Tina Memo No. 2013-001
Internal

Colorectal Liver Metastases Show Similar Microvascular Characteristics for Individual Patients.

N.A.Thacker, L. Horsley and A.Jackson.

Last updated
4 / 9 / 2013

Imaging Science and Biomedical Engineering Division,
Medical School, University of Manchester,
Stopford Building, Oxford Road,
Manchester, M13 9PT.

Colorectal Liver Metastases Show Similar Microvascular Characteristics for Individual Patient.

L. Horseley, N.A.Thacker, J. O'Connor and A.Jackson 4/9/2013

Abstract

The purpose of this study is to investigate the measurement of biological variation for colorectal liver metastases using DCE MRI. Such techniques can be used to obtain several parameters relating to microvascular characteristics of tissue. Here we demonstrate the ability to construct statistical similarity functions for these parameters. Key to this process is the construction of an empirical measurement model and subsequent re-mapping of the original measured data to Gaussian random variables, suitable for use in a chi-squared similarity function. We use this approach to evaluate within and between subject variation in metastases. The evidence provided suggests it may be possible to perform tumour discrimination on the basis of these variables.

Introduction

More than a hundred years ago, Paget published his findings that breast cancer had a propensity to spread to certain metastatic sites (ref), leading to the seed and soil theory of cancer metastases. However, a theory to explain the development of cancer cells was not proposed until 1976, when Nowell published his theory on clonal expansion of cancer cells [1], proposing that the majority of cancers originate from a single cell. Although both theories remain widely accepted, heterogeneity within tumours (intra-tumoral heterogeneity), has subsequently been well documented (ref). Here, non-genetic factors such as tumour plasticity according to the tumour environment are clearly important. In contrast, the relationship between primary tumours and metastases or different metastases (inter-tumoral heterogeneity), has been less widely studied. Recently, genetic inter- tumour heterogeneity has been demonstrated in 4 cases of renal cancer, where genetic sequencing and chromosomal analysis were carried out on both primary tumour and metastases [2]. This study demonstrated a branching theory of tumour evolution and concluded that analyses from a single tumour biopsy may underestimate the extent of mutations even within a single tumour. This is of particular relevance in the era of targeted therapy as the current convention utilises tissue from the primary tumour to guide the management of metastatic disease and the implications of cancer heterogeneity for personalised medicine and biological therapies have subsequently been reviewed (Fisher et al, 2013).

Angiogenesis refers to the formation of new capillaries from pre-existing vessels. Although this process is limited in healthy adults to only a few specific physiological conditions, it occurs pathologically in tumours and is recognised as one of the hallmarks of cancer [3] Dynamic Contrast Enhanced Magnetic Resonance Imaging (DCE-MRI) is a non-invasive technique which allows the quantitative assessment of microvascular properties within a given tissue. Tumours blood vessels are highly disorganised, resulting in heterogeneity of blood flow across tumours. In addition to this functional abnormality, vessels are also highly abnormal in there structure. These abnormalities include: defective or discontinuous basement membranes as well as abnormal endothelial cell growth and results in vessels which are leaky to a variable extent. This property is exploited in tracer kinetic modelling of contrast agent, which allows calculation of a number of parameters which describe the structure and function of the tumour micro-vasculature. In this way, DCE MRI has been widely used to characterise angiogenesis in tumours.

The liver is a common site for metastases in colorectal cancer and patients frequently have multiple lesions. Here we investigate the microvascular characteristics of liver metastases using DCE MRI and hypothesise that the microvascular characteristics of multiple colorectal liver metastases within an individual patient can be significantly different from a general population of these tumours.

Recent studies in breast cancer have suggested a correlation between tumour histological subtype and DCE-MRI characteristics [6, 7]. Unfortunately, the measured parameters obtained using DCE MRI are difficult to effectively summarise, due to the heterogeneity of tissue and variable measurement precision. As a consequence summary variables obtained from distributions of individual parameters are only very weakly correlated with biological changes. This observation is particularly relevant to liver metastases, where biological variation between different tumours is expected to be relatively small. In this work we introduce a multi-dimensional approach based upon the construction of a chi-squared statistic which supports the fusion of information from a set of such weak variables. We demonstrate an ability to differentiate between the biological differences observed for intra and inter-subject tumour tissue. We speculate that such methods may be able to provide a non invasive signature for an individual cancer.

Materials and Methods

Patient selection

The patients included in this study were undergoing imaging with Dynamic Contrast Enhanced Magnetic Resonance Imaging (DCE-MRI) as a part of a clinical trial running at our institution and had given written informed consent to participate in the study. The study had ethical approval and was carried out in accordance with standards of GCP. Patients were eligible providing they were over eighteen year of age with biopsy confirmed metastatic colorectal cancer, without previous therapy for metastatic disease and disease measuring >3cm. All patients had undergone 2 baseline DCE-MRI scans, median 4 days (range 2-7 days) prior to treatment. It is this data which has been used in the current work to investigate biological variation within tumour tissues. To help control for tumour microenvironment, only patients with liver metastases were included in our analyses. This resulted in a sample of 29 subjects with numbers of metastases varying between 2 and 6, giving 102 tumours in total.

MRI data acquisition and analysis

Data were acquired on a 1.5T Philips Intera system. The baseline T1 measurement consisted of 3 axial spoiled Fast Field Echo (gradient echo) volumes with flip angles 2, 10, 20 degrees, respectively and 4 signal averages. The dynamic series was acquired using the scanner whole body coil (Q body coil) for transmission and reception. The dynamic series consisted of 75 consecutively-acquired axial volumes with a flip angle of 20 degrees, 1 signal average, and a temporal resolution of 4.97 s. All studies maintained the same number of slices (25), field of view (375 mm 375 mm), matrix size (128 128), TR (4.0 ms), and TE (0.82 ms) for the baseline T1 measurement images and the dynamic series itself. Slice thickness was 4 mm for small target lesions or 8 mm for larger lesions, giving superior-inferior coverage of 100 mm or 200 mm, respectively. Gadoterate Meglumine (Dotarem) was injected intravenously (IV) by power injector at the time of the sixth dynamic acquisition at 0.2ml/kg, followed by a 20ml saline flush at a set rate of 3ml/sec. This was followed by acquisition of a post contrast T1-weighted image.

VOIs were delineated by an experienced radiographer on co-registered high resolution T1- and T2-weighted images. Whole TV was measured for each lesion. An arterial input function was measured where possible; in circumstances where this was not appropriate, a population derived input function was used [4]. Analysis was performed using in-house software (Manchester Dynamic Modelling) and the extended Tofts and Kermode pharmacokinetic model[5] was used to calculate the fractional volume of the extravascular extracellular space (V_e). The model free measurement, initial area under the gadolinium contrast curve at 60s (IAUC60) was calculated and voxels from tumour VOIs were included in the analysis if they demonstrated uptake of contrast, this was defined as an initial IAUC in the first 60 seconds (IAUC60)>0 mmol/s.

Statistical analysis

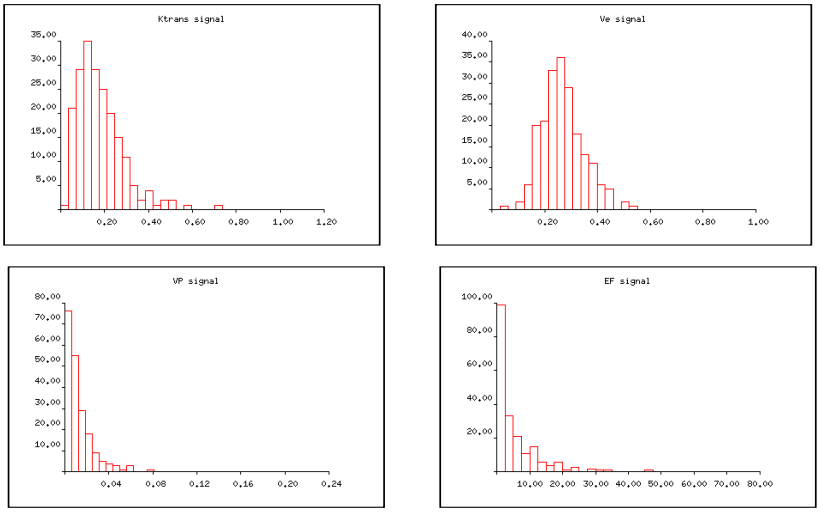

Figure 1: *The distribution of the median DCE MRI variables* K_{trans}, V_e, V_p *and* $100(1 - EF)$.

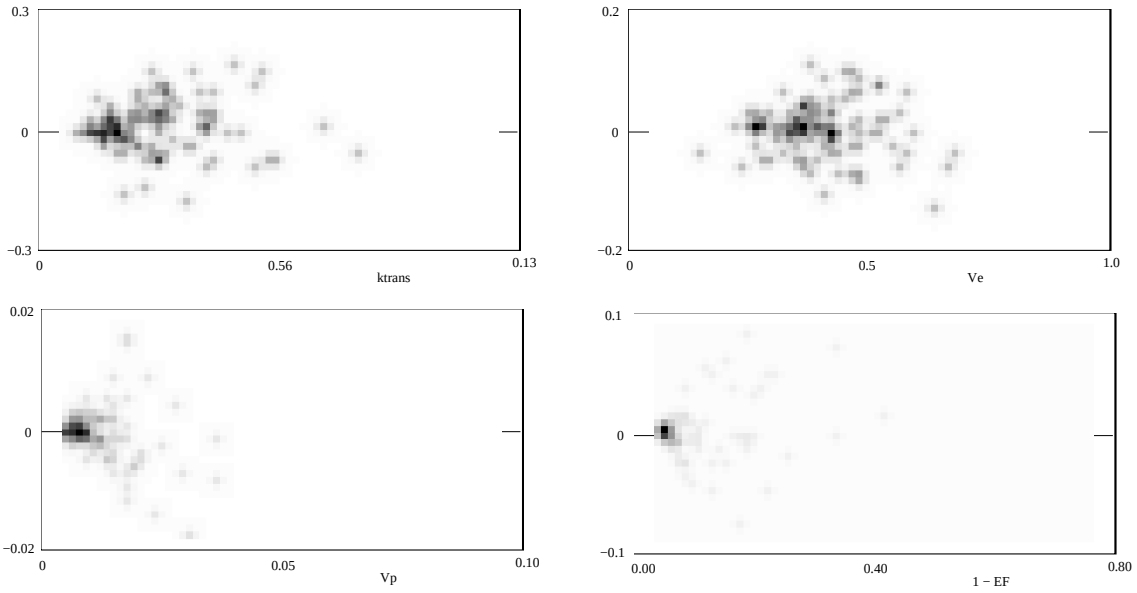

Figure 2: *Bland-Altman plots, showing reproducibility for median values of k_{trans}, V_p, V_e and EF (top left to bottom right). A clear parameter dependant accuracy is seen for all variables.*

Median values of the measured parameters K_{trans} V_p and V_e were extracted from distributions obtained from the 73 selected tumour regions. The enhancing fraction (EF) was also measured. Distributions of these values are shown in Figure 1. The repeat data was used to construct Bland Altman plots [8] (Figure 2), illustrating a parameter dependent measurement repeatability. As these repeat measurements were obtained from different scan sessions we can assume these estimates include all important aspects of the variation intrinsic to the process of measurement and consequently the accuracy with which we can quantify biological change. Unfortunately the data shown are unsuitable for the direct construction of a meaningful multi-dimensional statistical difference (such as summed squared difference) due to the high variation (inhomogeneity) in measurement accuracy. In some cases order of magnitude changes are visible in the standard deviation of the repeatability data across the dynamic range of the measured parameters.

For optimal statistical efficiency, the distance between two points in a measurement space needs to take into account the varying accuracy at locations along the shortest statistical path (Appendix A, see also for example [9]). This shortest path is not in general a straight line (Appendix C). This problem can be eliminated if the data can be monotonically mapped to a variable with homogenous Gaussian measurement (repeatability) errors. Under these circumstances the shortest statistical distance is a straight line and statistical differences are computable in multi-dimensional data as weighted Euclidean distances, i.e. a chi-square.

Using the Bland-Altman plots, approximate empirical models of measurement accuracy were constructed for the first three variables of the form $x \pm kx^{n/2}$ (for integer n). The EF value was expected to be distributed as a binomial random variable, but was found to be better modelled as a proportional error on $1 - EF$. The resulting empirical error models for these summary parameters are given in table 1. Each such error model $x \pm f(x)$ was used to derive a corresponding non-linear mapping function $y(x)$

$$y(x) \propto \int \frac{1}{f(x)}\, dx$$

The theory of error propagation (a first order approximation, see Appendix B) can be used to show that $y(x)$ will have approximate homogenous measurement error. For the mapping functions listed in Table 1, the corresponding Bland-Altman plots for these new variables are shown in Figure 3. The degree of uniformity of reproducibility in these plots (now within a factor of two) demonstrates the utility of the chosen mapping functions and was deemed sufficient for our purposes.

A chi-squared variable for 4 degrees of freedom for the measured difference between tumours i and j was then defined as the sum of the squares of the difference between changes in each derived DCE MRI variable i divided by its reproducibility variance.

$$\chi^2_{jk} = \sum_i^4 (y_{ij} - y_{ik})^2 / \sigma^2_{y_i}$$

The quantity $D_{jk} = \sqrt{\chi^2_{jk}/4}$ is expected to behave as a linear Euclidean distance of measurable statistical difference. A mean value of 1 will be generated for data j and k which differs only due to the presence of the modelled level of measurement error (σ_y).

parameter x	error model	transformation y	standard deviation σ_y
k_{trans}	$x \pm k\ x$	$y = -log(x)$	0.29
V_e	$x \pm k\ x$	$y = -log(x)$	0.18
V_p	$x \pm k\ x$	$y = -log(x)$	0.82
EF	$1-x \pm k\ (1-x)$	$y = log(1-x)$	0.74

Table 1: *Approximate error models determined from reproducibility data, and the non-linear functions required to generate approximate Gaussian random variables with the specified deviations.*

4 of the 102 repeat baseline data sets provided were found to have a chi-square of 16 for 4.0 degrees of freedom, indicating a problem with obtaining an equivalent repeat measurement. As we could not determine which of these was in error both of the baseline measurements were excluded from further study. The distribution of D for the reproducibility data was then found to have a mean of 0.91, which is very close to the theoretical value of 1.0 for independent y_i. The slight underestimate is thought to be due to quantization of EF values which leads to a zero difference for 100% values.

The statistical distance D was then constructed for differences between tumours of each subject, and also for differences between tumours from different subjects (Figure 5) giving mean values of 1.365 ± 0.036 (stat) ± 0.05 (sys) and 2.044 ± 0.021 (stat) ± 0.07 (sys). Differences were always taken to the measurement from the alternative baseline study so that the estimated reproducibility was correctly incorporated. Statistical errors for these quantities were estimated using boostrap resampling with replacement. Systematic errors were computed based upon accuracy of the reproducibility variances. Both of these values are statistically significantly different to the null hypothesis,(that the data can be accounted for by measurement error). They are also significantly different to each other on the basis of a z-score with a probability of $1 \times 10^{1-8}$.

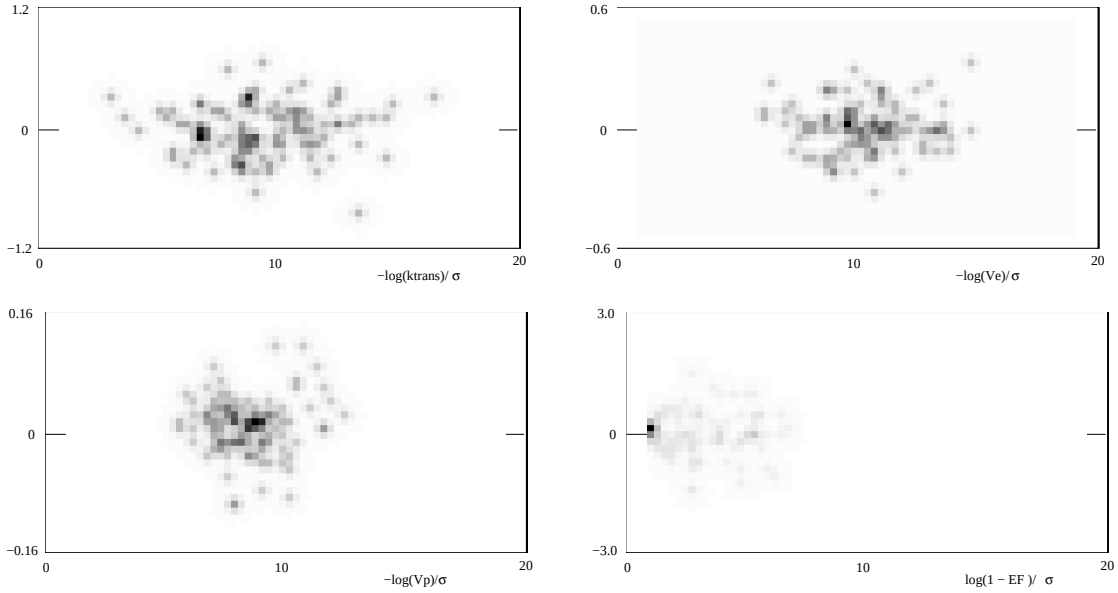

Figure 3: *Bland-Altman plots, showing reproducibility for transformed variables derived from k_{trans}, V_p, V_e and EF (top left to bottom right), scaled on the x axis to units of measured reproducibility σ. For a successful transformation the residual distributions (distribution of scatter above and below zero) should be independent of the variable. The high density value in the EF plot is due to the quantisation of this variable at 100% which causes identical values which cannot be separated by a transformation.*

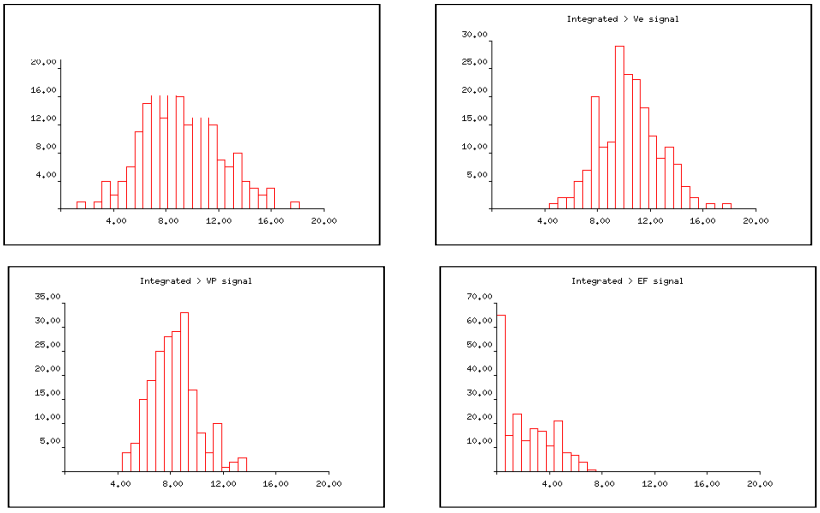

Figure 4: *Transformed variables scaled to reproducibility. Unlike the original parameter distributions (Figure 1), the spread of each variable (e.g. variance) is a measure of the associated information content. Gaussian random variables containing no signal (only noise) are expected to have a Gaussian distribution with unit variance, see Table 2.*

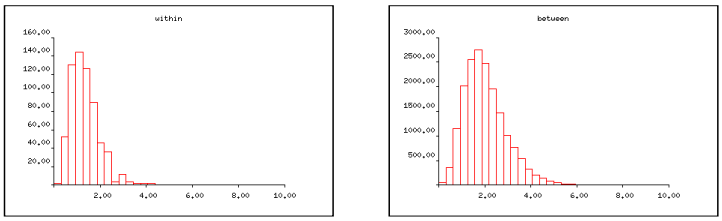

Figure 5: *The distributions of statistical distances D, for within subjects (left) and between subject (right) tumours. Computed using a chi squared statistic based on the transformed and scaled DCE MRI summary variables of median K_{trans}, V_e, V_p and EF (as described in the main text).*

Discussion and Conclusions

Our analysis of the reproducibility of DCE MRI parameters, K_{trans}, V_p, V_e and EF, has generated a set of approximate models which allow us to quantify the accuracy of these measurements. We have applied non-linear transformations in order to map these variables into Gaussian random variables. The distribution of transformed measurements from all 29 subjects were found to have standard deviations between 2.0 and 3.2 (Figure 2 and Table 2). As the expected value for pure noise is 1.0, this indicates (as expected) that the signal pertaining to biological variation for each individual measurement is quite weak, and insufficient to allow good biological separation of this tumour data. However, we have shown how the fusion of these measurements into a single chi-squared statistic allows us to differentiate between intra and inter subject tumour tissue.

parameter	transformation	signal S.D.
k_{trans}	$y = -log(x)$	2.88
V_e	$y = -log(x)$	2.25
V_p	$y = -log(x)$	1.75
EF	$y = log(1-x)$	1.99

Table 2: *Signal content of DCE MRI variables measured as the S.D. on the corresponding transformed Gaussian random variable (Figure 4). All variables provide only a weak quantification of tissue biology, the K_{trans} variable is the best, with an inherent signal which is over 3 times larger than the effects of measurement error (reproducibility). The V_p parameter is the weakest.*

The error models used are only effective approximations. Though the use of percentage errors (i.e. $x \pm kx$) are common for such parameters in the literature, the true source of repeatability errors associated with median values has many contributions which complicate any attempt to understand them. These contributions include acquisition standardisation, fitting error and sampling statistics for the biological variation under study. However, the assumption that there should be a monotonic dependency of the measurement error on the parameter is reasonable provided we are fitting similar data sets to similar values of parameters and for similar sized VOI's. Under these circumstances the distributions of measured values from which the medians are computed will vary only slowly as a function of the median value, and so too therefore will the accuracy with which these medians can be measured. Under such circumstances, the functional dependency of these accuracies could take a variety of forms, but we can expect them to be smooth, continuous, gradual, monotonic and non-linear. Consequently, the selection of a restricted set of simple power functions as error models is justified, and adequate for the amount of data available for the present study[1].

As analyses are aften conducted without re-mapping variables to achieve more Gausssian behaviour, we repeated our main analysis without the variable transformation stage. In this case the number of outliers identified in the repeat data increased from 4 to 12 (the main effect of the mappings are to reduce instabilities). The within and between distances were then computed as 1.273 ± 0.035 (stat) $\pm$ 0.05 (sys) and 1.995 ± 0.035 (stat) $\pm$ 0.07 (sys). The main difference to the original results is in the statistical error on the between group distance, which is 67% less accurate. As these errors are proportional to the intrinsic measurement errors we can say that this difference is equivalent to having improved the measurement accuracy of the original data by the same factor. Thus the transformed analysis is statistically more efficient. Although this improvement may seem unnecessary for a sample of 102 tumours, it would be valuable when trying to assess differences between individual samples.

The behaviour of K_{trans}, V_p and V_e seems to follow the expected behaviour of direct proportionality. EF is at least the easiest of these parameters to understand, as it is by definition a binomial sampling process for consistently sized VOI's if sampling is the dominant source of error. The additional error seen for low EF values therefore suggests additional biological variation. We would need to revisit these assumptions if the selected parameter mappings failed to generate approximate Gaussian distributions, or resulted in the identification of a large number of repeatability outliers.

The suitability of any non-linear mapping to achieve our aim of generating a Gaussian random variable may seem hit-and-miss. The advantage we have is that if we succeed in obtaining homogenous measurement error following the mapping then this is all we need to ensure the validity of subsequent statistical analysis. This is true even if we do not fully understand the origins of measurement variation. Better control over this process may ultimately require a more sophisticated approach to sampling and measurement. Ideally we would define parameters with which to characterise DCE MRI time course data which had these properties by construction rather than having to rely on empirical observation.

Obtaining accurate data for multiple variables for technically complex DCE MRI variables is always difficult. Despite a range of quality assurance measures, put in place as a standard procedure for the trial, 4 outliers were identified in the repeated baseline data. Finding 4 outliers in a repeat sample of 102 may be viewed as a large failure rate but it corresponds to only 0.5% of the 8 individual measurements per VOI and 2% of each group of DCE MRI measurements. Though checks were made, these data could not be attributed to errors of processing, acquisition or transcription. It was assumed that they were outliers caused by problems associated with either uncontrolled biological change or identification of equivalent tumour regions. The consequent number of un-reproducible measurements is an expected difficulty of such multi-dimensional analysis.

The mean chi-squared for reproducibility data is consistent with an assumption of independence between the DCE summary variables, so that only four scaling values (σ_y) are needed to construct a similarity measure. If this were not the case 10 separate values would be needed to characterise the measurement covariance for the construction of a Mahalanobis distance. A sample of twice the current size would be needed in order to attempt this to a useful precision.

In conclusion, biological variation is measurable within the metastases from an individual subject on the basis of our scaled parameters, but this heterogeneity is less (by a factor of 1.8) than the full variation seen across all tumours. This is evidence that biological variation is being measured and it may therefore be possible to categorise tumours on the basis of these variables and methods. This remains a topic for future work.

[1]The performance of better approximating models will not be statistically significant given the power of the data sample. Further, slight variation of the models around the assumed error dependences was found to make no difference to the main conclusions of the present study.

Appendix A: Heteroscedasticity and Measured Vector Difference

Those with a physics or mathematical background may find the motivation for homoscedastic transforms quite trivial. However, those with a more empirical approach to mathematics may find the idea perplexing. Therefore, this appendix has been written for the purposes of explaining the concept of statistical efficiency using a common engineering principle; the signal to noise ratio.

If we have two vector measurements ($\mathbf{x_1}$ and $\mathbf{x_2}$) of some arbitrary dimensionality, and wish to quantify changes in measured values ($x_i = x_{1i} - x_{2i}$), we might decide to do this by constructing a weighted Euclidean distance

$$D = \sqrt{\sum_i x_i^2 k_i} \qquad (A.1)$$

We seek a motivation for using this particular form and a choice of k_i which optimises our ability to detect measurement changes. Note that this is a different (but analogous) problem to that of building a vector classifier[2].

Assuming for now that the variations on x_i due to noise along the path between $\mathbf{x_1}$ and $\mathbf{x_2}$ are independent, Gaussian and constant (σ_i), we can apply error propagation to this quantity in order to determine its noise characteristics. As the derivative with respect to any measured difference is

$$\frac{\partial D}{\partial x_i} = \frac{x_i k_i}{D}$$

then the variance on D is

$$\sigma_D^2 = \sum_i (\frac{\partial D}{\partial x_i})^2 \sigma_i^2 = \frac{\sum_i x_i^2 k_i^2 \sigma_i^2}{D^2} \qquad (A.2)$$

As our weighted distance vector has uniform errors along its entire length, the ability to distinguish differences between measurements is optimal when D has the largest signal to noise ratio, or equivalently the largest z-score. This can be written as

$$S = \frac{D^2}{\sigma_D^2} = \frac{(\sum_i x_i^2 k_i)^2}{\sum_i x_i^2 k_i^2 \sigma_i^2} \qquad (A.3)$$

We can therefore determine the best weight factors k_j for detecting changes between x_1 and x_2 by maximising this expression in the usual way

$$\frac{\partial S}{\partial k_j} = \frac{2x_j^2 \sum_i x_i^2 k_i}{\sum_i x_i^2 k_i^2 \sigma_i^2} - \frac{2(\sum_i x_i^2 k_i)^2 x_j^2 k_j \sigma_j^2}{(\sum_i x_i^2 k_i^2 \sigma_i^2)^2}$$

so that the signal to noise ratio has an optima when

$$\frac{2x_j^2(\sum_i x_i^2 k_i)(\sum_i x_i^2 k_i^2 \sigma_i^2) - 2(\sum_i x_i^2 k_i)^2 x_j^2 k_j \sigma_j^2}{(\sum_i x_i^2 k_i^2 \sigma_i^2)^2} = 0 \qquad (A.4)$$

Eliminating the common multiplicative terms we can write this as the constraint

$$\sum_i (k_i \sigma_i^2) x_i^2 k_i = (k_j \sigma_j^2) \sum_i x_i^2 k_i \qquad (A.5)$$

i.e. in order to construct a weighted similarity metric (A.1) with fixed k_i so that the information in D is maximised, (A.5) must be satisfied for all x. By inspection of (A.5) we can see that (A.4) is true when

$$(k_i \sigma_i^2) = (k_j \sigma_j^2) \quad \rightarrow \quad k_i \propto \frac{1}{\sigma_i^2} \qquad (A.6)$$

At fist sight we have simply shown that a conventional Chi-squared test is appropriate, but we can also see that if we extend our vector $\mathbf{x}$ into a region where the measurement variances change, then A.6 will no longer hold. **For best statistical efficiency, the use of fixed weights k_i in (A.1) implies that x must be a homoscedastic space.** The next appendix will show how this can be achieved for some problems, the final appendix relates this specific result to more general concepts of statistical distance and probability. Clearly this result has implications for areas such as statistical data analysis or pattern recognition. The reader may wish to consider for example the construction of a nearest neighbour classifier, or the standard Chi-square as applied to histogram similarity.

[2]Where the mathematical solution is dominated by the distribution of signal within classes, rather than measurement uncertainty.

Appendix B: Transformations to Homoscedastic Space

Repeated measurements of data (x_1 and x_2) are related to the underlying measurement error ($var(x)$) by

$$var(x) \;=\; <(x_1 - x_2)^2> /2$$

Plots of $x_1 - x_2$ against $(x_1 + x_2)/2$ (a.k.a Bland-Altman plots) allow us to infer the functional dependency of the measurement $f(x)$. For an error model of measurement for a random variable x of the form

$$x \;=\; x' \pm f(x)$$

i.e. $var(x) \;=\; <(x - x')^2> \;=\; f(x)^2$, we can estimate errors for a monotonic continuous differentiable function $y \;=\; g(x)$ using error propagation, as follows

$$var(y) \;=\; \left[\frac{\partial g(x)}{\partial x}\right]^2 f(x)^2$$

Therefore for the specific choice $g(x) \;\propto\; \int \frac{1}{f(x)}\; dx$

$$var(y) \;\propto\; \left[\frac{\partial(\int \frac{1}{f(x)}\; dx)}{\partial x}\right]^2 f(x)^2 \;=\; \left[\frac{1}{f(x)}\right]^2 f(x)^2 \;=\; 1$$

i.e. $var(y) \;=\; <(y - y')^2>$ is constant ($y' \pm \; const$), and now independent of x and this behaviour should be observable in the corresponding Bland-Altman plot.

Median values are a good choice for input variables x to this approach, as the median is preserved under non-linear transformation ($g(x_{med}) \;=\; y_{med}$), while other measures (for example the mean) are not ($g(x_{mean}) \neq y_{mean}$). Therefore the methodology is independent of the initial choice of variable. They also provide a degree of robustness for distributions contaminated with outliers, such as fitting errors from DCE MRI analyses.

Appendix C: General Notes on Metric Spaces

This appendix is intended for those who are not already familiar with the concept of metric spaces and is written in the form of a general explanation for the approach taken in this document.

Given two vector measurements, $\mathbf{x_1}$ and, $\mathbf{x_2}$, how should we define their degree of similarity? This is a simple question which forms the basis of many statistical procedures, including Pattern Recognition systems. Simply taking a squared difference is the common choice, but this is not invariant under rescaling of the parameters. As such scaling is left to the choice of the experimenter this is unscientific. We could try using a weighted squared difference, and it is common to base this on the variance of each variable around the mean value. This is better, and also a common step in many Pattern Recognition papers, but is not statistically valid, for example it is not invariant under non-linear re-definition of the parameter.

At first sight this might not appear an important issue, as we tend to have preconceptions of which variables are important to us (mass, length), but this is only an illusion. If I am interested in the sizes of spheres and choose to measure the volume but someone else measures the radius, our measurements will behave differently in the subsequent analysis, due to the non-linear relationship between these variables, $v \;=\; \frac{4}{3}\pi r^3$. These issues of non-unique parameter definition are always present, only naivety leads us to think otherwise. We define processes such as free diffusion and might characterise them using parameters corresponding to a distance in or models, but we could have just as easily modelled using squared distance or distance cubed. Indeed there are an infinite number of equivalent approaches. Our use of one preferred approach is generally due to nothing more than convenience or convention.

The solution is to consider our measurement process as the definition of a metric space, such that at each point in the space there is a definable cost of moving a small distance in any direction which is a fixed property of that location[3]. If we understand the measurement accuracy of these techniques we can define similarity by taking the minimum statistical cost path as an integral of the covariance ($C_{\mathbf{x}}$) weighted Mahalanobis distance. With some

[3]The assumption made in Pattern Recognition is generally that measurement errors are negligible, but the magnitude of error is a completely different issue to the relative scalings and variations as we move across pattern space. In order to ignore measurement errors entirely we must be able to assume that they are both small **and** homogenous. The consequences of this assumption vary for different algorithms, but this is never tested, making the successful use of these methods somewhat a black art.

care, we can ensure with this approach that the estimate of statistical difference will always be the same, regardless of how we define our initial variables. Consequently, if we can find a space of measurement $\mathbf{y} = g(\mathbf{x})$ for which the associated measurement covariance is everywhere isotropic (errors are uniform, approximately Gaussian and isotropic, $C_{\mathbf{y}} \propto I$), the shortest cost path is by definition a straight line between $\mathbf{y_1}$ and $\mathbf{y_2}$, consistent with the definition of a Chi-squared variable. Further, the minimum cost path in the original variables can be found by transferring the line in $\mathbf{y}$ back into the original space $\mathbf{x}$. This is a useful shortcut, as in the original space $\mathbf{x}$ we would have had to construct every possible route and evaluate its statistical distance in order to find the shortest.

All credible methods for scientific analysis must ultimately be rooted in probability theory, and statistical metric spaces are no exception. By construction, the length of the shortest cost path in $\mathbf{y}$ can also be said to be the logarithm of the probability that a measurement located at $\mathbf{x_1}$ could be perturbed, due to the error on measurement, to a value at $\mathbf{x_2}$. Thus this definition of difference is specifically linked to the concept of measurement uncertainty. Hence the use of repeatability experiments to determine the required functional mappings. What we should now appreciate is that given the choice of "heteroscedastic $\mathbf{x}$" or "homoscedastic $\mathbf{y}$" the latter is fundamentally more useful to us as a way of summarising our measurements.

Returning to the comment above, just because we are familiar with the use of concepts such as permeability does not mean that K_{trans} is the best parameter to summarise an observation. The best parameters are those which can more readily support meaningful statistical tests. Correlations and other factors may make our initial biological parameters a very poor choice, particularly when seeking to combine them. Taking appropriate account of available evidence is expected to be the basis for the most powerful statistical tests. We therefore expect that statistical tests will generally be more informative if data are weighted via an appropriate scaled non-linear transformation. Making the effort to do this will have the largest benefit when the evidence is weak, such as the measures considered in this document.

Of course we can also conceive of data sets for which no metric space can be constructed, for example data from a completely unknown or unrepeatable measurement process. However, we should not use this as an argument against use of metric spaces, as the arguments for scientific credibility are fundamental, rather we should consider obtaining the data differently so that a credible analysis is possible.

References

[1] Nowell, P.C., The clonal evolution of tumour cell populations. Science, 1976. 194(4260): p. 23-28.

[2] Gerlinger, M., et al., Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med. 366(10): p. 883-892.

[3] Hanahan, D. and R.A. Weinberg, The hallmarks of cancer. Cell, 2000. 100(1): p. 57-70.

[4] Parker, G.J., et al., Experimentally-derived functional form for a population-averaged high-temporal-resolution arterial input function for dynamic contrast-enhanced MRI. Magn Reson Med, 2006. 56(5): p. 993-1000.

[5] Tofts, P.S., Modelling tracer kinetics in dynamic Gd-DTPA MR imaging. J Magn Reson Imaging, 1997. 7(1): p. 91-101.

[6] Martincich, L., et al., Correlations between diffusion-weighted imaging and breast cancer biomarkers. European Radiology. 22(7): p. 1519-1528.

[7] Youk, J., et al., Triple-negative invasive breast cancer on dynamic contrast-enhanced and diffusion-weighted MR imaging: comparison with other breast cancer subtypes. European Radiology. 22(8): p. 1724-1734.

[8] J M Bland and D G Altman, Statistical method for assessing agreement between two methods of clinical measurement., Lancet, 327, 307-310, 1986.

[9] N.A.Thacker, F.Ahearne and P.I.Rockett, 'The Bhattacharyya Metric as an Absolute Similarity Measure for Frequency Coded Data.' Kybernetika, 34, 4, 363-368, 1997.