OXFORD

Briefings in Bioinformatics, 2024, **25(5)**, bbae447

https://doi.org/10.1093/bib/bbae447

Review

Clinical bioinformatics desiderata for molecular tumor boards

Matteo Pallocca[1,*], Martina Betti[2], Sara Baldinelli[2], Ramona Palombo[1], Gabriele Bucci[3], Luca Mazzarella[4,5], Giovanni Tonon[6], Gennaro Ciliberto[7]

[1]Institute of Experimental Endocrinology and Oncology, National Research Council, Via Sergio Pansini, 5, 80131 Naples, Italy
[2]Biostatistics, Bioinformatics and Clinical Trial Center, IRCCS Regina Elena National Cancer Institute, Via Elio Chianesi, 53, 00144 Rome, Italy
[3]Center for OMICS Sciences, IRCCS San Raffaele Scientific Institute, Via Olgettina, 60, 20132 Milan, Italy
[4]Division of Gastrointestinal Medical Oncology and Neuroendocrine Tumors, IRCCS IEO - European Institute of Oncology, Via Adamello 16, 20139 Milan, Italy
[5]Department of Experimental Oncology, IRCCS IEO - European Institute of Oncology, Via Adamello 16, 20139 Milan, Italy
[6]Functional Genomics of Cancer Unit, Division of Experimental Oncology, and Center for Omics Sciences, IRCCS San Raffaele Scientific Institute, Via Olgettina, 60, 20132 Milan, Italy
[7]Scientific Direction, IRCCS Regina Elena National Cancer Institute, Via Elio Chianesi, 53, 00144 Rome, Italy

*Corresponding author. Institute of Experimental Endocrinology and Oncology, National Research Council, Via Sergio Pansini, 5, 80131 Naples, Italy.
E-mail: matteo.pallocca@cnr.it

Abstract

Clinical Bioinformatics is a knowledge framework required to interpret data of medical interest via computational methods. This area became of dramatic importance in precision oncology, fueled by cancer genomic profiling: most definitions of Molecular Tumor Boards require the presence of bioinformaticians. However, all available literature remained rather vague on what are the specific needs in terms of digital tools and expertise to tackle and interpret genomics data to assign novel targeted or biomarker-driven targeted therapies to cancer patients. To fill this gap, in this article, we present a catalog of software families and human skills required for the tumor board bioinformatician, with specific examples of real-world applications associated with each element presented.

Keywords: clinical bioinformatics; clinical informatics; molecular tumor board; decision support tools; molecular tumor registry; variant annotation; variant actionability; biobanking; data engineering

Introduction

Bioinformatics is a mandatory component of Molecular Tumor Boards (MTB) [1, 2]. MTBs have been defined in multiple ways, with a common ground being multidisciplinary teams able to discuss patient data, diagnosis, and results to recommend targeted therapies according to molecular alterations detected via -omics profiling, usually DNA or RNA sequencing [3, 4]. Specifically, employing automatized methods to annotate molecular alterations provides additional strength to any kind of -omics profiling. The knowledge of Clinical Bioinformatics empowers molecular pathologists to define tumor drug sensitivity and enables oncologists to consider additional lines of therapies based on patient history and molecular background [5].

As bioinformaticians became novel pillars of molecular research laboratories since the advent of massive-parallel sequencing, not the same happened a few years later with the introduction of routine cancer sequencing. The main reason for this asynchrony lies in the end-to-end engineering of formalin-fixed paraffin-embedded (FFPE) tumor sequencing solutions that enable pathology departments to perform all the multistep processes without deep computational knowledge (Fig. 1) [6, 7]. At the beginning of the precision oncology era, the number of actionable alterations was scarce, and the discrete combinations of sensibility or resistance profiles were easily human-intelligible without machine-aided annotations. Furthermore, the paucity of molecularly aided clinical algorithms led pathology departments to employ conditional qPCR instead of next-generation sequencing to test small indels (insertion/deletion) and mutations, as for KRAS, BRAF, and EGFR [8–10].

Fortunately, in recent years, the rate of clinical-grade approvals for immuno- and targeted therapy increased at a constant pace, due to the identification of many genetic aberrations and promising drug targets [11]. With the clinical and preclinical evidence increasing, the average genomic sampling carried out in routinary screening and target sequencing increased from up to hundreds of kilobases targets, accompanied by a systematic integration with other types of omics data. The application of full coding sequencing increased the data pipeline complexity and the probability of somatic and germline incidental findings. This novel scenario led on one hand to a series of novel treatment combinations for patients, while simultaneously posing strong challenges to hospital staff regarding variant analysis and reporting. The concept of variant interpretation itself is not trivial; on the contrary, the plethora of possible interpretations requires mastering the skill of effective integration and prioritization of variants [12, 13].

Interestingly, this molecular data expansion renders the latter a source of significant value comparable to clinical trial data. Such a phenomenon demands another revolution, the one associated with clinical data digitalization and the drive towards data

Received: December 18, 2023. **Revised:** June 28, 2024. **Accepted:** August 30, 2024

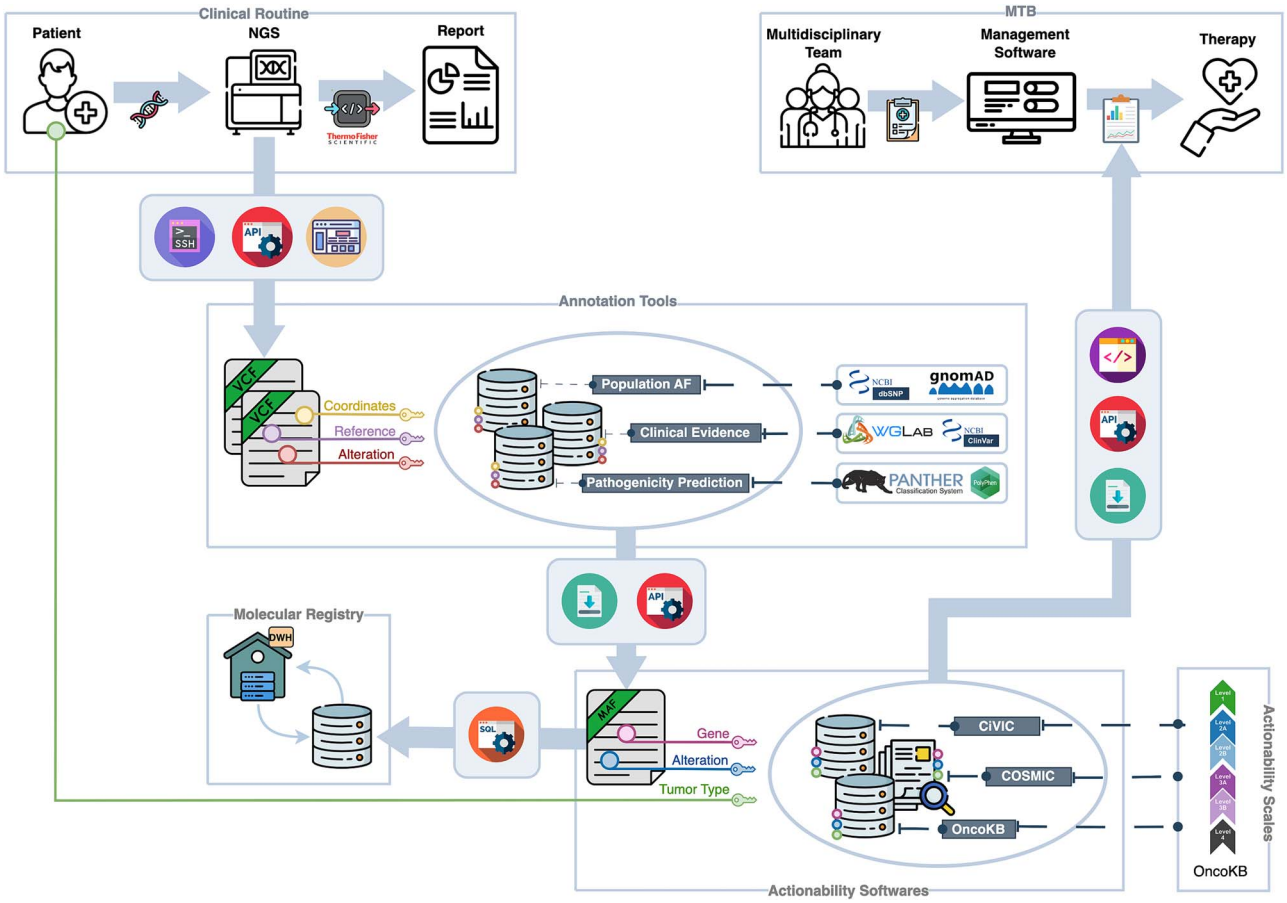

Figure 1. High-level representation of software families and the data flows in molecular tumor boards. Colored circles enclosed in boxes indicate the technologies and tools that allow for interaction among different entities. Dashed lines provide more details on entities. Keys indicate the mandatory fields that are required for database queries.

FAIR-ness and Federation for a novel era of digital medicine and clinical research [14]. Another challenge lies in connecting genomic alterations with the relevant patient data in Electronic Health Records (EHRs): for instance, genomics variants alone do not bring the same clinical significance when found in a tissue associated with a presurgery, postsurgery, or post–pre-remission phase.

Due to these driving improvements, Cancer Centers have been strongly driven by the request to empower both the Bioinformatics and the Clinical Informatics workload and to acquire software and digital tools accordingly. In this article, we map and define all the systems and know-how desiderata required for MTB data analysis and management to define the basics of how this new array of methodologies is to be integrated into the clinical practice.

Digital framework

Software families

Tools and software frameworks have been clustered into families in Fig. 1. Even if overlaps in features may occur across different families, the provided clustering is meant to assign specific tools to the most critical steps in MTB data processing from both the Bioinformatics and Precision Medicine points of view. The overall workflow is built with a medium to large-sized Cancer Centre setup in mind, stemming from the Italian Alliance Against Cancer experience, and other collaborations with European endeavors of international clinical genomics applications such as the DRUP trial and the DIGICORE consortia [15–18].

Certain aspects represent *de facto* standards, as illustrated in Table S1, such as the widely accepted VCF/MAFs formats for DNA variants. Meanwhile, others exhibit asymmetry concerning the diversity of biotechnological approaches. Some tools are specifically designed for Bioinformaticians and Data Engineers and require Unix systems and coding skills, whereas others, commonly integrated into user-friendly GUIs, aim to serve a wider audience, including the professional figures participating in the MTB [19].

Variant annotators

In a variant-calling pipeline, the raw Variant Call Files (VCFs) contain all the genomic alterations found on the sample concerning the reference genome or a matched control. The variant annotation task involves the association of these variations to biological consequences [20, 21]. The minimum requirement for a variant annotator is the ability to identify and predict all the potential effects of mutations, including those caused by mutations in coding sequences, splice sites, and regulatory sites. Despite the development of various tools for predicting mutational pathogenicity at the biochemical level [22–24], their widespread adoption and clinical utility remain limited. In the context of MTB discussion, the molecular biologist can often comment via these results whether the variant lies in the same exon of a specific known target drug, or whether the aminoacidic change impacts or not protein activity. Furthermore, variations in the sequence of a protein-coding gene can impact activity,

solubility, and cellular localization, and become fundamental in the case of a known drug target.

Literature and the software panorama are quite vast on annotators [25], with a recent interest in tools dedicated to Targeted Deep Sequencing analysis [26], while most of the developed software was associated with large-scale Whole-Exome and Whole-Genome sequencing data [27–29]. Most of this software is now embedded in the automatized end-to-end pipelines for clinical NGS analysis, but they are still of interest when a variant call file must be re-analyzed with updated databases for long-term storage or clinical research purposes (Table S1).

Artificial intelligence validation for variant calling and variant prioritization

Artificial intelligence (AI) and machine learning techniques are embedded in modern variant callers to improve their sensitivity, exploiting the wealth of data produced in the next-genomics era. Softwares such as DeepVariant and Clairvoyante utilize AI-driven methods to effectively remove artifacts [30, 31], whereas others like Mutect2 employ simpler models to accurately detect somatic variants, especially amidst tumor heterogeneity and sequencing artifacts [32]. In the context of variant annotation and prioritization, the abundance of evidence across multiple platforms represents a significant challenge. Tools such as VAAST and VarSeq leverage AI techniques to integrate diverse data sources, encompassing functional impact, evolutionary conservation, and clinical relevance. This enables molecular investigators to concentrate on variants with the greatest potential [33, 34].

Transcriptomics

Other kinds of -omics profiling approaches, predominantly RNA profiling, provide insight that can be useful at the clinical level for drug sensitivity or resistance, to provide a better-suited clinical pathway [35]. However, apart from some transcript-level fusion events [36], it remains a marginal diagnostic tool for most pathologies [37]. Several attempts have been carried out to distill robust biomarkers from high-throughput transcriptomics, albeit their level of reliability remains debated [38–41]. The two most promising scenarios are risk prediction algorithms to assess the utility of adjuvant chemotherapy, such as in the breast cancer setting (*PROSIGNA*) [42], and pan-cancer immunotherapy biomarkers, specifically for Immune-Checkpoint inhibitors towards solid tumors [40, 43]. On one side, the RNA landscape can derive information for both tumor intrinsic and tumor extrinsic factors; on the other side, the level of reproducibility is hindered at the biotechnological level due to RNA volatility and variability over time. Furthermore, an RNA picture suffers more from sampling bias than DNA profiling, given the heterogeneity of the tissue specimen cancer microenvironment in solid tumors. Other transcriptomics-based signatures and algorithms, such as neoantigens calling and allele-specific expression, are expected to gain more importance with the approval of cancer vaccines.

Liquid biopsy and other omics data

Furthermore, other –omics data sources will be fed and analyzed by clinical bioinformaticians in the near future. For instance, proteomics may include relevant information that may not be present in transcriptomics data, and proteome databases provide extensive data on molecular mechanisms and modulators of targets. The National Cancer Institute's Clinical Proteomic Tumor Analysis Consortium (CPTAC) provided the research community with mass spectrometry-based proteomic data to highlight the deregulated pathways of several cancer types. These web resources enable researchers to query cohorts and compare the MTB-derived profiles along with clinical annotations, as well as to compare the expected prognoses from -omics profiles. However, their applications remain preclinical [44–46].

From the metabolomics perspective, personalized therapeutic strategies to improve patient outcomes have been developed [47, 48]. However, several methodological flaws, such as the extreme heterogeneity across biological samples, and the lack of standardized technologies for samples and data processing, lead to the need to supplement this technique with transcriptomic and/or proteomic analyses. As a consequence, the quantity and quality of digital tools close to clinical applications for these approaches are still scarce.

Finally, all these biotech techniques have been gaining increasing interest when applied in a larger timeframe thanks to analytes tracked in fluids, referred to by the umbrella word liquid biopsy [49]. A specific landscape of bioinformatics workflows designed for liquid biopsies is usually related to proprietary platforms, often related to specific models able to discern vendor-specific background noise from signals. Nonetheless, a few tools have been developed with a broader application in mind [50–54]. We thus envision that with the increasing application of liquid biopsy for minimal residual disease monitoring and the improvement of liquid cancer screening techniques, novel standardized statistical, and engineering frameworks will be developed to be applied in the routinary MTB scenario.

Actionability annotation tools

Once the molecular variants are annotated at the biological level, the next step for the MTB workflow is to provide insights into the clinical value of said alterations. This analysis can be carried out with the usage of Bioinformatics resources integrating databases and annotation systems. We hereby define the boundaries and the requirements of said software. An Actionability Annotation Tool (AAT) can be defined as a deterministic system that given a triplet of a cancer type, a gene symbol, and an alteration, can rank all the drug sensitivities and resistances associated with that triplet, if available (Fig. 1). Many such frameworks have been recently developed [55], but not all harbor Application Programming Interfaces (APIs) to automatically embed said annotations in molecular reports, except for OncoKB and the community-driven CIViC [56, 57]. A useful feature is a public web portal to enable non-bioinformaticians to browse and search for sensitivity levels and actionability evidence. A critical separation must be made with Actionability Annotation Scales (AAS), which are theoretical frameworks in which the sensitivity levels are defined into multiple steps—usually the first level containing the regulatory approved drugs for the specific {*Tumour, Gene, Alteration*} combo and the last level associated with preclinical or *in vitro* evidence. Many scales have been defined but not all of them have been automatized in a software system (Table 1). Among the critical challenges associated with these tools are the capability of keeping up to date with the associations with regulatory bodies, and the ability to represent drug associations in several geographical locations. Most of the associations are related to FDA approvals that are usually not reflected, or not aligned with other continental references such as the EMA, although the scales are converging towards the same ranking structure (Table 1) [58].

MTB management systems

From the clinical logistics point of view, MTBs are hyperspecialized disease management teams with a specific focus on drug

Table 1. Most employed actionability scales and systems in clinical bioinformatics and regulatory approvals.

Scale	Sensitivity levels	Resistance levels	Consensus level	Efficacy level	Regulatory	API or web portal
COSMIC	4	1	No	No	FDA	Yes
OncoKB	4	2	No	No	FDA	Yes
CIViC	5	5	5	No	FDA	Yes
ESCAT	5 + lack of evidence	No	No	3	EMA	No

Sensitivity levels: levels of the evidence supporting the sensitivity hypothesis. Resistance levels: levels of the evidence supporting the resistance hypothesis. Consensus level: level of concordance among studies towards a given hypothesis. Efficacy level: level of effectiveness of a treatment. Regulatory: the regulatory body for which approvals are listed. API or web portal: whether the system employs application programming interfaces to be easily embedded in other workflows and/or a public web portal to be queried.

regulations and molecular results. According to the cancer centre's size and internal organization, MTB meetings may happen weekly, biweekly, or monthly [1]. Said meetings are usually held in a remote or hybrid setting. In addition to video-calling tools that experienced massive development and optimization in the postpandemic era [59, 60], management software is needed to keep track of meeting minutes, requests for additional analyses, and the production of a multidisciplinary report. Some of the proposed solutions semiautomatize the production of said reports in which the MTB summarizes the molecular and clinical evidence, and it recommends, if feasible, a treatment strategy.

Several software programs are available for this aim, from internally developed by OECI cancer centers to commercial ones [58, 61, 62]. Some of these tools have evolved into comprehensive data management frameworks that also enable radio- and pathological images to be shared and discussed. Since the focus remains on tracking cancer molecular evolution, the recent upgrades of the cBioPortal framework [63] must be cited. Other than the multicohort molecular analysis system that became one of the de facto standards thanks to his open-source framework, it recently expanded its features into a management system also including patient-level timelines and imaging analysis. Additionally, the molecular data infrastructure was empowered to also describe targeted sequencing panels against the classical Whole-Exome Sequencing coverage of TCGA cohorts [64].

Molecular tumor registries

One of the challenges in routinary genomics profiling of cancer patients is querying and filtering internal cohorts according to their molecular profiles. This step becomes particularly important when the MTB is implemented at the regional or national level, like a multicenter clinical trial. Additionally, the MTB could ask which patients exhibited a noncanonical mutation in an actionable gene, and how many of them were assigned or responded to a specific therapy. Unfortunately, most of the routinary screening is saved in static molecular report PDFs containing only actionable variants, hindering the value of retrospective genomics cohorts. Alternatively, somatic alterations may be digitally stored in systems not following basic FAIR guidelines, from Excel tables to proprietary databases not allowing batch export and workflow analysis transparency. To this aim, cancer centers have started developing 'Molecular registries' to reanalyze and store in databases genomics legacy data [65, 66]. Most of these implementations have anonymized access for research purposes and need to relate to broad informed consent for research use. These data are biotechnologically diverse, with several NGS panels of specific genomic targets employed over the years, and are not always complemented with raw data (FASTQ/uBAMs) or proper digitalized annotation. Said registries promise to relate to EHRs and internal data warehouses, and when applicable federated search to enable data FAIR-ness at the molecular level. Finally, said registries could be connected and interoperate with Biobank Data centers, to enable easier research for data and samples filtered at the molecular levels.

Knowledge bases

Linking molecular alterations to clinical trials is a critical step for MTBs. This information can be employed to try and find patients matching enrolment criteria or to mine past knowledge of drug sensitivity that may still not be included in AAS. Knowledge bases (KBs) could feature AI-enhanced search engines that via automated Natural Language Processing and Web Crawling techniques can summarize thousands of clinical trial records related to a '< Gene, Alteration > query'. These frameworks exist in stand-alone format [26, 67] and are embedded in some AAS as separate features [57]. Many challenges still exist to match human parsing specificity, especially for results about toxicity, adverse effects, and secondary objectives in Phase 2 trials.

Bioinformatician contribution to the molecular tumor board

All the aspects mentioned in the software family section of this review provide a high-level overview of what are the mandatory tools that could be exploited in the MTB context. Before dealing with the human skills requirements, we will discuss specifically what all the mentioned digital frameworks and topics mean for preparation and MTB discussion, including lessons that can be learned from the routinary molecular data workup.

'Variant annotators' are pivotal for the interpretation of quality metrics of internal or third-party molecular assays. In cases of dubious results, bioinformaticians enable the request for raw data and reanalysis in case of lack of validation of known alterations, or the cases of alterations scores close to thresholds. This aspect is of utter importance in cases of 'copy number variations' of low-middle ploidy. In the case of 'AI'-aided assays, it is fundamental to provide an interpretation of the AI subtype embedded in molecular assay analysis. This can range from classical automated statistical modeling up to more complex scenarios in the case of deep learning or neural network predictors. The explanation of the AI-derived molecular biomarkers' robustness can be provided for internal and especially for external, third-party parties, and commercial assays.

Regarding RNA and 'transcriptomics', a useful contribution is the interpretation and quality check of clinical-grade RNA data, mostly RNA fusions that have different kinds of thresholds and coverage metrics than DNA-based alterations and usually lack a universal baseline. On larger RNA-based assays, up to large and full transcriptomics, bioinformaticians are the MTB participants most aware of the complexity of RNA measurement, normalization, and unit of measurement. The questions hereby raised can

be related to the kind of internal or external artificial spike-in normalizers to reach an absolute value of a certain analyte or the relative abundance measured in classical transcript per million (TPM) levels.

When dissecting the type of mathematical and digital analysis embedded in a 'liquid biopsy' assay, a bioinformatician can better interpret the normalization strategy carried out for the background signal. If is a proprietary platform, explain specifically the type of validation carried out and how wild-type or mutated molecular tags are measured. Also in this case, when LB or minimal residual disease assays fail to predict recurrence or their results fall close to interpretation thresholds, dealing with raw data can provide an explanation and a better validation of the overall process.

When associating putative drugs to alteration via 'actionability annotation tools', which can be exploited via web interfaces and by nonbioinformaticians, a digital savvy expert can provide an interpretation of how actionability levels are associated across scales. These aspects lie at the intersection of clinical data accessibility and general data FAIRness, such as 'Molecular Tumor Registries' that enable the exploration of internal and external casuistries, and provide an estimate of the penetrance, frequency, and co-occurrence of molecular alterations with other DNA/RNA events, on the same amino acid, codon, or gene/protein target. In the same context, KB's can be exploited by the bioinformatician before the discussion to provide a specific amount of clinical validation and clinical research currently laid out for a specific gene or variant, via automated mining of clinical trials.

Human skills requirements

Most MTB definitions list a bioinformatician along with their minimal requirement but do not provide a specific list of skills that these bioinformaticians must have. Given the multifaceted definitions of Bioinformatics and Computational Biology [68–71], we hereby define a list of hard and soft skills required for the tumor board Bioinformatician (Table S2). The skill catalog starts from issues in Biobanking and sample storage towards more Bioinformatics and Data Engineering issues of results interpretation.

Firstly, tissue storage is a pivotal matter to be considered when dealing with Cancer Genomics, which still heavily relies on FFPE tissues, that, albeit heavily improved in quality and standardization thanks to Biobanking advancements, still represent a strong source of noise and false positives in mutational calling [72–74], well known to molecular pathology departments (Table S2). Additionally, strategies for sequencing panel design and optimization are critical to understanding how to track molecular alteration with a high signal-to-noise ratio. As an example, the usage of UMIs has strongly increased in Biotech to reduce duplication biases and better impute the specific molecule source in low-material sequencing, which was not part of the standard biotechnological package at the beginning of the targeted sequencing era (Table S2.2). After sequencing, raw analysis of data is required to transform reads into processed matrixes for downstream annotation. This step is increasingly treated as a commodity by Biotech/Pharma companies, while the art of creating an optimized workflow on a computing platform was a mandatory process for most bioinformaticians working on Genomics. The revolution of end-to-end virtualization techniques such as Docker and singularity enabled improved reproducibility of these complex workflows that would need to be finely understood, whether any problem arises in processed data results [75, 76] (Table S2.3, Table S2.10). Of note, the protection of raw sequence data is not part of the classical IT department skillset, and it needs to be integrated with the bioinformatician support, after the advent and application of GPDR regulation in Europe and similar legal frameworks internationally. After all the pre-processing and computing steps, it is required to annotate and interpret germline and somatic variations at the biological level. This part is already at the intersection of medical knowledge, but basic and intermediate technical concepts such as variant and population allele frequency, coverage, and intersection with available GWAS results must be in the Bioinformatician toolbox (Tables S2.4, S2.5, and S2.7). An additional task that is still required to interpret variants of uncertain significance that do not fall in hotspot regions is the manual inspection of reads via dedicated software like the integrative genome browser [77]. Manual inspection is an action complementary to AI-curated high-quality variant calling, which enables tracking artifacts when variants fall, for instance, in amplicon flanks or pseudogenes. This task is usually partitioned among the bioinformatician and the digital-savvy clinical biologist.

Apart from the actionability evidence levels that need basic translational knowledge to be interpreted, Biostatistics comes into the field when in need of interpreting clinical trials results associated with a particular target drug or complex biomarker (Table S2.8). Furthermore, classical pharmacogenomics is still not part of cancer routine workup, but it will become increasingly popular when whole-exome and whole-genome analyses are employed. Finally, with the increasing application scenarios of AI in precision medicine application, the clinical bioinformatics framework must feature workflows of increasing complexity, which employ machine learning models to link alterations with putative clinical utility [78–80]. The application of said frameworks must be handled by personnel able to discern evidence levels stemming from models having a level of embedded nondeterminism when handling evidence about the preclinical, *in silico* domain.

Digital skills education perspectives

Interdisciplinary contexts pose challenges in the education field. Students or personnel in training, interested in playing a role at the intersection of personalized medicine and computer science, would be likely uncertain of the main tracks to follow.

All the skills can be clustered into three macroareas processing, engineering, and interpretation (Fig. 2). With processing, we refer to the path that links the sample and the patient to data, engineering to large-scale analytics that requires high computational power and the whole data life cycle, and finally, interpretation with the processing of genomics information that ends up with results usable at the medical and biological level.

Depending on the minor-bachelor background of the Bioinformatician, the student can decide to invest time in improving one of these macrodomains with specific academic courses and online training. For processing and interpretation, many online and onsite courses are provided at the European Level by Research Infrastructures such as BBMRI-ERIC (Biobanking and Processing) and ELIXIR (variant analysis and interpretation) [81–84]. Interpretation courses can be found at EMBL [85], while the main bioinformatics education platform remains GOBLET [86].

Other more specifically digital skills have a steep ladder of complexity, from basic programming and scripting up to dealing with operating systems and virtualization techniques to ensure end-to-end reproducibility of the MTB process. All these aspects can be found both in academia and in other noninstitutional training platforms like Codeacademy [87].

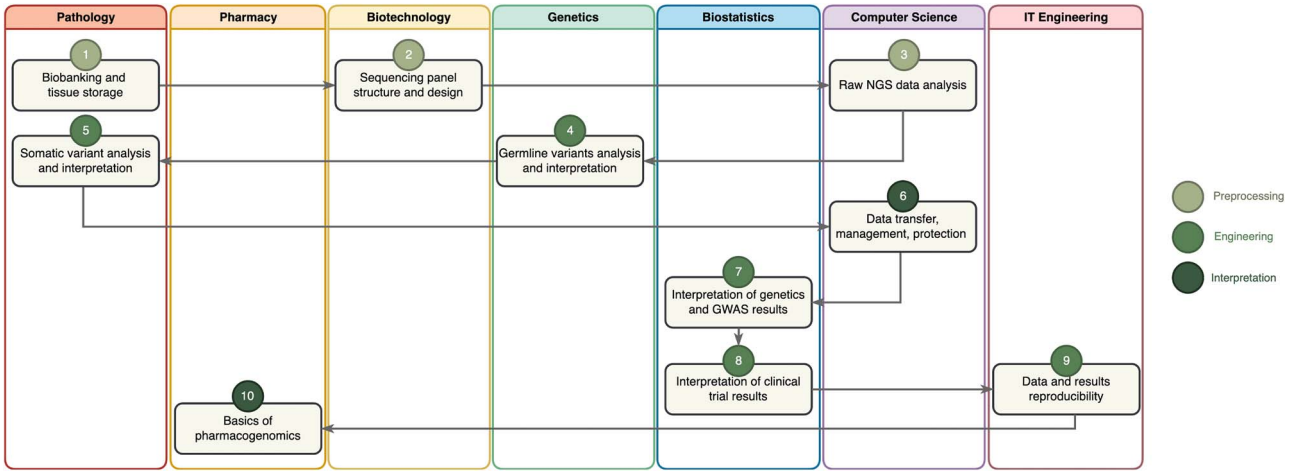

Figure 2. Graphical representation of the required human skillset, following sample processing, analysis, and interpretation. Color coding follows skill clustering as in Table S2.

From a general perspective, two main kinds of Bioinformatics skill sets can be identified. The first is more related to Data Science and Interpretation, which requires deep knowledge of one scripting language along with its framework and the second profile is closer to Genomics raw data processing, in that case operating systems and engineering techniques are required to transform massive-scale datasets into structured data matrixes.

The two kinds can co-exist in the same professional skillset, but a specialization could be required for facilities, hospitals, and MTBs that process a large number of biotechnological diverse assays.

The skill catalog is not meant to be comprehensive nor strictly minimal, and we acknowledge the variability of academic and professional trajectories that could end up in this workforce. Indeed, many bioinformaticians start from a molecular biology background and can become highly proficient in computational skills after their Master's degree, while for tasks closer to the Information Technology domain an Engineering or Computer Science foundation may be more fitting. All these aspects become a delicate matter when computational skills are intertwined in clinical decisions. Given these considerations, and since interdisciplinary professions follow different rules inside and outside Europe, we do not hereby provide a specific list of university degrees associated with these skills. To conclude, whether the Medical field will consider clinical bioinformaticians as mere technicians or higher level professionals with a specific path qualification of specialization still must be decided and defined in many European countries.

Conclusions

Clinical Bioinformatics is a novel interdisciplinary clinical framework and skillset that requires better definition, given its delicate intersection with the medical field. A critical risk to be avoided is to leave this whole field as a technical nuance to be subcontracted to Biotech and Pharma: this would lead to a critical loss of know-how and decision power of the whole biomedical field, in a critical moment where AI is starting to play an important and unprecedented role. Importantly, many of these novel tools will be regulated as medical devices. We thus envision that novel decision-makers in biomedicine must contain both computational-savvy medical doctors and clinical bioinformaticians.

Key Points

- Bioinformatics is a mandatory component of Molecular Tumor Boards.
- Digital tools and software frameworks are required for managing the complexity of Cancer Genomics and enable reproducible analysis and interpretation in the clinical setting.
- We cluster and describe the needed software and databases for Molecular Tumor Board Data Analysis and Management.
- We define for the first time a minimal skillset of human skills in MTB Clinical Bioinformatics, from Biobanking theory to IT Engineering.
- Educational and specialized training programs are necessary to equip future bioinformaticians with the required knowledge.

Acknowledgements

Part of this work has been shared and integrated with the EU-funded Cancer Mission CAN.Heal Project, specifically WP10 on Decision Support Tools for Molecular Tumor Board, on behalf of Alliance Against Cancer and DIGICORE.

Supplementary data

Supplementary data is available at *Briefings in Bioinformatics* online.

Author contributions

Matteo Pallocca (Conceptualization), Matteo Pallocca, Martina Betti, Sara Baldinelli (Data curation, Formal Analysis, Visualization), Matteo Pallocca, Martina Betti (Methodology and Investigation), Martina Betti (Resources), Matteo Pallocca (Writing—original draft), Matteo Pallocca, Martina Betti, Sara Baldinelli, Ramona Palombo, Gabriele Bucci, Luca Mazzarella, Giovanni Tonon, Gennaro Ciliberto (Writing—review & editing). Matteo Pallocca (Supervision), and Matteo Pallocca, Gennaro Ciliberto (Project administration, Funding acquisition); Matteo Pallocca accepted full responsibility for the study, had access to

the data and controlled the decision to submit for publication. All authors read and approved the final version of the manuscript.

Biographical note

Matteo Pallocca acted as a Bioinformatics Head of Regina Elena National Cancer Institute and he is currently the Data Engineering coordinator for the network of Italian Biobanks (BBMRI.it) at CNR, working towards federated -omics analysis.

The team behind this work represents large Real World Data networks such as DIGICORE (Gennaro Ciliberto, Giovanni Tonon), the Data Engineering Team of the Strenghtening BBMRI.it project (Matteo Pallocca, Ramona Palombo), and the Bioinformatics core of Alliance Against Cancer (Luca Mazzarella, Gabriele Bucci, Martina Betti, Sara Baldinelli). Gennaro Ciliberto is the scientific director of the IRCCS Regina Elena National Cancer Institute.

Conflict of interest: MP acts as a consultant for *Dexma srl* outside the scope of this work. All other authors state that they have no conflicts of interest.

Funding

This work has been supported by funding from the Italian Ministry of Health (Ministero della Salute, Ricerca Corrente), Ricerca Corrente Reti 2022 (Project ACCORD—The ACC OMICS REGISTRY, CUP Master J47G22000300001). The resources on biobanking were supported also by the European Union (Next Generation EU), Italian NRRP project code IR0000031—*Strengthening BBMRI.it*—CUP B53C22001820006.

Data availability

Not applicable.

References

1. Luchini C, Lawlor RT, Milella M. *et al.* Molecular tumor boards in clinical practice. *Trends Cancer* 2020;**6**:738–44. https://doi.org/10.1016/j.trecan.2020.05.008.
2. Tsimberidou AM, Kahle M, Vo HH. *et al.* Molecular tumour boards—Current and future considerations for precision oncology. *Nat Rev Clin Oncol* 2023;**20**:843–63. https://www.nature.com/articles/s41571-023-00824-4.
3. Jiménez-Santos MJ, García-Martín S, Fustero-Torre C. *et al.* Bioinformatics roadmap for therapy selection in cancer genomics. *Mol Oncol* 2022;**16**:3881–908. https://doi.org/https://onlinelibrary.wiley.com/doi/full/10.1002/1878-0261.13286.
4. Tamborero D, Dienstmann R, Rachid MH. *et al.* The molecular tumor board portal supports clinical decisions and automated reporting for precision oncology. *Nature Cancer* 2022;**3**:251–61. https://www.nature.com/articles/s43018-022-00332-x.
5. Schmid S, Jochum W, Padberg B. *et al.* How to read a next-generation sequencing report-what oncologists need to know. *ESMO Open* 2022;**7**:100570. https://doi.org/10.1016/j.esmoop.2022.100570.
6. MacConaill LE, Campbell CD, Kehoe SM. *et al.* Profiling critical cancer gene mutations in clinical tumor samples. *PloS One* 2009;**4**. https://doi.org/10.1371/annotation/613c7509-e4c9-42ac-82fb-fc504400d9e0.
7. Thomas RK, Baker AC, DeBiasi RM. *et al.* High-throughput oncogene mutation profiling in human cancer. *Nat Genet* 2007;**39**. 2007;39:347–51. https://www.nature.com/articles/ng1975.
8. Merrie AEH, Yun K, McCall JL. *et al.* Utilization of polymerase chain reaction technology in the detection of solid tumors [5] (multiple letters). *Cancer* 1999;**85**:248–9. https://doi.org/10.1002/(SICI)1097-0142(19990101)85:1<248::AID-CNCR39>3.0.CO;2-V.
9. Sokolenko AP, Imyanitov EN. Molecular diagnostics in clinical oncology. *Front Mol Biosci* 2018;**5**:402417.
10. Bernard PS, Wittwer CT. Real-time PCR Technology for Cancer Diagnostics. *Clin Chem* 2002;**48**:1178–85. https://dx.doi.org/10.1093/clinchem/48.8.1178.
11. Sinkala M. Mutational landscape of cancer-driver genes across human cancers. *Sci Rep* 2023;**13**:1–14. https://www.nature.com/articles/s41598-023-39608-2.
12. Nicora G, Zucca S, Limongelli I. *et al.* A machine learning approach based on ACMG/AMP guidelines for genomic variant classification and prioritization. *Sci Rep* 2022;**12**:1–12. [cited 8 April 2024]. https://www.nature.com/articles/s41598-022-06547-3.
13. Manshaei R, DeLong S, Andric V. *et al.* GeneTerpret: A customizable multilayer approach to genomic variant prioritization and interpretation. *BMC Med Genomics* 2022;**15**:1–11. https://bmcmedgenomics.biomedcentral.com/articles/10.1186/s12920-022-01166-3.
14. Sinaci AA, Núñez-Benjumea FJ, Gencturk M. *et al.* From raw data to FAIR data: The FAIRification workflow for Health Research. *Methods Inf Med* 2020;**59**:E21–32. https://doi.org/10.1055/s-0040-1713684.
15. Mahon P, Chatzitheofilou I, Dekker A. *et al.* A federated learning system for precision oncology in Europe: DigiONE. *Nat Med* 2024;**30**:334–7. https://doi.org/10.1038/s41591-023-02715-8.
16. Pallocca M, Molineris I, Berrino E. *et al.* Comprehensive genomic profiling on metastatic melanoma: Results from a network screening from 7 Italian cancer centres. *J Transl Med* 2024;**22**:29. https://doi.org/10.1186/s12967-023-04776-2.
17. Gregorc V, Mazzarella L, Lazzari C. *et al.* Prospective validation of the Italian alliance against cancer lung panel in patients with advanced non–small-cell lung cancer. *Clin Lung Cancer* 2021;**22**:e637–41. https://doi.org/10.1016/j.cllc.2020.12.007.
18. Hoes LR, van Berge Henegouwen JM, van der Wijngaart H. *et al.* Patients with rare cancers in the drug rediscovery protocol (DRUP) benefit from genomics-guided treatment. *Clin Cancer Res* 2022;**28**:1402–11. https://doi.org/10.1158/1078-0432.CCR-21-3752.
19. Reimer N, Unberath P, Busch H. *et al.* Challenges and experiences extending the cBioPortal for cancer genomics to a molecular tumor board platform. *Stud Health Technol Inform* 2021;**287**:139–43. https://doi.org/10.3233/SHTI210833.
20. D'Antonio M, D'Onorio De Meo P, Paoletti D. *et al.* WEP: A high-performance analysis pipeline for whole-exome data. *BMC Bioinformatics* 2013;**14**:1–11. https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-14-S7-S11.
21. Chiara M, Gioiosa S, Chillemi G. *et al.* CoVaCS: A consensus variant calling system. *BMC Genomics* 2018. cited 14 September 2023;**19**:1–9. https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-018-4508-1.
22. Alanazi M, Abduljaleel Z, Khan W. *et al.* In Silico analysis of single nucleotide polymorphism (SNPs) in human β-globin gene. *PloS One* 2011;**6**:25876. https://doi.org/10.1371/journal.pone.0025876.
23. Thomas PD, Ebert D, Muruganujan A. *et al.* PANTHER: Making genome-scale phylogenetics accessible to all. *Protein Sci* 2022;**31**.
24. Sim NL, Kumar P, Hu J. *et al.* SIFT web server: Predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res* 2012;**40**:W452–7. https://doi.org/10.1093/nar/gks539.
25. Borchert F, Mock A, Tomczak A. *et al.* Knowledge bases and software support for variant interpretation in precision

oncology. *Brief Bioinform* 2021;**22**:1–17. https://doi.org/10.1093/bib/bbab134.
26. Thermo OKR. [Internet]. https://www.thermofisher.com/order/catalog/product/A34298.
27. Li Q, Ren Z, Cao K. *et al.* CancerVar: An artificial intelligence–empowered platform for clinical interpretation of somatic mutations in cancer. *Sci Adv* 2022;**8**:1624. https://doi.org/10.1126/sciadv.abj1624.
28. Li Q, Wang K. InterVar: Clinical interpretation of genetic variants by the 2015 ACMG-AMP guidelines. *Am J Hum Genet* 2017;**100**:267.
29. Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010;**38**:e164. https://doi.org/10.1093/nar/gkq603.
30. Luo R, Sedlazeck FJ, Lam TW. *et al.* A multi-task convolutional deep neural network for variant calling in single molecule sequencing. *Nat Commun* 2019;**10**:998.
31. Poplin R, Chang PC, Alexander D. *et al.* A universal snp and small-indel variant caller using deep neural networks. *Nat Biotechnol* 2018;**36**:983–7. https://doi.org/10.1038/nbt.4235.
32. Tools VD. MuTect2. *GATK Manual* 2017. https://gatk.broadinstitute.org/hc/en-us/articles/360037593851-Mutect2.
33. Boudellioua I, Kulmanov M, Schofield PN. *et al.* DeepPVP: Phenotype-based prioritization of causative variants using deep learning. *BMC Bioinformatics* 2019;**20**:65. https://doi.org/10.1186/s12859-019-2633-8.
34. Ruscheinski A, Reimler AL, Ewald R. *et al.* VPMBench: A test bench for variant prioritization methods. *BMC Bioinformatics* 2021;**22**:543. https://doi.org/10.1186/s12859-021-04458-0.
35. Byron SA, Van Keuren-Jensen KR, Engelthaler DM. *et al.* Translating RNA sequencing into clinical diagnostics: Opportunities and challenges. *Nat Rev Genet* 2016;**17**:257–71. [cited 16 March 2024]. https://www.nature.com/articles/nrg.2016.10.
36. Li J, Lu H, Ng PKS. *et al.* A functional genomic approach to actionable gene fusions for precision oncology. *Sci Adv* 2022;**8**. [cited 16 March 2024], https://pubmed.ncbi.nlm.nih.gov/35138907/.
37. Emde-Rajaratnam M, Beck S, Benes V. *et al.* RNA-sequencing based first choice of treatment and determination of risk in multiple myeloma. *Front Immunol* 2023;**14**:1286700. https://doi.org/10.3389/fimmu.2023.1286700.
38. Tsimberidou AM, Fountzilas E, Bleris L. *et al.* Transcriptomics and solid tumors: The next frontier in precision cancer medicine. *Semin Cancer Biol* 2022;**84**:50–9. https://doi.org/10.1016/j.semcancer.2020.09.007.
39. Sun S, Xu L, Zhang X. *et al.* Systematic assessment of transcriptomic biomarkers for immune checkpoint blockade response in cancer immunotherapy. *Cancers (Basel)* 2021;**13**.
40. Pallocca M, Angeli D, Palombo F, Sperati F, Milella M, Goeman F., *et al.* Combinations of immuno-checkpoint inhibitors predictive biomarkers only marginally improve their individual accuracy. *J Transl Med* 2019;**17**:1–8. https://doi.org/10.1186/s12967-019-1865-8.
41. Merry E, Thway K, Jones RL. *et al.* Predictive and prognostic transcriptomic biomarkers in soft tissue sarcomas. *npj Precision Oncology* 2021;**5**:1–8. [cited 16 March 2024]. https://www.nature.com/articles/s41698-021-00157-4.
42. Buus R, Szijgyarto Z, Schuster EF. *et al.* Development and validation for research assessment of Oncotype DX® breast recurrence score, EndoPredict® and Prosigna®. *npj Breast Cancer* 2021;**7**:1–8. https://www.nature.com/articles/s41523-021-00216-w.
43. Jiang P, Gu S, Pan D. *et al.* Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat Med* 2018;**24**:1550–8. https://doi.org/10.1038/s41591-018-0136-1.
44. Monsivais D, Vasquez YM, Chen F. *et al.* Mass-spectrometry-based proteomic correlates of grade and stage reveal pathways and kinases associated with aggressive human cancers. *Oncogene* 2021;**40**:2081–95. [cited 216 March 2024]. https://www.nature.com/articles/s41388-021-01681-0.
45. Chen F, Chandrashekar DS, Varambally S. *et al.* Pan-cancer molecular subtypes revealed by mass-spectrometry-based proteomic characterization of more than 500 human cancers. Nature. *Communications* 2019;**10**:1–15. https://doi.org/10.1038/s41467-019-13528-0.
46. Kwon YW, Jo HS, Bae S. *et al.* Application of proteomics in cancer: Recent trends and approaches for biomarkers discovery. *Front Med (Lausanne)* 2021;**8**:747333. https://doi.org/10.3389/fmed.2021.747333.
47. Han J, Li Q, Chen Y. *et al.* Recent metabolomics analysis in tumor metabolism reprogramming. *Front Mol Biosci* 2021;**8**:763902. https://doi.org/10.3389/fmolb.2021.763902.
48. Stine ZE, Schug ZT, Salvino JM. *et al.* Targeting cancer metabolism in the era of precision oncology. *Nat Rev Drug Discov* 2021;**21**:141–62. [cited 16 March 2024]. https://www.nature.com/articles/s41573-021-00339-6.
49. Crowley E, Di Nicolantonio F, Loupakis F. *et al.* Liquid biopsy: Monitoring cancer-genetics in the blood. *Nat Rev Clin Oncol* 2013;**10**:472–84. https://doi.org/10.1038/nrclinonc.2013.110.
50. Wolf J, Rasmussen DK, Sun YJ. *et al.* Liquid-biopsy proteomics combined with AI identifies cellular drivers of eye aging and disease in vivo. *Cell* 2023;**186**:4868–4884.e12. https://doi.org/10.1016/j.cell.2023.09.012.
51. Luo H, Wei W, Ye Z. *et al.* Liquid biopsy of methylation biomarkers in cell-free DNA. *Trends Mol Med* 2021;**27**:482–500. https://doi.org/10.1016/j.molmed.2020.12.011.
52. Hou J, Li XT, Xie KP. Coupled liquid biopsy and bioinformatics for pancreatic cancer early detection and precision prognostication. *Mol Cancer* 2021;**20**:34. https://doi.org/10.1186/s12943-021-01309-7.
53. Brockley LJ, Souza VGP, Forder A. *et al.* Sequence-based platforms for discovering biomarkers in liquid biopsy of non-small-cell lung cancer. *Cancers (Basel)* 2023;**15**:2275. https://doi.org/10.3390/cancers15082275.
54. Jackson F, Lukasiewicz T. Deconvolution of cell-free DNA in cancer liquid biopsy using a deep AutoEncoder. *BCB '23: Proceedings of the 14th ACM International Conference on Bioinformatics, Computational Biology, and Health*. 2023;**32**:1–6. https://doi.org/10.1145/3584371.3612976.
55. Tamborero D, Dienstmann R, Rachid MH. *et al.* Support systems to guide clinical decision-making in precision oncology: The cancer Core Europe molecular tumor board portal. *Nat Med* 2020;**26**:992–4. [cited 25 September 2023]. https://www.nature.com/articles/s41591-020-0969-2.
56. Chakravarty D, Gao J, Phillips S. *et al.* OncoKB: A precision oncology Knowledge Base. *JCO Precis Oncol* **1**:1–16.
57. Griffith M, Spies NC, Krysiak K. *et al.* CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. *Nat Genet* 2017;**49**:170–4. [cited 25 September 2023]. https://www.nature.com/articles/ng.3774.
58. Introducing Simplified OncoKB™ Levels of Evidence. [cited 25 September 2023]. https://www.oncokb.org/news#12202019.
59. Security Guide. https://explore.zoom.us/docs/doc/Zoom-Security-White-Paper.pdf.
60. Microsoft teams calling for the modern workplace and worker simplifying voice and unifying systems with teams voice. https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RWSxAm.

61. Canfora M, Pallocca M, Allegretti M. *et al.* The MTB-orchestrator: A real-world clinical cloud system for collaborative molecular tumor board analysis. *Preprint*.
62. navify® Portal - Roche Website. [cited 25 September 2023], https://navifyportal.roche.com/.
63. Gao J, Aksoy BA, Dogrusoz U. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 2013;**6**:pl1. https://doi.org/10.1126/scisignal.2004088.
64. (2) cBioPortal webinar 2: Mutation details & patient view. YouTube [Internet]. https://www.youtube.com/watch?v=uJsp9kd2jIk.
65. Green MF, Bell JL, Hubbard CB. *et al.* Implementation of a molecular tumor registry to support the adoption of precision oncology within an Academic Medical Center: The Duke University experience. 2021;**5**:PO.21.00030. https://doi.org/10.1200/PO.21.00030.
66. Betti M, Maria Salzano C, Massacci A. *et al.* Development of a somatic variant registry in a National Cancer Center: Towards molecular real world data preparedness. *J Biomed Inform* 2023;**142**:104394. https://doi.org/10.1016/j.jbi.2023.104394.
67. Illumina Connected Analytics. https://www.illumina.com/products/by-type/informatics-products/connected-analytics.html.
68. Can T. Introduction to bioinformatics. *Methods Mol Biol* 2014;**1107**:51–71. https://doi.org/10.1007/978-1-62703-748-8_4.
69. Luscombe NM, Greenbaum D, Gerstein M. What is bioinformatics? A proposed definition and overview of the field. *Methods Inf Med* 2001;**40**:346–58. https://doi.org/10.1055/s-0038-1634431.
70. Diniz WJS, Canduri F. REVIEW-ARTICLE bioinformatics: An overview and its applications. *Genet Mol Res* 2017;**16**:16. https://doi.org/10.4238/gmr16019645.
71. Oulas A, Minadakis G, Zachariou M. *et al.* Systems bioinformatics: Increasing precision of computational diagnostics and therapeutics through network-based approaches. *Brief Bioinform* 2019;**20**:806–24. https://doi.org/10.1093/bib/bbx151.
72. Berrino E, Annaratone L, Miglio U. *et al.* Cold formalin fixation guarantees DNA integrity in formalin fixed paraffin embedded tissues: Premises for a better quality of diagnostic and experimental pathology with a specific impact on breast cancer. *Front Oncol* 2020;**10**. https://doi.org/10.3389/fonc.2020.00173.
73. Annaratone L, De Palma G, Bonizzi G. *et al.* Basic principles of biobanking: From biological samples to precision medicine for patients. *Virchows Arch* 2021;**479**:233–46. https://doi.org/10.1007/s00428-021-03151-0.
74. Guo Q, Lakatos E, Al BI. *et al.* The mutational signatures of formalin fixation on the human genome. *Nat Commun* 2022;**13**:1–14. https://www.nature.com/articles/s41467-022-32041-5.
75. MerkelDirk. Docker. *Linux Journal* 2014. https://dl.acm.org/doi/10.5555/2600239.2600241.
76. Kurtzer GM, Sochat V, Bauer MW. Singularity: Scientific containers for mobility of compute. *PloS One* 2017;**12**(5):e0177459. https://pubmed.ncbi.nlm.nih.gov/28494014/.
77. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative genomics viewer (IGV): High-performance genomics data visualization and exploration. *Brief Bioinform* 2013;**14**:178–92. https://doi.org/10.1093/bib/bbs017.
78. Bernasconi A, Pais RJ. Predictive modelling in clinical bioinformatics: Key concepts for Startups. *BioTech* 2022;**11**:35. https://www.mdpi.com/2673-6284/11/3/35/htm.
79. Taveira N, Fernandes AI, Pais RJ. *et al.* Predicting cancer prognostics from tumour Transcriptomics using an auto machine learning approach. *Medical Sciences Forum* 2023;**22**:6. https://www.mdpi.com/2673-9992/22/1/6/htm.
80. Filho UL, Pais TA, Pais RJ. Facilitating "omics" for phenotype classification using a user-friendly AI-driven platform: Application in cancer prognostics. *BioMedInformatics* 2023;**3**:1071–82. https://www.mdpi.com/2673-7426/3/4/64/htm.
81. Mayrhofer MT, Holub P, Wutte A. *et al.* BBMRI-ERIC: The novel gateway to biobanks: From humans to humans. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* 2016;**59**:379–84. https://doi.org/10.1007/s00103-015-2301-8.
82. Morgan SL, Palagi PM, Fernandes PL. *et al.* The ELIXIR-EXCELERATE train-the-trainer pilot programme: Empower researchers to deliver high-quality training. *F1000Res* 2017;**6**:1557. https://doi.org/10.12688/f1000research.12332.1.
83. BBMRI Italia. [cited 28 June 2024]. https://www.bbmri.it/.
84. BBMRI Accademy. [Internet]. [cited 28 June 2024]. https://www.bbmri-eric.eu/services/what-is-the-bbmri-eric-academy/.
85. EMBL Training. [cited 28 June 2024]. https://www.embl.org/training/.
86. Corpas M, Jimenez RC, Bongcam-Rudloff E. *et al.* The GOBLET training portal: A global repository of bioinformatics training materials, courses and trainers. *Bioinformatics* 2015;**31**:140–2. https://doi.org/10.1093/bioinformatics/btu601.
87. Code Accademy - Course catalog. [cited 28 June 2024]. https://www.codecademy.com/catalog.