

# BEYOND THE CSI EFFECT

## THE KEYS TO GOOD FORENSIC GENETICS COMMUNICATION

ÁNGEL CARRACEDO AND LOURDES PRIETO

Forensic genetics brings together all the genetic knowledge required to solve specific legal problems. In recent decades new techniques have shown the potential of DNA as a profiling system. These advances have arrived hand in hand with other improvements in terms of communication of test results, with the introduction of statistical evaluation. In the collective imagination, nourished by TV series such as *CSI*, forensic evidence is presented as one hundred percent certain, but the reality is different. However, statistical analysis has allowed us to turn from handcrafted forensic medicine based on intuition and experience, to tests based on evidence and data, where uncertainty is quantified in probabilistic terms.

Keywords: forensic genetics, DNA fingerprint, criminalistics, DNA, genetic polymorphism.

### ■ INTRODUCTION

Forensic genetics is a subfield of genetics and legal medicine that includes the set of genetic knowledge required to solve certain legal problems. The most commonly requested tests at forensic genetics laboratories include paternity tests, forensic biology tests (the analysis of biological remains of criminal interest, such as blood, sperm, sweat or saliva, hair, contact evidence, etc.), corpse and cadaveric remains identification, as well as other expert specialisations including non-human DNA (illegal trafficking of endangered species, food fraud, etc.).

In Europe there are around three hundred forensic genetics laboratories (more than fifty in Spain), but only a few follow the UNE-EN ISO/IEC 17025 standard (which guarantees technical competence and results reliability) and carry out criminal investigation tests. In the rest of the world there are approximately eight hundred laboratories, usually in economically and socially developed countries. Europe still leads the scientific investigation field, but the United States, Korea, and Australia-New Zealand are experiencing greater growth. Legal medicine, thanks to the impulse of forensic genetics,

is the only field in the Science Citation Index (SCI) led by Spanish teams.<sup>1</sup>

In forensic genetics, the importance of DNA databases in the identification of criminals is growing. They are considered by the law and have been implemented throughout the European Union, as well as in many other countries around the world, and involve the introduction of millions of DNA profiles every year.

The discovery of the so-called genetic fingerprint (that is, the analysis of DNA polymorphisms, which are highly variable between individuals) by Alec Jeffreys's team in 1985 (Jeffreys, Wilson, & Thein, 1985)

represented a radical change in the possibilities of forensic genetics laboratories. For a gene locus to be polymorphic, the allele (i.e., the variant) is assumed to be the most common for the locus, and its frequency must be lower than 99%. DNA minisatellites are tandem repeats of nucleotides with a very variable number of repeats between individuals; in other words, they are highly polymorphic.

«IN FORENSIC GENETICS,  
THE IMPORTANCE OF DNA  
DATABASES IN THE  
IDENTIFICATION OF  
CRIMINALS IS GROWING»

<sup>1</sup> <http://archive.sciencewatch.com/ana/fea/11julaugFea>



Before DNA was used, most paternity cases were solved using classical markers such as blood type or variants of blood proteins and enzymes. However, the use of DNA polymorphisms has simplified the test, and has made it cheaper and more reliable. They also offer better resolution possibilities in difficult cases, such as those in which the alleged father has died – meaning the paternity investigation must be carried out with cadaveric remains or with samples from direct relatives – or in prenatal paternity diagnoses (e.g., in rape cases). All these cases were difficult to address with the methodology available before the discovery of DNA repeat polymorphisms and, especially, of microsatellite polymorphisms. Microsatellites are short tandem repeats with between two and six base pairs (although in forensic genetics the ones with two or three base pairs are not used because they produce technical artefacts that make their analysis more complicated). Microsatellites or STRs (short tandem repeats) are less polymorphic than minisatellites, but they are preferred because they can be amplified through PCR (polymerase chain reaction). They allow the procedure to be automated and over twenty of them can be simultaneously analysed (selected and validated by forensic laboratories). They have huge discrimination potential and a high level of technical standardisation has been reached worldwide, allowing extensive data exchange, very rigorous quality controls, and high levels of analysis reliability.

The revolution in the identification of skeletal remains has also been important, although some cases are remarkably difficult because of DNA degradation in the samples. Sometimes we must rely on mitochondrial DNA analysis.

It is not as variable as nuclear DNA, but it contains more copies, so finding an intact fragment when the DNA is very degraded is more probable. Of course, they cannot be used in paternity tests because their lineage follows the maternal line, but they do allow us to reconstruct lineages. Indeed, it was first used to identify the Romanovs – the last Tsar of Russia and his family – who were assassinated during the Bolshevik revolution.

Many important cases around the world have been solved thanks to DNA, such as the identification of missing persons during Argentina’s dictatorship. Many mass disasters and historical enigmas have also been – and are still being – investigated.

In forensic biology the revolution was all-encompassing, particularly regarding the analysis of

**«NEXT-GENERATION  
SEQUENCING TECHNIQUES  
ARE PRODUCING A  
REVOLUTION»**



Jane Gitschler

In 1985, Alec Jeffreys’s discovery of the so-called genetic fingerprint (that is, the analysis of DNA polymorphisms, which are highly variable between individuals) represented a radical change in the possibilities of forensic genetics laboratories.

sperm smears, hair, saliva, or miniscule blood stains, because classic markers could offer very little information about the person these remains belonged to. Today, using a single hair, a minimal number of sperm cells, or an old blood stain, we can often provide

very valuable data regarding the individuality of those remains. This was unthinkable a few years ago.

The application of DNA polymorphism evidence in crimes against sexual liberty also deserves a special mention. In these crimes, when the alleged guilty party denies the crime and the only available evidence is circumstantial – from possible sperm on clothes or in the vaginal or anal cavity. Sperm is ideal for DNA analysis, but classic markers provided very little useful data, except in exceptional cases.

In the case of male-female mixtures with a low male component, the introduction of Y chromosome microsatellite analysis was hugely important because, if there is very little male DNA in the total sample, the microsatellite profile of its autosomal chromosomes would be undetectable because of a technical PCR



Sergio Moises Panel Pitrau



problem: the preferential amplification of the most abundant DNA type. Today we can even analyse the DNA left from contact with an object, although the low amounts of DNA and contamination often make it difficult to interpret these findings.

The potential of DNA as an identification system soon led to the proposal of creating data banks containing the DNA profiles of criminals. They were first created in England in 1995, followed by Northern Ireland and Scotland in 1996. New Zealand started their own in 1996, while the Netherlands, Slovakia, and Austria created theirs one year later, in 1997. The United States, Germany, and Slovenia were next, in 1998, and step by step other developed countries created them and started to develop specific legislation for them. In Spain, these data banks are regulated by Organic Law 10/2007, of 8 October, regulating the police's database of identifiers obtained from DNA (*Ley Orgánica 10/2007, de 8 de octubre, reguladora de la base de*

**«IT IS NOT THE FUNCTION  
OF GENETICS EXPERTS TO  
EXPRESS AN OPINION ABOUT  
THE GUILT OR INNOCENCE  
OF THE SUSPECT»**

*datos policial sobre identificadores obtenidos a partir del ADN*). In addition, in December 2008, the law allowing creation of the National Commission for the forensic use of DNA was passed (*Real Decreto 1977/2008, de 28 de noviembre*).

It is worth mentioning that, although the DNA microsatellites included in databases do not provide relevant medical information in most cases, they are not completely neutral either. They can provide data about chromosomal alterations, particularly those present on sexual chromosomes, and some rare diseases. Thus, they represent sensitive information.

Perhaps the most innovative application of current forensic genetics is what is known as forensic DNA phenotyping (Kayser & De Knijff, 2011), which can determine the geographical origin, physical characteristics, and age of the person to which the biological samples used in police investigations belong. To determine ancestry, specific single nucleotide polymorphisms (SNPs) are used. These are ancestry-informative markers (AIM) which are very different between populations. This type of test was successfully used for the first time after the Madrid train bombings on 11 March 2004, to predict the geographical origin of unidentified profiles found on important objects, and this evidence was used in the legal investigation of the case (Phillips et al., 2009). The model is very effective, to the point that in most cases it can predict with high probability whether a sample is from Southern Europe or from Northern Africa – two very close populations in geographical and historical terms.

SNPs are also important for predicting the physical characteristics of an individual based on a sample, which can then be used to aid a police investigation. SNP panels and mathematical prediction tools have been developed that can reliably discriminate eye colour using samples from biological remains. Another emerging field is the determination of an individual's age by analysing the methylation patterns in biological samples. About 20% of the variation in methylation in the human genome correlates with age. Trials using a select group of methylation markers have allowed



Many important cases around the world have been solved thanks to DNA, for example, the identification of missing persons during Argentina's dictatorship. The picture shows photographs in the memory of missing persons in the hall of the Regional Avellaneda School of the Argentine National Technological University in Villa Dominico.



increasingly accurate approximations of an individual's age to be obtained (the mean error is less than three years).

Analysis of the origin of biological fluids (i.e., semen, sperm cells, saliva, menstrual blood, etc.) is also progressing rapidly thanks to the analysis of microRNA or messenger RNA (mRNA) expression, and this evidence is becoming increasingly relevant in many criminal cases. Next-generation sequencing techniques are also producing a revolution, allowing us to simultaneously analyse microsatellites, SNPs, AIMs, and physical characteristics markers. This opens new possibilities for non-human DNA analysis (metagenomics, soil analysis, pollen, illegal trafficking of protected species, etc.). Through the massive analysis of complete genomes, experts have even managed to differentiate monozygotic twins, one of the oldest challenges in forensics.

Finally, it is important to highlight that standardisation and quality control are very important in this field. Experts from the International Society for Forensic Genetics and its working groups have facilitated the creation of these standards and controls. One of them, the Spanish and Portuguese-Speaking Working Group, have defined the best quality control system (proficiency testing) to date.

■ COMMUNICATING THE VALUE OF EVIDENCE

Probably the most important development in the history of forensic science was the introduction of statistical test assessment in forensic reports. This meant moving from handcrafted forensic medicine based on intuition and experience – which applied heuristic models and valued the voice of the expert the most – to tests based on evidence, where opinion is based on data and reasoning, and uncertainty regarding an opinion is quantified in probabilistic terms. This is precisely the difference between scientific evidence and expert opinion.

Forensic genetics pioneered the quantification of the value of evidence by using probabilities. When genetic polymorphisms are analysed in biological smears and we try to ascertain whether they correspond to an individual whose DNA is also analysed, we need to calculate the probability that



One of the most important problems in forensic medicine is the so-called «CSI effect». Most TV series present forensic evidence as infallible – one hundred percent reliable, with no margin for doubt – when reality is very different: the scientific validity of forensic tests is variable. In the picture, promotional material from the TV series CSI.

«MOST TV SERIES PRESENT FORENSIC EVIDENCE AS INFALLIBLE, WHEN REALITY IS VERY DIFFERENT»

they truly correspond. This information must then be offered to the judge so that it can be combined with other non-genetic information obtained during the investigation. This is possible when we evaluate tests from a Bayesian point of view.

Thus, forensic experts can evaluate the results of their analysis from two opposing and mutually exclusive perspectives (that of the prosecution and the defence) using a likelihood ratio (LR). For instance:

$H_p$  (hypothesis of the prosecution) = the traces found at the scene of the crime belong to the defendant.

$H_d$  (hypothesis of the defence) = the traces found at the scene of the crime do NOT belong to the defendant.

The LR measures the probability of obtaining specific results from the genetic analysis of the evidence and the sample from the defendant according to these two hypotheses. In other words, it measures how much more likely it is that the genetic results obtained are from the defendant compared to the likelihood that a different individual left the trace at the scene of the crime, and is formulated as follows:



Anna Mateu



CBS

$$LR = \frac{P(E/H_p)}{P(E/H_d)}$$

Likelihood of the evidence assuming  
that the trace belongs to the defendant.

---

Likelihood of the evidence assuming  
that the trace does NOT belong to the  
defendant

where E = evidence (the genetic result in the sample from the scene and the sample from the defendant) and P = probability.

A LR of 200 means that it is 200 times more likely that the genetic profile of the sample from the scene would be found if we assume it was from the defendant ( $H_p$ ) than if we assume it was from a different person ( $H_d$ ). In many cases, the LR obtained with the genetic test will be overwhelming (LRs into the millions, which are extremely favourable to the prosecutor's hypothesis), but this is not always the case. Sometimes the results of the analysis of the biological evidence are not good enough (because

of the poor conservation status of the DNA or because there was too little DNA).

This system for evaluating evidence allows lawyers to combine the results of the genetic analysis with other non-genetic results obtained in the investigation of the criminal offence; in other words, multiplying the value of the LR by the value of the non-genetic test (the *a priori* probability). The result of this multiplication is called the *a posteriori* probability (i.e., the probability of «guilt» according to the evidence, which is what the judge wants to know). Its formulation is:

$$Pa_{posteriori} = Pa_{priori} \times LR$$

To calculate the *a priori* probability, the judge must assess all the information from the investigation, looking at the odds. The judge has an idea about the «guilt» or «innocence» of the defendant before looking at the results from the genetic tests, thanks to other indications (witnesses who might have identified the defendant at the scene, lack of an alibi,



Forensic DNA is scientifically valid evidence, but the information it provides changes depending on the case. Thus, communicating the assessment of the evidence in probabilistic terms is very important.

**«STATISTICAL TEST ASSESSMENT IN  
FORENSIC REPORTS MEANT MOVING FROM  
HANDCRAFTED FORENSIC MEDICINE  
BASED ON INTUITION AND EXPERIENCE  
TO TESTS BASED ON EVIDENCE»**

**«THE EXPERTS' ASSESSMENT OF GENETIC  
EVIDENCE USING THE LR IS ASEPTIC,  
GUARANTEEING THAT IT IS NOT  
INFLUENCED BY OPINIONS»**

etc.). This information can be translated to a figure (for instance, 1,000 to 1 in favour of innocence if the judge thinks it is very likely that the defendant is innocent). The judge can integrate all the information simply by multiplying the *a priori* probability by the LR, to obtain the *a posteriori* guilt probability.

Thus, for example, if the judge has non-genetic evidence against the defendant (for instance, 1,000 to 1 in favour of the defendant being guilty) and, in addition, a bloodstain found on their clothes coincides with the victim's genetic profile (for example, with a LR = 1 million), the *a posteriori* probability of guilt will increase a lot (1,000 times 1 million) because of the LR – i.e., because of the scientific evidence.

Conversely, if a cigarette filter found at the scene (the victim's home) is being analysed, its genetic profile is complete, and it coincides with the defendant's profile, but the judge knows that the evidence might have ended up at the scene without it implying their guilt (for example, because the victim and the suspect lived together and there is no further information), its *a priori* probability must be low. Therefore, the scientific evidence would not increase



the *a posteriori* probability much, despite the high LR. In this extreme case, if the judge was sure the defendant were innocent, the *a posteriori* probability would be 0 despite having a LR in the millions.

These examples clearly show that the judge is responsible for assessing the evidence as a whole and the Bayesian approach can prevent experts from acting as judges. The genetics experts do not know the non-genetic information that the judge does, so it is not their function as experts to express an opinion about the guilt or innocence of the suspect. The experts' assessment of genetic evidence using the LR is aseptic, guaranteeing that it is not influenced by opinions or information they might have received by other means (e.g., from the press or TV).

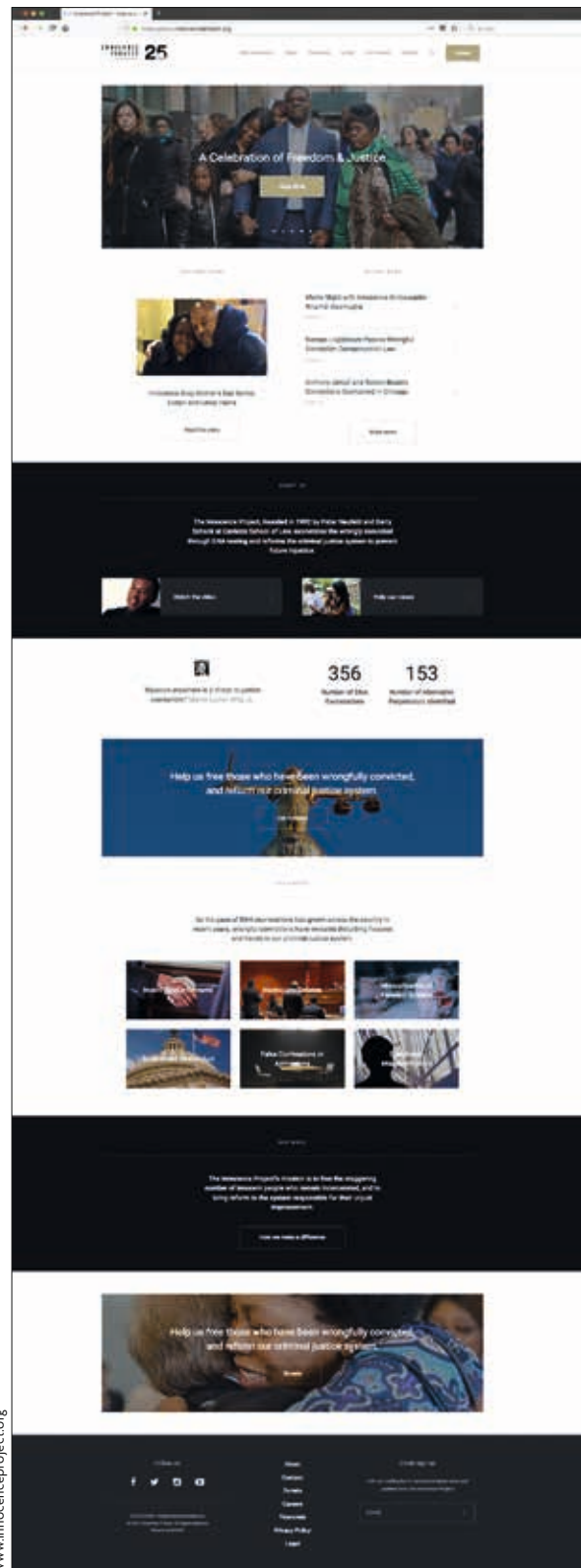
Despite the advantages of assessing evidence from the Bayesian point of view, this assessment is not free of mistakes and misunderstandings. One of the most common is to mix up the LR and *a posteriori* probability. For instance, the correct way to express a LR = 1,000 in words would be: «It is one thousand

**«EXPERTS ARE TRYING TO IMPROVE COMMUNICATION AND MAKE IT FAIRER AND LESS PRONE TO INTERPRETATION ERRORS»**

times more likely that the evidence from this genetic profile (the one resulting from the analysis) would be gathered at the scene if the profile belongs to the defendant than if it belongs to a different random Spanish person.» However, the LR is sometimes put into words incorrectly. For instance: «It is one thousand times more probable that this profile belongs to the defendant compared to it belonging to a different random Spanish person».

In the correct example we are assessing the evidence (the genetic profile found in the evidence) assuming two hypotheses (whether it belongs to the defendant or not). In mathematical terms, it translates to  $P(E/H_p) / P(E/H_d)$ , exactly the definition of LR. However, in the incorrect example we are talking about the probability of the hypotheses (whether the profile belongs to the defendant or not) without considering the evidence; i.e., we are defining something completely different. In mathematical terms, it would be  $P(H_p) / P(H_d)$ , which does not define the LR.

Intuitively, it is very easy to confuse the question that the judge considers with the question that the



www.innocenceproject.org

The Innocence Project initiative has exonerated more than three hundred falsely accused individuals thanks to modern DNA tests.



experts consider. The judge wonders what the probability of guilt is given the result of the DNA test, and the experts wonder what the probability is that the DNA test gave a specific result because it belongs to the defendant or to a different person. Mixing them up or communicating them incorrectly is known as a *transposed conditional* and is one of the biggest causes of interpretation errors (Carracedo & Prieto, 2014; Evett, 1995).

Through different initiatives, experts are trying to improve communication and make it fairer and less prone to interpretation errors, but a similar effort by the judiciary would also be necessary. Thus, the education of judges and prosecutors should include the interpretation and assessment of forensic evidence and, particularly, of forensic DNA.

#### ■ FORENSIC MEDICINE AND THE MEDIA

One of the most important problems in forensic medicine is the so-called «CSI effect». Most TV series present forensic evidence as infallible – one hundred percent reliable, with no margin for doubt – when reality is very different: the scientific validity of forensic tests is variable, as stated in the *PCAST* report (President’s Council of Advisors on Science and Technology) published in 2016 by the Executive Office of the President of the United States (The International Association for Identification, 2018). Forensic DNA is scientifically valid evidence, but the information it provides changes depending on the case. This is why communicating the assessment of the evidence in probabilistic terms is so important.

The EUROFORGEN (2017) network has promoted the guide *Making sense of forensic science* to explain everything about forensic DNA’s potential, as well as its limitations, using specific examples of how a bad interpretation can lead to errors which neither experts nor judges can prevent.

The Innocence Project<sup>2</sup> initiative has exonerated more than three hundred falsely accused individuals thanks to modern DNA tests. Although the most significant cause of errors is derived from witness identification, misinterpreted forensic expertise is not a minor problem.

The key, as in so many other matters, is education: for law professionals – especially judges and

prosecutors – and for the general population, so they can critically analyse the news. Regarding the news, the media should also strive not only to inform, but also to contribute to education through dissemination, especially in fields such as this one, which are prone to sensationalism. It would also be advisable for the media to adopt strict ethical standards for the dissemination of this sort of news so that, apart from respecting freedom of information, they also respect the independence of judges and experts, as well as the general principles of law. ☺

### «THE MEDIA SHOULD STRIVE NOT ONLY TO INFORM, BUT ALSO TO CONTRIBUTE TO EDUCATION THROUGH DISSEMINATION»

#### REFERENCES

- Carracedo, A. & Prieto, L. (2014). Valoración de la prueba genética. In M. Casado, & M. Guillén (Eds.), *ADN forense: Problemas éticos y jurídicos* (pp. 145–156). Barcelona: Observatori de Bioètica i Dret, Universitat de Barcelona.
- EUROFORGEN. (2017). *Making sense of forensic genetics*. London: Sense about Science. EUROFORGEN. Retrieved from <http://senseaboutscience.org/wp-content/uploads/2017/01/making-sense-of-forensic-genetics.pdf>
- Evett, I. W. (1995). Avoiding the transposed conditional. *Science and Justice*, 35(2), 127–131. doi: [10.1016/S1355-0306\(95\)72645-4](https://doi.org/10.1016/S1355-0306(95)72645-4)
- Jeffreys, A. J., Wilson, V., & Thein, S. L. (1985). Hypervariable minisatellite regions in human DNA. *Nature*, 314, 67–73. doi: [10.1038/314067a0](https://doi.org/10.1038/314067a0)
- Kayser, M., & De Knijff, P. (2011). Improving human forensics through advances in genetics, genomics and molecular biology. *Nature Reviews Genetics*, 12(3), 179–192. doi: [10.1038/nrg2952](https://doi.org/10.1038/nrg2952)
- Ley Orgánica 10/2007, de 8 de octubre, reguladora de la base de datos policial sobre identificadores obtenidos a partir del ADN. (2007). Retrieved from <https://www.boe.es/buscar/act.php?id=BOE-A-2007-17634>
- Phillips, C., Prieto, L., Fondevila, M., Salas, A., Gómez-Tato, A., Álvarez-Dios, J., ... Lareu, M. V. (2009). Ancestry analysis in the 11-M Madrid bomb attack investigation. *PLOS One*, 4(8), e6583. doi: [10.1371/journal.pone.0006583](https://doi.org/10.1371/journal.pone.0006583)
- Real Decreto 1977/2008, de 28 de noviembre, por el que se regula la composición y funciones de la Comisión Nacional para el uso forense del ADN. (2008). Retrieved from <https://www.boe.es/buscar/doc.php?id=BOE-A-2008-19992>
- The International Association for Identification. (2018). *PCAST report – Forensic science in criminal courts: Ensuring scientific validity of feature-comparison methods*. Retrieved from [https://www.theiai.org/president/201609\\_PCAST\\_Forensic\\_Science\\_Report\\_FINAL.pdf](https://www.theiai.org/president/201609_PCAST_Forensic_Science_Report_FINAL.pdf)

**Ángel Carracedo**. Full Professor of Legal Medicine of the University of Santiago de Compostela (Spain). He was the director of the Institute of Legal Medicine at the University of Santiago de Compostela between 1994 and 2012, and currently coordinates the Genomic Medicine Group at the same university and directs the Spanish National Genotyping Centre. He is a research team leader at the Centre for Biomedical Network Research on Rare Diseases (CIBERER), and has received many awards, including the Jaume I Award for Medical Research and the National Genetics Awards. He has been given honorary doctorates from several universities. ✉ [angel.carracedo@usc.es](mailto:angel.carracedo@usc.es)

**Lourdes Prieto**. PhD in Biology. She is a researcher affiliated to the Institute of Forensic Science at the University of Santiago de Compostela (Spain) and is the Vice President of GHEP-ESFG (the International Society for Forensic Genetics Spanish and Portuguese-Speaking Working Group). She has extensive experience in forensic genetics and has participated in international humanitarian identification efforts.

<sup>2</sup> [www.innocenceproject.org](http://www.innocenceproject.org)