Forensic Age Estimation with DNA Methylation

Combining the power and performance of next-generation sequencing with epigenetic insights to estimate chronological age

Highlights

- ***A screening tool for complex identifications***
  Knowing a person's age from analysis of remains or biological traces can guide identification.
- ***Accurate models for age prediction***
  DNA methylation sites are the most informative marker for aging, making methylation analysis the best way to predict age.
- ***Highly advantageous, cutting-edge technology***
  Sequencing is an accurate, sensitive method for analyzing age-associated CpG sites with minimal input DNA.

Introduction

Methylation sequencing shows great promise as a method of forensic age estimation. Over the past decade, researchers have identified and refined estimation markers with the goal of implementing an age-prediction model that translates age-associated DNA methylation patterns into chronological age. The challenge is developing a sufficiently sensitive and accurate method that can estimate age across a spectrum of human tissue types and samples of varying quality. With demonstrated robustness and exclusive multiplexing capability, next-generation sequencing (NGS) technology offers the capacity to produce the targeted data needed for forensic casework and provide access to DNA intelligence that can enhance physical descriptions of unidentified individuals and generate investigative leads (1–3).

Efforts to develop and validate NGS-based methods that quantify the methylation status of selected CpG sites are well underway, with the MiSeq FGx® Sequencing System uniquely positioned to empower this capability. This application note describes how DNA methylation correlates to human age and highlights research from a team at King's College London (KCL). The KCL team has made significant contributions to the understanding of methylation-based protocols and age estimation models, helping forensic laboratories better understand the value of NGS to estimate age and other phenotypic traits (1).

Accurate phenotypic trait prediction from DNA samples

Short tandem repeats (STRs) are the foundation of forensic databases. The databases provide investigative leads by linking DNA profiles from crime scenes or unidentified remains to DNA profiles in the database. To generate leads when database results are inconclusive, the forensic community is evaluating a variety of approaches, including phenotyping. Phenotyping analyzes single nucleotide polymorphisms (SNPs) to deduce phenotypic traits or externally visible characteristics (EVCs), such as eye color, hair color, and biogeographical ancestry (BGA), from samples of unknown origin.

QIAGEN already facilitates analysis of well-curated phenotyping markers with an end-to-end workflow of human identification tools, including the ForenSeq® Signature Plus Kit. The kit includes an extended complement of markers targeting 22 phenotype-informative SNPs (piSNPs) and 56 biogeographical ancestry-informative SNPs (aiSNPs). Results are produced from picogram amounts of DNA and visualized in Universal Analysis Software (UAS) after sequencing on the MiSeq FGx System (Figure 1).

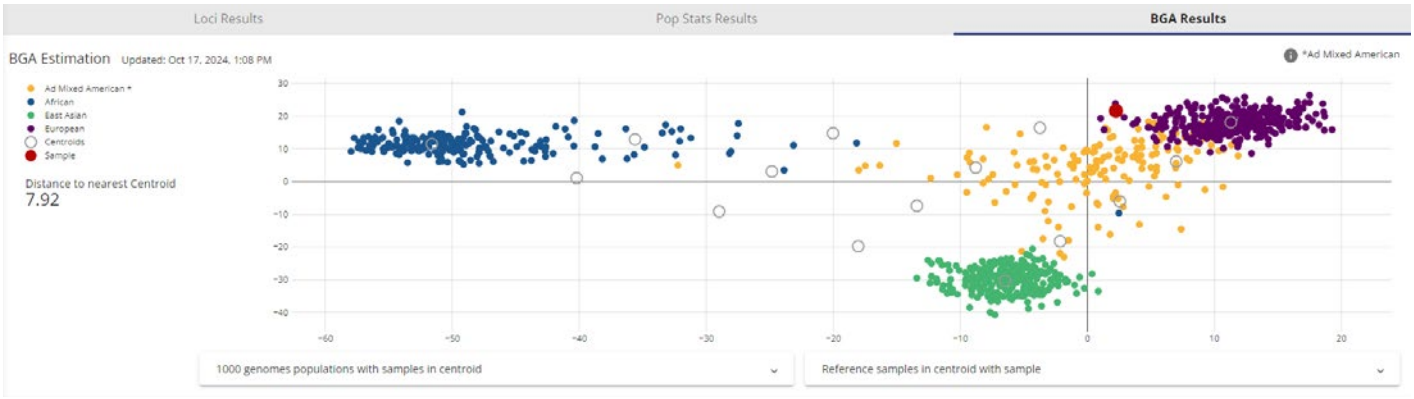

Figure 1.
Universal Analysis Software automatically analyzes sequencing data from the MiSeq FGx System and displays biogeographical ancestry to help open new avenues of investigation. The highly extensible system combines proven data quality with ease of use.

Multiple studies are under way to identify and evaluate an expanded range of EVC markers that can enhance investigative leads. Several teams are focusing on methylation sites in the human genome that strongly correlate with age and establishing a variety of effective age-prediction models, bringing the field of DNA intelligence closer to a universal model. Like other phenotypic traits, age estimation offers clear advantages for narrowing a suspect pool or defining select groups of interest. Age is a clear feature of human appearance that is difficult to disguise. Anyone can, for example, dye brown hair blonde, but concealing wrinkles or diminished muscle tone is significantly more challenging. Moreover, pinpointing an age boosts the estimation of age-related EVCs, such as male pattern baldness. In missing persons cases or disaster victim identification (DVI), estimating age from DNA can be a valuable screening tool that accelerates procedures or provides supporting data (2, 4, 5).

Exploring epigenetics and human aging

Initial approaches to human age estimation often included morphological inspection based on the examination of bones and teeth. While still informative for investigating human remains and making age-at-death estimations, this method is limited to solid tissues. This limitation also extends to traditional chemical methods, such as aspartic acid racemization and radiocarbon analysis. Although precise, these methods require dental specimens. Another constraint of morphological inspection is applicability to elderly subjects: after 65 years of age, prediction accuracy drops precipitously. Since 2010, researchers have been developing and advancing DNA-based tests to infer age from a biological trace, marking a shift in focus from DNA characteristics and proteins to epigenetic changes. This shift has led to the identification of age-correlated markers that in turn enable age-prediction models for forensic applications (2, 4).

Among the many features of epigenetic changes, DNA methylation is now firmly established as the most informative marker for aging and the optimum method to predict chronological age. Changes in patterns of epigenetic marks are considered the primary events that impact human aging, with different marks or signatures shaping the genome and undergoing dynamic changes to modulate gene expression. The result of continuous interactions between genes and the environment, these epigenetic changes exert significant influence on the aging process. The changes are cumulative and measurable, and therefore age-informative (2).

Forensic implementation and outlook

After establishing that the epigenetic signature of DNA methylation gradually changes over a lifespan, research highlighting DNA methylation changes correlating to age emerged and then grew to encompass methylation tests specific to forensic casework. Also noteworthy is the discovery of bisulfite conversion in the 1990s, heralding a breakthrough for DNA methylation. Almost all forensically relevant tests, from the discovery of age estimation markers to the implementation of tests, rely on bisulfite conversion (2).

Discovery of age-associated CpG sites

As a cornerstone technology, bisulfite conversion led to whole-genome bisulfite sequencing (WGBS) and the development of methylation microarrays. Both techniques use bisulfite-converted DNA and single-base resolution to enable genome-wide analysis of DNA methylation patterns. WGBS and microarrays cover many CpG sites throughout the genome and have produced large volumes of highly concordant data that continue to propel DNA methylation studies at the discovery phase. Microarrays are considered the best tool for finding relevant CpG sites. Neither technique is appropriate for forensic samples and validation and implementation require a more targeted approach (1, 2).

Implementing forensic markers for age estimation requires techniques compatible with low-level or degraded DNA from blood, semen, saliva, hair and other sample sources typical of forensic casework. Although WGBS has proven a useful NGS method for discoveries, it is too broad to apply to forensic casework. Pyrosequencing also held some promise and was once the prevailing technique for age-prediction models but is ultimately limited by lack of multiplexing. When analyzing samples with low-level DNA, multiplexing is desirable for its ability to target multiple markers in different genomic regions. More recent NGS methods support multiplexing while providing a targeted approach with minimal input requirements (2).

Targeted sequencing scaled for forensic casework

The MiSeq FGx System, an NGS platform designed for forensic applications, is gaining traction as the preferred solution for forensic age estimation. High sensitivity, the ability to multiplex small amplicons, and the established accuracy of Illumina® sequencing-by-synthesis (SBS) technology make the system well-suited to successful sequencing of degraded and otherwise low-quality samples. Dual sequencing modes support both workflows and a wide range of third-party and research chemistries, allowing laboratories to perform validated protocols such as STR and SNP analysis with the ForenSeq Signature Plus Kit alongside research-based applications – including methylation-based age estimation – on the same platform (1, 2).

An intuitive touch-screen interface and flexible protocols complete a simple but powerful solution for developing methods to estimate age and other EVCs. Fulfilling this potential, initial approaches to age estimation using MiSeq-based systems have already been completed, with further optimization and consensus the only remaining barriers to full forensic implementation (5).

King's College London

Materials and methods

Using publicly available microarray data from over 4000 individuals, KCL evaluated methylation markers previously correlated with chronological age. This statistical assessment highlighted 11 CpG sites with the highest potential for DNA methylation-based age estimation from forensic samples. The CpG markers were associated with 10 different genes, including ELOVL2 and FHL2.

After marker discovery, KCL developed a bisulfite sequencing assay for the marker set using the MiSeq FGx System. The assay design resembles a previously described the DNA methylation-based age estimation method (5). Based on bisulfite conversion, the assay starts with 50 ng of input DNA. PCR amplification of small amplicons with an average length of ~140 bp is followed by library prep and then sequencing on the MiSeq FGx System in Research Use Only mode with the Illumina MiSeq Reagent Kit v2 (300 cycles) (5).

Results

The proposed assay was employed for the analysis of 112 whole blood samples and the resulting data were applied to the training (n=77) and testing (n=35) of a support-vector machine model with polynomial function (SVMp). The mean absolute prediction error was calculated at 3.6 years for the training set and 3.3 years for the test set. More than 71% of the samples in the test set predicted age with an absolute error of fewer than 4 years, and 89% predicted age with an absolute error of fewer than 7 years (Figure 2). KCL observed similar accuracy for an independent set of 88 DNA extracts from whole blood samples obtained as part of a collaboration with the University of Santiago

de Compostela (USC) in Spain. This set averaged an error of ±3.8 years, closely matching the expected prediction accuracy based on results from the original training and test sets (Figure 3). The error for individuals under age 54 – an age range more relevant to forensics – was ±2.6 years with 81% of the samples predicting age with an absolute error of fewer than 4 years (5).

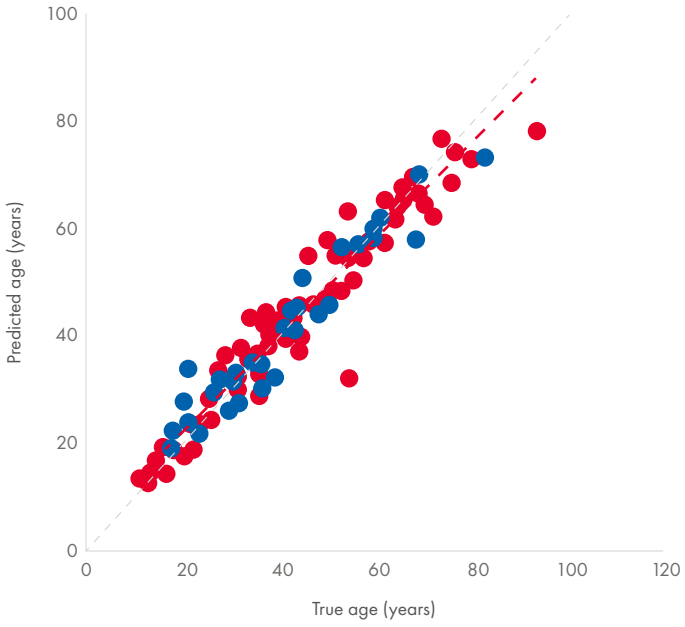

Figure 2.
Comparison between the predicted and true age for the training set (red, n=77) and blind test set (blue, n=35) in the SVMp model. The mean absolute prediction error was calculated at 3.6 and 3.3 years, respectively. The equation of the linear trendline fitting the training set (red dashed line) is visible on the graph. The gray dotted line represents "perfect" predictions where predicted and true age overlap (x=y).

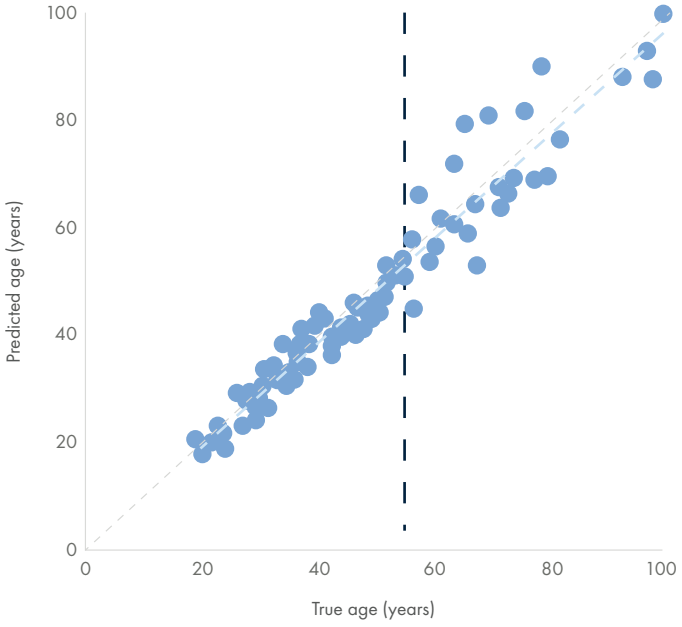

Figure 3.
Comparison between the predicted and true age for the USC sample set (blue, n=88) in the SVMp model trained on the KCL dataset (n=112). The mean absolute prediction error was calculated at 3.8 years. The equation of the linear trendline fitting the training set (blue dashed line) is visible on the graph. The grey dotted line represents "perfect" predictions where predicted and true age overlap (y=x). A vertical blue line represents the 54 years mark.

Furthermore, a sensitivity assessment revealed that age estimation accuracy was successfully retained down to 5 ng DNA input, with the prediction errors remaining practically identical for the 25 (±4.3 years), 10 (±4.3 years), and 5 ng (±3.9 years) inputs (Figure 4).

When accounting for template loss during bisulfite conversion (~52% recovery) and elution and reaction volumes, the inputs respectively translate to approximately 5, 2, and 1 ng in the PCR stage.
An independent evaluation of methylation markers for age estimation in saliva showed significant overlap between the most informative markers for both blood and saliva. Three out of 11 markers from the blood-based model were among the top 20 performing markers in saliva, suggesting high potential for transferability of this method between the two tissues (6).

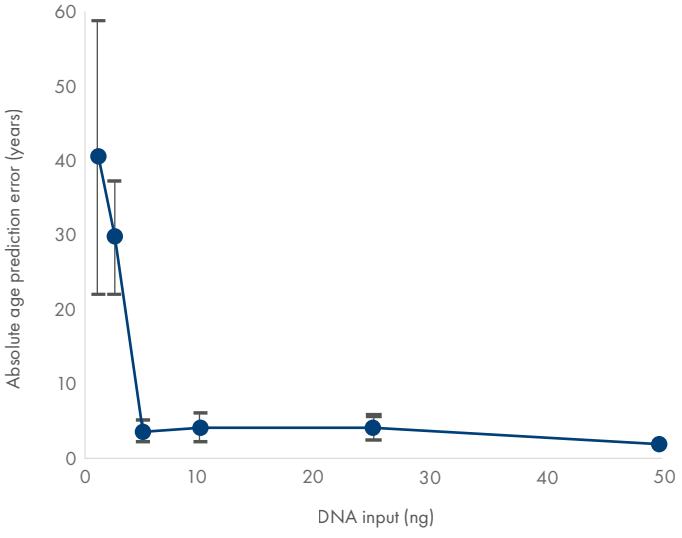

Figure 4.
Average absolute error in age prediction observed in a set of samples (n=6) analyzed at different DNA input amounts: 50, 25, 10, 5, 2.5, and 1 ng. Error bars represent the standard error of the prediction error between the six samples.

The age estimation assay developed on the MiSeq FGx System estimated the chronological age of a donor with an average error of approximately ±3 years for blood samples, an accuracy maintained as low as 5 ng DNA input. This is the template amount often recovered from medium and large blood stains associated with forensic investigations. Additionally, the small amplicon sizes are promising for the assessment of degraded blood samples while the significant marker overlap suggest important possibilities for assay use with multiple tissue types.

Conclusion

Forensic age estimation using DNA methylation analysis is a fast-expanding field with great potential for human identification. The correlation of methylation with chronological age has proven a key step in shoring up better, more robust estimation accuracy in forensic investigations. Achieving precise results requires targeted NGS methods to multiplex forensically relevant markers and tolerate the low input DNA amounts typical of forensic samples.

Although the methods are effective, traditional phenotyping requires large amounts of input DNA and relies heavily on complex bioinformatics, rendering them incompatible with most forensic genomic applications. To overcome this challenge, QIAGEN is empowering the development of myriad validated and research-driven but operationally relevant methods. This capability makes QIAGEN NGS technology uniquely suited to the widest range of applications and the only portfolio capable of supporting the full phenotyping suite in the forensic laboratory (4).

Ordering Information

Product	Contents	Cat. no.
ForenSeq Signature Plus Kit (96)	Includes all the required reagents to prepare sequencing libraries from forensic DNA samples; part of the MiSeq FGx Forensic Genomics Solution.	V16000213
ForenSeq Signature Plus Kit (384)	Includes all the required reagents to prepare sequencing libraries from forensic DNA samples; part of the MiSeq FGx Forensic Genomics Solution.	V16000214
MiSeq FGx Sequencing System	Desktop instrument with two run modes for a range of forensic genomics applications within a validated NGS workflow.	15048976
ForenSeq Analysis Software Server, Monitor	Software pre-installed as a dedicated server specific for forensic genomics for run setup, sample management, analysis and report generation. This product includes server, mouse, keyboard and monitor.	9003364
MiSeq FGx Reagent Kit	Supports up to 12.5 million paired-end reads for deep sequencing or high throughput sample processing.	15066817
MiSeq FGx Reagent Micro Kit	Supports up to 5 million paired-end reads for small batch sizes and faster turnaround times.	20021681

Learn more about NGS for HID in your lab. Visit **qiagen.com/SignaturePlus**

References:

1. Vidaki A, Ballard D, Aliferi A, Miller TH, Barron LP, Syndercombe Court D. DNA methylation-based forensic age prediction using artificial neural networks and next generation sequencing. *Forensic Sci Int Genet.* 2017; 28: 225–236. https://doi.org/10.1016/j.fsigen.2017.02.009
2. Freire-Aradas A, Phillips C, Lareu MV. Forensic individual age estimation with DNA: From initial approaches to methylation tests. *Forensic Sci Rev.* 2017; 29: 121–144.
3. Hamano Y, Manabe S, Morimoto C, Fujimoto S, Tamaki K. Forensic age prediction for saliva samples using methylation-sensitive high resolution melting: exploratory application for cigarette butts. *Scientific Reports* 2017; 7: 10444. https://www.nature.com/articles/s41598-017-10752-w
4. Parson W. Age estimation with DNA: From forensic DNA fingerprinting to forensic (epi)genomics: A mini-review. *Gerontology* 2018; 56: 326–332. https://doi.org/10.1159/000486239
5. Aliferi A, Ballard D, Gallidabino MD, Thurtle H, Barron L, Syndercombe Court D. DNA methylation-based age prediction using massively parallel sequencing data and multiple machine learning models. *Forensic Sci Int Genet.* 2018; 37: 215–226. https://doi.org/10.1016/j.fsigen.2018.09.003
6. Leontiou CA, Hadjidaniel MD, Mina P, Antoniou P, Ioannides M, Patsalis PC. Bisulfite conversion of DNA: Performance comparison of different kits and methylation quantitation of epigenetic biomarkers that have the potential to be used in non-invasive prenatal testing. *PLoS One* 2015; 10: e0135058. https://doi.org/10.1371/journal.pone.0135058