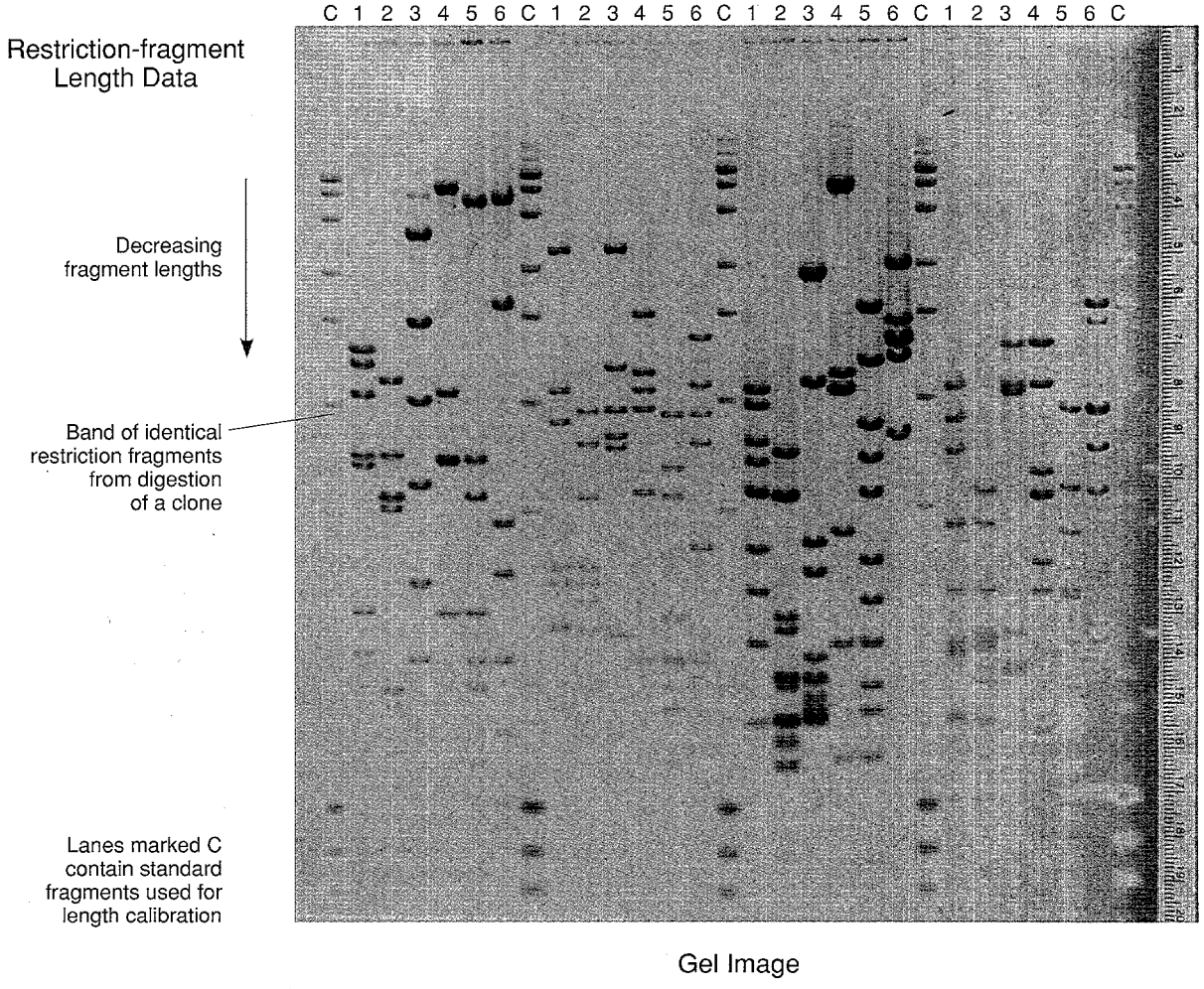
SCORE: *a program for computer-assisted scoring of Southern blots*

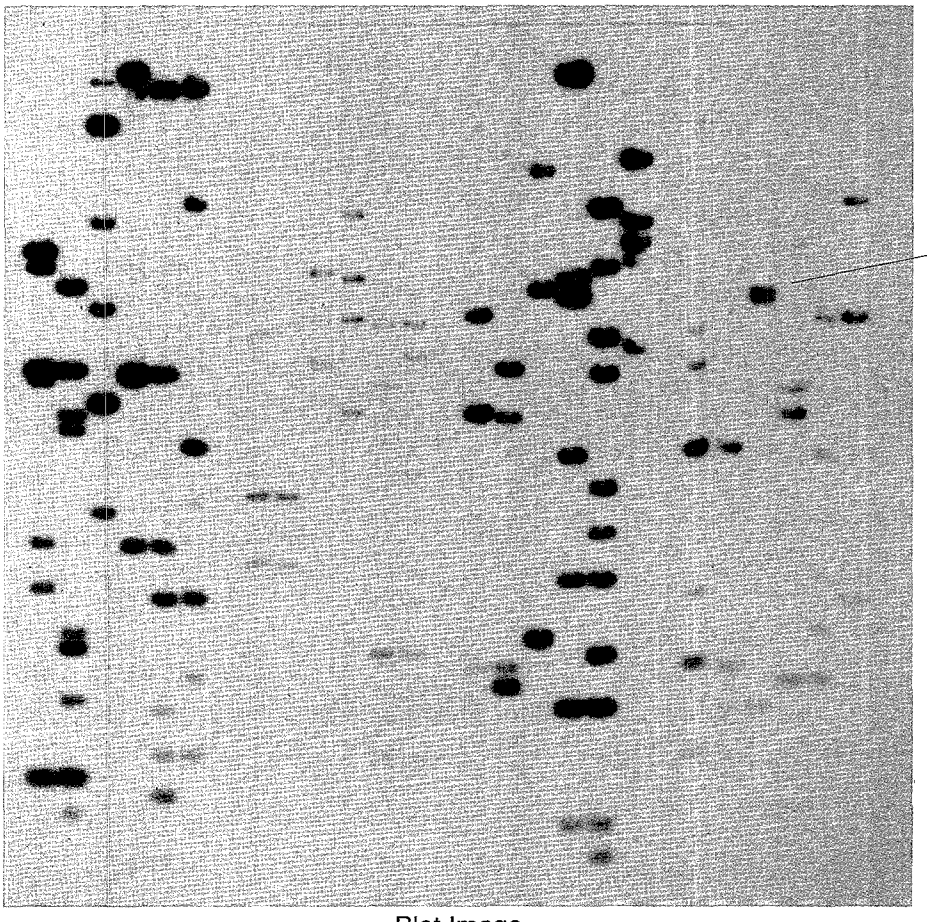
T. Michael Cannon, Rebecca J. Koskela, Christian Burks, Raymond L. Stallings, Amanda A. Ford, Philip E. Hempfner, Henry T. Brown, and James W. Fickett

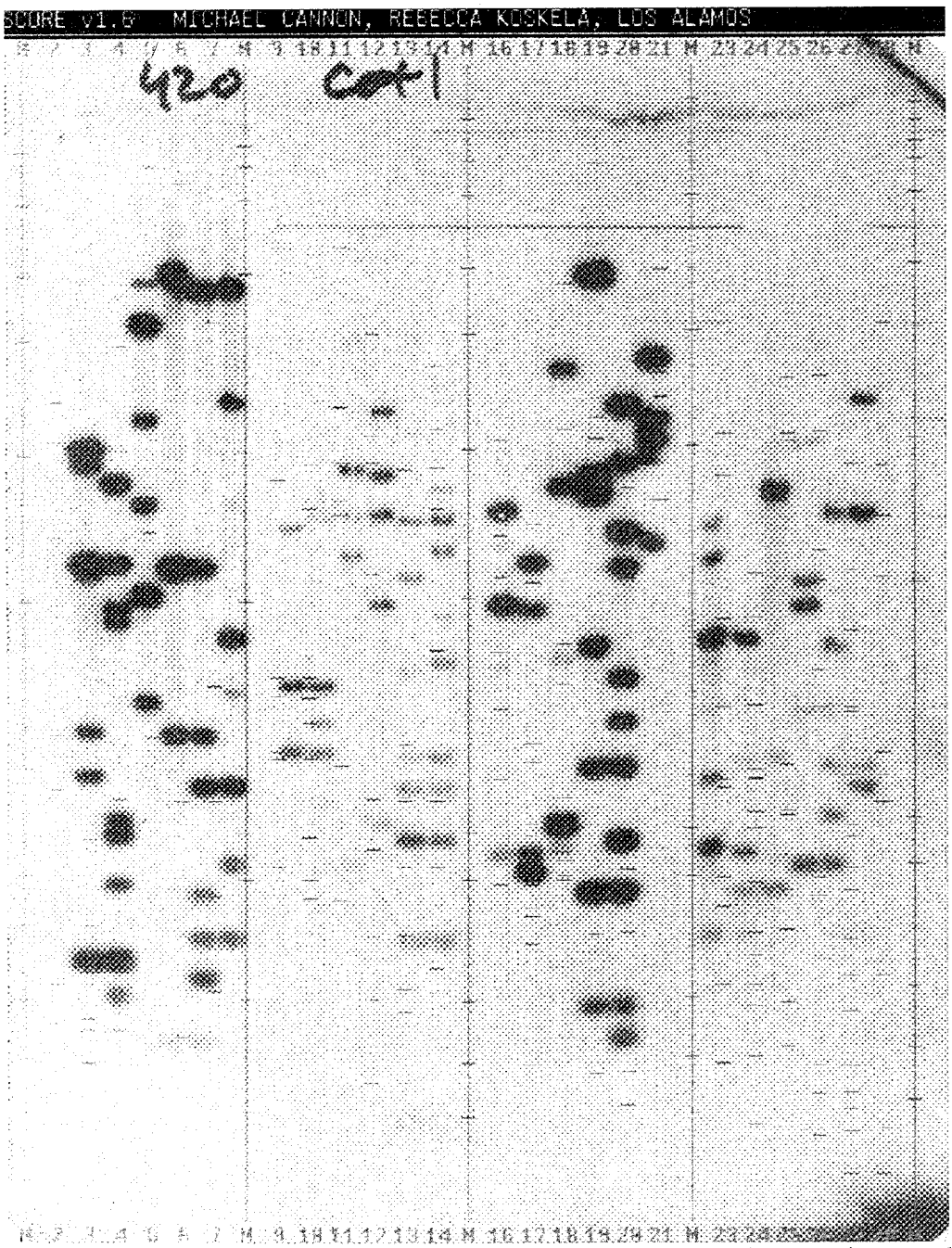
The Human Genome Project aims to collect unprecedented (for molecular biology) amounts of information, so the transfer of repetitive tasks to machines is essential. As part of the LANL physical-mapping effort, we have partially automated the task of entering clone-fingerprint data into computers. One aspect of the automation was the development of a simple image-manipulation program called SCORE. This program has improved the accuracy of the data entry and sped up the process by an order of magnitude.



As explained in "The Mapping of Chromosome 16," the Los Alamos physical-mapping project uses clone fingerprints that consist of two kinds of data. The first is a list of the lengths of DNA fragments obtained by digesting a larger cloned fragment with a restriction enzyme and then separating the restriction fragments by length using gel electrophoresis. On the previous page appears a sample photograph of a gel. The gel is divided into vertical lanes, each lane containing all the fragments of one digest of one clone. Every clone is subjected to three digests, so there are three lanes of fragments from each clone. Each fuzzy horizontal band within a lane consists of identical restriction fragments from the digest contained in the lane. The band's vertical position gives the length of the fragments in it.

The second kind of data is a Southern blot of the gel that indicates whether or not (or to what degree) certain repetitive sequences are present in each restriction fragment. The figure below is a blot image produced by hybridization of repetitive sequences to the gel shown on the left (see "Hybridization Techniques" in "Understanding Inheritance"). Bands of fragments produce a signal on the blot image only if they contain the particular repetitive element being tested for.

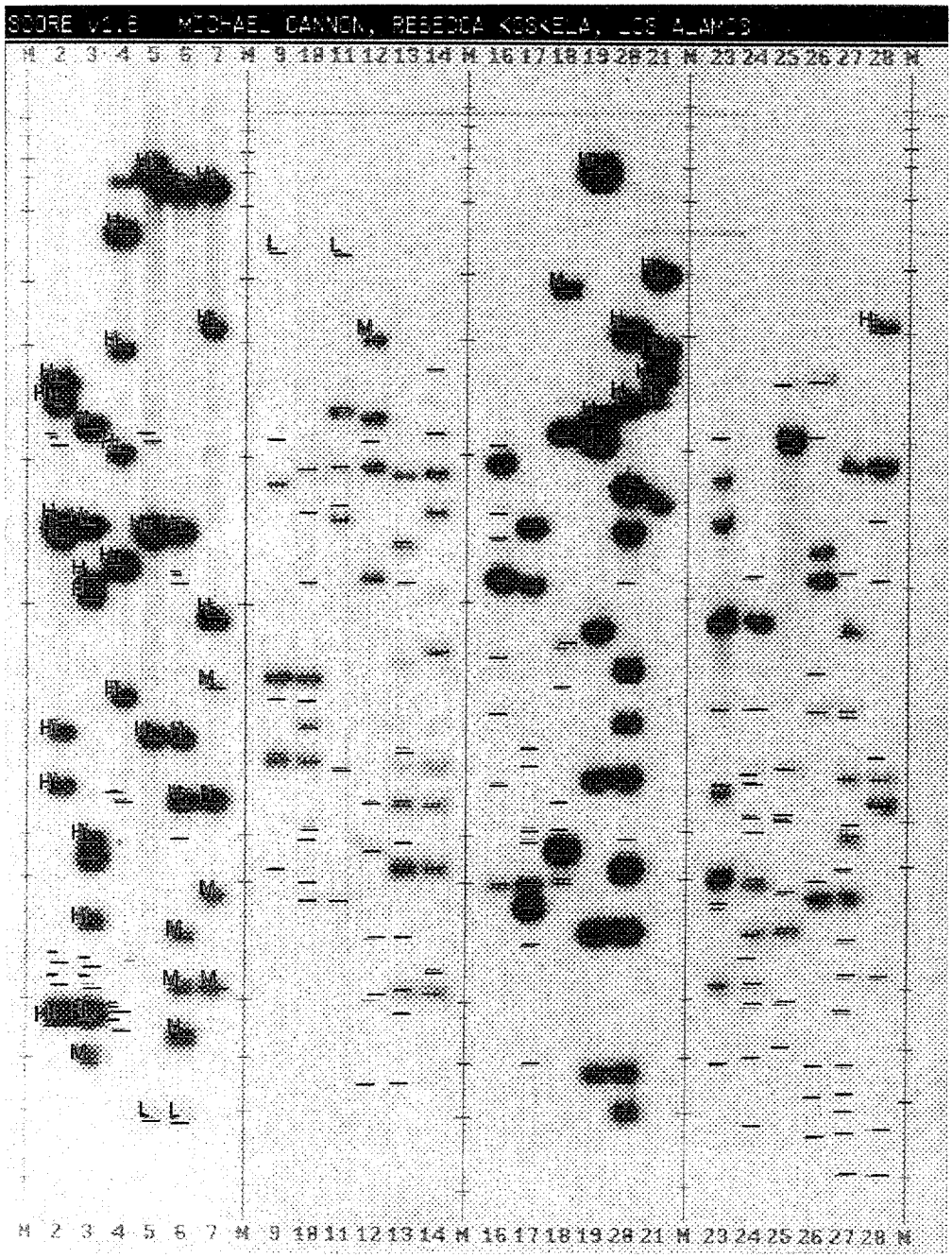




The blot image is used to assign a score to each band indicating the strength of its hybridization signal, a process known as scoring the blot. Therefore the bands on the blot image must be matched with the corresponding bands on the gel image. Formerly the two images were matched by hand, one region at a time. Each fragment was identified manually by numbering the lanes and bands on the photographs. After the scores were assigned, they were typed into our mapping database in a separate operation. Scoring the blot was the most labor-intensive part of fingerprinting. Now we score blots on a scientific workstation using the SCORE program.

Before SCORE is run, the fragment lengths are determined by a commercial image-processing workstation. Another program takes the report from the image processor and stores the lengths in the database. Also, the blot image is digitized using a desktop scanner. SCORE retrieves the fragment lengths from the database and constructs a schematic of the gel image in which the bands are denoted by colored horizontal lines positioned according to their length. The program then superimposes the digitized blot image on the schematic gel image. The figure above shows the two images on the previous pages as stored in the computer and superimposed; they match only approximately.

When the two images are on the screen, the user chooses two points on each image that should be aligned. The program then resizes and moves the digitized blot image to align it with the schematic gel image. The figure at right shows how the user sees the two images overlaid and matched on the computer screen.



At this point the actual scoring takes place. The user points to a band with a mouse, is given a menu of possible scores, and chooses one. Thus the program retains the use of expert human judgement where necessary. SCORE displays the score chosen, next to the band, for the rest of the session (colored letters in the figure). Any score may be revised at any time. If a band shows on the blot image but not on the gel image, the user may add a new fragment to the database. When all fragments have been scored, the program places their scores directly into the database, each score being associated with the proper fragment.

This program has not only cut the time needed for scoring the images by a factor of ten, but it has eliminated typographical errors in data entry. Using SCORE also has the advantage that the complete fingerprint data are in a database, easily accessible by network to the whole group working on the project and readable by the map-construction software, from the moment they are first determined.