

Getting More from Human Gene Expression Data

OmniViz Visualization Software Reveals New Gene Expression Patterns that can Influence Drug Development

Nicola C Day, PhD, Healthcare Consultant, BioWisdom
Sandra Williams, Department Head, Genomics, Asterand plc

The sequencing of the human genome has identified a vast number of potentially interesting targets for drug development. The challenge is to evaluate these targets and select those that are relevant to human disease. An important step for the evaluation process is to identify where these targets are expressed, and to link this expression to potential therapeutic uses. This evaluation is greatly aided by the provision of a) high quality gene expression data and b) sophisticated software tools for identifying targets with interesting expression patterns.

In this article, we describe the combination of Asterand's proprietary XpressWay® human gene expression data with BioWisdom's powerful visual analytics software, OmniViz. The XpressWay® dataset consists of more than 2000 gene expression profiles, the majority representing proteins that are potentially tractable drug targets. The OmniViz software provides easily interpreted visualizations and a suite of interactive tools that allows scientists to relate divergent data to uncover previously unknown associations and answer key scientific questions.

By analyzing XpressWay® data using OmniViz, the user can quickly select and stratify genes of interest, providing benefit in the following ways:

- Increasing confidence in a therapeutic approach
- Selecting the best targets for drug development
- Identifying new targets for an indication
- Exploiting opportunities to switch indications
- Identifying potential side effect liabilities
- Predicting translation of preclinical to clinical drug effects
- Gaining mechanistic insight by looking at the association of gene expression profiles

A 'whole body' scan of gene expression – revealing real differences between tissues

XpressWay® profiles consist of target expression in 72 different human tissues (see table 1) from 3 different donors (216 samples in total). These tissues are chosen to be representative of the major organ systems of the body, and comprise many sub-dissected regions. All the tissues are assessed as pathologically normal by pathologists. The breadth of tissues used for XpressWay® profiles holds advantages over other gene expression datasets which often consist of smaller sets of tissue.

Table 1. Tissues used in XpressWay® gene expression profiles

<i>heart : left atrium</i>	<i>liver : parenchyma</i>	<i>brain : dorsal raphe nucleus</i>	<i>bladder : trigone</i>
<i>heart : left ventricle</i>	<i>brain : cerebellum</i>	<i>spinal cord : cervical</i>	<i>ovary</i>
<i>blood vessel : coronary</i>	<i>brain : hippocampus</i>	<i>dorsal root ganglion : cervical</i>	<i>fallopian tube</i>
<i>oesophagus</i>	<i>brain : locus coeruleus</i>	<i>pineal gland</i>	<i>uterus : myometrium</i>
<i>stomach : fundus</i>	<i>brain : medulla</i>	<i>pituitary gland</i>	<i>uterus : cervix</i>
<i>stomach : body : whole</i>	<i>brain : amygdala</i>	<i>blood vessel : choroid plexus</i>	<i>prostate</i>
<i>stomach : antrum</i>	<i>brain : caudate</i>	<i>blood vessel : cerebral</i>	<i>vas deferens</i>
<i>stomach : pyloric canal</i>	<i>brain : hypothalamus - anterior</i>	<i>trachea</i>	<i>testis</i>
<i>duodenum</i>	<i>brain : hypothalamus - posterior</i>	<i>lung : parenchyma</i>	<i>spleen : parenchyma</i>
<i>jejunum</i>	<i>brain : cortex : cingulate - anterior</i>	<i>lung : primary bronchus</i>	<i>blood : mononuclear cells</i>
<i>ileum</i>	<i>brain : cortex : cingulate - posterior</i>	<i>lung : tertiary bronchus</i>	<i>lymph gland : tonsil</i>
<i>adipose</i>	<i>brain : cortex : frontal - lateral</i>	<i>blood vessel : pulmonary</i>	<i>skeletal muscle</i>
<i>blood vessel : mesenteric</i>	<i>brain : cortex : frontal - medial</i>	<i>kidney : cortex</i>	<i>skin</i>
<i>caecum</i>	<i>brain : cortex : occipital</i>	<i>kidney : medulla</i>	<i>adrenal gland</i>
<i>colon</i>	<i>brain : cortex : parietal</i>	<i>kidney : pelvis</i>	<i>thyroid gland</i>
<i>rectum</i>	<i>brain : cortex : temporal</i>	<i>blood vessel : renal</i>	<i>umbilical cord</i>
<i>gallbladder</i>	<i>brain : nucleus accumbens</i>	<i>ureter</i>	<i>placenta</i>
<i>pancreas</i>	<i>brain : substantia nigra</i>	<i>bladder : body</i>	<i>breast</i>

Total RNA is isolated from the tissues using standard methodologies, and has to pass several QC criteria before being considered as suitable for expression profiling (Barber et al, 2005).

Sensitive and standardized methodology – ensuring confidence in the data

The XpressWay® gene expression profiles are generated using quantitative real time-PCR (qRT-PCR; Bowen et al, 2000), a method that allows detection of absolute mRNA levels in tissues. This technique is different from other methodologies, such as chip-based assays, which measure relative abundance of mRNA. The target gene is multiplexed in the same well as GAPDH, which is used to confirm successful amplification. PCR amplification curves are analyzed to yield threshold values, and these are used to determine the starting mRNA copy number of both target and GAPDH genes by interpolation from a global standard curve. Rigorous pass/fail criteria are applied in order to ensure the quality of the data.

Extensive gene annotation – increasing your understanding of the data

The XpressWay® gene expression data are provided with BioWisdom annotation around gene synonyms, disease and process information. The provision of gene synonyms (Table 2) ensures that the user can find the gene of interest, and also facilitates connections to other data sources. For example, it is possible to send the synonyms out to PubMed in order to retrieve ALL the relevant publications around the target of interest. With one click, these documents can be imported into OmniViz, and, by employing the extensive analytical tools available, the important themes of the documents, and hence the nature of the target, can be revealed.

The inclusion of disease and process information provides context to the gene expression data, and allows the user to search for interesting targets by virtue of their functional role. Table 3 shows the diseases associated with the serotonin transporter.

Table 2. Gene target synonyms and related terms provided as annotation for the serotonin

<i>SERT</i>	<i>5-hydroxytryptamine transporter</i>	<i>Solute carrier family 6, member 4</i>
<i>SERT1</i>	<i>5-HTT</i>	<i>OCD1</i>
<i>Serotonin transporter</i>	<i>5HTT</i>	<i>A1323329</i>
<i>5-HT transporter</i>	<i>HTT</i>	<i>MGC124313</i>
<i>5HT transporter</i>	<i>SLC6A4</i>	<i>MGC124314</i>

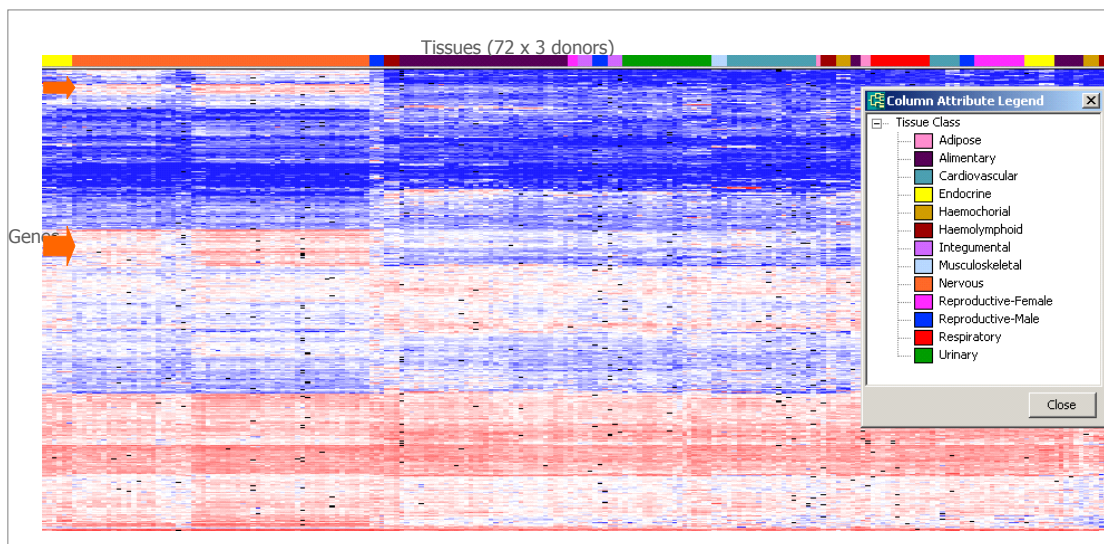
Table 3. Diseases provided as annotation for the serotonin transporter

<i>Autistic disorder</i>	<i>Suicide</i>
<i>Anxiety related disorder</i>	<i>Anxiety</i>
<i>Major depressive disorder</i>	<i>Bipolar disorder</i>
<i>Cardiovascular disorder</i>	<i>Schizophrenic disorder</i>
<i>Obsessive-compulsive disorder 1</i>	<i>Fibromyalgia</i>
<i>Mood disorder</i>	<i>Anxiety-related personality traits</i>

View the landscape of gene expression profiles

Using Omniviz to analyze XpressWay® data, it is possible to get an overview of the gene expression patterns of all ~2000 genes (Figure 1), and then to identify, very quickly, those gene expression profiles that possess features of interest. For example, if you are interested in developing a drug for a CNS disorder, you may be interested in the targets (see orange arrows on the left hand side) that have high expression (red) in nervous tissue and low expression (blue) in peripheral tissues.

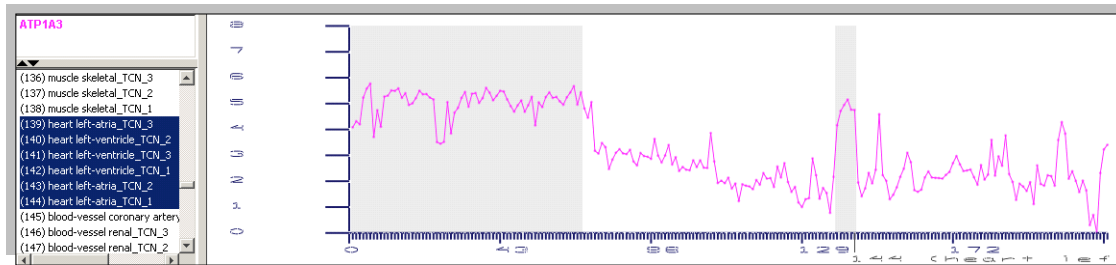
Figure 1. Omniviz Treescapes visualization providing an overview of all XpressWay® gene expression patterns (~2000 genes)



Find the expression profile for your drug target - reveal opportunities and threats

Individual gene expression profiles across the panel of 72 tissues can provide valuable information about potential drug action in those tissues. Within Omniviz, you can retrieve the expression profile for your drug target by a simple text query for the target name. The target expression profile in figure 2 shows high copy numbers in CNS and heart samples. Depending on your approach, this expression pattern may represent opportunities for the development of drugs for CNS or cardiac disorders, or may suggest potential side-effect liabilities in these areas. Knowledge of these data help you to make a rational decision on the future of a drug developed against this target.

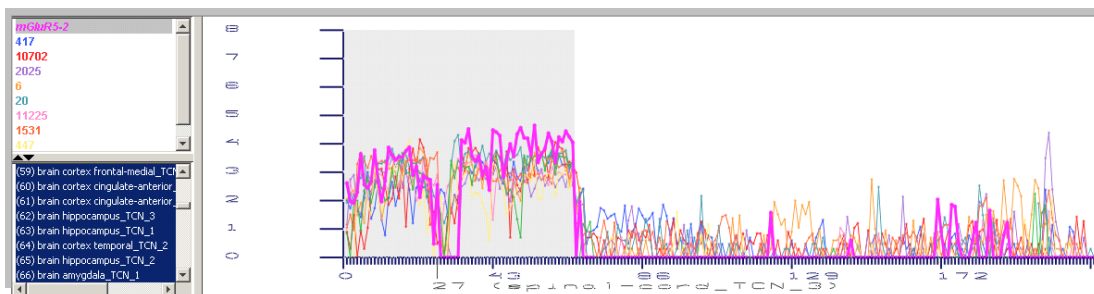
Figure 2. OmniViz Profile Plot showing expression of a single gene, ATP1A3, across the panel of tissues



Identify interesting targets using sophisticated queries

In addition to searching for genes by their name or by their functional role, OmniViz provides the capability to make complex numeric queries of the data. For example, you can identify targets with a particular expression profile, or profiles that are most similar to a chosen target. As an example, let's look for a new target for Alzheimer's disease (AD). It is known that the hippocampus atrophies in AD, and that it plays a role in processes that are affected in AD (e.g. memory). Within OmniViz, it is possible to set up a search for genes that have >10,000 mRNA copies in the hippocampus, but less than 100 mRNA copies in most of the other tissues. This particular query retrieves mGluR5-splice variant 2 (mGluR5-2; figure 3, bold pink line). As an additional step, it is possible to retrieve other gene profiles with the closest similarity to mGluR5-2, by correlation or Euclidean matching (see numbered targets in figure 3).

Figure 3. The expression profile of mGluR5-2 (bold pink line) and the 10 genes identified to have the most similar profiles



Furthermore, it is possible to assess the relevance of the retrieved targets by viewing diseases and processes associated with them. As can be seen from table 4, mGluR5-2 and its closest neighbors play a role in several processes (e.g. synaptic transmission, synaptic plasticity) that may be relevant to AD. Therefore, these may be interesting targets to progress for this disorder.

As demonstrated above, the ability to perform sophisticated queries of the data can prove useful in the identification of drug development opportunities.

Table 4. GO processes associated with mGlu5-2 and the 10 targets that have the most similar gene expression profiles

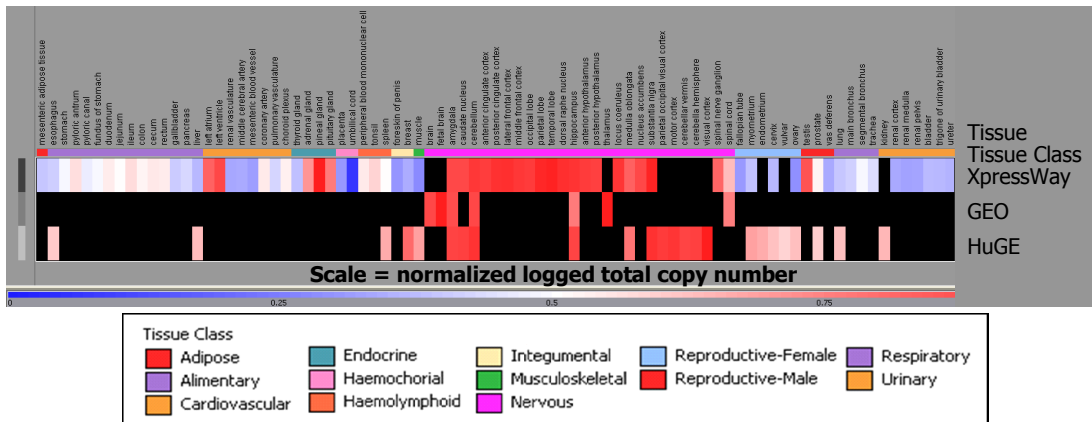
GO Process	No. Targets
synaptic transmission	6
signal transduction	5
G-protein coupled receptor protein signaling pathway	3
regulation of long-term neuronal synaptic plasticity	2
locomotory behavior	2
learning	2
ion transport	2
central nervous system development	2
sodium ion transport	1
regulation of synaptic transmission, glutamatergic	1
regulation of inhibitory postsynaptic membrane potential	1
regulation of excitatory postsynaptic membrane potential	1
positive regulation of synaptic transmission, GABAergic	1
positive regulation of gamma-aminobutyric acid secretion	1
peripheral nervous system development	1
neurotransmitter biosynthetic process	1
neuropeptide signaling pathway	1
negative regulation of synaptic transmission, glutamatergic	1

Compare XpressWay® data with other datasets

With OmniViz, you can easily import other datasets (internal or publicly available) to compare with XpressWay® gene expression data. These datasets can be linked to or merged with the XpressWay® dataset using common identifiers (e.g. gene ID).

As an example, figure 4 compares the expression profiles for the target, ATP1A3, in XpressWay®, GEO and HuGE datasets.

Figure 4. OmniViz Treescapes visualization comparing expression patterns of ATP1A3 across XpressWay®, GEO and HuGE datasets



Why XpressWay[®] data in OmniViz is a valuable asset for drug development

The XpressWay[®] dataset holds several advantages over other gene expression datasets. Firstly, the use of sensitive and standardized methodology, and the wide range of tissues used, allows you to have confidence in the presence (or not) of a target in a tissue, and the difference in expression between tissues. Secondly, the enhanced interrogation provided by OmniViz allows you to see previously unknown expression patterns, to find associations between targets, and to find targets with particular features of interest. Finally, your understanding of the data is augmented due to disease and process annotation, and the facility to bring related information (e.g. other gene expression data or PubMed articles on a target) into OmniViz for analysis.

For further information

To find out more about XpressWay[®] data in OmniViz, please contact sales@biowisdom.com and/or go to http://www.biowisdom.com/content/xpressway_gen where you will find a webinar on the subject.