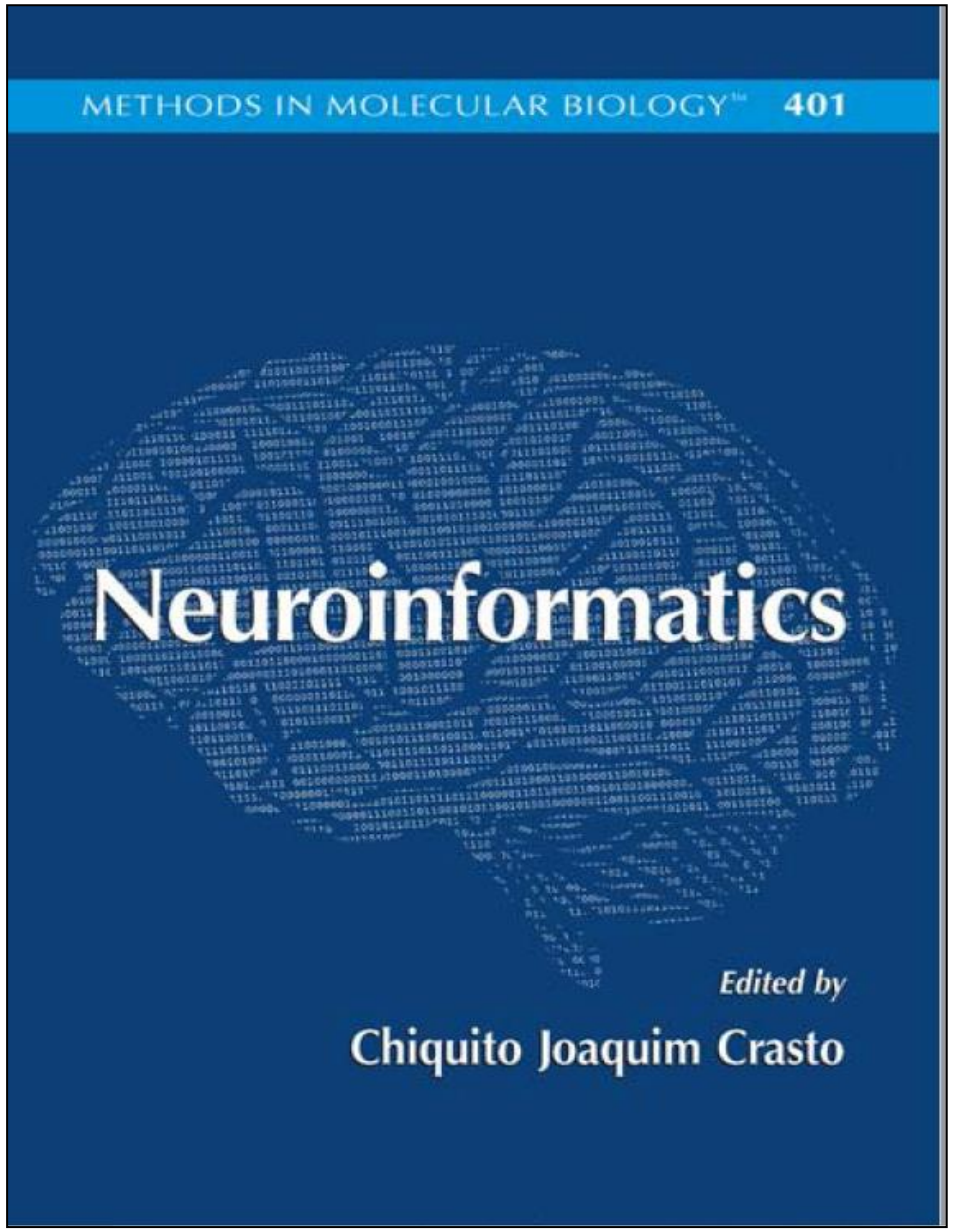
Integrating Genetic, Functional Genomic, and Bioinformatics Data:

Systems Biology Approach to Diseases.

Application to Schizophrenia



Neuroinformatics

Edited by

Chiquito Joaqium Crasto

Yale Center for Medical Informatics and Department of Neurobiology, Yale University School of Medicine, New Haven, CT

Current Address: Department of Genetics, University of Alabama at Birmingham, Birmingham, AL

Foreword by

Stephen H. Koslow

Biomedical Synergy
West Palm Beach, FL

HUMANA PRESS TOTOWA, NEW JERSEY

Introduction

Genetic Versus Complex Disease:

Simply put, a genetic disease is an unhealthy condition that is heritable. Whereas, if the occurrence of the disease requires a strong epigenetic or environmental trigger, then the disease is described as a “complex” disease.

Two clear-cut examples of simple and complex genetic diseases are *Huntington’s disease (HD)* and *schizophrenia*.

Approaches to Genetic Disease:

In any given search for disease-causing mutations, following are some key steps that need to be undertaken:

- **Choosing the Study Population**
- **Single Versus Multi-Family**
- **Case–Control**
- **Choosing the Variables**
- **Boosting Statistical Power**

Genetic Marker Analysis:

There are two primary methods of discovering causal mutations in a disease:

1. Direct re-sequencing of the DNA to discover known or novel mutations in affected individuals.
2. Performing genotyping of DNA to test whether subjects with a disease possess specific known DNA sequence alterations.

Define Recombination Events—Linkage Studies:

In whole genome linkage analysis, at least one generation of affected and unaffected individuals from single or multiple pedigrees are genotyped with a common set of markers.

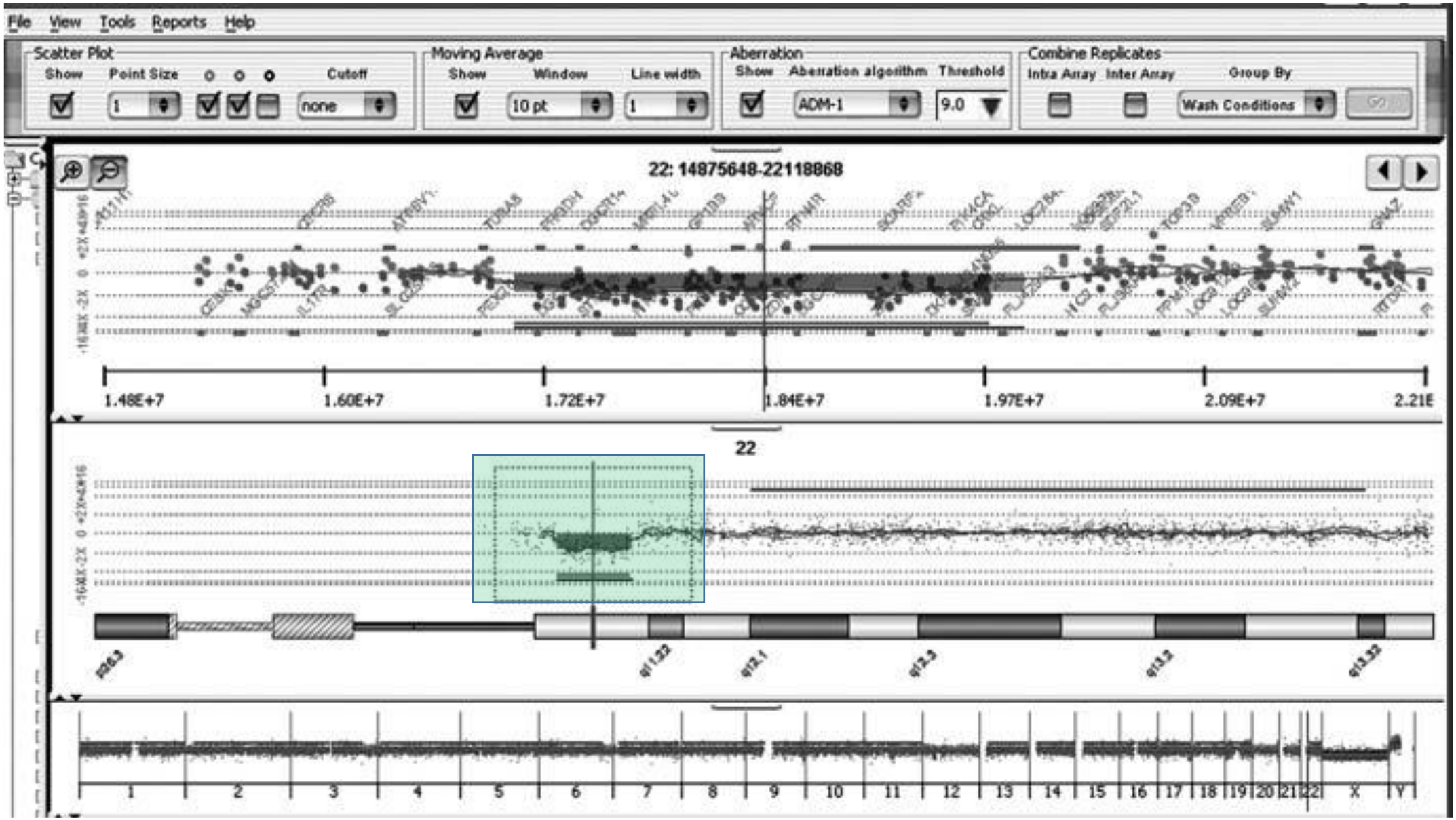
Define Cytogenetic/Chromosomal Abnormalities:

Changes in the copy number of genes or alleles have now been well established as a mechanism for causing neuropsychiatric disorders.

Chromosomal Abnormalities in Schizophrenia:

The most direct evidence for chromosomal abnormalities in schizophrenia comes from the study of **Velo-Cardio Facial Syndrome (VCFS).** VCFS is a condition caused by **a loss of one copy of a small piece (2–3 Mb)of chromosome 22** (*see* **Following image in the next slide**).

Chromosomal Abnormalities in Schizophrenia



Region of the velocardiofacial syndrome (VCFS) deletion on chromosome 22 detected by comparative genome hybridization (CGH).The deletion is clearly identified on the q-arm of chromosome 22, in the 17–20Mb region.

Functional Genomic Markers:

To help accelerate the discovery of disease-causing genes, it can often be helpful to directly screen for changes in the transcript level of different genes.

In brain disorders, the two most commonly used tissues are *postmortem brain tissue* and *peripheral blood leukocytes* (or transformed lymphoblasts).

Bioinformatics Data Sources:

With more than 30,000 full-length genes and tens of thousands of transcript variants, micro-RNAs, and noncanonical open reading frames (ORFs), the human genome annotation is undergoing constant development and refinement. Much of this information is now publicly accessible in major database interfaces, including the **Gene Ontology, InterPro, pFam, KEGG, UCSC, and NCBI sites.**

Example of Systems Biology Approach to Schizophrenia:

Portuguese Island Collection

Genetic Markers for Linkage and Family-Based Association:

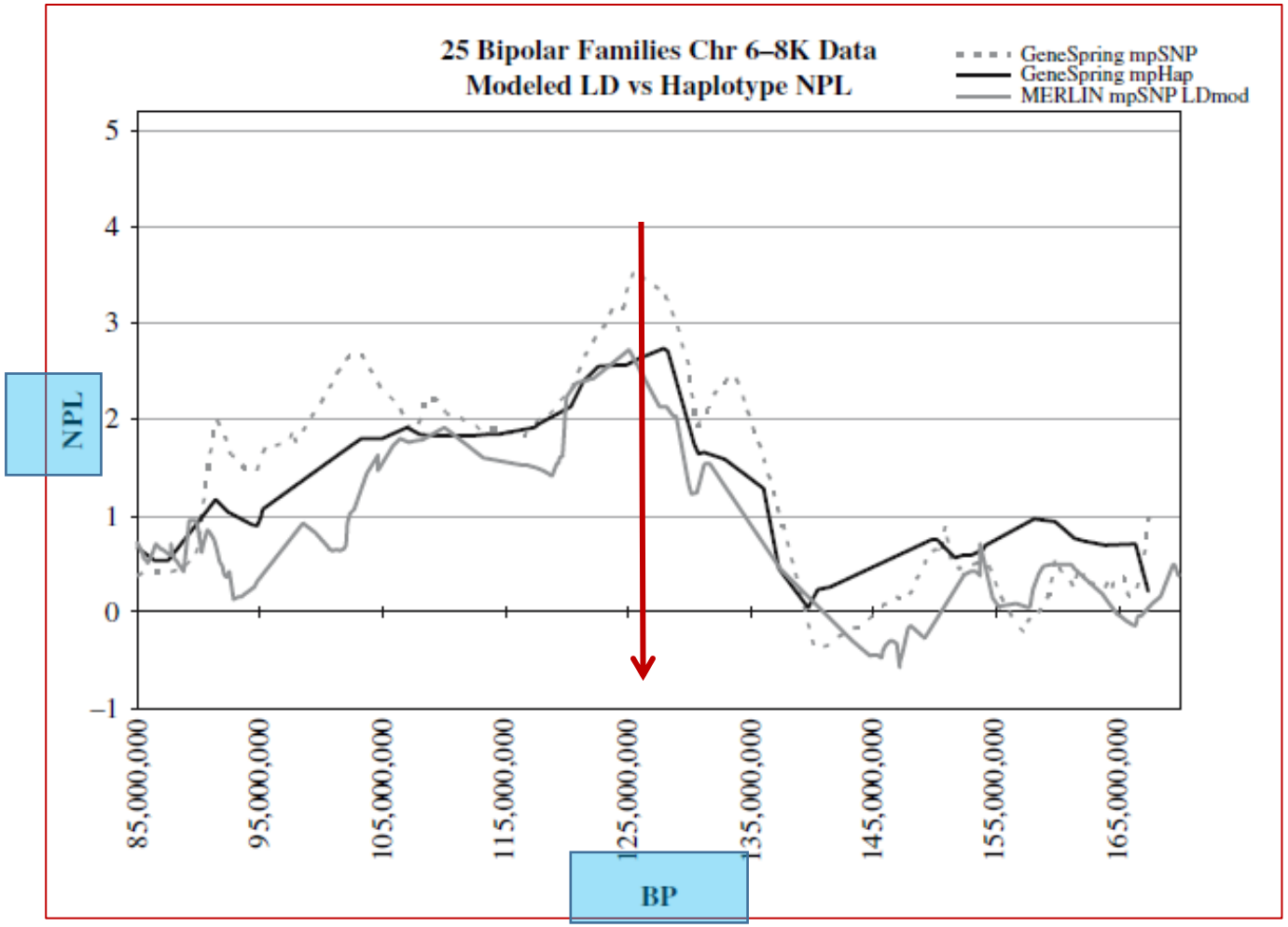
Single Nucleotide Polymorphisms(SNP).

Analysis of Dense SNP Data Sets: Single Nucleotide Polymorphism C → T

- Data obtained using genotypes from either the Affymetrix 10K or 50K HMAs for both family-based linkage and association.
- One free software platform **Multipoint Engine for Rapid Likelihood Inference (MERLIN)** and another commercial software program **GeneSpring GT (Agilent)** used for the study.

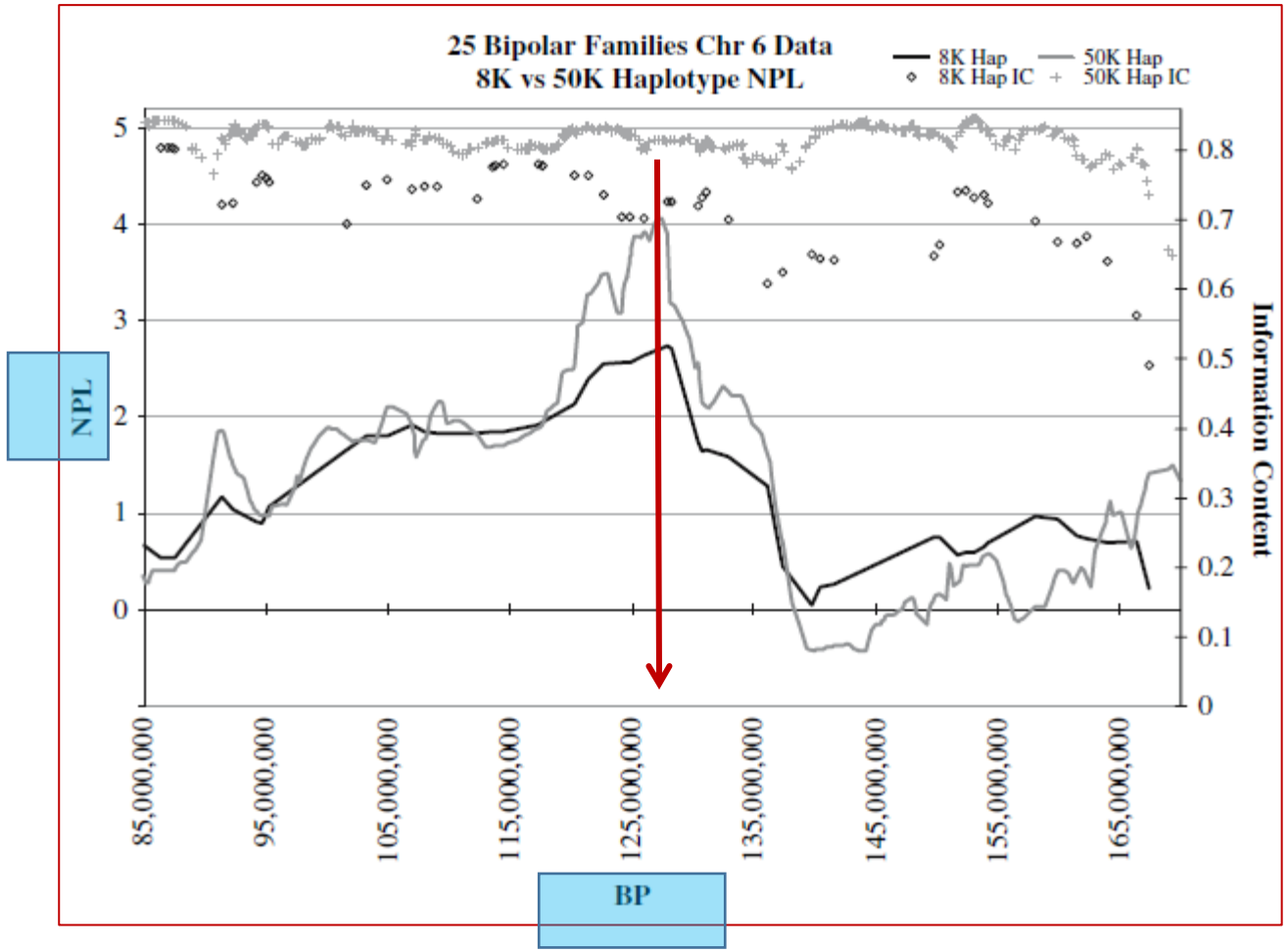
Box 1. Analysis Workflow Within GeneSpring GT
Data cleaning (error removal)
Haplotype map construction
Nonparametric haplotype-based linkage
Parametric haplotype-based linkage
Haplotype-based family-based association
Defining genes with major and minor effects

Nonparametric linkage results for Bipolar Disorder: Portuguese Island Collection



The results from multipoint SNP linkage with and without linkage disequilibrium (LD) modeling and multipoint haplotype-based linkage are shown for the region with the maximal score (chromosome 6q) in a study of 25 pedigrees segregating for bipolar disorder. These data are from a set of 7736 (8K) SNPs. The peaks obtained by LD modeling in MERLIN and haplotype-based analysis in GeneSpring are quite similar, while multipoint SNP results alone appear to inflate the linkage score.

Nonparametric linkage results for bipolar disorder: Portuguese Island Collection



The results of going from 8K to 50K SNP coverage in construction of the haplotype map clearly provides more information and boosts the linkage signal beyond 4.0 in the peak region, even with the intermarker LD correction provided by the use of haplotypes.

BP: base pair; **Hap**: haplotype; **IC**: information content; **mp**: multipoint; **NPL**: nonparametric linkage Z score.

Significant Family-Based Association Blocks in the Peak Linkage Region: Portuguese Island Collection

Chromosome	Base pair	Marker	D	D′	Rho	Chi square	*P* value
6	123901852	rs4307191	0.07	0.32	0.28	6.67	0.0098
6	123903977	rs10499125	0.07	0.32	0.28	6.67	0.0098
6	124984287	rs9375346	0.03	1.00	0.23	5.28	0.0216
6	124984549	rs9321003	0.03	1.00	0.23	5.28	0.0216
6	125278562	rs781748	0.04	0.38	0.29	7.04	0.0296
6	125292115	rs6910745	0.04	0.38	0.29	7.04	0.0296
6	125310084	rs4896606	0.04	0.38	0.29	7.04	0.0296
6	125364046	rs10485345	0.04	0.38	0.29	7.04	0.0296
6	129744486	rs6899448	0.04	0.31	0.26	6.34	0.0419
6	129754042	rs4599678	0.04	0.31	0.26	6.34	0.0419
6	130143572	rs10484280	0.08	0.81	0.88	25	0.0016
6	130144652	rs9321180	0.08	0.81	0.88	25	0.0016
6	130161191	rs2326869	0.08	0.81	0.88	25	0.0016
6	130236569	rs10499165	0.08	0.81	0.88	25	0.0016
6	130284551	rs7753901	0.08	0.81	0.88	25	0.0016

In the set of 25 bipolar families and an additional 15 bipolar families the haplotype-based haplotype relative risk (HHRR) algorithm was employed to test for overtransmission of specific risk haplotypes using family-based association in GeneSpring GT. The results from this analysis are shown only for the peak chromosome linkage region on chromosome 6.

Functional Group Overlap in Expressed Genes from Brain and Blood: Portuguese Island Collection

Functional gene group	Present in blood	Present in brain	Overlap in expression
Acetylcholine signaling	4	5	3/6 (50%)
Enzyme-Linked signaling	131	153	101/183 (55%)
Serotonin signaling	5	9	5/9(55%)
Integrin signaling	57	46	35/68 (51%)
GABA signaling	17	31	16/32 (50%)
Glutamate signaling	25	46	24/47 (51%)
Dopamine receptors	3	4	3/4 (75%)
Ligand binding/carriers	126	168	105/189 (56%)
Beta amyloid	27	39	23/43 (53%)
Synucleins	4	6	4/6 (75%)
Ribosomal	85	89	83/97 (86%)
Caspases	19	11	8/20 (40%)
Vacuolar	35	36	32/38 (84%)
Proteasome	53	56	52/57 (91%)
Oxidative phosphorylation	11	12	11/12 (92%)
Unfolded protein response	38	40	36/42 (86%)
Transcription factors	413	434	326/521 (63%)
Ubiquitin	111	113	97/127 (76%)
All mitochondria-related	298	368	273/393 (69%)

At the functional level, every neurotransmitter signaling group those were examined was also found to express many genes in common in the blood and the brain, and most groups even expressed a few genes in the blood that were not detectable in the brain.

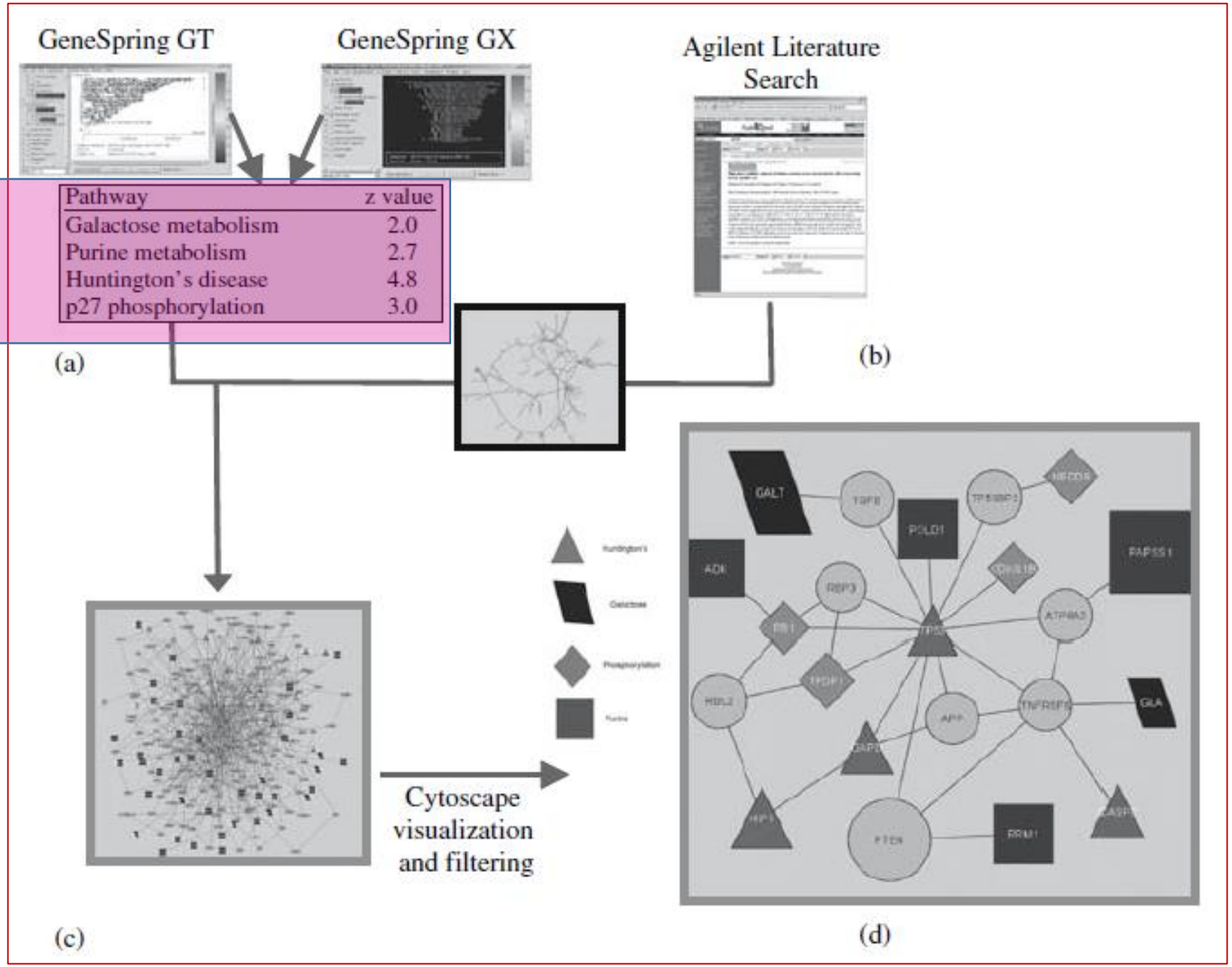
Bringing it All Together: Extended Comparisons of Expression Profiling, Comparative Genome Hybridization, and Genotyping

- In an extended analysis, a nonparametric whole genome linkage scan of schizophrenia pedigrees using **GeneSpring GT** identified 322 genes located within 2Mb of SNPs that showed evidence of linkage in a set of schizophrenia pedigrees analyzed at 10K density.
- Expression analysis using GeneSpring GX identified 2545 differentially expressed genes between the unaffected individuals and the affected siblings.

Box 2. Analysis Workflow Within GeneSpring GX

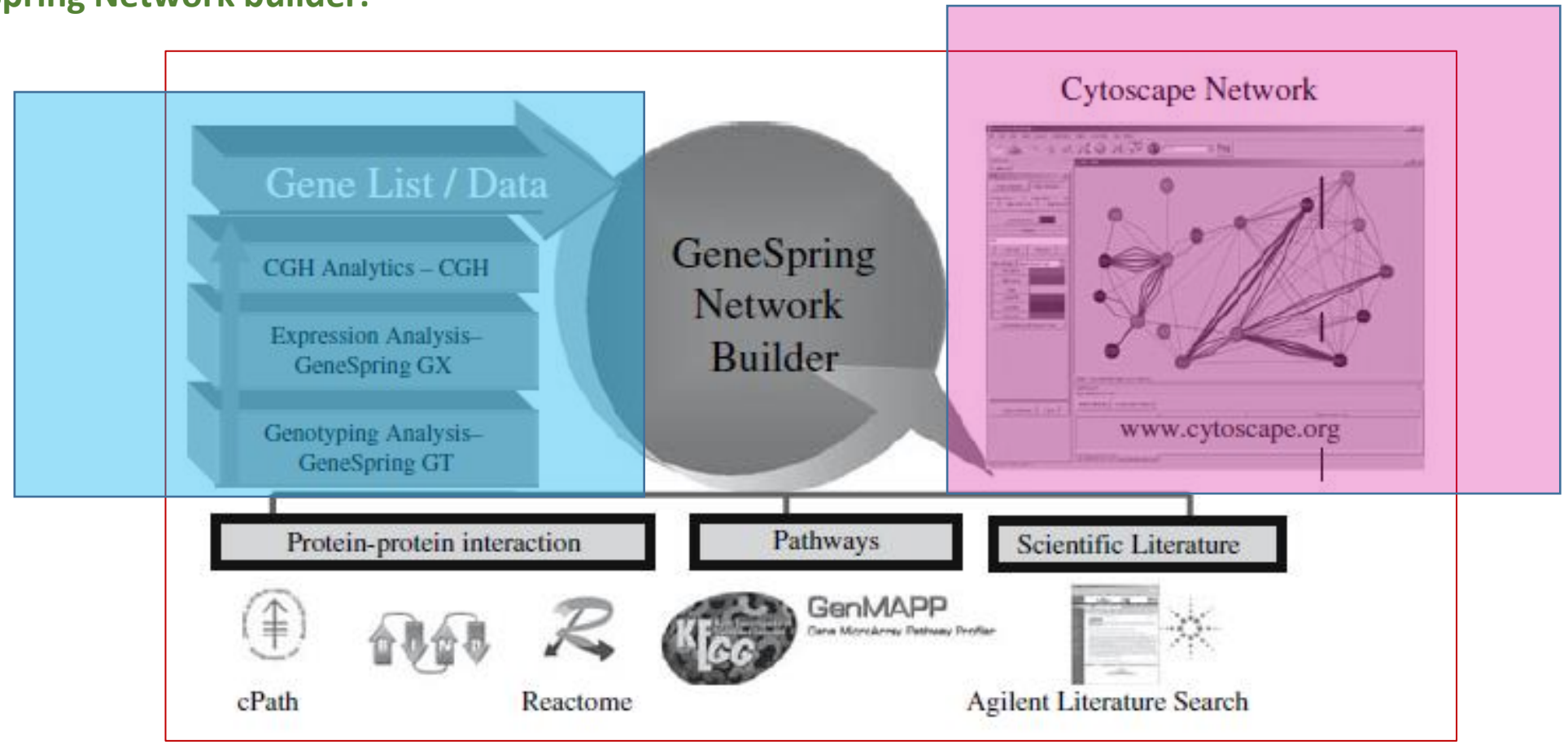
Data cleaning (removing hypervariable and unexpressed genes)
Pairwise or unpaired parametric testing
Correction for multiple comparisons
Class prediction analysis
Gene group analysis (biological and mathematical clusters)
Defining genes and gene groups with major and minor effects

Bringing it All Together: Network Building



Bringing it All Together: Comparative Genome Hybridization

GeneSpring Network builder:



Schematic representation of the workflow for the plug-in. A gene list is selected and the selected databases are searched for gene–gene interactions. A network is built using the Cytoscape viewer.

Bringing it All Together: Comparative Genome Hybridization

New network based on database info. for known associations between genes from the Huntington and ERK5 pathways: