

eScience Case Studies Using Taverna

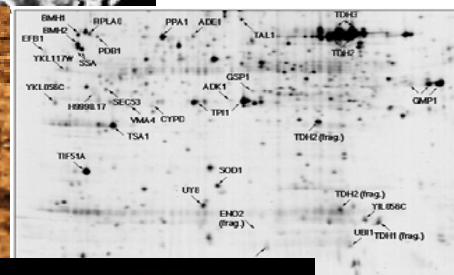
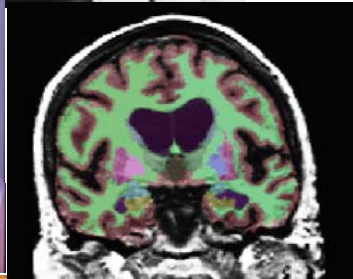
Dr. Georgina Moulton
The University of Manchester
(georgina.moulton@manchester.ac.uk)
(on behalf of the myGRID team)





Requirements

- Automation
- Reliability
- Repeatability
- Few programming skill required
- Works on distributed resources



Multi-disciplinary

- ~37000 downloads
- Ranked 210 on sourceforge
- Users in US, Singapore, UK, Europe, Australia,
- Systems biology
- Proteomics
- Gene/protein annotation
- Microarray data analysis
- Medical image analysis
- Heart simulations
- High throughput screening
- Phenotypical studies
- Plants, Mouse, Human
- Astronomy
- Aerospace
- Dilbert Cartoons

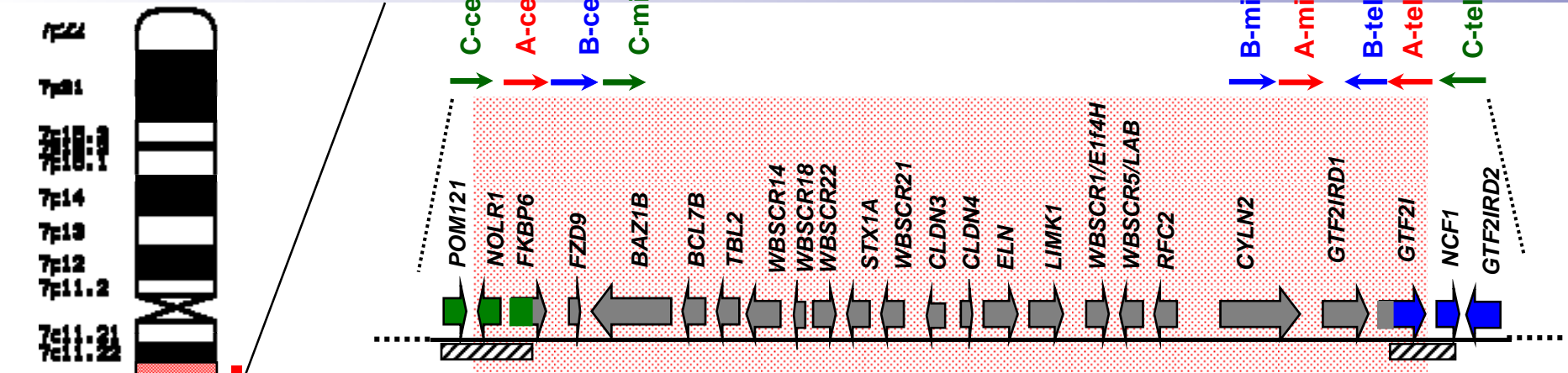
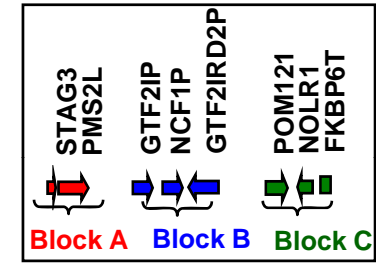
Williams-Beuren Syndrome (WBS)



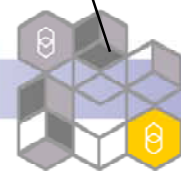
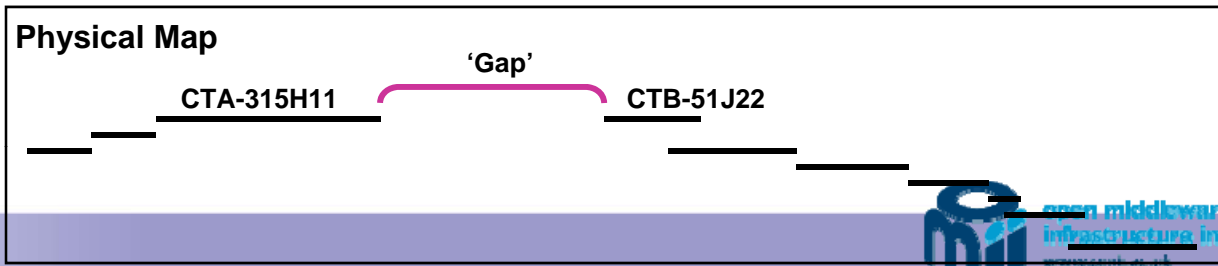
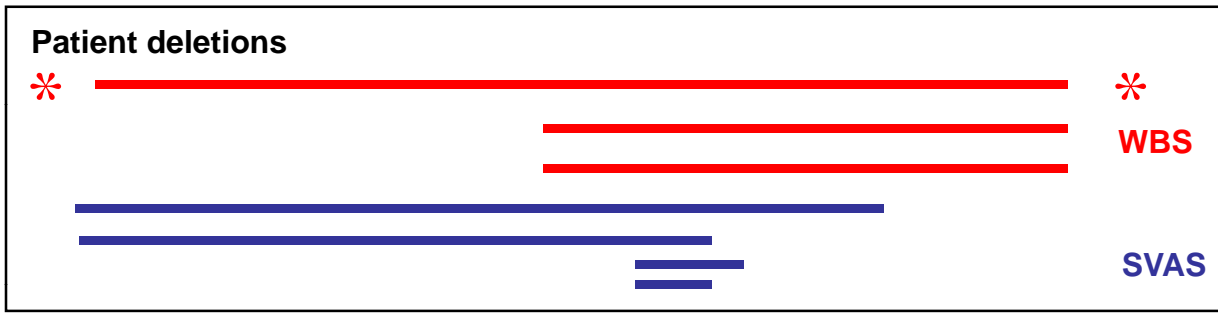
- Contiguous sporadic gene deletion disorder
- 1/20,000 live births, caused by unequal crossover (homologous recombination) during meiosis
- Haploinsufficiency of the region results in the phenotype
- Multisystem phenotype – muscular, nervous, circulatory systems
- Characteristic facial features
- Unique cognitive profile
- Mental retardation (IQ 40-100, mean~60, 'normal' mean ~ 100)
- Outgoing personality, friendly nature, 'charming'

Williams-Beuren Syndrome Microdeletion

Eicher E, Clark R & She, X An Assessment of the Sequence Gaps: Unfinished Business in a Finished Human Genome. *Nature Genetics Reviews* (2004) 5:345-354
 Hillier L et al. *The DNA Sequence of Human Chromosome 7. Nature* (2003) 424:157-164



7q11.23 ~1.5 Mb



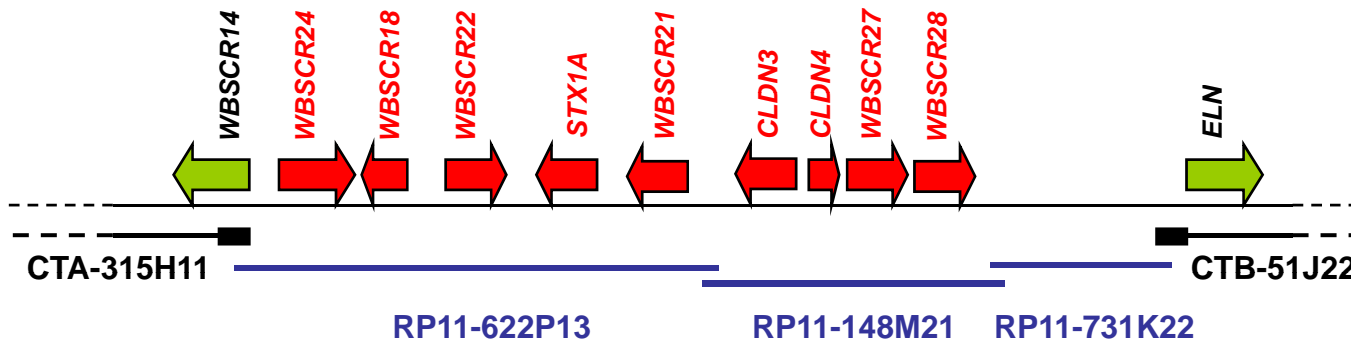
Filling a genomic gap *in silico*

- Two steps to filling the genomic gap:
 1. Identify new, overlapping sequence of interest
 2. Characterise the new sequence at nucleotide and amino acid level
- Number of issues if we are to do it the traditional way:
 1. Frequently repeated – info rapidly added to public databases
 2. Time consuming and mundane
 3. Don't always get results
 4. Huge amount of interrelated data is produced



The Biological Results

Four workflow cycles totalling ~ 10 hours
 The gap was correctly closed and all known features identified



314,004bp extension



All nine known genes identified
 (40/45 exons identified)

Case Study – Graves Disease

- Autoimmune disease that causes hyperthyroidism
- Antibodies to the thyrotropin receptor result in constitutive activation of the receptor and increased levels of thyroid hormone
- Original ^{my}Grid Case Study

Ref: Li P, Hayward K, Jennings C, Owen K, Oinn T, Stevens R, Pearce S and Wipat A (2004) Association of variations in NFKBIE with Graves? disease using classical and myGrid methodologies. UK e-Science All Hands Meeting 2004



Graves Disease

The experiment:

- Analysing microarray data to determine genes differentially-expressed in Graves Disease patients and healthy controls
- Characterising these genes (and any proteins encoded by them) in an annotation pipeline
- From affymetrix probeset identifier, extract information about genes encoded in this region.
- For each gene, evidence is extracted from other data sources to potentially support it as a candidate for disease involvement

Annotation Pipeline

Evidence includes:

- SNPs in coding and non-coding regions
- Protein products
- Protein structure and functional features
- Metabolic Pathways
- Gene Ontology terms

