Contig | Kazakhstan Society of Geneticists and Breeders

EDITORIAL

How to Cite: *Kairov, U. (2026). Bioinformatics in Central Asia at an inflection point: infrastructure, talent, and the case for sustainable core funding. Contig, 2: 202610. DOI: 10.66273/3134-6359.2026.1.2.003*

Received: 05/05/2026 | Accepted: 12/05/2026 | Published online: 12/05/2026

Bioinformatics in Central Asia at an Inflection Point: Infrastructure, Talent, and the Case for Sustainable Core Funding

Ulykbek Kairov[1*]

[1] Laboratory of Bioinformatics & Systems Biology, Center for Life Sciences, National Laboratory Astana, Nazarbayev University, Astana, Kazakhstan

* Corresponding author: ulykbek.kairov@nu.edu.kz

Abstract

Kazakhstan stands at a turning point in the development of bioinformatics. For the first time, the country combines the three prerequisites the discipline demands at scale modern sequencing capacity, substantial high-performance computing, and a rapidly expanding pool of computational talent. National Laboratory Astana at Nazarbayev University has become the region's first and most advanced bioinformatics and genomics centre, producing the first genomic reference resources for the Kazakh population and extending genomic methods into public health. National investments in supercomputing and artificial intelligence, alongside a fast-growing IT education sector, create an unusual opportunity to build an AI-native bioinformatics capability. Yet the field faces serious challenges: an analytical bottleneck, a shortage of cross-disciplinary specialists, immature data governance, and most consequentially a funding model in which the salaries of core specialists and laboratory heads, as well as the renewal of infrastructure, depend almost entirely on competitive grants. This editorial argues that the decisive intervention is stable, recurrent base funding for core bioinformatics infrastructure and personnel, complemented by dedicated training, sound data governance, and regional collaboration, so that recent gains mature into a durable national capability rather than remaining contingent on the next grant cycle.

Keywords: *bioinformatics; Kazakhstan; Central Asia; genomics; research infrastructure; sustainable funding; high-performance computing; workforce development*

Introduction

The life sciences are increasingly written in the language of data. Two decades of falling sequencing costs have turned genome generation into a routine operation, and the limiting step has moved decisively downstream: the constraint is no longer how much data we can produce, but how well we can store, analyse, interpret, and act on it. Bioinformatics the discipline at the intersection of biology, statistics, and computer science has therefore become the rate-limiting step of modern biomedicine, agriculture, and public health.

Central Asia has, until recently, been largely absent from this transformation. Kazakh and other Central Asian populations remain markedly underrepresented in the global reference databases on which precision medicine depends, a gap with real scientific and clinical consequences: variant interpretation tuned to European or East Asian cohorts does not transfer cleanly to a population shaped by millennia of admixture along the Great Silk Road. This absence is at once a scientific deficiency and an opportunity. For the first time, Kazakhstan possesses all three ingredients that bioinformatics requires at scale sequencing capacity, high-performance computing, and a fast-growing pool of computational talent. The question is no longer whether the country can do bioinformatics, but whether it can sustain the capacity it has built.

National Laboratory Astana: the regional flagship

The country’s first and, to date, most advanced bioinformatics and genomics centre is National Laboratory Astana (NLA) at Nazarbayev University. Its Center for Life Sciences brings interdisciplinary, internationally recruited teams together within a single modern research infrastructure that couples automated high-throughput human phenotyping, molecular omics, and dedicated IT systems. Two laboratories anchor the genomics effort the Laboratory of Bioinformatics and Systems Biology and the Laboratory of Genomic and Personalized Medicine supported by one of Kazakhstan’s longest-established and largest ex-situ gene banks.

The genomic platform is equipped to international standards. Sequencing spans both second- and third-generation technologies Illumina NovaSeq 6000 and Oxford Nanopore PromethION 48 backed by a fully equipped molecular biology laboratory and certified specialists drawn from medicine, biology, genetics, and biotechnology. The platform owns its own servers and storage, and analysis is carried out on a dedicated high-performance computing cluster (Q-Symphony; roughly 7.7 TFLOPS, 172 CPU cores, 3 TB RAM, and 186 TB of storage). This co-location of wet-lab, sequencing, and computation under one roof is the operational definition of a working bioinformatics core, and it remains, for now, unique in the country.

The scientific output validates the investment. The laboratory produced the first whole-genome sequences and variant analyses of ethnic Kazakh individuals [1, 2], and, most recently, a reference dataset describing the genomic landscape of healthy Kazakh individuals across the Great Steppe [3] the first endemic genomic reference resources for a population long missing from the global record. The same competencies now reach public health: whole-genome surveillance of drug-resistant tuberculosis, a disease that carries one of the world’s heaviest multidrug-resistant burdens in Kazakhstan, increasingly relies on the bioinformatic pipelines these groups have built [4].

The computing backbone

Kazakhstan has, over a short period, made a serious national commitment to computing and bioinformatics is a direct beneficiary. Nazarbayev University, which has expanded its supercomputing resources since 2018, in 2025 commissioned a next-generation system, Irgetas (“foundation”), reported to offer roughly forty times the capacity of its predecessors [5]. Beyond the campus, an academic supercomputer network on the order of forty petaflops is being assembled across Nazarbayev University, Al-Farabi Kazakh National University, the Eurasian National University, and Satbayev University, alongside Alem.cloud, described as the most powerful cluster in Central Asia [6].

Much of this capacity was justified by national artificial-intelligence priorities training Kazakh-language foundation models and deploying AI tools in education. Yet the same infrastructure large GPU clusters, fast storage, parallel CPUs is exactly what genome assembly, alignment, variant calling, and multi-omics integration demand. For bioinformatics in Kazakhstan, raw compute is, for the first time, no longer the binding constraint.

The talent pipeline

Kazakhstan is also building the human side of the digital economy at speed. Nazarbayev University trains computer scientists and AI specialists through its engineering and digital-sciences faculty and its Institute of Smart Systems and Artificial Intelligence. Astana IT University, through its School of AI and Data Science, offers programmes spanning computer science, big-data analysis, computational sciences, and applied artificial intelligence and data analytics, and in 2026 launched the first MIT IDSS educational partnership in Eurasia [7] all embedded in the wider Astana Hub innovation ecosystem. The country is producing software engineers, data scientists, and AI practitioners in rapidly growing numbers.

The gap is specific. Bioinformatics is an inherently hybrid discipline: it asks for fluency in molecular biology, statistics, software engineering, and domain knowledge simultaneously, and very few graduates arrive equipped in all of these at once. The talent exists in adjacent fields; it is simply not yet being channelled into the life sciences, because almost no dedicated training bridges the computing schools and the genomics laboratories. Closing this bridge is the single highest-leverage investment in human capital the field can make.

Opportunities from the AI–genomics convergence

The global frontier of bioinformatics has become AI-native. Foundation models for DNA and protein sequences, deep-learning systems for variant interpretation and clinical annotation, structural prediction at proteome scale, multi-omics integration, and AI-assisted target and drug discovery are reshaping what a research group can attempt. Because Kazakhstan is investing in artificial intelligence and high-performance computing in parallel, it is positioned to adopt these methods early rather than spending a decade catching up.

The concrete payoffs are regionally specific. Applied to a Central Asian genomic reference, modern AI methods enable population-specific precision medicine and pharmacogenomics calibrated to local variants; pathogen genomics for faster, sharper outbreak response; and agricultural and microbial genomics relevant to the country’s economy and food security. The opportunity to leapfrog to build an AI-first bioinformatics capability rather than retrofitting one is real and time-limited.

Current challenges in bioinformatics

The difficulties are concrete and compound one another. The first is the analysis bottleneck: sequencing throughput has outrun analytical capacity, so data accumulate faster than they can be interpreted into knowledge. The second is workforce: a genuine shortage of trained bioinformaticians, the hybrid-skill gap described above, and persistent competition for that talent from a better-paid domestic software industry and from employers abroad in short, brain drain. The third is data governance: the field still lacks mature, locally adopted standards for findable, accessible, interoperable, and reusable data; curated regional reference databases; reproducible analytical pipelines; and clear ethical and legal frameworks for handling human genomic information.

Beneath all of these sits a fourth problem that quietly determines the others the structure of how the work is funded.

The funding problem: sustaining core infrastructure and people

Today, almost the entire bioinformatics enterprise at National Laboratory Astana both the maintenance of its infrastructure and the salaries of its core specialists, including laboratory heads rests on competitive, time-limited research grants. Scientists are paid from the grants they win; when a grant ends, salary and continuity are immediately at risk. This “soft-money” arrangement is reasonable for funding individual projects, but it is structurally wrong for sustaining national core infrastructure and the expert teams who operate it.

Two features of bioinformatics make this acute. First, the expertise is cumulative and difficult to replace: a senior bioinformatician or laboratory head embodies ten to twenty years of training and institutional memory that cannot be reconstituted on a one-year grant cycle. Lose such a person in the gap between two grants and the capability itself disappears, not merely a salary line. Second, the physical infrastructure sequencers, servers, and computing clusters depreciates quickly and must be refreshed on roughly a five-year cycle to remain scientifically competitive. Project grants rarely fund recurrent salaries, and almost never fund systematic equipment renewal. The result is a paradox: a world-class facility perpetually one funding gap away from losing its people or falling technologically behind.

Recommendations

The remedies are known; what they require is the decision to implement them. They are set out below in order of leverage.

Stable core (base) funding. The decisive intervention is recurrent institutional funding for National Laboratory Astana that covers the salaries of core specialists and laboratory heads together with a planned, roughly five-year infrastructure-renewal budget distinct from, and additional to, competitive project grants. The bioinformatics and genomics core should be treated as national research infrastructure: a shared scientific utility, funded like one. Grants should pay for science; base funding should keep the people and the machines that make the science possible.

A dedicated bioinformatics training track. Joint, cross-disciplinary programmes linking the life-sciences and computing schools of Nazarbayev University, Astana IT University, and the Center for Life Sciences bringing biology to data scientists and computation to biologists supported by fellowships and competitive salaries that retain talent and slow brain drain.

National genomic data infrastructure and governance. A curated Central Asian and Kazakhstani reference-genome resource, FAIR-aligned data standards, reproducible pipelines, and a clear ethical and legal framework for human genomic data.

Channelling national AI compute toward the life sciences. Explicit allocation of a share of the new supercomputing capacity and AI expertise to genomics and biomedicine, so that infrastructure built for language models also advances human health.

Regional and international collaboration. A Central Asian bioinformatics network, paired with continued partnerships with leading institutions abroad, to share methods, standards, datasets, and training and to anchor the region in the global scientific commons.

Conclusions

The trajectory is genuinely promising. The convergence of artificial intelligence and genomics is arriving precisely as Kazakhstan acquires the compute, the early scientific output, and the rapidly expanding IT talent base needed to ride it. Within a decade the country could plausibly become the regional hub for genomic medicine and bioinformatics delivering population-specific precision medicine, pathogen surveillance for public health, and genomics for agriculture and food security across Central Asia.

But neither the technology nor the talent pipeline will, on its own, deliver this future. The binding constraint is institutional: the choice between funding the human and physical core of the country’s bioinformatics capacity in a stable, predictable way, and continuing to rebuild that capacity grant by grant, hoping that key specialists and critical infrastructure survive each interval between awards. The infrastructure is here. The talent is emerging. What remains is the decision to sustain them and that decision, more than any algorithm or instrument, will determine whether Kazakhstan’s bioinformatics moment becomes a lasting scientific capability.

Author Contributions

Conceptualization, writing original draft, and writing review & editing: U.K. The author read and approved the final manuscript.

Acknowledgments

The author thanks colleagues at the Center for Life Sciences and National Laboratory Astana, Nazarbayev University, and the wider Kazakhstani bioinformatics and genomics community for discussions that informed this perspective.

Funding

This editorial received no specific external funding. Research at the author’s laboratory has been supported by competitive grants from the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan. The funders had no role in the preparation or content of this editorial.

Competing Interests

The author has declared that no competing interests exist.

References

1. Kairov U, et al. Whole-genome sequencing data of Kazakh individuals. BMC Research Notes. 2021;14:97. https://doi.org/10.1186/s13104-021-05464-4
2. Kairov U, et al. Whole-Genome Sequencing and Genomic Variant Analysis of Kazakh Individuals. Frontiers in Genetics. 2022;13:902804. https://doi.org/10.3389/fgene.2022.902804
3. Kairov U, et al. Genomic landscape of the Great Steppe: genetic variants in healthy Kazakh individuals. Scientific Data. 2025. https://doi.org/10.1038/s41597-025-05964-z
4. Takenov N, et al. Genetic diversity, drug resistance, and transmission patterns of tuberculosis based on whole-genome sequencing in Almaty, Kazakhstan. Frontiers in Microbiology. 2025;16:1649137. https://doi.org/10.3389/fmicb.2025.1649137
5. Nazarbayev University. How supercomputers and artificial intelligence support researchers at Nazarbayev University (Irgetas supercomputer). 2025. Available from: https://nu.edu.kz/news-en/how-supercomputers-and-artificial-intelligence-support-researchers-at-nazarbayev-university/
6. Ministry of Science and Higher Education of the Republic of Kazakhstan. National academic supercomputer network and Alem.cloud cluster (government briefing). 2026. https://astanatimes.com/2026/04/kazakhstan-expands-supercomputer-and-data-center-to-boost-national-ai-infrastructure/
7. MIT Institute for Data, Systems, and Society. Astana IT University IDSS educational partnership. 2026. Available from: https://www.gov.kz/memleket/entities/maidd/press/news/details/1213780?lang=en
8. National Laboratory Astana. Laboratory of Bioinformatics and Systems Biology. Available from: https://nla.nu.edu.kz/laboratory_bsb

Author Information

Full Name	Ulykbek Kairov
Department	Laboratory of Bioinformatics & Systems Biology, Center for Life Sciences
Institution	National Laboratory Astana, Nazarbayev University, Astana, Kazakhstan
Email	ulykbek.kairov@nu.edu.kz
ORCID	0000-0001-8511-8064
Scopus ID	55549951000