Gene Logic’s Expression Databases:

BioExpress® System and ASCENDA® System
But first...genomics 101

- 6 billion bases of DNA in every nucleated human cell
- The ‘book of life’
- Quaternary code of A, G, C and T nucleotide bases in two complimentary strands wound in a double helix
- A.....T
- G.....C
- ‘Space junk’ or ‘dark genome’?
- Approximately 20,000 human genes

Cost ~$3bn to complete in 2001. Today $100k and coming down
The ‘$1,000 dollar genome’
Genes can be expressed at different levels in the course of normal biology, early development and in disease, and therefore make varying amounts of their protein.

We can monitor that by measuring the amount of mRNA specific to each gene in a cell or tissue.
High throughput large scale mRNA measurement

Affymetrix Gene Chip Array

GeneChip Probe Array

Hybridized Probe Cell

- Single stranded, labeled RNA target
- Oligonucleotide probe

Image of Hybridized Probe Array

- Millions of copies of a specific oligonucleotide probe
- >400,000 different complementary probes
• **Program Development**
  - Over $150 Million invested over a 7 year period
  - Assembled via a tissue accrual network of over 90 leading clinical and academic centers
  - BioExpress database contains gene expression profiles from over 22,000 samples of:
    - Diseased human tissue
    - Corresponding normal human tissue
    - Tissues from relevant animal models
  - Content maps to the following key Therapeutic Areas:
    - Normal Tissue
    - CNS Disorders, CVS Disorders
    - Inflammation, Metabolic Disease, Oncology
BioExpress System | Biorepository

- ~9 years in the making
- Tissue source: an international network of clinical and academic centers
- Collection under IRB approval and Gene Logic protocols
- Tissue Repository Information Management System (TRIMS) captures extensive clinical and demographic data
- Rigorous QC of collection procedures
- In-house board-certified pathologist review
- Samples run on Affymetrix GeneChip
- Residual sample material stored = resource for additional studies (PCR, SNP, miRNA, aCGH, etc.)
• **Demographics**
  - Age
  - Gender
  - Race / Ethnicity

• **Health Risk Factors**
  - Height / weight / BMI
  - Allergies / exposures
  - Diet / supplements
  - Smoking history
  - Alcohol use
  - Recreational drug use

• **OB/GYN History**
  - Menopausal status
  - Pregnancy history
  - Exogenous hormone use

• **Medical History**
  - Primary disease
  - Concurrent disease(s)
  - Prior history

• **Treatment History**
  - Current medications
  - Pertinent prior medications
  - Surgical procedure(s)
  - Anesthetics / perioperative agents
  - Radiation therapy
  - Other pertinent prior treatment

• **Family History**
  - Relative, disease, age of diagnosis

• **Diagnostic Tests**
  - Pre-operative labwork
  - Disease-specific studies
BioExpress System | Introduction

**Human Tissue Samples**
( Clinical Network)

- **Sample Annotations**
  - Clinical and Pathology data

- **Molecular Data**
  - Affymetrix GeneChip® gene expression

---

**BioExpress System Reference Database**
(>22,000 samples)

---

**Selected Diseased and Normal Human Samples for Bioinformatics Analysis**
BioExpress System | Disease Coverage

Sample Collection
> 22,000 Samples

ONCOLOGY
4131 Samples
Breast
Colon
Endometrium
Kidney
Leukemia
Lung
Lymphoma
Ovary
Pancreas
Prostate
Stomach
Other

INFLAMMATION
3584 Samples
Arthritis
Asthma
Cell Subsets
Cholecystitis
Colitis
Copd
Crohn’s
Dermatitis
Esophagitis
Lupus
Lymphoid
MS
Osteoarthritis
Psoriasis
Thyroiditis
Wegener

CNS
6491 Samples
Alzheimer’s
Parkinson’s
Anxiety
Depression
BiPolar
Schizophrenia
Cocaine

OTHER
800 Samples
Various

NORMAL
4214 Samples
80 Tissue Types
50+ Brain Regions

CVS
1839 Samples
Aneurism
Artherosclerosis
Arrhythmia
Cardiomyopathy
Carotid Endarectomy
Cell Subtypes
Congenital
Coronary Disease
Myocarditis
Peripheral Vascular
Transplants
Valvular Disease

METABOLIC
1078 Samples
Diabetes (Type I)
Diabetes (Type II)
Metabolic Syndrome
Obesity

BioExpress System and ASCENTA System
Biosolutions Inc. All rights reserved.
BioExpress System | 3 Individual Databases

- Gene Expression Data (RTE)
- Sample Annotations (Sample DB)
- Gene Annotations (Gene Index)
GX Explorer: A highly-integrated suite of analysis tools with which to simultaneously mine information from all three databases.

Sample & Gene Query Tools
Absolute Analysis
Comparative Analysis
Contrast Analysis
Correlation maps, PCA, scatter plots
Pathway and GO mapping tools
Visualization of clinical data
Export to other analysis tools

Sample DB

Gene Index

RTE
BioExpress | How is the data used?

• **Target Discovery:**
  - Identify novel gene targets that might be amenable to therapeutic intervention

• **Target Validation:**
  - Confirm that lead targets are expressed with sufficient specificity and to a sufficient degree to make them amenable to therapeutic intervention

• **Target Distribution**
  - Determine whether genes identified as potential targets in one disease/pathology show expression that would make them useful in other indications
  - Do target genes show an appropriate pattern of expression in animal models and cell lines that could then be used to test therapeutic utility
  - Selection of pathologies most likely to benefit from therapeutic intervention in clinical trials

• **Mechanistic studies**
  - What are the fundamental mechanisms underlying a given disease?
  - Pathway analysis can lead to the identification of additional “secondary” targets
Users of the Genesis Enterprise System have the benefits of:

- Access to all clinical details
- Complete freedom to assemble whatever sample sets meet their needs and interests
- Unsupervised workflows
  - Application of numerous user-defined filters
  - User-configurable data visualizations
- Access to individual expression values
- Ability to associate gene expression data with clinical data
- Ability to define gene sets and limit analyses to them
- Ability to simultaneously examine differential expression within multiple pair-wise comparisons
- Shared workspace
Challenges of using the Genesis Enterprise System include:

- Complexity of samples being coded by multiple SNOMED-based fields
- Unsupervised analysis - there are no “single-click” reports
- No library of pre-configured sample sets
- The number of user definable parameters can be “overwhelming”

The most productive use of the BioExpress database is therefore in the hands of power users who can dedicate significant time and effort to mining results.

So what about bench level researchers/occasional users with genes of interest?
• Grants users web-based access to Affymetrix GeneChip expression data residing in our BioExpress database

• Approximately 10,000 individual samples represented in total
  - Human, mouse and rat origin

• Human samples: comprehensive coverage of normal tissues alongside of extensive coverage of diseased tissues mapped to our 5 key therapeutic

• Rat and mouse samples: normal tissue from several strains, plus data from models of human disease

• System reports summary level expression data via a highly-intuitive, data-rich UI

• “Automated” reports allow data visualization with minimal user input

• Based entirely upon curated data
• All samples included are pre-assigned to sample sets
  - Pathologist-driven process
  - Dependent upon knowledge of factors that might influence a given pathology
  - Involves careful review of numerous clinical & demographic factors
    • Specific pathologies
    • Existence of secondary conditions
    • Lifestyle factors

• Primary curation factors
  - Sample site (topography)
  - Pathology
• In addition, our pathologists expertise allows the definition of more granular sub-sets
  - Sub-curation of IBD sets with respect to smoking history
  - Sub-curation of breast cancers with respect to the expression of certain hormone receptors
  - Sub-curation of Alzheimer’s content with respect to CERAD scoring
  - Etc.

• As subjective and labor-intensive as the curation process is, it enables end-users to make more detailed analyses of our data without them needing SME

• Pathologist-driven curation is then reinforced by statistical analysis to determine the consistency of gene expression within each sample set
• Use of a limited controlled vocabulary drives the categorization of sample sets

• Each pathologist-defined sample set is assigned to:
  - A tissue or organ
  - An organ system - a functional unit of the anatomy representing multiple organs
  - A disease - either a specific disease, or more often a term covering multiple related diseases
  - A therapeutic area

• Sample set nomenclature then captures important specifics
  - Actual disease state, specific pathology, specific tissue represented

• Net result:
  - Content that can be browsed simply by the selection of the desired organ/tissue or therapeutic area/disease terms
  - Results, i.e. expression patterns, that the system can sort and categorize by tissue and disease
From a gene-centric perspective, there are two absolutely fundamental questions that researchers want quick answers to:

- Where is a gene of interest expressed?
- Where/when is a gene of interest differentially expressed?

Pre-configured comparison pairs provide ASCENTA System users with a near instant answer to the latter question.

With a library of sample sets available for analysis, a corresponding library of pair-wise sample set comparisons can be created:

- NOT a comprehensive matrix of each sample set compared to all others
- INSTEAD, a library of “biologically relevant” comparisons
  - Majority = disease sets versus topography-matched normal sets
  - Also, more advanced/severe disease versus less advanced/sever disease

Comparisons are themselves then organized via assignment to disease- and organ system-based panels.

• Genes required for the proliferation and/or viability of three cancer cell lines were identified using a library of shRNAs targeting (i.e. preventing the expression of) several thousand “known” genes linked in some way to cellular proliferation

• Strategy identified 269 well characterized genes required for the proliferation of at least one of the three cell lines used

• Using the ASCENTA System, it’s possible to rapidly screen potential targets, and identify which ones might have the most therapeutic potential.....

• https://genomics.genelogic.com/ascenta
• Why is the database useful?
• What do researchers gain from the database?
• A clearer understanding of what is going wrong at the genomic level in disease
• Aberrations in gene expression generally lead to nothing good
• Potential new gene targets for drug therapy intervention
• Potential ‘biomarkers’..............
• For your genotype
• For your disease
• For your drug treatment
• For your early diagnosis

• **Personalized medicine......Pharmaceuticals into Pharmasuitables**
BioExpress System | Controlled Production Process

Nucleic Acid Extraction
- Quality: 28s/18s
- Integrity: RIN
- Purity: A260/A280

Sample Receipt
- Shipment QC
- Certified pathologist review
- Clinical data review and entry

Accession

Sample Release

Microarray Processing
- Quality: Agilent Bioanalyzer
- Consistency Checks
- Hybridization, fluidics LIMS controls

Product Delivery

Data Report and Storage
- QC Workbench
- Consistency checks

Fragment Target

Microarray Scan

Quality Control Points

BD Express System and ASCENTA System
Biosolutions Inc. All rights reserved.
• **Sample accrual via Gene Logic-trained clinical and academic centers**
  – Consistent sample collection protocols
  – Extensive clinical annotation
  – More controlled sample quality

• **Sample QC by pathologist and sample annotation QC by registered nurse**

• **Samples processed in single production Affymetrix facility applying 30 QC metrics**
  – Single platform - Affymetrix
  – Single industry-standard facility

• **Data produced at single site on single platform**
  – Reduced variability from array to array
  – Greater confidence in results

• **Extensive annotation allows the assembly of sample sets containing closely-matched samples**
  – Ask more defined questions
  – Greater confidence in results