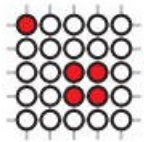
Data representation for clinical data and metadata



cancergrid

WP1: Data representation for clinical data and metadata

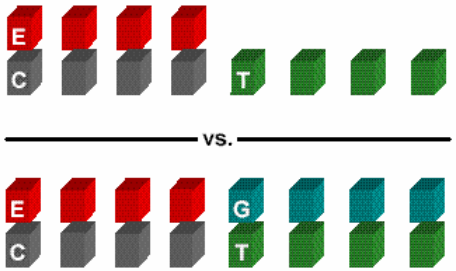
- Inconsistent terminology creates barriers to
  - identifying common clinical entities in disparate information systems and automatic linking
  - integrating clinical and genomic data
- Remove this barrier through the establishment of a consensus regarding domain-specific vocabularies, ontologies and data items.
- **WP1 acts as "Rosetta Stone" – data can be mobilised**

cancergrid

“Trials of the future”

tAnGo

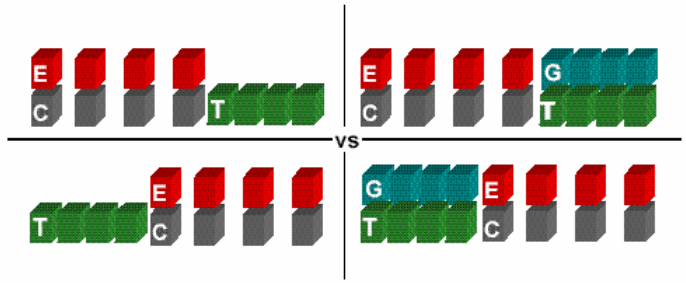
Paclitaxel , Anthracycline, Gemcitabine & Cyclophosphamide



A randomised phase III trial of gemcitabine in paclitaxel-containing, epirubicin-based, adjuvant chemotherapy for women with early stage breast cancer

Neo-tAnGo

Neoadjuvant Paclitaxel , Anthracycline, Gemcitabine & Cyclophosphamide



A neoadjuvant study of sequential epirubicin + cyclophosphamide and paclitaxel ± gemcitabine in the treatment of high risk early breast cancer with molecular profiling, proteomics and candidate gene analysis.

New data types - data integration – data sharing

cancergrid

Translation of clinical data to genomics and bioinformatics research

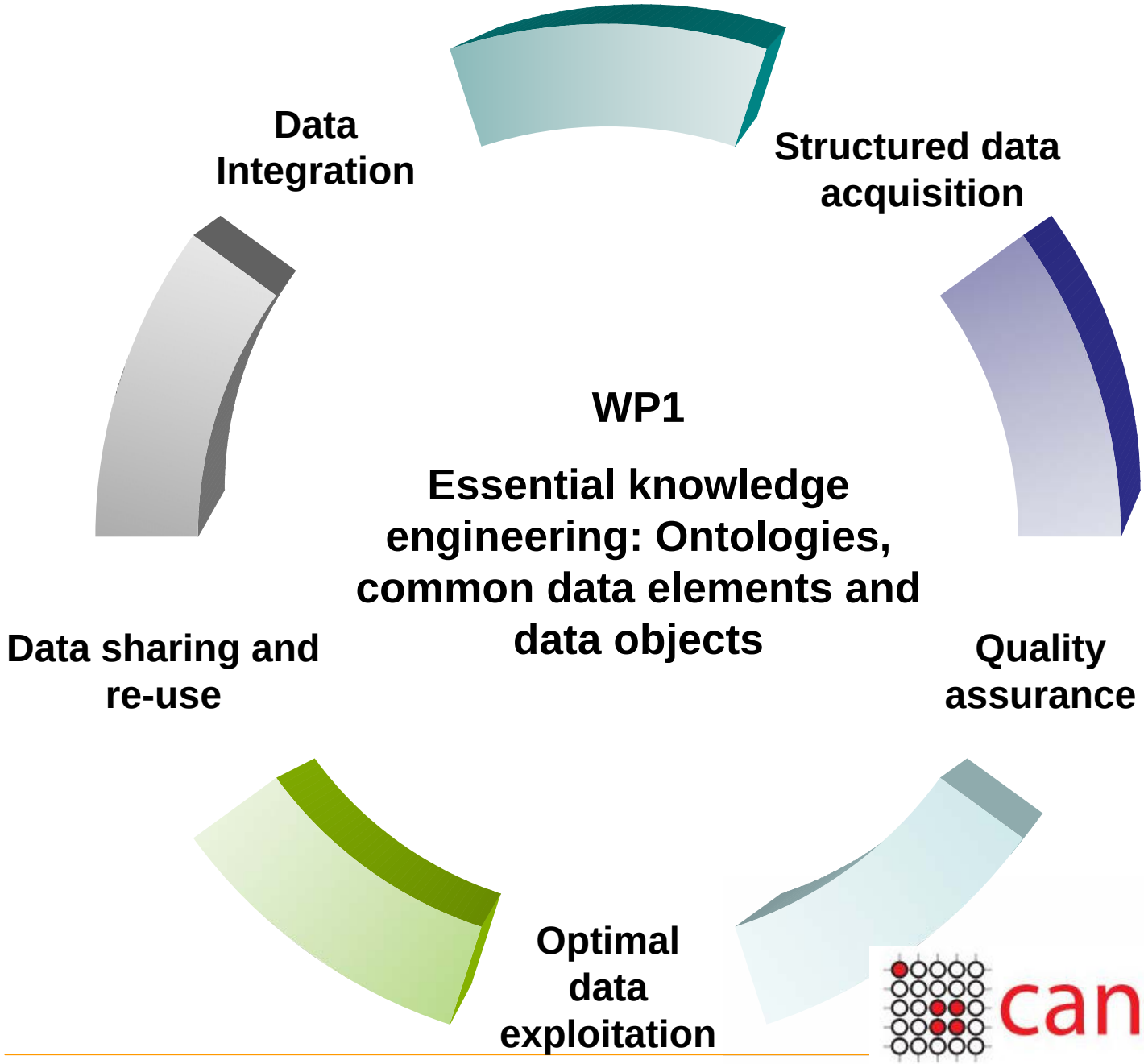
Molecular data  **Clinical data**

Ontology framework

Integrative analysis

Therapeutic systems

cancergrid


Data Integration
Structured data acquisition
WP1
Essential knowledge engineering: Ontologies, common data elements and data objects
Data sharing and re-use
Quality assurance
Optimal data exploitation

cancergrid

Systems biomedicine: Integrating physiology and genomics

Sylvia B. Nagl PhD

Cancer Systems Biology & Biomedical Informatics

Department of Oncology

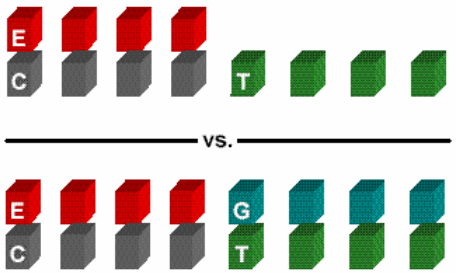
Translating clinical trials to genomics research

- Fast accumulating molecular data from fundamental and translational cancer research, genomics, proteomics etc.
- Clinical trials increasingly incorporate detailed molecular analysis of tumour and serum samples
- Challenges:
  - Elucidation of the relationships between molecular data and physiology
  - (Ontological) integration of multi-scale data
  - Translation of data into knowledge – understanding of cell and tumour systems

“Trials of the future”

tAnGo

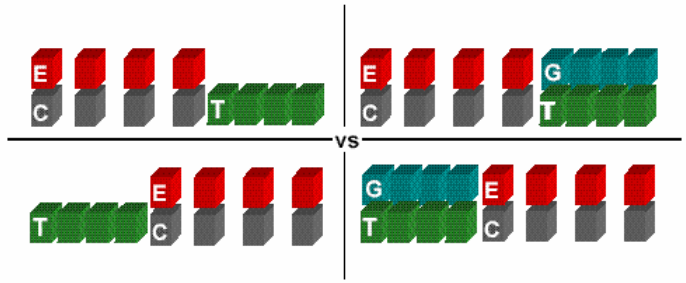
Paclitaxel , Anthracycline, Gemcitabine & Cyclophosphamide



A randomised phase III trial of gemcitabine in paclitaxel-containing, epirubicin-based, adjuvant chemotherapy for women with early stage breast cancer

Neo-tAnGo

Neoadjuvant Paclitaxel , Anthracycline, Gemcitabine & Cyclophosphamide



A neoadjuvant study of sequential epirubicin + cyclophosphamide and paclitaxel ± gemcitabine in the treatment of high risk early breast cancer with molecular profiling, proteomics and candidate gene analysis.

UCL

Neo-tAnGo

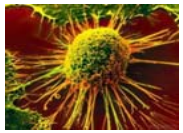
Secondary endpoints

- **Tumour tissue** analysis
  - **Prognostic and predictive markers and profiles**

    Molecular profiling using expression and DNA microarrays

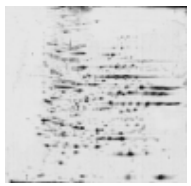
    Immunohistochemistry (IHC) and in situ hybridisation
- **Serum proteomics** - protein profiles

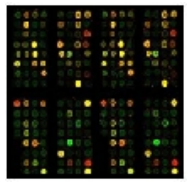
Markers and systems



clinical parameters

cell and tumour physiology



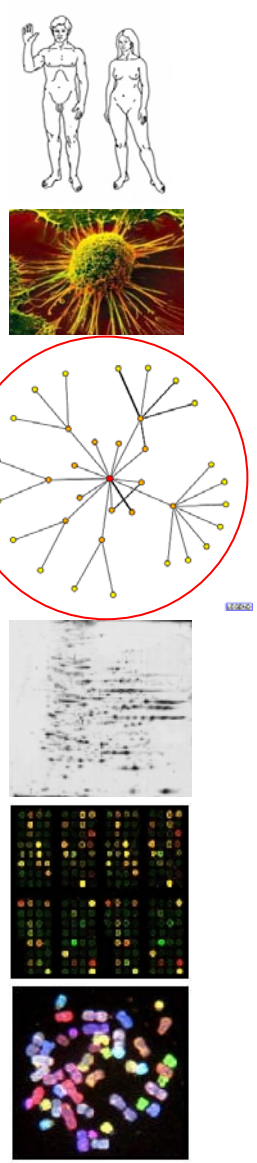


proteins

gene expression

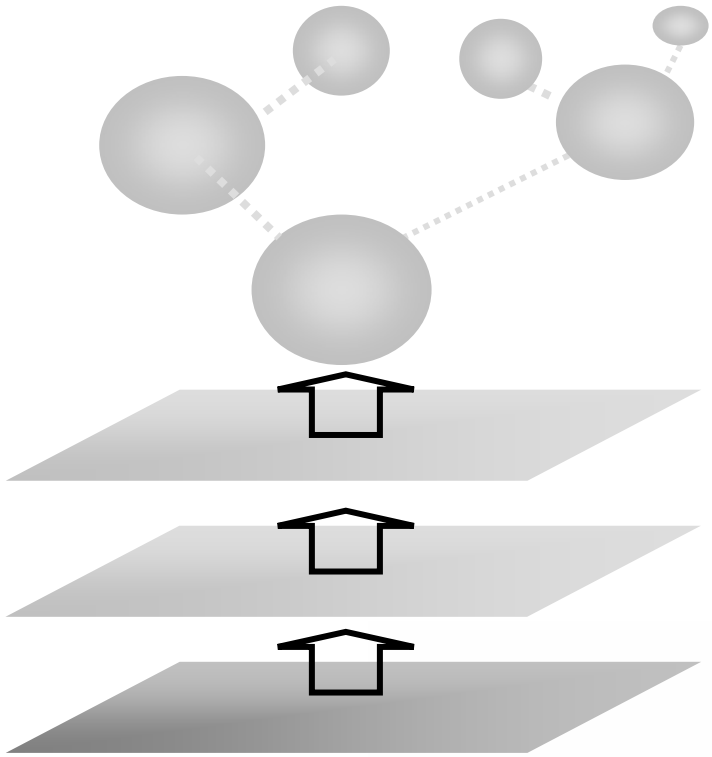
genome

Pathway modelling

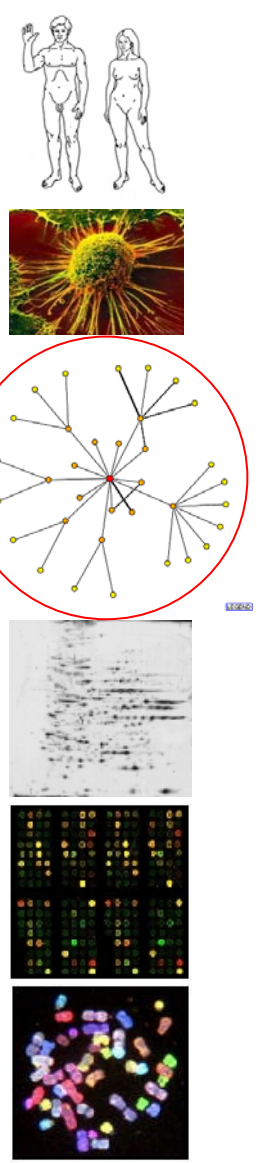


clinical parameters

cell and tumour physiology

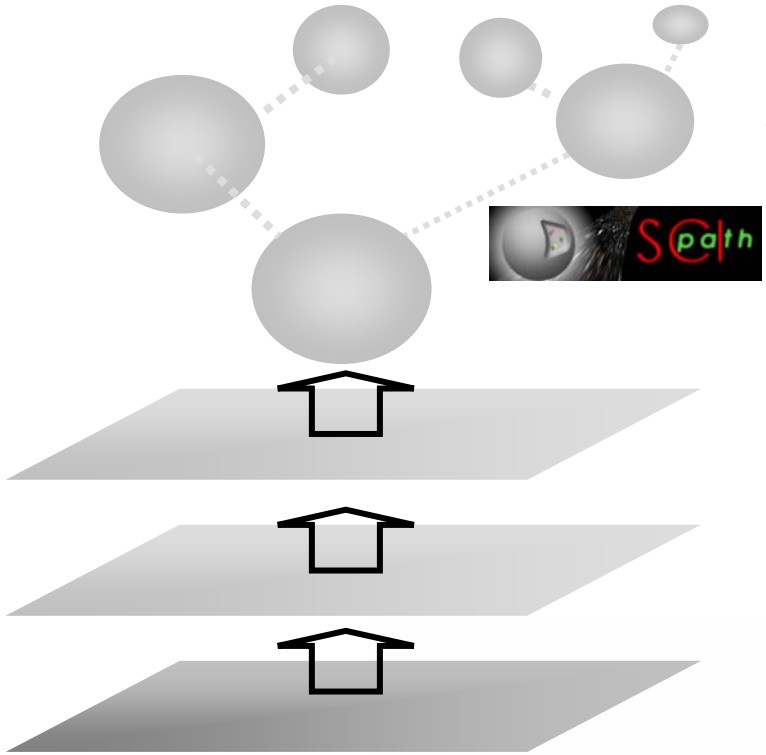


Pathway modelling



clinical parameters

cell and tumour physiology



SCIpath project

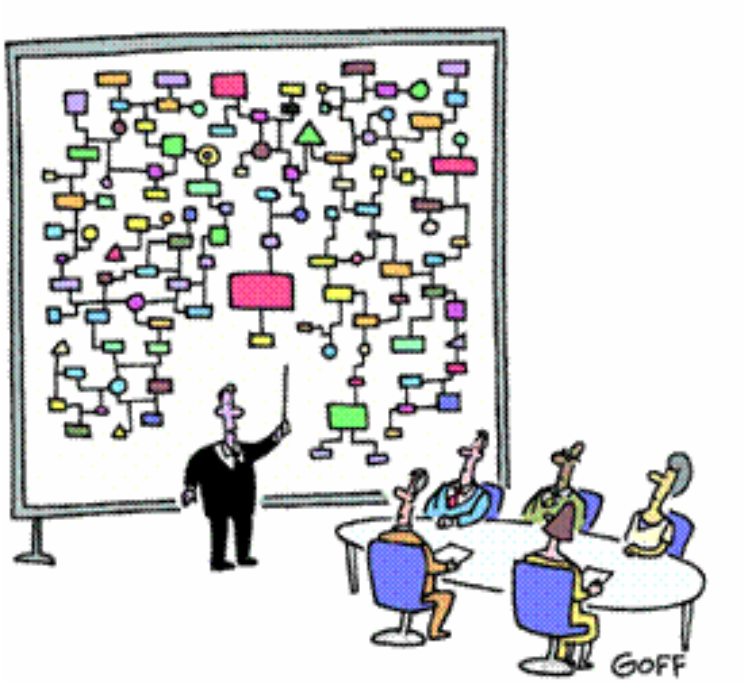
www.scipath.org.uk

UCL

Pathway modelling

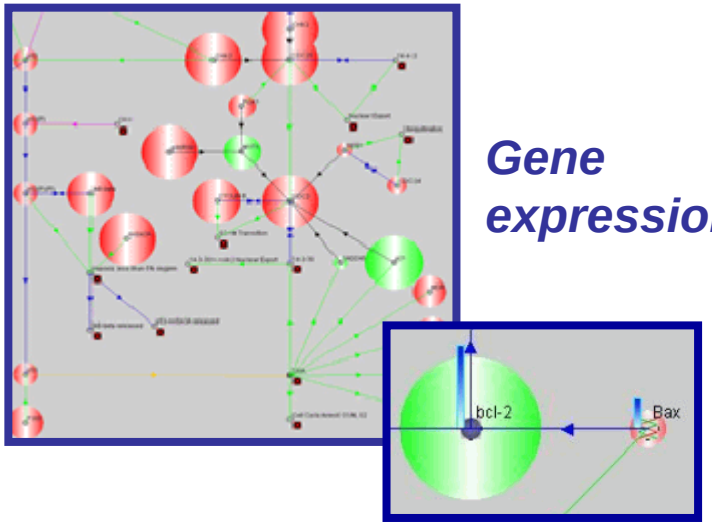
- Pathways located midway between the genome and the phenotype, and can be conceptualised as an ‘extended genotype’ or ‘elementary phenotype’.
- Intrinsically offer an integration framework for modelling of biological processes, and the translation of clinical studies of novel treatments to the molecular and genomic level.
- A formal framework for the expression of complex biological knowledge, assumptions and hypotheses in a form amenable to logical analysis and quantitative testing.
- Computational analysis is increasingly necessary as the scope and depth of information and knowledge, with the accompanying uncertainty, surpass the analytical capabilities of the unaided human mind.

Pathway modelling

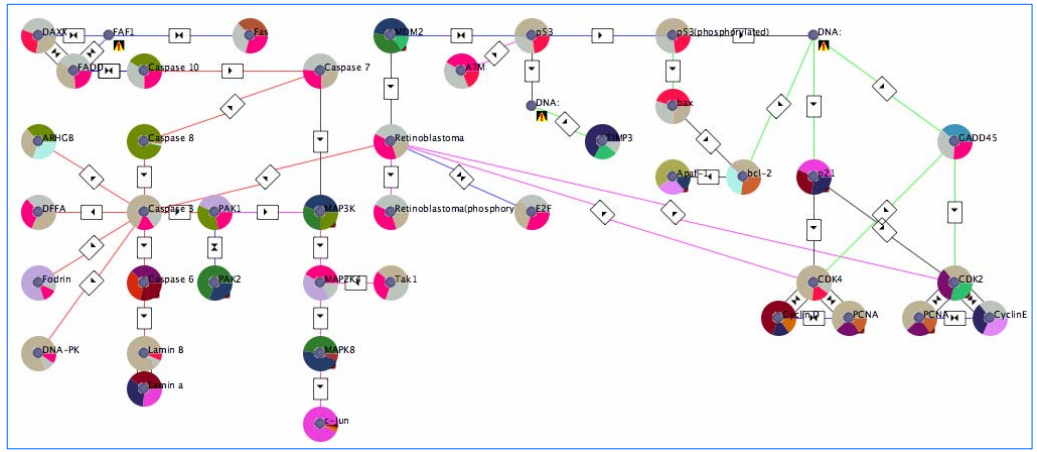


"And that's why we need a computer."

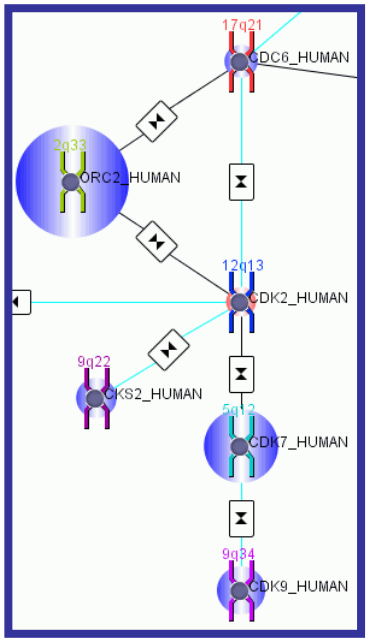
http://sbw.kgi.edu/



Gene expression



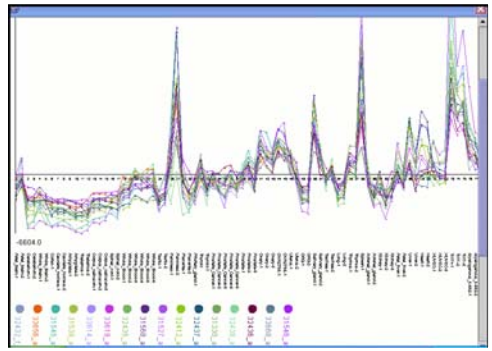
Co-expression patterns

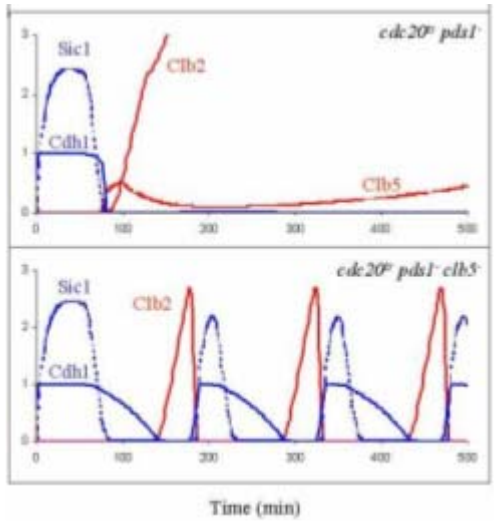


Gene-protein expression

Gene copy number

Cluster analysis

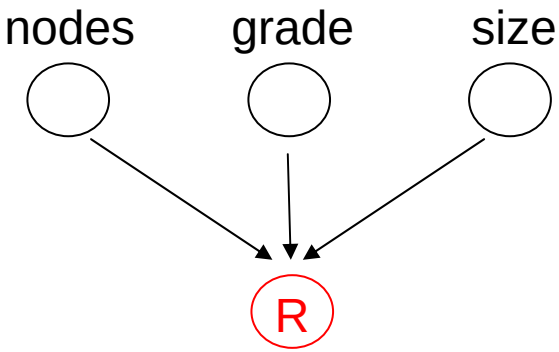




Simulation

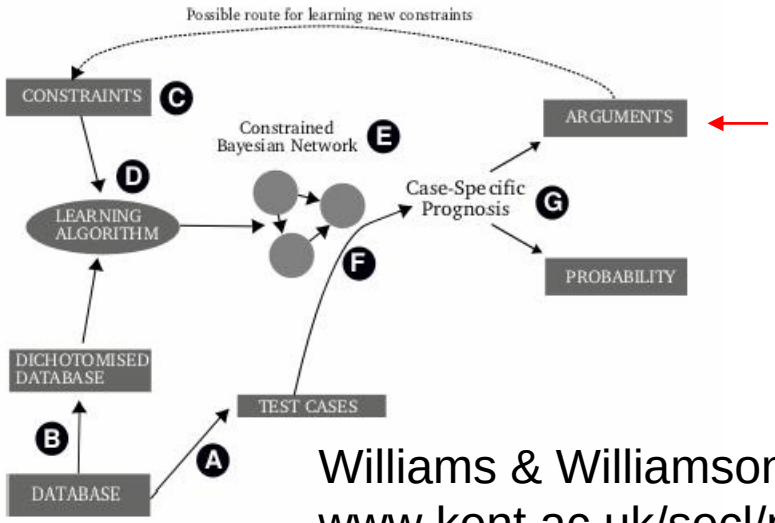
Bayesian network modelling

Decision support systems



Calculate prognostic probabilities

P(Recurrence | X)



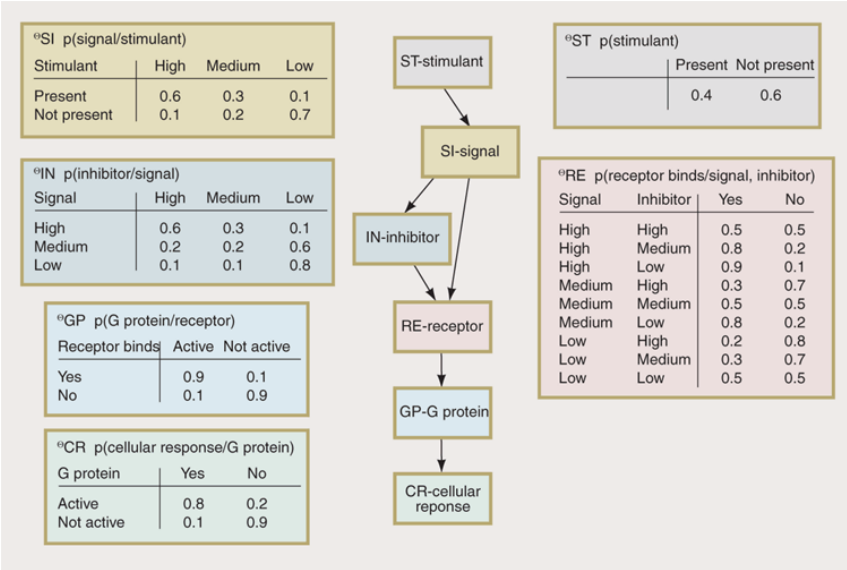
Logic

Rules

Williams & Williamson, 2005:
www.kent.ac.uk/secl/philosophy/jw/

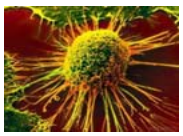
Bayesian nets as pathway models

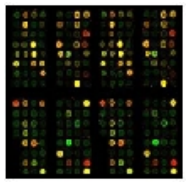
- **Bayesian networks are increasingly important for integrating biological data and for inferring cellular networks and pathways.**

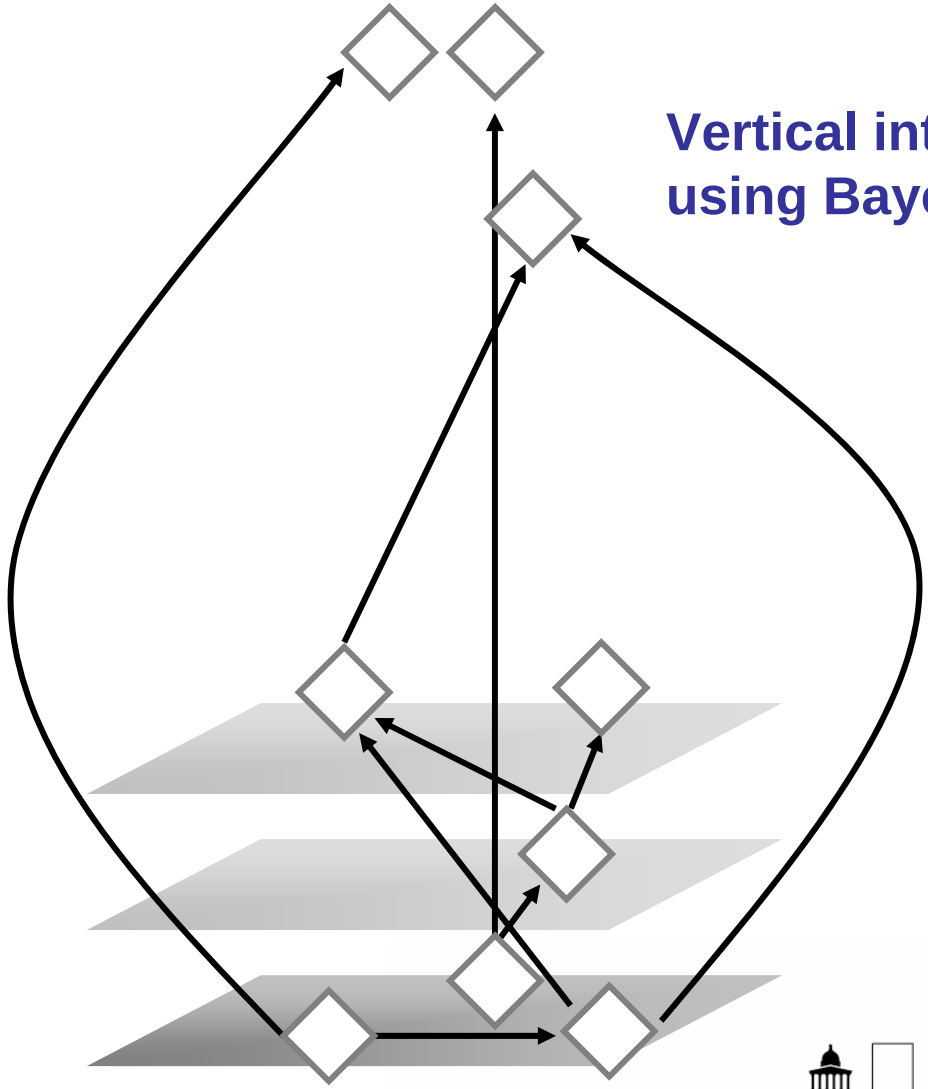


Needham et al., Nature Biotechnology **24**, 51 - 53 (2006)

Multi-scale predictive profiles







Vertical integration using Bayesian nets

Probability

Logic

Knowledge integration

- Development of Bayesian network integration methods
  - e.g., integration of predictive models from separate clinical trials
  - integration of predictive parameters from clinical and molecular levels
  - evolvability, integration of new knowledge, new data and new data types (e.g., epigenomics)

- Collaboration with Williams (CR UK) and Williamson (U of Kent)

UCL

Integration with Cancergrid

- SCIpath and Bayesian tools will implement CG standards

- Tools can be integrated in CG-caBIG

Conclusions

- Clinical trials incorporating molecular profiling create the challenge and opportunity for new integrative computational approaches
  - Pathway modelling – effects of differential gene expression
  - Bayesian networks – integration of molecular and clinical parameters for prediction, probability + logic
- Integration of genome structure, molecular processes and physiology (‘vertical integration’, multi-scale profiles)
- Systems understanding

s.nagl@ucl.ac.uk

UCL