Comments on FELIX results presentation by IARPA, Ginkgo Bioworks, Charles Stark Draper Laboratory and Concentric

https://www.youtube.com/watch?v=XQjR3mmFLhk 2022/10/17 and other links.

Yves Bertheau

Honorary INRA research director at the Muséum national d'Histoire naturelle, Paris

Background

Detecting biothreats is a major concern that was amplified after the events of 11th September. Numerous countries and regions, such as the EU, gathered experts in detection (in the fields of e.g., GMOs and microbiology) to determine future trends and decide which tools and concepts to develop against biological weapons. Attribution of genetically engineered sequences (identification and determination of their source) was one of the major challenges.

The problem of attribution of sequences and of organisms derived from synthetic biology, broadly defined, is not new (Lewis et al., 2020), with or without the use of artificial intelligence (AI) techniques.[1] Methods have been developed to enable such attributions, which can be used to determine the natural or non-natural origin of a sequence for biomonitoring purposes (Kunjapur et al., 2018), as well as to identify the laboratory, or even the technician handling the sequences (Alley et al., 2020; Nielsen and Voigt, 2018; Wang et al., 2021). The techniques allow an estimation of uncertainty, follow-up of the "genealogies" of the genetic engineering groups, and the identification of the nation in which the laboratory of origin is located.

Each of these capabilities holds promise for GMO identification, which would help to ensure food safety and uncontaminated food chains. An uncertainty estimation enhances robustness and can facilitate the integration of technical indicators with other available information to make a decision on attribution. Lineage tracking can identify other groups that knowingly or unknowingly aided the responsible actor, as well as identifying the nation of origin and/or the responsible laboratory, thus providing useful investigative clues in the absence of more specific information about the GMO.

Identification and attribution of the sequence, and thus engineered organism, is carried out using multiple possible sources of information on genomic variation, as in the

[1] With the usual dangers of projecting human priorities, overconfidence and miscalibration of the AI.

case of microorganisms (Budowle et al., 2005; Reisch and Prather, 2015), including genomic and epigenomic "scars" (unintended DNA damage and other traces) that some developers of new breeding techniques (NBT) or gene therapies are trying to avoid (Akella et al., 2021; Elison and Acar, 2018; Ellison et al., 2021; Ellison et al., 2020; Steyer et al., 2018). The genetic attribution of sequences can be used to detect biological threats as well as to protect industry's intellectual property resulting from patents on technologies and traits. It is already the subject of a competition[2], which aims to develop methods not only to detect engineered sequences and thus GMOs of any origin, but also to form a basis for proficiency tests, as already practised by GMO control laboratories (Crook et al., 2021).

The terminology of "scars" is used:

- in GMO detection, for traces left by the related techniques, including delivery tools with (for example) unintended Agrobaterium plasmid and genomic insertions, cell cultures and their somaclonal variations, selection markers and their excision, plant regeneration and its unintended mutations and epimutations – all scars which cannot be fully discarded by backcrosses with Elite varieties (Bertheau, 2019; 2022);
- in biodefence (cf. former FELIX researcher, 2020[3] and Gryphon[4] scientific poster, 2017) for the traces left by genome manipulations (Slezak et al., 2020; Slezak et al., 2009);
- in infectious pathology, because genotoxic bacteria, such as *Escherichia coli* or *Helicobacter pylori*, induce double-strand breaks in the genome of the infected host, causing scarring in the human genome (Cao et al., 2022; Cuevas-Ramos et al., 2010; Lemercier, 2014);
- in oncology for cancer diagnosis, following poor genome repair (for genetic and epigenetic reasons) by homologous recombination after DNA double-strand breaks (DSB) (Pacheco-Barcia et al., 2022; Rempel et al., 2022; van der Wiel et al., 2022; Watkins et al., 2014). These "natural" DSBs are similar to those brought about by genome editing through sequence specific nucleases. An accurate homologous recombination for repairing natural or artifactual DNA damage is a "Holy Grail" that has proven difficult to achieve through genome editing (Domingo-Prim et al., 2020; Scully et al., 2019);
- in microbiology, for manipulations carried out on plasmids (among other constructs), for which one tries to eliminate the traces of manipulation. As an example, enzymes called recombinases can be used to try to eliminate the markers of selection (antibiotic resistance genes) for genetically transformed cells (Bertheau, 2022; Fels et al., 2020).

These molecular biomarkers or scars range from single nucleotide polymorphisms (SNPs, a type of genetic variant) to large chromosomal rearrangements and aneuploidy (the presence of an abnormal number of chromosomes in a cell). They are used for diagnostic

---

[2] GEA challenge (Genetic Engineering Attribution challenge) https://www.drivendata.org/competitions/63/genetic-engineering-attribution/ and https://drivendata.co/blog/genetic-engineering-attribution-winners

[3] https://futurehuman.medium.com/how-do-we-know-if-a-virus-is-bioengineered-541ff6f8a48f

[4] https://www.iarpa.gov/images/PropsersDayPDFs/FELIX/Gryphon-Scientific-Poster.pdf

purposes or identification of past events, either as isolated elements or grouped into a convergent identification cluster (matrix approach/multivariate analysis). These scars have been considered since 2012 as constituting “mutational signatures” resulting from any type of DNA damage (Alexandrov et al., 2020; Degasperi et al., 2022; Wong et al., 2022; Zou et al., 2018). These mutational signatures are then analysed by one of the many computer tools available for detection, and a diagnosis is achieved by multivariate and AI approaches on all or some of the scars (Chevalier et al., 2021; Gulhan et al., 2019; Omichessan et al., 2019; Perner et al., 2020; Wang et al., 2022). These results can then be completed by other non-genomic information (such as other types of documentation for the GMO) and the whole exploited by decision support tools, such as decision support systems (DSS) or AI, for the purposes of detection of unknown GMOs (Bensadoun et al., 2016; Bohanec et al., 2013; Bohanec et al., 2016; Dobnik et al., 2018; Holst-Jensen et al., 2013; Hurel et al., 2020).

These problems of poor repair of on-target (at the intended edit site of the genome) and unintended off-target (elsewhere in the genome) double-strand breaks are one of the barriers to the use of genome editing in gene therapy.

I had already pointed out the similarity of the mutational scars resulting from gene editing to the scars that in combination define the "mutational signatures" of cancers in my 2019 paper, as the human/medical field is quite often ahead of the agricultural gene editing field, due to better funding (Bertheau, 2019).

FELIX

In the video published on Youtube, the presentations of the different partners contracted by IARPA, although very enthusiastic, are quite acceptable from the point of view of molecular and computer techniques. However, the first limitation of the work presented is the lack of peer-reviewed articles that will provide a precise idea of the quality of the work done, including its reproducibility,[5] even excluding the issue of access to DNA chips and computer tools. No publication was announced during the press conference, which does not bode well for the availability of the methods for GMO control laboratories.

The process can be summarised as follows:

- Creating a database of characteristic sequences of scars resulting from genetic manipulation;
- Conducting microarrays of these sequences in order to trap and enrich them;
- Sequencing (short, long, and mixed fragments) of enriched fragments;
- Sequence Analysis Pipeline (aligning the sequence reads against the reference sequences from the FELIX project database and non-GMO/natural organisms) and assignment of genetic scars to certain engineering tools;
- Analysis of the different scars for identification of the detected organisms.

[5] With the usual dangers of AI techniques.

If we limit ourselves to the use for the detection/identification of GMOs, the approach described above can be followed using known and easily accessible techniques. It is directly applicable to products resulting from genome editing, whether to enable ENGL[6] (European Network of GMO Laboratories) to detect and identify a non-authorised GMO, or for notifying firms within the framework of European GMO regulations. The use by FELIX contractors of private sequence banks gathered from the sequence and patent databases is therefore not prohibitive for the detection of GMOs intended for consumption. It is likely that with access to more pangenome sequences, and thus greater knowledge of the range of natural variability, these targeted sequence banks will be less and less necessary. All that is missing are applications to varieties derived from the NBT products.

Indeed, and this is the second limitation to the work of the firms involved in FELIX, the GM organisms used are not necessarily representative of the commercial varieties (cf. Figure 1, (Bertheau, 2022). Gingko Bioworks has been working since 2017[7] with Bayer and created a joint venture with Bayer on October 18, the day after the press conference, so the project had the opportunity of a real proof of concept, but only used "model" GM organisms. It would have been interesting and useful if the FELIX contractors had been closer to the competitors of the GEA challenge (Genetic Engineering Attribution challenge), a data science competition to identify engineered sequences and their original source. This would have increased the sample size (the number of sequences in the FELIX reference database), which is still small, as an IARPA representative regretted, and would probably have provided useful feedback (Crook *et al.*, 2021).

The third limitation identified during the press conference was the accessibility and analyses cost of these platforms.

Finally, the fourth limit for a direct application to the detection of GMOs is to determine, for pre-commercialised NBT products (Parisi and Rodriguez Cerezo, 2021), the single target, or a reduced subset of the detected targets, allowing the rapid and cheap identification and quantification of each GMO. The package developed by the FELIX project companies would only be used by notifying companies or – very occasionally – when detection results are challenged, as proposed in Figure 2 (Bertheau, 2022).

Can the results of FELIX Gingko be used for GMO detection? The work presented was carried out by companies, some of which are service providers. They can also aim, according to the economic model of spin-offs, to be bought by a large company in order to pay the initial contributors of ideas (Kampers et al., 2021). Gingko Bioworks has bought several other companies in recent years. To be able to exploit the tools developed by Gingko in the long term, and at the lowest cost, it would be necessary to develop similar tools, but with the risk that patents could hinder the use of the techniques by enforcement laboratories, as was attempted in the early 2000s for PCR. In addition, Gingko Bioworks may find itself in a

[6] https://gmo-crl.jrc.ec.europa.eu/ENGLabs

[7] https://investors.ginkgobioworks.com/news/news-details/2022/Bayer-and-Ginkgo-Bioworks-close-deal-creating-Agricultural-Biologicals-Powerhouse/default.aspx

conflict of interest because of the contract with Bayer. Finally, the company could also be tempted to help manufacturers hide their NBT products.

It appears from the press conference that the companies have worked efficiently and quickly to develop the FELIX programme. No conceptual or technical obstacle has prevented European laboratories involved in GMO detection from developing such technical packages since 2007, when COGEM issued its report on NBT, or 2012, when the CRISPR-Cas technique was published. Given that the majority of ENGL laboratories maintain databases of GMO techniques, products, and sequences of all origins, a sufficiently funded EU research programme would have allowed similar results to be directly mobilised by various labs and firms.

As well as the initial database of signatures, all the concepts and techniques developed for the characterization of sequences produced from “related techniques” can be adapted to the detection of the signatures of the NBT techniques used, and thus to the identification of NBT products as such (Bertheau, 2019; 2022). This identification is supported by the parallel attribution of the genetically engineered organism to a specific source.

Conclusion

In summary, the FELIX project reinforces the demonstration that the characterisation and identification of organisms obtained through any genetic modification method are quite accessible if one has the means and political will. Companies will not fail to carry out these tasks in order to protect their patents and patented varieties through the matrix approach mentioned above. The multivariate analyses of this matrix approach are of the same kind as those used in the varieties’ identification techniques currently standardised by the International Organization for Standardization (ISO) and International Union for the Protection of New Varieties of Plants (UPOV). A few steps still need to be developed to obtain a subset or even a single target, that can be cost-effectively used in routine analyses performed for the detection and identification of NBT products.

The results presented above also pave the way for the attribution of specific genetic engineering techniques to NBT products, by addressing the signatures of NBT products instead of the scars of genetic modifications. Here too, it will be necessary to determine a target that can be used cheaply and routinely, such as the sequence corresponding to the claimed trait.

In conclusion, the brief presentation of the results – to be confirmed on real NBT products – of this part of the FELIX project shows that third (non-EU) countries are not only giving themselves the means to detect unknown potentially dangerous GMOs, but that they are also giving companies producing patented varieties the means to defend their patents. European seed producers and farmers could be sued for patent infringement, for example, following contamination by patented GMO seeds. Due to a lack of foresight on the part of European governance, European seed producers and farmers will be subject to the claims and decisions of foreign companies. A lack of foresight of the European governance further reduces the competitive advantages of European non-GMO products, which are important

for our exports of seeds and food commodities. This is why the application of the European GMO regulation to NBT products is fundamental, as it will enable the detection and traceability of NBT products by obliging companies to provide reference material and identification techniques that they can no longer deny being able to develop. This work contradicts the assertions of the JRC (Joint Research Centre, the Commission's science and knowledge service), the Commission, and some scientists that it is impossible to detect and identify genome-edited products.

As pointed out by a representative of IARPA, the technologies necessary for the results presented on 17 October exist in other countries. The European Commission prides itself on promoting innovation, but the delays in implementing detection and identification technologies have been considerable since the meetings of European experts in detection that followed the events of September 11. As I pointed out to a European official of the JRC during a conference under the aegis of various European parliamentarians, it is time for EURL-GMFF, ENGL, and the JRC to get out of their comfort zone of transgenic GMOs with standardised methods and to move forward by studying without preconceived ideas the facts concerning the accuracy of genome editing methods, as well as the concepts and tools that need to be developed both for the identification of NBT products and for routine analyses. It is time for the proponents of GMOs to stop spreading false information so that European farmers and consumers can produce and consume the products they want.

References

Akella, S., Ma, X., Bacova, R., Harmer, Z.P., Kolackova, M., Wen, X., Wright, D.A., Spalding, M.H., Weeks, D.P., and Cerutti, H. (2021). Co-targeting strategy for precise, scarless gene editing with CRISPR/Cas9 and donor ssODNs in Chlamydomonas. Plant Physiol. *187*, 2637-2655. 10.1093/plphys/kiab418.

Alexandrov, L.B., Kim, J., Haradhvala, N.J., Huang, M.N., Tian Ng, A.W., Wu, Y., Boot, A., Covington, K.R., Gordenin, D.A., Bergstrom, E.N., et al. (2020). The repertoire of mutational signatures in human cancer. Nature *578*, 94-101. 10.1038/s41586-020-1943-3.

Alley, E.C., Turpin, M., Liu, A.B., Kulp-McDowall, T., Swett, J., Edison, R., Von Stetina, S.E., Church, G.M., and Esvelt, K.M. (2020). A machine learning toolkit for genetic engineering attribution to facilitate biosecurity. Nat. Commun. *11*, 6293. 10.1038/s41467-020-19612-0.

Bensadoun, A., Monod, H., Makowski, D., and Messéan, A. (2016). A Bayesian approach to model dispersal for decision support. Environ. Modell. Softw. *78*, 179-190. http://dx.doi.org/10.1016/j.envsoft.2015.12.018.

Bertheau, Y. (2019). New breeding techniques: detection and identification of the techniques and derived products. In Encyclopedia of Food Chemistry, Reference Module in Food Science, R.H. Stadler, ed. (Elsevier), pp. 320-336. https://doi.org/10.1016/B978-0-08-100596-5.21834-9.

Bertheau, Y. (2022). Advances in identifying GM plants. Toward the routine detection of "hidden" and "new" GMOs. In Developing smart-agrifood supply chains: using technology to improve safety and quality, L. Manning, ed. (Burleigh Dodds Science Publishing), pp. 87-150. http://dx.doi.org/10.19103/AS.2021.0097.22.

Bohanec, M., Bertheau, Y., Brera, C., Gruden, K., Holst-Jensen, A., Kok, E.J., Lécroart, B., Messéan, A., Miraglia, M., Onori, R., et al. (2013). The Co-Extra decision support system: a model-based integration of project results. In Genetically modified and non-genetically modified food supply chains: co-existence and traceability, Y. Bertheau, ed. (Wiley-Blackwell), pp. 461-490.

Bohanec, M., Boshkoska, B.M., Prins, T.W., and Kok, E.J. (2016). SIGMO: a decision support system for identification of genetically modified food or feed products. Food Control *71*, 168-177. 10.1016/j.foodcont.2016.06.032.

Budowle, B., Johnson, M.D., Fraser, C.M., Leighton, T.J., Murch, R.S., and Chakraborty, R. (2005). Genetic analysis and attribution of microbial forensics evidence. Critical Reviews in Microbiology *31*, 233-254. 10.1080/10408410500304082.

Cao, Y., Oh, J., Xue, M., Huh, W.J., Wang, J., Gonzalez-Hernandez, J.A., Rice, T.A., Martin, A.L., Song, D., Crawford, J.M., et al. (2022). Commensal microbiota from patients with inflammatory bowel disease produce genotoxic metabolites. Science *378*, eabm3233. doi:10.1126/science.abm3233.

Chevalier, A., Yang, S., Khurshid, Z., Sahelijo, N., Tong, T., Huggins, J.H., Yajima, M., and Campbell, J.D. (2021). The Mutational Signature Comprehensive Analysis Toolkit (musicatk) for the Discovery, Prediction, and Exploration of Mutational Signatures. Cancer Research *81*, 5813-5817. 10.1158/0008-5472.Can-21-0899.

Crook, O.M., Warmbrod, K.L., Lipstein, G., Chung, C., Bakerlee, C.W., Jr., T.G.M., Holland, S.R., Swett, J.L., Esvelt, K.M., Alley, E.C., and Bradshaw, W.J. (2021). Analysis of the first Genetic Engineering Attribution Challenge. 10.48550/arxiv.2110.11242.

Cuevas-Ramos, G., Petit, C.R., Marcq, I., Boury, M., Oswald, E., and Nougayrède, J.-P. (2010). *Escherichia coli* induces DNA damage in vivo and triggers genomic instability in mammalian cells. Proceedings of the National Academy of Sciences *107*, 11537-11542. doi:10.1073/pnas.1001261107.

Degasperi, A., Zou, X., Amarante, T.D., Martinez-Martinez, A., Koh, G.C.C., Dias, J.M.L., Heskin, L., Chmelova, L., Rinaldi, G., Wang, V.Y.W., et al. (2022). Substitution mutational signatures in whole-genome-sequenced cancers in the UK population. Science *376*, abl9283. doi:10.1126/science.abl9283.

Dobnik, D., Gruden, K., Zel, J., Bertheau, Y., Holst-Jensen, A., and Bohanec, M. (2018). Decision support for the comparative evaluation and selection of analytical methods: detection of Genetically Modified Organisms as an example. Food Analytical Methods *11*, 2105-2122. 10.1007/s12161-018-1194-1.

Domingo-Prim, J., Bonath, F., and Visa, N. (2020). RNA at DNA Double-Strand Breaks: The Challenge of Dealing with DNA:RNA Hybrids. BioEssays *42*, 1900225. https://doi.org/10.1002/bies.201900225.

Elison, G.L., and Acar, M. (2018). Scarless genome editing: progress towards understanding genotype-phenotype relationships. Curr Genet *64*, 1229-1238. 10.1007/s00294-018-0850-8.

Ellison, E.E., Chamness, J.C., and Voytas, D.F. (2021). Viruses as vectors for the delivery of gene-editing reagents. In Viruses as vectors for the delivery of gene editing reagents, E.E. Ellison, J.C. Chamness, and D.F. Voytas, eds. (Burleigh Dodds Publishing), pp. 1-36. http://dx.doi.org/10.19103/AS.2020.0082.28.

Ellison, E.E., Nagalakshmi, U., Gamo, M.E., Huang, P.-j., Dinesh-Kumar, S., and Voytas, D.F. (2020). Multiplexed heritable gene editing using RNA viruses and mobile single guide RNAs. Nature Plants *6*, 620-624. 10.1038/s41477-020-0670-y.

Fels, U., Gevaert, K., and Van Damme, P. (2020). Bacterial Genetic Engineering by Means of Recombineering for Reverse Genetics. Frontiers in Microbiology *11*. 10.3389/fmicb.2020.548410.

Gulhan, D.C., Lee, J.J.-K., Melloni, G.E.M., Cortés-Ciriano, I., and Park, P.J. (2019). Detecting the mutational signature of homologous recombination deficiency in clinical samples. Nature Genetics *51*, 912-919. 10.1038/s41588-019-0390-2.

Holst-Jensen, A., Berdal, K.G., Bertheau, Y., Bohanec, M., Bohlin, J., Chaouachi, M., Gruden, K., Hamels, S., Kok, E.J., Krech, A., et al. (2013). Towards detection of unknown GMOs. In Genetically modified and non-genetically modified food supply chains: co-existence and traceability, Y. Bertheau, ed. (Wiley-Blackwell), pp. 367-382. 10.1002/9781118373781.ch20.

Hurel, J., Schbath, S., Bougeard, S., Rolland, M., Petrillo, M., and Touzain, F. (2020). DUGMO: tool for the detection of unknown genetically modified organisms with high-throughput sequencing data for pure bacterial samples. BMC Bioinformatics *21*, 284. 10.1186/s12859-020-03611-5.

Kampers, L.F.C., Asin-Garcia, E., Schaap, P.J., Wagemakers, A., and Martins dos Santos, V.A.P. (2021). From Innovation to Application: Bridging the Valley of Death in Industrial Biotechnology. Trends in Biotechnology *39*, 1240-1242. https://doi.org/10.1016/j.tibtech.2021.04.010.

Kunjapur, A.M., Pfingstag, P., and Thompson, N.C. (2018). Gene synthesis allows biologists to source genes from farther away in the tree of life. Nat. Commun. *9*, 4425. 10.1038/s41467-018-06798-7.

Lemercier, C. (2014). Les infections bactériennes vues du génome eucaryote. Med Sci (Paris) *30*, 758-764.

Lewis, G., Jordan, J.L., Relman, D.A., Koblentz, G.D., Leung, J., Dafoe, A., Nelson, C., Epstein, G.L., Katz, R., Montague, M., et al. (2020). The biosecurity benefits of genetic engineering attribution. Nat. Commun. *11*, 6294. 10.1038/s41467-020-19149-2.

Nielsen, A.A.K., and Voigt, C.A. (2018). Deep learning to predict the lab-of-origin of engineered DNA. Nat. Commun. *9*, 3135. 10.1038/s41467-018-05378-z.

Omichessan, H., Severi, G., and Perduca, V. (2019). Computational tools to detect signatures of mutational processes in DNA from tumours: A review and empirical comparison of performance. PLOS ONE *14*, e0221235. 10.1371/journal.pone.0221235.

Pacheco-Barcia, V., Muñoz, A., Castro, E., Ballesteros, A.I., Marquina, G., González-Díaz, I., Colomer, R., and Romero-Laorden, N. (2022). The Homologous Recombination Deficiency Scar in Advanced Cancer: Agnostic Targeting of Damaged DNA Repair. Cancers *14*, 2950.

Parisi, C., and Rodriguez Cerezo, E. (2021). Current and future market applications of new genomic techniques. Joint Research Centre. https://publications.jrc.ec.europa.eu/repository/handle/JRC123830

https://publications.jrc.ec.europa.eu/repository/bitstream/JRC123830/210419_ngt_market_study_final.pdf

https://op.europa.eu/en/publication-detail/-/publication/8940fa16-a17e-11eb-b85c-01aa75ed71a1/language-en

Perner, J., Abbas, S., Nowicki-Osuch, K., Devonshire, G., Eldridge, M.D., Tavaré, S., and Fitzgerald, R.C. (2020). The mutREAD method detects mutational signatures from low quantities of cancer DNA. Nat. Commun. *11*, 3166. 10.1038/s41467-020-16974-3.

Reisch, C.R., and Prather, K.L.J. (2015). The no-SCAR (Scarless Cas9 Assisted Recombineering) system for genome editing in Escherichia coli. Scientific reports *5*, 15096. 10.1038/srep15096 http://www.nature.com/articles/srep15096#supplementary-information.

Rempel, E., Kluck, K., Beck, S., Ourailidis, I., Kazdal, D., Neumann, O., Volckmar, A.L., Kirchner, M., Goldschmid, H., Pfarr, N., et al. (2022). Pan-cancer analysis of genomic scar patterns caused by homologous repair deficiency (HRD). npj Precision Oncology *6*, 36. 10.1038/s41698-022-00276-6.

Scully, R., Panday, A., Elango, R., and Willis, N.A. (2019). DNA double-strand break repair-pathway choice in somatic mammalian cells. Nat Rev Mol Cell Biol *20*, 698-714. 10.1038/s41580-019-0152-0.

Slezak, T., Hart, B., and Jaing, C. (2020). Chapter 20 - Design of genomic signatures for pathogen identification and characterization. In Microbial Forensics (Third Edition), B. Budowle, S. Schutzer, and S. Morse, eds. (Academic Press), pp. 299-312. https://doi.org/10.1016/B978-0-12-815379-6.00020-9.

Slezak, T., Lenhoff, R., Allen, J., Borucki, M., Vitalis, E., and Gardner, S. (2009). TMTI Task 1.6 Genetic Engineering Methods and Detection. https://www.osti.gov/biblio/972121

https://www.osti.gov/servlets/purl/972121.

Steyer, B., Cory, E., and Saha, K. (2018). Developing precision medicine using scarless genome editing of human pluripotent stem cells. Drug Discovery Today: Technologies *28*, 3-12. https://doi.org/10.1016/j.ddtec.2018.02.001.

van der Wiel, A.M.A., Schuitmaker, L., Cong, Y., Theys, J., Van Hoeck, A., Vens, C., Lambin, P., Yaromina, A., and Dubois, L.J. (2022). Homologous Recombination Deficiency Scar: Mutations and Beyond' Implications for Precision Oncology. Cancers *14*, 4157. 10.3390/cancers14174157.

Wang, Q., Kille, B., Liu, T.R., Elworth, R.A.L., and Treangen, T.J. (2021). PlasmidHawk improves lab of origin prediction of engineered plasmids using sequence alignment. Nat. Commun. *12*, 1167. 10.1038/s41467-021-21180-w.

Wang, Z., Zhang, T., Wu, W., Wu, L., Li, J., Huang, B., Liang, Y., Li, Y., Li, P., Li, K., et al. (2022). Detection and Localization of Solid Tumors Utilizing the Cancer-Type-Specific Mutational Signatures. Frontiers in Bioengineering and Biotechnology *10*. 10.3389/fbioe.2022.883791.

Watkins, J.A., Irshad, S., Grigoriadis, A., and Tutt, A.N. (2014). Genomic scars as biomarkers of homologous recombination deficiency and drug response in breast and ovarian cancers. Breast Cancer Res *16*, 211. 10.1186/bcr3670.

Wong, J.K.L., Aichmüller, C., Schulze, M., Hlevnjak, M., Elgaafary, S., Lichter, P., and Zapatka, M. (2022). Association of mutation signature effectuating processes with mutation hotspots in driver genes and non-coding regions. Nat. Commun. *13*, 178. 10.1038/s41467-021-27792-6.

Zou, X., Owusu, M., Harris, R., Jackson, S.P., Loizou, J.I., and Nik-Zainal, S. (2018). Validating the concept of mutational signatures with isogenic cell models. Nat. Commun. *9*, 1744. 10.1038/s41467-018-04052-8.