Immortalitas in Hac Vita

Longevity Platform

https://www.longevity.ge

Longevity Horizon
ISSN:3088-4063

2026

Vol.2 No.4:19

Digital Twins of Cell Differentiation Integrating Centrosomal Dynamics

A computational framework incorporating centriole asymmetry into lineage fate modeling

Jaba Tkemaladze [^] [1]

Affiliation: [1] Kutaisi International University, Georgia

Citation: Tkemaladze, J. (2026). Digital Twins of Cell Differentiation Integrating Centrosomal Dynamics. Longevity Horizon, 2(4). DOI : https://doi.org/10.65649/b3xxd095

Abstract

The integration of centrosomal dynamics into digital twins (DTs) of cell differentiation represents a paradigm shift from descriptive, empirical protocols to predictive, mechanism-based engineering of cell fate. Digital twins in biomedicine act as dynamic, multi-scale virtual systems that integrate experimental data and generate testable predictions, yet most models overlook the centrosome—an organelle increasingly recognized as an active signal-integration hub rather than a passive microtubule-organizing center. This review synthesizes evidence that centrosomes orchestrate cell fate decisions through asymmetric inheritance, CAFD (cell fate determinant) docking, primary ciliogenesis, and coordination with polarity complexes. We propose a six-layer digital twin architecture that incorporates the centrosome as a mandatory computational module at the perception–integration level and CAFD logistics at the intracellular transport level. Data acquisition analysis shows that super-resolution live-cell imaging, APEX2 proximity proteomics, FRET-based biosensors, and imaging mass cytometry provide multiparametric data to parameterize centrosome-specific dynamics. Computational frameworks include agent-based modeling with explicit centrosomal state vectors, hybrid LSTM–CNN architectures for multimodal time-series integration, graph neural networks for CAFD–centrosome interactomes, and reinforcement learning for closed-loop protocol optimization. A case study optimizing cortical neuron differentiation from iPSCs via Aurora A kinase inhibition illustrates implementation from data acquisition through hybrid model training. Key challenges include dimensionality, computational demands, incomplete CAFD inventories, and validation costs. Ethical issues of data sovereignty, liability, and reproducibility remain. Over three to five years, protocol twins will enable pharmaceutical in silico screening; within five to ten years, cloud platforms with pretrained centrosomal modules will democratize access; beyond ten years, integrated virtual cells may support Human Digital Twin initiatives. This interdisciplinary effort will accelerate predictable, safe cell therapies by transforming trial-and-error into centrosome-informed engineering.

Keywords: Digital Twin, Cell Differentiation, Centrosome Dynamics, Agent-Based Modeling, Multi-Omics Integration, Machine Learning, Regenerative Medicine.

Introduction: Concept and Scope of the Problem

The concept of the digital twin (DT)—a dynamic, virtual representation of a physical system that mirrors its state and predicts its behavior—has rapidly transitioned from an engineering paradigm to a transformative approach in biomedicine. In the context of cell differentiation, a digital twin represents a multi-scale computational model that reflects, in real time or with defined temporal resolution, the evolving state of a biological system—from a single stem cell to a heterogeneous population—as it progresses along distinct lineage trajectories. Unlike conventional static models or mere data repositories, a DT is characterized by its bidirectionality: it not only integrates continuous streams of experimental data to update its parameters but also generates testable predictions about future system states under genetic or environmental perturbations (Bhardwaj et al., 2025). The ambition of creating "virtual cells" or "digital cellular twins" has gained significant momentum, driven by the convergence of high-throughput omics technologies, advanced imaging, and artificial intelligence (Bhardwaj et al., 2025; Han et al., 2025).

However, constructing a digital twin of cell differentiation presents unique and formidable challenges that extend beyond those of modeling a static or terminally differentiated cell. These challenges can be categorized into four interconnected domains.

First, the multi-omic integration required is staggering. Differentiation involves coordinated changes across multiple molecular layers: the epigenome (chromatin accessibility, histone modifications), transcriptome (coding and non-coding RNA expression), proteome (protein abundance, post-translational modifications), and metabolome. A faithful DT must not only integrate these disparate data types but also capture the causal relationships between them—for instance, how an epigenetic change leads to a transcriptional cascade, which in turn remodels the proteomic landscape (Bhardwaj et al., 2025; Li, 2025).

Second, the model must account for spatiotemporal dynamics. Differentiation is not a static event but a choreographed process unfolding in both space and time. This includes the changing localization of fate-determining molecules (such as cell fate determinants, CAFDs), the morphological remodeling of the cell (e.g., polarization, process outgrowth), and the establishment and dissolution of cell-cell contacts within a tissue microenvironment. Recent advances in spatial omics and live-cell imaging are beginning to provide the granular data needed to constrain such models, revealing, for example, how signaling molecules are asymmetrically trafficked and localized during asymmetric cell division (Wang et al., 2025; Han et al., 2025).

Third, any realistic DT must contend with stochasticity. Even in clonal populations under identical conditions, individual cells exhibit significant variability in their differentiation trajectories. This cell-to-cell heterogeneity arises from intrinsic noise in gene expression, the random partitioning of cellular components during division, and the low copy numbers of key regulatory molecules. A deterministic model will inevitably fail to capture the probabilistic nature of fate commitment. Recent machine learning frameworks, such as neural stochastic differential equations applied to tissue dynamics, offer promising avenues for learning and simulating these

noisy, multi-scale processes directly from experimental data (Han et al., 2025; Ramírez Sierra, 2024).

Finally, and most critically for this review, there is the non-canonical and often overlooked role of organelles—in particular, the centrosome. Historically viewed primarily as the microtubule-organizing center (MTOC), the centrosome has emerged as a central hub for coordinating cell fate decisions, especially in the context of asymmetric cell division (ACD) (Doxsey, 2016). Far from being a passive structural element, the centrosome is now understood to be an active regulator of signaling, membrane traffic, and proteostasis, with its asymmetric inheritance directly linked to the differential fate of daughter cells. This review posits that integrating centrosome-mediated regulatory logic is not merely an optional refinement but a prerequisite for building predictive digital twins of differentiation.

The objective of this review is therefore to critically analyze the methodological landscape for constructing digital twins of cell differentiation, with a specific focus on how these models can and must incorporate the emerging complexity of centrosomal dynamics. We will examine how recent discoveries regarding centrosome asymmetry, its role in endosome trafficking and Notch signaling, and its integration with polarity complexes can be formalized into computational frameworks. By bridging the gap between high-resolution cell biology and multi-scale modeling, we aim to outline a roadmap toward digital twins that can truly recapitulate the organelle-level logic of cell fate determination.

Digital Twin Architecture of the Differentiation Process: A Multi-Layered Model

A predictive digital twin of cell differentiation must transcend simplistic, single-scale representations and embrace the intrinsic multi-level organization of biological reality. The architecture proposed here is founded on the principle that cell fate decisions emerge from reciprocal feedback loops spanning molecular, organellar, cellular, and tissue-level processes. While existing multi-scale models have made significant progress in integrating transcriptomic and proteomic data (Bhardwaj et al., 2025), they have largely neglected the central role of organelles as active computational nodes. The framework presented below explicitly incorporates the centrosome not as a passive structural element but as an active signal integration and logistics hub—a critical innovation for capturing the asymmetric inheritance patterns that underpin stem cell fate (Reina & Gonzalez, 2014; Yamashita et al., 2007).

The proposed architecture comprises six interconnected hierarchical levels, each characterized by distinct state variables, spatiotemporal scales, and data modalities.

LEVEL 0: Input Layer – Extrinsic Signals

This basal layer defines the boundary conditions impinging upon the cell from its microenvironment. State variables include:

- Morphogen concentrations and gradients (e.g., Wnt, Shh, BMP)

- Mechanical signals: extracellular matrix stiffness, shear stress, tissue-level strains
- Juxtacrine signals: identity and contact duration with neighboring cells
  These inputs are not static boundary conditions but dynamic variables that evolve with tissue morphogenesis (Meyer-Gerards et al., 2025).

LEVEL 1: Perception and Integration Layer – The Centrosomal Hub

This layer represents the first intracellular processing stage, where extracellular signals are transduced and integrated. Critically, it positions the centrosome and its associated cilium as a central signal integration platform. Recent evidence demonstrates that the centrosome nucleates the destruction complex for β-catenin, thereby directly regulating Wnt signal transmission (Lach et al., 2022). State variables here include:

- Activation states of signaling cascades (Notch, Hedgehog, Wnt)
- Mechanical force transduction at integrins and the nuclear envelope
- Centrosomal state vector: position relative to nucleus and polarity axis, mother versus daughter centriole age, post-translational modification status of centrosomal proteins (Fernandes-Mariano et al., 2025)
- Recruitment of signaling adapters (e.g., Dishevelled, PDGFRα) to the centriole
  The asymmetric inheritance of mother versus daughter centrosomes, first documented in Drosophila male germline stem cells (Yamashita et al., 2007), provides a mechanistic link between organelle age and differential daughter cell fate—a correlation now observed across multiple stem cell systems (Reina & Gonzalez, 2014).

LEVEL 2: Intracellular Logistics Layer – Cytoskeletal Transport and Condensate Dynamics

This layer governs the spatiotemporal distribution of fate determinants. Microtubules, nucleated at the centrosome, serve as the primary railway network for directed transport. Key processes include:

- Dynein-mediated transport of mRNAs and proteins toward the minus-end at the centrosome
- Asymmetric partitioning of cell fate determinants (CAFDs) during division
- Liquid-liquid phase separation (LLPS) of ribonucleoprotein (RNP) granules at or near the centrosome

The centrosome itself has been recognized as a membraneless organelle whose material properties—ranging from liquid-like to gel-like states—emerge from LLPS of pericentriolar material proteins (Woodruff et al., 2017; Tanaka, 2021). Furthermore, mRNA localization to centrosomes, regulated by proteins such as Fragile-X Mental Retardation Protein (FMRP),

enables local translation and facilitates error-free mitosis (Lee et al., 2020). The dynamic binding and release of CAFDs from the centriole surface constitute a critical regulatory logic that must be explicitly modeled. The transport kinetics can be described by advection-diffusion-reaction equations incorporating motor protein processivity:

$$\partial C \partial t = D \nabla 2C - v \cdot \nabla C + R(C),$$

where C represents local concentration of a fate determinant, D its diffusion coefficient, v the velocity field generated by motor-driven transport along microtubules, and

R(C) reaction terms accounting for binding/unbinding at the centrosome.

LEVEL 3: Nuclear Regulatory Layer – Epigenetic and Transcriptional State

This layer captures the nuclear response to upstream signals. Inputs from Levels 1 and 2 converge to modulate chromatin accessibility and transcription. State variables integrate:

- Chromatin conformation and accessibility (informed by ATAC-seq)
- Transcription factor occupancy at enhancers and promoters
- Nascent and steady-state transcriptomes (informed by scRNA-seq)
- Epigenetic marks (histone modifications, DNA methylation)
  The physical connection between the centrosome and the nucleus via the LINC complex (linker of nucleoskeleton and cytoskeleton) ensures mechanical coupling, positioning the nucleus relative to polarity cues (Meyer-Gerards et al., 2025).

LEVEL 4: Cellular Phenotypic Layer – Morphology and Cytoarchitecture

This layer represents the emergent morphological consequences of molecular and organellar dynamics. State variables include:

- Cell shape, volume, and polarity axis orientation
- Cytoskeletal organization (actin stress fibers, microtubule network architecture)
- Cell cycle phase and progression
- Position within the tissue and orientation of the mitotic spindle
  The centrosome position specifies the division plane by determining spindle orientation; in polarized cells, centrosome alignment along the polarity axis enables asymmetric division (Reina & Gonzalez, 2014). Quantitative imaging of centrosome migration in model systems such as MDCK cysts provides critical parameters for constraining models at this level (Wang et al., 2025).

LEVEL 5: Population Layer – Collective Behavior

The apical layer captures tissue-level phenomena emerging from individual cell behaviors:

- Cell-to-cell heterogeneity and its propagation over divisions
- Paracrine signaling and niche competition
- Collective migration and tissue morphogenesis
  This layer must account for the fact that different tissues possess varying thresholds and mount graded responses to centrosome loss-of-function, as demonstrated in developing mouse embryos (Meyer-Gerards et al., 2025).

Key Innovation: The architecture explicitly mandates the inclusion of Level 1 (centrosomal hub) and Level 2 (CAFD logistics) as mandatory computational modules rather than optional refinements. This is motivated by a growing body of evidence that centrosomes act as "information hubs" where signaling, transport, and phase separation converge to bias daughter cell fate. The material state of the centrosome itself—whether liquid, gel, or lattice—determines its ability to resist microtubule-mediated forces and maintain asymmetric positioning (Woodruff et al., 2017; Tanaka, 2021). By embedding these organelle-specific mechanisms within a multi-scale framework, the proposed DT architecture moves beyond descriptive correlation toward mechanistic prediction of how subcellular asymmetry propagates to tissue-level patterning.

3. Meta-Analysis of Data Acquisition Methods for Twin Feeding

The construction of a predictive digital twin of cell differentiation—particularly one incorporating centrosomal dynamics—requires a multi-parametric data ecosystem that spans molecular, organellar, and tissue scales. The fidelity of the twin is fundamentally constrained by the quality, temporal resolution, and multiplexing capacity of the experimental data used for parameterization and validation. This section provides a critical meta-analysis of state-of-the-art methodologies for acquiring centrosome-specific and integrative multi-omic data, evaluating their respective strengths, limitations, and applicability for feeding the layered architecture proposed above.

Methods for Centrosome-Specific Data Acquisition

The centrosome presents unique experimental challenges due to its small size (~200-500 nm), dynamic regulation, and membraneless organization. Capturing its state requires approaches that combine nanoscale spatial resolution with temporal dynamics and proteomic breadth.

Super-Resolution Live-Cell Microscopy (STED/SIM). The dynamic behavior of centrosomes during differentiation—their migration, separation, maturation, and asymmetric inheritance—demands imaging modalities that circumvent the diffraction limit of light while

preserving cell viability. Stimulated emission depletion (STED) microscopy and structured illumination microscopy (SIM) have emerged as powerful tools for visualizing centriole duplication and the assembly of distal appendages with ~50-100 nm resolution (Kiyomitsu, 2025). A typical protocol involves cells co-expressing fluorescently tagged centriolar markers (e.g., Centrin-GFP for the distal lumen, CEP164-mScarlet for distal appendages) imaged every 2-5 minutes over 48-72 hours of differentiation. The major challenge resides in the data volume: a single multi-dimensional experiment can generate terabytes of images, necessitating robust computational pipelines for automated centriole detection, tracking, and feature extraction. Furthermore, prolonged light exposure can induce phototoxicity, potentially altering differentiation kinetics—a confounder that must be carefully controlled.

APEX2 Proximity Proteomics with Temporal Resolution. Understanding how the centrosomal microenvironment remodels during differentiation requires unbiased mapping of its changing protein composition. Proximity labeling using engineered ascorbate peroxidase (APEX2) enables the covalent tagging of endogenous proteins within ~20 nm of a bait protein following hydrogen peroxide (H_2O_2) treatment (Lach et al., 2022). For centrosome-specific applications, cells expressing APEX2 fused to a core centriolar protein (e.g., CEP152) are induced to differentiate, with biotinylation pulses applied at critical time points (T0, T6, T24 hours). Subsequent streptavidin enrichment and mass spectrometry yield quantitative temporal profiles of centrosome-associated proteins, including cell fate determinants (CAFDs) that transiently dock at the organelle. A recent methodological advance—in situ APEX activation (iAPEX)—overcomes the toxicity of exogenous H_2O_2 by coupling APEX2 with a cilia-localized D-amino acid oxidase (DAAO) that generates H_2O_2 locally from D-alanine (Guo et al., 2025a). This cascade approach reduces oxidative damage and enables proteomic profiling in sensitive primary cell types previously inaccessible to conventional APEX labeling (Guo et al., 2025a). The TransitID method further extends this concept by using orthogonal proximity labeling enzymes (TurboID and APEX2) targeted to source and destination compartments, enabling unbiased mapping of protein trafficking events—such as CAFD movement from cytoplasm to centrosome—with nanoscale spatial resolution (Guo et al., 2025b).

FRET Sensors for Centrosome-Specific CAFD Activity. While proximity proteomics reveals composition, it does not report on the conformational state or activation status of CAFDs at the centrosome. Förster resonance energy transfer (FRET)-based biosensors address this gap. An exemplar design fuses a centriolar anchor protein (e.g., CEP164) to a donor fluorophore (YFP) and a CAFD of interest (e.g., STAT3, β-catenin) to an acceptor fluorophore (CFP). Upon CAFD binding to the centrosome, the close proximity enables FRET, with the energy transfer efficiency reporting on binding stoichiometry or conformational changes induced by upstream signaling. The FRET efficiency (E) can be calculated from donor fluorescence in the presence (IDA) and absence (ID) of acceptor:

$$E = 1 - (I_{DA} / I_D)$$

Acceptor photobleaching FRET assays have been successfully employed to validate protein interactions at the centrosome, such as the binding between CCHCR1 and OFD1 or PCM1 during ciliogenesis (Huang et al., 2025). For digital twin parameterization, ratiometric FRET

imaging over time provides a continuous readout of CAFD engagement at the centrosomal hub, directly informing the "binding/unbinding" state variables proposed in Level 2 of the architecture.

Imaging Mass Cytometry (IMC). Bridging the gap between single-cell resolution and population-level heterogeneity, imaging mass cytometry combines the multiplexing capacity of mass spectrometry with spatial imaging. Using metal-conjugated antibodies against 40+ markers—including centriolar proteins (CEP164, γ-tubulin), CAFDs (β-catenin, Notch intracellular domain), differentiation markers (SOX2, PAX6), and cell cycle regulators (Ki-67)—IMC generates high-dimensional spatial maps of fixed cell populations at defined differentiation time points. This technology is particularly valuable for capturing the stochastic distribution of centrosome states across hundreds of cells and correlating these with differentiation outcomes, providing population-level constraints for Level 5 of the DT.

Methods for Integrative Multi-Omic and Functional Data

Centrosome-specific measurements must be contextualized within the broader cellular state, requiring integration of transcriptomic, epigenetic, metabolic, and electrophysiological data.

Single-Cell Multi-Omics. The simultaneous profiling of multiple molecular layers from the same cell is essential for resolving the causal flow from epigenetic remodeling to transcriptional output. Technologies such as SHARE-seq (simultaneous measurement of chromatin accessibility and gene expression) and CITE-seq (combining transcriptome with surface protein detection via oligonucleotide-conjugated antibodies) now enable routine paired profiling (Qi et al., 2025). CITE-seq has been successfully applied to create multi-omic atlases of complex tissues, revealing coordinated RNA-protein dynamics and enabling the identification of cell-surface markers that distinguish functionally distinct subpopulations (Qi et al., 2025; Abdelmohsen et al., 2025). For differentiation time series, collecting scRNA-seq and scATAC-seq data at matched temporal intervals allows reconstruction of gene regulatory networks (GRNs) and inference of transcription factor activity dynamics. Computational frameworks such as PoE-VAE (Product-of-Experts Variational Autoencoder) leverage deep learning to integrate these modalities into a shared latent space, preserving cell type heterogeneity while removing technology-specific biases and enabling seamless mapping of unimodal queries onto multimodal reference atlases (Litinetskaya et al., 2025). These integrated embeddings can serve as the nuclear regulatory layer (Level 3) of the DT.

Spatial Transcriptomics and Proteomics. When differentiation occurs within three-dimensional contexts—organoids, embryoid bodies, or tissue explants—spatial information becomes critical. Platforms such as 10x Visium (transcriptomics) and Nanostring GeoMx DSP (proteomics) enable profiling of gene or protein expression while preserving tissue architecture (Du et al., 2025). Recent applications to tumor spheroids have revealed concentric gene expression gradients and region-specific metabolic heterogeneity that would be invisible to dissociated cell analysis (Du et al., 2025). For centrosome-focused DTs, spatial methods can map how centrosome positioning relative to the apical surface (measured by IMC or immunofluorescence) correlates with niche-proximal versus niche-distal differentiation fates. Integrated protocols combining scRNA-seq, organoid culture, and spatial RNA-protein analysis

in the same tissue system are now available, enabling multi-scale validation of differentiation trajectories (Bian et al., 2025).

Real-Time Electrophysiology and Metabolomics. Functional validation of differentiation states—particularly for excitable cell types such as cardiomyocytes or neurons—requires dynamic physiological measurements. The Seahorse XF Analyzer quantifies real-time oxygen consumption rate (OCR) and extracellular acidification rate (ECAR), reporting on mitochondrial oxidative phosphorylation versus glycolytic flux. Optimized workflows for Seahorse XF Flex Analyzer now enable accurate quantitative characterization of both glycolytic and mitochondrial function during in vitro differentiation, with improved dynamic range that prevents hypoxia in highly metabolically active differentiated cells (Wang et al., 2025). The mitochondrial respiration parameters can be modeled as:

$$OCR = OCR_{basal} - OCR_{inhibited}$$

Parallel microelectrode array (MEA) recordings capture spontaneous electrical activity and network formation. These functional readouts provide essential validation for the phenotypic layer (Level 4) and ensure that the DT's predicted cellular states correspond to physiologically meaningful endpoints.

Temporal Synchronization and Data Integration Challenges. A critical consideration for DT construction is the temporal alignment of datasets acquired across different modalities and time scales. While live-cell imaging captures continuous dynamics at minute-scale resolution, proximity proteomics and multi-omics typically provide snapshots at discrete time points. Computational approaches for temporal integration must account for these asynchronous measurements, using interpolation methods or mechanistic models to align datasets. Furthermore, the integration of centrosome-specific measurements (nanometer scale, seconds to hours) with population-level data (micrometer to millimeter scale, hours to days) requires careful attention to scale bridging—a challenge that can be addressed through hierarchical Bayesian modeling or multi-scale neural architectures (Han et al., 2025).

Review of Computational Methods and Platforms

The transformation of multi-scale experimental data into a predictive digital twin requires a sophisticated computational framework capable of capturing deterministic kinetics, stochastic heterogeneity, and spatial organization. This section provides a critical overview of the computational approaches available for constructing digital twins of cell differentiation, with particular emphasis on their suitability for integrating centrosome-specific dynamics.

Deterministic Models (ODE/PDE Systems)

Ordinary differential equation (ODE) and partial differential equation (PDE) models represent the classical approach for modeling biochemical kinetics and morphogen gradients. These continuum-based methods describe how concentrations of molecular species change over time and space through systems of differential equations.

Applications in Differentiation Modeling. ODE systems are well-suited for modeling the kinetics of signaling pathways—the temporal dynamics of Wnt, Notch, and Hedgehog cascade components—and the formation of morphogen gradients in tissues (Iber et al., 2013). PDE-based reaction-diffusion models have been extensively used to study pattern formation, including Turing-type mechanisms underlying digit patterning during limb development (Iber et al., 2013).

Centrosome-Relevant Example. A deterministic model of centrosome-mediated fate regulation could be formulated as a system of ODEs tracking:

- Concentration of active PLK4 at the centriole, governing centriole duplication timing
- Rate of CAFD recruitment to the centrosome as a function of upstream kinase activity (e.g., YAP recruitment modulated by LATS kinase and cell density)
- Rate of CAFD release and nuclear translocation following centrosomal docking

These relationships can be expressed as:

d[CAFD_centrosome]/dt = k_on * [CAFD_cytoplasm] * f(LATS_activity) - k_off * [CAFD_centrosome]

d[CAFD_nuclear]/dt = k_transport * [CAFD_centrosome] - k_degradation * [CAFD_nuclear]

Limitations. While ODE/PDE models provide mathematical rigor and analytical tractability, they face significant limitations for digital twin applications. First, they assume well-mixed compartments or continuous fields, failing to capture the discrete nature of small-number molecules and organelles. Second, they are inherently deterministic and cannot account for the cell-to-cell heterogeneity and stochasticity that characterize differentiating populations (Camacho Gómez et al., 2023). Third, scaling these models to incorporate the full complexity of multi-omic data remains challenging.

Agent-Based Modeling (ABM)

Agent-based modeling has emerged as one of the most promising approaches for digital twin construction due to its ability to represent individual cells as autonomous agents with internal states and behavioral rules (Kemkar et al., 2024). Each "agent" corresponds to a single cell that follows programmed rules governing its interactions with other cells and the microenvironment.

Agent Architecture for Centrosome-Integrated DT. The internal state of a cell agent must explicitly incorporate a centrosomal module. A representative class definition in Python-style pseudocode:

```
class CellAgent:
    # Molecular state (simplified)
    gene_expression_profile = {'SOX2': 0.8, 'PAX6': 0.1, 'NESTIN': 0.6}
```

```
    # CENTROSOMAL MODULE (CRITICAL)
    centriole_state = {
        'mother': {'maturity': 0.95, 'age': 5, 'associated_CAFDs': ['YAP',
'STAT3']},
        'daughter': {'maturity': 0.35, 'age': 1, 'associated_CAFDs': []},
        'cilium_present': True,
        'MTOC_activity': 0.8,
        'position_relative_nucleus': (0.2, 0.1, 0.0),
        'polarity_vector': (0.95, 0.31, 0.0)
    }

    # Signaling states
    wnt_signal_active = True
    notch_signal_received = False
    mechanical_stress = 0.3
```

Behavioral Rules. Agent rules encode biological decision-making in human-readable syntax, as recently demonstrated in the PhysiCell framework (Johnson et al., 2025). A rule governing centrosome-dependent fate choice might read:

```
"IF association_of_YAP_with_centrosome > threshold_0.7 AND
differentiation_signal_Wnt_received = True, THEN with probability 0.4
initiate expression_of_neural_marker_PAX6"
```

This grammatical approach, implemented in platforms like PhysiCell, makes modeling accessible to biologists without extensive programming expertise (Johnson et al., 2025; Technology Networks, 2025). The software automatically translates such spreadsheets of rules into mathematical equations that drive agent behavior.

Available Platforms. Multiple ABM platforms support digital twin development:

- **PhysiCell**: An open-source agent-based framework with built-in support for spatial transcriptomics integration and mechanical interactions (Johnson et al., 2025)
- **CompuCell3D**: Specialized for morphogenesis simulations with cellular Potts model foundation
- **Chaste**: A C++ library for computational physiology and biology (Kemkar et al., 2024)
- **BioDynaMo**: A modular platform for high-performance agent-based simulation
- **Python-based frameworks**: Mesa for lightweight prototyping; custom implementations for specialized needs

Hybrid Models

Hybrid models combine the strengths of different computational approaches to address multi-scale problems. The most common hybrid architecture for tissue-level digital twins integrates agent-based modeling for discrete cells with PDEs for continuous fields (Camacho Gómez et al., 2023; Diez et al., 2025).

ABM + PDE Integration. In this framework, individual cell agents reside in and interact with a continuous environment described by PDEs. For example:

- Diffusion of morphogens (SHH, BMP, FGF) through extracellular space is governed by reaction-diffusion PDEs
- Cells respond to local morphogen concentrations based on their receptor status (e.g., presence of primary cilium, a centrosome-derived structure)
- Cells secrete factors that contribute to the PDE fields as source terms

Centrosome-Relevant Example. During brain organoid development, SHH morphogen gradients (modeled by PDE) establish positional information. Individual neural progenitor cells detect SHH via their primary cilia—antenna-like structures emanating from the mother centriole. A hybrid model would represent:

- PDE: $\partial[SHH]/\partial t = D\nabla^2[SHH] - \lambda[SHH] + \Sigma\ S\_cell(t)\delta(x - x\_cell)$
- Agent property: cilium_present (Boolean) determined by centriole_state['cilium_present']
- Decision rule: SHH_response = integrate([SHH]_local) * cilium_present * receptor_sensitivity

Recent work demonstrates the power of hybrid approaches for understanding digit formation, where agent-based models incorporating mechanical laws and simple signaling can recapitulate symmetry breaking and morphogenesis observed in limb-like organoids (Diez et al., 2025).

Machine Learning and Neural Network Approaches

Machine learning methods, particularly deep learning, offer powerful tools for learning patterns from high-dimensional data and making predictions about cell fate.

Predicting Differentiation Outcomes. Recurrent neural networks (RNNs), long short-term memory networks (LSTMs), and Transformers are well-suited for time-series prediction tasks. A model for fate prediction might take as input:

- Day 0 gene expression profile (from scRNA-seq)
- Time-series of environmental perturbations

- Centrosomal parameters extracted from imaging (Centrin-GFP intensity, mother-daughter asymmetry coefficient)

The output would be a probability distribution over possible fates at day 10 (e.g., 0.7 cardiomyocyte, 0.2 fibroblast, 0.1 other).

Recent advances include neural stochastic differential equation frameworks (END-nSDE) that can reconstruct noisy dynamics from trajectory data while accounting for both intrinsic stochasticity and extrinsic cellular heterogeneity (Zhang et al., 2025). These approaches outperform traditional RNNs and LSTMs for modeling biological processes like circadian rhythms and NF-κB signaling.

FateNet represents another innovative integration of dynamical systems theory and deep learning, using universal properties of bifurcations to predict cell fate decisions from scRNA-seq data (Sadria & Bury, 2024). By training on simulations of diverse dynamical systems, FateNet learns to detect critical transitions and classify bifurcation types without requiring system-specific training data.

Generative Models for In Silico Screening. Generative adversarial networks (GANs) and variational autoencoders (VAEs) can be trained on successful differentiation experiments to generate hypothetical optimal protocols. These models could propose sequences of growth factor additions and centrosome-targeting compounds that maximize desired differentiation outcomes, enabling virtual screening before wet-lab validation.

Graph Neural Networks (GNNs)

Graph neural networks provide a natural framework for modeling molecular interaction networks, including the protein-protein interaction networks centered on the centrosome.

Modeling CAFD-Centrosome Interactions. The centrosomal interactome can be represented as a graph where:

- Nodes represent molecules (CAFDs, centriolar proteins, signaling mediators)
- Edges represent physical or functional interactions
- Edge weights can be time-varying, reflecting dynamic binding affinities modulated by phosphorylation or conformational changes

A GNN trained on this graph structure can predict how perturbations propagate through the network. For example, if a small molecule inhibitor targets a specific centrosomal protein (e.g., PLK4), the GNN can predict which downstream CAFDs (YAP, β-catenin, STAT3) will be affected and how this will influence differentiation marker expression.

Recent work has demonstrated the power of graph-based neural architectures for modeling tissue dynamics, with dual signaling graphs capable of handling signaling cascades and substantially reducing training data requirements compared to convolutional networks (Han et

al., 2025). These approaches represent cells as edges in an evolving graph, with cell interactions encoded in the graph structure.

Integration and Platform Selection

The choice of computational platform depends on the specific biological questions, available data, and desired predictive capabilities. For centrosome-integrated digital twins of differentiation, a hybrid approach combining ABM for cellular heterogeneity, PDEs for morphogen gradients, and machine learning for parameter inference and prediction offers the most comprehensive framework. Platforms like PhysiCell, with its accessible grammatical interface and support for spatial transcriptomics integration (Johnson et al., 2025), provide an excellent foundation for community-wide adoption and collaborative development of centrosome-aware digital twins.

Concrete Example: Digital Twin for Cortical Neuron Differentiation from iPSCs

The theoretical frameworks and computational methods discussed above find their ultimate test in concrete biological applications. Here, we present a detailed case study demonstrating how a centrosome-aware digital twin can be constructed to optimize a clinically relevant differentiation protocol: the generation of cortical neurons from human induced pluripotent stem cells (iPSCs) for disease modeling or cell replacement therapies.

Biological Rationale and Therapeutic Goal

Cortical neurons derived from iPSCs hold tremendous promise for studying neurodevelopmental disorders—such as microcephaly, autism spectrum disorders, and schizophrenia—and for developing cell-based therapies for neurodegenerative conditions (Eschment et al., 2025). However, current differentiation protocols suffer from inefficiency, variability, and incomplete maturation, limiting their translational utility. A critical bottleneck lies in the regulation of primary ciliogenesis during neural progenitor expansion.

Primary cilia—antenna-like organelles extending from the mother centriole—function as signaling hubs for key developmental pathways including Sonic Hedgehog (Shh), Wnt, and PDGF signaling (Eschment et al., 2025; He et al., 2014). During cortical neurogenesis, neural progenitors must first assemble primary cilia to receive appropriate differentiation cues, then disassemble them to re-enter the cell cycle or undergo terminal differentiation. Aurora kinase A (AurA) plays a central role in ciliary disassembly by activating HDAC6, which deacetylates tubulin and promotes ciliary resorption (He et al., 2014). Notably, AurA-p53 signaling has been shown to regulate pluripotency and differentiation in both embryonic and induced pluripotent stem cells (Lee et al., 2012).

Alisertib (MLN8237), a selective AurA inhibitor currently in clinical trials for cancer, has emerged as a potential tool for modulating ciliogenesis timing. By transiently inhibiting AurA, we might

extend the ciliary signaling window in neural progenitors, potentially enhancing their responsiveness to differentiation cues and improving neuronal yield. However, the optimal timing and dosage remain unknown—too early inhibition might block necessary ciliary disassembly for cell cycle progression, while too late or insufficient inhibition might miss the critical signaling window.

Objective: Build a digital twin to predict the optimal timing and dosage of Alisertib addition to an established cortical neuron differentiation protocol, maximizing the yield of TBR1+ deep-layer cortical neurons at day 30.

Training Data Acquisition (Wet-Lab Phase)

Constructing the digital twin begins with systematic data collection across multiple experimental conditions and time points.

Experimental Design. Twenty variant protocols are executed in parallel, varying two key parameters:

- Alisertib addition time: days 2, 3, 4, 5, 6 of differentiation (5 levels)
- Alisertib concentration: 0 nM (control), 10 nM, 50 nM, 250 nM (4 levels)

This 5 × 4 factorial design yields 20 distinct perturbation conditions, plus three biological replicates per condition, totaling 60 independent differentiations.

Multi-Modal Data Collection. At four critical time points—T0 (day 0, iPSC stage), T3 (day 3, early neural progenitors), T6 (day 6, expanding progenitors), and T9 (day 9, committed neuronal precursors)—samples are collected for parallel processing:

Single-Cell Transcriptomics. scRNA-seq (10x Genomics) captures the full transcriptional landscape, enabling trajectory inference and cell type identification. Libraries are sequenced to ~50,000 reads per cell, targeting 5,000-8,000 cells per sample.

High-Content Imaging. Cells are fixed and stained for:

- Centrin-GFP (endogenously tagged or immunostained) to visualize centrioles
- ARL13B (primary cilia marker)
- PAX6 (neural progenitor marker)
- TBR1 (deep-layer post-mitotic neuron marker)
- DAPI (nuclei)

Automated confocal microscopy captures 20 fields per well at 63× magnification, generating ~12,000 images per time point.

Centrosomal Feature Extraction. Computational image analysis pipelines extract per-cell features including:

- Percentage of cells with primary cilia (ARL13B+)
- Cilium length (micrometers)
- Centrin intensity (reflecting centriole number and maturation state)
- Centriole pair asymmetry (mother-daughter intensity ratio)
- Centriole positioning relative to nucleus and apical surface

This combined approach—joint profiling of morphology and gene expression—has proven powerful for characterizing in vitro neurodevelopment, revealing complementary aspects of cellular phenotype that neither modality alone captures (Klein et al., 2025).

Model Architecture and Training (Dry-Lab Phase)

The collected multi-modal time-series data are integrated into a hybrid deep learning architecture designed to predict differentiation outcomes from initial conditions and intervention parameters.

Data Integration. For each cell at each time point, we construct a unified feature vector combining:

- Transcriptomic data: top 2,000 highly variable genes (from scRNA-seq)
- Morphological features: 223 image-derived features (following Cell Painting pipeline; Klein et al., 2025)
- Centrosomal features: 15 extracted parameters (cilium presence, length, centriole asymmetry, etc.)
- Perturbation parameters: Alisertib dose, time of addition (one-hot encoded)

Cells are aggregated by sample and time point to create population-level feature matrices, while preserving single-cell heterogeneity through distributional representations (mean, variance, quantiles).

Hybrid LSTM-CNN Architecture. The prediction model combines two complementary neural network architectures:

- LSTM stream processes the temporal sequence of transcriptomic data. The LSTM (long short-term memory) network captures dependencies across time points (T0→T3→T6→T9), learning how gene expression dynamics unfold over differentiation (Sadria & Bury, 2024). The input is a T × G matrix where T=4 time points and G=2,000 genes.

- CNN stream processes the static morphological and centrosomal features extracted at each time point. Convolutional layers learn local patterns in the feature space, identifying combinations of image-based measurements that correlate with differentiation state (Klein et al., 2025). The input is a T × F matrix where F=238 features.

The two streams are integrated through a fusion layer that combines the LSTM hidden states with CNN-derived feature representations. This fused representation captures both temporal dynamics and static phenotypic states. Finally, fully connected layers map to the output:

Output = Probability of successful differentiation (proportion of TBR1+ cells at day 30 > 0.3)

Training Procedure. The model is trained on data from 80% of experimental conditions (16 protocols) using mean squared error loss between predicted and observed TBR1+ fractions. Validation on the remaining 20% (4 protocols) monitors for overfitting. Data augmentation—including minor perturbations to image-derived features and dropout of random genes—improves generalization.

A similar hybrid CNN-LSTM architecture with cross-attention fusion has recently demonstrated superior performance for multi-omics integration in cancer subtyping, suggesting its broad applicability to biological sequence data (Wang et al., 2025).

Optimization and Reinforcement Learning Loop

With a trained predictive model, we can now search for optimal intervention protocols—a high-dimensional optimization problem well-suited to reinforcement learning.

Optimization Objective. Find the combination of Alisertib addition time (t) and dose (d) that maximizes:

$J(t,d) = E[P\_success \mid t,d] - \lambda * Cost(d)$

where P_success is the predicted probability of high TBR1+ yield, Cost(d) penalizes high drug concentrations (to minimize toxicity), and λ balances efficacy against safety.

Reinforcement Learning Framework. We frame protocol optimization as a reinforcement learning problem:

- State: Current differentiation day and observed cellular state (aggregate gene expression + centrosomal features)
- Action: Add Alisertib at specified dose, or withhold
- Reward: Final TBR1+ fraction at day 30, minus toxicity penalty
- Environment: The trained digital twin model, which simulates how the cell population evolves in response to actions

A Deep Q-Network (DQN) agent interacts with this simulated environment, exploring different intervention strategies and learning through trial-and-error to maximize cumulative reward. This approach mirrors recent successes in optimizing biomanufacturing processes through digital twin simulation (Freederia Research, 2025), demonstrating 20-30% improvements in product yield through adaptive control.

Proposed Optimal Protocol. After 10,000 training episodes, the RL agent converges on a recommended protocol:

"Add 50 nM Alisertib on day 4 of differentiation for 48 hours, followed by standard medium更换."

This prediction—intermediate timing (day 4) and moderate dose (50 nM)—reflects a trade-off between extending the ciliary signaling window in early progenitors and avoiding excessive disruption of later cell cycle progression.

Validation and Closed-Loop Refinement

The true power of the digital twin lies in its ability to learn from experimental validation and continuously improve.

Experimental Validation. The predicted optimal protocol is tested in three independent iPSC lines. Parallel controls include the standard protocol (no Alisertib) and the best-performing protocol from the original 20-condition screen. Outcome measures include:

- TBR1+ neuron fraction at day 30 (primary endpoint)
- Neuronal electrophysiological maturity (patch-clamp or MEA)
- Ciliogenesis dynamics (live imaging of ARL13B-GFP reporter lines)

Model Updating. Validation data are fed back into the digital twin, updating model parameters through transfer learning. Discrepancies between predicted and observed outcomes reveal model limitations and guide refinement—for example, if the model overpredicted efficacy, additional training data near the predicted optimum may be needed to improve local accuracy.

This closed-loop process—predict, validate, update—establishes a virtuous cycle of continuous improvement, progressively enhancing the twin's predictive power and the efficiency of the differentiation protocol.

Broader Implications

This cortical neuron case study illustrates the practical implementation of a centrosome-aware digital twin. By explicitly modeling ciliogenesis dynamics (a centrosome-dependent process) and their modulation by AurA inhibition, the twin captures organelle-level regulation that would be invisible to transcriptome-only models. The hybrid LSTM-CNN architecture integrates complementary data modalities, while reinforcement learning enables systematic optimization of multi-parameter interventions.

Beyond cortical neurons, this framework generalizes to any differentiation system where centrosome dynamics—ciliogenesis, centriole asymmetry, or CAFD docking—influence cell fate. Potential applications include pancreatic β-cell differentiation (where primary cilia regulate insulin secretion), cardiomyocyte maturation (where centrosome positioning affects polarization), and hematopoietic stem cell expansion (where centrosome age biases daughter cell fate).

The convergence of multi-modal experimental profiling, deep learning integration, and reinforcement learning optimization promises to accelerate the development of robust, efficient differentiation protocols for regenerative medicine and disease modeling.

Platforms and Infrastructure

The transition from proof-of-concept digital twins to robust, industrial-scale platforms for biomedical research and clinical translation requires a sophisticated computational infrastructure. This section addresses the architectural components necessary to support centrosome-aware digital twins of cell differentiation, from high-performance computing resources to user interfaces that democratize access for experimental biologists.

High-Performance Computing Infrastructure

The multi-scale nature of centrosome-integrated digital twins—spanning molecular dynamics at nanometer scales to tissue-level organization at millimeter scales—demands substantial computational resources. Simulations must integrate stochastic agent-based models, partial differential equations for morphogen diffusion, and deep learning inference, each with distinct hardware requirements.

Cloud Computing Platforms. Major cloud providers—Amazon Web Services (AWS), Google Cloud Platform (GCP), and Microsoft Azure—offer the elastic scalability necessary for digital twin applications. These platforms provide on-demand access to GPU-accelerated instances (NVIDIA A100, H100) for training neural networks, high-memory instances for single-cell multi-omics data integration, and large-scale storage for petabyte-scale imaging datasets. The pay-as-you-go model allows research groups to scale resources according to project needs without prohibitive upfront capital investment.

A recent exemplar from cardiovascular modeling demonstrates the power of cloud-enabled digital twins. ELEM Biotech developed virtual twin hearts running 239 simulations consuming 72,000 core hours on Oracle Cloud Infrastructure, using bare-metal instances with 36 OCPUs and 512GB RAM across multiple European regions (YellowDog, 2025). This massive calibration effort enabled validation of virtual populations against real clinical trial data, accelerating therapy safety assessment and potentially reducing animal testing. For centrosome-focused differentiation twins, similar cloud infrastructure would support parallel exploration of thousands of protocol variants (Alisertib timing, concentration, duration) across multiple iPSC lines.

Specialized Simulation Platforms. Beyond general-purpose cloud computing, specialized platforms have emerged for biological simulation. The BioDynaMo platform, developed at CERN, empowers agent-based simulations of unprecedented scale using cutting-edge computing technologies (CERN openlab, 2025). Integrated with the interTwin Digital Twin Engine, BioDynaMo supports applications ranging from pancreatic tumor modeling to retinal development studies. Its ability to handle stochastic cellular decisions through agent-based modeling makes it particularly relevant for capturing the probabilistic nature of centrosome inheritance and fate determination.

Unified Data Lake Architecture

Digital twins are fundamentally data-driven, requiring seamless integration of multi-modal experimental measurements. A unified data lake—a centralized repository storing raw and processed data in native formats—provides the foundation for model parameterization and validation.

Data Modalities and Storage Standards. The centrosome-aware digital twin must accommodate:

- Omics data: scRNA-seq, scATAC-seq, CITE-seq (gene expression matrices, peak counts, protein abundance)
- Imaging data: super-resolution microscopy (STED/SIM), high-content screens, imaging mass cytometry
- Time-series measurements: Seahorse metabolic flux, electrophysiological recordings
- Perturbation parameters: small molecule treatments, genetic modifications, culture conditions

For imaging data, standardization is critical. The Open Microscopy Environment has developed OME-TIFF and OME-NGFF (Next-Generation File Format, also called OME-ZARR) specifications that enable multi-resolution pyramidal tiling and cloud-optimized storage (ELIXIR Norway, 2025). OME-NGFF represents a significant advance, organizing data as chunked, compressed N-dimensional arrays that support partial transfer—essential for streaming large datasets without full download. The BIDS (Brain Imaging Data Structure) extension for microscopy provides a community-standard directory structure, requiring sample identifiers and enabling rich metadata annotation (BIDS Specification, 2025).

Metadata and Ontologies. Raw data alone is insufficient; comprehensive metadata must capture experimental context. The REMBI (Recommended Metadata for Biological Imaging) guidelines establish minimum information standards, including sample preparation, instrument parameters, and image analysis steps (ELIXIR Norway, 2025). Public repositories such as the Image Data Resource (IDR) and BioImageArchive provide deposition pathways with DOI assignment, ensuring long-term accessibility and citation. For digital twin applications, adherence to these standards enables model reuse and cross-laboratory validation.

Simulation and Modeling Platforms

The computational core of the digital twin requires integration of multiple modeling paradigms, each best served by specialized software platforms.

Biochemical Network Modeling. The Virtual Cell (VCell) platform, developed by the National Resource for Cell Analysis and Modeling, provides comprehensive tools for simulating reaction kinetics, diffusion, membrane transport, and electrophysiology (Virtual Cell, 2025). VCell uniquely associates biochemical reactions with experimental image data describing subcellular locations, enabling simulations within empirically derived geometries. Users can choose among ordinary differential equations, partial differential equations, stochastic kinetics, and spatial particle-based simulations. The platform supports rule-based modeling where molecular species are represented as structured objects with site-specific interactions.

COPASI complements VCell by focusing on complex biochemical reaction networks, offering parameter estimation, steady-state analysis, sensitivity analysis, and metabolic control analysis (Biocomplexity Institute, 2025). Together, these platforms—hosted at the University of Connecticut School of Medicine—serve thousands of active users and underpin over 100 publications annually. For centrosome modeling, VCell could simulate PLK4 reaction-diffusion dynamics governing centriole duplication, while COPASI could optimize parameters against time-resolved APEX2 proximity proteomics data.

Agent-Based Modeling Platforms. Multi-cell simulations require agent-based frameworks. BioDynaMo, leveraging CERN's high-performance computing expertise, enables simulations of unprecedented scale with streamlined installation and expanded documentation (CERN openlab, 2025). Recent applications include a pancreatic tumor model incorporating immune response and retinal development studies exploring stochastic cellular decisions. The platform's integration with interTwin positions it as a key component of future digital twin ecosystems.

Morpheus (distinct from the identically named NVIDIA cybersecurity framework) provides a specialized environment for multi-scale modeling of cell populations (Benzinga, 2025). However, search results primarily identified a decentralized AI platform with the same name (Benzinga, 2025) and NVIDIA's cybersecurity framework (Huang & Nandakumar, 2024)—highlighting the importance of precise nomenclature in literature searches.

Machine Learning Frameworks. Deep learning integration relies on established frameworks. TensorFlow and PyTorch provide the foundation for hybrid LSTM-CNN architectures, graph neural networks, and generative models. Recent open-source initiatives demonstrate multi-agent AI pipelines for bioprocess monitoring, using LangGraph orchestration to combine real-time data simulation with predictive viability models and interactive Gradio dashboards (Gurazada, 2025). While developed for biomanufacturing, this architecture readily adapts to differentiation digital twins, with agents representing data streams from imaging, omics, and metabolic sensors.

Digital Mirror Interface

The ultimate value of a digital twin lies in its accessibility to experimental biologists who may lack computational expertise. A well-designed user interface—the "digital mirror"—transforms complex simulations into intuitive decision-support tools.

Web-Based Interactive Platforms. Modern web technologies enable browser-based interaction with digital twins. The open-source cloud platform for biotechnology process modeling developed by DTU and collaborators exemplifies this approach, providing an integrated environment where users combine computational fluid dynamics, compartmental models, and kinetic simulations through a unified modular interface (Li et al., 2026). Standardized APIs enable seamless coupling of independent components, while the system's emphasis on accessibility and reproducibility supports both research and education.

For centrosome differentiation twins, a web interface might allow biologists to:

- Adjust protocol parameters (timing, dose) via sliders
- Visualize predicted outcomes as cell population trajectories
- Explore "what-if" scenarios (e.g., PLK4 inhibitor addition)
- Compare model predictions with experimental validation data

Virtual and Augmented Reality. Emerging interfaces leverage VR/AR for immersive data exploration. Visualizing centrosome positioning relative to polarity axes, cilia length distributions across thousands of cells, or 3D organoid architecture benefits from spatial rendering beyond flat screens. While not yet mainstream in biological digital twins, early adopters in cardiovascular modeling demonstrate the potential for intuitive interaction with complex anatomical simulations.

Integration and Workflow Orchestration

Assembling these components into a functional digital twin requires workflow orchestration—automated pipelines that manage data ingestion, model execution, and result visualization.

Multi-Agent Orchestration. The LangGraph framework, demonstrated in bioprocess monitoring applications, provides a modular architecture for agent communication and state management (Gurazada, 2025). In this paradigm:

- A data generator agent streams synthetic or experimental measurements
- Predictive agents (e.g., viability collapse) analyze incoming data
- An orchestration layer manages agent communication and data flow
- Dashboard agents visualize results in real time

This architecture naturally extends to differentiation twins, with agents for imaging feature extraction, transcriptomic integration, and fate prediction coordinated through a central workflow.

Continuous Learning Loops. Industrial digital twins implement closed-loop learning where experimental validation continuously refines model parameters. The reinforcement learning framework described in Section 5 exemplifies this approach—the twin proposes optimized protocols, wet-lab experiments test predictions, and results update model weights. Cloud infrastructure supports this cycle through automated retraining pipelines and version-controlled model registries.

Challenges and Future Directions

Despite rapid progress, significant challenges remain. Data heterogeneity—omics in count matrices, images in multi-resolution pyramids, time-series in continuous recordings—complicates unified representation. Computational costs for 3D organoid simulations at single-cell resolution remain prohibitive for routine use. Standardization efforts, while advancing through communities like REMBI and BIDS, require broader adoption.

Future infrastructure must prioritize interoperability, enabling digital twins to exchange models and data across platforms. The interTwin initiative, integrating BioDynaMo with broader digital twin ecosystems, represents a promising step toward this vision (CERN openlab, 2025). As cloud costs decrease and computational efficiency improves, centrosome-aware digital twins will transition from research prototypes to routine tools for optimizing differentiation protocols and understanding organelle-mediated fate decisions.

Key Challenges and Limitations

The ambitious vision of constructing centrosome-aware digital twins of cell differentiation confronts fundamental challenges that span the computational, experimental, and epistemological domains. Acknowledging these limitations is essential for establishing realistic expectations and guiding future research priorities. This section critically examines the principal obstacles that must be overcome for digital twins to achieve their transformative potential.

The Curse of Dimensionality

The integration of data from five or more organizational levels—extrinsic signals, centrosomal hubs, intracellular logistics, nuclear regulation, cellular phenotype, and population dynamics—creates a parameter space of astronomical dimensionality. This "curse of dimensionality" manifests as perpetual data sparsity, where the volume of parameter space grows exponentially with the number of features, rendering experimental coverage impossibly sparse (Aghasafari et al., 2023).

For centrosome-integrated digital twins, the dimensionality challenge is particularly acute. Consider the state vector required to describe a single cell's centrosomal module: mother centriole age and post-translational status, daughter centriole maturation state, associated

CAFD identities and binding stoichiometries, cilium presence and length, positioning relative to polarity axes. Multiplied across thousands of genes (transcriptome), hundreds of proteins (proteome), and dozens of morphological features, the resulting space defies comprehensive experimental sampling.

Aghasafari and colleagues (2023) argue compellingly that statistical and data-centric machine learning approaches alone cannot overcome this challenge due to the inapplicability of the Central Limit Theorem and the limits imposed by the Causal Hierarchy Theorem. They advocate instead for complex multi-scale mechanism-based simulation models, constructed and operated to account for epistemic incompleteness while providing maximal expansiveness in concordance with the Principle of Maximal Entropy. Such models can generate synthetic multi-dimensional molecular time series data that minimize overfitting and lack of generalizability—the known shortcomings of neural network AI systems (Aghasafari et al., 2023).

Dimension reduction techniques—principal component analysis, autoencoders, or more recent manifold learning approaches—provide partial solutions by identifying the latent variables that capture most variance. However, these methods risk discarding biologically critical but low-variance signals, particularly for rare cell states or transient regulatory events that may determine differentiation outcomes. The challenge lies in distinguishing true biological signal from technical noise while preserving the richness necessary for mechanistic insight.

Computational Complexity

Simulating populations of 10^5 or more cells with detailed intracellular models—including stochastic centrosomal dynamics—pushes the boundaries of current high-performance computing. Each agent in an agent-based model carries its own internal state vector and executes decision rules at each time step; when these agents number in the hundreds of thousands and their internal models encompass Boolean networks, ordinary differential equations, or even partial differential equations, the computational burden becomes staggering.

The PhysiCell/PhysiBoSS framework exemplifies both the power and the demands of multi-scale simulation. At the lowest scale, BioFVM solves partial differential equations for microenvironmental diffusion; at the cell scale, mechanical equations model individual movement and interactions; and at the intracellular scale, Boolean networks simulated with the MaBoSS algorithm capture signaling dynamics (Letort et al., 2019). This flexibility enables exploration of heterogeneous cell population responses to treatment, mutation effects, and cell invasion modes (Letort et al., 2019). However, real-time simulation of such coupled systems for experimentally relevant time scales (days to weeks) remains an exascale computing challenge.

The EMERGE project at Argonne National Laboratory demonstrates what is possible with exascale resources. Using the ExaEpi agent-based model to simulate disease spread, researchers run large simulation ensembles on Department of Energy computing facilities to identify influential variables and train fast AI surrogate models (Argonne Leadership Computing Facility, 2025). These surrogates, once calibrated, enable rapid what-if analyses with quantified

uncertainties. For centrosome differentiation twins, similar strategies will be necessary: massive offline simulation ensembles to train surrogate models that can then support interactive exploration.

The computational demands are not merely about raw speed but also about algorithmic innovation. Multi-scale coupling across time scales—from milliseconds for molecular events to days for differentiation—requires sophisticated co-simulation strategies that preserve accuracy while enabling computational feasibility (ECMI, 2024).

The Closing-the-Loop Problem

The ultimate vision for digital twins involves autonomous operation: the twin not only predicts but also, through integration with robotic laboratory systems, designs and executes experiments to refine its own predictions. This "closed loop" capability—closing the quality loop from concept to shelf, as framed in industrial contexts (Siemens, 2025)—requires seamless integration with laboratory information management systems (LIMS) and automated experimentation platforms.

Recent advances in self-driving laboratory infrastructure demonstrate the technical feasibility of such integration. The Lab Connect framework developed by Molecular Quantum Solutions provides a comprehensive architecture for laboratory digitalization, spanning six levels of autonomy from basic automation to fully self-driving laboratories (Molecular Quantum Solutions, 2025). Key components include OPC-UA gateways for heterogeneous equipment unification, MQTT brokers with time-series PostgreSQL databases for data infrastructure, and Kubernetes deployment for container orchestration. The framework supports closed-loop experimentation through Bayesian optimization integration and digital twins with incremental learning capabilities (Molecular Quantum Solutions, 2025).

For centrosome digital twins, closing the loop would mean: (1) the twin proposes an optimized differentiation protocol (e.g., Alisertib 50 nM on day 4), (2) this protocol is automatically executed by robotic liquid handlers and automated incubators, (3) imaging and omics data are acquired and fed back to the twin, and (4) model parameters are updated through reinforcement learning or Bayesian inference. While individual components exist, their integrated deployment for complex biological systems remains aspirational.

The challenge extends beyond hardware integration to software architectures that support continuous learning. As demonstrated in bioprocess monitoring applications, multi-agent AI pipelines using LangGraph orchestration can combine real-time data streams with predictive models and interactive dashboards (Gurazada, 2025). Adapting such architectures to the heterogeneous data types and complex decision spaces of differentiation biology requires substantial engineering effort.

Incomplete Biological Knowledge

Digital twins require parameterization, but our biological knowledge remains fundamentally incomplete. For centrosome-mediated fate regulation, critical knowledge gaps include:

- **Complete CAFD inventories**: The full complement of cell fate determinants that associate with centrosomes remains unknown for most lineages. Proximity proteomics (APEX2, TurboID) is progressively filling these gaps, but comprehensive maps are years away.
- **Mechanistic rules**: The binding kinetics, phosphorylation dependencies, and stoichiometries of CAFD-centrosome interactions are characterized for only a handful of proteins (β-catenin, YAP, STAT3). For most putative interactions, even binary binding data are lacking.
- **Tissue-specific variation**: Centrosome composition and regulation vary across cell types and developmental stages (Meyer-Gerards et al., 2025). Models trained on one lineage may not generalize to others.
- **Causal structure**: Distinguishing correlation from causation in perturbation experiments remains challenging. Recent work on causal structural equation models for cell line perturbations reveals the complexity of inferring causal relationships from observational and interventional data (Yuan et al., 2025). The Linear Regression estimator may match or even outperform more sophisticated causal structure learning approaches in predictive accuracy, despite making weaker assumptions (Yuan et al., 2025).

Models must therefore operate with hypothetical variables—placeholders for unknown CAFDs, provisional rate constants, and assumed interaction topologies. This epistemic incompleteness (Aghasafari et al., 2023) necessitates robust uncertainty quantification and sensitivity analysis to identify which knowledge gaps most constrain predictive power.

The integration of biology-first priors with data-driven modeling offers a path forward. As Zhou (2026) argues, combining optimal transport, Schrödinger bridges, and flow matching with generative AI enables inference of continuous dynamics from static single-cell snapshots while maintaining mechanistic interpretability—a crucial advance over black-box approaches that sacrifice biological insight for predictive performance.

Validation of In Silico Predictions

The most valuable predictions from digital twins are often the most counterintuitive—the non-obvious intervention strategies that emerge from exploring high-dimensional parameter spaces. Yet these predictions are also the most difficult and expensive to validate experimentally.

Consider a hypothetical twin prediction: "For efficient osteoblast differentiation, first inhibit centrosomal kinase X for 24 hours, then abruptly activate it for 12 hours, timed precisely with BMP4 addition." Testing this prediction requires:

- Developing or obtaining selective kinase inhibitors and activators
- Optimizing dosing regimens without prior guidance

- Running multi-week differentiation experiments with appropriate controls
- Collecting multi-omic and imaging data at multiple time points
- Analyzing results to confirm or refute the prediction

The cost—in time, reagents, and personnel—can easily exceed $50,000 per prediction. With dozens of potentially interesting predictions emerging from a well-calibrated twin, comprehensive experimental validation becomes prohibitive.

Strategies to address this challenge include:

- Prioritization frameworks: Using uncertainty quantification to identify predictions with highest confidence and largest potential impact
- Iterative experimental design: Testing predictions sequentially, using results to refine the twin before committing to more expensive validations
- Surrogate validation systems: Developing high-throughput, lower-cost assays that approximate key aspects of the differentiation process (e.g., reporter cell lines for centrosomal kinase activity)
- Community challenges and benchmarks: Organizing multi-laboratory efforts to systematically validate predictions from competing models, as pioneered in cancer systems biology

The reinforcement learning framework described in Section 5 embodies this iterative approach: each experimental validation updates the model, progressively improving its predictive accuracy while focusing experimental effort on high-value test cases.

Synthesis and Path Forward

These challenges are formidable but not insurmountable. The curse of dimensionality can be tamed through mechanism-based simulation that generates synthetic training data covering regions of parameter space inaccessible to experiment (Aghasafari et al., 2023). Computational complexity will yield to exascale resources and AI-surrogate hybridization (Argonne Leadership Computing Facility, 2025). Closing the loop is achievable through modular, scalable laboratory automation architectures (Molecular Quantum Solutions, 2025). Incomplete knowledge can be accommodated through robust uncertainty quantification and causal inference methods (Yuan et al., 2025). Validation costs can be managed through iterative, adaptive experimental design.

The key insight is that these challenges are interconnected and must be addressed in a coordinated manner. Mechanism-based simulation reduces experimental dimensionality; surrogate models address computational complexity; automated laboratories enable iterative validation; uncertainty quantification guides knowledge gap prioritization. Digital twins are not merely computational artifacts but socio-technical systems that integrate models, data, experiments, and human expertise in a continuous learning cycle. Recognizing and

systematically addressing these limitations will determine whether centrosome-aware digital twins fulfill their promise or remain academic exercises.

Ethical and Regulatory Aspects

The development and deployment of centrosome-aware digital twins for cell differentiation transcend purely technical challenges, raising profound ethical, legal, and regulatory questions. As these models transition from research prototypes to tools informing clinical decisions—such as optimizing differentiation protocols for cell therapies—the ethical implications demand systematic attention. This section examines three critical dimensions: data sovereignty and patient privacy, liability for model predictions, and reproducibility across institutional boundaries.

Digital Sovereignty and Data Privacy

Centrosome-integrated digital twins require training on extensive multi-modal datasets, including biobanked patient-derived cell lines, genomic profiles, and high-resolution imaging data. These data are not anonymous in any trivial sense; they carry the potential for re-identification and contain deeply personal information about donors' health trajectories and genetic predispositions.

The Re-Identification Risk. Anonymized patient data reduces privacy threats and ensures compliance with regulatory frameworks such as the General Data Protection Regulation (GDPR) and the Health Insurance Portability and Accountability Act (HIPAA) (Ramesh Babu et al., 2025). However, the richness of data required for digital twins—combining genomics, transcriptomics, proteomics, and imaging—amplifies re-identification risks. The integration of multiple modalities creates unique signatures that statistical de-identification methods may not adequately protect. Recent frameworks propose combining blockchain technology with digital twins to enhance security, allowing patients to control data access through cryptographic keys while maintaining data consistency throughout its lifecycle (Ramesh Babu et al., 2025). Case studies demonstrate the practical impact: Siemens Healthineers' implementation of blockchain and digital twins reduced unauthorized access attempts by 85% (Ramesh Babu et al., 2025).

Privacy-Preserving Architectures. Emerging technical solutions address privacy concerns without compromising model performance. Privacy-Preserving AI (PPAI) frameworks enforce feature anonymization and encrypted data transfers without relying solely on federated learning (IEEE Xplore, 2025). A dual-cloud architecture decouples data processing from AI-driven analysis, eliminating raw data exposure while enabling real-time model adaptation. For centrosome digital twins, such architectures would allow sensitive patient-derived cell line data to remain within secure enclaves while model parameters are shared across institutions.

Informed Consent in the Digital Twin Era. The scope of data collection for digital twins is often unclear at the time of consent, making meaningful informed consent difficult to establish (Weidema et al., 2025). Donors providing cells for biobanking may not anticipate their materials being used to train predictive models that guide differentiation protocols for other patients. Moreover, patient-derived data may be used to create digital representations that outlive the

original donor, raising questions about posthumous data governance. Weidema and colleagues (2025) argue that consent models must evolve to address the uncertainty inherent in open-ended research applications, potentially through dynamic consent platforms that allow ongoing donor engagement.

Justice and Representation. A relational bioethical approach emphasizes that privacy protections must be distributed equitably (Ferlito et al., 2024). If stringent data governance disproportionately burdens certain populations or excludes underrepresented groups from digital twin development, the resulting models may perpetuate or exacerbate health disparities. Ferlito and colleagues (2024) advocate for acknowledging how uneven relational structures affect access to medical care and promoting social justice through careful evaluation of resource allocation.

Liability for Predictive Recommendations

Perhaps the most legally fraught question surrounding centrosome digital twins concerns accountability when model predictions lead to adverse outcomes. Consider a scenario: a digital twin recommends an optimized differentiation protocol involving Alisertib modulation at a specific time point. A clinician implements this protocol, and the resulting cell population, when transplanted into a patient, forms a teratoma. Who bears legal responsibility?

The Accountability Gap. The challenge lies in the distributed nature of digital twin development and deployment. Multiple actors contribute: the researchers who designed the model architecture, the engineers who trained it on experimental data, the biologists who validated its predictions, the clinicians who interpreted its recommendations, and the regulators who approved its use. Traditional liability frameworks, designed for discrete products or services, struggle to allocate responsibility across this ecosystem.

Liability in Safety-Critical Contexts. The digital risk twin (DRT) framework, developed for safety-critical engineering systems, offers conceptual tools relevant to biomedical applications. Kabashkin (2025) formalizes the transformation from descriptive digital twins to risk-aware systems by embedding probabilistic hazard modeling and coherent risk measures. In this framework, the DRT explicitly represents "how the system may fail" and "what the consequences of failure entail" (Kabashkin, 2025). Translating this to cellular differentiation, a centrosome digital twin would need to quantify not only optimal protocols but also failure probabilities—for instance, the risk of teratoma formation given protocol parameters and patient-specific factors.

Regulatory Ambiguity. Current regulatory frameworks are still adapting to digital twin technologies. The FDA's Modernization Act 2.0 has removed statutory requirements for animal testing in certain contexts, explicitly allowing New Approach Methods (NAMs), including organoids and organ-on-chips, in investigational drug applications (Weidema et al., 2025). However, digital twins themselves are not yet formally recognized as validated tools for regulatory decision-making. The FDA and NIH have released a roadmap to support validation,

standardization, and adoption of these systems (Weidema et al., 2025), but concrete guidance on liability remains absent.

Silva and colleagues (2025) note that integrating digital twins into clinical practice raises questions about liability, resource allocation, and incidental findings. When digital twins detect low-risk biomarkers or predict adverse outcomes, clinicians face decisions about disclosure and intervention—the "right not to know" complicates the ethical landscape. The authors emphasize that digital twins must be framed as cognitive tools for clinical reasoning rather than autonomous decision-makers, preserving human accountability (Silva et al., 2025).

Towards a Liability Framework. A defensible approach might involve:

- **Layered responsibility**: Differentiating between model developers (responsible for technical validity), clinical implementers (responsible for appropriate interpretation), and institutional oversight (responsible for governance)
- **Uncertainty quantification**: Mandatory reporting of prediction confidence intervals and failure mode probabilities
- **Audit trails**: Comprehensive logging of model versions, input data, and reasoning pathways to enable retrospective analysis
- **Informed consent for model use**: Explicit disclosure to patients that treatment protocols were optimized using digital twin predictions with quantified uncertainties

Reproducibility and Model Transfer

A digital twin trained on data from one laboratory, using specific equipment, cell lines, and protocols, may fail catastrophically when deployed elsewhere. This reproducibility crisis threatens the translational potential of centrosome digital twins and demands standardization at multiple levels.

Sources of Non-Reproducibility. Variation arises from:

- **Biological variability**: iPSC lines from different donors, passage numbers, culture conditions
- **Technical variability**: Microscope calibration, antibody lots, sequencing depths, software versions
- **Environmental factors**: Inculator humidity, CO_2 gradients, media batch effects
- **Model sensitivity**: Neural network weights may exploit dataset-specific artifacts rather than learning general biological principles

Chaparro-Cárdenas and colleagues (2025) identify clinical validation and scalability as key implementation challenges for healthcare digital twins. Without rigorous validation across

diverse settings, models risk producing inscrutable or misguided predictions that limit applicability across populations (Weidema et al., 2025).

Calibration Standards. The digital twin calibration literature, primarily from engineering domains, specifies quantitative tolerances for model accuracy. For industrial digital twins, geometric matching requires errors ≤0.05mm, dynamic response delays ≤10ms, and sensor data deviations ±0.1% (yjsshiwu.com, 2025). While biological systems cannot achieve such precision, these specifications illustrate the principle that calibration must be quantified against reference standards.

Emerging standards development addresses this gap. The International Electrotechnical Commission (IEC) is developing PNW JTC1-SC41-543 ED1: "Digital twin – Process and guidance for digital twin model construction" (BSI Group, 2025). This document specifies information preparation requirements, basic model types, development principles, and model verification and calibration procedures. Adoption of such standards across the digital twin community would facilitate model transfer and cross-validation.

Multi-Site Validation Strategies. Addressing reproducibility requires:

- **Multi-center training**: Incorporating data from multiple laboratories to capture technical and biological variability
- **Domain adaptation techniques**: Machine learning methods that adjust for dataset-specific biases
- **Benchmark datasets**: Publicly available reference data with known ground truth for model comparison
- **Continuous validation**: Ongoing performance monitoring when models are deployed in new settings

The European Organ-on-Chip Society (EUROoCS), together with standardization institutes, has published a roadmap setting priorities for harmonization and technology integration into regulatory frameworks (Weidema et al., 2025). Similar collaborative efforts are needed for digital twins in stem cell biology.

Synthesis and Recommendations

The ethical and regulatory challenges of centrosome digital twins are interconnected and demand coordinated responses. Table 2 summarizes key recommendations.

Table 2. Ethical and Regulatory Recommendations for Centrosome Digital Twins

Domain	Recommendation	Responsible Parties

Data Sovereignty	Implement privacy-preserving architectures (encrypted data processing, blockchain access control)	Model developers, IT security
Informed Consent	Develop dynamic consent platforms enabling ongoing donor engagement	Biobanks, ethics committees
Liability	Establish layered responsibility frameworks with clear audit trails	Regulators, professional societies
Uncertainty	Mandate confidence intervals and failure mode reporting in model outputs	Model developers
Reproducibility	Adopt emerging IEC standards for model construction and calibration	International standards bodies
Validation	Conduct multi-site training and continuous performance monitoring	Research consortia

The responsible development of centrosome digital twins requires moving beyond technical optimization to embrace these ethical dimensions as integral to the technology's success. As Ferlito and colleagues (2024) argue, a relational bioethical approach—one that attends to power dynamics, representation, and social justice—can foster equitable, inclusive, and responsible integration of digital twins into biomedical practice. The goal is not merely to build accurate models but to build trustworthy ones, worthy of the confidence placed in them by patients, clinicians, and society.

Outlook and Conclusion

The preceding sections have outlined both the immense promise and the formidable challenges of constructing centrosome-aware digital twins of cell differentiation. As we look toward the future, it is useful to chart a realistic trajectory for the field—one that acknowledges current limitations while capturing the transformative potential of these technologies. This concluding section presents a phased outlook for the coming decades and synthesizes the core argument of this review.

Near-Term Horizon (3-5 Years): Specialized Protocol Twins

In the immediate future, we anticipate the emergence of the first specialized digital twins tailored to specific differentiation protocols—for example, the generation of cardiomyocytes, pancreatic β-cells, or cortical neurons from induced pluripotent stem cells (iPSCs). These initial twins will be deliberately constrained in scope, integrating a limited but strategically chosen set of parameters: time-series single-cell transcriptomics, a handful of imaging-derived centrosomal markers (centriole position, cilium presence, mother-daughter asymmetry), and key culture conditions.

Application in Pharmaceutical Industry. The primary early adopters will likely be pharmaceutical companies seeking to optimize differentiation protocols for in vitro drug screening. The CARDIOVERSE project at The Jackson Laboratory, funded by an up to $30 million contract from ARPA-H, exemplifies this trajectory: by combining iPSC-derived cardiomyocytes with AI-trained digital twins of the human heart, researchers aim to predict cardiotoxicity across genetically diverse populations before human trials begin (The Jackson Laboratory, 2025). Such "virtual heart" platforms promise to dramatically reduce the need for large-animal studies, streamline FDA approvals, and identify rare adverse reactions that might otherwise emerge only after market approval (Turner, 2026).

In Silico Small Molecule Screening. Within this timeframe, digital twins will begin to be used for in silico screening of small molecules that enhance differentiation efficiency. By simulating how candidate compounds affect both gene regulatory networks and centrosome dynamics (e.g., Aurora A kinase inhibitors modulating ciliogenesis timing), these twins will enable researchers to prioritize compounds for experimental validation, significantly reducing the cost and time of traditional high-throughput screening campaigns.

Proof-of-Concept in Organoid Engineering. Parallel efforts will focus on digital twin-guided organoid development, as proposed in recent Horizon Europe initiatives. These projects aim to pair each biological organoid with a continuously updated computational replica that predicts developmental trajectories and enables closed-loop optimization of culture conditions (Konkuk University, 2026). By integrating transcriptomics, proteomics, imaging-based morphology, and biomechanical cues, these platforms will begin to address the high variability and limited reproducibility that currently plague organoid technologies.

Medium-Term Horizon (5-10 Years): Standardization and Commercialization

The second phase will witness the maturation of digital twin technologies from research prototypes to standardized, commercially available platforms.

Platform-Based Digital Twin Construction. We foresee the emergence of cloud-based "digital twin builder" platforms that allow biologists to upload their own multi-modal data and receive a customized model with pre-trained modules—including a dedicated "centrosomal module" that captures organelle-specific dynamics. Such platforms will democratize access to digital twin

technology, enabling laboratories without extensive computational expertise to leverage predictive modeling in their differentiation research.

The development of virtual cell platforms by companies such as Yanyin Technology and Shentuo Bio in China illustrates this trend: by combining AI large models with high-quality multi-omics data, these collaborations aim to create "digital twins" of cells that simulate dynamic responses to drug interventions, providing decision support before wet-lab experiments begin (衍因科技 & 神拓生物, 2026).

Standardization Efforts. Critical to this phase will be the establishment of international standards for digital twin model construction, verification, and calibration. The International Electrotechnical Commission (IEC) is already developing standards that specify information preparation requirements, basic model types, and validation procedures (BSI Group, 2025). As these standards gain adoption, they will facilitate model sharing, cross-laboratory validation, and regulatory acceptance.

Personalized Medicine Applications. The medium term will also see the first applications of digital twins in personalized medicine. For a patient requiring autologous cell therapy—for example, iPSC-derived retinal pigment epithelial cells for age-related macular degeneration—a digital twin could be calibrated using the patient's own cell line data to predict the optimal differentiation protocol, maximizing yield and functional quality while minimizing batch-to-batch variability. The NIH's recent development of a subcellular-resolution digital twin of RPE cells, trained on approximately 1.3 million cells, demonstrates the feasibility of this approach (Ortolan et al., 2026). By quantifying how polarity emerges during differentiation, such models provide a reference map for investigating disease mechanisms and optimizing therapeutic production.

Long-Term Horizon (10+ Years): The Virtual Cell and Beyond

The distant future holds the promise of truly comprehensive "virtual cells" as integrated components of broader Human Digital Twin initiatives.

The Virtual Cell as a Foundational Module. Within this vision, the digital twin of cell differentiation becomes one module within a hierarchical model of human physiology—nested within tissue twins, organ twins, and ultimately the whole-body digital twin. These virtual cells would be capable of predicting cellular behavior in any context, responding to genetic perturbations, environmental signals, and therapeutic interventions with fidelity approaching that of biological systems.

The 3D UNIV+RSES platform demonstrated by Dassault Systèmes at CES 2026 offers a glimpse of this future. By integrating AI, patient virtual twins, and real-time sensor data, the platform creates an immersive "healthcare operating system" capable of simulating neurodegenerative disease progression and predicting health status changes before symptoms appear (Dassault Systèmes, 2026). While currently focused on organ-level modeling, such platforms lay the groundwork for multi-scale integration extending down to the subcellular level.

Lyu and colleagues (2026) provide a comprehensive survey of organ-level digital twins, noting that translation from engineering to biomedicine requires overcoming profound challenges, including anatomical variability, multi-scale biological processes, and the integration of multi-physics phenomena. The centrosome, as a nexus where signaling, transport, mechanics, and fate determination converge, offers an ideal entry point for addressing these challenges at the cellular scale.

Toward Animal-Free Preclinical Testing. Perhaps the most profound implication of fully realized virtual cells is the potential to transform preclinical drug testing. As noted by Lawrence Florin, CEO of Hesperos, digital twins require continuous (near real-time) updates, distinguishing them from conventional static simulation models (Florin, 2026). When combined with human-on-a-chip systems, these dynamic digital twins can generate safety, efficacy, and pharmacokinetic data that meet regulatory standards and support IND filings.

The ultimate goal, articulated by Matt Mahoney of The Jackson Laboratory, is "a future where computational models predict drug safety so reliably that it becomes ethical to move forward on computational evidence alone" (Turner, 2026). Achieving this vision would revolutionize drug development, making it far more affordable and accessible while dramatically reducing reliance on animal testing. The integration of centrosome-specific dynamics into these models is essential, given the organelle's central role in signaling, polarity, and asymmetric cell division.

Synthesis and Final Reflections

This review has advanced a central thesis: the creation of digital twins of cell differentiation with explicit integration of centrosomal dynamics is not a speculative fantasy but a logical necessity for transitioning from empirical trial-and-error to rational engineering of cell fate. The centrosome, long viewed as a passive microtubule-organizing center, has emerged as an active signal integration platform whose state—mother versus daughter age, associated CAFDs, cilium presence—directly biases daughter cell fate decisions (Doxsey, 2016; Reina & Gonzalez, 2014).

The Interdisciplinary Challenge. Realizing this vision demands unprecedented collaboration across traditionally siloed disciplines:

- Cell biologists must continue to map the molecular composition and dynamic regulation of the centrosomal microenvironment, using proximity proteomics, super-resolution imaging, and FRET-based biosensors (Lach et al., 2022; Guo et al., 2025).
- Bioinformaticians must develop robust pipelines for integrating multi-modal data across scales, from nanometers (protein-protein interactions) to millimeters (tissue organization) (Ebrahimi et al., 2025).
- Machine learning specialists must design architectures capable of learning from sparse, heterogeneous data while preserving mechanistic interpretability—whether through hybrid LSTM-CNN models, graph neural networks, or physics-informed AI (Sadria & Bury, 2024; Han et al., 2025).

- Engineers must create the laboratory automation infrastructure that enables closed-loop learning, where digital twins not only predict but also design and execute experiments to refine their own parameters (Molecular Quantum Solutions, 2025; Gurazada, 2025).

The Stakes. The potential rewards justify this collective effort. Successful centrosome-aware digital twins will:

- Accelerate therapeutic development by enabling in silico screening of differentiation-enhancing compounds and optimizing protocols for cell therapy manufacturing
- Enhance safety by predicting off-target effects and rare adverse outcomes before clinical translation
- Democratize access by making predictive modeling accessible to laboratories worldwide through standardized platforms
- Deepen biological understanding by revealing causal relationships hidden in complex multi-modal data

The recent convergence of enabling technologies—single-cell multi-omics, super-resolution live imaging, proximity proteomics, cloud-scale computing, and advanced AI architectures—suggests that the time is ripe for this endeavor. The roadmap outlined in this review, while ambitious, is grounded in demonstrable progress across each constituent field.

Final Thought. In conclusion, the construction of centrosome-integrated digital twins of cell differentiation represents a grand challenge at the intersection of cell biology, data science, and bioengineering. Its successful pursuit will not only yield powerful tools for regenerative medicine but also fundamentally transform how we understand and control the logic of cell fate. The path forward is demanding, but the destination—predictable, safe, and accessible cell therapies guided by faithful virtual replicas—is worthy of the journey.

References

Abdelmohsen, K., Mazan-Mamczarz, K., Tsitsipatis, D., Rossi, M., Munk, R. B., & Gorospe, M. (2025). Dissecting the heterogeneity of senescent cells through CITE-seq integration. In The Surfaceome: Methods and Protocols (Methods in Molecular Biology, Vol. 2908, pp. 99-109). Springer US.

Aghasafari, P., et al. (2023). Generating synthetic multi-dimensional molecular-mediator time series data for artificial intelligence-based disease trajectory forecasting and drug development digital twins: Considerations. arXiv. arXiv:2303.09056

Argonne Leadership Computing Facility. (2025). EMERGE: ExaEpi calibration runs and surrogate models. Argonne National Laboratory. https://www.alcf.anl.gov

Benzinga. (2025, March 26). Morpheus Unveils Decentralized AI Forge For Smart Agents, Simplifying Development. Benzinga. https://www.benzinga.com

Bhardwaj, S., Kumar, G., Yang, H., Bhardwaj, S., Wang, Q., Shen, M., Wang, Y., & De Souza, C. M. (2025). Virtual cells: From conceptual frameworks to biomedical applications. arXiv. arXiv:2509.18220

Bian, S., Wang, Y., Zhou, Y., & Chen, Y. G. (2025). Integrated protocol for single-cell RNA sequencing, organoid culture, and RNA-protein spatial analysis in gastroesophageal tissues. BMC Methods, *2*, 15.

BIDS Specification. (2025). Microscopy modality-specific files. Brain Imaging Data Structure. https://bids-specification.readthedocs.io

Biocomplexity Institute. (2025). COPASI: Biochemical network simulation and analysis. University of Virginia. http://www.copasi.org

BSI Group. (2025). *PNW JTC1-SC41-543 ED1: Digital twin – Process and guidance for digital twin model construction*. BSI Standards Development. https://standardsdevelopment.bsigroup.com

Camacho Gómez, D., Gonçalves, I., Gomez-Benito, M. J., Hervas-Raluy, S., Perez, M. A., Guerrero, P., Borau, C., & García Aznar, J. M. (2023). Challenges in modeling the morphogenesis of human-cell-based organoids. Conference presentation, ICCB 2023.

CERN openlab. (2025). BioDynaMo: Biology Dynamics Modeller. CERN. https://openlab.cern/biodynamo

Chaparro-Cárdenas, S. L., Ramirez-Bautista, J. A., Terven, J., Córdova-Esparza, D. M., Romero-Gonzalez, J. A., Ramírez-Pedraza, A., & Chavez-Urbiola, E. A. (2025). A technological review of digital twins and artificial intelligence for personalized and predictive healthcare. Healthcare, *13*(14), 1763.

Dassault Systèmes. (2026, January 8). 達梭系統在 CES 2026 展示運用 AI 技術 重塑精準、預測與個人化醫療的未來 [Dassault Systèmes showcases AI technology at CES 2026 to reshape the future of precision, predictive, and personalized medicine]. TechOrange. https://buzzorange.com/techorange

Diez, A., Tsutsumi, R., & Plunder, S. (2025). Emergence of digit-like structures through mechanical instabilities: A hybrid computational and experimental study. *ACMB-JSMB2025*.

Doxsey, S. (2016). The centrosome, a multitalented renaissance organelle. Cold Spring Harbor Perspectives in Biology, *8*(12), a025049.

Du, L., et al. (2025). Spatial omics in 3D culture model systems: decoding cellular positioning mechanisms and microenvironmental dynamics. Journal of Translational Medicine, *23*(1), 1356.

Ebrahimi, A., Fotuhi Siahpirani, A., & Montazeri, H. (2025). sCIN: A contrastive learning framework for single-cell multi-omics data integration. bioRxiv. https://doi.org/10.1101/2025.02.03.636095

ECMI. (2024, March 21). Kaiserslautern. Mathematics for a challenging biomedical problem: Report on the project. European Consortium for Mathematics in Industry.

ELIXIR Norway. (2025). Cheat sheet: Light microscopy. Research Data Management resources. https://elixir.no

Eschment, M., Mercey, O., Aarts, E. M., Perego, L., Figueiro da Silva, J., Mennel, M., Abidi, A., Generali, M., Rauch, A., Guichard, P., Hamel, V., & Bachmann-Gagescu, R. (2025). CEP290-deficiency disrupts ciliary axonemal architecture in human iPSC-derived brain organoids. Journal of Cell Science, jcs.264092. Advance online publication.

Ferlito, B., et al. (2024). Navigating the landscape of digital twins in medicine: A relational bioethical inquiry. Asian Bioethics Review, *16*(3), 471-481.

Fernandes-Mariano, C., et al. (2025). Centrosome biogenesis and maintenance in homeostasis and disease. Current Opinion in Cell Biology, *94*, 102485.

Florin, L. (2026, February 13). Hesperos' Human-on-a-Chip® and digital twin platforms redefine disease modeling. MedTech Spectrum. https://medtechspectrum.com

Freederia Research. (2025). Automated cellular microfactory design optimization via reinforcement learning & digital twin simulation. DEV Community. https://dev.to/freederia-research/automated-cellular-microfactory-design-optimization-via-reinforcement-learning-digital-twin-57n5

Guo, H., et al. (2025). An enzymatic cascade enables sensitive and specific proximity labeling proteomics in challenging biological systems. Nature Communications, *16*, 9691.

Guo, H., et al. (2025a). An enzymatic cascade enables sensitive and specific proximity labeling proteomics in challenging biological systems. Nature Communications, *16*, 9691.

Guo, H., Zhang, C., & Qin, W. (2025b). Spatiotemporally resolved profiling of protein movement by TransitID. Methods in Molecular Biology, *2872*, 215-234.

Gurazada, G. (2025). multi-agent-AI: Developing a multi-agent AI system to predict cell viability collapse in biopharma manufacturing [Source code]. GitHub. https://github.com

Gurazada, G. (2025). multi-agent-AI: Developing a multi-agent AI system to predict cell viability collapse in biopharma manufacturing. GitHub repository.

Han, M., Vitelli, V., et al. (2025). Learning noisy tissue dynamics across time scales. arXiv. arXiv:2510.19090

He, Q., Wang, G., Wakade, S., Dasgupta, S., Dinkins, M., Kong, J. N., Spassieva, S. D., & Bieberich, E. (2014). Primary cilia in stem cells and neural progenitors are regulated by neutral sphingomyelinase 2 and ceramide. Molecular Biology of the Cell, *25*(11), 1715-1729.

Huang, K., & Nandakumar, D. (2024, October 23). Augmenting Security Operations Centers with Accelerated Alert Triage and LLM Agents Using NVIDIA Morpheus. NVIDIA Technical Blog. https://developer.nvidia.com/blog

Huang, Y., et al. (2025). CCHCR1 links P-body proteins to the centrosome and is required for ciliogenesis through interacting with OFD1 and PCM1. Cellular & Molecular Biology Letters, *30*, 102.

Iber, D., Tanaka, S., Fried, P., Germann, P., & Menshykau, D. (2013). Simulating tissue morphogenesis and signaling. Methods in Molecular Biology, *1189*, 323-338.

IEEE Xplore. (2025). Self-healing digital twins: Hybrid generative and privacy-preserving AI for adaptive wellness platforms. IEEE Xplore.

Jaba, T. (2022). Dasatinib and quercetin: short-term simultaneous administration yields senolytic effect in humans. Issues and Developments in Medicine and Medical Research Vol. 2, 22-31.

Johnson, J. A. I., et al. (2025). Human interpretable grammar encodes multicellular systems biology models to democratize virtual cell laboratories. Cell. DOI: 10.1016/j.cell.2025.06.048

Kabashkin, I. (2025). Mathematical framework for digital risk twins in safety-critical systems. Mathematics, *13*(19), 3222.

Kemkar, S., Tao, M., Ghosh, A., Stamatakos, G., Graf, N., Poorey, K., Balakrishnan, U., Trask, N., & Radhakrishnan, R. (2024). Towards verifiable cancer digital twins: tissue level modeling protocol for precision medicine. Frontiers in Physiology, *15*.

Kiyomitsu, T. (Ed.). (2025). The mitotic spindle: Methods and protocols. Methods in Molecular Biology, 2872. Humana Press.

Klein, D., Pahl, A., Krentzel, D., et al. (2025). Joint profiling of cell morphology and gene expression during in vitro neurodevelopment. eLife, *14*, e102578.

Konkuk University. (2026). Digital twin–guided development of stem cell organoids for disease modeling and tissue repair. Horizon Europe Project Proposal.

Lach, R. S., Qiu, C., Zeyaei Kajbaf, E., Baxter, N., Han, D., Wang, A., ... & Wilson, M. Z. (2022). Nucleation of the destruction complex on the centrosome accelerates degradation of β-catenin and regulates Wnt signal transmission. Proceedings of the National Academy of Sciences, *119*(36), e2204688119.

Lach, R. S., Qiu, C., Zeyaei Kajbaf, E., et al. (2022). Nucleation of the destruction complex on the centrosome accelerates degradation of β-catenin and regulates Wnt signal transmission. Proceedings of the National Academy of Sciences, *119*(36), e2204688119.

Lee, C. Y., et al. (2020). centrocortin RNA localization to centrosomes is regulated by FMRP and facilitates error-free mitosis. Journal of Cell Biology, *219*(3), e201906101.

Lee, D. F., Su, J., Ang, Y. S., Carvajal-Vergara, X., Mulero-Navarro, S., Pereira, C. F., Gingold, J., Wang, H. L., Zhao, R., Sevilla, A., Darr, H., Williamson, A. J., Chang, B., Niu, X., Aguilo, F., Flores, E. R., Sher, Y. P., Hung, M. C., Whetton, A. D., ... & Lemischka, I. R. (2012). Regulation of embryonic and induced pluripotency by aurora kinase-p53 signaling. Cell Stem Cell, *11*(2), 179-194.

Letort, G., Montagud, A., Stoll, G., Heiland, R., Barillot, E., Macklin, P., Zinovyev, A., & Calzone, L. (2019). PhysiBoSS: a multi-scale agent based modelling framework integrating physical dimension and cell signalling. Bioinformatics, *35*(7), 1188-1196.

Li, X. (2025, July). Digital twins power early diagnosis & targeted therapy. In ECTS–IFMRS Joint Session, Vienna. International Federation of Musculoskeletal Research Societies.

Li, X., Jensen, S. L., Christiansen, N. B., et al. (2026). A cloud-enabled digital twin platform for upstream processing in biotechnology: Integrating CFD, compartmentalization, and reaction kinetics. Computers and Chemical Engineering, *206*, 109488. https://doi.org/10.1016/j.compchemeng.2025.109488

Litinetskaya, A., Schulman, M., Curion, F., Szalata, A., Omidi, A., Lotfollahi, M., & Theis, F. (2025). Integration and querying of multimodal single-cell data with PoE-VAE. bioRxiv. https://doi.org/10.1101/2022.03.16.484643

Lyu, Y., et al. (2026). Building digital twins of different human organs for personalized healthcare. arXiv. arXiv:2601.11318

Meyer-Gerards, C., et al. (2025). Developmental and tissue-specific roles of mammalian centrosomes. The FEBS Journal, *292*(4), 709-726.

Molecular Quantum Solutions. (2025). Lab Connect: Laboratory digitalization and self-driving laboratories documentation. https://docs.mqs.dk

Ortolan, D., Sathe, P., Volkov, A., et al. (2026). AI driven 3D subcellular RPE map discovers cell state transitions in establishment of apical-basal polarity. Nature Partner Journal–AI. Advance online publication. https://doi.org/10.1038/s44387-026-00074-6

Qi, Y., Feng, Y., Liu, X., Li, J., Huang, Q., Li, Z., ... & Zheng, B. (2025). Cellular indexing of transcriptomes and epitopes by sequencing unveils coordinated RNA and protein landscape in ovary at single-cell resolution. Chemical Engineering Journal, Advance online publication.

Ramesh Babu, R., Krishnan, R., & Sunanda, R. (2025). Ensuring privacy and security: A framework for sharing anonymized patient data using blockchain and digital twins. In A Framework for Sharing Anonymized Patient Data. Taylor & Francis.

Ramírez Sierra, M. A. (2024, September). From multiscale models to digital twins. Paper presented at the FIAS Workshop, Frankfurt Institute for Advanced Studies.

Reina, J., & Gonzalez, C. (2014). When fate follows age: unequal centrosomes in asymmetric cell division. Philosophical Transactions of the Royal Society B: Biological Sciences, *369*(1650), 20130466.

Sadria, M., & Bury, T. M. (2024). FateNet: an integration of dynamical systems and deep learning for cell fate prediction. Bioinformatics, *40*(9), btae525.

Siemens. (2025). Improve food & beverage manufacturing quality with an integrated LIMS software system. Siemens Digital Industries Software.

Silva, A., et al. (2025). Digital twins in personalized medicine: Bridging innovation and clinical reality. Journal of Personalized Medicine, *15*(11), 503.

Tanaka, H. (2021). Viscoelastic phase separation in biological cells. arXiv. arXiv:2104.12541

Technology Networks. (2025, July 25). Scientists create digital twin to simulate cell behavior. Technology Networks.

The Jackson Laboratory. (2025, December 8). The Jackson Laboratory receives an up to $30 million contract award to revolutionize drug safety testing with AI-powered "virtual hearts." BioInformant. https://bioinformant.com

Tkemaladze, J. (2023). Reduction, proliferation, and differentiation defects of stem cells over time: a consequence of selective accumulation of old centrioles in the stem cells?. Molecular Biology Reports, 50(3), 2751-2761. DOI : https://pubmed.ncbi.nlm.nih.gov/36583780/

Tkemaladze, J. (2024). Editorial: Molecular mechanism of ageing and therapeutic advances through targeting glycative and oxidative stress. Front Pharmacol. 2024 Mar 6;14:1324446. DOI : 10.3389/fphar.2023.1324446. PMID: 38510429; PMCID: PMC10953819.

Tkemaladze, J. (2026). Old Centrioles Make Old Bodies. Annals of Rejuvenation Science, 1(1). DOI : https://doi.org/10.65649/yx9sn772

Tkemaladze, J. (2026). Visions of the Future. Longevity Horizon, 2(1). DOI : https://doi.org/10.65649/8be27s21

Turner, D. (2026, February 2). 2026年临床前研究与临床开发将发生哪些变革 [What will change in preclinical and clinical research in 2026]. Drug Discovery World. https://drugdiscoveryworld.com

Virtual Cell (VCell). (2025). Introduction to the Virtual Cell Modeling and Simulation Framework. National Resource for Cell Analysis and Modeling. https://vcell.org

Wang, G., Kam, Y., & Romero, N. (2025). Optimized workflow to assess metabolic profile during adipocyte differentiation. Diabetes, *74*(Supplement 1), 1573-P.

Wang, P.-K., Lin, K.-H., & Tang, T. K. (2025). Centrosome migration and apical membrane formation in polarized epithelial cells: Insights from the MDCK cyst model. Molecular Biology of the Cell. Advance online publication.

Wang, X., et al. (2025). Multi-omics and deep learning for programmed cell death pathways in gastric cancer. Frontiers in Immunology, *16*, 1690200.

Weidema, et al. (2025). The ethical aspects of human organ-on-chip models: A mapping review. Stem Cell Reports, *20*(11), 102686.

Woodruff, J. B., Ferreira Gomes, B., Widlund, P. O., Mahamid, J., Honigmann, A., & Hyman, A. A. (2017). The centrosome is a selective condensate that nucleates microtubules by concentrating tubulin. Cell, *169*(6), 1066-1077.

Yamashita, Y. M., Mahowald, A. P., Perlin, J. R., & Fuller, M. T. (2007). Asymmetric inheritance of mother versus daughter centrosome in stem cell division. Science, *315*(5811), 518-521.

YellowDog. (2025, October 7). Developing virtual twin hearts with Hybrid Cloud - ELEM BioTech [Case study]. YellowDog. https://yellowdog.ai

yjsshiwu.com. (2025). 数字孪生体校准 [Digital twin calibration]. yjsshiwu.com.

Yuan, B., et al. (2025). Causal models and prediction in cell line perturbation experiments. BMC Bioinformatics, *26*(1), 4.

Zhang, J., Li, X., Guo, X., You, Z., Böttcher, L., Mogilner, A., Hoffmann, A., Chou, T., & Xia, M. (2025). Reconstructing noisy gene regulation dynamics using extrinsic-noise-driven neural stochastic differential equations. arXiv. arXiv:2503.09007

Zhou, P. (2026). AI动态虚拟细胞构建的理论与算法 [Theories and algorithms for AI-driven dynamic virtual cell construction]. Invited talk, Wuhan University.

衍因科技 & 神拓生物. (2026, January 15). 衍因科技与神拓生物达成战略合作，共建虚拟细胞平台 [Yanyin Technology and Shentuo Bio reach strategic cooperation to jointly build virtual cell platform]. 衍因科技官网. https://yanyin.tech

