YAC Library Pooling Scheme for PCR-Based Screening

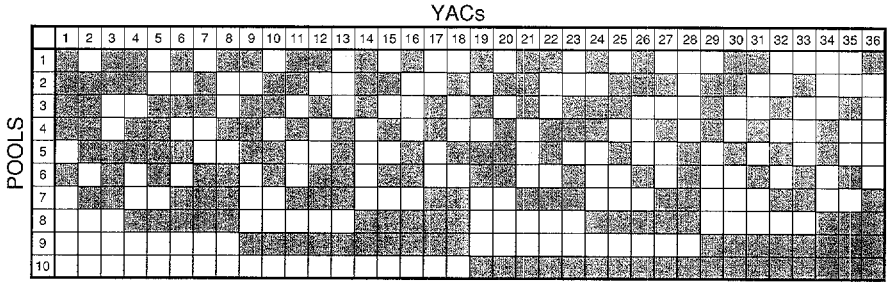
David J. Balding and David C. Torney

The PCR is a rapid method for screening a library of clones for the presence of clones containing an STS. Usually the library is divided into pools of clones, and the PCR is run on each pool. The problem we address here is to design efficient and robust pooling schemes for such PCR-based screening. Two questions are relevant: (1) Given an arbitrary unique sequence, how should one pool a library of clones to find rare positive clones (those containing this unique sequence), using a reasonable number of pools and a minimum number of pools queried per positive? (2) How can the design of the pooling scheme be robust to experimental errors (false positives, false negatives) when querying pools with PCR? Clearly, we want to do group testing in a way that gives correct results even in the presence of experimental errors.

In answer to these questions, we designed a pooling scheme called a J-detector, capable of indicating either which j clones are positive for $j \leq J$, or whether more than J clones are positive. The scheme works in the presence of K experimental errors provided any one clone in the J-detector occupies at least $K+1$ pools that are not among the pools jointly occupied by any set of J other clones. For example, if $J = 1$, and $K = 0$, we require that, among the pools containing clone$_i$, there is at least one pool that does not contain clone$_j$ for all $i \neq j$. Thus we can distinguish one positive from two positives.

From information theory we know that the number of pools in a J-detector must be at least $J \log N$, where N is the number of clones in the library. We believe that t-designs (Beth et al., 1986) constitute optimal J-detectors, therefore we focused our efforts on improved methods for the construction of t-designs. A t-design has three parameters: v, the number of pools; k, the number of pools each clone occupies; and t, the maximum number of pools any two clones jointly occupy.

The 1-Detector for PCR-Based Screening of the Chromosome 16 YAC Library



The design for a 1-detector containing 36 YAC clones has ten pools. Each clone is in 5 pools and two clones occur jointly in no more than 3 pools.

The chromosome 16-specific YAC library developed at Los Alamos contained 550 clones with an average insert size of approximately 215 kb, representing approximately a one-fold coverage of this chromosome. We chose to divide the library into 16 partitions each containing 36 clones and construct a 1-detector with $K = 1$ for the clones in each partition. In other words, the pooling scheme allows us to detect (1) which single clone among the 36 is positive for an STS or (2) whether there is more than one positive clone in the partition, even in the presence of an erroneous PCR reaction.

Assuming our YAC library represents uniform one-fold coverage of chromosome 16, the probability that more than one positive will occur in any of the 16 1-detectors is approximately 0.01. These 1-detectors (shown in the figure) are given by the t-design with parameters $v = 10$, $k = 5$, and $t = 3$. Note that the five pools containing one clone and the five pools containing another clone have at most three pools in common as $t = 3$.

Suppose only one clone in a 1-detector is positive for a given STS. Then even if one pool containing the positive clone yielded a false negative and only four pools containing that positive clone yielded positive results, one could use parsimony to tentatively identify the positive clone ($K = 1$). If the 1-detector contained two positive clones, at least seven pools would yield positive results (in the absence of experimental errors), a result readily distinguished from the five positive pools expected for a single positive clone. In fact, 4/7 of the time, only seven pools would be positive and all but three clones would be identified as negative. Thus, even when more than one clone in the 1-detector is positive for a given STS, the screening identifies a large number of negative clones, which can be eliminated from further consideration.

To identify which of the sixteen 1-detectors to screen, one could implement two levels of a four-way branching tree like that of Green and Olson (1990). Then, a maximum of 20 PCR reactions are required to identify each positive clone. Our pooling scheme has been successfully used to identify 30 YACs each containing a different STS. In almost all cases, PCR screening for each STS yielded five positive pools in a 1-detector, and the clone thereby identified as positive was always confirmed in subsequent analysis.

We plan to take advantage of the larger t-designs in future experiments. For example, the design with $v = 12$, $k = 6$, and $t = 4$ will accommodate 132 clones in its 12 pools. We found that the Biomek robot can create these pools given a bit-string representation. ■