Article

Decision Theory versus Conventional Statistics for Personalized Therapy of Breast Cancer

Michael Kenn [1], Rudolf Karch [1], Dan Cacsire Castillo-Tong [2], Christian F. Singer [2], Heinz Koelbl [3] and Wolfgang Schreiner [1,*]

[1] Section of Biosimulation and Bioinformatics, Center for Medical Statistics, Informatics and Intelligent Systems (CeMSIIS), Medical University of Vienna, Spitalgasse 23, 1090 Vienna, Austria; michael.kenn@meduniwien.ac.at (M.K.); rudolf.karch@meduniwien.ac.at (R.K.)

[2] Translational Gynecology Group, Department of Obstetrics and Gynecology Comprehensive Cancer Center, Medical University of Vienna, Waehringer Guertel 18-20, 1090 Vienna, Austria; dan.cacsire-castillo@meduniwien.ac.at (D.C.C.-T.); christian.singer@meduniwien.ac.at (C.F.S.)

[3] Department of General Gynecology and Gynecologic Oncology, Medical University of Vienna, Waehringer Guertel 18-20, 1090 Vienna, Austria; heinz.koelbl@meduniwien.ac.at

* Correspondence: wolfgang.schreiner@meduniwien.ac.at

Abstract: Estrogen and progesterone receptors being present or not represents one of the most important biomarkers for therapy selection in breast cancer patients. Conventional measurement by immunohistochemistry (IHC) involves errors, and numerous attempts have been made to increase precision by additional information from gene expression. This raises the question of how to fuse information, in particular, if there is disagreement. It is the primary domain of Dempster–Shafer decision theory (DST) to deal with contradicting evidence on the same item (here: receptor status), obtained through different techniques. DST is widely used in technical settings, such as self-driving cars and aviation, and is also promising to deliver significant advantages in medicine. Using data from breast cancer patients already presented in previous work, we focus on comparing DST with classical statistics in this work, to pave the way for its application in medicine. First, we explain how DST not only considers probabilities (a single number per sample), but also incorporates uncertainty in a concept of 'evidence' (two numbers per sample). This allows for very powerful displays of patient data in so-called ternary plots, a novel and crucial advantage for medical interpretation. Results are obtained according to conventional statistics (ODDS) and, in parallel, according to DST. Agreement and differences are evaluated, and the particular merits of DST discussed. The presented application demonstrates how decision theory introduces new levels of confidence in diagnoses derived from medical data.

Keywords: biomarkers; decision theory; gene expression; breast cancer; receptor status; precision medicine; personalized medicine; data science; mathematical oncology

1. Introduction

1.1. Biomarkers: A Cornerstone of Personalized Medicine for Breast Cancer

Biomarkers gain importance in selecting treatments optimized and personalized to the individual needs of patients, as envisaged in personalized medicine [1–4]. Breast cancer treatment has also benefited from personalized medicine, via molecular subtyping [5–8], the Gene expression Grade Index [9], pathway analysis and networks [10–14] and a plethora of expression signatures [15–22], dedicated to special questions and issues. Well-known indicators supporting therapy selection for breast cancer are PAM50 [23,24], PREDICT [25], and the Genomic Grade Index [26].

For breast cancer, the HER2-status (human epidermal growth factor receptor 2) of a patient is one of the most important prognostic factors [27,28]. The majority of patients (75–85%) are HER2-negative and, therefore, have a much better prognosis. In this work, we focus on these and disregard HER2-positive ones, in order to increase the homogeneity of data and precision of predictions. HER2 is routinely determined by immunohistochemistry (IHC). Numerous studies covered the significance and accuracy of estimates [29–36]. In a previous paper [37], we described, in detail, how to select patients who are HER2 negative to a high degree of confidence by using the ODDS method. We draw on the very same database in the current work.

Among HER2-negative patients, the hormone receptor status of estrogen (ER) and progesterone (PGR) are of focal importance. Clinically, patients are considered receptor-positive if at least one of both receptors (ER or PGR) is found positive. Since hormone receptors play a role in promoting cancer, hormone therapy has to be part of an effective treatment. In patients without metastases, hormone therapy may even render chemotherapy unnecessary.

However, if the receptor status is accidentally estimated as false positive, hormone treatment will not work and the patient may be deprived of life-saving chemotherapy. Therefore, numerous studies have evaluated the quality of receptor assessment [38,39] and revealed a possible rate of misclassification of 10% to 20% [40–42]. Although standard operation procedures have been implemented [40,43,44], the improvement of precision is still desirable [45].

One possibility is adding information from gene expression. Some approaches merely used visual inspection to set cut-points between positive and negative [46], some used frequency distributions of expression values [42,47], random sampling [48] or fuzzy rules [49] and other methodological advances in gene expression analysis [50–57].

We have elaborated and improved the above approaches [58,59] by introducing **Dempster–Shafer** decision Theory (DST) [60] into personalized medicine [37]. This is promising, since decision theory has demonstrated its benefit in many technical settings, such as self-driving cars [61–63], observing a driver's vigilance [62], aviation security [64,65] and also in some medical applications, such as image-based decisions [66], diagnosis of prostate [67] and breast cancer [68]. The specific merit of DST is the capability to handle unclear or contradicting information obtained from different sources about the same issue in question (e.g., receptor status). DST is able to combine multifactor and even diverging evidence, according to exact algorithms, with the potential to increase the precision of medical decisions.

In the present work, we draw on data from our previous paper [37] and elaborate on the key differences between classical statistics and DST. Ternary plots are introduced for the interpretation of probabilities, in case of contradicting evidence—a potent concept from technics is tailored to the needs of personalized medicine.

1.2. Basic Concepts of Decision Theory for Hormone Receptor Status Assessment

DST is a general theory for reasoning with uncertainty [69]. It starts with the outcome of measuring processes, rather than from 'true values' present in reality, as conventional statistics does. We present DST in a mode with only two statuses, '+' and '−'. This simplifies the formalism significantly.

Suppose some continuous variable, d, is being measured (e.g., $d \mathrel{\hat{=}}$ deepness of IHC-staining). Conventional statistics would derive a single number from d, the probability p for the receptor being positive, given the measured value of d. Consequently, $1 - p$ would be the probability for being negative.

DST provides *two* numbers to characterize possible predictions based on measuring d:

- The belief $\alpha(d)$, gives the probability (weight) that, upon measuring this particular value of d, the prediction 'positive' can be made based on the quality of measurement (classification 'with full right').
- The uncertainty $\theta(d)$, characterizing the probability (weight), that the prediction 'positive' could root in chance and not in quality of measurement. Belief and uncertainty taken together yield the total probability (termed 'plausibility') to obtain the prediction 'positive', given the measured value of d ($\alpha + \theta = pl$).
- Finally, a third number can be computed from belief and uncertainty, the probability $\beta(d)$ for yielding the prediction 'negative' by quality of measurement, given d. We always have: $\alpha + \theta + \beta = 1$; hence, β can be computed from α and θ.

In the special case of only two statuses, as considered here, the triple (α, β, θ) is equivalent to a piece of 'evidence'. DST not only yields probabilities for a positive versus negative outcome, but, additionally, incorporates the uncertainty of prediction [70]. This represents a significant surplus and motivates its introduction into personalized medicine.

A second advantage of DST is its capability of merging evidence from different sources (see also Figure 1 and the graphical abstract). In our case these will be:

- For estrogen (ER)
    - Receptor status predicted from expression of the receptor gene
    - Receptor status predicted from expression of a co-gene
        - Combining above evidence by Dempster evidence combination rule $\oplus_D$
    - Receptor status predicted from IHC
        - Combining evidence from gene expression and IHC by Yager evidence combination rule $\oplus_Y$

- For progesterone (PGR)
    - Receptor status predicted from expression of the receptor gene
    - Receptor status predicted from expression of a co-gene
        - Combining above evidence by Dempster evidence combination rule $\oplus_D$
    - Receptor status predicted from IHC
        - Combining evidence from gene expression and IHC by Yager evidence combination rule $\oplus_Y$

- Hormone receptor status is finally obtained by combining the statuses of estrogen and progesterone using a multiplicative combination rule $\otimes$

Details and references regarding the above workflow will be given in the following and are illustrated in Figure 1.

1.3. Ternary Plots: A Novel View on Evidence in Personalized Medicine

Another important point of the current work is to introduce so-called 'ternary plots', as a tailored tool to display not only probabilities, but also the uncertainties involved. We will evolve the framework step by step, to contrast conventional statistics against DST, thereby featuring the surpluses of DST.
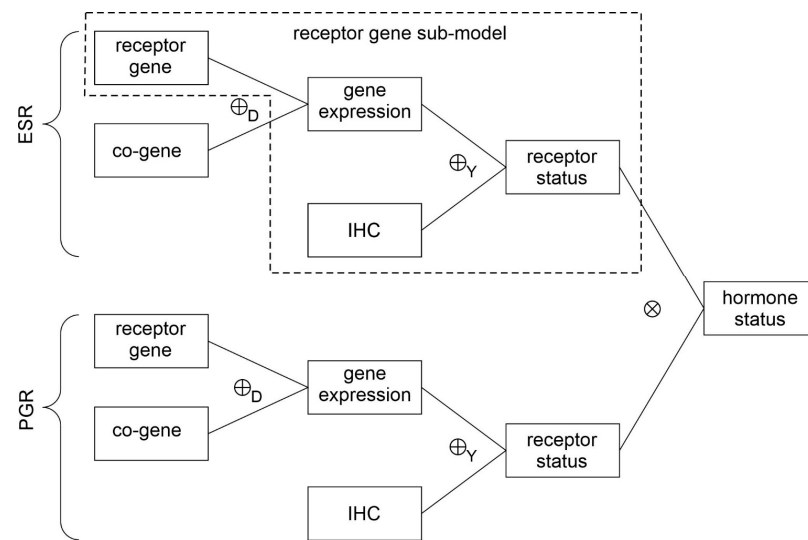
Figure 1. Combining evidence. For estrogen (ESR) and progesterone (PGR) similar procedures are applied to obtain receptor statuses. First evidence for receptor gene and co-gene are combined by Dempster rule ($\oplus_D$) and the result is combined with evidence for IHC by Yager rule ($\oplus_Y$). Finally, receptor statuses for estrogen and progesterone are combined ($\otimes$) to obtain hormone receptor status. For detailed illustration of evidence combination (Sections 2.2.1–2.2.4 and Appendix A.6) we start with a 'receptor gene sub-model' indicated by the dashed polygon.

2. Materials and Methods

2.1. Preliminaries on the Structure of the Methods' Section

For the sake of readability and to present this paper self-contained, data cleansing and the concept of responsibility functions are not expanded in the methods sections but recapitulated in the Appendices A.1–A.3. These computational procedures equal those detailed in our previous work [37]. Moreover, the methods' Sections 2.2.1–2.2.4 as well as Appendix A.6 are restricted to a single gene for didactical reasons (see the 'receptor gene sub-model' in Figure 1). In a thorough primary introduction, it seems important to demonstrate in detail how evidence from gene expression and IHC intermingle and eventually produce remarkable patterns in the data. These patterns demonstrate the dominant impact of IHC status on final predictions. Figures in Sections 2.2.1–2.2.4 and Appendix A.6 exemplify intricate features on data for estrogen but methods are general and identically apply to all other parts of the 'full' model.

In the final sections of methods (Section 2.3), we return to the full model shown in Figure 1, including co-genes. Line-like patterns are smeared out and do not remain visible as clearly as in the single-gene case. This full model was used to obtain the actual results for the patient cohort (Section 3).

2.2. Estrogen Receptor Gene Sub-Model

2.2.1. Logistic Regression as Prerequisite

Receptor status is related to gene expression (x_{Expr}) as follows: The responsibility function for positive receptor status, r_+, defines the probability for a positive receptor status, given the expression value x_{Expr}. Likewise, r_- relates to negative receptor status. We used logistic regression

$$
\begin{aligned}
\mathrm{r}_+\left(x_{\text{Expr}}|c_0, c_1\right) &= \frac{\exp\left(c_0 + c_1 x_{\text{Expr}}\right)}{1+\exp\left(c_0 + c_1 x_{\text{Expr}}\right)} \\
\mathrm{r}_-\left(x_{\text{Expr}}|c_0, c_1\right) &= 1 - \mathrm{r}_+\left(x_{\text{Expr}}|c_0, c_1\right)
\end{aligned}
\tag{1}
$$

and estimated the parameters c_0 and c_1 against IHC-measurements, separately for each gene and co-gene, for results see Table A2. Figure 2 shows the responsibility function r_+ for

positive estrogen (red dashed curve). r_-, for negative estrogen (blue dashed curve), is based on the same regression coefficients, see Equation (1). A similar analysis was performed for progesterone, see Figure A1 for graphics and Table A2 for numerical values.
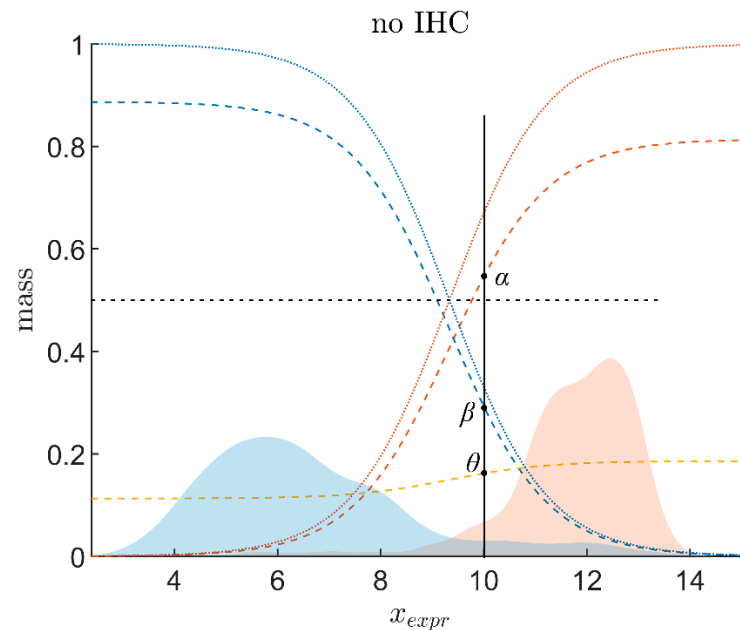
Figure 2. Logistic regression to obtain responsibility functions for decision theory evidence (data for estrogen, gene ESR1). Red-shaded area: distribution of gene expression for receptor positive (according to IHC) computed from density kernel estimates [71–73]. Blue shaded area: gene expression for negative IHC. IHC receptor status (IHC$^+$1, IHC$^-$0) was subjected to logistic regression versus gene expression (x_{Expr}). Responsibility functions for receptor positivity, r_+, (dotted red curve) and r_- (dotted blue) were thus obtained. It will be shown later (Equation (4)) that r_+ has to be multiplied by an upper limit, $\hat{\alpha}_{\text{Expr}}$, to obtain the actual belief α_{Expr}, see the dashed red curve. Likewise for β_{Expr} (dashed blue). Uncertainty: ochre. For a given expression value, e.g., $x_{\text{Expr}} = 10$, one can read off belief in positive (α), belief in negative (β) and uncertainty (θ). Note that analog concepts apply to any other gene of the full model.

2.2.2. Evidence of Receptor Status Based on Expression of Receptor Gene

Based on logistic regression, gene expression measurements lend themselves to derive evidence of receptor status according to Dempster–Shafer decision theory [70]. In the following, we formulate rules and principles in general terms of 'gene expression', x_{Expr}, to keep notation general (later on, the first example with real data will specifically refer to *estrogen* receptor diagnostics. Even later, the very same procedure will be applied to progesterone).

Assume the variable gene expression, x_{Expr}, is prognostic for receptor status. Given a measurement of x_{Expr}, DST attributes *two* independent numbers, as outlined below:

- $\alpha_{\text{Expr}}(x_{\text{Expr}})$: the belief (sometimes also called 'degree of belief' or 'credibility' [74]) for receptor status being positive on good grounds or by quality of the measuring method that has yielded x_{Expr};
- β_{Expr}: the belief (probability) for receptor status being non-positive (i.e., negative) on good grounds or by quality of the measuring method;
- θ_{Expr} is a third quantity considered: the probability that the receptor status is uncertain.

α, β and θ are also called 'masses' of the respective outcomes. They are by definition larger than or equal to zero ($\alpha_{\text{Expr}} \geq 0$, $\beta_{\text{Expr}} \geq 0$, $\theta_{\text{Expr}} \geq 0$), and if a mass equals zero, in

our setting zero corresponds to the 'empty set', i.e., an outcome that will never be found [75]. Masses always add up to unity, and hence we talk about normalized mass functions [76]:

$$\alpha_{\text{Expr}} + \beta_{\text{Expr}} + \theta_{\text{Expr}} = 1 \qquad (2)$$

Hence a third number is in fact redundant (may always be computed from the other two). Decision theory even considers a fourth quantity, called plausibility; it is also redundant but intuitive and useful:

$$pl_{\text{Expr}} = \alpha_{\text{Expr}} + \theta_{\text{Expr}} = 1 - \beta_{\text{Expr}} \qquad (3)$$

pl_{Expr} indicates the probability of a positive status being plausible, given the measurement x_{Expr} as is. The plausibility of a given outcome sums up everything either supportive or neutral, but excludes everything advocating the opposite outcome. The exactly opposite outcome is represented by β_{Expr}.

The output of above procedure is the evidence $(\alpha_{\text{Expr}}, \beta_{\text{Expr}})$ for receptor status, based on the expression (x_{Expr}) of a gene (in general); data in Figure 2 were shown for the receptor gene of estrogen (ESR1). Note that finally, in the 'full' model, 4 such pieces of evidence (4 pairs of numbers) will be obtained: (1) for the estrogen receptor gene and (2) its co-gene; (3) for the progesterone receptor gene and (4) its co-gene.

The beliefs in receptor positive, α_{Expr}, and negative, β_{Expr}, may be obtained from gene expression alone, x_{Expr}, as demonstrated above. Doing so, maximum expression corresponds to a responsibility function $r_+(x_{\text{Expr}})$ close to 1, see Figure 2. However, not even a gene expression that large can guarantee that the receptor is truly positive. Hence, the belief in positivity, α_{Expr}, actually must be less than 1.

We chose to model this fact by a factor, $\hat{\alpha}$, called 'upper limit for belief' in Table A2. For details of calculation see Appendices A.4 and A.5.

All in all we obtain the belief in receptor positivity after measuring x_{Expr}:

$$\begin{aligned} \alpha_{\text{Expr}}(x_{\text{Expr}}) &= \hat{\alpha}_{\text{Expr}} \cdot r_+(x_{\text{Expr}}|c_0, c_1) \\ \beta_{\text{Expr}}(x_{\text{Expr}}) &= \hat{\beta}_{\text{Expr}} \cdot r_-(x_{\text{Expr}}|c_0, c_1) \end{aligned} \qquad (4)$$

α_{Expr} is represented by the increasing solid red curve in Figure 2, β_{Expr} by the declining blue one. The remaining uncertainty, θ_{Expr}, is easily computed from reformulating Equation (2)

$$\theta_{\text{Expr}} = 1 - \alpha_{\text{Expr}} - \beta_{\text{Expr}} \qquad (5)$$

and is shown as ochre curve in Figure 2. The two numbers $(\alpha_{\text{Expr}}, \beta_{\text{Expr}})$ are collectively called 'evidence' of receptor status, given a measurement of the continuous variable 'gene expression', x_{Expr}. They enrich the information given by a single number, the probability p, known from conventional statistics, quantifying the chances of receptor statuses, a similar procedure applies to the receptor gene of progesterone, see Figure A1.

2.2.3. Combining Evidence from Receptor Gene Expression and IHC

To further increase precision of receptor status diagnostics, evidence from gene expression $(\alpha_{\text{Expr}}, \beta_{\text{Expr}})$ and evidence from IHC $(\alpha_{\text{IHC}}, \beta_{\text{IHC}})$ are combined by so-called 'evidence combination rules' (ECR). DST offers several such rules [69,74], out of which we consider two, the 'Dempster–Shafer' ECR, $\oplus_{\text{D}}$, and the Yager ECR, $\oplus_{\text{Y}}$ [77,78]. We chose the Yager rule, as it more easily accommodates contradicting items of evidence, see also Section 4.4 in the discussion. Performing some algebra, as detailed in our previous work [37], one finally obtains:

$$\begin{aligned} \alpha_{\text{Rez}} &= \alpha_{\text{Expr}}\alpha_{\text{IHC}} + \theta_{\text{Expr}}\alpha_{\text{IHC}} + \alpha_{\text{Expr}}\theta_{\text{IHC}} \\ \beta_{\text{Rez}} &= \beta_{\text{Expr}}\beta_{\text{IHC}} + \theta_{\text{Expr}}\beta_{\text{IHC}} + \beta_{\text{Expr}}\theta_{\text{IHC}} \\ \theta_{\text{Rez}} &= \theta_{\text{Expr}}\theta_{\text{IHC}} + \alpha_{\text{Expr}}\beta_{\text{IHC}} + \beta_{\text{Expr}}\alpha_{\text{IHC}} \end{aligned} \qquad (6)$$

As IHC-evidence is made up of two sets of constants, combination with gene expression yields two sets of curves, one for IHC$^-$ and one for IHC$^+$, see Figure 3.

(a) **(b)**
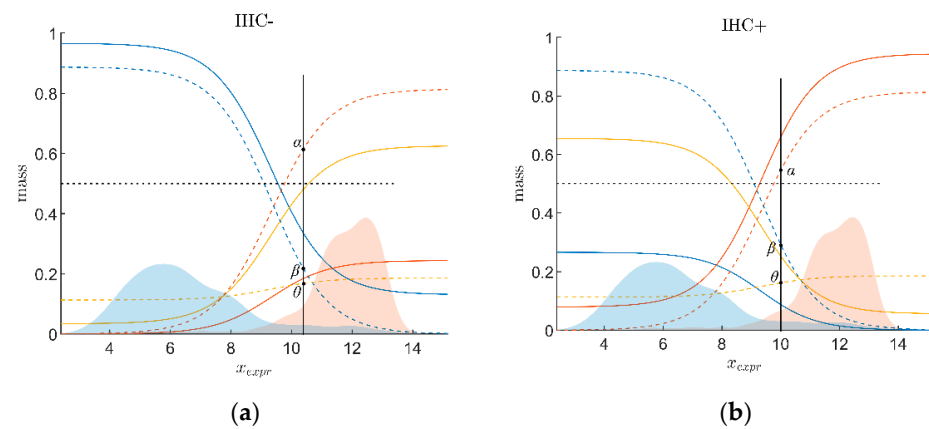
Figure 3. Combining receptor gene expression and IHC. Taking the estrogen receptor as an example, we demonstrate the principles of combining evidence from gene expression and IHC by the Yager evidence combination rule. (**a**) For IHC-negative estrogen receptor status. (**b**) For IHC-positive estrogen receptor status. Dotted lines represent beliefs merely based on gene expression, without considering IHC estimates (identical to the beliefs in Figure 2). Beliefs for gene expression combined with IHC estimates (via Yager evidence combination rule) are shown in solid lines. Clearly, a negative IHC estimate (**a**) strengthens the belief in negative (solid blue runs above dotted blue) and weakens the belief in positive (solid red runs below dotted red) for a given expression value, x_{Expr}. As opposed, a positive IHC estimate (**b**) strengthens the belief in positive (red) and weakens the belief in negative (blue).

A definite decision for positive receptor status is obtained if the combined evidence exceeds 0.5 ($\alpha > 0.5$). In that case the belief in positive surmounts the sum of both other beliefs ($\beta + \theta \le 0.5$) and dominates. Hence, the dotted line $\alpha = 0.5$ represents a decision border and will be analogously outlined in the following figures.

2.2.4. Ternary Plots of Evidence for Personalized Medicine: A Primer

Note that belief, plausibility and uncertainty are not independent but always sum up to unity for a given sample, see Equation (2). This mathematical property allows for a special graphic display, called 'ternary plot', as follows. When plotting these data in an ordinary 3-dimensional scatter plot with coordinates (α, β, θ), points of all samples lie within a single plane (of evidence), see Figure 4a. This is due to Equation (2), which—in mathematical terms—is nothing else than the equation of a plane in three dimensions [79]. This 'plane of evidence' may be viewed in orthogonal projection (https://en.wikipedia. org/wiki/Orthographic_projection (accessed on 26 March 2022)) which still contains all information but fits into two dimensions and is called 'ternary plot', see Figure 4b.

Ternary plots are widely used in technology and science but have only marginally entered the medical sciences [80]. They might also gain importance in personalized medicine but deserve some skillful understanding. Hence we provide a short primer.

A ternary plot is powerful whenever three quantities (hence the name 'ternary') add up to a constant, for each individual considered. For example, a biological fluid (say milk) may be composed of water, protein and fat (three components) and nothing else. Clearly, the percentages of water, protein and fat must then add up to 100%. For a set of milk samples, these percentages may be visualized by points in a 3D scatter plot, such as Figure 4a. If we consider mixtures of different composition (e.g., skimmed, normal and fat milk and many other possible kinds) and plot their corresponding 3D points, we will be surprised to realize that all these points lie within a single flat plane in 3-dimensional space. The reason is a mathematical one: if coordinates always add up to a given constant,

this is the very representation of a plane in mathematical terms [79]. This may be fruitfully exploited for personalized medicine as follows:
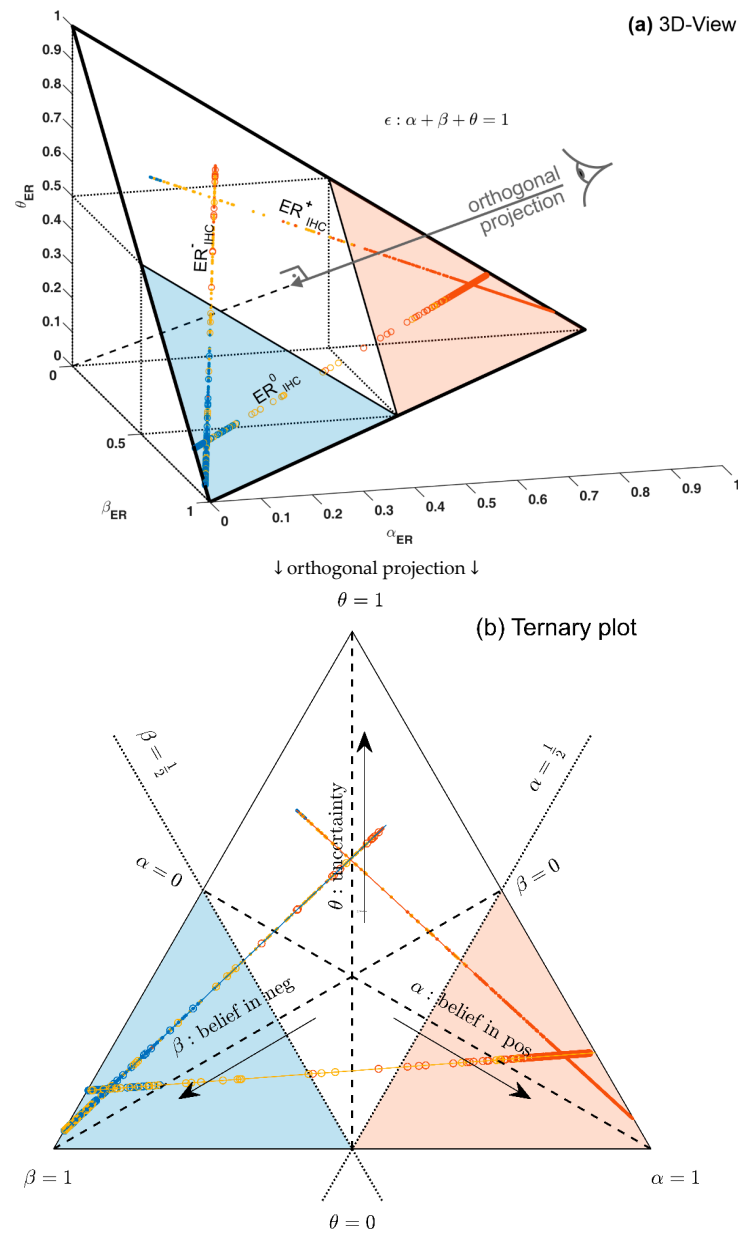

Figure 4. 3D plot and ternary plot of evidence for estrogen receptor status. (**a**) ordinary 3D plot. (**b**) ternary plot. Orthogonal projection of the 3D scatter plot (as indicated) in panel (**a**) yields the ternary version, panel (**b**). In each corner, one piece of evidence dominates and both others are zero, e.g., ($\alpha = 1$, $\beta = 0$, $\theta = 0$ in the lower right corner). Note that the baseline of the ternary runs along the diagonal through the bottom plane of the 3D plot: along this bottom side, α runs from zero to 1 from left to right and β in reverse (right to left), hence sides of a ternary triangle do not represent usual ordinate axes, please refer to the tutorial (2.2.4). Midway points of triangle sides mark decision boundaries, e.g., $\alpha = \beta = 0.5$ between positive and negative. Triangular areas contain definite results, either positive ($\alpha \geq 0.5$, shaded red) or negative ones ($\beta \geq 0.5$, shaded blue). The kite-shaped area (white) represents undecided status according to DST. For each value of ER_{IHC} see labeling (ER_{IHC}^{+}, ER_{IHC}^{-}, ER_{IHC}^{u}) evidence lies on a specific straight line due to mathematical reasons; samples with known IHC are shown as dots, samples with unknown IHC as circles. Coloring of samples *according to ODDS method, not DST*. Samples positive according to ODDS (red) may well lie within the undecided region according to IHC, etc.

What is true for three components of a substance (milk) is also true for evidence composed of three numbers, α, β and θ, since they also add up to unity due to Equation (2). Hence, points of evidence (α, β, θ) for any single patient, lie within the same (2-dimensional) plane in 3D. This plane always lies in the same, specific position and orientation, for the following reason: the point $(\alpha = 1, \beta = 0, \theta = 0)$, represents a valid point of evidence (adding up to unity) and must be part of the plane. Therefore, the plane cuts the α–axis at $\alpha = 1$, see Figure 4a. Likewise, the plane also cuts both other axes at $\beta = 1$ and $\theta = 1$, respectively. This uniquely defines an equilateral triangle in the 3D coordinate system, see Figure 4a.

Even though the plane of evidence lies embedded in a 3D coordinate system, it is by itself just a 2-dimensional object, as every flat plane is. Therefore, without any loss of information regarding the location of points (representing evidence) we may perform an orthogonal projection along the heavy arrow shown in Figure 4a. This yields a so-called 'isometric view'. The triangle, viewed face-on, appears equilateral, now in two dimensions, see Figure 4b.

A ternary plot does not have its axes at right angles, as ordinary plots do. To read off the coordinates of a point from such a ternary plot, several methods are available, out of which we propose the following (altitude method), illustrated in Figure 5:
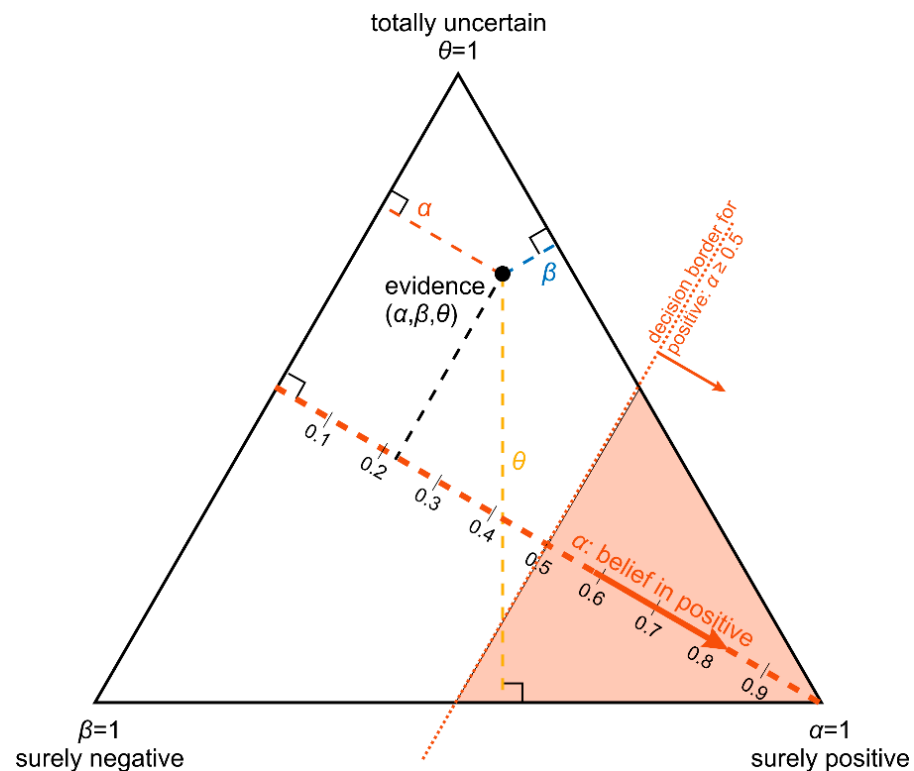
Figure 5. Principles of a ternary plot: obtaining coordinates by the altitude method. Decision border $\alpha \geq 0.5$: 'positive to the best of our knowledge' or 'positive is more likely than anything else'.

Each of the three components of evidence, e.g., α, the 'belief in positive', has its own scale, see the dashed heavy line in orange; it starts on the left with $\alpha = 0$ at a right angle from a triangle's left side and runs towards the opposing corner, where $\alpha = 1$ (indicating 'surely positive'). See also the scale with numerical values aside. The two other scales, for β and θ, are defined analogously (not shown for simplicity).

Given some point within the ternary plot (see the heavy black dot in Figure 5), corresponding evidence components (α, β, θ) can be read off as follows. Note the lines being drawn perpendicular to each side of the triangle (light dashed lines in red, blue and beige)—they represent the axes for quantification. Values α, β, θ (red, blue, and beige, respectively)

can be read off from the corresponding axis' scale. In this example (Figure 5), the plotted point of the evidence produces a reading of $\alpha \cong 0.22$.

Note also the following intriguing features of this ternary plot:

- Parallel lines at right angles with one axis represent constant values for the respective variable (as with ordinary right-angle axes). In particular, the line crossing the α–axis at $\alpha = 0.5$ (dotted red) discriminates points with $\alpha \leq 0.5$ (left upper) from those with $\alpha > 0.5$ (towards lower right corner), and hence represents a decision border; points right of this border are predicted 'positive', since their evidence for positive is greater than for all other options ('negative' and 'uncertain') taken together.
- Decision borders segregate subsets of samples. For example, all samples within the triangle in the lower right of $\alpha = 0.5$ (shaded light red) comprise samples predicted positive. Similarly, the subsets of negative and uncertain samples may be defined, see Figure 4b.
- In each corner one piece of evidence totally dominates, assuming a value of unity ($\alpha = 1$: 'surely positive'; $\beta = 1$: 'surely negative' and $\theta = 1$: 'totally uncertain').
- Conversely, the footing point of each axis (e.g., $\alpha = 0$) means that there is no indication whatsoever for the prediction at opposing corner. For example, $\alpha = 0$ along the left side of the triangle, means that there is no indication whatsoever for a 'positive' prediction. All evidence is shared between 'negative' and 'uncertain' (β and θ). In this case $\beta + \theta = 1$.
- A special role is played by the triangle's bottom edge, running from $\beta = 1$ (left) towards $\alpha = 1$ (right): for each sample along this line uncertainty θ equals zero, and all evidence is shared between belief in positive (α) and belief in negative (β), e.g., $\alpha = 0.6$ and $\beta = 0.4$, while $\theta = 0$. One may legitimately ask: "Does this mean that the prediction was made for sure?". Since $\alpha > 0.5$ and dominates both other options, we consider this prediction clearly positive. However, $\alpha = 0.6$ is no more than a probability and not that much larger than the probability of the opposite outcome, $\beta = 0.4$. In reality, the outcome may well result in a negative prediction. If $\theta = 0$, evidence masses revert back to ordinary probabilities: $p^+ = 0.6$ for positive and hence $p^- = 0.4$ for negative, without indicating any uncertainty about the estimates of these two numbers. Thus, for $\theta = 0$, decision theory's evidence coincides with ordinary probabilities. In DST terminology the evidence is said to turn 'Bayesian' [74].
- In general, for $\theta > 0$, decision theory not only gives estimates for probabilities (α, β) but additionally indicates the uncertainty of those (θ). It hence offers a wider scope of evidence, valuable in particular for personalized medicine.

Ternary plots allow for a highly transparent comparison of our two classification methods (ODDS versus DST) for each single sample:

- The *location of the point indicates the prediction according to DST* shown by the respective area: red triangular area for positive (+), blue for negative (-) and the white, kite-shaped area for inconclusive (inc).
- At the same time, *coloring of points indicates prediction according to ODDS*. For most samples, both predictions match. For some samples however, they differ, thus perfectly outlining the contrast between the two prediction methods.

Although ternary plots may seem somewhat unusual for medical application, they offer the unique capability to display *three* variables in *two* dimensions, provided their sum is constant, which is true for evidence and many other variables. We think it worth the effort to introduce ternary plots into the field of personalized medicine. They are the most adequate tool for quantitatively presenting evidence, and may in the future represent a cornerstone of personalized medicine.

2.3. Full Model: Evidence, Based on IHC, Genes, Co-Genes

In Sections 2.2.1–2.2.4 and Appendix A.6 description was restricted to the receptor gene (no co-gene considered) in order to explain more transparent details. Now we revert to the whole model, including co-genes, see the flow chart of evidence in Figure 1.

First, we supplement estrogen expression evidence $(\alpha_{\mathrm{Gen}}, \beta_{\mathrm{Gen}})$ by evidence $(\alpha_{\mathrm{Co}}, \beta_{\mathrm{Co}})$ from its co-gene, AGR3; the very same procedure outlined in Sections 2.2.1 and 2.2.2 is carried out to obtain these results, see Table A2.

2.3.1. Progesterone Evidence

Numerical results of the logistic regression for progesterone are shown in Table A2, for responsibility functions, see Figure A1. The co-gene of progesterone, incidentally, was estrogen, see Table A2.

2.3.2. Combining Evidence Form Genes and Co-Genes

Next, evidence from genes and co-genes are combined by the Dempster Evidence Combination Rule ($\oplus_{\mathrm{D}}$) to obtain the joint evidence from gene expression:

$$(\alpha_{\mathrm{Expr}}, \beta_{\mathrm{Expr}}) = (\alpha_{\mathrm{Gen}}, \beta_{\mathrm{Gen}}) \oplus_{\mathrm{D}} (\alpha_{\mathrm{Co}}, \beta_{\mathrm{Co}}) \tag{7}$$

In detail, the Dempster rule [77] reads:

$$
\begin{aligned}
\alpha_{\mathrm{Expr}} &= \frac{\alpha_{\mathrm{Gen}}\alpha_{\mathrm{Co}} + \theta_{\mathrm{Gen}}\alpha_{\mathrm{Co}} + \alpha_{\mathrm{Gen}}\theta_{\mathrm{Co}}}{1 - \alpha_{\mathrm{Gen}}\beta_{\mathrm{Co}} - \beta_{\mathrm{Gen}}\alpha_{\mathrm{Co}}} \\
\beta_{\mathrm{Expr}} &= \frac{\beta_{\mathrm{Gen}}\beta_{\mathrm{Co}} + \theta_{\mathrm{Gen}}\beta_{\mathrm{Co}} + \beta_{\mathrm{Gen}}\theta_{\mathrm{Co}}}{1 - \alpha_{\mathrm{Gen}}\beta_{\mathrm{Co}} - \beta_{\mathrm{Gen}}\alpha_{\mathrm{Co}}} \\
\theta_{\mathrm{Expr}} &= 1 - \alpha_{\mathrm{Expr}} - \beta_{\mathrm{Expr}} = \\
&= \frac{\theta_{\mathrm{Gen}}\theta_{\mathrm{Co}}}{1 - \alpha_{\mathrm{Gen}}\beta_{\mathrm{Co}} - \beta_{\mathrm{Gen}}\alpha_{\mathrm{Co}}}
\end{aligned}
\tag{8}
$$

Combination of gene and co-gene is carried out along the same lines for estrogen and progesterone.
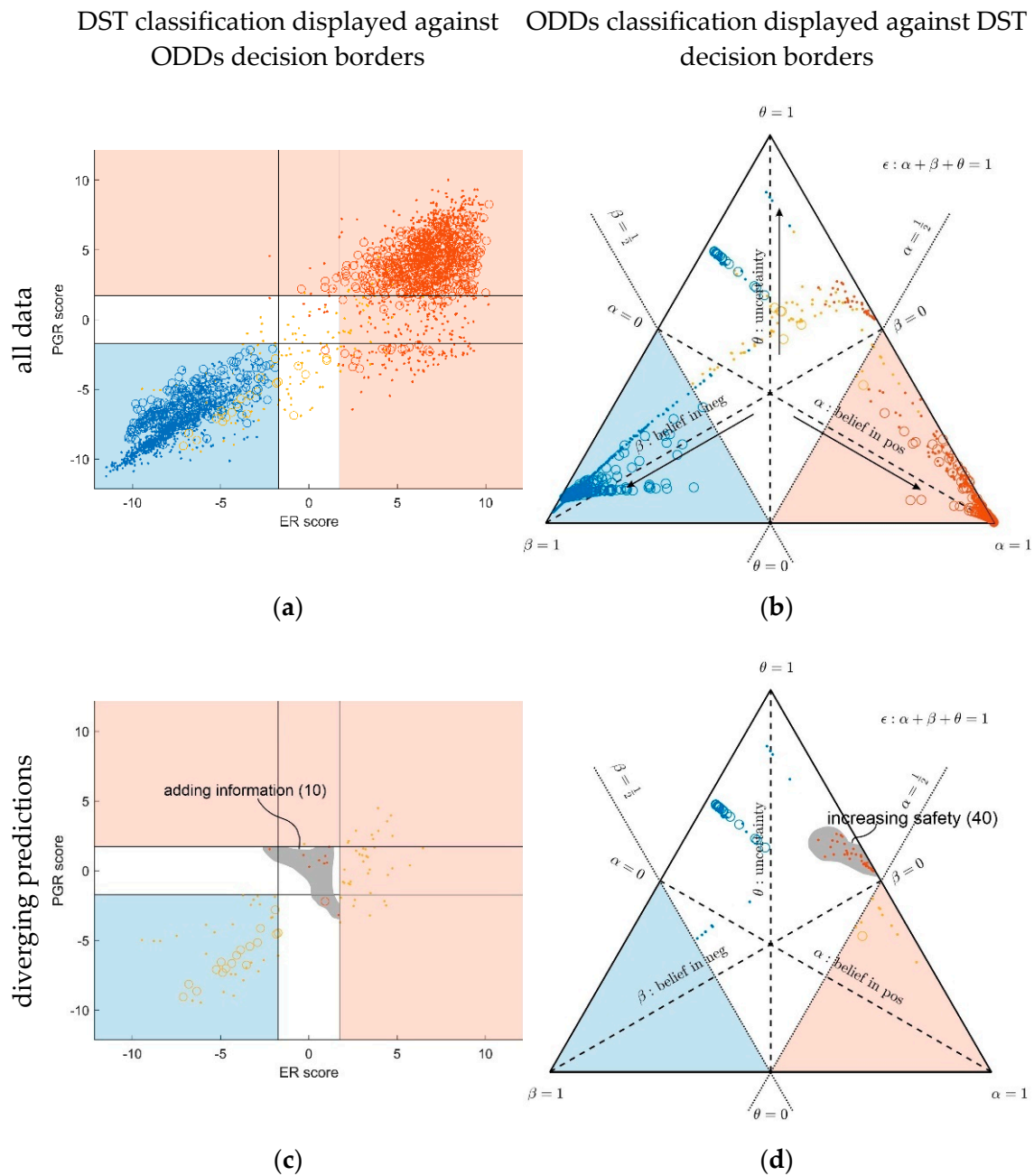
2.3.3. Combining Evidence from Gene Expression and IHC

As outlined in Section 2.2.3 for single gene case, we now combine the full gene evidence for estrogen with its IHC counterpart according to the Yager rule, see Equation (6), to obtain $(\alpha_{\mathrm{ER}}, \beta_{\mathrm{ER}}, \theta_{\mathrm{ER}})$. The very same is done for progesterone, yielding $(\alpha_{\mathrm{PGR}}, \beta_{\mathrm{PGR}}, \theta_{\mathrm{PGR}})$.

2.3.4. Combining Estrogen and Progesterone Receptor Status

In the step to follow, evidence for different targets—estrogen and progesterone—will be combined. Clinically, a breast cancer patient is considered receptor positive, if either the estrogen '***OR***' the progesterone receptor (or both) is/are positive, and treatment is assigned accordingly. Corresponding decision borders will be shown below (Figure 6). While clinical SOP (Standard Operating Procedure) draws on a crisp logical '***OR***', as implemented in the ODDS-method, DST offers a wider scope of possibilities. Evidence for estrogen $(\alpha_{\mathrm{ESR}}, \beta_{\mathrm{ESR}}, \theta_{\mathrm{ESR}})$ and progesterone $(\alpha_{\mathrm{PGR}}, \beta_{\mathrm{PGR}}, \theta_{\mathrm{PGR}})$ may be combined to obtain evidence for the overall hormone status $(\alpha_{\mathrm{H}}, \beta_{\mathrm{H}})$ as follows [37]:

$$
\begin{aligned}
\alpha_{\mathrm{H}} &= \alpha_{\mathrm{ESR}} + \alpha_{\mathrm{PGR}} - \alpha_{\mathrm{ESR}}\alpha_{\mathrm{PGR}} \\
\beta_{\mathrm{H}} &= \beta_{\mathrm{ESR}} \cdot \beta_{\mathrm{PGR}}
\end{aligned}
\tag{9}
$$

DST classification displayed against   ODDs classification displayed against DST
ODDs decision borders   decision borders

(**a**)

(**b**)

(**c**)

(**d**)

Figure 6. Hormone receptor status classified by ODDS versus Decision Theory. The same patient data were classified twice, along ODDS (left panels) and DST (right panels), for comparison. Data are shown as open circles if at least one IHC status is unknown. All other IHC statuses are shown as dots. Panels (**a**,**b**) include all 2519 patients of the cohort while panels (**c**,**d**) only display those 95 patients with predictions diverging between ODDS and DST, for easy comparison. Note the legends highlighting those patient samples which benefit from enhanced information or safety, respectively, conferred by DST. (**Left panels**): sample data arranged according to orthogonal ODDS score axes, but colored according to DST prediction. Light blue area: negative by ODDS. Light red area: positive. White L-shaped area: inconclusive by ODDS. (**Right panels**): sample data arranged according to ternary DST evidence axes, but colored according to ODDS prediction. Light blue triangle: negative by DST. Light red triangle: positive by DST. White kite-shaped area: inconclusive by DST. (**Lower panels**): only patients with diverging predictions are shown. Note two important groups marked by legends, corresponding to two cells in Table 1.

Table 1. Prediction by DST (Dempster–Shafer decision Theory) versus ODDS (conventional statistics). Predictions negative (neg), inconclusive (inc) and positive (pos). No samples are classified as fully contradicting (positive versus negative). Differences arise from samples predicted 'inconclusive': 144 (5.7%) via DST compared to 69 (2.7%) via ODDS. DST predicts more conservatively than ODDS. Cells with numbers in bold represent patients with 'gain of information (10)' and 'gain of safety (40)', respectively. As an overall measure of agreement we computed Cohen's kappa: $\kappa = 0.9287$ [81].

Number of Samples		DST			
		neg	**inc**	**pos**	**sum**
ODDS	neg	999	45	0	1044
	inc	0	59	**10**	69
	pos	0	**40**	1366	1406
	sum	999	144	1376	2519

percentage		DST			
		neg	**inc**	**pos**	**sum**
ODDS	neg	39.7%	1.8%	0.0%	41.5%
	inc	0.0%	2.3%	**0.4%**	2.7%
	pos	0.0%	**1.6%**	54.2%	55.8%
	sum	39.7%	5.7%	54.6%	100.0%

3. Results

3.1. Contrasting Predictions by ODDS versus DST

Predictions via conventional statistics (ODDS) and decision theory (DST) are directly compared for the whole patient cohort in Figure 6. To address clinical relevance, we highlight patients for which DST adds information (see legend), as well as those for which DST increases safety (see legend). For compactness, we abbreviate notation of the IHC receptor status, e.g., $ER^-_{IHC}, PGR^+_{IHC}(-,+)$ or $ER^+_{IHC}, PGR^u_{IHC}(+,0)$, with '0' representing 'undefined'. Likewise, we denote predictions (via ODDS or DST) as 'neg', 'pos' and 'inc', with 'inc' representing 'inconclusive'.

Note the following features in Figure 6:

- **In the left panels, samples are geometrically located according to ODDS scores, but color-coded according to DST prediction**.
- Decision borders in ODDS can be directly displayed in an orthogonal, 2-dimensional plot of 'scores', see Figure 6, left panels. Decision borders are defined by specific values for each receptor score (ER score, PGR score), see our previous paper [37], and, hence, appear as vertical lines for estrogen and as horizontal lines for progesterone, respectively. The rectangular region (in faint blue) denotes receptor status predicted definitely negative, the L-shaped stripe (no color) denotes inconclusive status, and the L-shaped stripe (in faint red) definitely positive predictions.
- ODDS scores incorporate IHC evidence in an additive fashion. Each of the nine possible IHC statuses $(++, --, +-, -+, +0, 0+, -0, 0-, 00)$ merely differ in shifts along the respective ODDS coordinate (ER score, PGR score). ODDS decision borders are, hence, valid for any combination of IHC statuses.
- **In the right panels, samples are geometrically located according to DST evidence, but color-coded according to ODDS**.
- Decision borders in DST are most appropriately displayed in ternary plots of evidence, see Figure 6, right panels. Decision borders run along evidence $\alpha = 0.5$ and $\beta = 0.5$, respectively, which appear as straight lines in a ternary plot. DST evidence also incorporates IHC information, and decision lines, hence, also represent unique borders in the ternary plot, valid for any combination of IHC statuses $(++, --, +-, -+, +0, 0+,$ etc.).
- **In the ternary plot, DST evidence for subsets of patient samples appear in polygonal areas**. In fact, these areas root in respective combinations of IHC statuses for estrogen and progesterone $(++, --, +-,$ etc.), as will be scrutinized in the appendix,

for those interested in mathematical details. Indeed, these polygonal areas are generalizations of those simple straight lines already seen with single gene expression data (Figure 4). Since each receptor may assume three values (+, −, 0), there are $3^2 = 9$ possible IHC status combinations for two receptors. Some IHC statuses give rise to very distinct arrangements of samples, such as 'lines'. Other IHC combinations give rise to more polygonal-shaped areas. Details will be discussed below. Data samples along these lines or polygons are seen to cross DST decision borders (dashed lines at $\alpha = 0.5$ and $\beta = 0.5$, respectively). For example, if such a subset of samples crosses from inconclusive to decided, this indicates that IHC on its own was inconclusive, but adding evidence from (increasing) gene expression finally rendered a decision:

○ A stripe of red points originates within the DST-inconclusive, kite-shaped area and protrudes into the positive triangle.

○ The stripe of blue points originates in the DST-inconclusive, kite-shaped area and protrudes into the negative triangle.

Crossing decision borders for given IHCs underpins the importance of information from gene expression being added.

3.2. Clinical Relevance of DST versus ODDS

Agreement and divergence between ODDS and DST are summarized in Table 1. Note that both methods never definitely contradict each another (positive versus negative predictions for a given sample); see the zero counts in the corners off diagonal. Differences only occur for samples predicted as inconclusive. In 59 cases, both methods agree in yielding 'inconclusive'. However, DST reports almost equal numbers of samples from ODDS-negative (45) and ODDS-positive (40) as DST-inconclusive, ending up with 144 inconclusive samples. Conversely, ODDS declares none from DST negative and only 10 from DST-positive as ODDS-inconclusive, ending up with just 69 samples rendered as inconclusive. In general, agreement between ODDS and DST is fine, with 999 + 59 + 1366 = 2424 out of 2519 samples (96.2%), as reflected by the high inter-rater agreement coefficient, Cohen's kappa: $\kappa = 0.9287$ [81].

Besides good overall agreement, possible advantages of DST may be seen twofold, cf. the cells outlined with bold face in Table 1. The very same groups of patients are highlighted with legends in Figure 6:

● For 10 patients, DST predicted a positive receptor status, whereas ODDS had predicted 'undecided'. Based on the additional information provided by DST, these patients may, upon careful reassessment, be candidates for milder therapies, possibly without chemotherapy (chemo). We, therefore, labelled this group with 'adding information' in Figure 6, panel (c).

● For 40 patients, DST predicted 'undecided', whereas ODDS had predicted 'positive'. 'Undecided' severely questions abstaining from chemo and calls for a re-assessment at least. We, therefore, labelled this group with 'increasing safety' in Figure 6, panel (d).

Hormone receptor diagnostics—in comparison with ODDS and DST—was evaluated regarding its impact on survival. Figure 7 shows survival, free from recurrence, for several relevant subgroups listed in Table 1. Acronyms in the legend of Figure 7 correspond to those in Table 1, and figures in the legend give the numbers of patients with survival data available and number of events (i.e., recurrences) in parenthesis. Naturally, the two largest groups are those that ODDS and DST found in agreement (neg/neg, pos/pos)—they exhibit rich survival curves, with many patients and numerous events. Subgroups with disagreement between ODDS and DST (fortunately) contain only few patients, reflecting the fact that, already, ODDS was an advanced, accurate prediction method. The point of largest possible merit is the subgroup pos/inc: 40 patients considered positive by ODDS could have been deprived of chemotherapy, although being eventually negative. Within this group, survival data were available only for seven, relegating statistical testing meaningless.
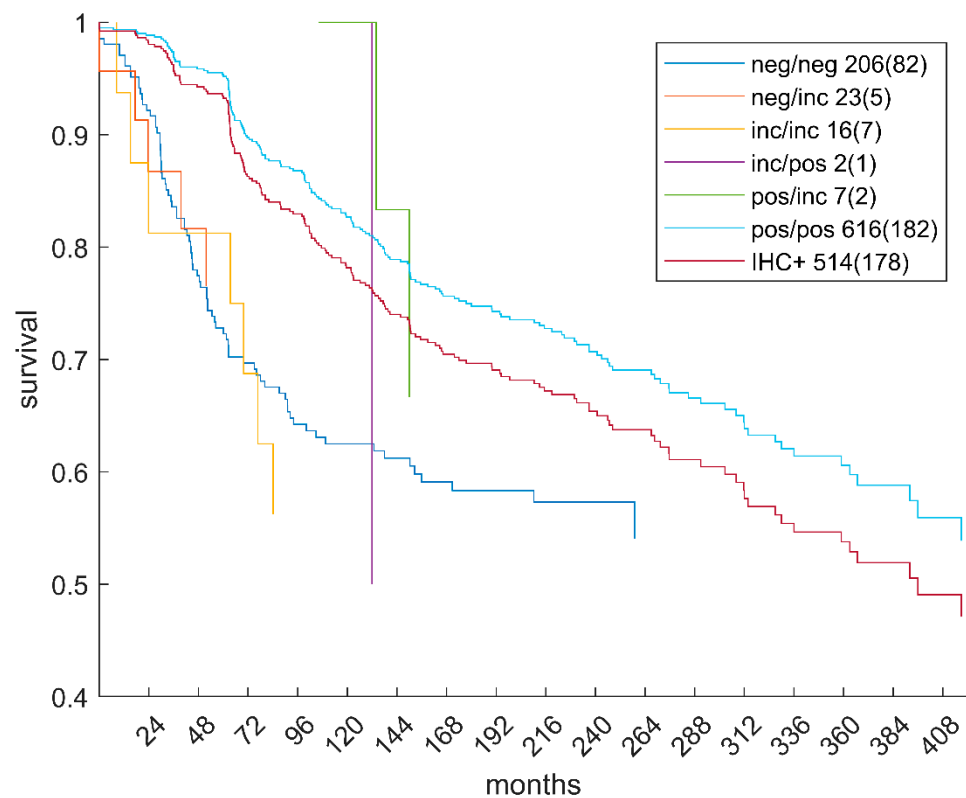
Figure 7. Survival free from recurrence. Kaplan–Meier estimates were obtained for several patient subgroups in Table 1. Legend acronyms refer to cells in Table 1 as ODDS/DST, figures give numbers of patients with survival data available and number of events (i.e., recurrences) in parenthesis. The curve 'IHC+' refers to patients diagnosed receptor positive according to current clinical standards, i.e., positive for estrogen or progesterone (or both).

For comparison, the IHC+ group was also evaluated, incorporating patients receptor positive either for estrogen *OR* progesterone, see Figure 7. Such patients are considered receptor positive and treated accordingly by 'conventional' clinical therapy allocation. Compared to these, our pos/pos group enjoyed definitely superior survival (log-rank $p = 0.03$). Since all patients considered in our study were actually treated according to conventional, clinical 'IHC+', we might speculate as follows: this actual, former treatment as 'IHC+' was confirmed post hoc in our study (by pos/pos) as correct and, hence, these patients experienced much better survival.

Over the years, hormone receptor status has become the most important predictive parameter, which allows for an identification of endocrine-sensitive invasive tumors. The use of hormone-receptor-targeted treatment strategies is associated with an approximately 50% reduction in recurrences and a reduction in breast-cancer-attributed deaths by approximately 30%, and receptor status assessment has, therefore, become the single most important biomarker in early and advanced breast cancer. A correct classification of endocrine sensitivity by receptor measurement is, therefore, critical for individualized treatment, since falsepositive results lead to overtreatment and therapy-associated side effects, which range from menopausal symptoms, infertility and depression, to bone loss and an increase in fractures, and other significant side effects. False negative results, on the other side, subject patients to under-treatment and a profound worsening of the long-term outcome. These profound clinical consequences are contrasted by a number of technical uncertainties: the hormone receptor status is presently assessed by immunohistochemistry, and different standards in tissue fixation, varying protocols, the myriad of commercially available antibodies, inter-observer variability and other technical issues compromise an objective assessment. Moreover, while some labs use a cut-off of 10% of hormone receptor

positive cells, others prefer a cutoff of 1%, thus, limiting the value of the current gold standard in receptor assessment. Within this context, prediction models, such as DST and ODDS, can add to further ascertainment of the receptor status. The decision of which model to use could be factored into the decision tree and allow for a more personalized treatment, in the sense that the more conservative DST could be applied in older and frail patients, in whom the significant side effects of endocrine therapy need to be balanced against competing mortalities and might lead to an omission of endocrine therapy, and an additional IHC, performed by an independent laboratory could be helpful in decision making and in potentially sparing patients from therapy-associated side effects. By contrast, ODDS with 0.4% inconclusive rates might be more appropriate in mainstream assessment, since the need for independent reassessment can be reduced.

3.3. Specific Differences in Prediction between ODDS and DST

As noted above, definite predictions were never seen contradicting between ODDS and DST. However, decisions deemed definite in ODDS were rendered inconclusive by DST and vice versa. This becomes evident by contrasting predictions *coded by location* versus predictions *coded by color* in Figure 6:

- Within the plane of ODDS scores (left panel), the L-shaped area (colored faint red) denotes samples definitely predicted positive by ODDS (according to location). However, some of them are inconclusive according to DST (colored beige); in fact, 40 DST-inconclusive samples invade the positive, and the other 45, the negative domain of ODDS scores, see Table 1.
- Conversely, the uncolored L-shaped area accommodates samples predicted inconclusive according to ODDS (according to location). However, 10 are colored red, i.e., according to DST, decided positive. In fact, these samples, definitely predicted positive by DST, invade the inconclusive region of ODDS scores and are labelled 'adding information', see Figure 6, panel (c) and Table 1.
- Within the ternary plot of DST evidence (right panel), the triangular shaped areas denote samples predicted negative (faint blue) and positive (faint red), respectively, according to DST (by location). However, some samples are color-coded beige, i.e., they were rendered inconclusive by ODDS. Note that the very same samples appear in dual roles along ODDS scores and ternary evidence, respectively (left and right panel).
- Conversely, the uncolored kite-shaped area denotes samples predicted inconclusive according to DST (by location). However, some of them are color-coded red or blue, i.e., definitely predicted as positive or negative according to ODDS. In fact, 40 samples definitely classified positive through ODDS intrude into the 'inconclusive' region of DST and have been labelled as 'increasing safety', see panel (d). Another 45 definitely predicted negative through ODDS intrude into the 'inconclusive' region of DST.

All in all, differences in prediction only occur with samples on the brink of predictability. While one method yields positive or negative, the other may yield 'inconclusive'. These differences turn up in the off-diagonal elements of Table 1, which are small; see also the percentages. Even if differences are small, they are important for the single patient and seen at the core of personalized medicine.

Moreover, visual inspection of the ternary plot in Figure 6 reveals samples not being evenly distributed over the triangular plane of evidence. Samples, rather, appear in groups, arranged in lines or lengthy polygons. The mechanisms behind the scenes, giving rise to these effects, are scrutinized in Appendices A.6 and A.7.

4. Discussion

Dempster–Shafer Decision Theory (DST) has been made available for the personalized therapy of breast cancer in a previous paper [37], in particular, to increase the precision of receptor status assessment. Unfortunately, we could not map with ground truth in our papers, since ground truth is not available for the data used. However, we were able to provide a sound comparison between ODDS and DST and pinpoint particular

differences in performance. To underpin the usefulness of DST, we have scrutinized the survival of patients with status corrected from positive or negative predictions by ODDS towards 'inconclusive' by DST, see Figure 7. Since only a small fraction of patients was to be 'corrected' (see Table 1), survival curves degenerate and were included only for completeness. Even if this percentage is small, it seems mandatory, considering the large number of breast cancer patients. In practice, patents rendered inconclusive should receive lab reassessment, in order to reduce false estimates and increase precision.

In addition, we compared patients considered receptor positive according to up-to-date clinical standards (IHC+, red curve) with those considered positive (pos/pos, light blue curve) according to both of our proposed methods, ODDS and DST. Patients positive according to the new methods experienced significantly better survival (log-rank $p = 0.03$) than those conventionally diagnosed positive, see the red versus the light blue curves in Figure 7.

Comparing ODDS and DST, DST was found to be somewhat more conservative than ODDS. Vice versa, patients considered 'positive' by DST, while being considered 'undecided' by ODDS, may benefit from this additional information inferred by DST. However, this gain of information has two sides: the 'positive' prediction might not really hold in the end, and relying on it may cause harm. Hence, re-evaluation remains the only safe advice in these cases.

4.1. Advantages of Evidence Compared to Probabilities in Conventional Statistics

In addition to our previous work, the implementation of DST is, here, unfolded in three steps:

First, we demonstrate the simplest case, starting with a single gene (the receptor gene) and demonstrate how to:

- Obtain DST evidence from gene expression.
- Obtain DST evidence from IHC.
- Fuse both items of evidence above, via the Yager evidence combination rule [78].
- Display results in a 'ternary' plot, a genuine format for presenting evidence.
- Show subgroups of patients with given IHC status, giving rise to specific patterns of samples in evidence space.

In a second step, we demonstrate how to create evidence from co-genes and join them with evidence of receptor genes and IHC (by Dempster and Yager Evidence Combination rule, respectively).

In the third step we demonstrate how to join evidence from estrogen with that from progesterone, using a formula imitating the clinical criterion 'positive ER or positive PGR' for 'receptor positivity', in terms of Dempster–Shafer mathematics.

This stepwise approach allows for a detailed introduction into ternary plots, demonstrating their applicability to clinical decision making, based on evidence. It becomes clear that evidence not only provides more information about the outcome of a measurement than conventional probability does, but that probabilities are supplemented by uncertainty. Evidence also has the property of three numbers summing up to unity for each single sample considered, and may be advantageously displayed in ternaries. Groups of patients (different IHC statuses) are segregated by the method itself (being either positive or negative).

Data quality is a crucial aspect of personalized medicine. In this work, we have never let gene expression overrule IHC. Technically, this was achieved by selecting the constants $\hat{\alpha}$ in our model very conservatively. As a consequence, positive IHC estimates were never converted into negative, not even a positive progesterone when estrogen was negative IHC = $(-, +)$. Such IHC estimates occurred in 15 samples, and gene expression by itself would turn them into $(-, -)$, if we had modeled less weight into IHC and more into gene expression.

4.2. How Uncertainty May Help Increase Correctness (Precision)

At first glance, this statement may seem paradoxical. However, DST—in comparison to ODDS—supports this concept, as can be seen from a vivid comparison:

Suppose we have a ballot between two options (pro, contra). If the voter turnout was 100%, we might obtain 75% for pro and 25% for contra (3:1), and with full right, consider this a clear decision. The option 'pro' would clearly be implemented, having the majority of voters on its side, see the top bar in Figure 8. Exactly this scenario corresponds to classical statistics, considering a probability p and the probability $1 - p$ for its opposite.
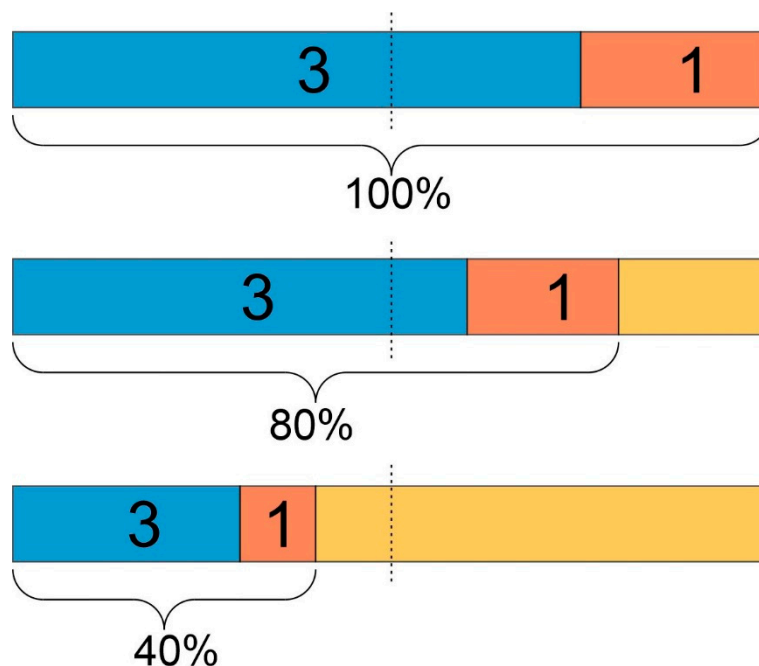
Figure 8. Uncertainty puts probabilities into perspective. Smaller voter turnout (e.g., 80%, 40%) in elections compares to increased uncertainty in DST. Dashed lines indicate 50%. From the very same ratio of votes for pro and contra (3:1 in each scenario), different consequences may be drawn in the light of high or low voter turnout, respectively. Likewise, probabilities of diagnoses may only be considered reliable if uncertainty, according to DST, is below some threshold.

Now suppose the voter turnout was only 80%, with exactly the same distribution between pro and contra, i.e., 60%:20% = 3:1, see the second bar in Figure 8. In this case also, we would consider it a valid decision, despite 20% non-voters, representing what is termed 'uncertainty' in DST. However, the result would not be considered as 'robust' as in the first case.

Finally, suppose a voter turnout of just 40%, again with the same ratio between pro and contra of 30%:10% = 3:1, see the third bar in Figure 8. Such a result would not be considered sound enough to draw conclusions from. An uncertainty of 60%—in terms of DST—may render the result 'un-trustable', even with a large ratio of probabilities $p : (1 - p) = 3 : 1$. After all, the relative 'majority' of 30% is far from absolute (50%). In such a case, a wise politician would not be confident to implement option 'pro', since opposition might emerge that is too strong to overcome.

Analog concepts hold for medical diagnostics. As DST introduces uncertainty as the third part of evidence [82,83], borderline or questionable results obtained by classical statistics may be relegated 'uncertain', suggesting further assessment and, thereby, increasing final correctness. In addition, significantly different risks may be inferred by falsepositive as compared to falsenegative decisions. For example, a falsepositive receptor status may lead to the avoidance of chemotherapy, in this case, the lifesaving therapy. Accordingly, one might request very low uncertainty, in order to 'take a positive status serious', regarding

therapeutic consequences. Conversely, a false-negative status might 'just' entail unnecessary chemo, a comparatively lower risk. All in all, it is but a clinical decision how much uncertainty seems acceptable.

To allow for evidence-based decisions, the explicit quantification of uncertainty seems utmost desirable.

4.3. Extensions of Decision Rules

The approach presented here may be expanded by considering more than one co-gene, since DST allows us to combine more than two items of evidence. Considerable increases in stability can be expected if such expanded markers are applied to new incoming data.

Another possible extension refers to combination rules.

One basic concept for combining evidence from different sources was introduced by Dubois [84], hence, termed "evidence combination rule (ECR) after Dubois and Prade". In the case of just two outcomes, this boils down to the Yager rule [78]. Smarandache [85] further generalized combination rules and defined the PCR5 combination rule, relevant for three (or more) outcomes. Fontani [66] proposed fusing the spaces of events in image processing and Denœux introduced weighted combination [69,74,75]. Chen defined distances between evidence [86]. Yang reviewed a framework of evidence combination rules and evidence weighting and discounting [65] and Sentz compiled all rules, in a comprehensive overview [87].

In the present work, we only used the Dempster Evidence Combination Rule (ECR) and the Yager rule [78]. However, this is not mandatory. In fact, a variety of ECRs exist, which differ in behavior in certain situations.

4.4. Modelling Sharp and Soft Clinical Decisions

The Dempster Evidence Combination rule advocates fierce decisions—leaving little uncertainty in the conclusions—even if both pieces of input evidence concede considerable uncertainty. As opposed, given the same input evidence, the Yager rule follows a much softer strategy, transmitting larger uncertainty into its conclusion. We illustrate this by a specific example.

Suppose we have two items of evidence for receptor status: The first piece of evidence, from gene expression ($\alpha = 0.8$, $\beta = 0.1$, $\theta = 0.1$), strongly favors 'positive', via large α and small β. Moreover, it claims to be quite 'sure' in terms of small θ. The second piece of evidence, from IHC ($\alpha = 0$, $\beta = 0.7$, $\theta = 0.3$), favors 'negative', with some larger uncertainty $\theta = 0.3$. Obviously, these pieces of evidence contradict each other quite strongly, and one may legitimately ask 'what should be the synthesis of these two?' The answer can be precisely modelled by decision combination rules, according to Dempster ($\oplus_D$) or Yager ($\oplus_Y$), which also exemplifies their difference in approach.

For the current example, the Dempster rule (Equations (7) and (8)) yields as combined evidence $E_D = (\alpha = 0.54, \beta = 0.38, \theta = 0.068)$, expressing a quite 'sharp' contradiction (large α, large β), without admitting much uncertainty (small θ). On the contrary, the Yager rule, Equation (6), yields combined evidence $E_Y = (\alpha = 0.24, \beta = 0.17, \theta = 0.59)$, expressing only 'soft' contradiction (small α, small β), with quite a lot of uncertainty (large θ).

How can these features be exploited for personalized medicine?

Clinical experts have always been looking for the most beneficial balance in decision making, based on SOPs, their personal experience, and also skill, or even educated guessing, in particularly difficult cases. It has always been the strength and fame of top clinicians to decide correctly in a percentage of cases far above average. However, it may not be fully transparent how such an outstanding clinical performance comes about and could be transferred to young doctors in training. Decision theory tries to bring such 'clinical expert competence' down to more formally applicable rules. Of course, it will remain the task of top clinicians to help define and select those rules, based on sound statistical evaluations of clinical studies. Such decision rules, once established, may be incorporated in SOPs and will improve their performance significantly.

While this work exemplifies the use of DST in personalized medicine, related to the very specific field of breast cancer receptor diagnostics, the methods described are universal. Decision theory, in particular, the fusion of diverging evidence (sometimes also called 'sensor-fusion'), as well as the professional incorporation of uncertainty into biomarker research, seem valuable for all fields of personalized medicine and medicine in general.

Appendix A

Appendix A.1. Download and Cleansing of Data

The Gene Expression Omnibus (GEO) [88,89] was screened for breast cancer studies using the Affymetrix chip U133A+2.0 [90] and found 38 studies. CEL files and clinical data (characteristics), such as estrogen and progesterone receptor statuses (ER, PGR) and HER2, all measured by immunohistochemistry (IHC) were downloaded and curated to arrive at a clean database [91,92], already described in our previous work [37]. Data cleansing meant, in particular:

- Only tumor samples were considered, controls excluded;
- Only tissue samples were considered, cell lines excluded;
- Replicates were removed;
- All samples were pairwise checked for being duplicates. CEL files with equal medical data (expression, clinical) may differ, just in format or container packing. Hence, actual expression values needed to be compared to safely locate duplicates;
- If duplicates in expression data were found to differ in metadata, these were curated manually;
- Some GSE studies have been 'enriched' with samples from other (previous) GSE studies. Such samples become duplicates if both of these studies were evaluated in combination. We always left such samples with their original study and removed its duplicate from the later GSE study;
- We detected damaged samples by RMAexpress [93] and removed them;

    After cleansing, 3753 samples remained to be used for joint evaluation [94].

A plethora of normalization methods for microarrays has been proposed [51,95–98], as evaluated by Bolstad [57]. Based on the results of our previous work [99], we performed RMA, using the implementation MATLAB affyrma. We had also evaluated several types of batch corrections [100–103] (Luo et al., 2010, Leek et al., 2012, Müller et al., 2016, Johnson et al., 2007) and surrogate variable analysis (SVA) [104–106] (Leek and Storey, 2007, Leek et al., 2017, Parker et al., 2014) for combined microarray studies. However, no clear benefit of explicit batch corrections could be clearly demonstrated [99]. Therefore, we preferred to perform 'global RMA normalization', over all studies being combined.

Appendix A.2. Selecting HER2 Negative Patients

Next, HER2-negative patients were selected as follows. Patients with positive IHC estimates for HER2 ($HER2^+_{IHC}$) were excluded right away and samples with negative HER2 ($HER2^-_{IHC}$) retained. For patients with missing IHC estimates for HER2, we attempted imputation via gene expression, using the ODDS method [37] (see also Table A3). Only 'safe' imputations, yielding $HER2^-_{IHC}$, were retained, in accordance with our previous work [37], yielding 2519 patients in all. The final set of studies considered is listed in Table A1 of Appendix A.1.

Appendix A.3. Selecting Genes and Probe Sets for Estrogen and Progesterone Receptors

Several genes are mentioned in the literature [22,107] to be relevant for estrogen and progesterone. In addition, we used the limma-package [55] to screen CEL files (of those 2519 HER2-negative patients) for probe sets, discriminating between positive and negative IHC statuses, separately for ER and PGR, and sorted results by ascending *p*-values. After mapping back from probe sets to genes, we finally adopted the very receptor gene and one co-gene, in addition, for ER and PGR, respectively. For details, please refer to our previous work [58], and Table A2.

Table A1. Microarray studies used. *N*: number of samples. All: total number of samples in study. ER_{IHC}: number of samples with IHC measurement of estrogen receptor status. PGR_{IHC}: likewise for progesterone. For reference see also our previous work [37,59,99].

		N		
Study	**City**	**All**	**ER_{IHC}**	**PGR_{IHC}**
GSE5460	Boston	17	17	0
GSE6532	Toronto	78	78	77
GSE12276	Rotterdam	118	0	0
GSE16391	Toronto	50	50	50
GSE16446	Toronto	84	84	0
GSE18728	Seattle	15	15	14
GSE18864	Lyngby	60	60	60
GSE19615	Manhattan	79	79	79
GSE20685	Taipei	163	0	0
GSE20711	Toronto	52	52	0
GSE22035	SAINT-CLOUD	27	27	0
GSE23177	Leuven	80	80	0
GSE26639	Paris Cedex 05	144	144	142
GSE27120	Brussels	26	26	26
GSE29431	Barcelona	23	23	23
GSE31448	Marseille	286	286	271
GSE36771	Auckland	86	86	85
GSE42568	Dublin	65	63	0
GSE43358	Brussels	43	43	43
GSE43365	Boston	98	98	98
GSE46222	Washington	26	26	0
GSE47389	Rotterdam	47	47	47
GSE48390	New Taipei City	34	34	0
GSE48905	Hørsholm	20	20	0
GSE50567	Gliwice	25	25	0
GSE58792	New York	33	33	0
GSE58812	Saint Herblain	107	107	107
GSE61304	Singapore	41	38	33
GSE65194	PARIS	70	67	41
GSE71258	Missouri	94	77	77
GSE76124	Houston	198	198	198
GSE76274	Houston	44	44	44
GSE87007	Brussels	24	24	24
GSE88770	Brussels	108	108	107
GSE95700	Taipei City	54	54	54
Σ		**2519**	**2213**	**1700**

Table A2. Probe sets and logistic regression for receptor genes and co-genes. c_0 and c_1 are coefficients from logistic regression for IHC values (dependent variable) versus gene expression (independent variable) [37,58]. Probe sets refer to the Affymetrix chip U133A + 2.0. For 'deviance of fit', see p. 118 in McCullagh [108]. Upper limits for beliefs ($\hat{\alpha}$, $\hat{\beta}$) are explained in Appendices A.4 and A.5. See Equations (A1) and (A2) for computing the limits, and Equation (4) for their application.

				Logistic Regression Parameters		Logistic Regression Quality		Upper Limits for Beliefs	
			Probe Set	c_0	c_1	Deviance of Fit	Number of Samples	$\hat{\alpha}$	$\hat{\beta}$
estrogen	gene	ESR1	205225_at	9.905	−1.061	1086.6	2213	0.814	0.887
	co-gene	AGR3	228241_at	5.582	−0.710	1253.1		0.794	0.840
progesterone	gene	PGR	208305_at	7.449	−0.983	1107.4	1700	0.753	0.702
	co-gene	ESR1	205225_at	8.617	−0.834	1249.8		0.618	0.817

Table A3. Probe sets and logistic regression for gene and co-gene of HER2. c_0 and c_1 are coefficients from logistic regression for IHC (dependent variable) versus gene expression (independent variable) [37,58]. Probe sets refer to the Affymetrix chip U133A + 2.0. For 'deviance of fit', see p. 118 in McCullagh [108].

				Logistic Regression Parameters		Logistic Regression Quality	
			Probe Set	β_0^{GE}	β_1^{GE}	Deviance of Fit	Number of Samples
HER2	gene	ERBB2	216836_s_at	15.963	−1.408	1421.2	2430
	co-gene	PGAP3	221811_at	17.756	−2.168	1330.6	
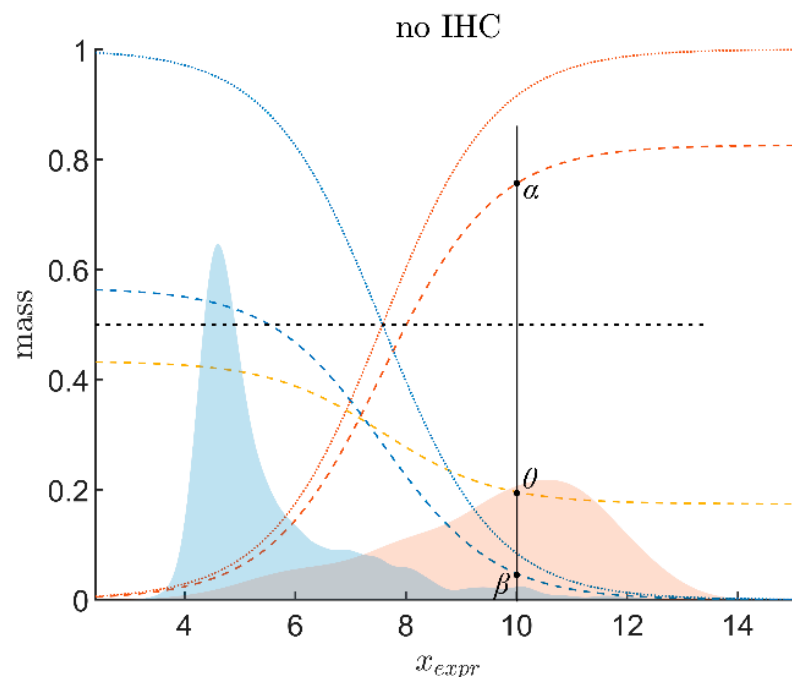
Figure A1. Logistic regression to obtain responsibility functions for progesterone, gene PGR. Distribution of gene expression for positive receptor (according to IHC) computed from density kernel estimates [71–73], shown red shaded, for negative IHC, blue shaded. Responsibility functions for receptor positivity, r_+ (dotted red curve) and r_- (dotted blue) were obtained in this way. Belief in positive (α): solid red, belief in negative (β): solid blue and uncertainty (θ): solid beige.

Appendix A.4. Tailoring Beliefs in Receptor Gene Expression to a Given Accuracy of IHC

It is intuitively understandable that such an upper limit for the belief in 'positive' must relate to true and false positive rate (*TP*, *FP*), as well as true and false negative rate (*TN*, *FN*) of the measuring process in question. It was one of the main achievements in our previous work [37], to coin this qualitative argument into the following Equation:

$$\hat{\alpha}_{\text{Expr}} = \frac{TP \cdot TN - FP \cdot FN}{(TP + FP) \cdot (TN + FP)} \tag{A1}$$

$$\hat{\beta}_{\text{Expr}} = \frac{TP \cdot TN - FP \cdot FN}{(TN + FN) \cdot (TP + FN)} \tag{A2}$$

$\hat{\alpha}_{\text{Expr}}$ and $\hat{\beta}_{\text{Expr}}$ quantify the remaining doubt, even if measurements seem perfectly clear (maximum gene expression). *TP*, *FP*, *TN* and *FN* can be obtained from the discrepancies between IHC and the prognosis obtained from gene expression, according to conventional statistics, using a cut-point of 0.5 in the logistic regression.

Appendix A.5. Formulating IHC Data in Terms of Evidence

Gene expression, x_{Expr}, is a continuous variable, and so is evidence derived thereof: $\alpha_{\text{Expr}}(x_{\text{Expr}}), \beta_{\text{Expr}}(x_{\text{Expr}}), \theta_{\text{Expr}}(x_{\text{Expr}})$, as shown in Figure 2. Opposed to that, IHC yields binary results (+/−) and, hence, evidence thereof are constants, one set for a positive IHC result ($\alpha_{\text{IHC}^+}\ \beta_{\text{IHC}^+} = 0\ \theta_{\text{IHC}^+}$) and a second set for a negative IHC result ($\alpha_{\text{IHC}^-}, \beta_{\text{IHC}^-}, \theta_{\text{IHC}^-}$). How shall these values be chosen?

For a start, we draw on the following findings: Quality assessments of IHC [38,39] revealed that approximately 85% of IHC estimates can be assumed to be correct and, consequently, 15% to be false [40–42].

To implement these findings in terms of DST, we first consider all *IHC measurements with positive outcome*, as illustrated in Figure A2, upper panel. Among these, some have resulted true positive, by quality of the measuring method, others resulted true positive by chance. Both taken together make up the (total) number of true positives (*TP*), i.e., 85% of all positive outcomes, according to the above data from the literature. The remaining 15% of positive IHC outcomes represent wrong results, namely false positives (*FP*), i.e., samples negative in reality. We may now assume (on good grounds) that 15% is also a reasonable estimate for the fraction of samples being true positive by chance, not by quality of the method, see Figure A2, upper panel.

Accordingly, given a positive IHC measurement (IHC⁺), the total evidence comes about as follows:

- Due to the positive IHC measurement, there is no evidence at all for the status being (truly) negative due to quality of the method, hence $\beta_{IHC^+} = 0$.
- Being measured as true positive by chance or as false positive by error represents all measurements not being true by quality of the method. Together they make up 30%, represented by $\theta_{IHC^+} = 0.3$. We assume that these split in equal parts into 15% true positive by chance and 15% false positive by error.
- Hence, cases being true by quality make up the remaining 70%, represented by $\alpha_{IHC^+} = 0.7$.
- Since all items add up to 1 (Equation (2)), we obtain $\beta_{\text{IHC}^+} = 0$, and the whole evidence after a positive IHC result is ($\alpha_{\text{IHC}^+} = 0.7, \beta_{\text{IHC}^+} = 0, \theta_{\text{IHC}^+} = 0.3$).

On the contrary, after a *negative* IHC measurement, IHC⁻, we obtain the evidence: $\alpha_{\text{IHC}^-} = 0.0, \beta_{\text{IHC}^-} = 0.7$ and $\theta_{\text{IHC}^-} = 0.3$, see panel (b) of Figure A2.

(**a**)

(**b**)
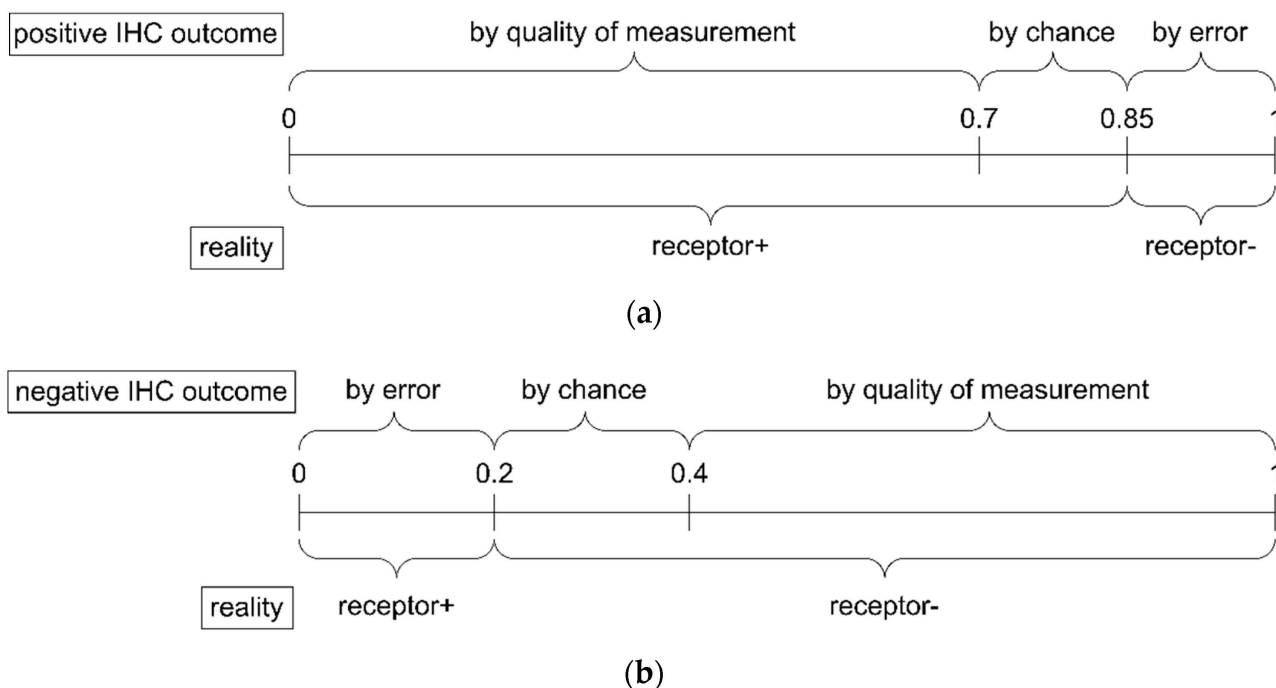
Figure A2. Results of measurements versus reality, seen along the concepts of Dempter–Shafer Theory. In panel (**a**) we focus on positive results only, yielded by IHC measurements (positive results represent 100%), see upper labels. Out of these, 70% are true positives (see lower labels) and result due to quality of the measuring technique. Accordingly, the belief in positive $\alpha = 0.7$. Another 15% have come out as true positives by chance. The remaining 15% of positive outcomes are due to error, i.e., they are false positives, being truly negative. Both together represent the uncertainty ($\theta_{\mathrm{IHC^+}} = 0.3$) of being receptor positive in reality. In panel (**b**) we focus on negative measurement outcomes only: 60% of these come about due to the quality of measurement, 20% were correctly negative by chance and another 20% false negatives, since they are positive in reality.

Appendix A.6. Ternary Plots Reflect Subgroups within Patient Cohort

After introducing the more general features of ternary plots in Section 2.2.4, we now describe specific features of actual patient data of this study within this framework, see also Figure 4. Considering *just one gene plus IHC* as evidence, it is easy to make subgroups of patients transparent, a possibly valuable feature for personalized medicine, illustrated by the following features:

a. Evidence for patients is not distributed evenly all over the 'triangle plain of evidence', but samples are grouped in 'traces', which deserves explanation: first, we note that exactly *three* lines appear and each sample belongs to one of these lines; no sample is found apart. The fact that we deal with three possible states of IHC values (+, −, inc) already points towards a possible reason, and this is in fact true: it is varying IHC statuses, which give rise to these lines. Suppose that, for a given IHC status, e.g., positive, we consider different values of gene expression, x_{Expr}. When computing corresponding evidence, $\alpha(x_{\mathrm{Expr}}), \beta(x_{\mathrm{Expr}}), \theta(x_{\mathrm{Expr}})$, these will appear along a straight line. This is visually obvious but can, in fact, be formally proven mathematically, resorting to Equations (1), (4) and (6). Hence, each of the specific lines may be labeled, accordingly ($\mathrm{ER}_{\mathrm{IHC}}^{+}$, $\mathrm{ER}_{\mathrm{IHC}}^{-}$ and $\mathrm{ER}_{\mathrm{IHC}}^{\mathrm{inc}}$), see Figure 4a.

b. Note also that the red line of $\mathrm{ER}_{\mathrm{IHC}}^{+}$ samples starts near the corner $\alpha = 1$, but *not exactly at* the corner: even a positive IHC and large gene expression cannot guarantee a positive prediction—some small uncertainty (θ) remains. At the same time, for such a sample, there is no evidence whatsoever for a negative status. Hence $\beta = 0$, and the

line starts at the ternary plot's side representing $\beta = 0$. Such a sample represents the total opposite to the lower left corner—where $\beta = 1$ (surely negative).

c. After originating close to the lower right corner of (marked with $\alpha = 1$) the line for ER_{IHC}^{+} (red), proceeds across the sub-area indicating receptor positive (shaded red). These samples have ER_{IHC}^{+} status (all dots, no circles), being confirmed by gene expression, ending up as positive predictions. After crossing the decision border at $\alpha = 0.5$, this line still represents samples with ER_{IHC}^{+}, which has obviously been questioned by gene expression; hence, prediction was rendered 'inconclusive' according to DST (samples lie within the kite-shaped area). Coloring these samples, according to ODDS, most vividly reveals differences in prediction: although located within the DST-inconclusive region, ODDS predicts some of these samples as positive, the majority as inconclusive (i.e., agrees with DST), but a few as negative (see the blue dots towards the end of the line in the upper left).

d. Note that lines for ER_{IHC}^{+} and ER_{IHC}^{-} never protrude into the opposing definite areas, for the following reason: given ER_{IHC}^{+}, gene expression can by no means reverse the prediction to surely negative. At the most, it may downgrade it to inconclusive. The same is true for ER_{IHC}^{-}. The white, kite-shaped area segregates the areas of positive and negative predictions, which is reasonable.

e. Only at one single point, two strongly opposing items of evidence might, in principle, become close to one another (at the point $\alpha = \beta = 0.5$, along the baseline of the ternary plot, see the tutorial Section 2.2.4 for further discussion). As a matter of fact, such samples do not occur in reality (in our cohort), and both lines meet farther outside, within the inconclusive region. In other words, if evidence incorporates contradiction, DST renders them inconclusive—as a precaution.

f. Finally, the line for ER_{IHC}^{inc} crosses the whole decision triangle, from surely positive (right side) through the inconclusive region (mid), towards surely negative (left side). Since no IHC status is available for these samples (shown as circles), gene expression is free to render this ample range of predictions.

The characteristics of ternary plots, enhanced data interpretation and its relevance for personalized medicine, have been introduced along a simple example—featuring only IHC status and the expression of one single gene—in order to be intuitively clear. In the following, the 'full' model (including co-genes), will be evaluated along the very same conceptual lines.

Appendix A.7. Evidence Patterns for Subsets of Patients

We have already demonstrated (Section 2.2.4 and Figure 4) for a single gene and IHC (as the only sources of evidence) that conspicuous arrangements of data points are rooted in the IHC status: for all patients with a given IHC status, evidence was seen to lie on straight lines, see Figure 4. Now, considering four genes (two receptor genes, two co-genes), the situation becomes more complex. More degrees of freedom in the input variables penetrate into the final prediction, and the lines (as seen for single genes) expand to lengthy polygons.

To scrutinize the underlying mechanism, we first display data separately for distinctive IHC statuses, e.g., for IHC = $(-, -)$ see Figure A3. Again, we display samples in ODDS coordinates (left column) side by side with DST ternary coordinates (right column). In the left panels, the locations of samples indicate their prediction according to ODDS, while their color indicates their prediction according to DST, and vice versa. Again, differences in prediction are read off easily. Note that the very same, specific subset of samples (IHC = $(-, -)$) is shown in all panels of Figure A3.

The following questions then arise: Do absolute, distinct boundaries exist for the evidence of samples with given IHC statuses? If yes, where are they located? To find out, we computed so-called 'maximum accessible prediction domains' (MPDs) as follows: artificial (simulated) samples were generated by scanning each gene and co-gene over the entire domain of measured expression values (in our data, 2.3 to 15.2) in 100 equidistant points, yielding $100^3 = 10^6$ generated samples. For each generated sample, we computed

predictions by both, ODDS and DST, and plotted them into the ODDS plane and the ternary triangle, respectively. These predictions spread out over much larger areas than the samples of actual patients did, and we, hence, termed these areas 'maximum accessible prediction domains' (MPD). Rather than showing all samples together, we displayed them separately for each prediction (negative, inconclusive or positive); see the rows 'negative', 'inconclusive' and 'positive' in Figure A3. Using 10,000 simulated samples, MPDs would be scrammed with points when being plotted. We do not display all of them (would look like filled areas) but only show the outline of these areas. In each row of Figure A3, the left panel shows the MPD of DST, arranged in ODDS coordinates. Vice versa, the right panel shows the MPD of ODDS in DST ternary coordinates. Note the following:

- Since an MPD represents a maximum area, no sample of the same color appears outside, e.g., no blue sample (predicted negative by DST) may lie outside the blue MPD in the left panel of Figure A3.
- No blue sample (predicted negative by ODDS) may lie outside the blue MPD in the right panel of Figure A3.
- While predictions coded in color transgress decision borders according to location, they never leave the maximum accessible prediction domain of their own prediction method.
- Samples of real patients were never seen to yield contradicting predictions (e.g., negative by DSST and positive by ODDS), but MPDs well intrude into contradicting domains. For example, the negative MPD of DST (outlined blue) not only reaches into the inconclusive region, but well overlaps, even with the positive area of ODDS (Figure A3, left column, row 1). A second example is the positive MPD of ODDS (outlined red), penetrating into the decisively negative domain of DST (Figure A3, right column, row 3).
- Note that these 'contradicting' overlaps are rooted in extreme expression values, occurring in generated samples only, but have never been seen in our real data. Thus, these potential areas of contradiction between ODDS and DST remain a theoretical possibility to be considered, which does not infringe, however, application of these methods to data of real studies.
- Note that the dots (evidence) of these 10,000 simulated samples are not evenly distributed over the MPD. This is similar to the evidence of real samples; these also appear in fairly restricted zones, well within the respective MPD. One could generate 2-dimensional histograms, showing the density of these simulated samples.

Other IHC statuses are covered in Figure A4 (+, +), Figure A5 (+, −) and Figure A6 (0, 0). More cases are shown in the Appendix, see Figure A7 (−, 0) and Figure A8 (+, 0).
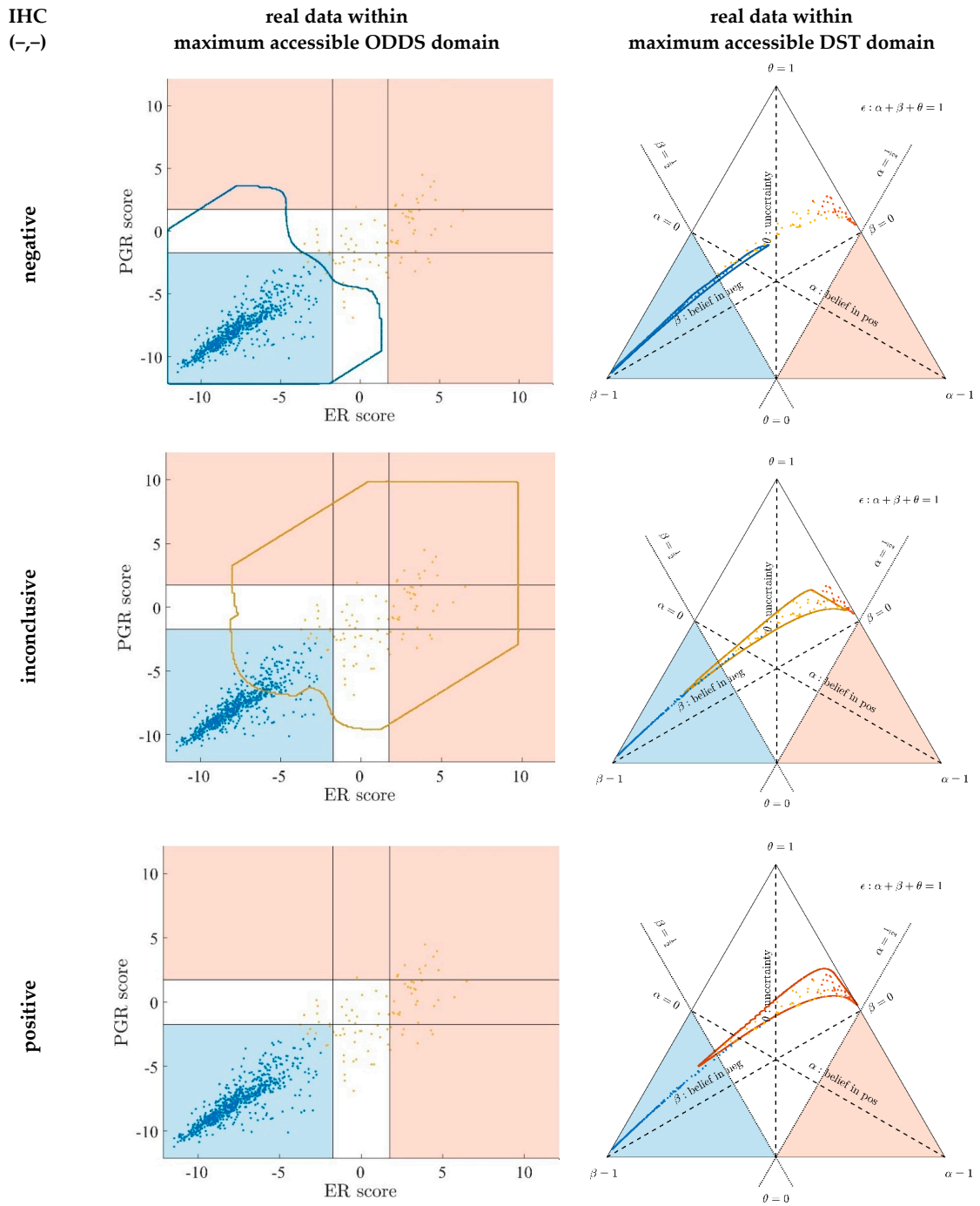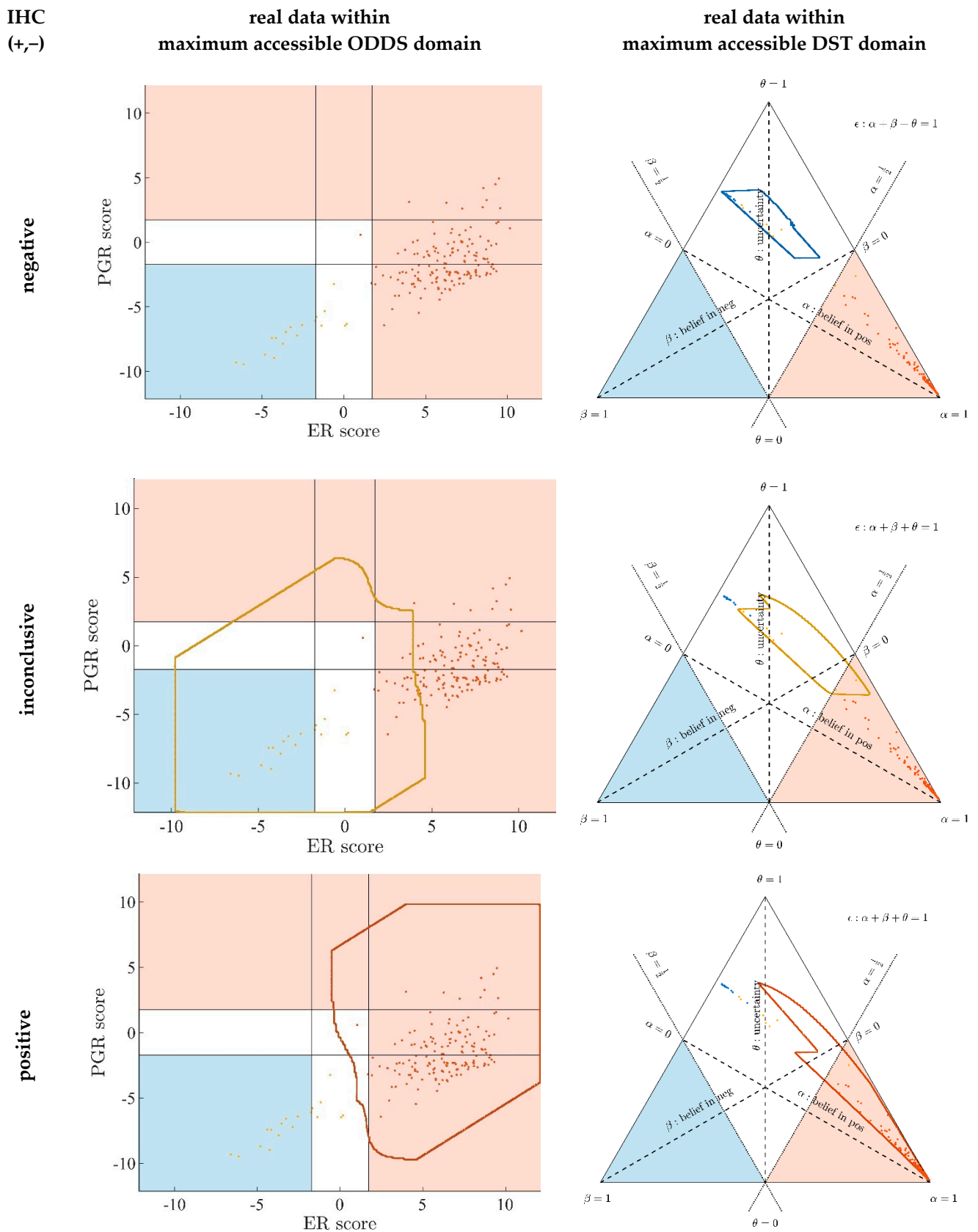
Figure A3. Real sample data within maximum accessible domains for IHC = $(-,-)$. **Left column:** Real data shown in coordinates of ODDS prediction scores. Square in light blue (bottom left): criterion for negative predictions according to ODDS. Area in light red: criterion for positive predictions according to ODDS. In between (white): area of inconclusive ODDS predictions. Outlined areas represent maximum domains accessible for DST predictions, displayed in ODDS coordinates. Sample data, *located* according to their ODDS prediction scores but *colored* according to DST prediction (neg $\hat{=}$ blue, inc $\hat{=}$ beige, pos $\hat{=}$ red). **Right column:** Real data shown in ternary coordinates of DST prediction evidence. Triangle in light blue (bottom left): criterion for negative predictions according to DST. Triangle in light red (bottom right): criterion for positive predictions according to DST. White, kite-shaped area: criterion for inconclusive DST predictions. Outlined areas represent maximum domains accessible for ODDS prediction, displayed in DST coordinates. Sample data, while being located according to DST evidence is colored, however, according to ODDS prediction. **Rows 1–3**: Negative, inconclusive and

positive predictions according to DST (left column) and ODDS (right column), respectively. **Purpose:** Differences between ODDS and DST predictions can easily be traced for real values as well as for maximum domains, e.g.: (1) Samples predicted negative by ODDS penetrate into the 'inconclusive' area (white kite) of DST (row 1, right panel). (2) The maximum domain for negative evidence by DST penetrates into the inconclusive (white) and also into the positive (light red) area of ODDS (row 1, left panel). For more extensive examples and extensive discussion, see text.
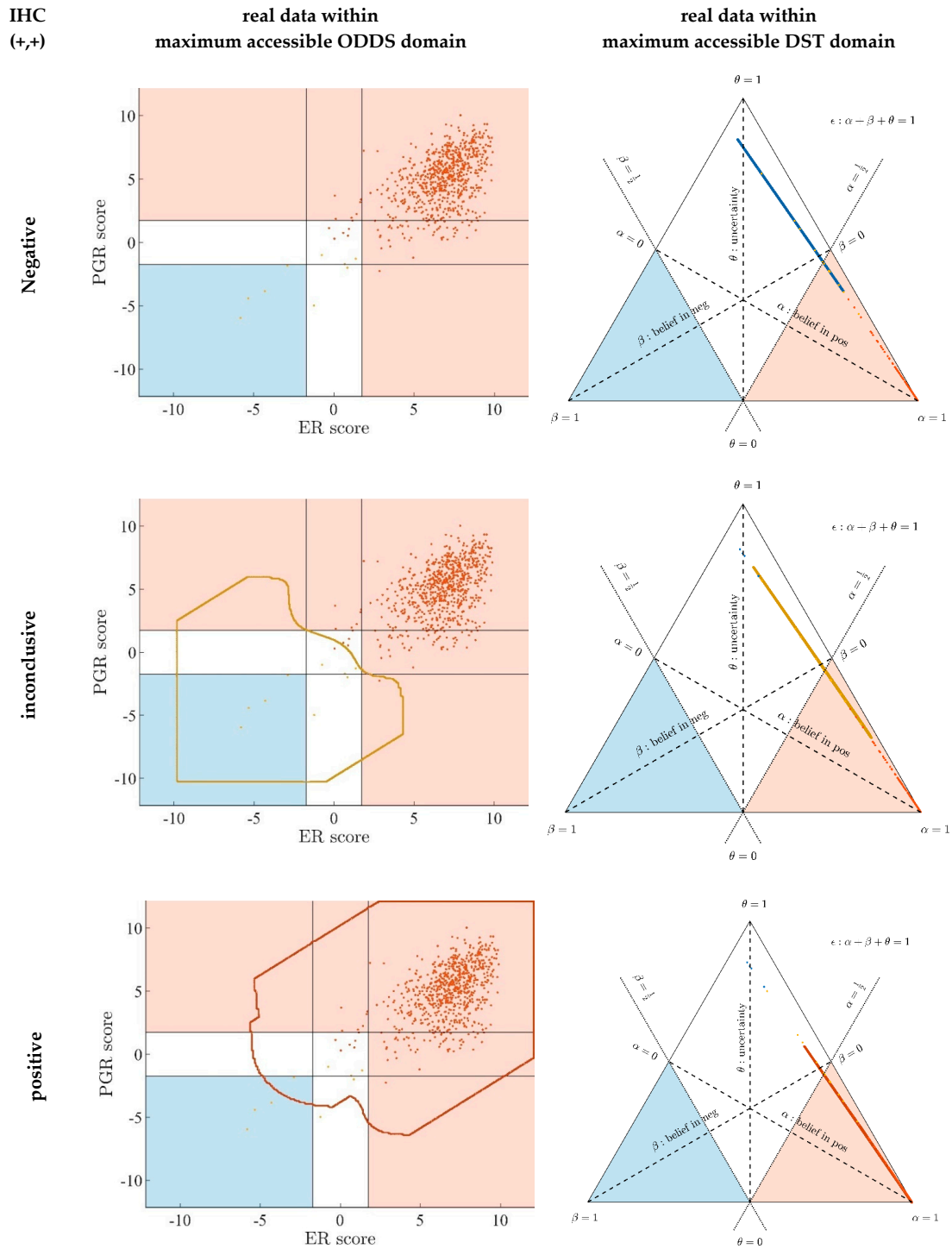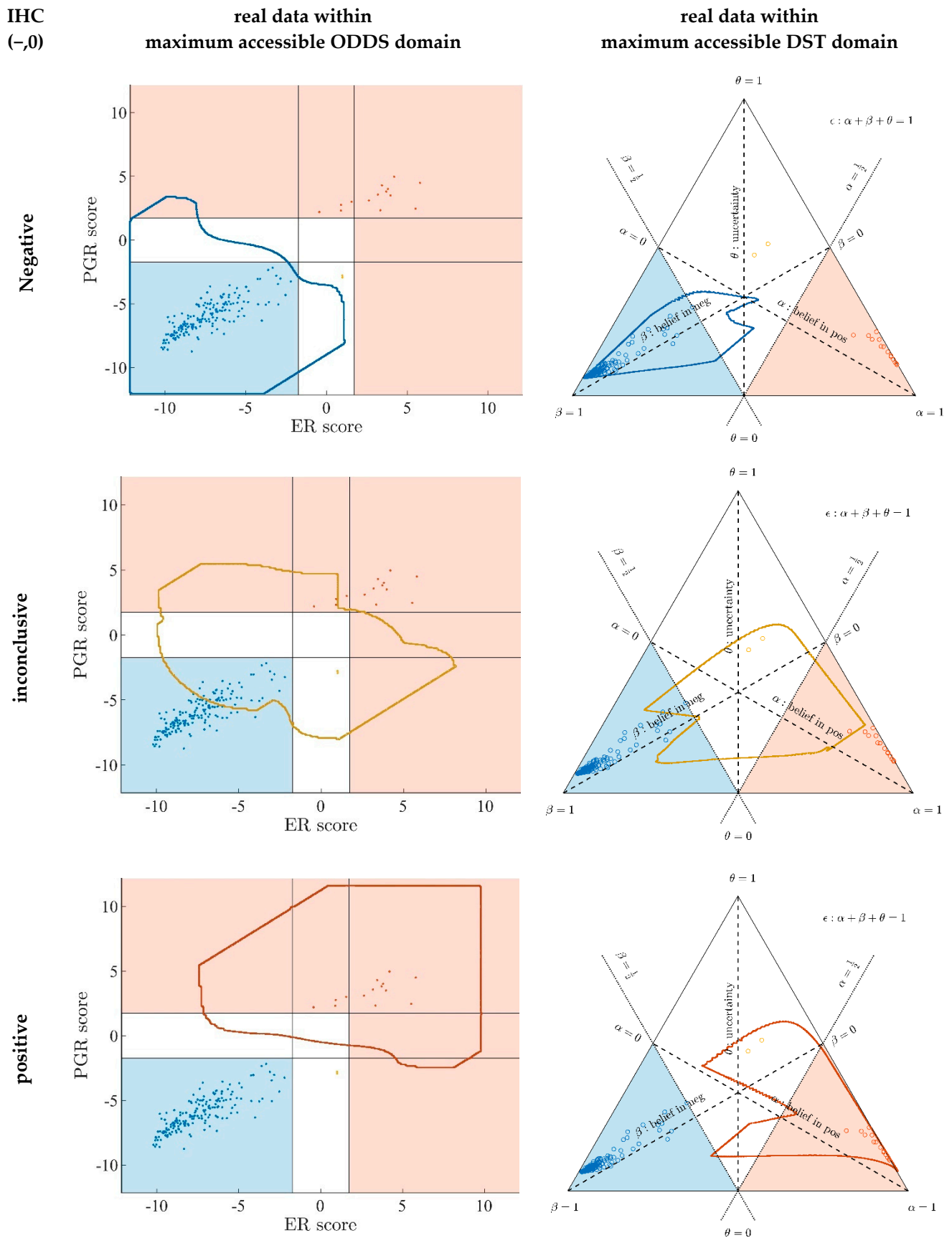

Figure A4. Real sample data within maximum accessible domains for IHC = (+,+). For general features of display see caption to Figure A3.

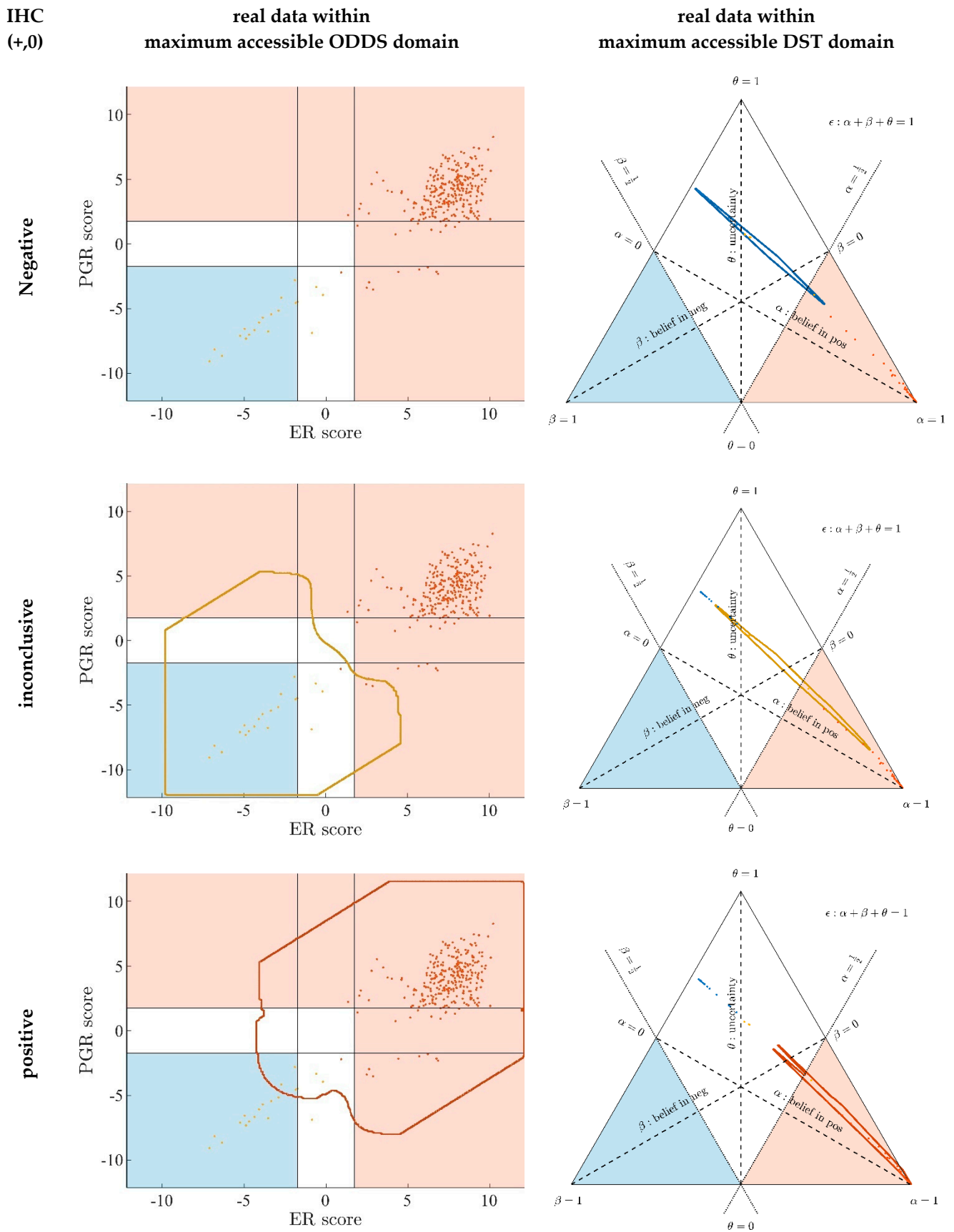| IHC (+,−)	**real data within maximum accessible ODDS domain**	**real data within maximum accessible DST domain**

Figure A5. Real sample data within maximum accessible domains for IHC = (+,-). For general features of display see caption to Figure A3.
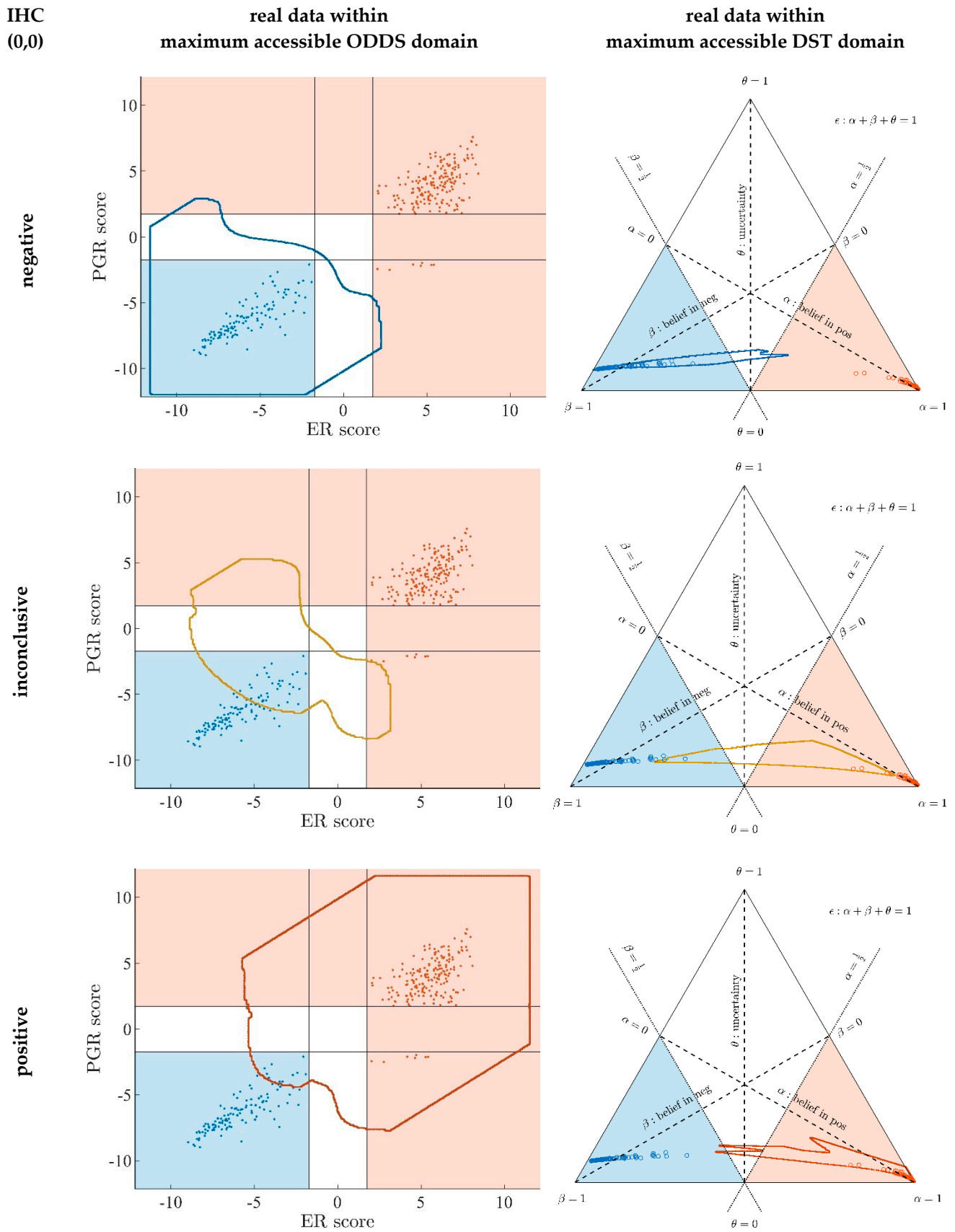
Figure A6. Real sample data within maximum accessible domains for totally unknown IHC = (0,0). For general features of display see caption to Figure A3.

Figure A7. Real sample data within maximum accessible domains for the IHC case = (−,0). For general features of display see caption to Figure A3.

Figure A8. Real sample data within maximum accessible domains for the IHC case = (+,0). For general features of display see caption to Figure A3.

References

1. Toss, A.; Cristofanilli, M. Molecular characterization and targeted therapeutic approaches in breast cancer. *Breast Cancer Res.* **2015**, *17*, 60. [CrossRef]
2. Dowsett, M.; Sestak, I.; Lopez-Knowles, E.; Sidhu, K.; Dunbier, A.K.; Cowens, J.W.; Ferree, S.; Storhoff, J.; Schaper, C.; Cuzick, J. Comparison of PAM50 risk of recurrence score with oncotype DX and IHC4 for predicting risk of distant recurrence after endocrine therapy. *J. Clin. Oncol.* **2013**, *31*, 2783–2790. [CrossRef]
3. Prat, A.; Ellis, M.J.; Perou, C.M. Practical implications of gene-expression-based assays for breast oncologists. *Nat. Rev. Clin. Oncol.* **2011**, *9*, 48–57. [CrossRef]
4. Huang, C.C.; Tu, S.H.; Lien, H.H.; Jeng, J.Y.; Huang, C.S.; Huang, C.J.; Lai, L.C.; Chuang, E.Y. Concurrent Gene Signatures for Han Chinese Breast Cancers. *PLoS ONE* **2013**, *8*, e76421. [CrossRef]
5. Desmedt, C.; Haibe-Kains, B.; Wirapati, P.; Buyse, M.; Larsimont, D.; Bontempi, G.; Delorenzi, M.; Piccart, M.; Sotiriou, C. Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes. *Clin. Cancer Res.* **2008**, *14*, 5158–5165. [CrossRef]
6. Kao, K.J.; Chang, K.M.; Hsu, H.C.; Huang, A.T. Correlation of microarray-based breast cancer molecular subtypes and clinical outcomes: Implications for treatment optimization. *BMC Cancer* **2011**, *11*, 143. [CrossRef]
7. Jezequel, P.; Loussouarn, D.; Guerin-Charbonnel, C.; Campion, L.; Vanier, A.; Gouraud, W.; Lasla, H.; Guette, C.; Valo, I.; Verriele, V.; et al. Gene-expression molecular subtyping of triple-negative breast cancer tumours: Importance of immune response. *Breast Cancer Res.* **2015**, *17*, 43. [CrossRef]
8. Burstein, M.D.; Tsimelzon, A.; Poage, G.M.; Covington, K.R.; Contreras, A.; Fuqua, S.A.; Savage, M.I.; Osborne, C.K.; Hilsenbeck, S.G.; Chang, J.C.; et al. Comprehensive Genomic Analysis Identifies Novel Subtypes and Targets of Triple-negative Breast Cancer. *Clin. Cancer Res.* **2015**, *21*, 1688–1698. [CrossRef]
9. Desmedt, C.; Giobbie-Hurder, A.; Neven, P.; Paridaens, R.; Christiaens, M.R.; Smeets, A.; Lallemand, F.; Haibe-Kains, B.; Viale, G.; Gelber, R.D.; et al. The Gene expression Grade Index: A potential predictor of relapse for endocrine-treated breast cancer patients in the BIG 1 Çô98 trial. *BMC Med. Genom.* **2009**, *2*, 40. [CrossRef]
10. Wu, G.; Stein, L. A network module-based method for identifying cancer prognostic signatures. *Genome Biol.* **2012**, *13*, R112. [CrossRef]
11. Liu, R.; Guo, C.X.; Zhou, H.H. Network-based approach to identify prognostic biomarkers for estrogen receptor–positive breast cancer treatment with tamoxifen. *Cancer Biol. Ther.* **2015**, *16*, 317–324. [CrossRef] [PubMed]
12. Korde, L.A.; Lusa, L.; McShane, L.; Lebowitz, P.F.; Lukes, L.; Camphausen, K.; Parker, J.S.; Swain, S.M.; Hunter, K.; Zujewski, J.A. Gene expression pathway analysis to predict response to neoadjuvant docetaxel and capecitabine for breast cancer. *Breast Cancer Res. Treat.* **2010**, *119*, 685–699. [CrossRef] [PubMed]
13. Clarke, C.; Madden, S.F.; Doolan, P.; Aherne, S.T.; Joyce, H.; O'Driscoll, L.; Gallagher, W.M.; Hennessy, B.T.; Moriarty, M.; Crown, J.; et al. Correlating transcriptional networks to breast cancer survival: A large-scale coexpression analysis. *Carcinogenesis* **2013**, *34*, 2300–2308. [CrossRef] [PubMed]
14. Aswad, L.; Yenamandra, S.P.; Ow, G.S.; Grinchuk, O.; Ivshina, A.V.; Kuznetsov, V.A. Genome and transcriptome delineation of two major oncogenic pathways governing invasive ductal breast cancer development. *Oncotarget* **2015**, *6*, 36652–36674. [CrossRef]
15. Filipits, M.; Rudas, M.; Jakesz, R.; Dubsky, P.; Fitzal, F.; Singer, C.F.; Dietze, O.; Greil, R.; Jelen, A.; Sevelda, P.; et al. A new molecular predictor of distant recurrence in ER-positive, HER2-negative breast cancer adds independent information to conventional clinical risk factors. *Clin. Cancer Res.* **2011**, *17*, 6012–6020. [CrossRef]
16. Zhao, X.; Rødland, E.A.; Sørlie, T.; Vollan, H.K.; Russnes, H.G.; Kristensen, V.N.; Lingjærde, O.C.; Børresen-Dale, A.-L. Systematic assessment of prognostic gene signatures for breast cancer shows distinct influence of time and ER status. *BMC Cancer* **2014**, *14*, 211. [CrossRef]
17. Van't Veer, L.J.; Dai, H.; van de Vijver, M.J.; He, Y.D.; Hart, A.A.; Mao, M.; Peterse, H.L.; van der Kooy, K.; Marton, M.J.; Witteveen, A.T. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **2002**, *415*, 530–536. [CrossRef]
18. Van de Vijver, M.J.; He, Y.D.; van't Veer, L.J.; Dai, H.; Hart, A.A.M.; Voskuil, D.W.; Schreiber, G.J.; Peterse, J.L.; Roberts, C.; Marton, M.J.; et al. A Gene-Expression Signature as a Predictor of Survival in Breast Cancer. *N. Engl. J. Med.* **2002**, *347*, 1999–2009. [CrossRef]
19. Paik, S.; Shak, S.; Tang, G.; Kim, C.; Baker, J.; Cronin, M.; Baehner, F.L.; Walker, M.G.; Watson, D.; Park, T. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N. Engl. J. Med.* **2004**, *351*, 2817–2826. [CrossRef]
20. Lu, X.; Lu, X.; Wang, Z.C.; Iglehart, J.D.; Zhang, X.; Richardson, A.L. Predicting features of breast cancer with gene expression patterns. *Breast Cancer Res. Treat.* **2008**, *108*, 191–201. [CrossRef]
21. Budczies, J.; Weichert, W.; Noske, A.; Muller, B.M.; Weller, C.; Wittenberger, T.; Hofmann, H.P.; Dietel, M.; Denkert, C.; Gekeler, V. Genome-wide gene expression profiling of formalin-fixed paraffin-embedded breast cancer core biopsies using microarrays. *J. Histochem. Cytochem.* **2011**, *59*, 146–157. [CrossRef] [PubMed]
22. Lin, C.Y.; Ström, A.; Vega, V.B.; Li Kong, S.; Li Yeo, A.; Thomsen, J.S.; Chan, W.C.; Doray, B.; Bangarusamy, D.K.; Ramasamy, A.; et al. Discovery of estrogen receptor α target genes and response elements in breast tumor cells. *Genome Biol.* **2004**, *5*, R66. [CrossRef] [PubMed]

23. Liu, M.C.; Pitcher, B.N.; Mardis, E.R.; Davies, S.R.; Friedman, P.N.; Snider, J.E.; Vickery, T.L.; Reed, J.P.; DeSchryver, K.; Singh, B.; et al. PAM50 gene signatures and breast cancer prognosis with adjuvant anthracycline- and taxane-based chemotherapy: Correlative analysis of C9741 (Alliance). *NPJ Breast Cancer* **2016**, *2*, 15023. [CrossRef] [PubMed]

24. Prat, A.; Bianchini, G.; Thomas, M.; Belousov, A.; Cheang, M.C.; Koehler, A.; Gomez, P.; Semiglazov, V.; Eiermann, W.; Tjulandin, S.; et al. Research-based PAM50 subtype predictor identifies higher responses and improved survival outcomes in HER2-positive breast cancer in the NOAH study. *Clin. Cancer Res.* **2014**, *20*, 511–521. [CrossRef]

25. Wishart, G.C.; Azzato, E.M.; Greenberg, D.C.; Rashbass, J.; Kearins, O.; Lawrence, G.; Caldas, C.; Pharoah, P.D.P. PREDICT: A new UK prognostic model that predicts survival following surgery for invasive breast cancer. *Breast Cancer Res.* **2010**, *12*, R1. [CrossRef]

26. Metzger-Filho, O.; Catteau, A.; Michiels, S.; Buyse, M.; Ignatiadis, M.; Saini, K.S.; de Azambuja, E.; Fasolo, V.; Naji, S.; Canon, J.L.; et al. Genomic Grade Index (GGI): Feasibility in Routine Practice and Impact on Treatment Decisions in Early Breast Cancer. *PLoS ONE* **2013**, *8*, e66848. [CrossRef]

27. Rhodes, A.; Jasani, B.; Balaton, A.; Miller, K. Immunohistochemical demonstration of oestrogen and progesterone receptors: Correlation of standards achieved on in house tumours with that achieved on external quality assessment material in over 150 laboratories from 26 countries. *J. Clin. Pathol.* **2000**, *53*, 292–301. [CrossRef]

28. Wolff, A.C.; Hammond, M.E.; Schwartz, J.N.; Hagerty, K.L.; Allred, D.C.; Cote, R.J.; Dowsett, M.; Fitzgibbons, P.L.; Hanna, W.M.; Langer, A.; et al. American Society of Clinical Oncology/College of American Pathologists Guideline Recommendations for Human Epidermal Growth Factor Receptor 2 Testing in Breast Cancer. *J. Clin. Oncol.* **2006**, *25*, 118–145.

29. Sparano, J.A.; Muss, H. Learning from big data: Are we undertreating older women with high-risk breast cancer? *NPJ Breast Cancer* **2016**, *2*, 16019. [CrossRef]

30. Harris, L.N.; Ismaila, N.; McShane, L.M.; Andre, F.; Collyar, D.E.; Gonzalez-Angulo, A.M.; Hammond, E.H.; Kuderer, N.M.; Liu, M.C.; Mennel, R.G.; et al. Use of Biomarkers to Guide Decisions on Adjuvant Systemic Therapy for Women with Early-Stage Invasive Breast Cancer: American Society of Clinical Oncology Clinical Practice Guideline. *J. Clin. Oncol.* **2016**, *34*, 1134–1150. [CrossRef]

31. Singer, C.F.; Tan, Y.Y.; Fitzal, F.; Steger, G.G.; Egle, D.; Reiner, A.; Rudas, M.; Moinfar, F.; Gruber, C.; Petru, E.; et al. Pathological Complete Response to Neoadjuvant Trastuzumab Is Dependent on HER2/CEP17 Ratio in HER2-Amplified Early Breast Cancer. *Clin. Cancer Res.* **2017**, *23*, 3676–3683. [CrossRef] [PubMed]

32. Harbeck, N.; Gnant, M. Breast cancer. *Lancet* **2016**, *389*, 1134–1150. [CrossRef]

33. Wirapati, P.; Sotiriou, C.; Kunkel, S.; Farmer, P.; Pradervand, S.; Haibe-Kains, B.; Desmedt, C.; Ignatiadis, M.; Sengstag, T.; Schütz, F.; et al. Meta-analysis of gene expression profiles in breast cancer: Toward a unified understanding of breast cancer subtyping and prognosis signatures. *Breast Cancer Res.* **2008**, *10*, R65. [CrossRef] [PubMed]

34. Hudis, C.A.; Barlow, W.E.; Costantino, J.P.; Gray, R.J.; Pritchard, K.I.; Chapman, J.A.W.; Sparano, J.A.; Hunsberger, S.; Enos, R.A.; Gelber, R.D.; et al. Proposal for standardized definitions for efficacy end points in adjuvant breast cancer trials: The STEEP system. *J. Clin. Oncol.* **2007**, *25*, 2127–2132. [CrossRef]

35. Regan, M.M.; Viale, G.; Mastropasqua, M.G.; Maiorano, E.; Golouh, R.; Carbone, A.; Brown, B.; Suurküla, M.; Langman, G.; Mazzucchelli, L.; et al. Re-evaluating Adjuvant Breast Cancer Trials: Assessing Hormone Receptor Status by Immunohistochemical Versus Extraction Assays. *J. Natl. Cancer Inst.* **2006**, *98*, 1571–1581. [CrossRef]

36. Kaufmann, M.; Pusztai, L.; Members, B.E.P. Use of standard markers and incorporation of molecular markers into breast cancer therapy: Consensus recommendations from an International Expert Panel. *Cancer* **2011**, *117*, 1575–1582. [CrossRef]

37. Kenn, M.; Cacsire Castillo-Tong, D.; Singer, C.F.; Karch, R.; Cibena, M.; Koelbl, H.; Schreiner, W. Decision theory for precision therapy of breast cancer. *Sci. Rep.* **2021**, *11*, 4233. [CrossRef]

38. Bartlett, J.M.; Campbell, F.M.; Ibrahim, M.; O'Grady, A.; Kay, E.; Faulkes, C.; Collins, N.; Starczynski, J.; Morgan, J.M.; Jasani, B.; et al. A UK NEQAS ISH multicenter ring study using the Ventana HER2 dual-color ISH assay. *Am. J. Clin. Pathol.* **2011**, *135*, 157–162. [CrossRef]

39. Lee, M.; Lee, C.S.; Tan, P.H. Hormone receptor expression in breast cancer: Postanalytical issues. *J. Clin. Pathol.* **2013**, *66*, 478–484. [CrossRef]

40. Hammond, M.E.; Hayes, D.F.; Wolff, A.C.; Mangu, P.B.; Temin, S. American Society of Clinical Oncology/College of American Pathologists Guideline Recommendations for Immunohistochemical Testing of Estrogen and Progesterone Receptors in Breast Cancer. *J. Oncol. Pract.* **2010**, *6*, 195–197. [CrossRef]

41. Wells, C.A.; Sloane, J.P.; Coleman, D.; Munt, C.; Amendoeira, I.; Apostolikas, N.; Bellocq, J.P.; Bianchi, S.; Boecker, W.; Bussolati, G.; et al. Consistency of staining and reporting of oestrogen receptor immunocytochemistry within the European Union-An inter-laboratory study. *Virchows Arch.* **2004**, *445*, 119–128. [CrossRef] [PubMed]

42. Laas, E.; Mallon, P.; Duhoux, F.P.; Hamidouche, A.; Rouzier, R.; Reyal, F. Low concordance between gene expression signatures in ER positive HER2 negative breast carcinoma could impair their clinical application. *PLoS ONE* **2016**, *11*, e0148957. [CrossRef] [PubMed]

43. Rakha, E.A.; Pinder, S.E.; Bartlett, J.M.; Ibrahim, M.; Starczynski, J.; Carder, P.J.; Provenzano, E.; Hanby, A.; Hales, S.; Lee, A.H.; et al. Updated UK Recommendations for HER2 assessment in breast cancer. *J. Clin. Pathol.* **2015**, *68*, 93–99. [CrossRef] [PubMed]

44. Allred, D.C.; Carlson, R.W.; Berry, D.A.; Burstein, H.J.; Edge, S.B.; Goldstein, L.J.; Gown, A.; Hammond, M.E.; Iglehart, J.D.; Moench, S.; et al. NCCN Task Force Report: Estrogen Receptor and Progesterone Receptor Testing in Breast Cancer by Immunohistochemistry. *J. Natl. Compr. Cancer Netw.* **2009**, *7* (Suppl. 6), S1–S21. [CrossRef] [PubMed]

45. Li, Q.; Eklund, A.C.; Juul, N.; Haibe-Kains, B.; Workman, C.T.; Richardson, A.L.; Szallasi, Z.; Swanton, C. Minimising immunohistochemical false negative ER classification using a complementary 23 gene expression signature of ER status. *PLoS ONE* **2010**, *5*, e15031. [CrossRef]

46. Bergqvist, J.; Ohd, J.F.; Smeds, J.; Klaar, S.; Isola, J.; Nordgren, H.; Elmberger, G.P.; Hellborg, H.; Bjohle, J.; Borg, A.L.; et al. Quantitative real-time PCR analysis and microarray-based RNA expression of HER2 in relation to outcome. *Ann. Oncol.* **2007**, *18*, 845–850. [CrossRef]

47. Chen, X.; Li, J.; Gray, W.H.; Lehmann, B.D.; Bauer, J.A.; Shyr, Y.; Pietenpol, J.A. TNBCtype: A subtyping tool for triple-negative breast cancer. *Cancer Inform.* **2012**, *11*, 147–156. [CrossRef]

48. Gong, Y.; Yan, K.; Lin, F.; Anderson, K.; Sotiriou, C.; Andre, F.; Holmes, F.A.; Valero, V.; Booser, D.; Pippen, J.; et al. Determination of oestrogen-receptor status and ERBB2 status of breast carcinoma: A gene-expression profiling study. *Lancet Oncol.* **2007**, *8*, 203–211. [CrossRef]

49. Lopez, F.J.; Cuadros, M.; Cano, C.; Concha, A.; Blanco, A. Biomedical application of fuzzy association rules for identifying breast cancer biomarkers. *Med. Biol. Eng. Comput.* **2012**, *50*, 981–990. [CrossRef]

50. Owzar, K.; Barry, W.T.; Jung, S.H.; Sohn, I.; George, S.L. Statistical Challenges in Pre-Processing in Microarray Experiments in Cancer. *Clin. Cancer Res.* **2008**, *14*, 5959–5966. [CrossRef]

51. Wu, Z. A Review of Statistical Methods for Preprocessing Oligonucleotide Microarrays. *Stat. Methods Med. Res.* **2009**, *18*, 533–541. [PubMed]

52. Wu, Z.; Irizarry, R.A. Preprocessing of oligonucleotide array data. *Nat. Biotechnol.* **2004**, *22*, 656–658. [CrossRef] [PubMed]

53. Wu, Z.; Irizarry, R.A.; Gentleman, R.; Martinez-Murillo, F.; Spencer, F. A Model-Based Background Adjustment for Oligonucleotide Expression Arrays. *J. Am. Stat. Assoc.* **2004**, *99*, 909–917. [CrossRef]

54. Zhang, B.; Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* **2005**, *4*, 17. [CrossRef] [PubMed]

55. Ritchie, M.E.; Phipson, B.; Wu, D.; Hu, Y.; Law, C.W.; Shi, W.; Smyth, G.K. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **2015**, *43*, e47. [CrossRef] [PubMed]

56. Wu, Z.; Irizarry, R.A. A Statistical Framework for the Analysis of Microarray Probe-Level Data. *Ann. Appl. Stat.* **2007**, *1*, 333–357. [CrossRef]

57. Bolstad, B.M.; Irizarry, R.A.; Astrand, M.; Speed, T.P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **2003**, *19*, 185–193. [CrossRef]

58. Kenn, M.; Cacsire Castillo-Tong, D.; Singer, C.F.; Cibena, M.; Kölbl, H.; Schreiner, W. Co-expressed genes enhance precision of receptor status identification in breast cancer patients. *Breast Cancer Res. Treat.* **2018**, *172*, 313–326. [CrossRef]

59. Kenn, M.; Schlangen, K.; Cacsire Castillo-Tong, D.; Singer, C.F.; Cibena, M.; Koelbl, H.; Schreiner, W. Gene expression information improves reliability of receptor status in breast cancer patients. *Oncotarget* **2017**, *8*, 77341–77359. [CrossRef]

60. Gordon, J.; Shortliffe, E.H. The Dempster-Shafer Theory of Evidence. In *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*; Buchanan, B.G., Shortliffe, E.H., Eds.; Addison-Wesley Publishing Company: Boston, MA, USA, 1984; pp. 832–838.

61. Högger, A. *Dempster Shafer Sensor Fusion for Autonomously Driving Vehicles: Association Free Tracking of Dynamic Objects*; KTH Royal Institut of Technology School of Electrical Engineering: Stockholm, Sweden, 2016.

62. Feng, R.; Zhang, G.; Cheng, B. An on-board system for detecting driver drowsiness based on multi-sensor data fusion using Dempster-Shafer theory. In Proceedings of the 2009 International Conference on Networking, Sensing and Control, Okayama, Japan, 26–29 March 2009; pp. 897–902.

63. Jugade, S.C.; Victorino, A.C. Grid based Estimation of Decision Uncertainty of Autonomous Driving Systems using Belief Function theory. *IFAC-PapersOnLine* **2018**, *51*, 261–266. [CrossRef]

64. Lu, C.; Wang, S.; Wang, X. A multi-source information fusion fault diagnosis for aviation hydraulic pump based on the new evidence similarity distance. *Aerosp. Sci. Technol.* **2017**, *71*, 392–401. [CrossRef]

65. Yang, J.; Huang, H.-Z.; He, L.-P.; Zhu, S.-P.; Wen, D. Risk evaluation in failure mode and effects analysis of aircraft turbine rotor blades using Dempster–Shafer evidence theory under uncertainty. *Eng. Fail. Anal.* **2011**, *18*, 2084–2092. [CrossRef]

66. Fontani, M.; Bianchi, T.; De Rosa, A.; Piva, A.; Barni, M. A Framework for Decision Fusion in Image Forensics Based on Dempster–Shafer Theory of Evidence. *IEEE Trans. Inf. Forensics Secur.* **2013**, *8*, 593–607. [CrossRef]

67. Chandana, S.; Leung, H.; Trpkov, K. Staging of prostate cancer using automatic feature selection, sampling and Dempster-Shafer fusion. *Cancer Inform.* **2009**, *7*, 57–73. [CrossRef]

68. Raza, M.; Gondal, I.; Green, D.; Coppel, R.L. Fusion of FNA-cytology and gene-expression data using Dempster-Shafer Theory of evidence to predict breast cancer tumors. *Bioinformation* **2006**, *1*, 170–175. [CrossRef]

69. Denœux, T. 40 years of Dempster–Shafer theory. *Int. J. Approx. Reason.* **2016**, *79*, 1–6. [CrossRef]

70. Denœux, T.; Smets, P. Classification using belief functions: Relationship between case-based and model-based approaches. *IEEE Trans. Syst. Man. Cybern. B* **2006**, *36*, 1395–1406. [CrossRef]

71. Parzen, E. On Estimation of a Probability Density Function and Mode. *Ann. Math. Stat.* **1962**, *33*, 1065–1076. [CrossRef]

72. Silverman, B.W. Using Kernel Density Estimates to Investigate Multimodality. *J. R. Stat. Soc. Ser. B* **1981**, *43*, 97–99. [CrossRef]
73. Rosenblatt, M. Remarks on Some Nonparametric Estimates of a Density Function. *Ann. Math. Stat.* **1956**, *27*, 832–837. [CrossRef]
74. Denœux, T. Decision-making with belief functions: A review. *Int. J. Approx. Reason.* **2019**, *109*, 87–110. [CrossRef]
75. Denœux, T. Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence. *Artif. Intell.* **2008**, *172*, 234–264. [CrossRef]
76. Shafer, G. *A Mathematical Theory of Evidence*; Princeton University Press: Princeton, NJ, USA, 1976.
77. Yang, J.B.; Xu, D.L. Evidential reasoning rule for evidence combination. *Artif. Intell.* **2013**, *205*, 1–29. [CrossRef]
78. Yager, R.R. On the dempster-shafer framework and new combination rules. *Inf. Sci.* **1987**, *41*, 93–137. [CrossRef]
79. Strang, G. *Introduction to Linear Algebra*; Wellesley-Cambridge Press: Wellesley, MA, USA, 2016.
80. Tapia, J.F.D.; Tan, R.R. Ternary Diagram for Visualizing Epidemic Progression. *Process Integr. Optim. Sustain.* **2021**, *5*, 687–691. [CrossRef]
81. Fleiss, J.L.; Levin, B.; Paik, M.C. *Statistical Methods for Rates and Proportions*, 3rd ed.; Wiley-Interscience: Hoboken, NI, USA, 2003; pp. 1–800.
82. Smets, P. The Combination of Evidence in the Transferable Belief Model. *IEEE Trans. Pattern Anal. Mach. Intell.* **1990**, *12*, 447–458. [CrossRef]
83. Smets, P.; Kennes, R. The transferable belief model. *Artif. Intell.* **1994**, *66*, 191–234. [CrossRef]
84. Dubois, D.; Prade, H. The logical view of conditioning and its application to possibility and evidence theories. *Int. J. Approx. Reason.* **1990**, *4*, 23–46. [CrossRef]
85. Smarandache, F.; Dezert, J. Proportional Conflict Redistribution Rules for Information Fusion. *Adv. Appl. DSmT Inf. Fusion* **2006**, *2*, 3–68.
86. Chen, L.; Diao, L.; Sang, J. Weighted Evidence Combination Rule Based on Evidence Distance and Uncertainty Measure: An Application in Fault Diagnosis. *Math. Probl. Eng.* **2018**, *2018*, 5858272. [CrossRef]
87. Sentz, K.; Ferson, S. Combination of Evidence in Dempster-Shafer Theory. In *Sandia Report*; Sandia National Laboratories: Albuquerque, NM, USA, 2002; Volume Sand 2002-0835.
88. Barrett, T.; Edgar, R. Mining microarray data at NCBI's Gene Expression Omnibus (GEO). *Methods Mol. Biol.* **2006**, *338*, 175–190. [PubMed]
89. Edgar, R.; Barrett, T. NCBI GEO standards and services for microarray data. *Nat. Biotechnol.* **2006**, *24*, 1471–1472. [CrossRef] [PubMed]
90. Irizarry, R.A.; Bolstad, B.M.; Collin, F.; Cope, L.M.; Hobbs, B.; Speed, T.P. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.* **2003**, *31*, e15. [CrossRef] [PubMed]
91. Rung, J.; Brazma, A. Reuse of public genome-wide gene expression data. *Nat. Rev. Genet.* **2013**, *14*, 89–99. [CrossRef] [PubMed]
92. Van Vliet, M.H.; Reyal, F.; Horlings, H.M.; van de Vijver, M.J.; Reinders, M.J.; Wessels, L.F. Pooling breast cancer datasets has a synergetic effect on classification performance and improves signature stability. *BMC Genom.* **2008**, *9*, 375. [CrossRef]
93. Bolstad, B.M. RMAExpress Users Guide. Available online: https://rmaexpress.bmbolstad.com/RMAExpress_UsersGuide.pdf (accessed on 26 March 2022).
94. McCall, M.N.; Bolstad, B.M.; Irizarry, R.A. Frozen robust multiarray analysis (fRMA). *Biostatistics* **2010**, *11*, 242–253. [CrossRef]
95. Gautier, L.; Cope, L.; Bolstad, B.M.; Irizarry, R.A. Affy—analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **2004**, *20*, 307–315. [CrossRef]
96. Stafford, P. *Methods in Microarray Normalization*; CRC Press: Boca Raton, FL, USA, 2008.
97. McCall, M.N.; Jaffee, H.A.; Irizarry, R.A. fRMA ST: Frozen robust multiarray analysis for Affymetrix Exon and Gene ST arrays. *Bioinformatics* **2012**, *28*, 3153–3154. [CrossRef]
98. Bolstad, B. Background and Normalization: Investigating the Effects of Preprocessing on Gene Expression Estimates. Available online: http://bmbolstad.com/stuff/BAUGM.pdf (accessed on 17 September 2021).
99. Kenn, M.; Cacsire Castillo-Tong, D.; Singer, C.F.; Cibena, M.; Kölbl, H.; Schreiner, W. Microarray Normalization Revisited for Reproducible Breast Cancer Biomarkers. *Biomed. Res. Int.* **2020**, *2020*, 1363827. [CrossRef]
100. Luo, J.; Schumacher, M.; Scherer, A.; Sanoudou, D.; Megherbi, D.; Davison, T.; Shi, T.; Tong, W.; Shi, L.; Hong, H.; et al. A comparison of batch effect removal methods for enhancement of prediction performance using MAQC-II microarray gene expression data. *Pharm. J.* **2010**, *10*, 278–291. [CrossRef]
101. Leek, J.T.; Johnson, W.E.; Parker, H.S.; Jaffe, A.E.; Storey, J.D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **2012**, *28*, 882–883. [CrossRef] [PubMed]
102. Müller, C.; Schillert, A.; Röthemeier, C.; Tregouet, D.A.; Proust, C.; Binder, H.; Pfeiffer, N.; Beutel, M.; Lackner, K.J.; Schnabel, R.B.; et al. Removing Batch Effects from Longitudinal Gene Expression-Quantile Normalization Plus ComBat as Best Approach for Microarray Transcriptome Data. *PLoS ONE* **2016**, *11*, e0156594. [CrossRef] [PubMed]
103. Johnson, W.E.; Li, C.; Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **2007**, *8*, 118–127. [CrossRef] [PubMed]
104. Leek, J.T.; Storey, J.D. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet.* **2007**, *3*, 1724–1735. [CrossRef] [PubMed]

105. Leek, J.T.; Johnson, W.E.; Parker, H.S.; Fertig, E.J.; Jaffe, A.E.; Storey, J.D.; Zhang, Y.; Torres, L.C. sva: Surrogate Variable Analysis. 2017. Available online: https://bioconductor.org/packages/release/bioc/manuals/sva/man/sva.pdf (accessed on 26 March 2022).

106. Parker, H.S.; Corrada Bravo, H.; Leek, J.T. Removing batch effects for prediction problems with frozen surrogate variable analysis. *PeerJ* **2014**, *2*, e561. [CrossRef] [PubMed]

107. Ikeda, K.; Horie-Inoue, K.; Inoue, S. Identification of estrogen-responsive genes based on the DNA binding properties of estrogen receptors using high-throughput sequencing technology. *Acta Pharmacol. Sin.* **2015**, *36*, 24–31. [CrossRef]

108. McCullagh, P.; Nelder, J.A. Generalized Linear Models. In *Monographs on Statistics and Applied Probability*, 2nd ed.; Chapman & Hall: London, UK; CRC: New York, NY, USA, 1989; Volume 37.