BMC Medical Education

RESEARCH

Open Access

A novel 'Blended Pedagogy' for integrating bioinformatics into the undergraduate medical curriculum: a pilot study on feasibility and student engagement

Ragini D Singh[1], Anupam Berwal[2], Lokvendra S Budania[3], Sagar Dholariya[1*], Ravindra Kumar[4] and Deepak Parchwani[1]

Abstract

Background Bioinformatics has become an integral part of genomic medicine; however, its formal inclusion in undergraduate medical education remains limited. Early, structured exposure is essential to bridge the gap between theoretical molecular biology and its computational applications in clinical contexts. This pilot study evaluated the feasibility and acceptability of integrating introductory, clinically contextualised bioinformatics sessions within the first-year molecular biology module of the MBBS curriculum, focusing on short-term changes in medical students' attitudes, perceived relevance, and engagement with bioinformatics and preparing the future physicians for data-driven clinical practice.

Methods A 10-week molecular biology module with scaffolded, clinically anchored bioinformatics sessions was implemented for first-year MBBS students. Blended Gagné's Nine Events of Instruction and Peyton's Four-Step Approach were used as instruction methods. Clinical case vignettes (e.g., TP53 mutation in Li-Fraumeni Syndrome) provided hands-on training on bioinformatics tools, linking sequence to structural and network analysis (BLAST, CLUSTAL W, PHYRE2, PolyPhen-2, HOPE, and STRING). Students' attitudes and skills were measured using an adapted Colorado Learning Attitudes about Science (CLASS-Bio) Survey with 28 items. Pre- and post-intervention changes in "Motivation and Relevance," "Problem-solving and confidence," and "Epistemological Beliefs" were assessed using paired t-tests with Bonferroni correction and Cohen's d for practical magnitude and clustering analysis to identify engagement patterns (Python 3.12.7).

Results The intervention improved '*Motivation and Relevance*' scores ($p < 0.001$), with gains in interest, perceived relevance, and willingness to pursue further bioinformatics learning. One item (interest in additional bioinformatics coursework) showed statistically significant improvement after Bonferroni correction ($p < 0.001$), while six other items in the same domain showed nominal significance ($p < 0.05$), including enjoyment, improved self-efficacy, and perceived tool utility. No significant change was observed in '*Problem-Solving and Confidence*' ($p = 0.39$) or

*Correspondence:
Sagar Dholariya
drsagar.dholariya@gmail.com

Full list of author information is available at the end of the article

'*Epistemological Beliefs*' domain ($p = 0.90$). Clustering analysis suggested three exploratory engagement profiles: 'enthusiasts', 'moderately positive, and 'cautious learners.'

Conclusions This pilot study supports the feasibility and acceptability of integrating clinically contextualised bioinformatics sessions into first-year molecular biology using a structured blended pedagogy. The module was associated with improved motivation and perceived relevance, while changes in confidence and epistemological beliefs were not observed. Larger multi-institutional studies incorporating objective learning outcomes and comparative designs are warranted to evaluate effectiveness and skill retention.

Highlights

- Integrating bioinformatics early in the medical curriculum enhances student motivation and perceived relevance of molecular biology.
- A blended pedagogy using Gagné's and Peyton's frameworks supports conceptual understanding and hands-on learning.
- Clinically anchored case vignettes (e.g., *TP53-Li-Fraumeni*) effectively link molecular concepts to bioinformatics practice.
- Stepwise exposure to tools from sequence to structure reduces cognitive load and fosters integrative learning.
- Adapting validated attitudinal tools (CLASS-Bio) enables systematic evaluation of interdisciplinary modules.

Keywords Bioinformatics, Medical curriculum, TP53, Pedagogical intervention, Computational tools, Curriculum reform

Introduction

Advancements in genomics and precision medicine have transformed biomedical science and clinical decision-making. Bioinformatics is pivotal to this change, as it aids in interpreting and analyzing the extensive molecular data generated by high-throughput sequencing and multi-omics technologies [1]. Bioinformatics algorithms and tools are now used to identify disease-linked mutations, predict protein structures, and map molecular interaction networks in biological systems. This enables clinicians and researchers to translate molecular findings into personalized diagnostics and targeted therapy [1, 2]. Furthermore, the increasing dependence on computational tools has likely unified and converged the roles and practices of clinical scientists in genetics, molecular pathology, and bioinformatics [3]. Consequently, a solid understanding of computational biology has become an educational imperative for training future physicians in the field. Additionally, for the modern physician, the ability to interpret genomic data is transitioning from a specialist skill to a core clinical competency, necessary for understanding genetic test results, pharmacogenomics, and the principles of precision medicine.

Despite the significant relevance of the discipline in current clinical practice, its integration into undergraduate medical education remains limited [4, 5]. In traditional undergraduate medical education curricula, molecular biology is often taught to first-year medical students, with a focus on basic concepts, including the central dogma, genetic code, and the relationship between protein structure and function. However, these foundational underpinnings fail to contextualize these concepts within the clinical utility and the advancing frameworks of computational biology. This disconnect often creates a skills gap between the acquired theoretical knowledge and its clinical application. Students understand the theoretical concepts of molecular biology but find it challenging to translate this knowledge into the ever-evolving real-world medical landscape, particularly in data-intensive clinical contexts [6–8].

Empirical research posits the importance of early, structured, scaffolded exposure to computational biology through pedagogical strategies that blend foundational biology understanding with practical skill development on diverse tools and data interpretation relevant to current clinical practice [9–11]. Moreover, structured educational interventions that combine authentic case-based learning with hands-on activities are essential to bridge the gap between theoretical knowledge and real-world application [12, 13]. Specifically, employing cognitive and procedural learning models, such as Gagné's Nine Events of Instruction and Peyton's Four-Step Approach, can enhance cognitive apprenticeship, deeper engagement, retention, and transfer of learning to newer clinical contexts [14, 15].

Consequently, in this pilot study, we designed, developed, and integrated introductory bioinformatics sessions (intervention) with the molecular biology module for first-year MBBS students at the All India Institute of Medical Sciences (AIIMS), Rajkot, India. Integration followed the guidelines of the International Society of Computational Biology (ISCB) Education Committee [2]. This committee recommends different sets of bioinformatics core competences tailored to learner's background. In essence, bioinformatics computational tools such as BLAST (nucleotide and protein), CLUSTAL W, PHYRE2,

PolyPhen-2, HOPE, and STRING were embedded within the framework of foundational molecular biology concepts. Clinically relevant case vignettes - such as *TP53* mutation in Li-Fraumeni syndrome [16–19], were specifically designed to guide the bioinformatics sessions' through a blended pedagogical approach i.e., Gagné's Nine Events of Instruction and Peyton's Four-Step. Notably, the primary focus of this evaluation was on implementation feasibility and attitudinal shifts rather than a summative assessment of bioinformatics conceptual knowledge or procedural proficiency. Specifically, our objectives were to: (i) describe the integration of this clinically anchored module within existing frameworks, and (ii) explore short-term changes in students' motivation, perceived relevance, and epistemological beliefs using an adapted 28-item CLASS-Bio survey [20]. Furthermore, by employing exploratory clustering of post-module response patterns, this study seeks to provide a scalable framework for curricular reform. Ultimately, we endeavor to evaluate whether structured, clinically contextualized bioinformatics sessions can enhance conceptual connection and engagement within the undergraduate medical education setting.

Materials and methods

Study design

We employed a single-arm, quasi-experimental pre-post design for this pilot study. The entire molecular biology module with integrated, synchronously in-person bioinformatics sessions (intervention) spanned 10 weeks, commenced in August 2024, in a computer-enabled classroom setting with real-time faculty facilitation and guided hands-on tool use. Bioinformatics comprised eight sessions (90 min each). As a single-group pilot without a comparison cohort, this study was not designed to support causal inferences regarding intervention effectiveness. Instead, the findings are intended to serve as exploratory evidence of implementation feasibility and within-cohort attitudinal shifts.

Study settings

This study was conducted at the All India Institute of Medical Sciences (AIIMS), Rajkot, India, a government-recognised institute of national importance. The intervention was implemented during Phase I of the MBBS program within the Biochemistry curriculum under the Competency-Based Medical Education framework, designed and approved by Standing Academic Committee (SAC) of AIIMS, Rajkot. Molecular biology is taught as a foundational component of biochemistry, and the integration of bioinformatics sessions within this structure reflects authentic first-year curricular implementation.

Participant recruitment

First-year MBBS students ($N = 50$) were the participants of the study. They were expected to have successfully completed a foundational molecular biology module as part of their undergraduate curricula. This ensured their familiarity with basic molecular biology concepts, such as the central dogma, genetic code, mutations, and protein structure-function relationships, and basic computer literacy was assumed.

Ethical considerations

This study was initiated after approval from the Institutional Ethics Committee (IEC: T/D/48/2024). Participation in the study was completely voluntary, with written informed consent. All collected data were anonymized, and personally identifiable information was removed before analysis to protect participant confidentiality. The survey tool responses (CLASS) were securely stored and accessible only to the research team.

Technical implementation

Intervention framework

The bioinformatics sessions were parallelly integrated with molecular biology concepts to create explicit connections between fundamental molecular biology and its computational applications (Fig. 1). Building on students' prior knowledge from their first-year MBBS molecular biology module, the bioinformatics sessions introduced key basic bioinformatics tools following a "*Molecular Biology Concept → Clinically relevant Case → Bioinformatics Tool*" framework. The sessions employed a 'scaffolded approach' to tool integration, ensuring students progressively built required computational skills while reinforcing core molecular biology concepts from sequence → structure → systems biology while balancing clinical relevance and accessibility. The sessions strategically introduced computational tools essential for creating the foundation for the discipline, such as BLAST, Clustal W, Phyre2, HOPE, PolyPhen-2, and STRING (Fig. 1).

The bioinformatics sessions were delivered by a multidisciplinary team from the author group. RDS (PhD, Medical Biochemistry) and SD (MBBS, MD Biochemistry) co-designed the module, developed worksheets, and led tool demonstrations. AB, Coordinator of the Medical Education Unit and Advanced Course in Medical Education (ACME-certified), ensured instructional design alignment and competency mapping within the Competency-Based Medical Education framework. RK (PhD, Bioscience and Engineering) provided specialist input in systems- and network-level bioinformatics interpretation. All sessions were conducted in person using standardized plans aligned with the blended Gagné–Peyton framework (Additional files 2–4). This

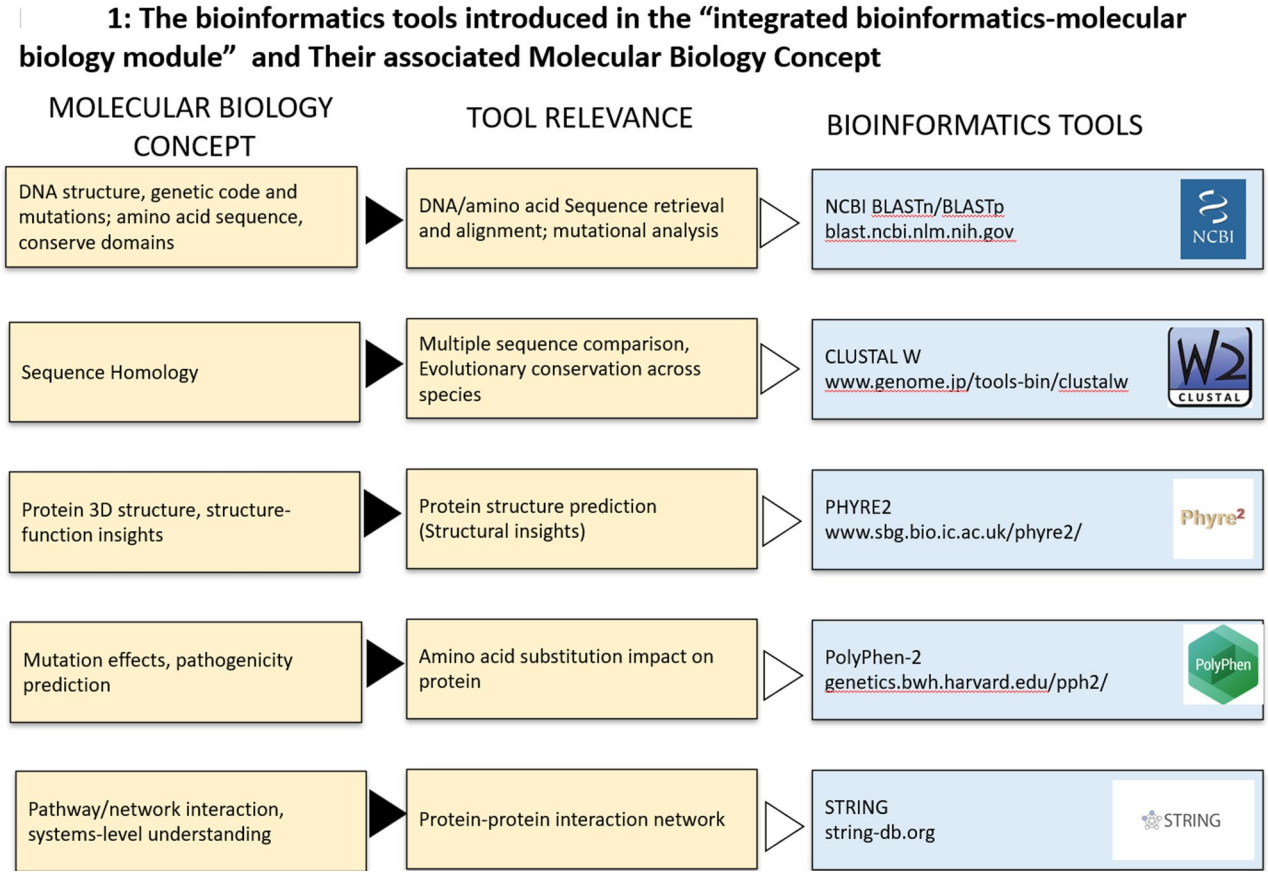

Fig. 1 The bioinformatics tools introduced in the "integrated bioinformatics-molecular biology module" and Their associated Molecular Biology Concept

was the first structured implementation of the module within first-year molecular biology teaching at our institution.

The content progression of each session followed the hybrid framework of Gagné's nine events of instruction to optimize cognitive processing with Peyton's approach for procedural skill acquisition for tool use (Additional file 2,3 and 4). This blended approach was selected specifically to address the dual challenge of introducing a novel *cognitive domain* (computational thinking) and a new *procedural skill* (tool use) to time-pressed medical students, a common scenario in crowded medical curricula.

Case-based integration of bioinformatics with molecular biology: the TP53-Li Fraumeni syndrome learning pathway

Clinically relevant cases were carefully constructed with the 'concept-connect' of both disciplines (Molecular Biology and Bioinformatics) in mind [17–19]. The idea was to develop fundamental practical bioinformatics skills through 'hands-on' experience with the selected tools while reinforcing key molecular biology principles. The clinical case chosen here focused on Li-Fraumeni syndrome, illustrating mutations in the tumor suppressor gene, *p53* (Additional file 2). The TP53–Li-Fraumeni syndrome vignette and accompanying guided worksheet (Additional file 2) were collaboratively developed by the lead author and corresponding author. Case design was informed by established literature on Li-Fraumeni syndrome and *TP53* mutation impact and supported by curated clinical resources (e.g., ClinVar Variation ID: VCV000012366).

The vignette was structured to balance clinical authenticity with accessibility for first-year MBBS students. Draft materials underwent iterative internal review by a multidisciplinary team with expertise in molecular biology, microbiology/genetics, bioinformatics, and medical education. Review focused on scientific accuracy, clarity of instructions, cognitive load appropriateness, and alignment with the intended scaffolded analytical sequence (BLASTn/p → CLUSTAL W → PHYRE2/HOPE → PolyPhen-2 → STRING). Revisions ensured that each analytical step explicitly mapped to a core molecular biology concept and its corresponding computational output. The finalized case vignette and worksheet are provided in Additional file 2 to support transparency and reproducibility.

Students began their learning with NCBI database alignment tools, BLASTn and BLASTp, retrieving and aligning the wild-type and mutant TP53 sequences. The purpose was to identify specific nucleotides and amino acids changes, respectively. Following this, the conservation of this gene and protein across species was analyzed using the CLUSTAL W tool. Subsequently, they used the PHYRE2 tool to visualize the structure of the p53 protein and assessed the potential pathogenicity of TP53 mutations, such as p. Arg273His using the PolyPhen-2 tool and structural impact using HOPE. This helped them link sequence changes in the protein to functional outcomes, deepening their understanding of how mutations might impair protein function and complementing structural predictions from Phyre2 and HOPE. Finally, they explored the broader implications of this alteration using STRING. This tool helped them map how these molecular changes affect the protein's interaction network with its key partners, such as MDM2 and p21, thereby connecting the molecular defect to the cancer predisposition phenotype of the syndrome (Fig. 1).

Learning strategies

A mix of 'active' learning strategies was incorporated into the bioinformatics sessions. The goal was to enhance student engagement, knowledge retention, and application within an introductory 'concept-connect' framework. The strategies were chosen based on the study's objective, student characteristics, their needs, and available resources [21]. The "Guided worksheets" provided novice learners with structured, step-by-step exercises and support that helped them navigate through basic bioinformatics tools and analyses while allowing for independent exploration. This approach was complemented by a "Jigsaw peer-teaching" activity that divided student groups into "experts" on a particular tool and/or case study, who then subsequently taught and shared their findings with their peers. This practice fosters collaborative learning, retention, and enhanced comprehension [22]. Additionally, "Case-based solving" served as the core instructional approach for the sessions, challenging students to apply their 'concept-connect' from both disciplines to address authentic real-world clinical scenarios, such as predicting mutation impacts. The blend of these complementary strategies collectively promoted higher-order thinking skills, with worksheets building technical proficiency, jigsaw activities helping develop communication, synthesis abilities, and collaboration, and case studies bridging theoretical knowledge with clinical applications. The combination of these strategies ensured that students not only mastered bioinformatics tools but also developed critical thinking skills regarding their application in biomedical contexts.

Assessment strategy: survey instrument adaptation -tailoring CLASS-bio for bioinformatics-molecular biology integration

Students' attitudes, motivation, problem solving and confidence and epistemological beliefs regarding the intervention were assessed using an adapted instrument. We adapted the Colorado Learning Attitudes about Science Survey for Biology (CLASS-Bio) following the guidelines by Daunert et al. [23] for the study. The adaptation process aimed to maximize the attainment of semantic, idiomatic, experiential, and conceptual equivalence between the source (CLASS-Bio) and target survey instrument [24], ensuring alignment with the pedagogical goals of our integrated bioinformatics sessions. As this was the first formal introduction to bioinformatics for our cohort, contextual modifications in CLASS-Bio were essential to preserve construct fidelity while enhancing clarity and relevance (Table 1). The theoretical framework of the original CLASS-Bio by Semsar et al. [20] was retained, maintaining its reliable structure across domains such as motivation, problem-solving, and real-world relevance (Table 1).

The adaptation process followed a "theory-driven" methodology. Items were scrutinized for fundamental bioinformatics learning objectives integrated with molecular biology, such as tool use, interdisciplinary connections, and real-world applications of bioinformatics. Terminology was adjusted to accurately reflect bioinformatics-molecular biology 'concept-connect' without altering the cognitive load or intent of the original item. To enhance data integrity, an attention-check item (Q:28; Additional file 1) was embedded in the survey. This dummy question instructed students to select a specific Likert-scale response to confirm their attentiveness.

The adapted instrument comprised 28 Likert-scale items in contrast to 32 Likert-scale items in CLASS-Bio and was administered both pre- and post-bioinformatics sessions (intervention) using a 5-point Likert scale (1 = Strongly Disagree to 5 = Strongly Agree). These items were structured into three major constructs/themes: (i) Motivation and Relevance, (ii) Problem-Solving and Confidence, and (iii) Epistemological Beliefs. This thematic structure was based on the curricular context and the intervention's educational goals. Additionally, 08 post-module exploratory items were included to capture feedback of the students on the integrated bioinformatics sessions.

Validation of the survey instrument

To ensure content validity and contextual alignment of the items in the adapted instrument with the bioinformatics-molecular biology integrated module, a structured validation process was conducted following the guidelines of Lynn [25] and Polit et al. [26]. An

Table 1 Likert Scale analysis of the items included in the adapted CLASS-Bio Survey (*Item-wise Paired t-test Analysis*)

	Items	Pre intervention Mean	Post intervention Mean	*p*-value	Cohen's 'd'	95% CI for Change
Q1.	My curiosity about the living world led me to study at least some molecular biology.	3.66±0.76	4.12±0.53	0.01	0.38	[0.11, 0.81]
Q2.	I think about and make connections between the molecular biology I learn in class with situations in everyday life.	3.37±0.81	3.75±0.82	0.05	0.29	[0.01, 0.75]
Q3.	At this stage, bioinformatics or computationally intensive aspects of molecular biology sometimes feel disconnected or hard to connect.	4.16±0.54	3.97±0.95	0.36	0.12	[-3.03, 1.23]
Q4.	If I had to solve a bioinformatics task today, I would probably struggle to explain what I'm doing or why.	3.67±1.03	3.05±0.98	0.02	0.35	[-1.04, -0.10]
Q5.	I relate molecular biology with bioinformatics to personal experiences	3.54±0.75	3.61±0.86	0.67	0.06	[-0.27, 0.41]
Q6.	I want to study molecular biology with bioinformatics to contribute to society.	3.94±0.77	4.26±0.66	0.04	0.30	[0.02, 0.62]
Q7.	I believe bioinformatics tools may be useful in certain situations.	3.12±0.60	3.51±0.80	0.01	0.38	[0.10, 0.68]
Q8.	I think I would enjoy using bioinformatics to answer questions in molecular biology	3.67±0.75	4.12±0.84	0.01	0.38	[0.11, 0.79]
Q9.	Learning bioinformatics involves software commands and basic computer science whereas for molecular biology I need to memorize facts and definitions	4.00±0.60	4.15±0.65	0.17	0.20	[-0.07, 0.37]
Q10.	The kind of thinking/reasoning skills used in molecular biology with integrated bioinformatics-like analyzing patterns or solving problems-can be useful in real-life situations too.	3.74±0.86	3.68±0.91	0.77	0.04	[-0.46, 0.34]
Q11.	It is a valuable use of my time to study some fundamental concepts in computer science that I may apply to computational analysis of molecular biological data in the future.	3.94±0.81	4.06±0.75	0.44	0.11	[-0.19, 0.43]
Q12.	If I had plenty of time, I would take a computer science class outside of my major requirements just for fun.	3.03±1.01	3.48±0.99	0.01	0.42	[0.15, 0.75]
Q13.	If I had plenty of time, I would take a bioinformatics course outside of my major requirements just for fun.	3.21±0.76	3.68±1.06	0.00	0.52	[0.21, 0.73]
Q14.	The subject of bioinformatics seems unrelated to real world experiences.	3.0±0.95	2.66±1.05	0.09	0.25	[-0.73, 0.05]
Q15.	When solving biology problems, I try to think about them in different ways-especially when new tools or approaches like bioinformatics are involved.	3.44±0.82	3.58±0.74	0.45	0.11	[-0.36, 0.16]
Q16.	If I get stuck on a biology question that involves the use of bioinformatics or computational biology, there is no chance I'll figure it out on my own.	3.00±1.08	2.63±0.86	0.08	0.26	[-0.78, 0.04]
Q17.	There is usually only one correct approach to solving a problem involving bioinformatics and large genomic data sets.	2.53±0.83	2.63±1.08	0.74	0.05	[-0.69, 0.49]
Q18.	When I am not pressed for time, I will employ additional analytical/mathematical/statistical/computational methods on a biology problem until I feel like I have the best answer as to why something works the way it does.	3.56±0.83	3.75±0.87	0.21	0.18	[-0.11, 0.49]
Q19.	Learning computational approaches that are not directly relevant or applicable to human health is not something I would prefer to invest much time in.	2.94±1.15	3.09±1.15	0.38	0.13	[-0.48, 0.18]
Q20.	Mathematical skills are important for understanding molecular biology and bioinformatics.	3.72±0.91	3.47±0.83	0.17	0.20	[-0.61, 0.11]
Q21.	Computer science skills are important for understanding molecular biology and bioinformatics.	3.61±0.89	3.64±0.77	0.80	0.04	[-0.21, 0.27]
Q22.	I enjoy explaining molecular biological ideas or concepts to others.	3.41±0.79	3.72±0.79	0.04	0.30	[0.02, 0.60]
Q23.	The general public misunderstands many biological ideas that require statistics, bioinformatics, or computational biology.	3.84±0.88	3.87±0.93	0.86	0.03	[-0.30, 0.36]
Q24.	Nearly everyone is capable of understanding bioinformatics if they work at it.	3.64±0.92	3.70±1.01	0.74	0.05	[-0.29, 0.41]
Q25.	I do not spend more than a few minutes stuck on a biology question that requires a lot of computation, bioinformatics or statistics before giving up or seeking help from someone else.	3.16±0.95	2.97±1.13	0.48	0.10	[-0.72, 0.34]
Q26.	Molecular Biology and bioinformatic principles are just to be memorized.	2.39±1.12	2.30±0.86	0.62	0.07	[-0.27, 0.45]
Q27.	I will study bioinformatics/biocomputing to learn knowledge that will be useful in my life outside of school.	3.91±1.19	4.03±0.76	0.60	0.07	[-0.34, 0.58]
Q28.	I only care about bioinformatics to understand existing facts-not to use it for my own investigations.	3.38±0.92	3.69±0.89	0.07	0.27	[-0.02, 0.64]

adaptation team was formed that included five subject matter experts from the disciplines of bioinformatics, medical education, and biochemistry and molecular biology. They independently reviewed the items for relevance and clarity using a 4-point ordinal scale (1 = *not relevant* to 4 = *highly relevant*). Content validity was assessed via expert review using the established Content Validity Index (CVI) thresholds [27]. The Item-Level CVI (I-CVI) was calculated as the proportion of experts rating an item ≥ 3. Twenty-five of the 28 items exceeded the accepted threshold (I-CVI ≥ 0.78); three items fell below the threshold and were revised for context, clarity, and simplified for language based on expert feedback. For example, Item 4: "Knowledge in biology consist of many disconnected topics," Item 19: "The subject of bioinformatics has little relation to what I experience in the real world," and Item 5: "When I am answering a biology question, I find it difficult to put what I know into my own words" (Additional file 1).

Given the pilot nature of this study and small sample size (N = 50), psychometric validation (e.g., Cronbach's alpha, factor analysis) was deferred [28, 29]. However, the instrument's theoretical grounding in the validated CLASS-Bio framework and robust content validity (Scale-Level CVI/Average = 0.91) support its preliminary use for exploratory research [27]. This meets the recommended threshold of ≥ 0.90 [27], indicating excellent scale-level content validity.

The adapted instrument was therefore validated both qualitatively through expert review and quantitatively via internal consistency testing. These steps ensured that the instrument was pedagogically sound, contextually appropriate, and statistically reliable for assessing the impact of bioinformatics integration in a first-year medical curriculum in preliminary research.

The surveys paper print copies were distributed before and after the module. Responses were anonymized and matched using unique identifiers to facilitate paired analyses. The instructions emphasized honest responses and were administered in a controlled environment without instructor supervision to minimize bias.

Data collection and analysis

Quantitative data analysis was conducted to assess changes in students attitudes, engagement, and satisfaction following participation in the bioinformatics module. The analysis aimed to (i) identify statistically significant differences between pre- and post-module responses in the adapted CLASS survey, (ii) evaluate broader thematic changes (grouped constructs), and (iii) evaluate exploratory response patterns of students based on post-intervention engagement patterns.

Because this study was conceived as a feasibility- and attitudes-focused pilot within a newly introduced module, no separate standardized knowledge or skills test in bioinformatics was administered for research purposes in this cohort. While routine course examinations included molecular biology content and basic bioinformatics-related components, these assessments were not designed or validated as research instruments and were therefore excluded from the formal analysis. Accordingly, the findings should be interpreted as reflecting shifts in student attitudes and engagement rather than direct evidence of gains in conceptual mastery or procedural tool proficiency.

Descriptive statistics

The mean and standard deviation were computed for all CLASS survey items at both the pre- and post-intervention stages. Paired-sample t-tests were performed to evaluate significant changes between pre- and post-module intervention responses at both the item level (individual survey questions) and at the theme/construct level.

A two-tailed significance level of $p < 0.05$ was considered statistically significant. To account for multiple comparisons across 28 paired t-tests, a Bonferroni correction was applied, adjusting the significance threshold to $p < 0.0018$ ($\alpha = 0.05/28 \approx 0.0018$). Findings were interpreted as strictly significant if they met the corrected threshold and as nominally significant if $p < 0.05$ but > 0.0018. Further, to assess the practical magnitude of the observed changes, Cohen's 'd' for paired samples was calculated for each item and domain. Standardized effect sizes were interpreted as small (0.2), medium (0.5), or large (0.8). Additionally, 95% confidence intervals (CI) for the mean differences were calculated to provide a range of the true effect of the intervention.

All reverse-coded items (e.g., Items 3,4,14,16,17,19,25,26,28) (Table 1) were appropriately recoded (i.e., 6- score) prior to analysis to ensure consistent directionality across constructs. They were transformed prior to aggregation to ensure the scoring direction ability.

Clustering analysis

Post-module feedback included the responses to eight Likert-scale statements (Items: Q29-36) assessing students' experiences with this integrated bioinformatics module. Responses were coded on a 1–5 scale (1 = strongly disagree; 5 = strongly agree). This was analysed using hierarchical clustering (items) and k-means clustering (participants) to explore response similarity and to identify descriptive engagement profiles, respectively. The optimal number of clusters (k) was determined by silhouette analysis [30]. Given the small sample size and ordinal Likert-scale data, these profiles are interpreted as exploratory response patterns rather than stable participant subgroups.

Visualization of the data

A radar plot was used to visualize domain-level pre- and post-intervention changes in student attitudes and beliefs, and a dendrogram and heatmap were constructed to further illustrate score differences in mean item responses across the clusters, facilitating pattern recognition.

Analyses were performed using Python (version 3.12.7) with libraries *Pandas*, *NumPy*, *Matplotlib*, *Seaborn*, *SciPy* and *Scikit-learn*.

Results

All the students enrolled in the study completed both the pre- and post-module surveys. As per the study objectives, results are presented in terms of feasibility, attitudinal change, and engagement patterns derived from survey responses.

Item-wise analysis: pre- and post-intervention comparison

Of the 28 items, eight showed statistically significant improvement ($p < 0.05$) in student responses post-intervention (Table 1). These items primarily belonged to the '*Motivation and Relevance*' domain, suggesting that the module positively influenced students' perceived value and interest in bioinformatics and its integration with molecular biology. Specifically, students reported greater interest in applying bioinformatics to molecular biology problems, increased appreciation of its societal relevance, and higher self-efficacy.

Notably, items such as Item Q8: *"I think I would enjoy using bioinformatics to answer questions in molecular biology"* ($p = 0.01$), Item Q13: *"If I had plenty of time, I would take a bioinformatics course outside of my major requirements… just for fun"* ($p < 0.001$), and Item Q12: *"If I had plenty of time, I would take a computer science class…"* ($p = 0.01$) indicate enhanced enthusiasm for computational learning beyond curricular requirements.

However, after applying the Bonferroni correction for multiple comparisons (adjusted $\alpha \approx 0.0018$), only one item-Q13-remained statistically significant, suggesting a robust increase in voluntary interest in deeper bioinformatics learning (Table 1). While the Bonferroni correction is conservative, several items demonstrated meaningful educational effects. For instance, Item 13 showed a medium effect size (Cohen's 'd': 0.52; 95% CI [0.21, 0.73]), and Item 12 showed an effect size of 0.42. Overall, the 'Motivation and Relevance' domain showed the most consistent improvement with a medium-range effect size of 0.49 (Table-2).

Theme-wise analysis: broader construct evaluation

At the thematic level, paired t-tests revealed a statistically significant improvement in the '*Motivation and Relevance*' domain ($p = 0.00072$), supporting the intervention's impact on increasing student engagement and perceived value of bioinformatics. In this domain, scores increased significantly from a pre-module mean of 3.44 to 3.70 post-module ($p = 0.0007$), reflecting improved student interest, recognition of real-world applications, and a desire to pursue bioinformatics further. This p-value survives Bonferroni correction ($\alpha = 0.05/3 \approx 0.0167$), confirming the robustness of the effect in this domain (Table 2; Fig. 2).

However, no statistically significant changes were found in the Problem-Solving and Confidence domain (Pre:2.91, Post: 2.93; $p = 0.90$), suggesting that while students became more motivated, their perceived ability to independently solve computational problems did not improve substantially over the short intervention (Table 2; Fig. 2).

Epistemological Beliefs also remained statistically unchanged (Pre: 3.25, Post: 3.35; $p = 0.39$), indicating that students' views about the nature of bioinformatics knowledge as interconnected or fragmented were unaffected by the module (Table 2 and Fig. 2). Beyond statistical significance, the practical magnitude of the change was evaluated using Cohen's 'd'. The 'Motivation and Relevance' domain demonstrated a medium effect size (0.49; 95% CI [0.21, 0.77]), suggesting a consistent and substantial shift in student engagement. In contrast, the 'Problem-Solving Confidence' and 'Epistemological Beliefs' domains showed negligible to small effects, reinforcing that these constructs remained stable over the 10-week period (Table 2).

Overall, the observed changes were domain-specific, with improvement primarily in motivation and perceived relevance, while perceived confidence and epistemological beliefs remained stable over the intervention period.

Table 2 The Domain wise trends observed in the items pre vs. post intervention (*Domain-wise Paired t-test Analysis*)

	Domain	Pre intervention Mean	Post intervention Mean	Mean Change	*p*-value	Significant Items ($p < 0.05$)	Cohen's d	95% CI for Change	Interpretation
1.	Motivation and Relevance	3.44	3.70	+0.26	0.00072*	7 out of 14 items	0.49	[0.21, 0.77]	Medium effect
2.	Problem-Solving Confidence	2.91	2.93	+0.02	0.909	1 out of 5 items	0.02	[-0.25, 0.29]	Negligible
3.	Epistemological Beliefs	3.25	3.35	+0.10	0.396	0 out of 9 items	0.12	[-0.15, 0.39]	Small

**Statistically significant after Bonferroni correction*

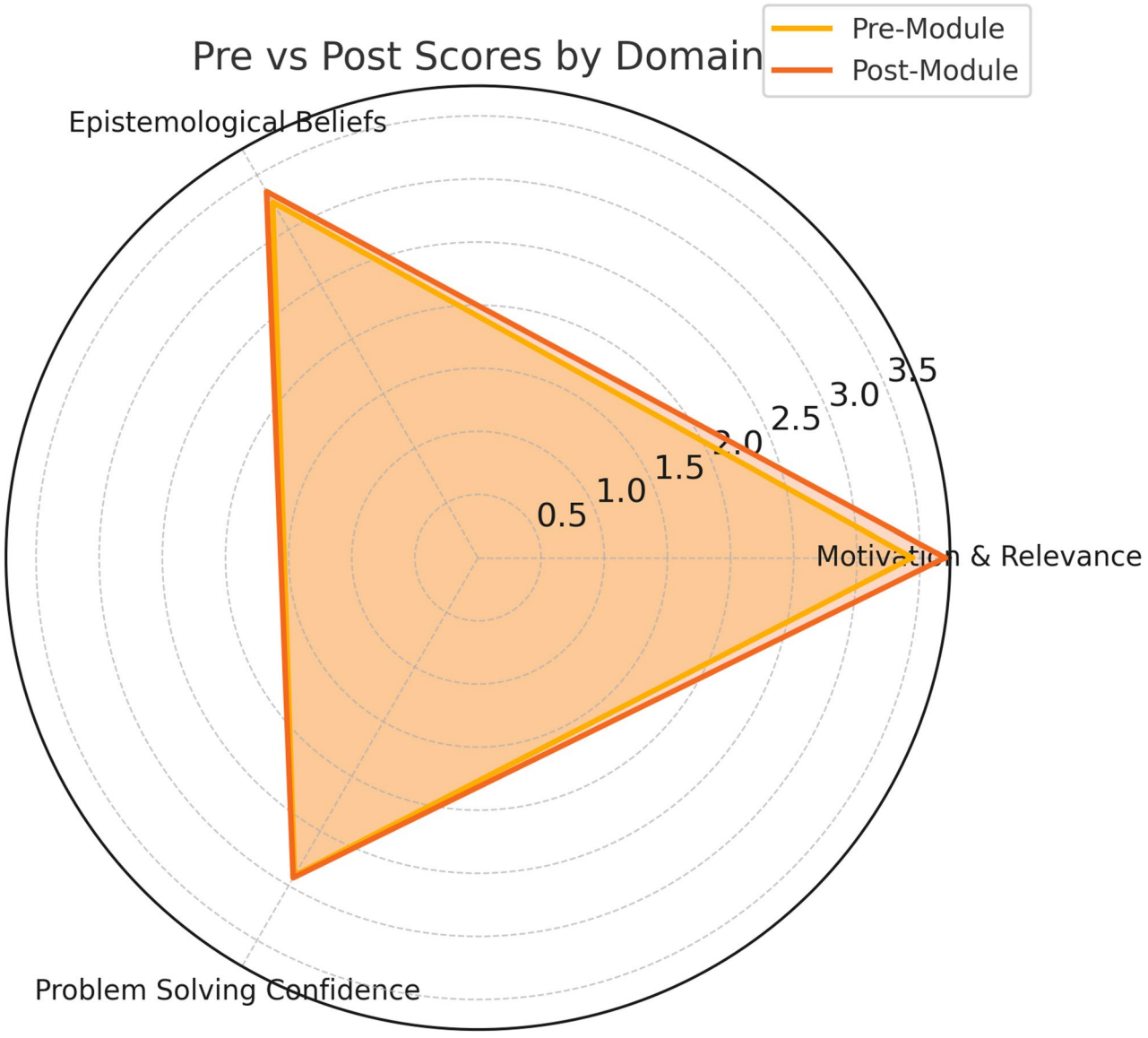

Fig. 2 Domain wise comparison of the pre vs. post-intervention scores in Student Attitudes assessed with *Adapted CLASS Survey*
The radar chart illustrates changes in mean Likert-scale scores (1 = Strongly Disagree to 5 = Strongly Agree) across three key domains: **Motivation & Relevance, Problem Solving Confidence, and Epistemological Beliefs**. Post-module responses (orange) show a visible improvement compared to pre-module responses (yellow), particularly in motivation and problem-solving confidence. Minimal change was observed in epistemological beliefs. The shaded region highlights the overall gain in student attitudes following the instructional module

Post-module student feedback analysis

A descriptive analysis of student responses to post-module items (Q29-Q36)(Table 3) demonstrated overall favorable attitudes toward the structured bioinformatics module. Students reported high levels of curiosity about bioinformatics (Q29: M = 4.05, SD = 0.81) and insights into the role of computer science in biology (Q30: M = 4.03, SD = 0.82), with over 68% rating these items ≥ 4 on the Likert scale (Table 3).

The highest-rated item was enjoyment of computer-based work (Q31: M = 4.25, SD = 0.77), with 80% agreement. Students also expressed strong appreciation for command-line-based learning, although only the differences between command line and web interfaces were given (Q35: M = 4.02, SD = 1.00), and 75% agreed that it was valuable to gain exposure to this modality. Regarding the future relevance of computational tools, 70% of students agreed that such skills would likely be important in their careers (Q36: M = 4.07, SD = 0.82).

Preferences for user-friendly graphical interfaces were evident (Q34: M = 3.57, SD = 0.91), with 62% agreement. Additionally, 65% of students expressed interest in taking a computer science course beyond their curriculum (Q33: M = 3.82, SD = 0.98), indicating that the module

Table 3 Descriptive analysis of the specific post-intervention questions (Feedback Items) (Q29-Q36)

S No.	Item	Item focusing on	Mean ± SD	% Agreement (≥ 4)	Interpretation
Q29	My curiosity about learning more bioinformatics has increased as a result of the module used in class.	Curiosity	4.05 ± 0.81	70%	High curiosity about bioinformatics
Q30	The module provided new insights to me about the uses of computer sciences in molecular biology and bioinformatics.	Insight into computer sciences (CS) in Biology	4.03 ± 0.82	68%	Strong insight gained through the module
Q31	I found the computer work during this module downright enjoyable.	Enjoyment	4.25 ± 0.77	80%	Very high enjoyment of module activities
Q32	Although I don't particularly enjoy computational analyses, I think that this module gave me a good idea of what scientists do with genome-size data sets.	Computational analysis even if not liked	3.87 ± 0.95	68%	Students found it meaningful despite disinterest
Q33	As a result of this module I am now considering taking a (or another) bioinformatics course in the future.	(Interest in bioinformatics course)	3.82 ± 0.98	65%	Moderate future interest in CS coursework
Q34	I like bioinformatics when using tools with user-friendly web-interfaces but do not like command-line style analysis.	Use of various interfaces	3.57 ± 0.91	62%	Preference leans toward user-friendly interfaces
Q35	I might not like command-line style analysis of large data sets but I am glad I got to know about it.	Glad to learn command line	4.02 ± 1.00	75%	Most students appreciated command-line exposure
Q36	Basic knowledge of command-line style bioinformatics/computational biology is likely important for my future career.	Command line's relevance to future career	4.07 ± 0.82	70%	Strong agreement on future relevance

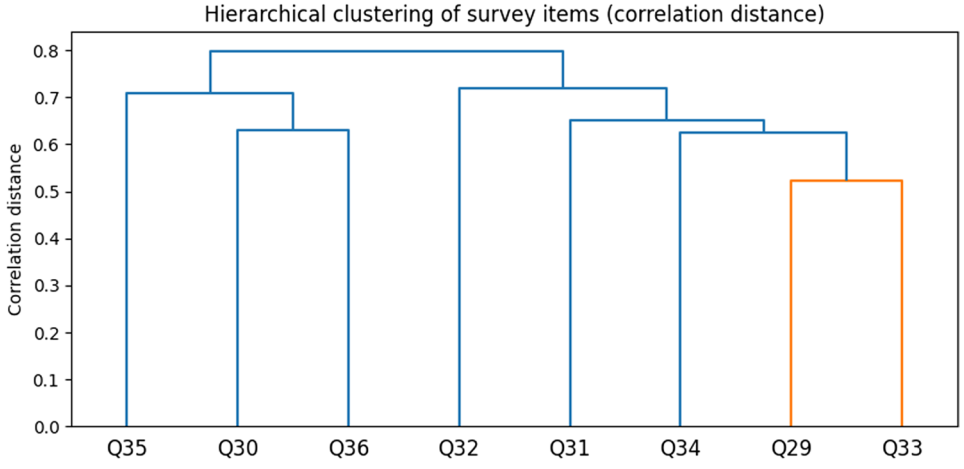

Fig. 3 Hierarchal clustering of the survey items
In a dendrogram, horizontal linkage distance reflects similarity - items joining at low heights are more closely related than those joining high up. The hierarchical tree shows three distinct groupings. Questions 29 (increased curiosity about bioinformatics) and 33 (considering taking another computer science course) join at the shortest distance, indicating similar responses. These items, along with questions 34 (web-based interfaces) and 31 (enjoyment of computer work), form a cluster reflecting affective responses. Questions 30 (insights into computer science in biology), 36 (importance of command-line skills), and 35 (learning command-line analysis) form a second cluster around module utility and command-line competence. Question 32 (understanding genome-scale data) merges only at a higher distance, suggesting independent variation from other themes. The dendrogram indicates that enjoyment and future interest form one dimension, practical utility forms a second, and understanding of scientific practice forms a third

may influence future academic pursuits. Even among those less inclined toward computational analysis, most acknowledged gaining a better understanding of how genome-scale data are used in real-world science (Q32: $M = 3.87$, $SD = 0.95$).

Furthermore, the responses to the eight Likert-scale items were subjected to both hierarchical clustering of items and k-means clustering of participants. "Hierarchical clustering" of items was conducted using correlation distance and average linkage [31]. The dendrogram (Fig. 3) revealed three interpretable item clusters. The first cluster grouped items Q29, Q33, and Q34, which addressed students' curiosity about bioinformatics, willingness to take further computing courses, and preference for web-based interfaces; Q31 (enjoyment of computer work) joined this group at a slightly higher correlation distance. The second cluster grouped items Q30, Q35, and Q36, all related to the perceived utility of the module. Q32 ("the module gave a good idea of what scientists do with genome-size data sets") did not cluster

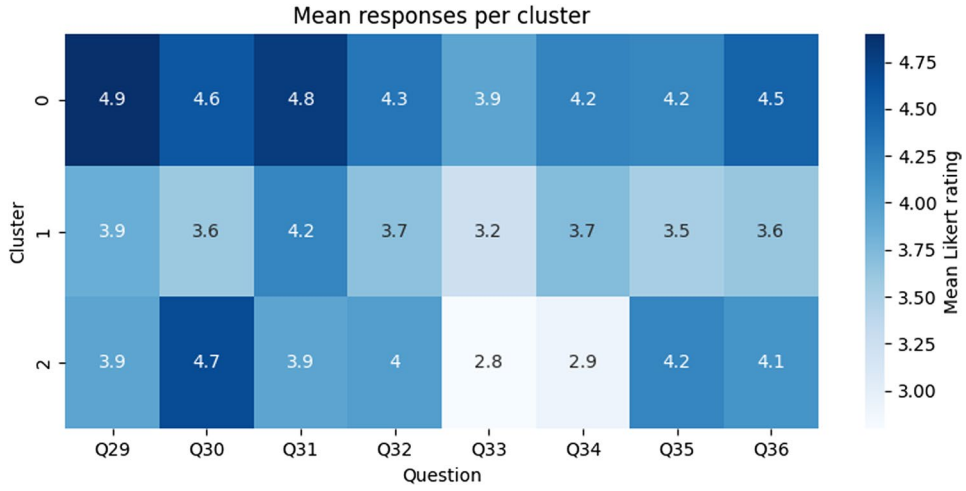

Fig. 4 Exploratory engagement profiles: Mean Likert ratings across three descriptive response patterns
The heatmap displays the mean post-module Likert-scale responses (1 = Strongly Disagree to 5 = Strongly Agree) across three student clusters for eight items (Q29–Q36) evaluating module-specific feedback and tool preferences

Table 4 Descriptive engagement profiles derived from k-means clustering of post-module responses (Q29–Q36)

Engagement Profile (exploratory)	Size	Characteristics of the students in specific clusters based on post intervention response to items Q29-36	Interpretation
Profile 0 - 'Enthusiasts'	12 students	This group gave very high ratings on most items, particularly the curiosity (Q29), new insights (Q30) and enjoyment (Q31) questions (mean ≈4.6–4.9). They rated the command-line-skills questions (Q35, Q36) positively (≈4.2–4.5) but showed lower preference for Q33 (considering another course) and especially Q34 (liking web-interfaces only).	The pattern suggests "enthusiastic students" who enjoyed the module, valued command-line experience and might already have computing backgrounds, so they are less worried about web interfaces.
Profile 1 -'Moderately positive'	23 students	These students rated most items above the midpoint but lower than Cluster 0. Their means range from about 3.5 (for Q34) to 4.2 (for Q31).	They were moderately curious and enjoyed the module, but were less likely to commit to further computer-science courses.
Profile 2 - 'Cautious learners'	25 students	This profile displayed mixed attitudes: high ratings for new insights (Q30) and enjoyment (Q31) but lower scores for Q33 (considering future courses) and Q34 (preference for web-interfaces). They also gave comparatively lower mean responses on Q32 (understanding what scientists do) and Q33 (course interest).	These students may value the information provided but feel less confident or motivated to pursue computational work beyond the module.

closely with any other item, indicating that it captured a distinct dimension of students' experiences.

"K-means clustering" was applied to the matrix of participants by items and three descriptive engagement profiles were observed. Cluster 0 (Profile 1)("*enthusiasts*") contained 12 participants and showed high mean ratings across all items, especially on curiosity (4.9±0.6) and enjoyment (4.8±0.5). Cluster 1 (Profile 2) ("*moderately positive*") included 23 participants who rated the module positively but not as strongly (mean scores ranged from 3.2 to 4.2). Cluster 2 (Profile 3) ("*cautious learners*") comprised 25 participants and displayed mixed attitudes: high scores on the "new insights" item (4.7±0.7) but lower scores on willingness to take further courses (2.8 ± 0.8) and preference for web interfaces (2.9 ± 0.9). These exploratory response patterns suggest a spectrum of engagement, from enthusiastic adopters to generally positive learners to students who value the information but are less motivated to pursue further computational training (Fig. 4; Table 4).

- Profile 0 represents students with consistently high ratings across all questions, indicating strong engagement and appreciation for the module.
- Profile 1 reflects a more neutral group with moderate ratings, showing reserved enthusiasm.
- Profiler 2 shows a polarized response—high interest in module content (e.g., Q30) but lower scores for command-line related items (e.g., Q33, Q34), suggesting a preference for GUI-based tools.

Discussion

Principal findings

This preliminary research addresses a critical gap by integrating first-year MBBS molecular biology with practical bioinformatics applications. The study aims to apply theoretical knowledge to real-world clinical practice.

This study is best interpreted as an implementation- and design-focused pilot evaluating feasibility and early learner engagement, rather than a definitive effectiveness trial. The integration builds on students' understanding of molecular biology concepts while introducing bioinformatic tools that demonstrate these principles. This approach aligns with constructivist learning theories [32] and clinically relevant skills. Findings suggest that integrating introductory bioinformatics within molecular-biology was feasible and improved students' motivation and perceived relevance of the subject. A structured, clinically contextualized approach can stimulate interest in computational thinking among non-programmers, the future physicians.

Design, development and implementation of the integrated module

Integrating bioinformatics education into medical curricula within molecular biology represents a critical step in preparing future healthcare professionals for personalized medicine. The module was designed to emphasize clinical relevance at the undergraduate level. Clinically anchored case vignettes were used to connect molecular biology concepts with bioinformatics applications. This case-based learning approach demonstrates bioinformatics utility, increases student interest, and highlights the relevance of molecular biology. We chose to blend Peyton's approach with Gagné's instructional events based on three core requirements: (a) structured skill sequencing: Gagné's model provides cognitive scaffolding crucial for introducing computational tools to novices, unlike Bloom's taxonomy which only classifies learning objectives [33]; (b) Procedural Mastery: Peyton's approach is preferred for its evidence-based methodology for teaching technical skills, with its four steps directly supporting hands-on tool proficiency [34]; (c) Clinical Training Alignment: Both frameworks are validated in medical education [35, 36], aligning with our goal of integrating molecular theory with clinical practice. Additionally, this framework adapts to different learning styles through its multimodal structure, benefiting visual learners through demonstrations, auditory learners through discussions, and kinaesthetic learners through software practice. The blended methodology addresses both the 'why' and 'how' of bioinformatics, advancing clinical relevance while maintaining adaptability to diverse student needs and fostering critical thinking abilities essential for clinical research settings.

Furthermore, the bioinformatics toolset was selected to be beginner-friendly, web-based, emphasizing DNA to protein pathway continuity, aligned with novice learners, and prioritizing clinical applicability. Tools were curated to create a progressive learning trajectory from sequence analysis to structural investigations, building basic skills while maintaining biological principles. Each tool was chosen for clinical relevance and integration into a cohesiveworkflow.ThetoolsNCBI→BLAST→CLUSTALW→PHYRE-2→HOPE→POLYPHEN-2→STRING were introduced sequentially, reflecting increasing biological complexity. Tool selection followed the 'Central Dogma of Molecular Biology', comparing DNA structure concepts with sequence alignment (BLAST), protein structure prediction (PHYRE-2 and HOPE), interaction networks (STRING) and mutation impact (PolyPhen-2) (Fig. 1). This vertical integration approach reflects scaffolded tool exposure for cognitive load management and conceptual continuity [37, 38].

The clinical case for bioinformatic sessions focused on TP53 mutation, which is part of the undergraduate cancer module and occurs in over 50% of human cancers [39]. It demonstrates the impact of nucleotide changes on protein structure and cellular pathways. Active learning strategies including guided worksheets and jigsaw peer teaching were implemented, following Fredricks et al.,2004's [40] findings on collaborative problem-solving in STEM. Students used BLAST for sequence analysis, PHYRE-2 for structural prediction, and STRING for network modeling. They compared mutant and wild-type TP53 sequences and modeled structural changes, experiencing the mutation-to-phenotype pipeline. This approach aligns with medical education literature supporting clinical case integration for cognitive retention [41, 42].

Assessment approach of the learning: adapted CLASS survey, clustering analysis and differentiated learning

An adapted CLASS survey assessed whether combining the disciplines improved students' ability to apply molecular biology knowledge using bioinformatics tools. The items of the instrument was divided into three constructs. The 'Motivation and Relevance' domain assessed whether students recognize how bioinformatics aids their understanding of molecular biology concepts, making abstract principles more concrete and clinically applicable. The 'Problem-Solving and Confidence' domain evaluated students' ability to use computational thinking for molecular biology challenges, building on prior learning. The 'Epistemological Beliefs' domain examined whether students developed a sophisticated understanding of molecular biology as an interconnected science rather than isolated facts. The study included "Module Feedback" to evaluate how bioinformatics exercises helped students reinforce molecular biology knowledge through practical applications. When students use tools like BLAST, PHYRE2, and PolyPhen-2, they see molecular biology knowledge in action. The survey determined whether this integrated approach helps students understand how computational

tools enhance their comprehension of molecular mechanisms in health and disease.

Analysis of the CLASS survey showed significant improvements in "Motivation and Relevance" domain scores post-module introduction, aligning with Semsar et al. [20]. Item 13 (Table 2), measuring interest in bioinformatics coursework, remained significant after Bonferroni correction. Other items showed nominal improvements in enjoyment, utility, motivation, interest, engagement, and communication confidence (Table 2). While 'problem-solving and confidence' domain scores were unchanged, some items showed nominal improvements, indicating reduced confusion. This suggests short-term interventions may not suffice for developing analytical skills in computational biology [43]. The unchanged "Epistemological Beliefs" scores indicate that shifts in understanding biology as a data-driven field may require prolonged exposure rather than single-module interventions. The absence of change in epistemological beliefs and perceived confidence is consistent with the expectation that these constructs require explicit instructional targeting and sustained longitudinal reinforcement. Interestingly, even though many individual item-level p-values did not survive the conservative Bonferroni correction, several items (e.g., Q12 and Q13) exhibited medium effect sizes ($d > 0.4$). This suggests that the clinically integrated module had a tangible impact on students' perceived value of bioinformatics in medicine, even if the sample size was not powered to show significance across every individual survey item. This finding aligns with the goal of the 'blended pedagogy' to bridge the gap between theoretical molecular biology and clinical practice through case-based learning.

Post-module feedback revealed effective reinforcement of molecular biology concepts through bioinformatics exercises. High scores in curiosity, enjoyment, and perceived relevance demonstrated student engagement, with particularly strong enthusiasm for computer-based work ($M = 4.25$) in molecular biology.

Additionally, the blended dual pedagogical approach in the study effectively bridged cognitive and procedural domains, guiding students from theoretical understanding to hands-on on computational tools. This aligns with findings that structural frameworks improve biomedical education outcomes [44]. The module's stepwise progression, combining concept reinforcement with bioinformatic tools, ensured cognitive alignment while encouraging deeper engagement [45]. Students built confidence gradually, as shown by improved practical assessments. Peyton's four-step method helped students use computational tools independently, transforming passive observers into active users. This supports Peyton's effectiveness in teaching procedural skills [46].

Hierarchical clustering of the survey items was performed, and the dendrogram analysis (Fig. 3) indicates that the module may simultaneously affect students' enthusiasm, perceived practical value, and comprehension of research practices. The clustering of students' based on post-module responses for the intervention highlighted a spectrum of engagement. One profile ("enthusiasts") rated most items highly, suggesting that they were both motivated and open to learning bioinformatics and computational skills. A second profile ("moderately positive") showed generally favorable but less pronounced attitudes, implying that they appreciated the module but might not actively pursue further training. For this group, integrating bioinformatics exercises directly into subsequent clinical case discussions (e.g., in Pathology or Pharmacology) could reinforce utility without increasing cognitive load, potentially moving them toward a more positive trajectory. The third profile ("cautious learners") valued the insights provided but were less inclined to take additional computing courses. Recognizing these exploratory engagement profiles could guide differentiated instructional strategies, for instance, offering advanced projects to enthusiasts, integrating bioinformatics more explicitly into the curriculum for moderately positive students, and providing additional support or confidence-building- activities for cautious learners. These profiles are descriptive and hypothesis-generating and should be validated in larger cohorts before being interpreted as stable learner typologies.

Implications for curriculum reform and policy

Our study indicates that a well-structured, 10-session module is sufficient to shift student attitudes and demonstrate feasibility of its integration and implementation in the existing molecular biology of the undergraduate curriculum. This can serve as a template for integration and make this upcoming generation of physicians future ready. We suggest that such a module be formally incorporated within the foundational molecular biology curriculum of the MBBS program by national medical regulatory bodies. The use of free, web-based tools and a clearly outlined pedagogical framework makes this model highly scalable and cost-effective, addressing common obstacles to curriculum change.

Limitations of the study and future directions

While this pilot study serves as a necessary and rigorous first step to establish feasibility and proof-of-concept, several limitations must be acknowledged. First, the single-center design and sample size limit the generalizability of our findings. Second, the single-group pre–post design without a comparison group limits causal inference and restricts interpretation to within-cohort changes. Furthermore, while the adapted CLASS-Bio

survey effectively captured shifts in student attitudes, the study did not include a standardized pre-post test of bioinformatics concepts or objective performance measures, such as rubric-based tool competency assessments. Consequently, the findings should be interpreted as reflecting shifts in student engagement and perceptions rather than direct evidence of gains in conceptual mastery or procedural proficiency. Additionally, given the pilot nature of the study, the focus was placed on content validity and feasibility rather than comprehensive psychometric validation of the instrument. While several safeguards were employed—including expert review and high item-level content validity indices—statistical reliability metrics such as Cronbach's alpha were not calculated for this cohort.

Future directions include planning longitudinal studies to assess the retention and clinical application of bioinformatics skills, as well as broader multi-institutional implementations to confirm the construct validity of the instrument and the generalizability of the pedagogical model. Future work shall also incorporate more rigorous comparative designs, such as controlled cohort comparisons or stepped-wedge implementations, alongside validated skill-based assessments aligned with ISCB core competencies. Finally, the k-means clustering was exploratory and descriptive; therefore, the identified engagement profiles should not be interpreted as definitive participant subgroups and require replication in larger samples.

Conclusion

By integrating bioinformatics practice with molecular biology theory, this study addresses a crucial need in medical education: helping students transition from passive learners of molecular biological concepts to active investigators who can harness computational tools to solve biomedical problems. This study provides empirical support for incorporating structured and clinically integrated bioinformatics training into preclinical medical curricula. By leveraging pedagogical frameworks (Gagné and Peyton), clinically relevant case studies (e.g., TP53), and carefully curated tool progression (NCBI→BLAST→CLUSTAL W→PHYRE2 → HOPE→POLYPHEN-2→STRING), the module significantly enhanced student motivation, perceived relevance, conceptual engagement and confidence to engage with bioinformatics tools. Tailoring future modules to accommodate diverse learner needs and incorporating longitudinal assessments will be key to sustaining and scaling such innovations in medical education.

Abbreviations

BLAST	Basic Local Alignment Search Tool
CBME	Competency-Based Medical Education
CI	Confidence Interval
CLASS-Bio	Colorado Learning Attitudes about Science Survey for Biology
CLUSTAL W	CLUSTAL W multiple sequence alignment program
CVI	Content Validity Index
HOPE	Have (y)Our Protein Explained
I-CVI	Item-level Content Validity Index
ISCB	International Society of Computational Biology
LFS	Li-Fraumeni syndrome
NCBI	National Center for Biotechnology Information
PHYRE2	Protein Homology/analogy Recognition Engine 2
PolyPhen-2	Polymorphism Phenotyping v2
S-CVI/Ave	Scale-level Content Validity Index (average)
SD	Standard Deviation
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
TP53	Tumor Protein 53 gene

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12909-026-09116-8.

Supplementary Material 1.

Supplementary Material 2.

Supplementary Material 3.

Supplementary Material 4.

Supplementary Material 5.

Acknowledgements

The authors acknowledge the institutional ethics committee for the permission to carry out this work.

Declaration of generative AI and AI-assisted technologies in the writing process

Statement: During the preparation of this work the author(s) used Paperpal in the writing process to improve the readability and language of the manuscript. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

Author contributions

Conceptualization: RDS (lead), AB and SD (supporting)Data curation: RDS (lead), SD (supporting)Formal analysis: RDS (lead), RK (supporting) Methodology: RDS (lead), SD (supporting), RK (supporting), AB (supporting) Project administration: RDS (lead), SD and DP (supporting)Resources: RDS (lead), SD (supporting), RK (supporting), LSB (supporting), Supervision: RDS (equal), SD (equal)Validation: RDS, RK, SD, DP and AB (all equal)Writing 6 original draft: RDS (lead), SD (supporting), LSB (supporting), AB (supporting) Writing 6 review & editing: RDS (equal), LBS (equal), AB (equal), RK (supporting), SD (supporting), DP (supporting).

Funding

No funding was received for this study.

Data availability

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

This study was initiated after approval from the Institutional Ethics Committee (IEC: T/D/48/2024) and adhered to Declaration of Helsinki and Good Clinical Practice guidelines. Participation in the study was completely voluntary, with written informed consent. The institute's ethics committee approved this study and the committee's reference number is ECR/1682/inst/GJ/2022. This study was approved by the Institutional Ethics Committee (Approved study project number: T/D/48/2024; IEC reference number: ECR/1682/inst/GJ/2022) and adhered to the Declaration of Helsinki and Good Clinical

Practice guidelines. Participation was voluntary, and written informed consent was obtained from all participants.

Consent for publication
Not Applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Biochemistry, All India Institute of Medical Sciences (AIIMS), Rajkot, India
[2]Department of Microbiology, Kalpana Chawla Government Medical College (KCGMC), Karnal, India
[3]Department of Anaesthesia, Kalpana Chawla Government Medical College (KCGMC), Karnal, India
[4]Department of Bioscience and Engineering, National Institute of Technology, Calicut, India

Received: 18 November 2025 / Accepted: 27 March 2026
Published online: 07 April 2026

References

1. Wolde T, Bhardwaj V, Pandey V. Current bioinformatics tools in precision oncology. MedComm (2020). 2025;6(7):e70243. https://doi.org/10.1002/mco2.70243.
2. Mulder N, Schwartz R, Brazas MD, Brooksbank C, Gaeta B, Morgan SL, et al. The development and application of bioinformatics core competencies to improve bioinformatics training and education. PLoS Comput Biol. 2018;14(2):e1005772. https://doi.org/10.1371/journal.pcbi.1005772.
3. Genomics Education Programme. *Clinical Bioinformatics Training*. Health Education England; 2019. Available from: https://www.genomicseducation.hee.nhs.uk/wp-content/uploads/2019/06/Clinical_Bioinformatics_Training.pdf. Cited 7 Aug 2025.
4. Madlung A. Assessing students' conceptual understanding and attitudes in introductory bioinformatics. PLoS Comput Biol. 2018;14(1):e1005872. https://doi.org/10.1371/journal.pcbi.1005872.
5. Tambi R, Bayoumi R, Lansberg M, Banerjee Y. Blended learning in bioinformatics: an insight into the student experience. JMIR Med Educ. 2018;4(2):e11122. https://doi.org/10.2196/11122.
6. BJ S, Parashar S, Saini R, Sharma R, Sharma D, Sharma D, et al. Electives in the medical curriculum as applicable to biochemistry: impacts and insights of medical students. BMC Med Educ. 2025;25:948. https://doi.org/10.1186/s12909-025-07226-3.
7. Hack C, Kendall G. Bioinformatics: student understanding and attitudes toward a new interdisciplinary field. Biochem Mol Biol Educ. 2005;33(2):82–5. https://doi.org/10.1002/bmb.2005.49403302082.
8. Yang C, Chen Y, Qian C, Shi F, Guo Y. The data-intensive research paradigm: challenges and responses in clinical professional graduate education. Front Med (Lausanne). 2025;12:1461863. https://doi.org/10.3389/fmed.2025.1461863.
9. Magana AJ, Arigye J, Udosen A, et al. Scaffolded team-based computational modeling and simulation projects for promoting representational competence and regulatory skills. Int J STEM Educ. 2024;11:34. https://doi.org/10.1186/s40594-024-00494-3.
10. Masava B, Nyoni CN, Botma Y. Scaffolding in health sciences education programmes: an integrative review. Med Sci Educ. 2022;33(1):255–73. https://doi.org/10.1007/s40670-022-01691-x.
11. Pevzner PA, Shamir R, Sharan R. Teaching computational thinking through bioinformatics to biology students. In: Proceedings of the 40th ACM Technical Symposium on Computer Science Education (SIGCSE '09). New York: ACM; 2009:188–191. https://doi.org/10.1145/1508865.1508932.
12. Varma B, Karuveettil V, Fernandez R, Halcomb E, Rolls K, Kumar SV, Aravind MS. Effectiveness of case-based learning on competencies and satisfaction among healthcare students: a systematic review. J Educ Health Promot. 2025;14:76. https://doi.org/10.4103/jehp.jehp_510_24.
13. Thistlethwaite JE, Davies D, Ekeocha S, Kidd JM, MacDougall C, Matthews P, et al. The effectiveness of case-based learning in health professional education: a BEME systematic review. Med Teach. 2012;34(6):e421–44. https://doi.org/10.3109/0142159X.2012.680939.
14. Al-Eraky MM. AM last page: Robert Gagné's nine events of instruction, revisited. Acad Med. 2012;87(5):677. https://doi.org/10.1097/ACM.0b013e318250e01d.
15. Walker M, Peyton JWR. Teaching in theatre. In: Peyton JWR, editor. Teaching and Learning in Medical Practice. Rickmansworth: Manticore Europe Limited; 1998. pp. 171–80.
16. National Center for Biotechnology Information (NCBI). ClinVar; TP53 variant p.Arg273His (VCV000012366) . Accessed 30 Mar 2026 https://www.ncbi.nlm.nih.gov/clinp53var/variation/VCV000012366
17. McLean SF. Case-based learning and its application in medical and health-care fields: a review of worldwide literature. J Med Educ Curric Dev. 2016;3:S20377.
18. Bougeard G, Renaux-Petel M, Flaman JM, Charbonnier C, Fermey P, Belotti M, et al. Revisiting Li-Fraumeni syndrome from TP53 mutation carriers. J Clin Oncol. 2015;33(21):2345–52. https://doi.org/10.1200/JCO.2014.59.5728.
19. Malkin D. Li-Fraumeni syndrome. Genes Cancer. 2011;2(4):475–84. https://doi.org/10.1177/1947601911413466.
20. Semsar K, Knight JK, Birol G, Smith MK. The Colorado Learning Attitudes about Science Survey (CLASS) for use in biology. CBE Life Sci Educ. 2011;10(3):268–78. https://doi.org/10.1187/cbe.10-10-0133.
21. Freeman S, Eddy SL, McDonough M, Smith MK, Okoroafor N, Jordt H, et al. Active learning increases student performance in science, engineering, and mathematics. Proc Natl Acad Sci U S A. 2014;111(23):8410–5.
22. Perkins DV, Tagler MJ. Jigsaw classroom. In: Miller RL, Amsel E, Marsteller-Kowalewski B, Beins BC, Keith KD, Peden BF, editors. Promoting Student Engagement – Volume 1: Programs, Techniques and Opportunities. Society for the Teaching of Psychology; 2011. pp. 195–7.
23. Daunert AL, Seel NM. Translating and adapting survey instruments: important steps that will define quality and optimize validity. Educ Psychol Rev. 2020;32:843–59. https://doi.org/10.1007/s10648-020-09518-1.
24. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine. 2000;25(24):3186–91. https://doi.org/10.1097/00007632-200012150-00014.
25. Lynn MR. Determination and quantification of content validity. Nurs Res. 1986;35(6):382–5. https://doi.org/10.1097/00006199-198611000-00017.
26. Polit DF, Beck CT, Owen SV. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. Res Nurs Health. 2007;30(4):459–67. https://doi.org/10.1002/nur.20199.
27. Polit DF, Beck CT. The content validity index: are you sure you know what's being reported? Res Nurs Health. 2006;29(5):489–97. https://doi.org/10.1002/nur.20147.
28. Thabane L, Ma J, Chu R, Cheng J, Ismaila A, Rios LP, et al. A tutorial on pilot studies: the what, why, and how. BMC Med Res Methodol. 2010;10:1. https://doi.org/10.1186/1471-2288-10-1.
29. Dunn TJ, Baguley T, Brunsden V. From alpha to omega: a practical solution to the pervasive problem of internal consistency estimation. Br J Psychol. 2014;105(3):399–412. https://doi.org/10.1111/bjop.12046.
30. Scikit-learn Developers. Scikit-learn: machine learning in Python [Internet]. Available from: https://scikit-learn.org/stable/. Accessed 30 Mar 2026. https://scikit-learn.org/stable/
31. Murtagh F, Contreras P. Algorithms for hierarchical clustering: an overview. Wiley Interdiscip Rev Data Min Knowl Discov. 2012;2(1):86–97.
32. Bruner JS. Toward a Theory of Instruction. Cambridge: Harvard University Press; 1966.
33. Anderson LW, Krathwohl DR. A Taxonomy for Learning, Teaching and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives. New York: Longman; 2001.
34. Kolb DA. Experiential Learning: Experience as the Source of Learning and Development. Englewood Cliffs, NJ: Prentice Hall; 1984.
35. Giacomino K, Caliesch R, Sattelmayer KM. The effectiveness of the Peyton's four-step teaching approach on skill acquisition of procedures in health professions education: a systematic review and meta-analysis. PeerJ. 2020;8:e10129. https://doi.org/10.7717/peerj.10129.
36. Li Y, Liang Z, Li Z, Yu Y, Yang Q, Li X. Effectiveness of Gagné's nine events of instruction in health professions education: a systematic review and meta-analysis. Front Med (Lausanne). 2025;12:1522830. https://doi.org/10.3389/fmed.2025.1522830.
37. Welch L, Lewitter F, Schwartz R, Brooksbank C, Radivojac P, Gaeta B, Schneider MV. Bioinformatics curriculum guidelines: toward a definition of core competencies. PLoS Comput Biol. 2014;10(3):e1003496. https://doi.org/10.1371/journal.pcbi.1003496.

38. Via A, Blicher T, Bongcam-Rudloff E, et al. Best practices in bioinformatics training for life scientists. Brief Bioinform. 2022;23(1):bbab456. https://doi.org/10.1093/bib/bbab456.
39. Olivier M, Hollstein M, Hainaut P. TP53 mutations in human cancers: origins, consequences, and clinical use. Cold Spring Harb Perspect Biol. 2010;2(1):a001008. https://doi.org/10.1101/cshperspect.a001008.
40. Fredricks JA, Blumenfeld PC, Paris AH. School engagement: potential of the concept, state of the evidence. Rev Educ Res. 2004;74(1):59–109. https://doi.org/10.3102/00346543074001059.
41. Thistlethwaite JE, Davies D, Ekeocha S, Kidd JM, MacDougall C, Matthews P, et al. The effectiveness of case-based learning in health professional education. Med Teach. 2012;34(6):e421–44. https://doi.org/10.3109/0142159X.2012.680939.
42. Wood DF. Problem-based learning. BMJ. 2003;326(7384):328–30. https://doi.org/10.1136/bmj.326.7384.328.
43. Madlung A, Bremer M. A study assessing the effectiveness of an online bioinformatics tool in teaching basic concepts of gene expression. CBE Life Sci Educ. 2011;10(4):456–67. https://doi.org/10.1187/cbe.10-07-0089.
44. Brownell SE, Tanner KD. Barriers to faculty pedagogical change: lack of training, time, incentives, and tensions with professional identity. CBE Life Sci Educ. 2012;11(4):339–46. https://doi.org/10.1187/cbe.12-03-0030.
45. Leppink J, van den Heuvel A. The evolution of cognitive load theory and its application to medical education. Perspect Med Educ. 2015;4(3):119–27. https://doi.org/10.1007/s40037-015-0192-2.
46. Nikendei C, Huber J, Stiepak J, Huhn D, Lauter J, Herzog W, Jünger J, Krautter M. Modification of Peyton's four-step approach for small group teaching: a descriptive study. BMC Med Educ. 2014;14:68. https://doi.org/10.1186/1472-6920-14-68.

Publisher's Note