ITMO Cancer Experts' Roadmap - Strategic directions

The field of cancer has radically changed due to a number of paradigm shifts in the last years.

From a molecular perspective, major advances due to large-scale, high-throughput and cell imaging techniques allowing a very large number of variables to be analysed simultaneously, applied to a growing number of tumour types, have expanded the classification of cancers. Hence, the stratification of cancer by means of "omic" technologies (genomics, epigenomics, transcriptomics, proteomics, metabolomics, etc.) is opening up a new dimension to shed light on its complexity; it can offer and predict the validity of therapeutic approaches, analyse and understand resistance, and identify biomarkers for diagnosis, theranostics and disease progression. Thus, the impact of technology on health is now a determining factor in cancer. Sequencing, the numerous technologies dedicated to biomarkers, and the impact of multimodal imaging (MRI, PET, ultrasound) are examples of the development of instruments which offer robust and reproducible ways of elucidating the mechanisms of oncogenesis and reaching diagnosis and successful therapy assessment and disease prognosis. In terms of treatment, bioproduction processes, cellular engineering, nanotechnologies, advances in minimally or non-invasive surgery, radiotherapy and its new variations (proton or carbon ion therapy) and new physical methods for cancer treatment, such as ultrasound, are opening up major prospects for improving the management of cancer patients.

These unfolding paradigm shifts are already being translated into treatment with spectacular results, for certain types of cancer, owing to immunotherapy or targeted therapy, insight into the action mechanisms of the various treatments, and the ability to predict their effects. Clinical research has access to the most advanced methodological instruments and is supported by competent imaging and biological platforms. Research in the field of human and social sciences has made it possible to place the patient at the centre of cancer research, and to highlight the real need for high-quality research.

The central role of the immune system in controlling pre-cancerous cells has been demonstrated and has led to the emergence of new areas of research to unleash the immune system imprisoned by cancer cells and strengthen immune surveillance to prevent disease. Due to the demonstration of tumour heterogeneity, the clonal structure of cancer cell populations, cell plasticity, and the increasingly robust arguments concerning dormancy, together with the existence of cancerous stem cells initiating cancer responsible for relapse or metastases, tumours are now seen in a different light, no longer as monolithic, uniform groups of cells, but rather as developing ecosystems, sensitive to the surrounding conditions and selection pressure.

Research in epidemiology, identifying combinatorial exposures (exposome), predisposition genes, or pre-neoplastic conditions, has resulted in preventive procedures based on scientific evidence, allowing measures to be envisaged to support and monitor subjects at risk. Secondary prevention (prevention of relapse and/or secondary cancer) owing to individual monitoring, based on more sensitive, specific and less harmful imaging or laboratory techniques (for example: detection and analysis of circulating cancer cells or circulating tumour DNA), is opening up extremely promising avenues with the aim to stay one step ahead of the emergence or resurgence of the disease. Lastly, research in human and social sciences, and interventional research highlights the shared responsibility of various risk factors, the need to optimise monitoring, document the efficacy of screening programmes, and support patients and their families faced with the disease and at the end of life.

While this overview is extremely positive, with a growing number of patients living with cancer and some cures achieved, the new knowledge acquired has shed light on the extraordinary complexity of cancer, necessitating the continued exploration of numerous new opportunities brought to light through fundamental research. This observation also highlights the technological limitations and the need for an ever-growing multidisciplinary approach to cancer research. The ability to identify and validate new parameters among the vast data has become indispensable. The ultimate holy grail is the in-depth characterisation of the biosignature of each tumour in space (its environment) and time (throughout its evolution, from initiation to possible relapse), enabling preventive measures, monitoring and optimised, individually tailored treatments to be envisaged.

Imperative goals to achieve in the next few years

A. Multidisciplinary fundamental research of excellence

Recent developments in cancer therapy clearly show that it is important to explore cancer cells on a molecular and cellular level in the living environment, in order to improve both prevention and treatment. The advent of targeted therapy is, in fact, entirely due to the discovery of "specific" markers for a given cancer cell, whether for mutated or modified proteins. So-called "fundamental" studies conducted simultaneously on the normal organ formation and the emergence of tumours enable the comparison of tissue-specific signalling pathways between normal cells and cancer cells and shed light on the key events in cancer initiation. **All technical or therapeutic advances are based on**

fundamental research work, the long-term results of which cannot be foreseen. Thus, high level of non-programmed research fields is necessary to offer new knowledge in the field of cancer. The study of the basic mechanisms of cell biology, including those examined in other scientific fields such as mathematics, physics or chemistry, must be supported so as to enable as yet unknown future advances.

Study the role of the non-coding genome in tumourigenesis.

The entire community of researchers and clinical practitioners agree on the fact that the sooner cancer is treated, the better the patient's chances. Blocking tumourigenesis in its early stages is therefore a major challenge. **All potential protagonists in carcinogenesis have yet to be identified.** The last two decades saw the collapse of the underlying concept behind genetics, according to which all cell processes were controlled by genes coding for proteins, themselves being the main effectors in the cells. We have recently learned that the non-coding genome, which represents 97 to 98% of the human genome, is not useless inactive "junk" (junk DNA) as it had been imagined but is predominantly transcribed into non-coding RNA. It is now known that expression of non-coding RNA is disrupted by certain tumours. Likewise, certain repetitive regions (centromeric or subtelomeric sequences, LINE or SINE sequences, etc.) have not been ruled out from playing a role in the carcinogenic process. The functions of the non-coding parts are only now starting to be explored and are thought to play key roles in various stages of gene expression (splicing, translation, etc.). Therefore, new study models need to be developed and more in-depth knowledge of these molecules in normal cells and cancer cells needs to be acquired, as a matter of urgency.

Epigenetic alterations of DNA and alterations of RNA synthesis, such as splicing are known to play a role in the emergence of a tumoural clone; however, the molecular links between genetic and epigenetic alterations, the role of interactions during DNA replication, transcription and translation, and particularly splicing and ribosome factors, require further clarification.

Lastly, the **mitochondria** has still not yet been sufficiently explored, whether in terms of its membrane parameters, from apoptosis to cellular metabolism, or its DNA and its own epigenetic regulations (see below).

Acquire more in-depth knowledge on the protein and metabolic characteristics of the cancer cell and its environment.

Alongside the genetic mechanisms of cancer, the role of **proteomic and metabolic dysregulation** is also crucial to understand the mechanisms of oncogenesis. Numerous protein synthesis and function do not originate form genomic or gene expression. The impact of post-translational protein modifications is a striking example of this, as is the case for immune checkpoint inhibitors, the efficacy of which is beginning to be predicted by proteomic biomarkers. Furthermore, proteomics holds back on technological perspectives. It is therefore necessary to develop sensitive approaches, which would enable proteomic analysis, even on a single cell, together with approaches for quantifying altered proteins and their post-translational modifications with CyTOF-type quantitative proteomic methods (cytometry coupled to mass spectrometry). The technology required to study and decipher the genome has now matured, as is the case for clinical-grade genomics. 3D analysis guided by new algorithms could be complemented by high-resolution cell imaging. However, functional analysis will remain the key element for - when proteins are expressed - determining the role of their alterations in the oncogenic stages, developing biomarkers for their detection and identifying new therapeutic targets. Techniques for studying the metabolome are also in development, making it possible to explore crucial mechanisms and acting in synergy with genomic abnormalities. Although the links between metabolism and cancer are now well established, with research shedding light on the role of mitochondrial DNA and gene or epigenetic alterations, the links between metabolism, nutrition and cancer, along with the links between metabolism and anti-tumour immune response are only just emerging. Further insight into these processes would allow new targets for prevention and targeted therapies to be identified.

Integration of these molecular data on a normal and pathological cellular scale, at the various stages of emergence of normal and pathological tissue, should then be validated. To successfully complete the mechanistic analyses which would solve these questions, and many others, functional study models should be developed, such as high-throughput mutant generation (using the CRISP/Cas9 system), *in vitro* but, even more importantly, in animals (such as PDX mice), including less conventional models (such as the zebrafish). At the other end of the analytical scale, "tumour-on-a-chip"-type microfluidic instruments will allow tumours to be analysed at cell by cell levels and will be powerful means of exploring tumour heterogeneity.

Continue to elucidate the mechanisms of cancer, the adaptive dynamic of tumours, genetic and non-genetic plasticity.

Studies conducted in the last decade, deeply rooted in the knowledge of normal cells, have demonstrated the **highly extensive specific plasticity** of cancer cells and their heterogeneity, notably with the discovery of cancer "stem" cells. These two characteristics contribute to the ability of cancer cells to elude conventional treatments. These advances are the fruit of research in various fields, ranging from cell biology to genomics, and using very diverse technologies.

The genetic plasticity of cancer cells is particularly important since it is related to intrinsic factors (cell ageing, reduced efficacy of DNA repair machinery) together with extrinsic factors (environmental mutagens). Tumour genome sequencing can still provide major information on the mechanisms at play in cancer cell instability, such as signatures specific to certain types of instability (for example,

related to BRCA1) or specific signatures related to certain mutagenic agents.

It is now also known that the mechanisms for cell plasticity and adaptation are not only of genetic origin but may also be non-genetic or epigenetic. Non-genetic plasticity is complex to understand, but probably crucial to explain certain aspects of tumour cell biology, such as resistance to conventional treatments. Advanced technologies, high-throughput or pan-genomic analyses (such as RNAseq, ChIPseq, ATAC-Seq etc.) allow massive exploration of these phenomena in tumour cells. It is vital to optimise these approaches on single cells so as to unravel the epigenetic aspects of tumour plasticity more clearly.

Elucidate and understand the role of the microenvironment; describe the early phases of anti-tumour immune response.

These cellular data cannot be fully understood if they are not integrated into **the cell environment**. The tumour microenvironment, in fact, helps maintain cancer cells at every stage of tumour progression, from the emergence of the first cancer cells (initiation) to the formation of metastases. For example, cell mutations in the bone marrow microenvironment may promote the emergence of leukaemia cell clones. Cells of mature tumour clone are known to act as immature cells, notably via regulation loops involving cytokines. The role of inflammation, biophysical factors (fibrosis, etc.), vascularisation, innervation, hypoxia, stromal alterations and more remote environmental factors, such as changes to the microbiota associated with various types of tissue (colon, lung, skin, etc.), in the emergence of tumour clones should be studied more closely so as to measure and control the microenvironmental impact on the emergence of tumour populations. Acting on the immune system to stimulate the anti-tumour response is now therefore a major avenue in the development of therapeutic solutions. The tolerance of tumour growth should be further described, and hierarchical classification and complete decoding of immune checkpoints should be continued. This also involves deciphering the respective roles of innate and acquired immunity in tumour emergence, together with the contextual impact (e.g.: hypoxia, influence of the microbiota, etc.) on immune response. It is now known that immune response is still operational in the early stages of tumourigenesis. It is nonetheless insufficient as the disease manages to develop.

Conversely, it is known that a tumour clone may affect remote tissue function; for example, a lung tumour disrupts the biological clock in the liver or clonal haematopoiesis promotes atherosclerosis. These processes have not, however, been described in detail, and the molecular mechanisms at play have yet to be elucidated. Lastly, **age**-related environmental characteristics should be taken into account. The emergence of tissue at precise moments in embryo development has an impact on the emergence, or indeed spontaneous regression, of certain tumours, particularly in children (mastocytosis, juvenile myelomonocytic leukaemia). Likewise, certain genetic abnormalities which appear with ageing could predict the emergence of cancerous disease, particularly blood, cardiac or vascular cancers.

Understand dormancy and the mechanisms of resistance.

The study of tumour heterogeneity and **resistance** to treatments clearly shows that tumour populations behave as a Darwinian-type system, sensitive to selection pressure. It is therefore necessary to study tumourigenesis from an evolutionary perspective, not only on cancer cell populations from the latent pre-cancerous condition to the confirmed cancerous condition (short-term evolution, on the scale of a human life), but also the evolution of genomes from the single-cell level to well-differentiated and coherently organised multi-cellularity. Drawing a parallel between evolution into multi-cellularity and evolution of the immune system should shed more light on the key stages of multi-cellularity and their presumed weaknesses, which could be responsible for cancer, whether genetically predisposed or sporadic. In this context, analysis of tumourigenesis in remote models may help elucidate the fundamental mechanisms of the processes at work. It would thus be worthwhile revisiting the findings of the research carried out in the Drosophila, zebrafish, axolotl and other models with a view to shed light on cancer mechanisms. By offering to code the biological variability of cancer cell populations in a continuous manner (resistance phenotypes, plasticity phenotypes), mathematical modelling (adaptive dynamic) should make it possible to predict phenotype evolution in cell populations and their response to treatments acting as selection pressure. These studies will make it possible to define the mechanisms for the (apparent) **dormancy** of certain cancer cell clones more clearly, and to decipher the mechanisms behind resistance to treatment and residual disease.

Integrating these molecular and cellular abnormalities found in malignant clones, and/or their environment, leads to identification of tumour heterogeneity in its ecosystem, one of the key aspects of resistance to treatment, while treatments cannot reach this level of complexity.

We are beginning to understand (infer) the way in which the structure of a tumour clone is organised during its evolution over time, based on genetic analyses, notably through recent single-cell studies. "New-generation sequencing (NGS)" technology, now applicable to single cells, makes it possible to explore genetic heterogeneity and the clonal evolution of tumour populations. This approach can be practically applied to patients via non-invasive methods, via circulating DNA analysis, which will improve diagnosis and patient follow-up, and will allow clonal evolution to be assessed *in situ*.

This research should be stepped up with a view to understanding how cells are able to promote the development or emergence of resistant clones which are highly diverse in nature, more than likely in a stochastic distribution of genetic and/or epigenetic abnormalities. We still need to understand how a cell becomes more competitive compared to its neighbouring cells. In this

context, we need to focus on the impact of the order of appearance of mutations, genetic/epigenetic structure, interactions between tumour/stroma (particularly on the symbiotic mode), tumour/endothelial cell, and also the influence of physical constraints and nerve impulses, such as the treatments received, inflammatory damage and changes in the anti-tumoural immune response.

Pursue the identification of risk factors: Genetic – Environment – Nutrition

Cancer research has shown that while driver mutations may be identified, these are often incapable of inducing the disease on their own. The context in which these occur is essential, and genetic predisposition factors with low penetrance appear to play a role. While continuing to elucidate the innermost mechanisms of tumourigenesis, it is therefore essential to expand the body of knowledge on **environmental, behavioural and constitutional genetic risk factors**. The proportion of cancers due to the hereditary transmission of a mutation is only evaluated at approximately 10%. The proportion resulting from interaction with the environment (all factors to which the body is exposed) is estimated at approximately 40% for the emergence of certain types of cancer, and at 35% for cancer deaths. This is attributed to exposure to various avoidable risk factors, related to lifestyle and behaviour (smoking, nutritional factors, sun, etc.) While these figures should not be perceived as definitive values, they nonetheless highlight possibilities for prevention, targeted monitoring and individually tailored screening.

Research into the impact of **environmental exposure resulting from human activity** (e.g.: industrial air pollutants or traffic-related pollutants, pesticides, endocrine disruptors, nanomaterials, electromagnetic fields), or of natural origin (e.g.: UV, terrestrial or cosmic gamma radiation, radon), or indeed medical origin (e.g.: CT scans, X-rays), has developed in recent decades. However, one of the major remaining issues is to shed light on the role of these factors in lifetime carcinogenesis, by investigating the specific vulnerable ages or periods (even prenatal) (e.g.: exposure to UV radiation during childhood and melanoma in adulthood), the effects of combined exposure, the effects of cumulative exposure over time, and the effects of low-level exposure. It is important to explore the variations in population vulnerability according to age, comorbidities, or genetic predisposition (gene-environment interaction), to study genomic and epigenetic changes induced by exposure, to identify markers for exposure in the general population and professional environment, to examine data on exposome and metabolomics, to use a wide range of databases, to model exposure and risks, and take into account the clinical and biological heterogeneity of cancer. If we consider the environment in a broad sense, we also face the challenge of understanding behaviours and their determining factors (e.g.: addictions, dietary behaviours, high-risk professional behaviours).

The impact of nutrition on the risk of cancer has been the subject of numerous epidemiological studies. More specifically, the proportion of cancer attributable exclusively to nutritional factors is estimated at nearly 20%. Hence, along with the fight against tobacco use, nutrition is an area in which the risk of cancer may be significantly reduced. Numerous studies have been conducted with a view to identify the nutritional factors (diet, but also alcohol use, physical exercise and weight) which play a role in the development of various types of cancer, either as risk factors (alcoholic beverages, overweight and obesity, red meat and processed meats, salt and salty foods, food supplements containing beta-carotene) or, conversely, as protective factors (physical exercise, fruit and vegetables, dietary fibre, dairy products, breastfeeding). Until now, these nutritional factors have been analysed and evaluated individually. It is now essential to incorporate the combined effect of these factors in a more global (e.g.: promoting effect of red meat and processed meats limited by the antioxidants in fruit and vegetables) and integrated approach (e.g.: interaction between nutritional factors and the host genome or microbiota) while combining experimental, epidemiological and clinical approaches. Furthermore, there is some research, although less extensive, evaluating the impact of nutritional factors (malnutrition, nutritional support, restrictive diets, etc.) on therapeutic efficacy, which should be continued, along with emerging research identifying risk factors among food additives and contaminants.

Study cancer through its evolution, particularly the early, pre-neoplastic and potentially reversible stages.

Pre-cancerous conditions are a preferential model for integrating genetic, epigenetic and environmental knowledge, both at cellular level and on a human scale, to shed more light on the early stages of carcinogenesis**.** Studies on constitutional genetic abnormalities, such as BCRA1/2, *TP53* (Li Fraumeni) and LYNCH syndrome, highlight the necessary combinatorics for evolution into cancer, either in preferential tissue (BRAC1/2 and LYNCH) or in several types of tissue (TP53). In haematology, pre-leukaemia conditions of the lymphoid cell line (MGUS, MBL) or myeloid cell line (MPN, MDS) remain insufficiently elucidated on a molecular level, except in certain specific cases (FANCONI, Noonan syndrome). The preponderance of epigenetic abnormalities, both in the patients themselves and in populations of healthy subjects followed up over the years, has already made it possible to define a new pre-neoplastic entity in haematology (CHIP). Longitudinal studies on these pre-neoplastic conditions will make it possible to identify molecular and cellular abnormalities and their (constitutional genetic, environmental or behavioural) origins, and may offer a more in-depth understanding of the complexity of cancer emergence, on a given human scale, and enable preventive approaches and treatment for these known pre-cancerous conditions including immune intervention.

1- Study cancer through its evolution, particularly the early, pre-neoplastic and potentially reversible stages.
2- Continue to elucidate the mechanisms of cancer, the adaptive dynamic of tumours, genetic and non-genetic plasticity.
3- Study the role of the non-coding genome in tumourigenesis.
4- Extend genome analysis to the study of the links between genetic abnormalities and cellular and/or intercellular functions, and epigenetic, circadian, inflammatory, metabolic and immunological regulation, etc.
5- Acquire more in-depth knowledge on the protein and metabolic characteristics of the cancer cell and its environment (role of metabolism in tumour cell plasticity).
6- Continue the study of tumour heterogeneity and its consequences.
7- Elucidate and understand the role of the microenvironment.
8- Describe the early phases of anti-tumour immune response.
9- Understand dormancy and the mechanisms of resistance.
10- Continue the identification of risk factors: Genetic – Environment – Nutrition

B. Translational and clinical research

Give patients a central role in research.

Transferring fundamental research results to the hospital bed is a priority in cancer research. There is an essential need to accelerate and validate the transfer process. The predictive nature of cellular or pre-clinical approaches continues to be minor and challenged in the light of their supporting clinical research trial results, although few identified biomarkers are actually used routinely. This transfer takes place through **translational and clinical research** which adheres to the same strict methodological and reproducibility requirements as fundamental research, guaranteeing the choice of specific fields in which a new approach could be introduced in the context of treatment or preventive medicine. Successfully transforming this research into a clinical reality is thus undoubtedly one of the greatest challenges in modern oncology. As patients and citizens are the focal point of these consequences, it is essential for them to have a key role in this transfer and be present in the various research bodies.

Develop and diversify animal models and their comparison; fine-tune pre-clinical models to validate therapies and understand the side effects of treatment.

The use of **animal models** is now questionable, not only for social/ethical reasons, but also and particularly because many of these models raise questions as to whether they can be readily extrapolated to human disease. Despite shedding some light on the fundamental mechanisms of cell function, tumourigenesis and the spread of metastases over the past forty years or so, the molecular, developmental and immune characteristics of murine (rat and mouse) carcinogenesis models can nonetheless be very different to cancer in humans. Before being used as a pre-clinical model, the murine tumour model still needs to be validated with respect to tumourigenesis in humans. The lack of validation could at least partly explain treatment failure for certain types of cancer. It has, in fact, emerged that, *in vivo*, cancerous disease cannot be considered outside its global context, comprising the tumour and its environment. Hence, targeting the immune system is now at the heart of certain research as its dialogue with cancer cells seems crucial to their growth. Likewise, the microbiota is attracting particular attention, having also been shown to interact with cancer cells and modify the immune system. It is therefore crucial to use specifically generated, dedicated pre-clinical models to take these dimensions into account.

The development of **more relevant pre-clinical models** is thus an essential requirement in order to better predict treatment side effects and efficacy. The development of syngeneic and humanised models, facilitated by **CRISPR approaches**, is a major opportunity. Transgenic mouse models expressing various types of functional proteins, or not, make it possible to predict protein tissue expression and to only target its cell function, without modifying the expression in its normal cell context. As a complementary approach, **spontaneous tumours** developing in animals enable the physiopathology of the disease to be studied in an authentic integrated system. Spontaneous models in dogs, for example, and developed models in pigs are thus in development. Development is crucial in order to evaluate biomedical technologies which require specimens on a human scale. These animal models therefore need to be developed, validated and used responsibly and under supervision on a national and international scale, according to the scientific field and medical need.

Develop alternative models.

Alternative or complementary models should be developed either, to replace or reduce in specific cases the number of animal models, and/or to provide additional pre-clinical information. Conventional cell models are also imperfect in liquid or 2D suspensions; however, alternative methods are in development, such as three-dimensional tumour reconstruction *in vitro* (also known as organoids) or *in vivo* (scaffold), which make it possible to integrate and control human tumour cells with immune cells, and components of the microbiota, bone and vessels, in a given environment. The impact of a unique or combined genetic abnormality may be integrated into a precise tissue differentiation context via an induced pluripotent stem cell model (iPSC) derived from patient cells. This makes it possible to measure the impact of genetic abnormalities on the phenotype, function, tumour cell secretome, dormancy and tumour resistance.

In addition, it is essential to encourage **new mathematical approaches to modelling and analysis** of cell proliferation

and its inhibition. Continuous models (ordinary differential equations, partial derivative equations) will make it possible to monitor the evolution of cancerous or non-cancerous cell populations and their interactions, the representation and control of reversible or non-reversible resistance to cancer treatments (particularly by taking into account tumour phenotype/genotype heterogeneity), therapeutic control exerted on functional targets (proliferation, cell death, differentiation) and the improvement in control using optimised treatment strategies (on a cell population scale), the representation of tumour/stroma interactions (competitive or mutualistic), and the representation and control of molecular targets at single-cell level.

Facilitate drug design and the reclassification of older molecules.

While transfer of the fruit of fundamental research in the therapeutic field (agents, radiotherapy, surgery) should continue in the same positive way as in recent decades, closer attention should be given to the development of pain relief treatments and the prevention of immediate and longer-term side effects of treatments. Interaction with other research disciplines will be essential to achieve this goal. New genomic and, particularly, proteomic and biochemical data should be compared with existing chemical libraries and serve as a basis for multidisciplinary drug design research.

Technology transfer for new biomarkers with a view to early diagnosis and monitoring (toxicity of treatment, particularly immunotherapy, prediction of relapse/resistance).

Since the advent of non-invasive tests such as imaging, the use of biomarkers in oncology has become increasingly complex over the years. Molecular biology and NGS (new-generation sequencing) of RNA, miRNA, whole genome, exome, methylome and epigenome are undoubtedly behind these developments, with nearly 600 genetic abnormalities (mutations, fusions, CNVs, etc.) potentially able to be identified at present. It is important to create new functional validation platforms for new identified mutations and **cellular genetic engineering platforms**, to bring pdx collections into general use so as to have access to all omic technologies and help clinical proteomics become firmly established. Technical approaches using tumour material, such as FISH, CISH and IHC, are expanding detection possibilities and helping to enhance the precision of biomarkers. Liquid biopsies from a single blood sample are now capable of detecting circulating tumour DNA and circulating tumour cells. Furthermore, the increasing number of combined therapeutic approaches is giving rise to the risk of major toxicity which warrants the development of various type of biomarkers. These must be able to evaluate the treatment efficacy for the patient's benefit, to detect the possibilities of secondary toxicity, emergence of immune resistance and risk of relapse, and to predict patient response. Immune phenotyping is dependent of several parameters: mutational load, measurement of immune checkpoint expression, characterisation of lymphocytic and myeloid infiltrates, measurement of soluble inhibitors, evaluation of tumour metabolism and evaluation of sensitivity to immune effectors.

Develop sensitive, precise and reproducible technologies.

Given the diverse technical possibilities, optimisation of biomarker panels which should include several parameters (comprising genetic and epigenetic factors) is essential in order to characterise patients who are candidates for the corresponding treatments. Mathematical approaches and new algorithms to integrate these multiple parameters need to be developed. All of these concepts represent the basis of "personalised" or “precision" medicine. The key challenge is probably the successful large-scale harmonisation of methods.

Harmonising procedures between various laboratories with a view to characterise a multiparameter profile for these biomarkers has now become a major challenge. Sample collection, choice of sample type, storage and transport protocols for clinical specimens and data processing are parameters which should ideally be harmonised. The introduction of validation platforms on a national scale, the creation of biobanks meeting standard quality criteria for the preparation, storage and annotation of samples, universal high-performance companion diagnostic tests and the construction of robust databases will help define the general guidelines for clinical decision-making and thus optimise treatments, while making them less costly. At the same time, data generation requires precise organisation of information so as to effectively capitalise on massive data resources, with a view to offering global patient management solutions. Multidisciplinary expertise and interaction between researchers, molecular biologists, clinical practitioners, pathologists and bioinformaticians are proving vital in facing up to these challenges along with the routine introduction of precision medicine.

Develop methods for stratification of responders and to predict clinical benefit.

Faced with the numerous diverse biological data and treatment possibilities, methods for stratification of responders are an essential part of cancer research. Introducing molecular stratification of the tumour and microenvironment, in particular, for instance, by genetic sequencing, phenotyping and imaging, may help define subgroups of responders. Faced with the unknown factors relating to the potential toxicity of new therapies or combinatorics, it will be crucial to develop predictive models for clinical benefit. Two predictive methods should be envisaged: 1) patient population studies (particularly population-based PK-PD) using public health-type statistical studies, and 2) studies modelling the response to treatment in organisms represented by interacting healthy and transformed stromal cancer cell populations. The latter aspect naturally involves mathematical modelling based on adaptive dynamic studies on the cell populations concerned. As is the case for fundamental research, translational and

clinical research must have access to sensitive, precise and reproducible technologies.

1- Give patients a central role in research.
2- Develop and diversify animal models and their comparison (advantages, disadvantages, integrated studies on the microbiota, immune response), highlight the importance of spontaneous tumour models in animals and create alternative methods to animal models (iPSC, organoids).
3- Fine-tune pre-clinical models to validate therapies and understand the side effects of treatment.
4- Develop alternative models.
5- Facilitate drug design and the reclassification of older molecules.
6- Identify predictive biomarkers for a successful therapeutic approach.
7- Technology transfer for new biomarkers with a view to early diagnosis and monitoring (toxicity of treatment, particularly immunotherapy, prediction of relapse/resistance).
8- Create new functional validation platforms for new identified mutations and cellular genetic engineering platforms, bring pdx collections into general use so as to have access to all omic technologies, and help clinical proteomics become firmly established.
9- Develop methods for stratification of responders and to predict clinical benefit.
10- Develop sensitive, precise and reproducible technologies.

C. Value creation and support for research

Scientific research in general and cancer research in particular are currently characterised by an abundance of hypotheses, models, investigation methods, collected and generated data, and channels for the publication of results. This wealth of resources is a remarkable asset for identifying the causal mechanisms, the signalling pathways involved in the development of the disease and the therapeutic targets, but also for evaluating new treatment strategies and predicting patient outcomes and certain long-term effects of treatment. It is also opening up new challenges and **has led to calls to rethink the way in which research is organised** in favour of reproducible and integrated research, and to ensure that the abundant methods, resources and data contribute to the growing knowledge on cancer and have prompt effects on patient management.
Three changes particularly illustrate the modern challenges facing cancer research:
- the number of publications on cancer indexed on WoS increased by approximately 8% per year in the past 10 years, and therefore doubled between 2007 and 2016. There are currently more than 250 scientific journals reporting results relating to cancer research. It is now impossible for scientists to reasonably understand all of these reliable results relating to cancer research. Scientists are now developing advanced expertise in very specific subdivisions of cancer research.
- the variety of data studied with a view to shedding light on cancer is constantly growing. It is now acknowledged that anatomopathological, omic, epigenetic, immunological, pharmacological, microbiotic, environmental, behavioural, nutritional and family data, or even data relating to physical exercise at various stages in life, should be jointly taken into account in order to describe and understand cancer. In addition to advanced research in each of these fields, integration, contextualisation and co-interpretation of the generated results are vital to elucidate the complexity of cancer, and effectively preventing and treating the disease.
- the evolution of digital technologies and communication tools now permits the collection of digital data banks (samples, images, clinical records) potentially containing a wealth of information, particularly when shared so as to reach cohort scales usually inaccessible to a single research centre or, indeed, a single country. These vast heterogeneous data banks constitute new experimental material, enabling the emergence of original knowledge using dedicated exploratory methods, based on artificial intelligence, statistical learning and semantic enrichment.

Generate high-quality, reproducible data and results, by prioritising information quality and representativeness over quantity.

Faced with these changes, it is now more than ever essential to:
- **generate quality-controlled data**, automatically matched with a description of the methods, allowing them to be reproduced and encouraging the use and improvement of reference methodologies for database production,
- encourage the **sharing of data** generated in the researcher community, to contribute to an international collective effort in building data banks which can be explored using dedicated methods enabling the identification of mechanisms unable to be detected on a small scale. Each team's work thus creates value on two levels, firstly via the new knowledge generated by the team, then via incorporation of these data into a data bank, for the contents to be reused on another scale.
- and, on the other hand, encourage the development of research using data from existing databases and enrichment of these databases.
- place each cancer research team in contact with integrated research laboratories working in oncology, to develop explanatory and predictive models using the full range of established knowledge.
- support and develop research into biomedical data mining methods using the most recent information technology to facilitate the emergence of original knowledge based on new experimental materials represented by vast data banks.

Value researchers' work based on quality and reusability rather than quantity

The advent of the massive data era is certainly one of the most impressive technological revolutions accompanying scientific achievements in the past ten years. It is consequently crucial that all data accumulated during massive sequencing of tumours can be utilised for research purposes insofar as these are a highly valuable resource for

Text prepared by
François Berger, Jean-Paul Borg, Irène Buvat-Guillemet, Salem Chouaib, Jean Clairambault, Jacqueline Clavel, Yvan De Launoit, Olivier Delattre, Annick Harel-Bellan, Claude Leclerc, Fabrice Pierre, Ariel Savina, Eric Solary, Muriel Altabef, Alain Eychène, Christine Chomienne, Members of the Expert Committee of the AVIESAN Multi-Organisation Thematic Institute (ITMO) on Cancer.
December 2017

furthering our knowledge of cancer, along with gene function, with very few genes having been annotated (molecular network, location, function, etc.), and for creating new hypotheses able to be tested in a laboratory setting, and useful in future human clinical practice. In this respect, data mining of genomic databases should be accessible to research workforce. France has an exceptional wealth of fundamental research teams, highly active in most fields of biology, chemistry, physics, mathematics, etc., able to take up the challenge in terms of functional decryption of genome alterations. This requires data to be adequately formatted; in particular, the associated clinical annotations need to be stated in a comprehensive, reliable, precise and, if possible, standard manner. Software to explore sequences, particularly non-coding sequences, not previously studied, needs to be developed. Samples need to be provided to be used for research purposes, and these need to be repeated, at various stages of the disease so as to integrate the analysis of genomic evolution and the dynamic of tumour cell populations. Lastly, dialogue with researchers is also required, before embarking on data collection.

Promote interdisciplinary training at doctoral schools, create interdisciplinary openings and develop continuous training specifically in cancer research

Researchers' range of skills must evolve at the same time. Hence, appropriate and, in particular, multidisciplinary doctoral and post-doctoral theoretical and practical training should be available. Furthermore, researchers clearly need access to regularly updated guidelines on required skills and appropriate training.

1- Generate high-quality, reproducible data and results, by prioritising information quality and representativeness over quantity.
2- Facilitate database creation and data sharing in compliance with referenced quality criteria, taking steps to ensure data bank interoperability
3- Develop and share reproducible data analysis methodologies for faster, objective problem-solving
4- Develop biomedical data mining methods using modern statistical, artificial intelligence and computer techniques
5- Encourage validation studies of these methods
6- Create explanatory and predictive models using heterogeneous data, subject to prospective, multi-centre validation
7- Where possible, encourage different approaches to given patients and/or given tumours
8- Promote cross-cutting and interdisciplinary approaches, by decompartmentalising actions and encouraging multidisciplinary cooperation
9- Promote interdisciplinary training at doctoral schools and create interdisciplinary openings dedicated to the study and treatment of cancer in universities and the *Grandes Ecoles*
10- Value researchers' work based on quality and reusability rather than quantity