Advanced Options
Selecting the “Advanced Options” will open a list of PCR and sequencing primer design settings. These allow the user to specify settings for primer selection that are different from the default settings. Clicking the “Advanced Options” a second time will hide these values. Selecting the “CLEAR” button will refresh the application to the default settings.

Clear
Selecting the “CLEAR” button will refresh the application to the default settings.

Find Gene Name
“Find Gene Name” opens a link to the search page for determining the locus symbol for either human or mouse genes as specified by the user. The "Find Gene Name" search box allows simple Boolean searches. It allows AND, OR, NOT. It uses '*' to indicate a wildcard character and quotation marks to designate a phrase. Queries are submitted using the “Submit Query” button.

HapMap
Selection of the “HapMap” search filter directs the application to screen genetic polymorphisms for those validated polymorphisms that are also included in the HapMap haplotyping mapping project provided by dbSNP. There are currently 35,126 human polymorphisms associated with this attribute.

Limit by Population Criteria
Selection of the “Limit by Population Criteria” filter directs the application to screen genetic polymorphisms by geographic origin of the sample. Selection of more than one population filter will return those polymorphisms that meet either of the criteria. The number of polymorphisms associated with each population group is indicated on the application’s homepage. See Table of Geographic Populations Provided by dbSNP at the end of this document for the definition of the various population groupings.

Limit by Function Type
When a variation is near a transcript or in a transcript interval but not in the coding region, then the functional class is defined by the position of the variation relative to the structure of the aligned transcript. A variation may be near a gene (locus region), in a UTR (mRNA-UTR), in an intron (intron), in a splice site (splice site). If the variation is in a coding region, then the functional class of the variation depends on how each allele may affect the translated peptide sequence. Selection of more than one of the function type filters will return those polymorphisms that meet either of the criteria. The number of polymorphisms associated with each function type, for human and mouse, is indicated on the application’s homepage.
See Table of Functional Classifications Provided by dbSNP at the end of this document for the definition of the various function attributes.

Open Saved Results
“Open Saved Results” opens a link to the list of saved files available on the SOPv2 server. Selection of a file provides a tab-delimited document consisting of locus symbol, reference sequence number, allele, heterozygosity value, chromosome location, primer sequences, and size of expected PCR product.

PCR Primer Design Settings
“Melting Temperature (Tm) C” default setting is 60C. Tm is calculated using the relationship (Tm = 16.6*log[cation concentration] + 41*(fraction of GC) + 81.5, where the concentration of cation was estimated to be 0.1 M) described by Schildkraut,C. and Lifson,S. (1965) “Dependence of the melting temperature of DNA on salt concentration” Biopolymers, 3, 195-208. User directed changes to this value instruct the application to design primers to the indicated minimum Tm value.

“Minimum Foldover” filter is set at a default of 5. This value reflects the minimum allowable number of contiguous base paired residues within a candidate PCR primer.

“Unique n’mer Maximum Length” filter specifies the number of 3’-end residues that meet the requirement of only occurring once within 1,000 nucleotides flanking either side of the polymorphism.

“A/T test” filter specifies that at least one A or T residue will occur in the last three 3' nucleotides of the PCR primer.

“Residue Thresholds” filter specifies the frequency range for each A, T, C, G nucleotide residue in the candidate PCR primer.

“Flank length (maximum 1000)” setting denotes the length of DNA flanking the polymorphism that is used to screen for suitable PCR primers.

"Use RepeatMasker" indicates whether the results of RepeatMasker (repeatmasker.org) will be used. When selected, residues that are identified as being included in a "repeat region" will be masked off and will be not considered when making the PCR primer. Thus, primers will not anneal to regions of repetitive DNA. Using this option will generally decrease the number of primers found, but will better ensure that primers do not anneal to multiple sites.

“PCR Product Size” determines the range of length of acceptable amplified PCR product.

Print
“Print” provides a link to the user’s print command for use in obtaining a hardcopy of the results.

Reporting the Results
After submitting a search the application will present the results at the bottom of the page. Each report consists of 4 subsections, i.e., attribute bar, candidate primer bar, the DNA sequence that was evaluated, and simulated pyrosequencing data.

“Attribute bar” provides information linked to the polymorphism consisting of reference sequence number, polymorphism identifier, locus symbol, chromosome location, heterozygosity value, and function attribute. If the filters for “Validated SNPs”, “HapMap”, or “Limit by Population Criteria” were selected that information will also appear.

“Candidate primer bar” contains the oligonucleotide sequence (written 5’ to 3’) for the candidate PCR forward and reverse primers as well as the sequencing primer.

“DNA sequence that was evaluated” is indicated in the third panel of the results. The user may view the “full” sequence that was evaluated, consisting of the polymorphism and 1,000 flanking residues. DNA
sequence may also be viewed in FASTA format by selecting the “text” button. A “legend” button providing a key to the color-coding is also provided.

“Simulated pyrosequencing data” is indicated in the bottom panel of each result. A simulated pyrogram is drawn for each genotype. A nucleotide dispensation order for evaluating these data in the pyrosequencer is provided above the pyrograms. Informative dispensations are indicated in yellow.

**Saving the Results**

“Save as” opens a text box for designating a file name and a “save” button for uploading the results to the SOP3v2 server. The saved file can be viewed by selecting the “Open Saved Results” button located at the top of the page. Saved files contain a tab delimited list of PCR and sequencing primers as well as SNP attributes such as reference sequence number, heterozygosity value, and function.

**Search**

Selecting the “SEARCH” button will activate the application.

**Search by Location/Range**

To query the SOP3v2 database by chromosomal range use the check boxes to specify whether the search will focus on human or mouse genomes. Next, use the “Chromosome number” dropdown menu to specify which chromosome to use. Enter lower and upper limits for the nucleotide positions that should be screened. Queries are restricted to regions not exceeding 100,000 nucleotides. The maximum allowable upper limit for screening chromosome range is indicated in the table located at the end of the user guide. Click the “SEARCH” button to activate the application. Results will appear at the bottom of the page. Selecting the “CLEAR” button will refresh the application to the default settings.

**Search by SNP id or Locus Name**

To query the SOP3v2 database enter either a list of locus symbols or polymorphism reference sequence numbers. Queries are restricted to no more than 100 entries per batch. Use the check boxes provided to specify whether the query is for human or mouse genomes. Click the “SEARCH” button to activate the application. Results will appear at the bottom of the page. Selecting the “CLEAR” button will refresh the application to the default settings.

**Sequencing Primer Design Settings**

“Melting Temperature (Tm) C” default setting is 40C. Tm is calculated using the relationship \( Tm = 16.6 \log \text{[cation concentration]} + 41 \text{[fraction of GC]} + 81.5 \), where the concentration of cation was estimated to be 0.1 M) described by Schildkraut,C. and Lifson,S. (1965) “Dependence of the melting temperature of DNA on salt concentration” *Biopolymers*, 3, 195-208. User directed changes to this value instruct the application to sequencing primer to the indicated minimum Tm value.

“T+A Percentage” evaluates the design of the candidate sequencing primer for % AT content ensuring that this value falls within the specified range.

“Distance from SNP” provides a tool for specifying the minimum number of bases away from the SNP to initiate design of the sequencing primer.

**Sorting the Results**

“Sort by” provides the option to view the results by SNP ID number (Reference Sequence Number), average heterozygosity value, chromosome and position, locus symbol, or function attribute. Results may be viewed in ascending (^) or descending (v) order.

**Upload**

“Upload” opens a text box, “Browse” button, and “submit” button for entering a batch of locus symbols or reference sequence numbers from a file located on the user’s computer. Selection of the “Browse” button prompts the user to indicate the location of a carriage return delimited text file. Selection of the “submit” button uploads the query batch to the application’s main entry box. The application is initiated by selecting the “SEARCH” button.
Validated SNPs
Selection of the “Validated SNPs” search filter directs the application to screen genetic polymorphisms for inclusion in the list of validated polymorphisms provided by dbSNP. There are currently 384,831 human polymorphisms associated with this attribute.

View Validated Primers
"View Validated Primers" opens a link to the primer trios that have been made available for general use.

Contact Information
SOP³v2 is maintained by the Division of Immunogenetics at Children’s Hospital of Pittsburgh. Inquiries concerning customized searches of the database or reports of errors in the application should be addressed to Steven Ringquist (Email: smr73@pitt.edu).
### Table of Functional Classifications Provided by dbSNP

| Functional Class                  | Description                                                                                                                                                                                                 |
|----------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| no rna_acc/protein_acc           | Variation is within 2 Kb 5' or 500 bp 3' of a gene feature (on either strand), but the variation is not in the transcript for the gene. There are 357,606 human and 30,373 mouse records for this classification. |
| coding                          | Variation is in the coding region of the gene. This class is assigned if the allele-specific class is unknown. No mapped human or mouse polymorphisms currently use this classification. Therefore, it is not included in the SOP^3v2 database. |
| synonymous change                | An allele receives this classification when substitution and translation of the allele into the codon makes no change to the amino acid specified by the reference sequence. A variation is a synonymous substitution if all alleles are classified as contig reference or coding-synon. There are 48,702 human and 9,548 mouse records for this classification. |
| nonsynonymous change             | An allele receives this class when substitution and translation of the allele into the codon changes the amino acid specified by the reference sequence. A variation is a nonsynonymous substitution if any alleles are classified as coding-nonsynon. There are 61,963 human and 5,478 mouse records for this classification. |
| untranslated region              | The variation is in the transcript of a gene but not in the coding region of the transcript. There are 668,282 human and 43,414 mouse records for this classification.                                                                 |
| intron                           | The variation is in the intron of a gene but not in the first two or last two bases of the intron. There are 3,704,388 human and 213,968 mouse records for this classification.                                 |
| splice-site                      | The variation is in the first two or last two bases of the intron. There are 971 human and 23 mouse records for this classification.                                                                               |
| contig-reference                 | The variation allele is identical to the contig nucleotide. Typically, one allele of a variation is the same as the reference genome. Not used by SOP^3v2 for filtering alleles.                                               |
| coding: syn unknown              | The variation is in the coding region of a gene, but the precise location cannot be resolved because of an error in the alignment of the exon. There are 34 human and 9 mouse records for this classification.                        |

There are a total of 10,079,771 human and 581,577 mouse polymorphisms reported for dbSNP Build 123 and contained within the SOP^3v2 database. Of these a subset of 4,841,946 human and 302,813 mouse polymorphisms are associated with a function classification.
| Population Class | Description |
|------------------|-------------|
| Cent. Asia       | Samples from Russia and its satellite Republics and from nations bordering the Indian Ocean between East Asia and the Persian Gulf regions. |
| Cent/S. Africa   | Samples from nations south of the Equator, Madagascar, and neighboring island nations. |
| Cent/S. America  | Samples from Mainland Central and South America and island nations of the western Atlantic, Gulf of Mexico, and Eastern Pacific. |
| E. Asia          | Samples from eastern and south eastern Mainland Asia and from Northern Pacific island nations. |
| Europe           | Samples from Europe north and west of Caucasus Mountains, Scandinavia, and Atlantic islands. |
| Multi-Nations    | Samples that were designated to maximize measures of heterogeneity or sample human diversity in a global fashion. |
| N. America       | All samples north of the Tropic of Cancer, including defined samples of U.S. caucasians, African Americans, Hispanic Americans, and the National Human Genome Research Initiative (NHGRI) polymorphism discovery resource. |
| N/E Africa & Mid. East | Samples collected from North Africa (including the Sahara desert), East Africa (south to the Equator), Levant, and the Persian Gulf. |
| Pacific          | Samples from Australia, New Zealand, Central and Southern Pacific Islands, and Southeast Asian peninsula/island nations. |
| Unknown          | Samples with unknown geographic provinces that are not global in nature. |
| W. Africa         | Sub-Saharan nations bordering the Atlantic north of the Congo River and central/southern Atlantic nations. |
| Human Chromosome | Maximum Range |
|------------------|--------------|
| 1                | 245203898    |
| 2                | 243315028    |
| 3                | 199411731    |
| 4                | 191610523    |
| 5                | 180967295    |
| 6                | 170740541    |
| 7                | 158431299    |
| 8                | 145908738    |
| 9                | 134505819    |
| 10               | 135480874    |
| 11               | 134978784    |
| 12               | 133464434    |
| 13               | 114151656    |
| 14               | 105311216    |
| 15               | 100114055    |
| 16               | 89995999     |
| 17               | 81691216     |
| 18               | 77753510     |
| 19               | 63790860     |
| 20               | 63644868     |
| 21               | 46976537     |
| 22               | 49476972     |
| X                | 152634166    |
| Y                | 50961097     |