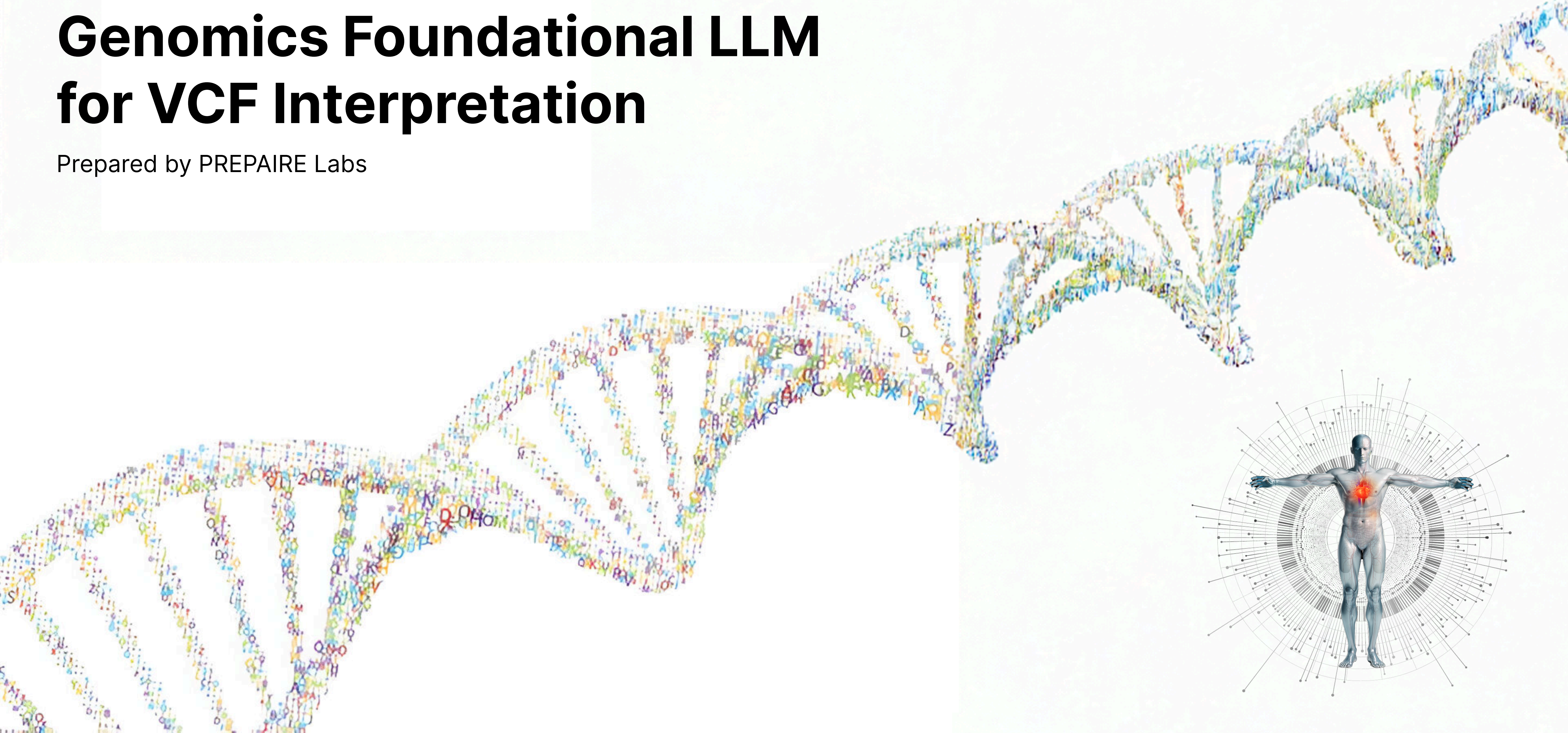




Genomics Foundational LLM for VCF Interpretation

Prepared by PREPAIRE Labs



Executive Summary

PREPAIRE Labs is a UAE-based AI and biotechnology innovation lab with a dual presence in Abu Dhabi Global Market (corporate HQ) and Masdar City (state-of-the-art R&D and wet-lab facilities). We specialize in end-to-end solutions that seamlessly integrate AI model development, secure genomic data handling, and wet-lab biological validation.

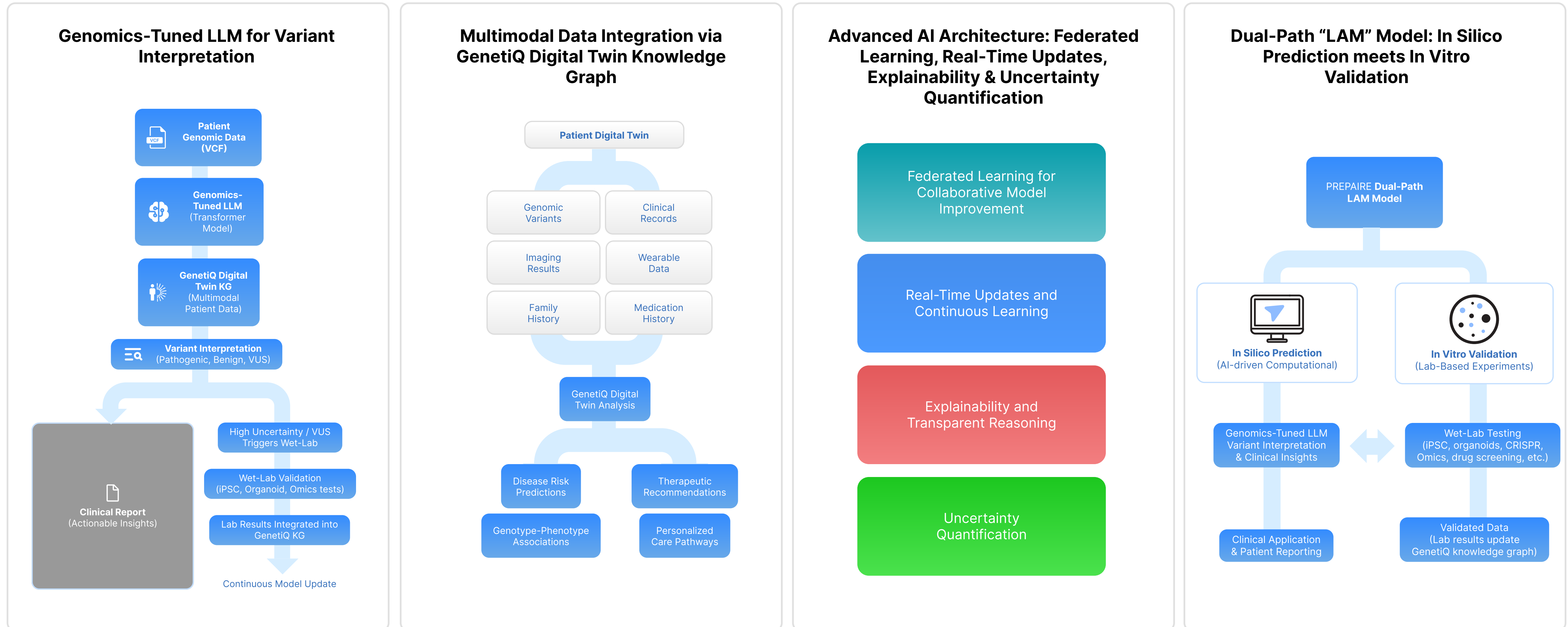
In response to the Department of Health's RFQ-6575161, PREPAIRE proposes a Genomics Foundational Large Language Model (LLM) for Variant Call Format (VCF) interpretation that aligns with Abu Dhabi's vision for advanced genomic medicine. Our solution will accelerate the interpretation of genomic variants into clinically actionable insights, leveraging local UAE infrastructure and adhering to the highest standards of data security and ethics.

PREPAIRE's value proposition lies in our holistic approach: we combine cutting-edge AI (large language models fine-tuned for genomics) with a robust "GenetiQ" digital twin knowledge graph and on-site wet-lab validation capabilities. Located in Masdar City, we are strategically positioned to collaborate with UAE healthcare stakeholders and contribute to the Emirate's genomic initiatives. Our platform directly supports the DoH's objectives by improving the speed and accuracy of variant classification for the Emirati Genome Program, all while ensuring data residency, patient privacy, and compliance with local regulations. PREPAIRE Labs' ongoing collaboration with leading healthcare providers (including a formal partnership with American Hospital Dubai) further demonstrates our commitment to deploying AI-driven genomics solutions within the UAE health system.

PREPAIRE Labs offers the Department of Health a reliable, innovative partner to develop and implement a genomics-focused LLM for VCF interpretation. Our solution will empower clinicians with rapid, explainable genetic variant analyses and recommendations, underpinned by a secure, UAE-based infrastructure and validated by world-class laboratory science. This proposal outlines our technical approach, infrastructure, validation strategy, team credentials, and roadmap to ensure the success of this initiative in alignment with Abu Dhabi's healthcare transformation goals.



PREPAIRE's technical approach combines advanced AI with comprehensive biomedical data integration and a unique feedback loop for continual learning. The system, called **GenetiQ**, is composed of the following key elements:



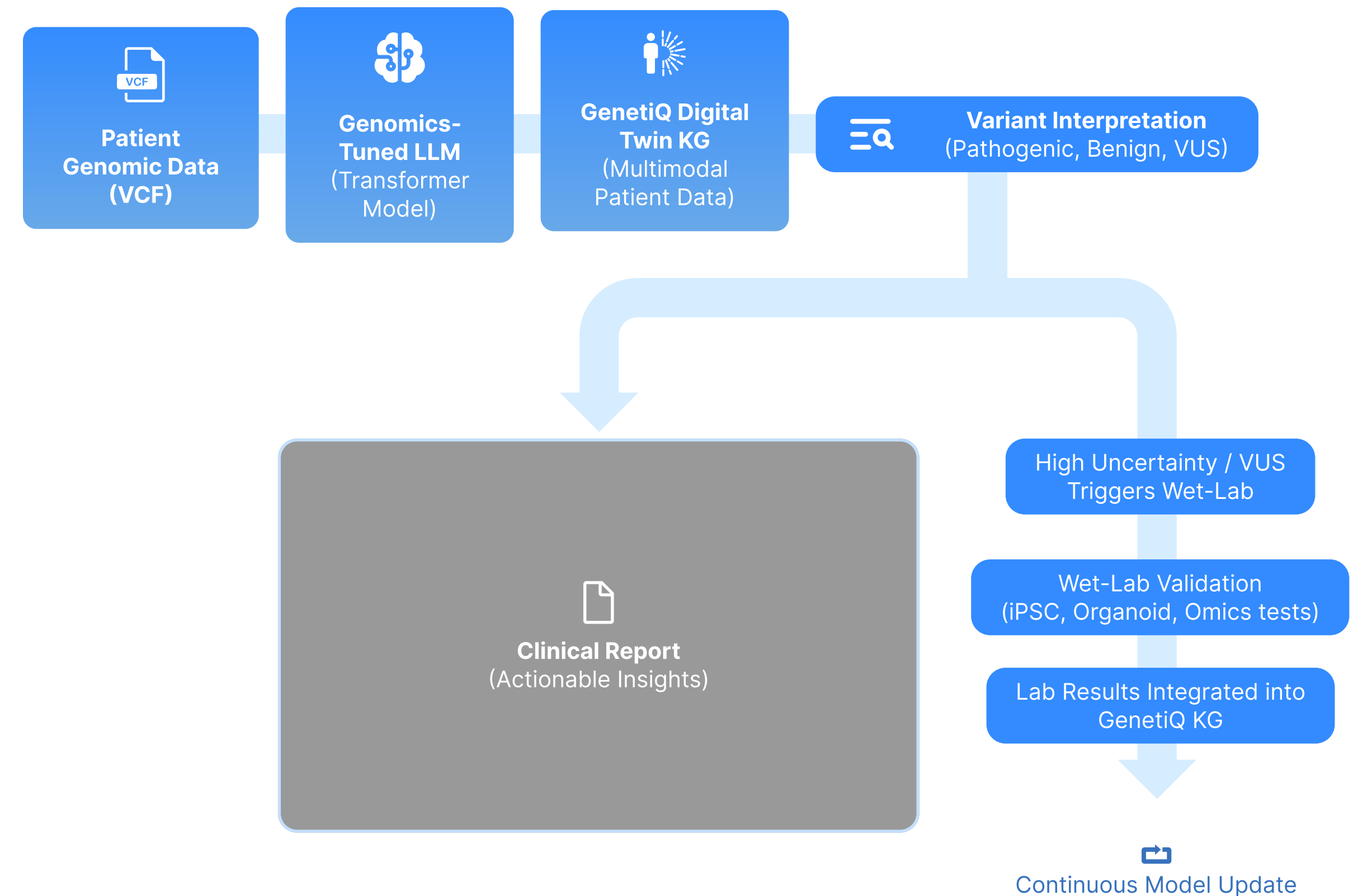
Genomics-Tuned LLM for Variant Interpretation

At the core of our solution is a Large Language Model fine-tuned for genomics. We will leverage state-of-the-art transformer-based LLM architectures and train them on vast genomic knowledge bases and clinical variant datasets. This model will be specialized in VCF interpretation, taking a patient's Variant Call Format data (genomic variants) as input and producing two primary outputs: (1) Variant Classification and (2) Clinical Report Generation.

- **Variant Classification:** The LLM will assess each genomic variant (e.g. SNPs, indels) for pathogenicity, pharmacogenomic relevance, or other clinical significance. Fine-tuning will be performed using labeled variant databases (such as ClinVar, gnomAD, HGMD) and expert-curated literature so that the model learns to classify variants according to standard guidelines (e.g. ACMG/AMP criteria). As a result, the model can distinguish benign variants from likely pathogenic ones and identify variants of uncertain significance (VUS) with calibrated confidence. The AI will also incorporate population-specific data (including Emirati genome reference information) to improve accuracy for local genomic variants.
- **Clinical Report Generation:** Beyond classification, the LLM will automatically generate a narrative report for each patient's genomic profile. This report will summarize key findings in clear clinician-friendly language, for example, highlighting pathogenic variants along with their known associated conditions, recommended confirmatory tests or specialist referrals, and potential therapy or management options if applicable. The LLM's natural language generation capabilities will ensure the reports are concise, coherent, and tailored to the patient's context. Each assertion the model makes (e.g. "Variant X is pathogenic for Condition Y") will be traceable to supporting evidence (literature citations or database references), providing transparency. This drastically reduces the manual burden on geneticists and enables faster turnaround for genomic reports, while maintaining interpretability.

By combining computational precision with language understanding, PREPAIRE's genomics-tuned LLM will serve as an intelligent assistant that can interpret raw genomic data and produce "ready-to-review" clinical genetics reports. Clinicians will remain in the loop to validate and adjust results as needed, and the LLM will learn from this expert feedback over time (continuous fine-tuning with supervised feedback is part of our approach).

PREPAIRE Genomics LLM Workflow Diagram



Multimodal Data Integration via GenetiQ Digital Twin Knowledge Graph

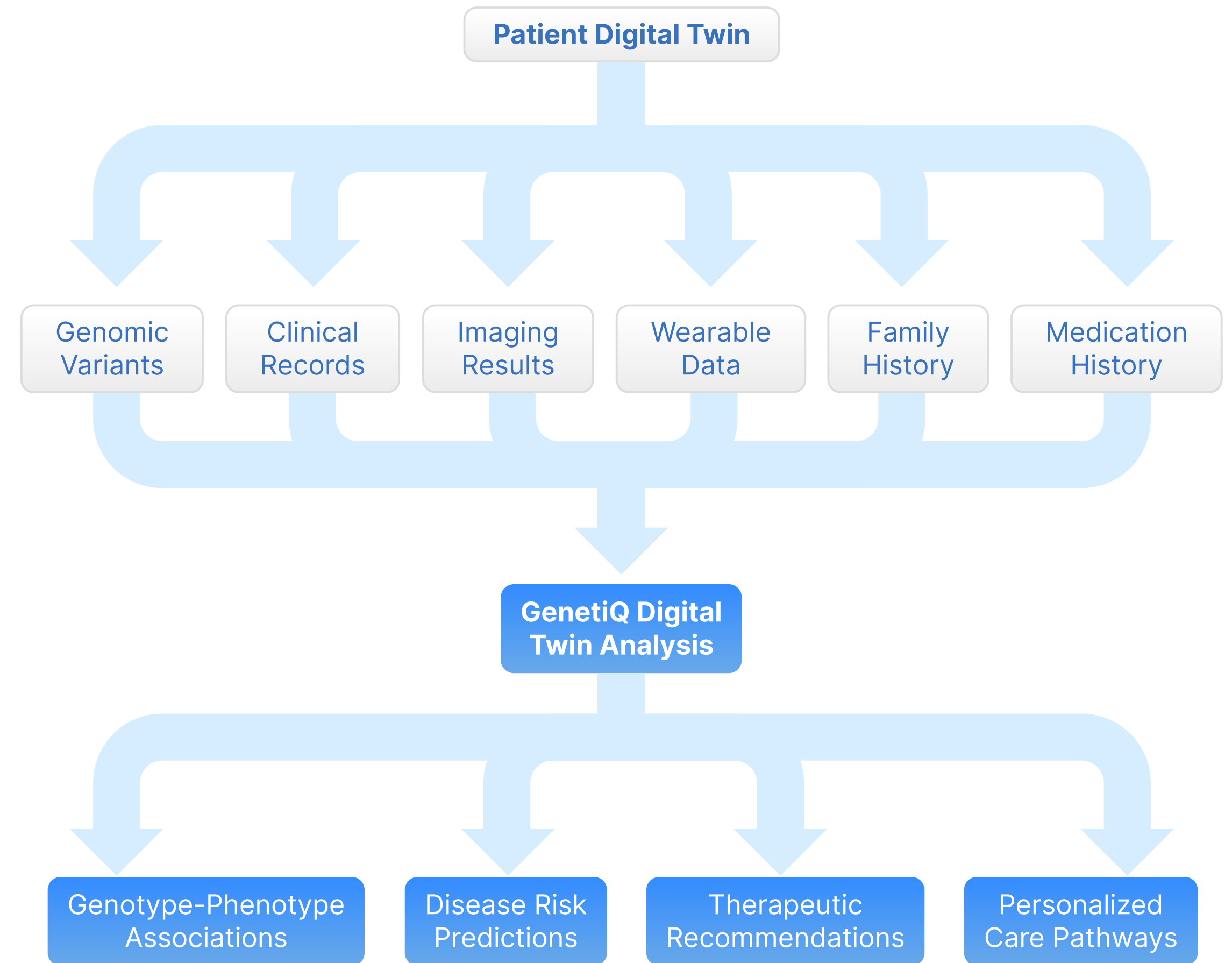
A distinguishing feature of our approach is the integration of the LLM with GenetiQ, PREPAIRE’s digital twin knowledge graph. GenetiQ acts as a comprehensive patient-centric knowledge graph that links genomic data with other health data modalities to provide rich context for interpretation. Our system will ingest multimodal data, not only genomic variants from VCF, but also the patient’s clinical records, laboratory results, imaging findings, wearable sensor data, family history, and more. All these data points are harmonized and represented in the knowledge graph as interconnected entities (e.g. genes, variants, lab test results, phenotypes, medications, pathways).

Integrating the LLM with this knowledge graph enables context-aware variant interpretation. For example, if a patient’s VCF shows a variant in a cancer susceptibility gene, the system can cross-reference the knowledge graph for that patient’s relevant history (family cancer incidence, current diagnoses, etc.) before assigning a pathogenicity score. If the variant’s impact might depend on other factors (like co-occurring mutations or environmental exposures), those relationships are captured in GenetiQ. Our approach ensures that genomic insights are not interpreted in isolation but in the full scope of the patient’s “digital twin.” This yields more accurate and personalized interpretations.

Technically, the LLM will interface with GenetiQ through structured queries and retrieval. Upon analyzing a variant, the model can query the knowledge graph for linked information (for instance, retrieving if the variant’s gene is part of a known metabolic pathway or if the patient’s lab tests show anomalies related to that gene’s function). This retrieval-augmented approach gives the LLM factual grounding and up-to-date patient-specific data, reducing hallucinations and improving the reliability of its outputs. All retrieved facts are incorporated into the model’s reasoning process, and relevant nodes in the graph can be cited in the model’s explanation for each variant. GenetiQ’s knowledge graph is continuously updated as new data comes in, meaning the LLM always works off the latest patient state.

In summary, multimodal integration via the GenetiQ knowledge graph allows our genomic LLM to operate as part of a larger intelligent system, one that “sees” the whole patient. This leads to more holistic and clinically insightful interpretations, bridging the gap between raw genomic data and practical health decision-making.

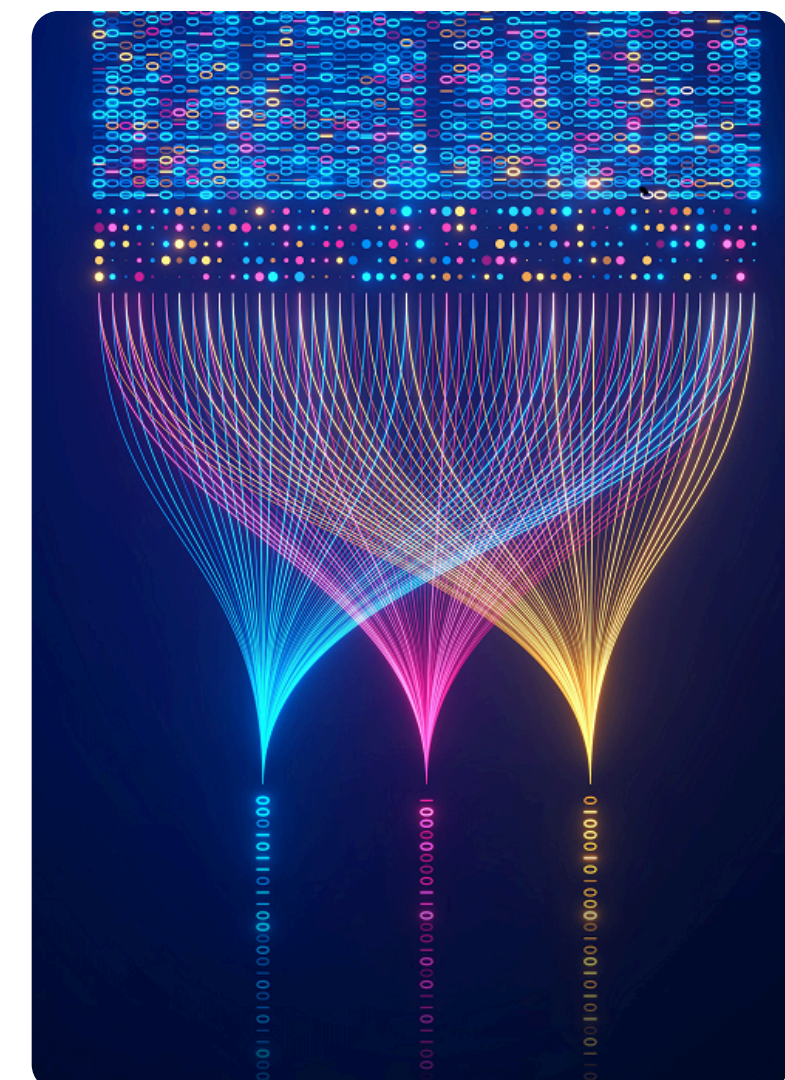
GenetiQ Digital Twin Knowledge Graph Infographic



Advanced AI Architecture: Federated Learning, Real-Time Updates, Explainability & Uncertainty Quantification

Our AI architecture is built for robustness, scalability, and trustworthiness. Key aspects of the design include:

- **Federated Learning for Collaborative Model Improvement:** To leverage data from across Abu Dhabi (and potentially the wider UAE) while preserving data privacy, PREPAIRE's solution uses a federated learning topology. Each participating hospital or institution will run a local instance of the GenetiQ system (including the LLM and knowledge graph) on-premises, training on their own genomic data. No raw genomic data or patient information leaves the local servers; instead, only encrypted model parameter updates are shared with a central aggregator. The central server hosts the global "Learning Across Medicine" (LAM) model, which periodically aggregates the updates from all sites to form an improved global model. This updated model is then sent back to each site to sync their LLM. This cycle (also known as Federated Averaging) allows the model to "learn" from the collective experience (e.g. variant interpretations) of all hospitals without breaching data sovereignty or privacy. Each hospital benefits from a more knowledgeable model (e.g. recognizing rare variants seen elsewhere) while ensuring patient genomic data remains local and confidential. Federated learning also means our system naturally scales as more data sources (hospitals, labs) are added, fostering a network effect in model accuracy. PREPAIRE's federated approach directly addresses the data silo challenges in healthcare, aligning with the UAE's strict data protection requirements.
- **Real-Time Updates and Continuous Learning:** The GenetiQ platform is designed for real-time operation so that the digital twin and LLM remain current with incoming data. Streaming data ingestion is enabled – as soon as new information is available (a new lab result, a newly published gene study, or a reclassification of a variant in global databases), the system updates the knowledge graph and notifies the LLM. For instance, if a patient's follow-up blood test or MRI result arrives, that data point is attached to their graph, and the LLM can immediately reconsider any relevant variant interpretations in light of the new evidence. Similarly, if global knowledge of a certain variant changes (e.g. a variant of unknown significance is now found to be pathogenic in recent literature), we will update the model's knowledge base promptly. Through this continuous update mechanism, the digital twin behaves like a "living model" that evolves with the patient's journey rather than a static one-off analysis. This is critical in genomics where interpretations can change as new research emerges. Real-time processing ensures clinicians always receive the most up-to-date and relevant insights.



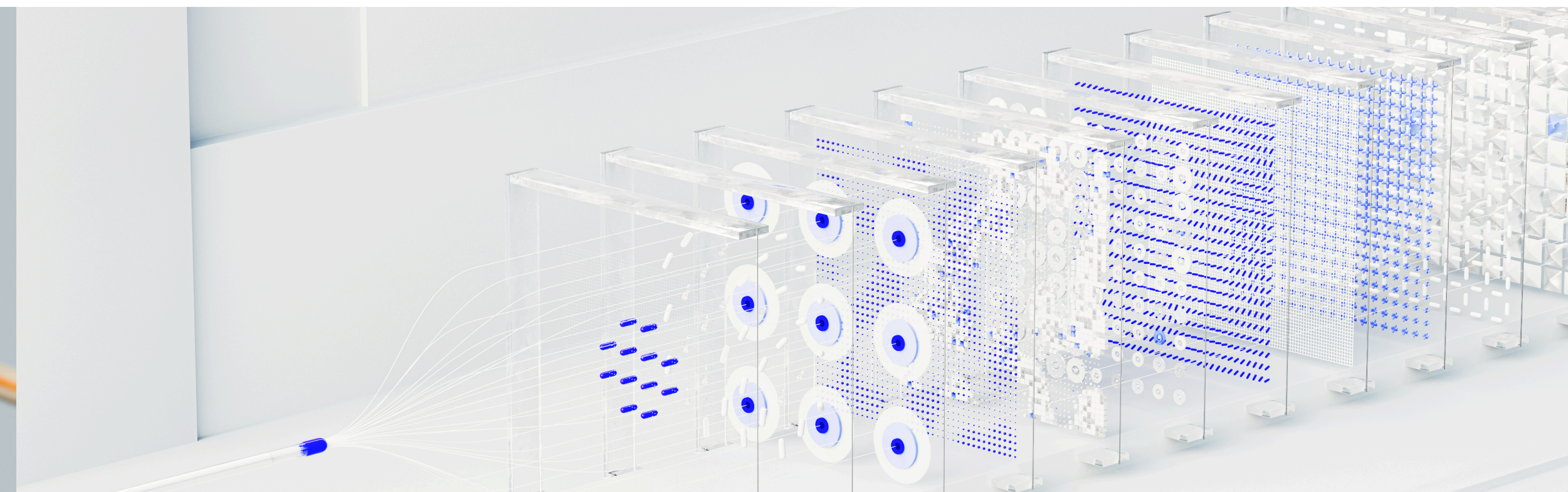
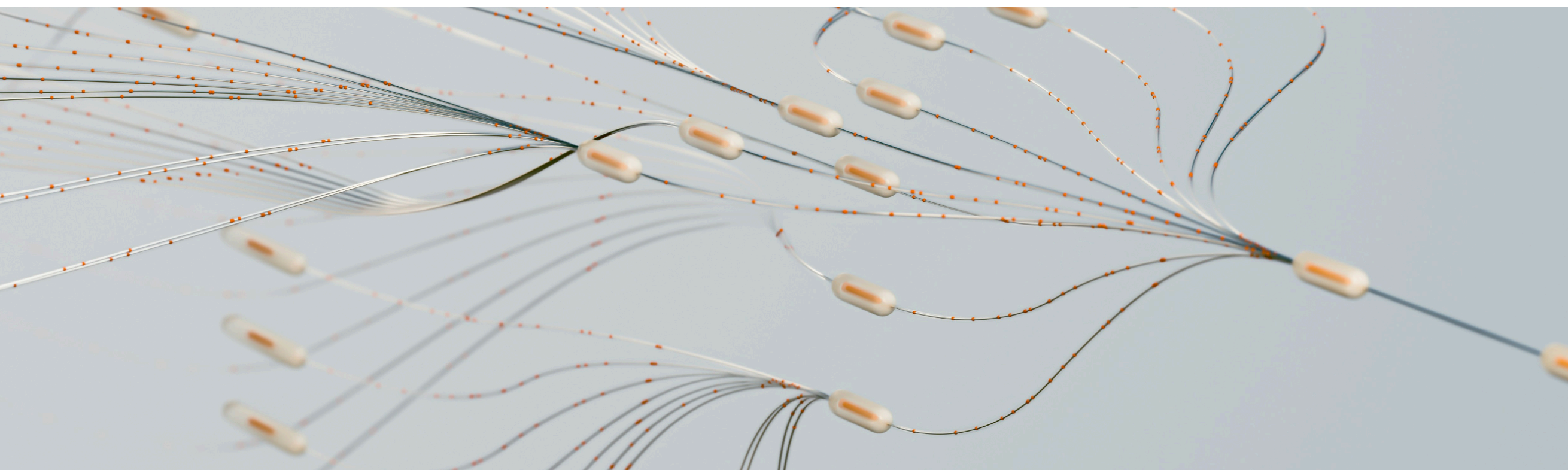
Advanced AI Architecture: Federated Learning, Real-Time Updates, Explainability & Uncertainty Quantification

- **Explainability and Transparent Reasoning:** We recognize that AI-driven genomic interpretations must be explainable for clinical acceptance. Our LLM and knowledge graph approach is inherently designed to provide clear justifications for each output. The system will surface evidence and reasoning paths for every variant classification or recommendation. For example, if the model flags a variant as pathogenic, it will also present the key supporting evidence: “Variant X is classified pathogenic due to [known functional studies], [frequency in population databases], and [patient’s phenotype correlation].” It may reference specific knowledge graph links such as a known disease-gene association or a published study node. This graph-based reasoning can even be visualized as a chain of connections (e.g. Variant → Gene → Pathway → Observed Phenotype) to make the decision process transparent. We will implement an Explainable AI (XAI) module that outputs a summary of factors and their weights contributing to each decision. Clinicians can drill down into these factors if needed. By designing the LLM to cite its sources (e.g. literature, database IDs) and by using the structured knowledge graph backbone, our solution ensures that nothing is a “black box.” This commitment to explainability builds trust and aligns with ethical AI principles.

Through this multi-faceted AI architecture, PREPAIRE’s solution will be smart, continuously improving, and trustworthy. The federated learning with a global LAM model guarantees the AI stays at the cutting edge of knowledge across all participating facilities, while real-time updates, explainability, and uncertainty handling make the system robust and acceptable in a clinical setting.

- **Uncertainty Quantification:** In genomic medicine, knowing what the AI does not know is as important as its predictions. Our model will provide confidence scores or probability estimates for each variant classification and highlight variants that are predictions with high uncertainty. For example, a variant of uncertain significance (VUS) may be flagged as such with a note that additional data or functional testing is required. The LLM will be trained to express uncertainty in its generated reports (e.g. “The pathogenicity of variant ABC is uncertain; it is recommended to interpret with caution and consider further testing”). Under the hood, we will use techniques like Monte Carlo dropout or ensemble methods with the model to gauge uncertainty intervals. Moreover, by cross-checking against the knowledge graph, the system can identify when it lacks information (say, the patient has a rare variant that isn’t in any database); those cases are elevated for human expert review or wet-lab validation. This calibrated uncertainty quantification ensures that the system’s outputs are safe and reliable for clinical use, by avoiding overconfident misclassifications. It also helps prioritize which variants or findings need additional evidence generation (triggering the wet-lab validation path, described next).

Through this multi-faceted AI architecture, PREPAIRE’s solution will be smart, continuously improving, and trustworthy. The federated learning with a global LAM model guarantees the AI stays at the cutting edge of knowledge across all participating facilities, while real-time updates, explainability, and uncertainty handling make the system robust and acceptable in a clinical setting.



Dual-Path “LAM” Model: In Silico Prediction meets In Vitro Validation

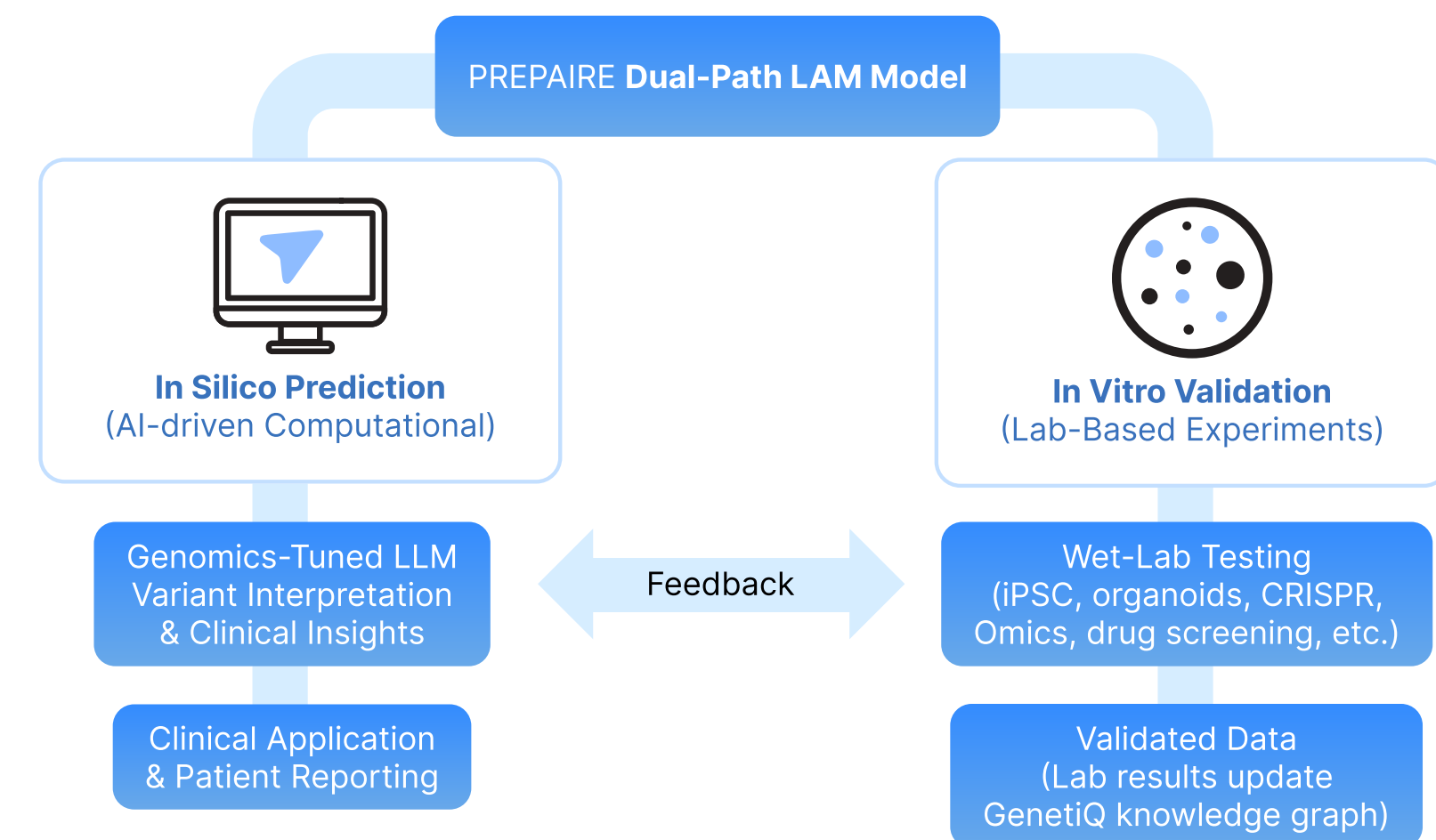
PREPAIRE Labs uniquely integrates in silico AI predictions with in vitro wet-lab validation, creating a powerful dual-path development model. At the heart of this approach is our “Learning Across Medicine (LAM)” model, which serves as the bridge between computational predictions and experimental confirmation. The LAM architecture has two complementary pathways:

- **Path 1:** In Silico Prediction (Digital Path), This is the AI-driven pipeline described above, where the LLM analyzes genomic data (and other patient data) to predict variant effects and suggest clinical actions. The output might include predictions such as “Variant X likely disrupts protein Y function and may cause Condition Z” or “A novel drug/remedy could ameliorate the patient’s genomic pathway imbalance.” These predictions are generated rapidly and iteratively as data updates, providing hypothesis and insight in silico.
- **Path 2:** In Vitro/Wet-Lab Validation (Physical Path), For critical or novel predictions, our platform triggers a real-world experimental validation using PREPAIRE’s wet-lab facilities in Masdar City. This means that if the AI identifies, for example, a previously uncharacterized variant as potentially pathogenic or recommends a novel therapeutic approach for a patient’s genomic profile, we can validate these findings in the lab. We leverage our in-house biotechnology services – including induced pluripotent stem cell (iPSC) technology, organoid models, and multi-omics assays – to test the AI’s hypotheses under controlled experimental conditions.

This dual-path workflow ensures that AI predictions are grounded in biological reality. Concretely, our system will utilize a “Wet-Lab Validation SDK and Feedback Loop” integrated with GenetiQ. For example, if the LLM flags a high-impact variant of uncertain significance, researchers can quickly design a functional assay: using CRISPR, introduce that variant into an iPSC-derived cell line or 3D organoid relevant to the affected tissue, then observe the impact on cell function or gene expression. Our SDK standardizes this process, from automating lab protocols to capturing the results. As another example, if the AI suggests a patient might respond to a particular drug or nutrient based on their genome, we can test that intervention on a patient-derived organoid or cell culture in our lab to see if the expected effect (say, reduction in a disease biomarker) occurs. These wet-lab experiments act as a safety net and refinement tool for the AI’s recommendations.

The results from in vitro tests flow back into the knowledge graph and LAM model as feedback. If the lab results confirm the AI prediction (e.g. the variant indeed causes a loss of function in the protein), the system gains confidence in that prediction and solidifies that knowledge for future cases. If the lab findings contradict the AI (perhaps the predicted drug had no effect or an unexpected effect), the model will adjust itself, lowering the confidence in that suggestion and analyzing the discrepancy to improve its algorithms. This continuous feedback loop between prediction and experimentation creates a self-correcting system that gets smarter and more accurate over time. It is a true closed-loop learning system where “the lab informs the model, and the model informs the lab.”

By supporting both in silico and in vitro paths, PREPAIRE ensures that the Genomics LLM is not only powerful, but also validated in a clinically relevant manner. This dual-path approach de-risks the deployment of AI in genomics: any novel or high-stakes recommendation is validated with real biological evidence before being fully adopted. It also accelerates translational research, as promising AI-driven discoveries can be rapidly tested in our Masdar City labs. In summary, the dual-path LAM model is a cornerstone of our approach, guaranteeing that the delivered solution will be both cutting-edge and empirically sound.



PREPAIRE Labs will deploy the Genomics LLM and GenetiQ platform on a secure, scalable infrastructure within the UAE, in full compliance with local data regulations and best-in-class security practices. Key infrastructure and security features include:

- **UAE-Based Cloud & On-Premises Hosting:** The solution can be hosted on secure UAE cloud environments or on-premises data centers as required by DoH. Our primary deployment will reside within UAE borders – for example, using Abu Dhabi-based Tier-III+ data centers or government-approved cloud services – to ensure all genomic data and patient information remain in-country. Additionally, each participating hospital can have an on-site server/appliance for the GenetiQ instance, aligning with federated learning design. This hybrid cloud architecture provides both central availability and local control. It is designed for high availability, redundancy, and scalability to accommodate the growth of the Emirati genomic program.
- **Quantum-Safe Network Architecture (Nokia ANYSEC/IP-MPLS):** PREPAIRE's network backbone is built in partnership with Nokia to be quantum-safe and ultra-secure. All data communications, whether between hospital sites and the central server or between our cloud and lab environments, are protected using Nokia's ANYSEC and MACsec encryption technologies over an IP/MPLS network. ANYSEC (Any Secure Transport) and MACsec (Media Access Control Security) provide end-to-end encryption at the network layer, ensuring that sensitive genomic data in transit is safeguarded against eavesdropping or interception. These technologies are chosen to be resilient even against future quantum computing threats, as they support next-generation cryptographic algorithms. The network design follows a defense-in-depth strategy, with multi-layer encryption and continuous authentication of nodes, creating a secure communication fabric for all genomic data exchanges. In practical terms, when model updates are sent in the federated learning process or when users access the system remotely, the underlying network encryption guarantees confidentiality and integrity of the data packets.

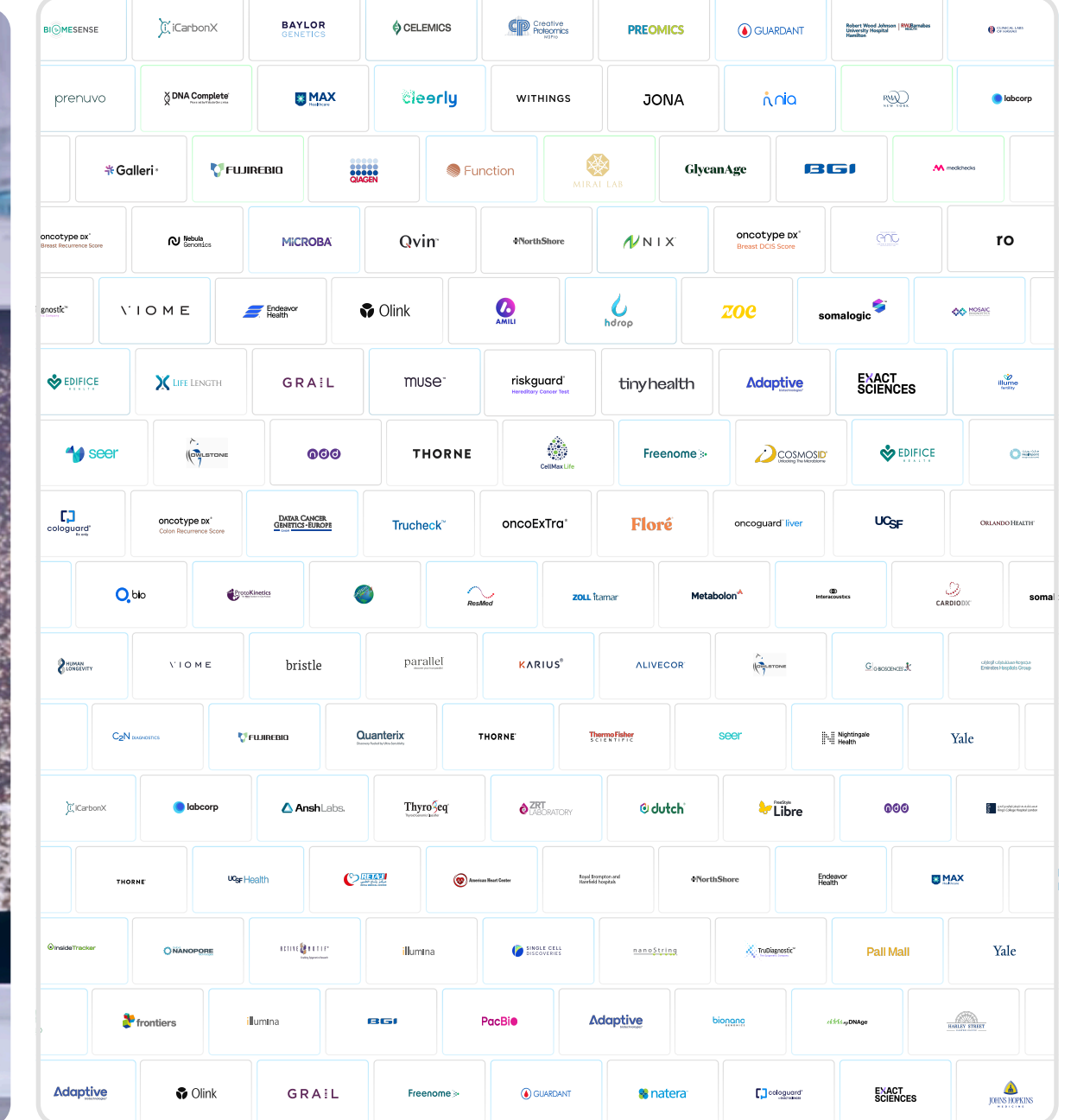
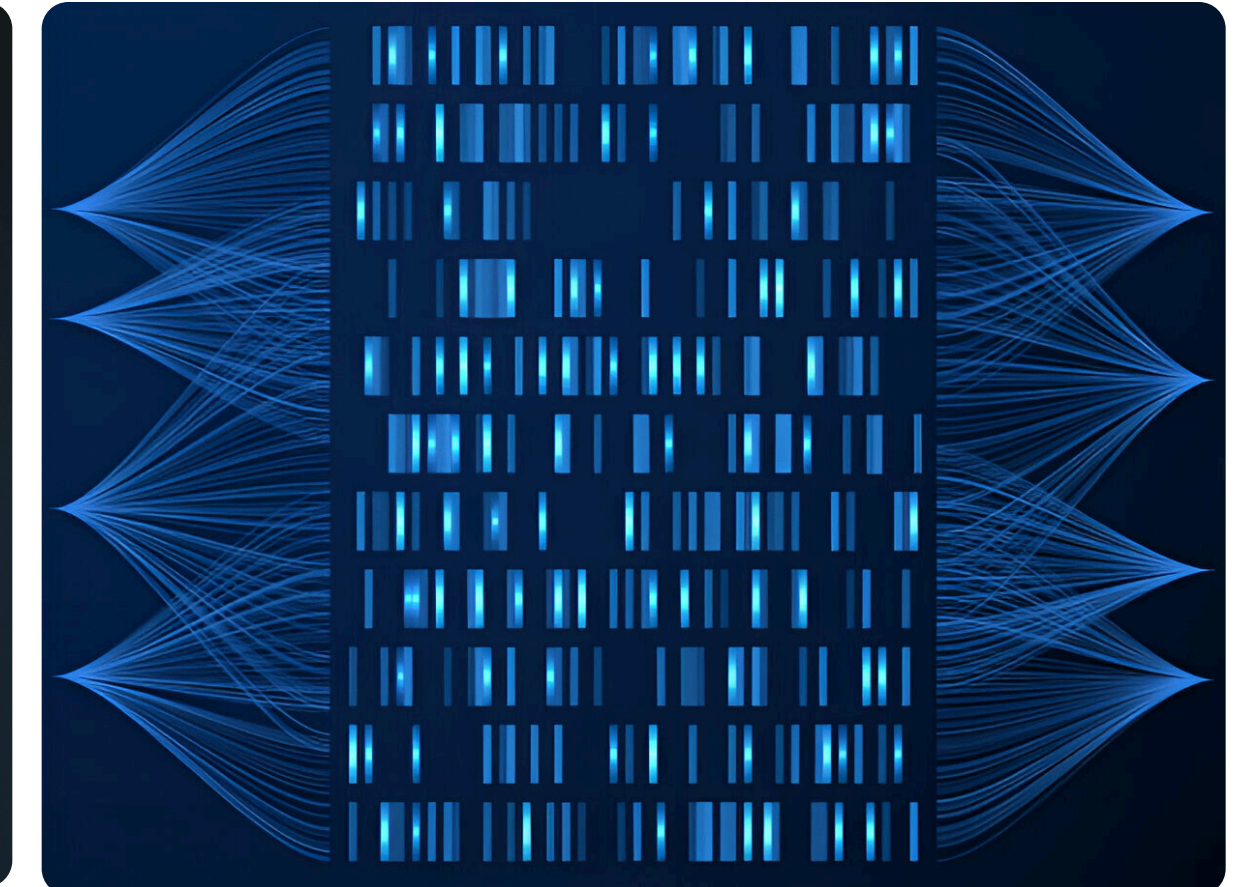
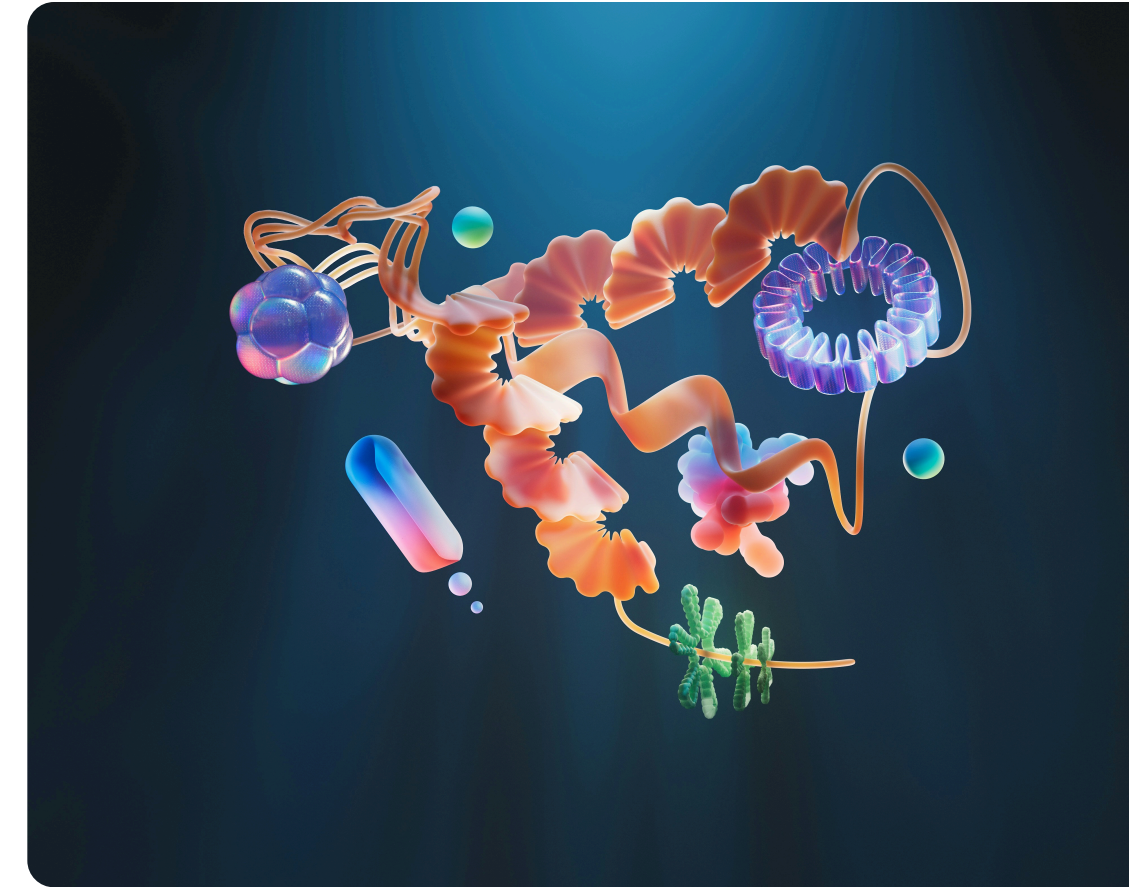
- **Hardware Security Modules (Entrust nShield HSMs):** To protect sensitive data at rest and cryptographic keys, we incorporate Entrust nShield Hardware Security Modules in our infrastructure. All encryption keys, digital certificates, and signing operations (for example, code signing of models or encryption of the genomic database) are handled inside FIPS 140-2 Level 3 certified HSMs. This ensures that private keys never leave the hardware device and are safe from compromise. HSMs also enable features like secure boot and tamper-evident logging for our servers. Genomic data stored in databases will be encrypted with strong algorithms (AES-256) with keys managed by the HSM (using techniques like key wrapping and secure key custody). By using HSMs, we greatly reduce the attack surface and comply with stringent healthcare security requirements for protecting patient data. This is especially important given the sensitivity of genetic information.
- **Compliance with UAE Data Residency Laws and NESAs Standards:** Our entire infrastructure and security setup is aligned with UAE federal and emirate-level regulations. All genomic data and patient identifiable information will reside within the UAE, in accordance with the UAE Genome Law (2023) requirement that genomic data be stored in the national databases within country borders. We also adhere to the Abu Dhabi Department of Health's data management policies and the National Electronic Security Authority (NESA) guidelines. The system's security controls (access control, monitoring, encryption, incident response) are mapped to the UAE IA Standards as defined by NESA, which include ~188 security controls aligned with ISO 27001 and other international benchmarks. From strong perimeter firewalls and intrusion detection systems to regular vulnerability assessments and audits, we will maintain a NESA-compliant posture. Moreover, data access within the platform will be governed by role-based access control and consent management in line with UAE health data privacy laws. In summary, PREPAIRE's solution will meet or exceed all UAE data residency, privacy, and cybersecurity requirements, giving DoH and patients full confidence that their genomic data is handled with utmost care and legal compliance.

In addition to the above, PREPAIRE's Masdar City facility itself is a secure environment (with controlled lab access, CCTV, and compliance to lab safety standards). The combination of robust cloud infrastructure, a quantum-safe network, HSM-based key management, and strict policy compliance makes our deployment highly secure and resilient against threats. DoH and its stakeholders can be assured that the genomic LLM platform will operate within a fortified IT environment engineered for healthcare.

Clinical and Laboratory Validation

A core strength of PREPAIRE Labs is our on-site wet-laboratory capabilities in Masdar City, which we will harness for rigorous clinical and experimental validation of the genomic LLM's outputs. This end-to-end capability, from data to bench, ensures that the AI's predictions are not only theoretically sound, but also empirically verified in biologically relevant systems. Our clinical and lab validation approach includes:

- **Advanced Wet-Lab Facilities in Masdar City:** PREPAIRE operates cutting-edge laboratory facilities co-located with our AI development teams. These include cell culture labs, molecular biology labs, and high-throughput screening platforms, all within Abu Dhabi. We have the ability to perform induced Pluripotent Stem Cell (iPSC) reprogramming, taking patient-derived cells (e.g. blood or skin cells) and reprogramming them into iPSCs, which can then be differentiated into specific cell types (such as cardiomyocytes, neurons, hepatocytes, etc.). This allows us to create patient-specific cellular models carrying the patient's exact genome. Additionally, we maintain 3D organoid culture systems (miniaturized organ-like structures grown from stem cells) for modeling diseases in vitro. These organoids (e.g. brain organoids, liver organoids) can mirror a patient's tissue and are invaluable for testing variant effects and drug responses in a physiologically relevant setting.
- **Multi-omics Profiling and Functional Assays:** Our lab is equipped for comprehensive omics profiling, including genomics (next-generation sequencing), transcriptomics (RNA-seq), proteomics (mass spectrometry), and metabolomics. When the AI flags a variant or suggests an intervention, we can design targeted wet-lab experiments to validate those findings. For example: if a variant is predicted to alter a metabolic pathway, we can measure changes in metabolite levels and gene expression in cells engineered with that variant. We routinely perform functional assays such as gene editing (CRISPR/Cas9) to introduce or correct variants in cell lines, enzymatic activity assays to see if a gene's function is disrupted, and cellular phenotype assays (proliferation, apoptosis, differentiation metrics) to observe any abnormal behavior caused by the genetic change. We also use high-content imaging (e.g. Cell Painting assays) to capture subtle cellular changes due to a genetic variant or treatment. All these laboratory techniques provide a means to experimentally confirm the significance of genomic variants that the LLM identifies as potentially pathogenic.



- **Validation of Variant Classification:** For variants of high clinical importance (pathogenic or novel VUS with strong suspicion), our protocol will involve functionally validating those variants. For instance, if the LLM suggests that a rare variant in the BRCA1 gene is likely pathogenic, we can collaborate with clinical geneticists to obtain patient-derived samples (with consent) and test DNA repair function in cells with that variant versus normal cells. Similarly, if a variant in a cardiac ion channel gene is flagged, we can create iPSC-derived cardiomyocytes with that variant to see if they exhibit irregular electrical activity. These laboratory validations provide definitive evidence to support the AI's classification, which can then be reported back to clinicians and even published as novel findings. This process transforms uncertain variants into well-characterized ones, improving the knowledge base for all.
- **Validation of Intervention Recommendations:** In some cases, the LLM (in conjunction with the knowledge graph) may recommend certain interventions, for example, suggesting a specific drug, gene therapy, or dietary supplement that could mitigate the effect of a patient's genetic variant (especially in the context of complex multi-genic conditions). PREPAIRE's wet-lab can test these recommendations in ex vivo models. If the AI suggests Drug A might reduce the harmful effects of Variant X, our team can treat patient-derived organoids or cell cultures with Drug A and measure outcomes (e.g. does a disease biomarker level decrease?). We follow standard pharmacology and toxicology protocols for any such testing to ensure safety and efficacy signals are properly evaluated. This gives clinicians an added layer of confidence before acting on AI-derived therapeutic suggestions. Essentially, our lab serves as a pre-clinical testbed for personalized interventions proposed by the digital twin, helping tailor truly personalized medicine.

- **Feedback Loop to AI Model:** Crucially, all results from the wet-lab validations are fed back into the GenetiQ knowledge graph and LLM training data. We have established data pipelines to capture experimental outcomes in a structured format. For example, if a lab experiment confirms that "Variant X → Loss of enzyme activity → Disease Y", that relationship is added as a new edge in the knowledge graph, and the LLM's training is updated to strengthen its association between Variant X and Disease Y. Conversely, if an experiment nullifies a predicted association, that information will adjust the model's parameters to reduce false associations. This closed-loop learning (as described in the Technical Approach) means the system gets more accurate with each validation. Over time, as we validate more variants and interventions, the need for manual lab validation will decline, because the model will have seen similar patterns before and "knows" the ground truth. In essence, the wet-lab work at Masdar serves to continuously calibrate and improve the AI model's performance in a way no purely digital system could.

Clinical Collaboration for Outcomes: Our validation workflow is conducted in partnership with clinicians and genetic counselors. Any lab test results are correlated with patient outcomes. If our lab finds a certain variant causes a cellular defect, we work with healthcare providers to see if that correlates with the patient's symptoms and if any clinical action was taken improved patient health. This allows us to also measure the real-world clinical impact of our AI-guided interventions. We will document cases and outcomes, contributing to evidence that can support regulatory approvals and broader adoption of the AI system in genomics practice.

Through the above strategies, PREPAIRE ensures that the delivered LLM solution is clinically validated end-to-end. We do not just stop at providing an AI prediction; we help verify those predictions in the lab and clinic, ensuring the Department of Health receives a solution that is scientifically sound and medically actionable. This integrated approach, combining digital intelligence with wet-lab evidence, is at the heart of PREPAIRE Labs' philosophy and a key differentiator that will greatly benefit the DoH's genomics initiative.

PREPAIRE Labs has assembled a multidisciplinary team of experts and forged key partnerships to guarantee the success of the Genomics LLM project. Our team and collaborators bring together AI prowess, genomic science expertise, clinical insight, and operational experience:

Executive Sponsors



Dr. Hoda Alkhzaimi

Co-Founder

Dr. Hoda Alkhzaimi is the Associate Vice Provost for Innovation & Entrepreneurship at NYU Abu Dhabi, with 20 years of experience in technology innovation, strategic investments, and secure computing. She currently leads research translation and innovation at NYU Abu Dhabi, specializing in transforming academic research into market-driven solutions. Dr. Alkhzaimi founded the Center for Cyber Security at NYUAD, significantly advancing cybersecurity, AI, and quantum technologies in the UAE. She advises international forums such as the World Economic Forum and the G20 on emerging technologies and data security. Dr. Alkhzaimi holds a doctorate in Computer Science and Mathematics from Denmark Technical University.



Dr. Vicent Ribas

Co-Founder and CSO

Dr. Vicent Ribas Ripoll is CSO at Prepaire Labs. He is a Senior Member of the IEEE and serves as the coordinator of the AGAUR-recognized research group on Data Analytics in Medicine and Omics Integration. He has also served as Co-Chair of the European Commission's FET Open program.

Dr. Ribas holds the title of Associate Professor accredited by AQU and has taught at both the University of Girona and the Polytechnic University of Catalonia (UPC). He began his professional career at Accenture, where he developed an early interest in artificial intelligence (AI) and its applications in healthcare. He also brings extensive experience as an entrepreneur and AI consultant.

His research has focused on sepsis, inflammatory diseases, cardiovascular modeling, and cancer. In 2012, he received the Best Scientific Poster Award at the international congress of the Society for Critical Care Medicine. Dr. Ribas holds a PhD in Artificial Intelligence, a Master's in Mathematical Engineering, and a Telecommunications Engineering degree from UPC. He also earned a Master's in Electrical Engineering from KTH Royal Institute of Technology in Stockholm.



Dr. Ashraf Allam

Co-Founder

Senior executive, with decades of experience in Pharma. Former CEO of the PIF (Pharmaceutical Investment Fund the largest Pharma Sovereign fund in the world), Regional Managing Director of MundiPharma, Regional Managing Director of AmGen for the Middle East, Africa and Turkey. Managing Director of J&J Egypt Head of Global Economic Affairs for Eli Lilly and Company in US. He has served as elected Chairman, Executive Committee for the association of American, European and Japanese research-based manufacturing pharmaceutical and Biotechnology companies (PhRMA) in the Middle East and Africa region. Executive board member for the medical devices and diagnostics association (Mecomed) for the MEA region Elected chairman of the board for the self-medication industry for the MENA region (MENAP-SMI)



Abdulla Al Mansoori

Co-Founder & Government Relations

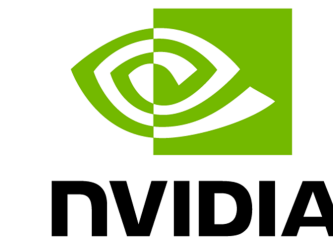
Abdulla Al Mansoori brings over 19 years of experience in senior leadership roles in prominent Dubai corporations including the Dubai International Financial Centre (DIFC), TECOM Investments, Dubai Biotechnology and Research Park (DuBiotech) and Dubai Media Office. Al Mansoori was Deputy CFO and Head of Finance Operations of DIFC and the Chief Executive Officer of DIFC Arbitration Centre. He has also served as Chief Operating Officer of RAPID, an innovative payment solutions initiative launched by DIFC.

Mr. Al Mansoori was also part of the creation of Global Response Aid during COVID-19 and created a successful alliance between Agility, Dr Readys and Fujifilm Toyama for the global distribution of an anti-viral called Avigan.

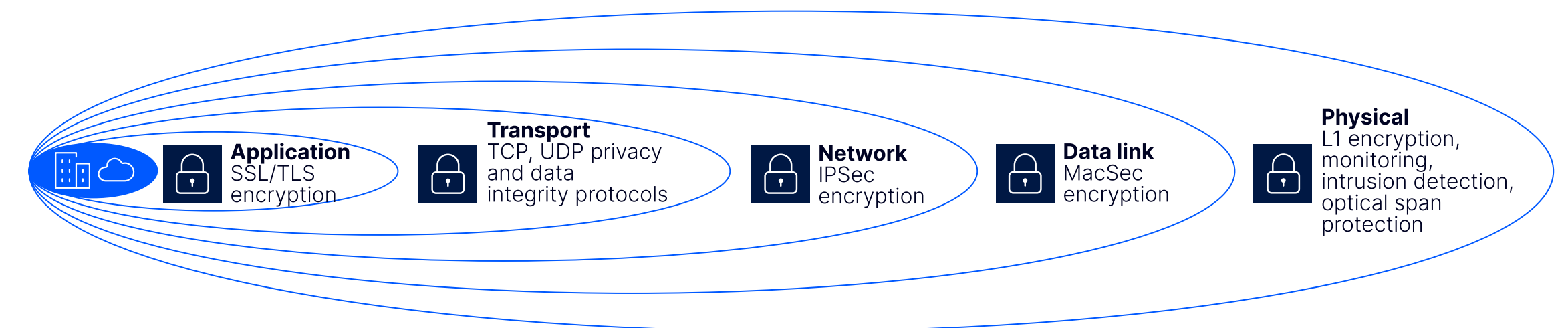
Together, this core team covers the spectrum from AI development to clinical interpretation to lab science, under strong leadership. The team will meet regularly with DoH's project representatives to report progress and incorporate feedback. In addition to these individuals, PREPAIRE Labs can draw on a broader staff of data scientists, software engineers, bioinformaticians, and lab technicians as needed during the project.

- **Masdar City and ADGM** – Strategic Location Partners: Being headquartered in ADGM and operating in Masdar City gives us strategic advantages, including strong support from the Abu Dhabi innovation ecosystem. While not a traditional “partner,” it is worth noting that our presence in these hubs aligns with UAE’s vision for healthcare innovation. We collaborate informally with entities in Masdar City (such as the Masdar Institute/Khalifa University labs) for knowledge exchange. The proximity to other tech and biotech companies fosters a collaborative environment that the DoH project can tap into if needed (for example, accessing specialized equipment or shared testing facilities in Masdar).
- **Industry Partners (Technology):** On the technology front, as described earlier, Nokia is a key partner providing the networking hardware and solutions for our quantum-safe infrastructure. Additionally, Entrust is our partner for security modules. We also have relationships with cloud providers (e.g. we utilize AWS for some components in a UAE region cloud, as noted in our AHD collaboration), however, for this project, all cloud usage will be strictly within UAE or via secure DCI to ensure compliance. We mention these to show that PREPAIRE has an ecosystem of reputable partners to support various project dimensions, from networking to hardware to cloud, ensuring we can deliver a robust end-to-end solution.

Through the combined expertise of our team and the support of our partners, PREPAIRE Labs is confident in executing this project successfully. We have the right people in place, individuals who are not only experts in their fields but also deeply committed to advancing genomic medicine in the UAE. Our partnerships further strengthen our capabilities by providing real-world clinical input and technical reinforcement. This collaborative approach will ensure that the DoH receives a solution that is scientifically sound, clinically validated, and seamlessly implementable in the healthcare system.



Quantum Encryption for Data Security



Regulatory Alignment



PREPAIRE Labs fully understands the critical importance of adhering to all relevant UAE laws, regulations, and ethical guidelines in the realm of genomics and AI. Our proposal and platform design have been crafted to ensure complete regulatory compliance and ethical integrity. Key areas of alignment include:

- **UAE Genome Law (Federal Decree Law No. 49 of 2023):** Our solution is built in strict compliance with the recently enacted UAE Genome Law of 2023, which governs the use of human genomic data. Specifically, we uphold informed consent requirements by ensuring that any use of patient genomic data in our system is authorized by the patient (or legal guardian) with a clear consent process integrated into the workflow. We also ensure the right to withdraw, patients can request their data or analysis to be removed, and our system will accommodate such requests, reflecting the law's patient-rights provisions. The law mandates that genomic data and information be stored in the UAE's national databases or secure local facilities; our data residency approach (UAE-based hosting) directly supports this, as we do not export genomic data abroad. Additionally, we support the Genome Law's vision of establishing a national genomic database and Emirati genome reference by structuring our knowledge graph to feed anonymized insights (with DoH's direction) into national repositories, thus contributing to the country's genomic knowledge base in a compliant manner. We also acknowledge the law's prohibition on unauthorized genetic modifications; our wet-lab validations do not involve any germline modifications or non-therapeutic alterations, and are purely for in vitro research purposes with regulatory oversight.
- **Department of Health Genomics Policy and Guidelines:** We align closely with the Abu Dhabi DoH's policy on genomics and related guidelines (e.g. for clinical genetic testing, research, data sharing). Our platform was designed with the DoH Genomics Policy Framework in mind, addressing its key pillars: Service, Data, Research & Innovation, Resources. For instance, Policy Objective 2.1 from DoH's genomics policy is to "ensure protection of genomic data during collection, storage & use", our security architecture and consent management directly fulfill this objective through strong protection controls and governance. Policy Objective 1.1 calls for scaling up genomic services and integrating precision medicine in healthcare; our LLM solution is a direct enabler for scaling genomic interpretation services to more patients efficiently. We also contribute to Policy Objective 3.2 (promote public-private partnerships) by virtue of this collaboration between DoH and PREPAIRE, a private innovation lab, to deliver public value in healthcare. All genomic analyses conducted by our system will follow clinical best practices, for example, using ACMG guidelines for variant classification and participating in any DoH-mandated quality assurance programs for genomic testing. If our system is used for research (translational genomics), we will comply with DoH's "Guidelines for Clinical & Translational Research in Genomics" which likely involve IRB approvals, periodic reporting, and so forth. Our internal ethics board will work with DoH's ethics committees to ensure all research-use of the platform (e.g. discovering new genotype-phenotype links) has proper oversight.

- **Ethical AI and UAE Artificial Intelligence Principles:** PREPAIRE Labs is committed to ethical AI development. Our Genomics LLM is designed following principles of fairness, transparency, and accountability. We ensure non-discrimination, the model is trained and tested to avoid biases, especially given the diverse genetic backgrounds; for example, it will not under-represent variants common in the Middle Eastern population. We incorporate transparency by providing explainability for the model's decisions, as described earlier, which aligns with emerging UAE and international guidelines on AI ethics. The UAE has a National AI Strategy that emphasizes responsible use of AI; our use of AI in healthcare is carefully governed to avoid any harm. We also abide by any specific AI guidelines from the UAE Ministry of Artificial Intelligence or relevant bodies. For instance, if there are UAE guidelines on AI in healthcare requiring continuous monitoring and human-in-the-loop control, we have those features (e.g. clinicians reviewing AI outputs). Accountability is maintained as well, we have clear human oversight in deployment (the AI assists rather than fully automates diagnoses), and any errors can be traced and corrected, with our team taking responsibility for model updates. In summary, our AI is a tool to augment professionals, and we treat patient safety and ethical considerations as paramount. We will also subject our solution to external audits if requested, and share algorithmic impact assessments with regulators to demonstrate compliance with ethical standards.
- **Medical Device and Health IT Regulations:** If the Genomics LLM platform is deemed a clinical decision support tool or medical device (software) under UAE regulations, we will ensure it meets those regulatory requirements for certification. We are prepared to work with regulatory authorities for classification and approval (for example, getting a UAE Ministry of Health approval or registration for the software if needed). Our development process follows software quality standards (ISO 13485 for medical device software lifecycle, if applicable, and IEC 62304 for software development) to ease this compliance. We are also attentive to data interoperability standards (like HL7 FHIR for any integration with EHRs such as Malaffi) to ensure the system can safely plug into clinical workflows.
- **Data Privacy and Patient Confidentiality:** In addition to Genome Law and NESAs, we adhere to UAE's general data protection regulations (which are converging towards international standards like GDPR). All patient data in our system is de-identified where possible and access is limited on a need-to-know basis. We implement privacy-by-design, meaning features like encryption, audit trails, and consent tracking are built into the software from the ground up. If any cross-border data transfer is required for support purposes, we will obtain explicit permission from DoH and relevant authorities, though our default stance is no data leaves the country. We also commit to storing data only for the necessary retention periods mandated by law or DoH policy, and will support data deletion/anonymization requests.

In essence, PREPAIRE's proposal is not only to deliver a technically advanced solution, but one that is responsible and compliant. We treat regulatory alignment as a first-order requirement, not an afterthought. By doing so, we minimize any legal or ethical risks for the Department of Health and ensure that the Genomics LLM platform can be confidently adopted as part of standard healthcare operations in Abu Dhabi.

Phase 1

Model Development, Integration & Pilot Deployment (Months 1–6)

Objective: Develop the core Genomics LLM system, integrate all components, and validate through a limited-scope pilot.

- **Kickoff & Requirements Finalization (Week 1-2):** Upon award, we will hold a kickoff meeting with DoH stakeholders to review requirements, refine success metrics, and establish governance. Detailed project plans, communication channels, and risk management strategies will be confirmed. We will also ensure all necessary data sharing agreements and ethical clearances (for using de-identified genomic data in development, etc.) are in place at the start.
- **Data Collection and Preparation (Month 1-2):** We will work with DoH and partners like AHD to gather relevant datasets for development. This includes curated variant interpretation datasets (from prior clinical reports, ClinVar entries, etc.), synthetic VCFs for testing, and any available multi-modal patient data for knowledge graph seeding. Data will be cleaned, de-identified, and annotated as needed. Concurrently, we will set up the secure data pipeline from AHD (and any other source) to our development environment in compliance with security protocols.
- **LLM Fine-Tuning and Knowledge Graph Construction (Month 2-3):** Using PREPAIRE's high-performance computing resources, we will fine-tune the base large language model on the genomics corpus. This involves supervised fine-tuning with example VCF interpretations and unsupervised tuning with genomic literature. We will iterate to ensure the model's outputs meet accuracy benchmarks (to be defined, e.g. >90% concordance with expert classifications on test set). In parallel, we will build the initial GenetiQ digital twin knowledge graph schema tailored to this project, defining nodes for patients, genes, variants, phenotypes, etc., and loading initial data (e.g., pathways, known gene-disease links, Emirates-specific genome reference data). Basic NLP and ETL pipelines will be implemented to feed new data into the graph.

- **System Integration & Alpha Prototype (Month 3-4):** We will integrate the LLM with the knowledge graph and the user interface. An alpha version of the GenetiQ platform will be deployed on a secure UAE cloud instance for internal testing. This includes the rules for federated learning (though in Phase 1, we might simulate a single hospital environment), the explainability module, and the uncertainty quantification features. By end of Month 4, we expect an alpha prototype that can take a sample VCF file, run it through the LLM, consult the knowledge graph, and produce a draft report with explanations.
- **Internal Testing and QA (Month 4-5):** The PREPAIRE team will conduct rigorous testing of the alpha system. We will use retrospective cases (if available from AHD or DoH) to see how the AI performs compared to known outcomes. Our clinical expert, Dr. Cronin, will review the AI-generated reports for clinical validity. Security testing will also be done at this stage, ensuring no data leakage, proper encryption functioning, etc. Any bugs or inaccuracies will be addressed in iterative development sprints. We aim to achieve a beta-ready system by the end of Month 5.
- **Pilot Deployment (Month 5-6):** In coordination with American Hospital Dubai (or another DoH-designated pilot site), we will deploy the beta system for real-world pilot testing. This will likely involve installing the system on AHD's premises (or a secure cloud enclave accessible to AHD) and running it on a limited number of cases. During the pilot, we will have a parallel workflow: clinicians will interpret the same cases manually as they normally do, and our LLM will interpret them as well, results will be compared for concordance, and any discrepancies analyzed. We will collect user feedback on the interface, report format, and usefulness of explanations. If possible, we will also pilot the wet-lab validation loop on a small scale (for example, take 1-2 novel findings from the pilot and run a lab test in our Masdar lab to demonstrate the feedback loop process). By the end of Phase 1, we expect to have: a validated LLM model, integrated system, and a successful pilot demonstration with measurable outcomes (e.g., time saved per report, additional variants found or clarified by the AI). We will document all results and lessons learned in a Phase 1 report to DoH, setting the stage for scale-up.

Phase 2

Full-Scale Deployment, Federated Learning Expansion & Continuous Improvement (Months 7–18)

Objective: Deploy the solution at scale across intended sites, implement the federated learning network, and incorporate continuous updates and wet-lab feedback for sustained performance.

Refinement and Production-Readiness (Month 7-9): Based on pilot feedback, we will refine the system. This may include UI/UX improvements, adding new features (e.g., custom report templates for different clinic needs), and further tuning the model. We will also harden the system for production: comprehensive documentation, user manuals, admin dashboards, backup/restore procedures, and integration with hospital IT systems (such as single sign-on with hospital AD accounts, integration with electronic medical record systems to pull patient demographics or push reports). By Month 9, we plan to have Version 1.0 of the product ready for broader release.

- **Federated Infrastructure Deployment (Month 7-12):** In parallel, we will work on the federated learning deployment across multiple sites. This involves setting up the central LAM server (likely hosted at DoH or a central data center) and installing local nodes at each participating hospital (could be AHD, SEHA hospitals in Abu Dhabi, etc. as determined by DoH). Each node will consist of a secured server with the GenetiQ software, knowledge graph instance, and a local copy of the LLM model. We will use containerization (e.g., Docker/Kubernetes) to make deployments reproducible. Networking will be configured with Nokia's assistance to ensure all communications between nodes and central server are via the quantum-safe channels. We will start with a small number of nodes (say 2-3 hospitals) to validate the federated learning process: each hospital will run the system on their data for a few weeks, then we trigger a federated learning round to aggregate model updates. We will monitor the process closely for convergence and any issues. Once validated, we will roll out to more hospitals. By Month 12, the aim is to have the platform deployed and actively used in all intended pilot sites or early adopter facilities in Abu Dhabi, with federated learning keeping the models in sync.
- **Training and Change Management (Month 9-12):** A crucial part of Phase 2 is ensuring end-users (clinical geneticists, lab scientists, physicians) are comfortable and proficient with the new system. We will conduct comprehensive training sessions on-site for each deployment. Training will cover how to input data, how to interpret the AI's report and flags, how to provide feedback/corrections, and how to request wet-lab validation for a case. We will also train IT administrators on managing the system. User feedback mechanisms will be established (hotline, support email, etc.). We'll work to integrate the solution into the clinical workflow smoothly, ensuring that it complements rather than disrupts current processes. This may involve an initial period where the AI's output is "for research use" and double-checked by humans before gradually being trusted as a primary tool. Change management will be guided by our clinical champion (Dr. Cronin) in collaboration with DoH's clinical leads.

- **Monitoring, Evaluation, and Optimization (Month 12-18):** Once the solution is in use, we will closely monitor key performance indicators (KPIs) and gather metrics: e.g., reduction in time to generate a genetic report, number of variants of unknown significance resolved, user satisfaction scores, etc. We will hold periodic review meetings with DoH (monthly or quarterly) to report these metrics. If any component underperforms or new needs arise, we will allocate development sprints to address them (this could include updating the model with new scientific knowledge, improving the variant explanation UI based on user suggestions, etc.). We also plan to implement routine model recalibration – for example, every 6 months retraining the LLM on an expanded dataset including all the new cases it has seen (taking care to do this in a privacy-preserving way, possibly at the central server with federated data). This ensures long-term accuracy and adaptation to any shifts (like new sequencing technologies or new treatment protocols that emerge).
- **Full-Scale Deployment & Handover (Month 16-18):** By the end of Phase 2, our goal is to have a fully deployed, fully functional Genomics LLM platform operational across the Abu Dhabi DoH network (or the intended scope of facilities). This includes a well-tested federated learning infrastructure, a robust AI model, and an active lab validation program. In the final months, we will focus on knowledge transfer and handover. We will ensure that DoH's IT teams and biomedical staff are well-equipped to maintain and continue using the system. Comprehensive documentation (technical docs, SOPs, user guides) will be delivered. We will also establish a support plan for post-deployment (whether it's in-house by PREPAIRE, or training DoH personnel to manage it). A final project report will be prepared, summarizing achievements, KPI outcomes, regulatory compliance confirmation, and suggestions for future improvements or expansions (for example, scaling beyond variant interpretation to drug discovery partnerships, etc.).

Throughout both phases, we will practice agile project management with frequent check-ins with DoH. This ensures transparency and the ability to adapt to any changing requirements or insights. Our iterative approach in Phase 1 reduces risk early by validating the concept in a controlled setting, and Phase 2 ensures a thoughtful scaling that includes user adoption and continuous improvement.

Flexibility: While the above roadmap is comprehensive, PREPAIRE Labs remains flexible to adjust the plan based on the Department's priorities. If DoH desires an accelerated timeline or a different sequencing (for instance, focusing first on a certain disease area's variants), we can accommodate that by re-prioritizing tasks. The phase breakdown is also aligned with typical procurement gateways, Phase 1 could be a proof-of-concept deliverable, and Phase 2 the full production deployment.

We anticipate achieving the project's goals on time and within scope. The end result will be a sustainable, state-of-the-art Genomics LLM service embedded in Abu Dhabi's healthcare system, providing long-term value in improving patient outcomes through precision medicine.

Conclusion and Next Steps



PREPAIRE Labs is excited about the opportunity to partner with the Abu Dhabi Department of Health on this pioneering genomics AI initiative. Our proposal demonstrates a robust plan to deliver a Genomics Foundational LLM for VCF Interpretation that meets the technical requirements and aligns strategically with Abu Dhabi's healthcare innovation goals. With our local presence in Masdar City, our end-to-end capabilities spanning AI to wet-lab, and our commitment to security and compliance, PREPAIRE is uniquely positioned to execute this project successfully. We look forward to the next steps, which would include detailed discussions on requirements, finalizing scope, and commencing the project upon award.

Pricing and Commercial: (As requested, we have kept the pricing/budget section flexible. A detailed commercial proposal can be provided separately.) PREPAIRE Labs is open to a flexible engagement model that can be tailored to DoH's budget and timeline. We propose to discuss and finalize the budget in collaboration with DoH, ensuring cost-effectiveness and value for money for the government.



Thank you for considering our proposal. We are committed to delivering a transformative solution that will position Abu Dhabi at the forefront of genomic medicine and AI.