# AGENDA FOR THE EDNAP MEETING

# SANTIAGO DE COMPOSTELA – 20 OCTOBER 2015

Expected duration: 09.00 - 17.00

Coffee: 10.30 – Lunch: 13.00-14.00 – Coffee: 15.30

---

Host: Angel Carracedo
Chairman: Niels Morling

| | |
|---|---|
| Welcome | Angel Carracedo |
| Update on activities concerning | |
| mtDNA SNP screening – two PCRs, 18 SNPs | Arnoud Kal |
| Updates from other groups | |
| EMPOP | Walther Parson |
| High quality DNA sequence database | Walther Parson |
| Nomenclature of STR sequences | Chris Phillips |
| EUROFORGEN-NoE | Peter Schneider |
| Identification of the MH17 victims | Arnoud Kal |
| EDNAP website update (www.isfg.org/EDNAP) | Peter Schneider |
| Future activities | |
| - Methylated DNA and age exercise | Athina Vidaki |
| EDNAP meeting in spring 2016 in Warsaw - April 2015 | Niels Morling |
| EDNAP meeting in autumn 2016 – Rome? | Niels Morling |
| Any other business | Niels Morling |

# EUROPEAN DNA PROFILING GROUP (EDNAP) MEETING

## Santiago de Compostela, Spain

## 20 October 2015

Host: Angel Carracedo
Chairman: Niels Morling

A list of participants is attached.

**Welcome**
Professor Angel Carracedo welcomed members to Santiago de Compostela.

**Update on exercises**

*A SNaPshot based method targeting18 common mtDNA mutations* Arnoud Kal
Arnoud Kal presented the results of the collaborative EDNAP exercise concerning typing of 18 mtDNA SNPs with the SNaPshot method (presentation attached).

**Updates from other groups**

*EMPOP* Walther Parson
Walther Parson gave an update on the new version of the database EMPOP. New developments include an event-based search engine (SAM-E), map-based description of database matches, a new tabular layout of search results and the provision of the haplogroup status of mtDNA sequences. EMPOP follows the concept of providing high-quality data and basic results of database searches. The interpretation of search results is the responsibility of the users. There are courses planned for training on mtDNA matters in association with congresses. Individual courses can be organized (presentation attached).

*STR nomenclature* Walther Parson
Walther Parson gave an update on nomenclature of STRs analyzed by Massively Parallel Sequencing (MPS). The board of the IFSG has established a commission to provide considerations for early adopters of STR MPS. The issue was discussed at the 26$^{th}$ ISFG conference held in September 2015 in Krakow, Poland. The ISFG will formulate considerations on nomenclature of DNA sequence data for the forensic genetic community. The considerations will include minimal requirements for the analysis and storage of data as well as the use of reference sequences and nomenclature systems that are regularly updated.

*Nomenclature of STR sequences* Chris Phillips
Chris Phillips gave an update on the structures of the forensic genetic STRs and the various nomenclatures of STR sequences with special focus on the new challenges associated with the complete DNA sequence data that are produced by massively parallel sequencing (presentation attached).

*EUROFORGEN-NoE – General update* Peter Schneider
Peter Schneider gave an update concerning the project (presentation attached).

*Identification of the MH17 victims* Arnoud Kal

Arnoud Kal gave an overview of the work with the identification by means of DNA of the victims of the airplane accident in 2015 in Ukraine (presentation attached).

**EDNAP website update** (www.isfg.org/EDNAP) Peter Schneider

Members are encouraged to visit the website. Suggestions are welcome.

**Future activities**

*Methylated DNA and age exercise* Athina Vidaki

Athina Vidaki and Denise Syndercombe Court suggested a collaborative EDNAP exercise of forensic age estimation based on investigation of the degree of DNA methylation of select nucleotides (presentation attached). The proposed method includes a bisulfite sequencing protocol using the Illumina MiSeq platform. The methylation status of 16 CpG sites will be investigated. Other chemistries and instruments may be used by participants. However, Kings College will send out the first set of samples, reagents, and detailed protocols that are compatible with the Illumina MiSeq before the end of December 2015. Participants will be asked to return results before the end of March 2016. The precise protocol and production plan will be formulated as soon as possible. Some of the procedures may be new to some of the laboratories. Therefore, the timeline may be considered. Athina Vidaki and Denise Syndercombe Court will contact laboratories.

**Next meeting** Niels Morling

The next EDNAP meeting will be held 19 April 2016 in Prague in connection with the CODIS meeting the same day and the ENFSI DNA WG Meeting 20-22 April 2016.

The colleagues from Laboratoria Genetica Forense, Universita Cattolica, Rome, kindly invited EDNAP members to hold the meeting in October/November 2016 in Rome. The EDNAP members happily accepted the invitation.

**Any other business** Niels Morling

Niels Morling thanked Roman Hradil for the excellent collaboration with the ENFSI DNA Working Group during his work as chairman and welcomed the new chairman, Sanders Kneppers, and invited to continuation of the professional and helpful collaboration between the groups.

There was no other business.

**Closing of the meeting**

The meeting closed with sincere thanks to Angel Carracedo, Chris Phillips, Ana Mosquera and all other colleagues, who helped to organise the meeting in Santiago de Compostela.

**Attachments are found at the EDNAP website** http://www.isfg.org/EDNAP/Meetings:

- List of participants
- Presentations
  - Arnoud Kal: Report on mtDNA SNP typing (waiting)
  - Walther Parson: EMPOP report
  - Walther Parson: High quality DNA sequence database
  - Chris Phillips: Nomenclature of STR sequences
  - Arnoud Kal: Identification of the MH17 victims (waiting)
  - Peter Schneider: EUROFORGEN-NoE report

- Ahina Vidaki: Methylated DNA and age exercise.

EDNAP Meeting, Santiago de Compostela, Spain, Oct 20, 2015

# EMPOP mtDNA database

**Version 3 - launch: July 27 2015**

Dr. Walther Parson

assoc. Professor, Institute of Legal Medicine, Innsbruck

adj. Professor, Forensic Program, Penn State University, PA, USA

walther.parson@i-med.ac.at

# EMPOP 3 - Why change to a new version?

**Improve forensic use**

Improved search engine (SAM-E)

# Block insertions (selection)

16032+TCTCTGTTCTTTCAT

104+CGGAGC

105+GGAGCA

209+GTGTGTT

241+TAA

286+TAACA

291+ACATCATAACAAAAAA

398+ACCAGATTTCAAAT

470+TACTACTA

524+$[AC]_{2-8}$

563+AACAAAGAAC...AAA (204)

8271+CCCCCTCTA

8280+CCCCCTCTA

8289+CCCCCTCTA

# EMPOP 3 - Why change to a new version?

**Improve forensic use**

- Improved search engine (SAM-E)
- Better visualization of results (distribution maps)
- Show more details of matching and neighbor haplotypes
- New tools (e.g. Haplogroup Browser)

**Introduce new concepts**

- Concept of haplogroups in practical forensic work
- Distribution of haplogroups as relevant information
- Cost model for mutations
- Stimulate alternative interpretation of haplotypes

EMPOP V3 R11
N=34,617
Metapops revised
Release updates

EMPOP V2.3 R11
will be available
until Dec 2015 (no
Release updates)

Registration by Email
to assign services
(e.g. search history,
NETWORK, ...)

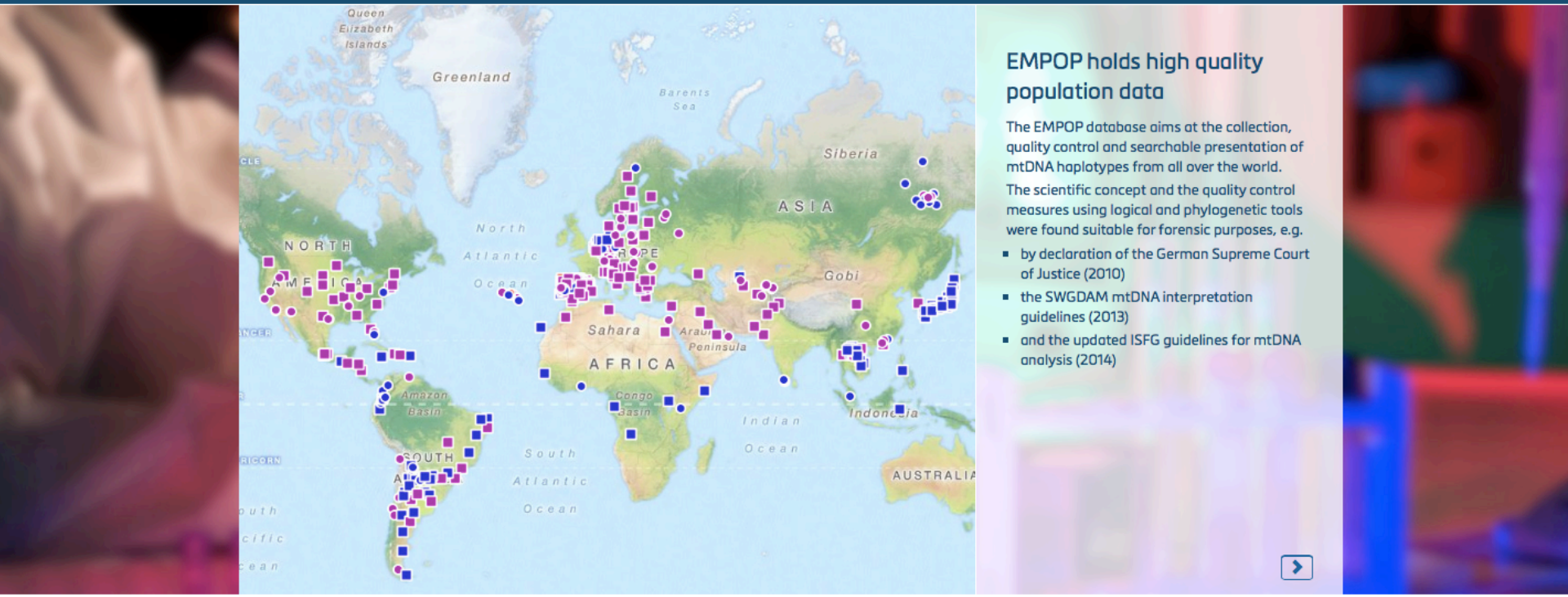
EMPOP mtDNA database, v3/R11 | GMI TIROL | ISFG

EMPOP v2.3 is still available here until Dec. 2015

walther.parson@gmail.com | your account | logout

Home Updates Use Methods Contribute Contributors Meet Terms of Use

QUERY POPULATIONS TOOLS




EMPOP slideshow provides information about novel features for mtDNA databanking and forensically relevant issues

# Updates

## 2015.07.27 | Welcome to EMPOP 3

EMPOP 3 represents a new mtDNA database version based on a novel programming concept and website layout. New features are included to accommodate Massively Parallel (Next Generation) Sequencing data.

### Good things that remain

- EMPOP only holds high quality mtDNA sequence data that underwent stringent quality control
- EMPOP uses SAM, an alignment-free search engine to guarantee that matches are found regardless of the alignment and nomenclature of query and database haplotypes
- MtDNA haplotypes are sorted by geographic and metapopulation categories that are relevant to forensics and informative for other scientific fields

### New developments in EMPOP 3

- The earlier distinction in forensic and literature data is no more required, as the QC tools applied to every uploaded dataset proved to be effective and reliable
- Tabular summaries of query results are presented in a more convenient format
- Statistical evaluation of matching haplotypes was improved by offering correction for sampling bias
- Neighbors are presented by Hamming distance and by costs (fluctuation rate)
- Haplogroup estimates of mtDNA sequences are provided based on maximum likelihood/minimum cost functions
- Geographical maps are provided to present the distribution of matching haplotypes and the distribution of haplogroups
- Haplogroup Browser is a new tool to search for and display the distribution of mitochondrial haplogroups

EMPOP mtDNA database, v3/R11

Home Updates Use Methods Contribute Contributors Meet Terms of Use

QUERY POPULATIONS TOOLS

# How to use EMPOP 3

The usage of EMPOP is described in the downloadable PDF file. For further questions and comments please contact info@empop.org.

pdf file
"how to use EMPOP"
- will be updated

# Methods

Description of methods used in EMPOP discussed at this workshop

## Alignment

MtDNA sequences are traditionally reported relative to the human reference sequence (rCRS). This format is short and convenient, however nucleotide sequence strings can be translated into more than one rCRS-coded haplotype and are therefore ambiguous. As a consequence, database searches may suffer from biased results when query and database haplotypes are aligned differently. In the forensic context that could lead to an underestimation of absolute and relative frequencies and thus to an overestimation of the statistical power of the evidence.

EMPOP uses SAM, a string-based search algorithm that converts query and database sequences into alignment-free nucleotide strings and thus eliminates the possibility that identical sequences will be missed in a database query.

Length variant hot spots can be ignored during a search (default setting). Length variant regions were reviewed and adapted in EMPOP 3 now including coding region length variants (see below). EMPOP 3 introduces an updated query engine that considers block insertions and block deletions (indels) as a single phylogenetic event. In the CA-repeat of the control region (positions 513 and 524) only tandem indels are observed, e.g. 523del 524del. While this tandem deletion nominally constitutes two individual differences to the rCRS it is considered as single event by the new version of SAM-E. This better reflects the phylogenetic nature of the mitochondrial molecule.

The following length variant regions are considered in EMPOP 3:

| Position | Type | Length of Ins/Del | Inserted/Deleted Block | Since SAM Version |
|---|---|---|---|---|
| 16193 | Length Variation | 16193del - 16193+CCC | Individual bases | 19 |
| 309 | Length Variation | 309-CCCC - 309+CCTCC | Individual bases | 19 |
| 455 | Length Variation | 455-T - 455+CC | Individual bases | 19 |
| 463 | Length Variation | 463-CC - 463+CCCCC | Individual bases | 20 |
| 573 | Length Variation | 573-C - 573+8C | Individual bases | 19 |
| 960 | Length Variation | 960-C - 960+CCCCC | Individual bases | 20 |
| 5899 | Length Variation | 5899-C - 5899+10C | Individual bases | 20 |
| 8276 | Length Variation | 8276-CCCC - 8276+CCCC | Individual bases | 20 |
| 8285 | Length Variation | 8285-CCCC - 8285+CCCC | Individual bases | 20 |

# Contribute

The board of the International Society of Forensic Genetics (ISFG) and the editors of Forensic Science International: Genetics and International Journal of Legal Medicine invited EMPOP to logistically organize and perform quality control (QC) of mtDNA sequences in the course of manuscript preparations for the journals. Before mtDNA population papers are put forward to the editors for review, the authors are requested to submit the data to EMPOP. After positive evaluation, the authors will be contacted by EMPOP with the respective EMPOP accession numbers that serve as indicator of successful QC for the editors and reviewers. The necessary steps for submission of mtDNA sequences to EMPOP are outlined below.

## Step 1

**Prepare your sequence data** as shown in the tutorial. An example can be downloaded from the EMPOP webpage here that can also be used as template. The emp-file is a tab delimited text file that can be created using standard text software or MS Excel (then, save file under .txt format and rename "txt" by "emp"). The first 2 lines specify individual details of the dataset and origin of the samples. Further lines list the mtDNA haplotypes headed by their sequence range(s). Note that a given sequence range is applied to all mtDNA haplotypes following this range until a new range is defined (range lines are defined by "#!"). Text lines, e.g. for individual comments, need to be identified using the "#" symbol.

## Step 2

Submit your file to EMPOP by Email to EMPOP. The data will be quality checked for format, clerical error, sequence range violation, reference error, designation of indels, phantom mutation using in-house software and Network.

# Contributors

EMPOP was established under the scientific umbrella of the European DNA Profiling Group (EDNAP). Its conceptual design has been open to scientific discussion, which many colleagues in the forensic and population genetic field joined. The current version of the database is a result of international teamwork and hopefully continues to evolve by fruitful collaboration in the future.

## Collaborative scientific groups

**EDNAP**
The European DNA Profiling Group
EDNAP Website
Publication: Parson 2004

**Ge.F.I.**
Gruppo Ematologi Forensi Italiani
Ge.F.I. Website
Publication: Turchi 2008

**GHEP-ISFG**
Grupo Español y Portugués de la ISFG / Grupo Espanhol e Português da ISFG
GHEP-ISFG Website
Publication: Prieto 2010

under construction
to be completed ...

please send comments

## Individual collaborators

**Argentina**
Carlos Vullo, Cecilia Bobillo, Daniel Corach, Laura Catelli

**Belgium**
Bernadette Hoste, Stijn Desmyter

**Brazil**
Cintia Fridman, Denilce R Sumita, Edna S. Miazato Iwamura, Greiciane G Paneto, Martin R Whittle, Regina M Cicarelli

**Chile**
Servicio Medico Legal

**China**
Xiang Yanbao

# EMPOP Tools

The EMPOP tools section provides a suite of software to support the analysis and interpretation of mitochondrial DNA sequence variation.

## Haplogroup Browser

The haplogroup browser represents the established most recent Phylotree haplogroups in convenient searchable format and provides the number of EMPOP sequences assigned to the respective haplogroups by EMMA. Note that EMPOP provides the MRCA haplogroup if multiple haplogroup assignments are feasible.

Individual haplogroups can also be found by querying differences to the rCRS in a database of > 20,000 mtGenome sequences.

## EMPcheck

EMPcheck is a tool to perform plausibility checks on a rCRS-coded data table.

The file format must meet the requirements described below and in Carracedo et al 2014

See here for the required file format.

## NETWORK

NETWORK is software to calculate and draw quasi-median networks. They are useful to examine the quality of an mtDNA dataset.

Details on how to use NETWORK and references can be found here.

**Example:**

query rCRS (empty profile) in CR range (16024-576)

QUERY POPULATIONS TOOLS

Query Result Details Neighbors

Sample ID rCRS Control Region

Ranges 16024-576

Profile

Subset of 26,127 CR haplotypes (R11)

| Entire Database | | Frequency | Clopper Pearson CI |
|---|---|---|---|
| | 11/26127 | 4.2102e-4 | [2.1019e-4, 7.5320e-4] |

| By Origin | | Frequency | Clopper Pearson CI |
|---|---|---|---|
| Africa | 0/1900 | 0.0000e+0 | [0.0000e+0, 1.9396e-3] |
| America | 6/13829 | 4.3387e-4 | [1.5924e-4, 9.4411e-4] |
| Asia | 3/6024 | 4.9801e-4 | [1.0271e-4, 1.4547e-3] |
| Europe | 2/4374 | 4.5725e-4 | [5.5380e-5, 1.6507e-3] |

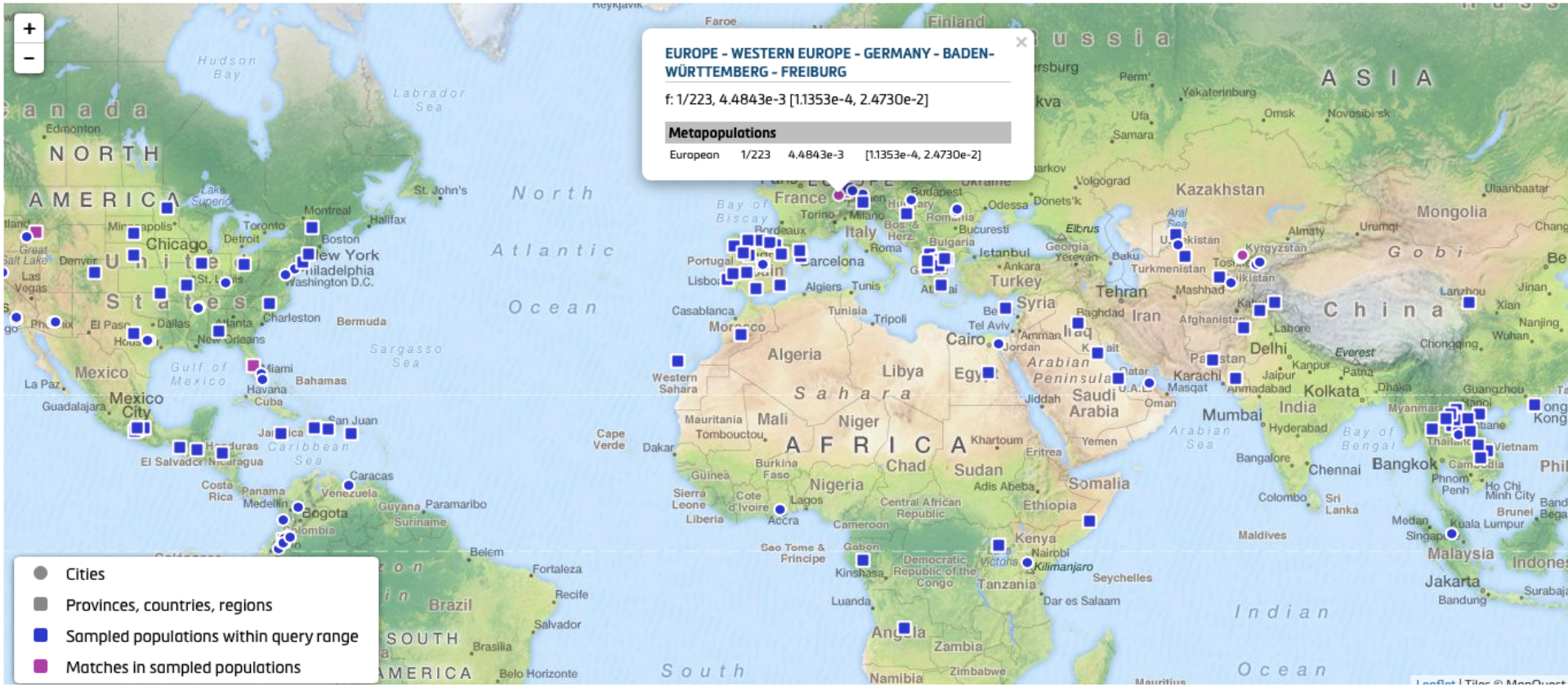
Find origin...

| By Metapopulation | | Frequency | Clopper Pearson CI |
|---|---|---|---|
| African | 2/3912 | 5.1125e-4 | [6.1920e-5, 1.8456e-3] |
| Westeurasian | 6/10687 | 5.6143e-4 | [2.0606e-4, 1.2216e-3] |
| South Asian | 0/644 | 0.0000e+0 | [0.0000e+0, 5.7117e-3] |
| East Asian | 0/2923 | 0.0000e+0 | [0.0000e+0, 1.2612e-3] |
| Southeast Asian | 0/1144 | 0.0000e+0 | [0.0000e+0, 3.2194e-3] |
| North Asian | 0/166 | 0.0000e+0 | [0.0000e+0, 2.1977e-2] |
| Native American | 0/1972 | 0.0000e+0 | [0.0000e+0, 1.8689e-3] |
| Native American Admixed | 2/2091 | 9.5648e-4 | [1.1586e-4, 3.4508e-3] |
| US Hispanics | 1/2588 | 3.8640e-4 | [9.7828e-6, 2.1510e-3] |

Find metapopulation...

# Example:

query rCRS (empty profile) in CR range (16024-576)



**Example:**

query rCRS (empty profile) in CR range (16024-576)

QUERY | POPULATIONS | TOOLS

Query | Result | Details | Neighbors

Sample ID rCRS Control Region
Ranges 16024-576
Profile

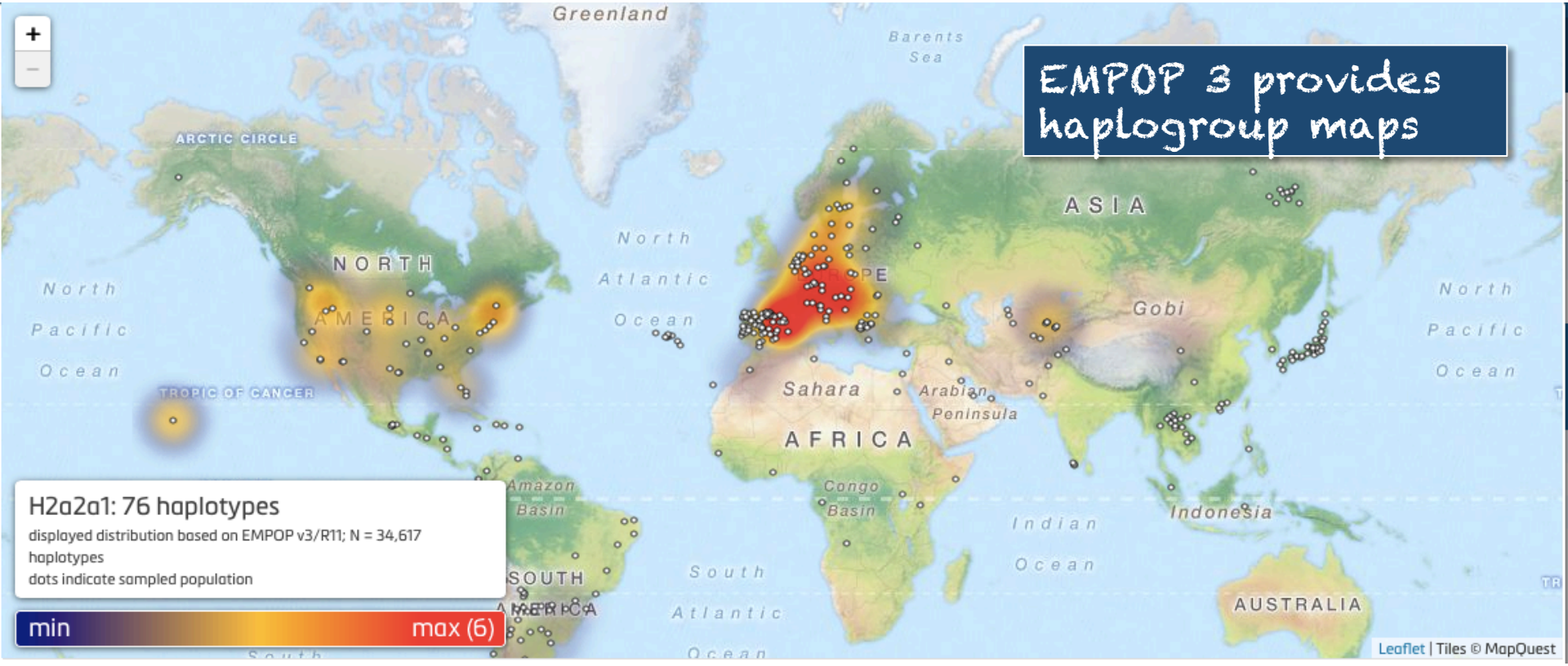
EMPOP 3 provides haplogroup info

11 of 11 haplotypes shown

| Origin | | | | | Metapopulation | Ignored Mutations | Haplogroup | | Publications |
|---|---|---|---|---|---|---|---|---|---|
| filter origins | | | | | filter metapopu | | filter haplogroup | | |
| Continent | Region | Country | Province | City | | | Rank 1 | Rank 2 | |
| Asia | Central Asia | Uzbekistan | Tashkent province | Chirchik | Westeurasian | | H2a2a1 | R0 | Irwin 2010 |
| Europe | Western Europe | Germany | Southwest Germany | | Westeurasian | | H2a2a1 | R0 | Lutz-Bonengel 2009 |
| Asia | Central Asia | Uzbekistan | Fergana | Ohunbabaev | Westeurasian | | H2a2a1 | R0 | Irwin 2010 |
| America | South America | Argentina | Misiones | | Native American Admixed | | H2a2a1 | R0 | Bobillo 2010 |
| Asia | Central Asia | Uzbekistan | Tashkent province | Chirchik | Westeurasian | | H2a2a1 | R0 | Irwin 2010 |
| America | Northern America | United States of America | Florida | | Westeurasian | | H2a2a1 | R0 | AFDIL 2011 |
| America | Northern America | United States of America | Idaho | | US Hispanics | | H2a2a1 | R0 | AFDIL 2012 |
| Europe | Western Europe | Germany | Baden-Württemberg | Freiburg | Westeurasian | | H2a2a1 | R0 | Lutz-Bonengel 2012 |
| America | Northern America | United States of America | Washington | | African | | H2a2a1 | R0 | AFDIL 2012 |
| America | Northern America | United States of America | California | Orange County | African | | H2a2a1 | R0 | AFDIL 2013 |
| America | South America | Argentina | Rio Negro | | Native American Admixed | | H2a2a1 | R0 | Bobillo 2010 |

**Example:**

query rCRS (empty profile) in CR range (16024-576)



# Example:

query rCRS (empty profile) in CR range (16024-576)

QUERY | POPULATIONS | TOOLS

Query | Result | Details | Neighbors

EMPOP 3 provides up to two step neighbours including cost values

Sample ID (none specified)
Ranges 16024-576
Profile

92 of 92 haplotypes shown

| Origin | | | | | Metapopulation | Cost | Count | Mutations | Ignored Mutations | Haplogroup | | Publications |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| filter origins | | | | | filter metapopu | | filter | | | filter haplogroup | | |
| Continent | Region | Country | Province | City | | | | | | Rank 1 | Rank 2 | |
| Europe | Southern Europe | Spain | Asturias | | Westeurasian | 2.41 | 2 | G263A (0.75) 315.1delC (0.77) | Y199T (0.00) 573.1delC 573.2delC 573.3delC (0.89) | H | R0 | Prieto 2011 |
| Europe | Northern Europe | Sweden | Uppsala | | Westeurasian | 1.52 | 2 | G263A (0.75) 315.1delC (0.77) | Y195T (0.00) | R0 | H | Tillmar 2010 |
| Europe | Eastern Europe | Hungary | Budapest | Budapest | Westeurasian | 1.52 | 2 | G263A (0.75) 315.1delC (0.77) | Y16056C (0.00) | R0 | H | Irwin 2007 |
| America | Northern America | United States of America | Colorado | | Westeurasian | 1.52 | 2 | G263A (0.75) 315.1delC (0.77) | Y152T (0.00) | R0 | H | AFDIL 2011 |
| America | Northern America | United States of America | California | Orange County | Westeurasian | 0.77 | 1 | 315.1delC (0.77) | R484A (0.00) | H2a2a1 | H2a2a1 | AFDIL 2013 |
| Asia | Western Asia | Kuwait | | | Westeurasian | 1.94 | 2 | G263A (0.75) 315.1delC (0.77) | R16183A (0.00) 309.1delC 309.2delC (0.42) | R0 | H | Scheible 2011 |

EMPOP accession number

Text

Geographic Affiliation: Spain

Authors

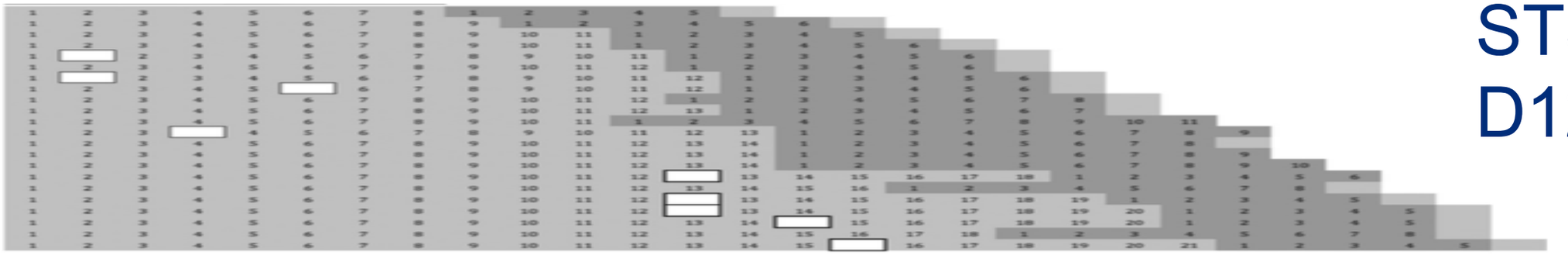
Metapopulation

Submit

1567 samples within 11 populations found.

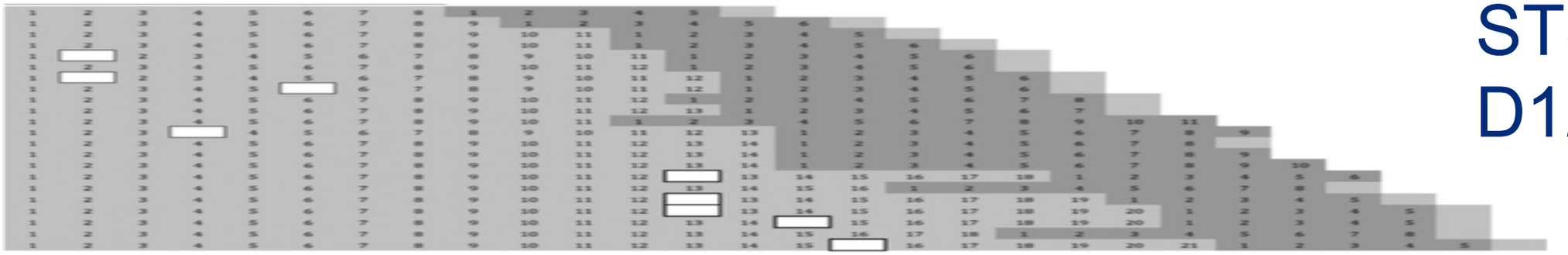
| EmpAcc# | Count | Origin | | | | | Metapopulation | | | Publication |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Continent | Region | Country | Province | City | L1 | L2 | L3 | |
| EMP00023 | 154 | Europe | Southern Europe | Spain | | | Westeurasian | European | | Crespillo 2006 |
| EMP00024 | 154 | Europe | Southern Europe | Spain | Basque Country | Madrid | Westeurasian | European | | Prieto 2011 |
| EMP00290 | 249 | Europe | Southern Europe | Spain | Madrid | Andalucia | Westeurasian | European | | Prieto 2011 |
| EMP00293 | 84 | Europe | Southern Europe | Spain | Basque Country | | Westeurasian | European | | Prieto 2011 |
| EMP00365 | 106 | Europe | Southern Europe | Spain | Basque Country | | Westeurasian | European | | Cardoso 2012 |
| EMP00400 | 61 | Europe | Southern Europe | Spain | Cantabria | | Westeurasian | European | | Cardoso 2010 |
| EMP00533 | 438 | Europe | Southern Europe | Spain | Val d'Aran | | Westeurasian | European | | Lopez Parra 2012 |

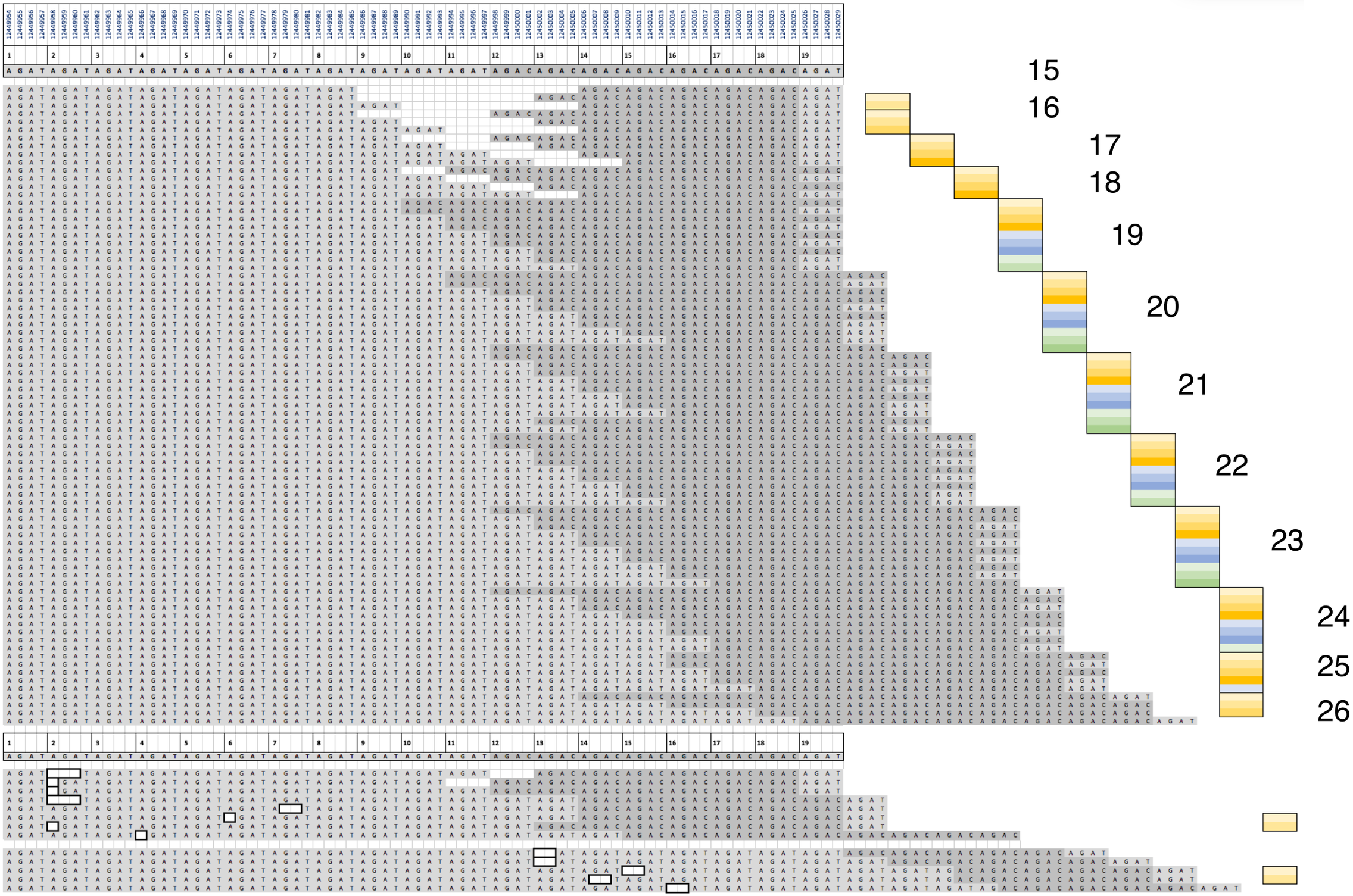
# STS to MPS
# D12S391



| | GAT | | | AGAT | | | | | | AGAC | | | AT | | | | AGAT | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 14 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 1 | 2 | 3 | 4 | 5 | | | | | | | | | | | |
| 16 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 1 | 2 | 3 | 4 | 5 | 6 | | | | | | | | | |
| 17 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 1 | 2 | 3 | 4 | 5 | | | | | | | | |
| 18 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 1 | 2 | 3 | 4 | 5 | 6 | | | | | | | |
| 18.3 | 1 | [ ] | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 1 | 2 | 3 | 4 | 5 | 6 | | | | | | |
| 19 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 1 | 2 | 3 | 4 | 5 | 6 | | | | | | |
| 19.3 | 1 | [ ] | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 1 | 2 | 3 | 4 | 5 | 6 | | | | | |
| 19.3 | 1 | 2 | 3 | 4 | 5 | [ ] | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 1 | 2 | 3 | 4 | 5 | 6 | | | | | |
| 21 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | | | | |
| 21 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | | | | |
| 22 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | | |
| 22.3 | 1 | 2 | 3 | [ ] | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | |
| 23 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | | |
| 23 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | |
| 24 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 24.2 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | [ ] | 13 | 14 | 15 | 16 | 17 | 18 | 1 | 2 | 3 | 4 | 5 |
| 25 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 25.2 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | [ ] | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 1 | 2 | 3 | 4 |
| 26.2 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | [ ] | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 1 | 2 | 3 |
| 26.2 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | [ ] | 15 | 16 | 17 | 18 | 19 | 20 | 1 | 2 | 3 |
| 27 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 1 | 2 | 3 | 4 | 5 | 6 |
| 27.2 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | [ ] | 16 | 17 | 18 | 19 | 20 | 21 | 1 | 2 |

# STS to MPS
# D12S391





# STR sequence variation analysis will add a lot of power
## ..in some markers, but not all

| STR ranked by repeat allele (RA) heterozygosity | Sequence heterozygotes detected in isometric homozygotes | Proportion of sequence heterozygotes in isometric homozygotes | No. of repeat alleles | No. of sequence alleles | No. of genotypes reported | Repeat allele heterozygosity | Sequence allele heterozygosity | Rise in power with MPS |
|---|---|---|---|---|---|---|---|---|
| **DXS10103** | 2.15% | 28% | 44 | 111 | 1244 | 92.31% | 98.22% | **6.4%** |
| **D12S391** | 5.72% | 36.73% | 25 | 91 | 1785 | 84.43% | 93.53% | **10.78%** |
| **D2S1338** | 3.18% | 19.87% | 12 | 61 | 1767 | 84% | 92.74% | **10.39%** |
| **D21S11** | 6.57% | 35.23% | 23 | 89 | 1772 | 81.36% | 92.95% | **14.25%** |
| **D8S1179** | 7.31% | 35.2% | 11 | 26 | 1761 | 79.24% | 90.57% | **14.31%** |
| **vWA** | 2.33% | 10.53% | 11 | 35 | 1701 | 77.86% | 85.15% | **9.36%** |
| **D13S317** | 7.2% | 30.91% | 9 | 25 | 1736 | 76.69% | 88.92% | **15.94%** |
| **D3S1358** | 8.05% | 30.65% | 10 | 22 | 1716 | 73.73% | 86.57% | **17.42%** |
| **D5S818** | 4.45% | 16.87% | 10 | 23 | 1681 | 73.62% | 85.42% | **16.02%** |
| **D1S1656** | 2.54% | 9.56% | 18 | 36 | 1657 | 73.41% | 90.48% | **23.24%** |
| **D4S2408** | 3.71% | 12.11% | 8 | 12 | 1634 | 69.39% | 79.95% | **15.22%** |
| **D9S1122** | 10.38% | 33.68% | 9 | 19 | 1695 | 69.17% | 83.78% | **21.12%** |
| **D2S441** | 3.6% | 11.04% | 15 | 27 | 1622 | 67.37% | 82.17% | **21.96%** |

# STR sequence nomenclature discussions - timelines

- **Discussed STR sequence nomenclature issues at EDNAP Zurich 2014**
  Katherine Gettings presented findings from NIST's MPS analysis of core STRs

# STR sequence nomenclature discussions - timelines

- **Discussed STR sequence nomenclature issues at EDNAP Zurich 2014**
  Katherine Gettings presented findings from NIST's MPS analysis of core STRs

- **Illumina launched the ForenSeq kit in January 2015**
  27 A-STRs, 24 Y-STRs, 7 X-STRs

# STR sequence nomenclature discussions - timelines

- **Discussed STR sequence nomenclature issues at EDNAP Zurich 2014**
  Katherine Gettings presented findings from NIST's MPS analysis of core STRs

- **Illumina launched the ForenSeq kit in January 2015**
  27 A-STRs, 24 Y-STRs, 7 X-STRs

- **2015: USC/KCL/Illumina analysed the 944 sample CEPH panel with *ForenSeq***
  Runs completed in August and data processing ongoing, but all STR sequence variation is now compiled - provides a good overview of 58 STRs

# STR sequence nomenclature discussions - timelines

- **Discussed STR sequence nomenclature issues at EDNAP Zurich 2014**
  Katherine Gettings presented findings from NIST's MPS analysis of core STRs

- **Illumina launched the ForenSeq kit in January 2015**
  27 A-STRs, 24 Y-STRs, 7 X-STRs

- **2015: USC/KCL/Illumina analysed the 944 sample CEPH panel with *ForenSeq***
  Runs completed in August and data processing ongoing, but all STR sequence variation is now compiled - provides a good overview of 58 STRs

- **September: Round table discussion at ISFG Krakow began collection of ideas on nomenclature and oversight of data**
  Several suggestions, but agreement to require flanking SNPs from suppliers

# STR sequence nomenclature discussions - timelines

- **Discussed STR sequence nomenclature issues at EDNAP Zurich 2014**
  Katherine Gettings presented findings from NIST's MPS analysis of core STRs

- **Illumina launched the ForenSeq kit in January 2015**
  27 A-STRs, 24 Y-STRs, 7 X-STRs

- **2015: USC/KCL/Illumina analysed the 944 sample CEPH panel with *ForenSeq***
  Runs completed in August and data processing ongoing, but all STR sequence variation is now compiled - provides a good overview of 58 STRs

- **September: Round table discussion at ISFG Krakow began collection of ideas on nomenclature and oversight of data**
  Several suggestions, but agreement to require flanking SNPs from suppliers

- **September: Katherine Gettings' FSIGEN review published and e-mail discussions to clarify STR sequence alignment issues and current repeat motif inconsistencies**

# STR sequence nomenclature discussions - timelines

- **Discussed STR sequence nomenclature issues at EDNAP Zurich 2014**
  Katherine Gettings presented findings from NIST's MPS analysis of core STRs

- **Illumina launched the ForenSeq kit in January 2015**
  27 A-STRs, 24 Y-STRs, 7 X-STRs

- **2015: USC/KCL/Illumina analysed the 944 sample CEPH panel with *ForenSeq***
  Runs completed in August and data processing ongoing, but all STR sequence variation is now compiled - provides a good overview of 58 STRs

- **September: Round table discussion at ISFG Krakow began collection of ideas on nomenclature and oversight of data**
  Several suggestions, but agreement to require flanking SNPs from suppliers

- **September: Katherine Gettings' FSIGEN review published and e-mail discussions to clarify STR sequence alignment issues and current repeat motif inconsistencies**

- **ISFG board has charged Walther with coordination of a guidelines document to highlight forensic needs and the challenges of annotating genomic data**

# Sequence alignment and sequence annotation

- **Alignment of an STR sequence string to a community-wide agreed human genome assembly**

  GRCh37 and GRCh38 - how easy is it to interchange these integer sets?

  Are the assemblies identical for forensic loci or is there extra sequence?

  How often will new assemblies be published in the future?

# Sequence alignment and sequence annotation

- **Alignment of an STR sequence string to a community-wide agreed human genome assembly**

  GRCh37 and GRCh38 - how easy is it to interchange these integer sets?

  Are the assemblies identical for forensic loci or is there extra sequence?

  How often will new assemblies be published in the future?

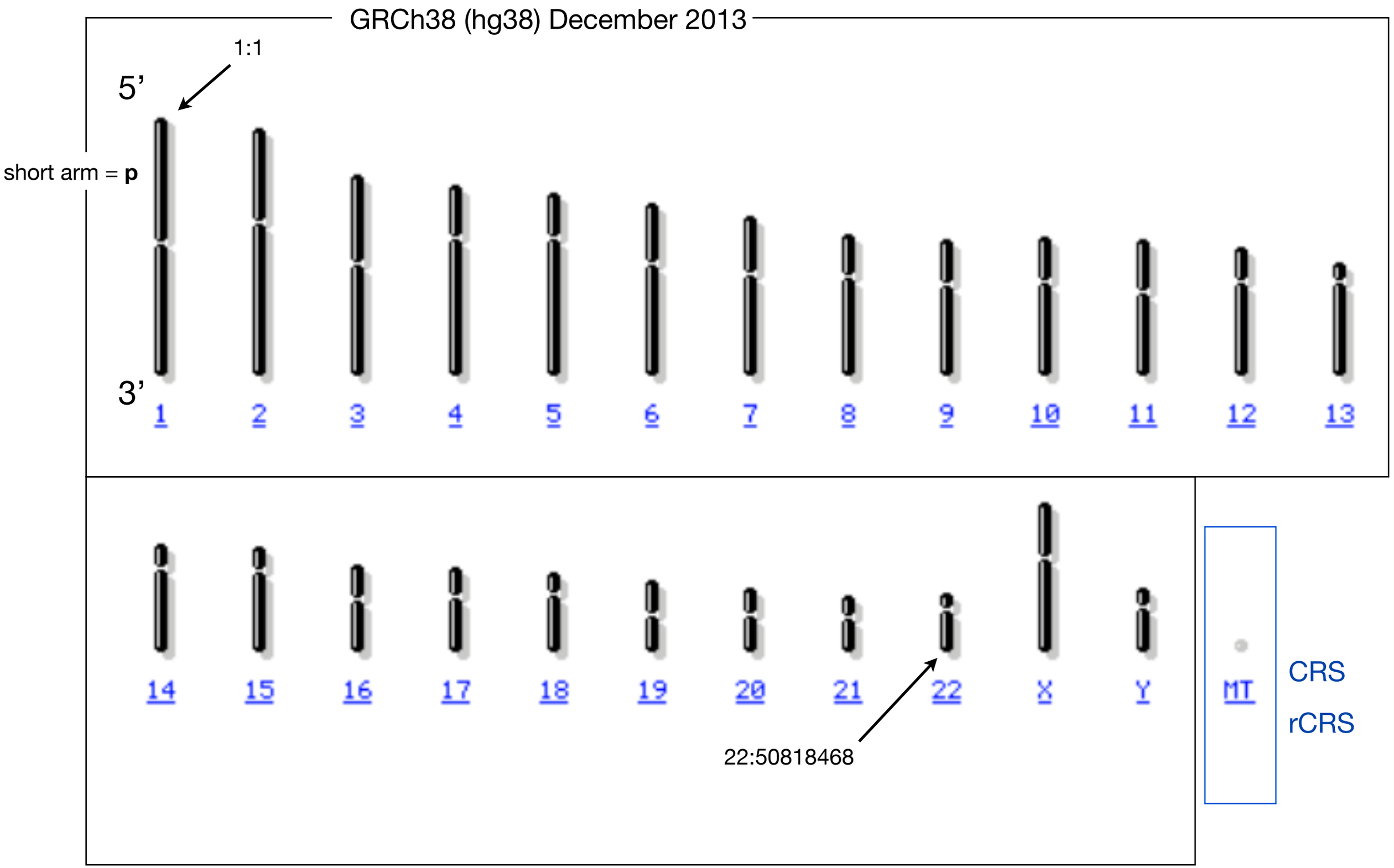
- **Annotation of repeat sequence features and SNPs and/or Indels in the flanking regions**

  Common flanking region markers generally have rs-numbers, but rare variants will also occur - but can use genomic coordinates  e.g. 17:4588194

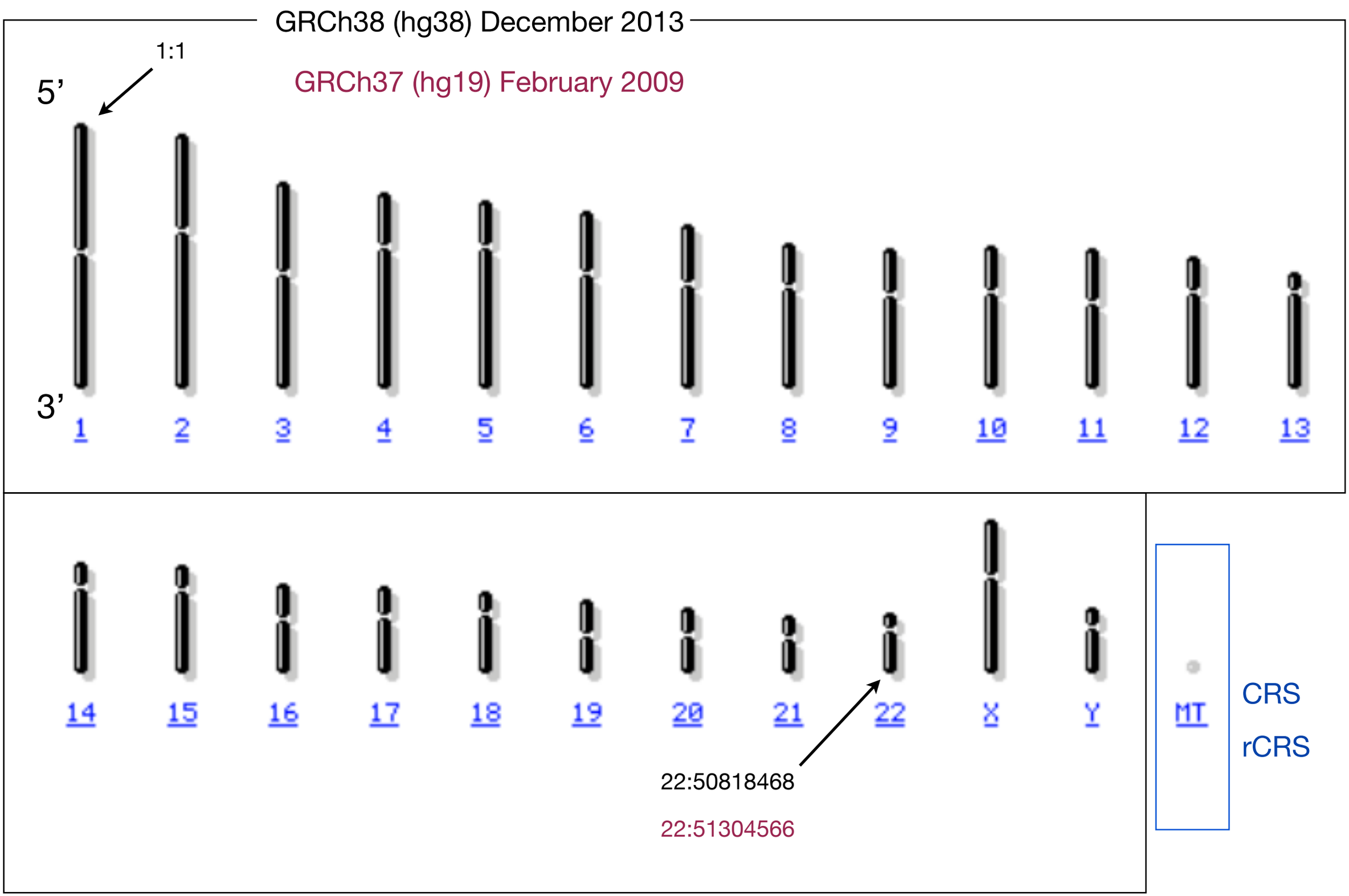
  Are there inconsistencies in the way repeat motifs are currently labelled?

  Can the current human assemblies be used to describe repeat motifs?

# Human genome assemblies



# Human genome assemblies

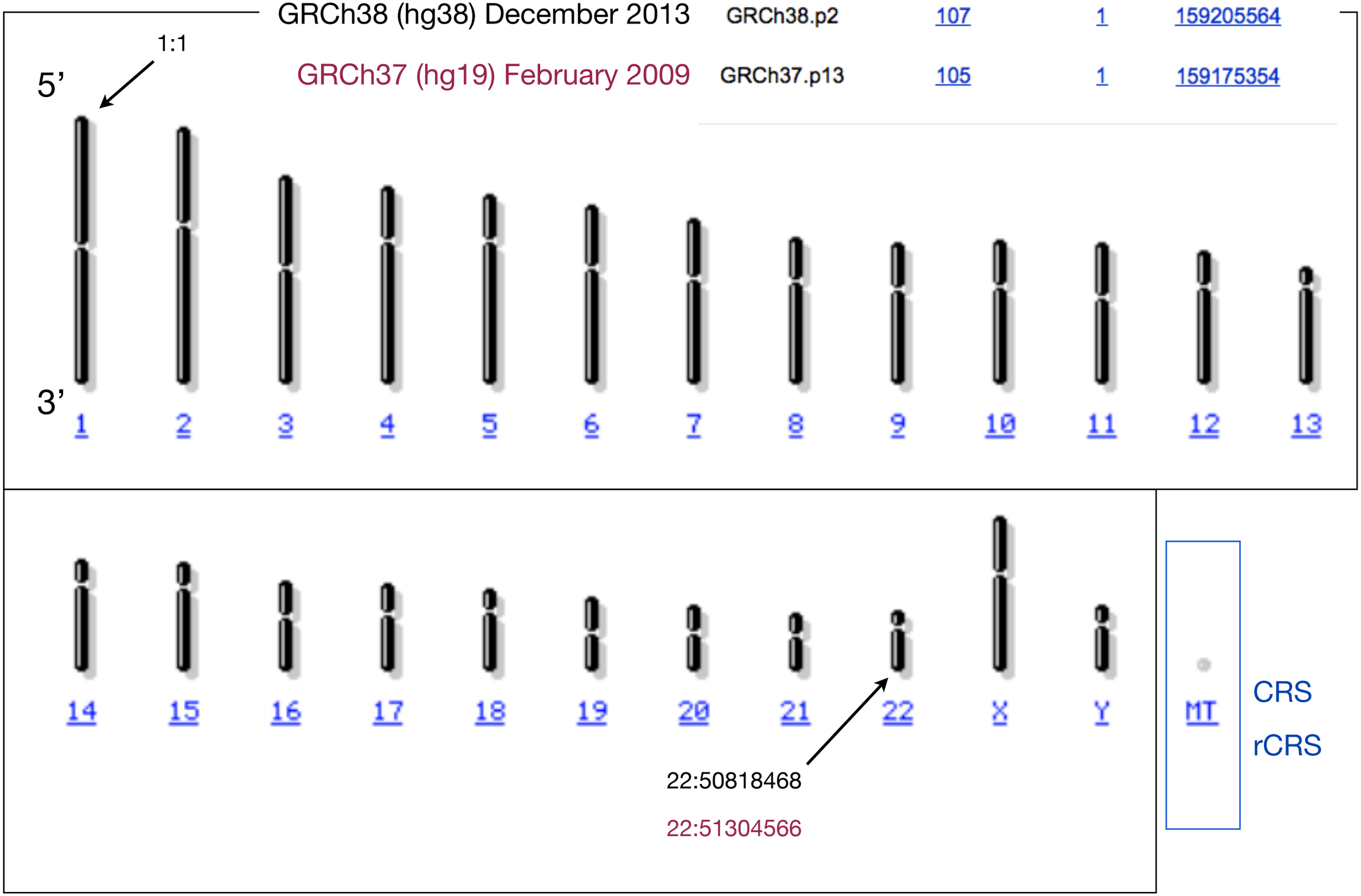


# Human genome assemblies

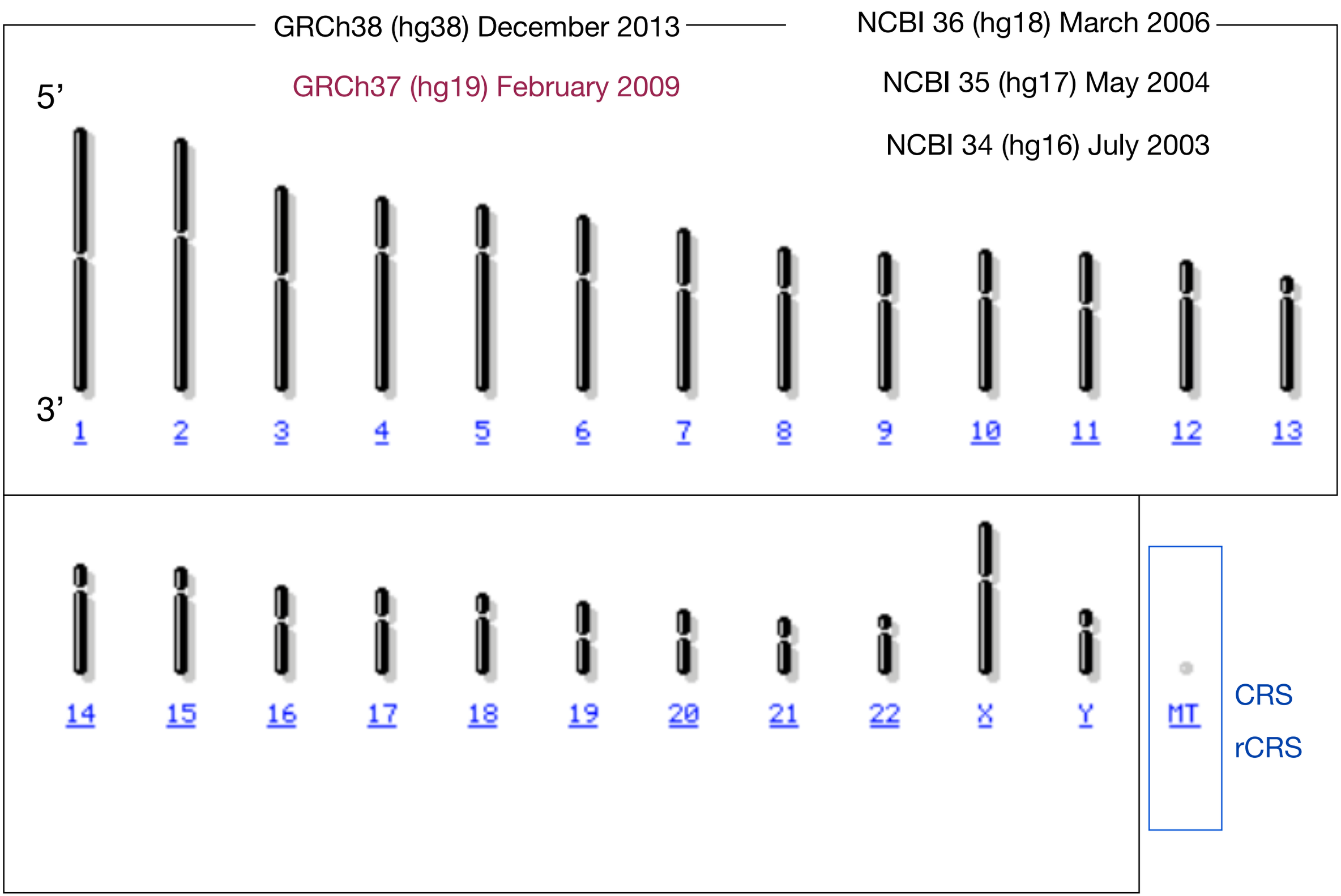
| Assembly | Annotation Release | Chr | Chr Pos |
|---|---|---|---|
| GRCh38.p2 | 107 | 1 | 159205564 |
| GRCh37.p13 | 105 | 1 | 159175354 |

GRCh38 (hg38) December 2013

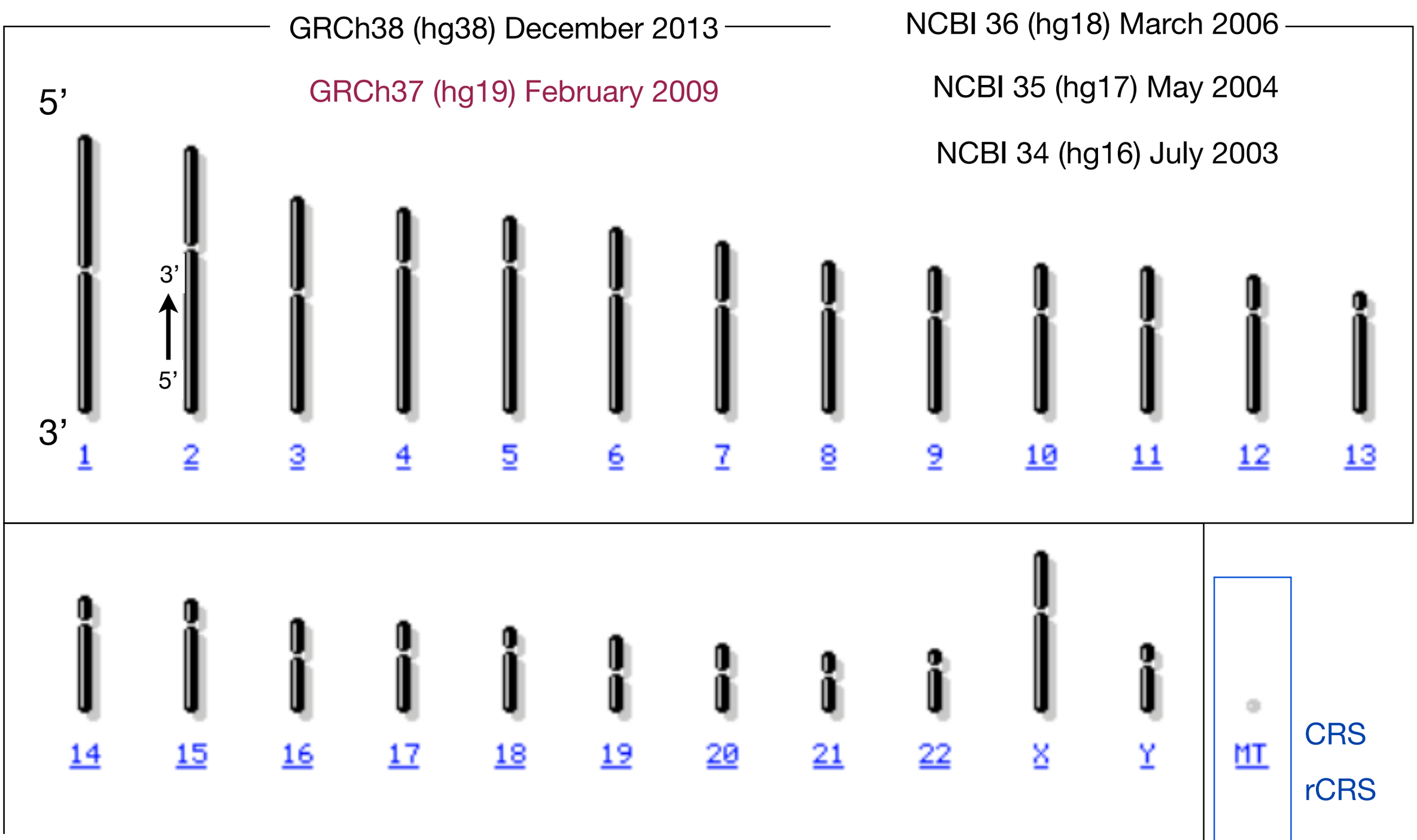
GRCh37 (hg19) February 2009



# Human genome assemblies - frequency of new builds



# Many established repeat descriptions are reverse strand

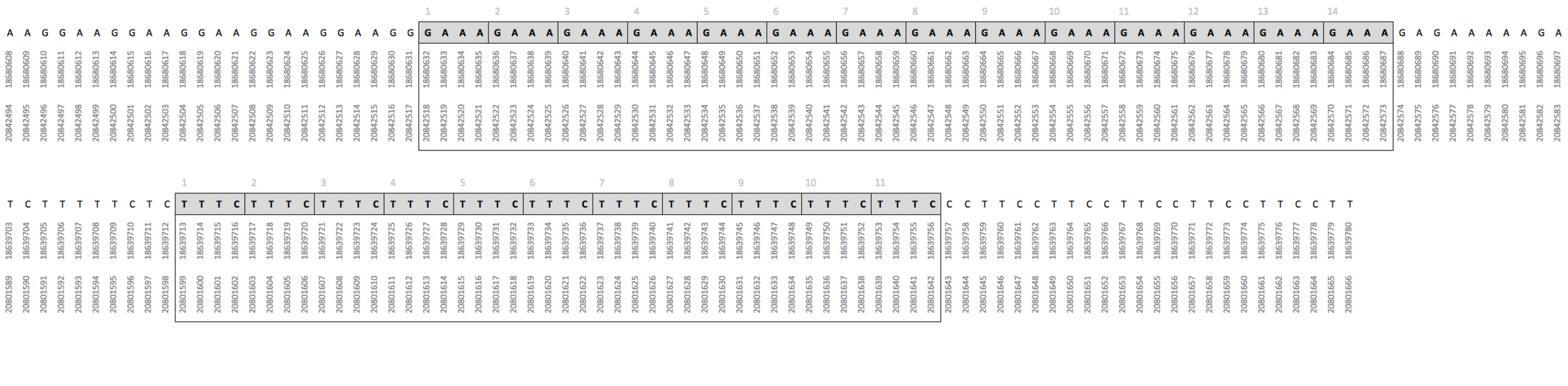



**D2S1338** [TGCC]a [TTCC]b - STRbase

**D2S1338** [GGAA]b [GGCA]a - reference genome

# DYS385 comprises two repeat regions that are inverses

GAAA




TTTC

# Both current assemblies map exactly for 55 of 58 STRs compared so far

Three core forensic STRs show nucleotide differences in their repeat regions between GRCh37 and GRCh38

Katherine Gettings' FSIGEN review refers to GRCh38 coordinates only (and no X/Y loci)

**GATA** Repeat region sequence (with line breaks)
**GATA** Extra sequence tracts in the genome assembly

DYS437 (GRCh38)
GGACTATGGGCGTGAGTGCATGCCCATCCGG
**TCTATCTATCTATCTATCTATCTATCTATCTATCTATCTGTCTGTCTATCTATCTAT**
**CTA**
TCATCTATCATCTGTGAATGATGTCTATCTA

DYS437 (GRCh37)
GGACTATGGGCGTGAGTGCATGCCCATCCGG
**TCTATCTATCTATCTATCTATCTATCTATCTATCTA**TCTA**TCTGTCTGTCTATCTAT**
**CTATCTA**
TCATCTATCATCTGTGAATGATGTCTATCTA

DYS438 (GRCh38)
AAAATTAGTGGGGAATAGTTGAACGGTAAACAGTATA
**TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTC**TTTTCTT
**TTC**
TATTTGAAATGGAGTTTCACTCTTGTTGCCCAGGCTG

DYS438 (GRCh37)
AAAATTAGTGGGGAATAGTTGAACGGTAAACAGTATA
**TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTC**
TATTTGAAATGGAGTTTCACTCTTGTTGCCCAGGCTG

DYS439 (GRCh38)
GATACATAGGTGGAGACAGATAGATGATAAATAGAA
**GATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA**
GAAAGTATAAGTAAAGAGATGATGGGTAAAAGAATT

DYS439 (GRCh37)
GATACATAGGTGGAGACAGATAGATGATAAATAGAA
**GATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA**GATA
GAAAGTATAAGTAAAGAGATGATGGGTAAAAGAATT

# Sequence alignment and sequence annotation

- **Alignment of an STR sequence string to a community-wide agreed human genome assembly**

  GRCh37 and GRCh38 - how easy is it to interchange these integer sets?

  Are the assemblies identical for forensic loci or is there extra sequence?

  How often will new assemblies be published in the future?

- **Annotation of repeat sequence features and SNPs and/or Indels in the flanking regions**

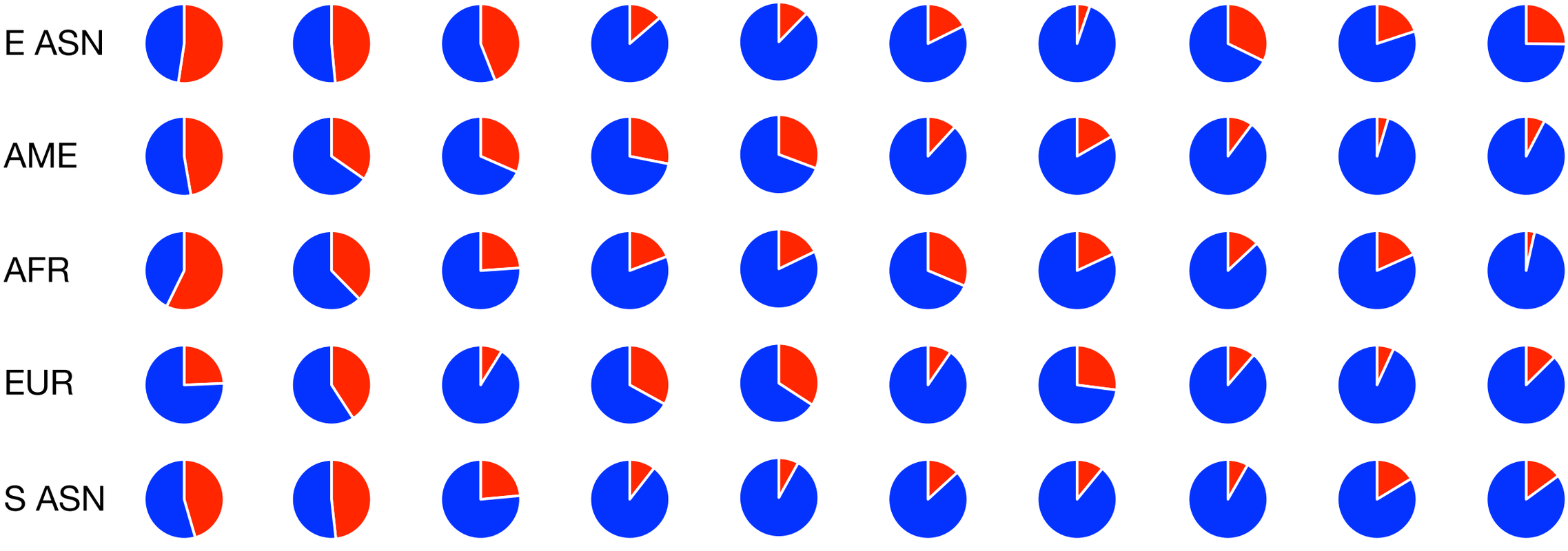
  Common flanking region markers generally have rs-numbers, but rare variants will also occur - but can use genomic coordinates  e.g. 17:4588194

  Are there inconsistencies in the way repeat motifs are currently labelled?

  Can the current human assemblies be used to describe repeat motifs?

# Ten informative flanking region SNPs in amplicons of +/-120 nt around the repeat region - suppliers not revealing primer-ends

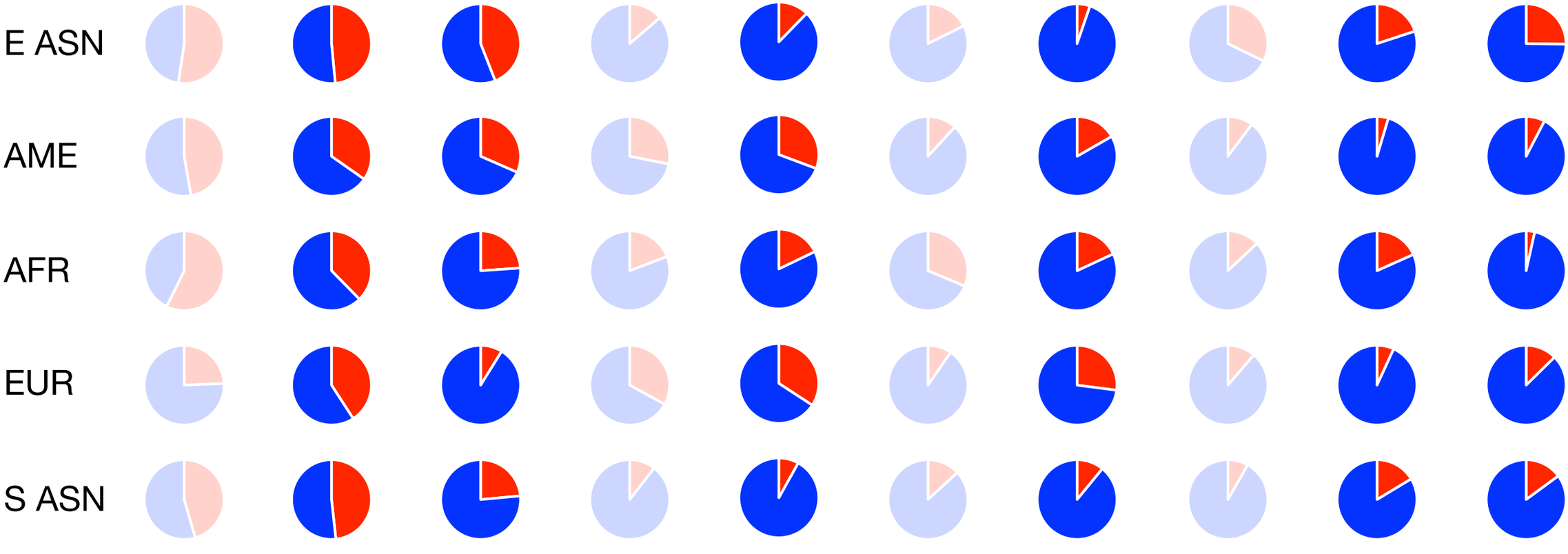
D2S441-rs58390469
D13S317-rs9546005
D16S539-rs11642858
D2S1338-rs6736691
D1S1656-rs4847015
D2S441-rs7953469
D5S818-rs25768
D10S1248-rs2246512
D7S820-rs16887642
VWA-rs75219269



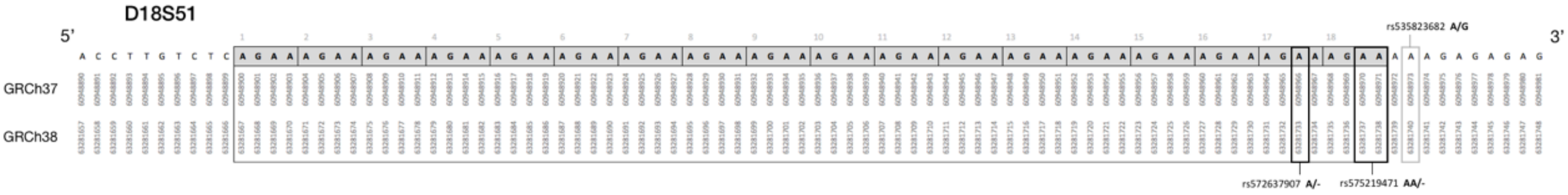
# Six flanking region SNPs in *ForenSeq* amplicons

Miniaturized STR primer designs will exclude some flanking SNPs

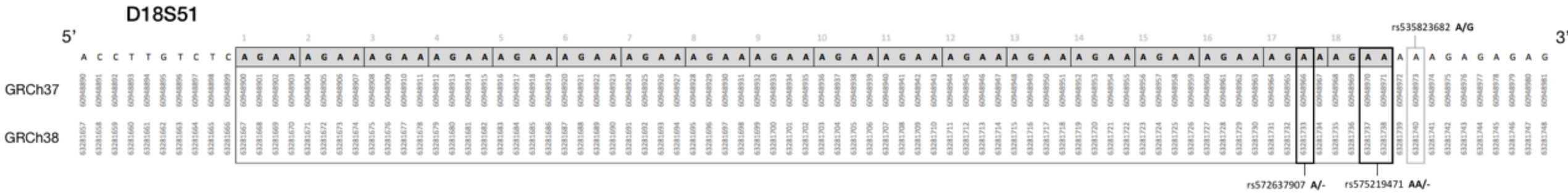
D2S441-rs58390469
D13S317-**rs9546005**
D16S539-**rs11642858**
D2S1338-rs6736691
D1S1656-**rs4847015**
D2S441-rs7934691
D5S818-**rs25768**
D10S1248-rs2246512
D7S820-**rs16887642**
VWA-**rs75219269**

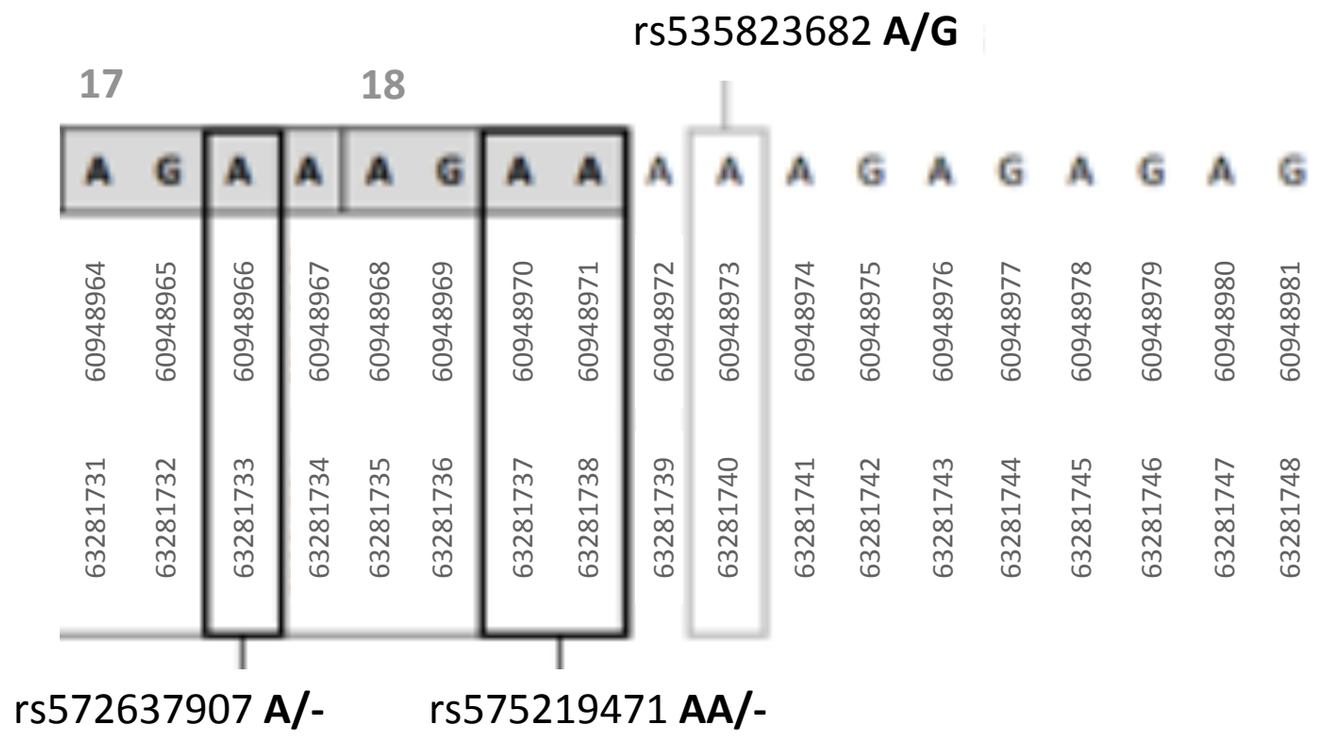


# Three STRs presenting challenges for motif description

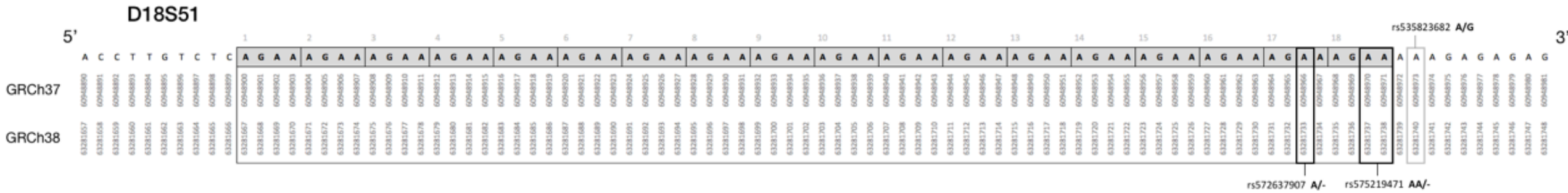


# Three STRs presenting challenges for motif description



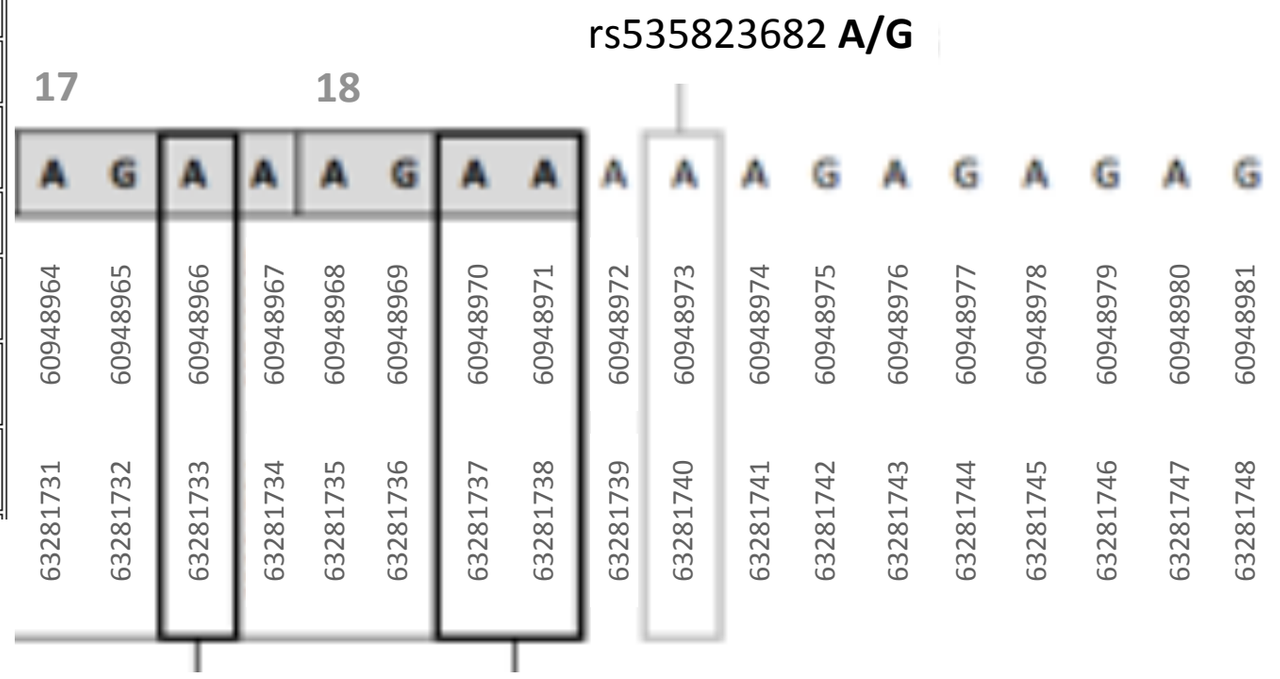



# Three STRs presenting challenges for motif description

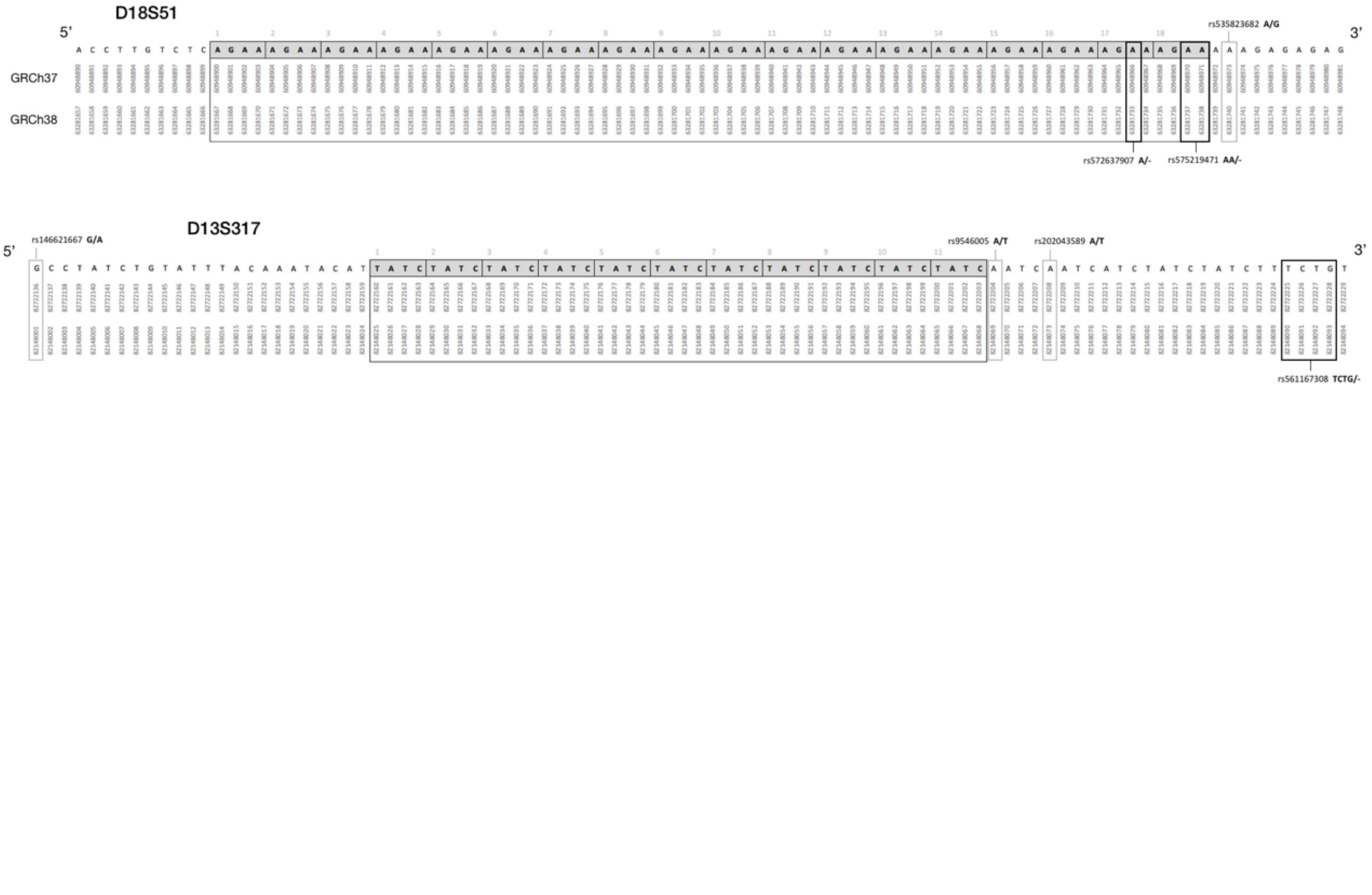



| | | | | | | |
|---|---|---|---|---|---|---|
| 13 | 286 bp | 288 bp | 310 bp | AAAG AGAGAG | $[AGAA]_{13}$ | 716 |
| 13.1 | 287 bp | 289 bp | 311 bp | | | *variant allele* |
| 13.2 | 288 bp | 290 bp | 312 bp | AAAG AG AGAGAG | $[AGAA]_{13}$ | 369 |
| 13.3 | 289 bp | 291 bp | 313 bp | | | *variant allele* |
| 14 | 290 bp | 292 bp | 314 bp | AAAG AGAGAG | $[AGAA]_{14}$ | 716 |
| 14.2 | 292 bp | 294 bp | 316 bp | AAAG AG AGAGAG | $[AGAA]_{14}$ | 369 |
| 15 | 294 bp | 296 bp | 318 bp | AAAG AGAGAG | $[AGAA]_{15}$ | 716 |

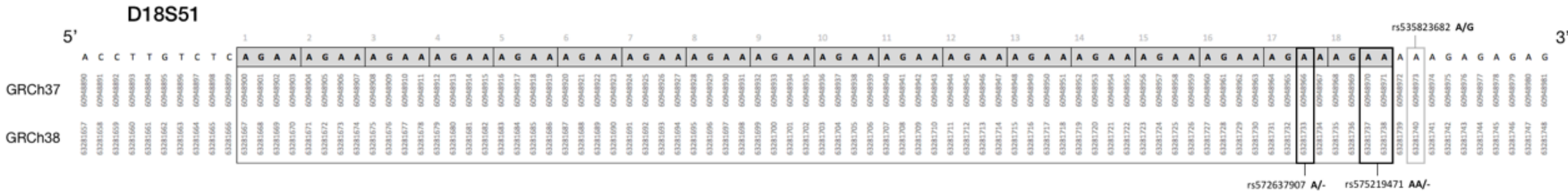
[AGAA]n AAAG AG AGAGAG



# Three STRs presenting challenges for motif description

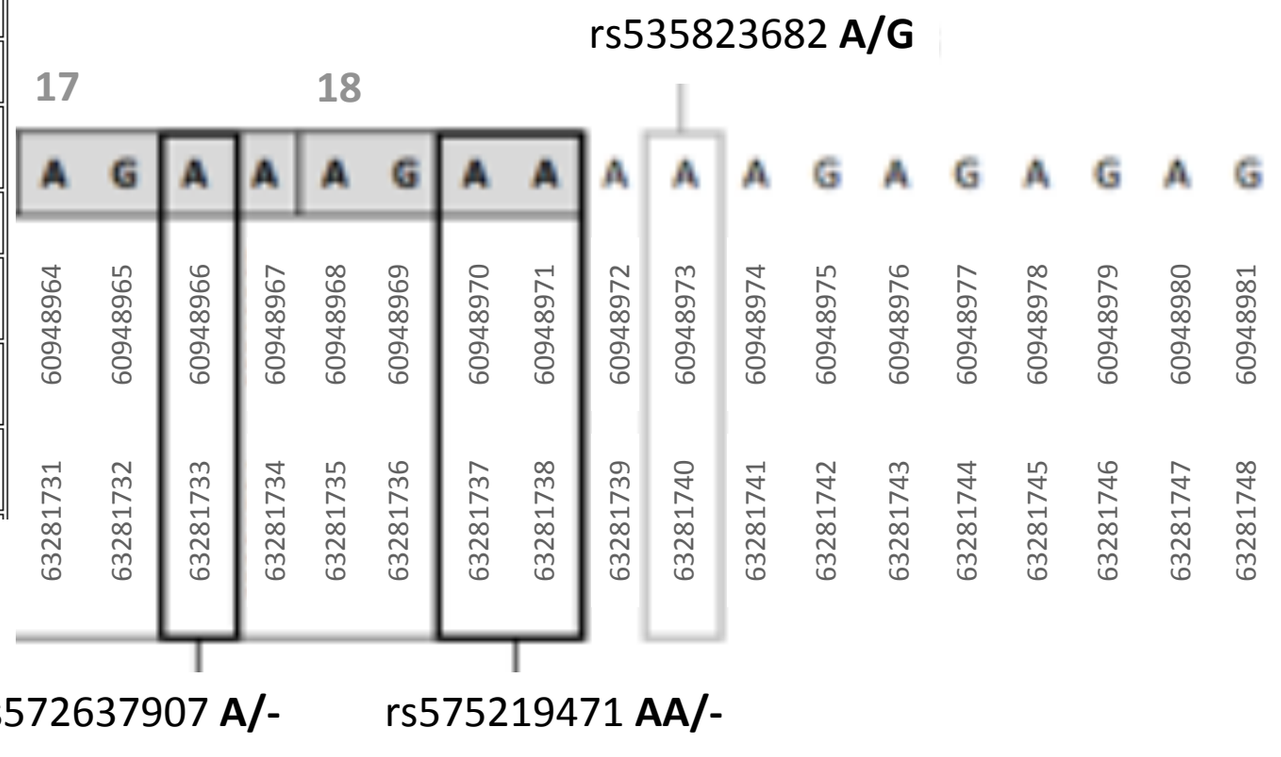


# Three STRs presenting challenges for motif description

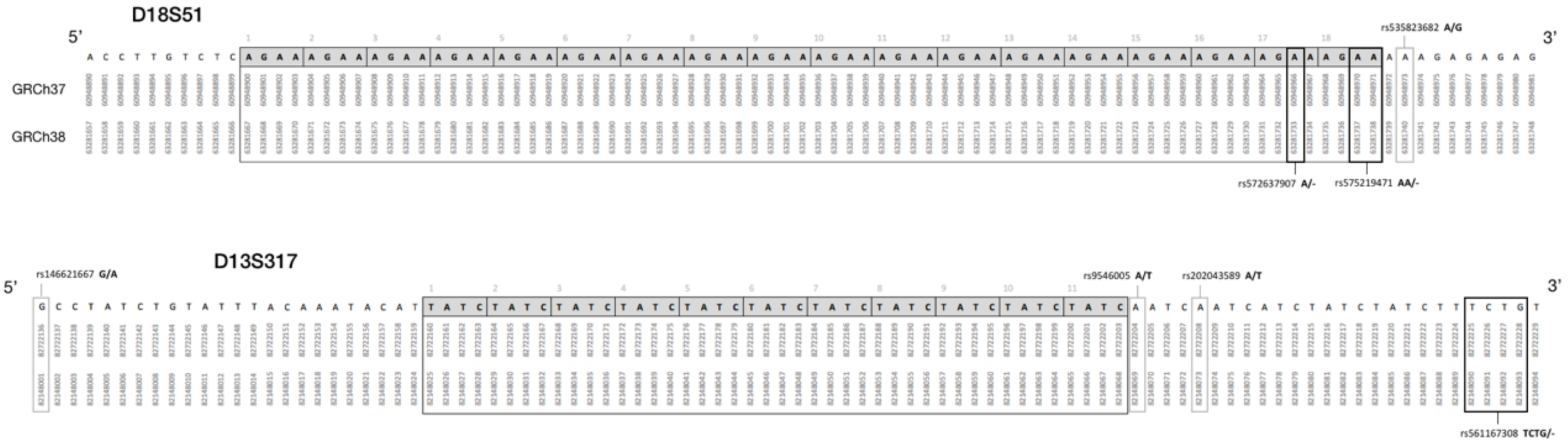


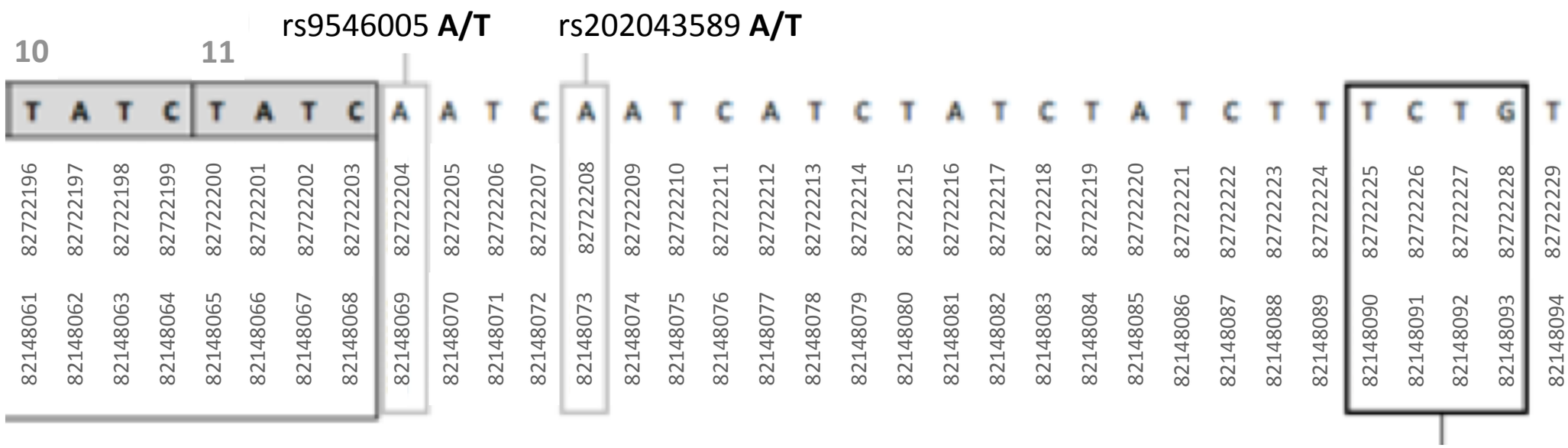

| | | | | | | |
|---|---|---|---|---|---|---|
| 13 | 286 bp | 288 bp | 310 bp | AAAG AGAGAG | $[AGAA]_{13}$ | 716 |
| 13.1 | 287 bp | 289 bp | 311 bp | | | *variant allele* |
| 13.2 | 288 bp | 290 bp | 312 bp | AAAG AG AGAGAG | $[AGAA]_{13}$ | 369 |
| 13.3 | 289 bp | 291 bp | 313 bp | | | *variant allele* |
| 14 | 290 bp | 292 bp | 314 bp | AAAG AGAGAG | $[AGAA]_{14}$ | 716 |
| 14.2 | 292 bp | 294 bp | 316 bp | AAAG AG AGAGAG | $[AGAA]_{14}$ | 369 |
| 15 | 294 bp | 296 bp | 318 bp | AAAG AGAGAG | $[AGAA]_{15}$ | 716 |

[AGAA]n AAAG AG AGAGAG



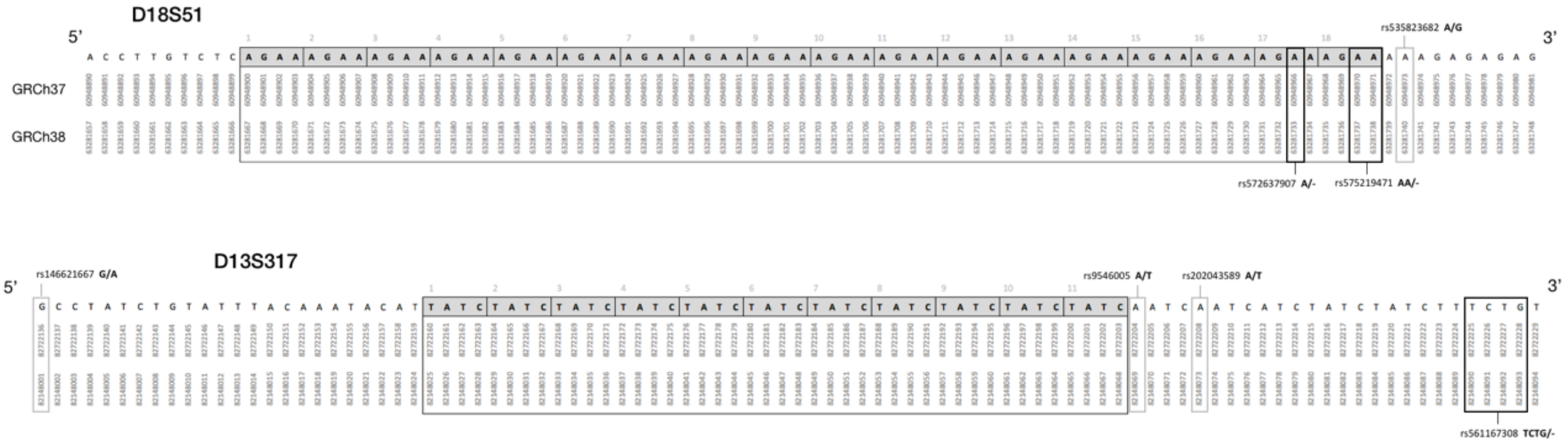
# Three STRs presenting challenges for motif description



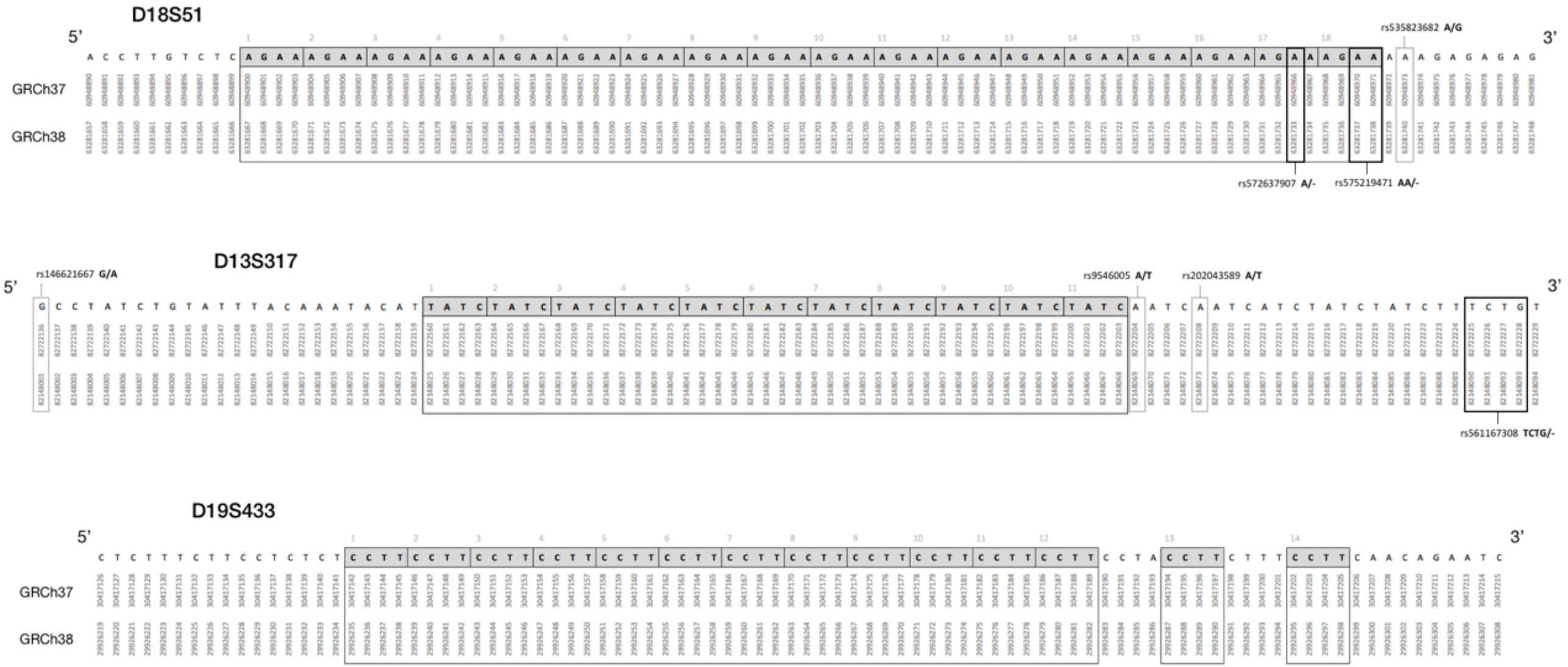




| | | | |
|---|---|---|---|
| 11 | Yes | 400 | TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCAATCAATCATCTATCTATCTTTCTGTCTGTC |
| 12 | Yes | 343 | TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCAATCAATCATCTATCTATCTTTCTGTCTGTC |

# Three STRs presenting challenges for motif description



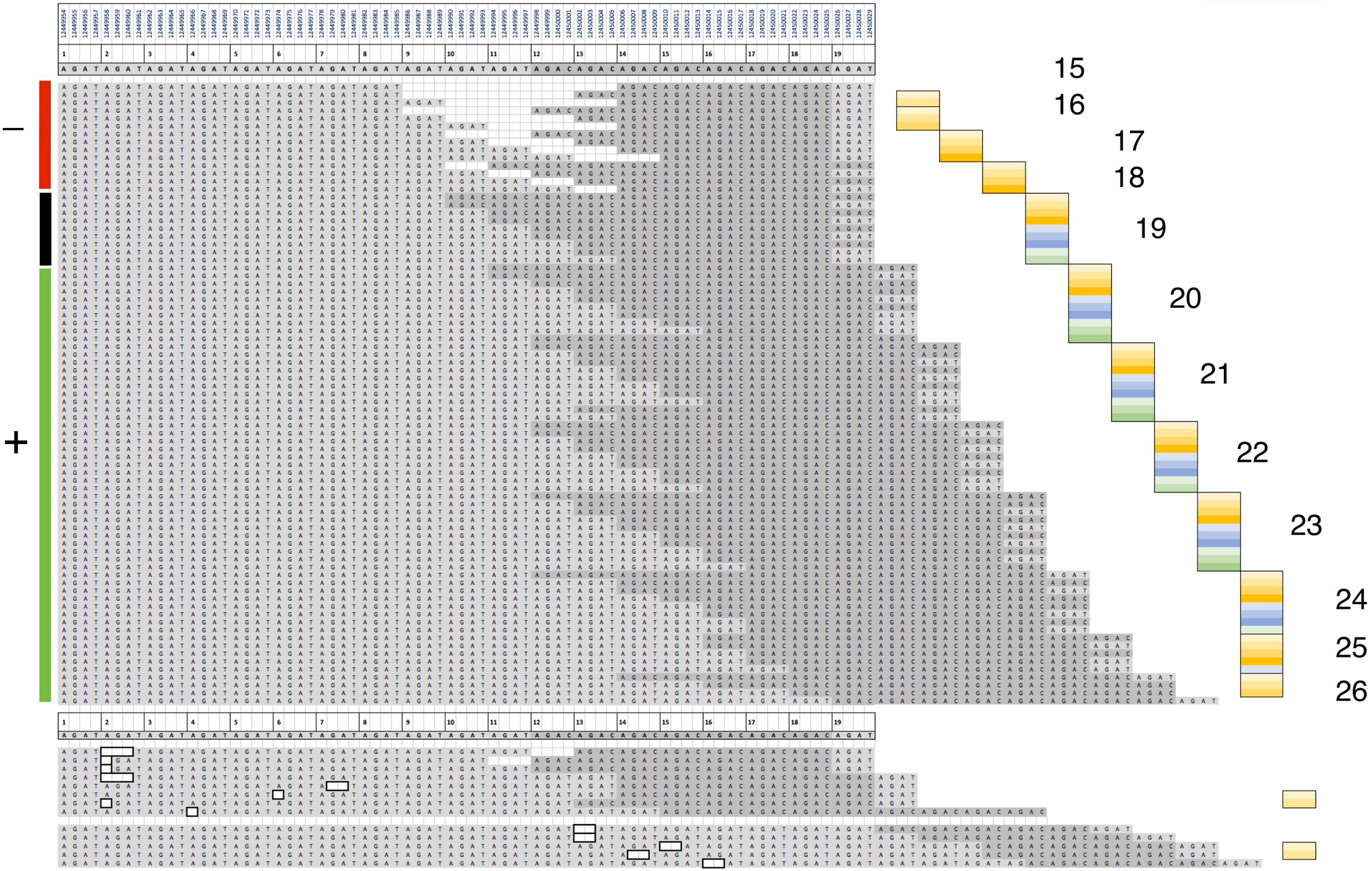
# Three STRs presenting challenges for motif description




ISFG guidelines on STR structure description indicate D19S433 repeats to be counted are [CCTT]  Kits that included this STR counted CCTA and CTTT

+2 allele numbers now populate all D19S433 in NDNADs

# Dealing with less or more nucleotides in the analysed repeat region compared to the reference sequence



```
D13S317-11      TCTAACGCCT ATCTGTATTT ACAAATACAT -TATC-TATC -TATC-TATC
D13S317[CE12]   ......A... .......... .......... -....-.... -....-....

D13S317-11      -TATC-TATC -TATC-TATC -TATC-TATC -TATC----- ------AATC
D13S317[CE12]   -....-.... -....-.... -....-.... -....-TATC -TATC-....

D13S317-11      -AATC-ATCT -ATCT-ATCT -TT
D13S317[CE12]   ------.... -....-.... -..
```

**D13S317[CE11]-Chr13-GRCh37-g.82.722.160:82.722.223-TATC[11]AATC[2]ATCT[3]**

**D13S317[CE12]-Chr13-GRCh37-g.82.722.160:82.722.223-TATC[13]AATC[1]ATCT[3]-g.x.136G>A**

**D13S317[CE12]**-Chr13-GRCh37-g.82.722.160:82.722.223-TATC[13]AATC[1]ATCT[3]-g.x.136G>A

**D13S317[CE12]: Locus name and capillary electrophoresis STR allele name**

D13S317[CE12]-**Chr13-GRCh37**-g.82.722.160:82.722.223-TATC[13]AATC[1]ATCT[3]-g.x.136G>A

**Chr13-GRCh37: Chromosome number and human genome reference version**

D13S317[CE12]-Chr13-GRCh37-**g.82.722.160:82.722.223**-TATC[13]AATC[1]ATCT[3]-g.x.136G>A

**g.82.722.160:82.722.211: STR repeat region coordinates of reference allele**

D13S317[CE12]-Chr13-GRCh37-g.82.722.160:82.722.223-**TATC[13]AATC[1]ATCT[3]**-g.x.136G>A

**TATC[13]AATC[1]ATCT[3]: full description of STR motif**

D13S317[CE12]-Chr13-GRCh37-g.82.722.160:82.722.223-TATC[13]AATC[1]ATCT[3]-**g.x.136G>A**

**g.x.136G>A: genome coordinate of derived SNP, where g.x refers to non-unique coordinate**

| | | | |
|---|---|---|---|
| 11 | Yes | 400 | TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCAATCAATCATCTATCTATCTTTCTGTCTGTC |
| 12 | Yes | 343 | TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATTAATCAATCATCTATCTATCTTTCTGTCTGTC |

# **Slide not shown:** 1000 Genomes lobSTR data released in late 2014

SATURDAY OCTOBER 18, 2014

## Short Tandem Repeats added to the 1000 Genomes Release #ASHG14

We have added two sets of STR predictions and genotypes to the 1000 Genomes dataset.

These are available in the supporting directory strs.

The call set were created using LobSTR and RepeatSeq respectively.

The sites are genotyped in all 2535 individuals who were used in our final release. This includes the 31 individuals related to other individuals in the main call set.

- A total of 670,646 micro-satellites in lobSTR format for 2,504 individuals. A bio-informatics challenge (mostly CA di-nucleotides)

- D2S1338; VWA; D12S391; D21S11 all missing from the dataset

- Long STRs in general have recognition problems, as lobSTR calls an STR by recognizing both flanking sequences plus the repeat region in reads of 100 bp, so longer STRs are missed

- Data failed to properly describe the repeat sequence variation known to occur in many core forensic STRs

# EUROFORGEN-NoE Update: EDNAP Meeting Santiago de Compostela 2015

**Peter M. Schneider**

**Institute of Legal Medicine**

**University of Cologne (Germany)**

EUROFORGEN-NoE is funded by the European Commission within the 7th Framework Programme

# Recent activities

- **Training workshops**
- **Recent EUROFORGEN-NoE publications**
- **Dissemination Conference 2016**

- **Three 'Train the Trainers' workshops in Copenhagen**

20-22 participants from all European countries

**2013 - 2015**

**Subject:** Statistical methods in forensic genetics

**Organized by:** Niels Morling

**Teachers:**

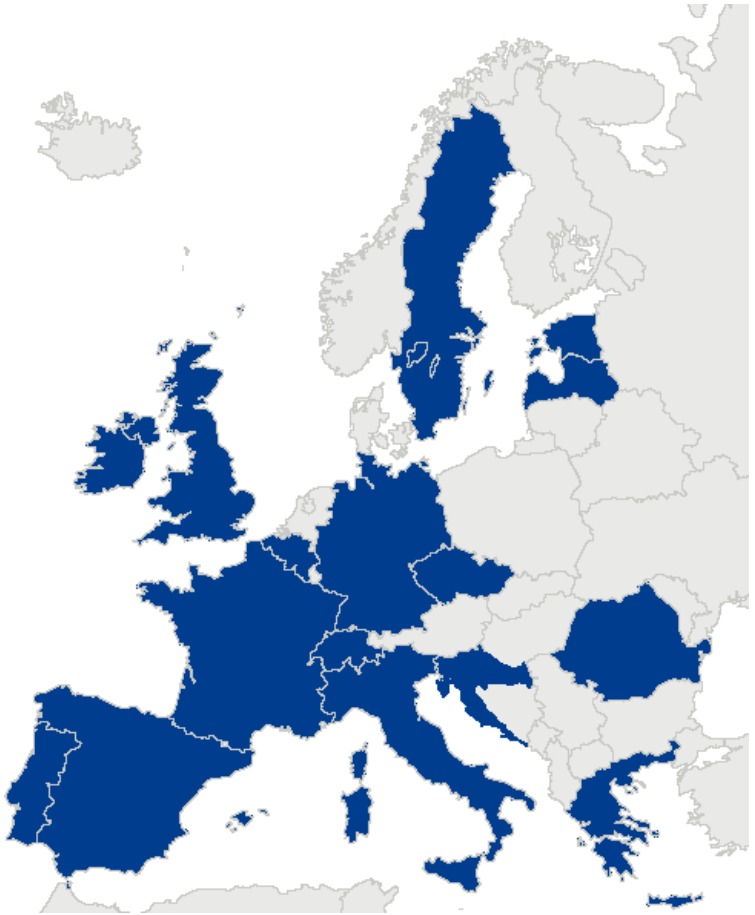
Thore Egeland (team leader),

Guro Dørum, Oskar Hansson,

Daniel Kling

EUROFORGEN-NoE is funded by the European Commission within the 7th Framework Programme

- … and >25 Satellite Training Workshops all over Europe!
  - Belgium 1 (+1)
  - Croatia 1
  - Czech Rep. 1 (+1)
  - Estonia (+1)
  - France 6 (+?)
  - Germany 5 (+2)
  - Greece 2
  - Ireland 1
  - Italy 1 (+1)
  - Latvia 1 (+1)
  - Portugal 2 (+1)
  - Romania 1
  - Slovenia 1 (+1)
  - Spain 1
  - Sweden 1
  - Switzerland 2 (+1)



EUROFORGEN-NoE is funded by the European Commission
within the 7th Framework Programme

# Consortium publications

# Consortium publications

- **Santos et al.: Forensic ancestry analysis with two capillary electrophoresis ancestry informative marker (AIM) panels: Results of a collaborative EDNAP exercise**
  - Forensic Sci Int Genet. 2015 Jun 15;19:56-67
- **Marcinska et al.: Evaluation of DNA Variants Associated with Androgenetic Alopecia and Their Potential to Predict Male Pattern Baldness**
  - PLoS One. 2015;10(5):e0127852
- **Eduardoff et al.: Inter-laboratory evaluation of SNP-based forensic identification by massively parallel sequencing using the Ion PGM™**
  - Forensic Sci Int Genet. 2015 Jul;17:110-21
- **Mapes et al.: Which traces to select for DNA analysis?**
  - Forensic Magazine Oct 2015 (in press)

EUROFORGEN-NoE is funded by the European Commission
within the 7th Framework Programme

# International dissemination conference "DNA in Forensics 2016" – Venue & Date


**23rd June 2016**

**21st - 24th June 2016**

There will be a special registration discount for EUROFORGEN participants

EUROFORGEN-NoE is funded by the European Commission within the 7th Framework Programme

# International dissemination conference "DNA in Forensics 2016" – Topics

- **FROM CRIME SCENE TO COURT ROOM**
  - Evidence challenges and advanced interpretation methods
  - The interpretation debate: miscarriages of justice
- **FROM GENOTYPE TO PHENOTYPE**
  - Eye or hair colour
  - Other externally visible characteristics
  - Biogeographic origin
- **FROM SCIENCE TO SOCIETY**
  - Ethical and legal aspects, the societal dimension of forensic genetics
  - Border security in Europe from a DNA perspective
  - Cost-benefits of DNA typing

EUROFORGEN-NoE is funded by the European Commission within the 7th Framework Programme

# The Short Term Fellowship Program

- **First Call 2013**
  - 14 fellowships awarded to 13 colleagues from 9 countries
  - Details on website

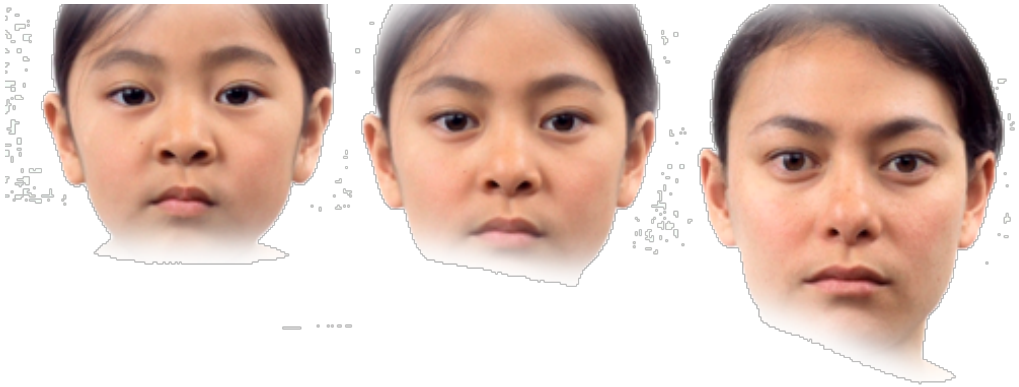
- **Second Call 2014-2015**
  - 20 new fellowships open
    - Laboratory visits for 3-5 days
    - Active participation in workshops related to EFG aims
    - Other research/training activities related to scope of WPs 2-5
  - Application details on the website
  - Travel support up to EUR 500


EUROFORGEN
Network of Excellence

find us on Facebook

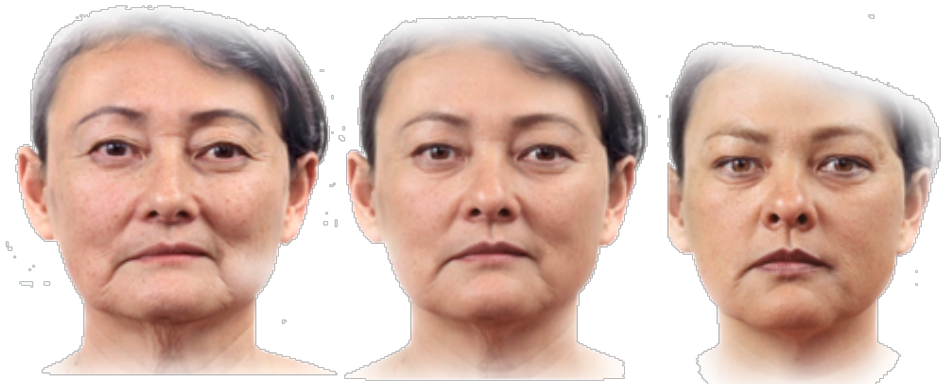
Please do not forget to join our Facebook group!
**… already 244 members!**

EDNAP meeting, 20 October 2015

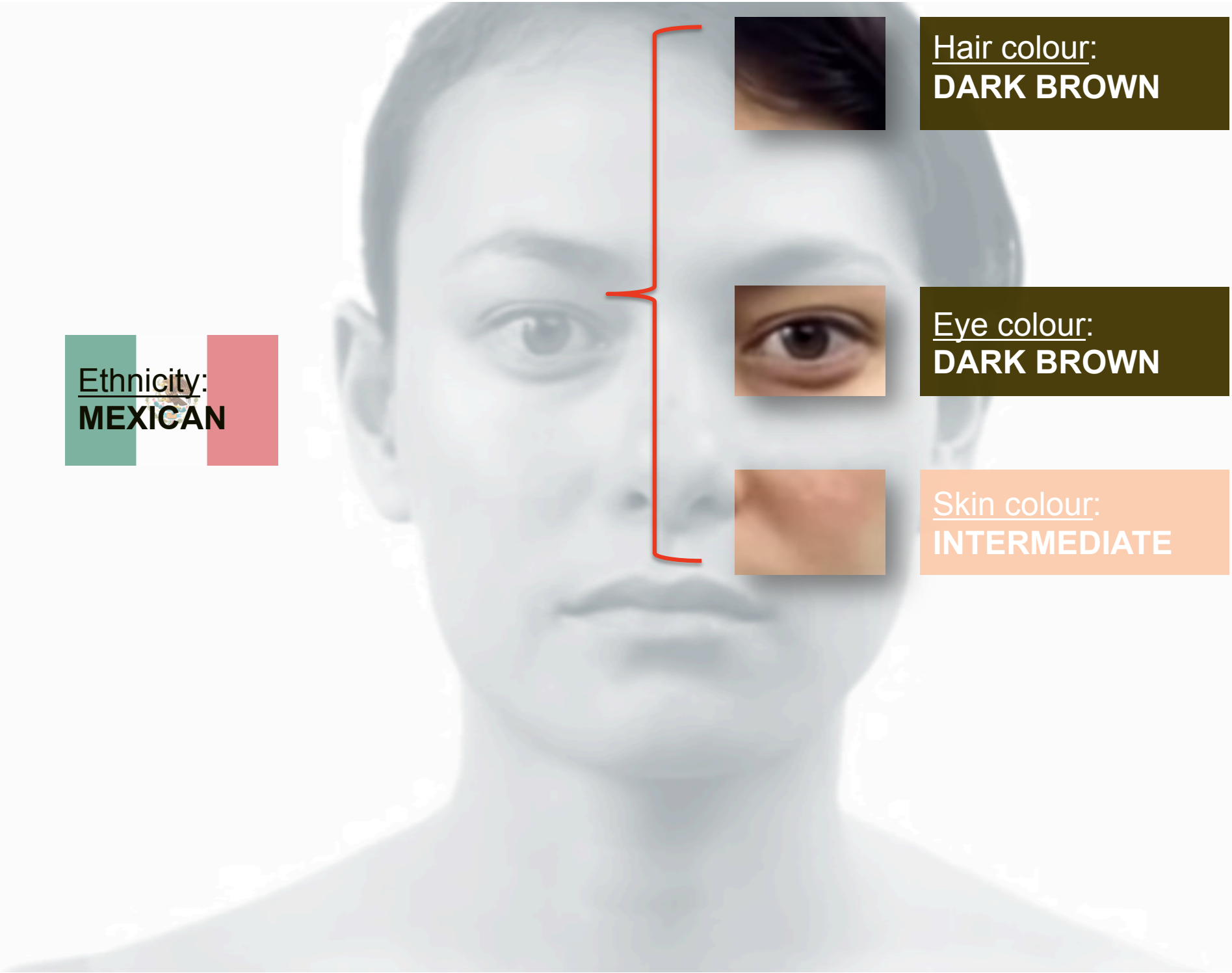


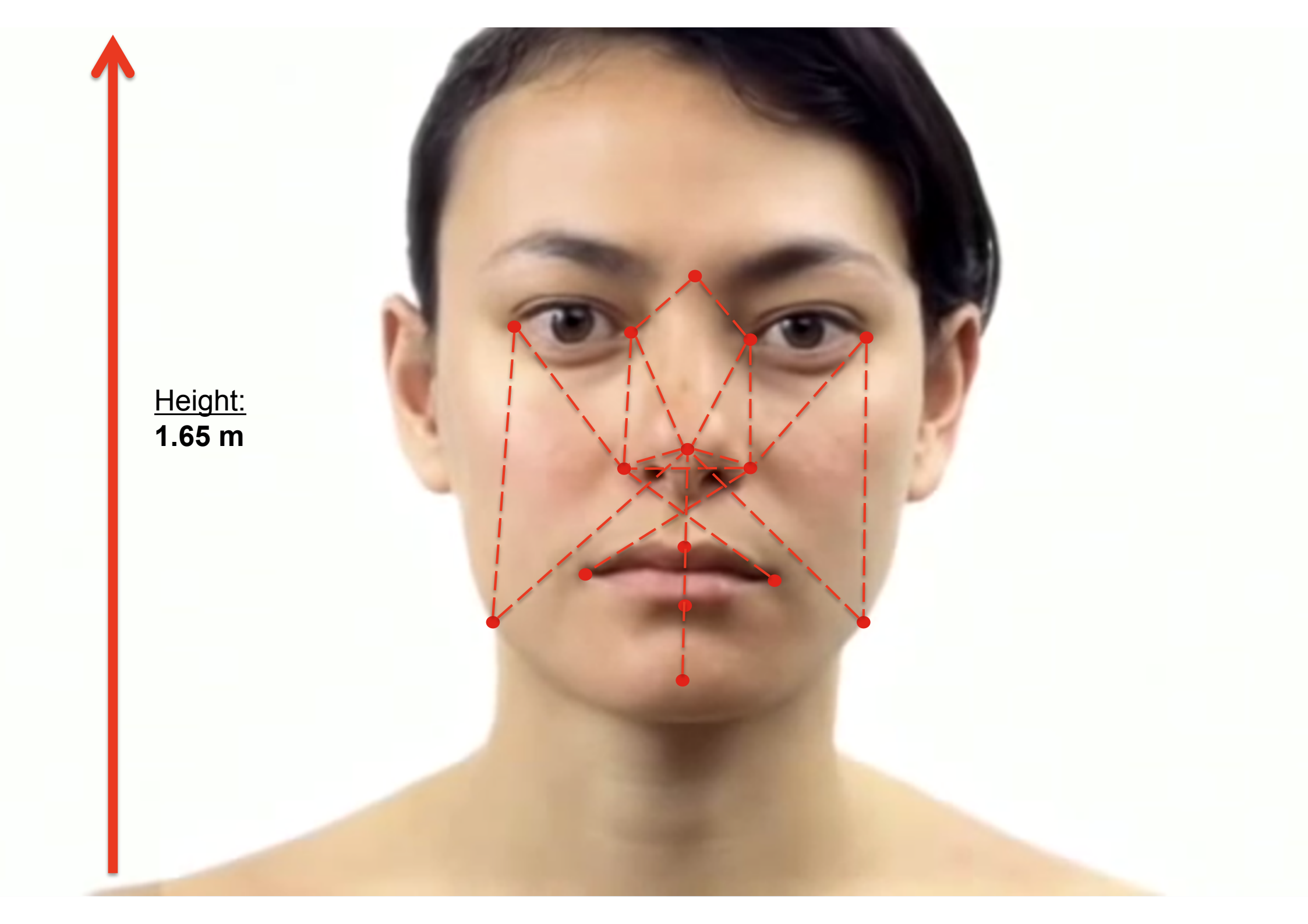
# DNA methylation-based NGS Method for Age Prediction

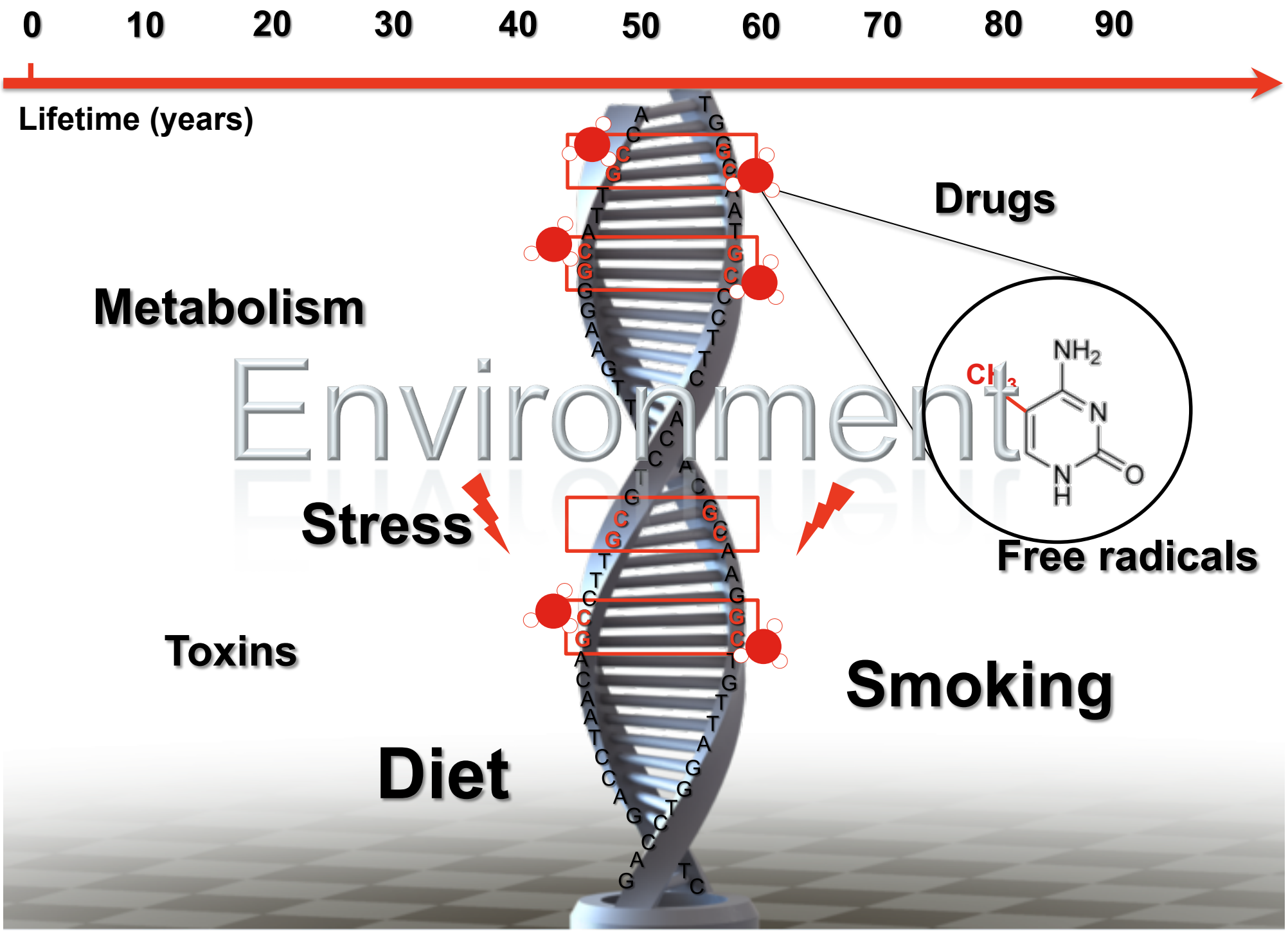
**A. Vidaki**, D. Ballard, D. Syndercombe Court



KING'S College LONDON

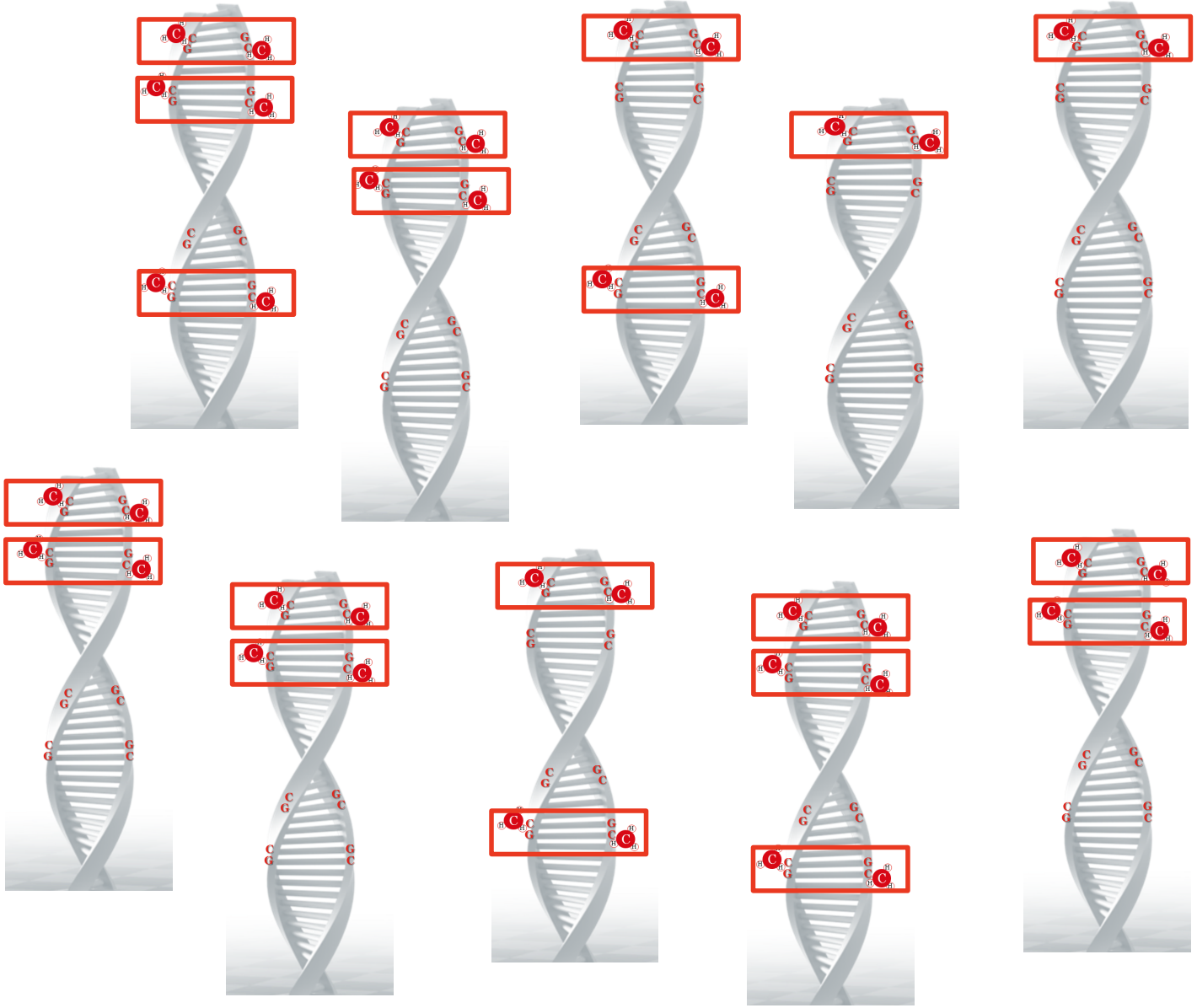

Hair colour:
DARK BROWN
Eye colour:
DARK BROWN
Skin colour:
INTERMEDIATE
Ethnicity:
MEXICAN


Height:
1.65 m


0
10
20
30
40
50
60
70
80
90
Lifetime (years)
Drugs
Metabolism
Environment
Stress
Free radicals
Toxins
Smoking
Diet

30

Lifetime (years)



**CpG1** – 100%

**CpG2** – 60%

**CpG3** – 0%

**CpG4** – 40%

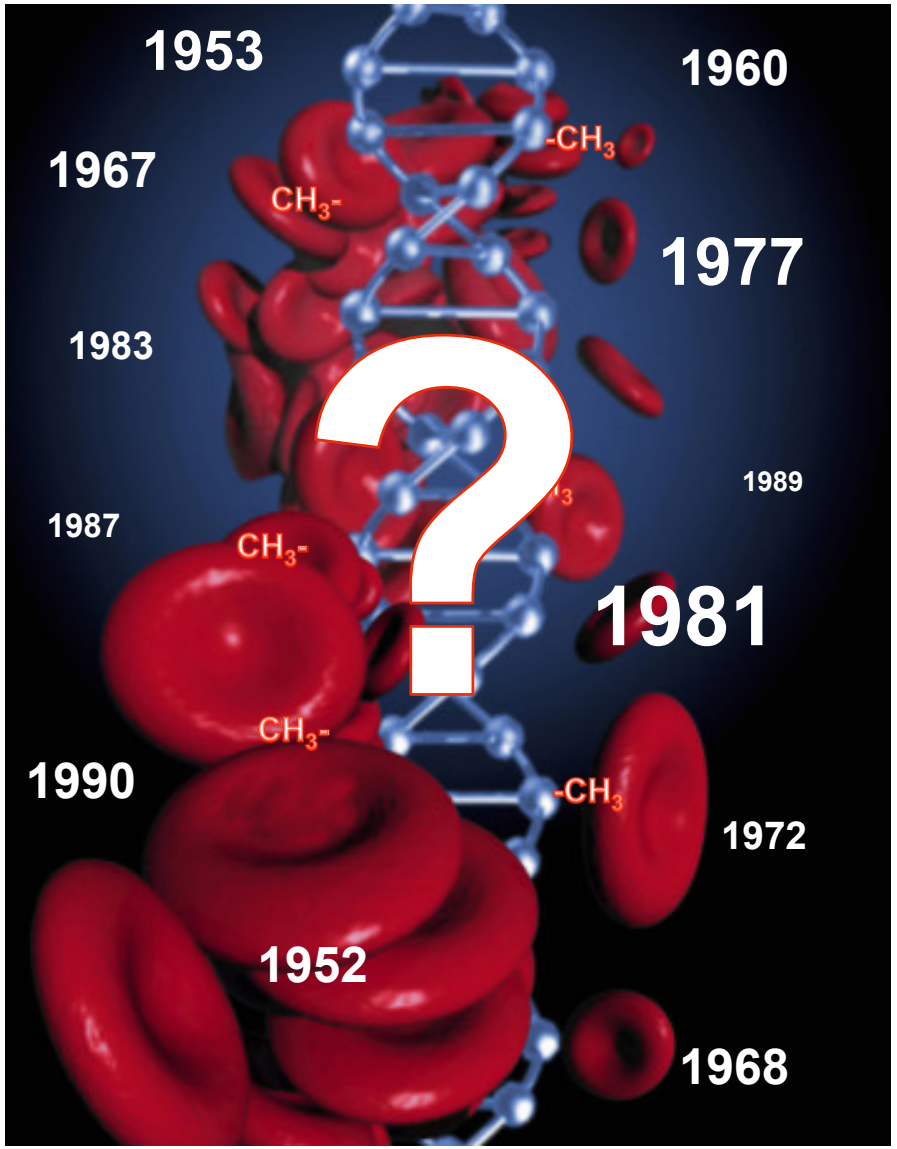
# Aim of the project

## Part 1

Age prediction in whole blood using a small set of age-associated differentially methylated CpG sites

## Part 2

Development of an NGS protocol for accurate and sensitive methylation quantification



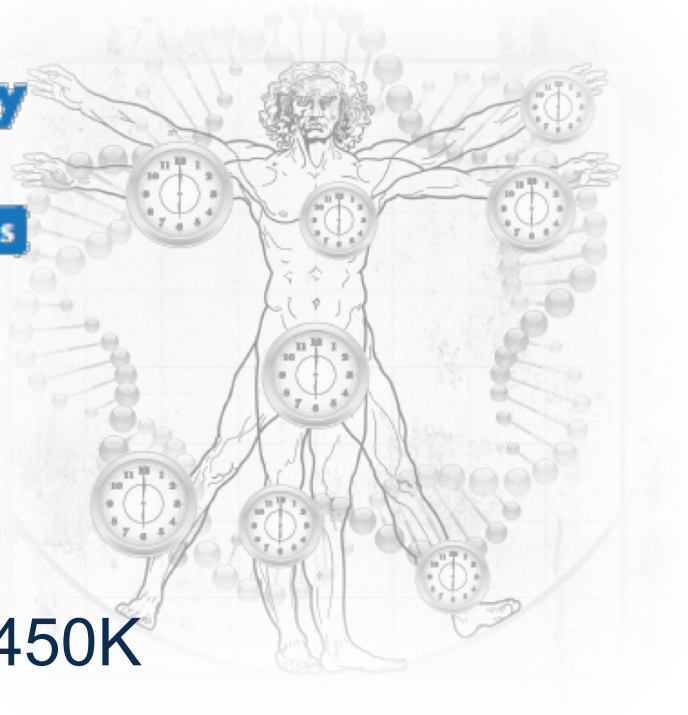
Horvath Genome Biology , 14:R115
http://genomebiology.com/2013/14/10/R115

Genome Biology

RESEARCH Open Access

# DNA methylation age of human tissues and cell types

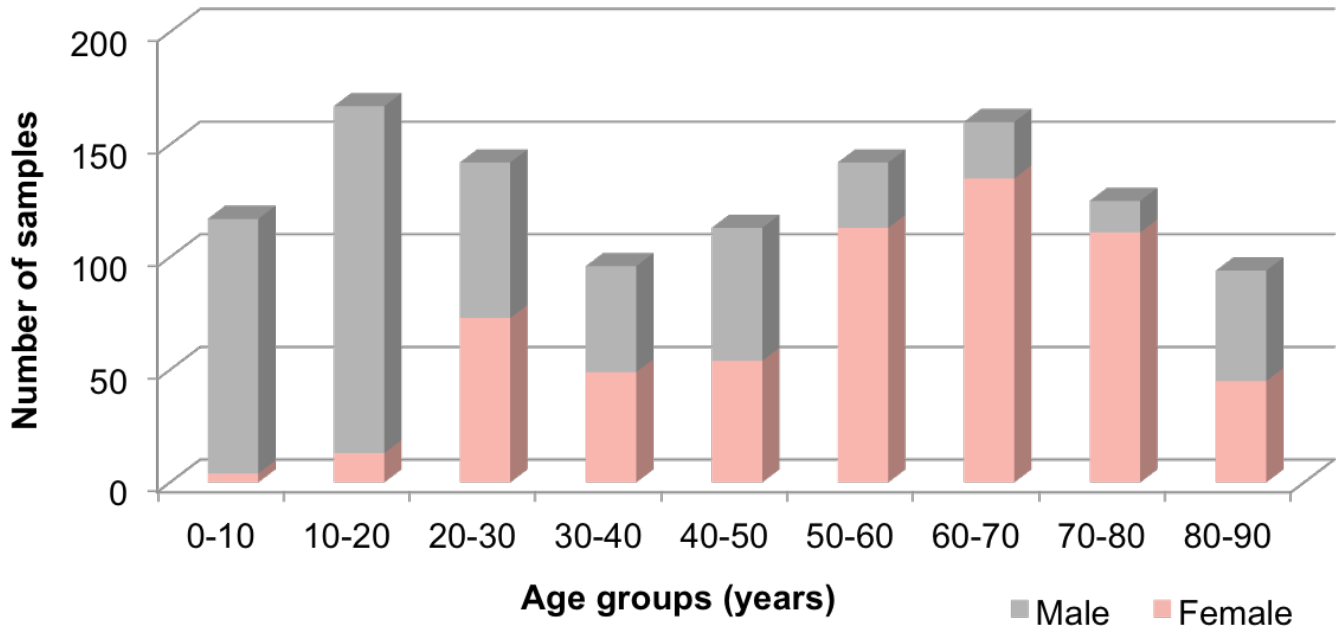
Steve Horvath[1,2]



- ~8,000 samples analysed by Illumina 27K/450K
- Multi-tissue age predictor
- 353 ‘clock’ CpG sites
- Whole blood → median error = 3.7 years

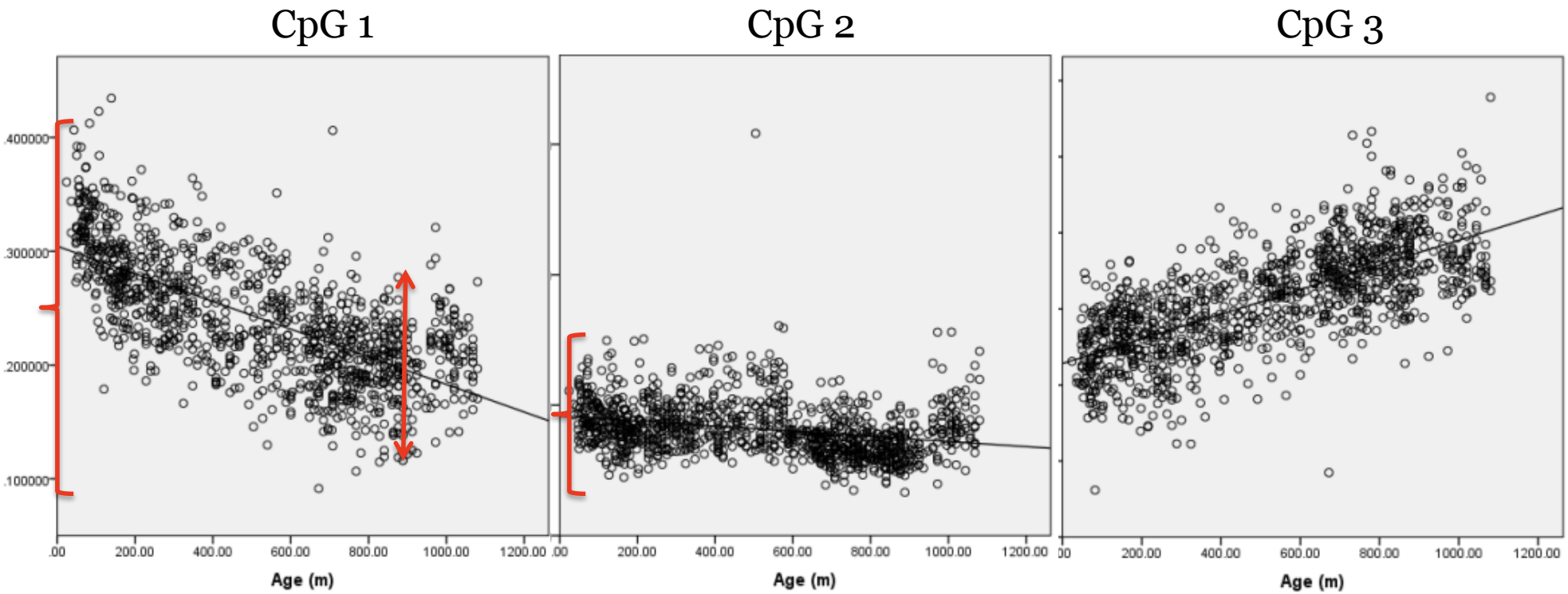
**45 CpG sites showing the strongest negative or positive age coefficient ranging from -1.719 to 3.067**

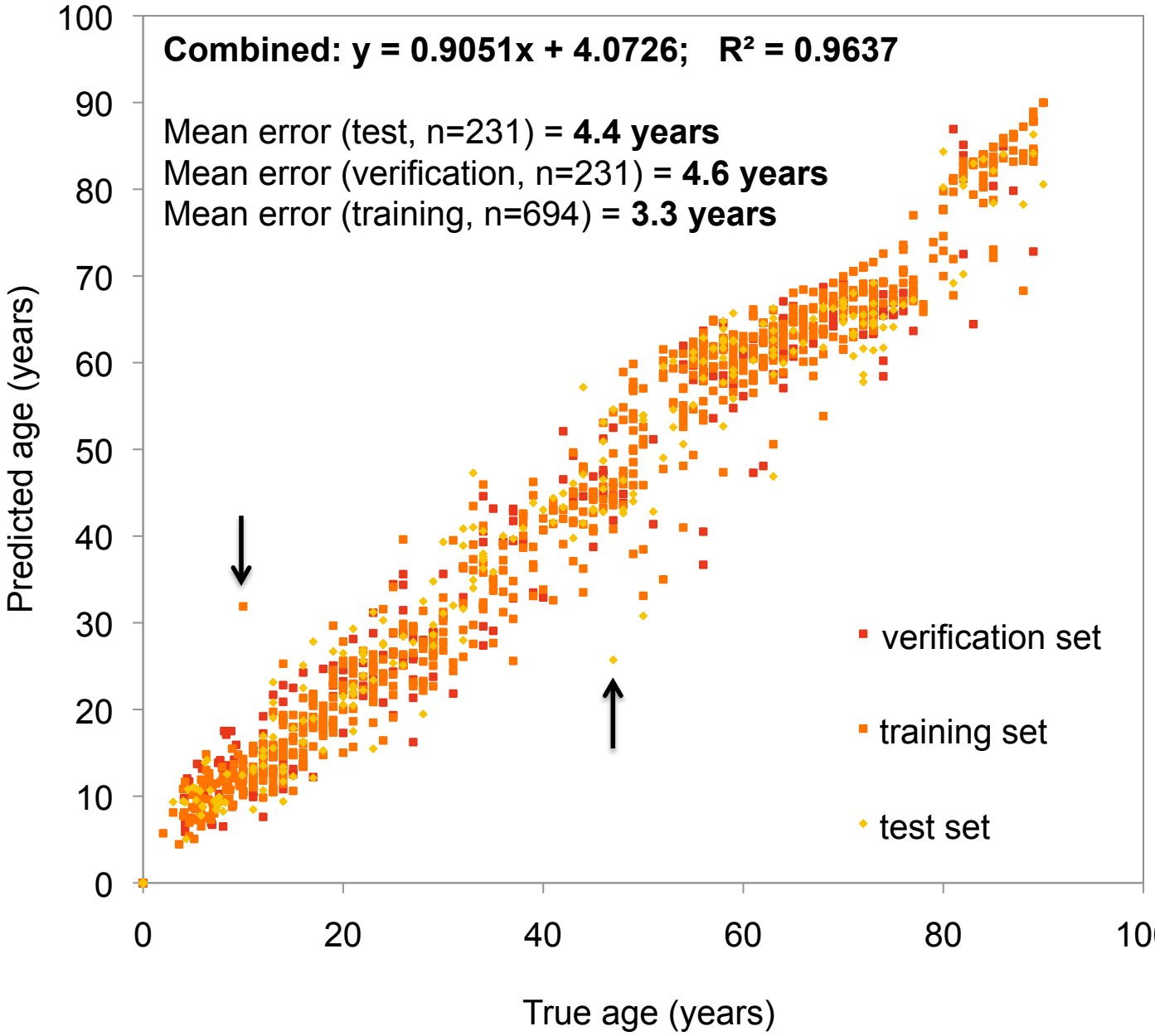
| Tissue | Set | Samples | ♀/♂ | Age range (mean) (years) | Platform |
|---|---|---|---|---|---|
| Whole blood | 1 | 235 | 0/235 | 4-18 (10) | 27K |
| | 2 | 24 | 12/12 | 2-35 (14) | 27K |
| | 3 | 170 | 77/93 | 32-90 (65) | 450K |
| | 4 | 385 | 187/198 | 16-88 (40) | 27K |
| | 5 | 33 | 12/21 | 18-65 (29) | 450K |
| | 6 | 91 | 91/0 | 49-74 (63) | 27K |
| | 7 | 218 | 218/0 | 52-78 (65) | 27K |
| | | **1,156** | **597/559** | **2-90 (44)** | |



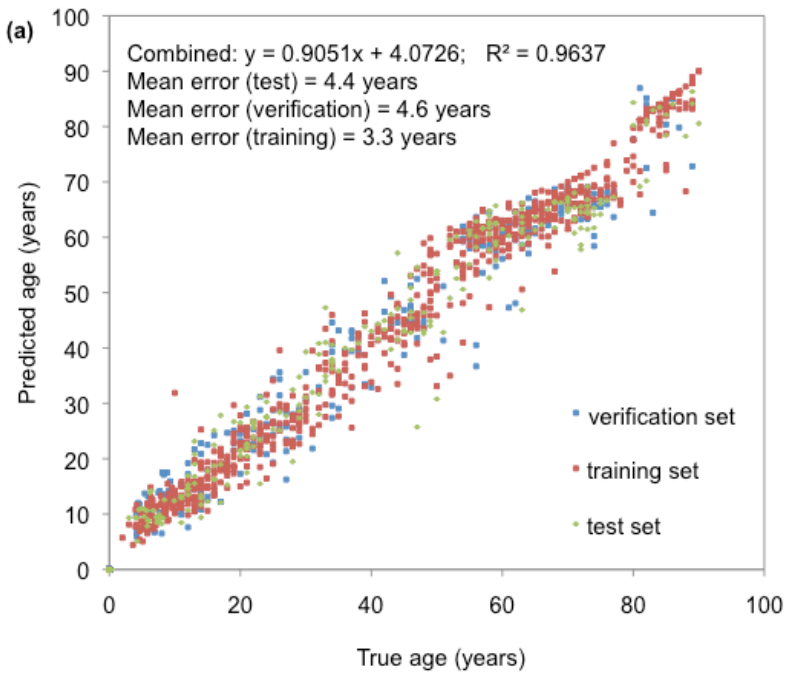
# Methylation patterns

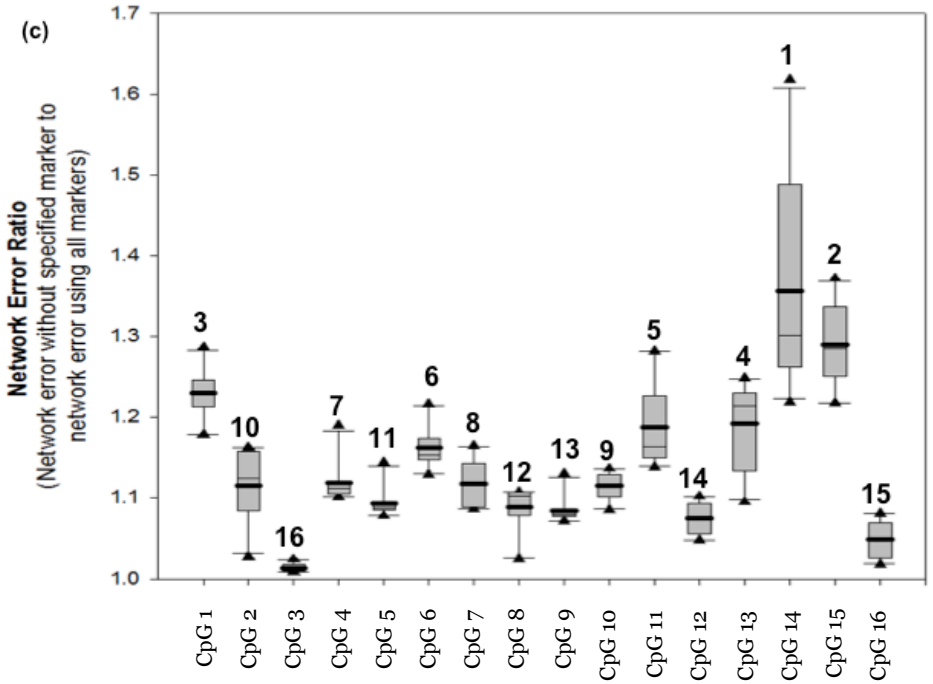
**24 out of 45 CpGs**
**p<0.05**

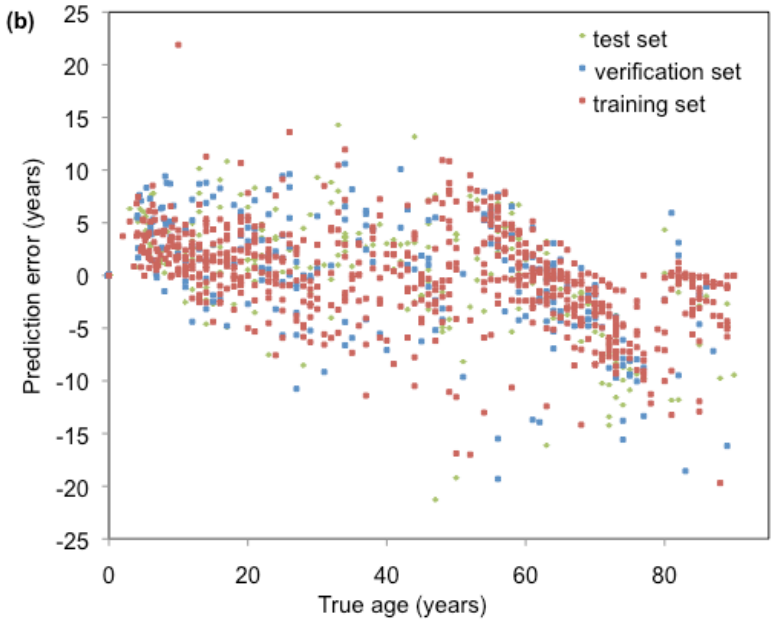


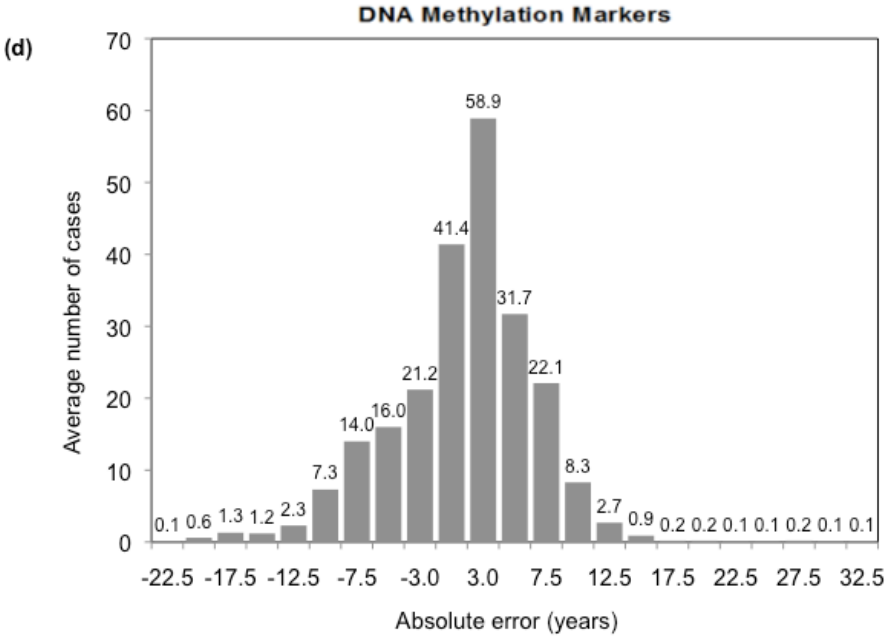



| Samples (n=1,156) | Error (years) |
|---|---|
| 95.6% | ±10 |
| 82.8% | ±5 |

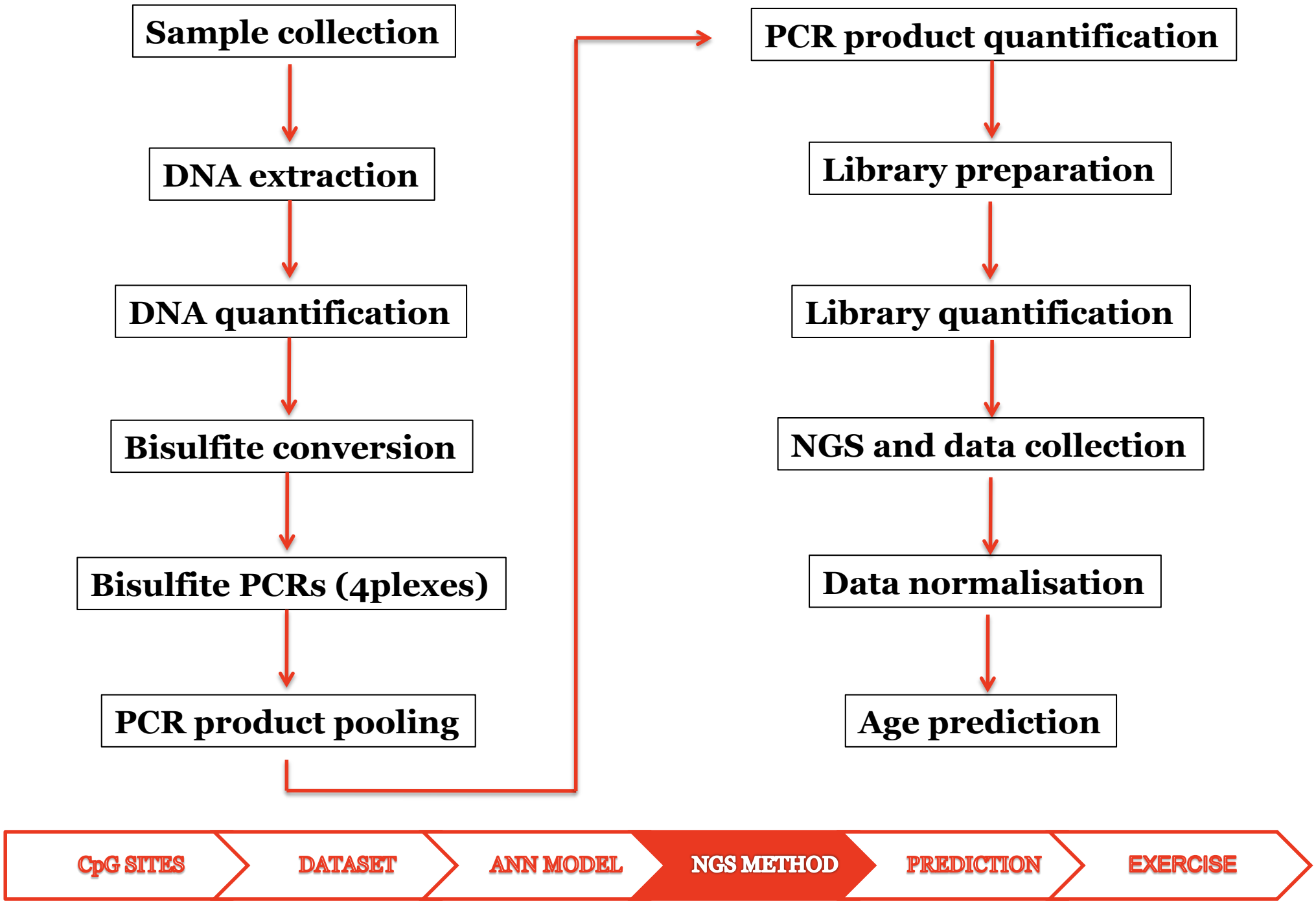


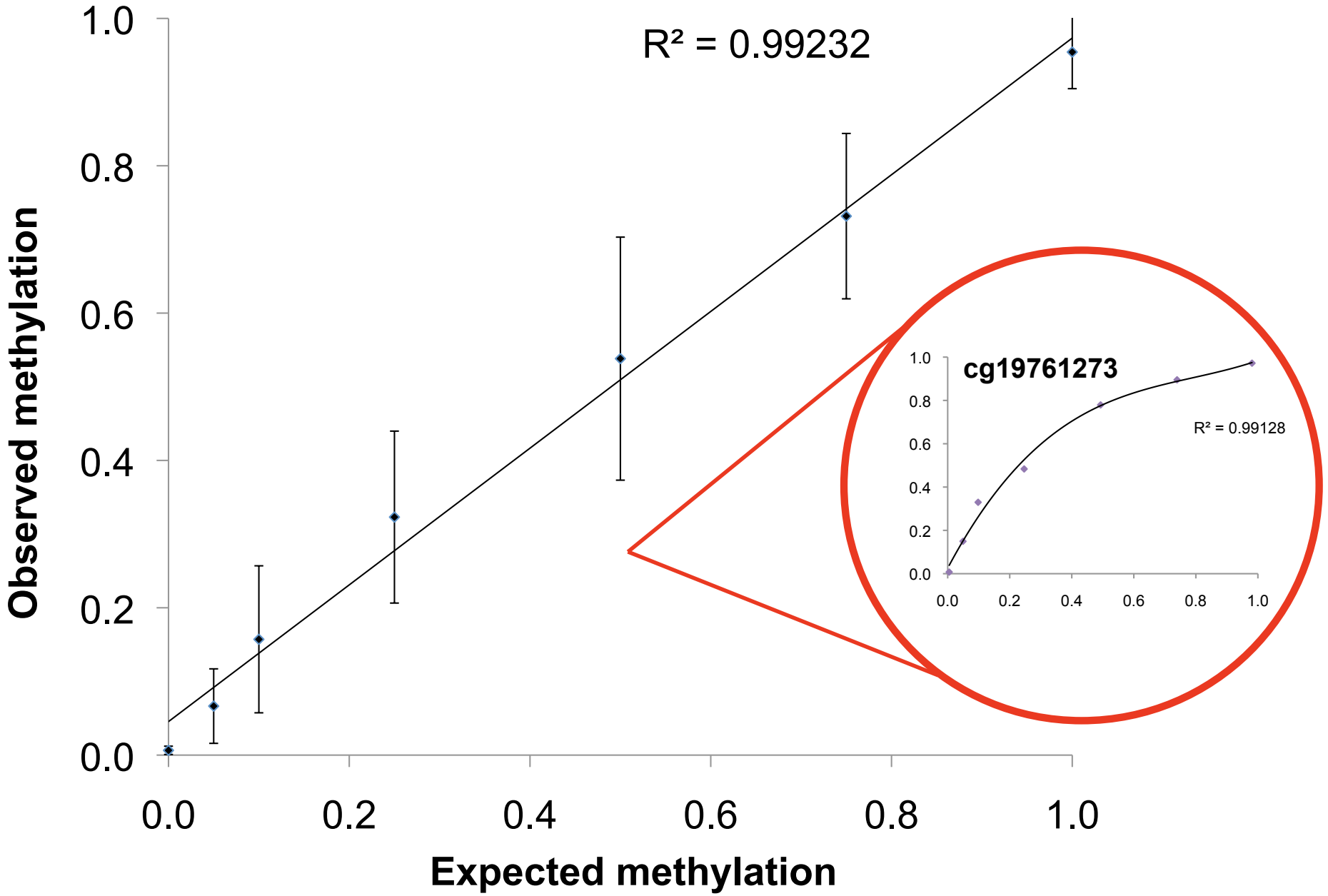



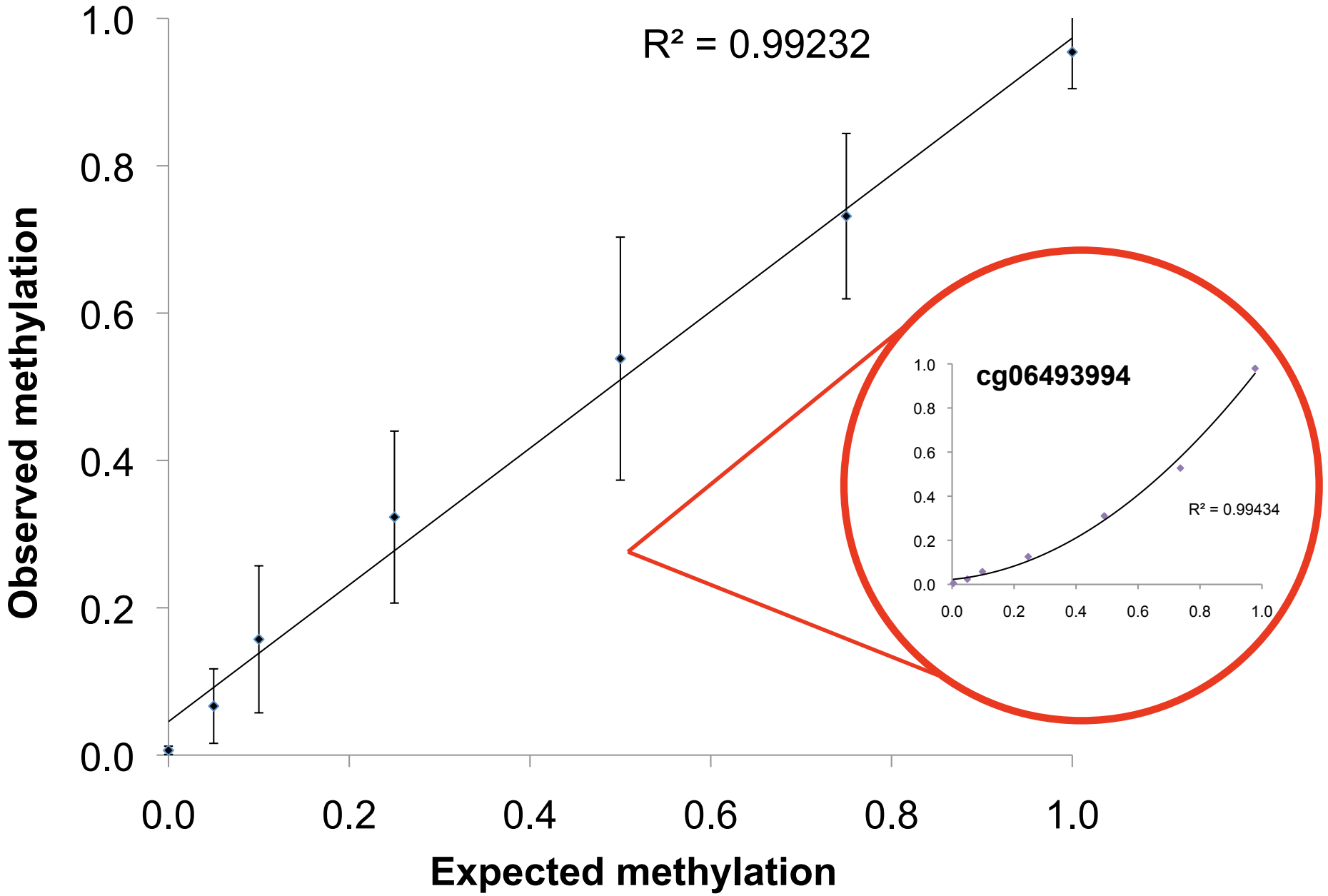



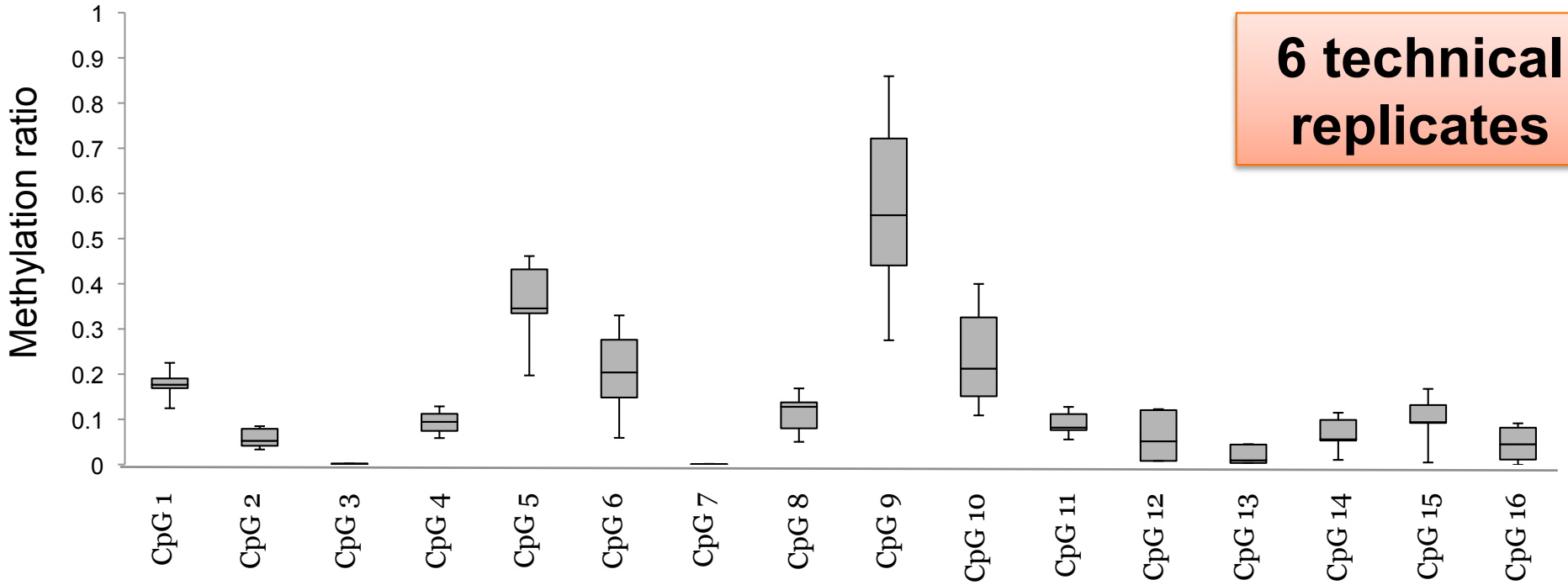




CpG SITES | DATASET | ANN MODEL | NGS METHOD | PREDICTION | EXERCISE


Sample collection
DNA extraction
DNA quantification
Bisulfite conversion
Bisulfite PCRs (4plexes)
PCR product pooling
PCR product quantification
Library preparation
Library quantification
NGS and data collection
Data normalisation
Age prediction
CpG SITES
DATASET
ANN MODEL
NGS METHOD
PREDICTION
EXERCISE


R² = 0.99232
Observed methylation
Expected methylation
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
cg19761273
R² = 0.99128


R² = 0.99232
Observed methylation
Expected methylation
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
cg06493994
R² = 0.99434



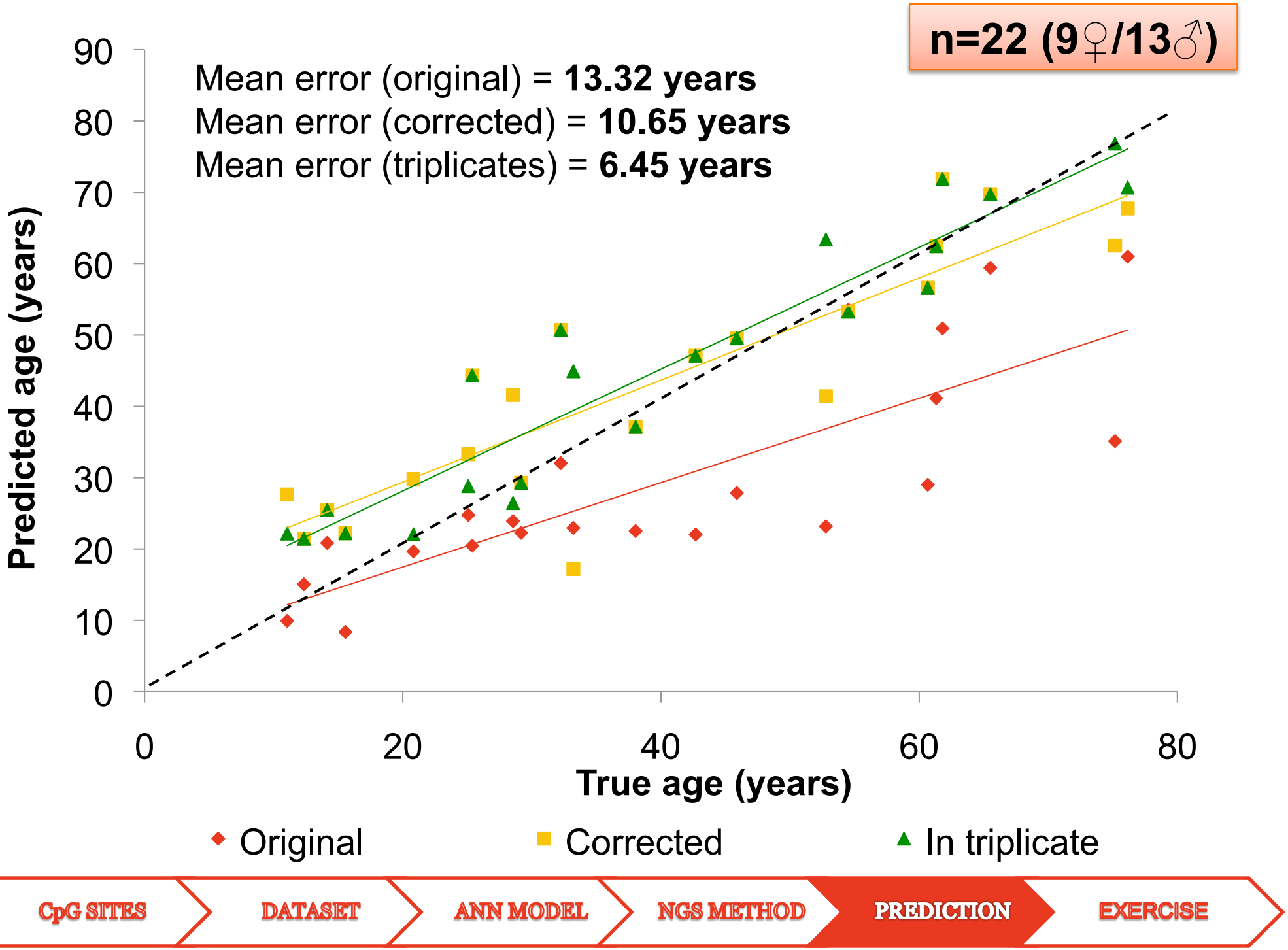

**6 technical replicates**

**True age**
**54.51**

**Predictions**

45.32
72.38
31.64
53.28
77.17
57.48

Mean prediction - 56.21
Mean error - 1.7 years


n=22 (9♀/13♂)
Mean error (original) = 13.32 years
Mean error (corrected) = 10.65 years
Mean error (triplicates) = 6.45 years
Predicted age (years)
True age (years)
0
10
20
30
40
50
60
70
80
90
0
20
40
60
80
Original
Corrected
In triplicate
CpG SITES
DATASET
ANN MODEL
NGS METHOD
PREDICTION
EXERCISE

# Current work

- ✓ Potential to improve the method's prediction by introducing different data normalisation strategies both on the genome-wide AND MiSeq data
- ✓ Testing different multiplex PCR combinations to further improve the method's sensitivity and the hands-on time
- ✓ Analysing more blood samples and mock casework stains
- ✓ Manuscript in preparation

# EDNAP Exercise Proposal

## Samples to be analysed:

5 blood samples (in triplicate)

7 DNA methylation standards (0, 5, 10, 25, 50, 75, 100%)

*Additional samples (optional)*

## KCL to provide:

DNA samples

PCR primers

Protocols

## Participating labs to provide:

Reagents

# EDNAP Exercise Proposal

## Potential dates:

*Part 1*

Samples to be sent out (controls) - December 2015
Data to be collected - March 2016
Presentation of results - April 2016

*Part 2*

Samples to be sent out (blood samples) - May 2015
Data to be collected - August 2016
Presentation of results - October 2016
Manuscript preparation/dissemination - End of 2016

## Points for discussion:

Type of data - FASTQ files or methylation values
Other PCR-based methods? PGM instrument?

# Interested in joining the exercise?

KING'S College LONDON

Please contact:

**Denise Syndercombe Court**
denise.syndercombe-court@kcl.ac.uk

**Athina Vidaki**
athina.a.vidaki@kcl.ac.uk