

# **HIRISPLEX-S, HIRISPLEX & IRISPLEX Eye, Hair and Skin colour DNA Phenotyping webtool USER MANUAL**

## **Table of Contents:**

### **Eye, hair and skin colour prediction using *HirisPlex-S***

#### *Single Individual*

---

Online Interface for on-screen result 1

#### *Multiple Individuals*

---

Upload File 4

### **Eye and hair colour prediction using *HirisPlex***

#### *Single Individual*

---

Online Interface for on-screen result 7

#### *Multiple Individuals*

---

Upload File 9

### **Eye colour prediction using *IrisPlex***

#### *Single Individual*

---

Online Interface for on-screen result 11

#### *Multiple Individuals*

---

Upload File 13

Interpreting the Prediction Outcomes 15

Understanding the Partial Prediction Result and AUC Loss 16

Online Database Information and Model Performance 18

Independent Performance Set 19

Acknowledgements 20

Disclaimer 20

Copyright 20

Citation 21

References 21

## Eye, Hair and Skin Colour Prediction using HirisPlex-S


### Single Individual

#### Online Interface for on-screen result

To use the HirisPlex-S portion of this website to predict the eye, hair and skin colour of an individual, please follow these instructions.

You can see the HirisPlex-S portion of the website below. All 41 gene names, single nucleotide polymorphisms (SNPs) and model input alleles belonging to the HirisPlex-S system are displayed.

**The HirisPlex-S System**



	Gene	SNP	Allele	No. of Alleles
1	MC1R	rs312262906	A	0 1 2 NA
2	MC1R	rs11547464	A	0 1 2 NA
3	MC1R	rs885479	T	0 1 2 NA
4	MC1R	rs1805008	T	0 1 2 NA
5	MC1R	rs1805005	T	0 1 2 NA
6	MC1R	rs1805006	A	0 1 2 NA
7	MC1R	rs1805007	T	0 1 2 NA
8	TUBB3	rs1805009	C	0 1 2 NA
9	MC1R	rs201326893	A	0 1 2 NA
10	MC1R	rs2228479	A	0 1 2 NA
11	MC1R	rs1110400	C	0 1 2 NA
12	SLC45A2	rs28777	C	0 1 2 NA
13	SLC45A2	rs16891982	C	0 1 2 NA
14	KITLG	rs12821256	G	0 1 2 NA
15	LOC105374875	rs4959270	A	0 1 2 NA
16	IRF4	rs12203592	T	0 1 2 NA
17	TYR	rs1042602	T	0 1 2 NA
18	OCA2	rs1800407	A	0 1 2 NA
19	SLC24A4	rs2402130	G	0 1 2 NA
20	HERC2	rs12913832	T	0 1 2 NA
21	PIGU	rs2378249	C	0 1 2 NA
22	LOC105370627	rs12896399	T	0 1 2 NA
23	TYR	rs1393350	T	0 1 2 NA
24	TYRP1	rs683	G	0 1 2 NA
25	ANKRD11	rs3114908	T	0 1 2 NA
26	OCA2	rs1800414	C	0 1 2 NA
27	BNC2	rs10756819	G	0 1 2 NA
28	HERC2	rs2238289	C	0 1 2 NA
29	SLC24A4	rs17128291	C	0 1 2 NA
30	HERC2	rs6497292	C	0 1 2 NA
31	HERC2	rs1129038	G	0 1 2 NA
32	HERC2	rs1667394	C	0 1 2 NA
33	TYR	rs1126809	A	0 1 2 NA
34	OCA2	rs1470608	A	0 1 2 NA
35	SLC24A5	rs1426654	G	0 1 2 NA
36	ASIP	rs6119471	C	0 1 2 NA
37	OCA2	rs1545397	T	0 1 2 NA
38	RALY	rs6059655	T	0 1 2 NA
39	OCA2	rs12441727	A	0 1 2 NA
40	MC1R	rs3212355	A	0 1 2 NA
41	DEF8	rs8051733	C	0 1 2 NA

Display Predicted Phenotype

Download Predicted Phenotype

As you can see, there are instructions on how to fill in the number of alleles, if the input allele mentioned under the 'allele' column is not present for that SNP, Click the '0' button on the page. If the input allele is present as a heterozygote genotype, you must click the '1' button. If the input allele is present in its homozygote state, you must click the '2' button.

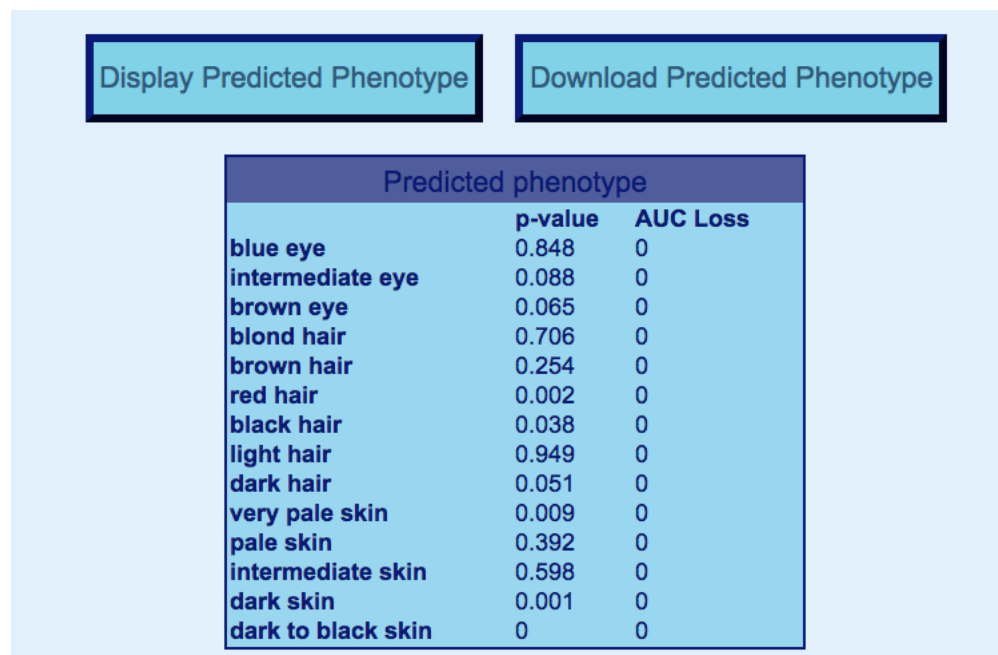
If you do not know the genotype of the SNP, then you must click 'NA'.

\*if you are using sequencing data, than you may need to flip strand orientation in your result before inputting into this prediction model. The only SNP that may cause confusion and therefore must be converted (from NCBI's forward orientation) is rs6119471. All other G/C, A/T SNPs are in the correct orientation for input or are opposite alleles i.e. SNP G/A, input requests C/T.

Please ensure all 41 SNPs have the correct input and then proceed with clicking the button '*Display Predicted Phenotype*'.

As you can see from the below example, all 41 SNP inputs were 0 and a prediction box appeared with the Predicted Phenotype result.

You can also download the prediction result as a file by clicking the '*Download Predicted Phenotype*' button to store on your computer. It includes details on missing input SNPs, Full AUC measures and loss in AUC measures.



The screenshot shows a user interface with two buttons at the top: "Display Predicted Phenotype" and "Download Predicted Phenotype". Below the buttons is a table titled "Predicted phenotype" with three columns: "Predicted phenotype", "p-value", and "AUC Loss". The table lists 13 categories with their respective p-values and AUC loss values.

Predicted phenotype	p-value	AUC Loss
blue eye	0.848	0
intermediate eye	0.088	0
brown eye	0.065	0
blond hair	0.706	0
brown hair	0.254	0
red hair	0.002	0
black hair	0.038	0
light hair	0.949	0
dark hair	0.051	0
very pale skin	0.009	0
pale skin	0.392	0
intermediate skin	0.598	0
dark skin	0.001	0
dark to black skin	0	0

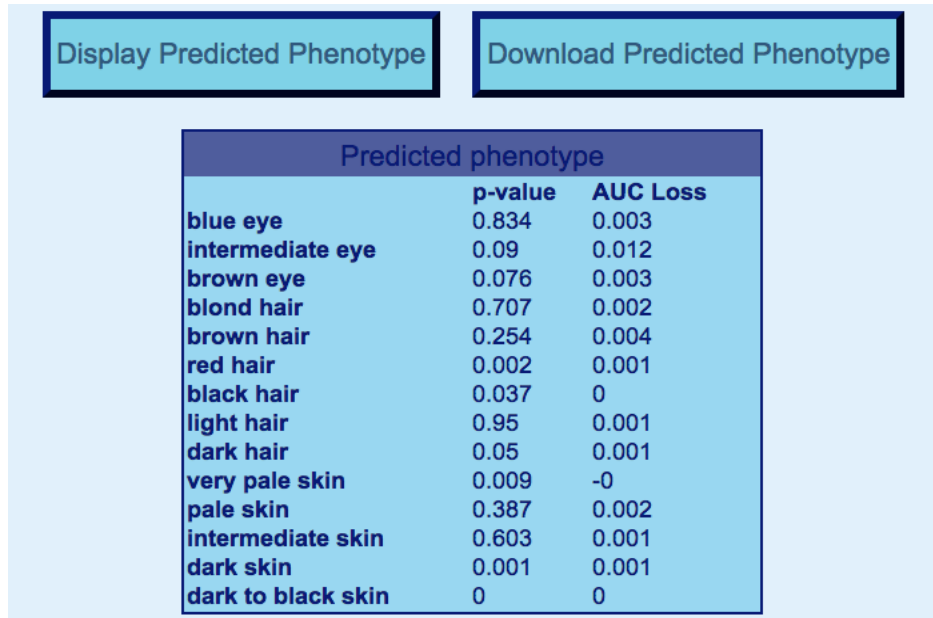
The **eye colour prediction** result consists of three categories that sum up to 1. These are a prediction probability (p-value) for Blue, Intermediate and Brown.

The **hair colour prediction** result consists of four categories that sum up to 1. These are a prediction probability (p-value) for Blond, Brown, Red and Black. There is also a shade prediction probability (p-value) for Light and Dark that sum up to 1.

The **skin colour prediction** result consists of five categories that sum up to 1. These are a prediction probability (p-value) for Very Pale, Pale, Intermediate, Dark, and Dark-to-Black.

The **AUC loss** for each of the predictions represent the loss in AUC accuracy when using all the required SNPs (6 for eye colour, 22 for hair colour and 36 for skin colour)

versus the loss of one or several SNPs. See example below where the SNP rs16891982 is missing. P-value results are still obtained but must be reported with the Loss in AUC due to missing input SNPs.



Predicted phenotype		
	p-value	AUC Loss
blue eye	0.834	0.003
intermediate eye	0.09	0.012
brown eye	0.076	0.003
blond hair	0.707	0.002
brown hair	0.254	0.004
red hair	0.002	0.001
black hair	0.037	0
light hair	0.95	0.001
dark hair	0.05	0.001
very pale skin	0.009	-0
pale skin	0.387	0.002
intermediate skin	0.603	0.001
dark skin	0.001	0.001
dark to black skin	0	0

Inputs missing the *HERC2* SNP (rs12913832 = NA) will not provide an eye colour prediction result, as the loss in AUC is too great.

Inputs missing all 11 *MC1R* SNPs will not provide a hair colour prediction result.

Inputs missing a combination of *HERC2-SLC45A2-IRF4* will not produce an eye or hair prediction result.

The reason for these clauses is that the loss in AUC is too great, and predictions could be severely incorrect.

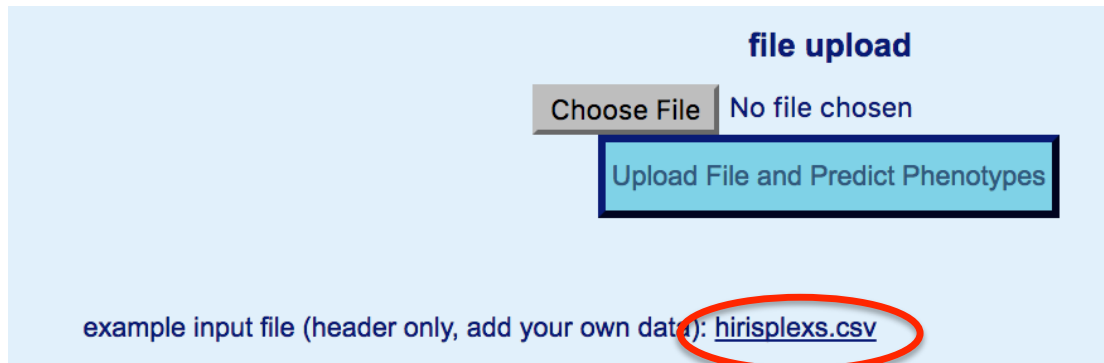
Although there is no strict cut off for skin colour prediction, missing any markers is not advised, in particular *SLC24A5* rs1426654, and *OCA2* rs1800414/*MC1R* rs3212355 as it will impact on discerning the difference between an Intermediate prediction versus a Dark/Dark-to-Black prediction, particularly in an Asian ancestral samples due to proposed convergent evolution.

## Eye, Hair and Skin Colour Prediction using HirisPlex-S

### Multiple Individuals

#### Upload File

Please download the HirisPlex-S input template from the file upload portion of the website, see image below. It is a .csv file and should stay in this format but the file can be renamed.



The *hirisplexs.csv* input file looks like this

	A	B	C	D	E	F	G	H	I	J	K	L
1	sampleid	rs312262906	rs11547464	rs885479_T	rs1805008_T	rs1805005_T	rs1805006_A	rs1805007_T	rs1805009_C	rs201326893	rs2228479_A	rs1110400_C
2												

Under SampleID, input the name of your sample. Under the SNP names, please enter the number of alleles of the input allele indicated in the column title.

i.e. rs1805007\_T  
rs1805007 is the SNP name  
T is the input allele.

If there are no T alleles in the rs1805007 genotype, input 0

If there is 1 T allele in the rs1805007 genotype, input 1

If there are 2 T alleles in the rs1805007 genotype, input 2 (example shown)

If the rs1805007 is missing, input NA

#### TEST Example input

	A	B	C	D	E	F	G	H	I	J
1	sampleid	rs312262906	rs11547464	rs885479_T	rs1805008_T	rs1805005_T	rs1805006_A	rs1805007_T	rs1805009_C	rs201326893_A
2	TEST	0	0	0	0	2	0	2	1	0
3										

For multiple individuals, insert into each row following the above instructions. When complete, save the file, do not change the .csv format, but you may rename.

Proceed with uploading the file by clicking the 'choose file' button and link to the created file on your computer.

Press the button 'Upload File and Predict Phenotypes' and the result file, containing all eye colour predictions for each individual, will automatically download to your computer under the title 'Result.csv'

The Results.csv prediction result file includes details on missing input SNPs, Full AUC measures and loss in AUC measures. See below for interpretation of file.

Sampleid:	Id of sample
PBlueEye:	Probability prediction of Blue eyes
PIntermediateEye:	Probability prediction of Intermed. eyes
PBrownEye:	Probability prediction of Brown eyes
Full_AUC_BlueEye:	Blue best AUC at a population level
Full_AUC_IntermediateEye:	Intermed. best AUC at a population level
Full_AUC_BrownEye:	Brown best AUC at a population level
Numb_missingSNPs_Eye:	Number of missing inputs for eye
Name_missingSNPs_Eye:	SNP Name of missing inputs for eye
AUC_Loss_BlueEye:	Loss in Blue AUC due to missing input
AUC_Loss_IntermediateEye:	Loss in Intermed. AUC due to missing input
AUC_Loss_BrownEye:	Loss in Brown AUC due to missing input
PBlondHair:	Probability prediction of Blond hair
PBrownHair:	Probability prediction of Brown hair
PRedHair:	Probability prediction of Red hair
PBlackHair:	Probability prediction of Black hair
Full_AUC_BlondHair:	Blond best AUC at a population level
Full_AUC_BrownHair:	Brown best AUC at a population level
Full_AUC_RedHair:	Red best AUC at a population level
Full_AUC_BlackHair:	Black best AUC at a population level
Numb_missingSNPs_Hair	Number of missing inputs for hair
Name_missingSNPs_Hair:	SNP Name of missing inputs for hair
AUC_Loss_BlondHair:	Loss in Blond AUC due to missing input
AUC_Loss_BrownHair:	Loss in Brown AUC due to missing input
AUC_Loss_RedHair:	Loss in Red AUC due to missing input
AUC_Loss_BlackHair:	Loss in Black AUC due to missing input
PLightHair:	Probability prediction of light hair colour
PDarkHair:	Probability prediction of dark hair colour
Full_AUC_HairShade:	Hair shade best AUC at a population level
Numb_missingSNPs_HairShade:	Number of missing inputs for hair shade
Name_missingSNPs_HairShade:	SNP Name of missing inputs for hair shade
AUC_Loss_HairShade:	Loss in Hair shade AUC due to missing input
PVeryPaleSkin:	Probability prediction of very pale skin colour
PPaleSkin:	Probability prediction of pale skin colour
PIntermediateSkin:	Probability prediction of intermediate skin colour
PDarkSkin:	Probability prediction of dark skin colour
PDarktoBlackSkin:	Probability prediction of dark-to-black skin colour
Full_AUC_VeryPaleSkin:	Very Pale skin best AUC at a population level
Full_AUC_PaleSkin:	Pale skin best AUC at a population level

Full_AUC_IntermediateSkin:	Intermediate skin best AUC at a population level
Full_AUC_DarkSkin:	Dark skin best AUC at a population level
Full_AUC_DarktoBlackSkin:	Dark-to-black skin best AUC at a population level
Numb_missingSNPs_Skin	Probability prediction of intermediate skin colour
Name_missingSNPs_Skin	Probability prediction of dark skin colour
AUC_Loss_VeryPaleSkin:	Loss in Very Pale AUC due to missing input
AUC_Loss_PaleSkin:	Loss in Pale AUC due to missing input
AUC_Loss_IntermedateSkin:	Loss in Intermediate AUC due to missing input
AUC_Loss_DarkSkin:	Loss in Dark AUC due to missing input
AUC_Loss_DarktoBlackSkin:	Loss in Dark-to-Black AUC due to missing input

## Eye and hair colour prediction using HirisPlex

### Single Individual

#### Online Interface for on-screen result

To use the HirisPlex portion of this website to simultaneously predict the eye and hair colour of an individual, please follow these instructions.

You can see the HirisPlex portion of the website below. All 24 gene names, single nucleotide polymorphisms (SNPs) and model input alleles belonging to the HirisPlex system are displayed.



	Gene	SNP	Allele	No. of Alleles
1	MC1R	rs312262906	A	0 1 2 NA
2	MC1R	rs11547464	A	0 1 2 NA
3	MC1R	rs885479	T	0 1 2 NA
4	MC1R	rs1805008	T	0 1 2 NA
5	MC1R	rs1805005	T	0 1 2 NA
6	MC1R	rs1805006	A	0 1 2 NA
7	MC1R	rs1805007	T	0 1 2 NA
8	TUBB3	rs1805009	C	0 1 2 NA
9	MC1R	rs201326893	A	0 1 2 NA
10	MC1R	rs2228479	A	0 1 2 NA
11	MC1R	rs1110400	C	0 1 2 NA
12	SLC45A2	rs28777	C	0 1 2 NA
13	SLC45A2	rs16891982	C	0 1 2 NA
14	KITLG	rs12821256	G	0 1 2 NA
15	LOC105374875	rs4959270	A	0 1 2 NA
16	IRF4	rs12203592	T	0 1 2 NA
17	TYR	rs1042602	T	0 1 2 NA
18	OCA2	rs1800407	A	0 1 2 NA
19	SLC24A4	rs2402130	G	0 1 2 NA
20	HERC2	rs12913832	T	0 1 2 NA
21	PIGU	rs2378249	C	0 1 2 NA
22	LOC105370627	rs12896399	T	0 1 2 NA
23	TYR	rs1393350	T	0 1 2 NA
24	TYRP1	rs683	G	0 1 2 NA

As you can see, there are instructions on how to fill in the number of alleles, if the input allele mentioned under the 'allele' column is not present for that SNP, Click the '0' button on the page. If the input allele is present as a heterozygote genotype, you must click the '1' button. If the input allele is present in its homozygote state, you must click the '2' button.

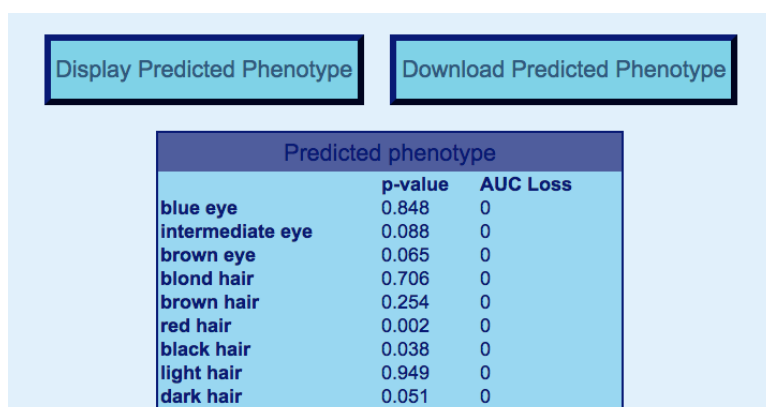
If you do not know the genotype of the SNP, then you must click 'NA'.

Please ensure all 24 SNPs have the correct input and then proceed with clicking the button 'Display Predicted Phenotype'.

As you can see from the below example, all 24 SNP inputs were 0 and a prediction box appeared with the Predicted Phenotype result.



You can also download the prediction result as a file by clicking the 'Download Predicted Phenotype' button to store on your computer. It includes details on missing input SNPs, Full AUC measures and loss in AUC measures.

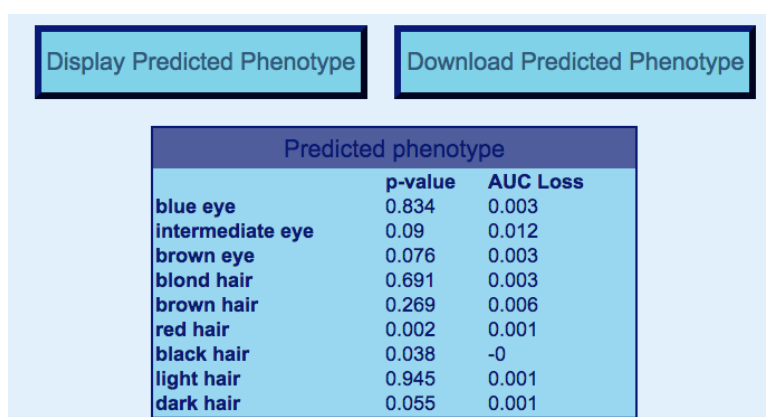


Predicted phenotype		
	p-value	AUC Loss
blue eye	0.848	0
intermediate eye	0.088	0
brown eye	0.065	0
blond hair	0.706	0
brown hair	0.254	0
red hair	0.002	0
black hair	0.038	0
light hair	0.949	0
dark hair	0.051	0

The eye colour prediction result consists of three categories that sum up to 1. These are a prediction probability (p-value) for Blue, Intermediate and Brown.

The hair colour prediction result consists of four categories that sum up to 1. These are a prediction probability (p-value) for Blond, Brown, Red and Black. And includes a separate Shade prediction of Light and Dark probabilities (p-values) that also sum to 1.

The AUC loss represent the loss in AUC accuracy when using all 24 SNPs versus the loss of one or several SNPs. See example below where the SNP rs16891982 and rs2402130 are missing. P-value results are still obtained but must be reported with the Loss in AUC due to missing input SNPs.



Predicted phenotype		
	p-value	AUC Loss
blue eye	0.834	0.003
intermediate eye	0.09	0.012
brown eye	0.076	0.003
blond hair	0.691	0.003
brown hair	0.269	0.006
red hair	0.002	0.001
black hair	0.038	-0
light hair	0.945	0.001
dark hair	0.055	0.001

Inputs missing the *HERC2* SNP (rs12913832 = NA) will not provide an eye colour prediction result.

Inputs missing all 11 *MC1R* SNPs will not provide a hair colour prediction result.

Inputs missing a combination of *HERC2-SLC45A2-IRF4* will not produce any prediction result.

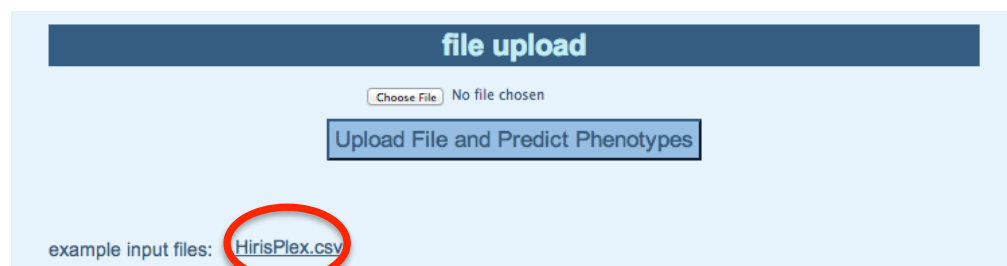
The reason for these clauses is that the loss in AUC is too great, and predictions could be severely incorrect.

## Eye & Hair Colour Prediction using HirisPlex

### Multiple Individuals

#### Upload File

Please download the IrisPlex input template from the file upload portion of the website, see image below. It is a .csv file and should stay in this format but the file can be renamed.



The *hirisplex.csv* input file looks similar to the *irisplex.csv* file, however contains columns for all 24 eye and hair colour input SNPs.

Under SampleID, input the name of your sample. Under the SNP names, please enter the number of alleles of the input allele indicated in the column title.

i.e. rs12203592\_T  
rs12203592 is the SNP name  
T is the input allele.

If there are no T alleles in the rs12203592 genotype, input 0  
If there is 1 T allele in the rs12203592 genotype, input 1  
If there are 2 T alleles in the rs12203592 genotype, input 2  
If the rs12203592 is missing, input NA

Example input if no T alleles present

rs12203592_T
0

For multiple individuals, insert into each row following the above instructions. When complete, save the file, do not change the .csv format, but you may rename.

Proceed with uploading the file by clicking the 'choose file' button and link to the created file on your computer.

Press the button 'Upload File and Predict Phenotypes' and the result file, containing all eye colour predictions for each individual, will automatically download to your computer under the title 'Result.csv'

The Results.csv prediction result file includes details on missing input SNPs, Full AUC measures and loss in AUC measures. See below for interpretation of file.

Sampleid:	Id of sample
PBlueEye:	Probability prediction of Blue eyes
PIntermediateEye:	Probability prediction of Intermed. eyes
PBrownEye:	Probability prediction of Brown eyes
Full_AUC_BlueEye:	Blue best AUC at a population level
Full_AUC_IntermediateEye:	Intermed. best AUC at a population level
Full_AUC_BrownEye:	Brown best AUC at a population level
Numb_missingSNPs_Eye:	Number of missing inputs for eye
Name_missingSNPs_Eye:	SNP Name of missing inputs for eye
AUC_Loss_BlueEye:	Loss in Blue AUC due to missing input
AUC_Loss_IntermediateEye:	Loss in Intermed. AUC due to missing input
AUC_Loss_BrownEye:	Loss in Brown AUC due to missing input
PBlondHair:	Probability prediction of Blond hair
PBrownHair:	Probability prediction of Brown hair
PRedHair:	Probability prediction of Red hair
PBlackHair:	Probability prediction of Black hair
Full_AUC_BlondHair:	Blond best AUC at a population level
Full_AUC_BrownHair:	Brown best AUC at a population level
Full_AUC_RedHair:	Red best AUC at a population level
Full_AUC_BlackHair:	Black best AUC at a population level
Numb_missingSNPs_Hair	Number of missing inputs for hair
Name_missingSNPs_Hair:	SNP Name of missing inputs for hair
AUC_Loss_BlondHair:	Loss in Blond AUC due to missing input
AUC_Loss_BrownHair:	Loss in Brown AUC due to missing input
AUC_Loss_RedHair:	Loss in Red AUC due to missing input
AUC_Loss_BlackHair:	Loss in Black AUC due to missing input
PLightHair:	Probability prediction of light hair colour
PDarkHair:	Probability prediction of dark hair colour
Full_AUC_HairShade:	Hair shade best AUC at a population level
Numb_missingSNPs_HairShade:	Number of missing inputs for hair shade
Name_missingSNPs_HairShade:	SNP Name of missing inputs for hair shade
AUC_Loss_HairShade:	Loss in Hair shade AUC due to missing input

## Eye Colour Prediction using IrisPlex


### Single Individual

#### Online Interface for on-screen result

To use the IrisPlex portion of this website to predict the eye colour of an individual, please follow these instructions.

You can see the IrisPlex portion of the website below. All 6 gene names, single nucleotide polymorphisms (SNPs) and model input alleles belonging to the IrisPlex system are displayed.

**The IrisPlex System**



Gene	SNP	Allele	No. of Alleles
1 <b>HERC2</b>	rs12913832	T	0 1 2 NA
2 <b>OCA2</b>	rs1800407	A	0 1 2 NA
3 <b>LOC105370627</b>	rs12896399	T	0 1 2 NA
4 <b>SLC45A2</b>	rs16891982	C	0 1 2 NA
5 <b>TYR</b>	rs1393350	T	0 1 2 NA
6 <b>IRF4</b>	rs12203592	T	0 1 2 NA

Display Predicted Phenotype

Download Predicted Phenotype

**Instructions**

If input allele is not present, put 0 into 'No. of alleles' box  
If one is present, heterozygote, put 1 in the box  
If the only allele present is the input allele, homozygote, put 2 in the box  
If there is a missing SNP, please enter NA  
Inputs missing **HERC2** rs12913832 will not produce an eye colour prediction result

As you can see, there are instructions on how to fill in the number of alleles, if the input allele mentioned under the 'allele' column is not present for that SNP, Click the '0' button on the page. If the input allele is present as a heterozygote genotype, you must click the '1' button. If the input allele is present in its homozygote state, you must click the '2' button.

If you do not know the genotype of the SNP, then you must click 'NA'.

Please ensure all 6 SNPs have the correct input and then proceed with clicking the button '*Display Predicted Phenotype*'.


As you can see from the below example, all 6 SNP inputs were 0 and a prediction box appeared with the Predicted Phenotype result.

You can also download the prediction result as a file by clicking the '*Download Predicted Phenotype*' button to store on your computer. It includes details on missing input SNPs, Full AUC measures and loss in AUC measures.

The eye colour prediction result consists of three categories that sum up to 1. These are a prediction probability (p-value) for Blue, Intermediate and Brown.

The AUC loss represent the loss in AUC accuracy when using all 6 SNPs versus the loss of one or several SNPs. See example below where the SNP rs16891982 is missing. P-value results are still obtained but must be reported with the Loss in AUC due to missing input SNPs.

The IrisPlex System



Gene	SNP	Allele	No. of Alleles
1 <b>HERC2</b>	rs12913832	T	0 1 2 NA
2 <b>OCA2</b>	rs1800407	A	0 1 2 NA
3 <b>LOC105370627</b>	rs12896399	T	0 1 2 NA
4 <b>SLC45A2</b>	rs16891982	C	0 1 2 NA
5 <b>TYR</b>	rs1393350	T	0 1 2 NA
6 <b>IRF4</b>	rs12203592	T	0 1 2 NA

Display Predicted Phenotype

Download Predicted Phenotype

Predicted phenotype		
	p-value	AUC Loss
blue eye	0.848	0
intermediate eye	0.088	0
brown eye	0.065	0

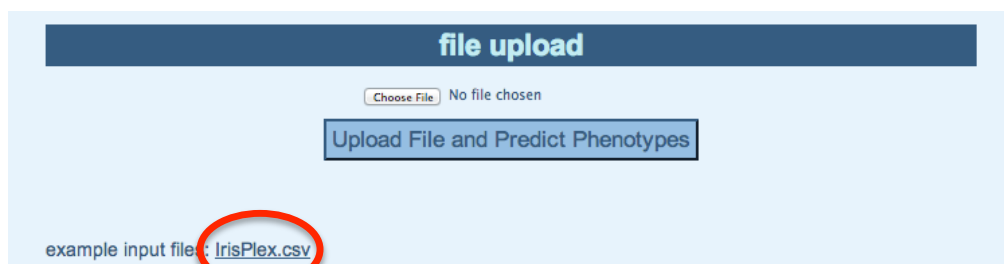
Inputs missing the *HERC2* SNP (rs12913832 = NA) will not provide a prediction result, as the loss in AUC is too great.

## Eye Colour Prediction using IrisPlex

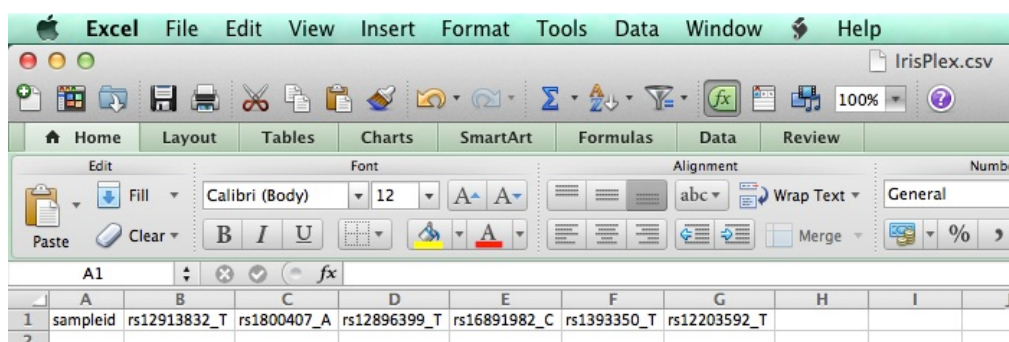
### Multiple Individuals

#### Upload File

Please download the IrisPlex input template from the file upload portion of the website, see image below. It is a .csv file and should stay in this format but the file can be renamed.



The *irisplex.csv* input file looks like this



	A	B	C	D	E	F	G	H	I	J
1	sampleid	rs12913832_T	rs1800407_A	rs12896399_T	rs16891982_C	rs1393350_T	rs12203592_T			
2										

Under SampleID, input the name of your sample. Under the SNP names, please enter the number of alleles of the input allele indicated in the column title.

i.e. rs12203592\_T  
rs12203592 is the SNP name  
T is the input allele.

If there are no T alleles in the rs12203592 genotype, input 0

If there is 1 T allele in the rs12203592 genotype, input 1

If there are 2 T alleles in the rs12203592 genotype, input 2

If the rs12203592 is missing, input NA

#### TEST Example input

	A	B	C	D	E	F	G	H
1	sampleid	rs12913832_T	rs1800407_A	rs12896399_T	rs16891982_C	rs1393350_T	rs12203592_T	
2	TEST	1	0	0	0	NA	0	

For multiple individuals, insert into each row following the above instructions.

When complete, save the file, do not change the .csv format, but you may rename.

Proceed with uploading the file by clicking the '*choose file*' button and link to the created file on your computer.

Press the button '*Upload File and Predict Phenotypes*' and the result file, containing all eye colour predictions for each individual, will automatically download to your computer under the title '*Result.csv*'

The Results.csv prediction result file includes details on missing input SNPs, Full AUC measures and loss in AUC measures. See below for interpretation of file.

Sampleid:	Id of sample
PBlueEye:	Probability prediction of Blue eyes
PIntermediateEye:	Probability prediction of Intermed. eyes
PBrownEye:	Probability prediction of Brown eyes
Full_AUC_BlueEye:	Blue AUC at a population level
Full_AUC_IntermediateEye:	Intermed. AUC at a population level
Full_AUC_BrownEye:	Brown AUC at a population level
Numb_missingSNPs_Eye:	Number of missing inputs
Name_missingSNPs_Eye:	SNP Name of missing inputs
AUC_Loss_BlueEye:	Loss in Blue AUC due to missing input
AUC_Loss_IntermediateEye:	Loss in Intermed. AUC due to missing input
AUC_Loss_BrownEye:	Loss in Brown AUC due to missing input

## Interpreting the Prediction Outcomes

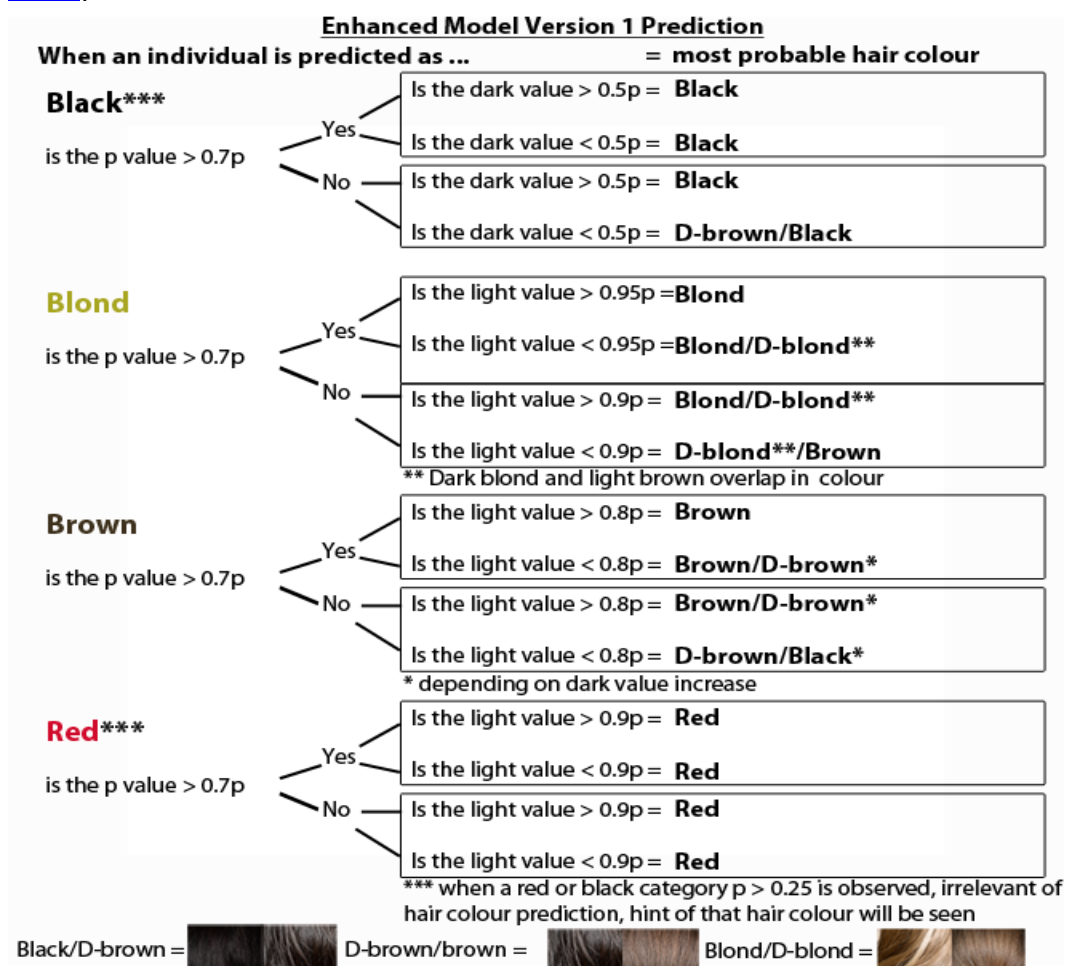
In terms of **eye colour**, the highest p-value is considered the predicted phenotype.



For more information, see [Walsh et al. \(2012\)](#) on threshold accuracy levels.

In terms of **hair colour**, there is a guide in place that utilises the highest p-value approach in combination with a step-wise model.




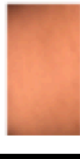
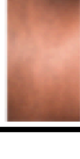
For more information, see the HirisPlex Developmental Validation paper ([Walsh et al. 2014](#))



For example, using the prediction from page 6 of this manual as an example. With a Blond p value of 0.742, it is the highest p value, hence the blond category p value > 0.7p Yes option above is chosen. The next step asks is the light value > 0.95p, the answer is No as it is 0.93p, hence the final prediction is blond/dark blond.



In terms of **skin colour**, there is a guide in place that utilises the highest p-value approach in combination with the second highest probability value.

	<b>Very Pale</b>	
	If highest probability > 0.9p	Very Pale predicted
	If highest probability > 0.7p	Very Pale is predicted however it will be affected by the second highest category if it is > 0.15 p and will appear darker Very Pale is predicted, unlikely to be affected by the second highest category if it is < 0.15 p
	If highest probability > 0.5p	Prediction significantly affected by second category, and will be a mix of the two highest categories (tends to represent the darker second highest category)
<b>Pale</b>		
	If highest probability > 0.9p	Pale predicted
	If highest probability > 0.7p	Pale is predicted however it will be affected by the second highest category if it is > 0.15 p (will appear darker if Intermediate, and lighter if pale) Pale is predicted, unlikely to be affected by the second highest category if it is < 0.15 p
	If highest probability > 0.5p	Prediction significantly affected by second category, and will be a mix of the two highest categories (tends to represent the darker second highest category) Not significantly impacted if second highest prediction is Very Pale
<b>Int.</b>		
	If highest probability > 0.9p	Intermediate predicted (*unless Dark-Black is the second highest category, then prediction is Dark)
	If highest probability > 0.7p	Intermediate is predicted however it will be affected by the second highest category if it is > 0.15 p (will appear darker if Dark/Dark-Black and lighter if Very Pale/Pale) Intermediate is predicted, unlikely to be affected by the second highest category if it is < 0.15 p (*unless it is Dark-Black, then prediction is Dark)
	If highest probability > 0.5p	Prediction significantly affected by second category, and will be a mix of the two highest categories (darker if Dark/Dark-Black represents the second highest category) Not significantly impacted if second highest prediction is Very Pale/Pale
<b>Dark</b>		
	If highest probability > 0.9p	Dark predicted
	If highest probability > 0.7p	Dark is predicted, unlikely to be affected by the second highest category if it is > 0.15 p (*unless it is Dark-Black, then prediction can be Dark-Black) Dark is predicted, unlikely to be affected by the second highest category if it is < 0.15 p (*unless it is Dark-Black, then prediction can be Dark-Black)
	If highest probability > 0.5p	Prediction significantly affected by second category, and will be a mix of the two highest categories (dark to black if Dark-Black represents the second highest category) Not significantly impacted if second highest prediction is Pale/Intermediate
<b>Dark-Black</b>		
	If highest probability > 0.9p	Dark to Black predicted
	If highest probability > 0.7p	Dark to Black is predicted, unlikely to be affected by the second highest category if it is > 0.15 p Dark to Black is predicted, unlikely to be affected by the second highest category if it is < 0.15 p
	If highest probability > 0.5p	Prediction affected by second category, and will be a mix of the two highest categories (Will be lighter than Dark to Black if Dark represents the second highest category)

For more information, see the HirisPlex-S Developmental Validation paper ([Chaitanya et al. 2018](#)).

## Understanding the Partial Prediction Result and AUC loss

As expected with low quality / quantity DNA samples, partial or incomplete IrisPlex/HirisPlex profiles may be observed in some forensic casework applications and any other applications where DNA quality and/or quantity is highly limited such as anthropological studies including ancient DNA studies.

With the use of this online prediction tool for partial IrisPlex/HirisPlex/HirisPlex-S profiles, the underlying prediction models (one for eye colour, another for hair colour including shade, and lastly one for skin colour) are re-estimated every time a partial profile is loaded in a way that the models are adjusted to represent only the DNA markers available from the input partial profile.

This approach provides a more realistic probability estimate from partial profiles because only the DNA variants present in the profile are used for model building

compared to the use of the complete model that is based on full profiles (all 24 SNPs) for probability estimations.

By re-running and creating a new model each time, based on the partial profile input SNPs, new AUCs are generated at the population level considering this lower SNP number. We define the result 'loss in AUC' as a subtraction of this partial AUC result from the full 24 SNP database model AUC values that can be found in this manual.

It shall be noted that different DNA variants used in the eye, hair and skin colour prediction models have different impacts on the eye, hair and skin colour prediction accuracies as determined by the differences in the strength of their predictive effects. Therefore, the accuracy obtainable with a partial profile is strongly determined by which DNA markers are present in a partial profile (or which are missing from a complete profile).

To illustrate: AUC for red hair drops from 0.92 AUC with all 11 *MC1R* markers in the HirisPlex profile to 0.64 AUC when all 11 HirisPlex DNA variants are missing from a partial profile, when using our revised hair colour prediction model introduced here. Keep in mind that AUC values of 0.5 represent random chance or the toss of a coin, and therefore no significant predictive enhancement. The closer the AUC value is to 1, the better its ability for accurate prediction at a population level. To read more, please see [Liu et al. \(2009\)](#)

As a partial profile has many possibilities and the missing genotypes would reflect different accuracy losses in particular colour categories, we allow the model to still produce probabilities based on the genotypes input, but at the same time strongly advise that the AUC losses due to missing DNA markers are reported together with the prediction outcomes. This allows a comparison for the analyst of the limitations of a particular partial profile relative to the full profile. The analyst can then make the report stating the result as acceptable or inconclusive based on these results.

## Online Database Information and Performance

For eye colour prediction, the model-underlying database now includes samples united from both publications by [Liu et al. \(2009\)](#) and [Walsh et al. \(2012\)](#), of 9,188 individuals from eight parts of Europe (Netherlands, Norway, Estonia, UK, France, Italy, Spain, and Greece), as well as 278 additional individuals from a US-based collection used in another study for skin colour prediction ([Walsh et al. 2017](#)).

Total number of individuals for categorical eye colour model = 9466

For hair colour prediction, the model-underlying database includes 1,601 individuals, from the publication [Walsh et al. \(2013\)](#) from Ireland, Greece and Poland and 50 individuals from a new Japanese set for which a black hair phenotype is recorded, as well as 277 additional individuals from a US-based collection used in another study for skin colour prediction ([Walsh et al. 2017](#)).

Total number of individuals for categorical hair colour model = 1878

For skin colour prediction, the model-underlying database includes 1423 individuals, from the publication [Walsh et al. \(2017\)](#) from Ireland, Greece and Poland, US-based (including parental birth places in Nigeria, Mexico, Columbia, India, Bangladesh, Palestine, Canada, China, Honduras, Germany, Philippines, Russia, Sudan, Japan, Saudi Arabia, Pakistan, El Salvador, Spain, Haiti, South Korea, Vietnam) as well as CEPH-HGDP individuals from Senegal, Nigeria, Kenya, and Papua New Guinea. AUCs reflect the full dataset being used for model assessment, which can also be found in Figure 2 of the publication.

Total number of individuals for categorical skin colour model = 1423

With such model-underlying reference database enlargements, below we provide overall population prediction accuracies as measured by the Area Under the receiver operating characteristic Curve (AUC) values using 100% of individuals/category in addition to confusion matrix measures assuming prediction >0.5p as 1 and <0.5p as 0 to measure Sensitivity (Sens), Specificity (Spec), Negative Predictive Value (NPV), Positive Predictive Value (PPV). Lastly, balanced accuracy (balACC) using the formula  $[1/2 (TP/P + TN/N)]$ , which can be defined as the average accuracy obtained on either class, is also given.

To note these performance measures take in highest probability only and not the prediction guide as used for hair and skin colour prediction.

<b>Eye Colour</b>	<b>AUC</b>	<b>Sens</b>	<b>Spec</b>	<b>PPV</b>	<b>NPV</b>	<b>bal ACC</b>
Blue	<b>0.94</b>	0.92	0.88	0.91	0.89	0.90
Intermediate	<b>0.74</b>	0	0.99	0	0.91	0.50
Brown	<b>0.95</b>	0.91	0.88	0.79	0.95	0.89

**\*Intermediate based on >0.45p for 1**

<b>Hair Colour</b>	<b>AUC</b>	<b>Sens</b>	<b>Spec</b>	<b>NPV</b>	<b>PPV</b>	<b>bal ACC</b>
Blond	<b>0.81</b>	0.66	0.78	0.64	0.80	0.72
Brown	<b>0.74</b>	0.66	0.67	0.59	0.73	0.66
Red	<b>0.93</b>	0.62	0.99	0.83	0.97	0.81
Black	<b>0.86</b>	0.35	0.98	0.69	0.91	0.66
Colour Shade	<b>0.91</b>	0.59	0.96	0.81	0.89	0.78

<b>Skin Colour</b>	<b>AUC</b>	<b>Sens</b>	<b>Spec</b>	<b>NPV</b>	<b>PPV</b>	<b>bal ACC</b>
Very Pale	<b>0.83</b>	0.09	0.99	0.64	0.94	0.54
Pale	<b>0.76</b>	0.67	0.69	0.64	0.74	0.69
Intermediate	<b>0.78</b>	0.58	0.81	0.66	0.75	0.69
Dark	<b>0.98</b>	0.53	0.99	0.68	0.98	0.76
Dark-to-Black	<b>0.99</b>	0.92	0.99	0.94	0.99	0.96

The addition of new individuals to the underlying databases are expected to provide additional strength with more difficult samples such as rare genotype-phenotype combinations as with increased database size come increased genotype combinations.

## Independent Performance Set

Using an independent test set of 194 US-based individuals who were not included in any model building and validation, an overall look at performance was undertaken using the HirisPlex-S models for eye, hair and skin colour.

### *Eye Colour*

Eye colour could be predicted correctly in 80% of cases (155 out of 194 predictions were correct). Worthy to note is that this value increases to 96% (156 out of 162) when only brown and blue phenotypes are assessed and the intermediate eye colour category was not included.

A word of caution is that blue eye colour predictions in particular are affected more than brown eye colour predictions due to these 33 'intermediate phenotypes', typically blue/green or blue/yellow described phenotypes. There were 26 incorrect predictions for blue (that had an Intermediate phenotype), while only 7 incorrect predictions for brown (that had an Intermediate phenotype). No intermediate individual was correctly predicted for their category.

### *Hair Colour*

Hair colour was predicted with 77% accuracy on average (149 out of 194), utilising the hair colour prediction guide from Walsh et al. 2014 and as seen below.

The errors came from predictions giving black but phenotype was recorded as Dark Brown (n=5) that can be explained by phenotypes originating from questionnaire data (as hair was dyed). In addition to predictions giving shades of Brown hair colour, that were incorrect (n=18), and lastly prediction giving Blond but phenotype was Brown (n=14) that has been linked with age-dependent hair colour changes. There were 8 instances of Red shade prediction that were incorrect; in fact the individuals were observed as light brown/blond.

### *Skin Colour*

Skin colour using a 3-point scale of White, Intermediate and Black was predicted with 91% accuracy on average (177 out of 194) utilizing the highest probability value.

However, as being able to distinguish the White category into further categories such as Very Pale versus Intermediate is very beneficial, it is best advised to use the top

two highest categories according to the prediction guide published in [Chaitanya et al. 2018](#) and also seen below, to give the final prediction. According to a 5-point scale and using this guide, 80% prediction accuracy on average was observed (156 out of 194).

The errors came from predicting the skin colour in a darker category than observed in 6 of the cases, and incorrectly predicting within the 'White' category in 32 of the cases. i.e. calling a prediction Very Pale to Pale when in fact it was Pale to Intermediate.

## **Acknowledgements**

Thank you to all current and past members of the Department of Genetic Identification, Erasmus MC, Rotterdam, The Netherlands and the Walsh Laboratory at IUPUI in Indiana, USA where these tools were designed and tested especially Professor Manfred Kayser, Assistant Professor Fan Liu (now also Professor at Key Laboratory of Genomic and Precision Medicine, Beijing China) and Assistant Professor Susan Walsh. A special thanks to our close collaborator Professor Wojciech Branicki, his team and all co-authors of the publications mentioned including the individuals who graciously gave their phenotype & genotype information as input for the databases.

Finally, a very special thanks to Sjozef van Baal for all the hard work needed to create this stellar website online tool.

The initial research and development that led to this resource was supported in part by funding from the Netherlands Forensic Institute (NFI), Erasmus MC University Medical Center Rotterdam, and by a grant from the Netherlands Genomics Initiative (NGI)/Netherlands Organization for Scientific Research (NWO) within the framework of the Forensic Genomics Consortium Netherlands (FGCN). The work of SW has funding support from the National Institute of Justice (Grant 2014-DN-BX-K031) and Indiana University Purdue University Indianapolis (IUPUI). FL is supported by the Thousand Talents Program for Distinguished Young Scholars China.

Development and maintenance of this web resource is supported by Erasmus MC University Medical Center Rotterdam.

## **Disclaimer**

Partial profiles do not reflect the full prediction potential of both IrisPlex and HirisPlex systems and should be used with caution. Also prediction values may differ from IrisPlex or HirisPlex Excel prediction values previously published. This is due to an increase in database size for use in the underlying model for both eye and hair colour in this online tool.

## **Copyright**

Copyright © of the Department of Genetic Identification of Erasmus MC, Rotterdam, The Netherlands. All Rights Reserved. Part of the knowledge used is protected by the European patent EP2195448.

No part of this website or manual may be reproduced without consent.

## Citation

Please make sure to cite the following publications if you use these website tools:

L. Chaitanya, K. Breslin, S. Zuñiga, L. Wirken, E. Pospiech, M. Kukla-Bartoszek, T. Sijen, P. de Knijff, F. Liu, W. Branicki, M. Kayser, S. Walsh. The HRISplex-S system for eye, hair and skin colour prediction from DNA: Introduction and forensic developmental validation. *Forensic Sci. Int. Genet.* <https://doi.org/10.1016/j.fsigen.2018.04.004>

S. Walsh, L. Chaitanya, K. Breslin, C. Muralidharan, A. Bronikowska, E. Pospiech, J. Koller, L. Kovatsi, A. Wollstein, W. Branicki, F. Liu, M. Kayser. Global skin colour prediction from DNA. *Human Genetics* 136 (2017) 847–863.

S. Walsh, L. Chaitanya, L. Clarisse, L. Wirken, J. Draus-Barini, L. Kovatsi, H. Maeda, T. Ishikawa, T. Sijen, P. de Knijff, W. Branicki, F. Liu, M. Kayser, Developmental validation of the HRISplex system: DNA-based eye and hair colour prediction for forensic and anthropological usage. *Forensic Sci. Int.: Genet.* 9 (2014) 150-161.

## References

1. F. Liu, K. van Duijn, J. R. Vingerling, A. Hofman, A. G. Uitterlinden, A. C. J. W. Janssens, M. Kayser, Eye color and the prediction of complex phenotypes from genotypes. *Current Biology* 19 (2009) R192-R193.
2. S. Walsh, F. Liu, K. Ballantyne, M. van Oven, O. Lao, M. Kayser, 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 (2011) 170 - 180.
3. S. Walsh, A. Lindenbergh, S. Zuniga, T. Sijen, P. de Knijff, M. Kayser, K. Ballantyne, Developmental validation of the IrisPlex system: determination of blue and brown iris colour for forensic intelligence. *Forensic Sci. Int.: Genet.* 5 (2011) 464 - 471.
4. S. Walsh, A. Wollstein, F. Liu, et al., DNA-based eye colour prediction across Europe with the IrisPlex system. *Forensic Sci. Int.: Genet.* 6 (2012) 330 - 340.
5. W. Branicki, F. Liu, K. van Duijn, J Draus-Barini, E Pospiech, S. Walsh, T. Kupiec, A. Wojas-Pelc and M. Kayser, Model-based prediction of human hair color using DNA variants. *Hum Genet* 129 (2011) 443 – 454.
6. S. Walsh, F. Liu, A. Wollstein, L. Kovatsi, A. Ralf, A. Kosiniak-Kamysz, W. Branicki, M. Kayser, The HRISplex system for simultaneous prediction of hair and eye colour from DNA. *Forensic Sci. Int. Genet.* 7 (2013) 98 - 115.
7. J. Draus-Barini, S. Walsh, E. Pospiech, T. Kupiec, H. Glab, W. Branicki, M. Kayser, Bona fide colour: DNA prediction of human eye and hair colour from ancient and contemporary skeletal remains. *Investigative Genetics* 4 (2013) 3.

8. S. Walsh, L. Chaitanya, L. Clarisse, L. Wirken, J. Draus-Barini, L. Kovatsi, H. Maeda, T. Ishikawa, T. Sijen, P. de Knijff, W. Branicki, F. Liu, M. Kayser, Developmental validation of the HIRisPlex system: DNA-based eye and hair colour prediction for forensic and anthropological usage. *Forensic Sci. Int. Genet.* 9 (2014) 150-161.
9. L. Chaitanya, I.Z. Pajnic, S. Walsh, J. Balažic, T. Zupanc, M. Kayser. Bringing colour back after 70 years: Predicting eye and hair colour from skeletal remains of World War II victims using the HIRisPlex system. *Forensic Sci. Int. Genetics* 26 (2017) 48-57.
10. S. Walsh, L. Chaitanya, K. Breslin, C. Muralidharan, A. Bronikowska, E. Pospiech, J. Koller, L. Kovatsi, A. Wollstein, W. Branicki, F. Liu, M. Kayser. Global skin colour prediction from DNA. *Human Genetics* 136 (2017) 847–863.
11. L. Chaitanya, K. Breslin, S. Zuñiga, L. Wirken, E. Pospiech, M. Kukla-Bartoszek, T. Sijen, P. de Knijff, F. Liu, W. Branicki, M. Kayser, S. Walsh. The HIRisPlex-S system for eye, hair and skin colour prediction from DNA: Introduction and forensic developmental validation. *Forensic Sci. Int. Genet.* <https://doi.org/10.1016/j.fsigen.2018.04.004>