LA-UR-97- 3500

Approved for public release;
distribution is unlimited

Title:

Optimal Screening Designs for Biomedical
Technology

Author(s):

D. C. Torney, T-10
D. J. Balding, University of Reading, England
W. J. Bruno, T-10
E. Knill, CIC-3
C. C. Whittaker, T-10

Submitted to:

DOE Office of Scientific and Technical Information (OSTI)

RECEIVED

OCT 01 1997

OSTI

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

MASTER

Los Alamos
NATIONAL LABORATORY

Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by the University of California for the U.S. Department of Energy under contract W-7405-ENG-36. By acceptance of this article, the publisher recognizes that the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, make any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

**Portions of this document may be illegible
in electronic image products. Images are
produced from the best available original
document.**

Optimal Screening Designs for Biomedical Technology

David C. Torney*, William J. Bruno,
Emmanuel Knill, and Clive C. Whittaker
Los Alamos National Laboratory

David J. Balding
University of Reading, England

Abstract

This is the final report of a three-year, Laboratory Directed Research and Development (LDRD) project at Los Alamos National Laboratory (LANL). Screening a large number of different types of molecules to isolate a few with desirable properties is essential in biomedical technology. For example, trying to find a particular gene in the Human genome could be akin to looking for a needle in a haystack. Fortunately, testing of mixtures, or *pools*, of molecules allows the desirable ones to be identified, using a number of experiments proportional only to the logarithm of the total number of types of molecules. We show how to capitalize upon this potential by using optimized pooling schemes, or *designs*. We propose efficient non-adaptive pooling designs, such as “random sets” designs and modified “row and column” designs. Our results have been applied in the pooling and unique-sequence screening of clone libraries used in the Human Genome Project and in the mapping of Human chromosome 16. This required the use of liquid-transferring robots and manifolds---for the largest clone libraries. Finally, we developed an efficient technique for finding the posterior probability each molecule has the desirable property, given the pool assay results. This technique works well, in practice, even if there are substantial rates of errors in the pool assay data. Both our methods and our results are relevant to a broad spectrum of research in modern biology.

Background and Research Objectives

This work was motivated by the need to screen large collections of arrayed clones of DNA sequences, to identify clones containing a particular sequence. These collections are called clone *libraries*---existing libraries have up to 10^5 clones. One motivation for such screenings was to construct clone maps, covering Human chromosomes [8]. In this setting, unique, or non-repeated, sequences form an unambiguous link between all clones which contain them [12]. It is, however, out of the question to perform individual

---

*Principal Investigator (E-mail: dct@lanl.gov)

experiments on every clone in the library to determine whether it contains a particular linking sequence, and, to make matters worse, to map a chromosome often requires thousands of linking sequences.

Fortunately, it is possible to identify all clones containing a particular sequence by performing *group tests*: ascertaining whether any clone in a group or *pool* contain the sequence of interest. The Polymerase Chain Reaction (PCR) enables accurate group tests, even with thousands of clones in the pool. The test outcomes from well-chosen groups allow correct inference of all of the *positive* clones---containing the sequence of interest---whose positive status is then individually verified. A complete specification of the clones in each pool is referred to as a pooling *design*.

In fact, group testing can be optimally efficient: each positive clone could be identified using a number of group tests proportional only to the logarithm of the number of clones in the collection. Whenever the proportion of positives is small, pooling experiments offer a dramatic reduction in the number of experiments required to identify the positives. The following example illustrates the great potential of group testing designs. A mnemonic for this design might be “divide and conquer”. If the group test on a group of clones indicates it contains a positive clone, this group is halved and a group test is performed on each half. Thus, if one first performs a group test on the entire library, and one follows positive “leads” down to individual clones, this approach will identify all positives in a number of experiments nearly equal to twice the number of positives multiplied by the logarithm, base two, of the number of clones. Although this would doubtless be the simplest approach conceptually, it would be a lot of work to create pools “on the fly”, more than could be performed in a reasonable length of time with state-of-the-art, commercially available robots. Thus, we proposed to develop non-adaptive pooling designs.

In non-adaptive designs, all pools are assayed in one pass. The group test results are used to infer all clones that contain the sequence of interest. Non-adaptive designs typically require construction of the smallest number of pools. Furthermore, as the same pools are used repeatedly, for all screening sequences, the cost of their construction is apportioned---and very little DNA is required by the PCR. Therefore, our results were of immediate practical significance in reducing the experimental effort of large-scale genomic experiments.

The following is an introductory example that demonstrates the principles and practices of pooling we have developed in the course of this project. Let there be 36 clones. These are put into subsets of 10 pools, according to the design shown in Figure 1. A dark square indicates that a clone occurs in a pool. Each clone occurs in five pools and

each pool contains 18 clones. Note that at most one clone occurs in any four pools and that three clones occur in any three pools. Thus, two clones occur together in at most three of their five pools. For future reference, this design is an example of a *t*-design, in which the number of pools two clones might occur in are also bounded [5]. In this case any single positive clone could be identified even if there were false-negative pool assay. This design would, however, perform poorly in the unusual circumstances that more than one pooled clone is positive.

This experiment was carried out on batches of 36 clones from a 550-clone, YAC library of Human chromosome 16. A Beckmann robot was used to construct the pools. An example of the PCR results used to identify a positive clone are shown in Figure 2, kindly provided by Norman Doggett. At the top, the results of PCR assays are shown for 15 batches of 36 clones. The PCR products are visualized after electrophoresis, and the desired PCR product has a size of approximately 150 basepairs, based upon the size standards. The batch denoted "d14" is the only batch which produces the product. In the bottom panel, at the left are shown the results of PCR assays of the ten pools created from the 36 clones in d14. The five pools that are positive (i.e., that contain the desired PCR product) are numbers 1, 2, 4, 5, and 8. With reference to Figure 1, the positive clone should be clone number 5, which was verified as shown at the bottom right. Note that the "negative" pools contain a pair of spurious bands, a frequent PCR artifact. Also, there is a trace of what might be the correct product in pool 9.

Only the small scale of the example described above and depicted in Figures 1 and 2 distinguishes it from the more production-oriented experiments of the Human Genome Project. The latter often involve of the order of 10^5 clones, combined into hundreds of pools, each containing approximately 10^3 clones. Under these circumstances, it requires a lot of effort to create pools, regardless of the pooling design. The main objective of this project is to improve the design of the pools---to optimize the performance of the design. To meet this objective, we created measures of design performance and the capability to simulate pooling experiments, so that the performances of various designs could be exhibited prior to their large-scale implementation. For instance, we found that some relatively simple "row and column" designs for pooling a BAC library of the Human Genome, which could be constructed manually, would perform nearly as well as more sophisticated designs, which would have required a prohibitive number of robot-hours for construction.

Our results are relevant to a much broader range than the screening of clones of DNA sequences because molecular biology is largely the study of combinations. Thus, extending the combinatorial design of experiments into this arena is natural and promising

[13], as might be seen in the following examples. For instance, the PCR is only one of many effective group tests; if a ligand-binding assay were used, drug discovery could be greatly facilitated by the "random sets" designs we advocate.

With the advent of new molecular technologies, routine medical examinations of individuals will become increasingly complex. The needles in the haystack are, in this situation, unhealthy tissues or inauspicious genes. Not only will the aggregates tested be heterogeneous, so will be the tests themselves. Some tests will be of blood biochemistry, others of genetic sequences, others of physiology, etc. It will become increasingly beneficial to optimize the design of these examinations. As an example, fixing the total or average cost of the examination, what types of data should be collected to yield the most comprehensive evaluation? This is an example of the type of desideratum whose solution lies distantly along the path we set out upon in this project.

Finally, the simplest genetic propensities---those involving one locus---have been successfully analyzed in recent years. Of more general significance, however, are phenomena which cannot be attributed to a single gene, but which are, essentially, a global property emerging from the combination of alleles of a large number of genes. Consider, for example, the various facets of intelligence or of aging, which, in all likelihood, will never be attributable to individual genes. We have studied how emergent properties can be attributed to combinations of genes, or other molecules, and it is quite natural to apply our methods and results in this arena as well [13].

Importance to LANL's Science and Technology Base and National R & D Needs

Efficient techniques for screening clone libraries are important for genome research and for mapping done in the Center for Human Genome Studies at Los Alamos. In addition, the programming of liquid-transferring Packard robots has many other applications as these are the best commercially available robots. The experiments of this project have the betterment of human health as their ultimate goal based, in part, upon improved technologies for molecular biological analyses. In addition, the theoretical aspects of the design of efficient screening experiments provide new avenues for mathematical research, particularly in combinatorics.

Scientific Approach and Accomplishments

Robots for Pooling. We adapted the Packard 204 robot for creating the pools of clones. This involved developing “scheduling” software, which determines, for example, how many pools are to be on the platform at one time, the volumes of the pools, and the order in which the arrayed clones are to be transferred into pools. The schedule is further expanded into detailed “elementary” instructions for the robot. We used this robot to make pools from a 1298-clone YAC library [6]. Because the PCR reaction amplifies minimal amounts of DNA, the probes or tips of the robot, which contact clone suspensions, must be fastidiously cleaned between transfers. We determined that a five-second rinse, including 0.5% bleach solution, was adequate.

Random Sets Pooling Designs. The main objective of this project was to characterize optimum non-adaptive pooling designs. Designs are often difficult to construct mathematically, and, with our practical objectives and requirements in mind, we decided to optimize the performance of a family of readily constructed designs. Based on information on clone libraries, it is a reasonable first approximation to assume, *a priori*, that each clone is independently positive with the same probability. Under this model, it follows that the number of positives is binomially distributed. In this situation, it is reasonable for each clone to occur in the same number of pools. How should this number be chosen? We were able to provide an answer by using a family of pooling designs called random k -sets designs. In these designs, each clone occurs in k pools, and these k are selected uniformly from all pools, independently for each clone. and there are no errors. The Barillot designs [4], which start with a hyper-cubic array of plates and add various types of diagonal pools, perform admirably but must be constructed using robots. A compromise between the ease of construction of “row-column” pools and the robustness of the Barillot systems is achievable by using pools that are diagonal when viewed at the resolution of plates, but consist of a single row or column within the plates on that diagonal. In effect, one makes simple row and column pools, and rather than shuffling all of the clones, as Barillot proposed, only the plates are shuffled.

Pools created in this way are not quite as good as the Barillot pools or as the optimized random k -sets pools. In particular, whenever there is more than one positive on a given plate, unresolved negatives from the same plate will also occur. Fortunately, for most libraries, there is only a small probability that two positives will occur, independently, on the same plate. On the other hand, this also means that extra care should be taken to prevent cross-contamination between wells on a plate or picking clones from the colony

more than once. A simple design of this type was implemented for a chromosome-specific YAC library, and a more elaborate one was used to pool the CEPH megaYAC library for mapping Human chromosome 16 [8].

Pools for BAC Human Genome Libraries. One of the project objectives was to design pools for a new Human genome clone library. This library, which offers many advantages for the sequencing and mapping of the Human genome, was created in Melvin Simon's laboratory at the California Institute of Technology. It is referred to as a BAC library. We have designed pools for this library but they await construction. There is no royal road to constructing pools, and it must be seen to be sufficiently advantageous before they will be constructed. A set of pools of this library was constructed at the Whitehead Institute Genome Center, using custom robots. Their pooling design has much in common with ours; for example, each clone occurs in five pools.

Consider a BAC library with 1000 plates and 96 clones on a plate. This library has 4.5-fold coverage, which means this is the average number of clones in the library that cover any "point" within the Human genome. It is desired to use 190 pools and to have each clone occur in five pools. Each plate is an 8x12 array of clones. Let the plates be arrayed in a rectangular parallelotope: $5 \times 3 \times 34 \times 2$. At the clone level, this array measures $40 \times 36 \times 34 \times 2$; thus, it can be thought of as two nearly cubic arrays. Each plate has coordinates in the range (0--4, 0--2, 0--33, 0--1). The first pools constructed are the 40-row, 36-column, and 34-level pools. Although the last coordinate is not used when making pools, *i. e.* the two cubes are pooled together into the same pools, this coordinate is used when rearranging plates, prior to constructing the remaining two sets of pools. If an additional set of row and column pools were constructed, the total number of pools would be 187. To get exactly 190 pools, we will use 38 levels instead of 35. This leaves us with more plate coordinates than plates, so we will distribute the plates among the coordinates pseudo-randomly or in such a way that the vacant coordinates are uniformly distributed relative to all of the pools.

The optimum value of k could be taken to be the value maximizing the probability that a positive clone will be *definite positive*: occurring in at least one pool with no other clone that might be considered to be positive. If most of the positives are definite positive, the pooling design will achieve its objectives.

Figure 3 illustrates these results [6]. The results are plotted for clone libraries of sizes between 10^3 and 10^5 and with the average number of positives between 1/4 and 16. It is required that the average number of definite positives be 95% of the average number of positives. The left-hand plot shows the minimum number of pools to achieve this result

using a random k -sets pooling design. The right-hand plot shows the optimum values of k . Similar results were obtained when it was desired that the average number of definite positives is half of the average number of positives. In both cases, our results indicate the minimum number of pools that would be required, using the optimum random k -sets design. These results provide a baseline: constructing more pools than our optimum is likely to be a waste of effort.

It is, of course, possible to modify the random-sets designs to improve performance. The key is to make sure that the maximum number of pools in which any two clones coincide is as small as possible. Such designs are called sets packings. The t -designs referred to above (see Figure 1) are special sets packings, which are especially recommended as pooling designs [1, 3, 10]. We implemented a 4-sets packing design for a library of 1298 clones. This implementation is described in more detail at the close of this report.

Using Manifolds to Construct Pools. As might be imagined, robots are virtually required to implement the very efficient pooling designs such as optimized random k -sets designs and k -sets packing designs. For the largest libraries, having the order of 10^5 clones, commercially available robots aren't fast enough to construct the pools in a reasonable length of time. Fifteen seconds is required for each clone-to-pool transfer. Thus, with 1,000 clones per pool, 10^8 such transfers, and, hence, nearly 7,000 robot-hours, would be required. Faster robots are highly desirable, but there is an alternative: "pooling plumbing". In practice, manifolds are constructed that direct the arrayed clones into pools. For example, pools could consist of the clones in the rows or columns or diagonals from an arrayed set of clones. Thus, four types of pools could easily be created from each array, using different manifolds. These pools are combined with similar pools from other arrays to achieve an acceptable number of clones in a pool. The labor involved should not, however, be underestimated. A pooling design of this type was implemented for the CEPH megaYAC Human genome library of 33,000 clones, used to map Human chromosome 16 [8]. Roughly one FTE-year and a truckload of plasticware were required to create the "sub-pools", such as the row and column pools from individual plates, and to combine them into the complete pools.

The easiest designs to construct are "row and column" pools. A commonly used design is one in which a stack of 8 to 12 plates are pooled as a 3-dimensional "cube", by means of row, column, and plate pools. This design works well only if there is no more than one positive clone and there are no errors. The Barillot designs, which start with a hyper-cubic array of plates and add various types of diagonal pools, perform

admirably but must be constructed using robots. A compromise between the ease of construction of “row and column” pools and the robustness of the Barillot systems is achievable by using pools that are diagonal when viewed at the resolution of plates, but consist of a single row or column within the plates on that diagonal. In effect, one makes simple row and column pools, and rather than shuffling all of the clones, as Barillot proposed, only the plates are shuffled.

Numerical simulations confirm the appropriateness of making one additional set of row pools and one additional set of column pools after transforming plate coordinates along the following lines. The original, unshuffled row pools correspond to multiplying (via inner product) the coordinates of a plate by $(1,0,0,0)$ to find which row a plate is in, and then making row pools from clones in the plates in each row. The additional row pools would be found by multiplying the plate's coordinates by $(3,1,1,0)$, and taking the result mod 5 to determine the rows. This could be interpreted as shuffling the plates according a linear transformation, as recommended by Barillot [4]. Furthermore, for these row pools, any plate whose fourth coordinate equals 1 should be included in the row of plates in reversed orientation, *i.e.* rotated 180° . For the final set of column pools, one would multiply the plate coordinates by $(1,5,1,1)$ to get a value a . The new plate column is $a \bmod 3$ and any plates for which the integer part of $a/3$ is odd are reversed. Although there are features of these rearrangements, such as plate rotation, which were not considered by Barillot *et al.*, the key requirements for good transformations were described in their work [4].

Clearly, the number of times that rows or columns from individual plates coincide in “large pools” is to be minimized. This motivates using prime numbers of plates on the axes of the parallelotope linear congruences for plate rearrangements. The predicted performance of pooling designs constructed using these techniques is surprisingly good. For example, with 190 pools, the expected number of definite positive clones would be 3.1 for an optimum random 5-sets design, whereas it would be 2.8 for the described “row column” design.

Miscellaneous Results. The length of clones can often vary considerably, and the prior probability of containing a screened DNA sequence is proportional to this length. This should be taken into account in the pooling design, and we were the first to propose pooling designs that make use of this prior information [7]. It turns out that clones that have a larger probability of being positive should occur in a smaller number of pools. A key collection of 33,000 clones for the Human genome that has clones between 10^5 to 10^6 basepairs in length is the CEPH megaYAC library. Assume the coverage, c —the average

number of times a “point” within the Human genome occurs in this collection---equals ten. Suppose we decide to use 250 pools [6]. How might the probabilities for a *random incidence design* be chosen for this collection of clones? Furthermore, is there a substantial difference between the optimum incidence probabilities for the shortest and longest clones?

The leading-order approximation to an optimum incidence probability equals $1/11$ for all clones, placing each clone into approximately 23 group tests. As the extent of a haploid Human genome is 3×10^9 basepairs, the probability a uniformly placed, short clone, of length 10^5 basepairs, covers any “point” within the genome is approximately equal to $(3 \times 10^4)^{-1}$, and the comparable probability for a long clone, of length 10^6 basepairs, is approximately equal to $(3 \times 10^3)^{-1}$. These probabilities are denoted $p_{\leq}$ and $p_{\geq}$, respectively. The difference between the first-order corrections to the optimum incidence probabilities for short and long clones is approximately equal to $.01 \log(p_{\geq}/p_{\leq})$. Substituting the probabilities given above, this yields $.01 \log(10)$. As described above, the leading-order term, equals $1/11$. The ratio of the difference between the two first-order corrections to the leading-order term is approximately equal to 0.25, indicating that these corrections to the incidence probabilities should not be neglected. In this setting, the long clones might occur in 20 pools, whereas the short clones could occur in 26 pools. Details are given in our manuscript [7] including the caveat that all incidences could profitably be proportionately reduced. For example, 23 might be reduced to 12, 20 to 10 and 26 to 13.

In a separate study, we considered the type of pooling design that would be appropriate for a set of clones that have been assembled into a map [2]. Imagine that these have been reduced to a “chromosome walk”, in which each clone overlaps only its two neighbors, and that each “point” of the genome is included in exactly one or two clones of this map. Such a map, comprising 1530 clones, has been constructed for the genome of *A. nidulans*. Although we found that the theoretical minimum number of pools required is twice the logarithm base two of the number of clones, we settled upon a design with substantially more pools because the complementation assays were not judged to be effective with more than 100 clones per pool. Thus, we proposed the Steiner system $S(3,5,65)$ [5], which has 65 pools, even though the theoretical minimum is 22 pools. In the Steiner system, each clone occurs in five pools, and two clones coincide in at most two pools. Thus, two positive clones would not obscure the status of any other clone. Thus, we have made use of a constraint on pool size to arrive at a design that has substantial error-correcting capabilities. This pooling design will be implemented in the laboratory of Jonathan Arnold at the University of Georgia using the robotic and computational techniques we have developed.

The more clones in a pool, the greater the error rates of the PCR reaction. Although we anticipated false negative PCR reactions, there are also false positive PCR reactions, largely attributable to cross-contamination of clones. That is, suspensions of clones that are adjacent to one another on plates can splash and overflow into one another, or drops could fall off of the robot's probes. These errors do not compromise the usefulness of the optimized random k -sets designs, but they required the development of new approaches for inferring the positive clones [10]. We implemented Bayes rule, calculating the posterior probability each clone is positive, given the error model and pool outcomes. This involves, formally, entertaining all subsets of the clones to be the set of positive clones. As the number of such subsets is typically too large to perform a computation for each subset, it is fortunate that the Markov chain Monte Carlo method, along with Gibbs sampling, samples the subsets according to their posterior probability of being positive. This procedure yields the probability that each clone is positive by counting the proportion of times that each clone occurs in the sets of positives generated by the Markov chain.

The Markov chain Monte Carlo Bayes analysis was used for a 1,298-clone YAC library, pooled using a 4-sets packing design on 47 pools. It was assumed that the errors were independent, pool by pool. The false positive and false negative pool-assay error rates were estimated to be 0.13 and 0.05, respectively, from the data. As one indication of the self-consistency of our analysis, screening the pools yielded 39 positives, whereas, the prediction of the Markov chain Monte Carlo was 40 ± 6 . The library coverage ensured that there should be an average of 2.6 positives per screening, whereas the Markov chain Monte Carlo technique predicted 3.06. The inconsistencies of the theoretical predictions were, however, surprisingly small because our model for the errors was not comprehensive. No other approach could have identified the positive clones. This approach was easily implemented, is generally applicable to any pooling design, and it makes the best possible use of the data. It is only in the last few decades that Markov chain Monte Carlo has begun to be broadly applied, and we are pleased to report that this relatively novel technique has now found its way into biological laboratories.

Conclusions

In conclusion, good techniques and designs for performing combinatorial experiments, such as the ones we have developed, allow the experimenter to rapidly identify the needles in their haystack. Although our approaches have already proven their worth for sequence-content screening of large collections of clones of DNA sequences,

they also contain the rudiments of more sophisticated, future tests for human genetics and human health.

Publications/References

- [1] Balding, D. J. and Torney, D. C. (1996) Optimal Pooling Designs with Error Detection. *J. Comb. Theory A* 74, 131-140.
- [2] Balding, D. J. and Torney, D. C. (1997) Pooling Designs for Screening Clone Maps. *J. Fungal Genomes and Biology*, in press.
- [3] Balding, D. J., Bruno, W. J., Knill, E. and Torney, D. C. (1996) A Comparative Survey of Non-Adaptive Pooling Designs. In *Genetic Mapping and DNA Sequencing*. IMA Volumes in Mathematics and its Applications, Springer-Verlag, 81 133-154.
- [4] Barillot, E., Lacroix, B., and Cohen, D. (1991) Theoretical analysis of library screening using a n-dimensional pooling strategy. *Nucl. Acids Res.* 19 6241-6247.
- [5] Beth, T., Jungnickel, D. and Lenz, H. (1985) *Design Theory*. B.-I. Wissenschaftsverlag, Mannheim.
- [6] Bruno, W. J, Knill, E., Balding, D. J., Bruce, D. C., Doggett, N. A., Sawhill, W. W., Stallings, R. L., Whittaker, C. C., and Torney, D. C. Efficient Pooling Designs for Library Screening (1995) *Genomics* 26, 21-30.
- [7] Bruno, W. J., Sun, F. and Torney, D. C. (1997) Optimizing Non-Adaptive Group Tests for Objects with Heterogeneous Priors. *SIAM J. of Applied Math.*, in press.
- [8] Doggett, N. A., Goodwin, L. A., Tesmer, J. G., Meinke, L. J., Bruce, D. C., Clark, L. M., Altherr, M. R., Ford, A. A., Chi, H.-C., Marrone, B. L., Longmire, J. L., Lane, S. A., Whitmore, S. A., Lowenstein, M. G., Sutherland, R. D., Mundt, M. O., Knill, E. H., Bruno, W. J., Macken, C. A., Torney, D. C., Wu, J. R., Griffith, J., Sutherland, R. D., Deaven, L. L., Callen, D. F., and Moyzis, R. K. ((1995) An integrated physical map of human chromosome 16. *Nature (London)* 377 335S-365S.
- [9] Knill, E., Bruno, W. and Torney, D. (1996) Non-Adaptive Group Testing in the Presence of Errors. *Discrete Applied Mathematics*, submitted.
- [10] Knill, E, Schliep, A., and Torney D. C. (1996) Interpretation of Pooling Experiments Using the Markov Chain Monte Carlo Method. *J. Comp. Biol.* 3 395-406.
- [11] McCormick, M. K., Buckler, A., Bruno, W., Campbell, E., Shera, K., Torney, D. Deaven, L., and Moyzis, R. (1993) Construction and Characterization of a YAC Library with a Low Frequency of Chimeric Clones from Flow Sorted Human Chromosome 9. *Genomics* 18, 553-558.
- [12] Torney, D. C. (1991) Mapping Using Unique Sequences. *J. Mol. Biol.*, 217 259-264.
- [13] Torney, D. C. (1996) Combinatorial Experiments for Complex Biological Systems. *Nature*, submitted.

Figure 1. The T-design for a 1-detector containing 36 YAC clones has 10 pools. Each clone is in 5 pools and two clones occur jointly in no more than 3 pools. Note: dark square indicates a clone is in a pool.

Figure 2. PCR results used to identify a positive clone. (a) Results of PCR assays for 15 batches of 36 clones. (b) Results for batch d14, the only batch with the desired PCR product of approximately 150 basepairs. (c) Results for single clone containing the desired PCR product.

Figure 3. (a) Minimum number of pools with the expected number of resolved positives equal to $0.95c$. The smallest value of v such that a random k -sets design can achieve the target expected number of $0.95c$ resolved positive clones is depicted. (b) Optimum k ; expected number of resolved positive clones equal to $0.95c$. The values of k achieving the maximum expected number of resolved positive clones are depicted, with the value of v depicted in (a). In both figures, the x axis is n , the number of clones to be pooled and the y axis is c , the coverage parameter. Both axes have a logarithmic scale, and the tics facing the inside of the plot are at integer values, while those on the outside are at values corresponding to half integers.