Data quality and personalized medicine

The collection, processing and banking of biospecimens

by Carolyn Compton, MD, PhD, FACP

ABSTRACT

The vision of personalized medicine begins with the collection, processing and banking of biospecimens. The quality of the specimens depends on a great number of pre-analytical factors. Tumour heterogeneity presents further challenges. Careful control of and information about variables that affect biospecimen data would enable rapid progress toward personalized medicine.

— Presented at the Personalized Medicine in the Management of CRC conference.

The vision of personalized/precision medicine from the point of view of the biobanker starts and ends with the patient: biospecimens (containing the biomarkers) are collected, processed and stored, then distributed to the investigator for analysis. Knowledge gained from the analysis of molecular data comes back to the patient in the form of new diagnostics and treatments. It is clear that the quality of what occurs during the collection/processing/banking phase of the process determines the quality of data obtained.

TUMOUR HETEROGENEITY THREATENS PROGRESS

Some utopic views anticipate scenarios in which outpatients will undergo a needle biopsy of a tumour and, within a short time, an active treatment will be devised for each patient on the basis of the tumour characteristics. The reality may not be so simple. Whether tumour tissue is used to manage patients or to study cancer biology, the heterogeneity of tumours (both innate and iatrogenic) poses a serious problem. Among the challenges:

• Most mutations found in single biopsies are not found in other regions of the same tumour.
• Favourable and unfavourable molecular profiles may be found in different parts of the same tumour.
• Widespread aneuploidy (alterations in the total number of chromosomes) — the sentinel feature of malignancy — is common.
• Different areas of the same tumour have varying mutations in the same genes.
• Metastases differ from each other and from the primary tumour.
• Treatment creates evolutionary pressures that increase heterogeneity.

This heterogeneity is in part responsible for our currently unsustainable situation in terms of the translational drug pipeline, which is characterized by massive attrition, long duration and staggering costs. It can take 12 years to develop a single new cancer drug, at an average cost of at least US$4 billion and perhaps as high as $11 billion. Further, the thousands of compounds being studied over this time may end with only one drug being approved by the US Food and Drug Administration (FDA). Time and attrition can be directly related to the lack of validated biomarkers.

IMPACT OF BIOMARKER IDENTIFICATION AND VALIDATION

The vision of precision medicine relies on biomarkers, defined as measurable characteristics that indicate a biologic state or condition. They must be quantifiable, reproducible and clinically relevant, and their potential for use in the following areas is extraordinary:

• Therapy development (markers of efficacy, toxicity)
• Early detection (broad or specific detection/corroboration of specific disease stage)
• Rational choice of treatments (patient stratification)
• Assessment of treatment effectiveness
• Prognosis (outcome), prediction (response to therapy)
• Prevention, surveillance
• Treatment/disease monitoring
• Surrogate endpoints for clinical trials that will predict activity of drug and impact on patient’s disease

In the current literature, there are an estimated 150,000 papers documenting thousands of claimed biomarkers, yet the number of biomarkers in routine clinical medicine is thought to be only 100.1 Several factors contribute to the spectacular failure rate in biomarker identification, verification and qualification, and the corresponding failure rate and expense in clinical trials and drug development. These include, in particular, the low reproducibility of academic publications related to poor access to rigorously annotated, fit-for-purpose biospecimens from stringently phenotyped sources (requiring lots of data not currently collected and provided

Carolyn Compton, MD, PhD, FACP is Chief Science/Medical Officer with the National Biomarkers Development Alliance, and Professor, School of Life Sciences, BioDesign Institute and Complex Adaptive Systems Initiative, at Arizona State University.
to investigators); and insufficient control of preanalytical parameters.

**DATA LACKING ON PREANALYTIC VARIABILITY**

The quality of biospecimens stored in biobanks and used to drive research is determined by the stringency/requirements of the analysis to be performed, the specific platform used for analysis, and the type, class and lability of the molecular species of interest. It is also determined by how carefully the chain of access, collection, processing and storage is driven by evidence-based standard operating procedures, and verified by quality control procedures and a total quality management system.

Ransohoff and Gourlay highlighted sources of bias that are relevant to specimen biobanking, storage, handling and analysis, and these reflect areas that can be approached in a prospective fashion and “fixed.”

In the life cycle of a biospecimen (Figure 1), many preanalytic factors affect the biospecimen’s molecular composition and integrity before it ever enters the research or clinical laboratory. Using CRC as an example, a specimen from a colectomy is first of all a part of a patient’s anatomy, and the patient may be exposed to many drugs and other factors before the operation. In-surgery variables such as handling of the specimen, stretching of nerves, clamping of arterial vessels and creation of anoxia also play a role before the pathologist even acquires the sample (Time 0). After acquisition by the pathologist, many other factors can affect the quality of the sample, e.g. time at room temperature, temperature of room, type of fixative used, time in fixative, rate of freezing etc.

Many people and processes contribute to this variation, and yet there is no requirement in clinical medicine to control or even record these variables. It can often be days before samples are recorded. We know more about grades of beef in the supermarket (which are regulated) than about the quality of most biosamples that come out of the clinical system!

Straightforward analysis of upregulation/downregulation of a simple protein biomarker for breast cancer (HER2) using immunohistochemistry (IHC) and fluorescence in situ hybridization (FISH) shows that the same specimen left out for 2 hours goes from a positive (which would make
patient eligible for lifesaving trastuzumab) to a negative biopsy result, in which case the patient would be denied the drug. High rates of false positives and false negatives due to preanalytical variation were recognized 5 years after trastuzumab entered the market; in precision medicine, it is clear that we cannot afford this kind of heterogeneity and misrepresentation. In another study of colon cancer, pMAPK IHC analysis demonstrated the opposite: a gain of biomarker signal with time to fixation. Without knowledge of tissue processing methods and assurance of rapid tissue fixation, protein expression data are unreliable and it is impossible to understand pathway activity. This is going to be a serious problem as we move more towards therapies that are targeted to labile proteins.

In postoperative tissue ischemia in colon cancer, patterns of gene expression are not predictable; even proteins thought of as stable, e.g. CEA and CK20, do in fact change rapidly over just 30 minutes.

In an experiment we conducted in Germany, where 2 surgeons did exactly the same colectomy procedure in the same hospital, we looked at groups of patients in whom the clamp time (between clamping of arterial vessel and removal of specimen and immediate stabilization) during surgery was known and recorded. Gene expression studies and clustering analysis showed that the strongest driver for clustering gene expression was the time the tumour spent anoxic in the operative site — another variable that is not recorded.

Mass spectroscopy analysis also shows extreme heterogeneity at the protein level: differences in expression of proteins can vary by as much as 40% depending on whether the sample was obtained from the peripheral or central region in the tumour, and again, this is information that is not usually available.

Blood is easily accessible and commonly banked, and is also affected by collection, processing and storage, e.g. factors such as tourniquet time, patient position, type of tube and centrifuge speed, which are all known to affect assay results.

**Getting the right stuff**

Making the transition from the era of single analytes, like IHC for a single protein, to an era of multianalytes and “omics” and network and systems analysis, will raise the bar for biospecimen quality even further. As the landscape becomes more challenging and our understanding of the complexity of cancer grows, we have to make sure that the data we are collecting are robust. There will be a need for a systems-based approach and a scientific rigour that biobanking has never before been subjected to. Despite the great advances in technology, it all starts with ensuring the quality of biospecimens — getting the “right stuff.”

**An opportunity for Canada**

If Canada proceeds to start a national biobank in CRC, there is a possibility of doing it right prospectively and perhaps being the first in the world to build a biobank fit for the era of personalized medicine and rapid progress. Sources of data a biobank could centralize and make available to investigators encompass data on preacquisition variables, postacquisition collection, process, storage, quality control and quality assurance, pathologic diagnosis verification, macro- and microscopic imaging, molecular analysis from samples, and clinical data about patients’ outcomes, which tells what the molecular features mean. Finally, without broad consent for unspecified future use, the best specimens and data cannot serve science, so paying attention to ethical and legal issues surrounding biobanking is paramount.

**References**