seen representing hundreds of DNA fragments. We have shown that this spectrum is dramatically different with DNAs from unrelated individuals, and the spectrum is markedly dependent on the choice of restriction enzyme, as expected. Repeated measurements on the same sample are highly reproducible. The ability of the method to detect a specific altered repeat length in a complex DNA sample has been validated by examining several individuals with normal or expanded repeat sequences in the Huntington's disease gene. One very powerful application of this method may be the analysis of potential DNA differences in monozygotic twins discordant for a genetic disease. This method can be used to capture genome subsets containing any interspersed repeat. It will also detect insertions and deletions nearby such repeats. Methylation differences between sensitive methylation samples are also detectable when restriction fragments are used.

Conventional analysis of triplet repeats is very laborious since individual repeats must be analyzed by electrophoresis on DNA sequencing gels. The decrease in effort for such analyses will scale linearly as the number of repeats that can be analyzed simultaneously, so we are potentially looking at something like a factor of 100 improvement if the above scheme under development can be effectively realized.

As an alternative approach, we are developing chip-based methods that can detect the length of a tandemly-repeating sequence without any need for gel electrophoresis. Here the goal is to build an array of all possible repeat sequence lengths flanked by single-copy DNA. When an actual sample is hybridized to such an array, the specific alleles in the sample will produce perfect duplexes at their corresponding points in the array and at mismatched duplexes elsewhere. Thus, the task of scoring the repeat lengths is reduced to the task of distinguishing perfect and imperfect duplexes. Currently we are exploring a number of different enzymatic protocols that offer the promise of making such distinctions reliably.

In other work we are using enzyme-enhanced sequencing by hybridization (SBH) as a device for the rapid preparation of DNA samples for mass spectrometry. For example, partially duplex DNA probes can capture and generate sequence ladders from any arbitrary DNA sequence. Current MALDI protocols allow sequence to be read to lengths of 50 to 60 bases. While this is probably insufficient for most de novo DNA sequencing, it is an extremely promising approach for comparative or diagnostic DNA sequencing.

DOE Grant No. DE-FG02-93ER61609.

Preparation of PAC Libraries

Joe Catanese, Baohui Zhao, Eirik Frengen, Chenyan Wu, Xiaoping Guan, Chira Chen, Eugenia Pietrzak, Panayotis A. Ioannou,[1] Julie Korenberg,[2] Joel Jessee,[3] and **Pieter J. de Jong**
Department of Human Genetics; Roswell Park Cancer Institute; Buffalo, NY 14263
de Jong: 716/845-3168, Fax: -8849
pieter@dejong.med.buffalo.edu
http://bacpac.med.buffalo.edu
[1]The Cyprus Institute of Neurology and Genetics; Nicosia, Cyprus
[2]Cedars Sinai Medical Center; Los Angeles, CA 90048
[3]Life Technologies, Gaithersburg, MD 20898

Recently, we have developed procedures for the cloning of large DNA fragments using a bacteriophage P1 derived vector, pCYPAC1 (Ioannou et al. (1994), Nature Genetics 6: 84-89). A slightly modified vector (pCYPAC2) has now been used to create a 15-fold redundant PAC library of the human genome, arrayed in more than 1,000 384-well dishes. DNA was obtained from blood lymphocytes from a male donor. The library was prepared in four distinct sections designated as RPCI-1, RPCI-3, RPCI-4 and RPCI-5, respectively, each having 120 kbp average inserts. The RPCI-1 segment of the library (3X; 120,000 clones, including 25% non-recombinant) has been distributed to over 40 genome centers worldwide and has been used in many physical mapping studies, positional cloning efforts and in various large-scale DNA sequencing enterprises. Screening of the RPCI-1 library by numerous markers results in an average of 3 positive PACs per autosome-derived probe or STS marker. In situ hybridization results with 250 PAC clones indicate that chimerism is low or non-existing. Distribution of RPCI-3 (3X, 78,000 clones, less than 1% non-recombinants, 4% empty wells) is now underway and the further RPCI-4 and -5 segments (< 5% empty wells) will be distributed upon request. To facilitate screening of the PAC library, we have provided the RPCI-1 PAC library to several screening companies and noncommercial resource centers. In addition, we are now distributing high-density colony membranes at cost-recovery price, mainly to groups having a copy of the PAC library. The combined RPCI-1 and -3 segments (6X) can be represented on 11 colony filters of 22x22 cm, using duplicate colonies for each clone. We are currently generating a similar PAC library from the 129 mouse strain.

To facilitate the additional use of large-insert bacterial clones for functional studies, we have prepared new PAC & BAC vectors with a dominant selectable marker gene (the blasticidin gene under control of the beta-actin promoter), an EBV replicon and an "update feature". This feature utilizes the specificity of Transposon Tn7 for the Tn7att sequence (in the new PAC and BAC vectors) to transpose marker genes, other replicons and other sequences into PACs