

# Bridging Genomics and Manufacturing: Middleware Solutions for Personalized Medicine Production

Lakshmi Vara Prasad Adusumilli  
Indsoft Inc., USA

---

Received: 11.02.2026

Accepted: 15.02.2026

---

## Abstract

Pharmaceutical companies, faced with stiff competition from the rapid progression of genomic medicine and next-generation sequencing (NGS) technology adoption, are beginning to move from their customary mass manufacturing model to a patient-centric therapeutic model. Customary centralized pharmaceutical manufacturing and supply chains are not suited to manufacture the customized formulations that patient-centric therapies require within clinically relevant time frames. A distributed middleware architecture is used to support the real-time data flows between the genomic sequencing centers and the geographically dispersed pharmaceutical manufacturing sites. Production digital twins are used to manage the production facilities. Digital twins of production facilities know the capabilities, equipment state, regulatory clearances, and queue for processing medical instruments in real time. Smart facility selection algorithms account for the network capacity of the routing path for patient-specific requests. Privacy-preserving computational models such as secure multi-party computation and homomorphic encryption can be used for genomic sequencing scheduling optimization. Event-driven interlinking of manufacturing metadata databases enables constraint-aware distributed scheduling for multi-objective optimization of patient wait time, facility utilization, regulatory compliance, and resource allocation. Discrete-event simulation modeling and piloted implementations show the feasibility of orchestrating secondary therapies across regional manufacturing networks, while offering cryptographic guarantees of privacy and regulatory auditability.

**Keywords:** Personalized Medicine Manufacturing, Distributed Middleware Orchestration, Genomic Data Integration, Privacy-Preserving Computation, Digital Twin Architecture

## 1. Introduction

The pharmaceutical industry is experiencing a revolutionary transformation due to personalized medicine, and the global personalized medicine market is currently growing rapidly. The fastest-growing segment of personalized medicine is genomic-based therapies. Adoption has been accelerated by the advent of next-generation sequencing (NGS) technologies, which have driven the cost of whole genome sequencing down from US\$2.7 billion to perform the Human Genome Project in 2003 to a range of \$80-200 in 2024. The cost of sequencing has fallen faster than Moore's Law since 2008, enabling the evaluation of how individual genetic variants affect drug response and delivery. Most drugs prescribed by clinicians have pharmacogenomic interactions, and precision oncology using whole genome sequencing can be offered at the cost equivalent to ordering both a cancer panel test and genetic screening specific for a fusion gene [1]. New markets will emerge for patient-specific therapeutics for time-sensitive applications (e.g., CART cell therapy) and the personalization of cancer therapies over customary therapeutic molecules [2].

10.48047/jocaaa.2026.35.02.29

Pharmaceutical supply chains are currently embedded in a centralized batch production model that is optimized towards distribution to the population, and the manufacturing of drugs uses the standard processes with long lead times and inventory systems optimized for fixed-mass standard pharmaceutical formulations. This will require building systems to route real-time genomic sequencing data from production facilities able to sequence 7 human genomes at 30x depth per hour to distributed manufacturing sites, whilst complying with regulatory requirements, protecting patient privacy, and delivering clinical treatments in days or hours rather than months. [1] The solution set will also include middleware architectures capable of interoperability between clinical genomic data and pharmaceutical manufacturing systems operating on distributed networks, as well as the interfaces between highthroughput bioinformatics computational pipelines and industrial computing execution systems operating on distributed computing models.



Figure 1: Genome Sequencing Cost Reduction (2003-2024) [1]

## 2. The Distributed Manufacturing Paradigm

### 2.1 Production Digital Twins

The variability of patient response to drugs and the need for an individualized approach to medicine lead to an inclination towards distributed personalized medicine manufacturing, as the so-called blockbuster model of standardized pharmaceuticals has many shortcomings in virtually every therapeutic area. About 38% of antidepressant, 40% of asthma, 43% of diabetes, 50% of arthritis, 70% of Alzheimer's, and up to 75% of cancer patients fail to see any benefit from these drugs due to patient-to-patient variability in drug metabolism, receptor binding, and pharmacokinetics [2]. This fundamentally alters the drug manufacturing model from a deterministic mass-production problem designed for economies of scale optimizing the production of standard formulations to a patient-specific manufacturing problem that requires real-time

10.48047/jocaaa.2026.35.02.29

dynamic manufacturing orchestration of the manufacturing of each formulation variant across the distributed network of manufacturing facilities to achieve the desired outcomes based on the genomic profile of the patient. This imposes unprecedented complexity on manufacturing networks to cleverly route and place facilities based on specialized capabilities, patient urgency, and maintaining production efficiency across the network. The middleware framework enables production digital twins to maintain real-time awareness of facility-specific formulation capabilities matched to patient genomic requirements to enable smart routing of production orders to facilities that are capable of making the specific therapeutic variant required by each patient's genetic profile.

### 2.2 Orchestration Workflow

Connected to the architecture are production digital twins, virtual representations of each manufacturing facility in the network. Production digital twins have been shown to minimize unplanned downtime and maximize overall equipment effectiveness in customary manufacturing facilities using industrial internet of things systems [3]. Each digital twin contains real-time information on the ability of a facility to make products, the status of its equipment including predictive maintenance, product certifications in various jurisdictions, the status of raw materials including reorder points, and a queue of production batches with expected completion dates. The twins are implemented using distributed data structures and replicated onto the devices of edge computing nodes and the cloud infrastructure. The twins are synchronized using an event-driven middleware architecture to provide an up-to-date view of the manufacturing capacity across all locations and facilities on the network. The digital twin framework is further augmented with machine learning models that predict availability windows for the facility, batch cycle times, and quality control violations based on historical production data [4].

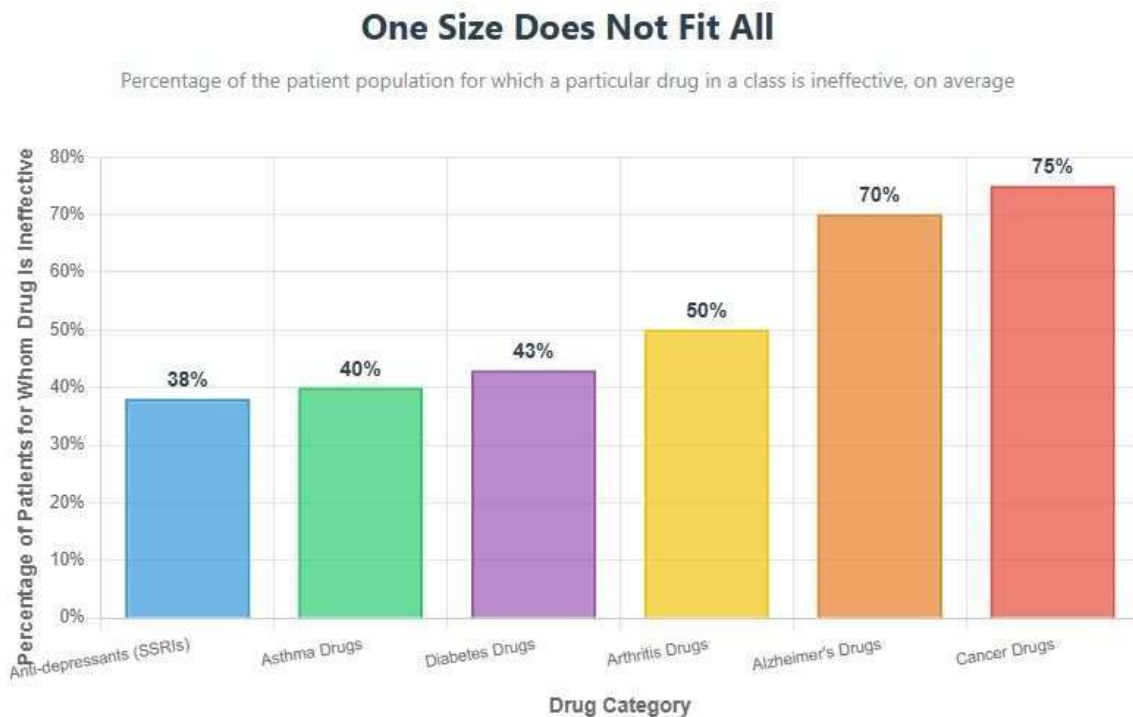


Figure 2: Percentage of the patient population for which a particular drug in a class is ineffective. [2]

### 3. Technical Architecture Components

#### 3.1 Polyglot Persistence Layer

Multiple data storage models can be employed in the architecture to meet different requirements; a polyglot persistence approach to distributed clinical computing is rightfully used in the field of biomedical informatics. Genomic sequences are stored using bioinformatics databases such as NoSQL document database systems that are optimized to meet the computational needs of very large genomic sequences. Genomic data storage needs efficient columnar data compression, while patient clinical data is stored using healthcare interoperability standards such as HL7 FHIR (Fast Healthcare Interoperability Resources) implementations in graph databases modeling complex relationships between clinical observations, medications, allergies, and genomic variants. Data structures support the high-throughput patient context queries required for formulation decision support. Manufacturing specifications are stored in industrial time-series databases that log real-time machine telemetry for temperature monitoring, pressure monitoring, and particle counts in clean room environments, with data retention policies that are compliant with regulations [6]. Regulatory documents for compliance and audit trail are cryptographically signed and timestamped on an immutable blockchain ledger. Each batch can then be tracked and linked to a complete chain of custody and quality control information, including raw materials' certificates of analysis, timestamps of equipment calibration, data from environmental sensors, and results of quality control tests, all stored as tamper-proof transaction records on the blockchain. The status data is stored in distributed in-memory data grids and run on edge computing nodes co-located with manufacturing facilities to allow for a low-latency response to scheduling requests while retaining high availability through the use of replicated services across multiple regions [5].

Data Category	Storage Model	Rationale
Genomic Sequences	NoSQL Document Databases	Optimized for large-scale columnar compression of sequencing data
Patient Clinical Records	Graph Databases (HL7 FHIR)	Models complex relationships among observations, medications, allergies, and variants
Manufacturing Telemetry	Industrial Time-Series Databases	Captures real-time machine sensor data with regulatorycompliant retention
Regulatory Audit Trails	Blockchain Ledger	Provides cryptographically signed, tamper-proof batch traceability
Facility Status Data	Distributed In-Memory Data Grids	Enables low-latency scheduling responses at edge computing nodes

Table 1: Data Storage Models in the Polyglot Persistence Architecture [5] [6]

#### 3.2 Event-Driven Communication

This is accomplished by using a cloud-native message streaming platform to implement publish-subscribe semantics via a distributed streaming platform that provides high volumes of messaging, enabling the distribution of events throughout the manufacturing network (e.g., batch start, quality control, batch end, facility status updates (e.g., current capacity utilization, equipment availability), and genomic analysis finish (to trigger orchestration workflows)) [6]. Smart event routing is enabled via topic-based partitioning, ensuring patient-protected health information only flows to systems authorized to receive sensitive

10.48047/jocaaa.2026.35.02.29

information via encrypted channels with end-to-end TLS encryption and certificate-based authentication. Production metadata (anonymized production metrics, readouts from equipment logs and sensors, and quality control) stream to cloud-based scheduling and optimization algorithms via separate topics. These topics have read-only access, with messages stored as timed retention policies for operational and regulatory audit events. Log compaction reduces storage usage while maintaining a full audit history [6]. The service also isolates systems for privacy while coordinating globally. Messages are delivered once and only once to critical production orders and at least once to monitoring telemetry. Data is not lost when there is network partitioning or a temporary service failure due to an automatic message resending mechanism with exponential backoff [5].

### 3.3 Constraint-Aware Distributed Scheduling

This layer is treated as a dynamic multi-objective optimization problem solved with MILP models in distributed computing clusters. The models are updated in real-time as new production requests or facility state changes within the distributed network. The conflicting objectives are minimizing wait times for patients with target time-to-treatment windows required by classification of therapeutic urgency, maximizing the rate of use of the facility in a balanced way with surge capacity against demand variations, complying with regulatory requirements through the restriction that production can only be performed in facilities with credentials compliant with regulatory requirements validated in real-time against regulatory databases, and minimizing waste of raw materials by batching production of compatible formulations with common excipients or active pharmaceutical ingredients. Hard constraints like maximum delays on critical therapies, geographical constraints imposed by regulations on crossborder shipments except for approved cold-chain supply chain corridors with verified temperaturecontrolled packaging, and specialized formulation and filling workflows like lyophilization and aseptic filling to matching facilities with certified clean rooms and qualified equipment are formulated as CSPs. Distributed consensus (e.g., Raft, Paxos, etc.) is used to prevent over-reservation of scarce shared resources like raw material inventories, limit allocation to facilities to avoid greedy, local decisions, and optimize globally across the network. These constraints are weighted by priority scores based on clinical needs, distance of patients from a priority patient, and facility throughput.

## 4. Privacy-Preserving Computation

Hence, the middleware must comply with HIPAA de-identification rules, Article 9 special category data of the GDPR, and HITRUST CSF certification requirements for healthcare handling [7]. The middleware utilizes SMC protocols based on Yao's garbled circuits and secret sharing schemes. When facilities participate in scheduling, they only receive production parameters without patients' identifying information. Homomorphic encryption methods for lattice public key algorithms, such as BGV (Brakerski-Gentry-Vaikuntanathan) and CKKS (Cheon-Kim-Kim-Song), could execute calculations that would perform dosage calculations, drug-drug interaction tests, or quality control parameters on genomic data without the need to decrypt it. Homomorphic encryption methods support the addition and multiplication of ciphertext with security at the same level as the AES encryption algorithm [8]. Genomic data is broken into two components. Identity data corresponding to the genomic data, including name, date of birth, medical record numbers, and geographical data, is always encrypted with standard AES-256 encryption during transmission and storage, with keys and devices managed by FIPS 140-2 Level 3 compliant hardware security modules (HSMs). The genetic variant data can be homomorphically encrypted, and the associated manufacturing systems and pipelines use encrypted computation protocols for drug formulation parameters such as the CYP2D6 metabolizer phenotype for dosage, TPMT enzyme activity for drug metabolism, and HLA-B\*5701 gene presence for hypersensitivity to abacavir [7].

10.48047/jocaaa.2026.35.02.29

Genomic data is encrypted and sent over the internet from sequencing laboratories to manufacturing execution systems. The transformation uses secret sharing among multiple remote key management servers and Shamir's secret sharing with threshold schemes. This maintains confidentiality and ensures that no single key server can decrypt patient data while maintaining operational systems. Homomorphic encryption of production parameters renders ciphertexts computationally usable for manufacturing tasks, as well as decryptable by authorized clinical systems for validation. Acceptable performance overhead of the privacy-preserving framework comes with cryptographically provable guarantees that patient genomic information never exists in plaintext form outside of secure sequencing laboratory environments [8].

Technique	Application Domain
Secure Multi-Party Computation (Yao's Garbled Circuits)	Scheduling participation without exposing patient identity to facilities
Homomorphic Encryption (BGV/CKKS)	Dosage calculation, drug interaction checks, and QC on encrypted genomic data
AES-256 Encryption with HSMs	Protection of identity-linked data in transit and at rest
Shamir's Secret Sharing	Distributed key management preventing single-point decryption of patient data
Data Bifurcation	Separates identity data from genetic variant data for differential protection

Table 2: Privacy-Preserving Techniques for Genomic Data in Manufacturing Workflows [7] [8]

## 5. Research Validation Approach

Design science research principles guide the design and development of the proposed artifact, and validation procedures are simulated using discrete-event modeling frameworks. Pilot implementations are deployed to pharmaceutical manufacturing partners, serving to show feasibility and quantify business value [9]. Prototyping is executed using agile software development processes via iterative sprint cycles, building the middleware framework as a containerizable microservices architecture in hybrid cloud infrastructure. On-premise data centers support sensitive genomic data processing, while the public cloud provides computational scalability. Container orchestration platforms instantiate services across available compute nodes. Continuous integration/continuous deployment pipelines instantiate automated test suites with unit tests, integration tests, and end-to-end workflow test scenarios that achieve target code coverage percentages [10].

Here, the simulations model the distributed manufacturing networks receiving and processing orders that are generated from synthetic patient genomic profiles, which are produced using population genetics algorithms in such a way that allele frequency distributions of genomic databases such as the 1000 Genomes Project and gnomAD are preserved. Numerical experiments are done with the systems under conditions of varying production requests in different time horizons and varying network topology and capacities, from the small, for example, compounding pharmacies serving local populations, to the large, for example, national networks serving larger population catchments [9]. Simulations test steady-state and seasonal surge requests (with base requests high for a longer time than for clinical trial recruitment) and outbreak response requests (pandemic-style requests that need extremely rapid scaling beyond base request levels). Failure injection testing evaluates the robustness of the system architecture via facility failures, as sections of the

10.48047/jocaaa.2026.35.02.29

network's capacity are randomly taken offline. Outages can be caused by short-term equipment failures or long-term maintenance on facilities. Performance is determined by the reallocation of workloads and patient wait times. Network partition events are used to test the eventual consistency guarantees, convergence of data reconciliation protocols, and lossless recovery mechanisms, ensuring that production orders are never lost during network partition events between facilities and central coordination systems [10].

Pilot implementations that seek to show distributed manufacture of individually tailored therapies, such as CAR-T cell therapy, personalized cancer vaccines, and pharmacogenomically guided chemotherapy, are now under way across a network of regional facilities. Implementation partners include hospital-based pharmaceutical manufacturing facilities with sterile compounding certification, academic medical center clean rooms with cell-based therapy capacity, and registered manufacturing facilities at contract manufacturing organizations for investigational new drug applications [9]. Time to treatment delivery improvements and patient wait time reductions are measured by comparing middleware-enabled treatment delivery times against historical baseline treatment delivery times by manual treatment delivery coordination, normalizing for patient and treatment complexity. Production scheduling efficiencies are measured by production facility uptime from predictive middleware-enabled production optimization, raw material waste from smart batching of compatible formulations, and labor efficiencies from automated production order routing eliminating manual coordination overhead [10]. Privacy protection validation involves third-party security audits for penetration testing, such as man-in-the-middle interception of communications, unauthorized access to the database, and side-channel attacks on the timing of encrypted computation to validate the protection of patient data from attacks. Regulatory compliance audits check electronic records compliance with requirements, the completeness of batch records for critical process parameters, and the integrity of audit trails by cryptographic validation to verify data integrity across production batches processed in the pilot (i.e., phase 3) [9].

## Conclusion

The middleware framework transforms pharmaceutical manufacturing from product-centric centralized systems to patient-centric distributed systems and enables participation in personalized medicine without huge investments in dedicated infrastructure by orchestrating heterogeneous manufacturing capabilities based on distributed capacity. Patients benefit from a shorter time-to-treatment of personalized therapies. When dealing with such an acute condition, techniques are presented regarding privacy-preserving distributed manufacturing coordination, constraint-aware scheduling algorithms for health care applications, and architectures around distributed architectural networks across the industry. Future endeavors include predictive analytics for capacity forecasting, federated learning between depots for continuous business process improvement, interfaces with other emerging biotechnologies, and equity considerations of accessing manufacturing capacity when demand exceeds available manufacturing capacity. As such, the ultimate vision for personalized medicine manufacturing networks is similar to distributed cloud computing for software: quickly and cost-effectively customized to individual patient needs while maintaining the scale, robustness, and regulatory rigor of pharmaceutical manufacturing. Another is to provide healthcare providers with new treatment delivery capabilities and regulatory agencies with increased visibility and audit trails as well as better compliance monitoring of the manufacturing process.

## References

- [1] Lyndsey Fletcher, "The \$100 Genome: Where's the Limit?" in *Front Line Genomics*, 2025. Available: <https://frontlinegenomics.com/the-100-genome-wheres-the-limit/>

10.48047/jocaaa.2026.35.02.29

- [2] S. Personalized Medicine Coalition, "Personalized Medicine Report: Opportunity, Challenges, and the Future," Personalized Medicine Coalition, 2020. Available: [https://www.personalizedmedicinecoalition.org/Userfiles/PMC-Corporate/file/PMC\\_The\\_Personalized\\_Medicine\\_Report\\_Opportunity\\_Challenges\\_and\\_the\\_Future.pdf](https://www.personalizedmedicinecoalition.org/Userfiles/PMC-Corporate/file/PMC_The_Personalized_Medicine_Report_Opportunity_Challenges_and_the_Future.pdf)
- [3] Fei Tao, et al., "Digital Twins and Cyber-Physical Systems toward Smart Manufacturing and Industry 4.0: Correlation and Comparison," *Engineering*, vol. 5, no. 4, pp. 653-661, 2019. Available: [https://www.sciencedirect.com/science/article/pii/S209580991830612X?ref=pdf\\_download&fr=RR-2&rr=9cb94fc4ffc12e96](https://www.sciencedirect.com/science/article/pii/S209580991830612X?ref=pdf_download&fr=RR-2&rr=9cb94fc4ffc12e96)
- [4] Vishal Sharma and Anand Sharma, "Blockchain Based Remote Patient Monitoring System for Healthcare Data Security," Mody University of Science and Technology. 2020. Available: [https://www.riverpublishers.com/pdf/ebook/chapter/RP\\_9788770228282C50.pdf](https://www.riverpublishers.com/pdf/ebook/chapter/RP_9788770228282C50.pdf)
- [5] Nane Kratzke and Peter-Christian Quint, "Understanding cloud-native applications after 10 years of cloud computing—A systematic mapping study," *Journal of Systems and Software*, vol. 126, pp. 1-16, 2017. Available: <https://www.sciencedirect.com/science/article/pii/S0164121217300018>
- [6] Paolo Bellavista, "A Survey of Context Data Distribution for Mobile Ubiquitous Systems," *ACM Computing Surveys*, vol. 44, no. 4, pp. 1-45, 2012. Available: <https://dl.acm.org/doi/epdf/10.1145/2333112.2333119>
- [7] Muhammad Naveed, et al., "Controlled Functional Encryption," in *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, 2014, pp. 1280-1291. Available: <https://dl.acm.org/doi/epdf/10.1145/2660267.2660291>
- [8] Jung Hee Cheon, et al., "Homomorphic Encryption for Arithmetic of Approximate Numbers," in *Advances in Cryptology – ASIACRYPT 2017*, T. Takagi and T. Peyrin, Eds. Cham: Springer, 2017, pp. 409-437. Available: [https://link.springer.com/chapter/10.1007/978-3-319-70694-8\\_15](https://link.springer.com/chapter/10.1007/978-3-319-70694-8_15)
- [9] Vijay Vaishnavi, et al., "Design Science Research In Information Systems," *designsciences*. Available: <https://designsciences.org/design-science-research-in-information-systems.pdf>
- [10] Diego Ongaro and John Ousterhout, "In Search of an Understandable Consensus Algorithm," in *Proceedings of the 2014 USENIX Annual Technical Conference*, 2014, pp. 305-319. Available: <https://www.usenix.org/system/files/conference/atc14/atc14-paper-ongaro.pdf>