Assessing the Performance of Different High-density Tiling Microarray Strategies for Mapping Transcribed Regions of the Human Genome

Olof Emanuelsson[1,+], Ugrappa Nagalakshmi[2], Deyou Zheng[1], Joel S. Rozowsky[1], Alexander E. Urban[2,3], Jiang Du[4], Zheng Lian[3], Viktor Stolc[5], Sherman Weissman[3], Michael Snyder[1,2,*], Mark Gerstein[1,4,*]

[1] Department of Molecular Biophysics and Biochemistry, Yale University, New Haven CT 06520-8114, USA

[2] Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven CT 06520-8103, USA

[3] Department of Genetics, Yale University School of Medicine, New Haven, CT 06520–8005, USA

[4] Department of Computer Science, Yale University, New Haven, CT 06520-8285, USA

[5] Center for Nanotechnology, NASA Ames Research Center, Moffett Field, CA 94035, USA

[*] To whom correspondence should be addressed: E-mail: michael.snyder@yale.edu (M.S.), mark.gerstein@yale.edu (M.G.)

[+] Present address: Stockholm Bioinformatics Center, AlbaNova University Center, Stockholm University, SE-10691 Stockholm, Sweden

Running title: High-density Tiling Microarray Strategies

ABSTRACT

Genomic tiling microarrays have become a popular tool for interrogating the transcriptional activity of large regions of the genome in an unbiased fashion. There are several key parameters associated with each tiling experiment (e.g., experimental protocols and genomic tiling density). Here, we assess the role of these parameters as they are manifest in different tiling array platforms used for transcription mapping. First, we analyze how a number of published tiling-array experiments agree with established gene annotation on human chromosome 22. We observe that the transcription detected from high-density arrays correlates substantially better with annotation than that from other array types. Next, we analyze the transcription-mapping performance of the two main high-density oligonucleotide array platforms in the ENCODE regions of the human genome. We hybridize identical biological samples and develop several ways of scoring the arrays and segmenting the genome into transcribed and non-transcribed regions, with the aim of making the platforms most comparable to each other. Finally, we develop a platform comparison approach based on agreement with known annotation. Overall, we find that the performance improves with more data points per locus, coupled with statistical scoring approaches that properly take advantage of this, where this larger number of data points arises from higher genomic tiling density, and the use of replicate arrays and mismatches. While we do find significant differences in the performance of the two high-density platforms, we also find that they complement each other to some extent. Finally, our experiments reveal a significant amount of novel transcription outside of known genes, and an appreciable sample of this was validated by independent experiments.

NOTES:

Gene Expression Omnibus (GEO) accession numbers are provided for the array data. Supplementary material is provided.

INTRODUCTION

Mapping transcribed regions of the human genome in an unbiased fashion is a crucial step towards understanding at a molecular level the organization of hereditary information and the specific functions of each human cell or tissue type. To this end, a number of approaches using genomic tiling microarrays have been tested and published over the last few years, including key studies by Kapranov et al. (2002), Rinn et al. (2003), Schadt et al. (2004), Bertone et al. (2004), and Cheng et al. (2005). While the strategies differ substantially in most of their details, they all share a basic array design concept: to construct an array whose probes (the molecules attached to the microarray at the manufacturing) cover all of the non-repetitive sequence of the genome or genomic region under investigation.

Kapranov et al. (2002) used a high density oligonucleotide array design containing perfect match probes of length 25 bp and corresponding mismatch probes. The arrays were synthesized *in situ* (directly on the supporting array material) using physical masks, and covered chromosomes 21 and 22 with probe starting positions spaced every 35 bp (genomic distance). They were hybridized with samples representing 11 cell lines. The data was later reanalyzed (Kampa et al., 2004) and a more sophisticated approach to genomic segmentation was introduced. We refer to this setup as the Affymetrix tiling array platform.

Rinn et al. (2003) mapped transcribed regions of chromosome 22 with an array of PCR products (amplicons), tiled end-to-end with a probe size range of 300-1,400 bp. This array represents the PCR tiling array platform and was hybridized with placenta poly(A)+ RNA (Rinn et al., 2003) and later with RNA from two cell lines (White et al., 2004).

Schadt et al. (2004) used tiling arrays where the probes were synthesized on the array using the Agilent ink-jet technology (Shoemaker et al., 2001). They tiled chromosomes 20 and 22 with 60-mers uniformly spaced every 30 bp. The statistical treatment of the data was presented in Ying et al. (2003).

Bertone et al. (2004) used oligonucleotide microarrays with 36 bp probes spaced every 46 bp to map transcribed regions of the entire non-repetitive portion of the human genome. The arrays are synthesized *in situ* using maskless technologies developed by NimbleGen Systems. We refer to this as the MAS (maskless array synthesis) tiling array platform (Singh-Gasson et al, 1999).

Cheng et al. (2005) used an updated version of the Affymetrix platform with a tighter spacing of the probes, every 5 bp, and covering 10 chromosomes of the human genome. Transcript maps were generated for polyadenylated cytosolic RNA from eight cell lines (and for one of these cell line, also non-polyadenylated RNA).

These different studies produced a wealth of data. However, the experiments represent very different choices in array design and manufacturing, RNA extraction and hybridization conditions, and data processing methods. As such, comparing the results from these studies is not trivial.

Here, we outline some of the key parameters differentiating the various studies. The array design parameters include the length and genomic spacing of the probes, the use of mismatch probes and whether to cover one or both genomic strands. For the oligonucleotide tiling experiments referenced above, the probe length varies between 25-70 bases. The genomic spacing of the probes is measured between probe initiation points, and can range from the smallest possible distance of one single base up to the length of the probe, or even further. At the design stage it is important to minimize potential cross-hybridization, self-pairing, and other probe sequence artifacts such as DNA secondary structure formation (SantaLucia and Hicks, 2004). Genomic regions considered as repeats (by, e.g., RepeatMasker (A.F.A. Smit and P. Green, unpublished)) are usually omitted from the design due to potential cross-hybridization. If some flexibility is allowed in the design process, probes may be chosen so as to achieve better probe thermodynamics. This is possible for arrays interrogating genes (Mathews et al, 1999; Hughes et al, 2001; Rouillard et al, 2003), but for tiling arrays with high genomic density probe optimization options are limited (Bertone et al., 2006).

The experimental protocols for extraction, labeling, and hybridization of the RNA sample to the array vary considerably. Choosing the type of target RNA (i.e., tissue or cell line, poly(A)+ or total RNA), and the reactions and conditions to use in the hybridization will affect the results. The number of technical and biological replicates is an additional crucial parameter, more replicates potentially enables greater certainty and detail in the interpretation of the results.

Once the tiling arrays have been designed, manufactured, hybridized with labeled RNA, and the hybridization intensities have been extracted, there are a number of ways to transform the raw

intensities into a score for each probe. This is usually done employing statistical methods such as a sign test or the t-test. Exactly what methods are available depends on the design features of the array, such as the presence of mismatch probes. The segmentation of the genome into transcribed and non-transcribed regions is then performed based on the scores.

Our goal is to assess different tiling microarrays that are currently used for transcription mapping, an area where no detailed comparison thus far has been performed, and ultimately to aid the ENCODE consortium when choosing strategy for the multiple tissue whole-genome transcription mapping of the human genome (The ENCODE Project Consortium, 2004). Previous work on comparing gene-based microarrays include papers by Tan et al (2003), Jarvinen et al (2004), Mah et al (2004), Park et al (2004), and Yauk et al (2004). Most of these indicate differences in the gene expression results from different microarray platforms, which have been attributed to differences in data processing or inadequate choice of comparison metrics (Larkin et al, 2005).

We start our microarray comparison by analysing a set of already published chromosome 22 transcription experiments. Overall, this study indicated that high-density oligonucleotide arrays perform significantly better than amplicon (PCR) arrays.

We then describe a direct comparison of the two *in situ*-synthesized oligonucleotide-based platforms *MAS* (Bertone et al, 2004) and *Affymetrix* (Kapranov et al, 2002) on the manually picked part of the ENCODE regions of the human genome (http://www.genome.gov/10005107). We hybridized identical biological samples to the arrays and developed a unified data processing scheme based in statistical treatment of the data. Using this approach, we compare the results from the two platforms with each other and with the recently generated Gencode gene annotation (Guigo et al., 2003; Ashurst et al., 2005; http://genome.imim.es/gencode/).

RESULTS AND DISCUSSION

Pilot Study: Comparison of Public Chromosome 22 Tiling Data

We carried out an initial comparison of previously published transcription maps of chromosome 22 generated from PCR-based tiling arrays (Rinn et al., 2003; White et al., 2004) and two

oligonucleotide tiling array platforms, MAS and Affymetrix (Bertone et al., 2004; Kapranov et al., 2002) (Figure 1). These maps were generated from 15 separate experiments (tissues or cell lines). We used the RefSeq annotation (Pruitt et al., 2005) as a benchmark since Gencode annotation is not yet available for the entire chromosome 22. For each experiment, we measure the consistency between gene annotation and transcribed regions identified by individual studies. The transcription data from oligonucleotide arrays agrees better with the RefSeq exon annotation than the data from PCR arrays, an observation that holds true across all experiments. The results have to be interpreted with some care since they were not obtained with the same biological samples or scoring schemes. Nonetheless, we conclude that PCR-based arrays are clearly less useful for a detailed transcription mapping study, possibly because of their lower genomic resolution. Spotted arrays (e.g., PCR-based) also have a significantly lower feature resolution on the array compared with arrays with *in-situ* synthesized probes. Therefore, we focus our subsequent experimental and analysis efforts on the oligonucleotide tiling microarrays.

APPROACH

Oligonucleotide Array Designs and Hybridizations

An oligonucleotide array containing 36 bp oligonucleotides that tile both strands of the nonrepetitive sequence of the ENCODE regions end-to-end (allowing some positional shifts to reduce self-complementarity) was prepared using maskless photolithography, MAS (maskless array synthesis). The MAS arrays cover both strands of the ENCODE regions ENm001-ENm011 (11.6 Mbases). An Affymetrix ENCODE array, which covers one strand of the entire ENCODE region on one array, tiled with 25-mer oligonucleotides with an average distance between oligonucleotide starts of 21 bases was obtained from the manufacturer. This array has both perfect match (PM) and mismatch (MM) probes.

As outlined in Table 1, five different hybridization experiments were carried out: Two different RNA targets (placenta poly(A)+ RNA and NB4 total RNA) were hybridized to the two different array types. (We follow the nomenclature of Royce et al. (2006), i.e., *target* or *sample* is the RNA extracted from a biological entity (tissue or cell line) and which is hybridized to the *probes* on the microarray.) The Affymetrix arrays were hybridized according to the manufacturer's recommendation. The MAS arrays were hybridized using two different experimental protocols,

MAS-B, described in Bertone et al. (2004), and MAS-N, a variant of the manufacturer's recommended protocol. The placental RNA was hybridized using both MAS protocols, the NB4 RNA only with MAS-N.

Generating Comparable Maps of Transcriptionally Active Regions (TARs)

(i) Development of consistent scoring schemes

To bring the outcomes from the two technologies MAS and Affymetrix into a comparable form, we developed ways of scoring them similarly. For each spot on the microarrays, a hybridization intensity was collected. For oligonucleotide tiling arrays, it is usually advantageous to aggregate the intensities from probes that are adjacent to each other in genomic space (Kampa et al., 2004; Cheng et al., 2005; Royce et al., 2005). This is done by applying a sliding genomic window encompassing multiple probes and converting the intensities within the window into a score, which is assigned to the middle probe. The windowed approach is logical since we are ultimately interested in obtaining a set of regions whose intensities are significantly higher than the background, and we expect those regions to be of the same length as exons (150-200 bp on average, depending on exon type) rather than of single probes (25-36 bp in this study).

We developed new ways of scoring the MAS arrays, and describe these in terms of three levels of scoring: single probe intensities, robust statistics within a sliding window, and robust statistics using paired data within a sliding window (Cawley et al., 2004).

Single probe intensities

Single probe intensity scoring uses the raw intensities from the arrays. By wisely choosing methods and parameters to deal with the genomic segmentation (see below) it is possible to obtain reasonable results from this approach (Bertone et al., 2004). In this approach, both intra- and inter-array normalization of the microarray data may be particularly important (Royce et al., 2005).

Robust non-parametric statistics within a sliding window

We employed the sign test for scoring MAS array data. The sign test is attractive since it is statistically robust and it does not assume normally distributed data. Comparing each intensity

within a sliding genomic window of a specified size with the array median, yields a measure or a score of the significance of the intensities (see Methods for details). It is easy to include multiple replicates in this scheme: each probe is simply compared to the median intensity of its own array, and no inter-array normalization is necessary. The number of available score levels is restricted, however, due to the discrete values introduced by the counting (it is a binomial), and it may not be sufficient in situations in which discerning the top scores (say, top 5%) from near-top scores is important. With an average genomic spacing of 36 bp between the starts of two adjacent probes, the window (160 bp) encompasses five probes. We also applied the sign test on the Affymetrix data as a part of our comparison.

Robust non-parametric statistics using paired data within a sliding window

When paired data is available, such as the perfect match (PM) and mismatch (MM) probe intensities on Affymetrix arrays, the paired Wilcoxon signed rank test is a more powerful option than the standard sign test. It was first used with tiling microarrays by Cawley et al. (2004) to score ChIP-chip data, and it is also immediately applicable to transcription data as is shown in Kampa et al. (2004) and Cheng et al. (2005). All pairwise PM-MM differences within the window are calculated and a p-value, which essentially measures how significantly the distribution of PM-MM differences is skewed to either side around zero, is calculated, along with the corresponding point estimate (the pseudomedian). While this approach is analogous to the standard sign test, it has considerably greater statistical power.

The MAS arrays did not contain proper mismatch probes. Instead, we tried to simulate these using the complementary strand oligo of the MAS arrays as the “mismatch” probe. We call this approach the Fwd-Rev scoring, and it is justified on the MAS-B (placenta) data since the correlation between forward and reverse strand probes is close to the correlation between PM and MM probes for the Affy placenta data (Table 2; below).

(ii) Segmentation of genomic regions

After obtaining one score value per oligonuclotide probe, the next step is to construct a transcription map based on these scores, i.e., to segment the genomic regions into transcribed and non-transcribed regions. We call the transcribed regions TARs ("Transcriptionally Active Regions"; Rinn et al, 2003), regardless of overlap with genes, exons, or other genomic features.

(Note, an alternate term, transfrag ("Transcriptional Fragment"), was introduced by Kampa et al. (2004)).

Maxgap/minrun segmentation

In Bertone et al. (2004), TARs were generated by requiring at least 5 adjacent probes with a raw intensity in the top 10% of all intensities of that slide. Thus, the threshold above which to consider a probe "positive" was the intensity value corresponding to the 90^{th} percentile, and any probe that was below the threshold immediately terminated the transcribed region. In the Affymetrix series of publications (Kapranov et al., 2002; Cheng et al., 2005), the threshold for generating TARs was based on setting a maximum false positive rate of the hybridization levels of negative bacterial controls, thus enabling an optimized percentile cutoff for each array set and biological sample. Furthermore, gaps were allowed, such that a maximum stretch of a certain number of nucleotides (called maximal gap, or maxgap for short) with a score below the threshold was allowed between probes whose scores were above the cutoff. Typically, the maxgap parameter allows one or two probes to be below the cutoff, while still being incorporated into the TAR. The total length of a TAR is then required to be of at least a certain length (a minimal run, or minrun), usually corresponding to at least two probes.

HMM segmentation

As an alternative to the maxgap/minrun segmentation, a hidden Markov model (HMM) (Rabiner, 1989; Li et al., 2005; Ji and Wong, 2005) was employed to predict TARs, given the derived probe scores (above). Each probe can be in one of four HMM states (TAR, non-TAR, and two intermediate transition states), emitting the assigned score (i.e., the emission spectrum is continuous). The parameters of the HMM can be estimated by learning from the sequences of probes which fall into regions with known transcription characteristics (e.g., according to gene annotation). The HMM can then be applied to sequences of probes bearing the same scoring protocol to determine the most likely corresponding state sequence, in order to identify TARs (Viterbi decoding).

Platform Comparison

We have analyzed the five microarray tiling experiments, representing the MAS and Affymetrix platforms, introduced in Table 1 at multiple stages throughout the data processing.

(i) First, we calculate a correlation coefficient between the raw hybridization intensities of technical replicates within each platform, to assess the level of basic experimental reproducibility. We also assess the overlap of preliminary TAR sets generated from technical replicates using single probe intensities.

(ii) Before proceeding to the next level, the TAR sets to be compared must be determined. This includes decisions about what scoring algorithm, what segmentation method, and what corresponding parameter settings to use for each of the five experimental data sets. In summary, the input data type (e.g., PM only or PM-MM), the number of replicates, the scoring algorithm and, if applicable, its corresponding genomic window size define the scoring scheme, which together with the segmentation algorithm and its parameters specify a particular TAR set.

(iii) The resulting sets of transcribed regions are compared with each other, both within and across microarray platforms and biological samples, and the degree of overlap between detected transcription and gene annotation is measured. For the annotation comparison, we have chosen to use the Gencode annotation, which aims at finding and verifying all protein-coding genes in the ENCODE regions. There are two main measurements: how much of the known annotated exons are covered by a detected transcribed region ("sensitivity"), and the degree to which the detected transcription falls within known exons ("positive predictive value", PPV). The PPV is defined as the number of nucleotides in TARs that overlap with exonic regions, divided by the total number of nucleotides in the TAR set (this is sometimes referred to as "specificity" (Burge and Karlin, 1997)). The sensitivity is defined as the number of nucleotides in annotated exons that overlap with TARs, divided by the total number of nucleotides in annotated exons. We do not expect a sensitivity of 100% since in any given tissue or cell line at any given time point, far from all annotated genes will be expressed. Also, we do not expect a PPV of 100% since Gencode, although arguably the most comprehensive and accurate gene annotation available, is incomplete.

(iv) The transcription status is assessed for all 1342 annotated splice variants of all 264 known genes in the Gencode annotation of regions ENm001-ENm01. For each splice variant, the scores of all exon overlapping probes are collected and the transcription status assessed using the sign test (by comparing each individual score with the median score) (Bertone et al, 2004). A gene is considered transcribed if at least one of its splice variants is deemed transcribed at the chosen significance level. We also assess for each transcript (with at least two exons) how many of its

exons are considered transcribed, based on the median intensity score of each exon within the transcript. Ideally, either all exons or no exons should be transcribed. We call this concept the multi-exon coherence, and a suitable quantitative measure is the percentage of transcribed transcripts that display all their exons as "on".

(v) Finally, we perform experimental validation of the results in placenta (MAS-B and Affy), using reverse transcriptase-PCR on a subset of the novel TARs (including both TARs unique to a platform and TARs present in both). We also perform experimental validation of a number of genes with differing transcription status in the two platforms, including as a negative control a set of genes that are considered off in both platforms.

The results are available at http://tiling.gersteinlab.org/platformcmp.

OUTCOMES

Conclusions about Optimal Scoring and Segmentation Systems

We first examined the effect of the segmentation threshold on the size of the resulting TAR sets. The results are in Figure 2a and Figure S1, and, as expected, the lower the segmentation threshold the greater the size of the TAR sets. The step-like pattern in these figures is because of the finite number of available scores. For the minrun/maxgap segmentation, the maxgap parameter was set to 50 for the Affymetrix data and 80 for the MAS data, thus including in a TAR a probe whose score is below the score threshold, if it is flanked on both sides by probes with scores above the threshold. Other maxgap settings were tested but the results did not improve in terms of gene annotation agreement (data not shown). The minrun parameter was set to 50 bp, i.e., the minumun length of a TAR is 50 bp. Thus, at least four probes have to be included in an Affymetrix TAR and three for the MAS array TARs.

Different scoring schemes give different results, and also differ from the results obtained using single probe intensity scores, when comparing to the gene annotation. As is clear from Figure 2b (and Figure S2a), the standard sign test provides the best performance for MAS-B (placenta) if a sensitivity at or above 25% is required, but its improvement in terms of PPV when increasing the segmentation threshold is modest. The Fwd-Rev scoring is more sensitive to the choice of

threshold, and it performs better than the sign test scoring for segmentation thresholds above the 94th percentile. For the standard sign test, we also tried applying different weights to the probes within a window, e.g., multiplying each score with a discretized Gaussian, but no improvement in performance was recorded (Figure S2).

For the Affymetrix data, Figure 2c (placenta) and Figure S3 (NB4) reveal that the use of mismatch probes improves performance, in particular for the NB4 total RNA experiment, and that the Wilcoxon scoring performs very well. Figure 2c shows that for sensitivities up to 35%, using mismatch probes is a better strategy than doubling the genomic density of the probes, and also that it is better to use a single array with a PM-MM setup than two replicates with PM probes only.

Figure 2d and Figure S3 show that the more elaborate scoring models using replicates outperform the single probe intensity scoring in terms of sensitivity and PPV for both MAS and Affy.

These two points (elaborate scoring with replicates and the advantage of using mismatches) are further illustrated in Figure 2e, where the positive predictive values (PPV) of the TAR sets when choosing a segmentation threshold that corresponds to a sensitivity of 30% have been plotted.

The analysis of different segmentation algorithms reveals that the TAR sets generated by the non-parametric HMM segmentation (Viterbi decoding) are biased towards a high sensitivity (Figure 2b and Figure S2) where it performs on par with the minrun/maxgap algorithm.

Results from comparison pipeline

(i) Replicate comparison of unprocessed hybridization intensities

As is shown in Table 2, we obtained Pearson correlation coefficients of 0.83 and 0.96 for placenta MAS-B and MAS-N data, respectively, measured on pairwise comparison of the raw hybridization intensities of the arrays. The figure for Affymetrix was 0.96, and NB4 results were similar. We also note that the correlation of PM and MM probes for Affy placenta is close to the correlation of Fwd and Rev probes for MAS-B. Comparing the preliminary TAR sets, generated

from single arrays, across technical replicates (Table S1) again indicates that the MAS-B data is the most variable.

(ii) Choosing TAR sets to include in comparison

For each of the five experiments, the best-performing scoring and segmentation algorithm was chosen, and the segmentation threshold was tuned to generate TAR sets of roughly equal size, as measured in number of bases. The chosen sets are presented in Table 3 (and its extended version Table S2), and the points corresponding to these sets in Figure 2a and Figure S1 have been circled. To enable a comparison of the TAR sets, MAS array TARs on the two strands were merged into one set of unstranded TARs for each biological sample (Affymetrix TARs do not have strand information). For MAS, the standard sign test scoring was chosen, and for Affymetrix, the Wilcoxon signed rank test (pseudo-median). The segmentation thresholds range from the 87^{th} to the 93^{rd} percentiles for the various sets. The resulting sizes of the sets included in the subsequent comparison range from 629 kb to 701 kb (between 2545 and 4674 TARs). The total number of bases in exons in the analyzed regions (ENm001-ENm011) is 1001 kb, which means that the chosen sets can reach a sensitivity of 63-70% at most (as discussed above). The length distributions of all five TAR sets are unimodal and decay roughly exponentially (Figure S6).

(iii) Compare TAR sets with each other and to Gencode annotation and conserved regions

Figure 2d and Figure S3 reveal that the agreement with annotation is better for the Affy sets than for the MAS sets, both placenta and NB4. While similar sensitivity levels are achievable, the Affy TAR sets reach significantly higher PPVs. Likewise, the agreement is larger for the placenta sets (MAS-B, MAS-N, and Affy) than for the NB4 sets (MAS-N and Affy). This is also summarized in Table 3.

Figure 3a shows the overlap between the placenta and NB4 TAR sets from the different experiments. As a measure of the overlap between the sets we calculate a ratio $R = |\cap|/|U|$ for each pairwise comparison, where the numerator represents the size of the intersection and the denominator represents the size of the union of the two sets under comparison. For two sets that agree completely, R equals 1. We find that the MAS-B placenta TAR set agrees better with the Affy placenta TAR set (R=0.22) than the MAS-N TAR sets do with Affy (R is 0.16-0.17). Between 62% and 72% of the nucleotides in the placenta sets are exclusive to a particular

experiment (pairwise comparison MAS-B vs. Affy and MAS-N vs. Affy). For the NB4 total RNA TAR sets (MAS-N and Affy), more than 70% of the nucleotides in either TAR set are exclusive to that set.

As shown in Figure 3b, the overlap across the different biological samples within each experimental technology is larger than the overlap within the same biological sample between the two experimental technologies. An extreme example is the *MAS-N* placenta and NB4 sets which agree much better (dashed-dotted brown line; R=0.67) than *NB4* MAS-N and Affy (solid black line; R=0.16). Restricting the overlap calculations to the subset of TARs that overlap conserved or exonic regions, or both (i.e., moving to the right in Figure 3b; Figure S8) yields higher values of R for the within-biological sets comparisons (black) and for the within-Affy comparisons (solid brown), but not for the within-MAS comparisons (non-solid brown). Consequently, there is no enrichment for conserved regions or known genes within the common parts of the MAS TAR sets.

The bimodal distribution in Figure 4 shows that most Gencode unique exons are either fully covered by a TAR (>90% of exon nucleotides overlap with a TAR) or not covered at all (<10%). This is true for all TAR sets (Figure S9). We notice a slight 3' bias for the Affymetrix poly(A)+ data (but not for the MAS data), detecting 32% of the 3' exons and 25% of the 5' exons entirely.

Table 4 and Figure S10 show a comparison of each of the five TAR sets to a set of conserved elements, generated from the union of conserved regions called by the Threader Blockset Aligner (TBA) (Blanchette et al., 2004) and MLagan (Brudno et al., 2003). The union set of conserved elements covers approximately 10% of the ENCODE regions. We find that most of the novel (intergenic) TARs do not overlap with conserved regions. Only 7-8% of Affymetrix novel TARs and 1-2% of MAS novel TARs overlap fully (>90%) with conserved regions.

(iv) Transcription status of known genes and exons

The transcription status of all known splice variants (transcripts) of the 264 Gencode genes in the regions ENm001-ENm011 was assessed, and the results are shown in Table 5. A gene is considered as "transcribed" if at least one of its transcripts is detected at significance level $p<0.001$. If a transcript has less than 10 probes it will be unable to reach a p-value below 0.001, and if this is true for all splice variants of a gene, that gene is in the "Too few probes" category. In

total 158 (69.3%) genes are considered transcribed according to both platforms, and 221 (83.7%) according to at least one platform (similar percentages on the transcript level). In Table 6, the multi-exon coherence (either all exons on or all exons off) is assessed and found to be better for Affy. For both placenta and NB4 there is an enrichment of multi-exon coherence in transcripts that are considered as transcribed in both platforms. One example is the Affy NB4 set for which in total 15.5% of all transcripts have all their exons transcribed, while 26.9% of the transcripts that are on in both NB4 sets have all their exons transcribed. The difference in score distribution between exons and introns is assessed (Figure S7), and for all five sets exons are indeed overrepresented at the high end of the score spectrum, but also many introns have high scores.

(v) Experimental validation of novel TARs and known genes

Experimental validation of the microarray transcription data is crucial to the interpretation of the results. Table 7 shows that we used RT-PCR to assess in total 144 regions experimentally in placenta. Of these, 98 were novel TARs (no overlap with known genes). The experiments verified the presence of 56.4% (22/39) of the assayed novel TARs that were exclusively found on the MAS platform (MAS-B), 66.7% (26/39) of the novel TARs that were exclusively found on the Affymetrix platform, and 85% (17/20) of the assessed novel TARs that were common to both. In total, 66.3% of all assessed novel TARs were verified.

Forty-three known genes were also validated. Genes that were completely off (i.e., none of their splice variants were considered transcribed) according to one of the platforms but not the other were assessed. In total 58.8% (10/17) of the MAS-B exclusive genes were verified, and 87.5% (7/8) of the Affymetrix exclusive genes were verified. For genes that were considered "off" in both platforms, 33.3% (6/18) were found in our experimental validation.

DISCUSSION

In this work we have attempted to assess the suitability of two oligonucleotide tiling microarray strategies for transcription mapping in human. We tried to overcome the inherent differences between the approaches through using the same biological samples and a unified scoring and TAR generation procedure, and we have produced, compared, and validated several sets of

transcribed regions. We conclude that many factors are significant for the outcome of the experiments. Here, we elaborate on some key findings.

Arrays are noisy

In the comparison between the two microarray tiling platforms, the Affymetrix platform yielded TARs that better agreed with the Gencode annotation (Figures 2d-e, 4, S3). A simple explanation for this would be a higher noise level for MAS arrays. Moreover, given that the Pearson's correlation coefficients between raw intensities of technical replicates of MAS-N and Affymetrix arrays (Table 2; see also Table S1) are similar, it is likely that the MAS-N noise was rather systematic than random, while the MAS-B data seem to have a larger component of random noise. The systematic noise hypothesis is further supported by the observation that the overlap of TARs is larger within platform than within biological sample (Figure 3). The noise could result, e.g., from probe sequence artifacts, sample contamination (after it was split into different aliquots for the experiments), sub-optimal hybridization parameters, or protocol-dependent labeling artifacts (Nazarenko et al, 2002). A related issue is the cross-hybridization. For a transcription experiment, the amount of different RNA species present in the sample is large. Since the target RNA is derived from the entire genome, cross-hybridization is potentially present at high levels. Longer probes can be hybridized at higher temperatures and are thus less sensitive to cross-hybridization, but at a given temperature these probes are more susceptible to non-specific binding.

A comparison of the NB4 total RNA and placenta poly(A)+ TAR sets within the two platforms (Tables 2-4 and 6, Figure 3) revealed that the agreement between the platforms, and between the results of each platform and annotation, is larger for the poly(A)+ sets. One possible reason is that for total RNA, introns may be labeled. This would explain the worse performance of the NB4 sets compared to placenta, but not the differences in performance between MAS and Affy arrays. The exon and intron score distributions (Figure S7) show that for both placenta poly(A)+ and NB4 total RNA, exons in general have higher scores than introns. For both MAS and Affy experiments there is a slight shift of the total RNA intron score distribution towards higher scores, as compared to poly(A)+ distribution, indicating the presence of intron labeling.

Counteract the noise: more data, appropriate scoring

Figures 2d-e show that using more replicates enables TAR sets that better agree with annotation. Figure 2c shows that the genomic probe density also is important – reducing the genomic density of the Affy array to 50% (excluding every other probe on the arrays) worsens the performance, specifically for sensitivities below 40-45%. Furthermore, the density determines how well the endpoints of the TARs can be defined. The theoretical uncertainty of where a transcribed region starts and ends has an upper limit in the genomic distance between two adjacent probes. It also influences the results through the scoring procedure where often a genomic window is used for the statistical calculations. A window that is significantly larger than the average size of an exon is not desired since it would likely contain both probes that represent actual transcription (exons) and probes that belong to truly non-transcribed regions (introns). Taken together we conclude that the number of recorded data points per genomic unit is a crucial parameter for tiling microarray transcription mapping – the more data the better results.

To take advantage of the data, appropriate probe scoring procedures are needed. We tried several scoring schemes for our array data and found, in Figures 2d-e, that statistically based scoring using replicates and a genomic window can significantly improve the results compared to using single probe intensities. Using the standard sign test scoring was ultimately deemed the best way to score the MAS data (in particular the more noisy MAS-B data), while the best way to score the Affy data was to use the pseudomedian from the Wilcoxon signed rank test. For MAS-B data, using a Fwd-Rev scoring algorithm as a surrogate for a true PM-MM scoring, improved agreement with annotation for high segmentation thresholds in the maxgap/minrun algorithm (Figure 2b). Exploring the Affy data showed that the sign test did not perform particularly well using PM-only data as input but quite well using PM-MM (Figure 2c). In fact, the PM-only Affy data scored with the standard sign test (ie., identical scoring as the MAS sign test) resulted in a sensitivity/*PPV* behavior very similar to that of MAS – a relatively low agreement with annotation, and reduced impact of increasing the segmentation threshold (*PPV* insensitive to threshold increases). These results indicate that mismatches can be very useful. Altogether, we conclude that regardless of array design, statistically based scoring taking into account the available data in an appropriate way is indispensable in the analysis of tiling microarray data.

In Figure 2d we also analyzed the trade-off between increasing the genomic density of probes versus using the array space for MM probes. We found that in the investigated genomic density

and sensitivity ranges, it is better to use half of the array features for MM probes (one PM/MM probe pair every 42 nts) than to double the genomic density (to 21 nt) and use PM probes only. This is true for both scored (sign test; Figure 2c) and unscored (Figure 2d) data. From Figure 2c we also observe that using a single array PM-MM setup is actually preferable to using technical replicates of PM-only data for sensitivities up to 40% (retaining the genomic density and using sign test scoring). These findings suggest that true mismatch probes is a straightforward way to significantly improving the signal-to-noise ratio of oligonucleotide tiling arrays.

Conclusions from the array platform comparison

According to our study, the current form of the Affymetrix tiling microarray platform is better suited than the MAS platform for detailed transcription mapping of the human genome. This is true in the sense that the agreement of the TARs with known annotation is larger (Figure 2), and also in the sense that the exons in multiple-exon transcripts are more coherently transcribed (Table 6). From our study, we attribute this foremost to the higher genomic density of the probes and the presence of mismatch probes, and how these can be used to reduce the impact of non-specific hybridization. However, we cannot entirely exclude the effects of the differing labeling and hybridization protocols. On the other hand, the two technologies are almost equal in their ability to detect novel transcription as indicated by our experimental validation of novel TARs: in total 66.1% of the novel MAS-B and 72.9% of the novel Affymetrix placenta TARs are validated using RT-PCR (Table 7). TARs supported by both platforms are even more reliable and 85% of these are validated. The overlap of genes that are considered transcribed by the different platforms is substantial. Experimental validation of a subset of the genes considered transcribed by only one of the two platforms indicated that the Affymetrix setup is ahead of MAS in this respect as well, although the sample sizes are relatively small. It is also clear that if a gene is not detected by either platform, it is less likely to actually be transcribed. Our validation study shows that the two technologies are complementary, since much transcription detected by only one array platform is in fact verified as transcribed. They also reinforce each other, in the sense that array-based transcriptional evidence (or lack thereof) from both platforms yields more reliable results.

While the results obtained from the Affy arrays agree better with the annotation and the validation results, the advantage of the MAS technology is that it allows for rapid manufacturing of customized designs and cost-effective production of small array series. Using true mismatches in

the MAS design may improve the results for MAS arrays as well, but there are currently no results publicly available. We conclude that oligonucleotide tiling microarrays are suitable to detect novel transcribed regions, and that the use of replicates and statistically based scoring schemes significantly improves the performance for all investigated oligonucleotide tiling microarray-based transcription mapping experiments

METHODS

Array designs

Affymetrix arrays

Arrays were designed and manufactured by Affymetrix, Inc., using a physical mask. Probes are 25 bp long with an average genomic spacing of 21 bp, and they cover one genomic strand with the exception of repeat regions, as defined by RepeatMasker (A.F.A. Smit and P. Green, unpublished). Each probe is present in a “perfect match” and a “mismatch” version. The mismatch probe contains a single substitution at the middle probe position (A->T, T->A, C->G, G->C). Each array contains in total approximately 1,400,000 features.

MAS arrays

Arrays were designed by us and manufactured by NASA using a NimbleGen maskless array synthesizer. Probes are 36 bp long with an average genomic spacing of 36 bp. Positional shifts were allowed to avoid self-complementarity at the probe ends (defined as at least 4 consecutive complementary nucleotides within the 6 5'/3' nucleotides). The probes cover both genomic strands with the exception of repeat regions. The design was done on the Ncbi v34 of the human genome build, and each array contains almost 390,000 features.

RNA extraction and array hybridization

Cell culture

The human NB4 cells were cultured in RPMI medium containing 20mM L-glutamine (Media Tech) and supplemented with 10% fetal bovine serum (Invitrogen), 100 IU/mL penicillin (Media Tech) and 100µg/mL Streptomycin (Media Tech). Cells were maintained at 37^0C under 5% CO_2/95% air in a humidified incubator.

RNA samples

Total RNA from the human NB4 cells was extracted using Qiagen RNA extraction kit according to the manufacturer's instructions. Human placental poly (A)+ mRNA (obtained from total RNA) was purchased from Ambion (Austin, TX).

Protocols

The Supplementary Material contains a detailed description of all three experimental protocols (MAS-B, MAS-N, Affy). The MAS-N protocol yields in-vitro transcribed, biotin-labeled, single-stranded cRNA, fragmented to an average size of 50-200 bp before hybridization. The MAS-B protocol yields Cy3-aminoallyl-labeled unfragmented single-stranded cDNA. The Affymetrix protocol yields end-labeled (bio-ddATP) double-stranded cDNA, fragmented to an average size of 50-100 bp before hybridization.

Scoring schemes

To obtain the desired statistical resolution, MAS array scoring was done pooling the data from all three biological samples (for both placenta and NB4). For placenta, this corresponds to seven measurements for each probe, and six for NB4. For Affymetrix, three technical replicates were used, corresponding to six measurements for each probe (three PM and three MM probes).

Sign test using array median intensity

The intensity of every probe within the window is compared to the median intensity of the slide and assigned a '1' if it is above and '0' otherwise. The number of ones within the window is

counted and the probability *p* of finding at least this number of '1's under the null hypothesis that half of the probes should be above the median, is calculated. The score assigned to the probe in the middle of the window is then defined as score = -log(*p*). Window sizes of 90-240 bp were tried, choosing 160 (5 probes) for the MAS data in this study. No inter-array normalization is performed since each intensity is compared to the median intensity on its own array only. A variant of this scoring approach is to weight the probes within the window differently, such that the central probe(s) becomes more important. For instance, the intensities within the window can be multiplied with a discretized gaussian envelope. Several parameter settings were tried (Figures S2-S5).

Paired Wilcoxon signed rank sum test

Inter-array normalization is undertaken through dividing each intensity with the array median (median normalization). Within a window, all pairwise differences between the intensities of a perfect match probe and its corresponding mismatch probe are calculated and ranked. A sign is assigned to each rank number depending on whether the PM or the MM intensity was greater, and a p-value is calculated from the sum of this signed ranking (keeping track of the rank sum of all negative ranks and the rank sum of all positive ranks). The p-value, which is a measure of how significantly the distribution of PM-MM differences is skewed to either side around zero, can then be used to compute the final score for the probe in the middle of the window (Kampa et al, 2004; Royce et al., 2005). The corresponding point estimate, the pseudomedian, is obtained by taking the median value of all the pairwise averages of PM-MM values within the window. The Affymetrix scores used in this study were calculated by Affymetrix using a window size of 101 nucleotides, corresponding to on average 5 probes in the window. For MAS arrays, the paired Wilcoxon signed rank test was applied using the probe corresponding to the reverse strand of the exact same genomic locus as mismatch probe (instead of a designed mismatch probe).

Segmentation of genomic regions

Maxgap/minrun segmentation

The transcribed regions were generated from scored data. The maxgap parameter was set to 50 for Affymetrix data and 80 for MAS data. The minrun parameter was set to 50 for both

approaches,. Other maxgap/minrun parameter settings were also tested (data not shown). We evaluated segmentation thresholds of 70-99th percentile.

HMM segmentation

The emission and transition probability distributions of the four-state HMM for each dataset were learned according to the scores of those probes which fall into known gene regions, where the score characteristics in the exon regions were used to estimate the parameters for the TAR state, and those in the intron regions for the non-TAR state. The parameters for the two intermediate transition states were obtained by investigating those probes containing both exon and intron regions. These emission distributions were fitted with mixed-Gaussian distributions to generate a continuous model. The Viterbi algorithm was utilized to identify TARs.

Assessing Transcription of Annotated Genes

The transcription status was assessed using the sign test as described for all annotated splice variants of all known genes in the Gencode annotation of regions ENm001-ENm011, accepting the exons with labels "VEGA_known", "VEGA_Novel_CDS", "VEGA_Novel_transcript_gencode_conf", and "VEGA_Putative_gencode_conf". For the exon/intron-based investigations (Table 6, Figure S7), the median probe score for each feature was used, with a percentile threshold for on/off calls as defined for each experiment in Table 3.

Choosing primer pairs for validation

Primer pairs were generated using Primer3 (Rozen and Skaletsky, 2000). Primers assessing novel TARs were required to define a genomic region with no overlap with any Gencode gene. When assessing known genes, the exon with the highest p-value based transcription score was chosen. Primer3 settings were as default or more stringent, e.g., GC content within 35-65%, primer size was forced to be between 20 and 28 nucleotides, and the resulting PCR products to be between 100 and 200 bp. Validation candidates were checked using UCSC In Silico PCR (http://genome.ucsc.edu/cgi-bin/hgPcr) against the Ncbi v35 human genome build to ensure that exactly one PCR product was possible; those that generated no or multiple hits were discarded.

Three regions that did not contain any verified or predicted transcription were chosen to act as negative controls. The experimental protocol of the PCR validation is in Supplementary Material.

Accessing data and results

The MAS ENCODE array platform has GEO (Gene Omnibus Expression, http://www.ncbi.nlm.nih.gov/geo/) accession number GPL2105; the corresponding data series has GEO accession number GSE2720 (placenta and untreated NB4). The Affymetrix anti-sense ENCODE array platform has GEO accession number GPL1789; the corresponding data series has accession number GSE2671 (placenta) and GSE2679 (untreated NB4). The TAR sets, the gene/transcript/exon transcription status, the validation results, and the raw data are available at or from http://tiling.gersteinlab.org/platformcmp.

ACKNOWLEDGEMENTS

O.E. is supported by a Knut and Alice Wallenberg Foundation postdoctoral fellowship. We acknowledge support from the NIH (1U01HG003156-01).

FIGURE LEGENDS

Figure 1:

Comparing human chromosome 22 transcription data sets with gene annotation. Transcription data sets were derived from previously published studies. They were generated from three different microarray platforms: PCR (red squares), MAS (blue diamond), and Affymetrix (green circles); in total 15 separate experiments (tissues or cell lines), each represented by a point in the figure . We used the RefSeq annotation as a benchmark to assess the quality of the data from each experiment. *X-axis*, the fraction of exonic probes that were identified to be transcribed in individual experiments (sensitivity). *Y-axis*, the fraction of transcribed probes overlapping with an exon (specificity). The PCR tiling array data were from placenta, fibroblast and B-cells (Rinn et al., 2003; White et al., 2004), the MAS data from liver (Bertone et al., 2004), and the Affymetrix sets were collected from Kapranov et al. (2002) representing 11 different cell lines. *Arrow*, the Affymetrix data from U87 cell line is not representative since a long section of chromosome 22 is identified as transcriptionally silent, suggesting this particular experiment probably did not work or something unusual about U87.

Figure 2:

(**A**) Number of nucleotides in placental TARs as a function of segmentation threshold (percentiles). TARs were generated with the maxgap/minrun algorithm based on the scored hybridization intensity data using a genomic window and technical replicates: MAS-B scored with standard sign test (*green*); MAS Fwd-Rev scoring using reverse strand as "mismatch" (*orange*), pseudomedian; Affymetrix scored using pseudomedian from PM-MM (*blue*). The data points corresponding to the data sets used in the Comparison section are circled: thresholds are 90^{th} percentile for Affy and 91^{th} percentile for MAS-B (sign test scoring). *X-axis*, the percentile score threshold for calling a probe "positive". *Y-axis*, the number of nucleotides in TARs (in megabasepairs). The dashed line corresponds to the number of nucleotides in exons in the analyzed region (1,001,238 nts). (**B**) Positive predictive value (*PPV*) versus sensitivity for three different ways of scoring and segmenting the MAS-B data, varying the segmentation threshold from 70^{th} percentile (to the right in the figure) to 99^{th} percentile (to the left) for the MAS-B set

scored with standard sign test (*green*); scored using reverse strand as "mismatch" (Fwd-Rev scoring, *orange*); and the result from HMM segmentation (Viterbi decoding) of sign test-scored data (*grey triangle*). *Sensitivity* (*x*-axis), defined as the percentage of bases in Gencode exonic regions that are covered by a TAR. *PPV* (*y*-axis), defined as the percentage of bases in the TARs that overlap with a Gencode exonic region. (**C**) *PPV* versus sensitivity for two different ways of scoring the placenta Affy data, using 3 replicates (6 array features) unless otherwise stated: Wilcoxon signed rank test (*blue circles*), and standard sign test (using PM-MM values: *cyan triangles*; using PM-only values: *yellow squares*). The result from reducing the genomic density of the Affy array to 50% (ie., removing the data from every second probe) is also shown, using PM-MM values (3 replicates: *cyan triangles, dashed line*; single replicate only: *grey triangles, dashed line*). (**D**) *PPV* versus sensitivity for MAS-B and Affy placenta data, varying the segmentation threshold from 70^{th} percentile (to the right in the figure) to 99^{th} percentile (to the left). The average results of TARs generated from raw intensities from single arrays for Affy (PM only (*blue squares*), and PM-MM (*blue triangles; solid line* actual genomic density, *dashed line* 50% genomic density)) and MAS-B (*green squares*) are plotted, as well as scored results for Affy (*blue circles*) and MAS-B (*green circles*). *Sensitivity* (*x*-axis), and *PPV* (*y*-axis), defined as above. The data points corresponding to the data sets used in the Comparison section are circled. The hatched area marks where a sensitivity of 30% is achieved for the various sets. (**E**) *PPV* for placental TAR sets when choosing a segmentation threshold that yields approximately 30% sensitivity (hatched area in (B)). Note that the actual sensitivity varies slightly between the sets.

Figure 3:

TAR set agreement. (**A**) Overlap of TAR sets, measured in number of overlapping nucleotides (kilobases). All three placenta TAR sets (MAS-B, MAS-N, Affy) and both NB4 TAR sets (MAS-N and Affy) . *R* is a measure of the size of the overlap. $R = |\cap| / |U|$ (calculated pairwise for the three placenta TAR sets). (**B)** Size of TAR set overlap, expressed in *R*, for comparisons within biological samples but across different array platforms (*black* lines), and comparisons within array platforms but across the biological samples (*brown* lines). Values in leftmost column of the graph are calculated with no further constraints. Second column, only TARs overlapping with conserved regions are included. Third column, only TARs overlapping with Gencode exons are included. Fourth column, only TARs overlapping with both conserved and exon regions included.

Figure 4:

Distribution of Gencode exon coverage by placenta TARs: all exons (MAS-B, g*reen squares,* and Affy, *blue squares*); 5' exons (Affy, *blue circles*); 3' exons (Affy, *blue triangles*). *X-axis*, the fraction to which an exon is covered by a TAR; 0.0-1.0 split up in 10 bins. *Y-axis*, the percentage of exons covered by a TAR to the fraction represented on the x-axis.

TABLES

Table 1

Outline of hybridization experiments.

--Experiment ID--		Sample	Number of technical replicates batch1+batch2+batch3	Hybridization protocol
Placenta	MAS-B	Poly(A)+	3+2+2	Bertone et al. (2004)
Placenta	MAS-N	Poly(A)+	3+2+2	derived from manufacturer's
Placenta	Affy	Poly(A)+	3+2+2	manufacturer's
NB4	MAS-N	total RNA	2+2+2	derived from manufacturer's
NB4	Affy	total RNA	2+2+2	manufacturer's

Table 2

Correlation of hybridization intensities. Average of absolute values of Pearson correlation coefficients (R), calculated from unprocessed hybridization intensities (excluding internal standards and grid alignment probes). Between arrays: between technical and between biological replicates. Within arrays: for Affymetrix arrays, between corresponding perfect match (PM) and mismatch (MM) values; for MAS arrays, between probes representing forward/leading (Fwd) and reverse/lagging (Rev) strands of the same genomic location.

--Experiment ID--		---------- Between arrays ----------		---------- Within arrays ----------	
		techn. repl.	biol. repl.	PM vs MM	Fwd vs Rev
Placenta	MAS-B	0.829	0.820	-	0.627
Placenta	MAS-N	0.955	0.953	-	0.046
Placenta	Affy	0.961	0.937	0.774	-
NB4	MAS-N	0.959	0.957	-	0.045
NB4	Affy	0.981	0.983	0.917	-

Table 3
Characteristics of TAR sets used in comparison (data for ENCODE regions ENm001-ENm011)

--Experiment ID--		Scoring method and Segmentation parameters		--Number of TARs and nucleotides-- ----Stranded------		---Unstranded---		Mean/ Median	Gencode cmp. Sens.	PPV
		threshold/minrun/maxgap		#TARs	#bases	#TARs	#bases	length	(%)	(%)
Placenta	MAS-B	Sign test win.160	91/50/80	4079	955k	2545	684k	269/180	24.6	35.9
	MAS-N	Sign test win.160	92/50/80	3853	768k	3248	701k	216/144	22.3	31.7
	Affy	PM-MM P-median	90/50/50	-	-	3694	629k	170/105	37.0	58.6
NB4	MAS-N	Sign test win.160	93/50/80	3520	697k	2936	632k	216/144	19.1	30.2
	Affy	PM-MM P-median	87/50/50	-	-	4674	629k	135/91	26.5	41.8
Gencode exon regions				2563	1018k	2482	1001k	403		

Table 4

Percentage of genic and intergenic TARs that overlap with conserved regions (>90% of TAR length within conserved region) or that do not overlap with conserved regions (<10% of TAR length within conserved region).

				Intergenic TARs		Genic TARs
			Overlaps conserved region:			
--Experiment ID--			**NO**	**YES**	**NO**	**YES**
Placenta	MAS-B		84%	2%	50%	8%
Placenta	Affy		77%	8%	41%	24%
NB4	MAS-N		85%	1%	54%	8%
NB4	Affy		79%	7%	49%	19%

Table 5

Transcribed placental genes (and in brackets: transcripts) in MAS-B and Affy experiments. A gene is considered as transcribed if at least one of its transcripts (splice variants) is transcribed at significance level $p<0.001$, using the sign test to score all transcripts. Probes that to at least 50% are within an exon of the transcript are considered. If a transcript/gene has less than 10 probes it will be unable to reach a p-value below 0.001, and is in the "Too few probes" category. Total number of genes (transcript) is 264 (1342).

		Affy		
		Yes	**No**	**Too few probes**
	Yes	158 (871)	37 (155)	15 (58)
MAS-B	**No**	9 (13)	24 (61)	7 (13)
	Too few probes	2 (51)	7 (36)	5 (44)

Table 6

Multi-exon coherence of transcripts with more than one exon. Percentage of transcripts where *all*, *some*, or *no* exons are considered transcribed according to median intensity of each exon. Exons were called on/off based on their median intensity compared with the experiment-specific score thresholds used for TAR generation (segmentation), specified in Table 3. The transcription status of entire transcripts was generated as in Table 5.

--Experiment ID--		**All exons**	**Some exons**	**No exons**
All transcripts (1298 transcripts):				
Placenta	MAS-B	6.3%	60.7%	33.0%
Placenta	MAS-N	3.4%	56.8%	39.8%
Placenta	Affy	36.8%	47.6%	15.6%
NB4	MAS-N	1.9%	51.7%	46.4%
NB4	Affy	15.5%	60.9%	23.7%
Intersection of placenta MAS-B and Affy transcribed transcripts (869):				
Placenta	MAS-B	7.8%	68.9%	23.2%
Placenta	Affy	50.1%	48.3%	1.6%
Intersection of NB4 MAS-N and Affy transcribed transcripts (543):				
NB4	MAS-N	2.9%	72.9%	24.1%
NB4	Affy	26.9%	61.9%	10.7%

Table 7

Results of experimental validation (reverse transcriptase PCR) in placenta of 144 regions: 98 novel TARs, 43 exons from known genes, 3 negative controls.

Set	Transcription status		Overlap with known exon?	Number assessed	Number positive
	MAS-B	Affy			
TAR	on	off	no	39	22 (56.4%)
TAR	off	on	no	39	26 (66.7%)
TAR	on	on	no	20	17 (85%)
Exons	on	off	yes	17	10 (58.8%)
Exons	off	on	yes	8	7 (87.5%)
Exons	off	off	yes	18	6 (33.3%)
Neg ctrl	off	off	no	3	1 (33.3%)

REFERENCES

Albert, T.J., Norton, J., Ott, M., Richmond, T., Nuwaysir, K., Nuwaysir, E.F., Stengele, K.P., and Green, R.D. 2003. Light-directed 5'-->3' synthesis of complex oligonucleotide microarrays. *Nucl. Acids Res.* **31:** e35

Ashurst, J.L., Chen, C.K., Gilbert, J.G., Jekosch, K., Keenan, S., Meidl, P., Searle, S.M., Stalker, J., Storey, R., Trevanion, S., Wilming, L., and Hubbard, T. 2005. The Vertebrate Genome Annotation (Vega) database. *Nucl. Acids Res.* **33 (Database issue):** D459-465

Bertone, P., Stolc, V., Royce, T.E., Rozowsky, J.S., Urban, A.E., Zhu, X., Rinn, J.L., Tongprasit, W., Samanta, M., Weissman, S., Gerstein, M., and Snyder, M. 2004. Global identification of human transcribed sequences with genome tiling arrays. *Science* **306:** 2242-2246

Bertone, P., Trifonov, V., Rozowsky, J.S., Schubert, F., Emanuelsson, O., Karro, J., Kao, M.-Y., Snyder, M., and Gerstein, M. 2006. Design Optimization Methods for Genomic DNA Tiling Arrays. *Genome Res.* accepted for publication. *Genome Res.* **16:** 271-281

Blanchette, M., Kent, W.J., Riemer, C., Elnitski, L., Smit, A.F.A., Roskin, K.M., Baertsch, R., Rosenbloom, K., Clawson, H., Green, E.D., Haussler, D., and Miller, W. 2004. Aligning Multiple Genomic Sequences With the Threaded Blockset Aligner. *Genome Res.* **14:** 708-715

Brudno, M., Do, C., Cooper, G., Kim, M.F., Davydov, E., Green, E.D., Sidow, A., and Batzoglou, S. 2003. LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. *Genome Res.* **13:**721-731

Burge, C., and Karlin, S. 1997. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268:** 78--94

Cawley, S., Bekiranov, S., Ng, H.H., Kapranov, P., Sekinger, E.A., Kampa, D., Piccolboni, A., Sementchenko, V., Cheng, J., Williams, A.J., et al. 2004. Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell* **116:** 499-509

Chen, J., Sun, M., Kent, W.J., Huang, X., Xie, H., Wang, W., Zhou, G., Shi, R.Z., and Rowley, J.D. 2004. Over 20% of human transcripts might form sense-antisense pairs. *Nucl. Acids Res.* **32:** 4812-4820

Cheng, J., Kapranov, P., Drenkow, J., Dike, S., Brubaker, S., Patel, S., Long, J., Stern, D., Tammana, H., Helt, G., et al. 2005. Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science* **308:** 1149-1154

The ENCODE Project Consortium. 2004. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **306:** 636-640

Guigo, R., Dermitzakis, E.T., Agarwal, P., Ponting, C.P., Parra, G., Reymond, A., Abril, J.F., Keibler, E., Lyle, R., Ucla, C., Antonarakis, S.E., and Brent, M.R. 2003. Comparison of mouse and human genomes followed by experimental verification yields an estimated 1,019 additional genes. *Proc. Natl. Acad. Sci.* **100:** 1140-1145

Horak, C., and Snyder, M. 2002. ChIP-chip: a genomic approach for identifying transcription factor binding sites. *Methods Enzymol.* **350:** 469-83

Hughes, T.R., Mao, M., Jones, A.R., Burchard, J., Marton, M.J., Shannon, K.W., Lefkowitz, S.M., Ziman, M., Schelter, J.M., Meyer, M.R., et al. 2001. Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat. Biotechnol.* **19:** 342-347

Jarvinen, A.-K., Hautaniemi, S., Edgren, H., Auvinen, P., Saarela, J., Kallioniemi, O.P., and Monni, O. 2004. Are data from different gene expression microarray platforms comparable? *Genomics* **83:** 1164-1168

Ji, H., and Wong, W.H. 2005. TileMap: create chromosomal map of tiling array hybridizations. *Bioinformatics* **21:** 3629-3636

Kampa, D., Cheng, J., Kapranov, P., Yamanaka, M., Brubaker, S., Cawley, S., Drenkow, J., Piccolboni, A., Bekiranov, S., Helt, G., et al. 2004. Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. *Genome Res.* **14:** 331-342

Kapranov, P., Cawley, S.E., Drenkow, J., Bekiranov, S., Strausberg, R.L., Fodor, S.P., and Gingeras, T.R. 2002. Large-scale transcriptional activity in chromosomes 21 and 22. *Science* **296:** 916-919

Kapranov, P., Drenkow, J., Cheng, J., Long, J., Helt, G., Dike, S., and Gingeras, T.R. 2005. Examples of the complex architecture of the human transcriptome revealed by RACE and high-density tiling arrays. *Genome Res.* **15:** 987-997

Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yap ,C.C., Suzuki, M., Kawai, J., et al. 2005. Antisense transcription in the mammalian transcriptome. *Science* **309:** 1564-1566

Kim, K., Page, G.P., Beasley, T.M., Barnes, S., Scheirer, K.E., and Allison, D.B. 2006. A proposed meetric for assessing the measurement quality of individual microarrays. *BMC Bioinformatics* **7**: e35

Larkin, J.E., Frank, B.C., Gavras, H., Sultana, R., and Quackenbush, J. 2005. Independence and reproducibility across microarray platforms. *Nat. Methods* **2:** 337-343

Li, W., Meyer, C.A., and Liu, X.S. 2005. A hidden Markov model for analyzing ChIP-chip experiments on genome tiling arrays and its application to p53 binding sequences. *Bioinformatics* **21 (suppl. 1):** i274-i282

Lipshutz, R.J., Fodor, S.P., Gingeras, T.R., and Lockhart, D.J. 1999. High density synthetic oligonucleotide arrays. *Nat. Genet.* **21:** 20-24

Mah, N., Thelin, A., Lu, T., Nikolaus, S., Kuhbacher, T., Gurbuz, Y., Eickhoff, H., Kloppel, G., Lehrach, H., Mellgard, B., Costello, C.M., and Schreiber, S. 2004. A comparison of oligonucleotide and cDNA-based microarray systems. *Physiol. Genomics* **16:** 361-370

Mathews, D.H., Burkard, M.E., Freier, S.M., Wyatt, J.R., and Turner, D.H. 1999. Predicting oligonucleotide affinity to nucleic acid targets. *RNA* **5:** 1458-1469

Mockler, T.C., Chan, S., Sundaresan, A., Chen, H., Jacobsen, S.E., and Ecker, J.R. 2005. Applications of DNA tiling arrays for whole-genome analysis. *Genomics* **85:** 1-15

Nazarenko, I., Pires, R., Lowe, B., Obaidy, M., and Rashtchian, A. {Effect of primary and secondary structure of oligodeoxyribonucleotides on the fluorescent properties of conjugated dyes. 2002. *Nucl. Acids Res.* **30:** 2089-2195

Nuwaysir, E.F., Huang, W., Albert, T.J., Singh, J., Nuwaysir, K., Pitas, A., Richmond, T., Gorski, T., Berg, J.P., Ballin, J., et al. 2002. Gene expression analysis using oligonucleotide arrays produced by maskless photolithography. *Genome Res.* **12:** 1749-1755

Park, P.J., Cao, Y.A., Lee, S.Y., Kim, J.W., Chang, M.S., Hart, R., and Choi, S. 2004. Current issues for DNA microarrays: platform comparison, double linear amplification, and universal RNA reference. *J. Biotechnol.* **112:** 225-245

Pruitt, K.D., Tatusova, T., and Maglott, D.R. 2005. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucl. Acids Res.* **33:** D501-D504

Rinn, J.L., Euskirchen, G., Bertone, P., Martone, R., Luscombe, N.M., Hartman, S., Harrison, P.M., Nelson, F.K., Miller, P., Gerstein, M., Weissman, S., and Snyder, M. 2003. The transcriptional activity of human Chromosome 22. *Genes Dev.* **17:** 529-540

Rouillard, J.M., Zuker, M., and Gulari, E. 2003. OligoArray 2.0: design of oligonucleotide probes for DNA microarrays using a thermodynamic approach. *Nucl. Acids Res.* **31:** 3057-3062

Royce, T.E., Rozowsky, J.S., Bertone, P., Samanta, M., Stolc, V., Weissman, S., Snyder, M., and Gerstein, M. 2005. Issues in the analysis of oligonucleotide tiling microarrays for transcript mapping. *Trends Genet* . **21:** 466-475

Royce, T.E., Rozowsky, J.S., Luscombe, N.M., Emanuelsson, O., Yu, H., Zhu, X., Snyder, M., and Gerstein, M. 2006. Extrapolating traditional DNA microarray statistics to the tiling and protein microarrays technologies. *Methods Enzymol.* accepted for publication

Rozen, S., and Skaletsky, H.J. 2000. Primer3 on the WWW for general users and for biologist programmers. In *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (eds. S. Krawetz and S. Misener), pp. 365-386. Humana Press, Totowa, NJ, USA

Schadt, E.E., Edwards, S.W., GuhaThakurta, D., Holder, D., Ying, L., Svetnik, V., Leonardson, A., Hart, K.W., Russell, A., Li, G., et al. 2004. A comprehensive transcript index of the human genome generated using microarrays and computational approaches. *Genome Biol.* **5:** R73

Shoemaker, D.D., Schadt, E.E., Armour, C.D., He, Y.D., Garrett-Engele, P., McDonagh, P.D., Loerch, P.M., Leonardson, A., Lum, P.Y., Cavet, G., et al. 2001. Experimental annotation of the human genome using microarray technology. *Nature* **409:** 922-927

Singh-Gasson, S., Green, R.D., Yue, Y., Nelson, C., Blattner, F., Sussman, M.R., and Cerrina, F. 1999. Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array. *Nat. Biotechnol.* **17:** 974-978

Tan, P.K., Downey, T.J., Spitznagel, E.L. Jr., Xu, P., Fu, D., Dimitrov, D.S., Lempicki, R.A., Raaka, B.M., and Cam, M.C. 2003. Evaluation of gene expression measurements from commercial microarray platforms. *Nucl. Acids Res.* **31:** 5676-5684

Van Gelder, R.N., von Zastrow, M.E., Yool, A., Dement, W.C., Barchas, J.D., and Eberwine, J.H. 1990. Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc. Natl. Acad. Sci. USA* **87:** 1663-1667

White, E.J., Emanuelsson, O., Scalzo, D., Royce, T., Kosak, S., Oakeley, E.J., Weissman, S., Gerstein, M., Groudine, M., Snyder, M., and Schübeler, D. 2004. DNA replication-timing analysis of human chromosome 22 at high resolution and different developmental states. *Proc. Natl. Acad. Sci. USA* **101**: 17771-17776

Yamada, K., Lim, J., Dale, J.M., Chen, H., Shinn, P., Palm, C.J., Southwick, A.M., Wu, H.C., Kim, C., Nguyen, M., et al. 2003. Empirical analysis of transcriptional activity in the Arabidopsis genome. *Science* **302:** 842-846

Yauk, C.L., Berndt, M.L., Williams, A., and Douglas, G.R. 2004. Comprehensive comparison of six microarray technologies. *Nucl. Adicds Res.* **32:** e124

Yelin, R., Dahary, D., Sorek, R., Levanon, E.Y., Goldstein, O., Shoshan, A., Diber, A., Biton, S., Tamir, Y., Khosravi, R., et al. 2003. Widespread occurrence of antisense transcription in the human genome. *Nat. Biotechnol.* **21:** 379-386

Ying, L., Schadt, E.E., Holder, S.V.D., Edwards, S., and Guhathakurtka, D. 2003. Identification of chromosomal regions containing transcribed sequences using microarray expression data. In *2003 Proc. of the American Statistical Association*, 4672-4677

FIGURES

Figure 1

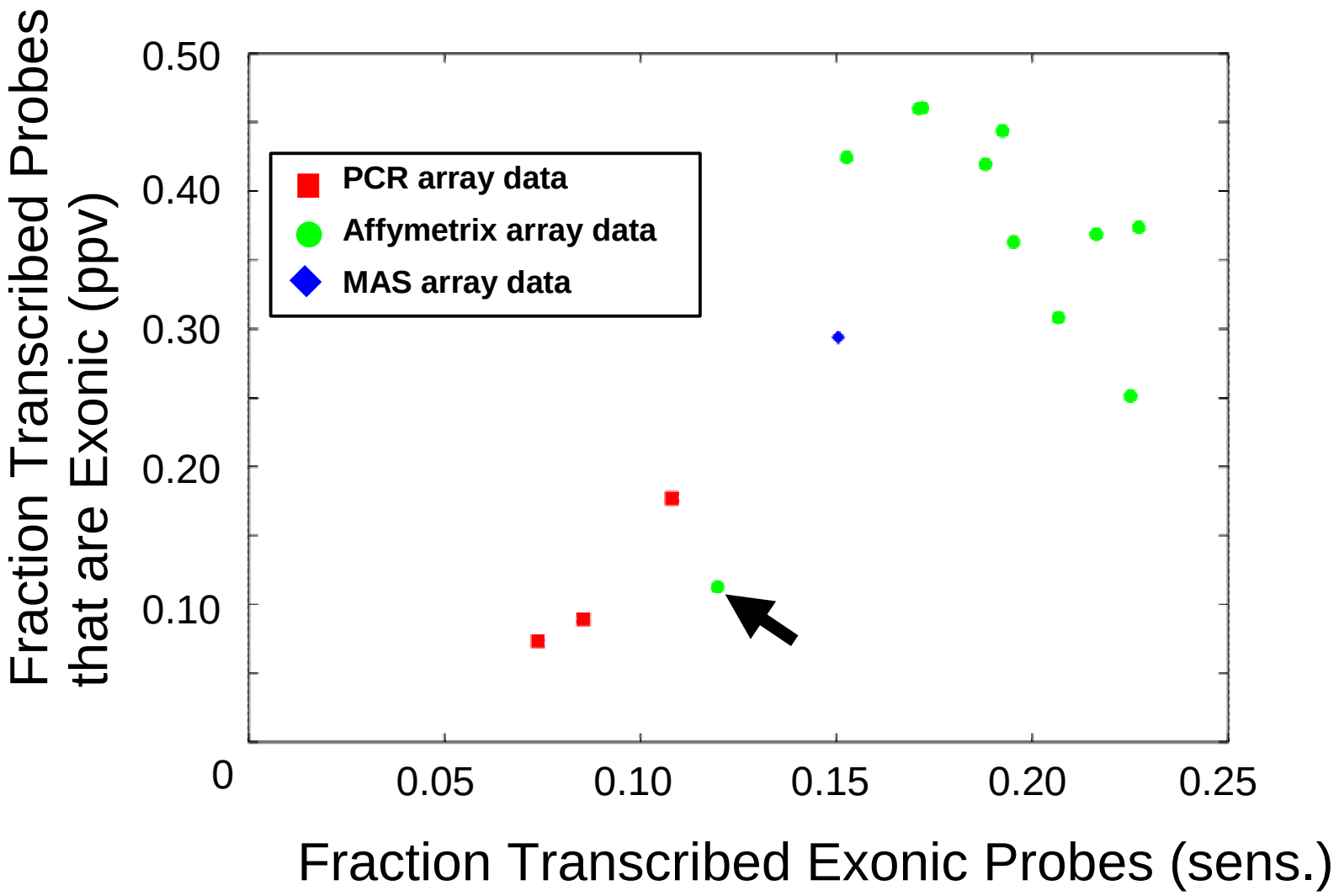

Fraction Transcribed Probes that are Exonic (ppv)
0.50
0.40
0.30
0.20
0.10
0
PCR array data
Affymetrix array data
MAS array data
0.05
0.10
0.15
0.20
0.25
Fraction Transcribed Exonic Probes (sens.)

Figure 2 A

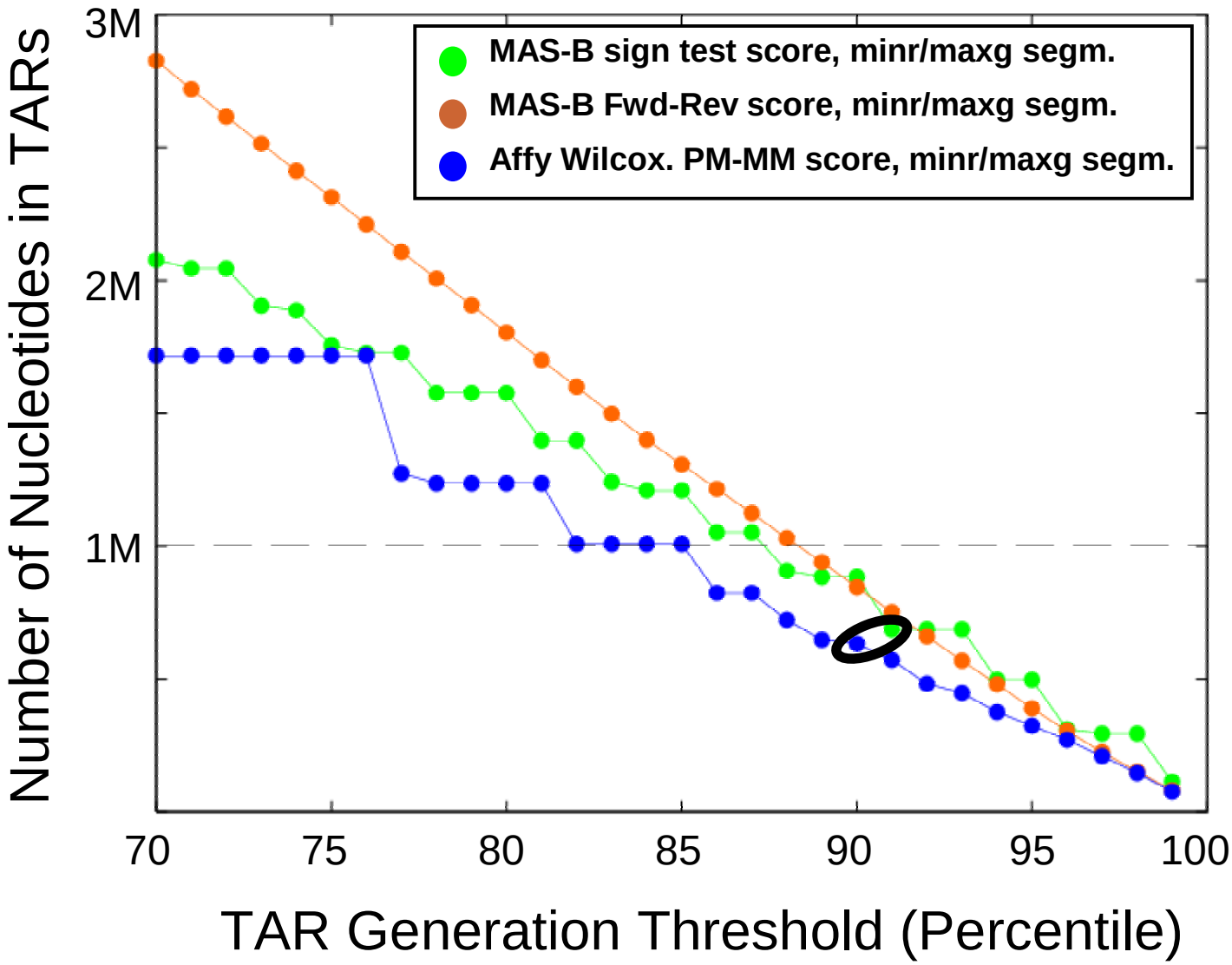

MAS-B sign test score, minr/maxg segm.
MAS-B Fwd-Rev score, minr/maxg segm.
Affy Wilcox. PM-MM score, minr/maxg segm.
Number of Nucleotides in TARs
3M
2M
1M
70
75
80
85
90
95
100
TAR Generation Threshold (Percentile)

Figure 2 B

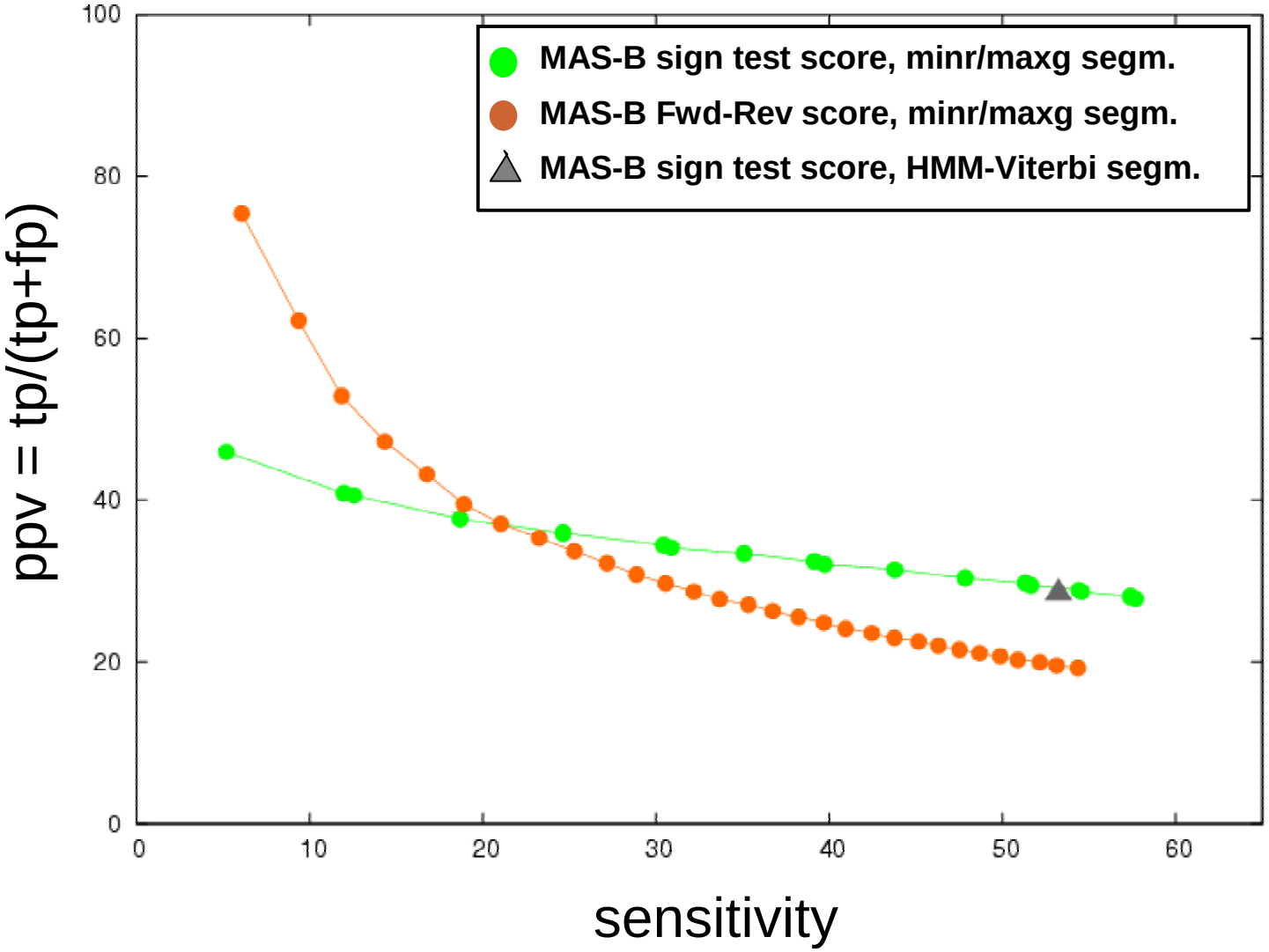

MAS-B sign test score, minr/maxg segm.
MAS-B Fwd-Rev score, minr/maxg segm.
MAS-B sign test score, HMM-Viterbi segm.
ppv = tp/(tp+fp)
sensitivity
0
20
40
60
80
100
0
10
20
30
40
50
60

Figure 2 C

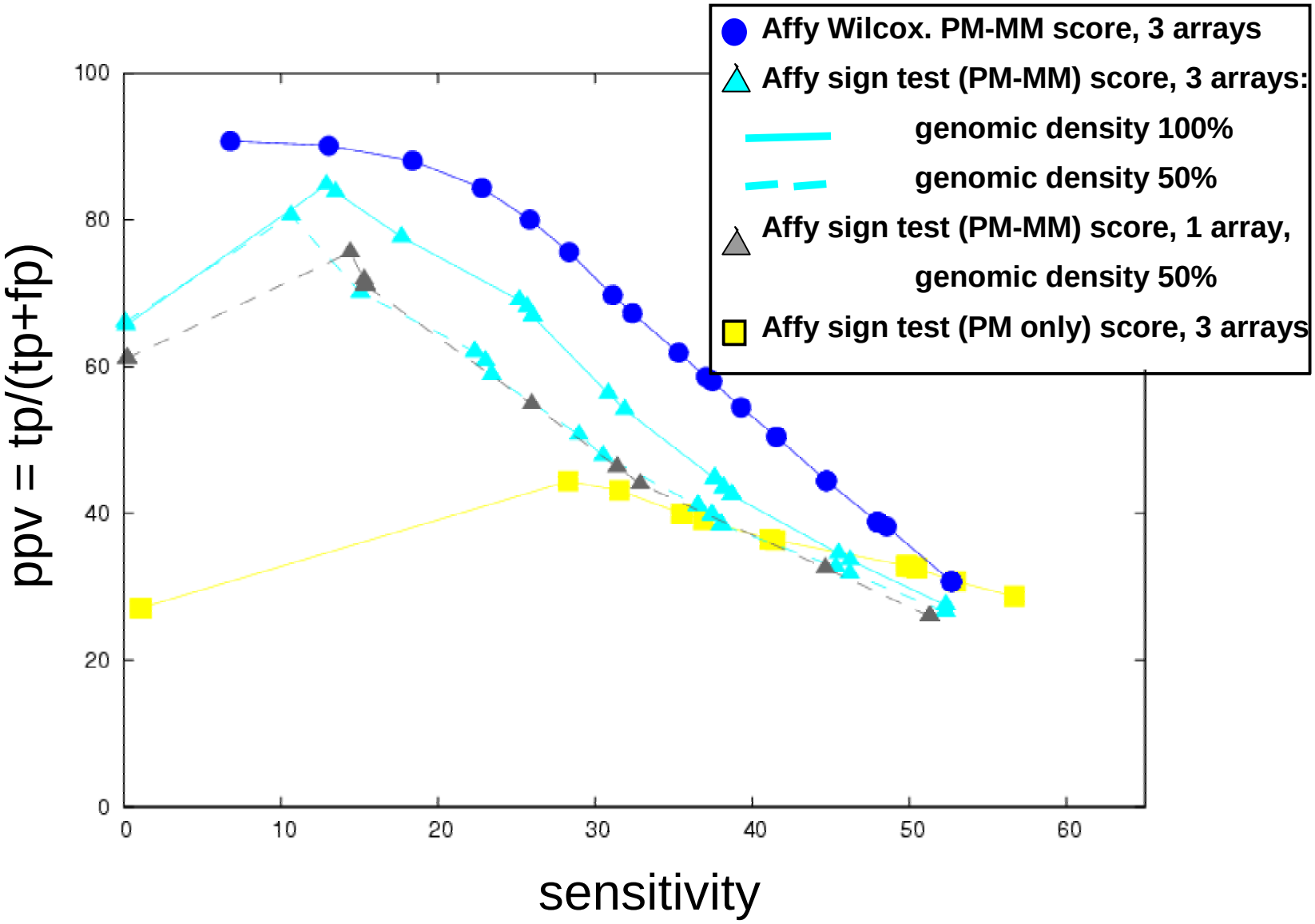

Affy Wilcox. PM-MM score, 3 arrays
Affy sign test (PM-MM) score, 3 arrays:
genomic density 100%
genomic density 50%
Affy sign test (PM-MM) score, 1 array,
genomic density 50%
Affy sign test (PM only) score, 3 arrays
ppv = tp/(tp+fp)
sensitivity
0
20
40
60
80
100
0
10
20
30
40
50
60

Figure 2 D

Emanuelsson et al.

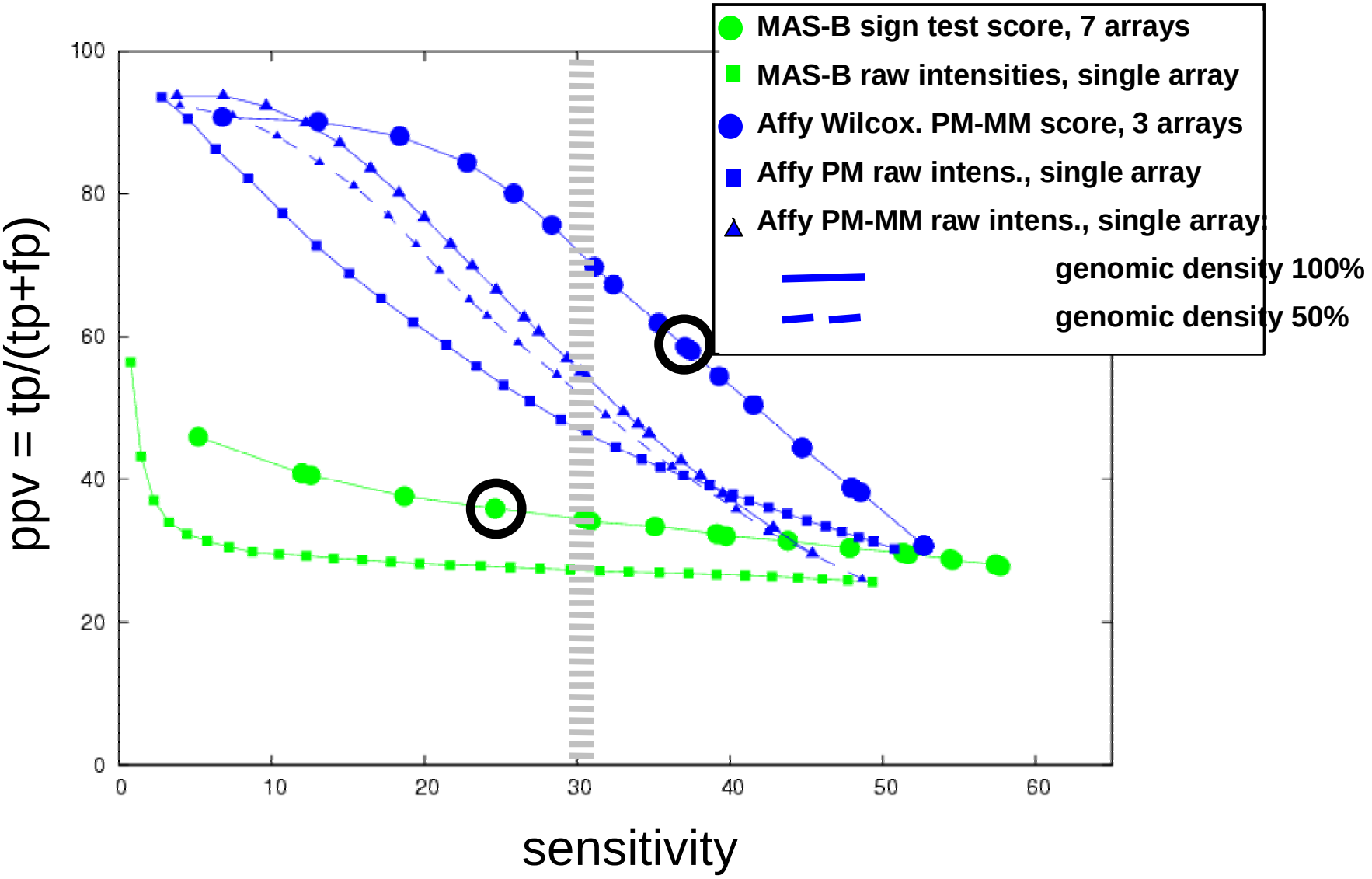
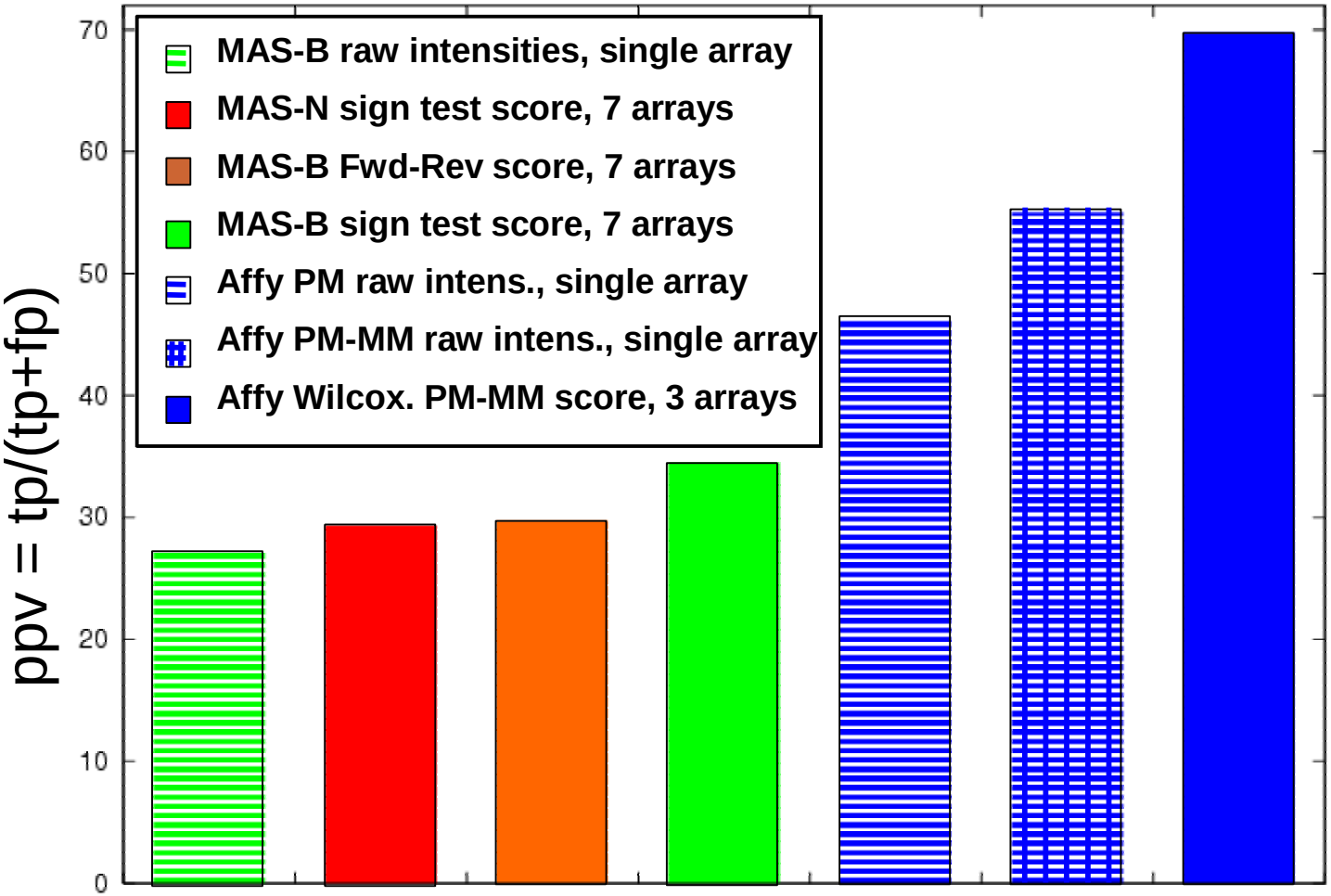


Figure 2 E



threshold: 81 89 88 89 85 85 93

sensitivity: 29.6 29.9 30.6 30.4 30.6 30.2 31.1

Figure 3 A

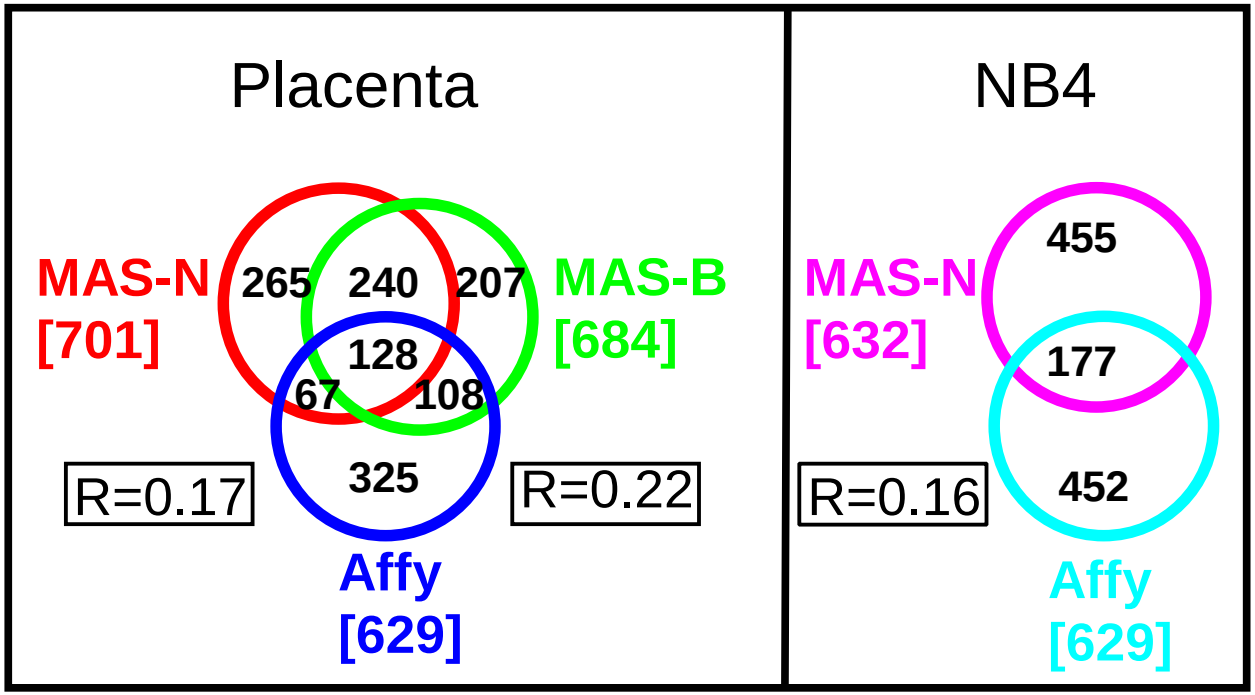
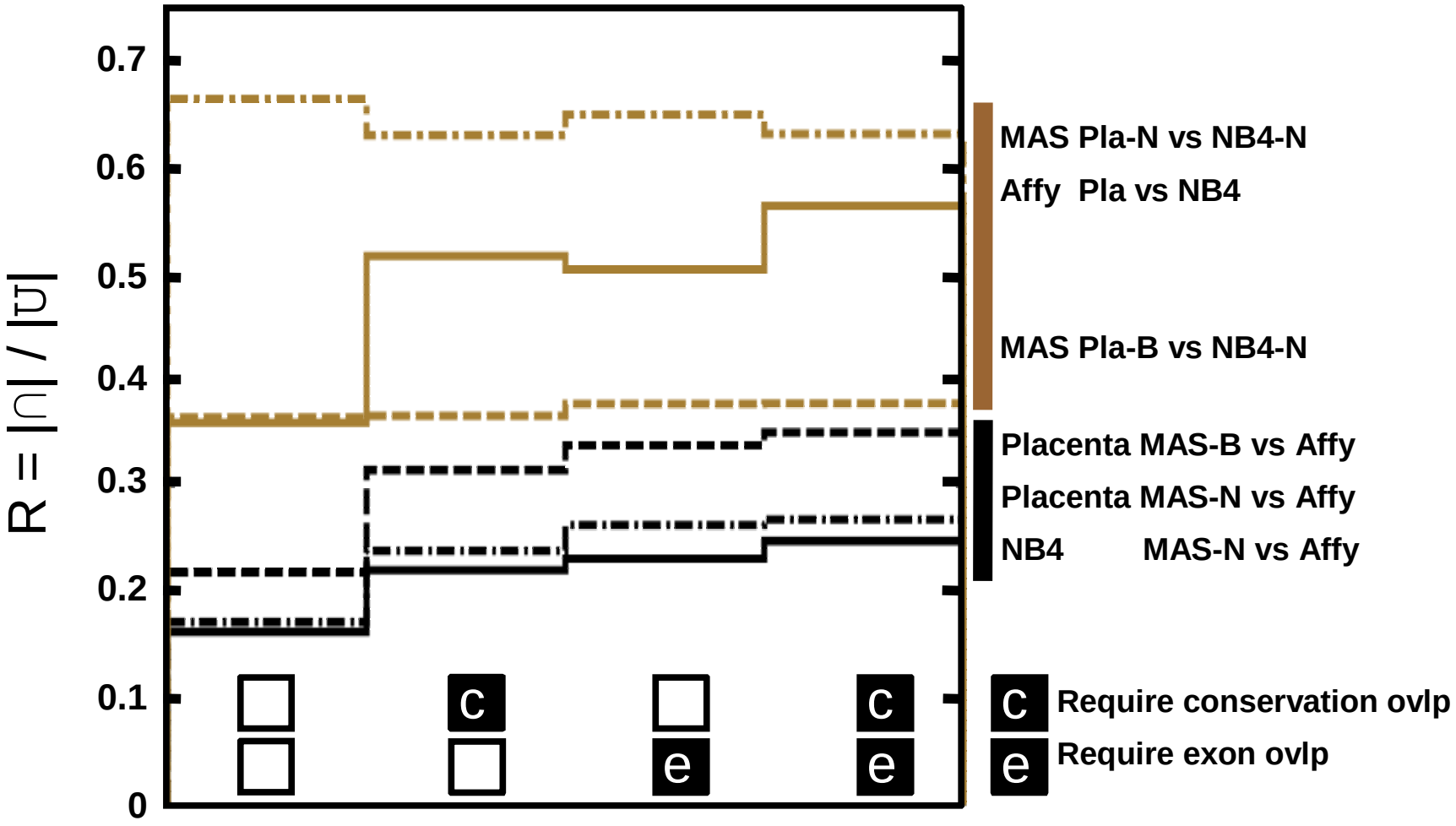
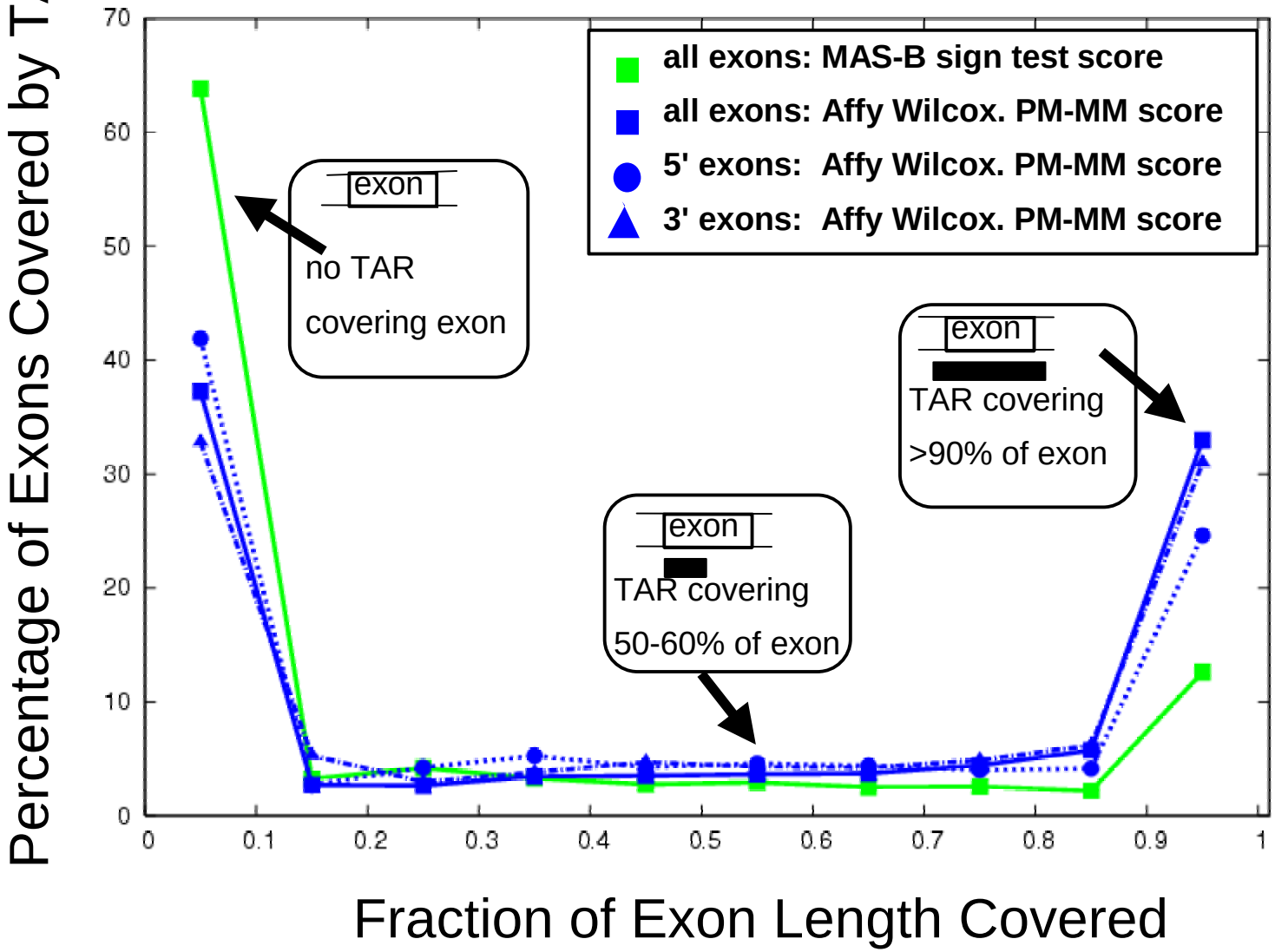

Placenta
MAS-N
[701]
265
240
207
MAS-B
[684]
128
67
108
325
R=0.17
R=0.22
Affy
[629]
NB4
MAS-N
[632]
455
177
452
R=0.16
Affy
[629]

Figure 3 B


R = |∩| / |∪|
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
c
e
c
e
MAS Pla-N vs NB4-N
Affy Pla vs NB4
MAS Pla-B vs NB4-N
Placenta MAS-B vs Affy
Placenta MAS-N vs Affy
NB4 MAS-N vs Affy
c Require conservation ovlp
e Require exon ovlp


Percentage of Exons Covered by TAR
Fraction of Exon Length Covered
all exons: MAS-B sign test score
all exons: Affy Wilcox. PM-MM score
5' exons: Affy Wilcox. PM-MM score
3' exons: Affy Wilcox. PM-MM score
exon
no TAR
covering exon
exon
TAR covering
>90% of exon
exon
TAR covering
50-60% of exon
0
10
20
30
40
50
60
70
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1