Published in final edited form as:

Anal Chem. 2012 January 17; 84(2): 1063–1069. doi:10.1021/ac2026496.

Analysis of Variance in Spectroscopic Imaging Data from Human Tissues

Jin Tae Kwak^{†,§}, Rohith Reddy^{†,§}, Saurabh Sinha^{*,†}, and Rohit Bhargava^{*,§,¶}

[†]Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

[§]Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA

[§]Department of Bioengineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

[¶]Department of Mechanical Science and Engineering and Electrical and Computer Engineering, Micro and Nanotechnology Laboratory and University of Illinois Cancer Center, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

Abstract

The analysis of cell types and disease using Fourier transform infrared (FT-IR) spectroscopic imaging is promising. The approach lacks an appreciation of the limits of performance for the technology, however, which limits both researcher efforts in improving the approach and acceptance by practitioners. One factor limiting performance is the variance in data arising from biological diversity, measurement noise or from another source. Here we identify the sources of variation by first employing a high throughput sampling platform of tissue microarrays (TMAs) to record a sufficiently large and diverse set data. Next, a comprehensive analysis of variance (ANOVA) model is employed to analyze the data. Estimating the portions of explained variation, we quantify the primary sources of variation, find the most discriminating spectral metrics, and recognize the aspects of the technology to improve. The study provides a framework for the development of protocols for clinical translation and provides guidelines to design statistically valid studies in the spectroscopic analysis of tissue.

Introduction

Fourier transform infrared (FT-IR) spectroscopic imaging¹ provides simultaneous chemical and structural information from heterogeneous materials of interest² and is being used increasingly for biomedical studies, especially involving cells and tissues.^{3, 4, 5, 6, 7, 8} Most biomedical samples, however, are chemically complex. Hence, their analysis often relies on treating the spectrum as a characteristic signature of the identity and/or physiologic state of the sample. Many studies seek to find the unique spectral signature or differences in spectral signatures between given classes of samples from a statistical, rather than purely biochemical, perspective. These classes may be tissue with different grades of disease or different cell types within the same tissue type, for example. Finding an IR imaging-based

*Corresponding Author. Rohit Bhargava: rxb@illinois.edu; address, 4265 Beckman Institute 405 N. Mathews Avenue, University of Illinois at Urbana-Champaign, Urbana IL 61801.

ASSOCIATED CONTENT

Supporting Information. Description of ANOVA, ANOVA models, tables, and symbols, samples and data preparation, and measurement error estimation.

approach that can distinguish between disease states is of tremendous technological and medical importance as it can potentially improve diagnostic information, reduce costs and prevent errors. The tasks in this approach would be to discover differences in spectral properties of classes and develop a computer algorithm such that every spectrum (pixel) can be classified into a particular class without using dyes, stains or human supervision.⁹ Though conceptually straightforward, this approach is exceptionally challenging not only because of the subtle differences between various components and disease states in tissue but also because of the variation in IR spectra that obscures differences between disease states. This variation may overwhelm differences due to disease states and is a prime cause of the failure of many analytical methods in providing robust diagnostic protocols. Quantification of the sources of analytic variability and redressing them, hence, are topics of much interest in IR spectroscopy¹⁰ and other analytical technologies.^{11,12,13}

Analytic variability can arise from (a) noise in signal measurement,^{10,14} (b) differences within the tissue that leads to differences both within a given sample and between samples from the same patient, (c) differences between patients due to biologic diversity, (d) differences due to sample handling in different clinical settings or research groups and (e) causes not falling into any of the above categories. The variation may also be understood to be biological, technical or residual. Biological variation arises from different biological characteristics of samples such as patients, tissues, cells, subcellular components, etc. It is natural and expected variation, and often of interest in an experiment. Technical variation is attributable to both sample preparation and analytical techniques. Potential sources of technical variation include tissue acquisition,^{15,16} fixation,¹⁷ and sectioning, placement of tissue section on the slide¹⁶ and post-preparation handling.¹⁸ The process of data acquisition also introduces variation, such as measurement noise.¹⁹ Although thoroughly identified, these potential sources of variation may not completely explain the total variation in a measurement. Residual variation refers to the unexplained variation in the experiment; for example, environmental conditions – room temperature and humidity – that may not be part of the sample or acquisition characteristics. Accordingly, residual variation will usually be present and, on occasion, can have a substantial impact on the analysis. In such a case, we may either re-examine potential sources of variation and/or re-design the experiment.

Understanding the relative importance of each of these factors and explaining the variance observed in large scale tissue studies is critical for developing any real-world application. While an understanding of the contributions of variance by various sources can result in improved protocol designs, the lack of such understanding brings into question the performance of any developed protocol²⁰. Hence, in this manuscript, we develop a framework to understand variability and its sources in IR spectroscopic imaging of tissue. This understanding may be extended to other analytical techniques and imaging modalities, in general, and may be used to improve the practice of IR spectroscopic imaging for biomedical analysis, in particular. The first challenge to understanding variability is to obtain a data set of sufficient diversity and size. Tissue microarrays (TMAs),²¹ to this end, are an excellent tool and have been used previously in a number of studies.^{22, 23, 24, 25} TMAs consist of many tissue samples arranged in a grid pattern (Fig. 1), in which multiple samples may be included from the same person and a population of different people is included. Multiple TMAs may further be employed to increase sample set diversity and size, both in the populations of patients as well as clinical settings and handling of samples. The second challenge is to quantify the effect of sources by determining their contribution to the total variance, which can be accomplished by applying analysis of variance (ANOVA) models to the acquired data set.

ANOVA is a popular statistical model for partitioning the total variance of the measured quantity in an experiment into various identifiable factors (or sources of variation). Consider

a TMA consisting of tissue samples from several patients, and n samples were taken from each patient. Here, patient is the only factor, and the difference between patients in terms of IR spectra is of interest. In this setting, the IR absorbance or any other combination of spectral features, which we term “metric”, can be expressed as $y_{jk} = \mu + \beta_j + \varepsilon_{jk}$ where y_{jk} is IR metric of the k th sample from the j th patient, μ is the overall mean, β is patient effect ($j = 1, \dots, n_p$), and ε is residual error effect. The total variance can be partitioned into two, $\sigma_{total}^2 = \sigma_{\beta}^2 + \sigma_{\varepsilon}^2$ where σ_{β}^2 and σ_{ε}^2 indicate the variance of patient effect and residual effect, respectively. Partitioning of variance can be carried out by computing sum of squares (SS) and mean squares (MS). The total SS is calculated as the sum of between-patient SS, sum of squared differences between the overall mean and patient means and within-patient SS (or residual SS), sum of squared differences between patient means and individual metric. Calculating MS, which is SS divided by degrees of freedom (df), the variances can be estimated by equating MS and expected mean square (EMS). EMS of patient effect and residual effect are $\sigma_{\varepsilon}^2 + n\sigma_{\beta}^2$ and σ_{ε}^2 , respectively. Dividing the estimated variances by the total variance allows us to obtain the portion of variance explained by patient effect. The larger the portion of variance due to patient effect, the bigger is the difference between IR metrics of the patients due to a characteristic of the patients themselves. Further, the significance of the differences can be assessed by conducting a hypothesis test, F -test. F -test statistic, which is the ratio of between-patient MS and within-patient MS, is computed, and is compared to the F -distribution with between-patient and within-patient df, resulting in p -value for the test. A low p -value denotes that the metric difference between patients is statistically significant. We note that the model can be extended to reflect additional variables; for example, including histologic class, the model becomes $y_{jkl} = \mu + \beta_j + \delta_l + \beta\delta_{jl} + \varepsilon_{jkl}$ where δ is histologic class effect ($l = 1, \dots, n_{\delta}$) and $\beta\delta$ is the interaction effect between patient effect and histologic class effect. Two factors interact if the effect of one factor changes with changes in contributions from the other factor. Both β and δ are designated as main effect, which is the effect of a factor averaged across the levels of other factors (see Supporting Information for details). Other analogous models, for example a *no-subcellular component model* and *within-histologic class model* (Table 1), can be constructed by adding measurement error and replacing patient and histologic class with core and subcellular component, respectively in these cases. ANOVA has been applied for analyzing several types of spectroscopic imaging data: chemical compounds^{26, 27}, collagen types²⁸, skin lesions²⁹, and plant species^{30, 31, 32}, but, to our knowledge, has not been applied to spectroscopic imaging data from tissues. To systematically apply this methodology for tissue analysis, we present appropriate ANOVA models (Table 1) for different experimental designs of IR imaging data from TMAs, evaluate the statistical significance of the sources of variance, estimate variance contributions of the identified sources, and quantify the relative contributions of the sources to the total variation in the data. Finally, after examining the effect of the sources of variance, we also find the most discriminative spectral metrics and address the aspects of FT-IR imaging and TMA techniques that can be improved for better diagnostic protocols.

RESULTS AND DISCUSSIONS

Four TMAs of prostate tissue samples from different sources and five ANOVA models (Table 1) are employed to examine various sources of variation in FT-IR-TMAs. The details of the TMA preparation and data acquisition and ANOVA models are available in Supporting Information.

Variance component analysis identifies discriminative metrics for histologic analysis

Three TMAs (i, ii, iii) were used in this experiment. From each TMA, 26 sample cores from 13 patients, containing a sufficiently large number of both epithelial and stromal pixels (>200), were selected, and 200 pixels were randomly chosen from each histologic class in a core. This data selection is necessary to eliminate bias that may arise from specific sets or patients with unequal representation. The histologic segmentation was conducted by a Bayesian classifier,²⁵ built on 18 spectral metrics and achieved >0.9 area under the curve (AUC) for the receiver operating characteristic (ROC) curve for most cell type classification. Although several histologic classes are present in these samples, we only consider epithelium and stroma as these are the two major functional cell types important in diagnosing prostate cancer, for simplicity afforded by the small model and to prevent data imbalance as some classes are not present in all samples. While epithelium may be expected to be rather uniform in chemical content, the stroma collectively consists of many cell types; hence, the within-class heterogeneity in stroma is likely to be much greater. Thus, the final reason for choosing this 2-class model is to examine both biochemically homogeneous and heterogeneous cellular populations. Using ANOVA table (Table S-2) for *between-histologic class model*, the portions of total variance due to the associated factors (Fig. 2A) were computed. An example of ANOVA table computation for one metric is shown in Table S-7. As described above, by equating MS and EMS for each factor, the variances of the identified factors were estimated. The ANOVA table was computed for the all metrics individually. As a result, we found that 21 out of 93 metrics (Table S-8) are dominated by variance due to histologic class differences, i.e., the portion of the variance due to histologic class was the largest among all the associated factors, and either array contribution or residual error introduced the most variation into the other 72 metrics. The largest contribution to variability in data arising from differences in spectral properties of histologic classes indicates that epithelium and stroma differ spectroscopically, independent of all other factors. Thus, the 21 histologic class-dominant metrics are capable of histologic analysis, and for the purpose of histologic discrimination, these metrics could serve as good candidates across wide populations of samples from different clinics. It must be noted that not all metrics that such an analysis provides will be useful for classification. In fact, comparing the 21 histologic class-dominant metrics and the 18 metrics, previously identified as being most useful for a (Bayesian) cell-type classification scheme²⁵, 6 of them were common (Table S-8), which is surprising.

Since this was a de novo exercise compared to the previous study for the sake of being completely independent, we further examined reasons for this surprising result. Among the 15 histologic class-dominant metrics that were not common, there are 10 absorbance ratios, 4 area of a spectral region, and 1 center of gravity of a spectral region metrics. In 10 absorbance ratio metrics, the numerator positions are either the same or close ($<30\text{cm}^{-1}$) to the numerator position(s) of one or more metrics in the 18 metrics for cell-type classification. These are likely associated with the same vibrational modes from similar functional group or chemical bonds; for example, the ratio of the absorbance at 1236cm^{-1} to the absorbance at 1545cm^{-1} and the ratio between absorbance at 1226cm^{-1} and absorbance at 1545cm^{-1} are included in the 18 metrics. Both absorbance at 1226cm^{-1} and 1236cm^{-1} are specific to nucleic acids. Denominator positions are also different for some cases, but these are often the most representative positions, for instance, amide I ($\sim 1650\text{cm}^{-1}$) and II ($\sim 1545\text{cm}^{-1}$). Although the absolute magnitude of the absorbance differs, the relative magnitude among different groups (e.g., histologic classes) tends to be the same, and thus the denominator positions may have little effect on the metric values. The 4 area of a spectral region metrics are not present in the 18 metrics, but the positions showing the highest spectral absorbance in the 4 areas are included in the 18 metrics. Lastly, the peak center of gravity metric, in fact, overlaps with one of the center of gravity of a spectral

region metrics among the 18 metrics. Thus, most of the 21 histologic class-dominant metrics revealed by ANOVA as candidate metrics for histologic discrimination are redundant with the 18 metrics for cell-type classification that were found by a completely independent classification-optimization scheme. It is assumed that a comprehensive variance analysis, of the type reported here, would have included all possible variation that may be encountered. If the addition of potentially useful metrics does not provide an improvement in classification accuracy, there may be an undiscovered pre-analytic variability in the validation datasets or redundancy in the metrics. The two confounding effects may be sorted out using the calibration data. That is, optimizing the classifier, the varying or redundant metrics tend to be avoided. Here, variance components only provide an indication of the likely ability of a single metric, not of the combined effects of multiple metrics or their relative importance. Hence, variance analysis is powerful in that it can identify whether a difference can be made between classes but is limited in identifying the set of metrics that are important for a classifier. Similarly, random contributors to the total variance can be combined to achieve better classification than that obtained using univariate analysis either due to an averaging effect. From this point of view, performing multivariate analysis (e.g., multivariate analysis of variance) will help to have a comprehensive understanding of the combined effects of the metrics and their importance.

Patient contributions have very little effect on the total variation of the data, indicating that inferences made or models built on this data would not be susceptible to the specific patient population. We emphasize, however, that this is not universally true and will change depending on the classification task. It is generally expected that there would be significant biochemical similarity in epithelial and stromal cells in men as these cells perform the same specific functions in all. Although its contribution to the total variance is small, there is significantly larger variance between cores than between patients suggesting that the selection of cores must be made carefully, especially with due attention to the architecture and biology of the tissue. The prostate is organized into zones,³³ which are known to be compositionally and functionally distinct. Yet, no effort was made to control for that variable in these TMAs. Hence, if finer epithelial analysis is required, the example illustrates that paying attention to zonal morphology is likely more important than considering patient variations in constructing TMAs. Since the size of a core is relatively small compared to the entire tissue or organ, it is likely that some of the selected cores are not representative of the zonal region of tissue. This stochastic effect is expected to be smaller as there is no biologic rationale for differences within a zone. We also note that the small number of core samples could affect the variance estimates. Hence, many array designs, including the ones used for training in our original prostate work,²⁵ used up to 8 cores per patient. We note further that interaction effects, by and large, were negligible except the interaction effect between core and histologic class. This is also a likely effect of the aforementioned zonal architecture of the prostate as both stromal and epithelial differences exist between zones. Interestingly, there were 19 metrics which were dominated by variance between arrays (Table S-8). This is likely a contribution of pre-analytical variability in array preparation and is a topic of much discussion in the biomarker community.³⁴ Should this be the dominant mechanism, there are several computational approaches that have been proposed to address array-based differences.³⁵ The discovery of a large number of metrics in which array variance dominates, though, emphasizes that this topic remains one that deserves attention. Further biochemical assessment should also be carried out to demonstrate the effects of sample preparation¹⁷ that are likely responsible for array-to-array differences. Finally, we assessed the statistical significance of histologic class effect by computing *F*-test statistics (Table S-7) and corresponding *p*-values (see Supporting Information for details). The lower *p*-values indicate statistically significant differences between histologic classes, thereby the metrics bearing lower *p*-values may be able to distinguish histologic classes. As shown in Fig. 2A, the larger the portions of explained variance, in general, the lower the observed *p*-

values. However, by definition, it is a relative significance of histologic class effect to the interaction effect immediately below it, not a significance of the effect regarding all the associated factors. Accordingly, it provides limited information and cannot capture the importance of the associated factors.

Variance component analysis reveals subcellular component-specific metrics

To examine the effect of subcellular components in a TMA (iv), pixels identified as epithelial were further divided into two subcellular components: cytoplasm-rich and nucleus-rich pixels. The division of epithelial cells into nuclear and cytoplasmic sub-units is important as tumors often lose the polarity (basal nucleus and cytoplasmic tip) as well as functional (cytoplasmic fraction of the cell) properties that can be observed via changes in morphology.³³ Due to cellular overlap and limited resolution, the two sub-units, in fact, share many of the same chemical compounds.³⁶ The chemical similarity may affect spectral analyses and the variance contributions of the associated factors. The two types of pixels were selected manually upon review of absorbance and corresponding hematoxylin and eosin (H&E)-stained tissue images. 40 cores from 40 patients were chosen, and 100 pixels for each of the two components were extracted to build *within-histologic class model*. As shown in Fig. 3A, subcellular component effect is the dominant source of variation for only 9 metrics (Table S-8), and residual error is the second dominant source for the 9 metrics and the most dominant factor for the rest of the metrics. Although it is the primary source of variation for the 9 metrics, the variance estimate of subcellular component effect does not overwhelm that of other effects. This is contrary to the large differences observed between histologic effect and other effects in *between-histologic class model*. Thus, with the selected cytoplasm-rich and nucleus-rich pixels, we observe only a weak spectroscopic difference.³⁶ This result is consistent with the previous work³⁷ where ~0.72 area under the curve (AUC) for an receiver operating characteristic (ROC) curve was obtained in classifying pixels into cytoplasm-rich and nucleus-rich pixels whereas >0.9 AUC was achieved in histologic segmentation. The lower classification performance or subcellular component contribution to the total variation is due to spectral similarity, errors in gold standard marking of pixels and limitations of IR imaging.³⁷ These cause both biological and technical variations in the data. Biological variations could be reduced by minimizing segmentation errors or obtaining high-definition IR imaging data.³⁸ One limitation of IR spectroscopic imaging is its spatial resolution due to both optical properties^{39,40} and instrumentation constraints.⁴¹ The limitation greatly affects the quality and selectivity of IR imaging;⁴² for example, the pixels between two different components such as epithelium and stroma are often contaminated or contributed by both components, thereby their spectra seem to be the average of the two components,^{9,38} leading to increase in biological variation. High-resolution IR images with reasonable signal-to-noise ratios (SNR) will substantially alleviate the averaging effect as well as biological variation. The finer IR image will, in turn, reduce the segmentation error. We could lessen technical variation by performing repeated measurements on the same TMAs or incorporating numerical methods⁴³ to correct for the data.^{44,45,46} In addition, we computed *F*-test statistic for subcellular component effect. Analogous to the results from *between-histologic class model*, the larger the explained variance due to subcellular component effect, the lower were the observed *p*-values; however, the *p*-values were often small, close to zero, which was not especially informative about the significance of subcellular component effect. Accordingly, *F*-test could not effectively provide discriminative metrics whereas variance components suggest weakly discriminating metrics. We also note that the metrics that capture high variation from histologic class effect in *between-histologic class model*, were not dominated by the subcellular component effect in general (Fig. 3A). This indicates that the chemical properties to distinguish histologic classes differ from the properties to differentiate subcellular components.

Biological variation is the main source of variation in residual error within a core and epithelium

Making no attempt to differentiate between histologic classes in a core, we fitted the TMAs data, used to build *between-histologic class model*, into *no-histologic class model*, and the portions of total variance explained by each factor were computed. As shown in Fig. 2B, residual error is, in general, the dominant source of variation over the 93 metrics. The effects of other factors were relatively small, and either array effect or core effect was mostly the second dominant source of variation. 16 metrics were dominated by array effect, of which 11 metrics were also array effect-dominant metrics in *between-histologic class model*. In comparison with *between-histologic class model*, combining histologic classes, residual error substantially increased in many metrics, especially for the 21 histologic-class dominant metrics. Similarly, constructing the *no-subcellular component model* using the data fitted into *within-histologic class model* and estimating the portions of variance, residual error dominated over the other effects in the entire 93 metrics (Fig. 3B). Compared to the variance components from *within-histologic class model*, we again observed significant increase in residual error for numerous metrics including the 9 subcellular component dominant metrics. Both histologic class and subcellular component factors group pixels similar in chemistry into the same group, as a result, decreasing biological variation. This suggests that biological variation is the main source of variation in residual error within a core and epithelium, especially for those 21 histologic-class dominant metrics_c and 9 subcellular component dominant metrics in the data. However, note that the interpretation of histologic class effect and subcellular component effect should be limited to the population under the experiment since both effects are fixed as described in Supporting Information.

Differences in the effect of the associated factors are observed across TMAs

In order to investigate the differences in variance estimates across TMAs, a *within-array model* is developed. The proportions of variance estimates were, in general, very similar across TMAs, and, comparing to *between-histologic class model*, similar trends were observed for the main effects; 16 out of the 21 histologic-class dominant metrics (*between-histologic class model*) showed high variability due to histologic class effect across all three TMAs; the rest of the metrics were mostly dominated by residual error across TMAs. Examining the 19 array-dominant metrics from *between-histologic class model* (Fig. 4a), we observed the differences in the variance components of not only histologic class effect but also other main and interaction effects across TMAs. In Fig. 4, for the first four metrics, although residual error was the most dominant source of variation, the relative orders of other factors varied greatly across TMAs; the next four metrics showed unusually high variability in Fig. 4b and moderate dominance in Fig. 4d from histologic class effect, but, in Fig. 4c, the effect was not dominant or its contribution is close to residual error; examining the last 11 metrics, the differences in the portions of variance due to both main and interaction effects were also observed. For histologic analysis, these 19 array-dominant metrics may be avoided. The four metrics, in particular, introducing high variation from histologic class effect (Fig. 4b) could be specific to the population represented by the TMA (i), and thus may distract the histologic analysis and its translation into clinical practice. Computing *p*-values of histologic class effect, as observed in *between-histologic class model*, metrics with higher variance components possess lower *p*-values. However, many of the metrics show very small *p*-values. Note that the computation of *F*-test statistic is not identical to *between-histologic class model*. Here, the denominator is the mean square of the interaction effect between histologic class and patient. As described in Supporting Information, the statistical significance is sensitive to the sample size. The degrees of freedom (df) of the denominator in *within-array model* is larger than that of the denominator in *between-histologic class model* whereas df of the numerators are the same. Hence, *p*-values become much smaller although the magnitude of the differences in histologic classes

remains the same or changes only slightly. Differences between TMAs, often from different sources, can be alleviated by standardizing the operating procedures or the management of tissue, for example maintaining biobanks.²⁰ These may help to reduce both technical and biological variations associated with TMA preparation. However, standardization of the process and quality management cannot be achieved without thoroughly evaluating the relevant factors and operations over the course of TMA preparation. In this regard, the variance analysis could also be an excellent tool to assess the procedures and to stabilize the processing and management protocol.

CONCLUSIONS

In this manuscript, ANOVA has been adopted to model IR imaging data from a large population and to identify the main sources of variation. Variation in recorded data arises from every aspect of the sample gathering (different biological characteristics of samples), processing (sample fixation, sectioning, and placement), data acquisition (measurement error) and analysis (baseline correction, normalization, and modeling) steps. The contributions of different factors varied across different spectral metrics, and the main source of variation was not identical, i.e., each of the associated factors affects different spectral metrics differently. Hence, thorough identification of the factors and careful quality control are indispensable to ensure the validity and reliability of tissue classification or analysis using spectroscopic imaging. Several metrics were found to be relevant for histologic classification from an analysis of variance and agreed closely with those reported previously using a pattern recognition approach. Importance of variance between data sets, within a patient population, within a single patient's sample(s) and residual sources were all quantified. Careful understanding of each presents an opportunity to improve the analytic ability of IR spectroscopic imaging, especially for tissue analyses. The approach is not specialized for IR imaging or TMA samples, which were used here, and minor modifications in the ANOVA models can be used to extend the analysis to different modalities or samples. Hence, the framework provided here should prove useful for other tissue types, problems and analytical techniques.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The work reported in this manuscript was supported by the National Institutes of Health via grant R01CA138882.

REFERENCES

1. Colarusso P, Kidder LH, Levin IW, Fraser JC, Arens JF, Lewis EN. *Appl Spectrosc.* 1998; 52(3): 106a–120a.
2. Levin IW, Bhargava R. *Annu Rev Phys Chem.* 2005; 56:429–474. [PubMed: 15796707]
3. Budinova G, Salva J, Volka K. *Appl Spectrosc.* 1997; 51(5):631–635.
4. Fabian H, Naumann D. *Methods.* 2004; 34(1):28–40. [PubMed: 15283913]
5. Petibois C, Deleris G. *Cell Biol Int.* 2005; 29(8):709–716. [PubMed: 15953739]
6. Malins DC, Polissar NL, Nishikida K, Holmes EH, Gardner HS, Gunselman SJ. *Cancer.* 1995; 75(2):503–517. [PubMed: 7812921]
7. Ly E, Piot O, Wolthuis R, Durlach A, Bernard P, Manfait M. *Analyst.* 2008; 133(2):197–205. [PubMed: 18227942]
8. Baker MJ, Gazi E, Brown MD, Shanks JH, Gardner P, Clarke NW. *Br J Cancer.* 2008; 99(11): 1859–1866. [PubMed: 18985044]

9. Bhargava R. *Anal Bioanal Chem.* 2007; 389(4):1155–1169. [PubMed: 17786414]
10. Chalmers, JM. *Mid-Infrared Spectroscopy: Anomalies, Artifacts and Common Errors.* John Wiley & Sons, Ltd; 2006.
11. Pusey E, Luffkin RB, Brown RK, Solomon MA, Stark DD, Tarr RW, Hanafsee WN. *Radiographics.* 1986; 6(5):891–911. [PubMed: 3685515]
12. vandeVen M, Ameloot M, Valeur B, Boens N. *J Fluoresc.* 2005; 15(3):377–413. [PubMed: 15986159]
13. Bowie, BT.; Chase, DB.; Lewis, IR.; Griffiths, PR. *Handbook of Vibrational Spectroscopy.* John Wiley & Sons, Ltd; 2006.
14. Humecki, HJ. *Practical guide to infrared microspectroscopy.* New York: Dekker; 1995. p. 472p x,
15. Williams AC, Barry BW, Edwards HGM, Farwell DW. *Pharmaceut Res.* 1993; 10(11):1642–1647.
16. Puppels GJ, Caspers PJ, Lucassen GW, Wolthuis R, Bruining HA. *Biospectroscopy.* 1998; 4(5):S31–S39. [PubMed: 9787912]
17. Zeng H, Huang ZW, McWilliams A, Lam S, English J, McLean DI, Lui H. *Int J Oncol.* 2003; 23(3):649–655. [PubMed: 12888900]
18. Shim MG, Wilson BC. *Photochemistry and Photobiology.* 1996; 63(5):662–671. [PubMed: 8628758]
19. Griffiths, PR.; Haseth, JAD. *Fourier Transform Infrared Spectrometry.* John Wiley & Sons; 2007.
20. Compton CC, Lim MD, Dickherber A. *Analytical Chemistry.* 2011; 83(1):8–13. [PubMed: 21114268]
21. Kononen J, Bubendorf L, Kallionimeni A, Barlund M, Schraml P, Leighton S, Torhorst J, Mihatsch MJ, Sauter G, Kallionimeni O-P. *Nat Med.* 1998; 4(7):844–847. [PubMed: 9662379]
22. Chung GG, Provost E, Kielhorn EP, Charette LA, Smith BL, Rimm DL. *Clin Cancer Res.* 2001; 7(12):4013–4020. [PubMed: 11751495]
23. Bremnes RM, Veve R, Gabrielson E, Hirsch FR, Baron A, Bemis L, Gemmill RM, Drabkin HA, Franklin WA. *J Clin Oncol.* 2002; 20(10):2417–2428. [PubMed: 12011119]
24. Hendriks Y, Franken P, Dierssen JW, de Leeuw J, Wijnen J, Dreef E, Tops C, Breuning M, Brouckx-Vriendt A, Vasen H, Fodde R, Morreau H. *Am J Pathol.* 2003; 162(2):469–477. [PubMed: 12547705]
25. Fernandez DC, Bhargava R, Hewitt SM, Levin IW. *Nature Biotechnology.* 2005; 23(4):469–474.
26. Mehrens SM, Kale UJ, Qu XG. *J Pharm Sci-U.S.* 2005; 94(6):1354–1367.
27. Corekci B, Malkoc S, Ozturk B, Gunduz B, Toy E. *Am J Orthod Dentofac.* 2011; 139(4):E299–E304.
28. Belbachir K, Noreen R, Gouspillou G, Petibois C. *Anal Bioanal Chem.* 2009; 395(3):829–837. [PubMed: 19685340]
29. McIntosh LM, Summers R, Jackson M, Mantsch HH, Mansfield JR, Howlett M, Crowson AN, Toole JWP. *J Invest Dermatol.* 2001; 116(1):175–181. [PubMed: 11168814]
30. Luthria DL, Mukhopadhyay S, Robbins RJ, Finley JW, Banuelos GS, Harnly JA. *J Agr Food Chem.* 2008; 56(14):5457–5462. [PubMed: 18572954]
31. Kokalj M, Krajsek SS, Bratusa JO, Kreft S. *J Chemometr.* 2010; 24(9–10):611–616.
32. Luthria DL, Mukhopadhyay S, Lin LZ, Harnly JM. *Appl Spectrosc.* 2011; 65(3):250–259. [PubMed: 21352644]
33. Epstein, JI.; Netto, GJ. *Biopsy interpretation of the prostate.* 4th ed.. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins; 2008. p. 358p x,
34. Parker RL, Huntsman DG, Lesack DW, Cupples JB, Grant DR, Akbari M, Gilks CB. *Am J Clin Pathol.* 2002; 117(5):723–728. [PubMed: 12090420]
35. Moreno-Torres JG, Llorà X, Goldberg DE, Bhargava R. *Information Sciences In Press, Corrected Proof.*
36. Lasch P, Pacifico A, Diem M. *Biopolymers.* 2002; 67(4–5):335–338. [PubMed: 12012461]
37. Bhargava R, Fernandez DC, Hewitt SM, Levin IW. *Biochimica Et Biophysica Acta-Biomembranes.* 2006; 1758(7):830–845.

38. Nasse MJ, Walsh MJ, Mattson EC, Reininger R, Kajdacsy-Balla A, Macias V, Bhargava R, Hirschmugl CJ. *Nat Methods*. 2011; 8(5):413–416. [PubMed: 21423192]
39. Sommer AJ, Katon JE. *Appl Spectrosc*. 1991; 45(10):1633–1640.
40. Carr GL. *Rev Sci Instrum*. 2001; 72(3):1613–1619.
41. Reffner JA. *Cell Mol Biol*. 1998; 44(1):1–7. [PubMed: 9551632]
42. Lasch P, Naumann D. *Biochim Biophys Acta*. 2006; 1758(7):814–829. [PubMed: 16875659]
43. Davis BJ, Carney PS, Bhargava R. *Anal Chem*. 2011; 83(2):525–532. [PubMed: 21158469]
44. Koenig JL, Bhargava R, Wang SQ. *Appl Spectrosc*. 1998; 52(3):323–328.
45. Diem M, Romeo M. *Vib Spectrosc*. 2005; 38(1–2):129–132.
46. Gardner P, Lee J, Gazi E, Dwyer J, Brown MD, Clarke NW, Nicholson JM. *Analyst*. 2007; 132(8): 750–755. [PubMed: 17646874]

Figure 1. Schematic of TMAs and potential sources of variation

(Lower row) a number of cores from different patients compose a TMA. (Top, left) A cell type classified image of a tissue sample where colors indicate cell types in prostate tissue. (Top, right) The corresponding IR spectra of cell types are shown. α , β , γ , δ , and ϕ denote array, patient, core, histologic class, and subcellular component effect, respectively. ω and ε refer to measurement error and residual error, respectively.

Figure 2. Portions of explained variance with and without histologic class factor

The portions of total variance explained by the associated factors are estimated for (Top, a) *between-histologic class model* and (b) *no-histologic class model* and plotted over 93 spectral metrics. (Bottom, a) p -values for histologic class effect are shown at the bottom. (a,b) The spectral metrics are ordered by the portion of total variance due to histologic class effect. Interaction effects are not shown for (a).

Figure 3. Portions of explained variance with and without subcellular component factor

The portions of total variance explained by the associated factors are estimated for (Top, a) *within-histologic class model* (b) *no-subcellular model* and plotted over 93 spectral metrics. (Bottom, a) p -values for subcellular component effect are shown at the bottom, and the blue boxes indicate that the corresponding metrics are histologic class-dominant. (a,b) The spectral metrics are ordered by the portion of total variance due to subcellular component effect.

Figure 4. Portions of explained variance for array-dominant metrics across TMs

Portions of variance are shown for (a) *no-histologic class model* and (b,c,d) *between-histologic class model* restricted to each of (i,ii,iii) TMs, respectively. Spectral metrics are ordered by the portion of total variance due to (a) array effect.

Table 1

Summary of ANOVA models in the manuscript. Here, y represents IR absorption of a pixel, and μ is the overall mean. α , β , γ , δ , and ϕ denote array, patient, core, histologic class, and subcellular component effect, respectively. $\alpha\delta$, $\beta\delta$, $\gamma\delta$, and $\gamma\phi$ are interaction effects. σ_α^2 , σ_β^2 , σ_γ^2 , σ_δ^2 , and σ_ϕ^2 indicate variance components assigned to array, patient, core, histologic class, and subcellular component effects respectively, and σ_ω^2 and σ_ε^2 are components of measurement error and residual error, respectively. Details of the symbols are presented in Table S-1 and models are explained in detail in Supporting Information.

<i>Between-histologic class</i>	Objective	Examines the effect of histologic class across TMAs. Helps determine whether a given metric can differentiate between histologic classes, overcoming all other sources of variance.
	Model	$y_{ijklw} = \mu + \alpha_i + \beta_{j(i)} + \gamma_{k(j(i))} + \delta_l + \alpha\delta_{il} + \beta\delta_{j(i)l} + \gamma\delta_{k(j(i))l} + \omega_{ijklw} + \varepsilon_{ijklw}$
	Variance	$\sigma_{total}^2 = \sigma_\alpha^2 + \sigma_{\beta(\alpha)}^2 + \sigma_{\gamma(\beta(\alpha))}^2 + \sigma_\delta^2 + \sigma_{\alpha\delta}^2 + \sigma_{\beta(\alpha)\delta}^2 + \sigma_{\gamma(\beta(\alpha))\delta}^2 + \sigma_\omega^2 + \sigma_\varepsilon^2$
<i>Within-array</i>	Objective	Examines the effect of histologic class within a TMA. Determines the ability of a metric to distinguish histologic classes within a data set, e.g. in calibration data.
	Model	$y_{jklw} = \mu + \beta_j + \gamma_{k(j)} + \delta_l + \beta\delta_{jl} + \gamma\delta_{k(j)l} + \omega_{jklw} + \varepsilon_{jklw}$
	Variance	$\sigma_{total}^2 = \sigma_\beta^2 + \sigma_{\gamma(\beta)}^2 + \sigma_\delta^2 + \sigma_{\beta\delta}^2 + \sigma_{\gamma(\beta)\delta}^2 + \sigma_\omega^2 + \sigma_\varepsilon^2$
<i>No-histologic class</i>	Objective	Examines residual error across TMAs. Helps determine the change in residual error in the absence of histologic class effect across TMAs in comparison with <i>between-histologic class</i> model.
	Model	$y_{ijkw} = \mu + \alpha_i + \beta_{j(i)} + \gamma_{k(j(i))} + \omega_{ijkw} + \varepsilon_{ijkw}$
	Variance	$\sigma_{total}^2 = \sigma_\alpha^2 + \sigma_{\beta(\alpha)}^2 + \sigma_{\gamma(\beta(\alpha))}^2 + \sigma_\omega^2 + \sigma_\varepsilon^2$
<i>Within-histologic class</i>	Objective	Examines the effect of subcellular component in a histologic class. Determines the discriminative ability of a metric to identify subcellular components in a histologic class.
	Model	$y_{kmw} = \mu + \gamma_k + \phi_m + \gamma\phi_{km} + \omega_{kmw} + \varepsilon_{kmw}$
	Variance	$\sigma_{total}^2 = \sigma_\gamma^2 + \sigma_\phi^2 + \sigma_{\gamma\phi}^2 + \sigma_\omega^2 + \sigma_\varepsilon^2$
<i>No-subcellular component</i>	Objective	Examines residual error in a core. Helps determine the change in residual error in the absence of subcellular component effect within a histologic class in a core, comparing with <i>within-histologic class</i> model.
	Model	$y_{kw} = \mu + \gamma_k + \omega_{kw} + \varepsilon_{kw}$
	Variance	$\sigma_{total}^2 = \sigma_\gamma^2 + \sigma_\omega^2 + \sigma_\varepsilon^2$