

Concepts of multi-level dynamical modelling: Understanding mechanisms of squamous cell carcinoma development in Fanconi anemia

Eunike Velleuer^{1,2}, Elisa Domínguez-Hüttinger³, Alfredo Rodriguez^{3,4}, Leonard Harris⁵, and Carsten Carlberg^{6,7}

¹Helios Klinikum

²Heinrich Heine University

³Universidad Nacional Autónoma de México

⁴Instituto Nacional de Pediatría

⁵University of Arkansas

⁶University of Eastern Finland

⁷Polish Academy of Sciences

July 25, 2023

Abstract

Fanconi anemia (FA) is a rare disease (incidence of 1:300,000) primarily based on the inheritance of pathogenic variants in genes of the FA/BRCA (breast cancer) pathway. These variants ultimately reduce the functionality of different proteins involved in the repair of DNA interstrand crosslinks and DNA double-strand breaks. At birth, individuals with FA might present with typical malformations, particularly radial axis and renal malformations, as well as other physical abnormalities like skin pigmentation anomalies. During the first decade of life, FA mostly causes bone marrow failure due to reduced capacity and loss of the hematopoietic stem and progenitor cells. This often makes hematopoietic stem cell transplantation necessary, but this therapy increases the already intrinsic risk of developing squamous cell carcinoma (SCC) in early adult age. Due to the underlying genetic defect in FA, classical chemo-radiation-based treatment protocols cannot be applied. Therefore, detecting and treating the multi-step tumorigenesis process of SCC in an early stage, or even its progenitors, is the best option for prolonging the life of adult FA individuals. However, the small number of FA individuals makes classical evidence-based medicine approaches based on results from randomized clinical trials impossible. As an alternative, we introduce here the concept of multi-level dynamical modelling using large, longitudinally collected genome, proteome- and transcriptome-wide data sets from a small number of FA individuals. This mechanistic modelling approach is based on the “hallmarks of cancer in FA”, which we derive from our unique database of the clinical history of over 750 FA individuals. Multi-omic data from healthy and diseased tissue samples of FA individuals are to be used for training constituent models of a multi-level tumorigenesis model, which will then be used to make experimentally testable predictions. In this way, mechanistic models facilitate not only a descriptive but also a functional understanding of SCC in FA. This approach will provide the basis for detecting signatures of SCCs at early stages and their precursors so they can be efficiently treated or even prevented, leading to a better prognosis and quality of life for the FA individual.

Introduction

Rare diseases are disorders that affect less than one case in 2000 people, *i.e.*, only a small percentage of the population. However, there are more than 6000 known rare diseases, affecting over 300 million people worldwide (1, 2). In 80% of the cases, the origin of a rare disease is one or multiple disadvantageous inherited variations of the genome (3). These are present in all cell types of the affected individual. Nevertheless, most rare diseases, which are not already prenatally lethal, are rather tissue specific. Pediatricians are more likely confronted with rare diseases than healthcare specialists from other disciplines, as those diseases frequently present symptoms early in life. Rare diseases are often referred to as “orphan diseases”, since in comparison to common non-communicable diseases, such as cardiovascular diseases and type 2 diabetes, there is less research and development of therapies for them. Fanconi anemia (FA) belongs to a small group of rare diseases that are investigated more intensively than most others (4). This is also the result of significant contributions from patient organizations in the United States, Germany and many other countries (5).

In general, evidence-based medicine aims to make optimal medical decisions by integrating the experience of a clinician with data from the individual patient and available scientific information on the respective disease (6). The latter information often derives from randomized clinical trials involving large numbers of cases and controls. Those trials are the source for the construction of statistical models, *i.e.*, to quantify mathematical relationships between non-random variables measured from the study participants (**Fig. 1, left**). For common diseases, there is no problem identifying a sufficiently large number of cases to achieve acceptable statistical power of the applied statistical model, *e.g.*, reflected by the p-value. However, this approach cannot be used for rare diseases due to the small number of cases. An alternative approach is to study a few individuals in very high detail by collecting longitudinal samples for many biological parameters (7) (**Fig. 1, right**). For example, multi-omic analyses provide many thousands of data points per individual, such as genome-wide DNA methylation, histone modifications and gene expression. These data, together with mechanistic information on biochemical and regulatory pathways from public databases, such as KEGG (Kyoto Encyclopedia of Genes and Genomes) (8), Wikipathways (9) and SPOKE (Scalable Precision Medicine Open Knowledge Engine) (10), can then be used to construct multi-level dynamical computational models (11, 12). These models can act as virtual platforms for identifying novel therapeutic targets and designing treatment and preventative protocols to improve individual patient outcomes. In this Perspective article, we introduce the concept of mechanistic modelling as a clinical decision support tool in FA, using the example of the multi-step tumorigenesis of squamous cell carcinoma (SCC) in FA individuals.

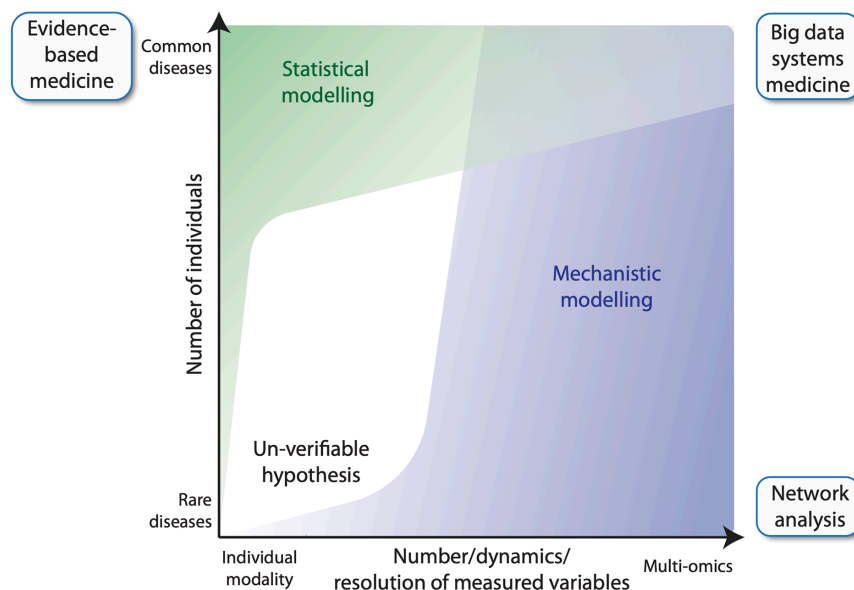


Figure 1: **Evidence-based medicine versus network analysis.** Number of participants of a clinical study versus the number of variables measured for each individual. In classical evidence-based medicine, many cases and controls are studied for a small number of variables (**top left**), whereas network analysis, can build upon fewer patients, for which many variables are measured longitudinally (**bottom right**). The latter benefits from using mechanistic modelling for analyzing data, while evidence-based medicine largely relies on statistical modelling. Big data systems medicine (**top right**) fuses both approaches by integrating large number of data points derived from several patients into predictive mechanistic models.

Fanconi anemia: a master example of a rare disease

FA is a rare disease (1 case in 300,000 persons) that was first described nearly 100 years ago by Guido Fanconi (13). The clinical characteristics of FA are: (i) congenital malformations including absent radius, thumb hypoplasia, disturbed skin pigmentation, as well as inner organ abnormalities most frequently found in the renal and cardiac system as well as disturbed skin pigmentation (14), (ii) progressive bone marrow failure already at childhood age (15, 16), and (iii) dramatically increased risk of developing cancers, such as acute myeloid leukemia (17) and SCC, especially of the head and neck, in early adulthood (18). FA individuals have defects in the molecular machinery of detection and repair of interstrand crosslinks (ICLs) and DNA double-strand breaks (DSBs), which are mostly due to the biallelic inheritance of recessive pathogenic variants in a subset of at least 20 *FANCA* genes (*FANCA*, *FANCC*, *FANCD1* (*BRCA2*), *FANCD2*, *FANCE*, *FANCF*, *FANCG*, *FANCI*, *FANCL*, *FANCM*, *FANCN* (*PALB2*), *FANCO* (*RAD51C*), *FANCP* (*SLX4*), *FANC* (*ERCC4*), *FANCS* (*BRCA1*), *FANCT* (*UBE2T*), *FANCU* (*XRCC2*), *FANCV* (*MAD2L2*), *FANCW* (*RFWD3*)) (19). Moreover, variants in the *FANCB* gene are inherited in an autosomal recessive X-linked manner, whereas the *FANCR* (*RAD51*) shows an autosomal dominant inheritance pattern and can also be spontaneously mutated (19-21). *FANC* genes encode for proteins that maintain genomic integrity during DNA replication, *i.e.*, their inactivation leads to accumulation of DSBs and genomic instability (22). However, patients with identical variants, such as siblings, often show significant differences in their clinical presentation, *i.e.*, there are more factors than the mutated *FANC* genes contributing to the disease (16, 23).

To date, hematopoietic stem cell transplantation is the only curative treatment option for the hematological complications of FA (24) and the main reason for improved life expectancy of young FA individuals (25). In addition, treatment with supra-pharmacological doses of testosterone analogs, such as Oxymetholone, Danazol, and others can stabilize declining blood counts and even improve them (26-29). Non-transplanted and, in particular, transplanted FA individuals have a several 100-fold increased risk for developing SCC, especially of their oral mucosa but also in their pharynx, larynx, esophagus, anus, and vulva (17), even without the main risk factors like alcohol and tobacco exposure. For adult FA individuals, developing SCC is the most life-threatening complication. Due to their dysregulated DNA repair machinery, FA patients cannot tolerate standard chemo-radiation therapies and treatment side effects are difficult to predict (30, 31). This makes non-surgical systemic therapeutical options very limited. Therefore, detecting SCC at an early stage and eliminating it surgically is effectively, at present, the best way to prolong the lives of FA individuals (32). Ultimately, prevention of disease progression should be the goal but is still far away from the clinical routine.

Based on its clinical and cellular phenotype, FA can also serve as a cellular model for the study of general molecular functions and physiological aspects, like aging, as well as other non-communicable diseases occurring in the general population. In that respect, the study of FA has had a major impact on the molecular understanding of breast/ovarian cancer (33). Moreover, *FANC* genes are frequently mutated or dysregulated in sporadic cancers (34), as well as in childhood cancers (35). Nevertheless, the enormous and quite specific cancer risk of SCC for FA individuals is poorly understood from a mechanistic standpoint.

Hallmarks of cancers arising in FA individuals

During tumorigenesis, the cells of all types of solid cancers arising in adults undergo a multi-step process from a healthy, non-transformed cell to low-grade and high-grade dysplasia, carcinoma *in situ*, and invasive cancer (36) (**Fig. 2A**). The early stages of this process are reversible, but when a “point of no return” is reached, an aggressive carcinoma forms, which *via* the spreading of metastases will eventually lead to patient mortality. In cancer patients in the general population, this multi-step tumorigenesis process takes decades, but in FA individuals it can take only months to a few years.

Transformed cells acquire a conserved set of cancer hallmarks During tumorigenesis, including self-sufficiency in growth signals, insensitivity to antigrowth signals, and tissue invasion and metastasis (37). These are altered functional capabilities, the accumulation of which allows malignant cells to survive, proliferate, and disseminate. The hallmarks are outcomes of specific cell fate-controlling regulatory pathways, which in turn are affected by cancer driver gene mutations (38). Environmental factors, like lifestyle decisions on diet, physical activity, and smoking, modulate these pathways. In addition, germline mutations may accelerate tumorigenesis like in the Li-Fraumeni syndrome (36). In contrast, 100% of all cancers arising in FA individuals (here termed “FA cancers”) are based on heredity of a defective FA/BRCA pathway of DNA repair. Therefore, genome instability & mutation prominently affects other physiological processes, which can result in the emergence of additional hallmarks, such as polymorphic microbiomes, tumor-promoting inflammation, deregulating cellular metabolism, activating invasion & metastasis, senescent cells, avoiding immune destruction, and non-mutational epigenetic reprogramming. These eight hallmarks of FA cancers (**Fig. 2B**) summarize clinical, cellular and molecular observations in the field (39) (**Box 1**). They overlap largely with the latest version of 14 hallmarks of cancer (40). However, while in the general population the order in which the cancer hallmarks are established, as well as the underlying mechanisms, varies significantly, in FA cancers the hallmark “genome instability & mutation” always emerges first. For example, we illustrate five real cases of tumorigenesis in FA individuals (**Fig. 2C, Box 1**). Note that these also demonstrate that in FA cancers there is variability in the order of appearance of subsequent hallmarks.

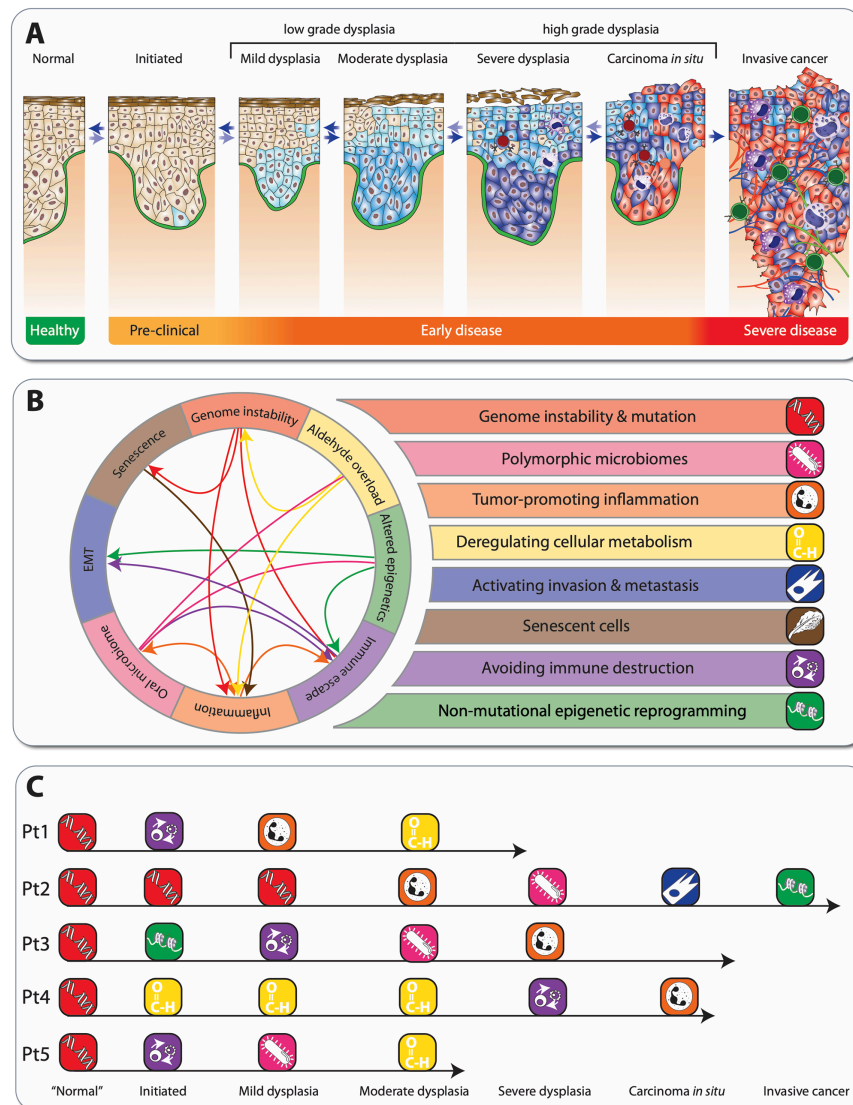


Figure 2: **Hallmarks of cancer in FA.** Depiction of the multi-step tumorigenesis process (**A**). Eight major hallmarks describe the onset and progression of cancer in FA individuals (**B**). In FA, genome instability is the first hallmark to emerge, which influences other hallmarks (arrows). The medical histories of five FA individuals are illustrated as examples such hallmark sequences (**C**).

Modelling FA cancer development

Cancer progression is a dynamical process ranging from early, mostly clinically asymptomatic, to late stages that are difficult to reverse. Each stage is characterized by a particular configuration (emergent behavior) of the tissue, such as infiltration of immune cells, the microbiome, cell cycle speed and the grade of cell dysplasia (**Fig. 3A**). Disease onset and aggravation emerge from the dynamic interplay between hallmarks of SCC in FA (**Fig. 2B**). The hallmarks are connected mechanistically by complex, multi-level regulatory networks that under homeostatic conditions maintain a healthy phenotype (**Fig. 3B**). At the cellular level, the stratified epithelium underlying the epithelial barrier function interplays with the oral microbiome and the infiltration

Principles of mechanistic modelling

The construction, calibration and validation of mechanistic mathematical models require a constant dialogue between the mathematical/computational implementation and analysis of the model, with the experimental and clinical data. As a first step, all available relevant empirical data describing the biological phenomenon are gathered from the literature together with in-house measurements from clinical, animal and *in vitro* studies. These data can be complemented using publicly available databases, such as String-DB (41), KEGG (8) and the Human Cell Atlas (42). In principle, all types of data, from low- to high-throughput, single-timepoint to dynamic and mean-field to single-patient-single-cell resolution, can be used (**Fig. 4A**).

After cleaning, filtering and pre-processing the data using standard statistical and bioinformatic methods (43, 44), regulatory networks at the cellular- and tissue-level scales are assembled from the data and visually represented using the Systems Biology Markup Language (SBML) (45). The networks are translated into mechanistic mathematical models using a chosen modeling formalism that depends on the scale and resolution of the data. Multi-scale mathematical models can be modularly constructed and assembled using a variety of formalisms from nonlinear dynamical systems (46), including Boolean networks for genetic regulatory networks, systems of nonlinear ordinary differential equations (ODEs) for signaling networks and interactions between cell types in the tissue, compartmental differential equations or agent-based models to explicitly model spatial coupling between different cell types, stochastic differential equations to simulate population-level distributions of cell markers and delay-differential equations for multi-level regulatory loops that control tissue homeostasis. In general, when constructing a model for a specific disease affecting a particular tissue (*e.g.*, FA-related SCC in the oral mucosa), a good starting point is to identify mathematical models previously proposed and experimentally validated, which can then be extended and adapted to reproduce specific experimental and clinical observations. For example, mathematical models of loss of epithelial homeostasis, originally developed for atopic diseases (47), could be extended to include SCC in FA-specific hallmarks. The quantitative models are then calibrated with a training set of experimental data, using global parametric optimization algorithms (48-50). This allows one to adapt the mathematical models to specific experimental conditions, as well as to integrate scattered experimental data into a formal and coherent framework that articulates all the individual regulatory players that are typically described in isolation, to understand how they give rise to different clinical manifestations. Finally, model validation is achieved by ensuring that the model can reproduce an additional set of empirical data (the validation set) that was not used for the calibration (**Fig. 4B**).

Next, the calibrated and validated models are analyzed extensively. For this, the nominal conditions, *i.e.*, the calibrated model, are perturbed, *e.g.*, by systematically altering the magnitude of the individual regulatory interactions (changing parameter values), or by structurally altering the different regulatory interactions (changing the equations). The output of the model, *e.g.*, the “phenotype”, is then collected. Examples of such perturbations-to-response mappings are robustness analysis, sensitivity analysis and bifurcation analysis. A robustness analysis tells us which fraction of variations results in a given phenotype (51). Sensitivity analysis weights each individual interaction (parameter) in terms of its contribution to the change in the model output (52). In bifurcation analysis, the model output is assessed as one (or more) parameters are gradually changed (53). Abrupt health-to-disease transitions mathematically correspond to qualitative changes in the model, and they occur typically at bifurcations. The value(s) of the parameter(s) at which such a bifurcation occurs, known as a bifurcation (set), is particularly interesting because it represents the critical threshold of a perturbation that can be tolerated by the regulatory structure. Furthermore, the appearance of these bifurcation sets can often be anticipated by early warning signals, such as an increased variance in recovery times of the system (54). Thus, early warning signals of bifurcating systems can be used to improve early detection strategies for abrupt disease transitions (**Fig. 4C**).

Finally, once the mathematical model has been exhaustively calibrated, validated and analyzed, it can be used for designing and optimizing personalized treatment strategies that consider the specific disease stage and patient characteristics (55). For this, one starts by identifying the potential targets of the intervention strategy, *e.g.*, growth hormone receptor inhibitors that reduce excessive proliferation of malignant cells, and

modifying the model accordingly. Next, analysis of the extended model, *i.e.*, the system without treatments plus the dynamics of the treatments, can be performed (**Fig. 4C**). For example, one can use bifurcation analysis to systematically explore how a given treatment affects the overall virtual population of patients (*e.g.*, considering the natural variations that occur in a population due to polymorphisms) by looking at how the bifurcation sets shift under a specific treatment, such as cetuximab (an epidermal growth factor receptor inhibitor). With a similar line of reasoning, it is possible to use bifurcation analysis to explore how different treatment combinations affect the phenotype of a given patient. Here, again, it is particularly relevant to look for bifurcation sets that separate the qualitatively different clinical phenotypes because these curves correspond to all the minimal treatment combinations that can effectively trigger a disease-to-health transition (56). Besides bifurcation analysis, other techniques, such as model predictive control (57), can also be applied to maximize treatment efficiency while minimizing duration, dosing, and negative side effects (**Fig. 4D**). In all stages of this modelling pipeline, model predictions must be verified by comparing them to empirical data.

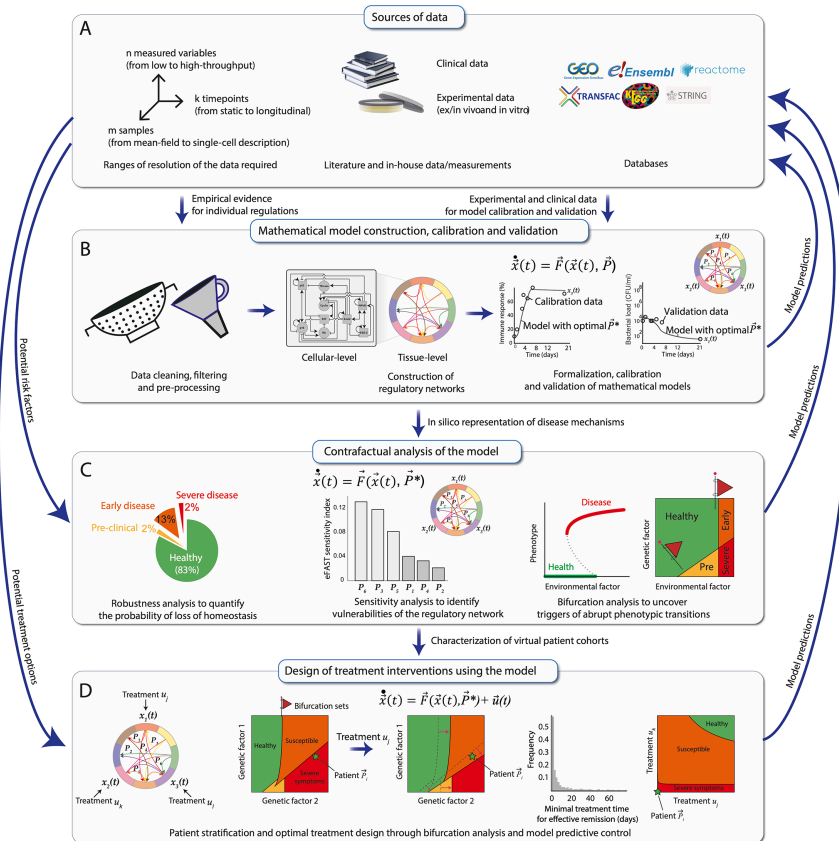


Figure 4: Principles of mechanistic modelling. The preconditions for performing mechanistic modelling of diseases, such as SCC in FA individuals, are clinical data derived directly from patients and experimental data obtained either *in vitro* from patient samples or *in vivo*. Additional data can be obtained from public databases and repositories (A). A mathematical model of regulatory networks is constructed after filtering and processing of the data on the level of either cells or whole tissues. The model is formalized using non-linear dynamical systems, which are calibrated and validated with the data (B). Next, the model is analyzed for robustness sensitivity, and ability to reflect abrupt phenotypic changes in response to perturbations, and resulting model predictions, e.g., the map between a risk factor and the disease severity, are compared with real-world data (C). The mathematical model can be used to systematically explore different types of treatment options for the SCC of an FA individual. Once validated, the most effective predicted treatments may then be applied to patients (D).

As a simple but illustrative example of the model construction process, we present a small biochemical model of the first few steps of ICL detection and repair by the FA/BRCA pathway (58, 59). During DNA replication, the presence of an ICL causes the replication fork to stall. This stressed fork is detected by the protein FANCM, which binds to the branched DNA structure caused by replication fork arrest and recruits the FA “core complex” to the damage site (Fig. 5A). The FA core complex is composed of three protein subcomplexes, each of which is composed of three proteins (Fig. 5B): AG20 is a complex of the proteins FANCA, FANCG, and FAAP20 (FA core complex associated protein 20); BL100 is a complex of FANCB, FANCL, and FAAP100; and CEF is a complex of FANCC, FANCE, and FANCF. For simplicity, we have chosen to represent in the model each of these subcomplexes as distinct molecular species that reversibly bind to each other to form the FA core complex. The complete computational model includes these subcomplex binding reactions, the binding of FANCM to the ICL, and binding of the FA core complex

to ICL-bound FANCM (**Fig. 5C**). By defining parameter values for the rates of each of these individual reactions, the model becomes executable, *i.e.*, the populations (or concentrations) of the constituent species can be simulated over time (**Fig. 5D**). In this way, hypotheses regarding the effects of mutations (changing parameter values) and/or adding external perturbagens (*e.g.*, drugs) can be explored *in silico*.

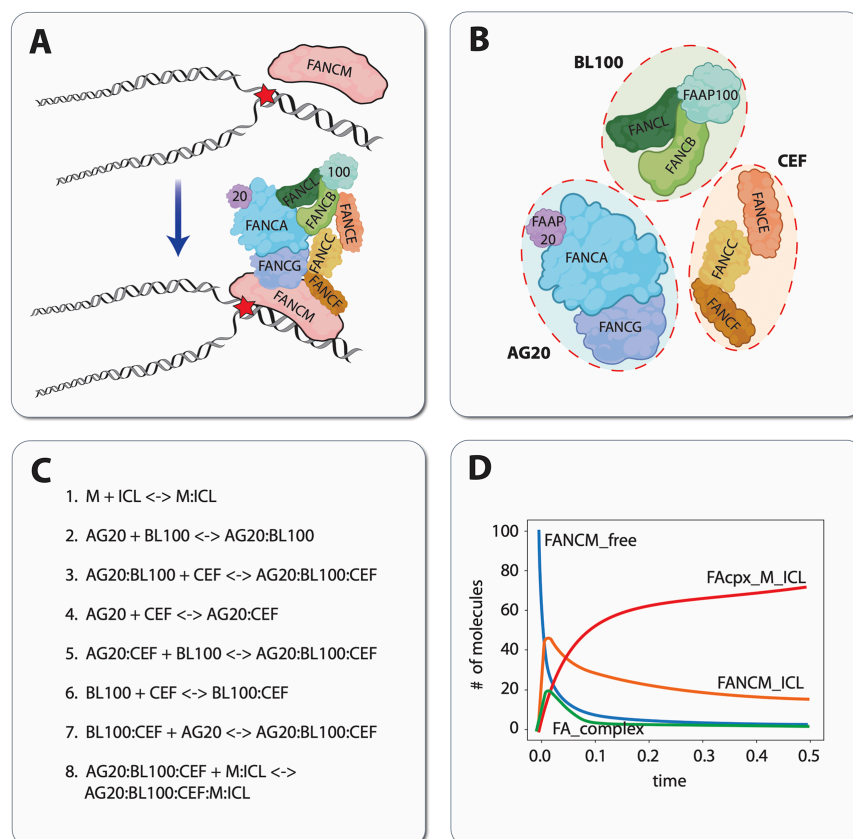


Figure 5: Illustrative example of mechanistic modeling in FA. Schematic of the first two steps of ICL detection and repair, involving binding of the protein FANCM to the DNA, followed by recruitment of the FA core complex (**A**). For demonstration purposes, we have chosen to model the protein subcomplexes AG20, BL100, and CEF as independent species that reversibly bind to form the FA core complex. Note that this choice of model resolution is at the modeler's discretion, *i.e.*, alternatively, each protein (FANCA, FANCG, FAAP20, etc.) comprising the subcomplexes could have been modeled as independent species (**B**). Eight reversible biochemical interactions (16 reactions total) can describe the ICL detection and FA core complex recruitment process (**C**). *In silico* time courses for different molecular species can be obtained by setting values of the binding/unbinding rate constants (all set to 1 in this case) and numerically integrating the resultant set of coupled ODEs (**D**). 'FANCM_free': unbound FANCM; 'FANCM_ICL': FANCM bound to the DNA around the ICL; 'FA_complex': FA core complex composed of AG20, BL100, and CEF that is not bound to FANCM; 'FAcp_M_ICL': FA core complex bound to FANCM, which is bound to the ICL.

Experimental data for FA model development and validation: Example of multi-omic analysis of a SCC lesion from one FA individual

The tumorigenesis process (**Fig. 6A**) results in a heterogeneous composition of tumors, *i.e.*, each tumor contains cells in various stages of the transformation process to aggressively metastasizing cells. Importantly, tumors are not only composed of malignant proliferating cells, but also by multiple cell types, thus making the tumor mass a complex ecosystem that includes immune cells of multiple types (B cells, T cells, macrophages, etc.), tumor-associated fibroblasts, endothelial cells (60) and even microbes, including bacteria and fungi (61). At the same time, a tumor is not only composed of cells, but also by extracellular matrix and secreted factors that can signal messages among cells (60). If malignant tumors from individuals with FA are to be characterized and this information used for accurate model building, all these factors must be accounted for. In this respect, high throughput multi-omics technologies can leverage the components of the tumor of interest, generating data in multiple modalities that need to be integrated and potentially exploited for discovering novel biomarkers and therapeutic targets for individuals with FA.

Of note, classical DNA sequencing, RNA sequencing (RNA-seq), and protein detection technologies are not able to deconvolute and deconstruct the above-mentioned complex composition of a given tumor, since they use the bulk content of the tumor or tissue and are, therefore, constrained to detect the mean expression of molecules, or the presence of a predominant DNA sequence, thus losing information of minor cell populations or incipient emergent cellular clones (62). However, we are witnessing the appearance, development, and refinement of multiple technologies with the capacity to resolve the cellular heterogeneity of tumors. Among these technologies, one of the most popular is single-cell RNA-seq (scRNAseq), which has given rise to a growing number of datasets from liquid and solid tumors (after tumor dissociation), as well as healthy tissue, leading to a compendium of single-cell-resolution gene expression atlases of multiple tissues and organs (63). Although scRNAseq is a technology that has revolutionized the resolution at which we analyze cell populations and tissues, it still lacks a critical component, *i.e.*, preservation of tissue architecture in its original context (62).

In the context of FA cancer, we are interested in the implementation of technologies that, in a multi-omics fashion, will generate single-cell-resolution data but will prevent tissue disaggregation and, therefore, maintain tissue architecture. The latter implies the preservation of cellular neighborhoods and cell-cell interactions, which are lost when the tissue is disaggregated. These technologies are known as spatial omics and include spatial transcriptomics, spatial proteomics, and spatial genomics, which combine molecular characterization with spatial resolution (64). The aim of these spatial resolution technologies is to assign omics information to spatial locations in the tissues, reaching cellular and subcellular resolution. Spatial genomics assigns DNA sequencing information, including copy-number variants and somatic mutations; spatial transcriptomics provides information on the number of transcripts of a certain gene per region; and spatial proteomics provides relative amounts of protein concentrations (64). The data obtained by these multi-omics technologies are highly dimensional in nature and require potent computational tools for their analysis. Although intense research is underway for improving all spatial omics technologies, the most developed are spatial transcriptomics and spatial proteomics. These technologies will allow for the detection and quantification of cell populations of interest, the discovery of new cell populations, the comparison of the abundance of cell populations across the carcinogenic progression and the quantitative and qualitative description of infiltrating immune cells (65, 66). These technologies have the capacity to compensate for the lack of resolution of bulk sequencing analyses, which has hampered the detection of premalignant clones at early stages in FA (67).

Here, we use as an example a hypopharynx cancer from a 41-year-old woman with FA. The hematoxylin- and eosin-stained tumor sample shows multistage carcinogenesis, ranging from low-grade dysplasia (yellow) to high-grade dysplasia (orange) and invasive carcinoma (red) (**Fig. 6B**). This type of formalin-fixed paraffin-embedded (FFPE) sample can be used for exploration and information retrieval using one or multiple of the multi-omics technologies discussed here (see also **Box 2**). If, for example, tissue-cyclic immunofluores-

cence (t-CycIF) is used, multiple sequential pictures of the tissue stained with fluorescent antibodies will be acquired and stitched. The composite image that is generated must first be segmented using artificial intelligence-based programs, such as ASHLAR (66), which recognize every cell nucleus and apply single-cell-level segmentation of the tumor (**Fig. 6C, upper left panel**). For every cell, we can feature-extract the expression of every marker of interest and proceed to non-supervised machine learning-based algorithms, such as uniform manifold approximation and projection (UMAP) (**Fig. 6C, upper right panel**), which generate clusters of cells based on the similarity of their expressed markers. This allows the separate visualization of cell populations (68), such as cancer and immune tumor-infiltrating cells. After feature extraction, we can explore the expression of markers of interest in every tumor population or across the tumor progression, for example the proportion of proliferating cells (**Fig. 6C, lower left panel**), the relative expression of p53 (**Fig. 6C, lower right panel**) or other markers of interest.

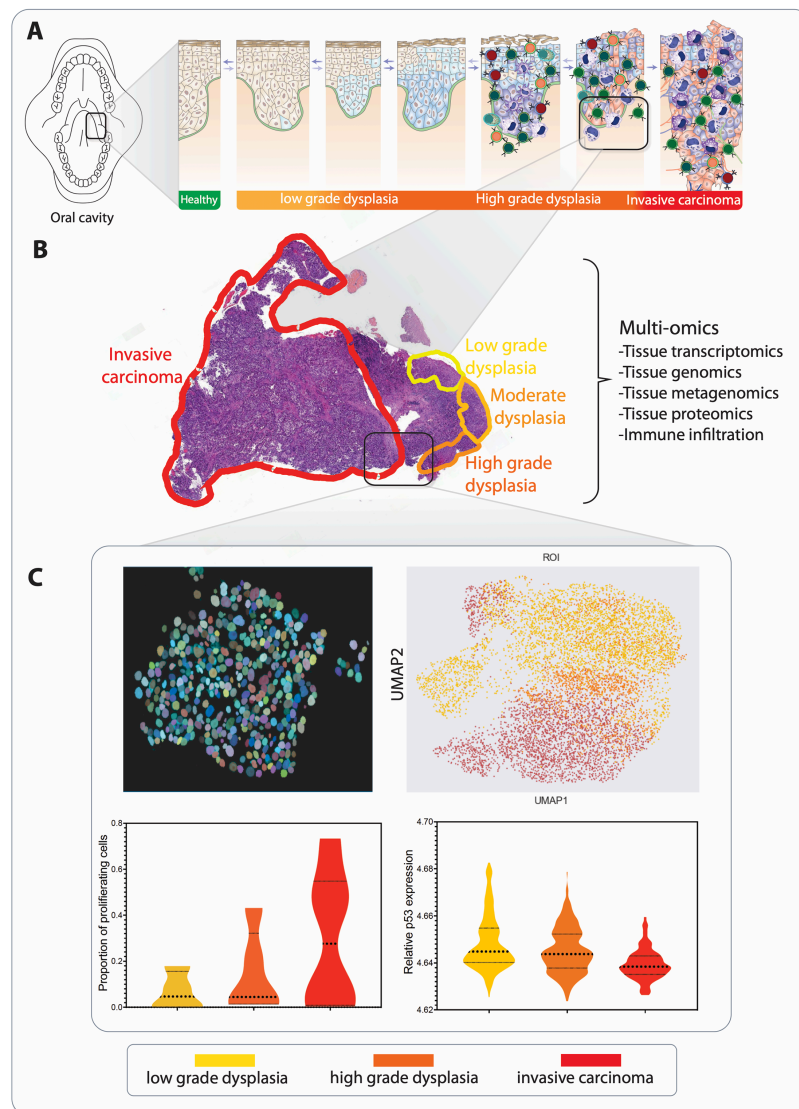


Figure 6: An SCC lesion from an FA patient analyzed by tCycIF imaging. The location within the oral cavity and stage of a hypopharynx tumor from a 41-year-old woman with FA is indicated (**A**). The hematoxylin- and eosin-stained tumor sample shows multistage carcinogenesis, ranging from low-grade dysplasia (yellow) *via* high-grade dysplasia (orange) to invasive carcinoma (red) (**B**). Multi-omic analysis of the tumor includes tissue transcriptomics, genomics, proteomics, and metagenomics for detection of pathogens inhabiting the tumor. Machine learning-based methods are applied in combination with single-cell level segmentation of the tumors and delineation of tumor neighborhoods (**C**). In this inset of invasive carcinoma, every circle represents an individual tumor cell, and its color indicates its stage within the multi-step tumorigenesis process. The data produced from tumor multi-omics can be processed using non-supervised machine learning algorithms, such as UMAP, for detection of commonalities and divergencies in the tumor sections from multiple patients, and information on markers expression can be extracted from every cell so as to generate graphs for comparing markers expression across the carcinogenic progression.

Toward ‘digital twins’ of FA individuals

Studies on SCC prevention in FA are limited by the small number of individuals with the condition, who are spread around the world. In addition to regular histopathological diagnosis of oral cancer development and cytology-based screening methods (32), reliable molecular markers are limited. Moreover, the scarcity of genotype-phenotype associations in FA makes it highly likely that each patient will respond in an individual way to drug treatments and/or lifestyle changes. Since robust predictive *in vitro* and *in vivo* FA models are lacking, drug screening and testing cannot be generalized for all FA individuals. For example, *in vitro* analysis of radiation sensitivity of fibroblasts from FA individuals does not correlate with the clinical response of the same patient to radiotherapy (69, 70) and the amount of chromosomal breaks found in lymphocyte cultures does not correlate with the severity of the disease, *e.g.*, bone marrow function of the individual. These issues motivate the effort to create multi-level, dynamical computational models of FA that can aid clinicians in tailoring therapies to each specific FA patient. Models of this type have been termed “medical digital twins” (71, 72).

Although a consensus definition of a medical digital twin does not yet exist, the concept of a digital twin is common in engineering disciplines (73). Often referred to as “industrial digital twins,” these models are computational replicas of complex devices or processes, such as jet engines or wind turbines, that are used to diagnose technical problems and guide interventions. Industrial digital twins are typically composed of multiple, interconnected computational models of the constituent components of an engineered system. Critically, this integrated “template” model of the base processes of the engineered system in question is subsequently tuned, or “calibrated,” to a specific instance of that system, *e.g.*, a particular jet engine, using performance data collected from sensors *in real time*. It is this “twinning” process, involving consistent feedback from real-time data streams, that differentiates a multi-level, computational model of a dynamical system from a true digital twin (74). Construction of digital twins for medical and clinical applications has been receiving increased interest lately (71, 72). However, biological systems are far more complex than engineered systems, making their practical implementation much more challenging. Nevertheless, there have been successful applications of medical digital twins for the treatment of type 1 diabetes (75) and pediatric cardiac patients (76). Furthermore, it is important to note that medical digital twins differ from alternative approaches gaining popularity in biomedical sciences, such as statistical and machine-learning models (77), in that they are based on a mechanistic understanding of the underlying biological system. As such, they are not constrained by the confines of the experimental data on which they are constructed, which in FA is sparse.

The utility of a FA medical digital twin will be to aid clinicians in determining best courses of action for individual patients in both the prevention and treatment of malignant tumors. The template model for an FA medical digital twin will comprise the biological processes mentioned previously, including microbiome interactions, DNA damage sensing and repair, EMT, cell cycle progression, and cell death, among others (**Fig. 3**). Calibrating the template model to individual FA patients will be challenging and require collecting spatially resolved, single-cell resolution multi-omics data, epigenome profiling, and metagenomics of the oral microbiome, from patients at regular intervals, *e.g.*, every three months in accordance with clinical care guidelines for FA individuals with potentially premalignant lesions in the epithelial tissue. Additional patient data, such as blood draws, genetic screens, and oral swabs, together with standard data from electronic health records, can also be integrated into the calibration data stream. Once the model is personalized in this way, it will be possible to test, *in silico*, numerous preventive and/or therapeutic options before applying them clinically to FA patients. Furthermore, medical digital twins should be flexible and extensible, able to grow in precision and predictive power as new knowledge is accrued and experimental data sets are generated. In this way, the FA digital twin will develop together with the patient and their clinician, ultimately forecasting, with high accuracy and precision, responses to novel personalized interventions and therapies.

Conclusions

For FA cancers, especially oral SCC, the use of multi-level dynamic mechanistic modelling provides a new perspective on early-stage diagnosis and decision support for the treatment of this rare disease. Such an approach is critical, since classical statistical models, using case studies and controls, cannot be applied due to the dearth of large patient groups. As such, we aim to build accurate computational models of tumorigenesis in a limited but representative number of FA patients. These mechanistic models will utilize pre-existing public knowledge on biochemical and regulatory pathways together with our knowledge of the life and disease course of more than 750 FA individuals, which will be essential for distinguishing the tumorigenesis process of FA cancer from that of regular cancer. In this way, our mechanistic models of FA cancer will take specific characteristics of this rare disease into account. Using longitudinal information about the lifestyles of FA individuals over years, together with multi-omics data at the genomic, transcriptomic, and proteomic levels, will lead to the construction of individual-specific models, or digital twins, that can be used to develop personalized treatment options. This approach has the potential to revolutionize the way FA individuals are treated clinically.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

All authors wrote and edited the manuscript and provided figures.

Funding

EV holds fundings from the Fanconi Anemia Research Fund (FARF), the Deutsche-Fanconi-Anämie-Hilfe e.V. and the Förderverein zugunsten krebskranker Kinder Krefeld e.V. EDH is funded by the Dirección General de Asuntos del Personal Académico (DGAPA-UNAM), grant PAPIIT UNAM IA207822, and by the Consejo Nacional de Ciencia y Tecnología (CONACyT), grant Ciencia de Frontera 2022 319600. AR is funded by PAPIIT-UNAM project number IA205022 and from CONACyT Ciencia de Frontera 2022, project number 319344. LAH was funded by an NIH/NCI Transition Career Development Award to Promote Diversity (K22-CA237857), the Vanderbilt University Center for Cancer Systems Biology, and the Arkansas Integrative Metabolic Research Center. This publication is part of a project that has received funding from the European Union's Horizon2020 research and innovation program under grant agreement no. 952601 and from the David and Amy Fulton Foundation (both to CC).

Acknowledgments

EV wants to express her deep gratitude to all FA individuals she has met, have been involved in research projects, and donated biological samples and clinical information. Moreover, she wants to thank all FA patient support organizations who contributed largely to the improvement of health care and increased survival in this rare disease. AR thanks Pablo Siliceo for providing part of Figure 6C and Dr. Anna Laury for pathologic annotation of Figure 6B. LAH acknowledges the NIH-sponsored Multiscale Modeling and Viral Pandemics working group and the organizers of the Building Immune Digital Twins workshop hosted at the Institut Pascal, Université Paris-Saclay from May 15–June 2, 2023.

References

1. The Lancet Diabetes E. Spotlight on rare diseases. *The lancet Diabetes & endocrinology*. 2019;7(2):75.
2. Boycott KM, Rath A, Chong JX, Hartley T, Alkuraya FS, Baynam G, et al. International cooperation to enable the diagnosis of all rare genetic diseases. *Am J Hum Genet*. 2017;100(5):695-705.
3. Wang Q, Dhindsa RS, Carss K, Harper AR, Nag A, Tachmazidou I, et al. Rare variant contribution to human disease in 281,104 UK Biobank exomes. *Nature*. 2021;597(7877):527-32.
4. Velleuer E, Carlberg C. Impact of epigenetics on complications of Fanconi anemia: the role of vitamin D-modulated immunity. *Nutrients*. 2020;12(5).
5. Wu ZH. The concept and practice of Fanconi Anemia: from the clinical bedside to the laboratory bench. *Transl Pediatr*. 2013;2(3):112-9.
6. Akobeng AK. Principles of evidence based medicine. *Arch Dis Child*. 2005;90(8):837-40.
7. Wheatley R, Diaz Caballero J, Kapel N, de Winter FHR, Jangir P, Quinn A, et al. Rapid evolution and host immunity drive the rise and fall of carbapenem resistance during an acute *Pseudomonas aeruginosa* infection. *Nature communications*. 2021;12(1):2460.
8. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res*. 2017;45(D1):D353-D61.
9. Pico AR, Kelder T, van Iersel MP, Hanspers K, Conklin BR, Evelo C. WikiPathways: pathway editing for the people. *PLoS Biol*. 2008;6:e184.
10. Himmelstein DS, Baranzini SE. Heterogeneous network edge prediction: a data integration approach to prioritize disease-associated genes. *PLoS computational biology*. 2015;11(7):e1004259.
11. Niarakis A, Helikar T. A practical guide to mechanistic systems modeling in biology using a logic-based approach. *Briefings in bioinformatics*. 2021;22(4).
12. Harris LA, Beik S, Ozawa PMM, Jimenez L, Weaver AM. Modeling heterogeneous tumor growth dynamics and cell-cell interactions at single-cell and cell-population resolution. *Curr Opin Syst Biol*. 2019;17:24-34.
13. Lobitz S, Velleuer E. Guido Fanconi (1892-1979): a jack of all trades. *Nat Rev Cancer*. 2006;6(11):893-8.
14. Fiesco-Roa MO, Giri N, McReynolds LJ, Best AF, Alter BP. Genotype-phenotype associations in Fanconi anemia: a literature review. *Blood Rev*. 2019;37:100589.
15. Alter BP. Inherited bone marrow failure syndromes: considerations pre- and posttransplant. *Blood*. 2017;130(21):2257-64.
16. Dufour C. How I manage patients with Fanconi anaemia. *Br J Haematol*. 2017;178(1):32-47.
17. Alter BP, Giri N, Savage SA, Rosenberg PS. Cancer in the National Cancer Institute inherited bone marrow failure syndrome cohort after fifteen years of follow-up. *Haematologica*. 2018;103(1):30-9.
18. Kutler DI, Singh B, Satagopan J, Batish SD, Berwick M, Giampietro PF, et al. A 20-year perspective on the International Fanconi Anemia Registry (IFAR). *Blood*. 2003;101(4):1249-56.
19. Wang AT, Smogorzewska A. SnapShot: Fanconi anemia and associated proteins. *Cell*. 2015;160(1-2):354-e1.
20. Ameziane N, May P, Haitjema A, van de Vrugt HJ, van Rossum-Fikkert SE, Ristic D, et al. A novel Fanconi anaemia subtype associated with a dominant-negative mutation in RAD51. *Nature communications*. 2015;6:8829.

21. Meetei AR, Levitus M, Xue Y, Medhurst AL, Zwaan M, Ling C, et al. X-linked inheritance of Fanconi anemia complementation group B. *Nat Genet.* 2004;36(11):1219-24.
22. Ceccaldi R, Sarangi P, D'Andrea AD. The Fanconi anaemia pathway: new players and new functions. *Nat Rev Mol Cell Biol.* 2016;17(6):337-49.
23. Tischkowitz MD, Hodgson SV. Fanconi anaemia. *J Med Genet.* 2003;40(1):1-10.
24. Gluckman E. Improving survival for Fanconi anemia patients. *Blood.* 2015;125(24):3676.
25. Bonfim C, Ribeiro L, Nichele S, Bitencourt M, Loth G, Koliski A, et al. Long-term survival, organ function, and malignancy after hematopoietic stem cell transplantation for Fanconi Anemia. *Biol Blood Marrow Transplant.* 2016;22(7):1257-63.
26. Paustian L, Chao MM, Hanenberg H, Schindler D, Neitzel H, Kratz CP, et al. Androgen therapy in Fanconi anemia: A retrospective analysis of 30 years in Germany. *Pediatr Hematol Oncol.* 2016;33(1):5-12.
27. Calado RT, Cle DV. Treatment of inherited bone marrow failure syndromes beyond transplantation. *Hematology Am Soc Hematol Educ Program.* 2017;2017(1):96-101.
28. Rose SR, Kim MO, Korb L, Wilson KA, Ris MD, Eyal O, et al. Oxandrolone for the treatment of bone marrow failure in Fanconi anemia. *Pediatr Blood Cancer.* 2014;61(1):11-9.
29. Scheckenbach K, Morgan M, Filger-Brillinger J, Sandmann M, Strimling B, Scheurlen W, et al. Treatment of the bone marrow failure in Fanconi anemia patients with danazol. *Blood Cells Mol Dis.* 2012;48(2):128-31.
30. Kutler DI, Patel KR, Auerbach AD, Kennedy J, Lach FP, Sanborn E, et al. Natural history and management of Fanconi anemia patients with head and neck cancer: A 10-year follow-up. *Laryngoscope.* 2016;126(4):870-9.
31. Lin J, Kutler DI. Why otolaryngologists need to be aware of Fanconi anemia. *Otolaryngol Clin North Am.* 2013;46(4):567-77.
32. Velleuer E, Dietrich R, Pomjanski N, de Santana Almeida Araujo IK, Silva de Araujo BE, Sroka I, et al. Diagnostic accuracy of brush biopsy-based cytology for the early detection of oral cancer and precursors in Fanconi anemia. *Cancer Cytopathol.* 2020.
33. Howlett NG, Taniguchi T, Olson S, Cox B, Waisfisz Q, De Die-Smulders C, et al. Biallelic inactivation of BRCA2 in Fanconi anemia. *Science.* 2002;297(5581):606-9.
34. Del Valle J, Rofes P, Moreno-Cabrera JM, Lopez-Doriga A, Belhadj S, Vargas-Parra G, et al. Exploring the role of mutations in Fanconi anemia genes in hereditary cancer patients. *Cancers.* 2020;12(4).
35. Pouliot GP, Degar J, Hinze L, Kochupurakkal B, Vo CD, Burns MA, et al. Fanconi-BRCA pathway mutations in childhood T-cell acute lymphoblastic leukemia. *PLoS ONE.* 2019;14(11):e0221288.
36. Carlberg C, Velleuer E. Cancer biology: how science works. Springer Textbook. 2021.
37. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell.* 2000;100:57-70.
38. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Jr., Kinzler KW. Cancer genome landscapes. *Science.* 2013;339(6127):1546-58.
39. Webster ALH, Sanders MA, Patel K, Dietrich R, Noonan RJ, Lach FP, et al. Genomic signature of Fanconi anaemia DNA repair pathway deficiency in cancer. *Nature.* 2022;612(7940):495-502.
40. Hanahan D. Hallmarks of Cancer: New Dimensions. *Cancer Discovery.* 2022;12(1):31-46.
41. Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, et al. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* 2021;49(D1):D605-D12.

42. Regev A, Teichmann SA, Lander ES, Amit I, Benoist C, Birney E, et al. The Human Cell Atlas. *Elife*. 2017;6.
43. Sinha PP. *Bioinformatics with R Cookbook*. Packt Publishing. 2014.
44. Sinha GB. *Fundamentals of Bioinformatics and Computational Biology*. Springer Berlin Heidelberg. 2015.
45. Keating SM, Waltemath D, König M, Zhang F, Dräger A, Chaouiya C, et al. SBML Level 3: an extensible format for the exchange and reuse of biological models. *Molecular systems biology*. 2020;16(8):e9110.
46. de Jong H. Modeling and simulation of genetic regulatory systems: a literature review. *J Comput Biol*. 2002;9(1):67-103.
47. Dominguez-Huttinger E, Christodoulides P, Miyauchi K, Irvine AD, Okada-Hatakeyama M, Kubo M, et al. Mathematical modeling of atopic dermatitis reveals "double-switch" mechanisms underlying 4 common disease phenotypes. *The Journal of allergy and clinical immunology*. 2017;139(6):1861-72 e7.
48. Balsa-Canto E, Banga JR, Egea JA, Fernandez-Villaverde A, de Hijas-Liste GM. Global optimization in systems biology: stochastic methods and their applications. *Adv Exp Med Biol*. 2012;736:409-24.
49. Tsiantis N, Balsa-Canto E, Banga JR. Optimality and identification of dynamic models in systems biology: an inverse optimal control framework. *Bioinformatics*. 2018;34(21):3780.
50. Shockley EM, Vrugt JA, Lopez CF. PyDREAM: high-dimensional parameter inference for biological models in python. *Bioinformatics*. 2018;34(4):695-7.
51. Dominguez-Huttinger E, Boon NJ, Clarke TB, Tanaka RJ. Mathematical Modeling of *Streptococcus pneumoniae* Colonization, Invasive Infection and Treatment. *Front Physiol*. 2017;8:115.
52. Zi Z. Sensitivity analysis approaches applied to systems biology models. *IET systems biology*. 2011;5(6):336-6.
53. Kuznetsov YA. *Elements of applied bifurcation theory*. Springer International Publishing. 2004.
54. Bargaje R, Trachana K, Shelton MN, McGinnis CS, Zhou JX, Chadick C, et al. Cell population structure prior to bifurcation predicts efficiency of directed differentiation in human induced pluripotent cells. *Proc Natl Acad Sci U S A*. 2017;114(9):2271-6.
55. Fey D, Halasz M, Dreidax D, Kennedy SP, Hastings JF, Rauch N, et al. Signaling pathway models as biomarkers: Patient-specific simulations of JNK activity predict the survival of neuroblastoma patients. *Science signaling*. 2015;8(408):ra130.
56. Tanaka G, Dominguez-Huttinger E, Christodoulides P, Aihara K, Tanaka RJ. Bifurcation analysis of a mathematical model of atopic dermatitis to determine patient-specific effects of treatments on dynamic phenotypes. *Journal of theoretical biology*. 2018;448:66-79.
57. Christodoulides P, Hirata Y, Dominguez-Huttinger E, Danby SG, Cork MJ, Williams HC, et al. Computational design of treatment strategies for proactive therapy on atopic dermatitis using optimal control theory. *Philos Trans A Math Phys Eng Sci*. 2017;375(2096).
58. Niraj J, Farkkila A, D'Andrea AD. The Fanconi Anemia Pathway in Cancer. *Annu Rev Cancer Biol*. 2019;3:457-78.
59. Semlow DR, Walter JC. Mechanisms of Vertebrate DNA Interstrand Cross-Link Repair. *Annu Rev Biochem*. 2021;90:107-35.
60. Anderson NM, Simon MC. The tumor microenvironment. *Curr Biol*. 2020;30(16):R921-R5.
61. Zong Z, Zhou F, Zhang L. The fungal mycobiome: a new hallmark of cancer revealed by pan-cancer analyses. *Signal Transduct Target Ther*. 2023;8(1):50.

62. Stark R, Grzelak M, Hadfield J. RNA sequencing: the teenage years. *Nat Rev Genet.* 2019;20(11):631-56.
63. Yuan H, Yan M, Zhang G, Liu W, Deng C, Liao G, et al. CancerSEA: a cancer single-cell state atlas. *Nucleic Acids Res.* 2019;47(D1):D900-D8.
64. Akhoundova D, Rubin MA. Clinical application of advanced multi-omics tumor profiling: Shaping precision oncology of the future. *Cancer cell.* 2022;40(9):920-38.
65. Karimi E, Yu MW, Maritan SM, Perus LJM, Rezanejad M, Sorin M, et al. Single-cell spatial immune landscapes of primary and metastatic brain tumours. *Nature.* 2023;614(7948):555-63.
66. Pelka K, Hofree M, Chen JH, Sarkizova S, Pirl JD, Jorgji V, et al. Spatially organized multicellular immune hubs in human colorectal cancer. *Cell.* 2021;184(18):4734-52 e20.
67. Rodriguez A, Zhang K, Farkkila A, Filiatrault J, Yang C, Velazquez M, et al. MYC Promotes Bone Marrow Stem Cell Dysfunction in Fanconi Anemia. *Cell stem cell.* 2021;28(1):33-47 e8.
68. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol.* 2018.
69. Marcou Y, D'Andrea A, Jeggo PA, Plowman PN. Normal cellular radiosensitivity in an adult Fanconi anaemia patient with marked clinical radiosensitivity. *Radiother Oncol.* 2001;60(1):75-9.
70. Alter BP. Radiosensitivity in Fanconi's anemia patients. *Radiother Oncol.* 2002;62(3):345-7.
71. Laubenbacher R, Sluka JP, Glazier JA. Using digital twins in viral infection. *Science.* 2021;371(6534):1105-6.
72. Masison J, Beezley J, Mei Y, Ribeiro H, Knapp AC, Sordo Vieira L, et al. A modular computational framework for medical digital twins. *Proc Natl Acad Sci U S A.* 2021;118(20).
73. Tao F, Qi Q. Make more digital twins. *Nature.* 2019;573(7775):490-1.
74. An G. Specialty Grand Challenge: What it Will Take to Cross the Valley of Death: Translational Systems Biology, "True" Precision Medicine, Medical Digital Twins, Artificial Intelligence and In Silico Clinical Trials. *Frontiers in Systems Biology.* 2022;2.
75. Kovatchev B. A century of diabetes technology: signals, models, and artificial pancreas control. *Trends Endocrinol Metab.* 2019;30(7):432-44.
76. Shang JK, Esmaily M, Verma A, Reinhartz O, Figliola RS, Hsia TY, et al. Patient-specific multiscale modeling of the assisted bidirectional glenn. *Ann Thorac Surg.* 2019;107(4):1232-9.
77. Swanson K, Wu E, Zhang A, Alizadeh AA, Zou J. From patterns to patients: Advances in clinical machine learning for cancer diagnosis, prognosis, and treatment. *Cell.* 2023;186(8):1772-91.
78. Marx V. Method of the Year: spatially resolved transcriptomics. *Nature methods.* 2021;18(1):9-14.
79. Lewis SM, Asselin-Labat ML, Nguyen Q, Berthelet J, Tan X, Wimmer VC, et al. Spatial omics and multiplexed imaging to explore cancer biology. *Nature methods.* 2021;18(9):997-1012.
80. Maniatis S, Petrescu J, Phatnani H. Spatially resolved transcriptomics and its applications in cancer. *Curr Opin Genet Dev.* 2021;66:70-7.
81. Stahl PL, Salmen F, Vickovic S, Lundmark A, Navarro JF, Magnusson J, et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science.* 2016;353(6294):78-82.
82. Merritt CR, Ong GT, Church SE, Barker K, Danaher P, Geiss G, et al. Multiplex digital spatial profiling of proteins and RNA in fixed tissue. *Nat Biotechnol.* 2020;38(5):586-99.

83. Nirmal AJ, Maliga Z, Vallius T, Quattrochi B, Chen AA, Jacobson CA, et al. The Spatial Landscape of Progression and Immunoediting in Primary Melanoma at Single-Cell Resolution. *Cancer Discov.* 2022;12(6):1518-41.
84. Giesen C, Wang HA, Schapiro D, Zivanovic N, Jacobs A, Hattendorf B, et al. Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nature methods.* 2014;11(4):417-22.
85. Hernandez S, Lazcano R, Serrano A, Powell S, Kostousov L, Mehta J, et al. Challenges and Opportunities for Immunoprofiling Using a Spatial High-Plex Technology: The NanoString GeoMx(©) Digital Spatial Profiler. *Frontiers in oncology.* 2022;12:890410.
86. Tang L. Spatially resolved DNA sequencing. *Nature methods.* 2022;19(2):139.
87. Zhao T, Chiang ZD, Morriss JW, LaFave LM, Murray EM, Del Priore I, et al. Spatial genomics enables multi-modal study of clonal heterogeneity in tissues. *Nature.* 2022;601(7891):85-91.

Box 1: Real-world FA patient examples

Patient 1: Diagnosed with FA at age 4 due to bone marrow failure and was transplanted with the bone marrow of his sister. Clinically, no signs for graft versus host disease (GvHD) had been observed but severe viral reactivation complicated the clinical course. He was treated both with anti-viral and immunosuppressive medication. Patient started to drink alcohol at a social occasion at age 18. A visible lesion developed at age 22 at the gingiva and was biopsied revealing moderate dysplasia. Since then, the patient stopped drinking.

Patient 2: Diagnosed with FA at age 6. Additionally, he was a carrier of an inherited mutation in the *APC* (*APC* regulator of WNT signaling pathway) gene. At age 8, the patient needed hematopoietic stem cell transplantation due to bone marrow failure. Unfortunately, he developed severe GvHD. At age 16, he was diagnosed with oral candida infection in a lesion at the gingiva. Due to persistence of the lesion, it was biopsied and the diagnosis of SCC was made. After local excision with clear margins, the patient developed three months later a local soft tissue metastasis. Further treatment including radiation and CD274 blockage were not able to save the patient and he deceased at age 19.

Patient 3: Diagnosed with FA at age 6 due to bone marrow failure. She was treated with anabolic steroids after a period of transfusions and severe infection. The treatment brought the blood counts up but due to the development of clonal hematopoiesis with pre-leukemia, she was transplanted at age 16. At age 21, an oral lesion at the tongue developed and the diagnosis of a candida infection was made. However, after initial treatment, the lesion came back showing signs of inflammation. Due to persistence of the patient, a biopsy was done at age 23 and a severe dysplasia was diagnosed.

Patient 4: Started social drinking and smoking at least one pack a day at age of 16. At age 20, a small lesion at the right side of the tongue was noticed by the patient. Clinical diagnosis of a local inflammation was made and the patient was treated over two months with immune suppression. Because there was no clinical benefit from the treatment, the patient stopped the medication on their own. Due to growing of the lesion and development of pain, eventually a biopsy was performed, confirming the diagnosis of T1 stage SCC at age 21. Due to the unusual age at presentation, investigations revealed the underlying FA diagnosis. Patient was treated with local excision.

Patient 5: Diagnosed with FA at age 8 and directly transplanted. Mild GvHD was clinically present. At age 28, an oral lesion at the gingiva developed. An infection with candida was diagnosed but the patient did not get any further treatment. At age 29, the lesion was biopsied due to increase in size. Histologically, a high-grade dysplasia and an invasive candida infection were seen.

Box 2: Multi-omic analysis methods

Tissue spatial transcriptomics

Tissue spatial transcriptomics allows the characterization of gene expression profiles keeping the tissue's spatial architecture intact. Multiple techniques have been developed for spatial transcriptomics, mainly based on *in situ* hybridization, *in situ* capturing, *in situ* sequencing or microdissection (78).

Fluorescent *in situ* hybridization (FISH)-based methods exploit the hybridization of fluorescent-labeled RNA-targeting probes with pre-defined transcripts of interest, followed by imaging, visualization, and quantification, which however is limited to the simultaneous detection of a small number of transcripts. Higher efficiency in mRNA detection has been reached with the usage of array-based *in situ* capturing methods. These arrays have attached barcoded oligonucleotides that capture, through complementarity, the mRNAs present in the sample. Capture is followed by reverse transcription to cDNA and NGS, allowing the detection of more than 10,000 targets (79, 80). The widely used Visium technology is an example of this approach (Visium spatial gene expression, 10X Genomics) (81).

Recent technologies allow to explore the transcriptome of specific regions of interest in FFPE samples through microdissection. The GeoMx Digital Spatial Profiler by Nanostring allows *in situ* capture of mRNAs using fluorescent-tagged RNA probes, which are linked to UV-photocleavable DNA oligonucleotides of known sequence. The fluorescent-tagged RNA probes are also known as imaging reagents since they will generate a fluorescent image that allows tissue visualization of regions where a specific mRNA is expressed. Once the investigator selects their regions of interest, these areas are exposed to UV light that cleaves the DNA tags in a region-specific manner. This releases indexing oligos that are collected via microcapillary aspiration and dispensed into a microplate and subject to Nanostring mediated counting, or NGS (82). The RNA from FFPE fixed samples very commonly suffers degradation. However, Visium and GeoMx technologies can retrieve good amounts of information from these tissue samples.

Tissue spatial proteomics

The most popular methods of tissue spatial proteomics have the advantage that FFPE samples can be used and, therefore, precious pathological archives can be studied. Strategies for exploring spatial proteomics are based on (i) immunofluorescence, (ii) imaging mass cytometry by time of flight and (iii) sequencing (79). Tissue cyclic immunofluorescence (tCycIF) is an immunofluorescence-based strategy. tCycIF uses FFPE tumor and tissue specimens mounted on glass slides that undergo staining cycles. In every cycle, the specimens are stained with fluorochrome-conjugated antibodies and imaged, followed by chemical inactivation of fluorochromes after each round of immunofluorescence (83). Conventional wide-field, confocal or super-resolution microscopes can be used for image acquisition. After multiple rounds of imaging, a final high-dimensional representation of all the images is assembled into a unique image using computational strategies. The final high-dimension image can be segmented into all individual cells composing the tissue to give single-cell resolution. Neighborhood analysis can also be performed to quantify cell-cell interactions. Of note, tCycIF does not require proprietary reagents, is robust and is a more economical option compared to other spatial proteomics strategies.

CyTOF is a mass spectrometry-based method. In this technology, cellular proteins are detected using antibodies that are conjugated to isotopes from the lanthanide series of rare metals. The sample is imaged using the Hyperion Imaging SystemTM, where these metal-tagged antibodies are laser ablated from regions of interest in the tissue and each ionized metal tag is detected based on differences in their mass instead of the wavelength emitted by a fluorochrome. This technology eliminates the autofluorescence inherent to biological specimens, since the rare metal tags with which the antibodies are conjugated are not present in cells. Also, compensation or background elimination is not needed, since there is no overlap among the signal produced by the ionized metals. In this technology, FFPE samples can be stained with an entire panel of

multiple antibodies in a single scanning round without the need for multiple staining and washing cycles. The image is analyzed using a proprietary software package (84).

Finally, GeoMx Digital Spatial Profiler by Nanostring can be adapted for detection of proteins instead of transcripts (described above). In this setting, the FFPE tissues are immunostained with UV-photocleavable oligonucleotide-labeled antibodies. The spatial location of proteins is again achieved by exposure of the region of interest to UV light that photocleaves the oligos, followed by retrieval of the oligos and sequencing. This provides an average count of oligonucleotides in every region of interest (82, 85).

Tissue spatial genomics

Technologies for spatial resolution of the genome that can preserve tissue architecture are less well developed. Nonetheless, using spatially resolved DNA sequencing will finally deliver information on the process of clonal evolution of solid tumors and provide a timeframe for when a specific mutation appeared. FFPE samples are especially problematic since DNA is very commonly degraded in these specimens (86).

Slide-DNA-seq is one new technology that works with cryosectioned intact tissues. Slide-DNA-seq uses cover slip arrays coated with 10 μ m DNA-barcoded polystyrene beads, each containing a unique DNA barcode corresponding to its spatial location in the cover slip. This is meant to provide spatial indexing. Then, a 10- μ m-thick fresh-frozen tissue section is placed onto the barcoded bead array, treated with HCl for histones removal and treated with the transposase Tn5 to generate DNA fragments that will be flanked with sequencing Illumina adapters. The barcodes are photocleaved from the beads and the resulting DNA sequencing library is amplified by PCR (86, 87).