

The Semantics of Genomic Analysis

Robert Stevens

BioHealth Informatics Group
School of Computer Science
University of Manchester

robert.stevens@manchester.ac.uk



The Collaboration

“Developing a GRID-based system for integrating and exploring data from comparative genomics, to discover biological knowledge that can not be discovered from any one source”

Collaborative BBSRC project

– 5 sites across the UK

<http://www.comparagrid.org>



Introduction

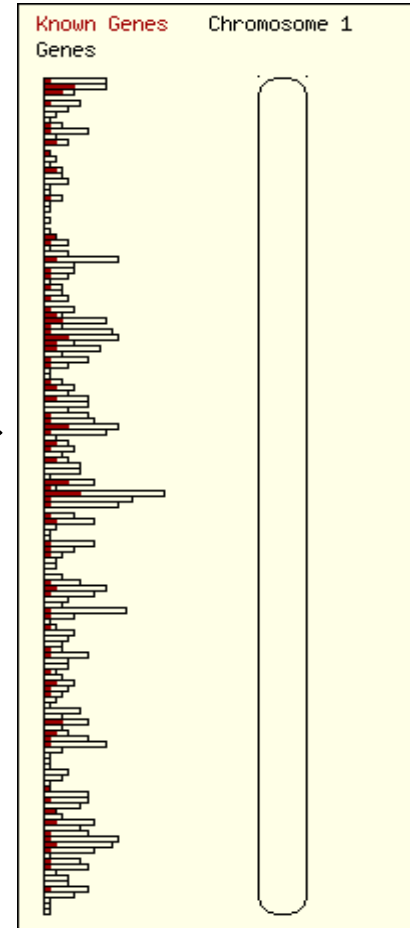
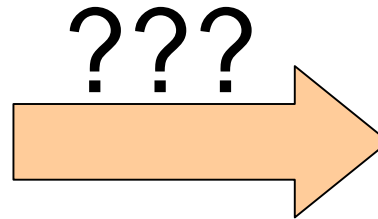
- The general problem
- An architecture for getting answers
- Forming the questions and making distinctions
- Some observations



Sleeping Cows: Unknown Genes

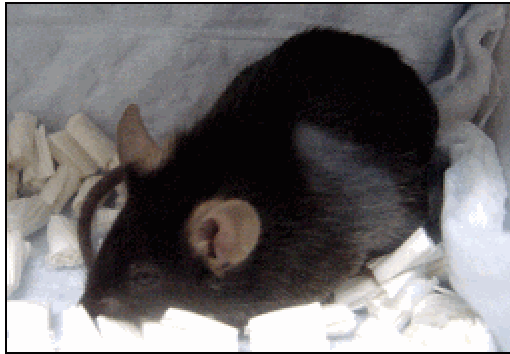


Cow Infected with African Trypanosomiasis

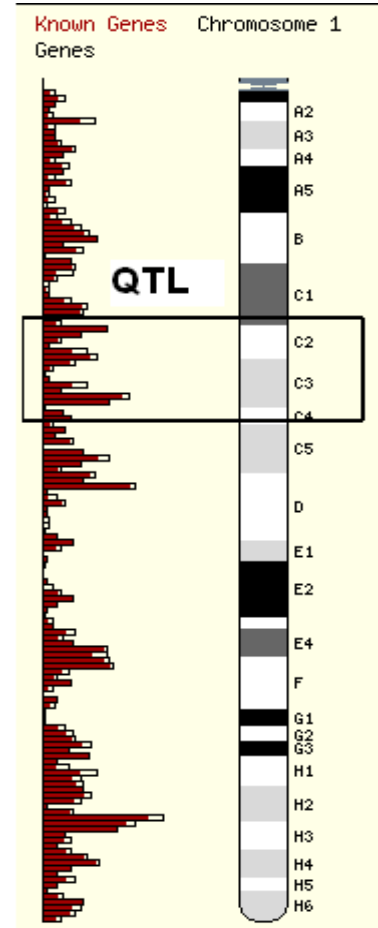
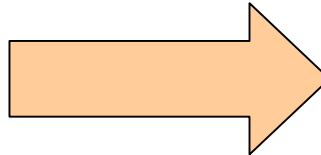


Cow chromosome and known Genes

Sleeping Mouse Model with QTL

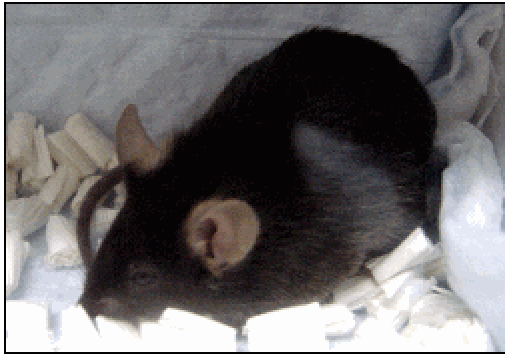


Mouse Infected with African Trypanosomiasis

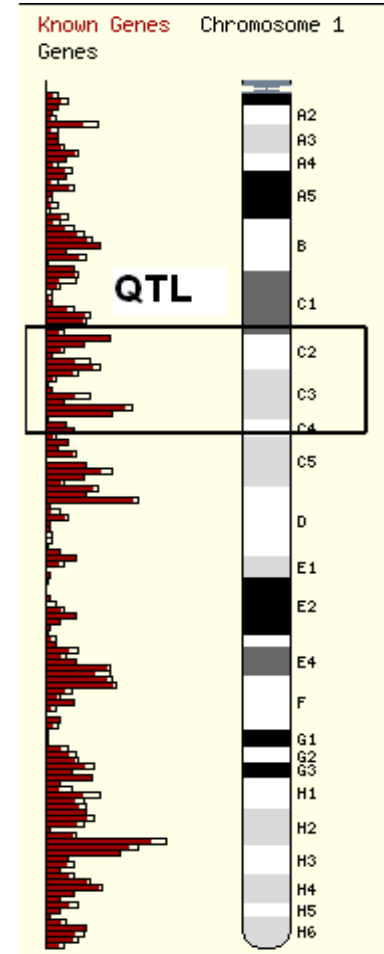
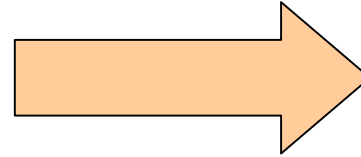


Mouse QTL region on chromosome, encompassing multiple genes

We can Find Out About Mouse Resistance



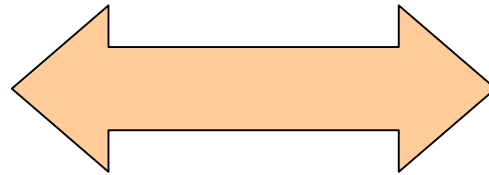
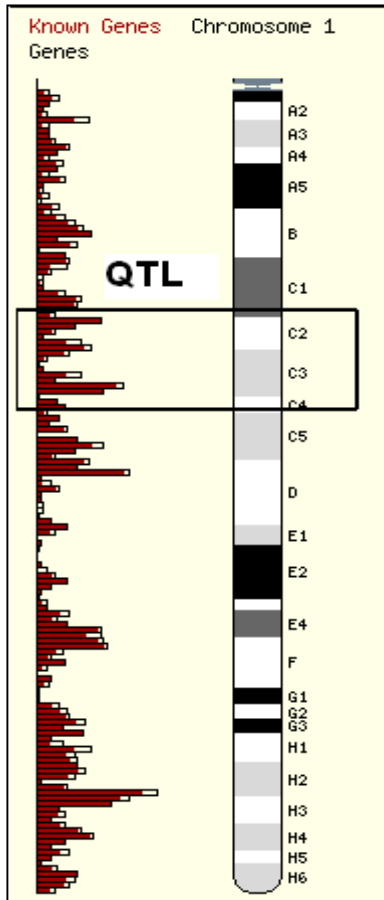
Mouse Infected with African Trypanosomiasis



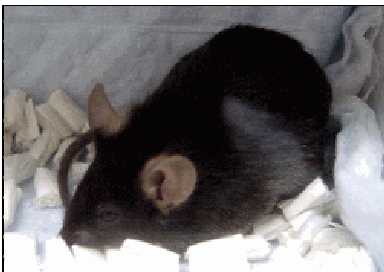
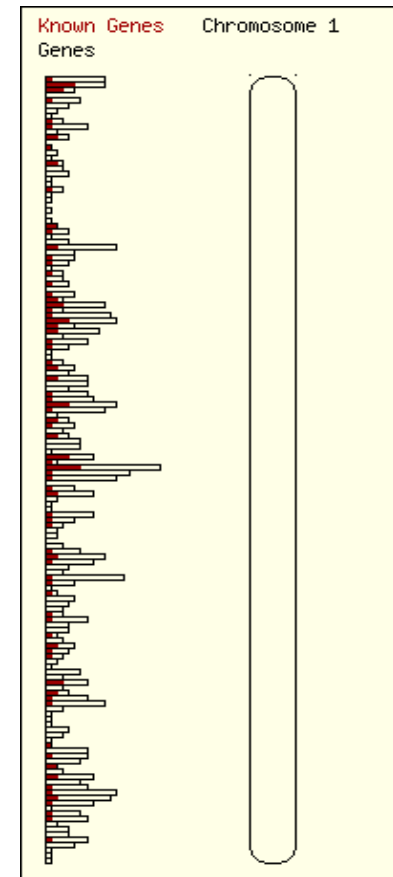
Mouse QTL region on chromosome, encompassing multiple genes

Genes A, B, C, D in QTL are involved in Trypanosomiasis resistance

Use Mouse to Tell us about Cow



Infer presence and
order of Cow
genes from
presence and order
of Mouse genes

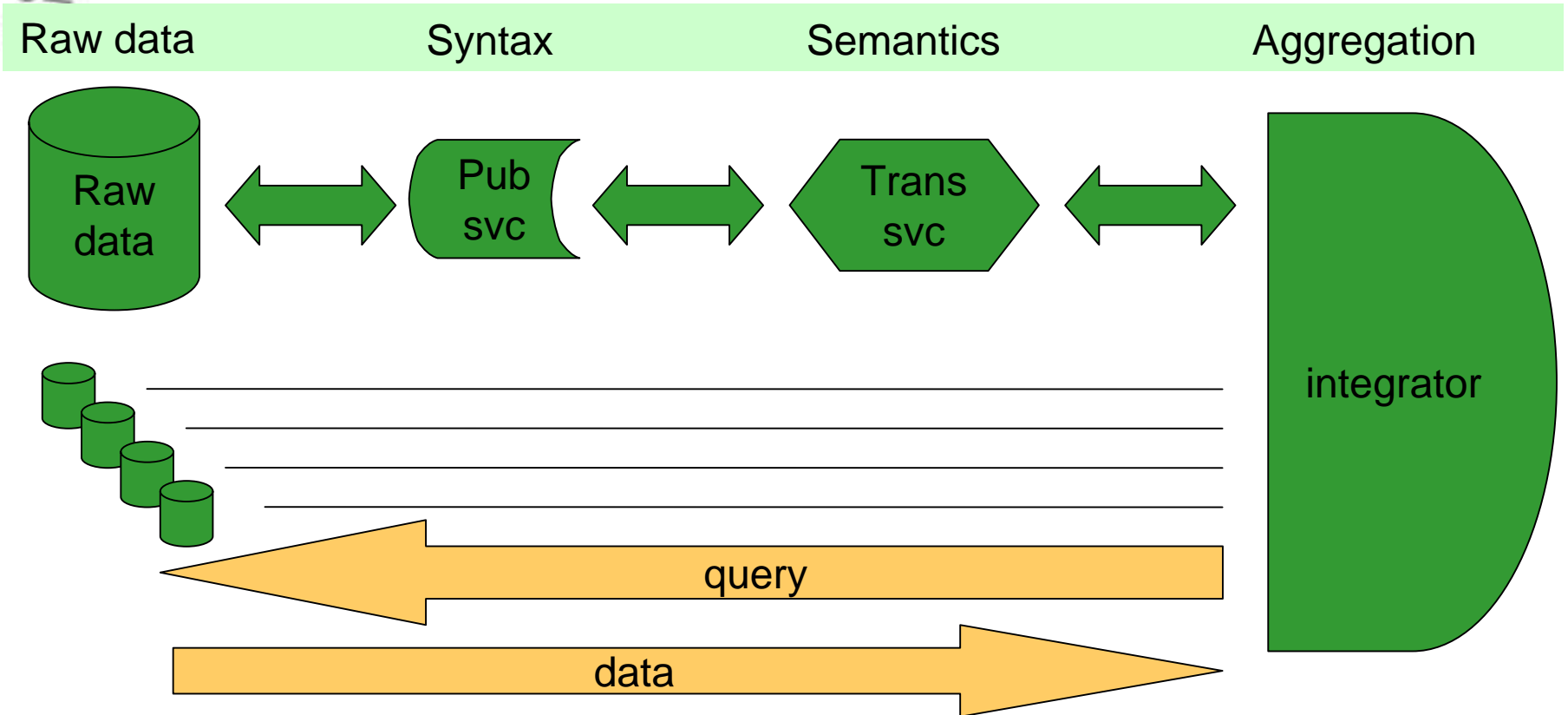


Requirements

- Ask for maps that contain *marker* from species x in species y, z, etc
- What chromosome or part of chromosome does a map model?
- What is the gene order in species x and species y?
- Also notions of similarity, homology, and, of course, synteny
- Orthology and paralogy



The Fluxion Stack



Query Semantics

- Query OWL class interpreted as
 - K query – (poor man's) epistemic closure of query
 - Against knowledge-base exposed by that data-source, not The World
- Result is a knowledge-base
 - All entailed by queried KB (it's a subset)
 - Can be statements in the original KB, or any statements that always follow
- Contains *at least* the statements needed to
 - Allow a reasoner to classify all the individuals who match the query correctly
 - Preferably using properties, not asserted types (a-box preferred over t-box, don't over-commit)
 - Speed for accuracy
 - Implementation complexity for data-volume
 - Return all instances of known classes e.g. db table with minimal filtering – if in doubt, return it

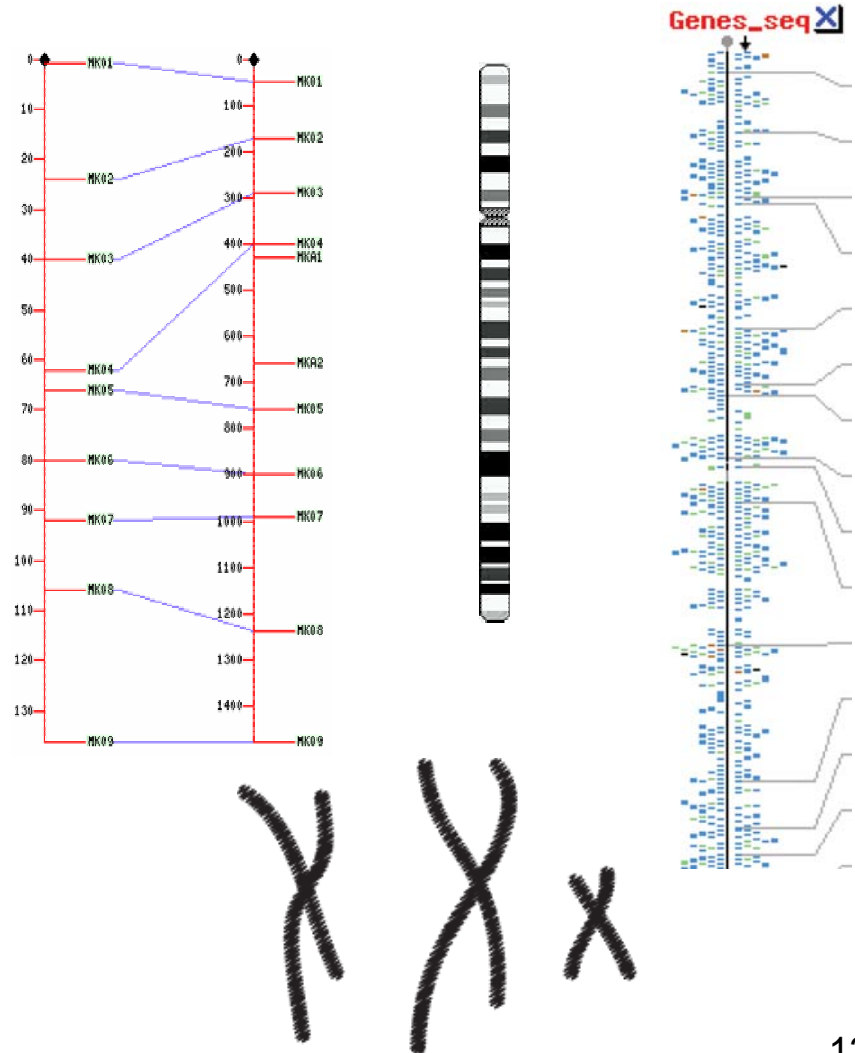
Role of Ontology in Fluxion

- Fluxion
 - Uses semantics of OWL
 - Not any domain-specific information
 - Any domain
- A domain ontology defines what Fluxion integrates
- Developing a ‘good’ domain ontology is
 - Hard work
 - Poorly scoped
 - No widely-validated methodology
 - Biologist \neq Modeller so language gap



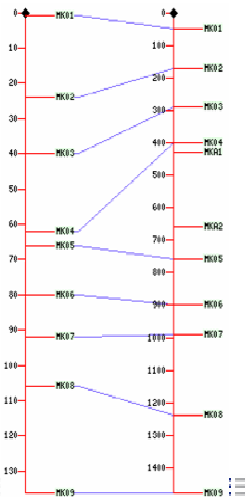
ComparaGRID Ontology Scope

- Genetics
 - Markers and Maps
- Genomics
 - Genomes and Sequences
- Comparative Aspects
 - Evolutionary relationships, Similarities
- Physical entities
 - Chromosome, Organism



Models and Representations

- We ‘know’ things about physical biological things
 - We know other things about the way we render information about these things
 - What reconciles a map and a sequence of a thing is the thing itself
 - A “knowledge anchor”



has model

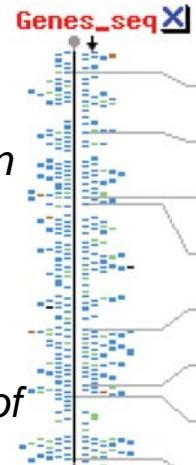
is model of



Physical Biological Entities

has representation

is representation of

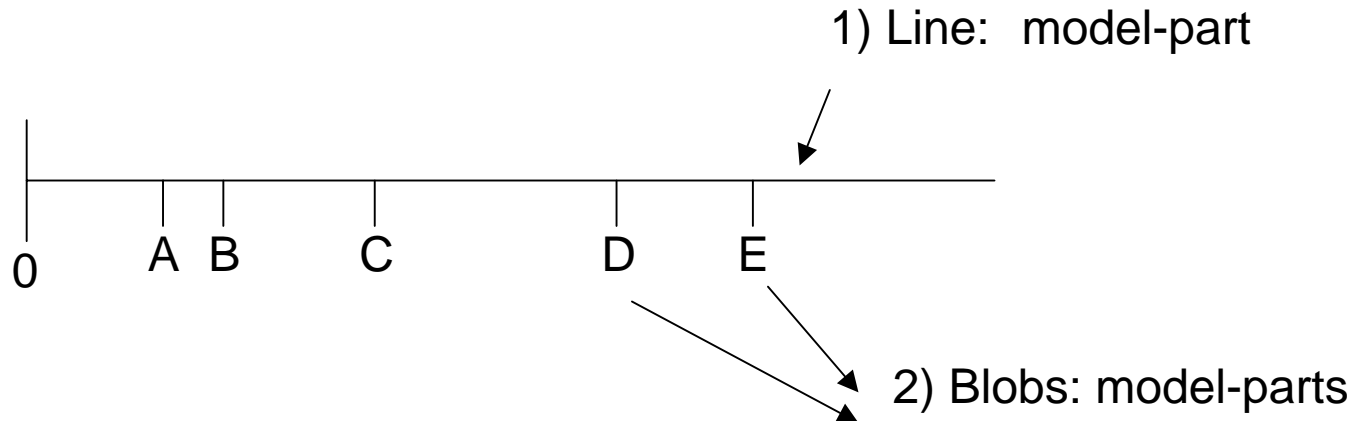


Sequences



Modelling Maps

- A map is an abstract model of a physical thing
 - Good for ordering limited knowledge
 - Minimal *explicit* biology in a map, just *implicit* biology in the labelling of things
 - Maps can be modelled as a combination of lines and blobs:

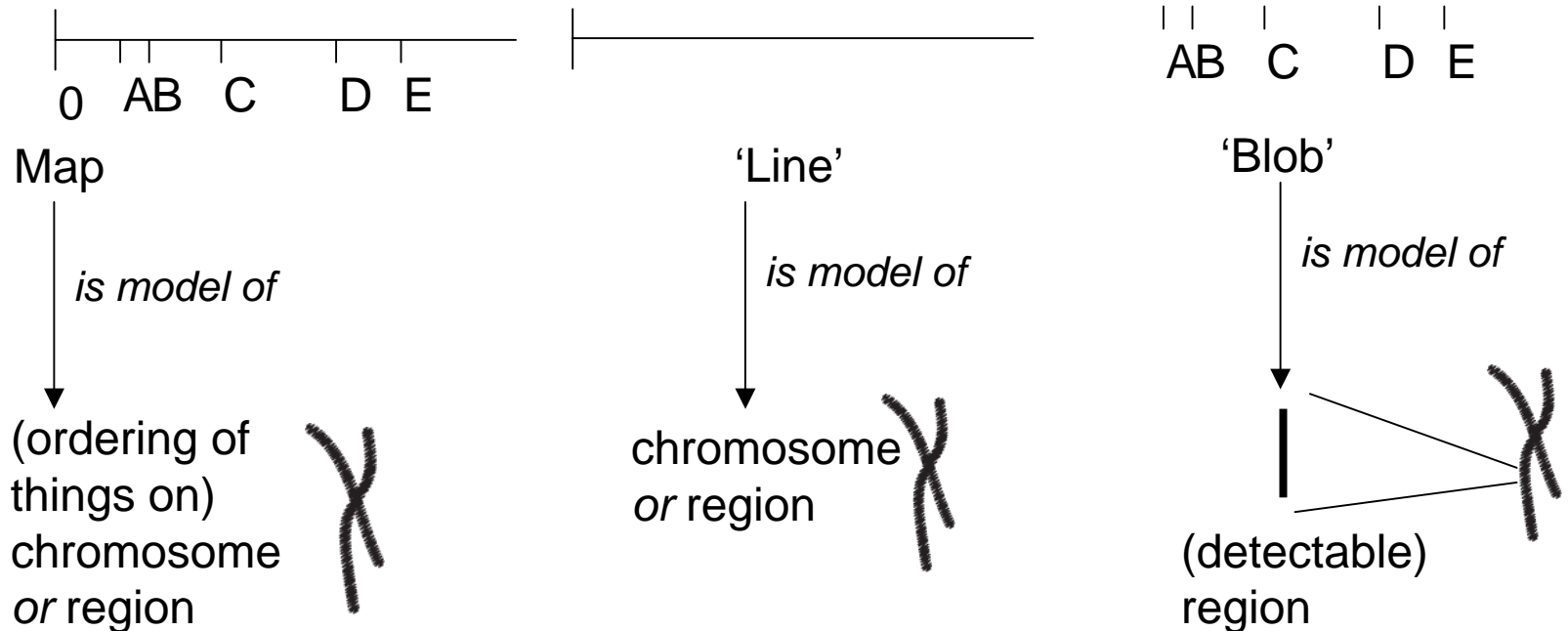


3) Map: Model of ordering of blobs within line

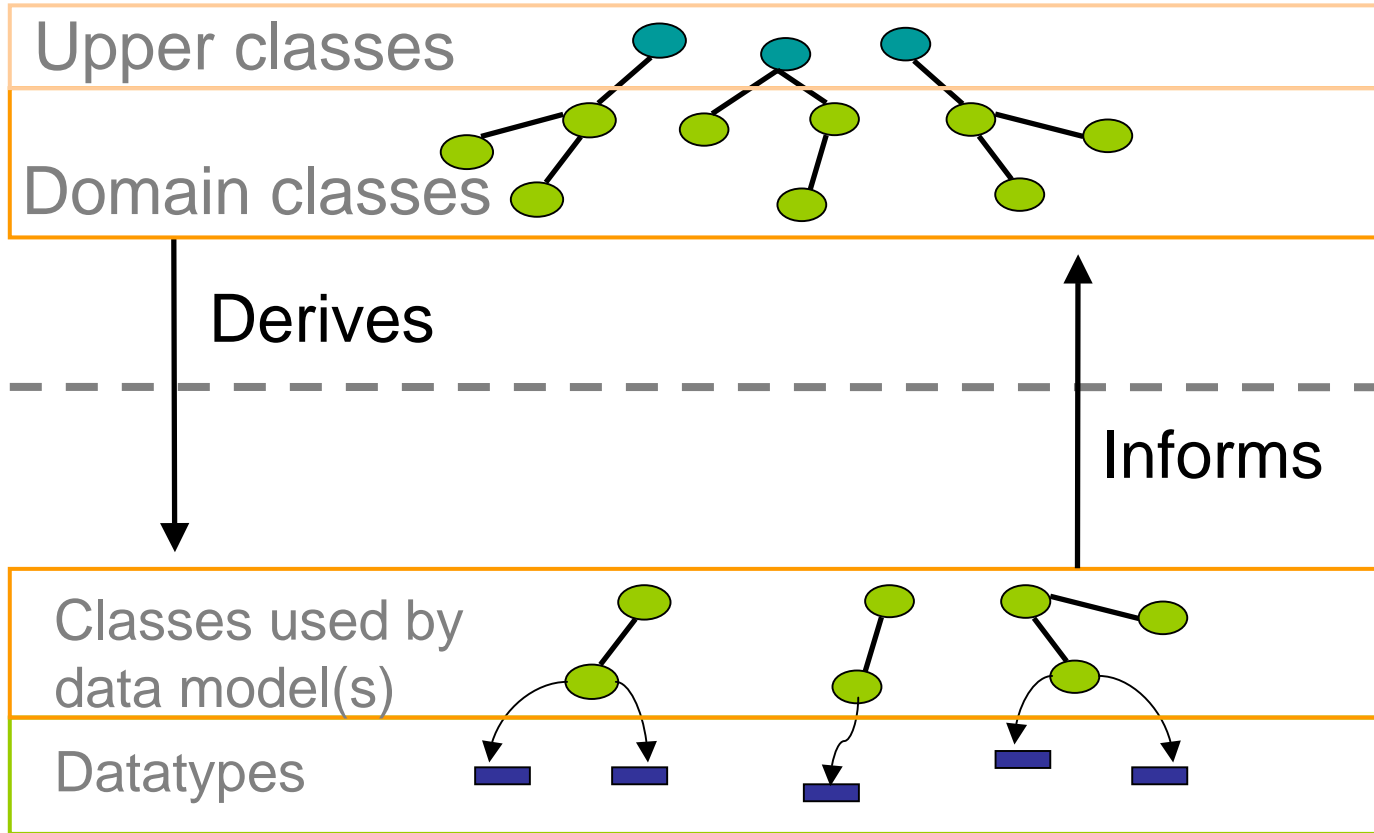


Modelling 'Markers'

- 'Marker' is overloaded term
- Meaning here is captured by linking lines and blobs up to physical things



Strategy: Domain and Application Ontologies



TAMBIS AllOver Again

- ..., but now we have different technologies; bio-ontologies galore; different resources.
- Mainly SQL queries over RDBMS
- Can actually answer the questions we ask
- Rather than mapping down to the resources; we map down from the global model and up from the resources into an OWL world



Main Points

- Its hard: Sophisticated ontologies don't appear by magic
- Open world good, open world bad
- Need to do local closure: A poor man's epistemic operator...
- It isn't completely correct
- A sophisticated ontology in a SW application
- Highly axiomatised ontologies for the Semantic Web?

