Expression-based Genetic/Physical Maps of Single-Nucleotide Polymorphisms Identified by the Cancer Genome Anatomy Project

Robert Clifford, Michael Edmonson, Ying Hu, Cu Nguyen, Titia Scherpbier, and Kenneth H. Buetow[1]

Laboratory of Population Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892 USA

SNPs (Single-Nucleotide Polymorphisms), the most common DNA variant in humans, represent a valuable resource for the genetic analysis of cancer and other illnesses. These markers may be used in a variety of ways to investigate the genetic underpinnings of disease. In gene-based studies, the correlations between allelic variants of genes of interest and particular disease states are assessed. An extensive collection of SNP markers may enable entire molecular pathways regulating cell metabolism, growth, or differentiation to be analyzed by this approach. In addition, high-resolution genetic maps based on SNPs will greatly facilitate linkage analysis and positional cloning. The National Cancer Institute's CGAP-GAI (Cancer Genome Anatomy Project Genetic Annotation Initiative) group has identified 10,243 SNPs by examining publicly available EST (Expressed Sequence Tag) chromatograms. More than 6800 of these polymorphisms have been placed on expression-based integrated genetic/physical maps. In addition to a set of comprehensive SNP maps, we have produced maps containing single nucleotide polymorphisms in genes expressed in breast, colon, kidney, liver, lung, or prostate tissue. The integrated maps, a SNP search engine, and a Java-based tool for viewing candidate SNPs in the context of EST assemblies can be accessed via the CGAP-GAI web site (http://cgap.nci.nih.gov/GAI/). Our SNP detection tools are available to the public for noncommercial use.

[The sequence data described in this paper have been submitted to the db SNP data library under accession nos. SS8196–SSI8418.]

SNPs (Single-Nucleotide Polymorphisms) are the most common form of DNA variation in humans. These variants occur at an estimated frequency of one per 1000 to 2000 base pairs (Cooper et al. 1995; Kwok et al. 1996; Wang et al. 1998; Cargill et al. 1999; Halushka et al. 1999), making it possible in principle to identify a genetic marker in every gene.

A collection of tens or hundreds of thousands of SNPs would serve as a valuable resource for the discovery of genetic factors affecting disease susceptibility and resistance. These markers can be used in association studies that assay how alleles of candidate disease loci correlate with particular diseases (Lander and Schork 1994; Lander 1996; Risch and Merikangas 1996). Likewise, an extensive collection of SNPs will be useful for identifying genetic variants involved in drug metabolism (Meyer and Zanger 1997); this information will enable clinicians to determine which pharmacological agent is most effective for treating a given patient's condition, as well as which compounds are least likely to produce an adverse reaction.

Because of their abundance, SNPs are the marker of choice for constructing high-resolution genetic maps used for linkage analysis (Lander and Schork 1994; Kruglyak 1997; Zhao et al. 1998) and positional clon-

ing (Collins 1995). High-density genetic maps are essential for studying complex traits such as predisposition to hypertension, diabetes, or asthma or susceptibility to infectious diseases such as malaria or acquired immune deficiency syndrome. Dense SNP-based maps also will prove valuable for loss-of-heterozygosity studies (Cavenee et al. 1983), which have played a critical role in deciphering the genetic changes involved in cancer initiation and progression. Understanding the genetic events that lead from immortalization to metastasis will improve cancer diagnosis and may reveal common genetic changes in apparently unrelated tumor types, thereby suggesting new therapies for certain forms of cancer.

Several large-scale SNP detection projects have been undertaken in recent years. The first, performed at the Whitehead Institute, was based on the hybridization of genomic PCR (Polymerase Chain Reaction) products to DNA oligonucleotide arrays (Wang et al. 1998). The Whitehead collection contains 3241 putative SNPs, 2227 of which have been placed on genetic maps. An alternative approach—examining high-throughput genomic sequence for nucleotide variants—was used by Taillon-Miller et al. (1998) to identify 153 potential SNPs in 200.6 kilobases of sequence from chromosomes 5, 7, and 13. More recently, SNP mining strategies based on the analysis of ESTs (Ex-

[1]Corresponding author.
E-MAIL buetowk@nih.gov; FAX (301) 435-8963.

pressed Sequence Tags) have been described (Buetow et al. 1999; Picoult-Newberg et al. 1999). Because the high error rate in EST sequences (~1%) makes it difficult to distinguish true genetic variants from sequencing artifacts, both Buetow et al. and Picoult-Newberg et al. used the basecalling program Phred (Ewing and Green 1998; Ewing et al. 1998) and the sequence assembly program Phrap (http://genome.washington.edu) to directly analyze EST sequencing traces. The two groups used different algorithms to filter out false-positives and validate predicted SNPs.

The goal of the National Cancer Institute's Cancer Genome Anatomy Project (CGAP) is to provide a comprehensive catalog of molecular differences distinguishing tumorous cells from their normal counterparts. Within CGAP, the Genome Annotation Initiative (CGAP-GAI) group seeks to identify allelic variants of genes involved in cancer initiation and progression. In our most recent round of SNP discovery, we used the SNPpipeline, a set of sequence analysis tools described in Buetow et al. (1999), to identify more than 10,000 high-probability candidate single nucleotide polymorphisms among publicly available EST sequences. Information about this collection of SNPs is accessible via the internet (http://cgap.nci.nih.gov/GAI/). To present these SNPs in a format useful to the human genetics community, we have placed >6800 predicted variants on integrated genetic/physical maps. We have produced maps showing the locations of SNPs in genes expressed in the breast, colon, kidney, liver, lung, or prostate in addition to a comprehensive integrated map. We provide a Java-based SNP viewer that displays sequence polymorphisms in the context of DNA sequence alignments and a search engine that retrieves SNPs by keyword, description, or gene symbol. Each SNP is linked to the extensive annotation maintained by the National Center for Biotechnology Information (NCBI). Our SNP prediction tools are publicly available for noncommercial use.

RESULTS AND DISCUSSION

SNP Prediction, Validation, and Confirmation

Using the SNPpipeline set of sequence analysis tools (Buetow et al. 1999), we have identified 10,243 high-probability (99% or better) potential SNPs among sequences contained in the 14 April 1999 release of Uni-Gene (Schuler et al. 1996). Candidates were derived from 6458 UniGene assemblies, 1862 of which corresponded to named genes. This set of candidate SNPs has been submitted to dbSNP (Sherry et al. 1999; http://www.ncbi.nlm.nih.gov/SNP/). We verified predicted SNPs in a two-step process. To validate a polymorphism, we showed that it is present in genomic DNA from eight individuals via direct sequencing or a RFLP (Restriction Fragment Length Polymorphism) as-

say. If validation was successful, we confirmed that the variant was transmitted in Centre d'Etude du Polymorphisme Humain pedigrees as a simple Mendelian trait. Confirmation is essential for distinguishing true allelic variants from false-positives produced by the assembly of ESTs from members of a gene family into a single contig and other artifacts.

As part of the confirmation procedure, we genetically mapped SNPs against a set of reference maps (Murray et al. 1994) by using the CRI-MAP 2.4 program (Green 1992). The first phase of genetic mapping involves assigning a polymorphism to a chromosome. This was accomplished by looking for pairwise linkage of the SNP and markers in the mapping panel via the CRI-MAP two-point option. In the absence of physical mapping data, LOD 4 (logarithm of odds) linkage to one genetic marker or LOD 3 linkage to two reference markers on the same chromosome is the minimum requirement for placing a SNP on a chromosome. If physical mapping information was available, LOD 3 linkage to a single reference marker on the same chromosome was sufficient for chromosome assignment. Once a SNP was assigned to a chromosome, we used the build option, which calculated a maximum likelihood score, to determine the "best" and "likely" map intervals. The best interval was that with the highest likelihood score, and the likely interval was the set of map intervals whose likelihood scores were within three orders of magnitude of the best score. Note that the best and likely locations may be identical and may encompass more than one marker interval.

Construction of Genetic/Physical Maps

Integrated genetic/physical maps for the 22 autosomes and the X chromosome are based on the CHLC/ABI (Cooperative Human Linkage Center/Applied Biosystems) Prism version 1 marker panel (http://www.chlc.org/ABI/ABIRefMaps.html) and the Gene-Map'98 version of the Genebridge4 radiation hybrid map (Gyapay et al. 1996; Deloukas et al. 1998; http://www.ncbi.nlm.nih.gov/genemap98). Genetic map distances between reference markers were obtained from the CHLC. Gender-averaged map positions were used for autosomal markers. Whenever possible, radiation hybrid map positions for reference markers were taken from GeneMap'98. Ninety-two of 359 markers, however, have not been localized on the reference physical map. To complete the linkage of the genetic and physical maps, we assigned these markers the radiation hybrid map position of a closely linked proxy marker. With the exception of D7S513, D11S987, and D12S336, proxies were chosen from the 1996 Genethon genetic map (Dib et al. 1996). For the above three markers, which are not on the Genethon map, proxies were chosen from the Marshfield Medical Center genetic map (Broman et al. 1998). To select a proxy

marker, we chose the closest marker on the genetic map (either proximal to or distal to the ABI version 1 marker), which has been positioned on the Gene-Map'98 Genebridge4 radiation hybrid map. With few exceptions, proxy markers lie within two centimorgans of the corresponding reference marker. The genetic and physical maps are colinear, with the exception of

inversions on chromosome 16 (markers D16S420 and D16S401), chromosome 18 (markers D18S462 and D18S70), and chromosome 20 (markers D20S173 and D20S171).

Placement of SNPs on the Integrated Map
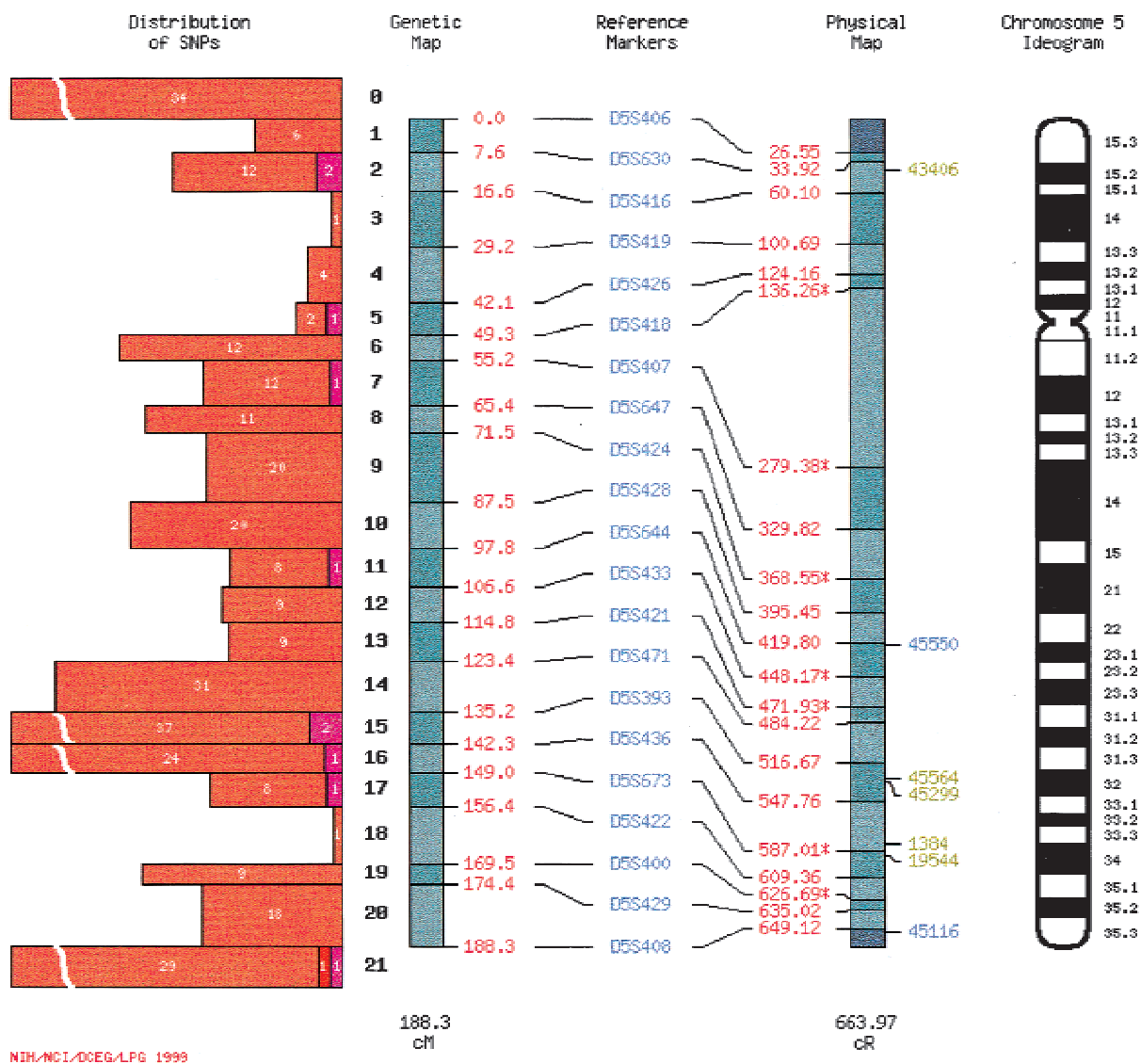Each SNP is associated with a template mRNA or ge-

Figure 1 SNP imagemap. Reference markers for the integrated map, which are from the CHLC/ABI Prism version 1 reference map (http://www.chlc.org/ABI/ABIRefMaps.html), are linked to CHLC annotation. Genetic map positions of reference markers, in red text, are to the right of the genetic map. Physical map positions of these markers, also in red text, are to the left of the physical map. Reference markers that have not been physically mapped are assigned the radiation hybrid map position of a closely linked proxy marker, indicated by a star next to the physical map position. Clicking on a starred physical map position opens a web page containing information about proxy markers. Numbers in boldface to the left of the genetic map denote genetic map intervals. The histogram to the left of the genetic map indicates the number of candidate (orange), validated (red), and confirmed (purple) SNPs mapping to each map interval. Clicking on a histogram bar opens an HTML summary page listing SNPs genetically and/or physically localized to that map interval. Physical map intervals are hyperlinked to web pages displaying the position of UniGene clusters with SNPs on the Genebridge4 radiation hybrid map. PCR primer set identification numbers of physically mapped; confirmed SNPs are located to the right of the physical map. Color coding is as follows: green (the physical map position lies within the best genetic mapping interval); blue (the physical map position lies within the likely genetic mapping interval); red (the physical map position lies outside the best and likely genetic mapping interval; not shown). PCR primer set identification numbers and the genetic map diagram are linked to the linkage imagemap web page (Fig. 2). Ideograms on are based on coordinates from Francke (1994).

nomic DNA sequence from one of the 21,993 parental UniGene clusters that served as the starting point of the SNP detection project. If the parental UniGene assembly has been mapped to a single chromosomal region (via one or more STSs within the cluster) on the GeneMap'98 Genebridge4 radiation hybrid map, we assign the SNP to the appropriate marker interval on the integrated genetic/physical map. If sequences within the UniGene cluster have been mapped to multiple locations on a single chromosome, we assign the mean map position to the SNP. If the cluster contains STSs that map to different chromosomes, we use cytogenetic mapping data for the cluster, if available, to resolve the inconsistency. Through this strategy, we positioned 6845 of 10,243 predicted SNPs on the integrated map.

In contrast to candidate and validated SNPs, both genetic and physical map data are used to place confirmed SNPs on the integrated map. If the physical map position (or a physical map position) of the parental UniGene assembly lies within the best or likely genetic map interval, we place the SNP in that marker interval. If no physical mapping data are available for the SNP or if the physical map position of the SNP does not correspond to the likely genetic map interval, we assign it to the lowest numbered marker interval within the best genetic map interval.

We also constructed expression-based integrated SNP maps. To determine whether a single nucleotide polymorphism lies in a gene expressed in one of the major sites of cancer (breast, colon, kidney, lung, liver, or prostate gland), we used information provided by the National Center for Biotechnology Information to ascertain whether one or more ESTs from the UniGene cluster associated with the SNP were isolated from a cDNA library derived from the tissue of interest. Therefore, only positive expression results are meaningful.

Features of the CGAP-GAI Web Site
SNP maps and related materials are accessible on the CGAP-GAI web site (http://cgap.nci.nih.gov/GAI/). Key features of the site are described below.

SNP Maps
SNP imagemaps (Fig. 1) contain a genetic map, physical map, and chromosome ideogram. A histogram adjacent to the genetic map shows the number of confirmed, validated, and candidate SNPs mapped to each marker interval. To the right of the physical map are PCR primer set identification numbers for confirmed SNPs that have been physically mapped. Each histogram and primer identification number is linked to a genetic map interval summary page (see below). Reference markers names are linked to annotation from the CHLC. In addition to a comprehensive map set containing every mapped SNP, we have generated expression-based sets of SNP maps. Our current set of tissue-specific maps show the locations of sequence variants identified in genes expressed in the breast, colon, kidney, liver, lung, and prostate.
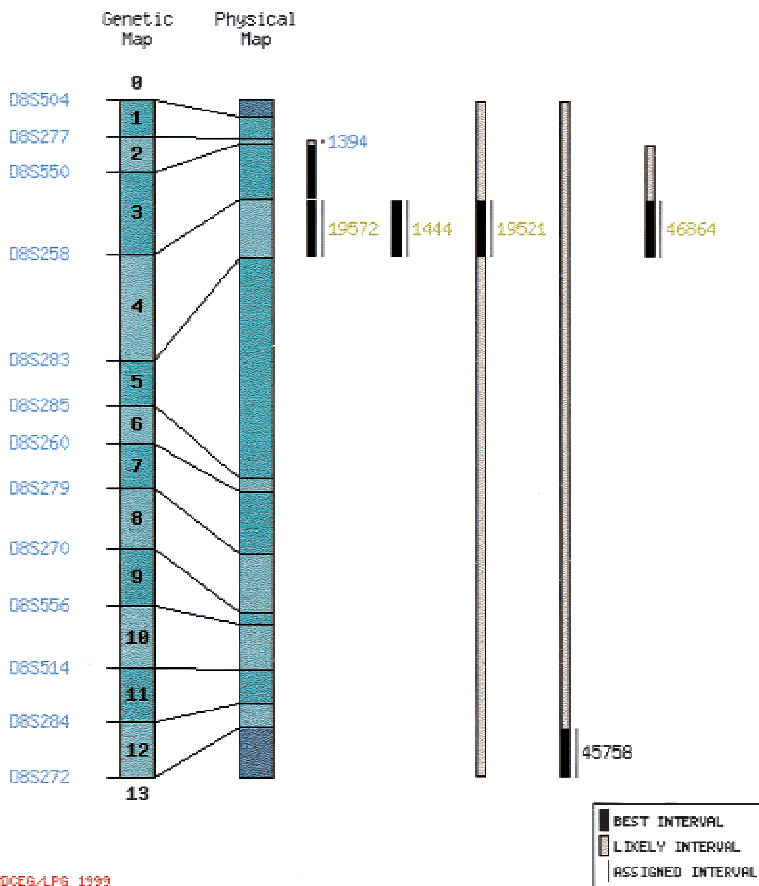
Figure 2 Linkage imagemap. Reference marker names are at the left margin of the figure. Numbers in boldface superimposed on the genetic map indicate genetic map intervals; clicking on a map interval number opens the corresponding SNP summary page. Bars to the right of the physical map diagram show the best (black) and likely (gray) genetic map interval for each SNP (see text for a description of best and likely intervals). Thin black lines to the right of the bars indicate the map interval to which we have assigned SNPs. PCR primer set numbers are color coded as described in Figure 1. In addition, black text indicates that the SNP has not been positioned on the physical map. Clicking on a PCR primer identification number opens the SNP summary page containing information about that SNP.

Linkage Maps

Linkage imagemaps (Fig. 2) show the best genetic mapping interval and the likely genetic mapping interval for each confirmed SNP in relation to the genetic/physical map. Primer set identification numbers identify SNPs. Genetic map interval numbers and primer identification numbers are linked to a genetic map interval summary page (see below).

Physical Maps

We provide a physical map of each interval of the integrated genetic/physical map. Physical maps display radiation hybrid mapped UniGene clusters containing candidate, validated or confirmed SNP as well as framework markers from the GeneMap'98 Genebridge4 map. Framework markers are hyperlinked to the corresponding GeneMap'98 chromosome map.

Summary Pages

Summary pages list confirmed, validated, and candidate SNPs. Information about each SNP also is provided; annotation includes the SNP identification number, a short description of the UniGene cluster associated with the SNP, the GenBank accession number of a template sequence from the UniGene assembly, and the gene symbol of the UniGene cluster if it corresponds to a named gene. Summary pages contain links to the SNP viewer (see below) and UniGene annotation, as well as RFLP and genetic mapping reports, where appropriate. Confirmed, validated, and candidate SNPs are listed separately on the summary pages, so SNPs within a single locus may be listed in three different locations on a page. In addition, we maintain a list of validated SNPs that have not been physically mapped.
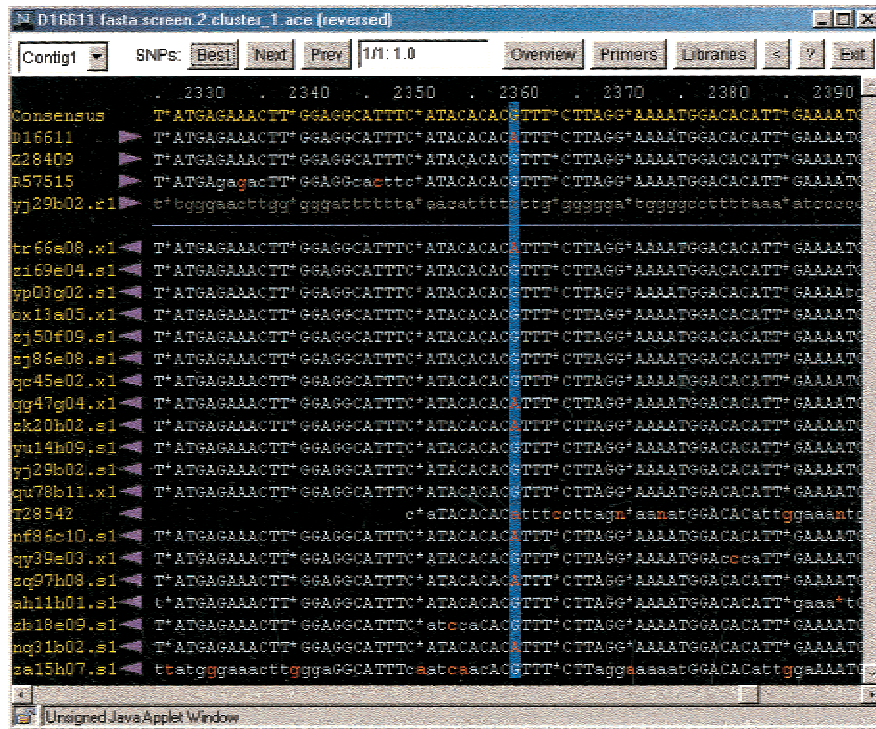

Figure 3 SNP viewer display of aligned sequences. In the window (black background), the consensus sequence of the assembly is displayed in yellow text. Gaps are indicated by asterisks. Coordinates of the assembly are above the consensus sequence in pale blue text. Individual sequences in the assembly are below in the consensus sequence. Sequence names are at the left in yellow text. Clicking on a sequence name launches a window containing a dbEST report for the sequence. The purple arrowhead to the right of a sequence name indicates the direction of the sequencing read. Sequences are to the right of the arrowheads. High quality basecalls are in uppercase pale blue text; low quality basecalls are in lowercase pale blue text; vector and low complexity sequences are in lowercase gray text. Basecalls differing from the consensus are in red. Clicking on a sequence launches a trace viewer. The candidate polymorphic site is indicated by dark blue vertical shading. In the panel at the top (white background), the Best, Next, and Prev buttons allow rapid navigation through the assembly to view candidate SNPs. The rank and probability score of the SNP are displayed in the small window at the center of the upper panel. The Overview button launches the assembly overview display (Fig. 4). The Primers button opens a link to the World Wide Web–based program Primer3 (Rozen and Skaletsky 1998), which can be used to design PCR primers for amplifying the region of DNA that contains the polymorphic site. The Libraries button opens a window listing the cDNA libraries that contain sequences defining the predicted SNP. Clicking the arrowhead button (<) reverses the contig.

SNP Viewer

Each SNP on a summary page is linked to a Java-based SNP viewer that displays two windows when launched. The first window shows the SNP in the context of a sequence alignment (Fig. 3), with minority residues at the polymorphic location shaded red. From this window, the user can retrieve additional information about the sequences in the assembly, obtain a list of cDNA libraries from which the sequences were derived, view the sequence traces, and access the PCR primer design program Primer3 (Rozen and Skaletsky 1998). The second window provides an overview of the sequence assembly (Fig. 4), displaying the locations of all SNPs in the assembly, SNP quality, contig depth, and position of the open reading frame in the assembly.

SNP Index

The SNP index search engine allows SNPs to be retrieved by keyword, GenBank accession number, or UniGene accession number. Search results are presented in a table that contains links to either a SNP summary page (for SNPs mapped on the Genebridge4 radiation hybrid map) or an integrated genetic/
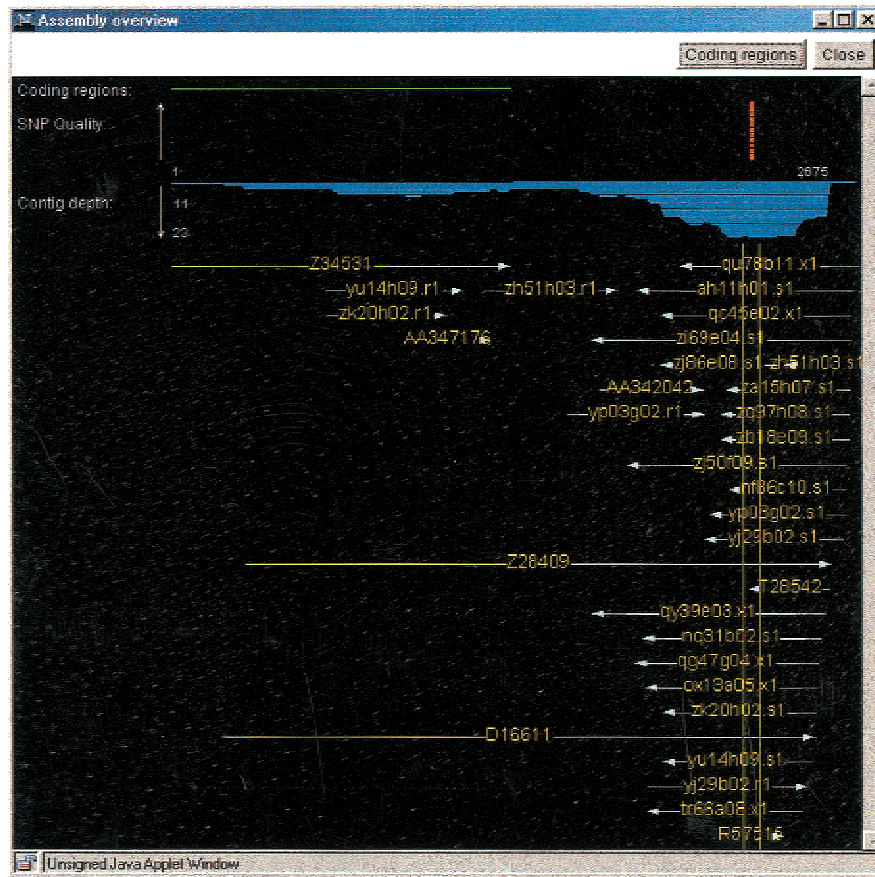
Figure 4 SNP viewer assembly overview display. Thin vertical green lines correspond to the portion of the assembly visible in the sequence-viewing window (Fig. 3). The overview display shows the physical position of each sequence within the assembly, the location of the open reading frame of the assembly, the location and quality of each SNP in the assembly, and contig depth. Clicking the Coding regions button adds horizontal green lines indicating the extent of the open reading frame in the contig and the mRNAs contained within the sequence assembly, and opens windows displaying mRNA protein translations.

physical map (for SNPs physically mapped by other means). The results table also is linked to the SNP viewer and the UniGene web site. Because only two-thirds of the candidate SNPs have been placed on the genetic/physical map, information about the remaining SNPs can be accessed via the search engine. The SNP index search engine also allows users to view assemblies that do not contain SNPs, thereby providing a graphical overview of EST coverage for a gene of interest.

SNP Lists
We also maintain information about SNPs as downloadable files. Candidate and confirmed SNPs in named genes are listed in hypertext format tables. The text file "all.fasta" displays each predicted SNP in the context of a published sequence, rather than an EST assembly consensus sequence; the set of published reference sequences is kept in the "summary.fasta" text file. SNP annotation is summarized in the tab-delimited "snps.all" file.

Cooperative Human Linkage Center
The CGAP-GAI home page contains a link to the CHLC web site, a repository of information about human genetic markers and genetic maps. The CHLC site contains information in a variety of formats about SNPs detected by the GAI.

SNP Finder
Access to our SNP detection tools for noncommercial use also is provided through the CGAP-GAI web site. Registered users may upload ABI or SCF format sequencing traces to our server for analysis. Submitted traces can be assembled with UniGene sequences to improve the sensitivity of SNP detection.

Future Directions
We plan to make a number of modifications to the SNP maps, including the incorporation of Stanford_G3 radiation hybrid map (Stewart et al. 1997) data that will increase the number of candidate and validated SNPs placed on the genetic/physical map. We also will construct a version of the SNP map based on the ABI Prism Linkage Mapping Set MD-10 (http://www2.perkin-elmer.com/ab/apply/dr/lmsv2/index.html). When the draft human genome sequence becomes available, we will incorporate this information into the integrated genetic/physical maps. We anticipate that users of the site will suggest many additional improvements.

METHODS

Platform
CGI (Common Gateway Interface) scripts and scripts that extract annotation from the NCBI UniGene web site, draw GIF (Graphic Interchange Format) images, and generate the HTML (Hypertext Markup Language) pages were written in Perl 5.005_02 (Wall et al. 1996). The LWP Perl module was used to access the NCBI web site, and the GD Perl module (http://stein.cshl.org/WWW/software/GD) was used to generate GIF images. The SNP viewer was written in Java 1.0 using the Linux port of Sun's Java Development Kit, version 1.0.2 (http://www.blackdown.org).

Availability
The CGAP-GAI site is accessible to the public at http://cgap.nci.nih.gov/GAI/. Pages displaying the integrated ge-

netic/physical maps and genetic map intervals contain client-side imagemaps that require Netscape 3.0 or Microsoft Internet Explorer 3.0 or higher. The SNP viewer is a Java applet, requiring a Java-capable browser.

Noncommercial use of the SNP finder is available to registered users. To obtain an account, contact K.H.B. (buetowk@nih.gov).

Data Storage and Retrieval

Integrated genetic/physical map pages and summary pages are maintained as static files. Linkage map pages and genetic mapping reports are generated via Perl CGI scripts from data in static files. Information in the CGAP-GAI site is indexed and searched using the Center for Networked Information Discovery and Retrieval Isearch-cgi 1.05 software (http://www.cnidr.org). The SNP viewer, SNP index, and RFLP report search engine are run from an Apache web server using the mod_perl extension (http://perl.apache.org). Data used by the SNP viewer and SNP index are retrieved from a Postgres database (http://www.quantum.de/~thh/postgres95/index.html) using the DBI Perl module (http://www.symbolstone.org/technology/perl/DBI/index.html), whereas the RFLP reports are stored as flat files.

ACKNOWLEDGMENTS

REFERENCES

Broman, K.W., Murray, J.C., Sheffield, V.C., White, R.L., and Weber, J.L. 1998. Comprehensive human genetic maps: Individual and sex-specific variation in recombination. *Am. J. Hum. Genet.* **63:** 861–869.

Buetow, K.H., Edmonson, M.N., and Cassidy, A.B. 1999. Reliable identification of large numbers of candidate SNPs from public EST data. *Nat. Genet.* **21:** 323–325.

Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., Patil, N., Lane, C.R., Lim, E.P., Kalayanaraman, N., Nemesh, J., et al. 1999. Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat. Genet.* **22:** 231–238.

Cavenee, W.K., Dryja, T.P., Phillips, R.A., Benedict, W.F., Godbout, R., Gallie, B.L., Murphree, A.L., Strong, L.C., and White, R.L. 1983. Expression of recessive alleles by chromosomal mechanisms in retinoblastoma. *Nature* **305:** 779–784.

Collins, F.S. 1995. Positional cloning moves from perditional to traditional. *Nat. Genet.* **9:** 347–350.

Cooper, D.N., Smith, B.A., Cooke, H.J., Niemann, S., and Schmidtke, J. 1995. An estimate of unique DNA sequence heterozygosity in the human genome. *Hum. Genet.* **69:** 201–205.

Deloukas, P., Schuler, G.D., Gyapay, G., Beasley, E.M., Soderlund, C., Rodriguez-Tomé, P., Hui, L., Matise, T.C., McKusick, K.B., Beckmann, J.S., et al. 1998. A physical map of 30,000 human genes. *Science* **282:** 744–746.

Dib, C., Faure, S., Fizames, C., Samson, D., Drouot, N., Vignal, A., Millasseau, P., Marc, S., Hazan, J., Seboun, E., et al. 1996. A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* **380:** 152–154.

Ewing, B. and Green, P. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8:** 186–194.

Ewing, B., Hillier, L., Wendl, M.C., and Green, P. 1998. Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. *Genome Res.* **8:** 175–185.

Francke, U. 1994. Digitized and differentially shaded human chromosome ideograms for genomic applications. *Cytogenet. Cell Genet.* **65:** 206–218.

Green, P. 1992. Construction and comparison of chromosome 21 radiation hybrid and linkage maps using CRI-MAP. *Cytogenet. Cell Genet.* **59:** 122–124.

Gyapay, G., Schmitt, K., Fizames, C., Jones, H., Vega-Czarny, N., Spillett, D., Muselet, D., Prud'Homme, J.F., Dib, C., Auffray, C., et al. 1996. A radiation hybrid map of the human genome. *Hum. Mol. Genet.* **5:** 339–346.

Halushka, M.K., Fan, J.-B., Bentley, K., Hsie, L., Shen, N., Weder, A., Cooper, R., Lipshutz, R., and Chakravarti, A. 1999. Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. *Nat. Genet.* **22:** 239–247.

Kruglyak, L. 1997. The use of a genetic map of biallelic markers in linkage studies. *Nat. Genet.* **17:** 21–24.

Kwok, P.-Y., Deng, Q., Zakeri, H., and Nickerson, D.A. 1996. Increasing the information content of STS-based genome maps: identifying polymorphisms in mapped STSs. *Genomics* **31:** 123–126.

Lander, E.S. 1996. The new genomics: global views of biology. *Science* **274:** 536–539.

Lander, E.S. and Schork, N.J. 1994. Genetic dissection of complex traits. *Science* **265:** 2037–2048.

Meyer, U.A. and Zanger, U.M. 1997. Molecular mechanisms of genetic polymorphisms of drug metabolism. *Ann. Rev. Pharmacol. Toxicol.* **37:** 269–296.

Murray, J.C., Buetow, K.H., Weber, J.L., Ludwigsen, S., Scherpbier-Heddema, T., Manion, F., Quillen, J., Sheffield, V.C., Sunden, S., and Duyk, G.M. 1994. A comprehensive human linkage map with centimorgan density. *Science* **265:** 2049–2054.

Picoult-Newberg, L., Ideker, T.E., Pohl, M.G., Taylor, S.L., Donaldson, M.A., Nickerson, D.A., and Boyce-Jacino, M. 1999. Mining SNPs from EST databases. *Genome Res.* **9:** 167–174.

Risch, N. and Merikangas, K. 1996. The future of genetic studies of complex human diseases. *Science* **273:** 1516–1517.

Schuler, G.D., Boguski, M.S., Stewart, E.A., Stein, L.D., Gyapay, G., Rice, K., White, R.E., Rodriguez-Tomé, P., Aggarwal, A., Bajorek, E., et al. 1996. A gene map of the human genome. *Science* **274:** 540–546.

Sherry, S.T., Ward, M., and Sirotkin, K. 1999. dbSNP—Database for single nucleotide polymorphisms and other classes of minor genetic variation. *Genome Res.* **9:** 677–679.

Stewart, E.A., McKusick, K.B., Aggarwal, A., Bajorek, E., Brady, S., Chu, A., Fang, N., Hadley, D., Harris, M., Hussain, S., et al. 1997. An STS-based radiation hybrid map of the human genome. *Genome Res.* **7:** 422–433.

Taillon-Miller, P., Gu, Z., Qun, L., Hillier, L., and Kwok, P.-Y. 1998. Overlapping genomic sequences: A treasure trove of single-nucleotide polymorphisms. *Genome Res.* **8:** 748–754.

Wall, L., Christiansen, T., and Schwartz, R.L. 1996. *Programming Perl*, 2nd ed. O'Reilly and Associates, Inc., Sebastopol, CA.

Wang, D.G., Fan, J.-B., Siao, C.-J., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., et al. 1998. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* **280:** 1077–1082.

Zhao, L.P., Aragaki, C., Hsu, L., and Quiaoit, F. 1998. Mapping of complex traits by single-nucleotide polymorphisms. *Am. J. Hum. Genet.* **63:** 225–240.

WWW RESOURCES

http://www-genome.wi.mit.edu/genome_software/other/primer3.html. Rozen, S. and Skaletsky, H.J. 1998. Primer3.