Centro de Investigación en Métodos de Producción de Software

Understanding the Human Genome: a Conceptual Modelling-based Approach

Prof. Oscar Pastor
ProS Research Center - Universidad Politécnica de Valencia, Spain

DEXA 2010, Bilbao

UNIVERSIDAD POLITECNICA DE VALENCIA

Centro de Investigación en Métodos de Producción de Software

Agenda

1. Why a Keynote on CM and the Human Genome?
2. Problem Statement
3. The Role of Conceptual Modeling
4. The Present
5. The Short-Term Future
6. Understanding the Domain (Problem Space)
7. Building the ER Model / Data Base (Solution Space)
8. Conclusions

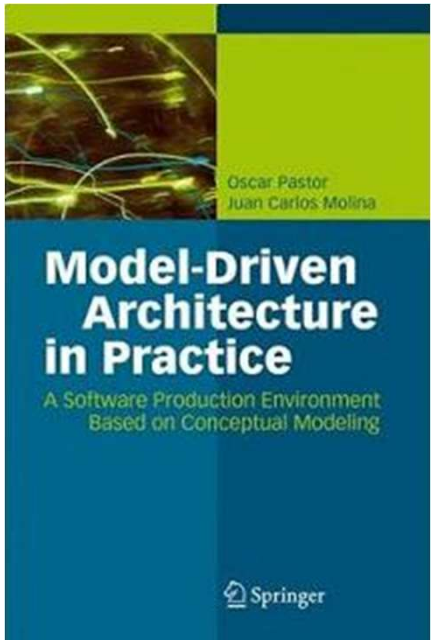
UNIVERSIDAD POLITECNICA DE VALENCIA

Centro de Investigación en Métodos de Producción de Software

Experience in Conceptual Modeling

- We have been building
  - Traditional Information Systems
  - Web-based Information Systems
  - SOA-based systems
  - Pervasive Systems

  - … but, what is next?



UNIVERSIDAD POLITECNICA DE VALENCIA

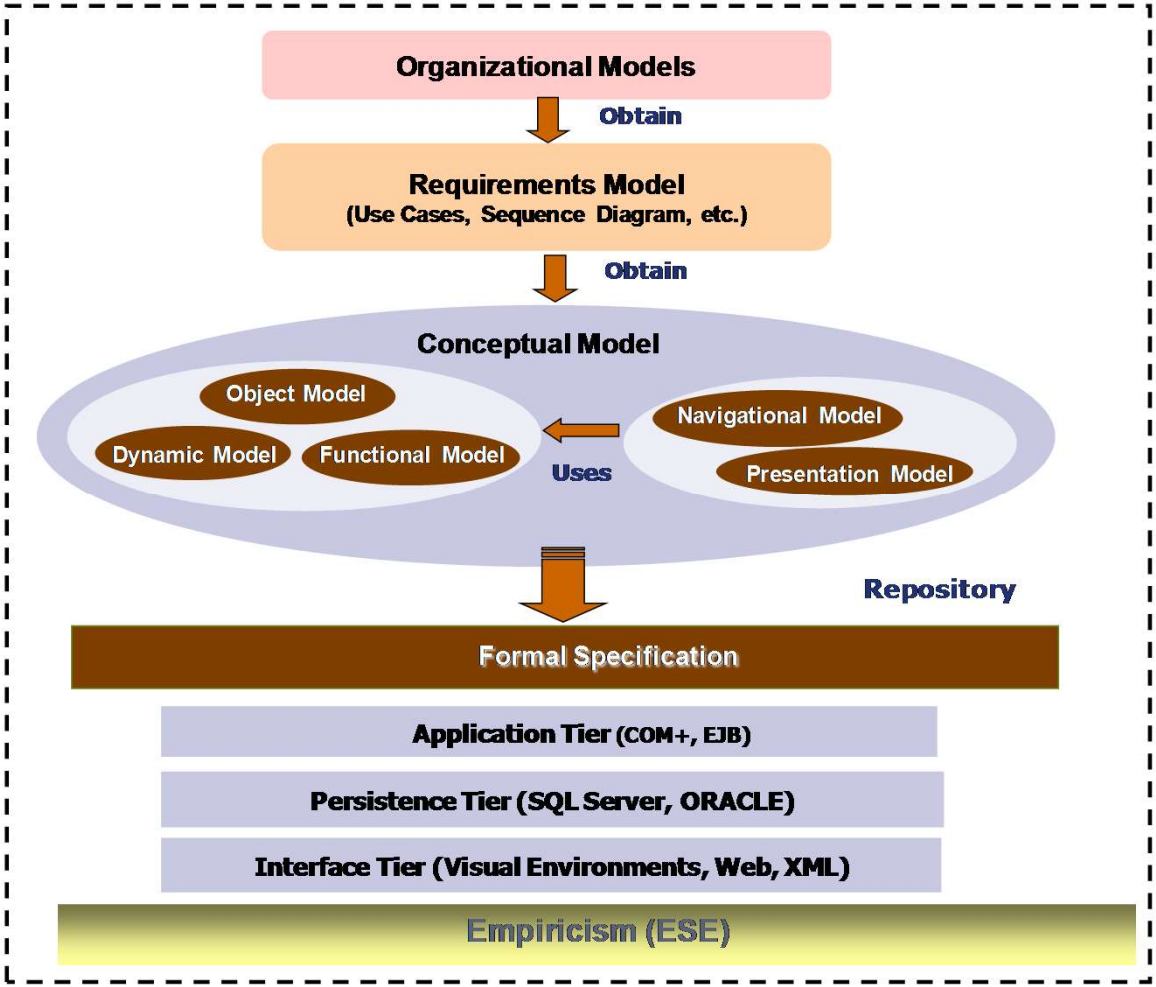
Centro de Investigación en Métodos de Producción de Software

The OO-Method Approach

Problem Space Level

Automated Translation

Solution Space Level

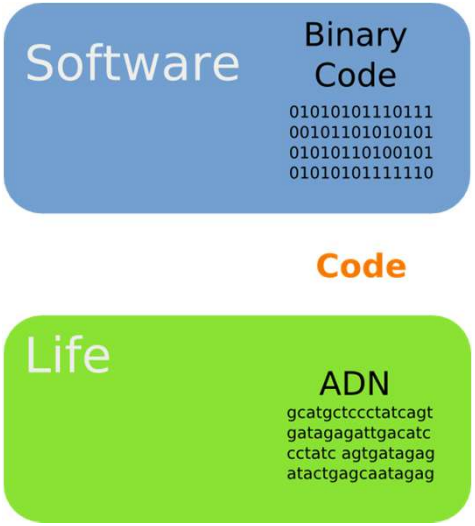


UNIVERSIDAD POLITECNICA DE VALENCIA

Centro de Investigación en Métodos de Producción de Software

A parallelism

- "A living organism is a *computer* or *machine* made up of genetic *circuits* in which DNA is the *software* that can be *hacked*." *— Drew Endy, MIT*



UNIVERSIDAD POLITECNICA DE VALENCIA

Centro de Investigación en Métodos de Producción de Software

Building life

- Synthetic Biology can create new forms of life from scratch
  - A microbe that would help in fuel production
  - Biological films as a basis of new forms of lithography for assembling circuits
  - Cell division counters to prevent cancer
  - Re-designed seeds that the tree is programmed to grow into a house

…but, how is this *"software"* developed?

UNIVERSIDAD POLITECNICA DE VALENCIA

Centro de Investigación en Métodos de Producción de Software

Building life

- First synthetic cell created (announced just last month)
- A tricky artificial cell
- Enormously useful as a proof of concept: alive cells can be generated from genetic sequences, that could create beings with different genomes…

…provided that the genome is fully understood!!!…

UNIVERSIDAD POLITECNICA DE VALENCIA

By the way....

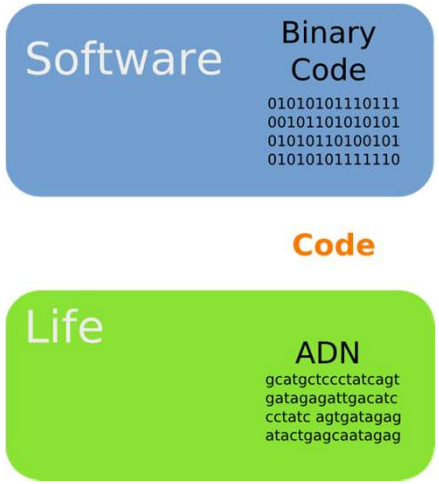
- Four enigmas with answer:

1. Crossing the “Rubicon” (point of no return): alive cells can be created from entirely artificial genomes
2. Bioterrorism threats
3. Does it mean creating life? Not from scratch: it is a copy of a preexistent cell
4. Will we create life? Not reason to answer no.

Centro de Investigación en Métodos de Producción de Software

Building life

- "Using a laptop computer, published gene sequence information and mail-order synthetic DNA, just about anyone has the potential to construct genes or entire genomes from scratch." *— Drew Endy, MIT*



UNIVERSIDAD POLITECNICA DE VALENCIA

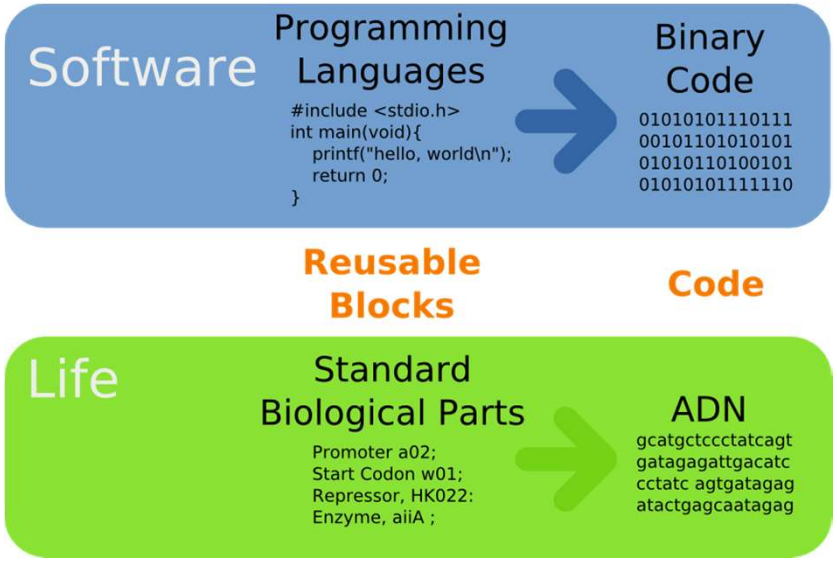
Abstraction as a solution

- Model Driven Development permits
  - Reason about the system prior to its construction
    - You can simulate the behavior to foresee the consequences of a system
  - Derivate the final system in an automatic way
    - Obtaining a consistent result

Centro de Investigación en Métodos de Producción de Software

First step: Assembling

- First abstraction step
  - Standard Biological Parts

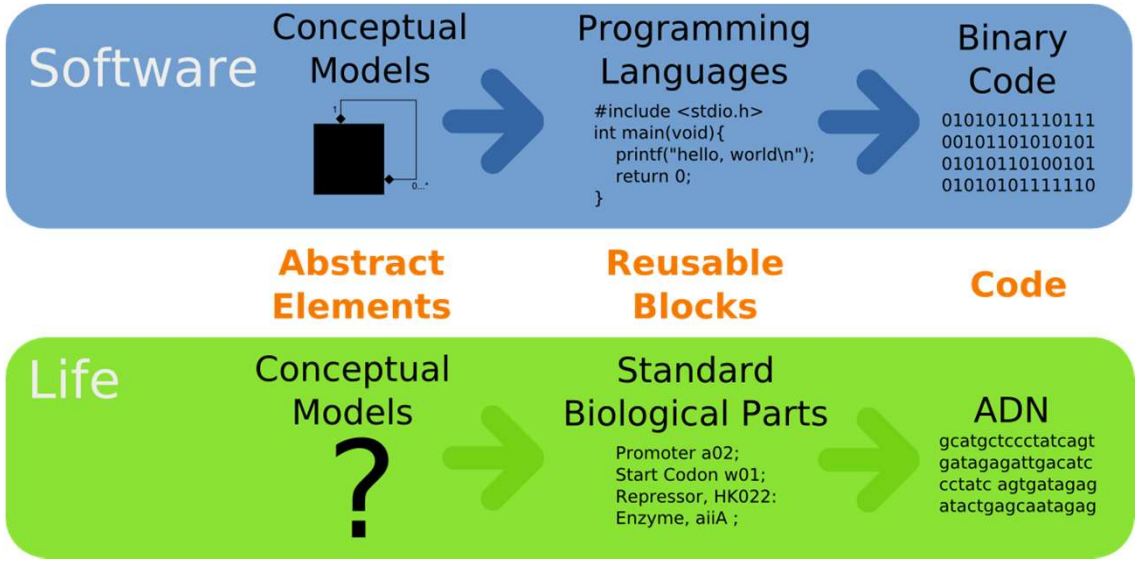


UNIVERSIDAD POLITECNICA DE VALENCIA

Centro de Investigación en Métodos de Producción de Software

One step further: Modeling

- Conceptual models are needed for a systematic development of biological systems



UNIVERSIDAD POLITECNICA DE VALENCIA

Centro de Investigación en Métodos de Producción de Software

From Genome To Reality

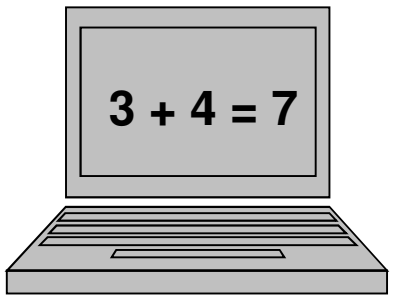
00010011	00000111	00000011	00001000

Physical Level

ADD $7 $3 $8

Instruction Level

Semantics: Add the values from the processor registers '3' and store the result in the register '8'



Representation Level

UNIVERSIDAD POLITECNICA DE VALENCIA

AUG	GAA	CAC	GAC	GAG	UAA

Physical Level

START Glu His Asp Glu STOP

Instruction Level

Semantics: Process a protein with the four selected aminoacids

However, ¿Why?

Representation Level

Centro de Investigación en Métodos de Producción de Software

One step further: Modeling

- Modeling benefits are needed for biological systems
  - Work at a higher abstraction level
    - Systems easy to specify
  - Reason about the system prior to construction
    - Foresee consequences in advance
    - Simulate, validate, etc.
  - Automate the development
    - In a systematic way

Centro de Investigación en Métodos de Producción de Software

(Initial) Conclusion

- With Conceptual Models targeted at digital elements, we can improve Information Systems Development
- With Conceptual Models targeted at life we can directly improve our living

Centro de Investigación en Métodos de Producción de Software

Translational Research

- Movement of discoveries in basic research (the Bench) to application at the clinical level (the Bedside)
- A significant barrier: the lack of uniformly structured data across related biomedical domains
- A potential solution: Semantic Web Technologies

UNIVERSIDAD POLITECNICA DE VALENCIA

- Information ecosystem
  - Scientific literature
  - Experimental data
  - Summaries of knowledge of gene products
  - Diseases
  - Compounds
  - Informal scientific discourse and commentary in a variety of forums
- This data has been provided in numerous disconnected DBs –data silos-

- The lack of uniformly structured data affects many areas of biomedical research
  - Drug discovery
  - Systems biology
  - Individualized medicine
- …all of which rely heavily on integrating and interpreting data sets produced by different experimental methods at different levels of granularity

Centro de Investigación en Métodos de Producción de Software

Example: Alzheimer's Disease (AD)

- Still no agreement on how it is caused, or where best to intervene to treat it or prevent it
- Recent hypothesis combines data from research in mouse genetics, cell biology, animal neuropsychology, protein biochemistry, neuropathology,… and other areas

Example: Huntington's Disease (HD)

- Relatively simple genetic basis, and a model for autosomal dominant neurogenetic disorders proposed …
- But the mechanisms by which the disorder causes pathology still not understood, what creates profound difficulties with existing treatments.

Centro de Investigación en Métodos de Producción de Software

How can the SW help biomedical research?

- Are Semantic Web Technologies the solution?
  - Thesauri, ontologies, rule systems, frame based representation systems,..
  - A query language (SPARQL)
  - RDF, OWL,…

UNIVERSIDAD POLITECNICA DE VALENCIA

Centro de Investigación en Métodos de Producción de Software

Some potential advantages

- Global scope of identifiers
- RDFS and OWL are
  - Self-descriptive languages
  - Flexible, extendable and decentralized
- Ability to do inference, classification and consistency checking
  - A review of GO gave up to 10% of obsolete terms for gene annotations

Centro de Investigación en Métodos de Producción de Software

Main objectives

- Identification of core vocabularies and ontologies to support effective access to knowledge and data
- Development of guidelines and best practices for unambiguously identifying resources such as docs and biological entities
- Development of strategies for linking to the information discussed in scientific pubs. from within those pubs.

UNIVERSIDAD POLITECNICA DE VALENCIA

Centro de Investigación en Métodos de Producción de Software

The caos of the genome data

- Currently there are **tons of data** from the genome publicly available
- Some of these databases are **free available** on the Web because owners doesn't know how to find relevant information
- Each database is defined with an specific schema, data format, identifications, etc.
- The **integration** of the different sources is a very difficult task

UNIVERSIDAD POLITECNICA DE VALENCIA

Centro de Investigación en Métodos de Producción de Software

Example: Looking for information about the NF1 Gene

- A genomic laboratory must perform an analysis to determine in the subject suffers from Neurofibromatosis
- Currently the genetic analyst must manually search in the different databases to elaborate the report
- As a first research exercise, we have been looking for information about the NF1 Gene that provokes the Neurofibromatosis disease
- Several databases have been consulted to understand how the data is stored and retrieved

UNIVERSIDAD POLITECNICA DE VALENCIA

HUGO

Core Data		Database Links		
Approved Symbol +	NF1	RefSeq IDs +		
Approved Name +	neurofibromin 1	NM_000267	GenBank	UCSC Browser
HGNC ID +	HGNC:7765	Rat Genome Database ID (mapped data supplied by RGD) +		
Status +	Approved	RGD:3168	RGD ID	
Chromosome +	17q11.2	Entrez Gene ID +		
Previous Symbols +		4763	Gene	Map Viewer
Previous Names +		CCDS IDs +		
Aliases +		CCDS11264.1	CCDS ID	
Name Aliases +	"neurofibromatosis", "von Recklinghausen disease", "Watson disease"	Pubmed IDs +		
Locus Type +	gene with protein product	1715669	PMID	
		VEGA IDs +		
Gene Symbol Links		OTTHUMG00000132871	VEGA GeneView	
GENATLAS GeneCards GeneClinics/GeneTests GoPubmed HCOP H-InvDB Treefam wikigenes		Ensembl ID (mapped data supplied by Ensembl) +		
		ENSG00000196712	Ensembl GeneView	
		OMIM ID (mapped data supplied by NCBI) +		
		162200	OMIM	
Specialist Database Links		UCSC ID (mapped data supplied by UCSC) +		
COSMIC Orphanet:16542		uc002hgg.1	UCSC Index	
		UniProt ID (mapped data supplied by UniProt) +		
Locus Specific Database Links		P21359	SwissProt	UniProt
NF1 International Mutation Databas, NF1 @ The Center for Medical Genetics				

Provides a common identification for a particular gene and the different alias used in another databases

Gene Ontology

NF1

Gene product information ⬇ Peptide sequence ⬇ Sequence information ⬇ 46 term associations ➡

Information

Symbol	NF1
Name(s)	Neurofibromin
Type	**protein**
Species	*Homo sapiens* (human)
Synonyms	NF1 IPI00299512 IPI00304235 IPI00220513 IPI00220514 NF1_HUMAN
Database	UniProtKB, UniProtKB:P21359
Sequence	View sequence; use as BLAST query sequence

Primary Peptide Sequence

Longest sequence shown.

RecName: Full=Neurof… …min truncated;
MAAHRPVEWVQAVVSRFDEQ…
ILKNVNNMRIFGEAAEKNLY…

Provides a controlled vocabulary to describe gene and gene product attributes in any organism. Useful to find relationships with a particular genomic term

Centro de Investigación en Métodos de Producción de Software

Entrez GENE

1: NF1 neurofibromin 1 [*Homo sapiens*]

GeneID: 4763 updated 03-Oct-2008

Summary

Official Symbol	NF1 provided by HGNC
Official Full Name	neurofibromin 1 provided by HGNC
Primary source	HGNC:7765
See related	Ensembl:ENSG00000196712; HPRD:01203; MIM:162200
Gene type	protein coding
RefSeq status	REVIEWED
Organism	*Homo sapiens*
Lineage	*Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo*
Also known as	WSS; NFNS; VRNF; FLJ21220; DKFZp686J1293
Summary	This gene product appears to function as a negative regulator of the ras signal transduction pathway. Mutations in this gene have been linked to neurofibromatosis type 1, juvenile myelomonocytic leukemia and Watson syndrome. The mRNA for this gene is subject to RNA editing (CGA>UGA->Arg1306Term) resulting in premature translation termination. Alternatively spliced transcript variants encoding different isoforms have also been described for this gene. [provided by RefSeq]

Genomic regions, transcripts, and products

Go to reference sequence details Try our new Sequence Viewer

NC_000017.9

Entrez Gene provides a unified query environment for *genes* provided by the NCBI. It can be considered ad the “facto” standard database to find information about a gene

UNIVERSIDAD POLITECNICA DE VALENCIA

Centro de Investigación en Métodos de Producción de Software

HGMD

Gene Symbol	Chromosomal location	Gene name	cDNA sequence	Extended cDNA	Splice junctions	Mutation viewer
NF1	17q11.2	Neurofibromatosis 1 protein (neurofibromin)	Get cDNA	BIOBASE Feature available to subscribers	Splice junctions	BIOBASE Feature available to subscribers

Mutation type	Number of mutations	Mutation data by type (register or log in)
Missense/nonsense	200	Get mutations
Splicing	149	Get mutations
Regulatory	0	No mutations
Small deletions	221	Get mutations
Small insertions	105	Get mutations
Small indels	12	Get mutations
Gross deletions	74	Get mutations
Gross insertions	8	Get mutations
Complex rearrangements	8	Get mutations
Repeat variations	0	No mutations
Public total (HGMD Professional 2008.2 total)	777 (1045)	

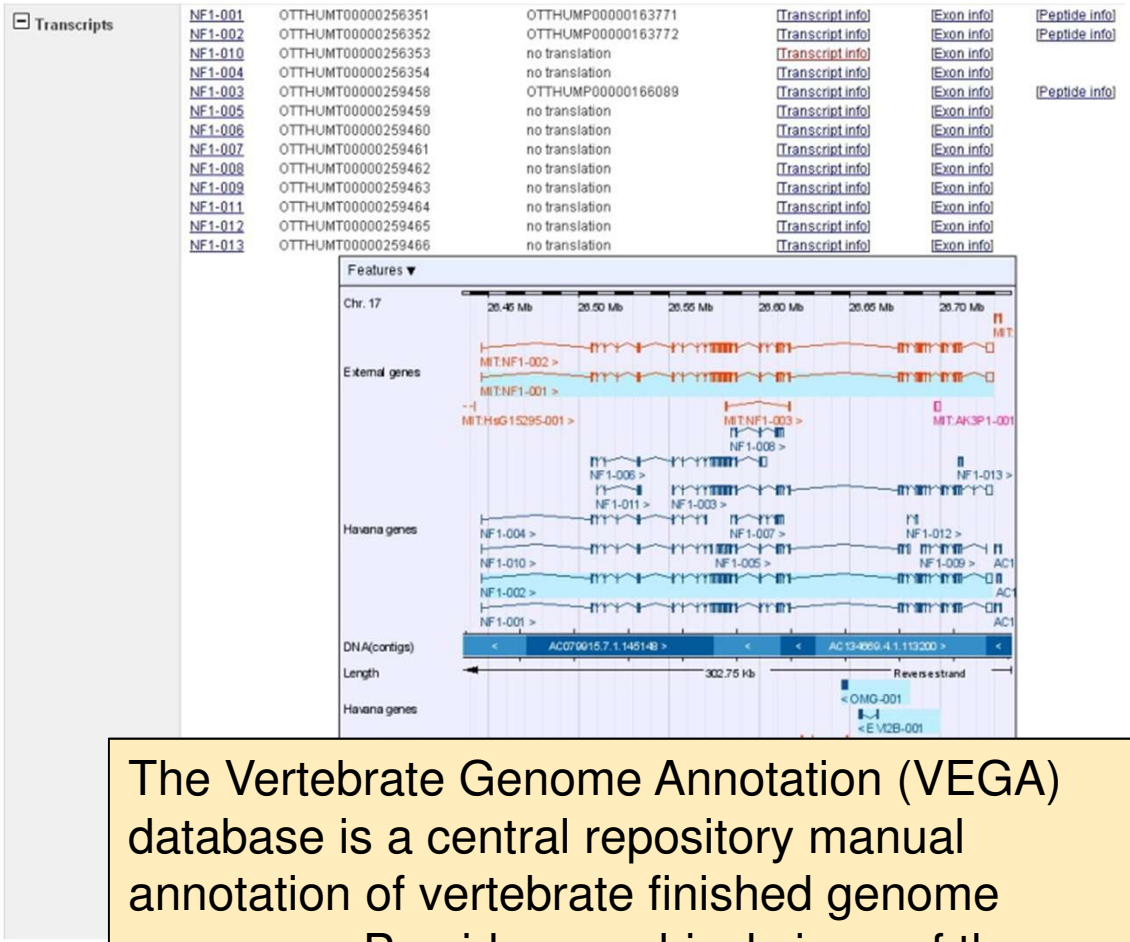
Disease/phenotype	Number of mutations	Mutation data by disease/phenotype
Neurofibromatosis 1		BIOBASE re available to subscribers
Neurofibromatosis-Noonan syndrome		BIOBASE re available to subscribers
Neurofibromatosis, spinal		BIOBASE re available to subscribers

The Human Gene Mutation Database comprises various types of mutation within the coding regions, splicing and regulatory regions of human nuclear genes causing inherited disease

UNIVERSIDAD POLITECNICA DE VALENCIA

Centro de Investigación en Métodos de Producción de Software

VEGA

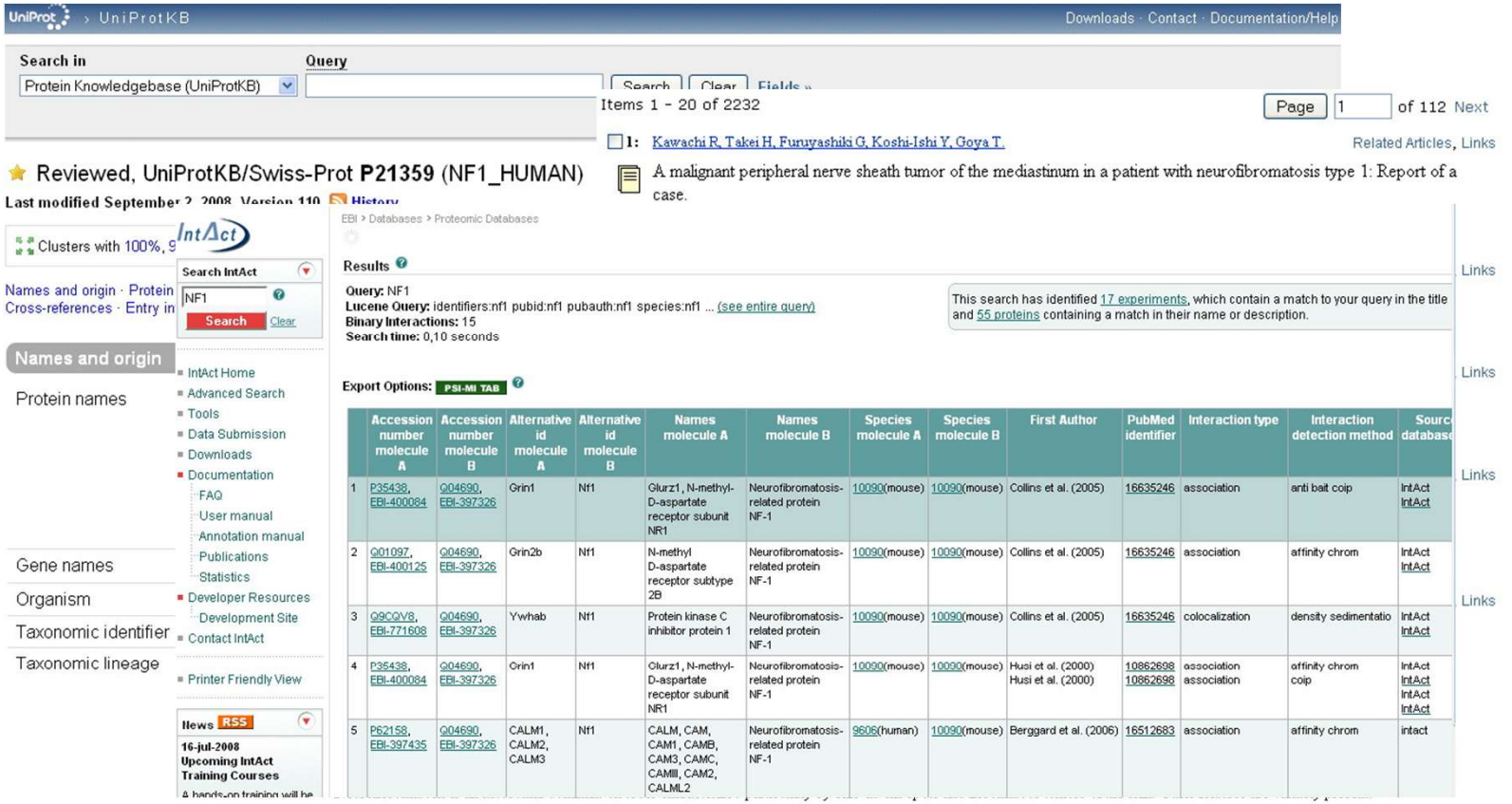


The Vertebrate Genome Annotation (VEGA) database is a central repository manual annotation of vertebrate finished genome sequence. Provides graphical views of the different gene transcripts

UNIVERSIDAD POLITECNICA DE VALENCIA

Centro de Investigación en Métodos de Producción de Software

And more...



Trovo-Marqui and Tajara (2006) provided a detailed review of neurofibromin and its role in neurofibromatosis.

Some patients with homozygous or compound heterozygous mutations in mismatch repair genes (see, e.g., MLH1; 120436 and MSH2; 609309) have a phenotype characterized by early onset malignancies and mild features of NF1, especially cafe-au-lait spots: see the mismatch repair cancer syndrome (276300), sometimes referred to as brain tumor-polyposis syndrome 1 or Turcot syndrome. These patients typically do not have germline mutations in the NF1 gene, although a study by Wang et al. (2003) suggested that biallelic mutations in mismatch repair genes may cause somatic mutations in the NF1 gene, perhaps resulting in isolated features resembling NF1.

CLINICAL FEATURES

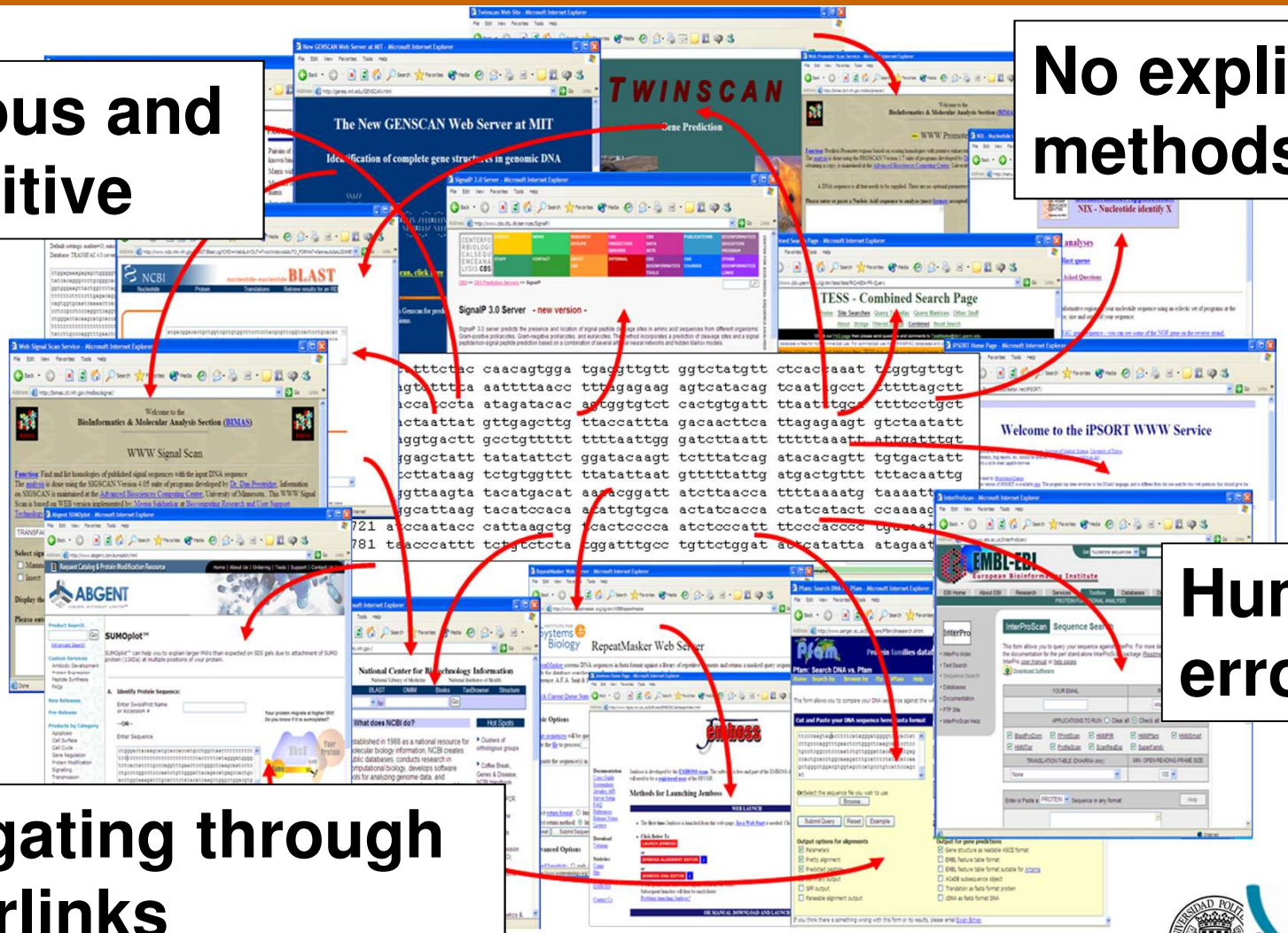
UNIVERSIDAD POLITECNICA DE VALENCIA

Centro de Investigación en Métodos de Producción de Software

Manual Methods of data analysis

Tedious and repetitive

No explicit methods



Human error

Navigating through hyperlinks

VALENCIA

BHIG

Centro de Investigación en Métodos de Producción de Software

Drawbacks observed

- Different identifications (ids) for the same disease gene
- The data is available on the Web but databases cannot always be directly queried
- The position (locus) of a particular gene depends on the genome sequenced
- Data is changing continuously
- High amount of information not well structured
- To provide a quality report about a gene disease several databases not interconnected must be manually consulted

UNIVERSIDAD POLITECNICA DE VALENCIA

Centro de Investigación en Métodos de Producción de Software

The short-term future

- The problem is getting worse !!!!!
- The DNA Sequencing hardware is evolving dramatically
- In next years, we will be able to sequence a complete human genome faster and cheaper

UNIVERSIDAD POLITECNICA DE VALENCIA

Centro de Investigación en Métodos de Producción de Software

The short-term future

- However, currently there is no software available to deal with the new challenges
- Software is required to:
  - Automatically find the mutations from a sequenced sample and store the new ones detected
  - Compare the genome of different subjects in order to determine all the differences between them
  - Trace the pathway from the genome code to the final phenotype of the individuals
- Conceptual modeling is required to produce quality software in this emerging domain

Centro de Investigación en Métodos de Producción de Software

Our Solution: Conceptual Modelling

- **Main goal:** provide Conceptual Models to represent the genome in order to enhance the Model-driven development of Biogenetic software
- The gene ontology is a useful resource to define a taxonomy but not to guide the software implementation
- The first step is to provide a common **E-R model** that will be able to support the genomic data complexity
- First approaches has been proposed by N.W. Paton et. Al[1], S.Ram [2], C.Tao and D.Embley [3]

[1] *N. W. Paton, S. A. Khan, A. Hayes, F. Moussouni, A. Brass, K. Eilbeck, C. A. Goble, S. J. Hubbard, and S. G. Oliver, "Conceptual modelling of Genomic Information," Bioinformatics, vol. 16, pp. 548-557, 2000.*

[2] Ram,S.: Toward Semantic Interoperability of Heterogeneous Biological Data Sources.CAiSE 2005 : 32-32

[3] *Tao,C.; Embley,D.: Seed-Based Generation of Personalized Bio-ontologies for Information Extraction. ER Workshops 2007: 74-84*

UNIVERSIDAD POLITECNICA DE VALENCIA

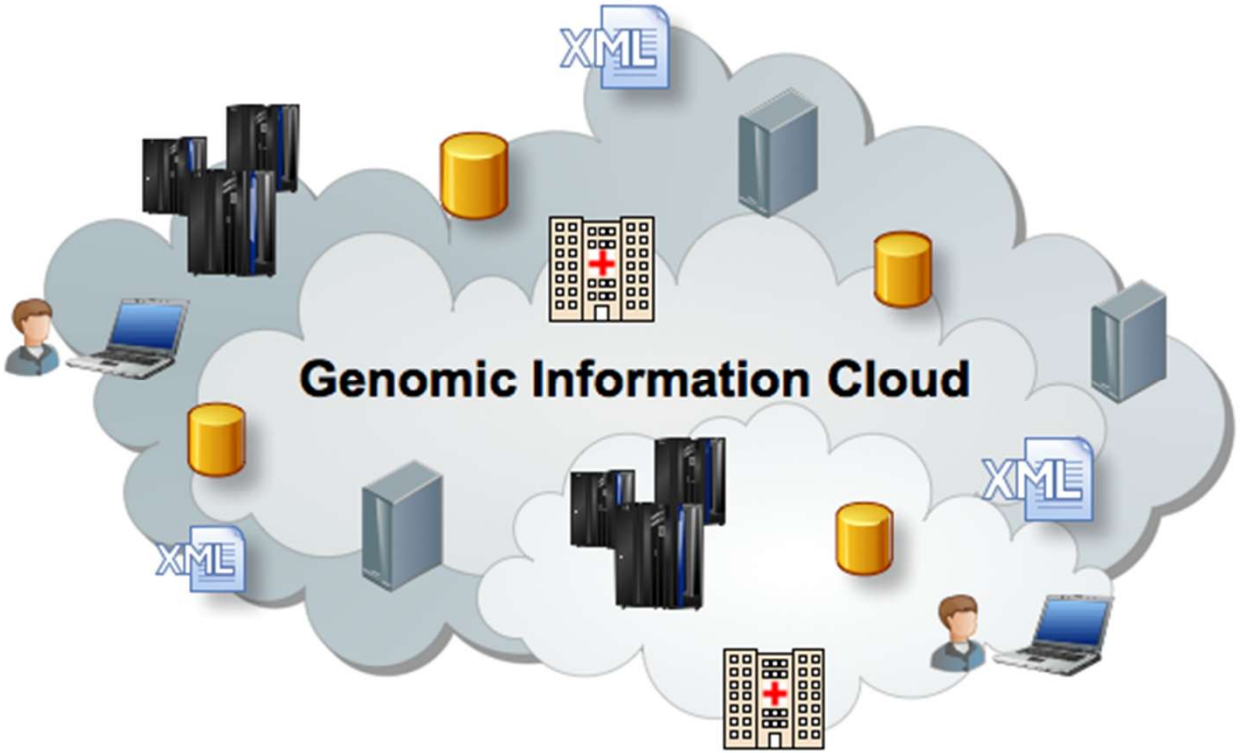
Centro de Investigación en Métodos de Producción de Software

The Genomic Data Chaos

UNIVERSIDAD POLITECNICA DE VALENCIA

Centro de Investigación en Métodos de Producción de Software

The Genomic Data Chaos

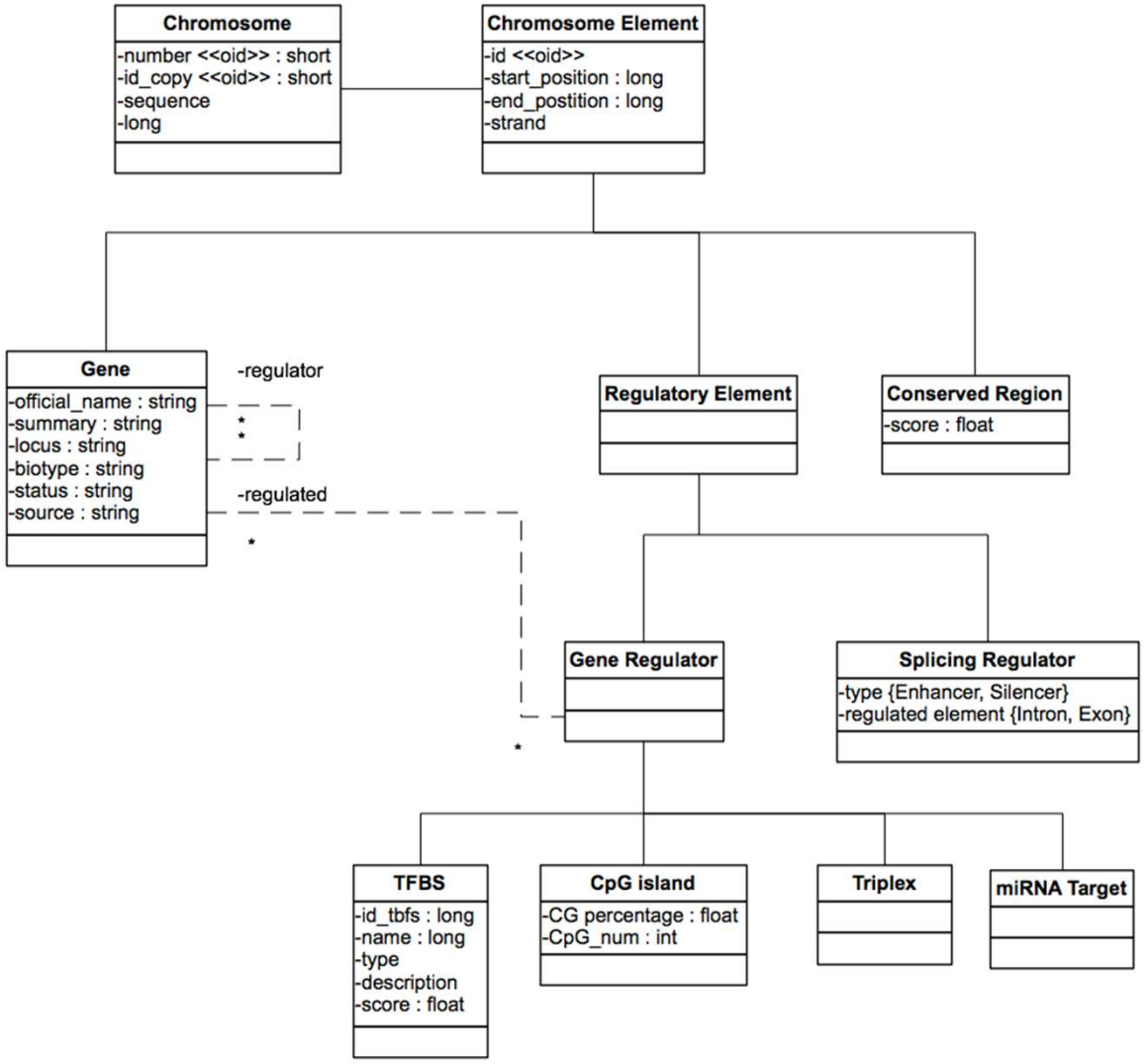


deling

UNIVERSIDAD POLITECNICA DE VALENCIA

Centro de Investigación en Métodos de Producción de Software

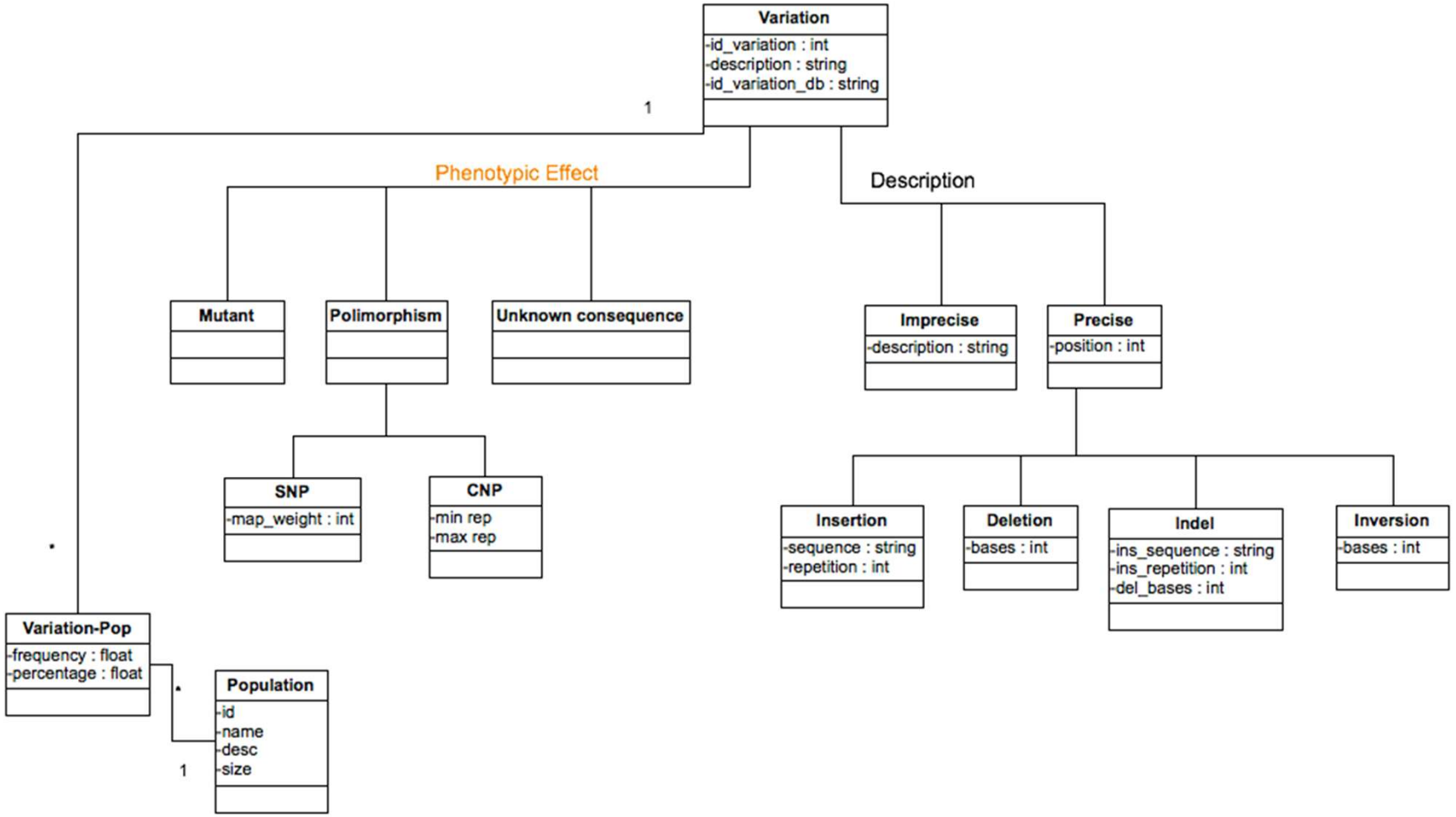
Conceptual Model: Gene View



UNIVERSIDAD POLITECNICA DE VALENCIA

Centro de Investigación en Métodos de Proc

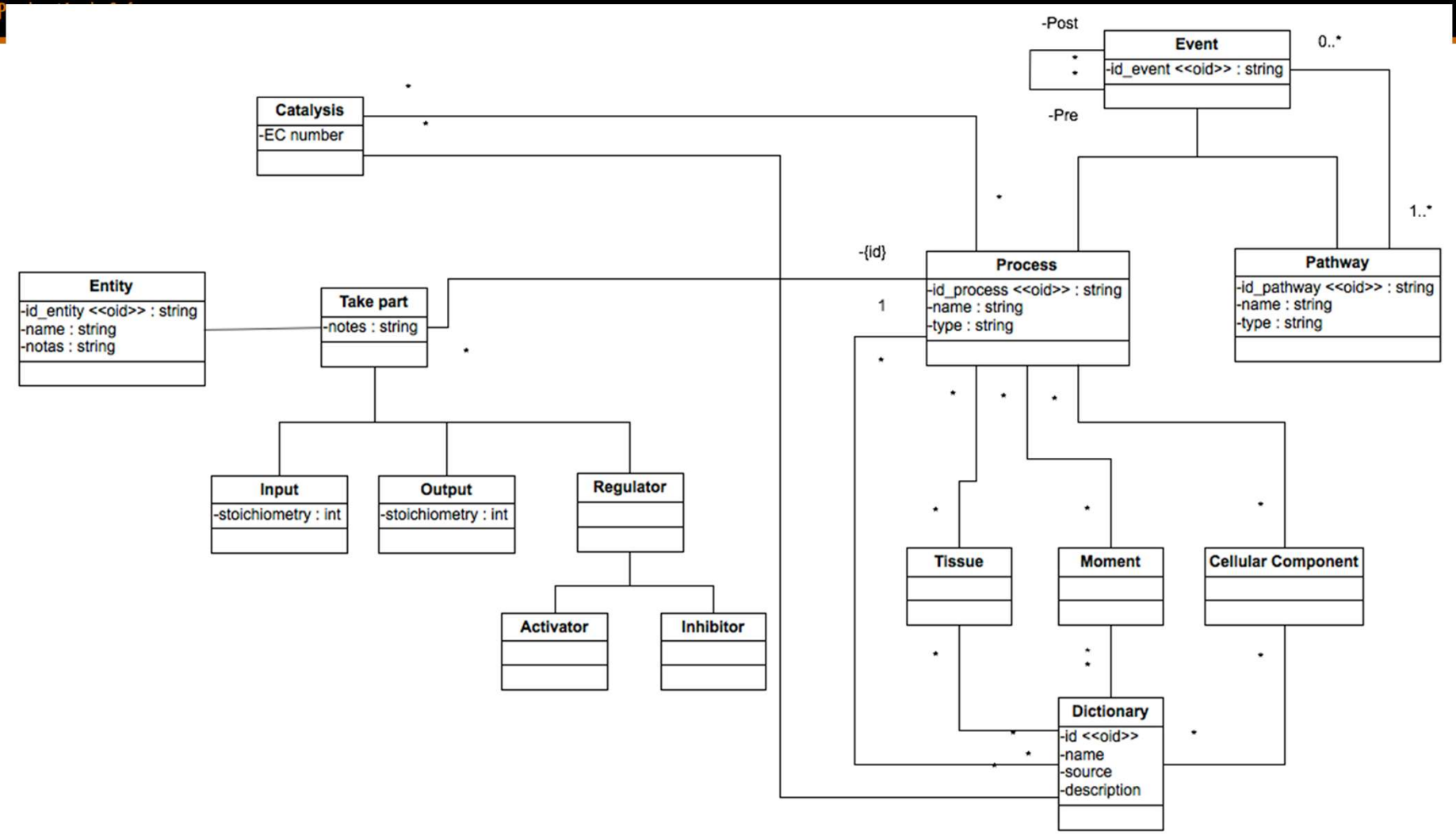
Conceptual Model: Variation View



UNIVERSIDAD POLITECNICA DE VALENCIA

Centro de Investigación en Métodos de

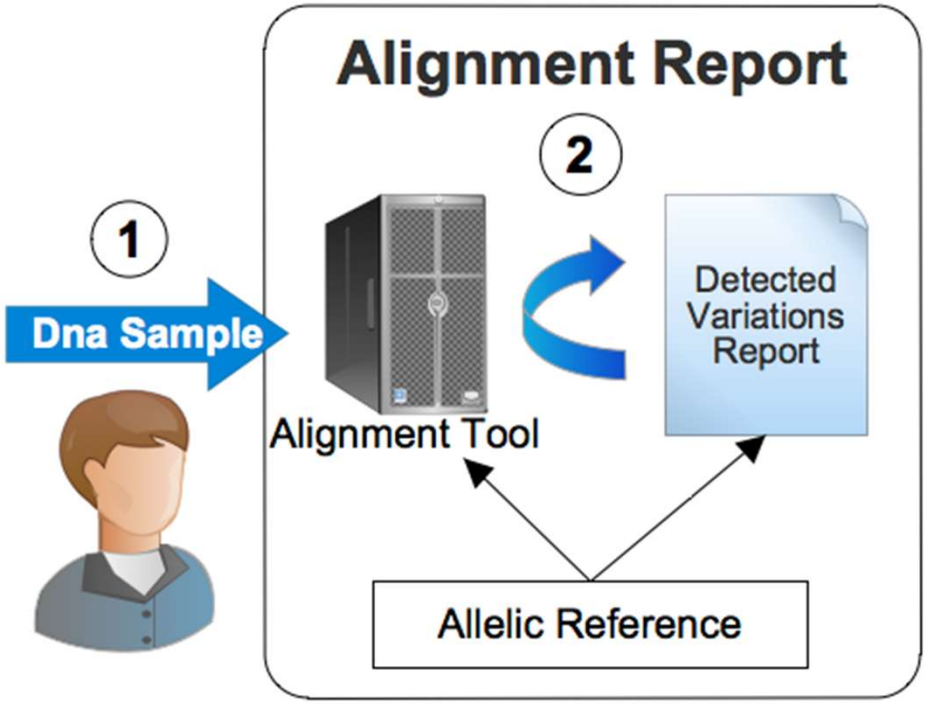
Conceptual Model: Pathway View



UNIVERSIDAD POLITECNICA DE VALENCIA

Centro de Investigación en Métodos de Producción de Software

Variation Analysis Process



The Input of the process is a DNA sample from a sequencing machine and an allelic reference sequence

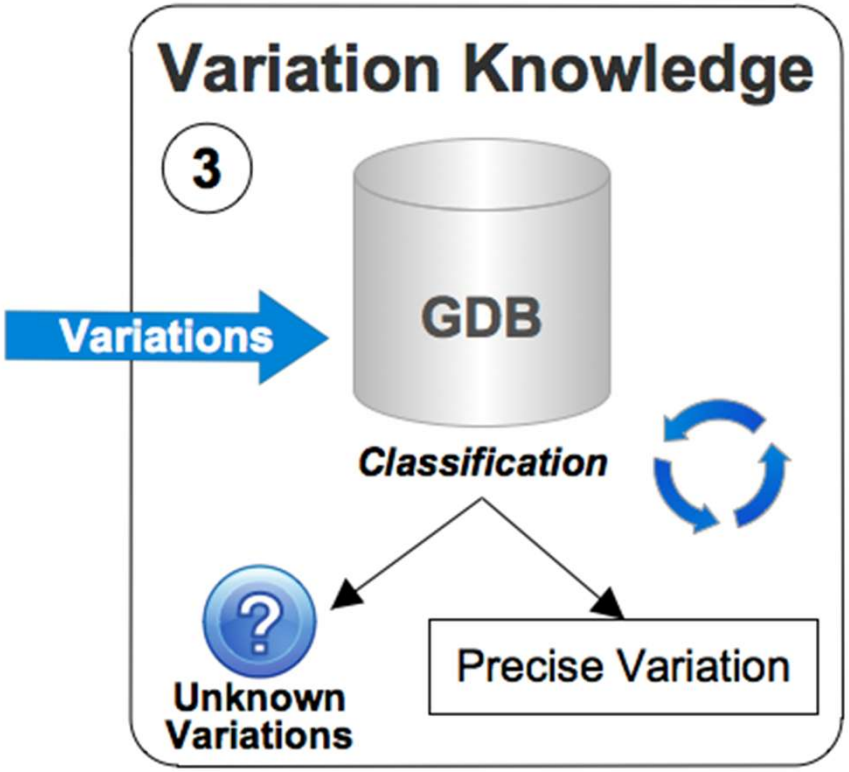
An alignment is performed using the BLAST tool

Each discovered difference is formalized as an instance of the variation entity. Then, a summarized report is generated.

UNIVERSIDAD POLITECNICA DE VALENCIA

Centro de Investigación en Métodos de Producción de Software

Variation Analysis Process



Founded Variations are searched in a database conforming to the genome conceptual model

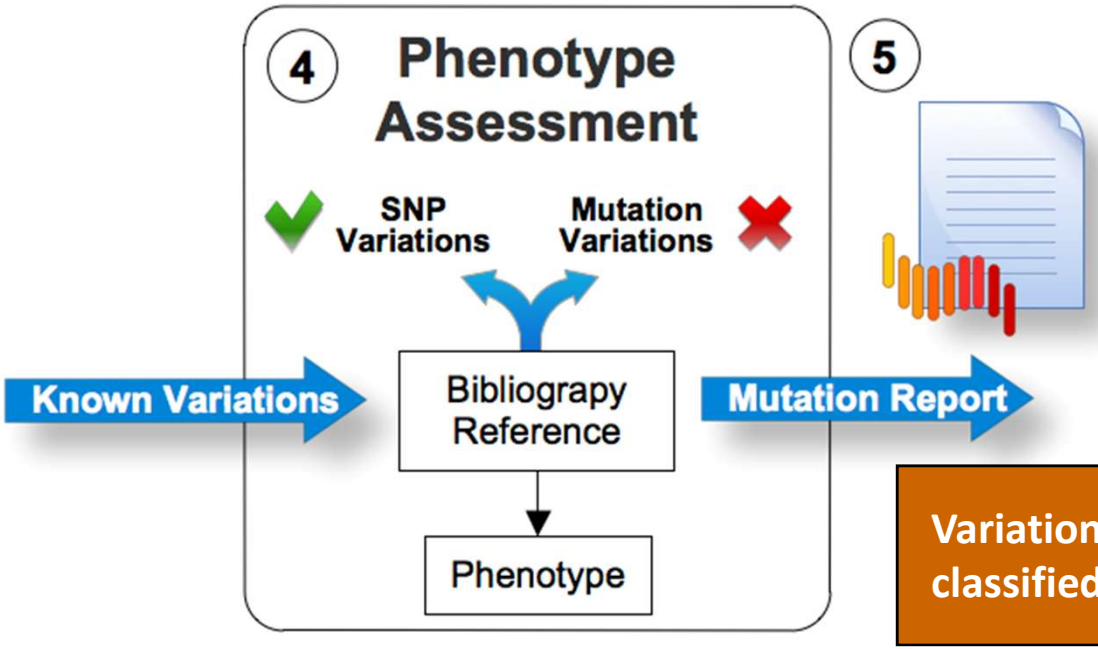
Known variations are classified into an specific type of sequence change (Insertion, Deletion, SNP, Indel).

Unknown variations are classified as non-silent if the variation produces an effect in the expected gene product

UNIVERSIDAD POLITECNICA DE VALENCIA

Centro de Investigación en Métodos de Producción de Software

Variation Analysis Process



In order to assess the phenotype of an specific variation, a research publication is required.

The conceptual model describes the bibliographical reference that supports the phenotype of a variation

Variations with a pathogenic phenotype are classified as mutations

Finally, the information is gathered in a report to support the clinical diagnosis

UNIVERSIDAD POLITECNICA DE VALENCIA

Genotype

The entire genetic identity of an individual that **does not show** any outward characteristics, *e.g.* Genes, mutations

Source: Paul Fisher -UMIST

Phenotype

(harder to characterise)

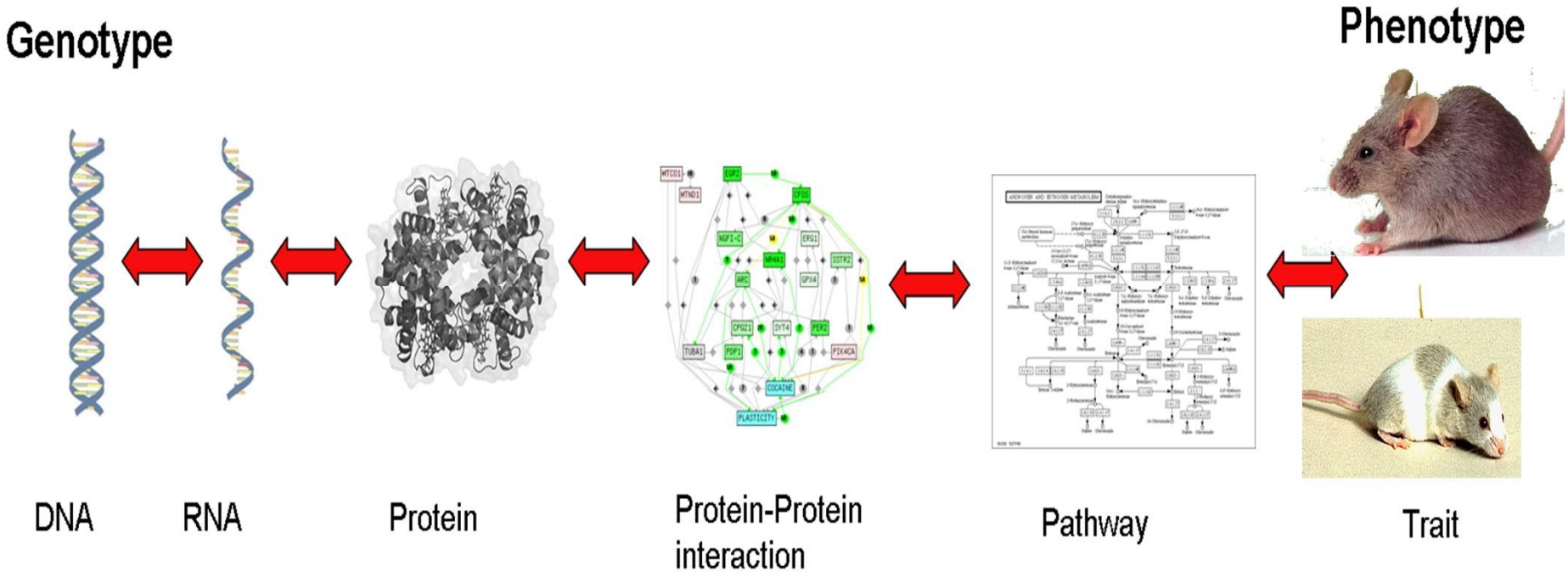
The observable expression of gene's producing **notable characteristics** in an individual, *e.g.* Hair or eye colour, body mass, resistance to disease

Brown

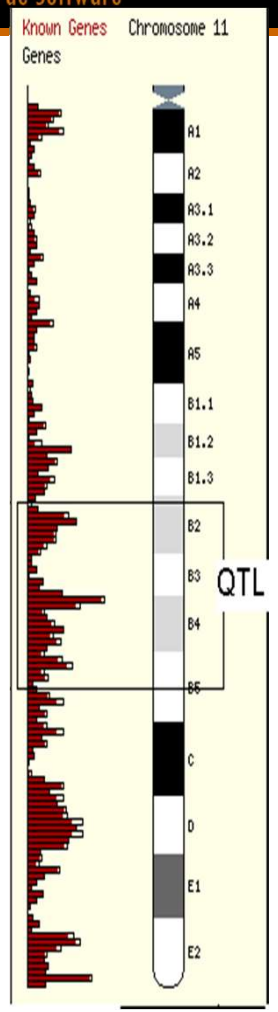
vs.

White and Brown

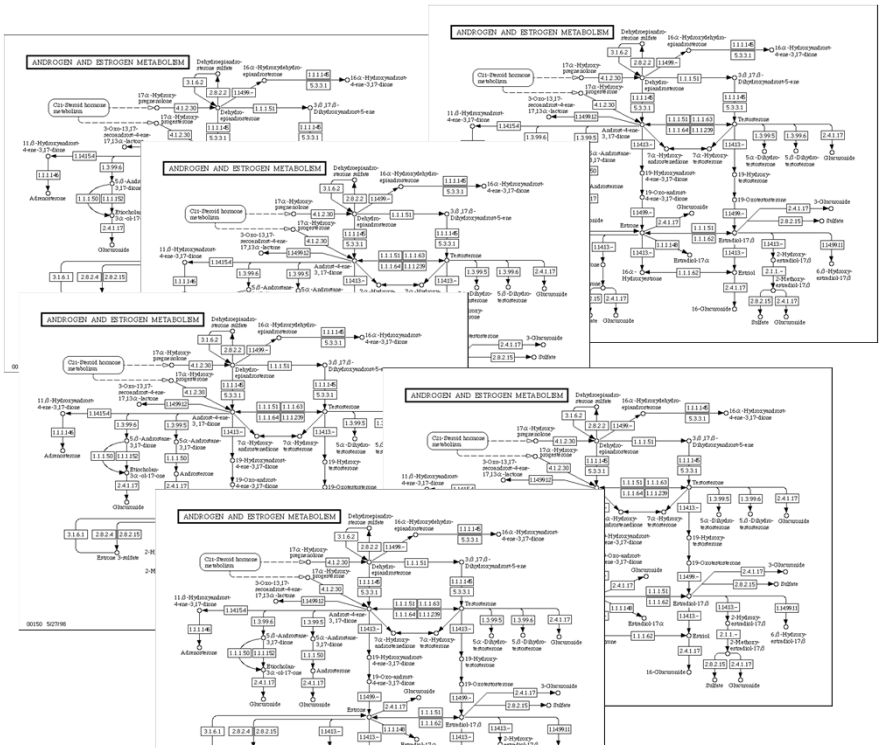
Source: Paul Fisher -UMIST



Source: Paul Fisher -UMIST



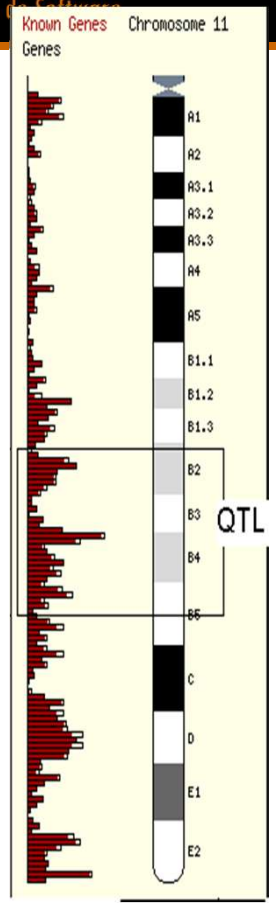
200



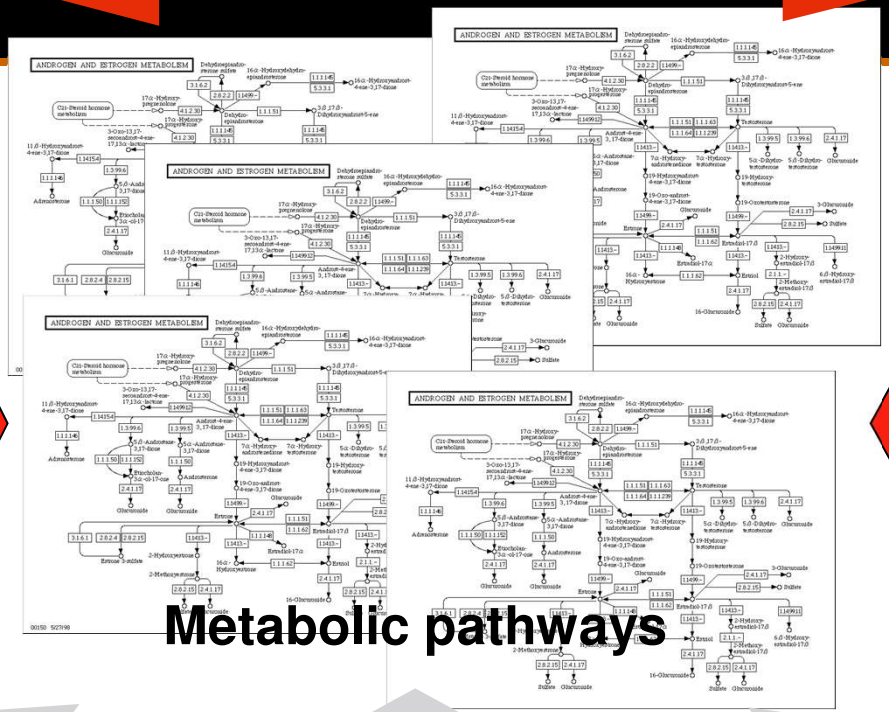
?

What processes to investigate?

Source: Paul Fisher -UMIST

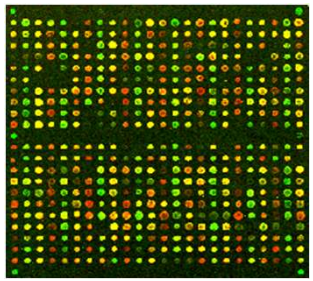


200



Metabolic pathways

?



Microarray + QTL

Genes captured in microarray experiment and present in QTL (**Quantitative Trait Loci**) region

Phenotypic response investigated using microarray in form of expressed genes or evidence provided through QTL mapping

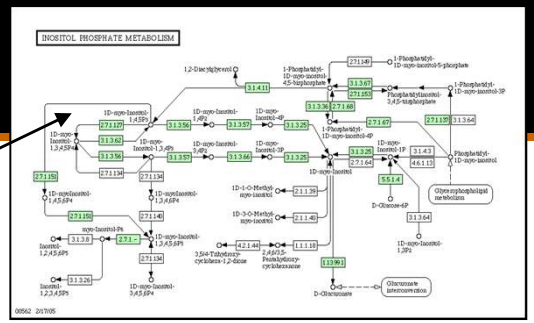
Centro de Investigación en Métodos de Producción de Soft

QTL

Gene A

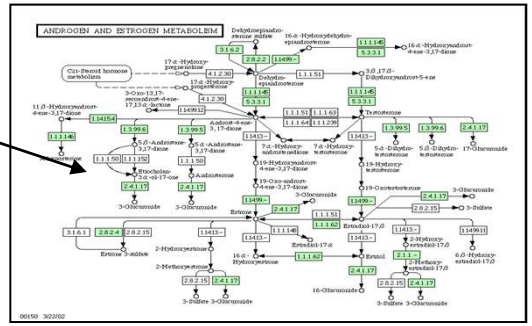
Gene B

Gene C



Pathway B

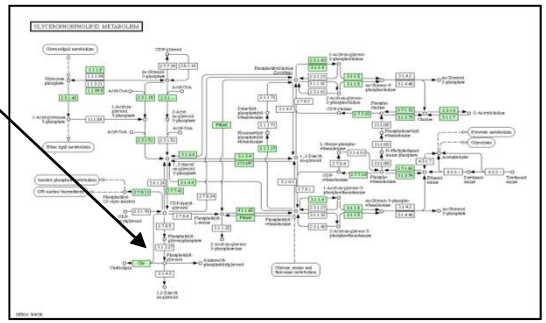
Pathway linked to phenotype – high priority



literature

Pathway not linked to phenotype – medium priority

Pathway C



literature

Pathway not linked to QTL – low priority

Genotype

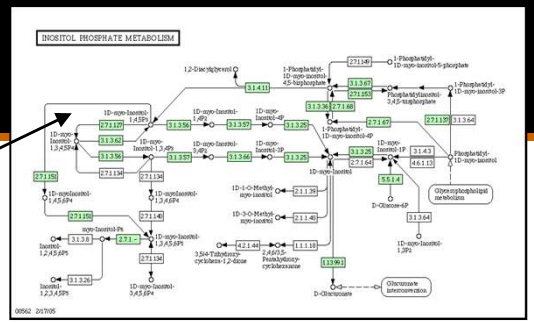
Centro de Investigación en Métodos de Producción de Soft

QTL

Gene A

Gene B

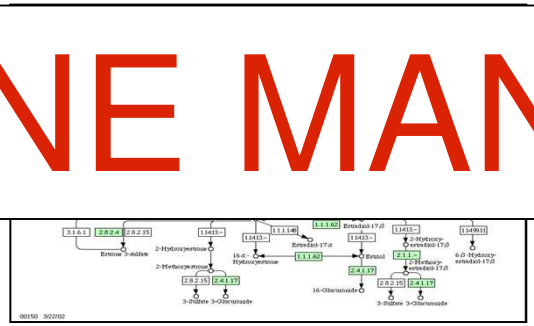
Gene C



Pathway B

Pathway linked to phenotype – high priority

DONE MANUALLY

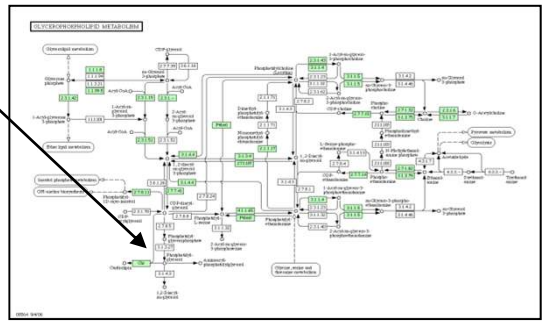


ked

medium priority

Pathway C

Genotype



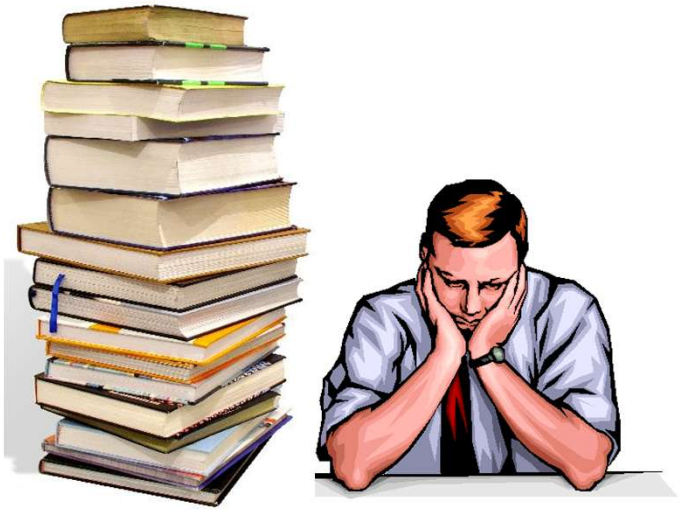
literature

Pathway not linked to QTL – low priority

Centro de Investigación en Métodos de Producción de Software

It can't be that hard, right?

- PubMed contains ~17,787,763 journals to date
- Manually searching is tedious and frustrating
- Can be hard finding the links



Source: Paul Fisher -UMIST

Computers can help with data gathering and information extraction – that's their job !!!

Centro de Investigación en Métodos de Producción de Software

Understanding the Domain (the Problem Space)

- Life as we know it is specified by the Genomes of the myriad organisms with which we share the planet.
- The nuclear genome comprises 3,2 G nucleotides of DNA, divided into 24 linear mollecules, the shortest 50M nucleotides, the longest 260M, each contained in a different chromosome.
- These 24 chromosomes consist of 22 autosomes and the two sex chromosomes, X and Y
- Some 35.000 genes are present in the human nuclear genome.

Centro de Investigación en Métodos de Producción de Software

Understanding the Domain (the Problem Space)

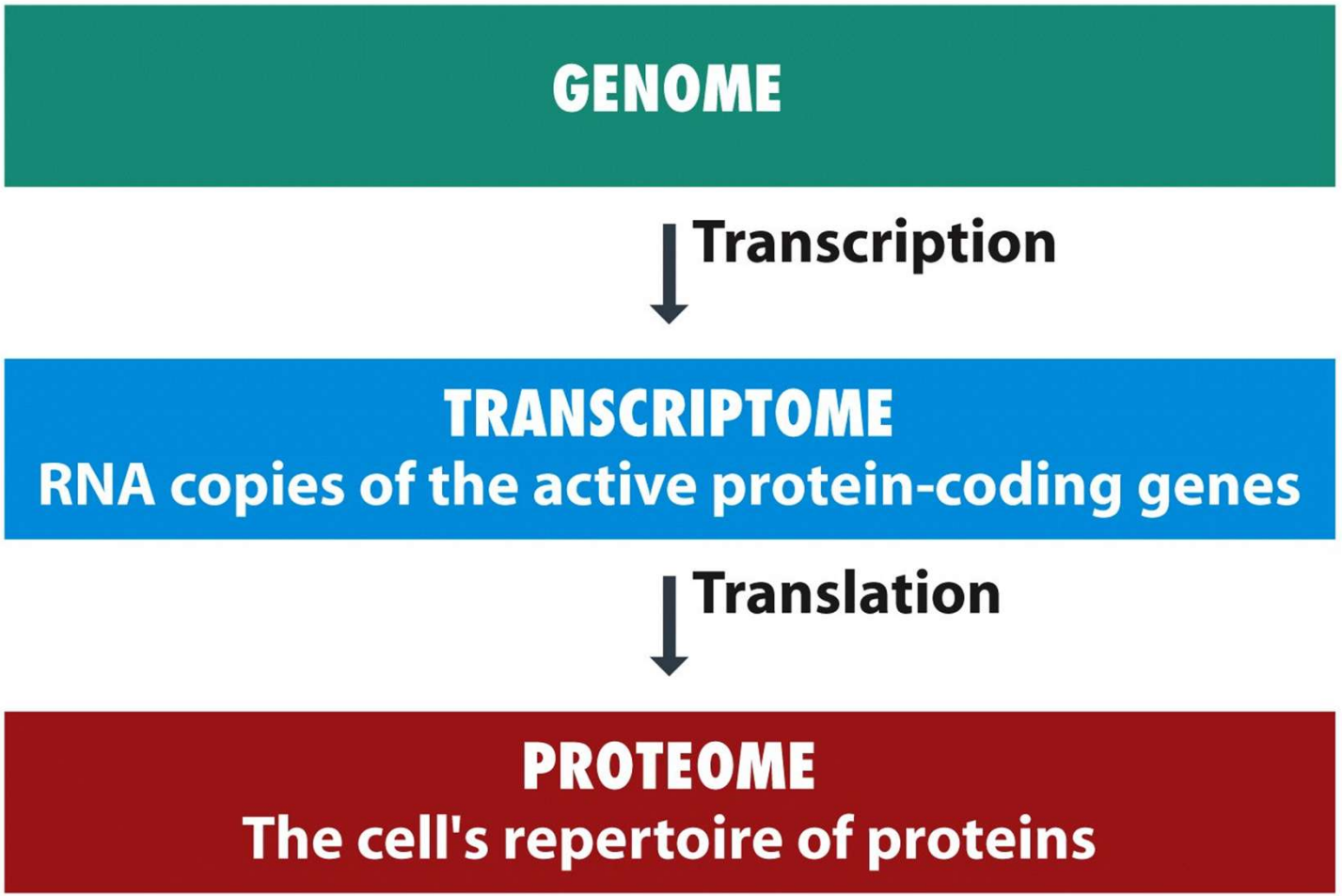


Figure 1.2 *Genomes 3* (© Garland Science 2007)

UNIVERSIDAD POLITECNICA DE VALENCIA

Centro de Investigación en Métodos de Producción de Software

Understanding the Domain (the Problem Space)

- Genes are made of DNA
- DNA is a linear, unbranched polymer in which the monomeric subunits are four chemically distinct nucleotides than can be linked in any order and in chains containing even millions of units in lenght

Centro de Investigación en Métodos de Producción de Software

From transcriptome to proteome

- Genetic code: how the nucleotide sequence of an mRNA is translated into the aminoacid sequence of a protein
- Proteins are made up from a set of 20 aminoacids
- Different sequences of amino acids result in different combinations of chemical reactivities
- Codon: codeword comprising three nucleotides
- Two-letter code is not enought, three-letter code provides 64 potential codons
- Code degeneracy
- Punctuation codons

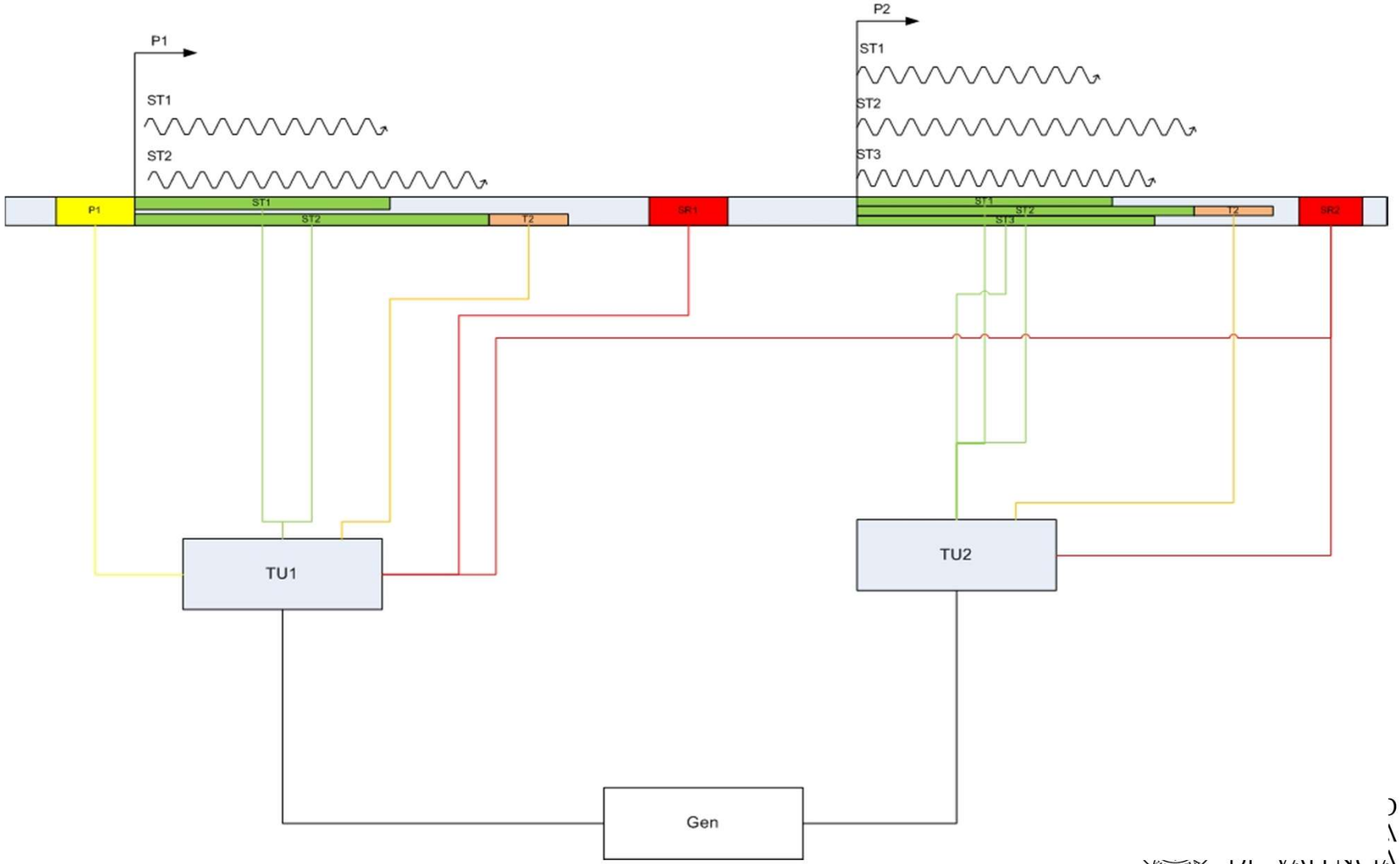
UNIVERSIDAD POLITECNICA DE VALENCIA

Building an ER Model

- Gene: A DNA segment containing biological information and hence coding for a RNA and/or polypedtide mollecule.
- Allele : One or two or more alternatives forms of a gene.

Centro de Investigación en Métodos de Producción de Software

Building an ER Model



Centro de Investigación en Métodos de Producción de Software

Genomic ER Model: Advantages

- Can be associated to different genomic databases and allows to use several gene identifications
- It has been described using terminology commonly used by biologists
- The definition of gene take into account that is not (always) a continuous sequence of bases
- The model does not include implementation details to a particular physical database schema

Centro de Investigación en Métodos de Producción de Software

Genomic ER Model: Advantages

- The Model is still to be refined and conceptually fixed…
- …but it provides a solid basis to incorporate contents in a precise and structured way
- … and the subsequent database can make possible an efficient use, content-oriented, where any human behaviour characteristic could be traced from fenotype to the involved gene(s)

UNIVERSIDAD POLITECNICA DE VALENCIA

Centro de Investigación en Métodos de Producción de Software

So many opportunities for the future!

- **Repairing Genetic Mutations With Lasers?**
  - *Physical base: DNA strands differ in their light sensitivity depending on their base sequences.*
  - *Conceptual base: need of understanding semantics behind given sequences of nucleotides*

- **Nature versus nurture**

UNIVERSIDAD POLITECNICA DE VALENCIA

- **Pre-implant Genetic Diagnosis**: a technique that allows to check if an embryo is/isn't healthy from a genetic perspective, before transfered to the maternal uterus.
  - Physical base: "assisted reproduction" technologies
  - Conceptual base: need to understand semantics of specific gene mutations

- Discovered a gene –**EYS** (for “Eyes Shut”) that **causes *inherited blindness***.
  - Physical base: mutation that gives rise to the problem
  - Conceptual base: why the mutation occurs? How to prevent it?

- Identified **295 potential therapeutics targets against AIDS**
  - Physical base: 295 human proteins that “probably” helps the AIDS to establish in the human cells
  - Conceptual base: “probably”? Under which conditions / interactions?

Centro de Investigación en Métodos de Producción de Software

Conclusions

- **Understanding the Human Genome** can become an extremely hard task if research is more and more oriented to the solution space
- Discovering “human” patterns in the genomic code is really like looking for a needle in a haystack.
- **Conceptual Modeling-based** approaches and techniques applied to this challenging domain should guide the efforts to succeed

UNIVERSIDAD POLITECNICA DE VALENCIA

Centro de Investigación en Métodos de Producción de Software

And more and more challenges to be explored…

- Linking diseases with genes with therapeutical purposes as a main application
- Gene mutations that enforce expression of some other genes while delaying or reducing the expression of others
- Gene regulators

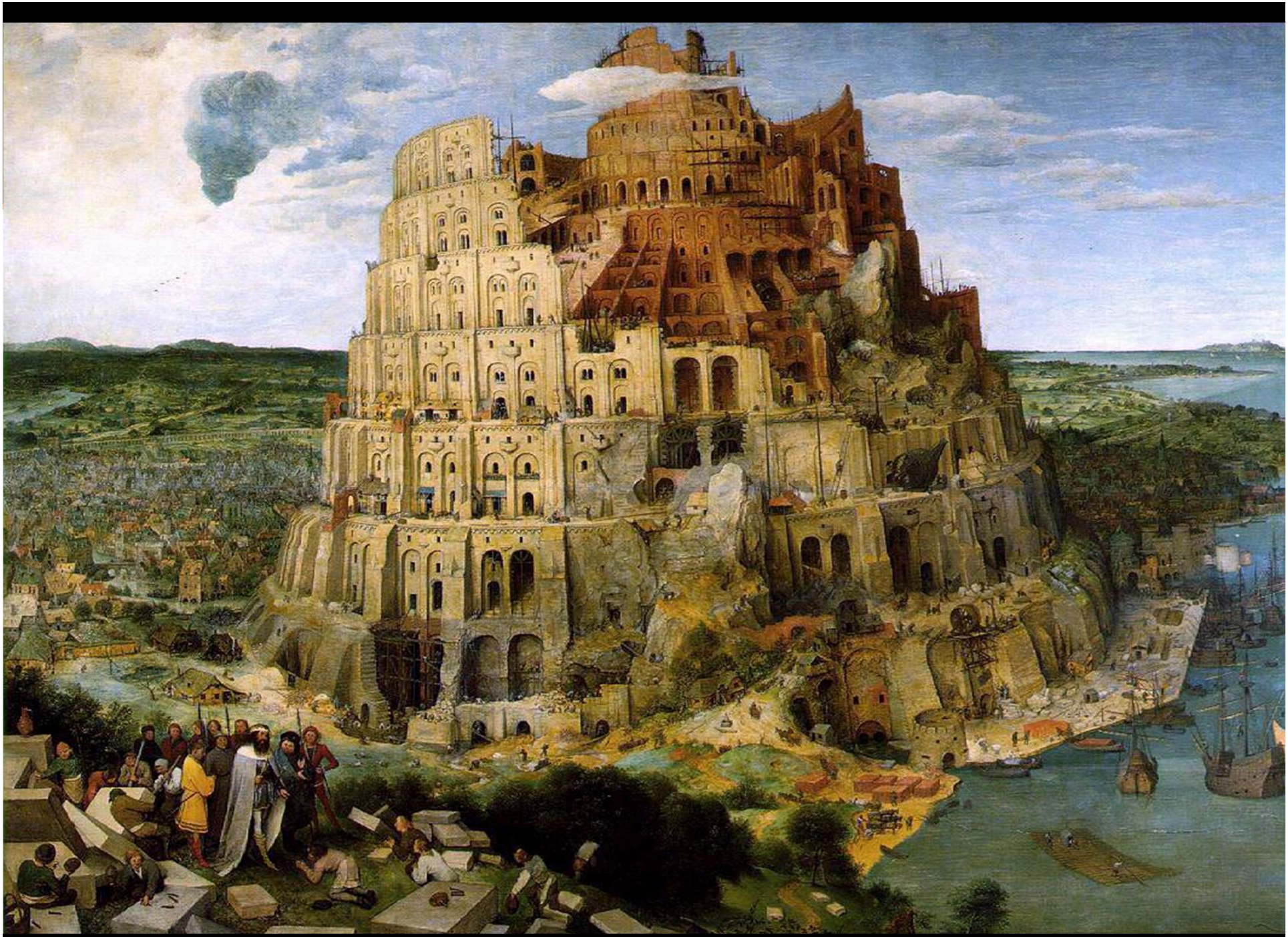
UNIVERSIDAD POLITECNICA DE VALENCIA

Centro de Investigación en Métodos de Producción de Software

Conclusions

Una polla xica, pica, pellarica, camatorta i becarica...
Immune system
Base pair
Protein
Transcribed sequence
RDF
Transcription
Genetic influences on female infidelity
Ontology
Exon
Human
Gene
Conceptual Modeling-based
Diagnosis
Cytosine
Cell
RNA polymerase
Conceptual model
Mutation
Terminator
Chromosome
Transcription unit
OO-Method
Genes against the malaria
ORF
Gene Ontology
Promoter
Guanine
Allele
Experiment
Nature versus nurture
Centromere
Intron
Neutral polimorphism
Regulator sequence
Chromosomic mutation
DNA
Aminoacid
GenoCAD
Widt type
Data bank
Proteone
OWL
Primary polipeptide
BioBricks
ORI
Genome
External identification
Inheritance
Allelic variant
Hydrogene bonds
Telomere
An 'infidelity' gene for men
Spliced transcript
Thymine
Research centre
Exon skipping
Intergenic region
HUGO
Enhanced sequence
Pre-implant genetic diagnosis
Ambient
Nucleotides
Mitocondrial genome
Adenine
Embryo
Entrez Gene
Vertebrate Genome Annotation
Genic mutant
Repairing genetic mutations with lasers
Codon
mRNA
'Fat' gene makes greedy
Major groove
Human Gene Mutation Database

UNIVERSIDAD POLITECNICA DE VALENCIA



Tower of Babel Pieter Bruegel the Old (Breda of Bree 1525 – Brussels 1569)

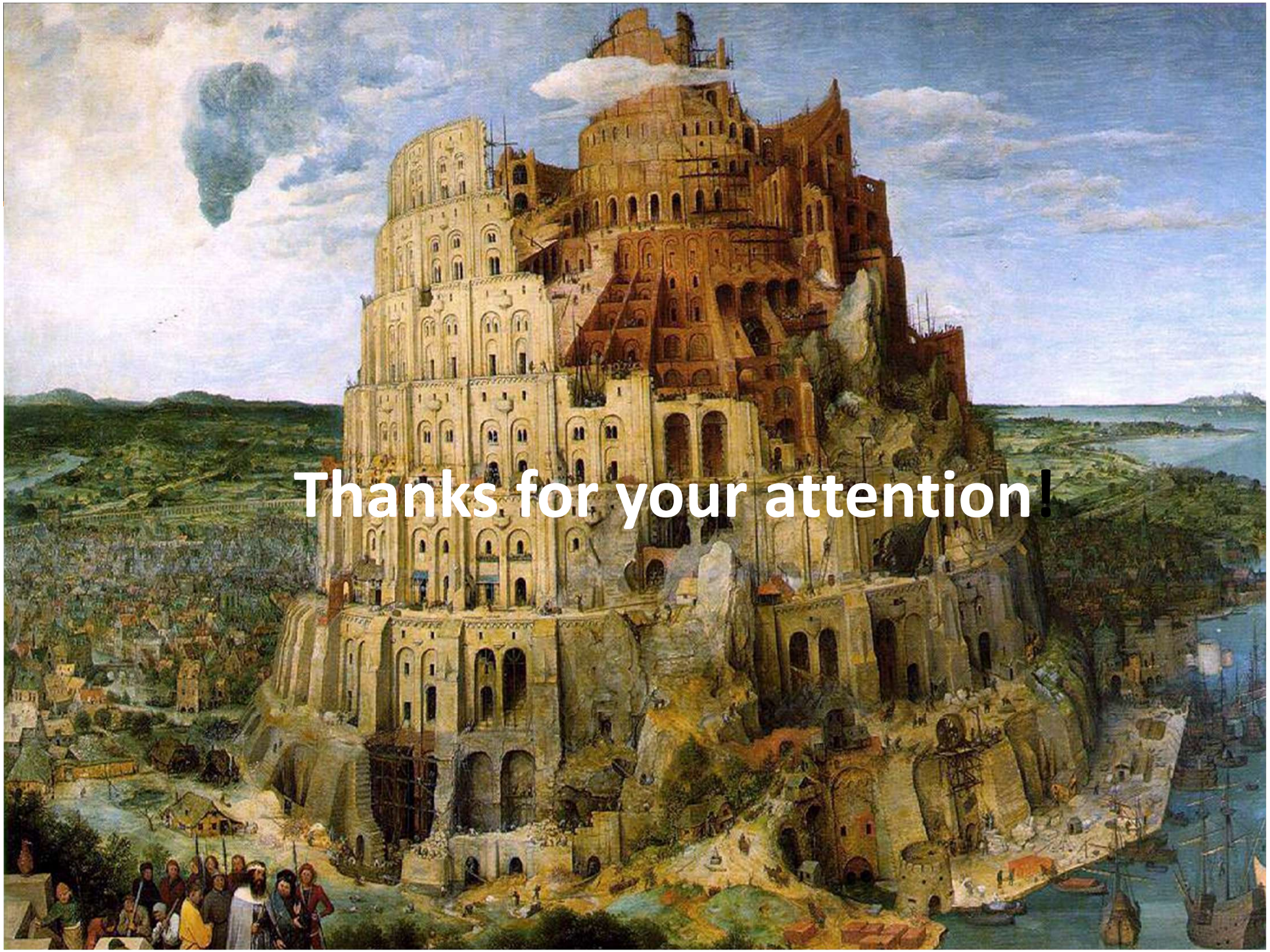
Centro de Investigación en Métodos de Producción de Software

Conclusions

- This is probably the most attractive challenge in the future of the Conceptual Modeling community:

Modeling the Real Life to understand why we are as we are, and how a human being can be seen as the "representation" of a Conceptual Model that can be specified in detail

UNIVERSIDAD POLITECNICA DE VALENCIA


Thanks for your attention!