Abstract

Motivation: The methods for analyzing overlap data are distinct from those for analyzing probe data, making integration of the two forms awkward. Conversion of overlap data to probe-like data elements would facilitate comparison and uniform integration of overlap data and probe data using software developed for analysis of STS data.

Results: We show that overlap data can be effectively converted to probe-like data elements by extracting maximal sets of mutually overlapping clones. We call these sets *virtual probes*, since each set determines a site in the genome corresponding to the region which is common among the clones of the set. Finding the virtual probes is equivalent to finding the maximal cliques of a graph. We modify a known maximal-clique algorithm such that it finds all virtual probes in a large dataset within minutes. We illustrate the algorithm by converting fingerprint and Alu-PCR overlap data to virtual probes. The virtual probes are then analyzed using double-linkage intersection graphs and structure graphs to show that methods designed for STS data are also applicable to overlap data represented as virtual probes. Next we show that virtual probes can produce a uniform integration of different kinds of mapping data, in particular STS probe data and fingerprint and Alu-PCR overlap data. The integrated virtual probes produce longer double-linkage contigs than STS probes alone, and in conjunction with structure graphs they facilitate the identification and elimination of anomalies. Thus, the virtual-probe technique provides (*i*) a new way to examine overlap data, (*ii*) a basis on which to compare overlap data and probe data using the same systems and standards, and (*iii*) a unique and useful way to uniformly integrate overlap data with probe data.

Availability: Freely available on request.

Contact: Eric Harley `eharley@cs.toronto.edu`