

Population genetics and the interpretation of Forensic DNA data - FROG-kb

http://frog.med.yale.edu/FROG-kb

Kenneth Kidd, PhD¹, Haseena Rajeevan, PhD¹, Katherine Moore, MS², Usha Soundararajan, MS¹, Richard Satcher, MS², Patricia Melton, PhD², Andrew Pakstis, PhD¹, and Jeri Roper-Miller, PhD²
¹Yale University School of Medicine, New Haven, CT; ²RTI International, Research Triangle Park, NC

Abstract

It is well recognized that individuals with ancestries from different parts of the world will have different allele frequencies. However, the information is usually very poorly incorporated into the interpretation of forensic data. We have developed the Forensic Research/Reference on Genetics knowledge base (FROG-kb) to assist in more rigorous interpretation for forensic data on SNPs. FROG-kb (<http://frog.med.yale.edu/FrogKB/>) is a freely accessible online database that provides tools for comparing user-provided data with underlying allele frequencies in populations, and it serves as a resource for teaching and research. Currently the SNP panels in FROG-kb emphasize Individual Identification (IISNPs) and Ancestry Inference (AISNPs). One Phenotype Interference (PISNP) set for inference of eye color is also included. SNPs can be valuable for investigative leads when data on the standard Short Tandem Repeat Polymorphisms (STRPs) are not available or no hits in such databases as the Combined DNA Index System (CODIS) are identified. When sample DNA is highly degraded SNPs will yield results when many STRPs do not. In such cases, the conventional problem of individual identification and matching samples can easily use SNPs with match probabilities of 10E-30 easily achievable. The investigative leads provided by SNPs can also include ancestry inference and individual identification. However, interpretation of the data on the published SNP panels require reference population data; such reference allele frequency data are provided in the FROG-kb database with the ability to calculate population specific random match probabilities and inference of which reference populations are the most likely ancestry for a forensic unknown. The FROG-kb web site went online in 2013 with most of its current functionality. It has recently undergone a major update in its interface and underlying data resources. The web site enhancements provide a more organized and easily navigated resource. There is also a comprehensive user manual that provides users with step-by-step directions on how to use and search FROG-kb for the different SNP panels. This poster provides an overview of FROG-kb design and functionality and examples of the use of FROG-kb.

FROG-kb Features

Figure 1a is a snapshot of the home page of FROG-kb. The buttons across the top of the page provide the main navigation tools. The pages for actual calculations for IISNPs, AISNPs, and PISNPs are entered through the respective buttons. Because FROG-kb is being continually updated, "Announcements" is used to note new functionality and new resources. Clicking on the "Documentation" button opens a page with several options:

- ◆ **Functionalities:** provides an overview for navigating through and using the SNP panel information, SNP panel example data, SNP panel data entry, and interpreting results of the calculations on the data entered.
- ◆ **Pipeline:** opens a graphical representation that illustrates the process for inputting, searching, and retrieving data from FROG-kb (shown in Figure 2 for IISNP)
- ◆ **FAQ:** presents Frequently Asked Questions about FROG-kb and answers.
- ◆ **Manual:** provides details on navigating, data entry, and searching the database.
- ◆ **Resources:** provides users with links (STR- and SNP-based) and references to potentially relevant databases and sites. This page also lists poster presentations pertaining to forensic use of SNPs, including FROG-kb, that have been presented at various conferences and that are freely accessible to users.

Figure 1a. FROG-kb website with Main Menu options



Welcome to the New FROG-kb

This new FROG-kb interface has all of the previous functionalities with supplementary reference material and "space" for additional documentation that will be added over the next few weeks. See [Announcements](#) for updates. New FROG-kb can be best viewed on latest Mozilla Firefox and Google Chrome. Some versions of Internet Explorer are not compatible. If you come across any issues or have suggestions use the [Contact Us](#) function to reach us. "If you see something - say something!"

FROG-kb seeks to make allele frequency data for SNPs and other genetic polymorphisms more useful in a forensic setting. The primary objective of FROG-kb is to provide a web interface that, from a forensic perspective, is useful for teaching and research and can serve as a tool facilitating forensic practice. The underlying data are housed in the already extensively used and referenced Allele Frequency Database (ALFRED). The FROG-kb interface makes the information usable for forensic purposes, including computational tools for comparing user-provided data with underlying allele frequencies in populations. These tools are organized by the methodology followed and the different published SNP/marker panels available.

FROG-kb considers three types of SNP panels: Individual Identification (IISNPs), Ancestry Inference (AISNPs), and Phenotype Interference (PISNPs). IISNP and AISNP panels provide the ability to enter genotypes of an individual at multiple SNPs and calculate probabilities of that multilocus genotype in each of several populations. The current PISNP panel is an eye color prediction function that provides relative probabilities of different eye colors. More information on each SNP panel can be found in Section XX of the User Manual along with information on searching each SNP panel within FROG-kb.

© FROGkb 2016 Contact Us Notice of Funding

Figure 1b. Options available from the homepage Documentation tab

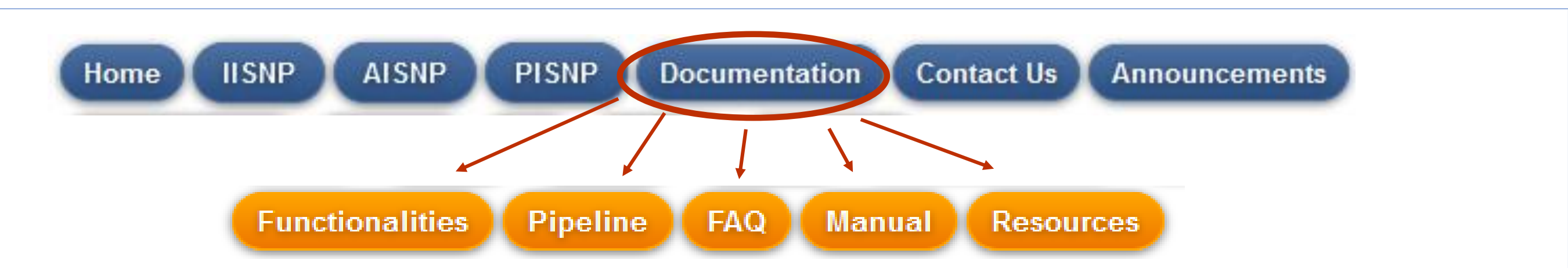


Figure 1c. Options available from the Panels tab



References

Rajeevan H, Soundararajan U, Pakstis AJ, Kidd KK. Introducing the Forensic Research/Reference on Genetics knowledge base, FROG-kb. Investigative genetics. 2012; 3(1):18.

Rajeevan H, Soundararajan U, Kidd JR, Pakstis AJ, Kidd KK. ALFRED: an allele frequency resource for research and teaching. Nucleic acids research. 2012; 40(Database issue):D1010-5.

ALFRED database url: <https://alfred.med.yale.edu/alfred>

FROG-kb database url: <http://frog.med.yale.edu/FrogKB/>

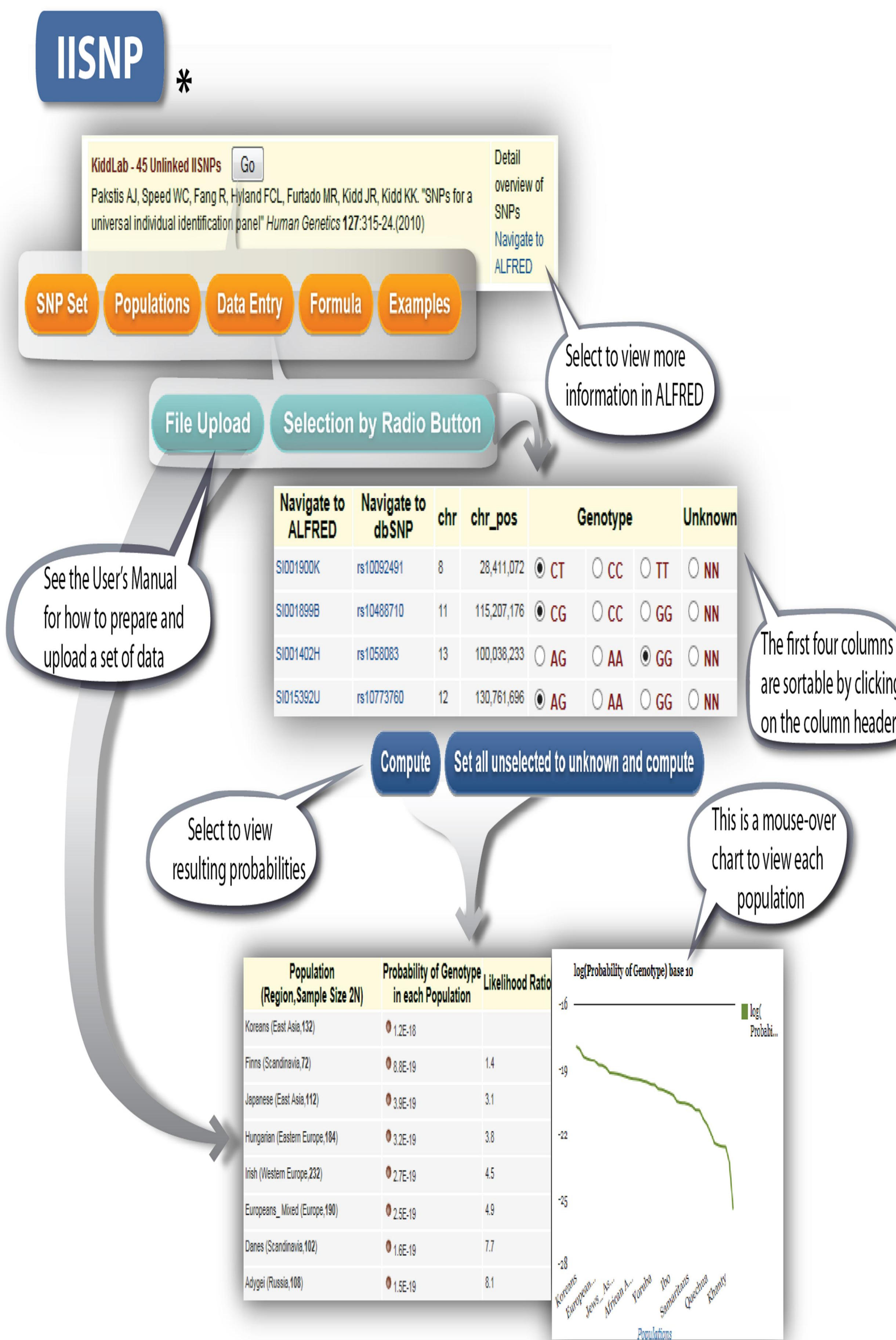
PISNP data

Phenotype Informative SNP Panels (PISNPs) provide the ability to predict a visible trait for user-specified genotypes for the relevant SNPs. FROG-kb currently has the Kayser's 6-plex for eye color prediction - IrisPlex panel implemented.

IISNP Data Entry and Data Output

IISNPs provide the ability to calculate population-specific random match probabilities for user defined genotypes. **Figure 2** provides an overview of data entry and data computation for the Kidd Lab - 45 Unlinked II SNPs. **Figure 3** shows the output result based on the IISNP panel calculation.

Figure 2. Overview of IISNP data entry and data output.



* The diagram shown is only an example for IISNP data entry, although you will find the same functionalities for AISNP data entry. More information on IISNP and AISNP can be found in the User's Manual

Figure 3. Eg. Calculation result output based on KiddLab 45 IISNP panel for a Korean

KiddLab - 45 Unlinked IISNPs

Population likelihoods based on 45 SNPs and 43 reference populations for the DNA profile: korean

Print Table Format View SNPs Used

Indicates the values are within an order of magnitude of the highest likelihood.

Population (Region, Sample Size 2N)	Probability of Genotype in each Population	Likelihood Ratio
Koreans (East Asia, 132)	1.224E-18	
Finns (Scandinavia, 72)	8.83E-19	1.39
Japanese (East Asia, 112)	3.943E-19	3.1
Hungarian (Eastern Europe, 184)	3.204E-19	3.82
Irish (Western Europe, 232)	2.729E-19	4.49
Europeans_Mixed (Europe, 190)	2.506E-19	4.88
Danes (Scandinavia, 102)	1.594E-19	7.68
Adygei (Russia, 108)	1.516E-19	8.07
Hakka (East Asia, 86)	1.169E-19	10.5
Lao Loum (Southeast Asia, 238)	6.975E-20	17.5
Jews_Ashkenazi (Southwest Asia, 166)	6.677E-20	18.3
Han (East Asia, 124)	6.162E-20	19.9
Han (East Asia, 100)	5.539E-20	22.1

Table 1. List of currently available II panels in FROG-kb

II panels in FROG-kb	Populations Included for likelihood calculation
KiddLab - 45 Unlinked IISNP	43
SNPforID 52-plex	20
Qiagen Investigator DIPplex kit	28

* Reference citations respective to each panel available in the database

AISNP Data Output

AISNPs provide the ability to calculate relative likelihoods of ancestry from different reference populations for user-specified genotype. The functionalities shown in **Figure 2** for the IISNPs are the same for the AISNP data entry and data computation.

Formula and Results of the Calculation:

For the IISNP and AISNP panels the calculations start by assuming Hardy-Weinberg proportions of the genotypes based on the allele frequencies available for each SNP in each population. The probability of each genotype is stored after being pre-calculated from the allele frequencies in ALFRED database. For the IISNP panels these are the population-specific match probabilities. For the AISNP panels these are the probabilities of the input multi-locus genotype arising in the different populations and can be considered the relative likelihoods of the populations being the origin of the multi-locus profile.

The results of the calculation are displayed as a table with three columns: the name of the population sampled with its geographic region and the sample size, the probability of the entered genotype in that population, and the likelihood ratio of the most probable population to the specified population. **Figure 4a** and **4b** illustrates the output result page for the AISNP calculation.

Figure 4a. Calculation result output based on Kidd Lab 55 AI SNP panel for a Maya individual

KiddLab - Set of 55 AISNPs

Population likelihoods based on 55 SNPs and 133 reference populations for the DNA profile: Maya

Print Table Format View SNPs Used

Indicates the values are within an order of magnitude of the highest likelihood.

Population (Region, Sample Size 2N)	Probability of Genotype in each Population	Likelihood Ratio
Ticuna (South America, 134)	7.557E-13	
Maya_Yucatec (North America, 106)	4.214E-13	1.79
Peruvian Quechuan (South America, 44)	1.253E-13	6.03
Gulhiba speakers (South America, 24)	1.219E-13	6.2
Peruvian (PEL) (South America, 170)	6.415E-14	11.8
Karitiana (South America, 114)	5.08E-15	149.0
Surui_Rondonia (South America, 100)	3.91E-16	1930.0
Mexican Pima (North America, 106)	1.096E-17	6900.0
Mongols (Inner-Mongolia) (Asia, 200)	2.498E-21	3.06E8
Outer Mongolians (Asia, 136)	2.294E-21	3.29E8
Yakut (Siberia, 102)	5.757E-23	1.31E10
Hui (Ningxia) (East Asia, 198)	4.065E-23	1.86E10

Figure 4b. Calculation result output based on Kidd Lab 55 AI SNP panel for a Finn individual

KiddLab - Set of 55 AISNPs

Population likelihoods based on 55 SNPs and 133 reference populations for the DNA profile: Finn

Print Table Format View SNPs Used

Indicates the values are within an order of magnitude of the highest likelihood.

Population (Region, Sample Size 2N)	Probability of Genotype in each Population	Likelihood Ratio
Danes (Europe, 102)	2.426E-11	
Finns (Europe, 72)	9.13E-12	2.66
Mixed Europeans (Europe, 190)	8.542E-12	2.84
Irish (Europe, 232)	5.275E-12	4.6
Finns (FIN) (Europe, 198)	1.034E-12	23.5
Mixed Europeans (CEU) (Europe, 198)	9.141E-13	29.5
British (GBR) (Europe, 182)	5.511E-13	44.0
Russians_Archangelski (Europe, 68)	3.906E-13	62.1
Russians (Europe, 96)	3.582E-13	67.7
Komi-Zyryan (Asia, 94)	3.24E-13	74.9
Hungarians (Europe, 184)	1.433E-13	169.0
Italians (IT) (Europe, 214)	2.498E-14	972.0
Toscani (TS) (Europe, 214)	6.295E-15	3850.0
Chuvash (Europe, 84)	2.35E-15	10300.0
Greeks (Europe, 104)	1.457E-15	16700.0
Ashkenazi Jews (Europe, 166)	1.774E-16	137000.0

Table 2. List of currently available AI panels in FROG-kb

AI panels in FROG-kb	Populations Included for likelihood calculation
Seldin's list of 128 AISNPs	70
SNPforID 34-plex	53
KiddLab - Set of 55 AISNPs	133
Kayser's set of 24 AI Markers	73
Daniele Podini's list of 32 AISNPs	111
Eurasiaplex 23 SNP Panel	76
Nievergelt's Set of 41AImS	123
Overlap set of AISNPs	72

* Reference citations respective to each panel available in the database

Acknowledgements

◆ This work is supported in part by grants awarded to K. Kidd: grant 2014-DN-BX-K030 awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice and by grant BCS-1444279 awarded by the U.S. National Science Foundation.

◆ Research Triangle Institute International (RTI) assisted the development of the new interface by providing feedback and documentation including user's manual, and pipeline, and by promoting and familiarizing the forensic community with the FROG-kb database through FTCoE. RTI was awarded the Forensic Technology Center of Excellence (FTCoE; NIJ award 2011-DN-BX-K564) to provide knowledge transfer and technology adoption to the criminal justice and forensic science communities.

◆ Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the US Department of Justice.