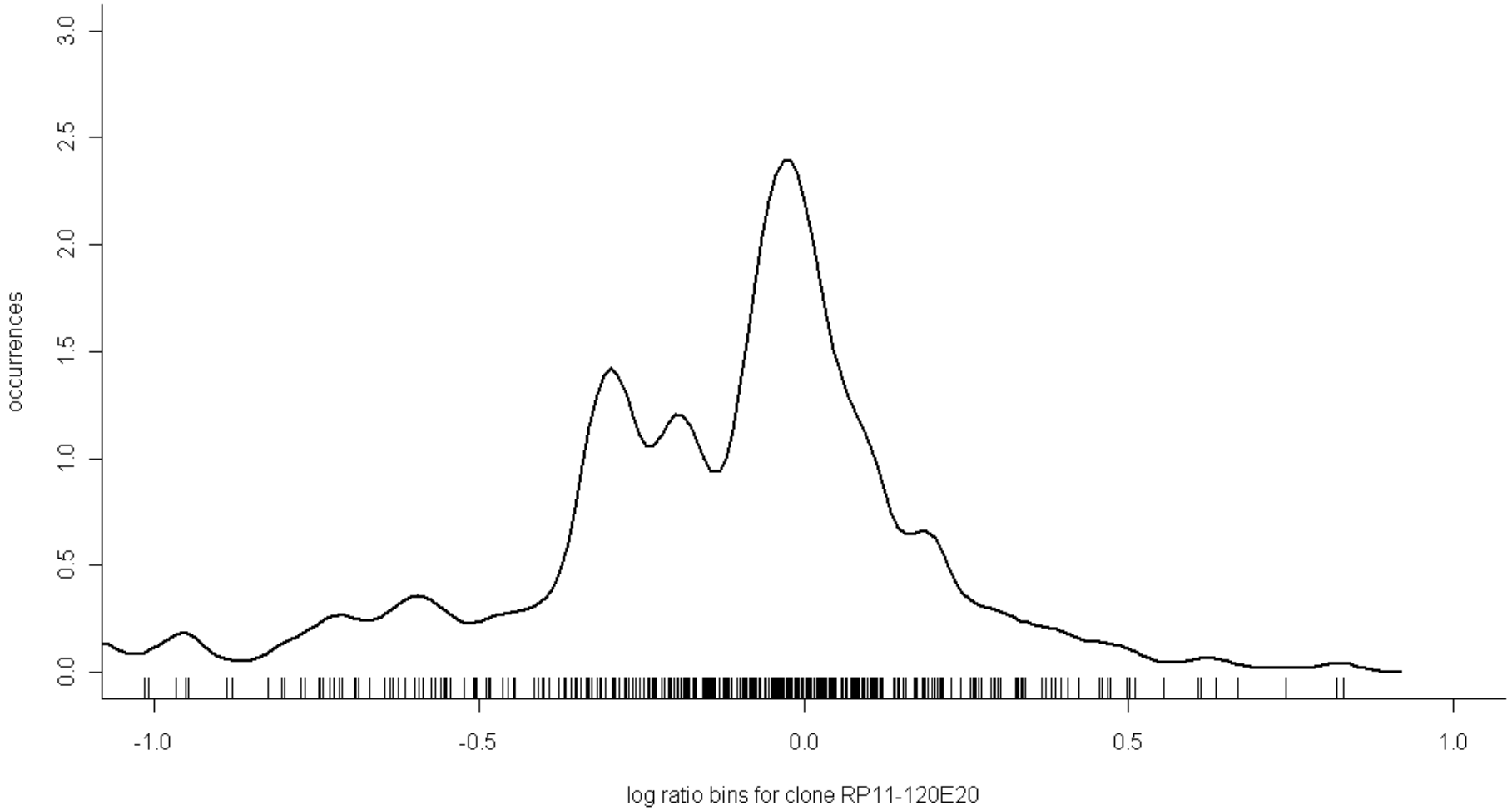
profile clone

Your input seems Ok.

Profile for clone [RP11-289D12]



Done


occurrences
0.0
0.5
1.0
1.5
2.0
2.5
3.0
-1.0
-0.5
0.0
0.5
1.0
log ratio bins for clone RP11-120E20

Publications

Menten B, Pattyn F, De Preter K, Robbrecht P, Michels E, Buysse K, Mortier G, De Paepe A, **Van Vooren S**, Vermeesch J, Moreau Y, De Moor B, Vermeulen S, Speleman F, Vandesompele J. **arrayCGHbase: an analysis platform for comparative genomic hybridization microarrays.** *BMC Bioinformatics.* 2005 May 23;6(1):124.

Vermeesch JR, Melotte C, Froyen G, **Van Vooren S**, Dutta B, Maas N, Vermeulen S, Menten B, Speleman F, De Moor B, Van Hummelen P, Marynen P, Fryns JP, Devriendt K. **Molecular karyotyping: array CGH quality criteria for constitutional genetic diagnosis.** *J Histochem Cytochem.* 2005 Mar;53(3):413-22.

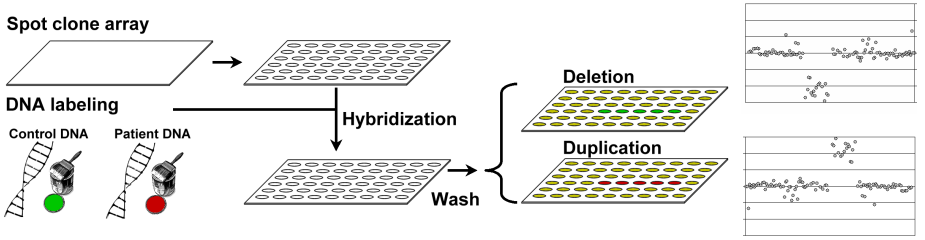
Publications

Maas N M C, **Van Vooren S**, Hannes F, Van Buggenhout G, Mysliwiec M, Moreau Y, Fagan K, Midro A, Engiz O, Balci, Parker M J, Sznajer Y, Devriendt K, Fryns J P, Vermeesch J R., **"The t(4;8) is mediated by homologous recombination between olfactory receptor gene clusters, but other 4p16 translocations occur at random"** Genet. Couns., 18(4):357-65, 2007.

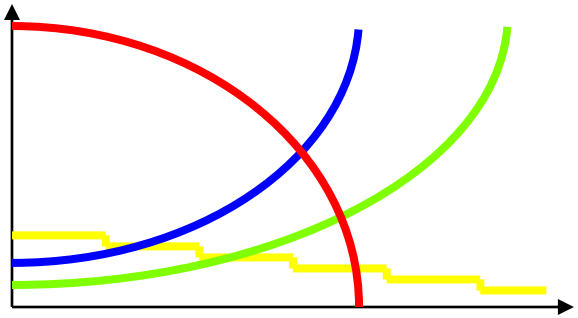
Menten B, Maas N, Thienpont B, Buysse K, Vandesompele J, Melotte C, de Ravel T, **Van Vooren S**, Balikova I, Backx L, Janssens S, De Paepe A, De Moor B, Moreau Y, Marynen P, Fryns JP, Mortier G, Devriendt K, Speleman F, Vermeesch J R., **"Emerging patterns of cryptic chromosomal imbalance in patients with idiopathic mental retardation and multiple congenital anomalies: a new series of 140 patients and review of published reports."** J Med Genet., 43(8):625-33, 2006.

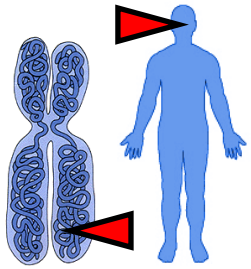
Chapter 3 – Functional interpretation of genetic information

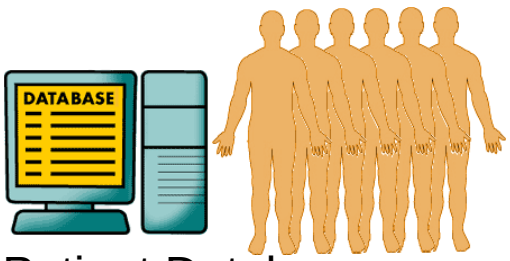
DOWNSTREAM ANALYSIS: INTERPRETATION IN CONTEXT OF GENE FUNCTION



DNA Microarray Experiments







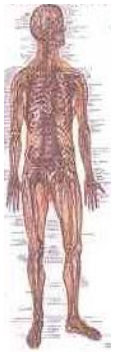
Patient Database

Biomed Literature

RP11-150A8
RP11-150B10
RP11-150C21
RP11-157P12
RP11-169B17
RP11-174G3
RP11-175H4
RP11-177E2
RP11-182I15
RP11-185P15
RP11-185H15
RP11-188O2
RP11-188O3
RP11-188O5
RP11-197I7
RP11-197B8
RP11-197A7
RP11-200K21
RP11-205J9
RP11-210N8
RP11-227G4
RP11-242A4
RP11-243O21
RP11-243H24
RP11-252D12
RP11-258F19
RP11-268E13
RP11-270L24
RP11-283N24
RP11-314M15

Patient Genotype

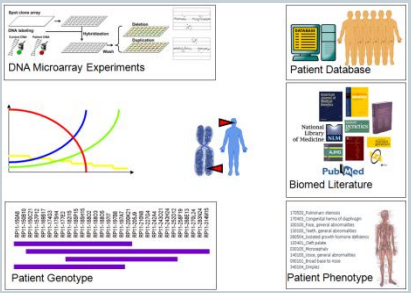
170520_Pulmonary stenosis
170403_Congenital hernia of diaphragm
100100_Face, general abnormalities
130100_Teeth, general abnormalities
280504_Isolated growth hormone deficiency
120401_Cleft palate
030105_Microcephaly
140100_Voice, general abnormalities
090101_Broad base to nose
340104_Dimples



Patient Phenotype

Chapter 3 – Functional interpretation of genetic information

- Downstream analysis: tools for interpretation
  - Association to phenotype and disease
    - Information from public databases and literature
    - Patient phenotype information
    - Other patients
  - Hunting for candidate genes
    - Computational approaches: a review
    - Literature as information source
    - Mapping concepts onto the genome
      - aBandApart
      - aGeneApart
    - Fusing multiple data sources
      - Endeavour



Chapter 3 – Functional interpretation of genetic information

- Downstream analysis: tools for interpretation
  - Association to phenotype and disease
    - Information from public databases and literature
    - Patient phenotype information
    - Other patients
  - Hunting for candidate genes
    - Computational approaches: a review
    - Literature as information source
    - **Mapping concepts onto the genome**
      - **aBandApart**
      - **aGeneApart**
    - Fusing multiple data sources
      - Endeavour

Rationale

- Brewer *et al.*:

A Chromosomal Duplication Map of Malformations: Regions of Suspected Haplo- and Triplolethality—and Tolerance of Segmental Aneuploidy—in Humans

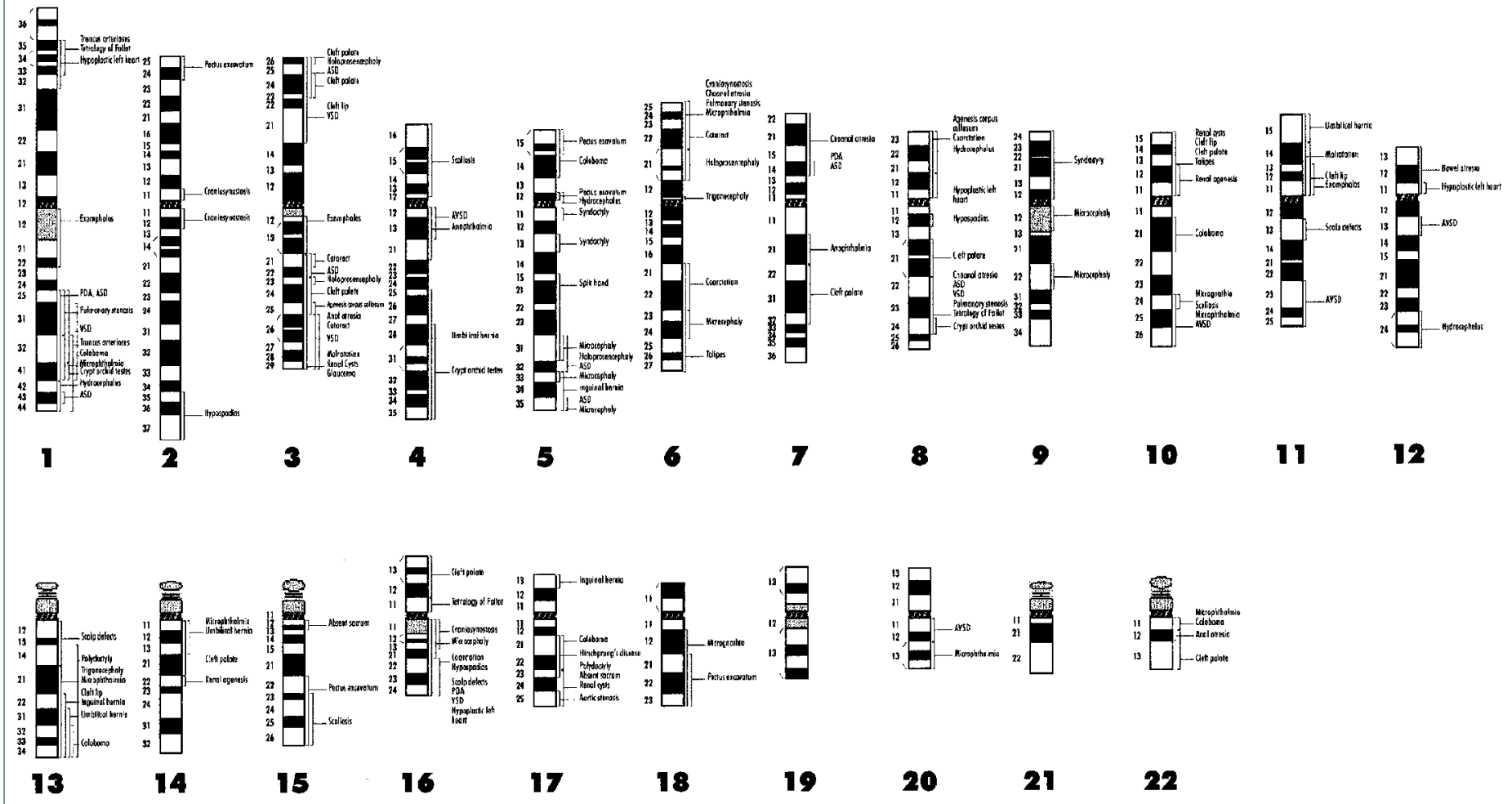
Carole Brewer,[1] Susan Holloway,[1] Paul Zawalnyski,[1] Albert Schinzel,[2] and David FitzPatrick[1]

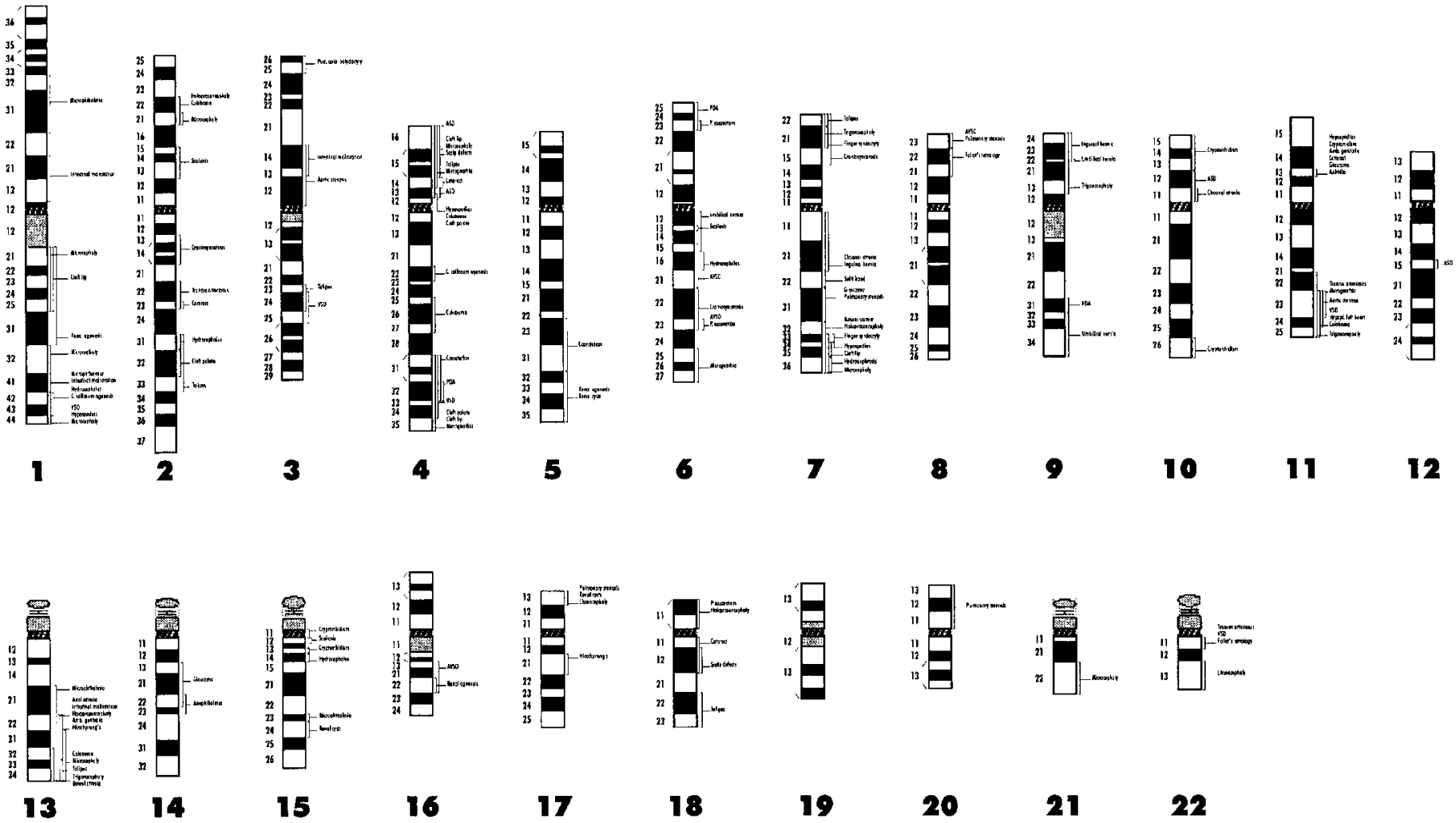
[1]Department of Human and Clinical Genetics, Molecular Medicine Centre, Western General Hospital, Edinburgh; and [2]Institute for Medical Genetics, University of Zurich, Zurich

A Chromosomal Deletion Map of Human Malformations

Carole Brewer,[1] Susan Holloway,[1] Paul Zawalnyski,[1] Albert Schinzel,[2] and David FitzPatrick[1]

[1]Department of Human and Clinical Genetics, MMC, Western General Hospital, Edinburgh; and [2]Institute for Medical Genetics, University of Zurich, Zurich


1
Truncus arteriosus
Tetralogy of Fallot
Hypoplastic left heart
Exomphalos
PDA, ASD
Pulmonary stenosis
VSD
Truncus arteriosus
Coloboma
Microphthalmia
Cryptorchidism testes
Hydrocephalus
ASD
2
Pectus excavatum
Craniosynostosis
Craniosynostosis
Hypospadias
3
Cleft palate
Holoprosencephaly
ASD
Cleft palate
Cleft lip
VSD
Exomphalos
Cataract
ASD
Holoprosencephaly
Cleft palate
Agenesis corpus callosum
Anal atresia
Cataract
VSD
Malrotation
Renal Cysts
Glaucoma
4
Scoliosis
AVSD
Anophthalmia
Umbilical hernia
Cryptorchidism
5
Pectus excavatum
Coloboma
Pectus excavatum
Hydrocephalus
Syndactyly
Syndactyly
Split hand
Microcephaly
Holoprosencephaly
ASD
Microcephaly
Inguinal hernia
ASD
Microcephaly
6
Craniosynostosis
Choanal atresia
Pulmonary stenosis
Microphthalmia
Cataract
Holoprosencephaly
Trigonocephaly
Coarctation
Microcephaly
Talipes
7
Choanal atresia
PDA
ASD
Anophthalmia
Cleft palate
8
Agenesis corpus callosum
Coarctation
Hydrocephalus
Hypoplastic left heart
Hypospadias
Cleft palate
Choanal atresia
ASD
VSD
Pulmonary stenosis
Tetralogy of Fallot
Cryptorchidism testes
9
Syndactyly
Microcephaly
Microcephaly
10
Renal cysts
Cleft lip
Cleft palate
Talipes
Renal agenesis
Coloboma
Micrognathia
Scoliosis
Microphthalmia
AVSD
11
Umbilical hernia
Malrotation
Cleft lip
Exomphalos
Scoto defects
AVSD
12
Bowel atresia
Hypoplastic left heart
AVSD
Hydrocephalus
13
Scoto defects
Polydactyly
Trigonocephaly
Microphthalmia
Cleft lip
Inguinal hernia
Umbilical hernia
Coloboma
14
Microphthalmia
Umbilical hernia
Cleft palate
Renal agenesis
15
Absent sternum
Pectus excavatum
Scoliosis
16
Cleft palate
Tetralogy of Fallot
Craniosynostosis
Microcephaly
Coarctation
Hypospadias
Scoto defects
PDA
VSD
Hypoplastic left heart
17
Inguinal hernia
Coloboma
Hirschsprung's disease
Polydactyly
Absent sternum
Renal cysts
Aortic stenosis
18
Micrognathia
Pectus excavatum
19
20
AVSD
Microphthalmia
21
22
Microphthalmia
Coloboma
Renal atresia
Cleft palate



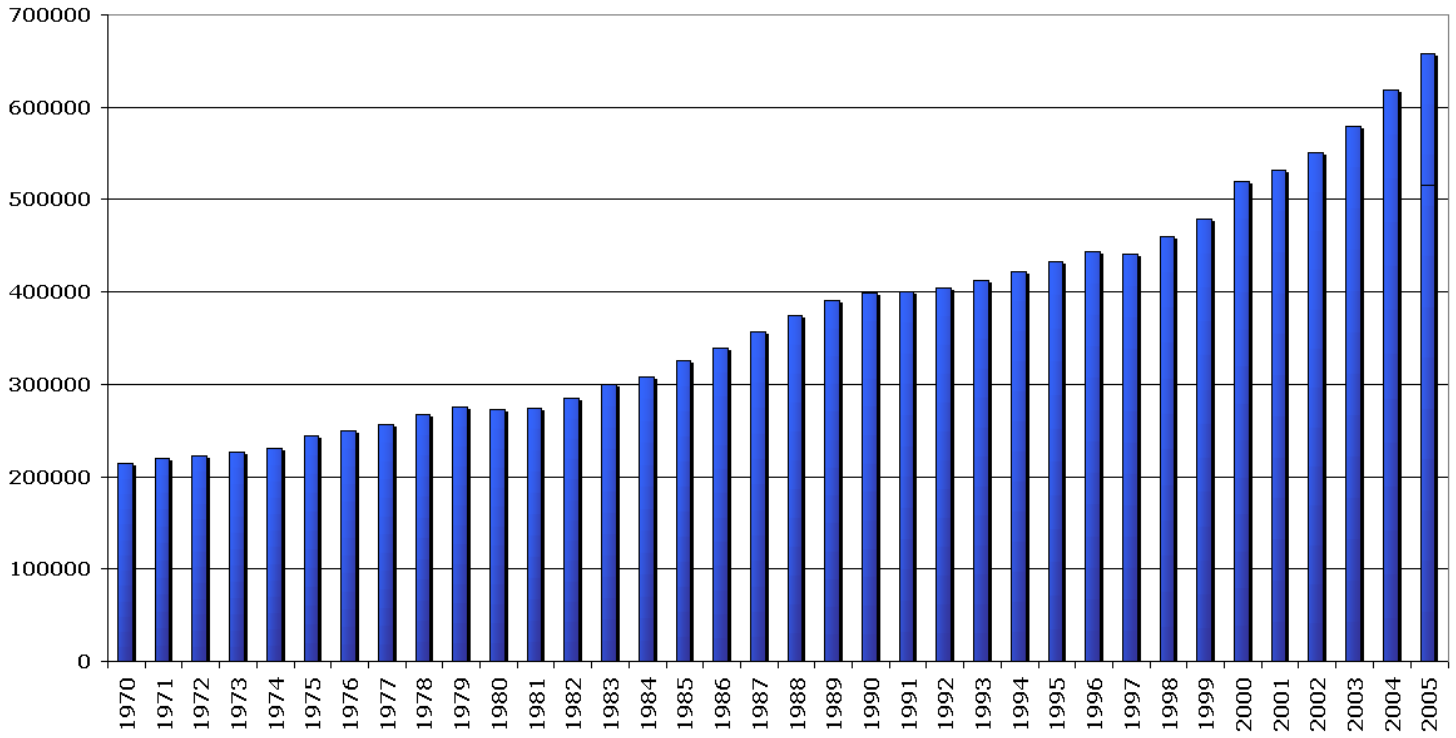
Key:

ASD	Atrial septal defect
AVSD	Atrio-ventricular septal defect
PDA	Patent ducts arteriosus
VSD	Ventricular septal defect

Rationale

- Public literature is evolving into a large phenotypic database
  - Case reports
  - Clinical research results
- Medline:
  - 4,800+ journals
  - 16 000 000+ abstracts
  - 1 185 000+ case reports
  - 500 000+ new each year

Rationale


700000
600000
500000
400000
300000
200000
100000
0
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005

Genomic Regions and Phenotypes

1: Prenat Diagn. 2005 Jun;25(6):451-5.

Related Articles, Links

Prenatal detection of a de novo terminal inverted duplication 4p in a fetus with the Wolf-Hirschhorn syndrome phenotype.

Beaujard MP, Jouannic JM, Bessieres B, Borie C, Martin-Luis I, Fallet-Bianco C, Portnoi MF.

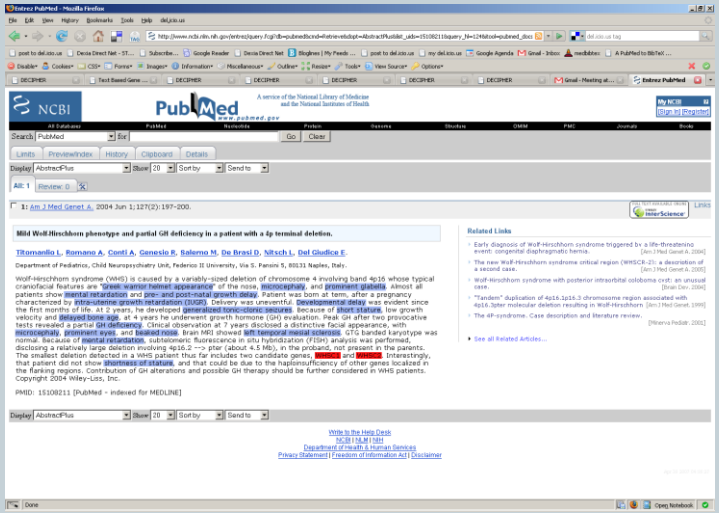
Laboratoire de Cytogenetique, Institut de Puericulture, Paris, France.

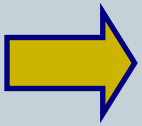
OBJECTIVES: To present the prenatal diagnosis of a de novo terminal inversion duplication of the short arm of chromosome 4 and a review of the literature. CASE: An amniocentesis for chromosome analysis was performed at 33 weeks' gestation because ultrasound examination showed a female fetus with multiple abnormalities consisting of severe intrauterine growth retardation, microcephaly, a cleft lip and renal hypoplasia. RESULTS: Cytogenetic analysis and FISH studies of the cultured amniocytes revealed a de novo terminal inversion duplication of the short arm of chromosome 4 characterized by a duplication of 4p14-p16.1 chromosome region concomitant with a terminal deletion 4p16.1-pter. The karyotype was thus: 46,XX, inv dup del (4)(:p14-->p16.1::p16.1-->qter). The parents opted to terminate the pregnancy. Fetopathological examination showed dysmorphic features and abnormalities consistent with a Wolf-Hirschhorn syndrome (WHS) diagnosis, clinical manifestations of partial 4p trisomy being mild. CONCLUSION: Although relatively rare, inverted duplications have been reported repeatedly in an increasing number of chromosomes. Only two previous cases with de novo inv dup del (4p) and one with tandem dup 4p have been reported, all of them associated with a 4pter deletion. We report the first case diagnosed prenatally. Breakpoints are variable, resulting in different abnormal phenotype. In our case, clinical manifestations resulted in a WHS phenotype.

Publication Types:
- Case Reports

PMID: 15966060 [PubMed - indexed for MEDLINE]

Selecting Text Features



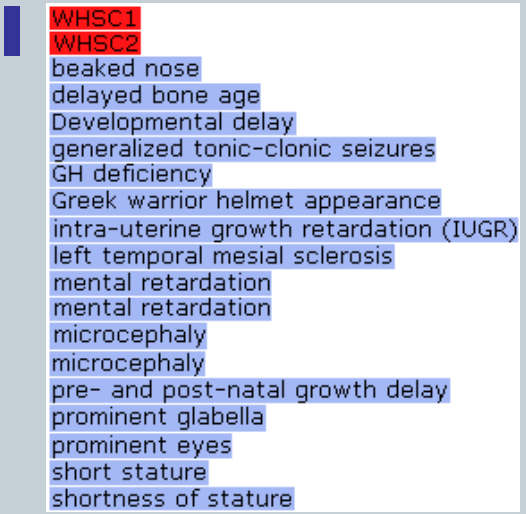


WHSC1
WHSC2
beaked nose
delayed bone age
Developmental delay
generalized tonic-clonic seizures
GH deficiency
Greek warrior helmet appearance
intra-uterine growth retardation (IUGR)
left temporal mesial sclerosis
mental retardation
mental retardation
microcephaly
microcephaly
pre- and post-natal growth delay
prominent glabella
prominent eyes
short stature
shortness of stature

Selecting Text Features

gene/protein	UniProt ID	Frequency in BNC
Had	Q9WVK7	402433
who	O01367	194312
how	O01367	94873
Last	Q9D032	73553
put	Q51841	59248
Yes	Q04736	58451
Set	Q63945	48232
Great	Q91ZZ5	45250
car	Q9Y1I2	34155

Source: EBI Rebholz group

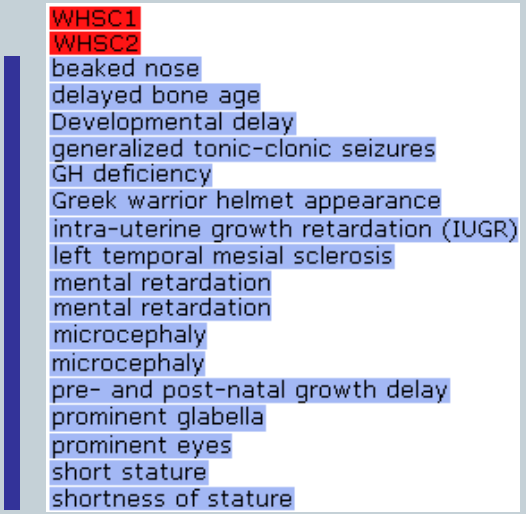


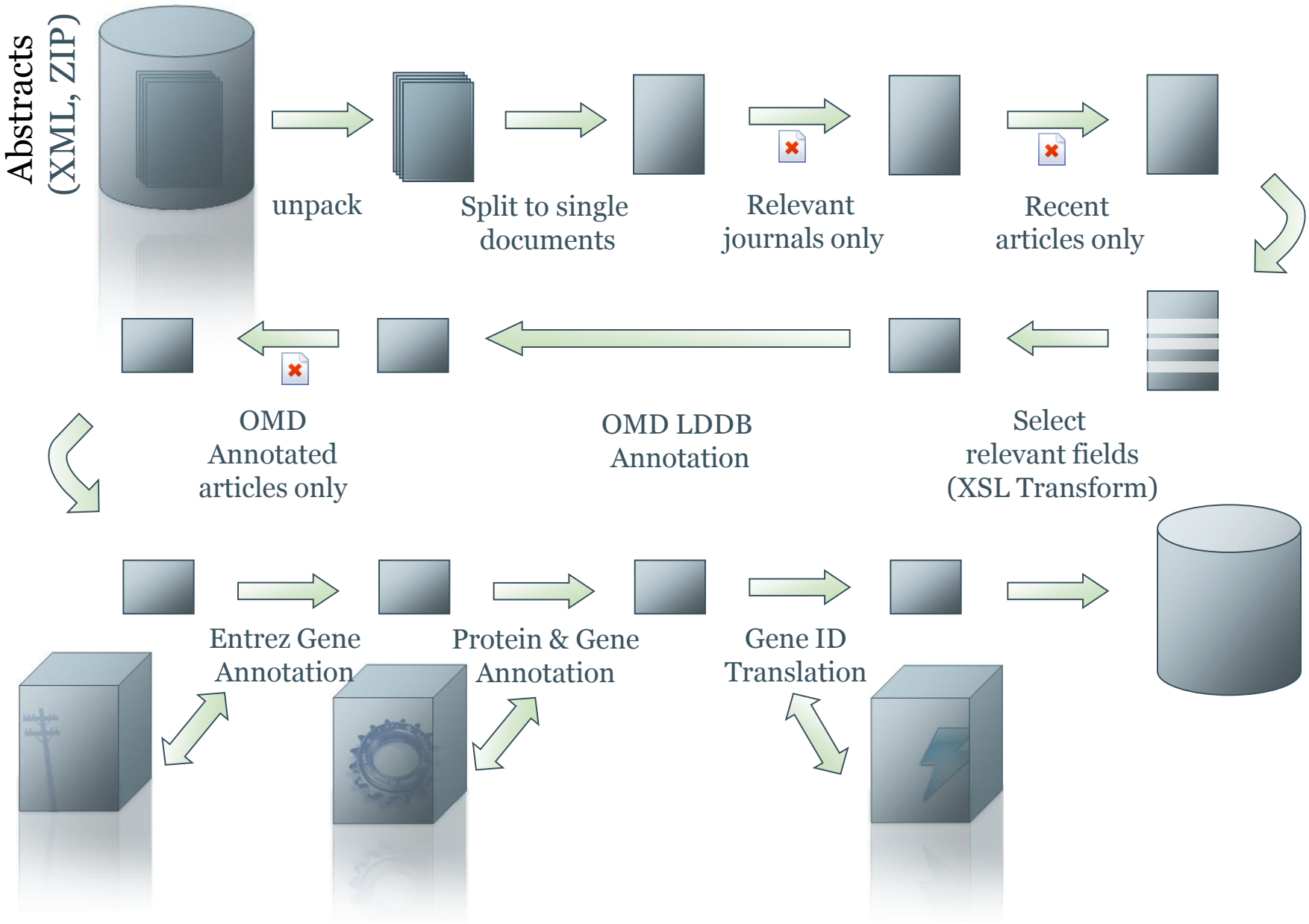
Selecting Text Features

included **beak nose**, microcephaly, and
and the **nose was beaked**. Further phen

showed **small head circumference**. In o
were **microcephalic**. To delineate a co
displaying **microcephaly**. In an earlie

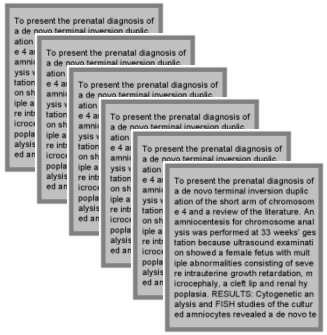
in **mentally retarded** patients with an
behaviour, **mental retardation**, and of




Abstracts
(XML, ZIP)
unpack
Split to single documents
Relevant journals only
Recent articles only
Select relevant fields (XSL Transform)
OMD LDDB Annotation
OMD Annotated articles only
Entrez Gene Annotation
Protein & Gene Annotation
Gene ID Translation

Gene to Concept associations

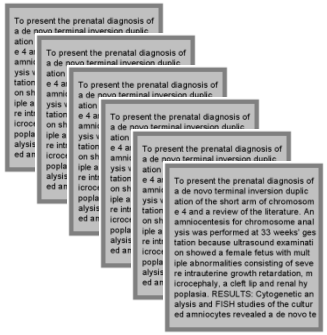
ENSG00000000001
ENSG00000000002
. . .
WHSC1 **ENSG00000109685**
. . .
ENSG00000024999
ENSG00000025000



PubMed

Gene to Concept associations

ENSG00000000001
ENSG00000000002
...
WHSC1 **ENSG00000109685**
...
ENSG00000024999
ENSG00000025000

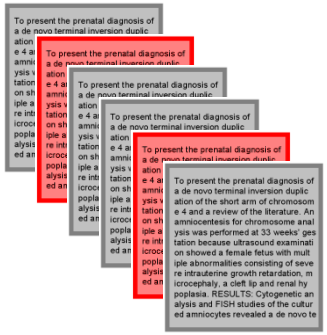


Microcephaly

PubMed

Gene to Concept associations

ENSG00000000001
ENSG00000000002
. . .
WHSC1 **ENSG00000109685**
. . .
ENSG00000024999
ENSG00000025000

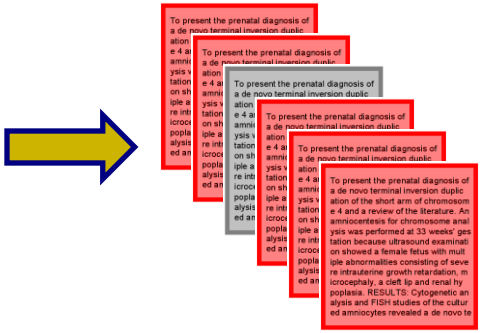


Microcephaly

PubMed

Gene to Concept associations

ENSG00000000001
ENSG00000000002
...
WHSC1 **ENSG00000109685**
...
ENSG00000024999
ENSG00000025000



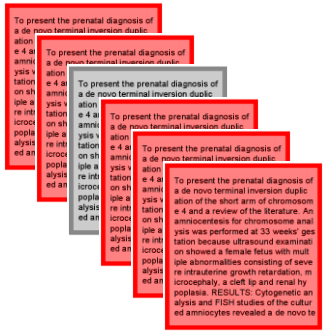
Microcephaly

overrepresented in document set for WHSC1 gene

PubMed

Gene to Concept associations

ENSG00000000001
ENSG00000000002
...
WHSC1 **ENSG00000109685**
...
ENSG00000024999
ENSG00000025000



$$p_{bc} = 1 - H_{\text{cdf}}(O_{bc}|A,B,C)$$

$$= 1 - \sum_{i=0}^{O_{bc}-1} \frac{\binom{B}{i}\binom{A-B}{C-i}}{\binom{A}{C}}$$

$$= \sum_{i=O_{bc}}^{\min(B,C)} \frac{\binom{B}{i}\binom{A-B}{C-i}}{\binom{A}{C}}$$

Gene = **WHSC1**
Concept = **microcephaly**

A (total abstracts) = 800.000
B (gene abstracts) = 30
C (concept abstracts) = 4000
O_{bc} (gene & concept) = 10

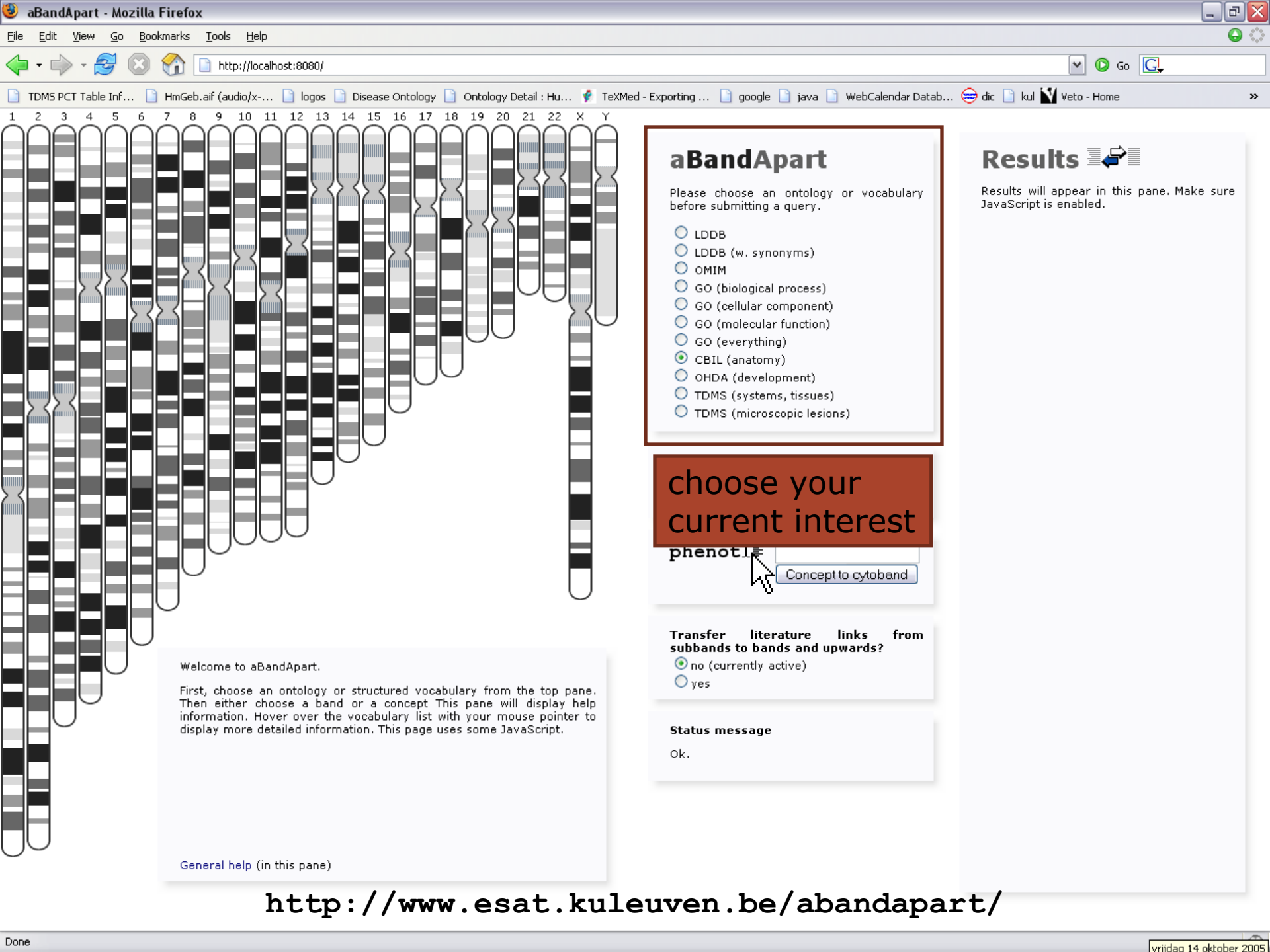
p_{bc} (*p*-value) = 1,57 x 10^{-6}

Gene to Concept associations

Results

WHSC2

concept	# abstr	*p*-value
32.04.02 (Mental retardation / developmental delay)	3	1.2445E-5
32.08.05 (Microcephaly)	2	1.5854E-5
32.06.00 (SEIZURES, general abnormalities)	3	2.2494E-5
05.01.01 (Prominent glabella)	1	4.2313E-5
09.01.07 (Convex / beaked profile of nose)	1	9.5199E-5
02.00.00 (STATURE)	2	1.7425E-4
32.08.00 (CRANIUM, general abnormalities)	2	3.5081E-4
32.04.00 (MENTAL,COGNITIVE FUNCTION, general abnormalities)	3	4.7173E-4
14.01.07 (Speech delay)	1	5.3936E-4
07.01.07 (Prominent eyes / proptosis)	1	5.3936E-4
33.01.11 (Delayed bone age)	1	6.6623E-4
05.01.00 (Forehead, prominent etc., general abnormalities)	1	1.0573E-3
32.06.14 (Tonic seizures)	1	1.987E-3
05.00.00 (FOREHEAD)	1	2.4409E-3
01.03.02 (Low birthweight (< 3rd centile))	1	3.4749E-3
01.03.00 (Thin or slender build, general abnormalities)	1	3.496E-3
14.00.00 (VOICE)	1	5.9083E-3
14.01.00 (Voice, general abnormalities)	1	5.9083E-3
07.01.00 (Eyes, general abnormalities (including spacing))	1	6.7498E-3
32.12.00 (CRANIAL BONES, general abnormalities)	1	8.2839E-3
32.04.05 (Hypotonia (non-myopathic))	1	8.6618E-3
02.01.00 (Short stature, general abnormalities)	1	1.6507E-2
32.00.00 (NEUROLOGY)	6	1.7382E-2
09.01.00 (Nose, general abnormalities)	1	2.0678E-2
09.00.00 (NOSE)	1	2.1062E-2
18.00.00 (ABDOMEN)	3	4.93E-2
01.00.00 (BUILD)	1	7.4564E-2
18.05.01 (Abnormal liver (including function))	2	8.3254E-2
18.05.00 (Liver / biliary system, general abnormalities)	2	8.4474E-2
28.05.04 (Isolated growth hormone deficiency)	1	1.0751E-1
18.03.00 (Colon, general abnormalities)	1	1.2606E-1
17.02.00 (Breasts, general abnormalities)	1	1.7483E-1
33.01.00 (Skeleton, general abnormalities)	1	1.7495E-1
33.00.00 (SKELETAL SYSTEM)	1	1.7691E-1
28.05.00 (Pituitary, general abnormalities)	1	1.8627E-1
07.00.00 (EYES, GLOBES)	1	2.9781E-1
17.00.00 (THORAX)	1	5.2725E-1
28.00.00 (ENDOCRINE)	1	6.5493E-1

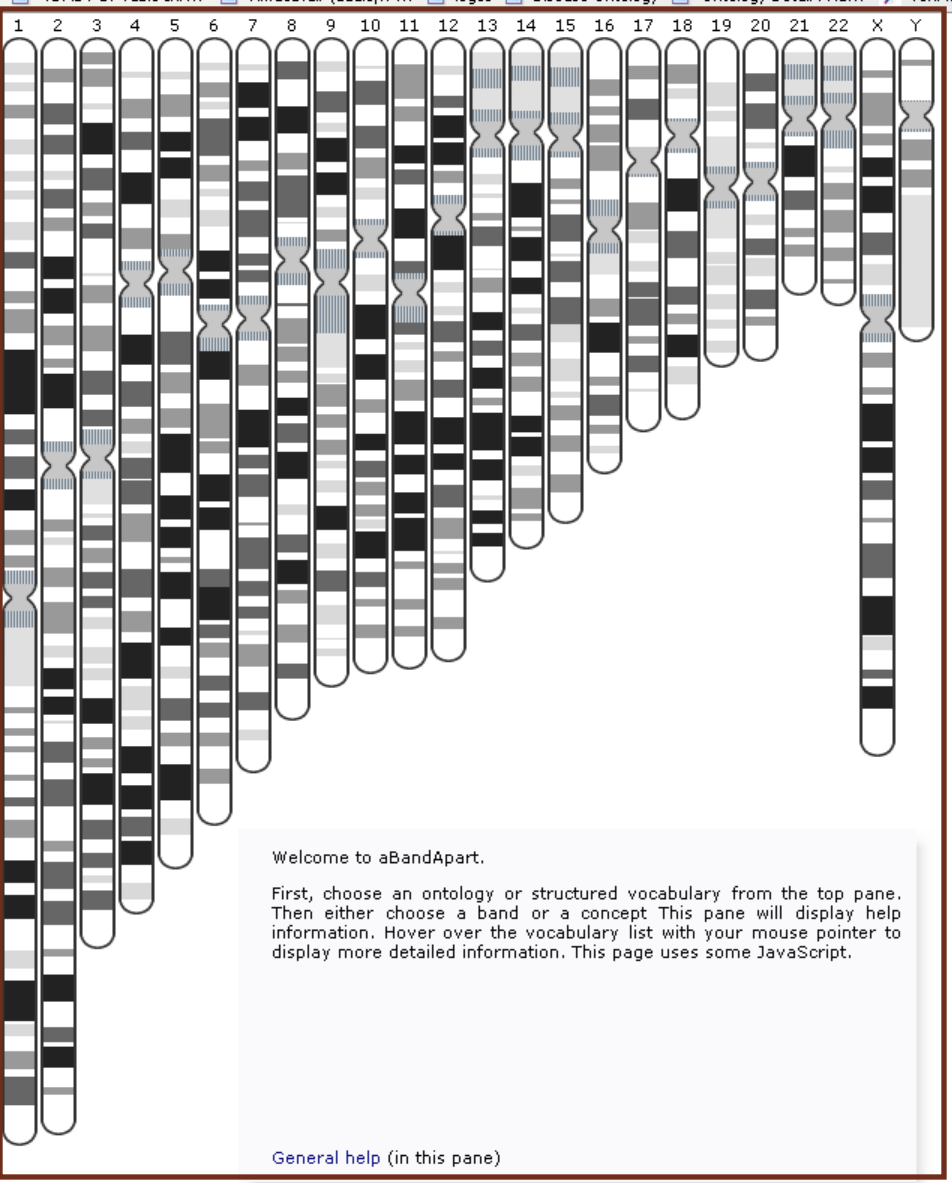

aBandApart - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://localhost:8080/
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y
aBandApart
Please choose an ontology or vocabulary before submitting a query.
LDDB
LDDB (w. synonyms)
OMIM
GO (biological process)
GO (cellular component)
GO (molecular function)
GO (everything)
CBIL (anatomy)
OHDA (development)
TDMS (systems, tissues)
TDMS (microscopic lesions)
Results
Results will appear in this pane. Make sure JavaScript is enabled.
choose your current interest
phenotl
Concept to cytoband
Transfer literature links from subbands to bands and upwards?
no (currently active)
yes
Status message
Ok.
Welcome to aBandApart.
First, choose an ontology or structured vocabulary from the top pane. Then either choose a band or a concept This pane will display help information. Hover over the vocabulary list with your mouse pointer to display more detailed information. This page uses some JavaScript.
General help (in this pane)
http://www.esat.kuleuven.be/abandapart/
Done

aBandApart - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://localhost:8080/

TDMS PCT Table Inf... HmGeb.aif (audio/x-... logos Disease Ontology Ontology Detail : Hu... TeXMed - Exporting ... google java WebCalendar Datab... dic kul Veto - Home



Welcome to aBandApart.

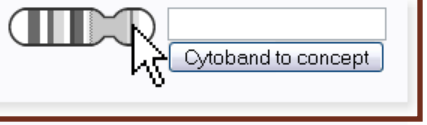
First, choose an ontology or structured vocabulary from the top pane. Then either choose a band or a concept This pane will display help information. Hover over the vocabulary list with your mouse pointer to display more detailed information. This page uses some JavaScript.

General help (in this pane)

aBandApart

Please choose an ontology or vocabulary before submitting a query.

- LDDB
- LDDB (w. synonyms)
- OMIM
- GO (biological process)
- GO (cellular component)
- GO (molecular function)
- GO (everything)
- CBIL (anatomy)
- OHDA (development)
- TDMS (systems, tissues)
- TDMS (microscopic lesions)



start from a certain cytoband

subbands to bands and upwards?

- no (currently active)
- yes

Status message

Ok.

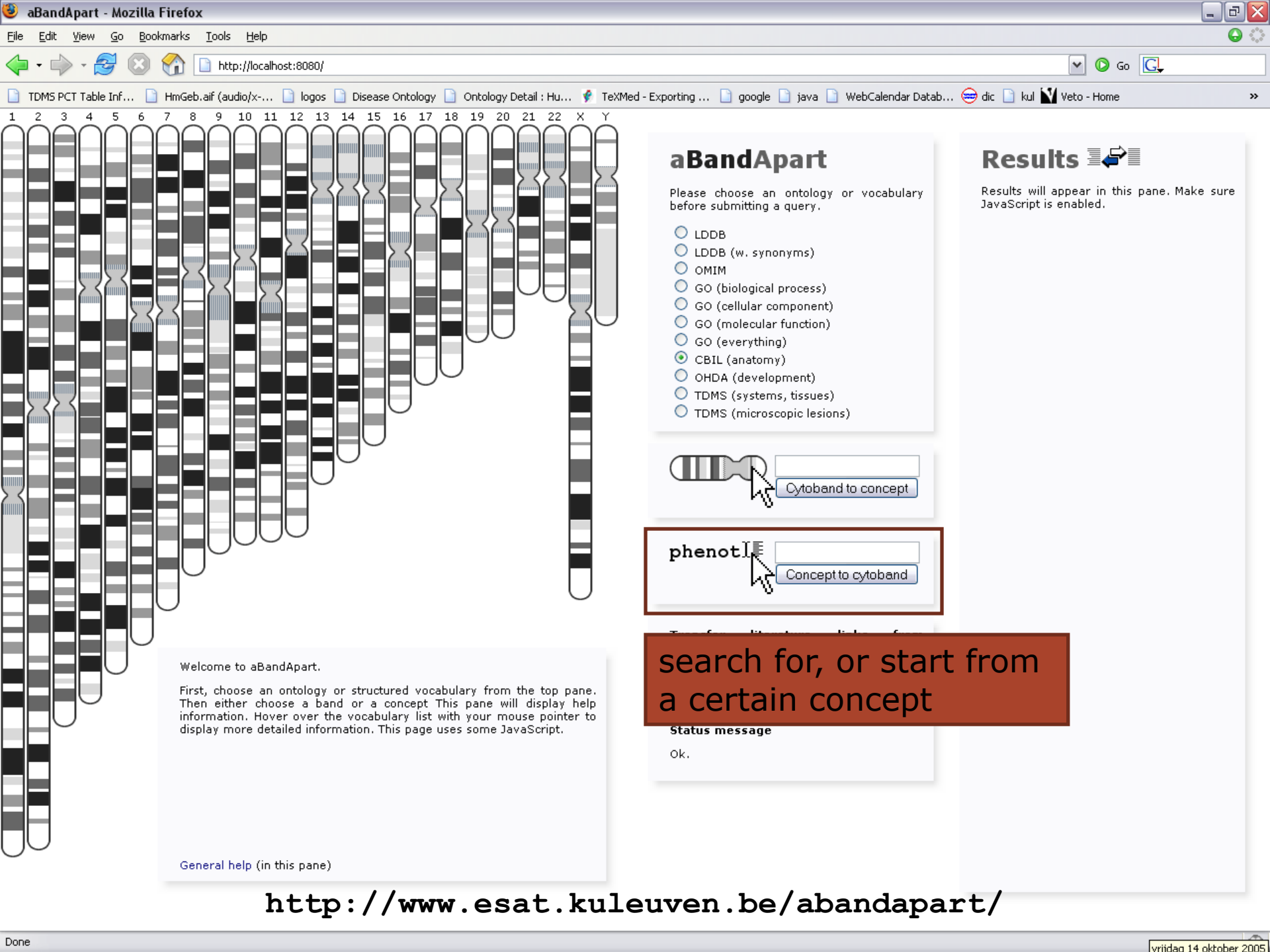
Results

Results will appear in this pane. Make sure JavaScript is enabled.

http://www.esat.kuleuven.be/abandapart/

Done

vrijdag 14 oktober 2005

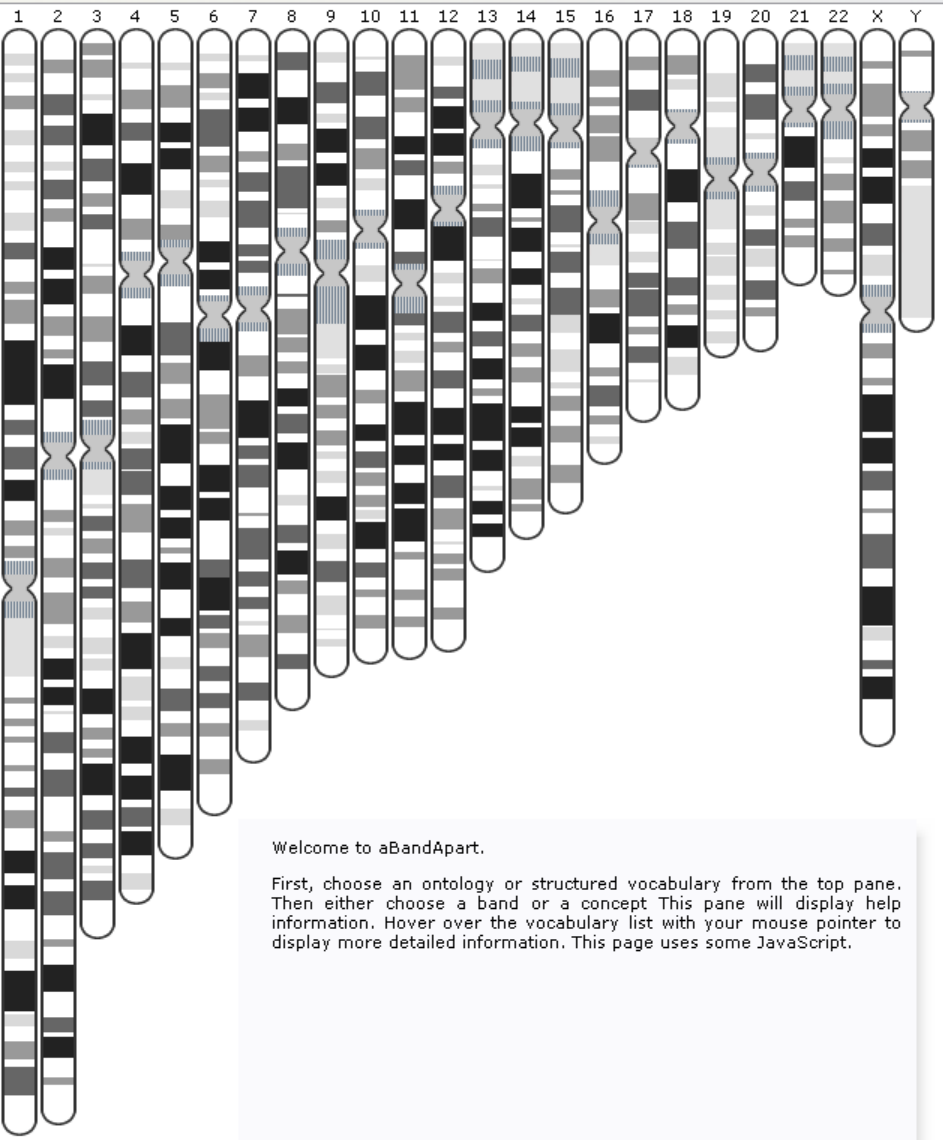

aBandApart - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://localhost:8080/
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y
aBandApart
Please choose an ontology or vocabulary before submitting a query.
LDDB
LDDB (w. synonyms)
OMIM
GO (biological process)
GO (cellular component)
GO (molecular function)
GO (everything)
CBIL (anatomy)
OHDA (development)
TDMS (systems, tissues)
TDMS (microscopic lesions)
Cytoband to concept
phenotI
Concept to cytoband
Results
Results will appear in this pane. Make sure JavaScript is enabled.
Welcome to aBandApart.
First, choose an ontology or structured vocabulary from the top pane. Then either choose a band or a concept This pane will display help information. Hover over the vocabulary list with your mouse pointer to display more detailed information. This page uses some JavaScript.
General help (in this pane)
Status message
Ok.
search for, or start from a certain concept
http://www.esat.kuleuven.be/abandapart/
Done

aBandApart - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://tomcatbackup.esat.kuleuven.be:8080/abandapart/

TDMS PCT Table Inf... HmGeb.aif (audio/x-... logos Disease Ontology Ontology Detail : Hu... TeXMed - Exporting ... google java WebCalendar Datab... dic kul Veto - Home



Welcome to aBandApart.

First, choose an ontology or structured vocabulary from the top pane. Then either choose a band or a concept This pane will display help information. Hover over the vocabulary list with your mouse pointer to display more detailed information. This page uses some JavaScript.

General help (in this pane)

aBandApart

Please choose an ontology or vocabulary before submitting a query.

- GO (biological process)
- GO (cellular component)
- GO (molecular function)
- GO (everything)
- (•) LDDB
- LDDB (w. synonyms)
- OMIM
- CBIL (anatomy)
- OHDA (development)
- TDMS (systems, tissues)
- TDMS (microscopic lesions)


4p16.3 Cytoband to concept

phenotI Concept to cytoband

Status message

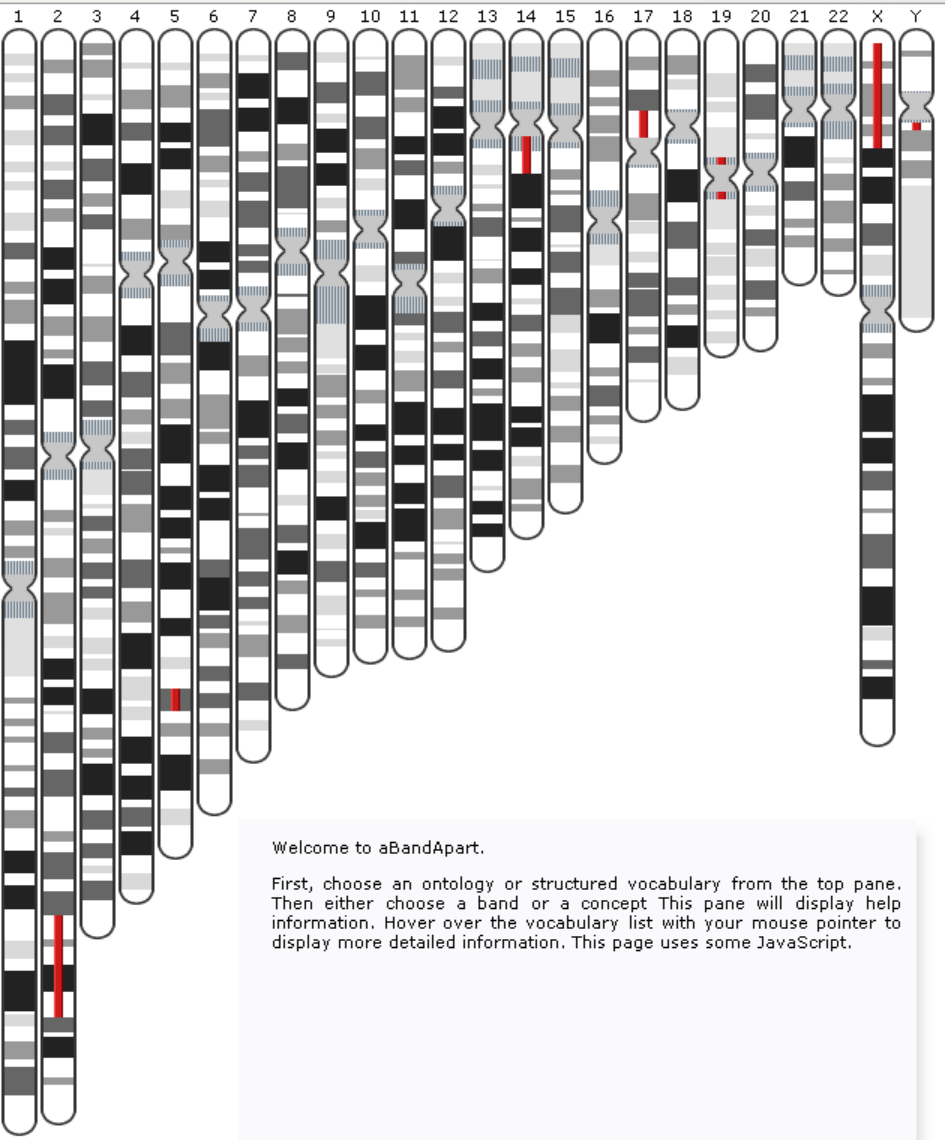
Ok. Band requested. Input validated. looking for **4p16.3**...

Band found.

Results

concept	*p*-value
wolf hirschhorn	0E0
whs	0E0
cherubism	2.6298E-10
maxilla	2.6213E-5
proptosis	1.6505E-4
macrocephaly	1.8656E-4
growth	2.7928E-4
hearing	3.271E-4
mandible	6.5273E-4
seizures	6.6031E-4
deep set eyes	1.6966E-3
basal ganglia	1.7647E-3
microcephaly	1.932E-3
ureters	4.5693E-3
back	6.6132E-3
short stature	8.5252E-3
lytic bone lesions	9.1179E-3
lordosis	9.1179E-3
choanal stenosis	9.1179E-3
mental retardation developmental delay	1.8153E-2
tremors	1.8153E-2
full cheeks	1.8153E-2
developmental delay	1.9372E-2
dementia	1.9485E-2
wolf syndrome	2.264E-2
globe	2.7106E-2
night blindness	3.1409E-2
respiratory infections	3.1552E-2
high nasal bridge	3.5978E-2
small head	3.5978E-2
otitis media	3.5978E-2
inguinal hernia	3.5978E-2
hand	3.6091E-2
hallucinations	4.0383E-2
deafness	4.0878E-2
growth retardation	4.6672E-2
low set ears	4.8071E-2
cerebral atrophy	4.9135E-2
femur	4.9135E-2
nystagmus	5.2642E-2
growth delay	6.2113E-2
finger	6.3372E-2
nose	6.5953E-2
beaked nose	6.64E-2
mental retardation	6.9843E-2
renal dysplasia	7.0667E-2
microcornea	7.4915E-2
severe growth retardation	7.9143E-2
feeding difficulties	7.9143E-2
hands	8.4296E-2
bladder	8.8269E-2
speech delay	9.5865E-2
eyes	9.6805E-2
ascites	9.9998E-2
retardation	1.0361E-1
breast tumour	1.0411E-1
facial dysmorphism	1.1127E-1

Done



Welcome to aBandApart.

First, choose an ontology or structured vocabulary from the top pane. Then either choose a band or a concept This pane will display help information. Hover over the vocabulary list with your mouse pointer to display more detailed information. This page uses some JavaScript.

General help (in this pane)

aBandApart

Please choose an ontology or vocabulary before submitting a query.

- GO (biological process)
- GO (cellular component)
- GO (molecular function)
- GO (everything)
- LDDB
- LDDB (w. synonyms)
- OMIM
- CBIL (anatomy)
- OHDA (development)
- TDMS (systems, tissues)
- TDMS (microscopic lesions)

4p16.3

Cytoband to concept

phenotI

ichthyosis

Concept to cytoband

Status message

Ok. Concept requested. Concept validated. looking for **ichthyosis**... Concept found.

Results

band	*p*-value
xp22	3.5305E-14
xp22.3	8.1934E-14
xp22.31	9.1351E-8
2q34	1.3124E-7
14q11	2.1914E-7
xpter	2.6043E-7
2q33.1	2.3192E-6
2q33.2	2.6074E-6
xp22.32	2.7628E-6
2q33.3	2.923E-6
2q35	1.0681E-5
2q33	3.027E-5
xptel	1.0491E-4
xp22.33	1.1846E-4
17p11.2	2.5604E-3
17p11	3.8756E-3
xp223	4.9231E-3
xp22.2	1.4125E-2
yq11.21	1.8337E-2
19cen	2.4375E-2
19p11	2.7979E-2
19q11	2.9178E-2
5q32	4.4136E-2
yq11.2	5.5175E-2
19p12	5.9827E-2
xq28	6.2565E-2
12q11	6.4456E-2
xq27.3	6.6007E-2
19p13.13	7.1356E-2
19p13.11	7.1356E-2
21q22.11	7.3645E-2
19p13.12	7.3645E-2
19q12	1.128E-1
xp11	1.1313E-1
12q12	1.2151E-1
xq27	1.2853E-1
19p13.1	1.3014E-1
yq11	1.3442E-1
12q13	1.3891E-1
xp11.22	1.4398E-1
14q	1.4925E-1
19p13.2	1.5029E-1
9q34.13	1.6175E-1
9q34.12	1.6175E-1
9q34.11	1.6175E-1
21q22.1	1.6278E-1
xp11.23	1.7101E-1
18q21.3	1.8721E-1
xqtel	1.8921E-1
9q34.1	1.9718E-1
1q21	1.9748E-1
3q28	2.0113E-1
9q33.3	2.0211E-1
3q27.2	2.031E-1
3q27.1	2.031E-1
3q27.3	2.0506E-1
xqter	2.2252E-1

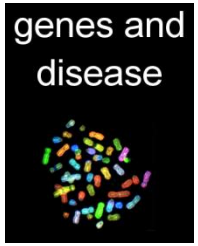
aBandApart

Linking genotypes and phenotypes

- The result is like *OMIM*, but
  - automated i.o. manual
  - a lot of vocabularies
- Vocabularies provide different views
- 10 vocabularies are available

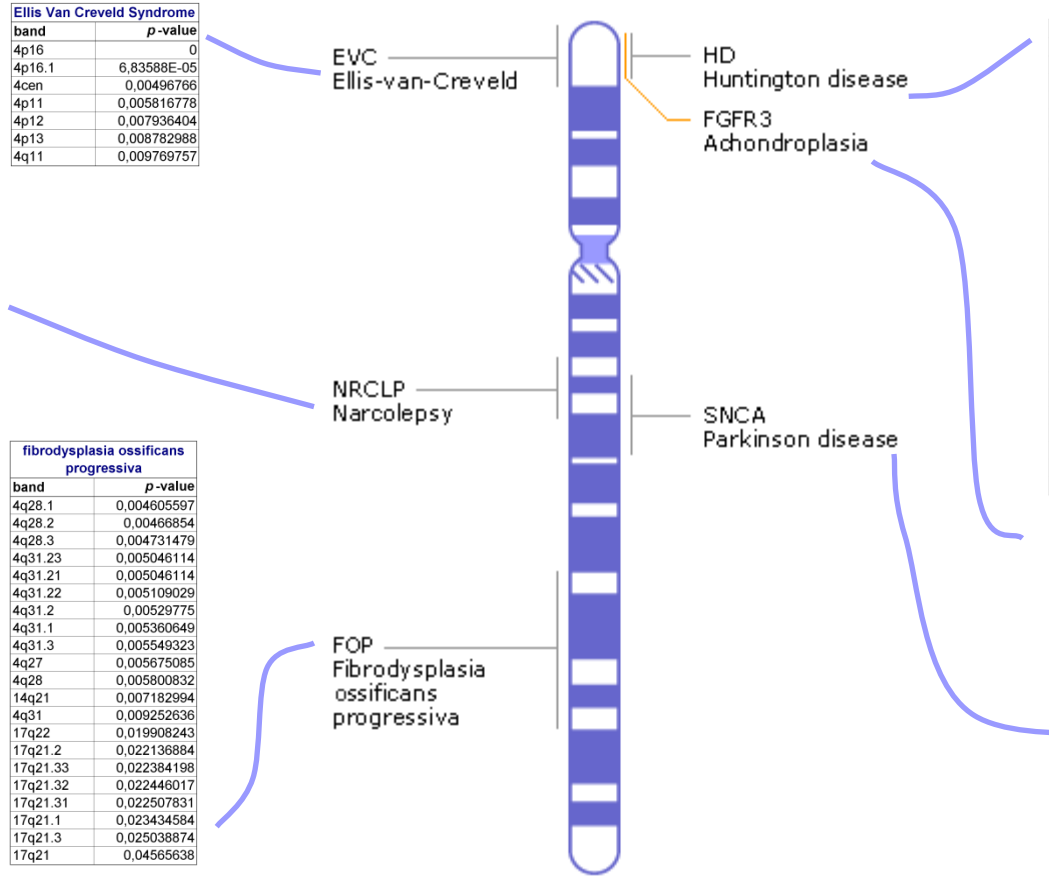
Validation

narcolepsy	
band	*p*-value
4cen	0,002762778
4p11	0,00323563
6cen	0,003471972
6q11.2	0,003944488
4p12	0,00441678
6q11.1	0,004574161
4p13	0,004888848
6q11	0,005046154
4q11	0,005439311
6p11.1	0,005832312
6p11.2	0,006068038
6q12	0,006617849
6p11	0,007167355
4q13.2	0,00779499
4q13.1	0,008030251
4q13.3	0,008108659
6p12.1	0,008343846
6p12.2	0,008343846
14q32.3	0,008500606
6q13	0,008500606
6p12.3	0,008735699
4q13	0,010379795
4q21.22	0,010458017
4q21.23	0,010536233
4q21.21	0,010536233
4q21.3	0,010849035
4q21.2	0,011083572
4q21.1	0,011318053
4q12	0,011474343
6p12	0,011786848
4q21	0,021193592
14q32	0,045792509



Ellis Van Creveld Syndrome	
band	*p*-value
4p16	0
4p16.1	6,83588E-05
4cen	0,00496766
4p11	0,005816778
4p12	0,007936404
4p13	0,008782988
4q11	0,009769757

fibrodysplasia ossificans progressiva	
band	*p*-value
4q28.1	0,004605597
4q28.2	0,00466854
4q28.3	0,004731479
4q31.23	0,005046114
4q31.21	0,005046114
4q31.22	0,005109029
4q31.2	0,00529775
4q31.1	0,005360649
4q31.3	0,005549323
4q27	0,005675085
4q28	0,005800832
14q21	0,007182994
4q31	0,009252636
17q22	0,019908243
17q21.2	0,022136884
17q21.33	0,022384198
17q21.32	0,022446017
17q21.31	0,022507831
17q21.1	0,023434584
17q21.3	0,025038874
17q21	0,04565638



Huntington disease	
band	*p*-value
4p16.3	1,60E-14
4p16	1,74E-14
4p16.1	3,02E-12
4pter	2,51E-11
4p15.3	3,29E-10
4ptel	8,78E-10
4p16.2	1,38E-09
4p15	4,96E-09
4p15.33	9,58E-09
4p15.31	9,58E-09
4p15.32	1,03E-08
4p15.1	1,46E-08
4p15.2	1,55E-08
14q	2,99E-04
20p12	0,005111011
20ptel	0,021001608
20pter	0,026652377
16q11,1	0,031336428
16cen	0,033203731
16q11,2	0,036927545
20p12.3	0,036927545
2p11,1	0,044332198
16q11	0,046174449
16p11,1	0,046174449
2p11,2	0,047094242

achondroplasia	
band	*p*-value
4p16	0,00E+00
4p16.3	0,00E+00
19p12	1,26E-02
13q	4,14E-02

Parkinson disease	
band	*p*-value
2p13	1,32E-05
1ptel	2,04E-05
1p36.32	2,12E-05
1p36.31	2,17E-05
1p36.21	2,40E-05
1p36.22	2,43E-05
1p36.33	2,52E-05
1p36.23	2,55E-05
1pter	2,74E-05
1p36.11	3,44E-05
1p36.12	3,51E-05
1p36.13	3,55E-05
1p36.2	3,71E-05
1p35.2	4,51E-05
1p35.3	4,51E-05
1p35.1	4,64E-05
1p36.1	5,41E-05
1p35	6,37E-05
1p36.3	6,54E-05
4q22.2	3,38E-04
4q22.3	3,38E-04
4q22.1	3,44E-04
4q21.22	4,22E-04
4q21.23	4,28E-04
4q21.21	4,28E-04
4q22	4,47E-04
4q21.3	4,54E-04
4q21.2	4,74E-04
4q21.1	4,94E-04
6qtel	6,01E-04
6qter	7,35E-04
6q25.2	8,38E-04
6q25.3	8,64E-04
6q26	0,001042487
1p36	0,001107022
6q27	0,001493136
6q25	0,001504812
4q21	0,001722449
4p16	0,003084742
2p11.1	0,010349236
2p11.2	0,011006178
4p16.1	0,018203814
2qtel	0,018638346
2p11	0,018855541
2q37.2	0,019506836
2p13.1	0,019506836
2q37.1	0,021024842
2qter	0,022540499
10q	0,023405535
2q37.3	0,023837765
4q23	0,024701653
2q36.2	0,026211617
2q36.3	0,026642606
2q36.1	0,026858029
14q	0,028579696
2q36	0,032013898
2q37	0,043726637
8p21.3	0,047107292

Wel(l)come

Welcome to the Text Based Prioritization Validation Application. If you are not part of EBI, Sanger or ESAT, you probably should go somewhere else.

This tool is under construction and shows text based gene profiling for human dysmorphy. Either choose a gene or a concept.

Examples:

- Microcephalin (MCPH1): ENSG00000147316
- Lissencephaly (LIS1): ENSG00000007168
- WHS Candidate 1 (WHSC1): ENSG00000109685
- SOTOS (NSD1): ENSG00000165671

[Contact] me about this tool.

ENSG00000169057
Gene to Dysmorphy

06.01.15
Dysmorphy to Gene

Status message

Ok. Gene requested. looking for ENSEMBL Gene **169057**...

Search Dysmorphology

This now works for both IE and Firefox.

auti
32.05.01: Autism / autistic behaviour

Search Gene

mec
ENSG00000116353: MECR
ENSG00000101935: AMMECR1
ENSG00000169057: MECP2

Results

concept	# abstr	*p*-value
32.05.01 (Autism / autistic behaviour)	42	0E0
32.04.00 (MENTAL,COGNITIVE FUNCTION, general abnormalities)	99	0E0
32.06.00 (SEIZURES, general abnormalities)	37	0E0
32.01.00 (ONSET OF NEUROLOGICAL SIGNS)	47	0E0
32.04.02 (Mental retardation / developmental delay)	93	0E0
32.00.00 (NEUROLOGY)	253	0E0
32.05.00 (BEHAVIOURAL PROBLEMS, general abnormalities)	58	5.1E-30
32.08.05 (Microcephaly)	12	3.0502E-18
16.01.06 (Scoliosis)	9	1.5409E-16
16.01.00 (Back and spine, general abnormalities)	13	1.1847E-13
32.08.00 (CRANIUM, general abnormalities)	14	9.0507E-13
16.00.00 (BACK AND SPINE)	13	1.9928E-12
32.06.02 (Infantile spasms)	6	1.3695E-10
32.06.17 (Spasms (not infantile))	6	1.518E-8
32.17.06 (Spasticity / brisk reflexes / Babinski)	8	2.0001E-8
32.17.00 (PYRAMIDAL SIGNS, general abnormalities)	11	1.5195E-7
32.16.00 (ATAXIA, general abnormalities)	10	9.384E-7
32.05.13 (Bruxism (teeth grinding))	2	1.7407E-5
32.07.20 (Intermittent tremor at rest)	4	1.0367E-4
32.35.04 (Cortex - nonspecific)	13	1.3582E-4
32.04.05 (Hypotonia (non-myopathic))	4	1.9534E-4
07.01.01 (Asymmetric eyes)	1	6.7689E-4
32.35.05 (Basal ganglia - nonspecific (excl. pigment))	4	7.4689E-4
32.36.01 (Multiple deletions)	2	7.9404E-4
18.01.07 (Feeding problems in infants)	2	8.4733E-4
32.06.06 (Myoclonus (for palatal and spinal myocl.i.v.))	3	1.0124E-3
32.35.00 (MICROSCOPIC CHANGES OF THE CNS, general abnorm.)	19	1.3275E-3
10.01.00 (Face, general abnormalities)	7	1.5255E-3
33.01.08 (Osteoporosis)	4	1.5325E-3
32.25.23 (EEG: *characteristic pattern of EEG / CN claimed)	1	2.3671E-3
10.00.00 (FACE)	7	2.615E-3
32.04.06 (Dyspraxia / apraxia including gait apraxia)	2	3.0051E-3
26.01.01 (Flat arches of feet)	1	3.0424E-3
32.35.06 (Cerebellum-nonspecific (see below for specific))	7	3.4779E-3
17.06.06 (Respiratory abnormality, unspecified)	20	4.701E-3
32.08.08 (Trigonocephaly)	1	6.7484E-3
17.06.00 (Lung, general abnormalities)	20	6.9399E-3
32.32.16 (Agenesis / hypoplasia of corpus callosum)	1	9.0996E-3
10.01.04 (Coarse facial features)	1	1.1445E-2
17.02.03 (Gynaecomastia)	1	1.2449E-2
32.16.02 (Truncal ataxia / non-cerebellar ataxia)	1	1.4787E-2
06.01.08 (Low-set ears)	1	2.3091E-2
32.32.10 (Polymicrogyria)	1	2.3752E-2
32.07.00 (PAROXYSMAL DISORDERS, general abnormalities)	4	2.7803E-2
32.23.00 (AUTONOMIC DYSFUNCTIONS, general abnormalities)	2	3.0927E-2
16.01.02 (Kyphosis)	1	3.1653E-2
17.00.00 (THORAX)	32	3.7074E-2
32.36.03 (Point mutation)	7	4.7795E-2
32.26.00 (ELECTROPHYSIOLOGY (except EEG), general abnorm.)	5	4.8974E-2
32.25.00 (EEG, general abnormalities)	1	5.0807E-2
29.01.13 (Chromosome instability / breakage)	2	5.2354E-2
08.05.00 (Palpebral fissures, general abnormalities)	1	7.6176E-2
18.01.06 (*Dysphagia - see CRANIAL NERVES)	1	8.6134E-2
17.02.00 (Breasts, general abnormalities)	10	8.7576E-2
22.01.00 (Upper limbs, general abnormalities)	1	9.5984E-2
01.00.00 (BUILD)	5	1.0382E-1

Gene-LDDB maps

```
./genes.sh LDDB170504
Looking for genes associated to LDDB170504...
129     1.464757e-129   ENSG00000049540         +           ELN         elastin (supravalvular aortic...
12      3.120289e-12    ENSG00000071462         +           WBSCR22     Williams Beuren syndrome ...
12      1.810960e-12    ENSG00000077809         +           GTF2I       general transcription factor II, i
11      6.238573e-11    ENSG00000106638         +           TBL2        transducin (beta)-like 2
11      4.967726e-11    ENSG00000006704         +           GTF2IRD1    GTF2I repeat domain containing 1
10      9.880883e-10    ENSG00000168542         +           COL3A1      collagen, type III, alpha 1 ...
09      5.514204e-09    ENSG00000009950         +           MLXIPL      MLX interacting protein-like
08      2.764015e-08    ENSG00000077800         +           FKBP6       FK506 binding protein 6, 36kDa
08      1.398369e-08    ENSG00000107438         -           PDLIM1      PDZ and LIM domain 1 (elfin)
07      4.414078e-07    ENSG00000165119         -           HNRPK       heterogeneous nuclear ...
```

Supravalvular Aortic Stenosis

Gene-LDDB maps

```
./genes.sh LDDB120401
Looking for genes associated to LDDB120401...
19      1.926391e-19    ENSG00000184058         +           TBX1        T-box 1
18      4.051166e-18    ENSG00000102271         +           KLHL4       kelch-like 4 (Drosophila)
15      9.455555e-15    ENSG00000119042         +           SATB2       SATB family member 2
15      3.299856e-15    ENSG00000101871         +           MID1        midline 1 (Opitz/BBB syndrome)
14      1.279493e-14    ENSG00000069399         -           BCL3        B-cell CLL/lymphoma 3
13      1.520270e-13    ENSG00000204248         +           COL11A2     collagen, type XI, alpha 2
12      1.585302e-12    ENSG00000099889         +           ARVCF       armadillo repeat gene in VCF
10      5.880943e-10    ENSG00000060718         +           COL11A1     collagen, type XI, alpha 1
10      2.295810e-10    ENSG00000172893         +           DHCR7       7-dehydrocholesterol reductase
10      1.824248e-10    ENSG00000163513         +           TGFBR2      transforming growth factor, ...
```

Cleft Palate

Conclusion

Extract Genotype – Phenotype links

- Using biomedical literature as a phenotype database,
- using available ontologies as sources of biomedical concepts within different views

Publications

Van Vooren S, Thienpont B, Menten B, Speleman F, De Moor B, Vermeesch J R, Moreau Y., **"Mapping biomedical concepts onto the human genome by mining literature on chromosomal aberrations"** *Nucleic Acids Res.*, 35(8):2533-43, 2007.

Van Vooren S, Coessens B, De Moor B, Moreau Y, Vermeesch J R., **"Array CGH and computational genome annotation in constitutional cytogenetics: suggesting candidate genes for novel submicroscopic chromosomal imbalance syndromes"** *Genet Med.*, 9(9):642-9, 2007.

Chapter 3 – Functional interpretation of genetic information

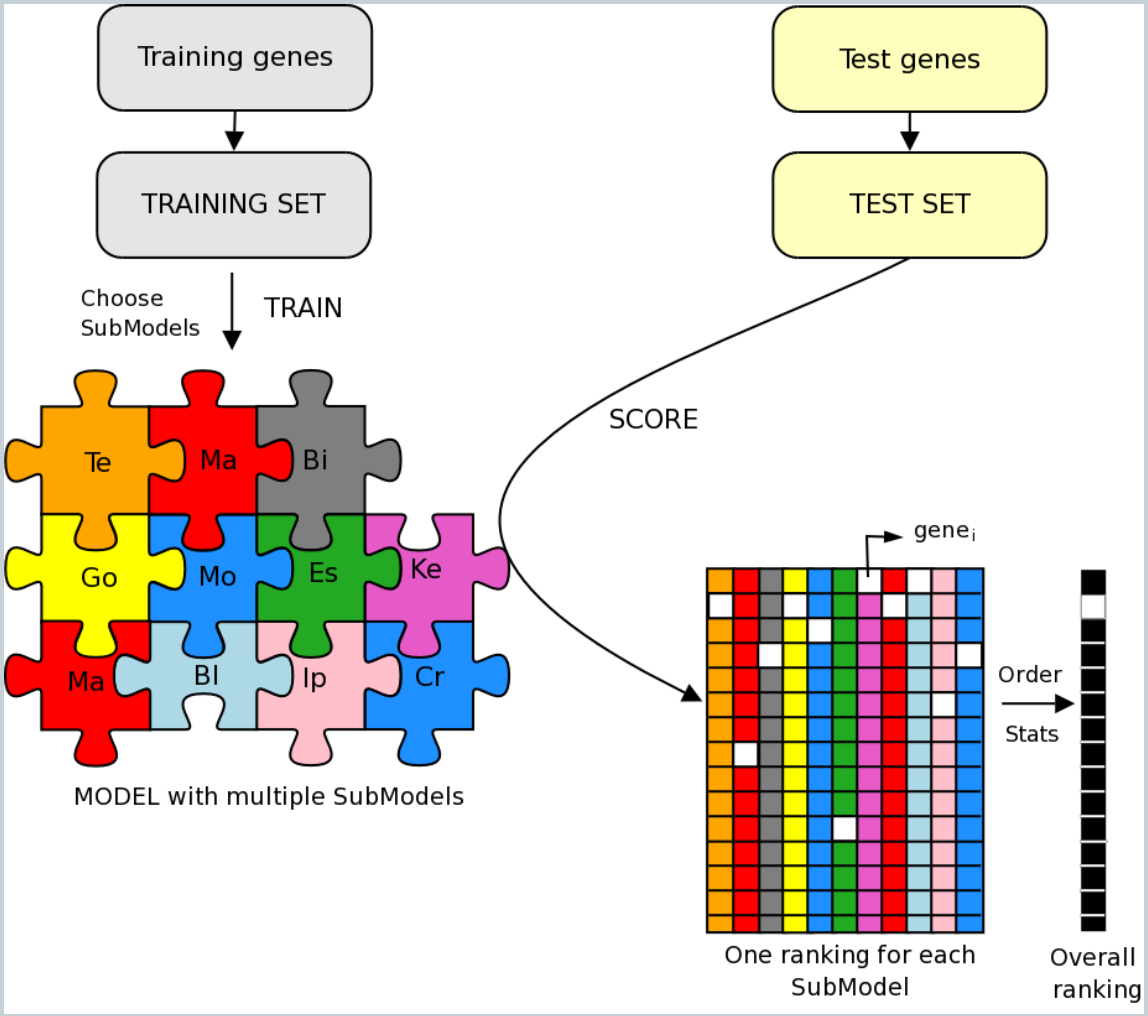
- Downstream analysis: tools for interpretation
  - Association to phenotype and disease
    - Information from public databases and literature
    - Patient phenotype information
    - Other patients
  - Hunting for candidate genes
    - Computational approaches: a review
    - Literature as information source
    - Mapping concepts onto the genome
      - aBandApart
      - aGeneApart
    - **Fusing multiple data sources**
      - **Endeavour**

What is Endeavour?

- Prioritisation of any list of candidate genes

- To identify disease genes
- To identify members of biological pathways
- To identify functionally similar genes
- ...

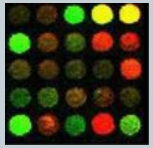
What is Endeavour?



What is Endeavour?

- Compares summaries of information of **test genes** with that of **training genes**
- Uses different **information sources**

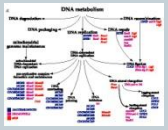
What is Endeavour?



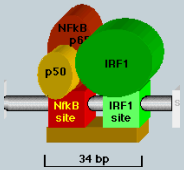
Gene expression

Anatomical expression

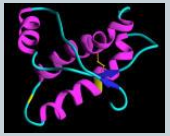
Literature



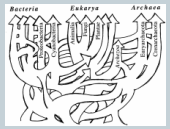
Biological process



Gene regulation

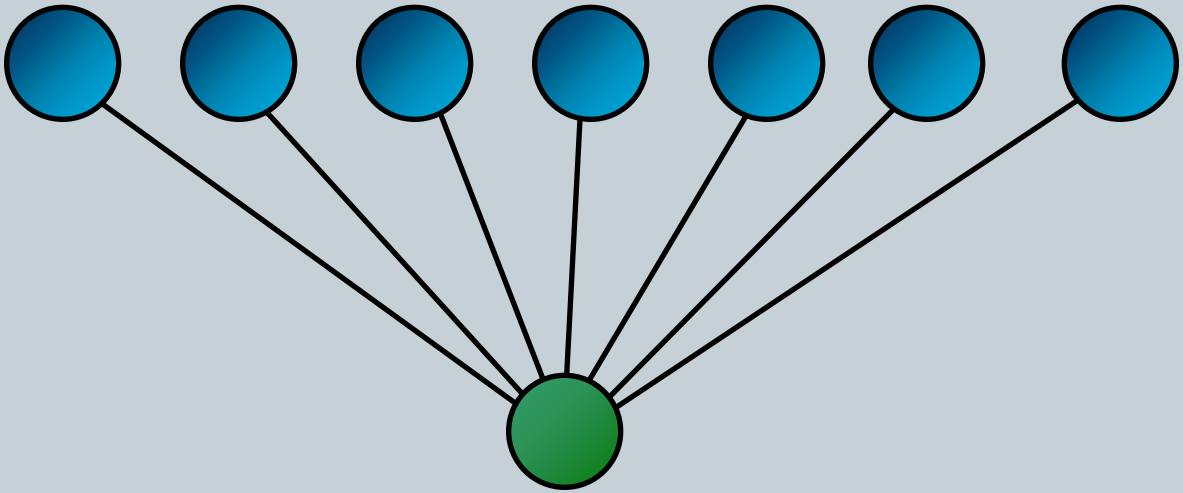


Protein domains



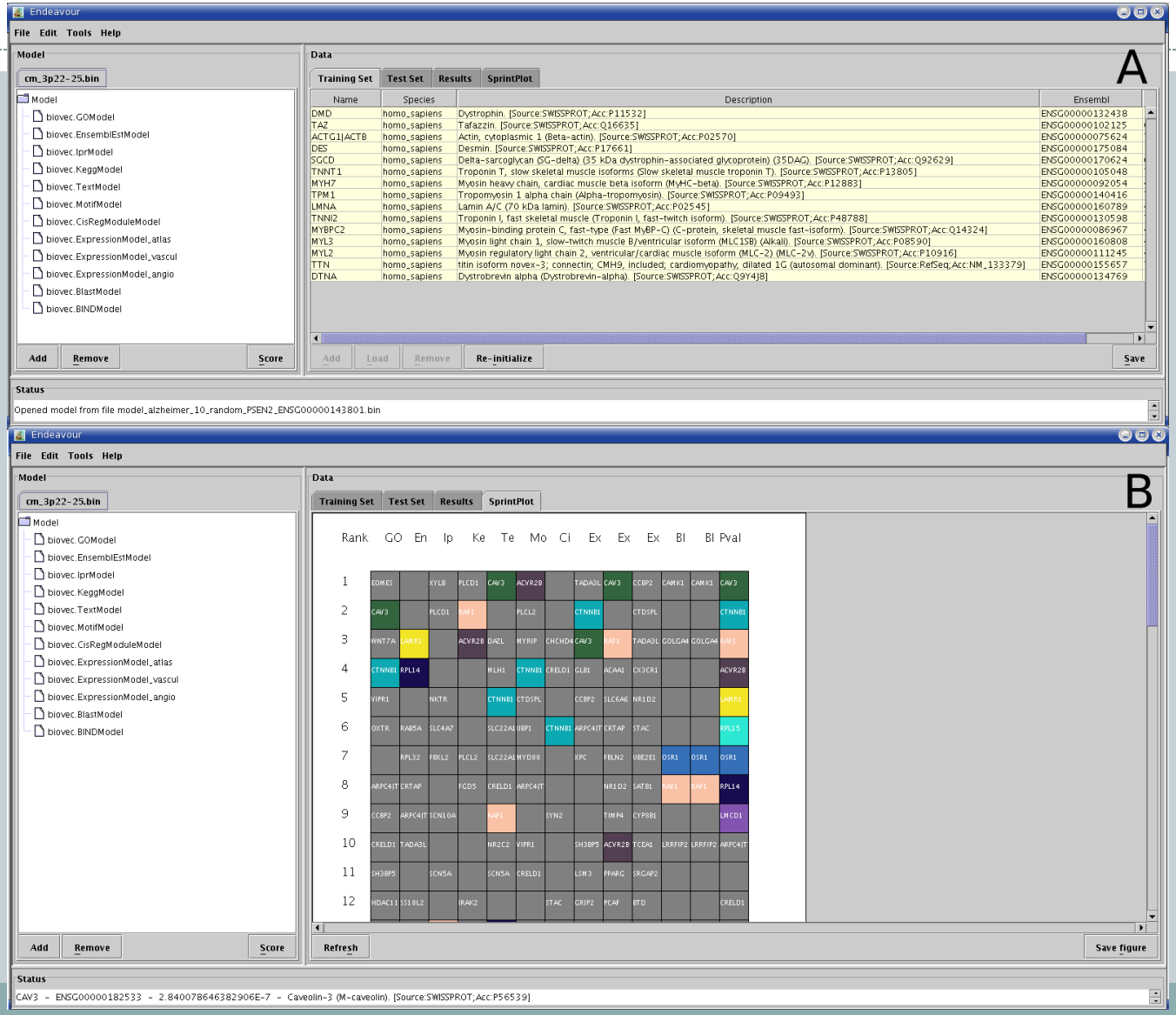
Evolutionary conservation

...

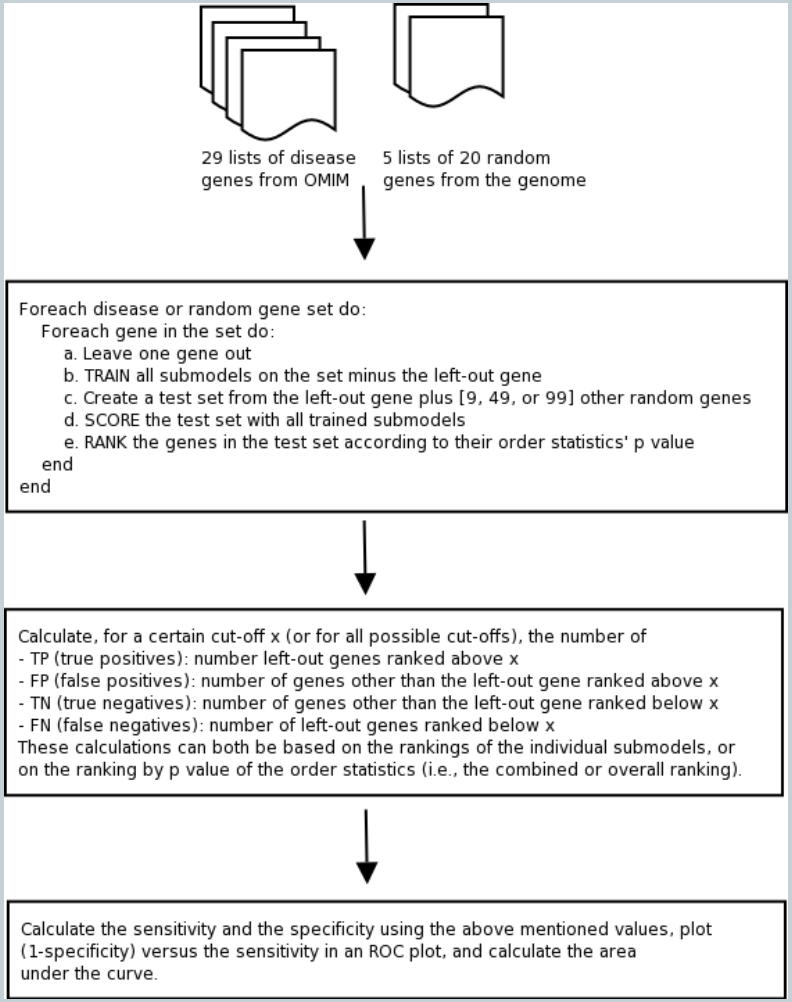


Pathology / …

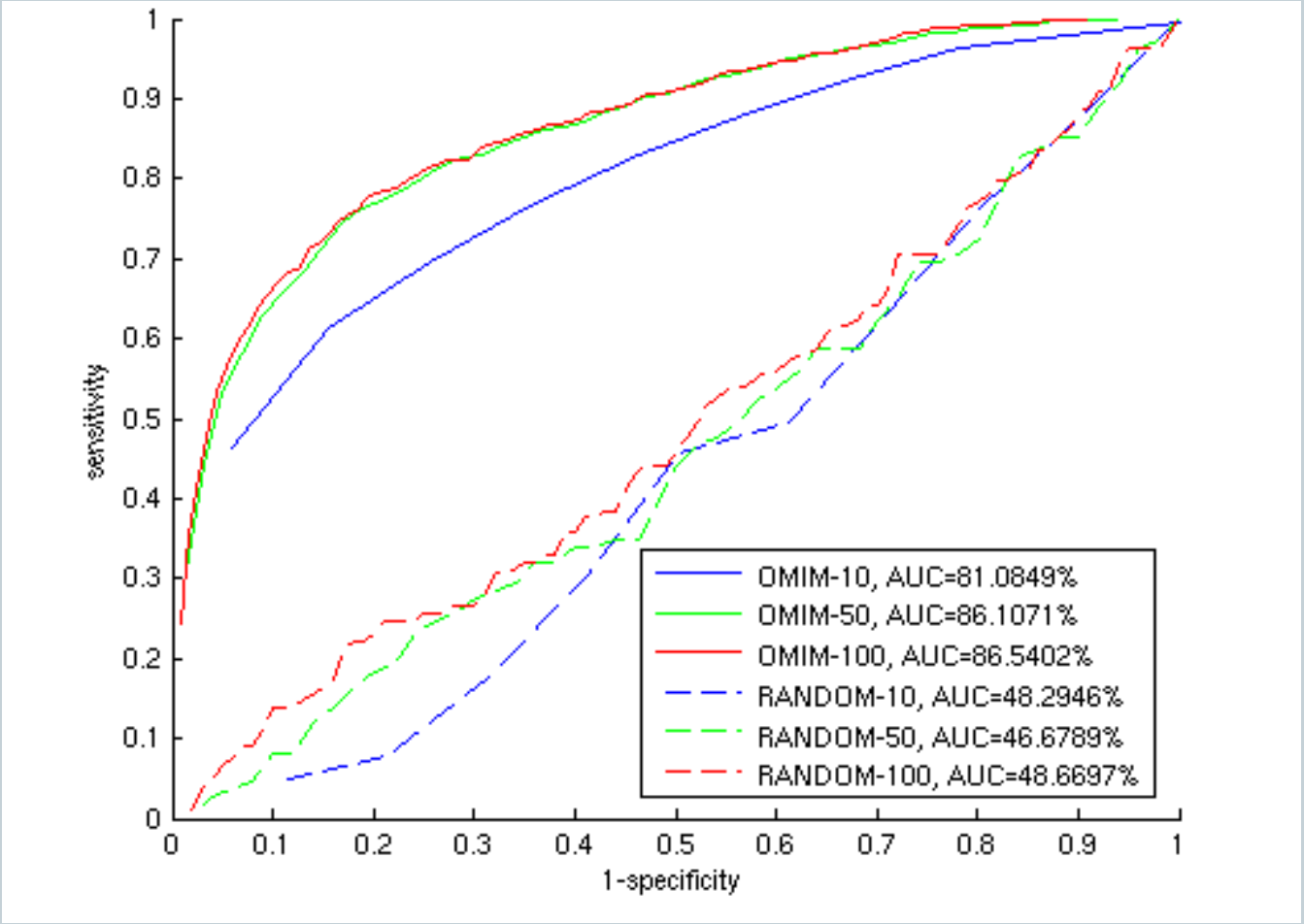
Application Screenshot



Validation Setup



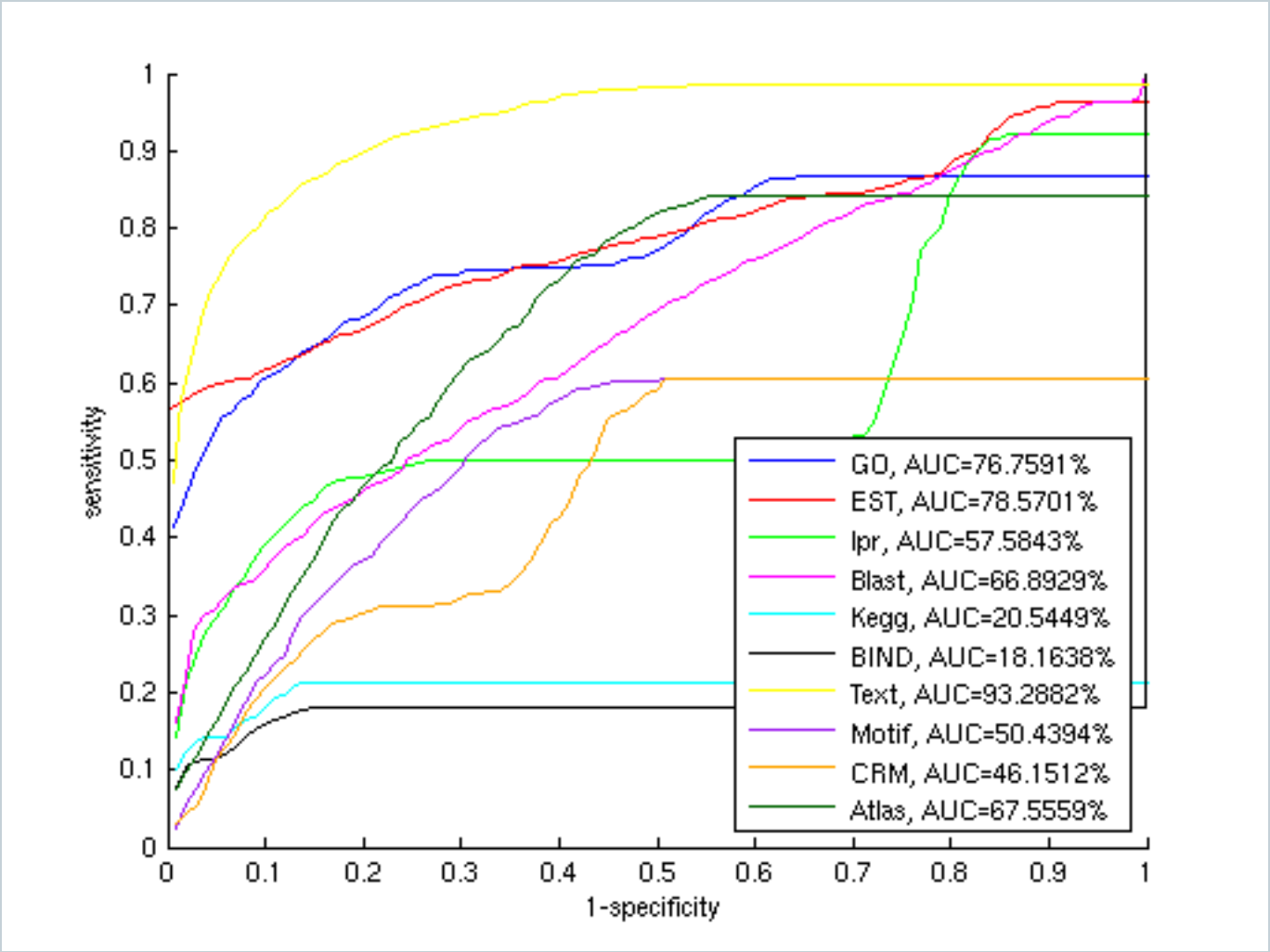
Validation Results



Sensitivity - the proportion of true positives or the proportion of cases correctly identified by the test **as meeting** a certain condition (e.g. in mammography testing, the proportion of patients **with** cancer who test positive).

Specificity - the proportion of true negatives or the proportion of cases correctly identified by the test as **not meeting** a certain condition (e.g. in mammography testing, the proportion of patients **without** cancer who test negative).

Validation Results



Publications

Tranchevent L.C., Barriot R., Yu S., **Van Vooren S.**, Van Loo P., Coessens B., De Moor B., Aerts S., Moreau Y., **"ENDEAVOUR update: a web resource for gene prioritization in multiple species [245]"** *Nucleic Acids Res.* 36:W377{384, 2008.

Balikova I, Martens K, Melotte C, Amyere M, **Van Vooren S**, Moreau Y, Vetrie D, Fiegler H, Carter N P, Liehr T, Vikkula M, Matthijs G, Fryns J P, Casteels I, Devriendt K, Vermeesch J R., **"Autosomal dominant microtia linked to five tandem copies of a copy number variable region at chromosome 4p16"** *Am J Hum Genet.*, 82(1):181-7, 2008.

Breckpot J, Takiyama Y, Thienpont B, **Van Vooren S**, Vermeesch J.R, Ortibus E, Devriendt K., **"A novel genomic disorder: a deletion of the SACS gene leading to Spastic Ataxia of Charlevoix-Saguenay"** Eur. J. Hum. Genet., 2008.

Chapter 4 – Intelligent databases for constitutional cytogenetics

CNV INTERPRETATION: DECIPHER, STORE+BENCH

Chapter 4 – Intelligent databases for constitutional cytogenetics

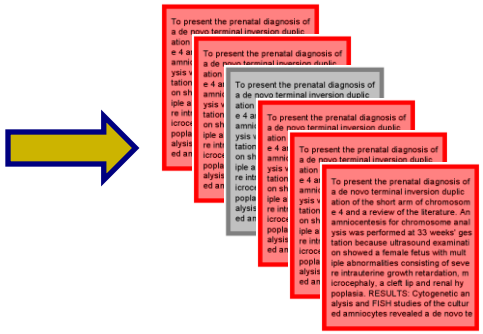
- ‘Downstream analysis’
  - DECIPHER
    - Consortium for aggregation of rare cases
    - Text based gene prioritisation
  - Store+Bench
    - Feature rich platform for downstream analysis

Chapter 4 – Intelligent databases for constitutional cytogenetics

- ‘Downstream analysis’
  - **DECIPHER**
    - **Consortium for aggregation of rare cases**
    - **Text based gene prioritisation**
  - Store+Bench
    - Feature rich platform for downstream analysis

Gene to Concept associations

ENSG00000000001
ENSG00000000002
...
ENSG00000109685
...
ENSG00000024999
ENSG00000025000



Microcephaly

overrepresented in document set for WHSC1 gene

PubMed

Wel(l)come

Welcome to the Text Based Prioritization Validation Application. If you are not part of EBI, Sanger or ESAT, you probably should go somewhere else.

This tool is under construction and shows text based gene profiling for human dysmorphy. Either choose a gene or a concept.

Examples:

- Microcephalin (MCPH1): ENSG00000147316
- Lissencephaly (LIS1): ENSG00000007168
- WHS Candidate 1 (WHSC1): ENSG00000109685
- SOTOS (NSD1): ENSG00000165671

[Contact] me about this tool.

ENSG00000169057
Gene to Dysmorphy

06.01.15
Dysmorphy to Gene

Status message

Ok. Gene requested. looking for ENSEMBL Gene **169057**...

Search Dysmorphology

This now works for both IE and Firefox.

auti
32.05.01: Autism / autistic behaviour

Search Gene

mec
ENSG00000116353: MECR
ENSG00000101935: AMMECR1
ENSG00000169057: MECP2

Results

concept	# abstr	*p*-value
32.05.01 (Autism / autistic behaviour)	42	0E0
32.04.00 (MENTAL,COGNITIVE FUNCTION, general abnormalities)	99	0E0
32.06.00 (SEIZURES, general abnormalities)	37	0E0
32.01.00 (ONSET OF NEUROLOGICAL SIGNS)	47	0E0
32.04.02 (Mental retardation / developmental delay)	93	0E0
32.00.00 (NEUROLOGY)	253	0E0
32.05.00 (BEHAVIOURAL PROBLEMS, general abnormalities)	58	5.1E-30
32.08.05 (Microcephaly)	12	3.0502E-18
16.01.06 (Scoliosis)	9	1.5409E-16
16.01.00 (Back and spine, general abnormalities)	13	1.1847E-13
32.08.00 (CRANIUM, general abnormalities)	14	9.0507E-13
16.00.00 (BACK AND SPINE)	13	1.9928E-12
32.06.02 (Infantile spasms)	6	1.3695E-10
32.06.17 (Spasms (not infantile))	6	1.518E-8
32.17.06 (Spasticity / brisk reflexes / Babinski)	8	2.0001E-8
32.17.00 (PYRAMIDAL SIGNS, general abnormalities)	11	1.5195E-7
32.16.00 (ATAXIA, general abnormalities)	10	9.384E-7
32.05.13 (Bruxism (teeth grinding))	2	1.7407E-5
32.07.20 (Intermittent tremor at rest)	4	1.0367E-4
32.35.04 (Cortex - nonspecific)	13	1.3582E-4
32.04.05 (Hypotonia (non-myopathic))	4	1.9534E-4
07.01.01 (Asymmetric eyes)	1	6.7689E-4
32.35.05 (Basal ganglia - nonspecific (excl. pigment))	4	7.4689E-4
32.36.01 (Multiple deletions)	2	7.9404E-4
18.01.07 (Feeding problems in infants)	2	8.4733E-4
32.06.06 (Myoclonus (for palatal and spinal myocl.i.v.))	3	1.0124E-3
32.35.00 (MICROSCOPIC CHANGES OF THE CNS, general abnorm.)	19	1.3275E-3
10.01.00 (Face, general abnormalities)	7	1.5255E-3
33.01.08 (Osteoporosis)	4	1.5325E-3
32.25.23 (EEG: *characteristic pattern of EEG / CN claimed)	1	2.3671E-3
10.00.00 (FACE)	7	2.615E-3
32.04.06 (Dyspraxia / apraxia including gait apraxia)	2	3.0051E-3
26.01.01 (Flat arches of feet)	1	3.0424E-3
32.35.06 (Cerebellum-nonspecific (see below for specific))	7	3.4779E-3
17.06.06 (Respiratory abnormality, unspecified)	20	4.701E-3
32.08.08 (Trigonocephaly)	1	6.7484E-3
17.06.00 (Lung, general abnormalities)	20	6.9399E-3
32.32.16 (Agenesis / hypoplasia of corpus callosum)	1	9.0996E-3
10.01.04 (Coarse facial features)	1	1.1445E-2
17.02.03 (Gynaecomastia)	1	1.2449E-2
32.16.02 (Truncal ataxia / non-cerebellar ataxia)	1	1.4787E-2
06.01.08 (Low-set ears)	1	2.3091E-2
32.32.10 (Polymicrogyria)	1	2.3752E-2
32.07.00 (PAROXYSMAL DISORDERS, general abnormalities)	4	2.7803E-2
32.23.00 (AUTONOMIC DYSFUNCTIONS, general abnormalities)	2	3.0927E-2
16.01.02 (Kyphosis)	1	3.1653E-2
17.00.00 (THORAX)	32	3.7074E-2
32.36.03 (Point mutation)	7	4.7795E-2
32.26.00 (ELECTROPHYSIOLOGY (except EEG), general abnorm.)	5	4.8974E-2
32.25.00 (EEG, general abnormalities)	1	5.0807E-2
29.01.13 (Chromosome instability / breakage)	2	5.2354E-2
08.05.00 (Palpebral fissures, general abnormalities)	1	7.6176E-2
18.01.06 (*Dysphagia - see CRANIAL NERVES)	1	8.6134E-2
17.02.00 (Breasts, general abnormalities)	10	8.7576E-2
22.01.00 (Upper limbs, general abnormalities)	1	9.5984E-2
01.00.00 (BUILD)	5	1.0382E-1

DECIPHER: Report for patient CHG00001206

Patient Number	CHG00001206 Edit
Record created	4th Jul 2006
Record last updated	5th Jul 2006
hide **details**	
External Reference	139215
Note	
Chromosomal Sex	46,XX
Age*(yrs)	12
Consent for display in Ensembl	N
hide **Citations**	

hide Phenotypes

Phenotypes

Primary	Secondary	Tertiary
NEUROLOGY	NEURORADIOLOGY CONT, general abnormalities	**Lissencephaly/pachygyria**
NEUROLOGY	MENTAL,COGNITIVE FUNCTION, general abnormalities	**Mental retardation/developmental delay**
NEUROLOGY	CRANIUM, general abnormalities	**Microcephaly**
MOUTH	Mouth, general abnormalities	**Open mouth appearance**
HANDS	Fingers, general abnormalities	**Thin fingers**
NEUROLOGY	SEIZURES, general abnormalities	**Tonic/clonic (grand-mal)**
EYES, GLOBES	Vision, general abnormalities	**Vision, non-specific impairment**

hide Arrays

Array Information

Array Type	Reference DNA	Slide #	Image Ref	
1Mb Clone Array (NCBI36)	other patients	2005033057		View

hide Feature Information

Feature Information

Origin	~~hideOrigin Unknown~~	~~showOrigin Familial~~	~~hideOrigin de novo~~
Origin counts	Unknown Counts:0	Familial Counts:0	Denovo Counts: 1

Chr	17
Upstream Flank	1882486 (RP11-233O10)
Start	2312031 (RP11-135N5) 2
End	2492162 (RP11-135N5) 2
Downstream Flank	3241556 (RP11-147K16)
Mean Ratio	-0.5869
Origin of Altered Region	de novo - parental origin of rearrangement undefined
Confirmation	FISH
Interval	1359070
View in genomic context	e! cytoview

Known Syndromes

Miller-Dieker syndrome (MDS)

show predicted | show known Ensembl Genes

~~show ALL~~ ~~hide OMIM~~ ~~show Imprinted~~ HUGO Gene Lists

HUGO Gene Names	HUGO Gene Descriptions
DPH1 (bp:1880156-1893469)	Description: DPH1 homolog (S. cerevisiae). Aliases: OVCA1 [Ensembl:DPH1] [OMIM:603527]
HIC1 (bp:1906354-1909070)	Description: hypermethylated in cancer 1. Aliases: ZBTB29 [Ensembl:HIC1] [OMIM:603825]
SRR (bp:2153998-2175304)	Description: serine racemase. Aliases: [Ensembl:SRR] [OMIM:606477]
MNT (bp:2234104-2251162)	Description: MAX binding protein. Aliases: ROX, MXD6, MAD6 [Ensembl:MNT] [OMIM:603039]
PAFAH1B1 (bp:2443686-2535638)	Description: platelet-activating factor acetylhydrolase, isoform Ib, alpha subunit 45kDa. Aliases: LIS1, MDCR, PAFAH [Ensembl:PAFAH1B1] [OMIM:601545]
OR1D2 (bp:2942102-2943040)	Description: olfactory receptor, family 1, subfamily D, member 2. Aliases: OR17-4 [Ensembl:OR1D2] [OMIM:164342]

show Prioritise genes by patient phenotype

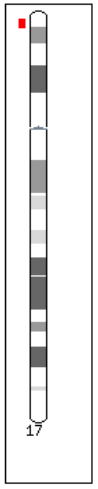
show Prioritise Genes per phenotype trait

hide Translocations

hide Karyotype

del(17)(p13.2;p13.3){1.5 Mb}

Key:
≥Mean Ratios >0 indicates a DUPLICATION
≥Mean Ratios <0 indicates a DELETION
>Indicates a translocation breakpoint



DECIPHER: Report for patient CHG00001206

Patient Number	CHG00001206 Edit
Record created	4th Jul 2006
Record last updated	5th Jul 2006
hide **details**	
External Reference	139215
Note	
Chromosomal Sex	46,XX
Age*(yrs)	12
Consent for display in Ensembl	N
hide **Citations**	

hide Phenotypes

Phenotypes

Primary	Secondary	Tertiary
NEUROLOGY	NEURORADIOLOGY CONT, general abnormalities	**Lissencephaly/pachygyria**
NEUROLOGY	MENTAL,COGNITIVE FUNCTION, general abnormalities	**Mental retardation/developmental delay**
NEUROLOGY	CRANIUM, general abnormalities	**Microcephaly**
MOUTH	Mouth, general abnormalities	**Open mouth appearance**
HANDS	Fingers, general abnormalities	**Thin fingers**
NEUROLOGY	SEIZURES, general abnormalities	**Tonic/clonic (grand-mal)**
EYES, GLOBES	Vision, general abnormalities	**Vision, non-specific impairment**

hide Arrays

Array Information

Array Type	Reference DNA	Slide #	Image Ref	
1Mb Clone Array (NCBI36)	other patients	2005033057		View

hide Feature Information

Feature Information

Origin	~~hideOrigin Unknown~~	~~showOrigin Familial~~	~~hideOrigin de novo~~
Origin counts	Unknown Counts:0	Familial Counts:0	Denovo Counts: 1

Chr	17
Upstream Flank	1882486 (RP11-233O10)
Start	2312031 (RP11-135N5) 2
End	2492162 (RP11-135N5) 2
Downstream Flank	3241556 (RP11-147K16)
Mean Ratio	-0.5869
Origin of Altered Region	de novo - parental origin of rearrangement undefined
Confirmation	FISH
Interval	1359070
View in genomic context	e! cytoview

Known Syndromes

Miller-Dieker syndrome (MDS)

show predicted show known Ensembl Genes

~~show ALL~~ ~~hide OMIM~~ ~~show Imprinted~~ HUGO Gene Lists

HUGO Gene Names	HUGO Gene Descriptions
DPH1 (bp:1880156-1893469)	Description: DPH1 homolog (S. cerevisiae). Aliases: OVCA1 [Ensembl:DPH1] [OMIM:603527]
HIC1 (bp:1906354-1909070)	Description: hypermethylated in cancer 1. Aliases: ZBTB29 [Ensembl:HIC1] [OMIM:603825]
SRR (bp:2153998-2175304)	Description: serine racemase. Aliases: [Ensembl:SRR] [OMIM:606477]
MNT (bp:2234104-2251162)	Description: MAX binding protein. Aliases: ROX, MXD6, MAD6 [Ensembl:MNT] [OMIM:603039]
PAFAH1B1 (bp:2443686-2535638)	Description: platelet-activating factor acetylhydrolase, isoform Ib, alpha subunit 45kDa. Aliases: LIS1, MDCR, PAFAH [Ensembl:PAFAH1B1] [OMIM:601545]
OR1D2 (bp:2942102-2943040)	Description: olfactory receptor, family 1, subfamily D, member 2. Aliases: OR17-4 [Ensembl:OR1D2] [OMIM:164342]

show Prioritise genes by patient phenotype

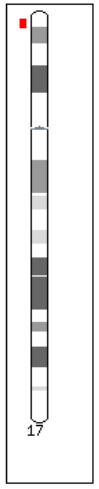
show Prioritise Genes per phenotype trait

hide Translocations

hide Karyotype

del(17)(p13.2;p13.3){1.5 Mb}

Key:
≥Mean Ratios >0 indicates a DUPLICATION
≥Mean Ratios <0 indicates a DELETION
>Indicates a translocation breakpoint



Chr	17
Upstream Flank	1882486 (RP11-233O10)
Start	2312031 (RP11-135N5) 2
End	2492162 (RP11-135N5) 2
Downstream Flank	3241556 (RP11-147K16)
Mean Ratio	-0.5869
Origin of Altered Region	de novo - parental origin of rearrangement undefined
Confirmation	FISH
Interval	1359070
View in genomic context	e! cytoview

Known Syndromes

Miller-Dieker syndrome (MDS)

show predicted | show known Ensembl Genes

~~show ALL~~ ~~hide OMIM~~ ~~show Imprinted~~ HUGO Gene Lists

HUGO Gene Names	HUGO Gene Descriptions
DPH1 (bp:1880156-1893469)	Description: DPH1 homolog (S. cerevisiae). Aliases: OVCA1 [Ensembl:DPH1] [OMIM:603527]
HIC1 (bp:1906354-1909070)	Description: hypermethylated in cancer 1. Aliases: ZBTB29 [Ensembl:HIC1] [OMIM:603825]
SRR (bp:2153998-2175304)	Description: serine racemase. Aliases: [Ensembl:SRR] [OMIM:606477]
MNT (bp:2234104-2251162)	Description: MAX binding protein. Aliases: ROX, MXD6, MAD6 [Ensembl:MNT] [OMIM:603039]
PAFAH1B1 (bp:2443686-2535638)	Description: platelet-activating factor acetylhydrolase, isoform Ib, alpha subunit 45kDa. Aliases: LIS1, MDCR, PAFAH [Ensembl:PAFAH1B1] [OMIM:601545]
OR1D2 (bp:2942102-2943040)	Description: olfactory receptor, family 1, subfamily D, member 2. Aliases: OR17-4 [Ensembl:OR1D2] [OMIM:164342]

hide Prioritise genes by patient phenotype

Gene Id	Log score	traits involved	Literature evidence	Gene information
ENSG00000007168	121.02	13	(157 citations)	Platelet-activating factor acetylhydrolase IB subunit alpha (PAF acetylhydrolase 45 kDa subunit) (PAF-AH 45 kDa subunit) (PAF-AH alpha) (PAFAH alpha) (Lissencephaly-1 protein) (LIS-1). [Source:Uniprot/SWISSPROT;Acc:P43034]
ENSG00000177374	119.64	8	(24 citations)	Hypermethylated in cancer 1 protein (Hic-1) (Zinc finger and BTB domain-containing protein 29). [Source:Uniprot/SWISSPROT;Acc:Q14526]
ENSG00000132361	29.62	5	(33 citations)	Putative eukaryotic translation initiation factor 3 subunit (eIF-3). [Source:Uniprot/SWISSPROT;Acc:O75153]
ENSG00000108963	7.61	3	(4 citations)	candidate tumor suppressor in ovarian cancer 2 [Source:RefSeq_peptide;Acc:NP_543012]
ENSG00000070366	2.68	1	(71 citations)	Telomerase-binding protein EST1A (Ever shorter telomeres 1A) (Telomerase subunit EST1A) (EST1-like protein A) (hSmg5/7a). [Source:Uniprot/SWISSPROT;Acc:Q86US8]
ENSG00000141258	1.27	1	(1 citation)	RUN and TBC1 domain containing 1 [Source:RefSeq_peptide;Acc:NP_055668]
ENSG00000167721	1.27	1	(1 citation)	TSR1, 20S rRNA accumulation, homolog [Source:RefSeq_peptide;Acc:NP_060598]
ENSG00000205813	0.73	1	(1 citation)	Olfactory receptor 3A2 (Olfactory receptor 17-228) (OR17-228). [Source:Uniprot/SWISSPROT;Acc:P47893]
ENSG00000180068	0.73	1	(1 citation)	Olfactory receptor 3A4 (Olfactory receptor 17-24) (OR17-24). [Source:Uniprot/SWISSPROT;Acc:P47883]
ENSG00000132359	0.47	1	(1 citation)	GTPase activating Rap/RanGAP domain-like 4 [Source:RefSeq_peptide;Acc:NP_055900]

show Prioritise Genes per phenotype trait

hide Prioritise genes by patient phenotype

Gene Id	Log score	traits involved	Literature evidence	Gene information
ENSG00000007168	121.02	13	(157 citations)	Platelet-activating factor acetylhydrolase IB subunit alpha (PAF acetylhydrolase 45 kDa subunit) (PAF-AH 45 kDa subunit) (PAF-AH alpha) (PAFAH alpha) (Lissencephaly-1 protein) (LIS-1). [Source:Uniprot/SWISSPROT;Acc:P43034]
ENSG00000177374	119.64	8	(24 citations)	Hypermethylated in cancer 1 protein (Hic-1) (Zinc finger and BTB domain-containing protein 29). [Source:Uniprot/SWISSPROT;Acc:Q14526]
ENSG00000132361	29.62	5	(33 citations)	Putative eukaryotic translation initiation factor 3 subunit (eIF-3). [Source:Uniprot/SWISSPROT;Acc:O75153]
ENSG00000108963	7.61	3	(4 citations)	candidate tumor suppressor in ovarian cancer 2 [Source:RefSeq_peptide;Acc:NP_543012]
ENSG00000070366	2.68	1	(71 citations)	Telomerase-binding protein EST1A (Ever shorter telomeres 1A) (Telomerase subunit EST1A) (EST1-like protein A) (hSmg5/7a). [Source:Uniprot/SWISSPROT;Acc:Q86US8]
ENSG00000141258	1.27	1	(1 citation)	RUN and TBC1 domain containing 1 [Source:RefSeq_peptide;Acc:NP_055668]
ENSG00000167721	1.27	1	(1 citation)	TSR1, 20S rRNA accumulation, homolog [Source:RefSeq_peptide;Acc:NP_060598]
ENSG00000205813	0.73	1	(1 citation)	Olfactory receptor 3A2 (Olfactory receptor 17-228) (OR17-228). [Source:Uniprot/SWISSPROT;Acc:P47893]
ENSG00000180068	0.73	1	(1 citation)	Olfactory receptor 3A4 (Olfactory receptor 17-24) (OR17-24). [Source:Uniprot/SWISSPROT;Acc:P47883]
ENSG00000132359	0.47	1	(1 citation)	GTPase activating Rap/RanGAP domain-like 4 [Source:RefSeq_peptide;Acc:NP_055900]

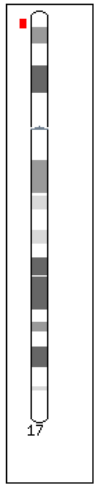
show Prioritise Genes per phenotype trait

hide Translocations

hide Karyotype

del(17)(p13.2;p13.3){1.5 Mb}

Key:
≥Mean Ratios >0 indicates a DUPLICATION
≥Mean Ratios <0 indicates a DELETION
>Indicates a translocation breakpoint



Note:Copy number changes are calculated from the start of the start clone to the end of the end clone to maintain constitency throughout DECIPHER.

hide Prioritise genes by patient phenotype

Gene Id	Log score	traits involved	Literature evidence	Gene information
ENSG00000007168	121.02	13	(157 citations)	Platelet-activating factor acetylhydrolase IB subunit alpha (PAF acetylhydrolase 45 kDa subunit) (PAF-AH 45 kDa subunit) (PAF-AH alpha) (PAFAH alpha) (Lissencephaly-1 protein) (LIS-1). [Source:Uniprot/SWISSPROT;Acc:P43034]

12885786: LIS1 missense mutations: variable phenotypes result from unpredictable altera..., *J. Biol. Chem.* (2003)
12551946: Targeted disruption of intracellular type I platelet activating factor-acetyl..., *J. Biol. Chem.* (2003)
11134054: Nudf, a fungal homolog of the human LIS1 protein, functions as a dimer in vivo., *J. Biol. Chem.* (2001)
9660828: Switching of platelet-activating factor acetylhydrolase catalytic subunits in..., *J. Biol. Chem.* (1998)
8537406: Cloning and expression of a cDNA encoding the beta-subunit (30-kDa subunit) o..., *J. Biol. Chem.* (1995)
7499404: Affinity chromatography demonstrates a direct binding between cytoplasmic dyn..., *J. Biol. Chem.* (1995)
7744858: Inhibition of neurogenic precursor proliferation by antisense alpha thyroid h..., *J. Biol. Chem.* (1995)
1400364: Kinesin and cytoplasmic dynein binding to brain microsomes., *J. Biol. Chem.* (1992)
16799048: Docosahexaenoic acid promotes photoreceptor differentiation without altering ..., *Invest. Ophthalmol. Vis. Sci.* (2006)
11687549: Effect of GDNF on neuroblast proliferation and photoreceptor survival: additi..., *Invest. Ophthalmol. Vis. Sci.* (2001)

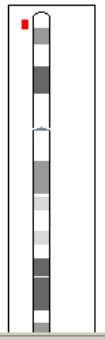
ENSG00000177374	119.64	8	(24 citations)	Hypermethylated in cancer 1 protein (Hic-1) (Zinc finger and BTB domain-containing protein 29). [Source:Uniprot/SWISSPROT;Acc:Q14526]
ENSG00000132361	29.62	5	(33 citations)	Putative eukaryotic translation initiation factor 3 subunit (eIF-3). [Source:Uniprot/SWISSPROT;Acc:O75153]
ENSG00000108963	7.61	3	(4 citations)	candidate tumor suppressor in ovarian cancer 2 [Source:RefSeq_peptide;Acc:NP_543012]
ENSG00000070366	2.68	1	(71 citations)	Telomerase-binding protein EST1A (Ever shorter telomeres 1A) (Telomerase subunit EST1A) (EST1-like protein A) (hSmg5/7a). [Source:Uniprot/SWISSPROT;Acc:Q86US8]
ENSG00000141258	1.27	1	(1 citation)	RUN and TBC1 domain containing 1 [Source:RefSeq_peptide;Acc:NP_055668]
ENSG00000167721	1.27	1	(1 citation)	TSR1, 20S rRNA accumulation, homolog [Source:RefSeq_peptide;Acc:NP_060598]
ENSG00000205813	0.73	1	(1 citation)	Olfactory receptor 3A2 (Olfactory receptor 17-228) (OR17-228). [Source:Uniprot/SWISSPROT;Acc:P47893]
ENSG00000180068	0.73	1	(1 citation)	Olfactory receptor 3A4 (Olfactory receptor 17-24) (OR17-24). [Source:Uniprot/SWISSPROT;Acc:P47883]
ENSG00000132359	0.47	1	(1 citation)	GTPase activating Rap/RanGAP domain-like 4 [Source:RefSeq_peptide;Acc:NP_055900]

show Prioritise Genes per phenotype trait

hide Translocations
hide Karyotype

del(17)(p13.2;p13.3){1.5 Mb}

Key:
≥Mean Ratios >0 indicates a DUPLICATION
≥Mean Ratios <0 indicates a DELETION
>Indicates a translocation breakpoint



Rank by score

Phenotype	PAFAH1B1	HIC1	eIF-3	...
CRANIUM, general abnormalities	12.58	-	-	...
Fingers, general abnormalities	1.38	41.95	0.98	...
Lissencephaly/pachygyria	28,54	16.01	-	...
Tonic/clonic seizures	29.41	-	-	...
Mouth	1.45	2.61	8.10	...
...	...	...	...	...
Sum of log scores	**121.02**	**119.64**	**29.62**	...

Σ

Lissencephaly / pachygyria: LNDB 32.32.09
PAFAH1B1 (LIS1): ENSG00000007168
***p*-value** 2.8951×10^{-29}
- log(*p*): 28,54

Example

Home | Centres | Studies | Array Types | Syndromes | Search

Logged in as public

Patient 797

Id	797
Chromosomal Sex	46,XX
Age at initial presentation	28
Consent	Y
Category	affected

Citations (0)

Phenotypes (4)

Primary	Secondary	Tertiary
FEET	Feet, general abnormalities	**Club foot, varus**
JOINTS	Joints, general abnormalities	**Joint laxity**
NEUROLOGY	MENTAL,COGNITIVE FUNCTION, general abnormalities	**Mental retardation/developmental delay**

Example

- 69 genes in deletion, 58 with HUGO name

Aberrations(bps) (1)

Chromosome
5
Start Position(bp)
98954115
End Position(bp)
125355179
Mean Ratio
-1
Classification
de novo - parental origin of rearrangement undefined
Interval(Mb)
26.40
Confirmation
FISH
Karyotype
del(5)(q21.1;q23.2){26.40 Mb}

e! cytoview

Graph | HGNC | OMIM (HGNC) | Imprinted (HGNC) | Ensembl Known
Ensembl Novel | Prioritise All Phenotypes | Prioritise Individual Phenotypes | Overlapping Syndromes | Overlapping Patients

58 listed

ST8SIA4 Chr:5Start:100170803End:100266869
ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 4. Aliases: PST, PST1
Ensembl:ENSG00000113532 Ensembl:ST8SIA4

SLCO4C1 Chr:5Start:101597589End:101660152
solute carrier organic anion transporter family, member 4C1. Aliases: SLC21A20, OATP4C1, OATPX, OATP-H
Ensembl:ENSG00000173930 Ensembl:SLCO4C1

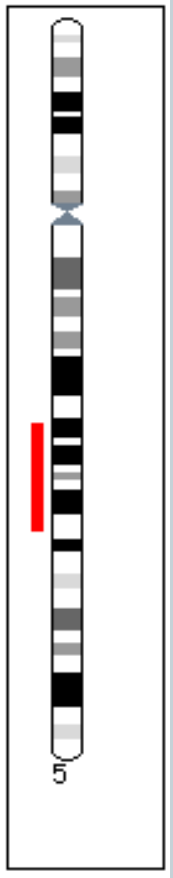
SLCO6A1 Chr:5Start:101735553End:101862619
solute carrier organic anion transporter family, member 6A1. Aliases: OATP6A1, OATPY, MGC26949
Ensembl:ENSG00000205359 Ensembl:SLCO6A1

PAM Chr:5Start:102229422End:102394708
peptidylglycine alpha-amidating monooxygenase. Aliases: PAL, PHM
Ensembl:ENSG00000145730 Ensembl:PAM

GIN1 Chr:5Start:102449603End:102483741
gypsy retrotransposon integrase 1. Aliases: FLJ20125, GIN-1, TGIN1
Ensembl:ENSG00000145723 Ensembl:GIN1

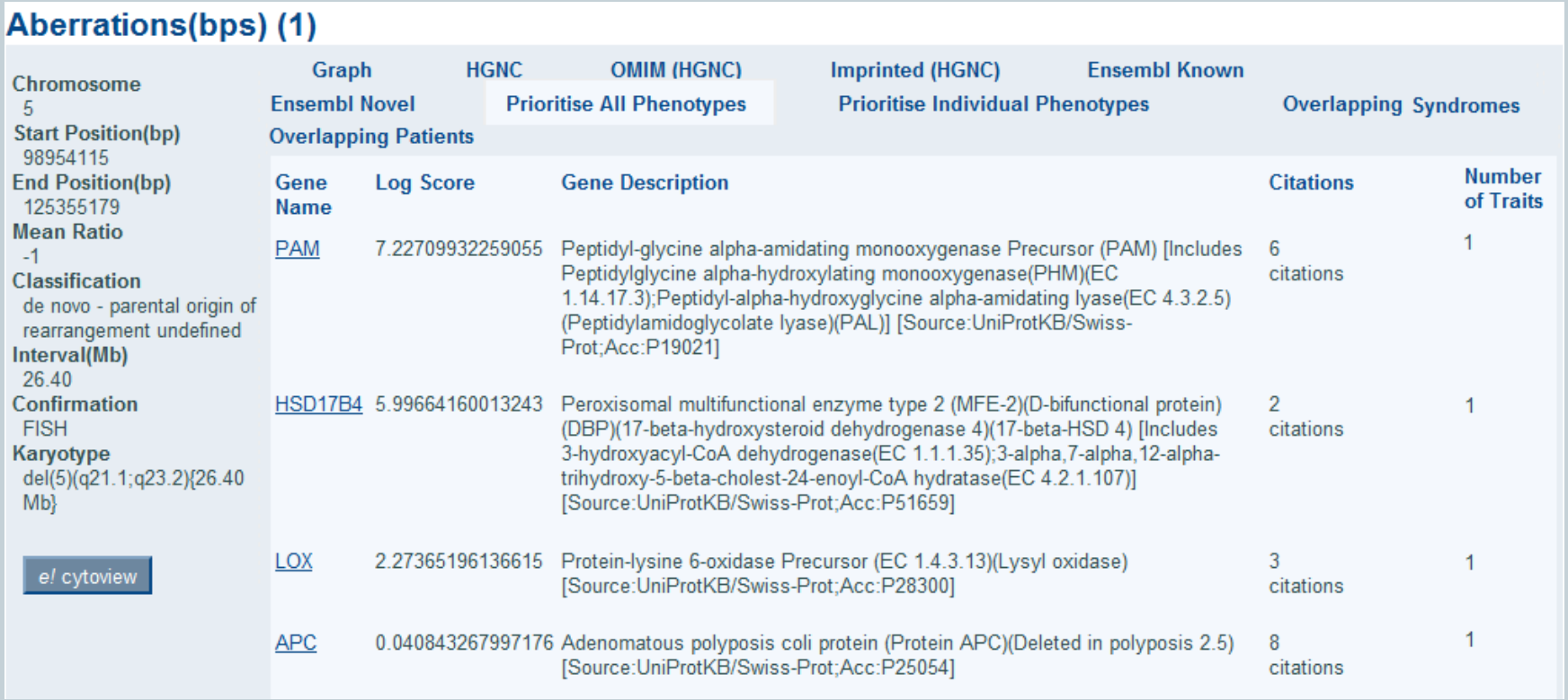
HISPPD1 Chr:5Start:102493156End:102566806
histidine acid phosphatase domain containing 1. Aliases: KIAA0433, VIP2
Ensembl:ENSG00000145725 Ensembl:HISPPD1

C5orf30 Chr:5Start:102622341End:102642260



Example

- 4 genes marked as potentially related to phenotype



Aberrations(bps) (1)

Chromosome
5
Start Position(bp)
98954115
End Position(bp)
125355179
Mean Ratio
-1
Classification
de novo - parental origin of rearrangement undefined
Interval(Mb)
26.40
Confirmation
FISH
Karyotype
del(5)(q21.1;q23.2){26.40 Mb}

e! cytoview

Graph | HGNC | OMIM (HGNC) | Imprinted (HGNC) | Ensembl Known | Ensembl Novel | Prioritise All Phenotypes | Prioritise Individual Phenotypes | Overlapping Syndromes | Overlapping Patients

Gene Name	Log Score	Gene Description	Citations	Number of Traits
PAM	7.22709932259055	Peptidyl-glycine alpha-amidating monooxygenase Precursor (PAM) [Includes Peptidylglycine alpha-hydroxylating monooxygenase(PHM)(EC 1.14.17.3);Peptidyl-alpha-hydroxyglycine alpha-amidating lyase(EC 4.3.2.5) (Peptidylamidoglycolate lyase)(PAL)] [Source:UniProtKB/Swiss-Prot;Acc:P19021]	6 citations	1
HSD17B4	5.99664160013243	Peroxisomal multifunctional enzyme type 2 (MFE-2)(D-bifunctional protein) (DBP)(17-beta-hydroxysteroid dehydrogenase 4)(17-beta-HSD 4) [Includes 3-hydroxyacyl-CoA dehydrogenase(EC 1.1.1.35);3-alpha,7-alpha,12-alpha-trihydroxy-5-beta-cholest-24-enoyl-CoA hydratase(EC 4.2.1.107)] [Source:UniProtKB/Swiss-Prot;Acc:P51659]	2 citations	1
LOX	2.27365196136615	Protein-lysine 6-oxidase Precursor (EC 1.4.3.13)(Lysyl oxidase) [Source:UniProtKB/Swiss-Prot;Acc:P28300]	3 citations	1
APC	0.040843267997176	Adenomatous polyposis coli protein (Protein APC)(Deleted in polyposis 2.5) [Source:UniProtKB/Swiss-Prot;Acc:P25054]	8 citations	1

Example

- Colonic polyps confirmed in patient

Phenotypes (4)

Primary	Secondary	Tertiary
ABDOMEN	Colon, general abnormalities	**Colonic tumours**
FEET	Feet, general abnormalities	**Club foot, varus**
JOINTS	Joints, general abnormalities	**Joint laxity**
NEUROLOGY	MENTAL,COGNITIVE FUNCTION, general abnormalities	**Mental retardation/developmental delay**

Example

- Colonic polyps confirmed in patient

Aberrations(bps) (1)

Chromosome
5
Start Position(bp)
98954115
End Position(bp)
125355179
Mean Ratio
-1
Classification
de novo - parental origin of rearrangement undefined
Interval(Mb)
26.40
Confirmation
FISH
Karyotype
del(5)(q21.1;q23.2){26.40 Mb}

e! cytoview

Graph | HGNC | OMIM (HGNC) | Imprinted (HGNC) | Ensembl Known
Ensembl Novel | Prioritise All Phenotypes | Prioritise Individual Phenotypes | Overlapping
Syndromes | Overlapping Patients

Gene Name	Log Score	Gene Description	Citations	Number of Traits
APC	119.891015131479	Adenomatous polyposis coli protein (Protein APC)(Deleted in polyposis 2.5) [Source:UniProtKB/Swiss-Prot;Acc:P25054]	57 citations	2
LOX	13.1774253190287	Protein-lysine 6-oxidase Precursor (EC 1.4.3.13)(Lysyl oxidase) [Source:UniProtKB/Swiss-Prot;Acc:P28300]	6 citations	2
PAM	7.22709932259055	Peptidyl-glycine alpha-amidating monooxygenase Precursor (PAM) [Includes Peptidylglycine alpha-hydroxylating monooxygenase(PHM)(EC 1.14.17.3);Peptidyl-alpha-hydroxyglycine alpha-amidating lyase(EC 4.3.2.5)(Peptidylamidoglycolate lyase)(PAL)] [Source:UniProtKB/Swiss-Prot;Acc:P19021]	6 citations	1
HSD17B4	5.99664160013243	Peroxisomal multifunctional enzyme type 2 (MFE-2)(D-bifunctional protein)(DBP)(17-beta-hydroxysteroid dehydrogenase 4)(17-beta-HSD 4) [Includes 3-hydroxyacyl-CoA dehydrogenase(EC 1.1.1.35);3-alpha,7-alpha,12-alpha-trihydroxy-5-beta-cholest-24-enoyl-CoA hydratase(EC 4.2.1.107)] [Source:UniProtKB/Swiss-Prot;Acc:P51659]	4 citations	1

Publications

Firth H V, Richards S, Bevan P, Clayton S, Corpas M, Rajan D, **Van Vooren S**, Moreau Y, Pettett R M, Carter N P., **"DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources"** [64]. Am J Hum Genet 84(4):524-33. 2009.

Chapter 4 – Intelligent databases for constitutional cytogenetics

- ‘Downstream analysis’
  - DECIPHER
    - Consortium for aggregation of rare cases
    - Text based gene prioritisation
  - **Store+Bench**
    - **Feature rich platform for downstream analysis**

Cartagenia Bench - Mozilla Firefox

File Edit View History Bookmarks Tools Help

https://darwin.cartagenia.com/demoserver/home.html

Google

Cartagenia Bench

To Folders

To Patients

To Toolbox

Admin

Help

Using: Ensembl v54 - Genome build NCBI 36, UCSC HG18 - DGV March 2009 - PubMed April 2008

Done

darwin.cartagenia.com

Proxy: None

Challenge

Manage all Array-CGH related information, from raw data over called aberrations to patient record.

Support routine clinical diagnostics workflow (2000 ppa) from array scan to final report, as well as ongoing biomedical research projects.

A web based application to store, manage, search, mine, analyze, report and visualize Array CGH-related data.
A joint effort of the Bio-informatics group and the Constitutional Cytogenetics group at the Center for Human Genetics of the University Hospital.

My Reports

Add all 30 cases to the selection

Project	Id	Primary ID	Secondary ID	External ID	Comment	Full text	Clinical	Traits	Keywords	Aberrations	Photo	Consent data	Consent images	In DECIPHER	Actions
chromosome_4_project	52	Patient 984375	HID M20090433	J.D.				7	0	1					
heart_disease_project	29	BC5697489	HumGen234	A.F.G.				6	1	4					
agilent-array_project	28	2462346	TEF3567	J.-P.L.	Heart-related phenotype			0	0	1					
bac-array_project	27	HumGen027	Patient_046223	R.B.	Xq28 (MECP2) duplication	✓		7	0	0					
bac-array_project	26	HumGen026	Patient_042351	I.J.	Wolf-Hirschhorn Syndrome	✓	✓	12	0	1	✓				
bac-array_project	25	HumGen025	Patient_064654	E.T.	Williams-Beuren Syndrome (WBS)			12	0	0					
bac-array_project	24	HumGen024	Patient_06312	S.T.	Sotos syndrome			7	0	1		✓	✓		
bac-array_project	23	HumGen023	Patient_064887	G.S.	Smith-Magenis Syndrome	✓		14	0	0			✓	✓	
bac-array_project	22	HumGen022	Patient_064454	D.N.	Prader-Willi syndrome (Type 1)	✓		7	0	0					
bac-array_project	21	HumGen021	Patient_064554	R.B.	NF1-microdeletion syndrome			6	0	0					
bac-array_project	20	HumGen020	Patient_004654	E.G.	Miller-Dieker syndrome (MDS)			6	0	0					
bac-array_project	19	HumGen019	Patient_004336	E.D.V.	Leri-Weill dyschondrostosis (LWD) - SHOX deletion			4	0	0					
agilent-array_project	18	HumGen018	Patient_007321	C.R.	Cri du Chat Syndrome (5p deletion)	✓		3	0	0					
agilent-array_project	17	HumGen017	Patient_006485	M.T.	Charcot-Marie-Tooth syndrome type 1A (CMT1A)	✓		9	0	0					
agilent-array_project	16	HumGen016	Patient_005461	T.K.	Angelman syndrome (Type 2)			6	0	0					
bac-array_project	15	HumGen015	Patient_000677	T.H.	Angelman syndrome (Type 1)			5	0	0					
bac-array_project	14	HumGen014	Patient_000047	P.G.R.	Adult-onset autosomal dominant leukodystrophy (ADLD)			0	0	0					
bac-array_project	13	HumGen013	Patient_000541	P.G.	9q subtelomeric deletion syndrome			0	0	0					
bac-array_project	12	HumGen012	Patient_009874	R.T.	8p23.1 deletion syndrome			0	0	0					
bac-array_project	11	HumGen011	Patient_002214	Y.M.	7q11.23 duplication syndrome			0	0	0					
bac-array_project	10	HumGen010	Patient_026497	P.R.	3q29 microdeletion syndrome	✓		0	0	0					
bac-array_project	9	HumGen009	Patient_004667	J.U.	2q37 monosomy			0	0	0		✓		✓	
bac-array_project	8	HumGen008	Patient_009754	W.H.	2q32.2 deletion syndrome			0	0	0					
bac-array_project	7	HumGen007	Patient_005466	E.W.	22q13 deletion syndrome (Phelan-Mcdermid syndrome)	✓		8	0	0					
bac-array_project	6	HumGen006	Patient_00457	G.L.	22q11 deletion syndrome (Velocardiofacial / DiGeorge syndrome)			8	0	0		✓	✓	✓	

Show 25

1 2 Next

Click here for help

Per project...
- agilent-array_project
- bac-array_project
- chromosome_4_project
- heart_disease_project
- INSERM

Show all...

Current selection
Selection is empty.

Actions on selection
- Save this selection
- Load existing selection
- Export to Excel
- Submit to database
- Draw karyogram
- Tally phenotype traits
- Tally genome variations
- Find similar cases
- Find related literature
- Find related diseases
- Cluster within selection

List view | Graphical view Upcoming

Label	Chromosome:start-stop	Size	Platform	Inheritance	Overlap with known CNVs: Redon WGTP	Redon 500K	DGV Regions	DGV Loci	Info	Genome browsers	
Not specified	[Deletion] **3**: 37979297 - 39113353	1,134.1 Kb	Agilent Oligo aCGH	Not specified	25%	0%	25%	25%			
Not specified	[Deletion] **3**: 45564680 - 46874572	1,309.9 Kb	Agilent Oligo aCGH	Not specified	27%	15%	27%	27%			
Not specified	[Multiplication] **4**: 76939086 - 78445285	1,506.2 Kb	Agilent Oligo aCGH	Not specified	0%	0%	18%	22%			
Not specified	[Deletion] **6**: 149439019 - 150662525	1,223.5 Kb	Agilent Oligo aCGH	Not specified	0%	0%	15%	15%			
Not specified	[Multiplication] **7**: 29801785 - 30737241	935.5 Kb	Agilent Oligo aCGH	Not specified	0%	0%	17%	16%			
Not specified	[Multiplication] **9**: 34177516 - 35599496	1,422 Kb	Agilent Oligo aCGH	Not specified	0%	0%	19%	19%			
Not specified	[Multiplication] **9**: 126263478 - 131429534	5,166.1 Kb	Agilent Oligo aCGH	Not specified	5%	4%	8%	17%			
Not specified	[Deletion] **10**: 111883181 - 116181061	4,297.9 Kb	Agilent Oligo aCGH	Not specified	0%	0%	10%	10%			

Deletion: -9.4
Amplification: 0.0
Cytoband: q25.2 - q25.3
Number of probes: 15.0

Not specified	[Deletion] **10**: 125496067 - 128256101	2,760 Kb	Agilent Oligo aCGH	Not specified	0%	0%	22%	22%			
Not specified	[Deletion] **11**: 2861232 - 4362913	1,501.7 Kb	Agilent Oligo aCGH	Not specified	55%	23%	55%	55%			
Not specified	[Multiplication] **11**: 62935488 - 67599022	4,663.5 Kb	Agilent Oligo aCGH	Not specified	13%	0%	16%	23%			
Not specified	[Multiplication] **11**: 71607196 - 74354785	2,747.6 Kb	Agilent Oligo aCGH	Not specified	39%	41%	41%	41%			
Not specified	[Multiplication] **14**: 19984867 - 24351391	4,366.5 Kb	Agilent Oligo aCGH	Not specified	11%	4%	31%	32%			
Not specified	[Multiplication] **15**: 22774234 - 23153211	379 Kb	Agilent Oligo aCGH	Not specified	0%	0%	44%	41%			

Disease	Start band	End band	OMIM	Gene symbol
Angelman syndrome	15q11.2	15q11.2	**105830**	**UBE3A**
Prader-Willi syndrome	15q11.2	15q11.2	**176270**	**SNRPN**
15q11 BP1-BP2 interval	15q11.2	15q11.2		**NIPA1 NIPA2**

Not specified	[Deletion] **15**: 49912495 - 50836651	924.2 Kb	Agilent Oligo aCGH	Not specified	0%	0%	0%	0%			
Not specified	[Multiplication] **17**: 30618145 - 36046391	5,428.2 Kb	Agilent Oligo aCGH	Not specified	28%	20%	28%	29%			
Not specified	[Multiplication] **17**: 43360826 - 51209655	7,848.8 Kb	Agilent Oligo aCGH	Not specified	6%	1%	6%	7%			
Not specified	[Multiplication] **17**: 53405918 - 56096367	2,690.4 Kb	Agilent Oligo aCGH	Not specified	23%	0%	23%	23%			
Not specified	[Multiplication] **17**: 76829283 - 78497822	1,668.5 Kb	Agilent Oligo aCGH	Not specified	11%	0%	11%	49%			
Not specified	[Multiplication] **20**: 32331534 - 34005593	1,674.1 Kb	Agilent Oligo aCGH	Not specified	0%	0%	0%	0%			
Not specified	[Multiplication] **20**: 42561114 - 43912664	1,351.6 Kb	Agilent Oligo aCGH	Not specified	0%	0%	4%	4%			
Not specified	[Multiplication] **20**: 45206149 - 47999390	2,793.2 Kb	Agilent Oligo aCGH	Not specified	17%	0%	17%	43%			
Not specified	[Multiplication] **20**: 51537918 - 57042885	5,505 Kb	Agilent Oligo aCGH	Not specified	0%	0%	17%	17%			

Done Cancel

Regions waived from exclusion

The following regions are waived from exclusion and will never be filtered out

Challenge

Describe patient phenotypes in a consistent way, ensuring easy case report retrieval (search).

Enable data mining for detection of similar patients, and clusters of patients with similar phenotypes as putative novel syndromes.

Use standard phenotype definitions (ontologies)

- Barraitser & Winter LNDB (human dysmorphy)
- AEPC codes (heart specific phenotypes)
- POSSUM codes (malformations & syndromes)

Extensible: allow any vocabulary to be plugged in (UMLS, ICD10, …)

Patients v4.0

Zoeken Patientrapport volgens Identificatie Start
Geavanceerd zoeken

Starten | Naar Folders | Naar Gereedschapskist | Nieuwe functionaliteiten
Mijn rapporten | Karyogram | Exporteer naar Excel | Beheer | Hulp | Instellingen | Afmelden
Nieuw | Toon | Rapporten
Rapport 'HumGen027' is actief voor bewerking. Aangemeld als svanvoor.

Casus: 433 - HumGen027 - Patient_046223 - R.B.

Gemaakt door: bcoessen Op Apr 22, 2009
Laatst gewijzigd door: svanvoor Op Sep 25, 2009
In project: BAC array project
Commentaar: Xq28 (MECP2) duplication

Overzicht | Beschrijving | Klinische gegevens | **Fenotype** | Genotype | Beelden | Verslagen

Patient phenotype information

Phenotype traits

LDDBLDDB:29.01.12 **Recurrent infections**
LDDBLDDB:32.19.07 **Hypotonia (myopathic)**
LDDBLDDB:32.06.00 **SEIZURES, general abnormalities**
LDDBLDDB:14.01.06 **Speech defect / dysarthria**
LDDBLDDB:32.04.05 **Hypotonia (non-myopathic)**
LDDBLDDB:32.17.06 **Spasticity / brisk reflexes / Babinski**
LDDBLDDB:32.08.05 **Microcephaly**
LDDBLDDB:32.04.02 **Mental retardation / developmental delay**
LDDBLDDB:07.04.00 **Cornea, general abnormalities**

Onset and clinical course

HPO OnsetHP:0003662 Adult onset has been reported

Fenotypes opslaan | Afbreken

Zoek phenotype

LDDB POSSUM HPO

microc

[LDDB] - Microcephaly
[LDDB] - Microcornea
[POSSUM] - Microcephaly
[POSSUM] - Microcornea
[HPO] - Alpha-thalassemia with microcytosis
[HPO] - hypochromic, microcytic anemia
[HPO] - Microcephaly
[HPO] - Microcephaly, acquired
[HPO] - Microcephaly, mild

Zoek verdere omschrijving

HPO Onset

d

[HPO Onset] - Adult onset has been reported
[HPO Onset] - Age-dependent penetrance
[HPO Onset] - Death in childhood
[HPO Onset] - Death in early childhood
[HPO Onset] - Death in infancy
[HPO Onset] - Death in majority of infants soon after birth
[HPO Onset] - Death in neonatal period or infancy
[HPO Onset] - Early death
[HPO Onset] - Early onset, mild and relatively uncomplicated course

Challenge

Slice and dice, report and visualize my cytogenetic case report database.
See patients in context (i.e. my own database, or genome browsers such as ENSEMBL).
Find candidate genes that are causal for the phenotype of my patient.

Similar patient detection by genotype and phenotype information

Gene-to-phenotype correlation based on analysis of biomedical literature (PUBMED)

Case report: 26 - HumGen026 - Patient_042351 - I.J.

Created by: bcoessen on Apr 15, 2009
Last updated by: bcoessen on May 26, 2009
In project: bac-array_project
Comment: Wolf-Hirschhorn Syndrome

Overview | Description | Clinical data | **Phenotype** | Genotype | Pictures | Reporting

Structured phenotype description

Add / Remove

Code	Label	Description	Association
LNDB.32.16.02	Truncal ataxia / non-cerebellar ataxia	Some ataxias in the literature are not typically cerebellar, but use this alone with caution. Rather combine with cerebellar tremor if uncertain.	positive
AEPC.14.03.33	Microcephaly	Microcephaly	positive
LNDB.32.04.00	MENTAL,COGNITIVE FUNCTION, general abnormalities	no detailed description for [MENTAL,COGNITIVE FUNCTION, general abnormalities]	positive
LNDB.01.03.02	Low birthweight (< 3rd centile)	Below the 3rd centile for gestational age.	positive
LNDB.32.06.00	SEIZURES, general abnormalities	no detailed description for [SEIZURES, general abnormalities]	positive
LNDB.32.08.00	CRANIUM, general abnormalities	no detailed description for [CRANIUM, general abnormalities]	positive
LNDB.32.04.05	Hypotonia (non-myopathic)	Floppiness not due to muscle disease or peripheral nerve disease, but whose origin is central (mostly in the brain).	positive
LNDB.32.04.02	Mental retardation / developmental delay	Includes mental handicap, learning disabled, but not dementia. It is not progressive and the delay should be present from birth.	positive
LNDB.32.08.05	Microcephaly	A reduction in head circumference of greater than 2SD below the mean. Parental measurements must be taken into account.	positive
POSSUM.186	Microcephaly		positive
LNDB.15.01.02	Goitre	Enlarged thyroid gland. See under 'endocrine - thyroid' if associated with hyper or hypothyroidism. Use together.	positive
LNDB.02.00.00	STATURE	no detailed description for [STATURE]	positive

Additional freetext keywords

Edit

Category	Text

This patient does not yet have keywords.

Cases with similar phenotype

- 52 - Patient 984375 - HID M20090433 - J.D.
- 15 - HumGen015 - Patient_000677 - T.H.
- 16 - HumGen016 - Patient_005461 - T.K.

Show all...

Bench v4.0

Search Case Report by Identifier Go

Case report: 26 - HumGen026 - Patient_042351 - I.J.

Created by: bcoessen on Apr 15, 2009
Last updated by: svanvoor on Jun 22, 2009
In project: bac-array_project
Comment: Wolf-Hirschhorn Syndrome

Overview | Description | Clinical data | Phenotype | Genotype | Pictures | Reporting

Chromosomal aberrations

Add Aberration

List view | Graphical view Upcoming

Label	Chromosome:start-stop	Size	Source	Platform	Inheritance	Cause of aberration	Confirmation	Flanking	Overlap with known CNVs: Redon WGTP	Redon 500K	DGV Regions	DGV Loci	Info	Genome browsers	Actions
Not specified	[Deletion] 4: 55628 - 2043467	1,987.8 Kb	Bench	BAC/PAC Array (Sanger clone sets)	De novo	Not specified	FISH	[]	28%	0%	28%	39%			

Disease	Start band	End band	OMIM	Gene symbol
Wolf-Hirschhorn syndrome	4p16.3	4p16.3	194190	WHSC1 WHSC2

Gene prioritisation based on phenotype descriptions

Warning: currently only general (LNDB) phenotypes are supported for gene prioritisation.
Note that negative associations are excluded by default.

Select: All - None

LNDB.11.04.03 - Prominent upper lip (association: positive)
LNDB.11.04.02 - Midline cleft upper lip (association: positive)
LNDB.32.06.00 - SEIZURES, general abnormalities (association: positive)
LNDB.32.08.00 - CRANIUM, general abnormalities (association: positive)
LNDB.32.04.05 - Hypotonia (non-myopathic) (association: positive)
LNDB.32.04.02 - Mental retardation / developmental delay (association: positive)
LNDB.32.08.05 - Microcephaly (association: positive)
LNDB.02.00.00 - STATURE (association: positive)

Minimal evidence for a gene-to-phenotype link: 2 abstracts

Prioritise by selected phenotype traits

Actions on region set
Upcoming:
Bulk edit regions
Export to Excel

Cases with similar genotype
28 - 2462346 - TEF3567 - J.-P.L.
52 - Patient 984375 - HID M20090433 - J.D.
24 - HumGen024 - Patient_06312 - S.T.
Show all...

Close

Gene prioritization result: 26 - HumGen026 - Patient_042351 - I.J.

Created by: bcoessen on Apr 15, 2009
In project: bac-array_project
Comment: Wolf-Hirschhorn Syndrome

Top ten of genes in affected regions

Score	Gene symbol	Type	Location	OMIM	Description	Evidence
180.41	**FGFR3**	protein_coding	4 (1):1764832-1780396	**134934**	Fibroblast growth factor receptor 3 Precursor (FGFR-3)(EC 2.7.10.1)(CD333 antigen) [Source:UniProtKB/Swiss-Prot;Acc:P22607]	**110 abstracts**
49.66	**WHSC2**	protein_coding	4 (-1):1954248-1981635	**606026**	Negative elongation factor A (NELF-A)(Wolf-Hirschhorn syndrome candidate 2 protein) [Source:UniProtKB/Swiss-Prot;Acc:Q9H3P2]	**12 abstracts**

Code	Label	Description	Go to PubMed
LNDB.32.06.00	SEIZURES, general abnormalities	no detailed description for [SEIZURES, general abnormalities]	**3 abstracts**
LNDB.32.04.02	Mental retardation / developmental delay	Includes mental handicap, learning disabled, but not dementia. It is not progressive and the delay should be present from birth.	**3 abstracts**
LNDB.02.00.00	STATURE	no detailed description for [STATURE]	**2 abstracts**
LNDB.11.04.02	Midline cleft upper lip	Often a small V-shaped indentation in the mid-line.	No evidence found
LNDB.32.04.05	Hypotonia (non-myopathic)	Floppiness not due to muscle disease or peripheral nerve disease, but whose origin is central (mostly in the brain).	**1 abstracts**
LNDB.32.08.05	Microcephaly	A reduction in head circumference of greater than 2SD below the mean. Parental measurements must be taken into account.	**2 abstracts**
LNDB.32.08.00	CRANIUM, general abnormalities	no detailed description for [CRANIUM, general abnormalities]	**2 abstracts**
LNDB.11.04.03	Prominent upper lip	A projecting upper lip. Use with thick (see below) if necessary.	No evidence found

39.60	**LETM1**	protein_coding	4 (-1):1784558-1827772	**604407**	LETM1 and EF-hand domain-containing protein 1, mitochondrial Precursor (Leucine zipper-EF-hand-containing transmembrane protein 1) [Source:UniProtKB/Swiss-Prot;Acc:O95202]	**9 abstracts**
36.35	**WHSC1**	protein_coding	4 (1):1842921-1953732	**602952**	Probable histone-lysine N-methyltransferase NSD2 (EC 2.1.1.43)(Nuclear SET domain-containing protein 2)(Wolf-Hirschhorn syndrome candidate 1 protein)(Multiple myeloma SET domain-containing protein)(Protein trithorax-5) [Source:UniProtKB/Swiss-Prot;Acc:O96028]	**11 abstracts**
14.22	**IDUA**	protein_coding	4 (1):970785-988316	**252800**	Alpha-L-iduronidase Precursor (EC 3.2.1.76) [Source:UniProtKB/Swiss-Prot;Acc:P35475]	**5 abstracts**

Top ten of all genes

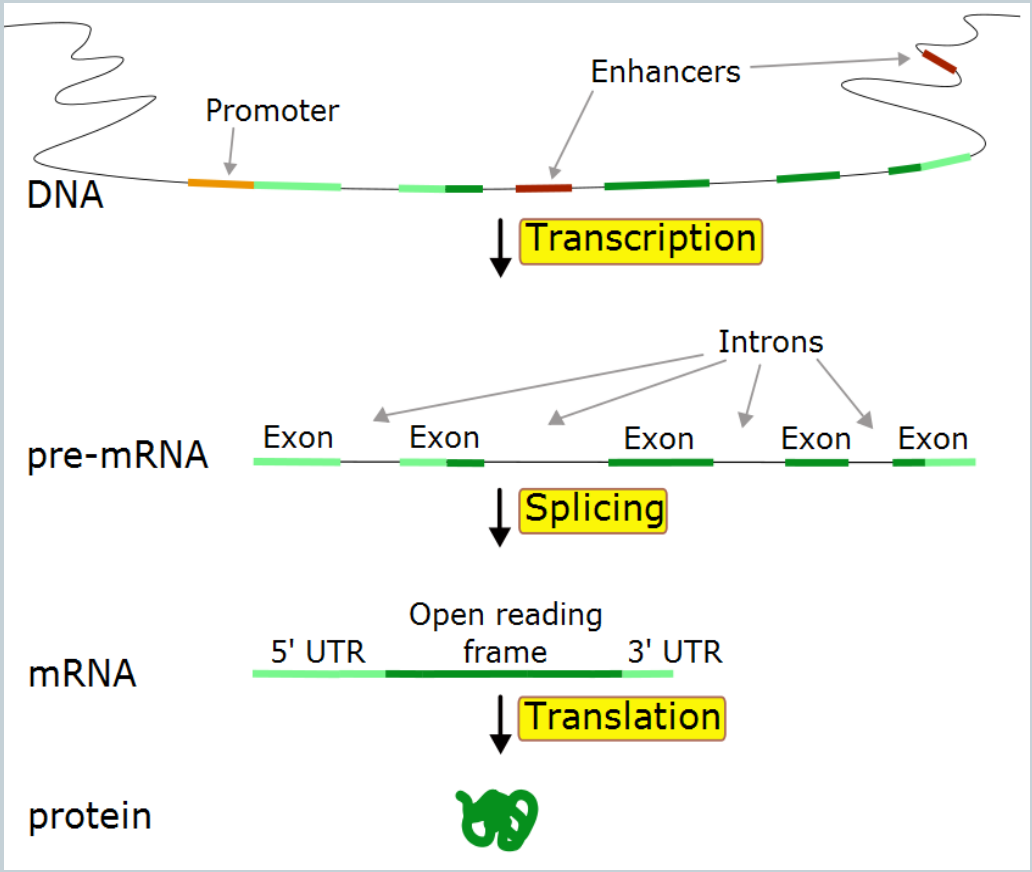
Score	Gene symbol	Type	Location	OMIM	Description	Evidence
377.92	**MTSS1**	protein_coding	8 (-1):125632212-125809911	**608486**	Metastasis suppressor protein 1 (Missing in metastasis protein)(Metastasis suppressor YGL-1) [Source:UniProtKB/Swiss-Prot;Acc:O43312]	**260 abstracts**
293.47	**S100A10**	protein_coding	1 (-1):150222011-150233338	**114085**	Protein S100-A10 (S100 calcium-binding protein A10)(Calpactin-1 light chain)(Calpactin I light chain)(p10 protein)(p11)(Cellular ligand of annexin II) [Source:UniProtKB/Swiss-Prot;Acc:P60903]	**276 abstracts**
228.60	**MECP2**	protein_coding	X (-1):152940218-153016406	**300005**	Methyl-CpG-binding protein 2 (MeCP-2 protein)(MeCP2) [Source:UniProtKB/Swiss-Prot;Acc:P51608]	**160 abstracts**
215.50	**DCTN3**	protein_coding	9 (-1):34603549-34610496	**607387**	Dynactin subunit 3 (Dynactin complex subunit 22 kDa subunit)(p22) [Source:UniProtKB/Swiss-Prot;Acc:O75935]	**155 abstracts**
208.20	**ARX**	protein_coding	X (-1):24932213-24943775	**300382**	Homeobox protein ARX (Aristaless-related homeobox) [Source:UniProtKB/Swiss-Prot;Acc:Q96QS3]	**81 abstracts**
206.19	**PQBP1**	protein_coding	X (1):48640139-48645363	**300463**	Polyglutamine-binding protein 1 (Polyglutamine tract-binding protein 1)(PQBP-1)(38 kDa nuclear protein containing a WW domain)(Npw38) [Source:UniProtKB/Swiss-Prot;Acc:O60828]	**45 abstracts**
197.80	**UBE3A**	protein_coding	15 (-1):23133489-23235221	**601623**	Ubiquitin-protein ligase E3A (EC 6.3.2.-)(E6AP ubiquitin-protein ligase)(Oncogenic protein-associated protein E6-AP)(Human papillomavirus E6-associated protein)(Renal carcinoma antigen NY-REN-54) [Source:UniProtKB/Swiss-Prot;Acc:Q05086]	**68 abstracts**
184.23	**ZEB2**	protein_coding	2 (-1):144862055-144994386	**605802**	Zinc finger E-box-binding homeobox 2 (Zinc finger homeobox protein 1b)(Smad-interacting protein 1)(SMADIP1) [Source:UniProtKB/Swiss-Prot;Acc:O60315]	**56 abstracts**
183.67	**MCPH1**	protein_coding	8 (1):6251529-6493434	**607117**	Microcephalin [Source:UniProtKB/Swiss-Prot;Acc:Q8NEM0]	**56 abstracts**
182.66	**CDK5R1**	protein_coding	17 (1):27838214-27842383	**603460**	Cyclin-dependent kinase 5 activator 1 Precursor (CDK5 activator 1)(Cyclin-dependent kinase 5 regulatory subunit 1)(Tau protein kinase II 23 kDa subunit)(p23)(TPKII regulatory subunit) [Contains Cyclin-dependent kinase 5 activator 1, p35(p35);Cyclin-dependent kinase 5 activator 1, p25(p25)] [Source:UniProtKB/Swiss-Prot;Acc:Q15078]	**144 abstracts**

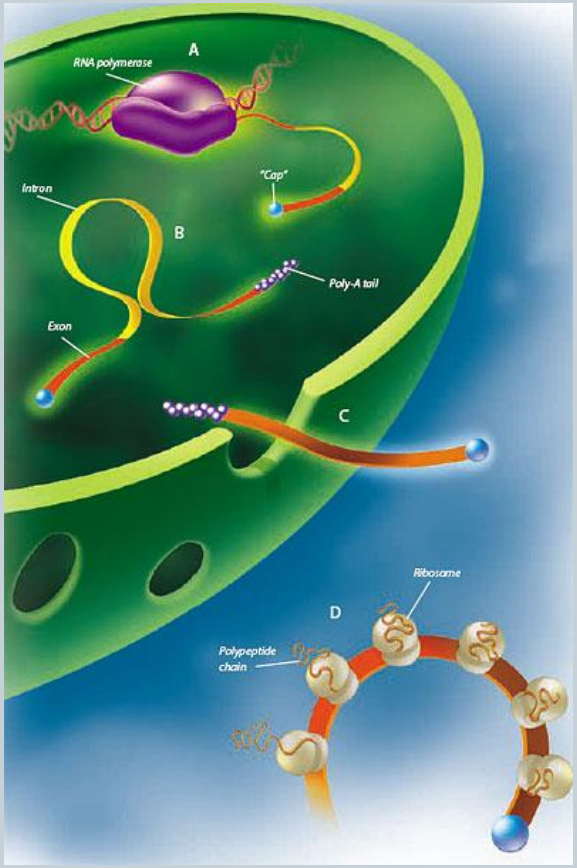
Chapter 5 – genome annotation

DISTRIBUTED ANNOTATION AND *CIS*-REGULATORY ELEMENTS: USING INFORMATION FROM LITERATURE

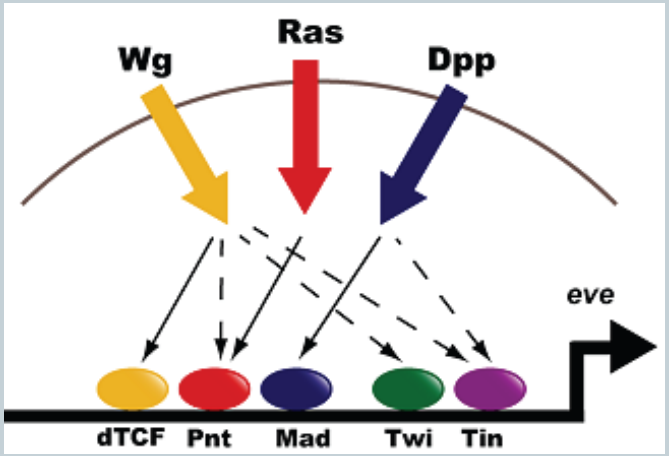
Regulatory elements

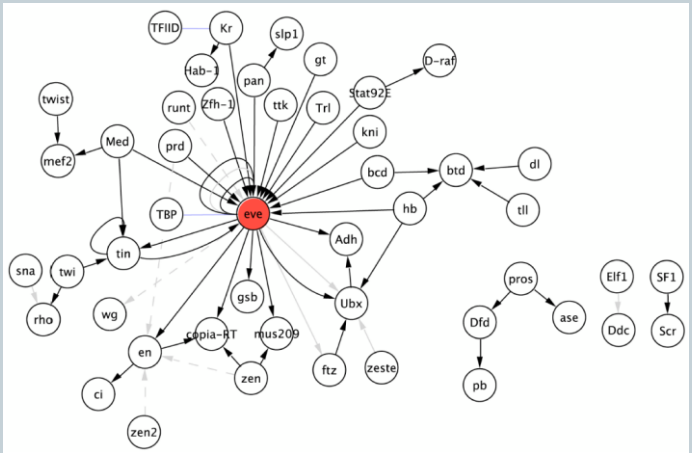
- Transcription regulation: gene activity switches





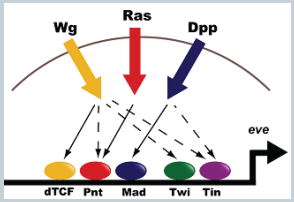
cis-regulatory elements

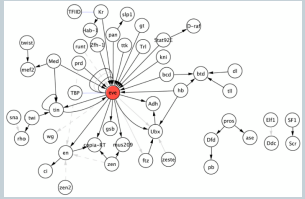




- Vast amount of discovered regulatory data remains locked in biomedical literature
- *Ad hoc* efforts: small teams curate organism- or process-specific datasets from the primary literature for short-term research purposes

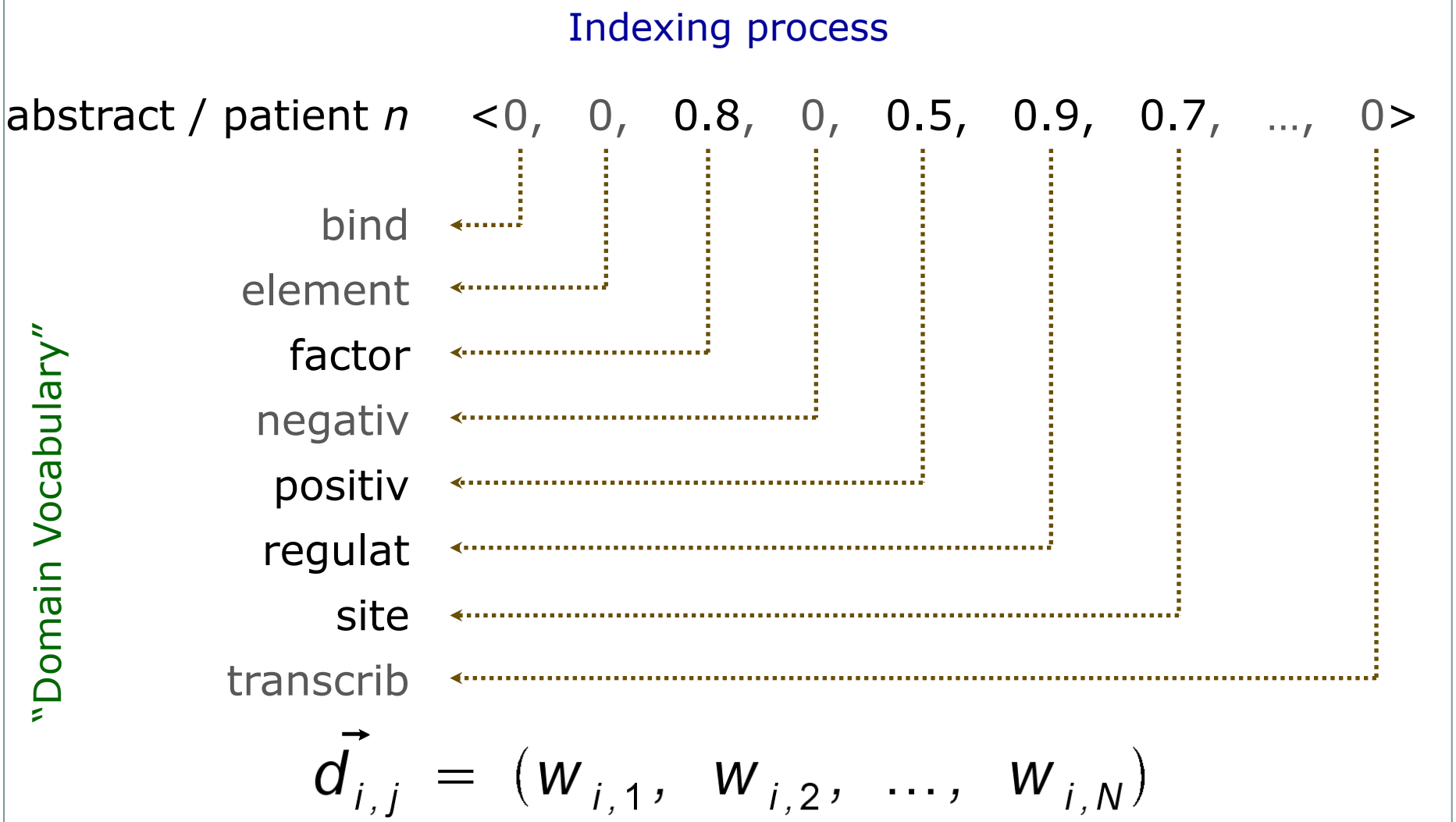
cis-regulatory elements





- ORegAnno and PAZAR
  - Manually curated database of regulatory elements
  - community annotation of regulatory data
- Population of curation pipeline by text-mining queue
  - text-mining to retrieve and extract relevant documents

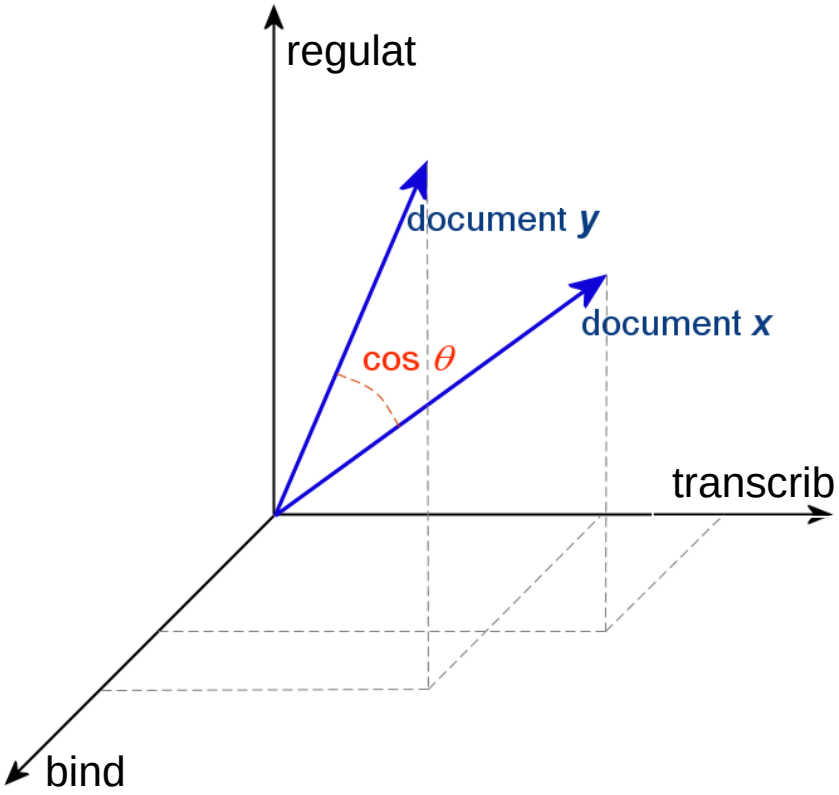
Vector Space Model



$$\vec{d}_{i,j} = (w_{i,1}, w_{i,2}, \ldots, w_{i,N})$$

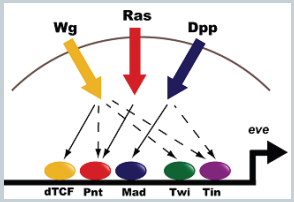
Vector Space Model: Similarity

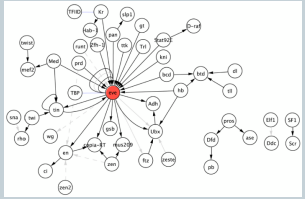
Ranking and scoring documents



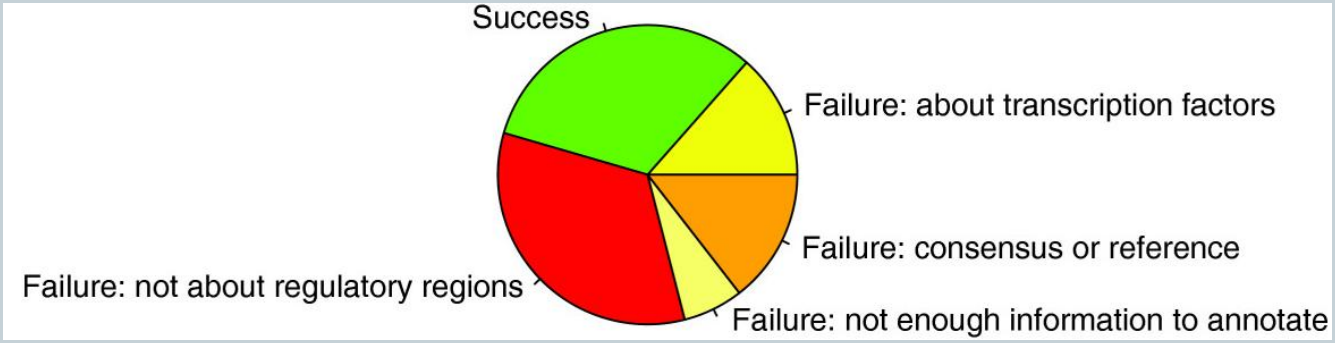
$$sim_{vs}(q, x_i) = Q \cdot X_i = \frac{\sum_{j=1}^{m} v_j \cdot w_{ij}}{\sqrt{\sum_{j=1}^{m} (v_j)^2 \cdot \sum_{j=1}^{m} (w_{ij})^2}}$$

Cis-regulatory elements





- ORegAnno and PAZAR
  - Manually curated database of regulatory elements
  - community annotation of regulatory data
- Population of curation pipeline by text-mining queue
  - text-mining to retrieve and extract relevant documents



Publications

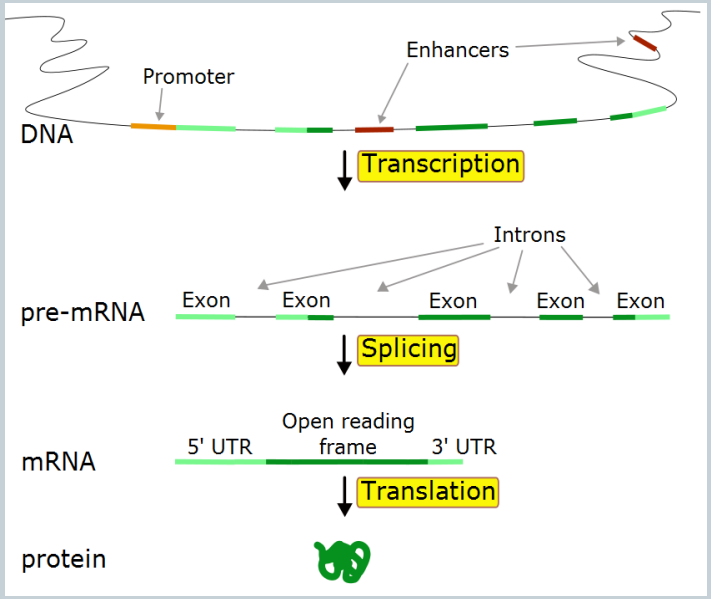
Aerts S., Haeussler M., **Van Vooren S.**, Grith O.L., Hulpiau P., Jones S.J., Montgomery S.B., Bergman C.M., The Open Regulatory Annotation Consortium, **"Text-mining assisted regulatory annotation"** Genome Biol. 9(R31):1-13, 2008.

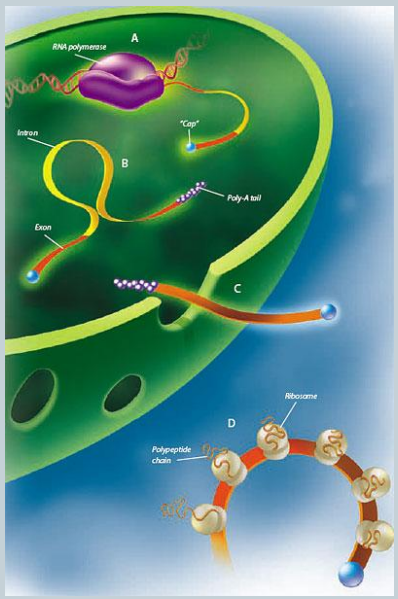
Chapter 6 – data fusion on biomedical text and gene expression arrays

TEXT DATA AS VALIDATION OF GENE CLUSTERS AND AS PRIOR FOR BAYESIAN NETWORKS

Chapter 6 – data fusion on biomedical text and gene expression arrays

- Background

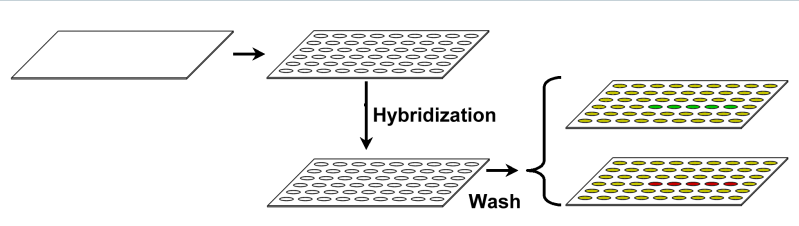


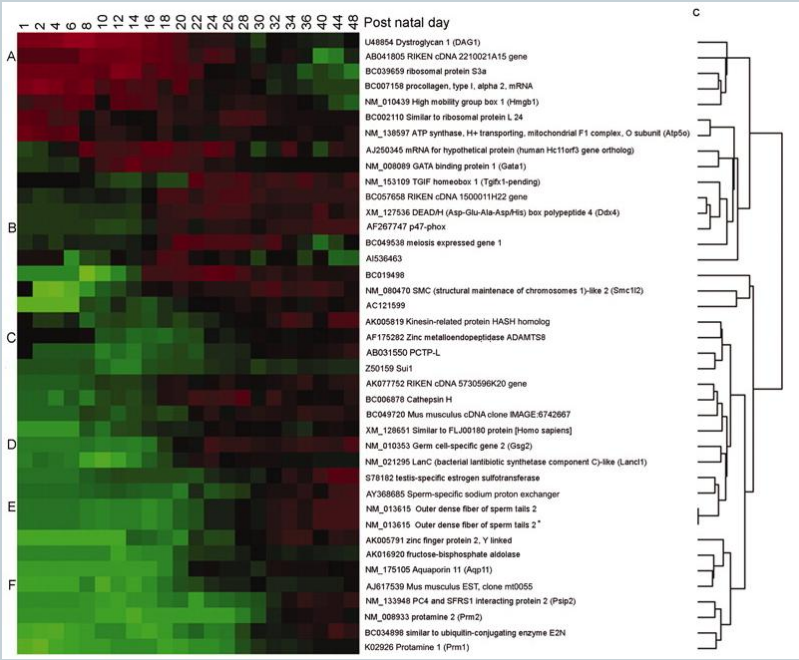


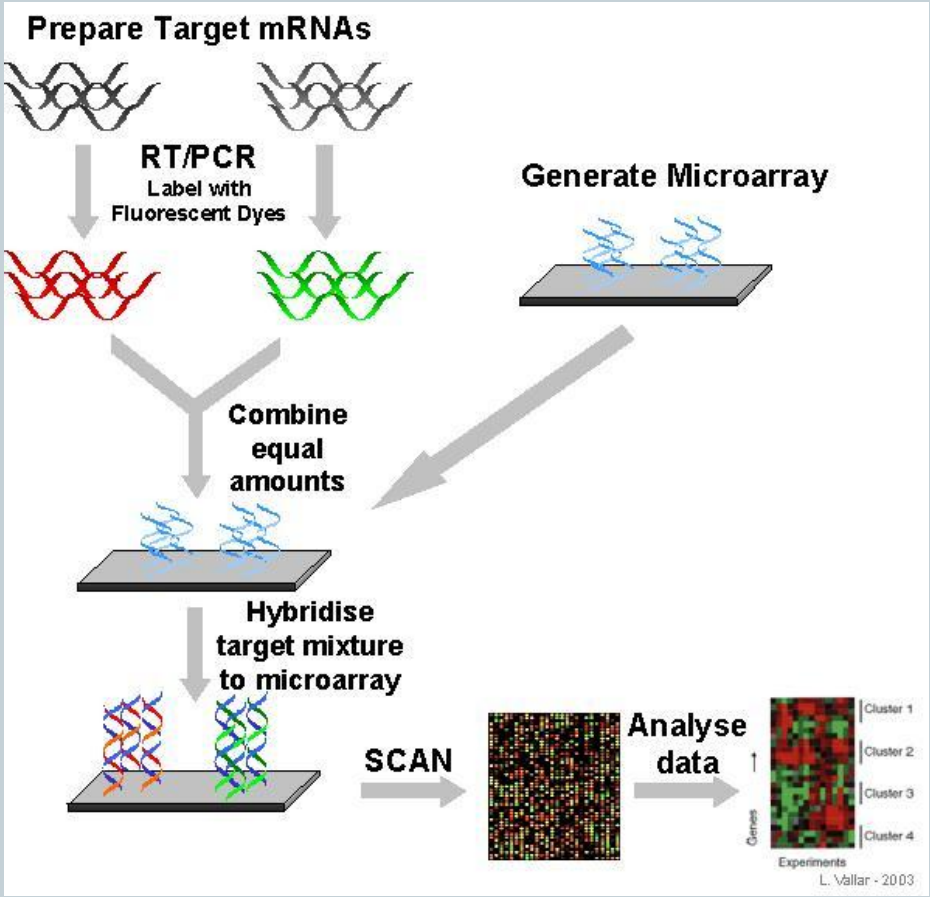
Smith *et al* - Noggin: important factor in brain and nerve development

Chapter 6 – data fusion on biomedical text and gene expression arrays

- Background







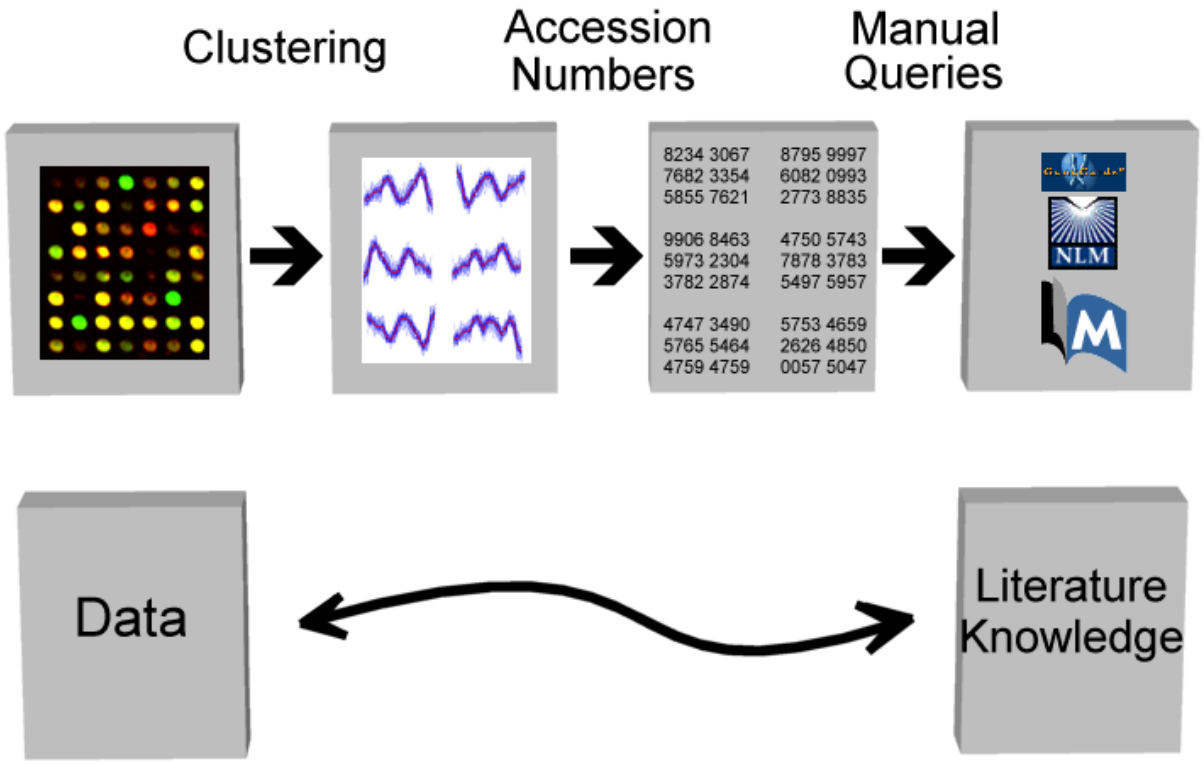
Chapter 6 – data fusion on biomedical text and gene expression arrays

- Application: expression arrays
  - Co-expressed genes: cluster validation
    - TXTGate overview
    - Validation
- Application: text data as Bayesian network prior
  - Context
  - results

Chapter 6 – data fusion on biomedical text and gene expression arrays

- Application: expression arrays
  - Co-expressed genes: cluster validation
    - **TXTGate overview**
    - **Validation**
- Application: text data as Bayesian network prior
  - Context
  - results

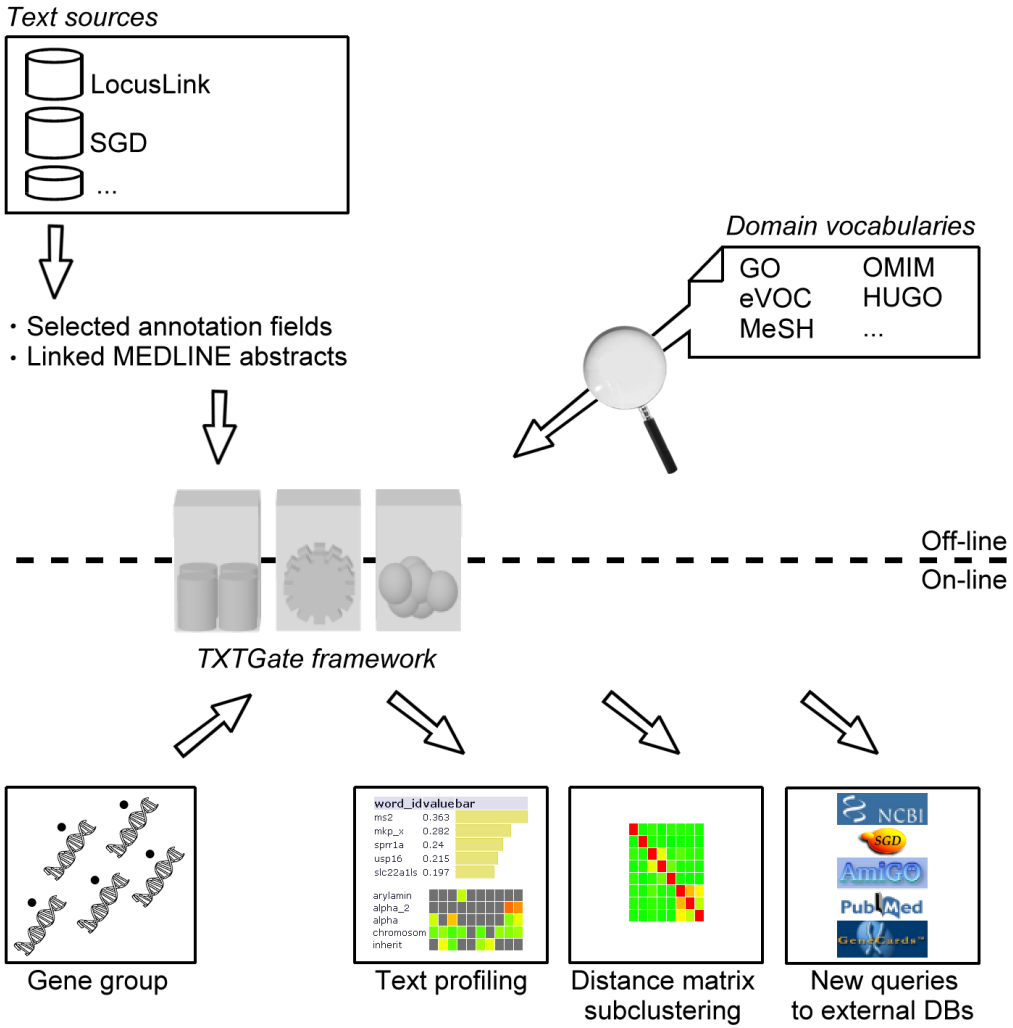
What is TXTGate?



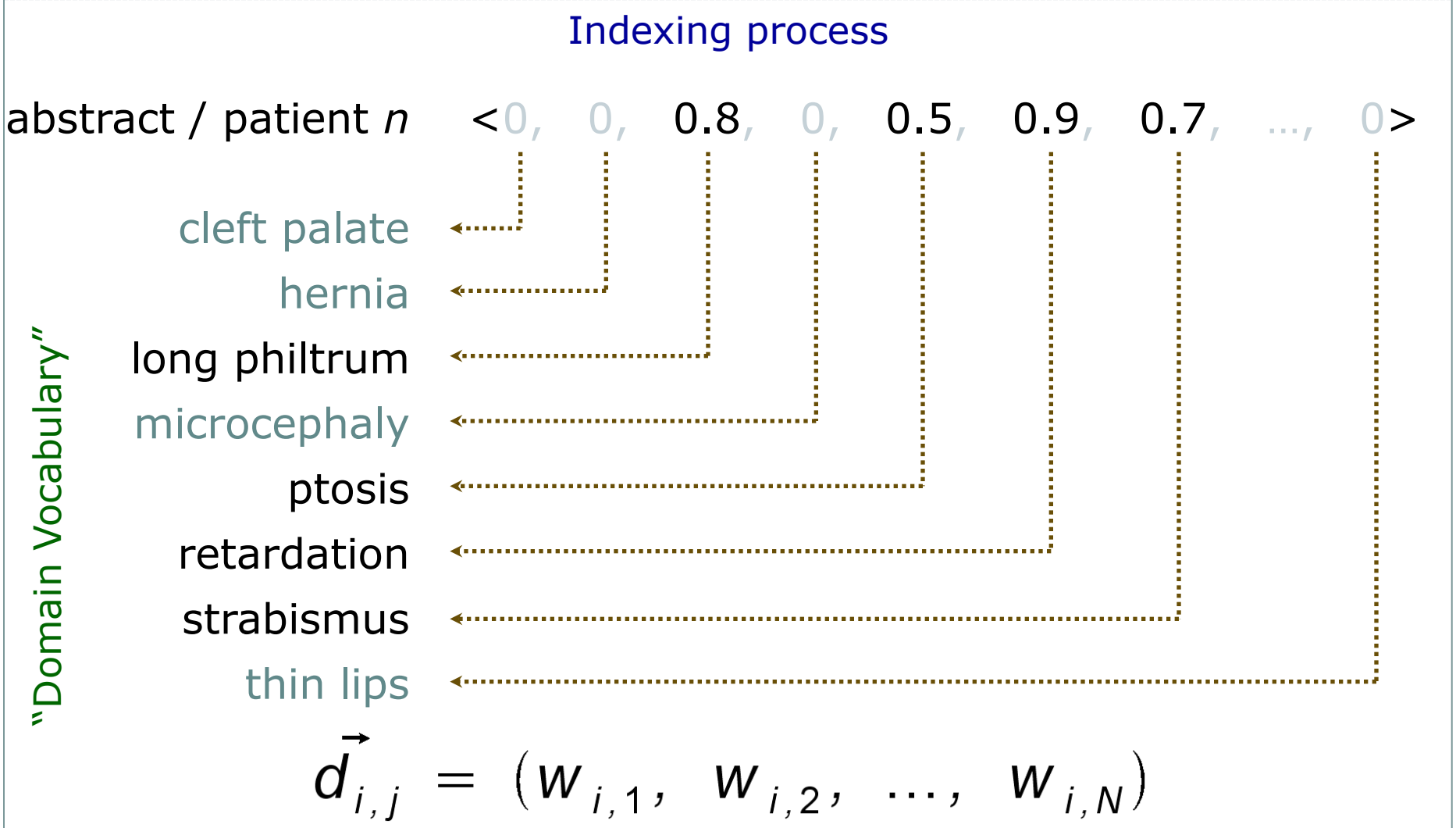
TXTGate:

Gene Cluster Analysis and Profiling Tool

Functional Overview



Vector Space Model



$$\vec{d}_{i,j} = (w_{i,1}, w_{i,2}, \ldots, w_{i,N})$$

Creating an index: details

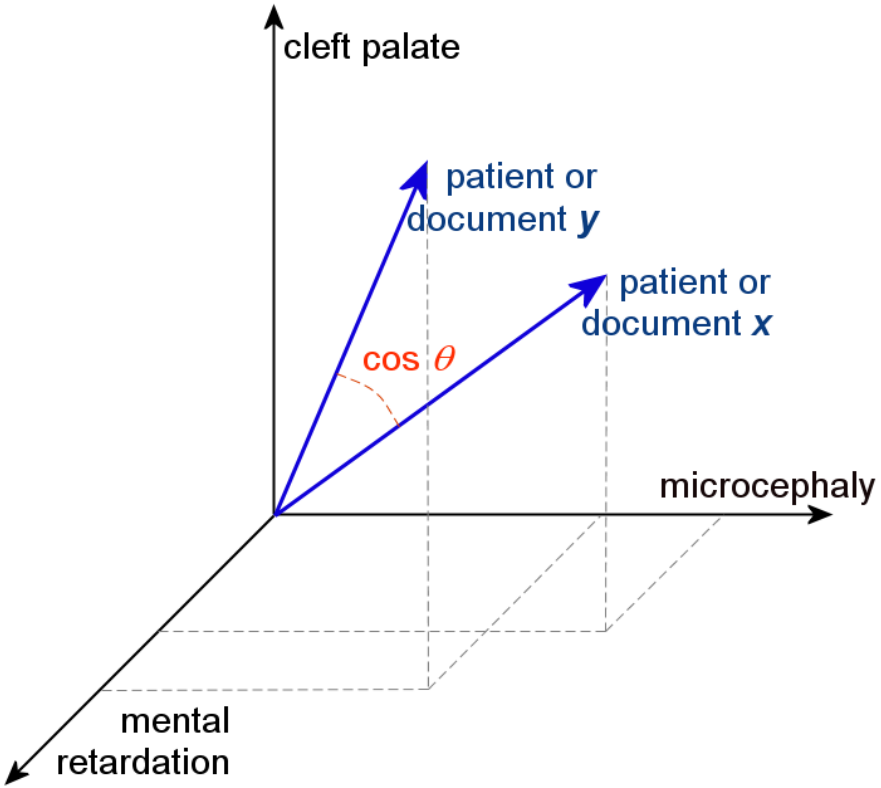
Creating an index involves:

- Stemming
  - treat “methylate”, “methylated”, “methylation” as one
- Remove stop words
  - “the”, “for”, “is” are not significant
- Construct phrases
  - “nucleic” and “acid” should be read as “nucleic_acid”
- Detect synonyms
  - “p53”, “tp53” should be treated as being identical
- Choose Weighting Scheme

$$tf_{i,j} = \frac{f_{i,j}}{\max_j f_{i,j}} \qquad idf_t = \log\left(1 + \frac{D}{D_t}\right) \qquad tf_{d,t} \times idf_{d,t}$$

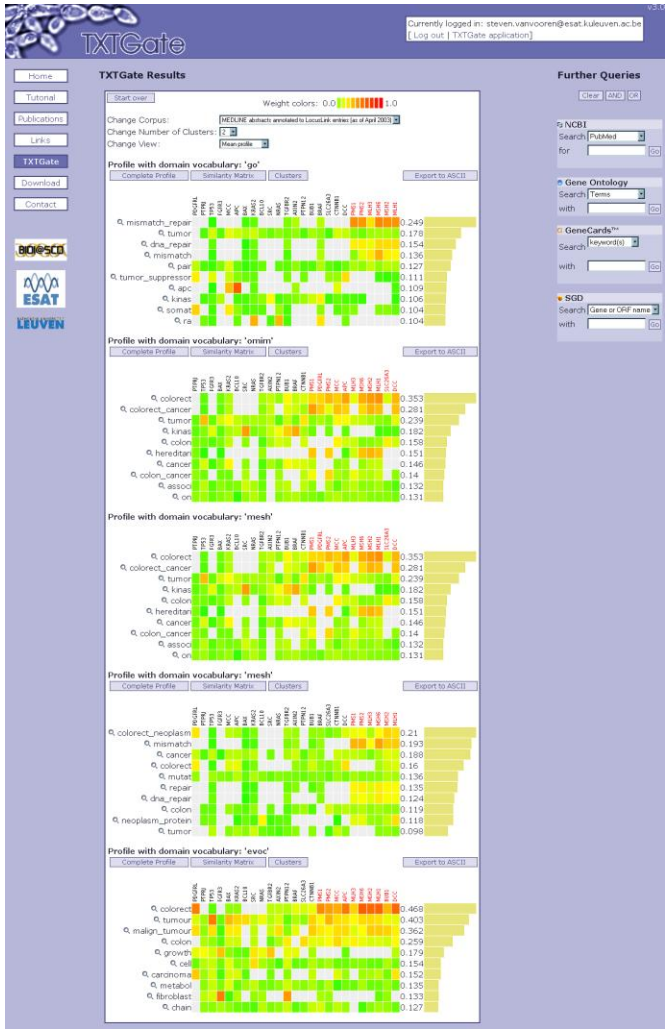
Similarity

Ranking, scoring and clustering of patients or documents



$$sim_{vs}(q, x_i) = Q \cdot X_i = \frac{\sum_{j=1}^{m} v_j \cdot w_{ij}}{\sqrt{\sum_{j=1}^{m} (v_j)^2 \cdot \sum_{j=1}^{m} (w_{ij})^2}}$$

Profiling and Clustering



Publications

Yu S, **Van Vooren S**, Tranchevent LC, De Moor B, Moreau Y **"Comparison of vocabularies, representations and ranking algorithms for gene prioritization by text mining"** *Bioinformatics* 24(16):i119-25, 2008.

Glenisson P, Coessens B, **Van Vooren S**, Mathys J, Moreau Y, De Moor B. **TXTGate: profiling gene groups with text-based information.** *Genome Biol.* 2004;5(6):R43. Epub 2004 May 28.

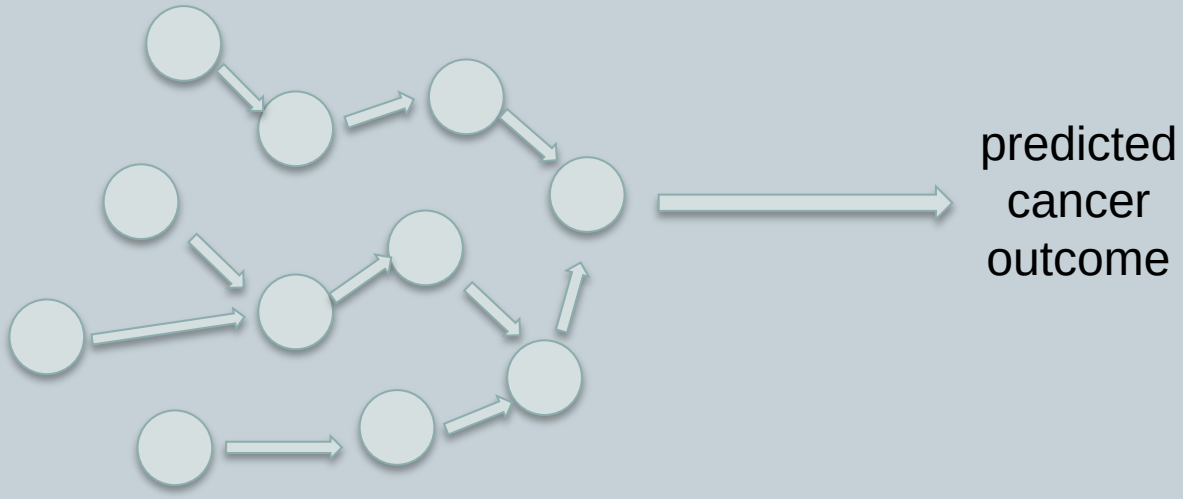
Glenisson P, Coessens B, **Van Vooren S**, Moreau Y, De Moor B **Text-based gene profiling with domain-specific views.** *in Proc. of the First International Workshop on Semantic Web and Databases (SWDB 2003), Berlin, Germany,* Sep. 2003, pp. 15-31.

Chapter 6 – data fusion on biomedical text and gene expression arrays

- Application: expression arrays
  - Co-expressed genes: cluster validation
    - TXTGate overview
    - Validation
- **Application: text data as Bayesian network prior**
  - Context
  - results

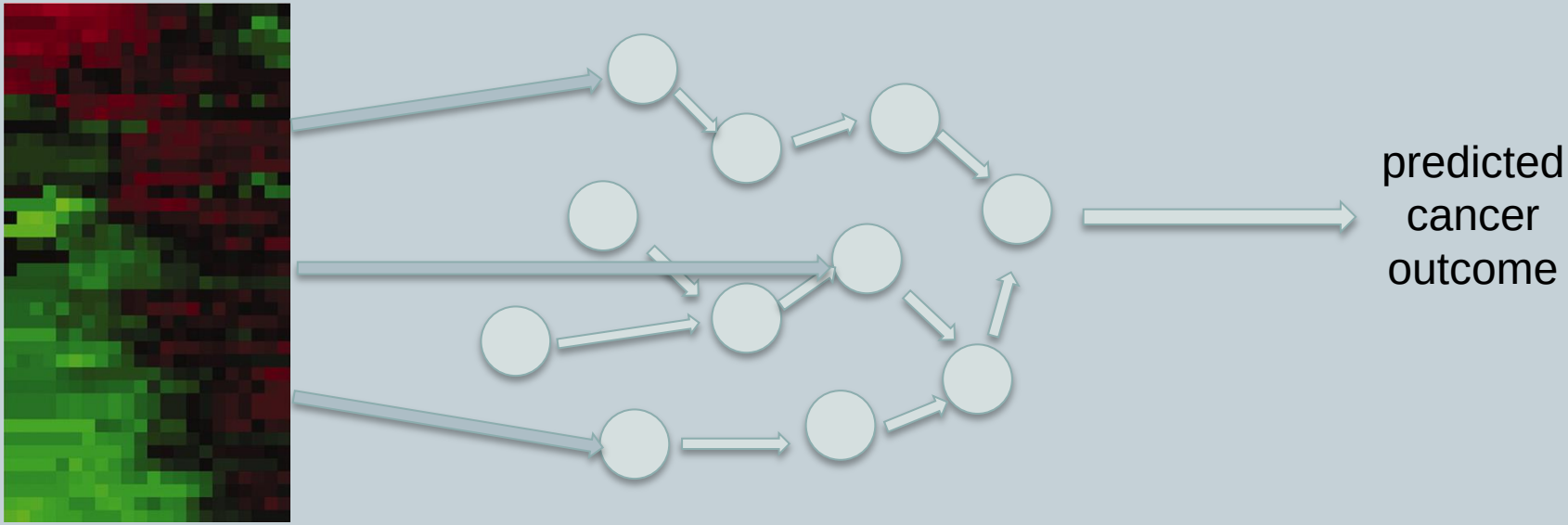
Structure improves with text prior

- Credit for Bayesian models: Olivier Gevaert
- Improve prognosis prediction in oncology
  - Build predictive network based on expression data



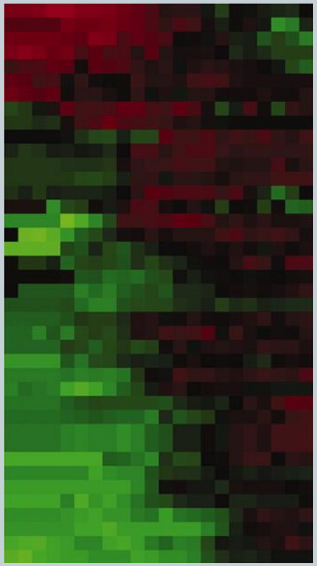
Structure improves with text prior

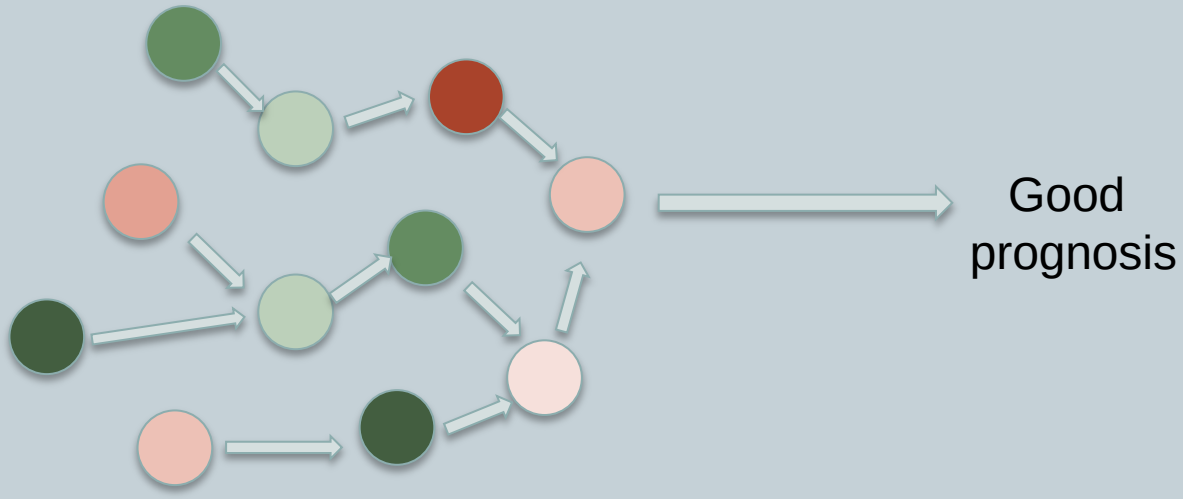
- Credit for Bayesian models: Olivier Gevaert
- Improve prognosis prediction in oncology
  - Build predictive network based on expression data



Structure improves with text prior

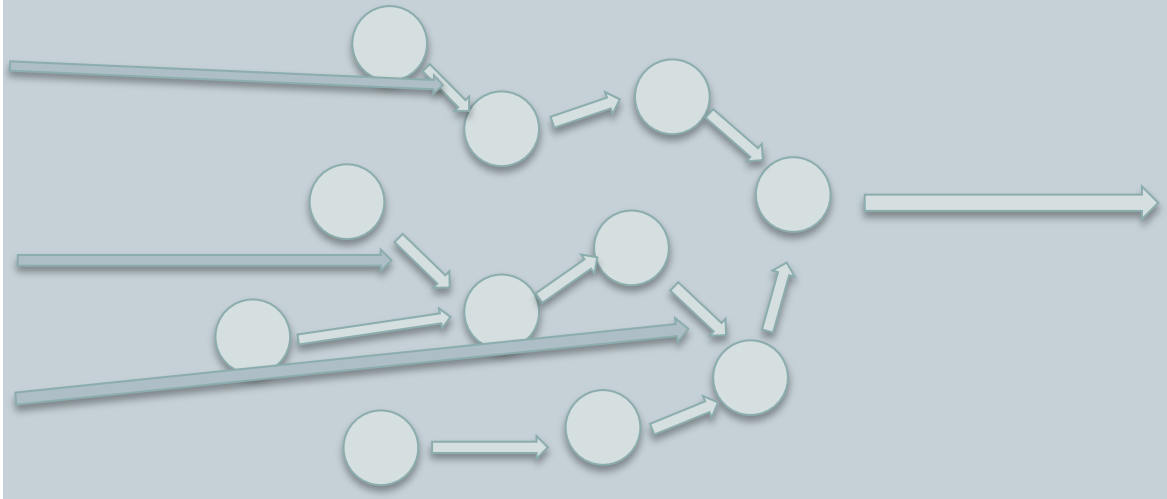
- Credit for Bayesian models: Olivier Gevaert
- Improve prognosis prediction in oncology
  - Build predictive network based on expression data





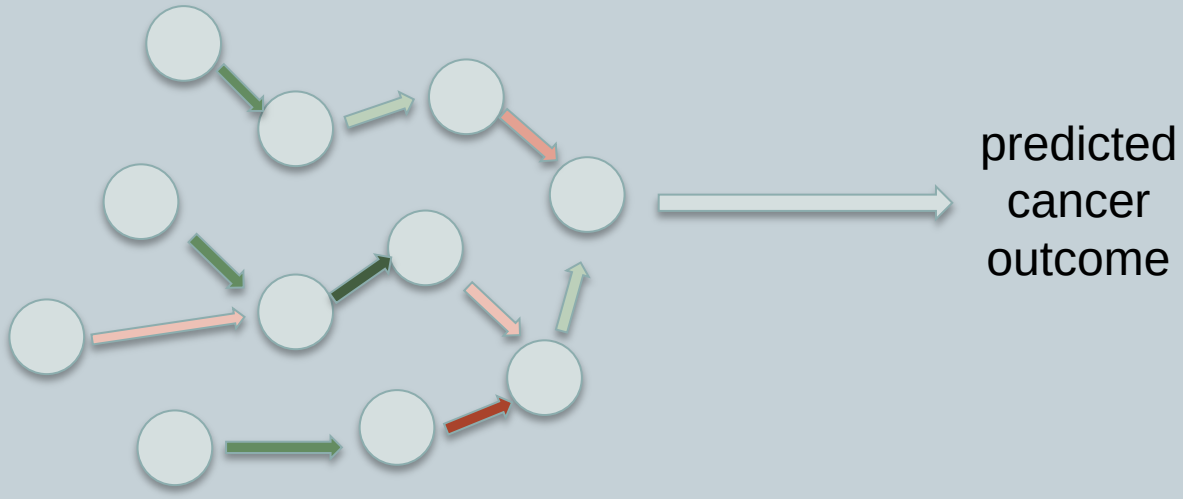
Structure improves with text prior

- Credit for Bayesian models: Olivier Gevaert
- Improve prognosis prediction in oncology
  - Build predictive network based on expression data



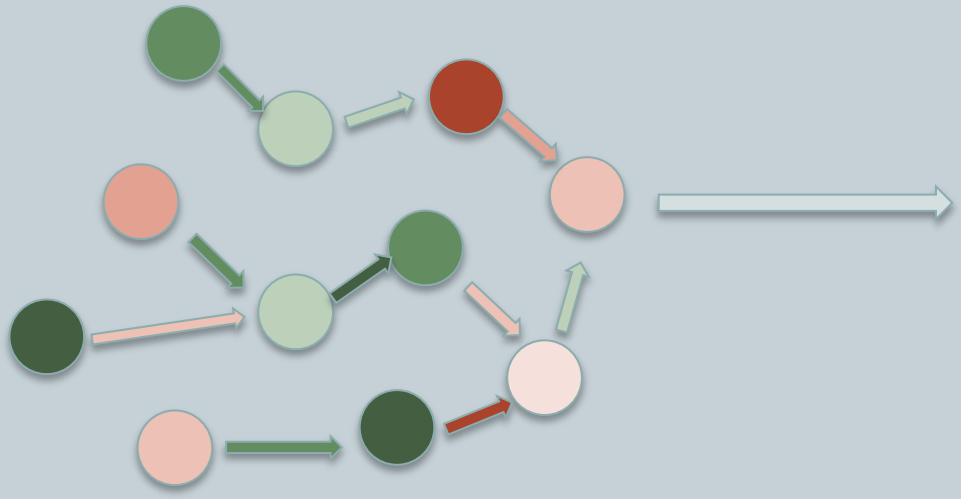
Structure improves with text prior

- Credit for Bayesian models: Olivier Gevaert
- Improve prognosis prediction in oncology
  - Build predictive network based on expression data



Structure improves with text prior

- Credit for Bayesian models: Olivier Gevaert
- Improve prognosis prediction in oncology
  - Build predictive network based on expression data



Structure improves with text prior

- Credit for Bayesian models: Olivier Gevaert
- Improve prognosis prediction in oncology
  - Build predictive network based on expression data
  - Use text vectors for gene distances as structure prior
- Bayesian networks: results
  - Model building benefits from text data as structure prior
  - Prediction with new data: validation on 3 sets
    - Breast, lung and ovarian cancer

Publications

Gevaert O, **Van Vooren S**, De Moor B., **"A framework for elucidating regulatory networks based on prior information and expression data"**, *Ann N Y Acad Sci.*, 1115:240-8, 2007.

Gevaert O, **Van Vooren S**, de Moor B., **"Integration of microarray and textual data improves the prognosis prediction of breast, lung and ovarian cancer patients [81]"** *Pac Symp Biocomput.*, 2008:279-90, 2008.

Chapter 7 – Conclusions and perspectives

ACCOMPLISHMENTS AND FUTURE PROSPECTS

Chapter 7 – Conclusions and perspectives

- Accomplishments
  - Increase Array CGH data interpretability by putting it in context
    - Data representation
    - Algorithms and statistics
    - Information (literature, other patients, public data sources, …)
  - Built a framework for data integration and mining
    - Biomedical text
    - Microarray assay data
    - Structured phenotype annotations
  - Built and extended platforms for
    - patient data aggregation and management
    - tools for clinical and functional interpretation
- Challenges
  - Future proof platforms - the data avalanche: next generation sequencing
  - Global initiatives to chart genome variation (CHERISH, national consortia, …)
  - Prepare for a clinical (r)evolution
    - Data collection and distribution will spread (from referrer to patient, eHealth)
    - Merger of different disciplines (SNP genotyping, expression, copy number variation, epigenetic assays, single gene polymorphisms, population studies… and environment)
    - Applications in oncology
    - Personalized medicine (early screening, targeted treatment)

Acknowledgements

- Prof. dr. ir. Yves Moreau and Prof. dr. ir. Bart De Moor
- Prof. dr. ir. Joris Vermeesch, Prof. dr. Koen Devriendt, dr. Bernard Thienpont, Dr. Thomy de Ravel
- Patrick Glenisson, Bert Coessens, Frizo Janssens, Joke Allemeersch, Stein Aerts, Olivier Gevaert
- Dr. Nigel Carter, Roger Pettett, Shola Richards, Helen Firth, Paul Bevan, Heike Fiegler
- Prof. dr. Frank Speleman, dr. Bjorn Menten
- Dr. Dietrich Rebholz-Schuhmann, Miguel Arregui, Sylvain Gaudan
- prof. dr. ir. M. Bruynooghe, prof. dr. ir. S. Vanhuffel
- prof. dr. ir. E. Aernoudt
- CME, BIOI, ESAT, LRD, EBI, WTSI, KUL, VIB
- …