

Case Study #2: Oligonucleotide Probe Design and Validation

In this study Vmatch demonstrates its ability to detect potential cross-hybridization, which may well prevent costly misinterpretation of oligoarray experiments!

A critical problem in oligonucleotide probe design is to identify unique regions in the target sequence set to avoid cross-hybridization with non-target regions. Vmatch provides an efficient solution to this problem even for very large target sequence sets. To illustrate this, we validate probe sets used in a recent study by Kapranov et al. (2002) to evaluate transcriptional activity in human chromosomes 21 and 22 (<http://www.sciencemag.org/cgi/content/full/296/5569/916>). Our validation consists of checking how many of the probes have exact matches on human chromosome 1, thus potentially confusing the original study by probe cross-hybridization to non-chromosome 21/22 transcripts. Full validation would merely involve running the same vmatch commands on all human chromosomes.

The probe sets were downloaded from http://www.affymetrix.com/transcriptome/sci_21_22_paper/rawdata.html. The 13249s_del set interrogates every base pair of the Di George critical region (362,901 bp) on the human chromosome 22. Using the clustering option of vmatch we show that 96.60% of the probes are non-redundant. Mapping to chromosome 1 revealed 1,632 distinct probes matching to 849,054 loci on that chromosome. The second probe set (chipA) consists of selected probes in order to minimize cross-hybridization, consisting of chromosome 21 and 22 sites with an average 35 bp gap between probes. This set is 99.57% non-redundant according to vmatch output, and it includes nine probes that perfectly match 272 distinct sites on chromosome 1. Not all of these sites will be within the transcriptome.