Forensic DNA Phenotyping Legislation Cannot be Based on "Ideal FDP"

A Response to Caliebe, Krawczak & Kayser (2017)

Buchanan N, Staubach F, Wienroth M, Pfaffelhuber P, Surdu M, Lipphardt A, Köttgen A, Syndercombe-Court D, Lipphardt V.

(accepted, but not final version)

Dear Editors,

In a recent correspondence, Caliebe, Krawczak & Kayser (2017) discussed our *Nature* correspondence (Staubach 2017) and an invited expert-advisory presentation we delivered at the German Ministry of Justice, Berlin (uploaded in manuscript form to our website (https://stsfreiburg.wordpress.com/)), alongside presentations by Dr. Kayser and others. We subsequently invited Dr. Caliebe to give a talk in Freiburg in July 2017. After extensive discussion with her, we have acknowledged that FDP might not be fully equivalent to diagnostic testing. Furthermore, immediately following this discussion, we formally stated so in a paragraph added to the talk's manuscript.

Nonetheless, it is necessary to respond to Caliebe, Krawczak & Kayser (2017) for two reasons: First, we would like to clarify the relevance of this dialogue for the ongoing public and political debate over Forensic DNA Phenotyping (FDP) in Germany. Second, we will address aspects that must be considered before we can know whether or not predictive values for FDP are prevalence-independent, as claimed.

We appreciate this scientific debate, as it enables decision-makers in Germany to better understand why their recent legislative initiatives to permit the technique would benefit from further consideration of the complexities of the techniques and the intricacies of how results are communicated and interpreted. Upon request, lawmakers disclosed that these initiatives relied on an review of FDP techniques by Kayser (2015) to make the argument that FDP could reliably predict hair (most colors 75-90%), eye (brown and blue eyes: 90-95%) and skin color (dark vs. light: 98%) (Bundesrat 2017). However, the values provided in that review are AUCs that, according to Caliebe, Walsh, Liu, Kayser & Krawczak (2017), cannot sensibly be used in criminal casework. Predictive values are often lower (Caliebe, Walsh, Liu, Kayser & Krawczak 2017). Interestingly, there is no statement or data presented in that review that would support the 98% for skin color prediction (light vs. dark), leaving it an open question what the source of this information in the law bill was.

Our key point in various publications and presentations is that too many open questions about FDP remain to implement it in investigative work in Germany, as envisioned. We understand the last sentence in Caliebe *et al.* (2017) ("these measures have to be determined empirically for each appearance phenotype, prediction model and target population of interest before they can be applied sensibly in criminal casework") as fully supporting this stance. However, the political and public support for these technologies in Germany rests heavily on the problematic notion that, in practice, many traits can be determined independently of context, with certainties as high as the AUCs given in Kayser (2015).

With regard to the predictive power of FDP, we maintain that it cannot be expressed sensibly by AUCs. Even though prevalence might not need to be considered in the same way it is done in diagnostic settings, we maintain that it remains unclear whether FDP will live up to what the authors call "ideal FDP" and prove to be population independent. We believe this for a number of reasons (some of which are also acknowledged by Caliebe, Krawczak & Kayser (2017)):

First, as the Caliebe, Krawczak & Kayser (2017) concede, current prediction models do not account for all causal factors for the determination of hair-, eye- or skin-color ("the etiological understanding of FDP-relevant appearance phenotypes is still incomplete, so are the prediction models used"). In other words, various co-factors, such as yet-to-be-accounted-for genetic markers, population affiliation, or environmental factors may influence predictive values (see e.g. section 4.2 of Caliebe, Walsh, Liu, Kayser

& Krawczak 2017). While the authors' argument that FDP is population-independent may hold true for "ideal FDP," the current state of the art obviously does not meet this criteria. Whether "ideal FDP" is possible at all will probably be contested by colleagues from science theory (see, e.g., Oreskes et al 1995).

Second, we are not convinced that the predictive values given in Caliebe, Walsh, Liu, Kayser & Krawczak (2017) are fully prevalence independent. That "the corresponding predictive values vary to a certain degree" (Caliebe, Krawczak & Kayser 2017) suggests a more complex relationship between population, prevalence and specific traits. For example, in the empirical data cited by Caliebe, Walsh, Liu, Kayser & Krawczak (2017), the positive predictive values for brown eyes are low in those countries where the prevalence of brown eyes can be assumed to be low (65% in Norway, 69% in Estonia, 68% in the Netherlands, and 67% in the UK; whereas 96%in Greece and 92% in Spain). This rather supports a model in which population/prevalence could play a role for predictive values. Moreover, these numbers are considerably lower than the numbers in the German law drafts that were informed by Kayser (2015).

We wish to emphasize that we do not subsume biogeographical ancestry (bga) under externally visible characteristics, in agreement with Walsh and Kayser (2017), yet in contrast to many statements in the debate in Germany.

We welcome the call to action by Kayser, Caliebe and Krawczak (2017) for further empirical studies to "determine" (or infer) these measures empirically, for each appearance phenotype, prediction model and target population. We also recognize what a significant task this will be. We suggest further inquiry into other aspects, for example, what kinds of populational data are being collected from individuals, and under what premises. In addition, the process of defining units under study (e.g. populations) should be made transparent. And finally, to allow addressing questions of replicability, data should be made available to other research teams to encourage methodological progress.

References

Bundesrat (2017). Entwurf eines Gesetzes zur Erweiterung des Umfangs der Untersuchungen von DNA-fähigem Material. Drucksache 117/17, February 3.

Caliebe, A., Walsh, S., Liu, F., Kayser, M. & Krawczak, M. (2017). Likelihood ratio and posterior odds in forensic genetics: Two sides of the same coin. *Forensic Science International: Genetics*, 28, 203-210.

Caliebe, A., Krawczak, M. & Kayser, M. (2017 [online, in press]). Predictive values in Forensic DNA Phenotyping are not necessarily prevalence-dependent. *Forensic Science International: Genetics.*

Kayser, M. (2015). Forensic DNA Phenotyping: Predicting human appearance from crime scene material for investigative purposes. *Forensic Science International: Genetics*, 18, 33-48.

Oreskes, N., Shrader-Frechette, K., & Belitz, K. (1995). "Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences." *Science* 263, 5147, 641–46.

Staubach, F. (2017). Correspondence: Germany: Note limitations of DNA legislation. *Nature*, 545, 30.

Walsh, S. & Kayser, M.(2017). Predicting Human Appearance from DNA for Forensic Investigations. In: Amorim, A., & Budowle, B. (Eds.), *Handbook of Forensic Genetics: Biodiversity and Heredity in Civil and*