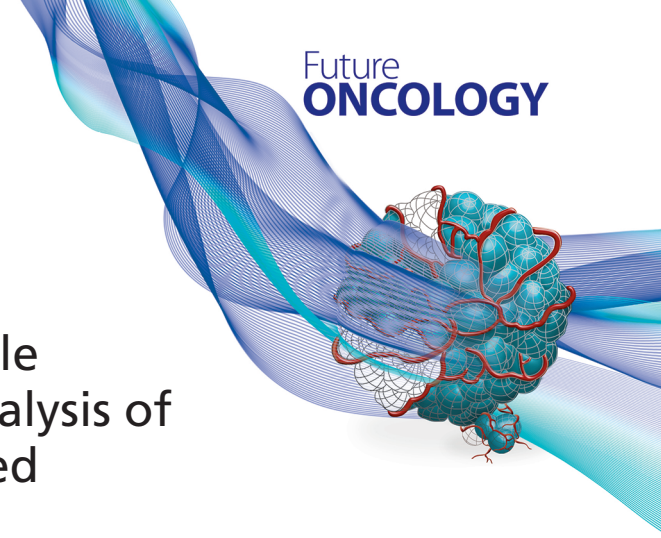
Commentary

For reprint orders, please contact: reprints@futuremedicine.com



Developing a growing cancer profile database based on quantitative analysis of protein biomarkers in formalin-fixed paraffin-embedded specimens

Jiandi Zhang*,[1] & Maozhou Yang[1]

[1]Quanticision Diagnostics, Inc., 400 Park Offices Drive, Room 110, Research Triangle Park, NC 27709, USA

*Author for correspondence: jiandi.zhang@outlook.com

Over a century of clinical practice has led to the accumulation of millions of archived formalin Fixed Paraffin Embedded (FFPE) cancer specimens with detailed medical records worldwide. Absolute quantitation of clinical protein biomarkers in these FFPE specimens allows individual specimens to be profiled at the population level, with the absolute nature of the measurements enabling the continuous processing of archived FFPE specimens over the time. A continuously growing cancer patient profile database is proposed here to support "big data" profiling of these protein biomarkers alone or in combination, enabling next-generation retrospective-prospective analytics into the field of clinical diagnostics.

First draft submitted: 13 May 2020; Accepted for publication: 9 July 2020; Published online: TBC

Keywords: big data • FFPE • profiling • QBD • quantitative diagnostics

Cancer heterogeneity exists at multiple levels, most evidently at the intra- and inter-tumor levels; in a certain sense, no two tumors are the same[1]. In addition, there are 17 million new cancer cases diagnosed worldwide each year [2]. For breast cancer alone, over 1.4 million new cases are reported each year. The heterogeneity of each tumor poses significant challenge for cancer treatment. Thus, clinical studies limited to only hundreds to thousands of cases have difficulty adequately representing the millions of different molecular identities. Consequently, under- or overtreatment becomes a common issue for cancer patients. Clearly, clinical studies on a much greater scale are needed to better represent the millions of different molecular identities at the population level, so that we may achieve the ultimate goal of personalized medicine for cancer patients.

For the majority of cancer patients, tumor tissues are surgically removed and archived in a formalin-fixed paraffin-embedded (FFPE) format in hospitals or other medical institutes. As a result, millions of archived FFPE specimens, accompanied by detailed medical records, including treatments administered and ensuing clinical outcomes, have accumulated as an enormous yet underutilized resource. The sheer number of these FFPE specimens allow comprehensive coverage of molecular identities at the individual level. It is reasonable to assume that, when combined with their known clinical outcomes, these specimens will become an unrivaled resource for clinical studies geared toward personalized medicine, where we may identify known cases with similar molecular identities for every single cancer patient in the world.

Quantitative & absolute

However, because of the limitations of current available techniques, especially at the protein level, this valuable resource has not been fully appreciated in clinical studies. One major challenge for studies using FFPE specimens is the heavy crosslinking of the proteins in these specimens. This remains an insurmountable challenge to Enzymatic linked Immunosorbent Assay (ELISA), as it significantly limits the amount of protein available for detection [3].

Other methods of protein analysis, including Immunohistochemistry (IHC), Western blot analysis [4], Reverse phase protein microarray (RPPA) [5] and mass spectrometry (MS) [6], have also been used to analyze FFPE specimens. Nonetheless, they become inadequate for evaluating the enormous quantity of FFPE specimens required for the outlined study: for both IHC and Western blot analysis, the qualitative nature of these two methods obscures

individual differences at the population level. For example, IHC, the primary method in daily clinical practice, categorizes the entire patient population into only 4 groups (0/1+/2+/3+) when used to assess protein levels of HER2, a commonly used biomarker for breast cancer patients. Yet, as indicated by several studies including our own, HER2 levels vary over 100-fold, even within the 3+ group alone [3,7].

Other methods may measure protein levels quantitatively to reveal individual differences at the population level, but offer relative results to limit the scale of the study. We may use an example to illustrate this limitation better. The expression level of a protein may be expressed in absolute terms (e.g., nmol/g) or in relative terms (% of a reference protein B). Although we can compare the protein levels in absolute terms easily across multiple analyses, it is harder to compare the results from analyses with varying levels of reference protein B in each analysis.

This is the case forMS and RPPA, in which results are expressed as values relative to a reference protein, which may vary for each individual study [5,8]. Thus, the scale of the studies based on these methods is limited by the number of specimens included in a single study and is unable to be expanded upon by incorporating results from other studies. The same also holds true for datasets generated using the Real-time PCR (RT-PCR) method.

Therefore, we believe that the method for accommodating the enormous amount of FFPE specimens must be capable of yielding datasets of a **continuous** and **absolute** nature. With respect to continuity, quantitative assessments are needed to distinguish the subtle differences among individual FFPE specimens at the population level; with respect to absoluteness, the quantitation of individual proteins should be consistent, regardless of location, timing, etc., to ensure that data can be reliably shared, cross-examined and combined to offer the much needed growth of the dataset to accommodate the enormous amount of FFPE specimens.

In this regard, the only two methods to meet this requirement so far are Selective reaction monitoring-mass spectrometry and Quantitative Dot Blot (QDB). Selective reaction monitoring-mass spectrometry has been reported to measure EGFR in non-small-cell carcinoma and HER2 in breast and gastric cancers [9]. The QDB method, by contrast, has been used to successfully measure HER2, Ki67, ER, PR and cyclin D1 levels in breast cancer [3,10,11].

A retrospective-prospective FFPE cancer profile database based on absolute quantitation of protein biomarkers

We propose here the development of an unprecedented retrospective-prospective FFPE cancer profile database or, more accurately, databases of different cancer types (breast, colorectal or prostate cancer) based on absolutely quantified protein biomarkers to take full advantage of the vast amount of FFPE specimens. Currently, the number of protein biomarkers assessed in routine practice varies depending on the cancer type. For breast cancer, around 20 protein biomarkers have been routinely used in clinical practice. When measured quantitatively and absolutely, their combinations are sufficient to differentiate individual FFPE specimens from millions of archived FFPE specimens. In a certain sense, the combination of these protein biomarkers can become a unique fingerprint for each FFPE specimen in the database.

This unique fingerprint can be used as a nucleus to anchor the matching clinical records, including the traditional clinicopathological factors, the treatments administered and the ensuing clinical outcomes, for a holistic picture of each FFPE specimen. All the information from these various aspects constitutes a cancer profile for every FFPE specimen in the proposed database. Any other clinical-related information may also be included in these cancer profiles. For example, genetic information, including small nucleotide variations, chromosomal alterations and scores of various genetic predictor assays, can be included in the cancer profiles.

The absolute nature of the database will ensure the continuing growth of the database. Cancer profiles, although from different sources, may be combined efficiently because of the absoluteness of the data. New cancer profiles will also be added and supplemented over the time. Over the years, this database is expected to accommodate a fair share of these archived FFPE specimens to assist the big data–supported clinical diagnosis.

Future applications of the retrospective FFPE database

The introduction of this database should have instant impact on both clinical research and practice. We now present several potential uses of this proposed database.

Proximity-based diagnostic method

The current diagnostic system is at the precision stage to separate cancer patients into a few subtypes with customized treatments. However, with widespread tumor heterogeneity existing at all levels, this system is far from meeting the individual needs of cancer patients. By contrast, considering the enormous number of cancer profiles included

in the proposed database, a novel diagnostic technique called the proximity-based diagnostic method is proposed [Zhang J, unpublished data; DOI: 10.17605/OSF.IO/2B4RE] [12]. It is based on the idea that for any new patient, there should be a group of past patients of high similarity. By comparing simultaneously the expression levels of a set of absolutely quantitated protein biomarkers measured from a new patient with those of every cancer profile in the database, a group of cancer profiles of high similarity to the new patient will be identified from millions of cancer profiles in the database. Analyzing the clinical outcomes and treatments administered to this group of cancer profiles should lead to a personalized prognosis and treatment plan for the new patient. Although the set of protein biomarkers used in this method remains to be defined through extensive outcome analysis of the enormous amount of cancer profiles in the proposed database, the concept of a proximity-based method represents a major step in clinical diagnosis.

Virtual randomized clinical trials

Prospective randomized clinical trials (RCTs) identify participants of the trial first and then follow their clinical outcomes over time. Retrospective RCTs, by contrast, examine participants with known clinical outcomes. Thus, prospective RCTs are considered the gold standard of clinical research, as participants are tightly controlled to maximally reduce the biases of the trial. However, prospective RCTs have obvious drawbacks, including high cost and requiring a long time to complete. Meanwhile, retrospective studies frequently lack stringent controls to ensure credibility of the conclusions. To accelerate the process, the concept of a retrospective-prospective clinical trial was proposed [13]. However, with the introduction of this enormous proposed database, one may develop a study protocol defined by vigorous requirements and clear aims and then search the database accordingly to identify those specimens meeting the predefined requirements. The documented clinical outcomes of these specimens can be used to test the hypothesis for clinical guidance. In other words, a virtual RCT may be achieved entirely with this proposed database.

In fact, as the proposed database expands to a certain size in the future, the virtual RCT would eventually become a real-world test of various hypotheses, without having to worry about various biases inherently associated with other RCTs. The conclusions from virtual RCTs may also be verified periodically with the continuing expansion of this database.

Accelerating the drug development process

For drug discovery, the proposed database also translates into significant savings in terms of both costs and developmental efforts. Although developing this database requires a collective global effort to obtain FFPE specimens with detailed medical records, the database offers comprehensive resources to accelerate the drug development process. Drug leads learned from cellular and animal studies may be verified directly or indirectly through analyzing expressions of various protein biomarkers, which helps avoid wasting resources during the preclinical stage. The unexpected prognostic/predictive roles of a biomarker may also be identified through extensive data mining. Companion diagnostic kits, which require RCTs currently, may also be developed virtually through evaluating the predictive role of biomarkers via outcome analyses of those cancer profiles receiving a specific treatment.

This database may also be helpful in several other aspects. For example, the database can be used to continuously re-evaluate and update existing clinical guidelines. It will also enable periodic re-evaluation of a specific drug or treatment plan through outcome analysis after approval by regulatory agencies.

Limitations & challenges of the proposed database

One major issue with this proposed database is that only routinely used protein biomarkers may be absolutely quantitated.In the case of breast cancer, no more than 15 protein biomarkers may be measured in our lab. Incidentally, 5–10 protein biomarkers are checked for individual patients in daily clinical practice. Obviously, the current database is more than sufficient to serve the field of clinical diagnostics for breast cancer patients. However, this issue significantly limits the use of this database for research of an exploratory nature. Hopefully, with improved technology and increased availability of high-quality antibodies, we can measure hundreds to thousands of different proteins in FFPE specimens to gain a better understanding of the underlying comprehensive protein networks leading to tumorigenesis at the population level.

This proposed database also requires a constant infusion of new information from other areas of clinical research to thrive. For example, it needs help from proteomic research to expand the list of protein biomarkers included in

the database. New drugs and treatment strategies need to be introduced into this database whenever available to upgrade the database continuously over time.

We also anticipate the enormous challenge of maintaining the consistency of the database. As the first database with protein biomarker levels measured in absolute and quantitative values, its consistency may easily be derailed by various factors. These factors include the quality of the FFPE specimen itself; the quality of the assay reagents, especially the protein standards used in the analysis; and how strictly Standard Operation Procedures (SOP) are executed. Extensive efforts must be devoted to these areas to minimize their impact on the overall quality of the database.

It should be mentioned that the proposed database, even with only thousands of cancer profiles available during the early stages, should be useful to aid current clinical research and practice. However, its power can only be fully appreciated when millions of cancer profiles are included in the database. To fulfill the goal of virtual RCTs, a global collaboration may be required in the future. How to encourage this transition would be another challenge for clinicians worldwide.

In conclusion, a growing cancer profile database with millions, or even tens of millions, of cancer profiles is proposed to make a profound impact on future clinical research and practice. We would like to suggest the name *'Quantitative Diagnostics'* to describe this new and exciting area in the field of clinical oncology.

Acknowledgments

The authors wish to thank the editor for his constructive comments as well as B Dou, G Chen and J Zhang for their helpful comments and assistance in writing the article.

Financial & competing interests disclosure

J Zhang is the founder of Quanticision Diagnostics Inc. and has filed several patents with regard to the QDB method and its application in cancer diagnosis.

No writing assistance was utilized in the production of this manuscript.

References

1 Roulot A, Héquet D, Guinebretière J-M *et al.* Tumoral heterogeneity of breast cancer. *Ann. Biol. Clin. (Paris)* 74(6), 653–660 (2016).

2 IARC. World cancer report. (2014).

3 Yu G, Zhang W, Zhang Y, Lv J, Zhang J, Tang F. Developing a routine lab test for absolute quantification of Her2 inFormalin Fixed Paraffin Embedded (FFPE) breast cancer tissues usingQuantitative Dot Blot (QDB) method. *bioRxiv* 584615 (2019).

4 Mansour AG, Khalil PA, Bejjani N, Chatila R, Dagher-Hamalian C, Faour WH. An optimized xylene-free protein extraction method adapted to formalin-fixedparaffin embedded tissue sections for western blot analysis. *Histol. Histopathol.*32(3), 307–313 (2017).

5 Boellner S, Becker K-F. Reverse Phase Protein Arrays-QuantitativeAssessment of Multiple Biomarkers in Biopsies for Clinical Use. *Microarrays (Basel)*4(2), 98–114 (2015).

6 O'Rourke MB, Padula MP. Analysis of formalin-fixed, paraffin-embedded(FFPE) tissue via proteomic techniques and misconceptions of antigen retrieval. *BioTechniques* 60(5), 229–238 (2016).

7 Nuciforo P, Thyparambil S, Aura C *et al.* High HER2 protein levelscorrelate with increased survival in breast cancer patients treated withanti-HER2 therapy. *Mol Oncol* 10(1), 138–147 (2016).

8 DeSouza LV, Siu KWM. Mass spectrometry-based quantification. *Clin. Biochem.* 46(6), 421–431 (2013).

9 Steiner C, Tille J-C, Lamerz J *et al.* Quantification of HER2 byTargeted Mass Spectrometry in Formalin-Fixed Paraffin-Embedded (FFPE) BreastCancer Tissues. *Mol Cell Proteomics* 14(10), 2786–2799 (2015).

10 Hao J, Lv Y, Zou J *et al.* Hao J, Lv Y, Zou J et al. Improving prognosis of surrogate assayfor breast cancer patients by absolute quantitation of Ki67 protein levelsusing Quantitative Dot Blot (QDB) method. *medRxiv* doi:2020.03.11.20034439 (2020).

11 Hao J, Zhang W, Lv Y *et al.* Combined use of absolutely quantitated cyclin D1 and Ki67 protein levels to improve prognosis of Luminal-like breast cancer. *medRxiv* doi:2020.04.15.20066993 (2020).

12 Zhang J. Proximity-based diagnostic method to provide personalizedtreatment for cancer patients. (2020).

13. Simon RM, Paik S, Hayes DF. Simon RM, Paik S, Hayes DF. Use of Archived Specimensin Evaluation of Prognostic and Predictive Biomarkers. 101(21), 1446–1452 (2009).