

EXPERT INSIGHT

Streamlining data management & process analytics for the manufacturing of cell & gene therapies

**Sébastien de Bournonville, Toon Lambrechts,
Thomas Pinna, Ioannis Papantoniou
& Jean-Marie Aerts**

With a likely increase in production volumes of cell and gene therapies, and the more widespread implementation of Quality by Design principles, the amount of process data to manage (i.e. collect, validate, analyze and interpret) will increase significantly. Data guides the manufacturing operations, it allows monitoring and control of the process, and it assures product quality and regulatory compliance. More recently big data, artificial intelligence and machine learning techniques promise a surge in data-driven insights, affecting nearly all aspects of R&D and manufacturing of cell and gene therapies. However, without proper tools and solid data foundations, data management will constitute a bottleneck in the production of these novel therapies. The aim of this article is to discuss steps needed in order to address the data management challenge.

Submitted: 25th September 2018 ► Published: xxx

THE DATA MANAGEMENT CHALLENGE

The traction for cell and gene therapies is fueled by encouraging clinical results and the involvement of

high profile pharmaceutical companies in the development, manufacturing and commercialization, such as Takeda's license for commercialization of Aloficel (developed by TiGenix), Celgene's acquisition of

Juno Therapeutics or Gilead's acquisition of Kite Pharma. It is expected that in coming years more potent and potentially more complex therapies will be reaching the market, and with these, the number

of treated patient populations will increase in size and complexity. This will not only apply additional pressure on the current manufacturing strategies and tools [1–3], but will also require improved management of process and regulatory data.

Together with data management, the need for enhanced data analytics is increasing rapidly as well. Using variable biology as a pretext to legitimize variable end products does not fit with the ambition of a maturing industry like the cell and gene therapy industry of which end users and investors expect consistent quality and effectiveness of its products. It reveals that the product, or at least the definition of the target product profile or stratification of the patient population, and/or the manufacturing processes are not yet understood completely to the extent that each patient can be treated effectively. This was recently illustrated by product specification complications in the manufacturing process of Novartis' Kymriah (tisagenlecleucel) [4,5]. The urge to manage process variability motivates the sector to adopt QbD (Quality by Design) principles under which the process conditions might vary (within validated limits, e.g. to compensate for differences in the starting cell material), but where

the final product and its effect in the patient are robust and reproducible [6]. Ultimately, the QbD principles allow to move away from 'the product is the process' paradigm that calls for dogmatic, static recipe type of manufacturing procedures. Instead, more agile manufacturing processes can be designed by adopting adaptive data-driven process controllers which make use of sensors and other analytical technologies that help to detect variation quickly and act upon those variations in real-time [7].

The increase in production volume, the adoption of QbD principles and more agile production processes lead to a direct increase of data to be collected, validated, analyzed and interpreted. While this should happen without slowing down the manufacturing processes, the lack of standardized and streamlined data management tools can cause a decrease in efficiency. The data management challenge is even more obvious for autologous therapies and decentralized manufacturing strategies. In these cases, an increase in production volume will always lead to a direct increase in data collection and analytics (in-process and release testing data, batch records, supply chain and sample tracking data, audit trails, comparability, etc.) (Figure 1).

► FIGURE 1 (OVERLEAF)

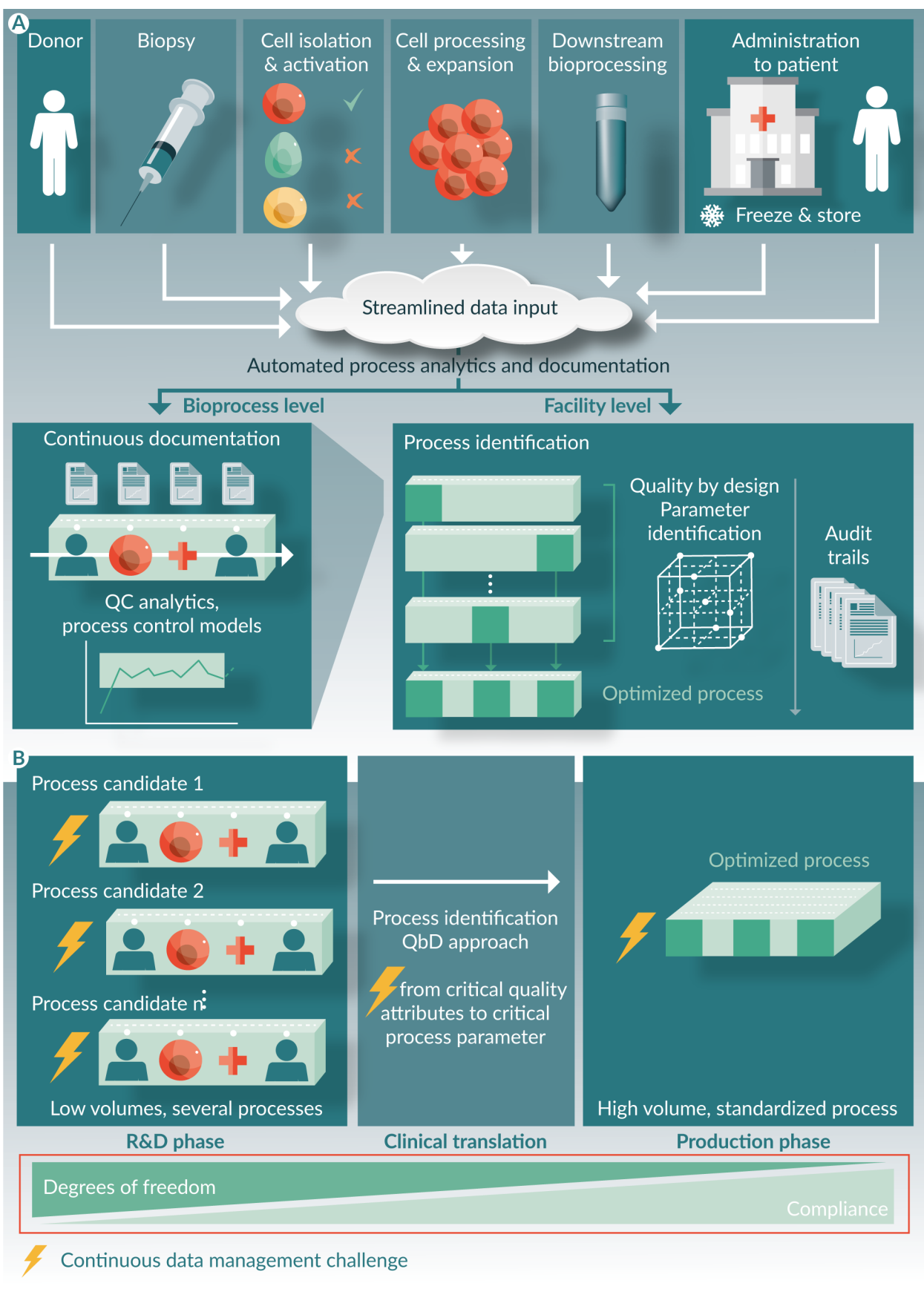
Visual representation of the data management challenge in cell & gene therapy manufacturing.

A) Typical steps of a cell or gene therapy manufacturing process. The manufacturing of cell and gene products is a complex sequence of manual and/or automated operations in a strictly regulated environment that therefore requires thorough process documentation and data collection at every step. The green arrows indicate data streams take place at all levels of the manufacturing process. Data management platforms can streamline the collection and analytics of those data and (automatically) provide valuable output (black arrows) at different organizational levels.

B) Typical phases required to commercialize a cell or gene therapy, from R&D to commercialization. The collected data forms the basis of the quality and compliance of the product, and it guides both the development and optimization of the production process. The data management challenge is present throughout the evolution from R&D to commercial production. During this evolution the focus changes from collecting a lot of diverse data that is required to optimize the process itself, to high volumes of more uniform data that is required to assure product quality and safety.

► **FIGURE 1**

Visual representation of the data management challenge in cell & gene therapy manufacturing.



STREAMLINING WHICH DATA TO COLLECT

While manufacturers of cell and gene therapies need to be able to collect a vast amount of data, the key is to collect only those data meaningful to the product under consideration, in the sense that it can be used to control the Critical Quality Attributes (CQAs). While for every production process there will be discussions on exactly which data to collect, it's imperative to plan ahead since in later clinical phases QC/QA testing becomes extremely expensive. Moreover, the easiest data to collect and store is potentially not the most ideal data to drive major process improvements or it cannot be linked with CQAs directly. The challenge is not necessarily about collecting as much data as possible, but rather to be able to collect all data that is meaningful to product quality, potency, safety and consistency.

The first aim in the collection of meaningful data should be to reduce the signal to noise ratio in the collected data, to make sure that the link between potentially critical process parameters (CPPs) and the product quality is not masked by variability. Reducing variability is discussed more often in literature and the main points to consider are using sensitive readouts, automating the process where possible and applying extreme standardization everywhere else, both on the direct handling of the production process itself as well as the documenting process [8-10]. Since most cell-based therapies are originating from research labs, most processes can be made leaner during scale-up, and thereby removing some sources of variability altogether. A typical

example would be the use of the recombinant enzyme TrypLE instead of trypsin for cell detachment as often it allows to eliminate a wash step before cell harvest and makes the timing of incubation steps less critical [11]. Additionally, care should be taken for variability that is imported via the suppliers of raw materials, for example a change in cell growth kinetics due to a different lot of serum, or an antibody lot dependent result in flow cytometry data.

The second aim in the collection of meaningful data is to pinpoint the minimum set of data that is required to actively control the CQAs. Before that, one has to figure out 1) what physical mechanisms are able to influence (the variability on) the CQAs, 2) which of these mechanisms can be actively controlled by adapting process parameters and 3) which techniques are available for on-line and real-time quantification of these influential factors. This quantification can be done via direct measurements or via a proxy-measure where (a combination of multiple) easy to measure process parameters are directly linked to the critical process parameters of interest via a data-based mathematical model. See [12] as an example where variations in oxygen concentrations were used as an on-line measure for the number of cells in a bioreactor, or [13] where variations in the lactate production rate of the cells were used to determine the end of the culture process. QbD tools such as Design of Experiment (DOE) and proper risk assessment (e.g. to determine which process parameters might need less stringent control) are indispensable to understand how these critical

process parameters (CPPs) relate to the CQAs of the product.

STREAMLINING DATA COLLECTION & MANAGEMENT

While the reduction in process variability and the determination of CPPs and CQAs are process dependent and therefore require a case by case approach [9], the data management and analytical requirements are more similar in nature for all cell and gene therapy manufacturing processes. However, very often manufacturers are still dependent on data management tools that are (partially) paper-based, require a lot of manual user input, involve operator-dependent decisions (e.g. “are the cells confluent and ready for harvest, or do we wait one more day?”) and require a significant amount of effort and oversight from the QA unit to remain compliant with regulation.

Additionally, the collected data often ends up in isolated data silos such as an ERP (enterprise resource planning) for inventory of raw materials, a LIMS (laboratory information management system) or ELN (electronic lab notebook) for lab results and sample tracking, a MES (manufacturing execution system) for electronic batch records, SCADA (supervisory control and data acquisition) for monitoring and controlling manufacturing devices. Analytical results and audit trails from specific unit operations are sometimes stored on the manufacturing machinery itself (e.g. the bioreactor, flow cytometer), while some data might be aggregated in excel files (often without access control), etc. Additionally, those systems, or the use of those systems, might

differ from department to department within the same facility and/or company. This results in a fragmented patchwork of data sources that grows more and more complex with the scale of the organization and that creates additional overhead in the QA department in order to put the pieces of information together. A fragmented information architecture, i.e. the lack of a single source of truth, leads to inefficient decision making and non-coherent audit trails. This type of data management situations works for low volume production, but it can easily become a bottleneck for large-scale production processes.

A streamlined data management and analytics strategy should at least consider the following four topics:

1. **Data integrity:** according to the ‘garbage in, garbage out’ principle, if the stored data is not complete, consistent and accurate [14], any conclusion drawn from this data cannot be trusted. Three main mechanisms exist for increasing the data integrity. First of all, automation decreases variability in input data as human errors are factored out of the documentation process (see point 2). For example the use of bar code scanners reduces errors compared to manually writing down lot numbers of used products. Second, where data documentation itself cannot be automated, the validation of the user input should be automated. For example by making use of electronic batch records, unit conversions, calculations, and checking of missing data and signatures can be automated. And lastly,

digital signatures (making use of encryption) should be implemented that assures that the collected data cannot be tinkered with. For example by encrypting the collected data within the signature, intentional or unintentional data tampering of information that is already in the database can be detected.

2. Automated data collection and centralization: this reduces both the workload and the chance for erroneous data manipulations. On the one hand there is the current trend for increased connectedness of all type of devices that should be leveraged (Industry 4.0 and Internet of Things). While it is technically still relatively challenging to achieve integration on a larger scale [15], and the data integrity issue could become even more prominent in cloud-based settings, most modern ERP, LIMS and MES systems have an API (application programming interface) that allows to exchange data according to fixed protocols. On the other hand, there are still quite some manual processing steps involved in cell and gene therapy manufacturing, going from refreshing media to connecting disposables and filling out batch records. Also those manual actions need to be documented and included in the audit trail for example. This is often (partially) done by means of paper-based systems in most manufacturing facilities. In some cases, electronic batch records could be a solution for the streamlined data collection of manual process steps, but

at least some kind of user interface is required where process operators can easily document their manual actions simultaneously with the physical action.

3. Process intelligence: While generic data management tools can be used that are not directly tailored to cell and gene therapy processes, the specific challenges of biological processes require knowledge that facilitates the data collection and certainly the data analytics. For example, it would make efficient use of the operators time if the system automatically processes and trends critical process parameters when new measurement data comes in (e.g. cell counts), or if analysis of in-process controls are automatically linked to potency assays. This would not only reduce the patchwork, but also lead to faster decision making.
4. Built-in compliance and data security: Note that under the previous topic of 'which data to collect', the collection of data in order to comply to regulation such as GMP (e.g. 21 CFR 11 on electronic records or 21 CFR Part 211 on manufacturing documentation) was not specifically mentioned. First of all one could argue that patient safety, product sterility etc. is an integral part of the target product profile and should therefore be reflected in the CQAs. Additionally, by using appropriate tools for data collection, the regulatory requirements and data security such as the audit trail, batch

records, signatures, permission management, etc. should be inherently built-in into the data management tool, effectively delivering compliance out of the box.

There is a long list of suppliers of generic data management platforms (e.g. SAP, Rockwell, Siemens, Lighthouse etc.). However, while they offer specialized packages for various verticals such as life sciences or pharmaceuticals, they still require significant configuration to meet the specific needs of cell and gene therapy production settings. More recently, highly specialized cell and gene therapy data management and data analytic platforms are arising, e.g., TrakCel and Vineti focusing on orchestrating the ‘needle to needle’ supply chain of cells and other raw materials, DataHow focusses on (big) process data analytics and descriptive process models, and MyCellHub focuses on the production process itself by the generation of real-time audit trails and actively managing workflows and documentation through intelligent work instructions (i.e. digital standard operating procedures and batch records).

STREAMLINING DATA ANALYTICS

It’s clear that a connected data management system does not only replace paper-based processes and manual note-taking, but is also able to actively manage workflows, guide documentation and facilitate regulatory compliance, which all ultimately leads to cost reductions.

These benefits alone could justify the investment in such data

management tools, but having a solid data management foundation is the first (and probably most important) step to be able to implement value generating and/or predictive data analytics.

Big data, artificial intelligence and machine learning techniques promise a surge in data-driven insights, and the use of *in silico* models for clinical trials are being discussed. In a research and development setting for cell and gene therapies substantial amount of time is dedicated on data analysis and interpretation, requiring a highly skilled labor force. Committing these resources is not always feasible for QC/QA teams in a commercial production environment, especially in small and medium enterprises. Again here automation of data analytics is key as it both improves throughput and reproducibility. Attention should be paid however to the type of analytics, specifically in the manufacturing environment and to a lesser extent in the R&D environment.

Completely ‘black box’ machine learning or artificial intelligence algorithms that are taking autonomous decisions on the quality of the product, often based on heuristics/statistics and data collected within a specific scope, might prove to be challenging in terms of (software as a medical device) regulation. Another challenge that comes up with these black box algorithms in commercial production of cell and gene products is “Who will be responsible for mistakes made by these algorithms?”

On the other hand there are the more classical ‘white box’ models, i.e., deterministic or mechanistic models. These models translate real *a priori* physical insight in the biological system into mathematics and

their use in regulated environments could therefore be less challenging. However, compared to data-based models they are much more difficult to develop and parameterize for complex systems, and generally cannot be used to process data in real-time due to a high computational load.

Since the order of complexity of the bio-manufacturing systems is so high that the rate of data collection surpasses the speed of analysis for deterministic or mechanistic models, hybrid data-based mechanistic models could prove useful. They are easier to compute (which is a benefit for process control algorithms that run in real time) and faster to develop as compared to purely mechanistic models, but their internals are easier to understand and verify when compared to machine learning or artificial intelligence algorithms [16,17]. These features will facilitate their regulatory compliance and therefore their integration in the production process.

TRANSLATIONAL INSIGHT

While a lot of work is done on scaling the physical production process of cell and gene therapies to accommodate their commercial use, it has to be taken care of that regulatory compliant data management and analytics does not become a new bottleneck.

This article listed key takeaways on streamlining data management and analytics for cell and gene therapy production:

- ▶ First make sure to reduce the process variability to a minimum so the relation between Critical Process Parameters and Critical

Quality Attributes are as clear as possible.

- ▶ Choose wisely which data to collect and focus on meaningful data i.e. make sure the collected data correlates to the Critical Quality Attributes in a direct or indirect way.
- ▶ Move away from paper based systems and aim to implement an integrated data management platform that automatically collects and centralizes data, assures data integrity and enforces appropriate regulation while being aware of the ongoing processes.
- ▶ Properly collected data and data management is the lifeblood of (predictive) analytics that are potentially able to transform the production process or the product for the better.
- ▶ Algorithms for data analysis of which the internal working is interpretable by physical concepts and that guide a human operators to take better decisions in less time (instead of autonomously intervening in the process itself), will be the norm for the foreseeable future.

Outcomes of better data management and process analytics for the manufacturing of cell and gene therapies :

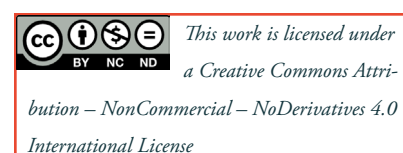
- ▶ Improves monitoring, control and optimization of the production process
- ▶ Faster (and potentially real-time and on-line) assessment

- of release criteria and potency testing
- ▶ Cost of goods reduction through automation and efficiency gains in the manufacturing process
- ▶ More effective scaling of manufacturing since changes in Critical Quality Attributes are understood earlier
- ▶ Improved quality assurance
- ▶ De-risking cell therapy through a robust and reliable delivery for every patient
- ▶ Improved production logistics and patient scheduling (e.g. cell harvest and infusion)

- ▶ Reduction of administrative workload through automated traceability and audit trails

FINANCIAL & COMPETING INTERESTS DISCLOSURE

Sébastien de Bournonville, Toon Lambrechts and Thomas Pinna are involved in the development of My-CellHub, a data management platform for cell-based products. Toon Lambrechts is funded by VLAIO project HBC.2016.0629. Sébastien de Bournonville is supported by a PhD grant of the Research Foundation Flanders (FWO, Grant No. 1S67217N, www.fwo.be).



REFERENCES

1. Lipsitz YY, Milligan WD, Fitzpatrick I et al. A roadmap for cost-of-goods planning to guide economic production of cell therapy products. *Cytotherapy* 2017; 19(12), 1383–1391.
2. Levine BL, Miskin J, Wonnacott K, Keir C. Global Manufacturing of CAR T Cell Therapy. *Mol. Ther. Methods Clin. Dev.* 2017; 4, 92–101.
3. Abraham E, Ahmadian BB, Holderness K, Levinson Y, McAfee E. Platforms for Manufacturing Allogeneic, Autologous and iPSC Cell Therapy Products: An Industry Perspective. *Adv. Biochem. Eng. Biotechnol.* 2017; doi: 10.1007/10_2017_14. [Epub ahead of print].
4. Novartis Has Production Hiccup for Kymriah CAR-T Therapy. *Pharm-Tech.com* 2018; [Online] www.pharmtech.com/novartis-has-production-hiccup-kymriah-car-t-therapy.
5. Stanton D. Lonza: CAR-T Manufacturing Glitch an Industry Problem, not Just Novartis's BioProcess International 2018; [Online] www.bioprocessintl.com/bioprocess-insider/therapeutic-class/lonza-car-t-manufacturing-glitch-an-industry-problem-not-just-novartiss/.
6. Lipsitz YY, Timmins NE, Zandstra PW. Quality cell therapy manufacturing by design. *Nat Biotechnol.* 2016; 34(4), 393–400.
7. Csaszar E, Kirouac DC, Yu M et al. Rapid expansion of human hematopoietic stem cells by automated control of inhibitory feedback signaling. *Cell Stem Cell* 2012; 10(2), 218–229.
8. Rafiq QA, Toward a scalable and consistent manufacturing process for the production of human MSCs. *Cell Gene Ther. Insights* 2016; 2(1), 127–140.
9. Papantoniou I, Lambrechts T, Aerts J–M. Bioprocess engineering strategies for autologous human MSC-based therapies: one size does not fit all. *Cell Gene Ther. Insights* 2017; 3(6), 469–482.
10. Lambrechts T, Papantoniou I, Rice B, Schrooten J, Luyten FP, Aerts J–M. Large-scale progenitor cell expansion for multiple donors in a monitored hollow fibre bioreactor. *Cytotherapy* 2016; 18(9), 1219–1233.
11. Viazzi S, Lambrechts T, Schrooten J, Papantoniou I, Aerts JM. Real-time characterisation of the harvesting process for adherent mesenchymal stem cell cultures based on on-line imaging and model-based monitoring. *Biosyst. Eng.* 2015; 138, 104–113.
12. Lambrechts T, Papantoniou I, Sonnaert M, Schrooten M, Aerts J–M. Model-based cell number quantification using online single-oxygen sensor

- data for tissue engineering perfusion bioreactors. *Biotechnol. Bioeng.* 2014; 111(10), 1982–1992.
13. Lambrechts T, Papantoniou I, Rice B, Schrooten J, Luyten FP, Aerts JM. Large-scale autologous stem cell expansion in monitored hollow fibre bioreactors: evaluation of donor-to-donor variability. *Cytotherapy* 2016; 18(9), 1219–1233.
 14. Food and Drug Administration. FDA Data Integrity and Compliance With CGMP – Guidance for Industry. 2016.
 15. Branke J, Farid SS, Shah N. Industry 4.0: a vision for personalized medicine supply chains? *Cell Gene Ther. Insights* 2016; 2(2), 263–270.
 16. Geris L, Lambrechts T, Carlier A, Papantoniou I. The future is digital: In silico tissue engineering. *Curr. Opin. Biomed. Eng.* 2018; 6, 92–98.
 17. von Stosch M, Davy S, Francois K et al. Hybrid modeling for quality by design and PAT-benefits and challenges of applications in biopharmaceutical industry. *Biotechnol. J.* 2014; 9(6): 719–726.

AFFILIATIONS

Sébastien de Bournonville

Prometheus, Division of Skeletal Tissue Engineering, KU Leuven, Leuven, Belgium
sebastien.debournonville@kuleuven.be

Toon Lambrechts

MyCellHub & Prometheus, Division of Skeletal Tissue

Engineering, KU Leuven, Leuven, Belgium

toon.lambrechts@mycellhub.com

Thomas Pinna

MyCellHub & Prometheus, Division of Skeletal Tissue Engineering, KU Leuven, Leuven, Belgium
thomas.pinna@mycellhub.com

Ioannis Papantoniou

Prometheus, Division of Skeletal Tissue Engineering & Skeletal Biology and Engineering Research Center, KU Leuven, Leuven, Belgium
ioannis.papantoniou@kuleuven.be

Jean-Marie Aerts

Animal and Human Health Engineering Division M3-BIORES, KU Leuven, Leuven, Belgium
jean-marie.aerts@kuleuven.be