Lucy Cherbas, Kevin Bogart, Yi Zou, Peter Cherbas, and Justen Andrews

DGRC-2: Spotted oligonucleotide transcriptome microarrays for the Drosophila community

CGB Technical Report 2006-01
doi:10.2506/cgbtr-200601

DGRC-2: Spotted oligonucleotide transcriptome microarrays for the Drosophila community

Lucy Cherbas, Kevin Bogart, Yi Zou, Peter Cherbas and Justen Andrews

Drosophila Genomics Resources Center, Center for Genomics and Bioinformatics, Indiana University, Bloomington, Indiana 47405

INTRODUCTION
The Drosophila Genomics Resource Center (DGRC) provides the research community with access to genomics resources including microarrays. The first generation transcriptome microarrays fabricated by the DGRC (DGRC-1 arrays) are spotted with amplified DNA fragments. The DNA fragments are amplified from genomic DNA template by polymerase chain reaction using gene-specific primers. The primers were designed against release 1 of the D. melanogaster genome annotation, and the amplified DNAs were validated to perform well as microarray targets (Johnston et al. 2004). In recent annotations the coverage of the amplicons fell to 88% of the predicted genes (Costello and Andrews, 2006).

Several recent studies have demonstrated the favorable performance of spotted oligonucleotide microarray platforms (Kane et al. 2000; Bozdech et al. 2003; Wang et al. 2003). Long oligonucleotides perform as well as cDNAs in detecting differential gene expression. Additionally, since oligonucleotides are shorter, they can be selected against unique sequences that distinguish between members of a gene family or alternative transcripts from a single gene. Motivated by the advantages of oligonucleotide microarrays and the need to update the existing Drosophila transcriptome platforms, the International Drosophila Array Consortium (INDAC) developed a set of oligonucleotides for microarray fabrication. This set of oligonucleotides is now being used by several Drosophila genomics facilities to fabricate transcriptome microarrays (FlyChip, CDMC, DGRC)

This report gives an overview of the DGRC-2 oligonucleotide transcriptome microarrays printed by the DGRC using the INDAC set of oligos. It is intended to inform DGRC users about the platform and provide initial overview of their performance characteristics.
RESULTS AND DISCUSSION

Oligonucleotide design and synthesis
The International Drosophila Array Consortium (INDAC) was formed to produce a widely available and uniform set of array reagents, so that microarray data collected from different studies may be more easily compared. On behalf of INDAC, the FlyChip group (Gos Micklem, Steve Russell, and their collaborators at the University of Cambridge) designed a set of long oligonucleotides to release 4.1 of the D. melanogaster genome, using a custom implementation of OligoArray2. Oligos were chosen to have minimal sequence similarity to other genes, size range between 69-65 nucleotides with a minimal Tm window, and location biased towards the 3'-end of transcripts. The final set of 15,158 oligonucleotides has the size distribution given in Table 1. The sequences of the oligonucleotides can be downloaded from DGRC platforms, and Flymine.

Table 1. Distribution of oligonucleotide lengths

<table>
<thead>
<tr>
<th>Length (nt)</th>
<th>Total #</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>69</td>
<td>9791</td>
<td>64.6</td>
</tr>
<tr>
<td>68</td>
<td>749</td>
<td>4.9</td>
</tr>
<tr>
<td>67</td>
<td>730</td>
<td>4.8</td>
</tr>
<tr>
<td>66</td>
<td>713</td>
<td>4.7</td>
</tr>
<tr>
<td>65</td>
<td>3175</td>
<td>20.9</td>
</tr>
</tbody>
</table>

Annotation
We have updated the annotation of the oligonucleotides to release v4.3 of the Drosophila genome (Rupali Patwardhan, James Costello, Justen Andrews, Peter Cherbas, unpublished). Full details can be found in the gene list master file which is posted for each DGRC-2 print run on the DGRC web-site; these files will be updated as new annotations of the Drosophila genome are released. Briefly, of the 15,158 oligonucleotides, 14,543 are designed against Drosophila sequences (the remainder being controls of various sorts); after correction for sequences that are repeated in the collection, there are 14,447 unique Drosophila oligo sequences. Of these, 14,047 are predicted to hybridize to transcripts defined in the v4.3 annotation of the Drosophila genome; 142 of them hybridize to transcripts from more than one gene. Altogether, of the 14486 genes in the v4.3 annotation of the Drosophila genome, 13548 (93.5%) are represented unambiguously in the INDAC collection of oligonucleotides. 72 of these genes express non-coding RNAs, some of which are not polyadenylated and therefore will not be detected with oligo-d(T)-primed fluorescent targets. In the INDAC set of oligonucleotides, no attempt was made to include probes specific for different transcripts.
from individual genes; specific transcripts predicted to hybridize to each oligonucleotide probe are listed in the gene list master file on the DGRC web-site.

The probe sequences were designed to be as close as possible to the 3’ end of transcripts. Probes generally hybridize within a few hundred nucleotides of the polyadenylation site of the transcripts against which they were designed, but some probes also hybridize to longer alternative transcripts. For each probe, we have calculated the distance from the 3’ end of the probe to the polyadenylation site of each predicted transcript to which it can hybridize; for all probe/transcript combinations, the mean distance is 723 N (S.D. = 576).

Internal controls and array layout
The set of INDAC oligonucleotides includes the following internal controls:

1. Exogenous targets: 9 from Arabidopsis, 2 from soybean, 1 from apple. These serve as negative controls or targets for mRNA spike-in controls.
2. Exogenous genes: GAL4, lacZ, modified GFP, and FLPase.
3. A series of oligos designed to test for the intactness of RNA targets. The array contains multiple copies of a set of oligos made from different parts of the eukaryotic initiation factor (eIF-4a) transcript. Since most labeled targets are made as oligo(dT)-primed cDNAs, mRNA degradation and/or incomplete reverse transcription will result in overrepresentation of sequences from the 3’ portion of an mRNA. Therefore, a degraded RNA sample may be detected by a relative decrease in hybridization to the more 5’ eIF-4a probes. We have not yet experimentally validated this test.

We have supplemented these controls with the following:

1. The ubiquitously expressed genes actin 5c (act5c) and eIF-4a printed at a characteristic location in each of the 48 sub-arrays. These positive control spots serve as an easily recognized registration mark and as a check on the uniformity of hybridization across the slide.
2. A spotting-buffer-only control printed in each of the 48 sub-arrays. This serves as a negative control for background and check for carry-over between samples.
3. Array Control Sense Oligo Spots (Ambion Inc.). These serve as additional negative controls and can also be utilized with spiking RNAs for normalization to verify the equivalence of labeling between channels.

The spotting buffer negative control, act5c and eIF-4a positive controls, and Array Control Sense spiking controls (Ambion) are printed by each of the 48 pins at the start of each
block. The deconvolution files describing the microarray lay-out can be downloaded from the DGRC website.

Fabrication conditions

Important parameters contributing to the quality of spotted oligonucleotide microarrays include spotting buffer composition, DNA concentration, slide surface chemistry, immobilization conditions, and post-processing protocols. These parameters are strongly interdependent. For instance, different combinations of spotting buffer and DNA concentration give optimal results in combination with different slide surface chemistries or slides with the same chemistry but from different manufacturers (Taylor et al., 2003; Wang et al. 2003). We therefore conducted a pilot experiment to optimize fabrication parameters. Since the INDAC oligo set fabricated by Illumina have a 5'-amino modification, we were particularly interested whether 5'-amino linkage to epoxide surfaces would give improved results compared to amino silane surfaces.

We selected a subset of 24 oligonucleotides from the full transcriptome set and printed these to Aminosilane, Epoxide, and Aldehyde microarray substrates from a variety of manufacturers (Schott Nexterion, Corning, ArrayIT, and Full Moon Biosystems) with a matrix of the following parameters:

- Target (spotted DNA) concentration: 10, 20, 40 µM.
- Spotting buffer: SSC + Betaine, SSC, and NaPO₄.
- Immobilization Conditions: UV crosslinking energy (400, 600, 800mJ).

Each of the microarrays was hybridized with aliquots from a common pool of labeled cDNA reverse transcribed using RNA isolated from whole adult males and females. All slides were hybridized, washed and scanned under identical conditions.

There was a wide range of performance with different combinations of surface chemistry/manufacturer, spotting buffer, DNA concentration, and immobilization conditions (data not shown). Among the slide types tested the Corning UltraGAPS and Schott Nexterion Slide E gave the best overall performance in terms of signal intensity, signal-to-noise ratios, and spot morphology (Figure 1). Focusing on these slide types, we observed the following. The NaPO₄ buffer gave the best signal-to-noise ratio. The signal-to-noise ratio increased as the probe concentration was increased from 10 to 20µM, but decreased slightly at higher probe concentrations (Figure 1A). The decrease in signal probably occurs when steric hindrance from very high probe concentrations on the slide decreases the availability of the probe for hybridization. UltraGAPS spots are square in shape (as are the printing tips) but are about 20µm larger and contain about 30 more pixels than Schott slide E (Figure 1B,C). More importantly, the spots are more
homogenous; the UltraGAPS spots had a lower SD and CV among pixels in the two channels (Table 2). Based on these observations we are printing the DGRC-2 microarrays on Corning UltraGAPS (aminosilane) slides, with probe DNA at 30µM dissolved in 150 mM NaPO4 (pH 8.5) buffer, and probes immobilized using 600mJ UV energy.

**Figure 1. Effects of fabrication conditions.** (A) Effect of buffer type and target concentration on background subtracted feature intensity. Individual spots ranked by intensity (x-axis) plotted with corresponding intensity values for each condition (y-axis). (B, C) Effect of slide type on spot morphology. Images of Corning UltraGaps (B) and Scott Slide E (C) slides fabricated, hybridized and scanned under identical conditions.
Table 2. **Spot morphology descriptive statistics.** Numerical averages of Diameter, Feature Pixels, Standard Deviation, and Coefficient of Variation on 2 slide chemistries.

<table>
<thead>
<tr>
<th></th>
<th>Diameter</th>
<th>F635SD</th>
<th>F635CV</th>
<th>F532 SD</th>
<th>F532 CV</th>
<th>F Pixels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schott Slide E</td>
<td>103.27</td>
<td>1954.70</td>
<td>105.01</td>
<td>452.79</td>
<td>77.90</td>
<td>86.45</td>
</tr>
<tr>
<td>UltraGAPS</td>
<td>119.16</td>
<td>952.31</td>
<td>36.77</td>
<td>209.49</td>
<td>31.78</td>
<td>115.54</td>
</tr>
</tbody>
</table>

**Conditions of hybridization**

We have obtained good results from oligo arrays using exactly the same conditions that we use for amplicon arrays. All of the analyses described below were done using dendrimer hybridizations (Genisphere Array50 Alexa647 and Alexa546 kits). We have done a few hybridizations with amino-allyl labeled sequences, with similar results.

In our experience, the oligo arrays are more sensitive to variations in experimental conditions than the amplicon arrays. For example, we have successfully hybridized amplicon arrays in an oven rather than a water bath, but the reduced humidity of the oven led to poor results with oligo arrays. We therefore recommend carefully following the protocols posted on the DGRC web-site.

**Dynamic range of the signal**

Using identical fluorescent labeled targets, we find that the level of fluorescent signal is slightly lower in the oligo slides than in the amplicon slides; under conditions where a PMT setting of 400 gives the full range of signal from an amplicon array, we use a PMT setting of 450 to give a similar range of signal. After making this small adjustment in the gain setting of the scanner, we find similar levels of between-spot backgrounds. Therefore, the signal is slightly less in the oligo arrays, but the signal-to-noise ratio is similar to that seen in amplicon arrays.

**Comparison between amplicon and oligonucleotide microarrays**

**Number of spots that give a detectable signal**

Despite having a similar dynamic range in signal intensities the oligonucleotide and amplicon arrays differ dramatically in the number of immobilized probes that give a detectable signal when hybridized to the same labeled target (Table 3). This difference cannot be explained by the relatively small difference in the coverage of the
transcriptome by the two probe sets. Rather, it must mean either that the oligonucleotide platform is less sensitive (i.e. that the missing signal represents false negatives) or that the amplicon array has a significant background signal (i.e. that the excess signal on the amplicon platform represents false positives). A priori, we would expect that the oligo array would have less non-specific signal in the spots, both because the shorter sequences of oligos are less likely to cross-hybridize to sequences from other genes than are the amplicon spots, and because the synthetic oligonucleotide DNA is more pure than the amplicon DNA.

Table 3. Number of spots detected by hybridization to fluorescent target. Only spots corresponding to Drosophila genes are shown. Spots listed as hybridizing gave a detectable signal in the majority of slides hybridized to sequences from Kc cells.

<table>
<thead>
<tr>
<th></th>
<th>total spots</th>
<th>detected spots</th>
<th>% detected spots</th>
</tr>
</thead>
<tbody>
<tr>
<td>amplicon array</td>
<td>14745</td>
<td>14534</td>
<td>99%</td>
</tr>
<tr>
<td>oligo array</td>
<td>14530</td>
<td>8697</td>
<td>60%</td>
</tr>
</tbody>
</table>

The inclusion of several classes of internal controls on the amplicon microarrays allows us to estimate the possible sources of non-specific signal. The amplicon arrays contain the following types of negative-control spots:

1. “Barcode Elements”, which are made by PCR amplification of an intronic sequence using a Drosophila genomic template.
2. No-primer spots, made from sham PCR reactions that contain all components in experimental reactions except primers. In all other respects (thermal cycling, purification etc) these are treated identically to experimental samples.
3. Various spikes and other foreign sequences (e.g. GAL4, FLP, lacZ) that are made by PCR amplification from plasmid templates.
4. Spotting buffer. These spots have no DNA at all.

Following hybridizations, in which all of the negative control spots are expected to give no specific signal we observe different signal intensities for the different classes of control spots (Table 4). The spotting buffer and foreign sequence control spots have similar, low mean signal intensities. Since the foreign sequences contain DNA and the spotting buffer controls do not, this indicates that the presence of DNA per se does not contribute significantly to non-specific signal. The very small difference between the spotting buffer and foreign sequence signals may be due to physical trapping of labeled target. The Barcode Element spots and the no-primer spots both have a greater mean signal intensity than do the spotting buffer and foreign sequence control spots. The Barcode Element and no-primer spots are distinguished from the foreign sequence spots by the probable presence of contaminating genomic Drosophila DNA template. This
suggests that the greater non-specific signal observed for the Barcode Element and no-primer spots is due to genomic DNA co-purified with the amplicon DNA. Alternatively, it is possible that the PCR reactions amplified a low level of non-specific DNA that was not detected by agarose gel electrophoresis. In any event, the non-specific signal observed in the Barcode Element and no-primer spots would also be expected to contribute to the signal in experimental probes. We therefore conclude that a significant component of in-spot non-specific background on the amplicon arrays is hybridization to contaminating genomic sequences and that this accounts for many of the genes that are detected on the amplicon array and not on the oligonucleotide array. As support for this notion, we point out that ovo, Salivary gland secretion 3 (Sgs3), eyeless (ey), Antennapedia (Antp), Abdominal B (Abd-B), and several cuticle protein genes are among the probes that give detectable signal on the amplicon arrays but not the oligo arrays when hybridized to RNA from Kc cells; all of these genes are known to have highly localized expression and their transcripts are expected to be found in Kc cells.

Table 4. Signal from negative control spots. The hybridization signal (A) is shown as an average (±SD) for each category of negative control spot. Amplicon and oligo microarrays were hybridized with labeled cDNAs made from the same samples of RNA.

<table>
<thead>
<tr>
<th></th>
<th>mean A (amplicon)</th>
<th>mean A (oligo)</th>
</tr>
</thead>
<tbody>
<tr>
<td>spotting buffer</td>
<td>2.0 (±1.7)</td>
<td>0.9 (±1.5)</td>
</tr>
<tr>
<td>foreign sequences</td>
<td>3.1 (±1.5)</td>
<td>3.0 (±2.0)</td>
</tr>
<tr>
<td>Barcode Element and no primer</td>
<td>5.9 (±1.8)</td>
<td>not applicable</td>
</tr>
</tbody>
</table>

Measurements of differential expression

To compare the measurements of differential expression in oligo and amplicon arrays, we estimated inductions following ecdysone treatment in Kc cells. Data compiled from 12 amplicon slides were compared to data from 6 oligo slides. In each case, multiple slides represented biological replicas; the same RNA samples were used for both amplicon and oligo arrays. The data were normalized by OLIN after background subtraction, and the best value of M was calculated using ArrayLOD (Cherbas et al., 2006). We considered 831 genes, chosen because they demonstrated a significant early ecdysone response in at least one cell line (of 10 that we tested), and had a detectable fluorescent signal for Kc cells in both platforms. Of these, 227 had a significant response in Kc cells, detectable in at least one of the two platforms. We observed the following:

When the amplicon arrays demonstrated a significant ecdysone response, the oligo arrays confirmed that response in virtually all cases. The reverse was not the case: We
found many examples of genes whose ecdysone response was clearly detectable in the oligo arrays, but not in the amplicon arrays. Our data, which will be published in more detail elsewhere, suggest that the non-specific background signal generated in amplicon microarrays may obscure differential expression of genes which generate a weak absolute fluorescent signal. In probes which give strong signals, the background contribution is insignificant, and the amplicon and oligo platforms give identical fluorescence ratios. In probes which give weak signals, the background contribution is significant, and in some cases dominates the observed fluorescence; in these cases, the ratio observed in amplicon microarrays is less than that observed in oligo microarrays. Several ecdysone inductions apparent in the oligo array but not in the amplicon array have been confirmed independently by Northern or RT-PCR analysis.

These observations are illustrated in Figure 2. Note that there are many points which fall along the line of identity or which show a lower absolute value of M for the amplicon array than for the oligo array (octants A and D), but very few in which M from the amplicon array has greater magnitude than M from the oligo array (octants B and E) or in which the sign of M is reversed between the two platforms (quadrants C and F).

**Figure 2. Comparison of differential expression measurements on oligo and amplicon arrays.** Each point represents a best value of M for the set of oligo arrays (X axis) and for the set of amplicon arrays (Y axis); the intensity of the shading is proportional to the number of superimposed data points, with the darkest points representing 20-40 probes.

**General conclusions**
Oligo microarrays will give results generally similar to those obtained from amplicon arrays. The major differences are as follows:
Fewer spots hybridize. This is in general a good thing, indicating decreased background from contaminating genomic DNA in the spots, cross-hybridization by related cDNA sequences and trapping of fluorescent cDNA.

Some estimates of differential expression are quite different in amplicon and oligo arrays. Our observations suggest that in most cases the differences can be explained by decreased background and cross-hybridization in the oligo array, plus possibly some differences in detected transcription units. Overall, differential expression observed in the amplicon array is confirmed by the oligo array, but the oligo array is more sensitive to differences in species which generate low signals.

**Acknowledgments**
We thank members of the International Drosophila Array consortium for cooperating in the oligonucleotide project; Steve Russell and Gos Micklem for design of the oligonucleotides; Tim Westwood and Mandy Lam, for sharing data and helpful discussions; and members of the DGRC and CGB, particularly James Costello and Rupali Patwardhan, for assistance. The project described was supported by Grant Number 1 P40 RR017093 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of NCRR or NIH.

**REFERENCES**

http://www.indac.net/
http://berry.engin.umich.edu/oligoarray2/
http://www.ambion.com/
http://www.illumina.com/
http://dgrc.cgb.indiana.edu/microarrays/downloads.html


