Roundtable Session 2 – Table 3 – Can AI Transform CGT Process Development?

Facilitator: Krzysztof Krawczyk, *F. Hoffmann -La Roche AG*

Scribe: Ane Quesada, *Viralgen*

Abstract:

While the term "AI" dominates today's headlines, the underlying concepts have been operating in process development under various names, such as multivariate data analysis, machine learning, and expert systems, for many decades.

Today, however, the paradigm is shifting from merely monitoring a process to actively designing it. Generative AI can navigate complex parameter spaces that traditional Design of Experiments (DoE) cannot easily solve, while digital twins combine mechanistic equations with machine learning to simulate cell growth, optimize transfection efficiency, and predict purification outcomes. Simultaneously, Large Language Models (LLMs) can ingest large numbers of unstructured documents - legacy batch records, deviation reports, shift logs, equipment manuals, and scientific literature, to facilitate root-cause analysis.

Yet, applying it in cell and gene therapy (CGT) process development presents severe operational hurdles. AI models thrive on large, standardized datasets that are rarely available in low-run, high-variability CGT processes. Meanwhile, "black-box" predictions and hallucination risks complicate defining robust control strategies. Finally, in a sector where "the process equals the product," integrating third-party AI introduces critical data confidentiality and IP risks. The recent controversy where private mathematical drafts on the Navier-Stokes problem allegedly influenced frontier model training serves as a stark cautionary tale for biopharma.

This roundtable discussion will explore whether AI can deliver transformational benefits to CGT process development today despite these challenges.

Questions:

1. How can teams train reliable models when constrained by small data sets and high donor heterogeneity?

2. Is the concept of a hybrid digital twin a viable option in CGT process development?

3. How do we establish a defendable design space using non-explainable ("black box") algorithms?

4. How can biopharma leverage external AI capabilities while safeguarding proprietary process trade secrets?

Notes:

1. Can AI fundamentally transform process development?

Define what we mean wit hAI: LL? Advanced statistic models? Machine learning?

Machine learning has been applied for years already in pharma, such as in epitope design.

LLM are maybe better suited to extract data from largaqr number of batches

A critical point that was raised was transparency: to be able to understand how the decisions and actions have been taken, understand its error and weaknesses. If AI is a black box is where the risk is, the regulator cannot review it or challenge it if it is a black box.

One approach to overcome ths is to build the model through an iterative process: building the model slowly and checking the optputs each time. Going back, reviewing it, continue building it: it is a slower but more secure approach. The objective is to have a model you can trust and understand how was built and with which controls.

Development is still needed for the implementation of AI for CMC: current AI guesses, skips steps, the output is my times not reliable, consistent or reproducible.

A big risk is to start accepting AI intuitive leaps as truth. It is key to understand which assumptions the AI has made, which references it has used, and wich steps it has skipped.

Digital twins: can be very useful to reduce experimental work at the stage of process development. AI can be applied to large amounts of data to define experimental work and reduce workload. The outputs have to be confirmed with experimental data.

Keay topic: reproducibility: ensure the results of the model you are using are reproducible and are not variable each time you run the simulations. Use of AI

to select for donors can include in the selection bias (ex. High socioeconomic status if it filters for reliability to get to appointments) that might be overlooked.

Key point: ensure the data that is using is homogeneous, the AI might treat different types of data as the same, when in reality they are not comparable (or example, results obtained with an analytical method that has been changing overtime).

To use LLM for the generation of CMC documentation and risk assessments: the regulators will review the documentation as if it was the sponsors, so the risk is assumed by the sponsor.