

Neuroinformatics at GSK and Application to Target Identification and Validation

Chris Larminie

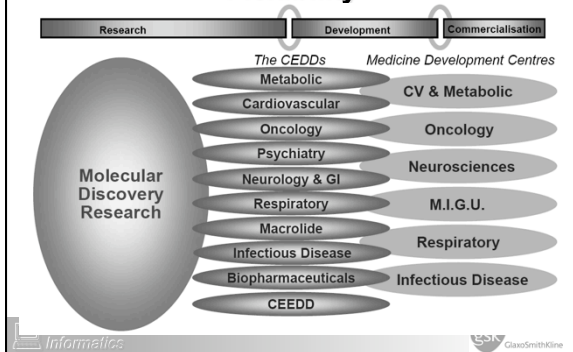


Presentation Summary

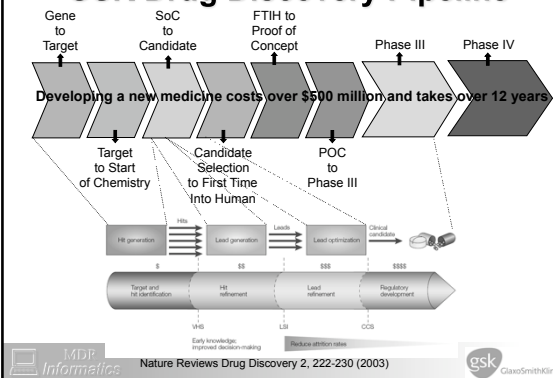
- QUICK overview of GSK R&D
- Integrating data
- Network / pathway analysis
- Neuroinformatics and issues pertinent to this field
- Informatics and Industry



R&D's Structure for Innovation & Flexibility



GSK Drug Discovery Pipeline



Nature Reviews Drug Discovery 2, 222-230 (2003)



MDR Informatics Areas

- Gene Identification, validation & classification
- HTS support, compound collections, (sub)structure searches, hard/soft filtering, HIT identification
- Data integration, application development
 - Unless integration is needed for a specific analysis project
- LIMS, data collection and pipelining
- Platform support (primer designs, SNP assay selection, etc.)
- Data analysis and interpretation
 - Sequence analysis
 - Platform data analysis
 - Network/Pathway analysis
 - Structural analysis
- New methods development, assessn



SYSTEMS BIOLOGY

- Progress in systems biology may be seen to rest on (*at least*) a three-way foundation:
 - **Pathways/Networks:** understanding the interactions that comprise biosystems
 - **Ontologies:** the effective representation of biological knowledge in all its richness
 - **Data Integration:** the ability to combine and analyse (*quality*) data from myriad sources



Data Integration: What data is there to integ

- Gene Relevant Annotation
 - Names & Aliases
 - Putative function
 - Representative sequence
 - Genetic context
 - Tracking variants

We are reliant on public domain sources but enhance this with GSK relevant information and standards

Synteny
Synteny between species means not only that genes in the same order on the genome, this indicating common ancestry. 90.2% of the human genome and 93.3% of the mouse genome lie in conserved syntenic segments. It is possible to dismiss homology as arising from similar function, but it is not possible to dismiss synteny this way.

Phylogeny
X2 Human, X3 Human, X1 Mouse, X1 Rat, X1 Human. Paralogues, Orthologues. A sequence in another species that shares a direct common ancestor with the current sequence. Orthologues are typically the most similar genes between 2 species. For some line after a speciation event the relationship is easily inferred from homology, as the two genes will differ only in time. As evolutionary time passes, the orthology relationship becomes more tenuous and eventually is lost through gene duplication and divergence.

UCSC, gsk, GlaxoSmithKline

Data Integration: What other data is there to integra

- Literature
 - key stimulant of interest
- Expression
 - is the gene expressed in tissues relevant to the disease?
 - Genome scale: Microarray technology
 - Focused follow up: TaqMan, ISH, IHC
- Phenotypic data:
 - Genetics
 - do people with *flaws* in this gene get sick more often?
 - Polymorphisms (SNPs) – genetic association, linkage analysis
 - Mouse KO/Transgenics
 - Identifying pathways from genetics
 - RNAi *in-vitro/in-vivo*
 - Protein Structure analysis/modelling

Microarray, Expression - protein, Immunohistochemistry (IHC), Expression - KOs, RNAi, KO mice. *Abnormalities follow in the wake of an LDLR knockout mouse.*

MDP Informatics, gsk, GlaxoSmithKline

Hypothesis-Driven Research

Hypothesis₁ → Experiment₁ → Dataset₁ → Result₁

Hypothesis₂ → Experiment₂ → Dataset₂ → Result₂

Hypothesis₃ → Experiment₃ → Dataset₃ → Result₃

⋮

Hypothesis_n → Experiment_n → Dataset_n → Result_n

MDP Informatics, gsk, GlaxoSmithKline

Combining validation data for finding targets... integrating validation data

Data from different sources can be combined to be more powerful than one source.

Permits prioritisation of targets to progress based on genetic, genomic and tractability data

Represents a systematic and stringent approach

Candidate genes in genetics study (human)

Expression in disease tissue according to IHC (rat)

Microarray study comparing KO mouse to wildtype (mouse)

☆ Drug targets

MDP Informatics, gsk, GlaxoSmithKline

Pathway studies: Immunoprecipitation, Y2H, Literature

Experimental pathway expansion/confirmation

Rich source of similar studies

Schizophrenia associated SNPs → Genes associated with schizophrenia

Schizophrenia: High-throughput candidate gene studies and WGA

Phenotype studies: Mouse knockouts, In vivo / in-vitro RNAi

Transcriptional profiling of EDG2 KO mouse

Differential expression in *in-vivo* and *in-vitro* models

Data generated from microarray analysis of antipsychotic treated rats and from isolation reared rats

Differential expression in human samples

Data generated from microarray and proteomic analysis of disease vs normal post-mortem brain samples

Distribution in normal tissues

Tissue profiles built from microarray data
Tissue profiles from Taqman data
Tissue distribution from ISH / IHC studies
Imaging Studies / translational medicine

MDP Informatics, gsk, GlaxoSmithKline

Disease view (PSTUD)

Different lines of evidence to find a drug target for a specific disease:

Gene	Candidate study	Associated loci	Linked to terms in literature	Mechanism	Comments	Expression array	Expression - IHC	Pathways & Y2H	Expression - <i>in-vitro</i>	Expression - <i>in-vivo</i>	Diff expression - VP	Diff expression - IH	Diff expression - Lit	Mouse Transgenics	Expression - RNAi	Proteomics
Gene X
TG1434

This Approach permits merging of general information with disease-specific/relevant information at gene level

MDP Informatics, gsk, GlaxoSmithKline

What is required

- Standard sample management/handling
- Standard data generation methods
- Standard QC analysis methods
- Standard significance/reporting cut-offs
- Agreed common standard for cross-reference
 - Gene?
 - Protein?
 - Genomic position?
- Pragmatism!
 - Balance standardisation with the need for the bespoke, avoid needless digression from established protocols
 - ◆ Some detail is inevitably lost through standardisation, determine cost-benefit and tactical versus strategic
 - ◆ If reasonable, report results from non-standard analyses and annotate as such
- Where one size does not fit all...

Ontologies

- Build standard vocabularies to accommodate these different data
- Effective and standardised representation of biological knowledge to permit mining and effective cross-querying

the obo foundry

The Open Biological and Biomedical Ontologies (OBO) Foundry is a collaborative experiment to produce well-structured vocabularies for shared use across different biological and medical domains. The OBO Foundry introduces a new paradigm for biomedical ontology development by the establishment of gold standard reference ontologies for individual domains of inquiry.

These involved compiling a group of biological researchers and ontology developers who agree in advance to the adoption of a growing set of principles specifying best practice in ontology development. These principles are designed to foster interoperability of ontologies within the broader OBO framework, and also to ensure a gradual improvement of quality and formal rigor to ontologies, in ways designed to meet the increasing needs of data and information integration in the biomedical domain.

By joining the OBO Foundry, the authors of an ontology commit to its maintenance in light of scientific advance, and to soliciting community feedback for its improvement. They also give an assurance that they will work with other Foundry members to ensure that, for any particular domain, there is community convergence on a single reference ontology. Application ontologies developed for specific purposes can then be referred back to this common reference, which will be updated in light of scientific advance. In this way application ontologies, too, for example the application ontologies developed for purposes of managing clinical trial data, can take advantage of the Foundry methodology.

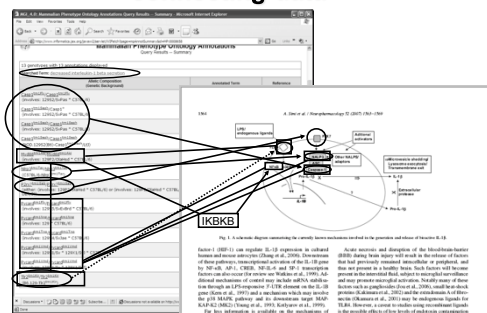
This site contains ontologies and points to some other efforts within the community. Ideally we see a range of ontologies being designed for biomedical domains. Some of these will be generic and apply across all organisms and others will be more restricted in scope, for example to specific taxonomic groups.

View the OBO Foundry ontologies in table form

Review the index of the OBO Foundry ontologies

- **Unify current data/knowledge more effectively**
- **Improve interpretation of data generated through new initiatives cited above**

Ontologies are powerful tools for connecting data



RED: Genes with an IL-1B secretion KO phenotype that relate to the inflammasome
GREEN: Genes with an IL-1B secretion KO phenotype that link directly to components of inflammasome

Pathway/Network Biology and Drug Discovery

- Pathway identification, extension, inference, modelling *can* help for:
 - Novel targets and biomarkers identification
 - Disease understanding
 - Mechanisms of action (MOA) understanding
 - Potential safety concerns
 - Combination therapies
 - Alternative indications
- Permits collapse of pertinent data around a series of genes linked by a common biological context
- Permits identification of common, known pathways represented in multiple platform datasets
 - Pathways are often more stably represented than their constituent genes
- Permits identification of "novel" disease mechanisms from datasets
- Essential for data driven polypharmacological approaches

Pathway/Network Analysis

- Do we have sufficient data within Neurosciences to power pathway/network based analysis approaches?
 - Advent of platform based approaches to studying neuroscience has dramatically increased data space over recent years
 - ◆ Are these data of sufficient quality?
 - ▶ Reproducibility/variability
 - ◆ Can we gain sufficient access to it?
 - ▶ Data storage and structure
 - ▶ Data silos/ Data access
 - ▶ Data reporting, sharing and distribution
 - ◆ Data annotation: can we find it?
 - ▶ Semantics of data
- We need to generate more data and improve reporting, description and storage of these data to maximise its benefit
 - Better industry, academic partnering
 - Support ontologies
 - Enforce journal and grant awarding body rules

What data is particularly relevant to Neuroscience?

- The 1990s: "Decade of the Brain"
 - Recognition of social and economic burden of brain disease
 - Increasing confidence amongst research community that brain disease is now a tractable problem
- Are we only now realising the benefits of this?
 - Broad Institute
 - Allen Brain Atlas
 - Genetic studies
 - Genomic studies
 - (Jackson Lab KO ontologies)
 - Neuroimaging and other translational medicine approaches
 - Better integration between these data / sites

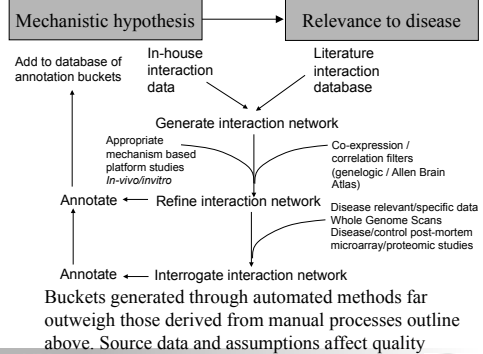
Pathway/Network Informatics

- Pathway Informatics (*and platform data*) require tools to analyze gene sets rather than individual genes, in the context of very large databases of **known** relationships and interactions:
 - Relationship data compiled into gene 'buckets' from multiple sources in multiple 'universes'
 - *involved in the same pathway, part of the same network, coexpressed, genetically linked, phylogenetically related, ...*
 - *In excess of 1m currently available*
 - Interaction data from different sources *binds, regulates, phosphorylates, degrades*
 - *Gene/protein/metabolite/compound*
- And determine if those observations are significant

MDP Informatics

gsk GlaxoSmithKline

Generating networks



MDP Informatics

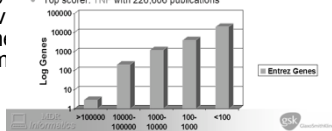
gsk GlaxoSmithKline

Literature

- Our molecular mechanistic understanding of CNS is improving
 - genes" returned (B4):

Using literature knowledge to identify CNS disease mechanisms will bias interpretation of your dataset

- 1967-1996: 96
- 1977-1986: 96
- Gene Ontology
- We can improve of molecular mechanisms elucidated from
- BUT...

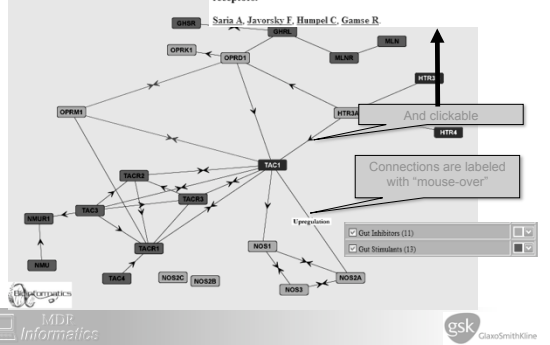


MDP Informatics

gsk GlaxoSmithKline

Building network from literature

Endogenous 5-hydroxytryptamine modulates the release of tachykinins and calcitonin gene-related peptide from the rat spinal cord via 5-HT3 receptors. Saria A, Javorovský F, Hampel C, Ganss R. Ann N Y Acad Sci. 1991;632:464-5.

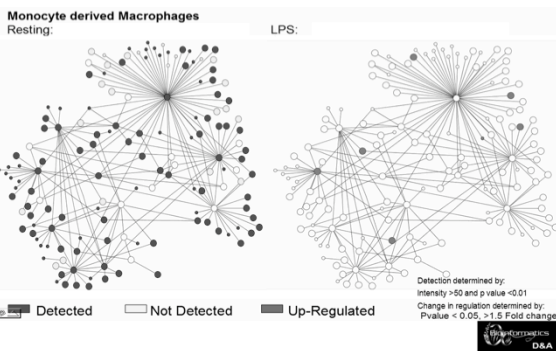


MDP Informatics

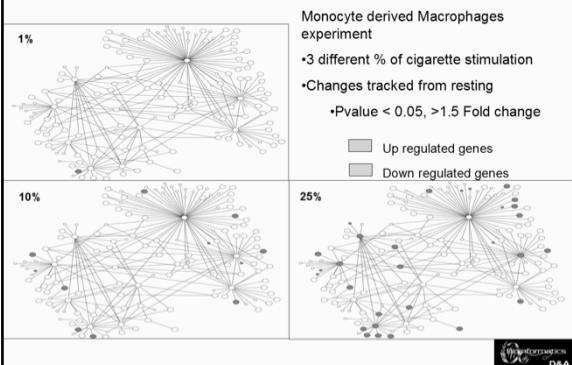
gsk GlaxoSmithKline

Identification of Pathways from Network

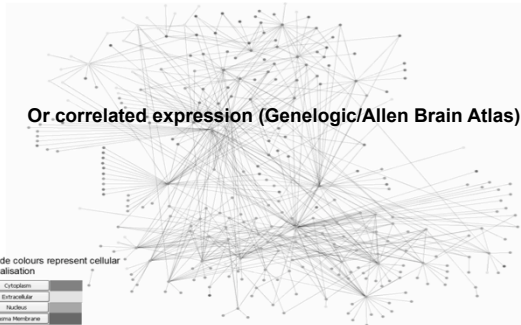
- Overlay specific cell type data to help with comparison of model systems/ cell lines



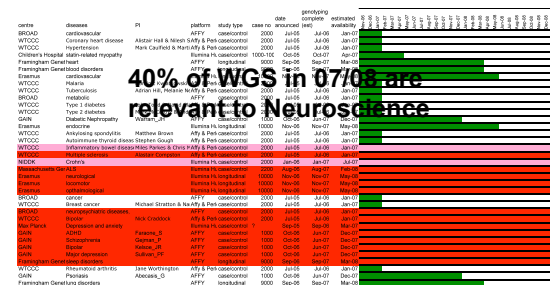
Dose Response Effect on Network



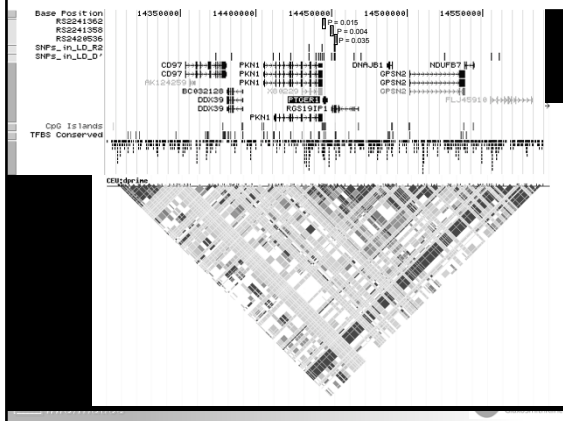
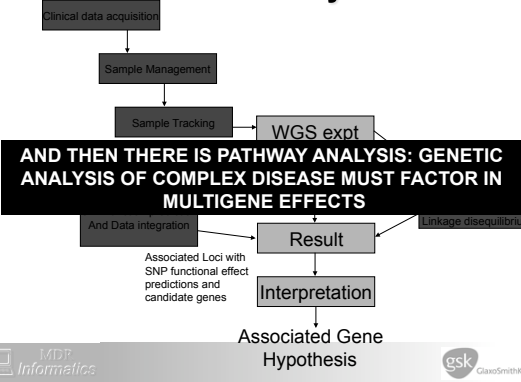
Compartmentalisation can help refine expanded Network



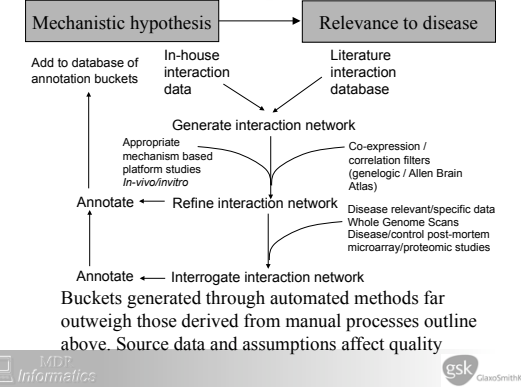
Whole Genome Scans



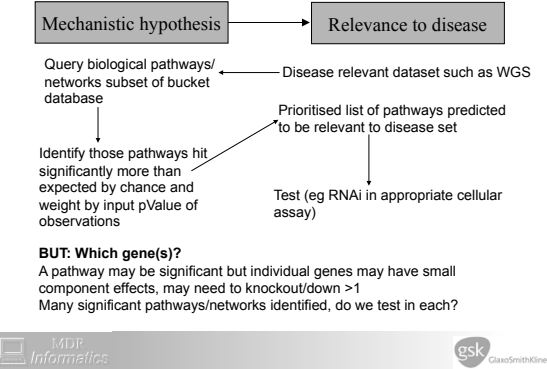
Genetic Analysis

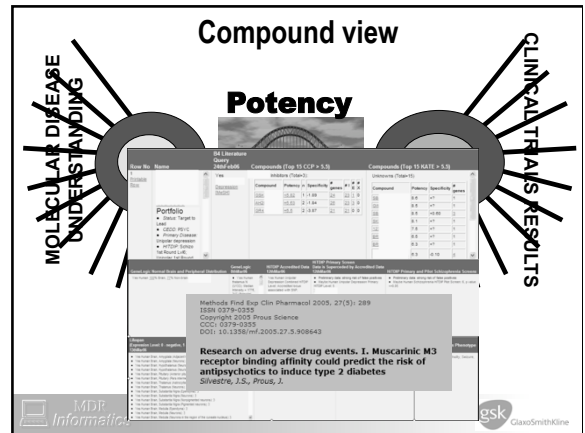
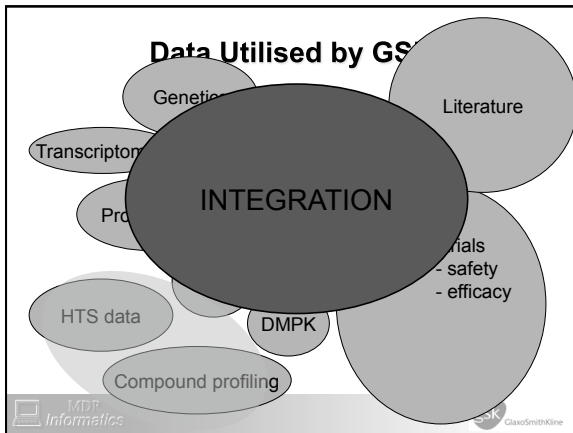


Generating networks

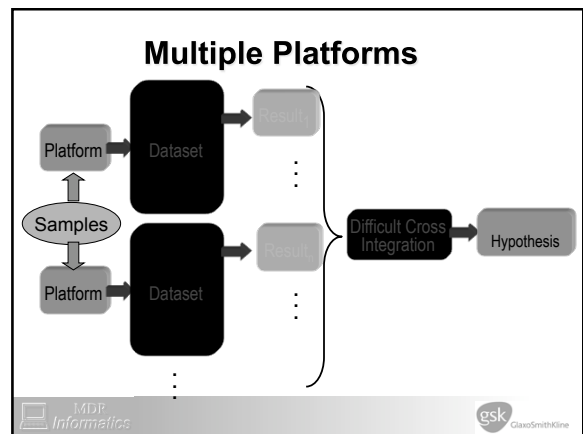


Identifying significant biological processes within complex disease relevant datasets



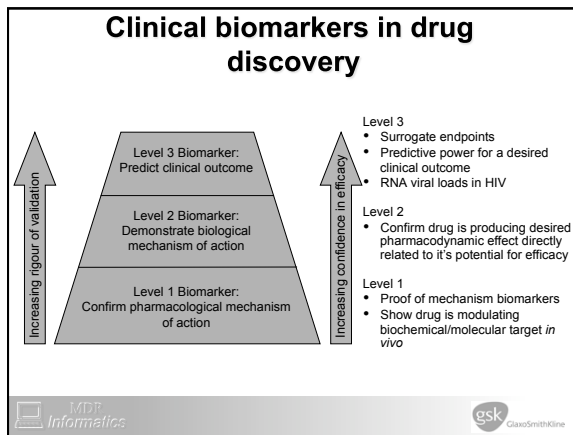


- ### Thanks to
- Past and present
 - CB / MDR Ix staff
 - MDR IT staff
- MDP Informatics | GSK GlaxoSmithKline



- BIOMARKERS / MULTIPLATFORM APPROACHES
- MDP Informatics | GSK GlaxoSmithKline

- ### Biomarkers
- "A characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes or pharmacological responses to a therapeutic intervention"
 - » NIH Biomarkers Definitions Working Group
 - Improve understanding of the therapeutic area and disease pathophysiology
 - Increase understanding of MOA of compound(s) and target both pre-clinically and clinically
 - Identify profiles characteristic of unwanted toxicity in early drug candidate screening
 - Provide evidence of drug efficacy and safety in early trials
 - Enhance experimental and clinical design
 - Provide means to make better clinical trial decisions earlier through use of surrogates
- MDP Informatics | GSK GlaxoSmithKline



- ### Identifying biomarker panels
- Pilot studies:
 - Depression
 - Alzheimers
 - Bloods from case/control populations
 - Clinical parameters
 - Genotype
 - Proteomics
 - Transcriptomics
 - Machine Learning approaches and training sets
 - “minimum panel of 9 analytes capable of segregating case/control with 78% accuracy”
- MDP Informatics gsk GlaxoSmithKline