Living instruments and theoretical terms: Xenografts as measurements in cancer research

Pierre-Luc Germain*

June 30, 2013

Abstract

I discuss the relationship between theoretical terms and measuring devices using a very peculiar example from biomedical research: cancer transplantation models. I do so through two complementary comparisons. I first show how a historical case study can shed light on a similar case from contemporary biomedical research. But I also compare both to a paradigmatic case of measurement in the physical sciences – thermometry – which reveals some of the most relevant epistemological issues. The comparison offers arguments for the recent debate on the operational definition of Cancer Stem Cells, and thereby suggests the relevance of a comparative approach in the history and philosophy of science.

As I argue in the first part of this paper, so-called xenograft "models" of cancer are often used not as models in the traditional, analogical sense, but as measuring devices. This prompts the question of what it is that they measure, and of the relationship they entertain with it. To investigate these issues, I compare two cases of xenograft as measurements with the prototypical example of a measuring device: the thermometer. I rely on the work by Hasok Chang on the history and epistemology of thermometry (Chang, 2004). Behind the apparent simplicity of thermometers lies a daunting epistemological problem, which he labels "the problem of nomic measurement": in a nutshell, there are a variety of thermometers giving inconsistent (not linearly correlated) readings, and we would need to know already what temperature is in order to know which one gives the right reading. I highlight some relevant similarities between his history of thermometry and the examples I will present from cancer research. In both cases, instruments and theories have a reciprocal stabilizing role: the instruments are at the same time means subordinated to theoretical understanding, and theoretical terms are means of bridging different instrumental and operational contexts. Finally, the comparison sheds some light on a contemporary debate in cancer research.

In the first part of this paper, I present the instrumental role that organisms sometimes play in biomedical research (section 1.1), and apply this concept to early xenograft models of cancer

*Università degli Studi di Milano & European Institute of Oncology (IEO), Campus IFOM-IEO, Via Adamello, 16, 20139 Milan, Italy. Contact: pierre.germain@ieo.eu

(section 1.2). I show how, in early xenograft experiments, transplantability was taken as a signal for an abstract quantity. Ultimately, this attempt at the mutual stabilization of operational and theoretical concepts failed. However, this failure may be informative for a similar attempt in contemporary cancer research. In order to better understand the epistemological issues involved, I review how the analogous problem was solved in the case of thermometry (section 2.1). I then present the more recent use of xenograft as measurements in the Cancer Stem Cell Framework, and the specific problem of establishing an operational definition (section 2.2). Finally, I show how some insights from the thermometry example can be used to inform, and even take position on, some of the issues relating to this problem (section 2.3).

1   Living instruments

1.1   The roles of laboratory animals in biomedical research

Animals have long been used in biological research aimed at learning about human biology, most often acting as "a surrogate for a human being" (ILAR and NRC, 1998, p.10). It is this surrogacy that has warranted calling them "models". However in the life sciences there seems to have been a conflation between this intensional meaning of the term and its coincidental extensional meaning, so that organisms *often* used as models came to permanently bear that label. While they are indeed often used as surrogates, this obscures a great variety of functions that organisms actually play in research. For instance, organisms are often factories for materials, as was the case in the first half of the 20th century when stocks of viruses were kept and grown in the lab by serial infection of host animals (e.g. rabbits). Nowadays, plasmids are routinely used for DNA cloning, and labs around the world still rely on the bleeding of mice (or rabbits, goats, etc.) for the production of antibodies. Calling these animals "models" would be so far-fetched as to rob the notion of any meaning.

Organisms can therefore have a variety of roles in biomedical research, which are not exhausted by the notion of model. The examples that I will discuss here represent a particular such function, which is that of measuring/detection device (which I will call the "instrumental" role for reasons of convenience).

What I mean by instrumental role can be illustrated with a simple example: the Ascheim-Zondek (A-Z) test for pregnancy invented in the 1920's. In this test, mice are injected with the urine of a female patient, and are dissected after two days. If the injection caused small blood stains on the mouse's ovarian follicles, then the patient is pregnant (Zondek, 1928). The mouse, here, is not used as a replica of the patient, not least because the phenotype actually being used by the test – the blood stains – are absent from the woman. Rather, it seems justified to talk of a measuring (or at least detection) device, as the animal has the function of detecting a signal in order to learn something about the woman. The mouse allows one to detect, in the input, something that was otherwise unobservable.

One can distinguish several kinds of observational instruments, and in order to talk of a measuring/detection device, there needs to be a kind of decoupling between the observable output of the device (the signal) and what this allows us to infer in the target system: not all

that is observable in the readout of an instrument is actually informative about the input. The fact that there is, for instance, an air bubble in the column of mercury is known to be irrelevant to the temperature one wishes to measure. Finally, the case of the A-Z test is not a measuring device in a strict sense, for its output is binary – it is a detection device. A measuring device should at least provide ordinal, if not quantitative, readouts. This implies that "measuring locates the target in a theoretically constructed logical space" (van Fraassen, 2008, p.2), and indeed I will be concerned here with this relationship between measuring devices on the one hand, and the structure and reference points of this theoretical space on the other.

Elsewhere (Germain, ming), I characterize the instrumental role of organisms in detail, and argue for its relevance in contemporary research. Here, I would like to pursue a slightly different goal, namely to push the analogy to a comparison with the classical example of the thermometer, in order to study the relation between these instruments and what it is that they should measure. Throughout this paper, I will discuss cases of such "living instruments" which are strikingly similar to the A-Z test, and yet still of high relevance in contemporary cancer research: xenografts as measurements of tumorigenicity.

1.2   Xenograft experiments in early cancer research

Following the nomenclature of Snell (1964), a xenograft – or xenotransplantation – is a case of tissue transplantation where the donor and recipient are of two different species. From the end of the 19th to the middle of the 20th centuries, transplantation was of big interest to the scientific community, and scientists attempted a disarraying diversity of transplant experiments. This was especially common in the field of cancer, in an attempt to domesticate tumours to the laboratory. Human tumours, if they were to be studied experimentally, needed to be studied outside their host. Even in the case of animal tumours, scientists were confronted with the simultaneous shortage of spontaneous tumours and inability to sustain a tumour beyond the death of its host. Hence transplantation became (and still is today, although for different reasons) among the most widespread ways of studying cancer in a lab.

There is some disagreement as to the first author to be credited with successful tumour transplantation. Claims go back at least to 1889-1898 (Hanau, 1889; Mayet, 1902; Ewing, 1919)[1], but were all strongly criticized – see for instance Hekzog (1902), who instead credited the feat to Loeb (see also Loeb 1945). More recently, a historical review of chemotherapy attributes the "first transplantable tumor systems in rodents" to Clowes in the early 1910's (DeVita and Chu, 2008, p.8643). The contention seems to hinge on what is stable enough to constitute a "system". Indeed, an important reason for the disagreement is that for a long time, the criteria on which to evaluate a successful graft were unclear (see Loeb 1945, chapter 12). In general, grafts lasted only for some time before resorbing under the pressure of the host's immune system (although, at that time, the explanation was that the foreign cells lacked specific "foods" – it was not until Medawar's work in the 1940's that the immunological

---

[1]As a matter of fact, Novinski (1876) showed even earlier the successful transplantation of the canine venereal tumour. However, because it was believed that a virus was transmitted, rather than the tumour itself, Novinsky's work was never interpreted as transplantation. It was only established recently that the tumour itself, and not some infectious agent, is transmitted (Murgia et al., 2006).

basis of rejection was firmly established). Therefore, a line had to be drawn somewhere to distinguish cells that have successfully engrafted, albeit only temporarily, and cells that are just "still there" from the injection. Some authors were already discussing histological criteria, for instance vascularization, but for a long time there was no established way to make the distinction. A related problem is that the injection caused an injury to the recipient that had important risks of infections, which (either because of the inflammation or of the death it brought) could easily pass for cancer.

The systematic, large-scale work of Loeb (especially from 1901 to 1910) was certainly of central importance in the establishment of transplantation systems (Witkowski, 1983), but the experiments were of limited success for a long time (see for instance Funk 1915). The main improvement in this respect came from the discovery that some locations in the host (the brain, the anterior chamber of the eye, etc) accepted grafts more readily. As grafts started to become more efficient, and transplantation systems were tamed, the possibility appeared of using them as tools for a variety of purposes.

The long established observation that only embryonic and cancer tissues were transplantable across species (normal adult tissues "did not take") lead to such instrumental uses. Some scientists proposed that "transplantability constitute[s] a biological test for cancer" (Greene, 1948, p.1364). Greene suggested that the "study of the transplants allows a more precise classification than is warranted from the morphologic features of the biopsy specimen" (*ibidem*). He was explicitly proposing a diagnostic tool to replace what he considered to be a 'coarse' and uninformed judgement of pathologists.

Importantly, transplantation was not simply believed to be a useful signal: if it was a good signal, it was because it was signaling *something*, and therefore giving access to some invisible differences between cancer cells:

> "The fact that a biological quality as fundamental as the ability to grow in an alien species differentiates morphologically identical tumors suggests that the tumors must also differ in metabolic or biochemical constitution. It would seem important, therefore, to distinguish tumors with respect to this property and to study the different groups formed rather than to consider morphological similarity a proof of constitutional identity." (Greene, 1952, p.41)

The very idea of using transplantation as a test implied that transplantation made visible a difference that was already in the tissues. More importantly, transplantability was not understood as binary: degrees of transplantability could be obtained either by resorting to statistics (the proportion of cases where the transplant was successful) or by assessing the pace, duration, and quality of the growth. Hence more than a tool to detect malignancy, transplantation was a tool to measure it. Arguably, this quantity was not numerical in a strong sense, but it was at least ordinal: by 1952, Green had ranked over one hundred tumours on the basis of their transplantability.

1.3 Inventing an abstract quantity

The key step I am interested in here is this invention of an abstract quantity to which transplantation provided access. Some, following Loeb, took this abstract quantity to be the "growth momentum":

> "Clinically, the growth momentum of a tumor, i.e., the rate of enlargement, infiltration, and metastasis, characterizes the degree of malignancy of a neoplasm. It has been shown experimentally with animal tumors that growth momentum is likewise one of the most important factors governing transplantability, particularly heterologous transplantability. [...] Accordingly, the determination of heterologous transplantability of a tumor would provide a measure of its growth momentum and, hence, the degree of malignancy." (Towbin, 1951, p.716)

To make the analogy plain, "growth momentum" was the unobservable value to be measured (the equivalent of temperature), and the growth of the transplant was the signal (the equivalent of the height of the mercury column). For some, transplantability was a proxy to "growth momentum", for others it was a measurement of the "autonomy" of the tumours. In both cases, these abstract quantities were already loaded with both conceptual content and experience. Indeed, the notion of autonomy was already used to explain both developmental processes and carcinogenesis (see for instance the work of Hugo Ribbert or John George Adami, where both are explained as differential responses to "tissue tension"). Similarly, the notion of "growth momentum"[2] was central to Leo Loeb's biology (Loeb, 1945). Loeb himself was using transplantation as a measuring procedure, although his approach was more complex[3] Importantly, in both cases the abstract notion was used to explain a variety of phenomena, both natural and artificial, both normal and pathological[4].

Given the diversity of transplantation procedures (host species, site of transplantation, assessment method, etc.), it should not come as a surprise that scientists produced different, conflicting classifications. This prompts the question of which transplantation system correctly tracks growth momentum (or the abstract quantity of choice). Here, we meet what has been the core problem of classical thermometry: without a direct access to temperature, how can one know which thermometer gives the right temperature?

---

[2]The notion seems to come from demographics, where it became especially popular in the 1920's. Should this have been Loeb's inspiration, it would be yet another early example of what would become very recurrent analogies between cancer and socio-economics: Bolshevik cells, anarchist cells, etc...

[3]Loeb considered his experiments to simultaneously measure different aspects of the phenomena which he called "differentials" (Loeb, 1945). Although I believe that the present discussion could equally apply to his work, the presence of multiple quantities in the same reading makes the matter less straightforward.

[4]Because it took cancer and physiology as variations of the same causes, this approach was part of what Michel Morange called "the Regulatory Vision of cancer", in which "cancer was conceived as a disease of development." (Morange, 1997, p.6)

2 Measurement and theoretical terms

But these questions already make an important assumption. Why did there have to be a single, true classification system – or, for that matter, a single, true temperature? Chang's "Ontological principle of single value", which states that "a real physical property can have no more than one definite value in a given situation." (Chang, 2004, p.90), simply displaces the question to why scientists believed that temperature was "a real physical property". At least part of the answer has to do with the fact that the notion of temperature already had an entrenched ancestry. It built upon both a long philosophical tradition that had already pre-conceptualized the notions of heat and "caloric", and an immediate, daily-life experience of variations in temperature[5]. Both of these loosely fitted the readings of thermometers, strongly suggesting that all were related to a common quantity.

To some extent, similar arguments can be made in the case of malignancy. In the quote from Towbin above, it is striking that the term "growth momentum" is associated to a disarraying variety of phenomena which, obviously, *could* point in different directions. But it was sufficient that they would align most of the time to postulate a common cause – after all, most scientific laws are *ceteris paribus*.

Moreover, the clinical context of xenotransplantation experiments already provided a very concrete notion of malignancy: the clinical outcomes of patients provided a grading of tumours. In a sense, therefore, the clinical could be understood as the (*de facto*) unobservable to which the instruments are giving access. In this context, transplantation experiments simply ought to approximate clinical outcomes. Indeed, scientists tinkered their transplantation procedures to approximate clinical knowledge[6]. Nevertheless, as will become obvious in the next section, scientists did not go all the way in this direction. Two important reasons can be given for this. The first, supported by the passage from Greene quoted above, is that they were not just after a predictive system, but an explanatory, or at least exploratory one, enabling an investigation of the "mechanisms of autonomy"(Greene, 1951, p.902). In this context, approximating the clinical outcome has the value of highlighting departures from it, and therefore of stabilizing these departures as objects of explanation.

A second reason is the lack of repeatability of what we might call the "clinical measurements" of malignancy. Malignancy, understood as the ability for pathological, neoplastic growth, is a relational property of cells. Indeed, cells can to some extent become malignant just because of differences in the surrounding tissue (Bhowmick and Neilson, 2004). This means that although each human tumour is associated to a medical history and clinical outcome, this history has many determinants that are external to the cancer cells. These can be due to the exact

---

[5]As Chang writes: "human sensation serves as a prior standard for thermoscopes" (Chang, 2004, p.42)

[6]The test was used in some laboratories (Towbin, 1951), but its sensitivity was criticized. Hence in the decades that followed, different methods were shown to make the hosts more receptive, including X-ray irradiation, cortisone treatment, and thymectomy, but the most important advance was certainly the discovery, in the early 1960's, of the *Nude* mouse mutant. Aside from its famous absence of hair, the nude mouse is characterized by its lack of a functional thymus and the corresponding massive reduction in T cells. As a consequence, it is largely unable to mount an immune rejection of the foreign tissues. In fact, even normal tissues successfully engrafted, but at that time Greene's idea of transplantability as a test of cancer was already forgotten.

site and micro-environment of the tumour, the patients's constitution or genetic background, treatment history, etc, so that the correlation between the nature of the cancer cells and the medical outcome is messily statistical rather than deterministic. Nevertheless, experience strongly suggested that among the factors of malignancy were differences that were intrinsic to the cancer cells: hence the invention of abstract notions such as "growth momentum". In the absence of an independent mean to group tumours together (which transplantability was trying to offer), this meant that in practice each clinical outcome was yet another poor approximation of the tumour's "intrinsic malignancy". The invention of such an intrinsic property, because it makes it transportable (transplantable), enables the phenomena to simultaneously become a theoretical variable and an object of experimental study[7].

To study cancer in a mouse, moreover in the highly artificial context of the laboratory, will never be the same as to study it in patients. However, the invention of an abstract property to which the measurement procedure would give imperfect access enabled the use of the mouse system to study the human system. In other words, the abstract concept provided a bridge between different material systems, by *assuming* that the material systems were simply two imperfect operationalizations of the same thing.

An analogous issue is ubiquitous in Chang's (2004) history of thermometry, although the question is never explicitly addressed: why did scientists need an abstract concept of temperature? Metallurgists were doing fine optimizing their furnaces with their solid thermometers, and scientists were designing steam engines on the basis of air thermometers. The only problem with purely operational concepts is that they are unable to afford semantic expansion, and therefore knowledge gained, say, at the casual temperature range could not be easily transferred to very high or very low temperatures. The concept of temperature, abstracted from any operationalization, provided such a bridge.

2.1 Reaching a thermometry consensus

Hasok Chang's history of thermometry (Chang, 2004) describes the many strategies with which scientists tried to stabilize the concept of temperature, many of which have an analog in the context of xenografts. For instance, a substantial part of the history of thermometry was aimed at establishing (or unsettling) "fixed points": the freezing point, melting point, boiling point, blood temperature, the first night frost, etc., up to the temperature of the cellars of Paris' Observatory (*Id.*, p.10). The same could be said of xenograft experiments: Greene, Towbin, Loeb and others spent considerable efforts establishing fixed points. The most obvious is the inability of healthy differentiated tissue to grow, but more interestingly a variety of clinical characteristics (e.g. whether the tumour was metastatic) were believed, in clinical experience, to correlate with the abstract quantity. Chang's tale of how the fixity of fixed points was challenged, until fixed points had to be manufactured, would find many echoes here. But for the purpose of this paper, it is more useful to go to a later episode of Chang's history

---

[7]In fact, most of experimental biology is about *turning* context-sensitive features into capacities or stable properties. Arguably, the stunning success of the strategies of decomposition and localization in biology (Bechtel and Richardson, 1993) is partly due to the fact that these strategies simultaneously provide understanding of the phenomena and constitute it as an epistemic thing (Rheinberger, 1997).

and briefly look at how the scientific community finally settled on what is nowadays taken for granted: absolute temperature.

In the end, most of the conundrum was solved when Thomson (Lord Kelvin) reasoned that the establishment of an absolute temperature required "a theoretical relation expressing temperature in terms of other general concepts", and relied on "the little-known theory of heat engines by the army engineer Sadi Carnot (1796-1832)" (*Id.*, p.175).

> "As Thomson was attempting to reduce temperature to a better established theoretical concept, the notion of mechanical effect (or, work) fitted the bill here. A theoretical relation between heat and mechanical effect is precisely what was provided by a theory of heat engines." (*Id.*, p.175)

The first step was therefore to postulate an abstract temperature as defined by its theoretical (and quantitative) relationship with another abstract term, "work", which was linked to operational concepts through mechanics. Interestingly, the second step was then a "deliberate conflation" of this absolute temperature and of the temperature given by any thermometer (air, mercury, etc): physicists assumed that the thermometers gave imperfect readings of this abstract quantity, and simply substituted one for the other in their formula (*Id.*, p.214). Obviously, the fit was not perfect, but discrepancies allowed scientists to recalibrate their instruments, and engage in successive steps of approximation and recalibration which Chang characterized as "epistemic iteration": "point-by-point justification of each and every step is neither possible nor necessary, what matters is that each stage leads on to the next one with some improvement." (*Id.*, p.215). While such iterations need not necessarily converge, when they do it vindicates both the instrument and the theoretical construction.

It is interesting to note that the theory provided both the motivation and the solution to the problem: scientists investigated thermometry to build a theory of temperature, and yet is it the theory which solved the problems of thermometry. There is no problem in this circularity: tools and theories that are rightly articulated gradually stabilize each other. But it means that there is considerable freedom in the starting point. Indeed, the theoretical relationship between work and temperature was enabling a quantitative theory of heat and an explanation of thermometry, therefore giving reasons to connect the different "operational temperatures". But at the same time, it provided ways of going from one thermometer to the other, therefore undermining the need for "the right thermometer".

2.2   Xenografts in the Cancer Stem Cell framework

Nowadays, saying that a tissue grows because of its growth momentum is as explanatory as saying that opium makes one sleep because of its "vertu dormitive". The situation was certainly different for scientists thinking within the theoretical context of Loeb (1945). Hence the fact that the notion of the "growth momentum" of cancer tissues did not catch on is most surely related to the demise of the notion in developmental biology. In any case, both growth momentum or Greene's "autonomy" lacked tractable relationship with other notions, and this proved critical for the establishment of a theory-instrument articulation.

From the 1960's on, and especially until the 1990's, cancer research was unified around a different notion, dissociated from physiology: tumorigenicity. While the concept of tumorigenicity would deserve a history of its own, it is an umbrella term gathering so many heterogeneous meanings that its treatment would only distract from the present discussion. Instead, I would like to discuss a more recent episode of xenograft experiments. Because it bears a strong resemblance to the previous example, its analysis can benefit from the previous discussion.

At the turn of 2000's, strong analogy between physiology and pathology resurfaced in cancer research under the form of the Cancer Stem Cell (CSC) hypothesis. I shall only briefly summarize the CSC model here – for a more detailed discussion, see Blasimme et al. (2013); Visvader and Lindeman (2012); Valent et al. (2012). Its core hypothesis is that cancer progression is driven by a small subpopulation of tumour cells with stem-cell-like properties. Like in normal tissues, only these cells are capable of infinite replication, and they therefore fuel a hierarchical tissue development. After the discovery that myeloid leukaemia followed such a model, a whole research programme developed with the aim of identifying and isolating such cells in other forms of cancer. The basic strategy is to divide the population in subpopulations according to some markers (typically on the cell's membrane, so that cells can be sorted through antibody-based methods), and assess whether these subpopulations differed in terms of some measurement. Once more, mice were recruited as measuring devices, and once more, a variety of transplantation procedures resulted in conflicting measurements.

One of the best examples of this conflict is the controversy regarding melanoma stem cells. In the field of melanoma research, scientists have proposed to speak of melanoma-initiating cells (MIC) as an operational definition of CSC: MIC are cells which, when serially transplanted into an immuno-deficient mouse, are able to produce tumours recapitulating the heterogeneity of the original tumour. Strictly speaking, scientists are most often not measuring whether the cells are able to produce tumours, but to what extent, and therefore the injected cells are not said to be all CSC, but to be enriched in CSC.

A few years ago, Schatton et al. (2008) identified a sub-population of cells, ABCB5+ cells (cells expressing the ABCB5 antigen at their surface) enriched in what they claimed to be CSC. In order to test it against the operational definition of MIC, they transplanted ABCB5- and ABCB5+ populations of cells from a human tumour into NOD/SCID immunodeficient mice ("non-obese diabetic/severe combined immunodeficiency") and looked at the tumour progression. After 8 weeks, hardly any tumour grew in the first case, and the majority steadily grew in the second. In other words, only a small proportion of tumour cells, strongly enriched in the ABCB5+ population, were able to initiate and sustain new tumours. They published an enthusiastic letter to nature which was heavily cited, and for a time it was proclaimed that CSC had been identified in melanoma.

Some months later, Morrison's lab (Quintana et al., 2008) published a paper attacking these claims. The most important for the present discussion is that they tried the same experiments with an even more immunocompromised mouse (the NOD/SCID Il2rd-/- mouse) and obtained radically different results. Injecting single cells, they found that one out of four was able to initiate palpable tumours, and trying a wide range of markers, they were not able to correlate this with any signature. They therefore concluded that there was no proof yet that the CSC model obtained in melanoma, and that experiments seeking tumour-initiating cells should beware of

relevant differences between the tumour environment in the patient and in the mouse host. The lesson, it seems, was that Schatton et al. (2008) had drawn a bad conclusion that was due to the particular mouse model they used, which happened to be unrepresentative of the human host. Quintana's paper was (and still is) a big success, being cited even more than the first, often as a methodological warning. Nevertheless, given how "unnatural" the dramatically immunocompromised mice are, there is still considerable debate as to which is the best (see for instance Civenni et al. 2011).

The mice, and in fact the whole experimental system, again acted as an instrument: they transformed an unobservable, yet causally relevant difference, into a visible signal, thus revealing this difference. But what difference exactly? Given the disagreements of two recipients, which signal faithfully informs us about tumorigenicity or "CSC-ness"?

The question becomes even more acute if we consider the rest of the story. Slightly more than a year after Quintana's paper, Schatton et al. (2010) published a follow-up paper in *Cancer Research*, apparently moving the topic: "Modulation of T-Cell Activation by Malignant Melanoma Initiating Cells". Taking the discrepancies between the two studies as a starting point, they addressed the question of why the difference between the mouse strains – the absence or presence of the interleukin-2 gamma receptor (Il2rg) – made such a difference to the apparent role of ABCB5+ cells. It turned out that ABCB5+ cells seem to block or reduce the proliferation of immune cells and the production of interleukin-2, thus modulating T-cell activity. Obviously, in a mouse which anyway lacks such an activity (and, as a matter of fact, that completely lacks interleukin-2 gamma receptors), one expects to find no difference between the subpopulations of cells. But in a mouse that has such an activity, only cells that are able to disrupt this mechanism can proliferate efficiently.

Assuming that ABCB5+ cells prove to also be more malignant in the case in humans, one might argue that the first model (the least immunocompromised) was a better model. However, Morrison's group would probably point out that this malignancy is not due to the tumorigenicity of the cells *per se*. But *on what ground can one exclude phenomena as part or not of such an abstract property*? On closer inspection, which instrument is the best depends on what it is that we wish to measure – in this context, on the understanding one has of tumorigenicity or CSC-ness.

Tumorigenicity is the capacity to form tumours and sustain growth, but in the presence or in the absence of an immune pressure? One the one hand, human tumours do not develop "in the void": cancer patients are seldom so immunodeficient, and immune response is an important part of cancer development and of variability in outcomes. A notion of tumorigenicity independent of this pressure seems to be an idealization that lacks practical relevance. On the other hand, it seems scientifically worthwhile to isolate the different components influencing the malignancy of cancer cells, so that we might want to exclude the effects of the immune system: tumorigenicity is one thing, evasion of immune surveillance is another. The problem with this reasoning is that many other causally relevant elements (many ways through which some cells might be more tumorigenic than others) could also be excluded. Therefore, one can legitimately ask why excluding this and not other causally relevant elements. The only reasoned

answer one can provide has to be linked to the adoption of a theoretical framework[8] – or what one could call the "theoretical grounding" of operations or instruments.

2.3 CSC and theoretical grounding of operations

In the first paper that I briefly described in the previous section, Quintana et al. (2008) reduce the CSC model to tumorigenicity: "the cancer stem-cell model has suggested that only small subpopulations of cancer cells have tumorigenic potential" (Quintana et al., 2008, p.593). There has been a general tendency, especially in the field of melanoma, to avoid the abstract talk of CSC in favor of the operational talk of melanoma-initiating cells – or cells that initiate melanoma when transplanted into an immuno-deficient mouse. Likewise, participants of the 2011 Working Conference on CSC have explicitly tried to split the conceptual and operational meanings (Valent et al., 2012) in order to avoid a conflation of the two. However, severing the connection between the two is equally problematic. An exclusive focus on tumour-initiating potential would be like a focus on the height of the mercury column: while it might be useful locally, it does not allow semantic extension. The abstract concept does. The CSC framework can potentially mediate between the material contexts. But this means that the problem of selecting the "right" xenograft model can only be solved if one has at least a tentative theoretical understanding of what the instrument should measure.

I believe the CSC framework can succeed where the notion of growth momentum has failed precisely because its meaning has theoretical implications which are not reducible to the operational definition of CSC. The seminal findings of Bonnet and Dick (1997) was not that some leukemic cells were more tumorigenic than others, but precisely that those were the cells possessing stem-cell like characteristics ("the differentiative and proliferative capacities and the potential for self-renewal" Bonnet and Dick 1997, p.730). In doing so, it established a parallel between cancer and normal development, suggesting that the physiological differentiation hierarchy can shed light on the dynamics of cancer. In other words, it also poses additional constraints as to the kind of measurements that ought to be linked to it. Schatton's findings of the modulation of T-cell activity by cancer cells may be extremely relevant for an understanding of cancer, but it has no physiological counterpart and is unrelated to the tissue hierarchy. As such, it is irrelevant to the identification of CSC. It is not biologically or clinically irrelavant, but irrelevant to what it is that the xenograft was supposed to measure. Insofar as the xenografts are used for the identification of CSC, the question of the "right" assay can only be answered with respect to the theoretical meaning of CSC: beyond its operationalizations and with full attention to the theoretical relationships it entertains.

3 Conclusion

There are several differences between thermometry and xenografting, or between thermodynamics and the CSC framework. The fact that physics is quantitative is perhaps the most

---

[8]See also Griesemer (1992) for discussion of how the appropriateness of a tool is necessarily linked to the adoption of a theoretical framework.

important one: ordinal measurements of growth momentum were not quantitative in the sense that they allowed no meaningful arithmetic operation between measurement results. Nevertheless, as I have tried to show, there are important similarities in the epistemological problems encountered, and the comparison can yield insights into the current problem of the choice of xenograft host. But I would now like to conclude with more general observations.

The comparison reveals that Chang's notion of "epistemic iteration" is not limited to quantitative cases. Instruments are often represented as sorting devices from which emerge classifications of reality (see for instance Buchwald 1992). At the same time, entertaining the full multiplicity of instruments (or procedures), and hence of competing classifications, would be counter-productive. What I have tried to highlight is that theoretical frameworks, however preliminary or vague they are, are needed to restrict this plurality. The first step is to *assume* an identity between an operational concept and a theoretical concept, and the importance of such bootstrapping assumptions was also emphasized in other fields such as experimental psychology (Sullivan, 2008; Feest, 2010). From that point on, the process is one of gradual correction of both terms to resolve inconsistencies. In the case of xenograft experiments, the use of increasingly immunodeficient mice, or of humanized mice (Maugeri and Blasimme, 2011), are gradual corrections of this kind. What I have been trying to emphasize, however, is that a lack of convergence is not a failure of the instrument: it is a failure of the whole articulation between instrument and theory.

Until a theoretical framework has been shown to be satisfactory, there is neither an epistemic ground nor even a reasonable motivation for operational monism. And once the theoretical framework is complete, such as in the case of thermodynamics, there is no more the need for operational monism – for the "true" temperature, since the values of one thermometer can be converted to those of another. It is in between these two moments that theoretical terms are the most productive. It is precisely because they are operationally vague, but not as vague as to defy transposition (we could say that they are operationally *suggestive*) that they allow mediation between material contexts. Yet to avoid a trivializing flexibility, their meaning has to be restricted through relations to other theoretical terms.

References

Bechtel, W. and R. C. Richardson (1993). *Discovering complexity: Decomposition and localization as strategies in scientific research*. Cambridge: The MIT Press.

Bhowmick, N. and E. Neilson (2004). Stromal fibroblasts in cancer initiation and progression. *Nature 432*(November), 332–337.

Blasimme, A., P. Maugeri, and P.-L. Germain (2013). What mechanisms can't do: Explanatory frameworks and the function of the p53 gene in molecular oncology. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, in print.

Bonnet, D. and J. Dick (1997). Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nature medicine 3*(7), 730–738.

Buchwald, J. (1992). Kinds and the wave theory of light. *Studies in History and Philosophy of Science Part A 23*, 39–74.

Chang, H. (2004). *Inventing Temperature: Measurement and Scientific Progress* (Oxford University Press ed.).

Civenni, G., A. Walter, N. Kobert, D. Mihic-Probst, M. Zipser, B. Belloni, B. Seifert, H. Moch, R. Dummer, M. van den Broek, and L. Sommer (2011, April). Human CD271-positive melanoma stem cells associated with metastasis establish tumor heterogeneity and long-term growth. *Cancer research 71*(8), 3098–109.

DeVita, V. T. and E. Chu (2008, November). A history of cancer chemotherapy. *Cancer research 68*(21), 8643–53.

Ewing, J. (1919). *Neoplastic diseases, a text-book on tumors*. Philadelphia: W.B. Saunders Company.

Feest, U. (2010). Concepts as Tools in the Experimental Generation of Knowledge in Cognitive Neuropsychology. *Spontaneous Generations: A Journal for the History and Philosophy of Science 4*(1), 173–190.

Funk, C. (1915). The transplantation of tumors to foreign species. *The Journal of Experimental Medicine XXI*.

Germain, P.-L. (forthcoming). From Replica to Instruments: Animal Models in Contemporary Biomedical Research. *History and Philosophy of the Life Sciences*.

Greene, H. S. (1951). A conception of tumor autonomy based on transplantation studies: a review. *Cancer research 11*(12), 899–903.

Greene, H. S. (1952). The significance of the heterologous transplantability of human cancer. *Cancer*, 24–44.

Greene, H. S. N. (1948). Identification of malignant tissues. *JAMA : the journal of the American Medical Association 137*(16), 1364–1366.

Griesemer, J. R. (1992). The Role of Instruments in the Generative Analysis of Science. In *The right tools for the job: at work in the Twentieth century life sciences*, pp. 47–76.

Hanau, A. (1889). Erfolgreiche experimentelle ubertragung von karzinom. *Fortschr Med*, 5–12.

Hekzog, M. (1902). On tumor transplantation and inoculation. *The Journal of medical research 1*(6), 74–84.

ILAR and NRC (1998). Biomedical Models and Resources: Current Needs and Future Opportunities. Technical report, Committee on New and Emerging Models in Biomedical and Behavioral Research, Institute for Laboratory Animal Research, Commission on Life Sciences, National Research Council.

Loeb, L. (1945). *The biological basis of individuality*. Springfield, III., Baltimore, Md: Charles C. Thomas.

Maugeri, P. and A. Blasimme (2011). Humanised models of cancer in molecular medicine: the experimental control of disanalogy. *History and philosophy of the life sciences 33*, 603–622.

Mayet, M. (1902). Production du Cancer chez les Rats blancs par Introduction dans leurs economies des Substances constituantes des Tumeurs malignes de l'Homme. *Gazette hebdomadaire de médecine et de chirurgie 6*.

Morange, M. (1997). From the Regulatory Vision of Cancer to the Oncogene Paradigm, 1975-1985. *Journal of the History of Biology 30*, 1–29.

Murgia, C., J. Pritchard, S. Kim, A. Fassati, and R. Weiss (2006). Clonal origin and evolution of a transmissible cancer. *Cell 126*, 477–487.

Novinski, M. (1876). Zur frage über die impfung der krebsigen geschwulste. *Zentralbl. Med. Wissensch. 14*, 790–791.

Quintana, E., M. Shackleton, M. S. Sabel, D. R. Fullen, T. M. Johnson, and S. J. Morrison (2008). Efficient tumour formation by single human melanoma cells. *Nature 456*(7222), 593–598.

Rheinberger, H.-J. (1997). *Toward a History of Epistemic Things: Synthesizing Proteins in the Test Tube*. Writing science; Variation: Writing science. Stanford University Press.

Schatton, T., G. F. Murphy, N. Y. Frank, K. Yamaura, A. M. Waaga-Gasser, M. Gasser, Q. Zhan, S. Jordan, L. M. Duncan, C. Weishaupt, R. C. Fuhlbrigge, T. S. Kupper, M. H. Sayegh, and M. H. Frank (2008). Identification of cells initiating human melanomas. *Nature 451*(7176), 345–349.

Schatton, T., U. Schütte, N. Y. Frank, Q. Zhan, A. Hoerning, S. C. Robles, J. Zhou, F. S. Hodi, G. C. Spagnoli, G. F. Murphy, and M. H. Frank (2010). Modulation of T-cell activation by malignant melanoma initiating cells. *Cancer Research 70*(2), 697–708.

Snell, G. D. (1964). The terminology of tissue transplantation. *Transplantation 2*, 655.

Sullivan, J. a. (2008). The multiplicity of experimental protocols: a challenge to reductionist and non-reductionist models of the unity of neuroscience. *Synthese 167*(3), 511–539.

Towbin, A. (1951). The heterologous transplantation of human tumors. *Cancer Research 11*, 716–722.

Valent, P., C. Eaves, D. Bonnet, R. De Maria, T. Lapidot, M. Copland, J. V. Melo, C. Chomienne, F. Ishikawa, J. J. Schuringa, G. Stassi, B. Huntly, H. Herrmann, J. Soulier, A. Roesch, G. J. Schuurhuis, S. Wöhrer, M. Arock, J. Zuber, S. Cerny-Reiterer, H. E. Johnsen, and M. Andreeff (2012, October). Cancer stem cell definitions and terminology: the devil is in the details. *Nature Reviews Cancer 12*(11), 767–775.

van Fraassen, B. C. (2008). *Scientific representation: Paradoxes of perspective*. Clarendon Press.

Visvader, J. and G. Lindeman (2012, June). Cancer Stem Cells: Current Status and Evolving Complexities. *Cell Stem Cell 10*(6), 717–728.

Witkowski, J. A. (1983). Experimental pathology and the origins of tissue culture: Leo Loeb's contribution. *Medical History 27*(03), 269–288.

Zondek, B. (1928). Die Schwangerschaftsdiagnose aus dem Harn durch Nachweis des Hypophysenvorderlappenhormons. *Die Naturwissenschaften 51*, 1088–1090.