FORENSIC TRANSCRIPTOMICS: PROVIDING CONTEXT TO DNA PROFILES

Jack Ballantyne, PhD and Erin Hanson, PhD, University of Central Florida
National Center for Forensic Science, PO Box 162367, Orlando, FL 32816

Correspondence should be sent to Jack Ballantyne (jack.ballantyne@ucf.edu)

Introduction

A crime scene DNA profile can have tremendous probative value in criminal investigations. Whether it does have value depends upon the context in which it was found. If it directly relates to the crime in question and either supports or doesn't support the posited sequence of events, then it could have significant value in helping establish facts that may be in question. However, absent any additional contextual information, the DNA profile does not necessarily indicate the relevance to the crime in question and does not inform about the actions that took place that resulted in the DNA being deposited at the scene. So how can one obtain contextual information about the DNA profile, independent of non-scientific meta data? Two different complementary approaches may help in helping determine DNA profile contextuality. These are studies on the transfer, persistence, prevalence and recovery (TPPR) of DNA traces [1] and determining the cellular (pheno)type from which the DNA originated. The latter can be important in supporting sexual versus social intercourse, for example, by distinguishing menstrual blood from peripheral blood on suspected perpetrators in sexual assault cases [2-3]. This short review is concerned with cellular phenotype determination.

Molecular methods for cellular phenotype determination (i.e. body fluid and tissue identification) include classical biochemical methods, mainly single-plex antigen-antibody reactions, as well more specific, multiplex methods based upon the transcriptome, the epigenome, the proteome and even the microbiome [2, 4-7]. While all of these "-omes" contain tissue specific biomarkers, the transcriptome is of particular interest in this era of Next Generation Sequencing (NGS) [8], since most of the genome (including sequence variation) is faithfully reproduced in multiple copies of RNA. Below we discuss what we know about the transcriptome (concentrating mainly on mRNA) and how it can be used in forensic genomics now or in the future.

Transcriptomics

What have fundamental studies revealed about the human transcriptome [9-12]. Although the 21,000 known protein coding genes constitute ~2% of the genome, >80% of the genome may be transcribed into different RNA species (although not in a single tissue). Primary transcription is recognized as the major driver of cellular specificity and therefore the transcriptome is an indicator of cellular phenotype. It is this fact that primarily makes RNA typing potentially useful in forensic genomics. Importantly, genes vary more across tissues (47% of total variance in gene expression) than individuals

(4% of total variance). Approximately 12% of genes are preferentially enriched in a single tissue (i.e. > 5X compared to other tissues) and 10% of transcripts in a tissue come from expression-elevated genes (differential expression in several tissues compared to most). Only a small percentage of genes are exclusively expressed in a single tissue (<200 total, i.e. <1%). On the other hand, almost half of all genes (44%) are expressed in all tissues (i.e. housekeeping genes). Additional tissue specific variation is obtained due to the fact that ~50% genes express tissue dependent isoforms because of alternative transcription start and termination sites, although not so much variation as initially believed due to alternative splicing.

Forensic Transcriptomics

Why is transcriptomics useful in forensic body fluid and tissue analysis? Despite its thermodynamic instability compared to DNA, RNA of sufficient quality and quantity for analysis is recoverable from evidentiary items typically found in forensic biology casework [2]. It is possible to co-extract DNA and RNA [2], and since the RNA is converted to DNA (cDNA) the same downstream chemical analysis as DNA can be conducted using the same analytical platforms. Contrary to genomic DNA where two copies of an autosomal DNA segment are present in a cell, multiple genome segments are copied into RNA species hundreds or thousands of times in a cell, a copy number somewhat akin to mtDNA. This increase in copy number may help ameliorate the reduction in RNA target number due to normal RNA degradation that takes place during the drying phase after the physiological fluid has been deposited *ex vivo*. Importantly for future investigations, genomic information is also directly encoded and reproduced in RNA transcripts. Thus, the emerging NGS/MPS technology permits genomic-encoded RNA data to be accessed. This digital gene expression technology permits the direct counting of transcripts and variants encountered in a sample and facilitates quantitative analysis. Today it is possible with RNA profiling to identify with a high degree of certainty the presence in dried stains of most of the commonly encountered forensically relevant body fluids and tissues including blood, semen, saliva, vaginal secretions, menstrual blood and skin [2]. Similarly, RNA methods have been developed to identify internal organ tissues, which may be useful in investigations involving significant trauma [13,14].

Future of Forensic Transcriptomics

Although it is now theoretically possible to definitively identify the most relevant body fluids in a non-physically separable mixture, it is not yet possible to directly link specific body fluids to individual DNA profiles in the mixture. The latter is needed to evaluate source versus sub-source level propositions as part of a robust and accurate mixture deconvolution and DNA interpretation process. However, due to the presence of coding region SNPs in body fluid mRNA transcripts it should be possible to do this by an NGS targeted transcriptome analysis [15]. Point-of-use, non-PCR based RNA methods currently under development may be used in the future for rapid (< 1h) definitive body

fluid identification as a triage for downstream DNA analysis. It is also possible that non-microscopical identification of spermatozoa will be possible due to the presence of sperm specific RNAs (e.g. protamines such as PRM1). It is likely that the transcriptome will be a constituent part of a (multi)omics/data integration process for evidence analysis that will include the not-isolated biological entities of the genome, transcriptome, methylome, proteome and microbiome. This will include mRNA and miRNA [2] since mRNA is regulated by miRNA binding and both exhibit tissue specific expression. It is expected that AI, such as machine learning, will play an increasing role in data analysis and interpretation of the transcriptome and other -omes. Socio-economic forces will probably result in the use of whole genome sequencing approaches (Whole Genome (WGS)/Transcriptome (WTS)/Exome (WES) exploiting the concept that one obtains the whole -ome data from which the relevant case dependent biomarker information is subsequently extracted. Since most of the genome is faithfully reflected in transcribed RNA, it is not out with the bounds of possibility that RNA could be used as a complement to DNA typing for the whole gamut of forensic genomic applications including personal Identification, providing context to the identification and obtaining a genetic eyewitness of the donor of a physiological stain (i.e. a phenotype that includes ethnicity, sex, age, external visible and behavioral traits etc.).

Acknowledgements

The authors would like to thank all students past and present who have worked on RNA profiling during the past 20 years, on funding agencies who have believed in the technology (NIJ, DoD, FBI) and on productive and fun international collaborators, especially Cordula Haas' group at the University of Zurich.

References

[1]. Van Oorschot R., Szkuta B., Meakin G., Kokshoorn B. and Goray M., DNA transfer in forensic science: A review, Forensic Sci Intl Genet., 38 (2019) 140-166.

[2]. Hanson E. and Ballantyne J., RNA profiling for the identification of the tissue origin of dried stains in forensic biology, Forensic Sci. Rev. 22(2) (2010) 145-157.

[3]. Tozzo P., Nespeca P., Spigarolo G. and Caenazzo L., The importance of distinguishing menstrual and peripheral blood in forensic casework: a case report, Am J Forensic Med Pathol. 39(4) (2018) 337-340.

[4]. Frumkin D., Wasserstrom A., Budowle B. and Davidson A., DNA methylation-based forensic tissue identification, Forensic Sci.Int.Genet. 5 (2011) 517-524.

[5]. Lee H.Y., Park M.J., Choi A., An J.H., Yang W.I., and Shin K.J., Potential forensic application of DNA methylation profiling to body fluid identification, Int.J.Legal Med. 126 (2012) 55-62.

[6]. Legg K.M., Powell R., Reisdorph, N., Reisdorph R., and Danielson P., Discovery of highly specific protein markers for the identification of biological stains, Electrophoresis. 35 (2014) 3069-3078.

[7]. Clarke T., Gomez A., Singh H., Nelson K. and Brinkac L., Integrating the microbiome as a resource in the forensic toolkit, Forensic Sci. Int. Genet., 30 (2017) 141-147.

[8]. Shendure J. and Ji H., Next-generation DNA sequencing, Nat.Biotechnol. 26 (2008) 1135-1145.

[9]. Uhlen M., Fagerberg L., Hallstrom BM, Lindskog C., Oksvold P., et. al, Tissue-based map of the human proteome, Science 347 (6220) (2015) 1260419.

[10]. Uhlen M., Oksvold P., Fagerberg L., Lundberg E., Jonasson K., et. al, Towards a knowledge-based Human Protein Atlas, Nat Biotechnol. 28 (12) (2010) 1248-1250.

[11]. The GTEx Consortium, The Genotype-Tissue Expression (GTEx) project, Nature Genetics 45(6) (2013) 580-585.

[12]. ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome, Nature 489 (7414) (2012) 57-74.

[13]. Hanson E. and Ballantyne J., Human organ tissue identification by targeted RNA deep sequencing to aid the investigation of traumatic injury, Genes 8 (2017) 319.

[14]. Lindenbergh A., van den Berge M., Oostra R., Cleypool C., Bruggink A., Kloosterman A. and Sijen T., Development of a mRNA profiling multiplex for the inference of organ tissue, Int. J. Legal Med. 127(5) (2013) 891-900.

[15]. Ingold S., Haas C., Dorum G., Hanson E. and Ballantyne J., Association of a body fluid with a DNA profile by targeted RNA/DNA deep sequencing, Forensic Sci. Intl. Genetics Supp. Ser. 6 (2017) e112-e113.