# IDENTIFYING VICTIM REMAINS FROM UNCERTAIN DATA

**Mark Perlin, Ph.D., M.D., Ph.D., Alexander Sinelnikov, Ph.D., Erin Vey, Matthew Legler, Meredith Clarke**
*Cybergenetics, Pittsburgh, PA*

A mass disaster can generate a vast amount of biological material. Such materials include the victim remains (VR) evidence at the disaster site, as well as the personal effects (PE) and family references (FR) of the missing people. These data may have considerable uncertainty, which occurs at many levels. Current reference sample expert systems do not address genotype uncertainty, so instead newer statistical computing methods are needed. This paper describes how computer tools can help forensic analysts overcome mass disaster data uncertainty, and thereby make reliable identifications, in the context of the World Trade Center (WTC) STR data.

At the DNA level, VR samples are often degraded, mixed, burned, contain minimal DNA, or are compromised in other ways. PE samples are collected from contaminated environments, and may include mixed or degraded DNA. With FR samples, there can be uncertainty in the family kinship relationships. Human visual review of the STR data is needed to determine which data are good, mixed or unusable.

Indeed, the STR lanes (or injections) associated with a particular sample may be derived from different individuals. To disambiguate the data, we use a Visual User Interface (VUIer™) tool that lets an analyst view all of the provided lanes for a given sample, and then group compatible lanes together to form one or more TrueAllele® genetic calculation requests for computer interpretation of that sample. In our WTC workflow, a VUIer operator can inspect data and generate interpretation requests for 30 VR samples every hour.

When inferring a profile from uncertain mass disaster data, the resulting DNA profile may include more than one genotype. This profile uncertainty can be represented by assigning probabilities to genotypes [1]. The forensic analyst can use the TrueAllele genetic calculator to infer genotypes in many common scenarios. For example, the calculator can infer profiles from mixtures having two unknown contributors, statistically combine uncertain data from multiple lanes, and infer a missing person's profile from kinship data.

To compare victim remains profiles with missing person profiles, the TrueAllele matching module computes a likelihood ratio that compares the probability of a victim remains match with a missing person to that with a random person [1]. Every match comparison in the system, whether kinship or STR, has a numeric likelihood ratio score that measures the degree of DNA identity.

To help the forensic analyst integrate all of this profile and match information, the TrueAllele VUIer program provides visual representations at each locus of the genotype contributor probabilities and match likelihood ratios. The original EPG data can be seen and explored by the user at all times. A match-directed sample review takes only a few minutes, since the user can interactively focus on interesting questions about data, genotypes or matches.

We have visually reviewed and statistically reanalyzed a subset of the WTC data for 18,251 VRs and 2,386 PEs. We have genetically reconstructed 2,347 missing person profiles from 6,660 FRs. In a mass disaster project of this magnitude, forensic analyst expertise is essential for examining the data to ask the correct questions, and assess the calculated answers. Computer tools, including a visual user interface and a 24 processor TrueAllele supercomputer for statistical genotype inference and matching, can perform genetic calculations that help analysts apply their expertise to obtain reliable identifications with less effort.

[1] Perlin MW, Kadane JB, Cotton RW. Forensic DNA Inference. In: Seventh International Conference on Forensic

Inference and Statistics; 2008 August; Lausanne, Switzerland.