Affymetrix Soybean Whole Transcript (WT) Array

Warning:

The Affymetrix Soybean Whole Transcript GeneChip array provided by the Goldberg lab is an ***antisense*** array that was designed specifically for our lab by Affymetrix. The array that you can order directly from Affymetrix is a ***sense*** array. The labeling protocol provided in this packet is different from the Sense Labeling kit available from Affymetrix. Please be aware that the labeled targets generated using the sense labeling kit ***will not work*** with this array and vice versa. Please follow the protocol provided in this packet to process the arrays provided by our lab.

Publication Acknowledgement:

The array was designed with collaboration from our lab (Goldberg Lab) and Affymetrix with advice and suggestions from other members of the soybean community, including Randy Shoemaker.

Please acknowledge the following people for the design of this array:

Goldberg Lab
Bob Goldberg, Brandon Le, Chen Cheng, Min Chen, and Anhthu Bui

Affymetrix
Gene Tanimoto, Christopher Davies, Stan Trask, Brant Wong, Eric Schell, Xue Mei Zhou, and Patricia Chan

Table of Contents

I. BACKGROUND

Motivation: We created this Soybean Whole Transcript (WT) Array to interrogate all the genes in the genome. The first generation Affymetrix Soybean Genome array was designed by the Soybean Consortium using publicly available soybean full-length cDNAs and ESTs. The Soybean Genome array consists of 37,000 probe sets interrogating ~ 25,000 distinct genes/transcripts. The release of the whole genome sequence of soybean[1] (available at Phytozome.net) allowed the creation of an array that can survey all the genes (both high and low confidence gene models)[1] in the genome.

Design: The design of the Soybean WT array is different from the Soybean Genome array. For the Soybean Genome array, probes were selected to correspond to the 3' end of the transcript or cDNA. However, for the Soybean WT array, probes were selected to span every exon of the predicted gene models/transcripts, if possible. This approach allows for the interrogation of the transcript (from 5' to 3') and can help determine exon usage in different splice variants that may be differentially expressed in specific tissues or compartments (**Fig. 1**). For information regarding this array design, please check out other references from Affymetrix[2].

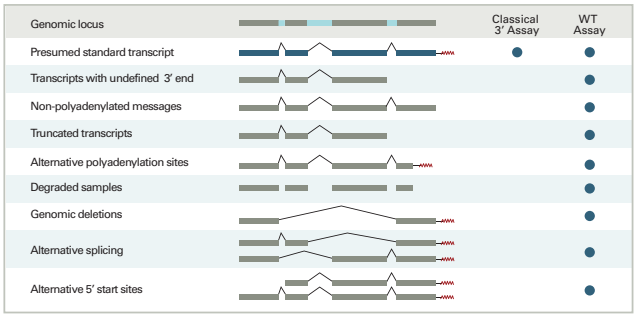

Figure 1. Types of transcripts captured by a whole-transcript array assay. Most of these transcripts cannot be distinguished with the classical 3' array assay. (Image taken from http://media.affymetrix.com:80/support/technical/appnotes/wt_appnote.pdf)

Probes were selected to interrogate one transcript only, although some probes might map to multiple transcripts (if no unique probes can be obtained for that exon region). The probe association file (see Table 1) does not indicate the number of transcripts interrogated by each probe.

Sequence Data: All sequence data used to design probes on the array were obtained from the Department of Energy - Joint Genome Institute (DOE-JGI) web site (phytozome: http://phytozome.net). Probes were designed from the first draft assembly of the soybean genome[1] (version 1.0). The probe selection algorithm was developed by Christopher Davies and Brant Wong at Affymetrix.

---

[1] Schmutz et al., Nature 463 pp. 178-83 (2010)

[2] http://media.affymetrix.com:80/support/technical/technotes/gene_1_0_st_technote.pdf

II. PROBE INFORMATION

Affymetrix provided several files for the design of the array (e.g. *.design file, *.mps file, and *.pgf file). To simplify downstream data analysis, we've created a text file that correlates Affymetrix probe ID with associated probe sequence, gene and exon information, etc. A sample table of the probe association file (**Table 1**) and an illustration of the probe assignment structure (**Fig. 2**) are shown below. The probe association file is available at our website[3].

Table 1. A snapshot of the probe association file showing the correlation of predicted gene models with probe and sequence information on the array.

Gene_Model[a]	Gene_ID[b]	Genomic _Start[c]	Genomic _End[c]	Exon_ID[d]	Probe_ID [e]	Probe_ Start[f]	Probe_ End[f]	GC[g]	Probe_Seq_On_Array[h]
Glyma01g00320	11764300	116300	127990	11764301	96462	116394	116418	11	GCAACATCACATATAGGACTTAGGG
Glyma01g00320	11764300	116300	127990	11764301	536743	116410	116434	12	GACTTAGGGCTAGCGTCTTTATCAC
Glyma01g00320	11764300	116300	127990	11764301	616296	116414	116438	11	TAGGGCTAGCGTCTTTATCACAATC

a. The predicted gene model obtained from the Phytozome web site (http://www.phytozome.net).
b. Affymetrix assigned identifier for the predicted gene model
c. Genomic start and end for the predicted gene model
d. Affymetrix assigned identifier for each exon of a given gene model. This identifier was previously labeled as "Probeset ID" but was renamed to "Exon_ID" by our group for simplicity in the interpretation of the array file information.
e. Affymetrix assigned identifier representing individual 25-mer oligonucleotide selected for the given gene model.
f. Indicates the genomic start and end for the selected probe.
g. GC content percentage for the 25-mer oligonucleotide probe.
h. Sequence of the probe on the array.

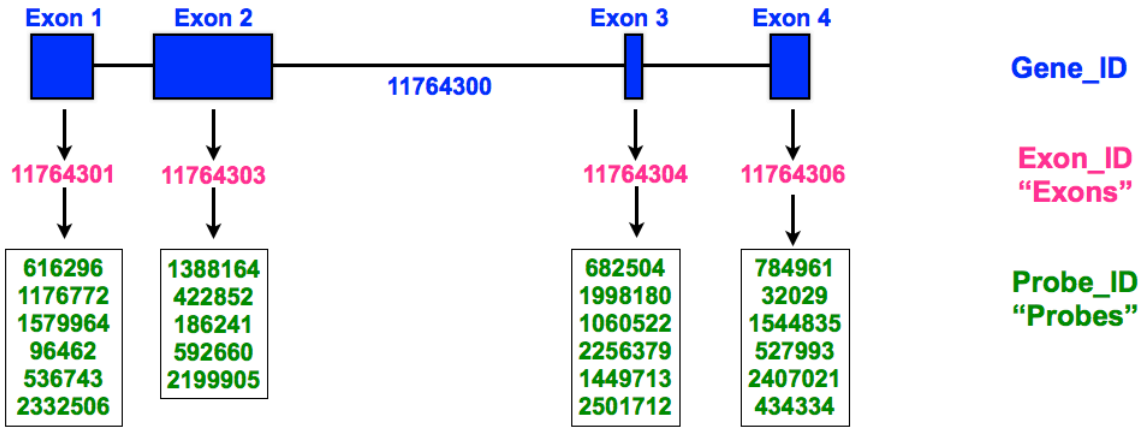

Figure 2. Illustration showing the organization of the Affymetrix identifiers (gene_id, exon_id, probe_id) for a given gene model. There are 23 probes selected for this gene model with an average of 6 probes per exon.

[3] http://seedgenenetwork.net/annotate#soybeanWT

A visual representation of the selected probes along the gene model is shown below in **Figure 3**.

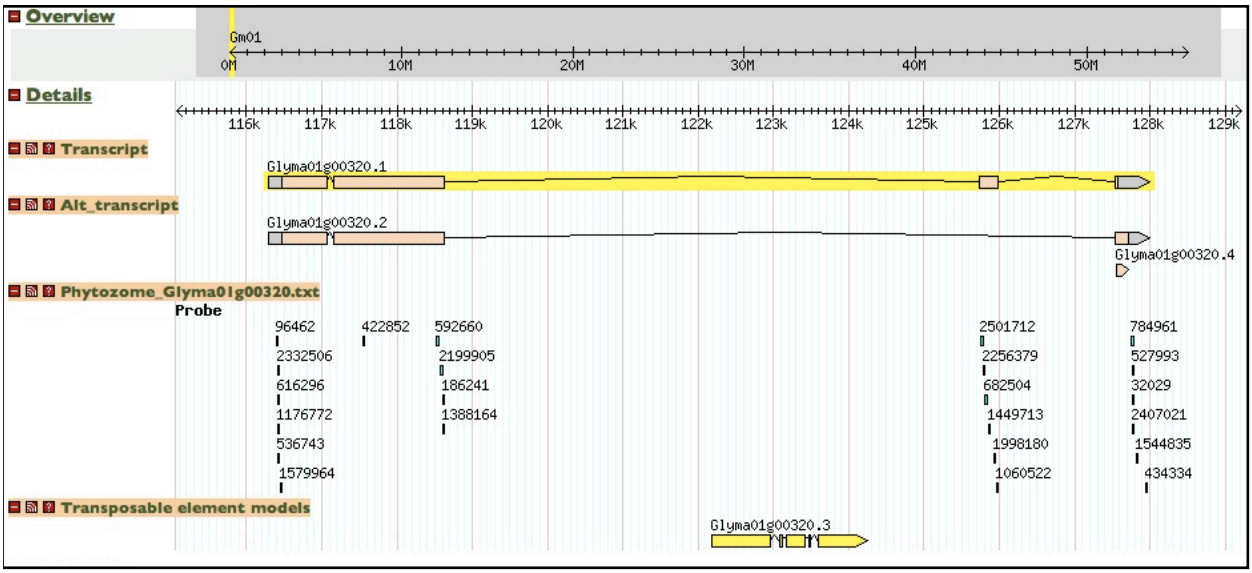

Figure 3. Visual representation of selected probes for gene, Glyma01g00320, in the soybean genome using Gbrowse.

There are several observations noted for this figure.

1. All of the exons have probes associated with them including probes designed in the untranslated regions (UTRs).
2. These probes were designed to interrogate two transcripts (Glyma01g00320.1 and Glyma01g00320.2). *Note there are no probes interrogating Glyma01g00320.3 or Glyma01g00320.4.*
3. There are six probes designed to cover exon 3. These probes will help distinguish between the two splice forms of the Glyma01g00320 gene.
4. The probes are not necessarily distributed equally within an exon.

III. SAMPLE PREPARATION AND LABELING PROTOCOL

Due to the design of this antisense array, a different protocol was created to prepare labeled targets for hybridization to this array compared to a sense array.

There are two methods for generating ***antisense*** WT targets for hybridization to the array[4]:

1. **Nugen Ovation Pico WTA System**
   a. This method allows you to start with 500 picograms of total RNA.
   b. This is a whole transcript labeling method and results in 6-10 µg of labeled antisense cDNA target, sufficient for hybridization to the Soybean WT array.
   c. Product information is available at http://www.nugeninc.com/nugen/index.cfm/products/amplification-systems/wt-ovation-pico/

2. **Merge Ambion WT Expression Kit with Affymetrix Second Strand cDNA synthesis**
   a. This method requires 50 ng of total RNA to start.
   b. This method results in double-stranded cDNA that has incorporated dU, which can be fragmented with Uracil DNA Glycosylase (UDG) and Apurinic/apyrimidinic endonuclease 1 (APE1) and end labeled using the Affymetrix Terminal Labeling Kit (P/N 702808 Rev. 3).
   c. This method requires an Ambion WT Expression kit, Affymetrix Fragmentation and Terminal Labeling kit, and second strand cDNA synthesis reagents from vendors provided in the protocol. Product information is available at http://www.affymetrix.com/estore/browse/products.jsp?productId=131557&categoryId=35699.
   d. The modified protocol for this method is available at our web site[5].

We have not tried either approach for generating the labeled targets for hybridization. Please contact Gene Tanimoto (gene_tanimoto@affymetrix.com) if you have questions or problems regarding these methods.

[4] Per email communication with Gene Tanimoto from Affymetrix.

[5] http://seedgenenetwork.net/annotate#soybeanWT

IV. FLUIDICS, IMAGE PROCESSING AND DATA ANALYSIS

A. WASHING AND STAINING

Follow the protocol for washing and staining using the fluidics protocol EuKGE-WS2v5_450 as described in Section 2 of the GeneChip Expression Analysis Technical Manual[6]. You will need to install the Soybean Antisense Library File (available at our web site[7]). Note: This library file will not work with the Soybean WT ***Sense*** array.

B. SCANNING AND DATA ANALYSIS

The GeneChip Operating System (GCOS) or Affymetrix GeneChip Command Console (AGCC) can be used to scan and visualize the image data. For probe level analysis, use the Affymetrix Expression Console (**Fig. 4**).

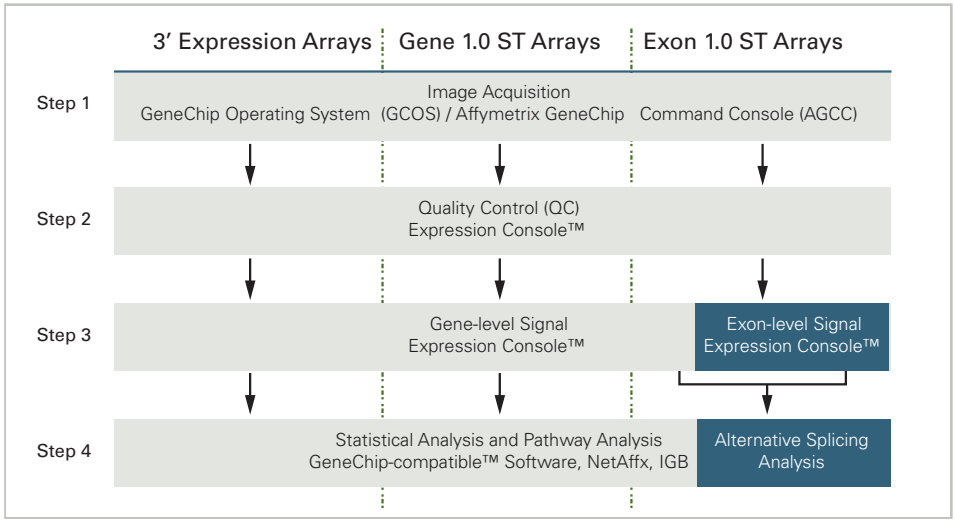

Figure 4. Data analysis workflow for Gene 1.0 ST and Exon 1.0 ST arrays using GCOS or AGCC, Expression Console. ST, sense target (Image taken from http://media.affymetrix.com:80/support/technical/appnotes/wt_appnote.pdf)

All software is freely available at the Affymetrix Software Download website[8]

[6] http://media.affymetrix.com:80/support/downloads/manuals/expression_analysis_manual.pdf

[7] http://seedgenenetwork.net/annotate#soybeanWT

[8] http://www.affymetrix.com/support/technical/software_downloads.affx?hightlight=true&rootCategoryId=34001