



FROG-kb: Forensic Resource/Reference on Genetics—knowledge base

<http://frog.med.yale.edu/>

Haseena Rajeevan^{1,2}, Usha Soundararajan¹, Kenneth K. Kidd¹
¹Department of Genetics and ²Center for Medical Informatics

This work is supported by National Institute of Justice Grant : 2010-DN-BX-K226

Overview

FROG-kb (<http://frog.med.yale.edu>) is an open access web application that is useful for teaching and research relevant to forensics and can serve as a tool facilitating forensic practice. The underlying data for FROG-kb are provided by the already extensively used and referenced Allele Frequency Database, ALFRED (<http://alfred.med.yale.edu>). In addition to displaying data in an organized manner, computational tools that use the underlying allele frequencies with user-provided data are implemented in FROG-kb. These tools are organized by the different published SNP/marker panels available. This web tool currently has implemented general functions for two types of SNP panels, individual identification and ancestry inference, and a prediction function specific to a phenotype informative panel for eye color.

FROG-kb Forensic Resource On Genetics knowledge base

Home

This is a pilot implementation of our new database and web interface. The structure and functionality of FROG-kb are being revised in an ongoing basis.

More background can be found under ABOUT. We have so far only implemented three functions; two general functions: the ability to enter genotypes of an individual at multiple SNPs and calculate likelihoods of that multilocus genotype in each of several populations, and an eye color prediction function specific to a PISNP panel. These functions are possible for three types of SNP panels, IISNPs, AISNPs, and PISNPs, described below. You can navigate by selecting options on the main menu at the left, as well as by links and buttons that appear on different pages.

- 1) For **Individual Identification SNPs (IISNPs)** this implementation provides examples and the ability to calculate match probabilities for user-specified genotypes. Two different published panels of IISNPs can be used to determine the probability of the user-specified genotype in each of several populations that have allele frequencies available for all SNPs in the panel. Click on the IISNP button on the left to use this function.
- 2) For **Ancestry Inference SNPs (AISNPs)** this implementation provides examples and ability to calculate relative likelihoods of ancestry from different populations for user-specified genotypes. Two published AISNP panel can be used as well as a provisional panel of 39 SNPs assembled specifically to test this functionality. Click on the AISNP button on the left to use this function.
- 3) For **Phenotype Informative SNPs (PISNPs)** this implementation provides ability to predict eye color for user-specified genotypes. One published PISNP panel can be used to test the eye color prediction functionality which is based on a published formula (Liu et al. & Walsh et al.). Click on the PISNP button on the left to use this function. We are implementing more SNP panels associated with various phenotypes.

FROG-kb is supported by National Institute of Justice grant 2010-DN-BX-K226

PISNP Sets

Kayser's 6-plex for eye color prediction - IrisPlex Go

Walsh S, Liu F, Ballantyne KN, van Oven M, Lao O, Kayser M. "IrisPlex: a sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information" *Forensic Sci Int Genet* 5:170-80.(2011)

Detail overview of SNPs
Navigate to ALFRED

Kayser's 6-plex for eye color prediction - IrisPlex

SNP Set | Data Entry | Hungarian | Keraites | Irish | Korean | Formula

KAYSER'S IRISPLEX PISNPs - PRESELECTED FOR AN INDIVIDUAL FROM HUNGARIAN SAMPLE

rsnumber	chr	chr_pos	Genotype	Unknown			
rs135310N	6	396,321	CC	CT	TT	CN	NN
rs168220T	14	92,773,663	GT	GG	TT	CN	NN
rs007119S	15	28,365,618	AG	GG	AA	CN	NN
rs196972I	11	89,011,046	AG	GG	AA	CN	NN
rs000363V	5	33,951,693	CC	GG	CG	CA	CN
rs663684H	15	28,230,318	CC	CT	TT	CN	NN

Set all unselected to unknown | Predict Eye Color

Print Format

Kayser's Eye Color Prediction - User Selected Genotype

rsnumber	genotype
rs12203592	CC
rs12896399	GT
rs12913832	AG
rs1293350	AG
rs16891982	GG
rs1800407	CC

Eye Color Predicted Probability

Blue	0.2095
Intermediate	0.1519
Brown	0.6386

CLICK HERE TO VIEW FORMULA

CLICK ON DATAENTRY BUTTON TO VIEW YOUR INPUT

The above calculation is based on the formula published (Liu et al. & Walsh et al.) on eye color prediction.

Liu F, van Duijn K, Vingerling JR, Hofman A, Uitterlinden AG, Janssens AC, Kayser M. Eye color and the prediction of complex phenotypes from genotypes. *Curr Biol*. 19(5):192-193 (2009)

Walsh S, Liu F, Ballantyne KN, van Oven M, Lao O, Kayser M. IrisPlex: a sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. *Forensic Sci Int Genet* 5:170-80.(2011)

AISNP Sets

Pilot Panel of 39 AISNPs from KiddLab Go

Kenneth K. Kidd et al. "Data unpublished"

Detail overview of SNPs
Navigate to ALFRED

Seldin's list of 128 AISNPs Go

Kosoy R, Nassir R, Tian C, White PA, Butler LM, Silva G, Kittles R, Alarcon-Riquelme ME, Gregersen PK, Belmont JW, De La Vega FM, Seldin MF. "Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America" *Hum Mutat* 30:69-78.(2009)

Detail overview of SNPs
Navigate to ALFRED

Kidd JR, Friedlander FR, Speed WC, Pakstis AJ, De La Vega FM, Kidd KK "Analyses of a set of 128 ancestry informative single-nucleotide polymorphisms in a global set of 119 population samples" *Investigative Genetics* 2:1.(2011)

SNPforID 34-plex Go

Phillips C, Salas A, Sánchez JJ, Fondevila M, Gómez-Tato A, Álvarez-Dios J, Calza M, Casares de Cal M, Ballard D, Lareu MV, Carracedo A - The SNPforID Consortium "Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs" *Forensic Science International: Genetics* 1:273-280.(2007)

Detail overview of SNPs
Navigate to ALFRED

Other SNPsets coming soon:
-Additional Ancestry Informative SNPs from KiddLab
-Qiagen Investigator DIPlex 30 INDEL panel

IISNP functionality

For each IISNP panels listed below there are several elements. Associated with each of the panel listings is a Detailed Overview of SNPs link into ALFRED precisely for a more detailed overview of the panel. That link opens the SNP Set page within ALFRED into a new browser window. The SNP Set module in ALFRED has multiple options, including the ability to see for each SNP a pie chart on Google Maps of frequencies for all populations with data.

After clicking on the GO button to enter an IISNP Set, several options are possible. Each has a link to the list of the SNPs in the panel. Within that SNP Set list each rs number is an active link to the dbSNP record for that SNP. The Populations button provides the list of populations for which comparable calculations can be made. This is the set of populations for which all SNPs in the set have allele frequency data. Note, within the SNP Set in ALFRED additional populations may have data for some, but not all populations; those populations are not included in the calculations. Each population name in this SNP Set list within FROG is an active link to information on the population stored within ALFRED; that page will open in a new browser window. Example options are also accessible using Ex 1 or similar buttons. These are screen shots to provide examples, but seem to load slowly—we are working on that.

More important is Data Entry that opens the ability to specify an individual multi-site genotype and then calculate the probability of that genotype in each of the populations. The genotype is entered by simply clicking on the radio button for the genotype at each SNP. Note, it is not necessary to click on the "NN" for missing data. At the bottom of the list are three buttons: Set all unselected to unknown, Print Format and Compile. The Print Format will generate a condensed version of the input data that can be printed as a permanent record of the input data. The Compile will initiate calculation and display the results. If there are SNPs with no selection, a warning will be sent and the option exists to examine which SNPs have no entry and to either enter a genotype or use the Set all unselected to unknown to fill those with "NN". Afterward, it is necessary to click on Compile again.

There are also buttons with population names for example Pima Mexican, Korean that will open a pre-entered data entry page for one individual from the specified population. Those should be tried before experimenting by entering a new genotype profile of an unknown or a forensic case.

Functionalities

KiddLab - 45 Unlinked IISNPs Go

Pakstis AJ, Speed WC, Fang R, Hyland FCL, Furtado MR, Kidd JR, Kidd KK. "SNPs for a universal individual identification panel" *Human Genetics* 127:315-24.(2010)

Detail overview of SNPs
Navigate to ALFRED

SNPforID 52-plex Go

Sánchez JJ, Phillips C, Borsting C, Balogh K, Bogus M, Fondevila M, Harrison CD, Musgrave-Brown E, Salas A, Syndercombe-Court D, Schneider PM, Carracedo A, Morling N. "A multiplex assay with 52 single nucleotide polymorphisms for human identification" *Electrophoresis*. 27:1713-1724.(2006)

Detail overview of SNPs
Navigate to ALFRED

SNP Set | Ex 2 | Populations | Data Entry | Clear Selection | Korean

Functionalities | Formula

How the probabilities are derived

For the IISNP and AISNP panels the formula used is as follows:
The probability of observing the given genotype at each locus examined in a population is

$$P_x(\text{homozygous allele 1}) = p^2$$

$$P_x(\text{heterozygous match}) = 2pq$$

$$P_x(\text{homozygous allele 2}) = q^2$$

These values are precalculated and stored in the database based on the allele frequencies in ALFRED for the locus and population.
Suppose the values for population P₁ are P₁(A), P₁(B) at SNP A and B,
the match probability of the genotypes in population P₁ is obtained by using the product rule,
P₁(A & B) = P₁(A) * P₁(B)

Panel of 45 IISNPs

SNP Set | Ex 2 | Populations | Data Entry

Functionalities | Formula

PANEL OF 45 IISNPs - SNP SET

rsnumber	chr	chr_pos
rs10072491	8	28,411,072
rs10488710	11	115,207,176
rs1058083	13	100,038,233
rs10773760	12	130,761,696
rs10776839	9	137,417,308
rs1109037	2	10,085,722
rs1294331	1	233,448,413
rs13218440	6	12,059,954
rs1338071	6	94,537,255
rs1498553	11	5,709,028

PANEL OF 45 IISNPs - POPULATIONS Geographic Region Map

Populations included in computing match probability	Geographic Region	Sample Size (2N)
Adyghe	Russia	108
African Americans	Africa	182
Ami	East Asia	80
Atayal	East Asia	84
Biaka	Central Africa	140
Cambodians, Khmer	Southeast Asia	52
Chagga	Eastern Africa	90
Chuvash	Russia	84
Danes	Scandinavia	102
Druze	Southwest Asia	212
Europeans, Mixed	Europe	190
Finn	Scandinavia	72
Hakka	East Asia	86
Han	East Asia	100
Han	East Asia	124
Hansa	Western Africa	78
Hungarian	Eastern Europe	184
Ibo	Western Africa	96
Irish	Western Europe	232
Japanese	East Asia	112
Jews, Ashkenazi	Southwest Asia	166

Link to Population page in ALFRED

Data input and output screen for IISNPs

SNP Set | Ex 2 | Populations | Data Entry | Clear Selection | Korean

Functionalities | Formula

PANEL OF 45 IISNPs - PRESELECTED FOR AN INDIVIDUAL FROM KOREAN SAMPLE

rsnumber	chr	chr_pos	Genotype	Unknown			
rs10072491	8	28,411,072	CC	CT	TT	CN	NN
rs10488710	11	115,207,176	CC	CG	GG	CN	NN
rs1058083	13	100,038,233	AG	GG	AA	CN	NN
rs10773760	12	130,761,696	AG	GG	AA	CN	NN
rs10776839	9	137,417,308	GT	GG	TT	CN	NN
rs1109037	2	10,085,722	AG	GG	AA	CN	NN
rs1294331	1	233,448,413	AG	GG	AA	CN	NN
rs13218440	6	12,059,954	AG	GG	AA	CN	NN
rs1338071	6	94,537,255	AG	GG	AA	CN	NN
rs1498553	11	5,709,028	AG	GG	AA	CN	NN

Set all unselected to unknown | Compile

Print Format

BASED ON 45 SNPs View SNPs Used

Print Table Format | Geographic Region Map

Population (Region, Sample Size 2N)	Probability of Genotype in each Population
Koreans (East Asia, 90)	1.026e-17
Finn (Scandinavia, 60)	7.894e-18
Japanese (East Asia, 90)	3.304e-18
Hungarian (Eastern Europe, 60)	2.779e-18
Irish (Western Europe, 60)	2.326e-18
Europeans_Mixed (Europe, 60)	2.213e-18
Danes (Scandinavia, 60)	1.501e-18

CLICK ON DATA ENTRY TO VIEW YOUR INPUT

log(Probability of Genotype) base 10

Koreans log(Probability of Genotype) base 10: -16.99

Populations