Empowering Genome Editing Through Standards

Samantha Maragh, Ph.D.

National Institute of Standards and Technology

A revolution is underway to reengineer the blueprint for life. The blueprint for life of every organism is the genetic code, whose sequence encodes identity and function. The genome (expressed in DNA base pairs) is the entire complement of an organism's genetic code, and is housed within the basic functional unit of life, the cell. This revolutionary field, called genome engineering, encompasses tools and techniques programmed to target a specific sequence within a genome and alter the genetic code (termed genome editing) or alter the chemical signatures associated with the genetic code. This technology operates by biochemical principles generally applicable to every kind of cell (Carroll 2014, Kim and Kim 2014). Genome editing is a highly pursued form of genome engineering to generate engineered/edited cells that are functional and have a permanent change to the genetic code (Doudna and Charpentier 2014, Gaj, Sirk et al. 2016). A generalized genome editing process can be described as: (1) Determine target location in the genome. (2) Design the editing system to bind to the intended target location (e.g., Zinc Finger Nucleases (ZFNs), Transcription activator-like effector nucleases (TALENs) and clustered regularly interspaced short palindromic repeats (CRISPR) with CRISPR associated proteins (CRISPR-Cas)) (Sander and Joung 2014, Urnov 2018). (3) Formulate and deliver the editing system into live cells. Once delivered into a cell, these systems find and bind to the target location in the genome and damage the DNA. This DNA damage gets recognized and repaired by the cell either perfectly, restoring the original genomic sequence, or imperfectly where one or more bases of sequence are changed, deleted, or inserted (Komor, Badran et al. 2017). Genome editing is achieved as a result of imperfect DNA break repair. (4) Confirm if and where the genome sequence has been changed, what sequence change has resulted, and the relative frequency of that change. (5) Determine if your engineered cells are fit for your purpose. Most often,

genome editing is performed directly on the cells of interest in a controlled laboratory setting, but for complex organisms like humans, editing systems may need to be delivered directly into the organism (*in vivo* editing), in which case the editing molecules must get to the tissues or cells that they are intended to edit, enter those cells, and appropriately edit the genomes of those cells (Maeder and Gersbach 2016, Yin, Kauffman et al. 2017).

Genome engineering is being pursued globally by government, academic, and private sectors to transform medicine and bioscience to enable previously impossible advances in areas such as basic biology research, gene therapy, synthetic biology, novel antimicrobial and antivirals, biomanufacturing, agriculture, and food production (Barrangou and Doudna 2016). In some human diseases, just a handful of incorrect letters in the genetic code or as little as a single incorrect letter at a specific position (e.g., Sickle Cell disease) of the ~6.6 billion letter human genetic code can cause a serious and/or deadly disease. Now, genome engineering proposes the first era of technology where the medical field isn't limited to solely managing symptoms and treating illness episodes, but where the cells of a patient could be engineered to 'fix' their disease at the genetic code level. The biomedical field is daring to think and even speak the word 'cure' for so many people that had few or no options previously (Porteus 2015, Fellmann, Gowen et al. 2017, Salsman, Masson et al. 2017). In addition, the agriculture industry has begun targeted engineering of crop genomes to have more favorable traits, better yield, and to be more disease resistant (Mao, Zhang et al. 2013, Yin, Gao et al. 2017). There are also large efforts within the environmental science community to engineer microbes with new abilities to produce biofuels, and act as biosensors (Ng, Tan et al. 2017).

The Edit's in the details – challenges and opportunities for standards

Concurrently with the wide global pursuit to leverage existing genome editing systems, there is an incredible level of innovation taking place towards new and expanded capabilities for high precision

targeting of any genomic sequence, to make any desired change, in any cell. The transformative capabilities of genome engineering are already beginning to be realized; however, there are crucial measurement challenges to address to facilitate transferring these technologies into trusted data and products. An editing process is carried out on many, often thousands to millions, of cells at a time and currently there is little technical control over the efficiency of editing and which sequence changes result. Each cell within this pool is a closed system where the resulting edits in each cell are independent events. Due to current technical limitations on measuring the sequence of individual cells at high throughput, edited sequence confirmation is a bulk measurement sampling of the genomes from this heterogeneously edited cell pool. In addition, due to the biological limitation, particularly for human therapeutics, this heterogeneous pool of cells may be your final product. It is also technically challenging to accurately parse the sequencing data from these bulk analyses since the number of edits detected can range into the tens or even hundreds for one genomic location. Bioinformatics pipelines to parse this data were benchmarked on what is observed in nature that only a handful, if that many, variants are expected to occur at any one location. It is therefore unclear how accurately sequencing analysis pipelines report on what is biologically present in the sample.

A prominent measurement challenge for human therapeutic genome editing is the issue of off-target editing (Fu, Foden et al. 2013, Tsai, Zheng et al. 2015). Off-target editing is when a sequence change(s) occurs at a site other than the intended target site. The concern is that off-target editing may change the cell in a way that makes the therapy or product unsuitable or unsafe, even in the presence of the desired edit at the intended target site. To assess this, the entire genome, or at least sites where there is reason to think off-target edits could occur, is sequenced after an editing process. There are technical limitations to sequencing the entirety of large genomes like the human genome with sufficient sensitivity to report any off-target edit at any location within the heterogeneous sample pool. Means of limiting where to sequence for off-target edits have and are continuing to be developed, but there is very

limited understanding of how accurate these tools are. With reliable sequence data in hand, there's still the challenge of interpreting the significance of unintended edits and understanding if edited cells are fit for the intended purpose. Ultimately a genome edited product must be manufactured. Traditional manufacturing approaches don't directly translate to this field where the input is live cells, which must be manipulated precisely and still be live and functional at the end of the process (Harrison, Ruck et al. 2017, Harrison, Ruck et al. 2018). Moreover, the product may be a personalized therapy for a patient with short storage life, therefore requiring small batch manufacturing with rapid distribution and use.

Figure 1: Overview of the types and roles of standards.

Standards play an essential role in the translation and durable adoption of technology (Figure 1) (Plant, Locascio et al. 2014, Plant, Becker et al. 2018). Standards within the genome engineering community will empower: innovation, technology adoption, evidence that new biological understanding is based on sound data, and that products generated with these technologies are suitable and "safe". Standards in the form of traceable materials can help address the challenges of having confidence a process can reliably report: where editing occurred, what edit(s) resulted, and the relative abundance of each edit. For this purpose, traceable material or control samples in the form of a well-qualified series of

cells or genomes containing a variety of edits at known relative abundance across the genome can serve as ground truth samples for assessing sequencing methods. For the genome engineering community, these control materials would enable comparability within and between: operators at the same site, operators at different sites or organizations, and state of the art vs. new innovative sequencing methods. Standards in the form of standard data sets and metadata norms can help address the challenges of understanding accuracy of sequencing data analysis pipelines. Shared standard datasets along with the associated metadata detailing the editing and data handling process will provide a means to benchmark data analysis pipelines, compare performance between iterations of a pipeline, and compare performance of different pipelines (Figure 2). Supporting these technical standards is the need for standardized definitions of key terms within the genome editing field enabling clear communication of results within the field and to regulatory agencies. The U.S. National Institute of Standards and Technology has launched an effort to work across the genome editing community to address developing the above standards. Standards for how to safely and efficiently manufacture engineered cell products will likely require a paradigm shift and disruptive technologies to address the particular challenges of manufacturing living cells.

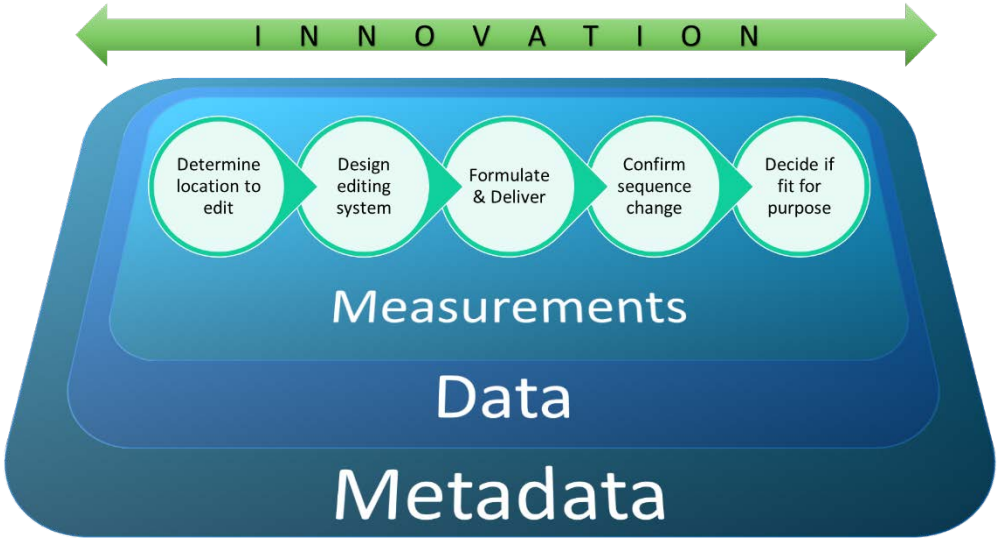

Figure 2: Genome editing process overview in context of opportunities for standardization.

Closing Thoughts

Translating the promise of genome engineering into production and medical practice requires robust quantitative assays, accurate data tools, and the associated standards and benchmarks to enable high confidence characterization of engineered genomes and cells. Steps towards addressing some of these needs through standards (physical controls, standard datasets and a standard lexicon) for this field are underway. Genome engineering technology that is not reliant on damaging the genome or changing the sequence at all is rapidly progressing (Thakore, Black et al. 2016, Liao, Hatanaka et al. 2017). Further progress can be made through developing a suite of tools and technology employing a multidisciplinary approach to address unmet measurement need such as single cell editing detection and in *vivo* tracking and monitoring of edited cells that have been introduced into the environment or a live organism (e.g., cells introduced into a human as a therapy). As the genome engineering field matures, there will need to be continuous evaluation of which new standards and norms can support rapid innovation and expansion of this field.

References

Barrangou, R. and J. A. Doudna (2016). "Applications of CRISPR technologies in research and beyond." Nat Biotechnol **34**(9): 933-941.

Carroll, D. (2014). "Genome engineering with targetable nucleases." Annu Rev Biochem **83**: 409-439.

Doudna, J. A. and E. Charpentier (2014). "Genome editing. The new frontier of genome engineering with CRISPR-Cas9." Science **346**(6213): 1258096.

Fellmann, C., B. G. Gowen, P. C. Lin, J. A. Doudna and J. E. Corn (2017). "Cornerstones of CRISPR-Cas in drug discovery and therapy." Nat Rev Drug Discov **16**(2): 89-100.

Fu, Y., J. A. Foden, C. Khayter, M. L. Maeder, D. Reyon, J. K. Joung and J. D. Sander (2013). "High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells." Nat Biotechnol **31**(9): 822-826.

Gaj, T., S. J. Sirk, S. L. Shui and J. Liu (2016). "Genome-Editing Technologies: Principles and Applications." Cold Spring Harb Perspect Biol **8**(12).

Harrison, R. P., S. Ruck, N. Medcalf and Q. A. Rafiq (2017). "Decentralized manufacturing of cell and gene therapies: Overcoming challenges and identifying opportunities." Cytotherapy **19**(10): 1140-1151.

Harrison, R. P., S. Ruck, Q. A. Rafiq and N. Medcalf (2018). "Decentralised manufacturing of cell and gene therapy products: Learning from other healthcare sectors." Biotechnol Adv **36**(2): 345-357.

Kim, H. and J. S. Kim (2014). "A guide to genome engineering with programmable nucleases." Nat Rev Genet **15**(5): 321-334.

Komor, A. C., A. H. Badran and D. R. Liu (2017). "CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes." Cell **168**(1-2): 20-36.

Liao, H. K., F. Hatanaka, T. Araoka, P. Reddy, M. Z. Wu, Y. Sui, T. Yamauchi, M. Sakurai, D. D. O'Keefe, E. Nunez-Delicado, P. Guillen, J. M. Campistol, C. J. Wu, L. F. Lu, C. R. Esteban and J. C. Izpisua Belmonte (2017). "In Vivo Target Gene Activation via CRISPR/Cas9-Mediated Trans-epigenetic Modulation." Cell **171**(7): 1495-1507 e1415.

Maeder, M. L. and C. A. Gersbach (2016). "Genome-editing Technologies for Gene and Cell Therapy." Mol Ther **24**(3): 430-446.

Mao, Y., H. Zhang, N. Xu, B. Zhang, F. Gou and J. K. Zhu (2013). "Application of the CRISPR-Cas system for efficient genome engineering in plants." Mol Plant **6**(6): 2008-2011.

Ng, I. S., S. I. Tan, P. H. Kao, Y. K. Chang and J. S. Chang (2017). "Recent Developments on Genetic Engineering of Microalgae for Biofuels and Bio-Based Chemicals." Biotechnol J **12**(10).

Plant, A. L., C. A. Becker, R. J. Hanisch, R. F. Boisvert, A. M. Possolo and J. T. Elliott (2018). "How measurement science can improve confidence in research results." PLoS Biol **16**(4): e2004299.

Plant, A. L., L. E. Locascio, W. E. May and P. D. Gallagher (2014). "Improved reproducibility by assuring confidence in measurements in biomedical research." Nat Methods **11**(9): 895-898.

Porteus, M. H. (2015). "Towards a new era in medicine: therapeutic genome editing." Genome Biol **16**: 286.

Salsman, J., J. Y. Masson, A. Orthwein and G. Dellaire (2017). "CRISPR/Cas9 Gene Editing: From Basic Mechanisms to Improved Strategies for Enhanced Genome Engineering In Vivo." Curr Gene Ther **17**(4): 263-274.

Sander, J. D. and J. K. Joung (2014). "CRISPR-Cas systems for editing, regulating and targeting genomes." Nat Biotechnol **32**(4): 347-355.

Thakore, P. I., J. B. Black, I. B. Hilton and C. A. Gersbach (2016). "Editing the epigenome: technologies for programmable transcription and epigenetic modulation." Nat Methods **13**(2): 127-137.

Tsai, S. Q., Z. Zheng, N. T. Nguyen, M. Liebers, V. V. Topkar, V. Thapar, N. Wyvekens, C. Khayter, A. J. Iafrate, L. P. Le, M. J. Aryee and J. K. Joung (2015). "GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases." Nat Biotechnol **33**(2): 187-197.

Urnov, F. D. (2018). "Genome Editing B.C. (Before CRISPR): Lasting Lessons from the "Old Testament"." CRISPR J **1**: 34-46.

Yin, H., K. J. Kauffman and D. G. Anderson (2017). "Delivery technologies for genome editing." Nat Rev Drug Discov **16**(6): 387-399.

Yin, K., C. Gao and J. L. Qiu (2017). "Progress and prospects in plant genome editing." Nat Plants **3**: 17107.