Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Biomedical Engineering

Current status and opportunities in adaptive data analysis for therapeutic cell manufacturing

Zhaonan Liu[1,2], Jialei Chen[1,3], Kan Wang[1], Ben Wang[1,2,3] and Chuck Zhang[1,3*]

Abstract

As an emerging technology, therapeutic cell manufacturing faces major challenges in three aspects, knowledge gap, regulations, and case-to-case variability. Among all variability sources, donor-to-donor variability is intrinsic to therapeutic cell manufacturing and can be very large. A few recent articles have addressed this variability, but enormous research opportunities remain. In this opinion article, we focus on the donor-to-donor variability and point out a new sub-field in the data analysis research, adaptive data analysis, as a potential solution to minimize the donor-to-donor variability. An adaptive data analysis framework adapts to each case and provides case-specific instructions on the manufacturing process. We present and discuss three specific adaptive data analysis approaches, including multi-task learning, Bayesian latent variable modeling, and representation learning. These modeling techniques may provide valuable solutions to challenging data analysis problems in cell manufacturing where large variabilities exist.

Addresses

[1] Georgia Tech Manufacturing Institute, Georgia Institute of Technology, Atlanta, 30332, United States

[2] School of Materials Science and Engineering, Georgia Institute of Technology, Atlanta, 30332, United States

[3] H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, 30332, United States

Corresponding author: Zhang, Chuck (chuck.zhang@gatech.edu)

Current Opinion in Biomedical Engineering 2021, **20:**100351

This review comes from a themed issue on **Molecular & Cellular Eng: Cell and Gene Therapy Manufacturing**

Edited by **Krishnendu Roy**

For complete overview of the section, please refer the article collection - Molecular & Cellular Eng: Cell and Gene Therapy Manufacturing

Received 4 June 2021, revised 1 September 2021, accepted 15 September 2021

https://doi.org/10.1016/j.cobme.2021.100351

Keywords

Cell manufacturing, Data analysis, Variability.

Introduction

Standardized mass production has enabled high-throughput pharmaceutical manufacturing in the past century, bringing trustworthy and affordable medicines to countless homes and families. However, limitations of conventional standardization started to surface as medical technologies advanced in the past few decades. Cell therapies are revolutionary technologies that treat diseases in a substantially different way than traditional pharmaceutics. Instead of chemical compounds, cell therapies treat patients with living cells. Current cell therapies mainly include stem cell therapies and immune cell therapies [1–3]. These cells may be genetically edited and thus possess abilities that the original cells of the patient lack and need. For example, chimeric antigen receptor T cells, genetically edited using the patient's original T cells, can recognize cancer cells and lead the immune system to eradicate them [4]. The new technologies bring possibilities to treat diseases that other therapies cannot treat effectively. However, the cell and gene therapies have inherited case-to-case variability that is difficult to standardize.

The challenge associated with the case-to-case variability in cell therapies is twofold, variability in patients and variability in the raw materials, that is, donor cells. The difference in individual patients is not new to conventional medicine. Doctors may adjust the type and dosage of medicine to accommodate each patient's conditions, for example, gender, age, weight, allergies, and medical history. Such considerations and adjustments may also apply to cell therapies but cell therapies require more. Alien living cells can have dramatic effects on the human body, such as severe immune rejection. A thorough inspection of the patient's immune system is necessary to ensure safe and effective cell therapy.

Unlike compound-based conventional medicines, cell therapies use living cells as raw materials and final products. Those living cells may come from a healthy donor, referred to as allogeneic, or they may come from the very same patient, referred to as autologous. For either allogeneic or autologous cell therapy, its raw materials come from a human body, which may vary significantly in viability, metabolic level, and many other aspects that impact the manufacturing process [5–8].

Data analysis in therapeutic cell manufacturing

Humans have learned to use cells to make bread, yogurt, and alcoholic beverages centuries ago, even though their makers may not have the slightest idea about cells. Yeast

and lactic acid bacteria are very adaptable and can even alter the environment in their favor. However, *ex vivo* human cells, without the protection of a human body, are sensitive to the environment and vulnerable to attacks from other microorganisms. Thus, therapeutic cell manufacturing is complicated and requires more than a profound understanding of cells. To successfully grow human cells massively in an artificial environment, deliberate monitoring and control of the manufacturing process parameters are essential [9].

Key challenges

As an emerging technology, therapeutic cell manufacturing is facing several challenges. First, knowledge and technology gaps hinder a seamless transition of therapeutic cell manufacturing from the lab scale to a full commercial scale. Life sciences have made significant progress in the past century, but there are still more unknowns than knowns. Furthermore, the current regulations for traditional pharmaceuticals are a challenge for this new medical paradigm. Unlike traditional synthetic pharmaceuticals, cell therapies' quality control and process control data only represent a small proportion of the product characteristics. A practical regulatory approach for advanced cell therapies is in grave need. Some protocols for therapeutic cell manufacturing have partially overcome the challenges above, and, as a result, the cell products have entered the market [10–12]. However, these protocols alone are insufficient for the production to scale and make the products accessible and affordable to those who need the treatment.

Data analysis is a promising research field that may provide practical solutions to the challenges mentioned previously. For the knowledge gap in fundamental sciences, data analysis has proven to be effective [13,14]. For manufacturing processes with vast variability, data analysis can potentially identify the underlying patterns. Data analysis also helps quantify and standardize the manufacturing processes, supporting the approval process by regulatory authorities.

Donor-to-donor variability

Variability is omnipresent in cell manufacturing. Different starting materials [15], culture media [16], and operator measurement [17] are all factors that lead to case-to-case variability. These variabilities bring challenges in cell manufacturing, especially process monitoring and process control. Among these variabilities, the donor-to-donor variability is intrinsic to therapeutic cell manufacturing because the starting materials of the manufacturing process may have to come from different donors [8]. It differentiates therapeutic cell manufacturing from conventional pharmaceutical manufacturing processes. The intrinsic variability in the starting materials further complicates the challenges mentioned in Section 2.1.

Quantification is one plausible way to address donor-to-donor variability partially. A recent article quantified the variability between cell culture processes using cells from different donors [15]. The results from the study show that donor-to-donor variability can be significant, but its impact on the final product is limited to some extent, given its magnitude. However, because this method does not reduce the variability itself, it may be insufficient for quality control when the donor-to-donor variability is enormous.

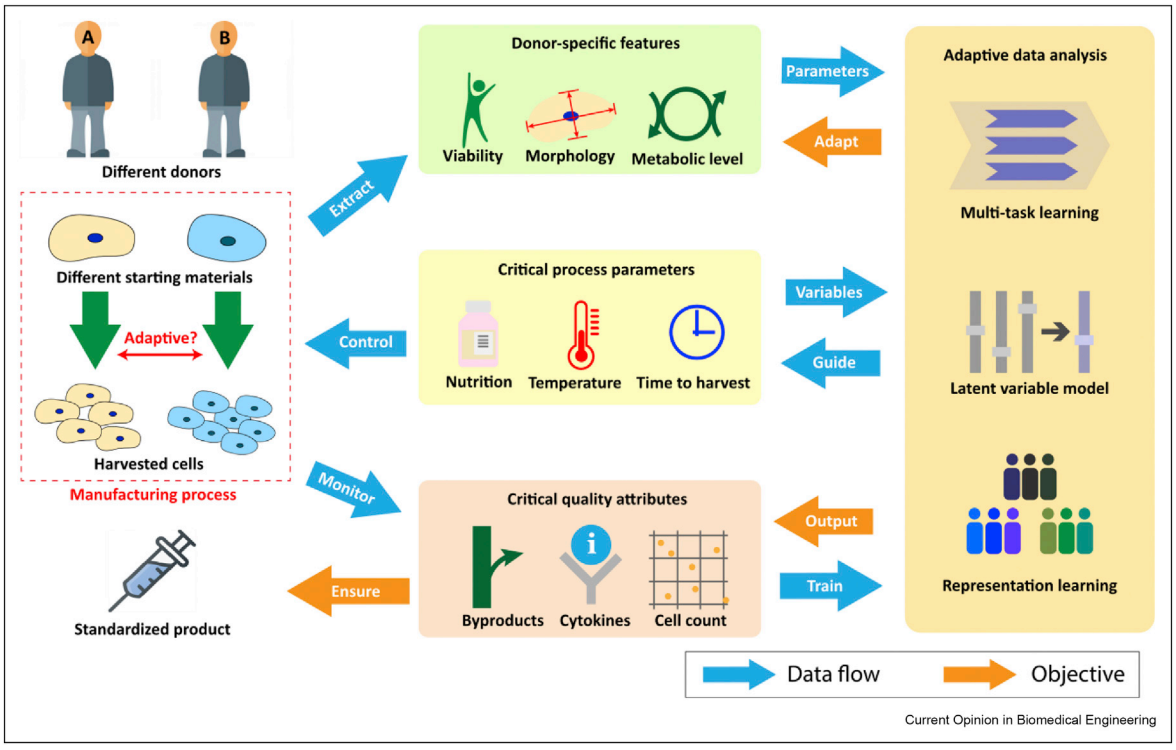
Donor-to-donor variability is governed by underlying patterns rather than random noise, which leads to a more promising approach than quantification alone. Cells collected from each donor bring unique features, such as viability, morphology, and metabolic level. These characteristics can sometimes be grouped by the phenotypes/haplotypes, which in adaptive data analysis may serve as a method to reduce the solution space. If the manufacturing process adapts to these features for each donor, the donor-to-donor variability can be reduced, resulting in more standardized cell products, which are safer for the patients and easier to get approval from regulators. Adaptive data analysis can identify the underlying patterns in massive data and render the manufacturing process adaptive to different donors (Figure 1).

Process monitoring: critical quality attributes and critical process parameters

Process monitoring is vital for any manufacturing process and more so in therapeutic cell manufacturing because living cells are complex, delicate, and difficult to inspect. To ensure the quality of cell therapy products, manufacturers must monitor critical quality attributes (CQAs) that reflect the condition and quality of the cells in culture. Typical CQAs for cell manufacturing include cell count and morphology, signaling molecules (cytokines), and concentrations in nutrients and byproducts. Critical process parameters (CPPs), the parameters that can impact CQAs, also need to be monitored for feedback control. Many CPPs need careful monitoring and controlling in therapeutic cell manufacturing, such as temperature, CO_2 concentration, nutrition, and time to harvest.

Most CQAs and CPPs in cell manufacturing are not directly observable and rely on indirect measurement [18,19]. Data analysis can help establish a model that describes the connection between sensor readings, CQAs, and CPPs. For example, the electric impedance of a cell culture system is considered an indicator of viable cell count (VCC). VCC impacts impedance readings, but the exact relationship is difficult to derive theoretically. With the help of data analysis, an optimal relationship can be found and thus enable more effective VCC monitoring [20,21]. Sometimes the data

Figure 1



Adaptive data analysis is a promising approach to address intrinsic donor-to-donor variability. By incorporating donor-specific features, adaptive data analysis helps the cell manufacturing process adapt to different starting materials.

analysis programs process multiple sensor readings together. These programs are commonly referred to as soft sensors, which have shown excellent performance in bioprocessing [22].

Process control

Combined with process monitoring and data analysis, process control is essential to establish a robust closed-loop control system in manufacturing processes. Because of the complexity of bioprocesses of living cells, process control in therapeutic cell manufacturing is particularly challenging. In therapeutic cell manufacturing, the cell culturing process can be controlled by adjusting CPPs [23]. To ensure that CPPs are moving in a desirable direction, a good understanding of the effects CPPs have on CQAs is essential. These effects can be complicated and may need many rounds of trial and error before arriving at a desirable protocol. Data analysis can help understand these effects in two ways (1) to improve its validity by learning a model to describe the effects [24], and (2) to expedite the trial-and-error process by optimizing the design of experiments to reduce the number of needed experiments, for example, by fractional factorial designs, to reach a robust understanding of the relationship between CQAs and CPPs [3].

Data analysis

Most modeling and monitoring tasks in cell manufacturing can be represented as a standard supervised learning problem. More specifically, for process control, we have input variables $\boldsymbol{x}_i$, denoting the collection of CPPs for the ith experiment. Meanwhile, we denote y_i the corresponding CQAs of interest, for example, VCC. The goal of data analysis is to find an input–output relationship $y(\cdot)$ that best fits the dataset $\{\boldsymbol{x}_i, y_i\}_{i=1}^n$:

$$y_i = y(x_i) + e_i \tag{1}$$

Here, e_i is the measurement error of ith experiment often assumed independently and identically drawn from some pre-specified distribution. Thanks to the development of machine learning, there are R/Python packages available. Many of them also leverage cloud computing platforms, including Amazon AWS and Microsoft Azura.

Adaptive data analysis

The modeling method in Section 2 introduces an assumption that the whole data set $\{\boldsymbol{x}_i, y_i\}_{i=1}^n$ is governed by the same underlying relationship $y(\cdot)$. As discussed

earlier, this assumption is often invalid in cell manufacturing applications because of the vast difference in the initial culturing material from different donors.

To this end, we believe a future direction would be adaptive models to the specific donor, thereby achieving better monitoring procedures and higher quality cell products. We call this 'adaptive data analysis.' More specifically, using the same notation in Section 2, we model the collected dataset $\{\boldsymbol{x}_i, y_i\}_{i=1}^n$ via different models $y_k(\bullet)$:

$$y_i = y_k(x_i) + e_i \tag{2}$$

Here, note that we use $k = 1, \cdots, K$ to denote different donors and $i = 1, \cdots, n$ to denote $n > K$ data points collected from the K donors. Furthermore, let n_k denote the amount of data from donor k, and obviously $n = n_1 + \cdots + n_K$.

One naïve work-around is to fit separate models, $y_k(\bullet)$ for the donor k using the data from the specific donor, via similar methods reviewed in Section 2 [25]. However, this strategy may not be helpful in many applications for two reasons. First, the amount of data $n_k$corresponding to the donor k might be smaller than the dimension of the input data, leading to poor predictive performance. Second and perhaps more importantly, in future applications, we will use our model for the new donors, denoted as, say $k+1$, whose input–output relationship $y_{k+1}(\bullet)$ is yet to be obtained.

One way to address the above issues is to rewrite the model (Equation (2)) via an additional variable $\boldsymbol{s}_k$:

$$y_i = y_k(x_i) + e_i = g(x_i, s_k) + e_i \tag{3}$$

Here, the $\boldsymbol{s}_k$ is donor-specific features, that is, a collection of the donor's information to differentiate him/her from other donors best. Intuitively speaking, $\boldsymbol{s}_k$ can reflect the historical disease status, age, and even the genetic material of the donor k.

The key idea behind the proposed adaptive model (Equation (3)) is that we use a donor-specific variable $\boldsymbol{s}_k$ to exploit the specific aspect of different donors while using a universal model $g(\bullet)$ to explore the similarity of the whole population to ensure a reasonable data amount for mining. While a latent variable $\boldsymbol{s}_k$ is introduced, it may not lead to overfitting. This is because we can now use all data for model fitting rather than n_K data for the patient k.

Now suppose we have two donors, A and B. For simplicity, we suppose A and B are twins while having different lifestyles and dietary habits. As a result, their cells behave differently in *ex vivo* culturing processes. Adaptive data analysis allows the manufacturer to process the massive donor-specific data collected from the two donors and their cells and arrive at different culturing protocols for each, e.g., different glucose levels and harvest time. The protocols are optimized for each donor to ensure high product quality, for example, cell count and viability, and all protocols are also optimized as a whole to ensure consistency between different batches.

We present three approaches (Table 1) to adopting the adaptive model (Equation (3)) in therapeutic cell manufacturing, with inspiration from the modern machine learning literature.

Multi-task learning

The first category of approaches uses the multi-task learning methods in the machine learning literature [26,27], which directly works on Equation (2). Here, each 'task' is to learn the model $y_k(\bullet)$ for the patient k. The key idea of multi-task learning methods is to assume a particular connection between the different tasks $y_k(\bullet)$ and then try to learn them together with the whole data set.

The simplest way to implement the multi-task learning idea is described in the following [28]. We first assume each donor is the same (i.e. the problem goes back to Equation (1)) and use a neural network to parametrize $y(\bullet)$. With this assumption, we can use the complete data $\{\boldsymbol{x}_i, y_i\}_{i=1}^n$ to train the neural network and obtain a model $\widehat{y}(\bullet)$. Then, for the specific donor k, we will use the corresponding small data of size n_k to fine-tune this $\widehat{y}(\bullet)$ and obtain $\widehat{y}_k(\bullet)$. Note that in practice, we would select neural network architecture according to the total data size n. For effective fine-tuning, we often fix most of the neural network parameters and only change a small subset (e.g. the last layer).

Some more sophisticated multi-task learning methods in the literature may be promising in practical cell manufacturing cases. For example, the model-agnostic meta-learning method can scale to an extensive data

Table 1

Pros and cons for the three proposed adaptive data analysis methods.

Methods	Advantage	Disadvantage
Multi-task learning	Easy to implement for complex data	Require many donors
Bayesian latent variable model	Provide uncertainty quantification	High computational cost
Representation learning	Easy to leverage patient's records	Difficult to interpret

set incorporating complex patient information [29]; cross-stitch networks are explicitly developed for image inputs, including computed tomography images of the donor and the microscopic cell images [30]; Hawkes relational meta-learning method can capture and leverage the underlying mixed-community patterns of the donors [31]; there are also non-neural–based models for better interpretability and provide uncertainty quantification [32–34]. It is noteworthy that most of those methods require many tasks, which means hundreds or thousands of donors in therapeutic cell manufacturing cases, which may pose a challenge to research in small laboratories.

Bayesian latent variable model

The second category of approaches is Bayesian latent variable models. Note that the critical challenge in Equation (3) is that the feature for the donor $\boldsymbol{s}_k$ can be complicated and high-dimensional. Therefore, we introduce latent variables $l_k = l(\boldsymbol{s}_k)$ and l_k is a much lower-dimensional variable, that is, 1D or 2D, in practice. With latent variable l_k, we can rewrite the model (3) in the following hierartical Bayesian flavor and assign some priors:

$$l(\cdot) \sim \pi_l,\ g(\cdot) \sim \pi_g,\ e_i \sim \pi_e,\ l_k = l(s_k),\ y_i = g(x_i, l_k) + e_i$$

Here, π_l, π_g, π_e denote the priors of different model aspects, specified according to prior knowledge. For functions $l(\cdot)$ and $g(\cdot)$, one can assume some parametric form similar to Bayesian linear regression [35] or directly use nonparametric priors, for example, Gaussian process [36]. For error e_i, a typical way is to assume they are independently drawn from a zero-mean Gaussian distribution. Finally, the parameter estimation can be conducted via the maximum a posteriori method for efficiency or Markov chain Monte Carlo methods for a complete account of uncertainty [37].

We would like to highlight one work in the literature that adopts this idea [21]. In this work, the authors model the donor-specific features as a calibration parameter for the uniqueness of the donor. They then use multiple biosensors to infer the underlying calibration parameter for the donor. Finally, they use the estimated model to recover the viable cell concentration for cell manufacturing monitoring and scale-up.

The advantage of the Bayesian latent variable model for cell manufacturing applications is twofold. First, it can naturally provide the quantification of uncertainty associated with the prediction. In biomedical applications, this uncertainty quantification can be as important as the prediction itself. Second, the latent value l_k provides information of different donors, which also applies to downstream applications, including matching the donor with donees.

There are also implications related to the Bayesian latent variable models. First, the Markov chain Monte Carlo method for parameter estimation is well-known for its high computational cost. This is particularly true when the latent dimension is high, that is, we have rich historical information of each donor. Therefore, many recent works leverage a more scalable expectation propagation [38] or variational inference methods [39]. In this way, the parameter estimation time can be reduced to less than 1 h, which is acceptable compared with days of culturing practice. Second, those methods require proper prior specifications to achieve a reliable model estimation. Such a prior specification usually leverages the domain-specific knowledge, and it might not be easy as the knowledge is often lacking in cell manufacturing applications. If no prior knowledge is available, normal distributions can be used.

Representation learning

Another approach is deep representation learning [40,41], which can also be viewed as a deep learning extension of the Bayesian latent variable model discussed earlier. Similarly, the idea is to find some variables $l_k = l(\boldsymbol{s}_k)$ for each donor. However, instead of using a hierartical Bayesian model, the representation learning method seeks to find the embedding by constructing an additional learning task. This additional learning task often leverages the patient's pathological features and historical disease records. Intuitively speaking, when conducting learning of this task, we can group the similar donors close together and separate those not similar. This additional learning task should be constructed on a case-by-case basis, with the available information of the donors at hand.

A straightforward way for a new task is to cluster the donors into J different groups, using the age, race, and also the historical disease course of the donors. We can then assign the same variables $l_k = e_j$ for the donors in the group j, where $e_j = [0, \cdots, 0, 1, 0, \cdots, 0]$ is one of the basis vectors of the J-dim space. In this way, the in-group distances are always zero, while the between-group distances are $\sqrt{2}$.

Some more sophisticated representation learning methods in the literature may also apply to cell manufacturing. For example, generative adversarial networks can be used to extract information from images [42], generative invertible networks can further provide a bidirectional mapping that appears to preserve pathological meaning [43], and some works mask a part of each datum and use the remaining input to predict the masked value [44].

Conclusion

Donor-to-donor variability is intrinsic in therapeutic cell manufacturing and an inevitable challenge. Recent research has partially addressed this challenge, which is plausible but insufficient. An emerging sub-field in data analysis research — adaptive data analysis — can be a potentially effective solution. An adaptive data analysis framework can adapt to the inherent variability and provide case- or patient-specific guidelines for the manufacturing process. The key idea of adaptive data analysis is to use specific models for each donor. Multi-task learning, Bayesian latent variable modeling, and representation learning are three potent candidates for finding donor-specific models. Multi-task learning assumes a particular connection between the different models and treats those models together as a whole learning task. Bayesian latent variable modeling tries to simplify the relationship between the donor-specific models by finding latent variables. Representation learning extends Bayesian latent variable learning with an additional learning task that groups similar donors. These adaptive data analysis methods can potentially improve the robustness and scalability of therapeutic cell manufacturing processes. In conclusion, adaptive data analysis with applications to therapeutic cell manufacturing is a promising research direction full of opportunities and in urgent need.

Declaration of competing interest

This work is supported by National Science Foundation CMaT ERC (Award # NSF EEC-1648035), and FDA (Award # 1R01FD006598-01).

References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest
** of outstanding interest

1. Wei X, Yang X, Han ZP, Qu FF, Shao L, Shi YF: **Mesenchymal stem cells: a new trend for cell therapy**. *Acta Pharmacol Sin* 2013, **34**:747–754.
2. Robinton DA, Daley GQ: **The promise of induced pluripotent stem cells in research and therapy**. *Nature* 2012, **481**:295–305.
3. Fesnak AD, June CH, Levine BL: **Engineered T cells: the promise and challenges of cancer immunotherapy**. *Nat Rev Canc* 2016, **16**:566–581.
4. Wang ZG, Wu ZQ, Liu Y, Han WD: **New development in CAR-T cell therapy**. *J Hematol Oncol* 2017, **10**.
5. Schaffer S, Lam VYM, Ernst IMA, Huebbe P, Rimbach G, Halliwell B: **Variability in APOE genotype status in human-derived cell lines: a cause for concern in cell culture studies?** *Gene Nutr* 2014, **9**.
6. Yang YLP, Wang SM, Li XG, Schor NF: **Cell line-dependent variability of coordinate expression of p75NTR and CRABP1 and modulation of effects of fenretinide on neuroblastoma cells**. *Oxidat Med Cell Longevit* 2016, **2016**.
7. Baer PC, Kuci S, Krause M, Kuci Z, Zielen S, Geiger H, Bader P, Schubert R: **Comprehensive phenotypic characterization of human adipose-derived stromal/stem cells and their subsets by a high throughput technology**. *Stem Cell Dev* 2013, **22**:330–339.
8. Heathman T: **Development of a controlled bioreactor process to drive the consistent manufacture of human mesenchymal stem cells from multiple donors**. *Cytotherapy* 2015, **17**:S81–S82.
9. Pigeau GM, Csaszar E, Dulgar-Tulloch A: **Commercial scale manufacturing of allogeneic cell therapy**. *Front Med* 2018, **5**.
10. Lee DW, Kochenderfer JN, Stetler-Stevenson M, Cui YZK, Delbrook C, Feldman SA, Fry TJ, Orentas R, Sabatino M, Shah NN, *et al.*: **T cells expressing CD19 chimeric antigen receptors for acute lymphoblastic leukaemia in children and young adults: a phase 1 dose-escalation trial**. *Lancet* 2015, **385**:517–528.
11. Neelapu SS, Locke FL, Bartlett NL, Lekakis LJ, Miklos DB, Jacobson CA, Braunschweig I, Oluwole OO, Siddiqi T, Lin Y, *et al.*: **Axicabtagene ciloleucel CAR T-cell therapy in refractory large B-cell lymphoma**. *N Engl J Med* 2017, **377**:2531–2544.
12. Trounson A, McDonald C: **Stem cell therapies in clinical trials: progress and challenges**. *Cell Stem Cell* 2015, **17**:11–22.
13. Falk T, Mai D, Bensch R, Çiçek Ö, Abdulkadir A, Marrakchi Y, Böhm A, Deubner J, Jäckel Z, Seiwald K, U-Net: **Deep learning for cell counting, detection, and morphometry**. *Nat Methods* 2019, **16**:67–70.
14. Ma J, Yu MK, Fong S, Ono K, Sage E, Demchak B, Sharan R, Ideker T: **Using deep learning to model the hierarchical structure and function of a cell**. *Nat Methods* 2018, **15**:290.
15. * Picken A, Harriman J, Iftimia-Mander A, Johnson L, Prosser A, Quirk R, Thomas R: **A Monte Carlo framework for managing biological variability in manufacture of autologous cell therapy from mesenchymal stromal cells therapies**. *Cytotherapy* 2020, **22**:227–238.
This paper quantifies donor-to-donor variability using a Monte Carlo method.
16. McGillicuddy N, Floris P, Albrecht S: **Bones J: examining the sources of variability in cell culture media used for biopharmaceutical production**. *Biotechnol Lett* 2018, **40**:5–21.
17. Grant R, Coopman K, Medcalf N, Silva-Gomes S, Campbell JJ, Kara B, Braybrook J, Petzing J: **Understanding the contribution of operator measurement variability within flow cytometry data analysis for quality control of cell and gene therapy manufacturing**. *Measurement* 2020, **150**.
18. Theuer L, Randek J, Junne S, Neubauer P, Mandenius CF, Beni V: **Single-use printed biosensor for L-lactate and its application in bioprocess monitoring**. *Processes* 2020, **8**.
19. Parate K, Rangnekar SV, Jing DP, Mendivelso-Perez DL, Ding SW, Secor EB, Smith EA, Hostetter JM, Hersam MC, Claussen JC: **Aerosol-jet-printed graphene immunosensor for label-free cytokine monitoring in serum**. *ACS Appl Mater Interfaces* 2020, **12**:8592–8603.
20. Xu Y, Li C, Mei W, Guo M, Yang Y: **Equivalent circuit models for a biomembrane impedance sensor and analysis of electrochemical impedance spectra based on support vector regression**. *Med Biol Eng Comput* 2019, **57**:1515–1524.
21. ** Chen J, Liu Z, Wang K, Jiang C, Zhang C, Wang B: **A calibration-free method for biosensing in cell manufacturing**. *IISE Transact* 2020:1–39.
This paper discusses a calibration-free method for recovering cell density under patient-to-patient / donor-to-donor variability.
22. Luttmann R, Bracewell DG, Cornelissen G, Gernaey KV, Glassey J, Hass VC, Kaiser C, Preusse C, Striedner G, Mandenius CF: *Soft sensors in bioprocessing: a status report and recommendations*. Wiley Online Library; 2012.
23. Kirouac DC, Zandstra PW: **The systematic production of cells for cell therapies**. *Cell Stem Cell* 2008, **3**:369–381.

24. Radrizzani M, Soncin S, Bolis S, Lo Cicero V, Andriolo G, Turchetto L: **Quality control assays for clinical-grade human mesenchymal stromal cells: validation strategy**. *Methods Mol Biol* 2016, **1416**:339–356.

25. Conti S, O'Hagan A: **Bayesian emulation of complex multi-output and dynamic computer models**. *J Stat Plann Inference* 2010, **140**:640–651.

26. Liu S, Johns E, Davison AJ: **End-to-end multi-task learning with attention**. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*; 2019:1871–1880.

27. * Zhang Y, Yang Q: **A survey on multi-task learning**. *IEEE Trans Knowl Data Eng* 2021.
This paper reviews many state-of-the-art multi-task learning methods, which has the potential for donor-to-donor variability therapeutic cell manufacturing.

28. Devlin J, Chang M-W, Lee K, Toutanova K: **Bert: pre-training of deep bidirectional transformers for language understanding**. In *Proceedings of NAACL-HLT 2019*; 2019:4171–4186.

29. Finn C, Abbeel P, Levine S: **Model-agnostic meta-learning for fast adaptation of deep networks**. In *International conference on machine learning*. PMLR; 2017:1126–1135.

30. Misra I, Shrivastava A, Gupta A, Hebert M: **Cross-stitch networks for multi-task learning**. In *Proceedings of the IEEE conference on computer vision and pattern recognition*; 2016:3994–4003.

31. Xie Y, Jiang H, Liu F, Zhao T, Zha H: **Meta learning with relational information for short sequences**. In *33rd Conference on Neural Information Processing Systems (NeurIPS 2019)*; 2019.

32. Kennedy MC, O'Hagan A: **Predicting the output from a complex computer code when fast approximations are available**. *Biometrika* 2000, **87**:1–13.

33. Lawrence ND, Platt JC: **Learning to learn with the informative vector machine**. In *Proceedings of the twenty-first international conference on Machine learning*, vol. 65; 2004.

34. Wan H-P, Ni Y-Q: **Bayesian multi-task learning methodology for reconstruction of structural health monitoring data**. *Struct Health Monit* 2019, **18**:1282–1309.

35. Raftery AE, Madigan D, Hoeting JA: **Bayesian model averaging for linear regression models**. *J Am Stat Assoc* 1997, **92**:179–191.

36. Ki Williams C: *Gaussian processes for machine learning*. Taylor & Francis Group; 2006.

37. Carlin BP, Louis TA: *Bayes and empirical Bayes methods for data analysis*. Chapman & Hall/CRC; 2000.

38. Li X, Li C, Chi J, Ouyang J, Wang W: **Black-box expectation propagation for Bayesian models**. In *Proceedings of the 2018 siam international conference on data mining*. SIAM; 2018:603–611.

39. Greff K, Kaufman RL, Kabra R, Watters N, Burgess C, Zoran D, Matthey L, Botvinick M, Lerchner A: **Multi-object representation learning with iterative variational inference**. In *International conference on machine learning*. PMLR; 2019:2424–2433.

40. Bengio Y, Courville A, Vincent P: **Representation learning: a review and new perspectives**. *IEEE Trans Pattern Anal Mach Intell* 2013, **35**:1798–1828.

41. Alley EC, Khimulya G, Biswas S, AlQuraishi M, Church GM: **Unified rational protein engineering with sequence-based deep representation learning**. *Nat Methods* 2019, **16**:1315–1322.

42. Zhang H, Goodfellow I, Metaxas D, Odena A: **Self-attention generative adversarial networks**. In *International conference on machine learning*. PMLR; 2019:7354–7363.

43. * Chen J, Xie Y, Wang K, Zhang C, Vannan MA, Wang B, Qian Z: **Active image synthesis for efficient labeling**. *IEEE Trans Pattern Anal Mach Intell* Nov. 2021, **43**:3770–3781.
This paper discussed a data analysis framework which can be viewed as a representation learning method with CT images and can be used in cell manufacturing applications.

44. Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, Neelakantan A, Shyam P, Sastry G, Askell A: **Language models are few-shot learners**. *arXiv preprint arXiv:2005.14165* 2020.